ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ
ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ ΚΑΙ
ΑΣΦΑΛΙΣΤΙΚΗΣ ΕΠΙΣΤΗΜΗΣ

Μερικά παρατηρήσιμες Μαρκοβιανές διαδικασίες αποφάσεων και εφαρμογές σε προβλήματα αντικατάστασης συστημάτων και επιλογής διδακτικών μεθόδων

Ιωάννης Γκουλιώνης

Διδακτορική Διατριβή
Υποβλήθηκε στο τμήμα Στατιστικής και Ασφαλιστικής επιστήμης του Πανεπιστημίου Πειραιώς

Πειραιάς 2007

Ευχαριστίες

Ευχαριστώ πρώτα το Θεό που μου έδωσε τη φώτιση και την αντοχή, ώστε να ολοκληρώσω αυτό το πόνημα. Θα επιθυμούσα να ευχαριστήσω θερμά τον επιβλέποντα του διδακτορικού μου, επίκουρο καθηγητή κύριο Στέγγο Δημήτριο, για τον χρόνο που διέθεσε,την ενθάρρυνσή του τις συμβουλές του, την υπομονή, καθώς και την ουσιαστική του συνεισφορά στην παρούσα εργασία. Χωρίς την συμβολή του κυρίου Στέγγου θα ήταν αδύνατη η παρουσίαση αυτής της διατριβής.Είμαι πολύ ευγνώμων για τη γενναιόδωρη υποστήριξή του, τις γόνιμες ιδέες του και την αυστηρή κριτική του που απέτρεψε σφάλματα και με οδήγησε στη σωστή κατεύθυνση.

Θα ήθελα να ευχαριστήσω θερμά επίσης τα μέλη της τριμελούς επιτροπής, καθηγητές κυρίους Μπένο Βασίλειο,και Γεωργακώδη Φώτη για την πολύτιμη βοήθεια, τα πολλαπλά σχόλια και την δημιουργική παρέμβασή τους στην τελική διαμόρφωση της παρούσας έρευνας.

Είμαι πολύ ευγνώμων προς τα μέλη της επιτροπής κρίσης του διδακτορικού μου καθηγητές Αρτίκη Θεόδωρο, Σίσκο Ιωάννη, Χαντζηκωνσταντινίδη Στάθη και Πολίτη Κώστα για την ενθάρρυνσή τους και τον χρόνο που διέθεσαν.

Ευχαριστώ επίσης τον Επίκουρο καθηγητή Πιτσέλη Γιώργο για τις εύστοχες παρατηρήσεις του στη δόμηση του παρόντος.

Με τον κίνδυνο να ξεχάσω κάποιους φίλους με τους οποίους είχα ευχάριστη αλληλεπίδραση όλα αυτά τα χρόνια αναφέρομαι με αγάπη στους: Σαχλά Θανάση, Γκίνη Δημήτρη, Παπαιωάννου Αποστόλη, Ψαράκο Γιώργο, και Μπερσίμη Σωτήρη.

Είμαι πολύ ευτυχής να έχω μια υπομονετική και ανεκτική στα όρια της αυτοθυσίας οικογένεια, που χωρίς την υποστήριξή της δεν θα ήταν δυνατή η παρουσίαση της παρούσας εργασίας.

Λίστα – αλγόριθμων

Ακρώνυμα

DP. Dynamic programming

δ.π. διάνυσμα πληροφορίας

f.t. finitely -transient

MDP Markov decision process

POMDP Partially observable Markov decision process

UMDP Unobservable Markov decision process

PWLC piecewise-linear-convex

SI Stochastically-increasing matrix

TP_2 Ολικά θετικός τάξεως 2

ACOE Εξίσωση βελτιστοποίησης για το μέσο κόστος ανά μονάδα χρόνου.

UB Ομοιόμορφο φράγμα

ΠΙΝΑΚΑΣ ΠΕΡΙΕΧΟΜΕΝΩΝ

ΚΕΦΑΛΑΙΑ

ΠΡΟΛΟΓΟΣ

Οι μερικά παρατηρήσιμες Μαρκοβιανές διαδικασίες αποφάσεων (POMDP) αποτελούν γενίκευση των Μαρκοβιανών διαδικασιών αποφάσεων (MDP),στις οποίες οι καταστάσεις του συστήματος δεν είναι παρατηρήσιμες. Ο decision maker λαμβάνει κάποιο μήνυμα από ένα σύνολο μηνυμάτων στην αρχή κάθε χρονικής περιόδου και ακολούθως παίρνει μια απόφαση από ένα σύνολο εναλλακτικών αποφάσεων.

Εκκινώντας από ένα διάνυσμα πληροφορίας (μία κατανομή πιθανότητας για τις καταστάσεις του συστήματος), αυτό τροποποιείται στην αρχή κάθε χρονικής περιόδου με την έλευση ενός μηνύματος μέσω του τύπου του Bayes,με βάση τον πίνακα μετάβασης καταστάσεων και τον πίνακα μηνυμάτων που αντιστοιχούν στην απόφαση που είχε ληφθεί την προηγούμενη χρονική περίοδο. Το διάνυσμα πληροφορίας ενσωματώνει όλη την πληροφορία της ιστορίας του συστήματος που είναι αναγκαία για την επιλογή μιας απόφασης στην αντίστοιχη χρονική περίοδο. Για προβλήματα κόστους (εσόδων) τα άμεσα κόστη (κέρδη) εξαρτώνται από την κατάσταση του συστήματος και από την απόφαση που επιλέγεται σε μία χρονική περίοδο. Σκοπός είναι ο υπολογισμός του ελάχιστου (μέγιστου) αναμενόμενου ολικού εκπίπτοντος κόστους (κέρδους) για πεπερασμένο ή άπειρο χρονικό ορίζοντα και ο προσδιορισμός της άριστης πολιτικής. Παρόλο που οι POMDP αποτελούν κατάλληλα υποδείγματα για πολλούς τομείς της ανθρώπινης δραστηριότητας, οι υπολογιστικές δυσκολίες καθιστούν την χρήση τους οριακή. Σε αυτό το πλαίσιο οι κύριοι στόχοι της διατριβής αυτής είναι οι ακόλουθοι:

Πρώτον, η ανάπτυξη ευέλικτων αλγόριθμων για την εύρεση άριστων ή σχεδόν άριστων λύσεων τόσο για πεπερασμένο όσο και για άπειρο χρονικό ορίζοντα.Δεύτερον, γενίκευση της συνθήκης Sondik, που εξασφαλίζει ότι μια στάσιμη πολιτική επάγει Μαρκοβιανή διαμέριση στον χώρο των διανυσμάτων πληροφορίας. Έτσι αν μία πολιτική ικανοποιεί αυτή τη συνθήκη, τότε η συνάρτηση του αναμενόμενου ολικού εκπίπτοντος κόστους για άπειρο χρονικό ορίζοντα είναι κατά τμήματα γραμμική και ο

υπολογισμός της ανάγεται στην επίλυση ενός γραμμικού συστήματος εξισώσεων. Τρίτον, εφαρμογή της POMDP σε προβλήματα συντήρησης/αντικατάστασης συστήματος όπου η κατάσταση (επίπεδο χειροτέρευσης) δεν είναι παρατηρήσιμη, αλλά λαμβάνονται μηνύματα που εξαρτώνται από την κατάσταση μέσω ενός μηχανισμού ελέγχου.

Τέταρτον, εφαρμογή της POMDP σε προβλήματα επιλογής διδακτικών μεθόδων,όπου η μαθησιακή κατάσταση της τάξης (βαθμός αφομοίωσης της διδασκόμενης ύλης) δεν είναι παρατηρήσιμη, αλλά λαμβάνονται μηνύματα τύπου επιτυχία/αποτυχία σε test.

Abstract

A Partially Observable Markov Decision Process (POMDP) is a natural extension of the Markov Decision processs (MDP).In POMDPs the state of the system is not observable and therefore unknown. Instead, the decision maker receives a random signal that depends on the state of the system at the beginning of each epoch and then he chooses an action from a finite set of actions.

Starting with an initial prior information vector, belief state,(i.e a probability distribution on the state space), it is updated at beginning of each time epoch just after the arrival of a signal.The new information vector (or belief state) is the posterior distribution on the state space using Bayes' rule that involves the transition and observation matrices assigned to the action selected at the previous time epoch.

It is well known that information vector incoporates the information of the history of the system when choosing an action at a time epoch. The immediate costs (rewards) depend on the current state and action.

The objective is the calculation of the optimal expected total discounted cost (reward) with respect to finite or infinite horizon and the determination of the optimal policy. Although POMDP may provide as suitable model for many applications they may be severely limited due to the computational complexity. Within this context the main goals of this thesis are as follows:

Firstly,the development of flexible algorithms for the determination on optimal or near optimal policies, as well as approximations of the optimal reward or (cost) functions for finite or infinite horizon.

Secondly, to find alternative conditions or generalize known conditions that ensure that a given stationary policy induces a Markovian partition of the belief state space. In this case the reward (or cost) function for infinite horizon is piecewise linear function and its evaluation is significantly simplified.

Thirdly, application of the POMDP model in problems of repair /replacement of the system. It is assumed that the system is monitored incompletely by a certain mechanism which gives the decision maker some information about the exact state of the system.

Fourthly, modeling a teaching methods selection problems as POMDP, where the state of the class (the degree of comprehension teaching material) is unknown to the teacher, and instead signals of success/failure type in tests are received.

ΠΕΡΙΛΗΨΗ

Η διατριβή οργανώνεται ως εξής:

Στο κεφάλαιο 1 περιγράφουμε τις Μαρκοβιανές διαδικασίες αποφάσεων (MDP), τις μερικά παρατηρήσιμες Μαρκοβιανές διαδικασίες αποφάσεων (POMDP) και συνοψίζουμε βασικά αποτελέσματα αναφορικά με τα διάφορα κριτήρια βελτιστοποίησης.Επίσης δίνουμε μία σύντομη ανασκόπηση της σχετικής βιβλιογραφίας.

Στο κεφάλαιο 2 περιγράφουμε την μέθοδο Smallwood-Sondik-Lovejoy για τον υπολογισμό της άριστης συνάρτησης κόστους (κέρδους) σε πεπερασμένο χρονικό ορίζοντα και επισημαίνουμε τις υπολογιστικές δυσκολίες αυτής της προσέγγισης.

Στο κεφάλαιο 3 παρουσιάζουμε μία νέα μέθοδο υπολογισμού της βέλτιστης συνάρτησης του αναμενόμενου ολικού εκπίπτοντος κόστους (ή κέρδους) για πεπερασμένο χρονικό ορίζοντα (αλγόριθμος των ακροτάτων σημείων). Σε γενικές γραμμές η βέλτιστη συνάρτηση κόστους (ή κέρδους) αντιπροσωπεύεται από ένα πεπερασμένο σύνολο διανυσμάτων *(«gradient vectors»)*, το οποίο δεν είναι γνωστό από την αρχή, αλλά «χτίζεται» σε διαδοχικά βήματα. Σε κάθε βήμα το σύνολο αυτό

εμπλουτίζεται με νέα *«gradient vectors»* και υπολογίζεται το μέγιστο σφάλμα προσέγγισης, που είναι φθίνουσα συνάρτηση του αριθμού των βημάτων. Η διαδικασία συνεχίζεται μέχρις ότου το μέγιστο σφάλμα προσέγγισης μηδενισθεί ή γίνει αρκούντως μικρό (δηλαδή μικρότερο ή ίσο από ένα *προκαθορισμένο σφάλμα*).Ο αλγόριθμος των ακροτάτων σημείων αναφέρεται στο πέρασμα από έναν χρονικό ορίζοντα στον επόμενο. *Το συσσωρευμένο σφάλμα προσέγγισης* για κάθε χρονικό ορίζοντα υπολογίζεται μέσω απλής αναγωγικής σχέσης. Το προκαθορισμένο σφάλμα προσέγγισης στον αλγόριθμο των ακροτάτων σημείων επιλέγεται έτσι ώστε το συσσωρευμένο σφάλμα προσέγγισης για οποιοδήποτε χρονικό ορίζοντα να μην υπερβαίνει ένα επιθυμητό φράγμα.

Στο κεφάλαιο 4 εξετάζονται προσεγγίσεις της βέλτιστης συνάρτησης του αναμενόμενου ολικού εκπίπτοντος κόστους (ή κέρδους) σε άπειρο χρονικό ορίζοντα και προσδιορίζονται σχεδόν άριστες πολιτικές εφαρμόζοντας επαναληπτικά τον αλγόριθμο των ακροτάτων σημείων. Μια παρεμφερής μέθοδος που μελετάμε, είναι η επιλογή άνω και κάτω φραγμάτων ως αρχικών προσεγγίσεων της άριστης συνάρτησης κόστους (ή κέρδους), η οποία συνοδεύεται από την επαναληπτική εφαρμογή του αλγορίθμου των ακροτάτων σημείων με σκοπό τη δημιουργία νέων προσεγγίσεων. Από τις προβαλλόμενες συναρτήσεις ελέγχου των νέων προσεγγίσεων, κατασκευάζονται σχεδόν άριστες πολιτικές. Επιτάχυνση της διαδικασίας είναι δυνατή αν η απόσταση των αρχικών φραγμάτων είναι μικρή. Σε κάθε περίπτωση υπολογίζεται ο απαιτούμενος αριθμός επαναλήψεων καθώς και το προκαθορισμένο σφάλμα του αλγορίθμου των ακροτάτων σημείων, έτσι ώστε να επιτυγχάνεται προσέγγιση με οποιαδήποτε επιθυμητή ακρίβεια.

Στο κεφάλαιο 5 παρουσιάζεται η επαναληπτική μέθοδος πολιτικής (policy-iteration) για τις POMDP αναφορικά με το κριτήριο βελτιστοποίησης του αναμενόμενου ολικού εκπίπτοντος κόστους (ή κέρδους) για άπειρο χρονικό ορίζοντα

Στο πλαίσιο αυτό, ο υπολογισμός της συνάρτησης κόστους (ή κέρδους) για άπειρο χρονικό ορίζοντα που αντιστοιχεί σε μία πολιτική απλουστεύεται σημαντικά αν η πολιτική αυτή επάγει Μαρκοβιανή διαμέριση στον χώρο των διανυσμάτων πληροφορίας. Στην περίπτωση αυτή η παραπάνω συνάρτηση είναι κατά τμήματα γραμμική και ο υπολογισμός της ανάγεται στην επίλυση ενός γραμμικού συστήματος

εξισώσεων. Μία γνωστή ικανή συνθήκη ώστε μία πολιτική να επάγει Μαρκοβιανή διαμέριση είναι:η πολιτική να είναι *πεπερασμένα μεταβατική* (συνθήκη του Sondik). Παρουσιάζουμε μια νέα ικανή συνθήκη: η πολιτική να είναι *περιοδική*. Επιπλέον διατυπώνουμε μία γενικότερη συνθήκη από την πεπερασμένη μεταβατικότητα και περιοδικότητα προκειμένου μια πολιτική να επάγει Μαρκοβιανή διαμέριση στον χώρο των διανυσμάτων πληροφορίας.

Στο κεφάλαιο 6 δίνεται γεωμετρική ερμηνεία της δομής της άριστης πολιτικής για άπειρο χρονικό ορίζοντα σε ένα πρόβλημα POMDP συντήρησης/αντικατάστασης συστήματος με πεπερασμένο πλήθος καταστάσεων (επιπέδων χειροτέρευσης), το οποίο παρατηρείται ατελώς μέσω μηνυμάτων ενός μηχανισμού ελέγχου και το οποίο μελετήθηκε από τους Ohnishi-Ibaraki. Η γεωμετρική ερμηνεία αναφέρεται σε τρία επίπεδα χειροτέρευσης στο πλαίσιο της μερικής διάταξης του λόγου πιθανοφανειών.

Στο κεφάλαιο 7 μελετάται μία ειδική περίπτωση του προβλήματος συντήρησης/ αντικατάστασης συστήματος των Ohnishi-Ibaraki που περιγράφεται στο κεφάλαιο 6, με δύο καταστάσεις και δύο μηνύματα. Ειδικότερα εξετάζεται η κλάση των control-limit πολιτικών στην οποία ανήκει και η άριστη πολιτική αναφορικά με το κριτήριο βελτιστοποίησης για άπειρο χρονικό ορίζοντα. Διερευνώνται συνθήκες κάτω από τις οποίες μία control-limit πολιτική είναι πεπερασμένα μεταβατική ή περιοδική. Δίνουμε αριθμητικά παραδείγματα περιοδικών πολιτικών..

Στο κεφάλαιο 8, εξετάζεται το πρόβλημα συντήρησης/αντικατάστασης συστήματος των Ohnishi-Ibaraki ως προς το κριτήριο του μακροπρόθεσμου μέσου κόστους αν μονάδα χρόνου. Μελετάται η δομή της άριστης πολιτικής με το παραπάνω κριτήριο στο ειδικότερο πρόβλημα με δύο καταστάσεις και δύο μηνύματα.

Στο κεφάλαιο 9 μελετάται ένα πρόβλημα επιλογής ανάμεσα σε δύο διδακτικές μεθόδους,μια συμβατική και φθηνή και μια εξειδικευμένη (π.χ.ενισχυτική, υποστηριζόμενη από υπολογιστές κ.λ.π.) και δαπανηρή. Το πρόβλημα τίθεται στην μορφή POMDP με δύο δυνατές μαθησιακές καταστάσεις αναφορικά με το βαθμό αφομοίωσης της διδασκόμενης ύλης από την τάξη και δύο μηνύματα (π.χ. επιτυχία/αποτυχία σε test).Υπολογίζεται αναλυτικά η συνάρτηση του ελάχιστου αναμενόμενου ολικού εκπίπτοντος κόστους για άπειρο χρονικό ορίζοτα και

προσδιορίζεται η άριστη πολιτική επιλογής διδακτικών μεθόδων σε δύο περιπτώσεις: α)περίπτωση πλήρους αβεβαιότητας, όπου το μήνυμα (π.χ. το αποτέλεσμα ενός test) είναι ανεξάρτητο από την μαθησιακή κατάσταση της τάξης είτε επιλέγεται η συμβατική είτε η εξειδικευμένη μέθοδος διδασκαλίας και β)περίπτωση πλήρους αβεβαιότητας όταν επιλέγεται η συμβατική μέθοδος και μερικής πληροφόρησης όταν επιλέγεται η εξειδικευμένη μέθοδος.

ΚΕΦΑΛΑΙΟ 1

Εισαγωγή

Περίληψη

Σε ένα σύστημα, όπου καλούμεθα να πάρουμε αποφάσεις, οι αποφάσεις αυτές μπορεί να βασίζονται πάνω σε πολλούς παράγοντες: Γνώση των άμεσων συνθηκών, για κάποιο εξειδικευμένο πρόγραμμα, προφανή εμπειρία περί των συνεπειών των ποικίλων αποφάσεων,εμπειρικοί-κανόνες, εγκατεστημένα πρωτόκολλα κ.λ.π. Για πολλές περιπτώσεις, αυτή η προσέγγιση εργάζεται καλά, ή τουλάχιστον αρκετά καλά, ώστε να μην υπάρχει λόγος για αλλαγή στον τρόπο και την συλλογιστική που λαμβάνονται οι αποφάσεις. Ωστόσο, όσο τα συστήματα γίνονται περισσότερο πολύπλοκα, η λήψη αποφάσεων δεν είναι απλή υπόθεση, διότι υπάρχει μεγάλη αλληλεπίδραση ανάμεσα σε πολλές παραμέτρους.Αυτή η δυσκολία αυξάνεται δραματικά σε συστήματα, όταν υπάρχει υψηλός βαθμός αβεβαιότητας. Για να αντιμετωπισθεί λοιπόν η παραπάνω δυσκολία, προβλήθηκε η ανάγκη αναζήτησης νέων θεωρητικών μοντέλων, που βασίζονται στη θεωρία πιθανοτήτων, ώστε να καλυφθεί η παραπάνω αβεβαιότητα.

Το μοντέλο των μερικά παρατηρήσιμων Μαρκοβιανών διαδικασιών απόφασης σύντομα POMDP αποτελεί ένα από τα κύρια θεωρητικά μοντέλα που έχουν σαν αντικείμενο τη λήψη αποφάσεων σε συνθήκες αβεβαιότητας. Στις POMDP οι καταστάσεις του συστήματος δεν είναι παρατηρήσιμες. Στη θέση τους ο decision maker λαμβάνει μηνύματα που συνδέονται πιθανοθεωρητικά με τις καταστάσεις του συστήματος. Επομένως με αυτή την έννοια υπάρχει μερική πληροφόρηση. Τα μοντέλα POMDP έχουν πολλές εφαρμογές στον σχεδιασμό και την αντιμετώπιση πολύπλοκων συστημάτων, με ατελή πληροφόρηση.

Στην ενότητα 1.1 περιγράφουμε το μοντέλο της Μαρκοβιανής διαδικασίας αποφάσεων σύντομα M D P.

Στις ενότητες 1.2 και 1.3 κάνουμε μια ιστορική αναδρομή και περιγράφουμε το μοντέλο των POMDPs.

Στην ενότητα 1.4 ορίζονται τα δ.π τα οποία ενσωματώνουν την ιστορία του συστήματος που είναι αναγκαία για τη λήψη αποφάσεων και παρέχεται η επικαιροποίησή τους με τον κανόνα του Bayes. Επίσης ορίζονται οι τελεστές που συνδέονται με συναρτήσεις ελέγχου,ο τελεστής ελαχιστοποίησης (για προβλήματα κόστους),ο τελεστής μεγιστοποίησης (για προβλήματα εσόδων)και δίνονται οι ιδιότητές τους. Στην ενότητα 1.5 παρουσιάζουμε διάφορους τύπους πολιτικών (κανόνων αποφάσεων) καθώς και τα βασικά κριτήρια βελτιστοποίησης για πεπερασμένο και άπειρο χρονικό ορίζοντα.

1.1.Μαρκοβιανές διαδικασίες αποφάσεων με πεπερασμένο πλήθος καταστάσεων και αποφάσεων.

Μία Μαρκοβιανή διαδικασία αποφάσεων (Markov Decision Process),σύντομα MDP, είναι ένα απλό στοχαστικό υπόδειγμα , στο οποίο σε κάθε χρονική περίοδο έχουμε γνώση της κατάστασης του συστήματος πριν από τη λήψη αποφάσεων. Οι MDPs έχουν μελετηθεί στα πλαίσια του άριστου στοχαστικού ελέγχου (optimal stochastic control) και του στοχαστικού δυναμικού προγραμματισμού από τους Howard [51], Derman [24], Ross [104], Bellman [9], Puterman [98], Bertseka [11] κ.α.

Θα περιγράψουμε τώρα το παραπάνω μοντέλο των MDPs. Θεωρούμε προς τούτο ένα δυναμικό σύστημα ,του οποίου η κατάσταση επιθεωρείται στις χρονικές περιόδους (time epochs) $t=0,1,2,\ldots$.

Το σύνολο S των δυνατών καταστάσεων θεωρείται πεπερασμένο,

$$S=\{1,2,\ldots,N\}.$$

Σε κάθε χρονική περίοδο, αφού παρατηρηθεί η κατάσταση του συστήματος,ο decision maker επιλέγει μία απόφαση, από ένα πεπερασμένο σύνολο εναλλακτικών αποφάσεων το οποίο συμβολίζουμε με A.

Έστω X_t η κατάσταση του συστήματος στον χρόνο t και Y_t η απόφαση που επιλέγεται στον χρόνο t. Η στοχαστική διαδικασία $\{X_t, t \in \mathbb{N}_0\}$ περιγράφεται από $N \times N$ πίνακα μετάβασης $P^{\alpha}=(p_{ij}^{\alpha})$, $\alpha \in A$,σύμφωνα με την ακόλουθη σχέση:

Για $i,j \in S, \alpha \in A$,

$$P[X_{t+1}=j \,/\, X_t=i, X_{t-1},\ldots, X_0; Y_t=\alpha, Y_{t-1},\ldots, Y_0]$$
$$= P[X_{t+1}=j \,/\, X_t=i, Y_t=\alpha] \equiv p_{ij}^{\alpha}, \; t \in \mathbb{N}_0.$$

Με άλλα λόγια η πιθανότητα μετάβασης του συστήματος σε μια κατάσταση την επόμενη χρονική περίοδο, εξαρτάται αποκλειστικά από την κατάσταση και την απόφαση που επιλέχθηκε στην τρέχουσα χρονική περίοδο (Μαρκοβιανή ιδιότητα). Επίσης εισάγεται μια δομή κέρδους (εσόδων) ή κόστους ανάλογα με το πρόβλημα. Για τα προβλήματα κέρδους θεωρούμε ότι $q(i,a)$ είναι το άμεσο κέρδος (immediate reward) στον χρόνο t ,όταν η κατάσταση του συστήματος είναι i και επιλέγεται η απόφαση α.

Για τα προβλήματα κόστους αντίστοιχα θεωρούμε ότι $c(i,a)$ είναι το άμεσο κόστος (immediate cost) στον χρόνο t ,όταν η κατάσταση είναι *i και επιλέγεται η απόφαση* α. Συνοψίζοντας μια MDP αναφορικά με πρόβλημα εσόδων, περιγράφεται από την τετράδα $(S,A,(P^a)_{a\in A}, q(.,.))$. Ανάλογα περιγράφεται μια MDP,που αναφέρεται σε πρόβλημα κόστους.

Η ιστορία του συστήματος στον χρόνο t συμβολίζεται με h_t και περιλαμβάνει τις καταστάσεις του συστήματος, καθώς και τις αποφάσεις που επιλέχθηκαν στους χρόνους 0,1,…,t,

δηλαδή

$$h_t = (X_0, Y_0, X_1, Y_1, \ldots, X_t, Y_t), \quad t = 1, 2, \ldots$$
$$h_0 = (X_0, Y_0).$$

Το πεδίο τιμών της ιστορίας h_t είναι το σύνολο

$$H_t = (S \times A)^{t+1}.$$

Μία πολιτική ή στρατηγική (policy-strategy) ορίζεται ένας μηχανισμός λήψης αποφάσεων στις χρονικές περιόδους $t = 0,1,2,\ldots$. Σε πλήρη γενικότητα η επιλογή της

απόφασης στον χρόνο t μέσω της πολιτικής δ γίνεται σύμφωνα με μια κατανομή πιθανότητας , που εξαρτάται από το ζεύγος (h_{t-1}, X_t):

$$\{\delta_t(a/h_{t-1}, X_t) : \alpha \in \mathrm{A}\}, \qquad \underline{\textbf{1.1.1}}$$

όπου $\delta_t(a/h_{t-1}, X_t) \geq 0 \quad \forall a \in A$ και $\sum_{a \in A} \delta_t(a/h_{t-1}, X_t) = 1.$

Με D συμβολίζουμε την κλάση όλων των πολιτικών.

Ορισμός 1.1.1:Μια πολιτική δ καλείται αμνήμων (memoryless policy),αν σε κάθε χρονική περίοδο t η κατανομή πιθανότητας (1.1.1) που επάγεται από τη δ εξαρτάται μόνο από την κατάσταση X_t, δηλαδή για κάθε $a \in A$, $h_{t-1} \in H_{t-1}$,

$$\delta_t(a/h_{t-1}, X_t) = \delta_t(a/X_t).$$

Με D_A συμβολίζουμε το σύνολο των αμνημόνων πολιτικών.

Ορισμός 1.1.2:Μια αμνήμων πολιτική δ καλείται γνήσια ή μη τυχαιοποιημένη (nonrandomized policy) αν σε κάθε χρονική περίοδο t η κατανομή πιθανότητας $\{\delta_t(a/X_t) : a \in A\}$ είναι εκφυλισμένη, δηλαδή $\delta_t(a/X_t)$=0 ή 1, $a \in A$.

Σημειώνουμε ότι η παραπάνω εκφυλισμένη κατανομή πιθανότητας μπορεί να εκφρασθεί ως συνάρτηση ελέγχου (control function) $\delta_t : S \to A$

με $\delta_t(i) = a^* \Leftrightarrow \delta_t(a^*/X_t = i) = 1.$

Συμπεραίνουμε ότι μία γνήσια πολιτική δ μπορεί να θεωρηθεί ως χρονική ακολουθία συναρτήσεων ελέγχου $\{\delta_t : t \in \mathbb{N}_0\}$ και παριστάνεται ως

$$\delta = (\delta_0, \delta_1, \ldots)$$

Με D_Γ συμβολίζουμε το σύνολο των γνήσιων πολιτικών.

Ορισμός 1.1.3:Μια γνήσια πολιτική δ καλείται στάσιμη (stationary policy), αν οι συναρτήσεις ελέγχου στις χρονικές περιόδους t=0,1,2,... ταυτίζονται:

$$\delta_t = \delta_0 \quad \forall t = 1, 2, \ldots, \text{ δηλαδή } \delta = (\delta_0, \delta_0, \ldots).$$

Συνήθως μία γνήσια στάσιμη πολιτική συμβολίζεται

$$\delta^{\infty} = (\delta, \delta, \ldots),$$

όπου δ είναι συνάρτηση ελέγχου (control function) $\delta : S \to A$.

Με D_{Σ} συμβολίζουμε το σύνολο των γνήσιων στάσιμων πολιτικών.

Προφανώς

$$D_{\Sigma} \subset D_{\Gamma} \subset D_A \subset D.$$

Θα περιγράψουμε εν συντομία δύο κριτήρια βελτιστοποίησης. Για περισσότερη ανάλυση βλέπε Bertseka [11] και Derman [24].

Περιοριζόμαστε σε MDP για προβλήματα εσόδων. Για προβλήματα κόστους έχουμε ανάλογη αντιμετώπιση.

1) Κριτήριο βελτιστοποίησης για πεπερασμένο χρονικό ορίζοντα

Θεωρούμε τον χρονικό ορίζοντα $T \geq 1$. Το αναμενόμενο ολικό εκπίπτον κέρδος για τον χρονικό ορίζοντα T, όταν η αρχική κατάσταση του συστήματος είναι $X_0 = i$ και εφαρμόζουμε την πολιτική δ, γράφεται:

$$E_{\delta}\left[\sum_{t=0}^{T-1} \beta^t . q(X_t, Y_t) + \beta^T . q(X_T) / X_0 = i \right], i \in S, \qquad \underline{1.1.2}$$

όπου $\beta > 0$ είναι ο συντελεστής έκπτωσης (discount factor) και q(j) είναι το (άμεσο) κέρδος τερματισμού (terminal reward), όταν η κατάσταση του συστήματος στον χρόνο περάτωσης T είναι j.

Επιθυμούμε να μεγιστοποιήσουμε την (1.1.2) πάνω στην κλάση όλων των πολιτικών D και να καθορίσουμε την άριστη πολιτική για την οποία επιτυγχάνεται το μέγιστο.(Για την ύπαρξη άριστης πολιτικής βλέπε Derman [24] και Denardo [23]).

Έστω $V_n(i)$ το βέλτιστο (μέγιστο) αναμενόμενο ολικό εκπίπτον κέρδος,όταν απομένουν $n \leq T$ χρονικές περίοδοι μέχρι το πέρας του χρονικού ορίζοντα T και η κατάσταση του συστήματος στον χρόνο T-n είναι i $(X_{T-n} = i)$. Η συνάρτηση $V_n(i), i \in S$ καλείται βέλτιστη συνάρτηση τιμών για χρονικό ορίζοντα n και υπολογίζεται από την ακόλουθη αναγωγική σχέση του δυναμικού προγραμματισμού.

$$V_n(i) = \max_a \{q(i,\alpha) + \beta . \sum_{j=1}^{N} p_{ij}^a V_{n-1}(j)\}, i \in S. \qquad \underline{1.1.3}$$

$$V_0(i) = q(i), i \in S$$

Η παράσταση εντός της αγκύλης στην (1.1.3)

$$q(i,\alpha)+\beta.\sum_{j=0}^{N} p_{ij}^{a} V_{n-1}(j)$$

εκφράζει το αναμενόμενο ολικό εκπίπτον κέρδος όταν απομένουν n χρονικές περίοδοι μέχρι το πέρας του χρονικού ορίζοντα Τ, στη χρονική περίοδο Τ-n η κατάσταση του συστήματος είναι i $(X_{T-n}=i)$, επιλέγεται η απόφαση α $(Y_{T-n}=a)$ και ακολουθούμε βέλτιστη πορεία για τις εναπομένουσες n-1 χρονικές περιόδους. Είναι φανερό ότι η άριστη πολιτική για τον πεπερασμένο χρονικό ορίζοντα Τ είναι η γνήσια μη στάσιμη πολιτική (non stationary policy) $\delta^{*}=(\delta_{T}^{*}, \delta_{T-1}^{*},\ldots,\delta_{1}^{*})$, όπου η συνάρτηση ελέγχου δ_{n}^{*} υπολογίζεται από τη σχέση:

$$\delta_{n}^{*}(i)=\arg\max_{a}\{q(i,\alpha)+\beta.\sum_{j=1}^{N} p_{ij}^{a} V_{n-1}(j)\}, i\in S, \qquad \underline{1.1.4}$$

$$n=1,2,\ldots,T.$$

2) Κριτήριο βελτιστοποίησης για άπειρο χρονικό ορίζοντα

Το αναμενόμενο ολικό εκπίπτον κέρδος για άπειρο χρονικό ορίζοντα, όταν η αρχική κατάσταση του συστήματος $X_0=i$ και εφαρμόζουμε την πολιτική δ, γράφεται:

$$V_{\delta}(i)=E_{\delta}\left[\sum_{t=0}^{\infty}\beta^{t}.q(X_t,Y_t)/X_0=i\right], i\in S, \qquad \underline{1.1.5}$$

όπου για τον συντελεστή εκπτώσεως υποθέτουμε ότι $\beta\in[0,1)$. Αποδεικνύεται εύκολα ότι για κάθε $\delta\in D$,

$$|V_{\delta}(i)|\le\frac{\Lambda}{1-\beta}\quad\forall\ i\in S,$$

όπου

$$\Lambda\equiv\max_{i,a}|q(i,a)|.$$

Η συνάρτηση $V_{\delta}(i)$, $i\in S$ αναφέρεται ως συνάρτηση τιμών για την πολιτική δ^{∞}. Επιθυμούμε να μεγιστοποιήσουμε την (1.1.5) πάνω στην κλάση όλων των πολιτικών D και να καθορίσουμε την άριστη πολιτική για την οποία επιτυγχάνεται το μέγιστο. Αποδεικνύεται ότι υπάρχει γνήσια στάσιμη πολιτική που είναι άριστη (δηλαδή μεγιστοποιεί την (1.1.5)).(βλέπε και Maitra (1968) [79]).

Επομένως μπορούμε να περιορισθούμε στην κλάση των γνήσιων στάσιμων πολιτικών D_Σ.

Θεωρούμε την γνήσια στάσιμη πολιτική $\delta^\infty = (\delta, \delta, ..., \delta)$. Η συνάρτηση τιμών V_δ για την πολιτική δ^∞ είναι η μοναδική λύση της εξίσωσης βελτιστοποίησης

$$V_\delta(i) = q(i,\delta(i)) + \beta \cdot \sum_{j=1}^{N} p_{ij}^{\delta(i)} V_\delta(j), \quad i=1,2,\ldots,N. \qquad \underline{1.1.6}$$

Η βέλτιστη συνάρτηση τιμών

$$V^*(i) = \sup_{\delta \in D} V_\delta(i) = \sup_{\delta^\infty \in D_\Sigma} V_\delta(i), \; i \in S.$$

Είναι η μοναδική λύση της εξίσωσης βελτιστοποίησης

$$V^*(i) = \max_a \{q(i,\alpha) + \beta \cdot \sum_{j=1}^{N} p_{ij}^a V^*(j)\}, \; i \in S. \qquad \underline{1.1.7}$$

Η παράσταση εντός της αγκύλης στην (1.1.7)

$$q(i,\alpha) + \beta \cdot \sum_{j=1}^{N} p_{ij}^a V^*(j)$$

εκφράζει το αναμενόμενο ολικό εκπίπτον κέρδος, όταν στον χρόνο t=0 η κατάσταση είναι $X_0 = i$, επιλέγεται η απόφαση $Y_0 = \alpha$ και κατόπιν ακολουθείται βέλτιστη πορεία. Η συνάρτηση ελέγχου της άριστης πολιτικής $(\delta^*)^\infty = (\delta^*, \delta^*, ...)$ προσδιορίζεται από τη σχέση:

$$\delta^*(i) = \arg\max_a [q(i,\alpha) + \beta \cdot \sum_{j=1}^{N} p_{ij}^a V^*(j)], \; i \in S. \qquad \underline{1.1.8}$$

Επομένως για τον προσδιορισμό της άριστης πολιτικής $(\delta^*)^\infty$ είναι αναγκαίος ο υπολογισμός της άριστης συνάρτησης τιμών V^*. Παραθέτουμε δύο μεθόδους υπολογισμού της V^*.

Α) Μέθοδος των διαδοχικών προσεγγίσεων ή επαναληπτική μέθοδος τιμών (method of successive approximations,value-iteration method).

Με τη μέθοδο αυτή επιλέγουμε αυθαίρετα μία συνάρτηση u (i), $i \in S$ και ακολούθως υπολογίζουμε την ακολουθία των συναρτήσεων $\{ V_n, n \in \mathbb{N} \}$ μέσω της αναγωγικής σχέσης

$$V_n(i) = \max_a \{q(i,\alpha) + \beta . \sum_{j=1}^{N} p_{ij}^a . V_{n-1}(j)\}, i \in S \qquad \underline{1.1.9}$$

$$V_0(i) = u(i), i \in S$$

Η συνάρτηση V_n είναι η βέλτιστη συνάρτηση τιμών για χρονικό ορίζοντα n με συνάρτηση κέρδους τερματισμού u.

Αποδεικνύεται ότι:

$$V_n \underset{n \to \infty}{\to} V^*$$

(βλέπε και Ross [107]).

Ορισμός 1.1.4: Μια πολιτική δ^∞ καλείται $\varepsilon - \acute{\alpha}ριστη$, όπου $\varepsilon > 0$, αν η συνάρτηση τιμών για την δ^∞, V_δ, αποτελεί προσέγγιση της άριστης συνάρτησης τιμών V^* με μέγιστο σφάλμα προσέγγισης μικρότερο ή ίσο του ε, δηλ.

$$\max_{i \in S} |V_\delta(i) - V^*(i)| \le \varepsilon$$

Πρόταση 1.1.1: Θεωρούμε τις συναρτήσεις V_n ,n=0,1,2,... οι οποίες υπολογίζονται μέσω της αναγωγικής σχέσης (1.1.9).

Αν για κάποιο $n \ge 1$,

$$\max_{i \in S} |V_n(i) - V_{n-1}(i)| \le \varepsilon \quad ,\text{όπου} \quad \varepsilon > 0 ,$$

τότε η πολιτική δ^∞ με συνάρτηση ελέγχου που υπολογίζεται από τη σχέση

$$\delta(i) = \arg\max_a [q(i,\alpha) + \beta . \sum_{j=1}^{N} p_{ij}^a . V_{n-1}(j)], i \in S$$

είναι $\frac{2\beta\varepsilon}{1-\beta}-\dot{a}\rho\iota\sigma\tau\eta$. (βλέπε και Bellman [9]).

Αλγόριθμος Α, (Value-iteration) (Howard [51],Bellman [9])

1. Input: Μιά αρχική αυθαίρετη συνάρτηση τιμών V_n με n=0,και μια παράμετρος $\varepsilon > 0$,για την επίτευξη ε-βέλτιστης πολιτικής.

2. Improve value function: Αυξάνουμε το n και για κάθε $i \in S$,

$$V_n(i)=\max_a q(i,\alpha)+\beta.\sum_j p_{ij}^a V_{n-1}(j),$$

3.Convergence -test: Αν

$$\max_{i\in S}\left|V_n(i)-V_{n-1}(i)\right| \le \frac{\varepsilon.(1-\beta)}{2\beta}$$

πήγαινε στο βήμα 4. Άλλως πήγαινε στο βήμα 2.

4.Output:Μια ε-βέλτιστη πολιτική δ^∞ με συνάρτηση ελέγχου

$$\delta(i) := \arg\max_a [q(i,\alpha)+\beta.\sum_j p_{ij}^a V_{n-1}(j)],\ i \in S.$$

Β) Επαναληπτική μέθοδος πολιτικής (policy iteration method).

Η μέθοδος αυτή αναπτύχθηκε από τους Howard [51] και Blackwell [15], [16] και περιλαμβάνει δύο στάδια. Κατά το πρώτο στάδιο (policy evaluation) υπολογίζεται η συνάρτηση τιμών V_δ μιας γνήσιας στάσιμης πολιτικής δ^∞,μέσω του συστήματος των εξισώσεων (1.1.6).Κατά το δεύτερο στάδιο (policy improvement) εντοπίζεται μία βελτιωμένη πολιτική $(\delta')^\infty$.Κατόπιν εφαρμόζεται το πρώτο στάδιο στην πολιτική $(\delta')^\infty$,και η διαδικασία συνεχίζεται μέχρις ότου καταλήξουμε σε άριστη πολιτική. Η πρόταση που ακολουθεί διευκρινίζει το δεύτερο στάδιο της βελτιωμένης πολιτικής.

Πρόταση 1.1.2: Έστω δ^{∞} γνήσια στάσιμη πολιτική και V_δ η αντίστοιχη συνάρτηση τιμών .Θεωρούμε την ακόλουθη συνάρτηση ελέγχου δ' :

$$\delta'(i) := \arg\max_a \{ q(i,\alpha) + \beta . \sum_{j=1}^{N} p_{ij}^{a} . V_\delta(j) , \quad i \in S .$$

Τότε για την συνάρτηση τιμών $V_{\delta'}$ της πολιτικής $(\delta')^{\infty}$ ισχύει:

$$V_{\delta'}(i) = q(i,\delta(i)) + \beta . \sum_{j=1}^{N} p_{ij}^{\delta(i)} . V_\delta(j) , \quad i=1,2,...,N .$$

$$V_{\delta'}(i) \geq V_\delta(i) , \forall\ i \in S ,$$

και εάν $V_{\delta'}(i) = V_\delta(i) \ \forall\ i \in S$ τότε:

$$V_{\delta'} = V_\delta = V^* \text{ (άριστη συνάρτηση τιμών).}$$

Βλέπε και Howard [51].

Αλγόριθμος A_2 (policy-Iteration)

1.**Input:** Μια αρχική γνήσια στάσιμη πολιτική $\delta^{\infty} = (\delta, \delta, ..., \delta)$.

2.**Policy evaluation:** Υπολογίζουμε την συνάρτηση τιμών V_δ, για την πολιτική δ^{∞}, λύνοντας το σύνολο των εξισώσεων (1.1.6).

3.**Policy improvement:** Για κάθε κατάσταση $i \in S$, αν υπάρχει κάποια απόφαση $\alpha \in A$ ώστε: $q(i,\alpha) + \beta . \sum_j p_{ij}^{a} . V_\delta(j) > V_\delta(i)$,

τότε $\delta'(i) = \alpha$, αλλιώς , $\delta'(i) = \delta(i)$.

4.**Convergence -test:** Αν δ' είναι η ίδια με την δ, τότε πάμε στο βήμα 5. Αλλιώς θέτουμε δ=δ' και πάμε στο βήμα 2

5. **Output:** Μία άριστη πολιτική δ^{∞} και άριστη συνάρτηση τιμών $V^* = V_\delta$.

Επειδή τα σύνολα S,A υποτέθηκαν πεπερασμένα ,το πλήθος των δυνατών συναρτήσεων ελέγχου είναι πεπερασμένο ($|A|^N$).Επομένως και το σύνολο των γνήσιων στάσιμων

πολιτικών D_Σ είναι πεπερασμένο με πληθάριθμο $|D_\Sigma| = |A|^N$.Επειδή σε κάθε επανάληψη βελτιώνεται η πολιτική στο βήμα 3,και το πλήθος των δυνατών γνήσιων στάσιμων πολιτικών είναι πεπερασμένο, ο αλγόριθμος A_2 τερματίζεται σε πεπερασμένο πλήθος επαναλήψεων.

1.2. Μερικά παρατηρήσιμες Μαρκοβιανές διαδικασίες αποφάσεων, (ιστορική αναδρομή).

Μια μερικά παρατηρήσιμη Μαρκοβιανή διαδικασία,σύντομα POMDP,είναι μια γενικευμένη (Markov decision process), που επιτρέπει ατελή πληροφόρηση του συστήματος των καταστάσεων. Η γενίκευση αυτή, είναι σημαντική σε προβλήματα όπου η αβεβαιότητα ως προς την κατάσταση είναι το κεντρικό και ουσιώδες. Στην πραγματικότητα έχουμε ένα ευρύ πεδίο εφαρμογών του μοντέλου POMDP, Goulionis [36,37,38,44,45] και το κοινό σημείο όλων αυτών των εφαρμογών, είναι η αβεβαιότητα ως προς την κατάσταση στην οποία βρίσκεται το σύστημα, και η επίδραση αυτής της αβεβαιότητας στην επιλογή μιας βέλτιστης πολιτικής. Το μοντέλο POMDP επίσης εξαναγκάζει τον ερευνητή να κάνει μια ξεκάθαρη διάκριση μεταξύ πραγματικών καταστάσεων και μηνυμάτων.

Παράλληλα ανάλογα με την «κατάσταση» που φαίνεται να βρίσκεται το σύστημα (belief-state) λαμβάνεται μια απόφαση. Όταν κάποιος καλείται να πάρει αποφάσεις βασίζεται σε ολόκληρη την ιστορία του συστήματος, ένα σύνολο δηλαδή από αποφάσεις και μηνύματα,που έχουν ήδη ληφθεί. Η POMDP συνήθως μετατρέπεται σε μια ισοδύναμη MDP,όπου ο χώρος καταστάσεων είναι η δεσμευμένη πιθανότητα κατανομής της κατάστασης του συστήματος, δοσμένης της ιστορίας του (Astrom 1965) [5],[6].

Έρευνα για τις POMDPs άρχισε την δεκαετία του 1960 από τους Howard [51] και Drake [26],που ανέπτυξαν το πιο απλό μοντέλο. Το 1965 ο Astrom [5] διατύπωσε το μοντέλο για τις μερικά παρατηρήσιμες MDPs σε πεπερασμένο χρονικό ορίζοντα.

Οι θεμελιωτές όμως της θεωρίας POMDP είναι οι Smallwood και Sondik [117,118,119,120],και κυρίαρχα ο δεύτερος.Αυτοί έδωσαν το έναυσμα για ικανοποιητικούς αλγόριθμους. Τα κύρια σημεία του μοντέλου είναι ο ορισμός μίας στοχαστικής διαδικασίας καταστάσεων (core-process)και μιας στοχαστικής διαδικασίας μηνυμάτων. Η στοχαστική διαδικασία καταστάσεων σχηματίζει μια Μαρκοβιανή διαδικασία, και δεν μπορεί να παρατηρηθεί απευθείας. Η στοχαστική διαδικασία μηνυμάτων είναι μια ακολουθία από καταστάσεις, που πραγματικά παρατηρούνται, αποφασίζεται μέσω της (core-process) και δεν είναι απαραίτητα Μαρκοβιανή.

Ο Sondik το 1971 [117] διατύπωσε τον αλγόριθμο ενός βήματος (one –pass-algorithm).Απέδειξε δύο ουσιαστικά πράγματα, που έκαναν την μέχρι τότε σχεδόν άβολη υπολογιστική διαδικασία αρκετά εφικτή, και την θεωρία γόνιμη και ρεαλιστική στην αντιμετώπιση προβλημάτων. Πρώτα απέδειξε ότι η βέλτιστη συνάρτηση τιμών σε πεπερασμένο χρονικό ορίζοντα, έχει δύο σημαντικές ιδιότητες, δηλαδή είναι κατά τμήματα γραμμική και κυρτή (piecewise –linear and convex) p.w.l.c. Κατόπιν ήλθε σαν άμεσο αποτέλεσμα της *κατά τμήματα γραμμικότητας και κυρτότητας*, ότι η παραπάνω συνάρτηση για κάθε χρονικό ορίζοντα T, μπορεί να αντιπροσωπευθεί χρησιμοποιώντας τα λεγόμενα «gradients vectors».

Επεκτείνοντας τις σκέψεις του στο πρόβλημα του άπειρου χρονικού ορίζοντα, εισήγαγε την κλάση των πεπερασμένα μεταβατικών πολιτικών, και ανέπτυξε προσεγγίσεις για κάθε στάσιμη πολιτική, που βασίζονται ακριβώς στις πεπερασμένα μεταβατικές πολιτικές, δείχνοντας παράλληλα ότι οι τομές των συναρτήσεων οφέλους, που βασίζονται σε μια τέτοια προσέγγιση, μπορούν να συμπεριληφθούν στον αλγόριθμο του Howard [51] (policy-improvement) με επακόλουθη σύγκλιση. Δηλαδή οι πεπερασμένα μεταβατικές πολιτικές παίζουν έναν ρόλο κλειδί, διότι γενικεύουν δυναμικές ισοδύναμες με εκείνες των MDPs. Ο Denardo [23] έδωσε μια πιο βολική μορφή στα «gradients-vectors» και εισήγαγε τους τελεστές H_δ,H στην αντιπροσώπευση της βέλτιστης συνάρτησης τιμών σε κάθε χρονικό ορίζοντα.

Πολλοί ερευνητές ασχολήθηκαν με την επίτευξη ενός πιο λειτουργικού από την άποψη μοντέλου, διότι όπως απέδειξε ο Mukherjee [87] ο αριθμός των « gradient-

vectors» αυξάνει εκθετικά καθώς αυξάνεται ο χρονικός ορίζοντας, με αποτέλεσμα η όλη διαδικασία του δυναμικού προγραμματισμού να είναι υπολογιστικά ανέφικτη και το μοντέλο μη γόνιμο στην αντιμετώπιση ρεαλιστικών προβλημάτων.

Ο Eagle (1984) [29] χρησιμοποίησε την POMDP προκειμένου να μελετήσει κινούμενο στόχο.

Ο Albright (1979) [1] έδωσε συνθήκες, ώστε ή βέλτιστη πολιτική για ένα σύστημα δύο καταστάσεων να είναι μονότονη ως προς την κατανομή πιθανότητας των καταστάσεων του συστήματος.

Ο Platzman (1980) [96] ανέπτυξε τις συνθήκες για να είναι καλά ορισμένο το πρόβλημα σε άπειρο χρονικό ορίζοντα (undiscounded -infinite - horizon POMDP).

Ο Lovejoy (1987)[76] παρείχε ικανοποιητικές συνθήκες, που αποφέρουν μονότονες βέλτιστες πολιτικές για το πρόβλημα POMDPs σε πεπερασμένο χρονικό ορίζοντα.

Ο Littman(1995) [69],[70],[71],[72] έδωσε μια πιο τυποποιημένη μορφή στο πρόβλημα του άπειρου χρονικού ορίζοντα, και απέδειξε πολύ απλά την ιδιότητα της κατά τμήματα γραμμικής συνάρτησης, ενώ σύνδεσε το εργαλείο που λέγεται δυναμικός προγραμματισμός, με τα (policy - trees).

Η απλούστερη έκδοση του μοντέλου που θα αναπτυχθεί είναι το μοντέλο Μαρκοβιανής εξέλιξης. Η πρώτη διατύπωσή του έγινε με τις εργασίες του Bellman το 1957 [9], μολονότι προηγήθηκαν οι εργασίες του Pollock [97] που αφορούσαν τα στοχαστικά παίγνια.

Βέβαια οι εργασίες των Howard [51] εφαρμόζοντας δυναμικό προγραμματισμό, Manne[80] εφαρμόζοντας γραμμικό προγραμματισμό(linear-programming formulation), Blackwell [14] που επέκτεινε το πρόβλημα σε αυθαίρετους χώρους καταστάσεων και Ross [104] ήταν οι θεμέλιοι λίθοι. Το μοντέλο των POMDPs είναι ένα τμήμα του δυναμικού προγραμματισμού και δίνει χρήσιμα και επιτυχημένα εργαλεία σε επιχειρήσεις, που ασχολούνται με ένα τέτοιο είδος πολύπλοκων αποφάσεων. Πέρα από αυτό,βρίσκεται στις παρυφές της (artificial-intelligence-community),με την συλλογιστική, ότι αρκετές φορές απαιτείται, ή είναι επιθυμητό, να πάρουμε μια ακολουθία αποφάσεων χωρίς την συμμετοχή ανθρώπινου παράγοντα, βλέπε Madani [78].Υπάρχει στενή σύνδεση ανάμεσα σε επιχειρησιακή έρευνα (δυναμικό

προγραμματισμό) και τεχνητή νοημοσύνη (artificial-intelligence) βλέπε και Zhang [142].

1.3.Μερικά παρατηρήσιμη Μαρκοβιανή διαδικασία αποφάσεων πεπερασμένου πλήθους καταστάσεων, περιγραφή.

Στην ενότητα αυτή θα περιγράψουμε την μερικά παρατηρήσιμη Μαρκοβιανή διαδικασία αποφάσεων πεπερασμένου πλήθους καταστάσεων (finite state partially observable Markov decision process) ή σύντομα POMDP.

Το σύνολο A των εναλλακτικών αποφάσεων, που έχει στη διάθεσή του ο decision maker (action space) θεωρείται πεπερασμένο.Οι αποφάσεις επιλέγονται σε διακριτούς χρόνους t=0,1,2,3…

Η στοχαστική διαδικασία των αποφάσεων συμβολίζεται με $\{Y_t, t\in \mathbb{N}_0\}$.

Θεωρούμε ότι το σύνολο των καταστάσεων του συστήματος είναι S={1,2,3…,N}.Η στοχαστική διαδικασία $\{X_t, t\in \mathbb{N}_0\}$ των καταστάσεων καλείται διαδικασία πυρήνα (core-process) και υποτίθεται είναι μια (πεπερασμένη) Μαρκοβιανή διαδικασία που περιγράφεται από έναν $N\times N$ πίνακα μετάβασης $P^{\alpha} = (p_{ij}^{a})$ $\alpha\in A$, σύμφωνα με την ακόλουθη σχέση:

Για $i, j \in S$,$\alpha\in A$,

$$p[X_{t+1}=j / X_t=i, X_{t-1}, \ldots, X_0; Y_t=\alpha, Y_{t-1}, \ldots, Y_0] = p[X_{t+1}=j / X_t=i, Y_t=\alpha] \equiv p_{ij}^{a}, \; t\in \mathbb{N}_0.$$

Με άλλα λόγια η πιθανότητα μετάβασης του συστήματος σε μία κατάσταση κάποια χρονική περίοδο (time epoch),εξαρτάται αποκλειστικά από την κατάσταση του συστήματος καθώς και από την απόφαση που επιλέχθηκε την προηγούμενη περίοδο. Θεωρούμε ότι η διαδικασία πυρήνα δεν είναι άμεσα παρατηρήσιμη,δηλαδή ο decision maker δεν λαμβάνει γνώση της κατάστασης του συστήματος στον χρόνο t=0,1,2,…

Ο decision maker λαμβάνει ωστόσο στον χρόνο t ένα μήνυμα από ένα σύνολο μηνυμάτων Θ={1,2,3,…M}.Η στοχαστική διαδικασία μηνυμάτων $\{Z_t, t\in \mathbb{N}_0\}$συνδέεται με τη διαδικασία πυρήνα $\{X_t, t\in \mathbb{N}_0\}$ μέσω της ακόλουθης σχέσης:

Για $i, j \in S$, $\theta \in \Theta$, $\alpha \in A$,

$p[Z_{t-1}=\theta/Z_t,\ldots, Z_1; X_{t-1}=i, X_t,\ldots,X_0; Y_1=\alpha, Y_{t-1},\ldots,Y_0]=p[Z_{t+1}=\theta/X_{t+1}=\iota, Y_t=\alpha] \equiv r_{i\theta}^{\alpha}$, $t \in \mathbb{N}_0$.

Με άλλα λόγια η πιθανότητα με την οποία λαμβάνεται ένα μήνυμα κάποια χρονική περίοδο (time epoch),εξαρτάται αποκλειστικά από την κατάσταση του συστήματος την ίδια χρονική περίοδο και την απόφαση που επιλέχθηκε την προηγούμενη περίοδο. Οι στοχαστικοί $N \times M$ πίνακες $R^a = (r_{i\theta}^{\alpha})$,καλούνται πίνακες μηνυμάτων. Σημειώνουμε ότι σε κάθε χρονική περίοδο, η απόφαση λαμβάνεται μετά τη λήψη του μηνύματος. Αναλυτικότερα η σειρά με την οποία συμβαίνουν τα γεγονότα θεωρείται η ακόλουθη:Αρχικά για (t=0) το σύστημα βρίσκεται στην κατάσταση X_0,επιλέγεται μια απόφαση Y_0 και στην αρχή της χρονικής περιόδου t=1 το σύστημα μεταβαίνει στην κατάσταση X_1, λαμβάνεται ένα μήνυμα Z_1 και ακολούθως επιλέγεται η απόφαση Y_1. Γενικά στην περίοδο t το σύστημα βρίσκεται στην κατάσταση X_t,λαμβάνεται ένα μήνυμα Z_t και κατόπιν επιλέγεται μία απόφαση Y_t.

Στην αρχή της περιόδου t+1 το σύστημα μεταβαίνει στην κατάσταση X_{t+1},λαμβάνεται μήνυμα Z_{t+1},κατόπιν επιλέγεται η απόφαση Y_{t+1} κ.ο.κ.

Για να γίνουν κατανοητά τα παραπάνω, δίνουμε ένα τυπικό παράδειγμα POMDP. Όταν εξετάζουμε την κατάσταση ενός ασθενούς που πάσχει από στεφανιαία νόσο, το αποτέλεσμα του λεγόμενου τέστ κόπωσης, το επίπεδο ισχαιμίας, καθώς και ο πόνος στο στήθος είναι μηνύματα, που συνυφαίνονται με την ζωτική κατάσταση του ασθενούς. Ωστόσο, δεν γνωρίζουμε την κατάσταση στην οποία βρίσκεται ο εν λόγω ασθενής.

Τέλος εισάγεται μία δομή κέρδους (εσόδων) ή δομή κόστους ,ανάλογα με το πρόβλημα. Για τα προβλήματα κέρδους, θεωρούμε ότι q(i,α) είναι το άμεσο κέρδος (immediate reward) στον χρόνο t, όταν η κατάσταση του συστήματος είναι i και λαμβάνεται η απόφαση α.Το διάνυσμα άμεσου κέρδους που αντιστοιχεί στην απόφαση α συμβολίζεται με q^{α} και θεωρείται διάνυσμα στήλη.

$$q^{\alpha}=(q(1,\alpha),q(2,\alpha),\ldots,q(N,\alpha))^{T}.$$

Με ανάλογο τρόπο σε προβλήματα κόστους εισάγεται το άμεσο κόστος (immediate cost) στον χρόνο t, όταν η κατάσταση του συστήματος είναι i και λαμβάνεται η απόφαση α. Το διάνυσμα άμεσου κόστους που αντιστοιχεί στην απόφαση α συμβολίζεται

$$c^{\alpha} = (c(1,\alpha), c(2,\alpha), \ldots, c(N,\alpha))^{T}.$$

Συνοψίζοντας, μία **POMDP**, αναφορικά με πρόβλημα εσόδων περιγράφεται από την εξάδα $(S,A,\Theta,(P^{a})_{a\in A}, (R^{a})_{a\in A}, (q^{\alpha})_{\alpha\in A})$. Για προβλήματα κόστους έχουμε ανάλογη αντιμετώπιση.

1.4. Μετατροπή μίας POMDP σε πλήρως παρατηρήσιμη MDP

Θεωρούμε μία **POMDP** όπως περιγράφθηκε στην ενότητα 1.3. Η κατανομή πιθανότητας της αρχικής κατάστασης του συστήματος θεωρείται γνωστή στον decision maker και συμβολίζεται με

$$\pi(0)= (\pi_1(0),\ldots, \pi_N(0))$$

όπου $\pi_i(0) \equiv p[X_0 = i], i = 1,2,\ldots,N$.

