On solutions of Kolmogorov's equations for non-homogeneous jump Markov processes & Sufficiency of Markov policies in continuous-time Markov decision processes

A Dissertation Presented

by

Manasa Mandava

to

The Graduate School

in Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

in

Applied Mathematics and Statistics

Stony Brook University

May 2015

Stony Brook University

The Graduate School

Manasa Mandava

We, the dissertation committee for the above candidate for the

Doctor of Philosophy degree, hereby recommend

acceptance of this dissertation.

Eugene Feinberg - Dissertation Advisor
Distinguished Professor, Department of Applied Mathematics and Statistics

Sevtlozar Rachev - Chairperson of Defense
Professor, Department of Applied Mathematics and Statistics

Jiaqiao Hu
Associate Professor, Department of Applied Mathematics and Statistics

Leon Takhtajan
Professor, Department of Mathematics

This dissertation is accepted by the Graduate School.

Charles Taber
Dean of the Graduate School

Abstract of the Dissertation

On solutions of Kolmogorov's equations for non-homogeneous jump Markov processes & Sufficiency of Markov policies in continuous-time Markov decision processes

by

Manasa Mandava

Doctor of Philosophy

in

Applied Mathematics and Statistics

Stony Brook University

2015

A basic fact in the theory of Discrete-time Markov Decision Processes is that for any policy there exists a Markov policy with the same marginal state-action distributions. This fact implies that the study of control problems with multiple criteria and constraints that are determined by marginal distribution (for e.g. expected total discounted and non-discounted costs, average cost per unit time) can be restricted to the set of Markov policies. This dissertation presents a similar result for Continuous-Time Markov Decision Processes (CTMDPs).

In CTMDPs with Borel state and action spaces, unbounded transition and cost rates, for an arbitrary policy, we construct a Markov policy such that the marginal distribution on the state-action pairs is the same for both the policies. This fact implies that the expected cost rates at each time instant are equal for these two policies. Thus, the constructed Markov policy performs equally

to the original policy for problems with multiple criteria and constraints that are determined by marginal distribution. The proof consists of two major steps: The first step describes the properties of solutions to Kolmogorov's equations for jump Markov processes. In particular, for given transition intensities, the three approaches to construct a jump Markov process: (i) via the compensator of the random measure of a multivariate point process, (ii) as a minimal solution of Kolmogorov's backward equation, and (iii) as a minimal solution of Kolmogorov's forward equation define the same transition function. If the jump Markov process associated with the transition function has no accumulation points, then it is the unique solution of both Kolmogorov's equations. The second step applies these results to CTMDPs and establishes that the marginal distribution on the state for both the policies satisfy Kolmogorov's forward equation defined by the Markov policy. This fact immediately implies that the marginal distributions on the state for both the policies coincide if the transition intensities corresponding to the Markov policy are bounded. In the general case, it is possible to consider a sequence of policies with bounded transition intensities and that converge to the original policy. The proof for the general case follows from these approximations.

Table of Contents

List of Figures

Acknowledgements

I thank my advisor Distinguished Professor Eugene Feinberg for his trust, patience, guidance and mentoring. The past five years of working with him have been the most fruitful and enriching period during my entire graduate study at Stony Brook University. I shall always treasure this experience and the time spent in our discussions. In every aspect, he has made me better than I was before.

I thank Professor Albert N. Shiryaev for his warm encouragement and helpful advice. I would also like to thank Professor Jiaqiao Hu, for his constant support during my academic stay in Stony Brook, and the rest of my thesis committee: Professor Sevtlozar Rachev and Professor Leon Takhtajan for their time, encouragement, and insightful comments.

This work was partially supported by NSF grant CMMI-0928490 and CMMI-1335296 and it is much appreciated. I would like to thank all my friends: Muqi, Jefferson, Katie, Swetha, Archana, and Lingling for the stimulating discussions, warm friendships, and constant encouragement. Last but not least, I thank my Husband Bharat Marella, my parents Mandava Ramalingeswara Prasad and Venigalla Bharathi, and my brother Mandava Vamsi Krishna for their caring, understanding, and unconditional support at all times. I dedicate this thesis to my advisor and family.

Chapter 1

Overview

1.1 Brief description of the results and related works

In this dissertation, we study the properties of solutions of Kolmogorov's backward and forward equations for non-homogeneous jump Markov processes and apply the results on Kolmogorov's forward equation to answer a fundamental control problem in Continuous-time Markov decision processes (CTMDPs). Both Kolmogorov's equations and CTMDPs have broad range of applications. For instance, Kolmogorov's equations are widely used in such applications as population growth, epidemics, queues, manufacturing systems, etc, and CTMDPs are widely used in such applications as inventory control, airline management, machine maintenance, smart grids, health care services, etc. In this section we give a preview of the results, and the background and earlier works to which they are related. We address existence and uniqueness of solutions of Kolmogorov's equations (Section 1.1.1) and sufficiency of Markov policies (or decision rules that depend only on the current state and time) to study CTMDPs (Section 1.1.2).

1.1.1 Kolmogorov's equations for non-homogeneous jump Markov processes

Our work answers the following questions, which are important for the theory of stochastic processes and their applications: (i) how a non-homogeneous jump Markov process can be defined for given transition intensities, called Q-functions, and (ii) how can its transition function be found as a solution of Kolmogorov's backward and forward equations? We answer these questions for measurable Q-functions when the jump Markov process takes values on a Borel state space.

The common approach used in the literature to address the first question is to construct a transition function for a given Q-function and show the jump property of the Markov process defined by the constructed transition function and an initial distribution by using the analytical properties of the transition function; see, for e.g., Anderson [1] and Reuter [30] for Q-functions that do not depend on time parameter and countable state spaces, Doob [5] for Q-functions that do not depend on time parameter and Borel state spaces. Our construction of the jump Markov process defined by the measurable Q-function and initial distribution is based on Jacod's theorem [17]. This approach to the construction of jump Markov process is motivated by the application of the theory of jump Markov processes to study CTMDPs controlled by a Markov policy.

The transition function of a non-homogenous jump Markov process as solution of Kolmogorov's equations was first studied by Feller [10]. Feller considered continuous Q-functions and provided the explicit formulae for the transition function and showed that it satisfies both Kolmogorov's backward and forward equations. In general, these equations can have multiple solutions. For Q-functions that do not depend on time parameter, Doob [5, Chap. 6] provided an explicit construction for multiple transition functions satisfying Kolmogorov's backward equation, and Kendall [22], Kendall and Reuter [23], and Reuter [30] gave examples with non-unique solutions to both the equations. Considering measurable Q-functions and a countable state space, Ye et al. [34] constructed the transition function satisfying both Kolmogorov's backward and forward equations. All of the above mentioned work on solutions of Kolmogorov's equations considered Q-functions satisfying certain boundedness conditions. For Borel state spaces, we consider more general classes of unbounded Q-functions and obtain the transition function of the jump Markov process as the minimal non-negative solution of Kolmogorov's backward and forward equations and provide a sufficient condition for its uniqueness.

Relation to stochastic process defined by a Markov policy in CTMDPs

Feller's [10] results on Kolmogorov's equations for non-homogeneous jump Markov processes are broadly used in the literature on CTMDPs to study the jump Markov process defined by a Markov policy, and this leads to the unnecessary assumption that decisions/actions depend continuously on time; see, e.g., Guo and Rieder [13, Definition 2.2]. For countable state problems, the results of Ye et al. [34] removed the necessity to assume this continuity. Our results on Kolmogorov's equations imply that this continuity assumption is unnecessary for CTMDPs with Borel state spaces. They also unify a body of research on jump Markov processes that can be traced back to the works by Feller [10] and Jacod [17]. Given an initial state, the non-homogeneous jump Markov process defined by a Markov policy is commonly constructed using one of the following two ways:

(i) Based on Jacod's [17] theorem via the compensator of the random measure of the multivariate point process; Kitaev [24], Kitaev and Rykov [25, Section 4.6], Feinberg [6, 7], Guo and Piunovskiy [12].

(ii) As the minimal non-negative solution of Kolmogorov's forward equation; Miller [28], Kakumanu [19], Guo and Hernández-Lerma [11].

The second approach is commonly used in the literature to study the jump Markov process defined by a Markov policy via its transition function, including in the monograph by Guo and Hernández-Lerma [11]. However, the first approach is used to construct the jump process associated with any policy, and in particular, with a Markov policy. Our results imply that for Markov policies these two constructions are equivalent for problems with Borel state spaces.

1.1.2 Sufficiency of Markov policies to study CTMDPs

In 1980, Yushkevich [35] introduced past dependent policies and constructed the jump process associated with them. Later, Kitaev [24] gave an equivalent construction for the jump process associated with a past dependent policy using the results by Jacod [17]. Since then, even though it is possible to consider past dependent policies for CTMDPs, most of the existing facts such as optimality of certain policies are established within the class of Markov policies; see Guo and Hernández-Lerma [11]. In this dissertation, we show that it is valid to restrict the study of CTMDPs to the class of Markov policies, and therefore, many of the previously existing results within the class of Markov policies hold within the class of all policies. A similar result on the sufficiency of Markov policies to study discrete-time Markov decision processes (DTMDPs) was given by Derman and Strauch [4].

Given any policy, we construct a Markov policy such that the marginal distribution on the state-action pairs at any time instant is the same for both the policies. This immediately implies that the expected cost rates at each time instant are equal for these two policies. Thus, the corresponding Markov policy performs equally or better than the original policy for problems with expected total discounted and non-discounted costs as well as with average costs per unit time. This is also true for problems with multiple criteria. We consider the state and action sets as standard Borel spaces, and the transition and cost rates may not be bounded. Thus, the results in this thesis are applicable to a wide class of problems. Two such important applications are noted below:
(i) Queuing control: state space is countable, cost and transition rate functions may not be bounded from above.
(ii) Inventory control: state and action spaces are uncountable, cost functions may not be bounded from above.

1.2 Contents and Organization

This thesis consists of two parts: Chapter 2 on Kolmogorov's equations for non-homogeneous jump Markov processes, and Chapters 3, 4 on CTMDPs.

Chapter 2: Kolmogorov's equations for non-homogeneous jump Markov processes
Chapter 2 concerns the construction of non-homogeneous jump Markov process on a general state space. For a given measurable Q-function satisfying certain boundedness conditions, it presents three equivalent ways to construct the non-homogeneous jump Markov process: (i) via the compensator defined by the Q-function and initial distribution, (ii) as a minimal non-negative solution of Kolmogorov's backward equation, and (iii) as a minimal non-negative solution of Kolmogorov's forward equation. The results on Kolmogorov's forward equation in this chapter lay the foundation for the proof of the main result of this dissertation presented in Chapter 4.

Chapter 3: Continuous-time Markov decision processes
In Chapter 3 we give a brief introduction of CTMDPs with general state and action spaces, including definition of the control model, induced stochastic process, and optimality criteria we are concerned with.

Chapter 4: Sufficiency of Markov policies in CTMDPs
Chapter 4 considers general state and action space CTMDPs with unbounded transition and cost rates and presents one of the main results of this thesis: given an initial probability measure on the state space, for any policy there exists a Markov policy that has the same marginal distribution on the state-action pairs at any time instant. It shows that if the optimality criteria depends only on the

marginal distribution, like the expected total discounted and non-discounted costs, average costs per unit time etc., one can restrict the search for optimal policies to the class of Markov policies. For completeness, we present the analogous result that is established for DTMDPs in Appendix 5.

Chapter 2

Kolmogorov's equations for Non-homogeneous jump Markov process

2.1 Introduction

This chapter answers the following questions, which are important for the theory of stochastic processes and their applications: how a non-homogeneous jump Markov process can be defined for given transition intensities, called Q-functions, and how can its transition probabilities be found as a solution of Kolmogorov's backward and forward equations? First we present a few definitions.

2.1.1 Basic definitions

For a topological space S, its Borel σ-field (the σ-field generated by open subsets of S) is always denoted by $\mathfrak{B}(S)$, and the sets in $\mathfrak{B}(S)$ are called *Borel subsets* of S. Let $\mathbb{R}$ be the real line endowed with the Euclidean metric. A topological space $(S,\mathfrak{B}(S))$ is called a *standard Borel space* if there exists a bijection f from $(S,\mathfrak{B}(S))$ to a Borel subset of $\mathbb{R}$ such that the mappings f and f^{-1} are measurable. In this dissertation, measurability and Borel measurability are used synonymously. Let $(X,\mathfrak{B}(X))$ be a standard Borel space (called the state space), and $[T_0,T_1[$ be a finite or an infinite interval of $\mathbb{R}_+ := [0,\infty[$.

Definition of jump Markov process and transition function:
A stochastic process $\{\mathbb{X}_t : t\in[T_0,T_1[\}$ with values in X, defined on the probability space $(\Omega,\mathscr{F},\mathbb{P})$ and adapted to the filtration $\{\mathscr{F}_t\}_{t\in[T_0,T_1[}$, is called *Markov* if $\mathbb{P}(\mathbb{X}_t\in B\mid\mathscr{F}_u)=\mathbb{P}(\mathbb{X}_t\in B\mid\mathbb{X}_u)$, $\mathbb{P}-a.s.$ for all $u,t\in[T_0,T_1[$ with $u<t$ and for all $B\in\mathfrak{B}(X)$. In addition, if the Markov process is a jump process, that is, if each sample path of the process is a right-continuous piecewise constant function in t that has a countable number of discontinuity points on $t\in[T_0,T_1[$, then the Markov process is called a *jump Markov process*.

A function $P(u,x;t,B)$, where $u,t\in[T_0,T_1[$, $u<t$, $x\in X$, and $B\in\mathfrak{B}(X)$, is called a *transition function* if it takes values in $[0,1]$ and satisfies the following properties:

(i) For all u,x,t the function $P(u,x;t,\cdot)$ is a measure on $(X,\mathfrak{B}(X))$.

(ii) For all B the function $P(u,x;t,B)$ is Borel measurable in (u,x,t).

(iii) $P(u,x;t,B)$ satisfies the Chapman-Kolmogorov equation

$$P(u,x;t,B) = \int_X P(s,y;t,B)P(u,x;s,dy), \qquad u < s < t. \tag{2.1}$$

A transition function P is called *regular* if $P(u,x;t,X) = 1$ for all u,x,t in the domain of P.

Each Markov process has a transition function P such that

$$\mathbb{P}(\mathbb{X}_t \in B \mid \mathbb{X}_u) = P(u,\mathbb{X}_u;t,B), \qquad \mathbb{P} - a.s, \tag{2.2}$$

and any probability measure γ on X and transition function P define a unique Markov process $\{\mathbb{X}_t, t \in [T_0,T_1[\}$ such that (2.2) holds; see Kuznetsov [26]. Thus, one can equivalently define a Markov process via the probability measure $\mathbb{P}$ or its transition function P.

Definition of Q-function:
A function $q(x,t,B)$, where $x \in X$, $t \in [T_0,T_1[$, and $B \in \mathfrak{B}(X)$, is called a *Q-function* if it satisfies the following properties:

(i) for all x,t the function $q(x,t,\cdot)$ is a signed measure on $(X,\mathfrak{B}(X))$ such that $q(x,t,X) \le 0$ and $0 \le q(x,t,B \setminus \{x\}) < \infty$ for all $B \in \mathfrak{B}(X)$;

(ii) for all B the function $q(x,t,B)$ is measurable in (x,t).

Let $q(x,t) := -q(x,t,\{x\})$ for all $x \in X$ and $t \in [T_0,T_1[$. In addition to properties (i) and (ii), if $q(x,t,X) = 0$ for all x,t, then the Q-function q is called *conservative*. Note that any Q-function can be transformed into a conservative Q-function by adding an absorbing state $\bar{x}$ to X with $q(x,t,\{\bar{x}\}) := -q(x,t,X)$, $q(\bar{x},t,X) := 0$, and $q(\bar{x},t,\{\bar{x}\}) := 0$, where $x \in X$ and $t \in [T_0,T_1[$. To simplify the presentation, we always assume that q is conservative. In Subsection 2.6.1, we explain how the main formulations change when the Q-function q is not conservative. A Q-function q is called *continuous* if it is continuous in $t \in [T_0,T_1[$.

A classical approach to the study of jump Markov processes is via the compensator of the random measure of a multlivariate point process. A conservative Q-function can be used to construct a predictable random measure. According to Jacod [17, Theorem 3.6], an initial state distribution and a predictable random measure define uniquely a multivariate point process. In this chapter, we show that the stochastic process associated with the multivariate point process defined by a conservative Q-function q and an initial state distribution is a jump Markov process, describe its transition function $\bar{P}$, and that $\bar{P}$ is the minimal non-negative solution to both Kolmogorov's backward and forward equations. The first study of jump Markov processes defined by the Q-function via Kolmogorov's equations was first undertaken by Feller [10].

2.2 Boundedness assumptions and description of the main results

In this section, we describe the boundedness assumptions on Q-functions and provide the general description of the results of this chapter. Let $\bar{q}(x) := \sup_{t \in [T_0,T_1[} q(x,t)$ for $x \in X$. Consider the following assumptions of boundedness of q in t. Feller [10] studied Kolmogorov's equations for continuous Q-functions under the following assumption.

Assumption 2.2.1 (Feller's assumption)**.** *There exists a sequence of measurable subsets* $\{B_n, n = 1,2,\ldots\}$ *of* X *such that* $\sup_{x\in B_n} \bar{q}(x) < n$ *for all* $n = 1,2,\ldots$ *and* $B_n \uparrow X$ *as* $n \to \infty$*.*

In this chapter, we consider the following assumptions.

Assumption 2.2.2 (Boundedness of q)**.** $\bar{q}(x) < \infty$ *for each* $x \in X$*.*

For $n = 1,2,\ldots$, consider the functions U_n from $(X \times [T_0,T_1])$ to $[0,\infty]$ defined by

$$U_n(x,t) := \int_{T_0}^{t} \mathbf{I}\{q(x,s) \geq n\}ds, \qquad x \in X, t \in [T_0,T_1] \ . \tag{2.3}$$

For each $t \in [T_0,T_1]$, let $X_n^t, n = 1,2,\ldots$, be the subsets of X such that

$$X_n^t = \{x \in X : U_n(x,t) = 0\}, \qquad n = 1,2,\ldots \ . \tag{2.4}$$

Since the functions $U_n(x,t)$ are measurable, the sets X_n^t are measurable subsets of X. Observe that $X_n^t \subseteq X_{n+1}^t, n = 1,2,\ldots$.

Assumption 2.2.3 (Almost everywhere local boundedness of q)**.** $X_n^t \uparrow X$ *as* $n \to \infty$ *for each* $t \in [T_0,T_1[$.

Assumption 2.2.4 (Local $\mathscr{L}^1$ boundedness of q)**.** *For all* $x \in X$*, the integral* $\int_{T_0}^{t} q(x,s)ds < \infty$ *for each* $t \in [T_0,T_1[$.

The following lemma compares Assumptions 2.2.1–2.2.4.

Lemma 2.2.1. *The following statements hold for a measurable Q-function* q :
(i) Assumptions 2.2.1 and 2.2.2 are equivalent;
(ii) Assumption 2.2.2 implies Assumption 2.2.3;
(iii) Assumption 2.2.3 implies Assumption 2.2.4.

Proof. (i) Let $\{B_n, n = 1,2,\ldots\}$ be a sequence of Borel subsets of X satisfying the properties stated in Assumption 2.2.1. Then for each $x \in X$ there exists an $n \in \{1,2,\ldots\}$ such that $x \in B_n$ and therefore $\bar{q}(x) < n$. Thus, Assumption 2.2.1 implies Assumption 2.2.2. To prove that Assumption 2.2.2 implies Assumption 2.2.1, define $C_n := \{x \in X : \bar{q}(x) \geq n\}, n = 1,2,\ldots$. Since $C_n = proj_X(\{(x,t) \in (X \times \mathbb{R}_+) \mid q(x,t) \geq n\})$ are projections of Borel sets, the sets C_n are analytic, $n = 1,2,\ldots$; see Bertsekas and Shreve [2, Proposition 7.39]. In addition, Assumption 2.2.2 implies that $\bigcap_{n=1}^{\infty} C_n = \emptyset$. Thus, in view of the Novikov separation theorem, Kechris [21, Theorem 28.5], there exist Borel subsets $Z_n, n = 1,2,\ldots$, of X such that $C_n \subseteq Z_n$ and $\bigcap_{n=1}^{\infty} Z_n = \emptyset$. This fact implies that $Z_n^c \subseteq C_n^c$ and $\bigcup_{n=1}^{\infty} Z_n^c = X$, where the sets Z_n^c and C_n^c are compliments of the sets Z_n and C_n, respectively. Let $B_n := \cup_{m=1}^{n} Z_m^c$ for all $n = 1,2,\ldots$. The Borel sets B_n, $n = 1,2,\ldots$, satisfy the properties stated in Assumption 2.2.1.

(ii) Let Assumption 2.2.2 hold. In view of Lemma 2.2.1(i), consider sets B_n, $n = 1,2,\dots,$ whose existence is stated in Assumption 2.2.1. Then $B_n \subseteq X_n^t$ and $B_n \uparrow X$, $n = 1,2,\dots,$ for each $t \in [T_0,T_1]$. Therefore, $X_n^t \uparrow X$ as $n \to \infty$ for each $t \in [T_0,T_1]$. Thus, Assumption 2.2.3 holds.

(iii) Under Assumption 2.2.3, for each $x \in X$ and $t \in [T_0,T_1[$ there exists an $n \in \{1,2,\dots\}$ such that $U_n(x,t) = 0$. That is, $\mu(w \in [T_0,t[: q(x,s) \geq n) = 0$, where μ is the Lebesgue measure on $\mathbb{R}_+$. This immediately implies that Assumption 2.2.4 holds. $\square$

In Section 2.3 we show in Theorem 2.3.2 that under Assumption 2.2.4 the compensator of a random measure defined by a Q-function defines a jump Markov process whose transition function $\bar{P}$ is described in (2.19). The function $\bar{P}$ was introduced in Feller [10]. Theorem 2.4.1 in Section 2.4 states that under Assumption 2.2.4 the transition function $\bar{P}$ is the minimal non-negative solution of Kolmogorov's backward equation, and Theorem 2.5.1 in Section 2.5 states that under Assumption 2.2.3 the transition function $\bar{P}$ is the minimal non-negative solution of Kolmogorov's forward equation. In Section 2.6, we consider non-conservative Q-functions and weaker boundedness conditions, Assumptions 2.6.1 and 2.6.2, than those presented in this section, and discuss how the main results of this paper change in these two scenarios. We also present some of the results of this chapter under Assumption 2.2.2 as corollaries. Assumption 2.2.2 means that jump intensities are bounded at each state for the time horizon $[T_0,T_1[$, and this assumption is natural for continuous-time Markov decision processes (CTMDPs). Hence, our results under Assumption 2.2.2 are useful for applying the results of this chapter to CTMDPs; see Chapter 4.

2.3 Relation between jump Markov processes and Q-functions

In this section, we show that a Q-function satisfying Assumption 2.2.4 defines a transition function for a jump Markov process.

Let $x_\infty \notin X$ be an isolated point adjoined to the space X. Denote $\bar{X} = X \cup \{x_\infty\}$ and $\bar{T} =]T_0,T_1]$. Consider the Borel σ-field $\mathfrak{B}(\bar{X}) = \sigma(\mathfrak{B}(X),\{x_\infty\})$ on $\bar{X}$, which is the minimal σ-field containing $\mathfrak{B}(X)$ and $\{x_\infty\}$. Let $(\bar{X} \times \bar{T})^\infty$ be the set of all sequences $(x_0,t_1,x_1,t_2,x_2,\dots)$ with $x_n \in \bar{X}$ and $t_{n+1} \in \bar{T}$ for all $n = 0,1,\dots$. This set is endowed with the σ-field generated by the products of the Borel σ-fields $\mathfrak{B}(\bar{X})$ and $\mathfrak{B}(\bar{T})$.

Denote by Ω the subset of all sequences $\omega = (x_0,t_1,x_1,t_2,x_2,\dots)$ from $(\bar{X} \times \bar{T})^\infty$ such that: (i) $x_0 \in X$; (ii) for all $n = 1,2,\dots$, if $t_n < T_1$, then $t_n < t_{n+1}$ and $x_n \in X$, and if $t_n = T_1$, then $t_{n+1} = t_n$ and $x_n = x_\infty$. Observe that Ω is a measurable subset of $(\bar{X} \times \bar{T})^\infty$. Consider the measurable space $(\Omega,\mathscr{F})$, where $\mathscr{F}$ is the σ-field of the measurable subsets of Ω. For all $n = 0,1,\dots$, let $x_n(\omega) = x_n$ and $t_{n+1}(\omega) = t_{n+1}$, where $\omega \in \Omega$, be the random variables defined on the measurable space $(\Omega,\mathscr{F})$. Let $t_0 := T_0$, $t_\infty(\omega) := \lim_{n\to\infty} t_n(\omega)$, $\omega \in \Omega$, and for all $t \in [T_0,T_1]$, let $\mathscr{F}_t := \sigma(\mathfrak{B}(X),\mathscr{G}_t)$, where $\mathscr{G}_t := \sigma(\mathbf{I}\{x_n \in B\}\mathbf{I}\{t_n \leq s\} : n \geq 1, T_0 \leq s \leq t, B \in \mathfrak{B}(X))$. Throughout this paper, we omit ω whenever possible.

For a given Q-function q satisfying Assumption 2.2.4, consider the random measure ν on $([T_0,T_1[\times X, \mathfrak{B}([T_0,T_1[) \times \mathfrak{B}(X))$ defined by

$$\nu(\omega;[T_0,t],B) = \int_{T_0}^{t} \sum_{n\geq 0} \mathbf{I}\{t_n < s \leq t_{n+1}\} q(x_n,s,B \setminus \{x_n\}) ds, \qquad t \in [T_0,T_1[, B \in \mathfrak{B}(X). \quad (2.5)$$

Note that $\nu(\{t\},X) = \nu([t_\infty,\infty[,X) = 0$ and (2.5) can be rearranged as

$$\nu([T_0,t],B) = \sum_{n\geq 0} \mathbf{I}\{t_n < t \leq t_{n+1}\} \left(\sum_{m=0}^{n-1} \int_0^{t_{m+1}-t_m} q(x_m, t_m+s, B\setminus\{x_m\})ds \right.$$
$$\left. + \int_0^{t-t_n} q(x_n, t_n+s, B\setminus\{x_n\})ds \right). \quad (2.6)$$

As the expression in the parentheses on the right hand side of (2.6) is an $\mathscr{F}_{t_n}$-measurable process for each $B \in \mathfrak{B}(X)$, it follows from Jacod [17, Lemma 3.3] that the process $\{\nu([T_0,t],B) : t \in [T_0,T_1[\}$ is predictable. Therefore, the measure ν is a predictable random measure. According to Jacod [17, Theorem 3.6], the predictable random measure ν defined in (2.5) and a probability measure γ on X define a unique probability measure $\mathbb{P}$ on $(\Omega,\mathscr{F})$ such that $\mathbb{P}(x_0 \in B) = \gamma(B), B \in \mathfrak{B}(X)$, and ν is the compensator of the random measure of the multivariate point process $(t_n,x_n)_{n\geq 1}$ defined by the triplet $(\Omega,\mathscr{F},\mathbb{P})$.

Consider the process $\{\mathbb{X}_t : t \in [T_0,T_1[\}$,

$$\mathbb{X}_t(\omega) := \sum_{n\geq 0} \mathbf{I}\{t_n \leq t < t_{n+1}\}x_n + \mathbf{I}\{t_\infty \leq t\}x_\infty, \quad (2.7)$$

defined on $(\Omega,\mathscr{F},\mathbb{P})$ and adapted to the filtration $\{\mathscr{F}_t, t \in [T_0,T_1[\}$. Observe that the process $\{\mathbb{X}_t : t \in [T_0,T_1[\}$ is a jump process. The main result of this section, Theorem 2.3.2, shows that the process $\{\mathbb{X}_t : t \in [T_0,T_1[\}$ is a jump Markov process and provides its transition function.

