ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΙΓΑΙΟΥ

ΤΜΗΜΑ ΜΑΘΗΜΑΤΙΚΩΝ

Πρόγραμμα Μεταπτυχιακών Σπουδών

Σπουδές στα Μαθηματικά

Μελέτη διαφόρων προβλημάτων βέλτιστης συντήρησης και αντικατάστασης μηχανημάτων με αριθμητικά παραδείγματα

Επιβλέπων καθηγητής: κ. Θεοδόσης Δημητράκος

Μέλη επιτροπής: κ. Νικόλαος Δαφνής και κ. Βασίλειος Κουκουλογιάννης

ΣΑΜΟΣ 2024

ΕΥΧΑΡΙΣΤΙΕΣ

Η παρούσα διπλωματική εργασία εκπονήθηκε στο Τμήμα Μαθηματικών της Σχολής Θετικών Επιστημών του Πανεπιστημίου Αιγαίου στα πλαίσια του μεταπτυχιακού προγράμματος «Σπουδές στα Μαθηματικά» με κατεύθυνση στα Εφαρμοσμένα Μαθηματικά.

Ευχαριστώ θερμά τον καθηγητή μου κ. Δημητράκο Θεοδόση, Αναπληρωτή Καθηγητή του Τμήματος Μαθηματικών του Πανεπιστημίου Αιγαίου, ο οποίος με τίμησε με την επίβλεψη της διπλωματικής μου εργασίας.

Θα ήθελα επίσης να ευχαριστήσω τους καθηγητές της επιτροπής που δέχτηκαν να με τιμήσουν αφιερώνοντας λίγο από το χρόνο τους για την αξιολόγηση της εργασίας μου.

Τέλος ευχαριστώ τους γονείς μου Μανώλη και Ελένη, τον αδερφό μου Κωνσταντίνο, καθώς και τα υπόλοιπα μέλη της οικογένειας μου, για την αγάπη τους και την υποστήριξη που μου δείχνουν σε κάθε στιγμή της ζωής μου.

ΠΕΡΙΛΗΨΗ

Στην παρούσα διπλωματική εργασία μελετήθηκαν διάφορα προβλήματα βέλτιστης συντήρησης μηχανημάτων με αριθμητικά παραδείγματα.

Αρχικά, στο πρώτο κεφάλαιο γίνεται μια ιστορική αναδρομή των Μαρκοβιανών διαδικασιών αποφάσεων (ΜΔΑ), συνοδευόμενη από τις βασικές έννοιες της θεωρίας αυτής. Οι Μαρκοβιανές διαδικασίες αποφάσεων μπορούν να μας υποδείξουν εκείνες τις ενέργειες οι οποίες ελαχιστοποιούν το κόστος για ένα δεδομένο πρόβλημα ή αντίστοιχα μεγιστοποιούν το κέρδος μας.

Στο δεύτερο κεφάλαιο, παρουσιάζεται η θεωρία βελτιστοποίησης τόσο σε πεπερασμένο όσο και σε άπειρο χρονικό ορίζοντα. Το κεφάλαιο αυτό πλαισιώνεται από τρεις διαφορετικούς τρόπους επίλυσης προβλημάτων βέλτιστης συντήρησης, μαζί με από τους αντίστοιχους αλγορίθμους.

Στο τρίτο κεφάλαιο περιέχονται ενδεικτικά παραδείγματα. Τα παραδείγματα αυτά που αναφέρονται στο συγκεκριμένο κεφάλαιο λύνονται με τους προαναφερθέντες αλγορίθμους.

Τέλος, στο τελευταίο κεφάλαιο παρουσιάζεται ένα επιλεγμένο παράδειγμα από το τρίτο κεφάλαιο, το οποίο έχει επιλυθεί και με την χρήση της MATLAB. Το παρόν κεφαλαίο έχει κομμάτια του κώδικα και κάποιες παρατηρήσεις που έχω κάνει λόγω ορισμένων δοκιμών στον κώδικα της MATLAB.

Η παρούσα διπλωματική εργασία είναι έργο ανθρώπινο. Εύχομαι με το έργο αυτό να βοηθήσω όποιον φοιτητή η συνάδελφο ενδιαφέρεται να γνωρίσει κάποιες εισαγωγικές πτυχές των Μαρκοβιανών διαδικασιών αποφάσεων για περαιτέρω έρευνα και εμβάθυνση τόσο στη θεωρία, όσο και στις εφαρμογές αυτού του τομέα των Μαθηματικών.

Λέξεις κλειδιά: Μαρκοβιανές Διαδικασίες Αποφάσεων, Βέλτιστη Συντήρηση, Αλγόριθμος Βελτίωσης των Πολιτικών, Αλγόριθμος Γραμμικού Προγραμματισμού, Αλγόριθμος Διαδοχικών Προσεγγίσεων.

ABSTRACT

In the present thesis, various problems of optimal maintenance of machines were studied with numerical examples. Initially, the first chapter provides a historical overview of Markov Decision Processes (MDPs), accompanied by the fundamental concepts of this theory. Markov Decision Processes can guide us in identifying actions that minimize costs for a given problem or, conversely, maximize our profits.

The second chapter presents the theory of optimization in both finite and infinite time horizons. This chapter is accompanied by three different approaches to solving optimal maintenance problems, along with their respective algorithms.

The third chapter contains illustrative examples. The examples mentioned in this chapter are solved using the algorithms mentioned earlier.

Finally, in the last chapter, a selected example from the third chapter is presented, which has been solved using MATLAB. This chapter includes snippets of code and some observations I have made due to certain tests on the MATLAB code.

This thesis is a human endeavor. With this work, I hope to assist any student or colleague interested in exploring introductory aspects of Markov Decision Processes for further research and deepening, both in theory and applications in the field of Mathematics.

Keywords: Markov Decision Processes, Optimal Maintenance, Policy-Iteration Algorithm, Linear Programming Algorithm, Value -Iteration Algorithm.

ΠΕΡΙΕΧΟΜΕΝΑ

Κεφάλαιο 1ο Εισαγωγή

1.1 Ιστορική αναδρομή

Οι Μαρκοβιανές αλυσίδες αποφάσεων (Markov decision processes) αποτελούν σημείο τομής δύο ιδιαίτερα εφαρμοσμένων κλάδων της Επιστήμης των Μαθηματικών: των Στοχαστικών Ανελίξεων και της Επιχειρησιακής Έρευνας.

Η Επιχειρησιακή Έρευνα πραγματεύεται την μαθηματική μοντελοποίηση και επίλυση προβλημάτων βελτιστοποίησης. Αυτό σημαίνει ότι στόχος της Επιχειρησιακής Έρευνας είναι η ανάπτυξη επιστημονικών μεθόδων για τον καλύτερο σχεδιασμό και έλεγχο οργανωμένων συστημάτων, ώστε η λειτουργία τους να επιτυγχάνει το βέλτιστο, κάθε φορά αποτέλεσμα. Με τον όρο σύστημα εννοούμε ένα σύνολο που αποτελείται από πράγματα με στενή σχέση ή αλληλεξάρτηση μεταξύ τους, που συνδυάζονται για να πραγματώσουν ένα στόχο. Ο όρος βέλτιστο αποτέλεσμα παραπέμπει συνήθως σε προβλήματα μεγιστοποίησης κέρδους ή ελαχιστοποίησης κόστους (αρνητικό κέρδος).

Το εύρος των εφαρμογών των Στοχαστικών Ανελίξεων είναι πραγματικά εντυπωσιακό, περιλαμβάνοντας μεταξύ άλλων εφαρμογές σε διάφορους κλάδους, όπως της Φυσικής, της Ιατρικής, της Βιολογίας, της Γενετικής, της Κοινωνιολογίας και της Ψυχολογίας. Ειδικότερα, για παράδειγμα στα Οικονομικά, μπορεί να σχετίζονται με τις ασφαλίσεις, τις επενδύσεις και τις χρηματοοικονομικές αναλύσεις, στην Επιχειρησιακή Έρευνα, με τις ουρές αναμονής και εξυπηρέτησης και στην Πληροφορική με τις τηλεπικοινωνίες και τα δίκτυα.

Όσον αφορά τις Μαρκοβιανές Διαδικασίες, με αυτές μελετούμε στοχαστικά συστήματα που έχουν την Μαρκοβιανή ιδιότητα, δηλαδή συστήματα των οποίων η μελλοντική εξέλιξη εξαρτάται μόνο από την παρούσα κατάστασή τους και όχι από τον παρελθόντα χρόνο. Ο Ρώσος μαθηματικός Andrey Andreyevich Markov (1856-1922) ήταν ο πρώτος που εισήγαγε και μελέτησε την εν λόγω ιδιότητα, η οποία χαρακτηρίζει πολλά πραγματικά συστήματα και ως εκ τούτου οι εφαρμογές της παρουσιάζουν μεγάλη ποικιλομορφία.

Οι Μαρκοβιανές διαδικασίες αποφάσεων (Μ.Δ.Α) αποτελούν υποπερίπτωση των Μαρκοβιανών Ανελίξεων με μεγάλη χρησιμότητα στην Επιχειρησιακή Έρευνα, από την οποία αντλούν διάφορα εργαλεία. Η αρχή τίθεται από την δεκαετία του 50 πιο συγκεκριμένα το 1957, όταν ο R. Bellman αναπτύσσει την θεωρία του δυναμικού προγραμματισμού και εν συνεχεία στην δεκαετία του εξήντα, ο R. Howard στηριζόμενος στην θεωρία των Μαρκοβιανών αλυσίδων και τον δυναμικό προγραμματισμό ανακαλύπτει τον αλγόριθμο βελτίωσης των πολιτικών. Από το σημείο αυτό και έπειτα, η ανάπτυξη των Μαρκοβιανών διαδικασιών αποφάσεων είναι ραγδαία.

Στην εποχή μας, η ραγδαία ανάπτυξη της τεχνολογίας και της βιομηχανίας αποτελεί κύριο χαρακτηριστικό ανταγωνισμού μεταξύ των επιχειρήσεων. Η αξιολόγηση του εξοπλισμού τους πρόκειται για το βασικό αντικείμενο μελέτης των επιχειρήσεων. Η μελέτη αυτή μπορεί να βοηθήσει τις επιχειρήσεις για την βελτίωση, την ανάπτυξη καθώς και την ομαλή λειτουργία του εξοπλισμού τους. Η βλάβη του εξοπλισμού μπορεί να αναδειχτεί μείζον πρόβλημα για την παραγωγική διαδικασία. Έτσι κάθε επιχείρηση προσπαθεί να βρει έναν βέλτιστο τρόπο λειτουργίας με το ελάχιστο κόστος. Μια σημαντική λειτουργία για την παραγωγική διαδικασία είναι η συντήρηση του εξοπλισμού.

Η συντήρηση επιδιώκει την καλή λειτουργία των συστημάτων παραγωγής και τα ελέγχει σε περίπτωση πιθανής βλάβης. Έτσι καθυστερείτε η φθορά των μηχανημάτων και αυξάνεται η διάρκεια ζωής τους. Υπάρχουν δυο είδη συντήρησης που θα μελετήσουμε. Αρχικά η επιδιορθωτική συντήρηση η οποία απαιτείται σε περίπτωση βλάβης και η προληπτική συντήρηση που χρησιμοποιείται για την προστασία των μηχανημάτων. Το κόστος συντήρησης και για τις δυο περιπτώσεις θέλουμε να ελαχιστοποιείται κάθε φορά διατηρώντας παράλληλα όμως την αξιοπιστία του εξοπλισμού. Συνεπώς κάθε επιχείρηση αντιμετωπίζει το πρόβλημα εύρεσης της πολιτικής που οδηγεί στην βέλτιστη συντήρηση και αντικατάσταση των μηχανημάτων της. Η επίλυση πολλών προβλημάτων που σχετίζονται με τη βέλτιστη συντήρηση ή την αντικατάσταση μηχανημάτων πραγματοποιείται με τη χρήση των Μαρκοβιανών διαδικασιών.

1.2 Βασικές έννοιες

Μία στοχαστική διαδικασία (ή ανέλιξη) από το T στο S είναι μία συλλογή από τυχαίες μεταβλητές (τ.μ.) $\{X(t,\omega)\}_{t\in T,\omega\epsilon\Omega}$,ορισμένες σε ένα χώρο πιθανότητας$(\Omega,\mathcal{F},P)$, όπου:

$T=\mathbb{N}_0=\{0,1,2,\dots\}$ ή $T=[0,+\infty)$ ο **χρονικός ορίζοντας** (συνήθως το T είναι το σύνολο χρόνων).

S: ο **χώρος καταστάσεων** της διαδικασίας, δηλαδή το πεδίο τιμών της $X(t,\omega)$για κάθε $t\in T$ και $\omega\in\Omega$. Αν το S είναι αριθμήσιμο ή πεπερασμένο, κάνουμε λόγο για **στοχαστική αλυσίδα**.

Ω: είναι ο **δειγματικός χώρος** (δ.χ.), πιο συγκεκριμένα το σύνολο όλων των δυνατών αποτελεσμάτων του υπό μελέτη πειράματος τύχης.

$\mathcal{F}$: **σ-άλγεβρα των ενδεχομένων**, θα λέγαμε ότι είναι το πεδίο ορισμού των ενδεχομένων του Ω (υπάρχουν περιπτώσεις δ.χ. Ω, όπου δεν μπορούμε να θεωρήσουμε κάθε υποσύνολό του ως ενδεχόμενο).

P: **μέτρο πιθανότητας**, δηλαδή μία συνάρτηση $P:\mathcal{F}\to\mathbb{R}$ τέτοια ώστε:

1. $P(\Omega)=1$ (κανονικότητα),
2. $P(A)\geq 0,\forall A\in\mathcal{F}$ (μη αρνητικότητα),
3. $P\left(\bigcup_{v=1}^{\infty}A_v\right)=\sum_{v=1}^{\infty}P(A_v)$, $\forall$ ακολουθία $\{A_v\}_{v\epsilon\mathbb{N}}$ της $\mathcal{F}$ ξένων ανά δύο ενδεχομένων $\left(A_i\cap A_j=\emptyset,\forall i\neq j\right)$.

Μία στοχαστική διαδικασία μπορεί να παρατηρηθεί κάτω από δύο οπτικές γωνίες. Πρώτον για κάθε $t\in T$, η τ.μ. $X_t=\Omega\to S$ με $\omega\mapsto X_t(\omega)$ περιγράφει την κατάσταση της διαδικασίας την χρονική στιγμή t, ενώ η συλλογή όλων των X_t δείχνει την κατάσταση του συστήματος για κάθε χρονική στιγμή $t\in T$.

Ανάλογα τον χώρο καταστάσεων, οι στοχαστικές διαδικασίες μπορούν να διακριθούν στις παρακάτω κατηγορίες:

1. Αν $S\subset(-\infty,+\infty)$ τότε θα λέμε ότι έχουμε **στοχαστικές διαδικασίες άπειρου χώρου καταστάσεων**.
2. Αν $S\subset\{0,\pm1,\pm2,\dots\}$ τότε θα λέμε ότι έχουμε **στοχαστικές διαδικασίες πεπερασμένου χώρου καταστάσεων**.

Ομοίως ανάλογα με τον χρόνο στον όποιο εξελίσσονται οι στοχαστικές διαδικασίες διακρίνονται στις εξής περιπτώσεις:

1. Αν $T=[0,\infty),(0,a)$ τότε ονομάζονται **στοχαστικές διαδικασίες συνεχούς χρόνου**.
2. Αν $T=\{0,1,2,\dots\}$ τότε ονομάζονται **στοχαστικές διαδικασίες διακριτού χρόνου**.

1.3 Μαρκοβιανή αλυσίδα σε διακριτό χρόνο

Μια ακολουθία τυχαίων μεταβλητών $\{X_n\}$, με $n \in \mathbb{N}$, με χώρο καταστάσεων S, καλείται Μαρκοβιανή αλυσίδα αν και μόνο αν αυτή έχει τη Μαρκοβιανή ιδιότητα, δηλαδή αν ισχύει ότι:

$P\{X_{n+1} = i_{n+1} | X_n = i_n, X_{n-1} = i_{n-1}, \dots, X_1 = i_1, X_0 = i_0\} = P\{X_{n+1} = i_{n+1} | X_n = i\} = P_{ij}$ για κάθε $i_0, i_1, \dots, i_{n+1} \in S$.

Η δεσμευμένη πιθανότητα $P\{X_{n+1} = j | X_n = i\} = P_{ij}$, $i, j \in S$, $\forall n \in \mathbb{N}$ ονομάζεται πιθανότητα μετάβασης από την κατάσταση i στην κατάσταση j στο (n+1)–οστό βήμα. Οι πιθανότητες μετάβασης P_{ij} καλούνται πιθανότητες μετάβασης πρώτης τάξης και ικανοποιούν τις εξής ιδιότητες:

i. $P_{ij} \geq 0, i, j \in S$ (1.1)

ii. $\sum_{j \in S} P_{ij} = 1, i \in S$

Οι πιθανότητες μετάβασης πρώτης τάξης γράφονται και με τη μορφή ενός πίνακα, ο οποίος καλείται πίνακας μετάβασης πρώτης τάξης της αλυσίδας. Ο πίνακας ικανοποιεί τις σχέσεις (1.1), συμβολίζεται με P, ονομάζεται στοχαστικός πίνακας και έχει την παρακάτω μορφή:

$$\mathrm{P}^{(n)} = \begin{pmatrix} P_{00}^{(n)} & \mathrm{P}_{01}^{(n)} & \cdots & P_{0N}^{(n)} \\ P_{10}^{(n)} & \mathrm{P}_{11}^{(n)} & & P_{1N}^{(n)} \\ \vdots & & \ddots & \vdots \\ P_{N0}^{(n)} & P_{N1}^{(n)} & \dots & P_{NN}^{(n)} \end{pmatrix}, \ n=0,1,\dots..$$

Αξίζει να σημειωθεί ότι μία Μαρκοβιανή αλυσίδα είναι πλήρως ορισμένη όταν δίνεται ο πίνακας P και η αρχική κατανομή $\pi_0(j) = P(X_0 = j)$ (συνάρτηση πιθανότητας της αρχικής τ.μ. X_0). Σε αυτήν την περίπτωση, είναι εύκολο να διαπιστώσουμε ότι για να βρεθεί η αλυσίδα στις καταστάσεις $i_0, i_1, \dots, i_{n-1}, i_n \in S$ διαδοχικά, η πιθανότητα είναι:

$P(X_0 = i_0, X_1 = i_1, \dots, X_{n-1} = i_{n-1}, X_n = i_n) = \pi_0(j) p_{i_0 i_1} \dots p_{i_{n-1} i_n}$ (1.2)

Ορίζουμε επίσης την κατανομή π_n της κατάστασης της αλυσίδας ως την κατανομή για την οποία ισχύει $\pi_n(j) = P(X_n = j) \forall j \in S$.Τότε από το Θεώρημα Ολικής Πιθανότητας, έχουμε:

$\pi_{n+1}(j) = \sum_{i \in S} P(X_n = i) P(X_{n+1} = j | X_n = i) = \sum_{i \in S} \pi_n \cdot p_{ij}$ (1.3)

ή ισοδύναμα με χρήση πινάκων: π_{n+1}=$\pi_n \mathbb{P}$, $\forall$ n$\in \mathbb{N}_0$ (1.4)

και επομένως, $\pi = \pi_0 \mathbb{P}^n$, όπου $\mathbb{P}^n = \{p_{ij}(n)\}, i, j \in S$ ο πίνακας πιθανοτήτων μετάβασης ν-οστής τάξης (ή n βημάτων), δηλαδή $p_{ij}(n) = P(X_n = j | X_0 = i)$. Μάλιστα, αν επιλέξουμε $\pi_0 = \pi$ τέτοια ώστε $\pi = \pi \cdot \mathbb{P}$, προκύπτει $\pi_n = \pi$, $\forall n \in \mathbb{N}$, δηλαδή η κατανομή κατάστασης της αλυσίδας είναι στάσιμη (ανεξάρτητη του n). Συνεπώς, για να βρούμε την στάσιμη κατανομή π αρκεί να λύσουμε το σύστημα:

$$\begin{cases} \pi = \pi \cdot \mathbb{P} \\ \sum_{i \in S} \pi(i) = 1 \quad \forall\, i \in S \quad (1.5) \\ \pi(i) \geq 0 \end{cases}$$

Ένα θέμα που εμφανίζει έντονο ενδιαφέρον είναι κατά πόσον η αλυσίδα μπορεί να μετακινείται από την μία κατάσταση σε άλλη, με ένα ή περισσότερα βήματα ή και όχι. Θα λέμε ότι δύο καταστάσεις $i, j \in S$ επικοινωνούν μεταξύ τους αν υπάρχουν $m, n \geq 0$: $p_{ij}^{(n)} > 0$ και $p_{ji}^{(m)} > 0$ (αν ισχύει μόνο ένα από τα δύο, τότε λέμε ότι μόνο μία έχει πρόσβαση στην άλλη). Η επικοινωνία των καταστάσεων είναι σχέση ισοδυναμίας που διαμερίζει τον S σε κλάσεις, ανοιχτές ή κλειστές. Σε μία ανοιχτή κλάση C ,υπάρχουν $i \in C$, $j \notin C$ και $n \geq 0$ με $p_{ij}^{(n)} > 0$, δηλαδή οι καταστάσεις της C έχουν πρόσβαση σε καταστάσεις εκτός της C. Αν μία κλάση δεν είναι ανοιχτή, θα λέγεται κλειστή. Είναι προφανές ότι αν η αλυσίδα βρεθεί σε μία κλειστή κλάση

$C(X_n = i, \in C)$ μένει για πάντα σε αυτήν.

Ωστόσο, παρατηρείται ένα επιπλέον θέμα που αξίζει να αναλυθεί ως προς το πόσο χρόνο δαπανά η αλυσίδα σε κάθε κατάσταση , δηλαδή πόσες φορές επισκέπτεται κάθε κατάσταση. Για κάθε κατάσταση $i \in C$ ορίζουμε το συνολικό πλήθος των επισκέψεων της αλυσίδας στην i ως $V(i) = \sum_{k=0}^{\infty} \mathbb{1}\{X_n = i\}$ (τ.μ. με τιμές στον $n \in \mathbb{N}_0 \cup \{\infty\}$). Για τα $\omega \in \Omega$ για τα οποία $V(i) < \infty$, η αλυσίδα εγκαταλείπει τελικά την i και δεν επιστρέφει ποτέ σ' αυτήν. Φυσικά, αν $V(i) = \infty$, η αλυσίδα δεν σταματά ποτέ να επισκέπτεται την i. Συνεπώς, αν $P(V(i) < \infty | X_0 = i) = 0$, η κατάσταση i καλείται επαναληπτική, ενώ αν $P(V(i) < \infty | X_0 = i) = 1$, η κατάσταση i καλείται παροδική. Αποδεικνύεται για μια κατάσταση i ότι είτε θα είναι παροδική είτε θα είναι επαναληπτική. Μία αλυσίδα θα καλείται μη-υποβιβάσιμη αν ο χώρος καταστάσεων S είναι ολόκληρος μία κλειστή κλάση.

Ορίζουμε ως χρόνο πρώτης επιστροφής στην κατάσταση $x \in S$ την τ.μ $T_i = inf\{k > 0 : X_k = i\}$.

Μάλιστα, για μία επαναληπτική κατάσταση x αν ισχύει επιπλέον $E_i[T_i] < +\infty$, τότε κάνουμε λόγο για μία γνήσια ή θετικά επαναληπτική κατάσταση. Οι ιδιότητες της παροδικότητας και επαναληπτικότητας (γνήσιας και μη) είναι ιδιότητες κλάσης, που σημαίνει ότι αν μία κατάσταση x έχει μία από τις ιδιότητες αυτές, τότε όλες οι καταστάσεις της κλάσης στην οποία η x ανήκει, έχουν την ίδια ιδιότητα. Είναι ευνόητο ότι οι παροδικές κλάσεις είναι ανοιχτές ισχύει και το αντίστροφο, ενώ οι επαναληπτικές κλάσεις είναι κλειστές. Αυτό σημαίνει ότι η αλυσίδα θα αφήσει κάποια στιγμή τις παροδικές καταστάσεις με πιθανότητα 1 και μετά από πεπερασμένο αριθμό βημάτων θα βρεθεί σε κάποια επαναληπτική κατάσταση, οπότε και θα παραμείνει για πάντα στην κλάση των επαναληπτικών καταστάσεων. Επιπλέον, αξίζει να σημειωθεί ότι $\pi(i) = 0$ για κάθε κατάσταση i που δεν είναι γνήσια επαναληπτική.

Σε αρκετές εφαρμογές, αν η αλυσίδα ξεκινήσει σε μία κατάσταση i θα την επισκέπτεται ανά τακτά χρονικά διαστήματα. Τότε λέμε ότι η κατάσταση i είναι περιοδική και ότι η αλυσίδα παρουσιάζει την ιδιότητα της περιοδικότητας. Ορίζουμε ως περίοδο μίας κατάστασης $i \in S$, τον μέγιστο κοινό διαιρέτη $d(i)$ του συνόλου $\left\{n \in \mathbb{N} : p_{ii}^{(n)} > 0\right\}$ (σύνολο δυνατών χρόνων επιστροφής στην i). Αν $d(i) = 1$, η κατάσταση i καλείται απεριοδική. Έπεται ότι αν $p_{ii} > 0$ τότε η κατάσταση i είναι απεριοδική και αν η αλυσίδα είναι μη-υποβιβάσιμη, όλες οι καταστάσεις είναι απεριοδικές, οπότε γίνεται λόγος για απεριοδική αλυσίδα.

Υποθέτουμε ότι η μαρκοβιανή αλυσίδα $\{X_n\}_{n \in \mathbb{N}_0}$ είναι μη-υποβιβάσιμη, γνησίως επαναληπτική και απεριοδική. Τότε, ανεξαρτήτως της αρχικής κατανομής της ισχύει $\lim_{n \to \infty} \pi_n = \pi$ και $\lim_{n \to \infty} p_{ij}^{(n)} = \pi(j), \forall i \in S$, όπου η $\pi(\cdot)$ είναι η λύση του συστήματος (1.5). Επίσης, το ποσοστό σε βάθος χρόνου που ξοδεύει η αλυσίδα σε κάθε κατάσταση $i \in S$, είναι $\pi(i)$. Δηλαδή:

$$P\left[\frac{V_n(i)}{n} \to \pi(i), \forall i \in S\right] = 1 \quad (1.6)$$

Τέλος, αν η μετάβαση από μία κατάσταση σε άλλη κατάσταση συνεπάγεται μία αμοιβή ή κάποιο κόστος (αρνητική αμοιβή) της μορφής $R_n = R_n(X_{n-1}, X_n), n \in \mathbb{N}$ για το n-οστό βήμα, τότε:

- Η συνολική αμοιβή στα πρώτα n βήματα είναι $C(n) = \sum_{s=1}^{n} R_s(X_{s-1}, X_s), n \in \mathbb{N}$.
- Η μέση αμοιβή για μετάβαση $i \to j$είναι $r_{ij} = E[R_n(i,j)] < \infty\, , i,j \in S$.
- Η μέση αμοιβή για ένα βήμα που αντιστοιχεί σε αναχώρηση της $\{X_n\}$ από την κατάσταση i είναι $r_i = \sum_j r_{ij} \cdot p_{ij}\ ,\ i \in S$.
- Αποδεικνύεται ότι για τον ρυθμό αμοιβής (μέση αμοιβή σ' ένα βήμα) ισχύει
  - $\lim\limits_{n\to\infty} \frac{C(n)}{n} = \sum_j r_j \cdot \pi_j$ με πιθανότητα 1,
  - $\lim\limits_{n\to\infty} \frac{E[C(n)]}{n} = \sum_j r_j \cdot \pi_j$ (1.7)

Οι παραπάνω τύποι ισχύουν εφόσον η αλυσίδα έχει αναλλοίωτη κατανομή π. Επιπλέον έχει γίνει η υπόθεση ότι οι R_n είναι ανεξάρτητες τ.μ. η κατανομή κάθε μίας από τις οποίες εξαρτάται από την εκάστοτε μετάβαση από κατάσταση σε κατάσταση και όχι από το χρονικό βήμα n.

1.4 Μαρκοβιανή αλυσίδα σε συνεχή χρόνο

Η μετάβαση από μία κατάσταση της διαδικασίας σε μία άλλη μπορεί να πραγματοποιηθεί οποιαδήποτε χρονική στιγμή, $t \geq 0$. Η ανάγκη τέτοιων αναλύσεων μας δίνει τον ακόλουθο ορισμό:

Ορισμός I

Μία στοχαστική διαδικασία σε συνεχή χρόνο $\{X(t), t \geq 0\}$, με διακριτό χώρο καταστάσεων το πεπερασμένο ή απείρως αριθμήσιμο σύνολο S καλείται Μαρκοβιανή αλυσίδα σε συνεχή χρόνο αν ισχύει ότι:

$$P\{X(t_n) = i_n | X(t_0) = i_0, \cdots, X(t_{n-1}) = i_{n-1}\} = P\{X(t_n) = i_n | X(t_{n-1}) = i_{n-1}\}$$

για οποιαδήποτε $0 \leq t_0 \leq \cdots \leq t_{n-1} \leq t_n$ και $i_0, i_1, \ldots, i_{n-1}, i_n \in S$.

Κεφάλαιο 2ο Μαρκοβιανές διαδικασίες αποφάσεων

2.1 Εισαγωγή

Το 1957 ο Bellman εισήγαγε τις Μαρκοβιανές διαδικασίες αποφάσεων (ΜΔΑ) που ήταν αποτέλεσμα του συνδυασμού της θεωρίας των Μαρκοβιανών διαδικασιών με τον δυναμικό προγραμματισμό. Οι Μαρκοβιανές διαδικασίες αποφάσεων είναι μοντέλα λήψης διαδοχικών αποφάσεων όταν το αποτέλεσμα των οποίων είναι αβέβαιο. Σε κάθε χρονική περίοδο αποφάσεων ή χρονική στιγμή, το σύστημα μας παρέχει την απαραίτητη πληροφορία ώστε από την κατάσταση στην οποία βρισκόμαστε να διαλέγουμε την κατάλληλη ενέργεια μέσα από ένα σύνολο ενεργειών A_i με $i \in S$. Επιλέγοντας μία ενέργεια σε μία κατάσταση, καθορίζουμε την κατάσταση στην οποία θα μεταβούμε την επόμενη χρονική περίοδο μέσω μιας εξίσωσης μετάβασης. Οι πιθανότητες μετάβασης εξαρτώνται από την κατάσταση αλλά και από την επιλογή της ενέργειας. Για μελλοντικές καταστάσεις, η πολιτική που θα ακολουθήσουμε είναι αυτή που θα μας δώσει τις απαραίτητες πληροφορίες για την επιλογή των ενεργειών μας.