Η ιστορία του συστήματος στον χρόνο t, συμβολίζεται με h_t και περιλαμβάνει όλη την πληροφορία (δεδομένα) που είναι διαθέσιμη πριν από τη λήψη απόφασης στον χρόνο t. Συγκεκριμένα η ιστορία h_t περιλαμβάνει την κατανομή πιθανότητας $\pi(0)$ της αρχικής κατάστασης, τα μηνύματα που πήραμε στους χρόνους 1,2,…,t καθώς και τις αποφάσεις που επιλέχθηκαν στους χρόνους 0,1,…,t-1 ,δηλαδή

$$h_t = (\pi(0), Y_0, Z_1, \ldots, Y_{t-1}, Z_t), t = 1,2,\ldots$$

$$h_0 = \pi(0).$$

Προφανώς $h_t = (h_{t-1}, Y_{t-1}, Z_t), t = 1,2,\ldots$

Το πεδίο τιμών της ιστορίας h_t είναι το σύνολο

$$H_t = \Pi \times (A \times \Theta)^{t}$$

όπου $\Pi = \{x \in \mathbb{R}^{N}: \sum_{i=1}^{N} x_i = 1, x_i \geq 0, i = 1,2,\ldots,N\}$

(το σύνολο των κατανομών πιθανότητας στον χώρο S).

As θεωρήσουμε ένα πρόβλημα εσόδων για πεπερασμένο χρονικό ορίζοντα $T \geq 1$. Αν η συνάρτηση $\upsilon_t(h_t), h_t \in H_t$ δηλώνει το βέλτιστο (μέγιστο) αναμενόμενο ολικό εκπίπτον όφελος από τον χρόνο t μέχρι το πέρας του χρονικού ορίζοντα T ($t \leq T$),τότε:

$$\upsilon_t(h_t) = \max_{Y_t \in A}\{E[q(X_t, Y_t) + \beta \upsilon_{t+1}(h_{t+1}) / h_t, Y_t]\}$$

$$= \max_{Y_t \in A} \{E\,[q(X_t, Y_t) / h_t, Y_t] + \beta . E\,[\upsilon_{t+1}(h_{t+1}) / h_t, Y_t]$$

$$= \max_{Y_t \in A} \left\{ \sum_{i=1}^{N} p\left(X_t = i \middle| h_t\right) . q(i, Y_t) + \beta \cdot \sum_{\theta \in \Theta} p\left(Z_{t+1} = \theta \middle| h_t, Y_t\right) \cdot \upsilon_{t+1}\left(h_t, Y_t, Z_{t+1} = \theta\right) \right\},$$

$$h_t \in H_t \ , \quad t=0,\ldots,T-1 \qquad \underline{1.4.1}$$

όπου β είναι ο παράγοντας έκπτωσης (discount factor). Υποθέτουμε ότι $\beta > 0$. Για τον χρόνο περατώσεως t=T παίρνουμε:

$$\upsilon_T(h_T) = \sum_{i=1}^{N} p[X_T = i / h_T]. q(i), \quad h_T \in H_T, \qquad \underline{1.4.2}$$

όπου $q(i)$ είναι το άμεσο κέρδος τερματισμού (terminal -reward),όταν η κατάσταση του συστήματος είναι η i.

Οι πιθανότητες που υπεισέρχονται στις (1.4.1),(1.4.2) υπολογίζονται στη συνέχεια αυτής της ενότητας.

Αν αντιμετωπίζουμε πρόβλημα κόστους και $\upsilon_t\left(h_t\right), h_t \in H_t$ είναι το βέλτιστο (ελάχιστο) αναμενόμενο ολικό εκπίπτον κόστος από τον χρόνο t έως τον χρονικό ορίζοντα T ($t \le T$),τότε παίρνουμε ανάλογες εκφράσεις με την (1.4.1) με τις προφανείς αλλαγές $c(X_t, Y_t)$ αντί $q(X_t, Y_t)$ και $\min_{Y_T \in A}$ αντί $\max_{Y_T \in A}$. Τέλος αν $c(i)$ είναι το άμεσο κόστος περατώσεως (terminal –cost ή salvage- cost) όταν η κατάσταση του συστήματος είναι η i παίρνουμε αντίστοιχη προς την (1.4.2) σχέση για τον χρόνο t=T. Ο υπολογισμός των συναρτήσεων υ_t γίνεται κατά την ανάδρομη χρονική φορά. Πρώτα υπολογίζεται η υ_T μέσω της (1.4.2) και κατόπιν υπολογίζονται αναγωγικά οι συναρτήσεις $\upsilon_{T-1}, \upsilon_{T-2}, \ldots, \upsilon_0$ μέσω της (1.4.1).Σημειώνουμε ότι η συνάρτηση υ_t στον χρόνο t πρέπει να υπολογισθεί για κάθε δυνατή ιστορία $h_t \in H_t$. Για δοσμένη κατανομή πιθανότητας π(0), το πλήθος των δυνατών ιστοριών στον χρόνο t είναι $\left|H_t\right| = (\left|A\right| . \left|\Theta\right|)^t$.

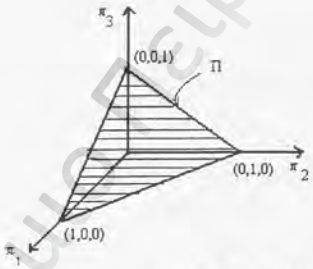
Οι υπολογιστικές απαιτήσεις ενός τέτοιου DP(dynamic-programming) αλγόριθμου ενδέχεται να είναι τεράστιες.(Οι λεπτομέρειες αυτού του προβλήματος σχολιάζονται στον Bertseka [11] και Bertseka-Shreve [115]).

Έστω $\pi_i(t) \equiv p[X_t=i / h_t]$,$i = 1, 2, ...,N.$

και $\pi(t) = (\pi_1(t), \pi_2(t), ..., \pi_N(t))$,

όπου $\sum_{i=1}^{N} \pi_i(t) = 1 \quad , \; 0 \le \pi_i(t) \le 1$.

Η κατανομή .πιθανότητας π(t) της κατάστασης του συστήματος στον χρόνο t καλείται διάνυσμα πληροφορίας «διάνυσμα πληροφορίας» information vector ή (belief-state) σύντομα δ.π .



Σχήμα 1.1: *Το σύνολο Π για N=3.*

Η ακολουθία $\{\pi(t), t \in \mathbb{N}_0\}$ αποτελεί στοχαστική διαδικασία επειδή εξαρτάται από τη στοχαστική διαδικασία της ιστορίας $\{h_t : t \in \mathbb{N}_0\}$.Αποδεικνύεται (Dynkin [28]) ότι:

$$\pi_i(t+1) = p\left(X_{t+1} = i \middle| h_t, Y_t, Z_{t+1}\right) = p\left(X_{t+1} = i \middle| \pi(t), Y_t, Z_{t+1}\right), i \in S$$

και

$$\Pr\left(Z_{t+1} = \theta \middle| h_t, Y_t\right) = \Pr\left(Z_{t+1} = \theta \middle| \pi(t), Y_t\right), \theta \in \Theta.$$

Αν είναι γνωστό το δ.π και η απόφαση στο χρόνο t ,καθώς και το μήνυμα στον χρόνο t+1,μπορούμε να υπολογίσουμε το δ.π στον χρόνο t+1.Πιο συγκεκριμένα, αν $\pi(t) = \pi, Y_t = \alpha, Z_{t+1} = \theta$, εφαρμόζοντας τον κανόνα Bayes' παίρνουμε:

$$T_j(\pi,\theta,\alpha) \equiv \pi_j(t+1) = \Pr(X_{t+1} = j / \pi(t) = \pi, Y_t = a, Z_{t+1} = \theta)$$

$$= \frac{p(Z_{t+1} = \theta / X_{t+1} = j, Y_t = a).p(X_{t+1} = j / \pi(t) = \pi, Y_t = a)}{p(Z_{t+1} = \theta / \pi(t), Y_t = a)}$$

$$= \frac{r^{\alpha}_{j\theta} \sum_{i=1}^{N} \cdot p^{\alpha}_{ij} . \pi_i}{\sum_{K=1}^{N} r^{\alpha}_{k\theta} \sum_{i=1}^{N} \cdot p^{\alpha}_{iK} . \pi_i}, \quad j=1,2,\ldots,N. \qquad \underline{1.4.3}$$

Το διάνυσμα πληροφορίας στον χρόνο t+1

$$T(\pi, \theta, \alpha) = (T_1(\pi(t),\theta,\alpha),\ldots,T_N(\pi(t),\theta,\alpha),$$

γράφεται σε μορφή πινάκων

$$T(\pi,\theta,\alpha) = \frac{\pi \cdot P^{\alpha} \cdot R^{\alpha}_{\theta}}{\pi \cdot P^{\alpha} \cdot R^{\alpha}_{\theta} \cdot 1} \qquad \underline{1.4.4}$$

όπου R^{α}_{θ} είναι ο $N \times N$ διαγώνιος πίνακας με τα διαγώνια στοιχεία (j,j) ίσα με $r^{\alpha}_{j\theta}$, δηλαδή:

$$R^{\alpha}_{\theta} = diag(r^{\alpha}_{1\theta}, r^{\alpha}_{2\theta}, \ldots r^{\alpha}_{N\theta}) = \begin{pmatrix} r^{\alpha}_{1\theta} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & r^{\alpha}_{N\theta} \end{pmatrix}$$

και **1** είναι το $N \times 1$ διάνυσμα στήλη με όλα τα στοιχεία του 1.

Ο παρανομαστής των 1.4.3) και (1.4.4) δηλώνει την πιθανότητα το επόμενο μήνυμα να είναι θ, δεδομένου ότι το τρέχον δ.π είναι π, και η τρέχουσα απόφαση είναι η α. και συμβολίζεται με {θ/π,α}, δηλ.

$$\{\theta/\pi,\alpha\} \equiv p(Z_{t+1} = \theta / \pi(t) = \pi, Y_t = a)$$

$$= \sum_{k=1}^{N} p(Z_{t+1} = \theta / X_{t+1} = k, Y_t = a) . p(X_{t+1} = k / \pi(t) = \pi, Y_t = a)$$

$$= \sum_{k=1}^{N} r^{\alpha}_{k\theta} \cdot \sum_{i=1}^{N} p^{a}_{ik} . \pi_i$$

$$= \pi . P^{\alpha} . R^{\alpha}_{\theta} . \mathbf{1} \qquad \underline{1.4.5}$$

Ισοδύναμα η ποσότητα {θ/π,α} δηλώνει την πιθανότητα το επόμενο δ.π να είναι T(π,θ,α) δοσμένου ότι το τρέχον δ.π είναι , π, και η τρέχουσα απόφαση είναι α, δηλαδή:

$$\{\theta/\pi,\alpha\} = p[\pi(t+1) = T(\pi,\theta,\alpha)/\pi(t) = \pi, Y_t = a].$$

Το δ.π, π(t), είναι μια επαρκής στατιστική για την ιστορία h_t, δηλαδή ενσωματώνει όλη την αναγκαία πληροφορία, προκειμένου να επιλεγεί μια απόφαση στον χρόνο t, βλέπε και (*Bertsekas* [11], *Monahan* [85], *Sondik* [117], *Striebel* [121]).

Επιπλέον ισχύει το ακόλουθο θεώρημα:

Θεώρημα 1.4.1: Για κάθε σταθερή ακολουθία αποφάσεων $Y_0, Y_1, \ldots$, η στοχαστική διαδικασία $\{\pi(t), t \in \mathbb{N}_0\}$ είναι Μαρκοβιανή, δηλαδή αν $\Gamma \subset \Pi$, τότε:

$$p\left(\pi(t+1) \in \Gamma \middle| \pi(0), \pi(1), \ldots \pi(t), Y_t\right) = Pr\left(\pi(t+1) \in \Gamma \middle| \pi(t), Y_t\right).$$

Βλέπε και Aoki [3], Astrom [6]. □

Με βάση τα παραπάνω αποτελέσματα, ένα πρόβλημα POMDP μετατρέπεται σε ένα ισοδύναμο (πλήρως παρατηρήσιμο) πρόβλημα MDP με χώρο καταστάσεων το σύνολο Π των κατανομών πιθανότητας στον χώρο S, το οποίο είναι το (N-1)-simplex του χώρου $\mathbb{R}^N$.

Μια σημαντική ιδιότητα της συνάρτησης μεταφοράς $T(\pi,\theta,\alpha), \pi \in \Pi$ είναι ότι μετασχηματίζει ευθύγραμμα τμήματα σε ευθύγραμμα τμήματα. Συγκεκριμένα ισχύει η ακόλουθη πρόταση.

Πρόταση 1.4.1: Έστω $\theta \in \Theta$, $a \in A$, $\pi^1, \pi^2 \in \Pi$, $0 \le \lambda \le 1$.

Τότε

$$T(\lambda.\pi^1 + (1-\lambda).\pi^2), \theta, \alpha) = \nu.\, T(\pi^1,\theta,\alpha) + (1-\nu).T(\pi^2,\theta,\alpha),$$

όπου $\nu = \lambda.\{\theta/\pi^1,\alpha\} / \lambda.\{\theta/\pi^1,\alpha\} + (1-\lambda).\{\theta/\pi^2,\alpha\}.$ □

Με βάση την παραπάνω πρόταση, το ευθύγραμμο τμήμα $\lambda.\pi^1 + (1-\lambda).\pi^2$, $0 \le \lambda \le 1$ του χώρου Π μετασχηματίζεται μέσω της $T(.,\theta,\alpha)$ *στο ευθύγραμμο τμήμα*

$$\nu.T(\pi^1,\theta,\alpha) + (1-\nu).T(\pi^2,\theta,\alpha), \quad 0 \le \nu \le 1.$$

Στη συνέχεια της ενότητας αυτής θα ορίσουμε τελεστές, που αποτελούν πολύ χρήσιμα εργαλεία στη μελέτη της POMDP.

Με $F(\Pi)$ συμβολίζουμε το σύνολο των πραγματικών συναρτήσεων με πεδίο ορισμού το σύνολο Π,

Με $B(\Pi)$ συμβολίζουμε το σύνολο των φραγμένων πραγματικών συναρτήσεων με πεδίο ορισμού το σύνολο Π,

$$B(\Pi) \subset F(\Pi).$$

Με $\|.\|$ συμβολίζουμε τη νόρμα supremum :

Για $u \in F(\Pi)$, $\|u\| := \sup_{\pi \in \Pi} |u(\pi)|$.

Θεωρούμε τη συνάρτηση $h : \Pi \times A \times B(\Pi) \to \mathbb{R}$, η οποία για προβλήματα εσόδων ορίζεται ως

$$h(\pi,\alpha,u) := \pi.q^{\alpha} + \beta. \sum_{\theta} \{\theta/\pi,\alpha\}.u(T(\pi,\theta,\alpha)),$$

$$(\pi, \alpha, u) \in \Pi \times A \times B(\Pi)$$

ενώ για προβλήματα κόστους ορίζεται ως

$$h(\pi,\alpha,u) := \pi.c^{\alpha} + \beta. \sum_{\theta} \{\theta/\pi,\alpha\}.u(T(\pi,\theta,\alpha)),$$

$$(\pi, \alpha, u) \in \Pi \times A \times B(\Pi),$$

όπου $\beta > 0$ είναι ο συντελεστής έκπτωσης .

Εισάγουμε τώρα τους ακόλουθους τελεστές:

1) *Θεωρούμε τη συνάρτηση ελέγχου* $\delta: \Pi \to A$.

Ο τελεστής $H_\delta : B(\Pi) \longrightarrow F(\Pi)$

ορίζεται ως εξής: Για $u \in B(\pi)$,

$$H_\delta u(\pi) := h(\pi, \delta(\pi), u), \pi \in \Pi.$$

Στην ειδική περίπτωση όπου η συνάρτηση ελέγχου δ είναι σταθερή, $\delta(\pi)=\alpha$ $\forall \pi \in \Pi (\alpha \in A)$, ο τελεστής συμβολίζεται με H_α και έχουμε

$$H_\alpha u(\pi) := h(\pi, \alpha, u), \pi \in \Pi.$$

2) Ο τελεστής $H : B(\Pi) \longrightarrow F(\Pi)$

για προβλήματα εσόδων ορίζεται ως εξής: Για $u \in B(\pi)$,

$$Hu(\pi) := \max_{\alpha \in A} H_\alpha u(\pi)$$

$$= \max_{a \in A} \{\pi.q^a + \beta \sum_{\theta} \{\theta / \pi, a\} .u\ (T(\pi/\theta,\alpha))\},\ \pi \in \Pi.$$

(τελεστής μεγιστοποίησης).

Για προβλήματα κόστους ορίζεται ως

$$Hu(\pi) := \min_{a \in A} H_a u(\pi)$$

$$= \min_{a \in A} \{\pi.c^a + \beta. \sum_{\theta} \{\theta / \pi, a\} .u\ (T(\pi/\theta,\alpha))\},\ \pi \in \Pi.$$

(τελεστής ελαχιστοποίησης).

Αποδεικνύεται εύκολα ότι οι τελεστές H_δ,H έχουν τις ακόλουθες ενδιαφέρουσες και χρήσιμες ιδιότητες:είναι φραγμένοι, ισότονοι και συστολές modulus β.(βλέπε Bertsekas [11]).Πιο συγκεκριμένα,

Αν L= H_δ,H τότε:

i) Φραγμένο: $\| Lu \| \leq \Lambda + \beta . \|u\| < \infty \quad \forall u \in B(\Pi)$,

όπου $\Lambda := \max_{i \in S, a \in A} |q(i,a)|$ για προβλήματα εσόδων,

και $\Lambda := \max_{i \in S, a \in A} |c(i,a)|$ για προβλήματα κόστους.

Επομένως $Lu \in B(\Pi)$, $\forall\ u \in B(\Pi)$.

ii) Ισοτονία: Αν $v,\ u \in B(\Pi)$ *με* $u \geq v$, *τότε* $Lu \geq Lv$.

iii) Συστολή: $\|Lu - Lv\| \leq \beta . \|u - v\| \quad \forall\ v,u \in B(\Pi)$.

Αν για τον συντελεστή έκπτωσης ισχύει $0 < \beta < 1$, συμπεραίνουμε ότι οι τελεστές H_δ , H έχουν μοναδικά σταθερά σημεία (fixed points).

Έστω $w \in B(\Pi)$ το σταθερό σημείο του τελεστή L

(όπου $L = H_\delta$,H), δηλαδή $w = Lw$.

Θεωρώντας την επαναληπτική σχέση

$$w_n = Lw_{n-1},\ n = 1, 2, \ldots$$

η ακολουθία $\{w_n\}$ συγκλίνει ομαλά όταν $n \to \infty$ στο σταθερό σημείο w ,ανεξάρτητα από την επιλογή της αρχικής συνάρτησης $w_0 \in B(\Pi)$.

1.5. Πολιτικές και κριτήρια βελτιστοποίησης για προβλήματα POMDP.

Μία πολιτική ή στρατηγική (policy,strategy) σε μία POMDP *ορίζεται ως ένας μηχανισμός λήψης αποφάσεων στις χρονικές περιόδους t=0,1,....*Σε πλήρη γενικότητα η επιλογή της απόφασης στον χρόνο t μέσω της πολιτικής δ γίνεται με μία κατανομή πιθανότητας η οποία εξαρτάται από την ιστορία του συστήματος $h_t \in H_t$

$$\delta_t(a/h_t), a \in A,$$

όπου $\delta_t(a/h_t) \geq 0, a \in A,$ και $\sum_{a \in A} \delta_t(a/h_t) = 1.$

Μπορούμε να θεωρήσουμε ισοδύναμα , ότι η κατανομή πιθανότητας εξαρτάται από το δ.π $\pi(t) \in \Pi$.

$$\delta_t(a/\pi(t)), a \in A.$$

επειδή το π(t) ενσωματώνει όλη την πληροφορία σχετικά με την ιστορία του συστήματος στον χρόνο t.

Με D συμβολίζουμε την κλάση όλων των πολιτικών.

Ορισμός 1.5.1:Μία πολιτική λέγεται γνήσια ή μη τυχαιοποιημένη(nonrandomized) αν για κάθε χρονική περίοδο t η κατανομή πιθανότητας $\delta_t(a/\pi(t)), a \in A.$ είναι εκφυλισμένη, δηλαδή

$$\delta_t(a/\pi(t)) = 0 \text{ ή } 1, \ a \in A.$$

Σημειώνουμε ότι η παραπάνω εκφυλισμένη κατανομή πιθανότητας μπορεί να εκφρασθεί ως συνάρτηση ελέγχου (control function) $\delta_t : \Pi \to A$

με $$\delta_t(\pi) = \alpha^* \Leftrightarrow \delta_t(a^*/\pi(t) = \pi) = 1.$$

Συμπεραίνουμε ότι μια γνήσια πολιτική δ μπορεί να θεωρηθεί ως χρονική ακολουθία συναρτήσεων ελέγχου $\{\delta_t : t \in \mathbb{N}_0\}$ και παριστάνεται ως

$$\delta = (\delta_0, \delta_1, \ldots).$$

Με D_Γ συμβολίζουμε το σύνολο των γνήσιων πολιτικών.

Ορισμός 1.5.2:Μία γνήσια πολιτική δ καλείται στάσιμη (stationary) αν οι συναρτήσεις ελέγχου στις χρονικές περιόδους t=0,1,... ταυτίζονται: $\delta_t = \delta_0 \ \forall t = 1,2,\ldots,$δηλαδή

$$\delta = (\delta_0, \delta_0, \ldots).$$

Συνήθως μία γνήσια στάσιμη πολιτική συμβολίζεται $\delta^{\infty} = (\delta, \delta, ...)$, όπου δ είναι συνάρτηση ελέγχου $\delta : \Pi \to A$.

Με D_{Σ} συμβολίζουμε το σύνολο των γνήσιων στάσιμων πολιτικών. Προφανώς

$$D_{\Sigma} \subset D_{\Gamma} \subset D.$$

Θα περιγράψουμε εν συντομία δύο κριτήρια βελτιστοποίησης. Περιοριζόμαστε σε POMDP για προβλήματα εσόδων. Η περιγραφή των κριτηρίων αυτών για προβλήματα κόστους είναι ανάλογη.

1) <u>Κριτήριο βελτιστοποίησης για πεπερασμένο χρονικό ορίζοντα.</u>

Θεωρούμε τον χρονικό ορίζοντα $T \geq 1$. Το αναμενόμενο ολικό εκπίπτον κέρδος για τον χρονικό ορίζοντα T, όταν το αρχικό δ.π είναι π(ο)=π και εφαρμόζουμε την πολιτική δ γράφεται:

$$V_T(\pi/\delta) \equiv \mathrm{E}_{\delta}\left[\sum_{t=0}^{T-1} \beta^t . q(X_t, Y_t) + \beta^T . q(X_T)/\pi(o) = \pi\right], \pi \in \Pi. \qquad \underline{1.5.1}$$

όπου $\beta > 0$ είναι ο συντελεστής έκπτωσης και q(j) είναι το (άμεσο) κέρδος τερματισμού (terminal reward), όταν η κατάσταση του συστήματος στον χρόνο περάτωσης T είναι j. Επιθυμούμε να μεγιστοποιήσουμε την (1.5.1), πάνω στην κλάση όλων των πολιτικών **D** και να καθορίσουμε την άριστη πολιτική για την οποία επιτυγχάνεται το παραπάνω μέγιστο.

Έστω $V_n(\pi)$ το βέλτιστο (μέγιστο) αναμενόμενο ολικό εκπίπτον κέρδος, όταν απομένουν $n \leq T$ χρονικές περίοδοι μέχρι το πέρας του χρονικού ορίζοντα T και το δ.π. στον χρόνο T-n είναι π (π(T-n)=π).

Η συνάρτηση $V_n(\pi)$, $\pi \in \Pi$ καλείται βέλτιστη συνάρτηση τιμών για χρονικό ορίζοντα n και υπολογίζεται από την ακόλουθη σχέση του δυναμικού προγραμματισμού. Για n=1,2,...,T

$$V_n(\pi) = \mathrm{H}V_{n-1}(\pi)$$

$$= \max_a \{\pi.q^a + \beta.\sum_{\theta} \{\theta/\pi, \alpha\}.V_{n-1}(T(\pi, \theta, \alpha))\}, \pi \in \Pi \quad \underline{1.5.2}$$

$$V_0(\pi) = \pi.q$$

όπου $q=(q(1),...,q(N))^T$ είναι το διάνυσμα των άμεσων κερδών τερματισμού.

Η παράσταση εντός της αγκύλης στην (1.5.2)

$$\pi.q^{\alpha} + \beta.\sum_{\theta}\{\theta/\pi,\alpha\}V_{n-1}(T(\pi,\theta,\alpha))$$

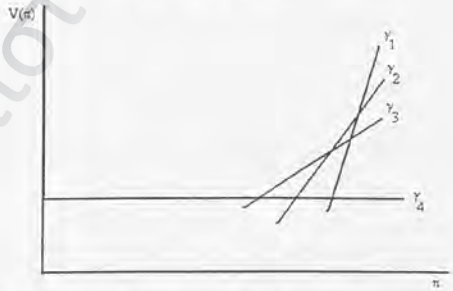
εκφράζει το αναμενόμενο ολικό εκπίπτον κέρδος, όταν απομένουν n χρονικές περίοδοι μέχρι το πέρας του χρονικού ορίζοντα Τ, στην χρονική περίοδο Τ-n ,το δ.π. είναι το π (π(Τ-n)=π),επιλέγεται η απόφαση α ,($Y_{t-n} = a$) και ακολουθείται βέλτιστη πορεία για τις εναπομένουσες n-1 χρονικές περιόδους. Είναι φανερό ότι η άριστη πολιτική για τον πεπερασμένο χρονικό ορίζοντα Τ είναι η γνήσια μη στάσιμη πολιτική $\delta^* = (\delta_T^*, \delta_{T-1}^*, ..., \delta_1^*)$, όπου η συνάρτηση ελέγχου δ_n^* υπολογίζεται από τη σχέση:

$$\delta_n^*(\pi) = \arg\max_{a}\{\pi.q^{\alpha} + \beta.\sum_{\theta}\{\theta/\pi,\alpha\}V_{n-1}(T(\pi,\theta,\alpha))\}, \pi \in \Pi, n = 1,2,...,T,$$

Οι Smallwood and Sondik [119] έδειξαν ότι για $n = 1,2,...,T$,η βέλτιστη συνάρτηση τιμών $V_n(\pi), \pi \in \Pi$ είναι συνεχής, κατά τμήματα γραμμική και κυρτή (piecewise linear and convex)(p.w.l.c),δηλαδή:

$$V_n(\pi) = \max\{\pi.\gamma : \gamma \in \Gamma_n\}, \pi \in \Pi \qquad \underline{1.5.3}$$

όπου Γ_n είναι πεπερασμένο σύνολο διανυσμάτων του χώρου $\mathbb{R}^N$.Τα διανύσματα $\gamma \in \Gamma_n$ θεωρούνται διανύσματα στήλες. Ο Lovejoy [74] εισήγαγε τον όρο «gradients-vectors» για τα διανύσματα γ. Με βάση τη σχέση (1.5.3) η συνάρτηση $V_n(\pi)$, $\pi \in \Pi$ είναι η άνω επιφάνεια (επιγραφή) των υπερεπιπέδων $\pi.\gamma$, $\gamma \in \Gamma_n$.



Σχήμα1.2:Κατά τμήματα γραμμική και κυρτή συνάρτηση με δύο καταστάσεις.

Στα προβλήματα κόστους η συνάρτηση $V_n(\pi)$, $\pi \in \Pi$ του ελάχιστου αναμενόμενου ολικού εκπίπτοντος κόστους *(βέλτιστη συνάρτηση τιμών)* για χρονικό ορίζοντα $n \geq 1$ υπολογίζεται με ανάλογο τρόπο:

$$V_n(\pi) = HV_{n-1}(\pi)$$

$$= \min_a \{\pi.c^a + \beta.\sum_\theta \{\theta/\pi,\alpha\}.V_{n-1}(T(\pi,\theta,\alpha))\}, \pi \in \Pi \qquad \underline{1.5.4}$$

$$V_0(\pi) = \pi.c\,,$$

όπου $c = (c(1),c(2),\ldots,c(N))^T$ είναι το διάνυσμα του άμεσου κόστους τερματισμού. Η άριστη πολιτική για χρονικό ορίζοντα T είναι η γνήσια μη στάσιμη πολιτική $\delta^*=(\delta_T^*, \delta_{T-1}^*, \ldots, \delta_1^*)$, όπου η συνάρτηση ελέγχου δ_n^* υπολογίζεται από τη σχέση:

$$\delta_n^*(\pi) = \arg\min_a \{\pi.c^\alpha + \sum_\theta \{\theta/\pi,\alpha\}V_{n-1}(T(\pi,\theta,\alpha))\}, \pi \in \Pi, n = 1,2,\ldots,T\,.$$

Επίσης η συνάρτηση $V_n(\pi)$ είναι συνεχής, κατά τμήματα γραμμική και κοίλη (piecewise linear and concave).

Δηλαδή

$$V_n(\pi) = \min\{\pi.\gamma : \gamma \in \Gamma_n\}, \pi \in \Pi\,. \qquad \underline{1.5.5}$$

Θα ασχοληθούμε με το κριτήριο αυτό και ειδικότερα με αλγόριθμους υπολογισμού της $V_n(\pi)$,$\pi \in \Pi$ στα κεφάλαια 2 και 3.

2) <u>Κριτήριο βελτιστοποίησης για άπειρο χρονικό ορίζοντα.</u>

Το αναμενόμενο ολικό εκπίπτον κέρδος για άπειρο χρονικό ορίζοντα, όταν το αρχικό δ.π είναι π(o)=π και εφαρμόζουμε την πολιτική δ γράφεται:

$$V(\pi/\delta) \equiv \mathrm{E}_\delta\left[\sum_{t=0}^{\infty} \beta^t.q(X_t,Y_t)/\pi(o) = \pi\right], \pi \in \Pi. \qquad \underline{1.5.6}$$

όπου για τον συντελεστή εκπτώσεως υποθέτουμε ότι $\beta \in (0,1)$. Αποδεικνύεται εύκολα ότι για κάθε $\delta \in D$

$$|\,V(\pi/\delta)\,| \le \frac{\Lambda}{1-\beta}, \quad \forall \pi \in \Pi\,.$$

όπου $\Lambda \equiv \max_{i,a} |q(i,a)|$.

Η συνάρτηση $V(\pi/\delta), \pi \in \Pi$ *αναφέρεται ως συνάρτηση τιμών για την πολιτική δ.* Αποδεικνύεται (Blackwell [15]), ότι υπάρχει γνήσια στάσιμη πολιτική ,η οποία είναι άριστη, δηλαδή μεγιστοποιεί την (1.5.4). Επιπλέον η βέλτιστη συνάρτηση τιμών

$$V^*(\pi) := \sup_{\delta \in D} V(\pi/\delta) = \sup_{\delta^\infty \in D_\Sigma} V(\pi/\delta), \pi \in \Pi,$$

είναι η μοναδική λύση της εξίσωσης βελτιστοποίησης

$$V^*(\pi) = \max_a \{\pi.q^a + \beta.\sum_\theta \{\theta/\pi,\alpha\}.V^*(T(\pi,\theta,\alpha))\}, \pi \in \Pi. \quad \underline{1.5.7}$$

Η παράσταση εντός της αγκύλης στην (1.5.7)

$$\pi.q^a + \beta.\sum_\theta \{\theta/\pi,\alpha\}.V^*(T(\pi,\theta,\alpha))$$

εκφράζει το αναμενόμενο ολικό εκπίπτον κέρδος, όταν στον χρόνο t=0 , το δ.π. είναι το π(0)=π, επιλέγεται η απόφαση $\alpha_1(Y_0 = a)$ και κατόπιν ακολουθείται άριστη πορεία. Η συνάρτηση ελέγχου στην άριστη πολιτική $(\delta^*)^\infty = (\delta^*, \delta^*, ...,)$, προσδιορίζεται από τη σχέση:

$$\delta^*(\pi) = \arg\max_a \{\pi.q^a + \sum_\theta \{\theta/\pi,\alpha\}V^*(T(\pi,\theta,\alpha))\}, \pi \in \Pi.$$

Χρησιμοποιώντας τον τελεστή μεγιστοποίησης Η (βλέπε ενότητα 1.4), η σχέση (1.5.7) γράφεται $V^* = HV^*$ **1.5.8**

Επομένως η βέλτιστη συνάρτηση τιμών V^* για άπειρο χρονικό ορίζοντα είναι το σταθερό σημείο του τελεστή Η. Λαμβάνοντας υπόψη την ιδιότητα της συστολής modulus β του τελεστή Η και την επαναληπτική σχέση (1.5.2), η ακολουθία βέλτιστων συναρτήσεων τιμών για πεπερασμένους χρονικούς ορίζοντες $\{V_n\}$ συγκλίνει ομαλά όταν $n \to \infty$ στην συνάρτηση V^* (βλέπε ενότητα 1.4). Επιπλέον οι ιδιότητες της κυρτότητας και της συνέχειας μεταφέρονται στο όριο. Με άλλα λόγια η συνάρτηση V^* είναι κυρτή και συνεχής.

Με ανάλογο τρόπο στα προβλήματα κόστους η βέλτιστη συνάρτηση τιμών V^* για άπειρο χρονικό ορίζοντα ικανοποιεί την εξίσωση αριστοποίησης:

$$V^*(\pi) = \min_a \{\pi.c^a + \beta.\sum_\theta \{\theta/\pi,\alpha\}V^*(T(\pi,\theta,\alpha))\}, \pi \in \Pi. \quad \underline{1.5.9}$$

Η (1.5.9) γράφεται:

$$V^* = HV^* \quad \underline{1.5.10}$$

όπου Η είναι τελεστής ελαχιστοποίησης.

Επομένως η συνάρτηση V^* είναι το σταθερό σημείο του τελεστή Η. Επίσης η V^* είναι συνεχής και κοίλη συνάρτηση. Σημειώνουμε ακόμη ότι η γνήσια στάσιμη πολιτική

$$(\delta^*)^\infty = (\delta^*, \delta^*, ...)$$

με συνάρτηση ελέγχου: $\delta^*(\pi) = \arg \min_a \{\pi.c^a + \sum_\theta \{\theta / \pi, \alpha\} V^*(T(\pi, \theta, \alpha))\}, \pi \in \Pi$,

είναι άριστη πολιτική. Με μεθόδους προσέγγισης της συνάρτησης V^* και της άριστης πολιτικής $(\delta^*)^\infty$ θα ασχοληθούμε στο κεφάλαιο 4. Καλές αρχικές αναφορές για την δομή των MDPs και επέκταση αυτών στις POMDPs υπάρχουν στις εργασίες των Puterman [98],[99],Bertseka [12],Monahan [85], Lovejoy [74] και κύρια του White [129,130, 131,132,133].

ΣΥΜΠΕΡΑΣΜΑΤΑ

Στο κεφάλαιο αυτό δώσαμε μία σύντομη ανασκόπηση της βιβλιογραφίας των μοντέλων MDP και POMDP.Επίσης περιγράψαμε τα παραπάνω μοντέλα, παρουσιάσαμε τα κύρια κριτήρια βελτιστοποίησης, και συνοψίσαμε βασικά αποτελέσματα που θα χρησιμοποιηθούν στην συνέχεια

ΚΕΦΑΛΑΙΟ 2

Υπολογισμός της άριστης συνάρτησης τιμών για πεπερασμένο χρονικό ορίζοντα σε μία POMDP με την μέθοδο των Smallwood –Sondik- Lovejoy.

Περίληψη

Στο κεφάλαιο αυτό θα περιγράψουμε τον υπολογισμό της άριστης συνάρτησης τιμών V_t για τον χρονικό ορίζοντα t , όταν είναι γνωστή η άριστη συνάρτηση τιμών V_{t-1} για τον χρονικό ορίζοντα t-1 σε μία POMDP για πρόβλημα εσόδων. Οι Smallwood-Sondik [119] έδωσαν αρχικά μια μέθοδο υπολογισμού που στη συνέχεια τροποποιήθηκε από τον Lovejoy [77] με την διόρθωση κάποιων ατελειών της αρχικής μεθόδου. Θα αναφερόμαστε στην τελευταία ως μέθοδο των Smallwood –Sondik-Lovejoy.

Στην ενότητα 2.1 δείχνουμε ότι το πρόβλημα υπολογισμού της V_t από την V_{t-1} ανάγεται στον υπολογισμό της Hυ αν η υ είναι γνωστή κατά τμήματα γραμμική και κυρτή συνάρτηση (p.w.l.c). Πιο συγκεκριμένα, δείχνουμε ότι ο υπολογισμός της Hυ ανάγεται στην εύρεση του συνόλου των λειτουργικών gradient vectors της Hυ αν γνωρίζουμε το σύνολο των λειτουργικών gradient vectors της υ.

Στην ενότητα 2.2 παρουσιάζουμε τον αλγόριθμο του ενός βήματος (one-pass algorithm) των Smallwood-Sondik [117] με τον οποίο υπολογίζεται η τιμή Hυ(π) για δοσμένο δ.π. π.

Στην ενότητα 2.3 περιγράφουμε μία μέθοδο καταγραφής του συνόλου G των " εν δυνάμει" gradient vectors της συνάρτησης Hυ και στη συνέχεια μέθοδο συρρίκνωσης του G σε μια ελάχιστη αντιπροσώπευση από λειτουργικά gradient vectors της Hυ, μέσω γραμμικών προγραμμάτων.

Τέλος στην ενότητα 2.4 προσδιορίζουμε την περιοχή του χώρου Π στην οποία ένα gradient vector για την Hυ είναι λειτουργικό με την μέθοδο των Smallwood-Sondik-Lovejoy μέσω γραμμικών προγραμμάτων.

2.1. Εισαγωγή

Είναι γνωστό ότι η βέλτιστη συνάρτηση τιμών για το πρόβλημα εσόδων σε πεπερασμένο χρονικό ορίζοντα είναι κατά τμήματα γραμμική και κυρτή (βλέπε και ενότητα 1.5).Επομένως η βέλτιστη συνάρτηση τιμών για ένα πρόβλημα t-1 ορίζοντα μπορεί να εκφρασθεί :

$$V_{t-1}(\pi)=\max\{\pi.\gamma:\gamma\in\Gamma_{t-1}\},\pi\in\Pi$$

όπου $\Gamma_{t-1}=\{\gamma^1_{t-1},\gamma^2_{t-1},\ldots\gamma^k_{t-1}\}$ένα πεπερασμένο σύνολο από διανύσματα του R^N που ονομάζονται ''gradient-vectors'' ή απλά ''gradients''.Τα διανύσματα $\pi\in\Pi$ θεωρούνται διανύσματα γραμμές, ενώ τα ''gradients'' διανύσματα στήλες.Ο αλγόριθμος των Smallwood-Sondik [119],ουσιαστικά βασίζεται στην εύρεση του $V_t(\pi)$, όταν είναι γνωστό το $V_{t-1}(\pi)$. Δηλαδή :

$$V_t(\pi)=HV_{t-1}(\pi), \qquad \underline{2.1.1}$$

όπου Η είναι ο τελεστής μεγιστοποίησης (βλ. ενότητα 1.4)

Πρόκειται δηλαδή για μια επέκταση τής μεθόδου των διαδοχικών προσεγγίσεων (value-iteration) από τις Μαρκοβιανές διαδικασίες αποφάσεων (MDP) (βλέπε παράγραφο 1.1) στις μερικά παρατηρήσιμες Μαρκοβιανές διαδικασίες αποφάσεων (POMDP). Η δυσκολία, που αντιμετωπίζουμε οφείλεται κύρια στο γεγονός ότι σε μια POMDP ο χώρος των δ.π ,Π, είναι συνεχής.

Το πρόβλημα ανάγεται στον υπολογισμό της συνάρτησης Hυ αν η συνάρτηση υ είναι κατά τμήματα γραμμική και κυρτή (p.w.l.c). Τότε η υ αντιπροσωπεύεται από ένα πεπερασμένο σύνολο από ''gradient-vectors'',έστω Γ,

$$\Gamma=\{\gamma^1,\gamma^2,\ldots\gamma^k\}$$

$$\upsilon(\pi) = \max_{1 \le j \le k} \pi.\gamma^j, \pi \in \Pi.$$

Ορισμός 2.1.1: Εστω $\gamma \in \Gamma$. Η περιοχή $R(\gamma,\Gamma)=\{\pi \in \Pi : \pi.\gamma \ge \pi.\bar{\gamma}, \forall\, \bar{\gamma} \in \Gamma\} \neq \emptyset$, λέγεται συσχετιζόμενη με το γ περιοχή, (ή υποστηρίζουσα το gradient γ).

Σημειώνουμε ότι η περιοχή $R(\gamma,\Gamma)$ είναι κυρτό πολύεδρο του Π.

Από τον ορισμό συνάγεται ότι :

Αν $\pi \in R(\gamma, \Gamma)$ τότε $\upsilon(\pi) = \pi.\gamma$.

Προφανώς ισχύουν :

$$\bigcup_{\gamma \in \Gamma} R(\gamma,\Gamma) = \Pi \quad \text{και} \quad \text{int}(R(\gamma,\Gamma)) \bigcap \text{int}(R(\bar{\gamma},\Gamma)) = \emptyset, \forall\ \gamma \neq \bar{\gamma}. \qquad \underline{2.1.2}$$

Έτσι ο χώρος Π διαμερίζεται σε συσχετιζόμενες περιοχές των ''gradient vectors'' $\gamma \in \Gamma$. Σημειώνουμε ακόμη ότι η συνάρτηση:

$$H\upsilon(\pi) = \max_{\alpha \in A}\{\pi.q^\alpha + \beta.\sum_{\theta}\{\theta/\pi,\alpha\}.\upsilon(T(\pi,\theta,\alpha)\}, \pi \in \Pi, \qquad \underline{2.1.3}$$

αποδεικνύεται ότι είναι p.w.l.c, βλέπε Smallwood–Sondik [117], και επομένως αντιπροσωπεύεται από ένα πεπερασμένο σύνολο ''gradient vectors'' έστω Γ_H ,έτσι ώστε:

$$H\upsilon(\pi) = \max\{\pi.\gamma' : \gamma' \in \Gamma_H\}. \qquad \underline{2.1.4}$$

Θα μας απασχολήσουν κύρια τρία προβλήματα .

1)Ο υπολογισμός του $H\upsilon(\pi)$ για δοσμένο δ.π $\pi \in \Pi$ (ενότητα 2.2).

2)Ο καθορισμός του συνόλου ,Γ_H, ως ελάχιστη αντιπροσώπευση από λειτουργικά ''gradient vectors'' για την συνάρτηση $H\upsilon(\pi), \pi \in \Pi$,(ενότητα 2.3).

3)Ο καθορισμός των συσχετιζόμενων περιοχών των λειτουργικών ''gradient vectors'' για την συνάρτηση Hυ (ενότητα 2.4).

Η σχέση (2.1.3) γράφεται:

$$H\upsilon(\pi) = \max_{\alpha \in A}\{\pi.q^\alpha + \beta.\sum_{\theta}\{\theta/\pi,\alpha\}.T(\pi,\theta,\alpha).\gamma^{l(\pi,\theta,\alpha)}\}$$

$$= \max_{\alpha \in A}\{\pi.(q^\alpha + \beta.P^\alpha.\sum_{\theta} R_\theta^\alpha . \gamma^{l(\pi,\theta,\alpha)})\}$$

$$= \max_{\alpha \in A} \{\pi.\gamma_\alpha(\pi)\}, \pi \in \Pi,$$

όπου $\gamma_\alpha(\pi) = q^\alpha + \beta.P^\alpha. \sum_\theta R_\theta^\alpha . \gamma^{l(\pi,\theta,\alpha)}$, $\forall \pi \in \Pi$, $\forall \alpha \in A$.

Ο δείκτης l(π,θ,α) είναι ο δείκτης μεγιστοποίησης στη σχέση:

$$T(\pi,\theta,\alpha).\gamma^{l(\pi,\theta,\alpha)} = \max_{1 \le j \le k} T(\pi,\theta,\alpha).\gamma^j,$$

η ισοδύναμα

$$\pi. P^\alpha.R_\theta^\alpha . \gamma^{l(\pi,\theta,\alpha)} = \max_{1 \le j \le k} \pi. P^\alpha.R_\theta^\alpha . \gamma^j \qquad \underline{2.1.5}$$

Με τον παρακάτω αλγόριθμο A_3, που είναι γνωστός σαν αλγόριθμος του ενός βήματος και οφείλεται στον Sondik [119] μπορούμε να υπολογίσουμε την τιμή $Hv(\tilde{\pi})$ για δοσμένο $\tilde{\pi} \in \Pi$.

2.2. Αλγόριθμος τού ενός βήματος

Σκοπός του αλγορίθμου του ενός βήματος (one–pass algorithm) είναι η εύρεση λειτουργικού «gradient» της Hv σε κάποιο δ.π , $\tilde{\pi}$, καθώς και της βέλτιστης απόφασης στο συγκεκριμένο $\tilde{\pi}$.Η ονομασία «αλγόριθμος του ενός βήματος», σχετίζεται με την εφαρμογή του στο ενός βήματος πέρασμα από τον χρονικό ορίζοντα t-1 στον χρονικό ορίζοντα t.

Αλγόριθμος (τού ενός βήματος) A_3 (Sondik [119])

ΒΗΜΑ 1: Για $\theta \in \Theta$, $\alpha \in A$ βρίσκουμε τον δείκτη $l(\tilde{\pi},\theta,\alpha)$ έτσι ώστε:

$$\tilde{\pi}.P^\alpha.R_\theta^\alpha.\gamma^{l(\tilde{\pi},\theta,\alpha)} = \max_{1 \le j \le k} \tilde{\pi}.P^\alpha R_\theta^\alpha \gamma^j$$

ΒΗΜΑ 2: Υπολογίζουμε τα gradients:

$$\gamma_\alpha(\tilde{\pi}) = q^\alpha + \beta.P^\alpha. \sum_\theta R_\theta^\alpha \gamma^{l(\tilde{\pi},\alpha,\theta)}, \alpha \in A.$$

ΒΗΜΑ3: Υπολογίζουμε την ποσότητα $Hv(\tilde{\pi})$ από τη σχέση:

$$Hv(\tilde{\pi}) = \max_\alpha \tilde{\pi}.\gamma_\alpha(\tilde{\pi}) = \tilde{\pi}.\gamma_{\alpha^*}(\tilde{\pi}).$$

Με τον παραπάνω αλγόριθμο πετυχαίνεται ο καθορισμός της άριστης απόφασης α^* *καθώς και το λειτουργικό gradient* $\gamma_{\alpha^*}(\tilde{\pi})$ της Hu *για το* ***συγκεκριμένο δ.π*** $\tilde{\pi}$.

2.3. Προσδιορισμός του συνόλου Γ_H των λειτουργικών gradient vectors για την συνάρτηση Hu.

Αποδεικνύεται ότι για δοσμένα, $\theta \in \Theta, \alpha \in A$ η συνάρτηση $l(.,\theta,\alpha)$ επάγει μία διαμέριση από κυρτές περιοχές (πολύεδρα) του χώρου Π, έτσι ώστε η $l(\pi,\theta,\alpha)$ να έχει μια μοναδική τιμή σε κάθε μια από αυτές. Η απόδειξη στηρίζεται στο γεγονός ότι η συνάρτηση $T(.,\theta,\alpha)$ μετασχηματίζει ευθύγραμμα τμήματα σε ευθύγραμμα τμήματα (βλέπε και πρόταση 1.4.1).

Συμβολίζουμε με $\mathbf{S}_{\alpha,\theta}$ την εν λόγω διαμέριση και $S^l_{\alpha,\theta}$, l=1,2,…,k τις περιοχές που την αποτελούν:

$$\mathbf{S}_{\alpha,\theta} = \{ S^1_{\alpha,\theta}, S^2_{\alpha,\theta}, \ldots, S^k_{\alpha,\theta} \}.$$

Ενδεχομένως κάποιες από τις παραπάνω περιοχές είναι κενές. Έτσι λοιπόν

$$\pi \in S^l_{\alpha,\theta} \Rightarrow l(\pi,\theta,\alpha) = l.$$

Επομένως

$$S^l_{\alpha,\theta} := \{\pi \in \Pi : T(\pi,\theta,\alpha).\gamma^l \geq T(\pi,\theta,\alpha).\gamma^j,\ 1 \leq j \leq k\} =$$

$$= \{\pi \in \Pi : \pi.P^\alpha.R^\alpha_\theta.\gamma^l \geq \pi.P^\alpha.R^\alpha_\theta.\gamma^j,\ 1 \leq j \leq k\} \qquad \underline{2.3.1}$$

Για δοσμένο $\alpha \in A$, σχηματίζονται οι διαμερίσεις $\mathbf{S}_{\alpha,1}, \mathbf{S}_{\alpha,2}, \ldots, \mathbf{S}_{\alpha,M}$ του χώρου Π, (όπου $M = |\Theta|$). Μπορούμε να σχηματίσουμε μια διαμέριση γινόμενο από αυτές τις διαμερίσεις $\mathbf{S}_{\alpha,\theta}$ δηλαδή:

$$\mathbf{S}_\alpha = \otimes_{\theta \in \Theta} \mathbf{S}_{\alpha,\theta} = \mathbf{S}_{\alpha,1} \otimes \mathbf{S}_{\alpha,2} \otimes \ldots \otimes \mathbf{S}_{\alpha,M}$$

που αποτελείται από τομές της μορφής

$$\mathbf{S}^s_\alpha = \bigcap_{\theta=1}^{M} S^{l_\theta}_{\alpha,\theta} \qquad s=(l_1,l_2,\ldots,l_M) \quad \text{όπου} \quad 1 \leq l_\theta \leq k, \theta = 1,2,3,\ldots M. \qquad \underline{2.3.2}$$

Οι τομές αυτές αποτελούν κυρτές περιοχές (κυρτά πολύεδρα) του χώρου Π.

Σημειώνουμε επίσης ότι για όλα τα $\pi \in \text{int}\, S_a^s$ αντιστοιχεί σταθερό gradient $\gamma_\alpha(\pi)$,

$$\gamma_\alpha(\pi)=q^\alpha+\beta.P^\alpha.\sum_\theta R_\theta^\alpha\ \gamma^{l(\pi,\theta,\alpha)}=q^\alpha+\beta.P^\alpha.\sum_\theta R_\theta^\alpha\ \gamma^{l_\theta},$$

που το συμβολίζουμε με ξ_a^s.

Δηλαδή $$\gamma_\alpha(\pi)=\xi_a^s=q^\alpha+\beta.P^\alpha.\sum_\theta R_\theta^\alpha\ \gamma^{l_\theta}, \qquad \underline{2.3.3}$$

όπου $$s=(l_1,l_2,\ldots,l_M).$$

Συμπεραίνουμε ότι για κάθε $\pi \in \text{int}\,(S_a^s)$, έχουμε:

$$H_\alpha \upsilon(\pi)=\pi.\xi_a^s$$

Πράγματι, $$H_\alpha \upsilon(\pi)=\pi.q^\alpha+\beta.\sum_\Theta\{\theta/\pi,\alpha\}.\upsilon(T(\pi,\theta,\alpha))$$

$$=\pi.(q^\alpha+\beta.P^\alpha.\sum_\theta R_\theta^\alpha.\gamma^{l(\pi,\alpha,\theta)})$$

$$=\pi.\gamma_\alpha(\pi)=\pi.\xi_a^s. \qquad \underline{2.3.4}$$

Εστω τώρα $$G=\bigcup_a G^\alpha,$$

Όπου $G^\alpha=\{\zeta_a^s:=q^\alpha+\beta.P^\alpha.\sum_\theta R_\theta^\alpha.\gamma^{j_\theta}: s=(j_1,j_2,\ldots,j_M), 1\le j_\theta\le k, 1\le\theta\le M\}, \alpha\in A$.

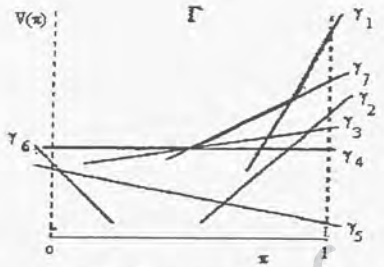
Κάθε «gradient» του συνόλου G^α, λέμε ότι έχει φέρουσα απόφαση την α.

Σημειώνουμε ότι G είναι το σύνολο των ΄΄εν δυνάμει΄΄*«gradient-vectors» για την συνάρτηση* H υ

Τότε $$H\upsilon(\pi)=\max\{\pi\gamma:\gamma\in G\}, \pi\in\Pi. \qquad \underline{2.3.5}$$

Οι πληθάριθμοι των συνόλων G^α, $\alpha\in A$ και G είναι αντίστοιχα

$$|G^\alpha|=|\Gamma|^{|\Theta|}=k^M,\ \alpha\in A \text{ και } |G|=|A|k^M.$$



Σχήμα 2.1:Ένα παράδειγμα μιας P.W.L.C συνάρτησης τιμών με άχρηστα (useless-vectors) τα $\gamma_2,\gamma_3,\gamma_5,\gamma_6$ στην ελάχιστη αντιπροσώπευση της.

Το παραπάνω σύνολο G ενδέχεται να είναι ένα πολυπληθές σύνολο. Αντικειμενικός μας σκοπός,είναι να οικοδομήσουμε ένα σύνολο Γ_H ελάχιστο στην αντιπροσώπευση της Η υ, δηλαδή ένα σύνολο:

$$\Gamma_H=\{\gamma: \gamma \in G \text{ και } R(\gamma,G)\neq \emptyset \} \qquad \underline{2.3.6}$$

Τότε βέβαια: $$[H\upsilon](\pi)=\max\{\pi\gamma: \gamma\in G\}$$

$$=\max\{\pi\gamma: \gamma\in\Gamma_H\}. \qquad \underline{2.3.7}$$

Με άλλα λόγια, ένα εν δυνάμει «gradient» είναι λειτουργικό και τοποθετείται στο σύνολο Γ_H, όταν η συσχετιζόμενη περιοχή $R(\gamma,G)\neq\emptyset$.

Η μέθοδος του γραμμικού προγραμματισμού μπορεί να χρησιμοποιηθεί για την περικοπή των μη λειτουργικών *«gradient-vectors» του συνόλου G με σκοπό να επιτύχουμε την ελάχιστη αντιπροσώπευση* Γ_H για την Hυ.

Έστω $\gamma' \in G$.Θεωρούμε το γραμμικό πρόγραμμα (γ.π)

$$z^* = \max z$$

υπό τους περιορισμούς

$$z \le \pi.(\gamma' - \gamma) \quad \forall \gamma \in G$$

$$\sum_{i=1}^{N} \pi_i = 1 , \quad \pi_i \ge 0, \quad 1 \le i \le N .$$

Το παραπάνω γ.π έχει N+1μεταβλητές (τις $z,\pi_1,\pi_2,\ldots,\pi_N$) και $|G|+N+1$ περιορισμούς. Το *gradient-vector* γ' είναι λειτουργικό ($\gamma' \in \Gamma_H$) αν και μόνον αν $z^* = 0$. Επομένως επιλύοντας το γ.π. τοποθετούμε το γ' *στο σύνολο* Γ_H αν $z^* = 0$. Επαναλαμβάνουμε την ίδια διαδικασία για όλα τα *gradient-vectors του συνόλου G, πράγμα που συνεπάγεται την επίλυση* $|G|$ το πλήθος γ.π. για τον καθορισμό του συνόλου Γ_H. Επισημαίνουμε την υπολογιστική δυσκολία που οφείλεται στο μεγάλο μέγεθος του συνόλου G ($|G|=|A|.k^M$).

2.4. Προσδιορισμός της υποστηρίζουσας περιοχής $R(\gamma^*,\Gamma_H)$

Εφαρμόζοντας τον αλγόριθμο του ενός βήματος της ενότητας 2.2 μπορούμε να βρούμε ένα συγκεκριμένο gradient ,γ^* ,που είναι λειτουργικό στο $\tilde{\pi}$ και την αντίστοιχη άριστη απόφαση α^* , ($\gamma^* = \gamma_{\alpha^*}(\tilde{\pi})$)

Στην τρέχουσα ενότητα έχουμε σαν αντικείμενο να βρούμε την περιοχή $R(\gamma^*,\Gamma_H)$,όπου το συγκεκριμένο gradient ,γ^* ,είναι λειτουργικό. Φανταζόμαστε προς τούτο ότι μετακινούμαστε σε ένα δ.π , π, λίγο πιο μακριά του $\tilde{\pi}$,και υπολογίζουμε με βάση τον αλγόριθμο του ενός βήματος το λειτουργικό gradient γ για το π. Εξακολουθούμε να μετακινούμαστε μέχρι να πετύχουμε $\gamma \neq \gamma^*$.
Σύμφωνα με τον Sondik [119] υποδεικνύονται δύο δυνατότητες να επιτευχθεί το παραπάνω $\gamma \neq \gamma^*$ μεταβαίνοντας από το $\tilde{\pi}$ στο π.

i) Να αλλάξει ο δείκτης $l(\tilde{\pi},\theta,\alpha^*)$ σε $l(\pi,\theta,\alpha^*) \neq l(\tilde{\pi},\theta,\alpha^*)$ για κάποιο $\theta \in \Theta$

ή

ii) Να αλλάξει η βέλτιστη απόφαση α^* που αντιστοιχεί στο $\tilde{\pi}$.

Για δοσμένο θ, η περιοχή του χώρου Π για την οποία ισχύει $l(\tilde{\pi},\theta,\alpha^*)= l(\pi,\theta,\alpha^*)$ δίνεται από τη σχέση : $\pi.P^{\alpha^*}.R_\theta^{\alpha^*}.(\gamma^{l(\tilde{\pi},\theta,\alpha^*)}-\gamma^j) \geq 0$ με $1 \leq j \leq k$.

Επαναλαμβάνοντας την ίδια διαδικασία για όλα τα θ=1,2,…,M η συνθήκη $l(\tilde{\pi},\theta,\alpha^*)= l(\pi,\theta,\alpha^*)$, $1 \leq \theta \leq M$ ισχύει αν και μόνο αν :

$$\pi.P^{\alpha^*}.R_\theta^{\alpha^*}.(\gamma^{l(\tilde{\pi},\theta,\alpha^*)}-\gamma^j) \geq 0 \quad \text{με } 1 \leq j \leq k, 1 \leq \theta \leq M \qquad \underline{2.4.1}$$

Οι γραμμικές ανισότητες (2.4.1) ορίζουν την περιοχή που η δυνατότητα (i) αποκλείεται. Σύμφωνα πάντα με τον Sondik [119], η απόφαση α^* είναι βέλτιστη στα δ.π της συσχετιζόμενης περιοχής του $\gamma^*, R(\gamma^*, \Gamma_H)$ αν

$$\pi.(\gamma^*-\gamma(\alpha)) \geq 0 \qquad \forall\ \alpha \in A, \qquad \underline{2.4.2}$$

όπου :

$$\gamma(\alpha)= q^{\alpha} +\beta.P^{\alpha} \sum_{\theta} R_{\theta}^{\alpha}\ \gamma^{l(\tilde{\pi},\theta,\alpha)}.$$

Σύμφωνα με τον παραπάνω συμβολισμό έχουμε ότι:

$$\gamma^* = \gamma(\alpha^*)= q^{\alpha^*} + \beta.P^{\alpha^*} . \sum_{\theta\in\theta} R_{\theta}^{\alpha^*}\ .\ \gamma^{l(\tilde{\pi},\theta,\alpha^*)}.$$

Ο Sondik [119] είχε θεωρήσει μόνο τις ανισότητες (2.4.1) και (2.4.2) για να προσδιορίσει την περιοχή $R(\gamma^*,\Gamma_H)$,το οποίο δεν είναι ορθό, όπως επισημάνθηκε από τον Lovejoy [77],διότι ενδέχεται σε κάποιο π,καθώς απομακρυνόμαστε από το $\tilde{\pi}$ να ικανοποιείται η ανίσωση:

$\pi.(\gamma^*-\gamma(\alpha)) \geq 0\ \ \forall\ \alpha \in A$ και να έχουμε ωστόσο $l(\pi,\theta,\alpha) \neq l(\tilde{\pi},\theta,\alpha)$, για κάποιο $\theta \in \Theta$, οπότε για το «gradient» $\gamma'(\alpha)= q^{\alpha} +\beta.P^{\alpha} \sum_{\theta} R_{\theta}^{\alpha}\ \gamma^{l(\pi,\theta,\alpha)}$ μπορεί να ισχύει η ανισότητα:

$$\pi.\gamma'(\alpha) > \pi.\gamma^* > \pi.\gamma(\alpha) \quad \text{για κάποιο } \alpha \in A.$$

Ο *Lovejoy* [77],τεκμηρίωσε το παραπάνω σφάλμα με αριθμητικό παράδειγμα. Επομένως οι σχέσεις (2.4.1) και (2.4.2) δεν επαρκούν για τον προσδιορισμό της συσχετιζόμενης περιοχής $R(\gamma^*,\Gamma_H)$ του γ^*. *Σύμφωνα με τον Lovejoy,σε περίπτωση που δεν γνωρίζουμε το σύνολο Γ_H των λειτουργικών «gradients» για την Hυ, η περιοχή $R(\gamma^*,\Gamma_H)$ καθορίζεται από τα $\pi \in \Pi$ για τα οποία:*

$$\pi.(\gamma^*-\gamma) \geq 0 \quad \forall \gamma \in G. \qquad \underline{2.4.3}$$

όπου G είναι το σύνολο των εν δυνάμει *«gradient vectors» για την Hυ* (βλέπε ενότητα 2.3).Σημειώνουμε ότι οι ανισώσεις (2.4.1) και (2.4.2) υπάγονται στην (2.4.3).Γενικά μόνο ένα υποσύνολο των περιορισμών (2.4.3) είναι απαραίτητο για τον καθορισμό της *περιοχής $R(\gamma^*,\Gamma_H)$.Για την εύρεση αυτού του* υποσυνόλου εφαρμόζουμε τη μέθοδο του γραμμικού προγραμματισμού.