For an $\mathscr{F}_t$-measurable stopping time τ, let $N(\tau) := \max\{n = 0,1,\ldots : \tau \geq t_n\}$. Since $N(\tau) = \infty$ and $\mathbb{X}_\tau = \{x_\infty\}$ when $\tau \geq t_\infty$, we follow the convention that $t_{\infty+1} = T_1$ and $x_{\infty+1} = x_\infty$. Denote by $G_\tau(\omega;\cdot,\cdot)$ and $H_\tau(\omega;\cdot)$ respectively the regular conditional laws of $(t_{N(\tau)+1}, x_{N(\tau)+1})$ and $t_{N(\tau)+1}$ with respect to $\mathscr{F}_\tau$; $H_\tau(\omega;\cdot) = G_\tau(\omega;\cdot,X)$. In particular, $G_{t_n}(\omega;\cdot,\cdot)$ and $H_{t_n}(\omega;\cdot)$, where $n = 0,1,\ldots$, denote the conditional laws of (t_{n+1},x_{n+1}) and t_{n+1} with respect to $\mathscr{F}_{t_n}$. We remark that the notations G_{t_n} and H_{t_n} correspond to the notations G_n and H_n in Jacod [17, p. 241].

Lemma 2.3.1. *For all $u,t \in [T_0,T_1[$, $u < t$,*

$$H_u([t,T_1]) = e^{-\int_u^t q(\mathbb{X}_u,s)ds}, \qquad N(u) < \infty, \quad (2.8)$$
$$G_u(dt,B) = e^{-\int_u^t q(\mathbb{X}_u,s)ds} q(\mathbb{X}_u,t,B\setminus\{\mathbb{X}_u\})dt, \qquad N(u) < \infty, B \in \mathfrak{B}(X). \quad (2.9)$$

Proof. In view of Jacod [17, Proposition 3.1], for all $t \in [T_0,T_1[$, $B \in \mathfrak{B}(X)$, and $n = 0,1,\ldots$

$$\nu(dt,B) = \frac{G_{t_n}(dt,B)}{H_{t_n}([t,\infty])}, \qquad t_n < t \leq t_{n+1}. \quad (2.10)$$

In particular, for $B = X$, from (2.10) and from the property that $x_{n+1} \in X$ when $t_{n+1} < T_1$,

$$\nu(dt,X) = \frac{G_{t_n}(dt,X)}{H_{t_n}([t,\infty])} = \frac{H_{t_n}(dt)}{H_{t_n}([t,\infty])}, \qquad t_n < t \leq t_{n+1}.$$

This equality implies that $\nu(dt, X)$ is the hazard rate function corresponding to the distribution H_{t_n} when $t_n < t \le t_{n+1}$. Therefore,

$$H_{t_n}([t,\infty]) = e^{-\nu(]t_n,t],X)\mathbf{I}\{t_n<t\le t_{n+1}\}}, \qquad t \in [T_0, T_1[\,, t > t_n. \tag{2.11}$$

From (2.5) and (2.11), for all $t \in [T_0, T_1[\,$,

$$H_{t_n}([t,\infty]) = e^{-\int_{t_n}^t q(x_n,s)ds}, \qquad t > t_n, \tag{2.12}$$

and from (2.5), (2.10), and (2.12), for all $t \in [T_0, T_1[\,$, $B \in \mathfrak{B}(X)$,

$$G_{t_n}(dt, B) = e^{-\int_{t_n}^t q(x_n,s)ds} q(x_n, t, B \setminus \{x_n\})dt, \qquad t > t_n. \tag{2.13}$$

To compute G_u, observe that for all $u, t \in [T_0, T_1[\,$, $u < t$, and $B \in \mathfrak{B}(X)$,

$$\begin{aligned} G_u(dt,B) &= \mathbb{P}(t_{N(u)+1} \in [t, t+dt[\,, x_{N(u)+1} \in B \mid \mathscr{F}_u) \\ &= \sum_{n\ge 0} \mathbb{P}(t_{N(u)+1} \in [t, t+dt[\,, x_{N(u)+1} \in B \mid \mathscr{F}_u)\mathbf{I}\{N(u)=n\} \\ &= \sum_{n\ge 0} \mathbb{P}(t_{n+1} \in [t, t+dt[\,, x_{n+1} \in B \mid \mathscr{F}_u, N(u)=n)\mathbf{I}\{N(u)=n\}, \end{aligned} \tag{2.14}$$

where the first equality follows from the definition of G_u, the second equality holds because $\{N(u)=\infty\} \cup \{N(u)=n\}_{n=0,1,\dots}$ is an $\mathscr{F}_u$-measurable partition of Ω and $x_{N(u)+1} = x_\infty \notin X$ when $N(u)=\infty$, and the third equality follows from $N(u)=n$ and from $\{N(u)=n\} \in \mathscr{F}_u$.

Observe that for any random variable Z on $(\Omega, \mathscr{F})$

$$\begin{aligned} \mathbb{P}(Z \mid \mathscr{F}_u, N(u)=n)\mathbf{I}\{N(u)=n\} &= \mathbb{P}(Z \mid \mathscr{F}_{t_n}, N(u)=n)\mathbf{I}\{N(u)=n\} \\ &= \mathbb{P}(Z \mid \mathscr{F}_{t_n}, t_n \le u, t_{n+1} > u)\mathbf{I}\{N(u)=n\} = \mathbb{P}(Z \mid \mathscr{F}_{t_n}, t_{n+1} > u)\mathbf{I}\{N(u)=n\} \\ &= \frac{\mathbb{P}(Z, t_{n+1} > u \mid \mathscr{F}_{t_n})}{\mathbb{P}(t_{n+1} > u \mid \mathscr{F}_{t_n})}\mathbf{I}\{N(u)=n\}, \end{aligned} \tag{2.15}$$

where the first equality follows from Brémaud [3, Theorem T32, p. 308], the second equality holds because $\{t_n \le u, t_{n+1} > u\} = \{N(u)=n\}$, the third equality holds because $\{t_n \le u\} \in \mathscr{F}_{t_n}$, and the last one follows from the definition of conditional probabilities. Let $Z = \{t_{n+1} \in [t, t+dt[\,, x_{n+1} \in B\}$, where $t \in [T_0, T_1[\,, B \in \mathfrak{B}(X)$. Then (2.14) and (2.15) imply

$$\begin{aligned} G_u(dt,B) &= \sum_{n\ge 0} \frac{\mathbb{P}(t_{n+1} \in [t, t+dt[\,, x_{n+1} \in B \mid \mathscr{F}_{t_n})}{\mathbb{P}(t_{n+1} > u \mid \mathscr{F}_{t_n})}\mathbf{I}\{N(u)=n\} \\ &= \sum_{n\ge 0} \frac{e^{-\int_{t_n}^t q(x_n,s)ds} q(x_n, t, B \setminus \{x_n\})dt}{e^{-\int_{t_n}^u q(x_n,s)ds}}\mathbf{I}\{N(u)=n\} \\ &= e^{-\int_u^t q(\mathbb{X}_u,s)ds} q(\mathbb{X}_u, t, B \setminus \{\mathbb{X}_u\})dt, \end{aligned} \tag{2.16}$$

where the first equality holds because $\{t_{n+1} \in [t, t+dt[\,, t_{n+1} > u\} = \{t_{n+1} \in [t, t+dt[\}$ when $t > u$, the second equality follows from (2.12) and (2.13), and the last equality holds since $x_n = \mathbb{X}_u$ when $N(u) = n$. For all $u, t \in [T_0, T_1[$, $u < t$, it follows from the property that $x_{N(u)+1} \in X$ when $t_{N(u)+1} < T_1$ and from (2.16) that $H_u([t, \infty])$ satisfies (2.8). $\square$

Following Feller [10, p. 501], for $x \in X$, $u, t \in [T_0, T_1[$, $u < t$, and $B \in \mathfrak{B}(X)$, define

$$\bar{P}^{(0)}(u, x; t, B) = I\{x \in B\} e^{-\int_u^t q(x,s)ds}, \tag{2.17}$$

and for $n \geq 1$ define

$$\bar{P}^{(n)}(u, x; t, B) = \int_u^t \int_{X \setminus \{x\}} e^{-\int_u^s q(x,\theta)d\theta} q(x, s, dy) \bar{P}^{(n-1)}(s, y; t, B) ds. \tag{2.18}$$

Set

$$\bar{P}(u, x; t, B) := \sum_{n=0}^{\infty} \bar{P}^{(n)}(u, x; t, B). \tag{2.19}$$

Observe that $\bar{P}$ is a transition function, if the Q-function q satisfies Assumption 2.2.4. For continuous Q-functions satisfying Assumption 2.2.1, Feller [10, Theorems 2, 5] proved that (a) for fixed u, x, t the function $\bar{P}(u, x; t, \cdot)$ is a measure on $(X, \mathfrak{B}(X))$ such that $0 \leq \bar{P}(u, x; t, \cdot) \leq 1$, and (b) for all u, x, t, B the function $\bar{P}(u, x; t, B)$ satisfies the Chapman-Kolmogorov equation (2.1). The proofs remain correct for measurable Q-functions satisfying Assumption 2.2.4. The measurability of $\bar{P}(u, x; t, B)$ in u, x, t for all $B \in \mathfrak{B}(X)$ is straightforward from the definitions (2.17), (2.18), and (2.19). Therefore, if q satisfies Assumption 2.2.4, the function $\bar{P}$ takes values in $[0, 1]$ and satisfies properties (i)-(iii) from the definition of a transition function.

Theorem 2.3.2. *Given a probability measure γ on X and a Q-function q satisfying Assumption 2.2.4, the jump process $\{\mathbb{X}_t : t \in [T_0, T_1[\}$ defined in (2.7) is a jump Markov process with transition function $\bar{P}$.*

Proof. Observe that if, for all $u, t \in [T_0, T_1[$, $u < t$, and $B \in \mathfrak{B}(X)$,

$$\mathbb{P}(\mathbb{X}_t \in B \mid \mathscr{F}_u) = \mathbb{P}(\mathbb{X}_t \in B \mid \mathbb{X}_u) = \bar{P}(u, \mathbb{X}_u; t, B), \qquad u < t_\infty, \tag{2.20}$$

then the jump process $\{\mathbb{X}_t : t \in [T_0, T_1[\}$ is a jump Markov process with transition function $\bar{P}$. To prove (2.20), we first establish by induction that for all $n = 0, 1, \ldots$, $u, t \in [T_0, T_1[$, $u < t$, and $B \in \mathfrak{B}(X)$

$$\mathbb{P}(\mathbb{X}_t \in B, N_{]u,t]} = n \mid \mathscr{F}_u) = \bar{P}^{(n)}(u, \mathbb{X}_u; t, B), \qquad u < t_\infty, \tag{2.21}$$

where $N_{]u,t]} := N(t) - N(u)$ when $u < t_\infty$ and $N_{]u,t]} := \infty$ when $u \geq t_\infty$. Equation (2.21) holds for $n = 0$ because for $u < t_\infty$

$$\begin{aligned} \mathbb{P}(\mathbb{X}_t \in B, N_{]u,t]} = 0 \mid \mathscr{F}_u) &= \mathbb{P}(\mathbb{X}_u \in B, t_{N(u)+1} > t \mid \mathscr{F}_u) \\ &= I\{\mathbb{X}_u \in B\} H_u(]t, \infty]) = I\{\mathbb{X}_u \in B\} e^{-\int_u^t q(\mathbb{X}_u, s)ds} = \bar{P}^{(0)}(u, \mathbb{X}_u; t, B), \end{aligned} \tag{2.22}$$

where the first equality holds because the corresponding events coincide, the second equality holds because $\{\mathbb{X}_u \in B\} \in \mathscr{F}_u$ and from the definition of H_u, the third equality is correct because of (2.8), and the last equality is (2.17).

For some $n \geq 0$, assume that (2.21) holds. Then for $u < t_\infty$

$$\begin{aligned}
&\mathbb{P}(\mathbb{X}_t \in B, N_{]u,t]} = n+1 \mid \mathscr{F}_u) \\
&\quad = \int_u^t \int_{X\setminus\{\mathbb{X}_u\}} \mathbb{P}(\mathbb{X}_t \in B, N_{]t_{N(u)+1},t]} = n \mid \mathscr{F}_u, t_{N(u)+1}, x_{N(u)+1}) G_u(dt_{N(u)+1}, dx_{N(u)+1}) \\
&\quad = \int_u^t \int_{X\setminus\{\mathbb{X}_u\}} \mathbb{P}(\mathbb{X}_t \in B, N_{]t_{N(u)+1},t]} = n \mid \mathscr{F}_{t_{N(u)+1}}) G_u(dt_{N(u)+1}, dx_{N(u)+1}) \qquad (2.23)\\
&\quad = \int_u^t \int_{X\setminus\{\mathbb{X}_u\}} q(\mathbb{X}_u, s, dy) e^{-\int_u^s q(\mathbb{X}_u,\theta)d\theta} \bar{P}^{(n)}(s,y;t,B)ds = \bar{P}^{(n+1)}(u,\mathbb{X}_u;t,B),
\end{aligned}$$

where the first equality is correct since $N_{]u,t]} = 1 + N_{]t_{N(u)+1},t]}$ for $N_{]u,t]} \geq 1$ and since $\mathbb{E}(\mathbb{E}(Z \mid \mathfrak{D})) = \mathbb{E}(Z)$ for any random variable Z and any σ-field $\mathfrak{D}$, the second equality holds because $\sigma(\mathscr{F}_u, t_{N(u)+1}, x_{N(u)+1}) = \mathscr{F}_{t_{N(u)+1}}$, the third equality follows from (2.9) and (2.21), and the last equality is (2.18). Equality (2.21) is proved.

Observe that for $u,t \in [T_0, T_1[$, $u < t$, $B \in \mathfrak{B}(X)$,

$$\begin{aligned}
\mathbb{P}(\mathbb{X}_t \in B \mid \mathscr{F}_u) &= \mathbb{P}(\mathbb{X}_t \in B \mid \mathscr{F}_u)\mathbf{I}\{u < t_\infty\} + \mathbb{P}(\mathbb{X}_t \in B \mid \mathscr{F}_u)\mathbf{I}\{u \geq t_\infty\} \\
&= \sum_{n\geq 0} \mathbb{P}(\mathbb{X}_t \in B, N_{]u,t]} = n \mid \mathscr{F}_u)\mathbf{I}\{u < t_\infty\} = \sum_{n\geq 0} \bar{P}^{(n)}(u,\mathbb{X}_u;t,B)\mathbf{I}\{u < t_\infty\} \qquad (2.24)\\
&= \bar{P}(u,\mathbb{X}_u;t,B)\mathbf{I}\{u < t_\infty\} = \bar{P}(u,\mathbb{X}_u;t,B)\mathbf{I}\{\mathbb{X}_u \in X\},
\end{aligned}$$

where the first equality holds since $\{\{u < t_\infty\}, \{u \geq t_\infty\}\}$ is a partition of Ω and $\{u < t_\infty\}, \{u \geq t_\infty\} \in \mathscr{F}_u$, the second equality holds since $\mathbb{X}_t \in X$ implies $t < t_\infty$, the third equality follows from (2.21), the fourth equality follows from (2.19), and the last one holds since $\{u < t_\infty\} = \{\mathbb{X}_u \in X\}$. As follows from (2.24), the function $\mathbb{P}(\mathbb{X}_t \in B \mid \mathscr{F}_u)$ is $\sigma(\mathbb{X}_u)$-measurable. Thus,

$$\mathbb{P}(\mathbb{X}_t \in B \mid \mathscr{F}_u) = \mathbb{P}(\mathbb{P}(\mathbb{X}_t \in B \mid \mathscr{F}_u) \mid \mathbb{X}_u) = \mathbb{P}(\mathbb{X}_t \in B \mid \mathbb{X}_u), \qquad (2.25)$$

where the second equality holds because $\sigma(\mathbb{X}_u) \subseteq \mathscr{F}_u$; see e.g. Brémaud [3, p. 280]. Thus, (2.20) follows from (2.24) and (2.25). □

Corollary 2.3.3. *Given a probability measure γ on X and a Q-function q satisfying Assumption 2.2.2, the jump process $\{\mathbb{X}_t : t \in [T_0, T_1[\}$ defined in* (2.7) *is a jump Markov process with transition function $\bar{P}$.*

Proof. This corollary follows from Lemma 2.2.1 and Theorem 2.3.2. □

The following lemma provides a simple statement that is needed for future references. Consider a Q-function q. Let q' be a Q-function such that

$$\mu(t \in [T_0, T_1[: q(x,t,B) \neq q'(x,t,B) \text{ for some } B \in \mathfrak{B}(X)) = 0, \qquad x \in X, \qquad (2.26)$$

where μ is the Lebesgue measure on $\mathbb{R}_+$.

Lemma 2.3.4. *For arbitrary T_0 and T_1, consider a Q-function q satisfying Assumption 2.2.4. Let q' be a Q-function satisfying* (2.26)*, and let $\bar{P}'(u,x;t,B)$, where $u,t \in [T_0,T_1[$, $u < t$, $x \in X$, and $B \in \mathfrak{B}(X)$, be the transition function $\bar{P}$ defined by* (2.19) *with q replaced by q'. Then, for a given probability measure γ on X, both the Q-functions q and q' define the same jump Markov process $\{\mathbb{X}_t : t \in [T_0,T_1[\}$ defined in* (2.7) *and*

$$\bar{P}'(u,x;t,B) = \bar{P}(u,x;t,B), \qquad u,t \in [T_0,T_1[, u < t, x \in X, B \in \mathfrak{B}(X). \tag{2.27}$$

Proof. Observe that the Q-function q' satisfies Assumption 2.2.4. This follows from (2.26). Then, it follows from Theorem 2.3.2 that, for a given a probability measure γ on X, each of the Q-functions q and q' define a jump Markov process $\{\mathbb{X}_t : t \in [T_0,T_1[\}$ defined in (2.7) with transition function $\bar{P}$ and $\bar{P}'$, respectively. Thus, if (2.27) holds, then the Q-functions q and q' define the same jump Markov process for a given probability measure γ on X. This is indeed true as explained below. It follows immediately from (2.17)–(2.19) that two Q-functions, that are equal almost everywhere in t with respect to the Lebesgue measure, define the same transition function. This fact and (2.26) imply that (2.27) holds. □

2.4 Kolmogorov's backward equation

In this section, by using the methods introduced by Feller [10, Theorems 2, 3] for continuous Q-functions, we show that the transition function $\bar{P}$ defined in (2.19) is the minimal non-negative solution of Kolmogorov's backward equation.

Definition 2.4.1. *A function f defined on $\mathbb{R}$ is called locally absolutely continuous on an interval $I \subseteq \mathbb{R}$, if for any closed bounded interval $[a,b] \subseteq I$, the function f is absolutely continuous on $[a,b]$.*

Definition 2.4.2. *For a Q-function q, a function $\hat{P}(u,x;t,B)$, where $t \in]T_0,T_1[$, $u \in [T_0,t[$, $x \in X$, and $B \in \mathfrak{B}(X)$, is a solution of Kolmogorov's backward equation* (2.29) *on the semi-interval $[T_0,T_1[$, if the function $\hat{P}(u,x;t,B)$ satisfies the following properties:*
(i) for each x, t, B, the function $\hat{P}(u,x;t,B)$ is locally absolutely continuous on $u \in [T_0,t[$ and satisfies the boundary condition

$$\lim_{u \to t-} \hat{P}(u,x;t,B) = I\{x \in B\}; \tag{2.28}$$

(ii) for each x,t,B,

$$\frac{\partial}{\partial u} P(u,x;t,B) = q(x,u)P(u,x;t,B) - \int_{X \setminus \{x\}} q(x,u,dy)P(u,y;t,B) \quad for\ almost\ every\ u \in [T_0,t[. \tag{2.29}$$

Theorem 2.4.1. *Under Assumption 2.2.4, the function $\bar{P}$ is the minimal non-negative solution of Kolmogorov's backward equation* (2.29) *on the semi-interval $[T_0,T_1[$. In addition, if $\bar{P}$ is a regular* *transition function (that is, $\bar{P}(u,x;t,X) = 1$ for all u,x,t in the domain of $\bar{P}$), then $\bar{P}$ is the unique non-negative solution of Kolmogorov's backward equation* (2.29) *on the semi-interval $[T_0, T_1[$ that is a measure on $(X, \mathfrak{B}(X))$ for fixed u, x, t with $u < t$ and takes values in $[0,1]$.*

We first show in Theorem 2.4.2 that the function $\bar{P}$ is a solution of Kolmogorov's backward equation (2.29), and then provide the proof of Theorem 2.4.1.

Theorem 2.4.2. *Under Assumption 2.2.4, the function $\bar{P}$ is a solution of Kolmogorov's backward equation* (2.29) *on the semi-interval $[T_0, T_1[$.*

Proof. For all $x \in X$, $u,t \in [T_0, T_1[$, $u < t$, and $B \in \mathfrak{B}(X)$,

$$
\begin{aligned}
\bar{P}(u,x;t,B) &= \sum_{n=0}^{\infty} \bar{P}^{(n)}(u,x;t,B) \\
&= I\{x \in B\} e^{-\int_u^t q(x,s)ds} + \sum_{n=1}^{\infty} \int_u^t e^{-\int_u^s q(x,\theta)d\theta} \int_{X\setminus\{x\}} q(x,s,dy) \bar{P}^{(n-1)}(s,y;t,B) ds \\
&= I\{x \in B\} e^{-\int_u^t q(x,s)ds} + \int_u^t e^{-\int_u^s q(x,\theta)d\theta} \int_{X\setminus\{x\}} q(x,s,dy) \sum_{n=1}^{\infty} \bar{P}^{(n-1)}(s,y;t,B) ds \qquad (2.30) \\
&= I\{x \in B\} e^{-\int_u^t q(x,s)ds} + \int_u^t e^{-\int_u^s q(x,\theta)d\theta} \int_{X\setminus\{x\}} q(x,s,dy) \bar{P}(s,y;t,B) ds,
\end{aligned}
$$

where the first equality is (2.19), the second equality follows from (2.17) and (2.18), the third equality is obtained by interchanging the integral and sum, and the last one follows from (2.19). For fixed x,t,B, equation (2.30) implies that $\bar{P}(u,x;t,B)$ is the sum of two locally absolutely continuous functions on u. Thus, $\bar{P}(u,x;t,B)$ is for fixed x,t,B locally absolutely continuous functions on u.

Observe that $\bar{P}^{(n)}(u,x;t,B) \le \bar{P}(u,x;t,B) \le 1$ for all $n \ge 0$, $x \in X, u,t \in [T_0, T_1[$, $u < t$, and $B \in \mathfrak{B}(X)$. Then from (2.18),

$$
\bar{P}^{(n)}(u,x;t,B) \le \int_u^t e^{-\int_u^s q(x,\sigma)d\sigma} q(x,s) ds = 1 - e^{-\int_u^t q(x,s)ds}, \qquad n \ge 1. \tag{2.31}
$$

Since $e^{-\int_u^t q(x,s)ds} \to 1$ as $u \to t^-$ for any Q-function q satisfying Assumption 2.2.4, the above inequality and (2.17) imply that

$$
\lim_{u\to t^-} \bar{P}^{(n)}(u,x;t,B) = 0 \quad \text{for all } n \ge 1 \quad \text{and} \quad \lim_{u\to t^-} \bar{P}^{(0)}(u,x;t,B) = I\{x \in B\}. \tag{2.32}
$$

Thus, it follows from (2.19) and (2.32) that (2.28) holds with $\hat{P} = \bar{P}$.

In addition, since locally absolutely continuous real-valued function is differentiable almost everywhere on its domain, for all x,t,B the function $\bar{P}(u,x;t,B)$ is differentiable in u almost everywhere on $[T_0, t[$. By differentiating (2.30), for almost every $u < t$,

$$
\begin{aligned}
\frac{\partial}{\partial u}\bar{P}(u,x;t,B) &= I\{x\in B\}e^{-\int_u^t q(x,s)ds}q(x,u) - \int_{X\setminus\{x\}} q(x,u,dy)\bar{P}(u,y;t,B) \\
&\qquad + \int_u^t \frac{\partial}{\partial u} e^{-\int_u^s q(x,\theta)d\theta} \int_{X\setminus\{x\}} q(x,s,dy)\bar{P}(s,y;t,B)ds \\
&= I\{x\in B\}e^{-\int_u^t q(x,s)ds}q(x,u) - \int_{X\setminus\{x\}} q(x,u,dy)\bar{P}(u,y;t,B) \\
&\qquad + \int_u^t e^{-\int_u^s q(x,\theta)d\theta} q(x,u) \int_{X\setminus\{x\}} q(x,s,dy)\bar{P}(s,y;t,B)ds.
\end{aligned} \tag{2.33}
$$

In view of (2.30), the sum of the first and the last terms in the last expression of (2.33) is equal to the first term on the right-hand side of (2.29). That is, the function $\bar{P}$ satisfies (2.29) for almost every $u \in [T_0,t[$. Therefore, $\bar{P}$ is a solution of Kolmogorov's backward equation (2.29) on the semi-interval $[T_0,T_1[$. □

Proof of Theorem 2.4.1. In view of Theorem 2.4.2, the function $\bar{P}$ is a solution of Kolmogorov's backward equation (2.29) on the semi-interval $[T_0,T_1[$. The proof of minimality of $\bar{P}$ is similar to the proof of Theorem 3 in Feller [10]. We provide it here for completeness. Consider a non-negative solution $P^*(u,x;t,B)$ of Kolmogorov's backward equation (2.29) on the semi-interval $[T_0,T_1[$. Integrating (2.29) from u to t and by using the boundary condition (2.28),

$$
P^*(u,x;t,B) = I\{x\in B\}e^{-\int_u^t q(x,s)ds} + \int_u^t \int_{X\setminus\{x\}} e^{-\int_u^s q(x,\theta)d\theta} q(x,s,dy)P^*(s,y;t,B)ds. \tag{2.34}
$$

Since the last term of (2.34) is non-negative,

$$
P^*(u,x;t,B) \geq I\{x\in B\}e^{-\int_u^t q(x,s)ds} = \bar{P}^{(0)}(u,x;t,B), \tag{2.35}
$$

where the last equality is (2.17). For all u,x,t,B with $u<t$, assume $P^*(u,x;t,B) \geq \sum_{m=0}^{n} \bar{P}^{(m)}(u,x;t,B)$ for some $n \geq 0$. Then from (2.34)

$$
\begin{aligned}
P^*(u,x;t,B) &\geq I\{x\in B\}e^{-\int_u^t q(x,s)ds} + \int_u^t \int_{X\setminus\{x\}} e^{-\int_u^s q(x,\theta)d\theta} q(x,s,dy) \sum_{m=0}^{n} \bar{P}^{(m)}(s,y;t,B)ds \\
&= \bar{P}^{(0)}(u,x;t,B) + \sum_{m=0}^{n} \bar{P}^{(m+1)}(u,x;t,B) = \sum_{m=0}^{n+1} \bar{P}^{(m)}(u,x;t,B),
\end{aligned}
$$

where the first equality follows from the assumption that $P^*(u,x;t,B) \geq \sum_{m=0}^{n} \bar{P}^{(m)}(u,x;t,B)$ for all u,x,t,B with $u<t$, the second equality follows from (2.17) and (2.18), and the third equality is straightforward. Thus, by induction, $P^*(u,x;t,B) \geq \sum_{m=0}^{n} \bar{P}^{(m)}(u,x;t,B)$ for all $n \geq 0$, $x \in X$, $u,t \in \mathbb{R}_+$, $u<t$, and $B \in \mathfrak{B}(X)$, which implies that $P^*(u,x;t,B) \geq \bar{P}(u,x;t,B)$ for all u,x,t,B.

