Constrained Markov Decision Processes with Application to Wireless Communications

Alexander Zadorojniy

Constrained Markov Decision Processes with Application to Wireless Communications

Final Paper

Submitted in Partial Fulfillment of the

Requirements for the Degree of

Master of Science in

Electrical Engineering

Alexander Zadorojniy

Submitted to the Senate of
the Technion – Israel Institute of Technology

SIVAN , 5764 HAIFA June 2004

The Final Paper Was Done Under

The Supervision of Professor A. Shwartz

in the Faculty of Electrical Engineering

I Would Like To Express My Deep Gratitude

To Professor A. Shwartz

For His Guidance and Help

Throughout All Stages of the Research.

THE GENEROUS FINANCIAL SUPPORT OF TECHNION IS GRATEFULLY ACKNOWLEDGED.

Contents

Page

Abstract……………………………………………………………………...…..1

List of Symbols and Notations……..…………………..…………………….....2

1. Introduction……………………………………..………………..…...4

2. Survey and Research Objectives………..…………..…………………..……5

3. Wireless Communications System Model………..……………………………..7

4. Constrained Markov Decision Processes (CMDP) and Linear Programming Approach……………...………………………………………..…..…10

4.1 CMDP……..…………………………………………………………..…10

4.2 Optimal Policies for CMDP Problem……….…………………………...…12

4.3 Occupation Measure and Linear Programming (LP)……...…………………13

5. The Optimal Policy Properties for CMDP……………...………………..……18

5.1 The Karush-Kuhn-Tucker (KKT) Conditions……………………………...……18

5.2 The Fundamental Theorems……………………………..………….……….22

6. The Optimal Policy Properties for Power Saving Problem……………….….29

6.1 CMDPS Algorithm for Two Control Levels..……………………..………35

6.1.1 CMDPS Algorithm…………………………….…………...……...37

6.1.2 CMDPS Algorithm Usage………...……………………………..39

6.1.3 Example of CMDPS Algorithm Application to CMDPS Problems....45

6.2 CMDPS Algorithm for Finite Number of Control Levels…………….……..50

Contents (Cont.)

Page

7. A New Approach to Optimization of Wireless Communications…..……….55

7.1 Communications Problem Optimization Using CMDP Approach…..……….58

7.2 CMDPS Algorithm Application in Wireless Communications……..………62

8. Conclusions and Further Research……………………………..……..…66

Appendix A: Linear Programming (LP) and Simplex Method Idea……………..…67

Appendix B: Principle of Optimality and Value Iteration Algorithm........…..…73

Appendix C: BPSK Modulation in Fading Environment……………..……..…75

References……………………………..…………………………………..…77

Hebrew Abstract………………………………………………………...……I

List of Figures

Number **Page**

1. Wireless Communications System Model………….. ……………………….…8

2. Markov Chain for Monotonicity Proof……………………………………..…30

3. Coupled System…………………………..…...……………………………….30

4. Randomization Changes………………….…………………………….…..……36

5. CMDPS Algorithm Diagram………………..……………….……………….40

6. Markov Chain for Example 1……………………………………………………….46

7. The Simplex Method Results……………………….………………………51

8. Randomization Behavior in Example 1………………….……………………..…..52

9. Markov Chain of the Wireless Communications Problem…..…………….……..56

10. Simplified Markov Chain of the Wireless Communications Problem ………..…57

11. Markov Chain for Communications Problem Example…………………….……63

12. Geometric Solution…………….……………………………………..……69

13. Simplex Method – Geometric Motivation…………………………….....……..70

Abstract

With the development of personal communication services, portable terminals such as mobile telephones and notebook computers are expected to be used more frequently and for longer times. Hence power consumption will become even more important than it is now. One of the major concerns in supporting such mobile applications is the energy conservation and managements in mobile devices. In wireless communications, low power may cause errors during the transmission. Therefore when we try to conserve power we have to care about the Quality-of-Service.

In recent years several solution methods were introduced. Usually these methods find the optimal power policy for minimal average delay under average power constraint. Some of them are for continuous time domain and other for discrete one. Methods which solve the problem in the continuous time domain are general but they are not always applicable for digital systems, and these type of systems are most popular in recent years. Methods which deal with discrete time domain, in many cases, are not general enough, only for two control levels, for example. As usual, in digital systems we are interested to find an optimal policy which minimizes the average delay under average and peak power constraints and uses only available discrete power levels. None of the existing methods can give us this capability.

This research deals with a derivation of new solution methods for constrained Markov decision processes and applications of these methods to the optimization of wireless communications. We intend to survey the existing methods of control, which involve control of power and delay, and investigate their effectiveness. We introduce a new type of power control in wireless communications, which minimizes the transmission delay under the average power constraint where at each slot uses one of the available discrete power levels, while the maximal power level is limited by a peak power constraint. Moreover we develop algorithm which, aside from the optimization problem solving, will be able to show sensitivity of the solution to changes in the average power level constraint.

List of Symbols and Notations

ACK	The acknowledge about successful transmission
$NACK$	The acknowledge about failure
SNR	The signal to noise ratio
AWGN	The additive white Gaussian noise
MDP	Markov decision processes
CMDP	Constrained Markov decision processes
CMDPS Algorithm	Constrained Markov decision processes solver algorithm
KKT Conditions	Karush- Kuhn Tucker Conditions
N	The buffer size
r_k	The random variable that represents a fading at slot k
n_k	The random variable that represents an AWGN at slot k
l_k	The codeword transmitted at slot k
y_k	The received word at slot k
R	The transmission rate
λ	The arrival rate
$\triangle t$	The slot duration
π	A policy
σ	The initial state
X	The state space
U	The action space
$P(u) = \{P_{ij}(u)\}$	The transition matrix when action $u \in U$ is taken
$c = c(x, u)$	The immediate cost at state $x \in X$ using action $u \in U$
$C = C(\sigma; \pi)$	The value of the criterion when starting at σ and using policy π

$d = d(x, u)$	The immediate cost, related to the constraint
$D = D(\sigma; \pi)$	The value of the constraint criterion
$\rho(x, u)$	The probability to be at state $x \in X$ and use action $u \in U$
β	The discount factor
$\mu_x(u)$	The ratio of using action $u \in U$ at $x \in X$ to all possible actions
α	The average available power
z	A vector of length n
b	A vector of length m
s	A vector of variables of length n
A	An $m \times n$ matrix
E_b	A transmitted signal energy per bit
T_b	A time duration of one bit transmission
f_c	A carrier frequency

1 Introduction

In this paper we consider a situation where one type of cost (delay, throughput,etc...) is to be minimized while keeping the other types of costs (power, delay, etc.) below some given bounds. Posed in this way, our control problem can be viewed as a constrained optimization problem over a given class of policies.

Telecommunications networks are designed to enable the simultaneous transmission of different types of traffic: voice, file transfers, interactive messages, video, etc. Typical performance measure are the transmission delay, power consumption, throughput, transmission error probabilities, etc. Different types of traffic differ from each other by their statistical properties, as well by their performance requirements. For example, for interactive messages it is necessary that the average end-to-end delay be limited. Strict delay constraints are important for voice traffic; there, we impose a delay limit of 0.1 second. When the delay increases beyond this limit, it becomes quickly intolerable. For non-interactive file transfer, we often wish to minimize delays or to maximize throughput.

A trade-off exists between achieving a small delay, on the one hand, and low power consumption on the other. Note that delay minimization and conserving power are two conflicting performance metrics. To minimize the delay we should transmit with the highest possible power because it will increase the probability of successful transmission and decrease the number of retransmissions. On the contrary, to decrease power consumption, we are interested to transmit with lowest possible power. The problem is formulated as a constrained MDP, where we wish to minimize the costs related to the delay subject to constraints on the average and peak power.

2 Survey and Research Objectives

We can divide the known work in the subject of saving energy to two types: first, the energy conservation control problem by assuming there is a finite supply of information to be transmitted and second, where the supply of information to be transmitted is infinite.

The first type of problems are considered in [12]. In this work the authors analyze the power control problem in transferring a finite-size data under a total energy constraint as well as delay constraints. In the paper they considered a problem with only two power levels: a constant power or zero power (i.e., no transmission). A randomized power control scheme is discussed, i.e., the transmitter can select either power level with a certain probability. The power control problem is (given the total energy and the file size) to find the optimal policy that maximizes the probability of success under either an average delay constraint or a strict delay constraint. These two problems form a constrained Markov decision problem and they can be solved via a dynamic programming algorithm. In order to solve them the authors used the techniques which were developed in [1] and [8].

The second type of problems are considered in ([5], Chapter 7.4). In this problem the author assumes that there exists a buffer so that the information rate into it is constant and a transmission rate is constant as well. The cost function is defined as buffer size at time k plus power, that was used in this time, multiplied by a Lagrange multiplier in order to control the average power. This optimization problem was solved by dynamic programming algorithm [4]. More precisely, the optimal equation was derived by dynamic programming techniques, and the equation itself was solved by the value iteration method (Appendix B), [4] and [8].

The first work helps us to better understand the optimal policy for problems with average and peak constraints for two control levels but the solution for general problem when we have constraints for both peak and average power (or delay) simultaneously, with finite, but arbitrary number of control levels is still unclear from this paper. The second one is much more general. When the author is solving the optimization problem he takes care of average power by Lagrangian multiplier so that the average power will be below some chosen level. The problem is that when the level of average power is given and we need to find the suitable Lagrangian multiplier - this can be a not trivial mathematical problem. In the paper the author shows what is a necessarily power levels in order to get an optimal solution, but it is not clear what happens if this power level is not available, in other words, if we have discrete levels of power and need to find the optimal solution by using only these power levels. The more complicated question can be asked as well, what happens to optimal policy when the level of average available power is changing. Or another question, if the policy will be changed in the same way when we have a small change in the average available power and when we have a big change in it. Or what is the behavior of the optimal solution for this kind of problems. Or does there exist common properties of the solution for different values of average power level constraints.

One way to solve the first two problems is, instead of using the value iteration method for optimality equation, to represent the problem in the linear programming form [1] and to solve it by the simplex method [3] and (Appendix A). In order to answer the rest of the questions we will derive and prove theorems and based on them algorithm that will solve the problem (sections 5,6). Moreover the optimal solution that is derived from the new developed algorithm will be compared to the solution obtained from the simplex method (section 6). We will show that in addition to the ability of the solution sensitivity investigation, the algorithm can be used as a tool for solving constrained Markov decision processes problems (sections 5,6). In section 7 the algorithm will be used in order to solve a wireless optimization problem that will be defined in section 3.

In this research we developed two fundamental theorems (section 5.2) which describe the structure of the optimal solution for general constrained Markov decision process problems. Two additional theorems, that describe the properties of the solution for constrained Markov decision process in power saving problems, were developed in section 6. In section 6.1.1 we developed innovative algorithm which solves constrained Markov decision process for power saving problems and which was applied to wireless communications problems in section 7.2.

3 Wireless Communications System Model

Our system model can be represented by five blocks (Figure 1): buffer, transmitter, fading channel, receiver and controller.

A buffer is a device that receives codewords with rate λ, stores and removes them depending on "Buffer Control" value.

A transmitter is a device that transmits codewords with rate R and power u which depends on the "Power Control" value.

A fading Channel is a channel where except for additive noise, there exists multiplicative disturbances that impact the amplitude of the transmitted signal.

A receiver is a device that receives the transmitted signal, decodes it, if needed, and transmits acknowledge (ACK or NACK) to the Controller in the transmitter side.

A controller is a device that has one input and two outputs. The input to the controller is a feedback from the receiver. The first output, named "Buffer Control", controls when the codewords are removed from the buffer or stored in it. The second output, named "Power Control", controls the power level that the transmitter should use.

This system works as follows: the transmitter transmits codewords from the buffer, afterward an amplitude of the transmitted signal is multiplied by disturbances and additive noise is added to it. In order to write a precise expression for this process, let's define the following:

Definitions:

- r_k is a random variable that represents a fading at slot k.
- n_k is a random variable that represents an additive white Gaussian noise ($AWGN$) at slot k.
- l_k is a codeword transmitted at slot k.
- y_k is a received word at slot k.

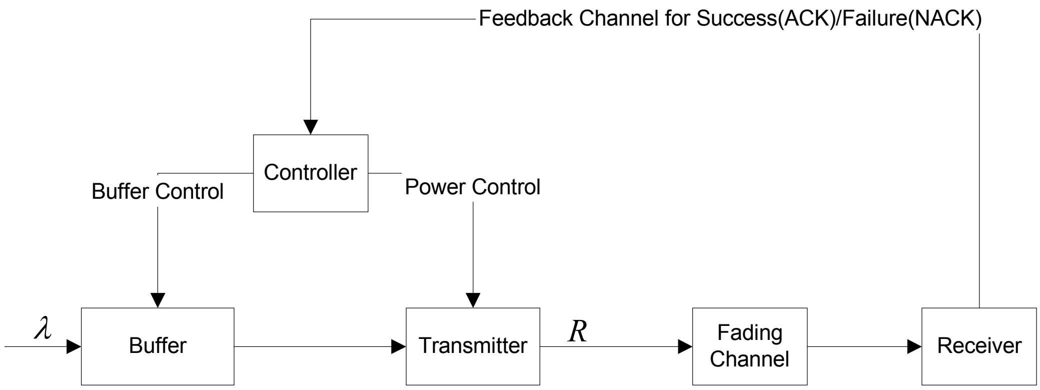


Figure 1: Wireless Communications System Model

- R is a transmission rate. It is constant over the whole period of transmission.
- λ is an arrival rate. It is constant over the whole period of transmission.
- $\triangle t$ is the slot duration.
- ACK is an acknowledge about successful transmission.
- $NACK$ is an acknowledge about failure.

In Communications the relationship between transmitted codeword and received codeword can be expressed by the following formula [7]

$$y_k = r_k l_k + n_k \tag{1}$$

Now we will describe a block fading Gaussian channel model:

- This model assumes that the fading-gain process is constant on blocks of N symbols.
- It is modeled as a sequence of independent random variables, each of which is the fading gain in a block.
- Except a fading there exists an additive white Gaussian noise ($AWGN$), where the fading and $AWGN$ are iid and independent from each other.

We assume a block fading Gaussian channel in which the fading of each slot is *iid* according to some distribution (such as Rayleigh, Rice etc...). At each slot a codeword is transmitted with constant rate R[codewords/time duration]. The information rate into the buffer is constant λ[codewords/time duration].

Define

$$\triangle t \triangleq \frac{1}{R}[codewords]$$

Transmission power is constant during each slot $\triangle t$ and may vary between slots. We assume that the average power of $AWGN$ and fading disturbances remain at the same level during the whole period of the transmission.

If a transmission succeeds an ACK is returned to the controller in the transmitter side using a feedback channel without errors and delay. Otherwise a $NACK$ is returned through the feedback channel without errors and delay, causing a retransmission of the data until an ACK is returned. The buffer stores the information that has to be transmitted. The information is removed from the buffer only if an ACK is returned. We assume that the buffer size is large enough so that we can neglect any loss of information due to overflow. We aim at finding the value of the optimal transmission power at each slot so that the average delay is minimal, the average power is below a given level α, and at each slot one of the available discrete power levels is used where maximal power level is limited by a peak power constraint. The important thing to notice is that at each slot only one codeword can be transmitted.

We will try to represent this model as Constrained Markov Decision Process (CMDP) model.

Let us now introduce CMDP and the new algorithm that will help us to optimize the solution of the communications problem. We will return to the communications problem in section (7).

4 Constrained Markov Decision Processes (CMDP) and Linear Programming Approach

Markov decision processes (*MDP*), also known as controlled Markov chains, constitute a basic framework for dynamically controlling systems that evolve in a stochastic way. We focus on discrete time models: we observe the system at times $t = 1, 2, ..., n$ where n is called horizon, and may be either finite or infinite. A controller has an influence on both the costs and the evolution of the system, by choosing at each time unit some parameters, called actions. As is often the case in control theory, we assume that the behavior of the system at each time is determined by what is called the 'state' of the system, as well as the control action. The system moves sequentially between different states in a random way; the current state and control action fully determine the probability to move to any given state in the next time unit.

MDP is thus a generalization of (non-controlled) Markov chains, and many useful properties of Markov chains carry over to controlled Markov chains. A key Markovian property is that conditioned on the state and action at some time t, the past states and the next one are independent.

The model that we consider in this paper is special in that more than one objective cost exists; the controller minimizes one of the objectives subject to constraints on the other. We will call this class of MDP Constrained MDP, or simply CMDP.

4.1 CMDP

Definition: A finite constrained Markov decision process is a 7-tuple $\{X, U, P, c, C, d,$D$\}$ [1] and [6] , where

- X is the state space that contains a finite number of states.
- U is the finite set of actions.

- $P(u) = \{P_{ij}(u)\}$ is the transition matrix when action u is taken.

- $c = c(x, u)$ is the immediate cost at state x using action u.

- $C = C(\sigma; \pi)$ is the value of the criterion when starting at σ and using policy π.

- $\overline{d} = \overline{d}(x, u) = \begin{pmatrix} d_1 \\ d_2 \\ . \\ . \\ d_n \end{pmatrix}$ is a vector of immediate costs, related to constraints, when at state x and using action u.

- $\overline{D} = \overline{D}(\sigma; \pi)$ is the vector of values of the constraint criteria when starting at σ and using policy π.

- $\geq$ for vectors means that each elements in the left hand vector is greater or equal to the corresponding element in the right hand one.

- Similar definitions hold for $\leq$ and $=$.

We now define the cost criteria. For any policy π and initial distribution σ at $t = 1$, the finite horizon cost for a horizon n is defined as [1]

$$C^n(\sigma, \pi) = \sum_{t=1}^{n} E_\sigma^\pi c(X_t, U_t)$$

An alternative cost that gives less importance to the far future is the discounted cost. For a fixed discount factor $\beta, 0 < \beta < 1$, define

$$\begin{aligned} C_\beta^n(\sigma, \pi) &= (1 - \beta) \sum_{t=1}^{n} \beta^{t-1} E_\sigma^\pi c(X_t, U_t) \\ C_\beta(\sigma, \pi) &= \overline{\lim_{n \to \infty}} C_\beta^n(\sigma, \pi) \end{aligned}$$

Since there are finitely many states and actions, the $\overline{lim}$ indeed exists as a limit and

$$C_\beta(\sigma,\pi) = (1-\beta)\sum_{t=1}^{\infty}\beta^{t-1}E_\sigma^\pi c(X_t,U_t)$$

Also, in a similar way we derive for the discounted cost, that is related to constraint, that

$$\overline{D}_\beta(\sigma,\pi) = (1-\beta)\sum_{t=1}^{\infty}\beta^{t-1}E_\sigma^\pi \overline{d}(X_t,U_t) \tag{2}$$

Quite frequently, the discounted cost is defined without the normalizing constant $(1-\beta)$. The techniques are the same for both cases, we can get one from another by multiplying or dividing the immediate cost by this factor. There are several advantages of using this normalization. First, we avoid the situation where, for fixed immediate cost c, the total discounted cost becomes very large if β is close to one. Second, with this normalization, the discounted cost will be seen to converge to the expected average cost when stationary policies are used. Finally, we shall see that the LP used to solve the discounted and the expected average costs has the same form when the normalization constant is used.

For a fixed real vector $\overline{V}$, we define the constrained control problem *COP* as:

Find a policy that minimizes $C_\beta(\sigma,\pi)$ subject to $\overline{D}_\beta(\sigma,\pi) \leq \overline{V}$.

4.2 Optimal Policies for CMDP Problem

Optimal policies are defined with respect to a given initial state. A policy that is optimal for one state might not even be feasible for another. In fact, there may be some initial states at which no policy is feasible, where feasible policy is a policy that satisfies the constraints. This is in contrast to non-constrained MDPs, in which there typically exist policies that are optimal for all initial states.

The class of Markov policies turns out to be rich in the following sense. For any policy, the exists an equivalent Markov policy that induces the same marginal probability measure, i.e., the same probability distribution of the pairs (X_t, U_t), $t = 1, 2, ...$ [1]

All cost criteria that we defined in the previous subsection have the property that they are functions of the distribution of these pairs. We conclude that Markov policies are sufficiently rich so that a cost that can be achieved by an arbitrary policy can also be achieved by a Markov policy. Moreover the Markov policies are dominating for any cost criterion which is a function of the marginal distribution of states and actions ([1], Theorem 2.1).