Έστω $\gamma' \in G$.Θεωρούμε το ακόλουθο γραμμικό πρόγραμμα (γ.π)

$$z = \min_{\pi} \pi.(\gamma^* - \gamma')$$

υπό τους περιορισμούς

$$\pi.(\gamma^* - \gamma) \geq 0 \ , \forall \gamma \in G.$$

$$\sum_{i=1}^{N} \pi_i = 1 \ ,$$

$$\pi_i \geq 0, \quad 1 \leq i \leq N \ .$$

Αυτό το γ.π έχει N μεταβλητές (τις $\pi_1, \pi_2, \ldots, \pi_N$) και $|G| + N + 1$ περιορισμούς. Αν z=0 τότε ο περιορισμός

$$\pi.(\gamma^* - \gamma') \geq 0 \qquad \underline{\textbf{2.4.4}}$$

είναι απαραίτητος για τον καθορισμό τής *περιοχής* $R(\gamma^*, \Gamma_H)$. Αντιθέτως αν $z \succ 0$, ο περιορισμός (2.4.4) μπορεί να παραλειφθεί ως πλεονάζων.

Επαναλαμβάνουμε την ίδια διαδικασία για όλα τα *gradient-vectors του συνόλου G, πράγμα που συνεπάγεται την επίλυση* $|G|$ το πλήθος γ.π. προκειμένου να βρούμε το υποσύνολο των περιορισμών στην (2.4.3) που είναι απαραίτητοι για τον καθορισμό του συνόρου του κυρτού πολυέδρου $R(\gamma^*, \Gamma_H)$. *Στο σημείο* αυτό επισημαίνουμε την υπολογιστική δυσκολία που συνδέεται με το μέγεθος του συνόλου G, ($|G| = |A| k^M$).

Αν το σύνολο Γ_H των λειτουργικών gradients για την Hυ είναι γνωστό (π.χ. με την μέθοδο που περιγράφηκε στην ενότητα 2.3), τότε από τον ορισμό 2.1.1 η περιοχή $R(\gamma^*, \Gamma_H)$ καθορίζεται από τα $\pi \in \Pi$ για τα οποία

$$\pi.(\gamma^* - \gamma) \geq 0 \ , \forall \gamma \in \Gamma_H. \qquad \underline{\textbf{2.4.5}}$$

Ομοίως στην περίπτωση αυτή εν γένει μόνο ένα υποσύνολο των περιορισμών (2.4.5) είναι απαραίτητο για τον καθορισμό του συνόρου της *περιοχής* $R(\gamma^*, \Gamma_H)$. Το υποσύνολο αυτό προσδιορίζεται με παρόμοια διαδικασία επίλυσης $|\Gamma_H|$ το πλήθος γ.π. Τέλος επισημαίνουμε ότι η άριστη απόφαση για την περιοχή $R(\gamma^*, \Gamma_H)$ *είναι η φέρουσα απόφαση* α^* *του gradient-vector* γ^*. Πράγματι, για κάθε $\pi \in R(\gamma^*, \Gamma_H)$ έχουμε

$$Hu(\pi) = \max\{\pi\,\gamma : \gamma \in \Gamma_H\} = \pi.\gamma^* .$$

ΣΥΜΠΕΡΑΣΜΑΤΑ

Στο κεφάλαιο αυτό περιγράφουμε τη μέθοδο των Smallwood –Sondik- Lovejoy για τον υπολογισμό της συνάρτησης Hυ αν υ είναι γνωστή κατά τμήματα γραμμική και κυρτή συνάρτηση. Ο υπολογισμός της Hυ ανάγεται στον καθορισμό του συνόλου Γ_H των λειτουργικών gradients της Hυ. Ειδικότερα:

- Παρουσιάζουμε τον αλγόριθμο του ενός βήματος των Smallwood –Sondik ,με τον οποίο υπολογίζεται η τιμή Hυ(π) για δοσμένο δ.π. π.
- Προσδιορίζουμε το σύνολο Γ_H περικόπτοντας μη λειτουργικά gradients από το σύνολο G των εν δυνάμει gradient-vectors της Hυ επιλύοντας $|G|$ το πλήθος γραμμικά προγράμματα.
- Για κάθε λειτουργικό gradient-vector της Hυ προσδιορίζουμε το σύνορο της συσχετιζόμενης περιοχής με την επίλυση γραμμικών προγραμμάτων.
- Επισημαίνουμε τις υπολογιστικές δυσκολίες, που οφείλονται στο μεγάλο μέγεθος του συνόλου G.

Στην κατεύθυνση λοιπόν αναζήτησης μιας εναλλακτικής μεθόδου υπολογισμού της Hυ, παρακάμπτοντας το σύνολο G,προτείνεται ένας νέος αλγόριθμος στο επόμενο κεφάλαιο.

ΚΕΦΑΛΑΙΟ 3

Αλγόριθμος των ακροτάτων σημείων

Περίληψη

Σκοπός του κεφαλαίου αυτού, είναι η παρουσίαση ενός νέου αλγόριθμου, για τον υπολογισμό της βέλτιστης συνάρτησης οφέλους σύμφωνα με το κριτήριο του αναμενόμενου εκπίπτοντος ολικού οφέλους, για πεπερασμένο χρονικό ορίζοντα στα πλαίσια της επαναληπτικής διαδικασίας τιμών (value-iteration). *Η βέλτιστη συνάρτηση οφέλους αντιπροσωπεύεται από ένα σύνολο διανυσμάτων* «gradients», που δεν είναι γνωστό από την αρχή αλλά χτίζεται σε διαδοχικά βήματα.

Αρχικά επιλέγονται αυθαίρετα κάποια διανύσματα πληροφορίας δ.π και μέσω γνωστών μεθόδων (αλγόριθμος ενός βήματος) υπολογίζονται τα αντίστοιχα λειτουργικά «gradients».Για καθένα από αυτά, προσδιορίζεται η αντίστοιχη διευρυμένη περιοχή δηλαδή η κυρτή περιοχή του χώρου των διανυσμάτων πληροφορίας, στα οποία το συγκεκριμένο «gradient» είναι λειτουργικό έναντι των άλλων.Ακολούθως υπολογίζεται το μέγιστο σφάλμα προσέγγισης, το οποίο όπως αποδεικνύεται, επιτυγχάνεται σε κάποιο ακρότατο (κορυφή) των διευρυμένων περιοχών.

Το αρχικό σύνολο «gradients» εμπλουτίζεται με νέα «gradients» που είναι λειτουργικά για την κορυφή στην οποία εμφανίζεται το μέγιστο σφάλμα προσέγγισης. Κατασκευάζονται νέες διευρυμένες περιοχές και αποδεικνύεται ότι το νέο μέγιστο σφάλμα προσέγγισης είναι μειωμένο σε σχέση με το προηγούμενο. Η διαδικασία συνεχίζεται μέχρις ότου το μέγιστο σφάλμα προσέγγισης μηδενισθεί ή γίνει αρκούντως μικρό,(δηλαδή μικρότερο ή ίσο από ένα προκαθορισμένο σφάλμα).Ο αλγόριθμος των ακροτάτων σημείων αναφέρεται στο πέρασμα από έναν χρονικό ορίζοντα στον επόμενο.Υπολογίζεται το συσσωρευμένο σφάλμα προσέγγισης για κάθε χρονικό ορίζοντα από μια απλή αναγωγική σχέση. Εξετάζεται επίσης πως ο προτεινόμενος

αλγόριθμος μπορεί να καλύψει την περίπτωση του υπολογισμού της συνάρτησης του ελαχίστου κόστους για πεπερασμένο χρονικό ορίζοντα.

3.1. Οι διευρυμένες περιοχές

As υποθέσουμε ότι για κάποιο χρονικό ορίζοντα $t \geq 1$ η συνάρτηση V_t του μέγιστου αναμενόμενου ολικού οφέλους είναι γνωστή. Τίθεται το πρόβλημα του υπολογισμού της συνάρτησης του μέγιστου αναμενόμενου ολικού οφέλους για τον χρονικό ορίζοντα t+1,

$$V_{t+1}=HV_t ,$$

(όπου H είναι τελεστής μεγιστοποίησης)

Γενικότερα έστω $\upsilon(\pi), \pi \in \Pi$ μια κατά τμήματα γραμμική και κυρτή συνάρτηση. Επομένως η συνάρτηση υ αντιπροσωπεύεται μέσω ενός πεπερασμένου συνόλου Γ από «gradients» και

$$\upsilon(\pi) = \max \{\pi.\gamma : \gamma \in \Gamma\}, \pi \in \Pi,$$

όπου συμβατικά τα δ.π $\pi \in \Pi$ θεωρούνται διανύσματα – γραμμές και τα «gradients» διανύσματα – στήλες. Ο χώρος Π διαμερίζεται σε κυρτές περιοχές $W_\gamma \gamma \in \Gamma$, έτσι ώστε:

$$\upsilon(\pi)= \pi.\gamma \quad \forall \pi \in W_\gamma .$$

Η συνάρτηση

$$H\upsilon(\pi) = \max_{a \in A} \left\{ \pi.q^a + \beta.\sum_{\theta \in \Theta} \{\theta / \pi, \alpha\}.\upsilon(\mathrm{T}(\pi, \theta, \alpha)) \right\}, \pi \in \Pi, \qquad \underline{3.1.1}$$

όπου q^a το διάνυσμα άμεσου οφέλους, που αντιστοιχεί στην απόφαση α (το οποίο θεωρείται διάνυσμα στήλη) και β ο συντελεστής έκπτωσης (discound-factor) $(0 < \beta \leq 1)$, είναι επίσης κατά τμήματα γραμμική και κυρτή.

Το πρόβλημα του υπολογισμού της συνάρτησης Hυ ανάγεται στον καθορισμό του συνόλου των λειτουργικών «gradients» Γ_H για την συνάρτηση Hυ και των αντίστοιχων περιοχών αντιπροσώπευσης της Hυ. Για να βρούμε το παραπάνω σύνολο των «gradients» Γ_H που υποστηρίζουν την Hυ ακολουθούμε την εξής πορεία. Επιλέγουμε αυθαίρετα ένα πεπερασμένο σύνολο δ.π. (Για παράδειγμα μπορούμε να επιλέξουμε τις κορυφές $e_1 = (1,0,...,0), e_2 = (0,1,...,0), \ldots e_N = (0,0,...,1)$ του χώρου Π των δ.π.

Εφαρμόζοντας τον αλγόριθμο A_3 του ενός βήματος του κεφαλαίου 2 υπολογίζουμε τα λειτουργικά «gradients» της Hυ που αντιστοιχούν σε αυτά τα διανύσματα πληροφορίας. As συμβολίσουμε με $\widetilde{\Gamma}_H$ το παραπάνω σύνολο των «gradients» της Hυ. Ορίζουμε τον τελεστή προσέγγισης $\widetilde{H}$:

$$\widetilde{H}\,\upsilon(\pi):=\max\{\pi.\gamma : \gamma\in\widetilde{\Gamma}_H\},\ \forall\ \pi\in\Pi. \qquad \underline{3.1.2}$$

Αν $\widetilde{H}\upsilon$ χρησιμοποιείται για να προσεγγίσει την Hυ, τότε το μέγιστο σφάλμα για την παραπάνω προσέγγιση είναι το:

$$\max_{\pi\in\Pi}\{H\upsilon(\pi)-\widetilde{H}\,\upsilon(\pi)\}. \qquad \underline{3.1.3}$$

Αν το παραπάνω σφάλμα δεν είναι 0,τότε ένα κατάλληλο «gradient» πρέπει να επιλεγεί και να συμπεριληφθεί στο $\widetilde{\Gamma}_H$, ώστε να φθάσουμε σε μια καλύτερη προσέγγιση. Αυτή η διαδικασία θα συνεχισθεί, μέχρις ότου το μέγιστο σφάλμα να γίνει μικρότερο ή ίσο από έναν προκαθορισμένο αριθμό $\varepsilon>0$ αρκετά μικρό, αν επιθυμούμε προσεγγιστικό αλγόριθμο, ενώ αν επιθυμούμε ακριβή αλγόριθμο, μέχρι μηδενισμού του σφάλματος.

Ορισμός 3.1.1: Έστω $\hat{\gamma}\in\widetilde{\Gamma}_H$. Ορίζουμε **"διευρυμένη περιοχή"** του «gradient» $\hat{\gamma}$ το κυρτό πολύεδρο του χώρου R^n:

$$\widetilde{R}\hat{\gamma}=\{\pi\in\Pi:\pi.\hat{\gamma}\geq\pi.\gamma\quad\forall\gamma\in\widetilde{\Gamma}_H\}. \qquad \underline{3.1.4}$$

Πρόκειται δηλαδή για την περιοχή, όπου το $\hat{\gamma}$ είναι "νικητής" έναντι των άλλων «gradients» του $\widetilde{\Gamma}_H$.

Το σύνολο των λειτουργικών «gradients» Γ_H που αντιπροσωπεύουν την συνάρτηση Hυ δεν είναι φυσικά γνωστό από την αρχή. Ξεκινώντας αρχικά με το σύνολο $\widetilde{\Gamma}_H\subset\Gamma_H$, με διαδοχικά βήματα το εμπλουτίζουμε με νέα λειτουργικά «gradients», όπως θα περιγράψουμε στη συνέχεια.

Εστω $\hat{\gamma}\in\widetilde{\Gamma}_H$. Η συσχετιζόμενη ή υποστηρίζουσα περιοχή του $\hat{\gamma}$, $R(\hat{\gamma},\Gamma_H)$ (βλέπε και παράγραφο 2.4), περιέχεται προφανώς στην διευρυμένη περιοχή του $\hat{\gamma}$.

Δηλαδή: $$R(\hat{\gamma},\Gamma_H)\subseteq\widetilde{R}\hat{\gamma}$$

Αν υπάρχουν k το πλήθος «gradients» στο $\widetilde{\Gamma}_H$,τότε θα υπάρχουν k διευρυμένες περιοχές. Αν η συνάρτηση $\widetilde{H}\upsilon(\pi):=\max\{\pi.\gamma:\gamma\in\widetilde{\Gamma}_H\},\forall\pi\in\Pi$ χρησιμοποιείται για να

προσεγγίσει την συνάρτηση $Hυ(π)=\max_{γ\in\Gamma_H} π.γ, π\in\Pi$ τότε το σφάλμα αυτής της προσέγγισης μπορεί να ορισθεί ως η συνάρτηση :

$$σ(π):=Hυ(π)-\widetilde{H}υ(π)\ ,\ π\in\Pi\ . \qquad \underline{3.1.5}$$

Προφανώς $σ(π)\geq 0 \quad \forall π\in\Pi.$

Το ουσιαστικό που αποδεικνύεται στο λήμμα 3.1.1 καθώς και στο θεώρημα 3.1.1 είναι το γεγονός, ότι το μέγιστο σφάλμα αυτής της προσέγγισης, ήτοι η μέγιστη τιμή του σ στο Π, θα πετυχαίνεται σε μία από τις κορυφές αυτών ακριβώς των «διευρυμένων περιοχών».

Η βασική ιδέα του αλγορίθμου που θα εκθέσουμε παρακάτω, με απλά βήματα δίνει το σύνολο $Γ_H$ κτίζοντας αυτό προοδευτικά, σε αντιδιαστολή με τους υπάρχοντες αλγόριθμους που ξεκινούν από το πολυπληθές σύνολο G των εν δυνάμει «gradients» για την Hυ, (βλέπε ενότητα 2.3), που στη συνέχεια με απαλοιφές των μη λειτουργικών «gradients» καταλήγει στην ελάχιστη αντιπροσώπευση $Γ_H$.

Λήμμα 3.1.1: *Ας είναι $\widetilde{R}_{\hat{γ}}$ η διευρυμένη περιοχή για ένα «gradient» $\hat{γ}\in\widetilde{Γ}_H$. Τότε η μέγιστη τιμή του σφάλματος σ(π) στην $\widetilde{R}\hat{γ}$ αντιστοιχεί σε μία από τις κορυφές της $\widetilde{R}_{\hat{γ}}$.*

Απόδειξη

Για $π\in\widetilde{R}\hat{γ}$, $π.\hat{γ}\geq π.γ$, για όλα τα $γ\in\widetilde{Γ}_H$.

Η συνάρτηση σφάλματος: $σ(π):=Hυ(π)-\max_{γ\in\widetilde{Γ}_H}\{π.γ\}$ μπορεί να ξαναγραφεί σαν :

$$σ(π)= Hυ(π)-π.\hat{γ} \text{ για όλα τα } π\in\widetilde{R}\hat{γ}\ .$$

Από το γεγονός ότι η Hυ είναι κυρτή συνάρτηση (βλέπε Sondik [117]) και $π.\hat{γ}$ είναι μια γραμμική συνάρτηση, η σ θα είναι κυρτή συνάρτηση στην περιοχή $\widetilde{R}\hat{γ}$.

Από την θεωρία των κυρτών συναρτήσεων, είναι γνωστό, ότι η μέγιστη τιμή για μια κυρτή συνάρτηση πάνω σε ένα κυρτό πολύεδρο, πετυχαίνεται σε μια από τις κορυφές (ακρότατα) του κυρτού πολύεδρου.

Πράγματι, ας είναι $\pi_1,\pi_2,\ldots\pi_s$ οι κορυφές της περιοχής $\widetilde{R}_{\tilde{\gamma}}$. Κάθε $\pi\in\widetilde{R}_{\tilde{\gamma}}$ εκφράζεται σαν κυρτός γραμμικός συνδυασμός των κορυφών, δηλ.

$$\pi=\sum_{i=1}^{s}\lambda_i.\pi_i,$$

όπου : $\lambda_i\geq 0, 1\leq i\leq s$ και $\sum_{i=1}^{s}\lambda_i=1.$

Έχουμε $\sigma(\pi)=\sigma(\sum_{i=1}^{s}\lambda_i.\pi_i)\leq\sum_{i=1}^{s}\lambda_i.\sigma(\pi_i)\leq\max_{1\leq i\leq s}\sigma(\pi_i).$

Αρα $\max_{\pi\in\widetilde{R}_{\tilde{\gamma}}}\sigma(\pi)=\max_{1\leq i\leq s}\sigma(\pi_i)$, οπότε η μέγιστη τιμή της σ στην περιοχή $\widetilde{R}_{\tilde{\gamma}}$ αντιστοιχεί σε μία από τις κορυφές της διευρυμένης περιοχής $\widetilde{R}_{\tilde{\gamma}}$. □

Θεώρημα 3.1.1: Η μέγιστη τιμή του σφάλματος σ στο Π θα είναι σε μία από τις κορυφές αυτών των γενικευμένων περιοχών που αντιστοιχούν σε gradients του $\widetilde{\Gamma}_{\text{H}}$.

Απόδειξη

Όπως δείχθηκε στο λήμμα 3.1.1,η μέγιστη τιμή του σ στην περιοχή $\widetilde{R}_{\tilde{\gamma}}$ λαμβάνεται σε μία από τις κορυφές $\widetilde{R}_{\tilde{\gamma}}$. Από τον ορισμό των παραπάνω διευρυμένων περιοχών, η ένωσή τους είναι το σύνολο Π. Επομένως, η μέγιστη τιμή του σφάλματος σ θα αντιστοιχεί σε μία από αυτές τις διευρυμένες περιοχές, οπότε με βάση το λήμμα 3.1.1,η μέγιστη τιμή του σ στο Π, θα είναι σε μία από τις κορυφές αυτών των διευρυμένων περιοχών.

□

Υποθέτουμε τώρα, ότι όλες οι κορυφές για τις διευρυμένες περιοχές απαρτίζουν ένα σύνολο Ε. Λόγω του θεωρήματος 3.1.1,το μέγιστο σφάλμα αυτής της προσέγγισης,

θα είναι σε μία από τις κορυφές του συνόλου Ε. Δηλώνουμε αυτή την κορυφή $\hat{\pi}$ και έχουμε :

$$\sigma(\hat{\pi}) = \max_{\pi \in E} \sigma(\pi) .$$

Αν $\sigma(\hat{\pi})=0$ τότε, δεν έχουμε καθόλου σφάλμα για την προσέγγιση αυτή, και έχουμε ακριβή λύση.

Αν $\sigma(\hat{\pi}) > 0$,τότε βρίσκουμε το «gradient η τα gradients» του Hu στο $\hat{\pi}$,εφαρμόζοντας τον αλγόριθμο A_3 (του ενός βήματος) της ενότητας 2.2. Συμβολίζουμε το σύνολο των gradients» της Hu στο $\hat{\pi}$ με $\Gamma_{\hat{\pi}}$. Σημειώνουμε επίσης ότι τα «gradients» του $\Gamma_{\hat{\pi}}$ δεν ανήκουν στο $\tilde{\Gamma}_H$,διότι :

$$Hu(\hat{\pi}) = \hat{\pi}.\hat{\gamma} > \max\{\hat{\pi}\gamma : \gamma \in \tilde{\Gamma}_H\}.$$

Επομένως, αν τα gradients του $\Gamma_{\hat{\pi}}$ συμπεριληφθούν στο $\tilde{\Gamma}_H$ μια καλύτερη προσέγγιση του Hu μπορεί να βρεθεί,και κατασκευάζονται πλέον καινούργιες **διευρυμένες περιοχές** για κάθε «gradient» του νέου συνόλου $\tilde{\Gamma}_H \cup \Gamma_{\hat{\pi}}$.

Από το θεώρημα 3.1.1, το μέγιστο σφάλμα αυτής της νέας προσέγγισης θα αντιστοιχεί σε μία από τις κορυφές των νέων διευρυμένων περιοχών, που ορίζονται μέσω των «gradients» του $\tilde{\Gamma}_H \cup \Gamma_{\hat{\pi}}$.Έτσι το μέγιστο σφάλμα για την καινούργια προσέγγιση αναζητείται στις κορυφές των νέων διευρυμένων περιοχών. Όπως θα δειχθεί στην επόμενη πρόταση, όλες οι κορυφές των νέων αυτών διευρυμένων περιοχών ανήκουν στο σύνολο $E \cup C$, όπου C το σύνολο από όλες τις κορυφές των διευρυμένων περιοχών για τα «gradients του $\Gamma_{\hat{\pi}}$. Αυτό πρακτικά σημαίνει ότι μόνο οι κορυφές των (ή της) διευρυμένων (διευρυμένης) περιοχών (περιοχής) για τα «gradients (ή gradient)» του $\Gamma_{\hat{\pi}}$ πρέπει να προσδιορισθούν στο νέο βήμα.

***Πρόταση* 3.1.1:** Ας είναι $\tilde{\Gamma}_H$ ένα σύνολο από «gradients» που περιγράφηκαν παραπάνω. Θεωρούμε τις διευρυμένες περιοχές που ορίζονται μέσω των «gradients» του συνόλου $\tilde{\Gamma}_H$: Για $\gamma \in \tilde{\Gamma}_H, \tilde{R}_\gamma = \{\pi \in \Pi : \pi.\gamma \geq \pi.\tilde{\gamma} \ \forall \ \tilde{\gamma} \in \tilde{\Gamma}_H\}$.Ας είναι τώρα ,Ε, το σύνολο των κορυφών των διευρυμένων περιοχών $\tilde{R}_\gamma$, $\gamma \in \tilde{\Gamma}_H$. Υποθέτουμε ότι $\hat{\pi} \in E$ είναι ένα δ.π. με $\sigma(\hat{\pi}) = \max_{\pi \in E} \sigma(\pi) > 0$, και $\Gamma_{\hat{\pi}}$ το σύνολο των «gradients» τής Hu στο $\hat{\pi}$.

Για $\gamma \in \widetilde{\Gamma}_H \cup \Gamma_{\hat{\pi}}$, συμβολίζουμε με $\widetilde{R}'_\gamma$, την διευρυμένη περιοχή:

$$\widetilde{R}'_\gamma = \{\pi \in \Pi: \pi.\gamma \geq \pi.\widetilde{\gamma} \quad \forall\ \widetilde{\gamma} \in \widetilde{\Gamma}_H \cup \Gamma_{\hat{\pi}}\}.$$

Αν επιπλέον C είναι το σύνολο από όλες τις κορυφές των **διευρυμένων περιοχών** $\widetilde{R}'_\gamma$ για $\gamma \in \Gamma_{\hat{\pi}}$ και E΄ το σύνολο των κορυφών των **διευρυμένων περιοχών** $\widetilde{R}'_\gamma$ για $\gamma \in \widetilde{\Gamma}_H \cup \Gamma_{\hat{\pi}}$ τότε ισχύει ότι:

$$C \subseteq E' \subseteq E \cup C.$$

Απόδειξη

Το γεγονός ότι $C \subseteq E'$ προκύπτει άμεσα από τον ορισμό των συνόλων C, E'.

Εστω $x \in E'$. Θα δείξουμε ότι $x \in E \cup C$.

Διακρίνουμε τις ακόλουθες περιπτώσεις

i) $\chi \in \tilde{R}'_{\hat{\gamma}}$ για κάποιο $\hat{\gamma} \in \Gamma_{\hat{\pi}}$.

Θα δείξουμε ότι $x \in C$. Θεωρούμε ότι $x \notin C$. Θα καταλήξουμε σε άτοπο.

Επειδή $\chi \in E'$ και $x \notin C$, το δ.π. χ_3 δεν είναι κορυφή τής διευρυμένης περιοχής $\tilde{R}'_{\hat{\gamma}}$, αλλά είναι κορυφή μίας διευρυμένης περιοχής $\tilde{R}'_{\gamma'}$ όπου $\gamma' \in \widetilde{\Gamma}_H$.

Έστω $$R = \tilde{R}'_{\hat{\gamma}} \cap \tilde{R}'_{\gamma'}$$

το κοινό σύνορο των περιοχών $\tilde{R}'_{\hat{\gamma}}, \tilde{R}'_{\gamma'}$, το οποίο είναι μη κενό (επειδή $\chi \in R$) και κυρτό σύνολο ως τομή κυρτών συνόλων (πολυέδρων του χώρου R^N).

Επειδή το χ είναι κορυφή της περιοχής $\tilde{R}'_{\gamma'}$ και $\chi \in R, R \subseteq \tilde{R}'_{\gamma'}$, έπεται ότι το χ είναι κορυφή του κοινού συνόρου R επίσης. Σημειώνουμε ότι:

$$R = \{\pi \in \Pi: \pi.\hat{\gamma} = \pi.\gamma' \geq \pi.\gamma \quad \forall\ \gamma \in \widetilde{\Gamma}_H \cup \Gamma_{\hat{\pi}}\}.$$

Επειδή το χ δεν είναι κορυφή της $\tilde{R}'_{\hat{\gamma}}$, υπάρχουν $\pi_1, \pi_2 \in \tilde{R}'_{\hat{\gamma}}, \pi_1 \neq \pi_2$ και $\lambda \in (0,1)$

έτσι ώστε : $x = \lambda.\pi_1 + (1-\lambda).\pi_2$

Σημειώνουμε ότι αποκλείεται αμφότερα τα π_1,π_2 να ανήκουν στο κοινό σύνορο R επειδή το χ είναι κορυφή του R.Χωρίς βλάβη της γενικότητας θεωρούμε ότι το $\pi_1 \notin R$.Τότε έχουμε:

$$\pi_1.\hat{\gamma} > \pi_1.\gamma' \quad \text{και} \quad \pi_2.\hat{\gamma} \geq \pi_2.\gamma'.$$

Επομένως

$$\chi.\hat{\gamma} = (\lambda.\pi_1 + (1-\lambda).\pi_2).\hat{\gamma} = \lambda.\pi_1.\hat{\gamma} + (1-\lambda).\pi_2.\hat{\gamma} > \lambda.\pi_1.\gamma' + (1-\lambda).\pi_2.\gamma' =$$

$$=(\lambda.\pi_1 + (1-\lambda).\pi_2).\gamma' = \chi.\gamma'$$

δηλαδή $\chi.\hat{\gamma} > \chi.\gamma'$,πράγμα πού αντίκειται στο γεγονός ότι $\chi \in R$. Αρα το $\chi \in C$.

ii) $x \notin \bar{R}'_\gamma \ \forall \gamma \in \Gamma_{\hat{\pi}}$ (και επομένως $x \notin C$). Θα αποδείξουμε ότι $\chi \in E$.

Επειδή $\chi \notin C$,το δ.π. ,χ, είναι κορυφή μιας διευρυμένης περιοχής $\bar{R}'_\gamma$,όπου $\gamma' \in \tilde{\Gamma}_H$.Σημειώνουμε ότι $\chi.\gamma' \geq \chi.\gamma \ \forall \gamma \in \tilde{\Gamma}_H$ και $\chi.\gamma' > \chi.\gamma \quad \forall \gamma \in \Gamma_{\hat{\pi}}$.

Αν το χ είναι εσωτερικό σημείο του χώρου Π, τότε $\chi = (\chi_1,\chi_2,\ldots,\chi_N)$ με $\chi_i > 0$, $1 \leq i \leq N$.Αν το χ είναι εσωτερικό σημείο μιας συνοριακής περιοχής του χώρου Π πού παράγεται από $s < N$ ακρότατα (κορυφές) του χώρου Π, τότε το χ έχει N-s μηδενικές συνιστώσες. Χωρίς βλάβη της γενικότητας και για απλοποίηση της απόδειξης θεωρούμε ότι:

$$x = (x_1, x_2, \ldots, x_s, 0, 0, \ldots, 0) \quad \text{αν} \quad s < N$$

και s=N αν το $x = (x_1, x_2, \ldots, x_N)$ είναι εσωτερικό σημείο του χώρου Π. Σε κάθε περίπτωση το δ.π προκύπτει ως μοναδική λύση ενός συστήματος s εξισώσεων. Υπάρχει ένα σύνολο από gradients$\{\gamma^1, \gamma^2, \ldots \gamma^{s-1}\}$,όπου $\gamma^i \in \tilde{\Gamma}_H$, i =1,2,3…,s-1 ,έτσι ώστε αυτό το σύστημα των εξισώσεων να γραφεί ως ακολούθως:

$$\pi.\gamma' = \pi.\gamma^i \text{ , όπου } i = 1,2,3\ldots,s-1$$

και

$$\sum_{k=1}^{s} \pi_k = 1. \qquad (\Sigma)$$

με μοναδική λύση $\pi = \chi$.Σημειώνουμε ότι $\gamma^i \in \tilde{\Gamma}_H$, i =1,2,3…,s-1.

Θα δείξουμε ότι το χ είναι κορυφή της περιοχής $\bar{R}_{\gamma'}$,με εις άτοπον απαγωγή.

Θεωρούμε ότι το χ δεν είναι κορυφή της $\tilde{R}_{\gamma'}$. Τότε υπάρχουν δ.π $\pi_1,\pi_2 \in \tilde{R}_{\gamma'}$, $\pi_1 \neq \pi_2$ και $\lambda \in (0,1)$

έτσι ώστε:

$$\chi = \lambda.\pi_1 + (1-\lambda).\pi_2$$

Αν $s < N$, από την παραπάνω σχέση προκύπτει ότι τα π_1,π_2 έχουν N-s τελευταίες συνιστώσες μηδενικές (όπως το x).Επειδή το χ είναι κορυφή της περιοχής $\tilde{R}'_{\gamma'}$ αποκλείεται αμφότερα τα π_1,π_2 να ανήκουν στην $\tilde{R}'_{\gamma'}$. Εστω $\pi_1 \in \tilde{R}_{\gamma'} - \tilde{R}'_{\gamma'}$. Τότε

$$\pi_1.\gamma' > \pi_1.\gamma^j \quad \text{για κάποιο } j \in \{1,2,3\ldots,s-1\}.$$

(γιατί αν $\pi_1.\gamma' = \pi_1.\gamma^i, 1 \leq i \leq s-1$,τότε το π_1 θα ήταν επίσης λύση του συστήματος (Σ) πράγμα πού αντίκειται στη μοναδικότητα της λύσης χ).

Επίσης προφανώς: $\pi_2.\gamma' \geq \pi_2.\gamma^j$ (επειδή $\pi_2 \in \tilde{R}_{\gamma'}$)

Επομένως,

$$\chi.\gamma' = (\lambda.\pi_1 + (1-\lambda).\pi_2).\gamma' = \lambda.\pi_1.\gamma' + (1-\lambda).\pi_2.\gamma' > \lambda.\pi_1.\gamma^j + (1-\lambda).\pi_2.\gamma^j =$$

$$= (\lambda.\pi_1 + (1-\lambda).\pi_2).\gamma^j = \chi.\gamma^j,$$

δηλαδή $\chi.\gamma' > \chi.\gamma^j$, πράγμα άτοπο.

Άρα το δ.π χ είναι κορυφή της περιοχής $\tilde{R}_{\gamma'}$,πράγμα πού συνεπάγεται ότι $\chi \in E$.

Από (i) και (ii) έχουμε: $E' \subseteq E \cup C$

□

Ορίζουμε το νέο σφάλμα προσέγγισης σ΄:

$$\sigma'(\pi) := H\upsilon(\pi) - \max_{\gamma \in \tilde{\Gamma}_H \cup \Gamma_{\tilde{\pi}}} (\pi.\gamma), \quad \pi \in \Pi.$$

Προφανώς $\sigma'(\pi) \leq \sigma(\pi) \quad \forall \pi \in \Pi$.

Σημειώνουμε επίσης ότι το δ.π, $\hat{\pi} \in E$ για το οποίο μεγιστοποιείται το αρχικό σφάλμα σ, δηλ. $\sigma(\hat{\pi}) = \max_{\pi \in E} \sigma(\pi)$,έχει νέο σφάλμα ίσο με 0, δηλ, $\sigma'(\hat{\pi}) = 0$.

Η επόμενη πρόταση μας δίνει την δυνατότητα να καθορίσουμε επακριβώς το σύνολο E΄ των κορυφών των νέων διευρυμένων περιοχών, πράγμα που είναι βασικό για τον αλγόριθμο που ακολουθεί.

***Πρόταση* 3.1.2:i)** Αν $\pi \in E$ και $\sigma'(\pi) = \sigma(\pi)$, τότε $\pi \in E'$

ii)Αν $\pi \in E, \pi \notin C$ και $\sigma'(\pi) < \sigma(\pi)$,τότε $\pi \notin E'$.

Απόδειξη

i) Επειδή $\pi \in E$, το δ.π π είναι κορυφή (ακρότατο) κάποιας διευρυμένης περιοχής $\tilde{R}_{\gamma'}$,όπου $\gamma' \in \tilde{\Gamma}_H$.

Επειδή $\sigma'(\pi) = \sigma(\pi)$,έχουμε

$$\pi.\gamma' = \max_{\gamma \in \tilde{\Gamma}_H} \{\pi.\gamma\} = \max_{\gamma \in \tilde{\Gamma}_H \cup \Gamma_{\hat{\pi}}} \pi.\gamma \text{ ,άρα } \pi \in \tilde{R}'_{\gamma'}.$$

Επειδή το π είναι κορυφή της περιοχής $\tilde{R}_{\gamma'}$,δεν υπάρχουν $\pi', \pi'' \in \tilde{R}_{\gamma'}, \pi' \neq \pi''$, $\lambda \in (0,1)$ έτσι ώστε $\pi = \lambda.\pi' + (1-\lambda).\pi''$.

Επειδή $\tilde{R}'_{\gamma'} \subseteq \tilde{R}_{\gamma'}$ και $\pi \in \tilde{R}'_{\gamma'}$, έπεται ότι π είναι επίσης κορυφή της διευρυμένης περιοχής $\tilde{R}'_{\gamma'}$.Άρα το $\pi \in E'$.

ii) Θεωρούμε ότι $\pi \in E'$. Θα καταλήξουμε σε άτοπο. Επειδή $\pi \notin C$,το **δ.π** ,π, είναι κορυφή κάποιας διευρυμένης περιοχής $\tilde{R}'_{\gamma'}$ όπου $\gamma' \in \tilde{\Gamma}_H$.Αυτό συνεπάγεται

$$\pi.\gamma' \geq \pi.\gamma \qquad \forall \gamma \in \tilde{\Gamma}_H \cup \Gamma_{\hat{\pi}}.$$

Επομένως

$$\pi.\gamma' = \max_{\gamma \in \tilde{\Gamma}_H \cup \Gamma_{\hat{\pi}}} \pi.\gamma = \max_{\gamma \in \tilde{\Gamma}_H} \pi.\gamma \quad \text{και}$$

$$\sigma'(\pi) := Hu(\pi) - \max_{\gamma \in \tilde{\Gamma}_H \cup \Gamma_{\hat{\pi}}} \pi.\gamma = Hu(\pi) - \max_{\gamma \in \tilde{\Gamma}_H} \pi.\gamma = \sigma(\pi),$$

που αντιβαίνει στην υπόθεση $\sigma'(\pi) < \sigma(\pi)$. Αρα $\pi \notin E'$.

□

Από τις προτάσεις 3.1.1 και 3.1.2 συνάγεται ότι:

$$E' = C \cup \{\pi \in E : \sigma'(\pi) = \sigma(\pi)\} \qquad \textbf{3.1.6}$$

Στο σημείο αυτό, θέλουμε να επισημάνουμε ότι μπορούμε να σταματήσουμε και πιο πριν τον αλγόριθμο, επιλέγοντας έναν μικρό θετικό αριθμό ε (προκαθορισμένο σφάλμα) και απαιτώντας το μέγιστο σφάλμα προσέγγισης να είναι μικρότερο ή ίσο του παραπάνω αριθμού ε (κριτήριο-τερματισμού).

3.2. Αλγόριθμος (ακρότατων σημείων) A_4

ΒΗΜΑ 0:Επιλέγουμε το προκαθορισμένο σφάλμα ε και ένα πεπερασμένο σύνολο από δ.π. $\widetilde{E}$ (π.χ. τις κορυφές του Π έστω $e_1, e_2, \ldots e_N$).Βρίσκουμε εφαρμόζοντας τον αλγόριθμο του ενός βήματος της ενότητας 2.2 τα "λειτουργικά gradients" και τα τοποθετούμε σε ένα σύνολο $\widetilde{\Gamma}_H$. Για κάθε $\bar{\gamma} \in \widetilde{\Gamma}_H$ βρίσκουμε την αντίστοιχη διευρυμένη περιοχή :

$$\widetilde{R}_{\bar{\gamma}} = \{\pi \in \Pi : \pi.\bar{\gamma} \geq \pi.\gamma \quad \forall \gamma \in \widetilde{\Gamma}_H\}.$$

καθώς και τις κορυφές (ακρότατα) αυτών των διευρυμένων περιοχών.
Τοποθετούμε αυτές τις κορυφές μέσα σε ένα σύνολο Ε.

ΒΗΜΑ 1:Βρίσκουμε το $H_U(\pi)$ για κάθε $\pi \in E$ με τον αλγόριθμο του ενός βήματος.

ΒΗΜΑ 2:Υπολογίζουμε το $\sigma(\pi) = H_U(\pi) - \max_{\gamma \in \widetilde{\Gamma}_H} \pi.\gamma$ για τα $\pi \in E$.

Αν όλα τα $\sigma(\pi)$ είναι μικρότερα ή ίσα του ε πάμε στο βήμα 5,αλλιώς επιλέγουμε την κορυφή από το Ε, στην οποία το σφάλμα προσέγγισης σ μεγιστοποιείται και συμβολίζουμε την κορυφή αυτή με $\hat{\pi}$.

ΒΗΜΑ3:Βρίσκουμε τα «gradients» ή «gradient» για το $H_U(\hat{\pi})$ εφαρμόζοντας τον αλγόριθμο του ενός βήματος και τα τοποθετούμε στο σύνολο $\Gamma_{\hat{\pi}}$.Ορίζουμε το σύνολο $\widetilde{\Gamma}'_H = \widetilde{\Gamma}_H \cup \Gamma_{\hat{\pi}}$.Βρίσκουμε τις νέες διευρυμένες περιοχές που αντιστοιχούν στα «gradients» του $\widetilde{\Gamma}'_H$:

Για $\gamma \in \widetilde{\Gamma}'_H$,

$$\widetilde{R}'_{\gamma} = \{\pi \in \Pi : \pi.\gamma \geq \pi.\gamma' \quad \forall \gamma' \in \widetilde{\Gamma}'_H\}.$$

Έστω C το σύνολο των κορυφών των νέων διευρυμένων περιοχών που αντιστοιχούν στα «gradients» του $\Gamma_{\hat{\pi}}$.

ΒΗΜΑ 4: Υπολογίζουμε $\sigma'(\pi)=Hu(\pi)-\max_{\gamma\in\Gamma_{\hat{\pi}}} \pi.\gamma$ για κάθε $\pi\in E\cup C$. Έστω E΄ το σύνολο των κορυφών που δίνεται από την (3.1.6).

Θέτουμε $\tilde{\Gamma}_H=\tilde{\Gamma}'_H$, E=E΄, σ=σ΄ και πηγαίνουμε στο βήμα 1.

ΒΗΜΑ 5:STOP. Η τιμή $\tilde{H}\upsilon(\pi)=\max\{\pi.\gamma:\gamma\in\tilde{\Gamma}_H\}$ είναι μια προσέγγιση της $Hu(\pi)$ με μέγιστο σφάλμα μικρότερο ή ίσο από έναν δεδομένο μικρό θετικό αριθμό ε. Το σύνολο E περιέχει όλες τις κορυφές των διευρυμένων περιοχών για τα «gradients» στο $\tilde{\Gamma}_H$. Σημειώνουμε στο σημείο αυτό ότι αν ε=0 τότε η συνάρτηση Hu καθορίζεται επακριβώς ($\tilde{H}\upsilon=Hu$) και $\Gamma_H=\tilde{\Gamma}_H$.

Παρατηρήσεις:

1) Ο παραπάνω αλγόριθμος τερματίζεται μετά από πεπερασμένο αριθμό επαναλήψεων. Πράγματι, έστω $\tilde{\Gamma}_H$ το σύνολο των λειτουργικών «gradients» για την συνάρτηση Hu που καθορίσθηκε σε κάποια επανάληψη και E το σύνολο των κορυφών των διευρυμένων περιοχών που αντιστοιχούν στα «gradients» του συνόλου $\tilde{\Gamma}_H$. Όταν επιλεγεί από το σύνολο E η κορυφή $\hat{\pi}$ στην οποία αντιστοιχεί το μέγιστο σφάλμα προσέγγισης

$$\sigma(\hat{\pi})=\max_{\pi\in E}\sigma(\pi)>\varepsilon.$$

(όπου $\varepsilon\geq 0$ είναι το προκαθορισμένο σφάλμα τερματισμού του αλγορίθμου), το σύνολο $\tilde{\Gamma}_H$ επικαιροποιείται με την προσθήκη του συνόλου $\Gamma_{\hat{\pi}}$ των «gradients» που είναι λειτουργικά για το δ.π $\hat{\pi}$. Επίσης επικαιροποιείται το σύνολο των κορυφών των νέων διευρυμένων περιοχών σύμφωνα με την σχέση 3.1.5. Επειδή $\tilde{\Gamma}_H\subseteq\Gamma_H$, $\Gamma_{\hat{\pi}}\subseteq\Gamma_H$ και το σύνολο Γ_H των λειτουργικών «gradients» που αντιπροσωπεύουν τη συνάρτηση Hu είναι πεπερασμένο, συνάγεται ότι απαιτείται πεπερασμένος αριθμός επαναλήψεων για να περατωθεί ο αλγόριθμος.

Στην τελική επανάληψη (βήμα τερματισμού) έχουμε:

$$\max_{\pi\in E}\sigma(\pi)\leq\varepsilon$$

Αν $\varepsilon = 0$ (μηδενικό προκαθορισμένο σφάλμα),τότε στην τελική επανάληψη έχουμε

$$\max_{\pi \in E} \sigma(\pi) = 0,$$

και από το θεώρημα 3.1.1 συνάγεται ότι:

$$\sigma(\pi) = Hu(\pi) - \tilde{H}\, u(\pi) = 0 \quad \forall\, \pi \in \Pi.$$

Επομένως επιτυγχάνεται ακριβής υπολογισμός της συνάρτησης Hu ($\tilde{H}\, u = Hu$) $\Gamma_H = \tilde{\Gamma}_H$ και οι διευρυμένες περιοχές συμπίπτουν με τις αντίστοιχες συσχετισμένες, δηλαδή:

$$\tilde{R}_\gamma = R(\gamma, \Gamma_H) \qquad \forall\, \gamma \in \tilde{\Gamma}_H\, (=\Gamma_H).$$

Αν $\varepsilon > 0$,τότε από το θεώρημα 3.1.1 προκύπτει ότι στην τελική επανάληψη,

$$0 \le \sigma(\pi) = Hu(\pi) - \tilde{H}\, u(\pi) \le \varepsilon \qquad \forall\, \pi \in \Pi.$$

Η προσέγγιση $\tilde{H}\, u$ της συνάρτησης Hu προσδιορίζεται από το σύνολο $\tilde{\Gamma}_H \subseteq \Gamma_H$. Επιπλέον οι διευρυμένες περιοχές δεν συμπίπτουν εν γένει με τις αντίστοιχες συσχετισμένες: $\tilde{R}_\gamma \supseteq R(\gamma, \Gamma_H),\ \gamma \in \tilde{\Gamma}_H$.

2)Άριστη ή σχεδόν άριστη συνάρτηση ελέγχου $\tilde{\delta}$ για τη συνάρτηση $\tilde{H}\, u$ επιτυγχάνεται όταν $\varepsilon = 0$ ή $\varepsilon > 0$ αντίστοιχα και καθορίζεται στο βήμα τερματισμού. Σε κάθε περίπτωση η άριστη (ή σχεδόν) άριστη συνάρτηση ελέγχου επιλέγει για μια διευρυμένη περιοχή τη φέρουσα απόφαση του αντίστοιχου «gradient».

Πιο συγκεκριμένα, για $\gamma \in \tilde{\Gamma}_H$ έχουμε:

$$\tilde{H}\, u(\pi) = \pi.\gamma \qquad \forall\, \pi \in \tilde{R}_\gamma$$

οπότε $$\tilde{\delta}(\pi) = \alpha, \qquad \forall\, \pi \in \tilde{R}_\gamma,$$

όπου ,α, είναι η φέρουσα απόφαση του «gradient» γ (πρβλ. κεφ. 2).

Σημειώνουμε ακόμη ότι:

$$\tilde{H}\, u(\pi) = H_{\tilde{\delta}}\, u(\pi) = \pi.q^{\tilde{\delta}} + \beta.\sum_{\theta \in \Theta} \{\theta / \pi, \tilde{\delta}\}.u(T(\pi, \theta, \tilde{\delta})), \pi \in \Pi.$$

3.3. Εφαρμογή του αλγορίθμου των ακροτάτων σημείων

Στο πρόβλημα που θα ακολουθήσει έχουμε τρεις δυνατές αποφάσεις, δύο μηνύματα. Οι πίνακες μεταφοράς, μηνυμάτων καθώς και τα άμεσα κόστη φαίνονται παρακάτω:

ΕΦΑΡΜΟΓΗ 3.1

$$p^1=\begin{bmatrix}0.8 & 0.2\\ 0.5 & 0.5\end{bmatrix} \qquad R^1=\begin{bmatrix}0.8 & 0.2\\ 0.6 & 0.4\end{bmatrix} \qquad q^1=\begin{bmatrix}-4\\ 5\end{bmatrix}$$

$$p^2=\begin{bmatrix}0.5 & 0.5\\ 0.4 & 0.6\end{bmatrix} \qquad R^2=\begin{bmatrix}0.9 & 0.1\\ 0.4 & 0.6\end{bmatrix} \qquad q^2=\begin{bmatrix}-2\\ 3\end{bmatrix}$$

$$p^3=\begin{bmatrix}0.6 & 0.4\\ 0.3 & 0.7\end{bmatrix} \qquad R^3=\begin{bmatrix}0.9 & 0.1\\ 0.2 & 0.8\end{bmatrix} \qquad q^3=\begin{bmatrix}-1\\ 1\end{bmatrix}$$

$\beta=1, \quad \Gamma=\{\gamma^1,\gamma^2 : \gamma^1=[4,5]^T \quad \gamma^2=[3,9]^T\}$

Θεωρούμε το σύνολο $\widetilde{E}=\{(0,1),(1,0)\}$.Εφαρμόζοντας τον αλγόριθμο του ενός βήματος βρίσκουμε τα λειτουργικά «gradients» στα εν λόγω σημεία και είναι για το (0,1) το $(0.2,11)^T$ και για το (1,0) το $(4.62, 7.91)^T$. Άρα θα έχουμε ότι:

$$\widetilde{\Gamma}_H=\{(0.2, 11)^T, (4.62, 7.91)^T\}.$$

Έστω $\widetilde{R}_1$ η διευρυμένη περιοχή για το $(0.2,11)^T$ και $\widetilde{R}_2$ η διευρυμένη περιοχή για το $(4.62, 7.91)^T$

όπου : $\widetilde{R}_1=\{\pi\in\Pi : \pi_1+\pi_2=1 \text{ και } 0.2\pi_1+11\,\pi_2\geq 4.62.\,\pi_1+7.92\,\pi_2\}$.

Το σύνολο των κορυφών των δύο διευρυμένων περιοχών είναι :

E={(0,1),(0.41,0.59),(1,0)}.Από τα βήματα 1 και 2 υπολογίζουμε το σφάλμα για όλες τις κορυφές και το μέγιστο σφάλμα παρατηρείται για την κορυφή $\hat{\pi}$=(0.41,0.59): $\sigma(\hat{\pi})$=0.74>0. Με το βήμα 3 βρίσκουμε το λειτουργικό «gradient» που αντιστοιχεί στην $\hat{\pi}$ εφαρμόζοντας τον αλγόριθμο του ενός βήματος και είναι το $(4, 9.6)^T$. Δηλαδή $\Gamma_{\hat{\pi}}=\{(4, 9.6)^T\}$.

Άρα $\tilde{\Gamma}'_H=\tilde{\Gamma}_H\cup\Gamma_{\hat{\pi}}=\{(0.2,11)^T,(4, 9.6)^T,(4.62, 7.91)^T\}$

Οι κορυφές για την διευρυμένη περιοχή του $(4,9.6)^T$ είναι (0.27,0.73)και (0.73,0.27). Τοποθετούμε τις δύο αυτές κορυφές στο σύνολο C. Υπολογίζουμε τα σ΄([0.27,0.73]) και σ΄([0.73,0.27]) που βγαίνουν 0. Αρα τα «gradients» για την αντιπροσώπευση του Hυ(π) είναι τα $(0.2,11)^T$, $(4, 9.6)^T$, $(4.62, 7.91)^T$. Οι φέρουσες αποφάσεις που αντιστοιχούν στα παραπάνω «gradients» είναι αντίστοιχα οι 1,2,3.Αφού έχουμε $\varepsilon=0$,επομένως οι τελικές συσχετιζόμενες ή υποστηρίζουσες περιοχές των

«gradients» $\Gamma_H=\widetilde{\Gamma}_H=\{(0.2,11)^T,(4,9.6)^T,(4.62,7.91)^T\}$ ταυτίζονται με τις διευρυμένες περιοχές. Δηλαδή:

$$\widetilde{R}_{\gamma} \equiv R(\gamma, \Gamma_H), \gamma \in \widetilde{\Gamma}_H=\Gamma_H$$

Η διευρυμένη περιοχή $\widetilde{R}_{(0.2,11)^T}$ έχει κορυφές τα δ.π :(0,1), (0.27, 0.73).

Η διευρυμένη περιοχή $\widetilde{R}_{(4,9.6)^T}$ έχει κορυφές τα δ.π :(0.27, 0.73), (0.73,0.27).

Η διευρυμένη περιοχή $\widetilde{R}_{(4.62,7.91)^T}$ έχει κορυφές τα δ.π :(0.73,0.27), (1,0).

Και οι 4 κορυφές στο σύνολο Ε΄, είναι οι (0,1), (0.27, 0.73), (0.73,0.27), (1,0).

3.4. Υπολογισμός του συσσωρευμένου σφάλματος της προσέγγισης για την συνάρτηση του μέγιστου αναμενόμενου ολικού οφέλους για δοσμένο χρονικό ορίζοντα.

Στην παράγραφο 3.1 περιγράψαμε έναν τρόπο υπολογισμού της συνάρτησης $Hυ(π), π\in\Pi$, όπου $υ(π)$, $π\in\Pi$ είναι γνωστή κατά τμήματα γραμμική και κυρτή συνάρτηση, σε διαδοχικά βήματα, υπολογίζοντας σε κάθε βήμα το μέγιστο σφάλμα προσέγγισης επιθεωρώντας τα ακρότατα σημεία διευρυμένων περιοχών. Μέσω του αλγορίθμου των ακροτάτων σημείων μπορούμε να βρούμε μια προσέγγιση $\widetilde{H}υ$ της Hυ με προκαθορισμένο σφάλμα $ε>0$ ή να υπολογίσουμε επακριβώς τη συνάρτηση Hυ (ε=0). Σε κάθε περίπτωση,

$$\widetilde{H}υ(π) \le Hυ(π) \le \widetilde{H}υ(π)+ε \quad \forall\ π\in\Pi.$$

Σημειώνουμε ότι η προσέγγιση $\widetilde{H}υ$ είναι κατά τμήματα γραμμική και κυρτή συνάρτηση.

Αν αντί για την συνάρτηση υ διαθέτουμε μια προσεγγιστική συνάρτηση $\tilde{υ}$,κατά τμήματα γραμμική και κυρτή, με σφάλμα προσέγγισης $ε'\ge 0$:

$$\tilde{υ}(π) \le υ(π) \le \tilde{υ}(π) + ε' \quad \forall\ π\in\Pi. \qquad \underline{3.4.1}$$

(για $ε'=0$, προφανώς $\tilde{υ}=υ$),θα υπολογίσουμε το συσσωρευμένο σφάλμα της προσέγγισης $\widetilde{H}\tilde{υ}$ για τη Hυ. Από τη σχέση (3.4.1)και λαμβάνοντας υπόψη ότι ο τελεστής Η είναι ισότονος, παίρνουμε:

$$H\tilde{υ}(π) \le Hυ(π) \le H(\tilde{υ}(π)+ε') \quad \forall\ π\in\Pi.$$

Το δεξί μέλος της παραπάνω σχέσης γράφεται:

$$H(\tilde{\upsilon}(\pi)+\varepsilon') = \max_{a\in A}\{\pi.q^a + \beta.\sum_{\theta\in\Theta}\{\theta/\pi,\alpha\}.(\tilde{\upsilon}\ (T(\pi,\theta,\alpha)+\varepsilon'))\} =$$

$$= \max_{a\in A}\{\pi.q^a + \beta.\sum_{\theta\in\Theta}\{\theta/\pi,\alpha\}.(\tilde{\upsilon}\ (T(\pi,\theta,\alpha)) + \beta.\varepsilon'\} =$$

$$= H\tilde{\upsilon}\ (\pi) + \beta.\ \varepsilon'$$

Επομένως,

$$H\tilde{\upsilon}\ (\pi) \leq H\upsilon(\pi) \leq H\tilde{\upsilon}(\pi) + \beta.\varepsilon'\ \forall\ \pi\in\Pi. \qquad \underline{3.4.2}$$

Για την μέσω του αλγορίθμου των ακρότατων σημείων προσέγγιση $\widetilde{H}\tilde{\upsilon}$ της $H\tilde{\upsilon}$ με προκαθορισμένο σφάλμα $\varepsilon \geq 0$,έχουμε:

$$\widetilde{H}\ \tilde{\upsilon}\ (\pi) \leq H\tilde{\upsilon}(\pi) \leq \widetilde{H}\tilde{\upsilon}(\pi) + \varepsilon \quad \forall\pi\in\Pi. \qquad \underline{3.4.3}$$

Από τις σχέσεις (3.4.2) και (3.4.3) παίρνουμε:

$$\widetilde{H}\ \tilde{\upsilon}\ (\pi) \leq H\tilde{\upsilon}(\pi) \leq H\upsilon(\pi) \leq H\tilde{\upsilon}(\pi) + \beta.\varepsilon' \leq \widetilde{H}\tilde{\upsilon}(\pi) + \varepsilon + \beta.\varepsilon', \forall\pi\in\Pi$$

δηλαδή

$$\widetilde{H}\ \tilde{\upsilon}\ (\pi) \leq H\upsilon(\pi) \leq \widetilde{H}\tilde{\upsilon}(\pi) + \varepsilon + \beta.\varepsilon', \forall\pi\in\Pi. \qquad \underline{3.4.4}$$

Συνοψίζοντας, αν πάρουμε την προσέγγιση $\tilde{\upsilon}$ της υ με σφάλμα προσέγγισης $\varepsilon' \geq 0$ (πρβλ. σχέση 3.4.1), τότε η προσέγγιση $\widetilde{H}\ \tilde{\upsilon}$ της $H\tilde{\upsilon}$ μέσω του αλγορίθμου των ακρότατων σημείων με προκαθορισμένο σφάλμα $\varepsilon \geq 0$,αποτελεί προσέγγιση της Hυ με σφάλμα $\varepsilon + \beta.\varepsilon'$ (σχέση 3.4.4).Το παραπάνω αποτέλεσμα εφαρμόζεται για να υπολογίσουμε το σφάλμα προσέγγισης της συνάρτησης V_t του μέγιστου αναμενόμενου ολικού οφέλους για τον χρονικό ορίζοντα t.

Συμβολίζουμε με σ_t το συσσωρευμένο σφάλμα της προσεγγιστικής συνάρτησης $\tilde{V}_t = \widetilde{H}\tilde{V}_{t-1}$ για τη συνάρτηση $V_t = HV_{t-1}$, t=1,2,3...,

Δηλαδή

$$\tilde{V}_t\ (\pi) \leq V_t(\pi) \leq \tilde{V}_t\ (\pi) + \sigma_t \quad \forall\pi\in\Pi. \qquad \underline{3.4.5}$$

Αν επιλέξουμε $\varepsilon \geq 0$ ως το προκαθορισμένο σφάλμα προσέγγισης στον αλγόριθμο των ακρότατων σημείων, τότε ισχύει η ακόλουθη αναγωγική σχέση.

$$\sigma_t = \varepsilon + \beta.\sigma_{t-1}\ ,\ t=1,2,... \qquad \underline{3.4.6}$$

Για t=0, ως συνάρτηση κέρδους στον τερματισμό επιλέγουμε συνήθως την μηδενική συνάρτηση ή γενικότερα μια γνωστή συνάρτηση (έστω u), δηλαδή:

$$V_0(\pi)=\pi.q \quad \forall \pi \in \Pi ,$$

όπου q είναι το διάνυσμα εσόδων στον τερματισμό.

Προφανώς το σφάλμα για t=0 είναι:

$$\sigma_0 = 0$$

Από την (3.3.6) παίρνουμε: $\sigma_t = (1+\beta+...+\beta^{t-1}).\varepsilon \; ,t \geq 1$

Επομένως

$$\sigma_t = \begin{cases} \dfrac{1-\beta^t}{1-\beta}.\varepsilon & ,\text{άν} \quad \beta \in (0,1) \\ t.\varepsilon & ,\text{άν} \quad \beta = 1 \end{cases} \qquad \underline{3.4.7}$$

As θεωρήσουμε ένα επιθυμητό άνω φράγμα $\eta > 0$ για το σφάλμα προσέγγισης της συνάρτησης V_T στον χρονικό ορίζοντα T. Μπορούμε τότε να επιλέξουμε το προκαθορισμένο σφάλμα $\varepsilon \geq 0$ του αλγορίθμου των ακρότατων σημείων ώστε:

$$\sigma_T \leq \eta$$

Από την (3.4.7) παίρνουμε: $\varepsilon \leq \dfrac{1-\beta}{1-\beta^T}.\eta$ αν $\beta \in (0,1)$

και $\varepsilon \leq \dfrac{\eta}{T}$ αν $\beta = 1$.