To prove the second part of the theorem, let the solution P^* be a measure on $(X,\mathfrak{B}(X))$ for fixed u,x,t and with values in $[0,1]$. Assume that $P^*(u,x;t,B) \neq \bar{P}(u,x;t,B)$ for at least one tuple (u,x,t,B). Then,

$$P^*(u,x;t,X) = P^*(u,x;t,B) + P^*(u,x;t,B^c) > \bar{P}(u,x;t,B) + \bar{P}(u,x;t,B^c) = \bar{P}(u,x;t,X) = 1,$$

where the inequality holds because $P^*(u,x,t,\cdot) \geq \bar{P}(u,x,t,\cdot)$ for all u,x,t. Since P^* takes values in $[0,1]$, the assumption that $P^*(u,x;t,B) \neq \bar{P}(u,x;t,B)$ for at least one tuple (u,x,t,B) leads to a contradiction. □

2.5 Kolmogorov's forward equation

Kolmogorov's forward equation (2.37) was studied by Feller [10, Theorem 1] for continuous Q-functions satisfying Assumption 2.2.1. In this section, we show that the function $\bar{P}$ is the minimal non-negative solution of Kolmogorov's forward equation (2.37) on the semi-interval $[T_0,T_1[$ if Assumption 2.2.3 holds, which, as stated in Lemma 2.2.1, is more general than Assumption 2.2.1.

Consider the sets X_n^t on which $q(x,\cdot) < n$ almost everywhere on $[T_0,t[$. These sets are defined in (2.4).

Definition 2.5.1. *For $t \in]T_0,T_1[$, a set $B \in \mathfrak{B}(X)$ is called (q,t)-bounded if $B \subseteq X_n^t$ for some $n = 1,2,\dots$.*

Definition 2.5.2. *For a given Q-function q, a function $\hat{P}(u,x;t,B)$, where $u \in [T_0,T_1[$, $t \in]u,T_1[$, $x \in X$, and $B \in \mathfrak{B}(X)$, is a solution of Kolmogorov's forward equation* (2.37)*, if for each u,x,s,B, such that $s \in]T_0,T_1[$, $u \in [T_0,s[$, $x \in X$, and the set B is (q,s)-bounded,*

(i) the function $\hat{P}(u,x;t,B)$ is locally absolutely continuous on $t \in]u,s[$ and satisfies the boundary condition

$$\lim_{t\to u+} P(u,x;t,B) = \mathbf{I}\{x \in B\}; \tag{2.36}$$

(ii) the function $\hat{P}(u,x;t,B)$ satisfies Kolmogorov's forward equation,

$$\frac{\partial}{\partial t}P(u,x;t,B) = -\int_B q(y,t)P(u,x;t,dy) + \int_X q(y,t,B\setminus\{y\})P(u,x;t,dy) \quad \text{for almost every } t \in]u,s[. \tag{2.37}$$

The main result of this section, Theorem 2.5.1, shows that under Assumption 2.2.3 the transition function $\bar{P}(u,x;t,B)$ is the minimal non-negative solution of Kolmogorov's forward equation (2.37) on the semi-interval $[T_0,T_1[$ and provides a sufficient condition for its uniqueness.

Theorem 2.5.1. *Under Assumption 2.2.3, the function $\bar{P}(u,x;t,B)$ is the minimal non-negative solution of Kolmogorov's forward equation* (2.37)*. Also, if $\bar{P}$ is a regular transition function, then $\bar{P}$ is the unique non-negative solution of Kolmogorov's forward equation* (2.37) *that takes values in* $[0,1]$.

We give the proof of Theorem 2.5.1 after presenting a few auxiliary results. Theorem 2.5.1 states that the function $\bar{P}$ satisfies Kolmogorov's forward equation (2.37) for (q,s)-bounded sets $B \in \mathfrak{B}(X)$. The following example demonstrates that, in general, it is not possible to extend (2.37) to all the sets $B \in \mathfrak{B}(X)$.

Example 2.5.2. *Kolmogorov's forward equation* (2.37) *may not hold for all sets* $B \in \mathfrak{B}(X)$. Let $X = \mathbb{Z}$, where $\mathbb{Z}$ denotes the set of integers, $q(0,t) = 1$, $q(0,t,j) = 2^{-(|j|+1)}$ for all $j \neq 0$, and $q(j,t,-j) = q(j,t) = 2^{|j|}$ for all $j \neq 0$. If $\mathbb{X}_u = 0$, then starting at time u the process spends an exponentially distributed amount of time at state 0, then it jumps to a state $j \neq 0$ with probability $2^{-(|j|+1)}$, and then it oscillates between the states j and $-j$ with equal intensities. Thus for all $u,t \in [T_0,T_1[$ with $u < t$,

$$\bar{P}(u,0;t,0) = e^{-(t-u)} \qquad \text{and} \qquad \bar{P}(u,0;t,j) = \frac{1-e^{-(t-u)}}{2^{|j|+1}}, \qquad j \neq 0,$$

which implies that

$$\begin{aligned}\int_X q(y,t,X \setminus \{y\})\bar{P}(u,0;t,dy) &= q(0,t)\bar{P}(u,0;t,0) + \sum_{j\neq 0} q(j,t,-j)\bar{P}(u,0;t,j) \\ &= e^{-(t-u)} + \sum_{j>0}(1-e^{-(t-u)}) = \infty.\end{aligned}$$

Thus, if $B = X$, then (2.37) does not hold because both integrals in (2.37) are infinite. □

Recall that $\bar{q}(x) := \sup_{t\in[T_0,T_1[} q(x,t)$ for all $x \in X$.

Definition 2.5.3. *A set* $B \in \mathfrak{B}(X)$ *is called q-bounded if* $\sup_{x\in B} \bar{q}(x) < \infty$.

For continuous Q-functions, Feller [10, Theorem 1] showed that the transition function $\bar{P}$ satisfies Kolmogorov's forward equation (2.37) for all q-bounded sets B. In order to show that the function $\bar{P}$ is a solution of Kolmogorov's forward equation (2.37), we show in Theorem 2.5.3 that this property is correct for measurable Q-functions.

Theorem 2.5.3. *Under Assumption 2.2.2, the following statements hold: for each* u,x,B, *such that* $u \in [T_0,T_1[$, $x \in X$, *and the set B is q-bounded,*

(a) the function $\bar{P}(u,x;t,B)$ *is locally absolutely continuous on* $t \in]u,T_1[$ *and satisfies the boundary condition* (2.36)*;*

(b) the function $\bar{P}(u,x;t,B)$ *satisfies* (2.37) *with* $s = T_1$.

Proof. (a) For all $x \in X, u,t \in [T_0,T_1[$, $u < t$, and $B \in \mathfrak{B}(X)$, equation (2.30) implies that the function $\bar{P}(u,x;t,B)$ is locally absolutely continuous on $t \in]u,T_1[$ for fixed u,x,B. Also, it follows from (2.17) and (2.31) that for any Q-function q satisfying Assumption 2.2.2,

$$\lim_{t\to u^+} \bar{P}^{(n)}(u,x;t,B) = 0 \quad \text{for all } n \geq 1 \quad \text{and} \quad \lim_{t\to u^+} \bar{P}^{(0)}(u,x;t,B) = I\{x \in B\} \tag{2.38}$$

uniformly with respect to B. Thus, (2.19) and (2.38) imply that the function $\bar{P}$ satisfies (2.36).

(b) From the last equality of (2.23) with $\mathbb{X}_u = x$ and the property that the jump process $\{\mathbb{X}_t : t \in [T_0, T_1[\}$ is a jump Markov process, for all $u \in [T_0, T_1[$, $t \in]u, T_1]$, $x \in X$, $B \in \mathfrak{B}(X)$, and $n = 1, 2, \ldots$,

$$\bar{P}^{(n)}(u,x;t,B) = \int_u^t \int_B q(y,s,dz \setminus \{y\}) e^{-\int_s^t q(z,\theta)d\theta} \bar{P}^{(n-1)}(u,x;s,dy)ds \tag{2.39}$$

Then, from (2.17), (2.19), and (2.39), we have

$$\begin{aligned}\bar{P}(u,x;t,B) &= \sum_{n=0}^{\infty} \bar{P}^{(n)}(u,x;t,B) = \bar{P}^{(0)}(u,x;t,B) \sum_{n=0}^{\infty} \bar{P}^{(n+1)}(u,x;t,B) \\ &= I\{x \in B\} e^{-\int_u^t q(x,s)ds} + \sum_{n=0}^{\infty} \int_u^t \int_B q(y,s,dz \setminus \{y\}) e^{-\int_s^t q(z,\theta)d\theta} \bar{P}^{(n)}(u,x;s,dy)ds \\ &= I\{x \in B\} e^{-\int_u^t q(x,s)ds} + \int_u^t \int_B q(y,s,dz \setminus \{y\}) e^{-\int_s^t q(z,\theta)d\theta} \bar{P}(u,x;s,dy)ds.\end{aligned} \tag{2.40}$$

Since $\bar{P}(u,x;t,B)$ is locally absolutely continuous function on $t \in [T_0, T_1[$ for fixed u, x, B, the derivative $\frac{\partial}{\partial t}\bar{P}(u,x;t,B)$ exists for almost every $t \in]u, T_1[$. By differentiating (2.40), for almost every $t > u$,

$$\begin{aligned}\frac{\partial}{\partial t}\bar{P}(u,x;t,B) = -I\{x \in B\} e^{-\int_u^t q(x,s)ds} q(x,t) + \int_B q(y,t,dz \setminus \{y\}) \bar{P}(u,x;t,dy) \\ + \int_u^t \frac{\partial}{\partial t} \int_X \int_B q(y,s,dz \setminus \{y\}) e^{-\int_s^t q(z,\theta)d\theta} \bar{P}(u,x;s,dy)ds.\end{aligned} \tag{2.41}$$

Observe that, for all q-bounded sets $B \in \mathfrak{B}(X)$,

$$\begin{aligned}\frac{\partial}{\partial t} \int_B q(y,s,dz \setminus \{dy\}) e^{-\int_s^t q(z,\theta)d\theta} &= \int_B q(y,s,dz \setminus \{y\}) \frac{\partial}{\partial t} e^{-\int_s^t q(z,\theta)d\theta} \\ &= -\int_B q(z,t) q(y,s,dz \setminus \{y\}) e^{-\int_s^t q(z,\theta)d\theta}.\end{aligned} \tag{2.42}$$

Combining (2.41) and (2.42), for all q-bounded sets B,

$$\begin{aligned}\frac{\partial}{\partial t}\bar{P}(u,x;t,B) = -I\{x \in B\} e^{-\int_u^t q(x,s)ds} q(x,t) + \int_X q(y,t,B \setminus \{y\}) \bar{P}(u,x;t,dy) \\ - \int_u^t \int_X \int_B q(z,t) q(y,s,dz \setminus \{y\}) e^{-\int_s^t q(z,\theta)d\theta} \bar{P}(u,x;s,dy)ds,\end{aligned} \tag{2.43}$$

for almost every $t > u$. By substituting $\bar{P}(u,x;t,dz)$ in the left-hand side of the following equality with the final expression in (2.40),

$$\int_B q(z,t)\bar{P}(u,x;t,dz) = I\{x\in B\}e^{-\int_u^t q(x,s)ds}q(x,t) \\ + \int_u^t \int_X \int_B q(z,t)q(y,s,dz\setminus\{y\})e^{-\int_s^t q(z,\theta)d\theta}\bar{P}(u,x;s,dy)ds. \tag{2.44}$$

Formulae (2.43) and (2.44) imply statement (b) of the theorem. □

It follows from Theorem 2.5.3 that under Assumption 2.2.2 the transition function $\bar{P}$ satisfies Kolmogorov's forward equation (2.37) for all q-bounded sets. To show that the function $\bar{P}$ satisfies (2.37) for all (q,s)-bounded sets, we consider the Q-functions q^s defined in (2.46) that satisfy Assumption 2.2.2 if the Q-function q satisfies the weaker Assumption 2.2.3.

Define the function $N : X\times[T_0,T_1[\to\{0,1,\dots\}$,

$$N(x,s) = \min\{n=1,2,\dots : x\in X_n^s\}, \qquad x\in X, s\in[T_0,T_1[, \tag{2.45}$$

where the sets X_n^s are defined in (2.4) and $\min\{\emptyset\}:=\infty$. For each x, the value of $N(x,s)$ is the minimum natural number n for which $q(x,t)<n$ for almost every $t\in[T_0,s[$. Consider the Q-functions q^s, $s\in]T_0,T_1[$, satisfying

$$q^s(x,t,B) = q(x,t,B)\mathbf{I}\{t\in[T_0,s[\}\mathbf{I}\{q(x,t)<N(x,s)\}, \qquad x\in X, t\in[T_0,T_1[, B\in\mathfrak{B}(X). \tag{2.46}$$

The difference between the Markov processes defined by the Q-functions q and q^s is that the latter stops at time s and does not move at times $t\in[T_0,s[$ and states $x\in X$ with $q(x,t)\ge N(x,s)$. At all other time instances, these two processes are controlled by the same Q-function q. Let $\bar{P}^s$ be the transition function $\bar{P}$ defined in (2.19) with q replaced with q^s.

Lemma 2.5.4. *For arbitrary T_0 and T_1, consider a Q-function q satisfying Assumption 2.2.3. Then, for each $s\in]T_0,T_1[$, the following statements hold:*

(a) the Q-function q^s satisfies Assumption 2.2.2.

(b) For any given probability measure γ on X, the Q-functions q and q^s define the same jump Markov process $\{\mathbb{X}_t : t\in[T_0,T_1[\}$ up to time s, and

$$\bar{P}(u,x;t,B) = \bar{P}^s(u,x;t,B), \qquad u\in[T_0,s[, t\in]u,s[, x\in X, B\in\mathfrak{B}(X). \tag{2.47}$$

Proof. (a) Assumption 2.2.3 implies that, for each $x\in X$ and $s\in]T_0,T_1[$, there exists an $n\in\{1,2,\dots\}$ such that $x\in X_n^s$, and therefore, $N(x,s)<\infty$. This fact and (2.46) with $B=X\setminus\{x\}$ imply that, for all $x\in X$,

$$q^s(x,t) = q(x,t)\mathbf{I}\{t\in[T_0,s[\}\mathbf{I}\{q(x,t)<N(x,s)\} < N(x,s) < \infty, \qquad t\in[T_0,T_1[. \tag{2.48}$$

Thus, the Q-function q^s satisfies Assumption 2.2.2.

(b) Assumption 2.2.3 and (2.45) imply that $x \in X^s_{N(x,s)}$ for all $x \in X$ and $s \in]T_0, T_1[$. Thus, from the definition of the sets X^s_n, formula (2.4), for all $s \in]T_0, T_1[$,

$$\mu(t \in [T_0, s[: q(x,t) \geq N(x,s)) = 0, \qquad x \in X. \tag{2.49}$$

The above equality and (2.46) imply that (2.26) holds with $T_1 = s$ and $q' = q^s$. This fact, Lemma 2.2.1(iii), and Lemma 2.3.4 with $T_1 = s$ and $q' = q^s$ imply statement (b) of the lemma. $\square$

For all $s \in]T_0, T_1[$ and $t \in [T_0, s[$, let $Y^1_{t,s} := \{x \in X : q(x,t) < N(x,s)\}$ and $Y^2_{t,s} := X \setminus Y^1_{t,s}$. For any set $B \in \mathfrak{B}(X)$, we denote by $B^1_{t,s} := B \cap Y^1_{t,s}$ and $B^2_{t,s} := B \cap Y^2_{t,s}$.

Lemma 2.5.5. *Under Assumption 2.2.3, for all $x \in X$, $s \in]T_0, T_1[$, and $u \in [T_0, s[$,*

$$\bar{P}(u, x; t, Y^2_{t,s}) = 0 \quad \textit{for almost every } t \in]u, s[. \tag{2.50}$$

Proof. Fix an arbitrary $x \in X$ and $s \in]T_0, T_1[$. To prove (2.50), we first show that (2.50) holds for the particular case when $u = T_0$. According to Lemma 2.5.4(b), given an initial state x, the Q-functions q and q^s define the same jump Markov process $\{\mathbb{X}_t : t \in [T_0, T_1[\}$ up to time s. This fact implies that the compensator corresponding to this process can be given by (2.5) with $T_1 = s$ or by (2.5) with $T_1 = s$ and with q replaced by q^s. However, it follows from Jacod [17, Theorem 2.1] that a compensator is unique up to a modification of a $\mathbb{P}$-null set. Thus, from Jacod [17, Theorem 2.1] and Lemma 2.5.4(b),

$$\int_{T_0}^{s}\int_X q(z,t)\bar{P}(T_0, x; t, dz)dt = \int_{T_0}^{s}\int_X q^s(z,t)\bar{P}(T_0, x; t, dz)dt = \int_{T_0}^{s}\int_X q(z,t)\mathbf{I}\{z \in Y^1_{t,s}\}\bar{P}(T_0, x; t, dz)dt,$$

where the first equality follows from Jacod [17, (1)] with $X(t,z) = \mathbf{I}\{z \in X, t \in]T_0, s[\}$, formula (2.5) with $T_1 = s$ given for the Q-functions q and q^s, and (2.7), and the last one follows from (2.46). The above equality implies that

$$\int_{T_0}^{s}\int_X q(z,t)\mathbf{I}\{z \in Y^2_{t,s}\}\bar{P}(T_0, x; t, dz)dt = 0. \tag{2.51}$$

Observe that

$$\int_{T_0}^{s}\bar{P}(T_0, x; t, Y^2_{t,s})dt \leq \int_{T_0}^{s}\int_X q(z,t)\mathbf{I}\{z \in Y^2_{t,s}\}\bar{P}(T_0, x; t, dz)dt. \tag{2.52}$$

This is true because $q(z,t) \geq N(z,s) \geq 1$ for all $z \in Y^2_{t,s}$ and $t \in [T_0, s[$. Thus, (2.50) with $u = T_0$ follows from (2.51) and (2.52).

Now, to prove (2.50) for all $u \in [T_0, s[$, fix an arbitrary $u \in [T_0, s[$. Consider the measurable Q-function $q_u(x,t,B)$ defined on the time domain $[u, T_1[$ and satisfying $q_u(x,t,B) = q(x,t,B)$ for all $x \in X$, $t \in [u, T_1[$, and $B \in \mathfrak{B}(X)$. Observe that the Q-function q_u satisfies Assumption 2.2.3 with $T_0 = u$, and that formula (2.19) defines the same transition function $\bar{P}(w, x; t, B)$ on the domain $w \in [u, T_1[$, $t \in]w, T_1[$, $x \in X$, $B \in \mathfrak{B}(X)$ if q is replaced by q_u. Starting at the point u instead of T_0, we have from the above arguments that (2.50) holds. Since u is chosen arbitrarily, (2.50) holds for all $u \in [T_0, s[$. $\square$

Lemma 2.5.6 provides a simple statement that is useful to prove Lemma 2.5.7 and Corollaries 2.6.2, 2.6.3.

Lemma 2.5.6. *A function* $\hat{P}(u,x;t,B)$*, where* $u \in [T_0,T_1[$*,* $t \in]u,T_1[$*,* $x \in X$*, and* $B \in \mathfrak{B}(X)$*, is a solution of Kolmogorov's forward equation* (2.37) *if and only if the function* $\hat{P}$ *satisfies, for all* $t \in]T_0,T_1[$*,* $u \in [T_0,t[$*,* $x \in X$*, and* (q,t)*-bounded sets* $B \in \mathfrak{B}(X)$*,*

$$P(u,x;t,B) = I\{x \in B\} - \int_u^t ds \int_B q(y,s)P(u,x;s,dy) + \int_u^t ds \int_X q(y,s,B \setminus \{y\})P(u,x;s,dy). \tag{2.53}$$

Proof. Integrating (2.37) from u to t and by using the boundary condition (2.36), we get (2.53). Thus, it follows from properties (i) and (ii) in Definition 2.5.2 that a solution of Kolmogorov's forward equation (2.37) satisfies (2.53) for all $u \in [T_0,T_1[$, $t \in]u,T_1[$, $x \in X$, and (q,t)-bounded sets $B \in \mathfrak{B}(X)$.

Consider a function $\hat{P}(u,x;t,B)$ that satisfies (2.53) for all $t \in]T_0,T_1[$, $u \in [T_0,t[$, $x \in X$, and (q,t)-bounded sets $B \in \mathfrak{B}(X)$. For each $s \in]T_0,T_1[$, a (q,s)-bounded set is (q,t)-bounded for all $t \in [T_0,s[$. This property implies that, for all u,x,s,B such that $s \in]T_0,T_1[$, $x \in X$, $u \in [T_0,s[$, and the set B is (q,s)-bounded, the function $\hat{P}(u,x;t,B)$ satisfies (2.53) for all $t \in]u,s[$. Thus, for all u,x,s,B such that $s \in]T_0,T_1[$, $x \in X$, $u \in [T_0,s[$, and the set B is (q,s)-bounded, the function $\hat{P}(u,x;t,B)$ is locally absolutely continuous on $t \in]u,s[$ and (2.36) holds. That is, the function $\hat{P}$ satisfies property(i) of Definition 2.5.2. Since a locally absolutely continuous function is differentiable almost every where on its domain, differentiating (2.53) we have that the function $\hat{P}$ satisfies property (ii) of Definition 2.5.2. □

The following lemma is useful in proving the minimality property of $\bar{P}$ stated in Theorem 2.5.1.

Lemma 2.5.7. *Let Assumption 2.2.3 hold. Consider a non-negative solution* $\hat{P}(u,x;t,B)$ *of Kolmogorov's forward equation* (2.37)*. Then, for all* $t \in]T_0,T_1[$*,* $u \in [T_0,t[$*,* $x \in X$*, and* (q,t)*-bounded sets* $B \in \mathfrak{B}(X)$*, the function* $\hat{P}(u,x;t,B)$ *satisfies*

$$P(u,x;t,B) = \mathbf{I}\{x \in B\}e^{-\int_u^t q(x,s)ds} + \int_u^t ds \int_X \int_B q(y,s,dz \setminus \{y\})e^{-\int_s^t q(z,\theta)d\theta}P(u,x;s,dy). \tag{2.54}$$

Proof. Observe that: (i) for each u,x,t,

$$\{\hat{P}(u,x;t+\frac{1}{n},\cdot)\}_{n=1,2,\dots} \textit{ converge setwise to } \hat{P}(u,x;t,\cdot), \tag{2.55}$$

and (ii) for any (q,s)-bounded set $B \in \mathfrak{B}(X)$, there exists a natural number m such that

$$\int_t^{t+\frac{1}{n}} q(y,\theta)d\theta \le m/n, \qquad t \in [T_0,s[, n > 1/(s-t), y \in B. \tag{2.56}$$

Formula (2.55) is correct since, for each u,x,B, the function $\hat{P}(u,x;t,B)$ is absolutely continuous in t; see Definition 2.5.2(i), and Formula (2.56) follows from the definition of a (q,s)-bounded set.

Then, for all $x \in X$, $s \in]T_0, T_1[$, $u \in [T_0, s[$, (q,s)-bounded sets $B \in \mathfrak{B}(X)$, and $t \in]u,s[$,

$$\frac{\partial}{\partial t}\int_B e^{\int_u^t q(y,\theta)d\theta}\hat{P}(u,x;t,dy) = \lim_{n\to\infty}\frac{\int_B e^{\int_u^{t+\frac{1}{n}} q(y,\theta)d\theta}\hat{P}(u,x;t+\frac{1}{n},dy) - \int_B e^{\int_u^t q(y,\theta)d\theta}\hat{P}(u,x;t,dy)}{\frac{1}{n}}$$

$$= \lim_{n\to\infty} n\int_B e^{\int_u^t q(y,\theta)d\theta}((1+\int_t^{t+\frac{1}{n}} q(y,\theta)d\theta + \frac{(\int_t^{t+\frac{1}{n}} q(y,\theta)d\theta)^2}{2!} + \ldots)\hat{P}(u,x;t+\frac{1}{n},dy) - \hat{P}(u,x;t,dy))$$

$$= \lim_{n\to\infty} n\int_B e^{\int_u^t q(y,\theta)d\theta}(\hat{P}(u,x;t+\frac{1}{n},dy) - \hat{P}(u,x;t,dy))$$

$$+ \lim_{n\to\infty} n\int_B e^{\int_u^t q(y,\theta)d\theta}(\int_t^{t+\frac{1}{n}} q(y,\theta)d\theta)\hat{P}(u,x;t+\frac{1}{n},dy) = J_1 + J_2 + J_3, \quad (2.57)$$

where the first equality follows from the definition of a partial derivative, the second equality is obtained by using the power series expansion of the exponential function $e^{\int_t^{t+\frac{1}{n}} q(y,\theta)d\theta}$, the third equality is correct since the limit of the higher order terms tend to zero in the limit due to (2.55), (2.56), and Lebesgue dominated convergence theorem, and

$$J_1 = \lim_{n\to\infty} n\int_B e^{\int_u^t q(y,\theta)d\theta}\left(\hat{P}(u,x;t+\frac{1}{n},dy) - \hat{P}(u,x;t,dy)\right),$$

$$J_2 = \lim_{n\to\infty} n\int_B e^{\int_u^t q(y,\theta)d\theta}\left(\int_t^{t+\frac{1}{n}} q(y,\theta)d\theta\right)\hat{P}(u,x;t,dy),$$

$$J_3 = \lim_{n\to\infty} n\int_B e^{\int_u^t q(y,\theta)d\theta}\left(\int_t^{t+\frac{1}{n}} q(y,\theta)d\theta\right)\left(\hat{P}(u,x;t+\frac{1}{n},dy) - \hat{P}(u,x;t,dy)\right).$$

Observe that, for $x \in X$, $s \in]T_0, T_1[$, $u \in [T_0, s[$, and (q,s)-bounded set $B \in \mathfrak{B}(X)$,

$$J_1 = \lim_{n\to\infty} n\int_B e^{\int_u^t q(y,\theta)d\theta}\int_t^{t+\frac{1}{n}}\left(-q(y,v)\hat{P}(u,x;v,dy) + \int_X q(z,v,dy\setminus\{z\})\hat{P}(u,x;v,dz)\right)dv$$

$$= \lim_{n\to\infty} n\int_t^{t+\frac{1}{n}}\left(\int_B e^{\int_u^t q(y,\theta)d\theta}\left(-q(y,v)\hat{P}(u,x;v,dy) + \int_X q(z,v,dy\setminus\{z\})\hat{P}(u,x;v,dz)\right)\right)dv$$

$$= \int_B e^{\int_u^t q(y,\theta)d\theta}\left(-q(y,t)\hat{P}(u,x;t,dy) + \int_X q(z,t,dy\setminus\{z\})\hat{P}(u,x;t,dz)\right), \quad (2.58)$$

for almost every $t \in]u,s[$, where the first equality follows from Lemma 2.5.6, the second equality is correct due to Fubini's theorem, and the last one follows from Lebesgue differentiation theorem [32, Theorem 7.10], and

$$J_2 = \lim_{n\to\infty} n\int_t^{t+\frac{1}{n}}\left(\int_B q(y,v)e^{\int_u^t q(y,\theta)d\theta}\hat{P}(u,x;t,dy)\right)dv = \int_B q(y,t)e^{\int_u^t q(y,\theta)d\theta}\hat{P}(u,x;t,dy), \quad (2.59)$$

for almost every $t \in]u,s[$, where the first equality is correct due to Fubini's theorem and the second equality follows from Lebesgue differentiation theorem.