4.3 Occupation Measure and Linear Programming (LP)

An occupation measure corresponding to a policy π is the total expected discounted time spent in different state-action pairs. It is thus a probability measure over the set of state-action pairs and it has the property that the discounted cost corresponding to that policy can be expressed as the expectation of the immediate cost with respect to this measure.

More precisely, define for any initial distribution σ, any policy π and any pair x, u:

$$f_\beta(\sigma, \pi; x, u) \overset{def}{=} (1-\beta) \sum_{t=1}^{\infty} \beta^{t-1} P_\sigma^\pi(X_t = x, U_t = u), \; x \in X, u \in U(x)$$

$f_\beta(\sigma, \pi)$ is then defined to be the set $\{f_\beta(\sigma, \pi; x, u)\}_{x,u}$. It can be considered as a probability measure, which we call the occupation measure, that assigns probability $f_\beta(\sigma, \pi; x, u)$ to the pair (x, u).

The discounted cost can be expressed as [1]

$$\begin{aligned}
C_\beta(\sigma,\pi) &= (1-\beta)E_\sigma^\pi\{\sum_{t=1}^{\infty}\beta^{t-1}c(x_t,u_t)\} \\
&= (1-\beta)\sum_{t=1}^{\infty}\beta^{t-1}E_\sigma^\pi c(x_t,u_t) \\
&= (1-\beta)\sum_{t=1}^{\infty}\sum_{x,u}\beta^{t-1}P_\sigma^\pi(x_t=x,u_t=u)c(x,u) \\
&= \sum_{x,u}(1-\beta)\sum_{t=1}^{\infty}\beta^{t-1}P_\sigma^\pi(x_t=x,u_t=u)c(x,u) \\
&= \sum_{x,u}(1-\beta)\sum_{t=1}^{\infty}\beta^{t-1}P_\sigma^\pi(x_t=x,u_t=u)c(x,u) \\
&= \sum_{x,u}f_\beta(\sigma,\pi;x,u)c(x,u)
\end{aligned}$$

$$C_\beta(\sigma,\pi) = \sum_{x\in X}\sum_{u\in U}f_\beta(\sigma,\pi;x,u)c(x,u) \tag{3}$$

Also, in a similar way we derive for the discounted cost, related to the constraint, that

$$\overline{D}_\beta(\sigma,\pi) = \sum_{x\in X}\sum_{u\in U}f_\beta(\sigma,\pi;x,u)\overline{d}(x,u) \tag{4}$$

Lemma ([10])

$$\sum_{y\in X}\sum_{u\in U(y)}f_\beta(\sigma,\pi;y,u)(\delta_x(y)-\beta P_{yx}(u)) = (1-\beta)\sigma(x), \forall x\in X \tag{5}$$

$$f_\beta(\sigma,\pi;y,u)\geq 0,\ \forall y,u \tag{6}$$

$$\sum_{y\in X}\sum_{u\in U(y)}f_\beta(\sigma,\pi;y,u) = 1 \tag{7}$$

Proof:

If $x\neq\sigma$ then $P_\sigma^\pi(X_{t=1}=x,U_{t=1}=u)=0$ so that

$$\begin{aligned}
\sum_{u\in U(x)} f_\beta(\sigma,\pi;x,u) &= \sum_{u\in U(x)}(1-\beta)\sum_{t=1}^{\infty}\beta^{t-1}P_\sigma^\pi(X_t=x,U_t=u)\\
&= \sum_{u\in U(x)}(1-\beta)\sum_{t=2}^{\infty}\beta^{t-1}P_\sigma^\pi(X_t=x,U_t=u)\\
&= (1-\beta)\sum_{t=2}^{\infty}\sum_{u\in U(x)}\beta^{t-1}P_\sigma^\pi(X_t=x,U_t=u)\\
&= (1-\beta)\beta\sum_{t=1}^{\infty}\sum_{u\in U(x)}\beta^{t-1}P_\sigma^\pi(X_{t+1}=x,U_{t+1}=u)\\
&= (1-\beta)\beta\sum_{t=1}^{\infty}\sum_{y\in X}\sum_{u\in U(y)}\beta^{t-1}P_\sigma^\pi(X_t=y,U_t=u)P_{yx}(u)\\
&= \beta\sum_{y\in X}\sum_{u\in U(y)}(1-\beta)\sum_{t=1}^{\infty}\beta^{t-1}P_\sigma^\pi(X_t=y,U_t=u)P_{yx}(u)\\
&= \beta\sum_{y\in X}\sum_{u\in U(y)}f_\beta(\sigma,\pi;y,u)P_{yx}(u)
\end{aligned}$$

Thus

$$\sum_{u\in U(x)} f_\beta(\sigma,\pi;x,u) = \beta\sum_{y\in X}\sum_{u\in U(y)}f_\beta(\sigma,\pi;y,u)P_{yx}(u), \quad \text{for } x\neq\sigma \tag{8}$$

If $x=\sigma$ then, in the first equality above we have an additional term

$$(1-\beta)P_\sigma^\pi(X_{t=1}=x,U_{t=1}=u)=(1-\beta)$$

Therefore for $x=\sigma$ we can rewrite (8) as

$$\sum_{u\in U(x)} f_\beta(\sigma,\pi;x,u) = \beta\sum_{y\in X}\sum_{u\in U(y)}f_\beta(\sigma,\pi;y,u)P_{yx}(u)+(1-\beta) \tag{9}$$

From (5) we can see that for $x\neq\sigma$ (5) equals to (8) and for $x=\sigma$ (5) equals to (9). (6) and (7) is established from the definition of $f_\beta(\sigma,\pi;y,u)$. Therefore the proof is completed.

■

Now suppose we have a set of numbers $\rho(y,u)$ that satisfy (5,6,7), define $Q^{\beta}(\sigma)$ to be the set of $\rho = \{\rho(y,u)\}$

$$Q^{\beta}(\sigma) = \left\{ \rho : \left(\begin{array}{c} \sum_{y \in X} \sum_{u \in U(y)} \rho(y,u)(\delta_x(y) - \beta P_{yx}(u)) = (1-\beta)\sigma(x), \forall x \in X \\ \rho(y,u) \geqslant 0, \forall y, u \end{array} \right) \right\} \tag{10}$$

where σ is an initial distribution.

In this paper we give ρ two different representations:

1. ρ is the set of $\rho = \{\rho(y,u)\}$ as was defined above

2. ρ is the vector of length $|X| * |U|$ where in each coordinate (y,u), $\{y \in X,\, u \in U\}$ we have the value of $\rho(y,u)$.

By summing the first constraint over x we note that $\sum_{y,u} \rho(y,u) = 1$, for $\rho \in Q^{\beta}(\sigma)$, so ρ, satisfying the above constraints, can be considered as a probability measure.

Since in many cases we are interested in which control to use in a given state, and not in the value of ρ in this state, or what is the probability of using a given control in a given state, then it is convenient to define $\mu_y(u)$, where

$$\mu_y(u) = \frac{\rho(y,u)}{\sum_{u \in U(y)} \rho(y,u)}, y \in X, u \in U(y) \tag{11}$$

Our optimization problem is:

Find π such that $C_{\beta}(\sigma,\pi)$ is minimal,

Subject to

$\overline{D}_{\beta}(\sigma,\pi) \leq \overline{V}$

It now follows from the definition of $Q^{\beta}(\sigma)$, from ([1], Theorem 3.2) and from the representation of the cost in (3) that the value of *COP* can be obtained using this program.

We express (σ,π) by $\rho \in Q^{\beta}(\sigma)$, we replace $C_{\beta}(\sigma,\pi)$ by $C_{\beta}(\rho)$ and $\overline{D}_{\beta}(\sigma,\pi)$ by $\overline{D}_{\beta}(\rho)$.

So we can rewrite this program as a linear program (LP) as follows

$LP_1^\beta(\sigma)$: Find ρ such that $C_\beta(\rho) = \sum_{x,u} c(x,u)\rho(x,u)$ is minimal,

Subject to

$\overline{D}_\beta(\rho) = \sum_{x,u} \overline{d}(x,u)\rho(x,u) \leq \overline{V}$

$\rho \in Q^\beta(\sigma)$.

The last constraint is linear by definition (10).

So we can derive the following theorem ([1], Theorem 3.3)

Equivalence between *COP* and the *LP*

Theorem ([1], Theorem 3.3): *Consider a finite CMDP, then*

- *For any $f_\beta(\sigma,\mu)$ there exists $\rho \in Q^\beta(\sigma)$ such that $\rho = f_\beta(\sigma,\mu)$, and conversely for any $\rho \in Q^\beta(\sigma)$ there exists μ such that $\rho = f_\beta(\sigma,\mu)$.*
- *$LP_1^\beta(\sigma)$ is feasible if and only if COP is. Assume that COP is feasible. Then there exists an optimal solution ρ^* for $LP_1^\beta(\sigma)$, and the stationary policy μ, that is related to ρ^* through (11), is optimal for COP.*

5 The Optimal Policy Properties for CMDP

In this chapter we will derive a number of theorems that describe important properties of CMDP problems solutions. The most valuable theorems are based on Karush-Kuhn Tucker conditions and prove in a very elegant way a statement that looks intuitive but not trivial for formal proof. We begin with considering a problem with two control levels and one constraint and extend it to arbitrary, but finite, number of control levels and one constraint.

5.1 The Karush-Kuhn-Tucker (KKT) Conditions

Definitions:

- z is a vector of length n.
- b is a vector of length m.
- s is a vector of variables of length n.
- A is an $m \times n$ matrix.
- $\cdot$ denotes a scalar product.

Consider the following linear programming problem.

$$Minimize \ \ z \cdot s \tag{12}$$

$$Subject\ to\ A \cdot s \geq b \tag{13}$$

$$s \geq 0 \tag{14}$$

The Karush-Kuhn-Tucker (KKT) conditions can be stated as follows:

There exist w and v so that

$$A \cdot s \geq b, \; s \geq 0 \tag{15}$$

$$w \cdot A + v = z, \; w \geq 0, \; v \geq 0 \tag{16}$$

$$w \cdot (A \cdot s - b) = 0, \; v \cdot s = 0 \tag{17}$$

The first condition (15) merely states that the candidate point must be feasible; that is, it must satisfy the constraints of the problem. This is usually referred to as primal feasibility. The second condition (16) is usually referred to as dual feasibility, since it corresponds to a feasibility of the problem closely related to the original one. Here w and v are called the Lagrangian multipliers (or dual variables) corresponding to the constraints $A \cdot s \geq b$ and $s \geq 0$ respectively.

Finally, the third condition (17) is usually referred to as complementary slackness.

Theorem ([3], KKT Conditions, inequality case):

Any solution s that satisfies conditions (15)-(17) is an optimal solution of the Linear Programming problem (12)-(14).

Consider the following linear programming problem with equality constraints.

$$Minimize \;\; z \cdot s \tag{18}$$

$$Subject \; to \; A \cdot s = b \tag{19}$$

$$s \geq 0 \tag{20}$$

By changing the equality into two inequalities of the form $A \cdot s \geq b$ and $-A \cdot s \geq -b$, the KKT conditions developed earlier would simplify to

$$A \cdot s = b,\ s \geq 0 \tag{21}$$

$$w \cdot A + v = z,\ w\ unrestricted,\ v \geq 0 \tag{22}$$

$$v \cdot s = 0 \tag{23}$$

Theorem ([3], KKT Conditions, equality case):

Any solution s that satisfies conditions (21)-(23) is an optimal solution of the Linear Programming problem (18)-(20).

For linear programming problems, these conditions are both necessary and sufficient and hence form an important characterization of optimality.

The main difference between these conditions for the inequality problem is that the Lagrangian multiplier vector (or dual vector) w corresponding to the constraint $A \cdot s = b$ is unrestricted in sign.

The major objective of this paper is to solve a communications problem (section 3), for which, as we will see later, only equality conditions are requiered. Therefore in this section we will derive general theorems for equality case only.

Now let's represent our optimization problem $(LP_1^{\beta}(\sigma))$ in the form described by (21)-(23).

Definitions:

- z is a vector of length $n = |X| * |U|$, this is a vector of the cost, which we represent as a row vector.

- b is a vector of length m,

$$b = \begin{bmatrix} 1-\beta \\ 0 \\ . \\ 0 \\ \alpha \end{bmatrix} \tag{24}$$

We limit b to the case where $\sigma(x) = \delta_{x=1}$. Later we will see that this is good enough.

- α is a value of the constraint, where we use only one constraint, in other words α is a vector 1x1.
- ρ is a vector of variables of length $n = |X| * |U|$ where each coordinate (x, u) gives the value of $\rho(x, u)$.
- A is an $m \times n$ matrix,

$$A = \begin{bmatrix} (1-\beta P_{x_1x_1}(u_0)) & (1-\beta P_{x_1x_1}(u_1)) & & -\beta P_{x_Nx_1}(u_0) & -\beta P_{x_Nx_1}(u_1) \\ -\beta P_{x_1x_2}(u_0) & -\beta P_{x_1x_2}(u_1) & & -\beta P_{x_Nx_2}(u_0) & -\beta P_{x_Nx_2}(u_1) \\ . & . & \ldots\ldots & . & . \\ -\beta P_{x_1x_N}(u_0) & -\beta P_{x_1x_N}(u_1) & & (1-\beta P_{x_Nx_N}(u_0)) & (1-\beta P_{x_Nx_N}(u_1)) \\ d(x_1,u_0) & d(x_1,u_1) & & d(x_N,u_0) & d(x_N,u_1) \end{bmatrix} \tag{25}$$

We can note that the first N rows of the matrix A represent the following part of equation (5)

$$A_{(N\times N)} = \sum_{y\in X} \sum_{u\in U(y)} (\delta_x(y) - \beta P_{yx}(u)), \quad \forall x \in X$$

The last row of the matrix A represents the cost vector related to the constraint, where u_0, u_1 are the available actions and $u_0, u_1 \in U$.

So by using (5-7) and definitions of $z, b, \alpha, \rho, A, C_\beta(\sigma, \pi)$ we can rewrite (18) - (20) as

$$Minimize \;\; z \cdot \rho \tag{26}$$

$$Subject\ to\ A \cdot \rho = b \tag{27}$$

$$\rho \geq 0 \tag{28}$$

and (21)-(23) as

$$A \cdot \rho = b, \;\; \rho \geq 0 \tag{29}$$

$$w \cdot A + v = z, \;\; w\ unrestricted, \;\; v \geq 0 \tag{30}$$

$$v \cdot \rho = 0 \tag{31}$$

5.2 The Fundamental Theorems

Now we will derive the fundamental theorems. These theorems describe the structure of the optimal solution for constrained Markov decision process problems and are used as a basis for innovative algorithm that will be derived later.

In order to derive our new theorems we will use ([1], Theorem 3.8).

Theorem ([1], Theorem 3.8) (*Bounds on the number of randomizations*)

If the constrained optimization problem is feasible then there exists an optimal stationary policy π such that the total number of randomizations that it uses is at most the number of constraints.

Let X_0 and X_1 be disjoint sets of states so that $|X_0| + |X_1| = N - 1$. Let x_i be the only state not in $X_0 \cup X_1$. Let

$$\pi^q(x) = \left\{ \begin{array}{c} \delta_{u_0}, \quad x \in X_0 \\ \delta_{u_1}, \quad x \in X_1 \\ (1-q)\delta_{u_0} + q\delta_{u_1}, \quad x = x_i \end{array} \right\} \tag{32}$$

Fix α_0 so that COP is feasible. By theorem ([1], Theorem 3.8), $\pi^{q_{\alpha_0}}$ is optimal, where $\pi^{q_{\alpha_0}}$ is randomized only in state x_i, and $\alpha_0 = D(\sigma, \pi^{q_{\alpha_0}})$.

Definitions:

- $u_0, u_1 \in U$.
- $0 \leq q_{\alpha_0} \leq 1$.
- $\alpha_{\min} = \inf\limits_{0\leq q\leq 1} D(\sigma, \pi^q)$.
- $\alpha_{\max} = \sup\limits_{0\leq q\leq 1} D(\sigma, \pi^q)$.
- π^{q_α} is the policy that is the same as $\pi^{q_{\alpha_0}}$ except in state x_i and π^{q_α} is chosen so that the value of the constraint is α, where $\alpha_{\min} \leq \alpha \leq \alpha_{\max}$. This is possible because $D(\sigma, \pi^{q_\alpha})$ is continuous in q_α. The continuity will be proven later.
- $\alpha = D(\sigma, \pi^{q_\alpha}), \;\; \alpha_{\min} \leq \alpha \leq \alpha_{\max}$.

Lemma 1: $D(\sigma, \pi^{q_\alpha})$ is continuous in q_α.

From Lemma 1 we can conclude that $\forall\ \alpha_{\min} \leq \alpha \leq \alpha_{\max}$, $\exists\ 0 \leq q_\alpha \leq 1$ s.t. $D(\sigma, \pi^{q_\alpha}) = \alpha$.

Proof. $D_\beta(\sigma, \pi^{q_\alpha}) = (1-\beta)\sum_{t=1}^{\infty} \beta^{t-1} E_\sigma^{\pi^{q_\alpha}} d(X_t, U_t)$

Given ε, fix N s.t.

$(1-\beta)\sum_{t=N}^{\infty} \beta^{t-1} \max_{X,U} |d(X_t, U_t)| < \frac{\varepsilon}{4}$

$\Rightarrow \left|D_\beta(\sigma, \pi^{q_\alpha}) - D_\beta(\sigma, \pi^{q_\alpha+\delta})\right| \leq \frac{\varepsilon}{2} +$

$+ \left|(1-\beta)\sum_{t=1}^{N-1} \beta^{t-1} E_\sigma^{\pi^{q_\alpha}} d(X_t, U_t) - (1-\beta)\sum_{t=1}^{N-1} \beta^{t-1} E_\sigma^{\pi^{q_\alpha+\delta}} d(X_t, U_t)\right|.$

$\left|(1-\beta)\sum_{t=1}^{N-1} \beta^{t-1} E_\sigma^{\pi^{q_\alpha}} d(X_t, U_t) - (1-\beta)\sum_{t=1}^{N-1} \beta^{t-1} E_\sigma^{\pi^{q_\alpha+\delta}} d(X_t, U_t)\right| =$

$|(1-\beta)\sum_{t=1}^{N-1}\sum_{x\in X}\sum_{u\in U} \beta^{t-1} P_{\sigma,x}^t(\pi^{q_\alpha}) d(x, \pi^{q_\alpha}(x)) -$

$(1-\beta)\sum_{t=1}^{N-1}\sum_{x\in X}\sum_{u\in U} \beta^{t-1} P_{\sigma,x}^t(\pi^{q_\alpha+\delta}) d(x, \pi^{q_\alpha+\delta}(x))|$

where $P_{yx}^t(\pi^q) = P_{yx}^{t-1}(\pi^q)[(1-q)P(\pi^0) + qP(\pi^1)]$.

Thus, the expression $\left|(1-\beta)\sum_{t=1}^{N-1} \beta^{t-1} E_\sigma^{\pi^{q_\alpha}} d(X_t, U_t) - (1-\beta)\sum_{t=1}^{N-1} \beta^{t-1} E_\sigma^{\pi^{q_\alpha+\delta}} d(X_t, U_t)\right|$

we can represent as a $\left|(1-\beta)\sum_{t=1}^{N-1}\beta^{t-1}poly(t,q_\alpha)-(1-\beta)\sum_{t=1}^{N-1}\beta^{t-1}poly(t,q_\alpha+\delta)\right|$, where $poly(t,q)$ is a polynom in q, of order t.

