A Markov Decision Process Model For Logistics in Supply Chain

BY

Jazeem Abdul Jaleel

A Thesis Presented to the
DEANSHIP OF GRADUATE STUDIES

KING FAHD UNIVERSITY OF PETROLEUM & MINERALS
DHAHRAN, SAUDI ARABIA

In Partial Fulfillment of the
Requirements for the Degree of

MASTER OF SCIENCE

In

INDUSTRIAL AND SYSTEMS ENGINEERING

DECEMBER 2015

KING FAHD UNIVERSITY OF PETROLEUM & MINERALS

DHAHRAN- 31261, SAUDI ARABIA

DEANSHIP OF GRADUATE STUDIES

This thesis, written by **Jazeem Abdul Jaleel** under the direction his thesis advisor and approved by his thesis committee, has been presented and accepted by the Dean of Graduate Studies, in partial fulfillment of the requirements for the degree of **MASTER OF SCIENCE IN INDUSTRIAL & SYSTEMS ENGINEERING**.

Feb, 8, 2016

Dr. Mohammad AlDurgam
(Advisor)

8-2-2016

Dr. Hesham K. Al-Fares
Department Chairman

8/2/2016

Dr. Shokri Selim
(Member)

Dr. Salam A. Zummo
Dean of Graduate Studies

8/Feb/2016

Dr. Salih Duffuaa
(Member)

8/2/16

Date

Dedication

I dedicate this work to my parents, wife, sisters and grandparents for their love and support during my studies abroad

ACKNOWLEDGMENTS

I thank God Almighty for his support in this journey of my life.

I thank King Fahd University of Petroleum and Minerals (KFUPM), for giving me this opportunity, and every facility required to complete my Master's degree program.

My sincere gratitude and thanks to my thesis advisor Dr. Mohammad AlDurgam for introducing me to the area of Markov Decision Process (MDP). His patience and unwavering support during the entire duration of my thesis has made it a memorable experience. His guidance helped me in the duration of this research and thesis write-up. I could not have imagined a better advisor for my thesis or a better mentor for the last two semesters of my Master's degree.

I thank my committee members, Prof. Shokri Selim and Prof. Salih Duffuaa for creating time to review my work and for their insights.

I thank all my professors in KFUPM with whom I have taken courses, Prof. Hesham K. Al-Fares, Prof. Muhammad Ben-Daya , Dr. M. Riaz and Dr. Syed Mujahid. I find myself very fortunate to be your student, and I have greatly enjoyed your classes.

I thank the staff in the Systems Engineering Department, Mr. Mervin Monsalve, Mr. Khasim Farooqui, and Mr. Hussain Suwaiket; for their continued and sincere assistance in my departmental needs.

I thank my parents Abdul Jaleel and Zubaida Jaleel, wife Hashma Nazeer, sisters Jumana Jaleel and Jazna Jaleel, and grandmothers Beevikutty Abdul Kader and Zuhara Abdulla for their prayers, well wishes and advices.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

MDP	:	Markov Decision Process
RL	:	Reverse Logistics
FL	:	Forward Logistics
SCM	:	Supply Chain Management
i.i.d	:	Independent and Identically Distributed

ABSTRACT

Full Name : Jazeem Abdul Jaleel

Thesis Title : A Markov Decision Process Model for logistics in supply chain.

Major Field : Systems Engineering

Date of Degree : [December, 2015

Markov Decision Process (MDP) models have been widely used in decision making under uncertainty. MDP has been applied in various fields of study – healthcare, maintenance management, transportations problems, production planning, robotics, and others.

In this thesis, two related-problems are addressed. First, an MDP model for reverse logistics (RL) published in the International Journal of Production Research, 2007 is studied, and corrected. A counter example is provided to show that the set of claimed sufficient conditions, to guarantee the existence of threshold policy, are incorrect. The correct way of approaching the problem is provided, and a new set of sufficient conditions for two-period planning horizon are provided, yet, the n-period problem is believed to be very complicated and difficult to characterize.

In the second part of the thesis, a generic MDP capacity planning model, which can be used in forward logistics capacity planning, is provided. The optimal policy is characterized over two different partially ordered state space, and based on the optimal policy's structural properties, a revised value iteration algorithm is provided.

ABSTRACT (ARABIC)

ملخص الرسالة

الاسم الكامل: جزيم عبدالجليل

عنوان الرسالة: نموذج ماركوفي متعدد الحالات للخدمات اللوجستية في ادارة سلاسل الامداد

التخصص: هندسة النظم

تاريخ الدرجة العلمية: ديسمبر 2015

نعتبر النماذج الماركوفية كاملة المشاهدة من الطرق واسعة الانتشار لنمذجة الانظمة متعددة الحالات، وذلك لدعم اتخاذ القرارات المثلى في حالة عدم اليقين. لقد تم استخدام النماذج الماركوفية في العديد من المجالات، مثل ادارة الصيانة، النقل، تخطيط الانتاج، الروبوتات وغيرها الكثير.

في هذا البحث تمت دراسة مسألتين ذوات صلة. الأولى تتعلق بأحد الأبحاث المنشورة في عام 2007 في المجلة العالمية لبحوث الانتاج. حيث طور بعض الباحثين نظرية وشروط لضمان الحصول على قرارات منتظمة الترتيب لنموذج ماركوفي متعدد الحالات، يعنى النموذج بتمثيل الخدمات اللوجستية العكسية ضمن ادارة سلاسل الامداد. في هذه الدراسة تم نقض تلك النظرية بواسطة مثال رياضي، ثم تم تعديل شروط النظرية الأصلية و تصحيحها، وذلك بعد توضيح الطريقة الصحيحة الواجب اتباعها وسبب الخطا في النظرية الأصلية.

في الجزء الثاني من الأطروحة، تم تطوير نموذج ماركوفي متعدد الحالات كأداة دعم قرار تخطيط الخدمات اللوجستية ضمن سلاسل الامداد، تم تطوير نظريات و شروط جديدة على عناصر المسألة وذلك لترتيب القرارات المثلى على عناصر المجال للمسألة. وأخيرا تمت الاستفادة من خصائص المسألة للحصول على طريقة حل سريعة وفعالة.

CHAPTER 1

INTRODUCTION

1.1 Introduction

This research addresses the application of Markov Decision Process (MDP) in logistics. By logistics capacity, we refer to the resources that enable us to meet the logistics demand. Consider a distribution warehouse, where the key resources represent the number of trucks used for distribution. Assuming random demand, a dynamic decision making framework is needed to deal with this capacity planning problem. The objective is to determine the optimal capacity level for each time epoch (decision making point) of the future; taking uncertainty into consideration. An MDP model facilitates the mathematical formulation of such a problem.

For a company marketing a product, logistics operations can be divided into Forward Logistics (FL) and Reverse Logistics (RL). Forward logistics involves the operations of furnishing the customers with their demand for the product. By contrast, Reverse Logistics (RL) involves the operations of facilitating the return of the products from the customers. There are many differences between the nature of FL and RL in terms of forecasting, costing, product quality etc., which makes it a complex operation (Tibben-Lembke and Rogers 2002). The reason for the returns could be numerous – repair, planned service, disposal, technical update, environmental responsibilities and so on (Nikolaou, Evangelinos, and Allan 2013). From a general standpoint, FL is linked with revenue generating operations, and RL with expenses. But if a company does not maintain a robust RL system it can lose its customers to other competitors (Rogers and Tibben-Lembke 2001).

1.2 General Statement of the problem

In this research, we address two problems. The first, a reverse logistics problem available in the literature is studied and corrected. In the second, an MDP model for forward logistics is presented and structural properties proved.

Here we give a brief description of the two problems:

1.2.1 A note on Serrato et al. (2007), and a corrected proof

Serrato et al. (2007) proposed a Markov Decision Process (MDP) model, to determine whether a company is to perform its reverse logistics activities either in-house, or by outsourcing them. In this part, through a counter example, it is shown that the theorem in Serrato et al. (2007) does not guarantee the existence of a structured optimal decision policy. Then, a new set of sufficient conditions are developed to guarantee the existence of structured optimal policy for a two-period problem.

1.2.2 An MDP model to optimize logistics capacity in forward logistics

In the second section of the thesis, a general MDP model is presented for optimizing logistics capacity for forward logistics. Further, the structural properties existing in the model under very realistic assumptions are investigated. The advantages of the structural properties in terms of computational effort is quantified and presented.

This general MDP model formulation and its structural properties are not only useful in optimizing logistics capacity, but also, it can be applied to other areas of Supply Chain Management (SCM).

1.3 Motivation

The motivations for pursuing this thesis are as follows:

1. The advantages of supply chain management include: improving resource allocation and customer satisfaction, reducing inventory and total cost of production, increasing system efficiency and profit margin, and more.
2. MDP has proven efficiency in dynamic decision making environments.
3. To the best of our knowledge, most MDP structural properties, especially for MDP models defined over multi-dimension sate space, are based on strict assumptions on the model parameters, which is not the case here.
4. To the best of our knowledge, the proposed mathematical model and its cost elements does not exist in the literature.

1.4 Thesis Objectives

1. To reproduce the results of Serrato et al. (2007) and fix few mistakes in their work. This include:

   - Through a counter example the conditions in the addressed paper are shown to be insufficient to guarantee structured optimal policy.

   - To correct the cost structure of the addressed paper.

   - Propose a new set of sufficient conditions, to guarantee a threshold policy in case of a two-period problem, and highlight the complexity of the problem in case of a multiperiod problem.

2. To formulate an MDP model for capacity planning, which can be utilized to determine the optimal capacity levels in forward logistics. .
3. Prove the existence of structured policy for the model in 2, over two different partially ordered state space, under very realistic assumptions on the proposed model parameters.
4. Present advantages of the developed structural properties and their uniqueness in comparison to the similar results found in the literature.

1.5 Thesis Contributions

1. Improving the work of Serrato et al. (2007),and highlighting the complexity of the problem to guarantee existence of threshold policy. Also, obtaining a set of conditions that guarantee threshold policy for a two-period problem.

2. Formulation of a general MDP model that can be utilized in other applications. Proving the existence of a structured optimal decision policy for the developed model over two different partially ordered state space.

1.6 Thesis Organization

The rest of this thesis is organized as follows: In chapter 2, an introduction to MDP modelling is provided, this is to introduce the reader to the different elements of the MDP. In chapter 3, we provide a literature review which focuses on the applications of MDP in supply chain. In chapter 4, Serrato et al. (2007) is addressed and revised. In chapter 5, a general MDP model to optimize logistics capacity is proposed. In chapter 6, the conclusion of this thesis work and some suggested future work is presented.

CHAPTER 2

MARKOV DECISION PROCESS

In this chapter, an introduction to the Markov Decision Process (MDP) is provided

2.1 Introduction

MDP provides a mathematical framework for dynamic programming. Problems where the outcomes are partly in the control of the decision maker and partly random can be formulated as an MDP model.

When a system's condition may be completely described by a set of information, it (information) is attributed to its state. In a Markov chain, the state transitions follow Markovian property. Thus, an MDP is a Markov chain whose state transitions can be influenced by a decision maker.

2.2 Elements of an MDP model

1. Decision Epochs t: These are points in time where actions are applied to the system. In discrete time problems, decision epochs are discrete, and the time between epochs are usually fixed. In continuous time problems, the decision epoch are usually random points in time when specific events occur.

2. State space S: At each decision epoch, the status of the system is fully described by its state s. It consists of all relevant information – system variables, events and actions; required to completely explain the system transitions and rewards on application of an action. The set of all possible states for a system is its state space.

3. Action space A: At every decision epoch, the decision maker observes the state of the system, and chooses an action a that influence state transition and rewards. The set of all actions available to the decision maker, in a given state, is called the action space for that state A_s .

4. Rewards $R_t(s, a)$: At a decision epoch t, when an action is chosen, the decision maker receives a reward, that is dependent on the current state and chosen action. The reward may also be dependent on the next state, in this case expected reward can be calculated using probability theory.

5. State transition probability $p_t(s'|s, a)$: is defined as the probability that a system will hold a particular state s' in the next decision epoch given its current state s and chosen action a.

2.3 Decision rules and policies in MDP

Given the state and time epoch, a decision rule specifies what action to implement. The decision rule is Markovian; because information on the previous system states is immaterial in deciding the action at an epoch. $d_t: S \rightarrow A_s$

A policy specifies the decision rule to be implemented at every decision epoch and system state. Hence a policy is a sequence of decision rules. A policy is said to be stationary if it is the same for all epochs.

2.4 Objective in MDP modeling

An objective is the maximization (or minimization) of total expected rewards (costs) during the planning horizon. Different possible objectives for an MDP model are

- Maximize (minimize) total expected rewards (costs),
- Maximize (minimize) discounted sum of rewards.

2.5 Solution Methodology

A finite horizon MDP model can be solved using backward value iteration method, wherein the decision maker takes into advantage the Markovian property of the MDP model, as in the current and future rewards are independent of past states of the system. In value iteration methodology, the decision maker derives the optimum action, and optimum reward starting from the last epoch, and works his way back to the first epoch of the planning horizon. Doing which, he obtains the optimum action for the current system, and also a mapping of the optimum actions to different possible states in the future epochs.

Evaluating an MDP model by value iteration is an arduous task and will require huge computational effort for large problems. But if it can be ascertained that the solution policy will follow a structure, say the action will be monotonically increasing or decreasing, the computational effort required for determining the optimal policy, will reduce significantly because the whole range of actions need not be probed in such cases. Several theorems have been developed in this regards (Puterman 2009).

2.6 Applications of the MDP

MDP is used extensively in describing dynamic decision making problems in many fields of study. In his surveys, White (1985, 1988, 1993) listed out many areas of MDP application, to name a few:

- Water Resources
- Maintenance Operations
- Inventory management
- Finance
- Robotics

- Manufacturing

A google search on MDP gives about 700,000 results at the time of this research, this shows how widely it is being used.

CHAPTER 3

LITERATURE REVIEW

3.1 Introduction

This chapter gives an idea of the various applications of MDP in optimizing SCM operations. The significance of the study of existence of structural properties in the decision policies is highlighted in this chapter.

Supply chain is a mapping of operations between different dependent businesses entities namely suppliers, manufacturers, distributors and retailers, that work ultimately for meeting the demands of customers in the most efficient and profitable manner.

Supply chain consists of a comprehensive group of activities, ranging from the procurement of raw materials, to manufacturing of products, satisfying/ furnishing customer demands and managing returns (if any).

Planning is required for every activity at different levels. The levels of supply chain management (SCM) decision making is often divided to three levels namely:

- Strategic Planning

  Strategic planning are high level planning keeping the organization's mission in mind. The planning is usually done by company stakeholders. The scope of the planning is usually for long periods. Strategic plans serve as input to tactical management and operational planning levels

- Tactical Management

  Tactical planning focuses on the implementation of the strategic plans. It is done usually by mid-level managers.

- Operational level

  Operational plans define the routine functioning and decisions of an industry. A clear cut plan on what decisions to take for the employees for their activities that will be in accordance to the strategic plans developed.

The application of MDP models in determining decision policies in different areas of supply chain is vast. In this literature review we focus on the application of MDP models in the tactical management and operational level planning of different sections of supply chain.

3.2 MDP models in Tactical management planning

Manufacturing problems:

Chien et al. (2012) works on a problem where several different type of a product (semi-conductor) are produced. Through empirical data the demand transition matrix was obtained for different demand states for each product. The machines have capacity for its respective product which maybe migrated to other products for a transfer rate and cost. The objective is to minimize the expected discounted cost over a finite horizon. The cost components include capacity shortage cost, capacity idling cost and capacity migration cost. The capacity expansion and migration decisions are made in each epoch. The model incorporates a lead time in the implementation of the expansion decision during which no decision can be made in further changing its capacity. The paper does not investigate for existence of any structure in the optimum decision policy.

He (He) takes a different take to the problem definition in Chien et al. (2012), here each product requires different operations for its production and the machines are categorized as per the operations they perform. The decisions are made for changing capacity of each machine type and also assignment of of machines to different operations. Costs are incurred for changing machine capacity, switching machine operations, inventory holding and for machine operations. The objective is to minimize the expected cost over a finite horizon. Solution is determined by backward value iteration and the paper does not investigate for existence of any structure in the optimum decision policy.

Wu and Chuang (2010) works on making optimal capacity decisions for a production industry considering price and demand fluctuations. The problem considers an industry making 2 products. There are two machines in the industry. One machine (dedicated) is used solely for the purpose of producing one product. A second machine is flexible and can produce either of the two products. The manner/ policy by which the machine capacity are allocated is fixed. That is the flexible machine will first use its capacity to satisfy the demand for the product that can only be manufactured by it before allocating its capacity to the other product. The action is the purchase quantity of each machine type at each epoch. The objective is to maximize profit in a finite horizon. The rewards include profits by selling products, purchase cost for increasing machines, shortage cost for not meeting demand, idling cost of machines and salvage value at the end of planning horizon. A more efficient algorithm than the standard backward value iteration algorithm was used to determine optimum decision policy. Structural properties in optimal solution was proof.

Ahiska and Kurtul (2014) considers a system where the product can be manufactured new or can be remanufactured from recovered units/ returned units. The demand and prices for manufactured and remanufactured items are different. As expected remanufactured items will cost lower to manufactured items. The demands for manufactured and remanufactured items and the number of returns at each period are stochastic and independent to each other. The paper also studies the effect of one-way substitution, which is once the remanufactured items are used up, its demand can be fulfilled by manufactured items sold at the price of remanufactured items. The decision to be made at each time epoch is the number of items to be manufactured and remanufactured. The problem is solved as an infinite horizon MDP. Howard Morton policy iteration (1971) method is applied to find optimal decision policy.

Garcia- Alvarado et al. (2014) introduces a manufacturing model that also aims at reducing the amount of carbon emissions to the atmosphere. This is done by making it a part of the total cost function and also by putting constraints on the decision variables. Different types of decision policies were analyzed. The paper does not study for any structural properties in the optimum policy.

Procurement:

Li (2013) deals with inventory systems with reverse logistics where the demand and returns are Markovian. Both finite and infinite horizon scenarios are addressed in the paper. Existence of structural decision policies were proved.

De Cuypere et al. (2013), an optimal order quantity is determined by accommodating fluctuations in demand and market price. The lead time is considered stochastic and no backlogging is allowed in the model. There can be no more than one ongoing order at any time and the demand is always unit item in this model. The time between two order deliveries is taken to be distributed geometrically and the demand is Bernoulli distributed. Price fluctuations is represented by three parameters that define the mean price as well as the spread (variation). The actions at decision epochs being the number of items to be placed in order. The objective is to minimize the long term discounted cost. The system state is explained by three variables – the onhand inventory, the present price level and the units in order. The paper does not study for any structural properties in the optimum decision policy.

Bendre and Nielsen (2013) works on a similar problem as De Cuypere et al. (2013), he analyzes the problem considering different properties for lead times. No structural properties were established.

Chen et al. (2010) model consists of a central warehouse whose inventory is replenished by the supplier/ manufacturer and the warehouse satisfies the demands of its subsidiaries. The action is taken every month on the set of delivery quantities for the subsidiaries from the warehouse and for the warehouse from the supplier. The system state is defined by the inventory levels at the subsidiaries and at the warehouse. The objective is to minimize long run expected discount cost over an infinite horizon. The costs include transportation costs from supplier to warehouse and warehouse to customers, warehouse operating costs, holding costs for warehouse and its subsidiaries, penalty cost for failing in meeting the subsidiary demand. Hence the inventory policy is to be modelled for both the warehouse and the subsidiaries. The paper used a modified policy iteration algorithm with action elimination procedures to help reach a near optimal solution.

Sales:

Thomas (1974), in his paper works on a model where the demand is dependent on the selling price. It reasonably assumes that the probability density function of demand stochastically reduces with increase in demand. Decision is made on the selling price and production amount in this finite horizon model. The state is represented by the inventory level. The objective is to minimize discounted cost in finite horizon. The costs incurred involves the revenue costs (negative), holding and backlog costs. The optimal policy was initially conjectured to be a structured type where production decision follows an (s,S) policy and the price decision follows with respect to it. But with some counterexamples it was suggested that the initial structured policy may only be assured under certain assumptions.

Federgruen (1999) work is a generalization of Thomas (1974) and provides a complete study of the MDP model. The paper studies separately cases for bi-directional price change and price markdowns. The paper addresses both finite and infinite horizon problems. The system state in the MDP at each period is represented by in-hand inventory. At every period decision is made on the inventory (units to order) and the pricing of the goods. The objective is to maximize the profits. The paper addresses optimal policies for several objectives – finite horizon discounted expected profit, infinite horizon discounted expected profit, and average expected profit. The costs involved are inventory holding costs, backordering costs, unit ordering costs and sales revenues. The paper proves the existence of structured policy through submodularity and convexity of its reward function and hence designs a modified value iteration technique to reach faster optimal decision policy.

3.3 MDP models in Operational level planning

Procurement:

In Kingsman (1969), the retailer is to determine a purchasing policy when the commodity costs are stochastic. The paper takes the case where the demand for the coming periods is perfectly forecasted but the commodity prices changes. The decision maker is to decide on the optimal order quantity. The objective is to minimize the expected cost of procuring a commodity over a finite horizon. Cost components are purchasing and inventory holding costs. The system state is explained by 2 variables - the inventory and the current price. The new system state depend on the decision made and the last state. The action is to decide the number of goods to purchase. The optimal decision policy are price breaks that decide how much period demands should the inventory purchase be done for.

Kalymon (1971) worked on a similar problem as Kingsman (1969) structure but the demand was uncertain and the commodity price varied following Markovian property. The states are described by current inventory and present price. The actions consist of two parts - (i) Is there any purchasing done in the epoch (ii) If yes, how much quantity will be ordered. The objective is to minimise the expected discounted cost. Cost components are purchasing cost, holding cost and shortage costs. Both finite and infinite planning horizon models studied. Structural properties are realized and exploited to reduce the computational effort in determining optimum decision policy.

Golabi (1985) problem was similar to Kingsman (1969), deals with deterministic demand and no shortages/ back ordering allowed. The differences being he considered period dependent cost components, addressed both finite and infinite horizon problems and also investigates the structural properties in the optimal decision policies. The ordering cost probability distribution may have different parameters for different time periods. States are represented as stocks in hand and the realized order price. The decision maker is to decide how many units to order each period. The objective is to minimizing average expected costs. Costs include ordering and inventory costs. The optimal decision policy is represented in terms of price breaks similar to Kingsman (1969).

Snyder (1975), worked on a continuous time review stock model to find optimal order quantities for a supplier with fixed lead times. The supplier makes order decisions each time it realizes a demand. The demand quantities are independent and identically distributed probability density function. The demand realization times are independent of the demand quantity and also follows an independent and identically distributed probability density function. The continuous time problem was reduced to a discrete stage MDP model. The state is defined by the system inventory. The decision is the amount of units to order at every inventory review point. The objective is to minimize expected cost over a finite horizon. Cost components are fixed ordering cost, variable order cost, inventory costs and backlog costs. The optimal policy was conjectured depending on the fixed ordering cost. If the fixed ordering order cost is 0, the optimum policy will tend to maintain an ideal system inventory; else it will follow an (s,S) ordering policy.

Puranam and Katehakis (2014) works on a problem where the firm builds up inventory by participating in auctions in order to fulfil its market demands. There are two phases in each cycle, in the first cycle bidding takes place in fixed number of auctions. The number of bidders in each bid is uncertain but is known right before the auction. In each bid, every person seals their bid amount in a sealed document and the person who makes the highest bid wins and his inventory increases by one unit. After all the bids take place phase 1 is complete, in phase two the demands are realized and it follows an independent identical distribution every period. The objective is to maximize the expected present value of profit for an infinite planning horizon through an optimal strategy. The models incorporates penalty for demands not fulfilled. The sales price is fixed and if any units are not sold they are stored in inventory incurring a holding cost. States are represented as a triplet – number of remaining auctions, number of bidders and current inventory level. In phase 2, the first two components of the state will be 0. In phase 1 the action is the bid amount, while in phase 2 there is no action. The paper derives the structural properties of the decision policy under certain assumptions.

Feng et al. (2014) works on a multi-product system with correlated demand and joint-replenishment costs for products. Products are divided into groups depending on their characteristics and each group experiences Poisson arrivals of demand. The number of units of each product in a particular demand has a joint density distribution. The problem has been formulated as an infinite horizon MDP to compute optimal policies. The structure of the optimum policy was analyzed through numerous numerical examples to assist developing an algorithm to reach a near-optimal decision policy.

Ahiska et al. (2013) works on a problem where we have one retailer who has two sources for its goods. One of which is a reliable source and the other is an unreliable one. The unreliable source may not meet the order but it has a lesser unit cost for the goods. If order has been made and not fulfilled the retailer incurs the loss of fixed ordering cost. The status of the unreliable source follows Markovian property. The decision maker has to decide how many goods to order from each source to meet its stochastic demand. The objective is to minimise the expected cost per period over an infinite horizon. The cost components includes fixed ordering costs, unit purchasing costs, inventory costs, backordering costs and lost sales costs. The problem has been modelled as an MDP with state defined as two random variables, the inventory of the retailer and the status of the unreliable source. The paper studies the structure of the optimum policies through numerical examples.

Transportation, shipments, freight:

Kleywegt et al. (2002) uses MDP to address a problem where daily inventory routing decision is to be made by the supplier by taking into consideration the inventory levels at its customers, customer stochastic demands and supplier transportation constraints. The supplier decides on the quantity of goods to be transported to each customer. The constraints are the number of vehicles available and their capacity, the inventory limits at the customers, time constraints (as delivery to different customers will take different times). The paper considers only direct deliveries that is transportation starts from the supplier to a single customer and then back. This assumption reduces the dimension of the action space to help solve the model. Nonetheless, the model is applicable for several practical scenarios. The hard problem was subdivided into sub-problems to get near optimal policies. The objective being to maximize the expected discounted value (revenue minus costs) over an infinite horizon. Cost components are sales revenues at each customer, inventory holding costs, transportation costs different for each customer and also depends on number of dispatches, and shortage cost. A general MDP model with states as the customer inventories and the actions as the transport quantities will be difficult to solve and will require huge computational time due to the large state and action space. The paper suggests different algorithms to find near optimal policies and help in value approximations of the optimal values. In the proposed algorithms, the problem is first divided into sub-problems for each customer assigning different transportation constraints and then the value function is approximated by solving the system as a knapsack problem that assigns transportation resources to each customer. MDP state in submodels is explained by customer inventories and number of visits available to each customer. In a Later paper Kleywegt et al. (2004) removed this assumption of direct deliveries making the problem NP hard.

Loading/ unloading

Rida (2014) proposed an MDP model to optimize loading/ unloading operations of trucks at different locations. In the problem defined, the containers on ships are unloaded using quay crane into trucks which then travel to the yard for offloading using yard crane. The crane service times are taken as exponential and the truck arrival rate at the crane queues are taken as Poisson distribution. The actions are taken with respect to addition/ reduction of shuttle trucks and allocation/ liberation of yard crane. Costs are incurred when trucks wait in the queue and when the cranes wait on the trucks. The objective is to minimize the expected discounted cost over an infinite horizon. The system state in the MDP formulation is represented by the shuttle trucks in each crane queues. The paper does not study for any structure in the optimum policy. The optimum policy is determined by application of either value iteration/ policy iteration techniques.

Kang, et al. (2008) also works optimizing loading acivities at the ports. Similar to Rida (2014), containers are offloaded at berth onto trucks which are then transported to yard for storage. The operation speed and bottlenecks depend on the number of active quayside cranes at the berth, the number of gantry cranes at the yard and the total number of trucks active for the system. The decision is made on the number of cranes at offloading and loading ends and on the number of trucks in the system. The objective is to minimize the total cost in unloading and storage of a given number of containers in a horizon subject to uncertainty. Cost function varies from Rida (2014) as it focusses on operation costs. Existence of structural properties were not evaluated.

Higginson and Bookbinder (1995) provides a MDP model to work on shipment consolidation, scenarios where decisions has to be made to satisfy demands either immediately or to keep accumulating demand till the transportation trucks are more filled for dispatch. The paper considers costs for keeping inventory for delayed shipments and costs for dispatching shipments. The paper does not prove the existence of any structured policy but conclude it through several numerical examples.

Hoffmann (2013) presents an aircraft cargo management problem. The aircraft has limited capacity in volume and weight. There is a fixed set of orders that can be realized during the planning horizon. Each order has its own volume and weight demands. Every period will see only one order demand being requested. The probability for each order is given. Acceptance of an order will result in reduction in the aircraft's cargo's availability in volume and weight for the subsequent periods. Each order has its own per unit margin income. The income is calculated either by an order's weight or volume requirement depending on which gives a better income. This is made possible by making use of a standard shipping weight to volume ratio. The decision maker is to decide to accept an order or reject it. The system state is defined in each epoch by available weight and volume. Through counterexamples the author proves that certain types of structured decision policy will not exist. The paper proposes certain heuristic methods to get near-optimal decision policies.

Production, manufacturing and rationing

In White (1965), we have a manufacturing process that is not perfect and the number of defectives is stochastic. The manufacturer gets an order quantity from a customer for a fixed number of good/quality products. The decision maker decides how much to produce taking into consideration expected defectives. If the demand is not met completely by non-defective products, another batch has to be processed to meet the remaining demand. The objective is to minimize the expected cost in meeting the customer's demand. Cost components are setup costs (fixed) and production cost (variable). System state is represented by the number of non-defective units remaining to be sent to the customer. The new state depends on the previous state, decision made and the defectives realized in the manufacturing process. MDP model is that of an absorbing state stochastic dynamic program. The optimal production sizes are determined by backward value iteration.

Huang and Iravani (2007) addresses a scenario with single manufacturer and two retailer scenario where goods are produced by manufacturer and kept in their storage. Each retailer demands goods from the manufacturer when their inventory is depleted (reorder point = 0). The order size for each retailer may be different but fixed. No backlogging is allowed in the model, if the manufacturer does not have the goods in inventory, it looks into alternate sources and facilitates the fulfilment of the retailer order. This would incur a penalty cost and can be different values for each retailers. The objective is to minimize the total discounted cost over a finite time horizon. Cost components are Inventory holding costs, and shortfall costs. The decision epochs are the demand arrival instances from the retailers and the production completion instances of the manufacturer. The state space are represented as the club of three inventories – the manufacturer and the two retailers. The manufacturer is to make decision on how to meet the demands of the retailers and when it should stop production or resume. Hence there are three decisions for the decision maker – 1.) keep the manufacturing system idle, 2.) production of goods and 3.) rationing goods as required when an order is received from the retailer. The retailer inventory is available for manufacturer to facilitate decision making. The customers arriving at the retailers follow independent Poisson process. When an order is received from a retailer that has relatively low penalty cost it may be optimal to fulfil only a part of it from the manufacturers inventory if we expect orders from retailer with a higher penalty cost which warrants importance of fulfilling them internally. Lead time for order fulfillment is taken as negligible. Existence of structured policy for production and rationing was shown based on the supermodularity and convexity of the cost function. Helper et al. (2010) works on a similar model and analyses the effect of different levels of information sharing between supplier and its 2 retailers. Unlike Huang and Iravani (2007) where rationing policy is a decision, here a fixed rationing policy is used. The paper considers a lead time of one time period to meet retailer demands and does not study the existence of any structured decision policy.

Benjaafar and Elhafsi (2012) MDP model deals with a supplier serving two different customer classes. One class consist of patient customers in the sense backordering is possible with their orders, the other is impatient, that is if the order for them is not immediately realized the sale is lost. In both the classes customers arrive following Poisson process with different rates. Processing times for goods follow exponential distribution. Since decision are made each time a demand is requested and when production is complete, the problem is inherently a continuous time MDP. Uniformization is used to convert continuous time MDP to discrete time MDP. The states are defined by 2 variables – onhand inventory and backorder levels. Decisions to be made on realizing an order are: should order requests be processed (fulfilled), backordered (for patient class) or should it be rejected. In addition, decision is to be made if items are to be produced and how should they be allocated once ready – kept in inventory or to satisfy a backorder. Cost components in the system are inventory costs, backordering costs, and rejection costs (same for lost sales). The objective is to minimize infinite horizon discounted cost. The decision policy for both production and allocation are completely characterized by threshold regions of the state space. This structured policy is enabled by the submodularity of the cost function. The paper proposes 5 heuristic methods that enables availing easier and practical inventory policies. These heuristics were compared with the optimal policies to give insight on their effectiveness.

Lin et al.(2014) gives an MDP formulation for an industry manufacturing different types of a given product (in this case-TFT-LCDs) using a set of common and specific resources. Each product type has its own Markovian demand. The industry has multiple manufacturing locations that can together work to meet the demand. Each manufacturing location may have its own type of raw materials (LCD sheet size). At each location there are tools dedicated for the production of each product. The tools for the same product may be different at different locations. Thus because of different sheet sizes and tool for the same product, the sheets are converted into goods at different consumption rates at each location. The profit in meeting a particular product demand can hence be different for different locations. The maximum production amount of a particular product in a period at a particular location is dependent on the capacity of the product tool as well as how the sheets (raw material) available at the location are allocated for different product types. The objective is to maximize overall profit in a finite horizon. The MDP state is defined by the quantity of each product tool available at different locations and demand state for each product. The action taken at each epoch are the purchasing of product tools at locations and tool allocation at every location to meet experienced demand. The optimal policy is determined by backward value iteration. Existence of any structured policy was not analyzed in the paper.

Sinha and Krishnamurthy () optimize an assemble to order production system furnishing multiple products. Each product having its subcomponents that can be processed either in-house or externally. The in-house facility has its own service rates and production costs for the different subcomponents it can manufacture. The external facility's service rates and costs for the subcomponents processing are provided for decision making. Subcomponents are stored in their buffer locations and when an order is received they are assembled instantaneously to give the product. The objective is to minimize the cost per unit time over an infinite horizon. Cost incurred for inventory, backlogging, and production. The actions are use of the in-house and external production facilities – the capacity in use and its configuration to which subcomponents they process. Solving the general model makes the computations tedious and impractical. For simplicity and ease in finding solution, the system was divided into subsystem and optimization done over each subsystem. Through numerical examples insights were given on the structural properties of the model.