In addition, for $x \in X$, $s \in]T_0,T_1[$, $u \in [T_0,s[$, and (q,s)-bounded set $B \in \mathfrak{B}(X)$,

$$\begin{aligned} J_3 &= \lim_{n\to\infty} \int_B e^{\int_u^t q(y,\theta)d\theta} \left(n\int_t^{t+\frac{1}{n}} q(y,\theta)d\theta\right) \int_t^{t+\frac{1}{n}} (-q(y,v)\hat{P}(u,x;v,dy) + \int_X q(z,v,dy\setminus\{z\})\hat{P}(u,x;v,dz))dv \\ &= \lim_{n\to\infty} \int_t^{t+\frac{1}{n}} \left(\int_B e^{\int_u^t q(y,\theta)d\theta} \left(n\int_t^{t+\frac{1}{n}} q(y,\theta)d\theta\right)(-q(y,v)\hat{P}(u,x;v,dy) + \int_X q(z,v,dy\setminus\{z\})\hat{P}(u,x;v,dz))\right)dv \\ &= 0, \quad (2.60) \end{aligned}$$

for almost every $t \in]u,s[$, where the first equality follows from Lemma 2.5.6, the second equality is correct due to Fubini's theorem, and the third equality is true since (2.56) holds and since, for each (q,s)-bounded set $B \in \mathfrak{B}(X)$, the integrals $\int_B q(y,v)\hat{P}(u,x;v,dy)$ and $\int_X q(z,v,B\setminus\{z\})\hat{P}(u,x;v,dz)$ are finite for almost every $v \in]u,s[$ as the function $\hat{P}$ is a solution of (2.37).

From (2.57), (2.58), (2.59), and (2.60), for all $x \in X$, $s \in]T_0,T_1[$, $u \in [T_0,s[$, and (q,s)-bounded sets $B \in \mathfrak{B}(X)$,

$$\frac{\partial}{\partial t}\int_B e^{\int_u^t q(y,\theta)d\theta}\hat{P}(u,x;t,dy) = \int_B e^{\int_u^t q(y,\theta)d\theta}\int_X q(z,t,dy\setminus\{z\})\hat{P}(u,x;t,dz),$$

for almost every $t \in]u,s[$. Integrating the above equality from u to t and using the boundary condition (2.36), for all $x \in X$, $t \in]T_0,T_1[$, $u \in [T_0,t[$, and (q,t)-bounded sets $B \in \mathfrak{B}(X)$,

$$\int_B e^{\int_u^t q(y,\theta)d\theta}\hat{P}(u,x;t,dy) - \mathbf{I}\{x\in B\} = \int_u^t d\theta \int_B e^{\int_u^\theta q(y,v)dv}\int_X q(z,\theta,dy\setminus\{z\})\hat{P}(u,x;\theta,dz)$$

which implies that,

$$\begin{aligned} &\int_B e^{\int_u^t q(y,\theta)d\theta}\left(\hat{P}(u,x;t,dy) - \mathbf{I}\{x\in dy\}e^{-\int_u^t q(y,\theta)d\theta}\right) \\ &\qquad = \int_B e^{\int_u^t q(y,\theta)d\theta}\left(\int_u^t ds\int_X q(z,s,dy\setminus\{z\})e^{-\int_s^t q(y,\theta)d\theta}\hat{P}(u,x;s,dz)\right). \end{aligned}$$

Since $e^{\int_u^t q(y,\theta)d\theta} > 0$ for all u,t and since every measurable subset of a (q,t)-bounded set is also (q,t)-bounded, it follows from Radon-Nikodym theorem and the above equality that (2.54) holds for all $x \in X$, $t \in]T_0,T_1[$, $u \in [T_0,t[$, and (q,t)-bounded sets $B \in \mathfrak{B}(X)$. □

Proof of Theorem 2.5.1. First, we show that the function $\bar{P}$ is a solution of Kolmogorov's forward equation (2.37). It follows from the arguments given in the proof of Theorem 2.5.3(a) that, for all u,x,B such that $u \in [T_0,T_1[$, $x \in X$, and $B \in \mathfrak{B}(X)$, the function $\bar{P}$ is an absolutely continuous function on $t \in]u,T_1[$ and (2.36) holds. Thus, the function $\bar{P}$ satisfies property (i) of Definition 2.5.2. Though it was assumed in Theorem 2.5.3(a) that Assumption 2.2.2 holds, the arguments there are correct for a Q-function satisfying Assumption 2.2.4. In view of Lemma 2.2.1(iii), the Q-function q satisfies Assumption 2.2.4 since Assumption 2.2.3 holds.

To prove that the function $\bar{P}$ satisfies property (ii) of Definition 2.5.2, fix an arbitrary $s \in]T_0,T_1[$ and consider the Q-function q^s defined in (2.46). This Q-function satisfies Assumption 2.2.2; Lemma 2.5.4(a). Then it follows from Theorem 2.5.3(b) that, for all $u \in [T_0,T_1[$, $x \in X$, and q^s-bounded sets $B \in \mathfrak{B}(X)$, the function $\bar{P}^s(u,x;t,B)$ satisfies (2.37) with $s = T_1$ and with the Q-function q replaced by q^s. This fact and (2.47) imply that, for all $u \in [T_0,s[$, $x \in X$, and q^s-bounded sets $B \in \mathfrak{B}(X)$,

$$\frac{\partial}{\partial t}\bar{P}(u,x;t,B) = -\int_B q^s(y,t)\bar{P}(u,x;t,dy) + \int_X q^s(y,t,B\setminus\{y\})\bar{P}(u,x;t,dy) \text{ for almost every } t \in]u,s[. \tag{2.61}$$

Observe that a set is q^s-bounded if and only if it is (q,s)-bounded. Suppose that a set $B \in \mathfrak{B}(X)$ is q^s-bounded. Then there exists a natural number n such that $q^s(x,t) < n$ for all $x \in B$ and $t \in [T_0,T_1[$. This fact and (2.48) imply that

$$q(x,t)\mathbf{I}\{q(x,t) < N(x,s)\} < n, \qquad x \in B, t \in [T_0,s[.$$

Thus, we have from the above inequality and (2.49) that $B \subseteq X_n^s$, and therefore (q,s)-bounded. Now, suppose that a set $B \in \mathfrak{B}(X)$ is (q,s)-bounded. Then there exists a natural number n such that $B \subseteq X_n^s$, and therefore, $N(x,s) \le n$ for all $x \in B$. This fact and (2.48) imply that the set B is q^s-bounded. Then for all $u \in [T_0,s[$, $x \in X$, and (q,s)-bounded sets $B \in \mathfrak{B}(X)$,

$$\begin{aligned}
\frac{\partial}{\partial t}\bar{P}(u,x;t,B) = &-\int_{B^1_{t,s}} q^s(y,t)\bar{P}(u,x;t,dy) + \int_{Y^1_{t,s}} q^s(y,t,B\setminus\{y\})\bar{P}(u,x;t,dy) \\
&-\int_{B^2_{t,s}} q^s(y,t)\bar{P}(u,x;t,dy) + \int_{Y^2_{t,s}} q^s(y,t,B\setminus\{y\})\bar{P}(u,x;t,dy) \\
= &-\int_{B^1_{t,s}} q(y,t)\bar{P}(u,x;t,dy) + \int_{Y^1_{t,s}} q(y,t,B\setminus\{y\})\bar{P}(u,x;t,dy) \\
= &-\int_{B} q(y,t)\bar{P}(u,x;t,dy) + \int_{X} q(y,t,B\setminus\{y\})\bar{P}(u,x;t,dy),
\end{aligned}$$

for almost every $t \in]u,s[$, where the sets $B^1_{t,s}, B^2_{t,s}, Y^1_{t,s}, Y^2_{t,s}$ are defined prior to Lemma 2.5.5, and the first equality follows from (2.61), the second equality follows from (2.46) and (2.50), and the last one is correct due to (2.50). Thus, the function $\bar{P}$ satisfies property (ii) from Definition 2.5.2, and is therefore a solution of Kolmogorov's forward equation (2.37).

To show the minimality property of $\bar{P}$, consider a non-negative solution P^* of Kolmogorov's forward equation (2.37). Due to Lemma 2.5.7, the function $P^*(u,x;t,B)$ satisfies (2.54) for all $t \in]T_0,T_1[$, $u \in [T_0,t[$, $x \in X$, and (q,t)-bounded sets $B \in \mathfrak{B}(X)$. Since the last term of (2.54) is non-negative, for all $t \in]T_0,T_1[$, $u \in [T_0,t[$, $x \in X$, and (q,t)-bounded sets $B \in \mathfrak{B}(X)$,

$$P^*(u,x;t,B) \ge I\{x \in B\}e^{-\int_u^t q(x,s)ds} = \bar{P}^{(0)}(u,x;t,B), \tag{2.62}$$

where the last equality is (2.17). For all $t \in]T_0,T_1[$, $u \in [T_0,t[$, $x \in X$, and (q,t)-bounded sets

$B \in \mathfrak{B}(X)$, assume $P^*(u,x;t,B) \geq \sum\limits_{m=0}^{n} \bar{P}^{(m)}(u,x;t,B)$ for some $n \geq 0$. This fact, (2.54) for the function P^*, and (2.39) imply that

$$P^*(u,x;t,B) \geq I\{x \in B\}e^{-\int_u^t q(x,s)ds} + \int_u^t ds \int_X \int_B q(y,s,dz \setminus \{y\}) e^{-\int_s^t q(z,\theta)d\theta} \sum_{m=0}^{n} \bar{P}^{(m)}(u,x;s,dy)$$
$$= \bar{P}^{(0)}(u,x;t,B) + \sum_{m=0}^{n} \bar{P}^{(m+1)}(u,x;t,B) = \sum_{m=0}^{n+1} \bar{P}^{(m)}(u,x;t,B).$$

Thus, by induction, $P^*(u,x;t,B) \geq \sum_{m=0}^{n} \bar{P}^{(m)}(u,x;t,B)$ for all $n \geq 0$, $t \in]T_0,T_1[$, $u \in [T_0,t[$, $x \in X$, and (q,t)-bounded sets $B \in \mathfrak{B}(X)$. This property and (2.19) imply that, for all $t \in]T_0,T_1[$, $u \in [T_0,t[$, and $x \in X$,

$$P^*(u,x;t,B) \geq \bar{P}(u,x;t,B) \qquad \text{for all } (q,t)\text{-bounded sets } B \in \mathfrak{B}(X). \tag{2.63}$$

Then for all $t \in]T_0,T_1[$, $u \in [T_0,t[$, and $x \in X$

$$P^*(u,x;t,B) = \lim_{n\to\infty} P^*(u,x;t,B \cap X_n^t) \geq \lim_{n\to\infty} \bar{P}(u,x;t,B \cap X_n^t) = \bar{P}(u,x;t,B), \quad B \in \mathfrak{B}(X), \tag{2.64}$$

where the first and last equalities are correct since the sets $X_n^t \uparrow X$ as $n \to \infty$ due to Assumption 2.2.3 and the functions $\bar{P}$ and P^* are measures on $(X,\mathfrak{B}(X))$ for fixed u,x,t, and the inequality holds due to (2.63). Thus, $\bar{P}$ is the minimal non-negative solution of Kolmogorov's forward equation (2.37).

To prove the uniqueness property of $\bar{P}$, let the solution P^* take values in $[0,1]$. If $\bar{P}(u,x;t,X) = 1$ for all u,x,t, then the uniqueness of $\bar{P}$ within the set of non-negative solutions with values in $[0,1]$ follows from (2.64) and from the same arguments as in the proof of uniqueness in Theorem 2.4.1. □

2.6 Additional results and comments

Assumption 2.2.2 means that the jump intensities are bounded at each state for the time horizon $[T_0,T_1[$, and this assumption is natural for continuous-time Markov decision processes (CTMDPs); see Feinberg et al. [8, 9]. In this section, we present particular results on the minimality and uniqueness properties of the solution $\bar{P}$ of Kolmogorov's forward equation under Assumption 2.2.2.

Theorem 2.6.1. *Let Assumption 2.2.2 hold. The function $\bar{P}(u,x;t,B)$ is the minimal non-negative function that satisfies, for each u,x,B such that $u \in [T_0,T_1[$, $x \in X$, and the set B is q-bounded, properties (a) and (b) given in Theorem 2.5.3. In addition, if $\bar{P}$ is a regular transition function, then $\bar{P}$ is the unique non-negative function satisfying properties (a) and (b) given in Theorem 2.5.3 and takes values in $[0,1]$.*

Proof. According to Theorem 2.5.3, the function $\bar{P}$ satisfies properties (a) and (b) given in Theorem 2.5.3. To prove the minimality of $\bar{P}$, consider a non-negative function P^* that satisfies properties (a) and (b) given in Theorem 2.5.3. Recall that a q-bounded set is (q,t)-bounded for all

$t \in]T_0, T_1[$. Then, in view of Theorem 2.5.3, it follows from the arguments given in Lemma 2.5.7 that, for all $t \in]T_0, T_1[$, $u \in [T_0, t[$, and $x \in X$, formula (2.54) holds for all q-bounded sets $B \in \mathfrak{B}(X)$. Therefore, it follows from Lemma 2.2.1(ii) and from the arguments given as in the proof of minimality in Theorem 2.5.1 that (2.63) holds, and therefore (2.64) holds for all $B \in \mathfrak{B}(X)$. The proof of uniqueness property of $\bar{P}$ is the same as the proof of uniqueness in Theorem 2.4.1. $\square$

The following two corollaries are useful for applying the results of this paper to continuous-time jump Markov decision processes; Feinberg et al. [8, Theorem 3.2] and Theorem 4.3.1.

Corollary 2.6.2. *Under Assumption 2.2.2, the following statements hold:*

(a) for all $x \in X$, $t \in]T_0, T_1[$, $u \in [T_0, t]$, and q-bounded sets $B \in \mathfrak{B}(X)$, the function $\bar{P}(u,x;t,B)$ satisfies (2.53).

(b) the function $\bar{P}$ is the minimal non-negative function for which statement (a) holds. In addition, if $\bar{P}$ is a regular transition function, then $\bar{P}$ is the unique non-negative function with values in $[0,1]$ for which statement (a) holds.

Proof. Integrating (2.37) from u to t and by using the boundary condition (2.36), we get (2.53). Thus, statements (a) and (b) of this corollary follow respectively from Theorem 2.5.3(a) and (b), and Theorem 2.6.1. $\square$

When x is fixed and $u = T_0$, (2.53) is an equation in two variables t and B. Hence, for simplicity, we write $P(t,B)$ instead of $P(T_0,x;t,B)$ for any function P on the domain of $\bar{P}$ when x is fixed and $u = T_0$, and (2.53) will be given as

$$P(t,B) = I\{x \in B\} + \int_{T_0}^{t} ds \int_X q(y,s,B \setminus \{y\}) P(s,dy) - \int_{T_0}^{t} ds \int_B q(y,s) P(s,dy). \qquad (2.65)$$

For fixed $x \in X$ and $u = T_0$, the function $\bar{P}(t,\cdot)$ is the marginal probability distribution on the state of the process $\{\mathbb{X}_t, t \in [T_0, T_1[\}$ given $\mathbb{X}_{T_0} = x$. Under Assumption 2.2.2, the following corollary describes the properties of the solution $\bar{P}(t,B)$ to (2.65).

Corollary 2.6.3. *Fix an arbitrary $x \in X$. Under Assumption 2.2.2, the following statements hold:*
(a) for all $t \in]T_0, T_1[$ and q-bounded sets $B \in \mathfrak{B}(X)$, the function $\bar{P}(t,B)$ satisfies (2.65)*;*
(b) the function $\bar{P}(t,B)$ is the minimal non-negative function for which statement (a) holds. In addition, if the intensities $q(z,t)$ are uniformly bounded in z and t, then $\bar{P}(t,B)$ is the unique non-negative function with values in $[0,1]$ for which statement (a) holds.

Proof. Statement (a) of the corollary follows immediately from Corollary 2.6.2(a) when $u = T_0$. To prove the minimality of the function $\bar{P}(t,B)$, consider a non-negative function $P^*(t,B)$, where $t \in]T_0, T_1[$ and $B \in \mathfrak{B}(X)$, satisfying (2.65) for each $t \in]T_0, T_1[$ and q-bounded set $B \in \mathfrak{B}(X)$. Define the function $f(u,z;t,B)$ with the same domain as $\bar{P}(u,z;t,B)$,

$$f(u,z;t,B) = \begin{cases} P^*(t,B), & \text{if} \quad u = T_0 \text{ and } z = x, \\ \bar{P}(u,z;t,B), & \text{otherwise.} \end{cases} \qquad (2.66)$$

Then, it follows from Corollary 2.6.2(a) and (2.66) that statement (a) of Corollary 2.6.2 holds for the function f. This fact, Corollary 2.6.2(b), and (2.66) imply that

$$P^*(t,B) = f(T_0,x;t,B) \geq \bar{P}(T_0,x;t,B) = \bar{P}(t,B), \qquad t \in]T_0,T_1[, B \in \mathfrak{B}(X). \tag{2.67}$$

Thus, the function $\bar{P}(t,B)$ is the minimal non-negative function for which statement (a) of this corollary holds.

To show the uniqueness property, let the function P^* take values in $[0,1]$. This fact and the property that the function $\bar{P}(u,z;t,B)$ takes values in $[0,1]$ for all u,z,t,B in the domain of $\bar{P}$ imply that the function f defined in (2.66) takes values in $[0,1]$. Observe that, if the intensities $q(z,t)$ are uniformly bounded in z and t, X is a q-bounded set. Then, from Corollary 2.6.2(a), we have $\bar{P}(u,z;t,X) = 1$ for all $u,t \in [T_0,T_1[$ with $u < t$ and $z \in X$. Therefore, it follows from Corollary 2.6.2(b) that $f(u,z;t,B) = \bar{P}(u,z;t,B)$ for all u,z,t,B in the domain of $\bar{P}$, which along with (2.67) implies the uniqueness property of $\bar{P}(t,B)$. □

2.6.1 Non-conservative Q-functions

The results of this chapter can be extended to non-conservative Q-functions. As mentioned in section 2.1, any non-conservative Q-function q can be transformed into a conservative Q-function by adding a state $\bar{x}$ to X with $q(x,t,\{\bar{x}\}) := -q(x,t,X)$, $q(\bar{x},t,X) := 0$, and $q(\bar{x},t,\{\bar{x}\}) := 0$, where $x \in X$ and $t \in \mathbb{R}_+$. According to Theorem 2.3.2, there is a transition function $\bar{P}$ of a jump Markov process with the state space $X \cup \{\bar{x}\}$, and this process is determined by the initial state distribution and by the compensator defined by the modified Q-function.

The proofs of the results of sections 2.4 and 2.5 do not use the assumption that the Q-function q is conservative. Therefore, these results remain valid for non-conservative Q-functions. However, the validity of the condition $\bar{P}(u,x;t,X) = 1$ for all x,u,t with $u < t$ in Theorems 2.4.1 and 2.5.1 is possible only if $q(x,t,X) = 0$ almost everywhere in t for each $x \in X$. Thus, in fact, q is conservative if $\bar{P}(u,x;t,X) = 1$ for all x,u,t with $u < t$. It is also easy to see that the minimal solutions of both Kolmogorov's backward and forward equations are equal to $\bar{P}(u,x;t,B)$, when $x \in X$ and $B \in \mathfrak{B}(X)$, where the transition function $\bar{P}$ is described in the previous paragraph for a broader domain.

2.6.2 Generalized boundedness assumptions

For some Q-functions, Assumptions 2.2.3 and 2.2.4 may not hold if the point T_1 is included. To study Kolmogorov's equations for such Q-functions, we excluded the point T_1 and defined the solution $\bar{P}(u,x;t,B)$ for all $T_0 \leq u < t < T_1$. Similarly, it is also possible to consider the situation when Assumptions 2.2.3 and 2.2.4 do not hold from the point T_0 but hold from any point $u > T_0$, as described below.

For $n = 1,2,\ldots$, consider the functions $Z_n : X \times [T_0,T_1] \times [T_0,T_1] \to [0,\infty]$ defined as

$$Z_n(x,u,t) = \int_u^t \mathbf{I}\{q(x,s) \geq n\} ds, \qquad x \in X, u,t \in [T_0,T_1], u < t. \tag{2.68}$$

For each $u,t \in [T_0,T_1]$ with $u<t$, let $X_n^{u,t}, n=1,2,\ldots$, be the subsets of X such that

$$X_n^{u,t} = \{x \in X : Z_n(x,u,t) = 0\}, \qquad n = 1,2,\ldots. \tag{2.69}$$

Since the functions $Z_n(x,u,t)$ are measurable, for each u,t the sets $X_n^{u,t}$ are measurable subsets of X. Observe that $X_n^{u,t} \subseteq X_{n+1}^{u,t}$, $n=1,2,\ldots$. A set $B \in \mathfrak{B}(X)$ is (q,u,t)-bounded if $B \subseteq X_n^{u,t}$ for some $n=1,2,\ldots$.

Consider the following weak boundedness assumptions:

Assumption 2.6.1 (General almost everywhere boundedness of q)**.** *$X_n^{u,t} \uparrow X$ as $n \to \infty$ for each $u,t \in]T_0,T_1[$ with $u<t$.*

Assumption 2.6.2 (General $\mathscr{L}^1$ boundedness of q)**.** *For all $x \in X$, the integral $\int_u^t q(x,s)ds < \infty$ for each $u,t \in]T_0,T_1[$ with $u<t$.*

Under Assumption 2.6.2, the transition function $\bar{P}(u,x;t,B)$ is well defined for all u,t satisfying $T_0 < u < t < T_1$, but the stochastic process $\{\mathbb{X}_t, t \in [T_0,T_1[\}$ defined in (2.7) may not be defined starting from $t = T_0$. The construction of the stochastic process $\{\mathbb{X}_t, t \in [T_0,T_1[\}$ is based on Jacod [17, Theorem 3.6], which requires that the function $\nu(\omega;[T_0,t],B)$ defined in (2.5) is finite for all $t \in]T_0,T_1[$ and $B \in \mathfrak{B}(X)$. This might not be the case if the Q-function does not satisfy Assumption 2.2.4 as demonstrated by the following example.

Example 2.6.4. *The random measure $\nu(\omega;[T_0,t],B)$ defined in* (2.5) *may not be finite when the Q-function satisfies Assumption 2.6.2.* Let $T_0 = 0$, $T_1 = \infty$, $X = \{1,2\}$, $q(i,t,j) = q(i,t) = \frac{1}{t}\mathbf{I}\{t>0\}$ for all $i,j \in X$ with $i \neq j$. From (2.5),

$$\nu(\omega;[0,t],X) = \int_0^t \sum_{n\geq 0} \mathbf{I}\{t_n \leq s < t_{n+1}\} q(x_n,s)ds = \int_0^t \frac{1}{s}ds = \infty, \qquad t \in]0,\infty[.$$

Observe that Assumptions 2.6.1 and 2.6.2 are respectively the same as Assumptions 2.2.3 and 2.2.4 with $T_0 = u$ for all $u \in]T_0,T_1[$. Hence, results similar to those in Sections 2.4 and 2.5 on Kolmogorov's equations remain valid under Assumptions 2.6.1 and 2.6.2. Under Assumption 2.6.2, the following statements hold:

(a) the function $\bar{P}(u,x;t,B)$ satisfies property (i) from Definition 2.4.2;

(b) for each x,t,B, such that $x \in X$, $t \in]T_0,T_1[$, and $B \in \mathfrak{B}(X)$, the function $\bar{P}(u,x;t,B)$ satisfies Kolmogorov's backward equation (2.29) for almost every $u \in]T_0,t[$.

In addition, $\bar{P}$ is the minimal non-negative function satisfying statements (a), (b) given above. If $\bar{P}(u,x;t,X) = 1$, then it is the unique non-negative function that satisfies these statements, takes values in $[0,1]$, and is a measure on $(X,\mathfrak{B}(X))$ for fixed u,x,t. This is true because any function $P^*(u,x;t,B)$ satisfying these statements can be extended to a solution of Kolmogorov's backward equation (2.29) by defining $P^*(T_0,x;t,B) = \mathbf{I}\{x \in B\}$ for all $x \in X$, $t \in]T_0,T_1[$, and $B \in \mathfrak{B}(X)$.

Under Assumption 2.6.1, the following statements hold:

(a) for each $x \in X, u \in]T_0,T_1[$, and $B \in \mathfrak{B}(X)$, the function $\bar{P}(u,x;t,B)$ satisfies property (i) from Definition 2.5.2;

(b) For each w, s, u, x, B, such that $w \in]T_0, T_1[, s \in]w, T_1[, u \in [w, s[$, $x \in X$, and the set B is (q, w, s)-bounded, the function $\bar{P}(u, x; t, B)$ satisfies Kolmogorov's forward equation (2.37).
In addition, $\bar{P}$ is the minimal non-negative function satisfying statements (a), (b) given above. This is true because any function $P^*(u, x; t, B)$ satisfying these statements is a solution of Kolmogorov's backward equation (2.29) on the semi-interval $[w, T_1[$ for each $w \in]T_0, T_1[$. Since a Q-function satisfying Assumption 2.6.1 satisfies Assumption 2.2.3 with $T_0 = w$ for all $w \in]T_0, T_1[$, we have from Theorem 2.5.1 that, for all $w \in]T_0, T_1[$,

$$P^*(u, x; t, B) \geq \bar{P}(u, x; t, B), \qquad u \in [w, T_1[, t \in]w, T_1[, x \in B, B \in \mathfrak{B}(X). \tag{2.70}$$

Therefore, $\bar{P}(u, x; t, B)$ is the minimal non-negative function for which statements (a) and (b) given above hold. In fact, if $\bar{P}(u, x; t, X) = 1$ then it is the unique non-negative function with values in $[0, 1]$ satisfying these statements because (2.70) holds with an equality.

Chapter 3

Continuous-time Markov decision processes

3.1 Introduction

A Markov decision process (MDP) is a mathematical framework used for sequential decision making under uncertainty. It is used for modelling decision-making (or control) problems whose uncontrolled version is a Markov chain. Hence, MDPs are also known as controlled Markov chains. They are commonly used in many fields, such as inventory control, queuing systems, manufacturing, system maintenance, appointment scheduling in healthcare, population models etc, to improve the performance characteristics.

In MDPs, the decision maker chooses actions/decisions to influence the evolution of underlying Markov chain to minimize certain long term costs. Depending on when the actions are chosen, MDPs can be classified into two types:

(i) Discrete-time MDP (DTMDP): Actions are chosen at fixed time points, say at $t = 0, 1, 2, \ldots$, and the uncontrolled version of the MDP is a discrete-time Markov chain.

(ii) Continuous-time MDP (CTMDP): Actions can be chosen at any time, say at any $t \in [0, \infty[$, and the uncontrolled version of the MDP is a continuous-time Markov chain.

In this chapter, we formally introduce CTMDPs, provide a brief construction of the stochastic processes induced by a policy, and define the optimality criteria that we are interested in. The reader can refer to Appendix 5 for a similar description of DTMDPs.

3.2 The CTMDP model

A CTMDP is defined by the multiplet $(X, A, A(x), q(\cdot|x,a), c(x,a), C(x,a,y))$, where

(i) X is the state space such that $(X, \mathfrak{B}(X))$ is a standard Borel space;

(ii) A is the action space such that $(A, \mathfrak{B}(A))$ is a standard Borel space;

(iii) $A(x)$ are the set of actions available at $x \in X$. It is assumed that $A(x) \in \mathfrak{B}(X)$ for all $x \in X$ and the set of feasible state-action pairs $Gr(A) = \{(x,a) : x \in X, a \in A(x)\}$ is a measurable subset of $(X \times A)$ containing the graph of a measurable mapping from X to A.

(iv) $q(\cdot|x,a)$ is the transition rate from $Gr(A)$ to X. It is a signed measure on $(X, \mathfrak{B}(X))$ for any $(x,a) \in Gr(A)$ such that $q(X|x,a) = 0$, $0 \leq q(Z \setminus \{x\}|x,a) < \infty$ for all $Z \in \mathfrak{B}(X)$, and $q(Z|x,a)$ is a measurable function on $Gr(A)$ for each $Z \in \mathfrak{B}(X)$.