Αν θέσουμε ένα άνω φράγμα $\eta > 0$ για το σφάλμα προσέγγισης της συνάρτησης V_t για κάθε χρονικό ορίζοντα t=1,2,..., τότε μπορούμε να επιλέξουμε το προκαθορισμένο σφάλμα ε του αλγορίθμου των ακρότατων σημείων ώστε:

$$\sigma_t \leq \eta \quad \forall t=1,2,3... \qquad \underline{3.4.8}$$

στην περίπτωση όπου ο συντελεστής έκπτωσης $\beta \in (0,1)$.

Η (3.4.8) ισοδυναμεί με την σχέση $\dfrac{1}{1-\beta}.\varepsilon \leq \eta$,όπου το προκαθορισμένο σφάλμα ε μπορεί να επιλεγεί ώστε :

$$\varepsilon \leq (1-\beta).\eta .$$

3.5.Εφαρμογή του αλγορίθμου των ακρότατων σημείων στον υπολογισμό της συνάρτησης ελάχιστου κόστους για πεπερασμένο χρονικό ορίζοντα.

As υποθέσουμε ότι για κάποιο χρονικό ορίζοντα $t \geq 1$ η συνάρτηση V_t του ελάχιστου αναμενόμενου ολικού κόστους είναι γνωστή.Τίθεται το πρόβλημα υπολογισμού της συνάρτησης του ελάχιστου αναμενόμενου ολικού κόστους για τον χρονικό ορίζοντα t+1,

$$V_{t+1}=HV_t$$

(όπου Η είναι τελεστής ελαχιστοποίησης)

Γενικότερα, έστω $\upsilon(\pi)$, $\pi \in \Pi$ μια κατά τμήματα γραμμική και κοίλη συνάρτηση. Επομένως η συνάρτηση υ αντιπροσωπεύεται μέσω ενός πεπερασμένου συνόλου Γ από «gradients» και

$$\upsilon(\pi)=\min\{\pi.\gamma:\ \gamma \in \Gamma\}.$$

Ο χώρος Π διαμερίζεται σε κυρτές περιοχές W_γ ,$\gamma \in \Gamma$ έτσι ώστε:

$$\upsilon(\pi)= \pi.\gamma \quad \forall \pi \in W_\gamma$$

Η συνάρτηση

$$H\upsilon(\pi) = \min_{\alpha \in A}\left\{\pi.c^{\alpha} + \beta.\sum_{\theta \in \Theta}\{\theta / \pi, \alpha\}.\upsilon(T(\pi,\theta,\alpha))\}\right\}, \pi \in \Pi,$$

όπου c^{α} το διάνυσμα άμεσου κόστους, πού αντιστοιχεί στην απόφαση α (το οποίο θεωρείται διάνυσμα στήλη) και β ο συντελεστής έκπτωσης (discound-factor) $(0<\beta \leq 1)$ είναι επίσης κατά τμήματα γραμμική και κοίλη.

Το πρόβλημα υπολογισμού της συνάρτησης Hυ ανάγεται στον προσδιορισμό του συνόλου των λειτουργικών «gradients» Γ_H για την συνάρτηση Hυ, και των αντίστοιχων περιοχών αντιπροσώπευσης της Hυ. Για να βρούμε το παραπάνω σύνολο των «gradients» Γ_H που υποστηρίζουν την Hυ ακολουθούμε την ίδια πορεία με εκείνη πού περιγράψαμε στην ενότητα 3.1.Οι μόνες αναγκαίες τροποποιήσεις αφορούν τους ορισμούς των διευρυμένων περιοχών και του σφάλματος προσέγγισης. Επιλέγουμε αρχικά ένα πεπερασμένο σύνολο δ.π. (Για παράδειγμα μπορούμε να επιλέξουμε τις κορυφές $e_1 = (1,0,...,0), e_2 = (0,1,...,0), ... e_N = (0,0,...,1)$ του χώρου

Π των δ.π). Εφαρμόζοντας τον αλγόριθμο A_3 του ενός βήματος του κεφαλαίου 2 υπολογίζουμε τα λειτουργικά «gradients» της Hυ που αντιστοιχούν σε αυτά τα δ.π. Συμβολίζουμε με $\tilde{\Gamma}_H$ το παραπάνω σύνολο των «gradients» της Hυ. Η **διευρυμένη περιοχή** ενός «gradient» $\hat{\gamma} \in \tilde{\Gamma}_H$ ορίζεται ως το κυρτό πολύεδρο του χώρου R^N:

$$\tilde{R}\hat{\gamma} = \{\pi \in \Pi : \pi.\hat{\gamma} \leq \pi.\gamma \quad \forall \gamma \in \tilde{\Gamma}_H\} \qquad \underline{3.5.1}$$

Αν η συνάρτηση $\tilde{H}\upsilon(\pi) = \min\{\pi.\gamma : \gamma \in \tilde{\Gamma}_H\}, \forall \pi \in \Pi$ χρησιμοποιείται για να προσεγγίσει τη συνάρτηση

$$H\upsilon = \min\{\pi.\gamma : \gamma \in \Gamma_H\}, \pi \in \Pi$$

τότε το σφάλμα αυτής της προσέγγισης ορίζεται ως η συνάρτηση

$$\sigma(\pi) := \tilde{H}\upsilon(\pi) - H\upsilon(\pi), \pi \in \Pi$$

Επειδή $\tilde{\Gamma}_H \subseteq \Gamma_H$, έχουμε $H\upsilon(\pi) \leq \tilde{H}\upsilon(\pi) \quad \forall \pi \in \Pi$

Και επομένως $\sigma(\pi) \geq 0, \forall \pi \in \Pi$

Οι προτάσεις της ενότητας 3.1 (λήμμα 3.1.1,θεώρημα 3.1.1,πρόταση 3.1.1, πρόταση 3.1.2) αληθεύουν στην περίπτωσή μας με προφανείς αλλαγές λόγω τροποποιήσεων στους ορισμούς των διευρυμένων περιοχών και του σφάλματος προσέγγισης. Εφαρμόζοντας τον αλγόριθμο των ακρότατων σημείων της ενότητας 3.2 βρίσκουμε μια προσέγγιση $\tilde{H}\upsilon$ της Hυ με προκαθορισμένο σφάλμα $\varepsilon \geq 0$:

$$\tilde{H}\upsilon(\pi) - \varepsilon \leq H\upsilon(\pi) \leq \tilde{H}\upsilon(\pi) + \varepsilon \quad \forall \pi \in \Pi$$

(Για $\varepsilon = 0$,προφανώς $H\upsilon = \tilde{H}\upsilon$ $\Gamma_H = \tilde{\Gamma}_H$).

Ακολουθώντας παρόμοια μέθοδο όπως στην ενότητα 3.4,υπολογίζουμε το συσσωρευμένο σφάλμα σ_t της προσεγγιστικής συνάρτησης $\tilde{V}_t = \tilde{H}\tilde{V}_{t-1}$ για την συνάρτηση $V_t = HV_{t-1}$ t=1,2,..

$$\tilde{V}_t(\pi) - \sigma_t \leq V_t(\pi) \leq \tilde{V}_t(\pi) \quad \forall \pi \in \Pi$$

Αν $\varepsilon \geq 0$ είναι το προκαθορισμένο σφάλμα στον αλγόριθμο των ακρότατων σημείων, τότε ισχύει η αναγωγική σχέση (3.3.6) από την οποία προκύπτει η (3.3.7).Όπως και στην ενότητα 3.4, μπορούμε να επιλέξουμε το προκαθορισμένο $\varepsilon \geq 0$,έτσι ώστε το

συσσωρευμένο σφάλμα προσέγγισης να είναι μικρότερο ή ίσο από ένα επιθυμητό φράγμα για οποιοδήποτε χρονικό ορίζοντα.

ΣΥΜΠΕΡΑΣΜΑΤΑ

Με τον αλγόριθμο των ακροτάτων σημείων αυτού του κεφαλαίου υπολογίζουμε επακριβώς ή προσεγγιστικά τη συνάρτηση Hυ -όπου Η είναι τελεστής μεγιστοποίησης (ελαχιστοποίησης) και υ είναι κατά τμήματα γραμμική κυρτή (κοίλη) συνάρτηση για προβλήματα εσόδων (κόστους)- σε πεπερασμένο αριθμό επαναλήψεων.

Σε κάθε επανάληψη προσδιορίζουμε τα ακρότατα (κορυφές) των διευρυμένων περιοχών των λειτουργικών «gradients»για την Hυ, που ήδη έχουμε εντοπίσει, και υπολογίζουμε το μέγιστο σφάλμα προσέγγισης επιθεωρώντας αυτά τα ακρότατα.

Ακολούθως εμπλουτίζουμε το σύνολο των «gradients» που ήδη έχουμε εντοπίσει, με νέα λειτουργικά «gradients» και προσδιορίζουμε τα ακρότατα των νέων διευρυμένων περιοχών. Το μέγιστο σφάλμα προσέγγισης μειώνεται σε διαδοχικές επαναλήψεις και η διαδικασία τερματίζεται όταν αυτό γίνει μικρότερο ή ίσο από ένα προκαθορισμένο αριθμό $\varepsilon \geq 0$.Η πορεία που ακολουθούμε με τον αλγόριθμο αυτό είναι αντίθετη από την μέθοδο των Smallwood-Sondik-Lovejoy,που ξεκινά με ένα πολυπληθές σύνολο από εν δυνάμει «gradients»και κατόπιν διαδοχικών απαλοιφών καταλήγει στην ελάχιστη αντιπροσώπευση Γ_H.

Με τη βοήθεια του αλγορίθμου επιτυγχάνεται προσέγγιση της συνάρτησης του μέγιστου (ελάχιστου) αναμενόμενου ολικού εκπίπτοντος κέρδους (κόστους) για οποιοδήποτε χρονικό ορίζοντα. Το συσσωρευμένο σφάλμα προσέγγισης υπολογίζεται μέσω απλής αναγωγικής σχέσης. Το βασικό πλεονέκτημα της μεθόδου αποτελεί η δυνατότητα επιλογής του προκαθορισμένου σφάλματος ε,έτσι ώστε το συσσωρευμένο σφάλμα προσέγγισης για οποιοδήποτε χρονικό ορίζοντα να μην υπερβαίνει ένα επιθυμητό φράγμα. Αυτή η ιδιότητα, θα αξιοποιηθεί στο κεφάλαιο 4 για την προσέγγιση της βέλτιστης συνάρτησης τιμών, αναφορικά με το κριτήριο του ολικού εκπίπτοντος οφέλους (κόστους) για άπειρο χρονικό ορίζοντα

ΚΕΦΑΛΑΙΟ 4

Αλγόριθμοι για το πρόβλημα των μερικά παρατηρήσιμων Μαρκοβιανών διαδικασιών απόφασης σε άπειρο χρονικό ορίζοντα.

Περίληψη

Οι Παπαδημητρίου και Τσιτσικλής [92] απέδειξαν ότι το πρόβλημα POMDP σε άπειρο χρονικά ορίζοντα δεν έχει στη γενική του έκφραση ακριβή λύση με πεπερασμένους αλγόριθμους.Αυτό αποτελεί για μας εφαλτήριο για αναζήτηση προσεγγιστικών λύσεων. Σκοπός του κεφαλαίου είναι η εύρεση προσεγγίσεων της άριστης συνάρτησης τιμών και της άριστης πολιτικής (προσδιορισμός σχεδόν άριστων πολιτικών) σε άπειρο χρονικά ορίζοντα, επεκτείνοντας αποτελέσματα των Bertsekas [11], Puterman [99] και Hauskrecht [48].Το κεφάλαιο αυτό οργανώνεται ως εξής:

Στην ενότητα 4.1. οι προσεγγίσεις υλοποιούνται με την επαναληπτική εφαρμογή του αλγορίθμου των ακροτάτων σημείων που περιγράψαμε στο κεφάλαιο 3.

Στην ενότητα 4.2. αναφέρονται διάφορα φράγματα για την άριστη συνάρτηση τιμών που απαντώνται στη βιβλιογραφία και τα οποία μπορούν να χρησιμεύσουν ως αρχικές προσεγγίσεις.

Στην ενότητα 4.3. εφαρμόζεται επαναληπτικά ο αλγόριθμος των ακρότατων σημείων στα αρχικά φράγματα της ενότητας 4.2 για την δημιουργία νέων προσεγγίσεων της άριστης συνάρτησης τιμών και σχεδόν άριστων πολιτικών. Σε κάθε περίπτωση υπολογίζεται ο απαιτούμενος αριθμός των βημάτων (επαναλήψεων), καθώς

και το προκαθορισμένο σφάλμα του αλγορίθμου, έτσι ώστε να επιτύχουμε προσέγγιση με οποιαδήποτε επιθυμητή ακρίβεια..

4.1. Προσέγγιση της άριστης συνάρτησης τιμών για άπειρο χρονικό ορίζοντα και εύρεση σχεδόν άριστων πολιτικών στα πλαίσια της επαναληπτικής μεθόδου τιμών (Value-Iteration).

Στην ενότητα αυτή θα εφαρμόσουμε τη μέθοδο των διαδοχικών προσεγγίσεων ή επαναληπτική μέθοδο τιμών (successive approximations or value iteration method) για την προσέγγιση της άριστης συνάρτησης τιμών V^* και της άριστης πολιτικής για άπειρο χρονικό ορίζοντα (βλ. ενότητα 1.5).Όπως θα διαπιστώσουμε στη συνέχεια, οι προσεγγίσεις αυτές είναι δυνατόν να υλοποιηθούν με οποιαδήποτε επιθυμητή ακρίβεια εφαρμόζοντας επαναληπτικά τον αλγόριθμο των ακροτάτων σημείων που περιγράψαμε στο κεφάλαιο 3. Σημειώνουμε ακόμη ότι στο πλαίσιο του κριτηρίου βελτιστοποίησης σε άπειρο χρονικό ορίζοντα, για τον συντελεστή εκπτώσεως (discount factor) υποθέτουμε $\beta \in (0,1)$.

Το επόμενο λήμμα είναι πολύ βασικό για την εφαρμογή της μεθόδου των διαδοχικών προσεγγίσεων και οφείλεται στον Denardo [23] (βλ.επίσης Blackwell [15]).

Λήμμα 4.1.1: *Έστω V^* η βέλτιστη συνάρτηση τιμών για άπειρο χρονικό ορίζοντα σε πρόβλημα εσόδων ή κόστους, H ο τελεστής μεγιστοποίησης ή ελαχιστοποίησης και $u \in B(\Pi)$ (βλ. ενότητα 1.4,1.5). Τότε*

i) *H συνάρτηση V^* είναι το μοναδικό σταθερό σημείο του τελεστή H, δηλ.*

$w \in B(\Pi)$, $Hw=w \Rightarrow w=V^*$.

ii) $\left\| H_n u - V^* \right\| \le \beta^n . \left\| u - V^* \right\|$, $n \ge 1$.

iii) $\left\| H_n u - V^* \right\| \le \frac{\beta^n}{1-\beta} \left\| Hu - u \right\|$, $n \ge 1$

όπου με H_n *συμβολίζουμε την επαναληπτική χρήση του τελεστού* H, n *φορές και* H_0 *είναι ο ταυτοτικός τελεστής:* $H_0 u = u$. □

Θα ασχοληθούμε με προσεγγίσεις σε προβλήματα εσόδων, οπότε ο Η θεωρείται τελεστής μεγιστοποίησης. Οι προσεγγίσεις για προβλήματα κόστους γίνονται με ανάλογο τρόπο.

Θεωρούμε $V_0 \in B(\pi)$ και

$$V_n = HV_{n-1} \quad , n=1,2,3,... \qquad \underline{4.1.1}$$

Η παραπάνω σχέση γράφεται:

$$V_n = H_n V_0 \, , \, n \ge 1$$

Το λήμμα 4.1.1 (ii) δείχνει ότι η ακολουθία $\{V_n\}$ συγκλίνει ομαλά, όταν το $n \to \infty$ στη βέλτιστη συνάρτηση τιμών, V^*, ανεξάρτητα από την επιλογή της αρχικής συνάρτησης V_0.

Αν $V_0 = 0$ (μηδενική συνάρτηση), τότε:

$$\left\| V_n - V^* \right\| \le \frac{\beta^n . \Lambda}{1-\beta} \quad \forall \; n \ge 1 \, , \qquad \underline{4.1.2}$$

όπου $\Lambda = \max_{i,a} \left| q(\overset{\bullet}{i}, \alpha) \right|$.

Πράγματι,

$$V_1(\pi) = HV_0(\pi) = \max_a \{\pi . q^a\} \le \max_{i,a} \left| q(i, \alpha) \right| = \Lambda \, , \qquad \forall \pi \in \Pi$$

Επομένως $\|V_1\| \le \Lambda$.

Εφαρμόζοντας το λήμμα 4.1.1 (iii) παίρνουμε:

$$\|V_n - V^*\| = \|H_n V_0 - V^*\| \le \frac{\beta^n}{1-\beta} \| HV_0 - V_0 \| = \frac{\beta^n}{1-\beta} \|V_1\| \le \frac{\beta^n}{1-\beta}\Lambda, \; n \ge 1$$

Η σχέση (4.1.2) είναι χρήσιμη στην περίπτωση που χρησιμοποιούμε τον αλγόριθμο των ακρότατων σημείων (Κεφ.3), για τον ακριβή υπολογισμό των συναρτήσεων V_n (προκαθορισμένο σφάλμα ε=0). Αν $\eta > 0$ είναι ένα επιθυμητό φράγμα για το σφάλμα προσέγγισης, δηλ.

$$\|V_n - V^*\| \le \eta$$

τότε, μπορούμε να επιλέξουμε χρονικό ορίζοντα ,*n*, έτσι ώστε :

$$\frac{\beta^n}{1-\beta}\Lambda \le \eta$$

δηλ. $$n \ge \frac{\ln(\frac{(1-\beta).\eta}{\Lambda})}{\ln \beta}.$$

Εργαζόμενοι με ανάλογο τρόπο μπορούμε να προσεγγίσουμε τη βέλτιστη συνάρτηση τιμών V^* εφαρμόζοντας τον προσεγγιστικό αλγόριθμο των ακρότατων σημείων με προκαθορισμένο σφάλμα ε> 0,σύμφωνα με την ακόλουθη πρόταση.

Πρόταση 4.1.2: Εστω $\tilde{V}_n$ η προσέγγιση της V_n μέσω του αλγορίθμου των ακρότατων σημείων, με προκαθορισμένο σφάλμα ε> 0 (βλέπε ενότητα 3.4). Τότε,

$$\|\tilde{V}_n - V^*\| \le \frac{1-\beta^n}{1-\beta}\varepsilon + \frac{\beta^n \Lambda}{1-\beta} \qquad \underline{4.1.3}$$

Απόδειξη

Από την τριγωνική ιδιότητα της νόρμας του supremum παίρνουμε:

$$\left\|\tilde{V}_n - V^*\right\| \le \left\|\tilde{V}_n - V_n\right\| + \left\|V_n - V^*\right\|.$$

Το συσσωρευμένο σφάλμα σ_n της προσέγγισης $\tilde{V}_n$ για την συνάρτηση V_n δίνεται από την σχέση (3.4.7)

Επομένως

$$\left\|\tilde{V}_n - V_n\right\| \le \sigma_n = \frac{1-\beta^n}{1-\beta}.\varepsilon$$

Η (4.1.3) συνάγεται άμεσα από την παραπάνω σχέση και την σχέση (4.1.2).

□

Αν $\eta > 0$ είναι ένα επιθυμητό άνω φράγμα για το σφάλμα προσέγγισης,

$$\left\|V_n - V^*\right\| \le \eta$$

τότε μπορούμε να επιλέξουμε προκαθορισμένο σφάλμα ε και χρονικό ορίζοντα n έτσι ώστε:

$$\frac{1-\beta^n}{1-\beta}.\varepsilon + \frac{\beta^n.\Lambda}{1-\beta} \le \eta$$

Μπορούμε π.χ να επιλέξουμε τα ε, η έτσι ώστε: $\frac{\varepsilon}{1-\beta} \le \frac{\eta}{2}$, $\frac{\beta^n.\Lambda}{1-\beta} \le \frac{\eta}{2}$,

δηλαδή: $\varepsilon \le \frac{(1-\beta).\eta}{2}$, $n \ge \frac{\ln(\frac{(1-\beta).\eta}{2\Lambda})}{\ln\beta}$

Έστω $\delta^{\infty}=(\delta,\delta,\ldots)$ μία στάσιμη πολιτική και $V(\pi/\delta)$, $\pi\in\Pi$, η συνάρτηση του αναμενόμενου ολικού εκπίπτοντος κέρδους για άπειρο χρονικό ορίζοντα εφαρμόζοντας την πολιτική δ^{∞}.

Η συνάρτηση V(./δ) αναφέρεται επίσης ως συνάρτηση τιμών για την πολιτική δ^{∞}.

Σημειώνουμε ακόμα ότι η V(./δ) είναι το μοναδικό σταθερό σημείο του τελεστή H_{δ}, ο οποίος είναι συστολή modulus β (βλέπε ενότητα 1.4):

$$H_{\delta}V(./\delta)=V(./\delta).$$

Με άλλα λόγια η συνάρτηση V(./δ) ικανοποιεί την εξίσωση:

$$V(\pi/\delta)=\pi.q^{\delta(\pi)}+\beta.\sum_{\theta}\{\theta/\pi,\delta\}.V(T(\pi,\theta,\delta)/\delta),\pi\in\Pi.$$

Θα στρέψουμε τώρα το ενδιαφέρον μας στον προσδιορισμό στάσιμων πολιτικών που προσεγγίζουν την άριστη πολιτική $(\delta^*)^{\infty}$ *(βλ. ενότητα 1.5).*

Ορισμός 4.1.1: Μια πολιτική δ^{∞}, λέγεται σχεδόν άριστη πολιτική με σφάλμα $\eta>0$ (ή, η-άριστη πολιτική) αν η συνάρτηση τιμών για την πολιτική δ^{∞}, V(./δ),αποτελεί προσέγγιση της άριστης συνάρτησης τιμών V* με μέγιστο σφάλμα προσέγγισης μικρότερο ή ίσο του αριθμού η, δηλαδή:

$$\left\|V(./\delta)-V^*\right\|\le\eta$$

Μπορούμε να προσεγγίσουμε την άριστη πολιτική $(\delta^*)^{\infty}$ με τη στάσιμη πολιτική $(\delta^n)^{\infty}$, όπου δ^n, είναι η άριστη συνάρτηση ελέγχου στον χρονικό ορίζοντα n :

$$V_n=HV_{n-1}=H_{\delta^n}V_{n-1}. \qquad \underline{4.1.4}$$

(όπου $V_0=0$, μηδενική συνάρτηση)

δηλαδή : $\delta^n(\pi)=\arg\max_a\{\pi.q^a+\beta.\sum_\theta\{\theta/\pi,\alpha\}.V_{n-1}(T(\pi,\theta,\alpha))\},\pi\in\Pi$.

Η συνάρτηση $V(./\delta^n)$ είναι μια καλή προσέγγιση της V* αν επιλέξουμε το n αρκετά μεγάλο, σύμφωνα με την ακόλουθη πρόταση.

Πρόταση 4.1.3: Έστω δ^n άριστη συνάρτηση ελέγχου για τον χρονικό ορίζοντα n σύμφωνα με την σχέση (4.1.4). Τότε η στάσιμη πολιτική $(\delta^n)^\infty$ είναι σχεδόν άριστη με σφάλμα $\frac{2.\beta^n}{1-\beta}.\Lambda$ $\left(\frac{2.\beta^n.\Lambda}{1-\beta}-\text{άριστη}\right)$, δηλαδή:

$$\|V(./\delta^n)-V^*\|\le\frac{2.\beta^n}{1-\beta}.\Lambda$$

όπου $\Lambda=\max_{s,a}|q(s,\alpha)|$. ***Bertsekas*** [10]

□

Αν επιθυμούμε η πολιτική $(\delta^n)^\infty$ να είναι η-άριστη για δοσμένο $\eta>0$, τότε μπορούμε να επιλέξουμε τον χρονικό ορίζοντα n έτσι ώστε:

$$\frac{2.\beta^n}{1-\beta}.\Lambda\le\eta$$

δηλαδή : $$n\ge\frac{\ln(\frac{(1-\beta).\eta}{2\Lambda})}{\ln\beta}$$

Σημειώνουμε, ότι η εφαρμογή της πρότασης 4.1.3 , είναι δυνατή μόνον στην περίπτωση όπου εφαρμόζουμε τον αλγόριθμο των ακρότατων σημείων με προκαθορισμένο σφάλμα ε=0,οπότε επιτυγχάνεται ακριβής υπολογισμός της συνάρτησης V_n και ο καθορισμός της άριστης συνάρτησης ελέγχου δ^n, μέσω της

σχέσης (4.1.4) είναι εφικτός .Αν το προκαθορισμένο σφάλμα στον αλγόριθμο των ακρότατων σημείων είναι $\varepsilon > 0$, τότε αντί για την δ^n προσδιορίζουμε μια συνάρτηση ελέγχου $\tilde{\delta}^n$ σύμφωνα με τη σχέση

$$\tilde{V}_n = \tilde{H}\,\tilde{V}_{n-1} = H_{\tilde{\delta}^n}\,\tilde{V}_{n-1} \qquad \underline{4.1.5}$$

(βλέπε επίσης παρατήρηση 2 στην ενότητα 3.2).

Μπορούμε να προσεγγίσουμε την άριστη πολιτική $(\delta^*)^\infty$ με την στάσιμη πολιτική $(\tilde{\delta}^n)^\infty$. Η πρόταση που ακολουθεί παρέχει το σφάλμα αυτής της προσέγγισης και αποτελεί επέκταση της πρότασης 4.1.3.

Πρόταση 4.1.4: Εστω $\tilde{V}_n$ η προσέγγιση της V_n μέσω του αλγορίθμου των ακρότατων σημείων με προκαθορισμένο σφάλμα $\varepsilon > 0$ και $\tilde{\delta}^n$ η συνάρτηση ελέγχου στον χρονικό ορίζοντα n σύμφωνα με τη σχέση (4.1.5).Τότε η στάσιμη πολιτική $(\tilde{\delta}^n)^\infty$ είναι $\phi(\varepsilon,n)-$*άριστη*,όπου

$$\phi(\varepsilon,n) := \frac{1+\beta-2\beta^n}{(1-\beta)^2}\,\varepsilon + \frac{2\beta^n\,\Lambda}{1-\beta}\,, \qquad \underline{4.1.6}$$

δηλαδή:

$$\left\| V(./\tilde{\delta}^n) - V^* \right\| \le \phi(\varepsilon,n).$$

Απόδειξη

Από την τριγωνική ιδιότητα της νόρμας του supremum προκύπτει ότι:

$$\|V(./\tilde{\delta}^n) - V^*\| \le \|V(./\tilde{\delta}^n) - \tilde{V}_n\| + \|\tilde{V}_n - V^*\|. \qquad \underline{4.1.7}$$

Εφαρμόζοντας πάλι την τριγωνική ιδιότητα παίρνουμε

$$\|V(./\tilde{\delta}^n) - \tilde{V}_n\| \le \|V(./\tilde{\delta}^n) - H_{\tilde{\delta}^n}\tilde{V}_n\| + \|H_{\tilde{\delta}^n}\tilde{V}_n - \tilde{V}_n\|. \qquad \underline{4.1.8}$$

Λαμβάνοντας υπόψη ότι η συνάρτηση $V(./\tilde{\delta}^n)$ είναι το σταθερό σημείο του τελεστή $H_{\tilde{\delta}^n}$ και ότι ο τελεστής είναι συστολή modulus β, έχουμε:

$$\|V(./\tilde{\delta}^n) - H_{\tilde{\delta}^n}\tilde{V}_n\| = \|H_{\tilde{\delta}^n}V(./\tilde{\delta}^n) - H_{\tilde{\delta}^n}\tilde{V}_n\| \le \beta.\|V(./\tilde{\delta}^n) - \tilde{V}_n\|. \qquad \underline{4.1.9}$$

Από τις (4.1.8) και (4.1.9) παίρνουμε

:
$$\|V(./\tilde{\delta}^n) - \tilde{V}_n\| \le \frac{1}{1-\beta}.\|H_{\tilde{\delta}^n}\tilde{V}_n - \tilde{V}_n\|. \qquad \underline{4.1.10}$$

Λαμβάνοντας υπόψη τη σχέση (4.1.9) και την ιδιότητα της συστολής του τελεστή $H_{\tilde{\delta}^n}$ προκύπτει ότι:

$$\|H_{\tilde{\delta}^n}\tilde{V}_n - \tilde{V}_n\| = \|H_{\tilde{\delta}^n}\tilde{V}_n - H_{\tilde{\delta}^n}\tilde{V}_{n-1}\| \le \beta.\|\tilde{V}_n - \tilde{V}_{n-1}\|. \qquad \underline{4.1.11}$$

Από την τριγωνική ιδιότητα της νόρμας supremum έχουμε:

$$\|\tilde{V}_n - \tilde{V}_{n-1}\| \le \|\tilde{V}_n - V_n\| + \|V_n - V_{n-1}\| + \|V_{n-1} - \tilde{V}_{n-1}\| \le \sigma_n + \sigma_{n-1} + \|V_n - V_{n-1}\|, \qquad \underline{4.1.12}$$

όπου σ_n, σ_{n-1} είναι τα συσσωρευμένα σφάλματα των προσεγγίσεων $\tilde{V}_n, \tilde{V}_{n-1}$ για τις V_n, V_{n-1} αντίστοιχα (βλέπε ενότητα 3.4).Εφαρμόζοντας διαδοχικά την ιδιότητα συστολής του τελεστή μεγιστοποίησης Η παίρνουμε:

$$\|V_n - V_{n-1}\| = \|H_{n-1}V_1 - H_{n-1}V_0\| \le \beta^{n-1}.\|V_1 - V_0\| \le \beta^{n-1}.\Lambda. \qquad \underline{4.1.13}$$

Η τελευταία ανισότητα προκύπτει από το γεγονός ότι έχουμε επιλέξει ως συνάρτηση οφέλους στον τερματισμό τη μηδενική συνάρτηση, δηλαδή $V_0=0$ και επομένως

$$\| V_1 - V_0 \| = \|V_1\| = \| HV_0 \| \leq \max_{i,a} |q(i,\alpha)| = \Lambda$$

Από τις σχέσεις (4.1.10),(4.1.11),(4.1.12),(4.1.13) λαμβάνοντας υπόψη και την (3.4.7) παίρνουμε:

$$\|V(./\tilde{\delta}^n) - \tilde{V}_n\| \leq \frac{\beta.(\sigma_n + \sigma_{n-1})}{1-\beta} + \frac{\beta^n}{1-\beta}.\Lambda = \frac{\beta.(2-\beta^n-\beta^{n-1})}{(1-\beta)^2}.\varepsilon + \frac{\beta^n}{1-\beta}.\Lambda . \qquad \underline{4.1.14}$$

Από τις σχέσεις (4.1.7),(4.1.14) και (4.1.3) έχουμε

$$\|V(./\tilde{\delta}^n) - V^*\| \leq \frac{\beta.(2-\beta^n-\beta^{n-1})}{(1-\beta)^2}.\varepsilon + \frac{1-\beta^n}{1-\beta}.\varepsilon + \frac{2\beta^n}{1-\beta}.\Lambda =$$

$$= \frac{2\beta^n}{1-\beta}.\Lambda + \frac{1+\beta-2\beta^n}{(1-\beta)^2}.\varepsilon = \phi(\varepsilon,n)$$

Επομένως η στάσιμη πολιτική $(\tilde{\delta}^n)^\infty$ είναι $\phi(\varepsilon,n)-\alpha\rho\iota\sigma\tau\eta$. □

Σημειώνουμε ότι για ε=0 η πρόταση 4.1.4 δίνει το ίδιο φράγμα όπως και η πρόταση 4.1.3: $\phi(0,n) = \frac{2\beta^n.\Lambda}{1-\beta}$.

Αν επιθυμούμε η πολιτική $(\tilde{\delta}^n)^\infty$ να είναι $\eta-\alpha\rho\iota\sigma\tau\eta$ για δοσμένο $\eta>0$,τότε μπορούμε να επιλέξουμε προκαθορισμένο σφάλμα ε και χρονικό ορίζοντα n έτσι ώστε:

$$\phi(\varepsilon,n) \leq \eta .$$

Μπορούμε π.χ. να επιλέξουμε τα ε, n έτσι ώστε:

$$\frac{1+\beta}{(1-\beta)^2}.\varepsilon \leq \frac{\eta}{2} \quad , \quad \frac{2\beta^n}{1-\beta}.\Lambda \leq \frac{\eta}{2}$$

δηλαδή: $\varepsilon \le \frac{(1-\beta)^2.\eta}{2.(1+\beta)}$, $n \ge \frac{\ln(\frac{(1-\beta).\eta}{4\Lambda})}{\ln\beta}$.

Ένας εναλλακτικός τρόπος εύρεσης σχεδόν άριστων πολιτικών σχετίζεται με το κατάλοιπο Bellman (Bellman residual) σε κάποιο χρονικό ορίζοντα n, που ορίζεται ως η μέγιστη απόλυτη διαφορά των συναρτήσεων V_n και V_{n-1} ,δηλαδή η ποσότητα $\| V_n - V_{n-1} \|$. Αν αυτή η ποσότητα είναι ΄΄αρκούντως μικρή΄΄,τότε στην περίπτωση αυτή μπορούμε να προσεγγίσουμε την άριστη πολιτική $(\delta^*)^\infty$ με την πολιτική $(\delta^n)^\infty$,όπου δ^n είναι η άριστη συνάρτηση ελέγχου για τον χρονικό ορίζοντα σύμφωνα με τη σχέση (4.1.4).Με άλλα λόγια η συνάρτηση του αναμενόμενου ολικού εκπίπτοντος κέρδους για άπειρο χρονικό ορίζοντα εφαρμόζοντας την πολιτική $(\delta^n)^\infty$,$V(./\delta^n)$, προσεγγίζει ικανοποιητικά τη βέλτιστη συνάρτηση τιμών $\mathbf{V}^*$.

Πρόταση 4.1.5: Έστω ότι για το κατάλοιπο Bellman σε κάποιο χρονικό ορίζοντα n ισχύει

$$\| V_n - V_{n-1} \| \le \eta \quad (\text{όπου } \eta > 0)$$

και δ^n είναι η άριστη συνάρτηση ελέγχου στον χρονικό ορίζοντα n ,σύμφωνα με τη σχέση (4.1.4). Τότε η στάσιμη πολιτική $(\delta^n)^\infty$ είναι $\frac{2\beta.\eta}{1-\beta}-$ *άριστη*,δηλαδή

$$\| V(./\delta^n) - V^* \| \le \frac{2\beta.\eta}{1-\beta} \qquad \underline{\mathbf{4.1.15}}$$

Puterman [98]

□

Είναι φανερό ότι η παραπάνω πρόταση είναι εφαρμόσιμη μόνο στην περίπτωση όπου εφαρμόζουμε τον αλγόριθμο των ακρότατων σημείων με προκαθορισμένο

σφάλμα $\varepsilon = 0$, οπότε επιτυγχάνεται ακριβής υπολογισμός των συναρτήσεων V_n, των καταλοίπων Bellman και των συναρτήσεων ελέγχου δ^n για n=1,2,...

Στην περίπτωση όπου εφαρμόζουμε τον αλγόριθμο των ακρότατων σημείων με προκαθορισμένο σφάλμα $\varepsilon > 0$, χρησιμοποιούμε το ''τροποποιημένο κατάλοιπο Bellman'',που ορίζεται ως η μέγιστη απόλυτη διαφορά των συναρτήσεων $\tilde{V}_n$ και $\tilde{V}_{n-1}$, δηλαδή η ποσότητα $\|\tilde{V}_n - \tilde{V}_{n-1}\|$.

Αν για κάποιο **n** η ποσότητα αυτή καθώς και το προκαθορισμένο σφάλμα ε είναι ''αρκούντως μικρά'', τότε στην περίπτωση αυτή μπορούμε να προσεγγίσουμε την άριστη πολιτική $(\delta^*)^\infty$ με την πολιτική $(\tilde{\delta}^n)^\infty$,όπου η συνάρτηση ελέγχου $(\tilde{\delta}^n)$ καθορίζεται σύμφωνα με τη σχέση (4.1.5). Η επόμενη πρόταση καλύπτει αυτή την περίπτωση και αποτελεί επέκταση της πρότασης 4.1.5.

Πρόταση 4.1.6: Έστω ότι για το τροποποιημένο κατάλοιπο Bellman, σε κάποια χρονικό ορίζοντα n ισχύει

$$\|\tilde{V}_n - \tilde{V}_{n-1}\| \le \eta \quad (\text{όπου } \eta > 0) \qquad \underline{4.1.16}$$

και $\tilde{\delta}^n$ είναι η συνάρτηση ελέγχου στον χρονικό ορίζοντα n σύμφωνα με τη σχέση (4.1.5). Τότε η στάσιμη πολιτική $(\tilde{\delta}^n)^\infty$ είναι $f(\varepsilon,\eta)-$*άριστη*,όπου

$$f(\varepsilon,\eta) := \frac{1+\beta-2\beta^n}{(1-\beta)^2}\varepsilon + \frac{2\beta\eta}{1-\beta}, \qquad \underline{4.1.17}$$

δηλαδή: $\| V(/\tilde{\delta}^n) - V^* \| \le f(\varepsilon,\eta)$

Απόδειξη

Από την τριγωνική ιδιότητα της νόρμας του supremum παίρνουμε

$$\left\| V(./\tilde{\delta}^n) - V^* \right\| \le \left\| V(./\tilde{\delta}^n) - \tilde{V}_n \right\| + \left\| \tilde{V}_n - V_n \right\| + \left\| V_n - V^* \right\| \le$$

$$\le \left\| V(./\tilde{\delta}^n) - \tilde{V}_n \right\| + \sigma_n + \left\| V_n - V^* \right\| \qquad \underline{4.1.18}$$

Εφαρμόζοντας πάλι την τριγωνική ιδιότητα έχουμε

$$\left\| V(./\tilde{\delta}^n) - \tilde{V}_n \right\| \le \left\| V(./\tilde{\delta}^n) - H_{\tilde{\delta}^n} \tilde{V}_n \right\| + \left\| H_{\tilde{\delta}^n} \tilde{V}_n - \tilde{V}_n \right\| \qquad \underline{4.1.19}$$

Λαμβάνοντας υπόψη ότι η συνάρτηση $V(./\tilde{\delta}^n)$ είναι το σταθερό σημείο του τελεστή $H_{\tilde{\delta}^n}$,ότι ο τελεστής είναι συστολή modulus β,και τις σχέσεις (4.1.5),(4.1.16), από την (4.1.19) παίρνουμε:

$$\left\| V(./\tilde{\delta}^n) - \tilde{V}_n \right\| \le \left\| H_{\tilde{\delta}^n} V(./\tilde{\delta}^n) - H_{\tilde{\delta}^n} \tilde{V}_n \right\| + \left\| H_{\tilde{\delta}^n} \tilde{V}_n - H_{\tilde{\delta}^n} \tilde{V}_{n-1} \right\| \le$$

$$\le \beta . \left\| V(./\tilde{\delta}^n) - \tilde{V}_n \right\| + \beta . \left\| \tilde{V}_n - \tilde{V}_{n-1} \right\| \le$$

$$\le \beta . \left\| V(./\tilde{\delta}^n) - \tilde{V}_n \right\| + \beta . \eta$$

Επομένως

$$\left\| V(./\tilde{\delta}^n) - \tilde{V}_n \right\| \le \frac{\beta . \eta}{1 - \beta} \qquad \underline{4.1.20}$$

Από το λήμμα 4.1.1 (iii) παίρνουμε:

$$\left\| V_n - V^* \right\| = \left\| H V_{n-1} - V^* \right\| \le \frac{\beta}{1 - \beta} \left\| H V_{n-1} - V_{n-1} \right\|$$

Άρα

$$\| V_n - V^* \| \le \frac{\beta}{1-\beta} \| V_n - V_{n-1} \| \qquad \underline{4.1.21}$$

Από την τριγωνική ιδιότητα και τη σχέση (4.1.16) έχουμε

$$\| V_n - V_{n-1} \| \le \| V_n - \tilde{V}_n \| + \| \tilde{V}_n - \tilde{V}_{n-1} \| + \| \tilde{V}_{n-1} - V_{n-1} \| \le \sigma_n + \eta + \sigma_{n-1} \qquad \underline{4.1.22}$$

Από τις σχέσεις (4.1.21),(4.1.22) έχουμε:

$$\| V_n - V^* \| \le \frac{\beta}{1-\beta} \| V_n - V_{n-1} \| \le \frac{\beta}{1-\beta} .(\sigma_n + \eta + \sigma_{n-1}) \qquad \underline{4.1.23}$$

Από τις σχέσεις (4.1.18),(4.1.20),(4.1.23) λαμβάνοντας υπόψη και την (3.4.7) παίρνουμε:

$$\| V(./\tilde{\delta}^n) - V^* \| \le \frac{\beta.\eta}{1-\beta} + \sigma_n + \frac{\beta.(\sigma_n + \sigma_{n-1} + \eta)}{1-\beta} =$$

$$= \frac{\sigma_n + \beta.\sigma_{n-1}}{1-\beta} + \frac{2\beta\eta}{1-\beta} = \frac{1+\beta-2\beta^n}{(1-\beta)^2} .\varepsilon + \frac{2\beta.\eta}{1-\beta} = f(\varepsilon,\eta)$$

Επομένως η στάσιμη πολιτική $(\tilde{\delta}^n)^\infty$ είναι $f(\varepsilon,\eta) - \alpha\rho\iota\sigma\tau\eta$. □

Σημειώνουμε ότι για ε=0 η πρόταση 4.1.6 δίνει το ίδιο φράγμα όπως και η πρόταση 4.1.5 : $f(0,\eta) = \frac{2\beta.\eta}{1-\beta}$. Αν επιθυμούμε η πολιτική $(\tilde{\delta}^n)^\infty$ να είναι $\lambda - \alpha\rho\iota\sigma\tau\eta$ για δοσμένο $\lambda > 0$, τότε μπορούμε να επιλέξουμε προκαθορισμένο σφάλμα ε και φράγμα η για το τροποποιημένο κατάλοιπο Bellman σε κάποιο χρονικό ορίζοντα έτσι ώστε:

$$f(\varepsilon,\eta) \le \lambda$$

Μπορούμε π.χ. να επιλέξουμε τα ε, η έτσι ώστε:

$$\frac{1+\beta}{(1-\beta)^2}.\varepsilon \le \frac{\lambda}{2} \quad , \quad \frac{2\beta.\eta}{1-\beta} \le \frac{\lambda}{2}$$

δηλαδή: $$\varepsilon \le \frac{(1-\beta)^2.\lambda}{2.(1+\beta)} \quad , \quad \eta \le \frac{(1-\beta).\lambda}{4\beta}$$

Οι προτάσεις αυτής της ενότητας ισχύουν επίσης για προσεγγίσεις της άριστης συνάρτησης τιμών V^* και της άριστης πολιτικής $(\delta^*)^\infty$ σε προβλήματα κόστους (με την προφανή αλλαγή $\Lambda = \max_{i,a} |q(\iota,\alpha)|$).

4.2. Κατασκευή φραγμάτων για την βέλτιστη συνάρτηση τιμών.

Στην ενότητα αυτή αναφέρουμε ορισμένα φράγματα για τη βέλτιστη συνάρτηση τιμών V* που απαντώνται στη βιβλιογραφία.Η διαδικασία δημιουργίας προσεγγίσεων της V* και της άριστης πολιτικής σε άπειρο χρονικό ορίζοντα από φράγματα θα μας απασχολήσει στην ενότητα 4.3. Θα περιοριστούμε στην κατασκευή φραγμάτων μόνο για προβλήματα POMDP μεγιστοποίησης εσόδων.

Α) Κατασκευή άνω φραγμάτων για την V*

1)Μέσω της βέλτιστης συνάρτησης τιμών της αντίστοιχης MDP.(Boutilier-Poole[17].

Τα απλούστερα μη τετριμμένα άνω φράγματα για την συνάρτηση V* μιας POMDP επιτυγχάνονται μέσω της βέλτιστης συνάρτησης τιμών V^*_{MDP} της αντίστοιχης (πλήρως παρατηρήσιμης Μαρκοβιανής διαδικασίας αποφάσεων (MDP)που ικανοποιεί την ακόλουθη εξίσωση αριστοποίησης.

$$V^*_{MDP}(i) = \max_a \{ q(i,a) + \beta . \sum_{j=1}^{j=N} p_{ij}^{\alpha} . V^*_{MDP}(j) \},\ i=1,2,\ldots,N$$

Η συνάρτηση V^*_{MDP} υπολογίζεται απλά και επακριβώς με την επαναληπτική μέθοδο πολιτικής (policy-iteration) σε πεπερασμένο αριθμό βημάτων (βλέπε Howard [51]). Σύμφωνα με τους Boutilier-Poole [17] έχουμε τα ακόλουθα άνω φράγματα για την συνάρτηση V*: Για κάθε $\pi=(\pi_1,\pi_2,\ldots\pi_N)\in\Pi$,

$$V^*(\pi) \le \max_a \sum_{i=1}^{N} \pi_i \, (q(i,a) + \beta . \sum_{j=1}^{j=N} p_{ij}^{\alpha} . V^*_{MDP}(j)) \qquad \underline{4.2.1}$$

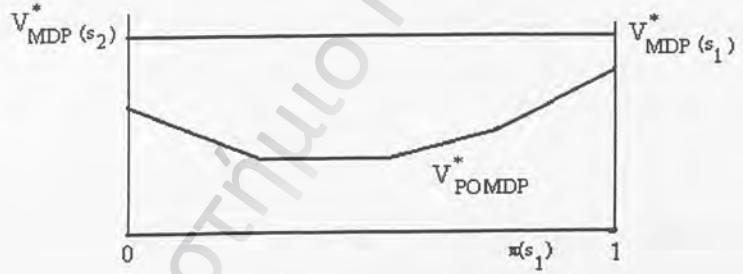
$$V^*(\pi) \le \sum_{i=1}^{N} \pi_i \, . V^*_{MDP}(i). \qquad \underline{4.2.2}$$

Επειδή

$$\max_a \sum_{i=1}^{N} \pi_i \, (q(i,a) + \beta . \sum_{j=1}^{j=N} p_{ij}^{\alpha} . V^*_{MDP}(j)) \le \sum_{i=1}^{N} \pi_i \, . \max_a (q(i,a) + \beta . \sum_{j=1}^{j=N} p_{ij}^{\alpha} . V^*_{MDP}(j)) =$$

$$= \sum_{i=1}^{N} \pi_i . V^*_{MDP}(i) \qquad \forall \, \pi = (\pi_1, \pi_2, \ldots \pi_N) \in \Pi$$

συνάγεται ότι το άνω φράγμα της (4.2.2) ως προσέγγιση της V* υπολείπεται του αντίστοιχου άνω φράγματος της (4.2.1),όμως ο υπολογισμός του είναι απλούστερος. Η ιδέα της προσέγγισης διευκρινίζεται στο σχήμα 4.1, όπου έχουμε μια POMDP δύο καταστάσεων s_1, s_2 και την MDP προσέγγισή της.



Σχήμα 4.1:Προσέγγιση βασιζόμενη σε μια πλήρως παρατηρήσιμη MDP για μια POMDP δύο καταστάσεων s_1, s_2.

2) Μέσω των συναρτήσεων V_n (Lovejoy [75])

Μια ακολουθία άνω φραγμάτων για την V* σχηματίζεται μέσω της ακολουθίας των συναρτήσεων $\{V_n\}$:

$$V_n = HV_{n-1}, n = 1,2,... \qquad \underline{4.2.3}$$

$$V_0 = 0 \, (\text{μηδενική συνάρτηση}),$$

όπου Η είναι ο τελεστής μεγιστοποίησης. Για n=0,1,2,... έχουμε:

$$V^*(\pi) \le V_n(\pi) + \frac{\beta^n}{1-\beta} q_{\max} \quad \forall \pi \in \Pi . \qquad \underline{4.2.4}$$

όπου $$q_{\max} = \max_{i,a} q(i,a) .$$

Για n=0 έχουμε την ειδική περίπτωση

$$V^*(\pi) \le \frac{1}{1-\beta} q_{\max} \quad \forall \pi \in \Pi .$$

Β) Κατασκευή κάτω φραγμάτων για την V*

1)Μέσω της βέλτιστης συνάρτησης τιμών της αντίστοιχης UMDP (unobservable MDP), μη παρατηρήσιμη Μαρκοβιανή διαδικασία αποφάσεων, χωρίς μηνύματα (Madani [78]).

Ένα κάτω φράγμα για την βέλτιστη συνάρτηση τιμών V* μιας POMDP είναι η βέλτιστη συνάρτηση τιμών V^*_{UMDP} της αντίστοιχης μη παρατηρήσιμης διαδικασίας αποφάσεων (UMDP),που ικανοποιεί την ακόλουθη εξίσωση αριστοποίησης:

$$V^*_{UMDP}(\pi) = \max_a \{\pi . q(i,a) + \beta V^*_{UMDP}(T(\pi,\alpha))\}, \pi \in \Pi$$

Όπου $$T(\pi,\alpha) = \pi . P^\alpha \ , \pi \in \Pi, a \in A.$$

Αποδεικνύεται ότι $$V^*_{UMDP} \le V^*(\pi) \quad \forall \pi \in \Pi .$$

Παρόλο που το πρόβλημα UMDP είναι απλούστερο από το αντίστοιχο POMDP, ο υπολογισμός της V^*_{UMDP} δεν είναι απλός και σε γενικές γραμμές ακολουθούνται οι ίδιες μέθοδοι προσέγγισης με την V^*.Επομένως το κάτω φράγμα V^*_{UMDP} για την V^* έχει μόνο θεωρητική αξία.

2) Μέσω των συναρτήσεων V_n (Lovejoy [76])

Μια ακολουθία κάτω φραγμάτων για την V^* σχηματίζεται μέσω της ακολουθίας των συναρτήσεων $\{V_n\}$ που ορίζονται από την (9.2.3). Για n=0,1,2,... έχουμε:

$$V_n(\pi) + \frac{\beta^n}{1-\beta} q_{\min} \le V^*(\pi) \quad \forall \pi \in \Pi . \qquad \underline{4.2.5}$$

$$q_{\min} = \min_{i,a} q(i,a) .$$

Για n=0 έχουμε την ειδική περίπτωση

$$\frac{1}{1-\beta}.q_{\min} \le V^*(\pi) \qquad \forall \pi \in \Pi .$$

3) Μέσω κάτω φραγμάτων των συναρτήσεων V_n (Lovejoy [76]).

Έστω $\upsilon(\pi), \pi \in \Pi$ μια κατά τμήματα γραμμική και κυρτή συνάρτηση. Όπως είναι γνωστό, ή συνάρτηση Hυ είναι επίσης κατά τμήματα γραμμική και κυρτή συνάρτηση, πού καθορίζεται μέσω ενός πεπερασμένου συνόλου Γ_H από gradients (διανύσματα) του χώρου R^N:

$$H\upsilon(\pi) = \max_{\gamma \in \Gamma_H} \pi.\gamma, \ \pi \in \Pi .$$

Έστω $\gamma(\pi) \in \Gamma_H$ το gradient της Hυ στο π, δηλαδή:

$$H\upsilon(\pi) = \pi.\gamma(\pi).$$

Το gradient $\gamma(\pi)$ προσδιορίζεται μέσω του αλγορίθμου του ενός βήματος (βλέπε ενότητα 2.2).

Ένα κάτω φράγμα για τη συνάρτηση Hυ κατασκευάζεται ως εξής:Θεωρούμε $G \subseteq \Pi$ ένα πεπερασμένο σύνολο από δ.π (grid points) και Γ_L το σύνολο των gradients της συνάρτησης Hυ στα δ.π του συνόλου G, δηλαδή

$$\Gamma_L = \{\gamma(\pi): \pi \in G\}.$$

Ορίζουμε τη συνάρτηση

$$H_L\upsilon(\pi) := \max_{\gamma \in \Gamma_L} \pi.\gamma, \ \pi \in \Pi .$$

Επειδή $\Gamma_L \subseteq \Gamma_H$ συνάγεται ότι

$$H_L\upsilon(\pi) \le H\upsilon(\pi) \ \ \forall \pi \in \Pi .$$

Ακολούθως κατασκευάζεται η ακολουθία των συναρτήσεων $\{V_{L_n}\}$

$$V_{L_n} = H_L V_{L_{n-1}}, n=1,2,3,\ldots$$

Όπου $V_{L_0} = 0$ (μηδενική συνάρτηση).

Σημειώνουμε ότι το σύνολο G παραμένει το ίδιο για όλα τα n.Με άλλα λόγια, το σύνολο των gradients της συνάρτησης V_{L_n} είναι το σύνολο των gradients της $H_L V_{L_{n-1}}$ στα δ.π του G. Αποδεικνύεται επαγωγικά ότι:

$$V_{L_n}(\pi) \le V_n(\pi) \quad \forall \pi \in \Pi .$$

Μια ακολουθία κάτω φραγμάτων για την V^* σχηματίζεται μέσω της ακολουθίας των συναρτήσεων $\{V_{L_n}\}$.Για n=0,1,2... έχουμε:

$$V_{L_n}(\pi) + \frac{\beta^n}{1-\beta} q_{\min} \le V^*(\pi) \quad \forall \pi \in \Pi . \qquad \underline{4.2.6}$$

Προφανώς το κάτω φράγμα της (4.2.6) ως προσέγγιση της V^* υπολείπεται του αντίστοιχου κάτω φράγματος της (4.2.5),όμως ο υπολογισμός του είναι γενικά πολύ απλούστερος.

Απαλοιφή αποφάσεων πού δεν είναι άριστες για κάποιο δ.π

Η εύρεση άνω και κάτω φραγμάτων για την βέλτιστη συνάρτηση τιμών V^* είναι δυνατόν να συμβάλει στην απαλοιφή μη άριστων αποφάσεων για κάποιο δ.π στο πρόβλημα POMDP για άπειρο χρονικό ορίζοντα.

Θεωρούμε τη συνάρτηση $h : \Pi \times A \times B(\Pi) \to \mathbb{R}$

πού ορίζεται ως:

$$h(\pi, a, u) := \pi.q^{\alpha} + \beta \sum_{\theta} \{\theta / \pi, \alpha\} .u\,(T(\pi,\theta,\alpha)\} , \ \forall \ \pi \in \Pi , \alpha \in A , u \in B(\pi) .$$

(βλέπε ενότητα 1.4).

Έστω V_L,V_U κάτω και άνω φράγμα αντίστοιχα για τήν V^* :

$$V_L(\pi) \le V^*(\pi) \le V_U(\pi) \quad \forall \ \pi \in \Pi .$$

Η επόμενη πρόταση παρέχει ένα κριτήριο απαλοιφής μη άριστων αποφάσεων για κάποιο δ.π στο πρόβλημα του άπειρου χρονικού ορίζοντα.

Πρόταση 4.2.1 :(MC Queen)[100]

Αν για κάποιο $\pi \in \Pi$ και κάποια απόφαση α ισχύει

$h(\pi, a, V_U) < HV_L(\pi)$,

τότε η απόφαση α δεν είναι άριστη για το π στο πρόβλημα του άπειρου χρονικού ορίζοντα. □

4.3. Προσεγγίσεις της άριστης συνάρτησης τιμών για άπειρο χρονικό ορίζοντα και προσδιορισμός σχεδόν άριστων πολιτικών μέσω φραγμάτων.

Στην παρούσα ενότητα θα ασχοληθούμε με προσεγγίσεις της άριστης συνάρτησης τιμών V^* καθώς και με τον προσδιορισμό σχεδόν άριστων πολιτικών, πού βασίζονται σε συναρτήσεις φραγμάτων της V^*.

Ο Hauskrecht [48], λαμβάνοντας δύο οποιαδήποτε άνω και κάτω φράγματα ως αρχικές προσεγγίσεις της V^* και εφαρμόζοντας την επαναληπτική μέθοδο τιμών (value-iteration) κατασκευάζει νέα φράγματα,που αποτελούν αυθαίρετα καλές προσεγγίσεις της V^* και οι προβαλλόμενες από αυτά πολιτικές (lookahead-controllers) είναι σχεδόν άριστες πολιτικές.Το πλήθος των βημάτων (επαναλήψεων) που απαιτούνται εξαρτάται από την «απόσταση» των αρχικών φραγμάτων και την επιθυμητή ακρίβεια της προσέγγισης. Εμείς επεκτείνουμε αυτά τα αποτελέσματα εφαρμόζοντας προσεγγιστικό αλγόριθμο των ακρότατων σημείων. Επιπλέον υπολογίζουμε τον αριθμό των βημάτων καθώς και το προκαθορισμένο σφάλμα του αλγορίθμου, ώστε να επιτύχουμε οποιαδήποτε επιθυμητή ακρίβεια προσέγγισης. Θα περιορισθούμε μόνο σε προσεγγίσεις προβλημάτων POMDP στα πλαίσια του κριτηρίου μεγιστοποίησης των ολικών εσόδων σε άπειρο χρονικό ορίζοντα. Οι αντίστοιχες προσεγγίσεις στα πλαίσια του κριτηρίου ελαχιστοποίησης του ολικού κόστους για άπειρο χρονικό ορίζοντα είναι ανάλογες.

Έστω $V_L(\pi)$, $V_U(\pi), \pi \in \Pi$ συνάρτηση κάτω και άνω φράγματος αντίστοιχα για την συνάρτηση V^*:

$$V_L(\pi) \le V^*(\pi) \le V_U(\pi), \pi \in \Pi .$$

Πρόταση 4.3.1: Milos Hauskrecht [48]

Ας είναι $\eta = \|V_U - V_L\| = \sup_{\pi} |V_U(\pi) - V_L(\pi)|$ και V οποιοδήποτε από τα δύο φράγματα, δηλαδή $V=V_L$ η $V=V_U$.Τότε

i) $\|V - V^*\| \le \eta$

ii) Αν δ είναι η συνάρτηση ελέγχου για την οποία

$$H_g V = HV$$

δηλαδή

$$\delta(\pi) := \arg\max_{a} \{ \pi q^{a} + \beta \sum_{\theta=1}^{M} \{\theta/\pi,\alpha\} V(T(\pi,\theta,\alpha))\}, \pi \in \Pi$$

τότε η στάσιμη πολιτική δ^{∞} είναι $\frac{2-\beta}{1-\beta}\eta$ *-άριστη* ,δηλαδή:

$$\| V(./\delta) - V^{*} \| \leq \frac{2-\beta}{1-\beta}\eta \qquad \square$$

Ο αριθμός η εκφράζει την απόσταση (μέγιστη απόλυτη διαφορά) ανάμεσα στα φράγματα V_L και V_U. Η πρόταση 4.3.1 δηλώνει ότι οποιοδήποτε από τα φράγματα V_L, V_U μπορεί να θεωρηθεί προσέγγιση της βέλτιστης συνάρτησης τιμών V^* με μέγιστο σφάλμα το πολύ η. Επίσης οποιοδήποτε από τα δύο φράγματα V_L, V_U μπορεί να χρησιμοποιηθεί ως εφαλτήριο για την εύρεση $\frac{2-\beta}{1-\beta}\eta$ -άριστης πολιτικής.

Η δ αντιπροσωπεύει την βέλτιστη συνάρτηση ελέγχου «για ένα βήμα μπροστά»(one-step lookahead controller) αναφορικά με το φράγμα V που επιλέγουμε.