Pang et al. (2014) considers a firm following make-to-stock production system for single product satisfying multiple demand classes. Demands from different classes follow Poisson distribution. On demand arrival the decision is made to accept or reject the demand request. A lost sales cost is incurred either when a demand request is rejected or when there is no inventory to meet the demand. If a demand is entertained the customer pays a price that is dependent on the demand class. The size of the production batch is fixed. The processing time for each batch is uncertain and follows a density function such that its failure rate increases over time. The objective is to maximize the expected discounted profit over an infinite horizon. The model incorporates fixed and variable costs for production, lost sales cost, inventory holding costs and revenues from sales. The actions made at each time epoch are rationing of demand requests and production order decision in case of no outstanding orders. The structure of the optimal policy for production and rationing was characterized. For production, an optimal policy follows a reorder point, and for rationing policy, it depends on time dependent critical inventory levels for each demand class.

Tiemessen et al. (2014) works on an industry with one production line that processes multiple products. Goods produced are stocked and used when demands are realized. Demand follows Poisson distribution and the lead time at production line is exponentially distributed at different rates for different products. Backlogging is allowed when stock-out happens. The model does not attribute for switching costs and switching times. The objective is to minimize average inventory holding and backorder costs in an infinite horizon. The decision maker decides the product being processed in the production line. The system state in the MDP model is the inventory of each product and the product in process at the production line. Decision epochs are at every instant a demand is experienced or when a production is complete. To facilitate determination of optimal policies the continuous time MDP is transformed to discrete time MDP through uniformization. The optimal structure was attained by relative value iteration and studied empirically. Deviation in the objective function when using base stock policies were empirically studied. Structural properties were analyzed using numerical examples and heuristic methods were proposed for determining near-optimal decision policies.

Nakashima et al. (2004) accounts for a single product production facility that supports remanufacturing, that is goods are also returned back by consumers for reprocessing, this may be repair of defective items, replacement of parts or simply servicing. And the amount of goods coming for remanufacturing is taken as a fixed percentage of goods in use by customers which is referred to as virtual memory. The goods coming for remanufacturing are automatically processed and added to inventory. The decision maker is to decide the number of goods to manufacture (new) at each epoch. The model allows backordering and objective is to minimize expected cost per period. Structural properties of optimal decision policy are not analyzed.

Sebnem Ahiska and King (2010) addresses a single product recoverable manufacturing system. The demand and goods returned are stochastic. There may be capacity constraints at remanufacturing or manufacturing inventory locations. The lead times and cost components for manufacturing and remanufacturing were varied and their effects in solution were studied. The objective being to reduce effective cost per unit time in a long run. Cost components include setup, unit production and holding costs each for manufacturing and remanufacturing process, Backordering costs, disposal costs for recoverable units, and cost due to lost sales. The decision at the epochs are to decide how many goods should be produced new and how many should be reworked. The paper compares near-optimal structured decision policies that are easy to implement with the optimum policies to give insights. In Ahiska and King (2010), the trends in the demand and return rates during a product's life cycle is addressed to facilitate manufacturing/ remanufacturing decisions.

Vercraene and Gayon (2013) model represents a production system requiring a sequence of operations and where returns are join in any stage of the production line. Each production stage has its own lead time that is exponentially distributed. After each stage the in-manufacture item is stored at a buffer till required by the next stage. Customer demands are realized following a Poisson process. Returned goods can join at any stage depending on the reason of its return. Thus returns are modelled as Poisson process with different parameter at each stage. The decision policy defines when to produce at each stage in this infinite horizon model. The paper analyzes the structure of the optimum policy.

Vila-Parrish et al. (2012) deals with the production of perishable goods. At the start of the planning horizon raw materials are purchased and kept in inventory. The raw materials have relatively long shelf life and their purchase is done only at the end of its shelf life which is taken as one cycle. Decisions made on processing of perishable goods from the raw materials. In case the demand in the period is greater than the goods produced, it may be addressed in two scenarios. One way being expediting the goods from an external source, and the second way: immediate processing of goods in-house provided the raw materials are available, if not they are acquired from external source. Each scenario is taken as a different problem. The quantity of raw materials to be purchased at the beginning of the planning horizon and the number of goods (having a shelf life of only one period) to be processed at each period are decision variables in this problem. The objective is to reduce the expected cost per cycle in this finite horizon problem. Structural properties of the optimal decision policy were proved. In an earlier work Vila-Parrish et al. (2008), the problem was such that different patient classes had different demand distributions for a medicine. Patients get admitted to the hospital following Poisson distribution with different rates at different patient classes. In time the patients may transition between different classes or stay in the same class or may exit the system (absorbing node). This is represented as a markov chain. At the start of each period decision is made as to the number of medicines to process and quantity of raw materials to be placed in order. In case of stockout of finished goods, there is a penalty per unit. If the raw material is also unavailable the penalty is larger. The objective is to minimize the cost over the finite planning horizon. The decision policies were suggested by simulation.

Haijema et al. (2005) considers a problem where there are two types of demand for a perishable product (blood platelets in a blood bank). One is a general demand independent of product age (as long as it is within shelf life), the second demand is for fresh inventory. First in- first out policy (FIFO) is used for the first demand while the second demand is met by last in first out (LIFO) policy. Both the demands are taken as Poisson distributions with different rates. The action taken is the production decision at start of every period. The objective is to minimize the expected cost per week in an infinite horizon. The system state in MDP formulation is the inventory levels by age. The complexity is very large to find optimal policies by value iteration. Near optimal policies are obtained through simulation. Two production policies were proposed that are near optimal and easy to implement. Their value functions were compared to give insights on the same.

Iravani, et al. (2012) considers a firm produces two levels of a single product (can be high quality and low quality). The production time for each follows exponential distribution. The demands for each product type follows Poisson distribution. The firm produces only one item at a time. The demand for low quality item may be substituted by high quality item. The decision maker decides on demand substitution and production decisions. The objective is to minimize the firm's total discounted cost over an infinite horizon. The paper studies the structural properties of the optimum decision policy.

Other Perishable inventory operations

Parlar (1985) considers the case where a vendor deals with perishable goods that has age based selling price. Goods are taken to expire after two periods. The selling costs for new items and one period old items are different and their demands to customers are also different. The Problem takes into account the consideration that some customers (given by a fixed probability) will be satisfied with the alternative (older or newer product) if their preferred goods (new/ one period old) are not available. Original demand for each age class is random and represented by a probability distribution function. Each period decision is to be made of the number of new items to process/ order. The objective is to maximize the expected profit per unit time for an infinite time horizon. Cost components are age based sales cost and order cost per unit. The paper using linear programming approach to determine optimum decision policy for the modelled MDP.

Haijema (2014) considered inventory policies for perishable goods with fixed shelf lives which required to take into account the age of the product and not just the quantity available when making orders. Demand is taken to be uncertain. Orders are made at the start of the period and received half-way into the period. Disposal of expired items are done at the end of each period. Each period is divided into 2 epochs. The start of the first epoch is when the order is placed for goods. The second epoch starts with the receipt of these goods and at its end, decision is made on the number of goods to dispose. Both amount of goods to order and dispose are decision variables. The objective function in this discrete time infinite horizon MDP is to minimize the expected cost per unit time. The cost component contains inventory costs, shortage costs, discount on aged goods and disposal costs. The state space has 3 parts: day of the week, period epoch and an age based inventory vector that helps track of the goods as per their remaining shelf lives. Depending on the type of organization/ industry the usage policy may be first in first out (FIFO) or Last in first out (LIFO). The optimal decisions were found by value iteration method as described in Puterman (2009). Results for 5 different policy types were tabulated and compared.

Sales planning:

Wu et al. (2011) considers a firm selling different levels of a single product (differentiated in quality). Each quality level goods cater to different customer group and have different uncertain demands. Downward demand substitution is allowed. The decision maker decides how a demand request is fulfilled. The objective is to maximize expected profit in a finite horizon. The structural properties of the optimum decision policy was studied.

Tiemessen and Van Houtum (2013) optimizes the scheduling of a repair center catering to a single product with different parts. Each part is kept in stock for replacement in the event a failure occurs. Failure of these parts follow a Poisson distribution each having their respective error rate. When a failure occurs the part is replaced and the failed part is placed in queue for repair, repair time for each part is exponentially distributed. If there is no stock of the part for replacement, backorder takes place and the system is in downtime till the part is furnished. The decision maker decides on the part being worked by the repair center. The objective is to attain a repair policy such that the annual average downtime is minimized. The paper analyze certain heuristic policies.

In Zhang and Kallesen (2008) we have two competitors marketing the same product. The problem is modelled in the perspective of one of the competitor. Customer choose the competitor with the lowest price every period. The probability of a customer arriving in a period depends on the minimum price offered in the period. The decision maker decides the selling price for a period. The objective is to maximize the revenue in a finite time horizon. System state represented by the inventory of the decision maker's company, and the price offered its competitor. Structural properties in decision policy were not analyzed.

Van Wijk et al. (2013) addresses a setup where we have several warehouses and the customer demands at different regions are facilitated at the appropriate warehouse. Each warehouse inventory are replenished by i.i.d (independent and identically distributed) exponential lead times. Of the set of warehouses, we have one quick response warehouse that in addition to facilitating customer demands has the ability to supply goods to other warehouses when their inventory has depleted to help furnish their demands. Customer demands at the local warehouses and quick response warehouse follow a poison process. The warehouses follow a base stock policy, the local warehouses and the quick response warehouse inventory experience one for one replenishment from a central supplier (or warehouse) with infinite inventory. When a local warehouse runs out, its demand may be fulfilled by a quick response warehouse (QRW) for a penalty cost if the QRW has inventory. In case the QRW does not have inventory or does not choose to fulfill a particular demand request from a local warehouse, the demand is fulfilled externally by emergency procedure (EP) at a higher penalty cost. The objective is to minimize the long run cost per unit time. The cost components are inventory costs at local warehouses and QRW, Quick response costs, and Emergency procedure costs. Taking into consideration expected customer demands to itself and the demand priorities for other warehouses, the quick response warehouse is to decide on whether to satisfy a request or not. The lead times for replenishment are exponentially distributed. Existence of structural properties were proved by showing the supermodularity and convexity of the value function.

CHAPTER 4

A NOTE ON SERRATO et al. (2007), AND A CORRECTED PROOF

4.1 Introduction

The purpose of this chapter is to correct the work of Serrato et al. (2007) presented in the paper titled "A Markov decision model to evaluate outsourcing in reverse logistics" published in the International journal of Production Research. In this paper, Serrato et al. (2007) presented their MDP formulation and presented a result for the existence of structural properties which is not correct and deserved correction. Section 4.3, states the results of Serrato et al. (2007) and shows it to be incorrect with a counter example. In section 4.2, the nomenclature, model assumptions, and the MDP formulation are presented as in Serrato et al. (2007). In section 4.4, a corrected formulation is proposed, which is in line with the reward definition stated in Serrato et al. (2007). New sufficient conditions to guarantee the existence of a threshold optimal policy for the two-period problem are presented in section 4.5. In section 4.6, we give the conclusion for this chapter.

4.2 Overview of Serrato *et al.* (2007)

This section provides a summary of the notations, assumptions, and model definitions as in Serrato et al. (2007)

4.2.1 Nomenclature

The notations in Serrato et al. (2007) are as follows:

k_t: Reverse Logistics (RL) capacity held by the firm at the beginning of period t

L: Length of the product life cycle: Duration of sales for a product.

n_t: Number of units sold and not returned at the end of period t, $n_t = S_t - w_t$

r: The probability that an unreturned sold item is returned in the next period

s_t: Amount of units sold by the firm during period t.

S_t: Cumulative sales experienced by the firm upto period t,

$$S_t = \sum_{i=1}^{t} s_i$$

T: Length of the planning horizon, T = L+W

t: Decision epoch, t=1,…, $T-1$, where the decision epoch t represents the end of period t

w_t: Cumulative number of units returned upto period t,

$$w_t = \sum_{i=1}^{t} x_i$$

W: Duration of time the returns for the product are managed after the sales period.

x_t: Number of units returned in period t.

X_l: Binomial Random variable with parameters l and r: The number of successes in l Bernoulli trials with probability of success $= r$.

4.2.2 Problem Definition and Model Assumptions

Serrato *et al.* (2007) considered the problem of making the decision of outsourcing (reverse Logistics) RL activities through a third party by studying the system parameters, namely, the existing system capacity and total items returned. When a company makes this strategic decision, the contract with the third party lasts till the end of the planning horizon (T). Serrato *et al.* (2007) proposed an MDP formulation for this problem with the objective of maximizing the profits.

The assumptions and conditions in Serrato *et al.* (2007) are as follows:

1. The sales in each period is known.
2. Every item that is sold and not yet returned has a fixed probability of return r in the following period.
3. The capacity of the firm's reverse logistics (RL) is taken as continuous.
4. If the returns exceeds the RL capacity, a penalty is incurred to facilitate its disposal or emergency service. No returns are carried forward to the next period.

5. If it is decided that the RL operations are continued internally, the capacity will be made equal to the expected returns in the following period.
6. Once RL is outsourced, the firm's capacity is salvaged and the decision is held for the remaining planning horizon(T).
7. All product returns after the planning horizon will incur a penalty, as the internal RL capacity is salvaged at the end of time T, and any third party RL contract, if undertaken, expires at the end of T.

4.2.3 Model Definition and formulation:

An MDP model consists of system states, actions, transition probabilities and reward/ cost functions as presented in Puterman (2009). Next we define these as in Serrato *et al.* (2007).

States: At any time epoch, t, of the planning horizon T, the system state is described by two state variables - the firm's RL capacity and the total number of goods returned until time t (k_t, w_t). The states are taken as partially ordered with respect to w_t in the MDP model. At the beginning $(t = 0)$, the system state is assumed to be $(k_0, 0)$.

Actions: The decision maker has two possible actions at the end of each time epoch. Action is described by 'a'.

$a_t(k, w)$: Is the action at the end of time t, when the system is in state(k_t, w_t).

a=0: RL continues to be done internally. The firm's capacity is changed to the expected number of returns in the next period i.e. k_{t+1} = E$[x_t]$ = $n_t r$.

a=1: RL is outsourced till the end of the study horizon. The firm's capacity is set to 0 and remains unchanged for the rest of the planning horizon, $k_{t+1} = k_{t+2} = \cdots k_T = 0$.

Transition Probabilities:

Now the state transition probabilities are defined, i.e. the probability of a future system state (in the next decision epoch) given the current system state and the action taken. Considering the fixed return probability r (assumption 2), the transition probabilities follow a binomial distribution. Let $P\left(X_{n_t} = j\right)$ be the probability of getting j successes in n_t Bernoulli trials, then

$$p_{t+1}[(n_t r, w_t + j)|(k_t, w_t), 0] = \begin{cases} P\left(X_{n_t} = j\right) = \binom{n_t}{j} r^j (1-r)^{n_t - j} & for\ j = 0,1,..,n_t \\ 0 & otherwise \end{cases} \tag{4.1}$$

Given state (k_t, w_t) and action 0 (continue internal RL), the probability of having j returns in the next period; hence $w_{t+1} = w_t + j$, is given by equation 4.1, where n_t represents the items that have been sold but not yet returned by the customers $(n_t = S_t - w_t)$. This can be computed, as the sales in each period is perfectly forecasted (assumption 1). As per assumption 5, we see that the new capacity is equal to the expected number of returns in the next epoch $(k_{t+1} = n_t r)$.

$$p_{t+1}[(0, w_t + j)|(k_t, w_t), 1] = \begin{cases} P\left(X_{n_t} = j\right) = \binom{n_t}{j} r^j (1-r)^{n_t - j} & for\ j = 0,1,..,n_t \\ 0 & otherwise \end{cases} \tag{4.2}$$

In case of opting for external RL, the system capacity is set to 0 as per assumption 6.

Rewards and model dynamics:

Cost components:

c_1 : Cost of increasing the firm's capacity by one unit – unit investment cost: ($/ capacity unit)

c_2 : Cost of decreasing the firm's capacity by one unit – unit disinvestment cost: ($/ capacity unit)

c_3 : Fixed internal capacity maintenance cost. ($/ capacity unit/ period)

c_4 : Labour cost. Processing cost per item returned under internal RL ($/ unit)

c_5 : Shortage cost. Penalty or emergency cost incurred per unit of the demand exceeding system capacity under internal RL. Penalty cost is also incurred by the system for the units returned after the planning horizon. ($/ unit)

c_6 : Capacity salvage value. Revenue, in salvaging unit capacity, of the system when shifting to third party RL or when salvaging capacity at the end of the planning horizon ($/ unit)

c_7 : Outsourcing cost. Processing cost per item returned under external RL ($/ unit)

At the end of each period, the decision maker decides, either, to continue RL internally, or opt for third party RL for the remaining planning horizon (T). The expected reward in the subsequent period is dependent on the system state and the decision made.

The reward function is represented as $R_{t+1}[(k_t, w_t), a]$.

$R_{t+1}[(k_t, w_t), a]$ is the expected reward in period t +1, after choosing an action a at the end of period t, when the system is in state (k_t, w_t).

The reward function for decision a=0, i.e continuing reverse logistics internally, is given as

$$R_{t+1}[(k_t, w_t), 0]$$
$$= -c_1(n_t r - k_t)^+ - c_2(k_t - n_t r)^+ - c_3 n_t r$$
$$- c_4 E[min(x_{t+1}, n_t r)] - c_5 E[(x_{t+1} - n_t r)^+]$$

Here the first term, c_1, accounts for costs incurred for an increase in capacity (investment); the second term, c_2, for any decrease in capacity (disinvestment); the third term, c_3, for maintenance of the system capacity; the fourth term; c_4, the cost in processing the RL in the period, t +1, and the fifth term, c_5, the penalty in case the number of returns exceeds the system's capacity in the period, t +1, as returns are not carried forward to the next period (assumption 4). The equation can be rewritten by substituting X_{n_t} for x_{t+1} to give:

$$R_{t+1}[(k_t, w_t), 0]$$
$$= -c_1(n_t r - k_t)^+ - c_2(k_t - n_t r)^+ - c_3 n_t r \qquad (4.3)$$
$$- c_4 E\left[min\left(X_{n_t}, n_t r\right)\right] - c_5 E\left[\left(X_{n_t} - n_t r\right)^+\right]$$

The expected reward function for decision a=1, which assumes outsourcing of RL operations until the end of the planning horizon, is given in Serrato *et al.* (2007) as:

$$R_{t+1}[(k_t, w_t), 1] = c_6 k_t - c_7 \left(n_t - \sum_{l=t+1}^{T} s_l (1 - (1 - r)^{T-l}) \right) \qquad (4.4)$$

In the above equation the first term, c_6, accounts for salvaging the system's capacity. The second term, c_7, accounts for the expected outsourcing cost for the entire remaining planning horizon.

Terminal reward is taken as

$$R_{T+1}[(k_T, w_T), a] = u_T(k_T, w_T) = c_6 k_T - c_5 n_T \tag{4.5}$$

The terminal reward accounts for the revenue in salvaging any remaining capacity, c_6, and the penalty costs in processing the remaining items after the study horizon c_5.

Given an initial system state (k_0, 0), the decision maker must determine the decision rule at the end of each period that will maximize the total expected reward. The value function that gives the expected reward from the end of period t till the end of the study horizon, when following optimum decision policy, is given as $u_t(k_t, w_t)$:

$$u_t(k_t, w_t) = max\begin{cases} R_{t+1}[(k_t, w_t), 0] + \sum_{j=0}^{n_t} p_{t+1}\big((n_t r, w_t + j)|\ (k_t, w_t), 0\big) u_{t+1}(n_t r, w_t + j) \\ R_{t+1}[(k_t, w_t), 1] \end{cases} \tag{4.6}$$

This function can be rewritten as:

$$u_t(k_t, w_t) = max\begin{cases} R_{t+1}[(k_t, w_t), 0] + \sum_{j=0}^{n_t} P\big(X_{n_t} = j\big) u_{t+1}(n_t r, w_t + j) \\ R_{t+1}[(k_t, w_t), 1] \end{cases} \tag{4.7}$$

Let the expected reward when opting for internal RL at period t be $g_t\big((k_t, w_t), 0\big)$ and the expected reward when opting for external RL at period t be $g_t\big((k_t, w_t), 1\big)$. Then,

$$g_t\big((k_t, w_t), 0\big) = R_{t+1}[(k_t, w_t), 0] + \sum_{j=0}^{n_t} P\big(X_{n_t} = j\big) u_{t+1}(n_t r, w_t + j)$$

$$g_t\big((k_t, w_t), 1\big) = R_{t+1}[(k_t, w_t), 1]$$

The optimal policies can be determined using the backward-value iteration policy algorithm (Puterman 2009).

4.3 A counter example to Serrato *et al.* (2007)

In this section we list the sufficient conditions stated in Serrato *et al.* (2007) to guarantee a structured optimal policy. Then, we provide a counter example when the conditions hold true, however, the resulting optimal decision policy is not structured.

4.3.1 Sufficient conditions in the paper

Serrato *et al.* (2007) states that, under the following cost assumptions and bounds on the item's return probability, there exists a structured optimal policy.

The assumptions:

$c_1 \geq |c_2|$ $\quad$ $c_3 \geq c_2$ $\quad$ $c_4 < c_7$ $\quad$ $c_7 < c_5$ $\quad$ $c_5 \geq c_1 + c_3 + c_4$

The bounds on the return probability are given as:

$$r \geq \frac{c_5 - c_7}{c_5 - c_1 - c_3 - c_4}$$

$$r \leq \frac{c_7 - c_4}{c_1 + c_3}$$

Under these conditions, the optimal decision policy is stated to be a monotone, non-decreasing in the system states partially ordered, over the cumulative units returned, *w*. That is, for a given period, if the optimal action for state $(k1, w1)$ is to outsource RL activities to a third party (a=1), then for all states $(k2, w2)$, such that $k2 = k1$ and $w2 \geq w1$, the optimal policy will be to outsource RL activities, a=1. Next, we provide a two-period, $T = 2$, counter example to show that these conditions are insufficient to guarantee a structured optimal policy as defined earlier.

4.3.2 Counter-example

Consider the following two-period problem, with initial conditions, cost parameters, and a return probability given as:

$T = 2$; c_1=20; c_2=18; c_3=20; c_4=30; c_5=150; c_6=0; c_7=100;

10 items already sold before the start of the planning horizon.

Hence, S=s_0=$S_0 = S_1 = S_2$=10; $s_1 = 0, s_2 = 0$

r=0.65; k_0=2;

We see that all the conditions stated in section 4.2.1 are satisfied:

For $T = 2$; equations 4.5 and 4.7 become:

$$u_2(k_2, w_2) = c_6 k_2 - c_5(S_2 - w_2)$$

$$u_1(k_1, w_1) = \max \begin{cases} g_1\big((k_1, w_1), 0\big) = R_2[(k_1, w_1), 0] + \sum_{j=0}^{n} P(X_n = j) u_2(n_1 r, w_1 + j) \\ g_1\big((k_1, w_1), 1\big) = R_2[(k_1, w_1), 1] \end{cases}$$

As in the original paper, the solution for the problem was obtained using the value iteration algorithm:

Table 1: Counter-example result

State (k,w)	**(2,0)**	**(2,1)**	**(2,2)**	**(2,3)**	**(2,4)**	**(2,5)**	**(2,6)**	**(2,7)**	**(2,8)**	**(2,9)**	**(2,10)**
$g_1\big((k,w),0\big)$	-1014.2	-910.4	-809.2	-708.7	-603.1	-502.5	-400.4	-293.2	-218.1	-136.6	-36.0
$g_1\big((k,w),1\big)$	-1000.0	-900.0	-800.0	-700.0	-600.0	-500.0	-400.0	-300.0	-200.0	-100.0	0.0
$u_1(k,w)$	-1000.0	-900.0	-800.0	-700.0	-600.0	-500.0	-400.0	-293.2	-200.0	-100.0	0.0
$a_1(k,w)$	1	1	1	1	1	1	1	0	1	1	1

The solution gives the following unstructured optimal decision rule $a_1(k, w)$:

For all w≠7, optimum action is: a=1 (3^{rd} party RL)

For ‘w’=7, optimum action is: a=0 (internal RL)

Hence, we arrive at a three-region policy, contrary to the, at most, two-region policy implied by (Serrato *et al.* (Serrato, Ryan, and Gaytán 2007)).

4.3.3 Explanation of this contradiction

The conditions in Theorem 4.7.4 in Puterman (2009) followed by Serrato *et al.* (2007) are sufficient to ascertain superadditivity of optimal value functions over a fully ordered state space. Since the stated definition in Serrato *et al.* (2007) is a partially ordered set, in terms of cumulative returned items, w_t, the proof need not stay valid, especially when the partial order doesn’t survive to the next period. Consider the optimality equation 4.7:

$$u_t(k_t, w_t) = \max \begin{cases} R_{t+1}[(k_t, w_t), 0] + \sum_{j=0}^{n_t} P(X_{n_t} = j) u_{t+1}(n_t r, w_t + j) \\ R_{t+1}[(k_t, w_t), 1] \end{cases}$$

Since a change in w also influences the system's capacity in the next period, $u_{t+1}(n_t r, w_t + j)$, under an internal RL action, the new system state $(n_t r, w_t + j\ )$ does not belong to the defined partial order. This is why Puterman's Theorem 4.7.4 is not sufficient for Serrato's problem.

4.4 Modified Formulation

In this section we correct a formula in Serrato *et al.* (2007).

When the decision is made to outsource RL activities, the paper by Serrato *et al.* (2007) follows that the decision is upheld till the end of the planning horizon (equation 4.4). The correct formulation of the expected reward function when opting for external RL should be:

$$\begin{aligned} R_{t+1}&[(k_t, w_t), 1] \\ &= c_6 k_t \\ &- c_7 \left(n_t (1 - (1-r)^{T-t}) + \sum_{l=t+1}^{T} s_l (1 - (1-r)^{T-l}) \right) \\ &- c_5 \left(n_t (1-r)^{T-t} + \sum_{l=t+1}^{T} s_l (1-r)^{T-l} \right) \end{aligned} \tag{4.8}$$

Where the, c_6, term accounts for revenue in salvaging capacity, the term, c_7, accounts for the RL costs for the expected returns till the end of the study horizon and, c_5, applies to the cost of managing the expected returns after the study horizon (assumption 7).

4.5 Structured optimal policy for a two period model

In this section, we provide a new set of sufficient conditions that guarantee the existence of an optimal monotone policy for Serrato's model. The optimal policy is characterized over the state space, partially ordered over the cumulative units returned, as discussed in section 2.

The nomenclature, reward function and model dynamics are as explained in the previous sections.

4.5.1 Model Formulation

For the final period, T=2, the value function described in equation 4.5 will be:

$$u_2(k,w) = c_6 k - c_5 n \tag{4.9}$$

For the first period (T=1), the optimal value function from equation 4.7 is given as:

$$u_1(k,w) = max \begin{cases} g_1\big((k,w),0\big) = R_2[(k,w),0] + \sum_{j=0}^{n} P(X_n = j) u_2(nr, w+j) \\ g_1\big((k,w),1\big) = R_2[(k,w),1] \end{cases} \tag{4.10}$$

Where $R_2[(k,w),0]$ is the expected reward in t =2, when doing RL internally from equation 4.3

$$R_2[(k,w),0] = -c_1(nr-k)^+ - c_2(k-nr)^+ - c_3nr - c_4E[\min(X_n,nr)] - c_5E[(X_n-nr)^+] \tag{4.11}$$

X_n is a random variable of the number of successes in n Bernoulli trials with the probability of success being r.

$R_2[(k,w),1]$ is the expected reward when doing RL externally, from equation 4.8

$$R_2[(k,w),1] = c_6k - c_7nr - c_5(n(1-r) + s_2) \tag{4.12}$$

$\sum_{j=0}^{n} P(X_n = j)u_2(nr, w+j)$ in equation 4.10 can be simplified using equation 4.9, as follows:

$$\sum_{j=0}^{n} P(X_n = j)u_2(nr, w+j) = \sum_{j=0}^{n} P(X_n = j)\big(c_6nr - c_5(S_2 - w - j)\big)$$

$$\sum_{j=0}^{n} P(X_n = j)u_2(nr, w+j) = c_6nr - c_5(S_2 - w - nr) \tag{4.13}$$

Next, we provide new set of sufficient conditions that guarantee a structured policy for this two-period problem.

4.5.2 Structured policy

In this section we state sufficient conditions on the problem parameters, that will guarantee the existence of a structured optimal policy.

We put forward the following Lemmas, not presented in Serrato *et al.* (2007), for the purpose of proving the main result as presented in Proposition 1.

Lemma 1:

If g is a superadditive function on X × Y, and for each x ∈ X, $\max_{y \in Y} g(x, y)$ exists. Then:

$$f(x) = max\left\{y' \in \arg\max_{y \in Y} g(x, y)\right\}$$

is monotone nondecreasing in x.

This lemma is stated and proved in Puterman (2009)(Lemma 4.7.1).

Lemma 2:

If $P(X_n = j)$ is the Binomial probability of getting j successes in n Bernoulli trials, when the probability of a success is r, then:

$$P(X_n = j) \leq \max(r, 1 - r)$$

Proof: Given in the appendix of this chapter – Section 4.6

Lemma 3:

If $P(X_n > j)$ is the probability of getting more than j successes in n Bernoulli trials, when the probability of a success in a Bernoulli trial is r, then:

$$P(X_{n+1} > j) = P(X_n > j) + rP(X_n = j)$$

Proof: Given in the appendix of this chapter – Section 4.6

Lemma 4:

If $P(X_n > j)$ is the probability of getting more than j successes in n Bernoulli trials, when the probability of a success in a Bernoulli trial is r, then:

$$P(X_{n+1} > j) = P(X_n > j-1) - (1-r)P(X_n = j)$$

Proof: Given in the appendix of this chapter – Section 4.6

Other Lemmas:

For the Binomial distribution where:

n: Number of Bernoulli trials

X_n: Random variable representing the number of successes in 'n' Bernoulli trials

r: The probability of success in a Bernoulli trial

We have the following relation:

Lemma 5: $E[min(X_n, nr)] - E[min(X_{n+1}, (n+1)r)] \geq -r \; \forall \, n > 0$

Lemma 6: $E[min(X_n, nr)] - E[min(X_{n+1}, (n+1)r)] = -r^2$ when $n = 0$

Lemma 7: $E[(X_n - nr)^+] - E[(X_{n+1} - (n+1)r)^+] \geq -r(1-r). \max(r, 1-r)$ $\forall \, n > 0$

Lemma 8: $E[(X_n - nr)^+] - E[(X_{n+1} - (n+1)r)^+] = -r(1-r)$ for $n = 0$

Proofs: Given in the appendix of this chapter – Section 4.6

Proposition 1:

If:

$$(1-r)\max(r, 1-r) \leq \frac{c_7 + c_6 - c_1 - c_3 - c_4}{c_5}$$

$$r \geq \frac{c_1 + c_3 + c_5 - c_7 - c_6}{c_5 - c_4}$$

Then, for the two-period problem, there exists an optimum decision policy $a_1(k, w)$, non-decreasing in (k, w), where (k, w) is partially ordered with respect to w.

That is, for a given period if the optimal action for state $(k1, w1)$ is to outsource RL activities to third party, $a = 1$, then for all states $(k2, w2)$, such that $k2 = k1$, and $w2 \geq w1$, the optimal policy will be to outsource RL activities.

Proof:

The model formulation is as follows:

$$u_1(k, w) = max\begin{cases} g_1\big((k, w), 0\big) = R_2[(k, w), 0] + \displaystyle\sum_{j=0}^{n} P(X_n = j) u_2(nr, w + j) \\ g_1\big((k, w), 1\big) = R_2[(k, w), 1] \end{cases} \tag{4.14}$$

For simplicity, we have removed the time subscripts on the state definition (equation 4.10.

As per lemma 1, if $g_1\big((k, w), a\big)$ is a superadditive function in W × A, then there exists a monotone non-decreasing policy in W.

Further, superadditivity of $g_1\big((k,w),a\big)$ in W × A means

$$\left[g_1\big((k,w),1\big)-g_1\big((k,w),0\big)\right]-\left[g_1\big((k,w-i),1\big)-g_1\big((k,w-i),0\big)\right]\geq 0$$

For all w and i.

The above condition is valid for all w and i if the equation holds true for all w and $i = 1$.

This can be easily shown by induction.

Hence, showing:

$$\left[g_1\big((k,w),1\big)-g_1\big((k,w),0\big)\right]-\left[g_1\big((k,w-1),1\big)-g_1\big((k,w-1),0\big)\right]\geq 0,$$

implies superadditivity of $g_1\big((k,w),a\big)$ in W × A. Expanding $[g_1(k,w,1) - g_1(k,w,0)] - [g_1(k,w-1,1) - g_1(k,w-1,0)]$, using equation 38, we find:

$$\begin{aligned}
&= R_2[(k,w),1]-R_2[(k,w-1),1] \\
&\qquad -\Bigg(R_2[(k,w),0]+\sum_{j=0}^{n}P(X_n=j)u_2(nr,w+j) \\
&\qquad -R_2[(k,w-1),0]-\sum_{j=0}^{n+1}P(X_{n+1}=j)u_2(nr,w+j-1)\Bigg)
\end{aligned} \tag{4.15}$$

Substituting equations 4.9, 4.11, 4.12, 4.13 in 4.15 and simplifying, we get:

$$[g_1(k,w,1)-g_1(k,w,0)]-[g_1(k,w-1,1)-g_1(k,w-1,0)] \tag{4.16}$$

$$= c_7 r + c_1\left[(nr-k)^+ - \left((n+1)r-k\right)^+\right] + c_2[(k - nr)^+$$

$$- (k - (n+1)r)^+] - c_3 r$$

$$+ c_4(E[min(X_n, nr)]$$

$$- E[min(X_{n+1}, (n+1)r)]) + c_5(E[(X_n - nr)^+]$$

$$- E[(X_{n+1} - (n+1)r)^+]) + c_6 r$$

Now, we intend to find the conditions that will guarantee the non-negativity of equation 4.16. For this purpose we divide the equation into three parts:

Part1: $c_7 r - c_3 r + c_6 r$.