Since this is a finite sum we can choose δ so that

$\left|(1-\beta)\sum_{t=1}^{N-1}\beta^{t-1}poly(t,q_\alpha)-(1-\beta)\sum_{t=1}^{N-1}\beta^{t-1}poly(t,q_\alpha+\delta)\right|<\frac{\varepsilon}{2}$

therefore $\left|D_\beta(\sigma,\pi^{q_\alpha})-D_\beta(\sigma,\pi^{q_\alpha+\delta})\right|<\varepsilon$. So $\forall\varepsilon>0$, $\exists\delta>0$

so that $\left|D_\beta(\sigma,\pi^{q_\alpha})-D_\beta(\sigma,\pi^{q_\alpha+\delta})\right|<\varepsilon$ therefore we proved the Lemma. ■

With a some abuse of notation we will use $\pi^{q_{\alpha_0}}$ instead of $\pi^{q_{\alpha_0}}(x)$.

Theorem 1 *Consider a constrained Markov decision process problem with two control levels and one constraint, where* $0<q_{\alpha_0}<1$. *Then for each* $\alpha_{\min}\le\alpha\le\alpha_{\max}$, π^{q_α} *is an optimal policy.*

Proof. The proof will be done by KKT conditions.

For convenience let us rearrange the states so that $X_0=\{x_1,...,x_{i-1}\}$, $X_1=\{x_{i+1},...,x_N\}$.

Since randomization is allowed only in state x_i then under $\pi^{q_{\alpha_0}}$ for any $x_j\neq x_i$ there exists at most one u so that $\rho(x_j,u)\neq 0$. Moreover, if $\rho(x_i,u)=0$ under $\pi^{q_{\alpha_0}}$, then this is true also under any π^{q_α}, $\alpha_{\min}\le\alpha\le\alpha_{\max}$.

From the ([1], Theorem 3.8) $\pi^{q_{\alpha_0}}$ is optimal so KKT conditions are satisfied for $\pi^{q_{\alpha_0}}$. Denote the Lagrange multipliers by $v^{\pi^{q_{\alpha_0}}}$, $w^{\pi^{q_{\alpha_0}}}$ for $\pi^{q_{\alpha_0}}$, and $v^{\pi^{q_\alpha}}$, $w^{\pi^{q_\alpha}}$ for π^{q_α}.

Arrange the elements of $\rho^{\pi^{q_{\alpha_0}}}=\rho(\pi^{q_{\alpha_0}})=f_\beta(\sigma,\pi^{q_{\alpha_0}})$ as

$$\begin{aligned}\rho^{\pi^{q_{\alpha_0}}} &= (\rho(x_1,u_0),\rho(x_1,u_1),...,\rho(x_i,u_0),\rho(x_i,u_1),...,\rho(x_N,u_0),\rho(x_N,u_1))\\ &= (\rho(x_1,u_0),0,...,\rho(x_{i-1},u_0),0,\rho(x_i,u_0),\rho(x_i,u_1),0,\rho(x_{i+1},u_1),...,0,\rho(x_N,u_1))\end{aligned}$$

From (31) $<v^{\pi^{q_{\alpha_0}}},\rho^{\pi^{q_{\alpha_0}}}>=0$. Because $\rho(x_i,u_0)>0$ and $\rho(x_i,u_1)>0$, then $v_i^0=0$ and $v_i^1=0$, and $v=0$ in all coordinates where $\rho(x,u)>0$, so

$$v^{\pi^{q_{\alpha_0}}}=(0,v_1^1,...,0,v_{i-1}^1,0,0,v_{i+1}^0,0,...,v_N^0,0)$$

Let's check KKT conditions (29-31) for policy π^{q_α}.

Since changing q_α does not change the structure, except perhaps to create more "0", it follows that $\rho^{\pi^{q_\alpha}} = f_\beta(\sigma, \pi^{q_\alpha})$ takes the form

$$\rho^{\pi^{q_\alpha}} = (\rho'(x_1, u_0), 0, ..., \rho'(x_{i-1}, u_0), 0, \rho'(x_i, u_0), \rho'(x_i, u_1), 0, \rho'(x_{i+1}, u_1), ..., 0, \rho'(x_N, u_1)) \tag{33}$$

We choose $w^{\pi^{q_\alpha}} = w^{\pi^{q_{\alpha_0}}}$and $v^{\pi^{q_\alpha}} = v^{\pi^{q_{\alpha_0}}}$

Equation (29) is satisfied because $\rho^{\pi^{q_\alpha}}$ is a feasible solution by definition.

Equation (30) is satisfied because $w^{\pi^{q_\alpha}} = w^{\pi^{q_{\alpha_0}}}$and $v^{\pi^{q_\alpha}} = v^{\pi^{q_{\alpha_0}}}$, and A and z are the same for each choice of α.

Equation (31) is satisfied because for $v^{\pi^{q_\alpha}} = v^{\pi^{q_{\alpha_0}}}$ it doesn't matter what is a ratio between $\mu_{x_i}(u_0)$ and $\mu_{x_i}(u_1)$, and because $\forall m \neq i$ $\mu_{x_m}^{\pi^\alpha}(u_0) = \mu_{x_m}^{\pi^{\overline{q}_{\alpha_0}}}(u_0)$ $(\mu_{x_m}^{\pi^\alpha}(u_1) = \mu_{x_m}^{\pi^{\overline{q}_{\alpha_0}}}(u_1))$, therefore

$$< v^{\pi^{q_\alpha}}, \rho^{\pi^{q_\alpha}} >=< v^{\pi^{q_{\alpha_0}}}, \rho^{\pi^{q_{\alpha_0}}} >= 0$$

■

Now we will extend this theorem from two control levels to an arbitrary, but finite, number of control levels.

Fix α_0 so that COP is feasible. By theorem ([1], Theorem 3.8), $\pi^{\overline{q}_{\alpha_0}}$ is optimal, where $\pi^{\overline{q}_{\alpha_0}}$ is randomized only in state x_i, and $\alpha_0 = D(\sigma, \pi^{\overline{q}_{\alpha_0}})$.

Definitions:

- $u_0, u_1, ..., u_s \in U$.
- Let X_0, $X_1, ..., X_s$ be disjoint sets of states so that $|X_0| + |X_1| + ... + |X_s| = N - 1$.

Let x_i be the only state not in $X_0 \cup X_1 \cup ... \cup X_s$. Let

$$\pi^{\overline{q}_{\alpha 0}}(x) = \left\{ \begin{array}{cc} \delta_{u_0}, & x \in X_0 \\ \delta_{u_1}, & x \in X_1 \\ & . \\ & . \\ & . \\ \delta_{u_s}, & x \in X_s \\ \sum_i q^i_{\alpha 0}\delta_{u_i}, & x = x_i \end{array} \right\}$$

- $0 \leq q^j_{\alpha 0} \leq 1, \forall i$ or in another representation $0 \leq \overline{q}_{\alpha 0} \leq 1$.
- $\sum_j q^j_{\alpha 0} = 1$, and at most two $q^j_{\alpha 0} > 0$.
- $\alpha_{\min} = \inf\limits_{\overline{q}_{\alpha 0}} D(\sigma, \pi^{\overline{q}_{\alpha 0}})$ s.t. the same two $q^j_{\alpha 0}$ as above are non zero.
- $\alpha_{\max} = \sup\limits_{\overline{q}_{\alpha 0}} D(\sigma, \pi^{\overline{q}_{\alpha 0}})$ s.t. the same two $q^j_{\alpha 0}$ as above are non zero.
- $\pi^{\overline{q}_\alpha}$ is the policy that is the same as $\pi^{\overline{q}_{\alpha 0}}$ except in state x_i and $\pi^{\overline{q}_\alpha}$ is chosen so that the value of the constraint is α, where $\alpha_{\min} \leq \alpha \leq \alpha_{\max}$, and zeros of $\overline{q}_\alpha$ are also zeros of $\overline{q}_{\alpha 0}$. This is possible because α, $\alpha \in [\alpha_{\min}, \alpha_{\max}]$, is continuous in $\overline{q}_\alpha$. The continuity will be proven later.
- $\alpha = D(\sigma, \pi^{\overline{q}_\alpha})$.

Lemma 2: $D(\sigma, \pi^{\overline{q}_\alpha})$ is continuous in $\overline{q}_\alpha$.

Proof. The proof is similar to the proof of Lemma 1. ■

From Lemma 2 we can conclude that $\forall\ \alpha_{\min} \leq \alpha \leq \alpha_{\max}$, $\exists\ 0 \leq \overline{q}_\alpha \leq 1$ where zeros of $\overline{q}_\alpha$ are also zeros of $\overline{q}_{\alpha 0}$, $\sum_j q^j_\alpha = 1$ s.t. $D(\sigma, \pi^{\overline{q}_\alpha}) = \alpha$.

Theorem 2 *Consider a constrained Markov decision process problem with arbitrary, but finite, number of control levels and one constraint, where* $0 < \overline{q}_{\alpha 0} < 1$. *Then for each* $\alpha_{\min} \leq \alpha \leq \alpha_{\max}$, $\pi^{\overline{q}_\alpha}$ *is an optimal policy.*

Proof. The proof will be done by KKT conditions.

For convenience let us rearrange the states so that $X_0 = \{x_1, ..., x_j\}, ...,$

$X_r = \{x_k, ..., x_{i-1}\}, .., X_s = \{x_s, ..., x_N\}$.

Since randomization is allowed only in state x_i then for any $x_j \neq x_i$ there exists at most one u so that $\rho(x_j, u) \neq 0$.

From the ([1], Theorem 3.8) $\pi^{\overline{q}_{\alpha_0}}$ is optimal so KKT conditions are satisfied for $\pi^{\overline{q}_{\alpha_0}}$. Denote the Lagrange multipliers by $v^{\pi^{\overline{q}_{\alpha 0}}}$, $w^{\pi^{\overline{q}_{\alpha 0}}}$ for $\pi^{\overline{q}_{\alpha_0}}$, and $v^{\pi^{\overline{q}_{\alpha}}}$, $w^{\pi^{\overline{q}_{\alpha}}}$ for $\pi^{\overline{q}_{\alpha}}$.

Arrange the elements of $\rho^{\pi^{\overline{q}_{\alpha 0}}} = \rho(\pi^{\overline{q}_{\alpha_0}}) = f_\beta(\sigma, \pi^{\overline{q}_{\alpha_0}})$ as

$$\begin{aligned}
\rho^{\pi^{q_{\alpha 0}}} &= (\rho(x_1, u_0), \rho(x_1, u_1), ..., \rho(x_1, u_s), ..., \rho(x_i, u_p), ..., \rho(x_i, u_j), ..., \rho(x_N, u_{s-1}), \rho(x_N, u_s)) \\
&= (\rho(x_1, u_0), 0, ..., 0, ..., \rho(x_i, u_p), 0, ..., 0, \rho(x_i, u_j), ..., 0, \rho(x_N, u_s))
\end{aligned}$$

From (31) $< v^{\pi^{\overline{q}_{\alpha 0}}}, \rho^{\pi^{\overline{q}_{\alpha 0}}} >= 0$. Because $\rho(x_i, u_p) > 0$, $\rho(x_i, u_j) > 0$ then $v_i^p = 0$, $v_i^j = 0$ and $v = 0$ in all coordinates where $\rho(x, u) > 0$. Thus,

$$v^{\pi^{q_{\alpha 0}}} = (0, v_1^1, ..., v_1^s, ..., 0, 0, ..., v_N^{s-1}, 0)$$

Let's check KKT conditions (29-31) for policy $\pi^{\overline{q}_{\alpha}}$.

Since changing $\overline{q}_\alpha$ does not change the structure, except perhaps to create more "0", it follows $\rho^{\pi^{\overline{q}_{\alpha}}} = f_\beta(\sigma, \pi^{\overline{q}_{\alpha}})$ takes the form

$$\rho^{\pi^{q_{\alpha}}} = (\rho'(x_1, u_0), 0, ..., 0, ..., \rho'(x_i, u_p), 0, ..., 0, \rho'(x_i, u_j), ..., 0, \rho'(x_N, u_s)) \qquad (34)$$

We choose $w^{\pi^{\overline{q}_{\alpha}}} = w^{\pi^{\overline{q}_{\alpha 0}}}$and $v^{\pi^{\overline{q}_{\alpha}}} = v^{\pi^{\overline{q}_{\alpha 0}}}$

Equation (29) is satisfied because $\rho^{\pi^{\overline{q}_{\alpha}}}$ is a feasible solution by definition.

Equation (30) is satisfied because $w^{\pi^{\overline{q}_{\alpha}}} = w^{\pi^{\overline{q}_{\alpha 0}}}$and $v^{\pi^{\overline{q}_{\alpha}}} = v^{\pi^{\overline{q}_{\alpha 0}}}$, and A and z are the same for each choice of α.

Equation (31) is satisfied because for $v^{\pi^{\overline{q}_\alpha}} = v^{\pi^{\overline{q}_{\alpha_0}}}$ it doesn't matter what is a ratio between $\mu_{x_i}(u_p)$ and $\mu_{x_i}(u_j)$, and $\forall m \neq i$ $\mu^{\pi^{\overline{q}_\alpha}}_{x_m}(u_p) = \mu^{\pi^{\overline{q}_{\alpha_0}}}_{x_m}(u_p)$, and $\alpha_{\min}$ and $\alpha_{\max}$ are obtained without changing actions therefore

$$< v^{\pi^{\overline{q}_\alpha}}, \rho^{\pi^{\overline{q}_\alpha}} >=< v^{\pi^{\overline{q}_{\alpha_0}}}, \rho^{\pi^{\overline{q}_{\alpha_0}}} >= 0$$

■

6 The Optimal Policy Properties for Power Saving Problem

Now we will extend our analysis to the power saving problem. We start with a two control levels problem and one constraint, and will extend it to arbitrary, but finite, number of control levels. Let define u_0 and u_1 as power control levels, where $u_0 < u_1$.

Assumption 1: For each α for which there exists a feasible solution, there exists a unique optimal solution.

Assumption 2 (monotonicity assumption): For strictly increasing value of the ratio $\frac{\mu_{x_i}(u_1)}{\mu_{x_i}(u_0)}$, the value of the constraint α is strictly increasing, where $\alpha \in [\alpha_{\min}, \alpha_{\max}]$ and x_i is the state with randomization.

In the problem defined by diagram on Figure 2 we will show analytically that Assumption 2 is satisfied.

Let's consider the problem: $\forall x_j$, where $0 \leq j < i$ the transmission is by u_1, $\forall x_j$, where $i < j \leq N$ the transmission is by u_0, at x_i we have randomization between u_0 and u_1. The transitions are allowed between neighborhood states only, where the state is defined as the number of messages in the buffer. Define two systems (Figure 3) where in system 1, $P(u = u_1|x = x_i) = \kappa$ and in system 2, $P(u = u_1|x = x_i) = \kappa'$ where $\kappa' > \kappa$. So we can say that ratio $\frac{\mu_{x_i}(u_1)}{\mu_{x_i}(u_0)}$ at system 2 is higher than at system 1. Using a coupling, we assume that at any time t, a success in system 1 implies a success in system 2. Two systems have the same buffer and channel. Server 1 transmits with $u = u_1$ in x_i with probability κ and server 2 transmits with $u = u_1$ in x_i with probability κ'. Assume that $P(success|u = u_1) > P(success|u = u_0) > 0.5$.

Lemma 3 Consider the two systems as described in Figure 2 and 3. The average power usage at the second system is higher than at the first one, where average power this is the value of the constraint α.

Proof. Till we are not at state x_i, the power usage in both systems is the same, because

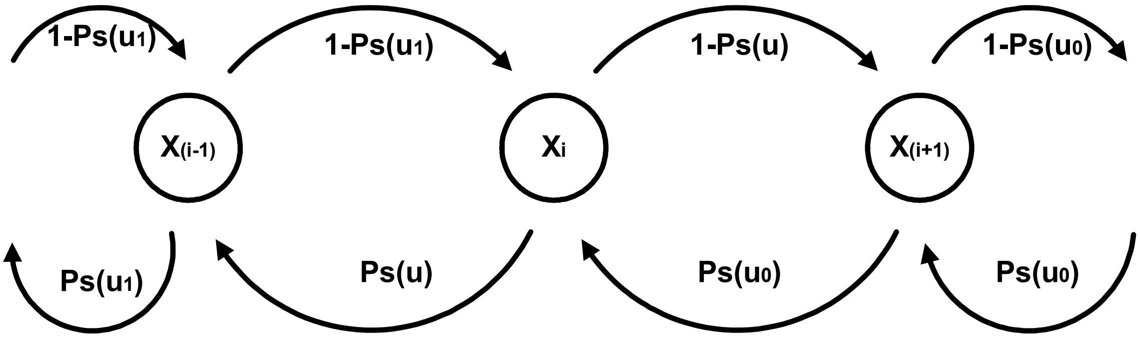


Figure 2: Markov Chain for Monotonicity Proof

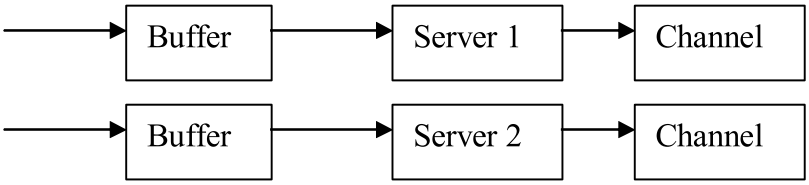


Figure 3: Coupled System

except state x_i systems are same. When we arrive to state x_i we have three possibilities for systems behaviors:

1). No one of the transmissions in both servers is successful, so the two systems continue to be the same.

2). Both systems have succeeded, so the two systems continue to be the same.

3). First system failed to transmit but the second succeeded to transmit.

Note, the two systems are initially at the same state, so the first time that 3) can happen, is when both systems are at state x_i.

Until 3) happens we have two equal systems. So let's concentrate on the third case.

If 3) has happened, at system 2 we always transmit with power u_1 till again we will arrive to state x_i, and at system 1 we always transmit with power u_0 till again we will arrive to state x_i. Let's denote the period of time that both systems are in different states by τ. During the period of time τ the power usage at the second system is higher that at the first one and when this time is finished both system always meet at the same state (from diagram 2 we can see that τ can finished only at states x_0 or x_N). So after time $\tau > 0$ both systems again at the same state and such we begin the same procedure from the beginning.

Because an average power usage is equal to the sum of all instantaneous power usages divided by a period of measured time then we can conclude that the average power usage at the second system is higher than at the first one as we required to prove. ■

For more complicates Markov chains we will show numerically that Assumption 2 is satisfied (section 7).

Fix $\alpha^{\pi_k^q}$ so that COP is feasible. By theorem ([1], Theorem 3.8), π_k^q is optimal, where π_k^q is randomized only in state x_i, and $\alpha^{\pi_k^q} = D(\sigma, \pi_k^q)$.

Definitions:

- k is a number of states where $\mu_y(u_1) = 1$ ($\mu_y(u_0) = 0$).

- Let X_0^1 and X_1^1 be disjoint sets of states so that $|X_0^1| = N - 1 - k$, and $|X_1^1| = k$. Let $x_{i(k)}$ be the only state not in $X_0^1 \cup X_1^1$, where $i(k)$ denotes index i as function of k.

$$\pi_k^{q_1}(x) = \left\{ \begin{array}{c} \delta_{u_0}, \quad x \in X_0^1 \\ \delta_{u_1}, \quad x \in X_1^1 \\ (1 - q_1)\delta_{u_0} + q_1\delta_{u_1}, \quad x = x_{i(k)} \end{array} \right\}$$

- Let X_0^2 and X_1^2 be disjoint sets of states so that $|X_0^2| = N - 2 - k$, and $|X_1^2| = k + 1$. Let $x_{i(k+1)}$ be the only state not in $X_0^2 \cup X_1^2$, where $x_{i(k+1)} \neq x_{i(k)}$, $X_0^1 = \{X_0^2, x_{i(k+1)}\}$, $X_1^2 = \{X_1^1, x_{i(k)}\}$.

$$\varsigma_{k+1}^{q_2}(x) = \left\{ \begin{array}{c} \delta_{u_0}, \quad x \in X_0^2 \\ \delta_{u_1}, \quad x \in X_1^2 \\ (1 - q_2)\delta_{u_0} + q_2\delta_{u_1}, \quad x = x_{i(k+1)} \end{array} \right\}$$

Note, for $q_2 = 0$ and $q_1 = 1$, $\varsigma_{k+1}^{q_2}(x)$ is equal to $\pi_k^{q_1}(x)$.

