Time-Dependence in Markovian Decision Processes

Jeremy James McMahon

Thesis submitted for the degree of

Doctor of Philosophy

in

Applied Mathematics

at

The University of Adelaide

(Faculty of Mathematical and Computer Sciences)

School of Applied Mathematics

September, 2008

This work contains no material which has been accepted for the award of any other degree or diploma in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text.

I consent to this copy of my thesis, when deposited in the University Library, being made available in all forms of media, now or hereafter known.

SIGNED: ..... DATE: .....

I extend sincere gratitude to my two supervisors, Professor Nigel Bean and Professor Michael Rumsewicz, who have both inspired and guided me to complete this work. Their input and encouragement has been extremely beneficial and I thank them wholeheartedly for their friendship and support over the last few years.

I am grateful to all three of my parents for the love and reassurance they have given me, not just throughout my time as a PhD student. Without them, any hope of reaching this point in my education would at best be a distant dream. I particularly appreciate the editorial contributions of my mother who, despite supposedly limited mathematical knowledge, provided invaluable feedback.

Lastly, I thank my beautiful wife who has endured much whilst I have been working on this thesis. I love and cherish Sarah for being by my side, encouraging me and always believing in me. Now we may begin the next chapter of our lives.

For Olive.

Contents

List of Figures

List of Tables

Abstract

The main focus of this thesis is Markovian decision processes with an emphasis on incorporating time-dependence into the system dynamics. When considering such decision processes, we provide value equations that apply to a large range of classes of Markovian decision processes, including Markov decision processes (MDPs) and semi-Markov decision processes (SMDPs), time-homogeneous or otherwise. We then formulate a simple decision process with exponential state transitions and solve this decision process using two separate techniques. The first technique solves the value equations directly, and the second utilizes an existing continuous-time MDP solution technique.

To incorporate time-dependence into the transition dynamics of the process, we examine a particular decision process with state transitions determined by the Erlang distribution. Although this process is originally classed as a generalized semi-Markov decision process, we re-define it as a time-inhomogeneous SMDP. We show that even for a simply stated process with desirable state-space properties, the complexity of the value equations becomes so substantial that useful analytic expressions for the optimal solutions for all states of the process are unattainable.

We develop a new technique, utilizing phase-type (PH) distributions, in an effort to address these complexity issues. By using PH representations, we construct a new state-space for the process, referred to as the phase-space, incorporating the phases of the state transition probability distributions. In performing this step, we effectively model the original process as a continuous-time MDP. The information available in this system is, however, richer than that of the original system. In the interest of maintaining the physical characteristics of the original system, we define

a new valuation technique for the phase-space that shields some of this information from the decision maker. Using the process of phase-space construction and our valuation technique, we define an original system of value equations for this phase-space that are equivalent to those for the general Markovian decision processes mentioned earlier. An example of our own phase-space technique is given for the aforementioned Erlang decision process and we identify certain characteristics of the optimal solution such that, when applicable, the implementation of our phase-space technique is greatly simplified.

These newly defined value equations for the phase-space are potentially as complex to solve as those defined for the original model. Restricting our focus to systems with acyclic state-spaces though, we describe a top-down approach to solution of the phase-space value equations for more general processes than those considered thus far. Again, we identify characteristics of the optimal solution to look for when implementing this technique and provide simplifications of the value equations where these characteristics are present. We note, however, that it is almost impossible to determine *a priori* the class of processes for which the simplifications outlined in our phase-space technique will be applicable. Nevertheless, we do no worse in terms of complexity by utilizing our phase-space technique, and leave open the opportunity to simplify the solution process if an appropriate situation arises.

The phase-space technique can handle time-dependence in the state transition probabilities, but is insufficient for any process with time-dependent reward structures or discounting. To address such decision processes, we define an approximation technique for the solution of the class of infinite horizon decision processes whose state transitions and reward structures are described with reference to a single global clock. This technique discretizes time into exponentially distributed length intervals and incorporates this absolute time information into the state-space. For processes where the state-transitions are not exponentially distributed, we use the hazard rates of the transition probability distributions evaluated at the discrete time points to model the transition dynamics of the system. We provide a suitable reward structure approximation using our discrete time points and guidelines for sensible truncation,

using an MDP approximation to the tail behaviour of the original infinite horizon process. The result is a finite-state time-homogeneous MDP approximation to the original process and this MDP may be solved using standard existing solution techniques. The approximate solution to the original process can then be inferred from the solution to our MDP approximation.