Η συνάρτηση τιμών που αντιστοιχεί στην πολιτική δ^{∞},$V(./\delta)$, προσεγγίζει τη βέλτιστη συνάρτηση τιμών V^* με μέγιστο σφάλμα το πολύ $\frac{2-\beta}{1-\beta}\eta$. *Η πολιτική δ^{∞} αναφέρεται ως προβαλλόμενη πολιτική από το φράγμα V (lookahead- policy).*

Θεωρούμε τις αναγωγικές σχέσεις:

$$V_L^k = HV_L^{k-1}, V_U^k = HV_U^{k-1}, k = 1,2,3,\ldots$$

$$V_L^0 = V_L \quad , V_U^0 = V_U$$

Εισάγοντας τον τελεστή H_k πού συμβολίζει την επαναληπτική χρήση του τελεστή Η k φορές, οι παραπάνω σχέσεις γράφονται:

$$V_L^k = H_k V_L \leq H_k V^* = V^* \leq H_k V_U = V_U^k$$

Επομένως οι συναρτήσεις V_L^k, V_U^k αποτελούν αντίστοιχα κάτω και άνω φράγμα για την V*. *Θα αναφερόμαστε σε αυτές ως φράγματα τάξεως k.* Επειδή ο τελεστής H_k είναι συστολή modulus β^k έχουμε:

$$\left\| V_U^k - V_L^k \right\| \le \beta^k \left\| V_L - V_U \right\| = \beta^k . \eta .$$

Συνοψίζοντας, από τα φράγματα V_L, V_U για την συνάρτηση V*,εφαρμόζοντας επαναληπτικά τον τελεστή H k φορές, σχηματίζονται τα φράγματα τάξεως k V_L^k, V_U^k για την V* των οποίων η απόσταση (μέγιστη απόλυτη διαφορά) είναι το πολύ $\beta^k . \eta$.

Από την πρόταση 4.3.1 (i) συνάγεται ότι τα φράγματα k τάξεως μπορούν να θεωρηθούν προσεγγίσεις της συνάρτησης V* με μέγιστο σφάλμα το πολύ $\beta^k . \eta$. Πιο συγκεκριμένα έχουμε:

$$\left\| V^k - V^* \right\| \le \beta^k . \eta , \qquad \underline{\textbf{4.3.1}}$$

όπου V^k είναι οποιοδήποτε από τα δύο φράγματα k τάξεως, δηλαδή $V^k = V_L^k$ ή $V^k = V_U^k$.

Αν $\lambda > 0$ είναι ένα επιθυμητό άνω φράγμα για το σφάλμα προσέγγισης

$$\left\| V^k - V^* \right\| \le \lambda$$

τότε μπορούμε να επιλέξουμε την τάξη k έτσι ώστε:

$$\beta^k . \eta \le \lambda$$

δηλαδή

$$k \ge \frac{\ln(\lambda / \eta)}{\ln \beta}$$

Εστω δ^k η βέλτιστη συνάρτηση ελέγχου για ένα βήμα μπροστά (one-step lookahead conroller) αναφορικά με το φράγμα V^k που επιλέγουμε , δηλαδή:

$$H_{\delta^k} V^k = \mathrm{H} V^k \qquad \underline{\textbf{4.3.2}}$$

Από την πρόταση 4.3.1 (ii) συνάγεται ότι η προβαλλόμενη από το φράγμα V^k πολιτική $(\delta^k)^\infty$ (lookahead policy) είναι $\frac{2-\beta}{1-\beta} . \beta^k . \eta - \acute{\alpha} \rho \iota \sigma \tau \eta .$

Αν επιθυμούμε η πολιτική $(\delta^k)^\infty$ να είναι λ-άριστη για δοσμένο $\lambda \succ 0$, τότε μπορούμε να επιλέξουμε την τάξη k έτσι ώστε:

$$\frac{2-\beta}{1-\beta}.\beta^k.\eta \le \lambda$$

δηλαδή

$$k \ge \frac{\ln(\lambda.(1-\beta)/\eta.(2-\beta))}{\ln\beta}$$

Σημειώνουμε ότι οι προσεγγίσεις πού περιγράψαμε είναι εφικτές μόνο στην περίπτωση όπου εφαρμόζουμε διαδοχικά τον ακριβή αλγόριθμο των ακρότατων σημείων του κεφ. 3 (προκαθορισμένο σφάλμα ε=0), οπότε επιτυγχάνεται ακριβής υπολογισμός των φραγμάτων k τάξεως V_L^k, V_U^k καθώς επίσης ο καθορισμός της συνάρτησης ελέγχου δ^k μέσω της σχέσης (4.3.2) είναι εφικτός.

Ανάλογες προσεγγίσεις είναι δυνατές εφαρμόζοντας διαδοχικά τον προσεγγιστικό αλγόριθμο των ακρότατων σημείων. Επιλέγοντας ως αρχικές συναρτήσεις τις συναρτήσεις φραγμάτων V_L, V_U και εφαρμόζοντας k φορές τον αλγόριθμο των ακροτάτων σημείων με προκαθορισμένο σφάλμα $\varepsilon > 0$ υπολογίζονται οι προσεγγίσεις $\tilde{V}_L^k, \tilde{V}_U^k$ των φραγμάτων k τάξης V_L^k, V_U^k αντίστοιχα, με συσσωρευμένο σφάλμα προσέγγισης σ_k σε κάθε περίπτωση (πρβλ. ενότητα 3.4). Συγκεκριμένα,

$$\tilde{V}_L^k = \tilde{\mathrm{H}}\tilde{V}_L^{k-1}, \qquad \tilde{V}_U^k = \tilde{\mathrm{H}}\tilde{V}_U^{k-1}, \quad k=1,2,3\ldots$$

$$\tilde{V}_L^0 = V_L, \qquad \tilde{V}_U^0 = V_U$$

Για κ =1,2,3,... έχουμε:

$$\tilde{V}_L^k(\pi) \le V_L^k(\pi) \le \tilde{V}_L^k(\pi) + \sigma_k \qquad \forall \pi \in \Pi$$

$$\tilde{V}_U^k(\pi) \le V_U^k(\pi) \le \tilde{V}_U^k(\pi) + \sigma_k \qquad \forall \pi \in \Pi \qquad \textbf{4.3.3}$$

Το συσσωρευμένο σφάλμα προσέγγισης σ_k ικανοποιεί την αναγωγική σχέση (3.4.6) και υπολογίζεται από την σχέση (3.4.7).

Σημειώνουμε ότι η προσέγγιση $\tilde{V}_L^k$ του κάτω φράγματος k τάξεως V_L^k είναι επίσης κάτω φράγμα για την συνάρτηση V^*. Δεν μπορούμε όμως να ισχυριστούμε το ίδιο για την προσέγγιση $\tilde{V}_U^k$ του άνω φράγματος k τάξεως V_U^k πού ενδέχεται να μην είναι άνω φράγμα της V^*. Ωστόσο αμφότερες οι συναρτήσεις $\tilde{V}_L^k, \tilde{V}_U^k$

μπορούν να ληφθούν ως προσεγγίσεις της συνάρτησης V^*. Το σφάλμα αυτών των προσεγγίσεων παρέχεται στην επόμενη πρόταση.

Παρατήρηση

Στα προβλήματα κόστους ισχύει το αντίστροφο. Πράγματι (βλέπε ενότητα 3.5) για k=1,2,3…έχουμε:

$$\tilde{V}_L^k(\pi) - \sigma_k \le V_L^k(\pi) \le \tilde{V}_L^k(\pi)$$

$$\tilde{V}_U^k(\pi) - \sigma_k \le V_U^k(\pi) \le \tilde{V}_U^k(\pi)$$

Επομένως η προσέγγιση $\tilde{V}_U^k(\pi)$ του άνω φράγματος $V_U^k(\pi)$ εξακολουθεί να είναι άνω φράγμα για την συνάρτηση V*,ενώ η προσέγγιση $\tilde{V}_L^k(\pi)$ του κάτω φράγματος $V_L^k(\pi)$,ενδέχεται να μην είναι κάτω φράγμα για την V*.

Ωστόσο και στην περίπτωση αυτή αμφότερες οι συναρτήσεις $\tilde{V}_L^k, \tilde{V}_U^k$,μπορούν να ληφθούν ως προσεγγίσεις της V*.

Πρόταση 4.3.2: Αν $\eta = \|V_U - V_L\| = \sup_{\pi} |V_U(\pi) - V_L(\pi)|$,τότε για k=1,2,3…

i) $\|\tilde{V}_L^k - V^*\| \le \frac{1-\beta^k}{1-\beta}\varepsilon + \beta^k \eta$,

ii) $\|\tilde{V}_U^k - V^*\| \le \frac{1-\beta^k}{1-\beta}\varepsilon + \beta^k \eta$.

Απόδειξη

i) Από την τριγωνική ιδιότητα της νόρμας supremum και τις σχέσεις (4.3.1),(4.3.3) και (3.4.7) παίρνουμε:

$$\|\tilde{V}_L^k - V^*\| \le \|\tilde{V}_L^k - V_L^k\| + \|V_L^k - V^*\| \le \sigma_k + \beta^k \eta =$$

$$= \frac{1-\beta^k}{1-\beta}\varepsilon + \beta^k \eta$$

ii) Παρόμοια □

Σημειώνουμε ότι είναι δυνατόν να επιτευχθούν προσεγγίσεις της συνάρτησης V* με οποιαδήποτε επιθυμητή ακρίβεια. Συγκεκριμένα αν $\lambda \succ 0$ είναι ένα επιθυμητό άνω φράγμα για το μέγιστο σφάλμα προσέγγισης, δηλαδή

$$\left\| \tilde{V}^k - V^* \right\| \le \lambda$$

όπου $\tilde{V}^k$ είναι οποιαδήποτε από τις δύο προσεγγίσεις $\tilde{V}_L^k, \tilde{V}_U^k$, τότε μπορούμε να επιλέξουμε προκαθορισμένο σφάλμα ε και τάξη k έτσι ώστε

$$\frac{1-\beta^k}{1-\beta}.\varepsilon + \beta^k.\eta \le \lambda$$

Μπορούμε π.χ να επιλέξουμε τα ε, k έτσι ώστε:

$$\frac{\varepsilon}{1-\beta} \le \frac{\lambda}{2}, \qquad \beta^k.\eta \le \frac{\lambda}{2}$$

δηλαδή

$$\varepsilon \le \frac{\lambda.(1-\beta)}{2}, \quad k \ge \frac{\ln(\frac{\lambda}{2\eta})}{\ln(\beta)}$$

Εστω $\tilde{\delta}^k$ η συνάρτηση ελέγχου που προκύπτει εφαρμόζοντας τον αλγόριθμο των ακρότατων σημείων με προκαθορισμένο σφάλμα ε> 0 στην προσέγγιση $\tilde{V}^k$, δηλαδή

$$H_{\tilde{\delta}^k}\tilde{V}^k = \tilde{H}\tilde{V}^k \qquad \underline{\textbf{4.3.4}}$$

(βλέπε επίσης παρατήρηση 2 στην ενότητα 3.2)

Η στάσιμη πολιτική $(\tilde{\delta}^k)^\infty$ μπορεί να ληφθεί ως προσέγγιση της άριστης πολιτικής $(\delta^*)^\infty$. Το σφάλμα αυτής της προσέγγισης παρέχεται στην πρόταση 4.3.3

Λήμμα 4.3.1: Αν $\eta = \left\| V_U - V_L \right\| = \sup_\pi \left| V_U(\pi) - V_L(\pi) \right|$, τότε για k=1,2,3...

$$\left\| HV^k - V^k \right\| \le \beta^k.\eta$$

όπου V^k είναι οποιοδήποτε από τα φράγματα k τάξεως V_L^k, V_U^k.

Απόδειξη

Επειδή

$$V_L^K \le V^* \le V_U^k \quad \text{και}$$

$$HV_L^K \le HV^* = V^* \le HV_U^k$$

συνάγεται ότι για κάθε $\pi \in \Pi$,

$$\left|HV^k(\pi) - V^k(\pi)\right| \le \max\{\left|HV^k(\pi) - V^*(\pi)\right|, \left|V^k(\pi) - V^*(\pi)\right|\} \le$$

$$\le \max\{\left\|HV^k - V^*\right\|, \left\|V^k - V^*\right\|\}$$

από την οποία προκύπτει ότι:

$$\left\|HV^k - V^k\right\| \le \max\{\left\|HV^k - V^*\right\|, \left\|V^k - V^*\right\|\} \qquad \underline{4.3.5}$$

Από τις σχέσεις (4.3.1),(4.3.5) και την ακόλουθη σχέση

$$\left\|HV^k - V*\right\| = \left\|HV^k - HV^*\right\| \le \beta.\left\|V^k - V^*\right\| \le \beta^{k+1}.\eta$$

συνάγεται ότι

$$\left\|HV^k - V^k\right\| \le \max\{\beta^{k+1}.\eta, \beta^k.\eta\} = \beta^k.\eta \qquad \square$$

Πρόταση 4.3.3: Αν $\eta = \left\|V_U - V_L\right\| = \sup_{\pi}\left|V_U(\pi) - V_L(\pi)\right|$, $\tilde{V}^k$ είναι οποιαδήποτε από τις προσεγγίσεις $\tilde{V}_L^k, \tilde{V}_U^k$ των φραγμάτων k-τάξεως V_L^k, V_U^k και $\tilde{\delta}^k$ η συνάρτηση ελέγχου πού ορίζεται από τη σχέση (4.3.4) τότε η στάσιμη πολιτική $(\tilde{\delta}^k)^\infty$ είναι f(ε,κ)-άριστη, όπου

$$f(\varepsilon,\kappa) = \frac{3-\beta-2\beta^k}{(1-\beta)^2}\cdot\varepsilon + \frac{2-\beta}{1-\beta}\cdot\beta^k\cdot\eta \qquad \underline{4.3.6}$$

δηλαδή:

$$\left\|V(./\tilde{\delta}^k) - V^*\right\| \le f(\varepsilon,k)$$

Απόδειξη

Από την τριγωνική ιδιότητα της νόρμας supremum παίρνουμε

$$\left\|V(./\tilde{\delta}^k) - V^*\right\| \le \left\|V(./\tilde{\delta}^k) - \tilde{V}^k\right\| + \left\|\tilde{V}^k - V^*\right\| \qquad \underline{4.3.7}$$

Εφαρμόζοντας πάλι την τριγωνική ιδιότητα έχουμε:

$$\left\|V(./\tilde{\delta}^k) - \tilde{V}^k\right\| \le \left\|V(./\tilde{\delta}^k) - H_{\tilde{\delta}^k}\tilde{V}^k\right\| + \left\|\mathrm{H}_{\tilde{\delta}^k}\tilde{V}^k - \tilde{V}^k\right\| \qquad \underline{4.3.8}$$

Λαμβάνοντας υπόψη ότι η συνάρτηση τιμών για την πολιτική $(\tilde{\delta}^k)^\infty$,$V(./\tilde{\delta}^k)$,είναι το σταθερό σημείο του τελεστή $H_{\tilde{\delta}^k}$ και ότι ο τελεστής είναι συστολή modulus β, έχουμε:

$$\left\|V(./\tilde{\delta}^k)-H_{\tilde{\delta}^\kappa}\tilde{V}^k\right\|=\left\|\mathrm{H}_{\tilde{\delta}^\kappa}V(./\tilde{\delta}^k)-H_{\tilde{\delta}^k}\tilde{V}^k\right\|\le\beta\cdot\left\|V(\cdot/\tilde{\delta}^k)-\tilde{V}^k\right\| \quad \underline{4.3.9}$$

Από τις σχέσεις (4.3.8),(4.3.9) παίρνουμε

$$\left\|V(./\tilde{\delta}^k)-\tilde{V}^k\right\|\le\left\|H_{\tilde{\delta}^k}\tilde{V}^k-\tilde{V}^k\right\|\Big/(1-\beta) \quad \underline{4.3.10}$$

Θεωρούμε $\mathrm{V}^k=V_L^k$ αν $\tilde{V}^k=\tilde{V}_L^k$

$\mathrm{V}^k=V_U^k$ άν $\tilde{V}^k=\tilde{V}_U^k$

Από την τριγωνική ιδιότητα, τις σχέσεις (4.3.3),(4.3.4) και το λήμμα 4.3.1 παίρνουμε:

$$\left\|H_{\tilde{\delta}^\kappa}\tilde{V}^k-\tilde{V}^k\right\|=\left\|\tilde{\mathrm{H}}\tilde{V}^k-\tilde{V}^k\right\|\le\left\|\tilde{H}\tilde{V}^k-H\tilde{V}^k\right\|+\left\|HV^k-V^k\right\|+\left\|V^k-\tilde{V}^k\right\|\le$$

$$\le\sigma_{k+1}+\beta^k.\eta+\sigma_k \quad \underline{4.3.11}$$

Από τις σχέσεις (4.3.10), (4.3.11) παίρνουμε

$$\left\|V(./\tilde{\delta}^k)-\tilde{V}^k\right\|\le\frac{\sigma_{k+1}+\sigma_\kappa+\beta^k.\eta}{1-\beta} \quad \underline{4.3.12}$$

Από την πρόταση 4.3.1 έχουμε:

$$\left\|\tilde{V}^k-V^*\right\|\le\sigma_k+\beta^k.\eta \quad \underline{4.3.13}$$

Από τις σχέσεις (4.3.7),(4.3.13) και λαμβάνοντας υπόψη τη σχέση (3.4.7) συνάγεται ότι:

$$\left\|V(./\tilde{\delta})-V^*\right\|\le\frac{\sigma_{k+1}+\sigma_\kappa+\beta^k.\eta}{1-\beta}+\sigma_k+\beta^k.\eta=\frac{(2-\beta).\sigma_k+\sigma_{\kappa+1}}{1-\beta}+\frac{2-\beta}{1-\beta}.\beta^k.\eta$$

$$=\frac{3-\beta-2\beta^k}{(1-\beta)^2}.\varepsilon+\frac{2-\beta}{1-\beta}\cdot\beta^k\cdot\eta=f(\varepsilon,k)$$

Επομένως η στάσιμη πολιτική $(\tilde{\delta}^k)^\infty$ είναι $f(\varepsilon,k)-\acute{\alpha}\rho\iota\sigma\tau\eta$. □

Αν επιθυμούμε η πολιτική $(\tilde{\delta}^k)^\infty$ να είναι λ-άριστη για δοσμένο $\lambda > 0$, τότε μπορούμε να επιλέξουμε το προκαθορισμένο σφάλμα ε και την τάξη k έτσι ώστε

$$f(\varepsilon, k) \le \lambda$$

Μπορούμε π.χ να επιλέξουμε τα ε,k έτσι ώστε:

$$\frac{3-\beta}{(1-\beta)^2}.\varepsilon \le \frac{\lambda}{2} \quad , \quad \frac{2-\beta}{1-\beta}\cdot \beta^k \cdot \eta \le \frac{\lambda}{2},$$

δηλαδή

$$\varepsilon \le \frac{\lambda'(1-\beta)^2}{2(3-\beta)} \quad , \quad k \ge \frac{\ln(\frac{\lambda(1-\beta)}{2(2-\beta).\eta})}{\ln \beta}. \qquad \square$$

Παρατηρήσεις

1) Σε όλες τις περιπτώσεις, είτε χρησιμοποιούμε τον ακριβή είτε τον προσεγγιστικό αλγόριθμο των ακρότατων σημείων, ο ελάχιστος αριθμός βημάτων (επαναλήψεων) k πού απαιτείται ώστε να επιτύχουμε οποιαδήποτε επιθυμητή ακρίβεια στην προσέγγιση της V^* ή της άριστης πολιτικής εξαρτάται από την απόσταση η των αρχικών άνω και κάτω φραγμάτων. Επομένως η διαδικασία προσέγγισης μπορεί να επιταχυνθεί σημαντικά αν το η είναι μικρό, δεδομένου ότι το η αποτελεί μέτρο του σφάλματος της αρχικής προσέγγισης της V^* (πρόταση 4.3.1(i)).

2) Επειδή έχουμε τη δυνατότητα ως αρχική προσέγγιση της V^* να επιλέξουμε ένα από τα δύο αρχικά φράγματα (είτε το κάτω είτε το άνω φράγμα),προφανώς είναι λογικό να επιλέξουμε το απλούστερο από αυτά. Σημαντικό πλεονέκτημα ως αρχική προσέγγιση της V^* παρουσιάζουν τα φράγματα πού υπολογίζονται μέσω της V^*_{MDP} (βλέπε παράγραφο 4.2) λόγω της απλότητάς τους.

ΣΥΜΠΕΡΑΣΜΑΤΑ

Προσεγγίσεις της άριστης συνάρτησης τιμών και της άριστης πολιτικής σε άπειρο χρονικό ορίζοντα επιτυγχάνονται με οποιαδήποτε επιθυμητή ακρίβεια μέσω διαδοχικών επαναλήψεων του αλγόριθμου των ακροτάτων σημείων που περιγράψαμε στο κεφάλαιο 3.Καλές προσεγγίσεις επιτυγχάνονται επίσης αν το "κατάλοιπο Bellman" σε κάποιο βήμα (επανάληψη) είναι αρκούντως μικρό.

Η μέθοδος των φραγμάτων αναφέρεται στην επιλογή άνω και κάτω φραγμάτων ως αρχικών προσεγγίσεων της άριστης συνάρτησης τιμών η οποία συνοδεύεται από την επαναληπτική εφαρμογή του αλγόριθμου των ακροτάτων σημείων για την δημιουργία νέων προσεγγίσεων.Από τις προβαλλόμενες συναρτήσεις ελέγχου των νέων προσεγγίσεων (lookahead controllers) κατασκευάζονται σχεδόν άριστες πολιτικές. Επιτάχυνση της διαδικασίας είναι δυνατή αν η απόσταση των αρχικών φραγμάτων είναι μικρή.

Σε κάθε περίπτωση υπολογίζεται ο απαιτούμενος αριθμός επαναλήψεων καθώς και το προκαθορισμένο σφάλμα του αλγορίθμου των ακροτάτων σημείων, έτσι ώστε να επιτυγχάνεται προσέγγιση με οποιαδήποτε επιθυμητή ακρίβεια.

ΚΕΦΑΛΑΙΟ 5

Επαναληπτική μέθοδος πολιτικής για προβλήματα POMDP σε άπειρο χρονικό ορίζοντα

Περίληψη

Στο προηγούμενο κεφάλαιο μελετήσαμε προσεγγίσεις της άριστης συνάρτησης τιμών και της άριστης πολιτικής για ένα πρόβλημα POMDP με την επαναληπτική μέθοδο τιμών (value iteration). Στο κεφάλαιο αυτό, παρουσιάζεται η επαναληπτική μέθοδος πολιτικής (policy-iteration),με την οποία σε κάθε επανάληψη επέρχεται βελτίωση της πολιτικής.

Στην ενότητα 5.1 παρουσιάζουμε την θεωρητική βάση της μεθόδου policy-iteration. Έχοντας υπόψη ότι για το πρόβλημα POMDP σε άπειρο χρονικά ορίζοντα υπάρχει μη τυχαιοποιημένη στάσιμη (nonrandomized stationary) άριστη πολιτική, θεωρούμε μη τυχαιοποιημένες στάσιμες πολιτικές. Μία εγγενής δυσκολία στην υπολογιστική διαδικασία είναι ότι ο χώρος των δ.π ,Π, είναι συνεχής. Κατά το στάδιο policy evaluation στο οποίο υπολογίζουμε την συνάρτηση τιμών μιας πολιτικής, το πρόβλημα απλουστεύεται σημαντικά αν η πολιτική επάγει μία Μαρκοβιανή διαμέριση του χώρου Π. Στην περίπτωση αυτή η συνάρτηση τιμών είναι κατά τμήματα

γραμμική και ο υπολογισμός της ανάγεται στην επίλυση ενός γραμμικού συστήματος εξισώσεων. Στις επόμενες ενότητες του κεφαλαίου αυτού παρουσιάζουμε ικανές συνθήκες , που εξασφαλίζουν ότι μια πολιτική επάγει Μαρκοβιανή διαμέριση του χώρου Π.

Στην ενότητα 5.2, παρουσιάζουμε μια ικανή συνθήκη που προτάθηκε από τον Sondik [120]:η πολιτική να είναι πεπερασμένα μεταβατική (finitely transient).

Στην ενότητα 5.3,προτείνουμε μια διαφορετική ικανή συνθήκη:η πολιτική να είναι περιοδική. Επιπλέον το βασικό αποτέλεσμα αυτής της ενότητας είναι η διατύπωση μιας γενικότερης συνθήκης από την πεπερασμένη μεταβατικότητα, και περιοδικότητα ,που όπως αποδεικνύουμε επάγει Μαρκοβιανή διαμέριση.

5.1. Εισαγωγή

Μία άλλη μέθοδος για επίλυση του προβλήματος μιας μερικά παρατηρήσιμης Μαρκοβιανής διαδικασίας POMDP σε άπειρο χρονικό ορίζοντα (εκτός από την μέθοδο των επαναληπτικών τιμών), είναι η μέθοδος επαναληπτικής βελτίωσης μιας πολιτικής *(policy-iteration)*.Η μέθοδος αυτή παρουσιάσθηκε για την περίπτωση μιας Μαρκοβιανής διαδικασίας αποφάσεων MDP με πεπερασμένο πλήθος καταστάσεων και αποφάσεων (βλέπε παράγραφο 1.1). Στην Μαρκοβιανή διαδικασία αποφάσεων MDP, η παραπάνω επαναληπτική μέθοδος στα πλαίσια του βήματος βελτίωσης της πολιτικής *(policy-improvement)*,βελτιώνει την πολιτική σε κάθε επανάληψη και τερματίζεται μετά από πεπερασμένο αριθμό επαναλήψεων με μια βέλτιστη πολιτική. Επέκταση αυτής της επαναληπτικής μεθόδου έχει γίνει από τον Blackwell [15],αλλά μονάχα από θεωρητική σκοπιά, ο βασικός ωστόσο αλγόριθμος οφείλεται στον Sondik [120]. Αυτός χρησιμοποίησε μια ανάλογη προσέγγιση για την εύρεση

μιας βέλτιστης πολιτικής σε μια μερικά παρατηρήσιμη Μαρκοβιανή διαδικασία αποφάσεων POMDP σε άπειρο χρονικό ορίζοντα.

Η επαναληπτική μέθοδος πολιτικής *(policy iteration)* περιλαμβάνει δύο στάδια Στο πρώτο στάδιο *(policy evaluation)* υπολογίζεται η συνάρτηση τιμών μίας στάσιμης πολιτικής δ^{∞}.Στο δεύτερο στάδιο *(policy-improvement)* αυτή η συνάρτηση τιμών χρησιμοποιείται ως εφαλτήριο για να βρούμε μια πολιτική με μεγαλύτερη συνάρτηση τιμών (βελτιωμένη πολιτική). Η επαναληπτική διαδικασία συνεχίζεται, μέχρις ότου μια διαφορετική πολιτική με μεγαλύτερη συνάρτηση τιμών να μην μπορεί πλέον να βρεθεί. *Η τελευταία αυτή πολιτική είναι τώρα βέλτιστη.*

Θα εξετάσουμε πρώτα το στάδιο *(policy evaluation)*.Έστω δ^{∞} μία (μη τυχαιοποιημένη) στάσιμη πολιτική και $V(\pi/\delta)$, $\pi \in \Pi$ η αντίστοιχη συνάρτηση του αναμενόμενου ολικού εκπίπτοντος κέρδους για άπειρο χρονικό ορίζοντα (συνάρτηση τιμών της δ^{∞}). Επειδή ο τελεστής H_{δ} (βλέπε κεφάλαιο 1) είναι συστολή modulus β (όπου β είναι ο συντελεστής έκπτωσης, $0 \le \beta < 1$), με μοναδικό σταθερό σημείο τη συνάρτηση $V(./\delta)$, η προσέγγιση της $V(./\delta)$ μπορεί να γίνει με επαναληπτική εφαρμογή του τελεστή H_{δ}:

$$V_n(\pi/\delta) = H_{\delta} V_{n-1}(\pi/\delta), \pi \in \Pi, \ n \ge 1,$$

όπου $V_0(\pi/\delta), \pi \in \Pi$ είναι τυχούσα φραγμένη συνάρτηση. Η ακολουθία $\{V_n(./\delta)\}$ συγκλίνει όταν $n \to \infty$ στην συνάρτηση $V(./\delta)$ ανεξάρτητα από την επιλογή της αρχικής συνάρτησης $V_0(./\delta)$ και η σύγκλιση είναι ομαλή στο χώρο Π. Όμως η παραπάνω επαναληπτική μέθοδος δύσκολα εφαρμόζεται επειδή δεν υπάρχουν αλγόριθμοι του ενός βήματος για τον τελεστή H_{δ} ανάλογοι με εκείνους για τον

τελεστή H που περιγράψαμε στα κεφάλαια 2 και 3. Αυτό οφείλεται στο γεγονός ότι ο τελεστής H_δ δεν διατηρεί την κατά τμήματα γραμμικότητα και κυρτότητα. Η δυσκολία αυτή παρακάμπτεται με μία διαδικασία που προτάθηκε από τον Sondik [120]. Η $V(./\delta)$προσεγγίζεται μέσω κατά τμήματα γραμμικών συναρτήσεων με οποιαδήποτε επιθυμητή ακρίβεια και βασίζεται σε θεμελιώδεις δομικές ιδιότητες της POMDP μάλλον παρά σε ιδιότητες σύγκλισης που σχετίζονται με τη συστολή του τελεστή H_δ.Στην περίπτωση που η πολιτική δ^∞ επάγει Μαρκοβιανή διαμέριση του χώρου Π τότε η $V(./\delta)$ υπολογίζεται επακριβώς μέσω της επίλυσης ενός γραμμικού συστήματος εξισώσεων (ενότητα 5.2). Επιπλέον στην ενότητα 5.3 γενικεύουμε την συνθήκη του Sondik που εξασφαλίζει Μαρκοβιανή διαμέριση.Το στάδιο ***policy-improvement*** στηρίζεται στο ακόλουθο θεώρημα

Θεώρημα 5.1.1: (*Howard-Blackwell* ***policy improvement***) [16].

Έστω $V(./\delta)$ η συνάρτηση τιμών για μια πολιτική δ^∞ και V^* η άριστη συνάρτηση τιμών. Αν δ^1 είναι η συνάρτηση ελέγχου, που για κάθε $\pi \in \Pi$ επιλέγει απόφαση που μεγιστοποιεί την

$$H_a[V(./\delta)] = \pi.q^a + \beta. \sum_\theta \{\theta / \pi, a\}.V[T(\pi,\theta,a)/\delta], a \in A$$

δηλαδή :

$$\delta^1(\pi) = \arg \max_a H_a[V(./\delta)], \pi \in \Pi.$$

Τότε :

$$V(\pi/\delta^1) \geq V(\pi/\delta), \forall \pi \in \Pi.$$

Επιπλέον, αν $V(./\delta) \neq V^*$, τότε υπάρχει κάποιο $\delta.\pi$,π, ώστε:

$$V(\pi/\delta^1) > V(\pi/\delta).$$ □

Επισημαίνουμε ότι το παραπάνω θεώρημα δεν εγγυάται την εύρεση άριστης πολιτικής

σε πεπερασμένο αριθμό επαναλήψεων, που οφείλεται στο γεγονός ότι ο χώρος Π των δ.π είναι συνεχής,*οπότε* το σύνολο των (μη τυχαιοποιημένων) στάσιμων πολιτικών *είναι μη αριθμήσιμο*.Αφού λοιπόν, δεν υπάρχει εγγύηση σύγκλισης σε πεπερασμένο αριθμό επαναλήψεων,μπορούμε να περιοριστούμε στην αναζήτηση μιας $\varepsilon-$*άριστης* πολιτικής ($\varepsilon > 0$).Το κριτήριο τερματισμού της *policy iteration* στηρίζεται τότε στο κατάλοιπο Bellman *μιας πολιτικής* δ^∞ *(Bellman –residual), που ορίζεται ως η μέγιστη απόλυτη διαφορά των συναρτήσεων* $V(./\delta)$ *και* $HV(./\delta)$, *δηλαδή η ποσότητα:*

$$\|HV(./\delta)-V(./\delta)\|$$

Αν το κατάλοιπο Bellman *για την πολιτική* δ^∞ είναι αρκούντως μικρό:

$$\|HV(./\delta)-V(./\delta)\| \le \varepsilon.(1-\beta),$$

τότε η πολιτική δ^∞ *είναι* $\varepsilon-$*άριστη* και η υπολογιστική διαδικασία τερματίζεται. Πράγματι από το λήμμα 4.1.1 (iii) παίρνουμε:

$$\|V^* - V(./\delta)\| \le \frac{1}{1-\beta}\|HV(./\delta)-V(./\delta)\| \le \varepsilon .$$

Επισημαίνουμε μία δυσκολία που εμφανίζεται στο στάδιο ***policy improvement*** και σχετίζεται με την εφαρμογή του τελεστού H στην συνάρτηση $V(./\delta)$, που γενικά δεν είναι κατά τμήματα γραμμική και κυρτή. Οι αλγόριθμοι όμως του ενός βήματος των κεφαλαίων 2 και 3 εφαρμόζονται μόνο σε συναρτήσεις που είναι κατά τμήματα γραμμικές και κυρτές (p.w.l.c.). Η δυσκολία αυτή αντιμετωπίσθηκε από τον Sondik [120] με την αντικατάσταση της $V(./\delta)$ ή μιας κατά τμήματα γραμμικής προσέγγισής της από την κυρτή θήκη της (convex hull).

Στο σημείο αυτό θέλουμε να επισημάνουμε, ότι αρκετοί συγγραφείς όπως οι Kakalic και Eckles [56] παρακάμπτουν τις παραπάνω δυσκολίες προσεγγίζοντας τον

χώρο Π με ένα σύνολο σημείων (grid of points).Ωστόσο οι υπολογιστικές δυσκολίες μιας τέτοιας προσέγγισης είναι τεράστιες.

5.2.Μαρκοβιανή διαμέριση και πεπερασμένα μεταβατικές πολιτικές

Από το γεγονός ότι η συνάρτηση τιμών μιας πολιτικής δ^∞, V(./δ),είναι το μοναδικό σταθερό σημείο του τελεστού H_δ, αποδεικνύεται εύκολα το ακόλουθο λήμμα

Λήμμα 5.2.1: Η V(π/δ) μπορεί να γραφεί σαν

$$V(\pi/\delta) = \pi.\gamma(\pi/\delta),\ \pi \in \Pi,$$

όπου γ(π/δ) είναι ένα διάνυσμα-στήλη $N \times 1$ (όπου N το πλήθος καταστάσεων του συστήματος) και μάλιστα η μοναδική φραγμένη λύση της διανυσματικής εξίσωσης

$$\gamma(\pi/\delta) = q^{\delta(\pi)} + \beta.P^{\delta(\pi)} \sum_{\theta} R_\theta^{\delta(\pi)}.\gamma[T(\pi, \theta, \delta)/\delta],\ \pi \in \Pi. \qquad \underline{5.2.1}$$

(απόδειξη Sondik)[120]. □

Βέβαια το παραπάνω λήμμα, δεν εκφράζει ότι η V(./δ) είναι κατά τμήματα γραμμική συνάρτηση. Στη συνέχεια θα μελετήσουμε συνθήκες κάτω από τις οποίες η συνάρτηση τιμών V(./δ) είναι κατά τμήματα γραμμική συνάρτηση (p.w.l.).

Ορισμός 5.2.1:Έστω δ^∞ μια στάσιμη πολιτική. Μια διαμέριση $\mathbf{V}=[V_1, V_2, \ldots]$ *του χώρου Π, λέγεται* ***Μαρκοβιανή διαμέριση*** *επαγόμενη από την δ^∞ αν ικανοποιούνται οι παρακάτω ιδιότητες:*

α)Σε όλα τα δ.π (information-vectors) που ανήκουν στο ίδιο κελλί αντιστοιχεί η ίδια απόφαση μέσω της δ.

β)Μέσω της συνάρτησης μεταφοράς $T(.,\theta,\delta)$ όλα τα δ.π που ανήκουν σε κάποιο κελλί της διαμέρισης έστω V_j, απεικονίζονται στο ίδιο κελλί $V_{j'}$, όπου ο δείκτης j', εξαρτάται αποκλειστικά από τα j και θ, $j' = \nu(j,\theta)$. Η συνάρτηση $\nu(j,\theta)$ *καλείται Μαρκοβιανή απεικόνιση (Markov mapping) επαγόμενη από την δ^∞.*

Σχετικά με την ιδιότητα (α), η απόφαση που αντιστοιχεί στο κελλί V_j μέσω της δ συμβολίζεται με δ_j :

$$\delta(\pi) \equiv \delta_j \quad \forall \pi \in V_j .$$

Σχετικά με την ιδιότητα (β), η σχέση ανάμεσα στα σύνολα V_j της διαμέρισης **V** μέσω της συνάρτησης μεταφοράς $T(.,\theta,\delta)$ *παρέχεται από την Μαρκοβιανή απεικόνιση* $\nu(j,\theta)$:

$$\pi \in V_j \Rightarrow T(\pi,\theta,\delta) \in V_{\nu(j,\theta)} .$$

Η Μαρκοβιανή απεικόνιση ν *και η Μαρκοβιανή διαμέριση* **V**, *πού επάγεται από την πολιτική* δ^∞, *λέμε ότι είναι* ***«ισοδύναμες μέσω της δ».***

Σημειώνουμε ότι στην περίπτωση που η πολιτική δ^∞ επάγει Μαρκοβιανή διαμέριση $\mathbf{V}=[V_1, V_2, \ldots V_m]$ η συνάρτηση τιμών $V(./\delta)$ είναι κατά τμήματα γραμμική (p.w.l) και υπολογίζεται μέσω επίλυσης ενός γραμμικού συστήματος εξισώσεων, που έχει πάντοτε λύση.

Πράγματι από το λήμμα 5.2.1 έχουμε:

$$V(\pi/\delta) = \pi.\gamma(\pi/\delta),$$

όπου το διάνυσμα-στήλη γ(π/δ) είναι η μοναδική φραγμένη λύση της διανυσματικής εξίσωσης (5.2.1) .

Σε κάθε κελλί της διαμέρισης αντιστοιχούμε σταθερό διάνυσμα γ(./δ) και θέτουμε:

$$\gamma(\pi/\delta) \equiv \gamma_j \quad \forall \pi \in V_j .$$

Επειδή για κάθε $\pi \in V_j$ έχουμε $\delta(\pi)=\delta_j$ και $\gamma(T(\pi,\theta,\delta)/\delta)=\gamma_{v(j,\theta)}$, $\theta \in \Theta$, η εξίσωση (5.2.1) ανάγεται στο ακόλουθο σύστημα διανυσματικών εξισώσεων

$$\gamma_j = q^{\delta_j} + \beta . \sum_{\theta} P^{\delta_j} . R_{\theta}^{\delta_j} . \gamma_{v(j,\theta)} , \; 1 \le j \le m . \qquad \underline{5.2.2}$$

Οι λύσεις για τα γ_j μέσω των σχέσεων (5.2.2) είναι μοναδικές (πρόκειται για vectors-contraction), και ισχύει:

$$V(\pi/\delta) = \pi . \gamma_j \qquad \forall \pi \in V_j . \qquad \underline{5.2.3}$$

□

Τίθεται τώρα το ερώτημα,πότε μια στάσιμη πολιτική δ^∞ ***επάγει μια Μαρκοβιανή διαμέριση.*** Ο Sondik [120] έδωσε ικανή συνθήκη, ώστε μια στάσιμη πολιτική δ^∞ ,να επάγει μια πεπερασμένη Μαρκοβιανή διαμέριση του χώρου Π. *Μια πολιτική, που ικανοποιεί αυτή τη συνθήκη λέμε ότι είναι, πεπερασμένα μεταβατική, σύντομα f.t (finitely transient).*Χονδρικά ως *πεπερασμένα μεταβατική* πολιτική, ορίζεται μια στάσιμη πολιτική ,δ^∞ ,για την οποία οποιοδήποτε δ.π δεν αποτελεί σημείο ασυνέχειας της συνάρτηση ελέγχου δ μετά από πεπερασμένο αριθμό χρονικών περιόδων. Έτσι τα δ.π στα οποία η συνάρτηση ελέγχου είναι ασυνεχής είναι ουσιαστικά «μεταβατικά» (*transient).*Για να ορίσουμε την έννοια αυτή αναλυτικότερα εισάγουμε τους ακόλουθους συμβολισμούς.

Αν δ^∞ είναι μια στάσιμη πολιτική και $B \subseteq \Pi$,ορίζουμε:

$$T_\delta(B) := \text{κλειστή θήκη}\{T(\pi, \theta, \delta): \pi \in B, \theta \in \Theta\}.$$

Με άλλα λόγια το $T_\delta(B)$ είναι το ελάχιστο κλειστό σύνολο που περιέχει τις δυνατές μεταβάσεις των δ.π. του συνόλου B την επόμενη χρονική περίοδο εφαρμόζοντας την πολιτική δ^∞.

Ορίζουμε επίσης : $S_\delta^0 := \Pi$

και

$$S_\delta^n := T_\delta(S_\delta^{n-1}) \ , n \geq 1,$$

Το S_δ^n περιέχει όλα τα δυνατά δ.π, στην $n^{στη}$ χρονική περίοδο μετά την έναρξη της λειτουργίας του συστήματος εφαρμόζοντας την πολιτική δ^∞.

Αποδεικνύεται επαγωγικά ότι :

$$S_\delta^n \subset S_\delta^{n-1} \ , n \geq 1. \qquad \underline{5.2.4}$$

Θεωρούμε

$$D_\delta := \text{κλειστή θήκη } \{\pi: \delta(\pi) \text{ είναι ασυνεχής στο } \pi\}.$$

Δηλαδή ,D_δ, είναι το ελάχιστο κλειστό σύνολο που περιέχει τις ασυνέχειες της συνάρτησης ελέγχου δ.

Ορίζουμε την ακολουθία των συνόλων D^n , $n=0,1,2...$ *ως εξής:*

$$D^0 := D_\delta$$

$$D^{n+1} := \{\pi: T(\pi,\theta,\delta) \in D^n \text{ για κάποιο } \theta\} \ n \geq 0.$$

Το σύνολο D^n *αναφέρεται σαν σύνολο ασυνεχειών n-τάξης της* δ^∞.

Ορισμός5.2.2:*Μια στάσιμη πολιτική* δ^∞ *είναι πεπερασμένα μεταβατική(finitely-transient) ,σύντομα f.t. αν υπάρχει ένας ακέραιος n , ώστε :*

$$D_\delta \cap S_\delta^n = \emptyset. \qquad \underline{5.2.5}$$

Ο μικρότερος ακέραιος, που πληροί την (5.2.5) σημειώνεται με n_δ και ονομάζεται δείκτης (index) της παραπάνω πολιτικής.

Αν η στάσιμη πολιτική, δ^∞,είναι πεπερασμένα μεταβατική με index n_δ τότε από την (5.2.4) προκύπτει άμεσα ότι:

$$D_\delta \cap S_\delta^n = \emptyset \quad \forall\ n \geq n_\delta.$$

Λήμμα 5.2.2: *Η πολιτική δ^∞ είναι πεπερασμένα μεταβατική f.t. με δείκτη ,n_δ, αν και μόνον αν D^{n_δ} είναι το πρώτο κενό σύνολο στην ακολουθία $D^0, D^1, D^2, \ldots$,*

Απόδειξη Sondik [120].

Από τα παραπάνω συνάγεται ότι:αν μια πολιτική δ^∞ είναι *f.t.* με δείκτη $n_\delta \geq 1$ τότε : $D^n \neq \emptyset$, $n < n_\delta$ και $D^n = \emptyset$, $n \geq n_\delta$.

Θεώρημα5.2.1:*Αν δ^∞ είναι f.t.,τότε αυτή επάγει πεπερασμένη Μαρκοβιανή διαμέριση και η συνάρτηση,$V(\pi/\delta)$, είναι κατά τμήματα γραμμική συνάρτηση (p.w.l).*

(απόδειξη Sondik) [120].

□

Παρατηρήσεις

1)Αν η συνάρτηση ελέγχου δ είναι σταθερή τότε προφανώς η δ δεν έχει σημεία ασυνέχειας και $D^\delta = \emptyset$. Η πολιτική δ^∞ είναι τετριμμένα *f.t.* με δείκτη $n_\delta = 0$. Η Μαρκοβιανή διαμέριση που επάγεται από την δ^∞ είναι τετριμμένη:αποτελείται από ένα μόνο σύνολο, τον χώρο Π.

2)Τα σύνορα ανάμεσα στα σύνολα (κελλιά) της Μαρκοβιανής διαμέρισης του χώρου Π που επάγεται από μία f.t . πολιτική δ^{∞} με δείκτη $n_{\delta} \geq 1$ σχηματίζονται από το σύνολο $\bigcup_{k=0}^{n_{\delta}-1} D^{k}$.

3)Το αντίστροφο του παραπάνω θεωρήματος 5.2.1, δεν αληθεύει, όπως αποδεικνύεται με αντιπαράδειγμα Sondik)[117].Επομένως αν έχουμε μια συνάρτηση τιμών,V(π/δ), που είναι κατά τμήματα γραμμική, δεν σημαίνει ότι η πολιτική που την επάγει είναι πεπερασμένα μεταβατική.

Εφαρμογή 5.2.1: (πηγή Sondik [117]). Θεωρούμε ένα πρόβλημα POMDP με δύο καταστάσεις, δύο μηνύματα και δύο αποφάσεις . S={1,2}, Θ={1,2}, A={1,2} με πίνακες μετάβασης και μηνυμάτων και διανύσματα άμεσου κέρδους:

	P^{α}	R^{α}	q^{α}
α=1	$\begin{pmatrix} 0.8 & 0.2 \\ 0.5 & 0.5 \end{pmatrix}$	$\begin{pmatrix} 0.8 & 0.2 \\ 0.6 & 0.4 \end{pmatrix}$	$\begin{pmatrix} 4 \\ -4 \end{pmatrix}$
α=2	$\begin{pmatrix} 0.2 & 0.8 \\ 0.8 & 0.2 \end{pmatrix}$	$\begin{pmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{pmatrix}$	$\begin{pmatrix} 0 \\ -3 \end{pmatrix}$

$$\delta(\pi) = \begin{cases} 1 & 0 \leq \pi_1 \leq 0.7 \\ 2 & 0.7 < \pi_1 \leq 1 \end{cases} \qquad \beta=0.9$$

Πηγή Sondik [117]

Τα δ.π $\pi=(\pi_1,\pi_2=1-\pi_1)$ καθώς και η συνάρτηση μεταφοράς $T(\pi,\theta,\alpha)=(T_1(\pi,\theta,\alpha),1-T_1(\pi,\theta,\alpha))$ θα εξετάζονται ως προς την πρώτη συνιστώσα. Από τον τύπο Bayes (1.4.4) παίρνουμε:

i) $T_1(\pi,\theta=1,\delta(\pi)=1)=\dfrac{0.4+0.24.\pi_1}{0.7+0.06.\pi_1},\quad 0\le\pi_1\le 0.7,$

γνήσια αύξουσα, με πεδίο τιμών $W_1=[0.5714,0.7655]$.

ii) $T_1(\pi,\theta=2,\delta(\pi)=1)=\dfrac{0.1+0.06.\pi_1}{0.3-0.06.\pi_1},\quad 0\le\pi_1\le 0.7,$

γνήσια αύξουσα, με πεδίο τιμών $W_2=[0.3333,0.5504]$.

iii) $T_1(\pi,\theta=1,\delta(\pi)=2)=\dfrac{0.56+0.42.\pi_1}{0.64-0.18.\pi_1},\quad 0.7<\pi_1\le 1,$

γνήσια φθίνουσα, με πεδίο τιμών $W_3=[0.3043,0.5175)$.

iv) $T_1(\pi,\theta=2,\delta(\pi)=2)=\dfrac{0.24-0.18.\pi_1}{0.36+0.18.\pi_1},\quad 0.7<\pi_1\le 1,$

γνήσια φθίνουσα, με πεδίο τιμών $W_4=[0.1111,0.2346)$.

Προφανώς $\mathbf{D^0=D_\delta=\{0.7\}}$.

Για να βρούμε το σύνολο ασυνεχειών $1^{\eta s}_{=}$ τάξης $\mathbf{D^1}$ της δ ,εξετάζουμε τις εξισώσεις

$$\mathbf{T_1(\pi,\theta,\delta)=0.7.}$$

Επειδή $0.7\in W_1$ ενώ $0.7\notin W_2, 0.7\notin W_3, 0.7\notin W_4$, μόνο η εξίσωση

$$T_1(\pi,\theta=1,\delta(\pi)=1)=\frac{0.4+0.24.\pi_1}{0.7+0.06.\pi_1}=0.7$$ έχει αποδεκτή λύση π_1=0.4545.

Αρα $\mathbf{D^1=\{0.4545\}}$.

Συνεχίζοντας παρόμοια, βρίσκουμε τα σύνολα ασυνεχειών $2^{\eta s}_{=}, 3^{\eta s}_{=}$ κλπ. τάξης της δ:

$\mathbf{D^2=\{0.4167,0.7957\}}$

$\mathbf{D^3=\{0.2941,0.8502\}}$

$\mathbf{D^4=\varnothing}$.

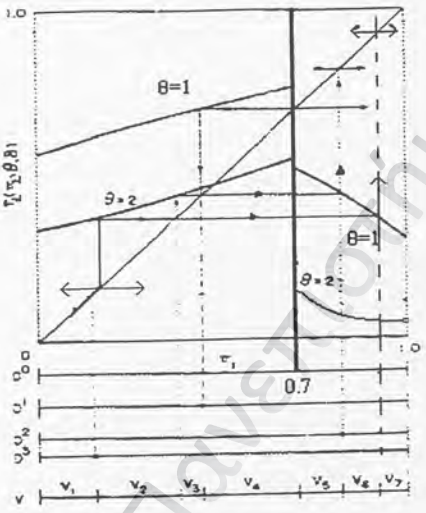
Άρα η στάσιμη πολιτική δ^{∞} είναι πεπερασμένα μεταβατική με δείκτη $n_{\delta} = 4$. Τα σύνολα D^n απεικονίζονται στο σχήμα 5.1. Το σύνολο D^1 βρίσκεται με αντανάκλαση του $D^0=\{0.7\}$,όπως αποδεικνύουν τα βέλη.Παρόμοια με αντανάκλαση του D^1 βρίσκεται το D^2 κ.ο.κ.

Η Μαρκοβιανή διαμέριση του διαστήματος [0,1] που επάγεται από την πολιτική δ^{∞} αποτελείται από 7 σύνολα (κελλιά) με συνοριακά σημεία τα στοιχεία του συνόλου $D^0 \cup D^1 \cup D^2 \cup D^3$ (βλέπε παρατήρηση 2 και σχήμα 5.1):

$V_1=[0,0.2941]$, $V_2=(0.2941,0.4167]$, $V_3=(0.4167,0.4545]$, $V_4=(0.4545,0.7]$,

$V_5=(0.7,0.7957]$, $V_6=[0.7957,0.8502)$, $V_7=[0.8502,1]$.

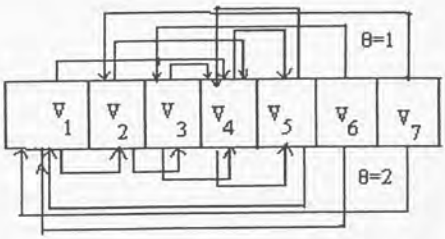
Μαρκοβιανή απεικόνιση



Σύνολο	ν(j,θ)	
V_j	θ=1	θ=2
1	4	2
2	4	3
3	4	4
4	5	4
5	4	1
6	3	1
7	2	1

Σχήμα 5.1: Διάγραμμα μεταφοράς στην εφαρμογή 5.2.1.

(πηγή Sondik [117])



Σχήμα 5.2: Διάγραμμα ροής της πολιτικής δ^{∞}.

Αφού η παραπάνω πολιτική δ^{∞} επάγει πεπερασμένη Μαρκοβιανή διαμέριση $\boldsymbol{V}=[V_1, V_2, \ldots V_7]$ *η συνάρτηση τιμών $V(./\delta)$ είναι κατά τμήματα γραμμική (p.w.l) και υπολογίζεται με επίλυση του παρακάτω πεπερασμένου γραμμικού συστήματος εξισώσεων μέσω του προγράμματος Mathematica 5.*

$$\gamma_1 = q^1 + \beta.P^1.R_1^1.\gamma_4 + \beta.P^1.R_2^1.\gamma_2$$
$$\gamma_2 = q^1 + \beta.P^1.R_1^1.\gamma_4 + \beta.P^1.R_2^1.\gamma_3$$
$$\gamma_3 = q^1 + \beta.P^1.R_1^1.\gamma_4 + \beta.P^1.R_2^1.\gamma_4$$
$$\gamma_4 = q^1 + \beta.P^1.R_1^1.\gamma_5 + \beta.P^1.R_2^1.\gamma_4$$
$$\gamma_5 = q^2 + \beta.P^2.R_1^2.\gamma_4 + \beta.P^2.R_2^2.\gamma_1$$
$$\gamma_6 = q^2 + \beta.P^2.R_1^2.\gamma_3 + \beta.P^2.R_2^2.\gamma_1$$
$$\gamma_7 = q^2 + \beta.P^2.R_1^2.\gamma_2 + \beta.P^2.R_2^2.\gamma_1$$

Βρίσκουμε λοιπόν τα gradients:

$$\gamma_1 = (11.2925, 0.6228)^T, \gamma_2 = (11.2483, 0.5694)^T, \gamma_3 = (11.0372, 0.3783)^T$$

$$\gamma_4 = (9.7884, -0.0588)^T, \gamma_5 = (7.3027, 1.6364)^T, \gamma_6 = (8.0691, 2.2040)^T,$$

$$\gamma_7 = (8.2115, 2.3315)^T.$$

$$V(\pi/\delta) = \pi.\gamma_j \quad , \quad \pi \in V_j \, , \, j=1,2,3,\ldots,7$$

5.3. Περιοδικές πολιτικές

Στην ενότητα αυτή, παρουσιάζουμε τη συνθήκη της περιοδικότητας η οποία πέραν από τη συνθήκη της πεπερασμένης μεταβατικότητας του Sondik εξασφαλίζει επίσης ότι μία στάσιμη πολιτική επάγει Μαρκοβιανή διαμέριση του χώρου Π.Το πιο σημαντικό όμως είναι ότι υπάρχει γενικότερη συνθήκη από τις δύο που αναφέραμε η οποία όπως θα διαπιστώσουμε στη συνέχεια εξασφαλίζει Μαρκοβιανή διαμέριση. Πρώτα όμως θα δώσουμε την έννοια της διαμέρισης τάξης k που επάγεται από μια στάσιμη πολιτική (βλέπε Sondik [120]).

Έστω δ^∞ μια στάσιμη πολιτική . Ορίζουμε

$$\bar{D}^n := \bigcup_{i=0}^{n} D^i, n = 0,1,2,...$$

Το $\bar{D}^n$ δηλώνει το σύνολο των δ.π που οδηγούνται σε ασυνέχειες της συνάρτησης ελέγχου δ μέσω της συνάρτησης Τ (.,.,δ) σε n το πολύ βήματα.

Σημειώνουμε ότι $\bar{D}^0 = D^0 \equiv D_\delta$ και $\bar{D}^n \subseteq \bar{D}^m \ \forall m > n$.

Θεωρούμε την ακολουθία διαμερίσεων του χώρου Π,$\mathbf{V^0}$,$\mathbf{V^1}$,...η οποία κατασκευάζεται ως εξής:Τα σύνορα ανάμεσα στα σύνολα (κελλιά) της αρχικής διαμέρισης $\mathbf{V^0}$ του χώρου Π σχηματίζονται από το σύνολο των ασυνεχειών της συνάρτησης ελέγχου δ, $D^0 = D_\delta$.Έτσι η συνάρτηση ελέγχου δ είναι σταθερή (παίρνει ως τιμή την ίδια απόφαση) σε κάθε κελλί της $\mathbf{V^0}$.Λέμε ότι η διαμέριση $\mathbf{V^0}$ επάγεται από την δ^∞.

Για $k \geq 1$, η διαμέριση $\mathbf{V^k}$ του χώρου Π κατασκευάζεται έτσι ώστε τα σύνορα ανάμεσα στα κελλιά αυτής να σχηματίζονται από το σύνολο

$$\bar{D}^k := \bigcup_{n=0}^{k} D^n.$$

Παρατηρούμε ότι η διαμέριση $\mathbf{V^k}$ αποτελεί λέπτυνση της διαμέρισης $\mathbf{V^{k-1}}$,δηλαδή για κάθε κελλί της $\mathbf{V^k}$ υπάρχει κελλί της $\mathbf{V^{k-1}}$ που το περιέχει. Επομένως συμπεραίνουμε επαγωγικά ότι η συνάρτηση ελέγχου δ είναι σταθερή σε κάθε κελλί της $\mathbf{V^k}$. Η $\mathbf{V^k}$ καλείται διαμέριση τάξης k που επάγεται από την δ^∞.

Σημειώνουμε ότι αν $\bar{D}^k \neq \bar{D}^{k-1}$, η διαμέριση $\mathbf{V^k}$ προκύπτει από την $\mathbf{V^{k-1}}$ μέσω του συνόλου των ασυνεχειών k τάξης της δ, D^k: τα κελλιά της $\mathbf{V^{k-1}}$ που τέμνουν το σύνολο D^k διαμερίζονται σε νέα σύνολα (κελλιά) τα οποία διαχωρίζονται με σύνορα από το D^k, ενώ τα κελλιά της $\mathbf{V^{k-1}}$ που δεν τέμνουν το D^k παραμένουν τα ίδια στη διαμέριση $\mathbf{V^k}$. Αν $\bar{D}^k = \bar{D}^{k-1}$, τότε η διαμέριση $\mathbf{V^k}$ συμπίπτει με την $\mathbf{V^{k-1}}$.

Πρόταση 5.3.1: *Έστω $\mathbf{V^k}$ διαμέριση τάξης k ($k \geq 1$) του χώρου Π που επάγεται από μια στάσιμη πολιτική δ^∞ και $\theta \in \Theta$. Τότε κάθε κελλί της $\mathbf{V^k}$ απεικονίζεται εντός κάποιου κελλιού της διαμέρισης $\mathbf{V^{k-1}}$ μέσω της $T(.,\theta,\delta)$. Με άλλα λόγια, αν V_i^k ,είναι κελλί της $\mathbf{V^k}$,τότε υπάρχει κελλί της $\mathbf{V^{k-1}}$, έστω V_j^{k-1} έτσι ώστε $T(V_i^k,\theta,\delta) \subseteq V_j^{k-1}$.*

Απόδειξη

Συμβολίζουμε με δ^i την απόφαση που παίρνει η δ στο κελλί V_i^k: $\delta^i \equiv \delta(\pi), \pi \in V_i^k$. Θεωρούμε δύο τυχόντα δ.π. π', π'' στο εσωτερικό της περιοχής V_i^k και υποθέτουμε ότι το π' απεικονίζεται στο κελλί V_j^{k-1} *της διαμέρισης* $\mathbf{V^{k-1}}$ μέσω της συνάρτησης

$T(.,\theta,\delta^i)$, δηλαδή $T(\pi',\theta,\delta^i)\in V_j^{k-1}$.Θα δείξουμε ότι το δ.π. π'' απεικονίζεται στο ίδιο κελλί, δηλαδή $T(\pi'',\theta,\delta^i)\in V_j^{k-1}$.Διακρίνουμε τις ακόλουθες περιπτώσεις:

ι) Υποθέτουμε ότι το ευθύγραμμο τμήμα

$$l:\lambda\pi'+(1-\lambda)\pi'' \ ,0\le\lambda\le 1$$

που συνδέει τα π',π'' περιέχεται εξ ολοκλήρου στο εσωτερικό της περιοχής V_i^k.Επειδή η συνάρτηση μεταφοράς T(.,θ,α), όπου $a\in A$,απεικονίζει ευθύγραμμα τμήματα σε ευθύγραμμα τμήματα (βλέπε πρόταση 1.4.1 ,Κεφ 1), συνάγεται ότι το σύνολο

$$T(l,\theta,\delta^i)=\{T(\pi,\theta,\delta^i):\pi\in l\}$$

είναι ευθύγραμμο τμήμα με άκρα $T(\pi',\theta,\delta^i)$ και $T(\pi'',\theta,\delta^i)$.As θεωρήσουμε ότι το π'' απεικονίζεται σε διαφορετικό κελλί της διαμέρισης V^{k-1} από αυτό που απεικονίζεται το π' μέσω της $T(.,\theta,\delta^i)$, δηλαδή $T(\pi'',\theta,\delta^i)\in V_s^{k-1}$,*όπου* $s\neq j$.Τότε το ευθύγραμμο τμήμα $T(l,\theta,\delta^i)$ διασχίζει κάποιο σύνορο ανάμεσα στις περιοχές V_j^{k-1},V_s^{k-1}. Επομένως υπάρχει $\pi^*\in l$, έτσι ώστε $T(\pi^*,\theta,\delta^i)\in\bar{D}^{k-1}$. Τότε $\pi^*\in\bar{D}^k$ το οποίο σημαίνει ότι το π^* είναι συνοριακό σημείο της περιοχής V_i^k. *Αυτό* όμως αντίκειται στην υπόθεσή μας ότι το ευθύγραμμο τμήμα l περιέχεται στο εσωτερικό της περιοχής V_i^k. Συνεπώς τα δ.π. π',π'' καθώς και το ευθύγραμμο τμήμα l που τα ενώνει, απεικονίζονται μέσω της $T(.,\theta,\delta^i)$ στο ίδιο κελλί της διαμέρισης V^{k-1} :

$$T(l,\theta,\delta^i)\subseteq V_j^{k-1},\quad T(\pi'',\theta,\delta^i)\in V_j^{k-1}.$$

ii) Θεωρούμε ότι το ευθύγραμμο τμήμα l που συνδέει τα π',π'' δεν περιέχεται εξ ολοκλήρου στο εσωτερικό της περιοχής V_i^k. Τότε υπάρχει διαδρομή από ευθ. τμήματα

$l_1, l_2, \ldots, l_m$ η οποία συνδέει τα π', π'', έτσι ώστε τα $l_1, l_2, \ldots, l_m$ να περιέχονται εξ ολοκλήρου στο εσωτερικό της περιοχής V_i^k. Συμβολίζουμε με π_t, π_{t+1} τα άκρα του ευθυγράμμου τμήματος l_t (t=1,2,…,m), όπου $\pi_1 \equiv \pi'$, $\pi_{m+1} \equiv \pi''$.