Part 2: $c_1\left[(nr-k)^+ - \left((n+1)r-k\right)^+\right] + c_2[(k - nr)^+ - (k - (n+1)r)^+]$

Finding the lower bound of part 2, we have three possible cases:

Case 1: $k > (n+1)r > nr$ gives

$$c_1\left[(nr-k)^+ - \left((n+1)r-k\right)^+\right] + c_2[(k - nr)^+ - (k - (n+1)r)^+] = c_2 r$$

Case 2: $(n+1)r > k > nr$ gives

$$c_1\left[(nr-k)^+ - \left((n+1)r-k\right)^+\right] + c_2[(k - nr)^+ - (k - (n+1)r)^+]$$

$$= (c_1 + c_2)(k - nr) - c_1 r$$

Case 3: $(n+1)r > nr > k$ gives

$$c_1\left[(nr-k)^+ - \left((n+1)r-k\right)^+\right] + c_2[(k - nr)^+ - (k - (n+1)r)^+] = -c_1 r$$

Hence, taking the least lower bound value for part 2:

$$min\left(c_1\left[(nr-k)^+ - \left((n+1)r-k\right)^+\right] + c_2[(k- nr)^+ - (k- (n+1)r)^+]\right) \geq -c_1 r \tag{4.17}$$

Part 3:

$$c_4\big[E[min(X_n, nr)] - E[min(X_{n+1},(n+1)r)]\big] + c_5\big[E[(X_n - nr)^+] - E[(X_{n+1} - (n+1)r)^+]\big]$$

This part has two cases:

Case 1: when $n > 0$ using Lemma 5 and Lemma 7 we get

$$min\Big(c_4\big[E[min(X_n, nr)] - E[min(X_{n+1},(n+1)r)]\big] + c_5\big[E[(X_n - nr)^+] - E[(X_{n+1} - (n+1)r)^+]\big]\Big) \geq -c_4 r - c_5 r(1-r).\max(r, 1-r) \tag{4.18}$$

Case 2: when $n = 0$ using Lemma 6 and Lemma 8 we get

$$min\Big(c_4\big[E[min(X_n, nr)] - E[min(X_{n+1},(n+1)r)]\big] + c_5\big[E[(X_n - nr)^+] - E[(X_{n+1} - (n+1)r)^+]\big]\Big) \geq -c_4 r^2 - c_5 r(1-r) \tag{4.19}$$

Hence, we get two conditions, $n \neq 0$ and $n = 0$

Condition 1: when $n \neq 0$. Substituting equations 4.17, 4.18 in equation 4.16, we get:

$$c_7 r - c_1 r - c_3 r - c_4 r - c_5 r(1-r).\max(r, 1-r) + c_6 r \geq 0$$

Gives:

$$(1-r)\max(r,1-r) \leq \frac{c_7 + c_6 - c_1 - c_3 - c_4}{c_5}$$

Condition 2: when $n = 0$. Substituting equations 4.17, 4.19 in equation 4.16, we get:

$$c_7 r - c_1 r - c_3 r - c_4 r^2 - c_5 r(1-r) + c_6 r \geq 0$$

This gives

$$r \geq \frac{c_1 + c_3 + c_5 - c_7 - c_6}{c_5 - c_4}$$

Hence, if the problem's parameters in the two-period problem satisfy the following conditions:

$$(1-r)\max(r,1-r) \leq \frac{c_7 + c_6 - c_1 - c_3 - c_4}{c_5} \tag{4.20}$$

$$r \geq \frac{c_1 + c_3 + c_5 - c_7 - c_6}{c_5 - c_4} \tag{4.21}$$

then $g_1\big((k,w),a\big)$ will be superadditive in W × A, and there exists a monotone non-decreasing policy in W.

4.5.3 Numerical Example

Consider the following problem two period problem with cost parameters and return probability.

T=2; c_1=20; c_2=14; c_3=20; c_4=30; c_5=150; c_6=17; c_7=90;

S=10; 10 items already sold at the start of planning horizon.

r=0.75; k_o=2;

We see that the sufficient conditions are satisfied – equations 4.20 & 4.21

$$(1-r)\max(r, 1-r) \leq \frac{c_7 + c_6 - c_1 - c_3 - c_4}{c_5}$$

$$r \geq \frac{c_1 + c_3 + c_5 - c_7 - c_6}{c_5 - c_4}$$

And there exists a structured monotone nondecreasing policy in w.

Table 2: Numerical example for structured policy for two period problem

state (k,w)	(2,0)	(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)	(2,7)	(2,8)	(2,9)	(2,10)
$g_1((k,w),0)$	-800.1	-718.3	-634.1	-556.8	-476.9	-393.7	-307.0	-229.7	-165.2	-102.3	-28.0
$g_1((k,w),1)$	-1016.0	-911.0	-806.0	-701.0	-596.0	-491.0	-386.0	-281.0	-176.0	-71.0	34.0
$u_1(k,w)$	-800.1	-718.3	-634.1	-556.8	-476.9	-393.7	-307.0	-229.7	-165.2	-71.0	34.0
$a_1(k,w)$	0	0	0	0	0	0	0	0	0	1	1

For w<9, the optimum action is $a = 0$ (internal RL)

For w≥9, the optimum action is $a = 1$ (3rd party RL)

4.6 Conclusion

Serrato *et al.* (2007) is the first paper to model reverse logistics outsourcing decision using Markov Decision Process. In this chapter, we provide a counterexample to show that the addressed paper's theorem does not guarantee the existence of a structured optimal policy. For a two-period problem, we put forward a new set of sufficient conditions to guarantee existence of structured optimal policy. Still, the study of an n-period problem is extremely challenging, and left as an open problem.

4.7 Chapter 4 Appendix

Lemma 2:

If $P(X_n = j)$ is the Binomial probability of getting j successes in n Bernoulli trials, when the probability of a success is r, then:

$$P(X_n = j) \leq \max(r, 1 - r)$$

Proof:

Conditioning on the outcome of the first trial:

$$P(X_n = j) = \; P(X_{n-1} = j)(1 - r) + P(X_{n-1} = j - 1)r$$

Since both $P(X_n)$ and $r \in [0,1]$, it follows that:

$$P(X_n = j) \leq \max(\, P(X_{n-1} = j), P(X_{n-1} = j - 1)) \tag{4.22}$$

Similarly, we have in the same principle:

$$P(X_{n-1} = j) \leq \max(P(X_{n-2} = j), P(X_{n-2} = j - 1)) \tag{4.23}$$

$$P(X_{n-1} = j - 1) \leq \max(P(X_{n-2} = j - 1), P(X_{n-2} = j - 2)) \tag{4.24}$$

From equations 4.22, 4.23 and 4.24 we get:

$$\begin{aligned} P(X_n = j) &\leq \max(P(X_{n-1} = j), P(X_{n-1} = j - 1)) \\ &\leq \; \max((X_{n-2} = j), P(X_{n-2} = j - 1), P(X_{n-2} = j - 2)) \end{aligned}$$

By induction, we have:

$$
\begin{aligned}
P(X_n = j) &\leq \max(P(X_{n-1} = j), P(X_{n-1} = j-1)) \\
&\leq \ \max((X_{n-2} = j), P(X_{n-2} = j-1), P(X_{n-2} = j-2)) \leq \cdots \\
&\leq \max\big(P(X_1 = 0), P(X_1 = 1)\big)
\end{aligned}
$$

Hence:

$$P(X_n = j) \leq \max(r, 1-r) \tag{4.25}$$

Lemma 3:

If $P(X_n > j)$ is the probability of getting more than j successes in n Bernoulli trials, when the probability of a success in a Bernoulli trial is r, then:

$$P(X_{n+1} > j) = \ P(X_n > j) + rP(X_n = j)$$

Proof:

Conditioning on the outcome of first trial, we have:

$$
\begin{aligned}
P(X_{n+1} > j) = P(X_{n+1} > j | X_1 = success)r \\
+ P(X_{n+1} > j | X_1 = Failure)(1-r)
\end{aligned} \tag{4.26}
$$

$$= P(X_n \geq j)r + P(X_n > j)(1-r)$$

Simplifying, we get:

$$P(X_{n+1} > j) = \ P(X_n > j) + rP(X_n = j) \tag{4.27}$$

This completes the proof of Lemma 3

Lemma 4:

If $P(X_n > j)$ is the probability of getting more than j successes in n Bernoulli trials, when the probability of a success in a Bernoulli trial is r, then:

$$P(X_{n+1} > j) = P(X_n > j-1) - (1-r)P(X_n = j)$$

Proof:

From Lemma 3, we have:

$$P(X_{n+1} > j) = P(X_n > j) + rP(X_n = j) \tag{4.28}$$

We know that:

$$P(X_n > j) = P(X_n > j-1) - P(X_n = j) \tag{4.29}$$

Substituting equation 4.29 in equation 4.28, we get:

$$P(X_{n+1} > j) = P(X_n > j-1) - P(X_n = j) + rP(X_n = j)$$

$$P(X_{n+1} > j) = P(X_n > j-1) - (1-r)P(X_n = j)$$

This completes the proof of Lemma 4

Other Lemmas:

For the Binomial distribution where:

n: Number of Bernoulli trials

X_n: Random variable representing the number of successes in ‘n’ Bernoulli trials

r: The probability of success in a Bernoulli trial

We have the following relation:

Lemma 5: $E[min(X_n, nr)] - E[min(X_{n+1}, (n+1)r)] \geq -r \; \forall \; n > 0$

Lemma 6: $E[min(X_n, nr)] - E[min(X_{n+1}, (n+1)r)] = -r^2$ when $n = 0$

Lemma 7: $E[(X_n - nr)^+] - E[(X_{n+1} - (n+1)r)^+] \geq -r(1-r).\max(r, 1-r)$

$\forall \; n > 0$

Lemma 8: $E[(X_n - nr)^+] - E[(X_{n+1} - (n+1)r)^+] = -r(1-r)$ for $n = 0$

Proofs:

Proof Lemma 5:

$$E[min(X_n, nr)] - E[min(X_{n+1}, (n+1)r)]$$

Note that:

$$E[min(X_n, nr)] = \sum_{j=0}^{\lfloor nr \rfloor} j\, P(X_n = j) + nr\, P(X_n > \lfloor nr \rfloor) \tag{4.30}$$

Where $\lfloor nr \rfloor$is the floor of nr.

Similarly we can write for $E[min(X_{n+1}, (n+1)r)]$:

$$E[min(X_{n+1}, (n+1)r)] = \sum_{j=0}^{\lfloor (n+1)r \rfloor} j\, P(X_{n+1} = j) + (n+1)r\, P(X_{n+1} > \lfloor (n+1)r \rfloor) \tag{4.31}$$

Conditioning on the outcome of the first Bernoulli trial, $P(X_{n+1} = j)$ can be expressed as:

$$P(X_{n+1} = j) = P(X_n = j)(1 - r) + P(X_n = j - 1)r \tag{4.32}$$

Substituting 4.32 in 4.31, we get:

$$E[min(X_{n+1}, (n+1)r)] = \sum_{j=0}^{\lfloor (n+1)r \rfloor} j \left(P(X_n = j)(1 - r) + P(X_n = j - 1)r\right) + (n+1)r\, P(X_{n+1} > \lfloor (n+1)r \rfloor) \tag{4.33}$$

To express $P(X_{n+1} > \lfloor (n+1)r \rfloor)$ in equation 4.25 in terms of $P(X_n > \lfloor nr \rfloor)$, we consider the two cases: $\lfloor nr \rfloor = \lfloor (n+1)r \rfloor$ and $\lfloor nr \rfloor \neq \lfloor (n+1)r \rfloor$.

Case 1: $\lfloor nr \rfloor = \lfloor (n+1)r \rfloor$

Using Lemma 3 for $P(X_{n+1} > \lfloor (n+1)r \rfloor)$, we have:

$$P(X_{n+1} > \lfloor (n+1)r \rfloor) = P(X_n > \lfloor nr \rfloor) + P(X_n = \lfloor nr \rfloor)r \tag{4.34}$$

Substituting 4.34 in 4.33 gives:

$$E[min(X_{n+1}, (n+1)r)] = \sum_{j=0}^{\lfloor (n+1)r \rfloor} j \left(P(X_n = j)(1 - r) + P(X_n = j - 1)r\right) + (n+1)r \left(P(X_n > \lfloor nr \rfloor) + P(X_n = \lfloor nr \rfloor)r\right) \tag{4.35}$$

Now, deducting 4.35 from 4.30:

$$E[min(X_n, nr)] - E[min(X_{n+1}, (n+1)r)]$$

$$= \left(\sum_{j=0}^{\lfloor nr \rfloor} j\, P(X_n = j) + nr\, P(X_n > \lfloor nr \rfloor) \right)$$

$$- \left(\sum_{j=0}^{\lfloor nr \rfloor} j\, \big(P(X_n = j)(1-r) + P(X_n = j-1)r\big) \right.$$

$$\left. + (n+1)r\, \big(P(X_n > \lfloor nr \rfloor) + P(X_n = \lfloor nr \rfloor)r\big) \right)$$

Simplifying,

$$E[min(X_n, nr)] - E[min(X_{n+1}, (n+1)r)]$$

$$= \sum_{j=0}^{\lfloor nr \rfloor} rj\, P(X_n = j) - \sum_{j=0}^{\lfloor nr \rfloor} rj\, P(X_n = j-1) \qquad (4.36)$$

$$- rP(X_n > \lfloor nr \rfloor) - (n+1)r^2 P(X_n = \lfloor nr \rfloor)$$

Combining the first two terms in equation 4.36, we get:

$$E[min(X_n, nr)] - E[min(X_{n+1}, (n+1)r)]$$

$$= r \sum_{j=0}^{\lfloor nr \rfloor - 1} \big(j - (j+1)\big)\, P(X_n = j) + \lfloor nr \rfloor r P(X_n = \lfloor nr \rfloor)$$

$$- rP(X_n > \lfloor nr \rfloor) - (n+1)r^2 P(X_n = \lfloor nr \rfloor)$$

$$= -rP(X_n < \lfloor nr \rfloor) + \lfloor nr \rfloor r P(X_n = \lfloor nr \rfloor) - rP(X_n > \lfloor nr \rfloor) - (n+1)r^2 P(X_n = \lfloor nr \rfloor)$$

$$= -r\big(1 - P(X_n = \lfloor nr \rfloor)\big) + \lfloor nr \rfloor r P(X_n = \lfloor nr \rfloor) - (n+1)r^2 P(X_n = \lfloor nr \rfloor)$$

$$= -r + rP(X_n = \lfloor nr \rfloor)(\lfloor nr \rfloor - (n+1)r + 1) \qquad (4.37)$$

Since we are studying the case $\lfloor nr \rfloor = \lfloor (n+1)r \rfloor$, the last equality follows

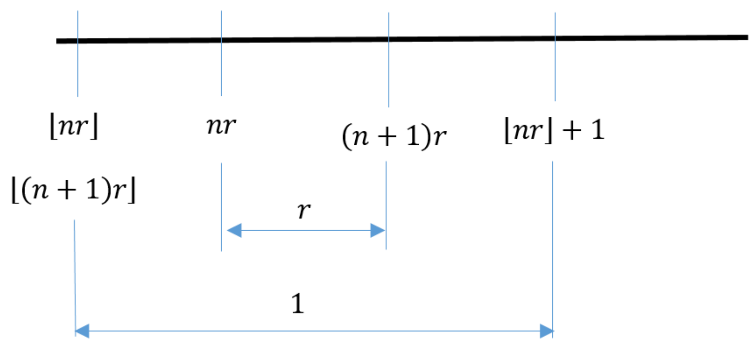

Figure 1: Lemma 5, case 1

Figure 1 shows that that the value of $(\lfloor nr \rfloor - (n+1)r + 1)$ in equation 4.37 is less than or equal to $(1 - r)$ and greater than or equal to 0.

Thus, we get:

$$E[min(X_n, nr)] - E[min(X_{n+1}, (n+1)r)]$$
$$= -r + rP(X_n = \lfloor nr \rfloor)(\lfloor nr \rfloor - (n+1)r + 1) \geq -r$$

This completes case 1.

Case 2: $\lfloor nr \rfloor + 1 = \lfloor (n+1)r \rfloor$

Using Lemma 4 for $P(X_{n+1} > \lfloor (n+1)r \rfloor)$, we have:

$$P(X_{n+1} > \lfloor (n+1)r \rfloor) = P(X_n > \lfloor nr \rfloor) - P(X_n = \lfloor nr \rfloor + 1)(1 - r) \qquad (4.38)$$

Substituting 4.38 in 4.33, gives:

$$E[min(X_{n+1}, (n+1)r)]$$

$$= \sum_{j=0}^{\lfloor nr \rfloor + 1} j\,(P(X_n = j)(1-r) + P(X_n = j-1)r) \tag{4.39}$$

$$+ (n+1)r\left(P(X_n > \lfloor nr \rfloor) - P(X_n = \lfloor nr \rfloor + 1)(1-r)\right)$$

Now, deducting 4.39 from 4.30:

$$= \left(\sum_{j=0}^{\lfloor nr \rfloor} j\,P(X_n = j) + nr\,P(X_n > \lfloor nr \rfloor)\right)$$

$$- \left(\sum_{j=0}^{\lfloor nr \rfloor + 1} j\,(P(X_n = j)(1-r) + P(X_n = j-1)r)\right.$$

$$\left. + (n+1)r\left(P(X_n > \lfloor nr \rfloor) - P(X_n = \lfloor nr \rfloor + 1)(1-r)\right)\right)$$

Simplifying:

$$= \sum_{j=0}^{\lfloor nr \rfloor + 1} rj\,P(X_n = j) - \sum_{j=0}^{\lfloor nr \rfloor + 1} rj\,P(X_n = j-1) - (\lfloor nr \rfloor + 1)P(X_n = \lfloor nr \rfloor + 1)$$

$$- rP(X_n > \lfloor nr \rfloor) + (n+1)r(1-r)P(X_n = \lfloor nr \rfloor + 1)$$

Combining the first two terms $\sum_{j=0}^{\lfloor nr \rfloor + 1} rj\,P(X_n = j) - \sum_{j=0}^{\lfloor nr \rfloor + 1} rj\,P(X_n = j-1) =$

$\sum_{j=0}^{\lfloor nr \rfloor}\left(rj - r(j+1)\right) P(X_n = j) + (\lfloor nr \rfloor + 1)rP(X_n = \lfloor nr \rfloor + 1)$, we have

$$= \sum_{j=0}^{\lfloor nr \rfloor} \big(rj - r(j+1)\big)\, P(X_n = j) + (\lfloor nr \rfloor + 1) r P(X_n = \lfloor nr \rfloor + 1)$$

$$- (\lfloor nr \rfloor + 1) P(X_n = \lfloor nr \rfloor + 1) - r P(X_n > \lfloor nr \rfloor)$$

$$+ (n+1) r (1-r) P(X_n = \lfloor nr \rfloor + 1)$$

$$= -rP(X_n \leq \lfloor nr \rfloor) + (\lfloor nr \rfloor + 1) r P(X_n = \lfloor nr \rfloor + 1) - (\lfloor nr \rfloor + 1) P(X_n = \lfloor nr \rfloor + 1)$$

$$- rP(X_n > \lfloor nr \rfloor) + (n+1) r (1-r) P(X_n = \lfloor nr \rfloor + 1)$$

$$= -r - (\lfloor nr \rfloor + 1)(1-r) r P(X_n = \lfloor nr \rfloor + 1) + (n+1) r (1-r) P(X_n = \lfloor nr \rfloor + 1)$$

Simplifying:

$$= -r + (1-r)\big((n+1)r - \lfloor nr \rfloor - 1\big) P(X_n = \lfloor nr \rfloor + 1) \geq -r$$

This completes case 2, hence:

$E[min(X_n, nr)] - E[min(X_{n+1}, (n+1)r)] \ \geq -r \ \forall \ n > 0$

Proof Lemma 6:

$$E[min(X_n, nr)] - E[min(X_{n+1}, (n+1)r)]$$

When n=0, $E[min(X_n, nr)] = 0$ & $E[min(X_{n+1}, (n+1)r)] = r^2$

This gives

$$E[min(X_n, nr)] - E[min(X_{n+1}, (n+1)r)] = \ -r^2$$

This completes the proof for lemma 6.

Proof Lemma 7:

$$\left[E[(X_n - nr)^+] - E[(X_{n+1} - (n+1)r)^+]\right]$$

Note that:

$$E[(X_n - nr)^+] = \sum_{j=\lceil nr \rceil}^{n} (j - nr)\, P(X_n = j) \tag{4.40}$$

Where $\lceil nr \rceil$ is the ceiling of nr.

Similarly, for $E[(X_{n+1} - (n+1)r)^+]$:

$$E[(X_{n+1} - (n+1)r)^+] = \sum_{j=\lceil (n+1)r \rceil}^{n+1} (j - (n+1)r)\, P(X_{n+1} = j) \tag{4.41}$$

Conditioning on the outcome of the first Bernoulli trial out of $n + 1$ trials, $P(X_{n+1} = j)$ can be expressed as:

$$P(X_{n+1} = j) = P(X_n = j)(1 - r) + P(X_n = j - 1)r$$

Substituting equation 4.40 in equation 4.41 gives:

$$E[(X_{n+1} - (n+1)r)^+]$$
$$= \sum_{j=\lceil (n+1)r \rceil}^{n+1} (j - (n+1)r)\, (P(X_n = j)(1 - r) \tag{4.42}$$
$$+ P(X_n = j - 1)r)$$

Subtracting 4.42 from 4.40:

$$E[(X_n - nr)^+] - E[(X_{n+1} - (n+1)r)^+]$$

$$= \sum_{j=\lceil nr \rceil}^{n} (j - nr)\, P(X_n = j)$$

$$- \sum_{j=\lceil (n+1)r \rceil}^{n+1} (j - (n+1)r)\, \big(P(X_n = j)(1-r) + P(X_n = j-1)r\big)$$

Taking summation inside, we get:

$$\begin{aligned} E[(X_n - nr)^+] - E[(X_{n+1} - (n+1)r)^+] \\ &= \sum_{j=\lceil nr \rceil}^{n} (j - nr)\, P(X_n = j) \\ &- \sum_{j=\lceil (n+1)r \rceil}^{n+1} (j - (n+1)r)\, P(X_n = j) \\ &+ \sum_{j=\lceil (n+1)r \rceil}^{n+1} r(j - (n+1)r)\, P(X_n = j) \\ &- \sum_{j=\lceil (n+1)r \rceil}^{n+1} (j - (n+1)r)\, rP(X_n = j-1) \end{aligned} \tag{4.43}$$

To prove Lemma 7, we again consider the two possible cases, $\lceil nr \rceil = \lceil (n+1)r \rceil$ and $\lceil nr \rceil \neq \lceil (n+1)r \rceil$.

Case 1: $\lceil nr \rceil = \lceil (n+1)r \rceil$

Equation 4.43 in this case simplifies into:

$$E[(X_n - nr)^+] - E[(X_{n+1} - (n+1)r)^+]$$

$$= \sum_{j=\lceil nr \rceil}^{n} (j - nr)\, P(X_n = j)$$

$$- \sum_{j=\lceil nr \rceil}^{n+1} (j - (n+1)r)\, P(X_n = j) + \sum_{j=\lceil nr \rceil}^{n+1} r(j - (n+1)r)\, P(X_n = j)$$

$$- \sum_{j=\lceil nr \rceil}^{n+1} (j - (n+1)r)\, rP(X_n = j - 1)$$

Combining the first two and the last two terms, we get:

$$E[(X_n - nr)^+] - E[(X_{n+1} - (n+1)r)^+]$$

$$= \sum_{j=\lceil nr \rceil}^{n} (j - nr - j + (n+1)r)\, P(X_n = j)$$

$$+ \sum_{j=\lceil nr \rceil}^{n} r(j - j - 1)\, P(X_n = j) \; - r(\lceil nr \rceil - (n+1)r) P(X_n = \lceil nr \rceil - 1)$$

$$= -r(\lceil nr \rceil - (n+1)r) P(X_n = \lceil nr \rceil - 1) \tag{4.44}$$

The last equality follows that, in the case $\lceil nr \rceil = \lceil (n+1)r \rceil$

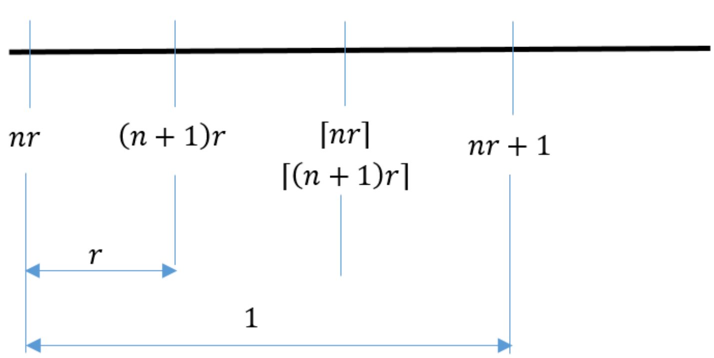

Figure 2: Lemma 7, case 1

Figure 2 shows that the maximum value $(\lceil nr \rceil - (n+1)r)$ can take is less than $(1-r)$, Hence:

$$E[(X_n - nr)^+] - E[(X_{n+1} - (n+1)r)^+] \geq -r(1-r)P(X_n = \lceil nr \rceil - 1)$$

Using Lemma 2, $P(X_n = j) \leq \max(r, 1-r)$

Hence, for case 1:

$$E[(X_n - nr)^+] - E[(X_{n+1} - (n+1)r)^+] \geq -r(1-r).\max(r, 1-r)$$

Case 2: $\lceil nr \rceil + 1 = \lceil (n+1)r \rceil$

Equation 4.43 in this case simplifies t:

$$\begin{aligned} &E[(X_n - nr)^+] - E[(X_{n+1} - (n+1)r)^+] \\ &= \sum_{j=\lceil nr \rceil}^{n} (j - nr)\, P(X_n = j) \\ &- \sum_{j=\lceil nr \rceil+1}^{n+1} (j - (n+1)r)\, P(X_n = j) \\ &+ \sum_{j=\lceil nr \rceil+1}^{n+1} r(j - (n+1)r)\, P(X_n = j) \\ &- \sum_{j=\lceil nr \rceil+1}^{n+1} (j - (n+1)r)\, rP(X_n = j-1) \end{aligned}$$

Combining the first and last two terms gives:

$$= (\lceil nr \rceil - nr)P(X_n = \lceil nr \rceil) - \sum_{j=\lceil nr \rceil+1}^{n} (j - nr - j + (n+1)r)\, P(X_n = j)$$

$$+ \sum_{j=\lceil nr \rceil+1}^{n} r(j - j - 1)\, P(X_n = j)$$

$$- r(\lceil nr \rceil + 1 - (n+1)r)P(X_n = \lceil nr \rceil)$$

The summation terms cancel out:

$$E[(X_n - nr)^+] - E[(X_{n+1} - (n+1)r)^+] \\ = \big(\lceil nr \rceil - nr - r(\lceil nr \rceil + 1 - (n+1)r)\big)P(X_n = \lceil nr \rceil) \tag{4.45}$$

The last equality follows since in the case : $\lceil nr \rceil + 1 = \lceil (n+1)r \rceil$

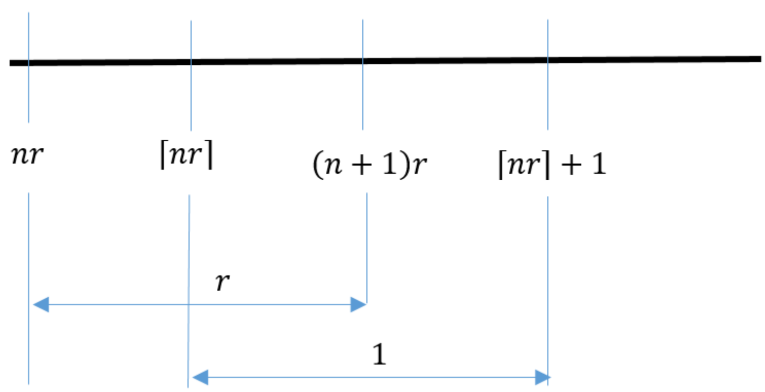

Figure 3: Lemma 7, case 2

Figure 3 shows that $\lceil nr \rceil \leq (n+1)r$

Representing $\lceil nr \rceil = nr + yr$, this implies $y \leq 1$

Substituting $\lceil nr \rceil = nr + yr$ in equation 4.45 gives:

$$E[(X_n - nr)^+] - E[(X_{n+1} - (n+1)r)^+]$$

$$= \big(\lceil nr \rceil - nr - r(\lceil nr \rceil + 1 - (n+1)r)\big)P(X_n = \lceil nr \rceil)$$

$$= \big(nr + yr - nr - r(nr + yr + 1 - (n+1)r)\big)P(X_n = \lceil nr \rceil)$$

Simplifying:

$$= \big(yr - r(yr + 1 - r)\big)P(X_n = \lceil nr \rceil)$$

Expanding:

$$= (yr - yr^2 + r - r^2)P(X_n = \lceil nr \rceil)$$

$$= -r(1-y)(1-r)P(X_n = \lceil nr \rceil)$$

The minimum is achieved at $y = 0$, and we have:

$$E[(X_n - nr)^+] - E[(X_{n+1} - (n+1)r)^+] \geq -r(1-r)P(X_n = \lceil nr \rceil)$$

Using Lemma 2, $P(X_n = j) \leq \max(r, 1-r)$

We have again for case 2

$$E[(X_n - nr)^+] - E[(X_{n+1} - (n+1)r)^+] \geq -r(1-r).\max(r, 1-r)$$

Proof Lemma 8:

When $n = 0$. $E[(X_n - nr)^+] = 0$ and $E[(X_{n+1} - (n+1)r)^+] = r(1-r)$

This gives

$E[(X_n - nr)^+] - E[(X_{n+1} - (n+1)r)^+] = -r(1-r)$ for n =0

CHAPTER 5

AN MDP MODEL FOR LOGISTICS CAPACITY PLANNING IN SUPPLY CHAIN

5.1 Introduction

The purpose of this chapter is to provide a general MDP model to optimize logistics capacity for forward logistics. The existence of structural properties for the optimal decision policy of the model is studied and presented. The advantages of these structural properties in terms of number of iterations/ computational effort in reaching the optimal policy is quantified and presented.

5.2 Problem Definition

In this section, the problem definition is presented. A forward logistics problem is considered; where a retailer caters to the demands of the customers in an area. The demand is taken as stochastic and Markovian. That is the probability distribution of the demand in the next epoch is defined by the demand realized in the current epoch. The logistics capacity of the retailer defines the maximum number of demands it may process internally. Demands in excess of the retailer's capacity incur a penalty for emergency processing. At every epoch, there will be a maintenance cost for maintaining the system capacity and a processing cost for the utilized system capacity. At the end of each time epoch, the retailer makes a decision on how much to increase or decrease its capacity. There is no backordering of demand, if the demand is not met by the system's Forward Logistics (FL) capacity, it is processed externally incurring a penalty cost. The decision policy states the optimum action at each time epoch for each system state, with the objective of minimizing system cost over the planning horizon.

5.3 Nomenclature

The notations used in this chapter are as follows:

CS_t: Retailer logistics capacity at time t

d_t: Demand at time t

MC: Maximum capacity possible for the retailer's FL

T: Planning horizon

t: Time epoch $t = 0,1,2 \ldots . T$

5.4 Model Definitions

An MDP model consists of system states, actions, transition probabilities and reward/ cost functions (Puterman 2009). Next, we define them for this problem:

System State: At any time epoch (t) of the planning horizon (T), the system state is described by two state variables - the firm's FL capacity, and the demand experienced for the period (CS_t, d_t). The states are taken as partially ordered with respect to CS_t in the MDP model. At the beginning of the planning horizon, $t = 0$, the system state is assumed to be (CS_0, d_0).

Action: (a_t) The decision maker decides on how much to increase or decrease the system capacity.

$a_t(CS_t, d_t)$ or simply $a_t(CS, d)$: Is the action at the end of time epoch 't' when the system is in state (CS_t, d_t) .

a_t=0 : No change in system's FL capacity. $CS_{t+1} = CS_t$

a_t=+ve : Increase in system's FL capacity. $CS_{t+1} = CS_t + a_t$

a_t=-ve : Decrease in system's FL capacity. $CS_{t+1} = CS_t + a_t$

The action space $A_t(CS_t, d_t)$ is the set of decisions available for the decision maker when the system is in state (CS_t, d_t)

$$A_t(CS_t, d_t) = \{-CS_t, -CS_t + 1, CS_t + 2 \dots MC - CS_t\}$$

This implies that the minimum capacity the system can have is 0 and the maximum capacity is MC.

Transition Probabilities:

This defines the state transition probabilities, the probability of a future system state (in the next decision epoch); given the current system state and the action taken.

$$p_{t+1}[(CS_{t+1}, d_{t+1})|(CS_t, d_t), a_t] = \begin{cases} P(d_{t+1}|d_t) & if\ CS_{t+1} = CS_t + a_t \\ 0 & otherwise \end{cases} \quad (5.1)$$

Given state $(\boldsymbol{CS_t}, \boldsymbol{d_t})$ and action $\boldsymbol{a_t}$, the probability of the state $(\boldsymbol{CS_{t+1}}, \boldsymbol{d_{t+1}})$ in the next period is equal to the demand transition probability if $\boldsymbol{P(d_{t+1}|d_t)}$ if $\boldsymbol{CS_{t+1} = CS_t + a_t}$ and 0 otherwise.

Rewards and model dynamics:

Cost components:

c_1: Cost of increasing the firm's FL capacity by one unit – unit investment cost: ($/ capacity unit)

c_2: Cost of decreasing the firm's FL capacity by one unit – unit disinvestment cost: ($/ capacity unit)

c_3: Fixed internal capacity maintenance cost. Maintenance Cost per unit of FL capacity for one period ($/ capacity unit/ period)

c_4: Labour cost. Processing cost per item for processing FL internally ($/ unit)

c_5: Shortage cost. Penalty or emergency cost incurred per unit of the demand exceeding FL capacity. ($/ unit)

c_6: Capacity salvage value. Revenue in salvaging unit FL capacity of the system at the end of the planning horizon ($/ unit)

All the structural properties derived in this chapter are mainly based on the following assumptions:

$$c_1 \geq max(0, -c_2)$$

$$c_5 > c_4 + c_3$$

The first assumption means that if revenue is generated in decreasing capacity, it shall be less than the cost in increasing capacity. The second assumption means that the cost of penalty per unit is more than the sum of the maintenance and labor cost per unit.