- Let X_0^3 and X_1^3 be disjoint sets of states so that $|X_0^3| = N - 1 - j$, and $|X_1^3| = j$. Let x_r be the only state not in $X_0^3 \cup X_1^3$.

$$\xi_j^{q_3}(x) = \left\{ \begin{array}{c} \delta_{u_0}, \quad x \in X_0^3 \\ \delta_{u_1}, \quad x \in X_1^3 \\ (1 - q_3)\delta_{u_0} + q_3\delta_{u_1}, \quad x = x_{r(j)} \end{array} \right\} \tag{35}$$

- $\overline{\alpha} = \sup\limits_{q_1} D(\sigma, \pi_k^{q_1})$ and $\underline{\alpha} = \inf\limits_{q_1} D(\sigma, \pi_k^{q_1})$.

In the following two theorems we are talking about range of α for which a feasible solution exists.

Theorem 3 *Consider a finite CMDP so that Assumptions 1 and 2 are satisfied. If policy* $\pi_k^{q_1}$ *is optimal for some* $\alpha^{\pi_k^{q_1}} \leq \overline{\alpha}$ *then there exists some* $\alpha^{\varsigma_{k+1}^{q_2}} > \overline{\alpha}$ *so that* $\varsigma_{k+1}^{q_2}$ *is optimal for* $\alpha^{\varsigma_{k+1}^{q_2}}$*.*

Proof. The proof will be done by contradiction.

By Theorem 1 $\pi_k^{q_1}$ is optimal for $\underline{\alpha} \leq \alpha^{\pi_k^{q_1}} \leq \overline{\alpha}$. From Monotonicity (Assumption 2), $\pi_k^{q_1}$ is optimal for $\alpha = \overline{\alpha}$ when $q_1 = 1$.

Assume that $\varsigma_{k+1}^{q_2}$ is not the optimal solution of the problem for any α_i such that $\alpha^{\varsigma_{k+1}^{q_2}} \geq \alpha_i > \overline{\alpha}$, where $\alpha^{\varsigma_{k+1}^{q_2}}$ this is a number so that $u_1 \geq \alpha^{\varsigma_{k+1}^{q_2}} > \overline{\alpha}$. Take $\alpha_i \downarrow \overline{\alpha}$. By Assumption 1 there exists another optimal solution for these α_i, say $\xi_j^{q_3}$, as defined in (35).

If $\xi_j^{q_3}$ is optimal, $\alpha_i \downarrow \overline{\alpha}$, then by Assumption 2 and Theorem 1, $\xi_j^{q_3}$ is optimal for $\alpha = \overline{\alpha}$ as well.

So for $\alpha = \overline{\alpha}$ we have gotten two different optimal solutions: $\xi_j^{q_3}$ and $\pi_k^{q_1=1}$, in contradiction to the Assumption 1, therefore we proved the Theorem. ■

Theorem 4 *Consider a finite CMDP so that Assumptions 1 and 2 are satisfied. For each* $k \in [0, N-1]$*, as defined above, there exists* α *so that* $\pi_k^{q_1}$ *is an optimal policy for constrained Markov decision process problem with two control levels and one constraint.*

Proof. The proof will be done by induction.

1. $k = 0$. This means $\forall (y \neq x_{i(k)}) \in [1, N]$, $\mu_y(u_1) = 0$, and only in one state $x_{i(k)}$, can be $\mu_{x_{i(k)}}(u_1) \neq 0$.

Let's denote by α' the minimal α that is needed for $\mu_{x_i}(u_1) = 1$. Assume that the available average power is less than α', denote it by α''. For α'', $\mu_y(u_1) \neq 1$. By theorem ([1], Theorem 3.8) we have that for $k = 0$ and $\alpha = \alpha''$, $\pi_{k=0}^{q_1}$ is the optimal policy.

2. Assume that for each $\alpha^{\pi_k^{q_1}}$, $0 \leq k = j < N$, $\pi_{k=j}^{q_1}$ is the optimal policy.

3. Let's prove that for $\alpha^{\pi_k^{q_1}}$, $k = j + 1$, $\pi_{k=j+1}^{q_1}$ is the optimal policy as well.

From Theorem 3, if $\pi_k^{q_1}$ is optimal for $k = j$ then there exists $\alpha^{\varsigma_{j+1}^{q_2}} > \alpha^{\pi_j^{q_1}}$ so that $\varsigma_{j+1}^{q_2}$ is an optimal policy, but $\varsigma_{j+1}^{q_2}$ is equal to $\pi_{j+1}^{q_1}$. When we say that $\varsigma_{j+1}^{q_2}$ is equal to $\pi_{j+1}^{q_1}$, we mean that $X_0^1 = X_0^2$, $X_1^1 = X_1^2$ and a randomization in the same state, q may be different. Therefore we proved the Theorem.

Because the length of Markov chain is N the maximal k is equal to $N - 1$. ■

Conclusions: If Assumptions 1 and 2 are satisfied, then by Theorems 1,3,4 can be concluded that if for each α there exists a unique optimal solution then it has the following form:

For $\alpha < u_0$ is no feasible solution because even if we always use at each state a power that equal to u_0 we get that the minimal average required power is

$$\sum_i \rho(x_i, u_0)u_0 = u_0 \sum_i \rho(x_i, u_0) = u_0$$

1. For $\alpha = u_0$, $\pi = (u_0, u_0, ..., u_0)$ is optimal.

2. For increasing α we begin with a randomization, afterward, the randomization is converted to u_1 only; for monotonically increasing α we begin the randomization in another state.

3. If once for $\alpha = \alpha'$ we have reached $\mu_y(u_1) = 1$ in state y, we always have $\mu_y(u_1) = 1$ in this state for non decreasing α.

4. For $\alpha = u_1$, $\pi = (u_1, u_1, ..., u_1)$ is optimal.

In the following figures we summarize these conclusions for $N = 3$ case:

In Figure 4 (a) we can see the situation for $\alpha = u_0$.

In Figure 4 (b) is represented the situation for $\alpha > u_0$, but less than needed to use u_1 in one of the states.

In Figure 4 (c) we can see the situation for α that enables for us using of u_1 in one of the states.

In Figure 4 (d) is represented the situation for α that enables for us using of u_1 in one of the states and a randomization in another one.

In Figure 4 (e) we can see the situation for α that enables for us using of u_1 in two states.

In Figure 4 (f) is represented the situation for α that enables for us using of u_1 in two states and randomization in another one.

In Figure 4 (g) we can see the situation for $\alpha = u_1$.

For $\alpha > u_1$ is no feasible solution because even if we always use at each state a power that equal to u_1 we get that the maximal average power is

$$\sum_i \rho(x_i, u_1)u_1 = u_1 \sum_i \rho(x_i, u_1) = u_1$$

6.1 CMDPS Algorithm for Two Control Levels

In this section we consider an algorithm that will help us to derive information about the optimal policy structure. Consider CMDP where Assumptions 1 and 2 are satisfied. From the previous section we know that there exist exactly $N + 1$ different optimal solutions without randomization for $N + 1$ different values of α in the range $u_0 \le \alpha \le u_1$. Two of them are trivial for $\alpha = u_0$ and $\alpha = u_1$, and $N - 1$ solutions are the optimal threshold solutions. Let's define what is an optimal threshold solution.

Definition: The optimal solution for $\alpha^k = D(\sigma, \pi_k^{q_1=1})$ is called the optimal threshold solution.

Now we are going to derive algorithm that will help us to obtain the following information:

- What are the values of α^k.
- What is the maximal variability in values of α that can be allowed, such that $\pi_k^{q_1}$ still an optimal solution for particularly chosen k, when $k \in [0, N]$.

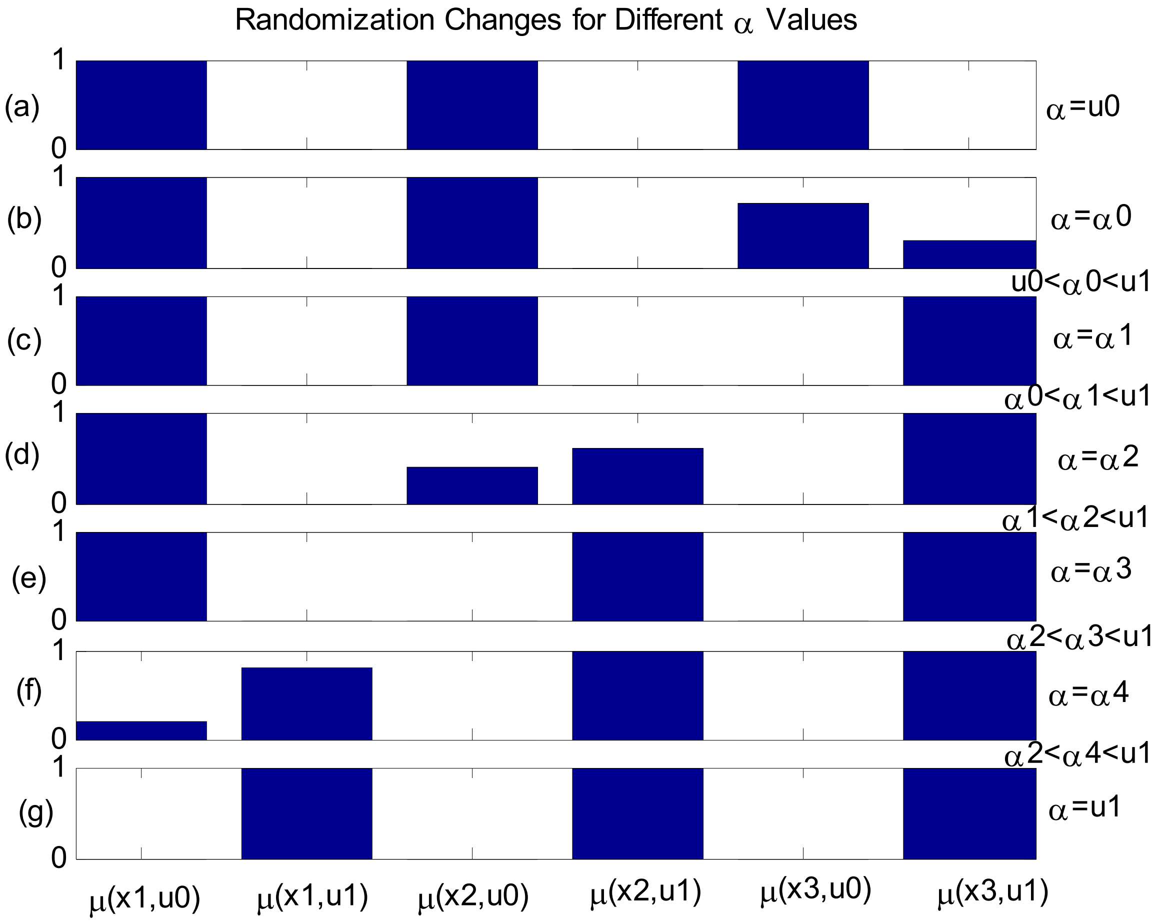


Figure 4: Randomization Changes

- For given α, what is an optimal solution π and what is the value of appropriate ρ, where ρ is a vector of variables.
- What is the minimal cost for given α.

6.1.1 CMDPS Algorithm

In order to derive this algorithm we will use KKT conditions (29-31).

Fix α and let ρ denote the optimizer in the LP, where LP is defined as in $LP_1^\beta(\sigma)$ on page 17.

Consider a finite CMDP so that $\forall x \in X$, $\sum_{u \in U} \rho(x,u) > 0$. Note, that the last assumption can be done without loss of generality because in case where $\exists x_l$ so that $\forall u$ $\rho(x_l, u) = 0$ we can throw away state x_l from the chain without any impact to the optimal policy. Thus, for each $\pi_k^{q_1}$ there exist exactly $N + 1$ entries in the vector ρ that are not equal to zero (N entries correspond to the N states in the chain and one more to the constraint) and $2N - N - 1 = N - 1$ entries that equal zero, so from (31) in vector $\overline{v}$ at least $N + 1$ entries equal to zero, in the same places where vector ρ has non-zero entries. Denote the set of these $N + 1$ entries by I_k, where N from these $N + 1$ entries correspond to N different states in the Markov chain and the last one to the state with randomization.

Define a matrix $A'_k = [a_i]_{i \in I_k}$, where a_i is a column i of matrix A defined in (25). So we get that A'_k is a matrix that consists of columns of A with indexes in I_k.

Define a matrix $A^c_k = [a_i]_{i \in I^c_k}$, where a_i is a column i of matrix A defined in (25). So we get that A^c_k is a matrix that consists from columns of A with indexes from I^c_k, where I^c_k is the complementary set to I_k.

From (30) we have that

$$wA + v = z, \; w \; unrestricted, \; v \geq 0 \tag{36}$$

We can do the following rearrangement:

First of all we will write entries of A, v and z from I_k and afterward entries from I_k^c. Clearly that it doesn't impact the equality and values of v and w.

A, rewritten as $[A'_k A_k^c]$

v, rewritten as v', where v' is a v after rearrangement

z, rewritten as z', where z' is a z after rearrangement

The order of w wasn't changed

After rearrangement we can rewrite equation (36) as follows:

$$w[A'_k A_k^c] + v' = z', \; w \; unrestricted, \; v' \geq 0$$

Because the first $N+1$ coordinates of v' are zero we can replace this equation by two equations

$$wA'_k = z'_{1....(N+1)}, \text{ and } wA_k^c + v'_{(N+2)...2N} = z'_{(N+2)...2N}, w \; unrestricted, \; v'_{(N+2)...2N} \geq 0$$

Now we have two possibilities for matrix A'_k

- The rows are independent, so $rank(A'_k) = N+1$.
- At least one of the rows can be represented as a linear combination of the others, so $rank(A'_k) \neq N+1$.

Conjecture 1: Consider a finite CMDP. If there exists a randomization in one of the states, $0 < q < 1$, then the rank of the matrix A'_k is full, $rank(A'_k) = N+1$. Moreover if there no randomization, a deterministic solution ($q = 0$ or $q = 1$), then $rank(A'_k) = N$ and the row that corresponds to the constraint can be removed without any impact to the solution. (the correctness of this conjecture will be shown numerically).

By some abuse of notation we will denote by A'_k, a matrix $(N+1 \times N+1)$, in which a last row corresponds to the constraint, for cases where $0 < q < 1$, and a matrix $(N \times N)$, where no rows for constraint exists, for cases where $q = 0$ and $q = 1$. Thus by Conjecture 1, we can say that matrix A'_k is invertible for both cases.

Because A'_k has a full rank, we have a unique solution for w,

$$w = z'_{1....(N+1)} * (A'_k)^{-1} \tag{37}$$

and from here there exists only one possible solution for $v'_{(N+2)...2N}$.

$$\begin{aligned} v'_{N+2...2N} &= z'_{(N+2)...2N} - wA^c_k & (38) \\ &= z'_{(N+2)...2N} - [z'_{1....(N+1)} * (A'_k)^{-1}]A^c_k & (39) \end{aligned}$$

If $v'_{(N+2)...2N} \geq 0$ then this is an optimal solution.

Assume that π is an optimal solution for a certain value of α. Since the space of states and actions is finite, then after a final number of checks of (37-39) the optimal solution will be found.

For example for $k = 0$ we have N places to begin randomization, therefore maximum as N possibilities must be checked. For $k = 1$ we have $N - 1$ places for randomization because one place we already found in the previous step. So it is easy to see that after maximum as $N - k$ checks of (37-39) the optimal solution will be found for each k.

On diagram 5 we can see a block representation of the CMDPS Algorithm.

In the following example the algorithm usage will be shown.

6.1.2 CMDPS Algorithm Usage

For simplicity let's assume that $N = 4$ and x_1 is an initial state and the $LP^{\beta}_1(\sigma)$ that we are required to solve looks as follows

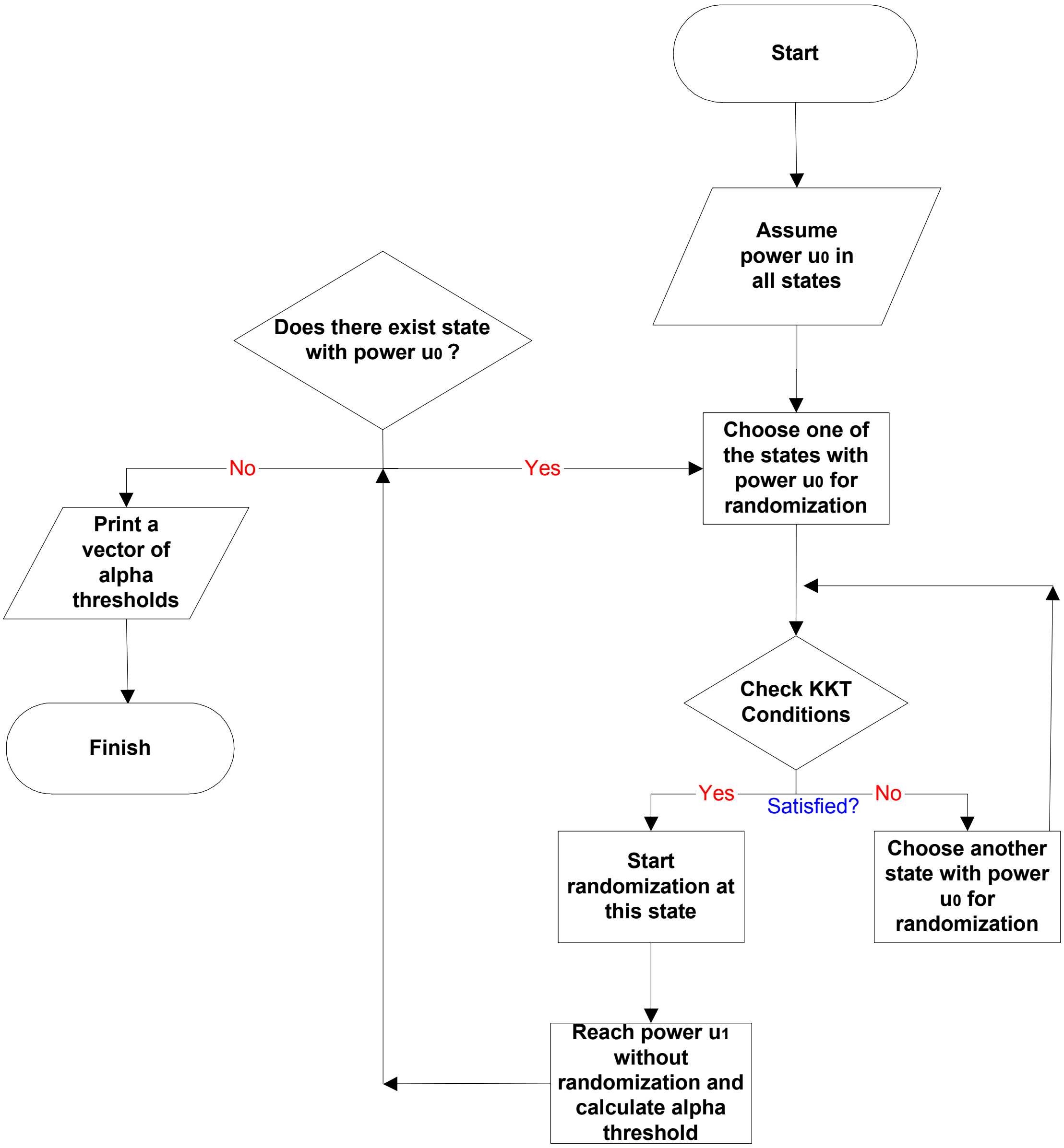

Start
Assume power u0 in all states
Does there exist state with power u0 ?
No
Yes
Choose one of the states with power u0 for randomization
Print a vector of alpha thresholds
Finish
Check KKT Conditions
Yes
Satisfied?
No
Start randomization at this state
Choose another state with power u0 for randomization
Reach power u1 without randomization and calculate alpha threshold

Figure 5: CMDPS Algorith Diagram

$$\min_{\rho}\left\{[z_1\ z_1\ z_2\ z_2\ z_3\ z_3\ z_4\ z_4]\begin{bmatrix}\rho(x_1,u_0)\\ \rho(x_1,u_1)\\ \rho(x_2,u_0)\\ \rho(x_2,u_1)\\ \rho(x_3,u_0)\\ \rho(x_3,u_1)\\ \rho(x_4,u_0)\\ \rho(x_4,u_1)\end{bmatrix}\right\}$$

Subject to

$$[u_0\ u_1\ u_0\ u_1\ u_0\ u_1\ u_0\ u_1]\begin{bmatrix}\rho(x_1,u_0)\\ \rho(x_1,u_1)\\ \rho(x_2,u_0)\\ \rho(x_2,u_1)\\ \rho(x_3,u_0)\\ \rho(x_3,u_1)\\ \rho(x_4,u_0)\\ \rho(x_4,u_1)\end{bmatrix}=\alpha$$

$$\begin{aligned} b &= \begin{bmatrix} 1-\beta \\ 0 \\ 0 \\ 0 \\ \alpha \end{bmatrix} \\ A &= \begin{bmatrix} (1-\beta P_{x_1x_1}(u_0)) & (1-\beta P_{x_1x_1}(u_1)) & -\beta P_{x_2x_1}(u_0) & -\beta P_{x_2x_1}(u_1) \\ -\beta P_{x_1x_2}(u_0) & -\beta P_{x_1x_2}(u_1) & (1-\beta P_{x_2x_2}(u_0)) & (1-\beta P_{x_2x_2}(u_1)) \\ -\beta P_{x_1x_3}(u_0) & -\beta P_{x_1x_3}(u_1) & -\beta P_{x_2x_3}(u_0) & -\beta P_{x_2x_3}(u_1) \\ -\beta P_{x_1x_4}(u_0) & -\beta P_{x_1x_4}(u_1) & -\beta P_{x_2x_4}(u_0) & -\beta P_{x_2x_4}(u_1) \\ u_0 & u_1 & u_0 & u_1 \\ -\beta P_{x_3x_1}(u_0) & -\beta P_{x_3x_1}(u_1) & -\beta P_{x_4x_1}(u_0) & -\beta P_{x_4x_1}(u_1) \\ -\beta P_{x_3x_2}(u_0) & -\beta P_{x_3x_2}(u_1) & -\beta P_{x_4x_2}(u_0) & -\beta P_{x_4x_2}(u_1) \\ (1-\beta P_{x_3x_3}(u_0)) & (1-\beta P_{x_3x_3}(u_1)) & -\beta P_{x_4x3}(u_0) & -\beta P_{x_4x_3}(u_1) \\ -\beta P_{x_3x_4}(u_0) & -\beta P_{x_3x_4}(u_1) & (1-\beta P_{x_4x_4}(u_0)) & (1-\beta P_{x_4x_4}(u_1)) \\ u_0 & u_1 & u_0 & u_1 \end{bmatrix} \end{aligned}$$

Now, by algorithm usage we will answer our questions.