Επειδή $T(\pi_1, \theta, \delta^i) \in V_j^{k-1}$, από την περίπτωση (i) συνάγεται ότι $T(l_1, \theta, \delta^i) \subseteq V_j^{k-1}$, $T(\pi_2, \theta, \delta^i) \subseteq V_j^{k-1}$. Εφαρμόζοντας διαδοχικά το ίδιο επιχείρημα για τα ευθύγραμμα τμήματα $l_2, \ldots, l_m$, παίρνουμε:

$$T(l_t, \theta, \delta^i) \subseteq V_j^{k-1},\ T(\pi_{t+1}, \theta, \delta^i) \subseteq V_j^{k-1}, t=2, \ldots, m$$

Άρα $T(\pi'', \theta, \delta^i) \in V_j^{k-1}$ και συμπεραίνουμε ότι το κελλί V_i^k *της διαμέρισης* $\mathbf{V^k}$ απεικονίζεται μέσω της $T(., \theta, \delta^i)$ εντός του κελλιού V_j^{k-1} *της διαμέρισης* $\mathbf{V^{k-1}}$. □

Πρόταση 5.3.2:

Οι ακόλουθες προτάσεις είναι ισοδύναμες

i) $\bar{D}^{n+1} = \bar{D}^n$ για κάποιο $n \in \mathbb{N}_0$.

ii) Υπάρχει $n \in \mathbb{N}_0$ έτσι ώστε $\bar{D}^m = \bar{D}^n \quad \forall m \geq n$.

Απόδειξη

$i \Rightarrow ii$). Ας υποθέσουμε ότι το *ii)* δεν αληθεύει. Θα καταλήξουμε σε άτοπο. Έστω $\bar{D}^m$ το πρώτο σύνολο στην ακολουθία $\bar{D}^{n+2}, \bar{D}^{n+3}, \ldots$, το οποίο διαφέρει από το $\bar{D}^n$, δηλαδή $\bar{D}^{m-1} = \bar{D}^n$ και $\bar{D}^n \subsetneq \bar{D}^m$.

Επομένως υπάρχει $\pi \in \bar{D}^m$ έτσι ώστε $\pi \notin \bar{D}^n$. Τότε υπάρχει $\theta \in \Theta$ έτσι ώστε $T(\pi, \theta, \delta(\pi)) \in \bar{D}^{m-1}$. Συνεπώς $T(\pi, \theta, \delta(\pi)) \in \bar{D}^n$ και $\pi \in \bar{D}^{n+1}$. Αυτό όμως οδηγεί σε αντίφαση επειδή υποθέσαμε $\bar{D}^{n+1} = \bar{D}^n$ και $\pi \notin \bar{D}^n$. Άρα το (ii) αληθεύει.

ii) $\Rightarrow$ *i)*. Είναι προφανές. □

Θα δώσουμε τώρα τον ορισμό της περιοδικής πολιτικής.

Ορισμός 5.3.1: Μία στάσιμη πολιτική δ^∞ λέγεται περιοδική αν η ακολουθία των συνόλων $D^0, D^1, D^2 \ldots$ δεν περιέχει κενό σύνολο και υπάρχουν φυσικοί αριθμοί $l \geq 0, s \geq 1$ έτσι ώστε :

$$D^{n+s} = D^n \quad \forall n \geq l.$$

Από τον παραπάνω ορισμό και το λήμμα 5.2.2 συνάγεται ότι μία περιοδική πολιτική αποκλείεται να είναι πεπερασμένα μεταβατική (f.t.) και αντιστρόφως. Με άλλα λόγια οι συνθήκες της περιοδικότητας και της πεπερασμένης μεταβατικότητας είναι αμοιβαία αποκλειόμενες.

Μία γενίκευση των συνθηκών περιοδικότητας και πεπερασμένης μεταβατικότητας είναι η ακόλουθη συνθήκη **(A)**: *Λέμε ότι μία στάσιμη πολιτική* δ^∞ *ικανοποιεί τη συνθήκη (A) αν:*

$$\bar{D}^{n+1} = \bar{D}^n \quad \textit{για κάποιο } n \in \mathbb{N}_0. \qquad \textbf{(A)}$$

Η γενίκευση αυτή προκύπτει από την ακόλουθη πρόταση.

Πρόταση 5.3.3: Αν η πολιτική δ^∞ είναι πεπερασμένα μεταβατική ή περιοδική τότε η δ^∞ ικανοποιεί τη συνθήκη **(A)**.

Απόδειξη

Αν η πολιτική δ^∞ είναι πεπερασμένα μεταβατική με δείκτη n_δ, τότε το συμπέρασμα έπεται προφανώς από το λήμμα 5.2.2 με $n = n_\delta$. Αν η πολιτική δ^∞ είναι περιοδική, τότε το συμπέρασμα έπεται από τον ορισμό 5.3.1 με $n = l + s - 1$.

Πράγματι,

$$\overline{D}^{n+1} = \overline{D}^{l+s} = \bigcup_{i=0}^{l+s} D^i = (\bigcup_{i=0}^{l+s-1} D^i) \cup D^{l+s}$$

$$= (\bigcup_{i=0}^{l+s-1} D^i) \cup D^l = \bigcup_{i=0}^{l+s-1} D^i = \overline{D}^{l+s-1} = \overline{D}^n .$$

□

Ορισμός 5.3.2: Έστω δ^∞ μία στάσιμη πολιτική *που ικανοποιεί τη συνθήκη (A).* Καλούμε δείκτη της δ^∞ τον ελάχιστο $n \in \mathbb{N}_0$ για τον οποίο $\overline{D}^{n+1} = \overline{D}^n$. Ο δείκτης συμβολίζεται με n_δ.

Από την πρόταση 5.3.2 έπεται ότι αν η πολιτική δ^∞ ικανοποιεί την συνθήκη (A) και έχει δείκτη n_δ, τότε

$$\overline{D}^n = \overline{D}^{n_\delta} \quad \forall n \geq n_\delta$$

Επομένως η διαμέριση του χώρου Π τάξης n_δ που επάγεται από την δ^∞, $\mathbf{V}^{n_\delta}$, είναι μια τελική διαμέριση, η οποία δεν μπορεί να λεπτυνθεί περαιτέρω. Με άλλα λόγια, όλες οι διαμερίσεις τάξης $k \geq n_\delta$ που επάγονται από την δ^∞ ταυτίζονται με την διαμέριση $\mathbf{V}^{n_\delta} : \mathbf{V}^k = \mathbf{V}^{n_\delta}$.

Σημειώνουμε ότι ο δείκτης μιας πεπερασμένης μεταβατικής πολιτικής όπως ορίσθηκε στην ενότητα 5.2 (βλέπε ορισμό 5.2.2 και λήμμα 5.2.2) δεν συμπίπτει με τον δείκτη σύμφωνα με τον ορισμό 5.3.2. Έτσι στην εφαρμογή 5.2.1 ο δείκτης της δ^∞ σύμφωνα με τον ορισμό 5.2.2 είναι $n_\delta = 4$ ενώ σύμφωνα με τον ορισμό 5.3.1 είναι $n_\delta = 3$, επειδή $\overline{D}^2 \neq \overline{D}^3$ και $\overline{D}^3 = \overline{D}^4$.

Το βασικό αποτέλεσμα αυτής της ενότητας δίνεται από την ακόλουθη πρόταση.

Πρόταση 5.3.4: Αν η πολιτική δ^∞ ικανοποιεί την συνθήκη (Α),τότε αυτή επάγει Μαρκοβιανή διαμέριση του χώρου Π και η συνάρτηση $V(\pi/\delta), \pi \in \Pi$ είναι κατά τμήματα γραμμική.

Απόδειξη

Έστω n_δ ο δείκτης της πολιτικής δ^∞. Θα δείξουμε ότι η διαμέριση του χώρου Π τάξεως n_δ που επάγεται από την δ^∞, $\mathbf{V}^{n_\delta}$,είναι Μαρκοβιανή. Πράγματι, η διαμέριση $\mathbf{V}^{n_\delta}$ (όπως και όλες οι διαμερίσεις $\mathbf{V}^k$ τάξης k=0,1,2... που επάγονται από την δ^∞) ικανοποιεί την ιδιότητα (α) του ορισμού 5.2.1: η συνάρτηση ελέγχου δ είναι σταθερή σε κάθε κελλί της διαμέρισης. Θεωρούμε την διαμέριση τάξεως $n_\delta + 1$, $\mathbf{V}^{n_\delta+1}$. Σύμφωνα με την πρόταση 5.3.1 κάθε κελλί της διαμέρισης $\mathbf{V}^{n_\delta+1}$ απεικονίζεται εντός κάποιου κελλιού της διαμέρισης $\mathbf{V}^{n_\delta}$ μέσω της συνάρτησης μεταφοράς $T(.,\theta,\delta)$,όπου $\theta \in \Theta$.Λαμβάνοντας υπόψη την ταύτιση των διαμερίσεων $V^{n_\delta}, V^{n_\delta+1}$ συνάγεται ότι η διαμέριση $\mathbf{V}^{n_\delta}$ ικανοποιεί την ιδιότητα (β) του ορισμού 5.2.1:κάθε κελλί της $\mathbf{V}^{n_\delta}$ απεικονίζεται εντός κάποιου κελλιού της ίδιας διαμέρισης μέσω της $T(.,\theta,\delta)$,όπου $\theta \in \Theta$.Επομένως η διαμέριση $\mathbf{V}^{n_\delta}$ είναι Μαρκοβιανή. Επίσης η συνάρτηση τιμών της δ^∞, $V(\pi/\delta), \pi \in \Pi$ είναι κατά τμήματα γραμμική (βλέπε σχέσεις (5.2.2),(5.2.3)). □

Από τις προτάσεις 5.3.3 και 5.3.4 συμπεραίνουμε ότι αν η πολιτική δ^∞ είναι πεπερασμένα μεταβατική ή περιοδική τότε αυτή επάγει Μαρκοβιανή διαμέριση στον χώρο Π και η συνάρτηση τιμών V(./δ) είναι κατά τμήματα γραμμική. Αριθμητικές

εφαρμογές περιοδικών πολιτικών δίνονται σε ένα πρόβλημα αντικατάστασης στο κεφάλαιο 7.

ΣΥΜΠΕΡΑΣΜΑΤΑ

Στο κεφάλαιο αυτό, δείξαμε ότι η συνθήκη της πεπερασμένης μεταβατικότητας του Sondik είναι πιο ισχυρή από αυτή που στην πραγματικότητα χρειάζεται για να επιβεβαιώσουμε ότι μία στάσιμη πολιτική επάγει Μαρκοβιανή διαμέριση και παράλληλα η αντίστοιχη συνάρτηση τιμών είναι κατά τμήματα γραμμική .Εναλλακτικά παρουσιάσαμε τη συνθήκη περιοδικότητας, η οποία είναι αμοιβαία αποκλειόμενη από τη συνθήκη πεπερασμένης μεταβατικότητας.

Το πιο βασικό αποτέλεσμα του κεφαλαίου αυτού είναι μία γενίκευση των παραπάνω συνθηκών (**η συνθήκη (Α)**, ενότητα 5.3),η οποία επάγει Μαρκοβιανή διαμέριση και παράλληλα συνάρτηση τιμών κατά τμήματα γραμμική (πρόταση 5.3.4).Αριθμητικές εφαρμογές περιοδικών πολιτικών δίνονται σε ένα πρόβλημα αντικατάστασης στο κεφάλαιο 7.

ΚΕΦΑΛΑΙΟ 6

Πρόβλημα POMDP για την άριστη πολιτική αντικατάστασης συστήματος σε άπειρο χρονικό ορίζοντα στα πλαίσια της διάταξης του λόγου πιθανοφανειών $\leq_L$

Περίληψη

Στο κεφάλαιο αυτό εξετάζουμε ένα πρόβλημα συντήρησης / αντικατάστασης συστήματος, το οποίο περιγράφεται από μια POMDP. Οι καταστάσεις αντιπροσωπεύουν τα επίπεδα χειροτέρευσης του συστήματος.Το πρόβλημα αυτό μελετήθηκε από τους Ohnishi-Ibaraki [91].

Θεωρούμε ότι το σύστημα είναι μερικά παρατηρήσιμο μέσω ενός μηχανισμού ελέγχου, που αποφέρει μηνύματα σχετιζόμενα με τις καταστάσεις (επίπεδα χειροτέρευσης) του συστήματος. Με βάση κάποιες υποθέσεις, που αφορούν τον χαρακτήρα χειροτέρευσης, το χαρακτήρα των μηνυμάτων που αντανακλούν τα επίπεδα χειροτέρευσης, και τη δομή του άμεσου κόστους, οι Ohnishi-Ibaraki έδωσαν τη δομή της άριστης πολιτικής αντικατάστασης του συστήματος για άπειρο χρονικό ορίζοντα.

Στην ενότητα 6.1 περιγράφουμε το πρόβλημα, τις υποθέσεις και την εξίσωση αριστοποίησης για άπειρο χρονικό ορίζοντα.

Στην ενότητα 6.2 παρουσιάζουμε αποτελέσματα των Ohnishi-Ibaraki σχετικά με τις διατάξεις S.I (στοχαστικής μονοτονίας) και λόγου πιθανοφανειών στον χώρο των διανυσμάτων πληροφορίας Π, καθώς επίσης και ιδιότητες της άριστης συνάρτησης του αναμενόμενου ολικού εκπίπτοντος κόστους για άπειρο χρονικό ορίζοντα.

Στην ενότητα 6.3 δίνουμε τη δομή της άριστης πολιτικής αντικατάστασης, που μελετήθηκε από τους Ohnishi-Ibaraki.

Τέλος στην ενότητα 6.4 δίνουμε γεωμετρικές ιδιότητες της στοχαστικής διάταξης του λόγου πιθανοφανειών.

6.1. Περιγραφή και υποθέσεις

Θεωρούμε ένα σύστημα του οποίου η εξέλιξη περιγράφεται από μία POMDP (S,A,Θ,P,R,c) (βλέπε ενότητα 1.3) ως εξής :

- S={1,2,3…,N} είναι το σύνολο των καταστάσεων του συστήματος. Οι καταστάσεις δηλώνουν τα επίπεδα χειροτέρευσης του συστήματος και θεωρούνται μη παρατηρήσιμες. Η κατάσταση 1 δηλώνει την βέλτιστη κατάσταση, (το σύστημα είναι καινούργιο) και η κατάσταση N δηλώνει την χείριστη κατάσταση.
- A={0,1} είναι το σύνολο των αποφάσεων ,όπου **0,1** είναι οι κωδικοποιήσεις:

  0: συνέχιση της λειτουργίας του συστήματος /συντήρηση

  1: αντικατάσταση του συστήματος με ένα καινούργιο.

- $P = (p_{ij})$ είναι ο $N \times N$ πίνακας μετάβασης καταστάσεων του συστήματος, που αντιστοιχεί στην απόφαση α=**0** (συνέχιση λειτουργίας/συντήρηση).Έτσι αν $\{X_t, t \in \mathbb{N}_0\}$ είναι η στοχαστική διαδικασία των καταστάσεων και $\{Y_t, t \in \mathbb{N}_0\}$ είναι η στοχαστική διαδικασία των αποφάσεων, έχουμε

$$p_{ij} \equiv p[X_{t+1} = j / X_t = i, Y_t = 0], \ i, j \in S, t \in \mathbb{N}_0.$$

Αν στην χρονική περίοδο t ληφθεί η απόφαση α=1 (αντικατάσταση), τότε στην αρχή της επόμενης περιόδου $t+1$ η κατάσταση του συστήματος είναι 1 (καινούργιο σύστημα).

- Θ={1,2,3,…,M}, είναι το σύνολο των μηνυμάτων.Ο decision maker, αντί για την κατάσταση του συστήματος, παρατηρεί ένα μήνυμα θ μέσω ενός μηχανισμού ελέγχου στην αρχή κάθε περιόδου.
- $R = (r_{i\theta})$ είναι NXM πίνακας μηνυμάτων που αντιστοιχεί στην απόφαση α=0. Αν $\{ Z_t , t \in \mathbb{N} \}$ είναι η στοχαστική διαδικασία μηνυμάτων, έχουμε:

$$r_{i\theta} \equiv p[Z_{t+1}=\theta / X_t=i , Y_t=0] , i \in S, \theta \in \Theta , \ t \in \mathbb{N}_0 .$$

- Τα διανύσματα άμεσου κόστους που αντιστοιχούν στις αποφάσεις α=0,1 συμβολίζονται αντίστοιχα με

$C^K=C^0=(c(1,0), c(2,0),\ldots,c(N,0))^T$ και $C^R=C^1=(c(1,1), c(2,1),\ldots, c(N,1))^T$,

όπου $c(i,a)$ είναι το άμεσο κόστος για κάθε χρονική περίοδο ,όταν η κατάσταση του συστήματος είναι i και λαμβάνεται η απόφαση $\alpha \in A$. (one-step-cost).

Αν το δ.π στην χρονική περίοδο t είναι π(t)= π, επιλεγεί η απόφαση α= 0 (Y_t =0) και στην αρχή της επόμενης περιόδου t+1 παρατηρηθεί το μήνυμα θ, ($Z_{t+1} = \theta$) τότε το νέο δ.π. π(t+1)=T(π,θ), δίνεται από τη σχέση (βλέπε (1.4.4), (1.4.5)):

$$T(\pi,\theta)=\frac{\pi.P.R_\theta}{\{\theta/\pi\}}, \qquad \underline{\textbf{6.1.1}}$$

όπου R_θ είναι ο NXN διαγώνιος πίνακας

$$R_\theta = diag(r_{1\theta}, r_{2\theta}, \ldots r_{N\theta}) = \begin{pmatrix} r_{1\theta} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & r_{N\theta} \end{pmatrix},$$

{θ/π} είναι η πιθανότητα το επόμενο μήνυμα να είναι θ, (Z_{t+1}=θ), δεδομένου ότι το τρέχον δ.π είναι το π (π(t)= π) και η τρέχουσα απόφαση είναι α=0 (Y_t=0) και δίνεται από την σχέση (βλέπε (1.4.5))

$$\{\theta/\pi\}= \pi.P.R_\theta.\mathbf{1}, \qquad \underline{\textbf{6.1.2}}$$

όπου **1** είναι το NX1 διάνυσμα στήλη με όλα τα στοιχεία του 1.Όπως προκύπτει εύκολα από την (6.1.2), η πιθανότητα {θ/π} γράφεται επίσης ως η θ-συνιστώσα του διανύσματος π.P.R , δηλαδή

$$\{\theta/\pi\}=(\pi.P.R)_\theta \qquad \underline{\textbf{6.1.3}}$$

Επίσης εύκολα διαπιστώνουμε από την (6.1.1) ότι η j-συνιστώσα του διανύσματος T(π,θ) γράφεται:

$$T(\pi,\theta)_j = \frac{(\pi.P)_j . r_{j\theta}}{(\pi.P.R)_\theta} \qquad \underline{\textbf{6.1.4}}$$

Όπως ήδη έχουμε προαναφέρει, αν στην περίοδο t ληφθεί η απόφαση α=1 (αντικατάσταση), τότε στην επόμενη περίοδο t+1 η κατάσταση του συστήματος είναι 1, ή ισοδύναμα , το δ.π. είναι e_1 =(1,0,0,…,0),δηλαδή

$$\pi(t) \underset{a=1}{\rightarrow} \pi(t+1) = e_1 .$$

Έστω V(π),π∈ Π το ελάχιστο αναμενόμενο ολικό εκπίπτον κόστος για άπειρο χρονικό ορίζοντα, όταν το αρχικό δ.π. είναι π(o)=π.(βέλτιστη συνάρτηση τιμών για άπειρο χρονικό ορίζοντα).Η συνάρτηση V(π), π∈ Π ικανοποιεί την εξίσωση βελτιστοποίησης:

$$V(\pi) = \min \begin{cases} \pi C^0 + \beta \sum_{\theta \in \Theta} \{\theta|\pi\} V(T(\pi,\theta)) \\ \pi C^1 + \beta V(e^1) \end{cases} \quad , \pi \in \Pi \qquad \underline{6.1.5}$$

όπου $\beta \in (0,1)$ ο συντελεστής έκπτωσης.

Για αυθαίρετη φραγμένη συνάρτηση $u(\pi)$, $\pi \in \Pi(u \in B(\Pi))$,*ορίζουμε τις ακόλουθες συναρτήσεις* $[Cu](\pi)$,$[Ru](\pi)$, $\pi \in \Pi$

$$[Cu](\pi) := \pi C^0 + \beta \sum_{\theta \in \Theta} \{\theta|\pi\} u(T(\pi,\theta)) \qquad \underline{6.1.6}$$

$$[Ru](\pi) := \pi C^1 + \beta u(e^1) \qquad \underline{6.1.7}$$

Λαμβάνοντας υπόψη τους παραπάνω συμβολισμούς η εξίσωση βελτιστοποίησης (6.1.5) γράφεται :

$$V(\pi)=\min \{[CV](\pi),[RV](\pi)\}, \pi \in \Pi . \qquad \underline{6.1.8}$$

Συμβολίζουμε με F^N *το σύνολο των N-διάστατων διανυσμάτων με μη ελαττούμενες συνιστώσες:*

$$F^N = \{(\chi_1, \chi_2, ..., \chi_N) \in \mathbb{R}^N : \chi_1 \le \chi_2 \le ... \le \chi_N\} .$$

ΥΠΟΘΕΣΕΙΣ

Υ1 **Ο Ν Χ Ν πίνακας μετάβασης καταστάσεων Ρ είναι ολικά θετικός τάξεως 2 (TP_2) που σημαίνει ότι:**

$$\begin{vmatrix} p_{im} & p_{in} \\ p_{jm} & p_{jn} \end{vmatrix} \ge 0 \qquad i<j \quad m<n$$

Υ2 **Ο Ν Χ Μ πίνακας μηνυμάτων R είναι TP_2, που σημαίνει ότι:**

$$\begin{vmatrix} r_{i\theta} & r_{ik} \\ r_{j\theta} & r_{jk} \end{vmatrix} \ge 0 \qquad i<j,\ \theta<\kappa.$$

Υ3 $C^R \in F^N$

Υ4 $C^K \in F^N$

Υ5 $C^K - C^R \in F^N$

- Η υπόθεση Y_1 υποδηλώνει, ότι αν το σύστημα συνεχίσει να λειτουργεί, τότε είναι περισσότερο πιθανό να μεταβεί σε υψηλότερο επίπεδο χειροτέρευσης.
- Η υπόθεση Y_2 υποδηλώνει, ότι αυξανόμενου του επιπέδου χειροτέρευσης του συστήματος αυξάνει η πιθανότητα λήψης αυξημένης τιμής μηνύματος. Επομένως υψηλή τιμή μηνύματος είναι ένδειξη υψηλού επιπέδου χειροτέρευσης.
- Οι υποθέσεις Y_3 και Y_4 υποδηλώνουν ότι αυξανόμενου του επιπέδου χειροτέρευσης αυξάνουν τα άμεσα κόστη αντικατάστασης και λειτουργίας/ συντήρησης του συστήματος.
- Τέλος η υπόθεση Y_5 υποδηλώνει, ότι αυξανόμενου του επιπέδου χειροτέρευσης, το εύρος ανάμεσα στα άμεσα κόστη λειτουργίας και αντικατάστασης του συστήματος μειώνεται.

6.2. Στοχαστικές διατάξεις στον χώρο Π.

Στην ενότητα αυτή ορίζουμε στοχαστικές διατάξεις η εισαγωγή των οποίων είναι απαραίτητη για την απόδειξη δομικών ιδιοτήτων της άριστης συνάρτησης τιμών σε άπειρο χρονικό ορίζοντα $V(\pi)$, $\pi \in \Pi$ και της άριστης πολιτικής.

Ορισμός 6.2.1: Η στοχαστική διάταξη ***(ordinary-stochastic-ordering)*** $\leq_D$ *στο χώρο Π των δ. π. ορίζεται:*

Για $\chi, \psi \in \Pi$, $\chi \leq_D \psi$ *(x καλύτερο από το* ψ *ως προς την* $\leq_D$*), αν και μόνον αν:*

$$\sum_{i=k}^{N} x_i \leq \sum_{i=k}^{N} \psi_i \quad , \ 1 \leq \kappa \leq N.$$

Ορισμός 6.2.2: *Η στοχαστική διάταξη λόγου πιθανοφανειών* ***(likelihood-ratio-ordering)*** $\leq_L$ *στον χώρο Π ορίζεται :*

Για $\chi, \psi \in \Pi$, $\chi \leq_L \psi$ *(*χ *καλύτερο από το* ψ *ως προς* $\leq_L$*), αν και μόνον αν:*

$$\begin{vmatrix} \chi_i & \chi_j \\ \psi_i & \psi_j \end{vmatrix} \geq 0 \quad \text{για} \quad 1 \leq i \leq j \leq N .$$

Ορισμός 6.2.3: Ένας $N \times M$ στοχαστικός πίνακας $A=(a_{ij})$ καλείται στοχαστικά αυξανόμενος (stochastically-increasing), σύντομα SI, αν και μόνον αν :

$$A_i \leq_D A_j \ , 1 \leq i \leq j \leq N ,$$

όπου A_i είναι η $i-$*γραμμή* του πίνακα A

(δηλαδή αν

$$\sum_{l=k}^{M} a_{il} \leq \sum_{l=k}^{M} a_{jl} \ , 1 \leq i \leq j \leq N \ , 1 \leq k \leq M).$$

Ορισμός 6.2.4: Ένας $N \times M$ στοχαστικός πίνακας $A=(a_{ij})$ καλείται ολικά θετικός τάξεως 2 (TP_2) αν και μόνον αν

$$A_i \leq_L A_j \ , 1 \leq i \leq j \leq N \ ,$$

(δηλαδή αν

$$\begin{vmatrix} a_{im} & a_{in} \\ a_{jm} & a_{jn} \end{vmatrix} \geq 0, \quad 1 \leq i \leq j \leq N \ , 1 \leq m \leq n \leq M .$$

Λήμμα 6.2.1 :

(a) *Για* $\chi, \psi \in \Pi$, *αν* $\chi \leq_L \psi$ *τότε* $\chi \leq_D \psi$.

(b) *Για* $\chi, \psi \in \Pi$, $\chi \leq_D \psi$ *αν και μόνον αν:*

$$\chi.\alpha \leq \psi.\alpha$$

για κάθε $N \times 1$ *διάνυσμα* $\alpha \in F^N$.

(c) *Για* $\chi, \psi \in \Pi$, $\chi \leq_L \psi$ *αν και μόνον αν ο* $2 \times N$ *πίνακας* $\begin{pmatrix} \chi \\ \psi \end{pmatrix}$ είναι TP_2.

(d) *Για* $\chi, \psi \in \Pi$, *αν* $\chi \leq_D \psi$ *και A είναι NXM στοχαστικός πίνακας S.I, τότε*

$$\chi.A \leq_D \psi.A.$$

(e) *Για* $\chi, \psi \in \Pi$, *αν* $\chi \leq_L \psi$ *και A ένας NXM στοχαστικός πίνακας* TP_2, *τότε*

$$\chi.A \leq_L \psi.A.$$

(f) *Αν οι πίνακες A και B είναι στοχαστικοί* TP_2 *με διαστάσεις NXM , MXK, τότε και το γινόμενο A.B είναι ένας NXK στοχαστικός* TP_2 *πίνακας.*

(Derman [24]).

Πρόταση 6.2.1: *Αν* $\chi,\psi\in\Pi$, $\chi\leq_L\psi$, *τότε* $(\{\theta/\chi\})_{\theta\in\Theta}\leq_L(\{\theta/\psi\})_{\theta\in\Theta}$.

(Ohnishi-Ibaraki [91])

Απόδειξη

Απο τη σχέση (6.1.3) παίρνουμε:

$$(\{\theta/\chi\})_{\theta\in\Theta}=((\chi.P.R)_1,(\chi.P.R)_2,\ldots,(\chi.P.R)_M)=\chi.P.R.$$

Επειδή από τις υποθέσεις Y_1 και Y_2 οι στοχαστικοί πίνακες P και R είναι TP_2 ,τότε σύμφωνα με το λήμμα 6.2.1 (f) το γινόμενο P.R είναι TP_2 .Επειδή $\chi\leq_L\psi$, από το λήμμα 6.2.1 (e) παίρνουμε:

$$\chi PR\leq_L\psi PR.$$

Αρα $(\{\theta/\chi\})_{\theta\in\Theta}\leq_L(\{\theta/\psi\})_{\theta\in\Theta}$. □

Η παραπάνω πρόταση δηλώνει ότι όσο το δ.π. (δηλ. η κατανομή πιθανότητας που εκφράζει την άποψη του decision maker για την κατάσταση του συστήματος) χειροτερεύει ως προς $\leq_L$, *υπάρχει μεγαλύτερη πιθανότητα να πάρουμε μήνυμα μεγαλύτερου βαθμού χειροτέρευσης.*

Πρόταση 6.2.2:Για αυθαίρετο $\chi\in\Pi$,ισχύει ότι:

$$T(\chi,\theta)\leq_L T(\chi,\kappa)\quad\text{για}\quad 1\leq\theta\leq\kappa\leq M.$$ *(Ohnishi-Ibaraki* [91])

Απόδειξη

Ας είναι z =x.P .Από τη σχέση (6.1.4) παίρνουμε:

$$T(x,\theta)_i=\frac{z_i.r_{i\theta}}{(z.R)_\theta}$$

Τότε για $1\leq i\leq j\leq N$, $1\leq\theta\leq\kappa\leq M$ και λαμβάνοντας υπόψη ότι ο πίνακας R είναι TP_2, (υπόθεση Y2) , έχουμε:

$$\begin{vmatrix} T(x,\theta)_i & T(x,\theta)_j \\ T(x,k)_j & T(x,\kappa)_j \end{vmatrix} = \frac{z_i \cdot z_j}{(zR)_\theta \cdot (zR)_k} \cdot \begin{vmatrix} r_{i\theta} & r_{ik} \\ r_{j\theta} & r_{jk} \end{vmatrix} \geq 0.$$

Η παραπάνω πρόταση δηλώνει, ότι το επικαιροποιημένο δ.π. θα χειροτερεύσει με βάση την διάταξη, $\leq_L$, όταν ο μηχανισμός ελέγχου δώσει μήνυμα υψηλότερου βαθμού χειροτέρευσης.

Πρόταση 6.2.3: *Αν* $\chi, \psi \in \Pi$, $\chi \leq_L \psi$ *τότε* $T(\chi,\theta) \leq_L T(\psi,\theta)$ *για όλα τα* $\theta \in \Theta$.

(Ohnishi-Ibaraki [91])

Απόδειξη

As είναι $z^1=\chi.P$ και $z^2=\psi.P$. Επειδή ο πίνακας P είναι TP_2 (υπόθεση Y1) και $\chi \leq_L \psi$, από το λήμμα 6.2.1 (e) έχουμε:

$$z^1=\chi.P \leq_L z^2=\psi.P. \qquad \underline{6.2.1}$$

Για $1 \leq i \leq j \leq N$, εφαρμόζοντας τις σχέσεις (6.1.4) και (6.2.1) παίρνουμε:

$$\begin{vmatrix} T(\chi,\theta)_i & T(\chi,\theta)_j \\ T(\psi,\theta)_i & T(\psi,\theta)_j \end{vmatrix} = \frac{z_i^1 . r_{j\theta}}{(z^1.R)_\theta} \cdot \frac{z_j^2 . r_{j\theta}}{(z^2.R)_\theta} - \frac{z_j^1 . r_{j\theta}}{(z^1.R)_\theta} \cdot \frac{z_i^2 . r_{j\theta}}{(z^2.R)_\theta}$$

$$= \begin{vmatrix} z_i^1 & z_j^1 \\ z_i^2 & z_j^2 \end{vmatrix} \cdot \frac{r_{j\theta} \cdot r_{j\theta}}{(z^1.R)_\theta \cdot (z^2.R)_\theta} \geq 0 \qquad , \theta \in \Theta.$$

Άρα $T(\chi,\theta) \leq_L T(\psi,\theta)$, $\theta \in \Theta$. □

Ορισμός 6.2.5: Μια πραγματική συνάρτηση g με πεδίο ορισμού το σύνολο Π λέγεται $\leq_L$ αύξουσα, *αν* $\chi, \psi \in \Pi$ με $\chi \leq_L \psi \Rightarrow g(\chi) \leq g(\psi)$.

***Θεώρημα* 6.2.1:** *Αν* $f(x,\theta)$ *είναι μια πραγματική συνάρτηση με πεδίο ορισμού το καρτεσιανό* $\Pi X \Theta$ *που ικανοποιεί τις ακόλουθες δύο ιδιότητες:*

1) Για χ *αυθαίρετο, σταθερό, η συνάρτηση* $f(x,\theta)$ *είναι αύξουσα συνάρτηση του* θ.

2) Για θ αυθαίρετο, σταθερό, η συνάρτηση f είναι $\leq_L$ αύξουσα,

τότε:

$$\sum_{\theta\in\Theta}\{\theta/\chi\}.f(x,\theta)\leq\sum_{\theta\in\Theta}\{\theta/\psi\}.f(\psi,\theta) \quad \text{για} \quad \chi\leq_L\psi .$$

(Ohnishi-Ibaraki [91]).

Απόδειξη

Ορίζουμε $f(x)=(f(x,1),f(x,2),\ldots,f(x,M))^T$, $x\in\Pi$. Από ιδιότητα 1 έχουμε $f(x)\in F^M$, $x\in\Pi$. Από τη σχέση (6.1.3) παίρνουμε :

$$\sum_{\theta\in\Theta}\{\theta/\chi\}.f(\chi,\theta)=\sum_{\theta\in\Theta}(x.P.R)_\theta.f(x,\theta)=\chi.P.R.f(\chi)$$

Θεωρούμε $\chi,\psi\in\Pi$ με $\chi\leq_L\psi$.

Επειδή οι πίνακες P,R είναι TP_2 (υποθέσεις Y1,Y2), εφαρμόζοντας διαδοχικά τα *(f),(e),(a) και (b)* του λήμματος 6.2.1 παίρνουμε: $\chi PR\ f(x)\leq\psi PR\ f(\chi)$.

Από την ιδιότητα (2) έχουμε:

$$f(x,\theta)\leq f(\psi,\theta) \text{ για κάθε } \theta\in\Theta .$$

Από τα παραπάνω συνάγεται ότι:

$$\sum_{\theta\in\Theta}\{\theta/\chi\}.f(\chi,\theta)\leq\sum_{\theta\in\Theta}\{\theta/\psi\}.f(\chi,\theta)\ \leq\ \sum_{\theta\in\Theta}\{\theta/\psi\}.f(\psi,\theta). \qquad \square$$

Συμβολίζουμε με , **H**, την κλάση όλων των πραγματικών συναρτήσεων $\upsilon(\chi),\chi\in\Pi$ που είναι:

1. ***Συνεχείς στο χώρο Π.***
2. ***$\leq_L$ αύξουσες***
3. ***Κοίλες.***

Θεώρημα 6.2.2: *Αν η συνάρτηση $\upsilon(\pi),\pi\in\Pi$, ανήκει στην κλάση **H** ,τότε οι συναρτήσεις Cυ (π), Rυ (π), $\pi\in\Pi$ ανήκουν επίσης στην κλάση **H**.*

(Ohnishi-Ibaraki [91]).

Απόδειξη

1)Από τις σχέσεις (6.1.6) (6.1.7) οι συναρτήσεις Cυ, Rυ, *είναι* συνεχείς στον χώρο Π.

2)Εστω $\pi^1,\pi^2\in\Pi$, $\pi^1\le_L\pi^2$.Αφού $C^K\in F^N$ εφαρμόζοντας διαδοχικά τα (α) και (b) του λήμματος 6.2.1 παίρνουμε $\pi^1\le_D\pi^2$

και

$$\pi^1.C^K\le\pi^2.C^K \qquad \underline{6.2.2}$$

Από την πρόταση 6.2.3 έχουμε:

$$T(\pi^1,\theta)\le_L T(\pi^2,\theta)\ ,\forall\theta\in\Theta.$$

Επειδή η υ είναι $\le_L$ αύξουσα έχουμε:

$$\upsilon(T(\pi^1,\theta))\le\upsilon(T(\pi^2,\theta))\quad\forall\theta\in\Theta. \qquad \underline{6.2.3}$$

Επιπλέον για $\pi\in\Pi$, $1\le\theta\le\kappa\le M$ έχουμε $T(\pi,\theta)\le_L T(\pi,k)$ (πρόταση 6.2.2), οπότε για σταθερό π η υ(T(π,θ)) είναι αύξουσα συνάρτηση του θ.Επομένως η συνάρτηση

$$f(\pi,\theta):=\upsilon(T(\pi,\theta)),\quad(\pi,\theta)\in\Pi\times\Theta,$$

ικανοποιεί τις ακόλουθες ιδιότητες:

1) για π σταθερό η $f(\pi,\theta)$ είναι αύξουσα συνάρτηση του θ.

2) για θ σταθερό η συνάρτηση $f(\pi,\theta)$ είναι $\le_L$ αύξουσα.

Σύμφωνα με το θεώρημα 6.2.1 παίρνουμε:

$$\sum_{\theta\in\Theta}\{\theta/\pi^1\}.\upsilon(T(\pi^1,\theta))\le\sum_{\theta\in\Theta}\{\theta/\pi^2\}.\upsilon(T(\pi^2,\theta))\ \text{αν}\ \pi^1\le_L\pi^2 \qquad \underline{6.2.4}$$

Από τις (6.1.6),(6.2.2), (6.2.4) έχουμε: $Cu(\pi^1)\le Cu(\pi^2)$, δηλαδή η συνάρτηση Cu είναι $\le_L$ αύξουσα

Επειδή $\pi^1\le_L\pi^2$ και $C^R\in F^N$, από τα (α) και(b) του λήμματος 6.2.1 παίρνουμε:

$$\pi^1.C^R\le\pi^2.C^R \qquad \underline{6.2.5}$$

Επομένως από τις σχέσεις (6.1.7),(6.2.5) η συνάρτηση Ru είναι $\le_L$ αύξουσα.

3) Έστω $\pi^1,\pi^2\in\Pi$. Για $\theta\in\Theta$ έχουμε :

$$\{\theta/\lambda.\pi^1+(1-\lambda).\pi^2\}=\lambda.\{\theta/\pi^1\}+(1-\lambda)\{\theta/\pi^2\}$$

και (βλέπε πρόταση 1.4.1)

$$T(\lambda.\pi^1(1-\lambda).\pi^2,\theta)=\frac{\lambda.\{\theta/\pi^1\}}{\{\theta/\lambda.\pi^1+(1-\lambda).\pi^2\}}.T(\pi^1,\theta)+\frac{(1-\lambda).\{\theta/\pi^2\}}{\{\theta/\lambda\pi^1+(1-\lambda).\pi^2\}}.T(\pi^2,\theta).$$

Επειδή η συνάρτηση υ υποτέθηκε κοίλη παίρνουμε:

$$\{\theta/\lambda.\pi^1+(1-\lambda).\pi^2\}.\upsilon(T(\lambda.\pi^1+(1-\lambda).\pi^2,\theta))=$$

$$\{\theta/\lambda.\pi^1+(1-\lambda).\pi^2\}.\upsilon(\frac{\lambda.\{\theta/\pi^1\}}{\{\theta/\lambda.\pi^1+(1-\lambda).\pi^2\}}.T(\pi^1,\theta)+\frac{(1-\lambda).\{\theta/\pi^2\}}{\{\theta/\lambda\pi^1+(1-\lambda).\pi^2\}}.T(\pi^2,\theta)$$

$$\ge\lambda.\{\theta/\pi^1\}.\upsilon(T(\pi^1,\theta)+(1-\lambda)\{\theta/\pi^2\}.\upsilon(T(\pi^2,\theta)).$$

Αρα για κάθε $\theta \in \Theta$, η συνάρτηση $\{\theta/\pi\}.\upsilon(T(\pi,\theta)), \pi \in \Pi$ είναι κοίλη.

Η συνάρτηση $\pi.C^K$ ως γραμμική είναι κοίλη. Επομένως από την 6.1.6 η συνάρτηση Cυ είναι κοίλη. Τέλος η Rυ είναι κοίλη ως γραμμική.

Επομένως Cυ, Rυ ανήκουν στην κλάση **H**. □

Θεώρημα 6.2.3: *Αν $V_n(\pi)$,το βέλτιστο προσδοκώμενο κόστος σε n περιόδους, δεδομένου ότι το αρχικό δ.π είναι το π, τότε:*

$$V_n(\pi) \xrightarrow[n\to\infty]{} V(\pi),\ \pi \in \Pi.$$

και μάλιστα η βέλτιστη συνάρτηση ,$V(\pi)$,είναι επίσης στοιχείο της κλάσης ***H****.*

Απόδειξη

Θεωρούμε τις συναρτήσεις $V_n(\pi)$ $\pi \in \Pi$, n=0,1,2…, που ορίζονται αναγωγικά :

$$V_0(\pi)=0$$

$$V_{n+1}(\pi)= \min\{ CV_n(\pi), RV_n(\pi)\}$$

Επειδή $V_0 \in$ **H,** και η πράξη min διατηρεί τις ιδιότητες 1,2,3 της κλάσης **H**, συμπεραίνουμε επαγωγικά ότι $V_n \in$ **H** (n=0,1,2…). Αφού η σύγκλιση

$$V_n(\pi) \xrightarrow[n\to\infty]{} V(\pi),\ \pi \in \Pi.$$

είναι ομαλή, συμπεραίνουμε ότι:$V \in$ **H** . □

Συνοψίζοντας, η βέλτιστη συνάρτηση τιμών για άπειρο χρονικό ορίζοντα, $V(\pi)$, $\pi \in \Pi$ είναι συνεχής, $\leq_L$ αύξουσα και κοίλη. Αποδεικνύεται εύκολα εφαρμόζοντας την υπόθεση Y5 ότι η συνάρτηση

$$CV(\pi) - RV(\pi) = \pi.(C^k - C^R) + \beta.\{\sum_{\theta\in\theta}\{\theta/\pi\}.V(T(\pi,\theta)) - V(e^1)\},\ \pi \in \Pi$$

ανήκει επίσης στην κλάση **H.**

6.3. Δομικές ιδιότητες της βέλτιστης πολιτικής αντικατάστασης

Τα κύρια αποτελέσματα της ενότητας αυτής παράγονται από το γεγονός ότι $V \in$**H**, CV-RV $\in$**H.**

Θεωρούμε τα ακόλουθα σύνολα

$$\mathfrak{I}_C=\{\pi\in\Pi: V(\pi)= CV(\pi)\}$$

$$\mathfrak{I}_R=\{\pi\in\Pi: V(\pi)=RV(\pi)\}$$

$\mathfrak{I}_C, \mathfrak{I}_R$ είναι οι περιοχές όπου είναι βέλτιστες οι αποφάσεις α=0 (συνέχιση λειτουργίας) και α=1 (αντικατάσταση) αντίστοιχα.

Πρόταση 6.3.1:

i) Το σύνολο $\mathfrak{I}_R$ είναι ένα κυρτό υποσύνολο του Π.

ii) Αν $\pi^0 \in \mathfrak{I}_R$, τότε $\pi^0 \leq_L \pi \Rightarrow \pi \in \mathfrak{I}_R$

iii) Αν $\pi^0 \in \mathfrak{I}_C$, τότε $\pi \leq_L \pi^0 \Rightarrow \pi \in \mathfrak{I}_C$

(Ohnishi-Ibaraki) [91]

Με απλά λόγια το *ii)* δηλώνει ότι αν για το καλύτερο εφαρμόζουμε αντικατάσταση, πόσο μάλλον για το χειρότερο.

Αντίστοιχα το *iii)* δηλώνει ότι αν για το χειρότερο δεν εφαρμόζουμε αντικατάσταση, πόσο μάλλον για το καλύτερο.

Κρίνουμε απαραίτητο στο σημείο αυτό να δώσουμε τους εξής ορισμούς:

Ορισμός 6.3.1: Ένα σύνολο $J \subseteq \Pi$ λέγεται $\leq_L$ αυξάνον, αν : $x \in J, x \leq_L \psi \Rightarrow \psi \in J$.

Ορισμός 6.3.2: Ένα σύνολο $J' \subseteq \Pi$ λέγεται $\leq_L$ φθίνον, αν : $\psi \in J', x \leq_L \psi \Rightarrow \chi \in J'$.

Σημειώνουμε ότι αν $J \neq \varnothing$ τότε $\mathbf{e}_N \in J$. Επίσης αν $J' \neq \varnothing$ τότε $\mathbf{e}_1 \in J'$.

Σύμφωνα με τον ορισμό 6.3.1 το σύνολο $\mathfrak{I}_R$ είναι $\leq_L$ αυξάνον και σύμφωνα με τον ορισμό 6.3.2 το σύνολο $\mathfrak{I}_C$ είναι $\leq_L$ φθίνον.

Παρατήρηση

Για N=2 καταστάσεις, η άριστη πολιτική αντικατάστασης του συστήματος, σύμφωνα με την πρόταση 6.3.1 είναι η control-limit πολιτική $(\delta^*)^\infty$, όπου

$$\delta^*(\pi)=\begin{cases}0 & \forall \pi=(1-\pi_2,\pi_2)\in\Pi \quad \mu\varepsilon \quad \pi_2 \leq p^* \\ 1 & \forall \pi=(1-\pi_2,\pi_2)\in\Pi \quad \mu\varepsilon \quad \pi_2 \geq p^*\end{cases}$$

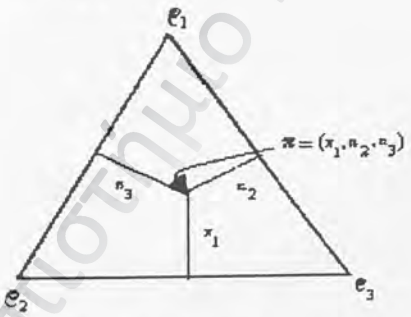
όπου $p^* \in (0,1)$ κατάλληλη κρίσιμη ποσότητα.

6.4. Γεωμετρική Ερμηνεία της διάταξης $\leq_L$, στην περίπτωση τριών καταστάσεων (N=3).

Στην ενότητα αυτή θα δώσουμε γεωμετρική ερμηνεία της μερικής διάταξης $\leq_L$ στην ειδική περίπτωση τριών καταστάσεων (N=3). Τότε ο χώρος των δ.π.

$$\Pi=\{\pi=(\pi_1,\pi_2,\pi_3): \pi_i \geq 0,\ i=1,2,3,\ \pi_1+\pi_2+\pi_3=1\}$$

παριστάνεται γραφικά με ένα ισόπλευρο τρίγωνο με κορυφές $e_1=(1,0,0), e_2=(0,1,0)$, $e_3=(0,0,1)$ και ύψη ίσα με την μονάδα. Ένα δ.π. $\pi=(\pi_1,\pi_2,\pi_3)$ απεικονίζεται στο σημείο του τριγώνου, που οι αποστάσεις από τις πλευρές e_2e_3, e_1e_3, e_1e_2 είναι αντίστοιχα π_1,π_2,π_3 (βλέπε σχήμα 6.1).



Σχήμα 6.1: Ο χώρος Π για N=3.

Πρόταση 6.4.1: *Εστω δ.π* $\pi=(\pi_1,\pi_2,\pi_3) \in \Pi$.

i) Το σύνολο των δ.π που είναι ''χειρότερα'' του π *,ως προς* $\leq_L$

$$D(\pi)=\{\pi' \in \Pi: \pi \leq_L \pi'\}$$

απεικονίζεται στο τρίγωνο($\pi e_3 \Gamma$*) με κορυφές* π, e_3, $\Gamma(0, \frac{\pi_2}{\pi_2+\pi_3}, \frac{\pi_3}{\pi_2+\pi_3})$.

ii) Το σύνολο των δ.π που είναι ''καλύτερα'' του π *ως προς* $\leq_L$

$$\aleph(\pi)=\{\pi' \in \Pi: \pi' \leq_L \pi\}$$

απεικονίζεται στο τρίγωνο $\pi e_1 A$ με κορυφές π, e_1, $A=(\frac{\pi_1}{\pi_1+\pi_2}, \frac{\pi_2}{\pi_1+\pi_2}, 0)$.

Απόδειξη

Σύμφωνα με τον ορισμό 6.2.2 έχουμε $\pi \leq_L \pi'$ αν και μόνον αν συναληθεύουν οι σχέσεις:

$$\begin{vmatrix} \pi_1 & \pi_2 \\ \pi_1' & \pi_2' \end{vmatrix} = \pi_1.\pi_2' - \pi_2.\pi_1' \geq 0 \qquad \underline{6.4.1}$$

$$\begin{vmatrix} \pi_1 & \pi_3 \\ \pi_1' & \pi_3' \end{vmatrix} = \pi_1.\pi_3' - \pi_1'.\pi_3 \geq 0 \qquad \underline{6.4.2}$$

$$\begin{vmatrix} \pi_2 & \pi_3 \\ \pi_2' & \pi_3' \end{vmatrix} = \pi_2.\pi_3' - \pi_2'.\pi_3 \geq 0 \qquad \underline{6.4.3}$$

Αρχικά θα εντοπίσουμε την περιοχή στην οποία απεικονίζεται το σύνολο των $\pi' \in \Pi$, που ικανοποιούν την (6.4.1).

Για την κορυφή $e_1=(1,0,0)$ έχουμε: $\begin{vmatrix} \pi_1 & \pi_2 \\ 1 & 0 \end{vmatrix} = -\pi_2 \leq 0$

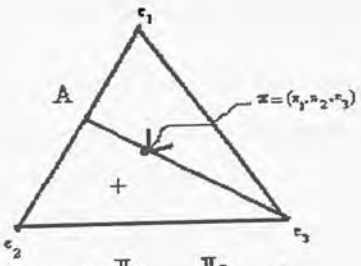
Για την κορυφή $e_2=(0,1,0)$ έχουμε: $\begin{vmatrix} \pi_1 & \pi_2 \\ 0 & 1 \end{vmatrix} = \pi_1 \geq 0$

Για την κορυφή $e_3=(0,0,1)$ έχουμε: $\begin{vmatrix} \pi_1 & \pi_2 \\ 0 & 0 \end{vmatrix} = 0$

Επομένως η κορυφή e_2 και οριακά η e_3 ικανοποιούν την (6.4.1).

Τα σημεία $(\lambda, 1-\lambda, 0)$ του ευθυγράμμου τμήματος e_1e_2 πού ικανοποιούν την (6.4.1) είναι εκείνα για τα οποία $0 \leq \lambda \leq \frac{\pi_1}{\pi_1+\pi_2}$.

Πράγματι $\begin{vmatrix} \pi_1 & \pi_2 \\ \lambda & 1-\lambda \end{vmatrix} = \pi_1 - (\pi_1+\pi_2).\lambda \geq 0 \Leftrightarrow \lambda \leq \frac{\pi_1}{\pi_1+\pi_2}$. Από τα ανωτέρω γίνεται φανερό ότι το σύνολο των $\pi' \in \Pi$ που ικανοποιούν την (6.4.1) απεικονίζεται στην περιοχή + (τρίγωνο Ae_2e_3) του σχήματος 6.2.

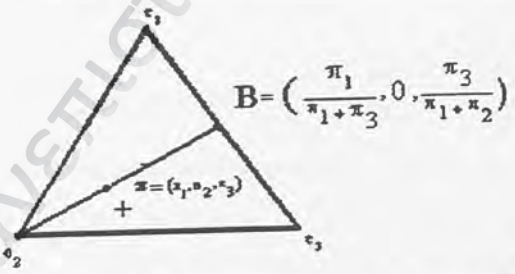


***Σχήμα 6.2:* Περιοχή(+) επαληθεύουσα την ανισοτική σχέση (6.4.1).**

Το ευθύγραμμο τμήμα Ae_3 , για το οποίο η σχέση (6.4.1) ισχύει σαν ισότητα, περιγράφεται σαν κυρτός γραμμικός συνδυασμός των σημείων $e_3=(0,0,1)$ και $A=(\frac{\pi_1}{\pi_1+\pi_2},\frac{\pi_2}{\pi_1+\pi_2},0)$,δηλαδή

$$Ae_3 : \lambda.(\frac{\pi_1}{\pi_1+\pi_2},\frac{\pi_2}{\pi_1+\pi_2},0)+(1-\lambda).(0,0,1)=(\lambda.\frac{\pi_1}{\pi_1+\pi_2},\lambda\frac{\pi_2}{\pi_1+\pi_2},1-\lambda),\ 0\le\lambda\le 1$$

Σημειώνουμε ότι το $\pi\in Ae_3$,πράγμα πού επαληθεύεται για $\lambda=\pi_1+\pi_2$.Παρόμοια το σύνολο των $\pi'\in\Pi$ που ικανοποιούν την (6.4.2) απεικονίζεται στην περιοχή + (τρίγωνο Be_2e_3) του σχήματος 6.3.

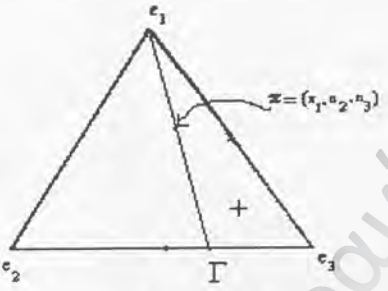


Σχήμα 6.3: Περιοχή επαληθεύουσα την ανισοτική σχέση (6.4.2).

Το ευθύγραμμο τμήμα Be_2 για το οποίο η σχέση (6.4.2) ισχύει με ισότητα, περιγράφεται σαν κυρτός συνδυασμός των σημείων $B=(\frac{\pi_1}{\pi_1+\pi_3},0,\frac{\pi_3}{\pi_1+\pi_3})$ και $e_2=(0,1,0)$,δηλαδή:

Be_2: $\lambda.\left(\frac{\pi_1}{\pi_1+\pi_3}, 0, \frac{\pi_3}{\pi_1+\pi_3}\right)+(1-\lambda).(0,1,0)=\left(\lambda.\frac{\pi_1}{\pi_1+\pi_3}, 1-\lambda, \lambda\frac{\pi_3}{\pi_1+\pi_3}\right)$, $0\le\lambda\le 1$

Σημειώνουμε ότι $\pi\in Be_2$, πράγμα που επαληθεύεται για $\lambda=\pi_1+\pi_3$.Τέλος το σύνολο των $\pi'\in\Pi$ που ικανοποιούν την ανισοτική σχέση (6.4.3) απεικονίζεται στην περιοχή + (τρίγωνο $e_1\Gamma e_3$) του σχήματος 6.4.



$$\Gamma=\left(0,\frac{\pi_2}{\pi_2+\pi_3},\frac{\pi_3}{\pi_2+\pi_3}\right)$$

Σχήμα 6.4: Περιοχή επαληθεύουσα την ανισοτική σχέση (6.4.3).

Το ευθύγραμμο τμήμα Γe_1 για το οποίο η σχέση (6.4.3) ισχύει με ισότητα περιγράφεται ως κυρτός γραμμικός συνδυασμός των σημείων

$\Gamma=\left(0,\frac{\pi_2}{\pi_2+\pi_3},\frac{\pi_3}{\pi_2+\pi_3}\right)$, και $e_1=(1,0,0)$, δηλαδή:

$$\Gamma e_1:\ \lambda\left(0,\frac{\pi_2}{\pi_2+\pi_3},\frac{\pi_3}{\pi_2+\pi_3}\right)+(1-\lambda).(1,0,0)=\left(1-\lambda,\lambda\frac{\pi_2}{\pi_2+\pi_3},\lambda\frac{\pi_3}{\pi_2+\pi_3}\right),\ 0\le\lambda\le 1$$

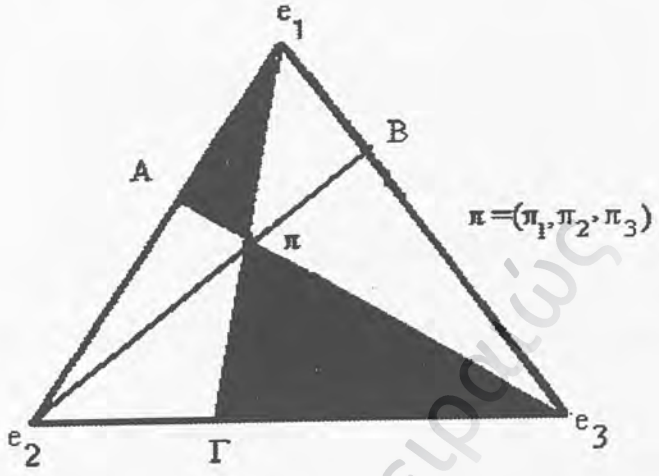
Σημειώνουμε ότι $\pi\in\Gamma e_1$,πράγμα που επαληθεύεται για $\lambda=\pi_2+\pi_3$.

Από τα ανωτέρω συνάγεται ότι τα ευθύγραμμα τμήματα Γe_1,Be_2,Ae_3 τέμνονται στο σημείο $\pi=(\pi_1,\pi_2,\pi_3)$.Το σύνολο $D(\pi)$ απεικονίζεται στο τρίγωνο $\pi e_3\Gamma$,το οποίο είναι η τομή των περιοχών που αντιστοιχούν στις ανισώσεις (6.4.1), (6.4.2), (6.4.3)

Με παρόμοια ανάλυση, το σύνολο $\aleph(\pi)$, απεικονίζεται στο τρίγωνο $\pi e_1 A$ του σχήματος 6.5. □

Σημειώνουμε ότι τα δ.π. που ανήκουν στις περιοχές ($A\pi\Gamma e_2$) και ($e_1\pi e_3$),δεν είναι $\le_L$ συγκρίσιμα με το π. Αυτό οφείλεται στο γεγονός ότι η διάταξη $\le_L$ **είναι μερική και όχι ολική διάταξη.**

Η παρακάτω πρόταση συνοψίζει τις κυριότερες ιδιότητες των συνόλων D(π) και αναφέρεται σε σύστημα με Ν δυνατές καταστάσεις.



Σχήμα 6.5: Περιοχές ℵ(π), D(π)

$$\mathbf{A}=\left(\frac{\pi_1}{\pi_1+\pi_2},\frac{\pi_2}{\pi_1+\pi_2},0\right),\ \mathbf{B}=\left(\frac{\pi_1}{\pi_1+\pi_3},0,\frac{\pi_3}{\pi_1+\pi_3}\right),\ \mathbf{\Gamma}=\left(0,\frac{\pi_2}{\pi_2+\pi_3},\frac{\pi_3}{\pi_2+\pi_3}\right)$$

D(π): τρίγωνο $\pi e_3\Gamma$, ℵ(π):τρίγωνο $\pi e_1 A$

Πρόταση 6.4.2:

i) *Το σύνολο, D(π), $\pi\in\Pi$ είναι κυρτό.*

ii) *Αν $\pi, \pi'\in\Pi$ με $\pi\leq_L\pi'$ τότε $D(\pi')\subseteq D(\pi)$.*

iii) *Για $\pi\in\Pi$, το σύνολο $D(\pi)$ είναι $\leq_L$ αυξάνον.*

iv) *Αν $X\subseteq\Pi$, μη κενό σύνολο, τότε τα σύνολα*

$\bigcup_{\pi\in X} D(\pi)$ *και* $\bigcap_{\pi\in X} D(\pi)$ *είναι $\leq_L$ αυξάνοντα.*

Απόδειξη

i) Εστω $\pi', \pi'' \in D(\pi)$ και $0 \le \lambda \le 1$. Θα αποδείξουμε ότι: $\lambda.\pi' + (1-\lambda).\pi'' \in D(\pi)$.

