Markov Processes and Controlled Markov Chains

Zhenting Hou, Jerzy A. Filar and Anyue Chen (Eds.)

Kluwer Academic Publishers

Markov Processes and Controlled Markov Chains

Markov Processes and Controlled Markov Chains

Edited by

Zhenting Hou

Research Department,
Changsha Railway University,
Changsha, China

Jerzy A. Filar

School of Mathematics,
University of South Australia,
Mawson Lakes, SA, Australia

and

Anyue Chen

School of Computing and Mathematical Sciences,
University of Greenwich,
London, U.K.

KLUWER ACADEMIC PUBLISHERS

DORDRECHT / BOSTON / LONDON

A C.I.P. Catalogue record for this book is available from the Library of Congress.

ISBN-13:978-1-4613-7968-3   e-ISBN-13:978-1-4613-0265-0
DOI: 10.1007/978-1-4613-0265-0

Published by Kluwer Academic Publishers,
P.O. Box 17, 3300 AA Dordrecht, The Netherlands.

Sold and distributed in North, Central and South America
by Kluwer Academic Publishers,
101 Philip Drive, Norwell, MA 02061, U.S.A.

In all other countries, sold and distributed
by Kluwer Academic Publishers,
P.O. Box 322, 3300 AH Dordrecht, The Netherlands.

Printed on acid-free paper

Contents

Preface

The general theory of stochastic processes and the more specialized theory of Markov processes evolved enormously in the second half of the last century. In parallel, and to a large extent independently, the theory of controlled Markov chains (or Markov decision processes) was being pioneered by control engineers and operations researchers. Since researchers in Markov processes and controlled Markov chains have been, for a long time, aware of the synergies between these two subject areas it was generally recognized that time was ripe to organize a conference that would bring together the leading practitioners in these fields.

In view of the above it could be argued, that an international conference devoted to the twin topics of Markov processes and controlled Markov chains was inevitable and that the only questions that needed to be settled were: when and where should such a meeting take place. We felt that 1999, the last year of the 20th century, the century during which the entire subject of probability has been formalized as a rigorous branch of mathematics was the right year to stage this conference. Furthermore, we felt that by holding it in China we would accomplish the important goal of facilitating a fruitful exchange of ideas between the international research community and the members of the vibrant Chinese school of probability. As a result, a decision was made to organize the International Workshop on Markov Processes and Controlled Markov Chains in Changsha, China, 22–28 August 1999.

The conference was a great success. It was attended by eminent scholars in their relevant disciplines, from eleven countries spanning four continents, including some of the leading Chinese experts. Stimulating plenary lectures by Professors Dynkin (Cornell University, USA), Watanabe (Kyoto University, Japan), Haurie (University of Geneva, Switzerland) and Hernandez-Lerma (CINVESTAV-PIN, Mexico) exposed the participants to some of the most important recent developments in Markov processes and controlled Markov chains. In total 94 research papers were presented at the workshop. There were also many lively discussions and new collaborative projects that resulted from this workshop.

A number of younger researchers and graduate students also actively participated in the conference.

Authors of the most interesting papers presented at the workshop were invited to submit their contributions for possible publication in this edited volume. All papers were refereed. The final selection which appears in the body of this book reflects both the maturity and the vitality of modern day Markov processes and controlled Markov chains. The maturity can be seen from the sophistication of the theorems, proofs, methods and algorithms contained in the selected papers. The vitality is manifested by the range of new ideas and new applications in such fields as finance and manufacturing.

As editors and workshop organizers we are very happy to express our thanks and appreciation to many people who have worked hard to make the workshop and this volume so successful. In particular, we are indebted to all the members of the International and Local Program Committees (IPC and LOC, respectively), and especially to the workshop secretary, Mr. Xiaobin Fang and Professor Hanjun Zhang who also helped to edit this volume. We are indebted to the many colleagues who reviewed the manuscripts and made suggestions for improvements. Ms Angela McKay and Mr Paul Haynes from the University of South Australia played an important role in converting the manuscripts into a consistent format. The thoughtful editorial oversight from Mr John Martindale from Kluwer is also gratefully acknowledged. Last but not least, the workshop was generously supported by Changsha Railway University, University of South Australia, Changsha Municipal Government, Xiangcai Securities Co., Ltd, National Science Foundation of China and the Bernoulli Society for Mathematical Statistics and Probability. Without their support the workshop and this volume would not have been possible.

ZHENTING HOU

JERZY A. FILAR

ANYUE CHEN

I

MARKOV PROCESSES

Chapter 1

BRANCHING EXIT MARKOV SYSTEM AND THEIR APPLICATIONS TO PARTIAL DIFFERENTIAL EQUATIONS*

E.B. Dynkin
Cornell University

1. Introduction

Connections between linear partial differential equations involving second order uniformly elliptic operators L and diffusion processes are known for a long time. Superdiffusions are related, in an analogous way, to equations involving semilinear differential operators $Lu - \psi(u)$.

Superdiffusions are a special case of superprocesses which were introduced (under the name continuous state branching processes) in pioneering work of Watenabe in 1968 [9]. Deep results on super-diffusion were obtained by Dawson, Perkins, Le Gall and others. Partial differential equations involving the operator $Lu - \psi(u)$ were studied independently by analysts, including Keller, Osserman, Loewner and Nirenberg, Brezis, Marcus and Veron, Baras and Pierre.

In earlier papers, a superdiffusion was interpreted as a Markov process X_t in the space of measures. This is not sufficient for the probabilistic approach to boundary value problems. A reacher model based on the concept of exit measures has been introduced in [1]. A model of a superprocess as a family of exit measures from time-space open sets was developed systematically in [3]. In particular, branching and Markov properties of such family were established and used for solving analytical problems. The central point of the present talk is to show that these two properties are sufficient to develop the entire theory of superprocesses.

*Partially supported by National Science Foundation Grant DMS-9970942

Z. Hou et al. (eds.), Markov Processes and Controlled Markov Chains, 3–13.

2. Exit systems associated with a diffusion

To every second order uniformly elliptic differential operator L there corresponds a Markov process $(\xi_t, \Pi_{r,x})$ in R^d with continuous paths and infinitesimal generator L. We call it $L-$diffusion. The process with the generator $L = \frac{1}{2}\Delta$ has the transition density

$$p(r, x; t, y) = [2\pi(t-r)]^{-d/2} e^{-|x-y|^2/2(t-r)}.$$

It is called the Brownian motion.

To every open set Q in time-space $S = R \times R^d$ there corresponds a random point (τ, ξ_τ) where $\tau = \inf\{t : (t, \xi_t) \notin Q\}$ is the first exit time from Q. If a particle starts at time r from a point x and if $(r, x) \in Q$, then the probability distribution of the exit point, given by the formula

$$k(r, x; B) = \Pi_{r,x}(\tau, \xi_\tau) \in B,$$

is concentrated on the boundary ? of Q. Moreover, it is concentrated on the set ? of regular points [a point $(s, c) \in$ is called regular if, for $s' > s, \Pi_{r,x}\{(t, \xi_t) \in Q$ for all $s < t < s'\} = 0$] If $(r, x) \notin Q$, then $k(r, x; \bullet)$ is concentrated at (r, x). For every bounded continuous function f,

$$u(r, x) = \Pi_{r,x} f(\tau, \xi_\tau) = \int k(r, x; ds, dy) f(s, y)$$

is a solution of the boundary value problem

$$\begin{aligned} \partial u/\partial r + Lu &= 0 \qquad \text{in } Q, \\ u &= f \qquad \text{on } \partial_{reg} Q. \end{aligned} \tag{2.1}$$

the family of random points $((\tau, \xi_\tau), \pi_{r,x})$ has the following strong Markov property: for every pre-τ $X \geq 0$ and every post-τ $Y \geq 0$,

$$\Pi_{r,x}(XY) = \Pi_{r,x}(X \Pi_{\tau, \xi_\tau} Y). \tag{2.2}$$

Pre-τ means depending only on the part of the path before τ. Similarly, post-τ means depending on the path after τ. To every measurable $\rho \geq 0$, there correspond a pre-τ random variable

$$X = \int_{-\infty}^{\tau} \rho(s, \xi_s) ds$$

and a post-τ random variable

$$Y = \int_{\tau}^{\infty} \rho(s, \xi_s) ds$$

Let τ and τ' be the first exit times from Q and Q'. Then $f(\tau', \xi_{\tau'})$ is a pre-τ random variable if $Q' \subset Q$ and it is a post-τ random variable if $Q \subset Q'$.

3. Exit systems associated with branching diffusion

Consider a system of particles moving in R^d according to the following rules:

(i) Each particle performs an L-diffusion.

(ii) It dies during time interval $(t, t+h)$ with probablility $kh + o(h)$, independently on its age.

(iii) If a particle dies at time t at point x, then it produces n new particles with probability $p_n(t, x)$.

(iv) The only interaction between the particles is that the birth time and place of offspring coincide with the death time and place of their parent.

(Assumption (ii) implies that the life time of every particle has an exponential probability distribution with the mean value $1/k$.)

We denote by $P_{r,x}$ the probability law corresponding to a process started at time t by a single particle located at point x. Suppose that particles stop to move and to procreate outside an open subset Q of S. In other words, we observe each particle at the first, in the family history (by the family history we mean the path of a particle and all its ancestors. If the family history starts at (r, x), then the probability law of this path is $\Pi_{r,x}$), exit time from Q. The exit measure from Q is defined by the formula

$$X_Q = \delta_{(t_1,y_1)} + \cdots + \delta_{(t_n,y_n)}$$

where $(t_1, y_1), \cdots, (t_n, y_n)$ are the states of frozen particles and $\delta_{(t,y)}$ means the unit measure concentrated at (t, y). We also consider a process started by a finite or infinite sequence of particles that "immigrate" at times r_i at points x_i. There is no interaction between their posterities and therefore the corresponding probability law is the convolution of P_{r_i,x_i}. We denote it P_μ where

$$\mu = \sum \delta_{(r_i,x_i)}$$

is a measure on S describing the immigration. We arrive at a family X of random measures $(X_Q, P_\mu), Q \in O, \mu \in M$ where O is a class of open subsets of S and M is the class of all integer-valued measures on S. Family X is a special case of a branching exit Markov system. A general definition of such systems is given in the next section.

4. Branching exit Markov systems

A random measure on a measurable space $(s, \mathcal{B}_S)$ is a pair (X, P) where $X(\omega, B)$ is a kernel (a kernel from a measurable space $(E_1, \mathcal{B}_1)$ to a measurable space $(E_2, \mathcal{B}_2)$ is a function $K(x, B)$ such that $K(x, \bullet)$ is a measure on $\mathcal{B}_2$ for every $x \in E_1$ and $K(x, B)$ is a $\mathcal{B}_1-$measurable function for every $B \in \mathcal{B}_2$) from an auxiliary measurable space $(\Omega, \mathcal{F})$ to $(S, \mathcal{B}_S)$. We assume that S is a Borel subset of a compact metric space and $\mathcal{B}_S$ is the class of all Borel subsets of S.

Suppose that:

(i) O is a subset of $\sigma-$algebra $\mathcal{B}_S$,

(ii) M is a class of measures on $(S, \mathcal{B}_S)$ which contains all measures $\delta_y,\ y \in S$,

(iii) to every $Q \in O$ and every $\mu \in M$, there corresponds a random measure (X_Q, P_μ) on $(S, \mathcal{B}_S)$.

Condition (ii) is satisfied, for instance, for the class $\mathcal{M}(S)$ of all finite measures and for the class $\mathcal{N}(S)$ of all integer-valued measures.

We use notation $< f, \mu >$ for the integral of f with respect to a measure μ. Denote by Z the class of functions

$$Z = \exp\left\{\sum_1^n < f_i, X_{Q_i} >\right\} \tag{4.1}$$

where $Q_i \in O$ and f_i are positive measurable functions on S. We say that $X = (X_Q, P_\mu), Q \in O, \mu \in M$ is a branching system if:

4.A For every $Z \in Z$ and every $\mu \in M$,

$$P_\mu Z = e^{-<u,\mu>} \tag{4.2}$$

where

$$u(y) = -log P_y Z \tag{4.3}$$

and $P_y = P_{\delta_y}$.

Condition 4.A (we call it the continuous branching property) implies that

$$P_\mu Z = \Pi P_{\mu_n} Z$$

for all $Z \in Z$ if $\mu_n, n = 1, 2, \cdots$ and $\mu = \sum \mu_n$ belong to M.

A family X is called an exit system if:

4.B For all $\mu \in M$ and $Q \in O$,

$$P_\mu\{X_Q(Q) = 0\} = 1$$

4.C If $\mu \in M$ and $\mu(Q) = 0$, then

$$P_\mu\{X_Q = \mu\} = 1.$$

Finally, we say that X is a branching exit Markov $[BEM]$ system, if $X_Q \in M$ for all $Q \in O$ and if, in addition to 4.A–4.C, we have:

4.D Markov property Suppose that $X \geq 0$ is measurable with respect to the $\sigma-$algebra $\mathcal{F}_{\subset Q}$ generated by $X_{Q'}, Q' \subset Q$ and $Y \geq 0$ is measurable with respect to the $\sigma-$algebra $\mathcal{F}_{\supset}$ generated by $X_{Q'}, Q \subset Q'$. Then

$$P_\mu(XY) = P_\mu(XP_{X_Q}Y). \tag{4.4}$$

It follows from the principles (i)–(iv) stated at the beginning of Section 3 that conditions 4.A–4.D hold for the systems of random measures associated with branching diffusions. For them $S = R \times R^d, M = \mathcal{N}(S)$ and O is a class of open subsets of S. In future, we deal with special classes $O_0 \subset O_1$: an open set Q belongs to O_0 if $Q \subset S_\Delta$ for a finite interval Δ, and it belongs to O_1 if $Q \subset S_{>t_0}$ for some $t_0 \in R$. (We put $S_A = A \times R^d$ for every $A \subset R$.)

5. Transition operator

Let $X = (X_Q, P_\mu)$, $Q \in O$, $\mu \in M$ be a family of random measures. Denote by B the set of all bounded positive $\mathcal{B}_S-$measurable functions. The transition operator of X is defined by the formula

$$V_Q(f)(y) = -\log P_y e^{-<f, X_Q>} \text{ for } f \in B \tag{5.1}$$

Note that

$$V_Q(0) = 0 \text{ for all } Q \tag{5.2}$$

Recall that the Laplace transform

$$\phi(\lambda) = Pe^{-\lambda Z}, \ \lambda \geq 0$$

determines uniquely the probability distribution of a positive random variable Z relative to P. Therefore the transition operator (5.1) defines uniquely the probability distribution of X_Q relative to P_y. If X is a branching system, then, for all $\mu \in M$,

$$P_\mu e^{-<f,X_Q>} = e^{-<V_Q(f),\mu>} \tag{5.3}$$

and therefore V_Q determines the probability distribution of X_Q relative to P_μ.

Theorem 5.1

(i) A branching system X is a branching exit system if and only if:

a)

$$V_Q(f) = V_Q(\bar{f}) \text{ if } f = \bar{f} \text{ on } Q^c.$$

b) For every $Q \in O$,

$$V_Q(f) = f \text{ on } Q^c.$$

(ii) A branching exit system is a BEM system if and only if:
For all $Q \subset \tilde{Q}$,

$$V_Q V_{\tilde{Q}} = V_{\tilde{Q}}.$$

If X is a BEM system associated with a branching diffusion, then

$$V_Q(f) = -\log w$$

where

$$w(r,x) = P_{r,x} e^{-<f,X_Q>}. \tag{5.4}$$

Consider the offspring generating function

$$\phi(r,x;z) = \sum_0^\infty p_n(r,x) z^n.$$

The four principles stated at the beginning of Section 3 lead to an equation

$$w(r,x) = \Pi_{r,x}\left[e^{-k(\tau-r)}e^{-f(\tau,\xi_\tau)} + k\int_r^\tau e^{-k(s-r)} ds \phi(s,\xi_s; w(s,\xi_s))\right] \tag{5.5}$$

where τ is the first exit time from Q. The first term in the brackets corresponds to the case when the particle started the process is still alive at time τ, and the second term corresponds to the case when it dies at time $s \in (r,\tau)$.

Formula (5.5) implies that $v = V_Q(f)$ satisfies the equation

$$e^{-v(r,x)} = \Pi_{r,x}\left[k\int_r^\tau \Phi\left(s,\xi_s; e^{-v(s,\xi_s)}\right) + e^{-f(\tau,\xi_\tau)}\right] \tag{5.6}$$

where $\Phi(r,x;z) = \phi(r,x;z) - z$.

6. β–transforms of BEM systems and their limits

To every BEM system $X = (X_Q, P_\mu)$, $Q \in O$, $\mu \in M$ and to every constant $\beta > 0$, there correspond a system $X^\beta = (X_Q^\beta, P_\mu^\beta)$, $Q \in O$, $\mu \in M^\beta$ where

$$M^\beta = \beta M, \; X_Q^\beta = \beta X_Q, \; P_\mu^\beta = P_{\frac{\mu}{\beta}}.$$

We call it the β–transform of X.

Put

$$V_Q^\beta(f)(y) = \beta^{-1} V_Q(\beta f)(y) \tag{6.1}$$

where $V_Q(f)$ is given by (5.1). Note that

$$P_\mu e^{-<f, X_Q^\beta>} = e^{-<V_Q^\beta(f), \mu>} \tag{6.2}$$

We construct a BEM system $X^0 = (X_Q^0, P_\mu^0)$ which is, in a certain sense, the limit of X^β as $\beta \to 0$. (If X is a BEM system associated with a branching diffusion, then X^β describes the evolution of the mass distribution assuming that all particles have mass β. The limit as $\beta \to 0$ reveals the behavior of a system of very small particles with very short lives.)

We put $||f|| = \sup_S |f(y)|$ and we denote by B_c the set of all positive $\mathcal{B}_S$-measurable functions f such that $||f|| \leq c$.

Theorem 6.1 *Suppose that operators V_Q^β defined by (6.1) satisfy, for every $Q \in O$, the conditions:*

(i) $V_Q^\beta(f)$ converge, as $\beta \to 0$, to a limit $V_Q^0(f)$ and the convergence is uniform on every set B_c, that is

$$\epsilon(c, \beta) = \sup_{B_c} ||V_Q^\beta(f) - V_Q^0(f)|| \to 0 \qquad \text{as } \beta \to 0 \tag{6.3}$$

(ii) $V_Q^0(f)$ satisfies the Lipschitz condition on every B_c, i.e., for every c, there exists a constant $a(c)$ such that

$$||V_Q^0(f) - V_Q^0(g)|| \leq a(c) ||f - g|| \qquad \text{for all } f, g \in B_c \tag{6.4}$$

Then there exists a BEM system $X^0 = (X_Q^0, P_\mu^0)$, $Q \in O$, $\mu \in \mathcal{M}(S)$ such that

$$P_\mu e^{-<f, X_Q^0>} = e^{-<V_Q^0(f), \mu>} \tag{6.5}$$

for all $\mu \in \mathcal{M}(S)$, $Q \in O$, $f \in B$.

7. Superdiffusion

We apply Theorem 6.1 to a BEM system $X = (X_Q, P_\mu)$, $Q \in O$, $\mu \in \mathcal{N}(S)$ associated with a branching diffusion.

It follows from (5.6) that $v^\beta = V_Q^\beta(f)$ satisfies the equation

$$e^{-\beta v^\beta(r,x)} = \Pi_{r,x}\left[\int_r^\tau k\Phi\left(s,\xi_s;e^{-\beta v^\beta(s,\xi_s)}\right)\,ds + e^{-\beta f(\tau,\xi_\tau)}\right] \tag{7.1}$$

which is equivalent to the equation

$$u^\beta(r,x) + \Pi_{r,x}\int_r^\tau \psi^\beta\left(s,\xi_s;u^\beta(s,\xi_s)\right)\,ds = \Pi_{r,x}f^\beta(\tau,\xi_\tau) \tag{7.2}$$

where

$$\begin{aligned} u^\beta &= \left[1 - e^{\beta v^\beta}\right]/\beta, \\ f^\beta &= \left[1 - e^{-\beta f}\right]/\beta, \\ \psi^\beta(r,x;u) &= \left[\phi^\beta(r,x;1-\beta u) - 1 + \beta u\right]k^\beta/\beta \end{aligned} \tag{7.3}$$

(We assume that k and ϕ depend on β.)

Note that, as $\beta \to 0$, $F^\beta \to f$. If $\psi^\beta \to \psi$, then we expect that $u^\beta \to u$ where u is a solution of the equation

$$u(r,x) + \Pi_{r,x}\int_r^\tau \psi\left(s,\xi_s;u(s,\xi_s)\right)\,ds = \Pi_{r,x}f(\tau,\xi_\tau). \tag{7.4}$$

We say that a BEM system $X = (X_Q, P_\mu)$, $Q \in O$, $\mu \in M$ is an (L,ψ)–superdiffusion if O is a class of open subsets of $S = R \times R^d$ and if transition operators V_Q satisfy the condition: for every $f \in B$, $u = V_Q(f)$ is a solution of the equation (7.4).

Equations (7.2) and (7.4) can be rewritten in the form

$$u^\beta + G_Q\Psi^\beta(u^\beta) = K_Q f^\beta \tag{7.5}$$

and

$$u + G_Q\Psi(u) = K_Q(f). \tag{7.6}$$

where the Poisson operator K_Q and the Green operator G_Q are defined by the formulae

$$K_Q f(r,x) = \Pi_{r,x} f(\tau,\zeta_\tau), \tag{7.7}$$

$$G_Q\rho(r,x) = \Pi_{r,x}\int_r^\tau \rho(s,\xi_s)\,ds \tag{7.8}$$

and

$$\begin{aligned} W^\beta(f)(r,x) &= \left[1 - e^{-\beta f(r,x)}\right]/\beta \\ \Psi^\beta(f)(r,x) &= \psi^\beta\left(r,x;f(r,x)\right), \\ \Psi(f)(r,x) &= \psi\left(r,x;f(r,x)\right). \end{aligned} \tag{7.9}$$

We prove:

Theorem 7.1 *Suppose that:*

(i) $\psi^\beta \geq 0$,

(ii) ψ^β *converges to* ψ *uniformly on every set* $\{(r,x) \in S, u \in [0,c]\}$, *and*

(iii) for every c, *there exists a constant* $q(c)$ *such that*

$$|\psi(r,x;u_1) - \psi(r,x;u_2)| \leq q(c)|u_1 - u_2|$$

for all $(r,x) \in S$ *and all* $u_1, u_2 \in [0,c]$.

Then operators V_Q^β *satisfy* Theorem 6.1(i) *and* (ii), *and* V_Q^0 *are transition operators of an* $(L,\psi)-$ *superdiffusion* X. *Finite-dimensional distributions of* X *are uniquely defined by* ψ.

By Theorem 7.1, an $(L,\psi)-$superdiffusion ψ exists (with $O = O_0, M = \mathcal{M}(S)$) if ψ satisfies condition (iii) and if there exist generating functions ϕ^β and constants k^β for which

$$\psi^\beta(r,x;u) = \left[\phi^\beta(r,x;1-\beta u) - 1 + \beta u\right] k^\beta/\beta \tag{7.10}$$

satisfy Theorem 7.1(i) and (ii). By constructing appropriate ϕ^β and k^β, we prove the following theorem:

Theorem 7.2 *An* $(L,\psi)-$*superdiffusion exists for every function*

$$\psi(r,x;u) = b(r,x)u^2 + \int_0^\infty \left(e^{-\lambda\mu} - 1 + \lambda\mu\right) n(r,x;\, du) \tag{7.11}$$

where b *is a positive function and* n *is a kernel from* $(S, \mathcal{B}_S)$ *to* R_+ *subject to the conditions*

$$b, \qquad \int_0^1 u^2 n(r,x;\, du) \quad \text{and} \quad \int_1^\infty u\, n(r,x;\, du) \tag{7.12}$$

are bounded on $\Delta \times R^d$ *for every finite interval* Δ.

We also prove that superdiffusions can be defined for wider classes O_1 (defined in Section 4) and M_1 which consists of all measures μ on S subject to the condition: $\mu(S_\Delta) < \infty$ for every finite interval Δ.

8. Applications to PDEs

Suppose that X is the (L,ψ)–superdiffusion described in Theorem 7.2. Then, for every $Q \in O_1$ and every $f \in B$,

$$u(r,x) = V_Q(f)(r,x) = -\log P_{r,x} e^{-<f,X_Q>} \tag{8.1}$$

is a solution of the integral equation (7.6). If f is bounded and continuous, then (7.6) implies

$$\begin{aligned} \partial u/\partial r + Lu &= \psi(u) \text{ in } Q, \\ u &= f \text{ on } \partial_{reg} Q \end{aligned} \tag{8.2}$$

See [2] or [4]. This is basis of a probabilistic theory of semi-linear parabolic and elliptic equations involving operator $Lu - \psi(u)$

For instance, the first boundary value problem

$$\begin{aligned} Lu - \psi(u) &= 0 \text{ in } D, \\ u &= f \text{ on } \partial D \end{aligned} \tag{8.3}$$

in a bounded domain DR^d with a smooth boundary can be solved by the formula

$$u(x) = -\log P_x e^{-<f,X_D>} \tag{8.4}$$

where $P_x = P_{0,x}$ and X_D is the exit measure from cylinder $Q = (0,\infty) \times D$.

The next step is a description of all positive solutions of equation $Lu = \psi(u)$ in an arbitrary domain D. It was shown in [3] that all such solutions can be obtained by the formula

$$u(x) = -\log P_x e^{-Z} \tag{8.5}$$

where

$$Z = \lim < u, X_{D_n} > . \tag{8.6}$$

Here D_n is a sequence of bounded smooth domains such that $\bar{D_n} \subset D_{n+1}$ and the union of D_n is equal to D. Formulae (8.5)–(8.6) establish a $1-1$ correspondence between the set $\mathcal{U}$ of all positive solutions in D and a closed convex cone Z of functionals of X. In particular,

$$Z = \begin{cases} \infty, & \text{if } X \text{ hits } \partial D \\ 0, & \text{otherwise} \end{cases}$$

corresponds to the maximal solution.

A more explicit description of $\mathcal{U}$ is based on the concept of a trace of u on ∂D (for a general domain, ∂D means its Martin boundary). This direction of research is a subject of Kuznetsov's article in the present volume.

Another important direction is the study of subsets of $D \cup \partial D$ which are removable singularly for $u \in \mathcal{U}$. It turns out that a set Γ belongs to this class if and only if it is not hit by the superdiffusion. An analytic characterization of removable singularities is done in terms of capacities. Recent results in this direction and references to earlier work can be found in [5, 6, 7, 8].

References

[1] E.B. Dynkin, A probabilistic approach to one class of nonlinear differential equations, Probab.Th.Rel.Fields 89 (1991), 89–115.

[2] _, Superdiffusions and parabolic nonlinear differential equations, Ann. Probab. 20 (1992), 942–962.

[3] _, Superprocesses and partial differential equations, Ann.Probab. 21 (1993), 1185–1262.

[4] _, Semilinear parabolic equations, diffusions and superdiffusions, J. Functional Analysis 158 (1998), 325–356.

[5] E. B. Dynkin and S. E. Kuznetsov, Fine topology and fine trace on the boundary associated with a class of quasilinear differential equations, Comm. Pure Appl. Math. 51 (1998), 897–936.

[6] S. E. Kuznetsov, Removable singularities for $Lu = \psi(u)$ and Orlicz capacities, J. Funct. Analysis, to appear.

[7] _, Polar boundary sets for superdiffusions and removable lateral singularities for nonlinear parabolic PDEs, Communications Pure Appl. Math. 51 (1998), 303–340.

[8] M. Marcus and L. Veron, The boundary trace of positive solutions of semilinear elliptic equations: The supercritical case, J. Math. Pure Appl. 77 (1998), 481–524.

[9] S. Watanabe, A limit theorem on branching processes and continuous state branching processes, J. Math. Kyoto Univ. 8 (1968), 141–167.

Chapter 2

FELLER TRANSITION FUNCTIONS, RESOLVENT DECOMPOSITION THEOREMS, AND THEIR APPLICATION IN UNSTABLE DENUMERABLE MARKOV PROCESSES

Anyue Chen
School of Computing and Mathematical Sciences,
University of Greenwich, London, UK

Hanjun Zhang
Research Department, Changsha Railway University,
Changsha 410075, China

Zhenting Hou
Research Department, Changsha Railway University,
Changsha 410075, China

Abstract This paper surveys the recent progresses made in the field of unstable denumerable Markov processes. Emphases are laid upon methodology and applications. The important tools of Feller transition functions and Resolvent Decomposition Theorems are highlighted. Their applications particularly in unstable denumerable Markov processes with a single instantaneous state and Markov branching processes are illustrated.

1. Introduction

Around 50 years ago Kolmogorov [29] raised the following challenging question.

Z. Hou et al. (eds.), Markov Processes and Controlled Markov Chains, 15–39.

Question 1.1 (Kolmogorov [29]) *Given a matrix $Q = \{q_{ij}\}$ on the non-negative integer $Z_+ = \{0, 1, 2, \cdots\}$ with the off-diagonal elements (here only the non-zero elements are specified)*

$$q_{0j} = 1 \qquad (\forall j \geq 1) \tag{1.1}$$

and

$$0 < q_{i0} \hat{=} a_i < +\infty \qquad (\forall i \geq 1) \tag{1.2}$$

together with the diagonal elements

$$q_i \hat{=} - q_{ii} = a_i \qquad (\forall i \geq 1) \tag{1.3}$$

then under what conditions does there exist an honest continuous time Markov Chain (CTMC) whose transition function P(t) satisfies $P'(0) = Q$?

Why is this question challenging? One of the reasons is, if there exists an honest transition function P(t) such that $P'(0) = Q$, then condition (1.1) forces

$$q_0 \hat{=} - q_{00} = +\infty. \tag{1.4}$$

That is, the state $\{0\}$ is an instantaneous state and thus the CTMC is an unstable one. Although fruitful results have been obtained for CTMC, they are almost all concerned with stable CTMC. Indeed, few results have been obtained for unstable CTMC even until now. See later.

Kendall and Reuter [27] provided the following answer to Question 1.1.

Theorem 1.1 (Kendall and Reuter [27]) *If $\sum_{i=1}^{\infty}(1/a_i) < +\infty$ then there exists an honest transition function $P(t)$ such that $P'(0) = Q$.*

It should be noted that (1.2) and (1.3) are assumed to be true in Kolmogorov's original question. That is, all positive states are assumed to be stable and conservative. Hence another interesting and challenging question naturally arises.

Question 1.2 *If there exists an honest function $P(t)$ whose (infinitesimal) intensity matrix $Q = \{q_{ij}\}$ satisfies* (1.1), *will all the positive states be stable and conservative?*

One has to wait for many years before Williams [46] answered 'yes' strongly in obtaining the following remarkable result.

Theorem 1.2 (Williams [46]) *Suppose a matrix $Q = \{q_{ij}\}$ defined on a countable set E satisfies the condition that there exists a state $b \in E$ such that*

$$\liminf_{i\to\infty} q_{bi} > 0. \tag{1.5}$$

If there exists an honest transition function $P(t) = \{p_{ij}(t); i, j, \in E\}$ such that

$$\lim_{t\to 0}(p_{bi}(t)/t) = q_{bi} \qquad (i \neq b) \tag{1.6}$$

then we have

$$\begin{aligned} q_b &\hat{=} -q_{bb} = +\infty && (1.7)\\ q_i &\hat{=} -q_{ii} < +\infty \qquad (\forall i \neq b) && (1.8)\\ q_i &= \sum_{j\neq i} q_{ij} \qquad (\forall i \neq b) && (1.9)\\ \sum_{j\neq b}\sum_{k\neq b} q_{bk}\phi_{kj}(\lambda) &< +\infty \qquad (\forall \lambda > 0) && (1.10) \end{aligned}$$

where $\Phi(\lambda) = \{\phi_{ij}(\lambda); i, j \neq b, \lambda > 0\}$ is the minimal Q_b-resolvent and Q_b is the restriction of Q on $E\backslash\{b\}$. Conversely, suppose a matrix $Q = \{q_{ij}\}$ defined on E satisfies (1.5) *and* (1.7)–(1.9), *and if* (1.10) *holds true then there exists an honest transition function $P(t) = \{p_{ij}(t)\}$ such that*

$$\lim_{t\to 0}(P(t) - I)/t = Q \tag{1.11}$$

Note that condition (1.5) means that there exists a finite subset $F \subset E$ and a positive number $\delta > 0$ such that for all $i \in E\backslash F$,

$$q_{bi} \geq \delta > 0 \tag{1.12}$$

which forces (1.7) to hold true under the condition that Q is an infinitesimal q-matrix. The interesting thing here is that (1.5) (or (1.12)) also forces (1.8) and (1.9) to be true, that is, all states except $\{b\}$ must be stable and conservative. Moveover, if (1.5) and (1.7)–(1.9) hold true, then Q becomes an infinitesimal q-matrix of some honest transition function if and only if (1.10) holds true.

Note also that if the requirement of an honest transition function is relaxed to be a transition function (not necessarily honest), then (1.8) must still hold true (i.e. all states except $\{b\}$ are stable), though (1.9) (conservativeness) may not be necessarily true.

Obviously, (1.1) is a special case of (1.5). From now on, a matrix $Q = \{q_{ij}\}$ satisfying (1.5) and (1.7) to (1.9) will be called a Kolmogorov–Williams q-matrix, or simply, K-W q-matrix. We use the same name for the corresponding q-processes.

William's proof for the existence of the K-W q-processes was a probabilistic one. The advantage is that the initiative meaning is clear. However it seems not easy to give more results from this proof. Therefore an analytic proof is hoped and more results are expected for such processes. We shall return to these questions later.

Reuter [38] once considered an example which is slightly more general than the Kolmogorov's q-matrix in replacing $q_{0j} = 1, (\forall j \geq 1)$ by $\sum_{j=1}^{\infty} q_{0j} = +\infty$.

Theorem 1.3 (Reuter [38]) *Suppose a matrix $Q = \{q_{ij}\}$ on the non-negative integer Z_+ is given by* (1.2)–(1.4) *(here, again, only the non-zero elements are specified) together with*

$$\sum_{j=1}^{\infty} b_j = +\infty \tag{1.13}$$

where $b_j = q_{0j}$. Then if

$$\sum_{j=1}^{\infty} (b_j/a_j) < +\infty \tag{1.14}$$

then there exists an honest transition function such that (1.11) *holds true. That is there exists an honest Q-process.*

Notice that the common feature in all the above examples is that there is one and only one unstable (or instantaneous) state. Before proceeding further, we first give the precise meaning of this term. Recall that a matrix $Q = \{q_{ij}\}$ defined on a countable set E is called a pre-q-matrix if the following D-K conditions are satisfied

$$0 \leq q_{ij} < +\infty \qquad (i \neq j; i,j \in E) \tag{1.15}$$
$$-\infty \leq q_{ii} \leq 0 \qquad (i \in E) \tag{1.16}$$

and

$$\sum_{j \neq i} q_{ij} \leq -q_{ii} \hat{=} q_i \qquad (i \in E). \tag{1.17}$$

If $q_i < +\infty$, then $i \in E$ is called stable while if $q_i = +\infty$, $i \in E$ is called instantaneous. If all $i \in E$ are stable then Q is called totally stable (TS). The meaning of totally instantaneous (TI) should be then clear. In the case of the existence of both stable and instantaneous states, Q is called a mixing pre-q-matrix. Both TI and mixing cases are called unstable.

Furthermore, a pre-q-matrix Q is called Conservative Uni-Instantaneous (CUI) if there exists a state $b \in E$ such that

$$\sum_{j \neq b} q_{bj} = q_b = +\infty \tag{1.18}$$

and that

$$\sum_{j \neq i} q_{ij} = q_i < +\infty \qquad (\forall i \neq b) \tag{1.19}$$

Note that all the above examples considered until now are special cases of CUI pre-q-matrix.

Also recall that a matrix $Q = \{q_{ij}\}$ defined on E is called a q-matrix if there exists a Markov transition function $P(t)$ such that (1.11) is satisfied. We shall apply all the above terms to q-matrix as well as the corresponding q-process (q-function, q-resolvent etc).

It is well-known that a q-matrix is a pre-q-matrix. However the converse may not be always true. Now several basic questions arise.

Question 1.3 (Existence) *Under what conditions does a given pre-q-matrix become a q-matrix?*

Question 1.4 (Uniqueness) *If a given Q is a q-matrix, under what conditions does there exist only one corresponding Q-process?*

Question 1.5 (Construction) *How do we construct all the Q-processes via a given q-matrix Q?*

Question 1.6 (Property) *How do we study all kinds of properties of Q-process in terms of the given q-matrix Q?*

No doubt, Question 1.6 is the most important question which has considerable significance both in theory and applications. However, Questions 1.3–1.5 are also of great importance since without solving them it is of little hope we could tackle Question 1.6 successfully.

If Q is totally stable, then the above questions were firstly systematically studied by J.L. Doob and W. Feller in the 1940s and then continuously investigated by many world-leading probabilists, including D.G. Kendall, G.E.H. Reuter, D. Williams, J.F.C. Kingman, Samuel Karlin and K.L. Chung. In particular, Feller [18] showed that a totally stable pre-q-matrix must be a q-matrix and constructed a solution for any totally stable q-matrix, which has a minimal property and bears his name today. Thus the existence Question 1.3 was solved completely. Doob [17] observed and investigated the non-uniqueness property of totally stable q-processes and then the uniqueness Question 1.4 was solved by Reuter [36], [37] for the conservative case and Hou [22] for the non-conservative case respectively. The construction Question 1.5 is closely related to the boundary theory (Feller boundary and Matrix boundary) of continuous time Markov chains, to which K.L. Chung and D. Williams contributed significantly. As to Question 1.6, fruitful results have been obtained and there are plenty of monographs and books to discuss totally stable q-processes. Until the publishing of Chung's foundation book [15], the theory of totally stable q-processes were viewed, by and large, as completed, though many other important topics such as reversibility, strong and exponential ergodicity, quasi-stationary distributions, monotonicity, duality, coupling, large deviation, and spectral theory have emerged and flourished since then and lasted even until today.

Now, how about the unstable case? It may be hard to believe that the above Theorems 1.1, 1.2 and 1.3 are essentially the only results obtained for the mixing case until the early 1980's. The picture of the totally instantaneous case is no better. Surprisingly, however, an elegant result was obtained by Williams [47] regarding the existence problem for TI q-processes. That is

Theorem 1.4 (Williams [47]) *Suppose $Q = \{q_{ij}\}$ is a totally instantaneous pre-q-matrix, then it becomes a q-matrix if and only if the following two conditions hold true.*

(i) $\sum_{j \neq \{a,b\}} (q_{aj} \wedge q_{bj}) < +\infty \qquad (\forall a \neq b, a, b \in E)$;

(ii) there exists an infinite subset I of E such that for all $i \in E$

$$\sum_{j \in I} q_{ij} < +\infty$$

See also Rogers and Williams [41]. Analysis of some examples of the totally instantaneous case can also be seen in Blackwell [4] and Kendall [26].

Of course, there exist a few books discussing the general theory concerning instantaneous states. The path behaviour of CTMC with instantaneous states is discussed in Chung [15]. Using the method of taboo probability to study properties of CTMC can also be found in Chung [15]. Another very important book containing discussion of instantaneous states is Rogers and Williams [42]. The monograph written by Freedman [19] is, perhaps, the only book to discuss instantaneous states exclusively.

Notwithstanding, the picture of the unstable case is still very poor, in particular, when compared with the stable case. This reflects the fact that the topic of unstable q-processes has essentially mathematical difficulty. This does not mean of course, that to study unstable q-processes has little significance in practical applications. On the contrary, to study unstable q-processes is of considerable significance both in theory and applications. Therefore, the right thing we should do is to find out some methods and techniques to overcome the mathematical difficulty and this will surely yield considerable progress in this challenging topic.

2. Feller transition function and resolvent decomposition theorem

Although there are few results obtained for unstable q-processes until early 1980's as we mentioned in the previous section, an interesting and closely related theory, called Feller transition function, has been already developed long before.

Note first that a countable set E with discrete topology is trivially an LCCB. Thus we may define the Markov semigroup $P = (P(t); t \geq 0)$, induced by a standard substochastic transition function $(p_{ij}(t); i, j \in E)$, as a Feller semigroup if $P(t)x \in C_0(E)$ whenever $x \in C_0(E)$ where $C_0(E)$ denote the Banach space of continuous functions on E vanished at infinity. The corresponding transition function is also called Feller.

Although this Feller property, one of the many kinds of Feller properties is well-known in the general theory of Markov processes, it is the only Feller property which may yield more interesting results for the countable state space. Indeed the more commonly used Feller property for general Markov semigroup, i.e., mapping $C(E)$ to $C(E)$, only yields unsatisfactory theory in the countable state space case.

The following interesting and important result was first announced (without proof) by Jurkat [25] and then proved and developed by Reuter and Riley [39].

Theorem 2.1 *The following statements are equivalent:*

(i) $P(t)$ *is Feller, i.e.,* $P(t)x \in C_0(E)$ *whenever* $x \in C_0(E)$,

(ii) $p_{ij}(t) \to 0$ *as* $i \to \infty$ *for all* $j \in E$ *and all* $t \geq 0$,

(iii) $r_{ij}(\lambda) \to 0$ *as* $i \to \infty$ *for all* $j \in E$ *and all* $\lambda > 0$

where $\{p_{ij}(t)\}$ *and* $\{r_{ij}(\lambda)\}$ *are transition and resolvent functions respectively. Moreover, if* $P(t)$ *is Feller, then its q-matrix* Q *must be totally stable and this* $P(t)$ *is actually the Feller minimal* Q*-function.*

The relationship between this remarkable result and Theorem 1.2 (and then Theorems 1.1 and 1.3) seems transparent. However, this result did not immediately lead to the research in the direction of unstable denumerable Markov processes. It reflects the fact that a gap exists between these two topics. The following decomposition theorem fills this gap and bridges the way to study the latter.

Theorem 2.2 *Suppose $Q = \{q_{ij}\}$ is a q-matrix on the state space E (that is there exists a transition function $P(t)$ such that $P'(0) = Q$). Suppose further that $R(\lambda) = \{r_{ij}(\lambda); i,j \in E, \lambda > 0\}$ is a Q-resolvent. Let F be a finite subset of E and denote $G = E\backslash F$. Then $R(\lambda)$ may be uniquely decomposed as follows*

$$R(\lambda) = \begin{pmatrix} 0 & 0 \\ 0 & \Psi(\lambda) \end{pmatrix} + \begin{pmatrix} A(\lambda) & A(\lambda)\eta(\lambda) \\ \xi(\lambda)A(\lambda) & \xi(\lambda)A(\lambda)\eta(\lambda) \end{pmatrix} \tag{2.1}$$

where

(i) $A(\lambda)$ is the restriction of $R(\lambda)$ on $F \times F$, i.e., $A(\lambda) = \{r_{ij}(\lambda); i,j \in F\}$ and

$$|A(\lambda)| > 0 \qquad (\forall \lambda > 0) \tag{2.2}$$

and thus $A(\lambda)$ is invertible for all $\lambda > 0$.

(ii) $\Psi(\lambda) = \{\psi_{ij}(\lambda); i,j \in G\}$ is a Q_G-resolvent where $Q_G = (q_{ij}; i,j \in G)$ is the restriction of Q on $G \times G$.

(iii) $\eta(\lambda) = \{\eta_{ij}(\lambda); i \in F, j \in G\}$ satisfies

$$\eta(\lambda) - \eta(\mu) = (\mu - \lambda)\eta(\lambda)\Psi(\mu), \qquad (\forall \lambda, \mu > 0) \tag{2.3}$$

and

$$0 \le \eta_{i\cdot}(\lambda) \in l_1 \qquad (\forall i \in F, \forall \lambda > 0) \tag{2.4}$$

(iv) $\xi(\lambda) = \{\xi_{ij}(\lambda); i \in G, j \in F\}$ satisfies

$$\xi(\lambda) - \xi(\mu) = (\mu - \lambda)\Psi(\lambda)\xi(\mu), \qquad (\forall \lambda, \mu > 0) \tag{2.5}$$

and

$$0 \le \xi(\lambda)\mathbf{1} \le \mathbf{1} - \lambda\psi(\lambda)\mathbf{1} \tag{2.6}$$

*(Here **1** is a column vector whose elements are all 1 and the dimension of which depends. For example, the first **1** in (2.6) is a finite dimensional vector on F while the other two are infinite dimensional vectors on G)*

(v)

$$\lim_{\lambda\to\infty} \lambda\eta(\lambda) = Q_{FG} \tag{2.7}$$

$$\lim_{\lambda\to\infty} \lambda\xi(\lambda) = Q_{GF} \tag{2.8}$$

where $Q_{FG} = (q_{ij}, i \in F, j \in G)$ *and* $Q_{GF} = (q_{ij}, i \in G, j \in F)$ *are the restriction of* Q *on* $F \times G$ *and* $G \times F$ *respectively.*

(vi)

$$\lim_{\lambda\to\infty} \lambda \sum_{j\in G} \eta_{ij}(\lambda)(1-\xi_{ji}) < +\infty \qquad (\forall i \in F) \tag{2.9}$$

(vii) There exists a constant matrix $C = \{c_{ij}; i,j \in F\}$ *such that*

$$A^{-1}(\lambda) = C + \lambda I + \lambda[\eta(\lambda), \xi] \tag{2.10}$$

and (thus the right hand side of (2.10) *is invertible)*

$$-c_{ij} = q_{ij} + \lim_{\lambda\to\infty} \lambda \sum_{k\in G} \eta_{ik}(\lambda)\xi_{kj} \qquad (\forall i,j \in F, i \neq j) \tag{2.11}$$

$$\sum_{j\in F} c_{ij} \geq \lim_{\lambda\to\infty} \lambda \sum_{k\in G} \eta_{ik}(\lambda)(1 - \sum_{j\in F} \xi_{kj}) \qquad (\forall i \in F) \tag{2.12}$$

(viii) If $i \in F$ *is unstable, i.e.,* $q_i = +\infty$*, then*

$$\lim_{\lambda\to\infty} \lambda \sum_{k\in G} \eta_{ik}(\lambda)\xi_{ki} = +\infty \qquad (\forall i \in F) \tag{2.13}$$

or, equivalently

$$\lim_{\lambda\to\infty} \lambda \sum_{k\in G} \eta_{ik}(\lambda) = +\infty \qquad (\forall i \in F) \tag{2.14}$$

while if $i \in F$ *is stable i.e.,* $q_i < \infty$*, then*

$$\lim_{\lambda\to\infty} \lambda \sum_{k\in G} \eta_{ik}(\lambda)\xi_{ki} < +\infty \tag{2.15}$$

and

$$q_i = c_{ii} + \lim_{\lambda\to\infty} \lambda \sum_{k\in G} \eta_{ik}(\lambda)\xi_{ki} \tag{2.16}$$

Here, $\xi = \{\xi_{ij}; i \in G, j \in F\}$ *in* (2.10)–(2.13) *and* (2.15)–(2.16) *is*

$$\xi = \lim_{\lambda\to 0} \xi(\lambda). \tag{2.17}$$

Moveover, if $R(\lambda)$ is honest, then (2.12) *and the second inequality in* (2.6) *become equalities, that is*

$$\sum_{j\in F} c_{ij} = \lim_{\lambda\to\infty} \lambda \sum_{k\in G} \eta_{ik}(\lambda)(1 - \sum_{j\in F} \xi_{kj}) \qquad (\forall i \in F) \tag{2.18}$$

and

$$\sum_{j\in F} \xi_{ij}(\lambda) = 1 - \lambda \sum_{k\in G} \psi_{ik}(\lambda) \qquad (\forall i \in G) \tag{2.19}$$

This extremely useful theorem has a very clear probabilistic meaning. It is just the Laplace transform version of first entrance — last exit decomposition theorem. Indeed, $\xi(\lambda)$ and $\eta(\lambda)$ are simply the Laplace transforms of the first entrance time to, and last exit time from, the subset F of the corresponding Markov chain and $\Psi(\lambda)$ is just the taboo-resolvent. See Chung [15] for the celebrated idea of taboo probability. This idea has been extensively developed by Syski [43], though the latter book concentrated on the Feller minimal chains and thus the q-matrix concerned is totally stable. It should be emphasized that the $A(\lambda)$ in (2.1) is the Laplace transform of the"transition function" of a quasi-Markov chain, a theory brilliantly developed by Kingman [28], in which the "Markov characterization problem" was tackled and solved.

Surely, the decomposition theorem 2.2 has a long history which can be at least traced to Neveu [30]–[32]. Based on Neveu and Chung's works, Williams systematically studied and raised it to a considerable high level, see Rogers and Williams [42].

However, it seems that people have paid less attention to the converse of theorem 2.2, which, in our opinion, has more applications, particularly, in the study of unstable chains. That is the following result.

Theorem 2.3 *Let $Q = (q_{ij}; i, j \in E)$ be a pre-q-matrix and F is a finite subset of E. Suppose there exists a Q_G-resolvent $\Psi(\lambda)$ and a $\eta(\lambda)$ and $\xi(\lambda)$ such that* (2.3)–(2.8) *and* (2.13)–(2.15) *are satisfied, where $G = E\backslash F$ and Q_G is the restriction of Q on $G\times G$, then Q is a q-matrix, that is there exists a Q-process. Moveover, if the above $\Psi(\lambda), \xi(\lambda)$ and $\eta(\lambda)$ further satisfy*

$$\sum_{k\in F} \xi_{ik}(\lambda) = 1 - \lambda \sum_{k\in G} \psi_{ik}(\lambda) \qquad (\forall i \in E)) \tag{2.20}$$

and for all stable $i \in F$

$$\sum_{j\in F} q_{ij} = -\lim_{\lambda\to\infty} \lambda \sum_{k\in G} \eta_{ik}(\lambda) \tag{2.21}$$

then there exists an honest Q-process. The corresponding Q-resolvent (honest and dishonest) may be constructed in using (2.10)–(2.17).

The important thing in Theorem 2.3 is that it not only gives the existence conditions but also yields uniqueness criteria. It also provides a method to construct the q-resolvents, by which the property of the corresponding q-processes may be analysed. This makes Theorems 2.2 and 2.3 useful even for the totally stable q-processes. In particular, if the underlying Q_G resolvent $\Psi(\lambda)$ is known, then the property of the Q-process may be easily derived. This idea stimulated some new research works. See, for example, Chen and Renshaw [10] in which the underlying structure is an M/M/1 (queue), and Chen and Renshaw [5, 8] in which the underlying structure is a simple branching process.

As existence conditions are concerned, Theorem 2.3 does not provide further information for the totally instantaneous chains. For the mixing case, however, Theorem 2.3 is quite informative. For example, if the instantaneous states form a finite (non-empty) set, then much more information may be obtained. In order to state such results, let us first denote $F = \{i \in E; q_i = +\infty\}$ where F is a finite set and, again, let $G = E \backslash F$. Further define a Q-process as almost B-type if

$$dP_{ij}(t)/dt = \sum_{k \in E} q_{ik} P_{kj}(t) \qquad (\forall i \in G, \forall j \in E) \tag{2.22}$$

or almost F-type if

$$dP_{ij}(t)/dt = \sum_{k \in E} P_{ik}(t) q_{kj} \qquad (\forall i \in E, \forall j \in G) \tag{2.23}$$

Now we have the following conclusion.

Theorem 2.4 *Let* $Q = \{q_{ij}\}$ *be a pre-q-matrix with a finite set* $F = \{i \in E, q_i = +\infty\}$.

(i) $R(\lambda)$ *is an almost B-type Q-resolvent if and only if the restricting* Q_G*-resolvent* $\Psi(\lambda)$ *is B-type.*

(ii) $R(\lambda)$ *is an almost F-type Q-resolvent if and only if the restricting* Q_G*-resolvent* $\Psi(\lambda)$ *is F-type.*

(iii) Suppose the pre-q-matrix Q satisfies the conditions

$$\sum_{j \neq i} q_{ij} = +\infty \qquad (\forall i \in F) \tag{2.24}$$

$$\sum_{j \neq i} q_{ij} = q_i < +\infty \qquad (\forall i \in G) \tag{2.25}$$

then Q is a q-matrix (i.e. there exists a Q-process) if and only if for some (and therefore for all) $\lambda > 0$, the following condition

$$\sum_{j\in G}\sum_{k\in G} q_{ik}\phi_{kj}(\lambda) < +\infty \qquad (\forall i \in F) \tag{2.26}$$

holds true, where $\Phi(\lambda) = \{\phi_{ij}(\lambda); i, j \in G\}$ is the Feller minimal Q_G-resolvent. Also if (2.26) is true, then there exists at least one honest Q-process.

For an elementary and purely analytic proof of all the above Theorems 2.2 to 2.4, see Chen and Renshaw [5, 7] for the case where F is a single point and consult Hou et al [24] for the general case where F is a finite set.

As we have already mentioned, Theorems 2.2 to 2.4 are just Laplace transform versions of the Neveu- Chung- Williams- Kingman's decomposition theorem. However, this Laplace transform version seems more informative and could yield more results, see the next section. To our knowledge, the above refined version of Theorems 2.2 to 2.4 was first stated and proved in Chen [9], but unfortunately, in a hardly accessible language, Chinese! This is one of the reasons we stated Theorems 2.2 to 2.4 in detail here.

3. Unstable chains with a single instantaneous state

In order to illustrate the application of Feller transition functions and resolvent decomposition theorems, we now discuss unstable Markov chains with a single instantaneous state. Recall that a pre-q−matrix Q is called CUI, if (1.18) and (1.19) are satisfied. In this case, the set $F = \{i \in E, q_i = +\infty\}$ becomes a singleton, denoted by $\{b\}$ say. Again, let $G = E\backslash\{b\}$. Now the following conclusion is a direct consequence of the resolvent decomposition theorems.

Theorem 3.1 (Chen and Renshaw [7]) *Suppose Q is a CUI pre-q-matrix. Then*

(i) It becomes a q−matrix if and only if for some (and therefore for all) $\lambda > 0$,

$$\sum_{k\in G}\sum_{j\in G} q_{bj}\phi_{jk}(\lambda) < +\infty \tag{3.1}$$

where $\Phi(\lambda) = \{\phi_{ij}(\lambda); i, j \in G\}$ is the Feller minimal Q_G−resolvent. Futhermore, if (3.1) is satisfied then there exists an honest almost $B \cap F$-type Q-process.

(ii) *If* (3.1) *is satisfied, then there always exist infinitely many of $Q-$ processes with at least one honest one. The honest Q-process is unique if and only if both equations*

$$Y(\lambda)(\lambda I - Q_G) = 0, \qquad 0 \le Y(\lambda) \in l \tag{3.2}$$

and

$$(\lambda I - Q_G)U(\lambda) = 0, \ \ 0 \le U(\lambda) \le 1 \text{ and } \sum_{j\in G} q_{bj}U_j(\lambda) < +\infty \tag{3.3}$$

have only the trivial solution for some (and therefore for all) $\lambda > 0$.

Conditions (3.1)–(3.3) are on one hand quite satisfactory. They may, for example, provide complete solutions for the three examples mentioned in section 1 see later. Recall that these three examples are essentially the only results obtained for mixing Markov chains for around 30 years until the early 1980's. For the general CUI pre-q-matrix, they also provide much useful information. For example, the following corollary is immediate and informative.

Corollary 3.1 (Chen and Renshaw [7]) *Suppose Q is a CUI q-matrix, then*

(i) $\sup_{i\in E} q_i = +\infty$,

(ii) $\sum_{j\in G} [q_{bj}/(1+q_j)] < \infty$,

(iii) $\inf_{i\in G} \lambda \sum_{j\in G} \phi_{ij}(\lambda) = 0 (\forall \lambda > 0)$

and thus the Q_G-process (totally stable chain) is not unique. In particular, either the equation

$$(\lambda I - Q_G)U(\lambda) = 0, \qquad 0 \le U(\lambda) \le 1 \tag{3.4}$$

has a non-trivial solution for some (and therefore for all) $\lambda > 0$, *or*

$$\sup_{i\in G} \sum_{j\in G} \phi_{ij}(\lambda) q_{jb} = 1 \tag{3.5}$$

which implies that $\sup_{i\in G} q_{ib} = +\infty$.

It seems to us that we may not be able to expect much more than conditions (3.1)–(3.3) for the general CUI q-matrix.

On the other hand, however, condition (3.1) is not totally satisfactory since the Feller minimal Q_G-resolvent rather than Q itself is involved here. It is usually not easy to check this condition. It is therefore useful if one can give a more exact condition for a possibly narrow sub-class of CUI q-matrices. See also later.

It is now the time for us to look back at the three examples mentioned in Section 1. Although this survey paper does not intend to give any proof for the results stated, we shall briefly explain how simply the proofs can be given for Theorems 1.1, 1.2 and 1.3 to emphasize the importance of Feller transition function and resolvent decomposition theorems.

Proof of William's Theorem 1.2 Suppose there exists an honest Q-resolvent where Q satisfies (1.5), then by Theorem 2.2 there is a Q_G-resolvent ($G = E\backslash\{b\}$) $\Psi(\lambda)$ satisfying (2.3) and (2.4). It follows immediately from (2.3), (2.4) and the Fatou Lemma that

$$\sum_{k\in G}\sum_{j\in G} q_{bj}\psi_{jk}(\lambda) < +\infty.$$

Now condition (1.5) implies that $\sum\limits_{k\in G}\sum\limits_{j\in G}\psi_{jk}(\lambda) < +\infty$ and thus, by Theorem 2.1(iii), $\Psi(\lambda)$ is Feller and thus any state in G is stable and $\Psi(\lambda)$ is the Feller minimal Q_G-resolvent. We have thus proved (1.8) and (1.10). (1.7) is a trivial consequence of (1.5) and, finally (1.9) follows from the fact that $R(\lambda)$ is honest. The proof is now complete. ■

Proof of Theorem 1.1 Note that for Kolmogorov pre-q-matrix Q, Q_G is diagonal and thus the Feller minimal Q_G-resolvent is just $\phi_{ij}(\lambda) = \frac{\delta_{ij}}{\lambda+a_i}$ and so Q becomes a q-matrix if and only if $\sum\limits_{i=1}^{\infty}\frac{1}{1+a_i} < +\infty$ (here we take $\lambda = 1$ and use the fact $q_{bj} = 1$). Now this condition implies that $\lim_{i\to\infty} a_i = +\infty$ and thus it holds true if and only if $\sum\limits_{i=1}^{\infty}(\frac{1}{a_i}) < +\infty$. ■

Note that the above simple proof yields more than Theorem 1.1. That is, condition $\sum\limits_{i=1}^{\infty}(\frac{1}{a_i}) < +\infty$ is not only sufficient but also necessary. Actually, the following further result is immediate.

Theorem 3.2 *Suppose* $Q = \{q_{ij}\}$ *is a Kolmogorov pre-q-matrix, i.e.* Q *is given in* (1.1)–(1.3). *Then*

(i) Q *becomes a q-matrix if and only if* $\sum\limits_{i=1}^{\infty}(\frac{1}{a_i}) < +\infty$.

(ii) When Q *is a q-matrix, i.e. the condition in* (i) *is satisfied, then the honest* Q*-process is unique.*

(iii) *This (unique) honest Q-process is recurrent and, furthermore, positive recurrent.*

(iv) *This (unique) honest Q-process is reversible.*

(v) *The equilibrium distribution* $\Pi = (\pi_i)$, *say, is given by*

$$\pi_0 = \left(1 + \sum_{k=1}^{\infty}\left(\frac{1}{a_k}\right)\right)^{-1}, \quad \pi_i = \left(\frac{1}{a_i}\right)\pi_0, \quad (i \geq 1).$$

Now, how about Theorem 1.3? Is the condition given by Reuter also necessary? This time, however, the answer is negative. Actually we have the following conclusion.

Theorem 3.3 *Suppose Q is a Reuter pre-q-matrix, then*

(i) *Q becomes a q-matrix if and only if* $\sum_{i=1}^{\infty} \frac{b_i}{1+a_i} < +\infty$.

(ii) *If* $\sum_{i=1}^{\infty} (\frac{b_i}{a_i}) < +\infty$ *then Q is a q-matrix. The converse, however, is not true.*

(iii) *When Q is a q-matrix, then the honest Q-process is unique.*

(iv) *This (unique) honest Q-process is always recurrent.*

(v) *This (unique) honest Q-process is positive recurrent if and only if* $\sum_{i=1}^{\infty} (\frac{b_i}{a_i}) < +\infty$, *and under this condition, the equilibrium distribution* $\Pi = (\pi_i)$, *say, is given by*

$$\pi_0 = \left(1 + \sum_{k=1}^{\infty}\left(\frac{b_k}{a_k}\right)\right)^{-1}, \quad \pi_i = \left(\frac{b_i}{a_i}\right)\pi_0, \qquad (i \geq 0).$$

(vi) *The (unique) honest Q-process is reversible if and only if* $\sum_{i=1}^{\infty} (\frac{b_i}{a_i}) < +\infty$.

Therefore, although Reuter's condition is not necessary for the existence of Q-process, it is an essential condition in the sense that it is the "if and only if" condition for reversibility and ergodicity. Of course, these later two concepts are closely related to each other.

Now we turn our attention to Williams-q-matrix. The existence theorem has been given above. Now two further questions arise namely what is the uniqueness condition and (if not unique) how to construct all the corresponding q-processes. Of course, we are mainly interested in the honest ones.

The uniqueness criterion has been given in Theorem 3.1(ii) already. For Williams-q-matrix, however, simpler conditions may be obtained.

Theorem 3.4 *Suppose Q is a Williams-pre-q-matrix, then*

(i) Q becomes a q-matrix if and only if (3.1) *holds true.*

(ii) If Q is a Williams-q-matrix, then there always exist infinite many of Q-functions.

(iii) The Equation (3.3) *always has only the trivial solution for some (and therefore for all) $\lambda > 0$, though the equation*

$$(\lambda I - Q_b)U(\lambda) = 0, 0 \leq U(\lambda) \leq 1$$

may still have non-trivial solution for all $\lambda > 0$.

(iv) There exists only one honest Q-function if and only if Equation (3.2) *has only the trivial solution for some (and therefore for all) $\lambda > 0$.*

(v) All the Q-resolvents can be easily constructed.

The proof of Theorem 3.4 together with the construction of all Q-resolvents can be found in Chen and Renshaw [7].

We may see how easily we could tackle the three examples introduced in Section 1 (recall, again, these three examples are essentially the only results obtained for the mixing case until early 1980s) if we use the theory of Feller transition function and resolvent decomposition theorems. Of course, the powerfulness of such theory and method is mainly reflected by the fact that it can handle new and more complicated models.

A particular interesting model is the so-called unstable piecewise birth and death (PBD) processes, whose pre-q-matrix $Q = \{q_{ij}; i, j \in Z_+\}$ is given by (here only the non-zero, off-diagonal elements are specified)

$$q_{ij} = \begin{cases} h_j, & if\ i = 0,\ j \geq 1 \\ a_i, & if\ i \geq 1,\ j = i - 1 \\ b_i, & if\ i \geq 1,\ j = i + 1 \end{cases} \tag{3.6}$$

together with

$$q_0 = -q_{00} = \sum_{j=1}^{\infty} h_j = +\infty \tag{3.7}$$

and

$$q_i = -q_{ii} = a_i + b_i \qquad (i \geq 1) \tag{3.8}$$

$$\text{where} \quad a_i > 0, \; b_i > 0 \qquad (\forall i \geq 1). \tag{3.9}$$

Such model was initially considered by Tang [44] and later investigated by several authors. We shall not list the corresponding results here but just content ourselves with pointing out the following two facts: firstly, the existence condition is given in terms of the PBD pre-q-matrix Q itself directly and thus easy to check and, secondly, the properties of such structure are also given. A particular interesting result is that any honest unstable PBD process is recurrent and ergodic, provided that, of course, the existence condition is satisfied.

4. Markov branching processes

An interesting application of Feller function is in Markov branching processes (MBP). A d-type MBP is a continuous time Markov chain (CTMC) on the state space $E = Z_+^d$ which possesses the branching property, i.e. "independence property". Standard references on MBP are Harris [21]. Athreya and Ney [3] and Asmussen and Hering [1]. The importance and many applications of such processes are so well-known that it would be superfluous if we should repeat it here.

Note that, however, for such a well-known structure there exist several basic questions, that seems less well-known. Without loss of generality, let us consider the case of $d = 1$.

As the definition of MBP is concerned, there are actually two basic definitions: a probabilistic one and an analytic one.

Definition 4.1 (Probabilistic) *A (one dimensional) MBP is a CTMC on the state space $E = Z_+$ whose transition function $P(t) = \{p_{ij}(t);\ i, j \in Z_+\}$ satisfies the branching property, i.e.*

$$\sum_{i=0}^{\infty} p_{ij}(t)s^j = \left(\sum_{j=0}^{\infty} p_{1j}(t)s^j\right)^i \qquad (\forall i \geq 0, |s| \leq 1) \tag{4.1}$$

Definition 4.2 (Analytic) *A (one dimensional) MBP is a CTMC on the state space $E = Z_+$ whose transition function $P(t) = \{p_{ij}(t); i, j \in Z_+\}$ satisfies the Kolmogorov forward equations*

$$P'(t) = P(t)Q \tag{4.2}$$

where $Q = (q_{ij}; i, j \in Z_+)$ *is a totally stable q-matrix and taking the form of*

$$q_{ij} = \begin{cases} ib_{j-i+1}, & \text{if } j \geq i-1, \\ 0, & \text{otherwise} \end{cases} \tag{4.3}$$

where

$$0 \leq b_j < +\infty \qquad (j \neq 1) \tag{4.4}$$

and

$$0 < \sum_{j \neq 1} b_j \leq -b_1 < +\infty \tag{4.5}$$

Relation (4.1) is called "branching property" since it is easy to see that (4.1) is equivalent to

$$p_{ij}(t) = \sum_{j_1 + \cdots + j_i = j} p_{1j_1}(t) \cdots p_{1j_i}(t) \qquad (\forall i, j \in Z_+, t \geq 0) \tag{4.6}$$

which states that different particles are independent in giving birth or death.

It has been a long history since people understood that Definition 4.2 implies Definition 4.1, see the proof in Harris [21] or Athreya and Ney [3].

However, how about the converse? In particular, if a CTMC satisfies the branching property, will the q-matrix Q of this CTMC be totally stable and even if so, will the transition function of the CTMC satisfy the Kolmogorov forward equation (4.2)? If we could not answer this question, we might have lost a large class of "new MBP"!

Another related question is whether there exists a so-called totally instantaneous (but one) branching process. That is, suppose a pre-q-matrix $Q = (q_{ij})$ is given as in (4.3)–(4.5) but with the amendment that $-b_1 = \infty$ in (4.5), then does there exists a standard transition function $P(t)$ such that $P(t)$ satisfies (4.1) and the condition $P'(0) = Q$ (i.e. $P(t)$ is a $Q-$process)? Of course, in order to make this question meaningful, we need first to illustrate that there exists a transition function $P(t)$ such that $P'(0) = Q$. This, however, can be answer by Williams' TI existence theorem (see Williams [47]), by which there exists a $P(t)$ such that $P'(0) = Q$ if and only if

$$\sum_{j=2}^{\infty} (b_j \wedge b_{j+k}) < +\infty \qquad (\forall k \geq 1) \tag{4.7}$$

It is fairly easy to find a sequence $\{b_j\}$ such that (4.7) is true. For example, if we adopt the sequence $\{1, 0, 1, 0, 0, 1, 0, 0, 0, 1, 0, \cdots\}$, then this sequence is not summable but satisfies (4.7).

Thus, the essential question is whether there exists a Q-function, where Q is a TI branching pre-q-matrix, such that the branching property holds true.

The above two questions can, actually, be answered by using the property of Feller transition function. Indeed, we have

Theorem 4.1 *The two* Definitions 4.1 *and* 4.2 *are equivalent.*

Sketch of the proof We only need to prove that Definition 4.1 implies Definition 4.2. By (4.1), we see that for any $t \geq 0$ and $0 < s < 1$, we have

$$p_{ij}(t)s^j \leq \sum_{j=0}^{\infty} p_{ij}(t)s^j = (\sum_{j=0}^{\infty} p_{1j}(t)s^j)^i \tag{4.8}$$

However, $\sum_{j=0}^{\infty} p_{1j}(t)s^j < 1$ since $0 < s < 1$ and thus let $i \to \infty$ in (4.8) immediately yields that $\lim_{i\to\infty} p_{ij}(t) = 0 (\forall j \in E)$. That is, $P(t)$ is actually the Feller minimal Q-function and thus satisfies the Kolmogorov forward function. Now an easy algebra yields the result that Q must take the form of (4.3)–(4.5).

Remark 4.1 *If the requirement of honesty is imposed to the transition function, then we can further prove that the Q matrix must be conservative and thus the second inequality in* (4.5) *becomes an equality.*

Now the following two corollaries immediately follow.

Corollary 4.1 *The Markov branching process is always unique no matter the q-matrix Q is regular or not.*

Corollary 4.2 *There exists no totally instantaneous (but one) branching processes.*

All the above results can be easily generalized to the d-type Markov branching processes.

In spite of Corollary 4.2, it is meaningful to consider the branching processes with the so-called instantaneous immigration. For example, we may consider a branching process with instantaneous immigration at state zero only. This is another interesting example of uni-instantaneous processes, in connection with the theory developed in the last section.

More specially, a pre-q-matrix Q is called a branching pre-q-matrix (with or without immigration or called resurrection), if $Q = (q_{ij})$ takes the form of

$$q_{ij} = \begin{cases} h_j, & \text{if } i = 0 \\ ib_{j-i+1}, & \text{if } i \geq 1,\ j \geq i-1 \\ 0, & \text{otherwise} \end{cases} \tag{4.9}$$

where

$$0 \leq -h_0 = \sum_{j=1}^{\infty} h_j \leq +\infty \tag{4.10}$$

and

$$0 < -b_1 = \sum_{j \neq 1} b_j < +\infty \tag{4.11}$$

Furthermore, Q is called "a branching q-matrix without resurrection" if $h_j = 0$; "a branching q-matrix with stable resurrection" if $0 < -h_0 < +\infty$, and "a branching pre-q-matrix with instantaneous resurrection" if $-h_0 = +\infty$. We shall apply all these terms to the corresponding processes (Q-functions; Q-resolvents) as well.

In order to guarantee that the underlying structure possesses the branching property, we shall define the branching process as the one that is F-type (almost F-type if instantaneous resurrection) i.e. the one which satisfies the Kolmogorov forward equations.

Branching processes with stable resurrection was considered by Yamazato [48]. This is a continuous version of the discrete time branching model investigated, nearly at the same time, by Foster [20] and Pakes [33]. In order to cite Yamazato's result, it is convenient to introduce generating functions of the two sequences $\{b_j\}$ and $\{h_j\}$ as

$$B(s) = \sum_{j=0}^{\infty} b_j s^j \tag{4.12}$$

$$H(s) = \sum_{j=1}^{\infty} h_j s^j \tag{4.13}$$

and

$$U(s) = h_0 + H(s) = \sum_{j=0}^{\infty} h_j s^j \tag{4.14}$$

Note that the two sequences $\{b_j\}$ and $\{h_j\}$ are the basic data of the branching processes with stable resurrection, and thus the above generating functions provide the full known information. Yamazato's main result can now be stated as follows.

Theorem 4.2 (Yamazato [48]) *For a branching process with stable resurrection, the following conclusions may be claimed.*

(i) The process is recurrent if and only if $B'(1) \leq 0$.

(ii) The process is positive recurrent if and only if $\int_0^1 \frac{U(s)}{B(s)} ds > -\infty$.

(iii) The moments of the process can be obtained.

The equilibrium distribution for the positive recurrent case was obtained by Pakes [34].

The analysis of branching processes with instantaneous resurrection has only been available quite recently. Interestingly, however, this model is still tractable, due to the effect of Chen and Renshaw [5, 6, 8].

Theorem 4.3 *Suppose Q is a branching pre-q-matrix with instantaneous resurrection. Then we have*

(i) There exists a branching process with instantaneous resurrection if and only if the following two conditions hold true.

a) $B'(1) > 0$ and thus $B(s) = 0$ has a unique root on $[0, 1)$. That is, there exists a q such that $0 \leq q < 1$ and $B(q) = 0$.

b)

$$\int_0^1 \frac{H(q) - H(s)}{B(s)} ds < +\infty \tag{4.15}$$

where $0 \leq q < 1$ is given in (a) *and $B(s)$ and $H(s)$ are given in* (4.12) *and* (4.13), *respectively.*

(ii) If the existence condition in (i) is satisfied, then there exists infinitely many of branching processes with instantaneous resurrection but only one of them is honest. That is, the honest branching process with instantaneous resurrection is unique.

(iii) The (unique) honest branching process with instantaneous resurrection is not only recurrent but also positive-recurrent.

For the positive recurrence case, the equilibrium distribution is, again, obtained by Pakes [34].

It is interesting to compare the conclusions in Theorems 4.2 and 4.3. For the Yamazato's model, $B'(1) \leq 0$ is a necessary condition for the recurrence while for the Chen-Renshaw model, $B'(1) > 0$ is required to guarantee the existence of a branching process with instantaneous resurrection and once the existence condition is satisfied, then the honest process is unique, recurrent, and positive recurrent.

Note that for the later model, $U(s)$ in (4.14) can not be defined. However, it can be proved that, as a necessary condition for the existence of the q-process, $H(s)$ is well defined for all $|s| < 1$. Also, the existence condition implies Harris' non-honest condition for the underlying branching process without resurrection. For the details, see Chen and Renshaw [5].

More recently, Chen [13] considered a much more general branching model by replacing the coefficient i in (4.9) by the general form v_i. Some interesting results have been obtained by wisely using the techniques and methods developed by Chen [11], [12]. Of course, only the stable resurrection has been considered until now. It will be interesting to investigate such general structure but with instantaneous resurrection. The general results obtained in Section 3 will be helpful in tackling such questions.

References

[1] S. Asmussen and H. Hering, Branching Processes (Birkhauser, Boston, 1983).

[2] K.B. Athreya and P. Jagers, Classical and Modern Branching Processes (Springer, Berlin, 1996).

[3] K.B. Athreya and P.E. Ney, Branching Processes (Springer, Berlin, 1992).

[4] Blackwell, D., Another countable Markov process with only instantaneous states, Ann.Math. Stat. 29, 313–316 (1958).

[5] A.Y. Chen and E. Renshaw, Markov branching processes with instantaneous immigration, Prob, Theor. Rel. Fields, 87 (1990) 209–240.

[6] A.Y. Chen and E. Renshaw, Recurrence of Markov branching processes with immigration, Stoch. Proc. Appl., 45 (1993a) 231–242.

[7] A.Y. Chen and E. Renshaw, Existence and uniqueness criteria for conservative uni-instantaneous denumerable Markov processes, Prob. Theor. Rel. Fields, 94 (1993b) 427–456.

[8] A.Y. Chen and E. Renshaw, Recurrence of Markov branching processes regulated by emigration and large immigration, Stoch. Proc. Appl., 57 (1995) 339–359.

[9] A.Y. Chen, Some Construction Problems in Q-processes with instantaneous states, Ph. D. Thesis (1988).

[10] A.Y. Chen and E. Renshaw, The M/M/1 Queue with Mass Exodus and Mass Arrivals when empty. J.Appl. Prob. 34, 192–207 (1997).

[11] M.F. Chen, On three classical problems for Markov chains with constituous time parameters, J. Appl. Probab. 28 (1991) 305–320.

[12] M.F. Chen, From Markov Chains to Non-Equilibrium Particle Systems(World Scientific, Singapore, 1992).

[13] R.R. Chen, An extended class of time-continuous branching processes, J. Appl. Probab. 34 (1997) 14–23.

[14] Chung, K.L., On the boundary theory for Markov chains. Acta. Math. 115, 111–163 (1966).

[15] Chung, K.L., Markov chains with stationary transition probabilities. Berlin Heidelberg New York: Springer 1967.

[16] Chung, K.L., Lectures on boundary theory for Markov chains. (Ann. Math. Stud., 65) Princeton University Press 1970.

[17] Doob, J.L., Markov chains - Denumerable case. Trans. Am. Math. Soc. 58, 455–473 (1945)

[18] Feller, W., On the integro-differential equations of purely discontinuous Markov processes. Trans. Am. Math. Soc. 48, 488–515 (1940)

[19] Freedman, D., Approximating countable Markov chains. Berlin, Heidelbeg, New York: Springer 1983.

[20] J.H. Foster, A limit theorem for a branching process with state-dependent immigration, Ann. Math. Ststist., 42 (1971) 1773–1776.

[21] T.E. Harris, The Theory of Branching Processes (Springer, Berlin, 1963).

[22] Hou, Z.T., The uniqueness criterion for Q-processes (in Chinese). Sinica 2, 115–130 (1974).

[23] Z.T. Hou and Q.F. Guo, Homogeneous Denumerable Markov Processes (Springer, Berlin. 1988).

[24] Z.T. Hou et al., The Q-matrix Problem (Human Sciences Process, Changsha, 1994).

[25] W.B. Jurkat, Analytic Investigations of Markov Chains, Syracuse University. Research Report 17 (1957).

[26] Kendall, D.G., A totally unstable denumerable Markov process. Q.J. Math., Oxford 9, 149–160 (1958).

[27] Kendall, D.G. and Reuter, G.E.H., Some pathological Markov processes with a denumerable infinity of states and the associated semigroups of operators on $\mathcal{L}$. Proc. Intern, Congr. Math. Amsterdam, Vol. III, 377–415. Amsterdam: North-Holland 1954.

[28] Kingman, J.F.C., Regenerative Phenomena Wiley, London (1972).

[29] Kolmogorov, A.N. On the differentiability of the transition probabilities in stationary Markov processes with a denumerable number of states (in Russian). Moskov. Gos. Univ. Ucenye Zapiski, Vol. 148, Mat. 4, 53–59 (1951).

[30] J. Neveu, Lattice methods and submarkovian processes. Proc. Fourth Berleley Symp. Math. Statist. Prob. 2, 347–391, University of California Press, Berkeley, California (1961).

[31] J. Neveu, Entrance, exit and fictitious states for Markov chains, Proc. Aarhus Colloq. Combin Prob 64–68 (1962a).

[32] J. Neveu, Sur les, etats d'entree ef les etats fictifs d'un processus de Markov, Ann. Inst. H. Poincare' 17, 323–337 (1962b).

[33] A.G. Pakes, A branching process with a state-dependent immigration component, Adv. Appl. Probab. 3(1971) 301–314.

[34] A.G. Pakes, Absorbing Markov and branching processes with instantaneous resurrection, Stochastic Process. Appl. 48, 85-106 (1993).

[35] Reuter, G.E.H., Denumerable Markov processes and the associated semigroup on $\mathcal{L}$. Acta. Math. 97, 1–46 (1957).

[36] Reuter, G.E.H., Denumerable Markov processes (II). J.Lond. Math. Soc. 34, 81-91 (1959).

[37] Reuter, G.E.H., Denumerable Markov processes (III). J.Lond. Math. Soc. 37, 63-73 (1962).

[38] Reuter, G.E.H., Remarks on a Markov chain example of Kolmogorov. Z. Wabrscheinlichkeits-theor. Verw. Geb. 13, 315–320 (1969).

[39] Reuter, G.E.H. and Riley, P.W., The Feller property for Markov semigroups on a countable state space. J.London. Math. Soc. (2)5, 267–275 (1972).

[40] Reuter, G.E.H., Denumerable Markov processes (IV). On C.T.Hou's uniqueness theorem for Q-semigroups. Z.Wahrscheinlichkeitstheor. Verw. Geb. 33, 309–315 (1976).

[41] Rogers, L.C.G. and Williams, D., Construction and approximation of transition matrix functions. Adv. Appl. Probab., Suppl. 133–160 (1986).

[42] L.C.G. Rogers and D. Williams, Diffusions, Markov Processes and Martingales, Vol.1. (2nd Edition), Wiley, Chichester (1994).

[43] Syski, R., Passage Times for Markov Chains, IOS Press. Amsterdam. (1992).

[44] Tang, L.Q., The construction of uni-instananeous birth and death processes. Chen Ann. Math. 8A(5), 565–570(1987).

[45] Williams, D., On the construction problem for Markov chains. Z. Wahrscheinlichkeitstheor. Verw. Geb. 3, 227–246 (1964).

[46] Williams, D., A note on the Q-matrices of Markov chains. Z. Wahrscheinlichkeitstheor. Verw. Geb. 7, (1967) 116–121.

[47] Williams, D., The Q-matrix problem. In Meye, P.A.(ed.) Seminaire de Probabilities X: (Lect. Notes Math., vol 511. 216–234) Berlin Heidelberg New York: Wiley (1976).

[48] M. Yamazato, Some results on continuous time branching processes with state-dependent immigration, J. Math. Soc. Japan. 27 (1975) 479–496.

[49] Yang, X.Q., The construction theory of denumerable Markov processes. New York: Wiley (1990).

Chapter 3

IDENTIFYING Q-PROCESSES WITH A GIVEN FINITE μ-INVARIANT MEASURE

P.K. Pollett
Department of Mathematics
The University of Queensland
Queensland 4072, Australia
pkp@maths.uq.edu.au

Abstract Let $Q = (q_{ij},\ i,j \in S)$ be a stable and conservative Q-matrix over a state space S consisting of an irreducible (transient) class C and a single absorbing state 0, which is accessible from C. Suppose that Q admits a *finite* μ-subinvariant measure $m = (m_j, j \in C)$ on C. We consider the problem of identifying all Q-processes for which m is a μ-invariant measure on C.

Keywords: Q-processes; quasi-stationary distributions; construction theory

1. Introduction

We begin with a totally stable Q-matrix over a countable set S, that is, a collection $Q = (q_{ij},\ i,j \in S)$ of real numbers which satisfies

$$\begin{aligned} 0 \le q_{ij} < \infty, &\quad i \ne j,\ i,j \in S, \\ q_i := -q_{ii} < \infty, &\quad i \in S, \\ \sum_{j \ne i} q_{ij} \le q_i, &\quad i \in S. \end{aligned} \tag{1.1}$$

The Q-matrix is said to be *conservative* if equality holds in (1.1) for all $i \in S$. For simplicity, we shall assume that Q is conservative. A set of real-valued functions $P(\cdot) = (p_{ij}(\cdot),\ i,j \in S)$ defined on $(0, \infty)$ is called

Z. Hou et al. (eds.), Markov Processes and Controlled Markov Chains, 41–55.

a *standard transition function* (or *process*) if

$$p_{ij}(t) \geq 0, \qquad i,j \in S,\ t > 0, \tag{1.2}$$

$$\sum_{j \in S} p_{ij}(t) \leq 1, \qquad i \in S,\ t > 0, \tag{1.3}$$

$$p_{ij}(s+t) = \sum_{k \in S} p_{ik}(s) p_{kj}(t), \quad i,j \in S,\ s,t > 0, \tag{1.4}$$

$$\lim_{t \downarrow 0} p_{ij}(t) = \delta_{ij}, \qquad i,j \in S. \tag{1.5}$$

P is then *honest* if equality holds in (1.3) for some (and then all) $t > 0$, and it is called a *Q-transition function* (or *Q-process*) if $p'_{ij}(0+) = q_{ij}$ for each $i,j \in S$. Under the conditions we have imposed, every Q-process P satisfies the *backward differential equations*,

$$p'_{ij}(t) = \sum_{k \in S} q_{ik} p_{kj}(t), \qquad t > 0, \qquad (\mathrm{BE}_{ij})$$

for all $i,j \in S$, but might not satisfy the *forward differential equations*,

$$p'_{ij}(t) = \sum_{k \in S} p_{ik}(t) q_{kj}, \qquad t > 0, \qquad (\mathrm{FE}_{ij})$$

for all $i,j \in S$. The classical construction problem is to find one and then all Q-processes. Feller's recursion (Feller [2]) provides for the existence of a *minimal* solution $F(\cdot) = (f_{ij}(\cdot),\ i,j \in S)$ to the backward equations (which also satisfies the forward equations); see also Feller [3] and Reuter [14]. This process is the *unique* Q-process if and only if the system of equations

$$\sum_{j \in S} q_{ij} x_j = \nu x_i, \qquad i \in S, \tag{1.6}$$

has no bounded, non-trivial solution (equivalently, non-negative solution) x for some (and then all) $\nu > 0$ (Reuter [14]); for the non-conservative case, see Hou [4]. When this condition fails, there are infinitely many Q-processes, including infinitely many honest ones (Reuter [14]), and the dimension d of the space of bounded vectors x on S satisfying (1.6), a quantity which does not depend on ν, determines the number of "escape routes to infinity" available to the process. A construction of all Q-processes was given by Reuter [15], [16] under the assumption that $d = 1$ (the *single-exit case*), and this was later extended to the *finite-exit case* ($d < \infty$) by Williams [22].

If (1.6) has infinitely many bounded non-trivial solutions, the problem of constructing all Q-processes remains unsolved; there are simply too many solutions of the backward equations to characterize. For this reason, variants of the classical construction have been considered in which various side conditions are imposed. The most recent work centres on an assumption that one is given an *invariant measure* for the Q-matrix, that is, a collection of positive numbers $m = (m_i,\ i \in S)$ which satisfy

$$\sum_{i\in S} m_i q_{ij} = 0, \qquad j \in S.$$

The problem is then to identify Q-processes with m as their invariant measure, that is

$$\sum_{i\in S} m_i p_{ij}(t) = m_j, \qquad j \in S,\ t > 0.$$

When does there exist such a Q-process, and, when is it a unique Q-process with the given invariant measure? This variant of the classical construction problem has particular significance when m is finite ($\sum m_i < \infty$), for then one is looking for a Q-process whose *stationary distribution* has been specified. The problem of existence, and then uniqueness in the single-exit case, was solved by Hou and Chen [5] under the assumption that Q is *m-symmetrizable*, that is,

$$m_i q_{ij} = m_j q_{ji}, \qquad i, j \in S,$$

(see Chen and Zhang [1] for the non-conservative case) and by myself in the general case (Pollett [10], [12]). Recently Han-jun Zhang announced a solution to the existence problem under more general circumstances; see Zhang et al. [23], [24].

In this paper we shall look at a slightly different kind of construction problem, where the state space can be decomposed into an irreducible class C and a single absorbing state, and we shall suppose, rather than an *invariant* measure, a μ-invariant measure on C is specified through Q. We seek to determine Q-processes for which m is a μ-invariant measure on C. Since here we shall assume that the μ-invariant measure is *finite*, we are effectively identifying Q-processes with a given *quasi-stationary distribution* (van Doorn [20]). And, since we will not necessarily require these processes to satisfy the forward equations, we shall relax the μ-invariance for Q to μ-*sub*invariance for Q.

Before proceeding, let me remark that in this introductory section I have restricted my attention to the *totally stable case* ($q_i < \infty$ for all $i \in S$). Of course, the problem of constructing Q-processes when all states, or a finite subset of states, are unstable is an important one, and can be traced back to Lévy and Kolmogorov; for an informative summary see Rogers and Williams [18].

2. Preliminaries

We shall suppose that $S = \{0\} \cup C$, where C is an irreducible class (for the minimal Q-process, and hence for any Q-process) and 0 is an absorbing state which is accessible from C, that is $q_0 = 0$ and $q_{i0} > 0$ for at least one $i \in C$. Then, if μ is some fixed non-negative real number, a collection of strictly positive numbers $m = (m_j,\ j \in C)$ is called a *μ-subinvariant measure(on C) for Q* if

$$\sum_{i \in C} m_i q_{ij} \leq -\mu m_j, \qquad j \in C, \tag{2.1}$$

and *μ-invariant* if equality holds for all $j \in C$. We shall suppose that Q admits a μ-subinvariant measure on C, and then identify Q-processes P such that m is a *μ-invariant (on C) for P*, that is,

$$\sum_{i \in C} m_i p_{ij}(t) = e^{-\mu t} m_j, \qquad j \in C,\ t > 0. \tag{2.2}$$

The relationship between (2.1) and (2.2) has been divined completely for the minimal Q-process F. It was shown by Tweedie [19] that if m is a μ-invariant measure for F, then it is μ-invariant for Q. Conversely, Pollett [8, 9], if m is a μ-invariant measure for Q, then it is μ-invariant for F if and only if the equations

$$\sum_{i \in C} y_i q_{ij} = -\nu y_j, \qquad 0 \leq y_j \leq m_j,\ j \in C,$$

have no non-trivial solution for some (and then all) $\nu < \mu$. If $\mu > 0$ and the measure m is assumed to be *finite*, that is $\sum_{i \in C} m_i < \infty$, then much simpler conditions obtain (Pollett and Vere-Jones [13], Nair and Pollett [7]). For example, if F is *honest* (and hence the unique Q-process), then a finite μ-subinvariant measure m for Q is μ-invariant for F *if and only if*

$$\mu \sum_{i \in C} m_i = \sum_{i \in C} m_i q_{i0}. \tag{2.3}$$

As we shall see, this condition guarantees, more generally, that there *exists* a Q-process P such that m is a μ-invariant measure for P; it is honest and satisfies (FE_{i0}) for $i \in C$. We note that, in determining such a P, we are effectively identifying a Q-process with a given *quasi-stationary distribution* (van Doorn [20]): a probability distribution $\pi = (\pi_j,\ j \in C)$ over C is called a quasi-stationary distribution if $p_j(t)/\sum_{i \in C} p_i(t) = \pi_j$ for all $t > 0$, where $p_j(t) = \sum_{i \in C} \pi_i p_{ij}(t)$, $t > 0$, so that, conditional on non-absorption, the state probabilities of the underlying continuous-time Markov chain are stationary. It was shown by Nair and Pollett [7] that a distribution $\pi = (\pi_j,\ j \in C)$ is a quasi-stationary distribution if and only if, for some $\mu > 0$, π is a μ-invariant measure for P, in which case if P is honest, then $p_{i0}(t) \to 1$ for all $i \in C$ as $t \to \infty$ (absorption occurs with probability 1).

3. The main result

We shall specify transition functions through their resolvents. If P is a given transition function, then the function $\Psi(\cdot) = (\psi_{ij}(\cdot),\ i,j \in S)$ given by

$$\psi_{ij}(\alpha) = \int_0^\infty e^{-\alpha t} p_{ij}(t)dt, \qquad \alpha > 0, \tag{3.1}$$

is called the *resolvent of* P. If $i,j \in C$, the integral in (3.1) converges for all $\alpha > -\lambda_P(C)$, where $\lambda_P(C)$ is the *decay parameter of* C (for P); see Kingman [6]. In particular, since C is irreducible, the integral (3.1) has the same abcissa of convergence for each $i,j \in C$. Notice also that, since 0 is an absorbing state, $\psi_{0j}(\alpha) = \delta_{0j}/\alpha$. Analogous to properties (1.2)–(1.5), Ψ satisfies

$$\psi_{ij}(\alpha) \geq 0, \qquad i,j \in S,\ \alpha > 0, \tag{3.2}$$

$$\sum_{j \in S} \alpha\psi_{ij}(\alpha) \leq 1, \qquad i \in S,\ \alpha > 0, \tag{3.3}$$

$$\psi_{ij}(\alpha) - \psi_{ij}(\beta) + (\alpha - \beta)\sum_{k \in S} \psi_{ik}(\alpha)\psi_{kj}(\beta) = 0, \qquad i,j \in S,\ \alpha,\beta > 0, \tag{3.4}$$

$$\lim_{\alpha \to \infty} \alpha\psi_{ij}(\alpha) = \delta_{ij}, \quad i,j \in S. \tag{3.5}$$

(Note that (3.4) is called the *resolvent equation.*) Indeed, any Ψ which satisfies (3.2)–(3.5) is the resolvent of a standard transition function (see Reuter [15], [16]). Further, (3.3) is satisfied with equality if and only if P is honest, in which case the *resolvent* is said to be honest. Also, the Q-matrix of P can be recovered from Ψ using the following identity:

$$q_{ij} = \lim_{\alpha \to \infty} \alpha(\alpha\psi_{ij}(\alpha) - \delta_{ij}). \tag{3.6}$$

And, a resolvent which satisfies (3.6) is called a Q-resolvent. The resolvent $\Phi(\cdot) = (\phi_{ij}(\cdot),\ i,j \in S)$ of the minimal Q-process has itself a minimal interpretation (see Reuter [14], [15]); it is the minimal solution to the equations

$$\alpha\psi_{ij}(\alpha) = \delta_{ij} + \sum_{k \in S} q_{ik}\psi_{kj}(\alpha), \qquad i,j \in S,\ \alpha > 0,$$

which are analogous to (BE_{ij}), and Φ is called the minimal Q-resolvent.

We can identify μ-invariant measures using resolvents. If P is a Q-process with resolvent Ψ and $m = (m_j,\ j \in C)$ is a μ-invariant measure for P, where of necessity $\mu \leq \lambda_P(C)$ (see Lemma 4.1 of Vere-Jones [21]), then, since the integral in (3.1) converges for all $\alpha > -\lambda_P(C)$, we have, for all $j \in C$ and $\alpha > 0$, that

$$\sum_{i\in C} m_i \alpha \psi_{ij}(\alpha - \mu) = m_j. \tag{3.7}$$

We refer to m as being μ-invariant for Ψ if (3.7) is satisfied. Finally, a simple extension of Lemma 4.1 of Pollett [11] establishes that m is μ-invariant for Ψ if it is μ-invariant for P, and, if $\mu \leq \lambda_P(C)$, then m is μ-invariant for P if it is μ-invariant for Ψ.

We are now ready to state our main result.

Theorem 3.1 *Let $\mu > 0$ and suppose that Q admits a finite μ-subinvariant measure. Then, if*

$$\mu \sum_{i\in C} m_i = \sum_{i\in C} m_i q_{i0}, \tag{3.8}$$

there exists a Q-process P for which m is μ-invariant. The resolvent $\Psi(\cdot) = (\psi_{ij}(\cdot),\ i,j \in S)$ of one such Q-process has the form

$$\psi_{ij}(\alpha) = \phi_{ij}(\alpha) + \frac{z_i(\alpha) d_j(\alpha)}{(\alpha+\mu)\sum_{k\in C} m_k z_k(\alpha)}, \qquad i,j \in S, \tag{3.9}$$

where $z(\cdot) = (z_i(\cdot),\ i \in C)$ is given by

$$z_i(\alpha) = 1 - \sum_{j\in S} \alpha \phi_{ij}(\alpha), \qquad i \in C,$$

with the interpretation that $\Psi = \Phi$ if z is identically 0, and $d(\cdot) = (d_i(\cdot),\ i \in S)$ is given by

$$d_i(\alpha) = m_i - \sum_{j\in C} m_j (\alpha+\mu)\phi_{ji}(\alpha), \qquad i \in C, \tag{3.10}$$

$$d_0(\alpha) = \frac{\mu}{\alpha}\sum_{j\in C} m_j - \sum_{j\in C} m_j(\alpha+\mu)\phi_{j0}(\alpha). \tag{3.11}$$

This process is honest and satisfies (FE_{i0}) for $i \in C$.

Proof First observe that if z is identically 0, the minimal Q-process F is honest and, by Theorem 3 of Pollett and Vere-Jones [13], (3.8) is necessary and sufficient for m to be μ-invariant for F (in which case d is identically 0 and, by Proposition 2 of Tweedie [19], m is μ-invariant for Q). Trivially, F satisfies (FE_{i0}) for $i \in C$.

Suppose that z is not identically 0. We will first show that m cannot be μ-invariant for F and, in so doing, establish that d is not identically 0. Suppose, by contradiction, that m *is* μ-invariant for F, so that

$$\sum_{i \in C} m_i \phi_{ij}(\alpha) = \frac{m_j}{\alpha + \mu}, \qquad j \in C. \tag{3.12}$$

Multiplying by α and summing over $j \in C$ gives

$$\sum_{i \in C} m_i \alpha \phi_{i0}(\alpha) + \sum_{i \in C} m_i z_i(\alpha) = \frac{\mu}{\alpha + \mu} \sum_{j \in C} m_j. \tag{3.13}$$

Now, since F satisfies (FE_{ij}) over S, we have in particular that

$$\alpha \phi_{i0}(\alpha) = \sum_{j \in C} \phi_{ij}(\alpha) q_{j0}, \qquad i \in C, \tag{3.14}$$

and so, again using (3.12), we get

$$\sum_{i \in C} m_i \alpha \phi_{i0}(\alpha) = \frac{1}{\alpha + \mu} \sum_{i \in C} m_i q_{i0}.$$

This expression combines with (3.13) and (3.8) to give $\sum_{i \in C} m_i z_i(\alpha) = 0$, which is a contradiction because z is not identically 0. We deduce that m cannot be μ-invariant for F. Moreover, we must have

$$\sum_{i \in C} m_i (\alpha + \mu) \phi_{ij}(\alpha) < m_j \tag{3.15}$$

for at least one $j \in C$, and hence, from (3.14) and (3.8),

$$\sum_{i \in C} m_i (\alpha + \mu) \phi_{i0}(\alpha) < \frac{1}{\alpha} \sum_{i \in C} m_i q_{i0} = \frac{\mu}{\alpha} \sum_{i \in C} m_i.$$

Thus, $d_0(\alpha) > 0$ and $d_j(\alpha) > 0$ for at least one $j \in C$.

Next we shall show that $\Psi(\cdot) = (\psi_{ij}(\cdot),\ i, j \in S)$, given by (3.9), is the resolvent of an honest Q-process P and that m is a μ-invariant measure for P. Clearly $\psi_{ij}(\alpha) \geq 0$ for all $i, j \in S$. Since m is finite, we have, from the definition of d, that

$$\alpha \sum_{j \in S} d_j(\alpha) = (\alpha + \mu) \sum_{j \in C} m_j z_j(\alpha)$$

and so $\sum_{j\in S}\alpha\psi_{ij}(\alpha) = 1$ for all $i \in S$. In order to prove that Ψ is the resolvent of a standard transition function P, we need only show that Ψ satisfies the resolvent equation (3.4); see Theorem 1 of Reuter [17]. We shall use the following identities:

$$z_i(\alpha) - z_i(\beta) + (\alpha - \beta)\sum_{k\in C}\phi_{ik}(\alpha)z_k(\beta) = 0, \qquad i \in C, \tag{3.16}$$

$$d_i(\alpha) - d_i(\beta) + (\alpha - \beta)\sum_{k\in C}d_k(\alpha)\phi_{ki}(\beta) = 0, \qquad i \in C, \tag{3.17}$$

$$\alpha d_0(\alpha) - \beta d_0(\beta) + (\alpha - \beta)\sum_{k\in C}d_k(\alpha)\beta\phi_{k0}(\beta) = 0 \tag{3.18}$$

and

$$(\alpha+\mu)\sum_{i\in C}m_i z_i(\alpha)-(\beta+\mu)\sum_{i\in C}m_i z_i(\beta) = (\alpha-\beta)\sum_{i\in C}d_i(\alpha)z_i(\beta). \tag{3.19}$$

The first three of these can be verified directly using the fact that Φ satisfies the resolvent equation and that $z_0(\alpha) = 0$. The fourth identity follows from the first on multiplying by m_i and summing over i. Using (3.16)–(3.19), together with the resolvent equation for Φ, it is easy to prove that Ψ satisfies its own resolvent equation.

Next we need to verify that P is indeed a Q-process, that is $p'_{ij}(0+) = q_{ij}$ for all $i, j \in S$. We shall use a remark of Reuter [15, page 83] (see also Feller [3, Theorem 3.1]): if one is given a standard transition function P, then it is a Q-process if and only if the backward equations hold, equivalently,

$$\alpha\psi_{ij}(\alpha) = \delta_{ij} + \sum_{k\in S}q_{ik}\psi_{kj}(\alpha), \qquad i, j \in S.$$

But, this follows almost immediately from the identity

$$\sum_{k\in S}q_{ik}z_k(\alpha) = \alpha z_i(\alpha), \qquad i \in S,$$

which can be deduced from the backward equations for Φ.

We have shown that Ψ is the resolvent of a Q-process P. To show that m is a μ-invariant measure for P, we again use the definition of d: it is elementary to check that

$$\sum_{i\in C}m_i(\alpha + \mu)\psi_{ij}(\alpha) = m_j, \qquad j \in C.$$

We have already seen that P is honest and so it remains only to show that P satisfies (FE_{i0}) for $i \in C$. But, since z is not identically 0, this happens when and only when

$$\sum_{i\in C} d_i(\alpha)q_{i0} = \alpha d_0(\alpha),$$

because it is easily verified that

$$\sum_{k\in C} \psi_{ik}(\alpha)q_{k0} = \alpha\psi_{i0}(\alpha) + C_\alpha z_i(\alpha)\left(\sum_{k\in C} d_k(\alpha)q_{k0} - \alpha d_0(\alpha)\right),$$

where

$$C_\alpha^{-1} = (\alpha+\mu)\sum_{k\in C} m_k z_k(\alpha).$$

On substituting for d, we find that

$$\alpha d_0(\alpha) - \sum_{i\in C} d_i(\alpha)q_{i0} = \mu\sum_{k\in C} m_k - \sum_{k\in C} m_k q_{k0} = 0,$$

and so the result follows. ■

Remark 3.1 *The final part of the theorem states that the process P we have identified satisfies* (FE_{i0}) *for $i \in C$. The remaining forward equations do not necessarily hold. By Nair and Pollett* [7, Theorem 3.1], *this happens when and only when m is μ-invariant for Q (rather than merely μ-subinvariant). Indeed, a simple calculation shows that, for all $j \in C$,*

$$\sum_{k\in C} \psi_{ik}(\alpha)q_{kj} = \alpha\psi_{ij}(\alpha) - \delta_{ij} + C_\alpha z_i(\alpha)\left(\sum_{k\in C} d_k(\alpha)q_{kj} - \alpha d_j(\alpha)\right),$$

and, for the given P,

$$\alpha d_j(\alpha) - \sum_{i\in C} d_i(\alpha)q_{ij} = -\mu m_j - \sum_{i\in C} m_i q_{ij}\ (\geq 0), \qquad j \in C,$$

this later quantity measuring the "μ-invariance deficit" of m for Q.

Remark 3.2 *A straightforward calculation shows that the given Ψ satisfies*

$$\sum_{i\in C} m_i(\alpha+\mu)\psi_{i0}(\alpha) = \mu\sum_{i\in C} m_i,$$

and hence m satisfies a set of "residual equations" for P, namely

$$\sum_{i\in C} m_i p_{i0}(t) = (1-e^{-\mu t})\sum_{i\in C} m_i, \qquad t>0, \tag{3.20}$$

which can be regarded as a "process counterpart" to (3.8). *(Since P is honest,* (3.20) *follows more directly on summing* (2.2) *over $j \in C$.)*

4. Necessary conditions

It would be tempting to conjecture that condition (3.8) is *necessary* for the existence of a Q-process for which the given measure is μ-invariant. However, while this is *not* the case, condition (3.8) turns out to be necessary when extra conditions are imposed.

Let P be a Q-process with C being an irreducible class (the conditions we have imposed on Q ensure that 0 is an absorbing state which is accessible from C) and suppose that $m = (m_j,\ j \in C)$ is a finite μ-invariant measure for P. Of necessity, m will be μ-subinvariant for Q, but does (3.8) necessarily hold? Under the conditions we have imposed, the *forward integral inequalities* are satisfied (Reuter [14]); in particular,

$$p_{i0}(t) \ge \sum_{k\in C}\int_0^t p_{ik}(s)q_{k0}\,ds, \qquad i\in C. \tag{4.1}$$

On multiplying by m_i and summing over $i \in C$, we find that

$$(1-e^{-\mu t})\sum_{k\in C} m_k q_{k0} \le \mu \sum_{i\in C} m_i p_{i0}(t)\ (<\infty). \tag{4.2}$$

If we divide by μ and let $t \to 0$, we may use dominated convergence to deduce that

$$\mu\sum_{i\in C} m_i a_i \ge \sum_{i\in C} m_i q_{i0},$$

where a_i (the probability of absorption starting in state i) is given by $a_i = \lim_{t\to\infty} p_{i0}(t)$. Thus, if a_i is strictly less than 1 for some (and then all) $i \in C$, (3.8) cannot hold.

If we were to assume that P satisfies (FE_{i0}) over C, then we would have equality in (4.1) and (4.2), and so

$$\mu\sum_{i\in C} m_i a_i = \sum_{i\in C} m_i q_{i0}.$$

If instead P were assumed to be *honest*, then we would have $a_i = 1$ for all $i \in C$. This can be seen as follows. Since m is a μ-invariant measure for P, we have, in particular, that $m_i p_{ij}(t) \le e^{-\mu t} m_j$ for $i, j \in C$, and so

$$1 - p_{i0}(t) = \sum_{j \in C} p_{ij}(t) \le e^{-\mu t} \frac{1}{m_i} \sum_{j \in C} m_j, \qquad i \in C.$$

Since m is finite and $\mu > 0$, $\lim_{t \to \infty}(1 - p_{i0}(t)) = 0$, and hence $a_i = 1$ for all $i \in C$. Thus, if P were honest, we would have

$$\mu \sum_{i \in C} m_i \ge \sum_{i \in C} m_i q_{i0}.$$

Neither the honesty of P, nor an assumption that P satisfies (FE_{i0}) over C, is enough on its own to establish (3.8); it is possible to construct examples of Q-processes which illustrate this. But, these conditions *together* imply (3.8).

We have therefore proved the following variant of Theorem 3.1:

Theorem 4.1 *Let $\mu > 0$ and suppose that Q admits a finite μ-subinvariant measure. Then, there exists an honest Q-process P satisfying* (FE_{i0}) *over C for which m is μ-invariant if and only if* (3.8) *holds. The resolvent of one such Q-process is given by* (3.9).

Next we shall examine the question of *uniqueness* under the assumption that Q is a single-exit Q-matrix. This was considered briefly in Section 5 of Nair and Pollett [7] under a condition weaker than (3.8). If Q is single exit and P is an arbitrary Q-process, then (Reuter [15]) either P is the minimal Q-process or otherwise its resolvent $\Psi(\cdot) = (\psi_{ij}(\cdot),\ i, j \in S)$ must be of the form

$$\psi_{ij}(\alpha) = \phi_{ij}(\alpha) + z_i(\alpha) y_j(\alpha), \qquad i, j \in S, \tag{4.3}$$

where $y(\alpha) = (y_j(\alpha),\ j \in S)$ is given by

$$y_j(\alpha) = \frac{\eta_j(\alpha)}{c + \sum_{k \in S} \alpha \eta_k(\alpha)}, \qquad j \in S, \tag{4.4}$$

c is a non-negative constant, and $\eta(\alpha) = (\eta_j(\alpha),\ j \in S)$ is a non-negative vector which satisfies

$$\sum_{k \in S} \eta_k(\alpha) < \infty, \tag{4.5}$$

$$\eta_j(\alpha) - \eta_j(\beta) + (\alpha - \beta) \sum_{k \in S} \eta_k(\alpha) \phi_{kj}(\beta) = 0, \quad j \in S. \tag{4.6}$$

Furthermore, Ψ is honest if and only if $c = 0$. Since we have assumed that 0 is an absorbing state, $z_0(\alpha) = 0$ and so (4.6) can be written

$$\eta_j(\alpha) - \eta_j(\beta) + (\alpha - \beta) \sum_{k \in C} \eta_k(\alpha)\phi_{kj}(\beta) = 0, \quad j \in C, \tag{4.7}$$

$$\alpha\eta_0(\alpha) - \beta\eta_0(\beta) + (\alpha - \beta) \sum_{k \in C} \eta_k(\alpha)\beta\phi_{k0}(\beta) = 0. \tag{4.8}$$

Once η is determined, a family of Q-processes, exactly one of which is honest, is obtained by varying c in the range $0 \le c < \infty$. Thus, the problem of identifying those Q-processes which satisfy a specified criterion, in our case, that a given measure is μ-invariant on C, amounts to determining which choices of η and c are admissible; the procedure is purely arithmetical.

Theorem 4.2 *Suppose that Q is single exit and suppose that, for a given $\mu > 0$, Q admits a finite μ-subinvariant measure. Then, there exists an honest Q-process P satisfying* (FE_{i0}) *over C for which m is μ-invariant if and only if* (3.8) *holds. It is the unique honest Q-process for which m is μ-invariant and its resolvent is given by* (3.9).

Proof In view of Theorem 4.1, we only need to establish uniqueness. If the minimal Q-process F is honest, then it is the unique Q-process, and, as we have already observed, (3.8) is necessary and sufficient for m to be μ-invariant for F.

Suppose then that F is dishonest, so that z is not identically 0. We will prove that if there *is* an honest Q-process P for which m is μ-invariant, then its resolvent must necessarily be given by (3.9).

Let d be given by (3.10) and (3.11). Since m is μ-invariant for P, multiplying (4.3) by $(\alpha + \mu)m_i$ and summing over $i \in C$ gives

$$m_j = \sum_{i \in C} m_i(\alpha + \mu)\phi_{ij}(\alpha) + (\alpha + \mu)y_j(\alpha) \sum_{i \in C} m_i z_i(\alpha),$$

for all $j \in C$. Since P is honest, we must set $c = 0$ and so in view of (4.4) we require

$$\frac{\eta_j(\alpha)}{\sum_{k \in S} \alpha\eta_k(\alpha)} = \frac{d_j(\alpha)}{(\alpha + \mu)\sum_{i \in C} m_i z_i(\alpha)}, \quad j \in C. \tag{4.9}$$

Notice that $d_j(\alpha) > 0$ for at least one $j \in C$: since m is μ-invariant for P, m cannot be μ-invariant for F, and so (3.15) holds for at least one $j \in C$. Furthermore, by the definition of d, we have that

$$\alpha \sum_{j \in S} d_j(\alpha) = (\alpha + \mu) \sum_{j \in C} m_j z_j(\alpha) < \infty, \tag{4.10}$$

which is consistent with (4.5). From (4.9) we see that $\eta_j(\alpha) = K(\alpha)d_j(\alpha)$, at least for $j \in C$, where K is some positive scalar function. Using the identity (3.17), together with the fact that η must satisfy (4.7), we find, on substituting $\eta_j(\alpha) = K(\alpha)d_j(\alpha)$ in (4.7), that $(K(\alpha)-K(\beta))d_j(\beta) = 0$. Hence, K must be constant, because $d_j(\beta) > 0$ for at least one $j \in C$. Now, using (4.9) again, we see that K must satisfy

$$K\left((\alpha+\mu)\sum_{k\in C} m_k z_k(\alpha) - \alpha\sum_{k\in C} d_k(\alpha)\right) = \alpha\eta_0(\alpha),$$

or equivalently, by (4.10),

$$K\alpha d_0(\alpha) = \alpha\eta_0(\alpha). \tag{4.11}$$

It is clear from (3.18) that η_0 satisfies (4.8) no matter what the value of K. It is also clear that there is no loss of generality in setting $K = 1$, for this is equivalent to replacing c in (4.9) by a different constant c/K. Hence $\eta_j = d_j$ for $j \in C$, and, from (4.11), $\eta_0 = d_0$.

We have proved that if Q is single exit and P is an honest Q-process with μ-invariant measure m, then its resolvent must be given by (3.9). ∎

Acknowledgments

The author thanks the organisers for their invitation to speak at the workshop and their kind hospitality during his visit to Changsha.

References

[1] Chen, A. and Zhang, H. (1987). Criterion for the existence of reversible Q-processes. *Acta Math. Sinica*, 3: 133–142.

[2] Feller, W. (1940). On the integro-differential equations of purely discontinuous Markoff processes. *Trans. Amer. Math. Soc.*, 48: 488–515.

[3] Feller, W. (1957). On boundaries and lateral conditions for the Kolmogoroff differential equations. *Ann. Math.*, 65: 527–570.

[4] Hou, C. (1974). The criterion for uniqueness of a Q-process. *Sci. Sinica*, XVII: 141–159.

[5] Hou, C. and Chen, M. (1980). Markov processes and field theory. *Kexue Tongbao*, 25: 807–811.

[6] Kingman, J. (1963). The exponential decay of Markov transition probabilities. *Proc. London Math. Soc.*, 13: 337–358.

[7] Nair, M. and Pollett, P. (1993). On the relationship between μ-invariant measures and quasistationary distributions for continuous-time Markov chains. *Adv. Appl. Probab.*, 25: 82–102.

[8] Pollett, P. (1986). On the equivalence of μ-invariant measures for the minimal process and its q-matrix. *Stochastic Process. Appl.*, 22: 203–221.

[9] Pollett, P. (1988). Reversibility, invariance and μ-invariance. *Adv. Appl. Probab.*, 20: 600–621.

[10] Pollett, P. (1991a). Invariant measures for Q-processes when Q is not regular. *Adv. Appl. Probab.*, 23: 277–292.

[11] Pollett, P. (1991b). On the construction problem for single-exit Markov chains. *Bull. Austral. Math. Soc.*, 43: 439–450.

[12] Pollett, P. (1994). On the identification of continuous-time Markov chains with a given invariant measure. *J. Appl. Probab.*, 31: 897–910.

[13] Pollett, P. and Vere-Jones, D. (1992). A note on evanescent processes. *Austral. J. Statist.*, 34: 531–536.

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

[15] Reuter, G. (1959). Denumerable Markov processes (II). *J. London Math. Soc.*, 34: 81–91.

[16] Reuter, G. (1962). Denumerable Markov processes (III). *J. London Math. Soc.*, 37: 63–73.

[17] Reuter, G. (1967). Note on resolvents of denumerable submarkovian processes. *Z. Wahrscheinlichkeitstheorie*, 9: 16–19.

[18] Rogers, L. and Williams, D. (1986). Construction and approximation of transition matrix functions. In Kendall, D., Kingman, J., and Williams, D., editors, *Analytic and Geometric Stochastics, Papers in honour of G.E.H. Reuter*, pages 133–160. Applied Probability Trust.

[19] Tweedie, R. (1974). Some ergodic properties of the Feller minimal process. *Quart. J. Math. Oxford*, 25: 485–495.

[20] van Doorn, E. (1991). Quasi-stationary distributions and convergence to quasi-stationarity of birth-death processes. *Adv. Appl. Probab.*, 23: 683–700.

[21] Vere-Jones, D. (1967). Ergodic properties on non-negative matrices I. *Pacific J. Math.*, 22: 361–386.

[22] Williams, D. (1964). On the construction problem for Markov chains. *Z. Wahrscheinlichkeitstheorie*, 3: 227–246.

[23] Zhang, H., Lin, X., and Hou, Z. (2001). Invariant distribution of Q-process (I). *Chin. Ann. Math.*, 22A: 323–330.

[24] Zhang, H., Lin, X., and Hou, Z. (2002). Invariant distribution of Q-process (II). *Chin. Ann. Math.*, 23A (to appear).

Chapter 4

CONVERGENCE PROPERTY OF STANDARD TRANSITION FUNCTIONS*

Hanjun Zhang
Research Department, Changsha Railway University, Changsha, Hunan 410075, China

Qixiang Mei
Research Department, Changsha Railway University, Changsha, Hunan 410075, China

Xiang Lin
Research Department, Changsha Railway University, Changsha, Hunan 410075, China

Zhenting Hou
Research Department, Changsha Railway University, Changsha, Hunan 410075, China

Abstract A standard transition function $P = (p_{ij}(t))$ is called ergodic (positive recurrent) if there exists a probability measure $\pi = (\pi_i; i \in E)$ such that

$$\lim_{t \to 0} p_{ij}(t) = \pi_j > 0, \qquad \forall i \in E \tag{0.1}$$

The aim of this paper is to discuss the convergence problem in (0.1). We shall study four special types of convergence: the so-called strong ergodicity, uniform polynomial convergence, L^2-exponential ergodicity

*Supported by the National Natural Science Foundation of China (19871006)

Z. Hou et al. (eds.), Markov Processes and Controlled Markov Chains, 57–67.

and exponential ergodicity. Our main interest is always to characterize these properties in terms of the q-matrix.

Keywords: strong ergodicity, uniform polynomial convergence, L^2-exponential convergence, exponential ergodicity, stochastic monotonicity

AMS Subject Classification(1991): 60J27

1. Introduction

Let E be a countable set, to be called the state space, and $P = (p_{ij}(t); i,j \in E,\ t \geq 0)$ be a standard transition function with stationary distribution $\pi = (\pi_i; i \in E)$. Ordinally, we have the following definitions for convergence:

1. Strong Ergodicity:

$$\sup_{i\in E} \sum_{j\in E} |p_{ij}(t) - \pi_j| \longrightarrow 0, \qquad t\to\infty \tag{1.1}$$

2. Uniform Polynomial Convergence: If there exists constants $C > 0,\ v > 0$ such that

$$\sup_{i,j\in E} t^v |p_{ij}(t) - \pi_j| \leq C < +\infty, \quad t \geq 0 \tag{1.2}$$

3. L^2-exponential Convergence: If there exists a constant $v > 0$ such that

$$||P(t)f - \pi(f)|| \leq e^{-vt}||f - \pi(f)||, \quad f \in L^2(\pi), t \geq 0 \tag{1.3}$$

   where

$$\begin{aligned} P(t)f &= \sum_{j\in E} p_{ij}(t) f_j \\ \pi(f) &= \sum_{i\in E} \pi_i f_i \\ ||f||^2 &= \sum_{i\in E} \pi_i f_i^2 \\ L^2(\pi) &= \{f : ||f|| < \infty\} \end{aligned}$$

4. Exponential Ergodicity: If there exist $v > 0$ and $C_{ij} > 0$ such that

$$|p_{ij}(t) - \pi_j| \leq C_{ij} e^{-vt}, \qquad \forall t \geq 0,\ i,\ j \in E \tag{1.4}$$

It is well-known that

$$\begin{aligned}\text{Strong ergodicity} &\implies \text{Uniform polynomial convergence} &(1.5)\\ \text{Strong ergodicity} &\implies L^2\text{-exponential convergence} &\\ &\implies \text{Exponential ergodicity} &(1.6)\end{aligned}$$

To study the convergence is an important topic in Markov processes, in particular, in the study of continuous time Markov chains (CTMC) and interacting particle systems. Good references are, among others, Chung [8], Hou and Guo [10], Yang [15] and Anderson [1] for the former and Liggett [11] and Chen [6] for the latter.

The main purpose of this paper is to consider the above convergence property for an important class of transition function, the stochastically monotone function. The close link among convergence, stochastically monotone function and another key concept, Feller-Reuter-Riley function is revealed.

For simplicity, we shall consider CTMC exclusively in this paper. Also, in most of cases, the state space E will be always assumed to be $Z_+ = \{0, 1, 2, \ldots\}$ with natural order. The monotonicity may then be simply defined as follows

Definition 1.1 *A standard transition function $P = (p_{ij}(t); i, j \in E, t \geq 0)$ is called stochastically monotone, if for any fixed $k \in Z_+$ and $t \geq 0$, $\sum_{j \geq k} p_{ij}(t)$ is a non-decreasing function of i, i.e.,*

$$\sum_{j\geq k} p_{ij}(t) \;\leq\; \sum_{j\geq k} p_{i+1,j}(t), \qquad k \in Z_+,\ t \geq 0 \tag{1.7}$$

Also due to the fact that the state space Z_+ is linear ordered, stochastic monotonicity is equivalent to another important concept, duality, as was revealed by Siegmund [13].

Proposition 1.1 (Siegmund [13]) *A standard transition function P is stochastically monotone if and only if there exists another standard transition function $\tilde{P} = (\tilde{p}_{ij}(t); i, j \in Z_+,\ t \geq 0)$, such that*

$$\sum_{k=0}^{j} \tilde{p}_{ik}(t) \;=\; \sum_{k=i}^{\infty} p_{jk}(t), \qquad \forall i, j \in Z_+,\ t \geq 0 \tag{1.8}$$

For the proof of Proposition 1.1, see Siegmund [13] or Anderson [1].

Definition 1.2 *A standard transition function P on Z_+ is called a Feller-Reuter-Riley transition function (henceforth referring to a FRR function) if*

$$\lim_{i\to\infty} p_{ij}(t) = 0, \qquad \forall j \in Z_+,\ t \geq 0 \tag{1.9}$$

For the interesting property of FRR functions, see Reuter-Riley [12].

Note that a transition function $P = (p_{ij}(t))$ is called standard if $\lim_{t\to 0} p_{ij}(t) = \delta_{ij}$, see Chung [8]. In this paper all transition functions are assumed to be standard and thus from now on, the modifier "standard" will be omitted.

The results of strong ergodicity and the uniform polynomial convergence, together with their proof, obtained in this paper are presented in Section 2. In Section 3, we have obtained the sufficient and necessary condition of exponential ergodicity of quadratic branching Q-processes.

2. Strong ergodicity and uniform polynomial convergence of monotone q-functions

We are now ready to state our main results and give their proofs.

Theorem 2.1 *If $P = (p_{ij}(t); i, j \in E,\ t \geq 0)$ is an FRR transition function, then it is neither strongly ergodic and nor uniformly polynomial convergent.*

Proof By (1.5), we need only to prove that if P is an FRR transition function, then it is not uniform polynomially convergent.

We may assume that the FRR transition function P is ergodic since otherwise nothing need to be proven. Since P is ergodic, it possesses an equilibrium distribution $\pi = (\pi_i; i \in E)$, say. Now suppose that P is uniformly polynomial convergent, by (1.2), there exist constants $C > 0$ and $v > 0$ such that

$$\sup_{i,j\in E} |p_{ij}(t) - \pi_j| \leq Ct^{-v}, \qquad t \geq 0 \tag{2.1}$$

Now, fix a state $j_0 \in E$, then (2.1) implies that there exists a $T < +\infty$ such that

$$\sup_{i\in E} |p_{ij_0}(t) - \pi_{j_0}| \leq \frac{1}{2}\pi_{j_0}, \qquad \forall t \geq T$$

hence

$$\inf_{i\in E} p_{ij_0}(t) \geq \frac{1}{2}\pi_{j_0}, \qquad \forall t \geq T$$

which contradicts the requirement of an FRR transition function since for an FRR transition function we have $p_{ij}(t) \to 0$, as $i \to \infty$ for each $j \in E$ and any $t \geq 0$. ■

Interestingly, for monotone transition functions, the converse of Theorem 2.1 also holds true. This is the following important conclusion.

Theorem 2.2 *Suppose* $P = (p_{ij}(t); i,j \in E,\ t \geq 0)$ *is an honest and stochastically monotone transition function. If* P *is ergodic, then it is strongly ergodic (or uniformly polynomial convergent) if and only if it is not an FRR transition function.*

Proof By the conclusion of Theorem 2.1 and (1.5), in order to finish the proof, we need only to prove that if an ergodic monotone function is not strong ergodic, then it is a FRR function. Now suppose $P = (p_{ij}(t); i,j \in E, t \geq 0)$ is an ergodic monotone function, then by writing $p_{ij}(t)$ as

$$p_{ij}(t) = \sum_{k=j}^{\infty} p_{ik}(t) - \sum_{k=j+1}^{\infty} p_{ik}(t) \tag{2.2}$$

we see that for any $j \in E$ and $t \geq 0$, the following limit, denoted by $c_j(t)$, exists

$$\lim_{i\to\infty} p_{ij}(t) = c_j(t) \tag{2.3}$$

since, by monotonicity, both terms in the right hand side of (2.2) are monotone function of i.

On the other hand, since P is not strongly ergodic, by Anderson [1, Proposition 6.3.1], we have, for any $t > 0$

$$\inf_{a,b\in E} \sum_{k\in E} (p_{ak}(t) \wedge p_{bk}(t)) = 0$$

which trivially implies that for any $j \in E$ and $t \geq 0$,

$$\inf_{a,b\in E} p_{aj}(t) \wedge p_{bj}(t) = 0 \tag{2.4}$$

but it is easy to see that (2.4) is equivalent to

$$\inf_{a\in E} p_{aj}(t) = 0. \tag{2.5}$$

Combining (2.3) with (2.5) shows that for any $t > 0$ and $j \in E$,

$$\lim_{i\to\infty} p_{ij}(t) = 0$$

i.e., $P = \{p_{ij}(t)\}$ is a FRR transition function. ∎

As a direct consequence, we immediately obtain the following useful and interesting result.

Corollary 2.1 *Suppose P is a non-minimal stochastically monotone transition function. Then P is strongly ergodic (or uniformly polynomial convergent) if and only if it is ergodic.*

Proof Note that FRR transition function must be Feller minimal, see Reuter and Riley [12], the conclusion immediately follows from Theorem 2.2.

■

As always the case in the study of CTMC, our main interest is to characterize convergence in terms of the inifinitesimal behaviour, i.e., the q-matrix. The following result answers this question satisfactorily for an important class of q-matrices.

Theorem 2.3 *Suppose the given q-matrix $Q = (q_{ij}; i, j \in E)$ is stable, conservative and monotone, i.e.,*

$$0 \leq q_{ij}, \qquad i \neq j$$

$$\sum_{j \neq i} q_{ij} = -q_{ii} < +\infty, \qquad i \in E \tag{2.6}$$

$$\sum_{j \geq k} q_{ij} \leq \sum_{j \geq k} q_{i+1,j}, \qquad k \neq i+1 \tag{2.7}$$

then

(i) *The minimal Feller Q-function is strongly ergodic (uniformly polynomial convergent) if and only if it is ergodic and at least one of the following two conditions hold true:*

a) *Q is not zero-entrance, i.e., the equation*

$$Y(\lambda I - Q) = 0, \qquad 0 \leq Y, \ Y1 < +\infty \tag{2.8}$$

has a non-zero solution for some (and therefore for all) $\lambda > 0$.

b) *Q is not an FRR q-matrix.*

(ii) *If $P = (p_{ij}(t); i, j \in E, \ t \geq 0)$ is a non-minimal stochastically monotone Q-function, then it is strongly ergodic (uniformly polynomial convergent) if and only if it is ergodic.*

Proof

- For part (ii), see Corollary 2.1.
- We now prove part (i).

  Suppose that Feller minimal Q-function is strongly ergodic (uniformly polynomial convergent), then it is surely ergodic and thus honest. It then implies the given q-matrix Q is regular, considering the Q is conservative. However, the given Q is also monotone by the assumption and thus by Chen and Zhang [4, Theorem 2.4], the Feller minimal Q-function is stochastically monotone. Thus by Theorem 2.2, the Feller minimal Q-function is not a FRR transition function. Now the conclusion follows from Zhang and Chen [16, Theorem 5.1] since our q-matrix Q is conservative, regular and monotone. ■

Remark 2.1 *A q-matrix Q is called a FRR q-matrix if*

$$q_{ij} \longrightarrow 0, \qquad \text{as } i \to \infty \text{ for every } j \tag{2.9}$$

so condition (b) *in* Theorem 2.3 *is quite easy to check.*

Remark 2.2 *The results about strong ergodicity and their application in birth and death processes can be seen in Zhang, Chen, Lin and Hou* [17].

3. Exponential ergodicity of quadratic branching Q-processes

Branching processes form one of the classical fields of probability theory and have a very wide range of applications. There are several specialized books devoted to this subject (see [2, 3, 9], for instance). On the other hand, the dual of a measure-valued process often leads to a modified model of the branching processes. For instance, the following model comes from a typical measure-valued process (the Fleming-Viot process), which was introduced to us by D.A. Dawson. The given q-matrix $Q = (q_{ij}; i, j \in E)$ is as follows:

$$q_{ij} = \begin{cases} i^2 p_{j-i+1}, & j \geq i-1,\ j \neq i \\ -i^2(1-p_1), & j = i \geq 1 \\ p_i, & j \geq i = 0 \\ p_0 - 1, & j = i = 0 \\ 0, & \text{elsewhere, } i, j \in Z_+ \end{cases} \tag{3.1}$$

where $P = (p_j; j \in E)$ is a probability distribution. This q-matrix is called a quadratic branching q-matrix.

Chen [7] discussed an extended class of q-matrices, that is,

$$q_{ij} = \begin{cases} r_i p_{j-i+1}, & j \geq i-1 \geq 0,\ j \neq i \\ -r_i(1-p_1), & j = i \geq 1 \\ q_{0j}, & j > i = 0 \\ -q_0, & j = i = 0 \\ 0, & \text{elsewhere, } i,j \in Z_+ \end{cases} \tag{3.2}$$

where $r_i > 0, i \geq 1$.

Chen obtained the criteria for the uniqueness, recurrence and positive recurrence (ergodicity) of the Q-processes. For quadratic branching q-matrix, Chen's results are as follows:

Theorem 3.1 (Chen [7]) *Suppose $Q = (q_{ij}; i,j \in E)$ be a quadratic branching q-matrix.*

(i) The Q-process is unique if and only if $M_1 = \sum\limits_{j=1}^{\infty} jp_j \leq 1$.

(ii) If $p_0 > 0$ and $p_k > 0$ for some $k \geq 2$, then the Feller minimal Q-process is recurrent if and only if $M_1 \leq 1$ and it is ergodic if and only if $M_1 \leq 1$.

(iii) The Feller minimal Q-process is exponentially ergodic if $M_1 < 1$.

Our main results are as follows:

Theorem 3.2 *Let $Q = (q_{ij}; i,j \in E)$ be a quadratic branching q-matrix with $p_0 > 0$ and $p_k > 0$ for some $k \geq 2$, then*

(i) The minimal Q-process is strongly ergodic if $M_1 < 1$.

(ii) The minimal Q-process is uniformly polynomial convergent if $M_1 < 1$.

(iii) The minimal Q-process is exponentially ergodic if and only if $M_1 \leq 1$.

Proof

- The proof of (i) and (ii) can be found in Zhang, Chen, Lin and Hou [18].
- Now we prove (iii). By Theorem 3.1, we need only to prove that if $M_1 = 1$, then the minimal Q-process is exponentially ergodic.

  Since

$$\sum_{k=0}^{\infty} p_k = 1 \text{ and } M_1 = \sum_{k=1}^{\infty} kp_k = 1,$$

we have

$$0 \ < \ p_0 = T = \sum_{k=1}^{\infty} k p_{k+1} < 1. \tag{3.3}$$

First, we show that there exists $\delta > 0$ such that the inequality

$$i^2(y_{i+1} - y_i) + i^2(y_{i-1} - y_i) + 1 + \delta y_i \leq 0, \quad i \geq 1 \tag{3.4}$$

has a finite solution.

In fact, let $y_0 = 0$, $y_i - y_{i-1} = \frac{8}{\sqrt{i}}$, $\delta = \frac{1}{16}$ then

$$y_n = \sum_{i=1}^{n}(y_i - y_{i-1}) = \sum_{i=1}^{n} \frac{8}{\sqrt{i}} \leq \int_0^n \frac{8}{\sqrt{x}}\, dx = 16\sqrt{n}.$$

Let

$$\begin{aligned} f(i) &= y_{i+1} - y_i - (y_i - y_{i-1}) \\ &= 8\left(\frac{1}{\sqrt{i+1}} - \frac{1}{\sqrt{i}}\right) \\ &= -\frac{8}{\sqrt{i}\sqrt{i+1}\,(\sqrt{i+1} + \sqrt{i})}, \end{aligned}$$

then

$$\frac{-4}{i\sqrt{i+1}} < f(i) < \frac{-4}{(i+1)\sqrt{i}}$$

and

$$\begin{aligned} f(i) + \frac{1}{i^2} + \frac{\delta y_i}{i^2} &\leq f(i) + \frac{1}{i^2} + \frac{1}{i\sqrt{i}} \\ &< \frac{-4}{(i+1)\sqrt{i}} + \frac{1}{i^2} + \frac{1}{i\sqrt{i}} \\ &\leq 2\left[\frac{1}{\sqrt{i}}\left(\frac{1-i}{(i+1)i}\right)\right] \\ &\leq 0, \end{aligned}$$

so equation (3.4) has a finite solution $y = (y_i; i \in Z_+)$.

Set $\bar{y}_n = \frac{1}{p_0} y_n$, $\bar{\delta} = p_0 \delta$, then the following inequalities hold

$$i^2 p_0(\bar{y}_{i+1} - \bar{y}_i) + i^2 p_0(\bar{y}_{i-1} - \bar{y}_i) + \bar{\delta}\bar{y}_i + 1 \leq 0, \ \forall i \geq 1 \tag{3.5}$$

so for $i \geq 1$

$$\begin{aligned}
&\sum_{j=0}^{\infty} q_{ij}(\bar{y}_j - \bar{y}_i) + \bar{\delta}\bar{y}_i + 1 \\
&= q_{i,i-1}(\bar{y}_{i-1} - \bar{y}_i) + \sum_{j\geq i+1} q_{ij}(\bar{y}_j - \bar{y}_i) + \bar{\delta}\bar{y}_i + 1 \\
&= i^2 p_0(\bar{y}_{i-1} - \bar{y}_i) + \sum_{j\geq i+1} i^2 p_{j-i+1} \sum_{k=i+1}^{j} (\bar{y}_k - \bar{y}_{k-1}) + \bar{\delta}\bar{y}_i + 1 \\
&\leq i^2 p_0(\bar{y}_{i-1} - \bar{y}_i) + \sum_{j\geq i+1} i^2 p_{j-i+1}(j-i)(\bar{y}_{i+1} - \bar{y}_i) + \bar{\delta}\bar{y}_i + 1 \\
&= i^2 p_0(\bar{y}_{i-1} - \bar{y}_i) + i^2(\bar{y}_{i+1} - \bar{y}_i)\sum_{k=1}^{\infty} k p_{k+1} + \bar{\delta}\bar{y}_i + 1 \\
&= i^2 p_0(\bar{y}_{i-1} - \bar{y}_i) + i^2 p_0(\bar{y}_{i+1} - \bar{y}_i) + \bar{\delta}\bar{y}_i + 1 \\
&\leq 0 \qquad (3.6)
\end{aligned}$$

Let $0 < \delta^* < min\{1-p_0, 1-p_1, \bar{\delta}\}$, then $\delta^* < \bar{\delta}$. By (3.6), we have

$$\sum_{j=0}^{\infty} q_{ij}(\bar{y}_j - \bar{y}_i) + \delta^* \bar{y}_i + 1 \leq 0, \qquad i \geq 1$$

and

$$\sum_{i\geq 1} q_{0i}\bar{y}_i = \sum_{i\geq 1} p_i \bar{y}_i \leq 16 \sum_{i\geq 1} p_i \sqrt{i} \leq 16 < +\infty \qquad (3.7)$$

Hence the minimal quadratic branching Q-process is exponentially ergodic, see Tweedie [14] or Anderson [1, Theorem 6.5]. ∎

Remark 3.1 *By* Theorem 3.1 *and* 3.2, *$M_1 = 1$ is a critical value, so to study the case of $M_1 = 1$ is very interesting.*

References

[1] W.J. Anderson (1991) *Continuous-Time Markov Chain.* Springer, Berlin.

[2] S. Asmussen and H. Hering (1983) Branching Processes. Birkhauser, Basel.

[3] K.B. Athreya and P.E. Neg (1972) Branching Processes. Springer, Berlin.

[4] A.Y. Chen and H.J. Zhang (1999a) Existence and Uniqueness of Stochastically Monotone Q-Processes. To appear in Southeast Asian Bulletin of Mathematics. 23:559-583

[5] A. Y. Chen and H.J. Zhang (1999b) Stochastic Monotonicity and Duality for Continuous Time Markov Chains with General Q-matrix. Southeast Asian Bulletin of Mathematics. 23: 383-408.

[6] M.F. Chen (1992) From Markov Chains to Non-Equilibrium Particle System. World Scientific, Singapore.

[7] R.R. Chen (1997) An Extended Class of Time-Continuous Branching Processes. J. Appl. Prob. 34, 14-23.

[8] K.L. Chung (1967) Markov Chains with Stationary Transition Probabilities. Springer-Verlag, Berlin/New York.

[9] T.E. Harris (1963) Branching Processes. Springer, Berlin.

[10] Z.T. Hou and Q.F. Guo (1988) Time-Homogeneous Markov Processes with Countable State Space. Springer-Verlag, New York.

[11] T.M. Liggett (1985) Interacting Particle System. Springer-Verlag, Berlin/New York.

[12] Reuter and Riley (1972) The Feller Property for Markov Semigroups on a Countable State Space. J. London Math. Soc.(2), 267-275.

[13] D. Siegmund (1976) The Equivalence of Absorbing and Reflecting Barrier Problems for Stochastically Monotone Markov Process. Ann. Prob. 4, No. 6, 914-924.

[14] R.L. Tweedie (1981) Criteria for Ergodicity, Exponential and Strong Ergodicity of Markov Processes. J. Appl. Prob. 18, 122-130.

[15] X.Q. Yang (1990) The Construction Theory of Denumerable Markov Processes. Wiley, New York.

[16] H.J. Zhang and A.Y. Chen (1999) Stochastic Comparability and Dual Q-Function. Journal of Mathematical Analysis and Applications 234, 482-499.

[17] H.J. Zhang, A.Y. Chen, X. Lin and Z.T. Hou (1999a) The Strong Ergodicity of Monotone q-Functions. Submitted.

[18] H.J. Zhang, A.Y. Chen, X. Lin and Z.T. Hou (1999b) Quadratic Branching Q-Matrix. Submitted.

Chapter 5

MARKOV SKELETON PROCESSES*

Hou Zhenting
Research Department,
Changsha Railway University,
Changsha,Hunan 410075,
P.R. China

Liu Zaiming
Research Department,
Changsha Railway University,
Changsha,Hunan 410075,
P.R. China

Zou Jiezhong
Research Department,
Changsha Railway University,
Changsha,Hunan 410075,
P.R. China

Chen Xuerong
Research Department,
Changsha Railway University,
Changsha,Hunan 410075,
P.R. China

Abstract In this paper, we introduce a new class of stochastic processes - Markov skeleton processes, which have the Markov property at a series of random times. Markov skeleton processes include minimal Q processes,

*Supported by the National Natural Science Foundation of China (19871006)

Z. Hou et al. (eds.), Markov Processes and Controlled Markov Chains, 69–92.

Doob processes, Q processes of order one, Markov processes, semi-Markov processes, piecewise determinate Markov processes, the input processes, the queuing lengths and the waiting times of the system GI/G/1, the insurance risk models, and the option pricing models, as particular cases. The present paper aims to fully expound the background and the history source of the introduction of Markov skeleton processes, and we deduce the forward and backward equation and use them as a powerful tool to obtain the criteria of regularity.

Keywords: Markov skeleton processes; the backward equations; the forward equations; the criteria for the regularity.

AMS Subject Classification(1991): 60J

1. Introduction

Markov processes are obviously of great importance. They have the Markov property at any constant stopping time (i.e. usual time). After further study in Markov processes, it was found that most of the Markov processes have the strong Markov property. The corresponding subclass of Markov processes is very rich . One may say that the research on Markov processes in fact deals with the strong Markov processes. Strong Markov processes have the Markov property at any stopping time. In actuality, it is not easy to determine whether a stochastic process is a Markov process or strong Markov process. Of course, many stochastic processes do not have the Markov property. However, there are many processes $\{X_t, t < \tau\}$ that are not (strong) Markovian, but there is a sequence of stopping times: $0 \equiv \tau_0 \leq \tau_1 \leq \tau_2 \leq \ldots \leq \tau_n \uparrow \tau$, such that the process $\{x_t\}$ has the Markov property at τ_n, $n \geq 0$. We call this property of $\{x_t\}$ or $\{\tau_n\}$ the property (H). The following are some examples:

Example 1.1

Ex 1. *Let $\{x_t, t < +\infty\}$ be a Markov process, set $\tau_n = n$, $n = 0, 1, \ldots$. Then $\{\tau_n, n \geq 0\}$ has property (H).*

Ex 2. *Let $\{x_t, t < \tau\}$ be the minimal homogeneous denumerable Markov process* [1], Figure 2, *denote the n^{th} jump point by τ_n, then $\tau_n \uparrow \tau$ and $\{\tau_n, n \geq 0\}$ is of property (H).*

Ex 3. *Let $\{x_t, t < \tau\}$ be a Doob process* [1], Figure 3, *and denote the n^{th} explosion of $\{x_t\}$ by τ_n, then $\tau_n \uparrow \tau$ and $\{\tau_n, n \geq 0\}$ has property (H), and the $\{X(\tau_n), n \geq 1\}$ has the same distribution.*

Ex 4. *Let $\{x_t, t < \tau\}$ be a Q-process of order one* [1], Figure 3, *and denote the n^{th} flying point of x_t by τ_n, then $\tau_n \uparrow \tau$ and $\{\tau_n, n \geq 0\}$ has property (H).*

Ex 5. *Let $\{N(t), t \geq 0\}$ be an input process to a GI/G/1 queue* [3]. *That is to say, $N(t)$ stands for the number of arrivals up to time t. Let $\tau_0 \equiv 0$, τ_n, $n \geq 1$ denote the arrival time of the n^{th} customer. Then $\tau_n \uparrow +\infty$, and it is easy to see that $\{\tau_n, n \geq 0\}$ has property (H), but $N(t)$ is not a Markov process unless $\tau_{n+1} - \tau_n$, $n \geq 0$ are independent and have exponential distributions.*

Ex 6. *Let $\{L(t), t \geq 0\}$ be a M/G/1 queuing process* [3]. *$L(t)$ stands for the queuing length at t, $\tau_0 \equiv 0$, τ_n, $n \geq 1$ denotes the exit time of the n^{th} customer. Then $\tau_n \uparrow \infty$, and $\{\tau_n, n \geq 0\}$ has property (H).*

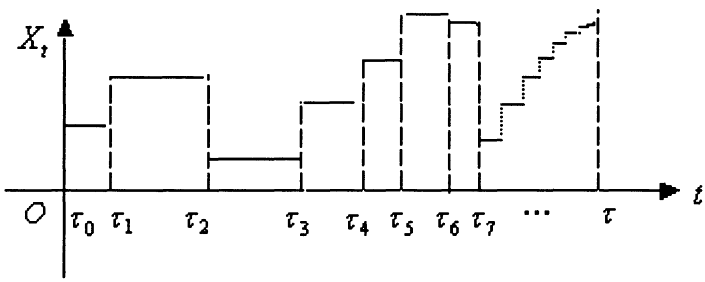


Figure 5.1. Minimal homogeneous denumerable Markov process

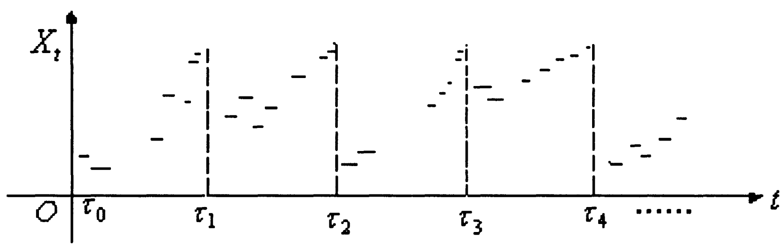


Figure 5.2. Doob process

Ex 7. Let $\{L(t), t \geq 0\}$ *be a GI/M/1 queuing process* [3]. $\tau_0 \equiv 0$, τ_n, $n \geq 1$ *stands for the arrival time of the* n^{th} *customer. Then* $\tau_n \uparrow \infty$, *and* $\{\tau_n, n \geq 0\}$ *has property* (H).

Ex 8. Let $\{L(t), t \geq 0\}$ *be a GI/G/1 queuing process* [3]. $\tau_0 \equiv 0$, τ_n, $(n \geq 1)$ *stands for the starting time of the* n^{th} *busy period. Then* $\tau_n \uparrow \infty$, *and* $\{\tau_n, n \geq 0\}$ *is of property* (H).

Ex 9. Let $\{W(t), t \geq 0\}$ *be a waiting process of a GI/G/1 queue* [3], *i.e.* $W(t)$ *stands for the waiting time of the customer who arrives at* t. $\tau_0 \equiv 0$, τ_n, $(n \geq 1)$ *denotes the arrival time of the* n^{th} *customer. Then* $\tau_n \uparrow \infty (n \uparrow \infty)$ *and* $\{\tau_n, n \geq 0\}$ *has property* (H). *See* Figure 9.

Ex 10. Risk decision model, Figure 10.

$$u_1(t) = u + ct - \sum_{i=1}^{N(t)} x_i$$

where $N(t)$ *is the number of claims occurred in* $[0, t]$, *and* x_i*'s are positive random variables.* $u(t)$ *has the property* (H).

Ex 11. Risk decision model with random disturbance.

$$u(t) = u + ct + W(t) - \sum_{i=1}^{N(t)} x_i$$

where $W(t)$ *stands for Brownian motion.* $u(t)$ *has the property* (H).

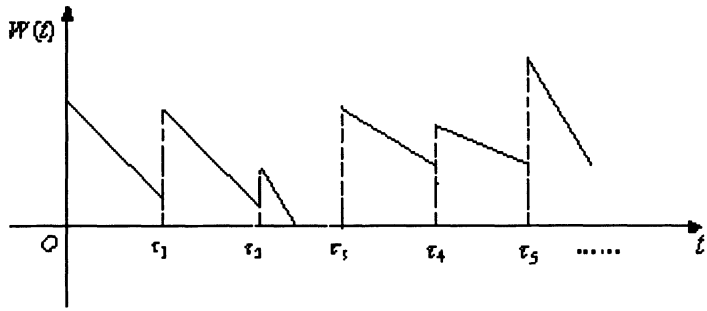


Figure 5.3. Waiting process

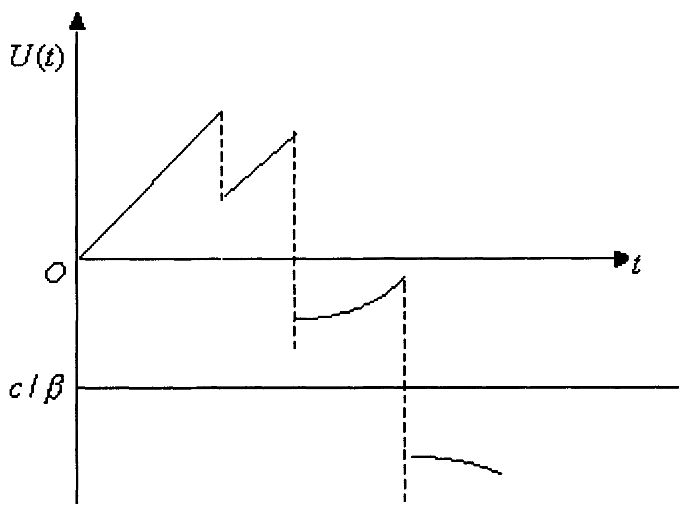


Figure 5.4. Risk decision model

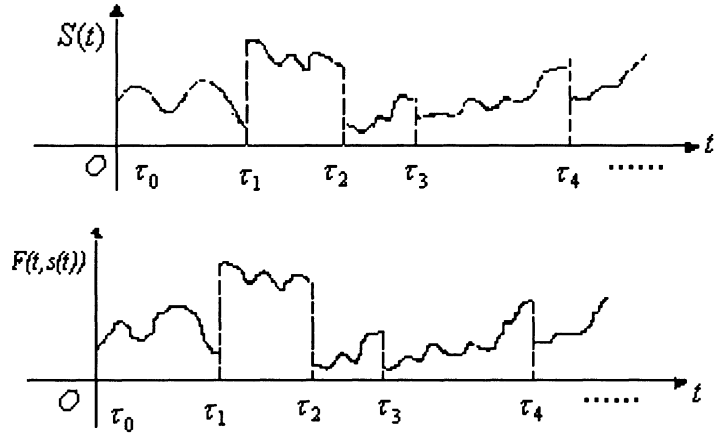


Figure 5.5. Option pricing model

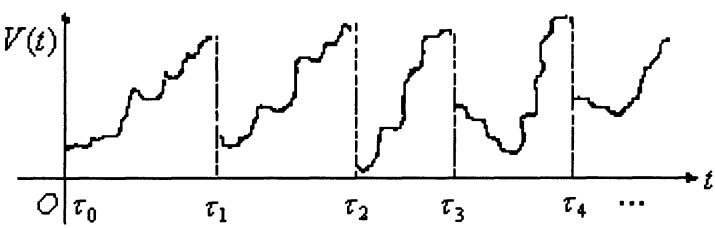


Figure 5.6. Reservoir model

Ex 12. Option pricing model, Figure 12. *The* $S(t)$ *denotes the price of stock,* $F[t, S(t)]$ *is a dualistic continuous function of* t, s *and stands for the option price at* t. $S(t)$ *and* $F(t, S(t))$ *have the property* (H).

Ex 13. Reservoir model, Figure 13. $V(t)$ *has the property* (H).

In fact, from the examples above, we see that many processes in practice have the property (H) but are usually not Markovian. From the study of Markov processes, it is easy to see that many results hold for those processes that only have the property (H). For example, the minimal Q-process satisfies both Kolmogorov backward and forward equations; the transition probability of a minimal Q-process satisfies both Kolmogorov backward and forward equations; the distributions and moments of the first arrival time and integral-type functional for minimal Q-processes and order 1 Q-processes are minimal nonnegative solutions of some nonnegative linear equations; and so on. All these results can be derived by the property (H) only. For these reasons, we consider the processes with the property (H) for a separate study and we call them Markov skeleton processes.

Next, we review the history from the introduction of Markov processes to that of Markov skeleton processes.

Markov chains, the original models of Markov processes were introduced in 1906 by the Russian mathematician A.A. Markov [30]. From then on, many scholars began their continuous investigation on Markov processes and many excellent results have been obtained, especially for strong Markov processes. The simplest one which we have studied in great detail is the minimal Markov chains $\{x(t,\omega), t < \tau\}$, Figure 2.

$0 \equiv \tau_0 \leq \tau_1 \leq \tau_2 \cdots \leq \tau_n \leq \cdots$, $\tau_n \uparrow \tau$, τ_i, $i = 1, 2, \cdots$ is the i^{th} jump point, τ is the explosion point. As we know,

(i) τ_n, $n = 0, 1, \cdots$ has the property (H);

(ii) $x(t) = x(\tau_n)$, $\tau_n \leq t < \tau_{n+1}$, $n = 0, 1, 2, \cdots$;

(iii) The distribution of $\tau_{n+1} - \tau_n$ is an exponential distribution depending on X_{τ_n}. i.e.

$$P(\tau_{n+1} - \tau_n \leq t | X_{\tau_n} = i) = \begin{cases} 1 - e^{-q_i t} & t \geq 0 \\ 0 & t \leq 0 \end{cases}$$

Conversely, if the above three conditions hold, then X(t) must be a Markov process.