Ορισμός ΙΙ

Αρχικά πριν δώσουμε τον ορισμό της ¨πολιτικής¨ σε αυτό το κομμάτι οφείλουμε να διακρίνουμε τα διάφορα είδη κανόνων απόφασης ή τρόπων λήψης απόφασης. Όταν $X_n = i$, οι τρόποι να διαλέξουμε μια απόφαση είναι οι εξής:

1. Μαρκοβιανός αιτιοκρατικός (ντετερμινιστικός): στην προκείμενη περίπτωση ο τρόπος λήψης απόφασης είναι μια συνάρτηση $d_n(i)$ που εξαρτάται μόνο από το παρόν βήμα n και την παρούσα κατάσταση i και επιλέγει μια απόφαση με πιθανότητα 1, δηλαδή $a_n = d_n(i)$, όπου

$$d_n : S \to A,\ i \mapsto d_n(i) \in A(i).$$

2. Παρελθοντικά εξαρτώμενος αιτιοκρατικός τρόπος: στην περίπτωση αυτή ο τρόπος λήψης απόφασης εξαρτάται από τον παρελθόντα χρόνο, οπότε αποτελεί μια συνάρτηση της μορφής:

$$h_n = (s_1, a_1, \dots, s_{n-1}, a_{n-1}, s_n),$$

όπου $s_k = X_k$, $a_k \in A(s_k), k = 1,2,\dots,n$, η κατάσταση και η ληφθείσα απόφαση την χρονική στιγμή k. Όταν $X_n = s_n = i$ η απόφαση $a_n = d_n(\cdot)$ λαμβάνεται από το σύνολο δυνατών αποφάσεων $A(s_n)$. Ως εκ τούτου έχουμε:

$$d_n : S \times A \times S \times \dots . \times A \times S \to A, h_n \mapsto d_n(h_n) \in A(s_n).$$

3. Τυχαιοποιημένος κανόνας λήψης αποφάσεων, ο οποίος διακρίνεται σε δυο υποκατηγορίες:

i. στον μαρκοβιανό τυχαιοποιημένο κανόνα, κατά τον οποίο όταν $X_n = i$ η λήψη της απόφασης $a \in A(i)$ γίνεται σύμφωνα με μία κατανομή πιθανοτήτων $q_n(i, a), a \in A(i)$.

ii. και στον παρελθοντικά εξαρτώμενο τυχαιοποιημένο κανόνα, κατά τον οποίο όταν $X_n = i$ η λήψη της απόφασης $\alpha \in A(i)$ εξαρτάται και από την παρελθούσα εξέλιξη h_n της διαδικασίας και των αντίστοιχων αποφάσεων ακολουθώντας την κατανομή $q_n(h_n, a), a \in A(i), i = s_n$

Ορισμός πολιτικής:

Μία ακολουθία τρόπων λήψης αποφάσεων για κάθε χρονική στιγμή n ονομάζεται **πολιτική**. Ουσιαστικά, πρόκειται για μία στρατηγική που μας εφοδιάζει με τον κανόνα επιλογής απόφασης για κάθε χρονική στιγμή και κάτω από οποιαδήποτε προηγούμενη εξέλιξη ή απόφαση της αλυσίδας. Σε περίπτωση που η πολιτική δεν εξαρτάται από το χρονικό βήμα, καλείται **στάσιμη πολιτική**.

Το ενδιαφέρον μας επικεντρώνεται στην κατηγορία των στάσιμων μαρκοβιανών αιτιοκρατικών πολιτικών δηλαδή στις ακολουθίες ενεργειών:

$$R = (R_i)\ i\ \in S \quad \text{με } R_i : S \to A, i \mapsto R_i \in A(i), i \in S$$

Έστω ότι έχουμε αποφασίσει να ακολουθήσουμε μια συγκεκριμένη τυχαία πολιτική R. Τη χρονική στιγμή t δε μπορούμε να γνωρίζουμε ακριβώς την κατάσταση στην οποία βρίσκεται το σύστημα, καθώς η κατάσταση του συστήματος είναι μια τυχαία μεταβλητή, την οποία έχουμε ήδη ορίσει ως X_t. Αυτόματα καταλήγουμε στο συμπέρασμα ότι τυχαία μεταβλητή θα είναι και η ενέργεια που θα επιλέξουμε τη χρονική στιγμή t . Έστω A_t η τυχαία μεταβλητή που δίνει την ενέργεια που επιλέγουμε τότε. Μας ενδιαφέρει η από κοινού κατανομή των $(X_t\, A_t)$ υπό την συγκεκριμένη πολιτική R, δηλαδή η $P_R\,(X_t = j, A_t = a|X_0 = i) = R(a|i)$. Έχοντας την συγκεκριμένη δεσμευμένη κοινή κατανομή θα επιχειρήσουμε να υπολογίσουμε το κόστος $C(X_t\, A_t)$. Το κόστος όμως είναι και αυτό μια τυχαία μεταβλητή, οπότε αυτή που μπορούμε να υπολογίσουμε είναι η αναμενόμενη τιμή του. Αυτή θα ισούται με:

$$E_R[C(X_t, A_t)|X_0 = i] = \sum_j \sum_{a \in A_j} C(j,a)\, P_R(X_t = j, A_t = a|X_0 = i)$$

και είναι το **αναμενόμενο μέσο κόστος** τη χρονική στιγμή t. Οι οικονομικές θεωρίες μας έχουν δείξει ότι ένα μελλοντικό κόστος είναι μικρότερο αν εκφραστεί σε σημερινά χρήματα. Εισάγουμε λοιπόν την έννοια του αποπληθωριστικού παράγοντα α. Ο αποπληθωριστικός παράγοντας είναι ένας αριθμός για τον οποίο ισχύει ότι είναι $0 < a \leq 1$ με τον οποίο πολλαπλασιάζονται τα κόστη κάθε χρονική στιγμή που περνάει. Στην περίπτωση $\alpha = 1$ προφανές ότι έχουμε μια εκφυλισμένη περίπτωση όπου δεν υπάρχει αποπληθωρισμός. Η τιμή 0 δεν είναι εφικτή γιατί αυτό θα μηδένιζε κάθε κόστος.

Λόγω της Μαρκοβιανής διαδικασίας απόφασης, υπάρχει βέλτιστη πολιτική η οποία εξασφαλίζει την ελαχιστοποίηση του συνολικού μέσου κόστους. Συνεπώς, στόχος των Μαρκοβιανών διαδικασιών αποφάσεων είναι η εύρεση μιας πολιτικής που ελέγχει τη διαδικασία με το βέλτιστο τρόπο. Η συνάρτηση του κόστους ορίζει το κριτήριο βελτιστοποίησης του προβλήματος. Τα κριτήρια βελτιστοποίησης τα οποία χρησιμοποιούνται πιο συχνά είναι η ελαχιστοποίηση του συνολικού αναμενόμενου αποπληθωρισμένου κόστους είτε σε πεπερασμένο χρονικό ορίζοντα, είτε σε άπειρο χρονικό ορίζοντα και η ελαχιστοποίηση του μακροπρόθεσμου αναμενόμενου μέσου κόστους ανά μονάδα χρόνου. Επίσης θα μας απασχολήσει το κριτήριο της ελαχιστοποίησης του συνολικού αναμενόμενου κόστους.

Έστω ότι είναι δεδομένη η αρχική κατάσταση i, ο χρονικός ορίζοντας n και μια πολιτική R. Τότε το αναμενόμενο συνολικό αποπληθωρισμένο κόστος χρονικού ορίζοντα μεγέθους n υπό την πολιτική R θα είναι $V_{R,\alpha,n}(i)$, όπου α ο αποπληθωριστικός παράγοντας. Επιτρέποντας την οριακή περίπτωση $n = 0$ έχουμε την διαδικασία να τερματίζεται στην κατάσταση i πριν ακόμα ξεκινήσει, οπότε $V(i) = F(i)$, ενώ για κάθε άλλη χρονική στιγμή $n \geq 1$ ορίζουμε:

$$V_{R,\alpha,n}(i) = \sum_{t=0}^{n-1} a^t\, E_R[C(X_t, A_t)|X_0 = i] + a^n E_R[F(X_n)|X_0 = i].$$

Η βέλτιστη συνάρτηση αναμενόμενου μέσου αποπληθωρισμένου κόστους για το χρονικό ορίζοντα μήκους n ορίζεται να είναι $V_{a,n}(i) = inf_R V_{R,a,n}(i)$, όπου το infimum διατρέχει όλες τις πιθανές πολιτικές για τον ορίζοντα μήκους n. Η ποσότητα $V_{a,n}(i)$ είναι το μεγαλύτερο κάτω φράγμα για όλα τα κόστη χρονικού ορίζοντα μήκους n, οπότε είναι το καλύτερο αποτέλεσμα. Έστω λοιπόν μια πολιτική R για το χρονικό ορίζοντα μήκους n. Τότε η πολιτική R είναι α-βέλτιστη για το κριτήριο αναμενόμενου μέσου αποπληθωρισμένου κόστους χρονικού ορίζοντα n αν και μόνο αν:

$$V_{R,a,n}(i) = V_{a,n}(i) \quad \forall\, i \in \mathrm{S}.$$

Αν ορίσουμε τον αποπληθωριστικό παράγοντα $a = 1$ τότε έχουμε μια εκφυλισμένη περίπτωση του προηγούμενου κριτηρίου με το αναμενόμενο συνολικό κόστος χρονικού ορίζοντα μήκους n να ορίζεται σαν $V_{R,0}(i) = F(i)$ αν $n = 0$ και για $n \geq 1$:

$$V_{R,n}(i) = \textstyle\sum_{t=0}^{n-1} E_R[C(X_t, A_t)|X_0 = i] + E_R[F(X_n)|X_0 = i].$$

Η αντίστοιχη βέλτιστη συνάρτηση αναμενόμενο μέσου κόστους χρονικού ορίζοντα n ορίζεται να είναι $V_n(i) = inf_R V_{R,n}(i)$, όπου το infimum διατρέχει όλες τις δυνατές πολιτικές. Μια πολιτική R λέγεται βέλτιστη για το κριτήριο αναμενόμενου μέσου κόστους χρονικού ορίζοντα μήκους n αν και μόνο αν:

$$V_{R,n}(i) = V(i), \forall i \in S.$$

Για να χρησιμοποιήσουμε το κριτήριο αναμενόμενου μέσου αποπληθωρισμένου κόστους σε άπειρο χρονικό ορίζοντα θα χρειαστεί να ελαχιστοποιήσουμε το αντίστοιχο συνολικό αναμενόμενο αποπληθωρισμένο κόστος υπό μια πολιτική R. Προς αποφυγήν της περιπτώσεως που όλα τα κόστη είναι ∞ ορίζουμε τον αποπληθωριστικό παράγοντα να είναι τέτοιος ώστε $a < 1$ αποκλείοντας την τιμή $a = 1$. Το αναμενόμενο συνολικό αποπληθωρισμένο κόστος θα προκύψει αν στο κόστος πεπερασμένου χρονικού ορίζοντα αφήσουμε n→ ∞ και θα είναι ίσο με:

$$V_{R,\alpha}(i) = E_R\left[\sum_{t=0}^{\infty} a^t C(X_t, A_t)|X_0 = i\right] = \sum_{t=0}^{\infty} a^t E_R[C(X_t, A_t)|X_0 = i].$$

Η βέλτιστη συνάρτηση αναμενόμενου μέσου αποπληθωρισμένου κόστους ορίζεται να είναι $V_a(i) = inf_R V_{R,a}(i)$, ενώ το infimum διατρέχει όλες τις δυνατές πολιτικές για τον άπειρο χρονικό ορίζοντα. Ως μέγιστο κάτω φράγμα για τα αναμενόμενα αποπληθωρισμένα κόστη είναι το καλύτερο δυνατό αποτέλεσμα. Μια πολιτική R λέγεται α-βέλτιστη για το κριτήριο αναμενόμενου μέσου αποπληθωρισμένου κόστους άπειρου χρονικού ορίζοντα αν και μόνο αν ισχύει:

$$V_{R,\alpha}(i) = V_a(i), \forall\, i \in S.$$

Το μακροπρόθεσμο αναμενόμενο μέσο κόστος ανά μονάδα χρόνου άπειρου χρονικού ορίζοντα υπό μια πολιτική R ορίζεται να είναι:

$$g_R(i) = \lim_{n\to\infty} sup \tfrac{1}{n} E_R[\textstyle\sum_{t=0}^{n-1} C(X_t, A_t)|X_0 = i] = \lim_{n\to\infty} \sup \tfrac{V_{R,n}(i)}{n}. \quad (2.1)$$

Η βέλτιστη συνάρτηση αναμενόμενου μακροπρόθεσμου μέσου κόστους θα ισούται με: g(i)= $inf_R g_R(i)$,όπου το infimum λαμβάνεται για όλες τις δυνατές πολιτικές για τον άπειρο χρονικό ορίζοντα. Μια πολιτική R ονομάζεται βέλτιστη για το κριτήριο αναμενόμενου μακροπρόθεσμου μέσου κόστους αν και μόνο αν ισχύει $R(i) = g(i) \forall i \in S$. Μια ακόμα καλύτερη προσέγγιση του ζητούμενου κόστους επιτυγχάνεται αν αντί για supremum πάρουμε το infimum της ίδιας συνάρτησης με το μακροπρόθεσμο αναμενόμενο μέσο κόστος ανά μονάδα χρόνου να ισούται στην προκείμενη περίπτωση με:

$$g^*{}_R(i) = \lim_{n\to\infty} inf \frac{1}{n} E_R\left[\sum_{t=0}^{n-1} C(X_t\,, A_t)|X_0 = i\right] = \lim_{n\to\infty} inf \frac{V_{R,n}(i)}{n}.$$

Η αντίστοιχη βέλτιστη συνάρτηση αναμενόμενου μακροπρόθεσμου μέσου κόστους να είναι ίση με $g^*{}_R(\iota) = inf_R g^*{}_R(i)$ και μια πολιτική R λέγεται βέλτιστη για το κριτήριο αναμενόμενου μακροπρόθεσμου μέσου κόστους αν και μόνο αν ισχύει $g^*{}_R(i) = g^*(i), \forall i \in S$.

2.2 Μαρκοβιανές διαδικασίες αποφάσεων σε πεπερασμένο χρονικό ορίζοντα

Ας υποθέσουμε ένα δυναμικό σύστημα το όποιο επανεξετάζεται σε χρονικές στιγμές $t = 0,1,2, \ldots$ οι οποίες απέχουν εξίσου μεταξύ τους. Σε κάθε επανεξέταση το σύστημα ταξινομείται σε ένα πιθανό αριθμό καταστάσεων και στην συνέχεια λαμβάνεται μια απόφαση που πρέπει να γίνει. Το σύνολο των πιθανών καταστάσεων συμβολίζεται με S. Για κάθε κατάσταση $i \in S$, δίνεται ένα σύνολο $A(i)$ αποφάσεων ή ενεργειών. Ο χώρος κατάστασης S και τα σύνολα δράσης $A(i)$ θεωρούνται ότι είναι πεπερασμένα. Οι οικονομικές συνέπειες των αποφάσεων που ελήφθησαν στους χρόνους επανεξέτασης αντικατοπτρίζονται στο κόστος. Αυτό το ελεγχόμενο δυναμικό σύστημα ονομάζεται Μαρκοβιανή διαδικασία αποφάσεων διακριτού χρόνου όταν η ακόλουθη Μαρκοβιανή ιδιότητα ικανοποιείται. Εάν σε μια χρονική στιγμή έχει επιλεγεί η ενέργεια α στην κατάσταση i τότε ανεξαρτήτως της προηγούμενης χρονικής στιγμής του συστήματος συμβαίνουν τα εξής:

- Προκύπτει άμεσο κόστος $C(i, a)$
- Την επόμενη χρονική στιγμή n το σύστημα θα βρίσκεται στην κατάσταση j με πιθανότητα $P_{ij}\ (a)$ η οποία εξαρτάται μόνον από την ενέργεια a και τις καταστάσεις i και j που:

$$\textstyle\sum_{j\in S} P_{ij}\ (a) = 1 \qquad \text{με i} \in S$$

Το κόστος ενός βήματος $C_i(a)$ και οι πιθανότητες μετάβασης ενός βήματος $P_{ij}\ (a)$ υποτίθεται ότι είναι χρονικά ομοιογενείς. Σε συγκεκριμένα προβλήματα το άμεσο κόστος $C_i(a)$ αντιπροσωπεύει το αναμενόμενο κόστος που προκύπτει μέχρι τον επόμενο χρόνο επανεξέτασης όταν επιλέγεται η ενέργεια a στην κατάσταση i. Επίσης θα πρέπει να τονιστεί ότι η επιλογή του χώρου καταστάσεων και του συνόλου ενεργειών συχνά εξαρτάται από την δομή του κόστους των συγκεκριμένων προβλημάτων που εξετάζεται. Για παράδειγμα σε ένα πρόβλημα παραγωγής-αποθεμάτων που συνεπάγεται σταθερό κόστος εγκατάστασης για την επανεκκίνηση της παραγωγής μετά από μια περίοδο αδράνειας η περιγραφή της κατάστασης θα πρέπει να περιλαμβάνει μια μεταβλητή κατάστασης που δείχνει αν η παραγωγή είναι ενεργοποιημένη η όχι. Πολλά πρακτικά προβλήματα ελέγχου μπορούν να μοντελοποιηθούν ως ένα Μαρκοβιανό μοντέλο διαδικασίας λήψης αποφάσεων με την κατάλληλη επιλογή του χώρου καταστάσεων και του συνόλου ενεργειών.

Ορίζουμε μια βοηθητική συνάρτηση που θα μας βοηθήσει να υπολογίσουμε το ελάχιστο κόστος που θέλουμε, που εξαρτάται από την κατάσταση i και την ενέργεια α ως εξής:

$$V_{a,n}(i,a) = C(i,a) + a\sum_{j} P_{ij}\ (a)V_{a,n-1}(j)\ , n \geq 1$$

Ορίζουμε επίσης το σύνολο που περιέχει τις ενέργειες που πραγματοποιούν το minimum ως εξής:

$$B_i(a\ , n\ ) = \{b \in A_i\ |V_{a,n}\ (i,b) = min_{a\in A_i}\{V_{a,n}(i,\alpha)\}\ \}$$

Στις περισσότερες των περιπτώσεων αυτό θα είναι μονοσύνολο, μα δεν είναι απαραίτητο να περιέχει μόνο μια ενέργεια κάθε φορά με συνέπεια η βέλτιστη πολιτική να μην είναι πάντοτε μοναδική. Είχαμε ορίσει το κόστος τερματισμού της διαδικασίας $V_{R\ ,a\ ,0}\ (i) = F\ (i)$. Άρα θα ισχύει $V_{a,0}\ = F$ ανεξάρτητα από την κατάσταση που βρισκόμαστε, καθώς στον ορίζοντα μήκους 0 δεν μπορούμε να επιλέξουμε καμία ενέργεια. Υποθέτουμε τώρα ότι $n\ \geq 1$ δηλαδή ότι η διαδικασία θα εξελιχθεί για τουλάχιστον ένα βήμα ακόμα κατά το όποιο θα επιλεγεί μια ενέργεια. Έστω ότι το μήκος του ορίζοντα είναι n και η αρχική κατάσταση i είναι δοσμένα .Κάποια ενέργεια $a \in A_i$ επιλέγεται επιφέροντας ένα κόστος $C(i,a)$ και η διαδικασία μεταβαίνει στην κατάσταση j με κάποια πιθανότητα $P_{ij}\ (a)$. Για να επιτύχουμε το καλύτερο δυνατό αποτέλεσμα θα πρέπει να λειτουργήσουμε βέλτιστα για το πρόβλημα ορίζοντα μήκους $n - 1$ με αρχική κατάσταση την j. Άρα το ζητούμενο κόστος θα δίνεται από την σχέση:

$$C(i,a) + a\sum_{j} P_{ij}\ (a)V_{a\ ,n-1}\ (j)$$

Η αρχική ενέργεια θα πρέπει να επιλέγεται με κριτήριο την ελαχιστοποίηση των ποσοτήτων που εμφανίζονται άρα $V_{a,n}(i) = min_{a\in A_i}\{u_{a,n}(i,a)\}$, η βέλτιστη ενέργεια δηλαδή θα πρέπει να περιορίζεται στο σύνολο $B_i\ (a,n)$. Εδώ βλέπουμε να εφαρμόζεται η αρχή βελτιστοποίησης του Bellman, η οποία λέει ότι για να μπορεί μια πολιτική να είναι βέλτιστη αν έχουν ήδη μεσολαβήσει οι ενέργειες για τις χρονικές στιγμές $0,1,2, \dots t - 1$, τότε οι υπόλοιπες ενέργειες πρέπει να αποτελούν μια βέλτιστη πολιτική για τον ορίζοντα μήκους $n - t$.

Τώρα λοιπόν μπορούμε να ορίσουμε το βασικό θεώρημα για την βέλτιστη συνάρτηση αναμενόμενου μέσου κόστους σε πεπερασμένο χρονικό ορίζοντα. Αυτή η συνάρτηση ικανοποιεί την ακόλουθη εξίσωση βελτιστοποίησης πεπερασμένου χρονικού ορίζοντα.

$$V_{a,n}\ (i) = \min_{a}\ \{C(\ i,a\ ) + a\ \textstyle\sum_j P_{ij}\ (a)\ \ V_{a\ ,n-1}(j)\}\quad \mathrm{i}\in S\quad n\ \geq 1$$

Μια πολιτική R είναι βέλτιστη για τον ορίζοντα μήκους 1 αν και μόνο αν με δεδομένη την αρχική κατάσταση i, η κατανομή $\theta(a|i)$ περιορίζεται στο σύνολο $B_i(a,1)$. Μια πολιτική R είναι βέλτιστη για τον ορίζοντα μήκους n≥ 2 αν και μόνο αν με δεδομένη την αρχική κατάσταση i, η κατανομή $\theta(a|i)$ περιορίζεται στο σύνολο $B_i(a,n)$ και με δεδομένη την κατάσταση j στην οποία μεταβαίνει η διαδικασία τη χρονική στιγμή $t - 1$ η πολιτική R θα παραμένει βέλτιστη για τον ορίζοντα μήκους $n - 1$ με αρχική κατάσταση την j.

Αν έχουμε την τιμή $a = 1$ για τον αποπληθωριστικό παράγοντα τότε ισχύουν ακριβώς τα ίδια αποτελέσματα στον υπολογισμό της βέλτιστης πολιτικής και της συνάρτησης κόστους με εξίσωση την ακόλουθη:

$$V_n(i) = \ min_a\{C(i,a) + \ \textstyle\sum_j P_{ij}\ (a)V_{n-1}\ (j)\}\quad i \in S\quad n \geq 1\ (2.2)$$

Αν αντί για κόστος $C(i, a)$ στο πρόβλημα μας έχουμε κέρδος $R(i, a)$ τότε η εξίσωση παίρνει τη μορφή:

$$V_{a,n}(i) = max_a \left\{ R(i, a) + \alpha \sum_j P_{ij}(a) V_{a,n-1}(j) \right\} \quad i \in S \; n \geq 1$$

Η εξίσωση που καταλήξαμε συχνά εμφανίζεται και ως εξίσωση δυναμικού προγραμματισμού. Απαραίτητη για την λύση είναι και η ύπαρξη κάποιων οριακών συνθηκών για την τιμή της συνάρτησης κόστους στο τέλος της διαδικασίας. Η βέλτιστη πολιτική υπολογίζεται με αναδρομική επίλυση της εξίσωσης δυναμικού προγραμματισμού και συνήθως υλοποιείται με την χρήση κάποιας γλώσσας προγραμματισμού. Στα επόμενα τρία κεφάλαια θα μελετήσουμε τρεις μεθόδους εύρεσης της βέλτιστης πολιτικής ως προς την ελαχιστοποίηση του μέσου ρυθμού κόστους. Αυτές είναι η μέθοδος βελτίωσης της πολιτικής, η μέθοδος του γραμμικού προγραμματισμού και η μέθοδος των διαδοχικών προσεγγίσεων. Τις μεθόδους συνοδεύουν οι αντίστοιχοι αλγόριθμοι και στο τέλος του εκάστοτε κεφαλαίου θα αναφερθούμε και στην σύγκλιση αντίστοιχα κάθε μεθόδου.

2.3 Η μέθοδος βελτίωσης της πολιτικής

Με την μέθοδο βελτίωσης της πολιτικής αρχίζοντας από μία οποιαδήποτε πολιτική R,ελέγχουμε αν αυτή είναι βέλτιστη ή όχι. Σε περίπτωση που δεν είναι, βρίσκουμε μία πολιτική R΄ με g(R΄)≤g(R) και ελέγχουμε αν είναι βέλτιστη. Γνωρίζοντας ότι το πλήθος των στάσιμων πολιτικών είναι πεπερασμένο (αφού τόσο ο χώρος καταστάσεων, όσο και ο χώρος αποφάσεων είναι πεπερασμένοι), μπορούμε να συνεχίσουμε την παραπάνω διαδικασία μέχρις ότου καταλήξουμε στην βέλτιστη πολιτική. Το πλεονέκτημα της μεθόδου αυτής είναι ότι αντί να ελέγξουμε όλες τις δυνατές στάσιμες πολιτικές, ελέγχουμε ένα συνήθως μικρό υποσύνολό τους, μεταβαίνοντας από μία πολιτική σε μία καλύτερη.

Η αρχή γίνεται από την σχέση (2.1): $g(R) = \lim_{n\to\infty} \frac{V_n(i,R)}{n}, \forall i \in S$ η οποία συνεπάγεται ότι ασυμπτωτικά (δηλαδή για n→∞: μεγάλα n) ισχύει:

$$V_n(i, R) \cong ng(R) + u_i(R), \forall i \in S \quad (2.3)$$

Οι ποσότητες $u_i(R)$ ονομάζονται σχετικές τιμές των καταστάσεων i όταν εφαρμόζεται μία στάσιμη πολιτική R και η διαφορά τους ισούται με:

$$u_i(R) - u_j(R) \cong V_n(i, R) - V_n(j, R), \forall i, j \in S$$

Η έννοια των σχετικών τιμών είναι θεμελιώδης για την μέθοδο διαδοχικών προσεγγίσεων. Αυτές εκφράζουν την παροδική επίδραση των αρχικών καταστάσεων στο αναμενόμενο συνολικό κόστος κάτω από την εφαρμογή της πολιτικής R. Τότε η ποσότητα $u_i(R) - u_j(R)$ εκφράζει την διαφορά στο μέσο συνολικό κόστος, αν η διαδικασία ξεκινήσει από την κατάσταση i σε σύγκριση με το αν ξεκινούσε από την κατάσταση j, όταν εφαρμόζεται η πολιτική R. Ισοδύναμα, η διαφορά αυτή ουσιαστικά αποτελεί το μέγιστο ποσό που συμφέρει κάποιον να διαθέσει, ώστε το σύστημα (η αλυσίδα) να ξεκινήσει από την κατάσταση j παρά από την i κάτω από την πολιτική R.

Μάλιστα, αν υποθέσουμε απεριοδική αλυσίδα $X_n(R)$, τότε τα όρια υπάρχουν $\lim_{n\to\infty} p_{ij}^{(k)}(R)$=$\pi(j)$ κάτι που θα ισχύει και για τα όρια:

$$\lim_{n\to\infty} V_n(i, R) = \lim_{n\to\infty} \sum_{k=0}^{n-1} \sum_{j\in S} p_{ij}^{(k)}(R) \cdot C\big(j, R_j\big),$$

οπότε και $u_i(R) - u_j(R) = \lim_{n\to\infty}[V_n(i,R) - V_n(j,R)]$ που εκφράζει την σε βάθος χρόνου διαφορά στο μέσο συνολικό κόστος, αν η διαδικασία ξεκινήσει από την κατάσταση i παρά από την κατάσταση i, κάτω από την πολιτική R.

Τόσο οι σχετικές τιμές $u_i(R)$, όσο και ο μέσος ρυθμός κόστους g(R) μπορούν να υπολογιστούν ταυτόχρονα από το Θεώρημα Ι, του οποίου θα περιγράψουμε την βασική ιδέα, προτού προχωρήσουμε στην διατύπωση και απόδειξή του.

Ας δούμε το μέσο συνολικό κόστος $V_n(i,R)$ στα n πρώτα βήματα από την αρχή. Η αλυσίδα ξεκινάει από την κατάσταση i, ενώ εφαρμόζεται η πολιτική R. Τότε στην i επιλέγεται η απόφαση $a = R_i$, προκύπτει ένα κόστος $C(i,a) = C(i,R_i)$ και η αλυσίδα μεταφέρεται στην κατάσταση j με πιθανότητα $p_{ij}(R_i)$. Επιπλέον, όταν η αλυσίδα βρεθεί στην κατάσταση j το μέσο συνολικό κόστος στα επόμενα $n-1$ βήματα είναι $V_{n-1}(j,R)$. Συνεπώς, $\forall n \geq 1, i \in S$ ισχύει:

$$V_n(i,R) = C(i,R_i) + \sum_{j\in S} p_{ij}(R_i) \cdot V_{n-1}(j,R) \quad (2.4)$$

Συνδυάζοντας τις (2.2) και (2.3) δεν είναι δύσκολο να δειχθεί ότι ασυμπτωτικά έχουμε $\forall i \in S$:

$u_i(R) + g(R) \cong C(i,R_i) + \sum_{j\in S} p_{ij}(R_i) \cdot u_j(R)$.

ή (2.5)

$u_i(R) \cong C(i,R_i) - g(R) + \sum_{j\in S} p_{ij}(R_i) \cdot u_j(R)$.

που αποτελεί το κυρίως θέμα προς απόδειξη του Θεωρήματος Ι. Στο σημείο αυτό θα ορίσουμε τις σχετικές τιμές. Αρχικά ας παραθέσουμε ορισμένους απαραίτητους ορισμούς.

$T_i(R)$: ο αναμενόμενος χρόνος μέχρι την πρώτη επίσκεψη στην r, δοθέντος ότι $X_0(R) = i$ (η αλυσίδα ξεκίνησε από την $i \in S$, κάτω από την πολιτική R).

$K_i(R)$: το αναμενόμενο κόστος μέχρι την πρώτη επίσκεψη στην r ,όταν $X_0(R) = i$.

Τότε για μία γνήσια επαναληπτική κατάσταση $r \in S$ έχουμε από την θεωρία των Ανανεωτικών διαδικασιών ότι: $g(R) = \frac{K_r(R)}{T_r(R)}$. (2.6)

Με βάση αυτό και την σχέση (2.3) ορίζουμε $\forall i \in S$ ως σχετικές τιμές που αντιστοιχούν στην πολιτική R, τις ποσότητες:

$$u_i(R) = K_i(R) - g(R)T_i(R), \quad (2.7)$$

που για $i = r$:γνήσια επαναληπτική κατάσταση, η (2.7) μέσω της (2.6) δίνει $u_r(R) = 0$.

Διατυπώνουμε και αποδεικνύουμε το Θεώρημα Ι.

Θεώρημα Ι (εύρεσης ρυθμού κόστους και σχετικών τιμών)

Δίνεται στάσιμη πολιτική R τέτοια ώστε η Μαρκοβιανή αλυσίδα $\{X_n(R)\}_{n\in\mathbb{N}_0}$ να έχει μοναδική κλειστή κλάση επικοινωνίας και έστω το γραμμικό σύστημα

$$\Sigma_1: u_i = C(i,R_i) - g + \sum_{j\in S} p_{ij}(R_i) \cdot u_j \text{, με } i \in S \quad (2.8)$$

ως προς $(g, u_i), i \in S$.

Ισχύουν τα ακόλουθα:

Α) Τα $(g, u_i) = \big(g(R), u_i(R)\big), \forall i \in S$ αποτελούν λύση του Σ_1 (δηλαδή το Σ_1 έχει μία τουλάχιστον λύση).

Β) Αν$(g, u_i), i \in S$ τυχαία λύση του Σ_1, τότε ισχύει:

$g(R) = g$ και , $u_i(R) = u_i + c, i \in S$ όπου c πραγματική σταθερά.

Γ) Το σύστημα

$$\Sigma_2 : \begin{cases} u_i = C(i, R_i) - g + \sum_{j \in S} p_{ij}(R_i) \cdot u_j \ , i \in S \\ u_s = 0 \end{cases}$$

έχει μοναδική λύση, όπου s: αυθαίρετη κατάσταση της Μαρκοβιανής αλυσίδας $\{X_n\}$.

Απόδειξη:

Α) Υπενθυμίζουμε ότι για την γνήσια επαναληπτική κατάσταση $r \in S$, ισχύει:

$$T_i(R) = 1 + \sum_{j \neq r} p_{ij}(R_i) \cdot T_j(R) \quad (2.9)$$

Όπου $T_i(R)$: ο αναμενόμενος χρόνος πρώτης επίσκεψης στην r, δεδομένου ότι η αλυσίδα ξεκίνησε από την κατάσταση i κάτω από την πολιτική R [$X_0(R)=i$], ενώ για το μέσο κόστος $K_i(R)$ της πρώτης επίσκεψης στην r, ισχύει αντίστοιχα:

$$K_i(R) = C(i, R_i) + \sum_{j \neq r} p_{ij}(R_i) \cdot K_j(R) \quad (2.10)$$

Συνδυάζοντας τις (2.9) και (2.10) προκύπτει:

$$K_i(R) - g(R) T_i(R) = C(i, R_i) - g(R) + \sum_{j \neq r} p_{ij}(R_i) \cdot \big[K_j(R) - g(R) \cdot T_j(R)\big]$$

$$u_i(R) = C(i, R_i) - g(R) + \sum_{j \neq r} p_{ij}(R_i) \cdot \big[K_j(R) - g(R) \cdot T_j(R)\big] + p_{ir}(R_i) \cdot u_r$$

$$u_i(R) = C(i, R_i) - g(R) + \sum_{j \neq r} p_{ij}(R_i) \cdot \big[K_j(R) - g(R) \cdot T_j(R)\big] + p_{ir}(R_i) \cdot [K_r(R) - g(R) \cdot T_r(R)]$$

αφού u_r=0, οπότε ισχύει τελικά:

$$u_i(R) = C(i, R_i) - g(R) + \sum_{j \in S} p_{ij}(R_i) \cdot \big[K_j(R) - g(R) \cdot T_j(R)\big]$$

ή αλλιώς $u_i(R) = C(i, R_i) - g(R) + \sum_{j \in S} p_{ij}(R_i), i \in S$ (2.11),

δηλαδή τα $(g, u_i) = \big(g(R), u_i(R)\big)$ αποτελούν λύση του Σ1 και το ζητούμενο αποδείχθηκε.