At the end of each period, the decision maker decides on how much to increase or decrease the system capacity. The reward in the period is dependent on the system state and this decision made.

The reward function is represented as $R_t[(CS_t, d_t), a_t]$.

$R_t[(k_t, w_t), a_t]$ is the immediate cost experienced in period, t after choosing action a_t in the period when the system is in state (CS_t, d_t).

The reward function for decision a_t is given as

$$\begin{aligned} R_t[(CS_t, d_t), a_t] \\ &= c_1(a_t)^+ + c_2(-a_t)^+ + \ c_3(CS_t) + c_4 min(d_t, CS_t) \\ &+ \ c_5(d_t - \ CS_t)^+ \end{aligned} \tag{5.2}$$

Here the first term, c_1, accounts for costs incurred for increase in capacity (investment), the second term, c_2, costs for any decrease in capacity (disinvestment), the third term, c_3, costs for maintenance of the system capacity, the fourth term, c_4, for the incurred cost in processing the FL in the period t and the fifth term, c_5, for the incurred penalty in case demand exceeds system capacity in the period t as demands are not carried to next period.

Terminal reward is taken as

$$R_T[(CS_T, d_T), a_t] = V_T^*(CS_T, d_T) = -c_6 CS_T \tag{5.3}$$

The terminal reward accounts for the revenue in salvaging system capacity at the end of the planning horizon.

The decision maker is to determine the decision policy, which will minimize the total expected cost over the planning horizon. The optimal value function that gives the expected cost from period t till the end of the planning horizon when following optimum decision policy is given as $V_t^*(CS_t, d_t)$

$$V_t^*(CS_t, d_t) = \begin{matrix} min \\ a \in A \end{matrix} \left\{ R_t[(CS_t, d_t), a_t] \right. \\ \left. + \sum_{d_{t+1}} p_{t+1}[(CS_{t+1}, d_{t+1}) | (CS_t, d_t), a_t] V_{t+1}^*(CS_{t+1}, d_{t+1}) \right\} \tag{5.4}$$

This can be rewritten using equation 5.1 as

$$V_t^*(CS_t, d_t) = \begin{matrix} min \\ a \in A \end{matrix} \left\{ R_t[(CS_t, d_t), a_t] \right. \\ \left. + \sum_{d_{t+1}} P(d_{t+1} | d_t)\, V_{t+1}^*(CS_t + a_t, d_{t+1}) \right\} \tag{5.5}$$

Hence the optimal action $a_t(CS_t, d_t)$

$$a_t(CS_t, d_t) = \begin{matrix} arg\min a \\ a \in A \end{matrix} \left\{ R_t[(CS_t, d_t), a_t] + \sum_{d_{t+1}} P(d_{t+1}|d_t)\, V_{t+1}^*(CS_t + a_t, d_{t+1}) \right\}$$

The value function that gives the expected cost from period t, to the end of the planning horizon, when taking action a_t at the end of period t; and henceforth following optimal decision policy, is given as $V_t\big((CS_t, d_t), a_t\big)$:

$$V_t\big((CS_t, d_t), a_t\big) = R_t[(CS_t, d_t), a_t] + \sum_{d_{t+1}} P(d_{t+1}|d_t)\, V_{t+1}^*(CS_t + a_t, d_{t+1}) \quad (5.6)$$

Solution Methodology

Optimum decision policy for a finite period MDP model can be determined using backward value iteration.

For this problem definition if the demand takes values from $[0, x]$. The number of iterations required is

$$T \times (x + 1)^3$$

As for each epoch, there are $(x + 1)^2$ states possible and for each state there are $(x + 1)$ possible actions.

5.5 Existence of structured optimal decision policy:

Existence of a structured decision policy helps reduce the computational effort in determining in the optimum decision policy and makes the policy practical and easier to follow. Next, the different structural properties of the optimal policy is provided.

Theorem 1:

In a particular time epoch t, if the optimum action for a state $(CS_t = CS1, d_t)$ is to increase the capacity to g $(g > CS1)$, then for all states $(CS_t = CS2, d_t)$ such that $CS1 < CS2 \leq g$, increasing the system capacity to reach the same capacity level g will be an optimal action.

That is if $a_t(CS_t = CS1, d_t) = g - CS1$ and $g > CS1$

Then $a_t(CS_t = CS2, d_t) = g - CS2 \;\; \forall \, CS1 \leq CS2 \leq g$

Proof

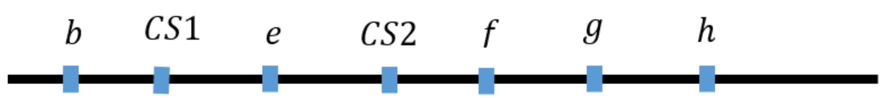

Figure 4: Illustration of theorem 1

The above figure graphically represents the relation between a set of capacities (b, $CS1, e, CS2, f, g, h$)

$$b < CS1 < e < CS2 < f < g < h$$

Given $a_t(CS_t = CS1, d_t) = x^* = g - CS1$ and $g > CS1$

Let x^{--} be any action such that $CS1 + x^{--} = b$; $b < CS1$

Let x^{-} be any action such that $CS1 + x^{-} = e$; $CS1 \leq e < CS2$

Let x^{+} be any action such that $CS1 + x^{+} = f$; $CS2 \leq f < g$

Let x^{++} be any action such that $CS1 + x^{++} = h;\ g < h$

Since $a_t(CS1, d) = x^*$is the optimum action at $(CS_t = CS1, d_t)$, we have the following relations

1.) $V_t^*(CS_t = CS1, d_t) = V_t\big((CS_t = CS1, d_t), x^*\big) \leq V_t((CS_t = CS1, d_t), x^{--})$

2.) $V_t^*(CS_t = CS1, d_t) = V_t\big((CS_t = CS1, d_t), x^*\big) \leq V_t((CS_t = CS1, d_t), x^{-})$

3.) $V_t^*(CS_t = CS1, d_t) = V_t\big((CS_t = CS1, d_t), x^*\big) \leq V_t((CS_t = CS1, d_t), x^{+})$

4.) $V_t^*(CS_t = CS1, d_t) = V_t\big((CS_t = CS1, d_t), x^*\big) \leq V_t((CS_t = CS1, d_t), x^{++})$

Taking one relation at a time, we get:

1.) $V_t\big((CS1, d_t), x^*\big) \leq V_t((CS1, d_t), x^{--})$

Substituting equation 5.6

$$R_t[(CS1, d_t), x^*] + \sum_{d_{t+1}} P(d_{t+1}|d_t)\, V_{t+1}^*(CS1 + x^*, d_{t+1})$$

$$\leq R_t[(CS1, d_t), x^{--}] + \sum_{d_{t+1}} P(d_{t+1}|d_t)\, V_{t+1}^*(CS1 + x^{--}, d_{t+1})$$

Substituting equation 5.2 for immediate expected reward

$$c_1 x^* + c_3 CS1 + c_4 min(d_t, CS1) + c_5(d_t - CS1)^+ + \sum_{d_{t+1}} P(d_{t+1}|d_t)\, V_{t+1}^*(g, d_{t+1})$$

$$\leq -c_2 x^{--} + c_3 CS1 + c_4 min(d_t, CS1) + c_5(d_t - CS1)^+$$

$$+ \sum_{d_{t+1}} P(d_{t+1}|d_t)\, V_{t+1}^*(b, d_{t+1})$$

Simplifying and grouping common terms we get:

$$c_1x^* + c_2x^{--} \leq \sum_{d_{t+1}} P(d_{t+1}|d_t)\big(V^*_{t+1}(b, d_{t+1}) - V^*_{t+1}(g, d_{t+1})\big)$$

$$c_1(g - CS1) + c_2(b - CS1) \qquad (5.7)$$

$$\leq \sum_{d_{t+1}} P(d_{t+1}|d_t)\big(V^*_{t+1}(b, d_{t+1}) - V^*_{t+1}(g, d_{t+1})\big)$$

2.) $V_t\big((CS1, d_t), x^*\big) \leq V_t((CS1, d_t), x^-)$

Substituting equation 5.6

$$R_t[(CS1, d_t), x^*] + \sum_{d_{t+1}} P(d_{t+1}|d_t)\, V^*_{t+1}(CS1 + x^*, d_{t+1})$$

$$\leq R_t[(CS1, d_t), x^-] + \sum_{d_{t+1}} P(d_{t+1}|d_t)\, V^*_{t+1}(CS1 + x^-, d_{t+1})$$

Substituting equation 5.2 for immediate expected reward

$$c_1x^* + c_3CS1 + c_4min(d_t, CS1) + c_5(d_t - CS1)^+ + \sum_{d_{t+1}} P(d_{t+1}|d_t)\, V^*_{t+1}(g, d_{t+1})$$

$$\leq c_1x^- + c_3CS1 + c_4min(d_t, CS1) + c_5(d - CS1)^+$$

$$+ \sum_{d_{t+1}} P(d_{t+1}|d_t)\, V^*_{t+1}(e, d_{t+1})$$

Simplifying and grouping common terms we get:

$$c_1(x^* - x^-) \leq \sum_{d_{t+1}} P(d_{t+1}|d_t)\big(V^*_{t+1}(e, d_{t+1}) - V^*_{t+1}(g, d_{t+1})\big)$$

$$c_1(g-e) \le \sum_{d_{t+1}} P(d_{t+1}|d_t)\big(V_{t+1}^*(e, d_{t+1}) - V_{t+1}^*(g, d_{t+1})\big) \tag{5.8}$$

3.) $V_t\big((CS1, d_t), x^*\big) \le V_t((CS1, d_t), x^+)$

Substituting equation 5.6

$$R_t[(CS1, d_t), x^*] + \sum_{d_{t+1}} P(d_{t+1}|d_t)\, V_{t+1}^*(CS1 + x^*, d_{t+1})$$
$$\le R_t[(CS1, d_t), x^+] + \sum_{d_{t+1}} P(d_{t+1}|d_t)\, V_{t+1}^*(CS1 + x^+, d_{t+1})$$

Substituting equation 5.2 for immediate expected reward

$$c_1x^* + c_3CS1 + c_4min(d_t, CS1) + c_5(d_t - CS1)^+ + \sum_{d_{t+1}} P(d_{t+1}|d_t)\, V_{t+1}^*(g, d_{t+1})$$
$$\le c_1x^+ + c_3CS1 + c_4min(d_t, CS1) + c_5(d_t - CS1)^+$$
$$+ \sum_{d_{t+1}} P(d_{t+1}|d_t)\, V_{t+1}^*(f, d_{t+1})$$

Simplifying and grouping common terms we get:

$$c_1(x^* - x^+) \le \sum_{d_{t+1}} P(d_{t+1}|d_t)\big(V_{t+1}^*(f, d_{t+1}) - V_{t+1}^*(g, d_{t+1})\big)$$

$$c_1(g-f) \le \sum_{d_{t+1}} P(d_{t+1}|d_t)\big(V_{t+1}^*(f, d_{t+1}) - V_{t+1}^*(g, d_{t+1})\big) \tag{5.9}$$

4.) $V_t\big((CS1, d_t), x^*\big) \le V_t((CS1, d_t), x^{++})$

Substituting equation 5.6

$$R_t[(CS1, d_t), x^*] + \sum_{d_{t+1}} P(d_{t+1}|d_t)\, V_{t+1}^*(CS1 + x^*, d_{t+1})$$

$$\leq R_t[(CS1, d_t), x^{++}] + \sum_{d_{t+1}} P(d_{t+1}|d_t)\, V_{t+1}^*(CS1 + x^{++}, d_{t+1})$$

Substituting equation 5.2 for immediate expected reward

$$c_1 x^* + c_3 CS1 + c_4 min(d_t, CS1) + c_5 (d_t - CS1)^+ + \sum_{d_{t+1}} P(d_{t+1}|d_t)\, V_{t+1}^*(g, d_{t+1})$$

$$\leq c_1 x^{++} + c_3 CS1 + c_4 min(d_t, CS1) + c_5 (d_t - CS1)^+$$

$$+ \sum_{d_{t+1}} P(d_{t+1}|d_t)\, V_{t+1}^*(h, d_{t+1})$$

Simplifying and grouping common terms we get:

$$c_1 (x^* - x^{++}) \leq \sum_{d_{t+1}} P(d_{t+1}|d_t)\big(V_{t+1}^*(h, d_{t+1}) - V_{t+1}^*(g, d_{t+1})\big)$$

$$c_1 (g - h) \leq \sum_{d_{t+1}} P(d_{t+1}|d_t)\big(V_{t+1}^*(h, d_{t+1}) - V_{t+1}^*(g, d_{t+1})\big) \tag{5.10}$$

Let y^{--} be any action such that $CS2 + y^{--} = b;\ b < CS1$

Let y^{-} be any action such that $CS2 + y^{-} = e;\ CS1 \leq e < CS2$

Let y^{+} be any action such that $CS2 + y^{+} = f; CS2 \leq f < g$

Let y^{++} be any action such that $CS2 + y^{++} = h;\ g < h$

Let y^{*} be any action such that $CS2 + y^{*} = g;$

If the optimum action was y^{--}, then:

$$V_t\big((CS2, d_t), y^*\big) > V_t((CS1, d_t), y^{--})$$

Substituting equation 5.6

$$R_t[(CS2, d_t), y^*] + \sum_{d_{t+1}} P(d_{t+1}|d_t)\, V_{t+1}^*(CS2 + y^*, d_{t+1})$$
$$> R_t[(CS2, d_t), y^{--}] + \sum_{d_{t+1}} P(d_{t+1}|d_t)\, V_{t+1}^*(CS2 + y^{--}, d_{t+1})$$

Substituting equation 5.2 for immediate expected reward

$$c_1 y^* + c_3 CS2 + c_4 min(d_t, CS2) + c_5(d_t - CS2)^+ + \sum_{d_{t+1}} P(d_{t+1}|d_t)\, V_{t+1}^*(g, d_{t+1})$$
$$> -c_2 y^{--} + c_3 CS2 + c_4 min(d_t, CS2) + c_5(d_t - CS2)^+$$
$$+ \sum_{d_{t+1}} P(d_{t+1}|d_t)\, V_{t+1}^*(b, d_{t+1})$$

Simplifying and grouping common terms we get:

$$c_1 y^* + c_2 y^{--} > \sum_{d_{t+1}} P(d_{t+1}|d_t)\big(V_{t+1}^*(b, d_{t+1}) - V_{t+1}^*(g, d_{t+1})\big)$$

$$c_1(g - CS2) + c_2(b - CS2) \tag{5.11}$$
$$> \sum_{d_{t+1}} P(d_{t+1}|d_t)\big(V_{t+1}^*(b, d_{t+1}) - V_{t+1}^*(g, d_{t+1})\big)$$

We can see that the RHS of equation 5.11 and equation 5.7 are the same and the LHS of equation 5.11 is less than or equal to the LHS of equation 5.7. Thus, we have a contradiction and hence y^{--} cannot be optimum decision.

If the optimum action was y^-, then:

$$V_t\big((CS2, d_t), y^*\big) > V_t((CS1, d_t), y^-)$$

Substituting equation 5.6

$$R_t[(CS2, d_t), y^*] + \sum_{d_{t+1}} P(d_{t+1}|d_t)\, V_{t+1}^*(CS2 + y^*, d_{t+1})$$

$$> R_t[(CS2, d_t), y^-] + \sum_{d_{t+1}} P(d_{t+1}|d_t)\, V_{t+1}^*(CS2 + y^-, d_{t+1})$$

Substituting equation 5.2 for immediate expected reward

$$c_1 y^* + c_3 CS2 + c_4 min(d_t, CS2) + c_5(d_t - CS2)^+ + \sum_{d_{t+1}} P(d_{t+1}|d_t)\, V_{t+1}^*(g, d_{t+1})$$

$$> -c_2 y^- + c_3 CS2 + c_4 min(d_t, CS2) + c_5(d_t - CS2)^+$$

$$+ \sum_{d_{t+1}} P(d_{t+1}|d_t)\, V_{t+1}^*(e, d_{t+1})$$

Simplifying and grouping common terms we get:

$$c_1 y^* + c_2 y^- > \sum_{d_{t+1}} P(d_{t+1}|d_t)\big(V_{t+1}^*(e, d_{t+1}) - V_{t+1}^*(g, d_{t+1})\big)$$

$$c_1(g - CS2) + c_2(e - CS2) \qquad (5.12)$$

$$> \sum_{d_{t+1}} P(d_{t+1}|d_t)\big(V_{t+1}^*(e, d_{t+1}) - V_{t+1}^*(g, d_{t+1})\big)$$

Again, we see that the RHS of equation 5.12 and equation 5.8 are the same and the LHS of equation 5.12 is less than or equal to the LHS of equation 5.8. Thus, we have a contradiction and hence y^- cannot be optimum decision.

If the optimum action was y^+, then:

$$V_t\big((CS2, d), y^*\big) > V_t((CS1, d), y^+)$$

Substituting equation 5.6

$$R_t[(CS2, d_t), y^*] + \sum_{d_{t+1}} P(d_{t+1}|d_t)\, V^*_{t+1}(CS2 + y^*, d_{t+1})$$

$$> R_t[(CS2, d_t), y^+] + \sum_{d_{t+1}} P(d_{t+1}|d_t)\, V^*_{t+1}(CS2 + y^+, d_{t+1})$$

Substituting equation 5.2 for immediate expected reward

$$c_1y^* + c_3CS2 + c_4min(d_t, CS2) + c_5(d_t - CS2)^+ + \sum_{d_{t+1}} P(d_{t+1}|d_t)\, V^*_{t+1}(g, d_{t+1})$$

$$> c_1y^+ + c_3CS2 + c_4min(d_t, CS2) + c_5(d_t - CS2)^+$$

$$+ \sum_{d_{t+1}} P(d_{t+1}|d_t)\, V^*_{t+1}(f, d_{t+1})$$

Simplifying and grouping common terms we get:

$$c_1y^* - c_1y^+ > \sum_{d_{t+1}} P(d_{t+1}|d_t)\big(V^*_{t+1}(f, d_{t+1}) - V^*_{t+1}(g, d_{t+1})\big)$$

$$c_1(g - f) > \sum_{d_{t+1}} P(d_{t+1}|d_t)\big(V^*_{t+1}(f, d_{t+1}) - V^*_{t+1}(g, d_{t+1})\big) \qquad (5.13)$$

We can see that equation 5.13 contradicts equation 5.9 hence y^+ cannot be optimum decision.

If the optimum action was y^{++}, then:

$$V_t\big((CS2, d_t), y^*\big) > V_t((CS1, d_t), y^{++})$$

Substituting equation 5.6

$$R_t[(CS2, d_t), y^*] + \sum_{d_{t+1}} P(d_{t+1}|d_t)\, V_{t+1}^*(CS2 + y^*, d_{t+1})$$
$$> R_t[(CS2, d_t), y^{++}] + \sum_{d_{t+1}} P(d_{t+1}|d_t)\, V_{t+1}^*(CS2 + y^{++}, d_{t+1})$$

Substituting equation 5.2 for immediate expected reward

$$c_1 y^* + c_3 CS2 + c_4 min(d_t, CS2) + c_5(d_t - CS2)^+ + \sum_{d_{t+1}} P(d_{t+1}|d_t)\, V_{t+1}^*(g, d_{t+1})$$
$$> c_1 y^{++} + c_3 CS2 + c_4 min(d_t, CS2) + c_5(d_t - CS2)^+$$
$$+ \sum_{d_{t+1}} P(d_{t+1}|d_t)\, V_{t+1}^*(h, d_{t+1})$$

Simplifying and grouping common terms we get:

$$c_1 y^* - c_1 y^{++} > \sum_{d_{t+1}} P(d_{t+1}|d_t)\big(V_{t+1}^*(h, d_{t+1}) - V_{t+1}^*(g, d_{t+1})\big)$$

$$c_1(g - h) > \sum_{d_{t+1}} P(d_{t+1}|d_t)\big(V_{t+1}^*(h, d_{t+1}) - V_{t+1}^*(g, d_{t+1})\big) \tag{5.14}$$

We can see that equation 5.14 contradicts equation 5.10 hence y^{++} cannot be optimum decision.

Hence we have the optimum action at $CS2$ to be y^*. This completes the proof for theorem 1.

Theorem 2:

In a particular time epoch t, if the optimum action for a state $(CS_t = CS2, d_t)$ is to decrease the system capacity to e, then for all states $(CS_t = CS1, d_t)$ such that $e \leq CS1 < CS2$, reaching the same capacity level e will be an optimal action.

That is if $a_t(CCS_t = CS2, d_t) = e - CS2$ and $e < CS2$

Then, $a_t(CS_t = CS1, d_t) = e - CS1 \;\; \forall\, e \leq CS1 < CS2$

Proof

The proof to theorem 2 is on the same lines as for theorem 1.

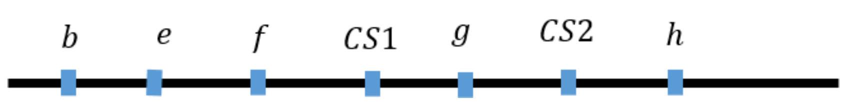

Figure 5: Illustration of theorem 2

The above figure graphically represents the relation between a set of capacities $(b, e, f, CS1, g, CS2,\; h)$

$$b < e < f < CS1 < g < CS2 < h$$

Given $a_t(CS2, d) = y^* = e - CS2$ and $e < CS2$

Let y^{--} be any action such that $CS2 + y^{--} = b;\; b < e$

Let y^{-} be any action such that $CS2 + y^{-} = f;\; e < f \leq CS1$

Let y^{+} be any action such that $CS2 + y^{+} = g; CS1 \leq g \leq CS2$

Let y^{++} be any action such that $CS2 + y^{++} = h;\; CS2 \leq h$

Since $a_t(CS2, d) = y^*$is the optimum action at $(CS2, d)$, we have the following relations

1.) $V_t^*(CS2, d_t) = V_t\big((CS2, d_t), y^*\big) \leq V_t((CS2, d_t), y^{--})$

2.) $V_t^*(CS2, d_t) = V_t\big((CS2, d_t), y^*\big) \leq V_t((CS2, d_t), y^{-})$

3.) $V_t^*(CS2, d_t) = V_t\big((CS2, d_t), y^*\big) \leq V_t((CS2, d_t), y^{+})$

4.) $V_t^*(CS2, d_t) = V_t\big((CS2, d_t), y^*\big) \leq V_t((CS2, d_t), y^{++})$

Taking one relation at a time, we get:

1.) $V_t\big((CS2, d_t), y^*\big) \leq V_t((CS1, d_t), y^{--})$

Substituting equation 5.6

$$R_t[(CS2, d_t), y^*] + \sum_{d_{t+1}} P(d_{t+1}|d_t)\, V_{t+1}^*(CS2 + y^*, d_{t+1})$$
$$\leq R_t[(CS2, d_t), y^{--}] + \sum_{d_{t+1}} P(d_{t+1}|d_t)\, V_{t+1}^*(CS2 + y^{--}, d_{t+1})$$

Substituting equation 5.2 for immediate expected reward

$$-c_2 y^* + c_3 CS2 + c_4 min(d_t, CS2) + c_5(d_t - \ CS2)^+$$
$$+ \sum_{d_{t+1}} P(d_{t+1}|d_t)\, V_{t+1}^*(e, d_{t+1})$$
$$\leq -c_2 y^{--} + c_3 CS2 + c_4 min(d_t, CS2) + c_5(d_t - \ CS2)^+$$
$$+ \sum_{d_{t+1}} P(d_{t+1}|d_t)\, V_{t+1}^*(b, d_{t+1})$$

Simplifying and grouping common terms we get:

$$-c_2 y^* + c_2 y^{--} \leq \sum_{d_{t+1}} P(d_{t+1}|d_t)\big(\, V_{t+1}^*(b, d_{t+1}) - \ V_{t+1}^*(e, d_{t+1})\big)$$

$$c_2(b-e) \leq \sum_{d_{t+1}} P(d_{t+1}|d_t)\big(\,V_{t+1}^*(b, d_{t+1}) - \,V_{t+1}^*(e, d_{t+1})\big) \tag{5.15}$$

2.) $V_t\big((CS2, d_t), y^*\big) \leq V_t((CS1, d_t), y^-)$

Substituting equation 5.6

$$R_t[(CS2, d), y^*] + \sum_{d_{t+1}} P(d_{t+1}|d)\, V_{t+1}^*(CS2 + y^*, d_{t+1})$$
$$\leq R_t[(CS2, d), y^-] + \sum_{d_{t+1}} P(d_{t+1}|d)\, V_{t+1}^*(CS2 + y^-, d_{t+1})$$

Substituting equation 5.2 for immediate expected reward

$$-c_2 y^* + c_3 CS2 + c_4 min(d, CS2) + c_5(d - \,CS2)^+ + \sum_{d_{t+1}} P(d_{t+1}|d)\, V_{t+1}^*(e, d_{t+1})$$
$$\leq -c_2 y^- + c_3 CS2 + c_4 min(d, CS2) + c_5(d - \,CS2)^+$$
$$+ \sum_{d_{t+1}} P(d_{t+1}|d)\, V_{t+1}^*(f, d_{t+1})$$

Simplifying and grouping common terms we get:

$$-c_2 y^* + c_2 y^- \leq \sum_{d_{t+1}} P(d_{t+1}|d)\big(\,V_{t+1}^*(f, d_{t+1}) - \,V_{t+1}^*(e, d_{t+1})\big)$$

$$c_2(f-e) \leq \sum_{d_{t+1}} P(d_{t+1}|d_t)\big(\,V_{t+1}^*(f, d_{t+1}) - \,V_{t+1}^*(e, d_{t+1})\big) \tag{5.16}$$

3.) $V_t\big((CS2, d_t), y^*\big) \leq V_t((CS1, d_t), y^+)$

Substituting equation 5.6

$$R_t[(CS2,d),y^*] + \sum_{d_{t+1}} P(d_{t+1}|d)\, V^*_{t+1}(CS2+y^*,d_{t+1})$$

$$\leq R_t[(CS2,d),y^+] + \sum_{d_{t+1}} P(d_{t+1}|d)\, V^*_{t+1}(CS2+y^+,d_{t+1})$$

Substituting equation 5.2 for immediate expected reward

$$-c_2y^* + c_3CS2 + c_4min(d,CS2) + c_5(d - \ CS2)^+ + \sum_{d_{t+1}} P(d_{t+1}|d)\, V^*_{t+1}(e,d_{t+1})$$

$$\leq -c_2y^+ + c_3CS2 + c_4min(d,CS2) + c_5(d - \ CS2)^+$$

$$+ \sum_{d_{t+1}} P(d_{t+1}|d)\, V^*_{t+1}(g,d_{t+1})$$

Simplifying and grouping common terms we get:

$$c_2(y^+ - y^*) \leq \sum_{d_{t+1}} P(d_{t+1}|d)\big(\, V^*_{t+1}(g,d_{t+1}) - \ V^*_{t+1}(e,d_{t+1})\big)$$

$$c_2(g - e) \leq \sum_{d_{t+1}} P(d_{t+1}|d_t)\big(\, V^*_{t+1}(g,d_{t+1}) - \ V^*_{t+1}(e,d_{t+1})\big) \tag{5.17}$$

4.) $V_t\big((CS2,d_t),y^*\big) \leq V_t((CS1,d_t),y^{++})$

Substituting equation 5.6

$$R_t[(CS2,d),y^*] + \sum_{d_{t+1}} P(d_{t+1}|d)\, V^*_{t+1}(CS2+y^*,d_{t+1})$$

$$\leq R_t[(CS2,d),y^{++}] + \sum_{d_{t+1}} P(d_{t+1}|d)\, V^*_{t+1}(CS2+y^{++},d_{t+1})$$

Substituting equation 5.2 for immediate expected reward

$$-c_2 y^* + c_3 CS2 + c_4 min(d, CS2) + c_5(d - CS2)^+ + \sum_{d_{t+1}} P(d_{t+1}|d)\, V_{t+1}^*(e, d_{t+1})$$

$$\leq c_1 y^{++} + c_3 CS2 + c_4 min(d, CS2) + c_5(d - CS2)^+$$

$$+ \sum_{d_{t+1}} P(d_{t+1}|d)\, V_{t+1}^*(h, d_{t+1})$$

Simplifying and grouping common terms we get:

$$-c_2 y^* - c_1 y^{++} \leq \sum_{d_{t+1}} P(d_{t+1}|d)\big(V_{t+1}^*(h, d_{t+1}) - V_{t+1}^*(e, d_{t+1})\big)$$

$$-c_2(e - CS2) - c_1(h - CS2) \qquad (5.18)$$

$$\leq \sum_{d_{t+1}} P(d_{t+1}|d_t)\big(V_{t+1}^*(h, d_{t+1}) - V_{t+1}^*(e, d_{t+1})\big)$$

Let $x^* = e - CS1$

Let x^{--} be any action such that $CS1 + x^{--} = b;\ b < e$

Let x^{-} be any action such that $CS1 + x^{-} = f;\ e \leq f < CS1$

Let x^{+} be any action such that $CS1 + x^{+} = g; CS1 < g \leq CS2$

Let x^{++} be any action such that $CS1 + x^{++} = h;\ CS2 < h$

If x^{--} was the optimum action, then

$$V_t\big((CS1, d_t), x^*\big) > V_t((CS1, d_t), x^{--})$$

Substituting equation 5.6

$$R_t[(CS1,d_t),x^*]+\sum_{d_{t+1}}P(d_{t+1}|d_t)\,V_{t+1}^*(CS1+x^*,d_{t+1})$$

$$>R_t[(CS1,d_t),x^{--}]+\sum_{d_{t+1}}P(d_{t+1}|d_t)\,V_{t+1}^*(CS1+x^{--},d_{t+1})$$

Substituting equation 5.2 for immediate expected reward

$$-c_2x^*+c_3CS1+c_4min(d_t,CS1)+c_5(d_t-\ CS1)^+$$

$$+\sum_{d_{t+1}}P(d_{t+1}|d_t)\,V_{t+1}^*(e,d_{t+1})$$

$$>-c_2x^{--}+c_3CS1+c_4min(d_t,CS1)+c_5(d_t-\ CS1)^+$$

$$+\sum_{d_{t+1}}P(d_{t+1}|d_t)\,V_{t+1}^*(b,d_{t+1})$$

Simplifying and grouping common terms we get:

$$c_2(x^{--}-x^*)>\sum_{d_{t+1}}P(d_{t+1}|d_t)\big(\,V_{t+1}^*(b,d_{t+1})-\ V_{t+1}^*(g,d_{t+1})\big)$$

$$c_2(b-e)>\sum_{d_{t+1}}P(d_{t+1}|d_t)\big(\,V_{t+1}^*(b,d_{t+1})-\ V_{t+1}^*(g,d_{t+1})\big) \tag{5.19}$$

We can see that equation 5.19 contradicts equation 5.15 hence x^{--} cannot be optimum decision.

If x^- was the optimum action, then

$$V_t\big((CS1,d_t),x^*\big)>V_t((CS1,d_t),x^-)$$

Substituting equation 5.6

$$R_t[(CS1,d),x^*] + \sum_{d_{t+1}} P(d_{t+1}|d)\, V^*_{t+1}(CS1 + x^*, d_{t+1})$$

$$> R_t[(CS1,d),x^-] + \sum_{d_{t+1}} P(d_{t+1}|d)\, V^*_{t+1}(CS1 + x^-, d_{t+1})$$

Substituting equation 5.2 for immediate expected reward

$$-c_2x^* + c_3CS1 + c_4min(d, CS1) + c_5(d - CS1)^+ + \sum_{d_{t+1}} P(d_{t+1}|d)\, V^*_{t+1}(e, d_{t+1})$$

$$> -c_2x^- + c_3CS1 + c_4min(d, CS1) + c_5(d - CS1)^+$$

$$+ \sum_{d_{t+1}} P(d_{t+1}|d)\, V^*_{t+1}(f, d_{t+1})$$

Simplifying and grouping common terms we get:

$$-c_2(x^* - x^-) > \sum_{d_{t+1}} P(d_{t+1}|d)\big(V^*_{t+1}(f, d_{t+1}) - V^*_{t+1}(e, d_{t+1})\big)$$

$$c_2(f - e) > \sum_{d_{t+1}} P(d_{t+1}|d_t)\big(V^*_{t+1}(f, d_{t+1}) - V^*_{t+1}(e, d_{t+1})\big) \tag{5.20}$$

We can see that equation 5.20 contradicts equation 5.16 hence x^- cannot be optimum decision.

If x^+ is the optimum action, then

$$V_t\big((CS1, d_t), x^*\big) > V_t((CS1, d_t), x^+)$$

Substituting equation 5.6

$$R_t[(CS1, d), x^*] + \sum_{d_{t+1}} P(d_{t+1}|d)\, V_{t+1}^*(CS1 + x^*, d_{t+1})$$

$$> R_t[(CS1, d), x^+] + \sum_{d_{t+1}} P(d_{t+1}|d)\, V_{t+1}^*(CS1 + x^+, d_{t+1})$$

Substituting equation 5.2 for immediate expected reward

$$-c_2 x^* + c_3 CS1 + c_4 min(d, CS1) + c_5(d - \ CS1)^+ + \sum_{d_{t+1}} P(d_{t+1}|d)\, V_{t+1}^*(e, d_{t+1})$$

$$> c_1 x^+ + c_3 CS1 + c_4 min(d, CS1) + c_5(d - \ CS1)^+$$

$$+ \sum_{d_{t+1}} P(d_{t+1}|d)\, V_{t+1}^*(g, d_{t+1})$$

Simplifying and grouping common terms we get:

$$-c_2 x^* - c_1 x^+ > \sum_{d_{t+1}} P(d_{t+1}|d)\big(\, V_{t+1}^*(g, d_{t+1}) - \ V_{t+1}^*(e, d_{t+1})\big)$$

$$-c_2(e - CS1) - c_1(g - CS1) \qquad (5.21)$$

$$> \sum_{d_{t+1}} P(d_{t+1}|d_t)\big(\, V_{t+1}^*(g, d_{t+1}) - \ V_{t+1}^*(e, d_{t+1})\big)$$

We see that the RHS of equation 5.21 and equation 5.17 are the same and the LHS of equation 5.21 is less than or equal to the LHS of equation 5.17. Thus, we have a contradiction and hence x^+ cannot be optimum decision.