But there are many stochastic processes, which have properties (i) and (ii) and do not have property (iii). In 1955, Levy [3] and some other authors gave up the property (iii), but they kept the properties (i) and (ii), and introduced the concept of semi-Markov processes to take up the study.

Until the 1980's, M.H.A. Davis [32, 33] relaxed the property (ii): the hypothesis that $x(t)$ took only one constant on $[\tau_n, \tau_{n+1})$ was replaced by taking a determinate smooth curve, but kept the Markov property on the jump time. Then with the aid of an auxiliary variable, he introduced the concept of piecewise deterministic Markov space, and obtained extended infinitesimal generators of this kind of processes. The above Examples Ex 9 and Ex 10 are typical examples of this kind. But many other important stochastic processes are still out of consideration. In general, Ex 9 and Ex 10 are not Markov processes, let alone piecewise deterministic Markov processes as defined by Davis. There are also some stochastic processes with property (H), or the above property (I), their paths between two adjacent Markov time τ_n and τ_{n+1} are pieces of deterministic smooth curves but pieces of stochastic processes. For example, Ex 3, Ex 4, Ex 6, Ex 7, Ex 8, Ex 11, Ex 12 and Ex 13 are typical models of such kind. Ex 3 and Ex 4 are Markov processes, but generally the other six examples are not Markov processes. As for Ex 6 and Ex 7, in the late 1950s, D.G. Kendall [6, 7] for the first time noticed that $L(\tau_n)$, $n \geq 0$ forms a Markov chain. L. Takaces [1] studied $L(\tau_n)$ in Ex 6, and used the property (H) to obtain the explicit expression of the generating function for the Laplace transform of the probability distribution of $L(t)$ and the expression of the stationary distribution. Later, Wu Fang [9], U.N. Bhat [11] obtained the same results for Ex 7 by the same method. In 1997, based on the results of these scholars and by laying an emphasis on the common character (H) of the above examples, Hou Zhenting, Liu Zaiming, Zou Jiezhong introduced the concept of Markov skeleton process (MSP), and obtained the backward and forward equations satisfying the probability distribution of Markov skeleton process. In recent years, we have carried out a basal study of the theories and applications of MSP, and thus built up the theoretic framework.

As we know the development process of everything is a repeat and alternate process which includes changes in quantity and quality. It makes a fresh start at a series of time when quality changes occur (i.e. of property H). The jump processes characterise the change of quality while diffusion processes characterise the change of quantity. In other words, at the moments when quality changes occur, things have the Markov property. It was for this reason that we introduced Markov skeleton process to provide an appropriate model for the study of these mixed stochastic processes. A wide application of such processes in queueing system, deposit system, reservoir management system, insurance and finance system, economy system, demography theory models and economic market can be found.

2. Definition and properties of Markov skeleton processes

Let $(\Omega, \mathcal{F}, P)$ be a complete probability space, $(E, \mathcal{E})$ a measurable space, $\{\mathcal{F}_t, t \geq 0\}$ a flow of σ-algebras of $\mathcal{F}$. Then $X = \{x(t,\omega), 0 \leq t < \tau(\omega)\}$ is $\mathcal{F}_t$-adaptive stochastic process defined on $(\Omega, \mathcal{F}, P)$ with values in $(E, \mathcal{E})$.

For convenience, we extend the state space E to $\hat{E} = E \cup \{b\}$ by adding an isolated state b to E. The process X is also extended to $\hat{X} = \{\hat{x}(t,\omega), 0 \leq t < \infty\}$, by

$$\hat{x}(t,\omega) = \begin{cases} x(t,\omega), & 0 \leq t < \tau(\omega), \\ b, & \tau(\omega) \leq t < \infty. \end{cases} \tag{2.1}$$

Definition 2.1 *The stochastic process $X = \{X(t,\omega), 0 \leq t < \tau)\}$ is called a process with Markov skeleton if there exists a sequence of stopping times $\{\tau_n\}_{n \geq 0}$, satisfying*

(i) $\tau_n \uparrow \tau$ with $\tau_0 = 0$, and for each $n \geq 0$, $\tau_n < \tau \Longrightarrow \tau_n < \tau_{n+1}$;

(ii) for every τ_n and any bounded $\hat{E}^{[0,\infty)}$-measurable function f defined on $\hat{\mathcal{E}}^{[0,\infty)}$

$$E[f(\hat{x}(\tau_n + \cdot, \omega))|\mathcal{F}_{\tau_n}] = E[f(\hat{x}(\tau_n + \cdot, \omega))|\hat{x}(\tau_n)] \qquad P\text{-a.e. on } \Omega_{\tau_n}, \tag{2.2}$$

where $\Omega_{\tau_n} = (\omega : \tau_n(\omega) < \infty)$, and

$$N_{\tau_n} \hat{=} \{A : \forall t \geq 0, A \cap (\omega : \tau_n(\omega) \leq t) \in \sigma\{\hat{x}_s, 0 \leq s \leq t\}\}$$

is the σ-algebra on Ω_{τ_n}.

We say that X is a homogeneous Markov skeleton process if the following equation holds in (ii)

$$\begin{aligned} E[f(\hat{x}(\tau_n + \cdot, \omega))|N_{\tau_n}] &= E(f(\hat{x}(\tau_n + \cdot, \omega))|\hat{x}_{\tau_n}] \\ &= E_{\hat{x}(\tau_n)}[f(\hat{x}(\cdot, \omega))], \quad \text{P-a.e. on } \Omega_{\tau_n}, \end{aligned} \tag{2.3}$$

where $E_{\hat{x}}(\cdot)$ denotes the expectation corresponding to $P(\cdot|\hat{x}(0) = \hat{x})$.

Remark 2.1 *In this article, suppose E to be a Polish space, $\mathcal{E}$ the Borel σ-algebra, and Ω to be a space of right-continuous of functions, defined on R_+ with values in E.*

Consider a right-continuous stochastic process $X = \{X(t,\omega), 0 \leq t < \tau(\omega)\}$, defined on $(\Omega, \mathcal{F}, P)$ with values in E. Because E is a Polish space, the Kolmogorov existence theorem assures that the above restriction of the Ω is reasonable. Let $\mathcal{F}^\theta = \{\mathcal{F}_t^\theta, t \geq 0\}$ be the σ-algebra flow generated by the process X initially, where $\mathcal{F}^\theta = \sigma(X_s, 0 \leq s \leq t)$, $\mathcal{F}_\infty^\theta = \vee_{t=0}^\infty \mathcal{F}_t^\theta$. Suppose that there exists a set of probability measures P_x on $(\Omega, \mathcal{F})$, $x \in E$, satisfying that $\forall A \in \mathcal{F}_\infty^\theta, x \to P_x(A)$ is $\mathcal{E}$-measurable and $\forall x \in E$,

$$P_x(A) = P(A|X_0 = x), \quad \forall A \in \mathcal{F}_\infty^\theta.$$

For any probability measure μ on $(E, \mathcal{E})$, we define the probability measure P_μ on $(\Omega, \mathcal{F}_\infty^\theta)$ as follows: $\forall A \in \mathcal{F}_\infty^\theta$, $P_\mu(A) = \int_E P_x(A)\mu(dx)$. Let $\mathcal{F}_t^\mu$ be completion of $\mathcal{F}_t^\mu$ about P_μ, and

$$\mathcal{F}_t = \cap_{\mu \in \mathcal{P}(E)} \mathcal{F}_t^\mu, \quad t \geq 0,$$

where $\mathcal{P}(E)$ denotes the set of all probability measures on $(E, \mathcal{E})$.

Remark 2.2 *Since $\hat{X}$ is a right-continuous process defined on a metric measurable space, then it is measurable step by step. Hence $\hat{X}(\tau_n, \omega)$ and $f(\hat{X}(\tau_n + \cdot, \omega))$ are measurable.*

Suppose that $X = \{X(t,\omega), 0 \leq t < \tau\}$ is a Markov skeleton process. Let $\eta_n = (\sigma_n, \hat{X}(\tau_n))$, $n \geq 0$, where $\sigma_0 = 0$, $\sigma_n = \tau_n - \tau_{n-1}$, $n \geq 1$ (agreeing that $\infty - \infty = 0$), then $\{\eta_n, n \geq 0\}$ is a series of random variables with values in measurable space $(R_+ \times E_\Delta, \mathcal{B}(R_+) \times \mathcal{E}_\Delta)$.

Theorem 2.1 *Suppose that $X = \{X(t,\omega), 0 \leq t < \tau\}$ is a Markov skeleton process, then $\{\eta_n, n \geq 0\}$ is a Markov sequence (Markov process), and the transition probability $P(\eta_{n+1} \in B|\eta) = P(\eta_{n+1} \in B|\hat{X}_{\tau_n})$ $(B \in \mathcal{B}(R_+) \times \mathcal{E}_\Delta, n \geq 0)$ is independent of the first component σ_n of η_n.*

Proof $\forall B \in \mathcal{B}(R_+) \times \mathcal{E}_\Delta$, $(\eta_{n+1} \in \mathcal{B}) \in \sigma(\hat{X}(\tau_n + t), t \geq 0)$. By the definition of Markov skeleton process (2.2), we have

$$\begin{aligned} P(\eta_{n+1} \in B|\eta_0, \eta_1, \cdots, \eta_n) &= E[P(\eta_{n+1} \in B|\mathcal{F}_{\tau_n})|\eta_0, \eta_1, \cdots, \eta_n] \\ &= E[P(\eta_{n+1} \in B|\hat{X}_{\tau_n})|\eta_0, \eta_1, \cdots, \eta_n] \\ &= P(\eta_{n+1} \in B|\hat{X}_{\tau_n}). \end{aligned}$$

So, $(\eta_n, n \geq 0)$ is a Markov process. ∎

Furthermore, if the Markov skeleton process X is homogeneous, by (2.3),

$$P(\eta_{n+1} \in B|\hat{X}_{\tau_n}) = P_{\hat{X}_{\tau_n}}(\eta_1 \in B)$$

and $\forall t \in R_+$, $x \in E$, $n \geq 0$, $P(\sigma_{n+1} > t|\mathcal{F}_{\tau_n}) = P_{\hat{X}_{\tau_n}}(\tau_1 > t)$, P_x- a.s. on Ω_{τ_n} where $\Omega_{\tau_n} = (\tau_n < \infty)$. So we get the statements (i) and (ii) of the following corollary, while the proof of (iii) is straightforward.

Corollary 2.1 *If the Markov skeleton process $X = (X(t,\omega), 0 \leq t < \tau(\omega))$ is homogeneous, then*

(i) $\{\eta_n, n \geq 0\}$ *is a homogeneous Markov process, and the transition probability is*

$$P_x(\eta_{n+1} \in B|\eta_n) = P_{\hat{X}_{\tau_n}}(\eta_1 \in B), \quad P_x\text{-a.s.}, x \in E;$$

(ii) $\forall t \in R_+$, $x \in E$, $n \geq 0$

$$P_x(\sigma_{n+1} \in B|\mathcal{F}_{\tau_n}) = P_{\hat{X}_{\tau_n}}(\tau_1 > t), \quad P_x\text{-a.s. on } \Omega_{\tau_n};$$

(iii) $\forall C_i \in \mathcal{B}(R_+)$, $i = 1, 2, \cdots, n$; $n \geq 0$,

$$\begin{aligned} &P_x(\sigma_1 \in C_1, \cdots, \sigma_n \in C_n | \hat{X}_0, \hat{X}_{\tau_1}, \cdots, \hat{X}_{\tau_n}) \\ &\quad = P_x(\sigma_1 \in C_1 | \hat{X}_0, \hat{X}_{\tau_1}) P_x(\sigma_2 \in C_2 | \hat{X}_{\tau_1}, \hat{X}_{\tau_2}) \cdots P(\sigma_n \in C_n | \hat{X}_{\tau_{n-1}}, \hat{X}_{\tau_n}), \\ &\qquad P_x\text{-}a.s.,\ x \in E. \end{aligned}$$

Theorem 2.1, Corollary 2.1, and the Markov property (2.2) and (2.3) with respect to the sequence of stopping time $(\tau_n)_{n\geq 0}$ are the reasons we call the stochastic process X a Markov skeleton process. The sequence $(\eta_n, n \geq 0)$ is called the Markov skeleton of the process X and $(X_{\tau_n}, n \geq 0)$ is called the embedded chain of X. Since $(\eta_n)_{n\geq 0}$ and $(\tau_n, \hat{X}_{\tau_n})_{n\geq 0}$ mutually determine each other, $(\tau_n, \hat{X}_{\tau_n})_{n\geq 0}$ may also be called the Markov skeleton of the process X.

Suppose $X = \{X(t,\omega), 0 \leq t < \tau(\omega)\}$ is the minimal Q-process, where τ is the first explosion. Obviously X is a homogeneous Markov skeleton process, whose Markov skeleton is $(\sigma_n, X_{\tau_n})_{n\geq 0}$, where $\sigma_n = \tau_{n+1} - \tau_n$ and τ_n is the n^{th} jump time, and $\{X_{\tau_n}\}$ is the embedded chain. We feel the Markov skeleton is much more important than the embedded chain for the minimal Q-process, because the embedded chain gives only the transfer states when a transition occurs without indicating how long the chain stayed on the state it left. But the Markov skeleton does both. In fact, the Markov skeleton of a minimal Q-process itself determine each other. Furthermore, from the viewpoint of the transition kernel, the transition matrix $\{q_{ij}/q_i\}$ cannot uniquely determine the Q-matrix (q_{ij}), but the transition kernel of the Markov skeleton and the Q-process can uniquely determine each other.

Next, we introduce jump process of the Markov skeleton process.

Definition 2.2 *The process $Y = (Y(t,\omega), 0 \leq t < \tau(\omega))$ is called the jump process of the Markov skeleton process $X = (X(t,\omega), 0 \leq t < \tau(\omega))$ if*

$$Y(t,\omega) = X_{\tau_n}, \qquad \text{when } \tau_n \leq t < \tau_{n+1},\ n \geq 0. \tag{2.4}$$

Let $(q_n(x, dt, dx))_{n\geq 0}$ denote the series of the transition kernels of the Markov series $(\eta_n)_{n\geq 0}$. (By Yan Jia'an [1, Note 11.50], we know $(\eta_n)_{n\geq 0}$ exists). And let $F_n(x, dt) = q_n(x, dt, E_\triangle)$, $n \geq 0$. $\forall B \in \mathcal{E}_\triangle$, $q_n(x, dt, dx) \ll F_n(x, dt)$, so

$$q_n(x, dt, B) = Q_n(x, t, B)\, F_n(x, dt),$$

where the Randon-Nikodym derivative $Q_n(x,t,B)$ of $q_n(x,t,B)$ with respect to $F_n(x,dt)$ may be chosen so that for fixed (x,t), $Q_n(x,t,\cdot)$ is a probability measure on $\mathcal{E}_\triangle$, and for fixed $B \in \mathcal{E}_\triangle$, $Q_n(\cdot,\cdot,B)$ is $\mathcal{E}_\triangle \times \mathcal{B}(R_+)$-measurable. In fact, we have

$$\begin{aligned} q_n(\hat{X}_{\tau_n}, dt, dx) &= P_x(\sigma_{n+1} \in dt, \hat{X}_{\tau_{n+1}} \in dx | \hat{X}_{\tau_n}), && P_x\text{-a.s.},\ x \in E_\triangle,\ n \geq 0; \\ F_n(\hat{X}_{\tau_n}, dt) &= P_x(\sigma_{n+1} \in dt | \hat{X}_{\tau_n}), && P_x\text{-a.s.},\ x \in E_\triangle,\ n \geq 0; \\ Q_n(\hat{X}_{\tau_n}, \sigma_{n+1}, dx) &= P_x(\hat{X}_{\tau_{n+1}} \in dx | \hat{X}_{\tau_n}), \sigma_{n+1}), && P_x\text{-a.s.},\ x \in E_\triangle,\ n \geq 0. \end{aligned}$$

Before we finish this section, we introduce a set of sub-processes of the Markov skeleton process $X = (X(t,\omega), 0 \leq t < \tau(\omega))$, $X^{(n)} = (X^{(n)}(t,\omega), 0 \leq t < \sigma_n(\omega))$, $n \geq 1$, as follows

$$X^{(n)}(t,\omega) = X(\tau_{n-1} + t, \omega), \quad 0 \leq t < \sigma_n(\omega)),\ n \geq 1$$

Obviously, the Markov skeleton process evolves as follows: starting from the initial state X_0, it first evolves according to the first sub-process $X^{(1)}$ until the time τ_1

(the distribution of the τ_1 is $F_0(X_0, \cdot)$), then jumps to the state X_{τ_1} according to the transition kernel $Q_0(X_0, \tau_1, \cdot)$; and starting from X_{τ_1} again, evolves according to the second sub-process $X^{(2)}$ until the time τ_2 (the distribution of the $\tau_2 - \tau_1$ is $F_1(X_{\tau_1}, \cdot)$), then jumps to the state X_{τ_2} according to the transition kernel $Q_1(X_{\tau_1}, \tau_2 - \tau_1, \cdot)$; continue this way until time τ, when the Markov skeleton process X stops.

3. Definition and the back-forward equations of normal Markov skeleton processes

Definition 3.1 *A homogeneous Markov skeleton process $X = \{x(t,\omega), 0 \le t < \tau(\omega)\}$ is said normal if there exist $(h(t,x,A))$ and $(q(t,x,A))$, with the stopping times $\{\tau_n\}_{n\ge 0}$ in* Definition 2.1, *satisfying the following conditions:*

(i) $P[x(\tau_n + t) \in A, \tau_{n+1} - \tau_n > t | x(\tau_n)] = h(t, x(\tau_n), A)$,*P-a.e.,* $A \in \mathcal{E}$, $t \ge 0$, $n \ge 0$;

(ii) $P[x(\tau_{n+1}) \in A, \tau_{n+1} - \tau_n \le t | x(\tau_n)] = q(t, x(\tau_n), A)$,*P-a.e.,* $A \in \mathcal{E}$, $t \ge 0$, $n \ge 0$.

For a fixed A, $h(t,x,A)$ is a measurable function of two variables; for fixed x and t, $h(t,x,A)$ is a quasi-distribution on $(E, \mathcal{E})$.

In particular,

$$\begin{aligned} h(t,x,A) &= P(x(t) \in A, \tau_1 > t | x(0) = x), \\ q(t,x,A) &= P(x(\tau_1) \in A, \tau_1 \le t | x(0) = x). \end{aligned}$$

Where $q(t,x,A)$ is the transition probability of (η_n), $q(t,x,A) = \int_0^t \int_A q(ds, x, dy)$.

From now on we consider the normal Markov skeleton processes only without mentioning the term "normal".

Let $\mathcal{M} \hat{=} \{R | R(x,A)$ be a nonnegative function defined on $E \times \mathcal{E}$; i.e. for fixed A, $R(x,A)$ is $\mathcal{E}$-measurable; for fixed x, $R(x,A)$ is a nonnegative measure on $(E,\mathcal{E})\}$. It is well known that convolution on $\mathcal{M}$ can be defined as follows: $\forall R, S \in \mathcal{M}$,

$$R \cdot S(x,A) \hat{=} \int_E R(x,dy) S(y,A), \quad x \in E,\ A \in \mathcal{E}. \tag{3.1}$$

Obviously, $R \cdot S \in \mathcal{M}$ and the multiplication in $\mathcal{M}$ satisfies the associative law. In particular, for any $R \in \mathcal{M}$

$$\begin{aligned} R^0(x,A) &\hat{=} \delta_A(x) \\ R^{n+1}(x,A) &\hat{=} \int_E R(x,dy) R^n(y,A) \\ &= \int_E R^n(x,dy) R(y,A), \quad x \in E,\ A \in \mathcal{E} \end{aligned} \tag{3.2}$$

Let

$$\begin{aligned} P(t,x,A) &= P(x(t) \in A | x(0) = x) \qquad & t \ge 0,\ x \in E,\ A \in \mathcal{E} \\ P_\lambda(x,A) &\hat{=} \int_0^\infty e^{-\lambda t} P(t,x,A)\, dt, \qquad & \lambda > 0,\ x \in E,\ A \in \mathcal{E} \end{aligned}$$

Theorem 3.1 *$\forall\lambda > 0$, $\{P_\lambda(x, A), x \in E, A \in \mathcal{E}\}$ is the minimal non-negative solution of the following non-negative equation*

$$X(x,A) = \int_E q_\lambda(x,dy)X(y,A) + h_\lambda(x,A), \quad x \in E,\ A \in \mathcal{E} \tag{3.3}$$

So

$$P_\lambda(x,A) = (\sum_{n=0}^{\infty} Q^n \cdot H)(x,A) \tag{3.4}$$

Where

$$\left.\begin{array}{rcl} H &=& (h_\lambda(x,A), x \in E, A \in \mathcal{E}), \\ Q &=& (q_\lambda(x,A), x \in E, A \in \mathcal{E}) \end{array}\right\} \tag{3.5}$$

$$\left.\begin{array}{rcl} h_\lambda(x,A) &=& \int_0^\infty e^{-\lambda t} h(t,x,A)dt, \\ q_\lambda(x,A) &=& \int_0^\infty e^{-\lambda t} dq(t,x,A) \end{array}\right\} \tag{3.6}$$

Remark 3.1 *By* Theorem 3.1, *the distributions of the process X are determined by H and Q, so X is called the (H,Q)-process. (H,Q) is called the binary characteristics or (H,Q)-pair of* Equation (3.3) *and process X.*

To prove Theorem 3.1, we need two lemmas.

Lemma 3.1 $\forall t \geq 0,\ A \in \mathcal{E},\ n \geq 0$

$$E[x(t) \in A, \tau_n \leq t < \tau_{n+1} | x(\tau_n), \tau_n, x(0)] = h(t-\tau_n, x(\tau_n), A) \cdot I_{\{\tau_n \leq t\}} \quad P\text{-}a.e. \tag{3.7}$$

where

$$I_C(\omega) = \begin{cases} 1, & \omega \in C \\ 0, & \omega \bar{\in} C \end{cases}$$

Proof First we prove (3.7) for closed set A. Suppose A is a closed set, let

$$\begin{array}{rcll} A_l &\hat{=}& \left\{x | d(x,A) < \frac{1}{l}\right\}, & l = 1,2,\cdots \\ \bar{A}_l &\hat{=}& \left\{x | d(x,A) \leq \frac{1}{l}\right\} & l = 1,2,\cdots \\ B_i^{(k)} &\hat{=}& \left\{\omega \in \Omega, \frac{i}{2^k}t \leq t - \tau_n < \frac{i+1}{2^k}t\right\} & i = 0,\cdots,2^k-1;\ k \geq 1 \end{array}$$

Noting that the paths of the X are right continuous and $A = \cap_{l=1}^{\infty} A_l = \cap_{l=1}^{\infty} \bar{A}_l$, we have

$$\begin{aligned} &\{x(t) \in A, \tau_n \leq t < \tau_{n+1}\} \\ &= \{x(\tau_n + t - \tau_n) \in A, \tau_{n+1} - \tau_n > t - \tau_n\} \cap \{\tau_n \leq t\} \\ &= \cap_{l=1}^{\infty}(\{x(\tau_n + t - \tau_n) \in A_l, \tau_{n+1} - \tau_n > t - \tau_n\} \cap \{\tau_n \leq t\}) \\ &\subset \cap_{l=1}^{\infty} \cup_{K=1}^{\infty} \cap_{k=K}^{\infty} \left(\cup_{i=0}^{2^k-1} \left\{ x\left(\tau_n + \frac{i+1}{2^k}t\right) \in A_l, \tau_{n+1} - \tau_n > \frac{i+1}{2^k}t \right\} \right. \\ &\qquad \left. \cap B_i^{(k)} \cap \{\tau_n \leq t\} \right) \end{aligned} \tag{3.8}$$

on the other hand,

$$
\begin{aligned}
&\{x(\tau_n + t - \tau_n) \in A, \tau_{n+1} - \tau_n > t - \tau_n\} \cap \{\tau_n \le t\}) \\
&= \cap_{l=1}^{\infty}(\{x(\tau_n + t - \tau_n) \in \bar{A}_l, \tau_{n+1} - \tau_n > t - \tau_n\} \cap \{\tau_n \le t\}) \\
&\supset \cap_{l=1}^{\infty} \cap_{K=1}^{\infty} \cup_{k=K}^{\infty} \left(\cup_{i=0}^{2^k-1} \left\{ x\left(\tau_n + \frac{i+1}{2^k} t\right) \in A_l, \tau_{n+1} - \tau_n > t - \tau_n \right\} \right. \\
&\qquad \left. \cap B_i^{(k)} \cap \{\tau_n \le t\} \right) \\
&\supset \cap_{l=1}^{\infty} \cap_{K=1}^{\infty} \cup_{k=K}^{\infty} \left(\cup_{i=0}^{2^k-1} \left\{ x\left(\tau_n + \frac{i+1}{2^k} t\right) \in A_l, \tau_{n+1} - \tau_n > \frac{i+1}{2^k} t \right\} \right. \\
&\qquad \left. \cap B_i^{(k)} \cap \{\tau_n \le t\} \right)
\end{aligned} \tag{3.9}
$$

Combining (3.8) with (3.9) and using indicator functions we have

$$
\begin{aligned}
&I_{\{x(t) \in A, \tau_n \le t < \tau_{n+1}\}} \\
&= \lim_{l \to \infty} \lim_{k \to \infty} \left(\sum_{i=0}^{2^k-1} I_{\left\{x(\tau_n + \frac{i+1}{2^k} t) \in A_l, \tau_{n+1} - \tau_n > \frac{i+1}{2^k} t\right\}} \cdot I_{B_i^{(k)}} \cdot I_{\{\tau_n \le t\}} \right) \quad P\text{-a.e.}
\end{aligned} \tag{3.10}
$$

Note that the limit in (3.10) is decreasing on l and

$$
\begin{aligned}
0 &\le \sum_{i=0}^{2^k-1} I_{\{x(\tau_n + \frac{i+1}{2^k} t) \in A_l, \tau_{n+1} - \tau_n > \frac{i+1}{2^k} t\}} \cdot I_{B_i^{(k)}} \cdot I_{\{\tau_n \le t\}} \\
&\le 1 \qquad P\text{-a.e.}
\end{aligned}
$$

By the monotone convergence and dominant convergence theorems and the properties of the conditional expectation, we have

$$
\begin{aligned}
&E[x(t) \in A, \tau_n \le t < \tau_{n+1} | x(\tau_n), \tau_n, x(0)] \\
&= \lim_{l \to \infty} \lim_{k \to \infty} \sum_{i=0}^{2^k-1} E\left[x\left(\tau_n + \frac{i+1}{2^k} t\right) \in A_l, \tau_{n+1} - \tau_n > \frac{i+1}{2^k} t | x(\tau_n), \tau_n, x(0) \right] \\
&\qquad \cdot I_{B_i^{(k)}} \cdot I_{\{\tau_n \le t\}} \\
&= \lim_{l \to \infty} \lim_{k \to \infty} \sum_{i=0}^{2^k-1} E\left[x\left(\tau_n + \frac{i+1}{2^k} t\right) \in A_l, \tau_{n+1} - \tau_n > \frac{i+1}{2^k} t | x(\tau_n) \right] \\
&\qquad \cdot I_{B_i^{(k)}} \cdot I_{\{\tau_n \le t\}} \\
&= \lim_{l \to \infty} \lim_{k \to \infty} \sum_{i=0}^{2^k-1} h\left(\frac{i+1}{2^k} t, x(\tau_n), A_l \right) \cdot I_{B_i^{(k)}} \cdot I_{\{\tau_n \le t\}} \\
&= \lim_{l \to \infty} h(t - \tau_n, x(\tau_n), A_l) \cdot I_{\{\tau_n \le t\}} \\
&= h(t - \tau_n, x(\tau_n), \cap_{l=1}^{\infty} A_l) \cdot I_{\{\tau_n \le t\}} \\
&= h(t - \tau_n, x(\tau_n), A) \qquad P\text{-a.e.}
\end{aligned}
$$

In the last third equality, we use

$$\begin{aligned} h(t-\tau_n, x(\tau_n), A_l)\cdot I_{\{\tau_n\le t\}} &\le \lim_{k\to\infty}\sum_{i=0}^{2^k-1} h\left(\frac{i+1}{2^k}t, x(\tau_n), A_l\right)\cdot I_{B_i^{(k)}}\cdot I_{\{\tau_n\le t\}} \\ &\le h(t-\tau_n, x(\tau_n), \bar{A}_l)\cdot I_{\{\tau_n\le t\}} \end{aligned}$$

and while $l\to\infty$, both sides have the same limits.

Up to here, we have proved (3.7) for close sets A. Noting that $h(t,x,A)$ is a quasi-distribution on A and the properties of the conditional expectation, using the monotone class theorem we can prove that for any $A\in\mathcal{E}$, (3.7) holds and this completes the proof of Lemma 3.1. ∎

Lemma 3.2 $\forall A\in\mathcal{E}$, $t\ge 0$, $x\in E$

$$P(x(\tau_n)\in A, \tau_n\le t|x(0)=x) = q^{*n}(t,x,A) \tag{3.11}$$

where

$$\begin{aligned} q^{*0}(t,x,A) &\hat{=} \delta_A(x), \\ q^{*1}(t,x,A) &\hat{=} q(t,x,A) \\ q^{*n}(t,x,A) &\hat{=} \int_E\int_0^t q^{*n-1}(ds,x,dy)q(t-s,y,A), \quad n\ge 2 \end{aligned} \tag{3.12}$$

Proof When $n=1$,

$$P(x(\tau_1)\in A, \tau_1\le t|x(0)=x) = q(t,x,A) = q^{*1}(t,x,A)$$

Assume that (3.11) holds for $n=k$. When $n=k+1$,

$$\begin{aligned} &P(x(\tau_{k+1})\in A, \tau_{k+1}\le t|x(0)=x) \\ &= \int_\Omega E\Big[x(\tau_{k+1})\in A, \tau_{k+1}-\tau_k\le t-\tau_k|x(\tau_k),\tau_k,x(0)=x\Big] \\ &\qquad\qquad \cdot I_{\{\tau_k\le t\}}\cdot P\Big(d\omega|x(0)=x\Big) \\ &= \int_\Omega q\Big(t-\tau_k, x(\tau_k), A\Big)\cdot I_{\{\tau_k\le t\}}\cdot P\Big(d\omega|x(0)=x\Big) \\ &= \int_E\int_0^t q(t-s,y,A)\cdot P\Big(x(\tau_k)\in dy, \tau_k\in ds|x(0)=x\Big) \\ &= \int_E\int_0^t q(t-s,y,A)\cdot q^{*k}(ds,x,dy) \\ &= q^{*k+1}(t,x,A) \end{aligned}$$

In the last third equality, we use the integral transformation:

$$T:\Omega\to E\times[0,\infty], \quad T(\omega)\hat{=}(X(\tau_k),\tau_k)$$

Thus the Lemma 3.2 is proved. ∎

Proof The proof of Theorem 3.1 is as follows:

First, we reduce the demonstration of Theorem 3.1 to proving that

$$P(x(t) \in A|x(0) = x) = \sum_{n=0}^{\infty} \int_E \int_0^t h(t-s, y, A) q^{*n}(ds, x, dy) \tag{3.13}$$

In fact, taking the Laplace transformation on both sides in (3.13), we have

$$\begin{aligned} P_\lambda(x, A) &= \int_0^\infty e^{-\lambda t} P(x(t) \in A|x(0) = x)\, dt \\ &= \sum_{n=0}^{\infty} \int_E h_\lambda(y, A) q_\lambda^{*n}(x, dy) \end{aligned} \tag{3.14}$$

where

$$\begin{aligned} q_\lambda^{*n}(x, A) &= \delta_A(x) \\ q_\lambda^{*n}(x, A) &= \int_0^\infty e^{-\lambda t} q^{*n}(dt, x, A) \\ &= \int_E \cdots \int_E q_\lambda(x, dy_1) q_\lambda(y_1, dy_2) \cdots q_\lambda(y_{n-1}, A), \quad n \geq 1 \end{aligned}$$

So, (3.14) becomes (3.4), i.e., the minimal non-negative solution of (3.3).

Next, we prove (3.13)

$$\begin{aligned} & P(x(t) \in A|x(0) = x) \\ &= \sum_{n=0}^{\infty} P\big(x(t) \in A, \tau_n \leq t < \tau_{n+1}|x(0) = x\big) \\ &= \sum_{n=0}^{\infty} \int_\Omega E\big[x(t) \in A, \tau_n \leq t < \tau_{n+1}|x(\tau_n), \tau_n, x(0)\big] P\big(d\omega|x(0) = x\big) \\ &= \sum_{n=0}^{\infty} \int_\Omega h(t-\tau_n, x(\tau_n), A) \cdot I_{\{\tau_k \leq t\}} \cdot P\big(d\omega|x(0) = x\big) \\ &= \sum_{n=0}^{\infty} \int_E \int_0^t h(t-s, y, A) \cdot P\big(x(\tau_n) \in dy, \tau_n \in ds|x(0) = x\big) \\ &= \sum_{n=0}^{\infty} \int_E \int_0^t h(t-s, y, A) \cdot q^{*n}(ds, x, dy) \end{aligned}$$

In the third, the fourth and the last equalities; we have respectively used Lemma 3.1, integral transformation and the Lemma 3.2.

So (3.13) holds and this completes the proof of Theorem 3.1. ∎

Definition 3.2 Equation (3.3) *is called the backward equation of the* (H, Q)*-process* X.

Definition 3.3 *If for any* $\lambda > 0$*, there exists* $\hat{Q} = (\hat{q}_\lambda(x, A), x \in E, A \in \mathcal{E}) \in \mathcal{M}$ *satisfying* $H \cdot \hat{Q} = Q \cdot H$*, i.e.*

$$\int_E h_\lambda(x, dy) \hat{q}_\lambda(y, A) = \int_E q_\lambda(x, dy) h_\lambda(y, A), \quad x \in E,\ A \in \mathcal{E} \tag{3.15}$$

then the following non-negative Equation (3.16) *is called the forward equation of the* (H, Q)*-process* X,

$$\hat{X}(x, A) = \int_E \hat{X}(x, dy)\hat{q}_\lambda(y, A) + h_\lambda(x, A), \quad x \in E, \ A \in \mathcal{E}, \ \lambda > 0 \qquad (3.16)$$

Proposition 3.1 *If* H *has right inverse element in* $\mathcal{M}$, *i.e. for any* $\lambda > 0$, *there exists* $H_r^{-1} \in \mathcal{M}$ *satisfying*

$$H \cdot H_r^{-1}(x, A) = \delta_A(x) \quad \forall x \in E, \ A \in \mathcal{E}$$

then there exists the forward Equation (3.16), *where* $\hat{Q} = H_r^{-1} \cdot Q \cdot H$.

Proof $\forall \lambda > 0$, there exists $H_r^{-1} \in \mathcal{M}$. Let

$$\hat{Q} = H_r^{-1} \cdot Q \cdot H$$

Noting that the multiplication in $\mathcal{M}$ satisfies the associative law, we have

$$\begin{aligned} H \cdot \hat{Q} &= H \cdot (H_r^{-1} \cdot Q \cdot H) \\ &= (H \cdot H_r^{-1}) \cdot (Q \cdot H) \\ &= (\delta_A(x)) \cdot (Q \cdot H) \\ &= Q \cdot H \end{aligned}$$

By the Definition 3.3, the forward Equation (3.16) exists. ■

Theorem 3.2 *If there exists a forward equation of the* (H, Q)*-process, then the minimal non-negative solutions of both the forward equation and the backward equation are identical. So* $\{P_\lambda(x, A), x \in E, A \in \mathcal{E}\}$ *is also the minimal non-negative solution of the forward equation, i.e.*

$$P_\lambda(x, A) = \left(\sum_{n=0}^{\infty} H \cdot \hat{Q}^n\right)(x, A)$$

Proof Obviously the minimal non-negative solutions of Equation (3.3) and Equation (3.16) can be obtained by the following

$$\begin{aligned} X(x, A) &= \lim_{n \to \infty} X^{(n)}(x, A), \\ \hat{X}(x, A) &= \lim_{n \to \infty} \hat{X}_n(x, A) \end{aligned}$$

where

$$\begin{aligned} X^{(0)}(x, A) &\equiv 0 && x \in E, \ A \in \mathcal{E} \\ X^{(n+1)}(x, A) &= \int_E q_\lambda(x, dy) X^{(n)}(y, A) + h_\lambda(x, A), && x \in E, \ A \in \mathcal{E}, \ n \geq 0 \\ \hat{X}^{(0)}(x, A) &\equiv 0 && x \in E, \ A \in \mathcal{E} \\ \hat{X}^{(n+1)}(x, A) &= \int_E \hat{X}^{(n)}(x, dy)\hat{q}_\lambda(y, A) + h_\lambda(x, A), && x \in E, \ A \in \mathcal{E}, \ n \geq 0 \end{aligned}$$

By (3.15), $X^{(0)} \equiv \hat{X}^{(0)}$ and the above equalities, we have

$$X^{(n)}(x, A) \equiv \hat{X}^{(n)}(x, A) \quad n \geq 0$$

Thus Theorem 3.2 is proved. ■

4. Regularity criterion of (H,Q)-process

Definition 4.1 *The (H,Q)-process $X = \{x(t,\omega), 0 \le t < \tau(\omega)\}$ is said to be regular if and only if for any $x \in E$, we have*

$$P(\tau = \infty | x(0) = x) \;=\; 1 \tag{4.1}$$

Theorem 4.1 *The (H,Q)-process $X = \{x(t,\omega), 0 \le t < \tau(\omega)\}$ is regular if and only if for any $x \in E$ and $t > 0$,*

$$p(t,x,E) \;=\; 1 \tag{4.2}$$

Equivalently, for any $x \in E$ and $\lambda > 0$,

$$\lambda p_\lambda(x,E) \;=\; 1 \tag{4.3}$$

Proof The conclusion is obvious. ■

Let $B_E \hat{=} \{f : f$ be a bounded measurable real-value function defined on $(E, \mathcal{E})\}$.

Lemma 4.1 *If $0 \le f \in B_E$ and for some $\lambda > 0$, there exists $0 \le u \in B_E$ such that*

$$f(x) - \int_E q_\lambda(x,dy) f(y) \;=\; \int_E h_\lambda(x,dy) u(y) \;\ge\; 0 \tag{4.4}$$

then

$$f(x) \;\ge\; \int_E P_\lambda(x,dy) u(y), \quad \forall x \in E \tag{4.5}$$

Furthermore, if the equation

$$\begin{cases} g(x) &= \int_E q_\lambda(x,dy) g(y), \quad x \in E \\ 0 &\le g \in B_E \end{cases} \tag{4.6}$$

only has a null solution, then (4.5) becomes an equality, i.e.

$$f(x) \;=\; \int_E P_\lambda(x,dy) u(y), \quad \forall x \in E \tag{4.7}$$

Proof For $\lambda > 0$, using the Theorem 3.1, we have that $\forall A \in \mathcal{E}$, $\{p_\lambda(x,A), x \in E\}$ is the minimal non-negative solution of the following equation

$$X(x) \;=\; \int_E q_\lambda(x,dy) X(y) + h_\lambda(x,A), \quad x \in E. \tag{4.8}$$

Using the method of finding the minimal non-negative solution in the proof of Theorem 3.2, we can prove that $\{\int_E p_\lambda(x,dy) u(y), x \in E\}$ is the minimal non-negative solution of the following

$$X(x) \;=\; \int_E q_\lambda(x,dy) X(y) + \int_E h_\lambda(x,dy) u(y), \quad x \in E. \tag{4.9}$$

By the conditions in the lemma, the following equality holds.

$$f(x) = \int_E q_\lambda(x,dy)f(y) + \int_E h_\lambda(x,dy)u(y), \quad x \in E.$$

So the equality (4.5) holds.

For any $x \in E$, let

$$g(x) \hat{=} f(x) - \int_E p_\lambda(x,dy)u(y)$$

Obviously, $0 \leq g \in B_E$ and g satisfies (4.6). If (4.6) has only a null solution, then $g \equiv 0$, i.e. (4.7) holds. ■

Theorem 4.2 *The $(H,Q)-$process X is regular if and only if that the following equation has only null solution*

$$\begin{cases} f(x) = \int_E q_\lambda(x,dy)f(y), & x \in E,\ \lambda > 0 \\ 0 \leq f \leq 1, & f \in B_E \end{cases} \tag{4.10}$$

Proof

Sufficiency: Let $u(y) \equiv \lambda$ in (4.9), we have that $\{\lambda p_\lambda(x,E), x \in E\}$ is the minimal non-negative solution of the following equation

$$\begin{cases} X(x) = \int_E q_\lambda(x,dy)X(y) + \lambda h_\lambda(x,E), & x \in E \\ 0 \leq X \leq 1, & X \in B_E \end{cases} \tag{4.11}$$

As (4.10) has only a null solution, thus (4.11) has the unique solution $\{\lambda p_\lambda(x,E),\ x \in E\}$.

By Definition 3.1, it is true that for any $x \in E$

$$q(t,x,E) + h(t,x,E) = 1.$$

So

$$\frac{1}{\lambda}\int_0^\infty e^{-\lambda t} dq(t,x,E) + \int_0^\infty e^{-\lambda t} h(t,x,E)\,dt = \frac{1}{\lambda}$$

i.e.

$$q_\lambda(x,E) + \lambda h_\lambda(x,E) = 1 \tag{4.12}$$

Hence $X(x) \equiv 1$ is also the solution of (4.11), and

$$\lambda p_\lambda(x,E) = 1, \quad \forall x \in E$$

So by Theorem 4.1, the $(H,Q)-$process X is regular.

Necessity: Because $\{\lambda p_\lambda(x,E), x \in E\}$ and $X(x) = 1$ are respectively the minimal solution and the maximal solution of (4.11), we can conclude that $\{1 - \lambda p_\lambda(x,E), x \in E\}$ is the maximal solution of (4.10). In fact, the maximal solution of (4.10) can be obtained by the following recurrent procedure

$$\begin{aligned}
f^{(0)}(x) &\equiv 1 \\
f^{(1)}(x) &= \int_E q_\lambda(x, dy) f^{(0)}(y) \\
&= q_\lambda(x, E) \\
&= 1 - \lambda h_\lambda(x, E) \\
f^{(2)}(x) &= \int_E q_\lambda(x, dy) f^{(1)}(y) \\
&= q_\lambda(x, E) - \lambda \int_E q_\lambda(x, dy) h_\lambda(y, E) \\
&= f^{(1)}(x) - \lambda \int_E q_\lambda(x, dy) h_\lambda(y, E) \\
&= 1 - \lambda \left[h_\lambda(x, E) + \int_E q_\lambda(x, dy) h_\lambda(y, E) \right] \\
&\vdots \\
f^{(n+1)}(x) &= \int_E q_\lambda(x, dy) f^{(n)}(y) \\
&= f^{(n)}(x) - \lambda \int_E q_\lambda(x, dy_1) \int_E \cdots \int_E q_\lambda(y_{n-1}, dy_n) h_\lambda(y_n, E) \\
&= 1 - \lambda \left(\sum_{k=0}^{n} Q^k \cdot H \right)(x, E)
\end{aligned}$$

So

$$\begin{aligned}
f(x) &= \lim_{n\to\infty} f^{(n)}(x) \\
&= 1 - \lambda \left(\sum_{k=0}^{\infty} Q^k \cdot H \right)(x, E) \\
&= 1 - \lambda P_\lambda(x, E)
\end{aligned}$$

Since the (H, Q)-process X is regular, by the Theorem 4.1 it is true that $\lambda p_\lambda(x, E) = 1$, (4.10) has only a null solution. ∎

The following is a sufficient condition for the (H, Q)-process to be regular which can be easily verified.

Theorem 4.3 *If $q_\lambda(x, A)$ satisfies the following condition*

$$\beta(\lambda) \hat{=} \sup_{x\in E} q_\lambda(x, E) < 1, \quad \forall \lambda > 0, \tag{4.13}$$

then the (H, Q)-process $X = \{x(t, \omega), 0 \le t < \tau\}$ is regular.

Proof By (4.13) and the method in solving the maximal solution of (4.10), we have

$$\begin{aligned}
f^{(1)}(x) &= q_\lambda(x,E) \\
&\le \beta(\lambda) \\
f^{(2)}(x) &= \int_E q_\lambda(x,dy) f^{(1)}(y) \\
&\le \beta(\lambda) q_\lambda(x,E) \\
&\le \beta^2(\lambda) \\
&\vdots \\
f^{(n+1)}(x) &= \int_E q_\lambda(x,dy) f^{(n)}(y) \\
&\le \beta^n(\lambda) q_\lambda(x,E) \\
&\le \beta^{n+1}(\lambda)
\end{aligned}$$

So

$$0 \le f(x) = \lim_{n\to\infty} f^n(x) \le \lim_{n\to\infty} \beta^n(\lambda) = 0, \quad \forall x \in E.$$

And (4.10) has only a null solution. By Theorem 4.2, X is regular. ■

Corollary 4.1 *If the number of the elements of state space E is finite and for any $x \in E$*

$$P(\tau_1 > 0 | x(0) = x) > 0 \tag{4.14}$$

Then the (H,Q)-process $X = \{x(t,\omega), 0 \le t < \tau\}$ is regular.

Proof By the above conditions (4.13) is true. ■

5. Some important special conditions

(A) $(H, G \cdot q)$-processes

The separation condition (D):

$$q(t,x,A) = G(t,x)\, q(x,A) \tag{5.1}$$

i.e.

$$P\big(x(\tau_1) \in A, \tau_1 \le t | x(0) = x\big) = P\big(\tau_1 \le t | x(0) = x\big) \cdot P\big(x(\tau_1) \in A | x(0) = x\big) \tag{5.2}$$

Definition 5.1 *If a (H,Q)-process X satisfies the condition (D), then it is called a $(H, G \cdot q)$-process.*

By Hou Zhenting and Guo Qingfeng [1, Lemma 9.3.1], the minimal Q-processes are $(H, G \cdot q)$-processes.

Let

$$G_\lambda(x) = \int_0^\infty e^{-\lambda t}\, dG(t,x), \tag{5.3}$$

then the separation condition (D) becomes

$$q_\lambda(x,A) = G_\lambda(x)\, q(x,A). \tag{5.4}$$

The backward equation of the $(H, G \cdot q)$-process becomes

$$X(x,A) = G_\lambda(x)\int_E q(x,dy)X(y,A)+h_\lambda(x,A), \quad \lambda>0,\ x\in E,\ A\in\mathcal{E}. \tag{5.5}$$

Theorem 4.2 (criteria for regularity) becomes the following theorem.

Theorem 5.1 *A $(H, G \cdot q)$-process X is regular if and only if the following equation has only a null solution.*

$$\begin{cases} f(x) = G_\lambda(x)\int_E q(x,dy)f(y) & x\in E,\ \lambda>0 \\ 0 \le f \le 1, & f\in B_E \end{cases} \tag{5.6}$$

(B) Generalized Doob processes

If the separation condition (D) holds and

$$q(x,A) = q(A), \quad x\in E,\ A\in\mathcal{E} \tag{5.7}$$

then the process X is called the generalized Doob process. It is obvious that the Doob processes in traditional Markov processes are the generalized Doob processes.

The backward equation of the generalized Doob process becomes

$$Z(x,t) = G_\lambda(x)\int_E q(dy)Z(y,A)+h_\lambda(x,A). \tag{5.8}$$

So

$$\int_E q(dy)X(y,A) = \int_E G_\lambda(y)q(dy)\cdot\int_E q(dy)X(y,A)+\int_E h_\lambda(y,A)q(dy). \tag{5.9}$$

And

$$\int_E q(dy)X(y,A) = \frac{\int_E h_\lambda(y,A)q(dy)}{1-\int_E G_\lambda(y)q(dy)} \tag{5.10}$$

By (5.8), (5.9), (5.10) and Theorem 3.1, we have

$$p_\lambda(x,A) = h_\lambda(x,A)+\frac{G_\lambda(x)\int_E h_\lambda(y,A)q(dy)}{1-\int_E G_\lambda(y)q(dy)} \tag{5.11}$$

(C) Semi-Markov processes

Definition 5.2 *Suppose that $X=\{x(t), t<\tau\}$ is a Markov skeleton process and $\{\tau_n\}_{n\ge 0}$ is the time component of its Markov skeleton. If the state space E is a denumerable set and*

$$x(t) = x(\tau_n), \quad \tau_n\le t<\tau_{n+1},\ n\ge 0, \tag{5.12}$$

then the X is called semi-Markov process.

(D) Piecewise determinate Markov skeleton processes

Definition 5.3 *Suppose that $X=\{x(t), t<\tau\}$ is a Markov skeleton process (maybe nonhomogeneous). If there exist measurable functions $\phi_n:[0,\infty)\times E\to E(n\ge 0)$ such that for any fixed $x\in E$, $\phi_n(t,x)$ is right-continuous on t and*

$$X(t,w) = \sum_{n=0}^{\infty}\phi_n(t-\tau_n, x(\tau_n))I_{\{\tau_n\le t<\tau_{n+1}\}}(w), \quad 0\le t<\tau,$$

then X is called a piecewise deterministic Markov skeleton process.

(E) Piecewise determinate Markov processes

Definition 5.4 *If a piecewise deterministic Markov skeleton process X is a Markov process, then the X is called a piecewise deterministic Markov process.*

(F) Piecewise deterministic Markov processes in Davis sense (PDMP)

Piecewise determinate Markov processes (PDMP) in Davis [1] is a proper subclass of the piecewise determinate Markov processes in Definition 5.4. One of the most important conditions for the piecewise deterministic Markov processes to be a PDMP in a Davis sense is that $F(x,t) = P(\tau_1 > t|X(0) = x)$ is absolutely continuous. At first sight, the PDMP model in a Davis sense, is limited by Markov processes and does not have rich content. While dealing with practical problems, it displays much generality and many advantages. By means of the additional variable, a piecewise deterministic Markov skeleton process with the absolutely continuous $F(x,t) = P\big(\tau_1 > t|X(0) = x\big)$, which is not a Markov process, can often be transformed to a PDMP model.

(G) Markov-type skeleton processes

If a Markov skeleton process X is a Markov process, then X is called a Markov-type skeleton process.

(H) Stochastic processes with jumps

The Markov skeleton processes which consist of Brownian motion, diffusion processes and denumerable Markov processes (or birth-and-death processes) and the Markov-type skeleton processes categorize diffusion processes with jump. Markov-type diffusion processes with jumps, Brownian birth-and-death processes, etc. are categorized as stochastic processes with jumps. This provides appropriate mathematical models to study the so-called evolution law (all things obey the repetitive and alternating development from change in quantity to change in quality) and make it meaningful to work on the theories and applications.

(I) Denumerable Markov skeleton processes

If the state space E is a denumerable set, then the Markov skeleton processes are called denumerable Markov skeleton processes. For example, the queue lengths $L(t)$ of the M/G/1 and GI/G/1 belong to these kind of processes.

The backward equation of these kind of processes becomes

$$X_{ij} = \sum_{k\in E} q_{ik}(\lambda)X_{ij} + h_{ij}(\lambda), \quad \lambda > 0,\ i,j \in E, \tag{5.13}$$

and the forward equation becomes

$$X_{ij} = \sum_{k\in E} X_{ik}\hat{q}_{kj}(\lambda)X_{ij} + h_{ij}(\lambda), \quad \lambda > 0,\ i,j \in E, \tag{5.14}$$

where

$$h_{ij}(\lambda) = \int_0^\infty e^{-\lambda t} P\big(X(t) = j,\, \tau_1 > t|X(0) = i\big)\, dt, \tag{5.15}$$

$$q_{ij}(\lambda) = \int_0^\infty e^{-\lambda t} P\big(X(\tau_1) = j,\, \tau_1 \le t|X(0) = i\big)\, dt, \tag{5.16}$$

$\hat{Q} = (\hat{q}_{ij}(\lambda))$ satisfies

$$\sum_{k\in E} h_{ik}(\lambda)\hat{q}_{kj}(\lambda) = \sum_{k\in E} q_{ik}(\lambda)h_{kj}(\lambda), \quad \lambda > 0,\ i,j \in E. \tag{5.17}$$

6. Supplements and notes

The concept of Markov skeleton processes (Definition 2.1) and their backward and forward equations were introduced in 1997 by Hou Zhenting, Liu Zaiming and Zou Jiezhong for the first time [2, 3, 5]. The main results of this paper are Theorems 3.1 and 3.2 which determine the one-dimensional probability distribution of the Markov skeleton processes . They are the key theorems in the theory of Markov skeleton processes, for in the study of any stochastic process, the first-line question is to determine its probability distribution, especially its one-dimensional distribution. As far as we know, Kolmogorov backward and forward equations were obtained only for pure-discontinuous (also called "jump") Markov processes and branching processes, while for semi-Markov processes, only the backward equation was established. Now all these become special cases of the backward equation (3.3) or the forward equation (3.16) in Theorem 3.1. In Hou Zhenting, Guo Qingfeng [1], the very simple formula (9.2.3) for the computation of the transition probability for minimal Q-processes and the rather complicated formula (10.2.16) for the computation for order-1 Q-processes are unified as the backward equation (3.3). The subsequent ones, such as the forward equation for semi-Markov processes, the backward equation for piecewise deterministic Markov processes, the queue length $L(t)$ of GI/M/1 queueing system, the waiting time $w(t)$ of GI/G/1 queueing system, and the forward equation for the queue length $L(t)$ of G/M/1 queueing system are all the new special cases of (3.3) or (3.6). So we may say that our backward and forward equations extend greatly the application range of that of Kolmogorov in Markov processes. In the deduction of the backward equation for the Markov skeleton processes (or rather, the deduction of Theorem 3.1) we use the same method as in the deduction of the original Kolmogorov backward equation, namely, by means of the Markov property of the stopping time τ_1 as this method is brief and clear. However, the deduction of the Kolmogorov forward equation for pure-discontinuous Markov processes use the Markov property thoroughly and with much difficulty. Since Markov skeleton processes have a much weaker Markov property than the traditional Markov processes, we give up the probability method, and turn to ideas used in operator theory to obtain the forward equation.

References

[1] Hou Zhenting, Guo Qingfeng, Homogeneous Denumerable Markov Processes. Beijing, Springer-verlag, Science Press, 1988.

[2] Yang Xiangqun, The Construction Theory of Denumerable Markov Processes. John Wiley & Son, 1990.

[3] Xu Guanghui, Random Serving System. Beijing. Science Press, 1980.

[4] M.H.A. Davis. Markov Models and Optimitation. London: Chapman & Hape, 1993.

[5] R.C. Merton. Option Pricing When Underlying Stock Return Are Discontinuous. Journal of Financial Economics, 1976 (3): 125–144.

[6] D.G. Kendall. Some Problems in the Theory of Queues. J Roy Statist Soc, B, 1951, 13, 151–185.

[7] D.G. Kendall. Stochastic Processes Occuring in the Theory of Queues and their Analysis by the Methods of the Imbedded Markov Chain. Ann Math Statist, 1953, 24: 338–354.

[8] L. takacs. The Transient Behavior of a single Server Queueing Process with a Poisson Input. Proc Fourth Berkeley Symp. On Math Statist and Prob Berkeley and Los Angeles, Univ. of California Press, 1961 (2): 535–567.

[9] Wu Fang, Queueing Process GI/M/n, ACTA Mathematica SINICA, 1961 (11): 295–305.

[10] Xu Guanghui, The Transient Properties of A Queuing Process GI/M/n, ACTA Mathematica SINICA, 1965 (15): 91–120.

[11] U.N.Bhat. Transient Behavior of Multiserver Queues with Recurrent Input and Exponential Service Times, J. Appl. Prob. ,1968 (5): 158–168.

[12] Hou Zhenting, Liu Zaiming, Zou Jiezhong. QNQL Processes: (H,Q)-processes and their Applications, Chinese Science Bulletin, 1997, 42 (11): 881–886.

[13] Hou Zhenting, Liu Zaiming, Zou Jiezhong. Stochastic Processes with Markov Framework, Mathematics in Economics, 1997, 14 (1): 1–13.

[14] Hou Zhenting, Liu Zaiming, Zou Jiezhong. Multidimensional Distributions of Stochastic Processes with Markov skeleton, Mathematics in Economics, 1997, 14 (2): 1–8.

[15] Hou Zhenting, Liu Zaiming, Zou Jiezhong. Markov Skeleton Processes. Chinese Science Bulletin, 1998, 43 (11), 881–889.

[16] Hou Zhenting, Zou Jiezhong, Yuan Chengui. Input Processes (the Intervals of Reaching times are i.i.d.), Mathematics in Economics, 1996, 13 (1): 1–8.

[17] Hou Zhenting, Liu Zaiming, Zou Yi. Input Processes (batch reaching case), Mathematics in Economics, 1996, 13 (2): 1–3.

[18] Hou Zhenting, Liu Zaiming. $(H_m, Q_m)_{m=1}^{\infty}$-processes and Applications in Queueing Systems. (To appear.)

[19] Hou Zhenting, Liu Zaiming. Random Models for Mathematical Ecology. (To appear in "Journal of the Biological Mathematics".)

[20] Hou Zhenting, Li Junping, Liu Zaiming et al. Functional Differential Equations with Random Pulse. (To appear.)

[21] Liu Wanrong, Liu Zaiming, Hou Zhenting. Characterisings of Markov Skeleton Processes. (To appear.)

[22] Liu Wanrong, Liu Zaiming, Hou Zhenting. Random Time Transformations for Markov Skeleton Processes. (To appear.)

[23] Hou Zhenting, Liu Guoxin, Zou Yi. Moments of First Passage Time of Semi-Markov Processes. Proceedings of the Second International Symposium on Semi-Markov Models: Theory and Applications. Universite De Technologic Compiegne. 1998.

[24] Hou Zhenting, Liu Guoxin et al. Semi-Markov Processes: the Forward and Backward Equations and Regularity. (To appear.)

[25] Hou Zhenting, Liu Guoxin et al. Semi-Markov Processes: Characters of Digits and Classifications of States. (To appear.)

[26] Hou Zhenting, Liu Guoxin et al. Semi-Markov Processes: Distributions and Moments of Integral-type Functionals. (To appear.)

[27] Tang Yourong, Liu Zaiming, Hou Zhenting. Semi-Markov birth-and-death Processes. (To appear in ACTA Mathematica SINICA.)

[28] Liu Guoxin, Hou Zhenting. Piecewise Deterministic Processes and Markov Modeling of Non-diffusion Stochastic Systems with Continuous Parameter. Proceedings of the Second International Symposium on Semi-Markov Models: Theory and Applications. Universite De Technologic Compiegne. 1998.

[29] Liu Guoxin. Piecewise deteministic Markov Processes. Ph.D thesis, Changsha Railway University. 1998.

[30] A.A. Markov. Extension of the law of large numbers to dependent quantities [in Russian]. IZV.FIZ-Matem. Obsch. Kazan Univ. (2nd Ser), 1906, 15: 135–156.

[31] Levy, P., Semi-Markovian Processes, Proc:III Internat. Congr. Math. (Amsterdam), 1954, 416–426.

[32] Davis, M.H.A., Piecewise-deterministic Markov processes: a general class of non-diffusion stochastic models, J. R. Statist. Soc. B. 1988, 46, 353–388.

[33] Davis, M.H.A., Markov Models and Optimization, London:Chaoman and Hall, 1993.

[34] Yan Jia'an, Introduction to Martingale and Stochastic Integral, Shanghai: Shanghai Science and Technology Press, 1981.

Chapter 6

PIECEWISE DETERMINISTIC MARKOV PROCESSES AND SEMI-DYNAMIC SYSTEMS*

Guoxin Liu
Department of Applied Mathematics,
Hebei University of Technology
Tianjin 300130, P.R.China
gxliu@hebut.edu.cn

Abstract This paper focuses on the generalised piecewise deterministic Markov processes (PDMPs), introduced by Liu and Hou [14], and studies some properties of PDMPs connected with the characteristic triple. It is pointed out that Davis' PDMP is the special case of the PDMP here, which restricts all the survivor functions so that they are absolutely continuous with respect to the time t. Furthermore, the state jump measure of a PDMP is introduced. When accompanied by the (state) transition kernal, it plays the same role to PDMPs as Q-matix to Q-processes.

Keywords: Piecewise deterministic Markov processes (PDMPs), the state jump measure of a PDMP, Stieltjes exponential, Stieltjes logarithm.

AMS 1991 classifications: 60J25,60G20.

1. Introduction

The terminology of *piecewise deterministic Markov processes* (PDMPs or PDPs) was initially introduced by Davis [4] as a general class of stochastic models. Since then, the stochastic control theory of PDMPs has been extensively studied by many authors. Among them we quote

*This research is supported by NSF of Hebei Province.

Z. Hou et al. (eds.), Markov Processes and Controlled Markov Chains, 93–107.

Davis [5], Dempster and Ye [7], and Schäl [16] and references therein. The theory of PDMPs is successfully applied in capacity expansion and risk theory by Davis, Dempster, Sethi and Vermes [6] and Dassios and Embrechts [3], and Embrechts and Schmidli [8] respectively. The stationary distribution for a PDMP has been studied by Costa [1]. The stability of PDMPs has recently been studied by Costa and Dufour [2].

Davis' PDMPs are a family of Markov processes following deterministic trajectories between random jumps. The motion of Davis' PDMP depends on three local characteristics, namely the flow ϕ, the jump rate λ and the transition measure Q, which specifies the post-jump location. Starting from x the motion of the process follows the flow $\phi(t, x)$ until the first jump time τ_1 which occurs either spontaneously in a Poisson-like fashion with rate $\lambda(\phi(t, x))$ or when the flow $\phi(t, x)$ hits the boundary of the state space. In either case the location of the process at the random jump time τ_1 is selected by the transition measure $Q(\phi(\tau_1, x), \cdot)$ and the motion restarts from this new point as before. It is well known that, when accompanied by the duration time since the last random jump as a supplementary variable, a semi-Markov process becomes a Markov process (refer to Gihman and Skorohod [9, page 295], and the latter is a Markov process following deterministic trajectories between random jumps too. Nevertheless, there is no restriction on its random jumps in so-called Poisson-like fashion. Hence, Davis' model is not general enough to cover this important case.

Liu and Hou [14] generalised the concept of Davis' PDMP by virtue of the ideas of Hou, Liu and Zou [11, 12, 13], in which they introduced the concept of *Markov skeleton processes* (MSPs) to discribe the general stochastic systems that are of Markov property at least at countable increasing (fixed or random) times. The generalised *piecewise deterministic processes* (PDPs) are, just as their name indicated, stochastic processes involving deterministic trajectories punctuated by random jumps. More precisely, there exists a sequence of random occurrances at fixed or random increasing times, $\tau_1, \tau_2, \cdots$, but there is no additional component of uncertainty between these times, and the only restriction is the Markov property of the processes at these times. Also they call a PDP a *piecewise deterministic Markov process* (PDMP) if it becomes a Markov process. This generalised PDMP overcame the shortage mentioned above. Liu and Hou [14] presented the necessary and sufficient conditions for a PDP to be a PDMP and pointed out that suitably chosen supplementary variables can make a PDP become a PDMP.

In this paper we provide the general properties of the generalised piecewise deterministic Markov processes. In Section 2 we give the main definitions and notations. In Section 3 we study the properties of PDMPs connected with the jump time characteristic of PDMP by virtue of the Stieltjes version of exponentials and logarithms, and present the representations of jump time characteristic of a PDMP. It is indicated, by the way, that the so-called 'Poisson-like jump', which is just the form of the jump time characteristic in Davis' PDMP model, is equivalent to restricting the jump time characteristic of a PDMP to being absolutely continuous with respect to the time t, and that the durations between jumps must be of exponential distribution with jump rate $\lambda(x)$, $x \in E$ in the cases of Markov jump processes. Section 4 introduces the concept of the state jump measure and state jump transition kernel of a PDMP, which play the same role to PDMPs as Q-matrix to Q-processes.

2. Definitions and notations

Let $X = \{x(t,\omega), 0 \le t < \tau\}$ be a stochastic process defined on a complete probability space $(\Omega, \mathcal{F}, P)$ with state space $(E, \mathcal{E})$, where $(E, \mathcal{E})$ is a Polish space. Let $F = (\mathcal{F}_t)_{t\ge 0}$ be the natural filtration of the process X, where $\mathcal{F}_t = \sigma(x_s, 0 \le s \le t)$, $\mathcal{F}_\infty = \bigvee_{t=0}^{\infty} \mathcal{F}_t$. Suppose that there exists a family of probability measures, P_x, $x \in E$, on $(\Omega, \mathcal{F})$ such that, for any $A \in \mathcal{F}_\infty$, the function $x \mapsto P_x(A)$ is $\mathcal{E}$-measurable and, for all $x \in E$,

$$P_x(A) = P(A|x_0 = x), \qquad A \in \mathcal{F}_\infty.$$

We add an isolated point Δ to state space E, and define a stochastic process $\hat{X} = \{\hat{x}(t,\omega), 0 \le t \le \infty\}$ on $\Omega \cup \{[\Delta]\}$ by

$$\hat{x}(t,\omega) = \begin{cases} x(t,\omega), & \text{if } 0 \le t < \tau(\omega), \\ \Delta, & \text{if } \tau \le t \le \infty. \end{cases}$$

Thus the process X can be thought of as a process well defined for all $t \ge 0$. We shall, by convenient use of notation, generally denote it X whenever $\tau(\omega) = \infty$ *a.s.* or not.

Definition 2.1 *A right continuous process $X = \{x(t,\omega), 0 \le t < \tau(\omega)\}$ with state space $(E, \mathcal{E})$ is called a (homogeneous)* ***piecewise deterministic process (PDPs)****, if there exist a strictly increasing sequence of nonnegative r.v.'s $\{\tau_n\}_{n\ge 0}$ with $\tau_0 = 0$, $\tau_n \uparrow \tau$, and a measurable map $\phi : (R_+ \times E, \mathcal{B}(R_+) \times \mathcal{E}) \to (E, \mathcal{E})$ with right continuity with respect to $t \in R_+$ and $\phi(0,x) = x$, $x \in E$ such that*

$$x(t,\omega) = \sum_{n=0}^{\infty} \phi(t - \tau_n, x(\tau_n)) I_{[\tau_n \le t < \tau_{n+1}]}, \qquad 0 \le t < \tau, \tag{2.1}$$

and the sequence $\{(\tau_n - \tau_{n-1}, x_{\tau_n})\}_{n\ge 0}$ ($\tau_{-1} = 0$ by convention) is a homogeneous Markov sequence with transition probabilities dependent on the second component only.

Denote transition kernel of Markov sequence, $\{(\tau_n - \tau_{n-1}, x_{\tau_n})\}_{n\geq 0}$, by $G(x, dt, dx)$. And denote $F(x, dt) = G(x, dt, E)$, $n \geq 0$. It follows from $G(x, dt, dx) \ll F(x, dt)$ that, for any $B \in \mathcal{E}$,

$$G(x, dt, B) = Q(x, t, B)F(x, dt),$$

where $Q(x, t, B)$ is the Radon-Nikodym derivative of $G(x, dt, B)$ with respect to $F(x, dt)$. $Q(x, t, B)$ can be selected such that $Q(x, t, \cdot)$ is a probability measure on $(E, \mathcal{E})$ for any fixed $(x, t) \in E \times R_+$ and $Q(\cdot, \cdot, B)$ is $\mathcal{E} \times \mathcal{B}(R_+)$-measurable for any fixed $B \in \mathcal{E}$. In fact, we have

$$\begin{aligned} G(x, dt, dy) &= P_x(\tau_1 \in dt, x_{\tau_1} \in dy), \quad && P_x - a.s.; \\ F(x, dt) &= P_x(\tau_1 \in dt), && P_x - a.s.; \\ Q(x, \tau_1, dy) &= P_x(x_{\tau_1} \in dy | \tau_1), && P_x - a.s., \end{aligned}$$

for all $x \in E$.

Now we are in position to see how a PDP evolves. Starting from the initial state x_0, a PDP X moves along $\phi(\cdot, x_0)$ until time τ_1 with distribution $F(x_0, \cdot)$ and then it jumps instantaneously to state x_{τ_1} according to the transition probability $Q(x_0, \tau_1, \cdot)$; and the process restarts from the state x_{τ_1} and moves along $\phi(\cdot, x_{\tau_1})$ until time $\tau_2, \cdots$. The process repeats in the similar way until it stops at the time τ. We can see that the motion of a PDP depends only on the three characteristics, ϕ,F and Q.

We call (ϕ, F, Q) the *characteristic triple* of a PDP; F and Q the *jump time characteristic* and *jump transition characteristic* respectively.

Definition 2.2 *A PDP $X = \{x(t, \omega), 0 \leq t < \tau\}$ is called a* ***piecewise deterministic Markov process (PDMP)****, if it is a Markov process.*

It is more convenient to use the following equivalent definition of PDMP by virtue of Liu and Hou [14, Theorem 3.1].

Definition 2.3 *A PDP $X = \{x(t, \omega), 0 \leq t < \tau\}$ with characteristic triple (ϕ, F, Q) is called a PDMP if there exists a function, $c : (E, \mathcal{E}) \to (\bar{R}_+ \backslash \{0\}, \mathcal{B}(\bar{R}_+ \backslash \{0\})$, such that, for any $x \in E$, $s, t \in R_+$ and $s + t \in (0, c(x))$, we have*

(i) ϕ is a semi-flow, i.e.,

$$\phi(0, x) = x; \qquad \phi(t, \phi(s, x)) = \phi(s + t, x). \tag{2.2}$$

(ii) the jump time characteristic F satisfies the following functional equation

$$F(x, 0) = 0; \qquad F(x, t + s) = F(x, t) \cdot F(\phi(t, x), s); \tag{2.3}$$

and $c(x) = \inf\{t : F(x,t) = 0\}$, *where the survivor function* $F(x,t) = F(x,(t,\infty])$.

(iii) the jump transition characteristic Q *satisfies*

$$Q(x,t,\{\phi(t,x)\}) = 0; \; Q(x,t+s,dy) = Q(\phi(t,x),s,dy). \quad (2.4)$$

Remark 2.1 *Further more,* Theorem 3.1 *of Liu and Hou* [14] *tells us that the above defined PDMP is also a strong Markov process.*

In the following we need the concepts of the Stieltjes version of exponentials and logarithms. we reserve the term **F-function** for a right continuous decreasing function $F : R_+ \to [0,1]$ such that $F(0) = 1$, and **Λ-function** for a right continuous increasing function $\Lambda : R_+ \to \bar{R}_+$ such that $\Lambda(0) = 0$ and $\Delta\Lambda(t) < 1$ for all t with $\Lambda(t) < \Lambda(\infty)$, possibly $\Delta\Lambda(t) = 1$ if $\Lambda(t) = \Lambda(\infty) < \infty$. (Here $\Lambda(\infty) := \sup \Lambda(t) = \lim_{t\uparrow\infty} \Lambda(t)$.)

It is easy to see that the survivor function $F(x,\cdot)$ for each $x \in E$ is an F-function.

For F-function F, let $c_F := \inf\{t : F(t) = 0\}$ and for Λ-function Λ, let $c_\Lambda := \inf\{t : \Lambda(t) = \infty \text{ or } \Delta\Lambda(t) = 1\}$. The **Stieltjes logarithm** of an F-function F is defined to be the function slog$F$, where

$$\text{slog}F(t) := \int_{(0,t\wedge c_F]} \frac{-dF(s)}{F(s-)}.$$

If Λ is a Λ-function, we may express Λ in a unique way as $\Lambda^c + \Lambda^d$, where Λ^c and Λ^d are Λ-functions which stop at c_Λ, Λ^c being continuous and Λ^d purely discontinuous. The **Stieltjes exponential** sexpΛ is defined by

$$\text{sexp}\Lambda(t) := e^{-\Lambda^c(t)} \prod_{0<u\le t} [1 - \Delta\Lambda(u)].$$

For the detailed properties of Stieltjes exponentials and Stieltjes logarithms, refer to Meyer [15] or Sharpe [17, Appendices].

3. The jump time characteristic of a PDMP

Let ϕ be a semi-flow with the state space $(E,\mathcal{E})$.

An F-function family, $\{F(x,\cdot) : x \in E\}$, is called **ϕ-multiplicative** if, for any $x \in E$, $s,t \in R_+$ and $s+t \in (0,c(x))$, we have

$$F(x,t+s) = F(x,t) \cdot F(\phi(t,x),s).$$

A Λ-function family,$\{\Lambda(x,\cdot) : x \in E\}$, is called **$\phi$-additive** if, for any $x \in E$, $s,t \in R_+$ and $s+t \in (0,c(x))$, we have

$$\Lambda(x,t+s) = \Lambda(x,t) + \Lambda(\phi(t,x),s).$$

Definition 2.3 shows us that the jump time characteristic of a PDMP is ϕ-multiplicative.

Lemma 3.1 *If An F-function family $\{F(x,\cdot)\}$ is ϕ-multiplicative, then the function $c(x) := \inf\{t : F(x,t) = 0\}$, $x \in E$ is $\bar{R}_+\backslash\{0\}$-valued and satisfies*

$$c(x) = t + c(\phi(t,x)), \qquad t \in [0, c(x)). \tag{3.1}$$

Furthermore, if the state x is periodic for the semi-flow ϕ, then $c(x) = \infty$.

Proof The positivity is directly from the right continuity at $t = 0$ of an F-function. Note that $F(x,t) > 0$ for each $t \in [0, c(x))$, we have

$$\begin{aligned} c(\phi(t,x)) &= \inf\{s > 0, F(\phi(t,x)), s) = 0\} \\ &= \inf\{s > 0, F(x, t+s)/F(x,t) = 0\} \\ &= \inf\{s > 0, F(x, t+s) = 0\} \\ &= \inf\{u > 0, F(x,u) = 0\} - t \\ &= c(x) - t. \end{aligned}$$

This proves equation (3.1).

Supposed that $x \in E$ be a periodic state of the semi-flow ϕ, then there exists a $T \in (0, c(x))$ such that $\phi(T,x) = x$. Hence,

$$c(x) = T + c(x),$$

by equation (3.1), and this implies $c(x) = \infty$. ∎

Lemma 3.2 *An F-function family $\{F(x,\cdot)\}$ is ϕ-multiplicative, if and only if the Λ-function family $\{\Lambda(x,\cdot)\}$ is ϕ-additive. Where $\Lambda(x,\cdot) := slogF(x,\cdot)$, $x \in E$ (i.e. $F(x,\cdot) = sexp\Lambda(x,\cdot)$, $x \in E$).*

Proof If the F-function family $\{F(x,\cdot)\}$ is ϕ-multiplicative, then we have by Lemma 3.1 that the Λ-function family $\{\Lambda(x,\cdot)\}$ satisfies, for

$t \in [0, c(x)]$, $s \in R_+$,

$$\begin{aligned}
\Lambda(x, t+s) &= \text{slog} F(x, t+s) \\
&= \int_{(0,(t+s)\wedge c(x)]} \frac{-dF(x,u)}{F(x,u-)} \\
&= \int_{(0,t\wedge c(x)]} \frac{-dF(x,u)}{F(x,u-)} + \int_{(t\wedge c(x),(t+s)\wedge c(x)]} \frac{-dF(x,u)}{F(x,u-)} \\
&= \Lambda(x,t) + \int_{(t,(t+s)\wedge c(x)]} \frac{-dF(x,u)}{F(x,u-)} \\
&= \Lambda(x,t) + \int_{(0,s\wedge c(\phi(t,x))]} \frac{-dF(x,u)}{F(x,u-)} \\
&= \Lambda(x,t) + \Lambda(\phi(t,x), s).
\end{aligned}$$

This proves the ϕ-additivity of the family $\{\Lambda(x,\cdot)\}$.

Conversely, Suppose that the Λ-function family $\{\Lambda(x,\cdot)\}$ is ϕ-additive. The ϕ-additivity implies that $\Delta\Lambda(x, t+s) = \Delta\Lambda(\phi(t,x), s)$, for each $x \in E$ and any $t \in (0, c(x)]$ and $s \in R_+$. Hence, we have for $t \in [0, c(x)]$, $s \in R_+$,

$$\begin{aligned}
F(x, t+s) &= \text{sexp}\Lambda(x, t+s) \\
&= e^{-\Lambda(x,t+s)} \prod_{0<u\leq t+s} [1 - \Delta\Lambda(x,u)] e^{\Delta\Lambda(x,u)} \\
&= e^{-\Lambda(x,t)} \prod_{0<u\leq t} [1 - \Delta\Lambda(x,u)] e^{\Delta\Lambda(x,u)} \\
&\quad \times e^{-\Lambda(\phi(t,x),s)} \prod_{t<u\leq t+s} [1 - \Delta\Lambda(x,u)] e^{\Delta\Lambda(x,u)} \\
&= F(x,t) e^{-\Lambda(\phi(t,x),s)} \prod_{0<u\leq s} [1 - \Delta\Lambda(\phi(t,x),u)] e^{\Delta\Lambda(\phi(t,x),u)} \\
&= F(x,t) F(\phi(t,x), s).
\end{aligned}$$

This is the ϕ-multiplicativity of the family $\{F(x,\cdot)\}$ and completes the proof of the lemma. ∎

Theorem 3.1 *Let $\{F(x,\cdot) : x \in E\}$ be ϕ-multiplicative. If $F(x,t)$ is absolutely continuous on $[0, c(x))$ for some $x \in E$, then so is $\Lambda(x,\cdot)$, the Stieltjes logarithm of $F(x,\cdot)$, and there exists a nonnegative function $\lambda(\cdot)$ on the trajectory $\{\phi(t,x) : 0 \leq t < c(x)\}$ such that*

$$\Lambda(x,t) = \int_0^t \lambda(\phi(u,x))\, du, \qquad t \in [0, c(x)), \tag{3.2}$$

or, equivalently,

$$F(x,t) = e^{-\int_0^t \lambda(\phi(u,x))\,du}, \qquad t \in [0, c(x)). \tag{3.3}$$

Proof It is obviously that $\Lambda(x,t)$ is absolutely continuous on $[0, c(x))$ by the definition of Stieltjes logarithm. Let

$$\lambda(y) = \lim_{t \downarrow 0} \frac{\Lambda(y,t)}{t}, \qquad y \in \{\phi(t,x) : 0 \le t < c(x)\},$$

if the right side limit above exists; and $= 0$ otherwise. Since $\{\Lambda(x,\cdot)\}$ is ϕ-additive, we have

$$\begin{aligned}
\frac{\partial^+ \Lambda(x,t)}{\partial t} &:= \lim_{s \downarrow 0} \frac{\Lambda(x,t+s) - \Lambda(x,t)}{s} \\
&= \lim_{s \downarrow 0} \frac{\Lambda(\phi(t,x), s)}{s} \\
&= \lambda(\phi(t,x)).
\end{aligned}$$

Also the formula (3.2) follows from the monotonousness and absolutely continuity of $\Lambda(x,\cdot)$, and so does the formula (3.3). This completes the proof. ■

Remark 3.1 Formula (3.3) *is just the form in Davis' PDMP for jump time characteristic, which indicates that Davis' PDMP is the special case of PDMP here. Also the so-called 'Poisson-like jump' is just to restrict the jump time characteristic of PDMP being absolutely continuous with respect to t.*

Theorem 3.2 *Let $\{F(x,\cdot) : x \in E\}$ be ϕ-multiplicative, and denote E_e the set of all equilibrium states for the semi-flow ϕ. Then there exists a nonnegative and finite function $\lambda(\cdot)$ on E_e such that, for any $x \in E_e$, the Stieltjes logarithm of $F(x,t)$,*

$$\Lambda(x,t) = \lambda(x)t, \qquad t \in R_+, \tag{3.4}$$

or equivalently,

$$F(x,t) = e^{-\lambda(x)t}, \qquad t \in R_+. \tag{3.5}$$

Proof It follows from Lemma 3.1 that $c(x) = \infty$ for $x \in E_e$, since equilibrium state is the special case of periodic state. In this case, the ϕ-additivity of $\{\Lambda(x,\cdot)\}$ yields

$$\Lambda(x,t+s) = \Lambda(x,t) + \Lambda(x,s) \qquad s,t \in R_+.$$

This functional equation has a unique Λ-function solution

$$\Lambda(x,t) = \lambda(x)t, \qquad t \in R_+,$$

where $\lambda(x) = \Lambda(x,1) = \frac{\partial^+\Lambda(x,t)}{\partial t}|_{t=0}$. Thus, one can get (3.4). Further (3.5) follows directly from the definition of Stieltjes exponential. This completes the proof. ■

Corollary 3.1 *If a PDMP reduces to a Markov jump process, then there exists a nonnegative and finite $\mathcal{E}$-measurable function $\lambda(\cdot)$ on E such that the survivor function*

$$F(x,t) = e^{-\lambda(x)t}, \qquad t \in R_+,$$

for each $x \in E$.

Proof Since $\phi(t,x) \equiv x$, for all $x \in E$, in the case of Markov jump processes, it follows that $E_e = E$. Hence, there exists a nonnegative and finite function $\lambda(\cdot)$ on E such that the equation (3.5) is satisfied for each $x \in E$. The $\mathcal{E}$-measurability of $\lambda(\cdot)$ is due to the $\mathcal{E}$-measurability of $F(\cdot,t) = P_\cdot(\tau_1 > t)$ for each $t \in R_+$. ■

4. Jump measure and jump transition kernel of a PDMP

In this section, we assume that ϕ is a flow on E instead of a semi-flow.

We add $c(x)$ to the domain of $\phi(\cdot,x)$ for each $x \in E$, which should keep ϕ a flow on $E \cup \partial_+ E$ if needed. Where

$$\partial_+ E := \{\phi(c(x),x) : x \in E\}, \tag{4.1}$$

represents those boundary points at which the flow exits from E.

Denote $c_-(x) := \inf\{t \in R : F(\phi(t,x),-t) > 0\}$, $x \in E$, and the subsets

$$E(x) := \{\phi(t,x) : c_-(x) < t \le c(x)\}, \qquad x \in E. \tag{4.2}$$

The set $E(x)$ is called a *state* (or *phase*) *trajectory* for any $x \in E$. In the case of a flow ϕ, $E(x_1) = E(x_2)$ if and only if $E(x_1) \cap E(x_2) \neq \emptyset$ $(x_1, x_2 \in E)$; and if $x \in E$ is a periodic state, then each state $y \in E(x)$ is periodic with same period.

Now we are in position to construct a measure along each state trajectory $E(x)$ for $x \notin E_e$ (i.e. the state x is not an equilibrium state).

Let $\{F(x,\cdot), x \in E\}$ be a ϕ-multiplicative family of F-functions.

Case 1: $x \in E$ is a *periodic state with minimal period* $T_x > 0$.

In this case, $\phi(\cdot, x)$, located on $(0, T_x]$, is a one-to-one map of $(0, T_x]$ to $E(x)$. So we get the inherited measure Λ_x along $E(x)$ with

$$\Lambda_x(\{\phi(u,x) : 0 < u \le t\}) = \Lambda(x,t), \qquad t \in (0, T_x]. \tag{4.3}$$

Case 2: $x \in E$ *is an aperiodic state.*

Denote

$$T_x := \inf\{t > 0 : \phi(t,x) \in E_e\} \wedge c(x),$$

which represents the hitting time to E_e or the boundary $\partial_+ E$ for ϕ starting from x. In Case 2, $\phi(\cdot, x)$ located on $(c_-(x), T_x]$ is a one-to-one map of $(c_-(x), T_x]$ to $E(x)$. We get the inherited measure Λ_x along $E(x)$, as for (4.3) and

$$\Lambda_x(\{\phi(u,x) : s < u \le 0\}) = \Lambda(\phi(s,x), -s), \; s \in [c_-(x), 0]. \tag{4.4}$$

Lemma 4.1 *Let* $x \in E \backslash E_e$. *For any* $y \in E(x) \backslash \partial_+ E$, *the measure* Λ_y *along* $E(y) = E(x)$ *coincides with* Λ_x.

Proof Let $y \in E(x) \backslash \partial_+ E$.

There exists unique $t_0 \in (0, T_x]$ in Case 1 such that $\phi(t_0, x) = y$, which implies $\phi(T_x - t_0, y) = x$ in this case. Therefore due to the ϕ-additive of $\{\Lambda(x,\cdot)\}$ we have

$$\begin{aligned}
\Lambda_y(\{\phi(u,x) : 0 < u \le t\}) &= \Lambda_y(\{\phi(T_x - t_0 + u, y) : 0 < u \le t\}) \\
&= \Lambda(y, T_x - t_0 + t) - \Lambda(y, T_x - t_0) \\
&= \Lambda(\phi(T_x - t_0, y), t) \\
&= \Lambda(x,t) \\
&= \Lambda_x(\{\phi(u,x) : 0 < u \le t\}),
\end{aligned}$$

for $t \in (0, T_x]$. Also there exists unique $t_0 \in (c_-(x), T_x]$ in Case 2 such that $\phi(t_0, x) = y$. Similarly, we have by (4.3) and (4.4),

$$\Lambda_y(\{\phi(u,x) : s < u \le t\}) = \Lambda_x(\{\phi(u,x) : s < u \le t\}),$$

for $c_-(x) < s \le t \le T_x$.
This completes the proof of the lemma. ∎

Therefore, we have defined unique measure along each state trajectory except for the trajectory $E(x) = \{x\}$ reducing to a single equilibrium point. We will omit the subscript x of the measure Λ_x for this reason and denote them Λ only. It is easy to see that,

(i) if $x \in E \backslash E_e$, then

a) $\Lambda(E(x)) < \infty$ in Case 1, and

b) $\Lambda(\{\phi(u,x) : s < u \leq t\} < \infty$ for any $c_-(x) < s \leq t < T_x$ in Case 2;

(ii) a) if $x \in E \backslash \partial_+ E$, then $\Lambda(\{x\}) < 1$;

b) if $x \in \partial_+ E$, then $\Lambda(\{x\}) = 1$ if and only if $\Lambda(\{\phi(u,x) : s < u < 0\}) < \infty$ and $c(\phi(s,x)) < \infty$ for some $s < 0$.

There are at most countable many states x's on a trajectory such that $\Lambda(\{x\}) > 0$.

Definition 4.1 *We call Λ, a measure along each state trajectory, the* ***state****(or* ***phase****)* ***jump measure*** *if the condition* (i) *and* (ii) *above are satisfied. A state $x \in E \cup \partial_+ E$ is a* ***positive jump state*** *if $\Lambda(\{x\}) > 0$.*

A nonnegative and finite $\mathcal{E}_e$-measurable function $\lambda(\cdot)$ on E_e is called a ***jump rate function*** *on E_e.*

Theorem 4.1 *Let ϕ is a flow on $(E, \mathcal{E})$.*

(i) Given a ϕ-multiplicative family of F-functions, $\{F(x,\cdot), x \in E\}$, there exists a unique state jump measure Λ such that (4.3) *and* (4.4) *are satisfied in* Case 1 *and* Case 2 *respectively; and unique jump rate function on E_e such that* (3.5) *is satisfied in case of $x \in E_e$.*

(ii) Conversely, given a state jump measure Λ and a jump rate function λ on E_e, there exists a unique ϕ-multiplicative family of F-functions, $\{F(x,\cdot), x \in E\}$, such that (4.3) *and* (4.4) *are satisfied in* Case 1 *and* Case 2 *respectively, and* (3.5) *in case of $x \in E_e$.*

Proof

(i) It follows directly from Lemma 4.1 and the discussion above.

(ii) Given a state jump measure Λ and a jump rate function λ on E_e, let

$$\Lambda(x,t) := \begin{cases} \lambda(x)t, & \text{if } x \in E_e; \\ \left[\frac{t}{T_x}\right] \Lambda(E(x)) + \Lambda(\{\phi(u,x) : 0 < u \leq t - [t/T_x]T_x\}), & \\ & \text{if in Case 1;} \\ \Lambda(\{\phi(u,x) : 0 < u \leq t\}), & \\ & \text{if in Case 2 and } \Lambda(\{\phi(c(x),x)\}) = 1; \\ \Lambda(\{\phi(u,x) : 0 < u \leq t\}) + \lambda(\phi(T_x,x))(t - T_x)I_{[T_x < t]}, & \\ & \text{if in Case 2 and } \Lambda(\{\phi(c(x),x)\}) < 1. \end{cases}$$

It is easy to see that the family of Λ-functions $\{\Lambda(x,\cdot) : x \in E\}$, defined above, is ϕ-additive and the family of F-functions $\{F(x,\cdot) := \text{sexp}\Lambda(x,\cdot) : x \in E\}$ is the only ϕ-multiplicative family of F-functions such that (4.3) and (4.4) are satisfied in Case 1 and Case 2 respectively, and (3.5) in case of $x \in E_e$.

This completes the proof. ■

Remark 4.1 *Roughly speaking, the state jump measure $\Lambda(dx)$ for a state x represents the possibility of occurrence of random jump just as the process hits x.*

A simple case is that $E \subset R$, $E(x) = E$ for any $x \in E$ and ϕ is continuous with respect t. In this case, the state jump measure Λ is just a measure on $(E, \mathcal{E})$ with $\Lambda([a,b]) < \infty$ for any $a, b \in E \backslash \partial_+ E$.

Theorem 4.2 *Let $\{F(x,\cdot) : x \in E\}$ be ϕ-multiplicative. If $F(x,t)$ is purely discontinuous on $[0, c(x))$ for some $x \in E$, then so is $\Lambda(x,\cdot)$, the Stieltjes logarithm of $F(x,\cdot)$, and there exists a $[0,1]$-valued function $p(\cdot)$ on the trajectory $E(x)$ such that*

$$\Lambda(x,t) = \sum_{0<u\leq t} p(\phi(u,x)), \quad t \in [0, c(x)), \tag{4.5}$$

or, equivalently,

$$F(x,t) = \prod_{0<u\leq t} [1 - p(\phi(u,x))], \quad t \in [0, c(x)). \tag{4.6}$$

Furthermore, $p(y) = 0$ except for, at most, at countable states y in $E(x)$.

Proof By the definitions, an F-function and its Stieltjes logarithm are purely discontinuous at same time. Let Λ be the state jump measure corresponding to the family of F-functions $\{F(x,\cdot) : x \in E\}$, and denote

$$p(y) := \Lambda(\{y\}), \qquad y \in E(x),$$

which is a $[0,1]$-valued function on $E(x)$. The deduction in the proof of Theorem 4.1 yields (4.5), (4.6) and $p(y) = 0$ except, for at most, at countable states y in $E(x)$. ■

Now let's turn to the jump transition characteristic Q, which specifies the post-jump location of a PDP. In fact, one has

$$Q(x_{\tau_n}, \tau_{n+1} - \tau_n, dy) = P(x_{\tau_{n+1}} \in dy | x_{\tau_n}, \tau_{n+1} - \tau_n),$$

for all $n \geq 0$. i.e. the distibution of the post-jump location depends upon both the last post-jump location and the time since the last jump. In the case of a PDMP, the jump transition characteristic Q also satisfies

$$Q(x, t, \{\phi(t, x)\}) = 0; \qquad Q(x, t + s, dy) = Q(\phi(t, x), s, dy), \quad (4.7)$$

for any $x \in E$, $s, t \in R_+$ and $s + t \in (0, c(x)]$. What does this property mean?

Let $K(y, B)$, $y \in E \cup \partial_+ E$ and $B \in \mathcal{E}$, be a Markov kernel with $K(y, \{y\}) = 0$, and let

$$Q(x, t, B) := K(\phi(t, x), B), \qquad x \in E, \ s, t \in R_+ \ (s+t) \in (0, c(x)]. \quad (4.8)$$

It is easy to see that this Q, defined by (4.8), satisfies (4.7). This is just the form of jump transition characteristic Q in Davis' model. Then, one may ask whether any jump transition characteristic Q should be in the form of (4.8). The answer is that it is not exactly.

Theorem 4.3 *Let ϕ be a flow on E. There exist a transition kernel $K(y, B)$, $y \in E \cup \partial_+ E$ and $B \in E$, with $K(y, \{y\}) = 0$ and a transition kernel $K_e(y, B)$, $y \in E_e$ and $B \in \mathcal{E}$,$K_e(y, \{y\}) = 0$ such that for:*
Case 1:

$$Q(x, t, B) = K(\phi(t, x), B), \qquad s, t > 0, \quad (4.9)$$

Case 2:

$$Q(x, t, B) = K(\phi(t, x), B), \qquad s, t \in R_+, \ (s + t) \in (0, T_x], \quad (4.10)$$

Case 2 or $x \in E_e$:

$$Q(x, t, B) = K_e(\phi(t, x), B), \qquad s, t \in R_+, \ (s + t) \in (T_x, \infty), \quad (4.11)$$

Proof If in Case 1, (4.7) implies that $Q(\phi(-t, x), t, B)$ is independent on the choice of $t > 0$. Denoting it by $K(y, B)$ we get (4.9).

If $x \in E_e$, (4.7) implies that $Q(x, t, B)$ is independent on the choice of $t > 0$. Denote it by $K_e(x, B)$ and , noticing that $T_x = 0$ in this case, we get (4.11).

Suppose that $t \in (T_x, \infty)$ in Case 2, then there exists an $s \in (T_x, t)$ such that $Q(x, t, B) = Q(\phi(s, x), t - s, B)$ and $\phi(t, x) = \phi(s, x) \in E_e$. Thus one gets

$$Q(x, t, B) = Q(\phi(s, x), t - s, B) = K_e(\phi(s, x), B) = K_e(\phi(t, x), B).$$

This is (4.11) in Case 2. Further, suppose that $t \in (0, T_x]$ in Case 2, and let $y := \phi(t, x)$, i.e. $x = \phi(-t, y)$, (4.7) implies $Q(x, t, B) =$

$Q(\phi(-u,y),u,B)$ for any $u \in (0, t - c_-(x))$. Denote it by $K(y,B)$ and one gets (4.10).

This completes the proof. ∎

Remark 4.2 *The jump transition characteristic Q should be in the form of* (4.8) *except for $E_e \cup \partial_+ E \neq \emptyset$. The distribution of a post-jump location of a PDMP conditioned on a pre-jump location may be the difference between just hitting E_e and staying in E_e for a while.*

We call $\{K(x,B) : x \in E \cup \partial_+ E,\ B \in \mathcal{E}\}$ the *state* (or, *phase*) *jump transition kernel* of a PDMP.

Finally, one can see that if ϕ is a flow, then the state jump measure Λ accompanied with the jump rate function λ on E_e plays the same role as the jump time characteristic F; and the state jump transition kernels K and K_e play the same role as the jump transition characteristic Q.

References

[1] Costa, O.L.V. (1990), Stationary distributions for piecewise deterministic processes. J. Appl. Prob., 27, 60-73.

[2] Costa, O.L.V. and Dufour (1999), Stability of piecewise deterministic Markov processes, SIAM J. Control and Optimization, 5,.

[3] Dassios, A. and Embrechts, P. (1989), Martingales and insurance risk. Commun. Statist.-Stochastic Models, 5(2), 181-217.

[4] Davis,M.H.A.(1984), Piecewise-deterministic Markov Processes: a general class of non-diffusion stochastic models (with discusstion), J. Royal Statist. Soc. (B), 46, 353-388.

[5] Davis,M.H.A.(1993), Markov Models and Optimization, Chapman & Hall, London.

[6] Davis, M.H.A., Dempster, M.A.H., Sethi, S.P. and Vermes, D. (1987), Optimal capacity expansion under uncertainty. Adv. Appl. Prob., 19, 156-176.

[7] Dempster, M.A.H. and Ye, J.J. (1995), Impulse control of piecewise deterministic Markov processes. Ann. Appl. Probab., 5, 399-423.

[8] Embrechts, P. and Schmidli, H. (1994), Ruin estimation for a general insurance risk model. Adv. Appl. Prob., 36, 404-422.

[9] Gihman,I.I. & Skorohod,A.V.(1983),The Theory of Stochastic Processes II, Springer-Verlag, Berlin.

[10] Hou,Z.T. and Guo,Q.F.(1988), Time-Homogeneous Markov Processes with Countable State Space. Springer-Verlag, Berlin and Science Press, Beijing.

[11] Hou,Z.T., Liu, Z.M. and Zou,J.Z.(1997a), Stochastic processes with Markov skelton. Mathematics in Ecnomics, Vol.14, 1, 1-13. [in Chinese]

[12] Hou,Z.T., Liu, Z.M. and Zou,J.Z.(1997b), QNQL processes: (H,Q)-processes and their applications. Kexue Tongbao, Vol.42, 9, 1003-1008.

[13] Hou,Z.T., Liu, Z.M. and Zou,J.Z.(1998), Markov skelton processes. Kexue Tongbao, Vol.43, 5, 457-467.

[14] Liu,G.X. and Hou, Z.T. (1998), Piecewise deterministic processes and Markov modeling of non-diffusion stochastic systems with continuous parameter. Proceedings of The 2nd International Symposium on Semi-Markov Models: Theory and Applications. France.

[15] Meyer,P.A. (1966), Quelques résultats sur les processus de Markov. Invent. Math. 1, 101-115.

[16] Schäl,M.(1998), On piecewise deterministic Markov control processes: Control of jumps and of risk processes in insurance. Insurance: Math.Econ. 22, 75-91.

[17] Sharpe,M.(1988), General Theory of Markov Processes. Academic Press, San Dicgo.

II

CONTROLLED MARKOV CHAINS AND DECISION PROCESSES

Chapter 7

AVERAGE OPTIMALITY FOR ADAPTIVE MARKOV CONTROL PROCESSES WITH UNBOUNDED COSTS AND UNKNOWN DISTURBANCE DISTRIBUTION*

J. Adolfo Minjárez-Sosa
Departamento de Matemáticas,
Universidad de Sonora.
Rosales s/n Col. Centro,
83000 Hermosillo,
Sonora, México.
aminjare@gauss.mat.uson.mx

Abstract We study the adaptive control problem for a class of discrete-time Markov control processes with Borel state and action spaces, and possibly unbounded one-stage costs. The processes evolve according to recursive equations $x_{t+1} = F(x_t, a_t, \xi_t)$, $t = 0, 1, ...$, with i.i.d. $\Re^k-$ valued random vectors ξ_t with unknown distribution. Assuming observability of ξ_t, we propose three different sets of conditions each of which allows us to prove average optimality of a type of adaptive control policies.

Keywords: Markov control process, discounted and average cost criteria, adaptive control policies.

AMS 1991 Subject Classification: 93E20, 90C40

1. Introduction

In this paper we introduce an average cost optimal adaptive policy for a class of discrete-time Markov control processes (MCPs), with possibly

*Work supported by Fondo para Retener y Repatriar a Investigadores Mexicanos del Consejo Nacional de Ciencia y Tecnología (CONACyT) Grant 980147.

Z. Hou et al. (eds.), Markov Processes and Controlled Markov Chains, 111–134.

unbounded costs, evolving according to the system equation

$$x_{t+1} = F(x_t, a_t, \xi_t), \quad t = 0, 1, \ldots \tag{1.1}$$

Here, F is a known function, x_t, a_t, and ξ_t are the state, action, and the random disturbance at time t, respectively. We suppose that $\{\xi_t\}$, the so-called "disturbance" or "driving" process, is a sequence of independent and identically distributed (i.i.d.) random vectors in $\Re^k$ having an unknown density ρ. Hence, the adaptive policies combine suitable statistical methods to estimate ρ and control actions a_t that depend on the estimators ρ_t of ρ.

In particular, to construct the adaptive policy in this paper, we take advantage of the procedure of statistical estimation of ρ proposed in [9] to obtain an asymptotically discounted optimal adaptive policy for the process (1.1), and then, having the estimators ρ_t we apply the "principle of estimation and control" [18, 20].

The average optimality of the adaptive policy is studied as a limit of discounted programs. For this, we propose three different conditions, C1, C2, C3, which, applying the so-called vanishing discount factor approach, ensure, among other things, the existence of a solution to the average cost optimality inequality (ACOI). These optimality conditions are variants of conditions used in previous works to study either non-adaptive MCPs (see, for instance, [4, 6, 8, 13, 14, 15, 16, 19, 23, 24, 26, 27, 28]) or non-controlled Markov process (see [16, 17]). A condition similar to C3, but more restrictive, was used in [10] and [21] to study also the nonparametric adaptive control problem for the average criterion.

On the other hand, it is well-known that to ensure the existence of average cost optimal stationary policies, under unbounded costs, it suffices to obtain a solution to the ACOI and its minimizers. However, to get such minimizer, typically we require rather restrictive continuity and compactness conditions on the control system (see, for instance, [6, 8, 12, 13, 14, 15, 16, 23, 24]). In contrast, the construction of the average cost optimal adaptive policy proposed in this paper is based on the existence of $\varepsilon-$ minimizers, for $\varepsilon > 0$, of the discounted cost optimality equation, which implies that, as opposed to previous works [5, 10], we need not to impose continuity and compactness conditions on the control model. That is, it can happen that under our assumptions average optimal stationary policies do not exist for the process (1.1) with a known density ρ, while our main results guarantee the existence of average cost optimal adaptive policies.

The remainder of the paper is organized as follows: In Section 2 we introduce the Markov control model we are concerned with, and some basic assumptions. Section 3 contains the condition C1 and some preliminary results, which are used to construct the average cost optimal adaptive policy in Section 4. Next in Section 5 we present the conditions C2 and C3, and finally, in Section 6 we illustrate our assumption and main results with examples on invariant systems [1], an autoregressive-like control model, and a queueing system with controlled service rate.

2. Markov control processes

Notation. Given a Borel space X (that is, a Borel subset of a complete and separable metric space) its Borel sigma-algebra is denoted by $\mathbb{B}(X)$, and "measurable", for either sets or functions, means "Borel measurable". Let X and Y be Borel spaces. Then a stochastic kernel $Q(dx \mid y)$ on X given Y is a function such that $Q(\cdot \mid y)$ is a probability measure on X for each fixed $y \in Y$, and $Q(B \mid \cdot)$ is a measurable function on Y for each fixed $B \in \mathbb{B}(X)$.

Markov control models. Let $(X, A, \Re^k, F, \rho, c)$ be a discrete-time Markov control model where the state space X, and the action space A are both Borel spaces. The dynamics is defined by the system equation (1.1). Here $F : X \times A \times \Re^k \to X$ is a given (known) measurable function, and $\{\xi_t\}$, is a sequence of independent and identically distributed (i.i.d.) random vectors (r.v.'s) on a probability space $(\Omega, \mathcal{F}, P)$, with values in $\Re^k$ and a common distribution with an unknown density ρ.

To each state $x \in X$, we associate a nonempty measurable subset $A(x)$ of A, whose elements are the *admissible controls* (or *actions*) when the system is in the state x. The set

$$\mathbb{K} = \{(x, a) : x \in X,\ a \in A(x)\}$$

of admissible state-action pairs is assumed to be a measurable subset of the Cartesian product of X and A. Finally, the cost-per-stage $c(x, a)$ is a possibly unbounded, nonnegative, real-valued measurable function on $\mathbb{K}$.

For each density μ on $\Re^k$, $Q_\mu(\cdot \mid \cdot)$ denotes the stochastic kernel on X given $\mathbb{K}$, defined as

$$Q_\mu(B \mid x, a) := \int_{\Re^k} 1_B[F(x, a, s)]\mu(s)\, ds, \quad B \in \mathbb{B}(X), \ (x, a) \in \mathbb{K}, \quad (2.1)$$

where $1_B(\cdot)$ stands for the indicator function of the set B. In other words, Q_μ represent the transition law corresponding to the controlled system (1.1) if the disturbance variables ξ_t have density μ.

Control policies. We define the space of admissible histories up to time t by $\mathbb{H}_0 := X$ and $\mathbb{H}_t := (\mathbb{K}\times\Re^k)^t \times X$ for $t \in\mathbb{N}:= \{1,2,\ldots\}$. A generic element of $\mathbb{H}_t$ is written as $h_t = (x_0, a_0, \xi_0, \ldots, x_{t-1}, a_{t-1}, \xi_{t-1}, x_t)$. A control policy $\pi = \{\pi_t\}$ is a sequence of measurable functions $\pi_t : \mathbb{H}_t \to A$ such that $\pi_t(h_t) \in A(x_t)$, for $h_t \in\mathbb{H}_t$ and $t \geq 0$. We denote by Π the set of all control policies, and by $\mathbb{F}\subset \Pi$ the subset of stationary policies. As usual, every stationary policy $\pi \in\mathbb{F}$ is identified with a measurable function $f : X \to A$ such that $f(x) \in A(x)$ for every $x \in X$, so that π is of the form $\pi = \{f, f, f, \ldots\}$. In this case we identify f with π, and use the notation

$$c(x,f) := c(x,f(x)), \quad F(x,f,s) := F(x,f(x),s), \text{ for } x \in X, \ s \in \Re^k.$$

Optimality criteria. When using a policy $\pi \in \Pi$, given the initial state $x_0 = x$, we define the total expected $\alpha-$discounted cost as

$$V_\alpha(\pi,x) := E_x^\pi\left[\sum_{t=0}^{\infty} \alpha^t c(x_t,a_t)\right], \tag{2.2}$$

where $\alpha \in (0,1)$ is the so-called discount factor, and E_x^π denotes the expectation operator with respect to the probability measure P_x^π induced by the policy π, given the initial state $x_0 = x$ (see, e.g., [3] for the construction of P_x^π). We also define the long-run expected average cost as

$$J(\pi,x) := \limsup_{n\to\infty} n^{-1} E_x^\pi\left[\sum_{t=0}^{n-1} c(x_t,a_t)\right]. \tag{2.3}$$

The functions

$$V_\alpha(x) := \inf_{\pi\in\Pi} V_\alpha(\pi,x) \quad \text{and} \quad J(x) := \inf_{\pi\in\Pi} J(\pi,x), \text{ for } x \in X, \tag{2.4}$$

are the optimal $\alpha-$discounted cost and the optimal average cost, respectively. A policy $\pi^* \in \Pi$ is said to be $\alpha-$discount optimal (or simply $\alpha-$optimal) if $V_\alpha(x) = V_\alpha(\pi^*,x)$ for all $x \in X$. Similarly, a policy $\pi^* \in \Pi$ is said to be average cost optimal (AC- optimal) if $J(x) = J(\pi^*,x)$ for all $x \in X$.

Assumptions. We shall require three sets of assumptions. The first one, Assumption 2.1, ensures the existence of $\delta-$optimal ($\delta > 0$) stationary policies for the discounted cost criterion (Lemma 2.1). Note that Assumption 2.1 allows a unbounded cost-per-stage function $c(x,a)$

provided that it is majorized by a "bounding" function W. Assumptions 2.2 and 2.3 are technical requirements on the unknown density ρ and the function W.

Assumption 2.1 (Bounds and semi-continuity)

(i) For every $x \in X$ the function $a \to c(x,a)$ is lower semi-continuous (l.s.c.) on $A(x)$. Moreover, there exists a measurable function $W : X \to \left[\bar{W}, \infty\right)$ such that $\sup_{A(x)} c(x,a) \leq W(x)$, for some $\bar{W} > 0$.

(ii) For each $x \in X$, $A(x)$ is a $\sigma-$ compact set.

Assumption 2.2 (On the density ρ) *Fix an arbitrary $\varepsilon \in (0, 1/2)$ and let $q := 1 + 2\varepsilon$.*

(i) $\rho \in L_q(\Re^k)$.

(ii) There exists a constant L such that for each $z \in \Re^k$

$$\|\Delta_z \rho\|_{L_q(\Re^k)} \leq L\,|z|^{1/q},$$

where $\Delta_z \rho(s) := \rho(s+z) - \rho(s)$, for $s \in \Re^k$, and $|\cdot|$ is the Euclidean norm in $\Re^k$;

(iii) There exists a nonnegative measurable function $\bar{\rho} : \Re^k \to \Re$ such that $\rho(s) \leq \bar{\rho}(s)$ almost everywhere with respect to the Lebesgue measure.

Assumption 2.3

(i) For every $s \in \Re^k$,

$$\varphi(s) := \sup_X [W(x)]^{-1} \sup_{A(x)} W[F(x,a,s)] < \infty. \tag{2.5}$$

(ii) $\int_{\Re^k} \varphi^2(s)\,|\bar{\rho}(s)|^{1-2\varepsilon}\,ds < \infty.$

The function φ in (2.5) might be nonmeasurable. In such a case we suppose the existence of a measurable majorant $\bar{\varphi}$ of φ for which Assumption 2.3(ii) holds.

To conclude this section we state an important consequence of Assumption 2.1 to be used in later sections.

Lemma 2.1 *Suppose that* Assumption 2.1 *holds, and let $\alpha \in (0,1)$ be an arbitrary but fixed discount factor.*

(i) [13] *If $V_\alpha(x) < \infty$ for every $x \in X$, then V_α satisfies the dynamic programming equation*

$$V_\alpha(x) = \inf_{a\in A(x)} \left[c(x,a) + \alpha \int_{\Re^k} V_\alpha[F(x,a,s)]\rho(s)\,ds \right] \quad \forall x \in X. \tag{2.6}$$

(ii) For each $\delta > 0$, there exists a policy $f \in I\!F$ such that

$$c(x,f) + \alpha \int_{\Re^k} V_\alpha[F(x,f,s)]\rho(s)ds \leq V_\alpha(x) + \delta \quad \forall x \in X. \tag{2.7}$$

From the fact that $Q_\rho(\cdot \mid \cdot)$ is a stochastic kernel (see (2.1)), it is easy to prove that for every non-negative measurable function u, and every $r \in \Re$, the set

$$\left\{ (x,a) : \int_{\Re^k} u[F(x,a,s)]\rho(s)ds \leq r \right\}$$

is a Borel subset of $I\!K$. Using this fact, part (ii) of Lemma 2.1 is a consequence of Corollary 4.3 in [25].

3. Optimality conditions

To prove average optimality of the adaptive policy constructed in the next section, we now need to impose conditions that ensure the existence of a solution to the ACOI, i.e., a pair $(j^*, h(\cdot))$ consisting of a real number j^* and a measurable function $h : X \to \Re$, satisfying, for all $x \in X$,

$$j^* + h(x) \geq \inf_{A(x)} \left[c(x,a) + \int_{\Re^k} h[F(x,a,s)]\rho(s)ds \right]. \tag{3.1}$$

In this section we state an average cost condition (AC - condition) ensuring (3.1).

Let $V_\alpha(\cdot)$ be the optimal $\alpha-$discounted cost (see (2.4)). Define $m_\alpha := \inf_x V_\alpha(x)$ and $g_\alpha(x) := V_\alpha(x) - m_\alpha$ for $x \in X$ and $\alpha \in (0,1)$.

Condition 3.1 (C1)

(i) There exists $\alpha^* \in [0,1)$ such that $\sup_{\alpha^*<\alpha<1} g_\alpha(x) < +\infty$ for every $x \in X$.

(ii) *There exist* $p > 1$, $\beta_0 < 1$ *and* $b_0 < \infty$, *such that, for every* $x \in X$ *and* $a \in A(x)$,

$$\int_{\Re^k} W^p[F(x,a,s)]\rho(s)ds \leq \beta_0 W^p(x) + b_0. \tag{3.2}$$

The Condition C1 is a combination of assumptions used in [14] and [19]. Indeed, supposing that

$$J(\bar{\pi}, \bar{x}) < \infty \text{ for some } \bar{\pi} \in \Pi \text{ and } \bar{x} \in X, \tag{3.3}$$

the Condition C1(i) was used in [14] (see also [4, 27]) to prove the existence of a solution to the ACOI, while C1(ii) is variant of a condition used by Lippman in [19] (see also [28]) to study semi-Markov control processes. Nevertheless, we can use C1 as a sufficient condition for (3.1) since, as is observed in Remark 3.1(i) below, (3.3) is a consequence of Condition C1(ii).

A comparison between several AC — optimality conditions has been presented in [24]. From these results we can deduce the equivalence of Condition C1 and the following Condition C1$'$.

Let $z \in X$ be an arbitrary, but fixed state. Define

$$h_\alpha(x) := V_\alpha(x) - V_\alpha(z) \text{ for } x \in X,\ \alpha \in (0,1). \tag{3.4}$$

Condition 3.2 (C1$'$) *There exist nonnegative constants* N *and* M, *a nonnegative (not necessarily measurable) function* G *on* X, *and* $\alpha^* \in (0,1)$ *such that*

(i) $(1-\alpha)V_\alpha(z) \leq M$ *for all* $\alpha \in [\alpha^*, 1)$;

(ii) $-N \leq h_\alpha(x) \leq G(x)$ *for every* $x \in X$ *and* $\alpha \in [\alpha^*, 1)$;

(iii) *Condition C1(ii) holds.*

Conditions C1$'$(i)–(ii) together with the assumption

$$V_\alpha(x) < \infty \text{ for every } x \in X \text{ and } \alpha \in (0,1), \tag{3.5}$$

were introduced in [26] for countable-state MCPs with finite control sets, and were extended to the Borel space case in [23]. Again, from Remark 3.1(i) below, (3.5) is a consequence of Condition C1$'$(iii).

Lemma 3.1 [9] *Suppose that* Assumption 2.1(i) *holds. Then* Condition C1(ii) *implies,*

(i) *for every* $(x,a) \in I\!\!K$

$$\int_{\Re^k} W[F(x,a,s)]\rho(s)ds \le \beta W(x) + b, \tag{3.6}$$

where $\beta = \beta_0^{1/p}$ *and* $b = b_0^{1/p}$;

(ii) $\sup_{t\ge 1} E_x^\pi[W(x_t)] < \infty$ *and* $\sup_{t\ge 1} E_x^\pi[W^p(x_t)] < \infty \ \forall \pi \in \Pi,\ x \in X$.

Remark 3.1

(i) *From* Assumption 2.1(i) *and* Lemma 3.1(ii), *it is easy to see that* $V_\alpha(\pi,x) < \infty$ *and* $J(\pi,x) < \infty$ *for each* $x \in X$, $\pi \in \Pi$. *In fact, in* [13] *it is proved that if* (3.6) *holds, then, under* Assumption 2.1(i), *we have*

$$V_\alpha(x) \le CW(x)/(1-\alpha)\ \forall x \in X,\ \alpha \in (0,1), \tag{3.7}$$

for some constant $C > 0$.

(ii) *Let* W *be the function introduced in* Assumption 2.1. *We denote by* L_W^∞ *the normed linear space of all measurable functions* $u : X \to \Re$ *with*

$$\|u\|_W := \sup_{x\in X} \frac{|u(x)|}{W(x)} < \infty. \tag{3.8}$$

Thus, from (3.7), $V_\alpha \in L_W^\infty$ *for all* $\alpha \in (0,1)$.

(iii) *Therefore, by* Condition C1 *and the fact that* $h_\alpha(\cdot) \le g_\alpha(\cdot)$ *for* $\alpha \in (\alpha^*,1)$,

$$\sup_{\alpha\in(\alpha^*,1)} \|h_\alpha\|_W < \infty. \tag{3.9}$$

The main conclusion of this section can now be stated as follows:

Theorem 3.1 *Suppose that* Assumption 2.1 *holds. Then the* Condition C1 *(or* C1′*) implies the existence of a solution* (j^*,h) *to the ACOI* (3.1) *with* $h \in L_W^\infty$. *Moreover,* j^* *is the optimal average cost, i.e.,* $j^* = \inf_{\pi\in\Pi} J(\pi,x)$ *for all* $x \in X$.

Remark 3.2 *Fix an arbitrary state* $z \in X$, *and let* $j_\alpha := (1-\alpha)V_\alpha(z)$ *for* $\alpha \in (0,1)$. *Then, following standard arguments in the literature on average cost MCPs (see, e.g.,* [6], [15], [23]*) it is possible to prove that*

$$\lim_{t\to\infty} j_{\alpha_t} = j^* \tag{3.10}$$

for any sequence $\{\alpha_t\}$ *of discount factor such that* $\alpha_t \nearrow 1$.

4. Adaptive policy

To construct the adaptive policy, we present first a method of statistical estimation of ρ, and then extend to the estimators ρ_t of ρ some assertions in the previous sections. This density estimation scheme was originally proposed in [9] to obtain an asymptotically discount optimal adaptive policy, and used again in [10] to construct an average optimal iterative adaptive policy under ergodicity assumption on the control model.

Density estimation. Let $\xi_0, \xi_1, ..., \xi_{t-1}$ be independent realizations (observed up to time $t-1$) of r.v.'s with the unknown density ρ. We suppose that ρ satisfies Assumption 2.2 and relation (3.6).

Let $\hat{\rho}_t(s) := \hat{\rho}_t(s; \xi_0, \xi_1, ..., \xi_{t-1})$, for $s \in \Re^k$, be an arbitrary estimator of ρ belonging to $L_q(\Re^k)$, and such that for some $\gamma > 0$

$$E\,\|\rho - \hat{\rho}_t\|_q^{\frac{qp'}{2}} = \mathbf{O}(t^{-\gamma}) \ \text{ as } \ t \to \infty, \tag{4.1}$$

where $1/p + 1/p' = 1$. To construct an average cost optimal adaptive policy we can not use, in general, the estimators $\hat{\rho}_t$ because they might not satisfy the right assumptions. Therefore, we estimate ρ by the projection ρ_t of $\hat{\rho}_t$ on the set of densities D in $L_q(\Re^k)$ defined as follows:

$$\begin{aligned} D := \{\mu \in L_q(\Re^k) : \mu \text{ is a density function}, \ \mu(s) \le \bar{\rho}(s) \ a.e., \text{ and} \\ \textstyle\int W[F(x,a,s)]\mu(s)ds \le \beta W(x) + b \ \forall (x,a) \in \mathbb{K}\}. \end{aligned} \tag{4.2}$$

See Lemma 3.1(i) for the constants β and b.

The existence (and uniqueness) of the estimator ρ_t is guaranteed because the set D is convex and closed in $L_q(\Re^k)$ [9]. Moreover, Assumption 2.1(i) and (3.6) ensure that the unknown density ρ is in D.

Examples of estimators satisfying (4.1) are given in [11]. On the other hand, from [9, 10] we known the following.

Lemma 4.1 [9, 10] *Suppose that* Assumptions 2.2 and 2.3 *hold. Then*

$$E\,\|\rho_t - \rho\|^{p'} = \mathbf{O}(t^{-\gamma}) \quad \text{as } t \to \infty, \tag{4.3}$$

where $\|\cdot\|$ *is the pseudo-norm on the space of all densities* μ *on* $\Re^k$ *defined as:*

$$\|\mu\| := \sup_X [W(x)]^{-1} \sup_{A(x)} \int_{\Re^k} W[F(x,a,s)]\mu(s)ds. \tag{4.4}$$

For an arbitrary density μ in $\Re^k$, the pseudo norm $\|\mu\|$ may be infinite. However, by (4.2), $\|\mu\| < \infty$ for μ in D.

Construction of the adaptive policy. Having the estimators ρ_t of ρ, we now define an adaptive control policy as follows.

Let $\{\alpha_t\}$ be an arbitrary nondecreasing sequence of discount factors such that $\alpha_t \nearrow 1$. For each fixed t, let

$$V_{\alpha_t}^{(\rho_t)}(\pi, x) := E_x^{\pi,\rho_t}\left[\sum\nolimits_{n=0}^{\infty} \alpha_t^n c(x_n, a_n)\right],$$

be the total expected α_t-discounted cost for the process (1.1) in which the r.v.'s $\xi_0, \xi_1, \ldots,$ have the common density ρ_t, and let $V_{\alpha_t}^{(\rho_t)}(x) := \inf_{\pi \in \Pi} V_{\alpha_t}^{(\rho_t)}(\pi, x)$, $x \in X$, be the corresponding value function. The sequences $h_{\alpha_t}^{(\rho_t)}(\cdot)$ and $j_{\alpha_t}^{(\rho_t)}$ are defined accordingly (see (3.4) and Remark 3.2).

Remark 4.1

(i) *The proof of* (3.7) *(given in* [13]*) shows that under* Assumption 2.1(i) *the following relations hold (because only inequality* (3.6) *is used here):*

$$V_{\alpha_t}(x) \le \frac{C}{1-\alpha_t} W(x), \quad V_{\alpha_t}^{(\rho_t)}(x) \le \frac{C}{1-\alpha_t} W(x), \; x \in X, \; t \in \mathbb{N}. \tag{4.5}$$

(ii) *For each $t \in \mathbb{N}$ and each density $\mu \in D$, we define the operator $T_{\mu,\alpha_t} \equiv T_\mu : L_W^\infty \to L_W^\infty$ as*

$$T_\mu u(x) := \inf_{A(x)} \left\{ c(x,a) + \alpha_t \int_{\Re^k} u[F(x,a,s)]\mu(s)\, ds \right\}, \tag{4.6}$$

for $x \in X$, $u \in L_W^\infty$. Now, under Assumption 2.1, *from* Lemma 2.1(i) *we have $T_\rho V_{\alpha_t} = V_{\alpha_t}$ and $T_{\rho_t} V_{\alpha_t}^{(\rho_t)} = V_{\alpha_t}^{(\rho_t)}$ for each $t \in \mathbb{N}$.*

(iii) *Moreover, from* Lemma 2.1(ii), *for each $t \in \mathbb{N}$ and $\delta_t > 0$, there exists a policy $f_t \in \mathbb{F}$ such that*

$$c(x, f_t) + \alpha_t \int_{\Re^k} V_{\alpha_t}^{(\rho_t)}[F(x, f_t, s)]\rho_t(s) ds \le V_{\alpha_t}^{(\rho_t)}(x) + \delta_t, \qquad x \in X. \tag{4.7}$$

We suppose that Condition C1 holds. To define the adaptive policy, first we fix an arbitrary nondecreasing sequence of discount factors $\{\hat{\alpha}_t\}$ on $(\alpha^*, 1)$ (see C1(a)) such that $1 - \hat{\alpha}_t = \mathbf{O}(t^{-\nu})$ as $t \to \infty$, and

$$\lim_{n\to\infty} \frac{\kappa(n)}{n} = 0, \tag{4.8}$$

where $0 < \nu < \gamma/(3p')$ (with γ and p' as in (4.3)) and $\kappa(n)$ is the number of changes of value of $\{\hat{\alpha}_t\}$ for $t = 0, 1, ..., n$.

Definition 4.1 *Let* $\left\{\hat{\delta}_t\right\}$ *be an arbitrary convergent sequence of positive numbers, and let* $\hat{\delta} := \lim_{t\to\infty} \hat{\delta}_t$. *In addition, let* $\left\{\hat{f}_t\right\}$ *be a sequence of functions in* $\mathbb{F}$ *satisfying* (4.7) *with* $\hat{\alpha}_t$ *instead of* α_t. *The adaptive policy* $\hat{\pi} = \{\hat{\pi}_t\}$ *is defined as* $\hat{\pi}_t(h_t) = \hat{\pi}_t(h_t; \rho_t) := \hat{f}_t(x_t)$ *for each* $t \in \mathbb{N}$, *where* $\hat{\pi}_0(x)$ *is any fixed action in* $A(x)$.

We are now ready to state our main result.

Theorem 4.1 *Suppose that* Assumptions 2.1, 2.2 *and* 2.3 *hold. Then, under* Condition C1 *(or* C1′*), the adaptive policy* $\hat{\pi}$ *is* $\hat{\delta}$*−average cost optimal, that is,* $J(\hat{\pi}, x) \le j^* + \hat{\delta}$ *for all* $x \in X$, *where* j^* *is the optimal average cost in* Theorem 3.1. *In particular, if* $\hat{\delta} = 0$, *then the policy* $\hat{\pi}$ *is average cost optimal.*

Throughout the proof of this theorem we will repeatedly use the following inequalities. For any $u \in L_W^\infty$ and any μ that satisfies (3.6), we have

$$|u(x)| \le \|u\|_W W(x) \tag{4.9}$$

and

$$\int_{\Re^k} u[F(x,a,s)]\mu(s)ds \le \|u\|_W [\beta W(x) + b] \tag{4.10}$$

for all $x \in X$ and $a \in A(x)$. The relation (4.9) is a consequence of the definition (3.8) of $\|\cdot\|_W$, and (4.10) holds because of (3.6) and (4.9).

The proof of Theorem 4.1 is based on the following lemma:

Lemma 4.2 *Under* Assumptions 2.1, 2.2 *and* 2.3, *and* Condition C1 *(or* C1′*), for each* $x \in X$ *and* $\pi \in \Pi$, *we have*

$$\lim_{t\to\infty} E_x^\pi \left\| V_{\hat{\alpha}_t} - V_{\hat{\alpha}_t}^{(\rho_t)} \right\|_W^{p'} = 0. \tag{4.11}$$

Proof of Lemma 4.2 For each $t \in \mathbb{N}$, define $\theta_t \in (\hat{\alpha}_t, 1)$ as $\theta_t := (1+\hat{\alpha}_t)/2$, and let $W_t(x) := W(x)+d_t$ for $x \in X$, where $d_t := b\,(\theta_t/\hat{\alpha}_t - 1)^{-1}$. Let $L^\infty_{W_t}$ be the space of measurable functions $u : X \to \Re$ with the norm

$$\|u\|_{W_t} := \sup_{x\in X} \frac{|u(x)|}{W_t(x)} < \infty, \quad t \in \mathbb{N}.$$

Using the fact that $d_t \leq 2b/(1-\hat{\alpha}_t)$, $t \in \mathbb{N}$, it is easy to see that

$$\|u\|_{W_t} \leq \|u\|_W \leq l_t \|u\|_{W_t}, \quad t \in \mathbb{N},$$

where $l_t := 1 + 2b/\left[(1-\hat{\alpha}_t)\inf_{x\in X} W(x)\right]$. Thus, (4.11) will follow if we show that

$$l_t^{p'} E_x^{\hat{\pi}} \left\| V_{\hat{\alpha}_t} - V_{\hat{\alpha}_t}^{(\rho_t)} \right\|_{W_t}^{p'} \to 0, \text{ as } t \to \infty. \tag{4.12}$$

A consequence of Lemma 2 in [28] is that, for each $t \in \mathbb{N}$ and $\mu \in D$, the inequality $\int_{\Re^k} W[F(x,a,s)]\mu(s)ds \leq W(x) + b$ implies that the operator T_μ defined in (4.6) is a contraction with respect to the norm $\|\cdot\|_{W_t}$, with modulus θ_t, i.e.,

$$\|T_\mu v - T_\mu u\|_{W_t} \leq \theta_t \|v-u\|_{W_t} \qquad \forall v,u \in L^\infty_W,\ t \in \mathbb{N}. \tag{4.13}$$

Hence, from (4.13) and Remark 4.1(ii) we can see that

$$\left\| V_{\hat{\alpha}_t} - V_{\hat{\alpha}_t}^{(\rho_t)} \right\|_{W_t} \leq \|T_\rho V_{\hat{\alpha}_t} - T_{\rho_t} V_{\hat{\alpha}_t}\|_{W_t} + \theta_t \left\| V_{\hat{\alpha}_t} - V_{\hat{\alpha}_t}^{(\rho_t)} \right\|_{W_t},$$

which implies that

$$l_t \left\| V_{\hat{\alpha}_t} - V_{\hat{\alpha}_t}^{(\rho_t)} \right\|_{W_t} \leq \frac{l_t}{1-\theta_t} \|T_\rho V_{\hat{\alpha}_t} - T_{\rho_t} V_{\hat{\alpha}_t}\|_{W_t} \qquad \forall t \in \mathbb{N}. \tag{4.14}$$

On the other hand, from definition (4.4), (4.5), and the fact that $[W_t(\cdot)]^{-1} < [W(\cdot)]^{-1}$ for all $t \in \mathbb{N}$, we obtain

$$\begin{aligned}
&\|T_\rho V_{\hat{\alpha}_t} - T_{\rho_t} V_{\hat{\alpha}_t}\|_{W_t} \\
&\leq \hat{\alpha}_t \sup_X [W_t(x)]^{-1} \sup_{A(x)} \int_{\Re^k} V_{\hat{\alpha}_t}[F(x,a,s)]\,|\rho(s) - \rho_t(s)|\,ds \\
&\leq \frac{C\hat{\alpha}_t}{1-\hat{\alpha}_t} \sup_X [W(x)]^{-1} \sup_{A(x)} \int_{\Re^k} W[F(x,a,s)]\,|\rho(s) - \rho_t(s)|\,ds \\
&\leq \frac{C}{1-\hat{\alpha}_t} \|\rho - \rho_t\|.
\end{aligned} \tag{4.15}$$

Now, observe that (by the definition of $\hat{\alpha}_t$ and θ_t),

$$\frac{1}{(1-\theta_t)(1-\hat{\alpha}_t)^2} = \mathbf{O}(t^{3\nu}) \textit{ as } t \to \infty. \tag{4.16}$$

Combining (4.14), (4.15), (4.16), and using the definition of l_t, we get

$$\begin{aligned} & l_t^{p'} \left\| V_{\hat{\alpha}_t} - V_{\hat{\alpha}_t}^{(\rho_t)} \right\|_{W_t}^{p'} \\ \leq \; & C^{p'} \left[\frac{1}{(1-\theta_t)(1-\hat{\alpha}_t)} + \frac{2b}{(1-\theta_t)(1-\hat{\alpha}_t)^2 \inf_X W(x)} \right]^{p'} \|\rho - \rho_t\|^{p'} \\ = \; & C^{p'} \mathbf{O}(t^{3p'\nu}) \|\rho - \rho_t\|^{p'} \text{ as } t \to \infty. \end{aligned} \tag{4.17}$$

Finally, taking expectation E_x^π on both sides of (4.17), and observing that $E_x^\pi \|\rho - \rho_t\|^{p'} = E \|\rho - \rho_t\|^{p'}$ (since ρ_t does not depend on x and π), we obtain (4.12) by virtue of Lemma 4.1 and the fact that $3\nu p' < \gamma$ (see the definition of $\hat{\alpha}_t$). This proves the Lemma. ■

Remark 4.2 *It is easy to prove that*

$$\lim_{t\to\infty} E_x^\pi \left\| V_{\hat{\alpha}_t} - V_{\hat{\alpha}_t}^{(\rho_t)} \right\|_W W(x_t) = 0 \text{ for } x \in X, \; \pi \in \Pi. \tag{4.18}$$

Indeed, denoting $\bar{C} := (E_x^\pi [W^p(x_t)])^{1/p} < \infty$ *[see* Lemma 3.1(ii)*], applying Hölder's inequality, and using* Lemma 4.2, *we obtain*

$$\begin{aligned} E_x^\pi \left\| V_{\hat{\alpha}_t} - V_{\hat{\alpha}_t}^{(\rho_t)} \right\|_W W(x_t) & \leq \bar{C} \left(E_x^\pi \left[\left\| V_{\hat{\alpha}_t} - V_{\hat{\alpha}_t}^{(\rho_t)} \right\|_W^{p'} \right] \right)^{1/p'} \\ & \to 0 \text{ as } t \to \infty. \end{aligned}$$

Proof of Theorem 4.1. Let $\{k_t\} := \{(x_t, a_t)\}$ be a sequence of state-action pairs corresponding to applications of the adaptive policy $\hat{\pi}$. We define

$$\begin{aligned} \Phi_t & := c(k_t) + \hat{\alpha}_t \int_{\Re^k} V_{\hat{\alpha}_t}[F(k_t, s)]\rho(s)\, ds - V_{\hat{\alpha}_t}(x_t) \\ & = c(k_t) + \hat{\alpha}_t E_x^{\hat{\pi}} \left[V_{\hat{\alpha}_t}(x_{t+1}) \mid k_t \right] - V_{\hat{\alpha}_t}(x_t). \end{aligned} \tag{4.19}$$

From definition of h_α and j_α (see (3.4) and Remark 3.2), it is easy to see that

$$\Phi_t = c(k_t) + \hat{\alpha}_t E_x^{\hat{\pi}} \left[h_{\hat{\alpha}_t}(x_{t+1}) \mid k_t \right] - j_{\hat{\alpha}_t} - h_{\hat{\alpha}_t}(x_t).$$

Hence, for $n \geq k \geq 1$

$$\begin{aligned} & n^{-1} E_x^{\hat{\pi}} \left[\sum_{t=k}^n c(k_t) - j_{\hat{\alpha}_t} \right] \\ = \; & n^{-1} E_x^{\hat{\pi}} \left[\sum_{t=k}^n \left(h_{\hat{\alpha}_t}(x_t) - \hat{\alpha}_t h_{\hat{\alpha}_t}(x_{t+1}) \right) \right] + n^{-1} E_x^{\hat{\pi}} \left[\sum_{t=k}^n \Phi_t \right]. \end{aligned} \tag{4.20}$$

On the other hand, from Lemma 3.1(ii), (3.9) and (4.9), we have $E_x^{\hat{\pi}}[h_\alpha(x_t)] < C'$ for $\alpha \in (\alpha^*, 1)$ and a constant $C' < \infty$. Thus, denoting $\alpha_1^*, \alpha_2^*, \dots, \alpha_{\kappa(n)}^*$, $n \geq 1$, the different values of $\hat{\alpha}_t$ for $t \leq n$, and using that $\{\hat{\alpha}_t\}$ is a nondecreasing sequence we have (see condition (4.8) and the definition of h_α)

$$\begin{aligned}
&n^{-1}E_x^{\hat{\pi}}\left[\sum_{t=k}^{n}(h_{\hat{\alpha}_t}(x_t) - \hat{\alpha}_t h_{\hat{\alpha}_t}(x_{t+1}))\right] \\
&= n^{-1}E_x^{\hat{\pi}}\left[\sum_{t=k}^{n}(h_{\hat{\alpha}_t}(x_t) - \hat{\alpha}_t h_{\hat{\alpha}_t}(x_t))\right] \\
&\quad + n^{-1}E_x^{\hat{\pi}}\left[\sum_{t=k}^{n}\hat{\alpha}_t\left(h_{\hat{\alpha}_t}(x_t) - h_{\hat{\alpha}_t}(x_{t+1})\right)\right] \\
&\leq (1-\alpha_k)C' + n^{-1}2C'\sum_{i=1}^{\kappa(n)}\alpha_i^* \\
&\leq (1-\hat{\alpha}_k)C' + 2C'\kappa(n)n^{-1} \\
&\leq (1-\hat{\alpha}_k)C' + \mathbf{o}(1), \ x \in X. \qquad (4.21)
\end{aligned}$$

Now, from (4.19) and (2.6) we have

$$\begin{aligned}
\Phi_t &= c(k_t) + \hat{\alpha}_t\int_{\Re^k}V_{\hat{\alpha}_t}[F(k_t,s)]\rho(s)\,ds \\
&\quad - \inf_{A(x_t)}\left[c(x_t,a) + \hat{\alpha}_t\int_{\Re^k}V_{\hat{\alpha}_t}[F(x_t,a,s)]\rho(s)\,ds\right] \\
&\leq |I_1(t)| + |I_2(t)| + |I_3(t)|,
\end{aligned}$$

where

$$\begin{aligned}
I_1(t) &:= \hat{\alpha}_t\int_{\Re^k}V_{\hat{\alpha}_t}[F(k_t,s)]\rho(s)\,ds - \hat{\alpha}_t\int_{\Re^k}V_{\hat{\alpha}_t}^{(\rho_t)}[F(k_t,s)]\rho(s)\,ds, \\
I_2(t) &:= \hat{\alpha}_t\int_{\Re^k}V_{\hat{\alpha}_t}^{(\rho_t)}[F(k_t,s)]\rho(s)\,ds - \hat{\alpha}_t\int_{\Re^k}V_{\hat{\alpha}_t}^{(\rho_t)}[F(k_t,s)]\rho_t(s)\,ds, \\
I_3(t) &:= c(k_t) + \hat{\alpha}_t\int_{\Re^k}V_{\hat{\alpha}_t}^{(\rho_t)}[F(k_t,s)]\rho_t(s)\,ds \\
&\quad - \inf_{A(x_t)}\left[c(x_t,a) + \hat{\alpha}_t\int_{\Re^k}V_{\hat{\alpha}_t}[F(x_t,a,s)]\rho(s)\,ds\right]
\end{aligned}$$

Using (4.9) and (4.10)

$$\begin{aligned} |I_1(t)| &\leq \hat{\alpha}_t \int_{\Re^k} \left| V_{\hat{\alpha}_t}[F(k_t,s)] - V_{\hat{\alpha}_t}^{(\rho_t)}[F(k_t,s)] \right| \rho(s)\, ds \\ &\leq \hat{\alpha}_t \left\| V_{\hat{\alpha}_t} - V_{\hat{\alpha}_t}^{(\rho_t)} \right\|_W [\beta W(x_t) + b]. \end{aligned} \tag{4.22}$$

Taking expectation $E_x^{\hat{\pi}}$ on both sides of (4.22) and using the Lemma 4.2 and (4.18), we get

$$E_x^{\hat{\pi}} |I_1(t)| \to 0, \text{ as } t \to \infty. \tag{4.23}$$

Now, from definition of $\hat{\alpha}_t$ and (4.5), $\left\| V_{\hat{\alpha}_t}^{(\rho_t)} \right\|_W = \mathbf{O}(t^\nu)$. Thus, from definition (4.4),

$$\begin{aligned} |I_2(t)| &\leq \hat{\alpha}_t \int_{\Re^k} V_{\hat{\alpha}_t}^{(\rho_t)}[F(k_t,s)]\, |\rho(s) - \rho_t(s)|\, ds \\ &\leq \hat{\alpha}_t W(x_t) \left\| V_{\hat{\alpha}_t}^{(\rho_t)} \right\|_W \|\rho - \rho_t\|. \end{aligned} \tag{4.24}$$

Hence, taking expectation and applying Hölder's inequality in (4.24) we get

$$\begin{aligned} E_x^{\hat{\pi}} |I_2(t)| &\leq \left([\mathbf{O}(t^\nu)]^{p'} E_x^{\hat{\pi}} \|\rho - \rho_t\|^{p'} \right)^{1/p'} \\ &= \left[\mathbf{O}(t^{\nu p' - \gamma}) \right]^{1/p'} \to 0 \text{ as } t \to \infty, \end{aligned} \tag{4.25}$$

due to the fact $\nu < \gamma/p'$ (see definition of $\hat{\alpha}_t$).

For the term $|I_3(t)|$, from the definition of the policy $\hat{\pi}$ and combining (2.6) and (4.7), adding and subtracting the term

$$\inf_{A(x_t)} \left\{ c(x_t,a) + \hat{\alpha}_t \int_{\Re^k} V_{\hat{\alpha}_t}^{(\rho_t)}[F(x_t,a,s)]\rho_t(s)\, ds \right\}$$

in $I_3(t)$, we get

$$\begin{aligned} |I_3(t)| \leq \hat{\delta}_t + \hat{\alpha}_t \sup_{A(x_t)} \bigg| &\int_{\Re^k} V_{\hat{\alpha}_t}^{(\rho_t)}[F(x_t,a,s)]\rho_t(s)\, ds \\ &- \int_{\Re^k} V_{\hat{\alpha}_t}[F(x_t,a,s)]\rho(s)\, ds \bigg| \end{aligned}$$

The latter inequality yields

$$\begin{aligned} |I_3(t)| &\leq \hat{\delta}_t + \hat{\alpha}_t \sup_{A(x_t)} \int_{\Re^k} V_{\hat{\alpha}_t}^{(\rho_t)}[F(x_t,a,s)]\, |\rho(s) - \rho_t(s)|\, ds \\ &\quad + \hat{\alpha}_t \sup_{A(x_t)} \int_{\Re^k} \left| V_{\hat{\alpha}_t}^{(\rho_t)}[F(x_t,a,s)] - V_{\hat{\alpha}_t}[F(x_t,a,s)] \right| \rho(s)\, ds. \end{aligned}$$

Thus, from (4.4),

$$\begin{aligned} |I_3(t)| &\leq \hat{\delta}_t + \hat{\alpha}_t W(x_t) \left\| V_{\hat{\alpha}_t}^{(\rho_t)} \right\|_W \|\rho - \rho_t\| \\ &\quad + \hat{\alpha}_t \left\| V_{\hat{\alpha}_t} - V_{\hat{\alpha}_t}^{(\rho_t)} \right\|_W [\beta W(x) + b]. \end{aligned}$$

Hence, from (4.22), (4.23), (4.24) and (4.25), we get $E_x^{\hat{\pi}} |I_3(t)| \to \hat{\delta}$, as $t \to \infty$. Therefore

$$E_x^{\hat{\pi}}[\Phi_t] \to \hat{\delta}, \text{ as } t \to \infty. \tag{4.26}$$

Finally, from (4.20), (4.21) and (4.26), we have for any $k \geq 1$ and $n \to \infty$,

$$n^{-1} E_x^{\hat{\pi}} \left[\sum_{t=k}^{n} c(k_t) - j_{\hat{\alpha}_t} \right] = (1 - \hat{\alpha}_k) C' + \mathbf{o}(1) + \hat{\delta}, \ x \in X.$$

It follows that (from (3.10), the fact that $\lim_{t\to\infty} \hat{\alpha}_t = 1$, and (2.3))

$$J(\hat{\pi}, x) \leq j^* + \hat{\delta}, \ x \in X.$$

This completes the proof of the theorem. ■

5. Additional optimality conditions

Besides Assumptions 2.1–2.3, the proof of the Theorem 4.1, as well as Lemma 4.2, is based on the following:

(i) the existence of a solution (j^*, h) to the ACOI (3.1) where j^* satisfies (3.10);

(ii) the Lippman-like hypotheses (3.2) which yields the results of Lemma 3.1;

(iii) the relation (3.9) in Remark 3.1(iii).

Thus, the average optimality of the adaptive policy constructed in the previous section (see Definition 4.1) can be proved under any condition ensuring the points (i)–(iii) (for instance C1 or C1′). In this section we state two additional sets of such conditions (C2 and C3), which are variants of conditions used in previous works to study either non-adaptive MCPs or non-controlled Markov processes. We illustrate the Conditions C1–C3, as well as our assumptions and main result, with three examples given in the next section.

Let us denote by M_W the space of all signed measure m on $\mathbb{B}(X)$ with a finite W−norm (see [16, 17]), which is defined as

$$\|m\|_{M_W} := \int_X W(x)\,|m|\,(dx), \tag{5.1}$$

where $|m|$ denotes the variation of the measure m.

Condition 5.1 (C2)

(i) There is a number $\beta_0 < 1$ such that, for some $p > 1$,

$$\left\|Q_\rho(\cdot \mid x,a) - Q_\rho(\cdot \mid x',a')\right\|_{M_{W^p}} \leq \beta_0 \left[W^p(x) + W^p(x')\right],$$

for each $x,\ x' \in X,\ a \in A(x),\ a' \in A(x')$.

(ii) There are $x^* \in X,\ a^* \in A(x^*)$ such that

$$\|Q_\rho(\cdot \mid x^*,a^*)\|_{M_{W^p}} < \infty.$$

Observe that Condition C2(i) is a generalization, to unbounded costs case, of the well-known ergodicity assumption (see, for instance, [12, 16]):

$$\left\|Q_\rho(\cdot \mid x,a) - Q_\rho(\cdot \mid x',a')\right\|_\tau \leq 2\beta_0,$$

for each $x,\ x' \in X,\ a \in A(x),\ a' \in A(x')$, where $\beta_0 < 1$ and $\|\cdot\|_\tau$ denotes the variation norm for signed measures, which is the same as (5.1) with $W(\cdot) \equiv 1$.

Condition 5.2 (C3) *There exists a probability measure m on $(X,\mathbb{B}(X))$ and a nonnegative number $\beta_0 < 1$ and, for every $f \in \mathbb{F}$, a nonnegative function $\psi_f : X \to \Re$ such that for any $x \in X$ and $B \in \mathbb{B}(X)$,*

(i) $Q_\rho(B \mid x,f) \geq \psi_f(x)m(B)$;

(ii) $\int_{\Re^k} W^p[F(x,f,s)]\rho(s)\,ds \leq \beta_0 W^p(x) + \psi_f(x)b_0$ for some $p > 1$, with $b_0 := \int_X W^p(y)m(dy) < \infty$;

(iii) $\inf_{f \in \mathbb{F}} \int_X \psi_f(x) m(dx) := \bar{\psi} > 0.$

Hypotheses of the type C2 and C3 were introduced in [17] for non-controlled Markov processes. For discrete-time average cost MCPs, Conditions similar to C2 and C3 were used, respectively, in [8] and [6] to show the existence of a solution to the ACOI, (see also [16] for a detailed study of these conditions). The procedure used in those works [6, 8, 16] is the so-called vanishing discount factor approach.

In contrast to C1, the key feature of C2, as well as C3, is that it ensures the geometric ergodicity of the state process with respect to the norm (5.1), when using stationary policies. Now, having geometric ergodicity we obtain (3.9) and (3.10) by a standard procedure (see [6, 8, 16]).

Finally, straightforward calculations show that each of the Conditions C2 and C3 implies the Lippman-like hypotheses (3.2).

6. Examples

In this section we consider special cases of the controlled system (1.1). To simplify the exposition we shall assume that the random disturbances ξ_0, $\xi_1, \ldots$, are real-valued, i.i.d. random variables with an unknown density $\rho \in L_q(\Re)$ that satisfies Assumption 2.2(ii).

In fact, when $k = 1$ it is not difficult to show (see [22, page 13]) that a sufficient condition for Assumption 2.2(ii) is the following: There is a finite set $H \subset \Re$ (possibly empty) and a constant $M \geq 0$ such that:

(i) ρ has a bounded derivative ρ' on $\Re \backslash H$ that belongs to $L_q(\Re)$;

(ii) the function $|\rho'(x)|$ is non-increasing for $x \geq M$ and nondecreasing for $x \leq -M$.

Note that H might include points of discontinuity of ρ if such points exist.

6.1 Invariant problems

A control problem is called invariant if the transition kernel $Q_\rho(\cdot \mid x, a)$ depends only on the control a; that is $Q_\rho(\cdot \mid x, a) = Q_\rho(\cdot \mid a)$ (see, e.g., [1, 27]). In this case the dynamics of the system can be represented as $x_{t+1} = F(a_t, \xi_t)$ for $t = 0, 1, \ldots$.

We consider a invariant control problem with state space $X = [0, \infty)$ and finite actions set $A(x) = A$, $x \in X$. We suppose that the random variables ξ_0, $\xi_1, \ldots$, are non-negative with a density ρ satisfying

$$\rho(s) \leq M_1 \exp(-\alpha s), \ \ s \geq 0,$$

for some constants $M_1 < \infty$ and $0 < \alpha < 1$.

Assumption 2.2(iii) is satisfied with $\bar{\rho}(s) := M_1 \exp(-\alpha s)$, $s \geq 0$; and taking, in particular, $W(x) := M_2 \exp(-\lambda x)$, $x \in [0, \infty)$ for some constants $M_2 > 0$ and $\lambda < \alpha(1-2\varepsilon)/2$, and supposing that $F(a^*, s) < s$, $s \geq 0$, with $a^* := \max A$, it is readily seen that Assumption 2.3(ii) (as well as Assumption 2.3(i) is satisfied.

The Assumption 2.1 is satisfied taking the one-stage cost c as any non-negative, l.s.c. function satisfying $\sup_A c(x,a) \leq M_2 \exp(-\lambda x)$ for $x \in [0, \infty)$.

Proposition 6.1 *The invariant problem satisfies the* Condition C1 *(or* C1′*).*

Proof From Lemma 2.1(i) and the fact that A is finite, there exists $\hat{a} \in A$ such that

$$V_\alpha(x) = c(x, \hat{a}) + \alpha \int_{\Re} V_\alpha[F(\hat{a}, s)]\rho(s)\, ds, \quad x \in X.$$

Thus, from definition of g_α, we have

$$\begin{aligned} g_\alpha(x) &= V_\alpha(x) - m_\alpha \\ &= c(x, \hat{a}) + \alpha \int_{\Re} V_\alpha[F(\hat{a}, s)]\rho(s)\, ds \\ &\quad - \inf_{x \in X} c(x, \hat{a}) - \alpha \int_{\Re} V_\alpha[F(\hat{a}, s)]\rho(s)\, ds \\ &= c(x, \hat{a}) - \inf_{x \in X} c(x, \hat{a}) \\ &\leq W(x) < \infty, \quad \alpha \in (0,1), \quad x \in X. \end{aligned}$$

Therefore, $\sup_{\alpha \in (0,1)} g_\alpha(x) < \infty$ for every $x \in X$. This yields Condition C1(i).

To conclude, the Condition C1(ii) follows from the fact

$$\int_{\Re} W^p[F(a, s)]\rho(s)\, ds \;\leq\; \max_{a \in A} \int_{\Re} W^p[F(a, s)]\rho(s)\, ds \;\leq\; b_0,$$

for some $p > 1$ and constant $b_0 < \infty$. ■

6.2 An autoregressive-like control process

We consider a process of the form

$$x_{t+1} = (\psi(a_t)x_t + \xi_t)^+, \; t = 0, 1, \ldots, \tag{6.1}$$

$x_0 = x$ given, with state space $X = [0, \infty)$, admissible controls set $A(x) = A$ for every $x \in X$, where $A \subset \Re$ is a compact set, and $\psi : A \to (0, \gamma]$ is a given measurable function with $\gamma < 1/2$. We suppose that the density ρ of the random variables ξ_0, $\xi_1, \ldots$, is a continuous function on $\Re$, satisfying $\rho(s) \le \bar{\rho}(s)$ for $s \in \Re$, where $\bar{\rho}(s) := M_1 \min\left\{1, 1/|s|^{1+r}\right\}$, for some constants $M_1 < \infty$ and $r > 0$, and moreover $E(\xi_0) \le 1/4 - (\gamma - 1)^2$. The one-stage cost c is an arbitrary nonnegative, l.s.c., measurable function satisfying $\sup_{a \in A} c(x, a) \le (x + \delta)^{1/p}$, $x \in X$, for some $p > 1$, where $\delta = (1 - 2\gamma)/2$.

For the latter, Assumptions 2.1, 2.2 and 2.3 are satisfied choosing $W(x) = (x + \delta)^{1/p}$ for $x \in [0, \infty)$ and appropriate $r > 0$ in $\bar{\rho}(\cdot)$.

Proposition 6.2 *The autoregressive-like control process satisfies* Condition C2.