- Let's find α^k. We start with α^0 because it's value is always known

$$\begin{aligned} \alpha^0 &= u_0 \\ \{\mu_{x_1}(u_0) &= 1, \mu_{x_1}(u_1) = 0, \mu_{x_2}(u_0) = 1, \mu_{x_2}(u_1) = 0, \\ \mu_{x_3}(u_0) &= 1, \mu_{x_3}(u_1) = 0, \mu_{x_4}(u_0) = 1, \mu_{x_4}(u_1) = 0\} \\ &\Longrightarrow \pi^1_{k=-1} = (1,0,1,0,1,0,1,0) \end{aligned}$$

where $\pi^1_{k=-1}(x) = \delta_{u_0}(x)$, $\forall x$.

Because $N = 4$ for $\pi^{q_1}_{k=0}$ there exist exactly $N = 4$ possibilities. Assume that $\pi^{q_1}_{k=0} =$

$(1,1,1,0,1,0,1,0)$ so at first step we suppose to do the following

$$I_0 = \{1,2,3,5,7\}, I_0^c = \{4,6,8\} \tag{40}$$

$$A_0' = \tag{41}$$

$$\begin{bmatrix} (1-\beta P_{x_1x_1}(u_0)) & (1-\beta P_{x_1x_1}(u_1)) & -\beta P_{x_2x_1}(u_0) & -\beta P_{x_3x_1}(u_0) & -\beta P_{x_4x_1}(u_0) \\ -\beta P_{x_1x_2}(u_0) & -\beta P_{x_1x_2}(u_1) & (1-\beta P_{x_2x_2}(u_0)) & -\beta P_{x_3x_2}(u_0) & -\beta P_{x_4x_2}(u_0) \\ -\beta P_{x_1x_3}(u_0) & -\beta P_{x_1x_3}(u_1) & -\beta P_{x_2x_3}(u_0) & (1-\beta P_{x_3x_3}(u_0)) & -\beta P_{x_4x3}(u_0) \\ -\beta P_{x_1x_4}(u_0) & -\beta P_{x_1x_4}(u_1) & -\beta P_{x_2x_4}(u_0) & -\beta P_{x_3x_4}(u_0) & (1-\beta P_{x_4x_4}(u_0)) \\ u_0 & u_1 & u_0 & u_0 & u_0 \end{bmatrix} \tag{42}$$

$$A_0^c = \begin{bmatrix} -\beta P_{x_2x_1}(u_1) & -\beta P_{x_3x_1}(u_1) & -\beta P_{x_4x_1}(u_1) \\ (1-\beta P_{x_2x_2}(u_1)) & -\beta P_{x_3x_2}(u_1) & -\beta P_{x_4x_2}(u_1) \\ -\beta P_{x_2x_3}(u_1) & (1-\beta P_{x_3x_3}(u_1)) & -\beta P_{x_4x_3}(u_1) \\ -\beta P_{x_2x_4}(u_1) & -\beta P_{x_3x_4}(u_1) & (1-\beta P_{x_4x_4}(u_1)) \\ u_1 & u_1 & u_1 \end{bmatrix}$$

$$v = (0,0,0,v_4,0,v_6,0,v_8)$$

$$z = [z_1\ z_1\ z_2\ z_2\ z_3\ z_3\ z_4\ z_4]$$

$$v' = (0,0,0,0,0,v_4,v_6,v_8)$$

$$z' = [z_1\ z_1\ z_2\ z_3\ z_4\ z_2\ z_3\ z_4]$$

From (37-39)

$$w = [z_1\ z_1\ z_2\ z_3\ z_4] * (A_0')^{-1}$$

$$v'_{6,7,8} = [z_2\ z_3\ z_4] - wA_0^c$$

If $v'_{6,7,8} \geq 0$ then this is an optimal solution, otherwise we continue to the next possible feasible solution of $\pi_{k=0}^{q_1}$ form.

After $\pi_{k=0}^{q_1}$ is found we can find the ρ and α^1. Assume that $\pi_{k=0}^{q_1} = (1,1,1,0,1,0,1,0)$ is an optimal solution, then from Theorem 1 for α^1

$$\begin{aligned}\pi_{k=0}^{q_1=1} &= (0,1,1,0,1,0,1,0) \\ &\implies \rho(x_1,u_0)=0, \rho(x_2,u_1)=0, \rho(x_3,u_1)=0, \rho(x_4,u_1)=0\end{aligned}$$

From (27)

$$\begin{aligned} A\rho &= b \\ &\implies \alpha^1 = [u_1\ u_0\ u_0\ u_0] \begin{bmatrix} \rho(x_1,u_1) \\ \rho(x_2,u_0) \\ \rho(x_3,u_0) \\ \rho(x_4,u_0) \end{bmatrix} \end{aligned} \tag{43}$$

Because $N = 4$ then for $\pi_{k=1}^{q_1}$ exist only 3 possibilities. Assume that $\pi_{k=1}^{q_1} = (0,1,1,1,1,0,1,0)$. Similarly to α^1 we can find α^2 and α^3 as well. It is clear that $\alpha^4 = u_1$. So the all α^k is found for $k \in [0,1,2,3,4]$.

- Let's find what is the maximal variability in the values of α that can be allowed such that $\pi_k^{q_1}$ is still an optimal policy for particularly chosen k, when $k \in [0,N]$.

Define d^k as

$$\begin{aligned} d^0 &= \alpha^1 - \alpha^0, d^1 = \alpha^2 - \alpha^1 \\ d^2 &= \alpha^3 - \alpha^2, d^3 = \alpha^4 - \alpha^3 \end{aligned}$$

From Theorem 1, for each d^k, $\pi_k^{q_1}$ is still an optimal solution. So d^k is the maximal variability in the values of α that can be allowed such that $\pi_k^{q_1}$ is still an optimal solution for particularly chosen k, when $k \in [0,N]$.

- For given α we can find an appropriate α^{k+1} and α^k, where $\alpha^{k+1} > \alpha > \alpha^k$. After α^{k+1} and α^k are known, we can find $\pi_k^{q_1}$. After $\pi_k^{q_1}$ is known the ρ will be found easily from (27).

- After ρ was found a cost can be calculated from the following formula

$$\text{cost} = [z_1 \; z_1 \; z_2 \; z_2 \; z_3 \; z_3 \; z_4 \; z_4] \begin{bmatrix} \rho(x_1, u_0) \\ \rho(x_1, u_1) \\ \rho(x_2, u_0) \\ \rho(x_2, u_1) \\ \rho(x_3, u_0) \\ \rho(x_3, u_1) \\ \rho(x_4, u_0) \\ \rho(x_4, u_1) \end{bmatrix} \tag{44}$$

6.1.3 Example of CMDPS Algorithm Application to CMDP Problems

For better understanding of the algorithm let's consider the following example.

We have four states, where for each state, its cost is equal to the number that is written in the circles on Figure 6. The transitions probabilities as in Figure 6 with values taken from (45). The values of control levels and β we can see at (46).

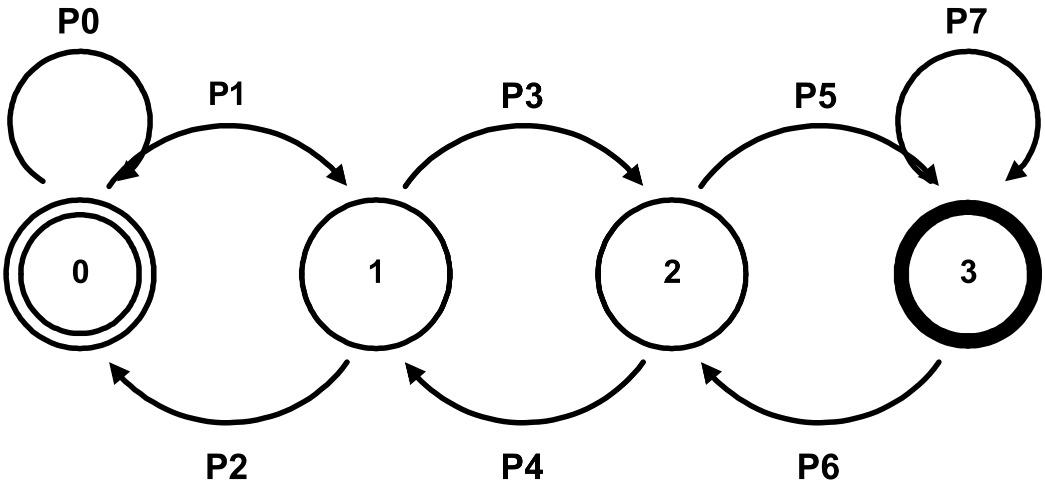


Figure 6: Markov Chain for Example 1

$$
\begin{aligned}
P0(u_0) &= 0.8, P0(u_1) = 0.9 \\
P1(u_0) &= 0.2, P1(u_1) = 0.1 \\
P2(u_0) &= 0.6, P2(u_1) = 0.7 \\
P3(u_0) &= 0.4, P3(u_1) = 0.3 \\
P4(u_0) &= 0.6, P4(u_1) = 0.7 \\
P5(u_0) &= 0.4, P5(u_1) = 0.3 \\
P6(u_0) &= 0.65, P6(u_1) = 0.75 \\
P7(u_0) &= 0.35, P7(u_1) = 0.25
\end{aligned} \tag{45}
$$

$$u_0 = 0.1,\ u_1 = 1,\ \beta = 0.99 \tag{46}$$

From (45) A is easily calculated.

Now let's answer our questions.

- Let's find α^k. We start with α^0 and continue according to the algorithm explained above

$$
\begin{aligned}
& \alpha^0 = 0.1 \\
& \pi^1_{k=-1} = (1,0,1,0,1,0,1,0) \\
& I_0 = \{1,2,3,5,7\}, I_0^c = \{4,6,8\} \\
& v = (0,0,0,v_4,0,v_6,0,v_8) \\
& z = [0\ 0\ 1\ 1\ 2\ 2\ 3\ 3] \\
& v' = (0,0,0,0,0,v_4,v_6,v_8) \\
& z' = [0\ 0\ 1\ 2\ 3\ 1\ 2\ 3] \\
& w = [0\ 0\ 1\ 2\ 3] * (A_0')^{-1} \\
v'_{6,7,8} &= [1\ 2\ 3] - wA_0^c \\
&= [-0.4345 \quad -0.4368 \quad -0.0023] \ngeq 0 \\
& I_0 = \{1,3,4,5,7\}, I_0^c = \{2,6,8\} \\
v'_{6,7,8} &= [0\ 2\ 3] - wA_0^c \\
&= [0.4345 \quad -0.0023 \ \ 0.4322] \ngeq 0 \\
& I_0 = \{1,3,5,6,7\}, I_0^c = \{2,4,8\} \\
v'_{6,7,8} &= [0\ 1\ 3] - wA_0^c \\
&= [0.4368 \ \ 0.0023 \ \ 0.4345] \geq 0 \\
&= > \pi^1_{k=0} = (1,0,1,0,0,1,1,0) \\
& \rho = [0.6302\ 0\ 0.2038\ 0\ 0\ 0.1141\ 0.0519\ 0] \\
& \alpha^1 = 0.2027
\end{aligned}
$$

In the same way we can find that

$$\begin{aligned}
\pi^1_{k=1} &= (1,0,0,1,0,1,1,0) \\
\alpha^2 &= 0.3461 \\
\pi^1_{k=2} &= (1,0,0,1,0,1,0,1) \\
\alpha^3 &= 0.3763 \\
\pi^1_{k=3} &= (0,1,0,1,0,1,0,1) \\
\alpha^4 &= 1
\end{aligned}$$

- Let's find what is the maximal variability in the values of α that can be allowed such that $\pi^{q_1}_k$ still an optimal solution for certainly k, when $k \in [0, N]$.

$$\begin{aligned}
d^0 &= \alpha^1 - \alpha^0 = 0.1027 \\
d^1 &= \alpha^2 - \alpha^1 = 0.1434 \\
d^2 &= \alpha^3 - \alpha^2 = 0.0302 \\
d^3 &= \alpha^4 - \alpha^3 = 0.6237
\end{aligned}$$

We can see that d^3 domain is a most stable domain.

- Assume that $\alpha = 0.5$, let's find what is an appropriated π and ρ.

Since $\alpha^4 > (\alpha = 0.5) > \alpha^3$, then

$$\pi = (1,1,0,1,0,1,0,1)$$

and from (27)

$$\rho = (0.5556 \quad 0.1624 \quad 0 \quad 0.1779 \quad 0 \; 0.0747 \quad 0 \; 0.0295) \tag{47}$$

- The cost can be calculated from (47) and (44).

$$\text{cost} = 0.4158 \tag{48}$$

Now we will compare the results which were gotten by CMDPS algorithm to one of the well known algorithms. The one of the most popular algorithms for LP solving is the Simplex algorithm, so the comparison will be done to this one.

First of all we are interested to compare the ρ and the cost for $\alpha = 0.5$. The results that were gotten by the Simplex algorithm coincide with (47) and (48) as expected.

Now let's compare the behavior of these two algorithms for the different values of α, where

$$\alpha \in (u_0, u_1) = (0.1, 1)$$

The results of CMDPS algorithm can be written in the following form:

For $\alpha = \alpha^0$, we use in all four states control level u_0.

For $(\alpha^0 = 0.1) < \alpha < (\alpha^1 = 0.2027)$, we have randomization in state 2 and in the rest states we use control level u_0.

For $\alpha = \alpha^1$, in the states 0, 1 and 3 we use control level u_0, and in the state 2 we use control level u_1.

For $(\alpha^1 = 0.2027) < \alpha < (\alpha^2 = 0.3461)$, we have randomization in state 1, in the states 0 and 3 we use control level u_0, and in the state 2 we use control level u_1.

For $\alpha = \alpha^2$, in the states 0 and 3 we use control level u_0, and in the states 1 and 2 we use control level u_1.

For $(\alpha^2 = 0.3461) < \alpha < (\alpha^3 = 0.3763)$, we have randomization in state 3, in the state 0 we use control level u_0, and in the states 1 and 2 we use control level u_1.

For $\alpha = \alpha^3$, in the state 0 we use control level u_0, and in the states 1, 2 and 3 we use control level u_1.

For $(\alpha^3 = 0.3763) < \alpha < (\alpha^4 = 1)$, we have randomization in state 0 and in the rest states we use control level u_1.

For $\alpha = \alpha^4$, we use in all four states control level u_1.

Actually from computational point of view, in order to get these results, we need to find only 3 values of $\alpha : \alpha^1, \alpha^2, \alpha^3$*. The values of* α^0 *and* α^4 *are known from the beginning because these values are equal to the minimal and maximal value of control level and so calculated in the trivial way.*

In order to derive the same solution properties by Simplex algorithm (Appendix A) we need to run a simulation with small enough step for α*, so in order to get a precise results we run the algorithm thousands times. Moreover even after this huge number of runs we don't know exactly what happens in the each possible value of* α *because the number of* α*'s is infinite and the number of runs is finite even if it is very big.*

On Figure 7 we can see the results that were derived by Simplex method for example 1.

On Figure 8 we can see the randomization behavior for example 1 that were derived by CMDPS algorithm. The results are the same.

By using a CMDPS algorithm we need to find only 3 values of α.

As was stated in this section, it is enough to find N-1 different points of α, in order to get the full solution structure.

6.2 CMDPS Algorithm for Finite Number of Control Levels

Now we will extend the analysis from two control levels for the arbitrary, but finite, number of control levels.

Here we should add one small correction for assumption 2 such that it will suitable for case with arbitrary number of control levels.