Β) Έστω (g, u_i), i ∈S οποιαδήποτε λύση του Σ1. Θα αποδείξουμε με επαγωγή ότι ισχύει:

$$u_i = \sum_{k=0}^{m-1} \sum_{j \in S} p_{ij}^{(k)}(R) C(j, R_j) - mg + \sum_{j \in S} p_{ij}^{(m)}(R) u_j \quad (2.12)$$

Αρχικά ισχύει για m=1:

$$u_i = \sum_{j \in S} p_{ij}^{(0)}(R) C\big(j, R_j\big) - g + \sum_{j \in S} p_{ij}(R) u_j$$

($P^{(0)} = \{p_{ij}{}^{(0)}\} = I$: μοναδιαίος πίνακας)

$u_i = 1 \cdot C(i, R_i) - g + \sum_{j \in S} p_{ij}(R_i) u_j$ που ισχύει επειδή τα (g, u_i) , i ∈S αποτελούν λύση του Σ_1.

Επαγωγικό βήμα:

Υποθέτουμε ότι η (2.12) ισχύει για κάποιο $m \in \mathbb{N}$. Τότε για $m+1$ έχουμε:

$$u_i = \sum_{k=0}^{m} \sum_{j \in S} p_{ij}^{(k)}(R) C\big(j, R_j\big) - (m+1) g + \sum_{j \in S} p_{ij}^{(m+1)}(R)\, u_j$$

$$u_i = \sum_{k=0}^{m-1} \sum_{j \in S} p_{ij}^{(k)}(R) C\big(j, R_j\big) + \sum_{j \in S} p_{ij}^{(m)}(R) C(j\,, R_j) - mg - g + \sum_{j \in S} \sum_{z \in S} p_{iz}^{(m)}(R) p_{zj}(R) u_j.$$

Προσθαφαιρούμε τον όρο $\sum_{j \in S} p_{ij}^{(m)}(R) u_j$ και προκύπτει:

$$u_i = \sum_{k=0}^{m-1} \sum_{j \in S} p_{ij}^{(k)}\,(R) C\big(j, R_j\big) - mg + \sum_{j \in S} p_{ij}^{(m)}(R) u_j$$

$$- \sum_{j \in S} p_{ij}^{(m)}(R) u_j + \sum_{j \in S} p_{ij}^{(m)}(R) C\big(j, R_j\big) - g + \sum_{j \in S} \sum_{z \in S} p_{iz}^{(m)}(R) p_{zj}(R) u_j.$$

Και λόγω της υπόθεσης του επαγωγικού βήματος έπεται ότι:

$$u_i = u_i - \sum_{j \in S} p_{ij}^{(m)}(R) u_j + \sum_{j \in S} p_{ij}^{(m)}(R) C(j, R_j) - g + \sum_{j \in S} \sum_{z \in S} p_{iz}^{(m)}(R) p_{zj}(R) u_j$$

και αντικαθιστώντας την (2.8) έχουμε:

$$0 = -\sum_{j \in S} p_{ij}^{(m)}(R) (C\big(j, R_j\big) - g + \sum_{z \in S} p_{jz}(R_j) u_z) + \sum_{j \in S} p_{ij}^{(m)}(R) C\big(j, R_j\big) - g + \sum_{j \in S} \sum_{z \in S} p_{iz}^{(m)}(R) p_{zj}(R) u_j$$

$$0 = -\sum_{j \in S} p_{ij}^{(m)}(R) C\big(j, R_j\big) + g \sum_{j \in S} p_{ij}^{(m)}(R) - \sum_{j \in S} p_{ij}^{(m)}(R) \sum_{z \in S} p_{zj}(R_j) u_z$$

$$+ \sum_{j \in S} p_{ij}^{(m)}(R) C\big(j, R_j\big) - g + \sum_{j \in S} \sum_{z \in S} p_{iz}^{(m)}(R) p_{zj}(R) u_j$$

και αφού $\sum_{j \in S} p_{ij}^{(m)}(R) = 1$ έπεται

$$0 = g - \sum_{j \in S} \sum_{z \in S} p_{ij}^{(m)}(R)\, p_{jz}\big(R_j\big) u_z - g + \sum_{j \in S} \sum_{z \in S} p_{iz}^{(m)}(R) p_{zj}(R) u_j$$

Συνεχίζοντας

$$0 = -\sum_{z\in S}\sum_{j\in S} p_{ij}^{(m)}(R)p_{jz}(R_j)u_z + \sum_{j\in S}\sum_{z\in S} p_{iz}^{(m)}(R)\, p_{zj}(R)u_j$$

που ισχύει, αφού η αλλαγή συμβολισμού δεν ακυρώνει το γεγονός ότι τα αθροίσματα είναι αντίθετα και άρα έχουν άθροισμα 0. Συνεπώς η επαγωγική διαδικασία ολοκληρώθηκε.

Υπενθυμίζοντας ότι το μέσο συνολικό κόστος είναι:

$$V_m(i,R) = \sum_{k=0}^{m-1}\sum_{j\in S} p_{ij}^{(k)}(R)C(j,R_j)$$

η (2.12) γίνεται

$$u_i = V_m(i,R) - mg + \sum_{j\in S} p_{ij}^{(m)}(R)u_j \;\; i \in S \quad (2.13)$$

Διαιρούμε με m κατά μέλη:

$$\frac{u_i}{m} = \frac{V_m(i,R)}{m} - g + \frac{\sum_{j\in S} p_{ij}^{(m)}(R)u_j}{m}$$

$$g = \frac{V_m(i,R)}{m} + \frac{\sum_{j\in S} p_{ij}^{(m)}(R)u_j}{m} - \frac{u_i}{m}$$

Και παίρνοντας όρια m→ ∞ κατά μέλη, έχουμε ότι:

$$\frac{u_i}{m} \to \infty$$

$$\frac{V_m(i,R)}{m} \to g(R)$$

ενώ $\left|\frac{\sum_{j\in S} p_{ij}^{(m)}(R)u_j}{m}\right| \leq \frac{|\sum_{j\in S} u_j|}{m} \leq \frac{h}{m} \to \infty$, με $h \in \mathbb{R}$ σταθερά.

Άρα g=g(R).

Μένει να αποδείξουμε ότι $u_i(R) = u_i + c\,,\;\; i \in S.$

Πράγματι, αν $(g, \underline{u_i})$ και $(g', \underline{u_i'})$ είναι δύο λύσεις του Σ_1, τότε ικανοποιούν και την (2.13) και επομένως g =g΄=g(R) όπως αποδείχθηκε παραπάνω. Δηλαδή έχουμε:

$$u_i = V_n(i,R) - mg + \sum_{j\in S} p_{ij}^{(m)}(R)u_j$$

$$u_i' = V_n(i,R) - mg + \sum_{j\in S} p_{ij}^{(m)}(R)u_j'$$

και αφαιρώντας κατά μέλη τις δύο προηγούμενες σχέσεις

$u_i - u_i' = \sum_{j\in S} p_{ij}^{(m)}(R_i)\cdot\left(u_j - u_j'\right)$ (2.14) με το δεύτερο μέλος να εξαρτάται από το i.

Αθροίζοντας για $m = 1,2,\dots,n$, στην (2.14) και αλλάζοντας την σειρά άθροισης έχουμε:

$$n(u_i - u_i') = \sum\nolimits_{k=1}^{n} \sum_{j\in S} p_{ij}^{(m)}(R_i)\cdot(u_j - u_j')$$

$$u_i - u_i' = \frac{1}{n}\sum_{k=1}^{n}\sum_{j\in S} p_{ij}^{(m)}(R_i)\cdot\left(u_j - u_j'\right)$$

$$u_i - u_i' = \sum_{j\in S}\left[\frac{1}{n}\sum_{k=1}^{n} p_{ij}^{(m)}(R_i)\right](u_j - u_j')$$

Ωστόσο, από τις υποθέσεις για την $\{X_n(R)\}$, $n\in\mathbb{N}_0$ (μοναδική κλειστή κλάση και πεπερασμένος χώρος καταστάσεων) έχουμε ότι για $m\to\infty$ ισχύει:

$$\tfrac{1}{n}\sum_{k=1}^{n} p_{ij}^{(m)}(R_i) \to \pi_j(R_i)\ \ \forall\, i\in S.$$

Μάλιστα η παραπάνω σχέση ισχύει ακόμη και για παροδικές καταστάσεις i, αφού μετά από πεπερασμένο αριθμό βημάτων θα βρεθούν στην μοναδική κλειστή κλάση (η οποία ως πεπερασμένη είναι γνησίως επαναληπτική).

Ας σημειωθεί εδώ ότι $\pi_j(R_i) = 0\ \forall\, j$ που δεν ανήκει στην μοναδική θετικά επαναληπτική κλάση S(R_i) του S . Εξάλλου $p_{ij}^{(m)}(R_i)\to 0 \forall i\in S$ και j∉S(R_i).

Επομένως, παίρνοντας όρια $m\to\infty$ κατά μέλη στην (2.14), έχουμε:

$$u_i - u_i' = \sum_{j\in S}\pi_j(R_i)\cdot\left(u_j - u_j'\right) = c \text{ , με } c\text{: σταθερά.}$$

Δηλαδή, $u_i = u_i' + c$. Αφού κάθε $u_i(R)$,$i\in S$ είναι λύση του Σ1, τελικά προκύπτει $u_i(R) = u_i + c\ , i\in S$ με $\underline{u_i}$ πιθανή λύση του Σ1.

Γ)Αρχικά θα αποδείξουμε ότι το Σ_2 έχει μία τουλάχιστον λύση, την οποία και θα κατασκευάσουμε.

Έστω $(g,\underline{u_i})$, i ∈ S τυχούσα λύση του Σ_1, τότε και η $(g',\underline{u_i'})$ με $u_i' = u_i + k$, με k∈ ℝ είναι λύση του Σ_1. Πράγματι, αφού ισχύει

$$u_i = C(i,R_i) - g + \sum_{j\in S} p_{ij}(R_i)\cdot u_j \quad (2.8)$$

Τότε

$$u_i' = C(i,R_i) - g + \sum_{j\in S} p_{ij}(R_i)\cdot u_j'$$

$$u_i + k = C(i,R_i) - g + \sum_{j\in S} p_{ij}(R_i)\cdot(u_j + k)$$

$$u_i + k = C(i, R_i) - g + \sum_{j \in S} p_{ij}(R_i) \cdot u_j + \sum_{j \in S} p_{ij}(R_i) \cdot k$$

Όμως αφού $\sum_{j \in S} p_{ij}(R_i) = 1$ έπεται ότι

$$u_i + k = C(i, R_i) - g + \sum_{j \in S} \big(p_{ij}(R_i) \cdot u_j\big) + k$$

$$u_i = C(i, R_i) - g + \sum_{j \in S} (p_{ij}(R_i) \cdot u_j)$$

που ισχύει από (2.8).

Επιλέγουμε $k = -u_s \in \mathbb{R}$. Σε αυτήν την περίπτωση $u_i' = u_i - u_s$ ικανοποιεί την (2.8), δηλαδή αποτελεί λύση του Σ_1 όπως αποδείχθηκε. Ταυτόχρονα ικανοποιεί και την συνθήκη $u_s' = 0$. Επομένως, η $u_i' = u_i - u_s$ αποτελεί λύση του Σ_2. Αυτό σημαίνει ότι το Σ_2 έχει μία τουλάχιστον λύση.

Απόδειξη της μοναδικότητας:

Έστω u_i'' η δεύτερη λύση του Σ_2. Τότε η u_i'' επαληθεύει την (2.8) και λόγω του ερωτήματος (Β) ισχύει

$$u_i'' = u_i' + c\,, c \in \mathbb{R}.$$

Όμως οι u_i', u_i'' αποτελούν λύσεις του Σ_2 και επομένως $u_s' = u_s'' = 0$. Με βάση αυτό η προηγούμενη σχέση δίνει c=0 και $u_i' = u_i''$ ∀ i ∈ S που αποδεικνύει το ζητούμενο.

Το Θεώρημα Ι αν και θεμελιώδες μας υποδεικνύει τον τρόπο εύρεσης των σχετικών τιμών και του μέσου ρυθμού κόστους, αλλά δεν μας παρέχει κάποιο τρόπο για να βελτιώσουμε την πολιτική που έχουμε επιλέξει. Κάτι τέτοιο θα γίνει με το Θεώρημα ΙΙ. Ουσιαστικά θέλουμε να πειραματιστούμε ως εξής: όταν η αρχική κατάσταση είναι η i επιλέγουμε μία απόφαση α και εν συνεχεία εφαρμόζουμε την πολιτική R. Τότε η (2.4) γίνεται:

$$V_n(i, a, R) = C(i, a) + \sum_{j \in S} p_{ij}(a) \cdot V_{n-1}(j, R)$$

και η (2.3) σε συνδυασμό με την (2.5) παίρνει την μορφή

$$C(i, a) + \sum_{j \in S} p_{ij}(a) \cdot V_{n-1}(j, R) \cong ng(R) + C(i, a) - g(R) + \sum_{j \in S} p_{ij}(a) \cdot u_j(R)\,.$$

Καθώς το n μεγαλώνει, για να ισχύει η ισότητα πρέπει να επιλέξουμε μία κατάλληλη απόφαση α=$\bar{R}_i$ που να ελαχιστοποιεί την ποσότητα:

$$C(i, a) - g(R) + \sum_{j \in S} p_{ij}(a) \cdot u_j(R)$$

την καλούμενη και ποσότητα βελτίωσης της πολιτικής.

Θεώρημα ΙΙ (εύρεσης βελτιωμένης πολιτικής)

Δίνονται g, u_i $i \in S$ και πολιτική $\bar{R}$ τέτοια ώστε:

$C(i, \bar{R}_i) - g + \sum_{j \in S} p_{ij}(a) \cdot u_j(R), i \in S$ (2.15).

Τότε για τον ρυθμό κόστους $g(\bar{R})$ που αντιστοιχεί στην πολιτική $\bar{R}$ ισχύει: $g(\bar{R}) \leq g$ (2.16)

Ομοίως, αν $C(i, \bar{R}_i) - g + \sum_{j \in S} p_{ij}(\bar{R}_i) \cdot u_j \geq u_i$, (2.17)

τότε $g(\bar{R}) \geq g$ (2.18)

Απόδειξη: Η απόδειξη είναι παραπλήσια με αυτήν του Θεωρήματος Ι (Β).

Αφού ισχύει η (2.14), με επαγωγή αποδεικνύεται ότι

$V_m(i, \bar{R}) - mg + \sum_{j \in S} p_{ij}^{(m)}(\bar{R}) u_j \leq u_i$ (2.19α)

ή $V_m(i, \bar{R}) - u_i + \sum_{j \in S} p_{ij}^{(m)}(\bar{R}) u_j \leq mg$ (2.19β)

Τελικά, διαιρώντας κατά μέλη με m και παίρνοντας όρια με $m \to \infty$, κατ' αναλογία με το Θεώρημα ΙΙ (Β) έπεται η (2.16).

Απολύτως όμοια είναι και η απόδειξη της (2.18) δεδομένης της (2.17).

Διαισθητικά το Θεώρημα ΙΙ θα μπορούσε να εξηγηθεί ως εξής:

Δεδομένου ενός κόστους ελέγχου $C(i, a_i) - g$ που προκαλείται όταν ευρισκόμενοι στην κατάσταση i λαμβάνουμε την απόφαση a_i, και ενός τελικού κόστους u_j που προκαλείται όταν ακολουθούμενης της πολιτικής $\bar{R}$, η αλυσίδα σταματήσει στην κατάσταση j, η (2.15) καταδεικνύει ότι μας συμφέρει να συνεχίσουμε στην κατάσταση j παρά να παραμείνουμε στην κατάσταση i. Αυτό θα συνεχίσει να ισχύει αν επαναλάβουμε την τακτική αυτή για m βήματα· οπότε προκύπτει η (2.19α) και αφού ο χώρος καταστάσεων είναι πεπερασμένος (το ίδιο θα συμβαίνει και με το πλήθος των στάσιμων πολιτικών), έπεται τελικά η (2.16). Δηλαδή ο τελικός ρυθμός κόστους $g(\bar{R})$ είναι μικρότερος από τον αρχικό g, που με την σειρά του συνεπάγεται βελτίωση της πολιτικής.

Συνεπώς, έχουμε βρει το κριτήριο με το οποίο μπορούμε να βελτιώνουμε την πολιτική μας, αρκεί να λαμβάνουμε αποφάσεις $a_i = \bar{R}_i$ που ικανοποιούν την (2.15). Τότε, η νέα πολιτική θα έχει μικρότερο ή ίσο ρυθμό μέσου κόστους από την υπάρχουσα.

Εν συντομία, η μέθοδος βελτίωσης της πολιτικής συνίσταται στα εξής:

- Για να βελτιώσουμε μία δεδομένη πολιτική R, χρησιμοποιούμε το Θεώρημα Ι ώστε να βρούμε τις σχετικές τιμές $u_i(R)$, i∈S και τον μέσο ρυθμό κόστους g(R).
- Ακολούθως εφαρμόζουμε το Θεώρημα ΙΙ με $g = g(R)$ και $u_i = u_i(R)$, i∈S.
- Κατ' αυτόν τον τρόπο ορίζουμε μία νέα πολιτική $\bar{R}$ για την οποία ισχύει:

  $C(i, \bar{R}_i) - g + \sum_{j \in S} p_{ij}(\bar{R}_i) \cdot u_j \leq u_i$ $i \in S$

  οπότε $g(\bar{R}) \leq g$, που σημαίνει ότι η $\bar{R}$ είναι καλύτερη ή (το λιγότερο) ισοδύναμη με την R.

Για κάθε $i \in S$, η απόφαση $a = \bar{R}_i$ επιλέγεται έτσι ώστε να ελαχιστοποιείται ως προς α∈A(i) η ποσότητα βελτίωσης της πολιτικής $C(i,a) - g(R) + \sum_{j\in S} p_{ij}(a) \cdot u_j(R)$ δηλαδή

$$a = argmin_{a\in A(i)}\{C(i,a) - g(R) + \sum_{j\in S} p_{ij}(a) \cdot u_j(R)\}.$$

Σε αυτό το σημείο παραθέτουμε τον αλγόριθμο βελτίωσης της πολιτικής, ο οποίος κωδικοποιεί την παραπάνω διαδικασία.

Αλγόριθμος Βελτίωσης της πολιτικής

Βήμα 1: Ορίζουμε μία αρχική στάσιμη πολιτική $R = (R_i)_{i\in S}$.

Βήμα 2: Υπολογίζουμε τα $g(R)\,, u_i(R)\,, i \in S$ λύνοντας το γραμμικό σύστημα:

$$\Sigma\colon \begin{cases} u_i = C(i,R_i) - g + \sum_{j\in S} p_{ij}(R_i) \cdot u_j \quad i \in S \\ u_s = 0 \end{cases}$$

Βήμα 3: Βελτιώνουμε την πολιτική, επιλέγοντας για κάθε i εκείνες τις αποφάσεις $a_i = \bar{R}_i$ ώστε να ελαχιστοποιείται η ποσότητα βελτίωσης της πολιτικής:

$$\min_{a\in A(i)} \left[C(i,a) - g(R) + \sum_{j\in S} p_{ij}(a) \cdot u_j(R) \right]$$

Η νέα στάσιμη πολιτική είναι η $\bar{R} = (\bar{R}_i)_{i\in S}$.

Βήμα 4: Αν $\bar{R} = R$, τότε ο αλγόριθμος τερματίζει και η R είναι η βέλτιστη πολιτική με ελάχιστο ρυθμό μέσου κόστους g(R) .

Αν $\bar{R} \neq R$, τότε πηγαίνουμε στο Βήμα 2 αντικαθιστώντας την R με την $\bar{R}$.

Τελικά, μετά την εφαρμογή του αλγορίθμου ουσιαστικά ισχύει:

$$u_j(R^*) = \min_{a\in A(i)} \left[C(i,a) - g(R^*) + \sum_{j\in S} p_{ij}(a) \cdot u_j(R^*) \right] \quad (2.20)$$

με R* την βέλτιστη πολιτική.

Η (2.20) καλείται εξίσωση βελτιστοποίησης αφού από αυτήν προκύπτει η βέλτιστη πολιτική R*.

Η μέθοδος βελτίωσης πολιτικής χρησιμοποιείται στην επιλογή της βέλτιστης πολιτικής σε προβλήματα ελέγχου. Η σύγκλιση της συγκεκριμένης μεθόδου εξαρτάται από τη σύγκλιση της τιμής της πολιτικής στο βέλτιστο σημείο.

2.4 Η μέθοδος του γραμμικού προγραμματισμού

Στην προηγούμενη παράγραφο η εύρεση βέλτιστης πολιτικής για την ελαχιστοποίηση του μέσου κόστους επιλύθηκε με την μέθοδο βελτίωσης πολιτικής. Ένας άλλος τρόπος είναι η θεώρησή του προβλήματος ως πρόβλημα γραμμικού προγραμματισμού (π.γ.π).Το πρώτο βήμα είναι η μορφοποίηση του π.γ.π που φαίνεται παρακάτω.

Μεταβλητές απόφασης:

$$x_{ia}\ , i \in S\ , a \in A(i)$$

Αντικειμενική συνάρτηση:

$$\min z = \min \sum_{i \in S} \sum_{a \in A(i)} c_{ia}(a) x_{ia}$$

με την επιφύλαξη ότι

$$\sum_{a \in A(j)} x_{ja} - \sum_{i \in S} \sum_{a \in A(i)} p_{ij}(a) x_{ia} = 0\ , j \in S$$

$$\sum_{i \in S} \sum_{a \in A(i)} x_{ia} = 1$$

$$x_{ia} \geq 0\ , i \in S\ , a \in A(i)\ .$$

Για λόγους συντομίας, την παραπάνω μορφοποίηση θα την ονομάσουμε ως (π.γ.π-1),ενώ η πρώτη ομάδα περιορισμών θα καλείται "πρώτος περιορισμός". Πριν συνεχίσουμε στο επόμενο βήμα είναι απαραίτητο να εξηγήσουμε τους παραπάνω συμβολισμούς.

Κατ' αρχάς, οι μεταβλητές απόφασης x_{ia} εκφράζουν το ποσοστό σε βάθος χρόνου των χρόνων απόφασης που το σύστημα βρίσκεται στην κατάσταση i$\in S$ και εκτελείται η απόφαση $a \in A(i)$. Συνεπώς, το διπλό άθροισμα:

$$\sum_{i \in S} \sum_{a \in A(i)} c_i(a) x_{ia}$$

εκφράζει το συνολικό μέσο ρυθμό κόστους από το γεγονός ότι η αλυσίδα βρέθηκε σε όλες τις δυνατές καταστάσεις κάτω από όλες τις δυνατές αποφάσεις. Ως εκ τούτου ο περιορισμός της μη-αρνητικότητας είναι προφανής, ενώ το ίδιο ισχύει και για τον δεύτερο περιορισμό (το σύνολο των ποσοστών οφείλει να αθροίζει στην μονάδα).

Τέλος, ο πρώτος περιορισμός γράφεται και ως εξής:

$$\sum_{a \in A(j)} x_{ja} = \sum_{i \in S} \sum_{a \in A(i)} p_{ij}(a) x_{ia}$$

για κάθε $j \in S$, που σημαίνει ότι σε βάθος χρόνου ο μέσος ρυθμός μετακινήσεων από την j πρέπει να είναι ίσος με τον μέσο ρυθμό μετακινήσεων προς την j, για κάθε $j \in S$.

Αλγόριθμος γραμμικού προγραμματισμού

Βήμα 1: Υπολογίζουμε μια βέλτιστη βασική λύση $\{x_{ia}^*\}$ του (π.γ.π-1) με χρήση της μεθόδου Simplex .

Βήμα 2: Για κάθε $i \in S^*$ θέτουμε $R_i^* := a$ με $a: x_{ia}^* > 0$, όπου $S^* = \left\{ i \middle| \sum_{a \in A(i)} x_{ia}^* > 0 \right\}$ μη-κενό σύνολο.

Βήμα 3: Αν $S = S^*$,η βέλτιστη πολιτική μέσου ρυθμού κόστους είναι η $R^* = \{R_i^*\}$ και ο αλγόριθμος τερματίζει. Αλλιώς, βρίσκουμε $i \notin S^*$ και κατάσταση $a \in A(i)$ τέτοια ώστε $p_{ij}(a) > 0$ για κάποιο j$\in S^*$,θέτουμε $R_i^* := a$ και $S^* = S^* \cup \{i\}$ και επαναλαμβάνουμε το 3^o βήμα.

Πριν από την ανάπτυξη της απόδειξης ότι ο προαναφερθείς αλγόριθμος επιλύει την εξίσωση βελτιστοποίησης, πρέπει να αναφέρουμε ότι με την μέθοδο του γραμμικού προγραμματισμού, επιτρέπεται η χαλάρωση της υπόθεσης ότι για κάθε πολιτική R, η αλυσίδα $\{X_n(R)\}_{n\in\mathbb{N}_0}$ έχει μοναδική κλειστή κλάση ή σύνολο. Εν προκειμένω, υποθέτουμε ότι για κάθε βέλτιστη πολιτική R, η $\{X_n(R)\}_{n\in\mathbb{N}_0}$ έχει μοναδική κλειστή κλάση ή σύνολο. Αυτό συνεπάγεται ότι για κάθε βέλτιστη πολιτική R υπάρχουν καταστάσεις $\in S$ που είναι επισκέψιμες από κάθε άλλη κατάσταση $j \in S$. Λόγω του πεπερασμένου χώρου καταστάσεων S, το μοναδικό κλειστό σύνολο περιέχει όλες τις γνησίως επαναληπτικές καταστάσεις, οι οποίες και είναι τελικά επισκέψιμες από κάθε άλλη κατάσταση -οι υπόλοιπες καταστάσεις είναι παροδικές. Επισημαίνουμε ότι πλέον δεν απαιτούμε τα παραπάνω να ισχύουν απαραίτητα για κάθε πολιτική R, αλλά μόνο για τις βέλτιστες πολιτικές πράγμα που σημαίνει ότι η υπόθεση είναι πιο ασθενής από την αρχική. Όσον αφορά τις μη-βέλτιστες πολιτικές, αυτές επιτρέπεται να ορίζουν και πολλαπλές κλειστές κλάσεις οι οποίες θα είναι ξένες μεταξύ τους. Βέβαια, αν αλλάξει η πολιτική τότε ενδεχομένως να αλλάξουν και οι ανοικτές ή και οι κλειστές κλάσεις της νέας αλυσίδας $\{X_n(R)\}_{n\in\mathbb{N}_0}$. Ωστόσο, αν η R είναι βέλτιστη, η κλειστή κλάση θα είναι αναγκαστικά μοναδική.

Κάτω από την προαναφερθείσα υπόθεση, η εξίσωση βελτιστοποίησης δίνει μοναδική λύση τον ελάχιστο αναμενόμενο ρυθμό κόστους, ο οποίος μάλιστα προκύπτει ανεξάρτητος της αρχικής κατάστασης i.

Θεώρημα ΙΙΙ

Έστω η μαρκοβιανή αλυσίδα $\{X_n\}_{n\in\mathbb{N}_0}$, για την οποία ισχύει ότι για κάθε βέλτιστη πολιτική R, η $\{X_n(R)\}_{n\in\mathbb{N}_0}$ έχει μοναδική κλειστή κλάση ή σύνολο. Τότε ο αλγόριθμος γραμμικού προγραμματισμού που αναφέρθηκε παραπάνω δίνει μία πολιτική που είναι βέλτιστη ως προς τον μέσο ρυθμό κόστους.

Απόδειξη:

Αρχικά υποθέτουμε ότι η εξίσωση βελτιστοποίησης έχει λύση την $(g^*, u_i^*), i \in S$. Δηλαδή ισχύει

$$u_i^* = \min_{a\in A(i)} \{c_i(a) - g^* + \sum_{j\in S} p_{ij}(a)\, u_j^*\} \quad (2.21), \; i \in S \text{ και}$$

$$g^* = \min_R g(R) \; (2.22)$$

όπου βέλτιστη είναι κάθε στάσιμη πολιτική R^* τέτοια ώστε για κάθε $i \in S$ η δραστηριότητα R_i ελαχιστοποιεί το δεύτερο μέλος της (2.21).

Από την (2.21) έπεται ότι οι ανισώσεις:

$$u_i \leq c_i(a) - g + \sum_{j\in S} p_{ij}(a) u_j \;\; , \forall i \in S \; \kappa\alpha\iota \; a \in A(i)$$

έχουν λύση (2.23). Οπότε χρησιμοποιώντας το Θεώρημα ΙΙ για κάθε $i \in$ S και α = R_i από την (2.17) έπεται ότι ισχύει η (2.18), δηλαδή:

$$g_i(R) \geq g \;\; \forall \, i \in S,$$

με $g_i(R)$ τον μέσο ρυθμό κόστους σε βάθος χρόνου κάτω από την στάσιμη πολιτική R και με αρχική κατάσταση i.

Τότε $\min_i g_i(R) \geq g$ και λόγω της (2.22) έχουμε $g^* \geq g$ ή $g \leq g^*$ (2.24)

Με λίγα λόγια ,οι (2.22) και οι (2.23) καταδεικνύουν ότι για να λύσουμε τις (2.21)-(2.22) αρκεί να βρούμε την μέγιστη τιμή g τέτοια ώστε οι ανισώσεις $u_i \leq c_i(a) - g + \sum_{j\in S} p_{ij}(a)u_j$, $\forall i \in S$ και $a \in A(i)$ να έχουν λύση.

Με λίγα λόγια αρκεί να λύσουμε το (π.γ.π-2): $\max z = \max g$

με περιορισμούς:

$g + u_i - \sum_{j\in S} p_{ij}(a)u_j \leq c_i(a)$, $\forall i \in S$ και $a \in A(i)$.

g, u_i: χωρίς περιορισμό.

Παρατηρούμε ότι το (π.γ.π-2) είναι το δυϊκό πρόβλημα του (π.γ.π-1). Πράγματι, το πρωτεύον (π.γ.π-1) είναι πρόβλημα ελαχιστοποίησης, άρα το δυϊκό θα είναι πρόβλημα μεγιστοποίησης. Στους περιορισμούς του (π.γ.π-1) αντιστοιχίζουμε τις μεταβλητές απόφασης του δυϊκού όπως φαίνεται παρακάτω:

Πίνακας 1: Αντιστοιχία πρωτεύοντος (π.γ.π-1) και δυϊκού (π.γ.π-2).