Proof First, observe that

$$\begin{aligned} \int_{\Re} W^p\left[(\psi(a)x + s)^+\right]\rho(s)\,ds &\le \delta P[\psi(a)x + \xi_0 \le 0] + \gamma x + \delta + E[\xi_0] \\ &\le (\gamma + 1/2)(x + \delta) \text{ for } x \in [0, \infty), \\ &\qquad a \in A. \end{aligned}$$

Thus, straightforward calculations of $\|Q_\rho(\cdot \mid x, a) - Q_\rho(\cdot \mid x', a')\|_{M_{W^p}}$ show that Condition C2(i) holds with $\beta_0 = \gamma + 1/2$. Moreover, since $E[\xi_0] < \infty$ the Condition C2(ii) is satisfied. ■

6.3 Controlled queueing systems

We consider a control process of the form

$$x_{t+1} = (x_t + a_t - \xi_t)^+, \quad t = 0, 1, \ldots, \tag{6.2}$$

$x_0 = x$ given, with state space $X = [0, \infty)$ and actions set $A(x) = A$ for every $x \in X$, where A is a compact subset of some interval $(0, \theta]$, with $\theta \in A$.

Equations (6.2) describe, in particular, the model of a single server queueing system of type $GI/D/1/\infty$ with controlled service rates $a_t \in A$. In this case x_t denotes the waiting time of the t^{th} customer, while ξ_t denotes the interarrival time between the t^{th} and the $(t+1)^{th}$ customers.

We suppose the random variables $\xi_0, \xi_1, \ldots$, having continuous density ρ such that $E(\xi_0)$ exist, and moreover

$$E(\xi_0) > \theta. \tag{6.3}$$

The latter assumption ensures ergodicity of the system when using the slowest services: $a_t = \theta\ t \geq 0$.

Proposition 6.3 *The controlled queueing systems satisfies* Assumptions 2.1–2.3 *and* Condition C3.

Proof Defining the function $\Psi(s) := e^{\theta s} E(e^{-s\xi_0})$ we find that (6.3) implies $\Psi'(0) < 0$, and so there is $\lambda > 0$ for which $\Psi(\lambda) < 1$. Also, by continuity of Ψ we can choose $p > 1$ such that

$$\Psi(p\lambda) := \beta_0 < 1. \tag{6.4}$$

To meet Assumption 2.1, we suppose that the one-stage cost $c(x, a)$ is any nonnegative measurable function which is l.s.c. in a and satisfying

$$\sup_A c(x, a) \leq \bar{b} e^{\lambda x}, \text{ for all } x \in [0, \infty),$$

where $\bar{b}$ is an arbitrary positive constant.

Now, supposing that $\rho(s) \leq \bar{\rho}(s)$, where $\bar{\rho}(s) := M_1 \min\{1, 1/s^{1+r}\}$ for all $s \in [0, \infty)$ and some constants $M_1 < \infty$, $r > 0$, and taking $W(x) = \bar{b}e^{\lambda x}$ for all $x \in [0, \infty)$ easy calculations shows that Assumptions 2.2 and 2.3 hold.

Finally, in [7] were taken advantages of (6.4) and definition $\psi_f(x) := P[x + f(x) - \xi_0 \leq 0]$, $f \in \mathbb{F}$, to verify, for this example, the Condition C3. ∎

7. Concluding remarks

In this paper we have constructed an average cost optimal adaptive policy for a class of discrete-time Markov control processes with unbounded costs assuming unknown density of the random disturbance. The basic idea has been to show the existence of $\varepsilon-$ minimizers, for $\varepsilon > 0$, of the discounted cost optimality equation, for which we need not to impose restrictive continuity and compactness assumptions on the control model. The average optimality of the adaptive policy was studied under three different optimality conditions applying the vanishing discount factor approach. The assumptions as well as the conditions of this work have been illustrated with examples of invariant systems, an autoregressive-like control process and a queueing system with controlled service rate.

In general, to construct an adaptive policy for systems of the form (1.1) when the disturbance distribution (say G) is unknown, we must combine statistical estimation methods of G and control procedures. A way to estimate G is to assume that it possesses a density function ρ on $\Re^k$, as in our work, that is,

$$G(B) = \int_B \rho(s)\, ds, \; B \in \mathbb{B}(X).$$

In this sense, the estimation of G is based on the estimation of the density function ρ, which in turn can be analyzed in a number of ways. This method has the disadvantage of excluding the case in which the disturbance distribution G is discrete, as can happen in some queueing systems.

Another way to estimate G is by means of the well-known empirical distribution:

$$G_t(B) := t^{-1} \sum_{i=0}^{t-1} 1_B(\xi_i), \; t \geq 1, \; B \in \mathbb{B}(X),$$

where $\xi_0, \xi_1, \dots, \xi_{t-1}$ are independent realizations (observed up to time $t-1$) of r.v.'s with the unknown distribution G. The construction of an adaptive policy via the empirical distribution is very general in the sense that G can be arbitrary. To the best of our knowledge, except when the one-stage cost is bounded and the discounted criterion is considered (see, e.g., [2, 12]), there are no works which treat construction of adaptive policies applying this approach. Thus future works in this direction might be of interest.

References

[1] D. Assaf, *Invariant problems in dynamic programming - average reward criterion*, Stoch. Processes Appl., 10 (1980), 313–322.

[2] R. Cavazos-Cadena, *Nonparametric adaptive control of discounted stochastic systems with compact state space*, J. Optim. Theory Appl.., 65 (1990), 191–207.

[3] E.B. Dynkin and A.A. Yushkevich, *Controlled Markov Processes*, Springer - Verlag, New York, 1979.

[4] D. Gatarek and L. Stettner, *On the compactness method in general ergodic impulsive control of Markov processes*, Stochastics, 31 (1990), 15–27.

[5] E.I. Gordienko, *Adaptive strategies for certain classes of controlled Markov processes*, Theory Probab. Appl., 29 (1985), 504–518.

[6] E.I. Gordienko and O. Hernández-Lerma, *Average cost Markov control processes with weighted norms: existence of canonical policies*, Appl. Math. (Warsaw), 23 (1995), 199–218.

[7] E.I. Gordienko and O. Hernández-Lerma, *Average cost Markov control processes with weighted norms: value iteration*, Appl. Math. (Warsaw), 23 (1995), 219–237.

[8] E.I. Gordienko, J.A. Minjárez-Sosa and R. Montes-de-Oca, *Average cost optimization in Markov control processes with unbounded costs: ergodicity and finite horizon approximation*, Reporte Interno, Departamento de Matemáticas, Universidad de Sonora, Rosales s/n, Col. Centro, 83000, Hermosillo, Sonora, México.

[9] E.I. Gordienko and J.A. Minjárez-Sosa, *Adaptive control for discrete-time Markov processes with unbounded costs: discounted criterion,* Kybernetika, 34 (1998a), 2, 217–234.

[10] E.I. Gordienko and J.A. Minjárez-Sosa, *Adaptive control for discrete-time Markov processes with unbounded costs: average criterion*, ZOR- Math. Methods of Oper. Res., 48, Iss. 2, 1998b.

[11] R. Hasminskii and I. Ibragimov, *On density estimation in the view of Kolmogorov's ideas in approximation theory*, Ann. of Statist., 18 (1990), 999–1010.

[12] O. Hernández-Lerma, *Adaptive Markov Control Processes*, Springer-Verlag, New York, 1989.

[13] O. Hernández-Lerma, *Infinite-horizon Markov control processes with undiscounted cost criteria: from average to overtaking optimality*, Reporte Interno 165, Departamento de Matemáticas, CINVESTAV-IPN, A.P. 14-740, 07000, México, D.F., México, 1994.

[14] O. Hernández-Lerma, *Average optimality in dynamic programming on Borel spaces — unbounded costs and controls*, Syst. Control Lett., 17 (1991), 237–242.

[15] O. Hernández-Lerma and J.B. Lasserre, *Discrete-Time Markov Control Processes: Basic Optimality Criteria*, Springer-Verlag, New York, 1996.

[16] O. Hernández-Lerma and J.B. Lasserre, *Further Topics on Discrete-Time Markov Control Processes,* Springer-Verlag, New York, 1999.

[17] N.V. Kartashov, *Inequalities in theorems of ergodicity and stability for Markov chains with common phase space II*, Theory Probab. Appl., 30 (1985), 507–515.

[18] M. Kurano, *Discrete-time markovian decision processes with an unknown parameter-average return criterion*, J. Oper. Res. Soc. Japan, 15 (1972), 67–76.

[19] S.A. Lippman, *On dynamic programming with unbounded rewards*, Manag. Sci., 21(1975), 1225–1233.

[20] P. Mandl, *Estimation and control in Markov chains*, Adv. Appl. Probab., 6(1974), 40–60.

[21] J.A. Minjárez-Sosa, *Nonparametric adaptive control for discrete-time Markov processes with unbounded costs under average criterion.* To appear in Appl. Math. (Warsaw).

[22] J.A. Minjárez-Sosa, *Adaptive Control for Markov Processes with Unbounded Costs,* Doctoral Thesis, UAM-Iztapalapa, México, 1998. (In Spanish).

[23] R. Montes-de-Oca and O. Hernández-Lerma, *Conditions for average optimality in Markov control processes with unbounded costs and controls*, J. Math. Syst., Estim., and Control, 4 (1995), 459–477.

[24] R. Montes-de-Oca, J.A. Minjárez-Sosa and O. Hernández-Lerma, *Conditions for average optimality in Markov control processes on Borel spaces*, Bol. Soc. Mat. Mexicana 39 (1994), 39–50.

[25] U. Rieder, *Measurable selection theorems for optimization problems*, Manuscripta Math., 24 (1978), 115–131.

[26] L.I. Sennott, *The average cost optimality equation and critical number policies*, Prob. Eng. Inform. Sci., 7 (1993), 47–67.

[27] M. Schal, *Average optimality in dynamic programming with general state space*, Math. Oper. Res., 18 (1993), 163–172.

[28] J.A.E.E. Van Nunen and J. Wessels, *A note on dynamic programming with unbounded rewards*, Manag. Sci., 24 (1978), 576–580.

Chapter 8

CONTROLLED MARKOV CHAINS WITH UTILITY FUNCTIONS

Seiichi Iwamoto
Department of Economic Engineering
Graduate School of Economics, Kyushu University 27
Fukuoka 812-8581, Japan
iwamoto@en.kyushu-u.ac.jp

Takayuki Ueno
Department of Economic Engineering
Graduate School of Economics, Kyushu University 27
Fukuoka 812-8581, Japan
ueno@en.kyushu-u.ac.jp

Toshiharu Fujita
Department of Electric, Electronic and Computer Engineering
Kyushu Institute of Technology
Kitakyushu 804-8550, Japan
fujita@comp.kyutech.ac.jp

Abstract In this paper we consider finite-stage stochastic optimization problems of utility criterion, which is the stochastic evaluation of associative reward through a utility function. We optimize the expected value of a utility criterion not in the class of Markov policies but in the class of general policies. We show that, by expanding the state space, an invariant imbedding approach yields an recursive relation between two adjacent optimal value functions. We show that the utility problem with a general policy is equivalent to a terminal problem with a Markov policy on the augmented state space. Finally it is shown that the utility problem has an optimal policy in the class of general policies on the original state space.

Z. Hou et al. (eds.), Markov Processes and Controlled Markov Chains, 135–149.

1. Introduction

In the theory of Markov decision process, the object is to maximize the expected value of additive function among the class of Markov policies (Markov class) ([1, 2, 3, 4, 8, 24, 25], and others). In this paper, we optimize the expected value of the utility function not in the Markov class but in the class of general policies (general class). The basic idea is how to use an invariant imbedding technique ([12, 14, 17]).

In Section 2, we propose a formulation of stochastic optimization problem with a utility criterion (utility problem) in general class.

In Section 3, we review the basic result on additive problem within Markov class, which is applied in the last section.

In Section 4, by use of the invariant imbedding method, we transform the utility problem into a terminal problem on an augmented state space.

In the last section we show that the utility problem with general class is equivalent to the terminal problem with Markov class on the augmented state space. Finally we show that the utility problem has an optimal policy in general class.

2. Utility problem

Throughout the paper, let $\{X_n, U_n\}$ be a controlled Markov chain on a finite state space X and a finite control space U with a transition law $p = \{p(y|x,u)\}$:

$$p(y|x,u) \quad \stackrel{\triangle}{=} \quad P(X_{n+1} = y \,|\, X_n = x, U_n = u).$$

Then we write

$$X_{n+1} \sim p(\cdot \,|\, x_n, u_n), \qquad 1 \leq n \leq N.$$

Given the data:

$$\begin{array}{lll} r : X{\times}U \to R^1 & \text{reward function,} & \\ k : X \to R^1 & \text{terminal function,} & \\ \circ : R^1{\times}R^1 \to R^1 & \text{associative binary operation} & \\ & \text{with left-identity element } \tilde{\lambda}, & (2.1) \\ \psi : R^1 \to R^1 & \text{utility function,} & \end{array}$$

we use the following notations:

$$\begin{array}{rcl} r_n &:=& r(X_n, U_n), \qquad\qquad k \;:=\; k(X_{N+1}) \\ p_n &:=& p(x_{n+1}|x_n, u_n), \qquad X^n \;:=\; X{\times}X{\times}\cdots{\times}X \quad (n\text{-times}), \\ H^n &:=& X{\times}U{\times}X{\times}U{\times}\cdots{\times}X{\times}U{\times}X \quad ((2n-1)\text{-factors}), \qquad (2.2) \\ h_{N+1} &:=& (x_1, u_1, x_2, u_2, \ldots, x_{N+1}) \in H^{N+1}. \end{array}$$

We consider three classes of policies. A *Markov* (resp. *general, primitive*) *policy*, $\pi = \{\pi_1, \pi_2, \ldots, \pi_N\}$ (resp. $\sigma = \{\sigma_1, \sigma_2, \ldots, \sigma_N\}$, $\mu = \{\mu_1, \mu_2, \ldots, \mu_N\}$) is a sequence of *Markov* (resp. *general, primitive*) *decision functions*:

$$\pi_n : X \to U \quad (\text{resp. } \sigma_n : X^n \to U,\ \mu_n : H^n \to U), \qquad 1 \le n \le N.$$

Let Π (resp. $\Pi(g)$, $\Pi(p)$) denote the set of all Markov (resp. general, primitive) policies. We call Π (resp. $\Pi(g)$, $\Pi(p)$) a *Markov* (resp. *general, primitive*) *class*. Then we note that

$$\Pi \subset \Pi(g) \subset \Pi(p). \tag{2.3}$$

Further, for n ($1 \le n \le N$), let Π_n (resp. $\Pi_n(g)$, $\Pi_n(p)$) denote the set of all corresponding policies which start from n^{th} stage on. For instance, $\Pi_n(p)$ denotes the set of all primitive policies $\mu = \{\mu_n, \ldots, \mu_N\}$ which begin at stage n. Now, let us consider the stochastic optimization problem of utility function:

$$\mathrm{P}_1(x_1) \quad \begin{cases} \text{Optimize} & E^{\sigma}_{x_1}[\psi(r_1 \circ r_2 \circ \cdots \circ r_N \circ k)] \\ \text{subject to} & \left.\begin{array}{ll} \text{(i)}_\text{n} & X_{n+1} \sim p(\cdot \,|\, x_n, u_n), \\ \text{(ii)}_\text{n} & u_n \in U, \end{array}\right\} 1 \le n \le N, \end{cases} \tag{2.4}$$

where $E^{\sigma}_{x_1}$ is the expectation operator on the product space X^{N+1} induced from the Markov transition law p, a general policy $\sigma = \{\sigma_1, \sigma_2, \ldots, \sigma_N\} \in \Pi(g)$, and an initial state $x_1 \in X$:

$$E^{\sigma}_{x_1}[W] = \sum_{(x_2,\ldots,x_{N+1}) \in X^N} \!\!\sum \cdots \sum W(h_{N+1}) p_1 p_2 \cdots p_N, \quad \forall W : H^{N+1} \to R^1, \tag{2.5}$$

where

$$\begin{aligned} p_1 &= p(x_2|x_1, u_1), \; p_2 = p(x_3|x_2, u_2), \; \ldots, \; p_N = p(x_{N+1}|x_N, u_N), \\ u_1 &= \sigma_1(x_1), \qquad u_2 = \sigma_2(x_1, x_2), \quad \ldots, \; u_N = \sigma_N(x_1, x_2, \ldots, x_N). \end{aligned}$$

3. Markov policies

In this section we review the basic results on additive problem in Markov class Π. The additive problem has always an optimal *Markov* policy in Π ([2, 3, 4, 5, 6, 7, 8, 9, 10, 18, 19, 22, 23]). This fact plays an impotant role at the last section.

3.1 Additive problem

Now, let us consider the *additive problem* in Π:

$$\mathrm{A}_1(x_1) \quad \left\{ \begin{array}{lll} \text{Optimize} & E^{\pi}_{x_1}[\,r_1 + r_2 + \cdots + r_N + k\,] & \\ \text{subject to} & \left.\begin{array}{ll} (\mathrm{i})_\mathrm{n} & X_{n+1} \sim p(\cdot\,|\,x_n, u_n), \\ (\mathrm{ii})_\mathrm{n} & u_n \in U, \end{array}\right\} & 1 \le n \le N, \end{array} \right. \tag{3.1}$$

where $E^{\pi}_{x_1}$ is defined through Markov policy π. Thus we have

$$E^{\pi}_{x_1}[r_1 + \cdots + r_N + k] = \sum_{(x_2,\ldots,x_{N+1})\in X^N}\sum\cdots\sum [r_1 + \cdots + r_N + k]p_1 p_2 \cdots p_N \tag{3.2}$$

where

$$u_1 = \pi_1(x_1), \quad u_2 = \pi_2(x_2), \quad \ldots, \quad u_N = \pi_N(x_N). \tag{3.3}$$

The conventional dynamic programming method solves the problem (3.1) as follows. It regards $\mathrm{A}_1(x_1)$ as one of the family of subproblems $\mathcal{A} = \{\mathrm{A}_n(x_n)\}$:

$$\mathrm{A}_n(x_n) \quad \left\{ \begin{array}{lll} \text{Optimize} & E^{\pi}_{x_n}[r_n + \cdots + r_N + k] & \\ \text{subject to} & (\mathrm{i})_\mathrm{m}, \quad (\mathrm{ii})_\mathrm{m}, & n \le m \le N, \\ & x_n \in X, & 1 \le n \le N+1. \end{array} \right. \tag{3.4}$$

Let $f_n(x_n)$ be the optimum value of $\mathrm{A}_n(x_n)$. Then we have the recursive formula between the optimum value $f_n(x_n)$ and its adjacent optimal value function $f_{n+1}(\cdot)$:

Theorem 3.1

$$\begin{aligned} f_n(x) &= \mathop{\mathrm{Opt}}_{u\in U}\left[r(x,u) + \sum_{y\in X} f_{n+1}(y)p(y|x,u)\right], \\ & \qquad\qquad\qquad\qquad x \in X,\ n = 1, 2, \cdots, N \qquad (3.5) \\ f_{N+1}(x) &= k(x), \qquad\qquad x \in X. \qquad (3.6) \end{aligned}$$

Further, we have an optimal policy as follows.

Theorem 3.2 *Let $\pi^*_n(x)$ be an opimizer of* (3.5). *Then policy π^* is optimal in Markov class; for all $\pi \in \Pi$*

$$E^{\pi^*}_{x_1}[r_1 + \cdots + r_N + k] \ \ge\ E^{\pi}_{x_1}[r_1 + \cdots + r_N + k], \quad \forall x_1 \in X. \tag{3.7}$$

3.2 Terminal problem

In this subsection, as a special case of the additive problem, we take the *terminal problem*:

$$\mathrm{T}_1(x_1) \quad \begin{cases} \text{Optimize} & E^{\pi}_{x_1}[k] \\ \text{subject to} & (\mathrm{i})_{\mathrm{n}}, \quad (\mathrm{ii})_{\mathrm{n}}, \quad 1 \le n \le N. \end{cases} \tag{3.8}$$

This is the case

$$r(x,u) = 0, \qquad \forall (x,u) \in X{\times}U,$$

in (3.1). We imbed $\mathrm{T}_1(x_1)$ into the family $\mathcal{T} = \{\mathrm{T}_n(x_n)\}$:

$$\mathrm{T}_n(x_n) \quad \begin{cases} \text{Optimize} & E^{\pi}_{x_n}[k] \\ \text{subject to} & (\mathrm{i})_{\mathrm{m}}, \quad (\mathrm{ii})_{\mathrm{m}}, \quad n \le m \le N, \\ & x_n \in X, \qquad 1 \le n \le N+1. \end{cases} \tag{3.9}$$

Then the optimum value $t_n(x_n)$ of $\mathrm{T}_n(x_n)$ satisfies the recursive formula:

Corollary 3.1

$$t_n(x) = \operatorname*{Opt}_{u \in U} \sum_{y \in X} t_{n+1}(y) p(y|x,u), \quad x \in X,\ 1 \le n \le N, \tag{3.10}$$

$$t_{N+1}(x) = k(x), \qquad x \in X. \tag{3.11}$$

Corollary 3.2 *Let* $\pi^*_n(x)$ *be an opimizer of* (3.10). *Then the policy* π^* *is* optimal *in Markov class; for all* $\pi \in \Pi$

$$E^{\pi^*}_{x_1}[k] \ \ge\ E^{\pi}_{x_1}[k], \qquad \forall x_1 \in X. \tag{3.12}$$

To conclude this section, we remark that the model is stationary; the state space, control space, reward function, and transition probability are all independent of stage n. However, all the results in this section are valid for non-stationary models.

4. Invariant imbedding approach

In this section we show how to imbed the original problem into an appropriate family of subproblems. Our imbedding process has two phases. The first phase is to introduce the past-value sets up to current stage. The second is to expand the original state space and to reduce the utility problem to a terminal problem over there. Both phases involve policies in a transliteration.

4.1 Past-value sets up to today

Now, we note that the left-identity element $\tilde{\lambda}$ implies

$$\psi(\tilde{\lambda} \circ r_1 \circ \cdots \circ r_N \circ k) = \psi(r_1 \circ \cdots \circ r_N \circ k).$$

For further discussion we take the *past-value set* up to the first stage:

$$\Omega_1 \overset{\triangle}{=} \{\tilde{\lambda}\}.$$

We define the *past-value set* up to n^{th} stage $(2 \leq n \leq N)$:

$$\begin{aligned}\Omega_n \overset{\triangle}{=} \{\tilde{\lambda} \circ r(x_1, u_1) \circ \cdots \circ r(x_{n-1}, u_{n-1}) \mid \\ (x_1, u_1, \ldots, x_{n-1}, u_{n-1}) \in X \times U \times \cdots \times X \times U\}. \quad (4.1)\end{aligned}$$

Then we have the forward recursive formula:

Lemma 4.1

$$\begin{aligned}\Omega_1 &= \{\tilde{\lambda}\}, \\ \Omega_{n+1} &= \{\lambda \circ r(x, u) \mid \lambda \in \Omega_n,\ (x, u) \in X \times U\}. \quad (4.2)\end{aligned}$$

Proof It is straightforward. ∎

4.2 Terminal problem on augmented state spaces

By attaching Ω_n, we expand the state space X to an *augmented state spaces* $\{Y_n\}$:

$$Y_n \overset{\triangle}{=} X \times \Omega_n, \qquad (n = 1,\ 2,\ \ldots, N), \quad (4.3)$$

and define a *new Markov transition law* $q = \{q_n\}$ there by

$$q_n((y; \mu) \mid (x; \lambda),\ u), \overset{\triangle}{=} \begin{cases} p(y|x,\ u), & \text{if } \lambda \circ r(x, u) = \mu, \\ 0, & \text{otherwise.} \end{cases} \quad (4.4)$$

Symbolically we express (4.4) as

$$(Y; \mu) \sim q_n(\cdot \,|(x; \lambda),\ u) \overset{\text{def}}{\Longleftrightarrow} \begin{cases} Y \sim p(\cdot\,|x,\ u), \\ \mu = \lambda \circ r(x, u). \end{cases} \quad (4.5)$$

Now, we consider Markov policy for the *augmented process* $\{\tilde{Y}_n, U_n\}$ where $\tilde{Y}_n = (X_n; \lambda_n)$. For the process, a *Markov policy* $\gamma = \{\gamma_1, \gamma_2, \ldots, \gamma_N\}$ is a sequence of *Markov decision functions*

$$\gamma_n : Y_n \to U, \qquad (n = 1,\ 2,\ \ldots, N).$$

Let $\tilde{\Pi}$ denote the set of all Markov policies (*augmented Markov class*).
We define the *terminal function* $T : Y_{N+1} \to R^1$ by

$$T(y_{N+1}) = T(x_{N+1}; \lambda_{N+1}) \stackrel{\triangle}{=} \psi(\lambda_{N+1} \circ k(x_{N+1})). \tag{4.6}$$

Now we consider a *relatively but not so large* family of *parametric* subproblems $\mathcal{Q} = \{Q_n(x_n; \lambda_n)\}$:

$$Q_n(x_n;\lambda_n) \left\{ \begin{array}{lll} \text{Optimize} & \tilde{E}^{\gamma}_{y_n}[T] & \\ \text{subject to} & \left.\begin{array}{l} (\mathrm{i})'_{\mathrm{m}} \quad \tilde{Y}_{m+1} \sim q_m(\cdot \mid y_m, u_m), \\ (\mathrm{ii})_{\mathrm{m}} \quad u_m \in U, \end{array}\right\} & n \le m \le N, \\ & x_n \in X, \quad \lambda_n \in \Omega_n, & 1 \le n \le N. \end{array} \right. \tag{4.7}$$

Here $\tilde{E}^{\gamma}_{y_n}$ is the expectation operator on the product space $Y_n \times Y_{n+1} \times \cdots \times Y_{N+1}$ induced from the Markov transition law $q = \{q_n, q_{n+1}, \ldots, q_N\}$, a Markov policy $\gamma = \{\gamma_n, \gamma_{n+1}, \ldots, \gamma_N\}$, and an initial state $y_n = (x_n; \lambda_n)$:

$$\tilde{E}^{\gamma}_{y_n}[T] = \sum_{(y_{n+1},\ldots,y_{N+1})}\sum \cdots \sum T(y_{N+1}) q^{\gamma}_n q^{\gamma}_{n+1} \cdots q^{\gamma}_N, \tag{4.8}$$

where

$$q^{\gamma}_m := q_m(y_{m+1}|y_m, u_m), \ u_m := \gamma_m(y_m), \ y_m = (x_m; \lambda_m), \ n \le m \le N. \tag{4.9}$$

We note that the multiple summation in (4.8) is taken over $Y_{n+1} \times Y_{n+2} \times \cdots \times Y_{N+1}$ and that the optimization in (4.7) is taken over all Markov policies $\gamma \in \tilde{\Pi}_n$. Thus we define the maximum value functions as follows:

$$u^n(y_n) := \operatorname*{Max}_{\gamma \in \tilde{\Pi}_n} \tilde{E}^{\gamma}_{y_n}[T], \qquad y_n \in Y_n, \quad 1 \le n \le N, \tag{4.10}$$

where

$$u^{N+1}(y_{N+1}) := T(y_{N+1}), \qquad y_{N+1} \in Y_{N+1}. \tag{4.11}$$

First we remark that the augmented subproblems $\mathcal{Q} = \{Q_n(x_n; \lambda_n)\}$ have a few equivalent forms.
Now let us in turn, decompose the unified transition law:

$(\mathrm{i})'_{\mathrm{m}} : \tilde{Y}_{m+1} \sim q_m(\cdot \mid y_m, u_m)$ in $Q_n(x_n; \lambda_n)$,

into the original transition law:

$(\mathrm{i})_{\mathrm{m}} : X_{m+1} \sim p(\cdot \mid x_m, u_m)$, and the deterministic transition $\lambda_m \overset{u_m}{\longmapsto} \lambda_{m+1}$. Then, from (4.3) and (4.4), we see that the problem $\mathrm{Q}_1(x_1;\tilde{\lambda})$ is also written in the form:

$$\mathrm{Q}_1(x_1;\tilde{\lambda}) \left\{ \begin{array}{lll} \text{Optimize} & \tilde{E}^{\gamma}_{y_1}[\,\psi(\tilde{\lambda}\circ r_1\circ\cdots\circ r_N\circ k)\,] & \\ \text{subject to} & \left.\begin{array}{ll} (\mathrm{i})_\mathrm{n} & X_{n+1}\sim p(\cdot\,|\,x_n,u_n), \\ (\mathrm{ii})_\mathrm{n} & u_n\in U, \\ (\mathrm{iii})_\mathrm{n} & \lambda_{n+1}=\lambda_n\circ r(x_n,u_n), \end{array}\right\} & 1\le n\le N. \end{array}\right. \tag{4.12}$$

Here we note that the additional sequential condition $\{(\mathrm{iii})_\mathrm{n}\}$ implies the equality

$$\psi(\tilde{\lambda}\circ r_1\circ\cdots\circ r_N\circ k) \quad = \quad \psi(\lambda_{N+1}\circ k). \tag{4.13}$$

Therefore, $\mathrm{Q}_1(x_1;\tilde{\lambda})$ is expressed as a *terminal* problem:

$$\mathrm{Q}_1(x_1;\tilde{\lambda}) \left\{ \begin{array}{ll} \text{Optimize} & \tilde{E}^{\gamma}_{y_1}[\,\psi(\lambda_{N+1}\circ k)\,] \\ \text{subject to} & (\mathrm{i})_\mathrm{n}, \quad (\mathrm{ii})_\mathrm{n}, \quad (\mathrm{iii})_\mathrm{n}, \quad 1\le n\le N, \end{array}\right. \tag{4.14}$$

provided that we view the pair $(x_n;\lambda_n)$ as a new state-variable y_n and that we consider the probability measure $\tilde{P}^{\gamma}_{y_1}(\cdot)$ on the augmented state spaces $\{Y_n\}$. Thus we can also imbed $\mathrm{Q}_1(x_1;\tilde{\lambda})$ into the family of subproblems $\mathcal{Q}=\{\mathrm{Q}_n(x_n;\lambda_n)\}$:

$$\mathrm{Q}_1(x_1;\tilde{\lambda}) \left\{ \begin{array}{ll} \text{Optimize} & \tilde{E}^{\gamma}_{y_n}[\,\psi(\lambda_n\circ r_n\circ\cdots\circ r_N\circ k)\,] \\ \text{subject to} & (\mathrm{i})_\mathrm{m}, \quad (\mathrm{ii})_\mathrm{m}, \quad (\mathrm{iii})_\mathrm{m}, \qquad n\le m\le N. \end{array}\right. \tag{4.15}$$

We note that $u^n(x_n;\lambda_n)$ denotes the optimum value of $\mathrm{Q}_n(x_n;\lambda_n)$ (see (4.10)). Then we have the backward recursive relation:

Theorem 4.1

$$u^n(x;\lambda) \quad = \quad \underset{u\in U}{\mathrm{Opt}} \sum_{y\in X} u^{n+1}(\,y\,;\lambda\circ r(x,u)\,)p(y|x,u),$$
$$x\in X,\ \ \lambda\in\Omega_n,\ \ n=1,2,\cdots,N, \tag{4.16}$$
$$u^{N+1}(x;\lambda) \quad = \quad \psi(\lambda\circ k(x)), \qquad x\in X,\ \ \lambda\in\Omega_{N+1}. \tag{4.17}$$

Proof It suffices to note that the family of subproblems $\mathcal{Q}=\{\mathrm{Q}_n(x_n;\lambda_n)\}$ is a terminal problem on $\{Y_n\}$. Thus, from Corollary 3.1, we have the desired recursive formula (4.16),(4.17). ■

Theorem 4.2 *Let* $\gamma^*_n(x;\lambda)$ *be the set of all maximizers in* (4.16). *Then policy* γ^* *is optimal in augmented Markov class: for any Markov policy* $\gamma\in\tilde{\Pi}$,

$$\tilde{E}^{\gamma^*}_{y_1}[T]\ge\tilde{E}^{\gamma}_{y_1}[T], \qquad \forall y_1\in\tilde{Y}_1. \tag{4.18}$$

Proof This is a direct transliteration of Corollary 3.2. ■

Remark 4.1 *There are two definitions:* the set of all maximizers *and* a maximizer. *As for optimal value, there is not much to choose between the former definition and the latter. For the clarification of optimal policies in the next section, we prefer the former to the latter.*

4.3 Subproblems associated with histories

In this subsection, we consider a *larger* family of subproblems $\mathcal{R} = \{\mathrm{R}_n(h_n)\}$:

$$\mathrm{R}_n(h_n) \begin{cases} \text{Optimize} & E^{\mu}_{h_n}[\psi(r_1 \circ \cdots \circ r_N \circ k)] \\ \text{subject to} & \left.\begin{array}{ll}(\mathrm{i})_{\mathrm{m}} & X_{m+1} \sim p(\cdot \,|\, x_m, u_m), \\ (\mathrm{ii})_{\mathrm{m}} & u_m \in U,\end{array}\right\} n \le m \le N, \\ & \quad h_n \in H_n, \quad 1 \le n \le N+1. \end{cases} \tag{4.19}$$

The subproblem $\mathrm{R}_n(h_n)$ starts at a given history $h_n \in H_n$ on the n^{th} stage (see also [13, 16]). The expectation operator $E^{\mu}_{h_n}$ is induced from the transition law p, a primitive policy $\mu = \{\mu_n, \ldots, \mu_N\} \in \Pi_n(p)$, and a history $h_n = (x_1, u_1, \ldots, u_{n-1}, x_n) \in H_n$:

$$E^{\mu}_{h_n}[W] = \sum_{(x_{n+1},\ldots,x_{N+1}) \in X^{N-n+1}} \!\!\sum \cdots \sum W(h_n, u_n, x_{n+1}, \ldots, u_N, x_{N+1}) \times P^{\mu}_{h_n}(x_{n+1}, \ldots, x_{N+1}), \tag{4.20}$$

where

$$\begin{aligned} u_n &= \mu_n(h_n), \quad u_{n+1} = \mu_{n+1}(h_{n+1}), \quad \ldots, \quad u_N = \mu_N(h_N), \\ h_m &= \big(h_n, u_n, x_{n+1}, u_{n+1}, \ldots, x_{m-1}, u_{m-1}, x_m\big). \end{aligned}$$

The objective function,

$$W := \psi(r_1 \circ \cdots \circ r_N \circ k) \tag{4.21}$$

is the evaluation of the process starting from a pair of stage and history (n, h_n) to the final stage $(N+1)$, and the conditional probability law on the product space X^{N-n+1},

$$P^{\mu}_{h_n}(x_{n+1}, \ldots, x_{N+1}),$$

is induced from the triplet (p, μ, h_n):

$$\begin{aligned} & P^{\mu}_{h_n}(x_{n+1}, \ldots, x_{N+1}) \\ &\quad := P^{\mu}(X_{n+1} = x_{n+1}, \ldots, X_{N+1} = x_{N+1} \,|\, h_n), \\ &\quad := P(X_{n+1} = x_{n+1}, U_{n+1} = u_{n+1}, \\ &\qquad\qquad \ldots, X_{N+1} = x_{N+1}, U_{N+1} = u_{N+1} \,|\, h_n), \end{aligned} \tag{4.22}$$

where $u_m = \mu_m(h_m)$, $n \le m \le N$. Thus we see that

$$P^{\mu}_{h_n}(x_{n+1}, \ldots, x_{N+1}) \;=\; p_n p_{n+1} \cdots p_N, \tag{4.23}$$

where

$$p_m := p(x_{m+1}|x_m, u_m), \qquad u_m = \mu_m(h_m), \quad n \le m \le N.$$

We note that the multiple summation in (4.20) is taken over X^{N-n+1}, and that the optimization in (4.19) is taken over all primitive policies $\mu \in \Pi_n(p)$. Thus we define,

$$w_n(h_n) := \operatorname*{Max}_{\mu \in \Pi_n(p)} E^{\mu}_{h_n}\left[\psi(r_1 \circ \cdots \circ r_N \circ k)\right], \quad h_n \in H_n, \; 1 \le n \le N, \tag{4.24}$$

where

$$w_{N+1}(h_{N+1}) := \psi(r_1 \circ \cdots \circ r_N \circ k), \qquad h_{N+1} \in H_{N+1}. \tag{4.25}$$

Then we have the backward recursive relation:

Theorem 4.3

$$w_n(h) \;=\; \operatorname*{Max}_{u \in U} \sum_{y \in X} w_{n+1}(h, u, y) p(y|x, u), \quad h \in H_n, \; 1 \le n \le N, \tag{4.26}$$

$$w_{N+1}(h) \;=\; \psi(r_1 \circ \cdots \circ r_N \circ k), \quad h \in H_{N+1}. \tag{4.27}$$

Proof This is straightforward. ■

Theorem 4.4 *Let $\mu^*_n(h)$ be the set of all maximizers in* (4.26). *Then policy μ^* is optimal in primitive class; for all $\mu \in \Pi(p)$,*

$$E^{\mu^*}_{x_1}\left[\psi(r_1 \circ \cdots \circ r_N \circ k)\right] \ge E^{\mu}_{x_1}\left[\psi(r_1 \circ \cdots \circ r_N \circ k)\right], \quad \forall x_1 \in X. \tag{4.28}$$

5. Equivalences and Optimality

In this section we establish two equivalent relations among the three related problems. Further, by use of the equivalences, we show that the desired optimal *general* policy is obtained by solving a recursive equation on the augmented process and by transforming the resultant optimal Markov policy.

5.1 Two Equivalent Relations

Now let us focus our attention on optimality relations among three optimization problems $\mathrm{P}_1(x_1)$, $\mathcal{Q} = \{\mathrm{Q}_n(x_n; \lambda_n)\}$ and $\mathcal{R} = \{\mathrm{R}_n(h_n)\}$.

Let $v_1(x_1)$ denote the *maximum value* of the original problem $\mathrm{P}_1(x_1)$ in (2.4). The first equivalence is among the policy classes.

Lemma 5.1 (Equivalence I)

(i) Any primitive policy μ generates a general policy σ which satisfies

$$E^{\sigma}_{x_1}\left[\psi(r_1\circ\cdots\circ r_N\circ k)\right] = E^{\mu}_{x_1}\left[\psi(r_1\circ\cdots\circ r_N\circ k)\right], \quad \forall x_1 \in X, \tag{5.1}$$

and vice versa. Thus we have

$$\underset{\sigma\in\Pi(g)}{\mathrm{Max}}\ E^{\sigma}_{x_1}\left[\psi(r_1\circ\cdots\circ r_N\circ k)\right] = \underset{\mu\in\Pi(p)}{\mathrm{Max}}\ E^{\mu}_{x_1}\left[\psi(r_1\circ\cdots\circ r_N\circ k)\right], \quad x_1 \in X. \tag{5.2}$$

(ii) Any Markov policy γ of $\mathrm{Q}_1(x_1;\tilde{\lambda})$ generates a general policy σ of $\mathrm{P}_1(x_1)$ which satisfies

$$E^{\sigma}_{x_1}\left[\psi(r_1\circ\cdots\circ r_N\circ k)\right] = \tilde{E}^{\gamma}_{y_1}[T], \qquad \forall x_1 \in X,\ y_1 = (x_1;\tilde{\lambda}). \tag{5.3}$$

Proof

(i) Any $\mu \in \Pi(p)$ is compressed to *the* $\sigma \in \Pi(g)$ by deletion of the dependency on the intermediate control(s). Conversely, any σ generates *a* μ with the same expected value.

(ii) Given γ, we define $\sigma_n(x_1, x_2, \ldots, x_n)$ as follows:

$$\begin{array}{rclrcl} u_1 & := & \gamma_1(x_1;\tilde{\lambda}), & \lambda_2 & := & \tilde{\lambda}\circ r(x_1,u_1),\\ u_2 & := & \gamma_2(x_2;\lambda_2), & \lambda_3 & := & \lambda_2\circ r(x_2,u_2),\\ \vdots & & & \vdots & & \\ u_{n-1} & := & \gamma_{n-1}(x_{n-1},u_{n-1}), & \lambda_n & := & \lambda_{n-1}\circ r(x_{n-1},u_{n-1}),\\ & & \sigma_n(x_1,x_2,\ldots,x_n) & & := & \gamma_n(x_n;\lambda_n). \end{array} \tag{5.4}$$

Then σ has the same expected value as γ. ■

We remark that both the probability measures are coincident:

$$P^{\sigma}_{x_1}(\cdot) = P^{\mu}_{x_1}(\cdot) \quad \text{on } X^N,\ x_1 \in X. \tag{5.5}$$

That is, (5.1) holds for any reward function:

$$E^{\sigma}_{x_1}[g] = E^{\mu}_{x_1}[g], \qquad \forall g : H_{N+1} \to R^1,\ x_1 \in X. \tag{5.6}$$

The second equivalence is between the optimal primitive policy and the optimal augmented Markov policy.

Theorem 5.1 (Equivalence II) *Let $\mu^* \in \Pi(p)$ and $\gamma^* \in \tilde{\Pi}$ be optimal, respectively. Then both optimal values are equal:*

$$E_{x_1}^{\mu^*}\left[\psi(r_1 \circ \cdots \circ r_N \circ k)\right] = \tilde{E}_{y_1}^{\gamma^*}[T], \qquad \forall x_1 \in X,\ y_1 = (x_1; \tilde{\lambda}). \quad (5.7)$$

Further, both optimal policies coincide on histories:

$$\mu_n^*(h_n) = \gamma_n^*(x_n; r(x_1, u_1) \circ \cdots \circ r(x_{n-1}, u_{n-1})) \\ \forall h_n = (x_1, u_1, \ldots, x_{n-1}, u_{n-1}, x_n) \in H_n,\ 1 \le n \le N. \quad (5.8)$$

Proof This is a backward induction on stage number. ■

Corollary 5.1 (Equivalence between primitive and augmented Markov classes)

(i) The optimal policy $\gamma^* \in \tilde{\Pi}$ satisfies

$$E_{x_1}^{\mu}\left[\psi(r_1 \circ \cdots \circ r_N \circ k)\right] \le \tilde{E}_{y_1}^{\gamma^*}[T],\ \forall \mu \in \Pi(p),\ x_1 \in X,\ y_1 = (x_1; \tilde{\lambda}). \quad (5.9)$$

(ii) Thus we have

$$\operatorname*{Max}_{\mu \in \Pi(p)} E_{x_1}^{\mu}\left[\psi(r_1 \circ \cdots \circ r_N \circ k)\right] = \operatorname*{Max}_{\gamma \in \tilde{\Pi}} \tilde{E}_{y_1}^{\gamma}[T],\ \forall x_1 \in X,\ y_1 = (x_1; \tilde{\lambda}). \quad (5.10)$$

To summarize these results we have:

Corollary 5.2 (Equivalence among three problems)

(i) Three optimal value functions are equal:

$$v_1(x_1) = w_1(x_1) = u^1(x_1; \tilde{\lambda}), \qquad \forall x_1 \in X,\ y_1 = (x_1; \tilde{\lambda}). \quad (5.11)$$

(ii) Let $\gamma^* \in \tilde{\Pi}$ and $\mu^* \in \Pi(p)$ be the optimal policies obtained by solving the recursive equations (4.16) and (4.26) respectively. Then both optimal policies coincide on histories:

$$\mu_n^*(h_n) = \gamma_n^*(x_n; r(x_1, u_1) \circ \cdots \circ r(x_{n-1}, u_{n-1})) \\ \forall h_n = (x_1, u_1, \ldots, x_{n-1}, u_{n-1}, x_n) \in H_n,\ 1 \le n \le N. \quad (5.12)$$

The general policy σ^*. compressed from μ^* through the deletion, is optimal in $\Pi(g)$. Furthermore the three optimal policies σ^*, μ^* and γ^* have the same expected value:

$$\begin{aligned} E_{x_1}^{\sigma^*}\left[\psi(r_1 \circ \cdots \circ r_N \circ k)\right] &= E_{x_1}^{\mu^*}\left[\psi(r_1 \circ \cdots \circ r_N \circ k)\right] \\ &= \tilde{E}_{y_1}^{\gamma^*}[T], \quad \forall x_1 \in X,\ y_1 = (x_1; \tilde{\lambda}). \end{aligned} \quad (5.13)$$

5.2 Optimal policies

Now we show that the optimal policy for Q in $\tilde{\Pi}$ yields an optimal policy for $P_1(x_1)$ in class $\Pi(g)$. The *optimum value* $v_1(x_1)$ of problem $P_1(x_1)$ is $u^1(x_1;\tilde{\lambda})$:

$$v_1(x_1) = u^1(x_1;\tilde{\lambda}). \tag{5.14}$$

Further, the *optimal Markov* policy γ^* generates a *general* policy σ^* of (2.4) through (5.4).

Theorem 5.2 *The policy σ^* is optimal in general class; for any general policy $\sigma \in \Pi(g)$,*

$$E^{\sigma^*}_{x_1}\left[\psi(r_1 \circ \cdots \circ r_N \circ k)\right] \geq E^{\sigma}_{x_1}\left[\psi(r_1 \circ \cdots \circ r_N \circ k)\right], \quad \forall x_1 \in X. \tag{5.15}$$

Proof This follows from Lemma 5.1(ii), Theorem 5.1 and Lemma 5.1(i). ■

References

[1] R.E. Bellman, *Dynamic Programming*, Princeton Univ. Press, NJ, 1957.

[2] D. Blackwell, Discounted dynamic programming, Ann. Math. Stat. **36** (1965), 226–235.

[3] E.V. Denardo, Contraction mappings in the theory underlying dynamic programming, SIAM Review **9** (1968), 165–177.

[4] E.V. Denardo, *Dynamic Programming : Models and Applications*, Prentice-Hall, N.J., 1982.

[5] T. Fujita, Re-examination of Markov policied for additive decision process, Bull. Informatics and Cybernetics, **29** (1997), 51–66.

[6] N. Furukawa and S. Iwamoto, Markovian decision processes with recursive reward functions, Bull. Math. Statist. **15** (1973), 79–91.

[7] N. Furukawa and S. Iwamoto, Dynamic programming on recursive reward systems, Bull. Math. Statist. **17** (1976), 103–126.

[8] K. Hinderer, *Foundations of Non-Stationary Dynamic Programming with Discrete Time Parameter*, Lect. Notes in Operation Research and Mathematical Systems, Vol. 33, Springer-Verlag, Berlin, 1970.

[9] R. A. Howard, *Dynamic Programming and Markov Processes*, MIT Press, Cambridge, Mass., 1960.

[10] S. Iwamoto, Discrete dynamic programming with recursive additive system, Bull. Math. Statist. **15** (1974), 49–66.

[11] S. Iwamoto, Associative dynamic programs, J. Math. Anal. Appl., **201** (1996), 195–211.

[12] S. Iwamoto, On expected values of Markov statistics, Bull. Informatics and Cybernetics, **30** (1998), 1–24.

[13] S. Iwamoto, Conditional decision processes with recursive reward function, J. Math. Anal. Appl., **230** (1999), 193–210.

[14] S. Iwamoto and T. Fujita, Stochastic decision-making in a fuzzy environment, J. Operations Res. Soc. Japan **38** (1995), 467–482.

[15] S. Iwamoto and M. Sniedovich, Sequential decision making in fuzzy environment, J. Math. Anal. Appl., **222** (1998), 208–224.

[16] S. Iwamoto, K. Tsurusaki and T. Fujita, Conditional decision-making in a fuzzy environment, J. Operations Res. Soc. Japan **42** (1999), 198–218.

[17] E. S. Lee, *Quasilinearization and Invariant Imbedding*, Academic Press, New York, 1968.

[18] D.M. Kreps, Decision problems with expected utility criteria, I,Math. Oper. Res. **2** (1977), 45–53.

[19] D.M. Kreps, Decision problems with expected utility criteria, II; stationarity, Math. Oper. Res. **2** (1977), 266–274.

[20] W. Lipfert, Über ein stochastisches dynamiches Entscheidungsmodell mit allegemeinen Ertragsfunktionalen, Optimization **16** (1985), 313–328.

[21] A.S. Nowak, On a general dynamic programming problem, Colloquium Mathematicum, **37** (1977), Fasc.1, 131–138.

[22] E. Porteus, An informal look at the princiole of optimality, Management Sci. **21** (1975), 1346–1348.

[23] E. Porteus, Conditions for characterizing the structure of optimal strategies in infinite-horizon dynamic programs, J. Opt. Theo. Anal. **36** (1982), 419–432.

[24] M.L. Puterman, *Markov Decision Processes*: Stochastic Models, Chap. VIII, D.P. Heyman and M.J. Sobel (Ed's), Handbooks in Operations Research and Management Science Vol. 2, Elsevier, Amsterdam, 1990.

[25] M.L. Puterman, *Markov Decision Processes : discrete stochastic dynamic programming*, Wiley & Sons, New York, 1994.

[26] U. Rieder, Non-cooperative dynamic games with general utility functions, *Stochastic Games and Related Topics*; T.E.S. Raghavan et. al.(eds), Kluwer Academic Publishers, 1991, 161–174.

[27] M. Sniedovich, *Dynamic Programming*, Marcel Dekker, Inc. NY, 1992.

Chapter 9

CLASSIFICATION PROBLEMS IN MDPS

L.C.M. Kallenberg
University of Leiden
The Netherlands

Abstract In this paper we investigate classification problems for Markov decision processes (MDPs). These MDPs can be classified in several ways. One way is based on the concept *communicating*, and distinguishes between communicating, weakly communicating and noncommunicating. Another way of classification is based on the *ergodic structure*. In this approach the distinction between completely ergodic, unichain and multichain is made. Furthermore, there is a classification based on *decomposition* of the state space. This decomposition distinguishes between several levels. At each level there is a set of recurrent classes and a (perhaps empty) set of transient states.

Classification of an MDP may be of interest, e.g. for undiscounted MDPs both in the unconstrained as in the constrained case. We review all these classification problems and present old and new results. It turns out that these problems, except one, can be solved in polynomial time; algorithms and complexity results are given. The only problem for which, to our knowledge, no polynomial-time algorithm is known, is the distinction between a unichain and a multichain MDP. For this problem, we have some partial results which can be obtained in polynomial time.

1. Introduction

Consider a discrete Markov decision problem (MDP) with finite state space E, finite action sets $A(i)$, $i \in E$, and transition probabilities $p_{ij}(a)$, $a \in A(i)$ and $i, j \in E$. Rewards are no subject of our study, so they are not mentioned here. A *deterministic policy* f is a function which assigns in a deterministic way an admissible action to each state, i.e. $f(i) \in A(i)$, $i \in E$. Any deterministic policy f induces a transition probability matrix $P(f)$ of a Markov chain with (i,j)th entry given by $p_{ij}(f(i))$; we shortly say the Markov chain of policy f, or the Markov chain $P(f)$. A

Z. Hou et al. (eds.), Markov Processes and Controlled Markov Chains, 151–165.

randomised policy π is a function which assigns in a randomised way an admissible action to each state, i.e. in state i action $a \in A(i)$ is chosen with probability $\pi_i(a)$. Hence, a deterministic policy is a special case of a randomised policy. Any randomised policy π induces a transition probability matrix $P(\pi)$ of a Markov chain with the (i, j)th entry given by $\sum_a p_{ij}(a)\pi_i(a)$; we shortly say the Markov chain of policy π, or the Markov chain $P(\pi)$. We assume the reader familiar with concepts from (finite) Markov chains as *accessible, communicating, recurrent (or ergodic) state, transient state, ergodic (or recurrent) class, closed set* and *complete ergodicity (or irreducibility).*

There are several ways to classify MDPs. A first one, introduced by Bather [1], distinguishes between *communicating* and *noncommunicating* MDPs. An MDP is communicating if for every $i, j \in E$ there exists a deterministic policy f, which may depend on i and j, such that in the Markov chain $P(f)$ state j is accessible from state i. An MDP is *weakly communicating* if $E = E_1 \cup E_2$, where $E_1 \cap E_2 = \emptyset$, E_1 is a closed communicating set under some randomised policy and E_2 is a (possibly empty) set of states which are transient under all (randomised) policies. The concept of weakly communicating was proposed by Platzman [8] under the name *simply connected.*

A second kind of classification concerns the ergodic structure. We distinguish between *completely ergodic (or irreducible), unichain* and *multichain* MDPs. An MDP is completely ergodic (or irreducible) if the Markov chain $P(f)$ is completely ergodic (or irreducible) for every deterministic policy f. We say that an MDP is unichain if for every deterministic policy f the Markov chain $P(f)$ has exactly one ergodic class plus a (possibly empty) set of transient states. An MDP is multichain if there exists a deterministic policy f for which the Markov chain $P(f)$ has (at least) two ergodic classes.

It is simple to verify the following relations:

1. irreducible $\rightarrow$ communicating $\rightarrow$ weakly communicating;
2. irreducible $\rightarrow$ unichain.

A reason to classify MDPs is, for instance, that a special structure may lead to simplified algorithms for solving these MDPs under the average reward criterion. MDPs with this criterion are also called undiscounted Markov decision problems.

For a single Markov chain it is easy to determine whether or not the Markov chain belongs to a certain class. Easy means polynomially solvable, i.e. the problem belongs in terms of complexity to the class $\mathcal{P}$ of polynomial-time problems. The classification of single Markov chains will be discussed in Section 2.

For an MDP we have $\prod_{i \in E} \#A(i)$ different deterministic policies and each policy induces a Markov chain. Therefore, MDPs are also called *Markov decision chains.* Even in the case that $\#A(i) = 2$ for every $i \in E$, there are $2^{\#E}$ policies which is exponential in $\#E$. Since, usually, $\#E$ is large, the approach to analyse all Markov chains separately is prohibitive. The problem to determine whether or not an MDP belongs to a certain class is a combinatorial problem.

These MDP problems are easy (i.e. polynomially solvable), except one, which belongs to the complexity class $\mathcal{NP}$ of non-deterministic polynomial-time problems; it is an open problem whether it belongs to $\mathcal{P}$, to $\mathcal{NPC}$ (the class of $\mathcal{NP}$-complete, i.e. the most difficult, problems) or to $\mathcal{NPI} := \mathcal{NP} - (\mathcal{P} \cup \mathcal{NPC})$. In Section 3 the classification of MDPs is studied. It is also possible to use simplified algorithms after a decomposition of the state space. This kind of approach was proposed by Bather [2] and by Ross and Varadarajan [9] who introduced the term *strongly communicating classes.* Both methods are similar. Section 4 deals with this decomposition.

2. Classification of Markov chains

Consider a Markov chain with transition matrix P. The classification of a Markov chain can be executed in the associated directed graph $G(P)$. $G(P)$ has as nodes, the states of the Markov chain, and (i, j) is an arc in $G(P)$ if and only if $p_{ij} > 0$. The recurrent classes $R_1, R_2, \dots, R_m$ and the set T of transient states can be determined by the following algorithm.

Algorithm 1
(Recurrent classes and transient states of a Markov chain)

1. Determine the strongly connected components of $G(P)$, say $C_1, C_2, \dots, C_k$.
2. Set $m = 0$ and $T = \emptyset$.
3. For $i = 1$ to k do:
   if C_1 is closed: $m = m + 1$ and $R_m = C_1$;
   else: $T = T \cup C_i$.

For a single Markov chain we have the following properties:

(i) the concepts of a irreducible and communicating Markov chain coincide and correspond to $m = 1$ and $T = \phi$;

(ii) the concepts of a unichain and weakly communicating Markov chain coincide and correspond to $m = 1$;

(iii) the concept of a multichain Markov chain corresponds to $m \geq 2$.

A subgraph of a directed graph is strongly connected if for every pair of distinct vertices i and j, there exists a path from i to j as well as a path from j to i. A maximal strongly connected subgraph of a directed graph is called a *strongly connected component.*

The determination of the strongly connected components of a graph can be done in $\mathcal{O}(M)$, where M is the number of positive entries in the matrix P. This algorithm was proposed by Tarjan [10] (in Moret and Shapiro [7] a Pascal program can be found). A related approach was introduced by Fox and Landi [5]. There is also an elegant $\mathcal{O}(M)$ algorithm for the computation of the period and the cyclic sets of an ergodic class. This algorithm is due to Denardo [3].

Let $N = \#E$. Notice that $M = \mathcal{O}(N^2)$. It is easy to verify that Algorithm 1 is correct and has complexity $\mathcal{O}(M)$. Hence, all classification problems of a single Markov chain are of $\mathcal{O}(M) = \mathcal{O}(N^2)$, so they are polynomial and belong to the class $\mathcal{P}$.

3. Classification of Markov decision chains

The methods for the determination of the chain structure of an MDP use two directed graphs, G^1 and G^2, both with E as vertex set. In G^1 there is an arc (i,j) if all Markov chains of the MDP have a positive one-step transition from i to j, i.e. $p_{ij}(a) > 0$ for every action $a \in A(i)$: $\min_{a \in A(i)} p_{ij}(a) > 0$. G^2 has an arc (i,j) if at least one of the Markov chains of the MDP has a positive one-step transition from i to j, i.e. $p_{ij}(a) > 0$ for at least one action $a \in A(i)$: $\max_{a \in A(i)} p_{ij}(a) > 0$. Loops have no sense for classification problems, so they are deleted. Notice that G^1 is a subgraph of G^2.

Let $A = \sum_{i \in E} \#A(i)$, the total number of actions in the MDP. Since for the construction of G^1 and G^2 each action has to be considered and this action can generate positive one-step transitions to all N states, the construction of G^1 and G^2 has complexity $\mathcal{O}(A \cdot N)$.

We also introduce the concept of the *condensed graph* G^1_c. The condensed graph G^1_c has a (compound) vertex for each strongly connected component of G^1. Let i and j be compound vertices of G^1_c corresponding to the strongly connected components C_k and C_ℓ, and let V_k and V_ℓ be the vertex sets in G^1 of C_k and C_ℓ respectively. Then, there is an arc from i to j in G^1_c if every Markov chain of the MDP has a positive one-step transition from some vertex of V_k to some vertex of V_ℓ, i.e. $\max_{r \in V_k} \{\min_{a \in A(r)} \sum_{s \in V_\ell} p_{rs}(a)\} > 0$. States in the same strongly connected component are mutual accessible under any policy. Hence, the arc (i,j) in G^1_c means that any $s \in V_\ell$ is accessible from any $r \in V_k$ under any policy. The construction of this condensed graph is of $\mathcal{O}(A \cdot N)$. The operation 'condensation' can be repeated until there are no changes in the graph. Let $(G_c^1)^*$ be the last graph that is obtained after repeated condensation.

Example 3.1 *Let*
$E = \{1,2,3,4\}$; $A(1) = \{1,2\}$, $A(2) = A(3) = A(4) = \{1\}$.
$p_{12}(1) = 1$; $p_{13}(2) = 1$; $p_{23}(1) = p_{24}(1) = 0.5$; $p_{32}(1) = p_{34}(1) = 0.5$; $p_{41}(1) = 0.5$, $p_{42} = p_{43} = 0.25$.
Consider $G^1 = (V^1, A^1)$:
$V^1 = \{1,2,3,4\}$; $A^1 = \{(2,3),(2,4),(3,2),(3,4),(4,1),(4,2),(4,3)\}$.
The strongly connected components of G^1 *are:* $C_1 = \{1\}$ *and* $C_2 = \{2,3,4\}$. $G_c^1 = (V_c^1, A_c^1)$ *with* $V_c^1 = \{1^*, 2^*\}$, *where* 1^* *corresponds to vertex* 1 *and* 2^* *to the vertices* $\{2,3,4\}$, *and* $A_c^1 = \{(1^*,2^*),(2^*,1^*)\}$. *If the graph* G_c^1 *is condensed then we obtain* $(G_c^1)^*$ *consisting of a single vertex.*

The next theorem shows that irreducibility is equivalent to the property that $(G_c^1)^*$ consists of a single vertex.

Theorem 3.1 *An MDP is irreducible if and only if the ultimate condensation* $(G_c^1)^\star$ *consists of a single vertex.*

Proof Suppose that $(G_c^1)^*$ consists of a single vertex. From the definition of condensation it follows that each two states communicate under any deterministic policy, i.e. the Markov chain is irreducible. Next, suppose that $(G_c^1)^\star$ has at least two vertices. Each component is a (compound) vertex and there is a vertex, say i, without an incoming arc (since the compound vertices are strongly connected components). Therefore, in every state of the compound vertices $j \neq i$ an action can be chosen with transistion probabilities 0 to the states of the compound vertex i. The Markov chain under this policy is not irreducible. ■

Theorem 3.1 yields the following algorithm for the irreducibility property.

Algorithm 2 (Irreducibility)

1. Construct the graph G^1 and let $G = G^1$
2. Determine the strongly connected components of G, say $C_1, C_2, \ldots, C_k$.
3. If all components consist of one vertex: go to step 4;
   Otherwise: construct the condensed graph G_c, let $G = G_c$ and go to step 2.

4. If $k = 1$: the MDP is irreducible;
   If $k \geq 2$: the MDP is not irreducible.

Theorem 3.2 *The time-complexity of* Algorithm 2 *is* $\mathcal{O}(A \cdot N^2)$.

Proof The construction of G^1 and G^1_c has complexity $\mathcal{O}(A \cdot N)$. The determination of the strongly connected components has as order the total number of arcs, which is $\mathcal{O}(N^2) \leq \mathcal{O}(A \cdot N)$. Hence, it is sufficient to show that the number of iterations is at most N. Each new iteration starts with a condensed graph for which the strongly connected components have to be determined. If each component consists of a single vertex, the algorithm terminates; if not, the next condensed graph has at least one fewer vertex. ■

Next, we consider the problem to decide whether an MDP is communicating. The result is based on the following theorem.

Theorem 3.3 *An MDP is communicating if and only if the graph* G^2 *is strongly connected.*

Proof Notice that (i, j) is an arc in G^2 if and only if $\max_{a \in A(i)} p_{ij}(a) > 0$. Hence, for any completely mixed stationary strategy π (i.e. $\pi_i(a) > 0$ for all i and a) $p_{ij}(\pi) > 0$ if and only if (i, j) is an arc in G^2. With this interpretation it is obvious that the concept of a communicating MDP is equivalent to the strongly connectness of G^2. ■

Note In Filar and Schultz [4] it is shown that communicating is also equivalent to the following condition: For every $b \in \mathbb{R}^N$ with $\sum_{i=1}^N b_i = 0$, there exists $y = \{y_i(a) \mid a \in A(i), i \in E\}$, where y may depend on b, such that $y_i(a) \geq 0$ for all $a \in A(i)$, $i \in E$, and $\sum_{i,a} [\delta_{ij} - p_{ij}(a)] y_i(a) = b_j$, $j \in E$. This last system is related to linear programs to solve undiscounted MDPs (see e.g. Kallenberg [6]).

Algorithm 3 (Communicating)

1. Construct the graph G^2.

2. Determine the strongly connected components of G^2, say $C_1, C_2, \ldots, C_k$.

3. If $k = 1$: the MDP is communicating;
   If $k \geq 2$: the MDP is not communicating.

Theorem 3.4 *The time-complexity of* Algorithm 3 *is* $\mathcal{O}(A \cdot N)$.

Proof The construction of G^2 has complexity $\mathcal{O}(A \cdot N)$ and the determination of the strongly connected components has order $\mathcal{O}(N^2) \leq \mathcal{O}(A \cdot N)$. ■

The investigation of the property weakly communicating can be done analogously by the following algorithm which was proposed in Platzman [8].

Algorithm 4 (Weakly communicating)

1. Construct de graph G^2
2. Determine the strongly connected components of G^2, say $C_1, C_2, \ldots, C_k$.
3. a) Set $m = 0$ and $T = \emptyset$.
   b) For $i = 1$ to k do:
   if C_i is closed: $m = m + 1$ and $R_m = C_i$;
   else: $T = T \cup C_1$.
4. If $m \geq 2$: the MDP is not weakly communicating;
   If $m = 1$:
   if $T = \emptyset$: the MDP is communicating, implying weakly communicating;
   else: go to step 5.
5. a) Let $c_i = 1$ for $i \notin T$ and $c_i = 0$ for $i \in T$;
   b) $S = \emptyset$;
   c) For every $i \in T$ do:
   if $\sum_j p_{ij}(a)c_j > 0$ for every $a \in A(i)$: $c_i = 1$ and $S = S \cup \{i\}$.
   d) If $S = \emptyset$: the MDP is not weakly communicating;
   else: $T = T \backslash S$ and go to 5e;
   e) if $T = \emptyset$: the MDP is weakly communicating;
   else: go to 5b.

Theorem 3.5
Algorithm 4 *is correct and the time-complexity is* $\mathcal{O}(A \cdot N^2)$.

Proof Weakly communicating means that each state is either transient under all policies or an element of a communicating class under some policy. If the algorithm ends in step 4 with $m = 1$ and $T = \emptyset$, then the MDP is communicating, so certainly weakly communicating. If the algorithm terminates in step 4 with $m \geq 2$, then there are two ergodic sets under all policies: the MDP is not weakly communicating. If the algorithm terminates in step 5d, then there is a state which does not belong to the maximal communicating class and will not reach this maximal communicating class under all policies, so it is recurrent for some policy: the MDP is not weakly communicating. Finally, when T becomes empty in step 5e, the states of the original set T are transient under all policies, so the MDP is weakly communicating.

For the complexity we remark that the steps 1 until 4 are executed only once and have complexity $\mathcal{O}(A \cdot N)$ as shown before. Step 5 can be executed at most N times, since at each iteration the set T becomes strictly smaller. In each iteration the computation $\sum_j p_{ij}(a)c_j$ has to be executed for every $i \in T$ and $a \in A(i)$. Each computation is $\mathcal{O}(N)$ and has to be done $\mathcal{O}(A)$ times: the overall computation of step 5 is $\mathcal{O}(A \cdot N^2)$. ■

We can conclude that the classification problem to decide whether an MDP is irreducible, communicating, weakly communicating or noncommunicating is polynomially solvable. Next, we continue with the distinction between unichain and multichain. Before we present the algorithm we first give some ideas of the method. We start with a description of the operation for states with outdegree 1.

States with outdegree 1

Suppose that Algorithm 2 (the algorithm for irreducibility) terminates with the conclusion 'not irreducible'. If there is a (compound) vertex with outdegree 1, say from i to j, then the states corresponding to the (compound) vertex j are accessible from the states of the (compound) vertex i under any policy. Therefore, if the MDP is multichain, then the states of i and j can not belong to different recurrent classes. Hence, the states of i and j can be merged into a new compound vertex. Then, it has to be considered whether there are new arcs in the new graph, similar as we did for condensation.

Starting with the graph G^1, the ultimate condensed graph $(G_c^1)^\star$ is constructed and the operation for states of outdegree 1 is executed. These two procedures are repeated sequentially. By $(G^1)^+$ we denote the graph which is finally obtained.

Theorem 3.6 *Let k^+ be the number of strongly connected components of the graph $(G^1)^+$.*

(i) If $k^+ = 1$, then the MDP is unichained.

(ii) If $k^+ = 2$, then the MDP is multichained.

(iii) If $k^+ \geq 3$, then the MDP is either multichained or unichained.

Proof

(i) Suppose that $k^+ = 1$ and the MDP is not unichain, i.e. there is a policy f such that the Markov chain $P(f)$ has at least two ergodic sets. By the above described construction these two ergodic sets will never be combined into one compound vertex. Hence, $k^+ \geq 2$: contradiction.

(ii) Suppose that $k^+ = 2$ and that the final graph has the (compound) vertices $i^\star$ and $j^\star$. Remark that there is no arc from $i^\star$ to $j^\star$ or vice versa. That implies that there are policies f^1 and f^2 such that $P(f^1)$ is closed in $i^\star$ and $P(f^2)$ in $j^\star$. Hence, f^1 and f^2 can be combined in a policy f^3 which is multichained.

(iii) Consider the following model: $E = \{1,2,3\}$; $A(1) = A(2) = A(3) = \{1,2\}$; $p_{12}(1) = p_{13}(2) = p_{21}(1) = p_{23}(2) = p_{31}(1) = p_{32}(2) = 1$. Notice that this model is unichain and that $G^1 = (V^1, A^1) : V^1 = \{1,2,3\}; A^1 = \emptyset$. Hence, $(G^1)^+ = G^1$ and $k^+ = 3$. So, $k^+ \geq 3$ and the MDP is unichained, is possible (that k^+ can be at least 3 for a multichain model is obvious). ■

Example 3.2 *Let*
$E = \{1,2,3,4,5\}$; $A(1) = A(2) = \{1,2\}, A(3) = A(4) = A(5) = \{1\}$.
$p_{13}(1) = 1$; $p_{11}(2) = p_{14}(2) = 0.5$; $p_{23}(1) = 1$; $p_{22}(2) = 1$; $p_{32}(1) = p_{31}(1) = 0.5$; $p_{43}(1) = 1$; $p_{53}(1) = p_{55}(1) = 0.5$.
Consider $G^1 = (V^1, A^1) : V^1 = \{1,2,3,4,5\}; A^1 = \{(3,1),(3,2),(4,3),(5,3)\}$. *The strongly connected components of* G^1 *are:* $C_i = \{i\}, 1 \leq i \leq 5 \cdot (G^1_c)^\star = G^1$.
Then, since the vertices 4 *and* 5 *have outdegree 1 (both to state 3), they are merged into the compound vertex* $3^\star$ *consisting of the states* $\{3,4,5\}$. *For the remaining problem we have* $G^1 = (V^1, A^1)$ *with* $V^1 = \{1,2,3^\star\}$ *and* $A^1 = \{(1,3^\star),(3^\star,1),(3^\star,2)\}$. *The strongly connected components of* G^1 *are:* $C_1 = \{1,3^\star\}$, $C_2 = \{2\}$. $(G^1_c)^\star = (V^1_c, A^1_c)$ *with* $V^1_c = \{1^\star, 2\}$ *where* $1^\star$ *corresponds to the vertices* 1 *and* $3^\star$, *and* $A^1_c = \{(1^\star,2)\}$. *Then, the vertices* $1^\star$ *and* 2 *are merged into one final compound vertex* $2^\star : k^+ = 1$. *From Theorem 3.6 it follows that the MDP is unichain.*

Note The counterintuitive difference between the property that $k^+ = 2$ yields a multichain MDP and $k^+ \geq 3$ can result in a unichain MDP is caused by the fact that in $k^+ = 2$ no arcs from $1^\star$ means that $1^\star$ can be a closed (compound) vertex and in $k^+ \geq 3$ no arcs from $1^\star$ can mean that arcs go either to a (compound) vertex $2^\star$ or to a (compound) vertex $3^\star$.

Algorithm 5 (Unichain or multichain)

1. Construct, by Algorithm 2, from graph G^1 the (repeatedly) condensed graph $(G_c^1)^\star$.

2. If there are (compound) vertices with outdegree 1:

    a) execute the merging operation;

    b) add actions from and to the new compound vertex (if they exist corresponding to the definition of G^1 for the new graph G^1).

3. a) Let $(G^1)^+$ be the graph obtained after the steps 1 and 2.

    b) If $(G^1)^+ = G^1$: go to step 4;
    else: $G^1 = (G^1)^+$ and go to step 1.

4. If $k^+ = 1$, then the MDP is unichained;
   if $k^+ = 2$, then the MDP is multichained;
   if $k^+ \geq 3$, then no decision can be made.

Theorem 3.7 *The complexity of* Algorithm 5 *is* $\mathcal{O}(A \cdot N^2)$.

Proof The construction of G^1, G_c^1 and the determination of the strongly connected components are of order $\mathcal{O}(A \cdot N)$. This can be done at most N times (because each time this has to be done, the number of states is strictly smaller). Furthermore, the total work for merging states and adding actions during one iteration is of order $\mathcal{O}(A)$. Hence, the overall complexity is $\mathcal{O}(A \cdot N^2)$. ■

Proof

1. It is obvious that the recognition problem 'is the MDP multichained' is in $\mathcal{NP}$: the certification can be done in $\mathcal{O}(N^2)$.

2. If the algorithm ends in step 4 with $k^+ \geq 3$, then each policy corresponding to the (in general smaller) MDP of the last graph $(G^1)^+$ can be analysed (as in Section 2) to decide whether the MDP is unichain or multichain.

3. If the MDP is *deterministic*, i.e. all transition probabilities $p_{ij}(a)$ are 0 or 1, then there is a one-to-one correspondence between the arcs of the graph G^2 and the action set of the MDP. Furthermore, the MDP is multichain if and only if the graph G^2 has two vertex-disjoint simple cycles. The simple cycles of a directed graph can be detected by the method described in Weinblatt [11]. Since a

graph may have an exponential number of simple cycles, the determination of all simple cycles is $\mathcal{NP}$-hard. However, the algorithm has experimentally been tested, and the tests indicate that the algorithm is reasonable fast (see Weinblatt [11]).

4. Decomposition of Markov decision chains

In Bather [2] a decomposition of the state space is described based on the accessibility between the states. The state space is divided into several levels. The first level L_1 contains the closed, communicating subsets of the state space. Hence, L_1 consists of the closed, strongly connected components of G^2 and can be determined in $\mathcal{O}(A \cdot N)$.

For the next step in the decomposition, we consider $E_1 := E \backslash L_1$, i.e. the states in the open strongly connected components of G. It will be useful to distinguish between the states from which for any policy absorption in L_1 will occur (the transient set T_1) and the states from which absorption in L_1 can be avoided by an appropriate choice of the policy (the 'new' set E_1). T_1 and the 'new' E_1 can be computed similarly to step 5 of Algorithm 4.

Algorithm 6 (Determination of T_1)

1. Let $c_i = 1$ for $i \in L_1$ and $c_i = 0$ for $i \in E_1$: let $T_1 = \emptyset$.
2. $S = \emptyset$.
3. For every $i \in E_1$ do:
   if $\sum_j p_{ij}(a)c_j > 0$ for every $a \in A(i)$: $c_i = 1$ and $S = S \cup \{i\}$.
4. If $S = \emptyset$: stop;
   else: $T_1 = T_1 \cup S, E_1 = E_1 \backslash S$ and go to step 5.
5. if $E_1 = \emptyset$: stop;
   else: go to step 2.

If the 'new' $E_1 \neq \emptyset$ then, by an appropriate choice of the deterministic policy f, E_1 is closed under $P(f)$. Hence, after deleting for $i \in E_1$ the actions $a \in A(i)$ with $\sum_{j \in L_1 \cup T_1} p_{ij}(a) > 0$, the resulting (to E_1) restricted model is again an MDP and can be treated in the same way, i.e. we can construct a second level with L_2 and a (possibly empty) set transient set T_2. In this way we proceed until all states are assigned to an L- or T-set.

Algorithm 7 (Bather decomposition of the state space)

1. a) $m = 0$; $E_m = E$.
   b) Construct the graph G^2 for the MPD corresponding to state space E_m.
   c) $m = m + 1$.
   d) Determine the set L_m of strongly connected and closed components of G^2.
   e) $E_m = E_{m-1} \backslash L_m$.

2. a) Let $c_i = 1$ for $i \in L_m$ and $c_i = 0$ for $i \in E_m$; let $T_m = \emptyset$.
   b) $S = \emptyset$.
   c) For every $i \in E_m$ do:
   if $\sum_{j \in E_{m-1}} p_{ij}(a)c_j > 0$ for every $a \in A)i)$: $c_i = 1$ and $S = S \cup \{i\}$.
   d) If $S = \emptyset$: go to step 3;
   else: $T_m = T_m \cup S, E_m = E_m \backslash S$ and go to step 2e.
   e) if $E_m = \emptyset$: stop;
   else: go to step 2b.

3. a) For $i \in E_m$ do
   for $a \in A(i)$ do if $\sum_{j \in L_m \cup T_m} p_{ij}(a) > 0 : A(i) = A(i) \backslash \{a\}$.
   b) go to step 1b.

Example 4.1 *Consider the following MDP:*
$E = \{1,2,3,4,5,6\}$; $A(1) = A(3) = \{1,2\}$, $A(2) = A(5) = A(6) = \{1\}, A(4) = \{1,2,3\}$; $p_{12}(1) = p_{14}(2) = p_{23}(1) = p_{32}(1) = p_{35}(2) = p_{43}(1) = p_{45}(2) = p_{44}(3) = p_{56}(1) = p_{66}(1) = 1$.

Level 1: $G^2 = (V, A)$ *with* $V = \{1,2,3,4,5,6\}$ *and*

$A = \{(1,2),(1,4),(2,3),(3,2),(3,5),(4,3),(4,5),(4,4),(5,6),(6,6)\}$. G^2 *has as strongly connected components:*

$C_1 = \{1\}$, $C_2 = \{2,3\}$, $C_3 = \{4\}$, $C_4 = \{5\}$ *and* $C_5 = \{6\}$.

$L_1 = \{6\}$; $E_1 = \{1,2,3,4,5\}$. $c_6 = 1, c_1 = c_2 = c_3 = c_4 = c_5 = 0$. $T_1 = \emptyset$.

$S = \emptyset$; $c_5 = 1$; $S = \{5\}$; $T_1 = \{5\}$; $E_1 = \{1,2,3,4\}$.

$S = \emptyset$; $A(3) = \{1\}$, $A(4) = \{1,3\}$.

Level 2: $G^2 = (V, A)$ *with* $V = \{1,2,3,4\}$ *and*

$A = \{(1,2),(1,4),(2,3),(3,2),(4,3),(4,4)\}$.

G^2 has as strongly connected components: $C_1 = \{1\}$, $C_2 = \{2,3\}$, $C_3 = \{4\}$.
$L_2 = \{2,3\}$; $E_2 = \{1,4\}$. $c_2 = c_3 = 1$; $c_1 = c_4 = 0$. $T_2 = \emptyset$.
$S = \emptyset$. $A(1) = \{2\}$, $A(4) = \{3\}$.

Level 3: $G^2 = (V, A)$ *with* $V = \{1,4\}$ *and* $A = \{(1,4),(4,4)\}$.
G^2 has as strongly connected components: $C_1 = \{1\}, C_2 = \{4\}$.
$L_3 = \{4\}; E_3 = \{1\}$. $c_4 = 1, c_1 = 0$. $T_3 = \emptyset$.
$S = \emptyset$. $c_1 = 1; S = \{1\}; T_3 = \{1\}; E_3 = \emptyset$.

The decomposition ends with the following levels:
E_1: $T_1 = \{5\} \rightarrow \{6\}$
E_2: $\{2,3\}$
E_3: $T_3 = \{1\} \rightarrow \{4\}$

Theorem 4.1 *The complexity of the Bather decomposition is $\mathcal{O}(A \cdot N^2)$.*

Proof Since in each iteration $m := m+1$ and $L_m \neq \emptyset$, the algorithm terminates after at most N iterations. Consider the complexity of one iteration.

Step 1: part b is of $\mathcal{O}(A \cdot N)$, part c of $\mathcal{O}(1)$ and part d of $\mathcal{O}(N^2)$.

Step 2: part a is of $\mathcal{O}(N)$, part b of $\mathcal{O}(1)$, part c of $\mathcal{O}(N \cdot A)$, part d of $\mathcal{O}(N)$ and part e of $\mathcal{O}(1)$.

Step 3: part a is of $\mathcal{O}(A \cdot N)$ and part b of $\mathcal{O}(1)$.

Hence, the complexity of one iteration is $\mathcal{O}(A \cdot N)$. Therefore, the overall complexity is $\mathcal{O}(A \cdot N^2)$. ■

Ross and Varadarajan [9] have presented a similar decomposition method. In this decomposition the state space is partitioned into strongly communicating classes $C_1, C_2, \ldots, C_m$ and a set T of transient states with the following properties:

(i) the states of T are transient under all stationary policies;

(ii) suppose $R(\pi)$ is a recurrent class under some stationary policy π, then $R(\pi) \subseteq C_i$ for some $1 \leq i \leq m$;

(iii) there exists a policy π such that $C_1, C_2, \ldots, C_m$ are the recurrent states of the Markov chain $P(\pi)$.

A set states $S \subseteq E$ is called a *strongly communicating* set if there exists a stationary policy π such that $S \subseteq R_i(\pi)$, where $R_i(\pi)$ is a recurrent class of the Markov chain $P(\pi)$. Hence, strongly communicating implies communicating. Example 4 (see below) shows that the reverse statement is not true. A (strongly) communicating set S is a (strongly) communicating *class* if S is maximal with respect to the property (strongly) communicating. Notice that a closed communicating class is a strongly communicating class.

Example 4.2 *Consider the Markov chain*

$$P = \begin{pmatrix} 0 & 1 & 0 \\ 0.5 & 0 & 0.5 \\ 0 & 0 & 1 \end{pmatrix}$$

$S = \{1,2\}$ *is a communicating class, but not strongly communicating.*

The sets E_i in the Bather decomposition are the strongly communicating classes C_i of Ross and Varadarajan. The union of the transient sets T_i in the Bather decomposition is the transient set T in the approach of Ross and Varadarajan. Hence, Algorithm 7 also gives the decomposition by Ross and Varadarajan. A formal proof of these properties can be found in Ross and Varadarajan [9].

References

[1] Bather, J.(1973a). "Optimal decision procedures for finite Markov chains. Part II: Communicating systems", Advances in Applied Probability **5**, 521-540.

[2] Bather, J. (1973b). "Optimal decision procedures for finite Markov chains. Part III: General convex systems", Advances in Applied Probability **5**, 541-553.

[3] Denardo, E.V. (1977). "Periods of connected networks and powers of nonnegative matrices", Mathematics of Operations Research **2**, 20-24.

[4] Filar, J.A. and T.A. Schultz (1988). "Communicating MDPs: equivalence and LP properties", Operations Research Letters **7**, 303-307.

[5] Fox, B.L. and D.M. Landi (1968). "An algorithm for identifying the ergodic subchains and transient states of a stochastic matrix", Communications of the ACM **11**, 619-621.

[6] Kallenberg, L.C.M. (1983). "Linear programming and finite Markovian control problems", Mathematical Centre Tract **148**, Amsterdam.

[7] Moret, B.M.E. and H.D. Shapiro (1990). " Algorithms from $\mathcal{P}$ to $\mathcal{NP}$, Volume I: Design and efficiency", Benjamin/Cummings, Redwood City.

[8] Platzman, L.K. (1977). "Improved conditions for convergence in undiscounted Markov renewal programming", Operations Research **25**, 529-533.

[9] Ross, K.M. and R. Varadarajan (1991). "Multichain Markov decision processes with a sample path constraint: a decomposition approach", Mathematics of Operations Research **16**, 195-207.

[10] Tarjan, R.E. (1972). "Depth-first search and linear graph algorithms", SIAM Journal of Computing **1**, 146-160.

[11] Weinblatt, H. (1972). "A new search algorithm for finding the simple cycles of a finite directed graph", Journal Association for Computing Machinery **19**, 43-56.

Chapter 10

OPTIMALITY CONDITIONS FOR CTMDP WITH AVERAGE COST CRITERION*

Xianping Guo
Department of Mathematics
Zhongshan University,
China
gxp@zsu.edu.cn

Weiping Zhu
Department of Computer Science and Electrical Engineering
The University of Queensland,
Australia
weiping@csee.uq.edu.au

Abstract In this paper, we consider continuous time Markov decision processes with (possibly unbounded) transition and cost rates under the average cost criterion. We present a set of conditions that is weaker than those in [5, 11, 12, 14], and prove the existence of optimal stationary policies using the optimality inequality. Moreover, the theory is illustrated by two examples.

Keywords: Continuous Time Markov Decision Processes (CTMDPs); Average Cost Criterion; Optimality Inequality (OIE); Optimal Stationary Policies.

1. Introduction

The research on continuous time Markov decision processes (CT-MDPs) is an important part in the theory and applications of Markov

*This research has been supported partially by Natural Science Foundation of Guangdong Province, by Foundation of Hongkong Zhongshan University Advanced Research Center, and by the University of Queensland under Grant No. 98/UQNSRG025G.

Z. Hou et al. (eds.), Markov Processes and Controlled Markov Chains, 167–188.

decision processes (MDPs). It has been studied by many researchers, such as Bather [1], Bertsekas [2], Doshi [3], Howard [4], Kakumanu [5], Lippman [6], Miller [7], Puterman [8], Yushkevich and Feinberg [9], Walrand [10], etc.. Both average and discounted reward (or cost) criteria are often used in the study of CTMDPs. This paper deals with the average (possibly unbounded) cost criterion for CTMDPs. We are concerned with axioms that guarantee the existence of an average cost optimal stationary policy. In [1]–[13], many sets of conditions for this purpose have been provided and the existence of optimal policies is proved using the optimality equation (OE). Guo and Liu [14] replaced the OE by the optimality inequality (OIE). They not only proved the existence of average optimal stationary policies under conditions weaker than those used in [5, 11, 12, 13] for the OE, but also gave an example to show that the conditions that ensure the existence of a solution of the OIE do not imply the existence of a solution of the OE. In the spirit of [8, 15, 16] on discrete time MDPs, we provide a new set of conditions, based on optimal discounted cost values, which is weaker than those used in [5, 11, 12, 14], and prove the existence of both the average cost optimal stationary policies and solutions to the OIE. Moreover, an admission control queueing model and controlled birth and death processes are given for which the new set of conditions holds, whereas the conditions in [5, 11, 12, 14] fail to hold. The conditions and results in this paper are very similar to those in Puterman [2] and Sennott [15, 16] on discrete time MDPs. Hence, this paper extends recent work to CTMDPs.

In Section 2, we present the model, notation and definitions. In Section 3 some results for CTMDPs with discounted cost criterion are provided. In Section 4, we present the conditions to establish OIE for CTMDPs with average cost criterion and prove the existence of optimal stationary policies. In Section 5 we give an admission control queueing model and controlled birth and death processes to illustrate the results of this paper.

2. Model, notation and definitions

We observe continuously a controlled system in which, when the system is at state i of a denumerable space S, a decision maker chooses an action a from a set $A(i)$ of available actions. There are two consequences for the action:

1. the decision maker pays a cost rate $r(i, a)$, and

2. the system moves to a new state j, $j \in S$, according to a transition rate $q(j|i, a)$.

The goal of the decision maker is to choose a sequence of actions to make the system perform optimally with respect to some predetermined performance criterion V. So the model can be described by a five-element tuple $\{S, (A(i), i \in S), r, q, V\}$ having the following properties:

(i) the state space S is denumerable;

(ii) every available action set $A(i)$ is a measurable subset of a measurable action space A, with σ-algebra T;

(iii) the cost rate r is a function bounded below on $K := \{(i,a)|i \in S, a \in A(i)\}$;

(iv) the transition rate q satisfies: $q(j|i,a) \geq 0 \ \forall i \neq j, \ a \in A(i), i,j \in S$, and furthermore, $\sum_{j \in S} q(j|i,a) = 0 \ \forall i \in S, \ a \in A(i)$, and $q(i) := \sup_{a \in A(i)}(-q(i|i,a)) < \infty$ for $i \in S$;

(v) V is a discounted (or average) cost criterion, which is defined below.

A randomized Markov policy π is a family $\{\pi_t, t \geq 0\}$ satisfying:

1. for any $t \geq 0$ and $i \in S$, $\pi_t(\cdot|i)$ is a probability measure on A such that $\pi_t(A(i)|i) = 1$;
2. for all $B \in T$ and $i \in S$, $\pi_t(B|i)$ is a Lebesgue measurable function in t on $[0, \infty)$.

The set of all randomized Markov policies is denoted by Π_m. A policy $\pi = \{\pi_t, t \geq 0\}$ is called randomized stationary if $\pi_t(B|i) \equiv \pi_0(B|i) \ \forall t \geq 0, \ B \in T, \ i \in S$. We denote this policy by π_0^∞. The set of all randomized stationary policies is denoted by Π_s. A policy $\pi = \pi_0^\infty \in \Pi_s$ is called stationary if there exists $f \in F := \{f | f : S \to A, f(i) \in A(i), i \in S\}$ such that $\pi_0(f(i)|i) = 1$ for every $i \in S$. We denote this policy by f^∞ (or f, for short). The set of all stationary policies is denoted by Π_s^d.

For any $\pi = \{\pi_t, t \geq 0\} \in \Pi_m$, let

$$q_{ij}(t,\pi) := \int_A q(j|i,a)\pi_t(da|i), \qquad i,j \in S, \ t \geq 0, \tag{2.1}$$

$$r(t,i,\pi) := \int_A r(i,a)\pi_t(da|i), \qquad i \in S, \ t \geq 0. \tag{2.2}$$

In particular, when $\pi = f^\infty \in \Pi_s^d$, we write $q_{ij}(t,\pi)$ and $r(t,i,\pi)$ as $q(j|i,f(i))$ and $r(i,f(i))$ respectively.

For each $\pi = \{\pi_t, t \geq 0\}$, let $\{Q(t,\pi) := (q_{ij}(t,\pi)), t \geq 0\}$ be the CTMDP infinitesimal generator. The minimum transition matrix with respect to $\{Q(t,\pi), t \geq 0\}$ is denoted by $P^{min}(s,t,\pi) = \left(p_{ij}^{min}(s,t,\pi)\right)$. Let $P^{min}(t,\pi) := P^{min}(0,t,\pi)$ for $t \geq 0$.

In this paper, we propose the following conditions.

Assumption 2.1 *There exist m nonnegative functions w_n, $n = 1,2,\ldots,m$, such that:*

(i) for all $i \in S$, $a \in A(i)$, and $n = 1,2,\ldots,m-1$,

$$\sum_{j \in S} q(j|i,a) w_n(j) \leq w_{n+1}(i); \tag{2.3}$$

(ii) for all $i \in S$ and $a \in A(i)$,

$$\sum_{j \in S} q(j|i,a) w_m(j) \leq 0. \tag{2.4}$$

Definition 2.1 *A function h on S is said to satisfy* Assumption 2.1 *if* $|h| \leq (w_1 + \cdots + w_m)$.

Assumption 2.2

(i) $R := (w_1 + \cdots + w_m) \geq 1$;

(ii) for all $i \in S$, $t > s \geq 0$ and $\pi \in \Pi_m$,

$$\int_s^t \sum_{j \in S} p_{ij}^{min}(s,u,\pi) q_j(u,\pi) R(j) du < \infty, \tag{2.5}$$

where $q_j(u,\pi) := -q_{jj}(u,\pi)$ for $j \in S$, $u \geq 0$, and w_n comes from Assumption 2.1.

Assumption 2.3

$$|r(i,a)| \leq MR(i), \qquad i \in S, \ a \in A(i), \ \text{for some } M \geq 0.$$

Assumption 2.4 Assumptions 2.1, 2.2 *and* 2.3 *hold.*

Remark 2.1 *In* Lemma 3.2 *below, we provide some conditions and examples that guarantee that* Assumption 2.4 *holds.*

Now we define the discounted cost criterion V_α and the average cost criterion $\bar{V}$, as well as the optimal cost values based on these two criteria.

For any $\pi \in \Pi_m$, $i \in S$ and $\alpha > 0$, let

$$V_\alpha(\pi, i) := \int_0^\infty e^{-\alpha t} \sum_{j\in S} p_{ij}^{min}(t,\pi) r(t,j,\pi)\,dt, \tag{2.6}$$

$$\bar{V}(\pi, i) := \limsup_{T\to\infty} \frac{1}{T} \int_0^T \sum_{j\in S} p_{ij}^{min}(t,\pi) r(t,j,\pi)\,dt, \tag{2.7}$$

$$V_\alpha^*(i) = \inf_{\pi\in\Pi_m} V_\alpha(\pi,i), \text{ and } \bar{V}^*(i) = \inf_{\pi\in\Pi_m} \bar{V}(\pi,i). \tag{2.8}$$

A policy $\pi^* \in \Pi_m$ is called discounted cost optimal if $V_\alpha(\pi^*, i) = V_\alpha^*(i)$ for all $i \in S$. Average cost optimal policies are defined similarly.

Remark 2.2 *Under* Assumptions 2.1 *and* 2.2, *we know that $P^{min}(t,\pi)$ is honest for every $\pi \in \Pi_m$. Hence, by* (2.6), (2.7) *and* (2.8), *we may increase r by adding constant without affecting the discussion on the existence of optimal policies. Therefore, we always assume $r \geq 0$.*

Remark 2.3 *Throughout this paper, a function on S is regarded as a column vector, and operations on matrices and vectors are componentwise.*

3. Discounted cost optimality

In this section we provide some results on the discounted cost criterion, which are essential to the discussion on the average cost criterion.

Lemma 3.1 *If* Assumption 2.1 *holds, then for any $\pi \in \Pi_m$ and $t \geq s \geq 0$, we have*

(i)
$$P^{min}(s,t,\pi)R \leq \sum_{k=1}^m \frac{1}{(k-1)!}(t-s)^{k-1} R_k; \tag{3.1}$$

(ii)
$$\int_s^\infty e^{-\alpha(t-s)} P^{min}(s,t,\pi) R\,dt \leq \sum_{k=1}^m \alpha^{-k} R_k \leq \left(\sum_{k=1}^m \alpha^{-k}\right) R. \tag{3.2}$$

where, $R_k := w_k + w_{k+1} + \cdots + w_m$ for $k = 1, 2, \ldots, m$, and $R_1 := R$.

Proof See [18, Lemma 2]. ■

Lemma 3.2 *If one of the following conditions holds, then* Assumption 2.4 *holds.*

(i) $\|r\| := \sup_{i\in S, a\in A(i)} r(i,a) < \infty$, $\|q\| := \sup_{i\in S} q(i) < \infty$.

(ii) Assumption 2.1 *holds,* $r \leq MR$ *for some* $M > 0$, *and* $\|q\| < \infty$.

(iii) Assumption 2.1 *holds,* $r(i,a) \leq MR(i)$ $\forall i \in S$ *and* $a \in A(i)$, *for some* $M > 0$, *and the function* $qR(i) := q(i)R(i)$ *on* S *satisfies* Assumption 2.1.

(iv) For all $i \in S \equiv \{0, 1, \cdots\}$, $A(i) := \{0, 1\}$:
let $q(0|0,0) = 0$ *and for* $i \geq 1$,

$$q(j|i,0) = \begin{cases} \mu i & \text{if } j = i-1, \\ -(\mu+\lambda)i & \text{if } j = i, \\ \lambda i & \text{if } j = i+1, \\ 0 & \text{otherwise}; \end{cases}$$

let $q(0|0,1) = -v$, $q(1|0,1) = v$, *and for* $i \geq 1$,

$$q(j|i,1) = \begin{cases} \mu i & \text{if } j = i-1, \\ -(\mu+\lambda)i - v & \text{if } j = i, \\ \lambda i + v & \text{if } j = i+1, \\ 0 & \text{otherwise}; \end{cases}$$

where $0 \leq \lambda \leq \mu$, $v \geq 0$, *there are* k *positive numbers* b_n, $n = 1, 2, \ldots k$ *such that* $r(i,a) \leq \sum_{n=1}^{k} b_n i^n$ *for all* $a \in A(i)$, $i \in S$.

(v) $S \equiv \{0, 1, \ldots\}$, $A(i) = \{0, 1\}$, $i \in S$:
let $q(0|0,0) = 0$ *and for* $i \geq 1$,

$$q(j|i,0) = \begin{cases} \mu & \text{if } j = i-1, \\ -\mu & \text{if } j = i, \\ 0 & \text{otherwise}; \end{cases}$$

let $q(0|0,1) = -(\sum_{k=1}^{\infty} \lambda_k), q(k|0,1) = \lambda_k$ $\forall k \geq 1$, *and for* $i \geq 1$,

$$q(j|i,1) = \begin{cases} \mu & \text{if } j = i-1, \\ -(\mu + \sum_{k=1}^{\infty} \lambda_k) & \text{if } j = i, \\ \lambda_k & \text{if } j = i+k, \\ 0 & \text{otherwise}; \end{cases}$$

where $\sum_{k=1}^{\infty} k\lambda_k < \infty$; $r(i,0) = pi + c$, $r(i,1) = pi$, $i \in S$, $p, c > 0$.

Proof Under condition (i), this lemma is obviously valid. Under (ii), the lemma follows from Lemma 3.1 and the condition $\sum_{j\in S} q(j|i,a) = 0$ $\forall i \in S$, $a \in A(i)$. On the other hand, since $q_j(u,\pi) \leq q(j)$ $\forall j \in S$,

$\pi \in \Pi_m$, and $u \geq 0$, by Lemma 3.1, we can also prove this lemma under condition (iii). Similarly, under condition (iv) or (v), this lemma can be proved by applying Lemma 3.1. The calculation is straightforward, but lengthy, and we shall omit the details here. ■

By Assumption 2.1, we can define

$$B(S) := \{u : |u(i)| \leq cR(i) \ \forall i \in S, \text{ for some constant } c > 0\}. \quad (3.3)$$

Lemma 3.3 *If* Assumption 2.4 *holds, then for all* $i \in S$ *and* $\pi \in \Pi_m$, *we have:*

(i)
$$\sum_{j\in S} p_{ij}^{min}(s,t,\pi) = 1, \qquad t > s \geq 0; \quad (3.4)$$

(ii)
$$|V_\alpha(\pi,i)| \leq \sum_{k=1}^{m} \alpha^{-k} R_k(i) \leq \left(\sum_{k=1}^{m} \alpha^{-k}\right) R(i); \quad (3.5)$$

(iii) $\pi \in \Pi_m$ *is discounted cost optimal if and only if* $V_\alpha(\pi)$ *is a solution of the following dynamic programming equation within* $B(S)$,

$$\alpha u(i) = \inf_{a\in A(i)} \left\{ r(i,a) + \sum_{j\in S} q(j|i,a)u(j) \right\}; \quad (3.6)$$

(iv) $V_\alpha^*(i) = \inf_{f\in F} V_\alpha(f,i)$, *and* V_α^* *is the unique solution of* (3.6) *within* $B(S)$;

(v) Any $f(\in F)$ *realizing the minimum on the right-hand of* (3.6) *is discounted cost optimal.*

Proof See [18, Theorem 1 and Lemmas 2 and 4]. ■

Remark 3.1 Lemma 3.3(i) *shows that the minimum transition matrix* $P^{min}(s,t,\pi)$ *is honest and unique for every* $\pi \in \Pi_m$, *and will be denoted by* $P(s,t,\pi)$. *That is,* $P^{min}(s,t,\pi) = P(s,t,\pi)$ *for* $t \geq s \geq 0$, *and,* $P^{min}(t,\pi) = P(t,\pi)$ *for* $t > s \geq 0$.

Lemma 3.4
If

(i) Assumption 2.4 *holds; and*

(ii) *for every* $i \in S$, $A(i)$ *is compact, and* $r(i,a)$ *and* $\sum_{j\in S} q(j|i,a)R(j)$ *are continuous in* a *on* $A(i)$,

then, for each $\alpha > 0$ *there exists a discounted cost optimal stationary policy* f_α^*.

Proof For every $i \in S$, by [17, Lemmas A.2 and A.3] and Lemma 3.3, we can obtain that $\sum_{j\in S} q(j|i,a)V_\alpha^*(j)$ is continuous in a on $A(i)$. Thus, by (ii), there exists $f \in F$ realizing the minimum on the right-hand side of (3.6). By Lemma 3.3, the result follows. ■

To study monotonicity properties of the discounted cost optimal value V_α^*, we take an arbitrary, but fixed, function m on S such that $m(i) \geq q(i)$, and $m(i) > 0$ $\forall i \in S$.

Let $\tilde{q}(k|i,a) := \sum_{j\geq k}\left[\frac{q(j|i,a)}{m(i)} + \delta_{ij}\right]$, $k, i \in S$, $a \in A(i)$; here $\delta_{ij} = 0$ for $i \neq j$, $\delta_{ij} = 1$ for $i = j$, $i, j \in S$.

Lemma 3.5 *Suppose that* Assumption 2.4 *holds and let* $A(i)$ *be finite for each* $i \in S$.

(i) *Let* $u_0(i) := 0$, *and*

$$u_{n+1}(i) := \inf_{a\in A(i)}\left\{\frac{r(i,a)}{m(i)+\alpha} + \frac{m(i)}{m(i)+\alpha}\sum_{j\in S}\left[\frac{q(j|i,a)}{m(i)} + \delta_{ij}\right]\right\} u_n(j)$$

for $i \in S$ *and* $n \geq 0$. *Then* $\lim_{n\to\infty} u_n(i) = V_\alpha^*(i)$ *for any* $i \in S$, $\alpha > 0$.

(ii) *If* $A(i) \equiv A$ $\forall i \in S := \{0, 1, \cdots\}$; *and* $\frac{r(i,a)}{m(i)+\alpha}$, $m(i)$ *and* $\tilde{q}(k|i,a)$ *are increasing functions in* i *for any fixed* $k \in S$, *and* $a \in A$, *then* $V_\alpha^*(i)$ *is increasing on* S.

Proof

(i) For $u \in B(S)$, let

$$Tu(i) := \min_{a\in A(i)}\left\{\frac{r(i,a)}{m(i)+\alpha} + \frac{m(i)}{m(i)+\alpha}\sum_{j\in S}\left[\frac{q(j|i,a)}{m(i)} + \delta_{ij}\right] u(j)\right\}. \tag{3.7}$$

Then we have $u_{n+1} = Tu_n$, $u_n = T^n 0$ and $u_n \le u_{n+1}$ for $n \ge 0$. By induction, we can get, for any $n \ge 1$,

$$\begin{aligned} u_n &\le \sum_{l=1}^{k} \alpha^{-l}(w_l + \cdots + w_k)M \\ &\le \left(\frac{1}{\alpha} + \frac{1}{\alpha^2} + \cdots + \frac{1}{\alpha^k}\right) MW. \end{aligned}$$

Hence, $\lim_{n\to} u_n := u$ exists and $u \in B(S)$. By monotone convergence and noting that $A(i)$ is finite for $i \in S$, we see that u satisfies (3.6). Thus, (i) follows from Lemma 3.3(iv).

(ii) By (i), to prove (ii), it suffices to show that, for $i_1, i_2 \in S$, $i_1 \ge i_2$, and $n \ge 0$,

$$u_n(i_1) \ge u_n(i_2). \tag{3.8}$$

By induction, when $n = 0$, (3.8) is obviously valid. Suppose now that (3.8) holds for $n = N$. With the notation $u_N(-1) := 0$, for any $a \in A$, $i_1 \ge i_2$, $i_1, i_2 \in S$, we have

$$\begin{aligned} &\sum_{j=0}^{\infty}\left(\frac{q(j|i_1,a)}{m(i_1)} + \delta_{i_1 j}\right) u_N(j) \\ &= \sum_{j=0}^{\infty}\left[\frac{q(j|i_1,a)}{m(i_1)} + \delta_{i_1 j}\right]\left[\sum_{i=0}^{j}(u_N(i) - u_N(i-1))\right] \\ &= \sum_{j=0}^{\infty}(u_N(j) - u_N(j-1))\left(\sum_{i=j}^{\infty}\left[\frac{q(i|i_1,a)}{m(i_1)} + \delta_{i_1 i}\right]\right) \\ &= \sum_{j=0}^{\infty}(u_N(j) - u_N(j-1))\tilde{q}(j|i_1,a) \\ &\ge \sum_{j=0}^{\infty}(u_N(j) - u_N(j-1))\tilde{q}(j|i_2,a) \\ &= \sum_{j=0}^{\infty}\left(\frac{q(j|i_2,a)}{m(i_2)} + \delta_{i_2 j}\right) u_N(j). \end{aligned}$$

Hence,

$$\frac{r(i_1,a)}{m(i_1)+\alpha}+\frac{m(i_1)}{m(i_1)+\alpha}\sum_{j\in S}\left(\frac{q(j|i_1,a)}{m(i_1)}+\delta_{i_1j}\right)u_N(j))$$
$$\geq \frac{r(i_2,a)}{m(i_2)+\alpha}+\frac{m(i_2)}{m(i_2)+\alpha}\sum_{j\in S}\left[\frac{q(j|i_2,a)}{m(i_2)}+\delta_{i_2j}\right]u_N(j).$$

Then, $u_{N+1}(i_1)\geq u_{N+1}(i_2)$, and so (3.8) is valid for $n=N+1$. This yields (ii). ■

Corollary 3.1 *Suppose that* $\|q\|<\infty$ *and that* Assumption 2.4 *holds. In addition suppose that for any* $i\in S\equiv\{0,1,2,\cdots\}$, $A(i)=A$, $r(i,a)$ *is increasing (or decreasing), and there exists a positive constant* $C\geq\|q\|$ *such that*

$$\tilde{p_c}(k|i,a):=\sum_{j\geq k}\left(\frac{q(j|i,a)}{C}+\delta_{ij}\right)$$

is increasing in i *for any fixed* $k\in S$ *and* $a\in A$. *Then* $V_\alpha^*(i)$ *is increasing (or decreasing).*

Proof Take $m(i):=C\ \forall i\in S$. The result then follows from Lemma 3.5. ■

In Lemma 3.4 we provided conditions under which there exists a discounted cost optimal stationary policy. Now we shall investigate more detailed results concerning the structure of an optimal policy.

Lemma 3.6 *Suppose that following conditions hold:*

(i) The hypotheses of Lemma 3.4 *are satisfied.*

(ii) For any $i\in S\equiv\{0,1,2,\cdots\}$, $A(i)=A$, A *is a partially ordered set, and* $r(i,a)$ *is nondecreasing in* i, *for any* $a\in A$.

(iii) $|q|\leq C$, *for some positive constant* C *and for any fixed* $k\in S$, *and* $a\in A$, $\tilde{p_c}(k|i,a)$ *is increasing in* i.

(iv) $r(i,a)$ *is a superadditive (subadditive) function (refer to* [8, page 103], *for instance) on* K.

(v) $\tilde{p_c}(k|i,a)$ *is a superadditive (subadditive) function on* K *for all fixed* $k\in S$.

Then there exists a discounted cost optimal stationary policy f_α^* *which is nondecreasing (non-increasing) in* i *on* S.

Proof We prove the result in the superadditive case.

By condition (v) and the definition of superadditivity, for $i_1 < i_2$, $a_1 < a_2$, and all $k \in S$, we have:

$$\sum_{j=k}^{\infty}\left[\left(\frac{q(j|i_1,a_1)}{\|q\|}+\delta_{i_1j}\right)+\left(\frac{q(j|i_2,a_2)}{\|q\|}+\delta_{i_2j}\right)\right]$$
$$\geq \sum_{j=k}^{\infty}\left[\left(\frac{q(j|i_1,a_2)}{\|q\|}+\delta_{i_1j}\right)+\left(\frac{q(j|i_2,a_1)}{\|q\|}+\delta_{i_2j}\right)\right].$$

By Lemma 3.5, $V_\alpha^*(i)$ is nondecreasing in i for all α. Applying [8, Lemma 4.7.2], we have

$$\sum_{j=0}^{\infty}\left[\left(\frac{q(j|i_1,a_1)}{\|q\|}+\delta_{i_1j}\right)\left(\frac{q(j|i_2,a_2)}{\|q\|}+\delta_{i_2j}\right)V_\alpha^*(j)\right]$$
$$\geq \sum_{j=0}^{\infty}\left(\frac{q(j|i_1,a_2)}{\|q\|}+\delta_{i_1j}\right)+\left(\frac{q(j|i_2,a_1)}{\|q\|}+\delta_{i_2j}\right)V_\alpha^*(j).$$

Thus, for each $\alpha > 0$, $\sum_{j=0}^{\infty}\left(\frac{q(j|i,a)}{|q|}+\delta_{ij}\right)V_\alpha^*(j)$ is superadditive on K. On the other hand, by condition (iv), r is superadditive. Hence, since the sum of superadditive functions is superadditive, the result follows from [8, Lemma 4.7.1]. ∎

4. Average cost criterion

In this section we establish the optimality inequality (OIE for short) for the average cost criterion $\bar{V}$ and prove the existence of average cost optimal policies. Throughout this section, we assume that the conditions in Lemma 3.4 hold. So, by Lemma 3.4, we can let f_α^* be a discounted cost optimal stationary policy with respect to the discounted rate $\alpha > 0$.

By the Tychonoff Theorem we have that F is a compact metric space. Hence, each sequence $\{f_{\alpha_n}^*, n \geq 1\} \subset F$, has a convergent subsequence. This means that there is a limit point $f^* \in F$ for $\{f_{\alpha_n}^*, n \geq 1\}$.

Let $k_0 \in S$ be fixed, and for any $\alpha > 0$ and $i \in S$, define

$$u_\alpha(i) \;:=\; V_\alpha^*(i) - V_\alpha^*(k_0).$$

Assumption 4.1 *For some decreasing sequence $\{\alpha_n\}$ tending to zero and some $k_0 \in S$, there exists a nonnegative function h and a constant N such that*

(i) $N \leq u_{\alpha_n}(i) \leq h(i)$ for all $n \geq 1$ and $i \in S$;

(ii) there exists an action $a' \in A(k_0)$ satisfying

$$\sum_{j\in S} q(j|k_0, a')h(j) \; < \; \infty.$$

Theorem 4.1 *Suppose that the following conditions hold,*

(i) Assumptions 2.4 *and* 4.1 *hold; and*

(ii) for every $i \in S$, $A(i)$ is compact, and $r(i,a)$ and $\sum_{j\in S} q(j|i,a)R(j)$ are continuous in a on $A(i)$.

Then we have:

(i) there exists a constant g^*, a function u on S, and a decreasing sequence $\{\alpha_m\}$ tending to zero, such that for $i \in S$:

$$a) \qquad g^* \; = \; \lim_{m\to\infty} \alpha_m V^*_{\alpha_m}(i), \qquad u(i) \; = \; \lim_{m\to\infty} u_{\alpha_m}(i);$$

$$b) \qquad g^* \; \geq \; r\big(i, f^*(i)\big) + \sum_{j\in S} q\big(j|i, f^*(i)\big)u(j)$$

$$\geq \; \min_{a\in A(i)} \left\{ r(i,a) + \sum_{j\in S} q(j|i,a)u(j) \right\}. \tag{4.1}$$

$$c) \qquad N \; \leq \; u \leq h. \tag{4.2}$$

(ii) f^* is an average cost optimal stationary policy and satisfies

$$\bar{V}(f^{*\infty}, i) \; = \; g^* \;\forall i \in S.$$

(iii) Any policy $f \in F$ realizing the minimum of the right-hand side of (4.1) is average cost optimal.

Proof

(i) For any $n \geq 1$, $i \in S$ and $a \in A(i)$, by Lemma 3.3 we have

$$\alpha_n V^*_{\alpha_n}(i) \; = \; r\big(i, f^*_{\alpha_n}(i)\big) + \sum_{j\in S} q\big(j|i, f^*_{\alpha_n}(i)\big) V^*_{\alpha_n}(j).$$

Hence,

$$\alpha_n V^*_{\alpha_n}(k_0) + \alpha_n u_{\alpha_n}(i) \; = \; r\left(i, f^*_{\alpha_n}(i)\right) + \sum_{j\in S} q\left(j|i, f^*_{\alpha_n}(i)\right) u_{\alpha_n}(j). \tag{4.3}$$

Since f^* is a limit point of $\{f^*_{\alpha_n}\}$, there exists a subsequence $\{\alpha_{n'}\}$ of $\{\alpha_n\}$ such that $\lim_{n'\to\infty} f^*_{\alpha_{n'}}(i) = f^*(i)$ for all $i \in S$. By Assumptions 4.1(ii) and (4.4), we have, for any $n' > 0$,

$$\begin{aligned} 0 &\le \left|\alpha_{n'} V^*_{\alpha_{n'}}(k_0)\right| \\ &\le \left|r(k_0, a')\right| + \sum_{j\in S} \left|q(j|k_0, a')\right| \left(h(j) + |N|\right) \\ &= r\left(k_0, a'\right) + \sum_{j\in S} q\left(j|k_0, a'\right) h(j) \\ &\qquad -2q\left(k_0|k_0, a'\right)\left(h(k_0) + |N|\right) \\ &< \infty. \end{aligned}$$

Hence, there exists a subsequence $\{\alpha_{m'}\}$ of $\{\alpha_{n'}\}$ such that

$$\lim_{m'\to\infty} \alpha_{m'} V_{\alpha_{m'}}(k_0) \doteq g^*.$$

By Assumption 4.1(i) and the Tychonoff Theorem, we have that $\{u_{\alpha_{m'}}\}$ is a sequence of the compact metric space $\prod_{i\in S}[N, h(i)]$. Thus, there exists a subsequence $\{\alpha_m\}$ of $\{\alpha_{m'}\}$ such that $\lim_{m\to\infty} u_{\alpha_m}(i) := u(i)\ \forall i \in S$. By Assumption 4.1, we have $\lim_{m\to\infty} \alpha_m = 0$; hence $\lim_{m\to\infty} \alpha_m u_{\alpha_m}(i) = 0$ for $i \in S$. We can then obtain that

$$\lim_{m\to\infty} \alpha_m V^*_{\alpha_m}(i) = \lim_{m\to\infty} \alpha_m V^*_{\alpha_m}(k_0) = g^*, \qquad \forall i \in S.$$

Hence, the conclusions (i)(a) and (i)(c) are valid.

To prove conclusion (i)(b), from (4.4), for any $m \ge 1$, we have

$$\begin{aligned} &\frac{\alpha_m V^*_{\alpha_m}(k_0)}{m(i)} + \frac{\alpha_m u_{\alpha_m}(i)}{m(i)} + u_{\alpha_m}(i) \\ &= \frac{r\left(i, f^*_{\alpha_m}(i)\right)}{m(i)} + \sum_{j\in S}\left[\frac{q\left(j|i, f^*_{\alpha_m}(i)\right)}{m(i)} + \delta_{ij}\right] u_{\alpha_m}(j). \end{aligned} \tag{4.4}$$

Since $\lim_{m\to\infty} f^*_{\alpha_m}(i) = f^*(i)$, we have that

$$\lim_{m\to\infty} r\left(i, f^*_{\alpha_m}(i)\right) = r\left(i, f^*(i)\right)$$

and

$$\lim_{m\to\infty} \sum_{j\in S} q\left(j|i, f^*_{\alpha_m}(i)\right) u(j) = \sum_{j\in S} q\left(j|i, f^*(i)\right) u(j).$$

By (4.4) and Fatou's Lemma, for $i \in S$, we have

$$\frac{g^*}{m(i)} + u(i) \geq \frac{r(i, f^*(i))}{m(i)} + \sum_{j \in S} \left[\frac{q(j|i, f^*(i))}{m(i)} + \delta_{ij} \right] u(j). \quad (4.5)$$

Hence,

$$\begin{aligned} g^* &\geq r(i, f^*(i)) + \sum_{j \in S} q(j|i, f^*(i))\, u(j) \\ &\geq \min_{a \in A(i)} \left\{ r(i,a) + \sum_{j \in S} q(j|i,a) u(j) \right\}. \end{aligned} \quad (4.6)$$

This means that (i)(b) is valid.

(ii) To prove (ii), from (4.6), there is a nonnegative function $c(f^*)$ on S such that

$$g^* = r(i, f^*(i)) + c(f^*)(i) + \sum_{j \in S} q(j|i, f^*(i))\, u(j). \quad (4.7)$$

By (4.2), (4.7) and Bather [1, Theorem 2.1.3], we have

$$\limsup_{T \to} \frac{1}{T} \int_0^T \sum_{j \in S} p_{ij}(t, f^*)(r(j, f^*(j)) + c(f^*)(j))\, dt \leq g^*, \quad i \in S. \quad (4.8)$$

Noting that $c(f^*) \geq 0$, from (2.7) and (4.8), we have

$$\bar{V}(f^*, i) \leq g^*, \qquad i \in S. \quad (4.9)$$

On the other hand, by a Tauberian Theorem [25, pp. 181–182], for $\pi \in \Pi_m$ and $i \in S$, we have

$$\begin{aligned} g^* &= \lim_{m \to \infty} \alpha_m V^*_{\alpha_m}(i) \\ &\leq \lim_{m \to \infty} \alpha_m V_{\alpha_m}(\pi, i) \\ &= \lim_{\alpha_m \searrow 0} \alpha_m \int_0^\infty e^{-\alpha_m t} \left(\sum_{j \in S} p_{ij}(t, \pi) r(t, j, \pi) \right) dt \\ &\leq \limsup_{T \to \infty} \frac{1}{T} \int_0^T \left(\sum_{j \in S} p_{ij}(t, \pi) r(t, j, \pi) \right) dt \\ &= \bar{V}(\pi, i). \end{aligned} \quad (4.10)$$

From (4.9) and (4.10), we have $\bar{V}(f^{*\infty}, i) = g^* \leq \bar{V}(\pi, i)$, for all $\pi \in \Pi_m$ and $i \in S$. The proof of (ii) is complete.

(iii) Similarly, we can prove (iii). ■

Theorem 4.2 *Suppose that the following conditions hold:*

(i) The hypotheses of Lemma 3.4 *and* Assumptions 4.1 *hold.*

(ii) For any $i \in S \equiv \{0, 1, 2, \cdots\}$, $A(i) = A$, A is a partially ordered set, and $r(i, a)$ is nondecreasing in i for any $a \in A$.

(iii) $\|q\| \leq C$, for some positive constant C, and for each fixed $k \in S$, $\tilde{p_c}(k|i, a)$ is increasing in i.

(iv) $r(i, a)$ is a superadditive (subadditive) function on K.

(v) $\tilde{p_c}(k|i, a)$ is a superadditive (subadditive) function on K, for all fixed $k \in S$.

Then there exists an average cost optimal stationary policy which is increasing (or decreasing) in state i.

Proof By Lemma 3.6, we have that f_α^* is increasing (or decreasing) on S for any $\alpha > 0$. Hence, the limit point f^* is also increasing (or decreasing) on S. By Theorem 4.1, f^* is also average cost optimal. The proof is complete. ■

5. Examples

In this section we provide two examples in which our Assumptions 2.4 and 4.1 hold, whereas the conditions in [5, 11, 12, 14] fail to hold.

Example 1

We observe continuously an admission control model for queueing system $M^X/M/1$. Let p_k, $k = 0, 1, 2, \ldots$, denote the arrival probability of k customers, and such that $\sum_{k=0}^{\infty} p_k = 1$ and $\sum_{k=0}^{\infty} kp_k < \infty$. The arrival rate of the system is λ. Let μ denote the exponential service rate of the system. At any arrival time, the controller decides whether to admit or reject all arriving customers. Rejected tasks are lost. Each accepted task generates a reward c. A nondecreasing function $r(i)$ denotes the cost rate for serving i customers. Let $p > 0$ denote the cost rate of serving a customer. Hence, we have $r(i) = pi$.

We formulate this model as a continuous-time Markov decision process. The system state i denotes the number of customers available for service in the system at any time (i.e., the queue length). So,

$S = \{0, 1, \ldots, i, \ldots\}$. For each $i \in S$, $A(i) = \{0, 1\}$, with action 0 corresponding to rejecting and action 1 corresponding to accepting all arriving customers. The cost rate function r satisfies $r(i, 0) = r(i)$, $r(i, 1) = r(i) - c\lambda$. By Remarks 2.2 and 3.1, the cost function $r(i, a)$ may be increased by adding a constant without affecting the discussion of average optimality. So, we may take that $r(i, 0) = r(i) + c\lambda$, and $r(i, 1) = r(i)$. The transition rate q satisfies:

$q(0|0, 0) = 0$ and for $i \geq 1$,

$$q(j|i,0) = \begin{cases} \mu & \text{if } j = i-1, \\ -\mu & \text{if } j = i, \\ 0 & \text{otherwise;} \end{cases}$$

$q(0|0, 1) = -\left(\sum_{k=1}^{\infty} \lambda p_k\right) = -\lambda(1 - p_0)$, $q(k|0, 1) = \lambda p_k \ \forall k \geq 1$, and for $i \geq 1$,

$$q(j|i,1) = \begin{cases} \mu & \text{if } j = i-1, \\ -(\mu + \lambda(1-p_0)) & \text{if } j = i, \\ \lambda p_k & \text{if } j = i+k, \\ 0 & \text{otherwise.} \end{cases}$$

For this model, we can derive that:

1. Assumption 2.4 holds. In fact, for all $i \in S$, we let $w_1(i) := pi + 1 + c\lambda$, and $w_2(i) := p\lambda\left(\sum_{k=1}^{\infty} kp_k\right)$, $i \in S$. By Lemma 3.2(v), we can then verify Assumption 2.4.

2. Assumption 4.1 holds. In fact, by Lemma 3.3, we have, for any $\alpha > 0$,

$$\begin{aligned} \alpha V_\alpha^*(i) &= \min_{a \in A(i)} \left\{ r(i,a) + \sum_{j \in S} q(j|i,a) V_\alpha^*(j) \right\} \\ &\leq r(i,0) + \mu V_\alpha^*(i-1) - \mu V_\alpha^*(i), \qquad i > 0. \\ \alpha V_\alpha^*(0) &\leq r(0,0) \\ &= c\lambda. \end{aligned}$$

So we have, for $i > 0$,

$$\begin{aligned} V_\alpha^*(i) &\leq \frac{r(i,0)}{\mu} + V_\alpha^*(i-1) \\ &\leq \frac{r(i,0)}{\mu} + \frac{r(i-1,0)}{\mu} + V_\alpha^*(i-2) \\ &\leq \cdots\cdots\cdots\cdots \\ &\leq \sum_{s=1}^{i} \frac{r(s,0)}{\mu} + V_\alpha^*(0) \\ &= \frac{pi(i+1)}{2\mu} + \frac{ic\lambda}{\mu} + V_\alpha^*(0). \end{aligned}$$

Hence, we take $k_0 = 0$ to get that $u_\alpha(i) \leq \frac{pi(i+1)}{2\mu} + \frac{ic\lambda}{\mu} := h(i)$, for all $i \in S$ and $\alpha > 0$. By Lemma 3.5, we see that $V_\alpha^*(i)$ is an increasing function on S. Thus, we have $u_\alpha(i) \geq 0$ for all $i \in S$ and $\alpha > 0$, and $\sum_{j \in S} q(j|0,0)h(j) = 0 < \infty$. Thus, Assumptions 4.1(i) and 4.1(ii) hold.

Hence, by Theorem 4.1, we have the following conclusion. For this admission control queue model, there exists an average cost optimal stationary policy.

Remark 5.1 *In* Example 1, *the cost rate is obviously unbounded. Hence, the assumption of bounds of the reward rate in* [7, 11, 12] *fails to hold. If we take $p_k > 0$ for all $k \geq 0$, then we can verify that the conditions in* [14] *fail to hold.*

We give next another example in which Assumptions 2.4 and 4.1 hold. Moreover, both the cost and the transition rates are unbounded.

Example 2

We consider an admission control birth and death process as follows: Let $S = \{0, 1, 2, \ldots\}$, $A(i) \equiv \{0, 1\}$, $i \in S$:

$q(0|0,0) = 0$, and for $i \geq 1$,

$$q(j|i,0) = \begin{cases} \mu i & \text{if } j = i-1, \\ -\mu i & \text{if } j = i, \\ 0 & \text{otherwise;} \end{cases}$$

$q(0|0,1) = -v$, $q(1|0,1) = v$, and for $i \geq 1$,

$$q(j|i,1) = \begin{cases} \mu i & \text{if } j = i-1, \\ -(\mu+\lambda)i - v & \text{if } j = i, \\ \lambda i + v & \text{if } j = i+1, \\ 0 & \text{otherwise;} \end{cases}$$

$\mu \geq \lambda \geq 0$. $r(i,0) = a_1 i^2 + a_2 i + c$, $r(i,1) = b_1 i^2 + b_2 i$, $i \in S$, $(c - 2a_1)(\lambda + \mu) \leq v(3a_1 + a_2)$, $c, a_1, a_2, b_1, b_2 \geq 0$.

Obviously, both the cost and the transition rates in this model are unbounded. On the other hand, we can obtain the following.

1. Assumption 2.4 holds. In fact, we verify this conclusion as follows: Let $w_1(i) := (a_1 + b_1)i^2$ for all $i \in S$, then

$$\sum_{j \in S} q(j|i,0)w_1(j) \leq 0, \qquad \forall i \in S;$$

and,

for $i = 0$:

$$\begin{aligned}\sum_{j \in S} q(j|0,1)w_1(j) &= v(a_1 + b_1) \\ &\leq (a_1 + b_1)(\mu + \lambda + 3v) + (a_2 + b_2);\end{aligned}$$

for $i \geq 1$:

$$\begin{aligned}&\sum_{j \in S} q(j|i,1)w_1(j) \\ &= \mu i w_1(i-1) - \mu i w_1(i) - \lambda i w_1(i) + \lambda i w_1(i+1) \\ &\quad + v w_1(i+1) - v w_1(i) \\ &= 2(a_1 + b_1)(-\mu + \lambda)i^2 + (a_1 + b_1)(\mu + \lambda)i \\ &\quad + v(a_1 + b_1)(2i + 1) \\ &\leq \big((a_1 + b_1)(\mu + \lambda + 3v) + (a_2 + b_2)\big)(i + 1).\end{aligned}$$

Let

$$w_2(i) := ((a_1 + b_1)(\mu + \lambda + 3v) + (a_2 + b_2))(i + 1), \quad i \in S.$$

Then we have

$$\sum_{j \in S} q(j|i,a)w_1(j) \leq w_2(i), \quad \forall i \in S, \ a \in \{0,1\}. \tag{5.1}$$

Moreover,

$$\sum_{j \in S} q(j|i,0)w_2(j) \leq 0, \quad \forall i \in S;$$

and,

for $i = 0$:

$$\sum_{j\in S} q(j|0,1)w_2(j) \quad = \quad v\big((a_1+b_1)(\mu+\lambda+3v)+(a_2+b_2)\big);$$

for $i \geq 1$:

$$\begin{aligned} &\sum_{j\in S} q(j|i,1)w_2(j) \\ &\quad = \mu i w_2(i-1) - \mu i w_2(i) - \lambda i w_2(i) + \lambda i w_2(i+1) \\ &\qquad + v w_2(i+1) - v w_2(i) \\ &\quad = \big((a_1+b_1)(\mu+\lambda+3v)+(a_2+b_2)\big)(-\mu i + \lambda i) \\ &\qquad + v\big((a_1+b_1)(\mu+\lambda+3v)+(a_2+b_2)\big) \\ &\quad \leq v\big((a_1+b_1)(\mu+\lambda+3v)+(a_2+b_2)\big). \end{aligned} \tag{5.2}$$

Let $w_3(i) := v\big((a_1+b_1)(\mu+\lambda+3v)+(a_2+b_2)\big)+c+1$ for all $i \in S$. Then we have

$$\sum_{j\in S} q(j|i,a)w_2(j) \quad \leq \quad w_3(i), \quad \forall i \in S,\ a \in \{0,1\}, \tag{5.3}$$

and

$$\sum_{j\in S} q(j|i,a)w_3(j) \quad \leq \quad 0, \quad i \in S,\ a \in \{0,1\}. \tag{5.4}$$

Hence, from (5.1),(5.3),(5.4), we get that Assumption 2.1 holds.

By a similar argument, we can obtain that the function qR on S satisfies Assumption 2.4(Assumption 2.1), where $R := w_1+w_2+w_3$. By Lemma 3.2, we can also get Assumption 2.2. Obviously $R \geq 1$ and $r \leq R$. Combining these facts we conclude that Assumption 2.4 holds.

2. Now, let $m(0) := 2(\lambda+\mu+v)$, and $m(i) = 2((\lambda+\mu+)i+v)$ for $i \geq 1$.

   $m(i)$, $\frac{r(i,a)}{m(i)+\alpha}$ and $\tilde{q}(k|i,a)$ all are increasing in i for any $\alpha > 0$ and $a \in \{0,1\}$. In fact, m is obviously an increasing function on S. Likewise, as $(c-2a_1)(\lambda+\mu) \leq v(3a_1+a_2)$, we can verify that $\frac{r(i,a)}{m(i)+\alpha}$ is increasing on S for any $\alpha > 0$ and $a \in \{0,1\}$. By definitions of $\tilde{q}$, we can also verify that $\tilde{q}(k|i,a)$ is increasing on S for any $\alpha > 0$, $a \in \{0,1\}$ and $k \in S$.

3. $u_\alpha(i) \geq 0$, with $k_0 = 0$. In fact, by Lemma 3.5 and conclusion (2), we have that $V_\alpha^*(i)$ is increasing in i. Hence, $u_\alpha(i) \geq 0$.

4. $u_\alpha(i) \leq h(i)\ \forall i \in S,\ \alpha > 0$, and $\sum_{j \in S} q(j|0,0)h(j) = 0 < \infty$, where $h(i) := \frac{(a_2+c)i}{\mu} + \frac{a_1 i(i+1)}{2\mu}$ for $i \in S$. In fact, by Lemma 3.3, we have $0 \leq \alpha V_\alpha^*(0) \leq r(0,0) = c$, and for $i \geq 1$,

$$\alpha V_\alpha^*(i) \leq r(i,0) + i\mu V_\alpha^*(i-1) - i\mu V_\alpha^*(i).$$

Hence,

$$\begin{aligned} V_\alpha^*(i) &\leq \frac{r(i,0)}{i\mu} + V_\alpha^*(i-1) \\ &\leq \frac{ia_1}{\mu} + \frac{a_2+c}{\mu} + V_\alpha^*(i-1) \\ &\leq \cdots\cdots\cdots \\ &\leq \frac{i(a_2+c)}{\mu} + \frac{a_1 i(i+1)}{2\mu} + V_\alpha^*(0). \\ u_\alpha(i) &= V_\alpha^*(i) - V_\alpha^*(0) \\ &\leq \frac{i(a_2+c)}{\mu} + \frac{a_1 i(i+1)}{2\mu} \\ &\equiv h(i). \end{aligned}$$

5. Assumption 4.1 holds. This follows from the conclusions (3) and (4).

Hence, by conclusions (1) to (5) and Theorem 4.1, we have the following. For the given controlled birth and death process, there exists an average cost optimal stationary policy.

Remark 5.2 *Obviously, in* Example 2, *the conditions in* [1]-[13] *are not satisfied, but our assumptions do hold.*

References

[1] Bather, J. Optimal stationary policies for denumerable Markov chains in continuous time. Adv, Appl. Prob., 8(1976), 144-158.

[2] Bertsekas, D. P. Dynamic Programming: Deterministic and Stochastic Models. Prentice-Hall, Englewood Cliffs, NJ, 1987.

[3] Doshi, B. T. Continuous time control of Markov processes on an arbitrary state space: average return criterion. Stochastic Processes and their Applications, 4(1976), 55-77.

[4] Howard, R. A. Dynamic Programming and Markov Processes. Wiley, New York, 1960.

[5] Kakumanu, P. Nondiscounted continuous time Markov decision processes with countable state space. SIAM J. Control, 10(1972), 210-220.

[6] Lippman, S. A. Applying a new device in the optimization of exponential queueing systems. Op. Res. 23(1975b), 687-710.

[7] Miller, R. L. Finite state continuous time Markov decision processes with an infinite planning horizon. J. Math. Anal. Appl., 22(1968), 552-569.

[8] Puterman, M. L. Markov Decision Processes. John Wiley & Sons Inc,1994.

[9] Yushkevich, A. A. and Feinberg, E. A. On homogeneous Markov model with continuous time and finite or countable state space. Theory. Prob. Appl., 24(1979), 156-161.

[10] Walrand, J. An Introduction to Queuing Networks. Prentice-Hall, Englewood Cliffs, NJ, 1988.

[11] Dong, Z. Q. Continuous time Markov decision programming with average reward criterion-countable state and action space. Sci. Chi.. SP ISS(II)(1979), 131-148.

[12] Song, J. S. Continuous time Markov decision programming with non-uniformly bounded transition rate. Scientia Sinica, 12(1987), 1258-1267.

[13] Zheng, S. H. Continuous time Markov decision programming with average reward criterion and unbounded reward rate. Acta Math. Appl. Sinica, 7(1991), 6-16.

[14] Guo, X. P. and Liu K. "Optimality inequality for continuous time Markov decision processes with average reward criterion". Preprint, Zhongshan University, Guangzhou, P.R. China, 1998(Submitted).

[15] Sennott, L. I. Another set of conditions for average optimality in Markov decision processes. System & control letters, 24(1995), 147-151.

[16] Sennott, L. I. Average cost optimal stationary policies in infinite state Markov decision processes with unbounded cost. Op. Res.,37(1989), 626-633.

[17] Cavazos-Cadena, R. and Gaucherand, Value iteration in a class of average controlled Markov chains with unbounded costs: Necessary and sufficient conditions for pointwise convergence, J. Appl. Prob., 33(1996), 986-1002.

[18] Guo, X. P. and Zhu, W. P. "Denumerable state continuous time Markov decision processes with unbounded cost and transition rates

under discounted cost criterion", To appear on Journal of Australian Math. Soci. series B.

[19] Wu, C. B. Continuous time Markov decision processes with unbounded reward and non-uniformly bounded transition rate under discounted criterion. Acta Math. Appl. Sinica, 20(1997), 196-208.

[20] Filar, J. A. and Vrieze, K. Competitive Markov Decision Processes. Springer-Verlag, New York, 1996.

[21] Chung, K. L. Markov Chains with Stationary Transition Probabilities. Springer-Verlag, Berlin, 1960.

[22] Haviv, M. and Puterman, M. L. Bias optimality in controlled queueing systems. J. Appl. Prob.., 35(1998), 136-150.

[23] Serfozo, R. Optimal control of random walks, birth and death processes, and queues. Adv. Appl. Prob.., 13(1981), 61-83.

[24] Anderson, W. J. Continuous Time Markov Chains. Springer-Verlag, New York, 1991.

[25] Widder, D.V. The Laplace Transform. Princeton University Press. Princeton, NJ, 1946.

Chapter 11

OPTIMAL AND NEARLY OPTIMAL POLICIES IN MARKOV DECISION CHAINS WITH NONNEGATIVE REWARDS AND RISK-SENSITIVE EXPECTED TOTAL-REWARD CRITERION*

Rolando Cavazos–Cadena
Departamento de Estadística y Cálculo,
Universidad Autónoma Agraria Antonio Narro,
Buenavista, Saltillo COAH. 25315, México
rcavazos@narro.uaaan.mx

Raúl Montes–de–Oca
Departamento de Matemáticas,
Universidad Autónoma Metropolitana,
Campus Iztapalapa,
Av. Michoacán y la Purísima s/n,
Col. Vicentina, México D.F. 09340, México
momr@xanum.uam.mx

Abstract This work considers Markov decision processes with discrete state space. Assuming that the decision maker has a non-null constant risk-sensitivity, which leads to grade random rewards via the expectation of an exponential utility function, the performance index of a control policy is the risk-sensitive expected total-reward criterion corresponding to a nonnegative reward function. Within this framework, the existence of

*The work of R. Cavazos-Cadena was supported in part by the PSF Organization under Grant No. 30-250–99–04.

Z. Hou et al. (eds.), Markov Processes and Controlled Markov Chains, 189–221.

optimal and approximately optimal stationary policies in the absolute sense is studied. The main results can be summarised as follows:
(i) An optimal stationary policy exists if the state and actions sets are finite, whereas an ε-optimal stationary policy is guaranteed when just the state space is finite.
(ii) This latter fact is used to obtain, for the general denumerable state space case, that ε-optimal stationary policies exist if the controller is risk-seeking and the optimal value function is bounded.
In contrast with the usual approach, the analysis performed in the paper does not involve the discounted criterion, and is completely based on properties of optimal value function, particularly, on the the strong optimality equation.

Keywords: Utility function,constant risk-sensitivity, Ornstein's theorem, strong optimality equation, risk-seeking controller.

1. Introduction

This note concerns Markov decision processes (MDPs) with discrete state space and nonnegative rewards. The fundamental assumption is that the attitude of the controller before a random reward is characterised by a *constant risk-sensitivity* coefficient $\lambda \neq 0$, which is associated to an exponential utility function (Pratt [13], Fishburn [7]). The performance of a control policy is measured by the corresponding *risk-sensitive expected total-reward criterion* introduced in Section 2, and the paper analyses the existence of optimal and ε-optimal stationary policies, i.e., stationary policies whose performance index differs from the optimal value by less than $\varepsilon > 0$.

Recently, there has been a great interest in controlled stochastic processes endowed with a risk-sensitive criterion; see, for instance, Flemming and Hernández–Hernández [8], Brau–Rojas [2], Ávila–Godoy [1], Cavazos–Cadena and Montes–de–Oca [4], as well as the references therein; the first two works deal with the risk-sensitive average index which, under a strong simultaneous Doeblin condition (Thomas [17]), was studied in Cavazos–Cadena and Fernández–Gaucherand [3] via the total-reward criterion considered in this article. Among other topics, the work by Ávila–Godoy generalizes results in risk-neutral negative dynamic programming to the risk-sensitive context; for instance, she has shown that if the reward function is non-positive (the negative dynamic programming framework), then a stationary policy obtained by maximising the right-hand side of the optimality equation is risk-sensitive optimal, extending a classical theorem by Strauch [16]. However, even under strong continuity-compactness conditions, this result does not hold for nonnegative rewards (Cavazos–Cadena and Montes–de–Oca [4]) so that, as in risk-neutral positive dynamic programming, searching for ε-optimal stationary policies is also an interesting problem within a risk-sensitive framework.

For MDPs with nonnegative rewards, the existence of optimal or ε-optimal stationary policies has been widely studied in the literature on the risk-neutral expected total-reward criterion. A key fact in the analysis of this problem is that if the state and action spaces are finite, then an optimal stationary policy exists, a result that is usually obtained via the *discounted* criterion (Puterman [14]). Using this result, the problem for MDPs with more general state space is approached by constructing approximations that allow one to obtain ε-optimal stationary policies (Ornstein [12], Hordijk [11, Chapter 13]). These ideas, based on the discounted criterion, were recently extended in Cavazos–Cadena and Montes–de–Oca [5], where for the risk-sensitive expected total-reward criterion and nonnegative rewards, the existence of optimal stationary policies was established for finite models, whereas, for general denumerable state space, the ε-optimality results were obtained whenever the optimal value function is bounded and the controller is risk-averse, i.e., when $\lambda < 0$.

This work has two *main objectives*: The first one is to establish the existence of risk-sensitive ε-optimal stationary policies for MDPs with denumerable state space when the controller is risk-seeking, a feature that corresponds to a positive risk-sensitivity coefficient. The result on this direction is stated below as Theorem 7.1, and is obtained under the assumption that the optimal value function is bounded. The *second goal* refers to the approach used establish Theorem7.1, which is based on the existence of optimal stationary policies for the finite state space case; the idea is to obtain this latter result focusing on the properties of the risk-sensitive expected total-reward criterion. The corresponding results, extending the analysis in Cavazos–Cadena and Montes–de–Oca [6] for risk-neutral dynamic programming, are contained in Sections 3–5. The approach is entirely based on the (usual) optimality equation, as well as on its strong version (see Lemmas 2.1 and 2.2). The idea behind this part of the work is to gain a better understanding of the properties of the risk-sensitive expected total-reward index.

The *organization* of the paper is as follows. In Section 2 the decision model is introduced, and the basic facts concerning the risk-sensitive optimality equations are stated. The effect of modifying a stationary policy at a single state is analysed in Section 3, whereas in Section 4 the optimal value functions of different MDPs are compared. These tools are used in Sections 5 and 6 to prove the existence of optimal and ε-optimal stationary policies for MDPs with finite state space, and this result is the basic ingredient to obtain ε-optimal stationary policies for models with general denumerable state space in Section 7. Finally, the paper concludes in Section 8 with some brief comments.

Notation Throughout the article, $\mathbb{R}$ and $\mathbb{N}$ stand for the set of real numbers and nonnegative integers, respectively. Given a function $C\colon S \to \mathbb{R}$, the corresponding supremum norm is denoted by

$$\|C\| \; := \; \sup_{w \in S} |C(w)|.$$

Finally, if W is an event, then $I[W]$ stands for the associated indicator function.

2. Decision model

Let $M = (S, A, \{A(x)\}, R, P)$ be the usual MDP, where the state space S is a (nonempty and) *denumerable* set endowed with the discrete topology, the metric space A is the control (or action) set, and for each $x \in S$, $\emptyset \neq A(x) \subset A$ is the *measurable* subset of admissible actions at state x. On the other hand, $R\colon \mathbb{K} \to \mathbb{R}$ is the reward function, where $\mathbb{K}\colon = \{(x, a) \,|\, a \in A(x), x \in S\}$ is the set of *admissible pairs*, and $P = [p_{x\,y}(\cdot)]$ is the controlled transition law. This model M has the following interpretation. At each time $t \in \mathbb{N}$ the state of a dynamical system is observed, say $X_t = x \in S$, and an action $A_t = a \in A(x)$ is chosen. As a consequence, a reward $R(x, a)$ is earned and, regardless of which states and actions were observed and applied up to time t, the state of the system at time $t + 1$ will be $X_{t+1} = y \in S$ with probability $p_{x\,y}(a)$, description that corresponds to the Markov property of the decision model.

Assumption 2.1 *For every $x, y \in S$,*

(i) $a \mapsto R(x, a)$ and $a \mapsto p_{x\,y}(a)$ are measurable functions on $A(x)$, and

(ii) the reward function is nonnegative: $R(x, a) \geq 0, \quad (x, a) \in \mathbb{K}$.

Utility function Given a real number λ, hereafter referred to as the (constant) *risk-sensitivity coefficient*, the corresponding utility function $U_\lambda\colon \mathbb{R} \to \mathbb{R}$ is determined as follows. For $x \in \mathbb{R}$,

$$U_\lambda(x)\colon \;=\; \begin{cases} \operatorname{sign}(\lambda) e^{\lambda x}, & \text{if} \quad \lambda \neq 0, \\ x, & \text{when} \quad \lambda = 0; \end{cases} \tag{2.1}$$

it is not difficult to verify that $U_\lambda(\cdot)$ is always a *strictly increasing* function which satisfies the basic relation,

$$U_\lambda(c + x) \;=\; e^{\lambda c} U_\lambda(x), \quad \lambda \neq 0, \; x, c \in \mathbb{R}. \tag{2.2}$$

It is assumed that the controller grades a random reward Y via the expectation of $U_\lambda(Y)$, in the following sense: if two decision strategies δ_1 and δ_2 lead to obtaining random rewards Y_1 and Y_2, respectively, δ_1 will be preferred if $E[U_\lambda(Y_1)] > E[U_\lambda(Y_2)]$, whereas the decision maker will be indifferent between δ_1 and δ_2 when $E[U_\lambda(Y_1)] = E[U_\lambda(Y_2)]$. Let Y be a given a random reward for which $U_\lambda(Y)$ has a well defined expectation, conditional that it is always valid when $\lambda \neq 0$. In this case, the *certain equivalent of* Y *with respect to* $U_\lambda(\cdot)$ is denoted by $E(\lambda, Y)$ and is implicitly determined by the

$$U_\lambda(E(\lambda,Y)) \;=\; E[U_\lambda(Y)] \tag{2.3}$$

equality that, via (2.1), leads to the explicit formula

$$E(\lambda,Y) \;=\; \begin{cases} \frac{1}{\lambda}\log\left(E\left[e^{\lambda Y}\right]\right), & \lambda \neq 0 \\ E[Y] & \lambda = 0, \end{cases} \tag{2.4}$$

where the usual conventions $\log(\infty) = \infty$ and $\log(0) = -\infty$ are enforced. Thus, for an observer with risk sensitivity λ, the opportunity of getting the random reward Y can be fairly interchanged by the certain amount $E(\lambda, Y)$. Suppose now that Y is a nonconstant random variable. When $\lambda > 0$ (resp. $\lambda < 0$) the utility function $U_\lambda(\cdot)$ in (2.1) is convex (resp. concave), and Jensen's inequality yields that $E(\lambda, Y) > E[Y]$ (resp. $E(\lambda, Y) < E[Y]$). A decision maker grading a random reward Y according to the certain equivalent $E(\lambda, Y)$ is referred to as risk–seeking if $\lambda > 0$, and risk–averse if $\lambda < 0$. If $\lambda = 0$, the controller is risk–neutral.

Remark 2.1 *The following simple properties of the certain equivalent* $E(\lambda, Y)$ *will be useful, Cavazos–Cadena and Fernández–Gaucherand [3]:*

(i) If $P[Y = c] = 1$ *for some* $c \in \mathbb{R}$, *then* $E(\lambda, Y) = c$.

(ii) Let Y *and* W *be two random variables satisfying* $P[Y \geq W] = 1$. *Since* $U_\lambda(\cdot)$ *is increasing, it follows that*

$$U_\lambda(E(\lambda,Y)) \;=\; E[U_\lambda(Y)] \;\geq\; E[U_\lambda(W)] \;=\; U_\lambda(E(\lambda,W)),$$

and then $E(\lambda, Y) \geq E(\lambda, W)$.

(iii) In particular, if $P[Y \geq 0] = 1$ *then* $E(\lambda, Y) \geq 0$.

Policies For each $t \in \mathbb{N}$, the space $\mathbb{H}_t$ of admissible histories up to time t is recursively defined by

$$\mathbb{H}_0 : = S, \qquad \mathbb{H}_t \;=\; \mathbb{K} \times \mathbb{H}_{t-1}, \quad t \geq 1, \text{ and}$$

a generic element of $\mathbb{H}_t$ is denoted by $h_t = (x_0, a_0, \ldots, x_{t-1}, a_{t-1}, x_t)$. An admissible *control policy* for model M is a special sequence $\pi = \{\pi_t\}$ of stochastic kernels: For each $t \in \mathbb{N}$ and $h_t \in \mathbb{H}_t$, $\pi_t(\cdot|h_t)$ is a probability measure on the Borel subsets of A satisfying $\pi_t(A(x_t)|h_t) = 1$, and for each Borel subset B of the space A, $h_t \mapsto \pi_t(B|h_t)$ is a measurable mapping on $\mathbb{H}_t$; the class of all policies is denoted by $\mathcal{P}$. Given the initial state $X_0 = x$ and the policy $\pi \in \mathcal{P}$ being used to drive the system, under Assumption 2.1(i) the distribution of the state-action process $\{(X_t, A_t)\}$ is uniquely determined via the Ionescu Tulcea's theorem (see, for instance, Hinderer [10], Ross [15], Hernández–Lerma [9], or Puterman [14]). Such a distribution is denoted by $P_x^\pi[\,\cdot\,]$, whereas $E_x^\pi[\,\cdot\,]$ stands for the corresponding expectation operator. Define $\mathbb{F}: = \prod_{x \in S} A(x)$, so that $\mathbb{F}$ consists of all (choice) functions $f: S \to A$ satisfying $f(x) \in A(x)$ for all $x \in S$. A policy π is stationary if there exists $f \in \mathbb{F}$ such that for each $t \in \mathbb{N}$ and $h_t \in \mathbb{H}_t$, the probability measure $\pi_t(\cdot|h_t)$ is concentrated on $\{f(x_t)\}$. The class of stationary policies is naturally identified with $\mathbb{F}$, and with this convention $\mathbb{F} \subset \mathcal{P}$.

Performance index Given $\lambda \in \mathbb{R}$, *the λ-sensitive expected-total reward at state $x \in S$ under policy $\pi \in \mathcal{P}$* is defined by

$$V_\lambda(\pi, x): = \begin{cases} \frac{1}{\lambda} \log \left(E_x^\pi \left[e^{\lambda \sum_{i=0}^\infty R(X_t, A_t)} \right] \right), & \text{if } \lambda \neq 0 \\ E_x^\pi \left[\sum_{i=0}^\infty R(X_t, A_t) \right], & \text{if } \lambda = 0. \end{cases} \tag{2.5}$$

Thus, when the system is driven by policy π starting at x, $V_\lambda(\pi, x)$ is the certain equivalent (with respect to U_λ) of the total reward $\sum_{t=0}^\infty R(X_t, A_t)$. Observe that the nonnegativity of the reward function implies that $V_\lambda(\pi, x) \geq 0$ always hold (see Remark 2.1(iii)). The *λ-optimal value function* is given by

$$V_\lambda^*(x): = \sup_{\pi \in \mathcal{P}} V_\lambda(\pi, x), \quad x \in S, \tag{2.6}$$

and a policy π^* is *λ-optimal* if $V_\lambda(\pi^*, x) = V_\lambda^*(x)$ for all $x \in S$. The case $\lambda = 0$ of this criterion has been widely studied in the literature (see, for instance, Puterman [14] and the references therein) and this paper concentrates in the risk-sensitive context $\lambda \neq 0$.

Under Assumption 2.1, the expected value in (2.5) is well defined and the inequalities $0 \leq V_\lambda(\pi, \cdot) \leq V_\lambda^*(\cdot)$ always hold, but it is possible to have that $V_\lambda^*(x)$ is not finite for some $x \in S$ which is excluded from the discussion.

Assumption 2.2 *For each $x \in S$, $V_\lambda^*(x)$ is finite.*

As already noted, even when the state space is finite and Assumption 2.1 is strengthened to require the continuity of the transition law and the reward function, as well as the compactness of the action sets, the finiteness of $V_\lambda^*(\cdot)$ does not generally ensures the existence of an optimal stationary policy (Cavazos–Cadena and Montes–de–Oca [4]). Under the present Assumptions 2.1 and 2.2, the following notion of 'nearly' optimal stationary policy, used by Hordijk [11] in the risk-neutral context, will be used.

Definition 2.1 *Let $\lambda \neq 0$, $\varepsilon > 0$ and $f \in \mathbb{F}$ be fixed.*

(i) Policy f is ε-optimal at state x if $V_\lambda(f, x) \geq V_\lambda^*(x) - \varepsilon$;

(ii) f is ε-optimal if it is ε-optimal at every state $x \in S$.

The existence of optimal and ε-optimal stationary policies will be analysed using the basic properties of the optimal value function stated in the following two lemmas.

Lemma 2.1 *Let $\lambda \neq 0$ be fixed. Under* Assumptions 2.1 *and* 2.2, *the following assertions are valid.*

(i) The optimal value function $V_\lambda^*(\cdot)$ in (2.6) *satisfies the following λ-optimality equation (λ–OE).*

$$U_\lambda(V_\lambda^*(x)) = \sup_{a \in A(x)} \left[e^{\lambda R(x,a)} \sum_y p_{x\,y}(a) U_\lambda\left(V_\lambda^*(y)\right) \right], \quad x \in S. \tag{2.7}$$

(ii) Moreover, if the function $W \colon S \to [0, \infty)$ is such that

$$U_\lambda(W(x)) \geq \sup_{a \in A(x)} \left[e^{\lambda R(x,a)} \sum_y p_{x\,y}(a) U_\lambda\left(W(y)\right) \right], \quad x \in S, \tag{2.8}$$

then $W(\cdot) \geq V_\lambda^*(\cdot)$.