(v) $c(x,a)$ is the cost rate incurred for choosing an action $a \in A(x)$ in state $x \in X$ and is assumed to be a bounded below measurable function on the $Gr(A)$.

(vi) $C(x,a,y)$ is the instantaneous cost incurred if the process jumps from state x to state y and action a was chosen at the jump epoch. It is assumed to be a bounded below measurable function on $Gr(A) \times X$.

For CTMDPs, it is possible to choose actions any time. At each time $t \in \mathbb{R}_+$, the decision maker observes the current state x of the stochastic system and chooses a particular action a from the set of actions $A(x)$ available at x. We give an informal description of the evolution of the system. Suppose that an action a is chosen in state x at time t. Then in the infinitesimal time interval $[t, t+dt]$, the decision maker incurs the cost $c(x,a)dt$ and the system transitions from state x to state $y \neq x$ under the control a with probability $q(y|x,a)dt + o(dt)$ or it stays in the state x with probability $1 - q(X \setminus \{x\}|x,a)dt + o(dt)$. If the transition occurs, the decision maker incurs the instantaneous cost $C(x,a,y)$. At each time t, the decision maker can also choose a probability distribution on the set of available actions $A(x)$. Such decisions are commonly called randomized. If the decision maker chooses a randomized action, or in other words chooses a probability distribution on the set of available actions, then the system evolves as if the decision maker choose an action whose associated transition rate, cost rate, and instantaneous cost are expectation of the corresponding values with respect to the measure defined by the randomized action. Intuitively, this means that, for any two actions a and b in state x and for any constant $\lambda \in (0,1)$ there is an action d in state x such that $q(\cdot|x,d) = \lambda q(\cdot|x,a) + (1-\lambda)q(\cdot|x,b)$, $c(x,d) = \lambda c(x,a) + (1-\lambda)c(x,b)$, and $C(x,d,y) = \lambda C(x,a,y) + (1-\lambda)C(x,b,y)$ for all $y \neq x$. This definition for randomized actions simply relaxes the control set rather than chose an action randomly from the set of actions available. Thus, we think the term 'relaxed' is more appropriate to describe such actions. the system spends in state x a random amount of time (also called the sojourn time) that has an exponential distribution with rate $q(x,a) := -q(\{x\}|x,a)$ for all $(x,a) \in Gr(A)$ and then makes the transition in to the next state. In particular, on the infinitesimal time interval $[t, t+dt]$: We give a simple example showing the various parameters corresponding to the chosen action.

Example 3.2.1. Consider the 3 state CTMDP in Figure 3.2.1: Then,

action	transition rate	cost rate	instantaneous costs
a	$q(2|1,a) = 1$, $q(3|1,a) = 0$	5	10
b	$q(2|1,b) = 0$, $q(3|1,b) = 2$	10	5
$d : p(a) = 0.4$	$q(2|1,d) = 0.4$, $q(3|1,d) = 1.2$	8	7

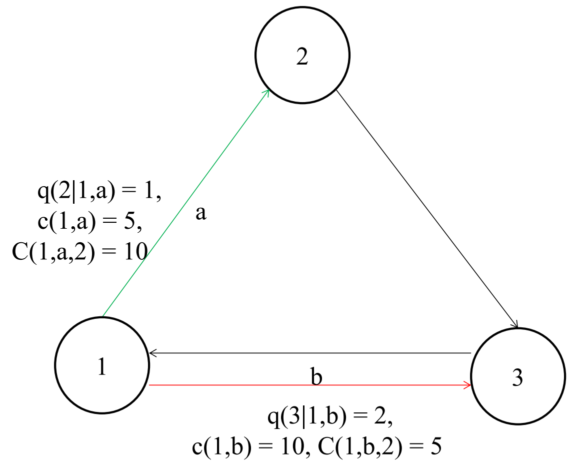

Figure 3.1: Example of a CTMDP

3.3 Jump process induced by a policy

In this section, we define different policies or decision rules considered in this thesis and provide a brief description of the construction of the jump process induced by them.

Adjoin an isolated point x_∞ to X, and let $\overline{X} := X \cup \{x_\infty\}$. Consider the Borel σ-algebra $\mathfrak{B}(\overline{X}) := \sigma(\mathfrak{B}(X), \{x_\infty\})$ on $\overline{X}$, which is the minimal σ-algebra containing $B(X)$ and $\{x_\infty\}$. Let $(\overline{X} \times \overline{\mathbb{R}}_+)^\infty$, where $\overline{\mathbb{R}}_+ :=]0,\infty]$, be the set of all sequences $(x_0, t_1, x_1, t_2, x_2, \ldots)$ with $x_n \in \overline{X}$ and $t_{n+1} \in \overline{\mathbb{R}}_+$ for $n = 0, 1, \ldots$. This set is endowed with the σ-algebra defined by the products of the Borel σ-algebras $\mathfrak{B}(\overline{X})$ and $\mathfrak{B}(\overline{\mathbb{R}}_+)$.

Denote by Ω the subset of all sequences $\omega = (x_0, t_1, x_1, t_2, x_2, \ldots)$ from $(\overline{X} \times \overline{\mathbb{R}}_+)^\infty$ such that: (i) $x_0 \in X$; (ii) for $n = 1, 2, \ldots$, $t_{n+1} > t_n$ if $t_n < \infty$ and $t_{n+1} = t_n$ if $t_n = \infty$; and, (iii) for $n = 1, 2, \ldots$, $x_n = x_\infty$ if and only if $t_n = \infty$. Observe that Ω is a measurable subset of $(\overline{X} \times \overline{\mathbb{R}}_+)^\infty$. Consider the measurable space $(\Omega, \mathscr{F})$, where $\mathscr{F}$ is the σ-algebra of the measurable subsets of Ω. Define the random variables $x_n(\omega) = x_n$, $t_{n+1}(\omega) = t_{n+1}$, $n = 0, 1, \ldots$, $t_0(\omega) = 0$, and $t_\infty(\omega) = \lim_{n\to\infty} t_n$ on the measurable space $(\Omega, \mathscr{F})$. Throughout our study on CTMDPs, we omit ω whenever possible. For all $t \in \mathbb{R}_+$ let $\mathscr{F}_t := \sigma(\mathfrak{B}(X), \mathscr{G}_t)$, where $\mathscr{G}_t := \sigma(I\{x_n \in Z\}I\{t_n \le s\} : n \ge 1, s \in [0,t], Z \in \mathfrak{B}(X))$, and let $\mathscr{P}$ be the predictable σ-algebra on $\Omega \times \mathbb{R}_+$ that is generated by sets of the form $\Gamma \times \{0\}$, $\Gamma \in \mathscr{F}_0$, and, $\Gamma \times]s,t]$, $\Gamma \in \mathscr{F}_s, s, t \in \mathbb{R}_+$ with $s < t$.

The jump process of interest, $\{\xi_t : t \in \mathbb{R}_+\}$, defined on $(\Omega, \mathscr{F})$ and adapted to the filtration $\{\mathscr{F}_t, t \in \mathbb{R}_+\}$ is given by

$$\xi_t(\omega) := \sum_{n\ge 0} I\{t_n \le t < t_{n+1}\}x_n + I\{t_\infty \le t\}x_\infty. \tag{3.1}$$

Along the trajectory ω, observe that $\xi_t(\omega)$ is right continuous piecewise constant in t and $\xi_{t^-}(\omega) =$

$\xi_t(\omega)$ (where $\xi_{0^-}(\omega) := \xi_0(\omega)$) for all $t \in \mathbb{R}_+$ except for a countable set of jump times $t_n, n = 1, 2, \ldots$. Thus, for notational convenience, we shall often replace ξ_{t^-} with ξ_t whenever this does not lead to a confusion.

For a standard Borel space $(S, \mathfrak{B}(S))$, denote by $\mathscr{P}(S)$ the set of all probability measures on $(S, \mathfrak{B}(S))$. We now describe the different classes of policies considered in this thesis. Adjoin an isolated point a_∞ to A, and let $\overline{A} := A \cup \{a_\infty\}$ and $A(x_\infty) := a_\infty$. Consider the Borel σ-algebra $\mathfrak{B}(\overline{A}) := \sigma(\mathfrak{B}(A), \{a_\infty\})$ on $\overline{A}$.

- A relaxed policy π is a transition probability from $(\Omega \times \mathbb{R}_+, \mathscr{P})$ to $(\overline{A}, \mathfrak{B}(\overline{A}))$ such that $\pi(A(\xi_{t^-})|\omega, t) = 1$.

- A relaxed policy φ is called a *relaxed Markov policy* if $\varphi(\cdot|\omega, t) = \varphi(\cdot|\xi_{t^-}, t)$ for $t \in \mathbb{R}_+$.

Observe that a relaxed policy π (respectively, a relaxed Markov policy φ) is a $\mathscr{P}$-measurable mapping from $\Omega \times \mathbb{R}_+$ (respectively, from $\overline{X} \times \mathbb{R}_+$) to the metric space $\mathscr{P}(\overline{A})$ that is endowed with the topology of weak convergence. It follows from Jacod [17, Lemma 3.3] that the $\mathscr{P}$-measurability of the relaxed policy π is equivalent to the existence of transition probabilities π_n, $n = 0, 1, \ldots$, from $(\Omega \times \mathbb{R}_+, \mathscr{F}_{t_n} \times \mathfrak{B}(\mathbb{R}_+))$ to $(A, \mathfrak{B}(A))$ concentrated on $A(x_n)$ such that, for all $n = 0, 1, \ldots$,

$$\pi(\cdot|\omega, t) = \pi_n(\cdot|x_0, t_1, x_1, \ldots, t_n, x_n, t - t_n) \quad \text{on} \quad \{t_n < t \leq t_{n+1}\}.$$

Hence, our terminology here is consistent with Feinberg [7].

3.3.1 Kitaev's construction of the probability measure defined by a policy

In this subsection, we provide a brief description of Kitaev's construction [24] of the probability measure of the jump process $\{\xi_t, t \in \mathbb{R}_+\}$ controlled by an arbitrary policy π.

Let $q(z,a) := q(X \setminus \{x\}|z, a)$ for all $(z,a) \in Gr(A)$ and $\overline{q}(z) := \sup_{a \in A(z)} q(z,a)$ for all $z \in X$. For all $Z \in \mathfrak{B}(X)$, $z \in \overline{X}$ and $p \in \mathscr{P}(\overline{A})$, let

$$q(Z|z,p) := \int_{A(z)} q(Z|z,a)p(da)\mathbf{I}\{z \in X\} \quad \text{and} \quad q(z,p) := q(X \setminus \{z\}|z,p). \tag{3.2}$$

Here, following the tradition, we use the same notation q on either side of the definitions in (3.2). The following assumption is necessary to define and analyse CTMDPs with relaxed actions, and we assume that it holds throughout our study on CTMDPs. In particular, for each $z \in X$, it guarantees that $q(Z|z,p) < \infty$ for all $Z \in \mathfrak{B}(X)$ and $p \in \mathscr{P}(\overline{A})$.

Assumption 3.3.1. $\overline{q}(z) < \infty$ *for each* $z \in X$.

For each policy π, let π_t be the measure corresponding to the probability distribution $\pi(\cdot|\omega, t)$. Then the random measure υ^π given by

$$\upsilon^\pi(\omega; [0,t], Z) := \int_0^t q(Z \setminus \{\xi_s\}|\xi_s, \pi_s)ds, \qquad t \in \mathbb{R}_+, Z \in \mathfrak{B}(X), \tag{3.3}$$

is predictable and $\upsilon^\pi(\{t\},X) = \upsilon^\pi([t_\infty,\infty[,X) = 0$; see, e.g., Kitaev [24]. According to Jacod [17], the predictable random measure υ^π defined in (3.3) and a probability distribution γ on $(X,\mathfrak{B}(X))$ define a unique probability measure $\mathbb{P}_\gamma^\pi$ on $(\Omega,\mathscr{F})$ such that $\mathbb{P}_\gamma^\pi(\xi_0 \in Z) = \gamma(Z)$, $Z \in \mathfrak{B}(X)$, and υ^π is the compensator (predictable projection) of the random measure μ on $(\mathbb{R}_+ \times X)$,

$$\mu(\omega;[0,t],Z) = \sum_{n\geq 1} I\{t_n \in [0,t]\}I\{x_n \in Z\}, \qquad t \in \mathbb{R}_+, Z \in \mathfrak{B}(X), \tag{3.4}$$

associated with the multivariate point process $(t_n,x_n)_{n\geq 1}$. If $\gamma(\{x\}) = 1$ for some $x \in X$, we shall write $\mathbb{P}_x^\pi$ instead of $\mathbb{P}_\gamma^\pi$. Let $\mathbb{E}_\gamma^\pi$ and $\mathbb{E}_x^\pi$ respectively denote the expectation with respect to the measures $\mathbb{P}_\gamma^\pi$ and $\mathbb{P}_x^\pi$.

We shall show in Chapter 4 that if π is a relaxed Markov policy, then the jump process $\{\xi_t, t \geq 0\}$ defined by the compensator ν^π satisfying (3.3) is a jump Markov process. This result follows from Corollary 2.3.3.

3.4 Cost Criteria

For all $y,z \in X$ with $y \neq z$ and $p \in \mathscr{P}(A)$, let

$$\begin{aligned} c(z,p) &:= \int_{A(z)} c(z,a)p(da) & (3.5)\\ C(z,p,y) &:= \int_{A(z)} C(z,a,y)p(da). & (3.6)\end{aligned}$$

We now give a brief description of the different cost criteria considered in this thesis for CTMDPs. Given an initial distribution γ on X, for any policy π:

(i) the *finite horizon expected total discounted cost* is given by

$$V_{\alpha,T}(\gamma,\pi) := \mathbb{E}_\gamma^\pi[\int_0^{T\wedge t_\infty} e^{-\alpha s} c(\xi_s,\pi_s)ds + \sum_{n=1}^{N(T)} e^{-\alpha t_n} C(\xi_{t_{n-1}},\pi_{t_n},\xi_{t_n})], \tag{3.7}$$

where T is the finite planning horizon, $N(T)$ is the number of jumps up to time T, and $\alpha \in]0,\infty[$ is the discount factor.

(ii) Formula (3.7) with $\alpha = 0$ defines the *finite-horizon expected total cost* denoted by $V_{0,T}(\gamma,\pi)$.

(iii) Formula (3.7) with $T = \infty$ defines the *expected total discounted cost* denoted by $V_\alpha(\gamma,\pi)$.

(iv) Formula (3.7) with $\alpha = 0$ and $T = \infty$ defines the *expected total cost* denoted by $V_0(\gamma,\pi)$.

(v) If $\mathbb{P}_\gamma^\pi(\xi_t \in X) = 1$ for all $t \in \mathbb{R}_+$, then the average cost per unit time is given by

$$W(\gamma,\pi) = \limsup_{T\to\infty} \frac{V_{0,T}(\gamma,\pi)}{T}. \tag{3.8}$$

Chapter 4

Sufficiency of Markov policies in CTMDPs

In this chapter, we consider Borel state and action CTMDPs with unbounded transition rates and present the main result of this dissertation, Theorem 4.2.2. We show that the search for optimal policies in CTMDPs can be restricted to Markov policies for optimality criteria that depend only on marginal distributions of state-action pairs, like the expected total discounted and non-discounted costs and average costs per unit time.

4.1 Introduction

The first consideration of relaxed history dependent policies is by Kitaev [24]. He observed that an arbitrary policy for the CTMDP defines a compensator in a natural way and constructed the stochastic process via the compensator and the initial state distribution based on Jacod [17, Theorem 3.6]. Even though it is possible to consider history dependent policies for CTMDPs, most of the literature on CTMDPs considered relaxed Markov policies as the most general class of policies and established many of the existing facts such as optimality of certain policies within the class of relaxed Markov policies; see Guo and Hernández-Lerma [11]. In this chapter, we show that for any arbitrary policy there exists a relaxed Markov policy that performs equally, and thus, extending the previously established results within the class of relaxed Markov policies to hold within the class of all policies.

This chapter is organized as follows. In Section 4.2, we introduce marginal distributions on the state-action pairs and on the states of CTMDPs. For a policy π we construct in Lemma 4.2.1 a relaxed Markov policy φ. Then, in Theorem 4.2.2, we state the main result of this dissertation that the marginal distributions coincide for these two policies. This theorem is similar to the well-known result by Derman and Strauch [4] for DTMDPs that states the sufficiency of Markov policies for objective criteria that depend only on the marginal distributions; see Theorem 5.2.1 in Appendix 5. The proof given by Derman and Strauch [4] for the discrete-time case is based on induction in the step number, and hence not applicable for continuous-time. The proof for the continuous-time case is based on the fact that the marginal distributions on the state for both the policies, π and φ, satisfy Kolmogorov's forward equation defined by the Markov policy φ; Lemma 4.4.4. In Section 4.3, we apply the results in Chapter 2 and show that the jump process defined by a Markov policy is a jump Markov process whose marginal distribution on the state of the process is the minimal non-negative solution of Kolmogorov's foward equation (4.10) defined by the Markov policy; Theorem 4.3.1. We also provide a sufficient condition for the marginal distribution on the state of the process to be the unique non-negative solution of Kolmogorov's forward equation (4.10) that takes values in $[0,1]$. The proof of the main result, Theorem 4.2.2, is provided in Section 4.4 after establishing few auxiliary results. Finally in Section 4.5, we characterize the equivalence between the classes of history-dependent and Markov policies for objective criteria such as the expected discounted and non-discounted total costs and average costs per unit time.

4.2 Main result

Given an initial distribution γ, for any policy π, consider

$$P_\gamma^\pi(t,Z,B) := \int_\Omega I\{\xi_t \in Z\}\pi(B|\omega,t)\mathbb{P}_\gamma^\pi(d\omega), \qquad t \in \mathbb{R}_+, Z \in \mathfrak{B}(X), B \in \mathfrak{B}(A), \quad (4.1)$$

$$P_\gamma^\pi(t,Z) := P_\gamma^\pi(t,Z,A) = \int_\Omega I\{\xi_t \in Z\}\mathbb{P}_\gamma^\pi(d\omega), \qquad t \in \mathbb{R}_+, Z \in \mathfrak{B}(X), \quad (4.2)$$

where the equality in (4.2) is correct since $\{\xi_t \in X\} = \{\pi(A|\omega,t) = 1\}$. Observe that (i) for fixed t, the functions $P_\gamma^\pi(t,\cdot,\cdot)$ and $P_\gamma^\pi(t,\cdot)$ are measures on $(X \times A, \mathfrak{B}(X) \times \mathfrak{B}(A))$ and $(X, \mathfrak{B}(X))$, respectively, and (ii) $P_\gamma^\pi(t,Z) = \mathbb{P}_\gamma^\pi(\xi_t \in Z)$ for all $t \in \mathbb{R}_+$ and $Z \in \mathfrak{B}(X)$. Similar to the notation $\mathbb{P}_x^\pi$, we shall write P_x^π instead of P_γ^π if $\gamma(\{x\}) = 1$ for some $x \in X$.

Lemma 4.2.1. *Given an initial distribution γ, for every policy π there exists a relaxed Markov policy φ that satisfies, for all $t \in \mathbb{R}_+$,*

$$\varphi(B|z,t) = \frac{P_\gamma^\pi(t,dz,B)}{P_\gamma^\pi(t,dz)}, \quad P_\gamma^\pi(t,\cdot) - a.e., z \in X, B \in \mathfrak{B}(A). \quad (4.3)$$

Proof. Fix $t \in \mathbb{R}_+$ such that $P_\gamma^\pi(t,X) > 0$. By Bertsekas and Shreve [2, Corollary 7.27.1], there exists a transition probability φ from $(X \times \mathbb{R}_+, \mathfrak{B}(X) \times \mathfrak{B}(\mathbb{R}_+))$ to $(A, \mathfrak{B}(A))$ satisfying

$$P_\gamma^\pi(t,Z,B) = \int_Z \varphi(B|z,t)P_\gamma^\pi(t,dz), \qquad Z \in \mathfrak{B}(X), B \in \mathfrak{B}(A), \quad (4.4)$$

which, by definition, is equivalent to (4.3). Since the measure $P_\gamma^\pi(t,\cdot,\cdot)$ is concentrated on $Gr(A)$, the transition probability φ can be defined in such a way that $\varphi(A(z)|z,t) = 1$ for all $z \in X$ and (4.4) holds. Therefore, if $P_\gamma^\pi(t,X) > 0$, then (4.3) holds. Alternatively, fix $t \in \mathbb{R}_+$ such that $P_\gamma^\pi(t,X) = 0$. Then, (4.3) holds for any relaxed Markov policy φ. Thus, (4.3) holds for all $t \in \mathbb{R}_+$. $\square$

The following theorem is the main result of this dissertation.

Theorem 4.2.2. *Let the initial distribution γ be fixed. For any policy π, consider a relaxed Markov policy φ satisfying* (4.3)*. Then*

$$P_\gamma^\varphi(t,Z,B) = P_\gamma^\pi(t,Z,B), \qquad Z \in \mathfrak{B}(X), B \in \mathfrak{B}(A), t \in \mathbb{R}_+. \quad (4.5)$$

In particular,

$$P^{\varphi}_{\gamma}(t,Z) = P^{\pi}_{\gamma}(t,Z), \qquad Z \in \mathfrak{B}(X), t \in \mathbb{R}_+. \tag{4.6}$$

The proof of Theorem 4.2.2 is given in Section 4.4.

4.3 Kolmogorov's forward equation for CTMDPs controlled by Markov policies

In this section, we apply the results on Kolmogorov's forward equation in Chapter 2 to CTMDPs controlled by Markov policies. Let φ be a Markov policy. Then, it follows immediately that the transition rate function $q(Z|z,\varphi_t)$, $Z \in \mathfrak{B}(X)$, $z \in X$, and $t \in \mathbb{R}_+$, is a Q-function (see Chapter 2 for definition of a Q-function). In view of Theorem 2.5.1, the minimal non-negative solution of Kolmogorov's forward equation defined by the Q-function $q(Z|z,\varphi_t), z \in X, t \in \mathbb{R}_+, Z \in \mathfrak{B}(X)$, is a transition function of a jump Markov process. This approach to construct the jump process defined by a Markov policy is adapted in many of the studies on CTMDPs controlled by Markov policies including the monograph Guo and Hernández-lerma [11]; see also Kakumanu [19], Miller [27, 28]. However, as mentioned in Chapter 3, the jump process defined by any policy can be constructed using the compensator of the random measure of the multivariate point process. The following theorem shows that the two ways to construct the jump process defined by a Markov policy: (i) via the compensator of the random measure of multivariate point process, and (ii) as the minimal non-negative solution of Kolmogorov's forward equation (4.10) are equivalent.

Consider the transition function $P^{\varphi}(u,z;t,Z)$, where $u,t \in \mathbb{R}_+$, $u < t$, $z \in X$, and $Z \in \mathfrak{B}(X)$, given below, that is obtained by replacing the generic Q-function $q(z,t,Z), z \in X, t \in \mathbb{R}_+, Z \in \mathfrak{B}(X)$, in (2.17)-(2.19) with the specific function $q(Z|z,\varphi_t), z \in X, t \in \mathbb{R}_+, Z \in \mathfrak{B}(X)$. For all $u,t \in \mathbb{R}_+$, $u < t$, $z \in X$, and $Z \in \mathfrak{B}(X)$, define

$$P^{(\varphi,0)}(u,z;t,Z) = I\{z \in Z\}e^{-\int_u^t q(z,\varphi_w)dw}, \tag{4.7}$$

and for $m = 1,2,\ldots$, define

$$P^{(\varphi,m)}(u,z;t,Z) = \int_u^t \int_{X\setminus\{z\}} e^{-\int_u^s q(z,\varphi_w)dw} q(dy|z,\varphi_s) P^{(\varphi,m-1)}(s,y;t,Z)ds. \tag{4.8}$$

Set

$$P^{\varphi}(u,z;t,Z) := \sum_{m=0}^{\infty} P^{(\varphi,m)}(u,z;t,Z). \tag{4.9}$$

Let Y_φ be the collection of Borel subsets of X such that

$$Y_\varphi = \{Z \in \mathfrak{B}(X) : \sup_{t\in\mathbb{R}_+, z\in Z} q(z,\varphi_t) < \infty\}.$$

Theorem 4.3.1. *Let the initial state x be fixed. For any Markov policy φ, the following statements hold:*

(i) the jump process $\{\xi_t : t \in \mathbb{R}_+\}$ *defined in* (3.1) *by the compensator* ν^φ *satisfying* (3.3) *is a jump Markov process with transition function* $P^\varphi(u,z;t,B)$ *defined in* (4.9).

(ii) the function $P_x^\varphi(t,Z)$ *is the minimal non-negative solution of Kolmogorov's forward equation,*

$$P(t,Z) = I\{x \in Z\} + \int_0^t \int_X q(Z|z,\varphi_s)P(s,dz)ds, \qquad t>0, Z \in Y_\varphi. \tag{4.10}$$

In addition, if $X \in Y_\varphi$*, then* $P_x^\varphi(t,Z)$ *is the unique non-negative function with values in* $[0,1]$ *for which statement (i) holds.*

Proof. (i) Assumption 3.3.1 and (3.2) imply that

$$\sup_{t\in\mathbb{R}_+} q(z,\varphi_t) = \sup_{t\in\mathbb{R}_+} \int_{A(z)} q(z,a)\varphi(da|z,t) \le \bar{q}(z) < \infty, \qquad z \in X. \tag{4.11}$$

That is, the transition rate function $q(Z|z,\varphi_t)$, $Z \in \mathfrak{B}(X)$, $z \in X$, $t \in \mathbb{R}_+$, is a Q-function satisfying Assumption 2.2.2 in Chapter 2. Thus, in view of Corollary 2.3.3, we have that statement (i) of the theorem holds.

(ii) From (4.2) and (2.2) with $u=0$ for the jump Markov process defined by the Markov policy φ,

$$P_x^\varphi(t,Z) = \mathbb{P}_x^\varphi(\xi_t \in Z) = P^\varphi(0,x;t,Z), \qquad t>0, Z \in \mathfrak{B}(X). \tag{4.12}$$

The above equality, (4.11), and Corollary 2.6.3 with $q(z,s,Z) = q(Z|z,\varphi_s)$, $T_0 = 0$, and $T_1 = \infty$ imply that the second statement of the theorem holds. $\square$

Remark 4.3.1. In view of Lemma 2.2.1(i), the collection of Borel subsets Y^φ contains, among others, a sequence of Borel subsets $X_\varphi^n \uparrow X$ as $n \to \infty$. Therefore, $Y^\varphi \neq \emptyset$.

4.4 Proof of Theorem 4.2.2

A set $Z \in \mathfrak{B}(X)$ is called a q-bounded set if $\sup_{z\in Z}\bar{q}(z) < \infty$. Given an initial state x and any policy π, for each $t \in \mathbb{R}_+$, a set $Z \in \mathfrak{B}(X)$ is called an (x,π,t)-bounded set if $\int_0^t \mathbb{E}_x^\pi q(\xi_s,\pi_s)I\{\xi_s \in Z\}ds < \infty$, and if the set $Z \in \mathfrak{B}(X)$ is (x,π,t)-bounded for all $t \in \mathbb{R}_+$ we say the set is (x,π)-bounded. Note that, if a set Z is a q-bounded set, then it is also an (x,π)-bounded set for all $x \in X$ and for any policy π. The jump process $\{\xi_t : t \in \mathbb{R}_+\}$ defined by a policy π is called non-explosive if $\mathbb{P}_x^\pi(\xi_t \in X) = 1$ for all $x \in X$ and $t \in \mathbb{R}_+$.