Assumption 2 for arbitrary number of control levels (monotonicity assumption): When the using of higher control levels in state x_i is increasing, the value of the

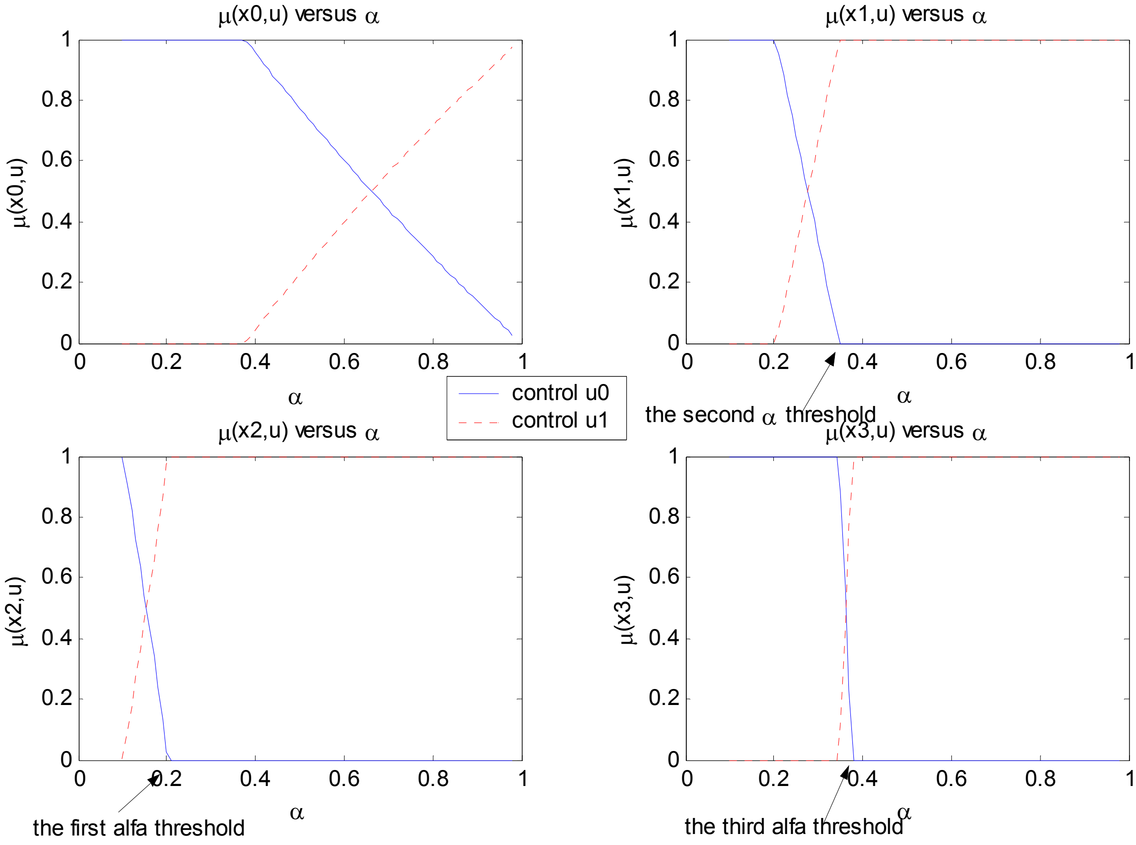


Figure 7: The Simplex Method Results

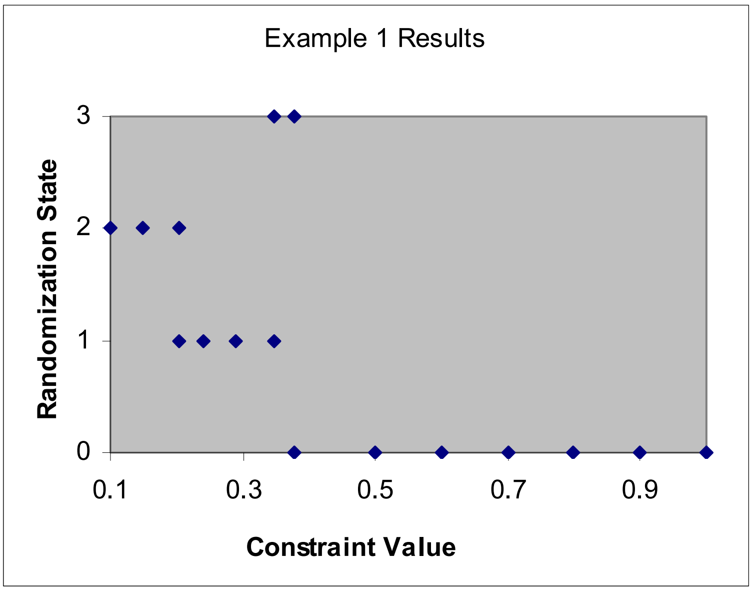


Figure 8: Randomization Behavior in Example 1

constraint α is increasing, where x_i is the state with randomization. We will say that effective power control level in state with randomization is higher.

By effective power control level of the state we mean the following:

$$u_{effective}(x_i) = \rho(x_i, u_0)u_0 + \rho(x_i, u_1)u_1 + ... + \rho(x_i, u_s)u_s \tag{49}$$

where x_i is the state and $u_0, u_1.., u_s$ are the power control levels.

With a some abuse of notations we will use Theorem 3 and Theorem 4 for the case with arbitrary, but finite, number of control levels as well. We can do this because from ([1], Theorem 3.8) we know that for a certain value of α there exists an optimal policy which has randomization only in one state and this randomization is only between two control levels.

Conclusions: From Theorem 2,3 and 4 we can conclude that if for each α there exists a unique optimal solution then it takes the following form:

1. For $\alpha = u_0$, $\pi = (u_0, u_0, ..., u_0)$ is optimal.

2. For increasing α we begin with a randomization, afterward, the randomization is converted to a certain deterministic control level. For monotonically increasing α we begin the randomization in another state or in a case that a previous deterministic control level was not a maximal one, it is possible that we will continue with the randomization in the same state but with two other control levels such that the effective control level will be higher.

3. If once for a certain value of α we have reached $\mu_y(u_s) = 1$ in state y, then we always have $\mu_y(u_s) = 1$ in this state for non decreasing values of α.

4. For $\alpha = u_s$, $\pi = (u_s, u_s, ..., u_s)$ is optimal.

Because CMDPS algorithm development is not constrained to the number of control levels the CMDPS algorithm is suitable for the case with arbitrary number of control levels without any changes.

Now we will consider the example from the previous subsection but only with $S = 3$ control levels.

$$\begin{aligned} P0(u_0) &= 0.8, P0(u_1) = 0.9, P0(u_2) = 0.95 \\ P1(u_0) &= 0.2, P1(u_1) = 0.1, P1(u_2) = 0.05 \\ P2(u_0) &= 0.6, P2(u_1) = 0.7, P2(u_2) = 0.75 \\ P3(u_0) &= 0.4, P3(u_1) = 0.3, P3(u_2) = 0.25 \\ P4(u_0) &= 0.6, P4(u_1) = 0.7, P4(u_2) = 0.75 \\ P5(u_0) &= 0.4, P5(u_1) = 0.3, P5(u_2) = 0.25 \\ P6(u_0) &= 0.65, P6(u_1) = 0.75, P6(u_2) = 0.8 \\ P7(u_0) &= 0.35, P7(u_1) = 0.25, P7(u_2) = 0.2 \end{aligned} \tag{50}$$

$$u_0 = 0.1,\ u_1 = 1,\ u_2 = 1.3,\ \beta = 0.99 \tag{51}$$

From (50) A is easily calculated.

Now let's answer our questions.

- Let's find α^k. We start with α^0 and continue according to the CMDPS algorithm

$$
\begin{aligned}
\alpha^0 &= 0.1 \\
\pi^1_{k=-1} &= (1,0,0,1,0,0,1,0,0,1,0,0) \\
\pi^1_{k=0} &= (1,0,0,1,0,0,0,0,1,1,0,0) \\
\alpha^1 &= 0.2307 \\
\pi^1_{k=1} &= (1,0,0,0,0,1,0,0,1,1,0,0) \\
\alpha^2 &= 0.4020 \\
\pi^1_{k=2} &= (1,0,0,0,0,1,0,0,1,0,0,1) \\
\alpha^3 &= 0.4265 \\
\pi^1_{k=3} &= (0,0,1,0,0,1,0,0,1,0,0,1) \\
\alpha^4 &= 1.3
\end{aligned}
$$

- Let's find what is the maximal variability in the values of α that can be allowed such that $\pi^{q_1}_k$ still an optimal solution for certainly k, when $k \in [0, N]$.

$$
\begin{aligned}
d^0 &= \alpha^1 - \alpha^0 = 0.2307 \\
d^1 &= \alpha^2 - \alpha^1 = 0.1713 \\
d^2 &= \alpha^3 - \alpha^2 = 0.0245 \\
d^3 &= \alpha^4 - \alpha^3 = 0.8735
\end{aligned}
$$

We can see that d^3 domain is a most stable domain.

- Assume that $\alpha = 0.5$, let's find what is an appropriated π and ρ.

Since $\alpha^4 > (\alpha = 0.5) > \alpha^3$, then

$$\pi = (1, 0, 1, 0, 0, 1, 0, 0, 1, 0, 0, 1)$$

and from (27)

$$\rho = (0.6667\ 0\ 0.0770\ 0\ 0\ 0.1795\ 0\ 0\ 0.0588\ 0\ 0\ 0.0181) \tag{52}$$

- The cost can be calculated from (52) and (44).

$$\text{cost} = 0.3514$$

7 A New Approach to Optimization of Wireless Communications

In this section we will discuss the application of CMDPS algorithm to solving of wireless communications problems.

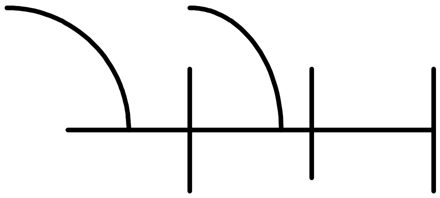
Let's consider a diagram shown in Figure 9 that describes our communications problem.

In this diagram we can see the representation of the following problem:

λ=2[codewords/time duration], R=3[codewords/time duration],

$T_{slot} = \frac{1}{R} = \frac{1}{3}$[time duration] and $N = 3$ (the buffer size is 2).

So in each slot a maximally one codeword can be transmitted. In the left upper corner of the diagram arcs denote arrivals and straight lines denote transmissions. As we can see in the first two slots after each arrival we have one transmission and in a third slot only transmission is presented. This structure can be represented by Markov chain. In this diagram (Figure 9), represented only Markov chain for $N = 3$, but it is clear that for general N it would be



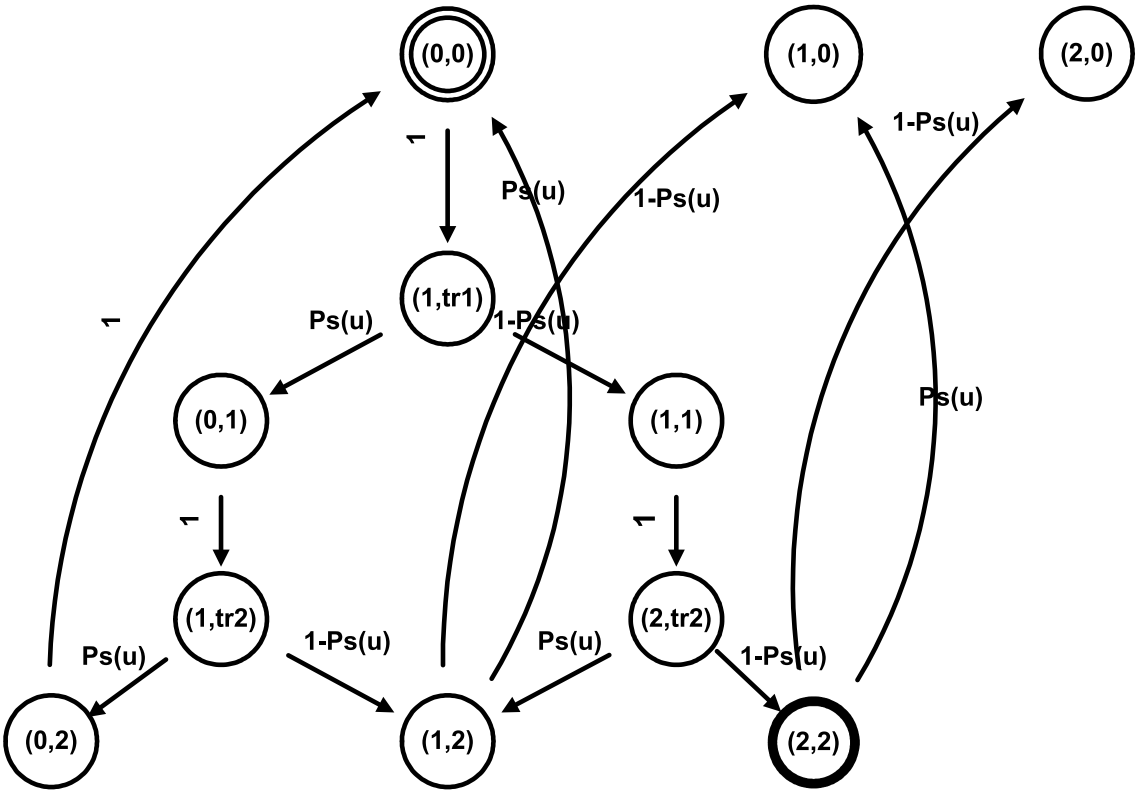


Figure 9: Markov Chain of the Wireless Communications Problem

the only duplicating of this replica. In each state we have two numbers (a, b), where a this is the number of messages in the buffer and b this is the number of the transmissions that were completed at the time duration.

Because our control problem is in which power level to use in each transmission, we can simplify the above diagram by removing states with arrivals only, without transmissions. There are $(1, tr1)$, $(2, tr1)$, $(1, tr2)$ states. The diagram in Figure 10 represents exactly the same problem but in a simpler way. We can see that for each main state, main states

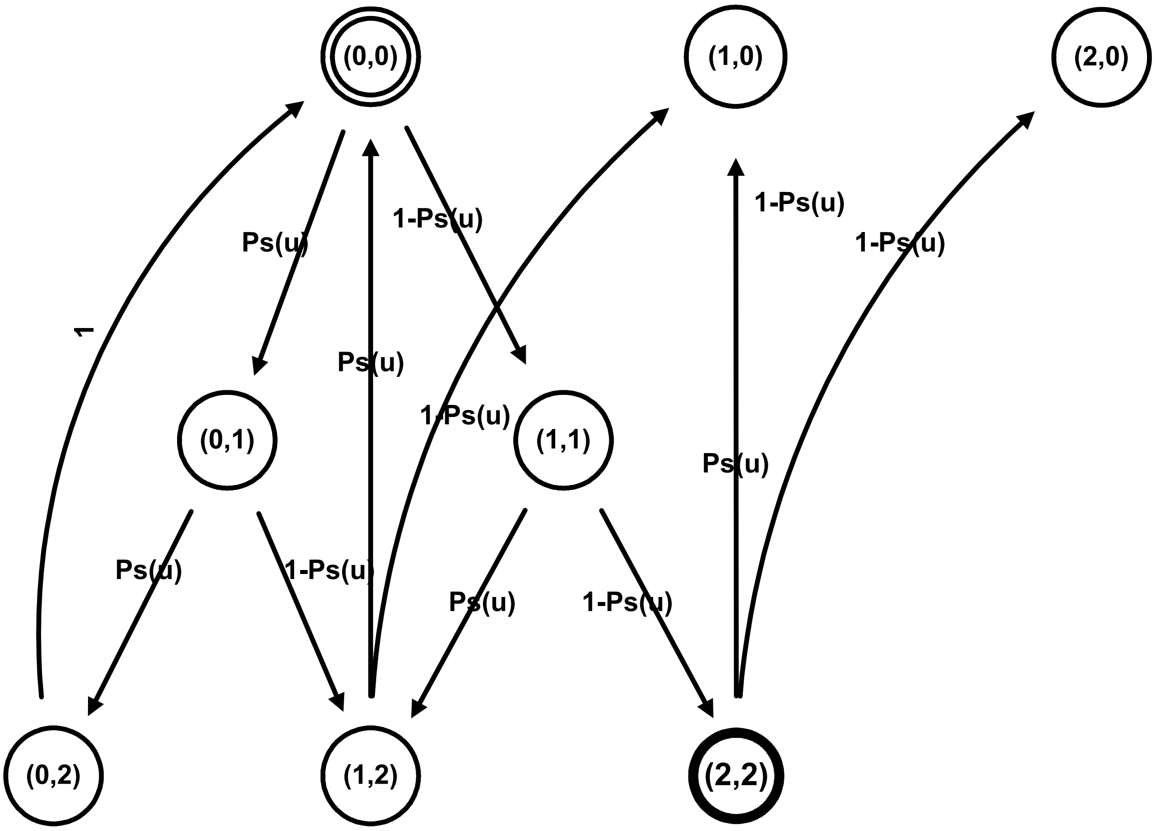


Figure 10: Simplified Markov Chain of the Wireless Communications Problem

are the states that represent different number of codewords in the buffer, we have $R-1$ additional states for each transmission during one time duration. So the maximal number of necessarily states is RN. The all states will be numerated in the following order: $(0,0)$ is a state number 1, $(0,1)$ is a state number 2, $(0,2)$ is a state number 3, ..., $(1,0)$ is a state number 4, ... $(N-1,0)$ is a state number $RN-R+1$ and finally $(N-1,R-1)$ is a state number RN.

7.1 Communications Problem Optimization Using CMDP Approach

We are interested to minimize the average delay of the transmissions while the average power remains below a given level α.

We assume a block flat fading Gaussian channel in which the fading of each slot is *iid* according to some distribution (such as Rayleigh, Rice etc...). Denote the fading disturbances at slot k by w_k.

Although it is more common to use average cost we are using the discounted cost. It simplifies the calculation and when the discount factor $\beta \to 1$ by Tauberian Theorem [10] we are getting the average cost. Consider a finite CMDP with arbitrary, but finite, number of control levels and one constraint.

Definitions:

- W is the set of all possible fading disturbances.
- $f_\beta(\sigma, \pi; x, u, w) = (1 - \beta) \sum_{t=1}^{\infty} \beta^{t-1} P_\sigma^\pi(X_t = x, U_t = u, W_t = w)$, where $w \in W$ and $0 < \beta < 1$.
- X_k is the buffer size at slot k.
- $u(x_k) \in U$ is a transmitted power at slot k where in the buffer we have x_k codewords waiting for transmission.
- Immediate cost at time k is the number of codewords waiting in the buffer after k steps when power u is used in slot k, therefore $c = c(x_k, u) = x_k$, where x_k is the number of codewords at slot k.

- The discounted cost can be defined using [1] and [4]

$$
\begin{aligned}
C &= C_{\beta}(\sigma,\pi) \\
&= (1-\beta)E_{\sigma}^{\pi}\{\sum_{k=1}^{\infty}\beta^{k-1}c(x_k,u_k)\} \\
&= (1-\beta)\sum_{k=1}^{\infty}\beta^{k-1}E_{\sigma}^{\pi}c(x_k,u_k) \\
&= (1-\beta)\sum_{k=1}^{\infty}\sum_{x,u,w}\beta^{k-1}P_{\sigma}^{\pi}(x_k=x,w_k=w,u_k=u)c(x,u) \\
&= \sum_{x,u}(1-\beta)\sum_{k=1}^{\infty}\sum_{w}\beta^{k-1}P_{\sigma}^{\pi}(x_k=x,w_k=w,u_k=u)c(x,u) \\
&= \sum_{x,u}(1-\beta)\sum_{k=1}^{\infty}\beta^{k-1}P_{\sigma}^{\pi}(x_k=x,u_k=u)c(x,u) \qquad (53)\\
&= \sum_{x,u}f_{\beta}(\sigma,\pi;x,u)c(x,u) \\
&= \sum_{x,u}f_{\beta}(\sigma,\pi;x,u)x
\end{aligned}
$$

- Immediate cost at time k, related to the constraint, is the number of codewords waiting in the buffer after k steps where power u used at slot k, therefore $d = d(x_k,u) = u$, where x_k is the number of codewords at slot k.

- The value of the constraint criterion when starting at σ and using policy π can be defined as follows

$$
\begin{aligned}
D &= D_{\beta}(\sigma;\pi) \\
&= \sum_{x\in X}\sum_{u\in U}f_{\beta}(\sigma,\pi;x,u)d(x,u) \\
&= \sum_{x\in X}\sum_{u\in U}f_{\beta}(\sigma,\pi;x,u)u
\end{aligned}
$$

In (53) we used property that $c(x,u)$ doesn't depend on w, therefore problem can be solved for average value of disturbances. (the same derivation for $d(x,u)$).