A proof of this result, using parallel arguments to those employed in the risk–neutral case, can be found, for instance, in Ávila–Godoy [1], Cavazos–Cadena and Fernández–Gaucherand [3], or in Cavazos–Cadena and Montes–de–Oca [4, 5]. The following generalization of Lemma 2.1 gives a strenghtened version of (2.7).

Lemma 2.2 *For each $n \in \mathbb{N}$, let $\mathcal{F}_n$ be the σ-field generated by*

$$(X_0, A_0, \ldots, X_{n-1}, A_{n-1}, X_n), \qquad \textit{and}$$

suppose that the positive random variable T is a stopping time with respect to $\{\mathcal{F}_n\}$, i.e., $P_x^\pi[T \in \{1,2,3,\ldots\} \cup \{\infty\}] = 1$ for every $x \in S$ and $\pi \in \mathcal{P}$, and for each $k \in \mathbb{N} \setminus \{0\}$, the event $[T = k]$ belongs to $\mathcal{F}_k$. In this case, the following strong λ–optimality equation is valid.

$$U_\lambda(V_\lambda^*(x)) = \sup_{\pi \in \mathcal{P}} E_x^\pi \left[e^{\lambda \sum_{t=0}^{T-1} R(X_t, A_t)} U_\lambda \left(V_\lambda^*(X_T) \right) \right], \quad x \in S. \tag{2.9}$$

Remark 2.2 *Throughout the paper the following convention is used. If T is a stopping time and $W: S \to \mathbb{R}$ is a given function, then $W(X_T) = 0$ on the event $[T = \infty]$. Thus, the expectation in the right hand side of* (2.9) *equals*

$$E_x^\pi \left[e^{\lambda \sum_{t=0}^{T-1} R(X_t, A_t)} U_\lambda \left(V_\lambda^*(X_T) \right) I[T < \infty] \right] \\ + E_x^\pi \left[e^{\lambda \sum_{t=0}^{\infty} R(X_t, A_t)} U_\lambda(0) I[T = \infty] \right].$$

A proof of Lemma 2.2, following the arguments presented in Hordijk [11], can be found in Cavazos–Cadena and Montes–de–Oca [5]. The above strong λ-OE, which will play an important role in the analysis of the existence of optimal and nearly optimal stationary policies, will be used when T is the first (positive) arrival time to a subset G of the state space. For $G \subset S$, define

$$T_G = \min\{n > 0 \mid X_n \in G\}, \tag{2.10}$$

where the minimum of the empty set is ∞. By convenience, the following notation is used when $G = \{x\}$ is a singleton.

$$T_{\{x\}} \equiv T_x, \quad x \in S. \tag{2.11}$$

3. A basic tool

As already mentioned, one of the objectives of this article is to analyse the existence of λ-optimal stationary policies for finite MDPs via the fundamental properties of the risk-sensitive expected total-reward criterion. In this section, the basic preliminary result to achieve this goal is stated as Theorem 3.1. First, let $\lambda \neq 0$, $f \in \mathbb{F}$ and $x_0 \in S$ be arbitrary but fixed, and consider the MDP $M_f = (S, A, \{A_f(x)\}, R, P)$, where $A_f(x) = \{f(x)\}$ for every $x \in S$, i.e., $f(x)$ is the single admissible action at state x with respect to model M_f. For every policy π in the class $\mathcal{P}_f$ of admissible policies for model M_f, it is clear that $\pi_t(A_f(x)|h_t) = 1$ always holds. Therefore it is not difficult to see that f is, essentially, the unique member of $\mathcal{P}_f$, in the sense that $P_x^\pi[\cdot] = P_x^f[\cdot]$ for every $x \in S$

and $\pi \in \mathcal{P}_f$. Consequently, $V_\lambda(f,\cdot)$ is the λ-optimal value function associated to M_f. Applying Lemma 2.2 with T_{x_0} and M_f instead of T and M, respectively, it follows that for every state x

$$\begin{aligned} U_\lambda\left(V(f,x)\right) &= E_x^f\left[e^{\lambda\sum_{t=0}^{T_{x_0}-1}R(X_t,A_t)}U_\lambda\left(V(f,X_{T_{x_0}})\right)\right] \\ &= E_x^f\left[e^{\lambda\sum_{t=0}^{T_{x_0}-1}R(X_t,A_t)}U_\lambda\left(V(f,x_0)\right)I[T_{x_0}<\infty]\right] \\ &\quad +E_x^f\left[e^{\lambda\sum_{t=0}^{T_{x_0}-1}R(X_t,A_t)}U_\lambda\left(0\right)I[T_{x_0}=\infty]\right], \end{aligned} \tag{3.1}$$

where the second equality is due to the convention in Remark 2.2. The main purpose of the section is to analyse the changes in this equality when policy f is modified at the given state x_0. Let $a \in A(x_0)$ be a fixed action, and define the new policy $\tilde{f} \in \mathbb{F}$ by

$$\tilde{f}(x) = \begin{cases} f(x), & x \neq x_0 \\ a, & x = x_0 \end{cases} \tag{3.2}$$

Since $V_\lambda(f,\cdot)$ is the optimal value function for model M_f, this definition of policy $\tilde{f}$ and Lemma 2.1(i) applied to M_f together yield

$$U_\lambda\left(V_\lambda(f,x)\right) = e^{\lambda R(x,\tilde{f}(x))}\sum_y p_{x,y}(\tilde{f}(x))U_\lambda(V_\lambda(f,y)), \quad x \in S\setminus\{x_0\}.$$

Although this equality is not generally satisfied when $x = x_0$, under Assumptions 2.1 and 2.2, it is not difficult to see that there exists $\delta \in \mathbb{R}$ be such that

$$U_\lambda\left(V_\lambda(f,x_0)\right) = e^{\lambda[\delta+R(x_0,\tilde{f}(x_0))]}\sum_y p_{x,y}(\tilde{f}(x_0))U_\lambda(V_\lambda(f,y)). \tag{3.3}$$

In the argument contained in Section 5, this equality will be satisfied for some $\delta \geq 0$. The following result extends Cavazos–Cadena and Montes–de–Oca [6, Lemma 3.1] to the present risk-sensitive framework.

Theorem 3.1 *Suppose that* Assumptions 2.1 *and* 2.2 *are valid, and let the policy* $\tilde{f} \in \mathbb{F}$ *and* $\delta \in \mathbb{R}$ *be as in* (3.2) *and* (3.3), *then*

$$\begin{aligned} &U_\lambda(V_\lambda(f,x_0)) \\ &= e^{\lambda\delta}\left(U_\lambda(V_\lambda(f,x_0))E_{x_0}^{\tilde{f}}\left[e^{\lambda\sum_{t=0}^{T_{x_0}-1}R(X_t,A_t)}I[T_{x_0}<\infty]\right]\right. \\ &\quad \left.+U_\lambda(0)E_{x_0}^{\tilde{f}}\left[e^{\lambda\sum_{t=0}^{\infty}R(X_t,A_t)}I[T_{x_0}=\infty]\right]\right). \end{aligned} \tag{3.4}$$

Proof Notice that (3.3) can be equivalently written as

$$
\begin{aligned}
&U_\lambda(V_\lambda(f,x_0)) \\
&= e^{\lambda[\delta+R(x_0,\tilde f(x_0))]}\Big(p_{x_0,x_0}(\tilde f(x_0))U_\lambda(V_\lambda(f,x_0)) \\
&\quad + \sum_{y\neq x_0} p_{x_0,y}(\tilde f(x_0))U_\lambda(V_\lambda(f,y))\Big) \\
&= e^{\lambda[\delta+R(x_0,\tilde f(x_0))]}p_{x_0,x_0}(\tilde f(x_0))U_\lambda(V_\lambda(f,x_0)) \\
&\quad + e^{\lambda[\delta+R(x_0,\tilde f(x_0))]}\sum_{y\neq x_0} p_{x_0,y}(\tilde f(x_0)) \\
&\qquad E_y^f\left[e^{\lambda\sum_{t=0}^{T_{x_0}-1}R(X_t,A_t)}U_\lambda\left(V(f,x_0)\right)I[T_{x_0}<\infty]\right] \\
&\quad + e^{\lambda[\delta+R(x_0,\tilde f(x_0))]}\sum_{y\neq x_0} p_{x_0,y}(\tilde f(x_0)) \\
&\qquad E_y^f\left[e^{\lambda\sum_{t=0}^{\infty}R(X_t,A_t)}U_\lambda\left(0\right)\right) I[T_{x_0}=\infty]\Big], \qquad (3.5)
\end{aligned}
$$

where (3.1) was used to obtain the second equality. On the other hand, since $V_\lambda(X_{T_{x_0}}) = V_\lambda(x_0)$ on the event $[T_{x_0} < \infty]$ (see (2.10) and (2.11)), it is clear that

$$
\begin{aligned}
&E_{x_0}^{\tilde f}\left[e^{\lambda\sum_{t=0}^{T_{x_0}-1}R(X_t,A_t)}U_\lambda(V_\lambda(X_{T_{x_0}})I[T_{x_0}=1]\right] \\
&= U_\lambda(V_\lambda(f,x_0))e^{\lambda R(x_0,\tilde f(x_0)}p_{x_0,x_0}(\tilde f(x_0)), \qquad (3.6)
\end{aligned}
$$

whereas, using the definition of the stopping time T_{x_0}, the Markov property yields

$$
\begin{aligned}
&E_{x_0}^{\tilde f}\left[e^{\lambda\sum_{t=0}^{T_{x_0}-1}R(X_t,A_t)}U_\lambda(V_\lambda(X_{T_{x_0}}))I[1<T_{x_0}<\infty]\,\middle|\, X_1=y\right] \\
&= \begin{cases} 0, & \text{if } y=x_0 \\ U_\lambda(V_\lambda(x_0))e^{\lambda R(x_0,\tilde f(x_0))} & \\ \quad E_y^{\tilde f}\left[e^{\lambda\sum_{t=0}^{T_{x_0}-1}R(X_t,A_t)}I[1\le T_{x_0}<\infty]\right], & \text{if } y\neq x_0. \end{cases} \qquad (3.7)
\end{aligned}
$$

Observe now that the expectation in right-hand side of this equality depends only on the actions selected at times $t < T_{x_0}$ and that $X_t \neq x_0$ when $1 \le t < T_{x_0}$ (see (2.10) and (2.11)), and in this case $f(X_t) = \tilde f(X_t)$. When $X_0 = y \neq x_0$, the latter equality also occurs at time $t = 0$, so that

$$
\begin{aligned}
&E_y^{\tilde f}\left[e^{\lambda\sum_{t=0}^{T_{x_0}-1}R(X_t,A_t)}I[1\le T_{x_0}<\infty]\right] \\
&= E_y^{f}\left[e^{\lambda\sum_{t=0}^{T_{x_0}-1}R(X_t,A_t)}I[1\le T_{x_0}<\infty]\right].
\end{aligned}
$$

Therefore, in the expectation in the right-hand side of (3.7), policy $\tilde{f}$ can be replaced by f, and taking the expected value with respect to X_1 in both sides of the resulting equation, it follows that

$$\begin{aligned} & E^{\tilde{f}}_{x_0}\left[e^{\lambda \sum_{t=0}^{T_{x_0}-1} R(X_t,A_t)} U_\lambda(V_\lambda(X_{T_{x_0}}))I[1 < T_{x_0} < \infty]\right] \\ = & \ U_\lambda(V_\lambda(x_0))e^{\lambda R(x_0,\tilde{f}(x_0))} \sum_{y\neq x_0} p_{x_0,y}(\tilde{f}(x_0)) \\ & \quad E^f_y\left[e^{\lambda \sum_{t=0}^{T_{x_0}-1} R(X_t,A_t)} I[1 \leq T_{x_0} < \infty]\right] \\ = & \ U_\lambda(V_\lambda(x_0))e^{\lambda R(x_0,\tilde{f}(x_0))} \sum_{y\neq x_0} p_{x_0,y}(\tilde{f}(x_0)) \\ & \quad E^f_y\left[e^{\lambda \sum_{t=0}^{T_{x_0}-1} R(X_t,A_t)} I[T_{x_0} < \infty]\right]. \end{aligned} \tag{3.8}$$

Recall that the inequality $T_{x_0} \geq 1$ always holds. Similarly, it can be proved that

$$\begin{aligned} & E^{\tilde{f}}_{x_0}\left[e^{\lambda \sum_{t=0}^{\infty} R(X_t,A_t)} I[T_{x_0} = \infty]\right] \\ = & \ e^{\lambda R(x_0,\tilde{f}(x_0))} \sum_{y\neq x_0} p_{x_0,y}(\tilde{f}(x_0)) E^f_y\left[e^{\lambda \sum_{t=0}^{\infty} R(X_t,A_t)} I[T_{x_0} = \infty]\right] \end{aligned}$$

Using this equation, together with (3.6) and (3.8), equation (3.5) yields that

$$\begin{aligned} & U_\lambda(V_\lambda(x_0)) \\ = & \ e^{\lambda\delta} U_\lambda\left(V_\lambda(x_0)\right) E^{\tilde{f}}_{x_0}\left[e^{\lambda \sum_{t=0}^{T_{x_0}-1} R(X_t,A_t)} I[T_{x_0} = 1]\right] \\ & + e^{\lambda\delta} U_\lambda\left(V_\lambda(x_0)\right) E^{\tilde{f}}_{x_0}\left[e^{\lambda \sum_{t=0}^{T_{x_0}-1} R(X_t,A_t)} I[1 < T_{x_0} < \infty]\right] \\ & + e^{\lambda\delta} U_\lambda\left(V_\lambda(0)\right) E^{\tilde{f}}_{x_0}\left[e^{\lambda \sum_{t=0}^{T_{x_0}-1} R(X_t,A_t)} I[T_{x_0} = \infty]\right] \end{aligned}$$

from which (3.4) follows. ■

4. Comparison of optimal value functions

This section presents an additional consequence of the strong optimality equation in Lemma 2.2, namely, if the action sets of several MDPs coincide except at some distinguished state x_0, then the corresponding optimal value functions can be compared.

Theorem 4.1 *Let $\lambda \neq 0$ and $x_0 \in S$ be fixed, and consider two MDPs $M_k = (S, A, \{A_k(x)\}, R, P)$, $k = 1, 2$, where $A_1(x_0)$ and $A_2(x_0)$ are nonempty (measurable) subsets of $A(x_0)$, and*

$$A_1(x) = A_2(x), \qquad x \in S \setminus \{x_0\}. \tag{4.1}$$

Let $\mathcal{P}_k$ be the class of admissible policies for model M_k and denote by $V^*_{\lambda,k}$ the corresponding optimal value function, i.e.,

$$V^*_{\lambda,k}(x) = \sup_{\pi \in \mathcal{P}_k} V_\lambda(\pi, x), \quad x \in S.$$

In this case,

(i) If δ is a nonnegative number,

$$V^*_{\lambda,2}(x_0) + \delta \ \geq \ V^*_{\lambda,1}(x_0) \Longrightarrow V^*_{\lambda,2}(\cdot) + \delta \ \geq \ V^*_{\lambda,1}(\cdot).$$

(ii) In particular,

$$V^*_{\lambda,2}(x_0) \ \geq \ V^*_{\lambda,1}(x_0) \Longrightarrow V^*_{\lambda,2}(\cdot) \ \geq \ V^*_{\lambda,1}(\cdot). \tag{4.2}$$

Proof Let $\pi \in \mathcal{P}_1$ be an arbitrary policy, select a stationary policy f_2 for model M_2, and define a new policy $\pi' \in \mathcal{P}_2$ as follows: $\pi'_t(\cdot|h_t) = \pi_t(\cdot|h_t)$ if $x_k \neq x_0$ for $k = 0, 1, \ldots, t$, whereas $\pi'_t(\{f_2(x_t)\}|h_t) = 1$ if $x_k = x_0$ for some $k \leq t$. In other words, if $X_0 \neq x_0$, policies π and π' coincide before the first visit to state x_0 (at time T_{x_0}), but once x_0 is reached at some time k, π' chooses actions according to f_2 from time k onwards. From (4.1) it follows that $\pi' \in \mathcal{P}_2$. Thus, given $X_0 = x \neq x_0$, the fact that π and π' coincide before time T_{x_0} implies that

$$\begin{aligned} U_\lambda\left(V^*_{\lambda,2}(x)\right) \ &\geq \ E^{\pi'}_x\left[e^{\lambda \sum_{t=0}^{T_{x_0}-1} R(X_t,A_t)} U_\lambda\left(V^*_{\lambda,2}(X_{T_{x_0}})\right)\right] \\ &= \ E^{\pi}_x\left[e^{\lambda \sum_{t=0}^{T_{x_0}-1} R(X_t,A_t)} U_\lambda\left(V^*_{\lambda,2}(X_{T_{x_0}})\right)\right], \end{aligned} \tag{4.3}$$

where Lemma 2.2 applied to model M_2 was used to obtain the inequality. Suppose now that

$$V^*_{\lambda,2}(x_0) + \delta \geq V^*_{\lambda,1}(x_0), \qquad \text{where } \delta \geq 0.$$

In this case,

$$V^*_{\lambda,2}(X_{T_{x_0}}) + \delta \geq V^*_{\lambda,1}(X_{T_{x_0}}). \tag{4.4}$$

In fact,

$$V^*_{\lambda,2}(X_{T_{x_0}}) + \delta \ = \ V^*_{\lambda,2}(x_0) + \delta \ \geq \ V^*_{\lambda,1}(x_0) \ = \ V^*_{\lambda,1}(X_{T_{x_0}})$$

on the event $[T_{x_0} < \infty]$ whereas, by the convention in Remark 2.2, $V^*_{\lambda,2}(X_{T_{x_0}}) + \delta = \delta \geq 0 = V^*_{\lambda,1}(X_{T_{x_0}})$ when $T_{x_0} = \infty$. Thus, multiplying both sides of (4.3) by $e^{\lambda\delta}$, (2.2) implies that, for $x \neq x_0$,

$$\begin{aligned} U_\lambda\left(V^*_{\lambda,2}(x) + \delta\right) &\geq E^\pi_x\left[e^{\lambda \sum_{t=0}^{T_{x_0}-1} R(X_t,A_t)} U_\lambda\left(V^*_{\lambda,2}(X_{T_{x_0}} + \delta)\right)\right] \\ &\geq E^\pi_x\left[e^{\lambda \sum_{t=0}^{T_{x_0}-1} R(X_t,A_t)} U_\lambda\left(V^*_{\lambda,1}(X_{T_{x_0}})\right)\right] \end{aligned}$$

where the second inequality uses (4.4) and the fact that the utility function is strictly increasing. Since $\pi \in \mathcal{P}_1$ is arbitrary, from another application of Lemma 2.2, then for model M_1, it follows that

$$U_\lambda\left(V^*_{\lambda,2}(x) + \delta\right) \geq U_\lambda\left(V^*_{\lambda,1}(x)\right), \quad x \neq x_0,$$

which is equivalent to

$$V^*_{\lambda,2}(x) + \delta \geq V^*_{\lambda,1}(x), \quad x \neq x_0. \tag{4.5}$$

In short, it has been proved that when $\delta \geq 0$, the inequality $V^*_{\lambda,2}(x_0) + \delta \geq V^*_{\lambda,1}(x_0)$ implies (4.5), establishing part (i), and then part (ii) equation (4.2) is obtained by setting $\delta = 0$ in part (i). ■

Corollary 4.1 *Given $\lambda \neq 0$, consider an MDP*

$$M = (S, A, \{A(x)\}, R, P)$$

satisfying Assumptions 2.1 *and* 2.2, *and let $x_0 \in S$ be a fixed state for which the corresponding action set is finite and has $r > 1$ elements, say*

$$A(x_0) = \{a_1, a_2, \ldots, a_r\}.$$

For each $k = 1, 2, \ldots, r$, define a new MDP $M_k = (S, A, \{A_k(x)\}, R, P)$ by setting

$$A_k(x) = A(x), \quad x \in S \setminus \{x_0\}, \quad \text{and} \quad A_k(x_0) = A(x_0) \setminus \{a_k\}. \tag{4.6}$$

Let $\mathcal{P}_k$ be the class of admissible policies for model M_k, and let $V^*_{\lambda,k}$ be the corresponding optimal value function, i.e.,

$$V^*_{\lambda,k}(x) = \sup_{\pi \in \mathcal{P}_k} V_\lambda(\pi, x), \quad x \in S,$$

so that $V^*_{\lambda,k}(x) \in [0, V^*_\lambda(x)]$, $x \in S$. In this case, there exists a permutation $(k_1, k_2, \ldots, k_r)$ of the set $\{1, 2, \ldots, r\}$ such that

$$V^*_{\lambda,k_1}(\cdot) \geq V^*_{\lambda,k_2}(\cdot) \geq V^*_{\lambda,k_3}(\cdot) \geq \cdots \geq V^*_{\lambda,k_r}(\cdot). \tag{4.7}$$

Proof Notice that the action sets for the different models M_k coincide except at the distinguished state x_0. Consider now the sequence of nonnegative numbers $(V_{\lambda,1}(x_0), V_{\lambda,2}(x_0), \ldots, V_{\lambda,r}(x_0))$. Since this is a sequence of real numbers, there exists a permutation $(k_1, k_2, \ldots, k_r)$ of the set $\{1, 2, \ldots, r\}$ satisfying

$$V^*_{\lambda,k_1}(x_0) \geq V^*_{\lambda,k_2}(x_0) \geq \cdots \geq V^*_{\lambda,k_r}(x_0),$$

and then Theorem 4.1 yields that (4.6) is satisfied by this permutation. ■

5. Optimality for finite models

In this section the existence of λ-optimal stationary policies for MDPs with finite state and action sets is established. As previously noted, this result has been recently obtained via a discounted dynamic operator associated to an auxiliary stochastic game (Cavazos–Cadena and Montes–de–Oca [5]). In contrast, the induction argument presented below depends solely on the the properties derived from the strong optimality equation in the previous sections.

Theorem 5.1 *Let $\lambda \neq 0$ be a fixed real number, and suppose that the MDP $M = (S, A, \{A(x)\}, R, P)$ satisfies* Assumptions 2.1(ii) *and* 2.2, *and*

$$|S| + \sum_{x \in S} |A(x)| < \infty \tag{5.1}$$

where, for each set B, $|B|$ denotes the number of elements of B. In this case, there exists a λ-optimal stationary policy.

Proof Consider the class $\mathcal{M}$ consisting of MDPs

$$M = (S, A, \{A(x)\}, R, P)$$

for which Assumptions 2.1(ii) and 2.2 are valid, and $|S|$ is a fixed positive number n. It will be proved, by induction of the value of $\sum_{x\in S} |A(x)|$, that an optimal stationary policy exists for each model $M \in \mathcal{M}$ whenever the summation is finite. To begin with, notice that, since each set $A(x)$ is nonempty, the inequality $\sum_{x\in S} |A(x)| \geq |S|$ always holds.

Initial step

The conclusion is valid when $\sum_{x\in S} |A(x)| = |S| = n$. In fact, in this case each set $A(x)$ is a singleton, and then $\mathbb{F}$ contains a single member, say $\mathbb{F} = \{f\}$. Moreover, f is essentially the unique element of $\mathcal{P}$, so that f is optimal; see the comments at the beginning of Section 3.

Induction step

Suppose that an optimal stationary policy exist for each MDP $\widetilde{M} = (S, A, \{\widetilde{A}(x)\}, R, P) \in \mathcal{M}$ satisfying that $\sum_{x\in S} |\widetilde{A}(x)| = m - 1 \geq |S| = n$. It will be proved that an optimal stationary policy exists for a model $M = (S, A, \{A(x)\}, R, P) \in \mathcal{M}$ satisfying that $\sum_{x\in S} |A(x)| = m > n = |S|$. To achieve this goal, first notice that, since $m > |S|$, there exists $x_0 \in S$ such that $|A(x_0)| = r \geq 2$, and write $A(x_0) = \{a_1, a_2, \ldots, a_r\}$. For each $k = 1, 2, \ldots, r$, let

$$M_k = (S, A, \{A_k(x)\}, R, P)$$

be the MDP defined in the statement of Corollary 4.1, and let the permutation $k_1, k_2, \ldots, k_r$ of the set $\{1, 2, \ldots, r\}$ be such that (4.7) holds. Observe now the following facts (i) and (ii):

(i) From the definition of the sets $A_k(\cdot)$ in (4.6), it follows that $\sum_{x\in S} |A_k(x)| = m - 1$, so that, by the induction hypothesis, each model M_k admits an optimal stationary policy. In particular, there exists a stationary policy f such that

$$f(x) \in A_{k_1}(x), \quad x \in S, \quad \text{and} \quad V_\lambda(f, \cdot) = V^*_{\lambda, k_1}(\cdot). \tag{5.2}$$

(ii) As it will be shown below,

$$V^*_{\lambda, k_1}(\cdot) = V^*_\lambda(\cdot). \tag{5.3}$$

From this equality and (5.2) it follows that $V_\lambda(f, \cdot) = V^*_\lambda(\cdot)$, so that f is optimal for model M, completing the induction argument. Thus, to conclude the proof it is necessary to verify (5.3). Since $V_\lambda(f, \cdot) \leq V^*_\lambda(\cdot)$ (see (2.6)), it is sufficient to show that $V_\lambda(f, \cdot) \geq V^*_\lambda(\cdot)$, an inequality that, using Lemma 2.1(ii) and the nonnegativity of $V_\lambda(f, \cdot)$, follows from

$$U_\lambda(V_\lambda(f, x)) \geq e^{\lambda R(x,a)} \sum_y p_{x\,y}(a) U_\lambda(V_\lambda(f, y)), \quad x \in S, \; a \in A(x). \tag{5.4}$$

To show that this statement is satisfied, observe that using the optimality equation for model M_{k_1} together with the equality in (5.2), it follows that the inequality in (5.4) occurs whenever $(x, a) \in \mathbb{K} \setminus \{(x_0, a_{k_1})\}$, so that to verify (5.4) it is sufficient to show that

$$U_\lambda(V_\lambda(f, x_0)) \geq e^{\lambda R(x_0, a_{k_1})} \sum_y p_{x\,y}(a_{k_1}) U_\lambda(V_\lambda(f, y)). \tag{5.5}$$

With this in mind, observe that the two numbers being compared in this inequality are finite and have the same sign, so that there exist $\delta \in \mathbb{R}$ such that

$$U_\lambda(V_\lambda(f,x_0)) = e^{\lambda[\delta + R(x_0,a_{k_1})]} \sum_y p_{x\,y}(a_{k_1}) U_\lambda(V_\lambda(f,y)), \qquad (5.6)$$

and it is not difficult to verify that, regardless of the sign of λ, (5.5) is equivalent to

$$\delta \geq 0, \qquad (5.7)$$

an assertion that can be verified as follows. Using a_{k_1} instead of a, construct the policy $\tilde{f}$ in (3.2) and observe that $\tilde{f}$ is not an admissible policy for model M_{k_1}, but belongs to the space of stationary policies for model M_{k_2}; notice that $a_{k_1} \in A(x_0)\setminus\{a_{k_2}\} = A_{k_2}(x_0)$. Thus, $V^*_{\lambda,k_2}(\cdot) \geq V_\lambda(\tilde{f},\cdot)$, so that (4.7) and (5.2) together yield that

$$V_\lambda(f,\cdot) \;=\; V^*_{\lambda,k_1}(\cdot) \;\geq\; V^*_{\lambda,k_2}(\cdot) \;\geq\; V_\lambda(\tilde{f},\cdot). \qquad (5.8)$$

On the other hand, (3.1) with $\tilde{f}$ and x_0 instead of f and x, respectively, yields that

$$\begin{aligned} &U_\lambda\left(V_\lambda(\tilde{f},x_0)\right)\left(1 - E^{\tilde{f}}_{x_0}\left[e^{\lambda\sum_{t=0}^{T_{x_0}-1} R(X_t,A_t)} I[T_{x_0} < \infty]\right]\right) \\ &= \; U_\lambda(0) E^{\tilde{f}}_{x_0}\left[e^{\lambda\sum_{t=0}^{\infty} R(X_t,A_t)} I[T_{x_0} = \infty]\right], \end{aligned} \qquad (5.9)$$

whereas, by equation (3.4) established in Theorem 3.1,

$$\begin{aligned} &U_\lambda\left(V_\lambda(f,x_0)\right)\left(1 - e^{\lambda\delta} E^{\tilde{f}}_{x_0}\left[e^{\lambda\sum_{t=0}^{T_{x_0}-1} R(X_t,A_t)} I[T_{x_0} < \infty]\right]\right) \\ &= \; e^{\lambda\delta} U_\lambda(0) E^{\tilde{f}}_{x_0}\left[e^{\lambda\sum_{t=0}^{\infty} R(X_t,A_t)} I[T_{x_0} = \infty]\right], \end{aligned}$$

which is equivalent to

$$\begin{aligned} &U_\lambda\left(V_\lambda(f,x_0)\right)\left(e^{-\lambda\delta} - E^{\tilde{f}}_{x_0}\left[e^{\lambda\sum_{t=0}^{T_{x_0}-1} R(X_t,A_t)} I[T_{x_0} < \infty]\right]\right) \\ &= \; U_\lambda(0) E^{\tilde{f}}_{x_0}\left[e^{\lambda\sum_{t=0}^{\infty} R(X_t,A_t)} I[T_{x_0} = \infty]\right]. \end{aligned}$$

Combining this equation with (5.9) and using the expression for the utility function in (2.1), it follows that

$$\begin{aligned} &e^{\lambda[V_\lambda(\tilde{f},x_0) - V_\lambda(f,x_0)]}\left(1 - E^{\tilde{f}}_{x_0}\left[e^{\lambda\sum_{t=0}^{T_{x_0}-1} R(X_t,A_t)} I[T_{x_0} < \infty]\right]\right) \\ &= \; e^{-\lambda\delta} - E^{\tilde{f}}_{x_0}\left[e^{\lambda\sum_{t=0}^{\infty} R(X_t,A_t)} I[T_{x_0} < \infty]\right]. \end{aligned} \qquad (5.10)$$

Observe now that $E_{x_0}^{\tilde f}\left[e^{\lambda\sum_{t=0}^{T_{x_0}-1}R(X_t,A_t)}I[T_{x_0}<\infty]\right]\le 1$. Indeed, the occurrence of the reverse inequality leads to the contradiction that both sides of (5.9) have different signs. To establish (5.7), consider the following three cases, which are exhaustive:

Case 1 $E_{x_0}^{\tilde f}\left[e^{\lambda\sum_{t=0}^{T_{x_0}-1}R(X_t,A_t)}I[T_{x_0}<\infty]\right]=1$.

In this situation (5.10) implies that

$$\begin{aligned} e^{-\lambda\delta} &= E_{x_0}^{\tilde f}\left[e^{\lambda\sum_{t=0}^{T_{x_0}-1}R(X_t,A_t)}I[T_{x_0}<\infty]\right] \\ &= 1, \end{aligned}$$

and then, recalling that $\lambda\neq 0$, it follows that $\delta=0$, so that (5.7) is certainly valid.

Case 2 $E_{x_0}^{\tilde f}\left[e^{\lambda\sum_{t=0}^{T_{x_0}-1}R(X_t,A_t)}I[T_{x_0}<\infty]\right]<1$ and $\lambda>0$.

Since $\lambda>0$, (5.8) implies that $e^{\lambda[V_\lambda(\tilde f,x_0)-V_\lambda(f,x_0)]}\le 1$, and then (5.10) yields

$$\begin{aligned} &1-E_{x_0}^{\tilde f}\left[e^{\lambda\sum_{t=0}^{T_{x_0}-1}R(X_t,A_t)}I[T_{x_0}<\infty]\right] \\ &\quad\ge e^{-\lambda\delta}-E_{x_0}^{\tilde f}\left[e^{\lambda\sum_{t=0}^{\infty}R(X_t,A_t)}I[T_{x_0}=\infty]\right], \end{aligned}$$

i.e, $1\ge e^{-\lambda\delta}$, and (5.7) follows combining this inequality with the positivity of λ.

Case 3 $E_{x_0}^{\tilde f}\left[e^{\lambda\sum_{t=0}^{T_{x_0}-1}R(X_t,A_t)}I[T_{x_0}<\infty]\right]<1$ and $\lambda<0$.

In this context, it follows from (5.8) that

$$e^{\lambda[V_\lambda(\tilde f,x_0)-V_\lambda(f,x_0)]}\ge 1,$$

since $\lambda<0$. Then (5.10) implies

$$\begin{aligned} &1-E_{x_0}^{\tilde f}\left[e^{\lambda\sum_{t=0}^{T_{x_0}-1}R(X_t,A_t)}I[T_{x_0}<\infty]\right] \\ &\quad\le e^{-\lambda\delta}-E_{x_0}^{\tilde f}\left[e^{\lambda\sum_{t=0}^{\infty}R(X_t,A_t)}I[T_{x_0}=\infty]\right], \end{aligned}$$

that is, $e^{|\lambda|\delta}=e^{-\lambda\delta}\ge 1$, inequality that yields (5.7).

In short, it has been proved that δ in (5.6) is nonnegative, which, as already noted, implies (5.5) and concludes the proof of the theorem. ■

Under Assumptions 2.1 and 2.2, Theorem 5.1 ensures that a λ-optimal policy exists when condition (5.1) holds. If the state space is finite but some action set is infinite, the existence result may fail even under strong continuity conditions on the transition-reward structure of the model (Cavazos–Cadena and Montes–de–Oca [4]); as a complement to this point, an example is now provided in which all the action sets are finite, the state space is denumerable and infinite, but an optimal stationary policy does not exist.

Example 5.1 *Suppose that $S = \mathbb{N}$ and $A = \{0,1\} = A(x)$ for every $x \in S$, and define the transition law and the reward function as follows:*

$$\begin{aligned} p_{x\,x+1}(0) &= 1 = p_{x\,0}(1), && x \neq 0, \text{ and} \\ p_{0\,0}(a) &= 1, && a = 0,1; \\ R(x,0) &= 0, && x \in \mathbb{N}, \\ R(x,1) &= 1 - \frac{1}{x}, && x \neq 0, \text{ and} \\ R(0,1) &= 0. \end{aligned}$$

Proposition 5.1 *In* Example 5.1, *assertions* (i) *and* (ii) *hold, where $W\colon \mathbb{N} \to \mathbb{R}$ is given by $W(x) = 1$ for $x \neq 0$ and $W(0) = 0$.*

(i) $V_\lambda^*(\cdot) = W(\cdot)$, *and*

(ii) A λ-optimal policy does not exist.

Proof Since state 0 is absorbing and $R(0,\cdot) \equiv 0$, it is clear that $P_0^\pi[R(X_t, A_t) = 0] = 1$ and then $V_\lambda(\pi, 0) = 0$ for every policy π, so that

$$V_\lambda^*(0) = 0 = W(0); \tag{5.11}$$

see (2.5) and (2.6). Next, suppose that $X_0 = x \in \mathbb{N} \setminus \{0\}$, let $\pi \in \mathcal{P}$ be arbitrary be fixed, and set

$$\tau = \min\{n \geq 0 \,|\, A_n = 1\}.$$

In this case, it is clear that, on the event $[\tau = \infty]$, $A_t = 0$ for every $t \in \mathbb{N}$, whereas the definition of the transition law yields on the event $[\tau < \infty]$, $A_t = 0$,

$$X_\tau = x + \tau, \text{ for } t < \tau, \quad \text{and} \quad X_t = 0, \text{ for } t > \tau.$$

Therefore, from the the definition of the reward function, it follows that

$$\begin{aligned} \sum_{t=0}^{\infty} R(X_t, A_t) &= \left(1 - \frac{1}{x + \tau}\right) I[\tau < \infty] \\ &< 1 \qquad P_x^\pi[\cdot]\text{-almost surely,} \end{aligned}$$

then, since $U_\lambda(\cdot)$ is increasing,

$$\begin{aligned} U_\lambda(V_\lambda(\pi,x)) &= E_x^\pi\left[U_\lambda\left(1-\frac{1}{x+\tau}\right)I[\tau<\infty]+U_\lambda(0)I[\tau=\infty])\right] \\ &< U(1), \end{aligned} \tag{5.12}$$

so that,

$$V_\lambda(\pi,x) < 1 = W(x). \tag{5.13}$$

Consider now the stationary policy f_n defined by $f_n(y)=0$ if $y\neq x+n$ and $f_n(x+n)=1$. In this case $P_x^{f_n}[\tau=n]=1$, and then the equality in (5.12) with f_n instead of π yields that $U_\lambda(V_\lambda(f_n,x)) = U_\lambda\left(1-\dfrac{1}{x+n}\right)$, i.e.,

$$V_\lambda(f_n,x) = 1-\frac{1}{x+n}.$$

Since $n\in\mathbb{N}$ and $\pi\in\mathcal{P}$ are arbitrary, this equality and (5.13) together yield, via (2.6), that $V_\lambda^*(x)=1=W(x)$. Since $x\in\mathbb{N}\setminus\{0\}$ was arbitrary in this argument, part (i) follows from this latter equality and (5.11), then part (ii) is obtained from the inequality in (5.13) which is valid for every $x\neq 0$ and $\pi\in\mathcal{P}$. ∎

In the remainder of the paper, attention concentrates on the existence of ε-optimal policies; see Definition 2.1.

6. Approximate optimality: part I

This section concerns MDPs with finite state space but, in contrast with Section 5, the action sets are assumed to be arbitrary (measurable) subsets of the action space. As previously noted, within this context, the existence of a stationary policy can not be generally ensured. However, as stated in Theorem 6.1 below, an ε-optimal stationary policy exists under Assumptions 2.1 and 2.2.

Theorem 6.1 *Suppose that* Assumptions 2.1 *and* 2.2 *hold, and that the state space is finite. In this case, given $\varepsilon>0$, there exists a stationary policy f which is ε–optimal, i.e.,*

$$V_\lambda(f,x) > V_\lambda^*(x)-\varepsilon, \quad x\in S.$$

The idea to establish this result consists in approximating the original MDP M by models whose action sets are appropriate finite sets, showing that in this reduction process, the optimal value function does not change 'substantially'. Although Theorem 6.1 was obtained in Cavazos–Cadena and Montes–de–Oca [5], the proof presented below, relying on the results of Section 4 and on the dominance property in Lemma 2.1(ii), is simpler.

Lemma 6.1 *Suppose that* Assumptions 2.1 *and* 2.2 *hold and let* $x_0 \in S$ *be fixed. For each nonempty and finite set* $G \subset A(x_0)$*, consider the new MDP* $M_G = (S, A, \{A_G(x)\}, R, P)$ *obtained by setting*

$$A_G(x) = A(x), \quad x \in S \setminus \{x_0\}, \text{ and } \quad A_G(x_0) = G. \tag{6.1}$$

Let $\mathcal{P}_G$ *be the class of admissible policies of* M_G *and denote by* $V^*_{\lambda,G}(\cdot)$ *the corresponding optimal value function, i.e.,*

$$V^*_{\lambda,G}(x) = \sup_{\pi \in \mathcal{P}_G} V_\lambda(\pi, x), \quad x \in S. \tag{6.2}$$

With this notation, assertions (i)–(iii) below are valid, where G *and* H *are nonempty and finite subsets of* $A(x_0)$.

(i) $G \subset H \Longrightarrow V^*_{\lambda,G}(\cdot) \le V^*_{\lambda,H}(\cdot)$.
Set

$$L(x) = \sup\left\{V^*_{\lambda,H}(x) \mid \emptyset \ne H \subset A(x_0),\ H \text{ is finite}\right\}, \quad x \in S. \tag{6.3}$$

(ii) Given $\varepsilon > 0$*, there exists a nonempty and finite set* $G = G(\varepsilon) \subset A(x_0)$ *such that*

$$V^*_{\lambda,G}(x) \ \ge\ L(x) - \varepsilon, \quad x \in S. \tag{6.4}$$

(iii) For each $\varepsilon > 0$*, the set* $G = G(\varepsilon)$ *in part (ii) satisfies*

$$V^*_{\lambda,G}(x) \ \ge\ V^*_\lambda(x) - \varepsilon, \quad x \in S, \tag{6.5}$$

Proof

(i) Using (6.1), it is clear that $G \subset H \Longrightarrow \mathcal{P}_G \subset \mathcal{P}_H$, and the assertion follows from (6.2).

(ii) Given $\varepsilon > 0$, select a finite set G such that $\emptyset \ne G \subset A(x_0)$ and $V^*_{\lambda,G}(x_0) + \varepsilon \ge L(x_0)$. In this case, from the definition of $L(\cdot)$ in (6.3), it follows that $V^*_{\lambda,G}(x_0) + \varepsilon \ge L(x_0) \ge V^*_{\lambda,H}(x_0)$ whenever H is a finite and nonempty subset of $A(x_0)$. Since models M_G and M_H have the same action sets, except at x_0, the above inequality implies, via Theorem 4.1(i), that $V^*_{\lambda,G}(\cdot) + \varepsilon \ge V^*_{\lambda,H}(\cdot)$, and, since the nonempty and finite subset H of $A(x_0)$ is arbitrary, (6.3) yields $V^*_{\lambda,G}(\cdot) + \varepsilon \ge L(\cdot)$.

(iii) By part (ii), it is sufficient to show that $L(\cdot) \geq V_\lambda(\cdot)$. With this in mind, given $\varepsilon > 0$ select a set $G = G(\varepsilon)$ as in (6.4), so that for every finite set H satisfying $G \subset H \subset A(x_0)$, part (i) and (6.2) together yield

$$L(\cdot) \geq V_{\lambda,H}(\cdot) \geq V_{\lambda,G}(\cdot) \geq L(\cdot) - \varepsilon.$$

Combining this fact with the optimality equation for model M_H and the monotonicity property of the utility function, it follows that

$$\begin{aligned} U_\lambda(L(x)) &\geq U_\lambda(V_{\lambda,H}(x)) \\ &\geq e^{\lambda R(x,a)} \sum_y p_{xy}(a) U_\lambda(V_{\lambda,H}(y)) \\ &\geq e^{\lambda R(x,a)} \sum_y p_{xy}(a) U_\lambda(L(y) - \varepsilon) \\ &= e^{\lambda\varepsilon} e^{\lambda R(x,a)} \sum_y p_{xy}(a) U_\lambda(L(y)), \\ &\qquad x \in S,\ a \in A_H(x), \end{aligned} \tag{6.6}$$

where (2.2) was used to obtain the equality. Since $A_H(x) = A(x)$ for $x \neq x_0$, this yields that

$$U_\lambda(L(x)) \geq e^{-\lambda\varepsilon} e^{\lambda R(x,a)} \sum_y p_{xy}(a) U_\lambda(L(y)), \quad x \neq x_0,\ a \in A(x) \tag{6.7}$$

whereas $A_H(x_0) = H$ and (6.6) together imply that

$$U_\lambda(L(x_0)) \geq e^{-\lambda\varepsilon} e^{\lambda R(x_0,a)} \sum_y p_{x_0 y}(a) U_\lambda(L(y)), \quad a \in H.$$

However, the finite set set H satisfying $G \subset H \subset A(x_0)$ is arbitrary, so that the above displayed relation implies

$$U_\lambda(L(x_0)) \geq e^{-\lambda\varepsilon} e^{\lambda R(x_0,a)} \sum_y p_{x_0 y}(a) U_\lambda(L(y)), \quad a \in A(x_0),$$

and combining this statement with (6.7), it follows that for every $x \in S$ and $a \in A(x)$, $U_\lambda(L(x)) \geq e^{\lambda\varepsilon} e^{\lambda R(x,a)} \sum_y p_{xy}(a) U_\lambda(L(y))$. Since $\varepsilon > 0$ is arbitrary, this yields

$$U_\lambda(L(x)) \geq e^{\lambda R(x,a)} \sum_y p_{xy}(a) U_\lambda(L(y)), \quad x \in S,\ a \in A(x)$$

and then, since $L(\cdot)$ is nonnegative, from Lemma 2.1(ii) it follows that $L(\cdot) \geq V_\lambda^*(\cdot)$. ■

Proof of Theorem 6.1 Given an MDP $M = (S, A, \{A(x)\}, R, P)$, set

$$\text{NFAS}(M) = \text{number of finite sets among the class } \{A(x)\,|x \in S\}.$$

Within the family $\mathcal{M}$ of MDPs for which Assumptions 2.1 and 2.2 hold, and whose state space has $n \geq 1$ members, consider the following proposition:

Proposition 6.1
$\mathbb{P}(k)$: *If* $M \in \mathcal{M}$ *is such that* $\text{NFAS}(M) = k$, *then for every* $\varepsilon > 0$ *there exists an* ε*-optimal stationary policy for model* $\mathcal{M}$.

To establish Theorem 6.1 it is clearly sufficient to prove that $\mathbb{P}(k)$ occurs for $k = 0, 1, 2, \ldots, n$, which will be verified by backward induction.

Initial step $\mathbb{P}(n)$ is valid.

When $M \in \mathcal{M}$ satisfies $\text{NFAS}(M) = n$, for every action set, $A(x)$ is finite, so that, by Theorem 5.1, there exists an optimal stationary policy f for model M. Clearly, such an f is ε-optimal for every $\varepsilon > 0$.

Induction step If $\mathbb{P}(k)$ holds for some $k > 0$, then $\mathbb{P}(k-1)$ occurs.

Suppose that $\mathbb{P}(k)$ is valid and assume that the MDP

$$M = (S, A, \{A(x)\}, R, P) \in \mathcal{M}(\text{where } |S| = n),$$

satisfies $\text{NFAS}(M) = k - 1$.
Write $S = \{x_1, x_2, \ldots, x_{k-1}, x_k, \ldots, x_n\}$ and without loss of generality suppose that $A(x_s)$ is finite for $s < k$. Given $\varepsilon > 0$, Lemma 6.1 applied to this model M with x_k instead of x_0 yields a nonempty and *finite* set G such that the new MDP $M_G = (S, A, \{A_G(x)\}, R, P)$ satisfies Assumptions 2.1 and 2.2, as well as the following properties (i) and (ii):

(i) $A_G(x) = A(x)$ for $x \neq x_k$ and $A(x_k) = G$, and

(ii)
$$V_{\lambda,G}^*(\cdot) \geq V_\lambda(\cdot) - \frac{\varepsilon}{2}, \tag{6.8}$$
where $V_{\lambda,G}^*(\cdot)$ is the optimal value function of M_G.

Since $\text{NFAS}(M) = k - 1$, the definition of the action sets $A_G(x)$ in (i) above yields that $\text{NFAS}(M_G) = k$, so that, by the induction hypothesis, there exists a stationary policy f such that $V_\lambda^*(f,\cdot) \geq V_{\lambda,G}^*(\cdot) - \varepsilon/2$ and combining this inequality with (6.8) it follows that

$$V_\lambda(f,\cdot) \geq V_\lambda(\cdot) - \varepsilon,$$

i.e., f is ε–optimal for model M; since M was an arbitrary MDP in the family $\mathcal{M}$ satisfying that NFAS$(M) = k-1$, it follows that $\mathbb{P}(k-1)$ is valid, completing the induction argument. ■

7. Approximate optimality: part II

This section concerns the existence of ε–optimal stationary policies for MDPs with general denumerable state space. The main objective is to establish Theorem 7.1 below, which extends results obtained in Cavazos–Cadena and Montes–de–Oca [5] for the risk-averse case $\lambda < 0$.

Theorem 7.1 *Let the risk-sensitivity coefficient λ be a positive number, and consider an MDP*

$$M = (S, A, \{A(x)\}, R, P)$$

with general denumerable state space satisfying Assumption 2.1 *as well as the condition that*

$$\|V_\lambda^*(\cdot)\| < \infty. \tag{7.1}$$

Then for every $\varepsilon > 0$ there exists a stationary policy f which is ε–optimal, i.e.,

$$V_\lambda(f\,\cdot) \;\geq\; V_\lambda^*(\cdot) - \varepsilon. \tag{7.2}$$

Remark 7.1

(i) For the risk-neutral case $\lambda = 0$, Ornstein [12] *obtained the following result (see also Hordijk* [11]*).*

Under Assumption 2.1, *the finiteness of the optimal value function implies that, for each $\varepsilon \in (0,1)$, there exists a stationary policy which is ε-optimal in the relative sense, i.e.,*

$$V_0(f,\cdot) \geq (1-\varepsilon)V_0^*(\cdot).$$

When $\|V_0^*(\cdot)\| < \infty$, this implies that for every $\varepsilon > 0$ it is possible to find a policy $f \in \mathbb{F}$ which is ε-optimal in the (absolute) sense of Definition 2.1(ii). *Thus,* Theorem 7.1 *is an extension of this latter result to the risk-seeking context $\lambda > 0$.*

(ii) When $\lambda < 0$, and the other conditions in Theorem 7.1 *occur, it was proved in Cavazos–Cadena and Montes–de–Oca* [5] *that for*

each $\varepsilon \in (0,1)$ there exists a stationary policy f which is ε-optimal in the relative sense, that is, $V^*_\lambda(f,\cdot) \geq (1-\varepsilon)V^*_\lambda(\cdot)$. Under (7.1), *this result implies that the conclusion of* Theorem 7.1 *also occurs for in the risk-averse case.*

(iii) The proof of Theorem 7.1 *presented below follows the route signaled by Ornstein* [12] *(see also Hordijk* [11, Chapter 13]*). This approach was adapted to the risk-averse case $\lambda < 0$ in Cavazos–Cadena and Montes–de–Oca* [5], *and it is interesting to point out that the key step of the proof, namely,* Lemma 7.2(iii) *below, requires a* substantially *different treatment in the risk-seeking and risk-averse cases.*

The strategy to prove Theorem 7.1 *needs two preliminary steps.*

Step 1 *It will be proved that, given a fixed state x_0, there exists a stationary policy which is ε-optimal at x_0, see* Lemma 7.1 *below.*

Step 2 *It will be shown in* Lemma 7.2 *that, for an appropriate subset E of S containing x_0, the action sets at states in E can be reduced to singletons without altering 'substantially' the optimal value function.*

Finally, *The proof of* Theorem 7.1 *is obtained by the successive application of this reduction process.*

Lemma 7.1 *Consider an MDP satisfying* Assumptions 2.1 *and* 2.2, *where the state space is an arbitrary denumerable set. Let $x_0 \in S$, the risk sensitivity coefficient $\lambda \neq 0$, and $\varepsilon > 0$, be fixed. Then there exists a stationary policy f which is ε-optimal at x_0:*

$$V_\lambda(f,x_0) \geq V^*_\lambda(x_0) - \varepsilon.$$

A proof of this Lemma, extending well-known ideas in risk-neutral dynamic programming to the risk sensitive context, was provided in Cavazos–Cadena and Montes–de–Oca [5]. For completeness, a short outline is given.

Proof of Lemma 7.1 Refer to (2.6) and select a policy $\pi \in \mathcal{P}$ such that

$$V_\lambda(\pi,x_0) + \frac{\varepsilon}{2} > V^*_\lambda(x_0). \tag{7.3}$$

Consider a finite set $G \subset S$ containing x_0, and observe that $T_{G^c} \nearrow \infty$ as $G \nearrow S$, where $G^c = S \setminus G$. Since

$$U_\lambda\left(V_\lambda(\pi,x_0)\right) = E^\pi_{x_0}\left[U_\lambda\left(\sum_{t=0}^{\infty} R(X_t,A_t)\right)\right],$$

the continuity and monotonicity of the utility function together imply, via the dominated convergence theorem, that

$$U_\lambda\left(V_\lambda(\pi, x_0)\right) = \lim_{G \nearrow S} E_{x_0}^{\pi}\left[U_\lambda\left(\sum_{t=0}^{T_{G^c}-1} R(X_t, A_t)\right)\right].$$

Then by (7.3), for some finite set G containing x_0, the following inequality holds,

$$E_{x_0}^{\pi}\left[U_\lambda\left(\sum_{t=0}^{T_{G^c}-1} R(X_t, A_t)\right)\right] > U_\lambda\left(V_\lambda^*(x_0) - \frac{\varepsilon}{2}\right). \tag{7.4}$$

Consider the new MDP $\widetilde{M} = (\widetilde{S}, A, \{\widetilde{A}(x)\}, \widetilde{R}, \widetilde{P})$ specified as follows.

For some object Δ outside of S, $\widetilde{S} = G \cup \{\Delta\}$, $\widetilde{A}(x) = A(x)$ (resp.$=$ $\{\Delta\}$) when $x \in G$ (resp., when $x = \Delta$). On the other hand, the transition law $\widetilde{P} = [\tilde{p}_{x\,y}(\cdot)]$ is given as follows. For $x \in G$ and $a \in A(x)$, $\tilde{p}_{x\,y}(a) = p_{x,y}(a)$ if $y \in G$ and $\tilde{p}_{x\,\Delta}(a) = 1 - \sum_{y \notin G} p_{x,y}(a)$ if $y = \Delta$; whereas $\tilde{p}_{\Delta,\Delta}(\cdot) = 1$. Finally, $\widetilde{R}(x,\cdot) = R(x,\cdot)$ for $x \in G$, and $\widetilde{R}(\Delta,\cdot) = 0$. Models M and $\widetilde{M}$ are closely related. In fact, starting at $x \in G$, the transitions and the reward streams of both models are identical as soon as the state stays in G, but once the system falls outside G, in model $\widetilde{M}$ the state remains equal to Δ and a null reward is earned forever. Given the policy π in (7.4), an admissible policy $\tilde{\pi}$ for $\widetilde{M}$ can be constructed as follows. If $\tilde{h}_t = (\tilde{x}_0, \tilde{a}_0, \ldots, \tilde{x}_t)$ is an admissible history up to time t for model $\widetilde{M}$, then $\tilde{\pi}_t(\cdot|\tilde{h}_t) = \pi_t(\cdot|\tilde{h}_t)$ if $\tilde{x}_s \in G$ for all $s \le t$, whereas $\tilde{\pi}_t(\{\Delta\}|\tilde{h}_t) = 1$ if $x_s = \Delta$ for some $s \le t$. Since $x_0 \in G$, from the relation between M and $\widetilde{M}$, it is not dificult to verify that $\widetilde{V}_\lambda(\tilde{\pi}, x_0)$, the λ-sensitive expected total-reward at x_0 corresponding to $\tilde{\pi}$, satisfies

$$U_\lambda\left(\widetilde{V}_\lambda(\tilde{\pi}, x_0)\right) = E_{x_0}^{\pi}\left[U_\lambda\left(\sum_{t=0}^{T_{G^c}-1} R(X_t, A_t)\right)\right]. \tag{7.5}$$

So that, by (7.4) and the strict monotonicity of $U_\lambda(\cdot)$,

$$\widetilde{V}_\lambda^*(x_0) \ge V_\lambda(\tilde{\pi}, x_0) > V_\lambda^*(x_0) - \frac{\varepsilon}{2}, \tag{7.6}$$

where $\widetilde{V}_\lambda^*(\cdot)$ is the optimal value function of model $\widetilde{M}$. The latter MDP has a finite state space, so that Theorem 5.1 yields an admissible stationary policy $\tilde{f}$ for $\tilde{M}$ satisfying $\widetilde{V}_\lambda(\tilde{f},\cdot) > \widetilde{V}_\lambda^*(\cdot) - \varepsilon/2$. Then by (7.6),

$$\widetilde{V}_\lambda(\tilde{f}, x_0) > V_\lambda^*(x_0) - \varepsilon. \tag{7.7}$$

To conclude, let $g \in \mathbb{F}$ be arbitrary and define the stationary policy f by setting $f(x) = \tilde{f}(x)$ for $x \in G$, and $f(x) = g(x)$ if $x \in S \setminus G$. In this case, it is not difficult to verify that equation (7.5) holds with $\tilde{f}$ and f replacing $\tilde{\pi}$ and π, respectively, so that

$$\begin{aligned} U_\lambda\left(\widetilde{V}_\lambda(\tilde{f}, x_0)\right) &= E_{x_0}^f\left[U_\lambda\left(\sum_{t=0}^{T_{G^c}-1} R(X_t, A_t)\right)\right] \\ &\le E_{x_0}^f\left[U_\lambda\left(\sum_{t=0}^{\infty} R(X_t, A_t)\right)\right] \\ &= U_\lambda\left(V_\lambda(f, x_0)\right). \end{aligned}$$

Recall that $R(\cdot,\cdot) \geq 0$ and (2.5). Thus, $V_\lambda(f, x_0) \geq \widetilde{V}_\lambda(\tilde{f}, x_0)$, by the strict monotonicity of U_λ. Hence f is ε-optimal at x_0 for model M, by (7.7). ■

Lemma 7.2 *Let the risk-sensitivity coefficient be positive, and suppose that* Assumption 2.1 *and condition* (7.1) *are valid. Given $x_0 \in S$ and $\varepsilon \in (0,1)$, consider the following construction. For a stationary policy f and $E \subset S$, define the MDP*

$$\widehat{M} = (S, A, \{\widehat{A}(x)\}, R, P), \tag{7.8}$$

where

$$\widehat{A}(x) = \begin{cases} A(x), & \text{if } x \notin E, \\ \{f(x)\}, & \text{if } x \in E. \end{cases} \tag{7.9}$$

Let $\widehat{\mathcal{P}}$ and $\widehat{V}_\lambda^*(\cdot)$ be the class of admissible policies and the optimal value function for $\widehat{M}$, respectively, so that

$$\widehat{V}_\lambda^*(x) = \sup_{\pi \in \widehat{\mathcal{P}}} V_\lambda(\pi, x), \qquad x \in S.$$

In this case, the policy f and the set E can be chosen so that the following assertions (i)—(iii) are valid.

(i) $x_0 \in E$,

(ii) $\widehat{V}_\lambda^*(\cdot) \geq V_\lambda^*(\cdot) - \varepsilon$.

(iii) Moreover, for every policy $\pi \in \widehat{\mathcal{P}}$,

$$V_\lambda(\pi, x_0) \geq V_\lambda^*(\cdot) - \varepsilon.$$

Proof Given $\delta \in (0,1)$, select a policy $f \in \mathbb{F}$ satisfying

$$V_\lambda(f, x_0) \geq (1 - \delta^2) V_\lambda^*(x_0). \tag{7.10}$$

When $V_\lambda^*(x_0) > 0$, the existence of such a policy follows from Lemma 7.1 with $\varepsilon = \delta^2 V_\lambda^*(x_0)$. Whereas if $V_\lambda^*(x_0) = 0$, then (2.6) yields that (7.10) is satisfied for every $f \in \mathbb{F}$. Define the set E by

$$E = \{x \in S \mid V_\lambda(f, x) \geq (1 - \delta)V_\lambda^*(x)\}. \tag{7.11}$$

It will be shown that the conclusions in the lemma are satisfied when f and E are selected in this way.

Assertion (i) Since $\delta \in (0, 1)$ and $V_\lambda^*(\cdot)$ is nonnegative, (7.10) implies that $x_0 \in E$.

Assertions (ii) and (iii) In the remainder of the proof, it will be shown that if δ is chosen appropriately assertions (ii) and (iii) are satisfied when model $\widehat{M}$ is constructed with f and E as in (7.10) and (7.11).

1. It will be proved that

$$\widehat{V}_\lambda(x) \geq V(x) - \delta\|V_\lambda^*\|, \quad x \in S. \tag{7.12}$$

To verify this assertion, notice that f is an admissible stationary policy for model $\widehat{M}$, so that, from the definition of E,

$$\widehat{V}_\lambda(x) \geq V_\lambda(f, x) \geq V_\lambda^*(x) - \delta\|V_\lambda^*\|, \qquad x \in E. \tag{7.13}$$

Since $\lambda > 0$ and recalling (2.1), the strong optimality equation in Lemma 2.2 yields,

$$e^{\lambda V_\lambda^*(x)} = \sup_{\pi \in \mathcal{P}} E_x^\pi \left[e^{\lambda \sum_{t=0}^{T_E - 1} R(X_t, A_t)} e^{\lambda V_\lambda^*(X_{T_E})} \right], \quad \forall x \in S. \tag{7.14}$$

As X_{T_E} belongs to the set E when $T_E < \infty$, using (7.11) and the fact that f belongs to the space of stationary policies for model $\widehat{M}$,

$$(1 - \delta)V_\lambda^*(X_{T_E}) \leq V_\lambda(f, X_{T_E}) \leq \widehat{V}_\lambda^*(X_{T_E}).$$

By the convention in Remark 2.2, this inequality remains valid when $T_E = \infty$. Therefore, (7.14) yields

$$\begin{aligned} & e^{\lambda V_\lambda^*(x)} \\ = \; & \sup_{\pi \in \mathcal{P}} E_x^\pi \left[e^{\lambda \sum_{t=0}^{T_E - 1} R(X_t, A_t)} e^{\lambda(1-\delta) V_\lambda^*(X_{T_E})} e^{\lambda \delta V_\lambda^*(X_{T_E})} \right] \\ \leq \; & \sup_{\pi \in \mathcal{P}} E_x^\pi \left[e^{\lambda \sum_{t=0}^{T_E - 1} R(X_t, A_t)} e^{\lambda \widehat{V}_\lambda^*(X_{T_E})} \right] e^{\lambda \delta \|V_\lambda^*(\cdot)\|}. \end{aligned} \tag{7.15}$$

Now suppose that $X_0 = x \notin E$, and notice

$$E_x^\pi\left[e^{\lambda\sum_{t=0}^{T_E-1} R(X_t,A_t)}e^{\lambda\widehat{V}_\lambda^*(X_{T_E})}\right]$$

depends only on the actions prescribed by π at times $k \in \{0,1,\ldots,T_E-1\}$, and that $X_k \notin E$ for these values of k. Since models M and $\widehat{M}$ have the same sets of admissible actions at the states in $E^c = S\setminus E$, the supremum over $\mathcal{P}$ after the inequality in (7.15) coincides with the supremum over $\widehat{\mathcal{P}}$. Thus, if $x \notin E$

$$\begin{aligned} e^{\lambda V_\lambda^*(x)} &\leq \sup_{\pi\in\mathcal{P}} E_x^\pi\left[e^{\lambda\sum_{t=0}^{T_E-1} R(X_t,A_t)}e^{\lambda\widehat{V}_\lambda^*(X_{T_E})}\right]e^{\lambda\delta\|V_\lambda^*(\cdot)\|} \\ &= \sup_{\pi\in\widehat{\mathcal{P}}} E_x^\pi\left[e^{\lambda\sum_{t=0}^{T_E-1} R(X_t,A_t)}e^{\lambda\widehat{V}_\lambda^*(X_{T_E})}\right]e^{\lambda\delta\|V_\lambda^*(\cdot)\|} \\ &= e^{\lambda[\widehat{V}_\lambda^*(x)+\delta\|V_\lambda^*(\cdot)\|]}, \end{aligned}$$

where the second equality stems from Lemma 2.2 applied to model $\widehat{M}$. Thus, since $\lambda > 0$, $V_\lambda^*(x) \leq \widehat{V}_\lambda^*(x) + \delta\|V_\lambda^*(\cdot)\|$ if $x \notin E$, which combined with (7.13) establishes (7.12).

2. As in Section 3, consider the MDP

$$M_f = (S, A, \{A_f(x)\}, R, P)$$

obtained by setting $A_f(x) = \{f(x)\}$, for which $V_\lambda(f,\cdot)$ is the corresponding optimal value function. Applying Lemma 2.2 to this model with the stopping time T_{E^c}, and using the condition $\lambda > 0$, it follows that

$$e^{\lambda V_\lambda(f,x)} = E_x^f\left[e^{\lambda\sum_{t=0}^{T_{E^c}-1} R(X_t,A_t)}e^{\lambda V_\lambda(f,X_{T_{E^c}})}\right], \quad x \in S. \tag{7.16}$$

Now observe that $V_\lambda(f, X_{T_{E^c}}) \leq (1-\delta)V_\lambda^*(X_{T_{E^c}})$, which follows from the definition of E when $T_{E^c} < \infty$, and from the convention in Remark 2.2 for $T_{E^c} = \infty$. Therefore, using that

$\lambda > 0$, (7.10) and (7.16) together imply

$$
\begin{aligned}
& e^{\lambda(1-\delta^2)V_\lambda^*(x_0)} \\
\le \;& E_{x_0}^f\left[e^{\lambda\sum_{t=0}^{T_{E^c}-1}R(X_t,A_t)}e^{\lambda(1-\delta)V_\lambda^*(X_{T_{E^c}})}\right] \\
= \;& E_{x_0}^f\left[e^{\lambda(1-\delta)\left[\sum_{t=0}^{T_{E^c}-1}R(X_t,A_t)+V_\lambda^*(X_{T_{E^c}})\right]}\right. \\
& \qquad \left. e^{\lambda\delta\sum_{t=0}^{T_{E^c}-1}R(X_t,A_t)}\right] \\
\le \;& \left(E_{x_0}^f\left[e^{\lambda\sum_{t=0}^{T_{E^c}-1}R(X_t,A_t)+\lambda V_\lambda^*(X_{T_{E^c}})}\right]\right)^{(1-\delta)} \\
& \qquad \left(E_{x_0}^f\left[e^{\lambda\sum_{t=0}^{T_{E^c}-1}R(X_t,A_t)}\right]\right)^{\delta} \\
& \text{By Hölder's inequality} \\
\le \;& e^{\lambda(1-\delta)V_\lambda^*(x_0)}\left(E_x^f\left[e^{\lambda\sum_{t=0}^{T_{E^c}-1}R(X_t,A_t)}\right]\right)^{\delta},
\end{aligned}
$$

where the last inequality comes from the strong optimality equation. Consequently, since $\delta \in (0,1)$, it follows that

$$
e^{\lambda(1-\delta)V_\lambda^*(x_0)} \le E_{x_0}^f\left[e^{\lambda\sum_{t=0}^{T_{E^c}-1}R(X_t,A_t)}\right]. \tag{7.17}
$$

Alternatively, the definition of the action sets $\widehat{A}(x)$ yields that a policy π in $\widehat{\mathcal{P}}$ prescribes the same actions as f on the set E. Since $x_0 \in E$, $E_{x_0}^\pi\left[e^{\lambda\sum_{t=0}^{T_{E^c}-1}R(X_t,A_t)}\right]$ depends only on the actions selected while the state of system remains in E, so that

$$
E_{x_0}^\pi\left[e^{\lambda\sum_{t=0}^{T_{E^c}-1}R(X_t,A_t)}\right] = E_{x_0}^f\left[e^{\lambda\sum_{t=0}^{T_{E^c}-1}R(X_t,A_t)}\right].
$$

Then, the positivity of λ and the nonnegativity of the reward function together yield that

$$
\begin{aligned}
e^{\lambda V_\lambda(\pi,x_0)} &= E_{x_0}^\pi\left[e^{\lambda\sum_{t=0}^{\infty}R(X_t,A_t)}\right] \\
&\ge E_{x_0}^\pi\left[e^{\lambda\sum_{t=0}^{T_{E^c}-1}R(X_t,A_t)}\right] \\
&= E_{x_0}^f\left[e^{\lambda\sum_{t=0}^{T_{E^c}-1}R(X_t,A_t)}\right] \\
& \text{by (7.17)} \\
&\ge e^{\lambda(1-\delta)V_\lambda^*(x_0)}.
\end{aligned}
$$

Thus,

$$\begin{aligned} V_\lambda(\pi, x_0) &\geq (1-\delta)V_\lambda^*(x_0) \\ &\geq V_\lambda^*(x_0) - \delta\|V_\lambda^*(\cdot)\|, \quad \pi \in \widehat{\mathcal{P}}. \end{aligned} \tag{7.18}$$

3. To conclude, set

$$\delta = \frac{\varepsilon}{1 + \|V_\lambda(\cdot)\|},$$

and observe that assertions (ii) and (iii) follow from (7.12) and (7.18). ■

Proof of Theorem 7.1 Given $\varepsilon \in (0,1)$, set $\varepsilon_n = \dfrac{\varepsilon}{2^{n+1}}$ and write

$$S = \{x_0, x_1, x_2, \ldots\}. \tag{7.19}$$

Consider now the following recursive construction of a sequence

$$\{M_n = (S, A, \{A_n(x)\}, R, P)\}.$$

1. Set $M_0 = M$, the original MDP described in Section 2.
2. Given M_k with $k \geq 0$, let M_{k+1} be the MDP $\widehat{M}$ constructed in Lemma 7.2 with M_k, x_k and ε_k instead of M, x_0 and ε respectively. Then by Lemma 7.2, the following assertions (a)–(d) are valid for every $k \in \mathbb{N}$.

    a) $A_{k+1}(x) \subset A_k(x)$ for every $x \in S$;

    b) $A_{k+1}(x_k)$ is a singleton.

    c) If $\mathcal{P}_k$ and $V_{\lambda,k}^*(\cdot)$ denote the class of admissible policies and the optimal value function of M_k, respectively, then

    $$V_{\lambda,k+1}^*(\cdot) \geq V_{\lambda,k}^*(\cdot) - \varepsilon_k, \quad \text{and}$$

    d) for each policy $\pi \in \mathcal{P}_{k+1}$, $V_\lambda(\pi, x_k) \geq V_{\lambda,k}^*(x_k) - \varepsilon_k$.

    By a simple induction argument, (c) clearly implies that

$$\begin{aligned} V_{\lambda,k}^*(\cdot) &\geq V_{\lambda,0}^*(\cdot) - \sum_{n=0}^{k-1} \varepsilon_n \\ &= V_\lambda^*(\cdot) - \sum_{n=0}^{k-1} \varepsilon_n, \quad k \in \mathbb{N}; \end{aligned} \tag{7.20}$$

For the equality, recall that $M_0 = M$. On the other hand, from (a) and (b) it follows that for each $x \in S = \{x_0, x_1, x_2, \ldots\}$, the intersection of the action sets $A_n(x)$ is a singleton, say

$$\bigcap_{n=0}^{\infty} A_n(x) = \{a_x\}.$$

Define $f(x) = a_x, \quad x \in S$, and observe that the inclusion $f(x) \in A_k(x)$ always holds, so that f is an admissible stationary policy for each model M_n.

To conclude, let $x \in S$ be arbitrary. In this case $x = x_k$ for some $k \in \mathbb{N}$ by (7.19), and assertion (d) implies that $V_\lambda(f, x_k) \geq V^*_{\lambda,k}(x_k) - \varepsilon_k$, which combined with (7.20) yields

$$\begin{aligned} V_\lambda(f, x) &= V_\lambda(f, x_k) \\ &\geq V^*_\lambda(x_k) - \sum_{n=0}^{k} \varepsilon_n \\ &> V^*_\lambda(x_k) - \varepsilon \\ &= V^*_\lambda(x) - \varepsilon. \end{aligned}$$

Hence f is ε-optimal. ∎

8. Conclusion

Under the basic structural assumption that the controller has a constant risk-sensitivity coefficient, this work considered MDPs with discrete state space and nonnegative rewards. When the performance index of a control policy is the risk-sensitive expected total-reward criterion, the paper addressed the existence of optimal and ε-optimal stationary policies assuming the finiteness of the optimal value function and the mild measurability condition in Assumption 2.1(i). In contrast with the usual approach to this problem, based on the discounted criterion, the arguments used in this work rely on the comparison of the optimal value functions associated to MDPs whose action spaces coincide, with the exception of a single state; see Theorems 3.1 and 4.1. Therefore, the paper faced a problem on the expected total-reward criterion, *entirely* within the framework of this performance index.

After establishing the existence of optimal stationary policies for MDPs with finite state and actions sets in Section 5, the result was used in Section 6 to obtain ε-optimal stationary policies, when only the state space is supposed to be finite, which are valid regardless of the sign of the risk-sensitivity coefficient. In Section 7 it was supposed that the decision maker is risk-seeking, and when the optimal value function is bounded, it was proved in Theorem 7.1 that the existence of an ε-optimal stationary policy is guaranteed, complementing results recently obtained for the risk-averse case; see Remark 7.1.

However, extending the result in Theorem 7.1 to the case of an unbounded optimal value function seems to be an interesting problem, and research on this direction is presently in progress.

References

[1] Ávila–Godoy, M.G. (1998). Controlled Markov Chains with Exponential Risk–Sensitive Criteria: Modularity, Structured Policies and Applications, Ph. D. Dissertation, Department of Mathematics, The University of Arizona, Tucson, AZ, USA.

[2] Brau–Rojas, A. (1998). Controlled Markov Chains with Risk–Sensitive Average Cost Criterion, Ph. D. Dissertation, Department of Mathematics, The University of Arizona, Tucson, AZ, USA.

[3] Cavazos–Cadena R., Fernández–Gaucherand E. (1999). Controlled Markov Chains with Risk-Sensitive Criteria: Average Cost, Optimality Equations, and Optimal Solutions, *Mathematical Methods of Operations Research*, **49**, 299–324.

[4] Cavazos–Cadena R., Montes–de–Oca, R. (2000). Optimal Stationary Policies in Risk–Sensitive Dynamic Programs with Finite State Space and Nonnegative Rewards. *Applicationes Mathematicae*, **27**, 167–185.

[5] Cavazos–Cadena R., Montes–de–Oca, R. (2000). Nearly Optimal Policies in Risk–Sensitive Positive Dynamic Programming on Discrete Spaces, *Mathematical Methods of Operations Research*, **52**, 133–167.

[6] Cavazos–Cadena R., Montes–de–Oca, R. (2001). Existence of Optimal Stationary Policies in Finite Dynamic Programs with Nonnegative Rewards: An Alternative Approach. *Probability in Engineering and Information Sciences*, **15**, 557–564.

[7] Fishburn, P.C. (1970). *Utility Theory for Decision Making*, Wiley, New York.

[8] Flemming, W.H., Hernández–Hernández D. (1997). Risk sensitive control of finite machines on an infinite horizon I, *SIAM Journal on Control and Optimization*, **35**, No. 5.

[9] Hernández-Lerma, O. (1989). *Adaptive Markov Control Processes*, Springer–Verlag, New York.

[10] Hinderer, K. (1970). *Foundations of Non-Stationary Dynamic Programming with Discrete Time Parameter*, Lecture Notes Oper. Res., **33**, Springer–Verlag, New York.

[11] Hordijk, A. (1974). *Dynamic Programming and Markov Potential Theory, Mathematical Centre Tracts*, No. 51, Amsterdam.

[12] Ornstein, D. (1969). On the existence of stationary optimal strategies, *Proceedings of the American Mathematical Society*, **20**, 563–569.

[13] Pratt, J.W. (1964). Risk aversion in the small and in the large, *Econometrica*, **32**, No. 1, 122–136.

[14] Puterman, M.L. (1994). *Markov Decision Processes*, Wiley, New York.

[15] Ross, S.M. (1970). *Applied Probability Models with Optimization Applications*, Holden–Day, San Francisco.

[16] Strauch, R. (1966). Negative Dynamic Programming, *Annals of Mathematical Statistics*, **37**, 871–890.

[17] Thomas L.C. (1980). Connectedness conditions for denumerable state Markov decision processes, in: R. Hartley, L.C. Thomas and D.J. White, Editors, *Recent Advances in Markov Decision Processes*, Academic press, New York.

Chapter 12

INTERVAL METHODS FOR UNCERTAIN MARKOV DECISION PROCESSES

Masami Kurano
Chiba University,
Chiba, Japan
kurano@math.e.chiba-u.ac.jp

Masami Yasuda
Chiba University,
Chiba, Japan
yasuda@math.s.chiba-u.ac.jp

Jun-ichi Nakagami
Chiba University,
Chiba, Japan
nakagami@math.s.chiba-u.ac.jp

Abstract In this paper, interval methods for uncertain Markov decision processes are considered. That is, a controlled Markov set-chain model with a finite state is developed by an interval arithmetic analysis, and we will find Pareto optimal policies which maximize the discounted or average expected rewards over all stationary policies under some partial order. The optimal policies are characterized by a maximal solution of an optimality equation including efficient set function.

Keywords: Controlled Markov set-chain, discounted reward, average reward, Pareto optimal, interval arithmetic.

Z. Hou et al. (eds.), Markov Processes and Controlled Markov Chains, 223–232.

1. Introduction and notation

In a real application of Markov decision processes (MDPs in short, see [7, 10, 16]), the required data must be estimated. The mathematical model of MDPs can only be viewed as approximations. It may be useful that the model is ameliorated so to be "robust" in the sense that it's reasonably efficient in rough approximations. How can be this situation modelled? One realistic answer to such a problem is to apply certain intervals containing the required data.

By Hartfiel's [4, 5, 6] interval method, Kurano et al. [13] has introduced a decision model, called a controlled Markov set-chain, which is robust for approximation of the transition matrix in MDPs. The discounted reward problem was developed in [12, 13]. The non-discounted case was treated in Hosaka et al. [8] and the average reward problem under contractive properties was studied Hosaka et al. [9]. However, the functional characterization of optimal policies is not given.

In this paper, applying an interval arithmetic analysis, we develop the functional characterization of Pareto optimal policies which maximize the discounted or average expected rewards over all stationary policies under some partial order.

In the remainder of this section, we shall introduce several notions referring to the works [4, 5, 6] on Markov set-chain. Refer [15] and [12, 13] for the interval arithmetic and formulation of a controlled Markov set-chain respectively.

Let R_+^{m*n} be the set of entry-wise non-negative $m*n$-matrix ($m, n \geq 1$). For any $\underline{B}, \overline{B} \in R_+^{m*n}$ with $\underline{B} \geq \overline{B}$ (component-wise), we denote by $< \underline{B}, \overline{B} >$ the set of stochastic matrices B such that $\underline{B} \geq B \geq \overline{B}$.

The set of all bounded and closed intervals on the non-negative numbers is denoted by $C(R_+)$, and $C(R_+)^n$ is the set of all n-dimensional column vectors whose elements are in $C(R_+)$, i.e.,

$$C(R_+)^n := \{D = (D_1, D_2, \ldots, D_n)' \mid D_i \in C(R_+)(1 \leq i \leq n)\}.$$

where d' denotes the transpose of a vector d.
If $D = ([\underline{d}_1, \overline{d}_1], \ldots, [\underline{d}_n, \overline{d}_n])'$, then it will be denoted by $D = [\underline{d}, \overline{d}]$, where $\underline{d} = (\underline{d}_1, \ldots, \underline{d}_n)'$, $\overline{d} = (\overline{d}_1, \ldots, \overline{d}_n)'$ and $[\underline{d}, \overline{d}] = \{d \in R_+^n \mid \underline{d} \leq d \leq \overline{d}\}$.

We will give a partial order $\succ, \succeq$ on $C(R_+)$ by the definition:
For $[c_1, c_2], [d_1, d_2] \in C(R_+)$,

- $[c_1, c_2] \succeq [d_1, d_2]$ if $c_1 \geq d_1$, $c_2 \geq d_2$, and
- $[c_1, c_2] \succ [d_1, d_2]$ if $[c_1, c_2] \succeq [d_1, d_2]$ and $[c_1, c_2] \neq [d_1, d_2]$.

For $\mathbf{v} = (v_1, v_2, \ldots, v_n)'$ and $\mathbf{w} = (w_1, w_2, \ldots, w_n)' \in C(R_+)^n$, we use the notation:

- $\mathbf{v} \succeq \mathbf{w}$ if $v_i \succeq w_i$, $1 \leq i \leq n$, and
- $\mathbf{v} \succ \mathbf{w}$ if $\mathbf{v} \succeq \mathbf{w}$ and $\mathbf{v} \neq \mathbf{w}$.

A controlled Markov set-chain consists of five objects:

$$(S, A, \underline{q}, \overline{q}, r),$$

where $S = \{1, 2, \ldots, n\}$ and $A = \{1, 2, \ldots, k\}$ are finite sets and for each $(i, a) \in S \times A$, $\underline{q} = \underline{q}(\cdot|i, a) \in R_+^{1\times n}$, $\overline{q} = \overline{q}(\cdot|i, a) \in R_+^{1\times n}$ with $\underline{q} \leq \overline{q}$, $\langle \underline{q}, \overline{q} \rangle \neq \emptyset$, and $r = r(i, a)$ a function on $S \times A$ with $r \geq 0$. Note that the notation used here obey the previous one ([8, 9]). We interpret S as the set of states of some system, and A as the set of actions available at each state.

When the system is in state $i \in S$ and we take action $a \in A$, we move to a new state $j \in S$ selected according to the probability distribution on S, $q(\cdot|i, a)$, and we receive an immediate return, $r(i, a)$, where we know only that $q(\cdot|i, a)$ is arbitrarily chosen from $\langle \underline{q}(\cdot|i, a), \overline{q}(\cdot|i, a) \rangle$. This process is then repeated from the new state j. Denote by F the set of functions from S to A.

A policy π is a sequence $(f_1, f_2, \ldots)$ of functions with $f_t \in F$, $(t \geq 1)$. Let Π denote the class of policies. We denote by f^∞ the policy $(h_1, h_2, \ldots)$ with $h_t = f$ for all $t \geq 1$ and some $f \in F$. Such a policy is called stationary, denoted simply by f, and the set of stationary policies is denoted by F.

We associate with each $f \in F$ the n-dimensional column vector $r(f) \in R_+^n$ whose i^{th} element is $r(i, f(i))$ and the set of stochastic matrices $\mathcal{Q}(f) := \langle \underline{Q}(f), \overline{Q}(f) \rangle$ where the (i, j) elements of $\underline{Q}(f)$ and $\overline{Q}(f)$ are $\underline{q}(j|i, f(i))$ and $\overline{q}(j|i, f(i))$ respectively.

First, we define the set of discounted total expected rewards. For any $\pi = (f_1, f_2, \cdots) \in \Pi$, and discount factor β $(0 < \beta < 1)$, let

$$\phi_T(\pi) := \left\{ r(f_1) + \sum_{i=1}^{T} \beta^i Q_1 Q_2 \cdots Q_i r(f_{i+1}) \mid \forall i, Q_i \in \mathcal{Q}(f_i) \right\}. \quad (1.1)$$

Since it is shown in [13] that $\{\phi_T(\pi)\}_{T=1}^{\infty}$ is a Cauchy sequence with respect to a metric on $C(R_+)^n$, the set of discounted expected total rewards from π in the infinite future can be defined by

$$\phi(\pi) := \lim_{T\to\infty} \phi_T(\pi) = [\underline{\phi}(\pi), \overline{\phi}(\pi)] \in C(R_+)^n, \quad (1.2)$$

where $\underline{\phi}(\pi), \overline{\phi}(\pi) \in R_+{}^n$.

Now, we define the set of average expected rewards. For any $\pi = (f_1, f_2, \ldots) \in \Pi$, let $\mathbf{v}_1(\pi) = r(f_1)$ and, by setting $Q_0 =$ identity,

$$\mathbf{v}_T(\pi) = \left\{ r(f_1) + \sum_{i=1}^{T} Q_1 Q_2 \cdots Q_i r(f_{i+1}) \mid \forall i, Q_i \in \mathcal{Q}(f_i) \right\} \tag{1.3}$$

for $T \geq 2$. It holds that $\mathbf{v}_T(\pi) \in C(R_+)^n$ for all $T \geq 1$. Let

$$\begin{aligned} \mathbf{v}(\pi) &:= \liminf_{T \to \infty} \mathbf{v}_T(\pi)/T & (1.4) \\ &= \left\{ x \in R^n \mid \limsup_{T\to\infty} \inf_{y \in \mathbf{v}_T(\pi)/T} \delta(x, y) = 0 \right\} \end{aligned}$$

where δ is a metric in R^n. Since $\mathbf{v}(\pi) \in C(R_+)^n$ ([13]), $\mathbf{v}(\pi)$ is written as $\mathbf{v}(\pi) = [\underline{v}(\pi), \overline{v}(\pi)]$. As the meaning of the values in (1.2) of a discounted case, (1.4) of an average case, they are the expected rewards under the corresponding behaviour in the worst or in the best respectively.

Definition 1.1 *A policy $f^* \in \Pi_F$ is called* discounted (average) optimal *if and only if for each $i \in S$, there does not exist $f \in F$ such that*

$$\phi(f^*)_i \prec \phi(f)_i \quad (\mathbf{v}(f^*)_i \prec \mathbf{v}(f)_i)\,. \tag{1.5}$$

where $\phi(f)_i$ $(\mathbf{v}(f)_i)$ is the i^{th} element of $\phi(f)$ $(\mathbf{v}(f))$.

In the above definition, we confine ourselves to the stationary policies, which simplifies our discussion in the sequel. In Section 2, discounted optimal policies are characterized by maximal solutions of optimality equation. The characterization of average optimal policies is done in Section 3.

2. Optimality for the discount case

In this section, we derive the optimality equation, by which discounted optimal policies are characterized. Associated with each $f \in F$ and $\beta \in (0, 1)$ is a corresponding operator $L(f)$, a mapping from $C(R_+)^n$ into itself, defined as follows. For $\boldsymbol{v} \in C(R_+)^n$,

$$L(f)\mathbf{v} := r(f) + \beta \mathcal{Q}(f)\mathbf{v} = \left[\underline{L}(f)\underline{v}, \, \overline{L}(f)\overline{v}\right] \in C(R_+)^n. \tag{2.1}$$

Note that $\mathbf{v} = [\underline{v}, \overline{v}]$ with $\underline{v} \leq \overline{v}$, $\underline{v}, \overline{v} \in R^n_+$, and $\underline{L}$ and $\overline{L}$ are operators from R^n_+ into R^n_+, defined by:

$$\begin{cases} \underline{L}(f)\mathbf{v} = r(f) + \beta \min_{Q \in \mathcal{Q}(f)} Q\mathbf{v}, \\ \overline{L}(f)\mathbf{v} = r(f) + \beta \max_{Q \in \mathcal{Q}(f)} Q\mathbf{v}. \end{cases} \tag{2.2}$$

In (2.2), each min/max represents component-wise minimization/maximizing. The following results are given in Kurano et al. [13].

Lemma 2.1 ([13]) *For any $f \in F$, we have:*

(i) $L(f)$ is monotone and contractive with modulus β, and $\phi(f)$ is a unique fixed point of $L(f)$, that is,

$$\phi(f) = L(f)\phi(f). \tag{2.3}$$

(ii) For any $h \in R_+^n$, $\phi(f) = \lim_{t\to\infty} L(f)^t h$ with respect to δ.

We have the following.

Lemma 2.2 *For $f, g \in F$, suppose that $\phi(f) \prec L(g)\phi(f)$. Then, it holds $\phi(f) \prec \phi(g)$.*

Proof By Lemma 2.1(i), we have that

$$\phi(f) \prec L(g)\phi(f) \preceq L(g)^t\phi(f), \quad \text{for all } t \geq 2.$$

By $t \to \infty$ in the above, from Lemma 2.1(ii) it follows that $\phi(f) \prec \phi(g)$, as required. ■

Let $\mathbf{q}(i,a) := \langle \underline{q}(\cdot|i,a), \overline{q}(\cdot|i,a)\rangle$ for $i \in S$ and $a \in A$. When $f \equiv a$ for some $a \in A$, the operator $L(f)$ will be denoted by L_a, that is, for $\mathbf{v} \in C(R_+)^n$ and $i \in S$, $(L_a\mathbf{v})_i = r(i,a) + \beta\mathbf{q}(i,a)\mathbf{v}$, and $L_a\mathbf{v} = ((L_a\mathbf{v})_1, \ldots, (L_a\mathbf{v})_n)$. For any $D \subset C(R_+)$, a point $u \in D$ is called an efficient element of D with respect to $\preceq$ on $C(R_+)$ if and only if it holds that there does not exist $v \in D$ such that $u \prec v$. We denote by eff(D) the set of all efficient elements of D. Let

$$\begin{aligned} \mathcal{L}(\mathbf{u})(i) &= \text{eff}\big(\{(L_a\mathbf{u})_i | a \in A\}\big), \quad i \in S = \{1, 2, \ldots, n\}, \\ \mathcal{L}(\mathbf{u}) &:= \big(\mathcal{L}(\mathbf{u})(1), \mathcal{L}(\mathbf{u})(2), \ldots, \mathcal{L}(\mathbf{u})(n)\big) \end{aligned}$$

for any $\mathbf{u} \in C(R_+)^n$. We note that $\mathcal{L}(\mathbf{u}) \subset C(R_+)^n$ holds.

Here, let us consider the following interval equation including efficient set-function $\mathcal{L}$ on $C(R_+)^n$: Find $\mathbf{u} \in C(R_+)^n$ such that

$$\mathbf{u} \in \mathcal{L}(\mathbf{u}). \tag{2.4}$$

The equation (2.4) may be called an optimality equation for the discounted case in this formulation of our model, by which discounted optimal policies are characterized. A solution $\mathbf{u}$ of the optimality equation is called maximal if at each $i \in S$ there does not exist any solution $\mathbf{v}$ such that $u_i \preceq v_i$, where $\mathbf{u} = (u_1, \ldots, u_n)$ and $\mathbf{v} = (v_1, \ldots, v_n)$.

In Theorem 2.1 below, discounted optimal policies are characterized by maximal solutions of the optimality equation. Lemmas 2.1 and 2.2 make the proof of Theorem 2.1 possible and the proof can be entirely done analogously to that of [3, Theorem 5.1] through a simple modification. So the proof is omitted.

Theorem 2.1 *A policy $f \in F$ is discounted optimal if and only if $\phi(f)$ is a maximal solution to the optimality equation.*

3. Optimality for the average case

In this section, we will give the optimality equation for the average case. Henceforth, the following assumption will remain operative.

Assumption 3.1 (Primitivity) *For any $f \in F$, each $Q \in \mathcal{Q}(f)$ is primitive, i.e., $Q^t > 0$ for some $t \geq 1$.*

Obviously, if $\underline{Q}(f)$ is primitive in the sense of non-negative matrix (see [17]), Assumption 3.1 holds.

The following facts on Markov matrices are well-known (see [2, 11]).

Lemma 3.1 *For any $f \in F$, let Q be any matrix in $\mathcal{Q}(f)$.*

(i) The sequence $(I + Q + \cdots + Q^t)/(t+1)$ converges as $t \to \infty$ to a stochastic matrix Q^* with $Q^*Q = Q^*$, $Q^* > 0$ and rank$(Q^*) = 1$.

(ii) The matrix Q^* in (i) is uniquely determined by $Q^*Q = Q$ and rank$(Q^*) = 1$.

Associated with each $f \in F$ is a corresponding operator $U(f)$, mapping $C(R_+)^n$ into $C(R_+)^n$, defined as follows.

For $\mathbf{v} = [\underline{v}, \overline{v}] \in C(R_+)^n$ with $\underline{v} \leq \overline{v}$, $\underline{v}, \overline{v}$,

$$U(f)\mathbf{v} := r(f) + \mathcal{Q}(f)\mathbf{v} = \left[\underline{U}(f)\underline{v}, \overline{U}(f)\overline{v}\right]. \tag{3.1}$$

where $\underline{U}$ and $\overline{U}$ are operators from R^n into itself, defined by:

$$\begin{cases} \underline{U}(f)v &= r(f) + \min_{Q \in \mathcal{Q}(f)} Qv, \\ \overline{U}(f)v &= r(f) + \max_{Q \in \mathcal{Q}(f)} Qv. \end{cases} \tag{3.2}$$

Let $\mathbf{e} := (1, 1, \ldots, 1)'$. Here, for any $f \in F$, we consider the interval equation:

$$r(f) + \mathcal{Q}(f)\mathbf{h} = \mathbf{v} + \mathbf{h}, \tag{3.3}$$

where $\mathbf{v} := [\underline{v}\mathbf{e}, \overline{v}\mathbf{e}]$, $\underline{v}, \overline{v} \in R$ and $\mathbf{h} = [\underline{h}, \overline{h}] \in C(R)^n$, $\underline{h}, \overline{h} \in R^n$ with $\underline{v} \leq \overline{v}$, $\underline{h} \leq \overline{h}$.

Obviously, the interval equation can be rewritten by their extremal points as

$$\begin{cases} r(f) + \min_{Q \in \mathcal{Q}(f)} Q\underline{h} &= \underline{v}\mathbf{e} + \underline{h} \\ r(f) + \max_{Q \in \mathcal{Q}(f)} Q\overline{h} &= \overline{v}\mathbf{e} + \overline{h} \end{cases} \tag{3.4}$$

with $\underline{v} \le \overline{v}$, $\underline{h} \le \overline{h}$ where $\underline{v}, \overline{v} \in R$, $\underline{h}, \overline{h} \in R^n$.

We have the following lemma.

Lemma 3.2 ([1, 8]) *For any $f \in F$, the interval equation* (3.3) *determines* $\mathbf{v}$ *uniquely and* $\mathbf{h}$ *up to an additive constant* $[c_1\mathbf{e}, c_2\mathbf{e}]$ *with* $c_1, c_2 \in R(c_1 < c_2)$.

Since the unique solutions $\mathbf{v}$ and $\mathbf{h}$ of (3.3) are dependent on $f \in F$, we will denote them respectively by $\mathbf{v}(f) := \mathbf{v}$ and $\mathbf{h}(f) := \mathbf{h}$. The following lemma can be proved similarly to [9, Corollary 3.1].

Lemma 3.3 ([9]) *For any $f \in F$, it holds that:*

(i) $\mathbf{v}(f) = [\underline{v}(f)\mathbf{e}, \overline{v}(f)\mathbf{e}]$.

(ii) $\underline{v}(f)\mathbf{e} = \min_{Q \in \mathcal{Q}(f)} Q^* r(f)$ *and* $\overline{v}(f)\mathbf{e} = \max_{Q \in \mathcal{Q}(f)} Q^* r(f)$.

Lemma 3.4 *For any f, g in F, suppose that*

$$\mathbf{v}(f) + \mathbf{h}(f) \left\{ \begin{array}{c} \preceq \\ = \\ \prec \end{array} \right\} r(g) + \mathcal{Q}(g)\mathbf{h}(f). \tag{3.5}$$

Then, it holds that

$$\mathbf{v}(f) \left\{ \begin{array}{c} \preceq \\ = \\ \prec \end{array} \right\} \mathbf{v}(g). \tag{3.6}$$

Proof The left and right extremal equation of (3.5) are given as follows.

$$\underline{v}(f)\mathbf{e} + \underline{h}(f) \quad \left\{ \begin{array}{c} \preceq \\ = \\ \prec \end{array} \right\} \quad r(g) + \min_{Q \in \mathcal{Q}(f)} Q\underline{h}(f) \tag{3.7}$$

$$\overline{v}(f)\mathbf{e} + \overline{h}(f) \quad \left\{ \begin{array}{c} \preceq \\ = \\ \prec \end{array} \right\} \quad r(g) + \max_{Q \in \mathcal{Q}(f)} Q\overline{h}(f). \tag{3.8}$$

By Lemma 3.3, there exists $\underline{Q} \in \mathcal{Q}(g)$ with $\underline{v}(g)\mathbf{e} = \underline{Q}^* r(g)$. Multiplying the both sides of (3.7) by $\underline{Q}^*$, we get from $\underline{Q}^*\underline{Q} = \underline{Q}^*$ and $\underline{Q}^* > 0$ that $\underline{v}(f)\mathbf{e} \left\{ \begin{array}{c} \preceq \\ = \\ \prec \end{array} \right\} \underline{Q}^* r(g)$. Thus $\underline{v}(f)\mathbf{e} \left\{ \begin{array}{c} \preceq \\ = \\ \prec \end{array} \right\} \underline{v}(g)\mathbf{e}$ follows. Similarly, we get $\overline{v}(f)\mathbf{e} \left\{ \begin{array}{c} \preceq \\ = \\ \prec \end{array} \right\} \overline{v}(g)\mathbf{e}$, which proves (3.6). ■

From Lemma 3.3, we observe that all elements of $\mathbf{v}(f)$ are equal, which implies that the set of average expected reward from $f \in F$ is independent of the initial state. So, a policy $f \in F$ is average optimal if and only if there is no $g \in F$ such that $\mathbf{v}(f) \prec \mathbf{v}(g)$. Keeping the above in mind, we can define an efficient point with respect to the partial order $\preceq$ on $C(R)^n$.

Let D be an arbitrary subset of $C(R)^n$. A point $\mathbf{u} \in D$ is called an efficient element of D with respect to $\preceq$ on $C(R)^n$ if and only if it holds that there does not exist $\mathbf{v} \in D$ such that $\mathbf{u} \prec \mathbf{v}$. We denote by eff$(D)$ the set of all elements of D efficient with respect to $\preceq$ on $C(R)^n$. For any $\mathbf{u} \in C(R)^n$, let

$$\mathcal{U}(\mathbf{u}) := \mathrm{eff}\Big(\{U(f)\mathbf{u} \mid f \in F\}\Big),$$

where $U(f)\mathbf{u} \in C(R)^n$ is defined in (3.1). We note that $\mathcal{U}(\mathbf{u}) \subset C(R)^n$ for any $\mathbf{u} \in C(R)^n$.

Here, we consider the following interval equations inducing efficient set-function $\mathcal{U}(\cdot)$ on $C(R)^n$.

$$\mathbf{v} + \mathbf{h} \in \mathcal{U}(\mathbf{h}), \tag{3.9}$$

where $\mathbf{v} = [\underline{v}\mathbf{e}, \overline{v}\mathbf{e}]$, $\mathbf{h} = [\underline{h}, \overline{h}] \in C(R)^n$ and $\underline{v} \le \overline{v}$, $\underline{h} \le \overline{h}$, $\underline{v}, \overline{v} \in R$, $\underline{h}, \overline{h} \in R^n$. The equation (3.9) is called *an optimality equation* for the average case, by which average optimal policies can be characterized. A solution $(\mathbf{v}, \mathbf{h})$ of the optimal equation is called *maximal* if there does not exist any solution $(\mathbf{v}', \mathbf{h}')$ of (3.9) such that $\mathbf{v} \prec \mathbf{v}'$.

Theorem 3.1 *A policy $f \in F$ is average optimal if and only if the pair $(\mathbf{v}(f), \mathbf{h}(f))$ given by* Lemma 3.2 *is a maximal solution to the optimality equation* (3.9).

Proof

The proof of the "only if" part is easily obtained from Lemma 3.4.

In order to prove the "if" part, suppose that $(\mathbf{v}(f), \mathbf{h}(f))$ is a maximal solution of (3.9) but f^∞ is not average optimal. Then, there exists $g \in F$ with $\mathbf{v}(f) \prec \mathbf{v}(g)$. If $(\mathbf{v}(g), \mathbf{h}(g)) \notin \mathcal{U}(\mathbf{h}(g))$, there exists $f^{(1)} \in F$ such that $\mathbf{v}(g) + \mathbf{h}(g) \prec L(f^{(1)})h(g)$, which implies from Lemma 3.4 that $\mathbf{v}(g) \prec \mathbf{v}(f^{(1)})$. Since F is a finite set, by repeating this method successively, we come to the conclusion that there exists $f^{(l)} \in F$ such that $\mathbf{v}(f) \prec \mathbf{v}\left(f^{(l)}\right)$ and $\left(\mathbf{v}\left(f^{(l)}\right), \mathbf{h}\left(f^{(l)}\right)\right)$ satisfies (3.9). However, this contradicts that $(\mathbf{v}(f), \mathbf{h}(f))$ is maximal. ∎

Remark 3.1 *For vector-valued discounted MDPs, Furukawa* [3] *and White* [18] *derived the optimal equation including efficient set-function* *on R^n, by which optimal policies are characterized. The form of the optimality equation* (3.9) *is corresponding to the average case of controlled Markov set-chains.*

As a simple example, a machine maintenance problem ([14, p.1, pp.17–18)], in the typical Markov decision processes can be formulated as this Markov set-chain version and possible to find a discounted or an average optimal policy by applying Theorems 2.1 or 3.1 respectively. However the details are omitted here.

References

[1] Bather, J. (1973) Optimal decision procedures for finite Markov chains, Part II : Communicating systems, *Adv. Appl. Prob.*, **5**, pp. 521–540.

[2] Blackwell, D. (1962) Discrete dynamic programming, *Ann. Math. Statist.* **33**, pp. 719–726.

[3] Furukawa,N. (1980) Characterization of Optimal Policies in Vector-valued Markovian Decision Process, Math.Oper.Res. vol.5, pp. 271–279.

[4] Hartfiel, D. J. (1993) Cyclic Markov set-chain. *J.Stat.Comp.Simul.*, **46**, pp. 145–167.

[5] Hartfiel, D. J. and Seneta, E. (1994) On the theory of Markov Set-chains, *Adv. Appl. Prob.* 26, pp. 947–964.

[6] Hartfiel, D. J. (1998) Markov Set-chains, Springer-Verlag, Berlin.

[7] Hinderer, K. (1970) *Foundations of Non-Stationary Dynamic Programming with Discrete Time Parameter.* Springer-Verlag, New York.

[8] Hosaka, M. and Kurano, M. (1999) Non-discounted Optimal policies in controlled Markov Set-chains, *Journal of Opern. Res. Soc. of Japan* 42, No.3., pp. 256–267.

[9] Hosaka, M., Horiguchi, M. and Kurano, M. (2001) Controlled Markov Set-chains under Average Criteria, *Applied Mathematics and Computation.*

[10] Howard, R. (1960) *Dynamic Programming and Markov processes,* 120, pp. 195–209, MIT Press, Cambridge MA.

[11] Kemeny, J. G. and Snell, J. L. (1960) *Finite Markov-Chains,* Van Nostrand, New York.

[12] Kurano, M., Nakagami, J. and Horiguchi, M. (1999) Controlled Markov Set-Chains with Set-valued rewards, *Proceeding of Inter-*

national Conference on Nonlinear Analysis and Convex Analysis(NACA98), Ed. by W.Takahashi and T.Tanaka, pp. 205–212

[13] Kurano, M. , Song, J. , Hosaka, M. and Huang, Y. (1998) Controlled Markov Set-Chains with Discounting, *J.Appl.Prob.*, pp. 293–302.

[14] Mine, H. and Osaki, S. (1970) *Markov Decision Processes,* Elsevier, Amsterdam.

[15] Nenmaier, A. (1984) New techniques for the analyses of linear interval equations, *Linear Algebra Appli.* 58, pp. 273–325.

[16] Puterman, M. L. (1994) *Markov decision processes: Discrete Stochastic Dynamic Programming,* John Wiley & Sons, INC.

[17] Seneta, E. (1981) *Nonnegative Matrices and Markov Chains,* Springer-Verlag, New York.

[18] White,D.J. (1982) Multi-objective infinite-horizon discounted Markov Decision Processes, J.Math.Anal.Appl., vol.89, pp. 639–647.

Chapter 13

CONSTRAINED DISCOUNTED SEMI-MARKOV DECISION PROCESSES*

Eugene A. Feinberg
Department of Applied Mathematics and Statistics
SUNY at Stony Brook
Stony Brook, NY 11794-3600
Eugene.Feinberg@sunysb.edu

Abstract This paper reduces problems on the existence and the finding of optimal policies for multiple criterion discounted SMDPs to similar problems for MDPs. We prove this reduction and illustrate it by extending to SMDPs several results for constrained discounted MDPs.

Keywords: Semi-Markov decision process, constrained optimization, discounted rewards.

AMS (MOS) subject classification: 90C40, 90C42, 90C39.

1. Introduction

This paper deals with multiple criterion discounted Semi-Markov Decision Processes (SMDPs). SMDPs are continuous-time generalizations of Markov Decision Processes (MDP). The main difference between these two models is that time intervals between jumps have arbitrary distributions in SMDPs and they all are equal to one in MDPs. Another difference is that strategies for SMDPs can use the information about the real time in addition to the information about the step numbers. For a discrete time MDP, the step number is the only time parameter. For many production, service, and telecommunication problems, SMDPs provide more realistic models than MDPs.

One-criterion discounted SMDPs can be reduced to discounted MDPs. This fact is well-known and its proof is based on the properties of optimality equations for discounted SMDPs; see Puterman [18, Section 11.3]. For multiple criterion models,

*This research was partially supported by NSF Grant DMI-9908258.

Z. Hou et al. (eds.), Markov Processes and Controlled Markov Chains, 233–244.

this proof fails because the optimality equation arguments are not valid anymore. The main focus of this paper is to establish the reduction of SMDPs to MDPs for multiple discounted criteria.

We prove this reduction in Section 3. In particular, for a given initial distribution and for an arbitrary policy, we construct a randomized Markov policy with the same performance vector; see Lemma 3.1. For a randomized Markov policy each decision depends only on the current state and on the current step number.

The mentioned construction generalizes the well-known construction of the equivalent randomized Markov policy for an MDP by Derman and Strauch; see Corollary 3.1 below. However, for a discrete-time MDP, such construction does not depend on the discount factor, and the resulting randomized Markov policy is the same for all discount factors including the discount factor equal to 1. For an SMDP, our construction leads to different equivalent randomized Markov policies for different discount rates. For an SMDP, a randomized Markov policy, which performance vectors are equal under all discount rates to performance vectors of a given arbitrary policy, may not exist.

For a multiple criterion discounted SMDP, we consider a discounted MDP with the same state and action sets, and with the same reward functions. In addition, the performance vectors coincide for these two models. Therefore, a randomized Markov policy, which is optimal for the corresponding discounted MDP, is also optimal for the original SMDP.

In Section 4 we illustrate this reduction by showing that several results recently established for MDPs hold also for SMDPs. As mentioned above, SMDPs are important for applications. In addition, the author's interest in SMDPs is motivated by their usefulness in studying Continuous Time Jump Markov Decision Processes; Feinberg [7, 8].

2. Definitions

The probability structure of an SMDP is specified by the four objects $\{X, A, D(x), Q(t, Y|x, a)\}$, where:

(i) X is a Borel state space;

(ii) A is a Borel action space;

(iii) $D(x) \subseteq A$ are Borel sets of actions available at $x \in X$;

(iv) $Q(\cdot|x, a)$ is a transition probability from $X \times A$ into $[0, \infty] \times X$.

It is assumed that

$$\text{graph}(D) = \{(x, a) : \ x \in X, a \in D(x)\}$$

is a Borel subset of $X \times A$ containing the graph of a Borel mapping from X to A. We denote by $\mathcal{X}$ and $\mathcal{A}$ the Borel σ-fields on X and A respectively.

We denote $Q(t, Y|x, a) = Q([0, t] \times Y|x, a)$ for any $0 \leq t < \infty$ and for any Borel $Y \subseteq X$. If action a is selected in state x then $Q(t, Y|x, a)$ is the joint probability that the sojourn time is not greater than $t \in R_+$ and the next state y is in Y, where $R_+ = [0, \infty)$.

Let ξ be the sojourn time. Then $P\{\xi \leq t\} = Q(t, X|x, a)$. Everywhere in this paper, we make the following standard assumption that implies that the system does not have accumulation points:

A1. There exist $\bar{\epsilon} > 0$ and $\bar{t} > 0$ such that $Q(\bar{t}, X|x, a) < 1 - \bar{\epsilon}$ for all $x \in X$ and for all $a \in A$.

Let $\mathbf{H}_n = X \times (A \times R_+ \times X)^n$, $n = 0, 1, \dots, \infty$, be the set of all histories up to n^{th} jump. Then $\mathbf{H} = \bigcup_{0 \le n < \infty} \mathbf{H}_n$ is the set of all histories that contain a finite number of jumps. The sets $\mathbf{H}_n$, $n = 0, 1, \dots, \infty$, and $\mathbf{H}$ are endowed with the σ-fields generated by the σ-fields $\mathcal{X}$, $\mathcal{A}$, and $\mathcal{B}(R_+)$; everywhere in this paper, for a Borel space E we denote by $\mathcal{B}(E)$ its Borel σ-field. A (possibly randomized) strategy π is defined as a transition probability from $\mathbf{H}$ to A such that $\pi(D(x_n) \mid \omega_n) = 1$ for each $\omega_n = x_0 a_0 \xi_0 \dots x_{n-1} a_{n-1} \xi_{n-1} x_n \in \mathbf{H}$, $n = 0, 1, \dots$.

To define a sample space that includes trajectories that have finite numbers of jumps over R_+, we add an additional point $\bar{x} \notin X$ to X and an additional point $\bar{a} \notin A$ to A. Let $\bar{X} = X \cup \{\bar{x}\}$ and $\bar{A} = A \cup \{\bar{a}\}$. We also define $D(\bar{x}) = \{\bar{a}\}$, $Q((\infty, \bar{x})|x, a) = 1 - Q(R_+ \times X|x, a)$ for $x \in X$, $a \in A$, and $Q((\infty, \bar{x})|x, a) = 1$ when either $x = \bar{x}$ or $a = \bar{a}$. We have that Q is a transition probability from $\bar{X} \times \bar{A}$ to $\bar{R}_+ \times \bar{X}$, where $\bar{R}_+ = [0, \infty]$.

Let $\bar{\mathbf{H}}_n = \bar{X} \times (\bar{A} \times \bar{R}_+ \times \bar{X})^n$, $n = 0, 1, \dots, \infty$. We also consider $\mathcal{B}(\bar{\mathbf{H}}_n) = \mathcal{B}(\bar{X}) \times (\mathcal{B}(\bar{A}) \times \mathcal{B}(\bar{R}_+) \times \mathcal{B}(\bar{X}))^n$. Any strategy π defines transition probabilities from $\bar{\mathbf{H}}_n$ to $\bar{\mathbf{H}}_n \times \bar{A}$ and Q defines transition probabilities from $\bar{\mathbf{H}}_n \times \bar{A}$ to $\bar{\mathbf{H}}_{n+1}$, $n = 0, 1, \dots$. Any initial distribution μ on X and any strategy π define a probability measure on the set $(\bar{\mathbf{H}}_\infty, \mathcal{B}(\bar{\mathbf{H}}_\infty))$; Neveu [15, Section 5.1]. We denote this measure by $\mathbb{P}_\mu^\pi$ and denote the expectation operator with respect to this measure by $\mathbb{E}_\mu^\pi$.

Let $\mathbf{h}_\infty = (x_0 a_0 \xi_0 x_1 a_1 \xi_1 \dots)$. We set $t_0 = 0$ and $t_n = t_{n-1} + \xi_{n-1}$, $n = 0, 1, \dots$. Let $N(t) = \sup\{n \ge 0 : t_n \le t\}$. **A1** implies that $N(t) < \infty$, ($\mathbb{P}_\mu^\pi$-a.s.) for all $t \in R_+$ and $t_n \to \infty$ ($\mathbb{P}_\mu^\pi$-a.s.) as $n \to \infty$ for all μ and π.

We may consider an SMDP as an object that has two time parameters. The first parameter is the actual continuous time $t = t_n$ at an n^{th} jump epoch. The second parameter is the jump number n. We say that a strategy is a *policy* if at each epoch t_n, $n = 0, 1, \dots$, the decision does not depend on the times $\xi_0, \dots, \xi_{n-1}$. A *randomized Markov* policy π is defined by a sequence of transition probabilities $\{\pi_n : n = 0, 1, \dots\}$ from X into A such that $\pi_n(D(x) \mid x) = 1$, $x \in X$, $n = 0, 1, \dots$. A *Markov policy* is defined by a sequence of mappings $\phi_n : X \to A$ such that $\phi_n(x) \in D(x)$, $x \in X$, $n = 0, 1, \dots$. A *randomized stationary* policy π is defined by a transition probability π from X into A such that $\pi(D(x) \mid x) = 1$, $x \in X$. A *stationary* policy is defined by a mapping $\phi : X \to A$ such that $\phi(x) \in D(x)$, $x \in X$.

The reward structure of an SMDP is specified by the three objects $\{\alpha, K, r_k(x, a)\}$, where:

a) $\alpha > 0$ is a discount rate;

b) $K = 0, 1, \dots$ is a number of constraints;

c) $r_k(x, a)$ is the expected discounted cumulative reward at the state x for the criterion $k = 0, \dots, K$ if the action a is selected. We assume that r_k are bounded above Borel functions on $X \times A$. We set $r_k(\bar{x}, \bar{a}) = 0$, $k = 0, \dots, K$.

Given an initial state distribution μ and a strategy π, the expected total discounted rewards over the infinite horizon are:

$$W_k(\mu, \pi) = \mathbb{E}_\mu^\pi \sum_{n=0}^{\infty} e^{-\alpha t_n} r_k(x_n, a_n), \qquad k = 0, \dots, K. \tag{2.1}$$

When we consider one criterion, or what we write is true for all criteria, we may omit indexes $k = 0, 1, \dots, K$. We assume everywhere that $0 \times \infty = 0$.

For a one-criterion problem, a strategy π is called *optimal* if $W(\mu,\pi) \geq W(\mu,\sigma)$ for any initial distribution μ and for any strategy σ. For a problem with multiple criteria, we fix the initial distribution μ and constants C_k, $k = 1,\ldots,K$. A strategy π is called *feasible* if $W_k(\mu,\pi) \geq C_k$ for all $k = 1,\ldots,K$. If there exists at least one feasible strategy, the SMDP is called *feasible.* A feasible strategy π is called *optimal* for a problem with multiple criteria if $W_0(\mu,\pi) \geq W_0(\mu,\sigma)$ for any feasible strategy σ.

A discrete time MDP is a particular case of an SMDP when all sojourn times ξ_i are deterministic and equal to 1. In this case, the transition mechanism is defined by transition probabilities $p(dy|x,a)$ instead of transition kernels Q; $p(X|x,a) = 1$. In other words, $Q(t,Y|x,a) = p(Y|x,a)I\{t \geq 1\}$, where I is the indicator function. Since all sojourn times are equal to 1, each strategy in an MDP is a policy and strategic measures are defined on $(H_\infty, \mathcal{B}(H_\infty))$. Consider the discount factor $\beta = e^{-\alpha}$. For MDPs, formula (2.1) has a simpler form:

$$W_k(\mu,\pi) = \mathbb{E}_\mu^\pi \sum_{n=0}^{\infty} \beta^n r_k(x_n,a_n), \qquad k = 0,\ldots,K. \tag{2.2}$$

Remark 2.1 *In this section we have defined a homogeneous SMDP. We can also consider a non-homogeneous SMDP when the action sets D, rewards r_k, and transition kernels Q depend on the step number. In this case, we have $D = D(x,n)$, $r = r(x,n,a)$, and $Q = Q(t,Y|x,n,a)$. A non-homogeneous SMDP can be reduced to the homogeneous SMDP by replacing the state space X with $X \times \{0,1,\ldots\}$. Then there is a one-to-one correspondence between (randomized) Markov policies for the original non-homogeneous SMDP and (randomized) stationary policies for the new homogeneous SMDP. Therefore, the existence of optimal (randomized) stationary policies for homogeneous SMDPs implies the existence of optimal (randomized) Markov policies for non-homogeneous SMDPs. A finite-step SMDP is an important example of a non-homogeneous SMDP. An important application of finite-step SMDPs is scheduling of a finite number of jobs with random durations; Ross* [19]*, Pinedo* [16]. *For a finite-step SMDP, the assumption $\alpha > 0$ can be omitted when the functions $r_k(x,a)$, $k = 0,\ldots,K$, are bounded above.*

It is also possible to define SMDPs with parameters depending on time t. We do not expect that the results of this paper can be applied to such models. For example, optimization of total rewards over the final time horizon $[0,T]$, in general, cannot be reduced to a finite-horizon MDP. For such problems, a natural approach is to use discrete-time approximations of continuous-time problems.

3. Reduction of SMDPs to MDPs

We define the regular nonnegative conditional measures on X,

$$\beta(Y|x,a) = \int_0^\infty e^{-\alpha t} Q(dt,Y|x,a).$$

For a strategy π, initial distribution μ, and epochs $n = 0,1,\ldots$, we define bounded non-negative measures $M_{\mu,n}^\pi$ on $X \times A$ and $m_{\mu,n}^\pi$ on X,

$$\begin{aligned} M_{\mu,n}^\pi(Y,B) &= \mathbb{E}_\mu^\pi \ e^{-\alpha t_n} I\{x_n \in Y, a_n \in B\}, \\ m_{\mu,n}^\pi(Y) &= \mathbb{E}_\mu^\pi \ e^{-\alpha t_n} I\{x_n \in Y\}, \end{aligned}$$

where $Y \in \mathcal{X}$ and $B \in \mathcal{A}$.

Since $m^\pi_{\mu,n}(Y) = M^\pi_{\mu,n}(Y, A)$, we have that m is a projection of M on X. In view of Corollary 7.27.2 in Bertsekas and Shreve [2], there is an ($m^\pi_{\mu,n}$-almost everywhere) unique regular transition probability from X to A such that

$$\sigma_n(da|x) = \frac{M^\pi_{\mu,n}(dx, da)}{m^\pi_{\mu,n}(dx)}. \tag{3.1}$$

By definition, (3.1) is equivalent to

$$M^\pi_{\mu,n}(Y, B) = \int_Y \sigma_n(B|x) m^\pi_{\mu,n}(dx)$$

for all $Y \in \mathcal{X}$, $B \in \mathcal{A}$. Since $M^\pi_{\mu,n}$ is concentrated on graph(D) then for every $n = 0, 1, \ldots$ we can select a version of σ_n such that $\sigma(D(x)|x) = 1$ for all $x \in X$. Then $\sigma = \{\sigma_n : n = 0, 1, \ldots\}$ is a randomized Markov policy. Let $R_n(\mu, \pi) = \mathbb{E}^\pi_\mu e^{-\alpha t_n} r(x_n, a_n)$.

Lemma 3.1 *Consider an SMDP. Let π be a strategy and μ be an initial distribution. Then for a randomized Markov policy σ defined by (3.1),*

$$M^\sigma_{\mu,n} = M^\pi_{\mu,n}, \qquad n = 0, 1, \ldots\,. \tag{3.2}$$

In addition, $R_n(\mu, \sigma) = R_n(\mu, \pi)$ for all $n = 0, 1, \ldots$ and therefore $W(\mu, \sigma) = W(\mu, \pi)$ for any bounded above Borel reward function r.

Proof We notice that the definition of $M^\pi_{\mu,n}$ implies that for any measurable on $X \times A$ step-function f,

$$\mathbb{E}^\pi_\mu e^{-\alpha t_n} f(x_n, a_n) = \int_X \int_A f(x, a) M^\pi_{\mu,n}(dx, da). \tag{3.3}$$

Therefore, (3.3) holds for any bounded above and measurable function f on $X \times A$. Thus,

$$R_n(\mu, \pi) = \int_X \int_A r(x, a) M^\pi_{\mu,n}(dx, da),$$

and the second statement of the lemma follows from the first one.

We shall prove (3.2) by induction. We have that $\sigma_0 = \pi_0$ (μ-a.s.) and thus (3.2) is obvious for $n = 0$. Let (3.2) hold for some n. First we show that

$$m^\sigma_{\mu,(n+1)} = m^\pi_{\mu,(n+1)}. \tag{3.4}$$

For any strategy γ,

$$\begin{aligned} m^\gamma_{\mu,(n+1)}(Y) &= \mathbb{E}^\gamma_\mu e^{-\alpha t_{n+1}} I\{x_{n+1} \in Y\} \\ &= \mathbb{E}^\gamma_\mu \mathbb{E}^\gamma_\mu \left[e^{-\alpha(t_n+\xi_n)} I\{x_{n+1} \in Y\} | t_n, x_n, a_n\right] \\ &= \mathbb{E}^\gamma_\mu e^{-\alpha t_n} \mathbb{E}^\gamma_\mu \left[e^{-\alpha \xi_n} I\{x_{n+1} \in Y\} | t_n, x_n, a_n\right] \\ &= \mathbb{E}^\gamma_\mu e^{-\alpha t_n} \int_0^\infty e^{-\alpha t} Q(dt, Y|x_n, a_n) \\ &= \mathbb{E}^\gamma_\mu e^{-\alpha t_n} \beta(Y|x_n, a_n) \\ &= \int_X \int_A \beta(Y|x, a) M^\gamma_{\mu,n}(dx, da). \end{aligned} \tag{3.5}$$

The last equality follows from (3.3) with $f(x,a) = \beta(Y|x,a)$. By setting $\gamma = \sigma$ and $\gamma = \pi$ we have that (3.2) implies (3.4).

Now we prove that (3.4) implies

$$M^{\sigma}_{\mu,(n+1)} = M^{\pi}_{\mu,(n+1)}. \tag{3.6}$$

From (3.1) and (3.4) we have

$$\begin{aligned} M^{\pi}_{\mu,(n+1)}(Y,B) &= \int_Y \sigma_{n+1}(B|x) m^{\pi}_{\mu,(n+1)}(dx) \\ &= \int_Y \sigma_{n+1}(B|x) m^{\sigma}_{\mu,(n+1)}(dx). \end{aligned}$$

For a randomized Markov policy σ,

$$\begin{aligned} M^{\sigma}_{\mu,(n+1)}(Y,B) &= \mathbb{E}^{\sigma}_{\mu}\; e^{-\alpha t_{n+1}} I\{x_{n+1} \in Y, a_{n+1} \in B\} \\ &= \mathbb{E}^{\sigma}_{\mu} \mathbb{E}^{\sigma}_{\mu} \left[e^{-\alpha t_{n+1}} I\{x_{n+1} \in Y\} I\{a_{n+1} \in B\} | t_{n+1}, x_{n+1} \right] \\ &= \mathbb{E}^{\sigma}_{\mu}\, e^{-\alpha t_{n+1}} I\{x_{n+1} \in Y\}\, \mathbb{E}^{\sigma}_{\mu} \left[I\{a_{n+1} \in B\} | t_{n+1}, x_{n+1} \right] \\ &= \mathbb{E}^{\sigma}_{\mu}\, e^{-\alpha t_{n+1}} I\{x_{n+1} \in Y\}\, \mathbb{P}^{\sigma}_{\mu}\{a_{n+1} \in B | x_{n+1}\} \\ &= \mathbb{E}^{\sigma}_{\mu}\, e^{-\alpha t_{n+1}} I\{x_{n+1} \in Y\} \sigma_{n+1}(B|x_{n+1}) \\ &= \int_Y \sigma_{n+1}(B|x) m^{\sigma}_{\mu,(n+1)}(dx), \end{aligned} \tag{3.7}$$

where the last equality follows from (3.3). So, (3.6) is proved. ■

We notice that, in general, formula (3.1) defines different randomized Markov policies σ for different discount rates α. For an MDP, $t_n = n$, (3.1) transforms into (3.8), and Lemma 3.1 transforms into the following well-known statement, in which in the equivalent Markov policy σ does not depend on the discount factor.

Corollary 3.1 (Derman and Strauch [6]) *Consider an MDP. Let π be a policy and μ be an initial distribution. Consider a randomized Markov policy σ such that for all $n = 0, 1, \ldots$ and for all $x_n \in X$,*

$$\sigma_n(da_n|x_n) = \frac{\mathbb{P}^{\pi}_{\mu}(dx_n da_n)}{\mathbb{P}^{\pi}_{\mu}(dx_n)}, \qquad (\mathbb{P}^{\pi}_{\mu} \text{ -- a. s.}). \tag{3.8}$$

Then $\mathbb{P}^{\sigma}_{\mu}(dx_n da_n) = \mathbb{P}^{\pi}_{\mu}(dx_n da_n)$, $n = 0, 1, \ldots$, and therefore $W(x,\sigma) = W(x,\pi)$.

We remark that (3.1) also implies (3.8) if t_n and (x_n, a_n) are $\mathbb{P}^{\pi}_{\mu}$-independent. Therefore, Corollary 3.1 also holds for SMDPs in which sojourn times do not depend on states and actions. In particular, this independence holds for uniformized Continuous Time Markov Decision Processes; see e.g. [3].

Let $\beta(x,a) = \beta(X|x,a)$, $\beta(X|\bar{x},\bar{a}) = 0$, $\beta(\bar{x}|\bar{x},\bar{a}) = 1$, and $\beta(\bar{x}|x,a) = 1 - \beta(x,a)$ for $x \in X$, $a \in D(x)$. We observe that $\beta(x,a) = 0$ means that the state x is absorbent under the action a. For $\bar{\epsilon}$ and $\bar{t}$ from **A1**

$$\begin{aligned}
\beta(x,a) &= \int_0^{\bar{t}} e^{-\alpha t} Q(dt, X|x,a) + \int_{\bar{t}}^{\infty} e^{-\alpha t} Q(dt, X|x,a) \\
&\leq Q(\bar{t}, X|x,a) + e^{-\alpha \bar{t}}(1 - Q(\bar{t}, X|x,a)) \\
&\leq e^{-\alpha \bar{t}} + Q\left(\bar{t}, X|x,a)(1 - e^{-\alpha \bar{t}}\right) \\
&\leq e^{-\alpha \bar{t}} + (1-\bar{\epsilon})\left(1 - e^{-\alpha \bar{t}}\right) \\
&= 1 - \bar{\epsilon}\left(1 - e^{-\alpha \bar{t}}\right) \\
&< 1.
\end{aligned}$$

Given a discounted SMDP, we shall construct an equivalent discounted MDP. We shall do it in three steps. At each step we define an MDP. All these MDPs have the same reward functions and the same sets of states, actions, and available actions as the original SMDP. Similarly to the original SMDP, they have an additional absorbent state $\bar{x}$ with zero rewards at it. In view of Lemma 3.1 and Corollary 3.1, in order to establish the equivalency, it is sufficient to show that, for any randomized Markov policy, the value of the appropriate criteria remain unchanged in all these models. At Step 1 we define a total-reward MDP with transition probabilities $\beta(Y|x,a)$ and with the expected total rewards. At step 2 we define the transition probabilities of the corresponding MDP by

$$p^{\alpha}(Y|x,a) = \begin{cases} \frac{\beta(Y|x,a)}{\beta(x,a)}, & \text{if } \beta(x,a) > 0; \\ \text{arbitrary}, & \text{otherwise}; \end{cases}$$

and consider the expected discounted total rewards with the discount factor $\beta(x,a)$. In order to use a constant discount factor, at step 3 we define the transition probabilities for the corresponding MDP by

$$\bar{p}(Y|x,a) = \begin{cases} \beta(Y|x,a)/\bar{\beta}, & \text{if } Y \in \mathcal{X},\ x \in X; \\ 1 - \beta(x,a)/\bar{\beta}, & \text{if } Y = \{\bar{x}\},\ x \in X; \\ 1, & \text{if } Y = \{\bar{x}\},\ x = \bar{x}; \end{cases}$$

where $\bar{\beta} < 1$ and $\bar{\beta} \geq \beta(x,a)$ for all $x \in X$ and for all $a \in A$.

We remark that, in each of these three MDPs, the transition probabilities depend on the discount rate α. For example, let X be finite and $p(y|x,a)$ be the probability that the next state of the SMDP is y if the action a is selected at the current state x. Then $p^{\alpha} \neq p$ if we do not assume that the distribution of the next state does not depend on the sojourn time. Except for special cases, this assumption does not take place in particular applications such as control of queues.

Thus, the reduction of a discounted constrained (or, in general, multiple-criterion) SMDP to the corresponding MDP can be conducted in the following three steps.

Step 1 We define the MDP with transition probabilities $\beta(Y|x,a)$. Let $\mathbf{P}_{\mu}^{\pi}$ be the probability measure on the sets of trajectories in this MDP defined by the initial distribution μ and policy π. Let $\mathbf{E}_{x}^{\pi}$ be the expectation operator with respect to this measure. The expected total rewards are $\mathbf{W}_k(\mu,\pi) = \mathbf{E}_{\mu}^{\pi} \sum_{n=0}^{\infty} r_k(x_n, a_n)$. Let π be a randomized Markov policy. The equality $M_{\mu,0}^{\pi}(dx) = \mu(dx)$, and formulas (3.5) and (3.7) imply that $\mathbf{P}_{\mu,n}^{\pi}(Y,B) = M_{\mu,n}^{\pi}(Y,B)$ for all $n = 0, 1, \ldots$

and for all measurable subsets Y and B of X and A respectively. This implies that $\mathbf{W}_k(\mu,\pi) = W_k(\mu,\pi)$.

Step 2 Consider the discounted MDP with the transition probability p^α and with the total discounted criteria $\tilde{W}$,

$$\tilde{W}_k(\mu,\pi) = \tilde{\mathbb{E}}_\mu^\pi \sum_{n=0}^{\infty} \left[\prod_{i=0}^{n-1} \beta(x_i,a_i)\right] r_k(x_n,a_n),$$

where $\tilde{\mathbb{E}}$ is the expectation operator in this MDP and the product from 0 to -1 is defined as 1. It is obvious that $\tilde{W}_k(\mu,\pi) = \mathbf{W}_k(\mu,\pi)$. Therefore $\tilde{W}_k(\mu,\pi) = W_k(\mu,\pi)$ for any randomized Markov policy π. Therefore, an optimal randomized Markov policy for this MDP is also optimal for the original SMDP.

The theory of discounted MDPs with discount factors depending on states and actions is similar to the theory of standard discounted MDPs. The only difference is that the optimality operator T, which is usually is defined as $T^a f(x) = r(x,a) + \beta \int_X f(y)p(dy|x,a)$ for MDPs with a constant discount factor β, has the form

$$T^a f(x) = r(x,a) + \int_X f(y)\beta(x,a)p^\alpha(y|x,a). \tag{3.9}$$

Condition **A1** implies that $\beta(x,a) \le 1 - \bar{\epsilon}(1 - e^{-\alpha\bar{t}}) < 1$. This provides contraction properties for the operator T defined in (3.9); see Denardo [5]. However, the next step reduces an SMDP to a standard discounted MDP.

Step 3 Consider the discounted MDP with the discount factor $\bar{\beta}$ and transition probabilities $\bar{p}$. Let $\overline{\mathbb{P}}_\mu^\pi$ and $\overline{\mathbb{E}}_\mu^\pi$ be the corresponding probability and expectation operators for this MDP. We consider the expected total discounted rewards

$$\overline{W}_k(\mu,\pi) = \overline{\mathbb{E}}_\mu^\pi \sum_{n=0}^{\infty} \bar{\beta}^n r_k(x_n,a_n).$$

It is obvious that $\bar{W}_k(\mu,\pi) = \tilde{W}_k(\mu,\pi)$. It is also obvious that $\overline{P}_\mu^\pi(x_n \in Y, a_n \in B) = \tilde{\beta}^n \tilde{P}_\mu^\pi(x_n \in Y, a_n \in B)$. The above results on the equivalence of models at steps 1–3 imply the following theorem which justifies for multiple criteria the reduction of a discounted SMDP to a standard discounted SMPD with the transition probabilities $\bar{p}$ and the discount factor $\bar{\beta}$.

We recall that a policy π for an SMDP is a policy for the corresponding MDP and vice versa.

Theorem 3.1 *Consider an SMDP and let an initial distribution μ and a policy π be given. Then the following statements hold:*

(i) $M_{\mu,n}^\pi(Y,B) = \bar{\beta}^n \overline{\mathbb{P}}_\mu^\pi(x_n \in Y, a_n \in B)$*, where* $n = 0,1,\ldots$, $Y \in \mathcal{X}$*, and* $B \in \mathcal{A}$*;*

(ii) $W_k(\mu,\pi) = \overline{W}_k(\mu,\pi)$ *for all* $k = 0,\ldots,K$*;*

(iii) A policy is optimal for an SMDP if and only if it is optimal for the MDP obtained from that SMDP by replacing the transition kernel Q and discount rate α with the transition probabilities $\bar{p}$ and discount factor $\bar{\beta}$.

Theorem 3.1 provides the justification for the reduction of discounted SMDPs to discounted MDPs. It also implies the sufficiency of randomized stationary policies for discounted SMDPs. Krylov [14] and Borkar [4] provided a formula, that, for a given initial measure, computes for an arbitrary policy a randomized stationary policy with the equal occupation measure; see Piunovskiy [17, Lemma 24 on p. 307], where this result is presented for Borel MDPs. For an SMDP, the occupation measure is $\nu_\mu^\pi = \sum_{n=0}^\infty M_{\mu,n}^\pi$. We set $\tilde{\nu}_\mu^\pi(Y) = \nu_\mu^\pi(Y, A)$ for $Y \in \mathcal{X}$. Lemma 3.1, Theorem 3.1(i), and the Krylov-Borkar theorem, applied to the MDP at step 3, imply the following result.

Corollary 3.2 (The Krylov-Borkar theorem for SMDPs)
Consider an SMDP. Let π be a strategy and μ be an initial distribution. Then $\nu_\mu^\sigma = \nu_\mu^\pi$ for a randomized stationary policy σ satisfying $\sigma(da|x) = \frac{\nu_\mu^\pi(dxda)}{\tilde{\nu}_\mu^\pi(dx)}$ and therefore $W(\mu, \sigma) = W(\mu, \pi)$.

4. Optimization of discounted SMDPs

The book by Altman [1] describes countable state constrained MDPs and the book by Piunovskiy [17] deals with uncountable constrained MDPs (mostly under the additional assumption that the state space is compact). As mentioned in the introduction, the optimality equations are not applicable to constrained problems. The analysis of multiple criterion problems is based mainly on properties of occupation measures. The major mathematical apparatus used for constrained MDPs is linear programming. In this section we mention three recent results for MDPs and provide their extensions to SMDPs. We shall use q to denote transition probabilities in MDPs.

Hernandez-Lerma and González-Hernández [13] studied Borel state and actions MDPs. They considered the following three additional conditions:

(i) reward functions r_k are upper semi-continuous;

(ii) for any finite number c the set $\{(x \in X, a \in D(x)|r_0(x,a) > c\}$ is compact; and

(iii) transition probabilities $q(dx|x,a)$ are weakly continuous on graph(D).

Under these three conditions they proved the existence of optimal policies and, under some additional assumptions, they formulated linear programs and studied their properties. In view of the Krylov-Borkar theorem for MDPs ([17, Lemma 24 on p. 307]), the paper by Hernandez-Lerma and González-Hernández [13] also implies the existence of optimal randomized stationary policies when conditions (i)–(iii) hold.

Theorem 3.1 implies that this existence results hold for SMDPs if conditions (i) and (ii) hold and the transition measure $\beta(\cdot|x,a)$ is weakly continuous on graph(D). The latter is true if $Q(\cdot|x,a)$ is weakly continuous on graph(D). In particular, the linear programs from Hernandez-Lerma and González-Hernández [13] remain the same for discounted SMDPs with the only change being that the product of $\beta q(dy|x,a)$ should be replaced with $\beta(dy|x,a)$, where the constant β is a discount factor for an MDP.

Feinberg and Piunovskiy [10] considered the following condition for a multiple-criterion total-reward MDP: μ is nonatomic and all measures $q(\cdot|x,a)$ are nonatomic, $x \in X$, $a \in D(x)$. It was proved in [10] that this condition implies that for any policy there exists a nonrandomized Markov policy with the same performance vector; see also [9] for earlier results. Theorem 3.1 implies that if μ and $\beta(\cdot|x,a)$ are nonatomic then for any policy in the discounted SMDP there exists a nonrandomized Markov policy with the same performance vector. We remark that $\bar{x}$ can be an atom of the transition measure $\beta(\cdot|x,a)$. However, $\bar{x}$ can be substituted with a an uncountable Borel set $\bar{X}$ and the probabilities $\beta(\cdot|x,a)$ could be corrected in a way that β does not have atoms on $\bar{X}$. It is also easy to see that if the probability distribution $p(dy|x,a)$ of the next state in the SMDP does not have atoms then the measure $\beta(\cdot|x,a)$ does not have atoms on X either.

If the nonatomic conditions do not hold, nonrandomized optimal policies may not exist; see e.g. Altman [1] and Piunovskiy [17]. The natural question is how to minimize the number of situations when the decision maker uses randomization procedures. Theorem 2.1 in Feinberg and Shwarts [12] describes the optimal policies of this type. Randomized stationary policies that use no more than K randomization procedures are called K-randomized stationary. However, even a 1-randomized stationary policy can use the infinite number of randomization procedures over the time horizon. Feinberg and Shwartz [12] introduces strong (K,n)-policies which satisfy the following conditions:

a) they are randomized Markov;

b) they are (nonrandomized) stationary from time epoch n onward; and

c) they use no more than K randomization procedures at all state-time couples (x,n).

The formal definitions of K-randomized stationary and strong (K,n)-policies are given in Feinberg and Shwartz [12].

Theorem 2.1 in Feinberg and Shwartz [12] establishes the existence of K-randomized stationary policies and the existence for some n of strong (K,n)-policies for discounted MDPs if the following conditions hold:

(i) X is countable or finite,

(ii) all sets of available actions $D(x)$, $x \in X$, are compact;

(iii) reward functions $r_k(x,a)$ are bounded above and continuous in $a \in D(x)$; and

(iv) transition probabilities $p(y|x,a)$ are continuous in $a \in D(x)$ for all $x,y \in X$.

Theorem 3.1 above and Theorem 2.1 in [12] imply Theorem 4.1. We remark that the weakly continuity of $Q(\cdot|x,a)$ on graph(D) implies condition (b) in Theorem 4.1.

Theorem 4.1 *Consider a discounted SMDPs such that:*

a) conditions (i)–(iii) *from the previous paragraph hold, and*

b) for all $x,y \in X$ the functions $\beta(y|x,a)$ and $\beta(x,a)$ are continuous in $a \in D(x)$.

If this SMDP is feasible then

(i) there exists an optimal K-randomized stationary policy; and

(ii) for some finite $n = 0,1,\ldots$ there exists an optimal strong (K,n)-policy.

As was mentioned above, Theorem 3.1 implies that the linear programs that are used for discounted MDPs can be applied, after a minor modification, to discounted SMDPs. If we consider nonhomogeneous SMDPs described in Remark 2.1 then Theorem 4.1(ii) implies the existence of randomized Markov policies which use no more than K randomization procedures and are nonrandomized after some epoch n. For finite-step SMDPs, Theorem 4.1(i) implies the existence of optimal randomized Markov policies which use no more than K randomization procedures at all state-time couples. If X and A are finite, these policies and optimal K-randomized stationary policies for homogeneous infinite-horizon models can be computed by applying linear programs; see Feinberg and Shwartz [11] for the LP formulation for a finite-step problem and Altman [1] for the infinite horizon case.

References

[1] Altman, E. (1999). *Constrained Markov Decision Processes.* Chapman & Hall/CRC, Boca Raton, Florida.

[2] Bertsekas, D.P. and Shreve, S.E. (1978). *Stochastic Optimal Control: The Discrete-Time Case*, Academic Press, New York; republished by Athena Scientific, 1997.

[3] Cassandras, C.G. (1993). *Discrete Event Systems.* IRWIN, Boston.

[4] Borkar V.S. (1988). A convex analytic approach to Markov decision processes, *Prob. Theor. Relat. Fields*, **78**, 583–602.

[5] Denardo, E.V. (1967). Contraction mappings in the theory underlying dynamic programming, *SIAM Rev.*, **9**, 165–177.

[6] Derman, C. and Strauch, R.E. (1966). A note on memoryless rules for controlling sequential control processes. *Ann. Math. Stat.*, **37**, 276–278.

[7] Feinberg, E.A. (1998). Continuous-time discounted jump-Markov decision processes: a discrete-event approach. Preprint. Department of Applied Mathematics and Statistics. State University of New York at Stony Brook. Stony Brook, New York.

[8] Feinberg, E.A. (1999). New approach to optimization of discounted stochastic continuous-time discrete-event systems. *Proceedings of the 38th IEEE Conference on Decision and Control*, **1**, 937 - 941, Phoenix, 7–10, December 1999.

[9] Feinberg, E.A. and Piunovskiy, A.B. (2000). Multiple objective nonatomic Markov decision processes with total rewards. *J. Math. Anal. Appl.*, **247**, 45–66.

[10] Feinberg, E.A. and Piunovskiy, A.B. (2000). Nonatomic total reward Markov decision processes with multiple criteria. *Proceedings of the 39th IEEE Conference on Decision and Control*, **1**, 723–728.

[11] Feinberg, E.A. and Shwartz, A. (1995). Constrained Markov decision models with weighted discounted rewards. *Math. Oper. Res.*, **20**, 302–920.

[12] Feinberg, E.A. and Shwartz, A. (1996). Constrained discounted dynamic programming. *Math. Oper. Res.*, **21**, 922–945.

[13] Hernández-Lerma, O. and González-Hernández, J. (2000). Constrained Markov control processes in Borel spaces: the discounted case. *Math. Meth. Oper. Res.*, **52**, 271–285.

[14] Krylov, N.V. (1985). Once more about the connection between elliptic operators and Itô's stochastic equations. In *Statistics and Control of Stochastic Processes, Steklov Seminar* (eds. N.V. Krylov, R.Sh. Liptser, and A.A. Novikov), Optimization Software, New York, 69–101.

[15] Neveu, J. (1965). *Mathematical Foundations of the Calculus of Probability.* Holden-Day, San Francisco.

[16] Pinedo, M. (1995). *Scheduling: Theory, Algorithms, and Systems.* Prentice Hall, Englewood Cliffs.

[17] Piunovskiy, A.B. (1997). *Optimal Control of Random Sequences in Problems with Constraints.* Kluwer, Boston.

[18] Puterman, M.L. (1994). *Markov Decision Processes.* John Wiley, New York.

[19] Ross, S.M. (1983). *Introduction to Stochastic Dynamic Programming.* Academic Press, New York.

Chapter 14

LINEAR PROGRAM FOR COMMUNICATING MDPS WITH MULTIPLE CONSTRAINTS*

Jerzy A. Filar
School of Mathematics
The University of South Australia
Mawson Lakes, SA 5095
Australia

Xianping Guo
The Department of Mathematics
Zhongshan University
Guangzhou 510275, P. R. China

Abstract In this paper, a mapping is developed between the 'multichain' and 'unichain' linear programs for average reward Markov decision processes (MDPs) with multiple constraints on average expected costs. Our approach applies the communicating properties of MDPs. The mapping is used not only to prove that the unichain linear program solves the average reward communicating MDPs with multiple constraints on average expected costs, but also to demonstrate that the optimal gain for the communicating MDPs with multiple constraints on average expected costs is constant.

Keywords: linear program; multiple constraints; communicating MDPs.

*This work has been supported, in part, by Professor Filar's Australia Research Council Grant # A49906132, by the Natural Science Foundation in China (No.19901038), and by the Natural Science Foundation of Guangdong Province, China.

Z. Hou et al. (eds.), Markov Processes and Controlled Markov Chains, 245–254.

1. Introduction

In this short note, we present a technical result for finite state, finite action space, average expected reward Markov decision processes (MDPs) with multiple constraints on average expected costs. This technical result shows that the communicating properties of MDPs can be applied to map variables from the relatively simple linear program LP 3.2, known to solve average reward unichain MDPs with multiple constraints on average expected costs, to more general linear program LP 3.1, which solves average reward multichain MDPs with multiple constraints on average expected costs. In addition to providing a structural link between the LP 3.2 and LP 3.1, this mapping can be used to a prove that the optimal gain for the communicating MDPs with multiple constraints on average expected costs is constant.

The literature concerning MDPs with average expected costs and linear programming, especially average reward constrained MDPs with particular ergodic structure is already quite extensive. Part of this literature concerns itself with the two linear programs LP 3.1 and LP 3.2 stated in Section 3. A single application of LP 3.1 solves average reward multichain MDPs with multiple constraints on average expected costs. This formulation can be found in Kallenberg [4]. It was proved in [4] that an optimal solution of LP 3.1 can be used to provide an optimal Markov policy. A single application of the simpler LP 3.2 solves average reward unichain MDPs with multiple constraints on average expected costs (see Kallenberg [4] and Puterman [6], etc.). It was shown in [4, 6] that an optimal solution to LP 3.2 provided optimal (deterministic, if desired) stationary policies, and a mapping was developed between solutions to LP 3.2 and optimal stationary policies. Generally, the assumption of unichain structure is difficult to verify.

In Bather [1], communicating MDPs were introduced (see Definition 3.1) and it was established that the optimal gain for an average reward communicating MDP without a constraint is a scalar (i.e., independent of starting state). The differences between communicating and unichain MDPs were exhibited in Kallenberg [4] and Ross and Varadarajan [7]. The communicating MDP's without constraint and their relationship to LP problems were investigated in Filar and Schultz [2]. In Ross and Varadarajan [7], the average reward communicating MDPs with sample path constraints were considered, the existence of $\epsilon > 0$ optimal stationary policies was proved, and the difference between a sample path constraint and an average expected constraint was discussed. This paper will deal with the LP problems and the existence of optimal policies for average reward communicating MDPs with multiple constraints on average expected costs. It will be shown that an optimal solution of the simpler LP 3.2 can be used to provide an optimal Markov policy (Theorem 4.1, Lemma 3.1). It will also be proved that the optimal gain for an average reward communicating MDP with multiple constraints on average expected costs is scalar.

This paper is organized as follows. In Section 2, the notation and the definitions are introduced. Some preparative results are given in Section 3. The mapping which proves the simplified linear programming solution method for these MDPs is developed in Section 4.

2. Model, notation and definitions

For our purpose, an MDP with multiple constraints is defined by a finite state space $S = \{1, \cdots, |S|\}$, finite action sets $A(i) = \{1, \cdots, m_i\}$, $i \in S$, a *reward law* $r=\{r(i,a) : a \in A(i), i \in S\}$, *costs laws* $c_n=\{c_n(i,a) : a \in A(i), i \in S\}$, $n = 1, 2, \cdots, K$, *constraint bounds* b_n, $n = 1, 2, \cdots, K$, and a *transition law* $q = \{q(j|i,a) : \sum_{j \in S} q(j|i,a) = 1, a \in A(i), i \in S\}$. Given that the process is in a state $i \in S$, the decision maker chooses an action $a \in A(i)$, receives reward $r(i,a)$, pays for implementing the action a n-th type of cost $c_n(i,a)$, $n = 1, 2, \cdots, K$, and the process moves to the next stage at state $j \in S$. The decision maker considers the situation where the average expected reward is to be maximized while keeping the K types of average expected costs c_n below the given bounds b_n.

For the average expected reward criterion with constraints on average expected costs considered below, a *randomized Markov policy* for the decision maker, denoted by $\pi = \{\pi_t, t = 0, 1, \cdots\}$, is sufficient to describe the decision maker's course of action, namely, choose action $a \in A(i)$ with probability $\pi_t(a|i), \sum_{a \in A(i)} \pi_t(a|i) = 1$ when the process is in state $i \in S$ at stage $t \geq 0$. A randomized Markov policy $\pi = \{\pi_t, t = 0, 1, \cdots\}$ is called *randomized stationary*, if $\pi_t(a|i) = \pi_0(a|i) := f(a|i)$, $a \in A(i)$, $i \in S$, $t \geq 0$, and denoted by f. A randomized stationary policy f is called *deterministic*, if $f(a|i) = 1$, for exactly one $a \in A(i)$ and each $i \in S$. The sets of all randomized Markov policies, all randomized stationary policies, and all deterministic policies, are denoted by Π_m, Π_{ms} and F, respectively.

We assume that $\beta = (\beta_1, \cdots, \beta_{|S|})$ is a known *initial distribution*, that is, $\beta_n \geq 0$, $n = 1, \cdots, |S|$, and $\sum_{j \in S} \beta_j = 1$. For any $\pi \in \Pi_m$, by Theorem of Ionescu-Tulcea (see Lerma and Laserre [5, pages 16 and 179]), there exist an unique probability measure P_π^β on $((S \times A)^\infty, (\mathcal{B}(S) \times \mathcal{B}(A))^\infty)$, and state and action variables at stage t denoted by X_t and Δ_t, respectively. The expectation operator with respect to P_π^β is denoted by E_π^β. The *average expected reward criterion* R and *average expected cost criteria* C_n respectively are defined well as follows:

$$R(\pi,\beta) := \limsup_{N\to\infty} \frac{\sum_{t=1}^{N} E_\pi^\beta r(X_t,\Delta_t)}{N}, \tag{2.1}$$

$$C_n(\pi,\beta) := \limsup_{N\to\infty} \frac{\sum_{t=1}^{N} E_\pi^\beta c_n(X_t,\Delta_t)}{N}. \tag{2.2}$$

Then, the average reward multichain MDP with multiple constraints on average costs is the following problem (denoted by $\Gamma(0)$) of choosing a $\pi^* \in \Pi_m$ to

$$maximize \quad R(\pi,\beta) \tag{2.3}$$

subject to

$$C_n(\pi,\beta) \le b_n, \quad n = 1,\cdots,K. \tag{2.4}$$

Let $U(\beta) = \{\pi \in \Pi_m : C_n(\pi,\beta) \le b_n, n = 1,\cdots,K\}$. A policy $\pi \in U(\beta)$ is called *feasible.* A policy π^* such that $R(\pi^*,\beta) = \sup_{\pi\in U(\beta)} R(\pi,\beta) := R^*(\beta)$ is called *optimal.* The quantity $R^*(\beta)$ is called *optimal gain* with respect to initial distribution β.

3. Multichain and unichain linear programs

In order to solve the problems of the existence and calculation of optimal policies, Kallenberg [4] proposed the following multichain linear programming.

LP 3.1

$$\max \quad \sum_{i\in S}\sum_{a\in A(i)} r(i,a)x(i,a) \tag{3.1}$$

subject to

$$\sum_{i\in S}\sum_{a\in A(i)} (\delta_{ij} - q(j|i,a))x(i,a) = 0, \quad j \in S \tag{3.2}$$

$$\sum_{a\in A(j)} x(j,a) + \sum_{i\in S}\sum_{a\in A(i)} (\delta_{ij} - q(j|i,a))y(i,a) = \beta_j, \quad j \in S \tag{3.3}$$

$$\sum_{i\in S}\sum_{a\in A(i)} c_n(i,a)x(i,a) \le b_n, \quad n = 1,\cdots K \tag{3.4}$$

$$x(i,a), y(i,a) \ge 0, \quad a \in A(i),\ i \in S. \tag{3.5}$$

and obtained the following results

Lemma 3.1

(i) $U(\beta) \neq \emptyset$ *if and only if* LP 3.1 *is feasible.*

(ii) The optima of the problems $\Gamma(0)$ *and* LP 3.1 *are equal.*

(iii) Let $F = \{f_1, f_2, \cdots, f_m\}$, *and* $P^*(f_k)(1 \leq k \leq m)$ *be the Cesaro limit of powers of transition matrix* $P(f_k) = (q(j|i,\ f_k(i)))$. *Suppose that* $((x(i,a), y(i,a)),\ a \in A(i),\ i \in S)$ *is an optimal solution of LP 3.1, and let* $x(i,a) = \sum_{k=1}^{m} \alpha_k x_k(i,a)$, *where* $x_k(i, f_k(i)) := (\beta P^*(f_k))_i$ *when* $a \neq f_k(i)$, *then,* $x_k(i,a) = 0$ *and* $\alpha_k \geq 0,\ k \in S,\ \sum_{k=1}^{m} \alpha_k = 1$. *If* $\pi \in \Pi_m$ *is the policy, introduced by* α_k *and* f_k, *such that*

$$\sum_{i \in S} \beta_i P_\pi^\beta(x_t = j, y_t = a | x_0 = i)$$
$$= \sum_{i \in S} \beta_i \sum_{k=1}^{m} \alpha_k P_{f_k}^\beta(x_t = j, y_t = a | x_1 = i),$$
$$t \geq 1, a \in A(i), i \in S,$$

then π *is an optimal policy of* $\Gamma(0)$.