4.4.1 Marginal distributions for an arbitrary policy satisfy Kolmogorov's forward equation

To study the jump process associated with a history dependent policy π and initial state x, Kitaev [24, Lemma 4] established the following equation for a uniformly bounded transition kernel

q. Given a policy π and initial state x, for all $t \in \mathbb{R}_+$, and $Z \in \mathfrak{B}(X)$,

$$P_x^\pi(t,Z) = I\{x \in Z\} + \mathbb{E}_x^\pi \left(\int_0^t q(Z|\xi_s, \pi_s) ds \right). \tag{4.13}$$

By imposing additional conditions on the transition kernel q to ensure that the jump process is non-explosive for any policy, Guo and Song [14, Theorem 3.1] showed that (4.13) holds for all q-bounded sets $Z \in \mathfrak{B}(X)$ and, under slightly weaker conditions for non-explosiveness than those considered by Guo and Song [14], Piunovskiy and Zhang [29, Proposition A.1] established (4.13) for all $Z \in \mathfrak{B}(X)$. Given a policy π and initial state x, Theorem 4.4.2 establishes (4.13) for all $t \in \mathbb{R}_+$ and (x,π,t)-bounded sets $Z \in \mathfrak{B}(X)$, and thus (4.13) holds for an explosive process.

The following lemma is used in the proof of Theorem 4.4.2. Given a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ with a right-continuous and complete filtration $\{\mathscr{F}_t\}_{t\in\mathbb{R}_+}$, a random measure $\nu : (\Omega \times \mathfrak{B}(\mathbb{R}_+) \times \mathfrak{B}(X)) \to \mathbb{R}_+$ is called a compensator of the random measure $\mu : (\Omega \times \mathfrak{B}(\mathbb{R}_+) \times \mathfrak{B}(X)) \to \mathbb{R}_+$ if (i) for each $Z \in \mathfrak{B}(X)$, the process $\{\nu(\omega;[0,t],Z)\}_{t\in\mathbb{R}_+}$ is measurable with respect to the predictable σ-algebra $\mathscr{P}$ of $(\Omega \times \mathbb{R}_+)$, and (ii) for any non-negative $\mathscr{P} \times \mathfrak{B}(X)$-measurable function $f(\omega,t,z)$,

$$\mathbb{E}(\int_0^\infty \int_X f(\omega,s,z)\mu(\omega;ds,dz)) = \mathbb{E}(\int_0^\infty \int_X f(\omega,s,z)\nu(\omega;ds,dz)),$$

where $\mathbb{E}$ denotes the expectation with respect to the probability measure $\mathbb{P}$; Kallenberg [20, p. 422].

Lemma 4.4.1. *Let the initial state x be fixed. For any policy π, the random measure $\tilde{\upsilon}^\pi$ given by*

$$\tilde{\upsilon}^\pi(\omega;[0,t],Z) := \int_0^t q(\xi_s,\pi_s) I\{\xi_s \in Z\} ds, \qquad t \in \mathbb{R}_+, Z \in \mathfrak{B}(X), \tag{4.14}$$

is a compensator for the random measure $\tilde{\mu}$ given by

$$\tilde{\mu}(\omega;[0,t],Z) := \sum_{n\geq 1} I\{t_n \in [0,t]\} I\{x_{n-1} \in Z\}, \qquad t \in \mathbb{R}_+, Z \in \mathfrak{B}(X), \tag{4.15}$$

with respect to the probability space $(\Omega, \mathscr{F}, \{\mathscr{F}_t\}_{t\geq 0}, \mathbb{P}_x^\pi)$.

Proof. See the proof of Lemma 4 in Kitaev [24]. □

Theorem 4.4.2. *Let the initial state $x \in X$ be fixed. For any policy π, formula* (4.13) *holds for all $t \in \mathbb{R}_+$ and (x,π,t)-bounded sets $Z \in \mathfrak{B}(X)$.*

Proof. For all $m = 0,1,\ldots$, $t \in \mathbb{R}_+$, and $Z \in \mathfrak{B}(X)$, as the number of jumps in the interval $[0, t\wedge t_m]$ is bounded by m and the random measures $\mu([0,t\wedge t_m],Z)$ and $\tilde{\mu}([0,t\wedge t_m],Z)$ defined in (3.4) and (4.15), respectively, give the number of jumps into and out of Z (including within itself) in the interval $[0,t\wedge t_m]$,

$$I\{\xi_{t\wedge t_m} \in Z\} = I\{x \in Z\} + \mu([0,t\wedge t_m],Z) - \tilde{\mu}([0,t\wedge t_m],Z). \tag{4.16}$$

Taking expectation with respect to $\mathbb{P}_x^\pi$ on both sides of the above equality, for all $t \in \mathbb{R}_+$ and $Z \in \mathfrak{B}(X)$,

$$\mathbb{P}_x^\pi(\xi_{t\wedge t_m} \in Z) = I\{x \in Z\} + \mathbb{E}_x^\pi(\mu([0,t\wedge t_m],Z)) - \mathbb{E}_x^\pi(\tilde{\mu}([0,t\wedge t_m],Z)), \tag{4.17}$$

which implies that

$$\mathbb{E}_x^\pi(\mu([0,t\wedge t_m],Z)) \leq \mathbb{E}_x^\pi(\tilde{\mu}([0,t\wedge t_m],Z)) + 1. \tag{4.18}$$

Next, we show that for each $t \in \mathbb{R}_+$ (4.17) holds with $m = \infty$ for all (x,π,t)-bounded sets $Z \in \mathfrak{B}(X)$. Fix an arbitrary $t \in \mathbb{R}_+$. For all (x,π,t)-bounded sets $Z \in \mathfrak{B}(X)$,

$$\begin{aligned}\mathbb{E}_x^\pi(\tilde{\mu}([0,t\wedge t_\infty],Z)) = \mathbb{E}_x^\pi(\tilde{\upsilon}^\pi([0,t\wedge t_\infty],Z)) = \mathbb{E}_x^\pi \int_0^{t\wedge t_\infty} q(\xi_s,\pi_s) I\{\xi_s \in Z\} ds \quad (4.19)\\ \leq \int_0^t \mathbb{E}_x^\pi q(\xi_s,\pi_s) I\{\xi_s \in Z\} ds < \infty,\end{aligned}$$

where the first equality is correct since $\tilde{\upsilon}^\pi$ is a compensator of the measure $\tilde{\mu}$; Lemma 4.4.1, the second equality is (4.14), the third inequality is straightforward, and the last one follows from the definition of (x,π,t)-bounded sets. Then (4.18), (4.19), and the property that $\mu([0,t\wedge t_m],Z)$ and $\tilde{\mu}([0,t\wedge t_m],Z)$ are non-decreasing in m for each Z imply that, for all (x,π,t)-bounded sets $Z \in \mathfrak{B}(X)$,

$$\lim_{m\to\infty} \mathbb{E}_x^\pi(\mu([0,t\wedge t_m],Z)) = \mathbb{E}_x^\pi(\mu([0,t\wedge t_\infty],Z)) < \infty, \tag{4.20}$$

$$\lim_{m\to\infty} \mathbb{E}_x^\pi(\tilde{\mu}([0,t\wedge t_m],Z)) = \mathbb{E}_x^\pi(\tilde{\mu}([0,t\wedge t_\infty],Z)) < \infty. \tag{4.21}$$

Observe that for all (x,π,t)-bounded sets $Z \in \mathfrak{B}(X)$

$$\begin{aligned}\lim_{m\to\infty} \mathbb{P}_x^\pi(\xi_{t\wedge t_m} \in Z) &= \mathbb{P}_x^\pi(\xi_t \in Z, t < t_\infty) + \lim_{m\to\infty} \mathbb{P}_x^\pi(\xi_{t_m} \in Z, t \geq t_m)\\ &= \mathbb{P}_x^\pi(\xi_t \in Z, t < t_\infty) = \mathbb{P}_x^\pi(\xi_{t\wedge t_\infty} \in Z),\end{aligned} \tag{4.22}$$

where the first equality holds because $t_m \uparrow t_\infty$ as $m \to \infty$, the second equality holds because $\lim_{m\to\infty} \mathbb{P}_x^\pi(\xi_{t_m} \in Z, t \geq t_m) = 0$, and the last one holds because $\{\xi_t \in X\} = \{t < t_\infty\}$. Indeed, let $\limsup_{m\to\infty} \mathbb{P}_x^\pi(\xi_{t_m} \in Z, t \geq t_m) = p > 0$. Then there exists a subsequence $\{m_k, k = 1,2,\ldots\}$ such that $\mathbb{P}_x^\pi(\xi_{t_{m_k}} \in Z, t \geq t_{m_k}) > \frac{p}{2}$ for all $k = 1,2,\ldots$. This fact, (3.4), and the property $\{t_m \in [0,t]\} = \{t_m \in [0,t\wedge t_\infty]\}$ for all $m = 0,1,\ldots$ imply that, for all (x,π,t)-bounded sets $Z \in \mathfrak{B}(X)$,

$$\mathbb{E}_x^\pi(\mu([0,t\wedge t_\infty],Z])) = \mathbb{E}_x^\pi(\mu([0,t],Z])) \geq \sum_{k=1,2,\ldots} \mathbb{P}_x^\pi(t_{m_k} \in [0,t], x_{m_k} \in Z) = \infty,$$

which contradicts (4.20). This contradiction implies that $\lim_{m\to\infty} \mathbb{P}_x^\pi(\xi_{t_m} \in Z, t \geq t_m) = 0$ for all (x,π,t)-bounded sets $Z \in \mathfrak{B}(X)$. Therefore, (4.22) holds which together with (4.17), (4.20), and (4.21) implies that, for each t, (4.17) holds with $m = \infty$ for all (x,π,t)-bounded sets $Z \in \mathfrak{B}(X)$.

To complete the proof, observe that, for all $t \in \mathbb{R}_+$ and $Z \in \mathfrak{B}(X)$, (i) $\mu([0,t],Z) = \mu([0,t\wedge$

$t_\infty], Z)$ and $\tilde{\mu}([0,t],Z) = \tilde{\mu}([0,t\wedge t_\infty],Z)$, and (ii) $P_x^\pi(t,Z) = \mathbb{P}_x^\pi(\xi_{t\wedge t_\infty} \in Z)$. We remark that (i) holds because $\{t_m \in [0,t]\} = \{t_m \in [0,t\wedge t_\infty]\}$, $m = 0,1,\ldots$, and (ii) follows from (4.2) and the property $\{\xi_{t\wedge t_\infty} \in Z\} = \{\xi_t \in Z, t < t_\infty\} = \{\xi_t \in Z\}$. Then, for all $t \in \mathbb{R}_+$ and for all (x,π,t)-bounded sets $Z \in \mathfrak{B}(X)$,

$$
\begin{aligned}
P_x^\pi(t,Z) &= I\{x \in Z\} + \mathbb{E}_x^\pi(\mu^\pi([0,t]\times Z) - \tilde{\mu}^\pi([0,t]\times Z)) \\
&= I\{x \in Z\} + \mathbb{E}_x^\pi(\upsilon^\pi([0,t]\times Z) - \tilde{\upsilon}^\pi([0,t]\times Z)) \\
&= I\{x \in Z\} + \mathbb{E}_x^\pi\left(\int_0^t q(Z\setminus\{\xi_s\}|\xi_s,\pi_s)ds - \int_0^t q(\xi_s,\pi_s)I\{\xi_s \in Z\}ds\right) \\
&= I\{x \in Z\} + \mathbb{E}_x^\pi\left(\int_0^t q(Z|\xi_s,\pi_s)ds\right),
\end{aligned}
$$

where the first equality follows from observations (i) and (ii) given above and from (4.17) with $m = \infty$, the second equality holds since υ^π and $\tilde{\upsilon}^\pi$ are respectively the compensator of the measures μ and $\tilde{\mu}$, the third equality follows from (3.3) and (4.14), and the last one follows from (3.2). □

Corollary 4.4.3 follows immediately from Theorem 4.4.2 and is useful to prove Lemma 4.4.4.

Corollary 4.4.3. *Let the initial state $x \in X$ be fixed. Then, for each policy π,* (4.13) *holds for all $t \in \mathbb{R}_+$ and (x,π)-bounded sets $Z \in \mathfrak{B}(X)$.*

Proof. The statement of this corollary follows immediately from Theorem 4.4.2 and the fact that an (x,π)-bounded set is (x,π,t)-bounded for all $t \in \mathbb{R}_+$. □

The following lemma plays the pivotal role in the proof of the main theorem, Theorem 4.2.2. It establishes that, given an initial state x, for any policy π the function $P_x^\pi(t,Z), t \in \mathbb{R}_+, Z \in \mathfrak{B}(X)$, satisfies Kolmogorov's forward equation (4.10) corresponding to relaxed Markov policy φ satisfying (4.3) with $\gamma(\{x\}) = 1$. In particular, it establishes that (4.6) holds if the transition rates $q(z,\varphi_t), z \in X, t \in \mathbb{R}_+$, are uniformly bounded.

Lemma 4.4.4. *Let the initial state x be fixed. For each policy π, consider a relaxed Markov policy φ satisfying* (4.3) *with $\gamma = \delta_x$. Then, (i) The function $P_x^\pi(t,Z), t \in \mathbb{R}_+, Z \in \mathfrak{B}(X)$, satisfies Kolmogorov's forward equation* (4.10)*, (ii) $P_x^\pi(t,Z) \geq P_x^\varphi(t,Z)$ for all $t \in \mathbb{R}_+$ and $Z \in \mathfrak{B}(X)$, and (iii) if $X \in Y_\varphi$, then* (4.6) *holds for all $t \in \mathbb{R}_+$ and $Z \in \mathfrak{B}(X)$.*

Proof. Observe that, for each t, the function $P_x^\pi(t,\cdot)$ is a nonnegative measure on $(X,\mathfrak{B}(X))$ and take values in $[0,1]$. Thus, if statement (i) of the lemma holds, it follows from the property that $P_x^\pi(0,Z) = P_x^\varphi(0,Z) = I\{x \in Z\}$ for all $Z \in \mathfrak{B}(X)$ and Theorem 4.3.1 that statements (ii) and (iii) of the lemma hold. The rest of the proof establishes statement (i) of the lemma.

For all $Y,Z \in \mathfrak{B}(X)$ and $t \in \mathbb{R}_+$,

$$
\mathbb{E}_x^\pi[q(Z\setminus\{\xi_t\}|\xi_t,\pi_t)I\{\xi_t \in Y\}] = \int_\Omega\int_A q(Z\setminus\{\xi_t\}|\xi_t,a)\pi(da|\omega,t)I\{\xi_t \in Y\}\mathbb{P}_x^\pi(d\omega)
$$

$$
\begin{aligned}
&= \int_A \int_\Omega q(Z\setminus\{\xi_t\}|\xi_t,a)\pi(da|\omega,t)I\{\xi_t\in Y\}\mathbb{P}_x^\pi(d\omega) = \int_A\int_Y q(Z\setminus\{z\}|z,a)P_x^\pi(t,dz,da) \\
&= \int_Y\int_A q(Z\setminus\{z\}|z,a)\varphi(da|z,t)P_x^\pi(t,dz) = \int_Y q(Z\setminus\{z\}|z,\varphi_t)P_x^\pi(t,dz),
\end{aligned} \tag{4.23}
$$

where the first equality follows from the definition of expectation and (3.2), the second equality is obtained by interchanging the order of integration, the third equality follows from (4.1), the fourth equality follows from (4.4) and from interchanging the order of integration, and the last one follows from (3.2). In particular, (4.23) with $Z=X$ and (3.2) imply that

$$
\mathbb{E}_x^\pi q(\xi_t,\pi_t)I\{\xi_t\in Y\} = \int_Y q(z,\varphi_t)P_x^\pi(t,dz), \qquad t\in\mathbb{R}_+, Y\in\mathfrak{B}(X). \tag{4.24}
$$

Then for all $Z\in\mathfrak{B}(X)$ and $t\in\mathbb{R}_+$

$$
\begin{aligned}
\mathbb{E}_x^\pi[q(Z|\xi_t,\pi_t)] &= \mathbb{E}_x^\pi[q(Z\setminus\{\xi_t\}|\xi_t,\pi_t)I\{\xi_t\in X\}] - \mathbb{E}_x^\pi[q(\xi_t,\pi_t)I\{\xi_t\in Z\}] \\
&= \int_X q(Z\setminus\{z\}|z,\varphi_t)P_x^\pi(t;dz) - \int_Z q(z,\varphi_t)P_x^\pi(t;dz) = \int_X q(Z|z,\varphi_t)P_x^\pi(t,dz),
\end{aligned} \tag{4.25}
$$

where the first and last equalities follows from (3.2), and the second equality follows from (4.23) with $Y=X$ and from (4.24) with $Y=Z$. Therefore, it follows from Corollary 4.4.3 and from (4.25) that the function $P_x^\pi(t,Z), t\in\mathbb{R}_+, Z\in\mathfrak{B}(X)$, satisfies (4.10) for all $t\in\mathbb{R}_+$ and (x,π)-bounded sets $Z\in\mathfrak{B}(X)$. In fact, it follows from (4.24) with $Y=Z$ that if $Z\in Y_\varphi$, then Z is an (x,π)-bounded set. Thus, the function $P_x^\pi(t,Z)$ defined in (4.2) satisfies Kolmogorov's forward equation (4.10). ☐

Lemma 4.4.5 shows that to prove the main theorem it is sufficient to show that (4.6) holds for the policies π and φ satisfying (4.3). The proof of Theorem 4.2.2 is then provided after giving few auxiliary results.

Lemma 4.4.5. *Given an initial distribution γ on X, for each policy π consider a relaxed Markov policy φ satisfying* (4.3)*. Then, formula* (4.6) *implies formula* (4.5)*.*

Proof. Assume that (4.6) holds. Then, for all $t\in\mathbb{R}_+$, $Z\in\mathfrak{B}(X)$, and $B\in\mathfrak{B}(A)$,

$$
\begin{aligned}
P_\gamma^\varphi(t,Z,B) &= \int_\Omega I\{\xi_t\in Z\}\varphi(B|\xi_t,t)\mathbb{P}_\gamma^\varphi(d\omega) = \int_Z \varphi(B|z,t)P_\gamma^\varphi(t,dz) \\
&= \int_Z \varphi(B|z,t)P_\gamma^\pi(t,dz) = P_\gamma^\pi(t,Z,B),
\end{aligned} \tag{4.26}
$$

where the first equality follows from (4.1) and the equality $\varphi(\cdot|\omega,t)=\varphi(\cdot|\xi_t,t)$, the second equality follows from (4.2), the third equality follows from the assumption that (4.6) holds for all $t\in\mathbb{R}_+$ and $Z\in\mathfrak{B}(X)$, and the last one holds due to (4.3). ☐

It follows from Lemma 4.4.4 and Lemma 4.4.5 that the main theorem holds if the initial state distribution is a Dirac measure and the transition rates $q(z,\varphi_t), z\in X, t\in\mathbb{R}_+$, defined by the relaxed Markov policy φ satisfying (4.3), are uniformly bounded. In the general case, we approximate the state space X by measurable subsets $X^n, n = 1,2,\ldots$, of X such that $X^n \uparrow X$ as $n \to \infty$, and consider policies $\pi^{(n)}, n = 1,2,\ldots$, such that $\pi^{(n)}$ coincides with π on X^n and all the states $z \in X \setminus X^n$ are absorbing under the policy $\pi^{(n)}$. For each policy $\pi^{(n)}, n = 1,2,\ldots$, we establish the existence of a relaxed Markov policy $\varphi^n, n = 1,2,\ldots$, such that the marginal distributions coincide for these two policies. The proof of the main theorem follows from these approximations; see the diagram below.

$$\begin{array}{ccc} P_x^{\pi^{(n)}}(t,Z,B) & \rightarrow & P_x^{\pi}(t,Z,B) \\ \| & & \\ P_x^{\varphi^n}(t,Z,B) & \rightarrow & P_x^{\varphi}(t,Z,B) \end{array} \Longrightarrow P_x^{\varphi}(t,Z,B) = P_x^{\pi}(t,Z,B).$$

Figure 4.1: Major steps of the proof of Theorem 4.2.2

Lemma 4.4.6. *There exists a sequence of measurable subsets X^n of X such that $\sup_{z\in X^n} \bar{q}(z) < n$ for all $n = 1,2,\ldots$ and $X^n \uparrow X$ as $n \to \infty$.*

Proof. The statement of the lemma follows from the arguments in the proof of Lemma 2.2.1(i) given for the measurable function $q(x,a)$ instead of for the measurable function $q(x,t)$. □

For an isolated point $\tilde{a} \notin A$, define $\tilde{A} := A \cup \{\tilde{a}\}$ and $\tilde{A}(x) := A(x) \cup \{\tilde{a}\}$, $x \in X$. Consider the Borel σ-algebra $\mathfrak{B}(\tilde{A}) := \sigma(\mathfrak{B}(A), \{\tilde{a}\})$ on $\tilde{A}$. For all $x \in X$, $a \in \tilde{A}(x)$, and $Z \in \mathfrak{B}(X)$, define the new transition intensity $\tilde{q}$ by

$$\tilde{q}(Z|x,a) := q(Z|x,a) I\{x \in X, a \in A(x)\}, \qquad Z \in \mathfrak{B}(X), x \in X, a \in \tilde{A}(x). \tag{4.27}$$

Consider the extended CTMDP $\{X, \tilde{A}, \tilde{A}(x), \tilde{q}(\cdot|x,a)\}$. Let $\tilde{q}(x,a) := \tilde{q}(X \setminus \{x\}|x,a)$, $x \in X$, $a \in \tilde{A}(x)$, and $\bar{q}(x_\infty) := 0$. Note that any policy π in the original CTMDP $\{X, A, A(x), q(\cdot|x,a)\}$ is a policy in the extended CTMDP that does not select the action $\tilde{a}$.

Consider the measurable subsets $X^n, n = 1,2,\ldots$, of X whose existence is stated in Lemma 4.4.6. For a policy π in the original CTMDP, let $\pi^{(n)}, n = 1,2,\ldots$, be a policy in the extended CTMDP such that, for all $B \in \mathfrak{B}(\tilde{A})$, $\omega \in \Omega$, and $t \in \mathbb{R}_+$,

$$\pi^{(n)}(B|\omega,t) := \pi(B \setminus \{\tilde{a}\}|\omega,t) I\{\xi_t \in X^n \cup \{x_\infty\}\} + I\{\tilde{a} \in B\} I\{\xi_t \in X \setminus X^n\}. \tag{4.28}$$

Lemma 4.4.7. *For a policy π in the original CTMDP, let $\pi^{(n)}, n = 1,2,\ldots$, be a policy in the extended CTMDP satisfying* (4.28)*. Then, given an initial distribution γ,*

$$P_\gamma^{\pi^{(n)}}(t,Z,B) \uparrow P_\gamma^{\pi}(t,Z,B) \quad \text{as} \quad n \to \infty, \tag{4.29}$$

for all $t \in \mathbb{R}_+, Z \in \mathfrak{B}(X)$, and $B \in \mathfrak{B}(A)$.

Proof. For $n = 1,2,\ldots$, define the function $\tau_n : \Omega \to \mathbb{R}_+$,

$$\tau_n(\omega) := \inf\{t \in \mathbb{R}_+ : \xi_t(\omega) \in X \setminus X^n\}, \qquad \omega \in \Omega, \tag{4.30}$$

where $\inf\{\emptyset\} := \infty$. Observe that, for all $n = 1,2,\ldots$, $t \in \mathbb{R}_+$, $Z \in \mathfrak{B}(X)$, and $B \in \mathfrak{B}(A)$,

$$P_\gamma^{\pi^{(n)}}(t,Z,B) = \int_\Omega I\{\xi_t \in Z, \tau_n > t\}\pi^{(n)}(B|\omega,t)\mathbb{P}_\gamma^{\pi^{(n)}}(d\omega)$$
$$+ \int_\Omega I\{\xi_t \in Z, \tau_n \le t\}\pi^{(n)}(B|\omega,t)\mathbb{P}_\gamma^{\pi^{(n)}}(d\omega) = \int_\Omega I\{\xi_t \in Z\}I\{\tau_n > t\}\pi(B|\omega,t)\mathbb{P}_\gamma^{\pi}(d\omega). \quad (4.31)$$

where the first equality follows from (4.1) for the policy $\pi^{(n)}$ and the second equality follows from (4.28) and (4.30) which imply that (i) if $\tau_n > t$ then $\pi^{(n)}(B|\omega,s) = \pi(B|\omega,s)$ for all $B \in \mathfrak{B}(A)$ and $s \le t$; and (ii) if $\tau_n \le t$ then $\pi^{(n)}(B|\omega,t) = 0$ for all $B \in \mathfrak{B}(A)$. Then, (4.29) follows from (4.1), (4.31), the almost sure convergence of $\tau_n(\omega) \to t_\infty$ as $n \to \infty$, and the property that $\{\xi_t \in X\} = \{t < t_\infty\}$. □

Lemma 4.4.8. *Let the initial state x be fixed. For any policy π in the original CTMDP, let $\pi^{(n)}, n = 1,2,\ldots,$ be a policy in the extended CTMDP satisfying* (4.28)*. Then, for each policy $\pi^{(n)}, n = 1,2,\ldots,$ there exists a relaxed Markov policy $\varphi^n, n = 1,2,\ldots,$ satisfying*

$$\varphi^n(\{\tilde{a}\}|z,t) = I\{z \in X \setminus X^n\}, \qquad t \in \mathbb{R}_+, z \in X, \quad (4.32)$$

and, for all $t \in \mathbb{R}_+$,

$$\varphi^n(B|z,t) = \frac{P_x^{\pi^{(n)}}(t,dz,B)}{P_x^{\pi^{(n)}}(t,dz)}, \qquad P_x^{\pi^{(n)}}(t,\cdot) - a.e., z \in X, B \in \mathfrak{B}(\tilde{A}). \quad (4.33)$$

Proof. In view of Lemma 4.2.1, for each policy $\pi^{(n)}, n = 1,2,\ldots,$ there exists a relaxed Markov policy $\psi^n, n = 1,2,\ldots,$ such that: for all $t \in \mathbb{R}_+$, there exists a subset $X_t \in \mathfrak{B}(X)$ such that

$$P_x^{\pi^{(n)}}(t, X \setminus X_t) = 0 \quad \text{and} \quad \psi^n(B|z,t) = \frac{P_x^{\pi^{(n)}}(t,dz,B)}{P_x^{\pi^{(n)}}(t,dz)}, \quad z \in X_t, B \in \mathfrak{B}(\tilde{A}). \quad (4.34)$$

Using the relaxed Markov policy $\psi^n, n = 1,2,\ldots$, we construct a relaxed Markov policy $\varphi^n, n = 1,2,\ldots,$ satisfying (4.32) and (4.33). For all $n = 1,2,\ldots$, $z \in X$, $t \in \mathbb{R}_+$, and $B \in \mathfrak{B}(\tilde{A})$, define

$$\varphi^n(B|z,t) = \psi^n(B \setminus \{\tilde{a}\}|z,t)I\{z \in X^n\} + I\{\tilde{a} \in B\}I\{z \in X \setminus X^n\}. \quad (4.35)$$

Then, it immediately follows from (4.35) that the function φ^n is a relaxed Markov policy satisfying (4.32).