Περιορισμός πρωτεύοντος	Μεταβλητή δυϊκού	Συντελεστής αντικειμενικής συνάρτησης δυϊκού
$\sum_{a\in A(i)} x_{ja} - \sum_{i\in S}\sum_{a\in A(i)} p_{ij}(a)\, x_{ia} = 0,\ j \in S$	u_j , $j \in S$	0 $j \in S$
$\sum_{i\in S}\sum_{a\in A(i)} x_{ia} = 1$	G	1

Άρα η αντικειμενική συνάρτηση του δυϊκού είναι $z = 1 \cdot g + 0 \cdot \sum_{j\in S} u_j = g$ και αφού έχουμε πρόβλημα μεγιστοποίησης, οι περιορισμοί θα έχουν την μορφή ανισώσεων με φορά $\leq c_i(a)$ $\forall$ $i \in$ S και $a \in A(i)$. Έπειτα από ορισμένες πράξεις αποδεικνύεται ότι οι περιορισμοί του (π.γ.π-2) παίρνουν τελικά την μορφή:

$$g + u_i - \sum_{j\in S} p_{ij}(a)u_j \leq c_i(a) \quad \forall i \in S \text{ και } a \in A(i) .$$

Εφόσον τα (π.γ.π-1) και (π.γ.π-2) έχουν βέλτιστη βασική εφικτή λύση, θα έχουν την ίδια βέλτιστη τιμή g^* της αντικειμενικής συνάρτησης που αποτελεί και τον ελάχιστο μέσο ρυθμό κόστους. Η σχέση αυτή (δυϊκού-πρωτεύοντος) φαινομενικά ελάχιστα σχετική με την απόδειξη, θα μας χρειαστεί στη συνέχεια όπως θα δούμε παρακάτω.

Μια εύλογη απορία που ενδεχομένως να δημιουργείται είναι κατά πόσον μία βέλτιστη βασική εφικτή λύση $\{x^*_{ia}\}$ του αλγορίθμου που αναφέρθηκε λίγο πιο πάνω ορίζει μία βέλτιστη πολιτική μέσου ρυθμού κόστους. Κάτι τέτοιο είναι αληθές και η απόδειξή του θα γίνει σε τρία βήματα.

Βήμα 1: Το σύνολο $S^* = \{i | \sum_{a\in A(i)} x^*_{ia} > 0\}$ είναι κλειστό (κλειστό σύνολο επικοινωνίας) κάτω από οποιαδήποτε στάσιμη πολιτική R.

Απόδειξη:

Ο δεύτερος περιορισμός του (π.γ.π-1) εξασφαλίζει ότι το σύνολο S^* είναι μη-κενό.

Έστω ότι υπάρχει κατάσταση $i \in S^*$και απόφαση $a' \in A(i)$, τέτοια ώστε $p_{ij}(a') > 0$ για κάποια j$\notin$ S^*. Αν δεν ισχύει κάτι τέτοιο, τότε δεν έχουμε τίποτα να αποδείξουμε. Επιλέγουμε την κατάσταση $i \in S^*$ ώστε $x^*_{ia'} > 0$. Είναι εύλογο να αναρωτηθούμε αν έχουμε την δυνατότητα μιας τέτοιας επιλογής. Ενδεχομένως για την κατάσταση i να ισχύει μεν $p_{ij}(a') > 0$ αλλά όχι και $x^*_{ia'} > 0$. Ωστόσο και σε αυτήν την περίπτωση, μπορούμε να βρούμε μία κατάσταση $l \in S^*$ τέτοια ώστε $x^*_{la''} > 0$ και $p_{li}(a'')$. Αν η j είναι προσβάσιμη από την i και η i είναι προσβάσιμη από την l, τότε και η j είναι προσβάσιμη από την l. Ας βρούμε μία τέτοια κατάσταση $l \in S^*$:

Από τον πρώτο περιορισμό του (π.γ.π-1),βάζοντας όπου j το $i \in S^*$ και όπου i το $l \in S^*$, λαμβάνοντας υπόψη ότι $x_{la} = 0$ για κάθε $l \notin S^*$ και $a \in A(l)$ έχουμε:

$$\sum_{l \in S^*} \sum_{a \in A(l)} p_{li}(a) x^*_{la} = \sum_{a \in A(i)} x^*_{ia} > 0$$

Επομένως υπάρχει $l \in S^*$ και $a \in A(l)$ τέτοια ώστε $p_{li}(a) x^*_{la} > 0$ δηλαδή η l έχει πρόσβαση στην $j \notin S^*$ και επιπλέον $x^*_{la} > 0$.

Τέλος, χρησιμοποιώντας ξανά τον πρώτο περιορισμό του (π.γ.π-1) για το $l \in S^*$ και το $j \notin S^*$ έχουμε:

$$\sum_{a \in A(j)} x^*_{ja} = \sum_{l \in S} \sum_{a \in A(l)} p_{lj}(a) x^*_{la} > 0$$

που συνεπάγεται $j \in S^*$: Άτοπο. Άρα, η κλάση S^*είναι κλειστή.

Βήμα 2:Το 3^{o} βήμα του αλγορίθμου ολοκληρώνεται σε πεπερασμένα βήματα. Αυτό σημαίνει ότι μπορούμε να βρούμε μια πολιτική R^*.

Απόδειξη:

Υπενθυμίζουμε ότι από την υπόθεση ισχύει ότι για κάθε βέλτιστη πολιτική R, η $\{X_n(R)\}_{n \in \mathbb{N}_0}$ έχει μοναδική κλειστή κλάση. Ωστόσο, από το 1ο βήμα έπεται ότι το σύνολο S* είναι κλειστό, άρα είναι και το μοναδικό κλειστό σύνολο και αφού είναι πεπερασμένο λόγω του πεπερασμένου χώρου καταστάσεων S, όλες οι επαναληπτικές καταστάσεις θα ανήκουν απαραίτητα σε αυτό. Άρα αναγκαστικά αν $i \notin S^*$, η κατάσταση i είναι παροδική κάτω από μια οποιαδήποτε βέλτιστη πολιτική μέσου ρυθμού κόστους. Συνεπώς αν υπάρχει παροδική κατάσταση i αυτή δεν θα ανήκει στην S^*, άρα μπορούμε να επιλέξουμε i$\notin$ S^* και κατάσταση $a \in A(i)$ τέτοια ώστε $p^*_{ij}(a) > 0$ για κάποιο $j \in S^*$ και να ολοκληρώσουμε το βήμα 3. Αν δεν υπάρχει παροδική κατάσταση, όλες οι καταστάσεις είναι επαναληπτικές. Επομένως, S^*=S και ο αλγόριθμος τερματίζει άμεσα.

Βήμα 3: Η πολιτική R* που κατασκευάστηκε με τον τερματισμό του αλγορίθμου είναι βέλτιστη ως προς τον μέσο ρυθμό κόστους.

Απόδειξη:

Έστω $\{x^*_{ia}\}, i \in S, a = R^*_i$, η βέλτιστη βασική (εφικτή) λύση του πρωτεύοντος (π.γ.π-1) και (g*$\underline{u}^*_i$) , i$\in$ S η αντίστοιχη λύση του δυϊκού (π.γ.π-2).

Τότε από το Θεώρημα συμπληρωματικότητας για τα δύο προβλήματα, έπεται ότι για κάθε $i \in S^*$ ισχύει:

$$\left[c_i(R_i^*) - (g^* + u_i^* - \sum_{j\in S} p_{ij}(R_i^*)u_j^*)\right] x_{ia}^* = 0$$

με $a = R_i^*$ να έχει επιλεχθεί έτσι ώστε $x_{ia}^* > 0$ (βήμα 2), οπότε έχουμε για κάθε $i \in S^*$:

$$c_i(R_i^*) - (g^* + u_i^* - \sum_{j\in S} p_{ij}(R_i^*)u_j^*) = 0$$

$$c_i(R_i^*) = g^* + u_i^* - \sum_{j\in S} p_{ij}(R_i^*)u_j^*.$$

Αν η αρχική κατάσταση $i \in S^*$ είναι επαναληπτική, τότε η αλυσίδα θα μείνει σε επαναληπτικές καταστάσεις και συνεπώς $j \in S^*$:επαναληπτική (προφανώς κάτω από την R^*). Συμβολίζοντας ως $S(R^*)$ το σύνολο των επαναληπτικών καταστάσεων του S κάτω από την πολιτική R^* , έχουμε την (2.25):

$$c_i(R_i^*) = g^* + u_i^* - \sum_{j\in S(R^*)} p_{ij}(R_i^*)u_j^*$$

για κάθε $i \in S^*$.

Συνεχίζοντας όπως στην απόδειξη του Θεωρήματος I (B), η (2.25) δίνει ότι η ποσότητα g^* είναι ο ελάχιστος ρυθμός μέσου κόστους κάτω από την πολιτική R* για κάθε αρχική κατάσταση i που είναι επαναληπτική. Άρα η R είναι βέλτιστη.

Στην περίπτωση που η αλυσίδα μας αρχίσει από παροδική κατάσταση, τότε μετά από πεπερασμένα βήματα θα βρεθεί σε επαναληπτική κατάσταση, οπότε θα παραμείνει για πάντα στην S(R*) και η (2.25) συνεχίζει να ισχύει και σε αυτήν την περίπτωση. Επομένως, ισχύουν απαράλλακτα τα προαναφερθέντα και η απόδειξη ολοκληρώθηκε.

Η περίπτωση των στοχαστικών περιορισμών

Σε αυτό το εδάφιο θα γίνει μία σύντομη αναφορά στην περίπτωση των στοχαστικών περιορισμών ή περιορισμών τυχαιότητας ή περιορισμών πιθανοτήτων. Το ερώτημα που τίθεται εδώ είναι κατά πόσο θα αλλάξει ο αλγόριθμος αν προστεθούν περιορισμοί πιθανοτήτων. Δεν είναι καθόλου σπάνιο να υπάρχουν περιορισμοί που επηρεάζουν συγκεκριμένες συχνότητες για κάποιες καταστάσεις της αλυσίδας. Αυτό σημαίνει ότι κάποιοι περιορισμοί δεν απασχολούν αιτιοκρατικά το σύστημά μας αλλά ασκούν επίδραση σε κάποιες πιθανότητες αυτού.

Για παράδειγμα κάτι τέτοιο είναι δυνατό να συμβεί όταν:

- σε ένα πρόβλημα συντήρησης, με τυχαίες αλλαγές στην κατάσταση του υπό συντήρηση μηχανήματος, υπάρχει περιορισμός στην συχνότητα εμφάνισης μίας βλάβης.
- σε πρόβλημα ελέγχου αποθεμάτων, με δυσκολία στον ακριβή υπολογισμό του κόστους έλλειψης μίας μονάδας, υπάρχουν περιορισμοί όσον αφορά τις πιθανότητες έλλειψης ή περιορισμοί στο ποσοστό των περιόδων στις οποίες η ζήτηση δεν καλύφθηκε από το διαθέσιμο απόθεμα.

Φυσικά τέτοιου είδους περιορισμοί επηρεάζουν την λήψη της βέλτιστης απόφασης και κατά συνέπεια αλλάζουν δραστικά το αποτέλεσμα.

Υπενθυμίζουμε τον ορισμό μιας στάσιμης τυχαιοποιημένης πολιτικής:

Μία πολιτική π καλείται στάσιμη τυχαιοποιημένη πολιτική αν μπορεί να περιγραφεί από μία κατανομή πιθανοτήτων $\{\pi_a(i), a \in A(i)\}$ για κάθε $i \in S$, τέτοια ώστε αν η αλυσίδα βρίσκεται στην κατάσταση i,η απόφαση $a \in A(i)$ θα επιλεγεί με πιθανότητα $\pi_a(i)$.

Προφανώς $\pi_a(i) \geq 0$ και $\sum_{a\in A(i)} \pi_a(i) = 1\ \forall i \in S$.

Ειδική περίπτωση: Αν οι $\pi_a(i)$ παίρνουν τις τιμές 0 ή 1 $\forall i \in S$ και $a \in A(i)$, τότε η στάσιμη τυχαιοποιημένη πολιτική π εκφυλίζεται στην γνωστή μας στάσιμη πολιτική, όπου οι επιλογές των αποφάσεων γίνονται αιτιοκρατικά.

Ορισμός ΙΙΙ

Για κάθε πολιτική π, ορίζουμε ως $f_{ia}(\pi)$ το ποσοστό των χρόνων απόφασης σε βάθος χρόνου στις οποίες η αλυσίδα βρισκόταν στην κατάσταση i, χρησιμοποιούνταν η πολιτική π και υπό αυτήν την πολιτική είχε ληφθεί η απόφαση α.

Υπενθυμίζουμε ότι ο στόχος παραμένει η ελαχιστοποίηση του μέσου ρυθμού κόστους. Ωστόσο, στην προκειμένη περίπτωση υπάρχουν γραμμικοί περιορισμοί στις ποσότητες fia (π), οι οποίοι μορφοποιούνται ως εξής:

$$\sum_{i\in S} \sum_{a\in A(i)} \gamma_{ia}(k) \cdot f_{ia}(\pi) \leq \beta(k)$$

για $k = 1,2, \dots, L$ με $\gamma_{\iota\alpha}(k), \beta(k)$: δοσμένες σταθερές.

Τελικά, το (π.γ.π-1) παίρνει την μορφή (π.γ.π-3):

Μεταβλητές απόφασης:

$$x_{ia}, i \in S\ , a \in A(i)$$

Αντικειμενική συνάρτηση:

$$\min z = \min \sum_{i\in S} \sum_{a\in A(i)} c_i(a) x_{ia}$$

Περιορισμοί:

$$\sum_{a\in A(i)} x_{ja} - \sum_{i\in S} \sum_{a\in A(i)} p_{ij}(a) \cdot x_{ia} = 0\ , j \in S$$

$$\sum_{i\in S} \sum_{a\in A(i)} x_{ia} = 1$$

$$\sum_{i\in S} \sum_{a\in A(i)} \gamma_{ia}(k) \cdot x_{ia} \leq \beta(k), k = 1,2, \dots, L$$

$$x_{ia} \geq 0, i \in S, a \in A(i)\ .$$

με $\gamma_{ia}(k), \beta(k)$: δοσμένες σταθερές

Στην συγκεκριμένη περίπτωση, οι μεταβλητές απόφασης x_{ia} υποκαθιστούν τις $f_{ia}(\pi)$ με π την ζητούμενη βέλτιστη τυχαιοποιημένη πολιτική, η οποία υπολογίζεται με βάση το Θεώρημα IV.

Θεώρημα IV

Έστω η μαρκοβιανή αλυσίδα $\{X_n\}_{n\in\mathbb{N}_0}$, για την οποία ισχύει ότι για κάθε πολιτική R, η $\{X_n(R)\}_{n\in\mathbb{N}_0}$ έχει μοναδική κλειστή κλάση ή σύνολο. Έστω $\{x_{ia}^*\}$ η βέλτιστη βασική εφικτή λύση του (π.γ.π-3) και $S^* = \left\{i \middle| \sum_{a\in A(i)} x_{ia}^* > 0\right\}$. Τότε μία τυχαιοποιημένη πολιτική που είναι βέλτιστη ως προς τον μέσο ρυθμό κόστους είναι η

$$\pi_a^*(i) = \begin{cases} x_{ia}^* \Big/ \sum_{a\in A(i)} x_{ia}^* & , a \in A(i) \; και \; i \in S^* \\ c \; αλλιώς \end{cases}$$

όπου c μία αυθαίρετη τιμή.

Προφανώς η ελάχιστη τιμή του μέσου κόστους δίνεται από την βέλτιστη τιμή της αντικειμενικής συνάρτησης του (π.γ.π-3). Αξίζει να αναφέρουμε ότι άλλη μία μέθοδος αντιμετώπισης των περιορισμών τυχαιότητας είναι με χρήση των πολλαπλασιαστών Lagrange, η μελέτη των οποίων ξεφεύγει από τους στόχους της παρούσας εργασίας. Σχετικά μα την μέθοδο αυτή παραπέμπουμε τον αναγνώστη στην μελέτη του Tijms.

Η μέθοδος γραμμικού προγραμματισμού αντιμετωπίζει ζητήματα βελτιστοποίησης που μπορούν να μοντελοποιηθούν με γραμμικούς περιορισμούς και γραμμικές αντικειμενικές συναρτήσεις. Στη συγκεκριμένη περίπτωση, η σύγκλιση αφορά μια βέλτιστη λύση μετά από έναν πεπερασμένο αριθμό επαναλήψεων.

2.5 Η μέθοδος διαδοχικών προσεγγίσεων

Η μέθοδος των διαδοχικών προσεγγίσεων βασίζεται στον δυναμικό προγραμματισμό. Ουσιαστικά με την μέθοδο αυτή υπολογίζεται σύμφωνα με έναν αναδρομικό τύπο η ακολουθία $V_n(i)$ η οποία αποτελεί μία πολύ καλή προσέγγιση του ελάχιστου ρυθμού κόστους $g(R)$. Αφετηρία της μεθόδου αυτής είναι η σχέση (2.4) με την διαφορά ότι στην συγκεκριμένη περίπτωση θέλουμε να επιλέξουμε την απόφαση α ώστε να ελαχιστοποιείται το μέσο συνολικό κόστος. Συνεπώς, αρχίζοντας από μία αρχική συνάρτηση $V_0(i), i \in S$ υπολογίζουμε αναδρομικά για n=1,2,… την

$$V_n(i) = \min_{a\in A(i)} \left\{ C(i,a) + \sum_{j\in S} p_{ij}(a) \cdot V_{n-1}(j) \right\},$$

για κάθε $i \in S$.

Η ποσότητα $V_n(i)$ εκφράζει το ελάχιστο μέσο συνολικό κόστος για τις n περιόδους που απομένουν, όταν η παρούσα κατάσταση είναι i και το σύστημά μας επιβαρύνεται με ένα τερματικό κόστος $V_0(i)$, για να τερματίσει στην κατάσταση j. Υπό αυτήν την οπτική γωνία, η διαφορά $V_n(i) - V_{n-1}(j)$ προσεγγίζει σε βάθος χρόνου (για μεγάλο n) τον ελάχιστο ρυθμό μέσου κόστους. Με βάση αυτή την ιδέα προκύπτει το παρακάτω θεώρημα.

Θεώρημα V (μεθόδου διαδοχικών προσεγγίσεων)

Υποθέτουμε ότι η Μαρκοβιανή αλυσίδα $\{X_n(R)\}_{n\in\mathbb{N}_0}$ έχει μοναδική κλειστή κλάση επικοινωνίας για κάθε στάσιμη πολιτική R και έστω στάσιμη πολιτική R(n) ∀ n∈ ℕ της οποίας οι αποφάσεις $R_i(n), i \in S$, ελαχιστοποιούν το δεύτερο μέλος της ακολουθίας τιμών για κάθε $i \in S$:

$$V_n(i) = \min_{a\in A(i)}\left\{C(i,a) + \sum_{j\in S} p_{ij}(a)\cdot V_{n-1}(j)\right\}, \quad (2.26)$$

με $V_0(i)$ αυθαίρετη, για κάθε $i \in S$.

Ορίζουμε $g^* = \min_R g(R)$ το ελάχιστο ρυθμό κόστους

$$m_n = \min_{j\in S}\{V_n(j) - V_{n-1}(j)\}\ n \in \mathbb{N} \text{ και}$$

$$M_n = \max_{j\in S}\{V_n(j) - V_{n-1}(j)\}\ n \in \mathbb{N}.$$

Τότε ισχύει:

$$m_n \leq g^* \leq g\big(R(n)\big) \leq M_n \quad (i \in S, n \in \mathbb{N}),$$

ενώ η ακολουθία $\{m_n\}_{n\in\mathbb{N}}$ είναι αύξουσα και η ακολουθία $\{M_n\}_{n\in\mathbb{N}}$ είναι φθίνουσα.

Απόδειξη:

Εξ' ορισμού για την $V_n(i)$ ισχύει:

$$V_n(i) \leq C(i,a) + \sum_{j\in S} p_{ij}(a)\cdot V_{n-1}(j), a \in A(i) \quad (2.27)$$

ενώ από τον ορισμό της $R(n)$ έπεται ότι:

$$V_n(i) = C\big(i, R_i(n)\big) + \sum_{j\in S} p_{ij}(R_i(n))\cdot V_{n-1}(j) \quad \text{με } i \in S \quad (2.28)$$

Ωστόσο,

$V_n(i) = V_{n-1}(i) + V_n(i) - V_{n-1}(i)$ (2.29)

με $V_n(i) - V_{n-1}(i) \leq M_n$ (2.30) και $V_n(i) - V_{n-1}(i) \geq m_n$ (2.31), $\forall n \in \mathbb{N}$.

Απόδειξη της ανισότητας $g\big(R(n)\big) \leq M_n$:

Ισχύει:

$$(2.28) \xrightarrow{(2.29)(2.30)} C\big(i, R_i(n)\big) + \sum_{j\in S} p_{ij}\big(R_i(n)\big)\cdot V_{n-1}(j) \leq V_{n-1}(i) + M_n$$

$$C\big(i, R_i(n)\big) - M_n + \sum_{j\in S} p_{ij}\big(R_i(n)\big)\cdot V_{n-1}(j) \leq V_{n-1}(i)$$

Συνεπώς, ισχύει η σχέση (2.15) και από το Θεώρημα II έπεται ότι $g\big(R(n)\big) \leq M_n$.

Επίσης είναι προφανές ότι $g^* \leq g\big(R(n)\big)$ (εξ' ορισμού του g^* ως ελαχίστου ρυθμού κόστους).

Απόδειξη της ανισότητας $m_n \leq g^*$:

Από την (2.27) αν επιλέξουμε απόφαση $a = R_i$ και χρησιμοποιώντας τις σχέσεις (2.29) και (2.31) έχουμε:

$$C(i, R_i) + \sum_{j \in S} p_{ij}(R_i) \cdot V_{n-1}(j) \geq V_n(i) = V_{n-1}(i) + V_n(i) - V_{n-1}(i)$$

$$C(i, R_i) + \sum_{j \in S} p_{ij}(R_i) \cdot V_{n-1}(j) \geq V_{n-1}(i) + m_n$$

$$C(i, R_i) - m_n + \sum_{j \in S} p_{ij}(R_i) \cdot V_{n-1}(j) \geq V_{n-1}(i)$$

Επομένως, ισχύει η σχέση (2.17) και από το Θεώρημα ΙΙ έπεται ότι $g(R) \geq m_n$.

Όμως $g(R) \geq m_n \Rightarrow \min_R g(R) \geq m_n \Rightarrow g^* \geq m_n$.

Απόδειξη ότι η $\{m_n\}_{n \in \mathbb{N}}$ είναι αύξουσα:

Αρκεί να δείξουμε ότι ισχύει: $m_k \leq m_{k+1}, \forall k \in \mathbb{N}$.

Αν αντικαταστήσουμε στην (2.27) όπου n=k και $a = R_i(k+1)$ έχουμε:

$$V_k(i) \leq C\big(i, R_i(k+1)\big) + \textstyle\sum_{j \in S} p_{ij}(R_i(k+1)) \cdot V_{k-1}(j) \text{ ή}$$

$$-V_k(i) \geq -C\big(i, R_i(k+1)\big) - \textstyle\sum_{j \in S} p_{ij}\big(R_i(k+1)\big) \cdot V_{k-1}(j) \quad (2.32)$$

Ενώ για $n = k+1$ η (2.22) γίνεται:

$$V_{k+1}(i) = C\big(i, R_i(k+1)\big) + \textstyle\sum_{j \in S} p_{ij}(R_i(k+1)) \cdot V_k(j) \quad (2.33)$$

Προσθέτοντας στην (2.32) την (2.33) έχουμε:

$$V_{k+1}(i) - V_k(i) \geq C\big(i, R_i(k+1)\big) + \sum_{j \in S} p_{ij}(R_i(k+1)) \cdot V_k(j) -$$

$$-C\big(i, R_i(k+1)\big) - \sum_{j \in S} p_{ij}(R_i(k+1)) \cdot V_{k-1}(j)$$

$$V_{k+1}(i) - V_k(i) \geq \sum_{j \in S} p_{ij}(R_i(k+1)) \cdot \{V_k(j) - V_{k-1}(j)\}$$

$$V_{k+1}(i) - V_k(i) \geq \sum_{j \in S} p_{ij}(R_i(k+1)) \cdot \min_{j' \in S}\{V_k(j') - V_{k-1}(j')\}$$

$$V_{k+1}(i) - V_k(i) \geq \min_{j' \in S}\{V_k(j') - V_{k-1}(j')\} \cdot \sum_{j \in S} p_{ij}(R_i(k+1))$$

$$V_{k+1}(i) - V_k(i) \geq \min_{j' \in S}\{V_k(j') - V_{k-1}(j')\}, \text{ αφού } \textstyle\sum_{j \in S} p_{ij}\big(R_i(k+1)\big) = 1.$$

$$V_{k+1}(i) - V_k(i) \geq m_k \text{ , άρα και } \min_{j \in S}\{V_{k+1}(i) - V_k(i)\} \geq m_k,$$

Δηλαδή $m_{k+1} \geq m_k$ για κάθε $k \in \mathbb{N}$ (η επιλογή του k ήταν τυχαία).

Απόδειξη ότι η $\{M_n\}_{n \in \mathbb{N}}$ είναι φθίνουσα:

Αρκεί να δείξουμε ότι ισχύει: $M_{k+1} \leq M_k$, $\forall k \in \mathbb{N}$.

Αν αντικαταστήσουμε στην (2.27) όπου $n = k + 1$ και $a = R_i(k)$ έχουμε:

$$V_{k+1}(i) \leq C\big(i, R_i(k)\big) + \textstyle\sum_{j \in S} p_{ij}(R_i(k)) \cdot V_k(j) \quad (2.34)$$

Ενώ για $n = k$ η (2.28) γίνεται:

$$V_k(i) = C\big(i, R_i(k)\big) + \sum_{j \in S} p_{ij}(R_i(k)) \cdot V_{k-1}(j)$$

$$-V_k(i) = -C\big(i, R_i(k)\big) - \textstyle\sum_{j \in S} p_{ij}\big(R_i(k)\big) \cdot V_{k-1}(j) \quad (2.35)$$

Προσθέτοντας στην (2.34) την (2.35) έπεται ότι:

$$V_{k+1}(i) - V_k(i) \leq C\big(i, R_i(k)\big) + \sum_{j \in S} p_{ij}\big(R_i(k)\big) \cdot V_k(j) - C\big(i, R_i(k)\big) - \sum_{j \in S} p_{ij}(R_i(k)) \cdot V_{k-1}(j)$$

$$V_{k+1}(i) - V_k(i) \leq \sum_{j \in S} p_{ij}(R_i(k)) \cdot \{V_k(j) - V_{k-1}(j)\}$$

$$V_{k+1}(i) - V_k(i) \leq \sum_{j \in S} p_{ij}(R_i(k)) \cdot \max_{j' \in S}\{V_k(j') - V_{k-1}(j')\}$$

$$V_{k+1}(i) - V_k(i) \leq \max_{j' \in S}\{V_k(j') - V_{k-1}(j')\} \cdot \sum_{j \in S} p_{ij}(R_i(k))$$

$$V_{k+1}(i) - V_k(i) \leq \max_{j' \in S}\{V_k(j') - V_{k-1}(j')\}$$

$$V_{k+1}(i) - V_k(i) \leq M_k \text{ και άρα } \max_{i \in S}\{V_{k+1}(i) - V_k(i)\} \leq M_k.$$

δηλαδή $M_{k+1} \leq M_k$ για κάθε $k \in \mathbb{N}$ και η απόδειξη ολοκληρώθηκε.

Από το Θεώρημα V, θα προσπαθήσουμε να προσεγγίσουμε το μέσο ρυθμό κόστους g^* από την ακολουθία $\mathrm{g}(R(n))$ για n μεγάλο. Για να επιτύχουμε κάτι τέτοιο, αρχικά υποθέτουμε ότι $C(i, a) \geq 0$. Αν δεν συμβαίνει κάτι τέτοιο, προσθέτουμε κατάλληλη σταθερά ώστε να το επιτύχουμε– τότε ο μέσος ρυθμός κόστους μεταβάλλεται κατά αυτήν την σταθερά C, αφού

$$g'(R) = \lim_{m \to \infty} \frac{V'_m(i, R)}{m} = \lim_{m \to \infty} \frac{\sum_{k=0}^{m-1} \sum_{j \in S} p_{ij}^{(k)}(R)\big[C\big(j, R_j\big) + C\big]}{m} =$$

$$\lim_{m \to \infty} \frac{mC + \sum_{k=0}^{m-1} \sum_{j \in S} p_{ij}^{(k)}(R) C(j, R_j)}{m} = g(R) + C.$$

Επιλέγουμε

$$V_0(i): 0 \leq V_0(i) \leq \min_{a \in A(i)} C(i,a) \quad \forall i \in S.$$

Τότε,

$$C(i,a) + \sum_{j \in S} p_{ij}(a) \cdot V_0(j) \geq V_0(i) \;\; \forall i \in S \; a \in A(i)$$

και

$$V_1(i) = \min_{a \in A(i)} \left\{ C(i,a) + \sum_{j \in S} p_{ij}(a) \cdot V_o(j) \right\} \geq V_0(i)$$

δηλαδή $V_1(i) \geq V_0(i) \quad \forall i \in S$

$V_1(i) - V_0(i) \geq 0 \; \forall i \in S$

Οπότε $m_0 = \min_{i \in S}\{V_1(i) - V_0(i)\}$ και αφού η m_k είναι αύξουσα ισχύει $m_k \geq 0 \; \forall k \in \mathbb{N}$.

Συνεπώς, έχουμε πλέον την ελαφρά αλλά σημαντικά αλλαγμένη σχέση:

$$0 \leq m_n \leq g^* \leq g\big(R(n)\big) \leq M_n \;\; (i \in S, n \in \mathbb{N}) \quad (2.36)$$

Επιπλέον υποθέτουμε ότι η μεταβολή $\frac{M_n - m_n}{m_n} \leq \varepsilon(*)$ είναι μικρότερη από ένα θετικό ε, το οποίο μπορεί να είναι οσοδήποτε μικρό.

Τότε $(*) \Leftrightarrow M_n \leq (\varepsilon + 1)m_n$ και από την (2.36) έχουμε:

$$0 \leq g^* \leq g(R(n)) \leq (\varepsilon + 1)m_n$$

$$0 \leq \frac{g\big(R(n)\big) - g^*}{g^*} \leq \frac{(\varepsilon + 1)m_n - g^*}{g^*} = (e + 1)\frac{m_n}{g^*} - 1 \leq \varepsilon$$

με την τελευταία ανισότητα να ισχύει καθώς από την (2.36) προκύπτει $\frac{m_n}{g^*} \leq 1$

και άρα $(\varepsilon + 1)\frac{m_n}{g^*} \leq \varepsilon + 1 \Leftrightarrow (\varepsilon + 1)\frac{m_n}{g^*} - 1 \leq \varepsilon$.

Τελικά, αποδείχθηκε ότι αν $\frac{M_n - m_n}{m_n} \leq \varepsilon \;\; \tau \acute{o} \tau \varepsilon \; 0 \leq \frac{g\left(R(n)\right) - g^*}{g^*} \leq \varepsilon$.

Αφού $g(R(n))$ ο ρυθμός κόστους στην ν-οστή επανάληψη και g^* ο ελάχιστος ρυθμός κόστους, συμπεραίνουμε ότι αν η μεταβολή της m_n από την M_n είναι μικρότερη από ένα θετικό ε, τότε και η μεταβολή του g^* από τον $g(R(n))$ είναι το πολύ της τάξεως του 100ε%. Ισοδύναμα, αν το ε είναι πολύ μικρό έχουμε μία πολύ καλή προσέγγιση του μέσου ρυθμού κόστους g^* από την $g(R(n))$.

Φαίνεται λογικό να θεωρήσουμε ότι οι ακολουθίες m_n, M_n συγκλίνουν και μάλιστα στον ίδιο αριθμό. Σε αυτήν την περίπτωση, η (2.36) δίνει από το κριτήριο παρεμβολής $\lim_{n \to \infty} g\big(R(n)\big) = g^*$ και παίρνουμε άμεσα ότι η επανάληψη της (2.36) για κάποιο (ίσως μεγάλο) αριθμό βημάτων n δίνει μια καλή προσέγγιση του μέσου ρυθμού κόστους. Δυστυχώς, οι ακολουθίες δεν συγκλίνουν απαραίτητα στον ίδιο αριθμό.

Γενικά, αν υποθέσουμε την ιδιότητα της απεριοδικότητας για το σύστημά μας, τότε οι m_n , M_n συγκλίνουν στο ίδιο όριο [από την (2.36) αυτό θα είναι το g^*] και άρα η κεντρική ιδέα της προσέγγισης του αναμενόμενου ρυθμού κόστους από τις m_nκαι M_n λειτουργεί. Ενδεικτικά ισχύει το παρακάτω θεώρημα.