If x^{++} is the optimum action, then

$$V_t\big((CS1, d_t), x^*\big) > V_t((CS1, d_t), x^{++})$$

Substituting equation 5.6

$$R_t[(CS1,d),x^*] + \sum_{d_{t+1}} P(d_{t+1}|d)\, V_{t+1}^*(CS1+x^*,d_{t+1})$$

$$> R_t[(CS1,d),x^{++}] + \sum_{d_{t+1}} P(d_{t+1}|d)\, V_{t+1}^*(CS1+x^{++},d_{t+1})$$

Substituting equation 5.2 for immediate expected reward

$$-c_2x^* + c_3CS1 + c_4min(d,CS1) + c_5(d-\ CS1)^+ + \sum_{d_{t+1}} P(d_{t+1}|d)\, V_{t+1}^*(e,d_{t+1})$$

$$> c_1x^{++} + c_3CS1 + c_4min(d,CS1) + c_5(d-\ CS1)^+$$

$$+ \sum_{d_{t+1}} P(d_{t+1}|d_t)\, V_{t+1}^*(h,d_{t+1})$$

Simplifying and grouping common terms we get:

$$-c_2x^* - c_1x^{++} > \sum_{d_{t+1}} P(d_{t+1}|d_t)\big(\, V_{t+1}^*(h,d_{t+1}) -\ V_{t+1}^*(e,d_{t+1})\big)$$

$$-c_2(e-CS1) - c_1(h-CS1) \tag{5.22}$$

$$> \sum_{d_{t+1}} P(d_{t+1}|d_t)\big(\, V_{t+1}^*(h,d_{t+1}) -\ V_{t+1}^*(e,d_{t+1})\big)$$

Again, we see that the RHS of equation 5.22 and equation 5.18 are the same and the LHS of equation 5.22 is less than or equal to the LHS of equation 5.18. Thus, we have a contradiction and hence x^{++} cannot be optimum decision.

Hence the optimal action at $CS1$ is x^*. This completes the proof for theorem 2.

We put forward the following Lemmas and theorems for the purpose of proving the main result in proposition 1.

Lemma 1:

If g is a superadditive (subadditive) function on X × Y and for each x ∈ X, $\min_{y \in Y} g(x, y)$ exists. then

$$f(x) = min\left\{y' \in \arg\min_{y \in Y} g(x, y)\right\}$$

is monotone non-increasing (non-decreasing) in x .

Definition of superadditivity from Puterman(2009): Let X and Y be partially ordered sets and $f(x, y)$ a real valued function on $X \times Y$. We say $f$is superadditive if for $x^{+} \geq x^{-}$in X and $y^{+} \geq y^{-}$in Y,

$$f(x^{+}, y^{+}) - f(x^{+}, y^{-}) \geq f(x^{-}, y^{+}) - f(x^{-}, y^{-})$$

This Lemma is stated and proved in Puterman (Puterman 2009_ENREF_34).

For subadditivity of $f(x, y)$ on $X \times Y$:

$$f(x^{+}, y^{+}) - f(x^{+}, y^{-}) \leq f(x^{-}, y^{+}) - f(x^{-}, y^{-})$$

Theorem 3:

$V_t\big((CS_t, d_t), a_t\big)$ is superadditive in $CS \times a \; \forall \; d_t, t$

That is for $CS2 \geq CS1$ in CS_t and $a2_t \geq a1_t$in a_t,

$$V_t\big((CS2, d_t), a2_t\big) - V_t\big((CS2, d_t), a1_t\big) \geq V_t\big((CS1, d_t), a2_t\big) - V_t\big((CS1, d_t), a1_t\big)$$

Proof:

The proof of this theorem is established using mathematical induction. This starts from the last time epoch T, where the capacity is to be salvaged.

From equation 5.3:

$$V_T\big((CS_T, d_T), a_T\big) = V_T^*(CS_T, d_T) = -c_6 CS_T \tag{5.23}$$

$V_T[(CS_T, d_T), a1_T] - V_T[(CS_T, d_T), a2_T] = 0$ for any CS, and any combination of $a1_T$ and $a2_T$. This gives $V_T\big((CS_T, d_T), a_T\big)$ superadditive in $CS \times a \;\forall\, d$

Writing equation 5.6

$$V_t\big((CS_t, d_t), a_t\big) = R_t\big((CS_t, d_t), a_t\big) + \sum_{d_{t+1}} P(d_{t+1}|d_t)\, V_{t+1}^*(CS_t + a_t, d_{t+1})$$

Considering the elements of this equation:

$R_t\big((CS_t, d_t), a_t\big) = c_1(a_t)^+ + c_2(-a_t)^+ + c_3(CS_t) + c_4 min(d_t, CS_t) + c_5(d_t - CS_t)^+$

$R_t\big((CS_t, d_t), a_t\big)$ is superadditive in $CS \times a \;\forall\, d_t, t$ because:

$R_t\big((CS_t, d_t), a2_t\big) - R_t\big((CS_t, d_t), a1_t\big)$ is a constant independent of $CS_t \;\forall\, a2_t \geq a1_t$.

Hence $V_t\big((CS_t, d_t), a_t\big)$ will be superadditive in $CS \times a$, if $\sum_{d_{t+1}} P(d_{t+1}|d)\, V_{t+1}^*(CS_t + a_t, d_{t+1})$ is superadditive in $CS \times a$.

Moreover, $\sum_{d_{t+1}} P(d_{t+1}|d)\, V_{t+1}^*(CS_t + a_t, d_{t+1})$ will be superadditive in $CS \times a$ if $V_{t+1}^*(CS_t + a_t, d_{t+1})$ is superadditive in $CS \times a$.

In conclusion, $V_t\big((CS_t, d_t), a_t\big)$ will be superadditive in $CS \times a \;\forall\, d$, if $V_{t+1}^*(CS_t + a_t, d_{t+1})$ is superadditive in $CS \times a$.

This will be shown by induction:

Since $V_T{}^*(CS_T, d_T) = -c_6.CS_T$, $V_T{}^*(CS_{T-1} + a_{T-1}, d_T)$ is superadditive in $CS \times a$, as for any $a2_{T-1} \geq a1_{T-1}$:

$$V_T{}^*(CS_{T-1} + a2_{T-1}, d_T) - V_T{}^*(CS_{T-1} + a1_{T-1}, d_T) = -c_6(a2_{T-1} - a1_{T-1})$$

is independent of CS_{T-1}.

Writing equation 5.6 for $t = T - 1$

$$\begin{aligned} V_{T-1}\big((CS_{T-1}, d_{T-1}), a_{T-1}\big) & \\ &= R_{T-1}[(CS_{T-1}, d_{T-1}), a_{T-1}] \\ &+ \sum_{d_T} P(d_T|d_{T-1})\, V_T^*(CS_{T-1} + a_{T-1}, d_T) \end{aligned} \tag{5.24}$$

Substituting equation 5.2 for $R_{T-1}[(CS_{T-1}, d_{T-1}), a_{T-1}]$ gives

$$\begin{aligned} V_{T-1}\big((CS_{T-1}, d_{T-1}), a_{T-1}\big) & \\ &= c_1(a_{T-1})^+ + c_2(-a_{T-1})^+ + c_3(CS_{T-1}) \\ &+ c_4 min(d_{T-1}, CS_{T-1}) + c_5(d_{T-1} - CS_{T-1})^+ \\ &+ \sum_{d_T} P(d_T|d_{T-1})\, V_T^*(CS_{T-1} + a_{T-1}, d_T) \end{aligned} \tag{5.25}$$

Substituting equation 5.23 in equation 5.25, we get

$$V_{T-1}\big((CS_{T-1}, d_{T-1}), a_{T-1}\big)$$
$$= c_1(a_{T-1})^+ + c_2(-a_{T-1})^+ + c_3(CS_{T-1}) + c_4 min(d_{T-1}, CS_{T-1})$$
$$+ c_5(d_{T-1} - CS_{T-1})^+ - \sum_{d_T} P(d_T|d_{T-1})c_6(CS_{T-1} + a_{T-1})$$

Simplifying:

$$V_{T-1}\big((CS_{T-1}, d_{T-1}), a_{T-1}\big) \tag{5.26}$$
$$= c_1(a_{T-1})^+ + c_2(-a_{T-1})^+ + c_3(CS_{T-1})$$
$$+ c_4 min(d_{T-1}, CS_{T-1}) + c_5(d_{T-1} - CS_{T-1})^+$$
$$- c_6(CS_{T-1} + a_{T-1})$$

The last term of equation 5.26, $-c_6(CS_{T-1} + a_{T-1})$ is superadditive in $CS \times A$

Hence $V_{T-1}\big((CS_{T-1}, d_{T-1}), a_{T-1}\big)$ is superadditive in $CS \times A$.

The following assumption on the cost parameter:

$$c_1 \geq max(0, -c_2)$$

concludes that there are three possible optimal actions at $t = T - 1$, namely

$a_{T-1}(CS_{T-1}, d_{T-1}) = 0$ if

$$c_2 \geq -c_6 \;\&\; c_1 \geq c_6$$

$a_{T-1}(CS_{T-1}, d_{T-1}) = -CS_{T-1}$ if

$$c_2 \leq -c_6 \;\&\; c_1 \geq c_6$$

$a_{T-1}(CS_{T-1}, d_{T-1}) = MC - CS_{T-1}$ if

$$c_1 \leq c_6$$

Hence $V_{T-1}^*(CS_{T-1}, d_{T-1})$ can be expressed, for these three possibilities as follows:

$$\begin{aligned}
&V_{T-1}^*(CS_{T-1}, d_{T-1}) = (c_3 - c_6)CS_{T-1} + c_4 min(d_{T-1}, CS_{T-1}) + \\
&c_5(d_{T-1} - CS_{T-1})^+ \text{ if } c_2 \geq -c_6 \,\&\, c_1 \geq c_6 \\
&V_{T-1}^*(CS_{T-1}, d_{T-1}) = (c_3 + c_2)CS_{T-1} + c_4 min(d_{T-1}, CS_{T-1}) + \\
&c_5(d_{T-1} - CS_{T-1})^+ \text{ if } c_2 \leq -c_6 \,\&\, c_1 \geq c_6 \\
&V_{T-1}^*(CS_{T-1}, d_{T-1}) = (c_3 - c_1)CS_{T-1} + c_4 min(d_{T-1}, CS_{T-1}) + \\
&c_5(d_{T-1} - CS_{T-1})^+ + (c_1 - c_6)MC \text{ if } c_1 \leq c_6
\end{aligned} \tag{5.27}$$

Next, we have the Bellman optimality equation for $t = T - 2$, which depends on $V_{T-1}^*()$ as follows:

$$\begin{aligned}
&V_{T-2}\big((CS_{T-2}, d_{T-2}), a_{T-2}\big) \\
&\qquad = R_{T-2}[(CS_{T-2}, d_{T-2}), a_{T-2}] \\
&\qquad + \sum_{d_{t+1}} P(d_{T-1}|d_{T-2})\, V_{T-1}^*(CS_{T-2} + a_{T-2}, d_{T-1})
\end{aligned}$$

As shown earlier, $V_{T-2}\big((CS_{T-2}, d_{T-2}), a_{T-2}\big)$ will be superadditive in $CS \times a \;\forall\, d$, if $V_{T-1}^*(CS_{T-2} + a_{T-2}, d_{T-1})$ is superadditive in $CS \times a$.

Let $a2_{T-2}, a1_{T-2}$ be two actions such that $a2_{T-2} \geq a1_{T-2}$

From equation 5.27, $V_{T-1}^*(CS_{T-2} + a2_{T-2}, d_{T-1}) - V_{T-1}^*(CS_{T-2} + a1_{T-2}, d_{T-1})$ is simplified by taking common terms to give:

$$
\begin{aligned}
&V^*_{T-1}(CS_{T-2} + a2_{T-2}, d_{T-1}) - V^*_{T-1}(CS_{T-2} + a1_{T-2}, d_{T-1}) = \\
&(c_3 - c_6)(a2_{T-2} - a1_{T-2}) + c_4\big(min(d_{T-1}, CS_{T-2} + a2_{T-2}) - \\
&min(d_{T-1}, CS_{T-2} + a1_{T-2})\big) + c_5((d_{T-1} - CS_{T-2} - a2_{T-2})^+ - \\
&(d_{T-1} - CS_{T-2} - a1_{T-2})^+) \text{ . if } c_2 \geq -c_6 \;\&\; c_1 \geq c_6 \\
&V^*_{T-1}(CS_{T-2} + a2_{T-2}, d_{T-1}) - V^*_{T-1}(CS_{T-2} + a1_{T-2}, d_{T-1}) = \\
&(c_3 + c_2)(a2_{T-2} - a1_{T-2}) + c_4\big(min(d_{T-1}, CS_{T-2} + a2_{T-2}) - \\
&min(d_{T-1}, CS_{T-2} + a1_{T-2})\big) + c_5((d_{T-1} - CS_{T-2} - a2_{T-2})^+ - \\
&(d_{T-1} - CS_{T-2} - a1_{T-2})^+) \text{ . if } c_2 \leq -c_6 \;\&\; c_1 \geq c_6 \\
&V^*_{T-1}(CS_{T-2} + a2_{T-2}, d_{T-1}) - V^*_{T-1}(CS_{T-2} + a1_{T-2}, d_{T-1}) = \\
&(c_3 - c_1)(a2_{T-2} - a1_{T-2}) + c_4\big(min(d_{T-1}, CS_{T-2} + a2_{T-2}) - \\
&min(d_{T-1}, CS_{T-2} + a1_{T-2})\big) + c_5((d_{T-1} - CS_{T-2} - a2_{T-2})^+ - \\
&(d_{T-1} - CS_{T-2} - a1_{T-2})^+) \text{ . if } c_1 \leq c_6
\end{aligned}
\tag{5.28}
$$

The first parts in equation 5.28: $(c_3 - c_6)(a2_{T-2} - a1_{T-2}) \mid (c_3 + c_2)(a2_{T-2} - a1_{T-2}) \mid (c_3 - c_1)(a2_{T-2} - a1_{T-2})$ are constants. Studying the second part, which is similar in all three possibilities of equation 5.28: $c_4\big(min(d_{T-1}, CS_{T-2} + a2_{T-2}) - min(d_{T-1}, CS_{T-2} + a1_{T-2})\big) + c_5((d_{T-1} - CS_{T-2} - a2_{T-2})^+ - (d_{T-1} - CS_{T-2} - a1_{T-2})^+)$ is increasing in CS from Lemma 2 given in the appendix of this chapter – section 5.6. Which means $V^*_{T-1}(CS_{T-2} + a2_{T-2}, d_{T-1}) - V^*_{T-1}(CS_{T-2} + a1_{T-2}, d_{T-1})$ is increasing in CS_{T-2}. This gives $V_{T-2}\big((CS_{T-2}, d_{T-2}), a_{T-2}\big)$ superadditive in $CS \times a$.

So far, the following results have been established:

1. $V_{T-2}\big((CS_{T-2}, d_{T-2}), a_{T-2}\big), V_{T-1}\big((CS_{T-1}, d_{T-1}), a_{T-1}\big), V_T\big((CS_T, d_T), a_T\big)$ is superadditive in $CS \times a$.
2. $V^*_{T-1}(CS_{T-2} + a_{T-2}, d_{T-1}),\ V^*_T(CS_T + a_T, d_T)$ is superadditive in $CS \times a$.

This gives the first step of the induction proof of theorem 3. Next assuming that $V_{t+1}\big((CS_{t+1}, d_{t+1}), a_{t+1}\big)$ and $V_{t+2}{}^*(CS_{t+1} + a_{t+1}, d_{t+2})$ are both superadditive in $CS \times a \ \forall\ d_{t+1},$ we need to prove that $V_{t+1}{}^*(CS_t + a_t, d_{t+1})$ is superadditive in $CS \times a$ to give $V_t\big((CS_t, d_t), a_t\big)$ superadditive in $CS \times a$.

That is to check if $V_{t+1}{}^*(CS_t + a2_t, d_{t+1}) - V_{t+1}{}^*(CS_t + a1_t, d_{t+1})$ is nondecreasing in CS for $a2_t \geq a1_t \forall\ d_{t+1}$

Let optimum action in time epoch $t + 1$ and state $(CS_{t+1} = 0, d_{t+1})$ be K.

$$a_{t+1}(0, d_{t+1}) = K$$

And optimum action in time epoch $t + 1$ and state $(CS_{t+1} = MC, d_{t+1})$ be $L - MC$

$$a_{t+1}(MC, d_{t+1}) = L - MC$$

From theorem 1 & 2 it is understood that $L \geq K$.

let $K > a2 > a1$ ($a2 > k > a1$ and $a2 > a1 > k$ will be special cases of this)

We have two possibilities:

Possibility 1: $K - a1_t \leq L - a2_t$

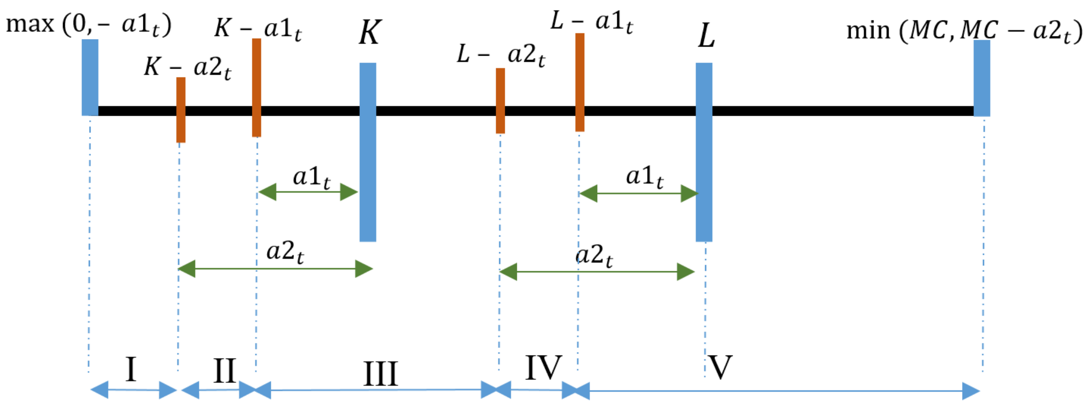

Figure 6: Illustration for Theorem 3, possibility 1 - CS_t

As shown in the figure above for possibility 1, the resulting regions are considered:

Region 1:

$CS_{t+1} \leq K$ in both functions $V_{t+1}{}^*(CS_t + a2_t, d_{t+1})$ & $V_{t+1}{}^*(CS_t + a1_t, d_{t+1})$

This implies:

$$\max(0, -a1_t) \leq CS_t \leq K - a2_t$$

From theorem 1 the optimal action will be to go to capacity K for both states $(CS_t + a2_t, d_{t+1})$ and $(CS_t + a1_t, d_{t+1})$, hence:

$$V_{t+1}{}^*(CS_t + a2_t, d_{t+1}) - V_{t+1}{}^*(CS_t + a1_t, d_{t+1})$$

$$= c_1(K - CS_t - a2_t) + c_3(CS_t + a2_t) + c_4 min(d_{t+1}, CS_t + a2)$$

$$+ c_5(d_{t+1} - \ CS_t - a2_t)^+$$

$$-\ (c_1(K - CS_t - a1_t) + c_3(CS_t + a1_t) + c_4 min(d_{t+1}, CS_t + a1_t)$$

$$+ c_5(d_{t+1} - \ CS_t - a1_t)^+)$$

Simplifying

$${V_{t+1}}^*(CS_t + a2_t, d_{t+1}) - {V_{t+1}}^*(CS_t + a1_t, d_{t+1})$$

$$= (c_3 - c_1)(a2_t - a1_t)$$

$$+ c_4\big(min(d_{t+1}, CS_t + a2_t) - min(d_{t+1}, CS_t + a1_t)\big)$$

$$+ c_5((d_{t+1} - CS_t - a2_t)^+ - (d_{t+1} - CS_t - a1_t)^+)$$

Lemma 2 shows that $c_4\big(min(d_{t+1}, CS_t + a2) - min(d_{t+1}, CS_t + a1_t)\big) +$ $c_5((d_{t+1} - CS_t - a2_t)^+ - (d_{t+1} - CS_t - a1_t)^+)$ is nondecreasing in CS_t.

Hence in region 1, ${V_{t+1}}^*(CS_t + a_t, d_{t+1})$ superadditive in $CS \times a$

Region 2:

$CS_{t+1} \leq K$ in the function ${V_{t+1}}^*(CS_t + a1_t, d_{t+1})$ and $L \geq CS_{t+1} \geq K$ in the function ${V_{t+1}}^*(CS_t + a2_t, d_{t+1})$

This implies:

$$K - a2_t \leq CS_t \leq K - a1_t$$

From theorem 1and theorem 2, the optimal action will be to go to capacity K for states $(CS_t + a1_t, d_{t+1})$ and to take action $a = 0$ for states $(CS_t + a2_t, d_{t+1})$

$$V_{t+1}{}^{*}(CS_t + a2_t, d_{t+1}) - V_{t+1}{}^{*}(CS_t + a1_t, d_{t+1})$$

$$= c_3(CS_t + a2_t) + c_4 min(d_{t+1}, CS_t + a2_t) + c_5(d_{t+1} - CS_t - a2_t)^+$$

$$+ \sum_{d_{t+2}} P(d_{t+2}|d_{t+1})\, V_{t+2}^{*}(CS_t + a2_t, d_{t+2})$$

$$- \Bigg(c_1(K - CS_t - a1_t) + c_3(CS_t + a1_t) + c_4 min(d_{t+1}, CS_t + a1_t)$$

$$+ c_5(d_{t+1} - CS_t - a1_t)^+ + \sum_{d_{t+2}} P(d_{t+2}|d_{t+1})\, V_{t+2}^{*}(K, d_{t+2}) \Bigg)$$

Simplifying

$$\begin{aligned} V_{t+1}{}^{*}(CS + a2, d) - V_{t+1}{}^{*}(CS + a1, d) & \\ &= c_3(a2_t - a1_t) \\ &+ c_4\big(min(d_{t+1}, CS_t + a2_t) - min(d_{t+1}, CS_t + a1_t)\big) \\ &+ c_5((d_{t+1} - CS_t - a2_t)^+ - (d_{t+1} - CS_t - a1_t)^+) \\ &- c_1(K - CS_t - a1_t) \\ &+ \sum_{d_{t+2}} P(d_{t+2}|d_{t+1})\big(V_{t+2}^{*}(CS_t + a2_t, d_{t+2}) \\ &- V_{t+2}^{*}(K, d_{t+2})\big) \end{aligned} \tag{5.29}$$

Studying equation 5.29 in parts:

1. $c_3(a2_t - a1_t) - \sum_{d_{t+2}} P(d_{t+2}|d)\, V_{t+2}^{*}(K, d_{t+2}) - c_1(K - a1)$ is fixed in this region
2. Lemma 2 shows that $c_4\big(min(d_{t+1}, CS_t + a2_t) - min(d_{t+1}, CS_t + a1_t)\big) + c_5((d_{t+1} - CS_t - a2_t)^+ - (d_{t+1} - CS_t - a1_t)^+)$ is nondecreasing in CS.

3. The last part:

$$c_1 CS_t + \sum_{d_{t+2}} P(d_{t+2}|d_{t+1})\, V_{t+2}^*(CS_t + a2_t, d_{t+2}) \tag{5.30}$$

In this region:

$$V_{t+1}\big((CS_t + a2_t, d_{t+1}), 0\big) \leq V_{t+1}\big((CS_t + a2_t, d_{t+1}), 1\big)$$

This gives:

$$\sum_{d_{t+2}} P(d_{t+2}|d_{t+1})\, V_{t+2}^*(CS_t + a2_t, d_{t+2})$$
$$\leq c_1 + \sum_{d_{t+2}} P(d_{t+2}|d_{t+1})\, V_{t+2}^*(CS_t + a2_t + 1, d_{t+2})$$

$$\sum_{d_{t+2}} P(d_{t+2}|d_{t+1})\, V_{t+2}^*(CS_t + a2_t + 1, d_{t+2}) - \sum_{d_{t+2}} P(d_{t+2}|d_{t+1})\, V_{t+2}^*(CS_t + a2_t, d_{t+2}) \geq -c_1 \tag{5.31}$$

Equation 5.31 implies that for unit increase in CS_t the decrease in

$\sum_{d_{t+2}} P(d_{t+2}|d_{t+1})\, V_{t+2}^*(CS_t + a2_t, d_{t+2})$ in equation 5.30 will not be more than c_1.

This gives the last part $c_1 CS_t + \sum_{d_{t+2}} P(d_{t+2}|d)\, V_{t+2}^*(CS_t + a2_t, d_{t+2})$ increasing in CS_t

Hence in region 2, ${V_{t+1}}^*(CS_t + a_t, d_{t+1})$ superadditive in $CS \times a$

Region 3:

$L \geq CS_{t+1} \geq K$ in both the functions ${V_{t+1}}^*(CS_t + a2_t, d_{t+1})$ and ${V_{t+1}}^*(CS_t + a1_t, d_{t+1})$

This implies:

$$K - a1_t \leq CS_t \leq L - a2_t$$

$$V_{t+1}{}^*(CS_t + a2_t, d_{t+1}) - V_{t+1}{}^*(CS_t + a1_t, d_{t+1})$$

$$= c_3(CS_t + a2_t) + c_4 min(d_{t+1}, CS_t + a2_t) + c_5(d_{t+1} - CS_t - a2_t)^+$$

$$+ \sum_{d_{t+2}} P(d_{t+2}|d_{t+1})\, V^*_{t+2}(CS_t + a2_t, d_{t+2})$$

$$-\Bigg(c_3(CS_t + a1_t) + c_4 min(d_{t+1}, CS_t + a1_t) + c_5(d_{t+1} - CS_t - a1_t)^+$$

$$+ \sum_{d_{t+2}} P(d_{t+2}|d_{t+1})\, V^*_{t+2}(CS_t + a1_t, d_{t+2})\Bigg)$$

Simplifying

$$V_{t+1}{}^*(CS_t + a2_t, d_{t+1}) - V_{t+1}{}^*(CS_t + a1_t, d_{t+1})$$

$$= c_3(a2_t - a1_t) + c_4\big(min(d_{t+1}, CS_t + a2_t) - min(d_{t+1}, CS_t + a1_t)\big)$$

$$+ c_5((d_{t+1} - CS_t - a2_t)^+ - (d_{t+1} - CS_t - a1_t)^+)$$

$$+ \sum_{d_{t+2}} P(d_{t+2}|d_{t+1})\big(V^*_{t+2}(CS_t + a2_t, d_{t+2}) - V^*_{t+2}(CS_t + a1_t, d_{t+2})\big)$$

$c_3(a2_t - a1_t)$ a fixed value. Lemma 2 shows that $c_4\big(min(d_{t+1}, CS_t + a2_t) - min(d_{t+1}, CS_t + a1_t)\big) + c_5((d_{t+1} - CS_t - a2_t)^+ - (d_{t+1} - CS_t - a1_t)^+)$ is nondecreasing in CS. $V^*_{t+2}(CS_t + a2_t, d_{t+2}) - V^*_{t+2}(CS_t + a1_t, d_{t+2})$ is given to be increasing in CS, because $V^*_{t+2}(CS_t + a_t, d_{t+2})$ is superadditive in $CS \times a$.

Hence in region 3, $V_{t+1}{}^*(CS_t + a_t, d_{t+1})$ superadditive in $CS \times a$

Region 4:

$L \geq CS_{t+1} \geq K$ in the function $V_{t+1}{}^{*}(CS_t + a1_t, d_{t+1})$ and $CS_{t+1} \geq L$ in the function $V_{t+1}{}^{*}(CS_t + a2_t, d_{t+1})$

This implies:

$$L - a2_t \leq CS_t \leq L - a1_t$$

$$\begin{aligned}
&V_{t+1}{}^{*}(CS_t + a2_t, d_{t+1}) - V_{t+1}{}^{*}(CS_t + a1_t, d_{t+1}) \\
&\qquad = c_2(CS_t + a2_t - L) + c_3(CS_t + a2_t) + c_4 min(d_{t+1}, CS_t + a2_t) \\
&\qquad + c_5(d_{t+1} - CS_t - a2_t)^{+} + \sum_{d_{t+2}} P(d_{t+2}|d_{t+1})\, V_{t+2}^{*}(L, d_{t+2}) \\
&\qquad - \Bigg(c_3(CS_t + a1_t) + c_4 min(d_{t+1}, CS_t + a1_t) + c_5(d_{t+1} - CS_t - a1_t)^{+} \\
&\qquad + \sum_{d_{t+2}} P(d_{t+2}|d_{t+1})\, V_{t+2}^{*}(CS_t + a1_t, d_{t+2}) \Bigg)
\end{aligned}$$

Simplifying:

$$V_{t+1}{}^*(CS_t + a2_t, d_{t+1}) - V_{t+1}{}^*(CS_t + a1_t, d_{t+1}) \tag{5.32}$$

$$= c_2(CS_t + a2_t - L) + c_3(a2_t - a1_t)$$

$$+ c_4\big(min(d_{t+1}, CS_t + a2_t) - min(d_{t+1}, CS_t + a1_t)\big)$$

$$+ c_5((d_{t+1} - CS_t - a2_t)^+ - (d_{t+1} - CS_t - a1_t)^+)$$

$$+ \sum_{d_{t+2}} P(d_{t+2}|d_{t+1})\big(V_{t+2}^*(L, d_{t+2})$$

$$- V_{t+2}^*(CS_t + a1_t, d_{t+2})\big)$$

Studying equation 5.32 in parts:

1. $c_3(a2_t - a1_t) + \sum_{d_{t+2}} P(d_{t+2}|d_{t+1})\, V_{t+2}^*(L, d_{t+2}) + c_2(a2_t - L)$ is fixed in this region
2. Lemma 2 shows that $c_4\big(min(d_{t+1}, CS_t + a2_t) - min(d_{t+1}, CS_t + a1_t)\big) + c_5((d_{t+1} - CS_t - a2_t)^+ - (d_{t+1} - CS_t - a1_t)^+)$ is nondecreasing in CS.
3. The last part:

$$c_2\, CS_t - \sum_{d_{t+2}} P(d_{t+2}|d_{t+1})\, V_{t+2}^*(CS_t + a1_t, d_{t+2}) \tag{5.33}$$

In this region:

$$V_{t+1}\big((CS_t + a1_t, d_{t+1}), 0\big) \leq V_{t+1}\big((CS_t + a1_t, d_{t+1}), -1\big)$$

$$\sum_{d_{t+2}} P(d_{t+2}|d_{t+1})\, V_{t+2}^*(CS_t + a1_t, d_{t+2})$$

$$\leq c_2 + \sum_{d_{t+2}} P(d_{t+2}|d_{t+1})\, V_{t+2}^*(CS_t + a1_t - 1, d_{t+2})$$

$$-c_2 \leq \sum_{d_{t+2}} P(d_{t+2}|d_{t+1})\, V_{t+2}^{*}(CS_t + a1_t - 1, d_{t+2})$$

$$- \sum_{d_{t+2}} P(d_{t+2}|d_{t+1})\, V_{t+2}^{*}(CS_t + a1_t, d_{t+2})$$

$$\sum_{d_{t+2}} P(d_{t+2}|d_{t+1})\, V_{t+2}^{*}(CS_t + a1_t, d_{t+2}) - \sum_{d_{t+2}} P(d_{t+2}|d_{t+1})\, V_{t+2}^{*}(CS_t + a1_t - 1, d_{t+2})$$

$$\leq c_2$$

$$-c_2 \leq \sum_{d_{t+2}} P(d_{t+2}|d_{t+1})\, V_{t+2}^{*}(CS_t + a1_t - 1, d_{t+2}) \tag{5.34}$$

$$- \sum_{d_{t+2}} P(d_{t+2}|d_{t+1})\, V_{t+2}^{*}(CS_t + a1_t, d_{t+2})$$

From equation 5.34, it implies for unit increase in CS_t, the last part equation 5.33 increasing in CS_t

Hence in region 4, ${V_{t+1}}^{*}(CS_t + a_t, d_{t+1})$ superadditive in $CS \times a$

Region 5:

$CS_{t+1} \geq L$ in both the functions ${V_{t+1}}^{*}(CS_t + a1_t, d_{t+1})$ and ${V_{t+1}}^{*}(CS_t + a2_t, d_{t+1})$

This implies:

$$L - a1_t \leq CS_t$$

$$V_{t+1}{}^{*}(CS_t + a2_t, d_{t+1}) - V_{t+1}{}^{*}(CS_t + a1_t, d_{t+1})$$

$$= c_2(CS_t + a2_t - L) + c_3(CS_t + a2_t) + c_4 min(d_{t+1}, CS_t + a2_t)$$

$$+ c_5(d_{t+1} - CS_t - a2_t)^+ + \sum_{d_{t+2}} P(d_{t+2}|d_{t+1})\, V_{t+2}^*(L, d_{t+2})$$

$$-\Bigg(c_2(CS + a1 - L) + c_3(CS_t + a1_t) + c_4 min(d_{t+1}, CS_t + a1_t)$$

$$+ c_5(d_{t+1} - CS_t - a1_t)^+ + \sum_{d_{t+2}} P(d_{t+2}|d_{t+1})\, V_{t+2}^*(CS_t + a1_t, d_{t+2})\Bigg)$$

$$V_{t+1}{}^{*}(CS_t + a2_t, d_{t+1}) - V_{t+1}{}^{*}(CS_t + a1_t, d_{t+1})$$

$$= (c_2 + c_3)(a2_t - a1_t)$$

$$+ c_4\big(min(d_{t+1}, CS_t + a2_t) - min(d_{t+1}, CS_t + a1_t)\big)$$

$$+ c_5((d_{t+1} - CS_t - a2_t)^+ - (d_{t+1} - CS_t - a1_t)^+)$$

$(c_2 + c_3)(a2_t - a1_t)$ constant with increase in CS_t. Lemma 2 shows that $c_4\big(min(d_{t+1}, CS_t + a2_t) - min(d_{t+1}, CS_t + a1_t)\big) + c_5((d_{t+1} - CS_t - a2_t)^+ - (d_{t+1} - CS_t - a1_t)^+)$ is nondecreasing in CS_t. Hence in region 5, $V_{t+1}{}^{*}(CS_t + a_t, d_{t+1})$ superadditive in $CS \times a$

This completes the proof for possibility 1.