Proof See Theorem 4.7.3 in Kallenberg [4]. ■

Remark 3.1 Lemma 3.1 *shows that* LP 3.1 *solves average reward multichain MDPs with multiple constraints on average expected costs. From* Lemma 3.1, *we can find that only* $x(i,a)$ *from the optimal solution to* LP 3.1 *is needed to construct an optimal policy. An algorithm for constructing the above optimal policy can be found in Kallenberg* [4]. *Unfortunately, this algorithm is computationally prohibitive.*

For the case of unichain MDPs, since it has particular ergodic structure, Kallenberg [4] and Puterman [6] give the simpler unichain linear programming formulation:

LP 3.2

$$\max \quad \sum_{i \in S} \sum_{a \in A(i)} r(i,a) x(i,a) \tag{3.6}$$

subject to

$$\sum_{i\in S}\sum_{a\in A(i)}(\delta_{ij}-q(j|i,a))x(i,a) = 0, \quad j\in S \tag{3.7}$$

$$\sum_{i\in S}\sum_{a\in A(i)} x(i,a) = 1 \tag{3.8}$$

$$\sum_{i\in S}\sum_{a\in A(i)} c_n(i,a)x(i,a) \leq b_n, n=1,\cdots K \tag{3.9}$$

$$x(i,a), y(i,a) \geq 0, \quad a\in A(i), i\in S. \tag{3.10}$$

and obtained many strong results (see [4, 6]). In particular an optimal stationary policy is easily constructed.

Now, we consider the case of communicating structure.

Definition 3.1 *An MDP is communicating if, for every pair of states $i, j \in S$, there exists a stationary policy $f \in F$ and an integer $l \geq 1$ (both f and l may depend on i and j) such that $P_{ij}^l(f)$ (the (i,j)-th entry of $[P(f)]^l$) is strictly positive.*

To establish that an MDP is communicating, we have

Lemma 3.2

(i) An MDP is communicating if and only if there exists a $f \in \Pi_{ms}$ such that $P(f)$ is irreducible.

(ii) An MDP is communicating if and only if $P(f)$ is irreducible for every randomized stationary policy f that satisfies $f(a|i) > 0$, $a \in A(i)$, $i \in S$.

Proof See [2, 6, 7]]. ∎

Hence, we have that communicating MDPs are more restrictive than multichain MDPs, but rather different from unichain MDPs in that the communicating property can be easily verified.

4. Linear program relationships

To solve average reward communicating MDPs with mutiple constraints on average expected costs, we set

$$f_*(a|i) = \frac{1}{m_i}, \qquad a\in A(i),\ i\in S,$$

$$P(f_*) = \left(\left(\sum_{a\in A(i)} q(j|i,a)f_*(a|i)\right), i,j\in S\right).$$

By Lemma 3.2, we have that $P(f_*)$ is irreducible. Let $v(f_*) :=$ [4] $(v_1, \cdots, v_{|S|}) > 0$ be the equilibrium distribution for $P(f_*)$. Hence, $P^*_{ij}(f_*) = v_j,\ i,j \in S$, and $Z(f_*) := [I - P(f_*) + P^*(f_*)]^{-1}$ exists (see [3], [6]). Let $\|Z(f_*)\| := \max\{|Z_{ij}(f_*)| : i,j \in S\}$ be the maximal value of absolute values of the elements of $Z(f_*)$, $v(f_*)(min) := \min\{v_k : 1 \le k \le |S|\}$. For any $b = (b_1, \cdots, b_{|S|})$ (all $b_k \in R$, being the set of real numbers), we define a map T as follows:

$$T(b)(i,a) := W(b)_i f_*(a|i), \qquad a \in A(i),\ i \in S. \tag{4.1}$$

where, $W(b) := bZ(f_*) + \lambda v(f_*)$, $\lambda := \frac{|S|\|b\|\|Z(f_*)\|}{v(f_*)(min)}$, $\|b\| := \max\{|b_k|, 1 \le k \le |S|\}$.

Obviously, we can derive that: $T(b)(i,a) \ge 0$, $a \in A(i)$, $i \in S$, $b \in R^{|S|}$.

We now derive our main results.

Theorem 4.1 *For a communicating MDP, we have that*

(i) If $\{(x^*(i,a), y^*(i,a)), a \in A(i), i \in S\}$ is an optimal solution to LP 3.1, *then $\{x^*(i,a), a \in A(i), i \in S\}$ is an optimal solution to* LP 3.2.

(ii) If $\{x^*(i,a), a \in A(i), i \in S\}$ is an optimal solution to LP 3.2, *then $\{(x^*(i,a), T(\beta - x^*)(i,a)), a \in A(i), i \in S\}$ is an optimal solution to* LP 3.1, *where $x^* := (x^*_i, i \in S)$ and $x^*_i := \sum_{a \in A(i)} x^*(i,a), i \in S$.*

Proof

(i) From (3.2), (3.3), (3.4) and (3.5), we can obtain that $\{x^*(i,a) : a \in A(i), i \in S\}$ is a feasible solution to LP 3.2. Suppose that $\{x^*(i,a) : a \in A(i), i \in S\}$ is not an optimal solutions to LP 3.2. Let $\{x(i,a) : a \in A(i), i \in S\}$ be an optimal solutions to LP 3.2, then

$$\sum_{i \in S} \sum_{a \in A(i)} r(i,a)x(i,a) > \sum_{i \in S} \sum_{a \in A(i)} r(i,a)x^*(i,a). \tag{4.2}$$

Since

$$v(f_*)[I - P(f_*)] = 0,$$

$$P(f_*)P^*(f_*) = P^*(f_*)P(f_*) = P^*(f_*)P^*(f_*) = P^*(f_*),$$

and

$$[I - P(f_*) + P^*(f_*)][I - P^*(f_*)] = [I - P(f_*)],$$

we have $Z(f_*)[I - P(f_*)] = [I - P^*(f_*)]$.
Recalling that $\sum_{i\in S}(\beta_i - x_i) = 0$, and $P^*_{ij}(f_*) = v_i, i, j \in S$, we observe that

$$\begin{aligned} W(\beta - x)[I - P(f_*)] &= [(\beta - x)Z(f_*) + \lambda v(f_*)][I - P(f_*)] \\ &= (\beta - x)[I - P^*(f_*)] \\ &= \beta - x, \end{aligned}$$

where the last equality follows from the fact that $P^*(f_*)$ has identical rows. Hence,

$$\begin{aligned} & W(\beta - x)_j - \sum_{i\in S} W(\beta - x)_i P_{ij}(f_*) \\ &\quad = \beta_j - x_j, j \in S, \\ & \sum_{a\in A(j)} T(\beta - x)(j,a) - \sum_{i\in S} W(\beta - x)_i \sum_{a\in A(i)} q(j|i,a) f_*(a|i) \\ &\quad = \beta_j - \sum_{a\in A(j)} x(j,a), j \in S, \\ & \sum_{a\in A(j)} T(\beta - x)(j,a) - \sum_{i\in S} \sum_{a\in A(i)} T(\beta - x)(i,a) q(j|i,a) \\ &\quad = \beta_j - \sum_{a\in A(j)} x(j,a), j \in S, \\ & \sum_{a\in A(j)} x(j,a) + \sum_{i\in S} \sum_{a\in A(i)} (\delta_{ij} - q(j|i,a)) T(\beta - x)(i,a) \\ &\quad = \beta_j, j \in S. \end{aligned} \tag{4.3}$$

Then, by (3.7), (3.9), (3.10) and (4.3), we can derive that $\{(x(i,a), T(\beta - x)(i,a)) : a \in A(i), i \in S\}$ is a feasible solution to LP 3.1. This with (4.2) contradict the optimality of $\{(x^*(i,a), y^*(i,a)), a \in A(i), i \in S\}$ to LP 3.1. Hence, (i) is valid.

(ii) To prove (ii), from the proof of (i), we have that $\{(x^*(i,a), T(\beta - x^*)(i,a)), a \in A(i), i \in S\}$ is a feasible solution to LP 3.1. Assume $\{(x(i,a), y(i,a)), a \in A(i), i \in S\}$ is any feasible solution to LP 3.1. By (3.2), (3.3), (3.4) and (3.5), $\{x(i,a), a \in A(i), i \in A\}$ is a feasible solution to LP 3.2. By the optimality of $\{x^*(i,a), a \in A(i), i \in A\}$ for LP 3.2, we have

$$\sum_{i\in S} \sum_{a\in A(i)} r(i,a) x^*(i,a) \geq \sum_{i\in S} \sum_{a\in A(i)} r(i,a) x(i,a).$$

So, (ii) is proved. ■

When $\beta_i = 1, \beta_j = 0, j \neq i, j \in S$, we denote $R^*(\beta)$ by $R^*(i)$, and call $R^*(i)$ the optimal gain for starting state i, $R^*(i)(i \in S)$ the optimal gain.

Theorem 4.2 *Suppose that the MDP is communicating. Then,*

(i) $U(\beta) \neq \emptyset$ *if and only if* LP 3.2 *is feasible;*

(ii) the optima of problems of $\Gamma(0)$, LP 3.1 *and* LP 3.2 *are equal;*

(iii) the optimal gain is a scalar (i.e., independent of starting state).

Proof

(i) From the proof of Theorem 4.1, we can obtain that LP 3.2 is feasible if and only if LP 3.1 is feasible. This with Lemma 3.1 show that (i) is valid.

(ii) From Theorem 4.1 and Lemma 3.1 , we can derive that (ii) holds.

(iii) To prove (iii), since the optimal solutions of LP 3.2 are independent of initial distribution β, the optimal gain of LP 3.2 is free of initial distribution β. By (ii), the optimal gain $R^*(\beta)$ is also independent of initial distribution β, and denoted by R^*. Hence, $R^*(i) = R^*, i \in S$. This means that (iii) is proved. ■

Remark 4.1 Theorem 4.1 *and* Remark 3.1 *show that average reward communicating MDPs with multiple constraints on average costs can be solved by the simpler* LP 3.2.

Acknowledgments

The second author is grateful to the hospitality of the Center for Industrial and Applicable Mathematics during his visit to The University of South Australia.

References

[1] Bather, J.,(1973). Optimal decision procedures for finite Markov chains. Part II: Communicating system. Advance in Applied Probability, 5, 521-540.

[2] Filar, J.A. and T.A. Schultz, (1988). Communicating MDPs: equivalence and LP properties. Oper. Res. Letters, 7(6), 303-307.

[3] Hou, Z. T. and Guo, X. P., (1998). Markov Decision Processes, Science and Technology Press of Hunan, Changsha, China.

[4] Kallenberg, L.C.M., (1983). Linear Programming and Finite Markovian Control Problems. Mathematisch Centrum Amsterdam.

[5] Lerma, O. and J.B. Laserre, (1996). Discrete-Time Markov Controlled Processes. Springer-Verlag, New York.

[6] Puterman, M.L., (1994). Markov Decision Processes, JOHN WILEY & SONS INC.

[7] Ross, K.W. and R. Varadarajan, (1989). Markov decision processes with sample path constraints: the communicating case. Oper. Res., 37, 780-790.

Chapter 15

OPTIMAL SWITCHING PROBLEM FOR MARKOV CHAINS

A.A. Yushkevich

University of North Carolina at Charlotte,
Department of Mathematics,
Charlotte, NC 28223, USA

Abstract We consider the following multi-step version of the optimal stopping problem. There is a Markov chain $\{x_t\}$ with a Borel state space X, and there are two functions $f < g$ defined on X; one may interpret $f(x_t)$ and $g(x_t)$ as the selling price and the purchase price of an asset at the epoch t. A controller selects a sequence of stopping times $\tau_1 \leq \tau_2 \leq \ldots$, and can be either in a position to sell or in a position to buy the asset. By selecting $\tau = \tau_k$, the controller, depending on the current position, either gets a reward $f(x_\tau)$ or pays a cost $g(x_\tau)$, and becomes switched to the opposite position. The control process terminates at an absorbing boundary, and the problem is to maximize the expected total rewards minus costs.

We find an optimal strategy and the value functions, and establish a connection to Dynkin games.

1. Introduction

We consider a generalization of the well-known optimal stopping problem for a Markov chain to the case when one may stop and get rewards many times[1]. It turned out that an interesting nontrivial generalization appears if there are two reward functions f and $-g$, with $g > f$, and every "stop" switches them. If $f > 0$, the scheme has the following financial interpretation: $f(x)$ and $g(x)$ are, respectively, the selling and purchase price of an asset when the system is at the state x, and a con-

[1]The literature on optimal stopping is enormous. Basic references are Snell [14], Chow, Robbins and Sigmund [2] and Shiryayev [12]. For an introductory exposition see Dynkin and Yushkevich [4, Chapter 3]

Z. Hou et al. (eds.), Markov Processes and Controlled Markov Chains, 255–286.

troller, who observes the Markovian evolution of the system, may sell or buy the asset at any time, but only in an alternating order; the aim of the controller is to maximize the expected profit. It also turned out that this model is closely related to stochastic games called Dynkin games. In this article we consider the expected total rewards criterion in the case of a discrete-time Markov chain with a Borel state space and an absorbing boundary. Other cases will be treated in subsequent papers.

The paper is organized in the following way. In Section 2 we state the optimal switching problem, define strategies and two value functions, corresponding to the selling and the buying positions of the controller. In Section 3 we discuss two ways to imitate the switching model by a Markov decision process, formally introduce the first of them called MDP1, and prove the measurability of the value functions of MDP1 by value iterations (Theorem 3.1). The goal of Section 4 is to establish a correspondence between policies in MDP1 and strategies in the switching problem, sufficient to reduce one optimization problem to the other. In Section 5 we shift the problem to MDP2, which is more convenient for further analysis (Theorem 5.1), justify the value iteration for MDP2 (Lemma 5.2), and characterize the value functions in terms of excessive envelopes (Theorem 5.2, Remark 5.1). In Section 6 we define a preference function as the difference of the two value functions, characterize this function by variational inequalities which coincide with those known in Dynkin games (Theorem 6.1), and describe it in terms of two supporting sets (Corollary 6.1). In Section 7 we prove that the supporting sets generate an optimal policy in MDP1, and that they are optimal switching sets in the original problem (Theorems 7.1, 7.2). In Section 8 we find the value functions (Theorem 8.1). The connection to a Dynkin game is treated in Section 9. In Section 10 we present examples in which the variational sense of the optimality inequalities is visual: the symmetric random walk and the birth and death process.

2. The optimal switching problem

To make the rewards and costs finite, we suppose that the underlying Markov chain reaches an absorbing boundary where the control actually stops. Keeping in mind an extension to the continuous-time case, we assume that every switching is performed instantaneously, so that it moves the controller from the selling position to the buying position or vice versa, but does not change the state of the Markov chain. As in the optimal stopping problem one may stop at the initial time 0, so in the switching problem we allow to switch at $t = 0$ and therefore, to have the Markov property of the system, we also allow two (and thus any denumerable number of) consecutive switchings at the time $t = 0$, and hence at any time $t > 0$.

We now turn to formal definitions. Let $(X, \mathcal{B})$ be a standard Borel space and let $P(x, E)$ ($x \in X$, $E \in \mathcal{B}$) be a measurable stochastic kernel in X (everywhere measurability means Borel measurability, if not stated otherwise). By $\{x_t\}$ we denote the corresponding Markov chain on X; $\mathbf{P}_x$ and $\mathbf{E}_x$ are the distribution and the expectation corresponding to the initial state $x_0 = x$ of this chain. As usual, Pf denotes the function

$$Pf(x) = \int_X f(y)\, P(x, dy), \qquad x \in X$$

(if only this integral is well defined). By $B(X)$ we denote the space of all bounded real-valued measurable functions on X. We assume that X consists of a boundary $B \in \mathcal{B}$ of absorbing states:

$$P(x, x) = 1, \qquad x \in B, \tag{2.1}$$

and of the set $X_0 = X \setminus B$ of interior points. As a sample space Ω we take, for simplicity, the set of all paths absorbed at B, i.e. the subset of X^∞ defined by the condition:

$$x_{t+1} = x_t \quad \text{if} \quad x_t \in B. \tag{2.2}$$

In Ω we consider the minimal σ-algebra $\mathcal{N}$ and filtration $\{\mathcal{N}_t\}$: $\mathcal{N}_t$ is generated by the random variables $x_0, \ldots, x_t$ $(t \geqslant 0)$, $\mathcal{N}$ is generated by all variables x_t. All stopping times $\tau(\omega)$, $\omega \in \Omega$ are understood with respect to this filtration, the range of τ consists of the integers $0, 1, 2, \ldots$ and ∞. By τ_E we denote the first entrance time into a set $E \in \mathcal{B}$. There are also a reward function $f \in B(X)$ and a cost function $g \in B(X)$.

Assumption 2.1 *The sets B and X_0 are nonempty, and*

$$\mathbf{E}_x \tau_B < \infty \qquad x \in X_0. \tag{2.3}$$

Under this assumption

$$\mathbf{P}_x \{\Omega_0\} = 1, \quad \Omega_0 = \{\omega \in \Omega : \tau_B(\omega) < \infty\}, \quad x \in X. \tag{2.4}$$

Assumption 2.2 *The reward and cost functions satisfy conditions*

$$-C \leqslant f < g \leqslant C, \tag{2.5}$$

$$f(x) \geq 0, \quad x \in B, \tag{2.6}$$

and

$$\inf_{x\in X_0} g(x) < \sup_{x\in X_0} f(x). \tag{2.7}$$

Denote by $\mathcal{S}$ the set of all sequences $\{\tau_k, k = 1, 2, \dots\}$ of stopping times such that

$$0 \le \tau_1(\omega) \le \tau_2(\omega) \le \dots, \quad \omega \in \Omega. \tag{2.8}$$

Let $\mathcal{S}_0$ be the subset of $\mathcal{S}$ specified by the condition

$$\tau_k(\omega) < \tau_{k+1}(\omega) \quad \text{if only} \quad \tau_k(\omega) < \infty.$$

We consider two random reward functionals

$$J_1 = J_1(\mathcal{T}, \omega) = \sum_{k=1}^{\infty} F_k(x_{\tau_k}) \mathbf{1}\{\tau_k < \infty\}, \tag{2.9}$$

$$J_2 = J_2(\mathcal{T}, \omega) = \sum_{k=1}^{\infty} F_{k+1}(x_{\tau_k}) \mathbf{1}\{\tau_k < \infty\}, \ \mathcal{T} \in \mathcal{S}, \ \omega \in \Omega, \tag{2.10}$$

where

$$F_k = \begin{cases} f & \text{if} \quad k = 1, 3, 5, \dots, \\ -g & \text{if} \quad k = 2, 4, 6, \dots. \end{cases}$$

Lemma 2.1 *For every $\mathcal{T} \in \mathcal{S}$ and $\omega \in \Omega_0$ the series J_i $(i = 1, 2)$ either contains a finite number of nonzero terms or diverges to $-\infty$, and*

$$J_i(\mathcal{T}, \omega) \leqslant C[\tau_B(\omega) + 1], \quad \omega \in \Omega_0, \quad i = 1, 2. \tag{2.11}$$

(In the case of divergence, both the sums of positive and negative terms can be infinite.)

Proof We consider J_1 (J_2 is treated in a similar way). Let $\omega \in \Omega_0$ be fixed. Then

$$J_1(\omega) = f(x_{t_1}) - g(x_{t_2}) + f(x_{t_3}) - \cdots \tag{2.12}$$

where all t_j are finite, and $0 \leqslant t_1 \leqslant t_2 \leqslant \cdots$. Let $N(\omega)$ be the finite or infinite number of terms in the series (2.12), and let $T = \tau_B(\omega)$; by (2.4) $T < \infty$.

If $N(\omega) < \infty$, the series (2.12) has a finite number of terms. If $N(\omega) = \infty$ and $t_j \uparrow \infty$, then, after t_j exceeds T, all x_{t_j} are equal to x_T by (2.2). If $N(\omega) = \infty$ and $\lim t_j = t^*$ is finite, then all t_j are equal starting from some j, and again x_{t_j} are equal for large numbers j. Thus in any case the series (2.12) consists of two parts: an initial finite sum of the same shape (2.12) and a (maybe absent) infinite remainder of the form $f(z) - g(z) + f(z) - \cdots$ [or $-g(z) + f(z) - g(z) + \cdots$]. The remainder diverges to $-\infty$ by (2.5), and in this case (2.11) trivially holds. Otherwise, the right side of (2.12) has a finite number of terms. By erasing any pair $f\left(x_{t_j}\right) - g\left(x_{t_{j+1}}\right)$ [or $-g\left(x_{t_j}\right) + f\left(x_{t_{j+1}}\right)$] with $x_{t_j} = x_{t_{j+1}}$ we may only increase the sum of all terms. After repeating this "cleaning", we get a sum of type (2.12) with $0 \leq t_1 < t_2 < \cdots \leq T$, i.e. a sum with no more than $T = \tau_B(\omega)$ terms, and the upper bound (2.11) follows from (2.4). ■

Lemma 2.2 *For every $\mathcal{T} \in \mathcal{S}$ there exists $\mathcal{T}' \in \mathcal{S}_0$ with the property*

$$J_i(\mathcal{T}, \omega) \leqslant J_i(\mathcal{T}', \omega), \quad \omega \in \Omega_0, \quad i = 1, 2. \tag{2.13}$$

Proof Let $n_t(\omega)$, $t = 0, 1, 2, \ldots$, $\omega \in \Omega$, be the number of stopping times τ_k in $\mathcal{T}$ such that $\tau_k(\omega) = t$. The possible values of the functions n_t are $0, 1, 2, \ldots$ and ∞, each n_t is $\mathcal{N}_t$-measurable, and if $n_{t_0}(\omega_0) = \infty$ for some t_0 and ω_0 then $n_t(\omega_0) = 0$ for all $t > t_0$; this follows from the structure of $\mathcal{T}$ and the definition of a stopping time. It is easy to see that conversely, any collection of functions n_t with such properties uniquely determines the corresponding element $\mathcal{T} \in \mathcal{S}$: namely, $\tau_k(\omega) = t$ if $\sum_{s=0}^{t-1} n_s(\omega) < k \leq \sum_{s=0}^{t} n_s(\omega)$, $t = 0, 1, 2, \ldots$ and $\tau_k(\omega) = \infty$ if $\sum_{s=0}^{\infty} n_s(\omega) < k$. One may define the needed $\mathcal{T}'$ by the corresponding functions

$$n'_t(\omega) = \begin{cases} 0 & \text{if } n_t(\omega) \text{ is even or is infinite,} \\ 1 & \text{if } n_t(\omega) \text{ is odd.} \end{cases}$$

This results in a "cleaning" of the functionals J_i similar to that in Lemma 2.1, and can only increase them. Since $n'_t(\omega)$ never exceeds 1, $\mathcal{T}' \in \mathcal{S}_0$. ■

Lemma 2.3 *For every $\mathcal{T} \in \mathcal{S}$ the expected rewards*

$$V(x, \mathcal{T}) = \mathbf{E}_x J_1(\mathcal{T}), \qquad W(x, \mathcal{T}) = \mathbf{E}_x J_2(\mathcal{T}), \quad x \in X, \tag{2.14}$$

are well defined, less than $+\infty$, and measurable.

Proof The existence of expectations (2.14) less than $+\infty$ follows from (2.3), (2.5) and Lemma 2.1. From (2.9)–(2.10) and the definition of stopping times it follows that $J_i(\mathcal{T}, \omega)$ is $\mathcal{N}$-measurable as a function of $\omega \in \Omega$ $(i = 1, 2)$. By general properties of Markov chains in Borel spaces, this implies measurability of V and W in $x \in X$. ■

Any pair $\sigma = \left(\mathcal{T}^1, \mathcal{T}^2\right)$ of elements of $\mathcal{S}$ is a *strategy* of the controller. By (2.9)–(2.10), if $\tau_1 = \infty$ then $V(x, \mathcal{T}) = W(x, \mathcal{T}) = 0$. Therefore and by Lemmas 2.2 and 2.3 the *value functions*

$$V(x) = \sup_{\mathcal{T}\in\mathcal{S}} V(x,\mathcal{T}) = \sup_{\mathcal{T}\in\mathcal{S}_0} V(x,\mathcal{T}), \quad x \in X, \tag{2.15}$$

$$W(x) = \sup_{\mathcal{T}\in\mathcal{S}} W(x,\mathcal{T}) = \sup_{\mathcal{T}\in\mathcal{S}_0} W(x,\mathcal{T}), \quad x \in X, \tag{2.16}$$

are well defined and nonnegative. By (2.1), (2.5) and (2.6), on the boundary

$$V(x) = f(x), \quad W(x) = 0, \quad x \in B.$$

A strategy $\sigma^* = \left(\mathcal{T}^1, \mathcal{T}^2\right)$ is *optimal* if

$$V\left(x,\mathcal{T}^1\right) = V(x), \quad W\left(x,\mathcal{T}^2\right) = W(x), \quad x \in X.$$

The *optimal switching problem* is to find an optimal policy and to evaluate the value functions V, W.

The following elements $\mathcal{T} \in \mathcal{S}_0$ play an important role in solving the optimal switching problem, and we introduce for them a special notation.

Definition 2.1 *For any two disjoint sets $F, G \in \mathcal{B}$, $\mathcal{T}_{FG}$ is an element of $\mathcal{S}_0$ in which $\tau_1 = \tau_F$, τ_2 is the first entrance time into G after τ_1, τ_3 is the first entrance time into F after τ_2, etc. in the alternating order.*

Remark 2.1 *If condition* (2.7) *fails to hold, the switching problem is still meaningful, but is of no independent interest.*

Indeed, in that case there is no sense to ever switch in the buying position, so that $w = 0$. In the selling position the problem becomes an *optimal stopping problem* with the reward function f. The solution of this problem under Assumptions 2.1–2.2 is well known (cf. references cited in the Introduction footnote). Namely, the value function v is a unique solution of the *optimality equation* $v = \max(Pv, f)$ with the boundary condition $v(x) = f(x)$, $x \in B$, and also is the *excessive envelope* of f, (i.e. the minimal function v with the properties $v \geq 0$, $Pv \leq v$ majorizing f). The *supporting set* $F = \{x : v(x) = f(x)\} \supset B$ defines an *optimal stopping time* $\tau^* = \tau_F$. Also, v is the unique *harmonic function* (i.e. a function with $Pv = v$) on the set $X \setminus F$ satisfying the boundary condition $v(x) = f(x)$, $x \in F$. All these features find their analogues in the optimal switching problem.

3. First imitating Markov decision process

To use the theory of Markov decision processes (MDPs), we imitate by them the switching problem of Section 2. This can be done in two ways. In the first MDP, let it be MDP1, the state of the Markov chain changes from x_t to x_{t+1} at every switching; in MDP2 the state x_t of the chain remains frozen while switching. In MDP2 all strategies of the switching problem are taken into account, and indeed we need the optimality equations of MDP2 to solve the switching problem. However, in MDP2 every stop requires an additional unit of time in comparison with the Markov chain, so that the time scales in these two processes become different, related in a random way, and this makes a formal description of the correspondence between the controls in them highly technical. Therefore we formally reduce our problem to MDP1, in spite of the fact that strategies σ with $\tau_{k+1} = \tau_k < \infty$ have no counterparts in it. Later, in Section 5, we transform the optimality equations of MDP1 into those of MDP2 by simple algebra.

An MDP is given by a state space, an action space, a transition function and a reward function[2]. The state space of MDP1 is $Y = X \times \{1,2\}$. For brevity, we use notations X^i for $X \times \{i\}$, and x^i for $y = (x,i) \in Y$, $i = 1,2$. The action space $A = \{0,1\}$, and both actions are admitted at every state y; here $a = 1$ corresponds to switching, $a = 0$ to nonswitching. Let

$$\tilde{\omega} = y_0 a_1 y_1 a_2 y_2 \ldots, \quad y_{t-1} \in Y, \quad a_t \in A, \tag{3.1}$$

be a path of MDP1, with $y_t = (x_t, i_t)$. The transition function is given by

$$\begin{aligned} \mathbf{P}\{x_{t+1} \in E,\ i_{t+1} = i | x_t = x,\ i_t = i,\ a_{t+1} = 0\} &= P(x,E), \\ \mathbf{P}\left\{x_{t+1} \in E,\ i_{t+1} = i + (-1)^{i-1} | x_t = x,\ i_t = i,\ a_{t+1} = 1\right\} &= P(x,E), \end{aligned} \tag{3.2}$$

$x \in X$, $E \in \mathcal{B}$, $i = 1,2$ (for other combinations of i_t, i_{t+1} and a_{t+1} the probabilities are zeros). In words, the x-component of $\{y_t\}$ develops precisely as the Markov chain $\{x_t\}$, while the i-component changes each time the action 1 is used. The reward function $r(y,a)$ is

$$\begin{aligned} &r(y,0) = 0 && y \in Y, \\ &r(x^1,1) = f(x), \qquad r(x^2,1) = -g(x), && x \in X. \end{aligned} \tag{3.3}$$

For the sample space $\tilde{\Omega}$ of MDP1 we take, for simplicity, only those paths (3.1) in which

(i) the i-component of y_{t+1} differs from the i-component of y_t if and only if $a_{t+1} = 1$, and

[2]In the terminology and notations we follow mostly Dynkin and Yushkevich [5]. Some other basic references on MDPs are Bertsekas and Shreve [1], Puterman [11], Hernández-Lerma and Lasserre [9]

(ii) similar to Ω, the x-component of y_{t+1} is equal to the x-component of y_t if the latter belongs to the boundary B.

A history h_t is any initial segment

$$h_t = y_0 a_1 \cdots y_t$$

of the path (3.1).

An arbitrary (in general, randomized and history dependent) policy π in MDP1 is defined by the (measurable in h_t) probabilities

$$p_{t+1}^{\pi}(h_t) = \mathbf{P}_y^{\pi}\{a_{t+1} = 1 | h_t\} \quad t = 0, 1, \ldots \tag{3.4}$$

(and the complimentary probabilities of the action 0), and this together with (3.2), as usually, determines the probability $\mathbf{P}_y^{\pi}$ in the space $\tilde{\Omega}$ corresponding to a policy π and an initial state $y \in Y$. The expectation corresponding to $\mathbf{P}_y^{\pi}$ is denoted $\mathbf{E}_y^{\pi} = \mathbf{E}_{x,i}^{\pi}$, $y = x^i$.

A policy π is called nonrandomized or deterministic, if the probabilities (3.4) assume only the values 0 and 1. Such a policy is specified by measurable functions $\pi_{t+1}(h_t)$ so that

$$a_{t+1} = \pi_{t+1}(h_t).$$

A stationary (deterministic) policy is determined by a measurable function $\varphi : Y \to A$ (a selector), so that $a_{t+1} = \varphi(y_t)$; such a policy we identify with φ. We denote by Π the set of all policies, and by Φ the set of all stationary policies (selectors). Let

$$\Phi_0 = \left\{\varphi \in \Phi : \varphi\left(x^1\right)\varphi\left(x^2\right) = 0,\ x \in X\right\}; \tag{3.5}$$

in words, $\varphi \in \Phi_0$ if, for every $x \in X$, the selector φ prescribes the action $a = 1$ at most at one of the states x^1, x^2. Each selector $\varphi \in \Phi$ is specified by *switching sets*

$$F_{\varphi} = \left\{x \in X : \varphi\left(x^1\right) = 1\right\}, \quad G_{\varphi} = \left\{x \in X : \varphi\left(x^2\right) = 1\right\}. \tag{3.6}$$

If $\varphi \in \Phi_0$, the sets F_{φ} and G_{φ} are disjoint.

The random total reward in MDP1 is

$$J(\tilde{\omega}) = \sum_{t=1}^{\infty} r_t = \sum_{t=1}^{\infty} r(y_{t-1}, a_t), \tag{3.7}$$

the expected rewards (if well defined) are

$$v(x, \pi) = \mathbf{E}_{x,1}^{\pi} J(\tilde{\omega}), \quad w(x, \pi) = \mathbf{E}_{x,2}^{\pi} J(\omega), \quad x \in X,\ \pi \in \Pi. \tag{3.8}$$

The *value functions* are

$$v(x) = \sup_{\pi \in \Pi} v(x, \pi), \quad w(x) = \sup_{\pi} w(x, \pi), \quad x \in X, \tag{3.9}$$

and a policy π is *optimal*, if $v(x, \pi) = v(x)$, $w(x, \pi) = w(x)$, $x \in X$.

By discarding in (3.1) all elements a_t and all components i of y_{t-1}, we get a natural mapping $\lambda : \tilde{\Omega} \to \Omega$, and we merely write ω instead of $\lambda(\tilde{\omega})$ where there should be no confusion. Functions on Ω, in particular the first entrance time $\tau_B(\omega)$, $\omega \in \Omega$ into the boundary, can be treated as functions on $\tilde{\Omega}$ with $\omega = \lambda(\tilde{\omega})$.

Lemma 3.1 *For every policy $\pi \in \Pi$ and initial state $y = x^i \in Y$, the $\boldsymbol{P}^{\pi}_{x,i}$-distribution on Ω induced by the mapping λ coincides with the $\boldsymbol{P}_x$-distribution as defined in* Section 2 *for the Markov chain $\{x_t\}$.*

Proof Follows directly from (3.2). ■

Lemma 3.2 *The expected rewards* (3.8) *are well-defined and measurable in x for every $\pi \in \Pi$. The value functions* (3.9) *satisfy bounds*

$$0 \leqslant w(x), \quad v(x) \leqslant C(1 + \mathbf{E}_x \tau_B), \quad x \in X. \tag{3.10}$$

Proof The sum of terms r_t in (3.7) over $1 \leqslant t \leqslant \tau_B(\omega)$ does not exceed $C\tau_B(\omega)$ (cf. (3.3) and (2.5)). The sum over $t > \tau_B$ is either 0 or of one of the forms $f(z) - g(z) + f(z) - g(z) + \cdots$ or $-g(z) + f(z) - \cdots$ (the number of terms may be finite or infinite); in any case it does not exceed C. Since $\mathbf{E}_x \tau_B < \infty$, and by Lemma 3.1, the expectations in (3.8) are well-defined and satisfy the upper bound in (3.10). By (3.9), the value functions satisfy the same bound. On the other hand, v and w are nonnegative because $v(x, \varphi) = w(x, \varphi) = 0$ for the stationary policy $\varphi(y) = 0$, $y \in Y$. ■

To prove that the value functions v, w are measurable and satisfy the Bellman optimality equations, we approximate them by value iterations. Let v_n, w_n be the expected rewards and value functions in the same model with a horizon n, i.e. with

$$J_n(\tilde{\omega}) = \sum_{t=1}^{n} r_t$$

instead of $J(\tilde{\omega})$ in formulas (3.8) and (3.9). Let T be the one-step Bellman operator defined on pairs (ξ, η) of nonnegative measurable functions on X by the formula

$$\begin{pmatrix} \xi' \\ \eta' \end{pmatrix} = T \begin{pmatrix} \xi \\ \eta \end{pmatrix} = \begin{pmatrix} \max(f + P\eta, P\xi) \\ \max(P\xi - g, P\eta) \end{pmatrix} \tag{3.11}$$

corresponding to the transition function (3.2) and the rewards (3.3). Evidently, T transforms such pairs (ξ, η) into similar pairs if we allow for nonnegative functions the value $+\infty$. It follows by induction from (3.2) and (3.3) (and general facts concerning optimality equations in MDPs with a Borel state space, a finite action space, and a bounded measurable reward function), that the value functions v_n, w_n are given by the formulas

$$\begin{pmatrix} v_0 \\ w_0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad \begin{pmatrix} v_{n+1} \\ w_{n+1} \end{pmatrix} = T\begin{pmatrix} v_n \\ w_n \end{pmatrix}, \quad n = 0, 1, 2, \ldots, \tag{3.12}$$

and are measurable. Here, by (3.11), $v_1 \geq Pv_1 = 0 = v_0$, $w_1 \geq Pw_0 = 0 = w_0$, and by the monotonicity of T

$$0 = v_0 \leqslant v_1 \leqslant v_2 \leqslant \cdots, \quad 0 = w_0 \leqslant w_1 \leqslant w_2 \leqslant \cdots. \tag{3.13}$$

Theorem 3.1 *The value functions v, w of MDP1 are equal to the limits*

$$v(x) = \lim_{n\to\infty} v_n(x), \quad w(x) = \lim_{n\to\infty} w_n(x), \quad x \in X, \tag{3.14}$$

and are measurable. They satisfy bounds (3.10), *optimality equations*

$$v = \max(f + Pw, Pv), \tag{3.15}$$
$$w = \max(Pv - g, Pw), \tag{3.16}$$

and boundary conditions

$$v(x) = f(x), \quad w(x) = 0, \quad x \in B. \tag{3.17}$$

Proof We prove (3.14) for v; the case of w is similar. Inequalities (3.13) imply the existence of the limit $\bar{v}(x) = \lim_{n\to\infty} v_n(x)$, $x \in X$. For arbitrary n and $\pi \in \Pi$, let π' be a policy equal to π at the initial steps $t = 1, 2, \ldots, n$, and assigning the action $a_t = 0$ at the steps $t > n$. Then

$$v_n(x, \pi) = v(x, \pi') \leqslant v(x), \qquad x \in X,$$

so that $v_n(x) \leqslant v(x)$, and hence $\bar{v} \leqslant v$.

To obtain the inverse inequality, observe that after the process $\{x_t\}$ reaches a point $z \in B$ at the time τ_B, the forthcoming reward is either 0, or $-\infty$, or a finite sum of alternating terms $f(z)$ and $-g(z)$; its maximum is $f(z)$ if $i_{\tau_B} = 1$ and is 0 if $i_{\tau_B} = 2$, and this maximum can be gathered at the first step of the control after the time τ_B. It follows that

$$v(x) = \sup_{\pi\in\Pi} v(x, \pi) = \sup_{\pi\in\Pi_0} v(x, \pi), \quad x \in X, \tag{3.18}$$

where Π_0 is the set of all policies for which $a_t = 0$ at the steps $t > \tau_B(\omega) + 1$, $\omega = \lambda(\tilde{\omega})$. For a policy $\pi \in \Pi_0$ all terms r_t in (3.7) with $t > \tau_B + 1$ are zeros, and since $|r| \leq C$, we have by Lemma 3.1

$$\mathbf{E}^{\pi}_{x,1} \sum_{n+1}^{\infty} |r_t| \leqslant C \sum_{t=n}^{\infty} \mathbf{P}_x \{\tau_B \geq t\}, \quad n = 0, 1, \ldots, \ \pi \in \Pi_0. \qquad (3.19)$$

For every policy $\pi \in \Pi$ evidently

$$v(x, \pi) \leqslant v_n(x) + \mathbf{E}^{\pi}_{x,1} \sum_{n+1}^{\infty} |r_t|,$$

and therefore (3.18) and (3.19) imply

$$v(x) \leqslant v_n(x) + C \sum_{t=n}^{\infty} P_x \{\tau_B \geq t\}, \quad n = 1, 2, \ldots, \ x \in X. \qquad (3.20)$$

Since $\mathbf{E}_x \tau_B < \infty$, the sum in (3.20) converges to 0 as $n \to \infty$, and in the limit (3.20) becomes $v \leqslant \bar{v}$. Thus, $v = \bar{v}$, and (3.14) is proved.

Measurability of v, w follows from (3.14) and Lemma 3.2. Relations (3.15)–(3.16) follow from the monotone convergence (3.14) and (3.11)–(3.12). Boundary conditions (3.17) follow from (2.1), (2.5)–(2.6) and (3.3). ■

4. Correspondence between strategies and policies

To show that a solution of MDP1 provides a solution to the switching problem, we establish a correspondence between some strategies and policies under which the expected rewards do not change. We perform this for classes of strategies and policies, sufficient to approximate the value functions.

Lemma 4.1 *To every strategy $\sigma = (\mathcal{T}^1, \mathcal{T}^2)$ with $\mathcal{T}^i \in \mathcal{S}_0$, $i = 1, 2$, there corresponds a nonrandomized policy π such that*

$$v(x, \pi) = V(x, \mathcal{T}^1), \quad w(x, \pi) = W(x, \mathcal{T}^2), \quad x \in X. \qquad (4.1)$$

Proof We construct the functions π_{t+1} for histories $h_t = y_0 a_1 \cdots y_t$ with $y_0 = (x_0, 1)$, so that the first of the relations (4.1) holds; the second, corresponding to $y_0 = (x_0, 2)$, is treated similarly.

We refer to the correspondence $\omega = \lambda(\tilde{\omega})$ and Lemma 3.1. Given the component $\mathcal{T}^1 = \{\tau_1 < \tau_2 < \cdots\}$ of the strategy σ, we define a mapping $\mu : \Omega \to \tilde{\Omega}$ by setting

$$\mu(\omega) = \mu(x_0 x_1 x_2 \cdots) = y_0 a_1 y_1 a_2 y_2 \cdots$$

where

$$\begin{cases} y_0 = (x_0, 1), & \\ a_t = 0, & y_t = (x_t, 1) \quad \text{if } 0 < t \leqslant \tau_1(\omega), \\ a_t = 1, & y_t = (x_t, 2) \quad \text{if } t = \tau_1(\omega) + 1, \\ a_t = 0, & y_t = (x_t, 2) \quad \text{if } \tau_1(\omega) + 1 < t \leqslant \tau_2(\omega), \\ a_t = 1, & y_t = (x_t, 1) \quad \text{if } t = \tau_2(\omega) + 1, \\ a_t = 0, & y_t = (x_t, 1) \quad \text{if } \tau_2(\omega) + 1 < t \leqslant \tau_3(\omega), \\ \dots, & \end{cases} \tag{4.2}$$

until we cover all $t = 0, 1, 2, \dots$. The paths $\tilde{\omega} \in \mu(\Omega)$ we call *marked* paths; marked histories h_t are initial segments of marked paths.

For a marked history h_t we set $\pi_{t+1}(h_t) = a_{t+1}$ where a_{t+1} is given in (4.2). In general, a marked history can belong to different marked paths. However, since τ_k are stopping times with respect to the minimal filtration in Ω, a_{t+1} is uniquely defined by $x_0, x_1 \dots, x_t$ (together with all $\tau_k \leq t$). For every unmarked history h_t we set, to be definite, $\pi_{t+1}(h_t) = 0$. The same argument shows that π_{t+1} are measurable functions of histories h_t. Thus $\pi = \{\pi_1, \pi_2, \dots\}$ is a nonrandomized policy.

By the construction (4.2) and by the definition of the reward $J(\tilde{\omega})$ (see (3.3) and (3.7)), we have for a marked path

$$J(\tilde{\omega}) = f(x_{\tau_1}) - g(x_{\tau_2}) + f(x_{\tau_3}) - \dots = J_1(\omega) \tag{4.3}$$

if $\tilde{\omega} = \mu(\omega)$ (see (2.9)). Evidently, $\lambda(\mu(\omega)) = \omega$ for every $\omega \in \Omega$. Also, from the construction (4.2) and by induction in t, it is easy to see that $\mathbf{P}^{\pi}_{x,1}\{\mu(\Omega)\} = 1$. Hence, by Lemma 3.1 and (4.3), $\mathbf{E}^{\pi}_{x,1} J(\tilde{\omega}) = \mathbf{E}_x J(\omega)$. ■

Lemma 4.2 *To every selector $\varphi \in \Phi_0$ there corresponds a strategy $\sigma = (\mathcal{T}^1, \mathcal{T}^2)$ with $\mathcal{T}^i \in \mathcal{S}_0$, $i = 1, 2$, such that*

$$V(x, \mathcal{T}^1) = v(x, \varphi), \quad W(x, \mathcal{T}^2) = w(x, \varphi), \quad x \in X;$$

namely, one may set $\mathcal{T}^1 = \mathcal{T}_{FG}$, $\mathcal{T}^2 = \mathcal{T}_{GF}$ where $F = F_\varphi$, $G = G_\varphi$ (see Definition 2.1*).*

Proof For $\mathcal{T}^1 = \mathcal{T}_{FG}$ and $a_{t+1} = \varphi(y_t)$ (i.e. $\pi_{t+1} = \varphi$) we have the same correspondence (4.2) between the marked paths and histories on one side and sample points ω on the other, as in the proof of Lemma 4.1, with the $\mathbf{P}^{\pi}_{x,1}$-probability of the set $\tilde{\Omega} \backslash \mu(\Omega)$ of non-marked paths equal to 0. Thus the concluding part of the proof of Lemma 4.1 extends to the present case, and $V(x, \mathcal{T}_{FG}) = v(x, \varphi)$. The case of $W(x, \mathcal{T}_{GF})$ is similar. ■

Remark 4.1 *It follows from* Lemmas 2.2 *and* 4.1, *that* $v \geq V$, $w \geq W$. *If*

$$\sup_{\varphi \in \Phi_0} v(x, \varphi) = \sup_{\pi \in \Pi} v(x, \pi), \ \sup_{\varphi \in \Phi_0} w(x, \varphi) = \sup_{\pi \in \Pi} w(x, \pi), \quad x \in X, \tag{4.4}$$

then also $v \leq V$, $w \leq W$, *so that indeed* $v = V$, $w = W$, *and an optimal selector* $\varphi \in \Phi_0$ *generates an optimal strategy* $\sigma = (\mathcal{T}_{FG}, \mathcal{T}_{GF})$ *(see* Lemma 4.2*).*

Relations (4.4) will be justified in Section 7.

5. Second imitating Markov decision process

In this section we show that optimality equations (3.15)–(3.16) of MDP1 are equivalent to optimality equations of MDP2, and obtain a characterization of the value functions in terms of excessive envelopes.

MDP1 was formally defined in the second paragraph of Section 3. The definition of MDP2 coincides with the above definition except at one point: the second of formulas (3.2) should be replaced by

$$\mathbf{P}\left\{x_{t+1} \in E, \ i_{t+1} = i + (-1)^{i-1} | x_t = x, \ i_t = i, \ a_{t+1} = 1\right\}$$
$$= \begin{cases} 1 & \text{if } x \in E, \\ 0 & \text{if } x \notin E. \end{cases}$$

In words, if the action $a = 1$ is used in a state $y = x^i$ in MDP2, then the system moves with probability 1 to the state x^j, while in MDP1 it moves to a state z^j, where z has the distribution $P(x, \cdot)$; in both cases j is different from i. We have no need to analyze MDP2 in detail, as we did in Section 3 with MDP1. However, equations (5.1)–(5.2), formally obtained below for the value functions of MDP1, and crucial for the optimal switching problem, are indeed Bellman equations of MDP2. In fact, we got them originally from MDP2 by a naive dynamic programming reasoning.

Theorem 5.1 *For finite, nonnegative measurable functions* v *and* w *on* X, *equations* (3.15)–(3.16) *imply the equations*

$$v = \max(f + w, Pv), \tag{5.1}$$
$$w = \max(v - g, Pw), \tag{5.2}$$

and vice versa. In particular, (5.1)–(5.2) *are true for the value functions of MDP1.*

Proof First assume (3.15)–(3.16). Fix $x_0 \in X$, and to simplify the writing, skip x_0 in $f(x_0)$, $v(x_0)$, $w(x_0)$, $Pv(x_0)$, etc. There are 4 cases compatible with (3.15)–(3.16).

Case 1 $v = Pv$, $w = Pw$.

Then (5.1)–(5.2) coincides with (3.15)–(3.16).

Case 2 $v > Pv$, $w > Pw$.

Then from (3.15) $v = f + Pw < f + w$ and from (3.16) $w = Pv - g < v - g$. Hence $v < f + v - g$ or $g < f$, and we have a contradiction with (2.4).

Case 3 $v = Pv$, $w > Pw$.

Here (5.2) coincides with (3.16). Since $w > Pw$, from (5.2) we get $w = v - g$, hence $v = w + g \geqslant w + f$, and this together with $v = Pv$ proves (5.1).

Case 4 $v > Pv$, $w = Pw$.

Now (5.1) follows from (3.15). Since $v > Pv$, (5.1) implies $v = f + w \leqslant g + w$, and this together with $w = Pw$ proves (5.2).

Now assume (5.1)–(5.2). We have the same 4 cases.

Case 1 Is again trivial.

Case 2 We get from (5.1)–(5.2) $v = f + w = f + v - g$, thus $f = g$, and this contradicts to (2.4).

Case 3 (3.16) holds automatically, while (3.15) reduces to $v \geqslant f + Pw$. This holds because $w > Pw$ and $v \geqslant f + w$ by (5.1).

Case 4 (3.15) holds automatically, and (3.16) reduces to $w \geqslant Pv - g$. The last inequality holds because $Pv < v$ and $w \geqslant v - g$ by (5.2).

The last assertion of the theorem follows from Theorem 3.1. ■

Let U be the Bellman operator corresponding to the optimality equations (5.1)–(5.2): for nonnegative measurable function ξ, η on X

$$\begin{pmatrix} \xi' \\ \eta' \end{pmatrix} = U \begin{pmatrix} \xi \\ \eta \end{pmatrix} = \begin{pmatrix} \max(f + \eta, P\xi) \\ \max(\xi - g, P\eta) \end{pmatrix} \tag{5.3}$$

(cf. the definition (3.11) of T). U is also a monotone operator, and analogous to

$$\begin{pmatrix} v_n \\ w_n \end{pmatrix} = T^n \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad n = 0, 1, 2, \ldots \tag{5.4}$$

(cf. (3.12)), we set

$$\begin{pmatrix} \bar{v}_n \\ \bar{w}_n \end{pmatrix} = U^n \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad n = 0, 1, 2, \ldots. \tag{5.5}$$

Lemma 5.1 *If* $\xi, \eta \geqslant 0$ *and*

$$\begin{pmatrix}\xi'\\ \eta'\end{pmatrix} = U\begin{pmatrix}\xi\\ \eta\end{pmatrix} \geqslant \begin{pmatrix}\xi\\ \eta\end{pmatrix} \tag{5.6}$$

then

$$U^2\begin{pmatrix}\xi\\ \eta\end{pmatrix} \geqslant T\begin{pmatrix}\xi\\ \eta\end{pmatrix}.$$

Proof By (5.3) $\eta' \geqslant P\eta$ and $\xi' \geqslant P\xi$, by (5.6) $P\xi' \geqslant P\xi$, $P\eta' \geqslant P\eta$. Thus

$$\begin{aligned} U^2\begin{pmatrix}\xi\\ \eta\end{pmatrix} &= U\begin{pmatrix}\xi'\\ \eta'\end{pmatrix} \\ &= \begin{pmatrix}\max(f+\eta', P\xi')\\ \max(\xi'-g, P\eta')\end{pmatrix} \\ &\geqslant \begin{pmatrix}\max(f+P\eta, P\xi)\\ \max(P\xi-g, P\eta)\end{pmatrix} \\ &= T\begin{pmatrix}\xi\\ \eta\end{pmatrix}. \end{aligned}$$

■

Lemma 5.2 *The functions* $\bar{v}_n$, $\bar{w}_n$ *defined in* (5.5) *are nondecreasing in* n *and converge to the value functions of MDP1:*

$$\bar{v}_n \uparrow v, \qquad \bar{w}_n \uparrow w. \tag{5.7}$$

Proof By (5.5) and (5.3)

$$\begin{pmatrix}\bar{v}_1\\ \bar{w}_1\end{pmatrix} = U\begin{pmatrix}0\\ 0\end{pmatrix} \geqslant \begin{pmatrix}0\\ 0\end{pmatrix} = \begin{pmatrix}\bar{v}_0\\ \bar{w}_0\end{pmatrix}, \tag{5.8}$$

and therefore, since U is a monotone operator, $\bar{v}_{n+1} \geqslant \bar{v}_n$, $\bar{w}_{n+1} \geqslant \bar{w}_n$. Multiplying by U^n the inequality

$$\begin{pmatrix}0\\ 0\end{pmatrix} \leqslant \begin{pmatrix}v\\ w\end{pmatrix}$$

we get by (5.1)–(5.2)

$$\begin{pmatrix}\bar{v}_n\\ \bar{w}_n\end{pmatrix} = U^n\begin{pmatrix}0\\ 0\end{pmatrix} \leqslant U^n\begin{pmatrix}v\\ w\end{pmatrix} = \begin{pmatrix}v\\ w\end{pmatrix}. \tag{5.9}$$

On the other hand, Lemma 5.1 and (5.8) imply the inequality

$$\begin{pmatrix}\bar{v}_2\\ \bar{w}_2\end{pmatrix} = U^2\begin{pmatrix}0\\ 0\end{pmatrix} \geqslant T\begin{pmatrix}0\\ 0\end{pmatrix} = \begin{pmatrix}v_1\\ w_1\end{pmatrix} \tag{5.10}$$

(see (5.4)). Similarly, Lemma 5.1 and the inequalities $\bar{v}_3 \geqslant \bar{v}_2$, $\bar{w}_3 \geqslant \bar{w}_2$, and after that (5.10), imply

$$\begin{pmatrix} \bar{v}_4 \\ \bar{w}_4 \end{pmatrix} = U^2 \begin{pmatrix} \bar{v}_2 \\ \bar{w}_2 \end{pmatrix} \geqslant T \begin{pmatrix} \bar{v}_2 \\ \bar{w}_2 \end{pmatrix} \geqslant T \begin{pmatrix} v_1 \\ w_1 \end{pmatrix} = \begin{pmatrix} v_2 \\ w_2 \end{pmatrix},$$

so that by an evident induction $\bar{v}_{2n} \geqslant v_n$, $\bar{w}_{2n} \geqslant w_n$. This together with (5.9) and (3.14) proves (5.7). ■

Theorem 5.2 *The pair (v, w) of the value functions of MDP1 is the minimal nonnegative measurable solution of the inequalities*

$$\begin{cases} v \geqslant \max(f + w, Pv), \\ w \geqslant \max(v - g, Pw). \end{cases} \tag{5.11}$$

Proof For any pair $(\tilde{v}, \tilde{w}) \geqslant 0$ satisfying (5.11) we have, in notations (5.5),

$$\begin{pmatrix} \bar{v}_0 \\ \bar{w}_0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \leqslant \begin{pmatrix} \tilde{v} \\ \tilde{w} \end{pmatrix}. \tag{5.12}$$

From (5.5), (5.12), and (5.11) applied to $(\tilde{v}, \tilde{w})$, by an evident induction in n

$$\begin{pmatrix} \bar{v}_n \\ \bar{w}_n \end{pmatrix} = U \begin{pmatrix} \bar{v}_{n-1} \\ \bar{w}_{n-1} \end{pmatrix} \leqslant U \begin{pmatrix} \tilde{v} \\ \tilde{w} \end{pmatrix} \leqslant \begin{pmatrix} \tilde{v} \\ \tilde{w} \end{pmatrix},$$

and by Lemma 5.2 $v \leq \tilde{v}$, $w \leq \tilde{w}$. ■

In other words, $(v,\ w)$ *is the minimal pair of excessive functions satisfying inequalities*

$$v \geqslant w + f, \quad w \geqslant v - g.$$

Remark 5.1 *Similar to the reasoning used in* Theorem 3.1, *one may show that v, w are also the value functions of MDP2.*

6. Preference function

We define the *preference function* as

$$u(x) = v(x) - w(x), \quad x \in X. \tag{6.1}$$

Since indeed the value functions of the switching problem and MDP1 are equal (Theorem 7.2 below), the preference function shows to what extent the selling position is more advantageous than the buying position at any state x.

From now on, it is convenient to use, instead of the transition operator P, the *generator* A of the Markov chain $\{x_t\}$:

$$A = P - I \tag{6.2}$$

where I is the identity operator. Harmonic functions are solutions of the equation $Ah = 0$, excessive functions are nonnegative functions h with $Ah \leqslant 0$.

Theorem 6.1 *The preference function u is the unique bounded measurable solution of the inequalities*

$$f(x) \leqslant u(x) \leqslant g(x), \qquad x \in X, \tag{6.3}$$
$$Au(x) \geqslant 0 \qquad \text{if } f(x) < u(x), \tag{6.4}$$
$$Au(x) \leqslant 0 \qquad \text{if } u(x) < g(x), \tag{6.5}$$

together with the boundary condition

$$u(x) = f(x), \quad x \in B. \tag{6.6}$$

Relations (6.3)–(6.5) are known in some stochastic games, the so-called Dynkin games (see Section 9 for more details). In connection with those games, the uniqueness of the solution to the system (6.3)–(6.6) was proved at various levels of generality, including the continuous-time case. For completeness of the paper, and since we do not have a proper reference covering the discrete-time case with a Borel state space, we present a simple proof of the uniqueness too.

Proof The measurability of u follows from (6.1) and Theorem 3.1. By Theorems 3.1 and 5.1, v and w satisfy equations (5.1)–(5.2). Subtracting w from both sides of (5.1) and using (6.2) on one hand, and multiplying (5.2) by -1, adding v and using (6.2) on the other, we get

$$\max(f, Av + u) = u = \min(g, u - Aw). \tag{6.7}$$

Now (6.3) follows immediately from (6.7), and (6.3) shows that u is bounded (cf. (2.5)). If $f(x) < u(x)$ at some point x, then by (6.7) $Av(x) + u(x) = u(x)$, so that $Av(x) = 0$; on the other hand, $Aw \leqslant 0$ everywhere by (5.1), and hence $Au(x) = Av(x) - Aw(x) \geqslant 0$. This proves (6.4). To get (6.5), use that $Av \leqslant 0$ by (5.1) and that $Aw(x) = 0$ if $u(x) < g(x)$ by (6.7). The boundary condition (6.6) follows from the boundary conditions (3.17) for v and w.

To prove the uniqueness, we observe that for any $u \in B(X)$ the process $\{\xi_t\}$:

$$\xi_0 = u(x_0), \quad \xi_t = u(x_t) - \sum_{k=0}^{t-1} Au(x_k), \quad (t \geqslant 1) \tag{6.8}$$

is a martingale (with respect to the minimal filtration $\{\mathcal{N}_t\}$ generated by the Markov chain $\{x_t\}$, see Section 2). Indeed, by the Markov property and (6.2), for any $t \geq 0$ and $x \in X$

$$\begin{aligned}
\mathbf{E}_x [\xi_{t+1} | \mathcal{N}_t] &= \mathbf{E}_x [u(x_{t+1}) | \mathcal{N}_t] - \sum_0^t Au(x_k) \\
&= Pu(x_t) - \sum_0^t Au(x_k) \\
&= u(x_t) - \sum_0^{t-1} A(u(x_k)) \\
&= \xi_t.
\end{aligned}$$

By Doob's theorem, for a stopping time τ

$$\mathbf{E}_x \xi_\tau = \xi_0 = u(x), \quad x \in X, \tag{6.9}$$

if only

$$\mathbf{E}_x |\xi_\tau| < \infty, \qquad \lim_{t \to \infty} \mathbf{E}_x |\xi_t| \mathbf{1}\{\tau > t\} = 0. \tag{6.10}$$

(See, for instance, Shiryayev [13]). We verify (6.10) for any stopping time $\tau \leqslant \tau_B$. By (6.3) and (2.4), $|u| \leqslant C$, hence by (6.2) $|Au| \leqslant 2C$, and therefore by (2.2)

$$\mathbf{E}_x |\xi_\tau| \leqslant C(1 + 2\mathbf{E}_x \tau_B) < \infty, \quad x \in X.$$

Also,

$$\mathbf{E}_x |\xi_t| \mathbf{1}\{\tau > t\} \leqslant C(2t+1) \mathbf{P}_x \{\tau_B > t\} \to 0 \quad \text{as } t \to \infty, \ x \in X,$$

because the expectation

$$\mathbf{E}_x \tau_B = \sum_1^\infty n \mathbf{P}\{\tau_B = n\}$$

converges and

$$t\mathbf{P}_x\{\tau_B > t\} \leqslant \sum_{n=t+1}^\infty n\mathbf{P}_x\{\tau_B = n\}.$$

Thus, (6.9) holds for any $\tau \leq \tau_B$.

We now suppose that there are two different solutions u_1 and u_2 to (6.3)–(6.6), and get a contradiction. Assume that $u_1(z) < u_2(z)$ at some $z \in X$. Let

$$D = \{x \in X : u_1(x) \geq u_2(x)\}. \tag{6.11}$$

The set D is measurable together with u_1 and u_2, so τ_D is a stopping time. By (6.6), $u_1 = u_2$ on the boundary, so that $B \subset D$, $\tau_D \leq \tau_B$, and therefore (6.9) is applicable to $\tau = \tau_D$. On the set $E = X \setminus D$ we have by (6.3) and (6.11)

$$f(x) \leqslant u_1(x) < u_2(x) \leqslant g(x),$$

hence by (6.4) and (6.5)

$$Au_2(x) \geqslant 0, \quad Au_1(x) \leqslant 0, \quad x \in E. \tag{6.12}$$

Since $x_t \in E$ for $0 \leqslant t < \tau_D$ and $x_t \in D$ for $t = \tau_D$, we obtain from (6.8), (6.9), (6.11) and (6.12) for $\tau = \tau_D$

$$u_1(z) \geqslant \mathbf{E}_z u_1(x_\tau) \geqslant \mathbf{E}_z u_2(x_\tau) \geqslant u_2(z)$$

in contradiction with the assumption $u_1(z) < u_2(z)$. ■

Another description of the preference function u can be given in terms of the *support sets*

$$F = \{x \in X : u(x) = f(x)\}, \quad G = \{x \in X : u(x) = g(x)\}. \tag{6.13}$$

Corollary 6.1 *The support sets F and G are disjoint, measurable, and F contains the boundary B and therefore is nonempty. The preference function u is the unique bounded measurable solution of the equation*

$$Au(x) = 0, \quad x \in X \setminus (F \cup G), \tag{6.14}$$

with boundary conditions

$$u(x) = f(x), \quad x \in F, \qquad u(x) = g(x), \quad x \in G. \tag{6.15}$$

Moreover,

$$Au(x) \leq 0, \quad x \in F, \qquad Au(x) \geq 0, \quad x \in G. \tag{6.16}$$

Proof Theorem 6.1 and the condition $f < g$ imply all the assertions except the uniqueness of the solution to (6.14)–(6.15). The latter follows from the representation of any bounded solution of (6.14) in the form

$$u(x) = \mathbf{E}_x u(x_\tau), \quad \tau = \tau_{F \cup G}, \quad x \in X \setminus (F \cup G),$$

obtained from (6.8), (6.9), and (6.14), combined with (6.15). Formula (6.9) is applicable because $B \subset F \cup G$ so that $\mathbf{E}_x \tau \leqslant \mathbf{E}_x \tau_B < \infty$. ■

In other words, u is the unique bounded function harmonic in $X \setminus (F \cup G)$ and satisfying boundary conditions (6.15).

7. Optimal policy and optimal strategy

We now show that the stationary policy

$$\varphi(x,i) = \begin{cases} 1 & \text{if } i=1,\ x \in F \quad \text{or} \quad i=2,\ x \in G, \\ 0 & \text{otherwise,} \end{cases} \tag{7.1}$$

where F and G are the support sets (6.13), is optimal in MDP1.

To shorten formulas, we write v^φ, w^φ in place of $v(\cdot,\varphi)$, $w(\cdot,\varphi)$. Consider in parallel to the Bellman operator T (see (3.11)) a similar operator T^φ corresponding to φ:

$$T^\varphi \begin{pmatrix} \xi \\ \eta \end{pmatrix}(x) = \begin{cases} \begin{pmatrix} f(x) + P\eta(x) \\ P\eta(x) \end{pmatrix} & \text{if } x \in F, \\ \begin{pmatrix} P\xi(x) \\ P\xi(x) - g(x) \end{pmatrix} & \text{if } x \in G, \\ \begin{pmatrix} P\xi(x) \\ P\eta(x) \end{pmatrix} & \text{if } x \in X \setminus (F \cup G). \end{cases} \tag{7.2}$$

Lemma 7.1 *The value functions v and w satisfy equations*

$$v(x) = \begin{cases} f(x) + w(x) = f(x) + Pw(x) & \text{if } x \in F, \\ Pv(x) & \text{if } x \in X \setminus F, \end{cases} \tag{7.3}$$

$$w(x) = \begin{cases} v(x) - g(x) = Pv(x) - g(x) & \text{if } x \in G, \\ Pw(x) & \text{if } x \in X \setminus G. \end{cases} \tag{7.4}$$

Proof Relations (7.3) with $w(x)$ and (7.4) with $v(x)$ follow directly from formulas (5.1)–(5.2) of Theorem 5.1, the definition (6.13) of the sets F and G, and the equation $u = v - w$. If $x \in F$, then $x \notin G$, therefore $w(x) = Pw(x)$ by the already proven part of (7.4), and hence we may replace $w(x)$ by $Pw(x)$ in (7.3). Similarly, if $x \in G$, then $v(x) = Pv(x)$ by (7.3), and this proves the remaining part of (7.4). ∎

Lemma 7.2 *The operator T^φ is conserving in MDP1, i.e.*

$$T^\varphi \begin{pmatrix} v \\ w \end{pmatrix} = \begin{pmatrix} v \\ w \end{pmatrix}. \tag{7.5}$$

Proof Evaluate the left side of (7.5) using (7.2) and compare with (7.3)–(7.4). ∎

Lemma 7.3 *The policy φ is equalizing in MDP1, i.e. with $x_n = x(y_n)$*

$$\lim_{n\to\infty} \mathbf{E}_y^\varphi \left[v(x_n)\,\mathbf{1}\{i_n = 1\} + w(x_n)\,\mathbf{1}\{i_n = 2\} \right] = 0, \quad y \in Y. \tag{7.6}$$

Proof We suppose that $y = (x, 1)$; the case of $y = (x, 2)$ is similar. If (7.6) does not hold, then, since $v \geqslant 0$ by Lemma 3.2, there exist $x \in X$ and $\varepsilon > 0$ such that

$$\mathbf{E}^{\varphi}_{x,1}\left[v(x_n)\mathbf{1}\{i_n = 1\} + w(x_n)\mathbf{1}\{i_n = 2\}\right] \geqslant 2\varepsilon, \quad n \in N = N(x, \varepsilon), \tag{7.7}$$

where the set $N = \{n_1 < n_2 < \cdots\}$ of integers is infinite. As in the proof of Theorem 3.1 (see (3.19)), it follows from Assumption 2.1 that the tail of the rewards converges to 0 uniformly in policies $\pi \in \Pi_0$, and therefore there exists an integer n_0 such that

$$\mathbf{E}^{\pi}_{x,1}\sum_{n}^{\infty}|r_{t+1}| < \varepsilon \qquad \text{for } n \geqslant n_0,\ \pi \in \Pi_0. \tag{7.8}$$

Let n be fixed, $n_0 \leqslant n \in N$, so that both (7.7) and (7.8) hold. According to general results on upper summable Borelian MDPs, for any $\varepsilon > 0$ and any probability measure μ on the state space X, there exists an (a.e. μ) ε-optimal policy; see, for example, Dynkin and Yushkevich [5, Chapters 3 and 5] (formally the MDP we consider is summable in the sense of Lemma 3.2, different from the upper or lower summability assumed in Dynkin and Yushkevich [5], but this causes no impact on the applicability of the general measurable selection theorems). Let $\sigma' \in \Pi$ be such a policy for the measure $\mu(dz) = \mathbf{P}_x\{x_n \in dz\}$. The same reasoning as used in the proof of (3.18) shows that σ' can be adjusted to a policy $\sigma \in \Pi_0$ without diminishing the random and hence the expected rewards. So we have a policy $\sigma \in \Pi_0$ with

$$\begin{cases} v(z,\sigma) & \geqslant & v(z) - \varepsilon & \text{(a.e. } \mu), \\ w(z,\sigma) & \geqslant & w(z) - \varepsilon & \text{(a.e. } \mu), \end{cases} \tag{7.9}$$

where $z \in X$.

Consider now a policy $\pi = \varphi^n\sigma$; this policy coincides with φ on the n initial steps of the control, and after that coincides with σ; in the notations of Section 3, $p^{\pi}_{n+t+1}(h_{t+n}) = p^{\sigma}_{t+1}\left(h'_t\right)$ where h'_t is obtained from h_{n+t} by erasing the initial elements $y_0 a_1 \dots y_n$. Since $\sigma \in \Pi_0$, also $\pi \in \Pi_0$, and (7.8) holds. On the other hand, by the structure of the policy π, Lemma 3.1 and the Markov property of the chain $\{x_t\}$,

$$\mathbf{E}^{\pi}_{x,1}\left(\sum_{n}^{\infty} r_{t+1}\right) = \int_X w(z,\sigma)\,\mu_2(dz) + \int_X v(z,\sigma)\,\mu_1(dz) \tag{7.10}$$

where

$$\mu_i(dz) = \mathbf{P}^{\pi}_{x,1}\{i_n = i,\ x_n \in dz\}, \quad i = 1, 2, \tag{7.11}$$

so that $\mu_1 + \mu_2 = \mu$. Using firstly (7.9) and (7.10), and secondly (7.11) and (7.7), we obtain

$$\begin{aligned}
\mathbf{E}^{\pi}_{x,1}\left(\sum_{r}^{\infty} r_{t+1}\right) &\geqslant \int_X v(z)\,\mu_1(dz) + w(z)\,\mu_2(dz) - \varepsilon \\
&= \mathbf{E}^{\varphi}_{x,1}\big[v(x_n)\,\mathbf{1}\{i_n = 1\} + w(x_n)\,\mathbf{1}\{i_n = 2\}\big] - \varepsilon \\
&\geqslant 2\varepsilon - \varepsilon \;=\; \varepsilon.
\end{aligned}$$

This contradicts to (7.8), and we are done. ■

Theorem 7.1 *The stationary policy* φ *defined in* (7.1) *is an optimal policy in MDP1.*

Proof Follows from Lemmas 7.2 and 7.3 and the general fact in the theory of MDPs that a conserving and equalizing stationary policy is optimal. Namely

$$\begin{pmatrix} v \\ w \end{pmatrix} = (T^{\varphi})^n \begin{pmatrix} v \\ w \end{pmatrix}$$

from (7.5), and this equation means that

$$v(x) = \mathbf{E}^{\varphi}_{x,1}\left[\sum_{1}^{n} r_t + v(x_n)\,\mathbf{1}\{i_n = 1\} + w(x_n)\,\mathbf{1}\{i_n = 2\}\right], \quad (7.12)$$

$$w(x) = \mathbf{E}^{\varphi}_{x,2}\left[\sum_{1}^{n} r_t + v(x_n)\,\mathbf{1}\{i_n = 1\} + w(x_n)\,\mathbf{1}\{i_n = 2\}\right], \quad (7.13)$$

(as follows from (7.2) and the structure (3.3) of rewards). Due to (7.6), in the limit (7.12) and (7.13) turn into $v(x) = v^{\varphi}(x)$ and $w(x) = w^{\varphi}(x)$. ■

We now return to the switching problem and refer to Definition 2.1 for notations.

Theorem 7.2

(i) The value functions (2.15)–(2.16) *of the switching problem and* (3.9) *of MDP1 (and hence also of MDP2) are equal:* $V = v$, $W = w$.

(ii) The strategy $\sigma = (\mathcal{T}_{FG}, \mathcal{T}_{GF})$ *is optimal in the switching problem.*

Proof

(i) Since the sets F and G are disjoint, the policy φ in (7.1) belongs to the class Φ_0 (see (3.5)). Also, φ is optimal in MDP1, and it remains to refer to Remark 4.1.

(ii) By Lemma 4.2 and (i) $V(x, \mathcal{T}_{FG}) = v(x, \varphi) = v(x) = V(x)$, $x \in X$. Similarly, $W(\cdot, \mathcal{T}_{GF}) = W$. ∎

8. The value functions

Theorems 3.1 and 5.2 give an implicit characterization of the value functions $V = v$ and $W = w$. In this section we get explicit formulas in terms of the preference function u and the support sets F and G.

Let

$$Q(x, D) = \mathbf{P}_x\{x_{\tau_B} \in D\}, \qquad x \in X, \quad D \subset B, \ D \in \mathcal{B}, \tag{8.1}$$

be the exit measures corresponding to the Markov process $\{x_t\}$; by Assumption 2.1 $Q(x, B) = 1$ for every x. Consider also the occupational measures

$$R(x, D) = \sum_{t=0}^{\infty} \mathbf{P}_x\{x_t \in D\}, \qquad x \in X, \ D \in \mathcal{B}, \tag{8.2}$$

and the corresponding operator R (the resolvent for $\lambda = 1$) defined for any nonnegative measurable function h on X:

$$Rh(x) := \int_X h(y) R(x, dy) = \sum_{t=0}^{\infty} P^t h(x), \qquad x \in X. \tag{8.3}$$

Mention that $R(x, X_0) = \mathbf{E}_x \tau_B < \infty$ for every x by Assumption 2.1, and that if h is bounded and $h = 0$ on the boundary B, then $Rh(x)$ is finite for every x.

Lemma 8.1 *We have*

$$\lim_{n\to\infty} P^n w(x) = 0, \qquad \lim_{n\to\infty} P^n v(x) = Qf(x) := \int_B f(y) Q(x, dy), \quad x \in X. \tag{8.4}$$

Proof Given x, consider the corresponding state $y = (x, 1)$ of MDP1. By Lemma 3.1, the following formula is valid for any policy π in MDP1, in particular for the optimal policy φ defined in (7.1):

$$P^n w(x) = \mathbf{E}_y^{\varphi}\big[w(x_n)\mathbf{1}\{i_n = 1\} + w(x_n)\mathbf{1}\{i_n = 2\}\big].$$

Now, since $w(z) = 0$ on the boundary B, and since everywhere $| v - w | = | u | \leq C$ (as follows from (6.3) and (2.5)),

$$w(z)\mathbf{1}\{i_n = 1\} \leq \begin{cases} 0 \leq v(z) & \text{if } z \in B, \\ v(z)\mathbf{1}\{i_n = 1\} + C & \text{if } z \in X_0. \end{cases}$$

Therefore

$$P^n w(x) \leq \mathbf{E}_y^{\varphi}\big[v(x_n)\mathbf{1}\{i_n = 1\} + w(x_n)\mathbf{1}\{i_n = 2\}\big] + C\mathbf{P}_y^{\varphi}\{x_n \in X_0\}.$$

Here the expectation tends to zero by (7.7), and the probability tends to zero by Assumption 2.1. Since $w \geq 0$, this proves (8.4) for w. For v we have

$$P^n v(x) = \mathbf{E}_x\big[v(x_n)\mathbf{1}\{x_n \in B\}\big] + \mathbf{E}_x\big[v(x_n)\mathbf{1}\{x_n \in X_0\}\big]. \tag{8.5}$$

Using the inequality $| v - w | \leq C$, we obtain

$$\mathbf{E}_x\big[v(x_n)\mathbf{1}\{x_n \in X_0\}\big] \leq \mathbf{E}_x\big[w(x_n)\mathbf{1}\{x_n \in X_0\}\big] + C\mathbf{P}_x\{x_n \in X_0\}.$$

As in the case of w, both terms here converge to zero as $n \to \infty$. Since $v \geq 0$, the last expectation in (8.5) tends to 0. On the boundary B we have $v = f$, and because all boundary states are absorbing states, and x_n reaches the boundary at the state x_{τ_B}, the first expectation in (8.5) converges to the integral in (8.4). ∎

Consider now the functions

$$h(x) = \begin{cases} -Au(x) & \text{if } x \in F, \\ 0 & \text{if } x \in X \setminus F, \end{cases} \quad l(x) = \begin{cases} Au(x) & \text{if } x \in G, \\ 0 & \text{if } x \in X \setminus G. \end{cases} \tag{8.6}$$

These functions are known together with u, and by (6.16) they are non-negative. On the boundary they both vanish (by (2.1), on the boundary $Af(x) = 0$ for any function f).

Theorem 8.1 *In the notations* (8.1)–(8.4) *and* (8.6), *the value functions v and w satisfy equations*

$$Av = -h, \qquad Aw = -l \tag{8.7}$$

and are given by the formulas

$$v = Rh + Qf, \qquad w = Rl. \tag{8.8}$$

Proof By (7.3) $Av = 0$ on the set $X \setminus F$. On the set $F \subset X \setminus G$ we have $Aw\,(x) = 0$ by (7.4), and therefore $Av = Aw + Au = Au$ (remember that $u = v - w$). By (8.6) this proves the first of equations (8.7). The second equation follows in a similar way from the relations: $Aw = 0$ on $X \setminus G$ (cf.(7.4)), and $Av = 0$ on $G \subset X \setminus F$.

Equations (8.7) mean that $v = h + Pv$, $w = l + Pw$. By iterations we get for any natural n

$$v = \sum_{t=0}^{n-1} P^t h + P^n v, \qquad w = \sum_{t=0}^{n-1} P^t l + P^n w.$$

Here all terms are nonnegative and finite. Due to (8.3) and (8.4), these relations turn into (8.8) as $n \to \infty$. ■

9. Relation to a Dynkin game

Zero-sum stochastic games with stopping times as strategies of the players were proposed by Dynkin [3], and one often calls them Dynkin games. Frid [6] studied the solution of the Dynkin game for a Markov chain with a finite number of states, Gusein-Zade [8] studied it for the Brownian motion process in a domain in R^n (as mentioned by Frid, Gusein-Zade solved also the Dynkin game for a finite Markov chain with an absorbing boundary). In those initial works, the stopping actions of the two players were restricted to two disjoint subsets E_1, E_2 of the state space, so that both players could never stop the process at the same time t. Correspondingly, there was only one reward function g, and the random gain of the first player (equal to the loss of the second player) was $R(\tau, \sigma) = g(x_\rho)$, $\rho = \min(\tau, \sigma)$, where τ and σ are the stopping times chosen by the players I and II. The value function of the game appeared to be a two-sided analogue of the value function of the optimal stopping problem, namely, a solution of a variational problem between two obstacles: an upper bound g on the set E_2 and a lower bound g on the set E_1.

It seems that Krylov [10] was the first who, in his study of the Dynkin game for a general diffusion process, replaced the two sets E_1, E_2 by two functions $f < g$, so that there remained no restriction on the stopping times τ and σ, but the gain of Player I took on a form

$$R(\tau, \sigma) = f(x_\tau) \mathbf{1}\{\tau \leqslant \sigma, \tau < \infty\} + g(x_\sigma) \mathbf{1}\{\sigma < \tau\}. \qquad (9.1)$$

Indeed, the original Dynkin's setting with the sets E_1, E_2 can be reduced to the form (9.1) by renaming g into f on the set E_1, and defining f close enough to $-\infty$ on the complement of E_1, and g close enough to $+\infty$ on the complement of E_2. The corresponding variational problem has precisely the form of inequalities (6.3)–(6.5) for the preference function

u we treated in Theorem 6.1, with A being the generator of the diffusion process. For an extensive exposition of such stochastic games in the continuous-time case we refer to Friedman [7, Chapter 16].

As an auxiliary result of our solution of the switching problem, we obtain a solution of the corresponding Dynkin game. This is not a new result but a new approach. We return to the setting of Section 2 and introduce the necessary definitions. The expected gain of the Player I is

$$u(x,\tau,\sigma) = \mathbf{E}_x R(\tau,\sigma), \quad x \in X.$$

The two functions

$$\underline{u}(x) = \sup_\tau \inf_\sigma u(x,\tau,\sigma), \qquad \bar{u}(x) = \inf_\sigma \sup_\tau u(x,\tau,\sigma), \quad x \in X,$$

are, respectively, the lower and upper values of the game. Always $\underline{u} \leqslant \bar{u}$, and if they are equal then the game has a *value function* $u = \bar{u} = \underline{u}$. A pair of stopping times (τ^*, σ^*) is a *saddle point* if

$$\max_\tau u(x,\tau,\sigma^*) = u(x,\tau^*,\sigma^*) = \min_\sigma u(x,\tau^*,\sigma), \quad x \in X. \tag{9.2}$$

If a saddle point (τ^*, σ^*) exists, then the value function u of the game also exists and is equal to

$$u(x) = u(x,\tau^*,\sigma^*), \quad x \in X. \tag{9.3}$$

The stopping times τ^*, σ^* satisfying (9.2)–(9.3) are *optimal strategies* of the players, and the triple (u, τ^*, σ^*) is called a *solution* of the game.

Theorem 9.1 *The preference function u defined in* Section 6 *is the value function of the game. The stopping times τ_F and τ_G (cf.* (6.13)) *are optimal strategies of the players I and II.*

The proof is based on the following verification lemma for the optimal stopping problem. It is not a new result, but it is easier to give a proof than to find an exact reference.

Lemma 9.1 *Let X, B, $\{x_t\}$ and stopping times τ be the same as in* Section 2. *If a measurable set $E \subset X$ and measurable bounded functions h and u on X satisfy conditions: $B \subset E$,*

$$u(x) = Pu(x) \geqslant h(x), \quad x \in X \setminus E, \tag{9.4}$$

$$u(x) = h(x) \geqslant Pu(x), \quad x \in E, \tag{9.5}$$

then

$$u(x) = \mathbf{E}_x h(x_{\tau_E}) = \sup_{\tau \leqslant \tau_B} \mathbf{E}_x h(x_\tau), \quad x \in X. \tag{9.6}$$

Proof As in the proof of Theorem 6.1, the process

$$\xi_t = u(x_t) - \sum_{k=0}^{t-1} Au(x_k), \quad t \geqslant 0, \tag{9.7}$$

is a martingale satisfying the conditions (6.10) because u is bounded, so that similar to (6.9)

$$u(x) = \mathbf{E}_x \xi_\tau, \quad x \in X, \tag{9.8}$$

for every $\tau \leqslant \tau_B$. By (9.4)–(9.5) $Au = Pu - u \leqslant 0$, hence (9.7) and (9.8), and then again (9.4)–(9.5) imply

$$u(x) \geqslant \mathbf{E}_x u(x_\tau) \geqslant \mathbf{E}_x h(x_\tau), \quad x \in X, \tag{9.9}$$

if only $\tau \leqslant \tau_B$. On the other hand, $\tau_E \leqslant \tau_B$ because $B \subset E$, so that (9.8) is valid for $\tau = \tau_E$, and for this stopping time $Au(x_k) = Pu(x_k) - u(x_k) = 0$ for $t < \tau$ by (9.4). Thus (9.7) and (9.8) imply

$$u(x) = \mathbf{E}_x \xi_{\tau_E} = \mathbf{E}_x u(x_{\tau_E}) = \mathbf{E}_x h(x_{\tau_E}), \quad x \in X, \tag{9.10}$$

where the last expression follows from the fact that $x_{\tau_E} \in E$ and (9.5). Relations (9.9) and (9.10) prove (9.6). ■

Proof of Theorem 9.1 Suppose that the second player uses the stopping time $\sigma = \tau_G$ (recall that G may be empty). Then the first player is in the following situation. On the set $X \setminus B'$ where $B' = B \cup G$, the process $\{x_t\}$ can be stopped only by him (or her), and if it is stopped at a state $x \in X \setminus B'$ the reward is $f(x)$. On the set B he (she) may stop the process or not, but it is optimal to stop and get the reward $f(x)$, $x \in B$ because $f \geqslant 0$ on the boundary B (see (2.5)), and because any state $x \in B$ is absorbing. On the set G, if he (she) stops, the reward is $f(x)$, if not, than the second player stops and the reward is $g(x) > f(x)$, $x \in G$ (see (9.1) and (2.5)). Hence it is definitely better for I not to stop and get the reward $g(x)$. The same happens if we forget about the player II, change every state $x \in G$ into an absorbing state, and change the reward of the first person from $f(x)$ to $g(x)$ at this state. Thus, indeed, if II uses the strategy τ_G, I faces an optimal stopping problem on the space X with an enlarged absorbing boundary $B' = B \cup G$, a modified reward function

$$h(x) = \begin{cases} f(x), & x \in X \setminus G, \\ g(x), & x \in G, \end{cases} \tag{9.11}$$

and the choice of stopping times τ reduced by the condition $\tau \leqslant \tau_{B'}$.

Let now $E = B' \cup F = B \cup G \cup F = G \cup F$ (by Corollary 6.1, $B \subset F$), and let u be the preference function from Section 6. By Corollary 6.1 $u(x) = Pu(x)$ if $x \in X \setminus E$ (see (6.14)), $u(x) \geqslant h(x)$ if $x \in X \setminus E$ (see (9.11) and (6.3)), $u(x) = h(x)$ if $x \in E$ (see (6.13) and (6.15)). To get all the conditions (9.4)–(9.5) of Lemma 9.1, it remains to check the inequality $u(x) \geqslant Pu(x)$ on the set E. It does not hold in a literal sense because we deal with a modified problem with an enlarged absorbing boundary $B' = B \cup G$. In this new process the kernel $P(x, dy)$ has changed from the original one, now we have $P(x, x) = 1$ for x not only in B, but also in $G \subset B'$, and for this new kernel the last condition in (9.5) holds. Thus we have all conditions of Lemma 9.1 with B changed to $B' = B \cup G$. By this lemma, τ_E is an optimal stopping time for Player I in the modified problem. By the relation between the just described optimal stopping problem and the original game with strategies of the second player reduced to the single stopping time $\sigma = \tau_G$, τ_F is the best reply of I to the choice τ_G of II, so that for $\tau^* = \tau_F$, $\sigma^* = \tau_G$ the left of the equations (9.2) holds.

The right of the equations (9.2) is proved in a symmetric way, with the player II maximizing the reward $-R(\tau_F, \sigma)$ over stopping times σ subject to the constraint $\sigma \leqslant \tau_F$ (actually, Player I stops the process on the set F, mandatory for the player II, and in place of (9.11) we now have $h(x) = -g(x)$ if $x \in X \setminus F$, $h(x) = -f(x)$ if $x \in F$). ■

10. Examples

10.1 Symmetric random walk

In this example the state space X is $\{0, 1, 2, \ldots, k, \ldots, n\}$, the states 1 and n are absorbing states, the transitions $k \to (k \pm 1)$ occur with probabilities 1/2, the reward and cost functions are vectors $\{f_0, f_1, \cdots, f_n\}$ and $\{g_0, g_1, \cdots, g_n\}$, where $f_k < g_k$ for all k, and $f_0 \geq 0, f_n \geq 0$. The optimality inequalities (6.3)–(6.5) of Theorem 6.1 take on the form

$$f_k \leq u_k \leq g_k, \tag{10.1}$$

$$\frac{1}{2}u_{k-1} + \frac{1}{2}u_{k+1} \geq u_k \quad \text{if } u_k > f_k, \tag{10.2}$$

$$\frac{1}{2}u_{k-1} + \frac{1}{2}u_{k+1} \leq u_k \quad \text{if } u_k < g_k \tag{10.3}$$

(here $k = 1, \ldots, n-1$), and the boundary conditions (6.6) become

$$u_0 = f_0, \qquad u_n = f_n, \tag{10.4}$$

(the values of g_0 and g_n are unessential). Consider the functional

$$\Phi[u] = \sum_{k=0}^{n-1} \eta(u_{k+1} - u_k)$$

where $\eta(x)$ is any twice differentiable function with $\eta''(x) > 0$, $-\infty < x < \infty$. For such η, the partial derivative

$$\frac{\partial \Phi}{\partial u_k} = \eta'(u_k - u_{k-1}) - \eta'(u_{k+1} - u_k)$$

is a strictly increasing function of u_k, $1 \leq k \leq n-1$. It follows that

$$u_k = \frac{1}{2}(u_{k-1} + u_{k+1})$$

is the unique solution of the equation $\frac{\partial \Phi}{\partial u_k} = 0$, and that relations (10.2)–(10.3) are necessary conditions of a minimum of the functional Φ subject to the constraints (10.1). In particular, one may take $\eta(x) = \sqrt{1+x^2}$. It is convenient to represent the functions f, u, g by broken lines connecting the points $(k, f(k))$, resp. $(k, u(k))$, resp. $(k, g(k))$ from $k = 0$ to $k = n$. It follows that *the graph of u is the shortest path between two obstacles: the graphs of f and g, connecting the points* $(0, f(0))$ *and* $(n, f(n))$. The optimal switching sets F and G consist of those k at which the graph of u touches the graphs of f, resp. g.

The exit probabilities and occupational measures in this example are known (or can be easily found from the corresponding difference equations). They are

$$Q(k,0) = \frac{n-k}{n}, \qquad Q(k,n) = \frac{k}{n},$$

$$R(k,j) = \begin{cases} \frac{2(n-k)j}{n}, & \text{if } 1 \leq j \leq k, \\ \frac{2k(n-j)}{n}, & \text{if } k \leq j \leq n-1. \end{cases}$$

By (8.6) and (8.8), we have for $k = 0, \ldots, n$

$$v(k) = \frac{n-k}{n}\left[f_0 + 2\sum_{j=1}^{k} j h_j\right] + \frac{k}{n}\left[2\sum_{k+1}^{n-1}(n-j)h_j + f_n\right], \tag{10.5}$$

$$w(k) = 2\left[\frac{n-k}{n}\sum_{j=1}^{k} j l_j + \frac{k}{n}\sum_{k+1}^{n-1}(n-j)l_j\right], \tag{10.6}$$

where

$$2h_j = \begin{cases} 2u_k - (u_{k-1} + u_{k+1}) \geq 0, & \text{if } k \in F \cap \{1, 2, \dots, n-1\}, \\ 0, & \text{otherwise,} \end{cases} \tag{10.7}$$

$$2l_j = \begin{cases} (u_{k-1} + u_{k+1}) - 2u_k \geq 0, & \text{if } k \in G, \\ 0, & \text{otherwise.} \end{cases} \tag{10.8}$$

10.2 Birth and death process

This model is similar to the preceding one, only the transition probabilities from k to $k+1$ and $k-1$ are now arbitrary numbers p_k and q_k satisfying conditions $p_k > 0$, $q_k > 0$, $p_k + q_k = 1$ $(k = 1, \dots, n-1)$. The picture becomes very similar to that in Example 10.1 if we introduce in X the so-called natural scale (cf. Dynkin and Yushkevich [4, Chapter 4]). We re-scale the states $k = 0, 1, \dots, n$ as $x_0, x_1, \dots x_n$ where

$$x_0 = 0, \qquad x_k = \sum_{i=1}^{k} \Delta_i, \quad k = 1, \dots, n,$$

and

$$\Delta_1 = 1, \qquad \Delta_i = \frac{q_1 q_2 \cdots q_{i-1}}{p_1 p_2 \cdots p_{i-1}}. \tag{10.9}$$

Instead of optimality inequalities (10.2)–(10.3) we now have

$$q_k u_{k-1} + p_k u_{k+1} \geq u_k \quad \text{if } u_k > f_k, \tag{10.10}$$

$$q_k u_{k-1} + p_k u_{k+1} \leq u_k \quad \text{if } u_k < g_k; \tag{10.11}$$

relations (10.1) and (10.4) remain unchanged. The appropriate functional is

$$\Phi[u] = \sum_{k=0}^{n-1} \Delta_{k+1} \eta \left[\frac{1}{\Delta_{k+1}} (u_{k+1} - u_k) \right]$$

where again $\eta'' > 0$. Relations (10.10)–(10.11) again are necessary conditions of a minimum of Φ under constraints (10.1). Indeed,

$$\frac{\partial \Phi}{\partial u_k} = \eta' \left[\frac{1}{\Delta_k} (u_k - u_{k-1}) \right] - \eta' \left[\frac{1}{\Delta_{k+1}} (u_{k+1} - u_k) \right],$$

and at a minimum point this partial derivative should be zero if $f_k < u_k < g_k$, nonpositive if $f_k < u_k = g_k$, and nonnegative if $f_k = u_k < g_k$. Since η' is strictly increasing, the equation $\frac{\partial \Phi}{\partial u_k} = 0$ means that

$$\frac{1}{\Delta_k} (u_k - u_{k-1}) = \frac{1}{\Delta_{k+1}} (u_{k+1} - u_k) \tag{10.12}$$

and this, by (10.9), is equivalent to $u_k = q_k u_{k-1} + p_k u_{k+1}$, so that both (10.10) and (10.11) hold. Similarly, $\frac{\partial \Phi}{\partial u_k} \leq 0$ implies (10.12) with the inequality $\leq$, i.e. (10.10), while $\frac{\partial \Phi}{\partial u_k} \geq 0$ implies (10.12) with the inequality $\geq$, i.e. (10.11). In particular, for $\eta(x) = \sqrt{1+x^2}$ the functional becomes

$$\Phi[u] = \sum_{k=0}^{n-1} \sqrt{\Delta_{k+1}^2 + (u_{k+1} - u_k)^2}\,;$$

this is the length of the graph of u in the natural scale. Thus the preference function has the same geometrical interpretation as in Example 10.1.

Exit probabilities are now (cf. [4])

$$Q(k,0) = \frac{x_n - x_k}{x_n}, \qquad Q(k,n) = \frac{x_k}{x_n}.$$

Occupational measures can be found from difference equations, and they are

$$R(k,j) = \begin{cases} \frac{(x_n - x_k)x_j}{\Delta_j q_j x_n}, & \text{if } 1 \leq j \leq k, \\ \frac{x_k(x_n - x_j)}{\Delta_j q_j x_n}, & \text{if } k \leq j \leq n-1. \end{cases}$$

Hence, analogous to (10.5)–(10.8), the value functions are

$$\begin{aligned} v(k) &= \frac{x_n - x_k}{x_n}\left[f_0 + \sum_{j=1}^{k} \frac{x_j h_j}{\Delta_j q_j}\right] + \frac{x_k}{x_n}\left[\sum_{k+1}^{n-1} \frac{(x_n - x_j)h_j}{\Delta_j q_j} + f_n\right], \\ w(k) &= \frac{x_n - x_k}{x_n}\sum_{j=1}^{k} \frac{x_j l_j}{\Delta_j q_j} + \frac{x_k}{x_n}\sum_{k+1}^{n-1} \frac{(x_n - x_j)l_j}{\Delta_j q_j} \end{aligned}$$

where

$$\begin{aligned} h_j &= \begin{cases} u_k - (q_k u_{k-1} + p_k u_{k+1}) \geq 0, & \text{if } k \in F \cap \{1, 2, \ldots, n-1\}, \\ 0, & \text{otherwise,} \end{cases} \\ l_j &= \begin{cases} q_k u_{k-1} + p_k u_{k+1} - u_k \geq 0, & \text{if } k \in G, \\ 0, & \text{otherwise.} \end{cases} \end{aligned}$$

References

[1] Bertsekas D.P. and Shreve S.E. (1978). *Stochastic Optimal Control: The Discrete Time Case,* Academic Press, N.Y.

[2] Chow Y.S., Robbins H. and Siegmund D. (1971). *Great Expectations: The Theory of Optimal Stopping,* Houghton Miffin, Boston (Dover edition, 1991).

[3] Dynkin E.B. (1969). A game version of an optimal stopping problem, *Soviet Math. - Doklady,* **10,** 270–274.

[4] Dynkin E.B. and Yushkevich A.A. (1969). *Markov Processes—Theorems and Problems,* Plenum Press, N.Y. (Russian edition, 1967).

[5] Dynkin E.B. and Yushkevich A.A. (1979). *Controlled Markov Processes,* Springer, N.Y. (Russian edition, 1975).

[6] Frid E.B. (1969). The optimal stopping rule for a two-person Markov chain with opposing interests, *Theory Probab. Applic.* **14,** 713–716.

[7] Friedman A. (1976). *Stochastic Differential Equations and Applications,* Academic Press, N.Y.

[8] Gusein-Zade S.M. (1969). A certain game connected with a Wiener process, *Theory Probab. Applic.* **14,** 701–704.

[9] Hernández-Lerma O. and Lasserre J.B. (1996). *Discrete-Time Markov Control Processes: Basic Optimality Criteria,* Springer, N.Y.

[10] Krylov N.V. (1971). Control of Markov Processes and W-spaces, *Math. USSR - Izvestiya* **5,** 233–266.

[11] Puterman M. (1994). *Markov Decision Processes: Discrete Stochastic Dynamic Programming,* Wiley, N.Y.

[12] Shiryayev A.N. (1978). *Optimal Stopping Rules,* Springer, N.Y. (Russian editions, 1969, 1976).

[13] Shiryayev A.N. (1984). *Probability,* Springer, N.Y. (Russian edition, 1980).

[14] Snell I.L. (1953). Applications of martingale system theorems, *Trans. Amer. Math. Soc.* **73**, 293–312.

Chapter 16

APPROXIMATIONS OF A CONTROLLED DIFFUSION MODEL FOR RENEWABLE RESOURCE EXPLOITATION

optional, to supply a shorter version of the title for the running head:

Sara Pasquali

Dipartimento di Matematica

Università degli Studi di Parma

Via M. D'Azeglio 85

43100 Parma

Italy

sara.pasquali@unipr.it

Wolfgang J. Runggaldier

Dipartimento di Matematica Pura ed Applicata

Università degli Studi di Padova

Via Belzoni 7

35131 Padova

Italy

runggal@math.unipd.it

Abstract We study the problem of a renewable resource exploitation as a problem of optimal stochastic control with the renewable resource being managed for social benefit.

The aim is to maximize the finite horizon total discounted utility by controlling the per capita consumption and extraction capacity. We suppose that the state is not directly observable, so we have an optimal stochastic control problem with partial observations. The exact solution is difficult to obtain, so we aim at a nearly optimal control determined via an approximation approach involving a discretization procedure in time and space.

Z. Hou et al. (eds.), Markov Processes and Controlled Markov Chains, 287–302.

1. Statement of the problem

Let $(\Omega, \mathbf{F}, P)$ be a probability space.

Let us denote by x_t the resource and by y_t the population at time t. Assume that x and y satisfy the following equations:

$$dx_t = \left[g(x_t) - F(x_t, s(y_t), D_t)\right] dt + \sigma_1\, dw_t^{(1)}, \tag{1.1}$$

$$dy_t = \left[\theta_t F(x_t, s(y_t), D_t) - (vD_t + C_t)\, s(y_t)\right] dt + \sigma_2\, dw_t^{(2)}, \tag{1.2}$$

with initial condition $x(0) = x_0$ and $y(0) = y_0$ respectively. Where $g(x_t)$ is the renewal rate of the resource, $F(x_t, s(y_t), D_t)$ is the total harvest of the resource, θ_t is the proportion of harvested resource used by each member of y_t, D_t is the potential harvest capacity, vD_t is the harvest cost, C_t is per-capita consumption, and $s(y)$ is a C^∞ transformation, bounded and Lipschitz in y. Moreover, σ_1 and σ_2 are two positive constants with σ_2 small, while $w_t^{(1)}$ and $w_t^{(2)}$ are independent Wiener processes. We observe that in equations (1.1) and (1.2) there is an additive noise term because both resource and population can be subject to random variations. We also remark that in equation (1.2), we have a small noise term because random fluctuations in population are smaller than those in the resource; in fact sharp variations in population are due only to a catastrophe. Model (1.1), (1.2) is a stochastic extension of a deterministic model studied by Regev, Gutierrez, Schreiber and Zilbermann in [2].

Furthermore, θ satisfies the following equation.

$$d\theta_t = \frac{(\theta_t - \lambda)(1 - \lambda - \theta_t)}{1 - 2\lambda}\left(\mu + \frac{1 - 2\theta_t}{2(1 - 2\lambda)}\right) dt + \frac{(\theta_t - \lambda)(1 - \lambda - \theta_t)}{1 - 2\lambda}\, dw_t^{(3)}, \tag{1.3}$$

with initial condition $\theta(0) = \theta_0$ and where μ is a constant parameter:

$$d\mu_t = 0, \qquad \mu(0) = \mu_0, \tag{1.4}$$

and $w^{(3)}$ is a Wiener process independent of $w^{(1)}$ and $w^{(2)}$. Drift and diffusion coefficients of equation (1.3) are such that $\theta \in (\lambda, 1 - \lambda)$ with λ small (this is possible by modelling θ by a C^∞ transformation from $\mathbb{R}$ to $(\lambda, 1 - \lambda)$ - see, for example, [4]) because θ represents a percentage. In order to have θ in a compact set, we restrict the interval $(\lambda, 1 - \lambda)$ by considering the interval $[\lambda + \epsilon, 1 - \lambda - \epsilon]$ with $\epsilon \to 0$. In what follows we shall denote, improperly, this closed set by $[\lambda, 1 - \lambda]$.

Assumption 1.1 *g is bounded and Lipschitz in x.*

Assumption 1.2 *F has the form $F(x_t, s(y_t), D_t) = D_t s(y_t) h\left(\frac{\alpha x_t}{D_t s(y_t)}\right)$, where h is the proportion of the potential demand for resources actually required, Lipschitz in (x, y) uniformly in D, and α (positive constant) is the technology parameter of resource harvesting (it represents the resource harvesting efficiency).*

Let $\xi_t := \frac{\alpha x_t}{D_t s(y_t)}$ and assume that the function $h(\xi)$ is concave,

$$\lim_{\xi\to\infty} \xi h'(\xi) = 0 \quad \textit{and} \quad h'(0) = \lim_{\xi\to\infty} h(\xi) = 1.$$

Assumption 1.3
$d \le D_t \le \overline{D}\ \forall t \in [0, \infty)$, for $\overline{D} > 0$, and for some $0 < d < \overline{D}$.

We consider the extended state (see [6]) $X_t = \left(X_t^{(1)}, \mu_t\right)$, with $X_t^{(1)} = (x_t, y_t, \theta_t)$. The components of the vector X_t satisfy the equations (1.1), (1.2), (1.3) and (1.4). As a consequence, $X_t^{(1)}$ satisfies an equation of the form:

$$dX_t^{(1)} = f^{(1)}(X_t, u_t)\, dt + \sigma(X_t)\, dW_t, \qquad X^{(1)}(0) = X_0^{(1)}, \tag{1.5}$$

with the obvious meaning of the symbols and with $u_t = (C_t, D_t)$ being the control at time t.

Equations (1.1), (1.2), (1.3) and (1.4) represent the dynamics for a stochastic control problem in which we want to find the supremum, over the controls C_t and D_t, of the reward,

$$J\big(C(\cdot), D(\cdot)\big) = E\int_0^T e^{-\delta t} s(y_t) U(C_t)\, dt, \tag{1.6}$$

where $U(C_t)$ is the per-capita utility function which is supposed to be Lipschitz in C and $\delta > 0$ is the discount factor.

The state is not fully observable, in particular for the first component (representing the renewable resource) we do not have precise information about the quantity of the available resource due both to errors in measurement and to incorrect information supplied by exploiters of the resource. We, thus, introduce the following assumption.

Assumption 1.4 *We suppose to have information on the state only through the observation process η_t satisfying the equation,*

$$d\eta_t = x_t\, dt + \epsilon\, d\widetilde{w}_t, \qquad \eta(0) = \widetilde{\eta}_0, \qquad \epsilon > 0, \tag{1.7}$$

where $\widetilde{w}_t$ is a Wiener process independent of $w_t^{(1)}$, $w_t^{(2)}$ and $w_t^{(3)}$, and ϵ is a small positive constant.

Control space Recall that, by Assumption 1.3 , $d \leq D_t \leq \overline{D}$. As for C_t, one can find, by simple calculation, that $C_t \leq D_t$, so we can take for C_t the upper bound of D_t. Thus the control space is

$$\mathbf{V} = \left[0, \overline{D}\right] \times \left[d, \overline{D}\right], \tag{1.8}$$

which is a compact metric space.

State space The processes x_t and y_t are modelled as solutions of stochastic differential equations, so their trajectories are continuous functions a.s., with values in $\mathbb{R}$; θ_t takes values in the interval $[\lambda, 1-\lambda]$ and μ_0 is uniformly distributed over an interval $[a, b]$. Consequently the state space is

$$\mathbf{X} = \mathbb{R} \times \mathbb{R} \times [\lambda, 1-\lambda] \times [a, b]. \tag{1.9}$$

Nevertheless, we are interested only in solutions of (1.1) and (1.2) satisfying the constraints $0 \leq x_t \leq \overline{x}$ and $\Delta \leq y_t \leq \overline{Y}$ for all t, namely such that the renewable resource does not become negative (otherwise we have exhaustion of the resource), and that the population does not go below level Δ (otherwise we have extinction of the species). Moreover, the resource must not exceed a level $\overline{X}$ and the population a level $\overline{Y}$, for $\overline{X}$ and $\overline{Y}$ sufficiently large. Since we are over a finite interval, both resource and population cannot explode. We observe that the upper limitations are introduced only in order to have x and y in a compact space, but it is possible to choose $\overline{X}$ and $\overline{Y}$ so large that in practice resource and population never reach these values.

Therefore we are interested in the trajectories contained in the space

$$\widehat{\mathbf{X}} = \left[0, \overline{X}\right] \times \left[\Delta, \overline{Y}\right] \times \left[\lambda, 1-\lambda\right] \times [a, b] \tag{1.10}$$

instead of in $\mathbf{X}$. We consider the following part of the boundary of $\widehat{\mathbf{X}}$:

$$\partial\widehat{\mathbf{X}} := \left(\left\{0, \overline{X}\right\} \times \left[\Delta, \overline{Y}\right] \times \left[\lambda, 1-\lambda\right] \times [a, b]\right)$$
$$\cup\left(\left[0, \overline{X}\right] \times \left\{\Delta, \overline{Y}\right\} \times [\lambda, 1-\lambda] \times [a, b]\right),$$

which we shall improperly call boundary of $\widehat{\mathbf{X}}$.

The aim is to maximize the total utility over the set $\widehat{\mathbf{X}}$, since outside this set the problem has no economic meaning. In order to achieve this goal, we introduce a new control component, the stopping time τ (with respect to the σ-algebra generated by the observations). We want the optimal τ to be near the hitting time of the boundary of $\widehat{\mathbf{X}}$ and near the final time T, otherwise we would have too quickly the exhaustion of the resource or the population extinction.

To this end we consider a penalization function $\overline{Q}(\tau, X_\tau)$ which depends both on the stopping time τ and on the state evaluated at the stopping time. We choose $\overline{Q}$ in such a way that it penalizes a τ far from both the hitting time of the set $\widehat{\mathbf{X}}$ and the final time T. Furthermore, we choose $\overline{Q}$ bounded by $\widetilde{Q}$ and Lipschitz in X.

Therefore, the new functional is

$$J(u,\tau) = E\left\{\int_0^{T\wedge\tau} e^{-\delta t} L(X_t, u_t)\, dt - \overline{Q}(\tau, X_\tau)\right\}, \tag{1.11}$$

where $L(X_t, u_t) := s(y_t)\, U(C_t)$, with $L : \mathbf{X} \times \mathbf{V} \to \mathbb{R}$ continuous in the state and the control and, for U bounded, L is bounded over the interval $[0, T]$ because $s(y)$ is bounded (we shall denote by $\widetilde{L}$ its bound). Moreover, since $s(y)$ is Lipschitz, we have that L is Lipschitz in X. Owing to the fact that U is Lipschitz in C, it follows that L is Lipschitz also in the control.

The aim is to find

$$\sup_{u(\cdot)\in\overline{\mathbf{V}},\tau\in[0,T]} J(u,\tau),$$

where $\overline{\mathbf{V}}$ is the set of the admissible controls $u(\cdot)$ (that is the set of the controls taking values in $\mathbf{V}$ and adapted to $\mathbf{F}_t^\eta = \sigma\{\eta_s, s \leq t\}$); the state equation is given by (1.4)–(1.5) where the function $f^{(1)} : \mathbf{X} \times \mathbf{V} \to \mathbb{R}^3$ is a continuous function of the state, Borel bounded and Lipschitz in X uniformly in u; and the function $\sigma : \mathbf{X} \to \mathcal{M}_d(3 \times 3)$, where $\mathcal{M}_d(3 \times 3)$ denotes the space of the diagonal 3×3 matrices with values in $\mathbb{R}$, is Borel bounded and Lipschitz in X.

The exact solution is difficult to obtain, so we aim at a nearly optimal control determined via an approximation approach involving a discretization procedure.

2. Discretization in time

In this section we approximate the continuous problem by a time discretized stochastic control problem.

2.1 Discretization of the state and of the control

For each fixed N, we consider the subset $\mathbf{V}_N \subset \overline{\mathbf{V}}$ of step controls corresponding to the deterministic splitting $0 = t_0 < t_1 < \ldots.. < t_N = T$ of the time interval with $|t_i - t_{i-1}| = \frac{T}{N}$, $\forall i = 1, ..., N$.

Therefore we have $u_t = u_n$, for $t \in [t_n, t_{n+1}[$, with $u_n \in \mathbf{V}$ and $\mathbf{F}_{t_n}^\eta$-measurable ($\mathbf{F}_{t_n}^\eta = \sigma(\eta_t, t \leq t_n)$ where η_t is the observation process defined in (1.7)). We discretize also the τ-space and, instead of $[0,T]$, we consider the discrete set of time points $\{t_0, t_1, ..., t_N\}$, so that $\tau \in \{n\frac{T}{N}; n = 0, ..., N\}$.

Proposition 2.1 *Given $\epsilon > 0$, for N sufficiently large*

$$\left| \sup_{\tau \in \{n\frac{T}{N},\ n=0,1,\dots,N\}} \sup_{u \in \mathbf{V}_N} J(u,\tau) - \sup_{\tau \in [0,T]} \sup_{u \in \overline{\mathbf{V}}} J(u,\tau) \right| \leq \epsilon. \tag{2.1}$$

Proof The left hand side of (2.1) is equivalent to

$$\left| \sup_{\tau \in \{n\frac{T}{N},\ n=0,1,\dots,N\}} \sup_{u \in \mathbf{V}_N} J(u,\tau) - \sup_{\tau \in R_T} \sup_{u \in \mathbf{V}_S} J(u,\tau) \right|, \tag{2.2}$$

where $R_T = [0,T] \cap \mathbb{Q}$ and $\mathbf{V}_S = \bigcup_N \mathbf{V}_N$ (see [4]).

It can be shown (see [4]) that, given $\epsilon > 0$, there exists $\overline{N}$ such that, for $N \geq \overline{N}$, (2.2) is less than ϵ. ■

Proposition 2.1 states that we can restrict ourselves to consider only controls $u \in \mathbf{V}_N$ and controls $\tau \in \{n\frac{T}{N}; n = 0, .., N\}$, for N sufficiently large. Therefore we have a simpler problem than the original one. The optimal step control of the corresponding time-discretized stochastic control problem will be shown to be an ϵ-optimal control for the original problem.

Corresponding to the splitting of the time interval into subintervals of the same width $\frac{T}{N}$, for each $N \in \mathbb{N}$, we consider a time discretized state:

$$X_t^N = \begin{cases} X_j^N & \text{for } t \in [j\frac{T}{N}, (j+1)\frac{T}{N}[\\ X_N^N & \text{for } t = T \end{cases} \tag{2.3}$$

where X_j^N $(j = 0, ..., N-1)$ and X_N^N depend on j and on T, respectively, and are obtained from an Euler discretization of (1.5) and (1.4).

What one usually does in these cases is a Girsanov change of measure in order to transform the original problem into a problem in which state and observations are independent. This allows one to work in a product space in which the distribution of the state is furthermore the same as in the original space (see, e.g., [4] and [6]).

We shall denote by P^0 the measure under which η_t is a Wiener process independent of X_t and X_t^N, and by P^N the measure under which η_t has the same form as under P, but as a function of the discretized state. More precisely,

$$d\eta_t = x_t^N \, dt + \epsilon \, dw_t^N$$

with w^N a P^N-Wiener process and x^N the first component of X^N (see (1.7)).

It can be shown that the process X_t^N converges to the continuous state X_t in fourth mean, both in the original measure P and in the transformed measure P^0 (see [4] or [6]).

2.2 Discretization of the reward functional

For each $N \in \mathbb{N}$ we define the following functional:

$$J^N(u,\tau) := E^0\left\{z^N(T)\left[\int_0^{T\wedge\tau} e^{-\delta t} L\left(X_t^N, u_t\right) dt - \overline{Q}\left(\tau, X_\tau^N\right)\right]\right\}, \tag{2.4}$$

where $z^N(T) = \frac{dP^N}{dP^0}$, $u \in \mathbf{V}_N$, that is $u = u_j$ in the interval $\left[j\frac{T}{N}, (j+1)\frac{T}{N}\right[$ with u_j measurable with respect to the σ-algebra generated by the increments of η_t up to time $j-1$ and $\tau \in \left\{n\frac{T}{N}; n = 0, 1, ..., N\right\}$ is a stopping time with respect to this same σ-algebra. It is important to notice that here we use the σ-algebra $\mathbf{F}_{j-1}^{\Delta\eta}$ generated by the increments $\eta_{j-1} := \eta\left(j\frac{T}{N}\right) - \eta\left((j-1)\frac{T}{N}\right)$, while in the continuous time case we used the σ-algebra $\mathbf{F}_t^\eta$ generated by the continuous process η_t. In [4] we show that, for the time discretized problem, there is no difference in taking controls adapted to $\mathbf{F}_t^\eta$ or to $\mathbf{F}_{j-1}^{\Delta\eta}$ since this does not modify the solution (see also [1]).

Proposition 2.2

$$\left|J(u,\tau) - J^N(u,\tau)\right| \leq \widetilde{K}\left(\frac{T}{N}\right)^{\frac{1}{2}} \tag{2.5}$$

uniformly in the control (u,τ), with $\widetilde{K}$ constant.

Proof Applying the change of measure, Jensen's inequality, Tonelli's theorem, and the Hölder inequality, and recalling that L and Q are bounded by $\widetilde{L}$ and $\widetilde{Q}$ respectively and Lipschitz, we obtain that the left hand side of (2.5) is bounded above by:

$$L' \int_0^T e^{-\delta t}\left\{E\left[\left\|X_t - X_t^N\right\|^4\right]\right\}^{\frac{1}{4}} dt$$
$$+\left(\widetilde{Q} + \widetilde{L}\frac{1-e^{-\delta T}}{\delta}\right)\cdot\left[E^0\left|z(T) - z^N(T)\right|^2\right]^{\frac{1}{2}}$$
$$+Q'\left[E\left\|X_t - X_t^N\right\|^4\right]^{\frac{1}{4}}$$

where $z(T) = \frac{dP}{dP^0}$, L' and Q' are the Lipschitz constants relative to L and Q, respectively.

As mentioned, the sequence X_t^N converges to X_t in fourth mean under both P and P^0; furthermore the sequence $z^N(T)$ converges to $z(T)$ in mean square under P^0. In particular we have $E\left\|X_t - X_t^N\right\|^4 \leq K_1\left(\frac{T}{N}\right)^2$ where K_1 is a constant (and the same upper bound is valid for the expectation E^0) and $E^0\left\|z(T) - z^N(T)\right\|^2 \leq K_2\frac{T}{N}$ with K_2 constant (see [1, 4]), so we immediately obtain the thesis. ■

Thanks to Proposition 2.2, we can approximate the initial reward J by the discretized reward J^N, for N sufficiently large. In this way, the discretized problem consists in maximizing J^N by choosing the controls $u \in \mathbf{V}_N$ and $\tau \in \left\{n\frac{T}{N} : n = 0, ..., N\right\}$. Due to the uniformity, in (u, τ), of the bound in (2.5), these controls are nearly optimal for the original problem.

We have now, a partially observable discrete time stochastic control problem over a finite horizon with discounted reward. We shall write the so-called separated problem associated to it, namely a corresponding problem with fully observed state, and then apply the DP algorithm to this problem after discretizing also in space.

2.3 The separated problem

Let

$$P\left(X_{n+1}, X_n, u_n, \eta_n\right) := P\left(X_{n+1}\left|X_n, u_n\right.\right) e^{\left[\frac{1}{\epsilon^2}x_n\eta_n - \frac{1}{2\epsilon^2}(x_n)^2\frac{T}{N}\right]} \qquad (2.6)$$

where

$$\begin{aligned} & P\left(X_{n+1}\left|X_n, u_n\right.\right) \\ = \; & \Phi_1\left(x_{n+1}; X_n, u_n\right)\Phi_2\left(y_{n+1}; X_n, u_n\right)\Phi_3\left(\theta_{n+1}; X_n, u_n\right) \end{aligned}$$

is the transition kernel obtainable as product of the conditional distributions of each component of $X^{(1)}$. The factors Φ_1, Φ_2 and Φ_3 are obtained from the discretization of the state equation (1.5) and, using an Euler discretization of the state equation, Φ_1 and Φ_2 are normal and Φ_3 can be transformed into a normal distribution by a change of variable (we have to consider the same C^∞ transformation mentioned in Section 1—see [4]).

Let $\eta^n = (\eta_0, ..., \eta_n)$ be the observation vector up to step n and $u^n = (u_0, ..., u_n)$ be the sequence of controls up to step n. Consider the process

$\{q_n(\cdot)\}$ defined recursively as follows.

$$\begin{aligned} q_0(X_0) &= p_0(X_0) \\ q_{n+1}(X_{n+1};\eta^n,u^n) &= \int_{\mathbf{X}} P(X_{n+1},X_n,u_n,\eta_n)\, q_n\left(X_n;\eta^{n-1},u^{n-1}\right) dX_n \end{aligned} \tag{2.7}$$

for $n = 0, ..., N-1$, where $p_0(X_0)$ is the initial distribution of the process $\{X_n\}$.

Assuming that the initial values of resource, population and percentage of the resource that is being used are known, and that the parameter μ is uniformly distributed over the interval $[a,b]$, the initial distribution of the state is:

$$p_0(X) = \delta(x-x_0)\,\delta(y-y_0)\,\delta(\theta-\theta_0)\,\frac{1}{b-a}. \tag{2.8}$$

We observe that (2.7) corresponds to a recursive Bayes' formula for computing an unnormalized conditional distribution $q_{n+1}(X_{n+1};\eta^n,u^n)$ of X_{n+1} given the observations η^n and the controls u^n.

Proposition 2.3 *Any function $\Phi(X_{j-1},u_{j-1},\tau)$, of one of the following types:*

$$\begin{aligned} &\Phi'(X_{j+1},u_{j+1})\, I_{]j\frac{T}{N},T]}(\tau); \qquad or \\ &\Phi''(X_{j+1},u_{j+1})\, I_{[j\frac{T}{N},(j+1)\frac{T}{N}[}(\tau), \end{aligned}$$

satisfies the property:

$$\begin{aligned} &E^0\left[\Phi(X_{j+1},u_{j+1},\tau)\, z^N(T)\,\middle|\eta^j,u^j\right] \\ &\quad= \int_{\mathbf{X}} \Phi(X,u_{j+1},\tau)\cdot q_{j+1}\left(X;\eta^j,u^j\right) dX \end{aligned} \tag{2.9}$$

where the control u_{j+1} is adapted to the σ-algebra generated by the increments η_n of the observations up to time j and the control τ is a stopping time with respect to this same σ-algebra. We recall that $z^N(T) = \frac{dP^N}{dP^0}$.

For the proof see [1].

Going back to the functional $J^N(u)$ in (2.4), recalling (2.3) and applying Proposition 2.3, we have:

$$\begin{aligned} J^N(u,\tau) &= E^0\Bigg\{\sum_{j=0}^{N-1}\int_{\mathbf{X}}\Bigg[\frac{e^{-\delta(j\frac{T}{N})}-e^{-\delta((j+1)\frac{T}{N})}}{\delta}L(X,u_j)\,I_{]j\frac{T}{N},T]}(\tau) \\ &\quad -\overline{Q}\left(j\frac{T}{N},X\right)I_{[j\frac{T}{N},(j+1)\frac{T}{N}[}(\tau)\Bigg]q_j\left(X;\eta^{j-1},u^{j-1}\right)dX \\ &\quad -\int_{\mathbf{X}}\overline{Q}(T,X)I_{\{T\}}(\tau)q_N\left(X;\eta^{N-1},u^{N-1}\right)dX\Bigg\}. \qquad (2.10) \end{aligned}$$

Now we have a completely observable stochastic control problem where the new state is the unnormalized conditional density $q_n\left(X_n;\eta^{n-1},u^{n-1}\right)$. The aim is to find the supremum of (2.10) subject to the state equation (2.7).

3. Further discretizations

At this point we are not yet able to apply the dynamic programming algorithm to the separated problem because the state is infinite-dimensional and takes a continuum of possible values. Therefore we discretize the state space $\widehat{\mathbf{X}}$.

3.1 State discretization

First we note that the state X can leave the set $\widehat{\mathbf{X}}$ so we have to define the problem on all of $\mathbf{X}$, but one can show that there exists a finite band $\overline{B}$ such that, with probability close to 1, X never leaves $\widehat{\mathbf{X}}\cup\overline{B}$. Consequently, for each positive integer m, we consider a partition $\{B_h\}_{h=1}^{\overline{M}}$ (with $\overline{M}=2^{4m}$) of the state space $\widehat{\mathbf{X}}\cup\overline{B}$ such that each B_h has width going to zero as we refine the partition and, from each of these subsets B_h, we select a representative element. The set of these representative elements forms the discretized state space.

Define the process $\left\{q_n^{(m)}\left(X_n;\eta^{n-1},u^{n-1}\right)\right\}$ having the same form as (2.7), but with P substituted by $P^{(m)}$, that is the analogue of (2.6) discretized in the state in such a way that

- $E^0\left\|P\left(X_{n+1},X_n,u_n,\eta_n\right)-P^{(m)}\left(X_{n+1},X_n,u_n,\eta_n\right)\right\|\leq\overline{H}_m$ where $\|\cdot\|$ is a norm in L^1 with respect to X_{n+1} and the expectation is with respect to η_n;
- $lim_{m\to\infty}\overline{H}_m=0$.

With these assumptions we can show the following result:

Proposition 3.1 *For N sufficiently large, for each $n = 0, 1, ..., N$, we have*

$$E^0 \left[\left\| q_n \left(X_n; \eta^{n-1}, u^{n-1} \right) - q_n^{(m)} \left(X_n; \eta^{n-1}, u^{n-1} \right) \right\| \right] \leq K_1 \overline{H}_m \quad (3.1)$$

uniformly in the control u, where K_1 is a constant, the norm (which is a norm in L^1) is with respect to X_n, the expectation is with respect to η^{n-1}.

Proof We proceed by induction. For $n = 0$, (3.1) is immediately verified. Suppose that (3.1) is true for n and consider $n+1$: we obtain that $E^0 \left[\left\| q_{n+1} \left(X_{n+1}; \eta^n, u^n \right) - q_{n+1}^{(m)} \left(X_{n+1}; \eta^n, u^n \right) \right\| \right]$ is bounded above by

$$\begin{aligned} & E^0 \int_{\mathbf{X}} \int_{\mathbf{X}} P^{(m)} \left(X_{n+1}, X_n, u_n, \eta_n \right) \\ & \quad \left| q_n \left(X_n; \eta^{n-1}, u^{n-1} \right) - q_n^{(m)} \left(X_n; \eta^{n-1}, u^{n-1} \right) \right| dX_n \, dX_{n+1} \quad (3.2) \\ & \quad + E^0 \int_{\mathbf{X}} \int_{\mathbf{X}} \left| P \left(X_{n+1}, X_n, u_n, \eta_n \right) - P^{(m)} \left(X_{n+1}, X_n, u_n, \eta_n \right) \right| \\ & \quad \quad \cdot q_n \left(X_n; \eta^{n-1}, u^{n-1} \right) dX_n \, dX_{n+1}. \quad (3.3) \end{aligned}$$

By Tonelli's theorem and observing that $E^0 \left\| P^{(m)} \left(X_{n+1}, X_n, u_n, \eta_n \right) \right\| \leq H_1$ where H_1 is a constant, (3.2) can be bounded above by

$$H_1 E^0 \left\| q_n \left(X_n; \eta^{n-1}, u^{n-1} \right) - q_n^{(m)} \left(X_n; \eta^{n-1}, u^{n-1} \right) \right\| \leq H_2 \overline{H}_m, \quad (3.4)$$

where we have applied the induction hypothesis. H_2 is a constant.

Applying again Tonelli's theorem, it can be shown (see [4]) that (3.3) is bounded above by $H_3 \overline{H}_m$, where H_3 is a constant. From here and (3.4) we obtain the thesis. ■

We denote by $J^{N,m}(u, \tau)$ the functional corresponding to (2.10) for the discretized state, that is obtained substituting the q_j by the $q_j^{(m)}$ in (2.10).

Proposition 3.1 allows us to state the following:

Proposition 3.2 *For N sufficiently large we have,*

$$\left| J^N(u, \tau) - J^{N,m}(u, \tau) \right| \leq N \left(\frac{1}{\delta} \tilde{L} + \tilde{Q} \right) K_1 \overline{H}_m \quad (3.5)$$

uniformly in $u \in \mathbf{V}_N$ *and* $\tau \in \left\{n\frac{T}{N}; n = 0, 1, ..., N\right\}$ *and where* K_1 *is a constant.*

It follows that for N and m sufficiently large we can approximate J^N by $J^{N,m}$ and look for an optimal control for the discretized problem. Again, by the uniformity of the bound in (u, τ), this latter control is nearly optimal for the original problem.

3.2 The discretized problem in alternative form

Just as q_n, also $q_n^{(m)}$ is infinite-dimensional but, since the function $P^{(m)}$ can be expressed as product of a function of the state at time $n+1$ and a function of the state and observation at time n, then the functions $q_n^{(m)}$ can be written in terms of finite-dimensional statistics of (η^{n-1}, u^{n-1}).

In fact, if we define, for $h = 1, \cdots, \overline{M}$,

$$d_1^{(h)}(\eta_0) := e^{\left[\frac{1}{\epsilon^2}x_0\eta_0 - \frac{1}{2\epsilon^2}x_0^2\frac{T}{N}\right]} I_{\pi_x(B_h)}(x_0)\, I_{\pi_y(B_h)}(y_0)\, I_{\pi_\theta(B_h)}(\theta_0), \quad (3.6)$$

where $\pi_k(B_h)$ is the projection of the set B_h on the k^{th} component,

$$d_{n+1}^{(h)}(\eta^n, u^{n-1}) := \sum_{\substack{h'=1 \\ \pi_\mu(B_{h'}) \equiv \pi_\mu(B_h)}}^{\overline{M}} d_n^{(h')}(\eta^{n-1}, u^{n-2})\, \varphi_{h,h'}(\eta_n, u_{n-1}), \quad (3.7)$$

and

$$\varphi_{h,h'}(\eta, u) := \int_A e^{\left[\frac{1}{\epsilon^2}x\eta - \frac{1}{2\epsilon^2}x^2\frac{T}{N}\right]} \Psi_{h'}((x, y, \theta), u)\, dx\, dy\, d\theta,$$

with $A = \pi_x(B_h) \times \pi_y(B_h) \times \pi_\theta(B_h)$ and $\Psi_h((x_{n+1}, y_{n+1}, \theta_{n+1}), u_n) := \Phi_1(x_{n+1}; X_h, u_n)\, \Phi_2(y_{n+1}; X_h, u_n)\, \Phi_3(\theta_{n+1}; X_h)$ and where X_h is the representative point of the set B_h, then it follows:

Proposition 3.3 *For each* $n = 1, ..., N$ *we have*

$$q_n^{(m)}(X_n; \eta^{n-1}, u^{n-1}) = \sum_{h=1}^{\overline{M}} d_n^{(h)}(\eta^{n-1}, u^{n-2})\, \Psi_h((x_n, y_n, \theta_n), u_{n-1}) \frac{1}{b-a} I_{\pi_\mu(B_h)}(\mu_n), \quad (3.8)$$

where $d_1^{(h)}(\eta^0, u^{-1}) = d_1^{(h)}(\eta_0)$.

The objective functional $J^{N,m}$ can be expressed as

$$E^0\left\{\frac{1-e^{-\delta\frac{T}{N}}}{\delta}s\left(y_0\right)U\left(C_0\right)I_{]0,T]}(\tau)-I_{[0,\frac{T}{N}[}\overline{Q}(0,X_0)\right.$$

$$+\sum_{n=1}^{N-1}\sum_{h=1}^{\overline{M}}d_n^{(h)}\left(\eta^{n-1},u^{n-2}\right)\left[\frac{e^{-\delta n\frac{T}{N}}-e^{-\delta(n+1)\frac{T}{N}}}{\delta}I_{]n\frac{T}{N},T]}\left(\tau\right)U\left(C_n\right)\right.$$

$$\cdot\int_{\mathbb{R}}s(y)\Phi_2\left(y;X_h,u_{n-1}\right)dy\int_{\pi_\mu(B_h)}\frac{1}{b-a}\,d\mu-I_{[n\frac{T}{N},(n+1)\frac{T}{N}[}(\tau)$$

$$\left.\cdot\int_{\mathbf{X}}\overline{Q}\left(n\frac{T}{N},X\right)\Psi_h\left((x,y,\theta),u_{n-1}\right)\frac{1}{b-a}I_{\pi_\mu(B_h)}(\mu)\,dX\right]$$

$$-\sum_{h=1}^{\overline{M}}d_N^{(h)}\left(\eta^{N-1},u^{N-2}\right)$$

$$\left.\cdot\int_{\mathbf{X}}\overline{Q}(T,X)I_{\{T\}}(\tau)\Psi_h\left((x,y,\theta),u_{N-1}\right)\frac{1}{b-a}I_{\pi_\mu(B_h)}(\mu)\,dX\right\} \quad (3.9)$$

where the expectation is with respect to the sequence $\{\eta_n\}$. We note that, under the measure P^0, $\{\eta_n\}$ is a sequence of i.i.d. random variables, normally distributed with mean 0 and variance $\epsilon^2\frac{T}{N}$.

The process $\left\{d_n^{(h)}\left(\eta^{n-1},u^{n-2}\right)\right\}$ is now finite-dimensional, but it still takes an infinite number of possible values since η_n and u_n do. It is then necessary to make some further approximations.

3.3 ϵ-optimality

The controls take values in a compact space $\mathbf{V}$. We can assume that also the observations take values in a compact space $\mathbf{S}$ since the expectation in (3.9) is finite, and if we restrict the values of η_n to a sufficiently large compact set $\mathbf{S}$, then, due to the boundedness of the costs, the corresponding change in the value of $J^{N,m}$ is, uniformly in u, negligible (see [6]).

Take finite partitions $\{\mathbf{V}_k\}_{k=1}^K$ and $\{\mathbf{S}_z\}_{z=1}^Z$ of the compact sets $\mathbf{V}$ and $\mathbf{S}$, respectively, and choose a representative element for each of the sets $v_k \in \mathbf{V}_k$ $(k=1,...,K)$ and $s_z \in \mathbf{S}_z$ $(z=1,...,Z)$.

We denote by $\mathbf{V}_N^K$ the set of the control sequences, taking as possible values the representative elements v_k and by $\mathbf{V}_N^{Kes}$ the set of the controls obtained from the controls of $\mathbf{V}_N^K$ by a step interpolation relative to the partition of the time interval into subintervals of the same width $\frac{T}{N}$.

Let $J^{N,m,Z}$ be the functional obtained from $J^{N,m}$ by substituting the observations and the controls with their discretized values.

From Propositions 2.1, 2.2 and 3.2, and the continuity in u and y of the functions in (3.9), we have the following two theorems (see [4]):

Theorem 3.1 *For $\epsilon > 0$ fixed, taking N, m, Z and K sufficiently large, we have*

$$\left| \sup_{\tau \in \left\{ n\frac{T}{N}; n=0,\dots,N \right\}} \sup_{u \in \mathbf{V}_N^K} J^{N,m,Z}(u,\tau) - \sup_{\tau \in [0,T]} \sup_{u \in \overline{\mathbf{V}}} J(u,\tau) \right| \leq \epsilon.$$

Theorem 3.2 *For each control $(u,\tau) \in \mathbf{V}_N \times \left\{ n\frac{T}{N}; n = 0, \dots, N \right\}$, for N, m, Z, K sufficiently large, we have*

$$\left| J^{N,m,Z}(u,\tau) - J(u,\tau) \right| \leq \epsilon.$$

Consequently, we have the following corollary.

Corollary 3.1 *An optimal control for the discretized control problem, extended in the sense given above (that is by a step interpolation), is ϵ-optimal for the original problem.*

From here it follows immediately that it sufficies to find an optimal control for the problem discretized in time and space; and the control obtained from the latter by a step interpolation (which is a control in $V_N^{K_{es}}$) will be nearly optimal for the original continuous time problem.

3.4 DP algorithm

Since the functional $J^{N,m,Z}$ has the same form as (3.9) but with the observations substituted by their discretized values, and since furthermore, the optimal value of the functional is calculated for controls $u \in \mathbf{V}_N^K$ with (3.6) and (3.7) as the state equations, we can write the DP algorithm to obtain an optimal control for this problem, in the following way:

$$\begin{aligned}
&\widehat{J}_N\left(\overline{\eta}^{N-1},\overline{u}^{N-1}\right)\\
&= -\sum_{h=1}^{\overline{M}} d_N^{(h)}\left(\overline{\eta}^{N-1},\overline{u}^{N-2}\right)\\
&\quad\left[\int_{\mathbf{X}} \overline{Q}(T,X) I_{\{T\}}(\tau)\Psi_h\left((x,y,\theta),\overline{u}_{N-1}\right)\frac{1}{b-a} I_{\pi_\mu(B_h)}(\mu)\,dX\right];
\end{aligned} \tag{3.10}$$

$$\begin{aligned}
&\widehat{J}_n\left(\overline{\eta}^{n-1},\overline{u}^{n-1}\right)\\
&= \max\left\{\max_{\overline{u}_n}\left[\sum_{h=1}^{\overline{M}} d_n^{(h)}\left(\overline{\eta}^{n-1},\overline{u}^{n-2}\right) U\left(\overline{C}_n\right)\right.\right.\\
&\quad\cdot\frac{e^{-\delta n\frac{T}{N}}-e^{-\delta(n+1)\frac{T}{N}}}{\delta}\int_{\mathbb{R}} s(y)\Phi_2\left(y;X_h,\overline{u}_{n-1}\right)dy\int_{\pi_\mu(B_h)}\frac{1}{b-a}\,d\mu\\
&\quad\left.+\sum_{z=1}^{Z} P^0\left(\eta\in S_z\right)\widehat{J}_{n+1}\left(\left(\overline{\eta}^{n-1},s_z\right),\left(\overline{u}^{n-1},\overline{u}_n\right)\right)\right];\\
&\quad-\sum_{h=1}^{\overline{M}-1} d_n^{(h)}\left(\overline{\eta}^{n-1},\overline{u}^{n-2}\right)\left[\int_{\mathbf{X}}\overline{Q}\left(n\frac{T}{N}X\right)\Psi_h\left((x,y,\theta),\overline{u}_{n-1}\right)\right.\\
&\quad\left.\left.\cdot\frac{1}{b-a} I_{\pi_\mu(B_h)}(\mu)\,dX\right]\right\},
\end{aligned} \tag{3.11}$$

for $n = N-1, ..., 1$.

$$\begin{aligned}
\widehat{J}_0 &= \max\left\{\max_{\overline{u}_0}\left[\frac{1-e^{-\delta\frac{T}{N}}}{\delta}s\left(y_0\right)U\left(\overline{C}_0\right)\right.\right.\\
&\quad\left.\left.+\sum_{z=1}^{Z} P^0\left(\eta\in S_z\right)\widehat{J}_1\left(s_z,\overline{u}_0\right)\right];-\overline{Q}\left(0,X_0\right)\right\}.
\end{aligned} \tag{3.12}$$

By $\overline{\eta}_n$ we denote the discretized observations and by $\overline{u}_n$ the discretized controls. We have used the fact that $\overline{\eta}_n$ can take only a finite number Z of possible values. We observe also that the dependence on the current control is in the term $U\left(\overline{C}_n\right)$.

Remark 3.1 *At each step it is necessary to calculate two rewards: one in the case in which we stop at that moment and the other in the case in which we decide to go on. We then choose the maximum between the two rewards in order to know the optimal stopping time τ for each pair of observations and controls.*

The DP algorithm provides us with the sequence of optimal controls (that is the optimal strategy) to be applied at times $\{n\frac{T}{N}; n = 0, 1, ..., N-1\}$ as function of the current statistic $\left(d_n^{(h)}\left(\overline{\eta}^{n-1}, \overline{u}^{n-2}\right), \overline{u}_{n-1}\right)$. Extending this sequence in the sense described previously we obtain a nearly optimal strategy for the initial problem.

This DP algorithm can be implemented in order to determine the optimal strategy. In [4] one can find a discussion of the numerical aspects as well as numerical results.

References

[1] A. Bensoussan, W. Runggaldier, An approximation method for stochastic control problems with partial observation of the state - a method for constructing ϵ-optimal controls, *Acta Appl. Math.*, **10** (1987), 145–170.

[2] A.P. Gutierrez, U. Regev, S.J. Schreiber, D. Zilbermann, Biological and Economic Foundation of Renewable Resource Exploitation, *Ecological Economics*, **26**(3) (1998), 227–242.

[3] H.J. Kushner, *Probability methods for approximations in stochastic control and for elliptic equations*, Academic Press, New York,1977.

[4] S.Pasquali, *A stochastic control problem for renewable resource exploitation*, Ph.D. Thesis, University of Padova, 1999.

[5] W.J. Runggaldier, L. Stettner, *Approximations of Discrete Time Partially Observed Control Problems*, Giardini Editiori e Stampatori in Pisa, 1994.

[6] W.J. Runggaldier, O. Zane, Approximations for discrete-time adaptive control: construction of ϵ-optimal controls, *Math. Control, Signals and Systems*, **4** (1991), 269-291.

III

STOCHASTIC PROCESSES AND MARTINGALES

Chapter 17

A FLEMING-VIOT PROCESS WITH UNBOUNDED SELECTION, II

S. N. Ethier
University of Utah

Tokuzo Shiga
Tokyo Institute of Technology

Abstract In a previous paper the authors studied a Fleming–Viot process with house-of-cards mutation and an unbounded haploid selection intensity function. Results included existence and uniqueness of solutions of an appropriate martingale problem, existence, uniqueness, and reversibility of stationary distributions, and a weak limit theorem for a corresponding sequence of Wright–Fisher models. In the present paper we extend these results to the diploid setting. The existence and uniqueness results carry over fairly easily, but the limit theorem is more difficult and requires new ideas.

1. Introduction

In a previous paper (Ethier and Shiga [1]), the authors studied a Fleming–Viot process with house-of-cards (or parent-independent) mutation and an unbounded haploid selection intensity function. More specifically, the set of possible alleles, known as the type space, is a locally compact, separable metric space E, so the state space for the process is a subset of $\mathcal{P}(E)$, the set of Borel probability measures on E; the mutation operator A on $B(E)$, the space of bounded Borel functions on E, is given by

$$Af = \tfrac{1}{2}\theta(\langle f, \nu_0 \rangle - f), \tag{1.1}$$

Z. Hou et al. (eds.), Markov Processes and Controlled Markov Chains, 305–322.

where $\theta > 0$, $\nu_0 \in \mathcal{P}(E)$, and $\langle f, \mu \rangle := \int_E f \, d\mu$; and the selection intensity (or scaled selection coefficient) of allele $x \in E$ is $\hat{h}(x)$, where $\hat{h}$ is a Borel function on E.

Assuming the existence of a continuous function $h_0 : E \mapsto [0, \infty)$ and a constant $\rho_0 \in (1, \infty]$ such that

$$|\hat{h}| \leq h_0, \qquad \langle e^{\rho h_0}, \nu_0 \rangle < \infty \quad \text{whenever} \quad 0 < \rho < \rho_0, \tag{1.2}$$

existence and uniqueness of solutions of an appropriate martingale problem were established, and a weak limit theorem for a corresponding sequence of Wright–Fisher models was proved (at least when $\hat{h}$ is continuous). Assuming also that $\rho_0 > 2$, existence, uniqueness, and reversibility of stationary distributions were obtained as well.

In the present paper we extend these results to the diploid setting, in which case $\hat{h}$ is replaced by a symmetric Borel function h on $E^2 := E \times E$, with $h(x, y)$ representing the selection intensity (or scaled selection coefficient) of the genotype $\{x, y\}$. We replace the first inequality in (1.2) by

$$|h(x, y)| \leq h_0(x) + h_0(y), \qquad (x, y) \in E^2, \tag{1.3}$$

but the other assumptions in (1.2) remain unchanged. This condition is in effect throughout the paper.

Overbeck *et al.* [2] introduced a more general type of selection, called interactive selection, in which $\hat{h}$ is allowed to depend on μ. (Diploid selection is just the special case $\hat{h}_\mu(y) := \langle h(\cdot, y), \mu \rangle$.) They too allowed for unbounded selection intensity functions. It seems unlikely that our results can be extended to this level of generality.

Our previous paper was motivated by the nearly neutral mutation model (or normal-selection model) of Tachida [3], which assumed additive diploid selection, that is,

$$h(x, y) = \hat{h}(x) + \hat{h}(y), \qquad (x, y) \in E^2. \tag{1.4}$$

This is, of course, mathematically equivalent to the haploid case treated in [1].

The generator of the Fleming–Viot process in question will be denoted by $\mathcal{L}_h$. It acts on functions φ on $\mathcal{P}(E)$ of the form

$$\varphi(\mu) = F(\langle f_1, \mu \rangle, \ldots, \langle f_k, \mu \rangle) = F(\langle \mathbf{f}, \mu \rangle), \tag{1.5}$$

where $k \geq 1$, $f_1, \ldots, f_k \in \overline{C}(E)$ (the space of bounded continuous functions on E), and $F \in C^2(\mathbf{R}^k)$, according to the formula

$$\begin{aligned}(\mathcal{L}_h\varphi)(\mu) &= \tfrac{1}{2}\sum_{i,j=1}^{k}(\langle f_if_j,\mu\rangle - \langle f_i,\mu\rangle\langle f_j,\mu\rangle)\,F_{z_iz_j}(\langle \mathbf{f},\mu\rangle)\\ &\quad+\sum_{i=1}^{k}\langle Af_i,\mu\rangle\,F_{z_i}(\langle \mathbf{f},\mu\rangle)\\ &\quad+\sum_{i=1}^{k}(\langle (f_i\circ\pi)h,\mu^2\rangle - \langle f_i,\mu\rangle\langle h,\mu^2\rangle)\,F_{z_i}(\langle \mathbf{f},\mu\rangle), \quad (1.6)\end{aligned}$$

where $\mu^2 := \mu\times\mu \in \mathcal{P}(E^2)$ and $\pi : E^2 \mapsto E$ is the projection map $\pi(x,y) = x$. This suffices if h is bounded, but if not, we need to restrict the state space to a suitable subset of $\mathcal{P}(E)$. We use the same state space as in [1], namely the set of Borel probability measures μ on E that satisfy the condition imposed on ν_0 in (1.2).

We therefore define

$$\mathcal{P}^\circ(E) = \{\mu \in \mathcal{P}(E) : \langle e^{\rho h_0},\mu\rangle < \infty \text{ for each } \rho \in (0,\rho_0)\} \qquad (1.7)$$

and, for $\mu,\nu \in \mathcal{P}^\circ(E)$,

$$d^\circ(\mu,\nu) = d(\mu,\nu) + \int_{(0,\rho_0)}\Big(1\wedge \sup_{0\leq\rho\leq r}|\langle e^{\rho h_0},\mu\rangle - \langle e^{\rho h_0},\nu\rangle|\Big)e^{-r}\,dr, \qquad (1.8)$$

where d is a metric on $\mathcal{P}(E)$ that induces the topology of weak convergence. Then $(\mathcal{P}^\circ(E), d^\circ)$ is a complete separable metric space and $d^\circ(\mu_n,\mu) \to 0$ if and only if $\mu_n \Rightarrow \mu$ and $\sup_n\langle e^{\rho h_0},\mu_n\rangle < \infty$ for each $\rho \in (0,\rho_0)$.

Section 2 establishes existence and uniqueness of solutions of the appropriate martingale problem for $\mathcal{L}_h$. Section 4 establishes existence, uniqueness, and reversibility of the stationary distribution of the resulting Fleming–Viot process. The proofs of these results are similar to those in the haploid case, so we point out only the necessary changes. Section 3 gives a precise description of the measure-valued Wright–Fisher model considered here and proves, assuming continuity of h, a weak convergence result that justifies the diffusion approximation of that model by the Fleming–Viot process with generator $\mathcal{L}_h$. The proof is more difficult than that in the haploid case, so most of the present paper will be concerned with this result.

2. Characterization of the process

Let $\Omega := C_{(\mathcal{P}(E),d)}[0,\infty)$ have the topology of uniform convergence on compact sets, let $\mathcal{F}$ be the Borel σ-field, let $\{\mu_t,\ t \geq 0\}$ be the canonical coordinate process, and let $\{\mathcal{F}_t\}$ be the corresponding filtration.

If $h_1, h_2 \in B(E^2)$, define $\Psi(h_1,h_2) \in B(E^3)$ and $\Upsilon h_2 \in B(E)$ by

$$\Psi(h_1,h_2)(x,y,z) = h_1(x,y)h_2(x,z), \qquad (\Upsilon h_2)(x) = h_2(x,x). \tag{2.1}$$

The analogue of Lemma 2.1 of [1] is as follows.

Lemma 2.1 *Let $h_1, h_2 \in B(E^2)$. If $P \in \mathcal{P}(\Omega)$ is a solution of the martingale problem for $\mathcal{L}_{h_1}$, then*

$$\begin{aligned} R_t := \exp\Bigg\{ & \tfrac{1}{2}\langle h_2, \mu_t^2\rangle - \tfrac{1}{2}\langle h_2, \mu_0^2\rangle \\ & - \int_0^t \Big[\tfrac{1}{2}(\langle \Psi(h_2,h_2), \mu_s^3\rangle - \langle h_2, \mu_s^2\rangle^2) \\ & \quad + \tfrac{1}{2}(\langle \Upsilon h_2, \mu_s\rangle - \langle h_2, \mu_s^2\rangle) + \tfrac{1}{2}\theta(\langle h_2, \mu_s \times \nu_0\rangle - \langle h_2, \mu_s^2\rangle) \\ & \quad + \langle \Psi(h_1,h_2), \mu_s^3\rangle - \langle h_1, \mu_s^2\rangle\langle h_2, \mu_s^2\rangle \Big] ds \Bigg\} \end{aligned} \tag{2.2}$$

is a mean-one $\{\mathcal{F}_t\}$-martingale on $(\Omega, \mathcal{F}, P)$. Furthermore, the measure $Q \in \mathcal{P}(\Omega)$ defined by

$$dQ = R_t\, dP \text{ on } \mathcal{F}_t, \qquad t \geq 0, \tag{2.3}$$

is a solution of the martingale problem for $\mathcal{L}_{h_1+h_2}$.

Informally, the integrand in (2.2) is simply $e^{-\frac{1}{2}\langle h_2,\mu^2\rangle}\mathcal{L}_{h_1} e^{\frac{1}{2}\langle h_2,\mu^2\rangle}$ at $\mu = \mu_s$. Strictly speaking, $e^{\frac{1}{2}\langle h_2,\mu^2\rangle}$ does not belong to the domain of $\mathcal{L}_{h_1}$ because it is not of the form (1.5), but the domain can be extended to include such functions. Of course, $\mu^3 := \mu \times \mu \times \mu$ in (2.2).

We will need the following simple observation.

Lemma 2.2 *For each $g \in B(E^2)$ and $\mu \in \mathcal{P}(E)$, we have $\langle \Psi(g,g), \mu^3\rangle - \langle g, \mu^2\rangle^2 \geq 0$, where $\Psi(g,g)$ is as in (2.1).*

Proof Let X, Y, Z be i.i.d. μ. Then, letting $\bar{g} = \langle g, \mu^2 \rangle$,

$$\begin{aligned}
\langle \Psi(g,g), \mu^3 \rangle - \langle g, \mu^2 \rangle^2 &= \mathbf{Cov}(g(X,Y), g(X,Z)) \\
&= \int_E \mathbf{E}[(g(x,Y) - \bar{g})(g(x,Z) - \bar{g})]\, \mu(dx) \\
&= \int_E \mathbf{E}[g(x,Y) - \bar{g}]\mathbf{E}[g(x,Z) - \bar{g}]\, \mu(dx) \\
&= \int_E (\mathbf{E}[g(x,Y) - \bar{g}])^2\, \mu(dx) \\
&\geq 0. \qquad (2.4)
\end{aligned}$$

■

We now define

$$\Omega^\circ = C_{(\mathcal{P}^\circ(E),d^\circ)}[0,\infty) \subset \Omega = C_{(\mathcal{P}(E),d)}[0,\infty), \qquad (2.5)$$

and let Ω° have the topology of uniform convergence on compact sets. The domain of $\mathcal{L}_h$ is the space of functions φ on $\mathcal{P}^\circ(E)$ of the form (1.5).

Theorem 2.1 *For each $\mu \in \mathcal{P}^\circ(E)$, the Ω° martingale problem for $\mathcal{L}_h$ starting at μ has one and only one solution.*

Proof The proof is similar to that of Theorem 2.5 of [1]. The only changes necessary are to equations (2.11)–(2.13) and (2.19)–(2.22) in the proofs of Lemmas 2.3 and 2.4 of [1]. Lemma 2.2 above disposes of the only awkward term, and otherwise the argument is essentially as before. ■

3. Diffusion approximation of the Wright-Fisher model

We begin by formulating a Wright–Fisher model with house-of-cards mutation and diploid selection. It depends on several parameters, some of which have already been introduced:

- E (a locally compact, separable metric space) is the set of possible alleles, and is known as the type space.
- M (a positive integer) is twice the diploid population size. (Most authors use $2N$ here, but we prefer to absorb the ubiquitous factor of 2. In fact, M need not be even.)
- u (in [0,1]) is the mutation rate (i.e., probability) per gene per generation.

- ν_0 (in $\mathcal{P}(E)$) is the distribution of the type of a new mutant; this is the house-of-cards assumption.

- $w(x, y)$ (a positive symmetric Borel function of $(x, y) \in E^2$) is the fitness of genotype $\{x, y\}$.