Θεώρημα VI

Υποθέτουμε ότι για κάθε βέλτιστη στάσιμη πολιτική R η μαρκοβιανή αλυσίδα $\{X_n(R)\}_{n\in\mathbb{N}_0}$ έχει μοναδική κλειστή κλάση επικοινωνίας και είναι απεριοδική. Αν g* ο ελάχιστος ρυθμός μέσου κόστους, τότε υπάρχουν $\alpha > 0$ και $0 < \beta < 1$:

i) $|M_n - m_n| \leq \alpha\beta^n$, $n \in \mathbb{N}$ και

ii) $\lim\limits_{n\to\infty} M_n = g^* = \lim\limits_{n\to\infty} m_n$.

Η υπόθεση της απεριοδικότητας στο προηγούμενο θεώρημα είναι ρεαλιστική, αφού μία περιοδική μαρκοβιανή αλυσίδα μπορεί να μετασχηματιστεί σε μία απεριοδική με την ίδια στάσιμη κατανομή και άρα τον ίδιο ρυθμό μέσου κόστους.

Κάτι τέτοιο επιτυγχάνεται αν η νέα μαρκοβιανή αλυσίδα $\overline{X_n}$ έχει πιθανότητες μετάβασης

$$\bar{p}_{ij} = \begin{cases} \tau p_{ij} \ , i \neq j \\ \tau p_{ii} + 1 - \tau \ , i = j \end{cases} ,0 < \tau < 1 .$$

Πράγματι, $\overline{p_{ii}} > 0$ και η $\overline{X_n}$ είναι απεριοδική, ενώ για να βρούμε την στάσιμη κατανομή της αρκεί να λύσουμε το σύστημα (1.5):

$$\pi_j = \sum_{i\in S} \pi_i \bar{p}_{ij} \Leftrightarrow \pi_j - \pi_j \bar{p}_{jj} = \sum_{i\neq j} \pi_i \bar{p}_{ij} \Leftrightarrow \pi_j(\tau - tp_{jj}) = \sum_{i\neq j} \pi_i \tau p_{ij} \Leftrightarrow$$

$$\pi_j = \pi_j \tau p_{jj} + \sum_{i\neq j} \pi_i \tau p_{ij} \overset{\tau>0}{\Leftrightarrow} \pi_j = \sum_{i\in S} \pi_i p_{ij} \ , \forall j \in S ,$$

που είναι ακριβώς το σύστημα (1.5) για την αρχική κατάσταση X_n.

Κωδικοποιούμε το Θεώρημα V και παίρνουμε τον παρακάτω αλγόριθμο.

Αλγόριθμος διαδοχικών προσεγγίσεων

Βήμα 1: Για κάθε i ∈ S επιλέγουμε αυθαίρετα $V_0(i)$ τέτοιο ώστε

$0 \leq V_0(i) \leq minC(i,a), \forall i \in S.$

Επίσης n=1 και επιλέγω ε > 0 την επιθυμητή ακρίβεια.

Βήμα 2: Υπολογίζουμε τις τιμές $V_n(i) = \min\limits_{a\in A(i)} \{C(i,a) + \sum_{j\in S} p_{ij}(a) \cdot V_{n-1}(i)\}$

και ορίζουμε την αντίστοιχη πολιτική $R(n) = \{R_i(n)\}_{i\in S}$, οι αποφάσεις της οποίας ελαχιστοποιούν το 2ο μέλος της παραπάνω ισότητας.

Βήμα 3: Υπολογίζουμε τα φράγματα $m_n = \min\limits_{j\in S}\{V_n(j) - V_{n-1}(j)\}$ και

$M_n = \max_{j \in S}\{V_n(j) - V_{n-1}(j)\}$. Ο αλγόριθμος τερματίζει με την στάσιμη πολιτική $R(n)$ όταν

$0 \leq M_n - m_n \leq m_n \cdot \varepsilon$. Ο ρυθμός μέσου κόστους g(R) δίνεται προσεγγιστικά από τις ποσότητες M_n, m_n .

Αν ο αλγόριθμος δεν τερματίσει, πηγαίνουμε στο Βήμα 4.

Βήμα 4: Θέτουμε n:=n+1 και πηγαίνουμε στο Βήμα 2.

Αξίζει να σημειωθεί ότι με την μέθοδο διαδοχικών προσεγγίσεων υπολογίζεται μία σχεδόν βέλτιστη πολιτική, κι αυτό λόγω της προσέγγισης του ελάχιστου ρυθμού μέσου κόστους από τον αλγόριθμο. Βέβαια, η επιλογή ενός πολύ μικρού $\varepsilon \geq 0$ καθιστά την προσέγγιση αυτή όσο καλή θέλουμε.

Η μέθοδος αυτή απαιτεί την επανάληψη της διαδικασίας βελτιστοποίησης μέχρι να επιτευχθεί μια αποδεκτή λύση. Η σύγκλιση της συγκεκριμένης μεθόδου στη βέλτιστη λύση εξαρτάται σε μεγάλο βαθμό από τον επιλεγόμενο αλγόριθμο βελτιστοποίησης αλλά και την αρχική εκτίμηση της βέλτιστης λύσης, ενώ μπορεί να επηρεαστεί από την εμφάνιση τοπικών βέλτιστων.

2.6 Μαρκοβιανές διαδικασίες αποφάσεων σε άπειρο χρονικό ορίζοντα

Ας υποθέσουμε στο παρόν εδάφιο ότι έχουμε να αντιμετωπίσουμε ένα πρόβλημα που εξελίσσεται σε άπειρο χρονικό ορίζοντα. Επίσης θεωρούμε ότι υπάρχει ένας θετικός πραγματικός αριθμός B τέτοιος ώστε για κάθε ενέργεια $\alpha \in A(i)$ και κάθε κατάσταση i της διαδικασίας να ισχύει ότι: $|C(i, a)| < B$.

Το συνολικό αναμενόμενο αποπληθωρισμένο κόστος $V_\pi(i)$ σε άπειρο χρονικό ορίζοντα, υπό τον έλεγχο της πολιτικής π, αν η αρχική κατάσταση της διαδικασίας είναι η κατάσταση i ορίζεται ως εξής:

$$V_\pi(i) = E_\pi[\sum_{t=0}^{\infty} \alpha^t C(X_t, a_t)|X_0 = i],$$

Όπου η σταθερά α ανήκει στο διάστημα (0,1), δηλαδή είναι γνησίως μικρότερος της μονάδας, ονομάζεται αποπληθωριστικός παράγοντας. Με E_π συμβολίζουμε τη δεσμευμένη μέση τιμή δοθείσης της πολιτικής π. Η ποσότητα $V_\pi(i)$ είναι καλώς ορισμένη, διότι το γεγονός ότι τα κόστη $C(i, a)$ είναι φραγμένα και ότι $\alpha < 1$, έχει ως συνέπεια $|V_\pi(i)| \leq \frac{B}{1-\alpha}$. Το κίνητρο της εισαγωγής του αποπληθωριστικού παράγοντα είναι οικονομικό διότι το κόστος που θα πληρώσουμε μελλοντικά έχει μικρότερη αξία από αυτό που πληρώνουμε σήμερα.

Το πρόβλημα που θα μας απασχολήσει είναι να βρούμε εκείνη την πολιτική που ελαχιστοποιεί το συνολικό αναμενόμενο αποπληθωρισμένο κόστος για κάθε αρχική κατάσταση της διαδικασίας. Μια πολιτική π^* ονομάζεται $\alpha - \beta έλτιστη$, αν $V_{\pi^*} = inf_\pi V_\pi(i)$ για όλα τα $i \geq 0$.

Έστω $V_\alpha(i) = inf_\pi V_\pi(i)$. Επομένως μία πολιτική π^*καλείται $\alpha - \beta έλτιστη$ αν $V_{\pi^*} = V_\alpha(i)$ για όλα τα $i \geq 0$. Η εξίσωση του επόμενου θεωρήματος είναι γνωστή ως εξίσωση βελτιστοποίησης για τα μοντέλα του άπειρου χρονικού ορίζοντα.

Θεώρημα VII

Γνωρίζοντας ότι ισχύει η ακόλουθη εξίσωση:

$$V_a(i) = \min_{a \in A(i)}\{C(i, a) + a \sum_{j=0}^{\infty} p_{ij}(a) V_a(j)\}, i \geq 0 \quad (2.37)$$

Απόδειξη: Έστω π μια τυχαία πολιτική. Ας υποθέσουμε ότι η πολιτική π κατά την χρονική στιγμή 0 επιλέγει την ενέργεια α με πιθανοτητα P_α , $\alpha \in A(i)$. Τότε:

$$V_\pi(i) = \sum_{\alpha \in A(\iota)} P_\alpha \left\{ C(i,a) + \sum_{j=0}^{\infty} p_{ij}(a) W_\pi(j) \right\},$$

Όπου $W_\pi(j)$ είναι το συνολικό αναμενόμενο αποπληθωρισμένο κόστος από τη χρονική στιγμή 1 και έπειτα, γνωρίζοντας ότι η κατάσταση κατά τη χρονική στιγμή 1 είναι η κατάσταση j. Από τους ορισμούς των $V_a(j)$ και $W_\pi(j)$ συνεπάγεται ότι:$W_\pi(j) \geq aV_a(j)$ και συνεπώς,

$V_\pi(i) \geq \sum_{a \in A(i)} P_a \left\{ C(i,a) + a \sum_{j=0}^{\infty} p_{ij}(a) V_a(j) \right\} \geq \sum_{a \in A(i)} P_a \min_{a \in A(i)} \left\{ C(i,a) + a \sum_{j=0}^{\infty} p_{ij}(a) V_a(j) \right\} = \min_{a \in A(i)} \left\{ C(i,a) + a \sum_{j=0}^{\infty} p_{ij}(a) V_a(j) \right\}$.

Από την παραπάνω ανισότητα λόγω του ότι η πολιτική π είναι τυχαία προκύπτει ότι:

$$V_a(i) \geq \min_{a \in A(i)} \left\{ C(i,a) + a \sum_{j=0}^{\infty} p_{ij}(a) V_a(j) \right\} \quad (2.38)$$

Για την απόδειξη της αντίστροφης ανισότητας θα λειτουργήσουμε ως εξής:

Έστω α_0 η ενέργεια που είναι τέτοια ώστε:

$$C(i,a) + a \sum_{j=0}^{\infty} p_{ij}(a_0) V_a(j) = \min_{a \in A(i)} \left\{ C(i,a) + a \sum_{j=0}^{\infty} p_{ij}(a) V_a(j) \right\} \quad (2.39)$$

Ας υποθέσουμε ότι π είναι η πολιτική που επιλέγεται από την ενέργεια α_0 κατά την χρονική στιγμή 0 και γνωρίζοντας αν η επόμενη κατάσταση είναι j τότε, θεωρώντας ότι η διαδικασία ξεκινάει από την j ακολουθεί μια πολιτική π_j τέτοια ώστε:

$$V_{\pi_j}(i) \leq V_a(j) + \varepsilon \ , \varepsilon > 0$$

Συνεπώς,

$$V_{\pi_j}(i) = C(i,a_o) + a \sum_{j=0}^{\infty} p_{ij}(a_0) V_{\pi_j}(j) \leq C(i,a_0) + a \sum_{j=0}^{\infty} p_{ij}(a_0) V_a(j) + a\varepsilon$$

Από την παραπάνω ανισότητα γνωρίζοντας ότι ισχύει $V_a(i) \leq V_\pi(i)$ έχουμε ότι:

$$V_a(i) \leq C(i,a_0) + a \sum_{j=0}^{\infty} p_{ij}(a_0) V_a(j) + a\varepsilon$$

Από την σχέση (2.39) προκύπτει ότι:

$$V_a(i) \leq \min_{a \in A(i)} \left\{ C(i,a) + a \sum_{j=0}^{\infty} p_{ij}(a) V_a(j) \right\} + a\varepsilon$$

Η εξίσωση (2.37) συνεπάγεται από την παραπάνω σχέση και την εξίσωση (2.38), επειδή το ε είναι ένας αυθαίρετος θετικός αριθμός.

Έστω $B(I)$ το σύνολο των φραγμένων πραγματικών συναρτήσεων επί του χώρου των καταστάσεων της διαδικασίας. Ισχύει ότι $V_\pi(i) \in B(I)$ για όλες τις πολιτικές π. Μια στάσιμη πολιτική είναι μια συνάρτηση $f: I \to A(i)$, όπου I είναι ο χώρος καταστάσεων της διαδικασίας με την ερμηνεία σύμφωνα με την οποία $f(i)$ είναι η ενέργεια που επιλέγεται οποτεδήποτε η κατασταση της διαδικασίας είναι η κατασταση i.

Ορίζουμε λοιπόν την συνάρτηση $T_f: B(I) \to B(I)$ για κάθε πολιτική f τέτοια ώστε:

$$\big(T_f u\big)(i) = C\big(i, f(i)\big) + a \sum_{j=0}^{\infty} p_{ij}\big(f(i)\big) u(j)$$

Δηλαδή η ποσότητα $T_f u$ για κάθε συνάρτηση $u \in B(I)$ είναι μία συνάρτηση της οποίας η τιμή στην κατασταση i δίδεται από τον παραπάνω τύπο. Η συνάρτηση $T_f u$ είναι φραγμένη και συνεπώς ανήκει στο σύνολο $B(I)$. Η τιμή $\big(T_f u\big)(i)$ στην κατασταση i είναι το αναμενόμενο κόστος αν χρησιμοποιήσουμε αρχικά την πολιτική f και μετά από κάποια χρονική στιγμή σταματώντας τη διαδικασία λαμβάνουμε ένα τελικό κόστος ίσο με $au(j)$, αν η τελική κατασταση είναι η j.

Ορισμός IV

Για οποιεσδήποτε συναρτήσεις $u, v \in B(I)$, ισχύει ότι $u \leq v$ αν $u(i) \leq v(i)$ για κάθε $i \geq 0$. Επίσης για $u_n, u \in B(I)$ ισχύει ότι $u_n \to u$, αν $u_n(i) \to u(i)$ καθώς $n \to \infty$, για κάθε $i \geq 0$. Είναι $T_f^1 = T_f$ και για $n > 1$, $T_f^1 = T_f(T_f^{n-1})$.

Ας αναφέρουμε ορισμένες σημαντικές ιδιότητες της T_f.

Λήμμα I

Αν $u, v \in B(I)$, και f μια στάσιμη πολιτική, ισχύει ότι:

(i) $u \leq v \Rightarrow T_f u \leq T_f v$

(ii) $T_f V_f = V_f$ και

(iii) $T_f^n u \to V_f$, για κάθε $u \in B(I)$.

Απόδειξη: το (i) συνεπάγεται από τον ορισμό της T_f.Το (ii) ισοδυναμεί με την εξίσωση:

$$V_f(i) = C\big(i, f(i)\big) + a \sum_{j=0}^{\infty} p_{ij}(f(i)) V_f(j).$$

Η παραπάνω εξίσωση αληθεύει αν δεσμευτούμε στην κατασταση στην οποία μεταβαίνει η διαδικασία κατά την χρονική στιγμή 1, αν χρησιμοποιηθεί η πολιτική f.

Για το (iii) Παρατηρούμε ότι : $T_f : B(I) \to B(I)$ τετοιο ώστε

$\big(T_f u\big)(i) = C\big(i, f(i)\big) + a \sum_{j=0}^{\infty} p_{ij}\big(f(i)\big) u(j)$

$\big(T_f u\big)(i) = (T_f^1 u)(i)$

$(T_f^2 u)(i) = T_f\big(T_f u\big)(i) = C\big(i, f(i)\big) + a \sum_{j=0}^{\infty} p_{ij}\big(f(i)\big)(T_f u)(j) =$

$= C\big(i, f(i)\big) + a \sum_{j=0}^{\infty} p_{ij}\big(f(i)\big)\big\{C\big(j, f(j)\big) + a \sum_{k=0}^{\infty} p_{jk}\big(f(j)\big) u(k)\big\} =$

$= C(i, f(i)) + a\sum_{j=0}^{\infty} p_{ij}(f(i))C(j, f(j)) + a^2 \sum_{j=0}^{\infty}\sum_{k=0}^{\infty} p_{ij}(f(i))p_{jk}(f(j))u(k)$.

$(T_f^3 u)(i) = T_f(T_f^2 u)(i) = T_f(C(i, f(i)) + a\sum_{j=0}^{\infty} p_{ij}(f(i))(T_f^2 u)(j) =$

$= C(i, f(i)) + a\sum_{j=0}^{\infty} p_{ij}(f(i))\{C(j, f(j)) + a\sum_{k=0}^{\infty} p_{jk}(f(j))(T_f u)(k)\} =$

$= C(i, f(is)) + a\sum_{j=0}^{\infty} p_{ij}(f(i))C(j, f(j)) + a^2 \sum_{j=0}^{\infty}\sum_{k=0}^{\infty} p_{ij}(f(i))p_{jk}(f(j))(C(k, f(k)) +$
$a\sum_{r=0}^{\infty} p_{kr}(f(k))u(r))$

$= C(i, f(i)) + a\sum_{j=0}^{\infty} p_{ij}(f(i))C(j, f(j)) + a^2 \sum_{j=0}^{\infty}\sum_{k=0}^{\infty} p_{ij}(f(i))p_{jk}(f(j))(C(k, f(k)) +$
$a^3 \sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\sum_{r=0}^{\infty} p_{ij}(f(i))p_{jk}(f(j))p_{kr}(f(k))u(r)$

Δηλαδή η $(T_f^3 u)$ είναι το αναμενόμενο κόστος αν χρησιμοποιήσουμε την πολιτική f και σταματησουμε τη διαδικασία μετά από τρεις περιόδους, λαμβάνοντας ένα τελικό κόστος ίσο με $a^3 u$. Με επαγωγή αποδεικνύεται ότι ο τελεστής $(T_f^n u)(i)$ είναι το αναμενόμενο κόστος αν χρησιμοποιήσουμε την πολιτική f και σταματήσουμε τη διαδικασία μετά από n περιόδους, λαμβάνοντας αντίστοιχα το τελικό κόστος $a^n u$. Επομένως αφού η συνάρτηση u είναι φραγμένη και $a < 1$ συνεπάγεται ότι $T_f^n u \to V_f$.

Θεώρημα VIII

Έστω f_a η στάσιμη πολιτική η οποία, όταν η διαδικασία βρίσκεται στην κατασταση i επιλέγει την ενέργεια εκείνη που ελαχιστοποιεί το δεξιό μέρος της (3.1), δηλαδή $f_a(i)$ είναι τέτοια ώστε :

$$C(i, f_a(i)) + a\sum_{j=0}^{\infty} p_{ij}(f_a(i))V_a(j) = \min_{a\in A(i)}\left\{C(i,a) + a\sum_{j=0}^{\infty} p_{ij}(a)V_a(j)\right\}, i \geq 0 .$$

Τότε $V_{f_a}(i) = V_a(i)$ για κάθε $i \geq 0$ και συνεπώς η πολιτική f_a είναι $\alpha - \beta\varepsilon\lambda\tau\iota\sigma\tau\eta$.

Απόδειξη: Είναι

$$(T_{f_a}V_a)(i) = C(i, f_a(i)) + a\sum_{j=0}^{\infty} p_{ij}(f_a(i))V_a(j) = = \min_{a\in A(i)}\left\{C(i,a) + a\sum_{j=0}^{\infty} p_{ij}(a)V_a(j)\right\}$$

Οπού η τελευταία ισότητα συνεπάγεται από το Θεώρημα Ι. Επομένως $T_{f_a}V_a = V_a$. Από την τελευταία ισότητα συνεπάγεται $T_{f_a}^2 V_a = T_{f_a}(T_{f_a}V_a) = T_{f_a}V_a = V_a$ και με επαγωγή προκύπτει ότι $T_{f_a}^n V_a = V_a$ για $n \geq 1$. Λαμβάνοντας το όριο για $n \to \infty$ και χρησιμοποιώντας την (iii) περίπτωση από το Λήμμα Ι έχουμε ότι

$V_{f_a} = V_a$.

Συγκεκριμένα αν καταφέρουμε να βρούμε το ελάχιστο συνολικό αναμενόμενο αποπληθωρισμένο κόστος $V_a(i), i \geq 0$, τοτε η στάσιμη πολιτική η οποία, όταν η διαδικασία βρίσκεται στην κατάσταση i επιλέγει εκείνη την ενέργεια που ελαχιστοποιεί την παράσταση $C(i,a) + a\sum_{j=0}^{\infty} p_{ij}(a)V_a(j)$ είναι $\alpha - \beta\varepsilon\lambda\tau\iota\sigma\tau\eta$.

Ας υποθέσουμε ότι για μια στάσιμη πολιτική f έχουμε υπολογίσει το αναμενόμενο κόστος $V_f(i)$, $i \geq 0$.Έστω f^* μια στάσιμη πολιτική η οποία, όταν η διαδικασία βρίσκεται στην κατάσταση i, επιλέγει την ενέργεια εκείνη που ελαχιστοποιεί την παράσταση $C(i,a) + a\sum_{j=0}^{\infty} p_{ij}(a)V_f(j)$, δηλαδη $f^*(i)$ είναι τέτοια ώστε:

$C(i, f^*(i)) + a\sum_{j=0}^{\infty} p_{ij}\big(f^*(i)\big)V_f(j) = \min_{a\in A(i)}\{C(i,a) + a\sum_{j=0}^{\infty} p_{ij}(a)V_f(j)\}, i \geq 0$.

Πόρισμα I

Ισχύει ότι $V_{f^*}(i) \leq V_f(i)$ για κάθε $i \geq 0$.

Απόδειξη: Είναι

$$\left(T_{f^*}V_f\right)(i) = C\big(i, f^*(i)\big) + a\sum_{j=0}^{\infty} p_{ij}\left(f^*(i)V_f(j)\right)$$

$$\leq C\big(i, f(i)\big) + a\sum_{j=0}^{\infty} p_{ij}\big(f(i)\big)V_f(j)$$

$$= V_f(i)$$

Όπου η ανισότητα οφείλεται στον ορισμό της πολιτικής f^* και η ισότητα στην (ii) περίπτωση του Λήμματος I. Συνεπώς $T_{f^*}V_f \leq V_f$ και εφαρμόζοντας T_{f^*} και στα δύο μέλη της παραπάνω ανισότητας από την μονοτονία της T_{f^*} περίπτωση (i) από το λήμμα προκύπτει ότι $T_{f^*}^2V_f \leq T_{f^*}V_f \leq V_f$ και με επαγωγή $T_{f^*}^nV_f \leq V_f$. Καθώς παίρνουμε το όριο για $n \to \infty$ και χρησιμοποιώντας την περίπτωση (iii) του προαναφερθέντος Λήμματος I προκύπτει το επιθυμητό αποτέλεσμα του πορίσματος. Η διαδικασία σύμφωνα με την οποία, ξεκινώντας από μια αρχική στάσιμη πολιτική και εφαρμόζοντας το πόρισμα, βρίσκουμε μια καλύτερη στάσιμη πολιτική ονομάζεται αλγόριθμος βελτίωσης των πολίτικων. Κάθε φορά σκοπός είναι να βελτιώνουμε την προηγούμενη στάσιμη πολιτική που είχε προκύψει μέχρις ότου βρεθεί η βέλτιστη.

2.7 Συστολές και αποπληθωρισμένος δυναμικός προγραμματισμός

Στη συνέχεια, θα εξετάσουμε τον τρόπο με τον οποίο η θεωρία συστολών μπορεί να εφαρμοστεί σε προβλήματα βελτιστοποίησης στοχαστικών μοντέλων. Ας ξεκινήσουμε αναφέροντας ορισμένες βασικές γνώσεις από την θεωρία συστολών. Για οποιαδήποτε συνάρτηση $u \in B(I)$ έστω $\|u\| = sup_{i\geq 0}|u(i)|$.

Ορισμός V

Μια συνάρτηση $T: B(I) \to B(I)$ καλείται συστολή αν $\|T_u - T_v\| \leq \beta\|u - v\|$,για κάποιο $\beta < 1$ και για κάθε $u, v \in B(I)$. Η συνάρτηση $u - v$ είναι συνάρτηση της οποίας η τιμή στο i είναι $u(i) - v(i)$.

Ας επισημάνουμε ένα σημαντικό θεώρημα το οποίο παρουσιάζουμε χωρίς απόδειξη.

Θεώρημα IX (Σταθερού Σημείου για Συστολές)

Αν $T: B(I) \to B(I)$ είναι συστολή, τότε υπάρχει μια και μοναδική συνάρτηση $g \in B(i)$

Τέτοια ώστε :$Tg = g$. Επιπρόσθετα, για κάθε $u \in B(i)$ ισχύει ότι,

$$T^n u \to g \text{ καθώς } n \to \infty$$

Για να εφαρμόσουμε το θεώρημα, ας ορίσουμε την συνάρτηση $T_a: B(I) \to B(I)$ ως εξής :

$$(T_a u)(i) = \min_{a\in A(i)}\left\{C(i,a) + a\sum_{j=0}^{\infty} p_{ij}(a)u(j)\right\} .$$

Από το θεώρημα Ι έπεται ότι $T_a V_a = V_a$.

Συνεπώς, αν καταφέρουμε να αποδείξουμε ότι η συνάρτηση T_a είναι συστολή, τότε αυτό θα συνεπάγεται ότι η V_a μπορεί να βρεθεί τουλάχιστον προσεγγιστικά αν διαδοχικά εφαρμόσουμε την T_a σε οποιαδήποτε αρχική συνάρτηση $u \in B(i)$. Αυτή η τεχνική είναι γνωστή ως μέθοδος των διαδοχικών προσεγγίσεων.

Θεώρημα Χ

Η συνάρτηση T_a είναι μία συστολή.

Απόδειξη: Για κάθε $u, v \in B(i)$ έχουμε ότι

$$(T_a u)(i) - (T_a v)(i)$$

$$= \min_{a \in A(i)} \left\{ C(i,a) + a \sum_{j=0}^{\infty} p_{ij}(a) u(j) \right\} - \min_{a \in A(i)} \left\{ C(i,a) + a \sum_{j=0}^{\infty} p_{ij}(a) v(j) \right\} =$$

$$= \min_{a \in A(i)} \left\{ C(i,a) + a \sum_{j=0}^{\infty} p_{ij}(a) u(j) \right\} - C(i,\bar{a}) - a \sum_{j=0}^{\infty} p_{ij}(\bar{a}) v(j)$$ οπού $\bar{a}$ είναι τέτοια ώστε:

$$C(i,\bar{a}) + a \sum_{j=0}^{\infty} p_{ij}(\bar{a}) v(j) = \min_{a \in A(i)} \left\{ C(i,a) + a \sum_{j=0}^{\infty} p_{ij}(a) v(j) \right\}.$$

Διαδοχικά, έχουμε:

$$(T_a u)(i) - (T_a v)(i) \leq a \sum_{j=0}^{\infty} p_{ij}(\bar{a}) u(j) - a \sum_{j=0}^{\infty} p_{ij}(\bar{a}) v(j)$$

$$= a \sum_{j=0}^{\infty} p_{ij}(\bar{a}) [u(i) - v(j)] \leq a \sum_{j=0}^{\infty} p_{ij}(\bar{a}) sup_j [u(j) - v(j)]$$

$$\leq a sup_j [u(j) - v(j)] = a \|u - v\|$$

Από την παραπάνω ανισότητα έχουμε ότι:

$$sup_{i \geq 0} \{ (T_a u)(i) - (T_a v)(i) \} \leq a sup_j [u(j) - v(j)] = a \|u - v\|$$

Με εναλλαγή των ρολών των u και v προκυπτει

$$sup_{i \geq 0} \{ (T_a v)(i) - (T_a u)(i) \} \leq a sup_j [v(j) - u(j)]$$

Συνδυάζοντας τις δυο παραπάνω ανισότητες έπεται ότι

$$sup_{i \geq 0} |(T_a u)(i) - (T_a v)(i)| \leq a sup_j |u(i) - v(i)|$$

ή ισοδύναμα

$$\|T_a u - T_a v\| \leq a \|u - v\|.$$

Άμεση συνέπεια του παραπάνω θεωρήματος και του θεωρήματος Σταθερού Σημείου, λαμβάνουμε το ακόλουθο πόρισμα.

Πόρισμα ΙΙ

Η συνάρτηση V_a είναι η μοναδική λύση της εξίσωσης:

$$V_a(i) = \min_{a\in A(i)} \left\{ C(i,a) + a\sum_{j=0}^{\infty} p_{ij}(a)V_a(j) \right\}, i \geq 0.$$

Επιπλέον, για κάθε συνάρτηση $u \in B(i)$ $T_a^n u \to V_a$, καθώς $n \to \infty$.

Παρατηρήσεις:

i. Μια εξαιρετικά χρήσιμη επιλογή για τη συνάρτηση u είναι να υποθέσουμε ως u τη μηδενική συνάρτηση, δηλαδή $u_i = 0$ $i \geq 0$. Τότε αν θεωρήσουμε $V_{n,a}(i) = (T_a^n 0)(i)$, προκύπτει ότι η ποσότητα $V_{n,a}(i)$ ισοδυναμεί με το ελάχιστο συνολικό αποπληθωρισμένο κόστος για το πρόβλημα του πεπερασμένου χρονικού ορίζοντα μήκους n. Συχνά σημαντικά αποτελέσματα που αφορούν την ποσότητα $V_a(i)$ μπορούν να αποδειχτούν, αφού πρώτα αποδειχθεί η ισχύ τους για την ποσότητα $V_{n,a}(i)$ και έπειτα πάρουμε το όριο για $n \to \infty$.

ii. Το πόρισμα μας επιτρέπει να δείξουμε ότι στον αλγόριθμο βελτίωσης των πολιτικών, δύο πράγματα είναι δυνατά. Πρώτον, η νέα πολιτική είναι «γνήσια» καλύτερη από την προηγούμενη και δεύτερον και οι δυο πολιτικές είναι βέλτιστες. Αυτός ο ισχυρισμός αποδεικνύεται ως εξής:

Αν $V_{f^*}(i) = V_f(i), i \geq 0$ τότε από την σχέση:

$$C(i, f^*(i)) + a\sum_{j=0}^{\infty} p_{ij}\left(f^*(i)V_f(j)\right) = \min_{a\in A(i)} \left\{ C(i,a) + a\sum_{j=0}^{\infty} p_{ij}(a)V_f(j) \right\}, i \geq 0$$

Και από την (ii) περίπτωση του Λήμματος Ι έπεται ότι η V_f ικανοποιεί το δεξιό μέλος της εξίσωσης βελτιστοποίησης (2.37) και συνεπώς από το πόρισμα ΙΙ έχουμε $V_f(i) = V_a(i)$ $i \geq 0$.

iii. Η συνάρτηση T_f είναι συστολή.

Πράγματι,

$$\|T_f u - T_f v\| = sup_{i\geq 0}\left|(T_f u)(i) - (T_f v)(i)\right| = a sup_{i\geq 0}\left|\sum_{j=0}^{\infty} p_{ij}(f(i))[u(j) - v(j)]\right| \leq a\|u - v\|$$

Για κάθε $u, v \in B(I)$. Αρά η T_f είναι μία συστολή. Επομένως από το Λήμμα Ι (ii) προκύπτει ότι η V_f είναι η μοναδική λύση της εξίσωσης:

$$V_f(i) = C(i, f(i)) + a\sum_{j=0}^{\infty} p_{ij}(f(i))V_f(j) \ \ i \geq 0.$$

iv. Αν το σύνολο των καταστάσεων της διαδικασίας είναι πεπερασμένο, ο αλγόριθμος βελτίωσης των πολιτικών λειτουργεί ως εξής: Επιλέγουμε μία αρχική στάσιμη πολιτική f. Έπειτα υπολογίζουμε τις ποσότητες $V_f(i)$ $i = 1, \ldots, n$ λύνοντας το παρακάτω γραμμικό σύστημα n εξισωσεων με n αγνώστους:

$$V_f(i) = C\big(i, f(i)\big) + a \sum_{j=0}^{\infty} p_{ij}(f(i)) V_f(j) \;\; i = 1, .., n$$

Στη συνέχεια βρίσκουμε τη στάσιμη πολιτική f^* η οποία, όταν η διαδικασία βρίσκεται στην κατασταση i επιλέγει την ενέργεια $a \in A(i)$ που ελαχιστοποιεί την παράσταση $C(i, a) + a \sum_{j=0}^{\infty} p_{ij}(a) V_f(j)$. Μετά από αυτά βρίσκουμε τις ποσότητες $V_{f^*}(i), i = 1, \ldots, n$ για να βελτιώσουμε την πολιτική f^* κ.ο.κ. Επειδή ο αριθμός των στάσιμων πολιτικών είναι πεπερασμένος, διότι ο αριθμός των καταστάσεων είναι πεπερασμένος, τελικά θα φτάσουμε σε μια στάσιμη πολιτική για την οποία δεν υπάρχει «γνήσια» βελτίωση. Αυτή είναι η $\alpha -$ βέλτιστη πολιτική.