Possibility 2: $K - a1_t > L - a2_t$

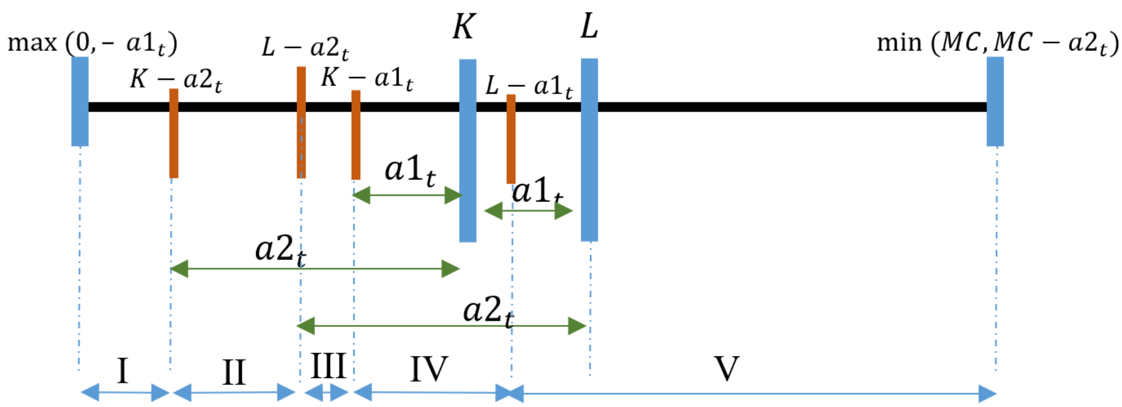

Figure 7: Illustration for theorem 3, possibility 2 - CS_t

As shown in the figure above for possibility 1, the resulting regions are considered:

Region 1:

$CS_{t+1} \leq K$ in both functions $V_{t+1}{}^{*}(CS_t + a2_t, d_{t+1})$ & $V_{t+1}{}^{*}(CS_t + a1_t, d_{t+1})$

This implies:

$$\max(0, -a1_t) \leq CS_t \leq K - a2_t$$

Same as region 1 for possibility 1

Hence in region 1, $V_{t+1}{}^{*}(CS_t + a_t, d_{t+1})$ is superadditive in $CS \times a$

Region 2:

$CS_{t+1} \leq K$ in the function $V_{t+1}{}^{*}(CS_t + a1_t, d_{t+1})$ and $L \geq CS_{t+1} \geq K$ in the function $V_{t+1}{}^{*}(CS_t + a2_t, d_{t+1})$

This implies:

$$K - a2_t \leq CS_t \leq L - a1_t$$

From theorem 1and theorem 2, the optimal action will be to go to capacity K for states $(CS_t + a1_t, d_{t+1})$ and to take action $a = 0$ for states $(CS_t + a2_t, d_{t+1})$

The region is given the same consideration as in region 2 of possibility 1.

$$
\begin{aligned}
&{V_{t+1}}^*(CS_t + a2_t, d_{t+1}) - {V_{t+1}}^*(CS_t + a1_t, d_{t+1}) \\
&\qquad = c_3(CS_t + a2_t) + c_4 min(d_{t+1}, CS_t + a2) + c_5(d_{t+1} - \ CS_t - a2_t)^+ \\
&\qquad + \sum_{d_{t+2}} P(d_{t+2}|d_{t+1})\, V_{t+2}^*(CS_t + a2_t, d_{t+2}) \\
&\qquad - \Bigg(c_1(K - CS_t - a1_t) + c_3(CS_t + a1_t) + c_4 min(d_{t+1}, CS_t + a1_t) \\
&\qquad + c_5(d_{t+1} - \ CS_t - a1_t)^+ + \sum_{d_{t+2}} P(d_{t+2}|d_{t+1})\, V_{t+2}^*(K, d_{t+2}) \Bigg)
\end{aligned}
$$

Simplifying

$$
\begin{aligned}
&{V_{t+1}}^*(CS_t + a2_t, d_{t+1}) - {V_{t+1}}^*(CS_t + a1_t, d_{t+1}) \\
&\qquad = c_3(a2_t - a1_t) \\
&\qquad + c_4\big(min(d_{t+1}, CS_t + a2_t) - min(d_{t+1}, CS_t + a1_t)\big) \\
&\qquad + c_5((d_{t+1} - \ CS_t - a2_t)^+ - (d_{t+1} - \ CS_t - a1_t)^+) \\
&\qquad - c_1(K - CS_t - a1_t) \\
&\qquad + \sum_{d_{t+2}} P(d_{t+2}|d_{t+1})\big(V_{t+2}^*(CS_t + a2_t, d_{t+2}) \\
&\qquad - \ V_{t+2}^*(K, d_{t+2})\big)
\end{aligned} \tag{5.35}
$$

Studying equation 5.35 in parts:

1. $c_3(a2_t - a1_t) + \sum_{d_{t+2}} P(d_{t+2}|d_{t+1})\, V^*_{t+2}(K, d_{t+2}) - c_1(K - a1_t)$ is fixed in this region
2. Lemma 2 shows that $c_4\big(min(d_{t+1}, CS_t + a2_t) - min(d_{t+1}, CS_t + a1_t)\big) + c_5((d_{t+1} - CS_t - a2_t)^+ - (d_{t+1} - CS_t - a1_t)^+)$ is nondecreasing in CS.
3. The last part:

$$c_1 CS_t + \sum_{d_{t+2}} P(d_{t+2}|d_{t+1})\, V^*_{t+2}(CS_t + a2_t, d_{t+2}) \tag{5.36}$$

In this region:

$$V_{t+1}\big((CS_t + a2_t, d_{t+1}), 0\big) \le V_{t+1}\big((CS_t + a2_t, d_{t+1}), 1\big)$$

This gives:

$$\sum_{d_{t+2}} P(d_{t+2}|d_{t+1})\, V^*_{t+2}(CS_t + a2_t, d_{t+2}) \le c_1 + \sum_{d_{t+2}} P(d_{t+2}|d_{t+1})\, V^*_{t+2}(CS_t + a2_t + 1, d_{t+2})$$

$$\sum_{d_{t+2}} P(d_{t+2}|d_{t+1})\, V^*_{t+2}(CS_t + a2_t + 1, d_{t+2}) - \sum_{d_{t+2}} P(d_{t+2}|d_{t+1})\, V^*_{t+2}(CS_t + a2_t, d_{t+2}) \ge -c_1 \tag{5.37}$$

Equation 5.37 implies that for unit increase in CS_t the decrease in $\sum_{d_{t+2}} P(d_{t+2}|d_{t+1})\, V^*_{t+2}(CS_t + a2_t, d_{t+2})$ in equation 5.36 will not be more than c_1.

This gives the last part '$c_1 CS_t + \sum_{d_{t+2}} P(d_{t+2}|d_{t+1})\, V_{t+2}^{*}(CS_t + a2_t, d_{t+2})$' increasing in CS_t

Hence in region 2, ${V_{t+1}}^{*}(CS_t + a_t, d_{t+1})$ superadditive in $CS \times a$

Region 3:

$CS_{t+1} \leq K$ in the function ${V_{t+1}}^{*}(CS_t + a1_t, d_{t+1})$ and $CS_{t+1} \geq L$ in the function ${V_{t+1}}^{*}(CS_t + a2_t, d_{t+1})$

This implies:

$$L - a2_t \leq CS_t \leq K - a1_t$$

$$\begin{aligned}
&{V_{t+1}}^{*}(CS_t + a2_t, d_{t+1}) - {V_{t+1}}^{*}(CS_t + a1_t, d_{t+1}) \\
&\quad = c_2(CS_t + a2_t - L) + c_3(CS_t + a2_t) + c_4 min(d_{t+1}, CS_t + a2) \\
&\quad + c_5(d_{t+1} - CS_t - a2_t)^{+} + \sum_{d_{t+2}} P(d_{t+2}|d_{t+1})\, V_{t+2}^{*}(L, d_{t+2}) \\
&\quad - \left(c_1(K - CS_t - a1_t) + c_3(CS_t + a1_t) + c_4 min(d_{t+1}, CS_t + a1_t) \right. \\
&\quad \left. + c_5(d_{t+1} - CS_t - a1_t)^{+} + \sum_{d_{t+2}} P(d_{t+2}|d_{t+1})\, V_{t+2}^{*}(K, d_{t+2}) \right)
\end{aligned}$$

Simplifying

$$V_{t+1}{}^{*}(CS_t + a2_t, d_{t+1}) - V_{t+1}{}^{*}(CS_t + a1_t, d_{t+1}) \tag{5.38}$$

$$= -c_1(K - CS_t - a1_t) + c_2(CS_t + a2_t - L)$$

$$+ c_3(a2_t - a1_t)$$

$$+ c_4\big(min(d_{t+1}, CS_t + a2_t) - min(d_{t+1}, CS_t + a1_t)\big)$$

$$+ c_5((d_{t+1} - CS_t - a2_t)^+ - (d_{t+1} - CS_t - a1_t)^+)$$

$$+ \sum_{d_{t+2}} P(d_{t+2}|d_{t+1})\big(V_{t+2}^*(L, d_{t+2}) - V_{t+2}^*(K, d_{t+2})\big)$$

Studying equation 5.38 in parts:

1. $-c_1(K - a1_t) + c_2(a2_t - L) + c_3(a2_t - a1_t) + \sum_{d_{t+2}} P(d_{t+2}|d_{t+1})\big(V_{t+2}^*(L, d_{t+2}) - V_{t+2}^*(K, d_{t+2})\big)$ is a fixed value in this region for increase in CS_t
2. Previously; it was shown that $c_4\big(min(d_{t+1}, CS_t + a2_t) - min(d_{t+1}, CS_t + a1_t)\big) + c_5((d_{t+1} - CS_t - a2_t)^+ - (d_{t+1} - CS_t - a1_t)^+)$ is nondecreasing in CS_t.
3. $(c_1 + c_2)CS_t$ increases with CS_t considering the assumption $c_1 \geq max(0, -c_2)$

   Hence in region 3, $V_{t+1}{}^{*}(CS_t + a_t, d_{t+1})$ superadditive in $CS \times a$

Region 4:

$L \geq CS_{t+1} \geq K$ in the function $V_{t+1}{}^{*}(CS_t + a1_t, d_{t+1})$ and $CS_{t+1} \geq L$ in the function $V_{t+1}{}^{*}(CS_t + a2_t, d_{t+1})$

This implies:

$$K - a1_t \leq CS_t \leq L - a1_t$$

$$V_{t+1}{}^{*}(CS_t + a2_t, d_{t+1}) - V_{t+1}{}^{*}(CS_t + a1_t, d_{t+1})$$

$$= c_2(CS_t + a2_t - L) + c_3(CS_t + a2_t) + c_4 min(d_{t+1}, CS_t + a2)$$

$$+ c_5(d_{t+1} - CS_t - a2_t)^+ + \sum_{d_{t+2}} P(d_{t+1}|d_{t+1})\, V_{t+2}^*(L, d_{t+2})$$

$$-\Bigg(c_3(CS_t + a1_t) + c_4 min(d_{t+1}, CS_t + a1_t) + c_5(d_{t+1} - CS_t - a1_t)^+$$

$$+ \sum_{d_{t+2}} P(d_{t+2}|d_{t+1})\, V_{t+2}^*(CS_t + a1_t, d_{t+2})\Bigg)$$

Simplifying:

$$V_{t+1}{}^{*}(CS_t + a2_t, d_{t+1}) - V_{t+1}{}^{*}(CS_t + a1_t, d_{t+1}) \tag{5.39}$$

$$= c_2(CS_t + a2_t - L) + c_3(a2_t - a1_t)$$

$$+ c_4\big(min(d_{t+1}, CS_t + a2_t) - min(d_{t+1}, CS_t + a1_t)\big)$$

$$+ c_5((d_{t+1} - CS_t - a2_t)^+ - (d_{t+1} - CS_t - a1_t)^+)$$

$$+ \sum_{d_{t+2}} P(d_{t+2}|d_{t+1})\big(V_{t+2}^*(L, d_{t+2})$$

$$- V_{t+2}^*(CS_t + a1_t, d_{t+2})\big)$$

Studying equation 5.39 in parts:

1. $c_3(a2_t - a1_t) + \sum_{d_{t+2}} P(d_{t+2}|d_{t+1})\, V_{t+2}^*(L, d_{t+2}) + c_2(a2_t - L)$ is fixed in this region
2. Previously; it was shown that $c_4\big(min(d_{t+1}, CS_t + a2_t) - min(d_{t+1}, CS_t + a1_t)\big) + c_5((d_{t+1} - CS_t - a2_t)^+ - (d_{t+1} - CS_t - a1_t)^+)$ is nondecreasing in CS.

3. The last part:

$$c_2\, CS_t - \sum_{d_{t+2}} P(d_{t+2}|d_{t+1})\, V_{t+2}^{*}(CS_t + a1_t, d_{t+2}) \tag{5.40}$$

In this region:

$$V_{t+1}\big((CS_t + a1_t, d_{t+1}), 0\big) \leq V_{t+1}\big((CS_t + a1_t, d_{t+1}), -1\big)$$

$$\sum_{d_{t+2}} P(d_{t+2}|d_{t+1})\, V_{t+2}^{*}(CS_t + a1_t, d_{t+2})$$

$$\leq c_2 + \sum_{d_{t+2}} P(d_{t+2}|d_{t+1})\, V_{t+2}^{*}(CS_t + a1_t - 1, d_{t+2})$$

$$-c_2 \leq \sum_{d_{t+2}} P(d_{t+2}|d_{t+1})\, V_{t+2}^{*}(CS_t + a1_t - 1, d_{t+2})$$

$$- \sum_{d_{t+2}} P(d_{t+2}|d_{t+1})\, V_{t+2}^{*}(CS_t + a1_t, d_{t+2})$$

$$\sum_{d_{t+2}} P(d_{t+2}|d_{t+1})\, V_{t+2}^{*}(CS_t + a1_t, d_{t+2}) - \sum_{d_{t+2}} P(d_{t+2}|d_{t+1})\, V_{t+2}^{*}(CS_t + a1_t - 1, d_{t+2})$$

$$\leq c_2$$

$$-c_2 \leq \sum_{d_{t+2}} P(d_{t+2}|d_{t+1})\, V_{t+2}^{*}(CS_t + a1_t - 1, d_{t+2}) \tag{5.41}$$

$$- \sum_{d_{t+2}} P(d_{t+2}|d_{t+1})\, V_{t+2}^{*}(CS_t + a1_t, d_{t+2})$$

From equation 5.41, it implies that for unit increase in CS_t the last part equation 5.40 increasing in CS_t

Hence in region 4, $V_{t+1}{}^{*}(CS_t + a_t, d_{t+1})$ superadditive in $CS \times a$

Region 5:

$CS_{t+1} \geq L$, in both the functions $V_{t+1}{}^{*}(CS_t + a1_t, d_{t+1})$ and $V_{t+1}{}^{*}(CS_t + a2_t, d_{t+1})$

This implies:

$$L - a1_t \leq CS_t$$

Same as region 5 in possibility 1. Hence in region 5, $V_{t+1}{}^{*}(CS_t + a_t, d_{t+1})$ is superadditive in $CS \times a$

This completes the proof for possibility 2.

This proves that $V_{t+1}{}^{*}(CS_t + a_t, d_{t+1})$ is superadditive in $CS \times a$ in the given model if $V_{t+1}\big((CS_{t+1}, d_{t+1}), a_{t+1}\big)$ and $V_{t+2}{}^{*}(CS_{t+1} + a_{t+1}, d_{t+2})$ is superadditive in $CS \times a$.

We have thus shown by induction that $V_t\big((CS_t, d_t), a_t\big)$ is superadditive in $CS \times a \; \forall \; d, t$

Proposition 1:

For given d_t, $a_t(CS_t, d_t)$ is nonincreasing in CS_t.

Rewriting the equations for $V_t^*(CS_t, d_t)$, $a_t(CS_t, d_t)$ and $V_t\big((CS_t, d_t), a_t\big)$ given in the previous section 5.3.

$$V_t^*(CS_t, d_t) = \begin{matrix} min \\ a \in A \end{matrix} \left\{ R_t[(CS_t, d_t), a_t] + \sum_{d_{t+1}} P(d_{t+1}|d_t)\, V_{t+1}^*(CS_t + a_t, d_{t+1}) \right\}$$

$$a_t(CS_t, d_t) = \begin{matrix} arg \min a \\ a \in A \end{matrix} \left\{ R_t[(CS_t, d_t), a_t] + \sum_{d_{t+1}} P(d_{t+1}|d_t)\, V_{t+1}^*(CS_t + a_t, d_{t+1}) \right\}$$

$$V_t\big((CS_t, d_t), a_t\big) = R_t[(CS_t, d_t), a_t] + \sum_{d_{t+1}} P(d_{t+1}|d_t)\, V_{t+1}^{*}(CS_t + a_t, d_{t+1})$$

We use lemma 1, to reach the result of proposition 1. As per Lemma 1:

If $V_t\big((CS_t, d_t), a_t\big)$is superadditive in $CS \times a \,\forall\, d_t$, then Proposition 1 follows.

Theorem 3 proves that $V_t\big((CS_t, d_t), a_t\big)$ is superadditive in $CS \times a \,\forall\, d_t$.

From Proposition 1, the computational effort is greatly reduced.

For this problem definition if the demand takes values from $[0, x]$. The number of iterations required using proposition 1 is

$$T(3x + 1)(x + 1)$$

Search done in $(x + 1)$ actions at the first and last capacity level, then apply theorem 1,2 and proposition 1 to get the optimal action at the remaining capacity levels.

Here are additional structural properties of the defined problem.

Theorem 4: In the defined problem $V_t\big((CS_t, d_t), a_t\big)$ is discrete convex in a_t $\forall\, CS_t, d_t, t$.

Proof:

From equation 5.6:

$$V_t\big((CS_t, d_t), a_t\big) = R_t[(CS_t, d_t), a_t] + \sum_{d_{t+1}} P(d_{t+1}|d_t)\, V_{t+1}^{*}(CS_t + a_t, d_{t+1})$$

Expanding for $R_t[(CS_t, d_t), a_t]$ from equation 5.2, gives

$$V_t\big((CS_t, d_t), a_t\big) = c_1(a_t)^+ + c_2(-a_t)^+ + c_3(CS_t) + c_4 min(d_t, CS_t) + c_5(d_t - CS_t)^+ + \sum_{d_{t+1}} P(d_{t+1}|d_t)\, V_{t+1}^*(CS_t + a_t, d_{t+1})$$

$\forall\ a_t < 0$, this gives

$$V_t\big((CS_t, d_t), a_t + 1\big) - V_t\big((CS_t, d_t), a_t\big) = -c_2 + \sum_{d_{t+1}} P(d_{t+1}|d_t)\big(\, V_{t+1}^*(CS_t + a_t + 1, d_{t+1}) - \ V_{t+1}^*(CS_t + a_t, d_{t+1})\big)$$

And $\forall\ a \geq 0$, this gives

$$V_t\big((CS_t, d_t), a_t + 1\big) - V_t\big((CS_t, d_t), a_t\big) = c_1 + \sum_{d_{t+1}} P(d_{t+1}|d_t)\big(\, V_{t+1}^*(CS_t + a_t + 1, d_{t+1}) - \ V_{t+1}^*(CS_t + a_t, d_{t+1})\big)$$

Considering the assumption: $c_1 \geq max(0, -c_2)$; $V_t\big((CS_t, d_t), a_t + 1\big) - V_t\big((CS_t, d_t), a_t\big)$ is nondecreasing in a_t if

$$V_{t+1}^*(CS_t + a_t + 1, d_{t+1}) - \ V_{t+1}^*(CS_t + a_t, d_{t+1})$$

is nondecreasing in a_t

From the proof of theorem 3, it is concluded that ${V_{t+1}}^*(CS_t + a_t, d_{t+1})$ is superadditive in $CS\ \times\ a\ \forall\, d, t$

Then $\forall\ CS2, CS1\ in\ CS_{t-1}$ & $\forall a2, a1\ in\ a_{t-1}$ such that $CS2 \geq CS1$ and $a2 \geq a1$, it follows

$$V_t^*(CS2 + a2, d_t) - V_t^*(CS2 + a1, d_t) \geq V_t^*(CS1 + a2, d_t) - V_t^*(CS1 + a1, d_t)$$

Let $CS1 = CS, CS2 = CS + 1, a1 = a$ & $a2 = a + 1$. This gives

$$\begin{aligned} &V_t^*(CS + a + 2, d_t) - V_t^*(CS + a + 1, d_t) \\ &\qquad \geq V_t^*(CS + a + 1, d_t) - V_t^*(CS + a, d_t) \end{aligned} \tag{5.42}$$

Equation 5.42 implies that $V_{t+1}^*(CS + a + 1, d_{t+1}) - V_{t+1}^*(CS + a, d_{t+1})$ is nondecreasing in a. This completes the proof of theorem 4.

Theorem 4 will assist in decreasing the number of iterations required to get the optimum policy to at-least:

$$2Tx(x + 1)$$

For this problem definition if the demand takes values from $[0, x]$.

In case of large values of x, we can employ modified Golden section search method or modified Fibonacci search that applies to discrete functions to reduce the steps/ time to determine optimum policy.

We put forward theorem 5 for the purpose of proving the main result in proposition 2.

Theorem 5:

In the MDP model presented if the demand transition matrix exhibit first order stochastic dominance then $V_t\big((CS_t, d_t), a_t\big)$ is subadditive in $d \times a\ \forall\ CS_t, t$

That is for $d2_t \geq d1_t$ in d_t and $a2_t \geq a1_t$in a_t,

$$V_t\big((CS_t, d2_t), a2_t\big) - V_t\big((CS_t, d2_t), a1_t\big) \leq V_t\big((CS_t, d1_t), a2_t\big) - V_t\big((CS_t, d1_t), a1_t\big)$$

Proof:

The proof of this theorem is established using mathematical induction. This starts from the last time epoch T, where the capacity is to be salvaged.

From equation 5.3:

$$V_T\big((CS_T, d_T), a_T\big) = V_T^*(CS_T, d_T) = \ -c_6 CS_T \tag{5.43}$$

$V_T[(CS_T, d_T), a1_T] - V_T[(CS_T, d_T), a2_T] = 0$ for any d_T, and any combination of $a1_T$ and $a2_T$. Hence we have $V_T\big((CS_T, d_T), a_T\big)$ superadditive in $d \times a \ \forall \ CS_T$

Writing equation 5.6

$$V_t\big((CS_t, d_t), a_t\big) = R_t\big((CS_t, d_t), a_t\big) + \sum_{d_{t+1}} P(d_{t+1}|d_t)\ V_{t+1}^*(CS_t + a_t, d_{t+1})$$

Considering the elements of this equation:

$$R_t\big((CS_t, d_t), a_t\big) \ = c_1(a_t)^+ + c_2(-a_t)^+ + c_3(CS_t) + c_4 min(d_t, CS_t) + c_5(d_t - \ CS_t)^+$$

$R_t\big((CS_t, d_t), a_t\big)$ is subadditive in $d \times a \ \forall \ CS_t, t$ because:

$R_t\big((CS_t, d_t), a2_t\big) - R_t\big((CS_t, d_t), a1_t\big)$ is a constant independent of $d_t \ \forall \ a2_t \geq a1_t$.

Hence $V_t\big((CS_t, d_t), a_t\big)$ will be subadditive in $d \times a$, if $\sum_{d_{t+1}} P(d_{t+1}|d)\ V_{t+1}^*(CS_t + a_t, d_{t+1})$ is superadditive in $d \times a$.

Moreover, Since $P(d_{t+1}|d_t)$ shows first order stochastic dominance

$\sum_{d_{t+1}} P(d_{t+1}|d)\, V_{t+1}^*(CS_t + a_t, d_{t+1})$ will be subadditive in $d \times a$ if $V_{t+1}^*(CS_t + a_t, d_{t+1})$ is subadditive in $d \times a$.

In conclusion, $V_t\big((CS_t, d_t), a_t\big)$ will be subadditive in $d \times a \; \forall \; d_t$, if $V_{t+1}^*(CS_t + a_t, d_{t+1})$ is subadditive in $d \times a$.

This will be shown by induction:

Since $V_T{}^*(CS_T, d_T) = -c_6.\, CS_T$, $V_T{}^*(CS_{T-1} + a_{T-1}, d_T)$ is subadditive in $d \times a$, as for $a2_{T-1} \geq a1_{T-1}$

$$V_T{}^*(CS_{T-1} + a2_{T-1}, d_T) - V_T{}^*(CS_{T-1} + a1_{T-1}, d_T) = -c_6(a2_{T-1} - a1_{T-1})$$

is independent of d_T.

From equation 5.26

$$\begin{aligned} V_{T-1}\big((CS_{T-1}, d_{T-1}), a_{T-1}\big) & \\ &= c_1(a_{T-1})^+ + c_2(-a_{T-1})^+ + c_3(CS_{T-1}) \\ &+ c_4 min(d_{T-1}, CS_{T-1}) + c_5(d_{T-1} - CS_{T-1})^+ \\ &- c_6(CS_{T-1} + a_{T-1}) \end{aligned} \tag{5.44}$$

$V_{T-1}\big((CS_{T-1}, d_{T-1}), a_{T-1}\big)$ is superadditive in $d \times A$.

From equation 5.27

$$V^*_{T-1}(CS_{T-1}, d_{T-1}) = (c_3 - c_6)CS_{T-1} + c_4 min(d_{T-1}, CS_{T-1}) + c_5(d_{T-1} - CS_{T-1})^+ \text{ if } c_2 \geq -c_6 \ \&\ c_1 \geq c_6$$

$$V^*_{T-1}(CS_{T-1}, d_{T-1}) = (c_3 + c_2)CS_{T-1} + c_4 min(d_{T-1}, CS_{T-1}) + c_5(d_{T-1} - CS_{T-1})^+ \text{ if } c_2 \leq -c_6 \ \&\ c_1 \geq c_6 \tag{5.45}$$

$$V^*_{T-1}(CS_{T-1}, d_{T-1}) = (c_3 - c_1)CS_{T-1} + c_4 min(d_{T-1}, CS_{T-1}) + c_5(d_{T-1} - CS_{T-1})^+ + (c_1 - c_6)MC \text{ if } c_1 \leq c_6$$

Next, we have the Bellman optimality equation for $t = T - 2$, which depends on $V^*_{T-1}()$ as follows:

$$\begin{aligned} V_{T-2}\big((CS_{T-2}, d_{T-2}), a_{T-2}\big) \\ &= R_{T-2}[(CS_{T-2}, d_{T-2}), a_{T-2}] \\ &+ \sum_{d_{t+1}} P(d_{T-1}|d_{T-2})\, V^*_{T-1}(CS_{T-2} + a_{T-2}, d_{T-1}) \end{aligned}$$

As shown earlier, $V_{T-2}\big((CS_{T-2}, d_{T-2}), a_{T-2}\big)$ will be subadditive in $d \times a\ \forall\ CS$, if $V^*_{T-1}(CS_{T-2} + a_{T-2}, d_{T-1})$ is superadditive in $d \times a$.

Let $a2_{T-2}, a1_{T-2}$ be two actions such that $a2_{T-2} \geq a1_{T-2}$

From equation 5.45 we can simplify $V^*_{T-1}(CS_{T-2} + a2_{T-2}, d_{T-1}) - V^*_{T-1}(CS_{T-2} + a1_{T-2}, d_{T-1})$ by taking common terms as follows:

$$V^*_{T-1}(CS_{T-2} + a2_{T-2}, d_{T-1}) - V^*_{T-1}(CS_{T-2} + a1_{T-2}, d_{T-1}) = (c_3 - c_6)(a2_{T-2} - a1_{T-2}) + c_4\big(min(d_{T-1}, CS_{T-2} + a2_{T-2}) - \tag{5.46}$$

$min(d_{T-1}, CS_{T-2} + a1_{T-2})\big) + c_5((d_{T-1} - \ CS_{T-2} - a2_{T-2})^+ -$

$(d_{T-1} - \ CS_{T-2} - a1_{T-2})^+)$. if $c_2 \geq -c_6$ & $c_1 \geq c_6$

$V_{T-1}^*(CS_{T-2} + a2_{T-2}, d_{T-1}) - \ V_{T-1}^*(CS_{T-2} + a1_{T-2}, d_{T-1}) =$

$(c_3 + c_2)(a2_{T-2} - a1_{T-2}) + c_4\big(min(d_{T-1}, CS_{T-2} + a2_{T-2}) -$

$min(d_{T-1}, CS_{T-2} + a1_{T-2})\big) + c_5((d_{T-1} - \ CS_{T-2} - a2_{T-2})^+ -$

$(d_{T-1} - \ CS_{T-2} - a1_{T-2})^+)$. if $c_2 \leq -c_6$ & $c_1 \geq c_6$

$V_{T-1}^*(CS_{T-2} + a2_{T-2}, d_{T-1}) - \ V_{T-1}^*(CS_{T-2} + a1_{T-2}, d_{T-1}) =$

$(c_3 - c_1)(a2_{T-2} - a1_{T-2}) + c_4\big(min(d_{T-1}, CS_{T-2} + a2_{T-2}) -$

$min(d_{T-1}, CS_{T-2} + a1_{T-2})\big) + c_5((d_{T-1} - \ CS_{T-2} - a2_{T-2})^+ -$

$(d_{T-1} - \ CS_{T-2} - a1_{T-2})^+)$. if $c_1 \leq c_6$

The first parts in equation 5.46: $(c_3 - c_6)(a2_{T-2} - a1_{T-2}) \mid (c_3 + c_2)(a2_{T-2} - a1_{T-2})|(c_3 - c_1)(a2_{T-2} - a1_{T-2})$ are constants. Studying the second part, which is similar in all three possibilities of equation 5.48: $c_4\big(min(d_{T-1}, CS_{T-2} + a2_{T-2}) - min(d_{T-1}, CS_{T-2} + a1_{T-2})\big) + c_5((d_{T-1} - \ CS_{T-2} - a2_{T-2})^+ - (d_{T-1} - \ CS_{T-2} - a1_{T-2})^+)$ is nonincreasing in d_{T-1}. Which means $V_{T-1}^*(CS_{T-2} + a2_{T-2}, d_{T-1}) - V_{T-1}^*(CS_{T-2} + a1_{T-2}, d_{T-1})$ is decreasing in d_{T-1}. This gives $V_{T-2}\big((CS_{T-2}, d_{T-2}), a_{T-2}\big)$ subadditive in $d \times a$.

So far, the following results have been established:

1. $V_{T-2}\big((CS_{T-2}, d_{T-2}), a_{T-2}\big), V_{T-1}\big((CS_{T-1}, d_{T-1}), a_{T-1}\big), V_T\big((CS_T, d_T), a_T\big)$ is subadditive in $d \times a$.
2. $V_{T-1}^*(CS_{T-2} + a_{T-2}, d_{T-1}),\ V_T^*(CS_T + a_T, d_T)$ is subadditive in $d \times a$.

This gives the first step of the induction proof of theorem 5. Next assuming that $V_{t+1}\big((CS_{t+1}, d_{t+1}), a_{t+1}\big)$, and $V_{t+2}{}^{*}(CS_{t+1} + a_{t+1}, d_{t+2})$ are subadditive in $d \times a$ $\forall\ CS_{t+1}$, we need to prove that $V_{t+1}{}^{*}(CS_t + a_t, d_{t+1})$ is subadditive in $d \times a$ to give $V_t\big((CS_t, d_t), a_t\big)$ subadditive in $d \times a$.