Επειδή $\pi \le_L \pi'$, $\pi \le_L \pi''$ έχουμε: $\begin{vmatrix} \pi_i & \pi_j \\ \pi_i' & \pi_j' \end{vmatrix} = \pi_i.\pi_j' - \pi_i'.\pi_j \ge 0$, $1 \le i \le j \le N$ **6.4.4**

$$\begin{vmatrix} \pi_i & \pi_j \\ \pi_i'' & \pi_j'' \end{vmatrix} = \pi_i.\pi_j'' - \pi_i''.\pi_j \ge 0 \quad , 1 \le i \le j \le N . \qquad \textbf{6.4.5}$$

Αρα

$$\begin{vmatrix} \pi_i & \pi_j \\ \lambda.\pi_i' + (1-\lambda).\pi_i'' & \lambda.\pi_j' + (1-\lambda).\pi_j'' \end{vmatrix} =$$

$$= \lambda. \begin{vmatrix} \pi_i & \pi_j \\ \pi_i' & \pi_j' \end{vmatrix} + (1-\lambda). \begin{vmatrix} \pi_i & \pi_j \\ \pi_i'' & \pi_j'' \end{vmatrix} \ge 0 \text{ για } 1 \le i \le j \le N .$$

Επομένως $\pi \le_L \lambda.\pi' + (1-\lambda).\pi'' \in D(\pi)$.

ii) Εστω $\pi'' \in D(\pi')$, δηλαδή $\pi' \le_L \pi''$. Επειδή $\pi \le_L \pi'$, από την μεταβατική ιδιότητα συνάγεται ότι : $\pi \le_L \pi''$, δηλαδή $\pi'' \in D(\pi) \Rightarrow D(\pi') \subseteq D(\pi)$.

iii) Εστω $\pi' \in D(\pi)$ και $\pi'' \in \Pi$ έτσι ώστε $\pi' \le_L \pi''$. Επειδή $\pi \le_L \pi'$, από μεταβατική ιδιότητα προκύπτει $\pi \le_L \pi''$, δηλαδή $\pi'' \in D(\pi)$. Αρα το σύνολο $D(\pi)$ *είναι* $\le_L$ αυξάνον.

iv) Προφανές, στηρίζεται στο iii). □

Με ανάλογο τρόπο συνοψίζονται οι κυριότερες ιδιότητες των συνόλων $\aleph(\pi)$ στην ακόλουθη πρόταση.

Πρόταση 6.4.3:

i) *Το σύνολο,* $\aleph(\pi)$, $(\pi \in \Pi)$ *είναι κυρτό.*

ii) *Αν* $\pi, \pi' \in \Pi$ *με* $\pi' \le_L \pi$ *τότε* $\aleph(\pi') \subseteq \aleph(\pi)$.

iii) *Για* $\pi \in \Pi$, *το σύνολο* $\aleph(\pi)$ *είναι* $\le_L$ *-φθίνον.*

iv) *Αν* $X \subseteq \Pi$, *μη κενό σύνολο, τότε τα σύνολα* $\bigcup_{\pi \in X} \aleph(\pi)$ *και* $\bigcap_{\pi \in X} \aleph(\pi)$ *είναι* $\le_L$ *φθίνοντα.* □

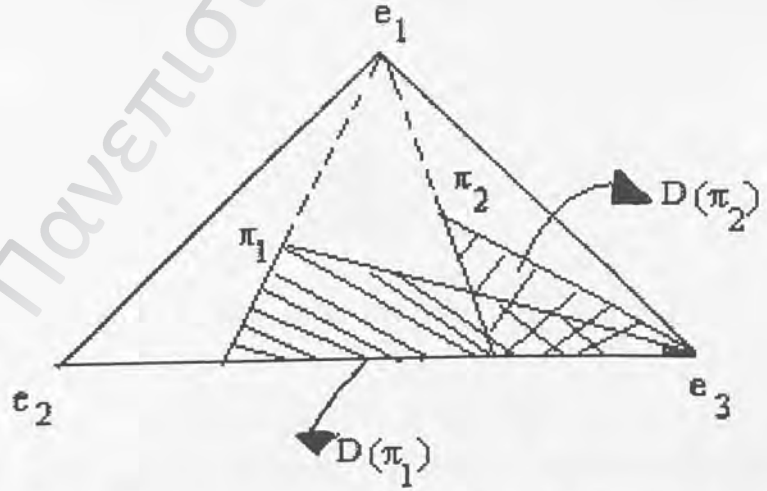
Παρατηρήσεις:

1) Ένα $\leq_L$ αυξάνον σύνολο δεν είναι πάντοτε κυρτό. Πράγματι, στο σχήμα 6.6 η ένωση $D(\pi_1) \cup D(\pi_2)$ είναι $\leq_L$ αυξάνον σύνολο, το οποίο όμως δεν είναι κυρτό.

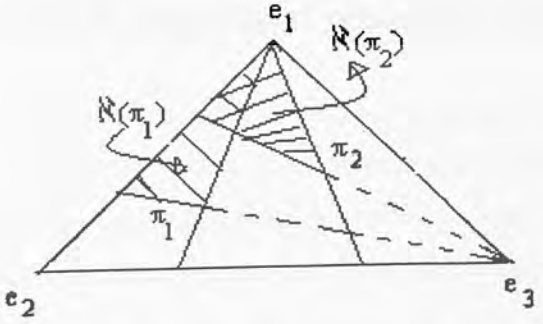
2) Ένα $\leq_L$ φθίνον σύνολο δεν είναι πάντοτε κυρτό. Πράγματι, στο σχήμα 6.7 η ένωση $\aleph(\pi_1) \cup \aleph(\pi_2)$ είναι $\leq_L$ φθίνον σύνολο, το οποίο όμως δεν είναι κυρτό.

3) Ένα κυρτό σύνολο που περιέχει το e_3 δεν είναι πάντοτε $\leq_L$ αυξάνον. Πράγματι, στο σχήμα 6.8 η περιοχή D (τετράπλευρο $\Delta\pi Ee_3$) περιέχει το e_3 και είναι κυρτή. Όμως το σύνολο D δεν είναι $\leq_L$ αυξάνον επειδή για το δ.π. $\pi' = (0, \frac{\pi_2}{\pi_2+\pi_3}, \frac{\pi_3}{\pi_2+\pi_3})$ έχουμε $\pi \leq_L \pi'$ και $\pi' \notin D$.

4) Ένα κυρτό σύνολο που περιέχει το e_1 δεν είναι πάντοτε $\leq_L$ φθίνον. Πράγματι, στο σχήμα 6.9 η περιοχή L (τετράπλευρο $Z\pi He_1$) περιέχει το e_1 και είναι κυρτή. Όμως το σύνολο L δεν είναι $\leq_L$ φθίνον επειδή για το δ.π. $\pi'' = (\frac{\pi_1}{\pi_1+\pi_2}, \frac{\pi_2}{\pi_1+\pi_2}, 0)$ έχουμε $\pi'' \leq_L \pi$ και $\pi'' \notin L$.

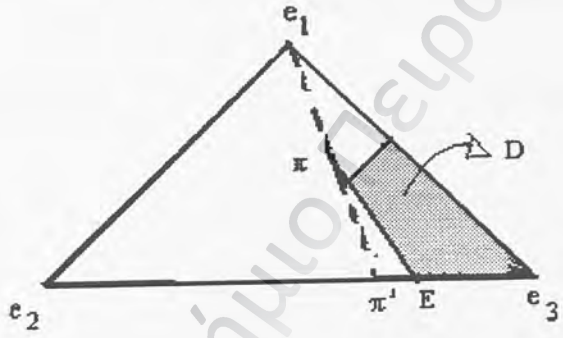


Σχήμα 6.6: **Το $\leq_L$ αυξάνον σύνολο $D(\pi_1) \cup D(\pi_2)$ δεν είναι κυρτό.**

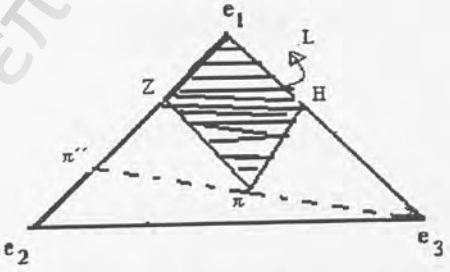


Σχήμα 6.7: **Το $\leq_L$ φθίνον σύνολο $\aleph(\pi_1)\cup\aleph(\pi_2)$ δεν είναι κυρτό.**

$$\pi'=(0,\frac{\pi_2}{\pi_2+\pi_3},\frac{\pi_3}{\pi_2+\pi_3}),\quad \pi=(\pi_1,\pi_2,\pi_3)$$



Σχήμα 6.8: **Το κυρτό σύνολο D, περιέχει το e_3 και δεν είναι $\leq_L$ *αυξάνον.***

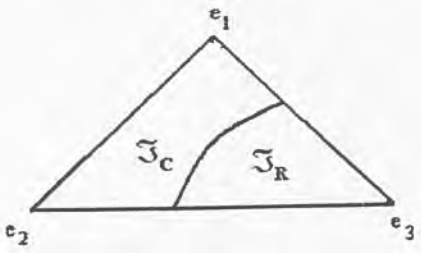


$$\pi''=(\frac{\pi_1}{\pi_1+\pi_2},\frac{\pi_2}{\pi_2+\pi_3},0),\quad \pi=(\pi_1,\pi_2,\pi_3).$$

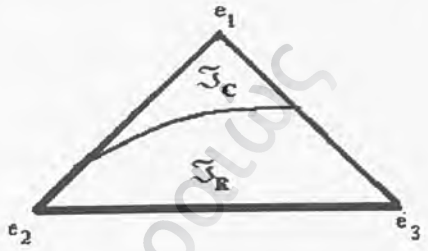
Σχήμα 6.9: **Το κυρτό σύνολο L, περιέχει το e_1 και δεν είναι $\leq_L$ *φθίνου.***

Γεωμετρική ερμηνεία της άριστης πολιτικής αντικατάστασης συστήματος

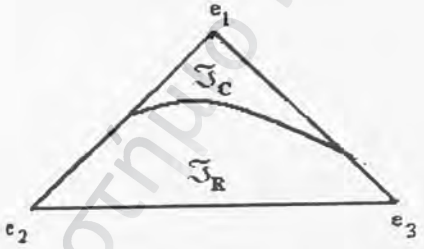
Με τα σχήματα 6.10-6.12 παρουσιάζουμε διάφορες εκδοχές της άριστης πολιτικής για άπειρο χρονικό ορίζοντα σχετικά με τη διαμέριση του χώρου Π στην $\leq_L$ φθίνουσα περιοχή συνέχισης λειτουργίας/συντήρησης $\Im_C$ και την $\leq_L$ αύξουσα κυρτή περιοχή αντικατάστασης $\Im_R$ (βλέπε πρόταση 6.3.1) για τρεις καταστάσεις (N=3).



Σχήμα 6.10



Σχήμα 6.11



Σχήμα 6.12

ΣΥΜΠΕΡΑΣΜΑΤΑ

Στο κεφάλαιο 6 παρουσιάσαμε ένα γενικό πρόβλημα αντικατάστασης συστήματος με πεπερασμένο πλήθος επιπέδων χειροτέρευσης , που παρατηρείται ατελώς μέσω μηνυμάτων ενός μηχανισμού ελέγχου, με δύο δυνατές αποφάσεις (συνέχιση της λειτουργίας/συντήρησης ή αντικατάστασης του συστήματος).Η δομή της άριστης πολιτικής μελετήθηκε από τους Ohnishi-Ibaraki στα πλαίσια της διάταξης λόγου

πιθανοφανειών $\leq_L$.Σύμφωνα με τη δομή αυτή η περιοχή αντικατάστασης είναι $\leq_L$ αυξάνον κυρτό σύνολο και η περιοχή συντήρησης είναι $\leq_L$ φθίνον σύνολο. Για την παραπάνω *δομή δίνουμε γεωμετρικές ιδιότητες στην* περίπτωση τριών καταστάσεων.

ΚΕΦΑΛΑΙΟ 7

Διερεύνηση control-limit πολιτικών σε προβλήματα αντικατάστασης συστήματος με δύο καταστάσεις δύο μηνύματα και δύο αποφάσεις

Περίληψη

Σε αυτό το κεφάλαιο μελετάμε μία ειδική περίπτωση του προβλήματος συντήρησης/αντικατάστασης των Ohnishi-Ibaraki που περιγράψαμε στο κεφάλαιο 6, με δύο καταστάσεις και δύο μηνύματα. Ειδικότερα μελετάμε την κλάση των control-limit πολιτικών, στην οποία ανήκει και η άριστη πολιτική με το κριτήριο βελτιστοποίησης για άπειρο χρονικό ορίζοντα και αναζητούμε συνθήκες κάτω από τις οποίες μία control-limit πολιτική επάγει Μαρκοβιανή διαμέριση στον χώρο των δ.π.

Στην ενότητα 7.1 δίνουμε τις ιδιότητες των συναρτήσεων μεταφοράς που αντιστοιχούν στα δύο μηνύματα και τις συνθήκες κάτω από τις οποίες μία control-limit πολιτική είναι πεπερασμένα μεταβατική. Επίσης παρουσιάζουμε διαγράμματα για τις διάφορες περιπτώσεις.

Στην ενότητα 7.2 μελετάμε συνθήκες κάτω από τις οποίες μία control-limit πολιτική ικανοποιεί τη συνθήκη **(A)** της ενότητας 5.3 καθώς και τις ειδικότερες συνθήκες κάτω από τις οποίες αυτή είναι περιοδική και δίνουμε αρκετά παραδείγματα περιοδικών πολιτικών.

7.1. Πεπερασμένα μεταβατικές control - limit πολιτικές

Θεωρούμε το πρόβλημα αντικατάστασης που περιγράψαμε στην ενότητα 6.1 με δύο καταστάσεις ((N=2),S={1,2}) και δύο μηνύματα ((M=2), Θ={1,2}) με βάση τις υποθέσεις $\mathbf{Y_1}$-$\mathbf{Y_5}$, που αντανακλούν τον Μαρκοβιανό χαρακτήρα της χειροτέρευσης (βλέπε παραγράφο (6.1)).

Με 1 και 2 κωδικοποιούμε την καλή (λειτουργική) και την κακή (μη λειτουργική) κατάσταση αντίστοιχα. Μήνυμα μπορεί να είναι ένα αποτέλεσμα της λειτουργίας του συστήματος, που αντανακλά την άγνωστη σε μας κατάστασή του (π.χ ποσοστό ελαττωματικών αντικειμένων, αριθμός παραγόμενων τεμαχίων ανά ώρα, κ.λ.π). Ετσι, ως μήνυμα 1 και 2 μπορούμε να κωδικοποιήσουμε χαμηλά ή υψηλά ποσοστά ελαττωματικών μονάδων που αντανακλούν τις καταστάσεις 1 και 2 αντίστοιχα.
Για να αποφύγουμε τετριμμένες περιπτώσεις, που δεν παρουσιάζουν ενδιαφέρον, περιοριζόμαστε στην τυπική περίπτωση όπου ο πίνακας μετάβασης καταστάσεων

$$P=\begin{pmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{pmatrix}$$

και ο πίνακας μηνυμάτων $R=\begin{pmatrix} r_{11} & r_{12} \\ r_{21} & r_{22} \end{pmatrix}$

i) είναι γνήσια TP_2,δηλαδή $|P|,|R|>0$.

ii) έχουν μη μηδενικά στοιχεία, δηλαδή : $p_{ij}\neq 0$, i,j=1,2,

$r_{i\theta}\neq 0$, i,θ=1,2.

Σε κάθε χρονική περίοδο επιλέγεται μια απόφαση από το σύνολο A={0,1},όπου οι κωδικοποιήσεις 0,1 είναι:

0:συνέχιση της λειτουργίας /συντήρηση του συστήματος

1:αντικατάσταση του συστήματος

Αν i είναι η κατάσταση του συστήματος και α η απόφαση, τα άμεσα κόστη c(i,α), i =1,2, α=0,1 είναι: C(1,0) = c_1, C(2,0) = c_2, C(1,1)= R_1, C(2,1)=R_2 ,για τα οποία υποθέτουμε ότι : $c_1 \le c_2$, $R_1 \le R_2$ και $c_1-R_1 \le c_2-R_2$.

Είναι βολικό να εργαζόμεθα με την δεύτερη συνιστώσα του δ.π: π = (1-p,p), όπου ,p, δηλώνει την *a priori* πιθανότητα το σύστημα να βρίσκεται στην κατάσταση 2. Παρόμοια, εργαζόμαστε με την δεύτερη συνιστώσα του a-posteriori δ.π :

$T(\pi,\theta)=(T_1(\pi,\theta), T_2(\pi,\theta))$ την οποία συμβολίζουμε με $T(p,\theta)$. Δηλαδή,

$$T(p,\theta)\equiv T_2(\pi,\theta)$$

εκφράζει την a-posteriori πιθανότητα το σύστημα στον επόμενο χρόνο (t+1) να βρεθεί στην κατάσταση 2,δοσμένου ότι στον ίδιο χρόνο (t+1) πήραμε το μήνυμα θ και ότι στον παρόντα χρόνο t επιλέξαμε την απόφαση α=0 (συνέχιση της λειτουργίας του συστήματος) και η πιθανότητα για την κατάσταση 2 είναι p.

Επίσης, {θ/p} είναι η πιθανότητα, ότι στον επόμενο χρόνο (t+1) το μήνυμα θα είναι θ, δοσμένου ότι στον παρόντα χρόνο t η πιθανότητα για την κατάσταση 2 του συστήματος είναι p και λαμβάνεται η απόφαση α=0 (συνέχιση της λειτουργίας). Θα περιοριστούμε στη μελέτη control-limit πολιτικών δ^{∞}, με συνάρτηση ελέγχου:

$$\delta(\pi)=\begin{cases}0\ (συνέχιση\ λειτουργίας)\ άν & 0\le p\le p_0.\\ 1\ (αντικατάσταση)\quad άν & p_0<p\le 1.\end{cases}$$

Είναι φανερό, ότι μια control -limit πολιτική δ^{∞} εξαρτάται αποκλειστικά από την κρίσιμη ποσότητα p_0. Η συνάρτηση κόστους V(p/δ) που αντιστοιχεί στην δ^{∞} γράφεται:

$V(p/\delta)= c_1+(c_2-c_1).p+\beta.\{1/p\}.V(T(p,1)/\delta)+ \beta.\{2/p\}.V(T(p,2)/\delta))$, αν $0\le p\le p_0$

και $V(p/\delta)=R_1+(R_2-R_1).p+ \beta.V(0/\delta)$, αν $p_0<p\le 1$.

Στην ενότητα αυτή θα εξετάσουμε συνθήκες που εξασφαλίζουν ότι μία control -limit πολιτική είναι πεπερασμένα μεταβατική (f.t). Εφαρμόζοντας τις σχέσεις (1.4.4) και (1.4.5), κατόπιν πράξεων, προκύπτει:

$$T(p,\theta)=\frac{a_\theta+\beta_\theta.p}{\gamma_\theta+\delta_\theta.p},\quad 0\le p\le 1 \qquad \underline{7.1.1}$$

$$\{\theta/p\}=\gamma_\theta+\delta_\theta.p,\quad 0\le p\le 1 \qquad \underline{7.1.2}$$

όπου

$\alpha_\theta= p_{12}.r_{2\theta}$,

$\beta_\theta=(p_{22}- p_{12}).\ r_{2\theta}=|P|.\ r_{2\theta}$

$\gamma_\theta= p_{11}.r_{1\theta}+ p_{12}.r_{2\theta}$

$\delta_\theta= p_{21}.r_{1\theta}+ p_{22}.r_{2\theta}- p_{11}.r_{1\theta}- p_{12}.r_{2\theta}=(r_{2\theta}-r_{1\theta}).|P|$,θ=1,2.

<u>7.1.3</u>

Προφανώς ισχύουν: $\alpha_\theta>0$, $\gamma_\theta>0$,θ=1,2.

Για θ=1:$\delta_1=(r_{21}- r_{22}).\ |P|=-|P|.|R|<0$ όπου είναι $-|R|=(r_{21}- r_{22})$.

Για θ=2:$\delta_2=(r_{22}- r_{12}).\ |P|=|P|.|R|>0$.

Επίσης $\beta_\theta=(p_{22}- p_{12}).r_{2\theta}=|P|.r_{2\theta}>0$,θ=1,2 επειδή:

$$|P|=\begin{vmatrix}p_{11} & p_{12}\\ p_{21} & p_{22}\end{vmatrix}=p_{22}-p_{12}>0$$

Αυτό προκύπτει διότι $p_{11}=1-p_{12}$ και $p_{21}=1-p_{22}$ ενώ ο πίνακας P είναι TP_2, επομένως η παραπάνω ορίζουσα είναι θετική.

Λήμμα 7.1.1:

i) Οι συναρτήσεις $T(p,1), T(p,2), 0 \le p \le 1$ είναι γνήσια αύξουσες

ii) Η συνάρτηση $T(p,1), 0 \le p \le 1$ είναι γνησίως κυρτή και

η συνάρτηση $T(p,2), 0 \le p \le 1$ είναι γνησίως κοίλη.

Απόδειξη

i) Η παράγωγος της $T(p,\theta)$, $0 \le p \le 1$ $(\theta=1,2)$ είναι:

$$T'(p,\theta)=\left(\frac{a_\theta+\beta_\theta.p}{\gamma_\theta+\delta_\theta.p}\right)'=\frac{\beta_\theta.\gamma_\theta-\alpha_\theta.\delta_\theta}{(\gamma_\theta+\delta_\theta.p)^2}\succ 0, 0 \le p \le 1,$$ δεδομένου ότι:

$$\beta_\theta.\gamma_\theta-\alpha_\theta.\delta_\theta=|P|r_{2\theta}.(p_{11}r_{1\theta}+p_{12}r_{2\theta})-p_{12}r_{2\theta}.(r_{2\theta}-r_{1\theta}).|P|=|P|r_{2\theta}r_{1\theta}>0$$

Επομένως η συνάρτηση $T(p,\theta), 0 \le p \le 1$ είναι γνήσια αύξουσα.

ii) Η δεύτερη παράγωγος της $T(p,\theta), 0 \le p \le 1$ $(\theta=1,2)$ είναι:

$$T''(p,\theta)=-2(\beta_\theta.\gamma_\theta-\alpha_\theta.\delta_\theta).\delta_\theta.(\gamma_\theta+\delta_\theta.p)^{-3} \quad 0 \le p \le 1$$

Επομένως $T''(p,1)>0, 0 \le p \le 1$, επειδή $\delta_1<0$.

$T''(p,2)<0, 0 \le p \le 1$, επειδή $\delta_2>0$. □

Λήμμα 7.1.2: Για τις συναρτήσεις $T(p,1)$, $T(p,2)$ ισχύουν:

i) $T(p,1)<T(p,2), 0 \le p \le 1$

ii) $T(1,\theta)<1$, $\theta=1,2$.

Απόδειξη

Αρκεί να αποδείξουμε ότι:

$$(\alpha_2+\beta_2.p).(\gamma_1+\delta_1.p)-(\alpha_1+\beta_1.p).(\gamma_2+\delta_2.p)>0, 0 \le p \le 1.$$

Λαμβάνοντας υπόψη ότι:

$$\{1/p\}+\{2/p\}=(\gamma_1+\delta_1.p)+(\gamma_2+\delta_2.p)=1,$$

το αριστερό μέλος της αποδεικτέας ανισότητας γράφεται:

$$(\alpha_2+\beta_2.p).(\gamma_1+\delta_1.p)-(\alpha_1+\beta_1.p).(\gamma_2+\delta_2.p)$$

$$=(\alpha_2+\beta_2.p).(1-(\gamma_2+\delta_2.p))-(\alpha_1+\beta_1.p).(\gamma_2+\delta_2.p)$$

$$=(\alpha_2+\beta_2.p)-(\alpha_2-\alpha_1+(\beta_2-\beta_1).p).(\gamma_2+\delta_2.p)$$

$$(p_{12}.r_{22}+|P|.r_{22}.p)-(p_{12}.(r_{22}-r_{21})+|P|.(r_{22}-r_{21}).p).(\gamma_2+\delta_2.p)$$

$$= r_{22}.(p_{12}+|P|.p)-(r_{22}-r_{21})(p_{12}+|P|.p).(\gamma_2+\delta_2.p)$$

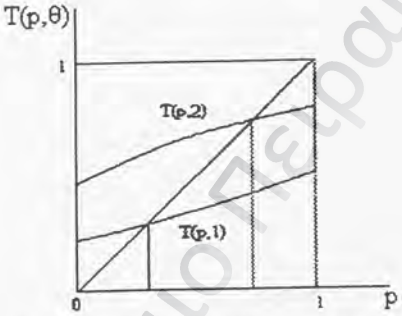
$$=(p_{12}+|P|.p).(r_{22}-(r_{22}-r_{21}).(\gamma_2+\delta_2.p))>0$$

Η τελευταία ανισότητα ισχύει επειδή:

Αν $r_{22}-r_{21}\geq 0, \tau\acute{o}\tau\varepsilon \quad r_{22}-(r_{22}-r_{21}).(\gamma_2+\delta_2.p)\geq r_{22}-(r_{22}-r_{21})=r_{21}>0$

Αν $r_{22}-r_{21}<0, \tau\acute{o}\tau\varepsilon \quad r_{22}-(r_{22}-r_{21}).(\gamma_2+\delta_2.p)\geq r_{22}>0$

ii) $T(1,\theta)=\dfrac{a_\theta+\beta_\theta}{\gamma_\theta+\delta_\theta}=\dfrac{p_{22}.r_{2\theta}}{p_{21}.r_{1\theta}+p_{22}.r_{2\theta}}<1, \qquad \theta=1,2$. □



Σχήμα 7.1: Οι συναρτήσεις T(p,θ), θ=1,2

Παρατήρηση: Το γεγονός ότι η συνάρτηση $T(p,\theta), 0\leq p\leq 1$ (θ=1,2)είναι αύξουσα είναι αναμενόμενο επειδή από την πρόταση 6.2.3 έχουμε:

$$\pi,\pi'\in\Pi, \pi\leq_L\pi'\Rightarrow T(\pi,\theta)\leq_L T(\pi',\theta).$$

Όμως για π=(1-p,p), π΄=(1-p΄,p΄) έχουμε:

$$\pi\leq_L\pi'\Leftrightarrow p\leq p' \text{ και } T(\pi,\theta)\leq_L T(\pi',\theta)\Leftrightarrow T(p,\theta)\leq T(p',\theta).$$

Αρα $p\leq p'\Rightarrow T(p,\theta)\leq T(p',\theta)$.Επίσης από την πρόταση 6.2.2 έχουμε:

$$T(\pi,1)\leq_L T(\pi,2), \forall\pi=(1-p,p)\in\Pi,$$

που ισοδυναμεί με τη σχέση:

$$T(p,1)\leq T(p,2)\ , \quad 0\leq p\leq 1$$

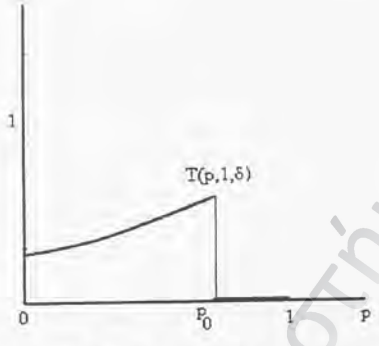
Το γεγονός ότι η παραπάνω σχέση ισχύει με γνήσια ανισότητα (λήμμα 7.1.2) και ότι η συνάρτηση $T(p,\theta), 0\leq p\leq 1$(θ=1,2) είναι γνήσια αύξουσα, οφείλεται στο γεγονός ότι υϊοθετήσαμε την τυπική περίπτωση στην οποία οι πίνακες P,R είναι γνησίως TP_2 *και έχουν* μη μηδενικά στοιχεία.΄Αλλες χρήσιμες ιδιότητες των

συναρτήσεων $T(p,\theta), 0 \le p \le 1$ (θ=1,2) θα διατυπωθούν στην πρόταση 7.1.1 που ακολουθεί.

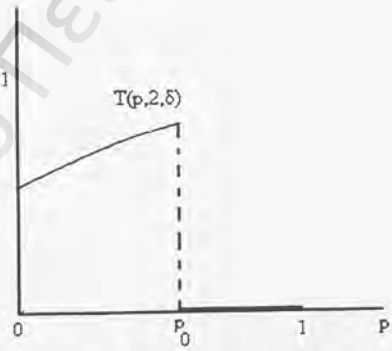
Στην συνέχεια της ενότητας αυτής θα αναφερόμαστε σε μια conrol-limit πολιτική δ^∞ με κρίσιμη πιθανότητα $p_0 \in (0,1)$. Συνδεδεμένες με την πολιτική δ^∞ είναι οι συναρτήσεις μεταφοράς $T(p,\theta,\delta), 0 \le p \le 1$ (θ=1,2), όπου T(p,θ,δ) εκφράζει την ***a-posteriori*** πιθανότητα το σύστημα στον επόμενο χρόνο (t+1) να βρεθεί στην κατάσταση 2, δοσμένου ότι στον ίδιο χρόνο (t+1) πήραμε το μήνυμα θ, και ότι στον παρόντα χρόνο t η πιθανότητα για την κατάσταση 2 είναι p και επιλέξαμε την απόφαση δ(p).Ετσι

$$T(p,\theta,\delta) = \begin{cases} T(p,\theta) & \text{άν} \quad 0 \le p \le p_0 \\ 0 & \text{άν} \quad p_0 \prec p \le 1 \end{cases}$$

Η συνάρτηση μεταφοράς T(p,θ, δ) είναι ασυνεχής στο σημείο p=p₀.



Σχήμα 7.2: Η συνάρτηση T(p,θ,δ) με θ=1



Σχήμα 7.3: Η συνάρτηση T(p,θ,δ) με θ=2

Εστω D_δ είναι το σύνολο των σημείων στα οποία η συνάρτηση ελέγχου δ παρουσιάζει ασυνέχεια.

Προφανώς $D_\delta = \{p_0\}$

Ορίζουμε τα σύνολα:

$$D^0 = D_\delta$$

$$D^n = \{p \in [0,1] : T(p,\theta,\delta) \in D^{n-1} \text{ για κάποιο } \theta \in \{1,2\}\},$$

$$n = 1,2,3\ldots$$

Το σύνολο D^n είναι το σύνολο των a-priori πιθανοτήτων για την κατάσταση 2 του συστήματος, που είναι δυνατόν σε n βήματα να μετασχηματισθούν *a posteriori* στο σημείο ασυνέχειας p_0 της συνάρτησης ελέγχου δ. Αν η πολιτική δ^∞ είναι πεπερασμένα μεταβατική ή περιοδική,(ή γενικότερα ικανοποιεί **την συνθήκη A** της

ενότητας 5.3), τότε η δ^{∞} επάγει Μαρκοβιανή διαμέριση στον χώρο Π των δ.π.. Σύμφωνα με το λήμμα 5.2.2 η πολιτική δ^{∞} είναι πεπερασμένα μεταβατική αν και μόνον αν είναι υπάρχει $n \geq 0$ έτσι ώστε: $D^n = \varnothing$.

Ο ακέραιος $n_\delta = \min\{n: D^n = \varnothing\}$ δηλώνει τον δείκτη της δ^{∞} και προφανώς $D^n = \varnothing$ $\forall n \geq n_\delta$.

Παρατήρηση: Αν η κρίσιμη πιθανότητα $p_0=1$, τότε εφαρμόζουμε την πολιτική δ^{∞}, ποτέ δεν αντικαθιστούμε το σύστημα (το αφήνουμε να λειτουργεί συνεχώς). Προφανώς η συνάρτηση ελέγχου δ είναι συνεχής και επομένως

$$D^0 = D_\delta = \varnothing.$$

Αρα η πολιτική δ^{∞} είναι πεπερασμένα μεταβατική με δείκτη $n_\delta=0$ στην περίπτωση αυτή.

Θεωρώντας $p_0 \in (0,1)$ είναι φανερό ότι:

$$D^0 = \{p_0\}$$

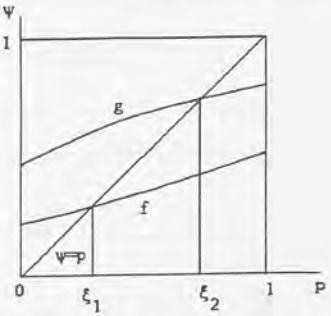
$$D^n = \{p \in [0, p_0] : T(p,\theta) \in D^{n-1} \text{ για κάποιο } \theta \in \{1,2\}\}, n = 1,2,3,\ldots$$

Για απλοποίηση των συμβολισμών θέτουμε:

$$f(p) \equiv T(p,1) = \frac{\alpha_1 + \beta_1 . p}{\gamma_1 + \delta_1 . p}, \quad 0 \leq p \leq 1$$

$$g(p) \equiv T(p,2) = \frac{\alpha_2 + \beta_2 . p}{\gamma_2 + \delta_2 . p}, \quad 0 \leq p \leq 1.$$

Στην πρόταση που ακολουθεί συγκεντρώνονται οι κυριότερες ιδιότητες των συναρτήσεων f,g / [0,1] και απεικονίζονται στο σχήμα 7.4 που ακολουθεί.



Σχήμα 7.4: Οι συναρτήσεις f,g με τα μοναδικά τους fixed-point ξ_1, ξ_2 αντίστοιχα.

Η απόδειξη ορισμένων από τις ιδιότητες αυτές στηρίζεται σε προτάσεις της στοιχειώδους ανάλυσης τις οποίες παραθέτουμε στο παράρτημα Α.

Πρόταση 7.1.1: Ιδιότητες των συναρτήσεων f,g/[0,1]

i) Οι συναρτήσεις f,g/[0,1] είναι γνήσια αύξουσες.

ii) Η συνάρτηση f /[0,1] είναι γνήσια κυρτή και η συνάρτηση g/[0,1] είναι γνήσια κοίλη.

iii) $0 < f(p) < g(p) < 1, 0 \le p \le 1$

iv) Η συνάρτηση f /[0,1] έχει μοναδικό σταθερό σημείο στο διάστημα (0,1) το οποίο συμβολίζουμε με ξ_1, δηλαδή $f(\xi_1)=\xi_1$. Επιπλέον,

$$p < f(p) < \xi_1 \quad \forall p \in [0, \xi_1), \quad \text{και} \quad \xi_1 < f(p) < p \quad \forall p \in (\xi_1, 1].$$

v) Η συνάρτηση g/[0,1] έχει μοναδικό σταθερό σημείο στο διάστημα (0,1) το οποίο συμβολίζουμε με ξ_2, δηλαδή $g(\xi_2)=\xi_2$. Επιπλέον,

$$p < g(p) < \xi_2 \quad \forall p \in [0, \xi_2), \quad \text{και} \quad \xi_2 < g(p) < p \quad \forall p \in (\xi_2, 1].$$

vi) $\xi_1 < \xi_2$.

Απόδειξη

Τα (i),(ii)και (iii) είναι επαναδιατύπωση των λημμάτων 7.1.1 και 7.1.2 με τους νέους συμβολισμούς. Τα (i.v)και(v) συνάγονται άμεσα από τα (i) και (ii),την συνέχεια των f,g/[0,1],το γεγονός ότι $0 < f(p) < 1, 0 < g(p) < 1, \forall p \in [0,1]$ και την πρόταση 2 του παραρτήματος Α.

vi) Αν $\xi_2 \le \xi_1$ τότε από (i.v) και (v) έχουμε:

$f(\xi_2) \ge \xi_2 = g(\xi_2)$ πράγμα άτοπο επειδή από (iii)συνάγεται $g(\xi_2) \succ f(\xi_2)$.

Επομένως $\xi_1 \prec \xi_2$. □

Σημειώνουμε ότι το σταθερό σημείο ξ_1 της συνάρτησης f /[0,1] προκύπτει ως η επιτρεπτή-δηλαδή η εντός του διαστήματος (0,1)-λύση της δευτεροβάθμιας εξίσωσης

$$\delta_1 \cdot \chi^2 + (\gamma_1 - \beta_1) \cdot \chi - \alpha_1 = 0$$

Παρόμοια, το σταθερό σημείο ξ_2 της συνάρτησης g/[0,1] προκύπτει ως η επιτρεπτή-δηλαδή η εντός του διαστήματος (0,1)- λύση της δευτεροβάθμιας εξίσωσης

$$\delta_2 \cdot \chi^2 + (\gamma_2 - \beta_2) \cdot \chi - \alpha_2 = 0$$

Εστω f_n /[0,1]η n-στη σύνθεση της f /[0,1] με τον εαυτό της,($f_1=f, f_2=fof$ και γενικά $f_n= f_{n-1}of$, n=2,3,…) και

g_n /[0,1] η n-στη σύνθεση της g /[0,1] με τον εαυτό της,($g_1=g, g_2=gog$ και γενικά $g_n= g_{n-1}og$, n=2,3,…).

Οι παραπάνω συνθέσεις είναι καλά ορισμένες, επειδή όπως προκύπτει άμεσα από την πρόταση 7.1.1, για τα πεδία τιμών R_f,R_g των συναρτήσεων f /[0,1],g /[0,1] ισχύει:

$$R_f=[f(0),f(1)] \subset (0,1), \qquad R_g=[g(0),g(1)] \subset (0,1).$$

Απο την πρόταση 7.1.1 (iii)συνάγεται ότι:

$$0 \prec f_n(p) \prec g_n(p) \prec 1, \ 0 \le p \le 1, \ n=1,2,...$$

Στα λήμματα 7.1.3 και 7.1.4 ανακεφαλαιώνονται οι κυριότερες ιδιότητες των συναρτήσεων f_n /[0,1] και g_n /[0,1] αντίστοιχα. Η απόδειξή τους είναι άμεση συνέπεια της πρότασης 3 του παραρτήματος Α.

Λήμμα 7.1.3: Ιδιότητες της συνάρτησης f_n /[0,1]

i) Για κάθε n= 1, 2, ... η συνάρτηση f_n /[0,1] είναι γνήσια αύξουσα, με πεδίο τιμών $[f_n(o),f_n(1)] \subset (0,1)$, και έχει σταθερό σημείο ξ_1 ,(το σταθερό σημείο της f /[0,1]).

ii) Για κάθε $p \in [0,\xi_1)$ η ακολουθία $\{f_n(p)\}$ είναι γνήσια αύξουσα,

$p \prec f_n(p) \prec \xi_1, n=1,2,...$ και $f_n(p) \nearrow \xi_1$,όταν $n \to \infty$.

Για κάθε $p \in (\xi_1,1]$ η ακολουθία $\{f_n(p)\}$ είναι γνήσια φθίνουσα,

$\xi_1 \prec f_n(p) \prec p, n=1,2,...$ και $f_n(p) \searrow \xi_1$, όταν $n \to \infty$. □

Λήμμα 7.1.4:Ιδιότητες της συνάρτησης g_n /[0,1].

Για κάθε n=1,2,...η συνάρτηση g_n /[0,1] είναι γνήσια αύξουσα, με πεδίο τιμών $[g_n(o),g_n(1)] \subset (0,1)$, και έχει σταθερό σημείο ξ_2 ,(το σταθερό σημείο της g/[0,1]).

Για κάθε $p \in [0,\xi_2)$ η ακολουθία $\{g_n(p)\}$ είναι γνήσια αύξουσα,

$p \prec g_n(p) \prec \xi_2, n=1,2,...$ και $g_n(p) \nearrow \xi_2$,όταν $n \to \infty$.

Για κάθε $p \in (\xi_2,1]$ η ακολουθία $\{g_n(p)\}$ είναι γνήσια φθίνουσα,

$\xi_2 \prec g_n(p) \prec p, n=1,2,...$ και $g_n(p) \searrow \xi_2$, όταν $n \to \infty$. □

Λήμμα 7.1.5: Έστω $\psi \in (0,1), n \ge 1$

i) Αν οι εξισώσεις $f_n(x)=\psi$, $f_{n+1}(x)=\psi$ έχουν λύσεις στο διάστημα [0,1] και τις συμβολίσουμε με x_n και x_{n+1} αντίστοιχα (δηλαδή $f_n(x_n)=\psi$, $f_{n+1}(x_{n+1})=\psi$), τότε $x_n=f(x_{n+1})$.

ii) Αν οι εξισώσεις $g_n(x)=\psi, g_{n+1}(x)=\psi$ έχουν λύσεις στο διάστημα $[0,1]$ και τις συμβολίσουμε με x_n' και x_{n+1}' αντίστοιχα (δηλαδή $g_n(x_n')=\psi$, $g_{n+1}(x_{n+1}')=\psi$), τότε $x_n'=g(x_{n+1}')$.

Απόδειξη

i) Προφανώς οι λύσεις x_n και x_{n+1} είναι μοναδικές.

Επειδή $f_{n+1}(x_{n+1})=f_n(f(x_{n+1}))=\psi$ και $f_n(x_n)=\psi$, συνάγεται ότι $x_n=f(x_{n+1})$

ii) Παρόμοια όπως το (i). □

Λήμμα 7.1.6: Εστω $\psi\in(0,1)$. Αν οι εξισώσεις $f(x)=\psi$ και $g(x)=\psi$ έχουν λύσεις στο διάστημα $[0,1]$ και τις συμβολίσουμε με x_1 και x_2 αντίστοιχα (δηλ. $f(x_1)=\psi$, $g(x_2)=\psi$), τότε $x_2<x_1$.

Απόδειξη

Προφανώς οι λύσεις x_1 και x_2 είναι μοναδικές. Αν $x_2\geq x_1$, τότε από την πρόταση 7.1.1 έχουμε $g(x_2)=\psi>f(x_2)\geq f(x_1)=\psi$, δηλαδή $\psi>\psi$, πράγμα άτοπο.

Αρα $x_2\prec x_1$. □

Εστω $p_0\in(0,1)$ η κρίσιμη ποσότητα που συνδέεται με μια control-limit πολιτική δ^∞.

Οι δυνατές περιπτώσεις πού χρήζουν διερεύνησης, λαμβάνοντας υπόψη και την πρόταση 7.1.1 είναι οι ακόλουθες:

ΠΕΡΙΠΤΩΣΗ 1

(1)	$0<p_0<\xi_1$	**με δυνατές υποπεριπτώσεις**
(1α)	$0<p_0<f(0)<\xi_1$	
(1b)	$f(0)\leq p_0<g(0)<\xi_1$	**αν** $g(0)<\xi_1$
(1c)	$g(0)\leq p_0<\xi_1$	**αν** $g(0)<\xi_1$

ΠΕΡΙΠΤΩΣΗ 2

(2)	$\xi_2<p_0<1$

ΠΕΡΙΠΤΩΣΗ 3

(3) $\xi_1 < p_0 < \xi_2$ **με δυνατές υποπεριπτώσεις**

(3α) $\xi_1 < p_0 < g(0)$ **αν** $g(0) > \xi_1$

(3b) $\xi_1 < f(p_0) < g(0) \le p_0 < g(p_0) < \xi_2$ **αν** $g(0) > \xi_1$

(3c) $\xi_1 < g(0) \le f(p_0) < p_0 < g(p_0) < \xi_2$ **αν** $g(0) > \xi_1$

(3d) $g(o) < \xi_1 < f(p_0) < p_0 < g(p_0) < \xi_2$ **αν** $g(0) < \xi_1$

ΠΕΡΙΠΤΩΣΗ 4

(4) $p_0 = \xi_1$

ΠΕΡΙΠΤΩΣΗ 5

(5) $p_0 = \xi_2$

Εισάγουμε την ακόλουθη ορολογία:

Έστω $\psi \in [o, p_o]$.

1) Λέμε ότι χ_1' είναι *άμεσος (ή πρώτης τάξης)* $f-$*απόγονος* του ψ, αν χ_1' είναι η λύση της εξίσωσης $f(\chi) = \psi$, δηλαδή $\chi_1' = f^{-1}(\psi)$, και $\chi_1' \in [o, p_o]$.

2) Λέμε ότι χ_1'' είναι *άμεσος (ή πρώτης τάξης)* $g-$*απόγονος* του ψ, αν χ_1'' είναι η λύση της εξίσωσης $g(\chi) = \psi$, δηλαδή $\chi_1'' = g^{-1}(\psi)$, και $\chi_1'' \in [o, p_o]$.

3) Λέμε ότι χ_1 είναι *άμεσος (ή πρώτης τάξης) απόγονος του* ψ αν χ_1 είναι άμεσος $f-$απόγονος ή $g-$απόγονος του ψ.

4) Λέμε ότι χ_n *είναι n-στης τάξης απόγονος του* ψ $(n \ge 2)$, αν $\chi_n \in [o, p_o]$ και υπάρχουν $\chi_1, \chi_2, \ldots, \chi_{n-1} \in [o, p_o]$, έτσι ώστε τα $\chi_1, \chi_2, \ldots, \chi_n$ να είναι άμεσοι (πρώτης τάξης) απόγονοι των $\psi, \chi_1, \chi_2, \ldots, \chi_{n-1}$ αντίστοιχα.

Σημειώνουμε, όπως προκύπτει άμεσα από την παραπάνω ορολογία, ότι αν χ_n είναι n-στης τάξης απόγονος του ψ και χ_{n-1} είναι άμεσος απόγονος του χ_n, τότε χ_{n+1} είναι n+1-τάξης απόγονος του ψ.

Με την εισαγωγή των συμβολισμών f,g /[0,1], στη θέση των T(p,1),T(p,2), $0 \le p \le 1$, τα σύνολα D^n γράφονται:

$$D^0 = \{p_0\}$$

$$D^n=\{p\in[0,p_0]: f(p)\in D^{n-1} \text{ η } g(p)\in D^{n-1}\},\ n=1,2,3,...$$

Τονίζουμε το γεγονός ότι στα παραπάνω σύνολα το διάστημα αναφοράς είναι το $[0,p_0]$.Το σύνολο D^n μπορεί να θεωρηθεί ως το σύνολο των η-στης τάξης ''απογόνων'' του p_0.

Εστω $z\in[0,p_0]$. *Ορίζουμε τα ακόλουθα σύνολα*:

$$D^0(z)=\{z\},$$

$$D^n(z)=\{p\in[0,p_0]: f(p)\in D^{n-1}(z) \text{ ή } g(p)\in D^{n-1}(z)\},\ n=1,2,3,...$$

Σημειώνουμε ότι στα παραπάνω σύνολα το διάστημα $[0,p_0]$ παραμένει ως διάστημα αναφοράς. Το σύνολο $D^n(z)$ μπορεί να θεωρηθεί ως το σύνολο των n-στης τάξης ''απογόνων'' του z. Προφανώς για $z=p_0$ έχουμε:

$$D^n(p_0)\equiv D^n,\ n=0,1,2,3,...$$

Στην πρόταση που ακολουθεί αποδεικνύουμε ότι αν $0<z<\xi_1$ και $z\le p_0$,τότε η ''γενιά''του z ''εκλείπει'' σε πεπερασμένο χρόνο.

Αυτή η ιδιότητα είναι ιδιαίτερα χρήσιμη, όπως θα διαπιστώσουμε στη συνέχεια.

Πρόταση 7.1.2: Αν $0\le z<\xi_1$, και $z\le p_0$, τότε υπάρχει φυσικός αριθμός $m\ge 1$ έτσι ώστε:

$$D^n(z)\neq\varnothing,\ 0\le n<m \text{ και } D^n(z)=\varnothing \quad \forall n\ge m.$$

Απόδειξη

Διακρίνουμε τις ακόλουθες περιπτώσεις:

1) $z<f(0)$

Τότε $z\notin[f(0),f(p_0)]$ και επομένως η εξίσωση $f(x)=z$ δεν έχει λύση στο διάστημα $[0,p_0]$. Λαμβάνοντας υπόψη ότι $f(o)<g(0)$ (πρόταση 7.1.1(iii)) έχουμε $z<g(0)$.Συνεπώς $z\notin[g(0),g(p_0)]$ και η εξίσωση $g(x)=z$ δεν έχει λύση στο διάστημα $[0,p_0]$.

Άρα στην περίπτωση αυτή το z δεν έχει απογόνους και $D^1(z)=\varnothing$ (m=1).

2) $f(o)\le z<\xi_1$

Επειδή $f_n(0)\nearrow\xi_1$ όταν $n\to\infty$ (λήμμα 7.1.3 (ii)), υπάρχει φυσικός αριθμός $n_0\ge 1$ έτσι ώστε: $f_n(o)\le z\ \forall n\le n_0$ και $f_n(0)>z\ \forall n>n_0$

Είναι φανερό ότι για κάθε $n > n_0$ η εξίσωση $f_n(x) = z$ δεν έχει λύση στο διάστημα [0,p₀] επειδή απλά το $z \notin [f_n(0), f_n(p_0)]$.

Για κάθε $n \le n_0$, λαμβάνοντας υπόψη το λήμμα 7.1.3 (ii), έχουμε:

$$f_n(0) \le z < f_n(z) < \xi_1 .$$

Επομένως $z \in [f_n(0), f_n(z)]$ και η εξίσωση $f_n(x) = z$ έχει μοναδική λύση στο διάστημα [0,z],(και κατά συνέπεια στο [0,p₀],επειδή $z \le p_0$),την οποία συμβολίζουμε με z_n ,δηλαδή:

$$f_n(z_n) = z \quad \forall n \le n_0 .$$

Αν n=1,τότε $f_1(z_1) = f(z_1) = z$

Αν $n \ge 2$,από λήμμα 7.1.5 (i) έχουμε $z_{n-1} = f(z_n)$.

Επειδή $z_n \le z < \xi_1$, έχουμε ότι $f(z_n) > z_n$ (πρόταση 7.1.1 (iv)).

Αρα $z_{n-1} > z_n$. Έτσι προκύπτει η ακόλουθη διάταξη:

$$z_{n_0} < z_{n_0-1} < \ldots < z_1 < z < \xi_1$$

Διακρίνουμε τις ακόλουθες υποπεριπτώσεις:

2α) $f(o) \le z < \min\{g(o), \xi_1\}$

Επειδή $z < g(o), έχουμε \quad z_n < z < g(o), \ 1 \le n \le n_0$.

Επομένως $z \notin [g(0), g(p_0)], z_n \notin [g(0), g(p_0)], 1 \le n \le n_0$,και οι εξισώσεις $g(x) = z$, $g(x) = z_n$,$1 \le n \le n_0$,δεν έχουν λύση στο διάστημα [0,p₀].Αυτό σημαίνει ότι δεν υπάρχουν g-απόγονοι του z και των f απογόνων του $z_1, z_2, \ldots z_{n_0}$. Συμπεραίνουμε ότι:

$$D^n(z) = \varnothing \quad \forall n \succ n_0 \ \ (m = n_0 + 1).$$

2b) $g(o) \le z \prec \xi_1$ (φυσικά υπό την προϋπόθεση ότι $g(o) < \xi_1$).

Επειδή $\xi_1 < \xi_2$ (πρόταση 7.1.1(vi)), έχουμε $z < g(z) < \xi_2$ (πρόταση 7.1.1(v)),οπότε $z \in [g(0), g(z)]$ και η εξίσωση $g(x) = z$ έχει την μοναδική λύση $\psi_1 = g^{-1}(z)$ στο διάστημα [0,z] (άρα και στο [0,p₀], επειδή $z \le p_0$).

Επομένως $D^1(z) = \{z_1, \psi_1\}$. Επειδή z_1 είναι η λύση της εξίσωσης $f(x) = z$, από το λήμμα 7.1.6 έχουμε $\psi_1 < z_1$.

Θα δείξουμε επαγωγικά ότι για οποιοδήποτε n-τάξεως απόγονο του z (όπου $n \le n_0$), $w \in D^n(z)$, ισχύει $w \le z_n$.

Ο ισχυρισμός ισχύει για n=1.As υποθέσουμε ότι ο ισχυρισμός ισχύει για n=k$< n_0$,δηλαδή:

$$w \in D^k(z) \Rightarrow w \leq z_k$$

Θα δείξουμε ότι ο ισχυρισμός ισχύει για $n = k+1 (\leq n_0)$. Η απόδειξη βασίζεται στην προφανή παρατήρηση ότι οι k+1 τάξεως απόγονοι του z προκύπτουν ως άμεσοι πρώτης τάξης απόγονοι των k τάξεως απογόνων του z.

Εστω $w \in D^k(z)$.Θεωρούμε τις εξισώσεις $f(x) = w, g(x) = w$ με λύσεις στο διάστημα [0,p_0]-αν υπάρχουν-$w' = f^{-1}(w)$ και $w'' = g^{-1}(w)$ αντίστοιχα. Με άλλα λόγια και εφόσον υπάρχουν $w' = f^{-1}(w)$ και $w'' = g^{-1}(w)$ είναι άμεσοι πρώτης τάξεως απόγονοι του w και επομένως $w', w'' \in D^{k+1}(z)$.Από το λήμμα 7.1.6 έχουμε $w'' < w'$. Εύκολα διαπιστώνουμε ότι : $w' \leq z_{k+1}$.

Πράγματι, αν $w' > z_{k+1}$,τότε $w = f(w') > f(z_{k+1}) = z_k$, που αντίκειται όμως στην υπόθεση της επαγωγής $w \leq z_k$. Επομένως $w'' \prec w' \leq z_{k+1}$ και ο ισχυρισμός ισχύει για $n = k+1 \leq n_0$.Αποδείξαμε συνεπώς ότι για $1 \leq n \leq n_0$, $w \in D^n(z) \Rightarrow w \leq z_n$.

Εστω $w \in D^{n_0}(z)$ (οπότε $w \leq z_{n_0}$). Θα δείξουμε ότι $z_{n_0} < f(o)$. Πράγματι, αν $z_{n_0} \geq f(o)$,τότε $z = f_{n_0}(z_{n_0}) \geq f_{n_0}(f(o)) = f_{n_0+1}(o)$,

δηλαδή $z \geq f_{n_0}(f(0)) = f_{n_0+1}(0)$, το οποίο αντίκειται στον ορισμό του n_0 ως $n_0 = \max\{n \in N : f_n(o) \leq z\}$. Επομένως $w \leq z_{n_0} < f(o)$.

Συμπεραίνουμε ότι $w \notin [f(0), f(p_0)], w \notin [g(0), g(p_0)]$ και οι εξισώσεις $f(x) = w$, $g(x) = w$ δεν έχουν λύσεις στο [o,p_0].Με άλλα λόγια το w δεν έχει απογόνους και συνεπώς $D^{n_0-1}(z) = \varnothing \quad (m = n_0 + 1)$. □

Πόρισμα 7.1.3: (Περίπτωση 1)

Αν $0 \prec p_0 \prec \xi_1$,τότε η control limit πολιτική δ^∞ με κρίσιμη ποσότητα p_0 είναι πεπερασμένα μεταβατική.

Απόδειξη

Από την πρόταση 7.1.2 , υπάρχει φυσικός αριθμός $m_0 \geq 1$ έτσι ώστε:

$D^n = D^n(p_0) \neq \varnothing \quad 0 \leq n < m_0$ και $D^n = D^n(p_0) = \varnothing \quad \forall n \geq m_0$. Επομένως η πολιτική δ^∞ είναι πεπερασμένα μεταβατική με δείκτη $n_\delta = m_0$. □

Πρόταση 7.1.4:(Περίπτωση 2)

Αν $\xi_2 < p_0 < 1$, τότε η control-limit πολιτική δ^∞ με κρίσιμη ποσότητα p_0 είναι πεπερασμένα μεταβατική με δείκτη $n_\delta = 1$.

Απόδειξη

Επειδή $p_0 \in (\xi_2, 1)$ έχουμε $\xi_2 < g(p_0) < p_0$ (πρόταση 7.1.1(v)).

Επομένως $p_0 \notin [g(0), g(p_0)]$ και η εξίσωση $g(x)=p_0$ δεν έχει λύση στο $[o,p_0]$. Επειδή $\xi_1 < \xi_2$ (πρόταση 7.1.1(vi)), έχουμε $p_0 \in (\xi_1, 1]$. Συνεπώς $\xi_1 < f(p_0) < p_0$ (πρόταση 7.1.1 (iv)),οπότε $p_0 \notin [f(0), f(p_0)]$ και η εξίσωση $f(x)= p_0$ δεν έχει λύση στο $[o,p_0]$. Συμπεραίνουμε λοιπόν, ότι το p_0 δεν έχει άμεσους πρώτης τάξεως απογόνους, δηλαδή $D^1 = \varnothing$. Επομένως η control-limit πολιτική δ^∞ με κρίσιμη ποσότητα p_0 είναι πεπερασμένα μεταβατική με δείκτη $n_\delta = 1$. □

Πρόταση 7.1.5:(Περίπτωση (3α))

Αν $\xi_1 \prec g(o)$ και $\xi_1 \prec p_0 \prec g(o)$, τότε η control-limit πολιτική δ^∞ με κρίσιμη ποσότητα p_0 είναι πεπερασμένα μεταβατική με δείκτη $n_\delta = 1$.

Απόδειξη

Επειδή $p_0 \notin [g(0), g(p_0)]$ η εξίσωση $g(x)=p_0$ δεν έχει λύση στο διάστημα $[o,p_0]$. Επειδή $p_0 \succ \xi_1$, έχουμε $\xi_1 < f(p_0) < p_0$ (πρόταση 7.1.1(iv)).

Επομένως, $p_0 \notin [f(0), f(p_0)]$ και η εξίσωση $f(x)= p_0$ δεν έχει λύση στο διάστημα $[o,p_0]$. Συμπεραίνουμε λοιπόν ότι το p_0 δεν έχει άμεσους πρώτης τάξεως απογόνους, δηλαδή $D^1 = \varnothing$.

Αρα, η control-limit δ^∞ με κρίσιμη ποσότητα p_0 είναι πεπερασμένα μεταβατική με δείκτη $n_\delta = 1$.

Πρόταση 7.1.6:(Περιπτώσεις (3b), (3c),(3d)).

Εστω ότι $g(0) \leq p_0$ και $\xi_1 < p_0 < \xi_2$

Αν υπάρχει φυσικός αριθμός $l \geq 1$, έτσι ώστε να ισχύει η συνθήκη

$$D^l \subset [0, \xi_1) \qquad (\Sigma)$$

Τότε η control-limit πολιτική δ^∞ με κρίσιμη ποσότητα p_0 είναι πεπερασμένα μεταβατική.

Απόδειξη

Εστω ότι ισχύει η συνθήκη (Σ) για $l \geq 1$. Αν $D^l = \varnothing$ τότε προφανώς η πολιτική δ^∞ είναι πεπερασμένα μεταβατική με δείκτη $n_\delta \leq l$.

Εστω ότι $D^l \neq \varnothing$. Προφανώς το πλήθος των l-τάξεως απογόνων του p_0 είναι πεπερασμένο; $|D^l| \leq 2^l$.

Για $z \in D^l$, επειδή $z \in [0, \xi_1)$, υπάρχει φυσικός αριθμός $m(z) \geq 1$:

$$D^n(z) \neq \varnothing, 0 \leq n < m(z) \quad \text{και} \quad D^n(z) = \varnothing \quad \forall n \geq m(z) \text{ (πρόταση 7.1.2).}$$

Επομένως η πολιτική $(\delta)^\infty$ είναι πεπερασμένα μεταβατική με δείκτη

$$n_\delta = \max\{m(z) : z \in D^l\} + l. \qquad \square$$

Ο έλεγχος ισχύος της συνθήκης (Σ) στην παραπάνω πρόταση είναι δύσκολο εγχείρημα, λόγω του πλήθους των δυνητικών απογόνων του p_0 (οι εν δυνάμει l-τάξεως απόγονοι του p_0 είναι 2^l).

Ωστόσο είναι σχετικά απλό να εξετάσουμε την ισχύ της συνθήκης (Σ) για l=1 και l=2.