To prove (4.33), observe that, for all $Z \in \mathfrak{B}(X)$, $t \in \mathbb{R}_+$, and $B \in \mathfrak{B}(\tilde{A})$,

$$P_x^{\pi^{(n)}}(t, Z \cap X^n, B) = \int_\Omega I\{\xi_t \in Z \cap X^n\}\pi^{(n)}(B|\omega,t)\mathbb{P}_x^{\pi^{(n)}}(d\omega)$$
$$= \int_\Omega I\{\xi_t \in Z \cap X^n\}\pi^{(n)}(B \setminus \{\tilde{a}\}|\omega,t)\mathbb{P}_x^{\pi^{(n)}}(d\omega) = P_x^{\pi^{(n)}}(t, Z \cap X^n, B \setminus \{\tilde{a}\}), \quad (4.36)$$

where the first and last equalities follows from (4.1) and the second equality follows from (4.28), and

$$P_x^{\pi^{(n)}}(t, Z\cap(X\setminus X^n), B) = \int_\Omega I\{\xi_t \in Z\cap(X\setminus X^n)\}\pi^{(n)}(B|\omega,t)\mathbb{P}_x^{\pi^{(n)}}(d\omega) \quad (4.37)$$
$$= I\{\tilde{a}\in B\}P_x^{\pi^{(n)}}(t, Z\cap(X\setminus X^n)),$$

where the first equality is (4.1) and the last one follows from (4.2) and (4.28). Then, from (4.34), (4.36), and (4.37), for all $t\in\mathbb{R}_+$,

$$\psi^n(B|z,t) = \begin{cases} \psi^n(B\setminus\{\tilde{a}\}|z,t), & \text{if} \quad z\in X_t\cap X^n,\ B\in\mathfrak{B}(\tilde{A}), \\ I\{\tilde{a}\in B\}, & \text{if} \quad z\in X_t\cap(X\setminus X^n),\ B\in\mathfrak{B}(\tilde{A}). \end{cases}$$

The above equality, (4.34), and (4.35) imply that (4.33) holds. □

Corollary 4.4.9. *Let the initial state x be fixed. For any policy π in the original CTMDP, let $\pi^{(n)}, n = 1,2,\ldots,$ be a policy satisfying* (4.28)*, and let $\varphi^n, n = 1,2,\ldots,$ be a relaxed Markov policy satisfying* (4.32) *and* (4.33)*. Then, for all $t\in\mathbb{R}_+, Z\in\mathfrak{B}(X)$, and $B\in\mathfrak{B}(A)$,*

$$P_x^{\varphi^n}(t,Z,B)\uparrow P_x^{\pi}(t,Z,B) \quad \text{as} \quad n\to\infty. \tag{4.38}$$

Proof. Observe that the transition rates $\tilde{q}(z,\varphi_t^n)$ are uniformly bounded by n for each $n = 1,2,\ldots.$ Indeed, for all $z\in X$ and $t\in\mathbb{R}_+$,

$$\tilde{q}(z,\varphi_t^n) = \int_{\tilde{A}(z)}\tilde{q}(z,a)\varphi^n(da|z,t) = \left(\int_{A(z)} q(z,a)\varphi^n(da|z,t)\right) I\{z\in X^n\} \quad (4.39)$$
$$\le \bar{q}(z)I\{z\in X^n\} < n,$$

where the first equality follows from (3.2) given for the extended CTMDP, the second equality follows from (4.27) and (4.35), and the last two inequalities are straightforward. That is, $X\in Y_{\varphi^n}$. This fact, Lemma 4.4.4(iii) and Lemma 4.4.5 for the policies $\pi^{(n)}$ and φ^n, and Lemma 4.4.7 imply (4.38). □

Lemma 4.4.10. *Let the initial state x be fixed. For each policy π, let φ be a relaxed Markov policy satisfying* (4.3) *with $\gamma(\{x\}) = 1$, and $\pi^{(n)}, n = 1,2,\ldots,$ be a policy in the extended CTMDP satisfying* (4.28)*. Then there exists a sequence of relaxed Markov policies $\{\varphi^n, n = 1,2,\ldots\}$ such that:*

(i) for all $n = 1,2,\ldots$, the relaxed Markov policy φ^n satisfies (4.32) *and* (4.33)*;*
(ii) for all $t\in\mathbb{R}_+$ and $z\in X$,

$$\varphi^n(B|z,t)\to\varphi(B|z,t) \quad \text{as} \quad n\to\infty \quad \text{for all} \quad B\in\mathfrak{B}(A). \tag{4.40}$$

Proof. In view of Lemma 4.4.8, for each policy $\pi^{(n)}, n = 1,2,\ldots,$ there exists a relaxed Markov policy $\tilde{\varphi}^n, n = 1,2,\ldots,$ such that (4.32) and (4.33) hold with φ^n replaced by $\tilde{\varphi}^n$. Using the relaxed Markov policies $\tilde{\varphi}^n$, we construct relaxed Markov policies φ^n, $n = 1,2,\ldots,$ for which both statements (i) and (ii) of the lemma hold.

For each $t \in \mathbb{R}_+$, let $C_t \in \mathfrak{B}(X)$ be the support of the measure $P_x^\pi(t,\cdot)$ on $(X,\mathfrak{B}(X))$, that is, C_t is the smallest closed set such that $P_x^\pi(t, X\setminus C_t) = 0$. For all $n = 1,2,\ldots$ and $t \in \mathbb{R}_+$, let $C_t^n := (X\setminus C_t)\cap X^n$. For all $n = 1,2,\ldots$, $z \in X$, $t \in \mathbb{R}_+$, and $B \in \mathfrak{B}(\tilde{A})$, define

$$\varphi^n(B|z,t) = \varphi(B\setminus\{\tilde{a}\}|z,t)I\{z \in C_t^n\} + \tilde{\varphi}^n(B|z,t)I\{z \in X\setminus C_t^n\}. \tag{4.41}$$

(i) Observe that the function φ^n defined in (4.41) is a relaxed Markov policy. Since (4.32) holds with φ^n replaced by $\tilde{\varphi}^n$, it immediately follows from (4.41) that (4.32) holds for the policy φ^n defined in (4.41).

Next, we show that the policy φ^n defined in (4.41) satisfies (4.33). From (4.1) and (4.31), for all $Z \in \mathfrak{B}(X)$, $B \in \mathfrak{B}(A)$, and $n = 1,2,\ldots$,

$$P_x^{\pi^{(n)}}(t,Z,B) \le P_x^\pi(t,Z,B). \tag{4.42}$$

Then, for all $n = 1,2,\ldots$,

$$P_x^{\pi^{(n)}}(t,C_t^n) = P_x^{\pi^{(n)}}(t,C_t^n,A) \le P_x^\pi(t,C_t^n,A) = P_x^\pi(t,C_t^n) = 0, \tag{4.43}$$

where the first equality follows from (4.2) and (4.36), the second inequality follows from (4.42), the third equality is (4.2), and the last one is true because $P_x^\pi(t,X\setminus C_t) = 0$. Thus, it follows from (4.33) with φ^n replaced by $\tilde{\varphi}^n$, (4.41), and (4.43) that the policy φ^n defined in (4.41) satisfies (4.33).

(ii) Fix an arbitrary $t \in \mathbb{R}_+$. To prove (4.40), observe that, for all $z \in X\setminus C_t$, we have $\varphi^n(B|z,t) = \varphi(B|z,t)$ for all $n > \overline{q}(z)$ and $B \in \mathfrak{B}(A)$. This follows from the property that $\overline{q}(z) < n$ for all $z \in X^n$, (4.41), and the definition of the set C_t^n. Thus, (4.40) holds for all $z \in X\setminus C_t$. To complete the proof, it remains to show that (4.40) holds for all $z \in C_t$. Since $\varphi^n(B|z,t) \le 1$ for all $B \in \mathfrak{B}(A), z \in X, t \in \mathbb{R}_+$, and $n = 1,2,\ldots$, to establish (4.40) for all $z \in C_t$, it is sufficient to show that every subsequence $\{n_k\}_{k=1,2,\ldots}$ has a subsequence $\{n_{k_m}\}_{m=1,2,\ldots}$ such that, for all $B \in \mathfrak{B}(A)$,

$$\varphi^{n_{k_m}}(B|z,t) \to \varphi(B|z,t) \quad \text{as} \quad m\to\infty \quad \text{for all} \quad z \in C_t. \tag{4.44}$$

The rest of the proof proves the existence of a subsequence $\{n_{k_m}\}_{m=1,2,\ldots}$ for every subsequence $\{n_k\}_{k=1,2,\ldots}$ of the sequence $\{1,2,\ldots\}$ such that (4.44) holds for all $B \in \mathfrak{B}(A)$.

Since $(A,\mathfrak{B}(A))$ is a standard Borel space, there exists a countable set $\mathscr{B} = \{B^1,B^2,\ldots\}$ of measurable subsets A such that $B^i\cap B^j = \emptyset$ if $i \ne j$ and any set $B \in \mathfrak{B}(A)$ can be represented as

$$B = \bigcup_{j=1,2,\ldots} B^{(j,1)} \quad \text{for some} \quad B^{(j,1)} \in \mathscr{B},\ j = 1,2,\ldots. \tag{4.45}$$

Choose an arbitrary subsequence $\{n_k\}_{k=1,2,\ldots}$ of the sequence $\{1,2,\ldots\}$. To prove (4.44) for all $B \in \mathfrak{B}(A)$, we first show the existence of a subsequence $\{n_{k_m}\}_{m=1,2,\ldots}$ of the sequence $\{n_k\}_{k=1,2,\ldots}$ such that (4.44) holds for all $B \in \mathscr{B}$. Observe that

$$\lim_{n\to\infty}\int_X |\varphi(B|z,t) - \varphi^n(B|z,t)|P_x^\pi(t,dz) = 0, \qquad B \in \mathfrak{B}(A), \tag{4.46}$$

that is, for fixed t,B, the sequence of random variables $\{\varphi^n(B|z,t), n=1,2,\ldots\}$ converge in $\mathscr{L}^1$ to $\varphi(B|z,t)$ with respect to the measure $P_x^\pi(t,\cdot)$ on $(X,\mathfrak{B}(X))$. Indeed, for all $Z\in\mathfrak{B}(X)$, $B\in\mathfrak{B}(A)$, and $n=1,2,\ldots,$

$$\begin{aligned}\int_Z[\varphi(B|z,t)-\varphi^n(B|z,t)]P_x^\pi(t,dz) &\le P_x^\pi(t,Z,B)-\int_Z\varphi^n(B|z,t)P_x^{\pi^{(n)}}(t,dz,A)\\ &= P_x^\pi(t,Z,B)-P_x^{\pi^{(n)}}(t,Z,B)+\int_Z\varphi^n(B|z,t)P_x^{\pi^{(n)}}(t,dz,\{\tilde a\})\\ &= P_x^\pi(t,Z,B)-P_x^{\pi^{(n)}}(t,Z,B)\le P_x^\pi(t,Z,A)-P_x^{\pi^{(n)}}(t,Z,A),\end{aligned} \tag{4.47}$$

where the first inequality follows from (4.2), (4.4), and (4.42), the second equality follows from (4.2) and (4.33), the third equality follows from (4.32) and (4.36), and the last one follows from (4.42) and from the property that, for fixed t, the functions $P_x^{\pi^{(n)}}(t,\cdot,\cdot)$ and $P_x^\pi(t,\cdot,\cdot)$ are measures on $X\times\tilde A$ and $X\times A$, respectively, and,

$$\begin{aligned}&\int_Z[\varphi^n(B|z,t)-\varphi(B|z,t)]P_x^\pi(t,dz)\le\int_Z\varphi^n(B|z,t)P_x^\pi(t,dz)-P_x^{\pi^{(n)}}(t,Z,B)\\ &\quad=\int_Z\varphi^n(B|z,t)[P_x^\pi(t,dz)-P_x^{\pi^{(n)}}(t,dz)]\le\int_Z\varphi^n(B|z,t)[P_x^\pi(t,dz,A)-P_x^{\pi^{(n)}}(t,dz,A)]\\ &\quad\le P_x^\pi(t,Z,A)-P_x^{\pi^{(n)}}(t,Z,A),\end{aligned} \tag{4.48}$$

where the first inequality follows from (4.4) and (4.42), the second equality follows from (4.33), the third inequality follows from (4.2) applied to the policies π and $\pi^{(n)}$, and the fourth inequality is correct because of (4.42) and since $\varphi^n(B|z,t)\le 1$. Then, for all $B\in\mathfrak{B}(A)$,

$$\begin{aligned}&\int_X|\varphi(B|z,t)-\varphi^n(B|z,t)|P_x^\pi(t,dz)\\ &\qquad=\int_{X^{(+,B)}}[\varphi(B|z,t)-\varphi^n(B|z,t)]P_x^\pi(t,dz)+\int_{X^{(-,B)}}[\varphi^n(B|z,t)-\varphi(B|z,t)]P_x^\pi(t,dz)\\ &\qquad\le P_x^\pi(t,X^{(+,B)},A)-P_x^{\pi^{(n)}}(t,X^{(+,B)},A)+P_x^\pi(t,X^{(-,B)},A)-P_x^{\pi^{(n)}}(t,X^{(-,B)},A)\\ &\qquad=P_x^\pi(t,X,A)-P_x^{\pi^{(n)}}(t,X,A),\end{aligned}$$

where $X^{(+,B)}:=\{z\in X:\varphi(B|z,t)\ge\varphi^n(B|z,t)\}$ and $X^{(-,B)}:=X\setminus X^{(+,B)}$ for each $B\in\mathfrak{B}(A)$, and the inequality follows from (4.47) with $Z=X^{(+,B)}$ and (4.48) with $Z=X^{(-,B)}$. Thus, (4.46) follows from the above inequality and (4.29) with $\gamma(\{x\})=1$.

Denote $n_{(0,k)}=n_k$ for all $k=1,2,\ldots$. For $j=1,2,\ldots$, from (4.46), Jacod and Protter [18, Theorems 17.2, 17.3], and from the property that $P_x^\pi(t,Z)>0$ for any measurable subset Z of C_t, there exists a subsequence $\{n_{(j,k)}\}_{k=1,2,\ldots}$ of the sequence $\{n_{(j-1,k)}\}_{k=1,2,\ldots}$ such that

$$\varphi^{n_{(j,k)}}(B^j|z,t)\to\varphi(B^j|z,t)\quad\text{as}\quad k\to\infty\quad\text{for all}\quad z\in C_t. \tag{4.49}$$

Let $n_{k_m}=n_{(m,m)}$, $m=1,2,\ldots$. As follows from Cantor's diagonal argument, (4.44) holds with

$B = B^j$ for all $j = 1,2,\ldots$. In other words, (4.44) is proved for all $B \in \mathscr{B}$.

Let B be an arbitrary set from $\mathfrak{B}(A)$ and $B^{(1,1)}, B^{(2,1)}, \ldots$ be disjoint elements of $\mathscr{B}$ satisfying (4.45). For all $z \in C_t$,

$$\begin{aligned}\liminf_{m\to\infty} \varphi^{n_{k_m}}(B|z,t) &= \liminf_{m\to\infty} \sum_{j=1,2,\ldots} \varphi^{n_{k_m}}(B^{(j,1)}|z,t) \geq \sum_{j=1,2,\ldots} \liminf_{m\to\infty} \varphi^{n_{k_m}}(B^{(j,1)}|z,t) \\ &= \sum_{j=1,2,\ldots} \varphi(B^{(j,1)}|z,t) = \varphi(B|z,t),\end{aligned} \tag{4.50}$$

where the first and last equalities follow from the countable additivity of probability measures, the third equality holds since (4.44) is correct for all $B \in \mathscr{B}$, and the inequality follows from Fatou's lemma [31, p. 226]. In addition, for all $z \in C_t$,

$$\begin{aligned}\limsup_{m\to\infty} \varphi^{n_{k_m}}(B|z,t) &= 1 - \liminf_{m\to\infty} \varphi^{n_{k_m}}(\tilde{A}\setminus B|z,t) = 1 - \liminf_{m\to\infty} \varphi^{n_{k_m}}(A\setminus B|z,t) \\ &\leq 1 - \varphi(A\setminus B|z,t) = \varphi(B|z,t),\end{aligned} \tag{4.51}$$

where the first equality follows from sub-additivity property of limit superior and the property that $\varphi^{n_{k_m}}(\,\cdot\,|z,t)$ is a measure on $(\tilde{A}, \mathfrak{B}(\tilde{A}))$ for each z,t, the second equality is correct because $\varphi^n(\tilde{a}|z,t) = 0$ for all $n \geq \overline{q}(z)$ (see (4.32)), the third inequality follows from (4.50) with $\overline{B} = A \setminus B$, and the last one is correct since $\varphi(\,\cdot\,|z,t)$ is a measure on $(A, \mathfrak{B}(A))$ for each z,t. Therefore, it follows from (4.50) and (4.51) that (4.44) holds for all $B \in \mathfrak{B}(A)$ and $z \in C_t$. $\square$

Lemma 4.4.11 is a special case of Theorem 4.2.2, and shows that Theorem 4.2.2 holds if the initial distribution γ is a Dirac measure.

Lemma 4.4.11. *Let the initial state x be fixed. For any policy π, consider a relaxed Markov policy φ satisfying* (4.3) *with $\gamma(\{x\}) = 1$. Then,* (4.5) *and* (4.6) *hold with $\gamma(\{x\}) = 1$.*

Proof. Let $\pi^{(n)}, n = 1,2,\ldots$, be a policy in the extended CTMDP satisfying (4.28). Then, from Lemma 4.4.10, there exists a sequence of relaxed Markov policies $\{\varphi^n, n = 1,2,\ldots\}$ such that, for all $n = 1,2,\ldots$, the relaxed Markov policy φ^n satisfies (4.32) and (4.33), and in addition (4.40) holds. Observe that if

$$\lim_{n\to\infty} P_x^{\varphi^n}(t,Z,A) = P_x^{\varphi}(t,Z), \qquad t \in \mathbb{R}_+, Z \in \mathfrak{B}(X), \tag{4.52}$$

then Lemma 4.4.5 and Corollary 4.4.9 imply the statement of this lemma. To prove (4.52), we first show by induction that, for all $m = 0,1,\ldots$, $u,t \in \mathbb{R}_+$, $u < t$, $z \in X$, and $Z \in \mathfrak{B}(X)$,

$$\lim_{n\to\infty} P^{(\varphi^n,m)}(u,z;t,Z\cap X^n) = P^{(\varphi,m)}(u,z;t,Z), \tag{4.53}$$

where the functions $P^{(\varphi^n,m)}$ and $P^{(\varphi,m)}$, $m = 0,1,\ldots$, are given by (4.7), (4.8).

For all $z \in X$, $s \in \mathbb{R}_+$, and $Z \in \mathfrak{B}(X)$,

$$\lim_{n\to\infty} \tilde{q}(Z|z,\varphi_s^n) = \lim_{n\to\infty} \int_{A(z)} q(Z|z,a)\varphi^n(da|z,s) = \int_{A(z)} q(Z|z,a)\varphi(da|z,s) = q(Z|z,\varphi_s), \tag{4.54}$$

where the first equality follows from (3.2) and (4.27), the second equality follows from Lemma 4.4.10, and the last one is (3.2). In particular, (4.54) with $Z = X\setminus\{z\}$ and (3.2) imply that

$$\lim_{n\to\infty}\tilde{q}(z,\varphi_s^n) = q(z,\varphi_s), \qquad z\in X, s\in\mathbb{R}_+. \tag{4.55}$$

This equality, (4.7) with $Z = Z\cap X^n$ for the relaxed Markov policy φ^n, and the property that $X^n\uparrow X$ as $n\to\infty$ imply that (4.53) holds for $m = 0$.

Assume that (4.53) holds for some $m\geq 0$. Then

$$\begin{aligned}\lim_{n\to\infty}P^{(\varphi^n,m+1)}(u,z;t,Z\cap X^n) &= \lim_{n\to\infty}\int\limits_u^t\int\limits_{X\setminus\{z\}} e^{-\int_u^s\tilde{q}(z,\varphi_w^n)dw}\tilde{q}(dy|z,\varphi_s^n)P^{(\varphi^n,m)}(s,y;t,Z\cap X^n)\,ds\\ &= \int_u^t\lim_{n\to\infty}\int_{X\setminus\{z\}} e^{-\int_u^s\tilde{q}(z,\varphi_w^n)dw}\tilde{q}(dy|z,\varphi_s^n)P^{(\varphi^n,m)}(s,y;t,Z\cap X^n)ds\\ &= \int_u^t\int_{X\setminus\{z\}} e^{-\int_u^s q(z,\varphi_w)dw}q(dy|z,\varphi_s)P^{(\varphi,m)}(s,y;t,Z)ds = P^{(\varphi,m+1)}(u,z;t,Z),\end{aligned}$$

where the first equality follows from (4.8) with $Z = Z\cap X^n$ for the relaxed Markov policy φ^n, the second and third equalities follow from (4.54), (4.55), the assumption that (4.53) holds for m, the Lebesgue dominated convergence theorem [31, p. 232], and from the property that the integrand is bounded by $\bar{q}(z)$, and the last one follows from (4.8) for the relaxed Markov policy φ. Equality (4.53) is proved.

Observe that, for all $t\in\mathbb{R}_+$ and $Z\in\mathfrak{B}(X)$,

$$\begin{aligned}\lim_{n\to\infty}P_x^{\varphi^n}(t,Z,A) &= \lim_{n\to\infty}P_x^{\varphi^n}(t,Z\cap X^n) = \lim_{n\to\infty}\sum_{m=0,1,\dots}P^{(\varphi^n,m)}(0,x;t,Z\cap X^n),\\ &= \sum_{m=0,1,\dots}P^{(\varphi,m)}(0,x;t,Z) = P_x^{\varphi}(t,Z),\end{aligned} \tag{4.56}$$

where the first equality follows from (4.1), (4.2), and (4.32), the second and the fourth equalities follow from (4.9) and (4.12) for the relaxed Markov policies φ^n and φ, respectively, and the third equality is correct due to Lebesgue dominated convergence theorem [31, p. 229] since (4.53) holds and $P^{(\varphi^n,m)}(0,x;t,Z\cap X^n)\leq 1$ for all $m = 0,1,\dots$ and $n = 1,2,\dots$. Thus, equation (4.52) holds. □

Proof of Theorem 4.2.2. For isolated points $x'\notin X$ and $a',a''\notin A$, let $X' := X\cup\{x'\}$, $A' := A\cup\{a',a''\}$, $A'(x) := A(x)\cup\{a''\}$ for all $x\in X$, and $A'(x') := \{a',a''\}$. For all $x\in X', a\in A'(x)$, and $Z\in\mathfrak{B}(X')$, define the new transition intensity q' by

$$q'(Z|x,a) := \begin{cases} q(Z\setminus\{x'\}|x,a), & \text{if } x\in X, a\in A(x),\\ \gamma(Z\setminus\{x'\}) & \text{if } x = x', a = a',\\ 0, & \text{if } x\in X', a = a''.\end{cases} \tag{4.57}$$

Consider the extended CTMDP $\{X', A', A'(x), q'(\cdot|x,a)\}$. Fix $T \in]0,\infty[$. To prove the statement of the theorem, we consider the evolution of the stochastic process defined by a policy and initial state x' over the time horizon $[-T,\infty[$. Similar to the measurable space $(\Omega,\mathscr{F})$ described in Section 3.3, let $(\Omega',\mathscr{F}')$ be the measurable space over which the stochastic process $\{\xi_t : t \in [-T,\infty[\}$ is defined in the extended CTMDP.

For any policy ϕ in the original CTMDP, let ϕ' denote a policy in the extended CTMDP that selects action a' up to time $\theta := \min\{t_1',0\}$, where t_1' denotes the time of the first jump starting from time $-T$, selects the absorbing action a'' for time $t \in]\theta,0]$ and for $t \in]0,\infty[$ if $\xi_0 = x'$, and plays the policy ϕ starting from time 0 if $\xi_0 \in X$. Then, for all $Z \in \mathfrak{B}(X)$, $B \in \mathfrak{B}(A')$, and $t \in \mathbb{R}_+$,

$$\begin{aligned}
P_{x'}^{\phi'}(t,Z,B) &= \int_{\Omega'} I\{\xi_t \in Z\}\phi'(B|\omega,t)\mathbb{P}_{x'}^{\phi'}(d\omega,\xi_0 \in X) \\
&\quad + \int_{\Omega'} I\{\xi_t \in Z\}\phi'(B|\omega,t)\mathbb{P}_{x'}^{\phi'}(d\omega,\xi_0 = \{x'\}) = \int_{\Omega'} I\{\xi_t \in Z\}\phi'(B|\omega,t)\mathbb{P}_{x'}^{\phi'}(d\omega,\xi_0 \in X) \\
&= \int_{\Omega} I\{\xi_t \in Z\}\phi(B\setminus\{a',a''\}|\omega,t)(1-e^{-T})\int_X \mathbb{P}_z^{\phi}(d\omega)\gamma(dz) = (1-e^{-T})P_{\gamma}^{\phi}(t,Z,B\setminus\{a',a''\}),
\end{aligned} \tag{4.58}$$

where the first equality follows from (4.1) for the policy ϕ', the second equality is correct because, under the policy ϕ', we have $\xi_t \in X$ for $t < t_\infty$ if and only if $\xi_0 \in X$, the third equality follows from (4.57) and the definition of policy ϕ' which imply that $\mathbb{P}_{x'}^{\phi'}(\xi_0 \in Z) = (1-e^{-T})\gamma(Z)$, $Z \in \mathfrak{B}(X)$ and, given $\xi_0 \in X$, the process is defined by the policy ϕ and initial state ξ_0, and the one follows from (4.1).

Applying (4.58) for the policies π and φ satisfying (4.3), for all $Z \in \mathfrak{B}(X)$, $B \in \mathfrak{B}(A')$, and $t \in \mathbb{R}_+$,

$$P_{x'}^{\pi'}(t,Z,B) = (1-e^{-T})P_{\gamma}^{\pi}(t,Z,B\setminus\{a',a''\}) \text{ and } P_{x'}^{\varphi'}(t,Z,B) = (1-e^{-T})P_{\gamma}^{\varphi}(t,Z,B\setminus\{a',a''\}). \tag{4.59}$$

Thus, (4.5) holds if

$$P_{x'}^{\pi'}(t,Z,B) = P_{x'}^{\varphi'}(t,Z,B), \qquad t \in \mathbb{R}_+, Z \in \mathfrak{B}(X), B \in \mathfrak{B}(A). \tag{4.60}$$

So, to complete the proof of the theorem, it is sufficient to verify (4.60).

Observe that, for all $t \in \mathbb{R}_+, Z \in \mathfrak{B}(X)$, and $B \in \mathfrak{B}(A')$,

$$\begin{aligned}
P_{x'}^{\pi'}(t,Z,B) &= (1-e^{-T})P_{\gamma}^{\pi}(t,Z,B\setminus\{a',a''\}) \\
&= (1-e^{-T})\int_Z \varphi(B\setminus\{a',a''\}|z,t)P_{\gamma}^{\pi}(t,dz,A) = \int_Z \varphi'(B|z,t)P_{x'}^{\pi'}(t,dz),
\end{aligned} \tag{4.61}$$

where the first equality follows from (4.58) with $\phi = \pi$, the second equality follows from (4.2) and (4.3), and the last one follows from (4.2) for the policy π' in the extended CTMDP, (4.59) with $B = A'$, and the definition of the policy φ'. In addition, for all $t \in [-T_0,0[$, $Z \in \mathfrak{B}(X)$, and

$B \in \mathfrak{B}(A')$,

$$P_{x'}^{\pi'}(t,Z,B) = \int_{\Omega'} I\{\xi_t \in Z\}\pi'(B|\omega,t)\mathbb{P}_{x'}^{\pi'}(d\omega) \\ = \int_{\Omega'} I\{\xi_t \in Z\}\varphi'(B|\xi_t,t)P_{x'}^{\pi'}(d\omega) = \int_Z \varphi'(B|z,t)P_{x'}^{\pi'}(t,dz), \tag{4.62}$$

where the first and last equalities follow respectively from (4.1) and (4.2) for the policy π' in the extended CTMDP, and the second equality holds by the definitions of the policies π' and φ' which imply that $\pi'(B|\omega,t) = \varphi'(B|\xi_t,t)$ for all $\omega \in \Omega'$, $t \in [-T,0[$, and $B \in \mathfrak{B}(A')$. Then, from (4.61) and (4.62),

$$P_{x'}^{\pi'}(t,Z,B) = \int_Z \varphi'(B|z,t)P_{x'}^{\pi'}(t,dz), \quad t \in [-T,\infty[, Z \in \mathfrak{B}(X'), B \in \mathfrak{B}(A'). \tag{4.63}$$

Therefore, (4.60) follows from Lemma 4.4.11. □

4.5 Sufficiency of relaxed Markov policies for particular objective criteria

In this section, using the results in Section 4.2, we characterize the equivalence of the class of history-dependent policies and the class of relaxed Markov policies for the above mentioned cost criteria. We tackle this problem in two steps: we consider (i) a CTMDP with zero instantaneous costs, and (ii) a CTMDP with non-zero instantaneous costs.