That is to check if $V_{t+1}{}^{*}(CS_t + a2_t, d_{t+1}) - V_{t+1}{}^{*}(CS_t + a1_t, d_{t+1})$ is nonincreasing in d for $a2_t \geq a1_t \forall\ CS_t$

Let optimum action in time epoch $t + 1$ and state $(CS_{t+1} = 0, d_{t+1})$ be K.

$$a_{t+1}(0, d_{t+1}) = K$$

And optimum action in time epoch $t + 1$ and state $(CS_{t+1} = MC, d_{t+1})$ be $L - MC$

$$a_{t+1}(MC, d_{t+1}) = L - MC$$

From theorem 1 & 2 it is understood that $L \geq K$.

let $K > a2 > a1$ ($a2 > k > a1$ and $a2 > a1 > k$ will be special cases of this)

We have two possibilities:

Possibility 1: $K - a1_t \leq L - a2_t$

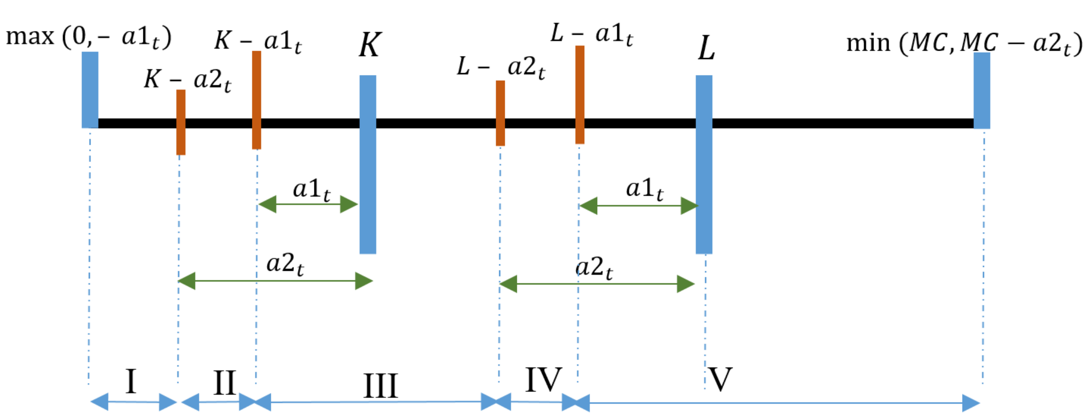

Figure 8: Illustration of theorem 5, possibility 1 - CS_t

As shown in the figure above for possibility 1, the resulting regions are considered:

Region 1:

$CS_{t+1} \le K$ in both functions $V_{t+1}{}^*(CS_t + a2_t, d_{t+1})$ & $V_{t+1}{}^*(CS_t + a1_t, d_{t+1})$

This implies:

$$\max(0, -a1_t) \le CS_t \le K - a2_t$$

From theorem 1 the optimal action will be to go to capacity K for both states $(CS_t + a2_t, d_{t+1})$ and $(CS_t + a1_t, d_{t+1})$, hence:

$$V_{t+1}{}^*(CS_t + a2_t, d_{t+1}) - V_{t+1}{}^*(CS_t + a1_t, d_{t+1})$$
$$= c_1(K - CS_t - a2_t) + c_3(CS_t + a2_t) + c_4 min(d_{t+1}, CS_t + a2)$$
$$+ c_5(d_{t+1} - CS_t - a2_t)^+$$
$$- (c_1(K - CS_t - a1_t) + c_3(CS_t + a1_t) + c_4 min(d_{t+1}, CS_t + a1_t)$$
$$+ c_5(d_{t+1} - CS_t - a1_t)^+)$$

Simplifying

$$V_{t+1}{}^*(CS_t + a2_t, d_{t+1}) - V_{t+1}{}^*(CS_t + a1_t, d_{t+1})$$
$$= (c_3 - c_1)(a2_t - a1_t)$$
$$+ c_4\big(min(d_{t+1}, CS_t + a2_t) - min(d_{t+1}, CS_t + a1_t)\big)$$
$$+ c_5((d_{t+1} - CS_t - a2_t)^+ - (d_{t+1} - CS_t - a1_t)^+)$$

Lemma 3 shows that $c_4\big(min(d_{t+1}, CS_t + a2) - min(d_{t+1}, CS_t + a1_t)\big) +$ $c_5((d_{t+1} - CS_t - a2_t)^+ - (d_{t+1} - CS_t - a1_t)^+)$ is nonincreasing in d_{t+1}.

Hence in region 1, $V_{t+1}{}^{*}(CS_t + a_t, d_{t+1})$ subadditive in $d \times a$

Region 2:

$CS_{t+1} \leq K$ in the function $V_{t+1}{}^{*}(CS_t + a1_t, d_{t+1})$ and $L \geq CS_{t+1} \geq K$ in the function $V_{t+1}{}^{*}(CS_t + a2_t, d_{t+1})$

This implies:

$$K - a2_t \leq CS_t \leq K - a1_t$$

From theorem 1and theorem 2, the optimal action will be to go to capacity K for states $(CS_t + a1_t, d_{t+1})$ and to take action $a = 0$ for states $(CS_t + a2_t, d_{t+1})$

$$\begin{aligned}
&V_{t+1}{}^{*}(CS_t + a2_t, d_{t+1}) - V_{t+1}{}^{*}(CS_t + a1_t, d_{t+1}) \\
&\qquad = c_3(CS_t + a2_t) + c_4 min(d_{t+1}, CS_t + a2_t) + c_5(d_{t+1} - CS_t - a2_t)^{+} \\
&\qquad + \sum_{d_{t+2}} P(d_{t+2}|d_{t+1})\, V_{t+2}^{*}(CS_t + a2_t, d_{t+2}) \\
&\qquad - \left(c_1(K - CS_t - a1_t) + c_3(CS_t + a1_t) + c_4 min(d_{t+1}, CS_t + a1_t) \right. \\
&\qquad \left. + c_5(d_{t+1} - CS_t - a1_t)^{+} + \sum_{d_{t+2}} P(d_{t+2}|d_{t+1})\, V_{t+2}^{*}(K, d_{t+2}) \right)
\end{aligned}$$

Simplifying

$$
\begin{aligned}
V_{t+1}{}^{*}(CS+a2,d)-V_{t+1}{}^{*}(CS+a1,d) \\
&= c_3(a2_t-a1_t) \\
&+ c_4\big(min(d_{t+1},CS_t+a2_t)-min(d_{t+1},CS_t+a1_t)\big) \\
&+ c_5((d_{t+1}-\ CS_t-a2_t)^{+}-(d_{t+1}-\ CS_t-a1_t)^{+}) \\
&- c_1(K-CS_t-a1_t) \\
&+ \sum_{d_{t+2}} P(d_{t+2}|d_{t+1})\big(\ V_{t+2}^{*}(CS_t+a2_t,d_{t+2}) \\
&-\ V_{t+2}^{*}(K,d_{t+2})\big)
\end{aligned}
\tag{5.47}
$$

Studying equation 5.47 in parts:

1. $c_3(a2_t-a1_t)\ -c_1(K-a1-CS_t)$ is fixed in this region
2. From Lemma 3, $c_4\big(min(d_{t+1},CS_t+a2_t)-min(d_{t+1},CS_t+a1_t)\big)+$ $c_5((d_{t+1}-\ CS_t-a2_t)^{+}-(d_{t+1}-\ CS_t-a1_t)^{+})$ is nonincreasing in d_{t+1}.
3. The last part:

$$\sum_{d_{t+2}} P(d_{t+2}|d_{t+1})\ V_{t+2}^{*}(CS_t+a2_t,d_{t+2})-\sum_{d_{t+2}} P(d_{t+2}|d_{t+1})\ V_{t+2}^{*}(K,d_{t+2})$$

Is nonincreasing in d_{t+1} because $P(d_{t+2}|d_{t+1})$ shows first order stochastic dominance and $V_{t+2}{}^{*}(CS_{t+1}+a_{t+1},d_{t+2})$ is subadditive in $d\ \times\ a$

Hence in region 2, $V_{t+1}{}^{*}(CS_t+a_t,d_{t+1})$ subadditive in $d\ \times\ a$

Region 3:

$L\geq CS_{t+1}\geq K$ in both the functions $V_{t+1}{}^{*}(CS_t+a2_t,d_{t+1})$ and $V_{t+1}{}^{*}(CS_t+a1_t,d_{t+1})$

This implies:

$$K - a1_t \leq CS_t \leq L - a2_t$$

$$V_{t+1}{}^*(CS_t + a2_t, d_{t+1}) - V_{t+1}{}^*(CS_t + a1_t, d_{t+1})$$

$$= c_3(CS_t + a2_t) + c_4 min(d_{t+1}, CS_t + a2_t) + c_5(d_{t+1} - CS_t - a2_t)^+$$

$$+ \sum_{d_{t+2}} P(d_{t+2}|d_{t+1})\, V^*_{t+2}(CS_t + a2_t, d_{t+2})$$

$$-\left(c_3(CS_t + a1_t) + c_4 min(d_{t+1}, CS_t + a1_t) + c_5(d_{t+1} - CS_t - a1_t)^+ \right.$$

$$\left. + \sum_{d_{t+2}} P(d_{t+2}|d_{t+1})\, V^*_{t+2}(CS_t + a1_t, d_{t+2}) \right)$$

Simplifying

$$V_{t+1}{}^*(CS_t + a2_t, d_{t+1}) - V_{t+1}{}^*(CS_t + a1_t, d_{t+1})$$

$$= c_3(a2_t - a1_t) + c_4\big(min(d_{t+1}, CS_t + a2_t) - min(d_{t+1}, CS_t + a1_t)\big)$$

$$+ c_5((d_{t+1} - CS_t - a2_t)^+ - (d_{t+1} - CS_t - a1_t)^+)$$

$$+ \sum_{d_{t+2}} P(d_{t+2}|d_{t+1})\big(V^*_{t+2}(CS_t + a2_t, d_{t+2}) - V^*_{t+2}(CS_t + a1_t, d_{t+2})\big)$$

$c_3(a2_t - a1_t)$ is a fixed value. Lemma 3 shows that $c_4\big(min(d_{t+1}, CS_t + a2_t) - min(d_{t+1}, CS_t + a1_t)\big) + c_5((d_{t+1} - CS_t - a2_t)^+ - (d_{t+1} - CS_t - a1_t)^+)$ is nonincreasing in d_{t+1}. $V^*_{t+2}(CS_t + a2_t, d_{t+2}) - V^*_{t+2}(CS_t + a1_t, d_{t+2})$ is given to be nonincreasing in d_{t+2}, because $V^*_{t+2}(CS_t + a_t, d_{t+2})$ is subadditive in $d \times a$.

Hence in region 3, $V_{t+1}{}^*(CS_t + a_t, d_{t+1})$ subadditive in $d \times a$

Region 4:

$L \geq CS_{t+1} \geq K$ in the function $V_{t+1}{}^{*}(CS_t + a1_t, d_{t+1})$ and $CS_{t+1} \geq L$ in the function $V_{t+1}{}^{*}(CS_t + a2_t, d_{t+1})$

This implies:

$$L - a2_t \leq CS_t \leq L - a1_t$$

$$\begin{aligned}
V_{t+1}{}^{*}(CS_t + a2_t, d_{t+1}) - V_{t+1}{}^{*}(CS_t + a1_t, d_{t+1}) \\
&= c_2(CS_t + a2_t - L) + c_3(CS_t + a2_t) + c_4 min(d_{t+1}, CS_t + a2_t) \\
&+ c_5(d_{t+1} - \ CS_t - a2_t)^{+} + \sum_{d_{t+2}} P(d_{t+2}|d_{t+1})\, V_{t+2}^{*}(L, d_{t+2}) \\
&- \Bigg(c_3(CS_t + a1_t) + c_4 min(d_{t+1}, CS_t + a1_t) + c_5(d_{t+1} - \ CS_t - a1_t)^{+} \\
&+ \sum_{d_{t+2}} P(d_{t+2}|d_{t+1})\, V_{t+2}^{*}(CS_t + a1_t, d_{t+2}) \Bigg)
\end{aligned}$$

Simplifying:

$$V_{t+1}{}^{*}(CS_t + a2_t, d_{t+1}) - V_{t+1}{}^{*}(CS_t + a1_t, d_{t+1}) \tag{5.48}$$

$$= c_2(CS_t + a2_t - L) + c_3(a2_t - a1_t)$$

$$+ c_4\big(min(d_{t+1}, CS_t + a2_t) - min(d_{t+1}, CS_t + a1_t)\big)$$

$$+ c_5((d_{t+1} - CS_t - a2_t)^+ - (d_{t+1} - CS_t - a1_t)^+)$$

$$+ \sum_{d_{t+2}} P(d_{t+2}|d_{t+1})\big(V_{t+2}^{*}(L, d_{t+2})$$

$$- V_{t+2}^{*}(CS_t + a1_t, d_{t+2})\big)$$

Studying equation 5.48 in parts:

1. $c_3(a2_t - a1_t) + c_2(CS_t + a2_t - L)$ is fixed in this region
2. Lemma 3 shows that $c_4\big(min(d_{t+1}, CS_t + a2_t) - min(d_{t+1}, CS_t + a1_t)\big) + c_5((d_{t+1} - CS_t - a2_t)^+ - (d_{t+1} - CS_t - a1_t)^+)$ is nonincreasing in d_{t+1}.
3. The last part:

$$\sum_{d_{t+2}} P(d_{t+2}|d_{t+1})\, V_{t+2}^{*}(L, d_{t+2}) - \sum_{d_{t+2}} P(d_{t+2}|d_{t+1})\, V_{t+2}^{*}(CS_t + a1_t, d_{t+2})$$

   Is nonincreasing in d_{t+1} because $P(d_{t+2}|d_{t+1})$ shows first order stochastic dominance and $V_{t+2}{}^{*}(CS_{t+1} + a_{t+1}, d_{t+2})$ is subadditive in $d \times a$

Hence in region 4, $V_{t+1}{}^{*}(CS_t + a_t, d_{t+1})$ subadditive in $d \times a$

Region 5:

$CS_{t+1} \geq L$ in both the functions $V_{t+1}{}^{*}(CS_t + a1_t, d_{t+1})$ and $V_{t+1}{}^{*}(CS_t + a2_t, d_{t+1})$

This implies:

$$L - a1_t \leq CS_t$$

$$V_{t+1}{}^{*}(CS_t + a2_t, d_{t+1}) - V_{t+1}{}^{*}(CS_t + a1_t, d_{t+1})$$

$$= c_2(CS_t + a2_t - L) + c_3(CS_t + a2_t) + c_4 min(d_{t+1}, CS_t + a2_t)$$

$$+ c_5(d_{t+1} - CS_t - a2_t)^{+} + \sum_{d_{t+2}} P(d_{t+2}|d_{t+1})\, V_{t+2}^{*}(L, d_{t+2})$$

$$-\Bigg(c_2(CS + a1 - L) + c_3(CS_t + a1_t) + c_4 min(d_{t+1}, CS_t + a1_t)$$

$$+ c_5(d_{t+1} - CS_t - a1_t)^{+} + \sum_{d_{t+2}} P(d_{t+2}|d_{t+1})\, V_{t+2}^{*}(CS_t + a1_t, d_{t+2})\Bigg)$$

$$V_{t+1}{}^{*}(CS_t + a2_t, d_{t+1}) - V_{t+1}{}^{*}(CS_t + a1_t, d_{t+1})$$

$$= (c_2 + c_3)(a2_t - a1_t)$$

$$+ c_4\big(min(d_{t+1}, CS_t + a2_t) - min(d_{t+1}, CS_t + a1_t)\big)$$

$$+ c_5((d_{t+1} - CS_t - a2_t)^{+} - (d_{t+1} - CS_t - a1_t)^{+})$$

$(c_2 + c_3)(a2_t - a1_t)$ constant with increase in CS_t. Lemma 3 shows that $c_4\big(min(d_{t+1}, CS_t + a2_t) - min(d_{t+1}, CS_t + a1_t)\big) + c_5((d_{t+1} - CS_t - a2_t)^{+} - (d_{t+1} - CS_t - a1_t)^{+})$ is nonincreasing in d_{t+1}. Hence in region 5, $V_{t+1}{}^{*}(CS_t + a_t, d_{t+1})$ subadditive in $d \times a$

This completes the proof for possibility 1.

Possibility 2: $K - a1_t > L - a2_t$

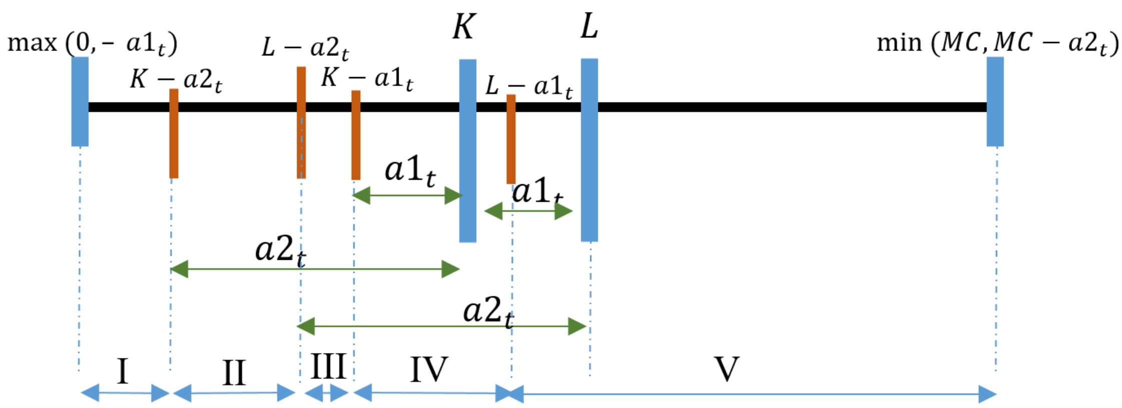

Figure 9: Illustration of theorem 5, possibility 2 - CS_t

As shown in the figure above for possibility 1, the resulting regions are considered:

Region 1:

$CS_{t+1} \leq K$ in both functions $V_{t+1}{}^{*}(CS_t + a2_t, d_{t+1})$ & $V_{t+1}{}^{*}(CS_t + a1_t, d_{t+1})$

This implies:

$$\max(0, -a1_t) \leq CS_t \leq K - a2_t$$

Same as region 1 for possibility 1

Hence in region 1, $V_{t+1}{}^{*}(CS_t + a_t, d_{t+1})$ subadditive in $d \times a$

Region 2:

$CS_{t+1} \leq K$ in the function $V_{t+1}{}^{*}(CS_t + a1_t, d_{t+1})$ and $L \geq CS_{t+1} \geq K$ in the function $V_{t+1}{}^{*}(CS_t + a2_t, d_{t+1})$

This implies:

$$K - a2_t \leq CS_t \leq L - a1_t$$

From theorem 1and theorem 2, the optimal action will be to go to capacity K for states $(CS_t + a1_t, d_{t+1})$ and to take action $a = 0$ for states $(CS_t + a2_t, d_{t+1})$

$$
\begin{aligned}
&{V_{t+1}}^*(CS_t + a2_t, d_{t+1}) - {V_{t+1}}^*(CS_t + a1_t, d_{t+1}) \\
&\quad = c_3(CS_t + a2_t) + c_4 min(d_{t+1}, CS_t + a2) + c_5(d_{t+1} - \; CS_t - a2_t)^+ \\
&\quad + \sum_{d_{t+2}} P(d_{t+2}|d_{t+1})\, V_{t+2}^*(CS_t + a2_t, d_{t+2}) \\
&\quad - \Bigg(c_1(K - CS_t - a1_t) + c_3(CS_t + a1_t) + c_4 min(d_{t+1}, CS_t + a1_t) \\
&\quad + c_5(d_{t+1} - \; CS_t - a1_t)^+ + \sum_{d_{t+2}} P(d_{t+2}|d_{t+1})\, V_{t+2}^*(K, d_{t+2}) \Bigg)
\end{aligned}
$$

Simplifying

$$
\begin{aligned}
&{V_{t+1}}^*(CS_t + a2_t, d_{t+1}) - {V_{t+1}}^*(CS_t + a1_t, d_{t+1}) \\
&\quad = c_3(a2_t - a1_t) \\
&\quad + c_4\big(min(d_{t+1}, CS_t + a2_t) - min(d_{t+1}, CS_t + a1_t)\big) \\
&\quad + c_5((d_{t+1} - \; CS_t - a2_t)^+ - (d_{t+1} - \; CS_t - a1_t)^+) \\
&\quad - c_1(K - CS_t - a1_t) \\
&\quad + \sum_{d_{t+2}} P(d_{t+2}|d_{t+1})\big(\, V_{t+2}^*(CS_t + a2_t, d_{t+2}) \\
&\quad - \; V_{t+2}^*(K, d_{t+2})\big)
\end{aligned}
\tag{5.49}
$$

Studying equation 5.49 in parts:

1. $c_3(a2_t - a1_t) - \sum_{d_{t+2}} P(d_{t+2}|d_{t+1})\, V^*_{t+2}(K, d_{t+2}) - c_1(CS_t + K - a1_t)$ is fixed in this region
2. Lemma 3 shows that $c_4\big(min(d_{t+1}, CS_t + a2_t) - min(d_{t+1}, CS_t + a1_t)\big) +$ $c_5((d_{t+1} - CS_t - a2_t)^+ - (d_{t+1} - CS_t - a1_t)^+)$ is nonincreasing in d_{t+1}
3. The last part:

$$\sum_{d_{t+2}} P(d_{t+2}|d_{t+1})\, V^*_{t+2}(CS_t + a2_t, d_{t+2}) - \sum_{d_{t+2}} P(d_{t+2}|d_{t+1})\, V^*_{t+2}(K, d_{t+2})$$

Is nonincreasing in d_{t+1} because $P(d_{t+2}|d_{t+1})$ shows first order stochastic dominance and ${V_{t+2}}^*(CS_{t+1} + a_{t+1}, d_{t+2})$ is subadditive in $d \times a$

Hence in region 2, ${V_{t+1}}^*(CS_t + a_t, d_{t+1})$ subadditive in $d \times a$

Region 3:

$CS_{t+1} \leq K$ in the function ${V_{t+1}}^*(CS_t + a1_t, d_{t+1})$ and $CS_{t+1} \geq L$ in the function ${V_{t+1}}^*(CS_t + a2_t, d_{t+1})$

This implies:

$$L - a2_t \leq CS_t \leq K - a1_t$$

$$V_{t+1}{}^{*}(CS_t + a2_t, d_{t+1}) - V_{t+1}{}^{*}(CS_t + a1_t, d_{t+1})$$

$$= c_2(CS_t + a2_t - L) + c_3(CS_t + a2_t) + c_4 min(d_{t+1}, CS_t + a2)$$

$$+ c_5(d_{t+1} - CS_t - a2_t)^{+} + \sum_{d_{t+2}} P(d_{t+2}|d_{t+1})\, V_{t+2}^{*}(L, d_{t+2})$$

$$-\Bigg(c_1(K - CS_t - a1_t) + c_3(CS_t + a1_t) + c_4 min(d_{t+1}, CS_t + a1_t)$$

$$+ c_5(d_{t+1} - CS_t - a1_t)^{+} + \sum_{d_{t+2}} P(d_{t+2}|d_{t+1})\, V_{t+2}^{*}(K, d_{t+2})\Bigg)$$

Simplifying

$$V_{t+1}{}^{*}(CS_t + a2_t, d_{t+1}) - V_{t+1}{}^{*}(CS_t + a1_t, d_{t+1}) \tag{5.50}$$

$$= -c_1(K - CS_t - a1_t) + c_2(CS_t + a2_t - L)$$

$$+ c_3(a2_t - a1_t)$$

$$+ c_4\big(min(d_{t+1}, CS_t + a2_t) - min(d_{t+1}, CS_t + a1_t)\big)$$

$$+ c_5((d_{t+1} - CS_t - a2_t)^{+} - (d_{t+1} - CS_t - a1_t)^{+})$$

$$+ \sum_{d_{t+2}} P(d_{t+2}|d_{t+1})\big(V_{t+2}^{*}(L, d_{t+2}) - V_{t+2}^{*}(K, d_{t+2})\big)$$

Studying equation 5.50 in parts:

1. $-c_1(K - a1_t) + c_2(a2_t - L) + c_3(a2_t - a1_t) + (c_1 + c_2)CS_t$ is a fixed value in this region
2. Lemma 3 shows that $c_4\big(min(d_{t+1}, CS_t + a2_t) - min(d_{t+1}, CS_t + a1_t)\big) + c_5((d_{t+1} - CS_t - a2_t)^{+} - (d_{t+1} - CS_t - a1_t)^{+})$ is nonincreasing in d_{t+1}.

3. $\sum_{d_{t+2}} P(d_{t+2}|d_{t+1})\left(V_{t+2}^{*}(L, d_{t+2}) - V_{t+2}^{*}(K, d_{t+2})\right)$ is nonincreasing with d_{t+1}because $P(d_{t+2}|d_{t+1})$ shows first order stochastic dominance and ${V_{t+2}}^{*}(CS_{t+1} + a_{t+1}, d_{t+2})$ is subadditive in $d \times a$

Hence in region 3, ${V_{t+1}}^{*}(CS_t + a_t, d_{t+1})$ subadditive in $d \times a$

Region 4:

$L \geq CS_{t+1} \geq K$ in the function ${V_{t+1}}^{*}(CS_t + a1_t, d_{t+1})$ and $CS_{t+1} \geq L$ in the function ${V_{t+1}}^{*}(CS_t + a2_t, d_{t+1})$

This implies:

$$K - a1_t \leq CS_t \leq L - a1_t$$

$$\begin{aligned}
&{V_{t+1}}^{*}(CS_t + a2_t, d_{t+1}) - {V_{t+1}}^{*}(CS_t + a1_t, d_{t+1}) \\
&\qquad = c_2(CS_t + a2_t - L) + c_3(CS_t + a2_t) + c_4 min(d_{t+1}, CS_t + a2) \\
&\qquad + c_5(d_{t+1} - CS_t - a2_t)^{+} + \sum_{d_{t+2}} P(d_{t+1}|d_{t+1})\, V_{t+2}^{*}(L, d_{t+2}) \\
&\qquad - \left(c_3(CS_t + a1_t) + c_4 min(d_{t+1}, CS_t + a1_t) + c_5(d_{t+1} - CS_t - a1_t)^{+} \right. \\
&\qquad \left. + \sum_{d_{t+2}} P(d_{t+2}|d_{t+1})\, V_{t+2}^{*}(CS_t + a1_t, d_{t+2}) \right)
\end{aligned}$$

Simplifying:

$$V_{t+1}{}^{*}(CS_t + a2_t, d_{t+1}) - V_{t+1}{}^{*}(CS_t + a1_t, d_{t+1}) \quad (5.51)$$

$$= c_2(CS_t + a2_t - L) + c_3(a2_t - a1_t)$$

$$+ c_4\big(min(d_{t+1}, CS_t + a2_t) - min(d_{t+1}, CS_t + a1_t)\big)$$

$$+ c_5((d_{t+1} - CS_t - a2_t)^+ - (d_{t+1} - CS_t - a1_t)^+)$$

$$+ \sum_{d_{t+2}} P(d_{t+2}|d_{t+1})\big(V^*_{t+2}(L, d_{t+2})$$

$$- V^*_{t+2}(CS_t + a1_t, d_{t+2})\big)$$

Studying equation 5.51 in parts:

1. $c_3(a2_t - a1_t) + c_2(CS_t + a2_t - L)$ is fixed in this region
2. Lemma 3 shows that $c_4\big(min(d_{t+1}, CS_t + a2_t) - min(d_{t+1}, CS_t + a1_t)\big) + c_5((d_{t+1} - CS_t - a2_t)^+ - (d_{t+1} - CS_t - a1_t)^+)$ is nonincreasing in d_{t+1}.
3. The last part:

   $\sum_{d_{t+2}} P(d_{t+2}|d_{t+1})\, V^*_{t+2}(L, d_{t+2}) - \sum_{d_{t+2}} P(d_{t+2}|d_{t+1})\, V^*_{t+2}(CS_t + a1_t, d_{t+2})$

   is nonincreasing with d_{t+1}because $P(d_{t+2}|d_{t+1})$ shows first order stochastic dominance and $V_{t+2}{}^{*}(CS_{t+1} + a_{t+1}, d_{t+2})$ is subadditive in $d \times a$

Hence in region 4, $V_{t+1}{}^{*}(CS_t + a_t, d_{t+1})$ is subadditive in $d \times a$

Region 5:

$CS_{t+1} \geq L$, in both the functions $V_{t+1}{}^{*}(CS_t + a1_t, d_{t+1})$ and $V_{t+1}{}^{*}(CS_t + a2_t, d_{t+1})$

This implies:

$$L - a1_t \leq CS_t$$

This is same as region 5 in possibility 1.

Hence in region 5, ${V_{t+1}}^*(CS_t + a_t, d_{t+1})$ subadditive in $d \times a$. This completes the case of possibility 2.

This gives that if $V_{t+1}\big((CS_{t+1}, d_{t+1}), a_{t+1}\big)$, and ${V_{t+2}}^*(CS_{t+1} + a_{t+1}, d_{t+2})$ are subadditive in $d \times a \; \forall \; CS_{t+1},$ then ${V_{t+1}}^*(CS_t + a_t, d_{t+1})$ is subadditive in $d \times a$ to give $V_t\big((CS_t, d_t), a_t\big)$ subadditive in $d \times a$.

This completes the proof for Theorem 5.

Proposition 2: If the demand transition follows first order stochastic dominance, then the optimum action is nondecreasing with increase in d.

That is, for given CS_t, $a_t(CS_t, d_t)$ is nondecreasing in d_t. If demand transition follows first order stochastic dominance.

That is $\forall \; d1, d2$ in d_t such that $d2 \leq d1$, if $P(d_{t+1} > d'|d2) \geq P(d_{t+1} > d'|d1) \; \forall \; d', t$

Then $a_t(CS_t, d2) \geq a_t(CS_t, d1) \; \forall \; CS_t, t$

By lemma 1, Proposition 2 follows if $V_t\big((CS_t, d_t), a_t\big)$is subadditive in $d \times a \; \forall \; CS_t$.

Theorem 5 proves that $V_t\big((CS_t, d_t), a_t\big)$is subadditive in $d \times a \; \forall \; CS_t$

5.6 Modified value iteration algorithm

This section gives the modified value iteration that will greatly reduce the computational efforts compared to standard backward value iteration.

5.6.1 Demand transition matrix does not show first order stochastic dominance

Step 1: $t = T$, calculate all $V_T^*(CS_T, d_T)$

Step 2: Reduce t by one unit

Step 3: $d_t = 0$

Step 4: Determine $a_t(CS_t = 0, d_t)$ and $V_t^*(CS_t = 0, d_t)$

Step 4.1: Set $x = 0$ and calculate $V_t\big((CS_t = 0, d_t), x\big)$

Step 4.2: Increase x by one unit.

If $x \leq MC$, Goto step 4.3

Else $a_t(CS_t = 0, d_t) = x - 1$ & $V_t^*(0, d_t) = V_t\big((0, d_t), x - 1\big)$. Step 5

Step 4.3: If $V_t\big((CS_t = 0, d_t), x\big) < V_t\big((CS_t = 0, d_t), x - 1\big)$, Goto step 4.2.

Else $a_t(CS_t = 0, d_t) = x - 1$ & $V_t^*(0, d_t) = V_t\big((0, d_t), x - 1\big)$. Step 5

Step 5: Determine $a_t(CS_t, d_t)$ and $V_t^*(CS_t, d_t)$ $\forall\ 0 < CS_t \leq x - 1$

$a_t(CS_t, d_t) = x - 1 - CS_t$ & $V_t^*(CS_t, d_t) = V_t\big((CS_t, d_t), x - 1 - CS_t\big)$

Step 6: Determine $a_t(CS_t = MC, d_t)$ and $V_t^*(CS_t = MC, d_t)$

Step 6.1: Set $y = MC - x - 1$ and calculate $V_t\big((CS_t = 0, d_t), y\big)$

Step 6.2: Increase y by one unit.

If $y \leq 0$, Goto step 6.3

Else $a_t(MC, d_t) = y - 1$ & $V_t^*(MC, d_t) = V_t\big((MC, d_t), y - 1\big)$. Step 7

Step 6.3: If $V_t\big((CS_t = 0, d_t), y\big) < V_t\big((CS_t = 0, d_t), y - 1\big)$, Goto step 6.2.

Else $a_t(CS_t = 0, d_t) = y - 1$ & $V_t^*(0, d_t) = V_t\big((0, d_t), y - 1\big)$. Step 7

Step 7: Determine $a_t(CS_t, d_t)$ and $V_t^*(CS_t, d_t)$ $\forall$ $MC > CS_t \geq MC + y - 1$

$$a_t(CS_t, d_t) = MC + y - 1 - CS_t \text{ \& } V_t^*(CS_t, d_t) = V_t\big((CS_t, d_t), MC + y - 1 - CS_t\big)$$

Step 8: Increase d_t by one unit.

If d_t $\leq$maximum demand, goto step 4

Else goto step 9

Step 9: If $t > 1$, goto step 2

Else END.

5.6.2 Demand transition matrix show first order stochastic dominance

Step 1: $t = T$, calculate all $V_T^*(CS_T, d_T)$

Step 2: Reduce t by one unit

Step 3: $d_t = 0$

Step 4: Determine $a_t(CS_t = 0, d_t)$ and $V_t^*(CS_t = 0, d_t)$

Step 4.1: If $d_t = 0$, Set $x = 0$ and calculate $V_t\big((CS_t = 0, d_t), x\big)$

Else, Set $x = a_t(CS_t = 0, d_t - 1)$

Step 4.2: Increase x by one unit.

If $x \leq MC$, Goto step 4.3

Else $a_t(CS_t = 0, d_t) = x - 1$ & $V_t^*(0, d_t) = V_t\big((0, d_t), x - 1\big)$. Step 5

Step 4.3: If $V_t\big((CS_t = 0, d_t), x\big) < V_t\big((CS_t = 0, d_t), x - 1\big)$, Goto step 4.2.

Else $a_t(CS_t = 0, d_t) = x - 1$ & $V_t^*(0, d_t) = V_t\big((0, d_t), x - 1\big)$. Step 5

Step 5: Determine $a_t(CS_t, d_t)$ and $V_t^*(CS_t, d_t)$ $\forall\ 0 < CS_t \leq x - 1$

$a_t(CS_t, d_t) = x - 1 - CS_t$ & $V_t^*(CS_t, d_t) = V_t\big((CS_t, d_t), x - 1 - CS_t\big)$

Step 6: Determine $a_t(CS_t = MC, d_t)$ and $V_t^*(CS_t = MC, d_t)$

Step 6.1: Set $y = MC - x - 1$ and calculate $V_t\big((CS_t = 0, d_t), y\big)$

Step 6.2: Increase y by one unit.

If $y \leq 0$, Goto step 6.3

Else $a_t(MC, d_t) = y - 1$ & $V_t^*(MC, d_t) = V_t\big((MC, d_t), y - 1\big)$. Step 7

Step 6.3: If $V_t\big((CS_t = 0, d_t), y\big) < V_t\big((CS_t = 0, d_t), y - 1\big)$, Goto step 6.2.

Else $a_t(CS_t = 0, d_t) = y - 1$ & $V_t^*(0, d_t) = V_t\big((0, d_t), y - 1\big)$. Step 7

Step 7: Determine $a_t(CS_t, d_t)$ and $V_t^*(CS_t, d_t)$ $\forall\ MC > CS_t \geq MC + y - 1$

$a_t(CS_t, d_t) = MC + y - 1 - CS_t$ & $V_t^*(CS_t, d_t) = V_t\big((CS_t, d_t), MC + y - 1 - CS_t\big)$

Step 8: Increase d_t by one unit.

If $d_t \leq$maximum demand, goto step 4

Else goto step 9

Step 9: If $t > 1$, goto step 2

Else END.

Note: in case of large action spaces we can use searches applicable to discrete convex functions that will reach the optimum decision action faster in steps 4 and 6.