The Wright–Fisher model is a Markov chain modelling the evolution of the population's composition. The state space for the process is

$$\mathcal{P}_M(E) := \left\{ \frac{1}{M} \sum_{i=1}^{M} \delta_{x_i} \in \mathcal{P}(E) : (x_1, \ldots, x_M) \in E \times \cdots \times E \right\} \quad (3.1)$$

with the topology of weak convergence, where $\delta_x \in \mathcal{P}(E)$ is the unit mass at x. Time is discrete and measured in generations. The transition mechanism is specified by

$$\mu := \frac{1}{M} \sum_{i=1}^{M} \delta_{x_i} \mapsto \frac{1}{M} \sum_{i=1}^{M} \delta_{Y_i}, \quad (3.2)$$

where

$$Y_1, \ldots, Y_M \text{ are i.i.d. } \mu^{**} \qquad \text{[random sampling]}, \quad (3.3)$$

$$\mu^{**} = (1-u)\mu^* + u\nu_0 \qquad \text{[house-of-cards mutation]}, \quad (3.4)$$

$$\mu^*(\Gamma) = \int_\Gamma \langle w(\cdot, y), \mu \rangle \, \mu(dy) \Big/ \langle w, \mu^2 \rangle \quad \text{[diploid selection]}. \quad (3.5)$$

This suffices to describe the Wright–Fisher model in terms of the parameters listed above.

However, since we are interested in a diffusion approximation, we further assume that

$$u = \frac{\theta}{2M}, \qquad w(x, y) = \exp\left\{ \frac{h(x, y)}{M} \right\}, \quad (3.6)$$

where θ is a positive constant and h is as in (1.3).

The aim here is to prove, assuming the continuity of h, that convergence in $\mathcal{P}^\circ(E)$ of the initial distributions implies convergence in distribution in Ω° of the sequence of rescaled and linearly interpolated Wright–Fisher models to a Fleming–Viot process with generator $\mathcal{L}_h$. We postpone a careful statement of the result to the end of the section.

The strategy of the proof is as in [1]. Lemmas 3.4 and 3.5 of [1] must be substantially modified however. For the two corresponding lemmas we require, as in [1], the infinitely-many-alleles assumption that every mutant is of a type that has not previously appeared. Mathematically, this amounts to

$$\nu_0(\{x\}) = 0, \qquad x \in E. \tag{3.7}$$

Let $\Xi_M := \mathcal{P}_M(E)^{\mathbf{Z}_+}$ have the product topology, let $\mathcal{F}$ be the Borel σ-field, let $\{\mu_n,\ n = 0, 1, \ldots\}$ be the canonical coordinate process, and let $\{\mathcal{F}_n\}$ be the corresponding filtration. For each $\mu \in \mathcal{P}_M(E)$, we denote by $P_\mu^{(M)}$ and $Q_\mu^{(M)}$ in $\mathcal{P}(\Xi_M)$ the distributions of the neutral and selective Wright–Fisher models, respectively, starting at μ.

Lemma 3.1 *Assume* (3.7). *Then, for each* $\mu \in \mathcal{P}_M(E)$,

$$dQ_\mu^{(M)} = R_n^{(M)}\, dP_\mu^{(M)} \text{ on } \mathcal{F}_n, \qquad n \geq 0, \tag{3.8}$$

where

$$\begin{aligned} R_n^{(M)} &= \exp\left\{ \sum_{k=1}^n \int_E 1_{\operatorname{supp}\mu_{k-1}}(y) M \log \left\langle e^{h(\cdot,y)/M}, \mu_{k-1}\right\rangle \mu_k(dy) \right.\\ &\qquad \left. - \sum_{k=1}^n \langle 1_{\operatorname{supp}\mu_{k-1}}, \mu_k\rangle M \log \left\langle e^{h/M}, \mu_{k-1}^2\right\rangle \right\}. \end{aligned} \tag{3.9}$$

Proof The proof is as in that of Lemma 3.4 of [1], except that, if $\mu_1 = M^{-1}\sum_{j=1}^M \delta_{y_j}$,

$$\begin{aligned} V^{(M)}(\mu_0,\mu_1) &:= \frac{\prod_{1\leq i\leq M: y_i \in \operatorname{supp}\mu_0} \langle w(\cdot, y_i), \mu_0\rangle}{\langle w, \mu_0^2\rangle^{|\{1\leq i\leq M: y_i\in \operatorname{supp}\mu_0\}|}} \\ &= \frac{\exp\{\int_E 1_{\operatorname{supp}\mu_0}(y) M \log\langle w(\cdot,y),\mu_0\rangle\, \mu_1(dy)\}}{\langle w, \mu_0^2\rangle^{M\langle 1_{\operatorname{supp}\mu_0},\mu_1\rangle}} \\ &= \exp\left\{ \int_E 1_{\operatorname{supp}\mu_0}(y) M \log\langle e^{h(\cdot,y)/M}, \mu_0\rangle\, \mu_1(dy) \right.\\ &\qquad \left. - \langle 1_{\operatorname{supp}\mu_0}, \mu_1\rangle M \log\langle e^{h/M}, \mu_0^2\rangle \right\}. \end{aligned} \tag{3.10}$$

■

We define the map $\Phi_M : \Xi_M \mapsto \Omega^\circ$ by

$$\Phi_M(\mu_0,\mu_1,\ldots)_t = (1-(Mt-[Mt]))\mu_{[Mt]} + (Mt-[Mt])\mu_{[Mt]+1}. \tag{3.11}$$

This transformation maps a discrete-time process to a continuous-time one with continuous piecewise-linear sample paths, re-scaling time by a factor of M.

We next show that the Girsanov-type formula for the Wright–Fisher model converges in some sense to the one for the Fleming–Viot process. Define $\hat{R}_t^{(M)}$ on Ω° for all $t \geq 0$ so as to satisfy

$$\hat{R}_t^{(M)} \circ \Phi_M = R_{[Mt]}^{(M)} \text{ on } \Xi_M, \qquad t \geq 0, \tag{3.12}$$

where $R_n^{(M)}$ is as in Lemma 3.1. Specifically, we take

$$\begin{aligned} &\hat{R}_t^{(M)} \\ &= \exp\left\{ \sum_{k=1}^{[Mt]} \int_E 1_{\operatorname{supp}\mu_{(k-1)/M}}(y) M \log \left\langle e^{h(\cdot,y)/M}, \mu_{(k-1)/M} \right\rangle \mu_{k/M}(dy) \right. \\ &\qquad \left. - \sum_{k=1}^{[Mt]} \langle 1_{\operatorname{supp}\mu_{(k-1)/M}}, \mu_{k/M} \rangle M \log \left\langle e^{h/M}, \mu_{(k-1)/M}^2 \right\rangle \right\}. \end{aligned} \tag{3.13}$$

We also define R_t on Ω° for all $t \geq 0$ as in Lemma 2.1 with $h_1 = 0$ and $h_2 = h$; specifically,

$$\begin{aligned} R_t := \exp\Big\{ &\tfrac{1}{2}\langle h, \mu_t^2 \rangle - \tfrac{1}{2}\langle h, \mu_0^2 \rangle \\ &- \int_0^t \left[\tfrac{1}{2}(\langle \Psi(h,h), \mu_s^3 \rangle - \langle h, \mu_s^2 \rangle^2) + \tfrac{1}{2}(\langle \Upsilon h, \mu_s \rangle - \langle h, \mu_s^2 \rangle) \right. \\ &\qquad \left. + \tfrac{1}{2}\theta(\langle h, \mu_s \times \nu_0 \rangle - \langle h, \mu_s^2 \rangle) \right] ds \Big\}. \end{aligned} \tag{3.14}$$

Lemma 3.2 *Let $\{\mu^{(M)}\} \subset \mathcal{P}_M(E) \subset \mathcal{P}^\circ(E)$ and $\mu \in \mathcal{P}^\circ(E)$ be such that $d^\circ(\mu^{(M)}, \mu) \to 0$. For simplicity of notation, denote $P_{\mu^{(M)}}^{(M)}$, which is defined as in the paragraph preceding* Lemma 3.1, *by just $P^{(M)}$. Assume that h is continuous and* (3.7) *holds, and let $T > 0$ be arbitrary. Then there exist Borel functions $F_M, G_M : \Omega^\circ \mapsto (0,\infty)$, a continuous function $F : \Omega^\circ \mapsto (0,\infty)$, and a positive constant G such that*

$$\hat{R}_T^{(M)} = F_M G_M, \qquad R_T = FG; \tag{3.15}$$

in addition, $F_M \to F$ uniformly on compact subsets of Ω°, and $G_M \to G$ in $P^{(M)}\Phi_M^{-1}$-probability.

Proof From (3.13) we get

$$\begin{aligned}\log \hat{R}_T^{(M)} &= \sum_{k=1}^{[MT]} \int_E M \log \left\langle e^{h(\cdot,y)/M}, \mu_{(k-1)/M} \right\rangle \mu_{k/M}(dy) \\ &\quad - \sum_{k=1}^{[MT]} \int_E 1_{(\mathrm{supp}\, \mu_{(k-1)/M})^c}(y) \\ &\qquad\qquad M \log \left\langle e^{h(\cdot,y)/M}, \mu_{(k-1)/M} \right\rangle \mu_{k/M}(dy) \\ &\quad - \sum_{k=1}^{[MT]} M \log \left\langle e^{h/M}, \mu^2_{(k-1)/M} \right\rangle \\ &\quad + \sum_{k=1}^{[MT]} \langle 1_{(\mathrm{supp}\, \mu_{(k-1)/M})^c}, \mu_{k/M} \rangle M \log \left\langle e^{h/M}, \mu^2_{(k-1)/M} \right\rangle \\ &=: S_1 - S_2 - S_3 + S_4. \end{aligned} \tag{3.16}$$

First

$$\begin{aligned} S_4 &= \sum_{k=1}^{[MT]} \left\{ M \langle 1_{(\mathrm{supp}\, \mu_{(k-1)/M})^c}, \mu_{k/M} \rangle - \tfrac{1}{2}\theta \right\} \\ &\qquad \times \log \left\langle e^{h/M}, \mu^2_{(k-1)/M} \right\rangle + \tfrac{1}{2}\theta \sum_{k=1}^{[MT]} \log \left\langle e^{h/M}, \mu^2_{(k-1)/M} \right\rangle \\ &=: S_4' + S_4''. \end{aligned} \tag{3.17}$$

By the argument using (3.43)–(3.46) of [1], S_4' goes to 0 in $P^{(M)}\Phi_M^{-1}$-probability. By a slight modification of (3.37) of [1], S_4'' converges to $\frac{1}{2}\theta \int_0^T \langle h, \mu_s^2 \rangle \, ds$ uniformly on compact sets (see the discussion following (3.39) of [1]).

Next

$$\begin{aligned} S_2 &= \sum_{k=1}^{[MT]} \left\{ \int_E 1_{(\mathrm{supp}\, \mu_{(k-1)/M})^c}(y) \right. \\ &\qquad \times M \log \langle e^{h(\cdot,y)/M}, \mu_{(k-1)/M} \rangle \, \mu_{k/M}(dy) \\ &\qquad \left. - \tfrac{1}{2}\theta \int_E \log \left\langle e^{h(\cdot,y)/M}, \mu_{(k-1)/M} \right\rangle \nu_0(dy) \right\} \\ &\quad + \tfrac{1}{2}\theta \sum_{k=1}^{[MT]} \int_E \log \left\langle e^{h(\cdot,y)/M}, \mu_{(k-1)/M} \right\rangle \nu_0(dy) \\ &=: S_2' + S_2''. \end{aligned} \tag{3.18}$$

We can argue as in (3.41) of [1] that S_2' is equal in $P^{(M)}\Phi_M^{-1}$-distribution to

$$\frac{1}{M}\sum_{k=1}^{[MT]}\Bigg(\sum_{l=1}^{X_k} M\log\left\langle e^{h(\cdot,\xi_{kl})/M},\mu_{(k-1)/M}\right\rangle - \tfrac{1}{2}\theta M\int_E \log\left\langle e^{h(\cdot,y)/M},\mu_{(k-1)/M}\right\rangle \nu_0(dy)\Bigg), \quad (3.19)$$

where X_k is the number of new mutants in generation k, and ξ_{kl} is the type of the lth new mutant in generation k. Note that X_k and (ξ_{kl}) are independent binomial$(M,\theta/(2M))$ and ν_0^∞-distributed, respectively, and independent of $\mu_{(k-1)/M}$ (but not of $\mu_{k/M}$). It can be shown, analogously to (3.42) of [1], that S_2' goes to 0 in probability, while of course S_2'' goes to $\frac{1}{2}\theta\int_0^T\langle h,\mu_s\times\nu_0\rangle\,ds$ uniformly on compact sets.

Finally

$$\begin{aligned} S_1 - S_3 &= \sum_{k=1}^{[MT]}\Bigg\{\int_E M\log\left\langle e^{h(\cdot,y)/M},\mu_{(k-1)/M}\right\rangle \mu_{k/M}(dy) \\ &\qquad -\langle h,\mu_{(k-1)/M}\times\mu_{k/M}\rangle \\ &\qquad -\left(M\log\left\langle e^{h/M},\mu^2_{(k-1)/M}\right\rangle - \left\langle h,\mu^2_{(k-1)/M}\right\rangle\right)\Bigg\} \\ &\quad + \sum_{k=1}^{[MT]}\left(\langle h,\mu_{(k-1)/M}\times\mu_{k/M}\rangle - \left\langle h,\mu^2_{(k-1)/M}\right\rangle\right) \\ &=: S_{13}' + S_{13}''. \end{aligned} \quad (3.20)$$

Using essentially [1, equation (3.37)], we have

$$\begin{aligned} S_{13}' &= \frac{1}{M}\sum_{k=1}^{[MT]}\Bigg\{\tfrac{1}{2}\int_E \left(\langle h^2(\cdot,y),\mu_{(k-1)/M}\rangle - \langle h(\cdot,y),\mu_{(k-1)/M}\rangle^2\right) \\ &\qquad \times\mu_{k/M}(dy) - \tfrac{1}{2}\left(\left\langle h^2,\mu^2_{(k-1)/M}\right\rangle - \left\langle h,\mu^2_{(k-1)/M}\right\rangle^2\right)\Bigg\} \\ &\quad + O(M^{-1}) \\ &= -\tfrac{1}{2}\int_0^T\left(\langle\Psi(h,h),\mu_s^3\rangle - \langle h,\mu_s^2\rangle^2\right)ds + o(1), \end{aligned} \quad (3.21)$$

and the convergence is uniform on compact sets.

It remains to consider S''_{13}. Note that, in the case of additive diploid selection (specifically, (1.4)), S''_{13} reduces to the telescoping sum

$$\sum_{k=1}^{[MT]} (\langle \hat{h}, \mu_{k/M} \rangle - \langle \hat{h}, \mu_{(k-1)/M} \rangle) = \langle \hat{h}, \mu_{[MT]/M} \rangle - \langle \hat{h}, \mu_0 \rangle. \qquad (3.22)$$

Here we have to work harder. By the symmetry of h and some algebra,

$$\begin{aligned} S''_{13} &:= \sum_{k=1}^{[MT]} \Big(\langle h, \mu_{(k-1)/M} \times \mu_{k/M} \rangle - \langle h, \mu^2_{(k-1)/M} \rangle \Big) \\ &= \sum_{k=1}^{[MT]} \Big(\tfrac{1}{2} \langle h, \mu_{(k-1)/M} \times \mu_{k/M} \rangle \\ &\qquad + \tfrac{1}{2} \langle h, \mu_{k/M} \times \mu_{(k-1)/M} \rangle - \langle h, \mu^2_{(k-1)/M} \rangle \Big) \\ &= \tfrac{1}{2} \sum_{k=1}^{[MT]} \Big(\langle h, \mu^2_{k/M} \rangle - \langle h, \mu^2_{(k-1)/M} \rangle \Big) \\ &\qquad - \tfrac{1}{2} \sum_{k=1}^{[MT]} \Big(\langle h, \mu^2_{k/M} \rangle - \langle h, \mu_{k/M} \times \mu_{(k-1)/M} \rangle \\ &\qquad\qquad - \langle h, \mu_{(k-1)/M} \times \mu_{k/M} \rangle + \langle h, \mu^2_{(k-1)/M} \rangle \Big) \\ &= \tfrac{1}{2} \Big(\langle h, \mu^2_{[MT]/M} \rangle - \langle h, \mu^2_0 \rangle \Big) - \tfrac{1}{2} \sum_{k=1}^{[MT]} \langle h, (\mu_{k/M} - \mu_{(k-1)/M})^2 \rangle \\ &=: \tfrac{1}{2} \Sigma_1 - \tfrac{1}{2} \Sigma_2. \end{aligned} \qquad (3.23)$$

The last sum, Σ_2, is the integral of h with respect to the quadratic variation of the Markov chain. Let us write

$$\begin{aligned} \Sigma_2 &= \sum_{k=1}^{[MT]} \Big\{ \langle h, (\mu_{k/M} - \mu_{(k-1)/M})^2 \rangle \\ &\qquad - M^{-1} \Big(\langle \Upsilon h, \mu_{(k-1)/M} \rangle - \langle h, \mu^2_{(k-1)/M} \rangle \Big) \Big\} \\ &\qquad + \frac{1}{M} \sum_{k=1}^{[MT]} \Big(\langle \Upsilon h, \mu_{(k-1)/M} \rangle - \langle h, \mu^2_{(k-1)/M} \rangle \Big) \\ &=: \Sigma'_2 + \Sigma''_2. \end{aligned} \qquad (3.24)$$

Of course, Σ''_2 goes to $\int_0^T (\langle \Upsilon h, \mu_s \rangle - \langle h, \mu_s^2 \rangle)\, ds$ uniformly on compact sets.

We claim that Σ_2' goes to 0 in $L^2(P^{(M)}\Phi_M^{-1})$, hence in $P^{(M)}\Phi_M^{-1}$-probability. Given $k \geq 1$, let $Y_1, \ldots, Y_M$ be, conditionally on μ_{k-1}, i.i.d. $\mu_{k-1}^* := (1-u)\mu_{k-1} + u\nu_0$. Then

$$
\begin{aligned}
&\mathbf{E}^{P^{(M)}}\left[\left\langle h, (\mu_k - \mu_{k-1})^2 \right\rangle\right] \\
&= \mathbf{E}^{P^{(M)}}\left[\left\langle h, \mu_k^2 \right\rangle - \langle h, \mu_k \times \mu_{k-1}\rangle - \langle h, \mu_{k-1} \times \mu_k\rangle + \left\langle h, \mu_{k-1}^2 \right\rangle\right] \\
&= \mathbf{E}^{P^{(M)}}\Bigg[\mathbf{E}\Bigg[\frac{1}{M^2}\sum_{i,j=1}^{M} h(Y_i, Y_j) - \frac{1}{M}\sum_{i=1}^{M}\langle h(Y_i, \cdot), \mu_{k-1}\rangle \\
&\qquad - \frac{1}{M}\sum_{j=1}^{M}\langle h(\cdot, Y_j), \mu_{k-1}\rangle + \left\langle h, \mu_{k-1}^2 \right\rangle\Bigg]\Bigg] \\
&= \mathbf{E}^{P^{(M)}}\Big[M^{-1}\left\langle \Upsilon h, \mu_{k-1}^* \right\rangle + (1 - M^{-1})\left\langle h, (\mu_{k-1}^*)^2 \right\rangle \\
&\qquad - \left\langle h, \mu_{k-1}^* \times \mu_{k-1} \right\rangle - \left\langle h, \mu_{k-1} \times \mu_{k-1}^* \right\rangle + \left\langle h, \mu_{k-1}^2 \right\rangle\Big] \\
&= M^{-1}\mathbf{E}^{P^{(M)}}\left[\left\langle \Upsilon h, \mu_{k-1}^* \right\rangle - \left\langle h, (\mu_{k-1}^*)^2 \right\rangle\right] \\
&\qquad + \mathbf{E}^{P^{(M)}}\left[\left\langle h, (\mu_{k-1}^* - \mu_{k-1})^2 \right\rangle\right] \\
&= M^{-1}\mathbf{E}^{P^{(M)}}\left[\langle \Upsilon h, \mu_{k-1}\rangle - \left\langle h, \mu_{k-1}^2 \right\rangle\right] + O\left(M^{-2}\right), \qquad (3.25)
\end{aligned}
$$

and this holds uniformly in $k \geq 1$. Consequently,

$$
\begin{aligned}
&\mathbf{E}^{P^{(M)}\Phi_M^{-1}}\left[(\Sigma_2')^2\right] \\
&= \mathbf{E}^{P^{(M)}}\Bigg[\Bigg(\sum_{k=1}^{[MT]}\Big\{\left\langle h, (\mu_k - \mu_{k-1})^2 \right\rangle \\
&\qquad - M^{-1}\left(\langle \Upsilon h, \mu_{k-1}\rangle - \left\langle h, \mu_{k-1}^2 \right\rangle\right)\Big\}\Bigg)^2\Bigg] \\
&= \sum_{k=1}^{[MT]}\mathbf{E}^{P^{(M)}}\Big[\Big\{\left\langle h, (\mu_k - \mu_{k-1})^2 \right\rangle \\
&\qquad - M^{-1}\left(\langle \Upsilon h, \mu_{k-1}\rangle - \left\langle h, \mu_{k-1}^2 \right\rangle\right)\Big\}^2\Big] + O\left(M^{-2}\right) \\
&\leq 2\sum_{k=1}^{[MT]}\mathbf{E}^{P^{(M)}}\left[\left\langle h, (\mu_k - \mu_{k-1})^2 \right\rangle^2\right] + O\left(M^{-1}\right). \qquad (3.26)
\end{aligned}
$$

It will therefore suffice to show that

$$
\mathbf{E}^{P^{(M)}}\left[\left\langle h, (\mu_k - \mu_{k-1})^2 \right\rangle^2\right] = O\left(M^{-2}\right), \qquad (3.27)
$$

uniformly in $k \geq 1$.

Let $Y_1, \ldots, Y_M$ be, conditionally on μ_{k-1}, i.i.d. $\mu_{k-1}^* := (1-u)\mu_{k-1} + u\nu_0$, and let

$$H_{ij}^u := h(Y_i, Y_j) - \langle h(Y_i, \cdot), \mu_{k-1}\rangle - \langle h(\cdot, Y_j), \mu_{k-1}\rangle + \langle h, \mu_{k-1}^2\rangle. \quad (3.28)$$

Then

$$\begin{aligned}
&\mathbf{E}^{P^{(M)}}\left[\langle h, (\mu_k - \mu_{k-1})^2\rangle^2\right] \\
&= \mathbf{E}^{P^{(M)}}\left[\mathbf{E}\left[\left(\frac{1}{M^2}\sum_{i,j=1}^M H_{ij}^u\right)^2\right]\right] \\
&\leq 2\mathbf{E}^{P^{(M)}}\left[\mathbf{E}\left[\left(\frac{1}{M^2}\sum_{i=1}^M H_{ii}^u\right)^2\right]\right] + 2\mathbf{E}^{P^{(M)}}\left[\mathbf{E}\left[\left(\frac{1}{M^2}\sum_{i\neq j} H_{ij}^u\right)^2\right]\right] \\
&=: 2\sigma_1 + 2\sigma_2. \qquad (3.29)
\end{aligned}$$

Now $\sigma_1 = O(M^{-2})$, and we claim that $\sigma_2 = O(M^{-2})$ as well.

To understand the latter, consider first the case in which $u = 0$. (Actually, our assumptions rule out this possibility, so this is merely to clarify the argument.) Notice that

$$\begin{aligned}
&\mathbf{E}^{P^{(M)}}\left[\mathbf{E}\left[H_{12}^0\right]\right] \\
&= \mathbf{E}^{P^{(M)}}\big[\mathbf{E}\big[\{h(Y_1, Y_2) - \langle h(Y_1, \cdot), \mu_{k-1}\rangle \\
&\qquad - \langle h(\cdot, Y_2), \mu_{k-1}\rangle + \langle h, \mu_{k-1}^2\rangle\}\big]\big] \\
&= \mathbf{E}^{P^{(M)}}\left[\langle h, \mu_{k-1}^2\rangle - \langle h, \mu_{k-1}^2\rangle - \langle h, \mu_{k-1}^2\rangle + \langle h, \mu_{k-1}^2\rangle\right] \\
&= 0 \qquad (3.30)
\end{aligned}$$

and

$$
\begin{aligned}
&\mathbf{E}^{P^{(M)}}[\mathbf{E}[H_{12}^0 H_{13}^0]] \\
&= \mathbf{E}^{P^{(M)}}\left[\int_E \mathbf{E}\left[\left\{h(y_1,Y_2) - \langle h(y_1,\cdot),\mu_{k-1}\rangle \right.\right.\right. \\
&\qquad \left. - \langle h(\cdot,Y_2),\mu_{k-1}\rangle + \left\langle h,\mu_{k-1}^2\right\rangle\right\} \\
&\qquad \left\{h(y_1,Y_3) - \langle h(y_1,\cdot),\mu_{k-1}\rangle \right. \\
&\qquad \left.\left.\left. - \langle h(\cdot,Y_3),\mu_{k-1}\rangle + \left\langle h,\mu_{k-1}^2\right\rangle\right\}\right] \mu_{k-1}(dy_1)\right] \\
&= \mathbf{E}^{P^{(M)}}\left[\int_E \mathbf{E}\left[\left\{h(y_1,Y_2) - \langle h(y_1,\cdot),\mu_{k-1}\rangle \right.\right.\right. \\
&\qquad \left.\left. - \langle h(\cdot,Y_2),\mu_{k-1}\rangle + \left\langle h,\mu_{k-1}^2\right\rangle\right\}\right] \\
&\qquad \mathbf{E}\left[\left\{h(y_1,Y_3) - \langle h(y_1,\cdot),\mu_{k-1}\rangle \right.\right. \\
&\qquad \left.\left.\left. -\langle h(\cdot,Y_3),\mu_{k-1}\rangle + \left\langle h,\mu_{k-1}^2\right\rangle\right\}\right] \mu_{k-1}(dy_1)\right] \\
&= 0, \qquad (3.31)
\end{aligned}
$$

since both of the inner expectations in the last integral are 0. (Note the similarity between this argument and the proof of Lemma 2.2.) In words, H_{12}^0 and H_{13}^0, although clearly *not* independent, are uncorrelated, mean 0 random variables. This fact allows us to conclude that

$$
\begin{aligned}
\mathbf{E}^{P^{(M)}}\left[\mathbf{E}\left[\left(\frac{1}{M^2}\sum_{i\neq j} H_{ij}^0\right)^2\right]\right] &= \mathbf{E}^{P^{(M)}}\left[\mathbf{E}\left[\frac{1}{M^4}\sum_{i\neq j}(H_{ij}^0)^2\right]\right] \\
&= O\left(M^{-2}\right), \qquad (3.32)
\end{aligned}
$$

at least when $u = 0$. Now it remains to show that the same approach works when $u = \theta/(2M)$.

In general, (3.30)–(3.32) become

$$\begin{aligned}\mathbf{E}^{P^{(M)}}\left[\mathbf{E}\left[H^u_{12}\right]\right] &= \mathbf{E}^{P^{(M)}}\Big[\left\langle h,(\mu^*_{k-1})^2\right\rangle - \langle h,\mu^*_{k-1}\times\mu_{k-1}\rangle \\ &\qquad - \left\langle h,\mu_{k-1}\times\mu^*_{k-1}\right\rangle + \left\langle h,\mu^2_{k-1}\right\rangle\Big] \\ &= O\left(M^{-1}\right), \end{aligned} \tag{3.33}$$

$$\begin{aligned}&\mathbf{E}^{P^{(M)}}\left[\mathbf{E}\left[H^u_{12}H^u_{13}\right]\right] \\ &= \mathbf{E}^{P^{(M)}}\Bigg[\int_E \mathbf{E}\Bigg[\Big\{h(y_1,Y_2) - \langle h(y_1,\cdot),\mu_{k-1}\rangle \\ &\qquad - \langle h(\cdot,Y_2),\mu_{k-1}\rangle + \left\langle h,\mu^2_{k-1}\right\rangle\Big\} \\ &\qquad \Big\{h(y_1,Y_3) - \langle h(y_1,\cdot),\mu_{k-1}\rangle \\ &\qquad -\langle h(\cdot,Y_3),\mu_{k-1}\rangle + \left\langle h,\mu^2_{k-1}\right\rangle\Big\}\Bigg]\,\mu^*_{k-1}(dy_1)\Bigg] \\ &= \mathbf{E}^{P^{(M)}}\Bigg[\int_E \mathbf{E}\Bigg[\Big\{h(y_1,Y_2) - \langle h(y_1,\cdot),\mu_{k-1}\rangle \\ &\qquad -\langle h(\cdot,Y_2),\mu_{k-1}\rangle + \left\langle h,\mu^2_{k-1}\right\rangle\Big\}\Bigg] \\ &\qquad \mathbf{E}\Bigg[\Big\{h(y_1,Y_3) - \langle h(y_1,\cdot),\mu_{k-1}\rangle \\ &\qquad -\langle h(\cdot,Y_3),\mu_{k-1}\rangle + \left\langle h,\mu^2_{k-1}\right\rangle\Big\}\Bigg]\,\mu^*_{k-1}(dy_1)\Bigg] \\ &= \mathbf{E}^{P^{(M)}}\Bigg[\int_E \Big\{\left\langle h(y_1,\cdot),\mu^*_{k-1}\right\rangle - \langle h(y_1,\cdot),\mu_{k-1}\rangle \\ &\qquad -\left\langle h,\mu_{k-1}\times\mu^*_{k-1}\right\rangle + \left\langle h,\mu^2_{k-1}\right\rangle\Big\}^2\,\mu^*_{k-1}(dy_1)\Bigg] \\ &= O\left(M^{-2}\right), \end{aligned} \tag{3.34}$$

and

$$\begin{aligned}
\sigma_2 &:= \mathbf{E}^{P^{(M)}}\left[\mathbf{E}\left[\left(\frac{1}{M^2}\sum_{i\neq j} H_{ij}^u\right)^2\right]\right] \\
&= \mathbf{E}^{P^{(M)}}\left[\mathbf{E}\left[\frac{1}{M^4}\sum_{i\neq j}(H_{ij}^u)^2\right]\right] + O\left(M^{-2}\right) \\
&= O\left(M^{-2}\right), \qquad (3.35)
\end{aligned}$$

as required.

We have verified the statement of the lemma with

$$\begin{aligned}
F_M &:= \exp\{S_{13}' + \tfrac{1}{2}\Sigma_1 - \tfrac{1}{2}\Sigma_2'' - S_2'' + S_4''\}, \\
G_M &:= \exp\{-\tfrac{1}{2}\Sigma_2' - S_2' + S_4'\}, \qquad (3.36)
\end{aligned}$$

$F := R_T$, and $G := 1$. This completes the proof. ■

For each $\mu \in \mathcal{P}_M(E)$, let $Q_\mu^{(M)} \in \mathcal{P}(\Xi_M)$ denote the distribution of the selective Wright–Fisher model starting at μ, and for each $\mu \in \mathcal{P}^\circ(E)$, let $Q_\mu \in \mathcal{P}(\Omega^\circ)$ denote the distribution of the selective Fleming–Viot process starting at μ.

We can now state the main result of this section.

Theorem 3.1 *Assume that h is continuous. Let $\{\mu^{(M)}\} \subset \mathcal{P}_M(E) \subset \mathcal{P}^\circ(E)$ and $\mu \in \mathcal{P}^\circ(E)$ satisfy $d^\circ(\mu^{(M)}, \mu) \to 0$. For simplicity of notation, denote $Q_{\mu^{(M)}}^{(M)}$ by just $Q^{(M)}$. Then $Q^{(M)}\Phi_M^{-1} \Rightarrow Q_\mu$ on Ω°.*

Proof The proof is similar to that of Theorem 3.7 of [1], except that we use Lemmas 3.1 and 3.2 above in place of Lemmas 3.4 and 3.5 of [1]. ■

4. Characterization of the stationary distribution

If h is bounded, then it is known that the Fleming–Viot process in $\mathcal{P}(E)$ with generator $\mathcal{L}_h$ has a unique stationary distribution $\Pi_h \in \mathcal{P}(\mathcal{P}(E))$, is strongly ergodic, and is reversible. In fact,

$$\Pi_0(\cdot) = \mathbf{P}\left\{\sum_{i=1}^{\infty} \rho_i \delta_{\xi_i} \in \cdot\right\}, \qquad (4.1)$$

where $\xi_1, \xi_2, \ldots$ are i.i.d. ν_0 and $(\rho_1, \rho_2, \ldots)$ is Poisson–Dirichlet with parameter θ and independent of $\xi_1, \xi_2, \ldots$. Furthermore,

$$\Pi_h(d\mu) = e^{\langle h, \mu^2 \rangle} \Pi_0(d\mu) \Big/ \int_{\mathcal{P}(E)} e^{\langle h, \nu^2 \rangle} \Pi_0(d\nu). \tag{4.2}$$

For h satisfying (1.3), the finiteness of the normalizing constant in (4.2) is precisely the condition needed in the work of Overbeck *et al.* [2]. Notice that

$$\begin{aligned} \int_{\mathcal{P}(E)} e^{\langle h, \nu^2 \rangle} \Pi_0(d\nu) &= \mathbf{E}\left[\exp\left\{\sum_{i,j=1}^{\infty} \rho_i \rho_j h(\xi_i, \xi_j)\right\}\right] \\ &\leq \mathbf{E}\left[\exp\left\{2\sum_{i=1}^{\infty} \rho_i h_0(\xi_i)\right\}\right] \\ &= \mathbf{E}\left[\prod_{i=1}^{\infty} \langle e^{2\rho_i h_0}, \nu_0 \rangle\right]. \end{aligned} \tag{4.3}$$

A sufficient condition for this to be finite is $\langle e^{2h_0}, \nu_0 \rangle < \infty$.

Here we impose a slightly stronger condition: E, ν_0, and h are arbitrary, subject to the condition that there exist a continuous function $h_0 : E \mapsto [0, \infty)$ and a constant $\rho_0 \in (2, \infty]$ such that (1.3) holds and $\langle e^{\rho h_0}, \nu_0 \rangle < \infty$ whenever $0 < \rho < \rho_0$. In other words, we now require $\rho_0 > 2$.

Theorem 4.1 *Under the above conditions, Π_h, defined by (4.2), is a reversible stationary distribution for the Fleming–Viot process with generator $\mathcal{L}_h$, and it is the unique stationary distribution for this process.*

Proof Define $\tilde{h}_1$ and $\tilde{h}_2$ on $(E \times E)^2$ not by (4.8) of [1] but by

$$\tilde{h}_i((x_1, x_2), (y_1, y_2)) = h(x_i, y_i). \tag{4.4}$$

With additional minor changes to equations (4.5), (4.6), and (4.13) of [1], the proof is otherwise the same as that of Theorem 4.2 of [1]. ■

References

[1] S.N. Ethier and T. Shiga (2000). A Fleming–Viot process with unbounded selection. *J. Math. Kyoto Univ.*, **40**, 337–361.

[2] L. Overbeck, M. Röckner and B. Schmuland (1995). An analytic approach to Fleming–Viot processes with interactive selection. *Ann. Probab.*, **23**, 1–36.

[3] H. Tachida (1991). A study on a nearly neutral mutation model in finite populations. *Genetics*, **128**, 183–192.

Chapter 18

BOUNDARY THEORY FOR SUPERDIFFUSIONS

Kuznetsov, S.E. *

Department of Mathematics,
University of Colorado at Boulder,
Campus Box 395, Boulder, CO 80309-0395, USA
sek@euclid.colorado.edu

Keywords: Positive solutions, moderate solutions, σ-moderate solutions, boundary singularities, fine topology, fine trace, diffusion, conditional diffusion, superdiffusion.

AMS Classification: Primary 35J67, 35J75, Secondary 60J50, 60J60, 60J85, 60H30

1. Introduction

Connections between diffusion processes and linear PDE involving second order uniformly elliptic operators L are known for a long time. Superdiffusions are related, in a similar way, to equations involving semilinear differential operators $Lu - \psi(u)$.

Positive solutions to a linear equation $Lu = 0$ in a bounded smooth domain $D \subset \mathbb{R}^d$, that is positive L-harmonic functions, can be represented as a Poisson integral. Corresponding formula establishes 1-1 correspondence between positive L-harmonic functions and finite measures on the boundary ∂D. The measure ν that corresponds to a function h is, in some sense, a weak boundary value of the function h. It is natural to call ν a trace of the function h on the boundary. If the measure ν has a density with respect to the surface measure on ∂D, then the corresponding function h admits a probabilistic representation in terms of L-diffusion.

*Partially supported by National Science Foundation Grant DMS-9971009

Z. Hou et al. (eds.), Markov Processes and Controlled Markov Chains, 323–340.

Analogous theory of a nonlinear equation $Lu = \psi(u)$ was developed independently by analysts, including Keller, Osserman, Loewner and Nirenberg, Brezis, Marcus and Véron, Baras and Pierre, and by probabilists, including Watanabe, Dawson, Perkins, Le Gall, Dynkin and others. However, the boundary behavior of solutions of a nonlinear equation can be more complicate. In particular, a solution may blow up on a substantial part of a boundary, even on the whole boundary. In 1993, Le Gall [12] found a characterization of all positive solutions of equation $\Delta u = u^2$ in the unit disk D by using the Brownian snake — a path-valued process introduced by him in an earlier work. To describe a boundary behavior of a solution of a nonlinear equation, it is necessary to split the boundary of the domain into two parts: a closed subset Γ which is a set of "significant" explosions, and its complement where the weak boundary value exists as a Radon measure. Le Gall established a 1-1 correspondence between the solutions and pairs (Γ, ν) such that Γ is a closed subset of ∂D and ν is a Radon measure on $\partial D \setminus \Gamma$. Moreover, every solution admits a probabilistic representation in terms of the Brownian snake. In [15], the results were extended to all smooth domains in $\mathbb{R}^2$. The pair (Γ, ν) that corresponds to a solution u is called the trace of u.

Numerous attempts to find a proper generalization of this fundamental result (Marcus and Véron, Dynkin and Kuznetsov) brought a partial success. Namely, similar result is valid for the equation $\Delta u = u^\alpha$ in a ball if the dimension d of the space satisfies the condition $d < \frac{\alpha+1}{\alpha-1}$ (so called subcritical case). The analytical part of this statement was done by Marcus and Véron [16], [17] and the probabilistic representation was established by Dynkin and Kuznetsov [8] (the probabilistic part is valid in a more general setting). If $d \geq \frac{\alpha+1}{\alpha-1}$ (the supercritical case), the situation becomes more delicate because of a new phenomena — polar sets on the boundary. For this reason, not every finite measure ν may serve as a boundary value (Gmira and Véron [9]). Moreover, the example by Le Gall [14] shows that the definition of the trace based on the Euclidean topology is not sufficient to describe all solutions (there exist different solutions with the same traces).

We present here a new approach to the problem. A breakthrough was made possible after we have replaced the Euclidean topology on the boundary by another one. Most of the results presented here were obtained in joint publications by Dynkin and Kuznetsov.

2. Diffusions and linear equations

Let L be a second order linear uniformly elliptic operator with smooth coefficients and no zero order term. Denote by (ξ_t, Π_x) the corresponding L-diffusion in $\mathbb{R}^d$. Let $D \subset \mathbb{R}^d$ be a bounded smooth domain and let $\phi \geq 0$ be a continuous function on ∂D. The function

$$h(x) = \Pi_x \phi(\xi_\tau) \tag{2.1}$$

is a unique solution of the boundary value problem

$$\begin{aligned} Lh &= 0 \quad \text{in } D, \\ h &= \phi \quad \text{on } \partial D. \end{aligned} \tag{2.2}$$

Here τ stands for the first exit time from D.

At the same time,

$$h(x) = \int_{\partial D} k(x,y)\phi(y)\sigma(dy) \tag{2.3}$$

where $k(x,y)$ is the Poisson kernel for L in D and $\sigma(dy)$ is the surface measure.

The analytic representation (2.3) can be extended to an arbitrary positive L-harmonic function h. Namely, to every positive L-harmonic function there corresponds a finite measure ν on the boundary such that

$$h(x) = \int_{\partial D} k(x,y)\nu(dy) \tag{2.4}$$

The equation (2.4) establishes a 1-1 correspondence between finite measures on the boundary and positive L-harmonic functions. However, a probabilistic formula (2.1) is possible only if the measure ν is absolutely continuous with respect to the surface measure.

3. Superdiffusion and the nonlinear equation

An (L,α)-superdiffusion is a measure-valued Markov processes (X_D, P_μ) related to the nonlinear operator $Lu - u^\alpha$, where $\alpha \in (1,2]$ is a parameter. Here a family of measures X_D is indexed by open subsets of $\mathbb{R}^d$ and a measure X_D characterizes the accumulation of mass on ∂D if all the particles are instantly frozen at the first exit from D. The measure P_μ stands for the corresponding probability distribution if the movement starts from initial mass distribution μ. As usual, we write P_x if the corresponding μ is a unit mass concentrated at x. A detailed discussion of the concept of (L,α)-superdiffusion could be found in a paper of Dynkin in this volume. A principal relation between the process

(X_D, P_μ) and the nonlinear operator $Lu - u^\alpha$ can be stated as follows. If $\phi(x)$ is a positive continuous function on the boundary of a bounded smooth domain, then the function

$$u(x) = -\log P_x \exp(-\langle \phi, X_D \rangle) \tag{3.1}$$

is a unique solution of the boundary value problem

$$\begin{aligned} Lu &= u^\alpha \quad \text{in } D, \\ u &= \phi \quad \text{on } \partial D. \end{aligned} \tag{3.2}$$

(Note that an analytic substitute for (3.1) is not an explicit formula, but an integral equation.)

4. Range of superdiffusion and polar sets on the boundary

The range $\mathcal{R}_D$ of a superdiffusion in D is a minimal closed set which supports all measures $X_{D'}$, $D' \subset D$. A compact set $\Gamma \subset \partial D$ is said to be *polar* for (L, α)-superdiffusion if $P_\mu(\mathcal{R}_D \cap \Gamma = \emptyset) = 1$ for all μ such that $\operatorname{supp} \mu$ is disjoint from ∂D.

Characterization of polar sets was given by Le Gall [13] in case $\alpha = 2$ and by Dynkin and Kuznetsov [4] for general $1 < \alpha \leq 2$. It was shown that the class of polar sets coincides with the class of removable boundary singularities for the equation (5.1), and also with the class of sets of capacity 0. In particular, we have

Theorem 4.1 (Dynkin and Kuznetsov [4]) *A closed set* $\Gamma \subset \partial D$ *is polar if and only if*

$$\int_D \left(\int_\Gamma K(x,y)\nu(dy) \right)^\alpha \rho(x)\, dx = \infty \tag{4.1}$$

for every non-trivial measure ν *concentrated on* Γ. *Here* $\rho(x)$ *stands for the distance to the boundary* ∂D.

5. Moderate solutions

Our goal is to describe all solutions to the equation

$$Lu = u^\alpha \tag{5.1}$$

in a bounded smooth domain D. The equation (5.1) was studied by analysts for decades. Keller [10] and Osserman [19] proved that there exists no non-trivial entire solution to (5.1) in the whole space. On the other hand, if D is a bounded smooth domain, then there exists a maximal solution to (5.1) which dominates all other solutions and blows up at the whole boundary. Dynkin [1, 2] proved that every solution to (5.1) can be uniquely represented as

$$u(x) = -\log P_x e^{-Z_u} \tag{5.2}$$

where $Z_u = \lim\langle u, X_{D_n}\rangle$ is the so called *stochastic boundary value* of u. (Here D_n is an arbitrary increasing sequence of smooth domains such that $\bar{D}_n \subset D$ and $\cup D_n = D$.)

We begin with a subclass of solutions of (5.1) which we call *moderate* solutions. Namely, a solution u is moderate if it is dominated by an L-harmonic function. For every moderate solution u, there exists a minimal L-harmonic function h such that $h \geq u$. We call it the *minimal harmonic majorant* of u. The solution u can be recovered from its minimal harmonic majorant as a maximal solution dominated by h. Let ν be the measure corresponding to h by the formula (2.1). We call ν the *trace* of the moderate solution $u = u_\nu$.

Theorem 5.1 ([13, 3, 5, 18]) *A measure ν is the trace of a moderate solution if and only if $\nu(\Gamma) = 0$ for all polar sets $\Gamma \subset \partial D$.*

Let now ν be a σ-finite measure on the boundary such that $\nu(\Gamma) = 0$ for all polar sets and let ν_n be an increasing sequence of finite measures with the limit ν. Formula $u_\nu = \lim u_{\nu_n}$ defines a solution of (5.1). It could be shown that u_ν does not depend on the choice of approximating sequence ν_n (however, the same solution may correspond to different ν). We denote by Z_ν the stochastic boundary value of u_ν.

6. Subcritical case

If $d < \frac{\alpha+1}{\alpha-1}$, then there is no non-trivial polar sets on the boundary. The complete characterization of all solutions to the equation (5.1) was first obtained by Le Gall [12, 15] in case of $\alpha = 2,\ d = 2$ and by Marcus and Véron [16, 17] for general $\alpha > 1$ (and D being a ball in $\mathbb{R}^d$). By combining their results with those in [8], we get

Theorem 6.1 ([15, 17, 8]) *Suppose D is a ball in $\mathbb{R}^d$. Formula*

$$u(x) = -\log P_x e^{-Z_\nu} 1_{\mathcal{R}_D \cap \Gamma = \emptyset} \tag{6.1}$$

establishes a 1-1 correspondence between the class of all solutions to (5.1) and all pairs (Γ, ν) such that Γ is a closed subset of ∂D and ν is a Radon measure on $\partial D \setminus \Gamma$. The pair (Γ, ν) is called the trace of u.

Remark 6.1 *In case $\alpha = 2,\ d = 2$, Le Gall proved that the exit measure X_D has a.s. continuous density with respect to the surface measure $\sigma(dx)$, and that $Z_\nu = \langle \frac{dX_D}{d\sigma}, \nu\rangle$ a.s.*

7. Supercritical case

The situation in the supercritical case is more complicated. Not every measure ν may serve as a component of a trace (cf. Theorem 5.1). Also, if Γ_1 and Γ_2 differ by a polar set, then the events $\{\Gamma_1 \cap \mathcal{R}_D = \emptyset\}$ and $\{\Gamma_2 \cap \mathcal{R}_D = \emptyset\}$ coincide a.s. and the solutions defined by (6.1) are equal to each other. These difficulties were taken into account in the definition of a trace given independently by Marcus, Véron [16], [18] and by Dynkin and Kuznetsov [8, 7]. All possible traces have been characterized. It was shown that the formula (6.1) gives a maximal solution with the given trace. However, an example given by Le Gall [14] shows that it is not possible to represent all the solutions of (5.1) in the form (6.1) if we restrict ourselves by closed sets Γ. In particular, different solutions may have the same trace.

8. σ-moderate solutions, singular points and fine topology

A new approach to the problem was suggested in [11, 6]. A solution u is called *σ-moderate* if there exists an increasing sequence of moderate solutions u_n such that $u_n \uparrow u$. For every Borel subset $B \subset \partial D$, we define u_B as a supremum of all moderate solutions u_ν with ν concentrated on B. It could be shown (see [11, 6]) that u_B is a σ-moderate solution of (5.1).

Let $y \in \partial D$. Denote by (ξ_t, Π_x^y) an L-diffusion conditioned to exit from D at y. Let ζ be the corresponding exit time. We call the point y a *singular point* for a solution u (cf. [2]) if

$$\int_0^\zeta u^{\alpha-1}(\xi_s)\, ds = \infty \qquad \Pi_x^y\text{-a.s.} \tag{8.1}$$

for some $x \in D$. We denote by $\mathrm{SG}(u)$ the set of all singular points of the solution u.

We define *finely closed* sets as sets $\Gamma \subset \partial D$ with the property $\mathrm{SG}(u_\Gamma) \subset \Gamma$ (cf. [11, 6]).

Finally, for every pair of solutions u, v, we define $u \oplus v$ as a maximal solution dominated by $u + v$ (it could be shown that $Z_{u\oplus v} = Z_u + Z_v$ a.s.).

9. Fine trace

Let u be a solution of (5.1). Denote $\Gamma = \mathrm{SG}(u)$. Next, consider the set of all moderate solutions u_μ such that $u_\mu \le u$ and $\mu(\Gamma) = 0$. Put

$\nu = \sup\{\mu : u_\mu \le u, \mu(\Gamma) = 0\}$. We call the pair (Γ, ν) the fine trace of u.

We prove:

Theorem 9.1 ([11, 6]) *The fine trace of every solution u has the following properties:*

1. *Γ is a Borel finely closed set.*
2. *ν is a σ-finite measure not charging polar sets and such that $\nu(\Gamma) = 0$ and $SG(u_\nu) \subset \Gamma$.*

Moreover,

$$u_{\Gamma,\nu} = u_\Gamma \oplus u_\nu \tag{9.1}$$

is the maximal σ-moderate solution dominated by u.

We say that pairs (Γ, ν) and (Γ', ν') are equivalent and we write $(\Gamma, \nu) \sim (\Gamma', \nu')$ if the symmetric difference between Γ and Γ' is polar and $\nu = \nu'$. Clearly, $u_\Gamma = u_{\Gamma'}$ and $u_{\Gamma,\nu} = u_{\Gamma',\nu'}$ if $(\Gamma, \nu) \sim (\Gamma', \nu')$.

Theorem 9.2 ([11, 6]) *Let (Γ, ν) satisfy* Conditions 1–2. *Then the fine trace of $u_{\Gamma,\nu}$ is equivalent to (Γ, ν). Moreover, $u_{\Gamma,\nu}$ is the minimal solution with this property and the only one which is σ-moderate.*

The existence of a non-σ-moderate solution remains an open question. If there is no such solutions, then Theorems 9.1 and 9.2 provide a complete answer to the problem. If such solutions exist, then we may have to refine the definition of the trace.

References

[1] E.B. Dynkin, *Superprocesses and partial differential equations*, Ann. Probab. **21** (1993), 1185–1262.

[2] E.B. Dynkin, *Stochastic boundary values and boundary singularities for solutions of the equation $Lu = u^\alpha$*, J. Functional Analysis **153** (1998), 147–186.

[3] E.B. Dynkin and S. E. Kuznetsov, *Linear additive functionals of superdiffusions and related nonlinear p.d.e.*, Trans. Amer. Math. Soc. **348** (1996), 1959–1987.

[4] E.B. Dynkin and S. E. Kuznetsov, *Superdiffusions and removable singularities for quasilinear partial differential equations*, Comm. Pure & Appl. Math **49** (1996), 125–176.

[5] E.B. Dynkin and S. E. Kuznetsov, *Nonlinear parabolic p.d.e. and additive functionals of superdiffusions*, Ann. Probab. **25** (1997), 662–701.

[6] E.B. Dynkin and S. E. Kuznetsov, *Fine topology and fine trace on the boundary associated with a class of quasilinear differential equations*, Comm. Pure Appl. Math. **51** (1998), 897–936.

[7] E.B. Dynkin and S. E. Kuznetsov, *Solutions of nonlinear differential equations on a Riemannian manifold and their trace on the Martin boundary*, Transact. Amer. Math. Soc. **350** (1998), 4521–4552.

[8] E.B. Dynkin and S. E. Kuznetsov, *Trace on the boundary for solutions of nonlinear differential equations*, Transact. Amer. Math. Soc. **350** (1998), 4499–4519.

[9] A. Gmira and L. Véron, *Boundary singularities of solutions of some nonlinear elliptic equations*, Duke Math.J. **64** (1991), 271–324.

[10] J.B. Keller, *On the solutions of* $\Delta u = f(u)$, Comm. Pure Appl. Math. **10** (1957), 503–510.

[11] S.E. Kuznetsov, *σ-moderate solutions of $Lu = u^\alpha$ and fine trace on the boundary*, C. R. Acad. Sci. Paris, Série I **326** (1998), 1189–1194.

[12] J.-F. Le Gall, *Solutions positives de $\Delta u = u^2$ dans le disque unité*, C.R. Acad. Sci. Paris, Série I **317** (1993), 873–878.

[13] J.-F. Le Gall, *The Brownian snake and solutions of $\Delta u = u^2$ in a domain*, Probab. Theory Relat. Fields **102** (1995), 393–402.

[14] J.-F. Le Gall, *A probabilistic approach to the trace at the boundary for solutions of a semilinear parabolic differential equation*, J. Appl.Math. Stochast. Analysis **9** (1996), 399–414.

[15] J.-F. Le Gall, *A probabilistic Poisson representation for positive solutions of $\Delta u = u^2$ in a planar domain*, Comm. Pure & Appl Math. (1997), 69–103.

[16] M. Marcus and L. Véron, *Trace au bord des solutions positives d'équations elliptiques non linéaires*, C.R. Acad.Sci Paris **321, ser I** (1995), 179–184.

[17] M. Marcus and L. Véron, *The boundary trace of positive solutions of semilinear elliptic equations, I: The subcritical case*, Arch. Rat. Mech. Anal. **144** (1998), 201–231.

[18] M. Marcus and L. Véron, *The boundary trace of positive solutions of semilinear elliptic equations: The supercritical case*, J. Math. Pures Appl. **77** (1998), 481–524.

[19] R. Osserman, *On the inequality* $\Delta u \geq f(u)$, Pacific J. Math, **7** (1957), 1641–1647.

Chapter 19

ON SOLUTIONS OF BACKWARD STOCHASTIC DIFFERENTIAL EQUATIONS WITH JUMPS AND STOCHASTIC CONTROL*

Situ Rong
Department of Mathematics,
Zhongshan University,
Guangzhou 510275, China

Abstract We relax conditions on coefficients given in [7] for the existence of solutions to backward stochastic differential equations (BSDE) with jumps. Counter examples are given to show that such conditions can not be weakened further in some sense. The existence of a solution for some continuous BSDE with coefficients $b(t, y, q)$ having a quadratic growth in q, having a greater than linear growth in y, and are unbounded in y belonging to a finite interval, is also obtained. Then we obtain an existence and uniqueness result for the Sobolev solution to some integro-differential equation (IDE) under weaker conditions. Some Markov properties for solutions to BSDEs associated with some forward SDEs are also discussed and a Feynman-Kac formula is also obtained. Finally, we obtain probably the first results on the existence of non-Lipschitzian optimal controls for some special stochastic control problems with respect to such BSDE systems with jumps, where some optimal control problem is also explained in the financial market.

1. Existence of solutions to BSDE with jumps under weaker conditions

Consider the following BSDE with jumps in R^d :

*This work is supported in part by The National Natural Science Foundation of China No. 79790130.

Z. Hou et al. (eds.), Markov Processes and Controlled Markov Chains, 331–340.

$$
\begin{aligned}
x_t = X &+ \int_{t\wedge\tau}^{\tau} b(s, x_s, q_s, p_s, \omega)\, ds - \int_{t\wedge\tau}^{\tau} q_s\, dw_s \\
&- \int_{t\wedge\tau}^{\tau}\int_Z p_s(z)\tilde{N}_k(ds, dz), \qquad 0 \le t, \qquad (1.1)
\end{aligned}
$$

where $w_t^T = (w_t^1, \dots, w_t^{d_1})$, $0 \le t$, is a d_1-dimensional standard Brownian motion (BM), w_t^T is the transpose of w_t; $k^T = (k_1, \dots, k_{d_2})$ is a d_2-dimensional stationary Poisson point process with independent components, $\tilde{N}_{k_i}(ds, dz)$ is the Poisson martingale measure generated by k_i satisfying

$$
\tilde{N}_{k_i}(ds, dz) = N_{k_i}(ds, dz) - \pi(dz)ds, \qquad i = 1, \cdots, d_2,
$$

where $\pi(.)$ is a σ-finite measure on a measurable space $(Z, \mathfrak{B}(Z))$, $N_{k_i}(ds, dz)$ is the Poisson counting measure generated by k_i, and τ is a bounded $\Im_t$ - stopping time, where $\Im_t$ is the $\sigma-$algebra generated (and completed) by $\{w_s, k_s, s \le t\}$. Let us assume that $0 \le \tau \le T_0$, where T_0 is a fixed number, and b in (1.1) is a R^d-valued function. It is known that the study of (1.1) is useful for the option pricing in the financial market [1]. For the precise definition of the solution to (1.1) we need the following notation:

$$
\begin{aligned}
S_{\Im}^2(R^d) &= \left\{ \begin{array}{c} f(t,\omega) : f(t,\omega) \text{ is } \Im_t - \text{adapted, } R^d - \text{valued} \\ \text{such that } E \sup_{t\in[0,\tau]} |f(t,\omega)|^2 < \infty \end{array} \right\}, \\
L_{\Im}^2(R^{d\otimes d_1}) &= \left\{ \begin{array}{c} f(t,\omega) : f(t,\omega) \text{ is } \Im_t - \text{adapted, } R^{d\otimes d_1} - \text{valued} \\ \text{such that } E\int_0^\tau |f(t,\omega)|^2\, dt < \infty \end{array} \right\}, \\
F_{\Im}^2(R^{d\otimes d_2}) &= \left\{ \begin{array}{c} f(t,z,\omega) : f(t,z,\omega) \text{ is } R^{d\otimes d_2} - \text{valued,} \\ \Im_t\text{-predictable such that} \\ E\int_0^\tau \int_Z |f(t,z,\omega)|^2\, \pi(dz)dt < \infty \end{array} \right\}.
\end{aligned}
$$

Definition 1.1 *(x_t, q_t, p_t) is said to be a solution of* (1.1), *if and only if $(x_t, q_t, p_t) \in S_{\Im}^2(R^d) \times L_{\Im}^2(R^{d\otimes d_1}) \times F_{\Im}^2(R^{d\otimes d_2})$, and it satisfies* (1.1).

Assumption 1.1 *For discussing the solution of* (1.1) *we make the following assumptions*

(i) $b : [0, T_0] \times R^d \times R^{d\otimes d_1} \times L_{\pi(\cdot)}^2(R^{d\otimes d_2}) \times \Omega \to R^d$ is jointly measurable and $\Im_t-$adapted, where

$$
L_{\pi(\cdot)}^2(R^{d\otimes d_2}) = \left\{ \begin{array}{c} f(z) : f(z) \text{ is } R^{d\otimes d_2} - \text{valued, and} \\ \|f\|^2 = \int_Z |f(z)|^2\, \pi(dz) < \infty \end{array} \right\};
$$

(ii) $N_{k_i}(\{t\},U)N_{k_j}(\{t\},U) = 0$, *as* $i \neq j$, $U \in \mathfrak{B}(Z)$ *with* $\pi(U) < \infty$.

We have

Theorem 1.1 *Assume that* $b = b_1 + b_2$, *and that*

(i) $b_i = b_i(t,x,q,p,\omega) : [0,T_0] \times R^d \times R^{d\otimes d_1} \times L^2_{\pi(\cdot)}(R^{d\otimes d_2}) \times \Omega \to R^d$, $i = 1,2$, *are* $\Im_t$*−adapted and measurable processes such that* $P-a.s.$

$$\begin{aligned} |b_1(t,x,q,p,\omega)| &\leq c_1(t)(1+|x|), \\ |b_2(t,x,q,p,\omega)| &\leq c_1(t)(1+|x|) + c_2(t)(1+|q|+\|p\|), \end{aligned}$$

where $c_1(t)$ *and* $c_2(t)$ *are non-negative and non-random such that*

$$\int_0^{T_0} c_1(t)\,dt + \int_0^{T_0} c_2(t)^2\,dt < \infty;$$

(ii) $(x_1 - x_2)\cdot(b_1(t,x_1,q_1,p_1,\omega) - b_1(t,x_2,q_2,p_2,\omega))$
$\leq c_1^N(t)\rho^N(|x_1-x_2|^2) + c_2^N(t)\,|x_1-x_2|\,(|q_1-q_2| + \|p_1-p_2\|)$, *and*

$$\begin{aligned} |b_1(t,x,q,p_1,\omega) - b_1(t,x,q,p_2,\omega)| &\leq c_2^N(t)\,\|p_1-p_2\|, \\ |b_2(t,x_1,q_1,p_1,\omega) - b_2(t,x_2,q_2,p_2,\omega)| & \\ \leq c_1^N(t)\,|x_1-x_2| + c_2^N(t)[|q_1-q_2| + \|p_1-p_2\|], & \end{aligned}$$

as $|x| \leq N$, $|x_i| \leq N$, $i = 1,2; N = 1,2\cdots$; *where for each* N, $c_1^N(t)$ *and* $c_2^N(t)$ *satisfy the same conditions as in (i); and for each* N, $\rho^N(u) \geq 0$, *as* $u \geq 0$, *is non-random, increasing, continuous and concave such that*

$$\int_{0+} du/\rho^N(u) = \infty;$$

(iii) $b_1(t,x,q,p,\omega)$ *is continuous in* (x,q,p);

(iv) $X \in \Im_T, E|X|^2 < \infty$.

Then (1.1) *has a unique solution.*

Here conditions in Theorem 1.1 are weaker than that in Theorem 1 of [7], where it assumes that $|X| \leq k_0$ and $\int_0^{T_0} |c_1(t)|^2\,dt < \infty$.

Let us give some counter examples and an example as follows:

Example 1.1 (Condition $\int_0^{T_0} c_1(t)dt < \infty$ can not be weakened) *Consider*

$$x_t = 1 + \int_t^T I_{s\neq 0} s^{-\alpha} x_s\, ds - \int_t^T q_s\, dw_s - \int_t^T \int_Z p_s(z)\tilde{N}_k(ds, dz), \quad 0 \le t \le T.$$

Obviously, if $\alpha < 1$, then by Theorem 1.1 it has a unique solution. However, if $\alpha \ge 1$, then it has no solution. Otherwise for the solution (x_t, q_t, p_t) one has

$$Ex_0^i = \infty, \qquad \forall i = 1, 2, \cdots, d, \ \text{as } \alpha \ge 1.$$

Example 1.2 (Condition $\int_0^{T_0} c_2(t)^2 dt < \infty$ cannot be weakened) *Now suppose that all processes appearing in BSDE* (1.1) *are real-valued. Let*

$$\begin{aligned} X &= \int_0^T I_{s\neq 0}(1+s)^{-1/2}(\log(1+s))^{-\alpha_2}\, dw_s \\ &\quad + \int_0^T \int_Z I_{s\neq 0}(1+s)^{-1/2}(\log(1+s))^{-\alpha_3} I_U(z)\widetilde{N}_k(ds, dz), \\ b &= I_{s\neq 0}(1+s)^{-1/2}(\log(1+s))^{-\alpha_1}(\widetilde{k}_1 |q| + \widetilde{k}_2 \int_Z |p(z)|\, I_U(z)\pi(dz)), \end{aligned}$$

where $\widetilde{k}_1, \widetilde{k}_2 \ge 0$ are constants, and we assume that $0 < \alpha_2$, $\alpha_3 < \frac{1}{2}$, and $0 < \pi(U) < \infty$. Obviously, if $0 < \alpha_1 < \frac{1}{2}$, then by Theorem 1.1, (1.1) *has a unique solution. However, if $\alpha_1 > \frac{1}{2}$, and $\widetilde{k}_1 > 0$, $\alpha_1 + \alpha_2 \ge 1$, or $\widetilde{k}_2 > 0$, $\alpha_1 + \alpha_3 \ge 1$, then* (1.1) *has no solution. Otherwise for solution (x_t, q_t, p_t)*

$$Ex_0 = \infty.$$

Example 1.3 *Let*

$$b = -I_{s\neq 0} s^{-\alpha_1} x\, |x|^{-\beta} + I_{s\neq 0} s^{-\alpha_2} q + I_{s\neq 0} s^{-\alpha_2} \int_Z p(z) I_U(z)\pi(dz),$$

where $\alpha_1 < 1$; $\alpha_2 < 1/2; 0 < \beta < 1$; $\pi(U) < \infty$; and assume that $X \in \Im_T$, $E|X|^2 < \infty$.

Obviously, by Theorem 1.1, (1.1) *has a unique solution. However, $c_1(s) = I_{s\neq 0} s^{-\alpha_1}$, $c_2(s) = I_{s\neq 0} s^{-\alpha_2}$ are unbounded in s, and $b_1 = -I_{s\neq 0} s^{-\alpha_1} x\, |x|^{-\beta}$ is also unbounded in s and x, and is non-Lipschitzian continuous in x. Note that here we have not assumed that X is bounded. (cf* [7]*).*

Theorem 1.1 can be shown by the approximation technique.

2. Existence of solutions to BSDE with coefficient having a greater than linear growth

Definition 2.1 (x_t, q_t, p_t) *is said to be a generalized solution of* (1.1), *if and only if* x_t *and* q_t *are* $\Im_t$*−adapted,* p_t *is* $\Im_t$*−predictable, and they satisfy* (1.1).

Consider BSDE without jumps in (1.1) with $d = 1$ as follows:

$$\begin{cases} dy_t &= \left(\widetilde{b}(t, y_t, \widetilde{q}_t, \omega)y_t^2 + 2I_{y_t\neq 0}\left|\widetilde{q}_t\right|^2/y_t\right)dt + \widetilde{q}_t dw_t, \\ y_\tau &= Y, \qquad 0 \le t \le \tau. \end{cases} \tag{2.1}$$

Denote $b(t, x, q, \omega) = \widetilde{b}(t, \frac{1}{x}, -\frac{q}{x^2}, \omega)$, where $b(t, 0, q, \omega) = \lim_{x\to 0} b(t, x, q, \omega)$ is assumed to exist and finite. We have the following theorem and example:

Theorem 2.1 *If* b *satisfies* Assumption (i)–(iv) *in* Theorem 1.1 *except that condition for* b_1 *in* (ii) *is cancelled and in* (i) *is weakened to be the same as* b_2*, and condition for* b_2 *in* (ii) *is strengthened to be that all* $c_1^N(t)$ *and* $c_1^N(t)$ *are the same for all* N*, moreover, if* $b^1(t, x, q, \omega) \ge 0$, $b^i(t, 0, 0, \omega) \ge 0$, $i = 1, 2$; *and* $X = 1/Y > 0$*, then BSDE* (2.1) *has a generalized solution. Furthermore, if* $X \ge r_0 > 0$*, where* r_0 *is a constant, then BSDE* (2.1) *has a solution* $(y_t, \widetilde{q}_t) \in S^2_{\Im}(R^1) \times L^2_{\Im}(R^{1\otimes d_1})$ *such that* $\frac{1}{\delta_0} \ge y_t > 0$, $\forall t \in [0, \tau]$*, where* δ_0 *is a constant, which exists.*

Example 2.1

$$\begin{aligned} &\widetilde{b}(t, y, \widetilde{q}, \omega)y^2 + 2I_{y\neq 0}\left|\widetilde{q}\right|^2/y \\ = \;& I_{s\neq 0}s^{-\alpha_1}\left|y\right|^{1+\beta} + I_{s\neq 0}s^{-\alpha_2}\left|\widetilde{q}\right|^{1-\beta_1}\left|y\right|^{2\beta_1} + 2I_{y\neq 0}\left|\widetilde{q}\right|^2/y \\ & + c_1(s)y - c_2(s)\widetilde{q} \end{aligned}$$

will satisfy all conditions in Theorem 2.1, *if in above* $\alpha_1 < 1$, $\alpha_2 < \frac{1}{2}$, $0 < \beta$, $\beta_1 < 1$; $c_1(t)$, $c_2(t)$ *satisfy condition in* (i) *of* Theorem 1.1. *However, such coefficient has a greater than linear growth in* y*, is unbounded in* y *belonging to any finite interval* $(-\varepsilon, \varepsilon)$*, and has a quadratic growth in* $\widetilde{q}$.

3. Application to integro-differential equations. Some Markov properties of solutions to some BSDEs and a Feynman-Kac formula

Applying Theorem 1.1 we can obtain an existence and uniqueness result on the Sobolev solution to some IDE under weaker conditions. Such IDE is useful in the stochastic optimal control problem [2].

Suppose that $D \subset R^d$ is a bounded open region, ∂D is its boundary, and denote $D^c = R^d - D$. Consider the following IDE

$$\begin{aligned}
&\pounds_{b,\sigma,c} u(t,x) \\
&\equiv \left(\frac{\partial}{\partial t} + \sum_{i=1}^{d} b_i(t,x) \frac{\partial}{\partial x_i} + \frac{1}{2} \sum_{i,j=1}^{d} a_{ij}(t,x) \frac{\partial^2}{\partial x_i \partial x_j} \right) u(t,x) \\
&\quad + \int_Z \left(u(t,x+c(t,x,z)) - u(t,x) - \sum_{i=1}^{d} c_i(t,x,z) \frac{\partial u(t,x)}{\partial x_i} \right) \pi(dz) \\
&= f(t,x,u(t,x),u'_x(t,x)\cdot\sigma(t,x),u(t,x+c(t,x,.))-u(t,x)), \qquad (3.1)
\end{aligned}$$

$$\begin{aligned}
u(T,x) = \phi(x), \quad u(t,x)|_{D^c} = \psi(t,x), \quad \psi(T,x) = \phi(x)|_{D^c}, \qquad (3.2) \\
u \in W_p^{1,2}([0,T]\times R^d)(= W_p^{1,2}([0,T]\times R^d; R^m)).
\end{aligned}$$

Also consider FSDE and BSDE as follows:
for any given $(t,x) \in [0,T] \times D$

$$\begin{aligned}
y_s &= x + \int_t^s b(r,y_r)\,dr + \int_t^s \sigma(r,y_r)\,dw_r \\
&\quad + \int_t^s \int_Z c(r,y_{r-},z)\tilde{N}_k(dr,dz), \qquad \text{as } t \le s \le T; \\
x_s &= I_{\tau<T}\psi(\tau,y_\tau) + I_{\tau=T}\varphi(y_\tau) - \int_{s\wedge\tau}^{\tau} f(r,y_r,x_r,q_r,p_r)dr \\
&\quad - \int_{s\wedge\tau}^{\tau} q_r dw_r - \int_{s\wedge\tau}^{\tau}\int_Z p_r(z)\tilde{N}_k(dr,dz), \qquad \text{as } t \le s \le T,
\end{aligned}$$

where $\tau = \tau_x = \inf\{s > t : y_s \notin D\}$, and $\tau = \tau_x = T$, for $\inf\{\phi\}$.

Assumption 3.1 Assumptions A.1–A.3, B.1', B.2–B.4 *and* (A)' *in* Section 4 *of* [7] *hold.*

We have the following theorem, which implies Theorem 10 in [7].

Theorem 3.1 *Suppose that* Assumption 3.1 *holds except that the condition* $|f(t,x,r,q,p)| \le k_0$ *in* B.1' *is weakened to be that*

$$|f(t,x,r,q,p)| \le k_0(1+|r|).$$

Then (3.1) *and* (3.2) *has a unique solution* $u(t,x) \in W_p^{1,2}([0,T]\times D)$ *such that*

$$\begin{aligned}
&\|u\|_{W_p^{1,2}([0,T]\times R^d)} \\
&\le c_0 \left(\|g^u\|_{L_p([0,T]\times D)} + \|\psi\|_{W_p^{1,2}([0,T]\times D^c)} + \|\phi\|_{W_p^{2(1-1/p)}(R^d)} \right),
\end{aligned}$$

where $c_0 \geq 0$ is a universal constant depending on T, the domain D, the dimensions d and m only. Moreover, one has that

$$\begin{aligned} x_s &= u(s, y_s), \\ q_s &= \sigma(s, y_s)\partial_x u(s, y_s), \\ p_s(\cdot) &= u(s, y_{s-} + c(s, y_{s-}, \cdot)) - u(s, y_{s-}), \end{aligned}$$

where y_s is the unique strong solution of the above FSDE, which is a Markov process, and

$$(x_s, q_s, p_s) \in S^2_{\Im}([t,\tau]; R^m) \times L^2_{\Im}([t,\tau]; R^{m\otimes d_1}) \times F^2_{\Im}([t,\tau]; R^{m\otimes d_2})$$

is the unique solution of the above BSDE. Hence we can say that (x_s, q_s, p_s) has a Markov property. Furthermore, we have a Feynman-Kac formula $u(t,x) = x_s|_{s=t}$.

Theorem 3.1 can be shown by the approximation technique, by using Itô's formula and Theorem 1.1.

4. Application to optimal stochastic control

In this section we obtain probably the first results on the existence of some non-Lipschitzian optimal controls for some stochastic control problems with respect to some BSDE systems with jumps.

Consider the following d-dimensional BSDE system: for $0 \leq t \leq T$,

$$x_t^u = X + \int_t^T u(s, x_s^u, q_s^u, p_s^u)\, ds - \int_t^T q_s^u\, dw_s - \int_t^T \int_Z p_s^u(z) \tilde{N}_k(ds, dz), \tag{4.1}$$

where $x_s^u, u(s, x_s^u, q_s^u, p_s^u) \in R^d$, $q_s^u \in R^{d\otimes d_1}$, $p_s^u(z) \in R^{d\otimes d_2}$, and $u \in U$,

$$U = \left\{ \begin{array}{l} u = u(t,x,q,p) : u(t,x,q,p) \text{ is jointly measurable} \\ \text{such that (4.1) has a unique solution } (x_t, q_t, p_t), \text{ and} \\ |u(t,x,q,p)| \leq |x|^\beta \end{array} \right\},$$

where $0 < \beta \leq 1$ is a given fixed constant. The following Theorem shows that a non-Lipschitzian feedback optimal stochastic control exists.

Theorem 4.1 *Define $u^0(x) = u^0(t,x,q,p) = -I_{x\neq 0} x / |x|^{1-\beta}$ and let*

$$J(u) = E\left(\frac{1}{2}|x_0^u|^2 + \frac{1}{2}\int_0^T (|q_s^u|^2 + \int_Z |p_s^u|^2 \pi(dz))ds + \int_0^T |x_s^u|^{1+\beta}\, ds\right),$$

where (x_t^u, q_t^u, p_t^u) is the unique solution of (4.1) for $u \in U$. Then *(i)* $u^0 \in U$,

(ii) $J(u) \geq J(u^0)$, *for all* $u \in U$.

The above target functional $J(u)$ can be explained as an energy functional. Now consider the BSDE system (4.1) with $x_t^u, u(s, x_s^u, q_s^u, p_s^u) \in R^1$, $q_s^u \in R^{1 \otimes d_1}$, $p_s^u(z) \in R^{1 \otimes d_2}$, and consider the admissible control set as

$$U = \left\{ \begin{array}{l} u = u(t,x,q,p) : u(t,x,q,p) \text{ is jointly measurable} \\ \text{such that (4.1) has at least a solution } (x_t, q_t, p_t), \text{ and} \\ |u(t,x,q,p)| \leq |x|^\beta \end{array} \right\}$$

where $0 < \beta \leq 1$ is a given fixed constant. Denote the target functional as

$$\begin{aligned} J_1(u) &= Max\left\{ E(2\int_0^T |x_s^u|^{1+\beta}\, ds - |x_0^u|^2 \right. \\ &\quad \left. - \int_0^T \left(|q_s^u|^2 + \int_Z |p_s^u|^2\, \pi(dz) \right) ds) \right\} \\ &= Max\left\{ E\left[\left(\int_0^T |x_s^u|^{1+\beta}\, ds \right. \right. \right. \\ &\quad \left. - \int_0^T \left(|q_s^u|^2 + \int_Z |p_s^u|^2\, \pi(dz) \right) ds \right) \\ &\quad \left. \left. + \left(\int_0^T |x_s^u|^{1+\beta}\, ds - |x_0^u|^2 \right) \right] \right\} \end{aligned}$$

for each $u \in U$,, where (x_t^u, q_t^u, p_t^u) is any solution corresponding to the same u. We have the following

Theorem 4.2 *Denote* $u^0(x) = I_{x \neq 0} x / |x|^{1-\beta}$. *Then*

(i) $u^0 \in U$,

(ii) $J_1(u) \leq J_1(u^0)$, $\forall u \in U$.

Both above theorems can be shown by using the Hamilton-Jacobi-Bellman equations.

Now let us explain Theorem 4.2 in the financial market. If we regard (4.1) as the equation for the wealth process x_t of a small investor, explain the control $u(t, x_t, q_t, p_t)$ as his feedback generalized consumption process, and $(q_t, p_t(\cdot))$ as his some generalized portfolio process for the stocks, then in case $\beta = 1$, we can illustrate the target functional $J_1(u)$ as a subtraction of the total summation of the square of wealths and the square of money of bonds on the whole time interval from the square of initial wealth (or say initial invest) for the investor. So $J_1(u)$ can be seen as some generalized utility functional for him. Theorem 4.2 tells the investor that he can get a maximum utility, if he chooses the consumption law as $u(x_t) = I_{x_t \neq 0} x_t / |x_t|^{1-\beta}$, i.e. when the wealth process $x_t \geq 0$, he should consume the money $I_{x_t \neq 0} x_t / |x_t|^{1-\beta}$; and when $x_t < 0$, he should borrow the money $-x_t / |x_t|^{1-\beta}$.

References

[1] Bardhan, I. and Chao, X. (1993). Pricing options on securities with discontinuous returns. *Stochastic Process. Appl.* 48, 123–137.

[2] Bensoussan, A. and J.L. Lions. (1984). *Impulse Control and Quasi-Variational Inequalities.* Gautheir-Villars.

[3] El Karoui, N., Peng, S., and Quenez, M.C. (1997). Backward stochastic differential equations in finance. *Math. Finance* 7, 1–71.

[4] Ladyzenskaja, O. A., Solonnikov, V. A. and Uralceva, N. N. (1968). *Linear and Quasilinear Equations of Parabolic Type.* Translation of Monographs 23, AMS Providence, Rode Island.

[5] Peng, S. (1993). Backward stochastic differential equations and applications to optimal control. *Appl. Math.Optim.* 27, 125–144.

[6] Situ Rong (1996). On comparison theorem of solutions to backward stochastic differential equations with jumps and its applications. *Proc. of the 3rd CSIAM Conf. on Systems and Control*, LiaoNing, China, 46–50.

[7] Situ Rong. (1997). On solutions of backward stochastic differential equations with jumps and applications. *Stochastic Process. Appl.* 66, 209–236.

[8] Situ Rong. (1999). Comparison theorem of solutions to BSDE with jumps and viscosity solution to a generalized Hamilton-Jacobi-Bellman equation. In *Control of Distributed Parameter and Stochastic Systems*, eds: S. Chen, X. Li, J. Yong, X.Y. Zhou, Kluwer Acad. Pub., Boston, 275–282.

[9] Situ Rong (1999). On comparison theorems and existence of solutions to backward stochastic differential equations with jumps and with discontinuous coefficients. *The 26th Conference on Stochastic Processes and their Applications*, 14–18, June, Beijing.

[10] Situ Rong (1999). *Reflecting Stochastic Differential Equations with Jumps and Applications.* CRC Research Notes in Mathematics 408, Chapman & Hall / CRC.

Chapter 20

DOOB'S INEQUALITY AND LOWER ESTIMATION OF THE MAXIMUM OF MARTINGALES

Li Zhichan
Department of Mathematics,
Hebei University of Technology,
Tianjin 30013)

Abstract For estimation of the maximum of submartingales, there are classical Doob's inequalities

$$E\sup_t|x_t|^p \leq q^p E|x_\infty|^p, \qquad p>1,\ \frac{1}{p}+\frac{1}{q}=1, \tag{0.1}$$

$$E\sup_t|x_t| \leq \frac{e}{e-1}\left(1+E|x_\infty|\log^+|x_\infty|\right), \qquad p=1. \tag{0.2}$$

The above two formulas used their ends x_∞, but ignored their beginnings x_0. This paper used both ends of $\{x_t, t\geq 0\}$, and gave Doob's inequality a more accurate improvement. For the maximum of martingales, we seldom see lower estimation except the trivial estimation:

$$E\sup_t|x_t|^p \geq E|x_\infty|^p, \qquad (p\geq 1).$$

[1] and [4] have given respectively a non-trivial estimation to the nonnegative continuous martingales for $p=1$ and 2. This paper considered lower estimation for all cases of $p\geq 1$, and got the corresponding inequalities.

Keywords: martingale, inequality of martingales, Dubins' and Gilat's conjectures.

1. Doob's inequality and lower estimation of the maximum of martingales

Suppose $\{F_t, t\geq 0\}$ is a filtration in probability space (Ω, F, P), $\{x_t, t\geq 0\}$ is a martingale (or nonnegative submartingale) which is right

Z. Hou et al. (eds.), Markov Processes and Controlled Markov Chains, 341–349.

continuous in (Ω, F, P), and adapts to $\{F_t, t \geq 0\}$. Let $x^* = \sup_t |x_t|$, $q > 1$ be the conjugate exponent of $p > 1$; $\frac{1}{p} + \frac{1}{q} = 1$.

Theorem 1.1 *Suppose $\{x_t, t \geq 0\}$ is a nonnegative submartingale. For $p > 1$, if*

$$Ex^{*p} < +\infty, \tag{1.1}$$

then

$$Ex^{*p} \leq q\,(q - c^*)^{p-1}\, Ex_\infty^p - (q-1)Ex_0^p = (q - c^*)^p\, Ex_\infty^p, \tag{1.2}$$

where $c^* \geq \frac{q-1}{q^{p-1}} \frac{Ex_0^p}{Ex_\infty^p}$. In particular,

$$Ex^{*p} \leq q^p Ex_\infty^p - (q-1)Ex_0^p. \tag{1.3}$$

Proof If $Ex_0^p = 0$, then (1.2) and (1.3) obviously hold, therefore we may assume $Ex_0^p > 0$. Under condition (1.1), there exists a limit $x_\infty = \lim\limits_{t \to +\infty} x_t$ (*a.s*), and $x_\infty \in L^p$. For any $\lambda > 0$, let $T_\lambda = \inf(t > 0 : t_i > \lambda + x_0)$. From the right continuity of x_t, it's not difficult to prove that T_λ is a stopping time, and $x_{T_\lambda} \geq \lambda + x_0$, (a.s) on $(T_\lambda < +\infty) = (x^* > \lambda + x_0)$. By using Fubini's Theorem,

$$\begin{aligned} Ex^{*p} &= \int_0^\infty EI\,(x^* > \lambda)\, d(\lambda^p) \\ &= E\int_0^\infty p\lambda^{p-1} I\,(x^* > \lambda)\, d\lambda \\ &= E\left[\int_0^{x_0} p\lambda^{p-1} I\,(x^* > \lambda)\, d\lambda + \int_{x_0}^\infty p\lambda^{p-1} I\,(x^* > \lambda)\, d\lambda\right]. \end{aligned} \tag{1.4}$$

Since $x^* \geq x_0$ (a.s),

$$E\int_0^{x_0} p\lambda^{p-1} I\,(x^* > \lambda)\, d\lambda = Ex_0^p. \tag{1.5}$$

Because $(x^* > \lambda + x_0) = (T_\lambda < +\infty) \in F_{T_\lambda}$, $(\lambda + x_0) \in F_0 \subset F_{T_\lambda}$, from the submartingality and the Stopping Theorem,

$$\begin{aligned} E(\lambda + x_0)^{p-1} I\,(x^* > \lambda + x_0) &\leq Ex_{T_\lambda}(\lambda + x_0)^{p-2} I\,(x^* > \lambda + x_0) \\ &\leq Ex_\infty(\lambda + x_0)^{p-2} I\,(x^* > \lambda + x_0)\,. \end{aligned}$$

Using the Fubini Theorem again, we get

$$\begin{aligned}
& E\int_{x_0}^{\infty} p\lambda^{p-1} I\left(x^* > \lambda\right) d\lambda \\
= \ & E\int_0^{\infty} p(\lambda + x_0)^{p-1} I\left(x^* > \lambda + x_0\right) d\lambda \\
= \ & p\int_0^{\infty} E(\lambda + x_0)^{p-1} I\left(x^* > \lambda + x_0\right) d\lambda \\
\le \ & p\int_0^{\infty} Ex_\infty(\lambda + x_0)^{p-2} I\left(x^* > \lambda + x_0\right) d\lambda \\
= \ & pEx_\infty \int_{x_0}^{\infty} \lambda^{p-2} I\left(x^* > \lambda\right) d\lambda \\
= \ & qEx_\infty \int_{x_0}^{x^*} d\lambda^{p-1} \\
= \ & q\left[Ex_\infty x^{*p-1} - Ex_\infty x_0^{p-1}\right] \\
\le \ & q\left(Ex_\infty x^{*p-1} - Ex_0^p\right). \qquad (1.6)
\end{aligned}$$

Synthesize (1.4), (1.5), (1.6), then we get

$$Ex^{*p} \le qEx_\infty x^{*p-1} - (q-1)Ex_0^p. \qquad (1.7)$$

Substitute the Hölder inequality, $Ex_\infty x^{*p-1} \le \left(Ex_\infty^p\right)^{1/p}\left(Ex^{*p}\right)^{1/q}$, into (1.7) and divide both sides by $\left(Ex^{*p}\right)^{1/q}$, to yield

$$\left(Ex^{*p}\right)^{1/p} \le q\left(Ex_\infty^p\right)^{1/p} - (q-1)\frac{Ex_0^p}{\left(Ex_\infty^p\right)^{1/q}}.$$

Choose a constant $c > 0$ such that

$$(q-1)Ex_0^p/(Ex^{*p})^{1/q} \ge c(Ex_\infty^p)^{1/p},$$

which can also be written as

$$Ex^{*p} \le \left(\frac{q-1}{c}\right)^q \frac{(Ex_0^p)^q}{(Ex_\infty^p)^{\frac{1}{p-1}}}. \qquad (1.8)$$

If (1.8) holds, then we have

$$Ex^{*p} \le (q-c)^p Ex_\infty^p. \qquad (1.9)$$

Let $c' = \frac{q-1}{(q-c)^{p-1}}\alpha$, ($\alpha = Ex_0^p/Ex_\infty^p$) then it is not difficult to calculate

$$(q-c)^p Ex_\infty^p = \left(\frac{q-1}{c'}\right)^q \frac{(Ex_0^p)^q}{(Ex_\infty^p)^{\frac{1}{p-1}}},$$

which shows that if (1.9) holds on c, then (1.8) holds on c'. Hence (1.9) holds on c'. In view of the above property, let $c_0 = 0$, and $c_{n+1} = (q-1)\alpha/(q-c_n)^{p-1}$, then (1.9) holds on c_n, $(n \geq 0)$. It is not difficult to check $0 < c_n < q-1$, $(n > 1)$, $c_n \uparrow$ $(n \to \infty)$, therefore there exists the limit $c^* = \lim_{n\to\infty} c_n$, $0 < c^* \leq q-1$, such that

$$c^* = \frac{(q-1)\alpha}{(q-c^*)^{p-1}} \geq \frac{q-1}{q^{p-1}} \frac{Ex_0^p}{Ex_\infty^p},$$

and

$$\begin{aligned} Ex^{*p} &\leq (q-c^*)^p \, Ex_\infty^p \\ &= q\,(q-c^*)^{p-1} \, Ex_\infty^p - c^* \,(q-c^*)^{p-1} \, Ex_\infty^p \\ &= q\,(q-c^*)^{p-1} \, Ex_\infty^p - (q-1)Ex_0^p, \end{aligned}$$

from which we get (1.2) and (1.3). This ends the proof. ∎

Dubins and Gilat conjectured in 1978 that the equality in Doob's inequality (0.1) held only if $x_t = 0$ (a.s) to any $t \geq 0$ (see [2] or [1, page 151]). Pitman [4] has proved the conjecture to be correct for $p = 2$. Paper [5] has solved this problem thoroughly and has proved the conjecture to be correct for all $p > 1$. The following corollary will show a more simple and direct proof than that given in [5].

Corollary 1.1 *Suppose $\{F_t,\ t \geq 0\}$ is right continuous, $\{x_t,\ t \geq 0\}$ is a right continuous martingale (or a nonnegative submartingale) which adapts to $\{F_t,\ t \geq 0\}$ and satisfies* (1.1). *Then, the equal-sign in (0.1) holds if and only if $x_t = 0$ (a.s) for any $t \geq 0$.*

Proof The sufficiency is obviously correct, we only need to prove the necessity.

Suppose the equality in (0.1) holds. Consider the right continuous version of martingale $E(|x_\infty| \mid F_t)$, $t \geq 0$ (it must exist under the condition that $\{F_t,\ t \geq 0\}$ is right continuous). Write $y_t = E(|x_\infty| \mid F_t)$, $t \geq 0$, then $y_t \geq 0$, $y_\infty = |x_\infty|$ (a.s), according to the submartingality of x_t, $y_t \geq |x_t|$ (a.s), $t \geq 0$. So $y^* = \sup_t y_t \geq \sup_t |x_t| = x^*$, from $y_\infty \in L^p$, we can conclude $y^* \in L^p$ (see [1] or [3]). Because $Ex^{*p} = q^p E|x_\infty|^p$,

$$q^p E|x_\infty|^p = Ex^{*p} \leq Ey^{*p} \leq q^p Ey_\infty^p = q^p E|x_\infty|^p,$$

from which we can get $Ey^{*p} = q^p Ey_\infty^p = q^p E|x_\infty^p|^p$. $\{y_t,\ t \geq 0\}$ satisfies Theorem 1.1. Put it into (1.2) or (1.3), then get $Ey_0^p = 0$, but $y_0 = E(|x_\infty| \mid F_0) \geq 0$ (a.s), so $y_0 = 0$ (a.s) and therefore $|x_\infty| = 0$ (a.s). From this we can conclude that for any $t \geq 0$, $x_t = 0$ (a.s). This ends the proof of the necessity. ∎

Theorem 1.2 *Suppose $\{x_t, t \geq 0\}$ is a nonnegative submartingale, and*

$$Ex^* < +\infty, \tag{1.10}$$

then

$$Ex^* \leq \frac{e}{e-1}\left(Ex_\infty \log^+ x_\infty - Ex_0 \log^+ x_0 + E(x_0 \vee 1)\right), \tag{1.11}$$

where $\log^+ x = \log(x \vee 1)$.

Proof Under condition (1.10), there exists the limit $x_\infty = \lim_{t\to\infty} x_t$ (a.s.). Without loss of generality, we may assume $Ex_\infty \log^+ x_\infty < +\infty$. Write $x_0^1 = x_0 \vee 1$, $\forall \lambda > 0$; let $T_\lambda = \inf(t > 0 :\ x_t > \lambda + x_0^1)$, T_λ is a stopping time, and according to the right continuity of x_t, $x_{T_\lambda} \geq \lambda + x_0^1$ (a.s.) on $(T_\lambda < +\infty)$,

$$\begin{aligned} Ex^* &= \int_0^\infty EI\,(x^* > \lambda)\,d\lambda \\ &= E\left[\int_0^{x_0^1} I\,(x^* > \lambda)\,d\lambda + \int_{x_0^1}^\infty I\,(x^* > \lambda)\,d\lambda\right] \\ &\leq Ex_0^1 + E\int_{x_0^1}^\infty I\,(x^* > \lambda)\,d\lambda. \end{aligned} \tag{1.12}$$

Because $\left(x^* > \lambda + x_0^1\right) = (T_\lambda < +\infty) \in F_{T_\lambda}$, $(\lambda + x_0^1) \in F_0 \subset F_{T_\lambda}$. From the submartingality and the Stopping Theorem, we get

$$\begin{aligned} EI\left(x^* > \lambda + x_0^1\right) &\leq E\frac{x_{T_\lambda}}{\lambda + x_0^1} I\left(x^* > \lambda + x_0^1\right) \\ &\leq E\frac{x_\infty}{\lambda + x_0^1} I\left(x^* > \lambda + x_0^1\right). \end{aligned}$$

Using the Fubini Theorem and $x^* \geq x_0$, we get

$$\begin{aligned} E\int_{x_0^1}^\infty I\,(x^* > \lambda)\,d\lambda &= E\int_0^\infty I\left(x^* > \lambda + x_0^1\right)d\lambda \\ &= \int_0^\infty EI\left(x^* > \lambda + x_0^1\right)d\lambda \\ &\leq \int_0^\infty E\frac{x_\infty}{\lambda + x_0^1} I\left(x^* > \lambda + x_0^1\right)d\lambda \\ &= Ex_\infty \int_{x_0^1}^\infty \frac{1}{\lambda} I\,(x^* > \lambda)\,d\lambda \\ &\leq Ex_\infty\left[\log\left(x^* \vee x_0^1\right) - \log x_0^1\right] \\ &= Ex_\infty \log\left(x^* \vee 1\right) - Ex_\infty \log x_0^1 \\ &\leq Ex_\infty \log^+ x^* - Ex_0 \log^+ x_0. \end{aligned} \tag{1.13}$$

Using the analytical inequality again

$$a\log^+ b \le a\log^+ a + b/e, \qquad a\ge 0,\ b\ge 0,$$

yields $Ex_\infty \log^+ x^* \le Ex_\infty \log^+ x_\infty + Ex^*/e$. Putting it into (1.13) and synthesizing with (1.12), we obtain (1.11). ∎

Remark 1.1 *Compare the right side of* (1.11) *with that of* (0.2), *it is not difficult to prove*

$$x_0^1 - x_0\log^+ x_0 \le 1.$$

So we can conclude that (1.11) *is more accurate than* (0.2). *In fact, the bigger* x_0, *the more superior formula* (1.11) *becomes. For example, take the martingale* $\{x_t \equiv e^2,\ t\ge 0\}$, *then the error of* (0.2) *is* $\left(e+e^2+e^3\right)/(e-1)$, *whereas the error of* (1.11) *is* $e^2/(e-1)$.

We now discuss the lower estimation of the maximum of martingales.

Theorem 1.3 *Suppose* $\{x_t,\ t\ge 0\}$ *is a nonnegative continuous martingale, and* $Ex^{*p} < +\infty$, $(p>1)$, *then*

$$Ex^{*p} \ge qEx_\infty^p - (q-1)Ex_0^p, \qquad (p>1). \tag{1.14}$$

Proof For any $\lambda > 0$, let $T_\lambda = \inf(t>0:\ x_t > \lambda + x_0)$, then T_λ is a stopping time. According to the continuity of x_t, $x_{T_\lambda} = \lambda + x_0$ (a.s.) on $(T_\lambda < +\infty)$. Since $x^* \ge x_\infty$, $x^*\ge x_0$ (a.s.), we have

$$\begin{aligned} Ex_\infty^p &\le Ex_\infty x^{*p-1} \\ &= Ex_\infty \int_0^{x^*} d\lambda^{p-1} \\ &= (p-1)Ex_\infty \int_0^\infty \lambda^{p-2} I\left(x^*>\lambda\right) d\lambda \\ &= (p-1)Ex_\infty \left[\int_0^{x_0} + \int_{x_0}^\infty \left(\lambda^{p-2} I\left(x^*>\lambda\right) d\lambda\right)\right] \\ &= Ex_\infty x_0^{p-1} + (p-1)Ex_\infty \int_{x_0}^\infty \lambda^{p-2} I\left(x^*>\lambda\right) d\lambda. \end{aligned} \tag{1.15}$$

By the martingality, we know $Ex_\infty x_0^{p-1} = Ex_0^p$. Since $(x^* > \lambda + x_0) = (T_\lambda < +\infty) \in F_{T_\lambda}$, $\lambda + x_0 \in F_0 \subset F_{T_\lambda}$, and thus by the Stopping Theorem, we obtain

$$\begin{aligned} Ex_\infty(\lambda+x_0)^{p-2} I(x^*>\lambda+x_0) &= Ex_{T_\lambda}(\lambda+x_0)^{p-2} I(x^*>\lambda+x_0) \\ &= E(\lambda+x_0)^{p-1} I(x^*>\lambda+x_0). \end{aligned}$$

Using Fubini's Theorem, we have

$$\begin{aligned}
& Ex_\infty \int_{x_0}^{\infty} \lambda^{p-2} I\left(x^* > \lambda\right) d\lambda \\
= & \int_0^{\infty} Ex_\infty(\lambda + x_0)^{p-2} I\left(x^* > \lambda + x_0\right) d\lambda \\
= & \int_0^{\infty} E(\lambda + x_0)^{p-1} I\left(x^* > \lambda + x_0\right) d\lambda \\
= & E \int_{x_0}^{\infty} \lambda^{p-1} I\left(x^* > \lambda\right) d\lambda \\
= & \frac{1}{p} E \int_{x_0}^{x^*} d\lambda^p \\
= & \frac{1}{p}\left(Ex^{*p} - Ex_0^p\right). \qquad (1.16)
\end{aligned}$$

Synthesize (1.15) and (1.16), we then get

$$q(Ex_\infty^p - Ex_0^p) \leq Ex^{*p} - Ex_0^p,$$

which is also formula (16). ∎

Remark 1.2 *Let $p = 2$ in (1.14), we then recover the result in [4].*

Theorem 1.4 *Suppose $\{x_t, t \geq 0\}$ is a nonnegative continuous martingale, and*

$$Ex_\infty \log^+ x_\infty < +\infty,$$

then

$$Ex^* \geq Ex_\infty \log^+ x_\infty - Ex_0 \log^+ x_0 + Ex_0. \qquad (1.17)$$

Proof For any $\lambda > 0$, let $T_\lambda = \inf(t > 0 : \ x_t > \lambda + x_0^1)$, so T_λ is a stopping time. From the continuity of x_t, $x_{T_\lambda} = \lambda + x_0^1$ on $(T_\lambda < +\infty)$, where $x_0^1 = x_0 \vee 1$. Write $x_\infty^1 = x_\infty \vee 1$, $x_1^* = x^* \vee 1$, then

$$\begin{aligned}
Ex_\infty \log^+ x_\infty &= Ex_\infty \log x_\infty^1 \\
&\le Ex_\infty \log x_1^* \\
&= Ex_\infty \int_1^{x_1^*} d(\log \lambda) \\
&= Ex_\infty \int_1^{x_1^*} \frac{1}{\lambda} d\lambda \\
&= Ex_\infty \int_1^{\infty} \frac{1}{\lambda} I\left(x_1^* > \lambda\right) d\lambda \\
&= Ex_\infty \int_1^{\infty} \frac{1}{\lambda} I\left(x^* > \lambda\right) d\lambda \\
&= Ex_\infty \left[\int_1^{x_0^1} + \int_{x_0^1}^{\infty} \left(\frac{1}{\lambda} I\left(x^* > \lambda\right) d\lambda\right)\right] \\
&\le Ex_\infty \log x_0^1 + Ex_\infty \int_{x_0^1}^{\infty} \frac{1}{\lambda} I\left(x^* > \lambda\right) d\lambda \\
&= Ex_0 \log^+ x_0 + \int_0^{\infty} Ex_\infty \frac{1}{\lambda + x_0^1} I\left(x^* > \lambda + x_0^1\right) d\lambda.
\end{aligned} \tag{1.18}$$

When calculating the integration of the right side of (1.18), we notice that $(x_1^* > \lambda + x_0^1) = (T_\lambda < +\infty) \in F_{T_\lambda}$, so $(\lambda + x_0^1)^{-1} \in F_0 \subset F_{T_\lambda}$. Using the Stopping Theorem of martingales, we get:

$$\begin{aligned}
Ex_\infty \frac{1}{\lambda + x_0^1} I(x^* > \lambda + x_0^1) &= Ex_{T_\lambda} \frac{1}{\lambda + x_0^1} I(x^* > \lambda + x_0^1) \\
&= EI(x^* > \lambda + x_0^1).
\end{aligned}$$

Substituting it into the right side of (1.18), gives

$$\begin{aligned}
\int_0^\infty Ex_\infty \frac{1}{\lambda + x_0^1} I\left(x^* > \lambda + x_0^1\right) d\lambda &= \int_0^\infty EI\left(x^* > \lambda + x_0^1\right) d\lambda \\
&\le \int_0^\infty EI\left(x^* > \lambda + x_0\right) d\lambda \\
&= E\int_{x_0}^{x^*} d\lambda \\
&= Ex^* - Ex_0.
\end{aligned} \tag{1.19}$$

Synthesizing (1.18) and (1.19) yields (1.17). ■

Remark 1.3 *Letting* $x_0 = 1$ *(a.s) in* (1.17), *we can get the result obtained by* [1, page 149].

References

[1] R. Durrett, Brownian motion and martingales in analysis. Wadsworth Inc., 1984.

[2] L.E. Dubins & D. Gilat, On the distribution of maximal of martingales. P.A.M.S., 68, 337–338, 1978.

[3] J.A. Yan, An introduction to martingales and stochastic integrals. Shanghai S&T press, 1981.

[4] J. Pitman, A note on L^2 maximal inequalities. In sem. XV.251–258. Springer LNM. 850, 1979.

[5] Z.C. Li, The proving of one conjecture on the martingales inequality.

[6] Z.C. Li, The extremal inequality of martingales and submartingales in L^2.

Chapter 21

THE HAUSDORFF MEASURE OF THE LEVEL SETS OF BROWNIAN MOTION ON THE SIERPINSKI CARPET*

Yuan Chenggui
Changsha Railway University,
China

Chen Xuerong
Changsha Railway University,
China

Abstract Let $L_t^x\{x \in F, t > 0\}$ be the local time of Brownian motion B on the Sierpinski carpet F, and $\varphi(h) = h^\beta(\log|\log h|)^{1-\beta}$, $\forall h \in (0, \frac{1}{4}]$, β is a constant. In this paper, we show that for each $x \in F$.

$$cL_t^x \le \varphi - m\{s : s \le t, B(s) = x\} \le CL_t^x, \text{ a.e. } \forall t > 0.$$

for some constants c and $C \in (0, \infty)$.

Keywords: Local time, Hausdorff measure, Level set.

1. Introduction

Let $\{X(t)\}_{t\geq 0}$ be a stable process with order $\alpha > 1$ on the line, and $A(t)$ be its local time at zero. In [2], Taylor and Wendel showed that

$$\psi - m\{s : s < t, X(s) = 0\} = C_1 A(t), \qquad a.e., \qquad \forall t > 0.$$

*Supported by NNSF of China (Grant No.19871006)

Z. Hou et al. (eds.), Markov Processes and Controlled Markov Chains, 351–361.

for some constant $C_1 \in (0,\infty)$, where $\psi - m(E)$ denotes the Hausdorff-ψ-measure of the set E, and

$$\psi(h) = h^{\beta}(\log|\log h|)^{1-\beta}, \qquad \forall h \in (0,1/4],\ \beta = 1 - \frac{1}{\alpha}.$$

Later, Perkins [3] improved the above result for the case of one dimensional Brownian motion, whereas, Zhou Xin Yin [4] gave the result for on the Sierpinski carpet.

Let F_n denote the n^{th} stage in the construction of the Sierpinski carpet, μ_n is the Hausdorff measure on F_n, $F = \cap_{n=1}^{\infty} F_n$. From [1], we know $\mu_n \Rightarrow \mu$, then μ is a Hausdorff measure on F.

In this paper, we concern with the Hausdorff measure problem for a class of processes defined on the Sierpinski carpet F. We consider the Brownian Motion B on F. As for the construction B, we refer to [5, 6, 7, 8]. In fact, Barlow and Bass have carried out many investigations about the process B. They showed that the process B, like the standard Brownian motion, also has a continuous symmetric transition density $p(t,x,y)$ with respect to the Hausdorff measure μ. Moreover, the function $p(t,x,y)$ has the following properties

Theorem 1.1 *There exists a function $p(t,x,y)$, $0 < t < \infty$, $x,y \in F$, such that*

(i) $p(t,x,y)$ is the transition density of X with respect to μ.

(ii) $p(t,x,y) = p(t,y,x)$ for all x,y,t.

(iii) $(t,x,y) \to p(t,x,y)$ is jointly continuous on $(0,\infty) \times F \times F$.

(iv) There exist constants $c_1, c_2, c_3, c_4 > 0$, and d_w such that, writing $d_s = 2d_f/d_w$,

$$\begin{aligned} & c_1 t^{-d_s/2} \exp\left(-c_2\left(|x-y|^{d_w}/t\right)^{1/(d_w-1)}\right) \\ \le\ & p(t,x,y) \\ \le\ & c_3 t^{-d_s/2} \exp\left(-c_4\left(|x-y|^{d_w}/t\right)^{1/(d_w-1)}\right). \end{aligned}$$

(v) $p(t,x,y)$ is Hölder continuous of order $d_w - d_f$ in x and y, and C^∞ in t on $(0,\infty)\times F\times F$. More precisely, there exists a constant c_5 such that

$$|p(t,x,y) - p(t,x',y)| \le c_5 t^{-1}|x-x'|^{d_w-d_f}, \quad \text{for } t>0,\ x,x',y \in F,$$

and for each $k \geq 1$, $\partial^k p(t,x,y)/\partial t^k$ is Hölder continuous of order $d_w - d_f$ in each space variable.

d_f is the Hausdorff dimension of F, d_s is the spectral dimension of F, d_w is unknown — we just have a definition in terms of the limiting resistances of the Sierpinski carpet. We also have $d_w = d_f + \overline{\xi}$, which connects the Hausdorff and spectral dimensions with the resistance exponent $\overline{\xi}$.

2. Preliminary

In this section, we make an additional study on the local time L_t^x, $\forall t > 0$, $\forall x \in F$. In fact, Barlow and Bass showed that L_t^x is jointly continuous with respect to $(t,x) \in R^+ \times F$. Moreover, the local time L_t^x satisfies the density of occupation formula

$$\int_0^{t\wedge\tau} g(B_s)ds = \int_F g(x) L_t^x \mu(dx) \tag{2.1}$$

where μ is defined in [1], $\tau = \inf\{t : B_t \in \partial F\}$

Lemma 2.1
Set

$$A_m(x) = \{y \in F : |y - x| \leq m^{-1}\}, \qquad \forall m \geq 1,\ \forall x \in F$$

Then for any $p \geq 1$

$$\lim_{m\to\infty} \left| L_t^x - [\mu(A_m(x))]^{-1} \int_0^{t\wedge\tau} \mathbf{I}_{\{|B(s)-x|\leq m^{-1}\}}\, ds \right|^p = 0 \tag{2.2}$$

Proof By (2.1) we have

$$\int_0^{t\wedge\tau} \mathbf{I}_{\{|B_s - x|\leq m^{-1}\}}\, ds = \int_{A_m(x)} L_t^y \mu(dy)$$

Since L_t^y is continuous with respect to y, one easily shows that

$$\lim_{m\to\infty} \left[\mu(A_m(x))\right]^{-1} \int_{A_m(x)} L_t^y \mu(dy) = L_t^x, \qquad \text{a.e.}$$

To prove (2.2), it suffices to show that

$$\sup_m E \left| \left[\mu(A_m(x))\right]^{-1} \int_0^t \mathbf{I}_{\{|B(s)-x|\leq m^{-1}\}}\, ds \right|^p < \infty, \qquad p \geq 1 \tag{2.3}$$

In fact, by the Markov property, we have

$$\begin{aligned}
& E\left|\int_0^t I_{\{|B(s)-x|\le m^{-1}\}}\, ds\right|^p \\
&= E\left[p!\int_0^t\int_0^{t_1}\cdots\int_0^{t_{p-1}} I_{\{|B(t_p)-x|\le m^{-1}\}}\cdots I_{\{|B(t_1)-x|\le m^{-1}\}} dt_1\cdots dt_p\right] \\
&= p!\int_0^t\int_0^{t_1}\cdots\int_0^{t_p-1}\int_F\cdots\int_F I_{\{|x_p-x|\le m^{-1}\}}\cdots I_{\{|x_1+\cdots+x_p-x|\le m^{-1}\}} \\
&= p_{t_p}(0,x_p)\cdots p_{t_1-t_2}(x_2,x_1)\mu(dx_1)\cdots\mu(dx_p)dt_0\cdots dt_p \qquad (2.4)
\end{aligned}$$

However, Theorem 1.1 tells us that,

$$\text{the right side of (2.4)} \le (C_3)^p p!\int_0^t\int_0^{t_1}\cdots\int_0^{t_{p-1}}$$

$$\begin{aligned}
&\int_{\{|x_p-x|\le m^{-1}\}}\cdots\int_{\{|x_1+\cdots+x_p-x|\le m^{-1}\}} t_p^{-d_s/2}(t_{p-1}-t_p)^{-d_s/2}\cdots(t_1-t_2)^{-d_s/2} \\
&\quad \exp\left(-C_4|x_p|^{d_w}/t\right)^{1/(d_w-1)}\cdots\exp\left(-C_4)|x_1-x_2|^{d_w}/t\right)^{1/(d_w-1)} \\
&\quad \mu(dx_1)\cdots\mu(dx_p)\,dt_1\cdots dt_p \\
&\le (C_3)^p p!\int_0^t\int_0^{t_1}\cdots\int_0^{t_{p-1}}\int_{\{|x_p-x|\le m^{-1}\}}\cdots\int_{\{|x_1+\cdots+x_p-x|\le m^{-1}\}} \\
&\quad t_p^{-d_s/2}(t_{p-1}-t_p)^{-d_s/2}\cdots(t_1-t_2)^{-d_s/2}\mu(dx_1)\cdots\mu(dx_p)\,dt_1\cdots dt_p \\
&\le (C_3^p)p!\int_0^t\int_0^{t_1}\cdots\int_0^{t_{p-1}}\int_0^{t_{p-1}} t_p^{-d_s/2}(t_{p-1}-t_p)^{-d_s/2} \\
&\quad \cdots(t_1-t_2)^{-d_s/2}(m^{-d_s})^p\int_{\{|x_p|\le 1\}}\cdots\int_{\{|x_1+\cdots+x_p|\le 1\}}\mu(dx_1)\cdots\mu(dx_p)\,dt_1\cdots dt_p \\
&< \infty
\end{aligned}$$

The proof is completed. ∎

Using this lemma, we can estimate $E|L_t^0|^p$ for $p \ge 1$

Lemma 2.2 *There exist finite positive constants c and C such that*

$$\begin{aligned}
c^p(t^{1-d_s/2^p})(p!)^{d_s/2} &\le E|L_t^0|^p \\
&\le C^p(t^{1-d_s/2})^p(p!)^{d_s/2},\quad \forall t>0,\ \forall p\ge 1 \qquad (2.5)
\end{aligned}$$

Proof By Lemma 2.1 it is sufficient to show that

$$\begin{cases} c^p \left(t^{1-d_s/2}\right)^p (p!)^{d_s/2} \le E\left|L_t^0\right|^p, \ \forall p \ge 1, \text{ and} \\ E\left|\left[\mu(A_m(x))\right]^{-1} \int_0^t I_{\{|B(s)-x|\le m^{-1}\}}\, ds\right|^p \le C^p (t^{1-d_s/2})^p (p!)^{d_s/2} \\ \quad \forall_w \ge 1, \ p \ge 1, \ t > 0. \end{cases} \tag{2.6}$$

From the Proof of Lemma 2.1, the second inequality of (2.6) holds.

On the other hand, by Theorem 1.1 and (2.4), we have

$$\begin{aligned} &E\left|\int_0^t I_{\{|B(s)-x|\le m^{-1}\}}\, ds\right|^p \\ \le\ & (C_1)^p p! \int_0^t \int_0^{t_1} \cdots \int_0^{t_{p-1}} \int_{\{|x_p-x|\le m^{-1}\}} \cdots \int_{\{x_1+\cdots+x_p-x\le m^{-1}\}} t_p^{-d_s/2} (t_{p-1}-t_p)^{-d_s/2} \\ & \cdots (t_1-t_2)^{-d_s/2} \exp\left(-C_2\left(|x_p|^{d_w}/t\right)^{1/(d_w-1)}\right) \\ & \cdots \exp\left(-C_2\left(|x_1-x_2|^{d_w}/t\right)^{1/(d_w-1)}\right) \mu(dx_1)\cdots\mu(dx_p)\, dt_1\cdots dt_p \\ =\ & (C_1)^p p! \left(m^{-d_f}\right) p \int_0^t \int_0^{t_1} \cdots \int_0^{t_{p-1}} \int_{|x_p-mx|\le 1\}} \\ & \cdots \int_{\{|x_1+\cdots+x_p-mx|\le 1\}} t_p^{-d_s/2} (t_{p-1}-t_p)^{-d_s/2} \cdots (t_1-t_2)^{-d_s/2} \\ & \exp\left(-C_2\left(m^{-1}|x_p|^{d_\omega}/t\right)^{1/(d_\omega-1)}\right) \\ & \cdots \exp\left(-C_2\left(m^{-1}|x_1-x_2|^{d_\omega}/t\right)^{1/(d_\omega-1)}\right) \mu(dx_1)\cdots\mu(dx_p) dt_1 \\ & \cdots dt_p \end{aligned}$$

Set $x = 0$, we can complete our proof. ■

The proof of Lemma 2.2 immediately yields that

$$E\left|L_{t+h}^0 - L_t^0\right|^p \le C^p \left(h^{1-d_s/2}\right)^p (P!)^{d_s/2}, \qquad \forall p \ge 1, \ \forall t > 0. \tag{2.7}$$

3. Lower bounds for Hausdorff measure

We begin with Lemmas 3.1 and 3.2.

Let

$$\beta = 1 - d_f/d_\omega, \varphi(h) = h^\beta (\log|\log h|)^{1-\beta}, \qquad \forall h \in (0, 1/4).$$

Lemma 3.1 *Suppose that μ is a completely additive measure defined on the real line Borel set and that E is a Borel set such that for each $x \in E$*

$$\lim_{h\to 0^+} \frac{\mu[x.x+h]}{\varphi(h)} \le \lambda < \infty \tag{3.1}$$

Then

$$\lambda\varphi - m(E) \ge \mu(E) \tag{3.2}$$

Proof See [2]. ∎

Set

$$
\begin{aligned}
{}_uB_v(t) &= \begin{cases} B(u+t) - B(u), & 0 \le t \le v-u \\ B(v) - B(u), & t > v-u \end{cases} \\
\Gamma &= \{f : C([0,\infty) \to R^1) \to R^1, \\
&\qquad \text{and } f \text{ is measurable and bounded.}\} \\
L(t) &= L_t^0
\end{aligned}
$$

Lemma 3.2 *Let $f \in \Gamma$, $A \in \mathcal{B}[0,1]$, then*

$$E\left(\int_A f({}_tB_1)\,dt\right) = E\int_A f({}_0B_{1-t})p_t(0,0)\,dt, \tag{3.3}$$

where $p_t(x,y)$ is the density function of $B(t)$.

Proof See [4]. ∎

Lemma 3.3 *For any fixed $t > 0$, there exists a constant $\lambda \in (0,\infty)$ such that*

$$\lim_{\lambda\to 0^+} sup \frac{L_{t+h}^0 - L_t^0}{\varphi(h)} \le \lambda \qquad a.e. \tag{3.4}$$

Proof We know from (2.7) that

$$\sup_{h\in(0.1)} E\left\{\exp\left[2^{-1}C\left(\frac{|L_{t+h}^0 - L_t^0|}{h^{1-d_s/2}}\right)^{2/d_s}\right]\right\} < \infty \tag{3.5}$$

for some constant $C \in (0,\infty)$.
Hence

$$
\begin{aligned}
&P\left[\frac{|L_{t+h}^0 - L_t^0|}{h^{1-d_s/2}} \ge a\right] \\
&\le P\left[\exp\left(2^{-1}C\left(\frac{|L_{t+h}^0 - L_t^0|}{h^{1-d_s/2}}\right)^{2/d_s}\right) \ge \exp\left(2^{-1}Ca^{2/d_s}\right)\right] \\
&\le \overline{C}\exp(-4^{-1}Ca^{2/d_s}), \quad \forall \in (0,1).
\end{aligned}
$$

We now choose

$$a_n = e^{-n/\log n}, \qquad n \geq 2$$

and set for any $\varepsilon > 0$,

$$F_n = \left\{L^0_{t+a_n} - L^0_t \geq \left(2C^{-1} + \varepsilon\right) a_{n+1}^{1-d_s/2} (\log|\log a_{n+1}|)^{-d_s/2}\right\}$$

$$G_n = \left\{L^0_{t+a_n} - L^0_t \geq \left(4C^{-1} + \varepsilon\right) a_n^{1-d_s/2} (\log|\log a_n|)^{-d_s/2}\right\}.$$

Obviously, we have

$$F_n \subseteq G_n \qquad n \geq 2.$$

and from (3.5)

$$\sum_{n=2}^{\infty} P(F_n) < \infty.$$

With the help of Borel-Cantelli lemma, we know that there exists an integer $N(\omega)$ for almost all ω, and F_n does not occur in case of $n \geq N(\omega)$.

If $a_{a+1} < a < a_n$ and $n \geq N(\omega)$, then

$$\frac{L^0_{t+a} - L^0_t}{\varphi(\lambda)} \leq \frac{L^0_{t+a} - L^0_t}{\varphi(a_n + 1)} < 2C^{-1} + \varepsilon.$$

Which ends the proof. ■

Theorem 3.1 *For any $T > 0$, there exists a constant C_1 , such that*

$$\varphi - m\{t \in (0,T) : B(t) = 0\} \geq C_1 L^0_T. \tag{3.6}$$

Proof

Set

$$\mu(dt) = L(dt)$$

by the help of Lemma 3.3, we have

$$\lim_{h \to 0^+} \frac{\mu[t, t+h]}{\varphi(h)} \leq \lambda < \infty \qquad \text{a.e.,} \quad \forall t \in [0,T]$$

So

$$P\text{-a.e. } L(dt)\text{-a.e. } \lim_{h \to 0^+} \sup \frac{\mu[t, t+h]}{\varphi(h)} \lambda < \infty.$$

Hence

$$L\left\{t \in (0,T) : \lim_{h \to 0^+} \sup \frac{\mu[t, t+h]}{\varphi(h)} > \lambda\right\} = 0.$$

Set

$$E = \left\{t \in (0,T) : B(t) = 0, \lim_{h \to 0^+} \sup \frac{\mu[t, t+h]}{\varphi(h)} \leq \lambda\right\}.$$

Then

$$L_T^0 = L(0,T) = L(E).$$

Therefore

$$\lambda\varphi - m\{t \in (0,T), B(t) = x\} \geq C_2 L_T^0 \qquad \text{a.e..}$$

■

4. Upper bounds for Hausdorff measure

According to [3], if $B(t)$ is a Brownian motion on F, then there exist constants M_1 and M_2, such that

$$M_1 \leq \lim_{\delta \downarrow 0} \sup_{\substack{0 \leq s < t \leq T \\ |s-t| < \delta}} \frac{|B(t) - B(s)|}{|s-t|^{1/d_\omega} \left(\frac{\log 1}{|s-t|}\right)^{1-1/d_\omega}} \leq M_2. \tag{4.1}$$

Set

$$\Lambda_m = \left\{ \left[k2^{-md_\omega}, (k+1)^{-md_\omega} \right]; 0 \leq k \leq 2^{md_\omega}, m = 1, 2, \cdots \right\}$$

Let I be an interval with length $|I| = 2^{-m}$. Define $B^{-1}(I) = \{t, |B(t)| \in I\}$. If $B^{-1}(I)$ meets one of the intervals of $\wedge_m$, then $B(k2^{-mdw})$ falls in an interval I', and concentric with I of length $\ll m^{1-d_\omega} \cdot 2^{-m}$,

Lemma 4.1 *It is almost surely that if I' is an interval of length $m^{-1-1/d_w} \cdot 2^{-m}$, then $B^{-1}(I')$ contains $\ll m^{(1-1/d_w)d_f + 2} 2^{(1-d_s/2)md_w}$ of the points $k \cdot 2^{-md_w}$.*

Proof Set $0 \leq k_1 < k_2 < \cdots < k_m \leq [2md_w]$
Let

$$\begin{aligned} & C_{k_1 \cdots k_m} \\ & = \left\{ \left| B\left(k_{i+1} 2^{-md_w}\right) - B\left(k_i 2^{-md_w}\right) \right| \leq m^{1-1/d_w} 2^{-m}, 1 \leq i \leq m \right\} \end{aligned}$$

$$\begin{aligned} & P\left\{ \left| B\left(k_{i+1} 2^{-md_w}\right) - B\left(k_i 2^{-md_w}\right) \right| \leq m^{1-1/d_w} 2^{-m} \right\} \\ = & \int_F \int_{\{|y-x| < m^{1-1/d_w} 2^{-m}\}} P_{k_i 2^{-md_w}}(0,x) P_{k_i} 2^{md_w}(0,x) P_{k_{i+1} 2^{-md_w}}(x,y)\, \mu(dx)\mu(dy) \\ \leq & \ C(k_{i+1} - k_i)^{-d_s/2} m^{(1-1/d_w)d_f} \end{aligned}$$

Then we have

$$P(C_{k_1 \cdots k_m}) \leq \left(C m^{1-1/d_w d_f} \right)^{m-1} (k_2 - K_1)^{-d_s/2} \cdots (k_m - k_{m-1})^{-d_s/2}$$

Hence

$$\sum_{0\le k_1<\cdots<k_m\le[2^{md_w}]} P(C_{k_1,\cdots,k_m}) \le \left[C'm^{(1-1/d_w)d_f+1}\right]^{m-1}\left(2^{md_w}\right)^{(1-d_s/2)m}$$

So we get our result. ∎

Lemma 4.2

$$P\left[\sup_{0\le s\le t}\left|L^x_{t+t_0}-\left(L^x_{t_0}-\left(L^y_{t+t_0}-L^y_{t_0}\right)\right)\right|\ge 2\delta\right] \le 2e^t\exp\left(\frac{-\delta 3^{nd_f}9^{-n}}{p\left|\frac{1}{3^n}x-\frac{1}{3^n}y\right|}\right)$$

where $p(v)=\sup_{|x-y|<v}(1-p_1(x,y)p_1(y,x))^{1/2}$, $\forall v>0$.

Proof

$$\begin{aligned} L^x_t - L^y_t &= \lim_{m\to\infty}\frac{1}{\mu(A_m(0))}\left[\int_0^t I_{\{|B(s)-x|\le m^{-1}\}}ds-\int_0^t I_{\{|B(s)-y|\le m^{-1}\}}ds\right]\\ &= \lim_{m\to\infty}\frac{1}{\mu(A_m(0))}9^n\left[\int_0^{9^{-n}t} I_{\{|\frac{1}{3^n}B(9^n s)-\frac{1}{3^n}x|\le(3m)^{-1}\}}ds\right.\\ &\qquad\left.-\int_0^{9^{-n}t} I_{\{|\frac{1}{3^n}B(9^n s)-\frac{1}{3^n}y|\le(3m)^{-1}\}}ds\right]\\ &= \lim_{m\to\infty}\frac{1}{\mu(A_m(0))}9^n\left[\int_0^{9^{-n}t} I_{\{|B(s)-\frac{1}{3^n}x|\le(3^n m)^{-1}\}}ds\right.\\ &\qquad\left.-\int_0^{9^{-n}t} I_{\{|B(s)-\frac{1}{3^n}y|\le(3^n m)^{-1}\}}ds\right]\\ &= 9^n/3^{nd_f}\left[L^{x/3^n}_{9^{-n}t}-L^{y/3^n}_{9^{-n}t}\right]. \end{aligned}$$

So

$$
\begin{aligned}
& P\left(\sup_{0\le s\le t}\left|L^x_{t+t_0} - L^x_{t_0} - \left(L^y_{t+t_0} - L^y_{t_0}\right)\right| \ge 2\delta\right) \\
= & \int_F P^z\left(\sup_{0\le s\le t}|L^x_s - L^y_s| \ge 2\delta\right) P_s(0,z)\mu(dz) \\
= & P\left(\sup_{0\le s\le t}\left|L^{x/3^n}_{9^{-n}s} - L^{y/3^n}_{9^{-n}s}\right| \ge 2\delta 3^{d_f} 9^{-n}\right) \\
\le & 2\exp\left(\frac{-\delta 3^{nd_f} 9^{-n}}{p\left|\frac{1}{3^n}x - \frac{1}{3^n}y\right|}\right).
\end{aligned}
$$

The proof is finished. ■

Set

$$
\begin{aligned}
a_n &= 2^{-n^{1+\delta}}, \qquad \delta \in (0,1), \\
t_A &= i\,a_n, \\
S_A &= (i+1)a_n, \\
& D(t_A, a_k) \text{ is the interval } [t_A - a_k, t_A + a_k], \\
A &= [ia_n, (i+1)a_n], \\
\Omega_n &= \left\{A = [ka_n, (k+1)a_n], 0 \le k \le 2^{n(1+\delta)}\right\}.
\end{aligned}
$$

Let T_n be the number of A belonging to Ω_n which intersects with $\{s : s \le 1, B(s) = 0\}$. By Lemma 4.1

$$
P\left[T_n \ge m^{(1-d_w)d_f+2} 2^{(1-d_s/2)m,^{1+\delta}}\right] \le 2^{-2m^{1+\delta}}.
$$

In view of reference [3] and [6], we have

$$
\lim_{n\to\infty} \inf \varphi(a_n)T_n = 0,
$$

and based on reference [5], $\varphi - m\left[s \le T : B(s) = 0\right\} \le C_3 L^0_T\right]$.

We get our result.

Acknowledgments

Many thanks to Professor X.Y. Zhou for his encouragement.

References

[1] Barlow, M.T., Bass, R.F.: The construction of Brownian motion on the Sierpinski carpet. Ann. Inst. Henri Poincare 25, 225–257 (1989).

[2] Taylor, S.J., Wendel, J.: The exact Hausdorff measure of the zero set of a stable process, Z.W. 6, 170–180 (1966).

[3] Pekins. E: The exact Hausdorff measure of the level sets of Brownian motion, Z.W. 58, 373–388 (1981).

[4] Zhou, X.Y.: The Hausdorff measure of the level sets of Brownian motion of the Sierpinski carpet. Nankai Series, Pure Appl. Math, Th. phs. prob., and stati., eds, Z.P. Jiang et. al, World Scientific Press (1992) 283–300.

[5] Barlow, M.T., Bass, R.F: Local times for Brownian motion on the Sierpinski carpet. Probab. Theory Relat. Field 85, 91–104 (1990).

[6] Barlow, M.T., Bass, R.F: On the resistance of the Sierpinski carpet. Proc. R. Soc. Lond. Ser. A 431, 345–360 (1990).

[7] Barlow, M.T., Bass, R.F, Sherwood, J.D: Resistance and spectral dimension of Sierpinski carpets. J. phys. A. 23, L_{253}–L_{258} (1990).

[8] Barlow, M.T., Bass, R.F: Transition densities for Brownian motion on the Sierpinski carpet. Probab. Theory Relat. Fields. (1992).

Chapter 22

MONOTONIC APPROXIMATION OF THE GITTINS INDEX*

Xikui Wang
Department of Statistics,
University of Manitoba
Winnipeg, Manitoba,
Canada R3T 2N2
xikui_wang@umanitoba.ca

Abstract The Gittins index is useful in the study of bandit processes and Markov decision processes, and can be approximated by finite horizon break-even values determined in the truncated finite horizon models. These break-even values are shown to form a nondecreasing sequence. A finite horizon optimal stopping solution is also derived.

Keywords: Markov decision processes; bandit processes; Gittins index; dynamic programming; geometric discounts; optimal stopping.

1. Introduction

The celebrated Gittins index, or dynamic allocation index, was introduced in Gittins and Jones [4] for the study of sequential designs of experiments. It has been very useful and powerful in the study of Markov decision processes and bandit problems (Gittins [3], Berry and Fristedt [1]).

The calculation of the Gittins index involves optimal stopping times and is formidable in most practice. Approximations for the Gittins index and error bounds have been discussed by Berry and Fristedt [1], Gittins [3], Wang [6], Chen and Katehakis [2], and Katehakis and Vei-

*This work was completed at the Division of Community Health, Memorial University of Newfoundland, and is supported by a grant from NSERC of Canada.

Z. Hou et al. (eds.), Markov Processes and Controlled Markov Chains, 363–367.

nott [5]. Especially, the Gittins index can be approximated by a sequence of finite horizon break-even values, which determine optimal strategies in finite horizon models obtained by truncating the infinite horizon geometric discount sequence. These break-even values can be found numerically by the method of dynamic programming.

This paper shows a monotonicity of these break-even values. In section 2, we introduce the bandit model, the relevant Markov decision process, and the Gittins index. The monotonicity property is shown in Section 3. This result accelerates the computations and offers some insight into the finite horizon decision making process. Moreover, a finite horizon optimal stopping solution is derived.

2. The bandit model, Markov decision process, and Gittins index

In a bandit with k independent unknown arms, each arm i ($i = 1, 2, \ldots, k$) consists of a sequence of conditionally independent and identically distributed random responses $\{X_{i,j}, j = 1, 2, \ldots\}$ with an unknown distribution G_i. One and only one arm is selected for observation at each time $n = 0, 1, 2, \ldots$. The objective is to maximize the expected total discounted responses $E_\pi(\sum_{n=1}^{\infty} \alpha_n Z_n)$, where $A = (\alpha_1, \alpha_2, \ldots)$, $\alpha_n \geq 0$, $\sum_{n=1}^{\infty} \alpha_n < \infty$, is the discount sequence, π is a strategy, and Z_n is the response resulted from the n^{th} selection specified by π.

For geometric discounts $A = (1, \alpha, \alpha^2, \ldots)$, Gittins and Jones [4] compare each unknown arm with a common known arm and a number called the Gittins index is calculated. The Gittins and Jones strategy, which selects an unknown arm with the largest Gittins index value, is optimal. It is shown in Berry and Fristedt [1] that the Gittins and Jones strategy is optimal if and only if the discount sequence is geometric $A = (1, \alpha, \ldots, \alpha^n, \ldots)$, $0 < \alpha < 1$. A k-armed bandit process then becomes a collection of k two-armed bandit processes.

Consider the approximation of the Gittins index for each two-armed bandit. Suppose that the random responses $X_1, X_2, \ldots$, on the unknown arm (denoted as arm 1) follow an unknown distribution G, which has a prior distribution F on $\mathcal{D}$ under the Bayesian approach. $\mathcal{D}$ is the space of all probability distributions on $(-\infty, \infty)$. The random responses on the known arm (denoted as arm 0) have a known mean λ. We call this a two-armed (F, λ, A)-bandit.

This two-armed bandit becomes a Markov decision process in the natural way: the state space consists of all possible F and the action space is $\{i = 1, 0\}$, indicating that arm i is selected for observation. For any observation x on the unknown arm, the mapping from the prior distribution F to the posterior $(x)F$ is measurable (Berry and Fristedt [1]). At any state F, the reward is $E(X|F) = \int_{\mathcal{D}} \int_{-\infty}^{\infty} x dG(x)\, dF(G)$ if arm 1 is selected or λ if arm 0 is selected.

In the (F, λ, A)-bandit, $\Delta(F, \lambda, A) = V^{(1)}(F, \lambda, A) - V^{(0)}(F, \lambda, A)$ determines the optimal initial selection, where $V^{(i)}(F, \lambda, A), i = 1, 0,$ is the worth of selecting arm i initially and then continuing with an optimal strategy. The Gittins index $\Lambda(F, A)$ is the solution for λ in $\Delta(F, \lambda, A) = 0$ (Berry and Fristedt [1]).

Let $A_n = (1, \alpha, \alpha^2, \ldots, \alpha^{n-1}, 0, 0, \ldots)$, $n = 1, 2, \ldots$, be the truncated discount sequence. By Theorem 5.3.1 in Berry and Fristedt [1], there is a $\lambda = \Lambda(F, A_n)$ for $\Delta(F, \lambda, A_n) = 0$ such that the unknown (the known) arm is optimal initially in the (F, λ, A_n)-bandit if and only if $\lambda \leq (\geq) \Lambda(F, A_n)$. $\Lambda(F, A_1) = E(X|F)$ may be found numerically since the myopic strategy is optimal for one selection. Moreover, $\lim_{n \to \infty} \Lambda(F, A_n) = \Lambda(F, A)$ (Berry and Fristedt [1]).

3. The monotonocity of the approximation

We show that for any F, $\Lambda(F, A_n), n = 1, 2, \ldots$, form a nondecreasing sequence. This indicates that the more selections to make, the more opportunity to choose the unknown arm. This is intuitive since we have to balance the competing goals of information gathering (understanding the unknown arms and making better informed selections in the future) and immediate payoff (making selections with high immediate payoffs). We need the following lemma.

Lemma 3.1 *Assume a two-armed (F, λ, A)-bandit with a geometric discount sequence $A = (1, \alpha, \alpha^2, \ldots)$.*
If $\Delta(F, \lambda, A_n) = 0$, then $\Delta(F, \lambda, A_{n+1}) \geq 0$.

Proof $\Delta(F, \lambda, A_n) = 0$ implies that $V(F, \lambda, A_n) = \lambda + \alpha V(F, \lambda, A_{n-1})$ and $E(X_1|F) - \lambda = \alpha V(F, \lambda, A_{n-1}) - \alpha E(V((x)F, \lambda, A_{n-1})|F)$. Therefore,

$$\begin{aligned}
&\Delta(F, \lambda, A_{n+1}) \\
&= E(X_1|F) - \lambda + \alpha E(V((x)F, \lambda, A_n)|F) - \alpha V(F, \lambda, A_n) \\
&= \alpha(1 - \alpha) V(F, \lambda, A_{n-1}) - \alpha\lambda \\
&\quad + \alpha E\{V((x)F, \lambda, A_n) - V((x)F, \lambda, A_{n-1})|F\}.
\end{aligned}$$

Let π^* be an optimal strategy for $V((x)F, \lambda, A_{n-1})$ and π^{**} be a strategy for the $((x)F, \lambda, A_n)$-bandit which follows π^* for the first $n - 1$ stages and then always select the known arm at the last stage. Then

$$V((x)F, \lambda, A_n) \geq W((x)F, \lambda, A_n; \pi^{**}) \geq V((x)F, \lambda, A_{n-1}) + \alpha^{n-1} k.$$

So,

$$E\{V((x)F,\lambda,A_n) - V((x)F,\lambda,A_{n-1})|F\} \geq \alpha^{n-1}k.$$

On the other hand, $V(F,\lambda,A_{n-1}) \geq \lambda + \alpha\lambda + \cdots + \alpha^{n-2}\lambda$. Hence,

$$\Delta(F,\lambda,A_{n+1}) \geq \alpha(1-\alpha)(1+\alpha+\cdots\alpha^{n-2})\lambda - \alpha\lambda + \alpha\alpha^{n-1}\lambda = 0.$$

■

Theorem 3.1
Assume a two-armed (F,λ,A)-bandit with $A = (1,\alpha,\alpha^2,\ldots)$. Then

$$\Lambda(F,A_1) \leq \Lambda(F,A_2) \leq \cdots \leq \Lambda(F,A_n) \leq \cdots \leq \Lambda(F,A).$$

Proof $\Delta(F,\Lambda(F,A_n),A_n) = 0$ implies $\Delta(F,\Lambda(F,A_n),A_{n+1}) \geq 0$. Now, $\Delta(F,\Lambda(F,A_{n+1}),A_{n+1}) = 0$ and $\Delta(F,\lambda,A_{n+1})$ is strictly decreasing in λ by Corollary 5.1.1 in Berry and Fristedt [1].
So $\Lambda(F,A_n) \leq \Lambda(F,A_{n+1})$. ■

Corollary 3.1
Assume a two-armed (F,λ,A)-bandit with $A = (1,\alpha,\alpha^2,\ldots)$. If the known arm becomes optimal for the (F,λ,A_n)-bandit, then it remains optimal for the rest of the stages.

Proof This is clear since the known arm is optimal if $\lambda \geq \Lambda(F,A_n)$, which implies that $\lambda \geq \Lambda(F,A_l)$ for $l = 1,\ldots,n-1$. □

4. A simulation example

Consider a Bernoulli bandit. The probability of success on the unknown (known) arm is θ (λ). λ is known and θ is either a or $b, 0 < b < a < 1$ with a prior $F = pI_{\{a\}} + (1-p)I_{\{b\}}$. Even in such a simple case, an explicit solution is prohibited for general (a,b) (example 5.4.1, Berry and Fristedt [1]).

Based on 5000 simulations, both the monotonicity and the convergence of the finite horizon break-even values have been observed. A part of the result is as follows, where $a = 0.8, b = 0.4, \alpha = 0.6$, and $p = 0.5$.

Horizon n	1	2	3	4	5	6
$\Lambda(F,A_n)$	0.6000	0.6177	0.6531	0.6828	0.7088	0.7273

Horizon n	7	8	9	10	11	12
$\Lambda(F,A_n)$	0.7362	0.7411	0.7444	0.7463	0.7473	0.7473

Acknowledgments

I am grateful to a referee for suggesting a better presentation of this paper and for pointing out similar monotonicity results in Chen and Katehakis [2] and an alternative approach in Katehakis and Veinott [5].

References

[1] Berry, D. A. and Fristedt, B. (1985). *Bandit Problems - Sequential Allocation of Experiments.* Chapman and Hall, London, New York.

[2] Chen, Y. R. and Katehakis, M. N. (1986). Linear programming for finite state multi-armed bandit problems. *Math. Oper. Res.* 11 (1), 180–183.

[3] Gittins, J. C. (1989). *Multi-armed bandit allocation indices.* John Wiley and Sons, Chichester.

[4] Gittins, J. C. and Jones, D. M. (1974). A dynamic allocation index for the sequential design of experiments, in *Progress in Statistics* (eds. J. Gani *et al.*), pp. 241–266, North-Holland, Amsterdam.

[5] Katehakis, M. N. and Veinott Jr. A. F. (1987). The multi-armed bandit problem: decomposition and computation. *Math. Oper. Res.* 12 (2), 262–268.

[6] Wang, Y. G. (1997). Error bounds for calculation of the Gittins indices. *Austral. J. Statist.* 39(2), 225–233.

IV

APPLICATIONS TO FINANCE, CONTROL SYSTEMS AND OTHER RELATED FIELDS

Chapter 23

OPTIMAL CONSUMPTION-INVEST -MENT DECISIONS ALLOWING FOR BANKRUPTCY: A BRIEF SURVEY*

Suresh P. Sethi
School of Management
The University of Texas at Dallas

Abstract This paper surveys the research on the optimal consumption and investment problem of an agent who is subject to bankruptcy that has a specified utility (reward or penalty). The bankruptcy utility, modelled by a parameter, may be the result of welfare subsidies, the agent's innate ability to recover from bankruptcy, psychic costs associated with bankruptcy, etc. Models with nonnegative consumption, positive subsistence consumption, risky assets modelled by geometric Brownian motion or semi-martingales are discussed. The paper concludes with suggestions for open research problems.

1. Introduction

This paper surveys the research on the optimal consumption-investment problem facing a utility maximizing agent (an individual or a household) that is subject to bankruptcy, the utility being associated with consumption and bankruptcy; for an in depth study of the problem, see Sethi [29]. The problem has its beginning in the classical works of Phelps [22], Hakansson [6], Samuelson [28], and Merton [18]. In a finite-horizon discrete-time framework, Samuelson [28] showed that for isoelastic marginal utility functions (i.e., $U'(c) = c^{\delta-1}, \delta < 1$), the opti-

*This paper is a shortened version of Sethi, S. P., Optimal Consumption-Investment Decisions Allowing for Bankruptcy: A Survey, in *Worldwide Asset and Liability Modeling*, William T. Ziemba and John M. Mulvey (Editors), Cambridge University Press, Cambridge, U.K., pp. 397-426. The research was supported in part by SSHRC Grant 410-93-042 and The University of Texas at Dallas.

Z. Hou et al. (eds.), Markov Processes and Controlled Markov Chains, 371–387.

mal portfolio decision is independent of wealth in each period and independent of the consumption decision. More specifically, the portfolio is re-balanced at each period so that the fraction of wealth invested in the risky asset remains a constant. Merton [18] confirmed the result in the continuous-time infinite horizon case.

A significant plateau was reached by Merton [19], who formulated many interesting problems in continuous time with geometric Brownian motions to model the uncertainties in the prices of risky assets. He chose the utility of consumption $U(c)$ to belong to the HARA (**H**yperbolic **A**bsolute **R**isk **A**version) class and obtained explicit solution in the case when the marginal utility at zero consumption is infinite (i.e., $U'(0) = \infty$). Among the important findings was the statement of the so-called *mutual fund theorem* that allows, under certain conditions, efficient separation of the decision to invest in the individual assets from the more macro allocational choices among classes of assets. This result represents a multi-period generalization of the well-known Markowitz-Tobin mean-variance portfolio rules.

Merton's [19, 20] analysis was erroneous in the case of HARA utility functions with $U'(0) < \infty$, as identified years later by Sethi and Taksar [32]. What Merton had done was to formally write the dynamic programming equation for the value function of the problem and provided an explicit solution of the equation. In the absence of a verification theorem, however, there is no guarantee that the solution obtained is the value function. Indeed, when $U'(0) < \infty$, not only his solution not the value function, but if it were, it would also imply negative consumption levels at some times. Missing in Merton's formulation were an all important boundary condition that the value function should satisfy, and the requirement that consumption be nonnegative. Without a boundary condition, it is not possible to obtain a verification theorem and without the nonnegativity requirement, negative consumption may occur.

A simple boundary condition specifies the value function at zero wealth. In addition to being mathematically expedient, the value function at zero wealth signifies the reward or penalty, or more generally utility, associated with bankruptcy. The value of the reward or penalty associated with bankruptcy will have consequences on agent's decisions. Lippman, McCall and Winston [16] underscore the importance of bankruptcy when they write,

> "Valid inferences concerning an agent's neutrality or aversion to risk must necessarily emanate from a highly robust model. Failure to include a constraint such as bankruptcy might very well produce the *maximally incorrect inference* (italics supplied)."

The specific value of the utility at bankruptcy depends on what is assumed to happen in its wake. In most modern societies, the agent can count on welfare if and when he goes bankrupt. In this case, the value may represent the discounted expected utility of future consumption stream provided by the government. In addition, as Gordon and Sethi [4, 5] indicate, bankruptcy may carry with it negative or positive psychic income, the former to the extent that shame attaches to going bankrupt or living on the dole and the latter to the extent that poverty may be a blessing to devoutly religious people. Mason [17] considered the case in which the agent might be re-endowed and allowed to restart the decision problem. Sethi and Taksar [33] consider a delayed recovery model of bankruptcy. Whatever the case, it is sufficient for mathematical purposes to assign a utility P to bankruptcy, and include P as a parameter of the problem. Karatzas, Lehoczky, Sethi, and Shreve [9] (KLSS hereafter) do this in their comprehensive treatment of the consumption-investment problem with nonnegative consumption requirement and bankruptcy.

We begin our survey with the discussion of the KLSS model in the next section. We also indicate how it generalizes the existing results, and discuss its implication for the agent's risk-aversion behavior as studied in Presman and Sethi [23]. In Section 3, we list models that require a subsistence or a minimum *positive* consumption rate, and the impact of this requirement on the risk-aversion behavior of the agent. Section 4 discusses briefly the influence of imposing borrowing and shortselling constraints. The constraints can give rise to more complicated value functions than the concave ones obtained earlier. See Sethi [30] for detailed versions of Sections 3 and 4. Section 5 concludes with a brief discussion of related research and open research problems.

2. Constant market coefficients with nonnegative consumption

In this section, we shall review models that assume constant interest rate, constant average mean rates of return on risky assets, and a variance-covariance matrix of constants. The models require nonnegative consumption rates. All models reviewed here allow explicitly for bankruptcy.

2.1 The KLSS model

KLSS consider a single agent attempting to maximize total discounted utility from consumption over an infinite horizon. The agent begins with an initial wealth x and makes consumption and investment decisions over time, which is assumed to be continuous. The agent has his wealth in $N+1$ distinct assets available to him. One is riskless (deterministic) with a rate of return $r > 0$, whereas the others are risky and are modelled by geometric Brownian motions. More specifically, the price dynamics of the available assets are given by

$$\frac{dP_0(t)}{P_0(t)} = rdt, \tag{2.1}$$

$$\frac{dP_i(t)}{P_i(t)} = \alpha_i dt + \mathbf{e}_i \mathbf{D}\, d\mathbf{w}_t^T, \tag{2.2}$$

where $P_0(t)$ is the price of the riskless asset and $\mathbf{P}(t) = (P_1(t), P_2, \ldots, P_N(t))$ is the vector of prices of N risky assets at time t, with given initial prices $P_0(0)$ and $\mathbf{P}(0)$. Furthermore, $\{\mathbf{w}_t,\ t \geq 0\}$ is an N-dimensional standard Wiener process given on the probability space $(\Omega, \mathcal{F}, \mathcal{P})$, $\mathbf{e}_i$ is the unit row vector with a 1 in the i^{th} position, α_i is the average rate of return on the i^{th} asset, the volatility matrix $\mathbf{D}$ is an $N \times N$ matrix with $\sum = \mathbf{D}\mathbf{D}^T$, a positive definite variance-covariance matrix, and $(^T)$ denotes the transpose operation.

The agent specifies a consumption rate c_t, $t \geq 0$, and an investment policy $\boldsymbol{\pi}_t = (\pi_1(t), \ldots, \pi_N(t))$, $t \geq 0$, where $\pi_i(t)$ denotes the fraction of wealth invested in the i^{th} investment at time t. The remaining fraction $\pi_0(t) = 1 - (\pi_1(t) + \pi_2(t) + \cdots + \pi_N(t))$ is invested in the riskless asset. The vector $\boldsymbol{\pi}_t$ is unconstrained, implying that unlimited borrowing and short-selling are allowed. We assume no transaction costs for buying and selling assets. The consumption rate must be nonnegative, i.e.,

$$c_t \geq 0, \qquad \text{a.s. } \omega, \quad \text{a.e. } t. \tag{2.3}$$

Both $C \triangleq \{c_t,\ t \geq 0\}$ and $\Pi \triangleq \{\boldsymbol{\pi}_t,\ t \geq 0\}$ must depend on the price vector $\{\mathbf{P}(t), t \geq 0\}$ in a non-anticipative way.

Given C and Π, it can be shown that the dynamics of the agent's wealth x_t, $t \geq 0$, satisfy the Itô stochastic differential equation

$$dx_t = (\boldsymbol{\alpha} - r\mathbf{1})\boldsymbol{\pi}_t^T x_t\, dt + (rx_t - c_t)\, dt + x_t \boldsymbol{\pi}_t \mathbf{D}\, d\mathbf{w}_t^T, \quad x_0 = x, \tag{2.4}$$

where $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_N)$ and $\mathbf{1} = (1, 1, \ldots, 1)$.

A complete formulation of the model requires some assumption concerning the options available to the agent if and when his wealth reaches zero, since further consumption would result in negative wealth. One possible, and quite general, treatment is to assign a value $P \in (-\infty, \infty)$ to bankruptcy and include it as a parameter of the model.

To define the agent's objective function, one needs to specify his utility function of consumption. This function U defined on $(0, \infty)$ is assumed to be strictly increasing, strictly concave, and *thrice* continuously differentiable. Extend U to $[0, \infty)$ by defining $U(0) = \lim_{c \downarrow 0} U(c)$. The agent chooses C and Π in order to maximize

$$V_{C,\Pi}(x) \triangleq E_x\left[\int_0^{T_x} e^{-\beta t}U(c_t)\,dt + Pe^{-\beta T_x}\right], \tag{2.5}$$

where $T_x = \inf\{t | x(t) = 0\}$ is the stopping time of bankruptcy when the initial wealth is x and $\beta > 0$ is the agent's discount rate. $P = U(0)/\beta$ is equivalent to continuing the problem indefinitely after bankruptcy with only zero consumption, and is termed the *natural payment.*

The value function is defined as,

$$V(x) = \begin{cases} \sup_{C,\Pi} \quad V_{C,\Pi}(x), & \text{if } x > 0, \\ \text{P}, & \text{if } x = 0. \end{cases} \tag{2.6}$$

Define the nonnegative constant $\gamma = (\frac{1}{2})(\boldsymbol{\alpha} - r\mathbf{1})\Sigma^{-1}(\boldsymbol{\alpha} - r\mathbf{1})^T$ and consider the quadratic equation $\gamma\lambda^2 - (r - \beta - \gamma)\lambda - r = 0$ with the solutions $\lambda_- < -1$ and $\lambda_+ > 0$ when $\gamma > 0$. When $\boldsymbol{\alpha} = r\mathbf{1}$ and $\beta < r$, define $\lambda_- = -r/(r - \beta)$. It is shown that $V(x)$ is finite for every $x > 0$ if

$$\int_c^{\infty} \frac{d\theta}{U'(\theta)^{\lambda_-}} < \infty, \qquad \forall c > 0. \tag{2.7}$$

Presman and Sethi [27] show that if the agent had an exponentially distributed random lifespan with the mortality rate λ, his problem could be reduced to the KLSS problem of an infinite horizon agent whose discount rate is $\beta + \lambda$.

2.2 The mutual fund theorem and the reduced model

In order to simplify the problem, choose any α and $\sigma > 0$ so that

$$\frac{(\alpha - r)^2}{2\sigma^2} = \gamma, \tag{2.8}$$

and consider the "reduced" problem with a single risky asset with drift α and variance σ^2, and the riskless asset with the rate of return r. The term $(\alpha - r)$ is known as the risk premium and $(\beta + \gamma)$ the risk-adjusted discount rate.

The mutual fund theorem states that, at any point in time, the agent will be indifferent between choosing from a linear combination of the above two assets or a linear combination of the original $(N + 1)$ assets. It is termed the mutual fund theorem, because the single risky asset can be thought of as a mutual fund. If one constructs a mutual fund which trades continuously using a self-financing strategy to maintain the proportions of the riskless and N risky assets given by the $(N+1)$-dimensional vector $\left(1-(\boldsymbol{\alpha}-r\mathbf{1})\sum^{-1}\mathbf{1}^T,\ (\boldsymbol{\alpha}-r\mathbf{1})\sum^{-1}\right)$, then the mutual fund has mean return $\alpha = r + 2\gamma$ and variance $\sigma^2 = 2\gamma$, which satisfy (2.8). Moreover, if $(\boldsymbol{\alpha}-r\mathbf{1})\sum^{-1}\mathbf{1}^T \neq 0$, then the mutual fund consisting only of risky stocks held with proportions $(\boldsymbol{\alpha}-r\mathbf{1})\sum^{-1}/(\boldsymbol{\alpha}-r\mathbf{1})\sum^{-1}\mathbf{1}^T$ also satisfies (2.8).

This important theorem was first stated by Merton [19] for the dynamic consumption-investment problem without bankruptcy considerations and without a rigorous proof. The rigorous proof is supplied by KLSS for all values of P. The theorem generalizes the Markowitz-Tobin separation theorem to multiple periods. Moreover, in the special case when $(\boldsymbol{\alpha}-r\mathbf{1})\sum^{-1}\mathbf{1}^T \neq 0$, the derived optimal portfolio policy has the same structure as that prescribed in the mean-variance model.