4.5.1 The case of zero instantaneous costs

Theorem 4.5.1. *Given an initial distribution γ on X, for any policy π there exists a relaxed Markov policy φ such that*

$$\mathbb{E}_\gamma^\varphi c(\xi_t,\varphi_t) = \mathbb{E}_\gamma^\pi c(\xi_t,\pi_t), \qquad \textit{if} \quad t < t_\infty, \tag{4.64}$$

and therefore, for the case of zero instantaneous costs, $V_{T,\alpha}(\gamma,\varphi) = V_{T,\alpha}(\gamma,\pi)$ and $V_\alpha(\gamma,\varphi) = V_\alpha(\gamma,\pi)$ for $\alpha \geq 0$, and $W(\gamma,\varphi) = W(\gamma,\pi)$ when $\mathbb{P}_\gamma^\varphi(\xi_t \in X) = \mathbb{P}_\gamma^\pi(\xi_t \in X) = 1$ for all $t \in \mathbb{R}_+$.

Proof. Observe that, for any policy ϕ

$$\mathbb{E}_\gamma^\phi[\int_0^{T\wedge t_\infty} e^{-\alpha s}c(\xi_s,\phi_s)ds] = \mathbb{E}_\gamma^\phi[\int_0^T e^{-\alpha s}c(\xi_s,\phi_s)I\{s < t_\infty\}ds], \qquad \alpha \geq 0, T > 0, \tag{4.65}$$

and, for all $t \in \mathbb{R}_+$,

$$\mathbb{E}_\gamma^\phi c(\xi_t,\phi_t)I\{t < t_\infty\} = \int_A \int_\Omega c(\xi_t,a)\phi(da|\omega,t)I\{\xi_t \in X\}\mathbb{P}_\gamma^\phi(d\omega) = \int_A \int_X c(z,a)P_\gamma^\phi(t,dz,da), \tag{4.66}$$

where the first equality follows from the definition of expectation, from exchanging the order of integration, (3.5), and the property that $\{\xi_t \in X\} = \{t < t_\infty\}$, and the last one follows from (4.1).

For each policy π, consider a relaxed Markov policy φ satisfying (4.3). In view of Theorem 4.2.2, we have that $P_\gamma^\varphi(t,Z,B) = P_\gamma^\pi(t,Z,B)$ for all $t \in \mathbb{R}_+$, $Z \in \mathfrak{B}(X)$, and $B \in \mathfrak{B}(A)$. This fact along with (4.66) applied to the policies φ and π implies that (4.64) holds. The second statement of the theorem immediately follows from the definitions of cost criteria given prior to the theorem, (4.64), and (4.65) applied to the policies π and φ. □

4.5.2 The case of non-zero instantaneous costs

Recall that, at each time $t < t_\infty$, the relaxed policy π selects the probability measure $\pi_t \in \mathscr{P}(A)$, and the cost rate at time t is equal to $c(\xi_t,\pi_t)$ and the instantaneous cost incurred if $t = t_n$ for some $n = 1,2,\dots$ is $C(\xi_{t_{n-1}},\pi_{t_n},\xi_{t_n})$. Given an initial distribution γ, for any policy π, consider the occupancy measure

$$H_\gamma^\pi(Z,\mathbf{B}) = \mathbb{E}_\gamma^\pi \int_0^{t_\infty} e^{-\alpha t} I\{\xi_t \in Z, \pi_t \in \mathbf{B}\}, \quad Z \in \mathfrak{B}(X), \mathbf{B} \in \mathfrak{B}(\mathscr{P}(A)). \tag{4.67}$$

It follows from Feinberg [6, Corollary 4.4] that the expected total discounted cost $V_\alpha(\gamma,\pi)$ satisfies

$$V_\alpha(\gamma,\pi) = \int_X \int_{P(A)} \left[c(z,p) + \int_{X\setminus\{z\}} C(z,p,y) q(dy|z,p) \right] H_\gamma^\pi(dz,dp).$$

This equality and (4.67) imply that

$$V_\alpha(\gamma,\pi) = \mathbb{E}_\gamma^\pi \int_0^{t_\infty} e^{-\alpha t} \left[c(\xi_t,\pi_t) + \int_{X\setminus\{\xi_t\}} C(\xi_t,\pi_t,y) q(dy|\xi_t,\pi_t) \right] dt. \tag{4.68}$$

The following theorem establishes the sufficiency of relaxed Markov policies for the expected total discounted cost criteria if the instantaneous cost function C depends only on the state prior to the jump epoch.

Theorem 4.5.2. *Consider a CTMDP such that the instantaneous cost function $C(x,a,y), x,y \in X$ with $x \neq y$ and $a \in A(x)$ depends only on x. Given an initial distribution γ, for any policy π there exists a relaxed Markov policy φ such that $V_\alpha(\gamma,\varphi) = V_\alpha(\gamma,\pi)$.*

Proof. Define

$$\hat{c}(z,p) = c(z,p) + C(z)q(z,p), \quad z \in X, p \in \mathscr{P}(A). \tag{4.69}$$

From the above equality and (4.68), for any policy ϕ,

$$V_\alpha(\gamma,\phi) = \mathbb{E}_\gamma^\pi \int_0^{t_\infty} e^{-\alpha t} \hat{c}(\xi_t,\pi_t) dt. \tag{4.70}$$

Thus, (4.70) and Theorem 4.5.1, with $c(z,p) = \hat{c}(z,p)$ imply that the statement of the theorem holds. □

In Theorems 4.5.1 and 4.5.2, the equivalent relaxed Markov policy φ corresponding to an arbitrary policy π is constructed using (4.4). In general, the expected total discounted costs need not be the same for the policy π and its corresponding relaxed Markov policy φ that satisfies (4.3). Consider the following example:

Example 4.5.3. *The relaxed Markov policy φ satisfying* (4.4) *may perform better than the policy π when the instantaneous costs C depend on the action chosen at the jump epochs.* Let $X = \{1,2\}, A = \{b,c\}, A(1) = A, A(2) = b, q(1,b) = q(2,b) = 4$, and $q(1,c) = 2$. The cost rate function $c(x,a) = 0$ for all $x \in X$ and $a \in A$, and the instantaneous costs C are given by $C(1,b,2) = 2$, $C(1,c,2) = 4$, and $C(2,b,1) = 0$. Consider a policy π satisfying

$$\pi(\omega,t) = \begin{cases} b, & \text{if } \xi_t(\omega) = 2, \\ b, & \text{if } \xi_t(\omega) = 1 \text{ and } N \text{ is even or } N = 0, \\ c, & \text{if } \xi_t(\omega) = 1 \text{ and } N \text{ is odd,} \end{cases} \tag{4.71}$$

where N represents the number of jumps into state 1 upto time t.

Let φ be a relaxed Markov policy such that, for all $t \in \mathbb{R}_+$,

$$\varphi(b \mid 2,t) := 1 \qquad \text{and} \qquad \varphi(\cdot \mid 1,t) := \begin{cases} \frac{P_1^\pi(t,1,\cdot)}{P_1^\pi(t,1)}, & \text{if } P_1^\pi(t,1) > 0, \\ \text{arbitrary}, & \text{otherwise.} \end{cases} \tag{4.72}$$

Note that the relaxed Markov policy φ satisfies (4.3) with $\gamma(\{1\}) = 1$, and, therefore it follows from Theorem 4.2.2 that $P_1^\pi(t;\cdot) = P_1^\varphi(t;\cdot)$ for all $t \in \mathbb{R}_+$.

The expected total discounted cost corresponding to the relaxed Markov policy φ and the initial state 1 is given by

$$\begin{aligned} V_\alpha(1,\varphi) &= \mathbb{E}_1^\varphi\left[\int_0^{t_\infty} e^{-\alpha s} C(1,\varphi_s,2) q(1,\varphi_s) I\{\xi_s = 1\} ds\right] \\ &= \int_\Omega \int_0^\infty e^{-\alpha s} C(1,\varphi_s,2) q(1,\varphi_s) I\{\xi_s = 1\} ds\, \mathbb{P}_1^\varphi(d\omega) \\ &= \int_0^\infty e^{-\alpha s} \int_\Omega C(1,\varphi_s,2) q(1,\varphi_s) I\{\xi_s = 1\} \mathbb{P}_1^\varphi(d\omega) ds \\ &= \int_0^\infty e^{-\alpha s} \left[2\frac{P_1^\pi(s,1,b)}{P_1^\pi(s,1)} + 4\frac{P_1^\pi(s,1,c)}{P_1^\pi(s,1)}\right]\left[4\frac{P_1^\pi(s;1,b)}{P_1^\pi(s,1)} + 2\frac{P_1^\pi(s,1,c)}{P_1^\pi(s,1)}\right] P_1^\varphi(s,1) ds \\ &= \int_0^\infty e^{-\alpha s} \frac{[2P_1^\pi(s,1,b) + 4P_1^\pi(s,1,c)][4P_1^\pi(s,1,b) + 2P_1^\pi(s,1,c)]}{P_1^\pi(s,1)} ds, \end{aligned} \tag{4.73}$$

where the first equality follows from (4.68), the second equality follows from the property that $\{\xi_s = 1\} \in \{s < t_\infty\}$ and from the definition of expectation, the third equality is obtained by interchanging the order of integration, the fourth equality follows from (3.2), (4.2), (3.6), and (4.72), and the last one holds since $P_1^\pi(t;1) = P_1^\varphi(t;1)$ for all $t \in \mathbb{R}_+$.

Similarly, the expected total discounted cost corresponding to the policy π and initial state 1 is given by

$$
\begin{aligned}
V_\alpha(1,\pi) &= \mathbb{E}_1^\pi\left[\int_0^{t_\infty} e^{-\alpha s} C(1,\pi_s,2) q(1,\pi_s) I\{\xi_s = 1\} ds\right] \\
&= \int_\Omega \int_0^\infty e^{-\alpha s} C(1,\pi_s,2) q(1,\pi_s) I\{\xi_s = 1\} ds \, \mathbb{P}_1^\pi(d\omega) \\
&= \int_\Omega \int_0^\infty e^{-\alpha s} [(2)(4) I\{\pi(\omega,s) = b\} + (4)(2) I\{\pi(\omega,s) = c\}] I\{\xi_s = 1\} ds \, \mathbb{P}_1^\pi(d\omega) \\
&= \int_0^\infty \int_\Omega 8 e^{-\alpha s} I\{\xi_s = 1\} \mathbb{P}_1^\pi(d\omega) ds = \int_0^\infty 8 e^{-\alpha s} P_1^\pi(s,1) ds,
\end{aligned}
\tag{4.74}
$$

where the first equality follows from (4.68), the second equality holds because $\{\xi_s = 1\} \in \{s < t_\infty\}$ and from the definition of expectation, the third equality follows from (4.71), the fourth equality is obtained by interchanging the order of integration, and the last one follows from (4.2).

For notational convenience, let $x_1 := P_1^\pi(s,1,b)$ and $x_2 := P_1^\pi(s,1,c)$. Then $P_1^\pi(s,1) = x_1 + x_2$ and

$$
V_\alpha(1,\varphi) = \int_0^\infty e^{-\alpha s} \frac{(8x_1^2 + 8x_2^2 + 20 x_1 x_2)}{x_1 + x_2} ds \quad \text{and} \quad V_\alpha(1,\pi) = \int_0^\infty e^{-\alpha s} \frac{(8x_1^2 + 8x_2^2 + 16 x_1 x_2)}{x_1 + x_2} ds.
$$

Therefore, $V_\alpha(1,\varphi) = V_\alpha(1,\pi) + \int_0^\infty 4 e^{-\alpha s} \frac{P_1^\pi(s,1,b) P_1^\pi(s,1,c)}{P_1^\pi(s,1)} ds$ which implies that $V_\alpha(1,\varphi) > V_\alpha(1,\pi)$.

Bibliography

[1] W.J. Anderson, Continuous-Time Markov Chains: An Applications-Oriented Approach, Springer Series in Statistics, Springer-Verlag, New York, 1991.

[2] D.P. Bertsekas, S.E. Shreve, Stochastic Optimal Control: The Discrete-time Case, Academic Press, New York, 1978.

[3] P. Brémaud, Point processes and queues: Martingale dynamics, Springer Series in Statistics, Springer-Verlag, New York, 1981.

[4] C. Derman and R.E. Strauch, A note on memoryless rules for controlling sequential control processes, Ann. Math. Statist. 37 (1966) 267–278.

[5] J.L. Doob, Stochastic Process, reprint of the 1953 original, John Wiley, New York, 1990.

[6] E.A. Feinberg, Continuous time discounted jump Markov decision processes: A discrete-event approach, Math. Oper. Res. 29 (3) (2004) 492–524.

[7] E.A. Feinberg, Reduction of Discounted Continuous-Time MDPs with Unbounded Jump and Reward Rates to Discrete-Time Total-Reward MDPs in *Optimization, Control, and Applications of Stochastic Systems, D. Hernandez and A. Minjarez (eds)*, 77–98, Birkhäuser/Springer, New York, 2012.

[8] E.A. Feinberg, M. Mandava, A.N. Shiryaev, Sufficiency of Markov policies for continuous-time Markov decision processes and solutions to Kolmogorov's forward equation for jump Markov processes, in: Proceedings of 2013 IEEE 52nd Annual Conference on Decision and Control, Florence, Italy, December 10-13, (2013) 5728–5732.

[9] E.A. Feinberg, M. Mandava, A.N. Shiryaev, On solutions of KolmogorovÊijs equations for nonhomogeneous jump Markov processes, J. Math. Anal. Appl. 411(1) (2014) 261–270.

[10] W. Feller, On the integro-differential equations of purely-discontinuous Markoff processes, Tran. Amer. Math. Soc. 48 (1940) 488–515. Errata, Trans. Amer. Math. Soc. 58 (1945) 474.

[11] X. Guo, O. Hernández-Lerma, Continuous-Time Markov decision processes: Theory and Applications, Stochastic Modeling and Applied Probability 62, Springer-Verlag, Berlin, 2009.

[12] X. Guo, A.B. Piunovskiy, Discounted continuous-time Markov decision processes with constraints: unbounded transition and loss rates, Math. Oper. Res. 36 (1) (2011) 105–132.

[13] X. Guo, U. Rieder, Average optimality for continuous-time Markov decision processes in Polish spaces, Ann. Appl. Probab. 16(2) (2006) 730–756.

[14] X. Guo, X. Song, Discounted continuous-time constrained Markov decision processes in Polish spaces, Ann. Appl. Probab. 21(5) (2011) 2016–2049.

[15] O. Hernández-Lerma, J.B. Lassarre, Discrete-time Markov Control Processes: Basic Optimality Criteria, Applications of Mathematics (New York), 30, Springer-Verlag, New York, 1966.

[16] A. Hordijk, F.A. van der Duyn Schouten, Discretization Procedures for Continuous Time Decision Processes in *Trans. Eighth Prague Conf. on Inform. Theory, Statistical Decision Functions, Random Processes, Prague, Czechoslovakia, 1978,* C (1979) 143–154.

[17] J. Jacod, Multivariate point processes: predictable projection, Radon-Nikodym derivatives, representation of martingales, Probab. Theory Related Fields 31 (3) (1975) 235–253.

[18] J. Jacod, P. Protter, Probability Essentials, second ed., Springer, New York, 2004.

[19] P. Kakumanu, Continuously Discounted Markov Decision Model with Countable State and Action Space. Ann. Math. Stat. 42(3) (1971) 919–926.

[20] O. Kallenberg, Foundations of Modern Probability, first ed., Springer, New York, 1997.

[21] A.S. Kechris, Classical descriptive set theory, Graduate Texts in Mathematics, (156), Springer, New York, 1995.

[22] D.G. Kendall, Some further pathological examples in the theory of denumerable Markov processes, Q. J. Math. 7 (1) (1956) 39–56.

[23] D.G. Kendall, G.E.H. Reuter, Some pathological Markov processes with a denumerable infinity of states and the associated semigroups of operators on l, in: Proceedings of the International Congress of Mathematicians, Vol. III, Amsterdam, 1954, pp. 377âĂŞ-415.

[24] M.Yu. Kitaev, Semi-Markov and jump Markov controlled models: average cost criterion, Theory Prob. Appl. 30 (2) (1986) 272–288.

[25] M.Yu. Kitaev, V.V. Rykov, Controlled queueing systems, CRC Press, Boca Raton, 1995.

[26] S.E. Kuznetsov, Any Markov process in a Borel space has a transition function, Theory Probab. Appl. 25 (2) (1981) 384–388.

[27] B.L. Miller, Finite State Continuous Time Markov Decision Processes with a Finite Planning Horizon, SIAM J. Control and Optim. 6 (1968) 266–280.

[28] B.L. Miller, Finite State Continuous Time Markov Decision Processes with an Infinite Planning Horizon, J. Math. Anal. Appl. 22 (1968) 552–569.

[29] A. Piunovskiy, Y. Zhang, Discounted continuous-time Markov decision processes with unbounded rates: the convex analytic approach, SIAM J. Control Optim. 49(5) (2011) 2032–2061.

[30] G.E.H. Reuter, Denumerable Markov processes and the associated contraction semigroups on l, Acta Math. 97 (1) (1957) 1–46.

[31] H.L. Royden, Real Analysis, Second edition, Macmillan, New York, 1963.

[32] W. Rudin, Real and Complex analysis, Third edition, McGraw-Hill, Inc., New York, 1987.

[33] R.E. Strauch, Negative dynamic programming, Ann. Math. Statist. 37(4) (1966 871–890.

[34] L. Ye, X. Guo, O. Hernández-Lerma, Existence and regularity of a nonhomogeneous transition matrix under measurability conditions, J. Theoret. Probab., 21 (3) (2008) 604–627

[35] A. A. Yushkevich, Controlled Markov Models with Countable State Space and Continuous Time, Theory Probab. Appl., 25 (1980) 244–266.

Chapter 5

Appendix A

In this appendix, we describe discrete-time Markov decision processes (DTMDPs), provide a brief construction of the stochastic processes induced by a policy, and define the common optimality criteria. We present the result on sufficiency of Markov policies in DTMDPs that was shown by Derman and Strauch [4].

5.1 Discrete-time Markov decision processes

A DTMDP is defined by the multiplet $(X, A, A(x), p(\cdot|x,a), c(x,a))$, where

(i) X is the state space such that $(X, \mathfrak{B}(X))$ is a standard Borel space;

(ii) A is the action space such that $(A, \mathfrak{B}(A))$ is a standard Borel space;

(iii) $A(x)$ are the set of actions available at $x \in X$. It is assumed that $A(x) \in \mathfrak{B}(X)$ for all $x \in X$ and the set of feasible state-action pairs $Gr(A) = \{(x,a) : x \in X, a \in A(x)\}$ is a measurable subset of $(X \times A)$ containing the graph of a measurable mapping from X to A.

(iv) $p(\cdot|x,a)$ is the transition probability from $Gr(A)$ to X, that is $p(\cdot|x,a)$ is a probability measure on $(X, \mathfrak{B}(X))$ for any $(x,a) \in Gr(A)$ and $p(Z|x,a)$ is a measurable function on $Gr(A)$ for any $Z \in \mathfrak{B}(X)$.

(v) $c(x,a)$ is the one step cost incurred for choosing an action $a \in A(x)$ in state $x \in X$ and is assumed to be a bounded below measurable function on the $Gr(A)$.

A DTMDP evolves as follows: At each time epoch $t = 0, 1, \ldots$, the decision maker observes the current state $x \in X$ of the stochastic system and chooses an action a from the set of actions $A(x)$ available at state x. After an action a is selected, the decision maker incurs the cost $c(x,a)$ and the system moves to the next state at time $t+1$ according to the probability law $p(\cdot|x,a)$.

The decision rules that specify how actions are chosen by the decision maker at every time t using the available information are called *policies*. In DTMDPs, the information available up to the current time t is $(x_0, a_0, \ldots, x_{t-1}, a_{t-1}, x_t)$, where x_n and a_n, $n = 0, 1, \ldots$, respectively denote the state of the process and the action chosen in this state at time n. At each time epoch, the decision maker may select a particular action or, in a more general way, a probability distribution on the set of available actions $A(x)$. Decisions of the first type are called non-randomized and decisions of the second type are called randomized. To formally define policies, let $H_t := (X \times A)^t \times X$ be the set of histories (or information vectors) up to and including time $t = 0, 1, \dots,$ and $\mathfrak{B}(H_t) = (\mathfrak{B}(X) \otimes \mathfrak{B}(A))^t \otimes \mathfrak{B}(X)$. We now define different classes of policies considered for DTMDPs in this thesis.

- A *(randomized) policy* π is a sequence of transition probabilities $\pi_t, t = 0, 1, \dots,$ from H_t to A such that $\pi_t(A(x_t)|x_0, a_0, \dots, x_{t-1}, a_{t-1}, x_t) = 1$.

- A policy π is called *(randomized) Markov* if $\pi_t(\cdot|x_0, a_0, \dots, x_{t-1}, a_{t-1}, x_t) = \pi_t(\cdot|x_t)$ for all $t = 0, 1, \dots$.

Let Δ be the set of all policies and Δ^M be the set of all Markov policies. Then, we immediately have, $\Delta^M \subseteq \Delta$. Also, let $H := \bigcup_{t=0,1,\dots} H_t$ be the set of histories that contain countable number of jumps. Observe that the set H is endowed with a σ-algebra defined by the products of the Borel σ-algebras $\mathfrak{B}(X)$ and $\mathfrak{B}(A)$. In view of Ionescu Tulcea theorem (Hernández-Lerma and Lasserre [15, p. 178]), an initial probability measure γ on X and a policy π define a unique probability measure $\mathbb{P}^\pi_\gamma$ on the space of histories H. We denote by $\mathbb{E}^\pi_\gamma$ the expectations with respect to the probability measure $\mathbb{P}^\pi_\gamma$.

5.1.1 Cost Criteria

We now give a brief description of the different cost criteria considered in this thesis for DTMDPs. Given an initial distribution γ on X, for any policy $\pi \in \Delta$,

(i) the *finite horizon expected total discounted cost* is given by

$$\mathbf{V}_{\beta,T}(\gamma, \pi) := \mathbb{E}^\pi_\gamma \sum_{t=0}^{T-1} \beta^t c(x_t, a_t), \tag{5.1}$$

where T is the finite planning horizon and $\beta \in [0, 1[$ is the discount factor.

(ii) Formula (5.1) with $\beta = 1$ defines the *finite-horizon expected total cost* denoted by $\mathbf{V}_{1,T}(\gamma, \pi)$.

(iii) Formula (5.1) with $T = \infty$ defines the *expected total discounted cost* denoted by $\mathbf{V}_\beta(\gamma, \pi)$.

(iv) Formula (5.1) with $\beta = 1$ and $T = \infty$ defines the *expected total cost* denoted by $\mathbf{V}_0(\gamma, \pi)$.

(v) the *average cost per unit time* is given by

$$\mathbf{W}(\gamma, \pi) = \limsup_{T \to \infty} \frac{\mathbf{V}_{1,T}(\gamma, \pi)}{T}. \tag{5.2}$$

5.2 Sufficiency of Markov policies in DTMDPs

The following theorem shows that, given an initial distribution γ on X, for any policy π there exists a Markov policy σ such that both the policies have the same marginal distributions on the state-action pairs. This fundamental result in DTMDPs given by Derman and Strauch [4] establishes that the search for optimal policies, when considering cost criteria that depend only on marginal distributions on the state-action pairs, can be restricted to the class of Markov policies.

Theorem 5.2.1 (Derman and Strauch [4], Strauch [33])**.** *Given an initial distribution γ on X, for any policy π consider the Markov policy satisfying, for all $t = 0,1,2,\dots,$*

$$\sigma_t(B|z) = \frac{\mathbb{P}^{\pi}_{\gamma}(x_t \in dz, a_t \in B)}{\mathbb{P}^{\pi}_{\gamma}(x_t \in dz)}, \quad \mathbb{P}^{\pi}_{\gamma}(x_t \in \cdot) - a.e., z \in X, B \in \mathfrak{B}(A). \tag{5.3}$$

Then, for all $t = 0,1,2,\dots,$

$$\mathbb{P}^{\pi}_{\gamma}(x_t \in Z, a_t \in B) = \mathbb{P}^{\sigma}_{\gamma}(x_t \in Z, a_t \in B), \qquad Z \in \mathfrak{B}(X), B \in \mathfrak{B}(A). \tag{5.4}$$

Therefore, for optimality criteria $G \in \{\boldsymbol{V}_\beta, \boldsymbol{V}_0, \boldsymbol{W}\}$,

$$G(\gamma,\pi) = G(\gamma,\sigma). \tag{5.5}$$

Proof. Observe that, for any policy ϕ and initial distribution γ,

$$\mathbb{E}^{\phi}_{\gamma} c(x_t,a_t) = \int_X \int_A c(x,a) \mathbb{P}^{\phi}_{\gamma}(x_t \in dx, a_t \in da).$$

This fact and the definitions of the optimality criteria in G imply that (5.5) is correct if (5.4) holds. The rest of the proof shows (5.4).

Observe that, if for any $t = 0,1,2,\dots,$

$$\mathbb{P}^{\sigma}_{\gamma}(x_t \in Z) = \mathbb{P}^{\pi}_{\gamma}(x_t \in Z), \qquad Z \in \mathfrak{B}(X), \tag{5.6}$$

then, for all $Z \in \mathfrak{B}(X)$, and $B \in \mathfrak{B}(A)$,

$$\mathbb{P}^{\sigma}_{\gamma}(x_t \in Z, a_t \in B) = \int_Z \sigma_t(B|z)\mathbb{P}^{\sigma}_{\gamma}(x_t \in Z) = \int_Z \sigma_t(B|z)\mathbb{P}^{\pi}_{\gamma}(x_t \in Z) = \mathbb{P}^{\pi}_{\gamma}(x_t \in Z, a_t \in B), \tag{5.7}$$

where the first equality holds since σ is a Markov policy, the second equality follows from (5.6), and the last one follows from (5.3). Thus, to complete the proof we establish (5.6). The proof of (5.6) is based on induction.

For $t = 0$, $\mathbb{P}^{\sigma}_{\gamma}(x_0 \in Z) = \mathbb{P}^{\pi}_{\gamma}(x_0 \in Z) = \gamma(Z)$ for all $Z \in \mathfrak{B}(X)$. Assume (5.6) holds for some $t \geq 0$. Then, for all $Z \in \mathfrak{B}(X)$,

$$
\begin{aligned}
\mathbb{P}^{\pi}_{\gamma}(x_{t+1} \in Z) &= \int_X \int_A \int_Z p(dy|z,a)\mathbb{P}^{\pi}_{\gamma}(x_t \in dz, a_t \in da) \\
&= \int_X \int_A \int_Z p(dy|z,a)\mathbb{P}^{\sigma}_{\gamma}(x_t \in dz, a_t \in da) = \mathbb{P}^{\sigma}_{\gamma}(x_{t+1} \in Z),
\end{aligned}
$$

where the first and last equalities are straightforward and the second equality holds due to (5.7) and (5.6). Thus, (5.6) holds for all $t = 0, 1, \ldots$. $\square$