Αξίζει να σημειωθεί ότι, όλα τα παραπάνω αποτελέσματα ισχύουν στην περίπτωση όπου, αντί της ελαχιστοποίησης του συνολικού αναμενομένου αποπληθωρισμένου κόστους, μας ενδιαφέρει η μεγιστοποίηση του συνολικού αποπληθωρισμένου κέρδους. Στην προκείμενη περίπτωση η αντίστοιχη εξίσωση βελτιστοποίησης είναι:

$$V_a(i) = \max_{a \in A(i)} \left\{ R(i, a) + a \sum_{j=0}^{\infty} p_{ij}(a) V_a(j) \right\}, i \geq 0$$

όπου $R(i, a)$ είναι το αναμενόμενο κέρδος, όταν βρισκόμαστε στην κατασταση i και λάβουμε την απόφαση a.

Εφαρμογή:

Θεωρούμε ένα μηχάνημα το οποίο μπορεί να βρίσκεται σε οποιαδήποτε από τις καταστάσεις 0,1,2…Υποθέτουμε ότι στην αρχή κάθε μέρας η κατασταση του μηχανήματος παρατηρείται και αποφασίζεται αν θα αντικατασταθεί ή αν δεν θα αντικατασταθεί. Αν επιλεχθεί να αντικατασταθεί, τότε υποθέτουμε ότι το μηχάνημα αντικαθίσταται αμέσως από ένα καινούριο. Η κατασταση του καινούριου μηχανήματος είναι η κατασταση 0 .Το κόστος αντικατάστασης του μηχανήματος είναι ίσο με R. Επιπλέον υποθέτουμε ότι υπάρχει ένα κόστος συντήρησης ίσο με $C(i)$ για κάθε μέρα που το μηχάνημα είναι στην κατασταση i.Επισης έστω P_{ij} η πιθανότητα ένα μηχάνημα που βρίσκεται στην κατασταση i στην αρχή κάποιας μέρας, να μεταβεί στην κατασταση j στην αρχή της επόμενης μέρας. Έχουμε ακόμα τις ακόλουθες υποθέσεις για το κόστος συντήρησης και τις πιθανότητες μετάβασης:

i. Η συνάρτηση $C(i), i \geq 0$ είναι φραγμένη και αύξουσα.

ii. Η έκφραση $\sum_{j=\kappa}^{\infty} P_{ij}$ είναι αύξουσα ως προς i για κάθε $\kappa \geq 0$

Η υπόθεση i. σημαίνει ότι το κόστος συντήρησης είναι μια αύξουσα συνάρτηση της κατάστασης. Η υπόθεση ii. σημαίνει ότι η πιθανότητα μετάβασης σε οποιοδήποτε σύνολο τιμών $\{\kappa, \kappa + 1, ..\}$ είναι μια αύξουσα συνάρτηση της παρούσας κατάστασης. Σε αυτό το μοντέλο οι ενέργειες αντιστοιχούν στις καταστάσεις 1,2,.. είναι η ενέργεια 1,που σημαίνει αντικατάσταση και η ενέργεια 2 που σημαίνει ¨μη αντικατάσταση¨. Στην κατασταση 0 αντιστοιχεί μόνο η ενέργεια 2. Από την περιγραφή του μοντέλου προκύπτουν τις ακόλουθες εκφράσεις για τις ποσότητες $C(i,a), P_{ij}(a)$:

$$C(i,1) = R + C(0)\,, C(i,2) = C(i), i \geq 1$$

$$C(0,2) = C(0), P_{ij}(1) = P_{0j}\,, P_{ij}(2) = P_{ij}\,, i \geq 1$$

$$P_{0j}(2) = P_{0j}$$

Το ακόλουθο λήμμα βασιζόμενο στην υπόθεση ii.,του οποίου η απόδειξη παραλείπεται, θα μας βοηθήσει για την εύρεση της μορφής της βέλτιστης πολιτικής.

Λήμμα II

Για οποιαδήποτε αύξουσα συνάρτηση $h(i)$, η συνάρτηση $\sum_{j=0}^{\infty} P_{ij}h(j)$ είναι επίσης αύξουσα ως προς i.

Για την απόδειξη του παρακάτω λήμματος θα χρειαστούμε την παρατήρηση i.

Λήμμα III

Αν οι υποθέσεις i. και ii. ισχύουν, τότε η ποσότητα $V_a(i), i \geq 0$ είναι αύξουσα.

Απόδειξη: Αρχικά σκοπός μας είναι να δείξουμε $V_{n,a}(i), i \geq 0$ είναι αύξουσα.

Η απόδειξη θα γίνει επαγωγικά ως προς n. Για $n = 1$, έχουμε

$V_{1,a}(i) = min\{R + C(0), C(i)\}, i \geq 1$ και $V_{1,a}(0) = C(0)$

Αφού $C(i), i \geq 0$ αύξουσα, έχουμε ότι $V_{1,a}(i), i \geq 0$ είναι αύξουσα. Έστω ότι $V_{n-1,a}(i), i \geq 0$ αύξουσα τότε

$V_{n,a}(i) = min\left\{R + C(0) + a\sum_{j=0}^{\infty} P_{0j}V_{n-1,a}(j)\,, C(i) + a\sum_{j=0}^{\infty} P_{ij}V_{n-1,a}(j)\right\}, i \geq 1$

$V_{n,a}(0) = C(0) + a\sum_{j=0}^{\infty} P_{0j}V_{n-1,a}(j)$

Από την επαγωγική υπόθεση και το λήμμα προκύπτει ότι $V_{n,a}(i), i \geq 0$ είναι αύξουσα ως προς i. Όμως $\lim_{n\to\infty} V_{n,a}(i) = V_a(i), i \geq 0$. Αρά $V_a(i), i \geq 0$ είναι αύξουσα.

Από το παρακάτω θεώρημα θα δοθεί η μορφή της βέλτιστης πολιτικής.

Θεώρημα XI

Υπάρχει ένας μη αρνητικός ακέραιος $i^*, i^* \leq \infty$, τέτοιος ώστε η $\alpha - \beta\varepsilon\lambda\tau\iota\sigma\tau\eta$ πολιτική αντικαθιστά το μηχάνημα όταν η κατάσταση του μηχανήματος είναι μεγαλύτερη του i^*.

Απόδειξη:

Η εξίσωση βελτιστοποίησης είναι:

$V_a(i) = min\left\{R + C(0) + a\sum_{j=0}^{\infty} P_{0j}V_a(j)\,, C(i) + a\sum_{j=0}^{\infty} P_{ij}V_a(j)\right\}, i \geq 0$

Έστω $i^* = max\left\{i: C(i) + a\sum_{j=0}^{\infty} P_{ij}V_a(j) \leq R + C(0) + a\sum_{j=0}^{\infty} P_{0j}V_a(j)\right\}$

Συνολικά από τα δυο προηγούμενα λήμματα συνεπάγεται ότι $C(i) + a\sum_{j=0}^{\infty} P_{ij}V_a(j)$ είναι αύξουσα ως προς $, i$.

Επομένως:

$$V_a(i) = \begin{cases} C(i) + a\sum_{j=0}^{\infty} P_{ij}V_a(j)\,, i \le i^* \\ R + C(0) + a\sum_{j=0}^{\infty} P_{0j}V_a(j)\,, i > i^* \end{cases}$$

2.8 Μέσο αναμενόμενο κόστος σε μοντέλα αποφάσεων άπειρου χρονικού ορίζοντα

Το μακροπρόθεσμο αναμενόμενο μέσο κόστος $g(i,\pi)$ ανά μονάδα χρόνου, υπό τον έλεγχο της πολιτικής π,δοθέντος ότι η αρχική κατάσταση της διαδικασίας είναι η κατασταση i, ορίζεται ως εξής:

$$g(i,\pi) = \lim_{n\to\infty} \frac{E_\pi\left[\sum_{J=0}^{n-1} C(X_j, a_j)\middle| X_0 = i\right]}{n}$$

Για κάθε κατάσταση i της διαδικασίας, μια πολιτική π^*είναι βέλτιστη, αν:

$$g(i,\pi^*) = \min_{\pi} g(i,\pi)\,, i \ge 0$$

Σε αντίθεση με το πρόβλημα της ελαχιστοποίησης του συνολικού αναμενόμενου αποπληθωρισμένου κόστους, στο πρόβλημα της ελαχιστοποίησης του μακροπρόθεσμου αναμενόμενου μέσου κόστους ανά μονάδα χρόνου η ύπαρξη μια βέλτιστης πολιτικής δεν είναι βέβαιη. Στον έλεγχο μιας στοχαστικής διαδικασίας, αντιμετωπίζουμε προβλήματα όπου η βέλτιστη πολιτική είτε δεν υπάρχει είτε και αν ακόμη υπάρχει δεν είναι μια στάσιμη πολιτική. Στο παρόν κεφάλαιο υποθέτουμε ότι υπάρχει ένας θετικός αριθμός M τέτοιος ώστε για κάθε ενέργεια a και κάθε κατάσταση i της διαδικασίας, ισχύει ότι: $|C(i,a)| < M$.Το θεώρημα XII που παρουσιάζεται παρακάτω παρέχει μια ικανή συνθήκη για την ύπαρξη μίας βέλτιστης στάσιμης πολιτικής.

Θεώρημα XII

Έστω ότι υπάρχει μία άνω φραγμένη ακολουθία αριθμών $\{h_i\}, i \ge 0$ και μια σταθερά g έτσι ώστε:

$$h_i = \min_{\alpha\in A(i)}\left\{C(i,\alpha) - g + \sum_{j=0}^{\infty} p_{ij}(a)h_j\right\}, i \ge 0 \qquad (2.40)$$

Τότε υπάρχει μια βέλτιστη πολιτική $f = \{f_i\}, i \ge 0$, η οποία, όταν η διαδικασία βρίσκεται στην κατάσταση i, επιλέγει εκείνη την ενέργεια $a = f_i$ που ελαχιστοποιεί το δεξιό μέλος της εξίσωσης (2.40). Επιπλέον η σταθερά g είναι ίση με $g(i,f), i \ge 0$, όπου $f = \{f_i\}, i \ge 0$.

Γνωρίζουμε ότι η εξίσωση (2.40) είναι γνωστή ως εξίσωση βελτιστοποίησης για το πρόβλημα της ελαχιστοποίησης του μακροπρόθεσμου αναμενόμενου μέσου κόστους ανά μονάδα χρόνου. Οι τιμές $\{h_i\}, i \ge 0$, είναι γνωστές ως οι σχετικές τιμές της βέλτιστης στάσιμης πολιτικής. Για κάθε κατάσταση i της διαδικασίας, το παρακάτω θεώρημα παρέχει μια συνθήκη η οποία εγγυάται την ύπαρξη μιας ακολουθίας των σχετικών τιμών $\{h_i\}$ μέσω της συνάρτησης $V_a(i), i \geq 0$.

Θεώρημα XIII

Έστω ότι για μια κατάσταση της διαδικασίας, για παράδειγμα την κατάσταση 0, υπάρχει μια σταθερά B τέτοια ώστε :$|V_a(i) - V_a(0)| < B$, για κάθε $i \geq 0$ και κάθε $a \in (0,1)$.Τότε:

i. Υπάρχει μια άνω φραγμένη ακολουθία αριθμών $\{h_i\}, i \geq 0$, και μια σταθερά g που ικανοποιούν την εξίσωση (2.40).

ii. Υπάρχει μια ακολουθία αριθμών $\{a_n\}$ τέτοια ώστε $\lim_{n\to\infty} a_n = 1$ για την οποία ισχύει ότι:

$$h_i = \lim_{n\to\infty}\{V_{a_n}(i) - V_{a_n}(0)\} \text{ και}$$

iii. $\lim_{a\to 1}(1-a)V_a(0) = g$.

Η απόδειξη του Θεωρήματος XIII βρίσκεται στο βιβλίο Ross S.M(1992). Από τα ii., του Θεωρήματος XIII προκύπτει ότι η συνάρτηση $h(i)$ λαμβάνει τη δομή της συνάρτησης $V_a(i)$. Για παράδειγμα, αν η συνάρτηση $V_a(i)$ είναι αύξουσα (ή κυρτή) τότε έπεται ότι το ίδιο συμβαίνει με την συνάρτηση $h(i)$.

Ας θεωρήσουμε την εφαρμογή του προηγούμενου κεφαλαίου, στο οποίο είχαμε:

$$V_a(0) = C(0) + a\sum_{j=0}^{\infty} P_{0j}V_a(j) \text{ και}$$

$$V_a(i) \leq R + C(0) + a\sum_{j=0}^{\infty} P_{0j}V_a(j) = R + V_a(0).$$

Επειδή $V_a(i)$ αύξουσα ως προς i προκύπτει ότι:

$$|V_a(i) - V_a(0)| \leq R.$$

Συνεπώς, υπάρχει μια σταθερά g και μια αύξουσα συνάρτηση h_i τέτοια ώστε:

$$g + h_i = \min\left\{R + C(0) + \sum_{j=0}^{\infty} P_{0j}h_j\,, C(i) + \sum_{j=0}^{\infty} P_{ij}h_j\right\}, i \geq 0$$

Και η πολιτική που επιλέγει εκείνες τις ενέργειες που ελαχιστοποιούν το δεξιό μέλος της παραπάνω σχέσης ,σε κάθε κατάσταση i της διαδικασίας είναι η βέλτιστη πολιτική ως προς την ελαχιστοποίηση του μακροπρόθεσμου αναμενόμενου μέσου κόστους ανά μονάδα χρόνου. Αν θέσουμε:

$$i^* = \max\left\{i | C(i) + \sum_{j=0}^{\infty} P_{ij}h_j \leq R + C(0) + \sum_{j=0}^{\infty} P_{0j}h_j\right\}$$

τότε η βέλτιστη πολιτική έχει την ίδια δομή με την $\alpha - \beta έλτιστη$ πολιτική. Στο πρόβλημα της ελαχιστοποίησης του μακροπρόθεσμου αναμενόμενου μέσου κόστους ανά μονάδα χρόνου, η ακόλουθη υπόθεση χρειάζεται να εισαχθεί έτσι ώστε να είναι εφικτός ο υπολογισμός μίας βέλτιστης στάσιμης πολιτικής.

Υπόθεση: Για κάθε στάσιμη πολιτική f υπάρχει μία κατάσταση r η οποία είναι τέτοια ώστε, ο αναμενόμενος χρόνος και το αναμενόμενο κόστος που απαιτούνται για την μετάβαση στην κατασταση r από οποιαδήποτε κατάσταση της διαδικασίας, υπό τον έλεγχο της πολιτικής f, είναι πεπερασμένο.

Η παραπάνω Υπόθεση εξασφαλίζει επίσης ότι το μακροπρόθεσμο αναμενόμενο μέσο κόστος ανά μονάδα χρόνου οποιασδήποτε στάσιμης πολιτικής που υιοθετείται για τον έλεγχο της διαδικασίας είναι ανεξάρτητο της αρχικής κατάστασης της διαδικασίας.

Όπως στο πρόβλημα της ελαχιστοποίησης του συνολικού αναμενόμενου αποπληθωρισμένου κόστους έτσι και στο πρόβλημα της ελαχιστοποίησης του μακροπρόθεσμου αναμενόμενου μέσου κόστους ανά μονάδα χρόνου, ο αλγόριθμος βελτίωσης των πολιτικών και η μέθοδος των διαδοχικών προσεγγίσεων αποτελούν τις βασικές υπολογιστικές τεχνικές για την εύρεση της βέλτιστης πολιτικής.

Θεώρημα XIV

Έστω ότι υπάρχει μια κατασταση r η οποία ικανοποιεί την Υπόθεση. Έστω g_f το μακροπρόθεσμο αναμενόμενο μέσο κόστος ανά μονάδα χρόνου, υπό τον έλεγχο της στάσιμης πολιτικής $f = \{f_i\}, i \geq 0$ και $h_i = K_i(f) - g_f T_i(f), i \geq 0$, όπου $T_i(f)$ και $K_i(f)$ είναι ο αναμενόμενος χρόνος και το αναμενόμενο κόστος, αντίστοιχα που απαιτούνται μέχρι η διαδικασία να επιστρέψει στην κατάσταση r, αν αρχικά βρισκόταν στην καταστασή i και η πολιτική f έχει υιοθετηθεί για τον έλεγχο της διαδικασίας.

Τότε οι ποσότητες $h_i(f), i \geq 0$ και g_f είναι η μοναδική λύση του ακόλουθου συστήματος των γραμμικών εξισώσεων με αγνώστους g και $h_i, i \geq 0$:

$$h_i = C(i, f_i) - g + \sum_{j=0}^{\infty} P_{ij}(f_i) h_j \, , i \geq 0$$

$$h_r = 0 \qquad (2.41)$$

Η εξίσωση (2.41) είναι γνώστη ως εξίσωση κανονικοποίησης.

Θεώρημα XV

Έστω ότι υπάρχει μια κατασταση r η οποία ικανοποιεί την Υπόθεση που αναφέραμε λίγο πιο πάνω. Έστω g_f και $g_{\tilde{f}}$ τα μακροπρόθεσμα αναμενόμενα μέσα κόστη ανά μονάδα χρόνου, υπό τον έλεγχο των στάσιμων πολιτικών $f = \{f_i\}$ και $\tilde{f} = \{\tilde{f_i}\}, i \geq 0$, αντίστοιχα. Ας υποθέσουμε ότι:

$$C(i, \tilde{f}_i) - g_f + \sum_{j=0}^{\infty} P_{ij}(\tilde{f}) h_j(f) \leq h_i(f), i \geq 0,$$

Όπου, οι ποσότητες $h_i(f), i \geq 0$, έχουν οριστεί μέσω της εξίσωσης:

$$h_i = k_i(f) - g_f T_i(f), i \geq 0.$$

Τότε: $g_{\tilde{f}} \leq g_f$.

Οι ποσότητες $h_i(f), i \geq 0$ είναι γνωστές ως οι σχετικές τιμές της πολιτικής f.

Έπειτα, περιγράφουμε σε βήματα, τον αλγόριθμο βελτίωσης των πολιτικών για το κριτήριο της ελαχιστοποίησης του μακροπρόθεσμου αναμενόμενου μέσου κόστους ανά μονάδα χρόνου.

Αλγόριθμος βελτίωσης των πολιτικών

Βήμα 0: Επιλέγουμε μία στάσιμη πολιτική R.

Βήμα1: Για αυτήν την πολιτική, υπολογίζουμε τη μοναδική λύση $\{g(R), h_i(R)\}$ του ακόλουθου συστήματος γραμμικών εξισώσεων:

$$h_i(R) = C(i, R_i) - g + \sum_{j=0}^{\infty} P_{ij}(R_i) h_j(R)$$

$$h_s(R) = 0$$

Όπου, s είναι μια αυθαίρετη επιλεγμένη κατάσταση της διαδικασίας.

Βήμα2: Για κάθε κατασταση i της διαδικασίας ,επιλέγουμε την ενέργεια a_i που επιτυγχάνει το ελάχιστο της παράστασης:

$$\min_{a \epsilon A(i)} \left\{ C(i, a) - g(R) + \sum\nolimits_{j=0}^{\infty} P_{ij}(a) h_j(R) \right\}.$$

Η νέα στάσιμη πολιτική $\bar{R}$ λαμβάνεται διαλέγοντας $\overline{R_\iota} = a_i$ για κάθε κατασταση i της διαδικασίας με την προϋπόθεση ότι η ενέργεια $\overline{R_\iota}$ επιλέγεται ίδια με την ενέργεια της προηγούμενης πολιτικής R_i όταν η ενέργεια αυτή ελαχιστοποιεί την ποσότητα ανάμεσα στα άγκιστρα.

Βήμα 3: Αν η καινούρια πολιτική $\bar{R}$ είναι ίδια με την προηγούμενη πολιτική R ο αλγόριθμος σταματά. Διαφορετικά επιστρέφουμε στο Βήμα 1 με την πολιτική R να αντικαθίσταται από την πολιτική $\bar{R}$.

Γνωρίζουμε ότι ο αλγόριθμος βελτίωσης των πολιτικών είναι ένας ευσταθής αλγόριθμος, ο οποίος συγκλίνει ,μετά από έναν πεπερασμένο αριθμό επαναλήψεων σε μία βέλτιστη πολιτική. Ο αριθμός των επαναλήψεων του αλγορίθμου είναι ανεξάρτητος του αριθμού των καταστάσεων της διαδικασίας και συνήθως κυμαίνεται μεταξύ 3 και 15.Επιπλεον το μακροπρόθεσμο αναμενόμενο μέσο κόστος ανά μονάδα χρόνου των πολιτικών που παράγονται από τις επαναλήψεις του αλγορίθμου επιτυγχάνει σημαντική βελτίωση στην τιμή του στις πρώτες επαναλήψεις του αλγορίθμου συγκλίνοντας στο ελάχιστο μακροπρόθεσμο αναμενόμενο μέσο κόστος ανά μονάδα χρόνου.

Η μέθοδος διαδοχικών προσεγγίσεων

Ο αλγόριθμος βελτίωσης των πολιτικών απαιτεί σε κάθε επανάληψη του, την επίλυση ενός συστήματος γραμμικών εξισώσεων με διάσταση ίδια με τη διάσταση του χώρου καταστάσεων της διαδικασίας για την οποία επιλύεται. Το γεγονός αυτό επιβαρύνει υπολογιστικά τους αλγορίθμους βελτίωσης των πολιτικών και τους κάνει λιγότερο ελκυστικούς σε προβλήματα Μαρκοβιανών διαδικασιών αποφάσεων με μεγάλους χώρους καταστάσεων. Ένας διαφορετικός αλγόριθμος αποφεύγοντας την επίλυση συστημάτων γραμμικών εξισώσεων, χρησιμοποιεί έναν αναδρομικό τρόπο επίλυσης των προβλημάτων και βασίζεται στον δυναμικό προγραμματισμό. Ο αλγόριθμος στηρίζεται στη μέθοδο των διαδοχικών προσεγγίσεων και υπολογίζει αναδρομικά μια ακολουθία τιμών συνάρτησης η οποία προσεγγίζει την τιμή του ελαχίστου μακροπρόθεσμου αναμενόμενου μέσου κόστους ανά μονάδα χρόνου. Αυτές οι συναρτήσεις παρέχουν άνω και κάτω φράγματα στην τιμή του ελάχιστου μακροπρόθεσμου αναμενόμενου μέσου κόστους και υπό μια μία συγκεκριμένη συνθήκη απεριοδικότητας της διαδικασίας, αυτά τα φράγματα προσεγγίζουν την τιμή του ελαχίστου μακροπρόθεσμου αναμενόμενου μέσου κόστους. Η μέθοδος των διαδοχικών προσεγγίσεων αποτελεί την καλύτερη υπολογιστική μέθοδο για την επίλυση προβλημάτων Μαρκοβιανών διαδικασιών αποφάσεων με μεγάλους χώρους καταστάσεων. Ο αλγόριθμος υπολογίζει αναδρομικά για $n = 1,2, . . ,$τη συνάρτηση $V_n(i)$ από τον τύπο:

$$V_n(i) = \min_{\alpha \in A(i)} \left\{ C(i,a) + \sum\nolimits_{j=0} p_{ij}(a) V_{n-1}(j) \right\}, i \geq 0.$$

Η ποσότητα $V_n(i)$ μπορεί να θεωρηθεί ως το ελάχιστο συνολικό αναμενόμενο κόστος όταν απομένουν n χρονικές περίοδοι για το τέλος της διαδικασίας. Η παρούσα κατάσταση είναι η κατασταση i και το τελικό κόστος είναι ίσο με $V_0(j)$ όταν η διαδικασία καταλήξει στην κατάσταση j. Σύμφωνα με την μέθοδο των διαδοχικών προσεγγίσεων για μεγάλες τιμές του n οι διάφορες $V_n(i) - V_{n-1}(i)$ θα είναι πολύ κοντά στην τιμή του ελάχιστου μακροπρόθεσμου αναμενόμενο μέσου κόστους ανά μονάδα χρόνου. Η στάσιμη πολιτική της οποίας οι ενέργειες ελαχιστοποιούν την παράσταση ανάμεσα στα άγκιστρα για κάθε κατασταση i της διαδικασίας στην παραπάνω εξίσωση, θα είναι η βέλτιστη. Παρακάτω περιγράφονται τα βήματα της μεθόδου των διαδοχικών προσεγγίσεων

Αλγόριθμος των διαδοχικών προσεγγίσεων

Βήμα 0: Αρχικά επιλέγουμε $V_0(i)$ με $0 \leq V_0(i) \leq \min_a C(i,a)$, για κάθε κατάσταση i της διαδικασίας.

Βήμα1: Υπολογίζουμε τη συνάρτηση $V_n(i)$ για κάθε κατάσταση i της διαδικασίας, από τον τύπο:

$$V_n(i) = \min_{\alpha \in A(i)} \left\{ C(i,a) + \sum_{j=0} p_{ij}(a) V_{n-1}(j) \right\}, i \geq 0$$

Και θέτουμε την πολιτική της οποίας οι ενέργειες ελαχιστοποιούν το δεξιό μέλος της συνάρτησης

$V_n(i), i \geq 0.$

Βήμα 2: Υπολογίζουμε τα φράγματα:

$M_n = \max_j \{V_n(j) - V_{n-1}(j)\}$ και $m_n = \min_j \{V_n(j) - V_{n-1}(j)\}, n = 1,2, \ldots$

Ο αλγόριθμος τελειώνει με την πολιτική $R(n)$, όταν $0 \leq M_n - m_n \leq \varepsilon \cdot m_n,$

Όπου ε είναι ένας προκαθορισμένος πολύ μικρός αριθμός.

Διαφορετικά, πηγαίνουμε στο Βήμα 3.

Βήμα 3: Θέτουμε $n \coloneqq n + 1$ και μεταβαίνουμε στο Βήμα 1.

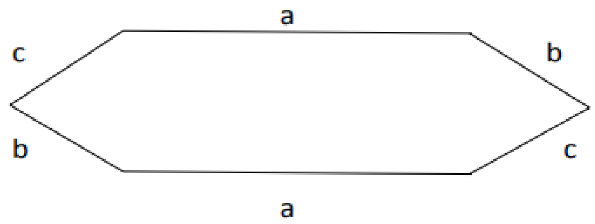
Ο αλγόριθμος των διαδοχικών προσεγγίσεων δεν έχει την ίδια ευστάθεια του αλγορίθμου βελτίωσης των πολιτικών. Ο αριθμός των επαναλήψεων που απαιτούνται από τη μέθοδο των διαδοχικών προσεγγίσεων για τη σύγκλιση του αλγορίθμου εξαρτάται από το πρόβλημα που εξετάζεται και συνήθως αυξάνεται όσο αυξάνεται ο αριθμός των καταστάσεων της διαδικασίας. Το όριο ανοχής ε και η συνάρτηση $V_0(i), i \geq 0$, επηρεάζει τον αριθμό των απαιτούμενων επαναλήψεων του αλγορίθμου.

Κεφάλαιο 3º Εφαρμογές των Μαρκοβιανών Διαδικασιών Αποφάσεων

3.1 Ενδεικτικά παραδείγματα των Μαρκοβιανών Διαδικασιών Αποφάσεων που αφορούν πεπερασμένο χρονικό ορίζοντα

Παράδειγμα 1 (ρύθμιση μηχανήματος κοπής)

Σε ένα εργοστάσιο κοπής μετάλλων, μία από τις μηχανές παράγει εξάγωνα πλαίσια της παρακάτω μορφής:



Η εν λόγω μηχανή αποτελείται από τρία ζεύγη κοπτήρων, καταχρηστικά ονομαζόμενα ως a,b,c και τα οποία χρησιμοποιούνται για την κοπή των πλευρών a,b,c του παραπάνω σχήματος αντίστοιχα. Δυστυχώς κάθε ζεύγος μπορεί να απορυθμίζεται από την κανονική του θέση κατά την διάρκεια κοπής ενός πλαισίου, αλλά κατόπιν υπάρχει η δυνατότητα επαναρρύθμισής του. Συγκεκριμένα, μετά από κάθε κοπή μπορούμε να ρυθμίζουμε την μηχανή ώστε όλα τα ζεύγη κοπτήρων να έχουν την κανονική τους θέση-διαδικασία που συνεπάγεται ένα κόστος 10 χρηματικών μονάδων. Αντιθέτως, αν κοπεί ένα πλαίσιο όταν k ζεύγη κοπτήρων δεν έχουν την κανονική τους θέση, προκύπτει ένα κόστος 4k χρηματικών μονάδων.

Έστω ότι οι δυνατές καταστάσεις στις οποίες μπορεί να βρεθεί η μηχανή είναι οι εξής πέντε:

1. και τα τρία ζεύγη βρίσκονται στην σωστή θέση,
2. μόνο τα ζεύγη b,c βρίσκονται στην σωστή θέση,
3. Μόνο το ζεύγος b βρίσκεται στην σωστή θέση,
4. Μόνο το ζεύγος c βρίσκεται στην σωστή θέση,
5. Κανένα ζεύγος δεν βρίσκεται στην σωστή θέση,

Ενώ ο πίνακας πιθανοτήτων μετάβασης ($1^{ης}$ τάξης) της μηχανής από μια κατάσταση σε μία άλλη είναι ο παρακάτω:

$$\mathbb{Q} = (q_{ij}) = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} \begin{array}{c} \begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} \\ \begin{pmatrix} 3/4 & 1/4 & 0 & 0 & 0 \\ 0 & 1/2 & 1/4 & 1/4 & 0 \\ 0 & 0 & 3/4 & 0 & 1/4 \\ 0 & 0 & 0 & 1/2 & 1/2 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \end{array}.$$

Ο στόχος είναι να βρεθεί μία πολιτική ρύθμισης των κοπτήρων με την οποία ελαχιστοποιείται ο ρυθμός κόστους.

Λύση:

Το σύστημά μας περιγράφεται από μία Μαρκοβιανή αλυσίδα αποφάσεων όπως φαίνεται παρακάτω:

- X_n: η κατάσταση της μηχανής (ως προς την θέση των κοπτήρων) μόλις πριν την κοπή του $(n+1)$ −οστού πλαισίου.
- Χώρος καταστάσεων $S = \{1,2,3,4,5\}$: με στοιχεία τις πέντε καταστάσεις i όπως αυτές ορίσθηκαν στην εκφώνηση της άσκησης.
- Δυνατές αποφάσεις $\alpha = \begin{Bmatrix} 0, \textit{καμμία επέμβαση στους κοπτήρες} \\ 1, \textit{επαναρρύθμισή όλων των κοπτήρων} \end{Bmatrix}$, αρά χώρος αποφάσεων

$A = \{0,1\}$ και μάλιστα $A(1) = \{0\}, A(i) = \{0,1\} \forall\, i \in S - \{1\}$.