The mutual fund theorem is based on the strict concavity of the value function $V(x)$, which, in turn, is brought about by the assumption that the investment vector $\boldsymbol{\pi}_t$ is unconstrained.

In view of the mutual fund theorem, it suffices to consider the reduced problem with the modified wealth dynamics

$$dx_t = (\alpha - r)\pi_t x_t\,dt + (rx_t - c_t)\,dt + x_t\pi_t\sigma\,dw_t, \quad x_0 = x, \tag{2.9}$$

in place of (2.4), where $\{w_t, t \geq 0\}$ is a standard Wiener process and π_t denotes the fraction of the wealth invested in the risky asset.

2.3 The HJB equation and the solution of the problem

From the theory of stochastic optimal control, it is known that the value function $V(x)$ must satisfy the HJB (Hamilton-Jacobi-Bellman) equation:

$$\begin{aligned}\beta V(x) &= \max_{c\geq 0,\pi}\left[(\alpha-r)\pi x V'(x) + (rx-c)V'(x) + \frac{1}{2}\pi^2\sigma^2x^2V''(x) + U(c)\right],\\ &\qquad x > 0,\ V(0) = P.\end{aligned} \tag{2.10}$$

Assume $\alpha \neq r$; see Lehoczky, Sethi, and Shreve [14] or Section 4 for the special case $\alpha = r$.

The optimal feedback policies for investment and consumption are respectively:

$$\pi(x) = -\frac{(\alpha-r)V'(x)}{\sigma^2 x V''(x)}, \quad \text{and} \tag{2.11}$$

$$c(x) = \max\{U'^{-1}(V'(x)), 0\}. \tag{2.12}$$

When (2.11) and (2.12) are substituted in (2.10), it results in a highly nonlinear differential equation, which appears to be very difficult to solve at first sight. However, KLSS discovered a change of variable that allowed them to convert the nonlinear equation into a linear second-order differential equation in a variable that represents the inverse of the marginal (indirect or derived) utility of wealth given by the first derivative $V'(x)$ of the value function. Since the resulting equation has many solutions depending on the constants of integration, one needs to identify the values of the constants that would yield the value function. Furthermore, when the candidate feedback policies are expressed in terms of the solution of the linear differential equation involving the constants, KLSS discovered surprisingly that the candidate marginal utility of wealth over time can be written as a process satisfying a linear Itô's stochastic differential equation. It is then a simple matter to evaluate the objective function value associated with the candidate policies and identify the one satisfying the HJB equation. The procedure yields the value function in view of the additional fact that any function satisfying (2.10) majorizes the value function as shown in KLSS.

Solutions for the general consumption utility functions have been obtained in KLSS. Because of the space limitation, we characterize the results in Table 23.1. In this table, q denotes the probability of bankruptcy under the optimal policy, and P^*, $\bar{x}$, and a are given as:

$$P^* = \frac{1}{\beta}U(0) - \frac{U'(0)^{1+\lambda_-}}{\beta\lambda_-}\int_0^\infty \frac{d\theta}{U'(\theta)^{\lambda_-}}, \tag{2.13}$$

$$\begin{aligned} \bar{x} &= \frac{[\beta P - U(0)]^{\frac{\lambda_- - \lambda_+}{1+\lambda_-}} - \left[\int_0^\infty \frac{d\theta}{U'(\theta)^{\lambda_-}}\right]^{\frac{1+\lambda_+}{1+\lambda_-}}}{\gamma(\lambda_+ - \lambda_-)(-\lambda_-)^{\frac{1+\lambda_+}{1+\lambda_-}}} U'(0)^{\lambda_+} \\ &\quad - \frac{\int_0^\infty \frac{d\theta}{U'(\theta)^{\lambda_-}}}{\gamma\lambda_-(\lambda_+ - \lambda_-)} U'(0)^{\lambda_-}, \end{aligned} \tag{2.14}$$

and a is given by the unique positive solution for c in the equation,

$$-\frac{U'(c)^{1+\lambda_-}}{\gamma\lambda_-(1+\lambda_-)}\int_c^\infty \frac{d\theta}{U'(\theta)^{\lambda_-}} - \frac{1+\lambda_+}{\beta}U(c) + \frac{\lambda_+}{r}cU'(c) = -\rho_+ P. \tag{2.15}$$

Formulas for the value function $V(x)$, modulo some transcendental equations, are derived in KLSS. Given $V(x)$, the optimal feedback policies can be obtained from (2.11) and (2.12).

Table 23.1. Characterization of optimal consumption and bankruptcy probability

	$U'(0) = \infty$	$U'(0) < \infty$
$P \leq \frac{1}{\beta}U(0)$	$c_t > 0$, $q = 0$.	$c_t = 0$, if $x_t \in (0, \bar{x}]$, $c_t > 0$, if $x_t \in (\bar{x}, \infty)$, $q = 0$.
$\frac{1}{\beta}U(0) < P \leq P^*$	$c_t > a > 0$, $0 < q < 1$, if $\beta < r + \gamma$, $q = 1$, if $\beta \geq r + \gamma$.	$c_t = 0$, if $x_t \in (0, \bar{x}]$, $_t > 0$, if $x_t \in (\bar{x}, \infty)$, ($\bar{x}$ when $P = P^*$.) $0 < q < 1$, if $\beta < r + \gamma$, $q = 1$, if $\beta \geq r + \gamma$.
$P^* < P < \frac{1}{\beta}U(\infty)$		$c_t > a > 0$, $0 < q < 1$, if $\beta < r + \gamma$, $q = 1$, if $\beta \geq r + \gamma$.
$\frac{1}{\beta}U(\infty) \leq P$	Consume quickly to bankruptcy.	No optimal policy. $V(x) = P$, $x \geq 0$

2.4 Solutions for the HARA utility class

The HARA utility functions on $(0, \infty)$ have the form:

$$U = (1/\delta)(c + \eta)^\delta, \qquad \delta < 1,\ \delta \neq 0,\ \eta \geq 0, \tag{2.16}$$
$$U = \log(c + \eta), \qquad \eta \geq 0. \tag{2.17}$$

The log utility function (2.17) is referred to as the HARA function with $\delta = 0$. In these cases, the growth condition (2.7) specializes to $\beta > r\delta + \gamma\delta/(1-\delta)$, which is weaker than $\beta > r\delta + \gamma(2-\delta)/(1-\delta)$ imposed by Merton [18, condition (41)].

Merton [19, 20] provides explicit solutions for $V(x)$ in these cases. His solutions, however, are correct *only* for $\eta = 0$, i.e., when $U'(0) = \infty$. For $\eta = 0$, these solutions are:

$$V_\delta(x) = \frac{1}{\delta}\left[\frac{1-\delta}{\beta - r\delta - \gamma\delta/(1-\delta)}\right]^{1-\delta} x^\delta, \qquad x \geq 0, \tag{2.18}$$
$$V_0(x) = (1/\beta)\log \beta x + \frac{(r - \beta + \gamma)}{\beta^2}, \qquad x \geq 0, \tag{2.19}$$

for utility functions (2.16) and (2.17), respectively. By (2.11) and (2.12), we have the optimal investment and consumption policies,

$$\pi(x) = \frac{\alpha - r}{(1-\delta)\sigma^2} \quad \text{and} \quad c(x) = \frac{1}{1-\delta}\left(\beta - r\delta - \frac{\gamma\delta}{1-\delta}\right)x. \tag{2.20}$$

2.5 Bankruptcy with delayed recovery

Sethi and Taksar [33] introduced a model of nonterminal bankruptcy that is equivalent to the KLSS model. In this model, an agent, upon going bankrupt, may recover from it after a temporary but random sojourn in bankruptcy. Such recovery may be brought about in a number of ways, e.g., the individual may generate an innovative idea having commercial value. The rate of such recovery reflects essentially his innate ability or resourcefulness. However, such a recovery is not instantaneous. The individual must stay in the bankruptcy state for a positive amount of time and during this time, his consumption rate must be zero. This type of bankruptcy can be modelled by a continuous diffusion process with a delayed reflection.

The wealth equation changes to

$$\begin{aligned} dx(t) &= [(\alpha - r)\pi(t)x(t) + rx(t) - c(t)]1_{x(t)>0}dt \\ &\quad + \mu 1_{x(t)=0}\, dt + x(t)\pi(t)\sigma\, dw(t), \qquad x(0) = x. \end{aligned} \tag{2.21}$$

The equation shows that the recovery rate μ can be viewed as the rate of wealth accumulation during the time when $x(t) = 0$; this permits the investor to leave the bankruptcy state.

Sethi and Taksar [33] show that for every recovery rate μ, there is a bankruptcy utility P that makes their model equivalent to the KLSS model, and vice versa.

In addition to providing an alternative model of bankruptcy, the nonterminal bankruptcy may be a way towards an eventual development of an equilibrium model that incorporates bankruptcy. Further discussion in this regard is deferred to Section 5.

2.6 Analysis of the risk-aversion behavior

While KLSS had obtained an explicit solution of the problem, the specification of the value function was still too complicated to examine the implied risk-aversion behavior in detail. The analysis was made possible by yet another change of variable introduced by Presman and Sethi [23]. They defined a variable equal to the logarithm of the inverse of the marginal utility of wealth. This allowed them to obtain a linear second-order differential equation in wealth as a function of the new variable, and whose solution can be obtained in a parametric form with the parameter standing for the utility of bankruptcy. In other words, given the bankruptcy utility P, there is a unique choice of this parameter that makes the solution of the differential equation correspond exactly to the value function. Thus, unlike in KLSS, it unifies the cases in which the optimal solution may or may not involve consumption at the boundary. Furthermore, it extends the KLSS analysis to utility functions that need only to be continuously differentiable rather than thrice so as assumed in KLSS.

Presman and Sethi [23] studied the Pratt-Arrow risk-aversion measures, namely the coefficient of the absolute risk aversion,

$$l_V(x) = -\frac{d\ln V'(x)}{dx} = -\frac{V''(x)}{V'(x)}, \tag{2.22}$$

and the coefficient of the relative or proportional risk aversion,

$$L_V(x) = -\frac{d\ln V'(x)}{d\ln x} = xl_V(x), \tag{2.23}$$

with respect to the value function $V(x)$ denoting the derived utility associated with the wealth level x. Note for later discussion purposes that (2.22) also defines the coefficient $l_U(c)$ associated with the consumption utility $U(c)$.

Merton [19] obtained some results relating the nature of the value function to the nature of the utility function for a consumption assumed to be of HARA type. When $\eta = 0$ (i.e., when $U'(0) = \infty$) and $P \leq U(0)/\beta$, the value function of the problem is also of HARA type with the same parameter as the one for the HARA utility of consumption used in the problem. Thus, the coefficient of absolute risk aversion decreases with wealth, while that of relative risk aversion is constant with value $(1 - \delta)$.

Merton's results obtained for the HARA case are not correct for $\eta > 0$ or $P > U(0)/\beta$. In these cases, Presman and Sethi [23] show that the agent's value function is no longer of HARA type; while Merton [21] recognizes the errors in Merton [19] as pointed out by Sethi and Taksar [32], he does not update the risk-aversion implications of the corrected solutions.

With regards to an agent's relative risk aversion, first we note that $L_U' > 0$ for $U(c)$ specified in (2.16) and (2.17) with $\eta > 0$. The agent's relative risk aversion increases with wealth provided $\eta > 0$ or $P > U(0)/\beta$. In other words, while not of HARA type, the value function inherits the qualitative behavior from the HARA utility of consumption used in the problem. However, for $\eta > 0$ and $P \leq U(0)/\beta$, the inheritance holds only at higher wealth levels, while at lower wealth levels, the agent's relative risk aversion remains constant.

The agent's absolute risk aversion behavior is more complicated for $\eta > 0$ or $P > U(0)/\beta$. If δ is sufficiently large, for which it is necessary that $\beta + \gamma - r > 0$, then absolute risk aversion decreases with wealth. For smaller values of δ and $\beta + \gamma - r \geq 0$ however, the absolute risk aversion decreases with wealth if the bankruptcy payment is sufficiently low; otherwise the risk aversion increases at lower levels of wealth, while it decreases at higher levels of wealth. Furthermore, if $\beta+\gamma-r < 0$, then for every δ and every $P > U(0)/\beta$, the absolute risk aversion increases at lower levels of wealth, while it decreases at higher levels of wealth.

From the above discussion, one may draw the following general conclusion regarding the risk aversion behavior in the HARA case with $\eta > 0$.

At higher wealth levels, the agent's absolute (relative) risk-aversion decreases (increases) with wealth. This qualitative behavior at high wealth levels is inherited from the agent's HARA type consumption utility, as the agents seem quite immune from the bankruptcy payment parameter P. Of course, what is considered to be a high wealth level itself may depend on P.

At lower wealth levels, the agent is no longer immune from the amount of payment at bankruptcy. His behavior at these wealth levels is somewhat complicated, and it results from the interaction of his consumption utility, the bankruptcy payment, and the relationship of his risk-adjusted discount and the risk-free rate of return; see Presman and Sethi [23] for details.

To describe the risk-aversion behavior with general concave utility functions, the situation is far more complex. The most surprising observation is that while the sign of the derivative of the coefficient of local risk-aversion depends on the entire utility function, it is nevertheless explicitly independent of U'' and U''' or even their existence. Both the absolute and relative risk aversions decrease as the bankruptcy payment P increases. Also derived for all values of P are some necessary and sufficient conditions for the absolute risk aversion to be decreasing and the relative risk aversion to be increasing as wealth increases. Furthermore, the relative risk aversion increases with wealth in the neighborhood of zero wealth. Moreover, if there exists an interval of zero consumption (which happens when $U(0)/\beta \leq P < P^*$), then the relative risk aversion increases with wealth in this interval. In the neighborhood of zero wealth and in the interval of zero consumption, the absolute risk aversion increases (decreases) with wealth accordingly as $\beta + \gamma - r < 0 (> 0)$ for $P < P^*$.

Presman and Sethi [23] also show that if $\beta+\gamma-r \leq 0$, then either the absolute risk aversion increases with wealth for all P, or for each wealth level there exists a bankruptcy payment $P(x)$ such that at x the risk aversion is decreasing for payments smaller than $P(x)$ and increasing for payments larger than $P(x)$.

Finally, contrary to the intuitive belief that the absolute risk aversion is non-increasing as wealth approaches infinity, the limiting behavior at infinity is much more complex.

3. Positive subsistence consumption

Sethi, Taksar, and Presman [35] provided an explicit specification of the optimal policies in a general consumption-investment problem of a single agent with subsistence consumption and bankruptcy. In doing so, they used the methods developed in KLSS and Presman and Sethi [23]. See also Presman and Sethi [24, 25, 26].

Cadenillas and Sethi [1] introduce random market parameters in the models of KLSS and Sethi, Taksar and Presman [35]; see also Karatzas [8]. Thus, their model also extends the models of Karatzas, Lehoczky, and Shreve [10] and Cox and Huang [2] to allow for explicit consideration of bankruptcy.

4. Borrowing and shortselling constraints

In this section we briefly discuss models with constrained borrowing and short-selling. These constraints give rise to value functions that may not be concave. Observe that in regions where value functions are convex, the agent will put all his investment in the riskiest asset available.

The model developed by Lehoczky, Sethi, and Shreve [14] can be related to the model of Sethi, Taksar and Presman [35] as follows. Impose an additional constraint that disallows short-selling, i.e., $0 \leq \pi_t \leq 1$, and set $\alpha = r$. While $0 \leq \pi_t \leq 1$ appears to permit no borrowing, a reformulation of the problem transforms it into a model that allows unlimited borrowing. Furthermore, $\alpha = r$ is imposed to simplify the solution and to focus entirely on the distortions caused by consumption constraints and bankruptcy, and thus eliminate other factors which might induce risk-taking behavior. See also Sethi, Gordon and Ingham [31].

Lehoczky, Sethi, and Shreve [15] have generalized their 1983 model by using the wealth dynamics,

$$x_t = x + \int_0^t r x_\tau \, d_\tau + S_t, \tag{4.1}$$

where S_t is a supermartingale with $S_0- = 0$, $-1 \leq t < 0$ and satisfying conditions DL. The condition allows decomposition of S_t into a martingale M_t and a cumulative consumption process C_t; see Karatzas and Shreve [13, pp. 24–25].

5. Open research problems and concluding remarks

We have reviewed the literature on consumption-investment problems that explicitly incorporate bankruptcy. This concluding section briefly refers to related research on consumption-investment problems that does not deal with the bankruptcy issue. This suggests some open research problems; see also Sethi [29, Chapter 16].

In all the papers discussed in this survey, there is no cost of buying and selling assets. Davis and Norman [3] and Shreve and Soner [34] have considered proportional transition costs in consumption-investment models with two assets and nonnegative consumption constraint. It would be interesting to incorporate a positive subsistence level and a bankruptcy utility in these models. Another extension would be to include fixed transaction costs; such a cost has not been considered in the consumption-investment context.

Karatzas, Lehoczky, Shreve, and Xu [12] and He and Pearson [7] have considered incomplete markets. One would like to incorporate such markets in consumption-investment models with bankruptcy and a subsistence requirement.

Finally, Karatzas, Lehoczky, and Shreve [11] have developed equilibrium models with many agents consuming and trading securities with one another over time. In these models, consumption utilities are chosen so that agents do not go bankrupt. This way if one begins with n agents, one stays with n agents throughout the horizon. Thus, there is no easy way to see how these models can be extended to allow for bankruptcy. Sethi and Taksar [33] introduced a concept of nonterminal bankruptcy as discussed in Section 2.5. This allows agents to stay in the system and may facilitate the eventual development of an equilibrium model with bankruptcy. Several important open research problems flow from these considerations.

The Sethi-Taksar nonterminal bankruptcy model needs to be extended to allow for random coefficients and subsistence consumption as in Cadenillas and Sethi [1]. It is not clear how to prove the equivalence between the terminal and nonterminal bankruptcies in the more general setup.

The Cadenillas-Sethi model treats an almost surely finite horizon agent. In addition, it deals with only nearly optimal policies. One needs to extend the model to allow for infinite horizon and to obtain optimal policies. If this problem is solved, and if it can be shown to be equivalent to a model with nonterminal bankruptcy as mentioned above, then we would have a single agent model as a starting point in the development of an equilibrium model with bankruptcy.

Another important consideration is how to provide for the bankruptcy value P if it consists of welfare or the subsistence consumption while in the state of bankruptcy. This would call for a different kind of agent, called the government, who must collect taxes and provide for welfare to agents who are in the bankruptcy state.

We hope that work will be carried out in addressing the open research problems described above, and that a suitable equilibrium model that allow for bankruptcy and subsistence consumption will eventually be developed.

References

[1] A. Cadenillas and S.P. Sethi (1997), "The Consumption Investment Problem with Subsistence Consumption, Bankruptcy, and Random Market Coefficients," *J. Opt. Theory & Appl.*, **93**, (2).

[2] J.C. Cox and C.F. Huang (1989), "Optimal Consumption and Portfolio Policies when Asset Prices Follow a Diffusion Process," *J. Econ. Theory*, **49**, 33–83.

[3] M.H.A. Davis and A. Norman (1990), "Portfolio Selection with Transaction Costs." *Math. Oper. Research*, **15**, 676–713.

[4] M.J. Gordon and S.P. Sethi (1997), "A Contribution to the Micro Foundation for Keynesian Macroeconomic Models," Chapter 11 in Sethi (1997), 217–244.

[5] M.J. Gordon and S.P. Sethi (1998), "Consumption and Investment When Bankruptcy is not a Fate Worse than Death," *Method, Theory and Policy in Keynes*, Philip Arestis, Editor, Edward Elgar, Cheltenham, U.K., 88–108.

[6] N.H. Hakansson (1969), "Optimal Investment and Consumption Strategies Under Risk, An Uncertain Lifetime, and Insurance," *International Econ. Review* **10**, 443–466.

[7] H. He and N. Pearson (1991), "Consumption and Portfolio Policies with Incomplete Markets and Short-Sale Constraints: The Infinite Dimensional Case." *J. Econ. Theory*, **54**, 259–304.

[8] I. Karatzas (1989), "Optimization Problems in the Theory of Continuous Trading." *SIAM J. Control Optim.* **27**, 1221–1259.

[9] I. Karatzas, J. Lehoczky, S.P. Sethi, and S. Shreve (1986), "Explicit Solution of a General Consumption/Investment Problem," *Math. Oper. Research*, **11**, No. 2, 261–294; reprinted as Chapter 2 in Sethi (1997).

[10] I. Karatzas, J. Lehoczky and S. Shreve (1987), "Optimal Portfolio and Consumption Decisions for a Small Investor on a Finite Horizon." *SIAM J. Control and Optim*, **25**, 1557–1586.

[11] I. Karatzas, J. Lehoczky and S. Shreve (1990), "Existence and Uniqueness of Multi-Agent Equilibrium in a Stochastic, Dynamic Consumption/Investment Model." *Math. Oper. Research*, **15**, 80–128.

[12] I. Karatzas, J. Lehoczky, S. Shreve and G. Xu (1991), "Martingale and Duality Methods for Utility Maximization in an Incomplete Market." *SIAM J. Control Optim.*, **29**, 702–730.

[13] I. Karatzas and S. Shreve (1988), "Brownian Motion and Stochastic Calculus." Springer-Verlag, New York.

[14] J. Lehoczky, S.P. Sethi and S. Shreve (1983), "Optimal Consumption and Investment Policies Allowing Consumption Constraints and Bankruptcy," *Math. Oper. Research*, **8**, 613–636; an unabridged version as Chapter 14 in Sethi (1997), 303–378.

[15] J. Lehoczky, S.P. Sethi and S. Shreve (1985), "A Martingale Formulation for Optimal Consumption/Investment Decision Making," *Optimal Control Theory and Economic Analysis 2*, G. Feichtinger, Editor, North-Holland; an unabridged version as Chapter 15 in Sethi (1997), 379–406.

[16] S.A. Lippman, J.J. McCall, W.L. Winston (1980), "Constant Absolute Risk Aversion, Bankruptcy, and Wealth-Dependent Decisions," *J. of Business,* **53** No. 3, Part 1, 285–296.

[17] S.P. Mason (1981), "Consumption and Investment Incentives Associated with Welfare Programs," Working Paper, No. 79–34, Graduate School of Business Administration, Harvard University, Boston, MA.

[18] R.C. Merton (1969), "Lifetime Portfolio Selection Under Uncertainty: The Continuous-Time Case," *Rev. Econ. Statist.*, **51**, 247–257.

[19] R.C. Merton (1971), "Optimum Consumption and Portfolio Rules in a Continuous Time Model," *J. Econ. Theory*, **3**, 373–413.

[20] R.C. Merton (1973), Erratum, *J. Econ. Theory*, **6**, 213–214.

[21] R.C. Merton (1990), *Continuous-Time Finance*, Basil Blackwell, Cambridge, MA.

[22] E.S. Phelps (1962), "The Accumulation of Risky Capital: A Sequential Utility Analysis", *Econometrica*, **30**, 729–743.

[23] E. Presman and S.P. Sethi (1991), "Risk-Aversion Behavior in Consumption/Investment Problem," *Mathematical Finance,* **1**, No. 1, 100–124; Erratum, **1**, P. 86.

[24] E. Presman and S.P. Sethi (1996), "Distribution of Bankruptcy Time in a Consumption/Portfolio Problem," *J. Econ. Dyn. and Control*, **20**, 471–477.

[25] E. Presman and S.P. Sethi (1997a), "Risk-Aversion Behavior in Consumption/Investment Problems with Subsistence Consumption," Chapter 8 in Sethi (1997), 155–184.

[26] E. Presman and S.P. Sethi (1997b), "Consumption Behavior in Investment/Consumption Problems," Chapter 9 in Sethi (1997), 185–205.

[27] E. Presman and S.P. Sethi (1997c), "Equivalence of Objective Functionals in Infinite Horizon and Random Horizon Problems," Chapter 10 in Sethi (1997), 207–216.

[28] P.A. Samuelson (1969), "Lifetime Portfolio Selection by Dynamic Stochastic Programming," *Rev. Econ. Statist.*, **51**, 239–246.

[29] S.P. Sethi (1997), *Optimal Consumption and Investment with Bankruptcy*, Kluwer, Norwell, MA.

[30] S.P. Sethi (1998), "Optimal Consumption-Investment Decisions Allowing for Bankruptcy: A Survey" in *Worldwide Asset and Liability Modeling*, W.T. Ziemba and J.M. Mulvey (Eds.), Cambridge University Press, Cambridge, UK, 387–426.

[31] S.P. Sethi, M.J. Gordon and B. Ingham (1979), "Optimal Dynamic Consumption and Portfolio Planning in a Welfare State," *Portfolio Theory, 25 Years After,* E.J. Elton and M.J. Gruber (Eds.), *TIMS Studies in the Management Sciences II*, 179–196; reprinted as Chapter 13 in Sethi (1997).

[32] S.P. Sethi and M. Taksar (1988), A Note on Merton's "Optimum Consumption and Portfolio Rules in a Continuous Time Model," *J. Econ. Theory*, **46**, 395–401; reprinted as Chapter 3 in Sethi (1997).

[33] S.P. Sethi and M. Taksar (1992), "Optimal Consumption and Investment Policies Modeled by a Diffusion Process with Delayed Reflections," *J. Opt. Theory and Appl.*, **74**, 333–346.

[34] S. Shreve and M. Soner (1994), "Optimal Investment and Consumption with Transaction Costs." *Annals of Appl. Probab.,* **4**, 609–692.

[35] S.P. Sethi, M. Taksar, and E. Presman (1992), "Explicit Solution of a General Consumption/Portfolio Problem with Subsistence Consumption and Bankruptcy," *J. Econ. Dyn. and Control,* **16** 747–768; Erratum, **19**, (1995), 1297–1298.

Chapter 24

THE HEDGING STRATEGY OF AN ASIAN OPTION

Zhaojun Yang
Changsha Railway University,
Changsha,410075
Hunan Taxation College,
Changsha,410075

Jiezhong Zou
Changsha Railway University,
Changsha,410075

Abstract By a generalized Clark formula, this paper provides a hedging strategy for the Asian option calculated with geometric averaging. The hedging strategy is uncomplicated and easy to operate.

Keywords: a generalized Clark formula, Asian option, a hedging strategy.

AMS Subject Classification (1991): 93E20, 90A09

1. Introduction

In financial economics, it is critically important to price options and derive the associated hedging strategies. Many results have been obtained in the options pricing, however, they only ascertain the existence of the hedging strategies and barely deal with how to construct them. Asian options are the common claims which depend on the mean prices of the basic assets in their life. Therefore, it is almost impossible for investors to change their options at will by manipulating the assets' prices in the near maturity date. As a result, Asian options avoid flaws of the European options in this respect. Asian options fall into two types, the

Z. Hou et al. (eds.), Markov Processes and Controlled Markov Chains, 389–395.

arithmetic type and the geometric type. The problems about their pricing are basically resolved [3, 6, 7], however, we have not found studies on constructing their hedging strategies. This paper establishes a hedging strategy for geometric Asian options by means of the generalized Clark formula. The strategy is uncomplicated and easy to operate. For the case of arithmetic Asian option, we refer to article [8].

2. The model and conclusions

Consider a complete probability space (Ω, F, P) and a standard 1-dimensional Brownian motion W=(W(t)), $0 \leq t \leq T$ defined on it. We shall denote by $\{F_t\}$ the P-augmentation of the natural filtration

$$F_t^w = \sigma(W(s); 0 \leq s \leq t), \quad 0 \leq t \leq T.$$

There are two assets on the market. One of the assets is risk-free bond with a constant deterministic interest rate r. The other is risky security (stock) on the space (Ω, F, P) with price process $S = (S(t))$, $0 \leq t \leq T$. The dynamics of the price process is determined by SDE

$$dS(t) = \mu S(t)dt + \sigma S(t)dW(t), \quad S(0) = S_0, \quad 0 \leq t \leq T. \tag{2.1}$$

According to the Girsanov theorem we may assume $\mu = r$ without loss of generality. Consequently Equation (2.1) can be equivalently to the following equation,

$$S(t) = S_0 \exp\left\{\sigma W(t) - \sigma^2 t/2 + rt\right\}, \quad 0 \leq t \leq T, \tag{2.2}$$

where the constants $r, \sigma, T > 0$.

Denote the investor's wealth and the amount invested in the stock at time t by $V(t)$ and $\pi(t)$ respectively. Assume the strategy to be self-financed, then the amount invested in the bond is $V(t) - \pi(t)$. It is easy to infer the process $V = (V(t))$ satisfies

$$dV(t) = rV(t)\,dt + \pi(t)\sigma\,dW(t), \quad V(0) = V_0, \quad 0 \leq t \leq T. \tag{2.3}$$

Definition 2.1 *A portfolio process $\pi = \{\pi(t), F_t, 0 \leq t \leq T\}$ is a measurable, adapted process for which*

$$\int_0^T \pi^2(t)\,dt < \infty \quad a.s. \tag{2.4}$$

Condition (2.4) ensures SDE (2.3) has a unique strong solution, which satisfies

$$V(t)\exp\{-rt\} = V_0 + \int_0^t \exp\{-ru\}\pi(u)\sigma\,dW(u), \quad 0 \leq t \leq T. \tag{2.5}$$

The payoff at maturity from a geometric Asian option is

$$f_T = \left[\exp\left\{ \frac{\int_0^T \log S(t)\,dt}{T} \right\} - q \right]^+, \tag{2.6}$$

where the constant $q > 0$. The hedging strategy for this option refers to a self-financing portfolio satisfying admissible condition [1], by which the investor's wealth determined in Equation (2.5) at time T is equal to f_T almost surely. Assuming that there exists no arbitrage opportunity , the article [1] proved

$$V_0 = E[f_T \exp\{-rt\}]. \tag{2.7}$$

The value V_0 is the fair price of the option f_T at time 0 which is not difficult to calculate [7]. We set that

$$A = \int_0^t W(u)\,du + W(t)(T-t) + \frac{\left(r - \sigma^2/2\right) T^2}{2\sigma} - \frac{T \log(q/s_0)}{\sigma},$$

and $\Phi(\bullet)$ is the standard normal distribution function. The main result of this paper is as follows.

Theorem 2.1 *The hedging strategy of option f_T is that the amount invested in the stock satisfies*

$$\pi(t) = \frac{V(t)(T-t)}{T} + \frac{q \exp\{-r(T-t)\}\Phi\left(\sqrt{3}A/(T-t)^{3/2}\right)(T-t)}{T}. \tag{2.8}$$

Consequently, the amount invested in the bond is $V(t) - \pi(t)$, where process $V = (V(t))$, $0 \le t \le T$ is the value process associated with the option, i.e., the investor's wealth process and

$$V(0) = V_0 = E\left[f_T \exp\{-rT\}\right]$$

Remark 2.1 *When we calculate A, according to identity* (2.2), *value $\int_0^t W(u)\,du$ is equal to*

$$\int_0^t \frac{\log(s(u)/s_0) + \sigma^2 u/2 - ru}{\sigma}\,du = \frac{4\int_0^t \log(s(u)/s_0)\,du + \left(\sigma^2 - 2r\right)t^2}{(4\sigma)}$$

Remark 2.2 $V(t) = E(f_T \exp\{-r(T-t)\}|F_t)$, *in addition, the amount invested in the stock is always positive.*

3. Proof of theorem 2.2

Lemma 3.1 $E(f_T^2) < \infty$

Proof It is enough to prove $E\left[\exp\left\{2\int_0^T \log S(t)\,dt/T\right\}\right] < \infty$. Considering that the sample paths of the process $S = (S(t))$ are continuous almost surely, for the partition $\prod = \{t_0, t_1, \ldots, t_n\}$ of $[0,T]$, with $0 = t_0 < t_1 < \cdots < t_n = T$, we conclude that

$$\lim_{\lambda_n \to 0} \left[\prod_{i=1}^{n} S(t_i)\right]^{1/n} = \lim_{\lambda_n \to 0} \exp\left\{\frac{\sum_{i=1}^{n} \log S(t_i)}{n}\right\} = \exp\left\{\frac{\int_0^T \log(S(t))\,dt}{T}\right\} \quad \text{a.s.}$$

$$\lim_{\lambda_n \to 0} \frac{\sum_{i=1}^{n} S(t_i)}{n} = \frac{\int_0^T S(t)\,dt}{T} \quad \text{a.s.}$$

where $\lambda_n = \max_{1 \le k \le n} |t_k - t_{k-1}|$ is the mesh of the partition. Because the arithmetic mean value is greater than or equal to the geometric mean value, we have

$$0 \le \exp\left\{\frac{\int_0^T \log S(t)\,dt}{T}\right\} \le \frac{\int_0^T S(t)\,dt}{T} \quad \text{a.s.}$$

then the lemma follows from (2.2) and the Hölder inequality and Fubini theorem. ∎

To introduce the generalized Clark formula, we first define Banach space $D_{p,1}$ and its gradient operator D [2, 4]. Consider a smooth functional, i.e., a function $F : \Omega \to R$ of the form $F(\omega) = f(W(t_1, \omega), \ldots, W(t_n, \omega))$ for some $n \in N$, $(t_1, \ldots, t_n) \in [0,T]^n$ and some element f in the space $C_b^\infty(R^n)$ of the functions with continuous and bounded derivatives of every order. The gradient DF(ω) of the smooth functional F is defined as the $L^2([0,T])$ - valued random variable showed as follows:

$$D_t F(\omega) = \sum_{j=1}^{n} \frac{\partial}{\partial x_j} f(W(t_1, \omega), \ldots, W(t_n, \omega)) 1_{[0,t_j]}(t), \quad 0 \le t \le T.$$

For every $p \ge 1$, introduce the norm $\| \bullet \|_{p,1}$ on the set S of smooth functionals by the formula

$$\|F\|_{p,1} = E\big[|F|^p + \|DF\|^p\big]^{1/p},$$

where $\| \bullet \|$ denote the $L^2([0,T])$ norm. We denote by $D_{p,1}$ the Banach space which is the completion of S under $\| \bullet \|_{p,1}$.

Lemma 3.2 *For every $F \in D_{1,1}$, we have*

$$F = E(F) + \int_0^T E(D_t F | F_t)\, dW(t) \qquad a.s. \tag{3.1}$$

Equation (3.1) is the generalized Clark formula whose proof can be found in article [2].

Lemma 3.3 *Let $F = (F_1, \dots, F_k) \in (D_{1,1})^k$. Let $\Phi \in C^1(R^k)$ be a real-valued function and assume that*

$$E\left\{ |\Phi(F)| + \left\| \sum \frac{\partial \Phi}{\partial x_i}(F) DF_i \right\| \right\} < \infty.$$

Then $\Phi(F) \in D_{1,1}$ and $D\Phi(F) = \sum (\partial \Phi / \partial x_i)(F) DF_i$.

For its proof, we refer to article [5].

Lemma 3.4 *Random Variable $\int_0^t W(u)\, du$ $(t \geq 0)$ is normally distributed with mean zero and variance $t^3/3$.*

Proof Applying Itô's rule, we obtain

$$d(uW(u)) = W(u)\, du + u\, dW(u).$$

It is easy to derive

$$\int_0^t W(u)\, du = \int_0^t (t-u)\, dW(u).$$

The lemma is proved. ∎

Proof of Theorem 2.1.
Let

$$g_T = \exp\left\{ \frac{\int_0^T [\log S_0 + \sigma W(t) - \sigma^2 t/2 + rt]\, dt}{T} \right\} - q, \text{ and}$$

function $\Phi(x) = x^+$, then $f_T = \Phi(g_T)$. Obviously, we have

$$D_t \left[\frac{\int_0^T (\log S_0 + \sigma W(u) - \sigma^2 u/2 + ru)\, du}{T} \right] = \frac{(T-t)\sigma}{T}.$$

Consequently, the proof of Lemma 3.1 and of Lemma 3.3 imply $g_T \in D_{1,1}$, and $D_t g_T = g_T(T-t)\sigma/T + q\sigma(T-t)/T$, $0 \le t \le T$. Define C^∞ function $\rho(x) = C1_{(0,2)}(x)\exp\{1/[(x-1)^2-1]\}$, where C is constant satisfying $\int_R \rho(x)\,dx = 1$. Let $\rho_n(x) = n\rho(nx)$, $\Phi_n(x) = \int_R \rho_n(x-y)\Phi(y)\,dy$. Thus, we have $\Phi_n(x) = \int_R \rho(z)\Phi(x-z/n)\,dz$, and $0 \le \Phi_n(x) \le \Phi(x)$, $\lim_{n\to\infty}\Phi_n(x) = \Phi(x)$, $0 \le \Phi_n'(x) \le 1$, $\lim_{n\to\infty}\Phi_n'(x) = D^-\Phi(x)$ (left-derivative). So we conclude $0 \le \Phi_n(g_T) \le \Phi(g_T) = f_T$, $|\Phi_n'(g_T)D_t g_T| \le f_T\sigma(T-t)/T + q\sigma(T-t)/T$. It follows from Lemma 3.1 and Lemma 3.3 that $\Phi_n(g_T) \in D_{1,1}$ and $D_t\Phi_n(g_T) = \Phi_n'(g_T)D_t g_T$, $0 \le t \le T$. We see that $\lim_{n\to\infty}\Phi_n(g_T) = f_T$ a.s., $\lim_{n\to\infty} D_t\Phi_n(g_T) = D^-\Phi(g_T)D_t g_T$ a.s.. Therefore,

$$\lim_{n\to\infty} E\{|\Phi_n(g_T) - f_T| + \|D\Phi_n(g_T) - D^-\Phi(g_T)Dg_T\|\} = 0$$

by dominated convergence theorem. Because D is a closed operator on $D_{1,1}$, we establish $f_T \in D_{1,1}$ and

$$\begin{aligned} D_t f_T &= D^-\Phi(g_T)D_t g_T \\ &= \frac{\sigma\big(f_T + D^-\Phi(g_T)q\big)(T-t)}{T}, \qquad 0 \le t \le T. \end{aligned} \tag{3.2}$$

By virtue of identity (2.5) and $V(T) = f_T$, the hedging strategy π satisfies

$$f_T \exp\{-rT\} = E\big(f_T\exp\{-rT\}\big) + \int_0^T \exp\{-ru\}\pi(u)\sigma\,dW(u) \tag{3.3}$$

and corresponding to this strategy, the wealth process, i.e., the value process of option f_T [1], V satisfies

$$V(t)\exp\{-rt\} = E\big(f_T\exp\{-rT\}|F_t\big), \qquad 0 \le t \le T. \tag{3.4}$$

On the other hand, since $f_T\exp\{-rT\} \in D_{1,1}$, it follows from Lemma 3.2 that

$$f_T\exp\{-rT\} = E\big(f_T\exp\{-rT\}\big) + \int_0^T E\big[D_t(f_T exp\{-rT\})|F_t\big]\,dW(t). \tag{3.5}$$

Comparing (3.3) with (3.5), we obtain

$$\pi(t) = \frac{\exp\{rt\}E\big[D_t(f_T\exp\{-rT\})|F_t\big]}{\sigma} \qquad \text{a.s.} \tag{3.6}$$

Applying (3.2) and (3.4), we have then

$$\begin{aligned} &E\left[D_t(f_T \exp\{-rT\})|F_t\right] \\ &= \frac{\sigma(T-t)\left[E(f_T \exp\{-rT\}|F_t) + q\exp\{-rT\}E(D^-\Phi(g_T)|F_t)\right]}{T} \\ &= \frac{\sigma(T-t)\left[V(t)\exp\{-rt\} + q\exp\{-rT\}P(g_T > 0|F_t)\right]}{T}. \end{aligned}$$

By solving inequalities and utilizing properties of Brownian motion and Lemma 3.4, it can be concluded that

$$\begin{aligned} P(g_T > 0|F_t) &= P\left(\int_0^{T-t} W(u)du < A\right) \\ &= \Phi\left(\sqrt{3}A/(T-t)^{3/2}\right). \end{aligned} \tag{3.7}$$

By summing up identities (3.6), (3.6) and (3.7), it is easy to derive (2.8). Because the hedging strategy is self-financed, the remaining part of the theorem is obvious. ■

References

[1] Karatzas, I., Shreve, S.E., 1988. Brownian motion and stochastic calculus. Springer, New York.

[2] Karatzas, I., Ocone, D.l., Li, J., 1991. An extension of Clark's formula. Stochastics and stochastic reports 37,127–131.

[3] Milevsky, M.A., Posner, S.E., 1998. Asian options, the sum of lognormals, and the reciprocal Gama distribution. Journal of Financial and Quantitative Analysis. 33, 409–422.

[4] Nualart , D., Pardoux, E., 1988. Stochastic calculus with anticipating integrands. Probability Theory and Related Fields 78, 535–581.

[5] Ocone, D.L., Karatzas, I., 1991. A generalized Clark representation formula, with application to optimal portfolios, Stochastics and Stochastic Reports 34, 187–220.

[6] Rogers, L.C.G., Shi,Z., 1995. The value of an Asian option. Journal of Applied Probability 32, 1077–1088.

[7] Wilmott, P., Dewynne, j., Howison, S., 1993. Option pricing:mathematical models and computation. Oxford Financial Press,UK.

[8] Yang, Z.J., Zou, J.Z., Hou, Z.T., The hedging strategy of arithmetic Asian option. Journal of Operations Research, to appear.(in Chinese).

Chapter 25

THE PRICING OF OPTIONS TO EXCHANGE ONE ASSET FOR ANOTHER

Chao Chen
Research Center of Mathematical Finance
Changsha Railway University,
Changsha 410075, China

Jiezhong Zou
Research Center of Mathematical Finance
Changsha Railway University,
Changsha 410075, China

Zhenting Hou
Research Center of Mathematical Finance
Changsha Railway University,
Changsha 410075, China

Abstract This paper deals with the pricing of options to exchange one asset for another. Under the assumption that the asset price processes are jump-diffusion processes, it deduces the partial equation that the option prices must satisfy, and then obtains the pricing formula of options.

Keywords: Option pricing; Underlying asset; Jump-diffusion process.

1. Introduction

Most of the recent literature on continuous finance has been based on an assumption of continuous price processes. The validity of the assumption depends on whether or not the change of the asset price satisfies a kind of local Markov property, i.e., in a short interval of time,

Z. Hou et al. (eds.), Markov Processes and Controlled Markov Chains, 397–404.

the asset price can only change by a small amount. In fact event studies suggest that certain public announcements of information are associated with jumps in asset prices.

Marhrabe obtains the pricing formula of options to exchange one asset for another when the price processes of two assets are geometric Brownian motions. This paper deals with the pricing of options to exchange one asset for another in the more-general case when the price processes are jump-diffusion processes.

2. The financial market

We suppose that there are three assets being traded continuously. One of these is a risk-free asset, with price $s_0(t)$ given by

$$\frac{ds_0}{s_0} = rdt \qquad\qquad s_0(0) = 1 \tag{2.1}$$

where r is the instantaneous rate of interest. The other two assets are risky assets, subject to the uncertainty in the market. The price of the ith asset $s_i(t)$ $(i = 1, 2)$ is governed by a stochastic differential equation

$$\frac{ds_i}{s_i} = (u_i - \lambda k_i)dt + \sigma_i dB_i + x_i dN, \qquad i = 1, 2 \tag{2.2}$$

where u_i $(i = 1, 2)$ is the instantaneous expected return on the ith asset; $\sigma_i^2(i = 1, 2)$ is the instantaneous variance of the return of the ith asset, conditional on no arrivals of important new information; $B_i(t)$ $(i = 1, 2)$ are standard Brownian motions, with a correlation coefficient ρ; N(t) is an Possion process with parameter λ; $k_i = \varepsilon(x_i)$, where x_i is the random variable percentage change in the ith asset price if the jump occurs; and ε is the expectation operator over the random variable x_i.

We suppose that u_i, k_i, σ_i and λ are constants (i=1,2), the solution to the equation (2.2) is

$$s_i(t) = s_i(0)exp\left(\left(u_i - \lambda k_i - \frac{\sigma_i^2}{2}\right)t + \sigma_i B_i(t)\right)\prod_{j=1}^{N(t)}(1 + x_{ij}) \tag{2.3}$$

where x_{ij} $(i = 1, 2,)$ are independent and identical distributions.

3. The option price dynamics

Suppose that the option price, w, can be written as a twice - continuously differentiable function of the assets s_1, s_2 and time t: namely, $w(t) = F(s_1, s_2, t)$. The option return dynamics can be written in a similar form as

$$\frac{dw}{w} = (\mu_w - \lambda k_w)dt + \sigma_{1w} dB_1 + \sigma_{2w} dB_2 + x_w dN \tag{3.1}$$

where μ_w is the instantaneous expected return on the option; $(\sigma_{1w}, \sigma_{2w})$ is the volatility; $k_w = \varepsilon(x_w)$, where x_w is the random variable percentage change in the option price if a jump occurs.

Using Itô lemma for the continuous part and analogous lemma for the jump part, we have

$$\begin{aligned}
\mu_w &= \Big[\frac{1}{2}\sigma_1^2 s_1^2 F_{11} + \frac{1}{2}\sigma_2^2 s_2^2 F_{22} + \rho\sigma_1\sigma_2 s_1 s_2 F_{12} \\
&\quad + (\mu_1 - \lambda k_1)s_1 F_1 + (\mu_2 - \lambda k_2)s_2 F_2 + F_t \\
&\quad + \lambda\varepsilon(F(s_1(1+x_1), s_2(1+x_2), t) - F(s_1, s_2, t))\Big] \\
&\quad /F(s_1, s_2, t) && (3.2) \\
\sigma_{1w} &= \frac{\sigma_1 s_1 F_1(s_1, s_2, t)}{F(s_1, s_2, t)} && (3.3) \\
\sigma_{2w} &= \frac{\sigma_2 s_2 F_2(s_1, s_2, t)}{F(s_1, s_2, t)} && (3.4) \\
x_w &= \frac{F(s_1(1+x_1), s_2(1+x_2), t) - F(s_1, s_2, t)}{F(s_1, s_2, t)} && (3.5)
\end{aligned}$$

where subscripts on $F(s_1, s_2, t)$ denote partial derivatives.

Consider a portfolio strategy which holds the assets s_1, s_2 and the option w in proportions π_1, π_2 and π_3, where $\pi_1 + \pi_2 + \pi_3 = 1$. If p is the value of the portfolio, then the return dynamics on the portfolio can be written as

$$\frac{dp}{p} = (\mu_p - \lambda k_p)dt + \sigma_{1p} dB_1 + \sigma_{2p} dB_2 + x_p dN \tag{3.6}$$

where μ_p is the instantaneous expected return on the portfolio; $(\sigma_{1p}, \sigma 2p)$ is the volatility; $k_p = \varepsilon(x_p)$, where x_p is the random variable percentage change in the portfolio value if the jump occurs.

From (2.2) and (3.1), we have that

$$\begin{aligned}
\mu_p &= \pi_1\mu_1 + \pi_2\mu_2 + \pi_3\mu_w && (3.7) \\
\sigma_{1p} &= \pi_1\sigma_1 + \pi_3\sigma_{2w} && (3.8) \\
\sigma_{2p} &= \pi_2\sigma_2 + \pi_3\sigma_{2w} && (3.9) \\
x_p &= \pi_1 x_1 + \pi_2 x_2 \\
&\quad + \pi_3 \frac{[F(s_1(1+x_1), s_2(1+x_2), t) - F(s_1, s_2, t)]}{F(s_1, s_2, t)} && (3.10)
\end{aligned}$$

We select $\pi_1 = \pi_1^*$, $\pi_2 = \pi_2^*$ and $\pi_3 = \pi_3^*$, so that $\pi_1^*\sigma_1 + \pi_3^*\sigma_{2w} = 0$ and $\pi_2^*\sigma_2 + \pi_3^*\sigma_{2w} = 0$. Let p^* denote the value of the portfolio, then from (3.6), we have that

$$\frac{dp^*}{p^*} = (\mu_p^* - \lambda k_p^*)dt + x_p^* dN \tag{3.11}$$

We suppose that the jump component of the asset's return represent 'non-systematic' risk. If the Capital Asset Pricing Model holds, then the expected return on the portfolio must equal the riskless rate r. Therefore $\mu_p^* = r$. Then, we have that

$$\begin{cases} \pi_1^*\mu_1 + \pi_2^*\mu_2 + \pi_3^*\mu_w & = r \\ \pi_1^*\sigma_1 + \pi_3^*\sigma_{1w} & = 0 \\ \pi_2^*\sigma_2 + \pi_3^*\sigma_{2w} & = 0 \end{cases} \tag{3.12}$$

But, (3.12), (3.2)–(3.5), and $\pi_1^* + \pi_2^* + \pi_3^* = 1$ imply that F must satisfy the following differential equation

$$\begin{aligned} &\frac{1}{2}\sigma_1^2 s_1^2 F_{11} + \frac{1}{2}\sigma_2^2 s_2^2 F_{22} + \rho\sigma_1\sigma_2 s_1 s_2 F_{12} \\ &\quad +(r - \lambda k_1)s_1 F_1 + (r - \lambda k_2)s_2 F_2 - rF + F_t \\ &\quad +\lambda\varepsilon[F(s_1(1 + x_1), s_2(1 + x_2), t) - F(s_1, s_2, t))] = 0 \end{aligned} \tag{3.13}$$

If $\lambda = 0$ i.e., if there are no jumps, then F must satisfy

$$\begin{aligned} &\frac{1}{2}\sigma_1^2 s_1^2 F_{11} + \frac{1}{2}\sigma_2^2 s_2^2 F_{22} + \rho\sigma_1\sigma_2 s_1 s_2 F_{12} \\ &\quad +rs_1 F_1 + rs_2 F_2 - rF + F_t = 0 \end{aligned} \tag{3.14}$$

4. The option pricing formula

Let $F(s_1, s_2, t)$ be the value of European call option to exchange one asset for another, then $F(s_1, s_2, t)$ satisfies equation(3.13), and subject to the boundary conditions

$$F(s_1, 0, t) = 0 \tag{4.1}$$
$$F(s_1, s_2, T) = max\{s_2 - s_1, 0\} \tag{4.2}$$

where T is the expiration time.

Define $H(s_1, s_2, t)$ to be the pricing formula of options to exchange one asset for another for the no-jump case. Then H will satisfy equation (3.14) subject to the boundary conditions (4.1) and (4.2). From Marhrabe's paper, H can be written as

$$H(s_1, s_2, t) = s_2\Phi(d_1) - s_1\Phi(d_2) \tag{4.3}$$

where

$$\begin{aligned}
\Phi(y) &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{y} e^{-\frac{x^2}{2}}\,dx, \\
d_1 &= \frac{ln(s_2/s_1)+v^2(T-t)}{v\sqrt{T-t}}, \text{ and} \\
d_2 &= d_1 - v\sqrt{T-t}
\end{aligned}$$

where

$$v^2 = \sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2$$

Theorem 4.1 *Suppose $F(s_1, s_2, t)$ is the value of European call option to exchange one asset for another, then*

$$\begin{aligned}
&F(s_1, s_2, t) \\
&= \sum_{n=0}^{\infty} \frac{e^{-\lambda(T-t)}(\lambda(T-t))^n}{n!}\varepsilon_n\left\{H\left(s_1\prod_{j=1}^{n}(1+x_{1j})e^{-\lambda k_1(T-t)},\right.\right. \\
&\qquad\qquad \left.\left. s_2\prod_{j=1}^{n}(1+x_{2j})\,e^{-\lambda k_2(T-t)}, t\right)\right\} \qquad (4.4)
\end{aligned}$$

where ε_n is the expected operation over $\prod_{j=1}^{n}(1+x_{1j})$ and $\prod_{i=1}^{n}(1+x_{2j})$.

5. Proof of the theorem

Let

$$\begin{aligned}
p_n(t) &= \frac{e^{-\lambda(T-t)}(\lambda(T-t))^n}{n!}, \\
v_n &= s_1\prod_{j=1}^{n}(1+x_{1j})e^{-\lambda k_1(T-t)}, \text{ and} \\
u_n &= s_2\prod_{j=1}^{n}(1+x_{2j})e^{-\lambda k_2(T-t)}.
\end{aligned}$$

Then

$$s_1F_1 = \sum_{n=0}^{\infty} p_n(t)\varepsilon_n\{v_nH_1\}, \quad (5.1)$$

$$s_1^2F_{11} = \sum_{n=0}^{\infty} p_n(t)\varepsilon_n\{v_n^2H_{11}\}, \quad (5.2)$$

$$s_2F_2 = \sum_{n=0}^{\infty} p_n(t)\varepsilon_n\{u_nH_2\}, \quad (5.3)$$

$$s_2^2F_{22} = \sum_{n=0}^{\infty} p_n(t)\varepsilon_n\{u_n^2H_{22}\}, \quad (5.4)$$

$$s_1s_2F_{12} = \sum_{n=0}^{\infty} p_n(t)\varepsilon_n\{v_nu_nH_{12}\}, \quad (5.5)$$

and

$$\begin{aligned} F_t &= \lambda F - \lambda\sum_{n=1}^{\infty}\frac{e^{-\lambda(T-t)}(\lambda(T-t))^{n-1}}{(n-1)!} + \lambda k_1\sum_{n=0}^{\infty} p_n(t)\varepsilon_n\{v_nH_1\} \\ &\quad +\lambda k_2\sum_{n=0}^{\infty} p_n(t)\varepsilon_n\{u_nH_2\} + \sum_{n=0}^{\infty} p_n(t)\varepsilon_n\{H_t\} \\ &= \lambda F + \lambda k_1s_1F_1 + \lambda k_2s_2F_2 + \sum_{n=0}^{\infty} p_n(t)\varepsilon_n\{H_t\} \\ &\quad -\lambda\sum_{m=0}^{\infty} p_m(t)\varepsilon_{m+1}\{H(v_{m+1}, u_{m+1}, t)\} \quad (5.6) \end{aligned}$$

$$\begin{aligned} &\varepsilon_{(1+x_1),(1+x_2)}\{F(s_1(1+x_1), s_2(1+x_2), t)\} \\ &= \varepsilon_{(1+x_1),(1+x_2)}\left[\sum_{n=0}^{\infty} p_n(t)\varepsilon_n\left\{H(v_n(1+x_1), u_n(1+x_2), t)\right\}\right] \\ &= \sum_{n=0}^{\infty} p_n(t)\varepsilon_{n+1}\left\{H(v_{n+1}, u_{n+1}, t)\right\} \quad (5.7) \end{aligned}$$

From(5.1)–(5.7), we have that

$$\begin{aligned}
&\frac{1}{2}\sigma_1^2 s_1^2 F_{11} + \frac{1}{2}\sigma_2^2 s_2^2 F_{22} + \rho\sigma_1\sigma_2 s_1 s_2 F_{12} + (r - \lambda k_1) s_1 F_1 + (r - \lambda k_2) s_2 F_2 \\
&\quad -rF + F_t \\
&= \sum_{n=0}^{\infty} p_n(t)\varepsilon_n \left\{ \frac{1}{2}\sigma_1^2 v_n^2 H_{11} + \frac{1}{2}\sigma_2^2 u_n^2 H_{22} + \rho\sigma_1\sigma_2 v_n u_n H_{12} + r v_n H_1 \right. \\
&\quad \left. + r u_n H_2 - rH + H_t \right\} + \lambda F \\
&\quad -\lambda \sum_{m=0}^{\infty} p_m(t)\varepsilon_{m+1}\{H(v_{m+1}, u_{m+1}, t)\} \\
&= \sum_{n=0}^{\infty} p_n(t)\varepsilon_n \left\{ \frac{1}{2}\sigma_1^2 v_n^2 H_{11} + \frac{1}{2}\sigma_2^2 u_n^2 H_{22} + \rho\sigma_1\sigma_2 v_n u_n H_{12} + r v_n H_1 \right. \\
&\quad \left. + r u_n H_2 - rH + H_t \right\} - \lambda\varepsilon[F(s_1(1+x_1), s_2(1+x_2), t)] \qquad (5.8)
\end{aligned}$$

because H satisfies equation (3.14) and therefore

$$\frac{1}{2}\sigma_1^2 v_n^2 H_{11} + \frac{1}{2}\sigma_2^2 u_n^2 H_{22} + \rho\sigma_1\sigma_2 v_n u_n H_{12} + r v_n H_1 + r u_n H_2 - rH + H_t = 0$$

for each n. It follows immediately from (5.8) that $F(s_1, s_2, t)$ satisfies equation (3.13). $s_2 = 0$ implies that $u_n = 0$ for each n. Furthermore, from (4.3) $H(v_n, 0, t) = 0$. Therefore, $F(s_1, 0, t) = 0$ which satisfies boundary condition (4.1).

$$\varepsilon_n\{H(v_n, u_n, T)\} = \varepsilon_n\{max(u_n - v_n, 0)\} \leq \varepsilon_n(u_n) = s_2(1 + k_2)^n \qquad (5.9)$$

Therefore , from (5.9)

$$\begin{aligned}
& \lim_{t\to T} \sum_{n=1}^{\infty} p_n(t)\varepsilon_n\{H(v_n, u_n, t)\} \\
&\leq \lim_{t\to T} \sum_{n=1}^{\infty} \frac{s_2((1+k_2)\lambda(T-t))^n e^{-\lambda(T-t)}}{n!} \\
&= \lim_{t\to T} s_2 e^{-\lambda(T-t)}(e^{(1+k_2)\lambda(T-t)} - 1) \\
&= 0 \qquad (5.10)
\end{aligned}$$

$$\begin{aligned}
\lim_{t\to T} F(s_1, s_2, t) &= \lim_{t\to T} [p_0(t)\varepsilon_0\{H(v_0, u_0, t)] \\
&= max\{s_2 - s_1, 0\} \qquad (5.11)
\end{aligned}$$

therefore $F(s_1, s_2, t)$ satisfies boundary condition (4.2). ■

References

[1] R.C.Merton, Option pricing when underlying stock returns are discontinuous, Journal of Financial Economics 3 (1976)125–144.

[2] I.Karatzas and S.E.Shreve, Brown motion and stochastic calculus, Springer-Verlag, World Publishing Corp. (1988)

[3] W.Marhrabe, A theory of forward and futures prices. Wharton School of Business. University of Pennsylvania, Amsterdam, (1976).

[4] J.C.Cox and S.A.Ross, The valuation of options for alternative stochastic process, J.Financ.Econom.3 (1985) 145–166.

[5] F.Black and M.Scholes, The pricing of options and corporate liabilities, J.Polit. Econom. 8 (1971) 637–659.

[6] K.K.Aose. Contingent claims valuation when the security price is a combination of Itô Process and a random point process. J.Stochastic Process and their Applications 28 (1988) 185–220.

Chapter 26

FINITE HORIZON PORTFOLIO RISK MODELS WITH PROBABILITY CRITERION*

Yuanlie Lin
Department of Mathematical Sciences,
Tsinghua University
Beijing, 100084,
PR China

Jerzy A. Filar
Center for the Industrial and Applicable Mathematics,
School of Mathematics,
University of South Australia,
Mawson Lakes,
SA 5095, Australia

Ke Liu
Institute of Applied Mathematics,
The Academy of Mathematics and System Sciences,
The Chinese Academy of Science,
Beijing, 100080,
PR China

Abstract We consider a consumption investment decision problem over a finite time horizon with respect to a probability risk criterion. That is, we wish to determine how to maximize the probability of an investor's

*This research was supported partially by the Australian Research Council (IREX)Grant and the National Natural Science Foundation of China No. 19871046, 19771085, 79930900 and 79970120.

Z. Hou et al. (eds.), Markov Processes and Controlled Markov Chains, 405–424.

wealth exceeding a given value at some finite stage T. Our model is different from traditional portfolio models in three aspects: Firstly, the model is based on discrete time, that is, the investor makes decisions at discrete time points and does not change his policy at any other moment. Secondly, only finitely many time stages are considered. Finally, the criterion is probabilistic which is different from the usual expectation criterion.

Keywords: Portfolio decision, probability criterion, investment decision.

AMS 1991 Subject Classification: 90A09, 90C47

1. Introduction

In this paper we consider a single agent, discrete-time multiperiod consumption investment decision problem with a special minimum risk criterion. The portfolio consists of two kinds of assets, one low-yield asset is "risk free" (we call it a "bond"), and the other higher-yield asset is "risky" (we call it a "stock").

We suppose that the investor's wealth and consumption at stage t are denoted by X_t and c_t respectively, with the initial wealth at stage zero denoted by x. Further, we assume that the investor has a given target value which he hopes his wealth should attain by stage T.

At stages t $(t = 0, 1, \ldots, T-1)$ the investor consumes a part of the wealth c_t $(c_t > 0)$. If at any time t, his wealth X_t cannot cover the consumption c_t, he is ruined and loses the opportunity to invest at the next stage. Otherwise, he distributes the remaining wealth into two parts. One part is the amount of the bond asset and the other part is the amount of the stock asset. Let θ_t denote the fraction allocated to the stock asset at stage t $(0 \leq \theta_t \leq 1)$. In this paper, we suppose that borrowing and short-selling is not allowed which means $0 \leq \theta_t \leq 1$.

The objective is to maximize the probability that the investor does not become ruined during the finite horizon and, at the same time, that his wealth at stage T exceeds the given target level l. We call the latter the *target-survival probability* and we call this problem, the *target-survival problem.*

As we know, Markowitz's mean-variance and Merton's Expected utility criteria are widely applied in portfolio selection problems. For instance, see Markowitz [13], Merton [14]. In recent years, there are many other authors continuing to do research work in this field (eg., [10], [11], [12]). In fact, the expectation criterion is insufficient to characterize the variability-risk features of dynamic portfolio selection (see [15], [16], [17] and [19]).

We assume that investors are interested in an objective that steers their wealth towards a given profit level (target) with maximal probability, over a finite and specified horizon. We also assume that an investor prefers to make decisions only at discrete time points (eg., once a month, week or day). His aim is to reach a given level by a given time. Intuitively, under our maximization of the target-survival probability, the decision made by the investor depends not only on system's state (his wealth) but also on the target value. By introducing the target into the description of the investor's state, we formulate the risk minimizing model. This formulation created a suitably constructed Markov Decision Process with target-percentile criterion, in the sense of [18], [19] and [6]. We derive a number of classical dynamic programming properties that our target-survival problem possesses. This sets the stage for future algorithmic developments.

2. Description of the model

2.1 Classical continuous time model

We first consider a classical continuous-time model of stock portfolio selection, sometimes referred to as Merton's portfolio problem [14]. The portfolio consists of two assets: one a "risk free" asset (called a bond) and the other "risky" asset (called a stock). The price b_s per share for the bond changes in time according to $db_s = ab_s ds$ while the price p_s of the stock changes in time according to $dp_s = p_s(\alpha ds + \sigma dW_s)$. Here a, α, σ are constants with $\alpha > a$, $\sigma > 0$ and W_s is a standard one-dimensional Brownian motion defined on a complete probability space $(\Omega, \mathcal{F}, P)$. The agent's wealth X_s at time s is governed by the stochastic differential equation

$$dX_s = (1-\theta_s)X_s a\, ds + \theta_s X_s(\alpha\, ds + \sigma\, dW_s) - c_s\, ds,$$

where $0 < s < T$, θ_s is the fraction of wealth invested in the stock at time s and $c_s \geq 0$ is the consumption rate. Assume that $\{c_s, s \geq 0\}$ satisfies $\int_0^T \theta_s^2 ds < +\infty$, $\int_0^T c_s ds < +\infty$.

We also assume that the investor observes his wealth, X_t at time $s = t$, and that at that time he also selects the functions representing his current consumption rate $c_s \geq 0$ and the fraction θ_s of his wealth that he allocates into the risky investment with the remaining fraction $(1-\theta_s)$ allocated into the safe one; throughout the time interval $[t, t+1]$. Then the investor's wealth X_s satisfies the Itô Stochastic Differential

Equation:

$$\begin{aligned} dX_s &= (1-\theta_s)X_s a\,ds + \theta_s X_s(\alpha\,ds + \sigma\,dW_s) - c_s\,ds, \\ & \qquad t \le s \le t+1, \quad t \in [0,T]. \end{aligned} \tag{2.1}$$

If $X_t = x_t$, then the explicit solution of the stochastic differential equation (2.1) has the form (to see [12]):

$$\begin{aligned} X_{t+1} = & \left\{ exp\left(\int_t^{t+1} \left[a + \theta_s(\alpha - a) - \frac{\sigma^2}{2}\theta_s^2 \right] ds + \sigma^2 \int_t^{t+1} \theta_s\,dW_s \right) \right\} \times \\ & \left\{ x_t - \int_t^{t+1} \left(exp\left[-\int_t^s \left[a + \theta_u(\alpha - a) - \frac{\sigma^2}{2}\theta_u^2 \right] du \right] \right) c_s\,ds \right\} \end{aligned} \tag{2.2}$$

2.2 Discrete time investment and consumption model

In reality, most investor's decision making processes are discrete. That is, an investor observes the price of a bond, a stock and his wealth only at discrete points of time. Similarly, we could assume that he makes consumption and allocation decisions only at those times. More precisely, our decision-maker (ie., investor) observes his wealth only at $t = nh$ and we define $x_n \hat{=} X_{nh}$ for each $n = 0, 1, 2, \ldots$). We also define a corresponding pair of decision variables $a_n = (\theta_n, c_n)$ that will remain constant during $[nh, (n+1)h)$. Without the loss of generality, we let $h = 1$. Now, given $x_n = x$, from the equation (2.2), we have

$$x_{n+1} = \rho(\theta_n) \cdot e^{\sigma^2 \xi(\theta_n)} \big(x_n - c_n \cdot \beta(\theta_n)\big), \tag{2.3}$$

for $n = 0, 1, 2, \ldots$), where the quantities

$$\begin{aligned} \delta(\theta_n) &= a + \theta_n(\alpha - a) - \sigma^2\theta^2/2, \\ \rho(\theta_n) &= e^{\delta(\theta_n)}, \\ \beta(\theta_n) &= \left(1 - e^{-\delta(\theta_n)}\right)/\delta(\theta_n), \\ \xi(\theta_n) &= \theta_n \int_n^{n+1} dW(u) \sim N(0, \theta_n^2) \end{aligned}$$

are obtained from the natural discretization of (2.2).

As mentioned in the Introduction, our goal in this paper is to find a policy which maximizes the probability that the wealth reaches a specified target value at stage $n = N$. In the following sections we show how this goal can be attained with the help of a discrete time Markov Decision Process (MDP, for short) with a probability criterion.

2.3 MDP model with probability criterion.

2.3.1 Standard Markov decision model.

A discrete-time MDP is a four tuple,

$$\Gamma_0 = (X, A, Q, r),$$

where $X = [0, +\infty)$ is the state space, x_n is state of the system at stage n ($n = 0, 1, 2, \ldots N$) which denotes the wealth of the investor at stage n. Let $A = [0, 1] \times [0, \infty)$ and the investor's action set in state x be denoted by a nonempty measurable subset $A(x) \subset A$. Here $A(x)$ denotes the set of feasible actions when the system is in state $x \in X$. A probabilistic transition law is denoted by q, that is a stochastic kernel on X. Given $X \times A$, a measurable function $r : X \times A \to R$ is called the reward-per-stage. In the classical formulation the decision-maker wants to maximize the expected value of total rewards. Below, we propose an alternative criterion that seems particularly relevant in the context of financial applications.

2.3.2 Target based MDP with probability criterion.

The MDP discussed in this paper belongs to the class of risk sensitive models (see [17, 5, 6, 19]). In our model, the decision maker considers not only the system's state but also his target value when making decisions and taking actions at each stage (see Filar et al [6] and Wu and Lin [18]) and wishes to maximize the probability of attaining that target.

As a consequence, the current decision made by the investor depends not only on system's state x, but also on the changing *current target level* y, which represents the difference between the current wealth x and the target wealth l which he wants to reach at stage N. More generally, we introduce the concept of a *target set* and denote it by L, for instance $L = R$, $L = [l_1, l_2] \subset R$ or $L = \{\{l_1\}, \{l_2\}\}$.

Since the current decisions will now depend on both the state of the investor's wealth and the target value, in the MDP model it is helpful to extend the decision-maker's state space to $E = \{e = (x, y) : x \in X, x + y \in L\}$. Let $e_n = (x_n, y_n)$, $a_n = (\theta_n, c_n)$ and $A(e_n) = A(\theta_n, c_n) = [0, 1] \times [m, x_n \cdot \beta^{-1}]$ where $m > 0$ is a positive number denoting the minimum amount required for consumption and

$$\beta = \max_{0 \le \theta \le 1} \beta(\theta) < \infty.$$

The set $A(e_n)(\forall e_n \in E)$ will be called the set of *feasible actions* at $e_n = (x_n, y_n)$.

Suppose the wealth of the investor at stage n is x and his current target value is y, that is, $x + y \in L$. If $x \cdot \beta^{-1} \le m$, he is ruined and loses the opportunity to invest in this stage and all following stages. Otherwise, he chooses an action $a = (\theta, c) \in A(x, y)$. That is, if at some stage n the decision-maker finds himself in state $e = (x, y)$ and he chooses the action $a = (\theta, c) \in A(x, y) = A(e) \neq \emptyset$, then his wealth at stage $n + 1$ is given by

$$x_{n+1} = \rho_n(\theta) \cdot e^{\sigma^2 \xi_n(\theta)} \big(x - c \cdot \beta_n(\theta)\big).$$

Denote the probability density function of $\xi_n(\theta)$ by

$$g(\theta, z) = \frac{1}{\sqrt{2\pi\theta}} e^{-z^2/2\theta^2} \sim N(0, \theta^2).$$

It follows that, x_{n+1} is a random variable whose distribution is determined by the distribution of $\xi_n(\theta)$. The probabilities of interest to us now have the form:

$$\begin{aligned} &P\big[x_{n+1} \in B | e_n = (x, y), a_n = (\theta, c)\big] \\ &\quad = \int I_B\big(\rho(\theta) e^{\sigma^2 z}\big(x - c \cdot \beta(\theta)\big)\big)\, g(\theta, z)\, dz, \end{aligned}$$

$\forall B \in B(X)$, where $B(X)$ denote the Borel σ-algebra.

With the above notation, we now see that when the investor chooses action a_n in state e_n he will receive a *current reward* of $r(e_n, a_n) = x_{n+1} - x$ for that stage. The current target value for the investor now changes to $y_{n+1} = y - (x_{n+1} - x) = x + y - x_{n+1}$.

Therefore, if the decision-maker's state is $e_n = (x, y) \in E$, and he takes the action $a_n = (\theta, c)$, then the next state is $e_{n+1} = (x_{n+1}, y_{n+1})$ with probability

$$\begin{aligned} &q_n\big(e_{n+1} \in B \times C | e_n = (x, y), a_n = (\theta, c)\big) \\ &= \begin{cases} I_{B \times C}(x, y), & \text{if } x \leq b_0 \beta(\theta), \\ \displaystyle\int I_{B \times C}\big(\rho(\theta) e^{\sigma^2 z}\big(x - c \cdot \beta(\theta)\big), x + y & \\ \qquad - \rho(\theta) e^{\sigma^2 z}\big(x - c \cdot \beta(\theta)\big)\big)\, g(\theta, z)\, dz, & \text{if } x > b_0 \beta(\theta), \end{cases} \end{aligned}$$

for any $B \in B(x), C \in B(R)$.

Let $K = \{(e, a) : e \in E, a \in A(e)\}$, $Q = (q_n, n \geq 0)$ be a sequence of stochastic kernels on E given K. We call

$$\Gamma = (E, A, r, Q)$$

the *target based MDP with probability criterion*, or T-MDP for short.

In order to define the usual hierarchy of policies in T-MDP we now define the sets of *histories* $H_0 = E$, $H_n = K \times H_{n-1}$ for $0 \leq n \leq N$. In particular H_n denotes the set of all admissible histories up to stage n with elements $h_n = (e_0, a_0, e_1, a_1, \ldots, e_n) \in H_n$, where $e_n = (x_n, y_n) \in E$, $a_n = (\theta_n, c_n) \in A(e_n)$.

A *policy* is a sequence $\pi = \{\pi_n, n \geq 0)$ of stochastic kernels π_n on A given h_n satisfying the constraint: $\pi_n(A(e_n)|h_n) = 1$, $\forall h_n \in H_n$, $n \geq 0$. The set of all policies is denoted by Π.

A *Markov policy* $\pi = (\pi_n, n \geq 0)$ is one in which each π_n depends only on the current state at stage n, that is, $\pi(\cdot|h_n) = \pi(\cdot|e_n)$ for all $h_n \in H_n$.

A *stationary policy* π is a Markov policy in which each *decision rule* $\pi_n = \pi_D$ and hence it is denoted by $\pi = \pi_D^\infty$.

A *deterministic Markov policy* π is one in which each π_n is non-randomized, that is, π_n is a measurable mapping from H_n to A such that $\pi_n(\cdot|h_n) \in A(e_n)$ for all $h_n \in H_n$; a *deterministic stationary policy* is similarly defined.

Let Π_m, Π_m^d, Π_s and Π_s^d denote the sets of all Markov-policies, all deterministic Markov policies, all stationary policies, and all deterministic stationary policies, respectively.

Let Π_0 denote the set of all policies which are independent of targets value y_n $(n \geq 0)$. For any $\pi = (\pi_n, n \geq 0) \in \Pi$ and a given single stage history $(e, a) = (x, y, a) \in H_1$, the *cut-head policy* of π with respect to (e, a) is defined by $\pi^{(e,a)} = (\pi_n^{(e,a)}, n \geq 0)$, where $\pi_k^{(e,a)}(\cdot|h_k) = \pi_{k+1}(\cdot|(e, a), h_k)$ for all $h_k \in H_k$, $k \geq 0$.

Let $\Omega = H_\infty$ and $\mathcal{F} = \sigma(H_\infty)$ be the corresponding product σ-algebra. Given $\pi \in \Pi$ and an initial state distribution α_0, according to the theorem of Ionescu-Tulcea, there exists a unique probability measure P on $(\Omega, \mathcal{F})$, which satisfies

$$\begin{aligned} P^\pi(e_0 \in B) &= \alpha_0(B), \\ P^\pi(a_n \in D|h_t) &= \pi_n(D|h_n), \\ P^\pi(e_{n+1} \in B|h_n, a_n) &= q_n(B|e_n, a_n), \end{aligned}$$

$\forall B \in B(E)$, $D \in \mathcal{B}(A)$, and $h_n \in H_n$, $n \geq 0$. We call the stochastic process $(e(\pi), a(\pi)) = \{(x_n, y_n, a_n), t \geq 0, \pi \in \Pi\}$ *target based Markov decision process.*

For any given $\pi = (\pi_n, n \geq 0)$ it will be sometimes convenient to suppress the dependence on the policy in the quantities $x_{n+1}(\pi)$, $\delta_n(\pi)$, $\rho_n(\pi)$, $\xi_n(\pi)$, $\beta_n(\pi)$, $c_n(\pi)$ and denote them simply by x_{n+1}, δ_n, ρ_n, ξ_n, β_n. Clearly, from (2.3) these still satisfy

$$x_{n+1} = \rho_n e^{\sigma^2 \xi_n}(x_n - c_n \cdot \beta_n),$$

where $\rho_n = e^{\delta_n}$ and

$$\beta_n = \begin{cases} \delta_n^{-1}\left(1-e^{-\delta_n}\right), & \delta_n \neq 0; \\ 1, & \delta_n = 0. \end{cases}$$

If $\theta_n = \theta$ $(n \geq 0)$, the random variables $(\xi_n, n \geq 0)$ are independent and identically distributed and $\xi_n \sim N\left(0, \theta^2\right)$.

Let L be the target level set. Clearly, $L \subset R$; take $l \in L$. Given $\pi = (\pi_n, n \geq 0)$ define the following dynamic programming type quantities

$$\begin{aligned} V_n(x,y,\pi) &= P^\pi\{b_0\beta < x_k, 0 \leq k \leq n-1, x_n \geq l | e_0 = (x,y)\}, \\ & \qquad\qquad\qquad\qquad\qquad 0 \leq n \leq N; \\ V_0(x,y,\pi) &= V_0^*(x,y) = I_{(y\leq 0)}, \qquad \forall (x,y) \in E,\ \pi \in \Pi; \\ V_n^*(x,y) &= \sup_{\pi\in\Pi} V_n(x,y,\pi), \qquad \forall (x,y) \in E,\ 0 \leq n \leq N. \end{aligned}$$

If $\pi^* \in \Pi$ is such that $V_N(x,y,\pi^*) = V_N^*(x,y)$ for all $(x,y) = e \in E$, $l = x + y \in L$, then π^* is called an *N-stage L-optimal policy* or an *N-stage optimal policy for minimizing risk with respect to L* (or, simply an *N-stage optimal policy with respect to L*).

Let $\Pi^*(L)$ be the set of all N-stage L-optimal policies,

$$\Pi^*(L) = \{\pi^* : V_N(x,y,\pi^*) = V_N^*(x,y), \forall (x,y) \in E, x+y \in L\}.$$

Obviously, it follows that the following two properties hold:

(i) If $L_1 \subset L_2$, $\Rightarrow \Pi^*(L_2) \subset \Pi^*(L_1)$,

(ii) For any index set K,

$$\Pi^*\left(\bigcup_{k\in K} L_k\right) = \bigcap_{k\in K} \Pi^*(L_k).$$

In particular, three types of the target level set L are considered in this paper:

1. $L = R$ for the complete stochastic order optimization model;
2. $L = [l_1, l_2]$, $0 \leq l_1 < l_2$ for the local stochastic order optimization model;
3. $L = \{l\}$ for the single point stochastic order optimization model.

If π^* is a N-stage L-optimal policy with respect to $L = R$ ($L = [l_1, l_2]$ or $L = \{l\}$), we shall also call π^* an optimal policy for a complete stochastic order (local stochastic order, single stochastic order).

The three models introduced above can be applied to three differential cases. If a decision-maker has a particular profit target in mind, he might want to use the single point stochastic order optimal policy, which attains that profit target with maximum probability. More generally, he might want to use the local stochastic order optimal policy, which ensures the maximum probability of attaining any target profit level with respect to the interval $L = [l_1, l_2]$, $0 \le l_1 < l_2$.

For example, the investor may wish that the probability of his wealth being more than l_2 should be no less than 0.95 at stage N, while the probability of the wealth being more than l_1 $(l_1 < l_2)$ should be no less than 0.99. Since these values 0.95 or 0.98 might be impossible to achieve, a reasonable approach is to maximize the probability of both $x_N \ge l_1$ and $x_N \ge l_2$.

Finally, the complete stochastic order optimization model $L = R$ is introduced only for the sake of mathematical completeness.

3. Finite horizon model

In this section we demonstrate that our target-survival problem possesses many of the desirable properties of standard MDPs. We begin by considering the properties of the n-stage value function $V_n(x, y, \pi)$. If we let $\pi(n) = (\pi_0, \pi_1, \ldots, \pi_n)$ denote the truncation of π to n stages, then it is clear that $V_n(x, y, \pi)$ $(0 \le n \le N)$ is determined by $\pi(n)$.

Lemma 3.1 *Let* $\pi = (\pi_n, n \ge 0) \in \Pi$, *then* $\forall (x, y) \in E$, $1 \le n \le N$,

$$V_n(x, y, \pi) = \begin{cases} \int_{A(x,y)} \pi\big(da|e_0 = (x, y)\big) \int_E V_{n-1}\left(u, v, \pi^{(x,y,a)}\right) \\ \qquad\qquad \times q(du \times dv|x, y, a), & x > b_0\beta \\ 0 & x \le b_0\beta. \end{cases} \tag{3.1}$$

$$V_0(x, y, \pi) = V_0^*(x, y) = I_{(y \le 0)}. \tag{3.2}$$

Proof The Equation (3.1) follows easily from the law of total probability and the properties of P^{π}. ■

Let $D = \{V : E \to [0, 1] | V \text{ a measurable function}\}$ and $\delta^{\infty} \in \Pi_s$. For each $u \in D$, $(x, y) \in E$, $a \in A(x, y)$, we define the operators $\mathcal{L}$, T^{δ} and T:

$$\mathcal{L}U(x,y,a) = \int_E U(du,dv)\, q(du \times dv|x,y,a); \tag{3.3}$$

$$T^{\delta}U(x,y) = \int_{A(x,y)} \delta(da|x,y)\, \mathcal{L}U(x,y,a); \tag{3.4}$$

$$TU(x,y) = \max_{a \in A(x,y)} \mathcal{L}U(x,y,a). \tag{3.5}$$

It follows from the definitions that

$$\begin{array}{rclrcl} (T^{\delta})^0 U &=& U, & (T^{\delta})^n U &=& T^{\delta}((T^{\delta})^{n-1}U), \\ T^0 U &=& U, & T^n U &=& T(T^{n-1}U), \end{array} \tag{3.6}$$

where $(T^{\delta})^n U$ means that the operator T^{δ} is applied to U n times and $(T^{\delta})^0$ is defined as the identity operator. Obviously, for any $\delta^{\infty} \in \Pi_s^d$ we have $T^{\delta}U(x,y) = \mathcal{L}U(x,y,\delta(x,y))$. We shall say that functions U, V in D satisfy the inequality $U \le V$ if $U(e) \le V(e)$ for every $e \in E$.

Lemma 3.2

(i) The operators $\mathcal{L}$, T^{δ}, T are monotone. That is, if $U, V \in D$, $U \le V$, then $\mathcal{L}U \le \mathcal{L}V$, $T^{\delta}U \le T^{\delta}V$, $TU \le TV$.

(ii) For any $U \in D$, if $U(x,y)$ is a non-increasing and a left continuous function with respect to y for any $x \in X$, then $TU(x,y)$ is also a non-increasing and a left continuous function with respect to y for each $x \in X$.

Proof The proof is obvious. ■

Since the right hand side of (3.1) is a little complex but corresponds to the decision maker using π_0 initially and then expecting a return of V_{n-1} thereafter, we shall extend the previous notation by setting $T^{\pi_0}(x,y)$ equal to the right side of (3.1).

Thus, the equation (3.1) can be re-written in a simpler, operator, form as:

$$V_n(\pi) = T^{\pi_0}V_{n-1}\left(\pi^{-0}\right), \quad 0 \le n \le N, \tag{3.7}$$

where $\pi^{-0} = (\pi_1, \pi_2, \ldots)$ is the cut-head policy obtained from a Markov policy $\pi = (\pi_0, \pi_1, \pi_2, \ldots)$ by deleting the initial decision rule π_0.

Similarly, when working with the optimal value function $V_n^*(x,y)$, $\forall (x,y) \in E$, $n \ge 0$ defined earlier, we can also use the simpler notation V_n^*, $n \ge 0$. The next result establishes the so-called optimality principle for V_n^*.

Theorem 3.1

(i) The optimal value function $\{V_n^*, 0 \le n \le N\}$ *satisfies the optimality equations:*

$$V_0^* = I_{(y \le 0)}, \qquad V_n^* = TV_{n-1}^*, \qquad (1 \le n \le N). \qquad (3.8)$$

(ii) For any $0 \le n \le N$, *and* $0 \le V_n^*(x,y) \le 1$, $V_n^*(x,y)$ *is a non-increasing and left continuous function of* y *for each* $x \in X$.

(iii) For any $0 \le n \le N$, *there exists a policy* $\pi \in \Pi_m^d$ *such that* $V_n(\pi) = V_n^*$, $0 \le n \le N$ *for any initial state* $e \in E$.

Proof We prove Theorem 3.1 by induction.

When $n = 0$ Theorem 3.1 is true by (3.2).

Assume that Theorem 3.1 holds when $n = k$. By inductive assumption (applied to all parts of the theorem), for any $x \in X$, $V_k^*(x,y)$ is a non-increasing and left continuous function of y and there exists $\sigma = (\sigma_k, k \ge 0) \in \Pi_m^d$ such that $V_k^* = V_k(\sigma)$. Also, because our criterion is a probability, we have that $0 \le V_k^* \le 1$.

Note that $A(x,y)$ is a closed set for any $e = (x,y) \in E$. By a measurable selection theorem (see [1], [7] or [8]), there exists a measurable mapping δ from E to A such that $\delta(x,y) \in A(x,y)$ and $\mathcal{L}V_k^*(x,y,\delta(x,y)) = TU_k^*(x,y)$ for all $(x,y) \in E$. That is, $\delta^\infty \in \Pi_s^d$ and $T^\delta V_k^* = TV_k^*$.

By the inductive assumption, there exists a policy $\sigma \in \Pi_m^d$ such that $V_k(\sigma) = V_k^*$. Let $\pi = (\delta, \sigma) = (\delta, \sigma_0, \sigma_1, \ldots)$, then $\pi \in \Pi_m^d$.

By Lemma 3.1 and equation (3.7) we have,

$$\begin{aligned} V_{k+1}^*(x,y) &\ge V_{k+1}(x,y,\pi) \\ &= T^\delta V_k(x,y,\sigma) \\ &= T^\delta V_k^*(x,y) \\ &= TV_k^*(x,y). \end{aligned} \qquad (3.9)$$

On the other hand, for any $\eta = (\eta_0, \eta_1, \ldots) \in \Pi$, by Lemma 3.1, (3.7) and the definition of T we have,

$$V_{k+1}(x,y,\eta) = T^{\eta_0} V_k(x,y,\eta) \le T^{\eta_0} V_k^*(x,y) \le TV_k^*(x,y)$$

and hence, by maximizing the left hand side of the above with respect to η we have,

$$V_{k+1}^*(x,y) \le TV_k^*(x,y).$$

Combining the latter with (3.9) we obtain,

$$V_{k+1}^*(x,y) = TV_k^*(x,y).$$

It now follows that $TV_k^* = V_{k+1}^* = V_{k+1}(\pi)$. Also it follows from Lemma 3.2 that $V_{k+1}^*(x,y)$ is a non-increasing and left continuous function of y.

By the above argument we have that the theorem also holds when $n = k+1$, thereby completing the induction. ■

Corollary 3.1 *It is possible to restrict the policy space to Markov deterministic policies. That is,*

$$V_n^*(x,y) = \sup_{\pi\in\Pi} V_k(x,y,\pi) = \sup_{\pi\in\Pi_m^d} V_k(x\ ,y,\pi),$$

$$(x,y) \in E,\ 0 \le n \le N.$$

Next, we shall discuss some properties of optimal policies. We define:

$$A_n^*(x,y) = \{a | a \in A(x,y) \text{ and } V_n^*(x,y) = \mathcal{L}V_{n-1}^*(x,y,a)\}, \quad \forall (x,y) \in E, \tag{3.10}$$

$$L_x = \{y : x+y \in L\}, \qquad \forall (x,y) \in E, \tag{3.11}$$

$$A_N^*(x) = \bigcap_{y\in L_x} A_N^*(x,y). \tag{3.12}$$

By Theorem 3.1 the set of optimal actions at state (x,y), $A_n^*(x,y) \neq \emptyset$ for any $n \ge 0$, $(x,y) \in E$. However, it is possible that $A_n^*(x) = \emptyset$.

Theorem 3.2 *Let δ_n be a measurable mapping from E to A which satisfies $\delta_n(x,y) \in A_n^*(x,y)$ for all $(x,y) \in E$, $0 \le n \le N$. Then any policy π which satisfies $\pi(N) = (\delta_N, \delta_{N-1}, \dots, \delta_1, \delta_0)$ is N-stage optimal for the target-survival problem with respect to L.*

Proof Note that because of the backward recursion of dynamic programming, $T^{\delta_n}V_{n-1}^* = V_n^*$ for all $n \ge 1$ and V_0^* is defined as in (3.2).

For $N = 1$, we have $\pi(1) = (\pi_0,\ \pi_1) = (\delta_1, \delta_0)$, and also $\pi(0) = (\pi_0) = (\delta_0)$. Then by equation (3.7), we have

$$V_1(\pi) = T^{\pi_0}V_0\left(\pi^{-0}\right) = T^{\delta_1}V_0^* = V_1^*,$$

where the second last equality follows from Lemma 3.1, which gives $V_0(\pi^{-0}) = V_0^*$.

Assume that Theorem 3.2 holds for $N = k$. Consider the case $N = k+1$.

Now, $\pi(k+1) = (\pi_0,\ \pi_1,\ \dots,\ \pi_{k+1}) = (\delta_{k+1},\ \delta_k,\ \dots,\ \delta_0)$. Since $\pi^{-0}(k+1) = (\delta_k, \delta_{k-1}, \dots, \delta_0)$ and by inductive hypothesis $V_k(\pi^{-0}(k+$

1)) $= V_k^*$, by Lemma 3.1 we obtain

$$\begin{aligned} V_{k+1}(\pi(k+1)) &= T^{\pi_0} V_k \left(\pi^{-0}(k+1)\right) \\ &= T^{\delta_{k+1}} V_k^* = V_{k+1}^*. \end{aligned}$$

By induction, Theorem 3.2 is proved. ■

Theorem 3.3 *Consider the set of Markov policies Π_m. For any given $(x,y) \in E$, $n \geq 1$, $V_n(x,y,\pi) = V_n^*(x,y)$ if and only if $\pi_0 \left(A_n^*(x,y)|x,y\right) = 1$ for every $n = 0,1,\ldots,N$ and $V_{n-1}(u,v,\pi^{(x,y,a)}) = V_{n-1}^*(u,v)$ for any $(u,v) \in B$ which satisfies,*

$$\int_{a \in A_n^*(x,y)} \pi_0(da|x,y) \int_{(u,v)\in B} q(du \times dv|x,y,a) > 0, \qquad B \in B(E). \tag{3.13}$$

Proof Assume that $V_n(x,y,\pi) = V_n^*(x,y)$ and $\pi = (\pi_0, \pi_1, \ldots, \pi_n)$. By Theorem 3.1, there exists an optimal policy $\sigma = (\sigma_k, k \geq 0) \in \Pi$, such that $V_{n-1}(x,y,\sigma) = V_{n-1}^*(x,y)$ for all $(x,y) \in E$. Hence, by Lemma 3.1, we have

$$\begin{aligned} V_n^*(x,y) &= V_n(x,y,\pi) \\ &= T^{\pi_0} V_{n-1}\left(x,y,\pi^{-0}\right) \\ &\leq T^{\pi_0} V_{n-1}^*(x,y) \\ &= T^{\pi_0} V_{n-1}(x,y,\sigma) \\ &= V_n(x,y,(\pi_0,\sigma)) \\ &\leq V_n^*(x,y), \end{aligned}$$

where $(\pi_0,\sigma) = (\pi_0, \sigma_0, \sigma_1, \ldots)$ and so,

$$V_n^*(x,y) = T^{\pi_0} V_{n-1}^*(x,y), \tag{3.14}$$

and

$$T^{\pi_0} V_{n-1}\left(x,y,\pi^{-0}\right) = T^{\pi_0} V_{n-1}^*(x,y). \tag{3.15}$$

From (3.14), with help of (3.3), (3.4) and (3.1), we have that $\forall (x,y) \in E$,

$$\begin{aligned} 0 &= T^{\pi_0} V_{n-1}^*(x,y) - V_n^*(x,y) \\ &= \int_{A_n(x,y)} \pi_0(da|x,y) \mathcal{L} V_{n-1}^*(x,y) - V_n^*(x,y) \\ &= \int_{A_n(x,y)} \pi_0(da|x,y) \left\{\mathcal{L} V_{n-1}^*(x,y,a) - V_n^*(x,y)\right\}. \end{aligned} \tag{3.16}$$

With the help of $\mathcal{L}V_{n-1}^*(x,y) \le V_n^*(x,y)$, (3.16), Theorem 3.1 and the definition of $A_n^*(x,y)$, we have

$$\pi_0(A_n^*(x,y)|x,y) \;=\; 1, \tag{3.17}$$

for all $(x,y) \in E$. Using the similar way, from (3.15), we have

$$\int_{A_n(x,y)} \pi_0(da|x,y) \int_E q(du \times dv|x,y,a) \left[V_{n-1}\left(u,v,\pi^{(x,y,a)}\right) - V_{n-1}^*(u,v)\right] = 0. \tag{3.18}$$

Thus, by (3.18) and (3.17), for any $B \subset B(E)$ such that

$$\int_{a \in A_n^*(x,y)} \pi_0(da|x,y) \int_B q(du \times dv|x,y,a) \;>\; 0,$$

we have

$$V_{n-1}\left(u,v,\pi^{(x,y,a)}\right) \;=\; V_{n-1}^*(u,v)$$

when $(u,v) \in B$, $a \in A_n^*(x,y)$ for all $(x,y) \in E$.

Hence the necessity of Theorem 3.3 is proved. Note that the preceding proof is reversible and so the sufficiency of the theorem also holds. ■

Remark 3.1 Theorem 3.3 *shows that a Markov policy π is optimal for a finite horizon model if and only if the action taken by π at each realizable state is an optimal action and the corresponding "cut-head" policy is also optimal at each stage. In general,* Theorem 3.3 *also holds for the general policies (the proof is similar to the proof of* Theorem 4 in [4].

Theorem 3.4

(i) *If there exists a policy $\pi \in \Pi_0$ such that $V_n(x,y,\pi) = V_n^*(x,y)$ for all (x,y) (i.e. $V_n(\pi) = V_n^*$), then $A_n^*(x) \neq \emptyset$ and $\pi_0(A_n^*(x)|x) = 1$ for any $x \in X$;*

(ii) *If $A_k^*(x) \neq \emptyset$ for all $x \in X$ and $0 \le k \le n$, then there exists a policy $\pi \in \Pi_0$ such that $V_n(\pi) = V_n^*$.*

Proof

(i) Let $\pi \in \Pi_0$ and $V_n(\pi) = V_n^*$. Then, by Theorem 3.3, $\pi_0(A_n^*(x,y)|x) = 1$ for all $x \in X$ and y such that $x + y \in L$. It follows that $\pi_0(A_n^*(x)|x) = 1$ for all $x \in X$. Hence, $A_n^*(x) \neq \emptyset$ for all $x \in X$.

(ii) Select $\delta_k : X \to A$ such that $\delta_k(x) \in A_k^*(x)$ for all $x \in X$ and $0 \le k \le n$. Then, by Theorem 3.3, any policy $\delta \in \Pi_0$ constructed as $\delta = (\delta_n, \delta_{n-1}, \dots, \delta_1, \delta_0)$, satisfies $V_n(\delta) = V_n^*$. ■

4. Three risk regions in the decision-maker's state space.

In this section, we demonstrate that the decision-maker's space can be divided into three distinct regions: the *risk free zone*, the *ruin zone* and the *risk zone.*

In the risk free region, the investor can find a risk free investment-consumption policy that reaches the desired target level with probability 1. In the ruin-zone, the investor cannot meet the minimal consumption requirement and is ruined. Hence, the most interesting region is the risk zone and the problem of finding an optimal target-survival policy with respect to the given target set L.

Let $\pi^f(N)$ denote a *riskless* investment-consumption policy. That is, $\pi^f(N)$ is a policy that always allocates all of the investment into the risk free asset (bonds). Clearly, under this policy, in the notation of Section 2.2, we have that $\theta_n = 0 \; \forall \; n$, and hence for all n, $\delta(0) = \delta(\theta_n) = a$ and $\beta_0 = \beta(0) = \beta(\theta_n) = (1 - e^{-a})/a$ and $\rho \hat{=} \rho(0) = e^a > 1$. Now, iterating equation (2.3) under this policy yields

$$x_n = \rho^n x_0 - C_n(d),$$

where

$$C_n(d) \;\hat{=}\; \beta_0 \cdot \sum_{i=0}^{n-1} c_i \rho^{n-i}. \tag{4.1}$$

Define

$$\Delta_n = (\rho^n - 1), \tag{4.2}$$

for $n \geq 1$. We may interpret $C_n(d)$ as the total discounted consumption from stage 0 to stage n when the investor adopts a riskless policy. Clearly, $x_n - x_0 = \Delta_n x_0 - C_n(d)$ under this riskless policy.

Next, we define the following sets:

$$\begin{cases} R_0 = \{(x,y) | x > b_0, y \leq 0\}, & \\ A_0^f(x,y) = \{a | a = (0,c), b_0 \leq c < x\}, & \forall (x,y) \in R_0, \\ R_n = \{(x,y) | x > \rho^{-n}(\beta_0 b_0 + C_n(d)), & \\ \quad\quad y \leq x\Delta_n - C_n(d), (x,y) \in R_{n-1}\}, & n \geq 1, \\ A_n^f(x,y) = \{a | a = (0,c), b_0 \leq c < x\beta_0^{-1}\}, & \forall (x,y) \in R_n, \\ \partial R_n = \{(x,y) | x > \rho^{-n}(\beta_0 b_0 + C_n(d)), & \\ \quad\quad y = x\Delta_n - C_n(d), (x,y) \in R_{n-1}\}, & n \geq 1. \end{cases} \tag{4.3}$$

We shall refer to R_n as the *risk free zone*, because of the following result.

Theorem 4.1 *Consider any state $(x,y) \in R_n$, $0 \le n \le N$, then under the riskless policy $\pi^f(n)$ we have that*

$$V_n\left(x, y, \pi^f(n)\right) = 1.$$

Proof By the definition, we have

$$R_n \subset R_{n-1} \subset \ldots \subset R_1 \subset R_0. \tag{4.4}$$

We prove Theorem 4.1 by induction.

When $n = 0$, $\forall(x,y) \in R_0$, we have $x > b_0$ and $y \le 0$. Then for any $b_0 \le c < x$,

$$a = (0, c) \in A_0^f(x,y) \neq \emptyset,$$

and for any $\pi^f(0)(\bullet)$, which is a probability distribution on the set of $A_0^f(x,y)$, we have

$$V_0\left(x, y, \pi^f(0)\right) = I_{(y \le 0)} = 1, \tag{4.5}$$

for all $(x,y) \in R_0$.

Inductively, assume that when $n = k$, we have

$$\begin{aligned} R_k = \Big\{&(x,y) | x > \rho^{-k}(\beta_0 b_0 + C_k(d)), \\ &y \le x\Delta_k - C_k(d) \text{ and } (x,y) \in R_{k-1}\Big\}; \end{aligned} \tag{4.6}$$

$$\begin{aligned} &V_k\left(x, y, \pi^f(k)\right) \\ &= P^{\pi^f(k)}\big(x_j > b_0, 0 \le j \le k-1, y_k \le 0 | e_0 = (x,y)\big) \\ &= 1, \end{aligned} \tag{4.7}$$

for all $(x,y) \in R_k$.

For the case of $n = k+1$, $\forall(x,y) \in R_{k+1}$, we have

$$x > \rho^{-(k+1)}\big(\beta_0 b_0 + C_{k+1}(d)\big), \tag{4.8}$$

$$y \le x\Delta_{k+1} - C_{k+1}(d), \tag{4.9}$$

and $(x, y) \in R_k$. Because $c_i \geq b_0$, for all $i = 0, 1, \ldots$, we have the following:

$$\begin{aligned} & 1 + \sum_{i=0}^{k} \rho^{k+1-i} &>& \rho^{k+1} \\ \Leftrightarrow \quad & 1 + \sum_{i=0}^{k} \frac{c_i}{b_0} \rho^{k+1-i} &>& \rho^{k+1} \\ \Leftrightarrow \quad & b_0 \sum_{i=0}^{k} c_i \rho^{k+1-i} &>& b_0 \rho^{k+1} \\ \Leftrightarrow \quad & \rho^{-(k+1)} + \big(\beta_0 b_0 + C_{k+1}(d)\big) &>& \beta_0 b_0. \end{aligned}$$

It follows from (4.5) that:

$$x\beta_0^{-1} \quad > \quad b_0. \tag{4.10}$$

That is the interval $[b_0, x\beta_0^{-1})$ is nonempty. Hence $A_{k+1}^{f} \neq \emptyset$.
From (4.1), we have

$$C_k(d) \quad = \quad \beta_0 \left(c_0 \rho^k + c_1 \rho^{k-1} + \cdots + c_{k-1} \rho \right)$$

and

$$C_{k+1}(d) \quad = \quad \beta_0 \left(c \rho^{k+1} + c_0 \rho^k + c_1 \rho^{k-1} + \cdots + c_{k-1} \rho \right). \tag{4.11}$$

For any policy $a = (0, c) \in A_{k+1}^{f}(x, y)$, we have

$$x_1 \quad = \quad \rho x - \beta_0 c \rho. \tag{4.12}$$

With the help of the following inequalities

$$\begin{aligned} x &>& \rho^{-(k+1)} \left(\beta_0 b_0 + C_k(d) + \beta_0 c \rho^{k+1} \right), \\ \Leftrightarrow \quad x - \beta_0 c &>& \rho^{-(k+1)} \big(\beta_0 b_0 + C_k(d) \big), \\ \Leftrightarrow \quad \rho(x - \beta_0) &>& \rho^{-k} \big(\beta_0 b_0 + C_k(d) \big); \end{aligned}$$

and (4.12), we have

$$x_1 > \rho^{-k} \big(\beta_0 b_0 + C_k(d) \big). \tag{4.13}$$

Now let us check the target variable. By assumption, we have

$$\beta_0 \in (0, 1) \text{ and } \rho > 1,$$

and

$$\begin{aligned}
\rho^k(1+\beta_0) - 2\beta_0 &\geq 0, \\
\Leftrightarrow \quad \beta_0\rho^k - 2\beta_0 + \rho^k &\geq 0, \\
\Leftrightarrow \quad \beta_0 c\rho^{k+1} - 2\beta_0 c\rho + c\rho^{k+1} &\geq 0, \\
\Leftrightarrow \quad x\rho^{k+1} - c\rho^{k+1} - \rho x + \beta_0 c\rho &\leq \beta_0 c\rho^{k+1} - \beta_0 c\rho + x\rho^{k+1} - \rho x, \\
\Leftrightarrow \quad x\rho^{k+1} - c\rho^{k+1} - \rho x + \beta_0 c\rho &\leq (\rho x + \beta_0 c\rho)\left(\rho^k - 1\right), \\
\Leftrightarrow x + x(\rho^{k+1} - 1) - c\rho^{k+1} - C_k(d) - \rho x + \beta_0 c\rho & \\
&\leq x_1\left(\rho^k - 1\right) - C_k(d).
\end{aligned} \tag{4.14}$$

It can now be seen that $y_1 = x + y - x_1 \leq$ left side of (4.14) which, combined with (4.13), leads to

$$(x_1, y_1) \in R_k.$$

With the help of the formula of total probability and the inductive assumptions it can now be checked that

$$\begin{aligned}
V_{k+1}\left(x, y, \pi^f(k+1)\right) &= P^{\pi^f(k+1)}\left(x_j > b_0, 0 \leq j \leq k, y_{k+1} \leq 0 | (x,y)\right) \\
&= 1,
\end{aligned} \tag{4.15}$$

for all $(x, y) \in R_{k+1}$. ■

Remark 4.1 Theorem 4.1 *means for all* $e_0 = (x, y) \in R_n$, *there exists a riskless policy* $\pi^f(\pi)$ *such that the investor can reach target level with probability 1. Therefore,* R_n *is called the risk free zone.*

Let

$$\begin{aligned}
D_0 &= \left\{(x, u), 0 \leq x < b_0, y > 0\right\} \text{ and more generally define} \\
D_n &= \left\{(x, y) : 0 \leq x \leq c_0\beta_0 + \rho^{-1}b_0, y > 0, (x, y) \in D_{n-1}\right\}, \quad n \geq 1.
\end{aligned}$$

The above set D_n is called the *ruin zone.*

Finally, define

$$\begin{aligned}
J_0 &= \{(x, y) : x > b_0, y > 0\}, \\
J_n &= \{(x, y) : x > \rho^{-n}(b_0 + c_n(d)), y > x\Delta_n - C_n(d), (x, y) \in J_{n-1}\}, \\
&\qquad n \geq 1.
\end{aligned}$$

The set J_n is called the *risk zone.*

In the risk zone J_N, if the investor follows the riskless policy, he will fail to reach the target value at stage N and he would not be ruined. Thus he will have to allocate some part of his wealth to the risky stock asset in order to maximize the probability of reaching target level at stage N.

References

[1] D.P. Bertsekas and S.E. Shreve, *Stochastic Optimal Control: the Discrete Time Case*, Academic Press, New York, 1978.

[2] S. Browne, Survival and growth with a liability: Optimal portfolio strategies in continuous time. *Math. OR*, 22: 468–493, 1997.

[3] J.X. Cvitamic and I. Karatzas, Hedging and Portfolio Optimization under Transaction Costs: A Martingale Approach. *Math. Finance*, 6: 133–166, 1996.

[4] Zeqing Dong and Ke Liu, Structure of optimal policies for discounted semi-Markov decision programming with unbounded rewards, *Scientia Sinica*, Series A, XXIX: 337–349, 1986.

[5] J.A. Filar, L.C.M. Kallenberg and H-M. Lee, Variance-Penalized Markov Decision Processes, *MOR*, 14: 147–161, 1989.

[6] J.A. Filar, D. Krass and K.W. Ross, Percentile Performance Criteria For Limiting Average Markov Decision Processes, *IEEE AC*, 40: 2–10, 1995.

[7] O. Hernandez-Lerma, *Adaptive Markov Control Processes*, Springer-Verlag, New York, 1989.

[8] O. Hernandez-Lerma and J.B. Lasserre, *Discrete-Time Markov Control Problems: Basic Optimality Criterion*, Springer-Verlag, New York, 1996.

[9] C.J. Himmelberg, T. Parthasarathy and F.S. Van Vleek, Optimal plan for dynamic programming problems, *Math. OR*, 390–394, 1976.

[10] I. Karatzas, Optimal portfolio and consumption decisions for a small investor on a Finite Horizon, *SIAM J.Control Optim.*, 25: 1557–1586, 1987.

[11] R.E. Kloeden and E. Platen, *Numerical Solution of Stochastic Differential Equations*, Springer-Verlag, New York, 1992.

[12] H.J. Kushner and P.G. Dupuis, *Numerical Methods for Stochastic Control Problems in Continuous Time*, Spring-Verlag, New York, 1992.

[13] H. Markowitz, *Portfolio Selection: Efficient Diversification of Investment*, J. Wiley and Sons, 1959.

[14] R.C. Merton, *Continuous-Time Finance*, Oxford and Cambridge: Basic Blackwell, 1990.

[15] A.D. Roy, Safety first and the holiday of assets, *Econometrica*, 220: 431–449, 1996.

[16] S. Roy, Theory of dynamic portfolio for survival under uncertainty, *Mathematical Social Sciences*, 30: 171-194, 1995.

[17] D.J. White Mean, Variance, and Probabilistic criterion in Finite Markov Decision Processes: A Review, *JOTA*, 56: 1–29, 1988.

[18] C. Wu and Y. Lin, Minimizing Risk Models in Markov Decision Processes with Policies Depending on Target Values, *J. of Math. Anal. and Appl.*, 231: 47–67, 1999.

[19] S. Yu, Y. Lin and P. Yan, Optimization Models for the Finite Arrival Target Distribution Function in Discrete Time, *J. of Math. Anal. and Appl.*, 225: 193–223, 1998.

Chapter 27

LONG TERM AVERAGE CONTROL OF A LOCAL TIME PROCESS*

Marta S. Mendiondo

Richard H. Stockbridge

Abstract This paper provides a tractable numerical method for long-term average stochastic control problems in which the cost includes a term based on the local time process of a diffusion. The control problem is reformulated as a linear program over the set of invariant distributions for the process. In particular, the long-term average local time cost is expressed in terms of the invariant distribution. Markov chain approximations are used to reduce the infinite-dimensional linear programs to finite-dimensional linear programs and conditions for the convergence of the optimal values are given.

Keywords: Linear programming, stochastic control, numerical approximation, long-term average criterion, local time process.

1. Introduction

The aim of this paper is to provide a tractable numerical method for a class of stochastic control problems in which the decision criterion includes a cost involving the local time process of a diffusion. We consider a long-term average criterion and reformulate the control problem as a linear program over the set of invariant distributions for the state and control process. In particular, this involves reformulating the long-term average cost for the local time process in terms of the invariant distribution. Markov chain approximations are used to reduce the infinite-dimensional linear programs to finite-dimensional linear programs and

*This research is partially supported by NSF under grant DMS 9803490.

Z. Hou et al. (eds.), Markov Processes and Controlled Markov Chains, 425–441.

the corresponding solutions are shown to converge to the optimal solution of the original problem.

We focus on a class of problems which arise in the modelling of semi-active suspension systems. The state $x(t) = \big(x_1(t), x_2(t)\big)$ satisfies a degenerate stochastic differential equation

$$\begin{aligned} dx_1(t) &= x_2(t)dt \\ dx_2(t) &= b\big(x(t), u(t)\big)\, dt + \sigma\, dW(t) \end{aligned} \tag{1.1}$$

in which $b(x,u) = -\big(\gamma_1 x_1 + u x_2 + \gamma_2 \text{sign}(x_2)\big)$, u is the control process taking values in some interval $U = [\underline{u}, \overline{u}]$ with $\underline{u} > 0$, W is a standard Brownian motion and $\sigma > 0$. The objective is to minimize

$$J(u) = \limsup_{t\to\infty} \frac{1}{t} E\left[\int_0^t c_1(x(s), u(s))\, ds + \int_0^t c_2(x(s))\, d\Lambda^{(2)}(s;0)\right] \tag{1.2}$$

in which $\Lambda^{(2)}(\cdot\,; x)$ denotes the local time process of $x_2(\cdot)$ at x; and c_1 and c_2 are nonnegative, bounded and continuous.

This model is obtained from a one-degree-of-freedom shock absorber system with dry friction in which $y = x_1(t)$ is the relative displacement and satisfies the equation,

$$m\ddot{y} + v\dot{y} + Ky + F\text{sign}(\dot{y}) = m\ddot{e}. \tag{1.3}$$

In this system, the control v is the shock absorber damping constant; $Ky + F\text{sign}(\dot{y})$ represents the restoring force, including the dry friction term; and $\ddot{e}$ is the random input of the system due to the road surface. The system (1.1) is obtained by setting $x_2 = \dot{y}$, $\gamma_1 = K/m$, $\gamma_2 = F/m$ and $u = v/m$.

This model has previously been studied by Campillo [2], Campillo, Le Gland and Pardoux [3] and Heinricher and Martins [5]. In [2, 3], the running cost was taken to be the absolute acceleration squared, $c_1(x,u) = |\gamma_1 x_1 + u x_2 + \gamma_2 \text{sign}(x_2)|^2$, and the local time process did not enter the cost. Heinricher and Martins, on the other hand, introduced the local time process in the cost function but replaced the long-term average criterion with a discounted criterion $\int_0^\infty e^{-\alpha s} c\, d\Lambda(s;0)$. In each of these papers, the authors used dynamic programming techniques with a Markov chain approximation of the original stochastic processes. Heinricher and Martins raised the question of how to determine the long-term average cost involving the local time process.

The motivation for including a cost based on the local time process arises from a particular analysis of the smoothness of the ride. Bumpy rides occur when the velocity makes significant changes in amplitude and direction. Consider a band of width 2ϵ centered at 0 for the velocity. A "bump" occurs when the velocity cycles from below $-\epsilon$ to above ϵ and back below $-\epsilon$. If the "level of discomfort" (the cost) of a cycle over $[-\epsilon, \epsilon]$ is proportional to ϵ, then the local time process arises in the limit as $\epsilon \to 0$ (see [10]). The reader is referred to [5] for additional motivation and explanation.

Costs associated with local time processes also arise in the heavy traffic diffusion limit for queueing systems. In this setting, the local time processes on the boundaries of the regions correspond to wasted capacity.

The main contributions of this paper are the analysis of long-term average control problems involving costs based on the local time process and the use of equivalent linear programming formulations in the solution. The reformulation of stochastic control problems as equivalent infinite-dimensional linear programming problems is given under very general conditions in [1], [7] and [11]. This paper uses the same approach but indicates how to include costs associated with the local time process.

The remainder of this paper is organized as follows. In the next section, we reformulate the stochastic control problem as an equivalent linear programming problem over the space of invariant distributions. Section 3 discusses the Markov chain approximations and convergence of the approximating solutions. The last section displays numerical examples.

2. Linear programming formulation

We consider, for the class of admissible controls, the set of transition functions $\eta : \mathbb{R}^2 \times \mathcal{B}[\underline{u}, \overline{u}] \to [0,1]$ for which the mean $u(x) = \int u\, \eta(x, du)$, as a function of the state x, satisfies the condition that there exists a finite number of submanifolds of $\mathbb{R}^2$ with dimension less than or equal to 1 outside of which u is continuous. The transition function η gives the conditional distribution on the control space $[\underline{u}, \overline{u}]$ given the state x and as such, is considered a randomized or relaxed control. Since the control enters linearly into the dynamics in (1.1), the mean $u(\cdot)$ is an admisssible control in the sense considered in [3]. Denote the collection of admissible controls by $\mathcal{U}$.

We begin by characterizing the invariant distributions for the processes. We require several results given in paper [3], in which the uniqueness of the invariant measure μ_η for each admissible η and the existence of a density with respect to Lebesgue measure for this measure are proved. We summarize this as a proposition and refer the reader to [3, Propositions 2.3, 2.6 and Lemma 2.5].

Proposition 2.1 *For any $\eta \in \mathcal{U}$, the diffusion process* (1.1) *admits a unique invariant measure μ_η on $\mathbb{R}^2$ which has a density $p(x)$ with respect to Lebesgue measure for which $p(x) > 0$ for almost every x.*

Theorem 2.1 (Characterization of invariant distributions)
Let η be an admissible control in $\mathcal{U}$. Then $\mu_\eta \in \mathcal{P}(\mathbb{R}^2)$ is an invariant distribution for the diffusion process (1.1) *having control η if and only if for each $f \in C_c^2(\mathbb{R}^2)$,*

$$\int\int Af(x_1, x_2, u)\, \eta(x_1, x_2, du)\, \mu_\eta(dx_1 \times dx_2) = 0, \tag{2.1}$$

where

$$Af(x_1, x_2, u) = x_2 f_{x_1}(x_1, x_2) + b(x, u) f_{x_2}(x_1, x_2) + (1/2)\sigma^2 f_{x_2 x_2}(x_1, x_2). \tag{2.2}$$

Proof We begin by showing the necessity of μ_η satisfying (2.1).

Let f be a twice-continuously differentiable function having compact support. Then Itô's formula implies that the quantity,

$$f(x_1(t), x_2(t)) - \int_0^t \int_{[\underline{u}, \overline{u}]} Af\big(x_1(s), x_2(s), u\big)\, \eta\big(x_1(s), x_2(s), du\big)\, ds, \tag{2.3}$$

is a martingale. Define the average occupation measure μ_t, for $t > 0$, to satisfy for each bounded, continuous function ϕ,

$$\begin{aligned} &\int \phi(x_1, x_2, u)\, \mu_t(dx_1 \times dx_2 \times du) \\ &= \frac{1}{t} E\left[\int_0^t \int_{[\underline{u}, \overline{u}]} \phi\big(x_1(s), x_2(s), u\big)\, \eta\big(x_1(s), x_2(s), du\big)\, ds\right]. \end{aligned}$$

Claim: *The collection of occupation measures $\{\mu_t : t > 0\}$ is tight and hence relatively compact.*

Proof By Lemma 2.1 of [3], there exists some constant C such that $E\big[|x(t)|^2\big] < C$ for all t and controls $\eta \in \mathcal{U}$. Since the space of controls $[\underline{u}, \overline{u}]$ is compact, given $\epsilon > 0$, by choosing $K > (C/\epsilon)^{1/2}$ an application of Markov's inequality shows that

$$\mu_t(\overline{B_K} \times [\underline{u}, \overline{u}]) > 1 - \epsilon, \qquad \forall t > 0,$$

in which B_K denotes the ball of radius K in $\mathbb{R}^2$ centered at the origin, which proves the claim. ∎

Since $\{\mu_t\}$ is relatively compact, uniqueness of the invariant measure implies that $\mu_t \Rightarrow \mu$ as $t \to \infty$, where

$$\mu(dx_1 \times dx_2 \times du) = \eta(x_1, x_2, du)\, \mu_\eta(dx_1 \times dx_2). \tag{2.4}$$

The fact that (2.3) is a martingale implies that

$$\begin{aligned}\int Af(x_1, x_2, u)\, \mu_t(dx_1 \times dx_2 \times du) \\ = \frac{1}{t}\left(E\Big[f\big(x_1(t), x_2(t)\big)\Big] - E\Big[f\big(x_1(0), x_2(0)\big)\Big]\right),\end{aligned}$$

and so letting $t \to \infty$, the invariant measure μ satisfies

$$\int Af(x_1, x_2, u)\, \mu(dx_1 \times dx_2 \times du) = 0 \tag{2.5}$$

for every $f \in C_c^2(\mathbb{R}^2)$. Note that we have used the fact that the set $\{(x_1, x_2) : x_2 = 0\}$ is a μ-null set in passing to the limit since Af is only discontinuous on this set.

To show sufficiency, let μ_η be any measure satisfying (2.1) and define μ as in (2.4). Theorem 2.2 of [7] gives the existence of a stationary process $x(t)$ for which the pair $(x(t), \eta(x(t), \cdot))$ makes (2.3) a martingale and μ is the one-dimensional distribution

$$\begin{aligned}E\left[\int_{[\underline{u},\overline{u}]} \phi(x_1(t), x_2(t), u)\, \eta\big(x_1(t), x_2(t), du\big)\right] \\ = \int \phi(x_1, x_2, u)\, \mu(dx_1 \times dx_2 \times du).\end{aligned}$$

A modification of Theorem 5.3.3 of [4] to include control then implies x is a solution of (1.2). ■

We now turn to expressing the long-term average cost associated with a control η in terms of η and the invariant measure μ_η.

Theorem 2.2 (Evaluation of the long-term average cost)
Let η be an admissible control and μ_η the corresponding invariant measure. Then (1.2) *is equal to*

$$\int \int_{[\underline{u},\overline{u}]} \left[c_1(x, u) - \frac{1}{2} c_2(x)\, \mathrm{sign}(x_2) b(x, u)\right] \eta(x_1, x_2, du)\, \mu_\eta(dx_1 \times dx_2).$$

Proof We concentrate on the contribution to the cost due to the local time process; the absolutely continuous cost term has been evaluated to be $\int c_1(x,u)\,\mu(dx \times du)$ in [7]. The key to reformulating the local time cost is the Tanaka-Meyer formula (see Karatzas and Shreve [6, page 220]):

$$|x_2(t)| = |x_2(0)| + \int_0^t \operatorname{sign}(x_2(s))\,dx_2(s) + 2\Lambda^{(2)}(s;0).$$

It then follows that

$$\begin{aligned}
&\int_0^t c_2(x(s))d\Lambda^{(2)}(s;0) \\
&= \frac{1}{2}\int_0^t c_2(x(s))d|x_2(s)| \\
&\quad -\frac{1}{2}\int_0^t \int_{[\underline{u},\overline{u}]} c_2(x(s)) \operatorname{sign}\big(x_2(s)\big)b(x(s),u)\,\eta(x(s),du\big)\,ds \\
&\quad -\frac{1}{2}\int_0^t c_2\big(x(s)\big)\sigma\operatorname{sign}\big(x_2(s)\big)\,dW(s)
\end{aligned}$$

which implies that

$$\begin{aligned}
&\frac{1}{t}E\left[\int_0^t c_2\big(x(s)\big)d\Lambda^{(2)}(s;0)\right] \\
&= \frac{1}{t}E\left[\int_0^t c_2\big(x(s)\big)\,d|x_2(s)|\right] \\
&\quad -\frac{1}{2}\int c_2(x)\,\operatorname{sign}(x_2)\,b(x,u)\,\mu_t(dx_1 \times dx_2 \times du).
\end{aligned}$$

Letting $t \to \infty$, Lemma 2.1 of [3] and the fact that c_2 is bounded imply that the first term becomes negligible yielding

$$\begin{aligned}
&\limsup_{t\to\infty}\frac{1}{t}E\left[\int_0^t \int_{[\underline{u},\overline{u}]} c_1\big(x(s),u\big)\,\eta\big(x(s),du\big)\,ds + \int_0^t c_2\big(x(s)\big)\,d\Lambda^{(2)}(s;0)\right. \\
&= \int \left[c_1(x,u) - \frac{1}{2}c_2(x)\,\operatorname{sign}(x_2)b(x,u)\right]\mu(dx_1 \times dx_2 \times du)
\end{aligned}$$

■

2.1 LP formulation

The results of Theorems 2.1 and 2.2 imply the following theorem.

Theorem 2.3 *The stochastic control problem of minimizing* (1.2) *over all solutions* $x(t)$ *of* (1.1) *is equivalent to finding a probability measure* μ *which minimizes*

$$\int \left[c_1(x,u) - \frac{1}{2}c_2(x)\,\mathrm{sign}(x_2)b(x,u)\right] \mu(dx_1 \times dx_2 \times du)$$

subject to the constraints that for each $f \in C_c^2(\mathbb{R}^2)$,

$$\int Af(x_1,x_2,u)\,\mu(dx_1 \times dx_2 \times du) = 0. \tag{2.6}$$

Moreover, an optimal control η^* *is given by the conditional distribution on the control space of an optimal measure* μ^**; that is,* η^* *is optimal if* μ^* *is optimal and*

$$\mu^*(dx_1 \times dx_2 \times du) = \eta^*(x_1,x_2,du)\,\mu_{\eta^*}(dx_1 \times dx_2 \times [\underline{u},\overline{u}]).$$

This is an infinite-dimensional linear program over the space of invariant distributions of the system (1.1).

3. Markov chain approximations

In order to obtain a numerical solution, it is necessary to reduce the LP problem to finite dimensions. We accomplish this by discretizing the state and control spaces and taking finite difference approximations to the differential operators. This follows the approach of the previous papers in that the approximating operators can be viewed as the generators of finite-state Markov chains which approximate the diffusion process. The approximating LP then gives the associated long-term average cost of these Markov chains.

Though our approximations are the same as in the previous works, our convergence results are based on the LP formulation rather than dynamic programming arguments. A similar approach was used in the setting of a compact state space in [9].

For each $n \geq 1$, let $h^{(n)}, k^{(n)}, m^{(n)} > 0$ denote the discretization sizes and let $K_1^{(n)}$ and $K_2^{(n)}$ be truncation limits of $\mathbb{R}^2$ in the x_1 and x_2 coordinates, respectively, where for simplicity we assume $K_1^{(n)} = M_1^{(n)} \cdot n \cdot h^{(n)}$, $K_2^{(n)} = M_2^{(n)} \cdot n \cdot k^{(n)}$ and $b-a = M_3^{(n)} \cdot n \cdot m^{(n)}$ for some positive integers $M_1^{(n)}, M_2^{(n)}$ and $M_3^{(n)}$. We assume that as $n \to \infty$, $K_1^{(n)}, K_2^{(n)} \to \infty$ and $h^{(n)}, k^{(n)}, m^{(n)} \searrow 0$. To simplify the notation, we will drop the superscript n from discretization parameters and discretized spaces and points in these spaces. Define

$$\begin{aligned} E &= E_1 \times E_2 \\ &= \{y = (y_1,y_2) = (ih, jk) : -M_1 n \leq i \leq M_1 n, -M_2 n \leq j \leq M_2 n\} \end{aligned}$$

to be the discretization of the state space $\mathbb{R}^2$ and

$$V = \{\underline{u} + lm : 0 \leq l \leq M_3 n\}$$

to be the discretization of the control space $U = [\underline{u}, \overline{u}]$.

Recall the generator of the two-dimensional diffusion is given by (2.2). For $(y_1, y_2) \in E$, (y_1, y_2) not on the boundary, and $v \in V$, we use the following approximations:

$$f_{x_1}(y_1, y_2) \approx \begin{cases} \dfrac{f(y_1 + h, y_2) - f(y_1, y_2)}{h}, & \text{if } y_2 \geq 0, \\ \dfrac{f(y_1 - h, y_2) - f(y_1, y_2)}{h}, & \text{if } y_2 < 0, \end{cases}$$

$$f_{x_2}(y_1, y_2) \approx \begin{cases} \dfrac{f(y_1, y_2 + k) - f(y_1, y_2)}{k}, & \text{if } b(y_1, y_2, v) \geq 0, \\ \dfrac{f(y_1, y_2 - k) - f(y_1, y_2)}{k}, & \text{if } b(y_1, y_2, v) < 0, \end{cases}$$

and

$$f_{x_2 x_2}(y_1, y_2) \approx \frac{f(y_1, y_2 + k) + f(y_1, y_2 - k) - 2f(y_1, y_2)}{k^2}.$$

Substituting into (2.2) and collecting the terms involving the test function at the various states in E we get (for (y_1, y_2) in the interior)

$$\begin{aligned} A_n f(y_1, y_2, v) &= \left(\frac{(y_2)^+}{h}\right) f(y_1 + h, y_2) + \left(\frac{(y_2)^-}{h}\right) f(y_1 - h, y_2) \\ &\quad + \left(\frac{b(y_1, y_2, v)^+}{k} + \frac{\sigma^2}{2k^2}\right) f(y_1, y_2 + k) \\ &\quad + \left(\frac{b(y_1, y_2, v)^-}{k} + \frac{\sigma^2}{2k^2}\right) f(y_1, y_2 - k) \\ &\quad - \left(\frac{|y_2|}{h} + \frac{|b(y_1, y_2, v)|}{k} + \frac{\sigma^2}{k^2}\right) f(y_1, y_2). \end{aligned} \tag{3.1}$$

A_n is the generator of a continuous time, finite state Markov chain. We need to define the generator on the boundary of E. We adopt the approach of Kushner and Dupuis [8] by initially allowing the Markov chain to exit E according to (3.1) then projecting the state onto the nearest point in E. This has the effect that the Markov chain becomes "sticky" at the boundary in that the state could transit to itself before it moves to another point in the space.

Due to the fact that $\sigma > 0$, this Markov chain is aperiodic and irreducible and thus has a unique invariant distribution for each choice of control policy.

We observe that as $n \to \infty$, $\sup_{(y,v) \in E \times V} |A_n f(y, v) - Af(y, v)| \to 0$.

3.1 Approximating LP

The approximating linear program could, in fact, have been determined simply by discretizing the spaces and taking the finite difference approximations to the differential operators. However, we will use the knowledge of the underlying stochastic processes to provide existence and uniqueness of feasible measures for the approximating LP for given admissible control policies.

We state the linear program (for each n) for the approximating Markov chains which "approximates" the original LP given in Theorem 2.3.

Approximating linear programs Find a probability measure ν on $E \times V$ which minimizes

$$\int \left[c_1(y_1, y_2, v) - \frac{1}{2}c_2(y_1, y_2)\ \mathrm{sign}(y_2)\, b(y_1, y_2, v)\right] \nu(dy_1 \times dy_2 \times dv) \tag{3.2}$$

subject to the constraints that for each $f \in C_c^2(\mathbb{R}^2)$,

$$\int A_n f(y_1, y_2, v)\, \nu(dy_1 \times dy_2 \times dv) = 0. \tag{3.3}$$

Moreover, an optimal control η^* is given by the conditional distribution on the control space of an optimal measure ν^*; that is, η^* is optimal if ν^* is optimal and

$$\nu^*(dy_1 \times dy_2 \times dv) = \eta^*(y_1, y_2, dv)\, \nu_{\eta^*}(dy_1 \times dy_2).$$

Since there are only a finite number of states in the approximating problem, it is only necessary to evaluate (3.3) for a finite number of functions f. For each point $(y_1, y_2) = (i, j) \in E$, consider a function $f_{ij} \in C_c^2(\mathbb{R}^2)$ such that $f_{ij}(i, j) = 1$; and for all (x_1, x_2) for which either $|x_1 - i| \geq h/2$ or $|x_2 - j| \geq k/2$, $f_{ij}(x_1, x_2) = 0$. This choice of f has the effect of taking $f_{ij} \in C_c^2(\mathbb{R}^2)$ so that when restricted to $E_1 \times E_2$ the function is the indicator of the point (i, j). As a result, the constraints (3.3) of the approximating LP become

$$\int A_n I_{\{(i,j)\}}(y_1, y_2, v)\, \nu(dy_1 \times dy_2 \times dv) = 0$$

for each $(i, j) \in E$.

3.2 Convergence results

It is necessary to relate controls for the original problem to controls for the approximating problems and vice versa in order to establish the convergence results.

Observe that $E \times V \subset \mathbb{R}^2 \times U$; that is, the discretized space for the controlled Markov chain is a subset of the space for the original diffusion process. One aspect of this imbedding is that we can view each $\nu \in \mathcal{P}(E \times V)$ for the approximating problem as a probability measure on $\mathbb{R}^2 \times U$. Our goal, however, is to define controls for the approximating LP corresponding to each admissible control and also to define an admissible control for the original problem for each control of the approximating LP.

Define the mapping $\phi_n^1 : \mathbb{R} \to E_1$ by

$$\phi_n^1(x) = \begin{cases} -M_1nh, & \text{for } x < -M_1nh + h/2, \\ ih, & \text{for } ih - h/2 \le x < ih + h/2, \\ & \quad -M_1n + 1 \le i \le M_1n - 1, \\ M_1nh, & \text{for } M_1nh - h/2 \le x, \end{cases}$$

similarly define $\phi_n^2 : \mathbb{R} \to E_2$, and finally define $\phi_n^3 : U \to V$ by

$$\phi_n^3(u) = \begin{cases} \underline{u}, & \text{for } u < \underline{u} + m/2, \\ \underline{u} + km, & \text{for } \underline{u} + km - m/2 \le u < \underline{u} + km + m/2, \\ & \quad 1 \le k \le M_3n - 1, \\ \overline{u}, & \text{for } \overline{u} - m/2 \le u. \end{cases}$$

The function $\Phi_n = (\phi_n^1, \phi_n^2)$ takes the partition of $\mathbb{R}^2$ consisting of rectangles with each rectangle containing exactly one point of E and maps the points of the rectangle to this point of the discretization. In like manner, ϕ_n^3 maps each interval in U to the unique point in the discretization V contained in the interval. We observe that as $n \to \infty$,

$$\sup_{u \in U} |u - \phi_n^3(u)| \to 0 \qquad \text{and} \qquad |x - \Phi_n(x)| \to 0 \quad \text{for each } x \in \mathbb{R}^2.$$

Let η be an admissible control for the original problem. Define the corresponding control η_n for the approximating problems by setting

$$\eta_n(y, \{v\}) = \eta(y, (\phi_n^3)^{-1}(\{v\})). \tag{3.4}$$

Note that each control on the discretized space $E \times V$ is a transition function $\eta_n : E \times \mathcal{B}(V) \to [0, 1]$.

We now start with a control η_n for the approximating LP and extend it to an admissible control $\overline{\eta_n}$ on $\mathbb{R}^2 \times U$. First extend η_n to a transition function on $E \times U$ by setting $\eta_n(E \times V^c) = 0$. Now require $\overline{\eta_n}$ to satisfy

$$\int_U h(u)\, \overline{\eta_n}(x_1, x_2, du) = \int_U h(u)\, \eta_n\left(\phi_n^1(x_1), \phi_n^2(x_2), du\right) \tag{3.5}$$

for each $h \in C(U)$. The control $\overline{\eta_n}$ is piecewise constant.

Turning to convergence, the paper [3] establishes several results concerning the Markov chain and the diffusion process. In particular, they prove the following convergence result (see Lemma 4.8) about the invariant distributions of the approximating Markov chains and the original diffusion processes.

Proposition 3.1 *For each admissible control η, let μ_η denote the invariant measure corresponding to η; and define η_n by* (3.4) *and ν_{η_n} to be the invariant measure of the Markov chain satisfying* (3.3). *Then $\nu_{\eta_n} \Rightarrow \mu_\eta$.*

We now use this result to show that the optimal cost of the approximating LP provides an asymptotic lower bound on the value of the optimal cost of the original diffusion. We also show that if the optimal controls of the approximating LPs converge to an admissible control then, in fact, the approximating optimal costs converge to the optimal cost of the original LP and the limiting control is optimal. The first result establishes that for each admissible control η and induced controls η_n, the costs for the approximating problems converge to the cost of the original problem.

Proposition 3.2 *For each admissible η, let η_n be given by* (3.4) *and let ν_{η_n} and μ_η denote the invariant distributions satisfying* (3.3) *and* (2.6), *respectively. Then*

$$\int\int \left[c_1(y,v) + c_2(y,v)\,\text{sign}(y_2)\,b(y,v)\right] \eta_n(y,dv)\,\nu_{\eta_n}(dy)$$
$$\to \int\int \left[c_1(x,u) + c_2(x,u)\,\text{sign}(x_2)\,b(x,u)\right] \eta(x,du)\,\mu_\eta(x,du).$$

Proof This follows immediately from Proposition 3.1 and the fact that $\mu_\eta\{x_2 = 0\} = 0$. ■

We now use this result to establish the asymptotic lower bound on the optimal cost.

Theorem 3.1 *Let $\nu_n^* \in \mathcal{P}(E \times V)$ denote an optimal invariant measure for the approximating LP problem. Then*

$$\limsup_{n\to\infty} \int \left[c_1(y,v) + c_2(y,v)\,\text{sign}(y_2)\,b(y,v)\right] \nu_n^*(dy \times dv)$$
$$\leq \inf_{\eta\in\mathcal{U}} \int\int \left[c_1(x,u) + c_2(x,u)\,\text{sign}(x_2)\,b(x,u)\right] \eta(x,du)\,\mu_\eta(dx).$$

Proof For each admissible η, let μ_η denote the invariant distribution corresponding to η. Let η_n be given by (3.4) and ν_{η_n} be the invariant distribution corresponding to η_n. Then by optimality of ν_n^*,

$$\int \left[c_1(y,v) + c_2(y,v)\,\mathrm{sign}(y_2)\,b(y,v)\right] \nu_n^*(dy \times dv)$$
$$\leq \int \left[c_1(y,v) + c_2(y,v)\,\mathrm{sign}(y_2)\,b(y,v)\right] \eta_n(y,dv)\,\nu_{\eta_n}(dy).$$

It follows that

$$\limsup_{n\to\infty} \int \left[c_1(y,v) + c_2(y,v)\,\mathrm{sign}(y_2)\,b(y,v)\right] \nu_n^*(dy \times dv)$$
$$\leq \lim_{n\to\infty} \int\int \left[c_1(y,v) + c_2(y,v)\,\mathrm{sign}(y_2)\,b(y,v)\right] \eta_n(y,dv)\,\nu_{\eta_n}(dy)$$
$$= \int\int \left[c_1(x,u) + c_2(x,u)\,\mathrm{sign}(x_2)\,b(x,u)\right] \eta(x,du)\,\mu_\eta(dx),$$

and the result follows upon taking the infimum over the admissible controls. ∎

Finally, we consider the case in which the optimal controls of the approximating LPs converge to an admissible control for the original problem.

Theorem 3.2 *Let η_n^* denote an optimal control for the approximating LP and define $\overline{\eta_n^*}$ by (3.5). Suppose there exists an admissible η^* for the original problem such that*

$$\overline{\eta_n^*}(x,\cdot) \Rightarrow \eta^*(x,\cdot)$$

for almost every x (in Lebesgue measure). Then η^* is an optimal control for the original problem.

Proof Let $\mu_{\overline{\eta_n^*}}$ and μ_{η^*} denote the invariant distributions corresponding to $\overline{\eta_n^*}$ and η^*, respectively. The proof of Lemma 4.8 of [3] establishes the tightness of $\{\mu_{\overline{\eta_n^*}}\}$ (in fact, it establishes tightness for the collection of all invariant distributions of admissible controls). As a result, by defining the measures $\mu_n \in \mathcal{P}(\mathbb{R} \times U)$ as $\mu_n(dx \times du) = \overline{\eta_n^*}(x,du)\mu_{\overline{\eta_n^*}}(dx)$, it immediately follows that $\{\mu_n\}$ is tight and hence relatively compact. Thus there exists some subsequence $\{n_k\}$ and some measure μ which is a weak limit of μ_{n_k}. For simplicity of notation, we may assume $\{n_k\}$ is the entire sequence.

Let μ_0 denote the marginal of μ and η a regular conditional distribution of u given x under μ so that

$$\mu(dx \times du) = \eta(x, du)\mu_0(dx).$$

Then for every bounded continuous h,

$$\begin{aligned}\int\int h(x,u)\,\overline{\eta_n^*}(x,du)\,\mu_{\overline{\eta_n^*}}(dx) \quad &\to \quad \int h(x,u)\,\mu(dx \times du) \\ &= \quad \int\int h(x,u)\,\eta(x,du)\,\mu_0(dx)\end{aligned}$$

Since the Bounded Convergence theorem implies

$$\left|\int h(x,u)\,\overline{\eta_n^*}(x,du) - \int h(x,u)\eta^*(x,du)\right| \to 0 \qquad a.e.\ x,$$

it follows that

$$\int\int h(x,u)\,\eta^*(x,du)\,\mu_{\overline{\eta_n^*}}(dx) \to \int\int h(x,u)\,\eta(x,du)\,\mu_0(dx). \tag{3.6}$$

Since $\mu_n \Rightarrow \mu$ implies $\mu_{\overline{\eta_n^*}} \Rightarrow \mu_0$, the continuous mapping theorem [4, Corollary 3.1.9] implies

$$\int\int h(x,u)\,\eta^*(x,du)\,\mu_{\overline{\eta_n^*}}(dx) \to \int\int h(x,u)\,\eta^*(x,du)\,\mu_0(dx). \tag{3.7}$$

Comparing (3.6) and (3.7) and writing $h(x,u) = h_1(x)h_2(u)$, we have, for every bounded continuous h_1 and h_2,

$$\int h_1(x) \int h_2(u)\,\eta^*(x,du)\,\mu_0(dx) = \int h_1(x) \int h_2(u)\,\eta(x,du)\,\mu_0(dx),$$

which implies that $\int h_2(u)\,\eta^*(x,du) = \int h_2(u)\,\eta(x,du)$ for almost every x and hence

$$\eta^*(x,\cdot) = \eta(x,\cdot) \qquad a.e.\ \ x.$$

Since the invariant distribution for this control is unique, $\mu^* = \mu_0$ and hence, for every bounded continuous h,

$$\int\int h(x,u)\,\overline{\eta_n^*}(x,du)\,\mu_{\overline{\eta_n^*}}(dx) \to \int\int h(x,u)\,\eta^*(x,du)\,\mu^*(dx).$$

The continuous mapping theorem again implies

$$\begin{aligned}&\lim_{n\to\infty} \int\int \left[c_1(x,u) + c_2(x,u)\,\text{sign}(x_2)\,b(x,u)\right]\,\overline{\eta_n^*}(x,du)\,\mu_{\overline{\eta_n^*}}(dx) \\ &\quad = \quad \int\int \left[c_1(x,u) + c_2(x,u)\,\text{sign}(x_2)\,b(x,u)\right]\,\eta^*(x,du)\,\mu_{\eta^*}(dx)\end{aligned}$$

and the result follows from Theorem 3.1. ■

4. Numerical example

We now illustrate the LP methods using a particular choice for the parameters. We consider the case in which the only cost is that associated with the local time process and assume that the cost rate c_2 is constant. Thus, $c_1(x, u) \equiv 0$ and $c_2(x) \equiv c_2$ for some constant c_2. We also restrict the model by assuming $\gamma_2 = 0$. This is the model studied by Heinricher and Martins with a discounted criterion.

We implemented our numerical approximation in SAS.

4.1 Test case

To test the accuracy of the numerical solution, we further restrict the model by fixing $\gamma_1 = 0$ and only allow a single control value $u = 1$. Thus the dynamics are reduced to

$$\begin{aligned} dx_1(t) &= x_2(t)dt \\ dx_2(t) &= -x_2(t)dt + \sigma dW(t). \end{aligned}$$

We take $c_2 = 2$ to compensate for the fraction $1/2$ in the objective function. The objective function for the test case is

$$\int |x_2| \, \mu(dx_1 \times dx_2).$$

In this test case, it is clear that only the x_2-process is important to the analysis. This process is an Ornstein-Uhlenbeck process for which the invariant distribution is unique and easily determined to be normally distributed with mean 0 and variance $\sigma^2/2$ and the objective function value is $\dfrac{\sigma}{2\sqrt{\pi}}$.

Figure 27.1 illustrates the results of the numerical approximations when $\sigma = 3$ using discretization size $h = k = 0.6, 0.3$ and 0.1 together with the $N(0, 3^2/2)$ density function. In addition, Table 27.1 presents the objective function values obtained in these three cases. It is very clear looking at this data that the approximating invariant distributions as well as the approximating objective value are close to the invariant distribution and objective value of the original process.

4.2 General example

In the general setting, $\gamma_1 \neq 0$ and the control is not fixed. We selected $\gamma = 2$ and $\sigma = 2$ and used a discretization size of $h = k = 0.2$ over the truncated square $[-2, 2] \times [-2, 2]$. We chose $[\underline{u}, \overline{u}] = [0.5, 1.5]$ and allowed the control to take the values 0.5, 1.0 and 1.5. Figure 27.2 illustrates

Table 27.1. Objective Function Values

	Mesh Size	Objective Function Value
Approximating Markov Chain	$h = .6$	.89189
	$h = .3$	.87102
	$h = .1$	.85505
Diffusion Process		.84628

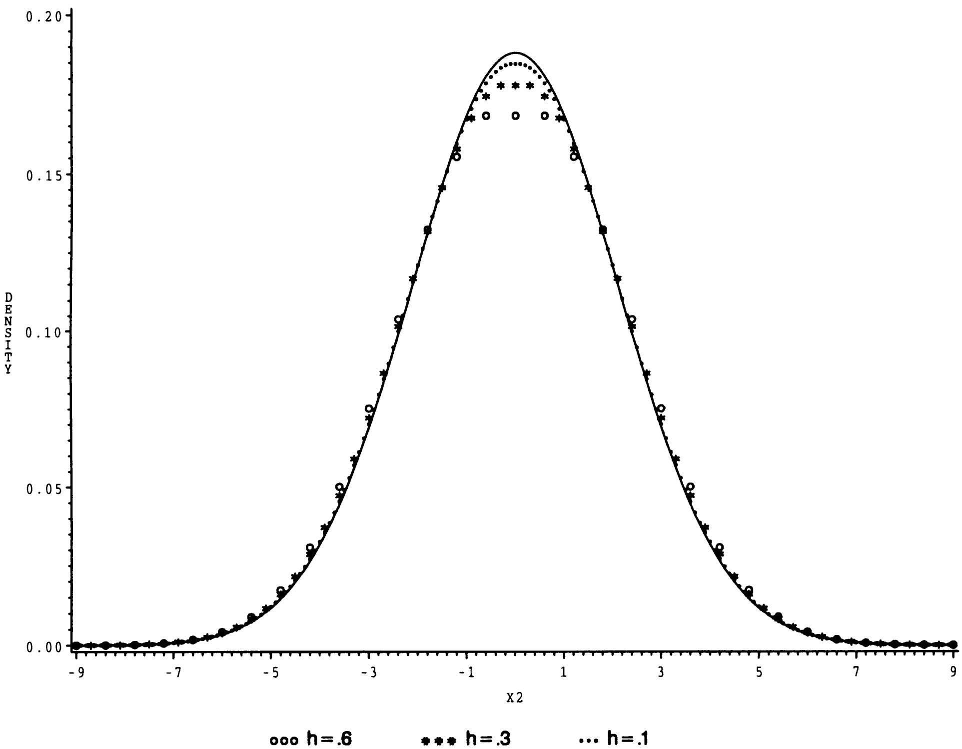


Figure 27.1. Invariant distributions for the test case

the resulting optimal control. Notice that the optimal control takes the smallest possible value of u whenever $x_2 \neq 0$. The only change in control occurs where $x_2 = 0$. This behavior of the optimal control was consistent throughout both coarser and finer discretizations of the state and control spaces.

We conjecture that an optimal control for the diffusion process is to use the maximum control whenever the velocity is zero and to use minimum control otherwise.

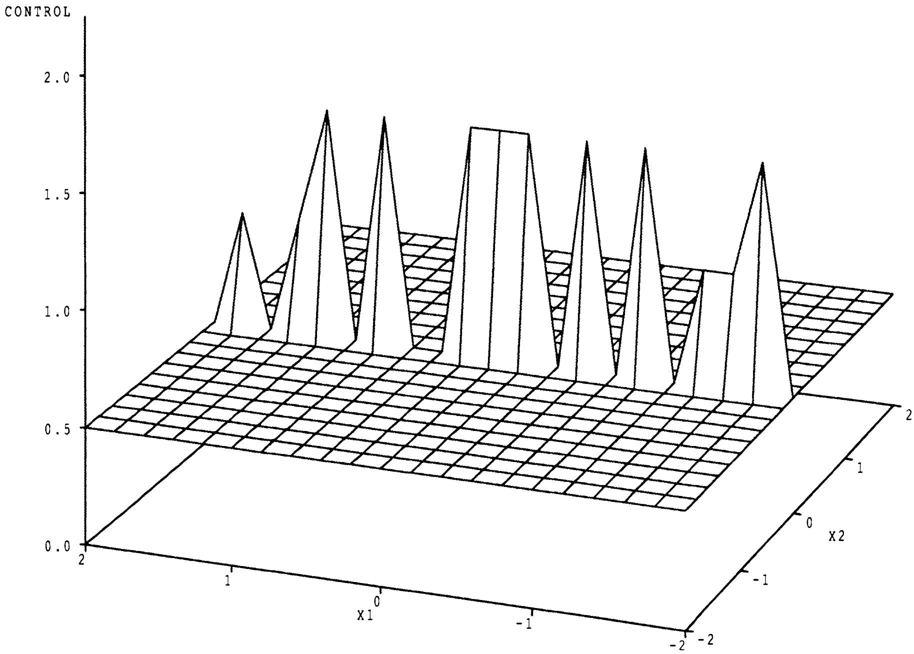


Figure 27.2. Optimal control for general example, $h = k = .02$, $u = 0.5, 1, 1.5$

References

[1] A.G. Bhatt and V.S. Borkar, Occupation measures for controlled Markov processes: Characterization and optimality, *Ann. Probab.*, **24** (1996), 1531–62.

[2] F. Campillo, Optimal ergodic control for a class of nonlinear stochastic systems application to semi-active vehicle suspensions, *Proc. 28th Conf. Dec. Cont.*, (1989), 1190–1195.

[3] F. Campillo, F. Le Gland and E. Pardoux, Approximation of a stochastic ergodic control problem, *New trends in nonlinear control theory (Nantes, 1988),* Lect. Notes in Control and Inform. Sci. **122**, Springer, Berlin (1989), 379–395.

[4] S.N. Ethier and T.G. Kurtz, *Markov Processes: Characterization and Convergence*, Wiley, New York (1986).

[5] A.C. Heinricher and F. Martins, Optimal control for the local time of a diffusion: numerical approximation, in the *Stochastic Dynamics and Reliability of Nonlinear Ocean Systems,* DE Vol. 77, ASME 1994, 7–14.

[6] I. Karatzas and S.E. Shreve, *Brownian Motion and Stochastic Calculus*, Springer-Verlag, New York, (1988).

[7] T.G. Kurtz and R.H. Stockbridge, Existence of Markov Controls and Characterization of Optimal Markov Controls, *SIAM J. Cont. Optim.*, **36** (1998), 609–653.

[8] H.J. Kushner and P.G. Dupuis, *Numerical Methods for Stochastic Control Problems in Continuous Time*, Springer-Verlag, New York (1992).

[9] M.S. Mendiondo and R.H. Stockbridge, Approximation of Infinite-Dimensional Linear Programming Problems which Arise in Stochastic Control, *SIAM J. Cont. Optim.* **36** (1998), 1448–1472.

[10] D. Revus and M. Yor, *Continuous martingales and Brownian motion*, Springer-Verlag, New York, 1991.

[11] R.H. Stockbridge, Time-average control of martingale problems: A linear programming formulation, *Ann. Probab.*, **18** (1990), 206–217.

Chapter 28

SINGULARLY PERTURBED HYBRID CONTROL SYSTEMS APPROXIMATED BY STRUCTURED LINEAR PROGRAMS*

A. Haurie
Logilab-HEC,
University of Geneva,
Switzerland.

F. Moresino
Logilab-HEC,
University of Geneva,
Switzerland and
Cambridge University,
UK.

J.-P. Vial
Logilab-HEC,
University of Geneva,
Switzerland.

1. Introduction

The aim of this tutorial paper is to present the relationship that exists between the control of singularly perturbed hybrid stochastic systems and the decomposition approach in structured linear programs. The mathematical sophistication is voluntarily kept at a low level by

*This research has been supported by The Swiss Science Foundation (FNRS # 1214-057093.99).

Z. Hou et al. (eds.), Markov Processes and Controlled Markov Chains, 443–463.

avoiding the full development of the theorems demonstrations that can be found in papers already published or to appear shortly. On another hand, since it corresponds to a new application of a convex optimization method that has been successfully applied in other contexts, we give a rather detailed account of the decomposition technique used in the numerical approximation method and of the comparison with a direct linear programming method.

The class of systems we are interested in are characterized by an hybrid state, the continuous part evolving in a fast time scale according to controlled diffusion processes and the discrete part evolving in a slow time scale as a finite state jump process. The diffusion and the jump processes are coupled. We consider the ergodic control of such a system. To illustrate this type of structure we propose an economic production model where the continuous state corresponds to stocks of production factors whereas the discrete state describes different market structures that determine the demand for the produced good. When the ratio between the fast and slow time scale tends to 0 the problem becomes singular. However, under sufficient ergodicity assumptions, one can exploit the fact that, between two successive jumps of the slow process, the fast diffusion process has enough time to reach a steady state, or, more precisely an invariant state probability measure. This permits us to define a limit control problem in the form of a finite state controlled Markov chain that is well behaved and gives a good approximation of the optimal value when the time scale ratio is close to zero.

When we implement a numerical approach, the singular perturbation generally yields an ill-conditioned problem. This is the case when one uses an approximation by controlled Markov chains as these chains will exhibit strong and weak interactions. But here again, we can identify a limit problem that is well conditioned and which yields a good approximation to the solution when the time scale ratio is close to 0. Furthermore, the limit problem yields to a structured block angular linear program that is amenable to an efficient decomposition technique. The decomposition technique implements a dialogue between a master program that distributes a dual information obtained at the analytical center of a localization set and an oracle that proposes cutting planes obtained via a policy iteration algorithm run on a local reduced size MDP.