In the next equations we summarize our optimization problem. The optimization problem is:

$$\min_{\pi} C_{\beta}(\sigma,\pi) = \min_{\pi} \sum_{x,u} f_{\beta}(\sigma,\pi;x,u)x$$

Subject to

$$D_{\beta}(\sigma,\pi) = \sum_{x,u} f_{\beta}(\sigma,\pi;x,u)u \leqslant \alpha$$

We can do one more simplification. The inequality constraint can be replaced by equality because in practise we are interested to use as much power as we have in order to improve the delay (higher power will decrease the error probability of the transmission), therefore

$$D_{\beta}(\sigma,\pi) = \sum_{x,u} f_{\beta}(\sigma,\pi;x,u)u = \alpha$$

Now by using $LP_1^{\beta}(\sigma)$ we can rewrite our problem in a simpler form.

$$\min_{\rho} \sum_{x,u} \rho(x,u)x \tag{54}$$

Subject to

$$\sum_{x,u} \rho(x,u)u = \alpha \tag{55}$$

$$\left\{ \begin{array}{c} \sum_{y\in X}\sum_{u\in U(y)} \rho(y,u)(\delta_x(y) - \beta P_{yx}(u)) = (1-\beta)\sigma(x), \forall x \in X \\ \rho(y,u) \geqslant 0, \forall y,u \end{array} \right\}$$

Define $\rho_x = \rho_x(u) = \sum_x \rho(x,u)$, and $\rho_u = \rho_u(x) = \sum_u \rho(x,u)$, so we can rewrite (54) and (55) as following

$$\begin{aligned} & \min_{\rho} \sum_{x,u} \rho(x,u)x \\ = & \min_{\rho} \sum_{x} x \sum_{u} \rho(x,u) \\ = & \min_{\rho} \sum_{x} \rho_u(x)x \end{aligned}$$

$$\begin{aligned} & \sum_{x,u} \rho(x,u)u \\ = & \sum_{u} u \sum_{x} \rho(x,u) \\ = & \sum_{u} u\rho_x(u) = \alpha \end{aligned}$$

$\Rightarrow$

$$\min_{\rho} \langle \overline{x}, \rho_u(x) \rangle \tag{56}$$

Subject to

$$\langle \overline{u}, \rho_x(u) \rangle = \alpha \tag{57}$$

$$\left\{ \begin{array}{c} \sum_{y\in X} \sum_{u\in U(y)} \rho(y,u)(\delta_x(y) - \beta P_{yx}(u)) = (1-\beta)\sigma(x), \forall x \in X \\ \rho(y,u) \geqslant 0, \forall y, u \end{array} \right\} \tag{58}$$

Define an average delay (ad) in the mean of the number of codewords in the buffer. Denote the number of codewords in the buffer by "nc", a probability for k codewords in the buffer by $P(nc = k)$ and express an average delay as follows

$$ad = \sum_{k=0}^{\infty} P(nc = k)k \tag{59}$$

In our design assignment we are interested to answer the following questions:

1. What are the values of α^k.

2. What is the maximal variability in average power α that can be allowed such that $\pi_k^{q_1}$ still an optimal solution for certain value of k, when $k \in [0, N]$.

3. For given α what is an optimal solution π and what is the value of appropriate ρ.

4. What is an average delay.

The CMDPS algorithm will help us to answer these questions.

7.2 CMDPS Algorithm Application in Wireless Communications

In our model we assume a BPSK modulation without any coding and our Markov diagram as in Figure 11. We can see that the transitions at final states a little bit differ from the other because when the buffer is full and arriving is happening we lost an arrival word immediately. The transition probabilities are calculated according to (65) in Appendix C.

Simulation Conditions:

$$\lambda = 2\ [words/\sec], R = 3\ [words/\sec]$$
$$T_{slot} = \frac{1}{R} = \frac{1}{3}\ [sec]$$
$$\beta = 0.99,\ N = 3$$
$$Eb_0 = 0.1\ [Joules],\ Eb_1 = 1\ [Joules] \quad \text{(Appendix C)}$$
$$x_1 = 0\ [words],\ x_2 = 1\ [words],\ x_3 = 2\ [words],\ x_4 = 3\ [words]$$
$$\frac{N_0}{2} = 1,\ \zeta^2 = 1 \quad \text{(Appendix C)}$$
$$Tb = 0.1\ [sec] \quad \text{(Appendix C)}$$
$$u_0 = \frac{Eb_0}{Tb} = 1\ [W],\ u_1 = \frac{Eb_1}{Tb} = 10\ [W]$$

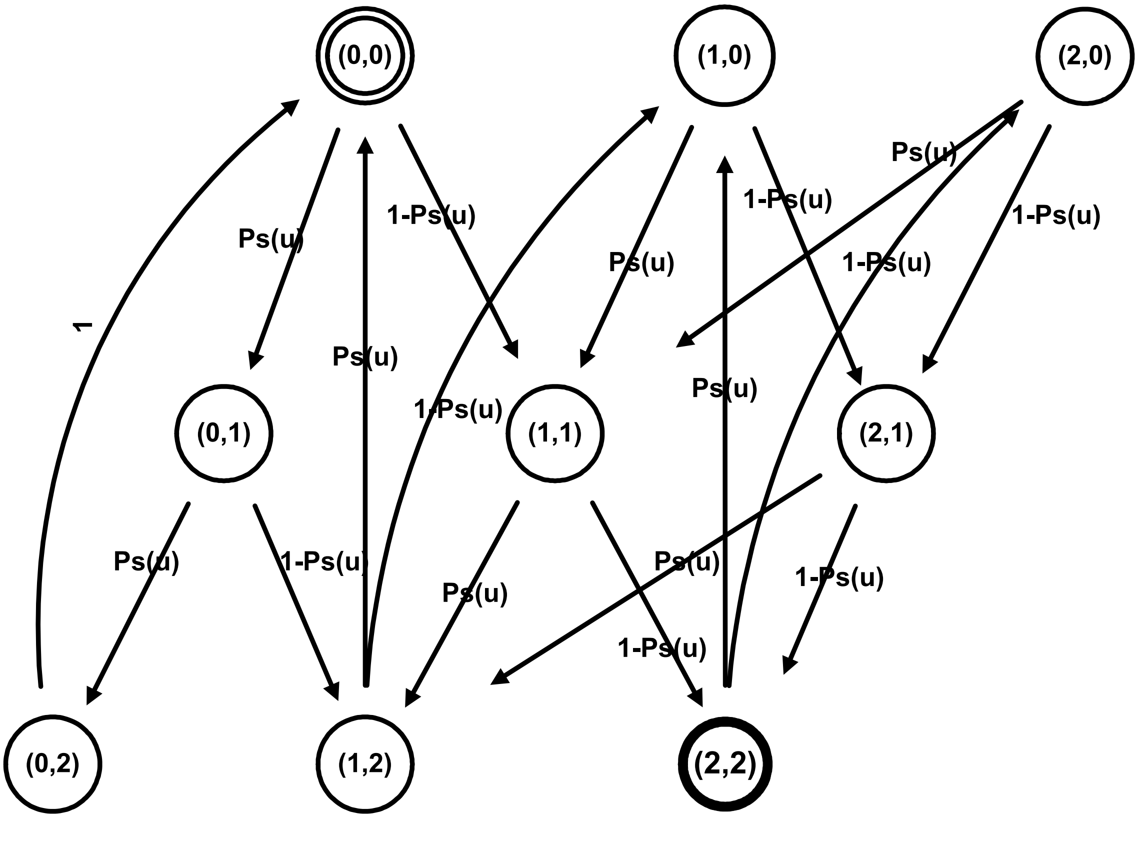


Figure 11: Markov Chain for Communications Problem Example

For simplicity we will run simulation for two control levels but as we saw in section 6.2 it could be done for the arbitrary, but finite, number of control levels.

state number	label on the diagram	randomization order
1	(0,0)	1
2	(0,1)	5
3	(0,2)	9
4	(1,0)	7
5	(1,1)	4
6	(1,2)	2
7	(2,0)	6
8	(2,1)	3
9	(2,2)	8

Questions 1 and 2:

state number	α^k	d^k
1	$\alpha^1 = 0.3141$	$d^0 = 0.3141$
6	$\alpha^2 = 0.4844$	$d^1 = 0.1703$
8	$\alpha^3 = 0.5053$	$d^2 = 0.0210$
5	$\alpha^4 = 0.5806$	$d^3 = 0.0753$
2	$\alpha^5 = 0.7685$	$d^4 = 0.1878$
7	$\alpha^6 = 0.7716$	$d^5 = 0.0031$
4	$\alpha^7 = 0.7900$	$d^6 = 0.0184$
9	$\alpha^8 = 0.7990$	$d^7 = 0.0090$
3	$\alpha^9 = 1$	$d^8 = 0.2010$

Questions 3 and 4:

Assume that $\alpha^3 > \alpha = 0.5 > \alpha^2$, then

$$\pi = (0, 1, 1, 0, 1, 0, 1, 0, 1, 0, 0, 1, 1, 0, 1, 1, 1, 0)$$

$$\begin{aligned}\rho &= (0, 0.2871, 0.2426, 0, 0.1563, 0, 0.0396, 0, 0.0736, 0, \\ &\quad 0, 0.1448, 0.0100, 0, 0.0046, 0.0125, 0.0288, 0)\end{aligned}$$

$$\begin{aligned}ad &= \sum_{k=0}^{\infty} P(nc = k)k \\ &= \sum_{k=1}^{4} \sum_{u} \rho(x_k, u)x_k \\ &= 0.3699\ [words]\end{aligned}$$

8 Conclusions and Further Research

In this research we derived a new solution methods for constrained Markov decision processes and considered applications of these methods to the optimization of wireless communications. In this paper we concentrated on the problem of infinite amount of information transfer over a fading channel. We seceded to minimize the transmission delay under the average power constraint where at each slot we use one of the available discrete power levels, while maximal power level is limited by a peak power constraint. Moreover we developed an innovative algorithm which ,aside of optimization problem solving, is able to show sensitivity of the solution to changes in the average power level. The results show that by using a new developed algorithm we can get a general and simple solution of the problem.

For further research we would like to propose the following:

- Extension of the fundamental Theorems to the case with arbitrary number of constraints.
- Extension of the constraints to inequality cases.
- Extension to finite horizon problems.

Appendix A: Linear Programming (LP) and Simplex Method Idea

In this Appendix we will discuss about LP and usage of Simplex Method for its solution.

Consider the following linear programming problem.

$$Minimize \qquad c_1x_1 + c_2x_2 + ... + c_nx_n$$

$$\begin{array}{rrcl} Subject\ to & a_{11}x_1 + a_{12}x_1 + ... + a_{1n}x_n & \geq & b_1 \\ & a_{21}x_1 + a_{22}x_1 + ... + a_{2n}x_n & \geq & b_2 \\ & & & . \\ & & & . \\ & & & . \\ & a_{m1}x_1 + a_{m2}x_1 + ... + a_{mn}x_n & \geq & b_m \\ & x_1, \qquad x_2, .., . \qquad , x_n & \geq & 0 \end{array}$$

Here $c_1x_1 + c_2x_2 + ... + c_nx_n$ is the objective function to be minimized. The coefficients $c_1, c_2, ..., c_n$ are the cost coefficients and $x_1, x_2, ..., x_n$ are the decision variables to be determined. The inequality $\sum_{j=1}^{n} a_{ij}x_j \geq b_i$ denotes the ith constraint. The coefficients a_{ij} for $i = 1, 2, ..., m, \quad j = 1, 2, ..., n$ are called the technological coefficients. These technological coefficients form the constraint matrix A.

$$A = \begin{bmatrix} a_{11} & a_{12} & . & . & . & a_{1n} \\ a_{21} & a_{22} & . & . & . & a_{2n} \\ . & . & & & & . \\ . & . & & & & . \\ . & . & & & & . \\ a_{m1} & a_{m2} & . & . & . & a_{mn} \end{bmatrix}$$

The column vector whose ith component is b_i, which is referred to as the right-hand side vector, represents the minimal requirements to be satisfied. The constraints $x_1, x_2, ..., x_n \geq 0$ are the nonnegativity constraints. A set of variables $x_1, x_2, ..., x_n$ satisfying all the constraints is called a feasible point or a feasible vector. The set of all such points constitutes the feasible region or the feasible space.

The linear programming problem can be stated as follows: Among all feasible vectors, find one that minimizes (or maximizes) the objective function.

The idea of the finding of the LP problem solution can be represented Geometrically in the Figure 12. Let consider the following problem

$$\begin{aligned} &Minimize \quad c \cdot x \\ &s.t. \quad A \cdot x \geq b \\ &x \geq 0 \end{aligned}$$

A feasible region consists of all vectors x satisfying $Ax \geq b$ and $x \geq 0$. Among all such points we wish to find a point with minimal cx value. In order to minimize the cost we need to move in the direction that do this minimization. This direction is $-c$, and hence the plane is moved in this direction as much as possible. This process is illustrated in Figure 12, where x^* is an optimal point. Needless to say, for a maximization problem, we need to move as much as possible in the direction c.

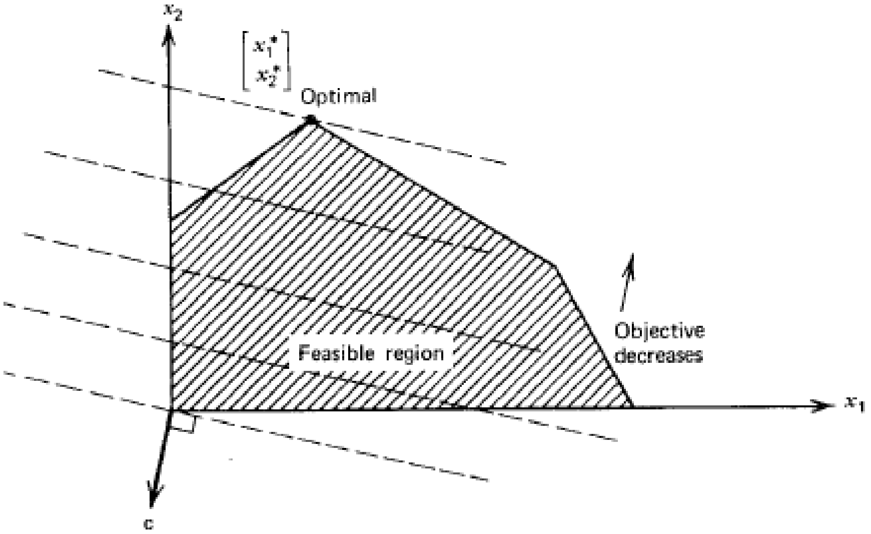


Figure 12: Geometric Solution

Now we will give a brief explanation of the Simplex method idea.

Consider the system $A \cdot x = b$ and $x \geq 0$, where A is an $m \times n$ matrix and b is a vector of length m. Suppose that $rank(A, b) = rank(A) = m$. After possibly rearranging the columns of A, let $A = [B, N]$ where B is an $m \times m$ invertible matrix and N is an $m \times (n-m)$ matrix. The solution $x = \begin{bmatrix} X_B \\ X_N \end{bmatrix}$ to the equations $Ax = b$, where

$$X_B = B^{-1} \cdot b$$

and

$$X_N = 0$$

is called a basic solution of the system. If $X_B \geq 0$, then x is called a basic feasible solution of the system. Here B is called the basic matrix (or simply the basis) and N is called the nonbasic matrix. The components of X_B are called basic variables and the the components of X_N are called nonbasic variables.

The key to the simplex method lies in recognizing the optimality of a given basic feasible

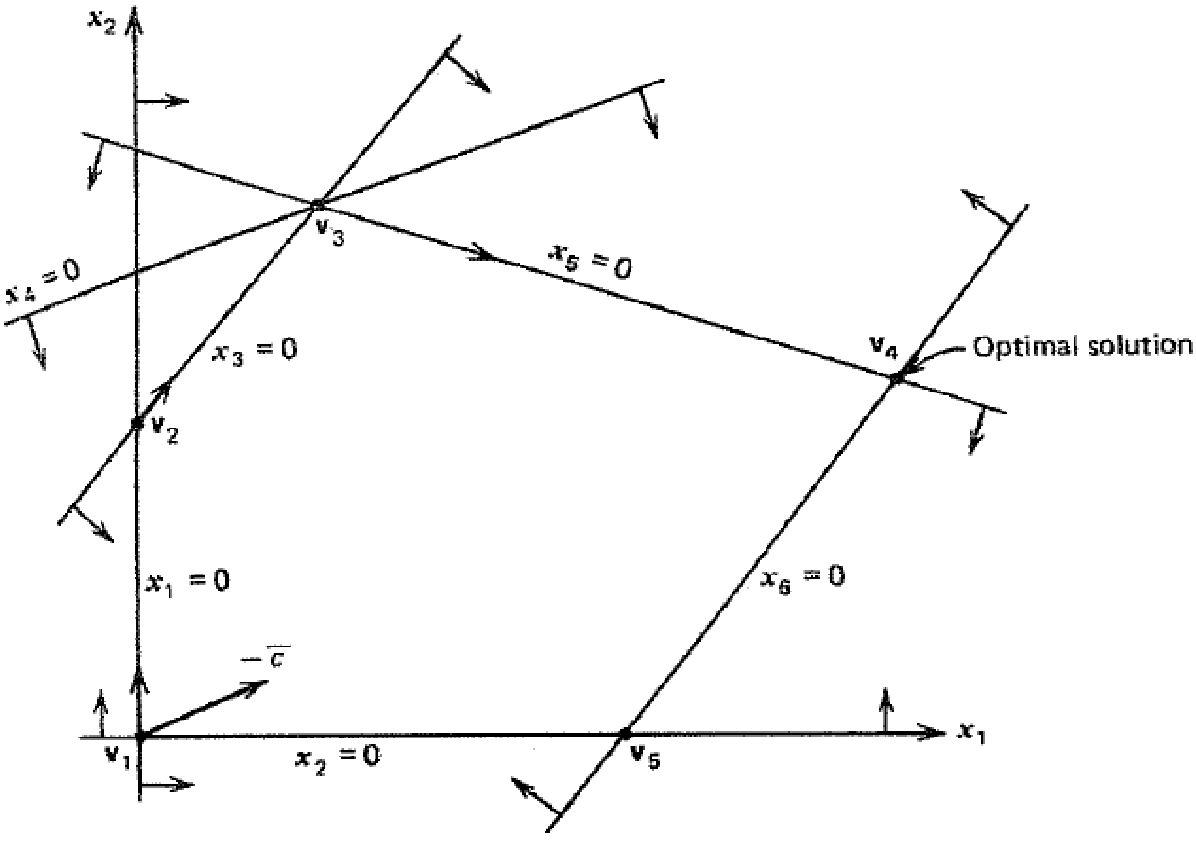


Figure 13: Simplex Method - Geometric Motivation

solution (extreme point solution) based on local considerations without having to (globally) enumerate all basic feasible solutions.

Let discuss a Geometric motivation of Simplex method. Consider the following LP problem

$$\begin{aligned}&\min_x \ c\cdot x\\ &A\cdot x=b\\ &x\geq 0\end{aligned}$$

Figure 13 illustrates the situation in 2 dimensions, where $n = 6$, $m = 4$ and where $X_B = \{x_3, x_4, x_5, x_6\}$. The defining variables associated with $n = 6$ hyperplanes on this figure. The vertices or extreme points of the feasible region are labeled as $v_1, ..., v_5$ and $\overline{c}$ is the cost vector. The current basic feasible solution corresponds to the vertex v_1,that is, the origin. Now consider the origin v_1 and examine the two defining hyperplanes associated with the corresponding nonbasic variables. Holding one of these hyperplanes binding and moving in a direction, feasible to the remaining hyperplane, takes us along a one-dimensional ray with vertex at v_1. There are two such rays. Holding $x_2 = 0$ and increasing x_1 takes us along the x_1 axis. Similarly, holding $x_1 = 0$ and increasing x_2 takes us along the x_2 axis. Assume that latter direction is better (from point of minimization view) then we move along the x_2 axis. Naturally we would like to move as far as possible along this edge since as far we are going as more we minimize the objective function (the angle between x_2 and $-\overline{c}$ is acute). However, our motion is blocked by the hyperplane $x_3 = 0$, since x_3 has been driven to zero and moving any further would drive x_3 negative. If no such blocking hyperplane existed, then the optimal solution value would have been unbounded. Furthermore, since one linearly independent hyperplane was binding all along and we were blocked by a new hyperplane, we now have two linearly independent hyperplanes binding, and so we are at another extreme point of the feasible region. At v_2 the nonbasic variables are x_1 and x_3, and the remaining variables are basic. We have now completed one step known as the iteration or pivot of the simplex method. In this step the variable x_2 is called the entering variable since it entered the set of basic variables, and the variable x_3 is called the blocking variable, or the leaving variable, since it blocked our motion or left the set of basic variables at v_1.