- Πιθανότητες μετάβασης 1ης τάξης:
  - $p_{ij}(0) = q_{ij} \forall i,j \in S$,
  - $p_{ij}(1) = q_{1j} \forall i,j \in S$ γιατί αν η διαδικασία βρίσκεται σε οποιαδήποτε κατάσταση i, τότε κάτω από την απόφαση $a = 1$, επιστρέφει στην κατάσταση 1, οπότε για να μεταβεί στην j έχει ξεκινήσει από την 1.
- Κόστη απόφασης(αναμενόμενα):
  - $C(1{,}0) = p_{11}(0) \cdot (4 \cdot 0) + p_{12}(0) \cdot (4 \cdot 1) = 1$ αν $X_n = 1$ τότε $\alpha = 0$ και $X_{n+1} = 1$

    ($k = 0$ ζεύγη απορυθμισμένα) με πιθανότητα $p_{11}(0)$ ή $X_{n+1} = 2$ ($k = 1$ ζεύγος απορυθμισμένο) με πιθανότητα $p_{12}(0)$.
  - $C(2{,}0) = p_{22}(0) \cdot (4 \cdot 1) + p_{23}(0) \cdot (4 \cdot 2) + p_{24}(0) \cdot (4 \cdot 2) = 6$.
  - $C(3{,}0) = q_{33} \cdot (4 \cdot 2) + q_{35} \cdot (4 \cdot 3) = 9$.
  - $C(4{,}0) = q_{44} \cdot (4 \cdot 2) + q_{45} \cdot (4 \cdot 3) = 10$.
  - $C(5{,}0) = q_{55} \cdot (4 \cdot 3) = 12$.
  - $C(i{,}1) = 10 + p_{i1}(1) \cdot (4 \cdot 0) + p_{i2}(1) \cdot (4 \cdot 1) = 10 + q_{12} \cdot 4 = 11$, $\forall\, i = 2,3,4,5$.

Το ζητούμενο είναι η ελαχιστοποίηση του $g(R) = \sum_{j \in S} C(j, R_j) \cdot \pi_j(R)$ και η εύρεση της αντίστοιχης πολιτικής R.

Παράδειγμα 2 (Μαρκοβιανή Διαδικασία Αποφάσεων)

Ας υποθέσουμε ότι έχουμε μία Μ.Δ.Α με χώρο καταστάσεων που αποτελείται μόνο από δύο καταστάσεις s_1, s_2 με αντίστοιχα σύνολα αποφάσεων $a_{1,1}, a_{1,2}, a_{1,3}$ και $\alpha_{2,1}$. Όταν $X_n = s_1$ και επιλέγεται η απόφαση $a_{1,1}$ τότε προκύπτει μία αμοιβή 5 μονάδων και η αλυσίδα μεταβαίνει την επόμενη χρονική στιγμή στην s_2 με πιθανότητα 0.5, ενώ παραμένει στην s_1 με πιθανότητα επίσης 0.5. Αν $X_n = s_1$ και επιλεγεί η $a_{1,2}$, τότε η διαδικασία καταλήγει με βεβαιότητα στην s_2 και η αμοιβή είναι 10 μονάδες. Αν επιλεγεί η $a_{1,3}$ δεν υπάρχει αμοιβή ή κόστος αλλά η αλυσίδα παραμένει στην s_1. Τέλος, αν $X_n = s_2$,τότε θα επιλεγεί αναγκαστικά η $\alpha_{2,1}$ και θα προκύψει κόστος μίας μονάδας, ενώ η διαδικασία θα παραμείνει σίγουρα στην s_2 .

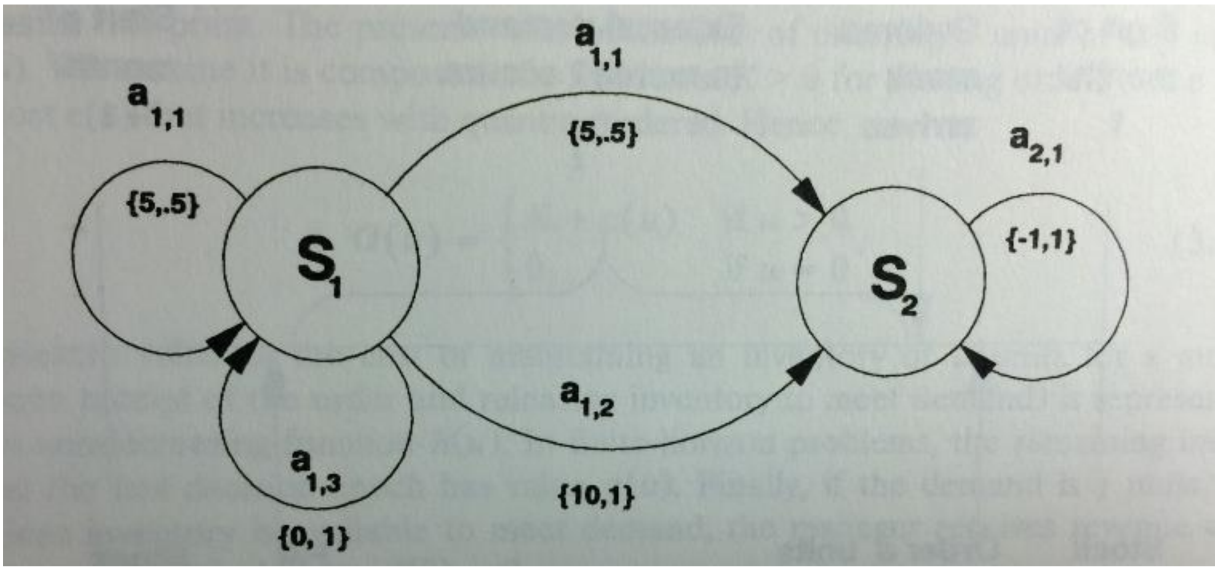
Λύση:

Εν συντομία η Μ.Δ.Α μοντελοποιείται ως εξής:

- Χώρος καταστάσεων $S = \{s_1, s_2\}$,
- Χώρος αποφάσεων $A = \{a_{1,1}, a_{1,2}, a_{1,3}, a_{2,1}\}$ και μάλιστα:
  - $A(s_1) = \{a_{1,1}, a_{1,2}, a_{1,3}\}$
  - $A(s_2) = \{a_{2,1}\}$
- Πιθανότητες μετάβασης 1ης τάξης $\mathbb{P}(a)$:
  - $p_{s_1 s_1}(a_{1,1}) = 0.5, p_{s_1 s_2}(a_{1,1}) = 0.5, p_{s_1 s_2}(a_{1,2}) = 1, p_{s_1 s_1}(a_{1,3}) = 1$
  - $p_{s_2 s_2}(a_{2,1}) = 1$
- Κόστη αποφάσεων:
  - $C(s_1, a_{1,1}) = +5$, $C(s_1 a_{1,2}) = +10$, $C(s_1, a_{1,3}) = 0$
  - $C(s_2, a_{2,1}) = -1$

Σημειώνουμε ότι εδώ, τόσο οι πιθανότητες μετάβασης όσο και τα κόστη δεν εξαρτώνται από το χρονικό βήμα. Επομένως, αυτή η αλυσίδα χαρακτηρίζεται ως χρονικά ομογενής.

Στο σχήμα 1.1 φαίνεται ο γράφος της συγκεκριμένης Μ.Δ.Α. Να σημειώσουμε ότι οι καταστάσεις βρίσκονται σε κύκλο, ενώ μετάβαση από την κατάσταση i στην j συμβολίζεται με καμπυλοειδές βέλος. Σε κάθε βέλος αντιστοιχεί ένα άγκιστρο του οποίου το πρώτο στοιχείο αντιπροσωπεύει την αμοιβή ή το κόστος της σχετικής μετάβασης, ενώ το δεύτερο στοιχείο είναι η πιθανότητα της μετάβασης αυτής.



Σχήμα 1.1 :Ο γράφος του παραδείγματος

Παράδειγμα 3 (επιδείνωσης ενός συστήματος)

Ένα μηχάνημα ελέγχεται στην αρχή κάθε ημέρας και κατηγοριοποιείται σε μια κατάσταση λειτουργίας i με $i = 1,2, \dots N$. Ο εξοπλισμός θα βρίσκεται σε μια από αυτές τις καταστάσεις $i = 1,2, \dots N$ όπου η κατάσταση i είναι καλύτερη από την κατάσταση $i + 1$. Ωστόσο υπάρχει και η δυνατότητα επισκευής του μηχανήματος. Αν η παρούσα κατάσταση του μηχανήματος είναι i και το μηχάνημα δεν επισκευάζεται , τότε στην αρχή της επόμενης ημέρας το μηχάνημα θα βρίσκεται στην κατάσταση j (που προφανώς δεν είναι καλύτερη από την κατάσταση i άρα j$\geq i$) με πιθανότητα q_{ij} $j \geq i$.Φυσικά $q_{ij} = 0$ για j< 1 (το μηχάνημα δεν μπορεί να βρεθεί σε καλύτερη κατάσταση την επόμενη μέρα εφόσον δεν επισκευάζεται) και $\sum_{j\geq 1} q_{ij} = 1$.

Η κατάσταση $i = N$ δηλώνει βλάβη του μηχανήματος, οπότε το μηχάνημα δεν μπορεί να λειτουργήσει αν δεν επισκευαστεί. Η αναγκαστική επισκευή απαιτεί δυο ημέρες, στο τέλος των οποίων το επισκευασμένο μηχάνημα επιστρέφει στην κατάσταση 1. Για τις υπόλοιπες καταστάσεις ($i = 2, .., N - 1$) υπάρχει η προληπτική επισκευή ή η επιλογή ότι το μηχάνημα λειτουργεί κανονικά. Η προληπτική επισκευή του μηχανήματος διαρκεί μια ημέρα και φέρνει το μηχάνημα πάλι στην κατάσταση 1. Προφανώς κατά την διάρκεια των ημερών επισκευής (ολικής ή μερικής) το μηχάνημα δεν λειτουργεί. Επιθυμούμε να προσδιορίσουμε έναν κανόνα συντήρησης που θα ελαχιστοποιεί το μακροπρόθεσμο μέσο κόστος επισκευής ανά ημέρα.

Λύση:

Αρχικά το σύστημα μπορεί να μοντελοποιηθεί και να περιγράφει από μια Μαρκοβιανή διαδικασία αποφάσεων όπως φαίνεται:

- Ορίζουμε X_n: την κατάσταση του μηχανήματος στην αρχή της n-οστης ημέρας.
- Χώρος καταστάσεων S= $\{1, .., N, N + 1\}$ οπού η κατάσταση $N + 1$ συμβολίζει ότι το μηχάνημα βρίσκεται στην δεύτερη μέρα ολικής επισκευής του. Η συγκεκριμένη επισκευή άρχισε στην κατάσταση N.
- Δυνατές αποφάσεις $a = \begin{cases} 0 \ \textit{δεν γίνεται επισκευή} \\ 1 \ \textit{προληπτική επισκευή} \\ 2 \ \textit{αναγκαστική επισκευή} \end{cases}$

Το σύνολο των πιθανών αποφάσεων είναι:

- $A(1) = \{0\}$ δηλαδή στην κατάσταση 1 δεν επισκευάζω τον εξοπλισμό του μηχανήματος
- $A(i) = \{0,1\}$ με $i = 2, .., N - 1$ δυνατότητα επισκευής η όχι
- $A(N) = A(N + 1) = 2$ το μηχάνημα είναι στην κατάσταση της αναγκαστικής επισκευής.
  - Πιθανότητες μετάβασης 1ης τάξης P(R)
    - $P_{i1}(1) = 1$ για $1 < i < N$ σε περίπτωση επισκευής το μηχάνημα μεταφέρεται στην κατάσταση 1 υποχρεωτικά.
    - $P_{N,N+1}(2) = P_{N+1,1}(2) = 1$
    - $P_{ij}(0) = q_{ij}$ για $1 \leq i < N$ και $j \geq i$

- Αλλιώς $P_{ij}(a) = 0$

Το κόστος επισκευής ανάλογα με την απόφαση a δίνεται από την σχέση:

$c_i(0) = 0$,$c_i(1) = c_{P_i}$,$c_N(2) = c_f$,$c_{N+1}(2) = 0$,

οπού c_f είναι το κόστος αναγκαστικής επισκευής και c_{p_i} είναι το κόστος προληπτικής συντήρησης.

Ο γενικός πίνακας μετάβασης θα έχει τη μορφή

$$\mathbb{p}(R)= \begin{array}{c} \begin{array}{cccccc} 1 & 2 & 3 & \dots & N & N+1 \end{array} \\ \left(\begin{array}{cccccc} q_{11} & q_{12} & q_{13} & & q_{1N} & 0 \\ 1 & 0 & 0 & \cdots & 0 & 0 \\ 1 & 0 & 0 & & 0 & 0 \\ & \vdots & & \ddots & \vdots & \\ 1 & 0 & 0 & & 0 & 0 \\ 0 & 0 & 0 & \dots & 0 & 1 \\ 1 & 0 & 0 & & 0 & 0 \end{array}\right) \end{array} \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ \vdots \\ N-1 \\ N \\ N+1 \end{array}$$

Όπως φαίνεται όλες οι καταστάσεις επικοινωνούν μεταξύ τους, οπότε ο χώρος καταστάσεων είναι μια κλειστή κλάση. Ας βρούμε λοιπόν την στάσιμη κατανομή π. Για να συμβεί κάτι τέτοιο αρκεί να επιλύσουμε το σύστημα:

$$\begin{cases} \pi = \pi P \\ \sum_{i\in S} \pi(i) = 1 \end{cases} \Leftrightarrow \begin{cases} \pi_1 = \pi_1 q_{11} + \pi_2 + \pi_3 + \cdots + \pi_{N-1} + \pi_{N+1} \\ \pi_2 = \pi_1 q_{12} \\ \pi_3 = \pi_1 q_{13} \\ \vdots \\ \pi_N = \pi_1 q_{1N} \\ \pi_{N+1} = \pi_N \\ \pi_1 + \pi_2 + \pi_3 + \cdots + \pi_N + \pi_{N+1} = 1 \end{cases}$$

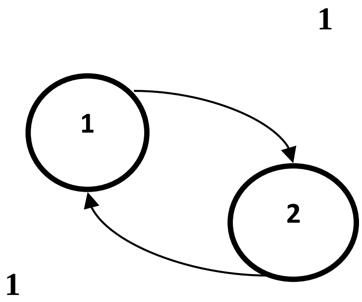
Συνεπώς $g(R) = \sum_{i\in S} C(i, R_i) \cdot \pi_i(R)$.

Παράδειγμα 4 (Ειδική περίπτωση σύγκλισης μιας Μαρκοβιανής Διαδικασίας Αποφάσεων)

Δίνεται η μαρκοβιανή διαδικασία αποφάσεων $\{X_n(R)\}_{n\in\mathbb{N}_0}$ με χώρο καταστάσεων S={1,2},χώρο δυνατών αποφάσεων A={a_0},κόστη αποφάσεων $C(1, a_0) = 1$, $C(2, a_0) = 0$ και πίνακα πιθανοτήτων μετάβασης πρώτης τάξης

$$\mathbb{P}(R) = \mathbb{P}(a_o) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

Λύση:



Σχήμα 1.2: Ο γράφος των δυνατών μεταβάσεων για την αλυσίδα κάτω από την μοναδική απόφαση a_0 του παραδείγματος 4

Στην συγκεκριμένη περίπτωση, η αλυσίδα έχει μία μοναδική κλειστή κλάση (για την ακρίβεια είναι μη-υποβιβάσιμη) κάτω από την μοναδική απόφαση a_0.

Τότε, ορίζουμε αυθαίρετα $V_n(i) = 0$, και

$$V_n(i) = \min_{a\in A(i)} \left\{ C(i,a) + \sum_{j\in S} p_{ij}(a)\cdot V_{n-1}(j) \right\}, i = 1,2$$

οπότε:

$$V_1(1) = C(1,a_0) + 0 = 1 \quad , V_1(2) = C(2,a_0) + 0 = 0$$

$$V_2(1) = C(1,a_0) + \sum_{j\in S} p_{ij}(a_0)\cdot V_1(j) = 1 + p_{12}(a_0)\cdot V_1(2) = 1$$

$$V_2(2) = C(2,a_0) + \sum_{j\in S} p_{2j}(a_0)\cdot V_1(j) = p_{21}(a_0)\cdot V_1(1) = 1$$

$$V_3(1) = C(1,a_0) + \sum_{j\in S} p_{1j}(a_0)\cdot V_2(j) = 1 + p_{12}(a_0)\cdot V_2(2) = 1 + 1 = 2$$

$$V_3(2) = 0 + p_{21}(a_0)\cdot V_2(1) = 1$$

$\vdots$

και τελικά $\forall k \in \mathbb{N}$ ισχύει:

$$V_{2k}(1) = k\,, k \geq 0 \qquad\qquad V_{2k}(2) = k\,, k \geq 0$$

$$V_{2k-1}(1) = k\,, k \geq 1 \qquad\qquad V_{2k-1}(2) = k-1\,, k \geq 1$$

και

$$m_n = \min_{j\in S}\{V_n(j) - V_{n-1}(j)\} = V_n(1) - V_{n-1}(1) = 0 \xrightarrow{n\to\infty} 0$$

$$M_n = \max_{j\in S}\{V_n(j) - V_{n-1}(j)\} = V_n(2) - V_{n-1}(2) = 1 \xrightarrow{n\to\infty} 1$$

και ως εκ τούτου οι ακολουθίες m_n , M_n δεν συγκλίνουν στο ίδιο όριο.

Στην προκείμενη περίπτωση, γίνεται εύκολα αντιληπτό ότι η αλυσίδα είναι περιοδική με περίοδο δύο (2) – αυτό είναι και ένα από τα πιο συνηθισμένα εμπόδια στα περισσότερα θεωρήματα συγκλίσεων στις στοχαστικές ανελίξεις.

Ακολουθεί ένα παράδειγμα που αφορά τον Γραμμικό Προγραμματισμό.

Παράδειγμα 5 (Ζήτηση ενός προϊόντος)

Η ζήτηση D ενός προϊόντος Π είναι τυχαία μεταβλητή με κατανομή

$$P(D=x)=\begin{cases}\frac{1}{6}, & x=0\\ \frac{1}{6}, & x=1\\ \frac{2}{3}, & x=2\end{cases}$$

Οι τ.μ. D_i είναι ανεξάρτητες για διαδοχικές μέρες i. Το θέμα που προκύπτει είναι το πόσο θα πρέπει να παραγγείλουμε από το προϊόν Π ώστε να καλύψουμε την ζήτηση. Η παραγγελία μας αυτή γίνεται στην αρχή κάθε ημέρας και η παράδοση των μονάδων που παραγγείλαμε γίνεται άμεσα (ακαριαία-ο χρόνος παράδοσης είναι αμελητέος). Το κόστος παραγγελίας κάθε μονάδας του Π είναι c χρηματικές μονάδες. Επίσης οι μονάδες που μένουν απούλητες καταστρέφονται ως παρωχημένες.

Ο στόχος μας είναι η ελαχιστοποίηση του μέσου ημερήσιου κόστους παραγγελίας με περιορισμό η ζήτηση να ικανοποιείται με ποσοστό (πιθανότητα) τουλάχιστον 1/3.

Λύση:

Έχουμε δύο τρόπους να αντιμετωπίσουμε το συγκεκριμένο πρόβλημα: ο $1^{ος}$ είναι ο αιτιοκρατικός (ντετερμινιστικός έλεγχος), ενώ ο $2^{ος}$ είναι ο στοχαστικός (τυχαιοποιημένος έλεγχος). Στα παρακάτω θεωρούμε ως S το ημερήσιο απόθεμα μας για την κάλυψη της ζήτησης. Στο συγκεκριμένο παράδειγμα αυτό ισοδυναμεί με την ποσότητα της παραγγελίας που έχουμε κάνει.

$1^{ος}$ τρόπος (ντετερμινιστικός έλεγχος): παραγγέλνουμε μία (1) μονάδα ημερησίως.

Ο περιορισμός απαιτεί $P (S \geq D) \geq 1/3$. Πράγματι, $S=1$ και

$P (S \geq D) = P (1 \geq D) = 1-P (D >1) = 1-P (D=2) = 1-2/3 =1/3$.

Εύκολα θα υποστηριζόταν ότι αυτός είναι ο βέλτιστος τρόπος αφού φαινομενικά παραγγέλνουμε τον ελάχιστο απαιτούμενο αριθμό προϊόντων Π για να καλύψουμε την ζήτηση. Μάλιστα, το ημερήσιο κόστος είναι c χρηματικές μονάδες- φαινομενικά το ελάχιστο δυνατό.

$2^{ος}$ τρόπος (τυχαιοποιημένος έλεγχος): παραγγέλνουμε 0 μονάδες με πιθανότητα 4/5 και 2 μονάδες με πιθανότητα 1/5.

Τότε εφαρμόζοντας το Θεώρημα Ολικής Πιθανότητας έχουμε:

$$P(S\geq D)=P(S\geq D|S=0)\cdot P(S=0)+P(S\geq D|S=2)\cdot P(S=2)$$

$$=P(D=0)\cdot P(S=0)+P(D\leq 2)\cdot P(S=2)$$

$=\frac{1}{6}\cdot\frac{4}{5}+1\cdot\frac{1}{5}=\frac{1}{3}$

που σημαίνει ότι ο περιορισμός ικανοποιείται. Το αναμενόμενο (μέσο) ημερήσιο κόστος παραγγελίας είναι

$E(c\cdot S)=\sum_{s\in R_s} c\cdot s\cdot P(S=s)=c\cdot 0\cdot\frac{4}{5}+c\cdot 2\cdot\frac{1}{5}=0{,}4c$.

Συγκρίνοντας τους δύο τρόπους ελέγχου συμπεραίνουμε ότι ο δεύτερος από αυτούς είναι καλύτερος. Αυτό σημαίνει ότι ο τυχαιοποιημένος έλεγχος είναι προτιμότερος και αναιρεί το πρώτο φαινομενικό συμπέρασμα ότι η ημερήσια παραγγελία μίας μονάδας επιτυγχάνει το ελάχιστο ημερήσιο κόστος παραγγελίας.

3.2 Ενδεικτικά παραδείγματα των Μαρκοβιανών Διαδικασιών Αποφάσεων που αφορούν μοντέλα άπειρου χρονικού ορίζοντα κάτω από το κριτήριο μέσου αναμενόμενου κόστους

Παράδειγμα 6 (επιδείνωσης ενός συστήματος)

Θεωρούμε ένα μηχάνημα το οποίο επιδεινώνεται με την πάροδο του χρόνου και επιθεωρείται στην αρχή κάθε ημέρας για να διαπιστωθεί ο βαθμός επιδείνωσης του. Ας υποθέσουμε ότι το μηχάνημα, μετά από κάθε επιθεώρησή του, βρίσκεται σε μία από της καταστάσεις i , $i=1,2,\dots,N$,όπου κάθε κατάσταση i αναπαριστά το βαθμό επιδείνωσης του μηχανήματος και θεωρείται καλύτερη από την κατασταση $i+1$. Αν η παρούσα κατάσταση του μηχανήματος είναι ίση με i και το μηχάνημα δεν επισκευάζεται, τότε στην αρχή της επόμενης μέρας το μηχάνημα μεταβαίνει στην κατάσταση j με πιθανότητα q_{ij}. Υποθέτουμε ότι το μηχάνημα δε μπορεί να επιδιορθωθεί από μόνο του. Δηλαδή ισχύει ότι $q_{ij}=0$ για $j<i$ και $\sum_{j\geq 1} q_{ij}=1$.

Στην κατάσταση $i=N$ το μηχάνημα δυσλειτουργεί και απαιτείται μία επισκευή του η οποία διαρκεί δυο ημέρες. Στις ενδιάμεσες καταστάσεις $i, 1<i<N$, μπορούμε να επιλέξουμε είτε να θέσουμε σε λειτουργία μία προληπτική συντήρηση του μηχανήματος είτε να μην επέμβουμε στη λειτουργία του μηχανήματος. Η προληπτική συντήρηση διαρκεί μία ημέρα. Το επιδιορθωμένο μηχάνημα μεταβαίνει στην κατάσταση$i=1$. Ο σκοπός μας είναι να καθορίσουμε έναν κανόνα συντήρησης ο οποίος να ελαχιστοποιεί το μακροπρόθεσμο ποσοστό του χρόνου στον οποίο, το μηχάνημα θα βρίσκεται υπό επισκευή. Υποθέτοντας ότι το κόστος για κάθε ημέρα επισκευής του μηχανήματος είναι ίσο με τη μονάδα, το μακροπρόθεσμο μέσο κόστος ανά ημέρα, είναι το μακροπρόθεσμο ποσοστό ημερών κατά τις οποίες το σύστημα θα βρίσκεται υπό επισκευή. Επειδή η επισκευή του μηχανήματος όταν βρίσκεται στην κατάσταση N διαρκεί δύο ημέρες και σε ένα, Μαρκοβιανό μοντέλο αποφάσεων σε διακριτό χρόνο, η κατάσταση πρέπει να ορίζεται στην αρχή κάθε ημέρας, θα πρέπει να εισάγουμε μία βοηθητική κατάσταση για την περίπτωση κατά την οποία η επισκευή του μηχανήματος, βρίσκεται σε εξέλιξη στην αρχή κάποιας ημέρας κατά την οποία το μηχάνημα βρίσκεται στην κατάσταση N. Το σύνολο όλων των δυνατών καταστάσεων της διαδικασίας είναι:$S=\{1,2,\dots,N,N+1\}$. Η κατάσταση $i, 1\leq i\leq N$ αντιστοιχεί στις περιπτώσεις κατά τις οποίες, μετά από μια επιθεώρηση, ο βαθμός επιδείνωσης του μηχανήματος είναι ίσος με i, ενώ η κατάσταση $N+1$ αντιστοιχεί στην περίπτωση κατά την οποία η επισκευή του μηχανήματος είναι ήδη σε εξέλιξη για μια ημέρα, όταν το μηχάνημα διαπιστωθεί ότι βρίσκεται στην κατάσταση $N+1$. Οι δύο δυνατές ενέργειες είναι: $a=1$, αν αποφασιστεί επισκευή και

$a=0$, διαφορετικά. Το σύνολο των ενεργειών για κάθε κατάσταση $i\in S$ είναι:

$$A(1)=\{0\}, A(i)=\{0,1\}, 1<i<N, A(N)=A(N+1)=\{1\}.$$

Οι πιθανότητες μετάβασης $p_{ij}(a)$ είναι:

$$p_{i1}=1\ ,\ 1<i<N, p_{N,N+1}(1)=1, p_{N+1,1}(1)=1, p_{ij}(0)=q_{ij}, 1\leq i<N, j\geq i.$$

Οι υπόλοιπες πιθανότητες μετάβασης είναι $p_{ij}(a) = 0$.Επιπλέον τα κόστη $C(i, a)$ για κάθε κατάσταση $i \in S$ και κάθε ενέργεια $a \in \{0,1\}$ είναι $C(i, 1) = 1$ και $C(i,0) = 0$.

Υποθέτουμε ότι $N = 5$.Οι πιθανότητες επιδείνωσης του μηχανήματος q_{ij} δίνονται από τον πίνακα:

$$(q_{ij})_{i,j} = \begin{pmatrix} 0,75 & 0,2 & 0,05 & 0 & 0 \\ 0 & 0,5 & 0,2 & 0,2 & 0,1 \\ 0 & 0 & 0,5 & 0,25 & 0,25 \\ 0 & 0 & 0 & 0,3 & 0,7 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

Ο αλγόριθμος βελτίωσης των πολιτικών ξεκινά με την πολιτική η οποία ορίζει την ενέργεια της επισκευής του μηχανήματος $a = 1$ σε κάθε κατάσταση, εκτός της κατάστασης 1.

Λύση:

1η Επανάληψη

Βήμα1

Για την πολιτική $R_1^{(1)}$ της οποίας οι ενέργειες είναι:

$R_1^{(1)} = 0, R_2^{(1)} = R_3^{(1)} = R_4^{(1)} = R_5(1) = R_6^{(1)} = 1$ και το μέσο κόστος $g(R^{(1)})$ και οι σχετικές τιμές $h_i(R^{(1)})$ μπορούν να υπολογιστούν ως μοναδική λύση του συστήματος των γραμμικών εξισώσεων:

$$h_1(R^{(1)}) = 0 - g + 0.75h_1(R^{(1)}) + 0.2h_2(R^{(1)}) + 0.05h_3(R^{(1)})$$

$$h_k(R^{(1)}) = 1 - g + h_1(R^{(1)}), k = 2,3,4$$

$$h_5(R^{(1)}) = 1 - g + h_6(R^{(1)}), h_6(R^{(1)}) = 1 - g + h_1(R^{(1)})$$

$$h_6(R^{(1)}) = 0$$

όπου η κατάσταση $s = 6$ επιλέγεται για την εξίσωση κανονικοποίησης. Οι λύσεις αυτών των γραμμικών εξισώσεων δίνουν:

$g(R^{(1)}) = 0.2$,$h_1(R^{(1)}) = 0.8$, $h_2(R^{(1)}) = 0, h_3(R^{(1)}) = 0, h_4(R^{(1)}) = 0, h_5(R^{(1)}) = 0.8, h_6(R^{(1)}) = 0$

Βήμα 2

Κατάσταση i	Ενέργεια a	$C(i, a) - g(R^{(1)}) + \sum_{j=1}^{6} p_{ij}(a)h_j(R^{(1)})$
2	0	$0 - 0.2 + 0.1 * 0.8 = -0.12$
	1	$1 - 0.2 + 0.1 * (-0.8) = 0$
3	0	$0 - 0.2 + 0.25 * 0.8 = 0$
	1	$1 - 0.2 + 1 * (-0.8) = 0$
4	0	$0 - 0.2 + 0.7 * 0.8 = 0.36$

	1	$1-0.2+1*(-0.8)=0$

Παρατηρούμε ότι για την κατάσταση 3, η ενέργεια $a=1$, που καθορίζεται από την παρούσα πολιτική $R^{(1)}$ είναι μία από τις ενέργειες που ελαχιστοποιούν την ποσότητα βελτίωσης των πολιτικών. Η ενέργεια που λαμβάνεται στην κατάσταση 3 είναι η ενέργεια $a=1$(θα μπορούσε να επιλεγεί και η ενέργεια $a=0$).

Από τον παραπάνω πίνακα, λαμβάνουμε τη νέα πολιτική $R^{(2)}$ της οποίας οι ενέργειες είναι:

$R_1^{(2)}=R_2^{(2)}=0, R_3^{(2)}=R_4^{(2)}=R_5^{(2)}=R_6^{(2)}=1$

Βήμα 3

Η νέα πολιτική $R^{(2)}$ είναι διαφορετική από την προηγούμενη πολιτική $R^{(1)}$, συνεπώς ο αλγόριθμος συνεχίζεται με τη δεύτερη επανάληψη.

2η Επανάληψη

Βήμα 1

Για την παρούσα πολιτική $R^{(2)}$ το μέσο κόστος $g\left(R^{(2)}\right)$ και οι σχετικές τιμές $h_i\left(R^{(2)}\right)$ μπορούν να υπολογιστούν ως η μοναδική λύση του ακόλουθου συστήματος των γραμμικών εξισώσεων:

$$h_1\left(R^{(2)}\right)=0-g+0.75h_1\left(R^{(2)}\right)+0.2h_2\left(R^{(2)}\right)+0.05h_3\left(R^{(2)}\right)$$

$$h_2\left(R^{(2)}\right)=0-g+0.5h_2\left(R^{(2)}\right)+0.2h_3\left(R^{(2)}\right)+0.2h_4\left(R^{(2)}\right)+0.1h_5\left(R^{(2)}\right)$$

$$h_k\left(R^{(2)}\right)=1-g+h_1\left(R^{(2)}\right), k=3,4$$

$$h_5\left(R^{(2)}\right)=1-g+h_6\left(R^{(2)}\right)$$

$$h_6\left(R^{(2)}\right)=1-g+h_1\left(R^{(2)}\right)$$

$$h_6\left(R^{(2)}\right)=0$$

Η λύση του συστήματος δίνει:

$$g\left(R^{(2)}\right)=0.172, h_1\left(R^{(2)}\right)=-0.828, h_2\left(R^{(2)}\right)=-0.178, h_3\left(R^{(2)}\right)=0, h_4\left(R^{(2)}\right)=0,$$

$$h_5\left(R^{(2)}\right)=0.828, h_6\left(R^{(2)}\right)=0$$

Βήμα 2

Το βήμα βελτίωσης των πολιτικών για την πολιτική $R^{(2)}$ φαίνεται στον παρακάτω πίνακα:

Κατάσταση i	Ενέργεια a	$C(i,a)-g\left(R^{(2)}\right)+\sum_{j=1}^{6} p_{ij}(a)h_j\left(R^{(2)}\right)$
2	0	$0-0.172+0.5*(-0.178)+0.1*(0.828)=-0.178$
	1	$1-0.172+1*(-0.828)=0$
3	0	$0-0.172+0.25*(0.828)=0.035$

	1	$1 - 0.172 + 1 * (-0.828) = 0$
4	0	$0 - 0.172 + 0.7 * (0.828) = 0.408$
	1	$1 - 0.20.172 + 1 * (-0.828) = 0$

Από τον πίνακα βλέπουμε ότι η νέα πολιτική $R^{(3)}$ της οποίας οι ενέργειες είναι:

$R_1^{(3)} = R_2^{(3)} = 0,\ R_3^{(3)} = R_4^{(3)} = R_5^{(3)} = R_6^{(3)} = 1$ είναι ίδια με την προηγούμενη πολιτική $R^{(2)}$ συνεπώς είναι η βέλτιστη ως προς την ελαχιστοποίηση του μέσου κόστους ανά μονάδα χρόνου. Το ελάχιστο ποσοστό ημερών στις οποίες το μηχάνημα είναι υπό επισκευή είναι ίσο με 0.172 .