The paper is organized as follows. In Section 2 we recall the theory of ergodic control for a singularly perturbed, two-time scale hybrid stochastic system. In Section 3 we give the main result concerning the definition of a limit control problem for a class of well behaved feedback controls. This limit control problem is a finite state Markov decision process that gives a good approximation of the optimal value when the time scale ratio tends to 0. For this we rely mostly on Filar and Haurie, 1997 [6]. In Section 4 we recall the numerical technique that can be used for the solution of an ergodic control of an hybrid stochastic system. The fundamental reference is Kushner and Dupuis, 1992 [12] where an approximating controlled Markov chain is used to compute numerically the solution of stochastic control problems. In Section 5 we observe that the approximating controlled Markov chain has also the structure of a singularly perturbed Markov decision process (MDP) with strong and weak interactions. This is the occasion to recall the results of Abbad, 1991 [1] and Abbad et al., 1992 [2] showing that the limit control problem for the singularly perturbed MDP can be formulated as a structured block-angular linear program. We are able to show the close similarity between the limit control problem defined in Section 3 and the structured LP obtained in the numerical approach.

In Section 6 we implement a decomposition technique for the solution of the limit control problem, using the analytic center cutting plane method (ACCPM), initially proposed in Goffin et al., 1992 [9] as a general method for solving nondifferentiable convex programming problems.

2. A two-time-scale hybrid stochastic control system

In this section we describe a control system characterized by an hybrid state (y, ζ) where the continuous state variable y is "moving fast" according to a diffusion process while the discrete state variable ζ is "moving slowly" according to a continuous time stochastic jump process. The diffusion and controlled processes are coupled.

2.1 The dynamics

We consider a hybrid control system described by the hybrid controlled process $(y, \zeta)(\cdot)$ where $y(\cdot)$ is "moving faster" than $\zeta(\cdot)$ and described formally by the stochastic state equation.

$$\begin{aligned} \varepsilon dy(t) &= f^{\zeta(t)}(y(t), v(t))dt + \sqrt{\varepsilon}\sigma^{\zeta(t)}\, dw(t), \\ v(t) &\in U^{\zeta(t)}. \end{aligned}$$

More precisely, we consider the following specifications:

- A set of Itô equations

$$\begin{aligned} \varepsilon dy(t) &= f^i(y(t), v(t))dt + \sqrt{\varepsilon}\sigma^i dw(t), \\ v(t) &\in U^i, \end{aligned}$$

where $y \in \text{int}(X)$, the interior of a compact subset of $\mathbb{R}^n$, is the *continuous* state variable and $i \in E$, a given finite set, is the *discrete state variable.* For each $i \in E$, the control constraint set U^i is a compact set, σ^i is an $n \times n$ matrix which, for simplicity, is taken as diagonal, the function $f^i(y, v)$ is continuous in both arguments and $\{w(t) : t \geq 0\}$ is an n-dimensional Wiener process.

- The perturbation parameter ε is a positive scalar which will eventually tend to 0. It can be viewed as the ratio between the fast and the slow time scales.

- Some *reflecting* boundary conditions, as those detailed in [12], section 1.4 are imposed on ∂X.

- For each pair $(i, j) \in E \times E$, $i \neq j$, let be given a continuous function $q_{ij}(y, v)$, where $v \in U^i$, is the *conditional transition rate* from i to j of a jump process $\{\zeta(t) : t \geq 0\}$. We assume that the following holds

  $$\mathrm{P}[\zeta(t + dt) = j | \zeta(t) = i, \ y(t) = y, \ v(t) = v] = q_{ij}(y, v)dt + o(dt),$$

  where
  $$\lim_{dt \to 0} \frac{o(dt)}{dt} = 0$$
  uniformly in (y, v).

- For each $i \in E$, let $L^i(y, v)$ be a continuous function of (y, v) describing the cost rate for the control system.

The class $\mathcal{U}$ of admissible controls is the set of $\mathcal{F}_t$-adapted processes $\{v(t) : t \geq 0\}$, where $\{\mathcal{F}_t : t \geq 0\}$ is the σ-field describing the history of the hybrid process $\{(y, \xi)(t) : t \geq 0\}$ and $v(t) \in U^i$, whenever $\xi(t) = i$.

2.2 Change of time scale

It will be convenient to work with a "stretched out" time scale, by defining the trajectory

$$(x(t), \xi(t)) = (y(\varepsilon t), \zeta(\varepsilon t)).$$

The process dynamics now become

$$dx(t) = f^{\xi(t)}\big(x(t), u(t)\big)dt + \sigma^{\xi(t)}dz(t),$$

where $z(t) = \frac{1}{\sqrt{\varepsilon}} w(\varepsilon t)$, and the $\xi(\cdot)$ process has transition rates given by $\varepsilon q_{ij}(x, u)$. The differential operator of the $(x, \xi)(\cdot)$ process is denoted

A^u and defined by

$$\begin{aligned}(A^u\phi)(x,\xi) \\ = \quad \frac{\partial}{\partial x}\phi(x,\xi)f^{\xi}(x,u) + \frac{1}{2}\frac{\partial^2}{\partial x^2}\phi(x,\xi)(\sigma^{\xi})^2 + \varepsilon\sum_{j\in E} q_{\xi,j}(x,u)\phi(x,j).\end{aligned}$$

2.3 A restricted control class and the associated performance criterion

The class of controls $\mathcal{U}$ is too large for our purpose and we shall consider a restricted class of controls defined as follows.

For each $i \in E$ let Θ^i be a compact parameter set. With each $\theta \in \Theta^i$ is associated a *piecewise continuous feedback admissible control*, denoted $\tilde{v}_\theta(\cdot) : \mathbb{R}^n \mapsto U^i$. We assume that this feedback controls varies continuously with θ. A policy is a mapping $\gamma : E \mapsto \Theta^i$. Once a policy is chosen, the feedback control used between two successive random times t_n and t_{n+1} is defined by

$$v(t) \quad = \quad \tilde{v}_{\gamma(\zeta(t_n))}(y(t)).$$

We assume that the process has good ergodicity properties and in particular

Assumption 2.1 *For each admissible feedback control $\tilde{u}_\theta(\cdot)$, $\theta \in \Theta^i$, the set of functions $\{A^{\tilde{u}_\theta}(\cdot), h(\cdot) \in C_0^2(X)\}$ is measure determining. Equivalently, the equation*

$$(\nu^i(t), h) \quad = \quad (\nu^i(0), h) + \int_0^t (A_i^{\tilde{u}_\theta} h, \nu^i(s))\, ds, \quad h(\cdot) \in C_0^2(X)$$

has a unique weakly continuous probability measure valued solution $\nu^i_{\tilde{u}_\theta}(\cdot)$ for each initial (probability measure) condition $\nu^i(0)$.

It will be convenient to permit randomization of the parameter choice at any decision time. We consider that for each $i \in E$ the parameter set Θ^i belongs to a probability space and that a policy associates with each possible mode $i \in E$ a probability distribution $m(i, d\theta)$ over Θ^i.

Associated with an admissible randomized policy $m(\cdot)$ we define the *long term average reward*

$$\begin{aligned}J_\varepsilon(m(\cdot)) \\ = \quad \liminf_{T\to\infty} \mathrm{E}_{m(\cdot)}\left[\frac{1}{T}\int_0^T \int_{\Theta^{\xi(t)}} L^{\xi(t)}(x(t), \tilde{u}_\theta(x(t))m(\xi(t), d\theta)\, dt | x(0), \xi(0)\right].\end{aligned}$$

We are interested in the behavior of the infimum value

$$J_\varepsilon^* = \sup_{m(\cdot)} J_\varepsilon(m(\cdot)), \tag{2.1}$$

when $\varepsilon \to 0$.

3. Convergence to a limit-control problem

In this section we define a limit-control problem, when the time-scale ratio ε tends to zero and recall the convergence theorem obtained in Filar and Haurie, 1997 [6]. This theorem asserts that the optimal average reward of the perturbed problem converges to the optimal average reward of the limit-control problem, when the time-scale ratio tends to zero.

3.1 The fixed-ξ control process

Consider the fixed-ξ control process $x(\cdot|\xi)$, when $\xi = i$, which is associated with the Itô equation

$$dx(t) = f^i(x(t), u(t))dt + \sigma^i dz(t). \tag{3.1}$$

$$u(t) \in U^i \tag{3.2}$$

The differential operator of the fixed-ξ process is denoted A_i^u and defined by

$$(A_i^u\psi)(x) = \frac{\partial}{\partial x}\psi(x)f^i(x,u) + \frac{1}{2}\frac{\partial^2}{\partial x^2}\psi(x)(\sigma^i)^2.$$

Assumption 3.1 *For each admissible feedback control $\tilde{u}_\theta(\cdot)$, $\theta \in \Theta^i$ and each initial condition $x(0) = x_0$, (3.1) has a unique, weak sense, solution and a unique invariant measure $\nu_\theta^i(\cdot)$.*

3.2 The limit-control problem

We make the following strong ergodicity assumption on the ξ-process.

Assumption 3.2 *For any x and vector $(u_i)_{i\in E} \in \Pi_{i\in E}U^i$ the discrete state continuous time Markov chain with transition rates $q_{ij}(x, u_i)$ has a single recurrence class.*

For each possible discrete state $i \in E$ we consider the fixed-ξ controlled diffusion process associated with an admissible feedback control $\tilde{u}_\theta(\cdot)$. Its Itô equation is

$$dx(t) = f^i(x(t), \tilde{u}_\theta(x(t)))dt + \sigma^i dz(t).$$

According to Assumption 3.1 there corresponds an invariant measure on $\mathbb{R}^n$, denoted $\nu_\theta^i(dx)$, such that

$$\int_X \nu_\theta^i(dx) \;=\; 1.$$

We can then construct a Markov Decision Process (MDP) with state space E and action space Θ^i, $i \in E$, where the transition rates and the cost rates are given by

$$\begin{aligned} \bar{B}_{ij}(\theta) &= \int_X q_{ij}(x, \tilde{u}_\theta(x))\nu_\theta^i(dx), \quad i, j \in E \\ \bar{L}(i,\theta) &= \int_X L^i(x, \tilde{u}_\theta(x))\nu_\theta^i(dx), \quad i \in E, \end{aligned}$$

respectively. Now, due to the strong ergodicity property of Assumption 3.2, we can associate with a randomized policy $m(\cdot)$, an invariant measure on E denoted $\{\mu^i(m);\, i \in E\}$ and verifying

$$\begin{aligned} 0 &= \sum_{i\in E} \mu^i(m) \int_{\Theta^i} \bar{B}_{ij}(\theta) m(i, d\theta) \\ 1 &= \sum_{i\in E} \mu^i(m). \end{aligned}$$

Since we are interested in the limiting behavior as $\varepsilon \to 0$, the natural *Limit-Control Problem* to solve is the following finite state ergodic cost *MDP*

$$\begin{aligned} \bar{J}^* &= \inf_m \bar{J}(m) \\ &= \inf_m \sum_{i\in E} \mu^i(m) \int_{\Theta^i} \bar{L}(i,\theta) m(i, d\theta). \end{aligned}$$

The two following assumptions are needed to insure convergence in the space of probability measures.

Assumption 3.3 *For any sequence of admissible randomized policies* $\{m^\varepsilon(\cdot)\}$, *there is a function* $0 \le \hat{g}(x) \to \infty$ *when* $|x| \to \infty$, *a* $K_1 < \infty$ *and* $\Delta_\varepsilon \to 0$ *as* $\varepsilon \to 0$ *such that* $\frac{\varepsilon}{\Delta_\varepsilon} \to 0$ *and*

$$\sup_{t\geq 0} \frac{1}{\Delta_\varepsilon} \int_t^{t+\Delta_\varepsilon} \mathrm{E}\left[\hat{g}\left(x^\varepsilon(s)\right)\right] ds \;\le\; K_1 \;<\; \infty.$$

Assumption 3.4 *For each $\varepsilon_n > 0$ of a decreasing sequence $\varepsilon_n \to 0$ there is an optimal (for the reward functional* (2.1)*) admissible randomized policy $m^*_{\varepsilon_n}(\cdot)$ such that the corresponding set $\{(x^*_{\varepsilon_n}(\cdot), \xi^*_{\varepsilon_n}(\cdot)), n = 0, 1, \dots\}$ is tight.*

The proof of the following theorem can be found in [6].

Theorem 3.1 *Under* Assumptions 2.1–3.4 *the following holds*

$$\lim_{\varepsilon \to 0} |\bar{J}^* - J^*_\varepsilon| = 0.$$

This result means that the optimal average reward for the perturbed hybrid problem converges, when ε tends to zero, to the optimal average reward of the limit-control problem.

4. Numerical approximation scheme

In this section, following Kushner and Dupuis, 1992 [12], we propose a numerical approximation technique for the hybrid control problem. Then, following Abbad et al., 1992, [2], we derive a limit problem when ε tends to zero.

4.1 The Markov decision problem

The ergodic cost stochastic control problem identified in the previous section is an instance of the class of controlled switching diffusion studied by Ghost et al., 1997 [8]. The dynamic programming equations, established in the previous reference as a necessary optimality condition take the form

$$J = \max_{u \geq 0} \Big\{ L^i(x,u) + \varepsilon \sum_{j \neq i} q_{ij}(x,u)[V(x,j) - V(x,i)] + \frac{\partial}{\partial x} V(x,i) f^i(x,u) + \frac{1}{2}\sigma^2 \frac{\partial^2}{\partial x^2} V(x,i) \Big\}, \quad i \in E \quad (4.1)$$

where $V(x,\cdot)$ is C^2 in x for each i in E and represents a potential value function and J is the maximal expected reward growth rate.

This system of Hamilton-Jacobi-Bellman (HJB) equations cannot, in general, be solved analytically. However a numerical approximation technique can be implemented following a scheme described in [12]. The space of the continuous state is discretized with mesh h. That means that the variable x_k belongs to the grid $\mathcal{X}_k = \{x_k^{\min}, x_k^{\min} + h, x_k^{\min} + 2h, \dots, x_k^{\max}\}$. Denote e_k the unit vector on the x_k axis and $\mathcal{X} = \mathcal{X}_1 \times \mathcal{X}_2 \times \cdots \times \mathcal{X}_K$. We approximate the first partial derivatives by finite differences, taken "in the direction of the flow", as follows:

$$\frac{\partial}{\partial x_k}V(x) \to \begin{cases} \frac{V(x+e_kh)-V(x)}{h} & \text{if } \dot{x}_k \geq 0 \\ \frac{V(x)-V(x-e_kh)}{h} & \text{if } \dot{x}_k < 0. \end{cases} \tag{4.2}$$

The second partial derivatives are approximated by

$$\frac{\partial^2}{\partial x_k^2}V(x) \to \frac{V(x+e_kh)+V(x-e_kh)-2V(x)}{h^2}. \tag{4.3}$$

We define the interpolation interval as

$$\Delta t_h = \frac{h^2}{\tilde{Q}_h},$$

where

$$\begin{aligned} Q_h(x,i,u) &= \varepsilon q^i(x,u)h^2 + \sum_{k=1}^{K}\left\{\sigma_k^2 + h\left|f_k^i(x,u)\right|\right\}, \\ q^i(x,u) &= \sum_{j\neq i} q_{ij}(x,u) \end{aligned}$$

and

$$\tilde{Q}_h = \max_{x,i,u} Q_h(x,i,u).$$

We define transitions probabilities to neighboring grid points as follows

$$p_h[(x,i),(x\pm e_kh,,i)|u] = \frac{\frac{\sigma_k^2}{2} + hf^i(x,u)^{\pm}}{\tilde{Q}_h}, \tag{4.4}$$

$$p_h[(x,i),(x,j)|u] = \varepsilon h^2\frac{q_{ij}(x,u)}{\tilde{Q}_h} \qquad i\neq j, \tag{4.5}$$

$$p_h[(x,i),(x,i)|u] = 1-\frac{Q(x,i,u)}{\tilde{Q}_h}, \tag{4.6}$$

where $f^i(x,u)^+ = \max\{f^i(x,u);0\}$ and $f^i(x,u)^- = \max\{-f^i(x,u);0\}$. The other transitions probabilities are equal to zero. The possible transitions are represented in Figure 28.1, for an example with card(E)=3 and $K=2$.

If we substitute in the HJB-equations (4.1) the finite differences (4.2) and (4.3) to the partial derivatives, after regrouping terms and using the

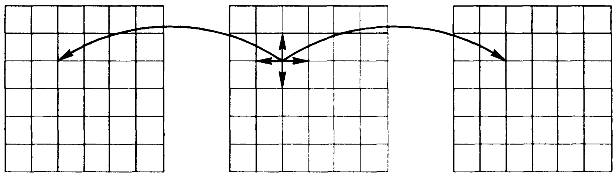


Figure 28.1. Transitions in the grid set.

transition probabilities (4.4), (4.5) and (4.6), we can formulate the following associated discrete state MDP dynamic programming equation:

$$\begin{aligned} g_h \Delta t_h + W(x,i) &= \max_{u \geq 0} \Big\{ \sum_{x'} p_h[(x,i),(x',i)|u] W(x',i) \\ &\quad + \sum_{j \neq i} p_h[(x,i),(x,j)|u] W(x,j) + \Delta t_h L^i(x,u) \Big\}, \\ &\quad x \in \mathcal{X}, \quad i \in E. \end{aligned} \tag{4.7}$$

In this discrete state MDP, the term g approximates the maximal expected reward growth rate J and the functions $W(x,j)$ approximate, in the sense of weak convergence, the potential value functions $V(x,j)$. Solving this MDP gives thus a numerical approximation to the solution of the HJB-equation (4.1).

If we discretize the space of the control with mesh h_u ($u_k \in \mathcal{U}_k = \{u_k^{\min}, u_k^{\min} + h_u, u_k^{\min} + 2h_u, \ldots, u_k^{\max}\}$),we obtain an MDP with finite state and action spaces. The optimal control law of this MDP can be obtained through the solution of the following linear program (see de Ghellinck, 1960 and Manne, 1960 [4] and [14]):

$$\max \sum_i \sum_x \sum_u L^i(x,u) Z^i(x,u) \tag{4.8}$$

s.t.

$$\sum_i \sum_x \sum_u G_h^\varepsilon[(x,i),(x',j)|u] Z^i(x,u) = 0 \quad x' \in \mathcal{X},\ j \in E \tag{4.9}$$

$$\sum_i \sum_x \sum_u Z^i(x,u) = 1 \tag{4.10}$$

$$Z^i(x,u) \geq 0, \tag{4.11}$$

where $G_h^\varepsilon[(x,i),(x',j)|u]$ denotes the generator of the MDP, defined as follows:

$$G_h^\varepsilon[(x,i),(x',j)|u] = \begin{cases} p_h[(x,i),(x,i)|u] - 1 & \text{if } (x,i) = (x',j) \\ p_h[(x,i),(x',j)|u] & \text{otherwise.} \end{cases}$$

Then the steady state probabilities will be defined as

$$P[x,i] = \sum_u Z^i(x,u)$$

and the conditional steady-state probabilities, given a mode i are

$$P[x|i] = \frac{\sum_u Z^i(x,u)}{\sum_x \sum_u Z^i(x,u)}.$$

One should notice that the linear program (4.8-4.11) will tend to be ill-conditioned when ε tends to be small since coefficients with difference of an order of magnitude $\frac{1}{\varepsilon}$ appear in the same constraints.

4.2 The limit Markov decision problem

The generator of the MDP can be written

$$G_h^\varepsilon[(x,i),(x',j)|u] = B_h[(x,i),(x',j)|u] + \varepsilon D_h[(x,i),(x',j)|u] + o(\varepsilon),$$

where $B_h[(x,i),(x',j)|u]$ is the generator of a completely decomposable MDP, with card(E) subprocesses which do not communicate one with the other (*i.e.* if $i \neq j \quad B_h[(x,i),(x',j)|u] \equiv 0 \quad \forall x, x'$) and $\varepsilon D_h[(x,i),(x',j)|u]$ is a perturbation that links together these card(E) sub-blocks.

For singularly perturbed systems, the optimal solution of the limit MDP is, in general, different from the optimal solution of the initial MDP where ε has been replaced by zero. However, the theory developed by Abbad, Filar and Bielecki (in [1] and [2]) offers tools to handle the limit of singularly perturbed MDP. Concretely, when ε tends to zero the optimal control law of the MDP (4.7) can be obtained through the solution of the following linear program (see [2]):

$$\max \sum_i \sum_x \sum_u L^i(x,u) Z^i(x,u) \tag{4.12}$$

s.t.

$$\sum_x \sum_u B_h[(x,i),(x',i)|u] Z^i(x,u) = 0 \quad x' \in \mathcal{X},\ i \in E \qquad (4.13)$$

$$\sum_i \sum_{x'} \sum_x \sum_u D_h[(x,i),(x',j)|u] Z^i(x,u) = 0 \quad j \in E \qquad (4.14)$$

$$\sum_i \sum_x \sum_u Z^i(x,u) = 1 \qquad (4.15)$$

$$Z^i(x,u) \geq 0 \qquad (4.16)$$

Indeed this linear program exhibits a typical bloc-diagonal structure in the constraints (4.13). The constraints (4.14–4.15) are the so-called coupling constraints. In Section 5 we will apply a decomposition technique to exploit this structure. It should be noticed that the ill-conditioning has vanished since the variable ε doesn't appear in the linear program.

5. A decomposition approach for the limit MDP

In this section, following Filar and Haurie, 1997 [6] and Filar and Haurie, 2001 [7], we derive for the MDP (4.12–4.16) a decomposition approach which exploits the bloc-diagonal structure. We then explain how this decomposition can be implemented using the Analytic Center Cutting Plane Method (ACCPM).

5.1 The dual problem

The dual problem of the linear program associated with the limit MDP (4.12–4.16) writes

$$\min_{\psi,\phi,\Upsilon} \Upsilon \qquad (5.1)$$

s.t.

$$\Upsilon \geq L^i(x,u) - \sum_{x'} B_h[(x,i),(x',i)|u]\phi(x',i)$$
$$- \sum_j \sum_{x'} D_h[(x,i),(x',j)|u]\psi(j) \quad i \in E,\ x \in \mathcal{X},\ u \in \mathcal{U}. \qquad (5.2)$$

The constraint matrix in the left-hand-side of (5.2) has a special structure. The terms associated with the variables $\phi(x',i)$ form independent blocks along the main diagonal. The terms associated with the variables $\psi(j)$ form a rectangular matrix that links all the blocks together.

In this formulation we may also recognize the approach proposed in [2], under the name *Aggregation-Disaggregation.* Indeed, if we define the modified costs

$$\Pi(\psi, x, i, u) = L^i(x, u) - \sum_j \sum_{x'} D_h[(x, i), (x', j)|u]\psi(j) \tag{5.3}$$

then the expression (5.2) corresponds to a set of card(E) decoupled MDPs. More precisely, the problem can be rewritten as

$$\min_{\psi, \phi, \Upsilon} \Upsilon \tag{5.4}$$

s.t.

$$\Upsilon \geq \Pi(\psi, x, i, u) - \sum_{x'} B_h[(x, i), (x', i)|u]\phi(x', i)$$
$$i \in E,\ x \in \mathcal{X},\ u \in \mathcal{U}. \tag{5.5}$$

Now, for each $i \in E$, (5.5) defines a decoupled MDP with modified transition cost (5.3).

The formulation (5.4)–(5.5) is also amenable to Benders decomposition (see Benders, 1962 [3]). Indeed, fixing the variables $\psi(j)$, the minimization in $\phi(x', i)$ and Υ is equivalent to

$$\chi(\psi) \quad = \quad \max_{i \in E} \chi_i(\psi) \tag{5.6}$$

where the functions $\chi_i(\psi)$, given by

$$\chi_i(\psi) = \min_{\phi, \Upsilon} \Upsilon \tag{5.7}$$

s.t.

$$\Upsilon \geq \Pi(\psi, x, i, u) - \sum_{x'} B_h[(x, i), (x', i)|u]\phi(x', i) \quad x \in \mathcal{X}, u \in \mathcal{U} \tag{5.8}$$

are the value functions of card(E) independent ergodic MDPs with cost $\Pi(\psi, x, i, u)$ and transition kernel $B_h[(x, i), (x', i)|u]$. It is easy to show that the functions $\chi_i(\psi)$ are convex and so is $\chi(\psi)$ as the pointwise maximum of convex functions.

Since the functions are also optimal values of linear programs, one should realize that the optimal dual variables make it possible to compute a subgradients $X_i(\tilde{\psi}) \in \partial\chi_i$ at $\tilde{\psi}$, for $i \in E$ as well as a subgradient[1] $X(\tilde{\psi}) \in \partial\chi$ at $\tilde{\psi}$ with the property

$$\chi(\psi) \geq \chi(\tilde{\psi}) + \langle X(\tilde{\psi}), \psi - \tilde{\psi}\rangle. \tag{5.9}$$

[1]The elements of $\partial\chi$ can be computed as follows: Let $\mathcal{A}(\tilde{\psi}) \subset \{1, \ldots, \text{card}(E)\}$ be the set of indices of active functions, i.e. those that satisfy $\chi_i(\tilde{\psi}) = \chi(\tilde{\psi}) = \max_{i \in E} \chi_i(\tilde{\psi})$. Then $X(\tilde{\psi}) \in \partial\chi$ iff $X(\tilde{\psi}) = \sum_{i \in \mathcal{A}(\tilde{\psi})} \lambda_i X_i(\tilde{\psi})$, for $X_i(\tilde{\psi}) \in \partial\chi_i$ and $\lambda_i \geq 0$, $\sum_{i \in \mathcal{A}(\tilde{\psi})} \lambda_i = 1$.

The optimization problem $\min_\psi \chi(\psi)$ is convex and nondifferentiable. A procedure that computes the value $\chi(\psi)$ and the associated subgradient $X(\psi)$ is called an oracle. The subgradient inequality (5.9) defines a so-called *cutting plane.*

5.2 ACCPM

There are many possible approaches for solving the convex nondifferentiable problem $\min_\psi \chi(\psi)$ (see Goffin and Vial, 1999 [10] for a short survey of these methods). In the present case we used ACCPM, a method developed by Goffin et al. 1992 [9] around the concept of analytic center (Sonnevend 1988)[2]. This is a cutting plane method, in which the query points are the analytic centers of a shrinking sequence of *localization sets.* Let $\{\psi^n\}_{n\in N}$ be a set of query points at which the oracle has been called. The answers $\chi(\psi^n)$ and $X(\psi^n)$ define a piecewise linear approximation $\underline{\chi}^N : \mathbb{R}^{\mathrm{card}(E)} \to \mathbb{R}$ to the convex function χ,

$$\underline{\chi}^N(\psi) = \max_{n\in N} \{\chi(\psi^n) + \langle X(\psi^n), \psi - \psi^n\rangle\}. \tag{5.10}$$

Since $\chi(\psi) \leq \chi^N(\psi)$, any lower bound π_l

$$\pi_l \leq \min_\psi \underline{\chi}^N(\psi) = \min_{\zeta,\psi} \{\zeta | \zeta \geq \chi(\psi^n) + \langle X(\psi^n), \psi - \psi^n\rangle, \forall n \in N\} \tag{5.11}$$

is also a lower bound for $\chi(\psi)$.

On the other hand, the best solution in the generated sequence provides an upper bound π_u for the convex problem, i.e.

$$\pi_u = \min_{n\in N} \{\chi(\psi^n)\}. \tag{5.12}$$

For a given upper bound π_u, we call *localization set* the following polyhedral approximation

$$\mathcal{L}^N(\pi_u) = \{(\varsigma, \psi) : \ \pi \geq \zeta, \ \zeta \geq \chi(\psi^n) + \langle X(\psi^n), \psi - \psi^n\rangle, \ \forall n \in N\}. \tag{5.13}$$

Note that, for any optimal solution ψ^*, the pair $(\chi(\psi^*), \psi^*)$ belongs to all the sets $\mathcal{L}^N(\pi_u)$. The analytic center of $\mathcal{L}^N(\pi_u)$ is defined as the unique pair (χ, ψ) which maximizes the product of the distances to the $N + 1$ linear constraints defining (5.13).

We can now summarize the ACCPM algorithm for our special case.

[2]It is beyond the scope of this paper to detail ACCPM. Interested readers will find a full account of the theory in [10]. The implementation is described at length in the thesis of Du Merle 1995 [5], while a library of programs to implement the method is presented in [11].

1. Compute the analytic center $(\bar{\zeta}, \bar{\psi})$ of the localization set $\mathcal{L}^N(\pi_u)$ and an associated lower bound $\underline{\pi}$.

2. Call the oracle at $(\bar{\zeta}, \bar{\psi})$. The oracle returns $\chi(\bar{\psi})$ and an element $X(\bar{\psi}) \in \partial\chi(\bar{\psi})$ that defines a valid cutting plane.

3. Update the bounds:

   a) $\pi_u = \min\{\chi(\bar{\psi}), \pi_u\}$

   b) $\pi_l = \max\{\underline{\pi}, \pi_l\}$.

4. Update the localization set with the new upper bound and the new cutting plane.

These steps are repeated until a point is found such that $\pi_u - \pi_l$ falls below a prescribed optimality tolerance.

The above procedure introduces one cut at a time. However one should note that the oracle always computes the card(E) values $\chi_i(\bar{\psi})$ and the vectors $X_i(\bar{\psi})$. Furthermore the inequalities

$$\zeta \geq \chi_i(\tilde{\psi}) + \langle X_i(\tilde{\psi}), \, \psi - \tilde{\psi} \rangle$$

are valid in the sense that they do not exclude optimal solutions. This information can therefore be added to the definition of $\mathcal{L}^N(\pi_u)$ to accelerate convergence.

6. Example

We propose to study an example of a plant producing one good with the help of two production factors and subject to random changes of the market price. This example is a special instance of the class of the two-time-scale hybrid stochastic systems we presented in Section 2. The discrete variable ξ describes the state of the market, which influences the profit derived from the produced good. We suppose that we have four different market states, so the ξ-process takes value in the set $E = \{1, 2, 3, 4\}$. The continuous variable $x \in (\mathbb{R}^+)^2$ describes the accumulated stock of the two different production factors. More precisely, x_k, $k = 1, 2$ corresponds to the number of employees of type k.

The output is determined by a CES production function[3]

$$Y(x_1, x_2) = \left(\eta[x_1]^{-\beta} + (1-\eta)[x_2]^{-\beta}\right)^{-\frac{1}{\beta}},$$

[3]See, for example, Layard and Waters, 1978 [13] Section 9-4

where $-1 < \beta < \infty$ is the substitution parameter ($\beta \neq 0$) and $0 < \eta < 1$ is the distribution parameter. The profit rate structure is described by the function

$$\begin{aligned} & L^{\xi(t)}(x_1(t), x_2(t), u_1(t), u_2(t)) \\ &= c(\xi(t))Y(x_1(t), x_2(t)) - a_1x_1(t) - a_2x_2(t) - A_1x_1^2(t) - A_2x_2^2(t) \\ &\qquad - b_1u_1(t) - b_2u_2(t) - B_1u_1^2(t) - B_2u_2^2(t), \end{aligned}$$

where $c(i)$ is the selling price, given the market is in state $i \in E$, $a_kx_k(t) + A_kx_k^2(t)$ is a cost function, related to the holding of a stock $x_k(t)$ of employees and $b_ku_k(t) + B_ku_k^2(t)$ is a cost function related to the enrollment effort, $u_k(t)$, of new employees.

We assume that the the price is influenced by the level of production of the firm. We rank the 4 market states by increasing selling price and we suppose that only jumps to neighboring market states can occur. More precisely, the ξ-process transition rates are defined by

$$\begin{aligned} \varepsilon q_{i(i+1)}(x_1, x_2) &= \varepsilon\left(E_i - e_iY(x_1, x_2)\right) \\ \varepsilon q_{i(i-1)}(x_1, x_2) &= \varepsilon\left(D_i + d_iY(x_1, x_2)\right). \end{aligned}$$

The parameter ε is the time-scale ratio that will, eventually, be considered as very small. The positive terms e_i, E_i, d_i and D_i are parameters which depend on the market state $i \in E$. We see that the transition rate toward a highest market price is negatively correlated to the production level, whereas the transition rate toward a lowest market price is positively correlated to the production level.

The dynamics of the employees is described by

$$dx_k(t) = [u_k(t) - \alpha_kx_k(t)]dt + \sigma_kd\omega_k(t), \quad k = 1, 2.$$

We consider the set of parameter values given in Table 28.1. We solved the limit model, when ε tends to zero, with the decomposition method described in Section 5.

The steady state probabilities obtained from the solution of the limit control problem in the approximating MDP are shown in Figure 28.2. As expected, the higher the selling price the higher the production level. For comparison, we considered also the model associated with a fixed ξ-process, that is, the model where the selling price stays the same forever. For the fixed ξ-process, the steady state probabilities are shown in Figure 28.3. Given a market state, the production level is higher for the model associated with the fixed ξ-process than for the limit model. This comes from the fact that, when the price can change, the probability that it will increase, resp. decrease, is negatively, resp. positively,

Table 28.1. List of parameter values for the numerical experiments.

					e_i	=	0.002			$\forall i \in E$
ν	=	1.0			E_i	=	0.4			$\forall i \in E$
η	=	0.5			d_i	=	0.004			$\forall i \in E$
β	=	−0.6			D_i	=	0.15			$\forall i \in E$
a_1	=	a_2	=	0.4	α_1	=	α_2	=	0.05	
A_1	=	A_2	=	0.004	σ_1	=	σ_2	=	3.0	
b_1	=	b_2	=	0	$x_1^{\max}$	=	$x_2^{\max}$	=	100	
B_1	=	B_2	=	0.05	$x_1^{\min}$	=	$x_2^{\min}$	=	0	
$c(1)$	=	1.3			h	=	10/3			
$c(2)$	=	1.6			$u_1^{\max}$	=	$u_2^{\max}$	=	10	
$c(3)$	=	1.9			$u_1^{\min}$	=	$u_2^{\min}$	=	0	
$c(4)$	=	2.2			h_u	=	2			

correlated with the production level. The effect of the production level on the price can be seen in Figure 28.4. In this Figure, we displayed, for the limit model, the steady state probabilities as a function of the state for two policies, namely the optimal policy and the optimal policy of the model with fixed ξ-process (note that this second policy is, in general, not optimal for the limit model). We see distinctly that the price tends to be higher in the first case (where the production level is lower) than in the second case.

The maximal expected reward growth rate J equals 27.6. The potential value functions are shown in Figure 28.5, for the case when the market is in the state $i = 3$. For the other states, the value functions are similar and therefore not displayed.

The optimal policy for the enrollment of new employees is shown in Figure 28.6, when the market is in state $i = 3$. For the other states, the optimal policies are similar and therefore not displayed.

7. Concluding remarks

In this paper we have implemented a decomposition method for the resolution of hybrid stochastic models with two time scales. This method, which was proposed by Filar and Haurie, 1997 [6] and by Filar and Haurie, 2001 [7], reformulates the initial problem as an approximating singularly perturbed MDP that can be solved as a structured linear programming problem. The originality of this paper was the coupling of ACCPM with a policy improvement algorithm to achieve a decomposition in order to exploit the special bloc-diagonal structure.

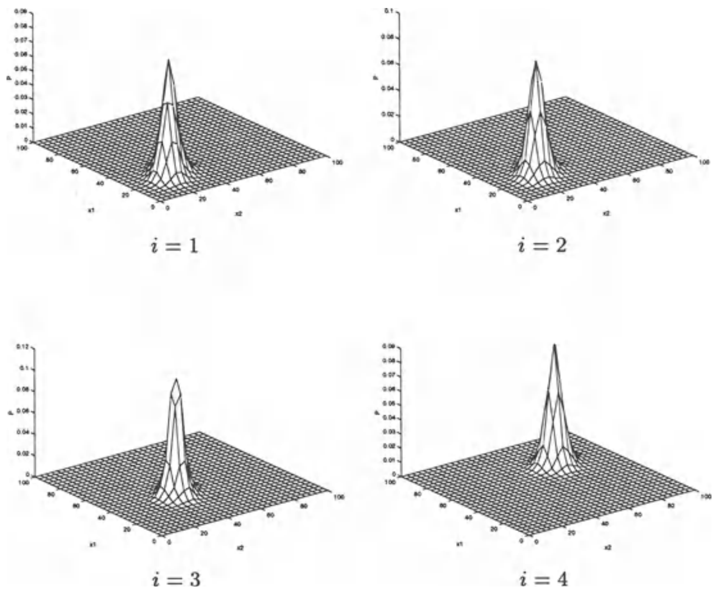


Figure 28.2. Steady state probabilities for the limit model, given the market state i.

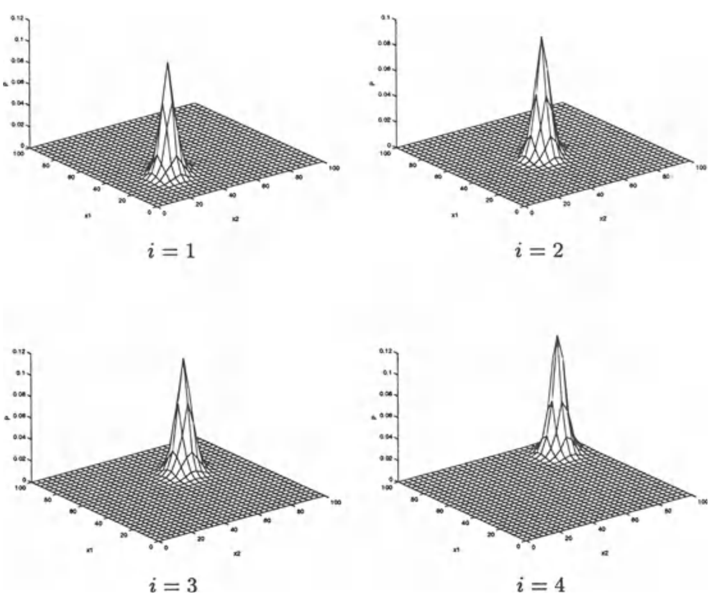
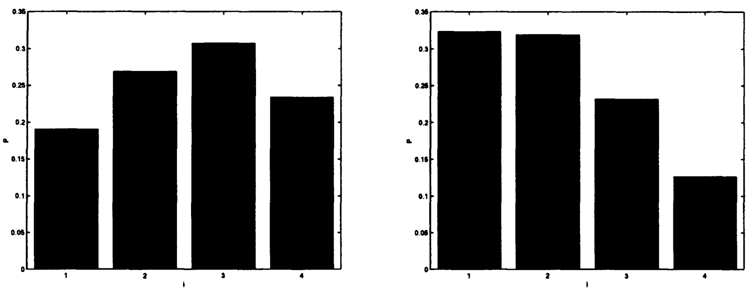


Figure 28.3. Steady state probabilities for the fixed ξ-process, given $\xi(t) = i \quad \forall t \geq 0$.



Optimal policy | Acting as if the price would never change

Figure 28.4. Steady state probabilities as a function of the state i.

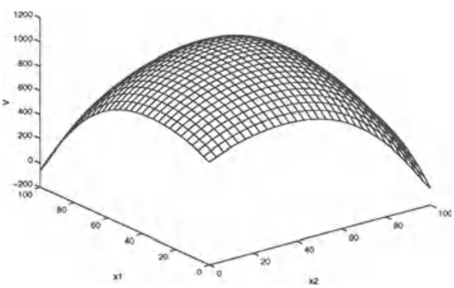
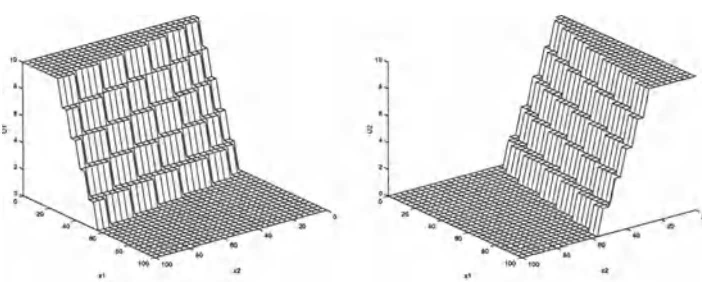


Figure 28.5. Value function $V(x, 3)$



$u_1(x)$ | $u_2(x)$

Figure 28.6. Optimal policy $u(x)$, $i = 3$.

References

[1] Abbad, M. (1991). *Perturbation and Stability Theory for Markov Control Problems.* Ph.D. thesis, University of Maryland at Baltimore County, Dept. of Mathematics.

[2] Abbad, M., Filar, J.A. and Bielecki, T.R. (1992). Algorithms for singularly perturbed limiting average markov control problems. *IEEE Transactions on Automatic Control*, 37: 1421–1425.

[3] Benders, J.F. (1962). Partitioning procedures for solving mixed-variables programming problems. *Numerische Mathematik*, 4: 238–252.

[4] de Ghellinck, G.T. (1960). Les problè de décision séquentielles. *Cahier du Centre d'Etudes de Recherche Operationnelle*, 2: 161–179.

[5] du Merle, O. (1995). *Interior points and cutting planes: development and inmplementation of methods for convex optimization and large scale structured linear programming.* PhD thesis, Department of Management Studies, University of Geneva, Switzerland. (in French.)

[6] Filar, J.A., and Haurie, A. (1997). Optimal ergodic control of singularly perturbed hybrid stochastic systems. *Lectures in Applied Mathematics*, 33: 101–126.

[7] Filar, J.A., and Haurie, A. (2001). A two factor stochastic production model with two time scales. *Automatica*, 37: 1505–1513.

[8] Ghost, M.K., Arapostathis, A., and Marcus, S.I. (1997). Ergodic control of switching diffusion. *SIAM Journal of Control and Optimization*, 35(6): 1952–1988.

[9] Goffin, J.-L., Haurie, A. and Vial, J.-P. (1992). Decomposition and nondifferentiable optimization with the projective algorithm. *Management Science*, 38: 284–302.

[10] Goffin, J.-L. and Vial, J.-P. (1999) Convex nondifferentiable optimization: a survey focussed on the analytic center cutting plane method. Technical report, Logilab, Department of Management Studies, University of Geneva, Switzerland. To appear in *Optimization Methods and Software.*

[11] Gondzio, J., du Merle, O., Sarkissian, R., and Vial, J.-P. (1996). ACCPM—a library for convex optimisation based on an analytic center cutting plane method. *European Journal of Operational Research,* 94: 206–211.

[12] Kushner, H.J. and Dupuis, P.G. (1992). *Numerical Methods for Stochastic Control Problems in Continuous Time.* Springer Verlag, New York.

[13] Layard, P.R.G. and Walters, A.A. (1978) *Microeconomic Theory,* McGraw Hill, New York.

[14] Manne, A. (1960). Linear programming and sequential decisions. *Management Science,* 6 (3): 259–267.

Chapter 29

THE EFFECT OF STOCHASTIC DISTURBANCE ON THE SOLITARY WAVES*

Junping Li
Changsha Railway University
Hunan, 410075, P.R.China

Zhenting Hou
Changsha Railway University
Hunan, 410075, P.R.China

Weiguo Zhang
University of Shanghai for Science and Technology
Shanghai, 200000, P.R.China

Zaiming Liu
Changsha Railway University
Hunan, 410075, P.R.China

Abstract This paper is devoted to studying the effect of stochastic disturbance on the kink profile solitary wave solution of the equation: $u_{tt} - \delta u_{xxt} - ku_{xx} + au_t + buu_t = 0 (\delta > 0)$ by using the theory of Markov Skeleton processes established recently by Z.T. Hou, Z.M. Liu and J.Z. Zou [1]. The transition probability and stability of the solution are given.

Keywords: Markov Skeleton process, Stochastic disturbance, Kink profile solitary wave.

*Supported by the national natural science fund of China (19871006).

Z. Hou et al. (eds.), Markov Processes and Controlled Markov Chains, 465–474.

1. Introduction

The non-linear wave equation

$$u_{tt} - \delta u_{xxt} - ku_{xx} + au_t + buu_t = 0, \qquad (\delta > 0) \tag{1.1}$$

is an important mathematical model of studying quantum mechanics, vibration of a viscous rod, and nerve conduct, etc. [4, 5, 6, 7, 8].

Recently, W.G. Zhang [3] successfully provided the kink profile solitary solution of equation (1.1). It is well known that, the developing process of every thing should be a stochastic process because of the disturbance from some random factors around it. This paper is devoted to studying the effect of stochastic disturbance on kink profile solitary wave solutions of equation (1.1).

Let $(\Omega, \mathcal{F}, P)$ be a complete probability space, $(E, \mathcal{E})$ be a polish space and $X \triangleq \{X(t,\omega); t < \tau(\omega)\}$ be a right-continuous and left-limit stochastic process on $(\Omega, \mathcal{F}, P)$, with values in $(E, \mathcal{E})$.

For convenience, we extend the state space E to $\hat{E} = E \cup \{b\}$ by adding an isolated state b to E, as usual, we get a new polish space $(\hat{E}, \hat{\mathcal{E}})$, and the process X is also extended to $\hat{X} = \{\hat{X}(t,\omega); 0 \le t < \infty\}$ by

$$\hat{X}(t,\omega) = \begin{cases} X(t,\omega), & 0 \le t < \tau(\omega) \\ b, & \tau(\omega) \le t < \infty \end{cases} \tag{1.2}$$

Definition 1.1 *The stochastic process $X = \{X(t,\omega); 0 \le t < \tau(\omega)\}$ is called a Markov Skeleton process if there exists a series of stopping times $\{\tau_n; n \ge 0\}$ such that*

(i) $0 = \tau_0 \le \tau_1 \le \tau_2 \cdots$, $\tau = \lim_{n\to\infty} \tau_n$, *P-a.e..*

(ii) For each τ_n and any bounded $\hat{\mathcal{E}}^{[0,\infty)}$-measurable function f on $\hat{E}^{[0,\infty)}$,

$$E[f(\hat{X}(\tau_n+\cdot))|\mathcal{N}_{\tau_n}] = E[f(\hat{X}(\tau_n+\cdot))|\hat{X}(\tau_n)], \qquad P\text{-}a.e. \text{ on } \Omega_{\tau_n}$$

where $\Omega_{\tau_n} = \{\omega; \tau_n(\omega) < \infty\}$, and

$$\mathcal{N}_{\tau_n} \triangleq \left\{A; A \cap \{\tau_n \le t\} \in \sigma\{\hat{X}(s); 0 \le s \le t\} \text{ for any } t \ge 0\right\}$$

is the σ-algebra on Ω_{τ_n}.

Definition 1.2 *A Markov Skeleton process X is called a non-homogeneous (H,Q) – process, if there exist $\{h^{(n)}(t,x,A); n \ge 1\}$ and $\{q^{(n)}(t,x,A); n \ge 1\}$ such that*

(i) *For* $n \geq 0,\ t \geq 0,\ A \in \mathcal{E}$

$$E[X(\tau_n + t) \in A, \tau_{n+1} - \tau_n > t | X(\tau_n)] = h^{(n+1)}(t, X(\tau_n), A)$$

(ii) *For* $n \geq 0,\ t \geq 0,\ A \in \mathcal{E}$

$$E[X(\tau_{n+1}) \in A, \tau_{n+1} - \tau_n \leq t | X(\tau_n)] = q^{(n+1)}(t, X(\tau_n), A)$$

where $\{\tau_n; n \geq 0\}$ *is the series of stopping times as in* Definition 1.1, $h^{(n)}(t,x,A)$ *and* $q^{(n)}(t,x,A)$ *are non-negative binary measurable functions for fixed* A, *and pre-distributions for fixed* t *and* x.

Let $q^{(n)}(x,A) = \lim_{t\to\infty} q^{(n)}(t,x,A)$, then by Definition 1.2(i), one can see that $\{X(\tau_n); n \geq 0\}$ is a non-homogeneous Markov chain with transition probability $\{q^{(n)}(x,A); x \in E, A \in \mathcal{E}, n \geq 1\}$.

Define

$$\begin{aligned} \mathcal{U}_E \ \triangleq\ \{R | & R(x,A) \text{ is non-negative}, \\ & R(\cdot, A) \text{ is } \mathcal{E} - \text{measurable for fixed } A, \\ & R(x, \cdot) \text{ is a measure on } (E, \mathcal{E}) \text{ for fixed } x\}, \end{aligned}$$

the product operation in $\mathcal{U}_E$ is defined by

$$R \cdot S(x,A) = \int_E R(x,dy) S(y,A), \qquad \text{for } R, S \in \mathcal{U}_E$$

For $x \in E$, $A \in \mathcal{E}$, define

$$\begin{aligned} P(t,x,A) &= P(X(t) \in A | X(0) = x) \\ P^{(n)}(t,x,A) &= P(X(\tau_n + t) \in A | X(\tau_n) = x), \qquad n \geq 0 \\ P_\lambda(x,A) &= \int_0^\infty e^{-\lambda t} P(t,x,A)\, dt \\ P_\lambda^{(n)}(t,x,A) &= \int_0^\infty e^{-\lambda t} P^{(n)}(t,x,A)\, dt, \qquad n \geq 0 \\ h_\lambda^{(n)}(x,A) &= \int_0^\infty e^{-\lambda t} h^{(n)}(t,x,A)\, dt, \qquad n \geq 1 \\ q_\lambda^{(n)}(x,A) &= \int_0^\infty e^{-\lambda t} dq^{(n)}(t,x,A), \qquad n \geq 1 \end{aligned}$$

Theorem 1.1 (Hou, Liu and Guo [2])
$\{P_\lambda^{(n)}(x,A); x \in E, A \in \mathcal{E}, n \geq 0\}$ *is the minimal non-negative solution of non-negative equation*

$$X^{(n)}(x,A) = \int_E q_\lambda^{(n+1)}(x,dy) X^{(n+1)}(y,A) + h_\lambda^{(n+1)}(x,A), \qquad n \geq 0, x \in E, A \in \mathcal{E} \qquad (1.3)$$

thus

$$P_\lambda^{(n)}(x,A) = \left(\sum_{m=1}^{\infty}\left(\prod_{k=1}^{m} Q_{n+k}\right)\cdot H_{n+m+1} + H_{n+1}\right)(x,A) \quad (1.4)$$

in particular

$$\begin{aligned} P_\lambda(x,A) &= P_\lambda^{(0)}(x,A) \\ &= \sum_{m=0}^{\infty}\left(\prod_{k=0}^{m} Q_k\right)\cdot H_{m+1}(x,A) \end{aligned} \quad (1.5)$$

where

$$\begin{aligned} Q_0 &= (\delta_A(x)) \in \mathcal{U}_E, & \\ Q_m &= (q_\lambda^{(m)}(x,A)) \in \mathcal{U}_E, & m \geq 1, \\ H_m &= (h_\lambda^{(m)}(x,A)) \in \mathcal{U}_E, & m \geq 1. \end{aligned}$$

Proof Refer to [2]. ∎

From Theorem 1.1, we know that, the distribution of X is determined uniquely by $(H_m, Q_m)_{m=1}^{\infty}$, so we also call X a $(H_m, Q_m)_{m=1}^{\infty}$-process.

In the case that $Q_m = Q_1$, $H_m = H_1 (m \geq 1)$, the associated process is homogeneous (H,Q)-process.

Definition 1.3 *A Markov Skeleton process $X = \{X(t); t < \tau\}$ is called a piecewise deterministic $(H_m, Q_m)_{m=1}^{\infty}$ -process, if*

(i) X is a $(H_m, Q_m)_{m=1}^{\infty}$-process with respect to $\{\tau_n; n \geq 0\}$;

(ii) There exists a series of measurable functions $\{f_n(x,t); n \geq 0\}$ defined on $E \times [0,\infty)$, such that for each $n \geq 0$,

$$X(t) = f_n(X(\tau_n), t - \tau_n), \quad t \in [\tau_n, \tau_{n+1}] \quad (1.6)$$

2. The effect of stochastic disturbance on the solitary waves

Now we discuss the effect of stochastic disturbance on the kink profile solitary wave solutions of equation (1.1).

It is easy to see that the solitary wave solution $u(x,t) = u(x - vt) = u(\xi)$ of (1.1) must solve the ordinary differential equation

$$u''(\xi) + \frac{v^2 - k}{v\delta}u'(\xi) - \frac{a}{\delta}u(\xi) - \frac{b}{2\delta}u^2(\xi) = c \quad (2.1)$$

where c is a constant. by [3, Theorem 5.3.1 and Theorem 5.4.2], (1.1) has a unique bounded solitary wave solution which is strictly monotone, if $v^2 \neq k$ and

$$\left(\frac{v^2-k}{v}\right)^2 \geq 4\delta\sqrt{a^2-2\delta bc} \tag{2.2}$$

For convenience, we only consider the stochastic disturbance on the half line $\xi \geq 0$.

Assume $v^2 \neq k$ and (2.2) hold, we also assume that $\frac{v^2-k}{v} < 0$ and $b < 0$ without loss of generality. By [3, Theorem 5.3.1], $u(\xi)$ is strictly decreasing and moreover

$$\begin{cases} u(-\infty) & = \lim_{\xi\to-\infty} u(\xi) & = -\frac{a}{b} - \frac{1}{b}\sqrt{a^2-2\delta bc} \\ u(+\infty) & = \lim u_{\xi\to+\infty}(\xi) & = -\frac{a}{b} + \frac{1}{b}\sqrt{a^2-2\delta bc} \end{cases} \tag{2.3}$$

Let $f(y,\xi)$ denote the solitary wave solution satisfying $u(0) = y$.

Suppose $\{\tau_n; n \geq 0\}$ is a series of random times defined on a complete probability space $(\Omega, \mathcal{F}, P) : 0 = \tau_0 < \tau_1 < \tau_2 < \cdots,\ \tau_n \uparrow +\infty$. At each τ_n, the solitary wave of (1.1) has *a* jump. Suppose the distribution after nth jump is $\pi^{(n)}$, then the solitary wave solution of (1.1) must be a stochastic process, say $X(\xi)$, thus

$$P\Big(X(\tau_n) \in A | X(\tau_{n-1}), \tau_{n-1}, \tau_n) = \pi^{(n)}(A), A \in \mathcal{B}((u+\infty), u(-\infty)\Big) \tag{2.4}$$

and for each $n \geq 0$

$$X(\xi) = f(X(\tau_n), \xi - \tau_n), \qquad \xi \in [\tau_n, \tau_{n+1}) \tag{2.5}$$

By Definition 1.3, $X(\xi)$ is a piecewise deterministic $(H_m, Q_m)_{m=1}^{\infty}$-process with respect to $\{\tau_n; n \geq 0\}$. So the discussion of $f(y,\xi)$ $(\xi \geq 0)$ under the effect of stochastic disturbance is equivalent to the discussion of $X(\xi)$.

Let

$$\begin{aligned} G^{(n+1)}(\xi, y) &\triangleq P(\tau_{n+1} - \tau_n \leq \xi | X(\tau_n) = y), \\ G_\lambda^{(n+1)}(y) &\triangleq \int_{0-}^{\infty} e^{-\lambda\xi} dG^{n+1}(\xi, y), \end{aligned}$$

so

$$\begin{aligned} h^{(n+1)}(\xi, y, A) &= P(X(\tau_n + \xi) \in A, \tau_{n+1} - \tau_n > \xi | X(\tau_n) = y) \\ &= \delta_A(f(y,\xi)) \cdot (1 - G^{(n+1)}(\xi, y)) \qquad (2.6) \\ q^{(n+1)}(\xi, y, A) &= P(X(\tau_{n+1}) \in A, \tau_{n+1} - \tau_n \leq \xi | X(\tau_n) = y) \\ &= G^{(n+1)}(\xi, y) \cdot \pi^{(n+1)}(A). \qquad (2.7) \end{aligned}$$

Noting that $Q_0 \cdot H_1(y,A) = h_\lambda^{(1)}(y,A)$ and

$$\left(\prod_{k=0}^{m} Q_k\right) \cdot H_{m+1}(y,A)$$
$$= \begin{cases} G_\lambda^{(1)}(y) \int \pi^{(1)}(dz) h_\lambda^{(2)}(z,A), & \text{if } m=1 \\ G_\lambda^{(1)}(y) \left[\prod_{k=1}^{m-1} \int \pi^{(k)}(dz) G_\lambda^{(k+1)}(z)\right] \cdot \int \pi^{(m)}(dz) h_\lambda^{(m+1)}(z,A), & \\ & \text{if } m>1 \end{cases}$$

By Theorem 1.1, we have

Theorem 2.1 *For every $y \in (u(+\infty), u(-\infty))$, the transition probability of $X(\xi)$ is given by*

$$P_\lambda(y,A) = h_\lambda^{(1)}(y,A) + G_\lambda^{(1)}(y) \sum_{m=1}^{\infty} \left[\prod_{k=1}^{m-1} \int \pi^{(k)}(dz) G_\lambda^{(k+1)}(z)\right] \cdot \int \pi^{(m)}(dz) h_\lambda^{(m+1)}(z,A) \tag{2.8}$$

where $\prod_{k=1}^{0} = 1$.

Theorem 2.2 *Suppose that $\frac{v^2-k}{v} < 0$, $b<0$ and (2.2). Let $\{\pi^{(n)}; n \geq 1\}$ be a series of probability measures on $(u(+\infty), u(-\infty))$. Then for every $y \in (u(+\infty), u(-\infty))$*

$$\begin{aligned} E_{y,\lambda}X &\triangleq \int_0^\infty e^{-\lambda\xi} E_y[X(\xi)]\, d\xi \\ &= \int_0^\infty e^{-\lambda\xi} f(y,\xi)(1-G^{(1)}(\xi,y))\, d\xi \\ &\quad + G_\lambda^{(1)}(y) \sum_{m=1}^{\infty} \left[\prod_{k=1}^{m-1} \int \pi^{(k)}(dz) G_\lambda^{(k+1)}(z)\right] \\ &\quad \cdot \int_0^\infty e^{-\lambda\xi} \int f(z,\xi) \left(1 - G^{(m+1)}(\xi,z)\right) \cdot \pi^{(m)}(dz)\, d\xi \end{aligned} \tag{2.9}$$

In particular, if $G^{(n)}(\xi,y) = 1 - e^{-\xi y} (n \geq 1)$, then

$$\begin{aligned} E_{y,\lambda}X &= \int_0^\infty e^{-(\lambda+\mu)\xi} f(y,\xi) d\xi \\ &\quad + \sum_{m=1}^{\infty} \left(\frac{\mu}{\lambda+\mu}\right)^m \int_0^\infty \int e^{-(\lambda+\mu)\xi} f(z,\xi) \pi^{(n)}(dz)\, d\xi \end{aligned} \tag{2.10}$$

thus

$$\begin{aligned} E_y X(\xi) &= e^{-\mu\xi} f(y,\xi) \\ &\quad + e^{-\mu\xi} \sum_{m=1}^{\infty} \frac{\mu^m}{(m-1)!} \int_0^{\xi} r^{m-1} \int f(z,\xi-r)\pi^{(m)}(dz)\,dr \end{aligned} \tag{2.11}$$

Proof By (2.6) and (2.8)

$$\begin{aligned} E_{y,\lambda} X &= \int_0^{\infty} e^{-\lambda\xi} E_y[X(\xi)]\,d\xi \\ &= \int z h_\lambda^{(1)}(y,dz) + G_\lambda^{(1)}(y) \sum_{m=1}^{\infty} \left[\prod_{k=1}^{m-1} \int \pi^{(k)}(dz) G_\lambda^{(k+1)}(z) \right] \\ &\quad \cdot \int\!\!\int z h_\lambda^{(m+1)}(u,dz)\pi^{(m)}(du) \\ &= \int_0^{\infty} e^{-\lambda\xi} f(y,\xi)\left(1 - G^{(1)}(\xi,y)\right) d\xi \\ &\quad + G_\lambda^{(1)}(y) \sum_{m=1}^{\infty} \left[\prod_{k=1}^{m-1} \int \pi^{(k)}(dz) G_\lambda^{(k+1)}(z) \right] \\ &\quad \cdot \int_0^{\infty} e^{-\lambda\xi}\,d\xi \int f(u,\xi)\left(1 - G^{(m+1)}(\xi,u)\right) \pi^{(m)}(du) \end{aligned}$$

In particular, if $G^{(n)}(\xi,y) = 1 - e^{-\mu\xi} (n \geq 1)$, then $G_\lambda^{(n)}(y) = \frac{\mu}{\lambda+\mu}$, this proves (2.10). Secondly, note that

$$\left(\frac{\mu}{\lambda+\mu}\right)^m = \int_0^{\infty} e^{-(\lambda+\mu)\xi} \frac{\mu^m}{(m-1)!} \xi^{m-1}\,d\xi$$

It is easy to get (2.11) by using Laplace transform. ∎

Theorem 2.3 *Suppose that the conditions of* Theorem 2.2 *hold and moreover,* $G^{(n)}(\xi,y) = 1 - e^{-\mu\xi}$ $(n \geq 1)$

(i) *If* $E[X(\tau_n)] = \int y\pi^{(n)}(dy) \to u(+\infty)$ *as* $n \uparrow \infty$, *then*

$$\lim_{\xi\to+\infty} E_y[X(\xi)] = u(+\infty), \textit{ for any } y \in (u(+\infty), u(-\infty)) \tag{2.12}$$

(ii) *If* $\pi^{(n)}(\cdot) = \pi(\cdot)(n \geq 1)$, *then for any* $y \in (u(+\infty), u(-\infty))$,

$$\lim_{\xi\to+\infty} E_y[X(\xi)] = \mu \int_0^{\infty} e^{-\mu r} \int f(z,r)\pi(dz)\,dr \tag{2.13}$$

Proof It can be proved from (2.11). ∎

3. The effect of stochastic disturbance at fixed site

Now we turn to study the effect of stochastic disturbance at fixed site x.

For fixed x, the solitary wave solution of (1.1) is $u(x,t) = u(x - vt)$. Let $u(x,t,y)$ denote the solitary wave solution satisfying $u(x,0) = y$.

Suppose $\{\tau_n; n \geq 0\}$ is a series of random times defined on a complete probability space $(\Omega, \mathcal{F}, P) : 0 = \tau_0 < \tau_1 < \tau_2 < \cdots$, $\tau_n \uparrow +\infty. X = \{X(t); t \geq 0\}$ is a $(H_m, Q_m)_{m=1}^{\infty}$-process with respect to $\{\tau_n; n \geq 0\}$ such that

$$X(t) = u(x, t - \tau_n, X(\tau_n)), \qquad \text{for } t \in [\tau_n, \tau_{n+1}) \tag{3.1}$$

and moreover, there exists a series of probability measures $\{\pi^{(n)}; n \geq 1\}$ on $(u(+\infty), u(-\infty))$ such that

$$P(X(\tau_n) \in A | X(\tau_{n-1}), \tau_{n-1}, \tau_n) = \pi^{(n)}(A) \tag{3.2}$$

Let

$$\begin{aligned} G^{(n+1)}(t,y) &\hat{=} P(\tau_{n+1} - \tau_n \leq t | X(\tau_n) = y), \text{ and} \\ G_\lambda^{(n+1)}(y) &= \int_{0-}^{\infty} e^{-\lambda t} dG^{(n+1)}(t,y), \end{aligned}$$

so

$$\begin{aligned} h^{(n+1)}(t,y,A) &= P(X(\tau_n + t) \in A, \tau_{n+1} - \tau_n > t | X(\tau_n) = y) \\ &= \delta_A(u(x,t,y)) \cdot (1 - G^{(n+1)}(t,y)), \qquad (3.3) \\ q^{(n+1)}(t,y,A) &= P(X(\tau_{n+1}) \in A, \tau_{n+1} - \tau_n \leq t | X(\tau_n) = y) \\ &= G^{(n+1)}(t,y)\pi^{(n+1)}(A). \qquad (3.4) \end{aligned}$$

Theorem 3.1 *Suppose the conditions of* Theorem 2.2 *hold, and* $\{\pi^{(n)}; n \geq 1\}$ *is a series of probability measures on* $(u(+\infty), u(-\infty))$.

Then for every $y \in (u(+\infty), u(-\infty))$,

$$\begin{aligned} E_{y,\lambda}X & \\ &= \int_0^\infty e^{-\lambda t} E_y[X(t)]\,dt \\ &= \int_0^\infty e^{-\lambda t} u(x,t,y)(1 - G^{(1)}(t,y))\,dt \\ &\quad + G_\lambda^{(1)}(y) \sum_{m=1}^{\infty} \left[\prod_{k=1}^{m-1} \int \pi^{(k)}(dz) G_\lambda^{(k+1)}(z) \right] \\ &\quad \cdot \int_0^\infty e^{-\lambda t} \int u(x,t,z) \left(1 - G^{(m+1)}(t,z)\right) \cdot \pi^{(m)}(dz)\,dt \end{aligned} \tag{3.5}$$

In particular, if $G^{(n)}(t,y) = 1 - e^{-\mu t} (n \geq 1)$, *then*

$$\begin{aligned} E_{y,\lambda}X &= \int_0^\infty e^{-(\lambda+\mu)t} u(x,t,y)\,dt \\ &\quad + \sum_{m=1}^{\infty} \left(\frac{\mu}{\lambda+\mu} \right)^m \int_0^{+\infty} e^{-(\lambda+\mu)t} \int u(x,t,z) \pi^{(m)}(dz)\,dt \end{aligned} \tag{3.6}$$

thus

$$\begin{aligned} E_y[X(t)] &= e^{-\mu t} \cdot u(x,t,y) \\ &\quad + e^{-\mu t} \sum_{m=1}^{\infty} \frac{\mu^m}{(m-1)!} \int_0^t s^{m-1} \int u(x,t-s,z) \pi^{(m)}(dz)\,ds \end{aligned} \tag{3.7}$$

Proof The proof is similar to that of Theorem 2.2. ■

Theorem 3.2 *Suppose the conditions of* Theorem 2.2 *hold, and* $v > 0$, $G^{(n)}(t,y) = 1 - e^{-\mu t} (n \geq 1)$.

(i) *If* $\lim_{m\to\infty} E[X(\tau_m)] = \lim_{m\to\infty} \int z\pi^{(m)} dz = u(-\infty)$, *then*

$$\lim_{t\to\infty} E_y[X(t)] = u(-\infty)$$

(ii) *If* $\pi^{(n)}(\cdot) = \pi(\cdot)(n \geq 1)$, *then*

$$\lim_{t\to\infty} E_y[X(t)] = \mu \int_0^\infty e^{-\mu s} \int u(x,s,z) \pi(dz)\,ds$$

Proof Note that $u(x,s,y)$ is increasing in s if $v > 0$, we can get the conclusion from (3.7). ■

References

[1] Z.T. Hou, Z.M. Liu and J.Z. Zou, Markov skeleton processes, Chinese Science Bulletin (in Chinese), 1998, Vol.43 (5), 457–466.

[2] Z.T. Hou, Z.M. Liu and X.P. Guo, Non-homogeneous (H,Q)-processes and their applications in queueing, 1999, (in print).

[3] W.G. Zhang, Long time behavior of the non-linear wave equations and the effect of the stochastic disturbance on the solitary wave, (Doctor's Thesis, Changsha Railway university, Changsha, China), 1999

[4] Nagume J., Arimoto S. and Yoshizawa S., An active pulse transmission line simulating nerve axon, Proc.IRE, 50 (1962), 2061–2070

[5] Pao C.V., A mixed initial boundary-value problem arising in neurophysidogy, J. Math. Anal. Appl., 52 (1975), 105–119

[6] Fitzgibbon W.E., Strongly damped quasilinear evolution equations, J. Math. Anal. Appl. 79 (1981), 536–550

[7] Lomdah P.S., Serensen O.H. and Christiansen P.L., Soliton excitations in Josephson tunnel junctions, Phys. Rev. B, 25 (1982), 5737–5748

[8] Showalter R.E., Regularization and approximation of second order evolution equations, SIAM J. Math. Anal., 7 (1976), 461–472

Chapter 30

INDEPENDENT CANDIDATE FOR TIERNEY MODEL OF H-M ALGORITHMS

Peide Chen
Institute of Applied Mathematics, Academia Sinica
Research Institute, Railway Campus,
South-Central University, Changsha

Abstract In Tierney's extended model, if the candidate kernels are independent of the present states, we found all possible acceptance functions, and distinguished a subclass for which the associated Markov chains converge uniformly with some nice rate. We also distinguished some other easily treated subclasses with some desirable properties of the associated chains.

Keywords: H-M algorithms; Tierney model

AMS Classfication: 60J05; 60C05

1. Introduction

Originally the motivation of H-M algorithm came from the following situation: one wants to generate a Monte Carlo sample from a distribution Π either discrete or having density π in R^k. But the distribution *prod* is not easy to sample from directly, and it is quite possible that the functional form of $\prod$ may be known only up to some unspecified normalizing constant which will be inconvenient to compute directly. One strategy is to pick some other distribution (which may depend on the "present" state x) $q(x, \cdot)$ which is easy to sample from, and then define the Markov transition kernel $P(x, \cdot)$ with density on $R^k \backslash \{x\}$

$$p(x, y) = \alpha(x, y)q(x, y)$$

Z. Hou et al. (eds.), Markov Processes and Controlled Markov Chains, 475–487.

and

$$P(x,\{x\}) = 1 - \int_{R^k\setminus\{x\}} p(x,y)dy,$$

where $\alpha(x,y)$, the so-called acceptance probability, is defined by

$$\alpha(x,y) = \begin{cases} \frac{\pi(y)q(y,x)}{\pi(x)q(x,y)} \wedge 1 & \text{if } \pi(x)q(x,y) > 0, \\ 1 & \text{if } \pi(x)q(x,y) = 0. \end{cases}$$

Two special cases have been used much more frequently: one, proposed originally by Metropolis and his co-authors in 1953 [1], requires a symmetric candidate $q(x,y) = q(y,x)$; the other, proposed originally by Hastings in 1970 [2], requires an independent candidate $q(x,y) = q(y)$, which does not depend on x.

In papers of recent years, some authors realized that the concepts work well for general state spaces with a σ-finite reference measure μ (see for example, Tierney, 1994 [5]). In his paper [6], Tierney even worked on general state spaces without reference measure. He studied a general acceptance probability $\alpha(x,y)$ ensuring the resulting Markov chain is reversible, this implies that $\prod$ is the invariant measure of the chain, i.e., the transition kernel

$$P(x,dy) = \alpha(x,y)Q(x,dy) + \delta_x(dy)\int_{F\setminus\{x\}} [1-\alpha(x,u)]Q(x,du)$$

satisfies the detailed balance relation

$$\prod(dx)P(x,dy) = P(y,dx)\prod(dy),$$

or equivalently

$$\prod(dx)\alpha(x,y)Q(x,dy) = \alpha(y,x)Q(y,dx)\prod(dy).$$

Here (F,$\mathcal{F}$) denotes the general state space, $\prod$ the target stationary distribution, Q the candidate transition probability kernel.

Let

$$\lambda(dx,dy) = \prod(dx)Q(x,dy),$$

Tierney decomposed $F \times F$ into a disjoint union of symmetric subset $S \in \mathcal{F} \times \mathcal{F}$ and its complement S^c such that on S, λ and its transpose $\lambda^T : \lambda^T(dx,dy) = \lambda(dy,dx)$ are mutually absolutely continuous, while on S^c, they are mutually singular. Thus restricted on S, there exists a version of the density

$$s(x,y) = \frac{\lambda_S(dx,dy)}{\lambda_S^T(dx,dy)}$$

such that $0 < s(x,y) < \infty$ and $s(x,y) = \frac{1}{s(y,x)}$ holds for all $x, y \in F$ (outside S, define $s(x,y) = 1$), where λ_S and λ_S^T denote the restriction of λ and λ^T on S respectively. With these notations, Tierney obtained the following sufficient and necessary condition to ensure the reversibility of the Markov chain: the detailed balance condition

$$\alpha(x,y)\lambda(dx,dy) = \alpha(y,x)\lambda(dy,dx)$$

holds if and only if $\alpha(x,y)s(x,y) = \alpha(y,x)$ on S, and $\alpha(x,y) = 0$ on S^c, $\lambda - a.e.$

For an independent candidate, i.e., when $Q(x,dy) = Q(dy)$ does not depend on x, we find in this paper all possible α described by associating them with a symmetric function

$$\sigma : \sigma(x,y) = \alpha(x,y)\frac{1-h(y)}{h(y)}, \text{ where } h(x) = \frac{d\prod}{d(\prod + Q)}.$$

A chain is called uniformly ergodic, if the convergence from the n-step transition probability measure $P^n(x,\cdot)$ to the invariant distribution $\prod$ is uniformly geometric, i.e.,

$$\left\| P^n(x,\cdot) - \prod \right\| \leq Cr^n$$

for some constants r, C such that $0 < r < 1$, $0 < C < \infty$. We call r a convergence rate, and C a controlled coefficient. They are not unique, but of course we prefer to choose them as small as possible. If we allow the controlled coefficient C to depend on x, the chain is called geometrically ergodic. These definitions are not the original ones but equivalent to them. Mengersen and Tweedie (1994) obtained an easy to check condition to ensure the independent Hastings algorithm uniformly ergodic: suppose $\prod \ll Q$, then the chain is uniformly ergodic, if and only if $\pi(x) = \frac{d\prod}{dQ} \leq W < \infty$, $Q - a.s.$, $x \in F$ for some positive real number $W > 0$, if this is the case,

$$\left\| P^n(x,\cdot) - \prod \right\| \leq (1 - \frac{1}{W})^n,$$

otherwise the chain is not even geometrically ergodic. It turns out that their method works well for general state spaces and even for general acceptance probabilities. Thus we need to discuss how to choose an acceptance probability making the chain uniformly ergodic with the best rate, and we distinguish a subclass of acceptance probabilities making the corresponding Markov chains uniformly ergodic with the nice rate ensured by this result.

We also distinguish some other subclasses with special structure which are easy to treat in some sense and the corresponding Markov chains having some desirable properties.

2. Independent case in Tierney model

If the candidate kernel $Q(x, dy) = Q(dy)$ does not depend on the present state x, the detailed balance condition becomes

$$\alpha(x,y)\prod(dx)Q(dy) = \alpha(y,x)Q(dx)\prod(dy).$$

Let

$$M(dx) = \prod(dx) + Q(dx), \qquad h(x) = \frac{d\prod}{dM},$$

then $\frac{dQ}{dM} = 1 - h(x)$, and it is clear that $S_{\prod} = \{x : h(x) > 0\}$ is a support of $\prod$ and $S_Q = \{x : h(x) < 1\}$ is a support of Q. Let

$$S_\Lambda = S_{\prod} \cap S_Q = \{x : 0 < h(x) < 1\},$$

then $S = S_\Lambda \times S_\Lambda$ is the symmetric subset of $F \times F$ in Tierney decomposition for $\lambda(dx, dy) = \prod(dx)Q(dy)$, and the density

$$s(x,y) = \begin{cases} \frac{\lambda_S(dx.dy)}{\lambda_S^T(dx,dy)} = \frac{h(x)[1-h(y)]}{h(y)[1-h(x)]} & \text{on S,} \\ s(x,y) = 1 & \text{elsewhere.} \end{cases}$$

Thus the Tierney's extended acceptance probabilities $\alpha(x, y)$ satisfy

$$\alpha(x,y)h(x)[1-h(y)] = \alpha(y,x)h(y)[1-h(x)] \text{ on S}$$

and $\alpha(x, y) = 0$ elsewhere. To solve it, we consider first the case when $\alpha(x, y)$ is "factorized", i.e., $\alpha(x, y) = \alpha_1(x)\alpha_2(y)$. We ignore the "zero" solution $\alpha(x, y) \equiv 0$, so $\alpha_1(x) > 0$ for some $x \in S_\Lambda$, and $\alpha_2(y) > 0$ for some $y \in S_\Lambda$, thus from

$$\alpha_1(x)\alpha_2(y)h(x)[1-h(y)] = \alpha_1(y)\alpha_2(x)h(y)[1-h(x)]$$

we know $\alpha_1(y)$ and $\alpha_2(x) > 0$ too. Furthermore from

$$\frac{\alpha_1(x)}{\alpha_2(x)}\frac{h(x)}{1-h(x)} = \frac{\alpha_1(y)}{\alpha_2(y)}\frac{h(y)}{1-h(y)} = c$$

for all $x, y \in S_\Lambda$, the constant c is positive, therefore $\alpha_1(x) > 0$ for all $x \in S_\Lambda$, and $\alpha_2(y) > 0$ for all $y \in S_\Lambda$. Also $\alpha_2(x) = \frac{1}{c}\frac{h(x)}{1-h(x)}\alpha_1(x)$. The positive constant c is not important, if $\alpha(x, y) = \alpha_1(x)\alpha_2(y)$ is a factorized representation with

$$\alpha_2(x) = \frac{1}{c}\frac{h(x)}{1-h(x)}\alpha_1(x)$$

for $c > 0$, let $\alpha_1^*(x) = \frac{1}{\sqrt{c}}\alpha_1(x)$, $\alpha_2^*(x) = \sqrt{c}\alpha_2(x)$, then $\alpha(x,y) = \alpha_1^*(x)\alpha_2^*(y)$ is another factorized representation with $c = 1$.

It is more convenient to use another choice of c in finding all factorized acceptance probabilities. Notice that $\alpha_1(\cdot)$ must be bounded due to the fact $\alpha(x,y) \le 1$, so we may "normalize" $\alpha_1(\cdot)$ by requiring $\sup_{x\in S_\Lambda} \alpha_1(x) = 1$, then $\alpha_2(x) = \frac{1}{c}\frac{h(x)}{1-h(x)}\alpha_1(x)$, and

$$\alpha(x,y) = \frac{1}{c}\alpha_1(x)\alpha_1(y)\frac{h(x)}{1-h(x)} \le 1$$

implies $\sup_{x\in S_\Lambda} \alpha_1(y)\frac{h(y)}{1-h(y)} \le c$, in other words, $\alpha_1(y) = O(1-h(y))$ for those y such that h(y) close to one. Now we reach the structure statement for the factorized acceptance probabilities.

Proposition 2.1 *Given any non-negative function $\alpha_1(x)$ on F with S_Λ as its support and such that*

$$\sup_{x\in S_\Lambda} \alpha_1(x) = 1, \qquad W \hat{=} \sup_{x\in S_\Lambda} \alpha_1(x)\frac{h(x)}{1-h(x)} < \infty.$$

Then for any $c \ge W$, let $\alpha_2(x) = \frac{1}{c}\frac{h(x)}{1-h(x)}\alpha_1(x)$, we get a factorized acceptance probability $\alpha(x,y) = \alpha_1(x)\alpha_2(y)$. Conversely any factorized acceptance probability can be obtained by this approach.

Usually we take $c = W$ since larger c will make the chain with worse rate in uniform convergence. For example, if $c = \sup_{x\in S_\Lambda} \frac{h(x)}{1-h(x)} < \infty$, we may take $\alpha_1 \equiv 1$, $\alpha_2(x) = \frac{1}{c}\frac{h(x)}{1-h(x)}$, and get a factorized acceptance probability $\alpha(x,y) = \frac{1}{c}\frac{h(y)}{1-h(y)}$. In this example we require $h(x) \ge \varepsilon > 0$ for some positive ε on S_Λ, which means that we should choose the candidate Q not "too far away from" the target $\prod$, otherwise the chain will not be uniformly ergodic, this is natural. To give another example of factorized acceptance probability, we may take $\alpha_1(x) = \frac{1-h(x)}{1-\inf_{u\in S_\Lambda} h(u)}$ and $\alpha_2(x) = \frac{h(x)}{\sup_{u\in S_\Lambda} h(u)}$.

Now turn to the non-factorized case, the structure of general acceptance probabilities is similar.

Proposition 2.2 *Given any non-negative symmetric measurable function $\sigma(x,y)$ on $F \times F$ with support S and such that*

$$M = \sup_{(x,y)\in S} \sigma(x,y)\frac{h(y)}{1-h(y)} < \infty,$$

then for any $c \geq M$,

$$\alpha(x,y) = \frac{1}{c}\sigma(x,y)\frac{h(y)}{1-h(y)}$$

is an acceptance probability. Conversely any acceptance probability can be obtained by this approach.

Proof The first part is straight forward. For the second part, let $\alpha(x,y)$ be any given acceptance probability. From

$$\alpha(x,y)s(x,y) = \alpha(x,y)\frac{h(x)}{1-h(x)}\frac{1-h(y)}{h(y)} = \alpha(y,x)$$

we know

$$\sigma(x,y)\hat{=}\alpha(x,y)\frac{1-h(y)}{h(y)} = \alpha(y,x)\frac{1-h(x)}{h(x)} = \sigma(y,x)$$

is a symmetric function on S and satisfies

$$\sup_{(x,y)\in S} \sigma(x,y)\frac{h(y)}{1-h(y)} \leq 1 < \infty,$$

and we extend it to the whole space by defining it as 0 elsewhere. This σ will give the original α by the approach. ■

For independent candidate, the standard H-M algorithm has acceptance probability

$$\alpha_s(x,y) = \frac{[1-h(x)]h(y)}{h(x)[1-h(y)]} \wedge 1,$$

the corresponding symmetric function is $\alpha_s(x,y) = \frac{[1-h(x)]}{h(x)} \wedge \frac{h(x)}{[1-h(y)]}$, here the subscription s represents that it is standard but not a general one. While for factorized acceptance probability $\alpha(x,y) = \alpha_1(x)\alpha_2(x)$, the corresponding symmetric function is $\frac{1}{c}\alpha_1(x)\alpha_1(y)$, which is factorized too. The following lemma shows that the acceptance probability of the standard H-M algorithm is usually not a factorized one.

Lemma 2.1 *If $\alpha_s(x,y) = \frac{[1-h(x)]h(y)}{h(x)[1-h(y)]} \wedge 1$ is factorized, then $\alpha_s(x,y) = 1$.*

Proof Suppose $\alpha_s(x,y) = \alpha_1(x)\alpha_2(y)$, if h attains its maximum at some $y_0 \in F$, then $\alpha_s(x,y_0) = \alpha_1(x)\alpha_2(y_0) = 1$, so $\alpha_1(x) = \frac{1}{\alpha_2(y_0)}$ does not depend on x and we may take $\alpha_1 \equiv 1$. But then $\alpha_2(y) = \frac{\alpha_s(y,y)}{\alpha_1(y)} = 1$

for all $y \in F$. In the case where $\sup_{x \in F} h(x)$ is unreachable, take $y_n \in F$ such that $\lim_{n\to\infty} h(y_n) = \sup_{x\in F} h(x)$, then

$$E_n = \{x : h(x) \le h(y_n)\} \uparrow F \qquad \text{as } n \to \infty,$$

and $\alpha_1 \equiv 1$ on E_n, so $\alpha_1(x) \equiv 1$. ■

Clearly factorized functions are much easier to calculate than others, so factorized acceptance probabilities may hopefully provide better algorithms in the sense of reducing the total amount of calculations. Next theorem gives this a theoretical support.

Theorem 2.1 *If the acceptance probability is factorized, then for all positive integers n, the n-step transition probability densities of the chain are also factorized.*

Proof The one-step density is clearly factorized: with $p_1(x) = \alpha_1(x)$ and $p_2(y) = \alpha_2(y)q(y)$ by

$$p(x,y) = \alpha(x,y)q(y) = \alpha_1(x)\alpha_2(y)q(y).$$

Then inductively, if the n-step transition probability density is factorized

$$p^{(n)}(x,y) = p_1^{(n)}(x)p_2^{(n)}(y),$$

then

$$\begin{aligned} &p^{(n+1)}(x,y) \\ &= \int_F p_1(x)p_2(z)p_1^{(n)}(z)p_2^{(n)}(y)\mu(dz) + P_r(x,\{x\})p_1^{(n)}(x)p_2^{(n)}(y) \\ &= \left[p_1(x)\int_F p_2(z)p_1^{(n)}(z)\mu(dz) + P_r(x,\{x\})p_1^{(n)}(x)\right] p_2^{(n)}(y) \end{aligned}$$

shows that the $(n+1)$-step transition probability density is factorized too. ■

Here μ is any reference measure on F such that $\prod \ll \mu$ and $Q \ll \mu$; while

$$P_r(x,\{x\}) = 1 - \int_F p(x,y)\mu(dy)$$

is the probability the chain stayed in put by rejection.

Here we distinguish two kinds of "stay in put": by rejection and by acceptance, the later $P_a(x,\{x\}) = p(x,x)\mu(\{x\})$ denotes the probability that the chain stays at the same state x because it happens to be the "new" state by sampling and we accept it. If the reference measure μ is non-automatic, this does not happen.

We see that $p_2^{(n)}(y) = p_2(y) = \alpha_2(y)q(y)$ does not depend on n. Besides if we take $\alpha_1(x) = 1$ (or any constant in (0,1), then both $p_1(x) = \alpha_1(x)$ and

$$P_r(x, \{x\}) = 1 - p_1(x) \int_F p_2(y)\mu(dy)$$

do not depend on x too, then inductively, $p_1^{(n)}(x) = c_n$ are constants for all $n \in N$. Therefore $p^{(n)}(x, y) = c_s p_2(y)$ for all $y \neq x$. Thus the study of convergence of n-step transition probability reduce to that of a sequence of real numbers.

3. Convergence rate of independence algorithms

Now we discuss the convergence of the associated chain. As we know from the general theory of Markov chains, to ensure the n-step transition probabilities converge to the stationary distribution, the usual starting point is to assume that the chain is irreducible and aperiodic. Here and later, we follow the concepts and terminology from Meyn and Tweedie [4] when our discussion relates to general theory of Markov chains. For the standard H-M acceptance probability, the chain is irreducible if and only if $\prod \ll Q$; and if this is the case, the chain is automatically aperiodic (in fact it is strongly aperiodic). For general acceptance probabilities, to ensure irreducibility, it is still necessary to have $\prod \ll Q$. In the rest of this paper, we will assume so.

To begin with, we restate the result of Mengersen and Tweedie [3] in a version we want, it is an extension, but the demonstration is almost the same, so we omit.

Theorem 3.1 *For any acceptance probability $\alpha(x, y)$, if*

$$p(x, y) = \alpha(x, y)[1 - h(y)] \geq \beta h(y)$$

for some positive real number β and all $(x, y) \in S$, then the chain is uniformly ergodic with the rate $1 - \beta$. If $\alpha(x, y)$ is "mixfactorized" in the sense of having the form

$$\alpha(x, y) = \min_{1 \leq i \leq k} [f_i(x) g_i(y)],$$

the condition is also necessary, in fact if the condition does not hold, the chain can not be even geometrically ergodic.

Clearly the standard H-M acceptance probability

$$\alpha_s(x,y) = \frac{[1-h(x)]h(y)}{h(x)[1-h(y)]} \wedge 1$$

is mix-factorized. For a given $\alpha(x,y)$, if we choose the largest possible β, denoted by β_α, then the chain is uniformly ergodic with the rate β_α and controlled coefficient 1. It is natural to choose α so that β_α become as large as possible. Denote the largest possible β_α by β_0, an acceptance probability α is called "nice" if $\beta_\alpha = \beta_0$. All such acceptance probabilities form a subclass, we will call it "the nice rate subclass". We use "nice" but not "best", because the best rate $1-\beta_\alpha$ provided by Theorem 3.1 may not be "sharp", the following examples shows that the chain may uniformly converge at a better rate than the one ensured by the theorem. Consider the finite state space with three states, let $\pi = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$, $q = (\frac{1}{6}, \frac{1}{3}, \frac{1}{2})$, then $w = (2, 1, \frac{2}{3})$ and $\beta_0 = \frac{1}{2}$. Choose

$$\sigma = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{2} \end{pmatrix},$$

then

$$\alpha = \begin{pmatrix} 1 & \frac{1}{2} & \frac{1}{6} \\ 1 & \frac{1}{2} & \frac{1}{3} \\ \frac{1}{2} & \frac{1}{2} & \frac{1}{3} \end{pmatrix}, \text{ and } P = \begin{pmatrix} \frac{9}{12} & \frac{2}{12} & \frac{1}{12} \\ \frac{2}{12} & \frac{8}{12} & \frac{2}{12} \\ \frac{1}{12} & \frac{2}{12} & \frac{9}{12} \end{pmatrix},$$

so the largest possible β is $\frac{1}{4}$, because $P_{13} = \frac{1}{12} = \frac{1}{4}\pi_3$, (for convenience of calculation, we use counting measure as the reference measure in the case of finite state space, so the condition reduces to $P_{ij} \geq \beta\pi_j$). Then it is easy to see

$$|\lambda I - P| = (\lambda - 1)\left(\lambda - \frac{1}{2}\right)\left(\lambda - \frac{2}{3}\right)$$

gives the three eigenvalues 1, $\frac{2}{3}$, and $\frac{1}{2}$. Therefore the convergence rate $\frac{2}{3}$ is better then $\frac{3}{4} = 1 - \beta$, the "best" one ensure by Theorem 3.1.

In general, the supremum of a bounded subset may not be attainable. But that of $\{\beta_\alpha\}$ is attainable, the following two theorems give a constructive way to get β_0. First we define: $\beta_0 = \inf_{u \in S_\Lambda} \frac{1-h(u)}{h(u)}$, then in Theorem 3.2, we prove that it is the best β_α in the class of all factorized acceptance probabilities. Finally in Theorem 3.3 we prove that it is the best β_α in the class of all acceptance probabilities.

Theorem 3.2 *In the class of all factorized acceptance probabilities,*

$$\alpha(x,y) = \beta_0 \frac{h(y)}{1-h(y)} \quad \textit{on } S$$

uniquely provides a chain which is uniformly ergodic with the best rate $1-\beta_0$ ensured by Theorem 3.1.

Proof Take $\alpha_1(x) = 1$ on S_Λ, then

$$p(x,y) = \alpha(x,y)[1-h(y)] = \beta_0 h(y),$$

we see that $\beta_\alpha = \beta_0$. We need to prove that for any other factorized acceptance probability α, $\beta_\alpha < \beta_0$. Suppose

$$\alpha(x,y) = \alpha_1(x)\alpha_2(y) = \frac{1}{c}\alpha_1(x)\alpha_1(y)\frac{h(y)}{1-h(y)} \text{ on S},$$

where $0 < \alpha_1 \le 1$ on S_Λ, $\sup_{x\in S_\Lambda}\alpha_1(x) = 1$, $\rho = \inf_{x\in S_\Lambda}\alpha_1(x) < 1$ and $c = \sup_{x\in S_\Lambda}\alpha_1(x)\frac{h(x)}{1-h(x)}$. Clearly $c \ge \rho \sup_{x\in S_\Lambda}\frac{h(x)}{1-h(x)} = \frac{\rho}{\beta_0}$, so

$$p(x,y) = \alpha(x,y)[1-h(y)] = \frac{1}{c}\alpha_1(x)\alpha_1(y)h(y) \ge \beta\, h(y)$$

if and only if $\alpha_1(x)\alpha_1(y) \ge \beta c$ for all $(x,y) \in S$, if and only if $\rho^2 \ge \beta c$. Thus we have $\beta_\alpha = \frac{\rho^2}{c} \le \rho\beta_0 < \beta_0$. ■

Theorem 3.3 *Based on* Theorem 3.1, $1-\beta_0$ *is still the best rate even in the whole class of all acceptance probabilities. An acceptance probability $\alpha(x,y)$ provides a chain with this rate if and only if*

$$\beta_0\frac{h(y)}{1-h(y)} \le \alpha(x,y) \le \left[\frac{1-h(x)}{h(x)}\frac{h(y)}{1-h(y)}\right] \wedge 1,$$

i.e. if and only if $\alpha(x,y)$ lies between the best factorized acceptance probability and the standard H-M acceptance probability.

Proof For any acceptance probability $\alpha(x,y)$, we know that

$$\sigma(x,y) = \alpha(x,y)\frac{1-h(y)}{h(y)}$$

is symmetric. So on S,

$$\begin{aligned} p(x,y) &= \alpha(x,y)[1-h(y)] \\ &= \alpha(x,y)\frac{1-h(y)}{h(y)}h(y) \\ &= \alpha(y,x)\frac{1-h(x)}{h(x)}h(y) \\ &\ge \beta h(y) \end{aligned}$$

if and only if $\alpha(x,y) \geq \beta\frac{h(y)}{1-h(y)}$. Thus $\beta \leq \frac{1-h(y)}{h(y)}$ for all $y \in S_\Lambda$, therefore $\beta \leq \beta_0$. To attain this rate, we must have $\alpha(x,y) \geq \beta_0\frac{h(y)}{1-h(y)}$. On the other hand, $\alpha(x,y) \leq 1 \Rightarrow \sigma(x,y)\frac{1-h(y)}{h(y)}$, and $\sigma(x,y) = \sigma(y,x) \leq \frac{1-h(x)}{h(x)}$, therefore $\sigma(x,y) \leq \frac{1-h(x)}{h(x)} \wedge \frac{1-h(y)}{h(y)}$. Return to $\alpha(x,y)$, we get

$$\begin{aligned}\alpha(x,y) &\leq [\frac{1-h(x)}{h(x)} \wedge \frac{1-h(y)}{h(y)}] \cdot \frac{h(y)}{1-h(y)} \\ &= [\frac{1-h(x)}{h(x)} \cdot \frac{h(y)}{1-h(y)}] \wedge 1\end{aligned}$$

as required for necessity. The sufficiency is a direct consequence of Theorem 3.1. ■

4. Some other subclasses of acceptance probabilities

We have discussed the subclass of all factorized acceptance probabilities and the nice rate subclass. Their intersection consists of a single member, the best one of the factorized acceptance probabilities $\alpha(x,y) = \beta_0\frac{h(y)}{1-h(y)}$ on S.

If we regard $\wedge$ as an operation to replace the ordinary multiplication, we may define another kind of "factorized" acceptance probabilities, called Λ-factorized acceptance probabilities, with the standard H-M acceptance probability as a typical example. An acceptance probability $\alpha(x,y)$ is called Λ-factorized, if the corresponding symmetric function $\sigma(x,y) = \alpha(x,y)\frac{1-h(y)}{h(y)}$ has the form $\sigma^*(x) \wedge \sigma^*(y)$. For example, the standard H-M acceptance probability is Λ-factorized with $\sigma^*(x) = \frac{1-h(x)}{h(x)}$.

The decomposition of a factorized function $f(x,y) = f_1(x)f_2(y)$ is essentially unique (up to a constant multiplier), while the decomposition of a Λ-factorized function $f(x,y) = f_1(x) \wedge f_2(y)$ is various. But if we restrict ourselves on symmetric functions, there exists a unique Λ-factor decomposition, which is the smallest one.

Lemma 4.1 *Suppose Λ-factorized function $f(x,y) = f_1(x) \wedge f_2(y)$ is symmetric, then there uniquely exists a function g on F such that $f(x,y) = g(x) \wedge g(y)$. This Λ-factor decomposition is the smallest one in the following sense: if*

$$f(x,y) = f_1^*(x) \wedge f_2^*(y)$$

is another Λ*-factor decomposition, then*

$$g(x) \leq f_1^*(x) \wedge f_2^*(x).$$

Proof It is straightforward and omitted. ■

We will call the class of all Λ-factorized acceptance probabilities the Λ-factorized subclass. The members of this subclass are characterized not by the acceptance probabilities themselves, but by the associated symmetric functions due to the symmetry of the later. Each member corresponds to a measurable function $\sigma^*(x)$ on $(F, \mathcal{F})$, a measurable function $\sigma^*(x)$ on $(F, \mathcal{F})$ corresponds to a member of this subclass if and only if

$$\sigma^*(x) \leq \frac{1 - h(x)}{h(x)}.$$

So the standard H-M acceptance probability happens to be the member with the largest σ^*.

Combine these two kinds of factorized functions, we get the concept of mix-factorized functions with the general form

$$f(x, y) = \min_{1 \leq i \leq k} [f_i(x) g_i(y)].$$

Clearly both factorized functions and Λ-factorized functions are all mix-factorized functions. And the class of all mix-factorized functions if closed under both multiplication and the operation of taking minimum. The class of all mix-factorized acceptance probabilities will be called the mix-factorized subclass. This is obviously a container of both the factorized subclass and Λ-factorized subclass.

Finally the largest class discussed in this paper is the sign-factorized class. A function or an acceptance probability $\alpha(x, y)$ is called sign-factorized if $sgn[\alpha(x, y)]$ is factorized. Since $\alpha(x, y)$ is non-negative, so $sgn[\alpha(x, y)]$ is in fact the indicator of the support $S_\alpha = \{(x, y) : \alpha(x, y) > 0\}$ of α. It is trivial to see that an indicator 1_A is factorized if and only if A is a measurable rectangle. So any non-negative mix-factorized function $f(x, y) = \min_{1 \leq i \leq k}[f_i(x) g_i(y)]$ is sign-factorized since its support

$$S_f = \left\{ x : \prod_{i=1}^{k} f_i(x) > 0 \right\} \times \left\{ y : \prod_{i=1}^{k} g_i(y) > 0 \right\}$$

is a measurable rectangle. For this we state the following proposition to end this section and the whole paper.

Proposition 4.1 *For a sign-factorized acceptance probability* $\alpha(x,y)$*, the chain is* $\prod$*-irreducible, if and only if* $\prod \times \prod(S_\alpha) = 1$*. Besides if* $\alpha(x,y)$ *is mix-factorized, then* $\prod \times \prod(S_\alpha) = 1$ *also implies that the chain is strongly aperiodic.*

References

[1] Metropolis, N., Rosenbluth, A. W. Rosenbluth, M. N., Teller, A. H., and Teller, E.,(1953). Equations of state calculation by fast computing machines. J.Chemical Physics 21 1087-1091.

[2] Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their application. Biometrika 57 97-109.

[3] Mengersen, K. L., and Tweedie, R. L. (1996). Rate of convergence of the Hastings and Metroplis algorithm. Ann. Statist. 24 no.1 101-121.

[4] Meyn, S. P., and Tweedie, R. L. (1993). Markov chains and stochastic stability. Springer-Verlag, London.

[5] Tierney, L. (1994). Markov chains for exploring posterior distribution (with discussion). Ann. Statist. 22 1071-1786

[6] Tierney, L. (1998). A note on Metropolis-Hastings kernel for general state space. Ann. Appl. Probab. 8 no.1 1-8

Chapter 31

HOW RATES OF CONVERGENCE FOR GIBBS FIELDS DEPEND ON THE INTERACTION AND THE KIND OF SCANNING USED

Yuzhi Cai
Department of Mathematics and Statistics
University of Surrey
Guildford, Surrey, GU2 5XH, UK
y.cai@surrey.ac.uk

Abstract In this paper we describe recent empirical work using perfect simulation to investigate how rates of convergence for Gibbs fields might depend on the interaction between sites and the kind of scanning used. We also give some experiment results on Kendall's [8] perfect simulation method for area-interaction process, which show that the repulsive case could be quicker or slower than the attractive case for different choices of the parameters.

Keywords: McMC, Gibbs sampler, Perfect simulation, coalescence time, attractive, repulsive, area-interaction

1. Introduction

Development in the area of perfect simulation is rapid. Many perfect simulation methods have been proposed. There are mainly two main types: one based on the idea of coupling from the past (CFTP) as proposed by Propp & Wilson [12], the other is the interruptible method proposed by Fill [3]. The majority of recent work has focussed on the CFTP idea: Murdoch & Green's [11] work, Kendall [6, 7, 8], Kendall & Møller [9], Kendall & Thönnes [10] and Cai & Kendall [2] work on point processes and stochastic geometry, Häggström & Nelander [5] work on Markov random fields etc.

Z. Hou et al. (eds.), Markov Processes and Controlled Markov Chains, 489–498.

Clearly, perfect simulation is a very powerful tool in the area of simulation. One direction is to consider how it can be used to investigate empirical (and more broadly applicable) methods. For example, can we use perfect simulation to investigate how rates of convergence for Gibbs fields might depend on the interaction between sites and the kind of scanning used?

Without using perfect simulation, Roberts & Sahu [13] investigated many convergence issues concerning the implementation of the Gibbs sampler. They conclude that for Gaussian target distribution with inverse dispersion matrix satisfying certain conditions, a random scan will take approximately twice as many iterations as a lexicographic order scan to achieve the same level of accuracy.

Greenwood et al [4] investigated information bounds (which is the minimal asymptotic variance of estimators of $E_\pi[f]$) for Gibbs samplers. Suppose we want to estimate

$$E_\pi[f] \approx \frac{1}{n}\sum_{i=1}^{n} f(x_i) = E_n f,$$

where the x_i are obtained by using a Gibbs sampler based on either a random scan or a systematic (deterministic) scan. Empirically the $E_n f$ has noticeable smaller variance for a deterministic scan. The variance bound for a random scan is smaller than that for a deterministic scan except when π is continuous, in which case the bounds coincide. Furthermore, the information bound for a deterministic scan does not depend on the details of the order of the sweep. The asymptotic variance of the empirical estimator under a random scan is no more than twice that under a deterministic sweep.

Both the Roberts & Sahu and the Greenwood et al results suggest that the way of scan does affect the rate of the convergence. Their work motivated the current work, i.e. to investigate empirically how rates of convergence for Gibbs fields might depend on the interaction and the kind of scan used, using perfect simulation.

The construction of this paper is as follows. In Section 2 we introduce perfect simulation methods for Ising model. The empirical simulation results will be presented in Section 3. In Section 4 we give some experimental results on Kendall's [8] perfect simulation method for the area-interaction process. Conclusions are presented in Section 5.

2. CFTP for Ising model

The Ising model is a simple magnetic model, in which spins σ_i are placed on the site i of a lattice. Each spin takes two values: $+1$ and -1.

If there are N sites on the lattice G, then the system can be in 2^N states, and the energy of any particular state is given by the Ising Hamiltonian:

$$H \quad = \quad -\tilde{J}\sum_{i\sim j}\sigma_i\sigma_j - B\sum_i \sigma_i$$

where $\tilde{J}$ is an interaction energy between nearest neighbor spins, $i \sim j$ indicates the existence of an edge connecting i and j, and B is an external magnetic field. Many interesting problems about the Ising model can be investigated by performing simulations in zero magnetic field $B = 0$. We will only consider this case here. Then the Ising measure π for G is a probability measure on $\{-1,1\}^V$, where V is the site set of G, which to each configuration $\sigma \in \{-1,1\}^V$ assigns probability

$$\pi(\sigma) = \frac{1}{Z_G^J} e^{-2J\sum_{i,j\in V, i\sim j} 1_{\sigma_i \neq \sigma_j}} \tag{2.1}$$

where $J = \tilde{J}/K_B\tilde{T}$, K_B is Boltzmann's constant, $\tilde{T}$ is the temperature, Z_G^J is a normalizing constant.

We can implement a Gibbs sampler to obtain an *a*pproximate sample from the equilibrium distribution (2.1) of the Ising model. But how do we get a perfect sample from π?

It is noted that the state space of the Ising model considered here is finite. We can define a partial order $\preceq$ on the state space as follows: we say $\xi \preceq \eta$ if $\xi_i \leq \eta_i$ for all $i \in V$. Corresponding to this partial order, there exists a maximum element $\hat{1}$ such that $\hat{1}_i = 1$ for all $i \in V$, and a minimum element $\hat{0}$ in the state space such that $\hat{0}_i = -1$ for all $i \in V$.

There are several well-known results with respect to the partial order, see Cai [1]. Using these well-known results, we can construct a monotone CFTP method for Ising model to obtain a perfect sample from π. The details about the monotone CFTP algorithms for attractive and repulsive cases are given in Cai [1].

Note that for four neighbouring Ising models on a square lattice, the repulsive interaction is actually the same as the attractive interaction (so no need for "Ising repulsive CFTP"). However we will consider other neighbourhood structures for which there is a difference. In the next section, we present our empirical results.

3. Experimental results for Ising model

In order to investigate how the rates of convergence for Gibbs fields might depend on the interaction between sites and the kind of scan used, we consider the coalescence time T_c of our perfect Gibbs sampler. Large values of T_c are related (though not exactly so) to a slower convergence rate. The following points, which might effect the convergence rate, have been considered.

We will consider two Ising models: One is an Ising model on a square lattice with four neighbours. We call this Model 1. The other is an Ising model on a triangular lattice with six neighbours. We call this Model 2. For these two models, we know the theoretical critical values J_c of J: for Model 1, $J_c \approx 0.44$, while for Model 2, $J_c \approx 0.27$. Propp & Wilson [12] showed how to simulate critical Ising models. Their method is rather complicated. In fact their perfect samples from the critical Ising models are obtained as a byproduct of a perfect simulation for random cluster models. Here our experiments will be based on simple cases, i.e. sub-critical values of J.

We will consider the effect of different scans. Two types of scans have been considered: one is a random scan, the other is a systematic scan which includes scans with lexicographic order, miss one out in lexicographic order (or chess-board scan), miss two out in lexicographic order and alternating lexicographic order.

We will also consider the effect of the interaction on the coalescence time, but this will only be done for Model 2, because, as we have pointed out in Section 2, for Model 1 the attractive interaction is actually the same as that for the repulsive interaction.

The data is collected in the following way. We set G to be a 100×100 grid. For Model 1 we take $J = 0.05, 0.10, 0.15, 0.2, 0.25, 0.3$. Corresponding to each value of J, we collect 100 independent coalescence times for the random scan and 25 independent coalescence times for each systematic scan — hence we also have 100 independent coalescence times for systematic scan. For Model 2 we take $J = \pm 0.05, \pm 0.10, \pm 0.15, \pm 0.2$. Then we collect 100 independent coalescence times for random scan and 100 for systematic scan in the same way as we did for Model 1.

Plots of the sample mean of the log coalescence times and the corresponding $5th$ and $95th$ percentiles versus the values of J are given in Figure 31.1 for Model 1 and in Figure 31.2 for Model 2.

These plots suggest that the random scan is slower than the systematic scan. The statistical test on the difference of the two mean coalescence times based on large samples confirmed this point. The Roberts & Sahu [13] and Greenwood et al's [4] work suggest that the scan method does have an effect on the convergence rate. Our results agree with this.

But how much slower for the random scan compared with the systematic scan? Figure 31.3 gives the plots of the ratios of the mean coalescence times for the systematic scan to that of the random scan versus the value of J for Model 1 and Model 2.

Both plots show increased ratio with $|J|$.

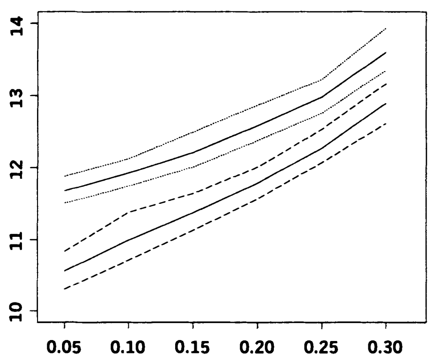


Figure 31.1. Graph of the sample mean of the log coalescence times and the corresponding empirical $5th$ and $95th$ percentiles versus the values of J for Ising model on square lattice with 4 neighbours and $J > 0$. The upper solid curve corresponds to random scan, the two dotted curves are the corresponding empirical $5th$ and $95th$ percentile curves. The lower solid curve corresponds to the systematic scan, the two dashed curves are the corresponding empirical $5th$ and $95th$ percentile curves.

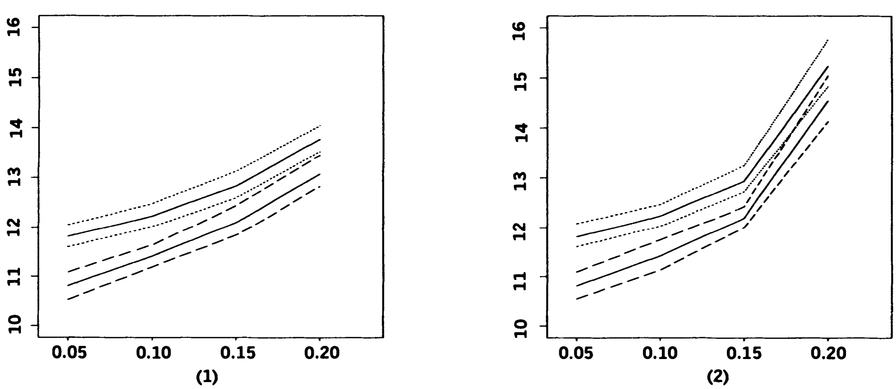


Figure 31.2. Graph of the sample mean of the log coalescence time and the corresponding empirical $5th$ and $95th$ percentiles versus the values of $|J|$ for Ising model on triangular lattice with 6 neighbours. The upper solid curve corresponds to random scan, the two dotted curves are the corresponding empirical $5th$ and $95th$ percentile curves. The lower solid curve corresponds to the systematic scan, the two dashed curves are the corresponding empirical $5th$ and $95th$ percentile curves. (1) Attractive case: $J > 0$. (2) Repulsive case: $J < 0$.

Does the interaction effect the coalescence time for Model 2?

A simple analysis of variance shows that both the scan method and the sign of J have a significant effect on the coalescence time at 5% level when $|J| \geq 1.5$. In the case of $|J| \leq 0.1$, the analysis suggests that only the way of scan has a significant effect on the coalescence time at

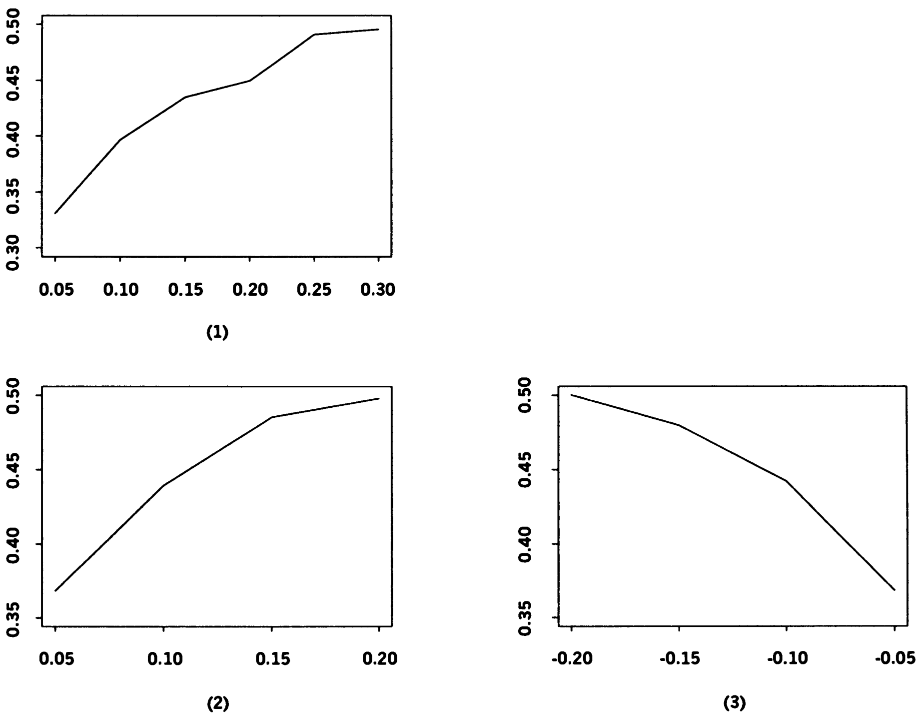


Figure 31.3. Graph of the ratios of the mean coalescence time for the systematic scan to that of the random scan versus the values of J. (1) For Model 1 with $J > 0$. (2) For Model 2 with $J > 0$. (3) For Model 2 with $J < 0$.

5% level. Further statistical tests on the mean difference also suggest that when $|J| \leq 0.1$, there is no significant difference between the mean coalescence time of the repulsive and attractive Model 2 at 5% level. The following heuristic analysis shows some of the reasons for the the above situation.

In the attractive case, let p_1 be the probability for the upper process to take value -1 at site i, p_2 be the probability for the lower process to take value -1 at site i. In the repulsive case, define p_1' and p_2' similarly.

Then we can prove that $p_1 \leq p_2$, $p_1' \leq p_2'$. Furthermore, we have the following result.

Theorem 3.1 *Consider Ising model on a triangular lattice with six neighbours. Suppose* $0 < J < J_c$. *Let* $\hat{p}_1 = (1 + e^{12J})^{-1}$ *and* $\hat{p}_2 = (1 + e^{-12J})^{-1}$. *Then* $p_1, p_2, p_1', p_2' \in [\hat{p}_1, \hat{p}_2]$.

The proof of the theorem is given in Cai [1]. ■

Now it is observed that if the difference $(p_2 - p_1)$(or $(p'_2 - p'_1)$) is large, then the upper and lower processes tend to stay where they are for the attractive (or repulsive) case. If $(p_2 - p_1)$(or $(p'_2 - p'_1)$) is small, then the upper and lower processes tend to take the same value. On the other hand, it follows from Theorem 3.1 that $p_1,\ p_2,\ p'_1,\ p'_2 \in [\hat{p}_1,\ \hat{p}_2]$. Furthermore when J is small, $\hat{p}_2 - \hat{p}_1$ is small, hence $p_2 - p_1$ and $p'_2 - p'_1$ are all small. So the upper and lower processes in both attractive and repulsive cases tend to move closer at the same time. Consequently, the coalescence time of the attractive case are not significantly different from that of the repulsive case when J is small. This is in good agreement with our statistical tests above. However, when J is large, $\hat{p}_2 - \hat{p}_1$ is large also. Hence the situation becomes very complicated now. It needs further investigation in the future.

4. Experimental results for the area-interaction process

The area interaction point process is a random process X of points in R^d with distribution having a Radon-Nikodym density $p(X)$ with respect to the unit-rate Poisson process restricted to a compact window, where

$$p(X) \quad = \quad \alpha\lambda^{\#(X)}\gamma^{-m_d(X\oplus C)} \tag{4.1}$$

where α is a normalizing constant, λ and γ are positive parameters and the grain C is a compact (typically convex) subset of R^d. The set $X \oplus C$ is given by

$$X \oplus C \quad = \quad \bigcup \{x \oplus C : x \in X\}.$$

The parameter γ controls the area interaction between the points of X: $\gamma > 1$ is attractive case and $\gamma < 1$ is the repulsive case.

Kendall [8] developed a perfect simulation method to obtain a perfect sample from the area-interaction process. He constructs a maximal and a minimal process, both of which are based on a dominated birth and death process that is in equilibrium. A perfect sample is obtained once the maximal and minimal processes coalesce by time 0. He observed that the perfect simulation method coalesces quicker for the repulsive case than that for the attractive case for the parameters he used. Is this true for any repulsive and attractive processes?

To answer his question, we carried on further experiments on the perfect simulation method.

First note that in the attractive case, we set $\gamma_1 = \gamma$ and in the repulsive case, we set $\gamma_2 = \gamma$. Then $\gamma_1 > 1$ and $\gamma_2 < 1$. In our experiments with the perfect simulation method, we use deliberately chosen values of γ_1 and γ_2 according to some theory we obtained. Our experiments suggest that it is possible to find some regions of (γ_1, γ_2), such that within those regions, we have that the repulsive case is quicker than the attractive case. It is also possible to find some other regions of (γ_1, γ_2) within which the repulsive case is slower than the attractive case. The experiment results (see Table 31.1) agree with what we expected. For the details about the experiments and the analysis about the experiment results, see Cai [1]. We hope we will be able to use our theory to give a precise answer to Kendall's [8] question in the near future.

Table 31.1. The mean and the standard deviation of the computation time (unit:second)

Experiments	Attractive case	Repulsive case
1	0.2844 (0.0084)	0.0282 (0.0083)
2	0.2538 (0.0088)	0.0316 (0.0084)
3	0.4624 (0.0139)	0.0198 (0.0068)
4	0.5948 (0.0111)	0.9564 (0.0344)

5. Conclusion

We have presented our empirical investigation results on the coalescence time of perfect Gibbs sampler based on different types of scans and different interactions by using perfect simulation method. Also, we have presented some experimental results on the perfect simulation method for area-interaction process. Our main results can be summarized as follows:

- No matter which types of interactions we have, the way of scanning (systematic or random) has a significant effect on the coalescence time. Specifically, the sampler with random scan has a larger coalescence time than that with systematic scan. The larger the value of $|J|$, the slower the random scan.

- The sampler for the repulsive case is not necessarily quicker than the attractive case. Generally speaking, it is possible to find some regions of the parameter space in which the repulsive case is quicker than the attractive case; it is also possible to find some regions in which the repulsive case is slower than the attractive case. Furthermore, the way of deciding the corresponding regions is model dependent.

References

[1] Cai, Y. (1999) How Rates of Convergence for Gibbs Fields Depend on the Interaction and the Kind of Scanning Used. *Research report 360, Department of Statistics, University of Warwick, Coventry, CV4 7Al, UK*

[2] Cai, Y. & Kendall, W.S. (2002) Perfect Simulation for Correlated Poisson Random Variables Conditioned to be Positive. *Statistics and Computing.* To appear.

[3] Fill, J. (1998) An Interruptible Algorithm for Exact Sampling via Markov Chains. *Annals of Applied Probability*, 8:131–162.

[4] Greenwood, P.E., McKeague, I.W. & Wefelmeyer, W. (1998) Information Bounds for Gibbs Samplers. *The Annals of Statistics*, 26, No.6:2128–2156.

[5] Häggström, O. & Nelander, K. (1999) On Exact Simulation of Markov Random Fields Using Coupling from the Past. *Scandinavian Journal of Statistics.* 26(3): 395–411.

[6] Kendall, W.S. (1997a) On Some Weighted Boolean Models. In D. Jeulin, editor, *Advances in Theory and Applications of Random Sets*, pages 105–120, Singapore. World Scientific Publishing Company.

[7] Kendall, W.S. (1997b) Perfect Simulation for Spatial Point Processes. In *Bulletin ISI,* 51^{st} *session proceedings, Istanbul (August 1997)*, volume 3, pages 163–166.

[8] Kendall, W.S. (1998) Perfect Simulation for the Area-interaction Point Process. In L. Accardi and C.C. Heyde, editors, *Probability Towards 2000*, pages 218–234, New York. Springer.

[9] Kendall, W.S. & Møller, J. (1999) Perfect Simulation of Point Processes Using Metropolis-Hastings Algorithms. *Research report 347, Department of Statistics, University of Warwick.*

[10] Kendall, W.S. & Thönnes, E. (1999) Perfect Simulation in Stochastic Geometry. *Pattern Recognition*, 32(9):1569–1586.

[11] Murdoch, D.J. & Green, P.J. (1998) Exact Sampling from a Continuous State Space. *Scandinavian Journal of Statistics*, 25(3):483–502.

[12] Propp, J.G. & Wilson, D.B. (1996) Exact Sampling With Coupled Markov Chains and Applications to Statistical Mechanics. *Random Structures and Algorithms*, 9:223–252.

[13] Roberts, G.O. & Sahu, S.K. (1997) Updating Scheme, Correlation Structure, Blocking and Parameterization for Gibbs Sampler. *J. R. Statist. Soc. B*59, No.2, 291–317.

Chapter 32

EXPECTED LOSS AND AVAILABILITY OF MULTISTATE REPAIRABLE SYSTEM*

Yubo GE
Department of Mathematical Sciences,
Tsinghua University
Beijing, 100084, China
ybge@math.tsinghua.edu.cn

Abstract Expected loss and availability of a multistate repairable system, allowing bulk-repair and bulk faults, with an overhaul period in the schedule are discussed. A new class of integral functional is first considered. Their moments and L-transformations are given and two ratio functions are defined as numerical indexes of availability of the system and are used to find out an optimum overhaul period. Discussion of the paper will be useful for traditional industry and control systems, computer information science (for example, computer recognition of speech and picture) and economy management. The obtained results are not only exact but also can be approximated for use on the computer.

Keywords: System reliability, jump process, integral function.

AMS Subject Classification: 90B52 60J75 62N05

1. Introduction

In reliability theory the traditional binary theory describing a system and components just as functioning or failed is being replaced by the theory for multistate system with multistate components (see for example [15]). Article [11] studied the association in time when performance

*Project supported by NSFC (19871045) and by Tsinghua Univ. (JC1999042)

Z. Hou et al. (eds.), Markov Processes and Controlled Markov Chains, 499–511.

process of each component is Markovian. [6] gave the expected loss and availability of system, which could be described by a general birth-death process.

Now, this paper discusses the expected loss of a multistate repairable system, allowing bulk-repair and bulk faults, with an overhaul period t in schedule, and defines two ratio functions to analyze the system availability and answer the following questions: How should we estimate the system availability? To save expenses, when is the opportune moment to overhaul completely or even to renew the whole system? It will be targeted in the present paper to study a class of integral functional,

$$Y_{tA} = \int_0^{t\wedge h_A} V(x_s)\, ds, \qquad t \wedge h_A = \min\{t, h_A\}, \tag{1.1}$$

where V is an arbitrary non-negative function on state space E and h_A is the first passage time entering state set A.

Assumption 1.1 *Our basic assumptions are following:*

(i) A system state at time s, denoted by x_s, indicates the number of faults (or, in other cases, a failure level of the system). The state space $E = \{0, 1, \ldots, \}$. The state process is a Markov jump process $X = \{x_s,\ s \geq 0\}$ with conservative $Q-$matrix $Q = (q_{ij})$.

(ii) State 0 shows a perfect functioning and every state in a set A is regarded as hardly functioning because of there being too many failures in the system or too high expenses for maintenance. So the system in the state of A cannot help taking an emergence overhaul ahead of schedule. Then $t \wedge h_A = \min\{t, h_A\}$ is a practical overhaul time. Take $A = [N, \infty) := \{N, N+1, \ldots\}$.

As you know, if $V(j)$ is the loss cost of the system with j faults within a unit of time, $j \in E$, then Y_{tA} indicates a total loss of the system before the practical overhaul time. If $V(\cdot) = \cdot$, Y_{tA} is an accumulation of duration of every state j before $t \wedge h_A$, weighted by the fault numbers.

In this paper, Section 2 first shows that the mean loss of the system before the practical overhaul time can be expressed in terms of the distribution of the relative last quitting time and prohibition potential; and then defines two ratio functions regarded as numerical indexes of the system availability. Maximizing them, one can find the optimum overhaul period t in schedule. All probability quantities concerning the theorems given in Section 2 are calculated in Section 3. Finally, Section 4 studies the high order moments of these functionals, and provides L-transformations. The discussion in this paper will be useful not only for traditional industry and control system, but also for computer information science (for example, computer recognition of speech and picture) and economy management. These results are exact and convenient to approximate and use.

Integral functionals

$$\int_0^t V(x_s)\,ds \quad \text{and} \quad Y_A := \int_0^{h_A} V(x_s)\,ds. \tag{1.2}$$

have been studied since the early 1960s. The first one is also called additive functional, see [1, 12]. [17] gave the systems of equations that distribution functions and L-transforms of Y_A should satisfy. [11] and [19] considered other functionals associated with jump points and problems about non-negative solutions corresponding system of equations. [16] discussed functional about extremum-times and extreme values for jump process. For birth-death processes, refer to [4], [7, 8, 9], [13], [14], [15, 16], [18], and so on.

In Section 2, we first show two main theorems to calculate the mean loss of the system during the practical overhaul period, which will be expressed in terms of a distribution of the relative last quitting time, prohibition potential and additive functional of prohibition probability. Then we define two ratio functions to be regarded as numerical indexes of availability of the system. All probability quantities concerned with these theorems are calculated in Section 3. Finally, Section 4 discusses the l-order of these functionals.

2. Expected loss and availability of system

Firstly, we show two main theorems to calculate mean loss of the system during the practical overhaul period, $[0, t \wedge h_A]$, which will be expressed in terms of a distribution of relative last quitting time ${}_AL_j$, prohibition potential ${}_Ag(k,j)$ and additive functional of prohibition probability ${}_AP_{kj}(s)$. Then we'll define two ratio functions that can be regarded as numerical indexes of availability of the system. Some quantities concerned in both theorems and ratios will be given in the next section.
Let

$$\begin{cases} {}_AP_{kj}(s) = P_k(x_s = j,\ s < h_A), & {}_Ag(k,j) = \int_0^\infty {}_AP_{kj}(s)\,ds, \\ {}_AP_{kj}(\lambda) = \int_0^\infty e^{-\lambda s}{}_AP_{kj}(s)\,ds, & Re(\lambda) > 0, \\ p_k(j,A) = P_k(h_j < h_A), & {}_AL_j = \sup\{s : x_s = j, 0 < s < h_A\}. \end{cases} \tag{2.1}$$

And let I_B or $I(B)$ be a characteristic function of set B, simply, $I_j := I_{\{j\}}$.

Suppose $X = \{x_s,\ s \geq 0\}$ is a strong Markovian jump process ([15] or [2]). Now using a shift operator θ_t in [3] and Markov property, one can obtain the following exact formulae of mean loss of the system.

Theorem 2.1
For $k < N$,

$$E_k Y_{tA} = \sum_0^{N-1} V(j) \int_0^t {}_AP_{kj}(s)\, ds \tag{2.2}$$

$$= E_k Y_A \sum_0^{N-1} {}_AP_{kj}(t)\, E_j Y_A, \tag{2.3}$$

where ${}_AP_{kj}(s)$ and $E_k Y_A$ are given by (3.9) *and* (3.11) *respectively in* Section 3.

The mean loss of the system can be also expressed in terms of distribution of relative last quitting time ${}_AL_j$ and prohibition potential ${}_Ag(k,j)$.

Theorem 2.2
For $k < N$,

$$E_k Y_{tA} = \textstyle\sum_0^{N-1} V(j)\,[\,{}_Ag(k,j) - {}_Ag(j,j)\; P_k({}_AL_j > t)\,] \tag{2.4}$$

$$= E_k Y_A - \textstyle\sum_0^{N-1} E_j(\int_0^{h_A} V(x_s)\, P_k({}_AL_{x(s)} > t)\, ds) \tag{2.5}$$

where ${}_Ag(k,j)$, $P_k({}_AL_j > t)$ and $E_k Y_A$ are given by (3.10), (3.12) *and* (3.11), *respectively.*

Proof

(i) First prove

$$P_k(j,\, A) = {}_Ag(k,j)/{}_Ag(j,j), \qquad k, j < N. \tag{2.6}$$

In fact, using strong Markov property, we have

$$\begin{aligned} {}_Ag(k,j) &= E_k \textstyle\int_0^{h_A} I_j(x_s)\, ds \\ &= E_k \left(\textstyle\int_{h_j}^{h_A} I_j(x_s)\, ds\,,\; h_j < h_A \right) \\ &= p_k(j, A)\, E_j \textstyle\int_0^{h_A} I_j(x_s)\, ds\,. \end{aligned}$$

This is just (2.6).

(ii) Moreover prove that

$$E_k \left(I_{(t<h_A)}\, E_{x_t} \textstyle\int_0^{h_A} I_j(x_s)\, ds \right) = {}_Ag(j,j)\; P_k({}_AL_j > t). \tag{2.7}$$

Using (2.6), the left-hand side of (2.7) equals

$$\begin{aligned} E_k(I_{(t<h_A)}\,{}_Ag(j,j)p_{x(t)}(j,N)) &= E_k(I_{(t<h_A)}\,\theta_t I_{(h_j<h_A)}) \\ &= {}_Ag(j,j)P_k(t<h_A,\,\theta_t\tilde{h}_j<\infty) \\ &= {}_Ag(j,j)P_k({}_AL_j>t). \end{aligned}$$

Where $\tilde{h}_j = \inf\{s :< h_A,\, x_s = j\}$.

(iii) Finally we have,

$$\begin{aligned} E_kY_{tA} &= \textstyle\sum_0^{N-1} V(j) \\ &\quad E_k\left(\int_0^{h_A} I_j(x_s)\,ds - I_{(t<h_A)}\int_t^{h_A} I_j(x_s)\,ds\right) \\ &= \textstyle\sum_0^{N-1} V(j)\left[\,{}_Ag(k,j) - {}_Ag(j,j)\;P_k({}_AL_j>t)\right]. \end{aligned}$$

Noting that ${}_Ag(i,j) = E_i\int_0^{h_A} I_j(x_s)\,ds$, we can immediately obtain (2.5).

The proof is complete. ■

Using Theorem 2.1, we can obtain the mean practical overhaul time, the mean faultless time and the mean accumulation of sojourn at states $0, 1, \ldots n$ ($<< N$), regarding the system as well run, and by taking $V = I$, $V = I_0$ and $V = I_{[0,n]}$ respectively. In fact, taking $V = I$, then from (2.3) obtain

$$E_k(t \wedge h_A) = E_kh_A - \textstyle\sum_0^{N-1}\,{}_AP_{kj}(t)\;E_jh_A. \tag{2.8}$$

Since $E_k(t \wedge h_A) = tP_k(h_A > t) + E_k(h_A,\, t > h_A)$, we can calculate the mean of the emergence overhaul time before the overhaul period in schedule, t.

From (2.2) and (2.3), taking $V = I_0$ and $V = I_{[0,n]}$, we respectively have

$$\begin{aligned} E_k\int_0^{t\wedge h_A} I_0(x_s)\,ds &= \int_0^t {}_AP_{k0}(s)\;ds \\ &= {}_Ag(k,0) - \textstyle\sum_0^{N-1}\,{}_AP_{kj}(t)\,{}_Ag(j,0) \qquad (2.9) \\ E_k\int_0^{t\wedge h_A} I_{[0,n]}(x_s)\,ds &= \textstyle\sum_0^{N-1}\,{}_AP_{kj}(t)\,{}_Ag(j,0) \qquad (2.10) \end{aligned}$$

Now we define the following ratios

$$\alpha_t = E_k\int_0^{t\wedge h_A} I_0(x_s)\,ds/E_k(t\wedge h_A), \text{ and} \tag{2.11}$$

$$\beta_t = E_k\int_0^{t\wedge h_A} I_{[0,n]}(x_s)\,ds/E_k(t\wedge h_A). \tag{2.12}$$

We suggest regarding them as numerical indexes of availability of the system with overhaul period according to schedule, t, and an allowed limit to fault number (or failure level), N. Differentiating with respect to t one may find maximum values of t, i.e optimum overhaul period in schedule maximizing the availability in both two senses above. It is also natural and useful to consider other ratios substituting $V(j)$ for I_j in (2.11) and (2.12).

3. Basic probability quantities in above section

In this section we calculate the following probability quantities concerned in Theorems 2.1 and 2.2 and definitions of the ratios given in the preceding section: ${}_AP_{kj}(s)$, $E_k\int_0^{h_A} V(x_s)\,ds$, *d.f.* of h_A and ${}_AL_j$, from which ${}_Ag(k,j)$ and $E_k h_A$ can be also obtained. Suppose $\sum_{j>i} q_{ij} > 0$, $i \in [0, N)$ for the following.

We first give a useful algebra conclusion as a lemma, which can be proved by induction. For this, let

$$\begin{cases} a_{ij} \geq 0,\ i \neq j,\ -\infty < a_{ii} = -A_i,\ A_i \geq \sum_{j\neq i} a_{ij} > 0, \\ j \in [1, n+m],\ i \in [1,n],\ \text{and}\ D_n = det(a_{ij})_{n\times n}. \end{cases} \tag{3.1}$$

Moreover, substituting the j^{th} column of D_n with

$$(a_{1\,n+i}, a_{2\,n+i}, \cdots, a_{n\,n+i})',$$

we denote the obtained determinant by $D_n^{(j)}(i)$,

$$D_n^{(j)}(i) = \det \begin{vmatrix} -A_1 & & & a_{1\,n+i} & & & \\ & \ddots & & \vdots & & & \\ & & -A_{j-1} & a_{j-1\,n+i} & & & \\ & & & a_{j\,n+i} & & & \\ & & & a_{j+1\,n+i} & -A_{j+1} & & \\ & & & \vdots & & \ddots & \\ & & & a_{n\,n+i} & & & -A_n \end{vmatrix} \tag{3.2}$$

where $j \in [1,n]$ and $i \in [1,m]$.

Lemma 3.1

$$(-1)^{n+1} D_n^{(j)}(i) \geq 0, \qquad j \in [1,n]\ \textit{and}\ i \in [1,m], \tag{3.3}$$

$$(-1)^n D_n \geq (-1)^{n+1} \textstyle\sum_{i=1}^m D_n^{(j)}(i), \qquad j \in [1,n], \tag{3.4}$$

$$(-1)^n D_n > 0, \qquad \textit{if}\ \textstyle\sum_{k>i} a_{ik} > 0\ \textit{and}\ \forall i \in [1,n]. \tag{3.5}$$

The conclusion given below as Lemma 3.2 is known, for example by Keilson [12], which can be proved by using methods similar to those used for the proofs of the following Theorems.

Lemma 3.2 *Let* ${}_AP(t) = ({}_AP_{ij}(t))_{N\times N}$ *and* $Q_N = (q_{ij})_{N\times N}$. *Then* ${}_AP(t)$ *satisfies forward equation*

$$\frac{d}{dt}\,{}_AP(t) \;=\; {}_AP(t)\,Q_N. \tag{3.6}$$

Because Q_N is finite, (3.6) shows that ${}_AP(t)$ has an exponential form. The sign Q_N will also denote $\det(q_{ij})_{N\times N}$ where there is no confusion.

Theorem 3.1 *Suppose* $k,\ j < N$, *then*

$${}_AP_{kj}(\lambda) \;=\; Q^{(k)}_{N\lambda}(-\delta_{\cdot j})/Q_{N\lambda}, \qquad Re(\lambda) > 0, \tag{3.7}$$

where $Q_{N\lambda}$ *denotes the result substituting* $-q_i$ *with* $-(\lambda + q_i)$ *in* $Q_N = (q_{ij})_{N\times N}$, $i \in [0, N)$, *and* $Q^{(k)}_{N\lambda}(-\delta_{\cdot j})$ *the result substituting the* k^{th} *column of* $Q_{N\lambda}$ *with column vector* $\delta_{\cdot j} = (0, 0, \ldots, 0, 1, 0, \ldots, 0)'$ *with the* j^{th} *element being 1.*

Proof Let β be the first jump of X and θ_β be the shift operator in Dynkin [3]. Noting the conditional independence of β and x_β, and the strong Markov property of X, and letting real $\lambda > 0$, we have

$$\begin{aligned} {}_AP_{kj}(\lambda) &= E_k\textstyle\int_0^{h_A} e^{-\lambda s} I_j(x_s)\, ds \\ &= E_k\textstyle\int_0^{\beta} e^{-\lambda s} I_j(x_s)\, ds + e^{-\lambda\beta}\,\theta_\beta \textstyle\int_0^{h_A} e^{-\lambda s} I_j(x_s)\, ds \\ &= \delta_{kj}/(\lambda + q_k) + \textstyle\sum_{i\notin A\cup\{k\}} (q_k/(\lambda + q_k))\ {}_AP_{ij}(\lambda)\,, \\ &\qquad\qquad\qquad\qquad\qquad\qquad k, j < N. \end{aligned}$$

Therefore

$$\textstyle\sum_{i\notin A\cup\{k\}} q_{ki}\, {}_AP_{ij}(\lambda) - (\lambda + q_k)\, {}_AP_{kj}(\lambda) \;=\; -\delta_{kj}, \quad k, j < N. \tag{3.8}$$

Letting $a_{ki} = q_{ki}$, $A_k = \lambda + q_k$ and fixing $j < N$, the system of equations (3.8) has unique system of solutions by Lemma 3.1, which is (3.7). ■

Corollary 3.1

$$\begin{aligned} {}_AP_{kj}(s) &= \textstyle\sum_{i=1}^{n_i} res_{\alpha_i}\left(e^{\lambda s} Q^{(k)}_{N\lambda}(-\delta_{\cdot j})/Q_{N\lambda}\right) &(3.9)\\ {}_Ag(k, j) &= Q^{(k)}_{N}(-\delta_{\cdot j})/Q_N. &(3.10) \end{aligned}$$

Where $-\alpha_i$ is the i^{th} zero of order n_i of $Q_{N\lambda}$, $\alpha_i > 0$, $\sum_i n_i = N$ and $Q_N(-\delta_{\cdot j})$ is a determination substituting q_{kj} with $-\delta_{kj}$ in Q_N.

Proof From Lemma 3.1, $(-1)^N Q_{N\lambda} > 0$ if real $\lambda > 0$. So polynomial $Q_{N\lambda}$ of degree N for complex λ has only negative zero. Taking the Laplace inverse transform, (3.9) is verified. Letting $\lambda \to 0_+$, one can obtain (3.10) from Lemma 3.1. ∎

Since $E_k \int_0^{h_A} V(x_s)\, ds = \sum_0^{N-1} V(j)\ {}_A g(k,j)$, using (3.10) it follows that:

Corollary 3.2 *Let $V(\cdot) = (V(0), V(1), \cdots, V(N-1))'$,*

$$\begin{aligned} E_k Y_A &= Q_N^{(k)}(-V(\cdot))/Q_N, \\ \textit{especially } E_k h_A &= Q_N^{(k)}(-I)/Q_N. \end{aligned} \tag{3.11}$$

Theorem 3.2 *Suppose k, $j < N$, then*

$$P_k({}_A L_j > t) = \textstyle\sum_0^{N-1} {}_A P_{ki}(t)\ p_i(j, A), \tag{3.12}$$

$$p_k(j, A) = \begin{cases} \bar{Q}_{Nj}^{(k)}(q_{\cdot j})/\bar{Q}_{Nj}, & \textit{if } k \neq j; \\ 1 & \textit{otherwise.} \end{cases} \tag{3.13}$$

Where $\bar{Q}_{Nj}$ is $N-1) \times (N-1)$ determination omitted both the j^{th} row and the j^{th} column in Q_N and $\bar{Q}_{Nj}^{(k)}(q_{\cdot j})$ is a result substituting the k^{th} column of $\bar{Q}_{Nj}$ with $q_{\cdot j} = (q_{1j}, q_{2j}, \cdots, q_{j-1j}, q_{j+1j}, \cdots, q_{N-1j})'$.

Proof

$$\begin{aligned} P_k({}_A L_j > t) &= E_k(I_{(t<h_A)}\, \theta_t I_{(h_j<h_A)}) \\ &= \textstyle\sum_0^{N-1} E_k(I_{(x_t=i,\, t<h_A)})\, E_i I_{(h_j<h_A)}. \end{aligned}$$

So (3.12) is true. If $k \neq j$,

$$\begin{aligned} p_k(j\ A) &= E_k\, \theta_\beta\, I_{(h_j<h_A)} \\ &= E_k\, E_{x(\beta)}\, I_{(h_j<h_A)} \\ &= \textstyle\sum_{i \notin A\cup\{k\}} (q_{ki}/q_k)\, p_k(j\ A) + q_{ki}/q_k. \end{aligned}$$

In (3.2), take that

$$a_{ij} = \begin{cases} q_{ij} & \text{if } i < j; \\ q_{i+1\,j} & \text{if } i > j. \end{cases}$$

$$A_i = \begin{cases} q_i & \text{if } i < j; \\ q_{i+1} & \text{if } i > j. \end{cases}$$

From Lemma 3.1, (3.13) is established. ■

Note:

$$P_k(h_A > t) = \sum_{j \notin A} P_k(x_t = j,\ t < h_A) = \sum_0^{N-1} {}_AP_{kj}(t)\ .$$

Thus up to now, all probabilities connected with both theorems and ratios in Section 2 are given.

4. Lth order moment

Let us return to discuss the loss of the system and to find the 2nd order moment of Y_{tA}. For simple, $\int_0^t := \int_0^t V(x_s)ds$.

Theorem 4.1 *For $k < N$*

$$\begin{aligned} E_k(Y_{tA})^2 &= E_k(Y_A)^2 - 2\sum_{j=0}^{N-1} E_k Y_A\, I(t < h_A, x_t = j)\, E_j Y_A \\ &\quad + \sum_0^{N-1} {}_AP_{kj}(t)\ E_j(Y_A)^2, \quad k < N. \end{aligned} \tag{4.1}$$

Proof Since $Y_{tA}^2 = \left(Y_A - I_{(t<h_A)} \int_t^{h_A}\right)^2$, expanding its right hand side and using Markov property, we find that

$$\begin{aligned} E_k(Y_{tA})^2 &= E_k(Y_A)^2 - 2E_k(Y_A\, I_{(t<h_A)}\, E_{x(t)} Y_A) \\ &\quad + E_k\left(I_{(t<h_A)}\, E_{x_t}(Y_A)^2\right). \end{aligned} \tag{4.2}$$

From this obtain that just as desired.
Let

$$\left\{ \begin{aligned} w_{tA}(k,j) &:= E_k\left(\int_0^t V(x_s)ds\, I(t < h_A, x_t = j)\right), \\ W_{tA} &= (w_{tA}(k,j))_{N\times N}, \quad P(t) = ({}_AP_{kj}(t))_{N\times N}, \\ V_N &= (V(0), V(1), \cdots, V(N-1))' \cdot I_{N\times N}, \quad k,j \in [0,N). \end{aligned} \right. \tag{4.3}$$

■

Theorem 4.2 *For $k < N$*

$$E_k(Y_A\, I(t < h_A,\ x_t = j)) = {}_AP_{kj}(t)\ E_j Y_A + w_{tA}(k,j) \tag{4.4}$$

and W_{tA} satisfies the following differential equation:

$$\frac{d}{dt} W_{tA} = W_{tA}\, Q_N + {}_AP(t) V_N, \qquad W_{0A} = 0.$$

So $W_{tA} = \exp\{Q_N t\} \int_0^t {}_AP(s)\, V_N \exp\{-Q_N s\}\, ds.$

Proof
$E_k(Y_A I(t < h_A,\, x_t = j)) = E_k \left(\int_0^t + \theta_t \int_0^{h_A}\right) I(x_t = j,\, t < h_A).$
Using Markov property one can obtain (4.4).

B_0, B_1 and B_2 respectively denote the events that there is no jump, only unique up-jump and down-jump in the interval $(t, t + \Delta t)$, then omitted $o(\Delta t)$

$$w_{t+\Delta t\, A}(k, j) = \sum_0^2 E_k \left(\int_0^{t+\Delta t} I(t + \Delta t < h_A,\, x(t + \Delta t) = j,\, B_i)\right).$$

The 1st term of right-hand side equals

$$E_k \left(\int_0^t I(t < h_A,\, x_t = j,\, \theta_t\beta > \Delta t)\right)$$
$$+V(j)\Delta t\, E_k I(t < h_A,\, x_t = j,\, \theta_t\beta > \Delta t).$$

Thus on B_0

$$\begin{aligned}&\left(w_{t+\Delta t\, A}(k, j) - w_{tA}(k, j)\right)/\Delta t\\ &= V(j)\ {}_AP_{kj}(t)\ P_j(\beta > \Delta t) + o(\Delta t)/\Delta t\\ &\to V(j)\ {}_AP_{kj}(t), \qquad \text{as } \Delta t \to 0.\end{aligned}$$

The 2nd term

$$\begin{aligned}&= \sum_{i<j} E_k(\int_0^t I(t < h_A,\, x_t = j)\, \theta_t I(\beta < \Delta t,\, x_\beta = j))\\ &\quad +\sum_{i<j} V(i)\Delta t\ {}_AP_{kj}(t)\ P_i(\beta < \Delta t,\, x_\beta = j).\end{aligned}$$

So on B_1,

$$\begin{aligned}&(w_{t+\Delta t\, A}(k, j) - w_{tA}(k, j))\, /\, \Delta t\\ &= \sum_{i<j} w_{tA}(k, i)\, P_i(\beta < \Delta t,\, x(\beta) = j)/\Delta t\\ &\quad +\sum_{i<j} V(j)\, {}_AP_{ki}(t)\ P_i(\beta < \Delta t,\, x(\beta) = j)/\Delta t\\ &\quad -\sum_{i>j} w_{tA}(k, i)\, P_j(\beta < \Delta t,\, x(\beta) = i)/\Delta t\\ &\to \sum_{i<j} q_{ij}\, w_{tA}(k, i) - \sum_{i>j} q_{ji}\, w_{tA}(k, j), \qquad as\ \Delta t \to 0.\end{aligned}$$

Similarly, on B_2, the limit is

$$\sum_{i>j} q_{ij}\, w_{tA}(k, i) - \sum_{i<j} q_{ji}\, w_{tA}(k, j).$$

Summing up the above results and noting $w_{tA}(k, j) = 0,\ j \in A$, (4.4) is true. ■

Using Theorem 4.2, the second moment of $E_k Y_{tA}$ is calculated from (4.1). Noting $e^{-\lambda Y_A}$ is $\mathcal{N}_{h_A}$-measurable (*cf* Wang [15]), it is easy prove that $e^{-\lambda Y_A} I_{(t<h_A)}$ is $\mathcal{N}_t$-measurable. For the reasons similar to the proof of Theorem 4.1, we have the following conclusion.

Theorem 4.3 *For* $l = 1, 2, \cdots$

$$E_k(Y_{tA})^l = E_k(Y_A)^l \sum_{i=1}^{l}(-1)^i C_l^i E_k\left((Y_A)^{l-i} I_{(t<h_A)} E_{x_t}(Y_A)^i\right). \tag{4.5}$$

Then calculate $E_k(Y_A^i I(x_t = j,\ t < h_A)$ similarly to Theorem 4.2. Finally, we have the higher ordinal moments of $E_k Y_{tA}$.

Another way to obtain higher ordinal moments is to consider the L-S transforms. Let $\Phi_{kA}(\lambda) = E_k e^{-\lambda Y_{tA}}$ and $\Psi_{kA}(\lambda) = E_k\, e^{-\lambda Y_A}$.

Theorem 4.4 *The L-S transformations*

$$\begin{aligned}\Psi_{kA}(\lambda) &= \Phi_{kA}(\lambda) + \textstyle\sum_{j=0}^{N-1}(\Phi_{jA}(\lambda) - 1)\, E_k\big(I(t < h_A,\ x_t = j) \\ &\quad \exp\{-\ \lambda\ Y_A\}\big), \end{aligned} \tag{4.6}$$

$$\Phi_{kA}(\lambda) = Q_{N\lambda V}^{(k)}(-\bar{q}.)/Q_{N\lambda V}. \tag{4.7}$$

Where $Q_{N\lambda V}$ *denotes the result substituting* $-q_i$ *with* $-(\lambda\ V(i) + q_i)$ *in* Q_N *and* $Q_{N\lambda V}(-\bar{q}.)$ *denotes the result substituting the* k^{th} *column of* $Q_{N\lambda V}$ *with a column vector of which the* k^{th} *element is* $q_k - \sum_{i=0, i\neq k}^{N-1} q_{ki}$.

Proof

$$E_k e^{-\lambda Y_{tA}} = E_k(e^{-\lambda Y_A},\ t < h_A) + E_k(\exp\{-\ \lambda\ \textstyle\int_0^t\},\ t > h_A).$$

Using shift and Markovian property, one can show that the right term is equal to

$$E_k\left(I_{(t<h_A)}\, e^{-\lambda Y_A}\, E_{x(t)} e^{-\lambda Y_A}\right).$$

Substituting and putting right, obtain (4.6).

Being similar to the proof of Theorem 3.1 and noting $\Phi_{iA}(\lambda) \equiv 1$, $i \in A$, we have

$$\sum_{i\notin A\cup\{k\}} q_{ki}\Phi_{iA}(\lambda) - (\lambda\ V(k) + q_k)\,\Phi_{kA}(\lambda) = -\left(q_k - \sum_{i\notin A\cup\{k\}} q_{ki}\right), \quad k < N. \tag{4.8}$$

Because of $q_k - \sum q_{ki} > 0$, from Lemma 3.1, (4.8) has unique system of solutions (4.7). ■

The following theorem is easily checked. Therefore, differentiating (4.6) and letting $\lambda \to 0_+$, one can obtain any ordinal moment of Y_{tA}.

Theorem 4.5 *Let* $U_{kj}(t) = E_k(I(x_t = j,\ t < h_A) \exp\{-\lambda Y_A\})$. *Then* $U_{kj}(0) = \delta_{kj}\, \Phi_{kA}(\lambda)$ *and* $U_{kj}(t)$ *satisfies the forward equation.*

References

[1] Blumenthal, R. and Getoor, R. (1968), *Markov Processes and Potential Theory*, New York, Acandemic Press.

[2] Chung, K.L. (1967), *Markov Chain with Stationary Transition Probabilities*. Springer.

[3] Dynkin, E.B. (1963), *Markov Processes*. Moscow.

[4] Gani, J. and McNeil, D.R. (1971), "Joint distributions of random variables and their integrals for certain birth-death and diffusion processes". *Adv. Appl. Prob.*, **3**, 339–352.

[5] Ge, Yubo (1987) "Expected loss and reliability of system with overhaul period and break quota", *Reliability Theory and Applications*. World Scientific, Singapore, 101–112.

[6] Ge, Yubo (1984), "The number of crossings and a group of probability formulae in birth-death processes". *Chin. Ann. of Math.*, **Ser. A, 5:3**, 371–378.

[7] Ge, Yubo (1985a), "Distributions of V-sojourn times on high-levels and the first passage times for birth-death processes". *Acta Math. Sinica*, **28:1**, 112–121.

[8] Ge, Yubo (1985b), "Moments of V-sojourn times for birth-death processes". *Acta Math. Sinica*, **28:3**, 414–426.

[9] Ge, Yubo (1985c), "On the last exit times of birth-death processes". *Acta Math. Sinica*, **28:4**, 565–572.

[10] Hjort, N.L., Natvig, B. and Funnemark, E. (1985), "The association in time of a Markov process with application to multistate reliability theory". *J. Appl. Prob.* **22**, 473–479.

[11] Hou, Zhenting and Guo, Tsingfeng (1978), *Homogeneous Countable Markovian Processes*. Chin. Sci. Publishing House.

[12] Keilson, R. (1979), *Markov Chain Models — Rarity and Exponentiality* . Applied Mathematical Sciences Series **28**, Springer-Verlag, Berlin.

[13] McNeil, R. (1970), "Integral functionals of birth and death processes and related limiting distribution", *Anna. Math. Statis.*, **41** (3), 480–485.

[14] Natvig, B. (1985), Multistate coherent system, In Encyclopedia of Statistical Sciences **5**, ed. N.Johnson and S. Kots. Wiley, New York.

[15] Wang, Zikun(1980a), *Birth-Death Processes and Markov Chains.* Chin. Sci. Publishing House.

[16] Wang, Zikun (1980b), "Sojourn times and first passage times for birth and death processes". *Scientia Sinica* **23**, 269–279.

[17] Wu, Lide (1963), "Distribution of Integral type of functional of homogeneous denumerable Markov processes". *Acta Math Sinica* **13** (1), 86–93.

[18] Wu, Rong (1980), "Distribution of the functional of birth and death processes". *Acta Math Sinica* **24** (3), 337–358.

[19] Yang, Chaoqun (1964), "Integral Type of Functional of Denumerable Markov Process and Boundary Properties of Birth and Death processes with Both-Side Boundaries", *Adv. Math Sinica* **7:4**, 397–424.