Repeating this process at v_2, we would of course not like to enter x_3 into the basis since it will only take us back along the reverse of the previous direction. However, holding $x_3 = 0$ and increasing $x_1$takes us along an improving edge because this direction is making an acute angle with $-\bar{c}$ direction (Figure 13). Proceeding in this direction we notice that more than one hyperplane blocks our motion. Suppose we arbitrarily choose one of them, namely $x_4 = 0$ as the blocking hyperplane. Hence x_4 is the leaving variable, and for current basis representation of v_3, x_3 and x_4 are the nonbasic variables. Now if we hold $x_4 = 0$ and move in a direction along which x_3 increases, the objective function value decreases since this direction is making acute angle with $-\bar{c}$ direction. However, this direction is not feasible. We are blocked by the hyperplane $x_5 = 0$ even before we begin to move. That is, x_5 leaves the basis giving x_4 and x_5 as our new non basis variables, while we are still at the same vertex v_3. With x_4 and x_5 as the nonbasic variables at v_3, holding $x_5 = 0$ and increasing x_4 takes us along an improving, feasible edge direction. The blocking variable that gets driven to zero is x_6, and the new nonbasic variables are x_5 and x_6. Observe that corresponding to this basis, none of the two rays, defined by holding one of the nonbasic variables equal to zero and increasing the remaining nonbasic variable, lead to an improvement in the objective function, and so our "key result" declares v_4 to optimal solution.

Appendix B: Principle of Optimality and Value Iteration Algorithm

The dynamic programming (DP) technique rests on a very simple idea, the principle of optimality. Roughly, the principle of optimality states the following rather obvious fact.

Principle of Optimality: [4]

Let $\pi^* = \{\pi_0, \pi_1, ..., \pi_{N-1}\}$ be an optimal policy for the basic problem, and assume that when using π^*, a given state x_i occurs at time i with positive probability. Consider the subproblem whereby we are at x_i and wish to minimize the "cost-to-go" from time i to time N.

$$E\left\{g_N(x_N) + \sum_{k=i}^{N-1} g_k(x_k, \pi_k(x_k), w_k)\right\}$$

. Then the truncated policy $\{\pi_i, \pi_{i+1}, ..., \pi_{N-1}\}$ is optimal for this subproblem.

The intuitive justification of the principle of optimality is very simple. If the truncated policy $\{\pi_i, \pi_{i+1}, ..., \pi_{N-1}\}$ were not optimal as stated, we would be able to reduce the cost further by switching to an otptimal policy for the subproblem once we reach x_i. The principle of optimality suggests that an optimal policy can be constructed in piecemeal fashion, first constructing an optimal policy for the "tail subproblem" involving the last stage, then extending the optimal policy to the "tail subproblem" involving the last two stages, and continuing in this manner until an optimal policy for the entire problem is constructed.

Value iteration is the most widely used and best understood algorithm for solving discounted Markov decision problems. The following value iteration algorithm finds a stationary ε-optimal policy, $(d_\varepsilon)^\infty$, and an approximation to its value.

Value Iteration Algorithm: [8]

1. Select $v^0 \in V$, specify $\varepsilon > 0$, and set $n = 0$.

2. For each $x \in X$, compute $v^{n+1}(x)$ by

$$v^{n+1}(x) = \max_{u \in U(x)} \{r(x,u) + \sum_{j \in X} \beta P(j|x,u) v^n(j)\} \tag{60}$$

where $r(x,u)$ is a reward at state $x \in X$ using action $u \in U$, and $P(j|x,u)$ is a transition probability to state j from state x using action u.

3. If

$$\left\|v^{n+1} - v^n\right\| < \varepsilon \frac{1-\beta}{2\beta} \tag{61}$$

go to step 4. Otherwise increment n by 1 and return to step 2.

4. For each $x \in X$, choose

$$d_\varepsilon(x) \in \arg \max_{u \in U(x)} \{r(x,u) + \sum_{j \in X} \beta P(i|x,u) v^n(j)\} \tag{62}$$

and stop.

The following theorem provides the main results regarding convergence of the above value iteration algorithm.

Theorem ([8], Theorem 6.3.1)

Let $v^0 \in V$, $\varepsilon > 0$, and let $\{v^n\}$ satisfy (60) for $n \geq 1$. Then

- v^n converges in norm to v^*_β,
- finite N for which (61) holds for all $n \geq N$,
- the stationary policy $(d_\varepsilon)^\infty$ defined in (62) is ε-optimal, and
- $\left\|v^{n+1} - v^*_\beta\right\| < \varepsilon/2$ whenever (61) holds.

Appendix C: BPSK Modulation in Fading Environment

In this section we will give a brief description of the BPSK modulation [6] and an impact of the channel fading for the error probability calculation [7].

Definitions:

- E_b is the transmitted signal energy per bit.
- T_b is the time duration of one bit transmission.
- f_c is the carrier frequency.

BPSK (Binary phase-shift keying) is a modulation where, the pair of signals, $s_1(t)$ and $s_2(t)$, used to represent binary symbols 1 and 0, respectively, are defined by

$$\begin{aligned} s_1(t) &= \sqrt{\frac{2E_b}{T_b}}\cos(2\pi f_c t) \qquad (63) \\ s_2(t) &= \sqrt{\frac{2E_b}{T_b}}\cos(2\pi f_c t + \pi) \\ &= -\sqrt{\frac{2E_b}{T_b}}\cos(2\pi f_c t) \qquad (64) \end{aligned}$$

A pair of sinusoidal waves, $s_1(t)$ and $s_2(t)$, differing only in phase by 180 degrees, as defined above, are named antipodal signals. Using Gram-Schmidt orthogonalization we find a single basis function, namely

$$\phi_1(t) = \sqrt{\frac{2}{T_b}}\cos(2\pi f_c t), \qquad 0 \le t \le T_b$$

Thus, we may express the transmitted signals $s_1(t)$ and $s_2(t)$ in terms of $\phi_1(t)$ as follows

$$\begin{aligned} s_1(t) &= \sqrt{E_b}\phi_1(t) \\ s_2(t) &= -\sqrt{E_b}\phi_1(t) \end{aligned}$$

Here we will skip the part of probability of error derivation because it quite technical and can be seen in [6]. We will give only the final formula:

$$P_r\{\varepsilon\} = Q\left(\sqrt{\frac{2E_b}{N_0}}\right)$$

where

$$Q\left(x\right) = \frac{1}{\sqrt{2\pi}}\int_x^{\infty} \exp(-\frac{z^2}{2})dz$$

For a fading channel we have one additional modification, this is factor r^2. For block fading Gaussian channel in which the fading of each slot is *iid* according to Rayleigh distribution, the error probability for one bit transmission can be expressed by a following formulas [6] and [7]:

$$P_r\{\varepsilon\} = Q\left(\sqrt{\frac{2r^2E_b}{N_0}}\right)$$

where Rayleigh distribution of random variable r is expressed as follows

$$f(r) = \frac{r}{\sigma^2}e^{-\frac{r^2}{2\sigma^2}}, r \geq 0$$

Because we used letter σ for another purpose (initial distribution) before, we will replace σ^2 here by letter ζ^2, and from now on

$$f(r) = \frac{r}{\zeta^2}e^{-\frac{r^2}{2\zeta^2}}, r \geq 0$$

$$\begin{aligned} P_r\{\varepsilon\}_{avr} &= \int_0^{\infty} Q\left(\sqrt{\frac{2r^2E_b}{N_0}}\right) f(r)dr = \frac{1}{2}\left[1 - \sqrt{\frac{\overline{E}_b/N_0}{1+\overline{E}_b/N_0}}\right] \\ \text{where } \overline{E}_b &= E\{r^2E_b\} = 2\zeta^2E_b \end{aligned}$$

⇒ A probability of one bit successful transmission is expressed as follows

$$P_{succ} = P_s = 1 - \frac{1}{2}\left[1 - \sqrt{\frac{\overline{E}_b/N_0}{1+\overline{E}_b/N_0}}\right] \tag{65}$$

References

[1] E. Altman, Constrained Markov Decision Processes, Chapman&Hall/CRC,1998

[2] E. Altman, Applications of Markov Decision Processes in Communication Networks, INRIA B.P. 1993

[3] M.S. Bazaraa, J.J. Jarvis, H.D. Sherali, "Linear Programming and Network Flows", John Wiley&Sons, 1990

[4] D.P. Bertsekas, Dynamic Programming and Optimal Control, Athena Scientific, Belmont, Massachusetts, 1995

[5] I. Bettesh, Information and Network Theory Aspects of Communication Systems in Fading Environment, PhD Thesis, October 2002

[6] Simon Haykin, Communication Systems, Second Edition, John Wiley&Sons, 1983

[7] I. Kalet, Lecture notes - Cellular Communications, Winter 2003

[8] Martin L. Puterman, Markov Decision Processes, John Wiley&Sons, 1994

[9] S.M. Ross, Introduction to Stochastic Dynamic Programming,Academic Press, 1983

[10] A. Shwartz, Lecture notes - Control of Stochastic Processes, October 2003

[11] J. Walrand, An Introduction to Queueing Networks, Prentice Hall, 1988

[12] H. Wang and N.B. Mandayam, Opportunistic File Transfer over a Fading Channel under Energy and Delay Constraints, Asilomar Conference on Signals, Systems and Computers, Monterey, California, November 2001

תהליכי החלטה מרקובים מאולצים
ויישומם בתקשורת אלחותית

אלכסנדר זדורוז׳ני

תהליכי החלטה מרקובים מאולצים ויישומם בתקשורת אלחוטית

חיבור על עבודת גמר

לשם מילוי חלקי של הדרישות לקבלת התואר

מגיסטר למדעים בהנדסת חשמל

אלכסנדר זדורוז'ני

הוגש לסנט הטכניון – מכון טכנולוגי לישראל

סיוון תשס"ד חיפה יוני 2004

עבודת הגמר נעשתה בהנחיית פרופ׳ אדם שורץ

בפקולטה להנדסת חשמל.

תודתי והוקרתי לפרופ׳ אדם שורץ

על הדרכתו ועזרתו

בכל שלבי המחקר .

אני מודה לטכניון על תמיכה הכספית הנדיבה

בהשתלמותי.

תוכן העניינים

עמוד

תקציר..1

רשימת סמלים וקיצורים...2

1. מבוא..4

2. סקר ספרות ומטרות המחקר..5

3. מודל של תקשורת אלחוטית...7

4. תהליכי החלטה מרקובים מאולצים (CMDP) וגישה של תכנות ליניארי..10

4.1 CMDP..10

4.2 מדיניות אופטימאלית עבור בעיות של CMDP......................................12

4.3 מידת ההסתברות ותכנות ליניארי (LP)..13

5. מדיניות אופטימאלית עבור CMDP..18

5.1 תנאי קרוש-קון-תקר (KKT)..18

5.2 משפטים יסודיים...22

6. תכונות של מדיניות אופטימאלית עבור בעיות של חיסכון בהספק...29

6.1 אלגוריתם CMDPS עבור שתי רמות בקרה...35

6.1.1 אלגוריתם CNDPS...37

6.1.2 שימוש באלגוריתם CMDPS..39

6.1.3 דוגמא לשימוש באלגוריתם CMDPS בבעיות CMDP...............................45

6.2 אלגוריתם CMDPS עבור מספר סופי של רמות הבקרה............................50

תוכן העניינים (המשך)

עמוד

7. גישה חדשה עבור אופטימיזציה בתקשורת אלחוטית................55

7.1 אופטימיזציה של בעיית תקשורת ע"י שימוש בגישת CMDP.................58

7.2 שימוש באלגוריתם CMDPS בתקשורת אלחוטית 62

8. מסקנות ומחקר עתידי...66

נספח A – תכנות ליניארי ורעיון של שיטת הסימפלקס...................67

נספח B – עיקרון האופטימאליות ואלגוריתם Value Iteration.......73

נספח C – אפנון BPSK בערוצי דעיכה...75

מקורות...77

תקציר (עברית)..I

רשימת ציורים

מספר **עמוד**

1. מודל של תקשורת אלחוטית.....8
2. שרשרת מרקוב עבור הוכחה של מונוטוניות.....30
3. זוג של מערכות.....30
4. שינוי הרנדומיזציה.....36
5. דיאגראמה של אלגוריתם CMDPS.....40
6. שרשרת מרקוב עבור דוגמא ראשונה 46
7. תוצאות של שיטת הסימפלקס.....51
8. התנהגות הרנדומיזציה בדוגמא הראשונה.....52
9. שרשרת מרקוב עבור בעיה של תקשורת אלחוטית.....56
10. שרשרת מרקוב מפושטת עבור בעיה של תקשורת אלחוטית.....57
11. שרשרת מרקוב עבור דוגמא של בעיית תקשורת.....63
12. פתרון גיאומטרי.....69
13. שיטת הסימפלקס – מוטיבציה גיאומטרית.....70

תקציר

ההתפתחות בשירותי תקשורת אישיים והשימוש בטלפונים הסלולאריים ומחשבים ניידים גוברים משנה לשנה. לכן חיסכון בצריכת הספק הסוללה באמצעי תקשורת אלה הוא בעל חשיבות רבה. אחד הגורמים המרכזיים בתמיכה בתקשורת אלחוטית הוא חיסכון מרבי בצריכת ההספק. מצד אחד חיסכון זה הוא מאד חשוב, ומצד שני הוא יכול לגרום לבעיה רצינית בתפקודה של תקשורת אלחוטית אמינה – הספק נמוך יכול לגרום לשגיאות שידור ופגיעה באיכות התקשורת. לכן בזמן שמנסים לחסוך הספק חייבים לדאוג לאיכות השירות של המשתמש.

בשנים האחרונות הוצעו מספר שיטות הפתרון לבעיה זו. בדרך כלל פתרונות אלה מוצאים מדיניות אופטימאלית עבור מזעור ההשהיה הממוצעת תוך אילוץ על ההספק הממוצע. כמה מהן פותחו עבור פתרון של הבעיות בזמן רציף, אחרות עבור זמן בדיד. שיטות שמטפלות בבעיות בזמן רציף הן כלליות יותר אבל לא תמיד אפשר ליישם אותן בתקשורת ספרתית שהוא סוג התקשורת הנפוץ ביותר בשנים האחרונות. שיטות שפותחו עד עכשיו, עבור זמן בדיד, לא מספיק כלליות בהרבה מקרים. למשל הן מוגבלות רק עבור שתי רמות בקרה. בדרך כלל בתקשורת ספרתית אנחנו מתעניינים במציאת מדיניות האופטימאלית שממזערת השהיה ממוצעת תחת אילוץ על הספק הממוצע ומקסימאלי ומשתמשת רק ברמות הבקרה הנתונות. אף אחד מהשיטות הקיימות לא יכולה לספק את זה.

במחקר הזה אנחנו מפתחים שיטות פתרון חדשות עבור תהליכי החלטה מרקובים מאולצים ודנים בישומן בבעיות אופטימיזציה של תקשורת אלחוטית. אנחנו סוקרים את שיטות הבקרה הקיימות עם ניתוח ביצועיהן. במחקר אנחנו מציעים שיטה חדשה של בקרת הספק בתקשורת אלחוטית שממזערת את ההשהיה הממוצעת תחת אילוץ על הספק ממוצע ומקסימאלי, כאשר אנחנו מוגבלים לשימוש רק ברמות הבקרה הנתונות. השיטה הזאת מבוססת על טכניקה שפותחה במהלך המחקר עבור תהליכי החלטה מרקובים עם אילוצים.

יותר מכך אנחנו פיתחנו אלגוריתם שמסוגל חוץ מפתרון של בעיית האופטימיזציה גם לנתח את הרגישות של הפתרון האופטימאלי לשינוים ברמת ההספק הממוצע באילוץ.
בעבודה זאת אנחנו מתארים מצב כאשר סוג אחד של מחיר (השהיה, קצב העברה וכדומה) צריך להיות ממוזער תוך כדי שסוג אחר של מחיר (הספק, השהיה...) נשאר מתחת לרמה נתונה. לכן זוהי בעיית אופטימיזציה מאולצת.

רשתות תקשורת בנויות בצורה כזאת המאפשרת שידור סימולטאני של סוגי מידע שונים: קול, העברת קבצים, וידיאו... בדרך כלל מדד לביצועים הוא השהיית השידור, צריכת ההספק, הסתברות השגיאה בשידור... סוגים שונים של מידע מובדלים אחד מהשני לפי תכונות סטטיסטית שלהם וגם ע"י דרישות שונות עבור הביצועים. לדוגמא, עבור שידור אינטראקטיבי של הודעות דרוש שההשהיה הממוצעת מקצה לקצה תהיה מוגבלת. ההגבלה על ההשהיה חשובה מאד בשידורי מידע קולי- ההשהיה מוגבלת ל 0.1 שניות. במידע וההשהיה עוברת את הגבול המותר, איכות הקול יורדת בצורה דרמטית. עבור העברת קבצים לא אינטראקטיבית אנחנו בדרך כלל רוצים למזער את ההשהיה ולמקסם את קצב ההעברה.

בבעיות אופטימיזציה מאולצות תמיד קיימת ניגודיות בין מזעור (או מקסום) של סוג אחד של מחיר, למשל השהיה או קצב העברה, תוך כדי שסוג אחר של המחיר (צריכת הספק או השהיה) נמצא מתחת לרמה הנדרשת. כדי למזער את ההשהיה אנחנו צריכים לשדר בהספק הכי גבוה שאפשר, מפני שזה יגדיל את הסתברות ההצלחה של השידור, שיקטין את מספר השידורים החוזרים.
בניגוד לזה, על מנת להקטין את ההספק הנצרך אנחנו מעוניינים לשדר בהספק המינימאלי האפשרי. הבעיה שלנו ניתנת לניסוח כבעיית של תהליכי ההחלטה מרקובים עם אילוצים, בה אנחנו שואפים למזער את המחיר המיוחס להשהיה כתלות באילוץ על ההספק הממוצע והמקסימאלי.

אנחנו יכולים לחלק את העבודות בתחום של חיסכון באנרגיה לשני סוגים.
ראשון – בקרת הצריכה באנרגיה בהנחה שיש לנו הספקה סופית של מידע לשידור.
שני – בקרת הצריכה באנרגיה בהנחה שיש לנו הספקה אינסופית של מידע לשידור.
במחקר הזה אנחנו מתרכזים בסוג שני של הבעיות – הספקה אינסופית של מידע לשידור. בעבודה הזו אנחנו מנסים להרחיב את השיטות הקיימות למקרים יותר כלליים – הרחבה משתי רמות הבקרה למספר כלשהו, אבל סופי, של רמות הבקרה. בנוסף אנחנו מנתחים את המבנה האופטימאלי של הפתרון ממבט שלא היה ידוע עד עכשיו - "רגישות" של המדיניות האופטימאלית לשינויים באילוץ. בהקשר של בעיית התקשורת זה מושג חדש שהוגדר תוך כדי המחקר. את הנושא הזה אנחנו הצלחנו לנתח בצורה די כללית – גם מבחינה אנליטית וגם מבחינה נומרית, ובגישה חדשה. התוצאות שהתקבלו הן שעבור שינויים מספיק קטנים ברמת האילוץ, המדיניות האופטימאלית, במצבים שהיא דטרמיניסטית, לא משתנה. הצלחנו גם למצוא את התחומים המדויקים עבור השינויים האלה.

במהלך המחקר אנחנו גם הצלחנו לפתח שני משפטים שמתארים את המבנה של הפתרון האופטימאלי עבור תהליכי החלטה מרקובים מאולצים כלליים. שני משפטים נוספים פותחו עבור הנושא של חיסכון בהספק. במהלך המחקר גם פותח אלגוריתם מקורי שבאמצעותו אנחנו יכולים לפתור בעיות של תהליכי החלטה מרקובים מאולצים וגם לנתח רגישות הפתרון לשינויים ברמת האילוץ. התוצאות שהתקבלו ע"י הרצת האלגוריתם הזה הושווה לתוצאות שהתקבלו ע"י הפעלה של שיטת הסימפלקס עם יתרון ברור לאלגוריתם החדש שפיתחנו.

בסופו של דבר הצלחנו ליישם את שיטת הפתרון החדשה שפיתחנו עבור בעיות בתקשורת אלחוטית. היתרונות העיקריים של שיטה החדשה הם פשטות, כלליות, וחידוש יחסית לשיטות הקודמות.