Παράδειγμα 7 (εργοστάσιο με αποθηκευτικό χώρο)

Ένα εργοστάσιο έχει έναν αποθηκευτικό χώρο με χωρητικότητα ίση με 4 κυβικά μέτρα για την προσωρινή αποθήκευση σκουπιδιών που παράγονται εξαιτίας της λειτουργίας του εργοστασίου. Κάθε εβδομάδα το εργοστάσιο παράγει k κυβικά μέτρα σκουπιδιών με πιθανότητα p_k,όπου $p_0 = 1/8\,, p_1 = 1/2\,, p_2 = 1/4$ και $p_3 = 1/4$. Αν η ποσότητα που παράγεται σε μια εβδομάδα ξεπερνά την υπολειπόμενη χωρητικότητα του αποθηκευτικού χώρου, το περίσσευμα μεταφέρεται αλλού με ένα κόστος ίσο με 30 ανά κυβικό μέτρο. Στο τέλος κάθε εβδομάδας υπάρχει η δυνατότητα της μετακίνησης των σκουπιδιών από τον αποθηκευτικό χώρο με ένα σταθερό κόστος ίσο με 25 και ένα μεταβλητό κόστος ίσο με 5 ανά κυβικό μέτρο. Η ακόλουθη πολιτική υιοθετείται. Αν στο τέλος κάθε εβδομάδας ο αποθηκευτικός χώρος περιέχει περισσότερα από 2 κυβικά μέτρα ο αποθηκευτικός χώρος αδειάζει, διαφορετικά καμία ποσότητα σκουπιδιών δεν μετακινείται. Με την χρήση του αλγόριθμου βελτίωσης των πολιτικών θα υπολογίσουμε το μακροπρόθεσμο μέσο κόστος ανά μονάδα χρόνου .

Λύση:

Το πρόβλημα είναι ένα πρόβλημα Μαρκοβιανής διαδικασίας αποφάσεων σε διακριτό χρόνο στο οποίο μία απόφαση λαμβάνεται στο τέλος κάθε εβδομάδας. Η κατάσταση του συστήματος περιγράφεται από την ποσότητα σκουπιδιών στο αποθηκευτικό χώρο. Ο χώρος καταστάσεων είναι το σύνολο $I = \{0,1,2,3,4\}$. Οι δύο πιθανές ενέργειες είναι $a = 0$ και $a = 1$ που αντιστοιχούν στην μη-μετακίνηση ποσότητας σκουπιδιών από τον αποθηκευτικό χώρο και στην μετακίνηση όλης της ποσότητας σκουπιδιών από τον αποθηκευτικό χώρο. Μπορούμε να γράψουμε ότι:$A(0) = \{0\}$ και $A(i) = \{0,1\}$ για $1 \le i \le 4$. Το άμεσο κόστος και το αναμενόμενο κόστος για κάθε δυνατή επιλογή ενέργειας a σε κάθε κατάσταση i δίνονται από τις ακόλουθες εκφράσεις:

$$c_i(a = 0) = 30 \sum_{k>4-i}(i + k - 4)p_k\,, c_i(a = 1) = 25 + 5i + 30 \sum_{k>4}(k - 4)\,p_k.$$

Οι πιθανότητες μετάβασης ενός βήματος $p_{ij}(a)$ δίνονται ως εξής:

$$p_{ij}(a = 0) = p_{j-i}\,, i \le j < 4, p_{i4}(a = 0) = \sum_{k\ge 4-i} p_k, p_{ij}(a = 1) = p_j, 0 \le j \le 3.$$

Πριν εφαρμόσουμε τον αλγόριθμο βελτίωσης των πολιτικών ας συνοψίσουμε τα δεδομένα του προβλήματος σε έναν πίνακα:

		$c_i(a)$	$p_{i0}(a)$	$p_{i1}(a)$	$p_{i2}(a)$	$p_{i3}(a)$	$p_{i4}(a)$
$i = 0$	$a = 0$	0	1/8	½	1/4	1/8	0
$i = 1$	$a = 0$	0	0	1/8	1/2	1/4	1/8
	$a = 1$	30	1/8	½	1/4	1/8	0
$i = 2$	$a = 0$	3.75	0	0	1/8	1/2	3/8
	$a = 1$	35	1/8	½	1/4	1/8	0
$i = 3$	$a = 0$	15	0	0	0	1/8	7/8
	$a = 1$	40	1/8	½	1/4	1/8	0
$i = 4$	$a = 0$	41.25	0	0	0	0	1
	$a = 1$	45	1.8	½	1/4	1/8	0

Ξεκινάμε τον αλγόριθμο των πολιτικών με την πολιτική $R^{(1)}$ που σε κάθε κατάσταση i μια ενέργεια a για την οποία τα κόστη $c_i(a)$ είναι τα μικρότερα. Συνεπώς επιλέγουμε ως πολιτική $R^{(1)}$ την πολιτική

$R^{(1)} = (0,0,0,0,0)$.

1η Επανάληψη

Βήμα 1

Για την πολιτική $R^{(1)}$ το μέσο κόστος $g(R^{(1)})$ και οι σχετικές τιμές $h_i(R^{(1)})$ υπολογίζονται επιλύοντας το ακόλουθο σύστημα των γραμμικών εξισώσεων:

$$h_0 = 0 - g + \frac{1}{8}h_0 + \frac{1}{2}h_1 + \frac{1}{4}h_2 + \frac{1}{8}h_3$$

$$h_1 = 0 - g + \frac{1}{8}h_1 + \frac{1}{2}h_2 + \frac{1}{4}h_3 + \frac{1}{8}h_4$$

$$h_2 = 3.75 - g + \frac{1}{8}h_2 + \frac{1}{2}h_3 + \frac{3}{8}h_4$$

$$h_3 = 15 - g + \frac{1}{8}h_3 + \frac{7}{8}h_4$$

$$h_4 = 41.25 - g + h_4$$

$$h_4 = 0$$

Η μοναδική λύση αυτών των γραμμικών εξισώσεων δίνεται από :

$$g\left(R^{(1)}\right) = 41.25\ , h_0\left(R^{(1)}\right) = -120,\ h_1\left(R^{(1)}\right) = -90, h_2\left(R^{(1)}\right) = -60, h_3\left(R^{(1)}\right) = -30, h_4\left(R^{(1)}\right) = 0$$

Βήμα 2

Το βήμα βελτίωσης των πολιτικών για την πολιτική $R^{(1)}$ οδηγεί στις ακόλουθες αριθμητικές τιμές για την ποσότητα βελτίωσης των πολιτικών $c_i(a) - g\big(R^{(1)}\big) + \sum_j p_{ij}(a) h_j\big(R^{(1)}\big)$:

	$\boldsymbol{i = 1}$	$\boldsymbol{i = 2}$	$\boldsymbol{i = 3}$	$\boldsymbol{i = 4}$
$\boldsymbol{a = 0}$	0	0	0	0
$\boldsymbol{a = 1}$	-90	-85	-80	-75

Ελαχιστοποιώντας για κάθε κατάσταση i την ποσότητα βελτίωσης των πολιτικών λαμβάνουμε την νέα πολιτική $R^{(2)} = (0,1,1,1,1)$.

2η Επανάληψη

Βήμα 1

Για την πολιτική $R^{(2)}$ το μέσο κόστος $g\big(R^{(2)}\big)$ και οι σχετικές τιμές $h_i\big(R^{(2)}\big)$ υπολογίζονται επιλύοντας το ακόλουθο σύστημα των γραμμικών εξισώσεων:

$h_i = \gamma_i - g + \frac{1}{8} h_0 + \frac{1}{2} h_1 + \frac{1}{4} h_2 + \frac{1}{8} h_3 \, , i = 0,1,2,3,4$

$h_4 = 0$

όπου $\gamma_0 = 0, \gamma_1 = 30, \gamma_2 = 35, \gamma_3 = 40, \gamma_4 = 45$. Η λύση είναι:

$g\big(R^{(2)}\big) = 28.75 \, , h_0\big(R^{(2)}\big) = -45, \, h_1\big(R^{(2)}\big) = -15, h_2\big(R^{(2)}\big) = -10, h_3\big(R^{(2)}\big) = -5, h_4\big(R^{(2)}\big) = 0$

Βήμα 2

Το βήμα βελτίωσης των πολιτικών για την πολιτική $R^{(2)}$ οδηγεί στις ακόλουθες αριθμητικές τιμές για την ποσότητα βελτίωσης των πολιτικών $c_i(a) - g\big(R^{(2)}\big) + \sum_j p_{ij}(a) h_j\big(R^{(2)}\big)$:

	$\boldsymbol{i = 1}$	$\boldsymbol{i = 2}$	$\boldsymbol{i = 3}$	$\boldsymbol{i = 4}$
$\boldsymbol{a = 0}$	-36.875	-28.75	-14.375	15.5
$\boldsymbol{a = 1}$	-15	-10	-5	0

Ελαχιστοποιώντας για κάθε κατάσταση i την ποσότητα βελτίωσης των πολιτικών λαμβάνουμε την νέα πολιτική $R^{(3)} = (0,0,0,0,1)$.

3η Επανάληψη

Βήμα 1

Για την πολιτική $R^{(3)}$ το μέσο κόστος $g\big(R^{(3)}\big)$ και οι σχετικές τιμές $h_i\big(R^{(3)}\big)$ υπολογίζονται επιλύοντας το ακόλουθο σύστημα των γραμμικών εξισώσεων:

$h_0 = 0 - g + \frac{1}{8}h_0 + \frac{1}{2}h_1 + \frac{1}{4}h_2 + \frac{1}{8}h_3$

$h_1 = 0 - g + \frac{1}{8}h_1 + \frac{1}{2}h_2 + \frac{1}{4}h_3 + \frac{1}{8}h_4$

$h_2 = 3.75 - g + \frac{1}{8}h_2 + \frac{1}{2}h_3 + \frac{3}{8}h_4$

$h_3 = 15 - g + \frac{1}{8}h_3 + \frac{7}{8}h_4$

$h_4 = 45 - g + \frac{1}{8}h_0 + \frac{1}{2}h_1 + \frac{1}{4}h_2 + \frac{1}{8}h_3$

$h_4 = 0$

Η μοναδική λύση αυτών των γραμμικών εξισώσεων δίνεται από:

$g\left(R^{(3)}\right) = 18.1398$, $h_0\left(R^{(3)}\right) = -45$, $h_1\left(R^{(3)}\right) = -32.3255, h_2\left(R^{(3)}\right) = -18.4959, h_3\left(R^{(3)}\right) = -3.5883, h_4\left(R^{(3)}\right) = 0$

Βήμα 2

Το βήμα βελτίωσης των πολιτικών για την πολιτική $R^{(3)}$ οδηγεί στις ακόλουθες αριθμητικές τιμές για την ποσότητα βελτίωσης των πολιτικών $c_i(a) - g\left(R^{(3)}\right) + \sum_j p_{ij}(a)h_j\left(R^{(3)}\right)$:

	$\boldsymbol{i=1}$	$\boldsymbol{i=2}$	$\boldsymbol{i=3}$	$\boldsymbol{i=4}$
$\boldsymbol{a=0}$	-32.33	-18.50	-3.59	13.11
$\boldsymbol{a=1}$	-15	-10	-5	0

Ελαχιστοποιώντας για κάθε κατάσταση i την ποσότητα βελτίωσης των πολιτικών λαμβάνουμε την νέα πολιτική $R^{(4)} = (0{,}0{,}0{,}1{,}1)$.

4η Επανάληψη

Βήμα 1

Για την πολιτική $R^{(4)}$ το μέσο κόστος $g\left(R^{(4)}\right)$ και οι σχετικές τιμές $h_i\left(R^{(4)}\right)$ υπολογίζονται επιλύοντας το ακόλουθο σύστημα των γραμμικών εξισώσεων:

$h_0 = 0 - g + \frac{1}{8}h_0 + \frac{1}{2}h_1 + \frac{1}{4}h_2 + \frac{1}{8}h_3$

$h_1 = 0 - g + \frac{1}{8}h_1 + \frac{1}{2}h_2 + \frac{1}{4}h_3 + \frac{1}{8}h_4$

$h_2 = 3.75 - g + \frac{1}{8}h_2 + \frac{1}{2}h_3 + \frac{3}{8}h_4$

$h_3 = 40 - g + \frac{1}{8}h_0 + \frac{1}{2}h_1 + \frac{1}{4}h_2 + \frac{1}{8}h_3$

$h_4 = 45 - g + \frac{1}{8}h_0 + \frac{1}{2}h_1 + \frac{1}{4}h_2 + \frac{1}{8}h_3$

$h_4 = 0$

Η μοναδική λύση αυτών των γραμμικών εξισώσεων δίνεται από:

$g(R^{(4)}) = 17.7687, \qquad h_0(R^{(4)}) = -45, \qquad h_1(R^{(4)}) = -32.5234, h_2(R^{(4)}) = -18.8785, h_3(R^{(4)}) = -5, h_4(R^{(4)}) = 0.$

Βήμα 2

Το βήμα βελτίωσης των πολιτικών για την πολιτική $R^{(4)}$ οδηγεί στις ακόλουθες αριθμητικές τιμές για την ποσότητα βελτίωσης των πολιτικών $c_i(a) - g(R^{(4)}) + \sum_j p_{ij}(a)h_j(R^{(4)})$:

	$\boldsymbol{i=1}$	$\boldsymbol{i=2}$	$\boldsymbol{i=3}$	$\boldsymbol{i=4}$
$\boldsymbol{a=0}$	-32.52	-18.88	-3.39	23.48
$\boldsymbol{a=1}$	-15	-10	-5	0

Ελαχιστοποιώντας για κάθε κατάσταση i την ποσότητα βελτίωσης των πολιτικών λαμβάνουμε την πολιτική $R^{(5)} = (0,0,0,1,1)$. Επειδή η νέα πολιτική συμπίπτει με την προηγούμενη πολιτική $R^{(4)}$, η πολιτική $R^* = (0,0,0,1,1)$ είναι η βέλτιστη ως προς το κριτήριο της ελαχιστοποίησης του μακροπρόθεσμου αναμενόμενου μέσου κόστους ανά μονάδα χρόνου. Σύμφωνα με τη βέλτιστη πολιτική ο αποθηκευτικός χώρος είναι άδειος όταν περιέχει περισσότερα από 2 κυβικά μέτρα σκουπιδιών, διαφορετικά καθόλου ποσότητα σκουπιδιών δεν μεταφέρεται εκτός του αποθηκευτικού χώρου του εργοστασίου. Το ελάχιστο μακροπρόθεσμο αναμενόμενο μέσο κόστος ανά εβδομάδα είναι ίσο με 17.7687.

Κεφάλαιο 4ο Εφαρμογή ενός προβλήματος συντήρησης με την χρήση MATLAB

4.1 Περιγραφή του κώδικα

Σε αυτό το κεφάλαιο θα ασχοληθούμε με το Παράδειγμα 3 εφαρμόζοντας τον αλγόριθμο διαδοχικών προσεγγίσεων για την εύρεση της βέλτιστης συντήρησης ενός συστήματος, με την εκτέλεση του αντίστοιχου προγράμματος MATLAB. Ο αλγόριθμος ξεκινά με την αρχικοποίηση των παραμέτρων όπως φαίνονται παρακάτω:

- **N:** Το συνολικό πλήθος των καταστάσεων.
- **a:** Η ενέργεια που πρόκειται να ληφθεί.
- **Q:** Ο πίνακας πιθανοτήτων μετάβασης από την τρέχουσα κατάσταση σε όλες τις καταστάσεις, λαμβάνοντας υπόψη την ενέργεια a.
- **C:** Ο πίνακας που περιέχει τις τιμές για το κόστος μετάβασης από τη μια κατάσταση στην άλλη.
- **i:** Η τρέχουσα κατάσταση.
- **j:** Η επόμενη κατάσταση.
- **el:** Συντελεστής ακρίβειας.
- **VN:** Ελάχιστο συνολικό αναμενόμενο κόστος.
- **M:** Ανώτερο όριο.
- **m:** Κατώτερο όριο.
- **R:** Η πολιτική που θα ακολουθήσουμε.
- **average_costs:** Ο πίνακας που αποθηκεύει τις διάφορες τιμές για το μέσο κόστος.
- **perform_iteration:** Η συνάρτηση που μας βοηθάει να εκτελέσουμε τις δοκιμές.
- **gR:** Ρυθμός μέσου κόστους, όταν εφαρμοσθεί η πολιτική R.

```
function perform_test()

   % Test 1

   N1 = 8;

   C1 = [0, 6, 6, 7, 7, 7, 5, 10];

   Q1 = [0.9, 0.1, 0, 0, 0, 0, 0, 0;

         0, 0.8, 0, 0.1, 0, 0, 0.05, 0.05;

         0, 0.7, 0, 0, 0, 0, 0.1, 0.2;

         0, 0, 0, 0, 0, 0, 0.5, 0.5;

         0, 0, 0, 0, 0, 0.3, 0.35, 0.35;

         0, 0, 0, 0, 0, 0.2, 0.4, 0.4;

         0, 0, 0, 0, 0, 0.1, 0.45, 0.45];

   el1 = [0.0001, 0.001, 0.01, 0.1];

   perform_iteration(N1, C1, Q1, el1, 1);

   % Test 2
```

```
N2 = 7;
C2 = [0, 6, 7, 7, 7, 5, 10];
Q2 = [0.9, 0.1, 0, 0, 0, 0, 0;
      0, 0.8, 0, 0.1, 0, 0.05, 0.05;
      0, 0.7, 0, 0, 0, 0.1, 0.2;
      0, 0, 0, 0, 0, 0.5, 0.5;
      0, 0, 0, 0, 0.3, 0.35, 0.35;
      0, 0, 0, 0, 0.2, 0.4, 0.4];
el2 = [0.0001, 0.001, 0.01, 0.1];
perform_iteration(N2, C2, Q2, el2, 2);
% Test 3
N3 = 6;
C3 = [0, 6, 7, 7, 5, 10];
Q3 = [0.9, 0.1, 0, 0, 0, 0;
      0, 0.8, 0, 0.1, 0.05, 0.05;
      0, 0.7, 0, 0, 0.1, 0.2;
      0, 0, 0, 0, 0.5, 0.5;
      0, 0, 0, 0.3, 0.35, 0.35];
el3 = [0.0001, 0.001, 0.01, 0.1];
perform_iteration(N3, C3, Q3, el3, 3);
 % Test 4
N4 = 5;
C4 = [0, 7, 7, 5, 10];
Q4 = [0.9, 0.1, 0, 0, 0;
      0, 0.8, 0.1, 0.05, 0.05;
      0, 0, 0.7, 0.1, 0.2;
      0, 0, 0, 0.5, 0.5];
el4 = [0.0001, 0.001, 0.01, 0.1];
perform_iteration(N4, C4, Q4, el4, 4);
```

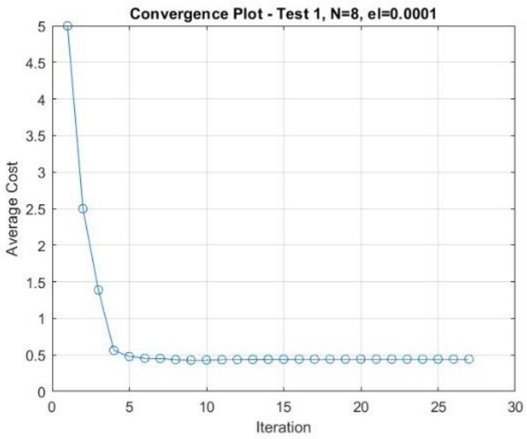
```
end
function perform_iteration(N, C, Q, el_values, test_number)
    for el_idx = 1:length(el_values)
        el = el_values(el_idx);
        VN_1 = zeros(N + 1, 1);
        VN = VN_1;
        VN_a = zeros(N - 2, 2);
        R = zeros(1, N + 1);
        n = 0;
        M = 0;
        m = 0;
       average_costs = []; % Added to store average cost for each iteration
          while (n == 0) || ((M - m) <= 0 || (M - m) >= (el * m))
           s = 0;
           for j = 1:N
               s = s + Q(1, j) * VN_1(j);
           end
           VN(1) = s;
           R(1) = 0;
           for i = 2:(N - 1)
               for a = 0:1
                   s = 0;
                   for j = 1:N
                       s = s + prob611(i, j, a, N, Q) * VN_1(j);
                   end
                  VN_a(i, a + 1) = cost611(i, a, N, C) + s;
               end
               [min_V, index_a] = min(VN_a(i, :));
               VN(i) = min_V;
```

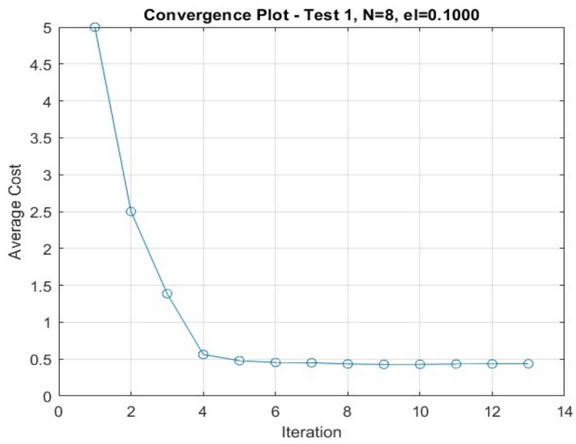
```
        R(i) = index_a - 1;
    end
    VN(N) = C(N) + VN_1(N + 1);
    VN(N + 1) = VN_1(1);
    R(N) = 2;
    R(N + 1) = 2;
    D = VN - VN_1;
    M = max(D);
    m = min(D);
    VN_1 = VN;
    n = n + 1;
    % Calculate and store the average cost for each iteration
    average_costs = [average_costs, (m + M) / 2];
  end
  gR = (m + M) / 2;
  % Print the results
  fprintf('For Test %d and N=%d with el=%.4f:\n', test_number, N, el);
  fprintf('The optimal policy is given by the following vector\n');
  disp(R);
  fprintf('The optimal average cost under the above policy is %.4f\n', gR);
  fprintf('And the number of iterations needed for the algorithm to converge is %d\n', n);
  fprintf('----------------------------------\n');
% Plot the convergence
  figure;
  plot(1:n, average_costs, '-o');
  xlabel('Iteration');
  ylabel('Average Cost');
  title(sprintf('Convergence Plot - Test %d, N=%d, el=%.4f', test_number, N, el));
 grid on;
```

```
        drawnow; % Display the plot immediately
    end
end
function [p] = prob611(i, j, a, N, Q)
    if a == 0 && i < N
        p = Q(i, j);
    elseif a == 1 && j == 1 && i < N
          p = 1;
    elseif a == 2 && ((i == N && j == (N + 1)) || (i == (N + 1) && (j == 1)))
        p = 1;
    else
        p = 0;
    end
end
function [C] = cost611(i, a, N, C)
    if i == 1
        C = 0;
    elseif i > 1 && i < N
        if a == 1
            C = C(i);
        else
             C = 0;
        end
    elseif a == 2 && i == N
        C = C(N);
    else
        C = 0;
    end
end
```

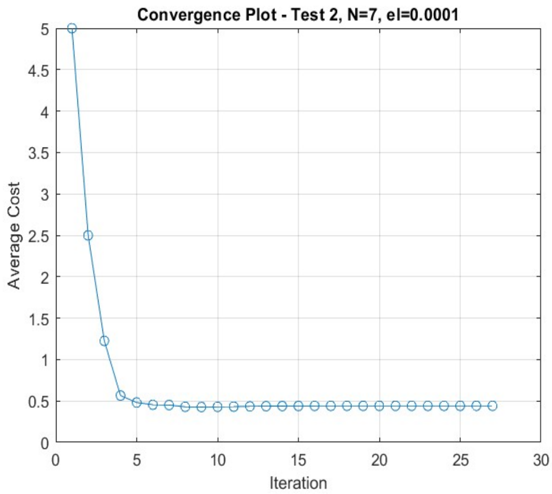
4.2 Ανάλυση ευαισθησίας και παρουσίαση αποτελεσμάτων

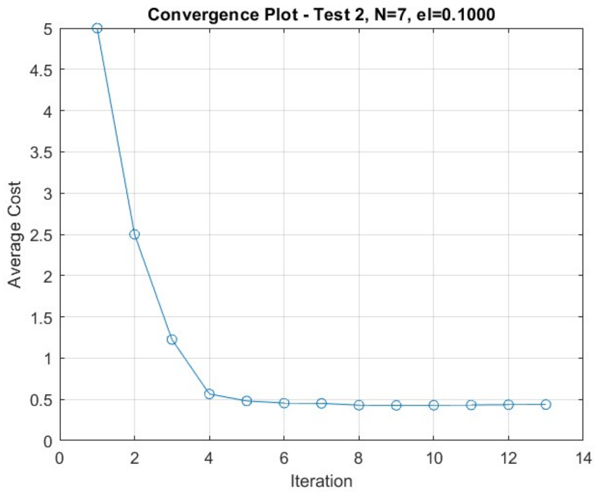
Ακολουθούν ορισμένα γραφήματα από τις δοκιμές που έκανα πάνω στο κώδικα. Ενδεικτικά για πλήθος καταστάσεων N=8 θα έχω τα παρακάτω γραφήματα.



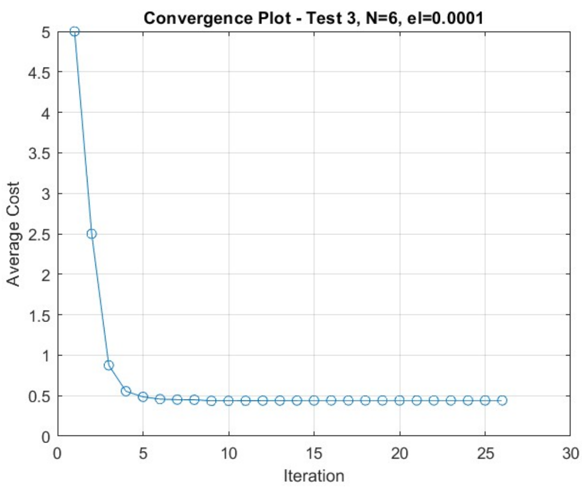


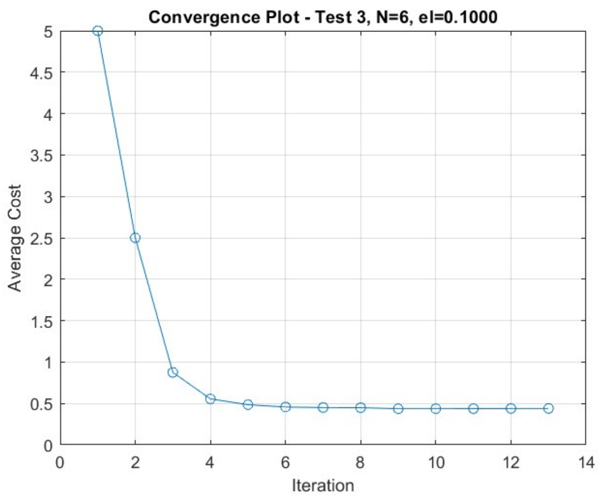
Για N=7 προκύπτουν τα ακόλουθα γραφήματα:



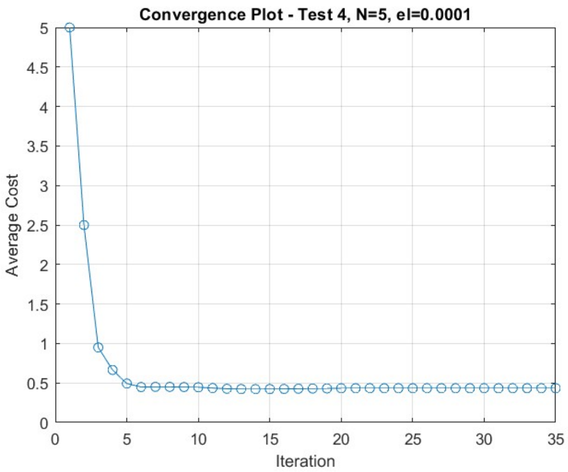


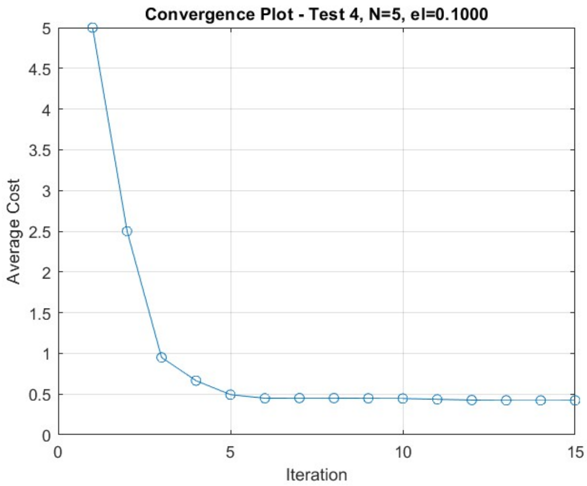
Αντίστοιχα για N=6 θα έχουμε:





Τέλος για πλήθος καταστάσεων N=5 έχουμε τα εξής:





Ο αριθμός των επαναλήψεων που απαιτούνται από τη μέθοδο των διαδοχικών προσεγγίσεων για τη σύγκλιση του αλγορίθμου εξαρτάται από το πρόβλημα που εξετάζεται και συνήθως αυξάνεται όσο αυξάνεται ο αριθμός των καταστάσεων της διαδικασίας. Αξίζει να σημειωθεί ότι με την μέθοδο διαδοχικών προσεγγίσεων υπολογίζεται μία σχεδόν βέλτιστη πολιτική, κι αυτό λόγω της προσέγγισης του ελάχιστου ρυθμού μέσου κόστους από τον αλγόριθμο. Βέβαια, η επιλογή ενός πολύ μικρού $el \geq 0$ επιτρέπει τη βελτίωση της προσέγγισης ανάλογα με τις απαιτήσεις του προβλήματος.

.

Βιβλιογραφία

(1) ROSS S.M (1992) Applied Probability Models with Optimization Application, Dover, New York.

(2) PUTERMAN M.L (1994) Markov Decision Process: Discrete Stochastic Dynamic Programming, Wiley, New York.

(3) TIJMS H.C (1994) Stochastic Models: An Algorithmic Approach, Wiley, New York.

(4) TIJMS H.C (2003) A First Course in Stochastic Models, Wiley, New York.

(5) SENNOT L.I (1999) Stochastic Dynamic Programming and the Control of Queueing Systems, Wiley, New York.

(6) HILLERS F.S-LIEBERMAN G.J (1985) Εισαγωγή στην Επιχειρησιακή Έρευνα, Παπαζήση, Αθήνα.

(7) ΦΑΚΙΝΟΥ Δ. – ΟΙΚΟΝΟΜΟΥ Α. (2003) Εισαγωγή στην Επιχειρησιακή Έρευνα, Συμμετρία, Αθήνα.

(8) ΧΡΥΣΑΦΙΝΟΥ Ο. (2008) Εισαγωγή στις Στοχαστικές Ανελίξεις, Σοφία, Θεσσαλονίκη.

(9) ΒΑΣΙΛΕΙΟΥ Π-Χ.Γ (2000) Στοχαστικές Μέθοδοι στις Επιχειρησιακές Έρευνες, Ζήτη.