5.7 Numerical Example

Problem 1:

Problem parameters

c_1= 13; $c_2 = 4$; c_3= 12; c_4= 25; c_5= 65; c_6= 3; T=10; MC=10

Demand transition matrix $P(d_{t+1}|d_t)$

	$d_{t+1} = 0$	1	2	3	4	5	6	7	8	9	10
$d_t = 0$	.2	.1	.3		.1						.3
1	.1	.2	.2	.4	.1						
2	.1	.1	.2		.1	.1			.4		
3	.1	.1	.2		.2	.1					.3
4	.1	.1			.2	.2				.3	.1
5	.1			.2	.2	.2	.1			.2	
6		.1	.1	.2	.1	.2	.1	.1	.1		
7			.1		.1	.1	.1	.1	.3	.2	
8				.1		.1	.1	.1	.2	.2	.2
9						.1	.1	.2	.2	.2	.2
10					.1	.1	.1	.1	.2	.2	.2

Optimum solution

$$a_1(CS,d) = a_2(CS,d) = a_3(CS,d)$$

	$CS = 0$	1	2	3	4	5	6	7	8	9	10
$d = 0$	4	3	2	1	0	0	0	0	0	0	0
1	3	2	1	0	0	0	0	0	0	−1	−2
2	8	7	6	5	4	3	2	1	0	−1	−2
3	5	4	3	2	1	0	0	0	0	0	0
4	8	7	6	5	4	3	2	1	0	0	−1
5	5	4	3	2	1	0	0	0	0	0	−1
6	5	4	3	2	1	0	0	0	0	−1	−2
7	8	7	6	5	4	3	2	1	0	0	−1
8	8	7	6	5	4	3	2	1	0	0	−1
9	8	7	6	5	4	3	2	1	0	0	−1
10	8	7	6	5	4	3	2	1	0	0	−1

Table 3: Problem 1, optimal action at t=1,2,3

$$a_4(CS,d) =$$

	$CS = 0$	1	2	3	4	5	6	7	8	9	10
$d = 0$	4	3	2	1	0	0	0	0	0	0	0
1	3	2	1	0	0	0	0	−1	−2	−3	−4
2	8	7	6	5	4	3	2	1	0	−1	−2
3	5	4	3	2	1	0	0	0	0	0	0
4	8	7	6	5	4	3	2	1	0	0	−1
5	5	4	3	2	1	0	0	0	0	0	−1
6	5	4	3	2	1	0	0	0	0	−1	−2
7	8	7	6	5	4	3	2	1	0	0	−1
8	8	7	6	5	4	3	2	1	0	0	−1
9	8	7	6	5	4	3	2	1	0	0	−1
10	8	7	6	5	4	3	2	1	0	0	−1

Table 4: Problem 1, optimal action at t=4

$$a_5(CS, d) =$$

	$CS = 0$	1	2	3	4	5	6	7	8	9	10
$d = 0$	4	3	2	1	0	0	0	0	0	0	0
1	3	2	1	0	0	0	−1	−2	−3	−4	−5
2	8	7	6	5	4	3	2	1	0	−1	−2
3	5	4	3	2	1	0	0	0	0	0	0
4	6	5	4	3	2	1	0	0	0	0	−1
5	5	4	3	2	1	0	0	0	0	0	−1
6	5	4	3	2	1	0	0	0	0	−1	−2
7	8	7	6	5	4	3	2	1	0	0	−1
8	8	7	6	5	4	3	2	1	0	0	−1
9	8	7	6	5	4	3	2	1	0	0	−1
10	8	7	6	5	4	3	2	1	0	0	−1

Table 5: Problem 1, optimal action at t=5

$$a_6(CS, d) =$$

	$CS = 0$	1	2	3	4	5	6	7	8	9	10
$d = 0$	4	3	2	1	0	0	0	0	0	0	0
1	3	2	1	0	0	0	−1	−2	−3	−4	−5
2	6	5	4	3	2	1	0	0	0	−1	−2
3	5	4	3	2	1	0	0	0	0	0	0
4	5	4	3	2	1	0	0	0	0	0	−1
5	5	4	3	2	1	0	0	0	0	0	−1
6	5	4	3	2	1	0	0	0	0	−1	−2
7	8	7	6	5	4	3	2	1	0	0	−1
8	8	7	6	5	4	3	2	1	0	0	−1
9	8	7	6	5	4	3	2	1	0	0	−1
10	8	7	6	5	4	3	2	1	0	0	−1

Table 6: Problem 1, optimal action at t=6

$$a_7(CS, d) =$$

	$CS = 0$	1	2	3	4	5	6	7	8	9	10
$d = 0$	4	3	2	1	0	0	0	0	0	0	0
1	3	2	1	0	0	−1	−2	−3	−4	−5	−6
2	5	4	3	2	1	0	0	0	0	−1	−2
3	5	4	3	2	1	0	0	0	0	0	0
4	5	4	3	2	1	0	0	0	0	0	−1
5	5	4	3	2	1	0	0	0	0	0	−1
6	5	4	3	2	1	0	0	0	−1	−2	−3
7	8	7	6	5	4	3	2	1	0	0	−1
8	8	7	6	5	4	3	2	1	0	0	−1
9	8	7	6	5	4	3	2	1	0	0	−1
10	8	7	6	5	4	3	2	1	0	0	−1

Table 7: Problem 1, optimal action at t=7

$$a_8(CS, d) =$$

	$CS = 0$	1	2	3	4	5	6	7	8	9	10
$d = 0$	2	1	0	0	0	0	0	0	0	0	0
1	2	1	0	0	−1	−2	−3	−4	−5	−6	−7
2	4	3	2	1	0	0	0	0	0	−1	−2
3	4	3	2	1	0	0	0	0	0	0	0
4	5	4	3	2	1	0	0	0	0	0	−1
5	4	3	2	1	0	0	0	0	0	0	−1
6	4	3	2	1	0	0	0	0	−1	−2	−3
7	7	6	5	4	3	2	1	0	0	0	−1
8	8	7	6	5	4	3	2	1	0	0	0
9	8	7	6	5	4	3	2	1	0	0	0
10	8	7	6	5	4	3	2	1	0	0	0

Table 8: Problem 1, optimal action at t=8

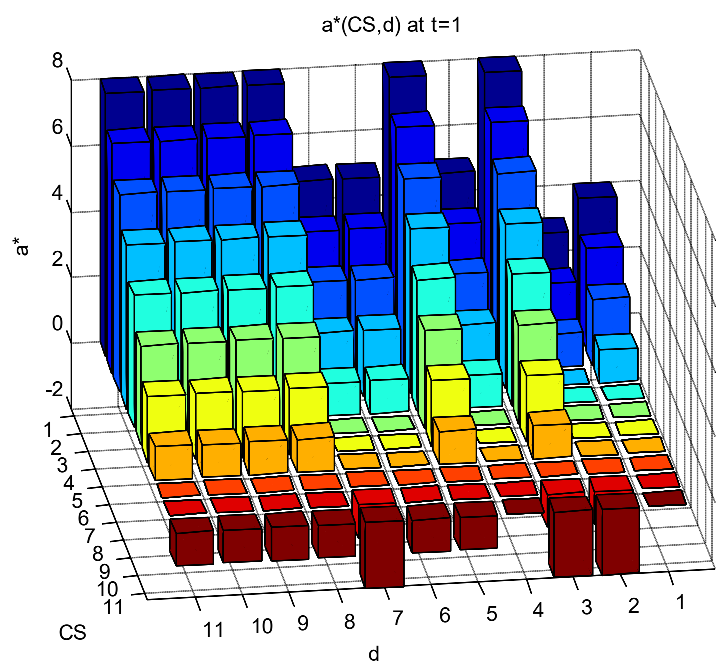

Figure 10: Graphical representation (i) of optimal solution at t=1 – Problem 1

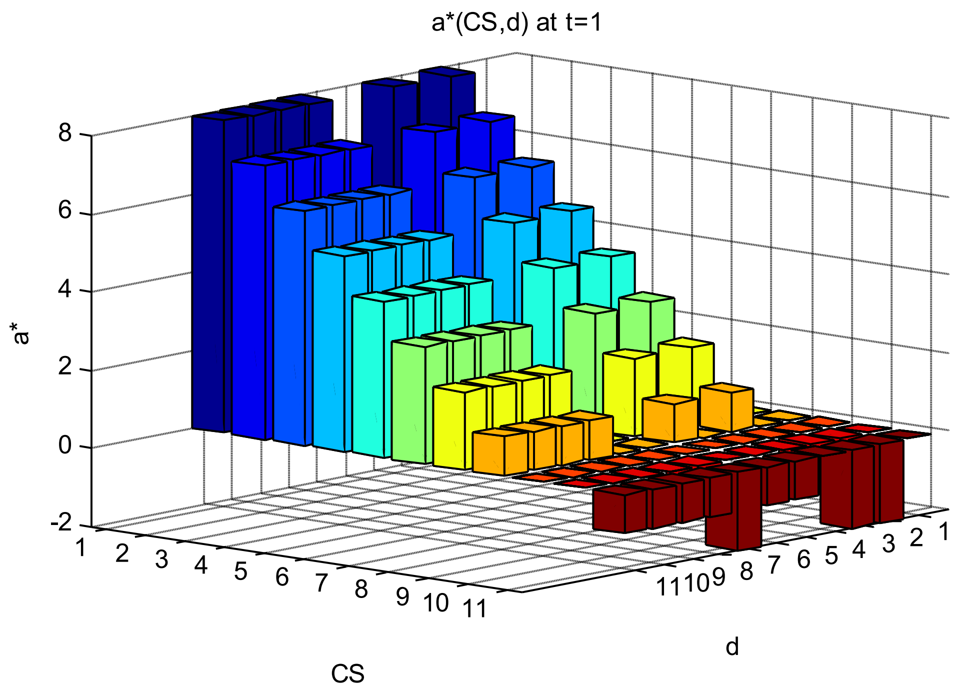

Figure 11: Graphical representation (ii) of optimal solution at t=1 – Problem 1

Problem 2:

The same problem with demand transition matrix having first order stochastic dominance given below

	$d_{t+1}=0$	1	2	3	4	5	6	7	8	9	10
$d_t=0$	.1	.1	.1	.1	.1	.1	.1	.1	.1	.1	
1		.1	.1	.1	.1	.1	.1	.1	.1	.1	.1
2			.2	.1	.1	.1	.1	.1	.1	.1	.1
3			.1	.2	.1	.1	.1	.1	.1	.1	.1
4				.2	.1	.2	.1	.1	.1	.1	.1
5					.1	.2	.2	.2	.1	.1	.1
6						.2	.2	.3	.1	.1	.1
7						.1	.3	.3	.1	.1	.1
8						.1	.2	.3	.1	.2	.1
9						.1	.1	.3	.2	.2	.1
10						.1	.1	.2	.2	.2	.2

Gives the following result

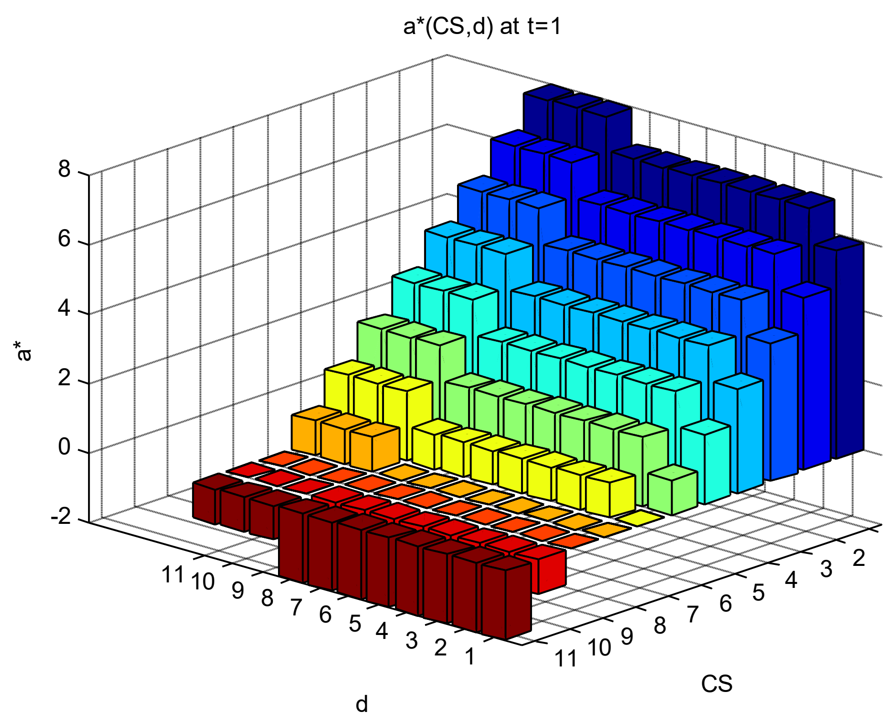

Figure 12: Graphical representation of optimal solution at t=1 – Problem 2

5.8 Chapter 5 Appendix

Lemma 2:

The function:

$$c_4 min(d_t, CS_{t-1} + a_{t-1}) + c_5(d_t - CS_{t-1} - a_{t-1})^+$$

is superadditive in $CS \times a \; \forall \; d_t$.

Proof:

$c_4 min(d_t, CS_{t-1} + a_{t-1}) + c_5(d_t - CS_{t-1} - a_{t-1})^+$ superadditive in $CS \times a \; \forall \; d_t$ means $\forall \; a2_{t-1} \geq a1_{t-1}$in a_{t-1}:

$c_4\big(min(d_t, CS_{t-1} + a2_{t-1}) - min(d_t, CS_{t-1} + a1_{t-1})\big) + c_5((d_t - CS_{t-1} - a2_{t-1})^+ - (d_t - CS_{t-1} - a1_{t-1})^+)$ is nondecreasing in CS_{t-1}

$c_4\big(min(d_t, CS_{t-1} + a2_{t-1}) - min(d_t, CS_{t-1} + a1_{t-1})\big) + c_5((d_t - CS_{t-1} - a2_{t-1})^+ - (d_t - CS_{t-1} - a1_{t-1})^+)$ has 3 regions:

1. $CS_{t-1} < d_t - a2_{t-1}$
2. $d_t - a2_{t-1} \leq CS_{t-1} < d_t - a1_{t-1}$
3. $CS_{t-1} \geq d_t - a1_{t-1}$

Region 1: $CS_{t-1} < d_t - a2_{t-1}$

$$\begin{aligned} &c_4\big(min(d_t, CS_{t-1} + a2_{t-1}) - min(d_t, CS_{t-1} + a1_{t-1})\big) \\ &\qquad + c_5((d_t - CS_{t-1} - a2_{t-1})^+ - (d_t - CS_{t-1} - a1_{t-1})^+) \\ &\qquad = (c_5 - c_4)(a1_{t-1} - a2_{t-1}) \end{aligned}$$

Region 2: $d_t - a2_{t-1} \leq CS_{t-1} < d_t - a1_{t-1}$

$$c_4\big(min(d_t, CS_{t-1} + a2_{t-1}) - min(d_t, CS_{t-1} + a1_{t-1})\big)$$
$$+ c_5((d_t - \; CS_{t-1} - a2_{t-1})^+ - (d_t - \; CS_{t-1} - a1_{t-1})^+)$$
$$= (c_5 - c_4)(CS_{T-2} + a1_{t-1} - d_t)$$

Region 3: $CS_{t-1} \geq d_t - a1_{t-1}$

$$c_4\big(min(d_t, CS_{t-1} + a2_{t-1}) - min(d_t, CS_{t-1} + a1_{t-1})\big)$$
$$+ c_5((d_t - \; CS_{t-1} - a2_{t-1})^+ - (d_t - \; CS_{t-1} - a1_{t-1})^+) = 0$$

Note from the cost parameter assumption $c_5 > c_4 + c_3$, the relation $c_5 > c_4$ is implicit.

Hence, $c_4\big(min(d_t, CS_{t-1} + a2_{t-1}) - min(d_t, CS_{t-1} + a1_{t-1})\big) + c_5((d_t - \; CS_{t-1} - a2_{t-1})^+ - (d_t - \; CS_{t-1} - a1_{t-1})^+)$ increasing in CS_{t-1}.

This completes the proof.

Lemma 3:

The function:

$$c_4 min(d_t, CS_{t-1} + a_{t-1}) + c_5(d_t - \; CS_{t-1} - a_{t-1})^+$$

is subadditive in $d \times a \; \forall \; CS_{t-1}$.

Proof:

$c_4 min(d_t, CS_{t-1} + a_{t-1}) + c_5(d_t - \; CS_{t-1} - a_{t-1})^+$ superadditive in $d \times a \; \forall \; CS_{t-1}$ means $\forall \; a2_{t-1} \geq a1_{t-1}$in a_{t-1}:

$c_4\big(min(d_t, CS_{t-1} + a2_{t-1}) - min(d_t, CS_{t-1} + a1_{t-1})\big) + c_5((d_t - \; CS_{t-1} - a2_{t-1})^+ - (d_t - \; CS_{t-1} - a1_{t-1})^+)$ is nonincreasing in d_t

$c_4\big(min(d_t, CS_{t-1} + a2_{t-1}) - min(d_t, CS_{t-1} + a1_{t-1})\big) + c_5((d_t - \ CS_{t-1} - a2_{t-1})^+ - (d_t - \ CS_{t-1} - a1_{t-1})^+)$ has 3 regions:

1. $d_t \leq CS_{t-1} + a1_{t-1}$
2. $CS_{t-1} + a1_{t-1} \leq d_t \leq CS_{t-1} + a2_{t-1}$
3. $d_t \geq CS_{t-1} + a2_{t-1}$

Region 1: $d_t \leq CS_{t-1} + a1_{t-1}$

$$c_4\big(min(d_t, CS_{t-1} + a2_{t-1}) - min(d_t, CS_{t-1} + a1_{t-1})\big) + c_5((d_t - \ CS_{t-1} - a2_{t-1})^+ - (d_t - \ CS_{t-1} - a1_{t-1})^+) = 0$$

Region 2: $CS_{t-1} + a1_{t-1} \leq d_t \leq CS_{t-1} + a2_{t-1}$

$$\begin{aligned} &c_4\big(min(d_t, CS_{t-1} + a2_{t-1}) - min(d_t, CS_{t-1} + a1_{t-1})\big) \\ &\quad + c_5((d_t - \ CS_{t-1} - a2_{t-1})^+ - (d_t - \ CS_{t-1} - a1_{t-1})^+) \\ &\quad = (c_5 - c_4)(CS_{t-1} + a1_{t-1} - d_t) \end{aligned}$$

Region 3: $d_t \geq CS_{t-1} + a2_{t-1}$

$$\begin{aligned} &c_4\big(min(d_t, CS_{t-1} + a2_{t-1}) - min(d_t, CS_{t-1} + a1_{t-1})\big) \\ &\quad + c_5((d_t - \ CS_{t-1} - a2_{t-1})^+ - (d_t - \ CS_{t-1} - a1_{t-1})^+) \\ &\quad = c_5(a1_{t-1} - a2_{t-1}) \end{aligned}$$

Note from the cost parameter assumption $c_5 > c_4 + c_3$, the relation $c_5 > c_4$ is implicit.

Hence we have $c_4\big(min(d_t, CS_{t-1} + a2_{t-1}) - min(d_t, CS_{t-1} + a1_{t-1})\big) + c_5((d_t - \ CS_{t-1} - a2_{t-1})^+ - (d_t - \ CS_{t-1} - a1_{t-1})^+)$ decreasing in d_t $\forall t$.

CHAPTER 6

SUMMARY AND CONCLUSIONS

6.1 Summary

This work applies MDP modeling to optimize logistics capacity in supply chain. In the first part of the thesis, a previous work on optimizing RL operations is studied. It is shown that the method of proof to guarantee the existence of monotone policies in the original work was inappropriate. A simple counterexample is provided to support this statement. Further, a formulation was corrected to correctly represent the stated problem. A new set of conditions to guarantee the existence of threshold policy in case of a two-period problem is then provided.

In the second part of the thesis, an MDP model for forward logistics was formulated and a complete study of the structural properties of the optimal decision policy was performed. The advantages in terms of computational effort due to the structural properties were quantified. Further, a modified value iteration algorithm is provided that applies these structural properties.

6.2 Future Research

Few of several research opportunities arising from this work are as follows:

1. A complete study of the structural properties for the multi period Reverse Logistics (RL) problem of Serrato et al. (2007).
2. Generalize the Forward Logistics problem by introducing to it fixed cost (switching cost) which is incurred whenever a decision is made to change its capacity.
3. Generalize the RL problem by allowing the decision maker to decide on the change in capacity when opting for internal RL.
4. Introducing the outsourcing option to Forward Logistics problem.

References

Ahiska, S Sebnem, Samyuktha R Appaji, Russell E King, and Donald P Warsing Jr. 2013. "A Markov decision process-based policy characterization approach for a stochastic inventory control problem with unreliable sourcing." *International Journal of Production Economics* 144 (2):485-96.

Ahiska, S Sebnem, and Russell E King. 2010. "Life cycle inventory policy characterizations for a single-product recoverable system." *International Journal of Production Economics* 124 (1):51-61.

Ahiska, S Sebnem, and Emre Kurtul. 2014. "Modeling and analysis of a product substitution strategy for a stochastic manufacturing/remanufacturing system." *Computers & Industrial Engineering* 72:1-11.

Bendre, Abhijit Bhagwan, and Lars Relund Nielsen. 2013. "Inventory control in a lost-sales setting with information about supply lead times." *International Journal of Production Economics* 142 (2):324-31.

Benjaafar, Saif, and Mohsen Elhafsi. 2012. "A production-inventory system with both patient and impatient demand classes." *Automation Science and Engineering, IEEE Transactions on* 9 (1):148-59.

Chen, Shuang, Stephen CH Leung, and Kin Keung Lai. 2010. "Integration of inventory and transportation decisions in a logistics system." *Transportation Research Part E: Logistics and Transportation Review* 46 (6):913-25.

Chien, Chen-Fu, Cheng-Hung Wu, and Yu-Shian Chiang. 2012. "Coordinated capacity migration and expansion planning for semiconductor manufacturing under demand uncertainties." *International Journal of Production Economics* 135 (2):860-9.

De Cuypere, Eline, Koen De Turck, Herwig Bruneel, and Dieter Fiems. 2013. "Optimal inventory management in a fluctuating market." In *Analytical and Stochastic Modeling Techniques and Applications*, 158-70. Springer.

Federgruen, Awi, and Aliza Heching. 1999. "Combined pricing and inventory control under uncertainty." *Operations Research* 47 (3):454-75.

Feng, Haolin, Qi Wu, Kumar Muthuraman, and Vinayak Deshpande. 2014. "Replenishment Policies for Multi-Product Stochastic Inventory Systems with Correlated Demand and Joint-Replenishment Costs." *Production and Operations Management*.

García-Alvarado, Marthy S, Marc Paquet, and Amin Chaabane. 2014. *An Inventory Model with Recovery and Environment Considerations*.

Golabi, Kamal. 1985. "Optimal inventory policies when ordering prices are random." *Operations Research* 33 (3):575-88.

Haijema, Rene. 2014. "Optimal ordering, issuance and disposal policies for inventory management of perishable products." *International Journal of Production Economics*.

Haijema, René, Jan van der Wal, and Nico M van Dijk. 2005. Blood platelet production: A multi-type perishable inventory problem. Paper presented at the Operations Research Proceedings 2004.

He, Ying. "A Markov Decision Process Model for Capacity Expansion and Allocation." Review of.

Helper, Chanel Marchèi, Lauren B Davis, and Wenbin Wei. 2010. "Impact of demand correlation and information sharing in a capacity constrained supply chain with multiple-retailers." *Computers & Industrial Engineering* 59 (4):552-60.

Higginson, James K, and James H Bookbinder. 1995. "Markovian decision processes in shipment consolidation." *Transportation Science* 29 (3):242-55.

Hoffmann, Rainer. 2013. "Dynamic capacity control in cargo revenue management–A new heuristic for solving the single-leg problem efficiently." *Journal of Revenue & Pricing Management* 12 (1):46-59.

Huang, Boray, and Seyed MR Iravani. 2007. "Optimal production and rationing decisions in supply chains with information sharing." *Operations Research Letters* 35 (5):669-76.

Iravani, Seyed MR, Bora Kolfal, and Mark P Van Oyen. 2012. "Process flexibility and inventory flexibility via product substitution." *Flexible Services and Manufacturing Journal*:1-24.

Kalymon, Basil A. 1971. "Stochastic prices in a single-item inventory purchasing model." *Operations Research* 19 (6):1434-58.

Kang, Seungmo, Juan C Medina, and Yanfeng Ouyang. 2008. "Optimal operations of transportation fleet for unloading activities at container ports." *Transportation Research Part B: Methodological* 42 (10):970-84.

Kingsman, BG. 1969. "Commodity purchasing." *OR*:59-79.

Kleywegt, Anton J, Vijay S Nori, and Martin WP Savelsbergh. 2002. "The stochastic inventory routing problem with direct deliveries." *Transportation Science* 36 (1):94-118.

Kleywegt, Anton J, Vijay S Nori, and Martin WP Savelsbergh. 2004. "Dynamic programming approximations for a stochastic inventory routing problem." *Transportation Science* 38 (1):42-70.

Li, Xiaoming. 2013. "Managing Dynamic Inventory Systems with Product Returns: A Markov Decision Process." *Journal of Optimization Theory and Applications* 157 (2):577-92.

Lin, James T, Tzu-Li Chen, and Hsiao-Ching Chu. 2014. "A stochastic dynamic programming approach for multi-site capacity planning in TFT-LCD manufacturing under demand uncertainty." *International Journal of Production Economics* 148:21-36.

Morton, Thomas E. 1971. "Technical Note—On the Asymptotic Convergence Rate of Cost Differences for Markovian Decision Processes." *Operations Research* 19 (1):244-8.

Nakashima*, Kenichi, Hiroyuki Arimitsu, Toyokazu Nose, and Sennosuke Kuriyama. 2004. "Optimal control of a remanufacturing system." *International Journal of Production Research* 42 (17):3619-25.

Nikolaou, Ioannis E, Konstantinos I Evangelinos, and S Allan. 2013. "A reverse logistics social responsibility evaluation framework based on the triple bottom line approach." *Journal of Cleaner Production* 56:173-84.

Pang, Zhan, Houcai Shen, and TCE Cheng. 2014. "Inventory Rationing in a Make-to-Stock System with Batch Production and Lost Sales." *Production and Operations Management*.

Parlar, Mahmut. 1985. "Optimal ordering policies for a perishable and substitutable product: A Markov decision model." *Infor* 23 (2):182-95.

Puranam, Kartikeya S, and Michael N Katehakis. 2014. "On optimal bidding and inventory control in sequential procurement auctions: the multi period case." *Annals of Operations Research* 217 (1):447-62.

Puterman, Martin L. 2009. *Markov decision processes: discrete stochastic dynamic programming*. Vol. 414: John Wiley & Sons.

Rida, Mohamed. 2014. "Modeling and Optimization of Decision-Making Process During Loading and Unloading Operations at Container Port." *Arabian Journal for Science and Engineering* 39 (11):8395-408.

Rogers, Dale S, and Ronald Tibben-Lembke. 2001. "An examination of reverse logistics practices." *Journal of business logistics* 22 (2):129-48.

Sebnem Ahiska, S, and Russell E King. 2010. "Inventory optimization in a one product recoverable manufacturing system." *International Journal of Production Economics* 124 (1):11-9.

Serrato, Marco A, Sarah M Ryan, and Juan Gaytán. 2007. "A Markov decision model to evaluate outsourcing in reverse logistics." *International Journal of Production Research* 45 (18-19):4289-315.

Sinha, Ashesh Kumar, and Ananth Krishnamurthy. "Optimal Control Policy for Multi-Product Assemble-to-Order Systems." Review of.

Snyder, Ralph D. 1975. "Technical Note—A Dynamic Programming Formulation for Continuous Time Stock Control Systems." *Operations Research* 23 (2):383-6.

Thomas, L Joseph. 1974. "Technical Note—Price and Production Decisions with Random Demand." *Operations Research* 22 (3):513-8.

Tibben-Lembke, Ronald S, and Dale S Rogers. 2002. "Differences between forward and reverse logistics in a retail environment." *Supply Chain Management: An International Journal* 7 (5):271-82.

Tiemessen, Harold, Moritz Fleischmann, and Geert-Jan Van Houtum. 2014. "Dynamic Control in Multi-Item Production/Inventory Systems." *Inventory Systems (November 7, 2014)*.

Tiemessen, HGH, and GJ Van Houtum. 2013. "Reducing costs of repairable inventory supply systems via dynamic scheduling." *International Journal of Production Economics* 143 (2):478-88.

van Wijk, ACC, Ivo JBF Adan, and GJ Van Houtum. 2013. "Optimal allocation policy for a multi-location inventory system with a quick response warehouse." *Operations Research Letters* 41 (3):305-10.

Vercraene, Samuel, and Jean-Philippe Gayon. 2013. "Optimal control of a production-inventory system with productreturns." *International Journal of Production Economics* 142 (2):302-10.

Vila-Parrish, Ana R, Julie S Ivy, Russell E King, and Steven R Abel. 2012. "Patient-based pharmaceutical inventory management: a two-stage inventory and production model for perishable products with Markovian demand." *Health Systems* 1 (1):69-83.

Vila-Parrish, Ana R, Julie Simmons Ivy, and Russell E King. 2008. A simulation-based approach for inventory modeling of perishable pharmaceuticals. Paper presented at the Simulation Conference, 2008. WSC 2008. Winter.

White, DJ. 1965. "Dynamic programming and systems of uncertain duration." *Management Science* 12 (1):37-67.

White, Douglas J. 1985. "Real applications of Markov decision processes." *Interfaces* 15 (6):73-83.

White, Douglas J. 1988. "Further real applications of Markov decision processes." *Interfaces* 18 (5):55-61.

White, Douglas J. 1993. "A survey of applications of Markov decision processes." *Journal of the Operational Research Society*:1073-96.

Wu, Cheng-Hung, and Ya-Tang Chuang. 2010. "An innovative approach for strategic capacity portfolio planning under uncertainties." *European Journal of Operational Research* 207 (2):1002-13.

Wu, Hua-Hsuan, Cheng-Hung Wu, and James T Lin. 2011. "Dynamic selling of quality-graded products under demand uncertainties." *Computers & Industrial Engineering* 61 (1):141-9.

Zhang, Dan, and Royce Kallesen. 2008. "Incorporating competitive price information into revenue management." *Journal of Revenue & Pricing Management* 7 (1):17-26.

Vitae

JAZEEM ABDUL JALEEL

Pazhoor House, 363 B Cheroor Road, Chembukavu, Trichur, Kerala, 680020, India
Tel: +966-50-3051789 (present), +91 - 9946- 052382 **E-mail**:jazeemj@gmail.com

BIO INFORMATION

Date of Birth: 22nd August 1987
Marital Status: Married
Country of Citizenship: India
Languages: English, Hindi, Malayalam

AREAS OF INTEREST

Operations Research and applications, supply chain management, maintenance management, quality control, simulation, probabilistic models

EDUCATIONAL QUALIFICATIONS

Course	Year	University/Board	CGPA
M.Sc. Industrial & Systems Engineering	2013 – 2015	KFUPM, Dhahran, Saudi Arabia	3.89/4
B.Tech. Production engineering	2005-2009	NIT Calicut Kerala, India	7.76/10

WORK EXPERIENCE

HSE Officer

Rowan Drilling Company (Qatar and Saudi Arabia)

Rig : Rowan Paris 023, Rowan Hank Boswel 057, Rowan Mississippi 074

Location : Al Shaheen Field (Offshore Qatar); Rastanura, Tanajib(Offshore Saudi Arabia)

Time : From June 2009 till June 2013

Job Scope : Ensuring worksite compliance with company safety policies, supervise permits and worksites, crew training, equipment inspections, safety equipment procurement and inventory management, safety auditing and incident investigations

RESEARCH EXPERIENCE

A Markov Decision Process Model for logistics in supply chain (M.Sc. Thesis)

PAPER SUBMITTED FOR PUBLICATION	Abdul Jaleel , J. and AlDurgam, M., A note on "A Markov decision model to evaluate outsourcing in reverse logistics", and a corrected proof". (Submitted for publication-IJPR)
PAPERS IN PROGRESS	Abdul Jaleel , J. and AlDurgam, M., "A Markov decision model for capacity planning in forward logistics ". AlDurgam, M. and Abdul Jaleel , J., "A count-based maintenance model for a process with two types of inspection and repairs"

COMPUTER SKILLS

- Programming Languages: C++
- Engineering (basics): Pro-E, Catia, Ansys
- Engineering (advanced): Matlab, Arena, Minitab, Lingo

TEST SCORES

- **GRE** – 14 September 2012

Verbal – 160/170	Quantitative – 164/170	Analytical Writing – 3.0/6.0

- **TOEFL** – 21 August 2011 (110/120)

Reading - 28	Listening – 26	Speaking – 28	Writing – 28

INDUSTRIAL TRAINING

ROWAN DRILLING COMPANY

Duration	: 1 month
Time	: July 2007
Location	: Offshore Qatar
Study	: Familiarize with the various operations on location by being placed under different supervisors during the duration of the training

WORK BASED TRAININGS/ CERTIFICATIONS

- **Fundamentals of Well Control** from SG Petroleum
- **Advanced Safety Certificate** from National Safety council
- **Fundamentals of Industrial Hygiene** from National Safety Council
- **Safety Training Methods** from National Safety Council
- **Certificate in Principles of Safety** from National Safety Council
- **Emergency Care Program** from National Safety Council
- **Introduction to OSHA** from Aberdeen Drilling School

ADDITIONAL TRAININGS/ CERTIFICATION

- **Six Sigma Green Belt certification** from IIE (2015)
- **SAP TERP10 certification** from SAP (2015)
- **SAP Project Management certification** from SAP (2015)

PROJECTS

Mechanization of firm making agricultural implements (undergraduate)

Abstract: Metal Industries Ltd. Shoranur (A Kerala Government Undertaking, Estd: 1928, India) is in the business of manufacturing agricultural implements and products. We studied their operations, determined the bottle neck process and optimized the system. Here are the few highlights of the project:

- The design of dies and punches for seven different kinds of hoes using engineering softwares like Pro-E.
- Calculation of press capacity and selection of suitable press
- Cost estimation and Break even analysis