Πρόταση 7.1.7:

Εστω $g(0) \leq p_0$ και $\xi_1 < p_0 < \xi_2$. Τότε

i) η συνθήκη (Σ) ισχύει για $l = 1$ αν και μόνον αν:

$$p_0 < g(\xi_1)$$

ii) Αν η συνθήκη (Σ) δεν ισχύει για $l = 1$ $(g(\xi_1) \leq p_0)$, τότε η (Σ) ισχύει για $l = 2$ αν και μόνο αν:

$$g(f(p_0)) < p_0 < g_2(\xi_1).$$

Απόδειξη

Οι δυνατές περιπτώσεις όταν $g(0) \le p_0$ και $\xi_1 \prec p_0 \prec \xi_2$ είναι οι (3b),(3c) και (3d) . Επειδή

$$p_0 \notin [f(0), f(p_0)],\ p_0 \in [g(0), g(p_0)]$$

από τις δύο εξισώσεις $f(\chi) = p_0$, $g(\chi) = p_0$ μόνο η δεύτερη έχει λύση $\psi = g^{-1}(p_0)$ στο διάστημα $[0, p_0]$. Συνεπώς ο μοναδικός άμεσος (πρώτης τάξεως) απόγονος του p_0 είναι το ψ και $D^1 = \{\psi\}$. Διαπιστώνουμε ότι $\psi \neq p_0$.

Πράγματι, επειδή $p_0 < \xi_2$ έχουμε:

$$g(\psi) = p_0 < g(p_0) \quad \text{(πρόταση 7.1.1 (v))}$$

από την οποία προκύπτει $\psi < p_0$.

i) η συνθήκη (Σ) ισχύει για $l = 1$, τότε και μόνο τότε αν:

$$\psi < \xi_1 \quad \text{η ισοδύναμα αν}$$

$$g(\psi) = p_0 < g(\xi_1)$$

ii) Αν δεν ισχύει η συνθήκη (Σ) για $l = 1$, τότε:

$$\xi_1 \le \psi < p_0.$$

Αν η εξίσωση $f(\chi) = \psi$ έχει λύση $\psi_1 = f^{-1}(\psi)$ στο διάστημα $[0, p_0]$ τότε $\psi_1 \ge \psi$.

Πράγματι επειδή $\psi \ge \xi_1$, από την πρόταση 7.1.1 (iv) έχουμε:

$f(\psi_1) = \psi \ge f(\psi)$, από την οποία προκύπτει $\psi_1 \ge \psi$.

Αν η εξίσωση $g(\chi) = \psi$ έχει λύση $\psi_2 = g^{-1}(\psi)$ στο διάστημα $[0, p_0]$ τότε $\psi_2 < \psi$.

Πράγματι επειδή $\psi < \xi_2$, από την πρόταση 7.1.1 (v) έχουμε

$g(\psi_2) = \psi < g(\psi)$, από την οποία προκύπτει $\psi_2 < \psi$.

Εξετάζοντας πότε οι παραπάνω εξισώσεις έχουν λύσεις στο διάστημα $[0, p_0]$ διακρίνουμε τις ακόλουθες περιπτώσεις:

α) $(f(o) < \xi_1 \le)\psi \le f(p_0) < g(0)$

ή $(f(o) < \xi_1 \le)\psi \le g(0) \le f(p_0)$

Επειδή $\psi \in [f(0), f(p_0)]$, $\psi \notin [g(0), g(p_0)]$ μόνο η εξίσωση $f(\chi) = \psi$ έχει λύση $\psi_1 = f^{-1}(\psi)$ στο διάστημα $[0, p_0]$. Άρα $D^2 = \{\psi_1\}$. Επειδή $\psi_1 \ge \psi$ συνάγεται ότι $\psi_1 \ge \xi_1$ και επομένως η συνθήκη (Σ) δεν ισχύει για $l = 2$.

b) $f(p_0) < \psi < g(0)$

Επειδή $\psi \notin [f(0), f(p_0)]$, $\psi \notin [g(0), g(p_0)]$ οι εξισώσεις $f(\chi) = \psi$, $g(\chi) = \psi$ δεν έχουν λύση στο διάστημα $[0, p_0]$.

Επομένως $D^2 = \varnothing$ και η συνθήκη (Σ) ισχύει τετριμμένα για $l = 2$.

Σημειώνουμε ακόμη ότι:

$$f(p_0) < \psi < g(0) \Leftrightarrow g(f(p_0)) < g(\psi) = p_0 < g_2(0).$$

c) $(f(0) <) g(0) \leq \psi \leq f(p_0) (< p_0 < g(p_0) < \xi_2)$

Επειδή $\psi \in [f(0), f(p_0)]$, $\psi \in [g(0), g(p_0)]$, οι εξισώσεις $f(\chi) = \psi$, $g(\chi) = \psi$ έχουν λύσεις $\psi_1 = f^{-1}(\psi)$, $\psi_2 = g^{-1}(\psi)$ στο διάστημα $[0, p_0]$. Άρα $D^2 = \{\psi_1, \psi_2\}$ και $\psi_2 \prec \psi \leq \psi_1$. Συνάγεται ότι $\psi_1 \geq \xi_1$ και επομένως η συνθήκη (Σ) δεν ισχύει για $l = 2$.

d) $f(p_0) \prec g(0) \leq \psi (\prec p_0 \prec g(p_0) \prec \xi_2)$

ή $g(0) \leq f(p_0) \prec \psi (\prec p_0 \prec g(p_0) \prec \xi_2)$

Επειδή $\psi \notin [f(0), f(p_0)]$, $\psi \in [g(0), g(p_0)]$, μόνο η εξίσωση $g(\chi) = \psi$ έχει λύση $\psi_2 = g^{-1}(\psi)$ στο διάστημα $[0, p_0]$. Επομένως $D^2 = \{\psi_2\}$ και $\psi_2 < \psi$. Η συνθήκη (Σ) ισχύει για $l = 2$, αν και μόνον αν $\psi_2 < \xi_1$ ή ισοδύναμα αν ισχύει: $p_0 = g_2(\psi_2) < g_2(\xi_1)$.

Σημειώνουμε επίσης ότι:

$$f(p_0) < g(0) \leq \psi \Leftrightarrow g(f(p_0)) \prec g_2(0) \leq g(\psi) = p_0$$

και

$$g(0) \leq f(p_0) \prec \psi \Leftrightarrow g_2(0) \leq g(f(p_0)) \prec g(\psi) = p_0$$

Συνοψίζοντας από τις περιπτώσεις (α),(b),(c),(d) συνάγεται ότι η συνθήκη (Σ) ισχύει για $l = 2$ τότε και μόνο τότε αν:

$$g(f(p_0)) \prec p_0 \leq g_2(\xi_1) \qquad \square$$

Ο έλεγχος της συνθήκης (Σ) για $l = 3$ επιτυγχάνεται με παρόμοιο τρόπο, όμως επισημαίνουμε ότι οι περιπτώσεις που πρέπει να εξετασθούν είναι περισσότερες. Δίνουμε το ακόλουθο αποτέλεσμα παραλείποντας την απόδειξη.

Πρόταση 7.1.8:

Εστω $g(0) \leq p_0$ και $\xi_1 < p_0 < \xi_2$. Αν η συνθήκη (Σ) δεν ισχύει για $l = 2$, τότε η (Σ) ισχύει για $l = 3$ αν και μόνον αν:

1) $p_0 \le g((f(p_0))$, $p_0 < g_2(\xi_1)$,

$g((f_2(p_0)) < p_0 < g(f(g(\xi_1)))$

ή

2) $g((f(p_0)) < p_0$, $p_0 \ge g_2(\xi_1)$,

$g_2(f(p_0)) < p_0 < g_3(\xi_1)$ □

Πρόταση 7.1.9 :*(Περίπτωση (4))*

Αν $p_0 = \xi_1$,τότε η control-limit πολιτική δ^∞ με κρίσιμη ποσότητα p_0 είναι περιοδική.

Απόδειξη

Η εξίσωση $f(\chi) = p_0$ έχει λύση στο διάστημα $[0, p_0]$ το σταθερό σημείο της f, $p_0 = \xi_1$. Επομένως

$$\xi_1 \in D^n, \text{ n} = 0,1,2,\ldots$$

Διακρίνουμε τις ακόλουθες περιπτώσεις :

i) $p_0 = \xi_1 \prec g(0)$.

Έχουμε $p_0 \notin [g(0), g(p_0)]$ και η εξίσωση $g(\chi) = p_0$ δεν έχει λύση στο διάστημα $[0, p_0]$.Επομένως $D^n = \{\xi_1\}$ ∀ n=0,1,2…και η πολιτική δ^∞ είναι τετριμμένα περιοδική.

ii) $p_0 = \xi_1 \ge g(0)$.

Επειδή $\xi_1 \prec \xi_2$ έχουμε $\xi_1 \prec g(\xi_1)$ (πρόταση 7.1.1). Αρα $p_0 \in [g(0), g(p_0)]$ και η εξίσωση $g(\chi) = p_0 = \xi_1$ έχει λύση $\psi = g^{-1}(\xi_1)$ στο διάστημα $[0, p_0]$.

Επειδή $g(\psi) = \xi_1 \prec g(\xi_1)$,συνάγεται ότι $\psi \prec \xi_1 = p_0$.

Σύμφωνα με την πρόταση 7.1.2 ,υπάρχει φυσικός αριθμός $m \ge 1$ έτσι ώστε:

$$D^n(\psi) \neq \varnothing, 0 \le n < m \quad \text{και} \quad D^n(\psi) = \varnothing, \forall n \ge m.$$

Συμπεραίνουμε ότι : $D^0 = \{\xi_1\}$

$$D^n = \{\xi_1\} \cup \bigcup_{k=0}^{n-1} D^K(\psi), \text{n=1,2,}\ldots$$

Σημειώνουμε ότι $D^n=D^m$ $\forall n \geq m$.

Επομένως η πολιτική δ^∞ είναι περιοδική. □

Πρόταση 7.1.10 :(Περίπτωση (5))

Αν $p_0 = \xi_2$,τότε η control-limit πολιτική δ^∞ με κρίσιμη ποσότητα p_0 είναι περιοδική.

Απόδειξη

Η εξίσωση $g(\chi) = p_0$ έχει λύση στο διάστημα $[0, p_0]$ το σταθερό σημείο της g, $p_0 = \xi_2$.

Επομένως

$$\xi_2 \in D^n, \text{ n} = 0,1,2,\ldots$$

Επειδή $\xi_2 = p_0 > \xi_1$ έχουμε $f(p_0) < p_0$ (πρόταση 7.1.1 (iv)).

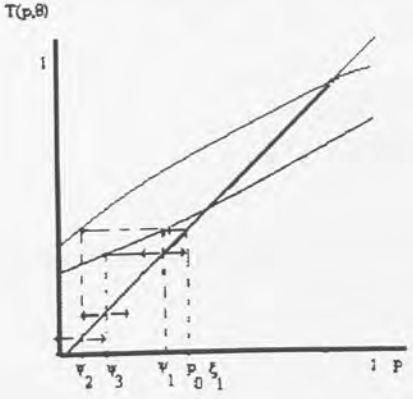
Επομένως $p_0 \notin [f(0), f(p_0)]$ και η εξίσωση $f(\chi) = p_0$ δεν έχει λύση στο διάστημα $[0, p_0]$.

Συμπεραίνουμε ότι:

$D^n = \{\xi_2\}$, n = 0,1,2,... και η πολιτική δ^∞ είναι τετριμμένα περιοδική. □

Τα διαγράμματα που ακολουθούν διαφωτίζουν τις διάφορες περιπτώσεις 1-5.Ενδεικτικά παραθέτουμε επίσης τη Μαρκοβιανή διαμέριση τη Μαρκοβιανή απεικόνιση και το διάγραμμα ροής της control limit πολιτικής για κάθε μία από τις περιπτώσεις.

Περίπτωση (1): $0 < p_0 < \xi_1$ (πρβλ. πρόταση 7.1.3)



Σχήμα 7.5

$$\psi_1 = f^{-1}(p_0) = \frac{-a_1 + \gamma_1 . p_0}{\beta_1 - \delta_1 . p_0} \qquad D^0 = \{p_0\},$$

$$\psi_2 = g^{-1}(p_0) = \frac{-a_2 + \gamma_2 . p_0}{\beta_2 - \delta_2 . p_0} \qquad D^1 = \{\psi_1, \psi_2\},$$

$$\psi_3 = f^{-1}(\psi_1) = \frac{-a_1 + \gamma_1 . \psi_1}{\beta_1 - \delta_1 . \psi_1} \qquad D^2 = \{\psi_3\},$$

$$D^3 = \varnothing, n_\delta = 3$$

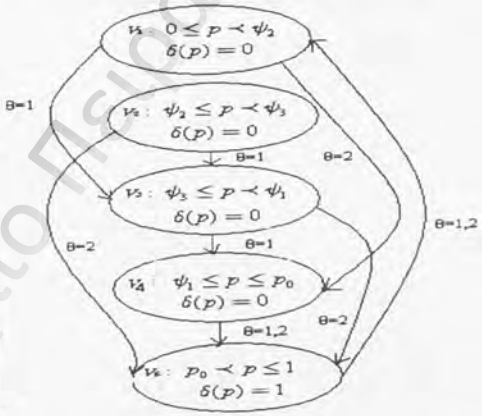
Μαρκοβιανή διαμέριση του διαστήματος [0,1]:

$V_1=[0,\psi_2), V_2=[\psi_2,\psi_3), V_3=[\psi_3,\psi_1), V_4=[\psi_1,p_0], V_5=(p_0,1)$.

Μαρκοβιανή απεικόνιση ν(j,θ)

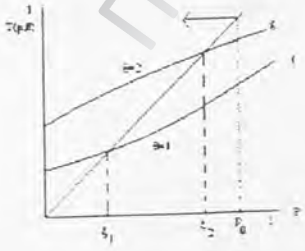
j \ θ	1	2
1	3	4
2	3	5
3	4	5
4	5	5
5	1	1

Διάγραμμα ροής για την πολιτική δ^∞



□

Περίπτωση (2): $\xi_2 \prec p_0 \prec 1$ (πρβλ. πρόταση 7.1.4).



$D^0=\{p_0\}$

$D^1 = \varnothing$, $n_\delta=1$

Μαρκοβιανή διαμέριση του διαστήματος [0,1]:

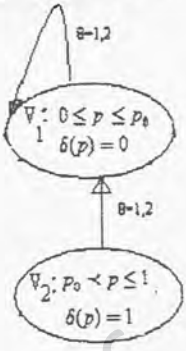
$V_1=[0, p_0]$, $V_2=(p_0,1)$

Σχήμα 7.6

Μαρκοβιανή απεικόνιση ν(j,θ)

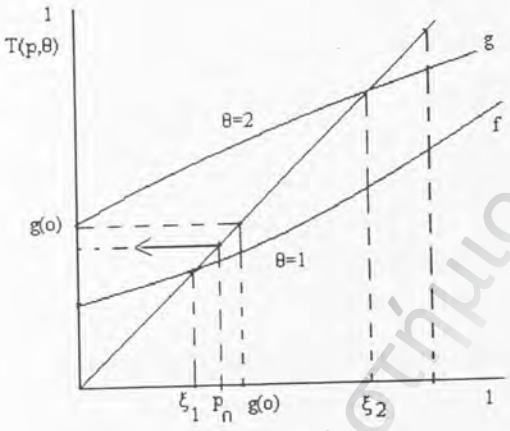
j \ θ	1	2
1	1	1
2	1	1

Διάγραμμα ροής για την πολιτική δ^∞



□

Περίπτωση (3α): $g(0) \succ \xi_1$, $\xi_1 \prec p_0 \prec g(0)$.(πρβλ. πρόταση 7.1.5).



$D^0=\{p_0\}$
$D^1=\varnothing$, $n_\delta=1$

Σχήμα 7.7

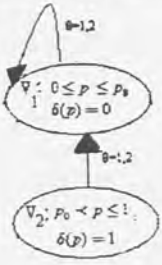
Μαρκοβιανή διαμέριση του διαστήματος [0,1]:

$V_1=[0, p_0]$, $V_2=(p_0,1)$

Μαρκοβιανή απεικόνιση ν(j,θ)

j \ θ	1	2
1	1	1
2	1	1

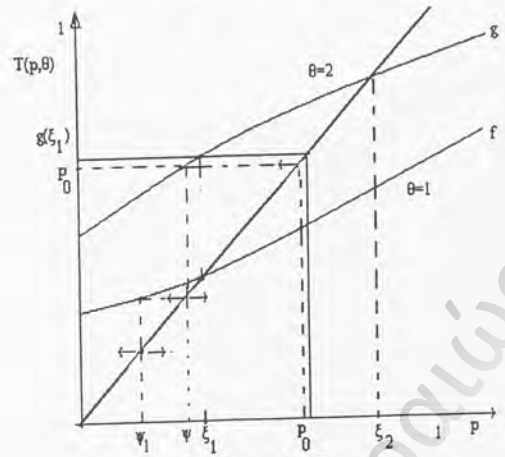
Διάγραμμα ροής για την πολιτική δ^∞



□

Πρόταση 7.1.7,περίπτωση 1: $g(0) \le p_0, \xi_1 \prec p_0 \prec \xi_2, p_0 \prec g(\xi_1)$

Η Συνθήκη (Σ) ισχύει για l=1.



Σχήμα 7.8

$$\psi = g^{-1}(p_0) = \frac{-a_2 + \gamma_2 \cdot p_0}{\beta_2 - \delta_2 \cdot p_0} < \xi_1$$

$$\psi_1 = f^{-1}(\psi) = \frac{-\alpha_1 + \gamma_1 \cdot \psi}{\beta_1 - \delta_1 \cdot \psi}$$

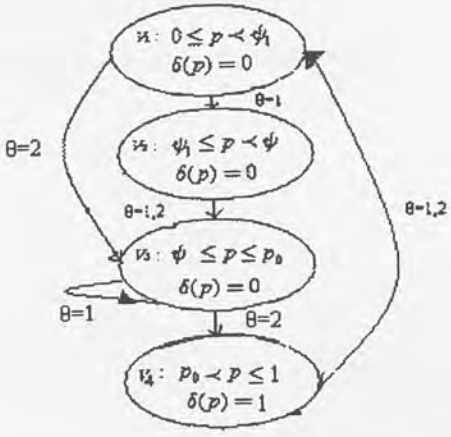
$D^0=\{p_0\}$, $D^1=\{\psi\} \subset [0,\xi_1)$, $D^2=\{\psi_1\}$, $D^3=\varnothing$, $n_\delta=3$.

Μαρκοβιανή διαμέριση του διαστήματος [0,1]:

$V_1=[0,\psi_1)$, $V_2=[\psi_1,\psi)$, $V_3=[\psi,p_0]$, $V_4=(p_0,1]$

Μαρκοβιανή απεικόνιση ν(j,θ) **Διάγραμμα ροής για την πολιτική** δ^∞

j \ θ	1	2
1	2	3
2	3	3
3	3	4
4	1	1

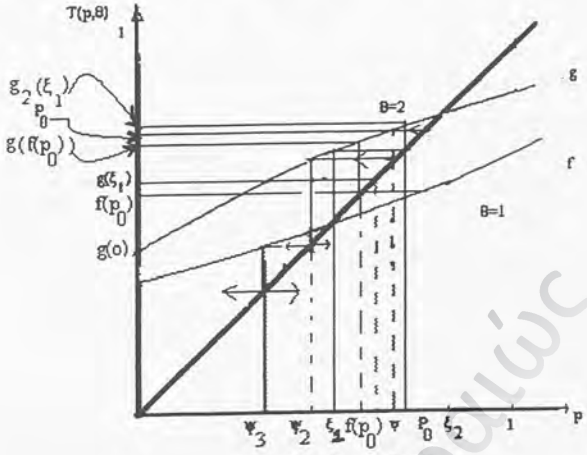


□

Πρόταση 7.1.7,περίπτωση 2: $g(0) \le p_0, \xi_1 \prec p_0 \prec \xi_2, p_0 \ge g(\xi_1)$.

Η Συνθήκη (Σ) δεν ισχύει για l=1: $g(f(p_0)) \prec p_0 \prec g_2(\xi_1)$

Η Συνθήκη (Σ) ισχύει για l=2



Σχήμα 7.9

$$\psi = g^{-1}(p_0) = \frac{-a_2 + \gamma_2 . p_0}{\beta_2 - \delta_2 . p_0}, \quad \xi_1 \le \psi < p_0$$

$$D^0=\{p_0\}, D^1 = \{\psi\}, D^2=\{\psi_2\} \subset [0, \xi_1),$$

$$\psi_2 = g^{-1}(\psi) = \frac{-a_2 + \gamma_2 . \psi}{\beta_2 - \delta_2 . \psi} \prec \xi_1 \qquad D^3 = \{\psi_3\},$$

$$\psi_3 = f^{-1}(\psi_2) = \frac{-a_1 + \gamma_1 . \psi_2}{\beta_1 - \delta_1 . \psi_2} \qquad D^4=\varnothing,\ n_\delta=4$$

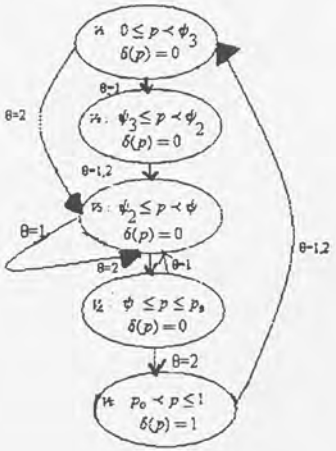
Μαρκοβιανή διαμέριση του διαστήματος [0,1]:

$$V_1=[0,\psi_3),\ V_2=[\psi_3,\psi_2),V_3=[\psi_2,\psi),V_4=[\psi,p_0],V_5=(p_0,1).$$

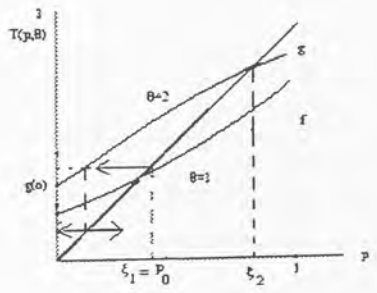
Μαρκοβιανή απεικόνιση ν(j,θ)

j \ θ	1	2
1	2	3
2	3	3
3	3	4
4	3	5
5	1	1

Διάγραμμα ροής για την πολιτική δ^∞



Περίπτωση (4) : $p_0 = \xi_1$, **(πρβλ. πρόταση 7.1.9),** $g(0) \prec \xi_1$.



$$p_0 = \xi_1 \ , \ g(0) \prec \xi_1$$
$$\psi_1 = f^{-1}(p_0) = p_0 = \xi_1$$

$$\psi_2 = g^{-1}(p_0)$$

$$D^0 = \{\xi_1\}, D^n = \{\xi_1, \psi_2\}, n = 1,2,3,...$$

Σχήμα 7.10

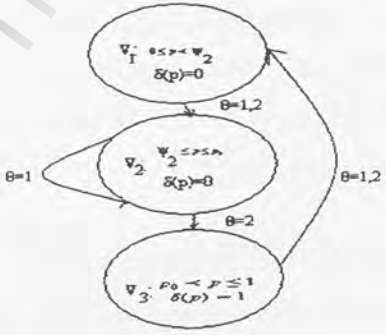
Μαρκοβιανή διαμέριση του διαστήματος [0,1]

V_1=[0,ψ_2), V_2 =[ψ_2, p_0], V_3=(p_0,1).

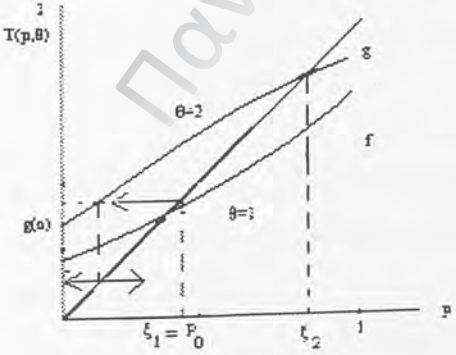
Μαρκοβιανή απεικόνιση ν(j,θ)

j \ θ	1	2
1	2	2
2	2	3
3	1	1

Διάγραμμα ροής για την πολιτική δ^∞



Περίπτωση (5): $p_0 = \xi_2$ (πρβλ. πρόταση 7.1.9).



$$\psi_1 = f^{-1}(p_0) = p_0 = \xi_1$$
$$g^{-1}(p_0) = p_0 = \xi_2$$
$$D^n = \{\xi_2\},\ n = 0,1,2,...$$

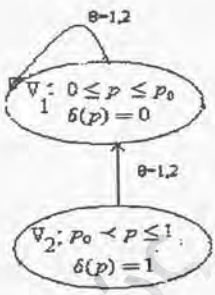
Σχήμα 7.11

Μαρκοβιανή διαμέριση του διαστήματος [0,1]

$V_1=[0,p_0]$, $V_2=(p_0,1]$.

Μαρκοβιανή απεικόνιση ν(j,θ) **Διάγραμμα ροής για την πολιτική** δ^∞

j \ θ	1	2
1	1	1
2	1	1



7.2. Περιοδικές control-limit πολιτικές

Στην ενότητα αυτή θα αναζητήσουμε control-limit πολιτικές, οι οποίες να ικανοποιούν την συνθήκη (Α) της ενότητας 5.3 και ειδικότερα περιοδικές control-limit πολιτικές πέραν των τετριμμένων περιπτώσεων όπου η κρίσιμη ποσότητα είναι $p_0=\xi_1$ ή $p_0=\xi_2$ (βλέπε προτάσεις 7.1.9, 7.1.10). Όπως θα διαπιστώσουμε στη συνέχεια, τέτοιες πολιτικές συνδέονται με τα σταθερά σημεία πεπερασμένων συνθέσεων των συναρτήσεων μεταφοράς f, g ,όπως ορίστηκαν στην ενότητα 7.1. Επομένως πρώτα θα μελετήσουμε τέτοιες συνθέσεις, τα σταθερά τους σημεία καθώς και τη διαδικασία υπολογισμού τους.

Θεωρούμε $n \ge 2$ συναρτήσεις $h_1, h_2, \ldots h_n$ ορισμένες στο διάστημα [0,1],όπου $h_i = f$ ή $h_i = g$, i=1,2,3,…,n. Επειδή τα πεδία τιμών των συναρτήσεων f, g /[0,1] περιέχονται στο διάστημα [0,1], οι συνθέσεις

$$w_k = h_1 o h_2 o \ldots o h_k \text{ /[0,1]}, \qquad 2 \le k \le n$$

είναι καλά ορισμένες και τα πεδία τιμών τους περιέχονται στο διάστημα [0,1].Σημειώνουμε ακόμα ότι:

$$w_k = w_{k-1} o h_k \text{ /[0,1]}, \qquad 2 \le k \le n.$$

Επειδή οι συναρτήσεις f, g /[0,1] είναι γνήσια αύξουσες, συνεχείς και ομογραφικές, συνάγεται εύκολα ότι και οι συναρτήσεις w_k /[0,1], $2 \le k \le n$ είναι γνήσια αύξουσες, συνεχείς και ομογραφικές:

$$w_k(\chi) = \frac{A_k + B_k \cdot \chi}{\Gamma_k + \Delta_k \cdot \chi}, \qquad 0 \le \chi \le 1$$

Επειδή

$$w_k(\chi) = w_{k-1}(h_k(\chi)) = \frac{A_{k-1} + B_{k-1} h_k(\chi)}{\Gamma_{k-1} + \Delta_{k-1} h_k(\chi)}, \qquad 0 \le \chi \le 1$$

οι παράμετροι $A_k, B_k, \Gamma_k, \Delta_k$ υπολογίζονται, όπως διαπιστώνεται εύκολα, μέσω των αναγωγικών σχέσεων:

$$A_k = A_{k-1}.\gamma_\theta + B_{k-1}.a_\theta$$

$$B_k = A_{k-1}.\delta_\theta + B_{k-1}.\beta_\theta$$

$$\Gamma_k = \Gamma_{k-1}.\gamma_\theta + \Delta_{k-1}.\alpha_\theta$$

$$\Delta_k = \Gamma_{k-1}.\delta_\theta + \Delta_{k-1}.\beta_\theta \ , \ 2 \le k \le n,$$

όπου θ=1 αν $h_k = f$ και θ=2 αν $h_k = g$,

$$A_1 = a_\theta \ , B_1 = \beta_\theta \ , \Gamma_1 = \gamma_\theta , \Delta_1 = \delta_\theta$$

όπου θ=1 αν $h_1 = f$ και θ=2 αν $h_1 = g$.

Τέλος οι συναρτήσεις w_k /[0,1], $2 \le k \le n$, ως ομογραφικές είναι κοίλες ή κυρτές. Ανακεφαλαιώνοντας, μια τυχούσα σύνθεση μήκους $n \ge 2$, w_n /[0,1], των συναρτήσεων f, g /[0,1] είναι γνήσια αύξουσα, συνεχής, ομογραφική και κοίλη ή κυρτή συνάρτηση,

$$w_n(\chi) = \frac{A_n + B_n.\chi}{\Gamma_n + \Delta_n.\chi}, \quad 0 \le \chi \le 1,$$

όπου οι παράμετροι υπολογίζονται μέσω των αναγωγικών σχέσεων που αναφέραμε προηγούμενα. Επειδή

$$0 \prec f(x) \prec g(x) \prec 1 \ , \ 0 \le \chi \le 1$$

συνάγεται ότι:

$$0 \prec f_n(x) \le w_n(\chi) \le g_n(x) \prec 1 \ , \ 0 \le \chi \le 1.$$

Επομένως $w_n(0) \succ 0, w_n(1) \prec 1$ και σύμφωνα με την πρόταση 1 του παραρτήματος Α η συνάρτηση w_n /[0,1] έχει μοναδικό σταθερό σημείο $\xi \in (0,1)$. Με άλλα λόγια η εξίσωση $w_n(\chi) = \chi$,

έχει μοναδική λύση ξ στο διάστημα (0,1) και υπολογίζεται ως η επιτρεπτή, δηλαδή η εντός του διαστήματος (0,1) λύση της δευτεροβάθμιας εξίσωσης

$$\Delta_n.\chi^2 + (\Gamma_n - B_n).\chi - A_n = 0.$$

Σύμφωνα με την πρόταση 2 του παραρτήματος Α έχουμε:

$$\chi < w_n(\chi) < \xi \qquad \forall \chi \in [0,\xi) \qquad \underline{7.2.1}$$

$$\xi < w_n(\chi) < \chi \qquad \forall \chi \in (\xi,1]. \qquad \underline{7.2.2}$$

Ειδικές περιπτώσεις τέτοιων συνθέσεων μήκους $n \geq 2$ είναι η f_n /[0,1] (n-στη σύνθεση της f) και η g_n /[0,1] (n-στη σύνθεση της g),που έχουν σταθερά σημεία ξ_1 (το σταθερό σημείο της f) και ξ_2 (το σταθερό σημείο της g) αντίστοιχα. (βλέπε και λήμματα 7.1.3, 7.1.4).

Μια σύνθεση μήκους $n \geq 2, w_n$ /[0,1] των συναρτήσεων f, g θα αναφέρεται ως μη τετριμμένη αν $w_n \neq f_n, g_n$.

Στην επόμενη πρόταση αποδεικνύεται ότι το σταθερό σημείο *μιας μη τετριμμένης* σύνθεσης των συναρτήσεων f, g ανήκει στο διάστημα (ξ_1,ξ_2).

Πρόταση 7.2.1

Θεωρούμε $n \geq 2$ συναρτήσεις $h_1, h_2, ...h_n$ ορισμένες στο διάστημα [0,1],όπου $h_i = f$ ή $h_i = g$, i=1,2,3,...,n, όχι όλες του ίδιου τύπου, και τη (μη τετριμμένη) σύνθεση

$$w_n = h_1 o h_2 o ... o h_n \text{ /}[0,1].$$

Έστω $\xi \in (0,1)$ το μοναδικό σταθερό σημείο της w_n.

Τότε

i) $\xi_1 < \xi < \xi_2$

ii) $\xi_1 < w_n(\xi_1) < \xi < w_n(\xi_2) < \xi_2$

Απόδειξη

i) Επειδή η συνάρτηση w_n /[0,1] είναι μη τετριμμένη σύνθεση μήκους n των συναρτήσεων f, g, από την πρόταση 7.1.1(iii) παίρνουμε:

$$f_n(\chi) < w_n(\chi) < g_n(\chi),\ 0 \leq \chi \leq 1$$

Αν $\xi_1 \geq \xi$,τότε από την σχέση (7.2.2) παίρνουμε

$$\xi \leq w_n(\xi_1) \leq \xi_1 = f_n(\xi_1)$$

πράγμα άτοπο, επειδή $w_n(\xi_1) \succ f_n(\xi_1)$.

Αν $\xi_2 \leq \xi$,τότε από την σχέση (7.2.1) παίρνουμε

$$g_n(\xi_2) = \xi_2 \leq w_n(\xi_2) \leq \xi,$$

πράγμα άτοπο, επειδή $g_n(\xi_2) > w_n(\xi_2)$.Επομένως $\xi \in (\xi_1,\xi_2)$.

ii) Συνάγεται άμεσα από το (i) και τις σχέσεις (7.2.1), (7.2.2). □

As θεωρήσουμε μια οποιαδήποτε μη τετριμμένη σύνθεση πεπερασμένου μήκους των συναρτήσεων f, g καθώς και τις συνθέσεις που παράγονται με κυκλικές εναλλαγές των συναρτήσεων που συμμετέχουν στην αρχική σύνθεση. Στην πρόταση που ακολουθεί παρέχονται χρήσιμες σχέσεις που συνδέουν τα σταθερά σημεία αυτών των συνθέσεων. Όπως θα δείξουμε στη συνέχεια (πρόταση 7.2.3) η control-limit πολιτική με κρίσιμη ποσότητα το μέγιστο από αυτά τα σταθερά σημεία ικανοποιεί υπό προϋποθέσεις την συνθήκη (A) της ενότητας 5.3.

Πρόταση 7.2.2:

Θεωρούμε $n \geq 2$ *συναρτήσεις* $h_1, h_2, \ldots h_n$ *ορισμένες στο διάστημα* $[0,1]$, *όπου* $h_i = f$ *ή* $h_i = g$, $i=1,2,3,\ldots,n$, *όχι όλες του ίδιου τύπου. Θεωρούμε τη συνάρτηση* $\sigma_1 /[0,1]$ *που ορίζεται ως η σύνθεση των* $h_1, h_2, \ldots h_n$ *και τις συναρτήσεις* $\sigma_2, \sigma_3, \ldots, \sigma_n /[0,1]$ *που παράγονται από τις συνθέσεις των κυκλικών εναλλαγών των* $h_1, h_2, \ldots h_n$, *δηλαδή*

$\sigma_1 = h_1 \circ h_2 \circ \ldots \circ h_n$

$\sigma_2 = h_2 \circ h_3 \circ \ldots \circ h_n \circ h_1$

.....

.....

$\sigma_n = h_n \circ h_1 \circ \ldots \circ h_{n-1}$

Αν $\chi_1, \chi_2, \ldots, \chi_n$ *είναι τα σταθερά σημεία των συναρτήσεων* $\sigma_1, \sigma_2, \sigma_3, \ldots, \sigma_n /[0,1]$ *αντίστοιχα, δηλαδή* $\sigma_i(\chi_i) = \chi_i, i = 1,2,3,\ldots n$, τότε:

i) $\xi_1 \prec \chi_i \prec \xi_2$, $i=1,2,\ldots,n$

ii) $\chi_i = h_i(\chi_{i+1}), i = 1,2,3,\ldots n,$

όπου $\chi_{n+1} \equiv \chi_1$

iii) Αν $h_i = f$ τότε $\chi_i \prec \chi_{i+1}$,

Αν $h_i = g$ τότε $\chi_i \succ \chi_{i+1}, i = 1,2,3,\ldots n.$

Απόδειξη

i)Επειδή οι συναρτήσεις $\sigma_1, \sigma_2, \sigma_3, \ldots, \sigma_n /[0,1]$ είναι μη τετριμμένες συνθέσεις μήκους n των συναρτήσεων f, g ,τα σταθερά τους σημεία $\chi_1, \chi_2, \ldots, \chi_n$ ανήκουν στο διάστημα (ξ_1, ξ_2) σύμφωνα με την πρόταση 7.2.1 (i).

ii) Για i=1, 2,3,…,n έχουμε

$$h_i(\chi_{i+1}) = h_i(\sigma_{i+1}(\chi_{i+1})) = (h_i \circ \sigma_{i+1})(\chi_{i+1}) = (h_i \circ h_{i+1} \circ ... \circ h_n \circ h_1 \circ ..h_i)(\chi_{i+1}) =$$

$$= (h_i \circ h_{i+1} \circ ... \circ h_n \circ h_1 \circ ..h_{i-1})(h_i(\chi_{i+1})) = \sigma_i(h_i(\chi_{i+1})).$$

Επομένως $h_i(\chi_{i+1})$ είναι σταθερό σημείο της συνάρτησης σ_i /[0,1] και επειδή αυτό είναι μοναδικό συμπεραίνουμε ότι $\chi_i = h_i(\chi_{i+1})$.

iii) Για i =1, 2, 3,…,n έχουμε:

Αν $h_i = f$ τότε από το (ii), $\chi_i = f(\chi_{i+1})$.

Επειδή $\chi_{i+1} > \xi_1$ έχουμε $\chi_i = f(\chi_{i+1}) < \chi_{i+1}$ (πρόταση 7.1.1 (iv)).

Αν $h_i = g$ τότε από το (ii), $\chi_i = g(\chi_{i+1})$.

Επειδή $\chi_{i+1} < \xi_2$ έχουμε $\chi_i = g(\chi_{i+1}) > \chi_{i+1}$ (πρόταση 7.1.1 (v)). □

Παρατηρήσεις

1)Η πρόταση 7.2.2 παρέχει έναν απλό τρόπο υπολογισμού των σταθερών σημείων των συναρτήσεων $\sigma_1, \sigma_2, \sigma_3, ..., \sigma_n$ /[0,1], όπως αυτές ορίσθηκαν στην ίδια πρόταση. Αφού υπολογίσουμε το σταθερό σημείο κάποιας από τις παραπάνω συναρτήσεις με τη διαδικασία που περιγράψαμε στην αρχή της ενότητας, για τα υπόλοιπα $n-1$ σταθερά σημεία μπορούμε να αποφύγουμε αυτή τη διαδικασία υπολογισμού, και να εφαρμόσουμε απλές αναγωγικές σχέσεις που υπαγορεύονται από το τμήμα (ii) της πρότασης 7.2.2. Συγκεκριμένα, αν για παράδειγμα έχουμε υπολογίσει το σταθερό σημείο χ_1 της συνάρτησης σ_1 με τη γνωστή διαδικασία, τότε τα υπόλοιπα σταθερά σημεία $\chi_2, \chi_3, ..., \chi_n$ των συναρτήσεων $\sigma_2, \sigma_3, ..., \sigma_n$ υπολογίζονται μέσω των αναγωγικών σχέσεων

$$\chi_{i+1} = h_i^{-1}(\chi_i) = \frac{-a_\theta + \gamma_\theta \cdot \chi_i}{\beta_\theta - \delta_\theta \cdot \chi_i},$$

όπου $\theta = 1$ αν $h_i = f$ και $\theta = 2$ αν $h_i = g$, i =1, 2, 3,…,n-1 .

2)Αν η διάταξη των συναρτήσεων f, g που συμμετέχουν σε μία σύνθεση παρουσιάζει περιοδικότητα, τότε προφανώς το σταθερό σημείο αυτής της σύνθεσης ταυτίζεται με το σταθερό σημείο της σύνθεσης του περιοδικού τμήματος (δηλαδή του τμήματος της διάταξης που επαναλαμβάνεται).

Για παράδειγμα, το σταθερό σημείο της σύνθεσης

$$f \circ g \circ f \circ f \circ g \circ f \circ f \circ g \circ f \circ f \circ g \circ f,$$

ταυτίζεται με το σταθερό σημείο της σύνθεσης $f \circ g \circ f$.

Επίσης αν η διάταξη των f,g στη σύνθεση σ_1 είναι περιοδική, τότε και οι διατάξεις των f,g στις συνθέσεις $\sigma_2,\sigma_3,...,\sigma_n$ -που προκύπτουν με κυκλική εναλλαγή-παρουσιάζουν και αυτές περιοδικότητα. Από τα παραπάνω γίνεται φανερό ότι το πρόβλημα σχετικά με τα σταθερά σημεία συνθέσεων των συναρτήσεων f,g απλοποιείται, αν περιοριστούμε σε συνθέσεις διατάξεων των f,g που δεν παρουσιάζουν περιοδικότητα.

Πρόταση 7.2.3:

Θεωρούμε $n \geq 2$ *συναρτήσεις* $h_1,h_2,...h_n$ *ορισμένες στο διάστημα [0,1],όπου* $h_i = f$ *ή* $h_i = g, i=1,2,3,...,n$, *όχι όλες του ίδιου τύπου, έτσι ώστε η διάταξη να μην παρουσιάζει περιοδικότητα. Θεωρούμε τη συνάρτηση* $\sigma_1/[0,1]$ *πού ορίζεται ως η σύνθεση των* $h_1,h_2,...h_n$ *και τις συναρτήσεις* $\sigma_2,\sigma_3,...,\sigma_n/[0,1]$ *πού παράγονται από τις συνθέσεις των κυκλικών εναλλαγών των* $h_1,h_2,...h_n$, *(όπως στην πρόταση 7.2.2).Εστω* $\chi_1,\chi_2,...,\chi_n$ *τα σταθερά σημεία των συναρτήσεων* $\sigma_1,\sigma_2,\sigma_3,...,\sigma_n/[0,1]$. *Θέτουμε* $p_0 = \max\{\chi_1,\chi_2,...,\chi_n\}$ *και θεωρούμε τη control-limit πολιτική* δ^∞ *με κρίσιμη ποσότητα* p_0. *Τότε:*

i) Για i =1, 2, 3,…,n έχουμε:

Αν $h_i = f$,τότε το χ_{i+1} είναι άμεσος $f-$απόγονος του χ_i και $\chi_{i+1} \succ \chi_i$.$(\chi_{n+1} \equiv \chi_1)$.

Αν $h_i = g$, τότε το χ_{i+1} είναι άμεσος $g-$απόγονος του χ_i και $\chi_{i+1} \prec \chi_i$.

ii)Τα $\chi_2,...,\chi_n,\chi_1$ είναι άμεσοι απόγονοι των $\chi_1,\chi_2,...,\chi_n$ αντίστοιχα. Θα αναφερόμαστε σε αυτά ως άμεσους περιοδικούς απογόνους των $\chi_1,\chi_2,...,\chi_n$.

iii) Αν ισχύει η συνθήκη **(Σ΄)**, τότε η πολιτική δ^∞ *ικανοποιεί τη συνθήκη* **(Α)** της ενότητας 5.3. Η συνθήκη **(Σ΄)** διατυπώνεται ως εξής:

Για i =1, 2, 3,…,n, σε περίπτωση που το σταθερό σημείο χ_i έχει-πέραν του άμεσου περιοδικού απογόνου χ_{i+1}-άμεσο μη περιοδικό απόγονο, έστω ψ_i, υπάρχει $I_i \geq 1$ έτσι ώστε:

$$D^{I_i}(\psi_i) = \varnothing \qquad (\Sigma')$$

Ειδικότερα, αν κανένα από τα $\chi_1,\chi_2,...,\chi_n$ δεν έχει άμεσο μη περιοδικό απόγονο, τότε η πολιτική δ^∞ είναι περιοδική.

Απόδειξη

i) Από την πρόταση 7.2.2 (i), (ii) και την επιλογή της κρίσιμης ποσότητας p_0 ως το μέγιστο των σταθερών σημείων $\chi_1, \chi_2, \ldots, \chi_n$, έχουμε:

Για i =1, 2, 3,…,n,

$$\chi_{i+1} = h_i^{-1}(\chi_i) \in (\xi_1, p_0].$$

Συμπεραίνουμε ότι αν $h_i = f$, τότε το χ_{i+1} είναι άμεσος $f-$απόγονος του χ_i, ενώ αν $h_i = g$, τότε το χ_{i+1} είναι άμεσος g-απόγονος του χ_i. Στην πρώτη περίπτωση έχουμε $\chi_{i+1} \succ \chi_i$, ενώ στην δεύτερη $\chi_{i+1} \prec \chi_i$ (πρόταση 7.2.2)(iii)).

ii) Άμεση συνέπεια του (i).

iii) Θεωρούμε τα σύνολα

$$\bar{D}^m = \bigcup_{k=0}^{m} D^k \text{ ,m=0,1,2,…}$$

Για $m \geq 1$, το σύνολο, $\bar{D}^m$, δηλώνει το σύνολο των απογόνων του p_0 τάξεως μικρότερης ή ίσης του m.

Αν ισχύει η συνθήκη (Σ΄),τότε σε συνδυασμό με το (ii) συνάγεται ότι:

$$\bar{D}^m = \bar{D}^l \quad \forall m \geq l$$

όπου $$l := n + \max_{1 \leq i \leq n}\{l_i\}$$

(Αν το χ_i δεν έχει άμεσο μη περιοδικό απόγονο, θέτουμε $l_i = 0$.)

Άρα η πολιτική δ^∞ ικανοποιεί τη συνθήκη **(Α)** της ενότητας 5.3. Σημειώνουμε ότι η δ^∞ αποκλείεται να είναι πεπερασμένα μεταβατική επειδή το (ii) συνεπάγεται $D^m \neq \varnothing \quad \forall m = 0, 1, 2, \ldots$

Θεωρούμε τώρα την ειδική περίπτωση όπου κανένα από τα $\chi_1, \chi_2, \ldots, \chi_n$ δεν έχει μη περιοδικό απόγονο. Θα δείξουμε ότι η δ^∞ είναι περιοδική. Πράγματι, έστω

$$p_0 = \chi_j = \max\{\chi_1, \chi_2, \ldots, \chi_n\}.$$

Τότε από το (ii) συνάγεται ότι

$$D^m = \{p_0\} = \{\chi_j\} \text{ , } m \equiv 0 (\mathrm{mod}\, n)$$

$$D^m = \{\chi_{j+1}\} \text{ , } m \equiv 1 (\mathrm{mod}\, n)$$

……

$$D^m = \{\chi_{j+n-1}\} \text{ , } m \equiv n-1 (\mathrm{mod}\, n),$$

όπου $\chi_{n+k} \equiv \chi_k$, $1 \le k \le n-1$

Άρα η δ^∞ είναι περιοδική. □

Παρατηρήσεις

1)Επειδή $\xi_1 \prec \chi_i \prec \xi_2$, i =1, 2, 3,…,n,(πρόταση 7.2.2 (i)) και η κρίσιμη ποσότητα της πολιτικής δ^∞ είναι $p_0 = \max\{\chi_1, \chi_2, ..., \chi_n\}$ έχουμε $\xi_1 < p_0 < \xi_2$.

2)Το τμήμα (iii) της πρότασης 7.2.3 δηλώνει ότι αν οι «γενιές» των άμεσων μη περιοδικών απογόνων των σταθερών σημείων $\chi_1, \chi_2, ..., \chi_n$ -εφόσον υπάρχουν τέτοιοι – εκλείπουν σε πεπερασμένο αριθμό χρονικών περιόδων (Συνθήκη (Σ΄)), τότε η πολιτική δ^∞ υπακούει στη συνθήκη **(A)** της ενότητας 5.3 και επομένως επάγει Μαρκοβιανή διαμέριση στο χώρο Π (βλ. πρόταση 5.3.4)

3) As θεωρήσουμε ότι ο άμεσος περιοδικός απόγονος χ_{i+1} του χ_i είναι τύπου f ,δηλαδή $\chi_{i+1} = f^{-1}(\chi_i)$.

Από την πρόταση 7.2.3 (i) έχουμε $\chi_{i+1} > \chi_i$.Επομένως $\chi_i < p_0$.Προφανώς ο δυνάμει άμεσος μη περιοδικός απόγονος του χ_i είναι τύπου g. Επομένως για να διαπιστώσουμε αν υπάρχει άμεσος μη περιοδικός απόγονος του χ_i, εξετάζουμε αν η λύση $g^{-1}(\chi_i)$ της εξίσωσης $g(\chi) = \chi_i$ ανήκει στο διάστημα [0,p_0] ή ισοδύναμα, αν το $\chi_i \in [g(0), g(p_o)]$.

Επειδή $p_0 < \xi_2$, από την **πρόταση** 7.1.1(v) παίρνουμε:

$$\chi_i < p_0 < g(p_0)$$

Άρα το χ_i έχει άμεσο μη περιοδικό απόγονο $\psi_i \equiv g^{-1}(\chi_i)$ αν $\chi_i \ge g(0)$.

Αποδεικνύεται εύκολα το ακόλουθο:

Αν $\chi_i < g(\xi_1)$ τότε το χ_i ικανοποιεί τη συνθήκη (Σ΄).

Πράγματι η σχέση $\chi_i < g(\xi_1)$ ισοδυναμεί με τη σχέση $g^{-1}(\chi_i) < \xi_1$.

- Αν $g^{-1}(\chi_i) < 0$,τότε το χ_i δεν έχει άμεσο μη περιοδικό απόγονο και επομένως ικανοποιεί τετριμμένα την (Σ΄).
- Αν $0 < g^{-1}(\chi_i) < \xi_1 (< p_0)$, τότε το χ_i έχει άμεσο μη περιοδικό απόγονο $\psi_i = g^{-1}(\chi_i)$. Επειδή $\psi_i \in [o, \xi_1)$, σύμφωνα με την πρόταση 7.1.2, υπάρχει $l_i \ge 1$ έτσι ώστε $D^{l_i}(\psi_i) = \varnothing$.

4)Ας θεωρήσουμε ότι ο άμεσος περιοδικός απόγονος χ_{i+1} του χ_i είναι τύπου g, δηλαδή $\chi_{i+1}=g^{-1}(\chi_i)$. Τότε προφανώς ο δυνάμει άμεσος μη περιοδικός απόγονος του χ_i είναι τύπου f. Επομένως για να διαπιστώσουμε αν υπάρχει άμεσος μη περιοδικός απόγονος του χ_i, εξετάζουμε αν η λύση $f^{-1}(\chi_i)$ της εξίσωσης $f(\chi)=\chi_i$ ανήκει στο διάστημα $[0,p_0]$ή ισοδύναμα αν το χ_iανήκει στο διάστημα $[f(0),f(p_o)]$.

Επειδή $\chi_i > \xi_1 > f(0)$, συμπεραίνουμε ότι το χ_i έχει άμεσο μη περιοδικό απόγονο αν $\chi_i \le f(p_0)$.

Αν $\chi_i \succ f(p_0)$ τότε το χ_i δεν έχει άμεσο περιοδικό απόγονο και επομένως ικανοποιεί τετριμμένα τη **(Σ΄)**.

5)Ας θεωρήσουμε ότι χ_j είναι το μέγιστο από τα σταθερά σημεία $\chi_1,\chi_2,...,\chi_n$, οπότε η κρίσιμη ποσότητα της πολιτικής δ^∞ είναι $p_0=\chi_j$.Τότε

- Ο άμεσος περιοδικός απόγονος χ_{j+1} του p_0 είναι τύπου g.

Πράγματι, αν χ_{j+1} ήταν f-απόγονος του p_0,τότε θα είχαμε $\chi_{j+1} > \chi_j = p_0$,πράγμα άτοπο.

- Το p_0 δεν έχει άμεσο μη περιοδικό απόγονο.

Πράγματι, ο δυνάμει μη περιοδικός απόγονος του p_0 είναι τύπου f.

Επειδή $p_0 > \xi_1$, από την πρόταση 7.1.1 (iv) παίρνουμε $p_0 > f(p_0)$.

Επομένως $p_0 \notin [f(0),f(p_0)]$ και η εξίσωση $f(\chi)=p_0$ δεν έχει λύση στο διάστημα $[0,p_0]$.Έτσι το p_0 δεν έχει άμεσο μη περιοδικό απόγονο και συμπεραίνουμε ότι το p_0 ικανοποιεί τετριμμένα τη **(Σ΄)**.

Παραδείγματα

Στα παραδείγματα που ακολουθούν θεωρούμε τον πίνακα μετάβασης καταστάσεων P και τον πίνακα μηνυμάτων R,που αντιστοιχούν στην απόφαση α=0 (συνέχιση της λειτουργίας /συντήρησης του συστήματος):

$$P=\begin{pmatrix}0.8 & 0.2\\ 0.3 & 0.7\end{pmatrix},\ R=\begin{pmatrix}0.8 & 0.2\\ 0.4 & 0.6\end{pmatrix}$$

Εφαρμόζοντας τις σχέσεις (7.1.1), (7.1.3) παίρνουμε:

$f(p)=T(p,1)=\dfrac{2+5p}{18-5p}$, $0\le p\le 1$, γνήσια αύξουσα, κυρτή με πεδίο τιμών $[0.1111, 0.5385]$ και σταθερό σημείο $\xi_1 = 0.16421833$.

$g(p)=T(p,2)=\dfrac{6+15p}{14+10p}$, $0\le p\le 1$, γνήσια αύξουσα, κοίλη με πεδίο τιμών $[0.4286, 0.8750]$ και σταθερό σημείο $\xi_2 = 0.826208734$.

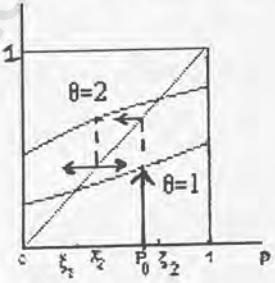
Παράδειγμα 7.2.1: Για $h_1 = g, h_2 = f$ παίρνουμε:

$\sigma_1(p)=g(f(p))=\dfrac{138+45p}{272-20p}$, $0\le p\le 1, \uparrow$, κυρτή, με πεδίο τιμών $[0.5074, 0.7262]$ και σταθερό σημείο $\chi_1 = 0.64453034$.

$\sigma_2(p)=f(g(p))=\dfrac{58+95p}{222+105p}$, $0\le p\le 1, \uparrow$, κοίλη, με πεδίο τιμών $[0.2613, 0.4679]$ και σταθερό σημείο $\chi_2 = 0.353422792$. $p_0 = \max\{\chi_1, \chi_2\} = \chi_1$.

Επειδή $\chi_2 = h_1^{-1}(\chi_1) = g^{-1}(p_0)$, $p_0 = \chi_1 = h_2^{-1}(\chi_2) = f^{-1}(\chi_2)$,
το χ_2 είναι άμεσος περιοδικός g-απόγονος του p_0 και το p_0 είναι άμεσος περιοδικός f-απόγονος του χ_2.
Επειδή $p_0 > f(p_0) = \chi_2, \chi_2 < g(0) = 0.4286$,
το p_0 δεν έχει άμεσο (μη περιοδικό) f-απόγονο και το χ_2 δεν έχει άμεσο (μη περιοδικό) g-απόγονο. □

$D^n = \{p_0\}, n \equiv 0.(\mathrm{mod}\,2)$
$D^n = \{\chi_2\}, n \equiv 1.(\mathrm{mod}\,2)$



Σχήμα 7.12: Διάγραμμα παραδείγματος 7.2.1

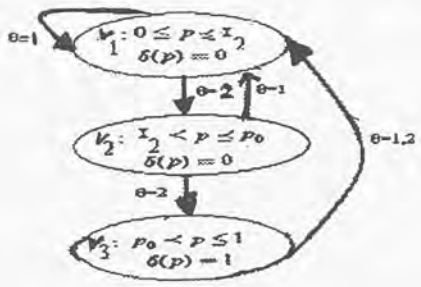
Η πολιτική δ^∞ με κρίσιμη ποσότητα p_0 είναι περιοδική.

Μαρκοβιανή διαμέριση του διαστήματος [0,1]

$V_1=[0, \chi_2]$, $V_2=(\chi_2, p_0]$, $V_3=(p_0, 1]$

Μαρκοβιανή απεικόνιση ν(j,θ) **Διάγραμμα ροής για την πολιτική δ^{∞}**

j \ θ	1	2
1	1	2
2	1	3
3	1	1



Παράδειγμα 7.2.2: Για $h_1 = g, h_2 = g, h_3 = f$ παίρνουμε:

$$\sigma_1(p) = g_2(f(p)) = \frac{3702 + 555.p}{5188 + 170p},\ 0 \le p \le 1, \uparrow, \text{κοίλη},$$

με πεδίο τιμών [0.7136,0.7945] και σταθερό σημείο $\chi_1 = 0.776903019$.

$$\sigma_2(p) = g(f(g(p))) = \frac{2202 + 2055.p}{3688 + 2420p},\ 0 \le p \le 1, \uparrow, \text{κοίλη},$$

με πεδίο τιμών [0.5971,0.6970] και σταθερό σημείο $\chi_2 = 0.674410542$.

$$\sigma_3(p) = f(g_2(p)) = \frac{1382 + 2005.p}{3738 + 3795p},\ 0 \le p \le 1, \uparrow, \text{κοίλη},$$

με πεδίο τιμών [0.3697,0.4496] και σταθερό σημείο $\chi_3 = 0.416883666$.

$p_0 = \max\{\chi_1, \chi_2, \chi_3\} = \chi_1$.

Επειδή

$\chi_2 = h_1^{-1}(\chi_1) = g^{-1}(p_0), \chi_3 = h_2^{-1}(\chi_2) = g^{-1}(\chi_2), p_0 = \chi_1 = h_3^{-1}(\chi_3) = f^{-1}(\chi_3)$, τα χ_2, χ_3 είναι άμεσοι περιοδικοί g-απόγονοι των p_0, χ_2 αντίστοιχα, και το p_0 είναι άμεσος περιοδικός f-απόγονος του χ_3.

Επειδή

$$p_0 > f(p_0) = \chi_3, \chi_2 > f(p_0) = \chi_3, \chi_3 < g(0) = 0.4286,$$

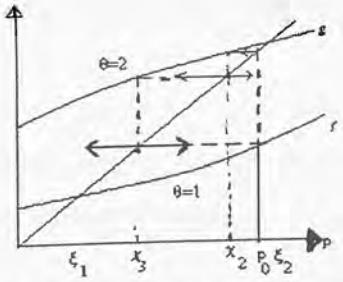
τα p_0, χ_2 δεν έχουν άμεσο (μη περιοδικό) f-απόγονο, και το χ_3 δεν έχει άμεσο (μη περιοδικό) g-απόγονο.

$$D^n = \{p_0\}, n \equiv 0 (\text{mod}\, 3)$$

$$D^n = \{\chi_2\}, n \equiv 1 (\text{mod}\, 3)$$

$$D^n = \{\chi_3\}, n = 2 (\text{mod}\, 3)$$

Η πολιτική δ^{∞} με κρίσιμη ποσότητα p_0 είναι περιοδική.



Σχήμα7.13: διάγραμμα παραδείγματος 7.2.2

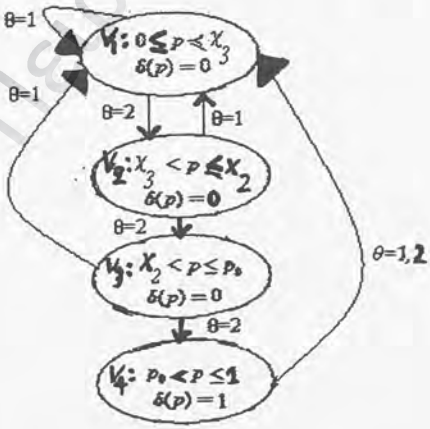
Μαρκοβιανή διαμέριση του διαστήματος [0,1]

$V_1=[0,\chi_3]$, $V_2=(\chi_3,\chi_2]$, $V_3=(\chi_2,p_0]$, $V_4=(p_0,1]$

Μαρκοβιανή απεικόνιση ν(j,θ)

j \ θ	1	2
1	1	2
2	1	3
3	1	4
4	1	1

Διάγραμμα ροής για την πολιτική δ^∞



Παράδειγμα 7.2.3: Για $h_1=g, h_2=f, h_3=f$ παίρνουμε:

$$\sigma_1(p)=g(f_2(p))=\frac{2574-465.p}{4856-1460p},\ 0\le p\le 1,\uparrow,\text{κυρτή},$$

με πεδίο τιμών [0.5301,0.6210] και σταθερό σημείο $\chi_1=0.574214278$.

$$\sigma_2(p)=f_2(g(p))=\frac{734+685.p}{3706+1415.p},\ 0\le p\le 1,\uparrow,\text{κοίλη},$$

με πεδίο τιμών [0.1981,0.2771] και σταθερό σημείο $\chi_2=0.220245338$.

$$\sigma_3(p)=f\,(g(f(p)))=\frac{1234+185.p}{4206-585.p},\ 0\le p\le 1,\uparrow,\text{κυρτή},$$

με πεδίο τιμών [0.2934,0.3919] και σταθερό σημείο $\chi_3=0.321970677$.

$$p_0 = \max\{\chi_1, \chi_2, \chi_3\} = \chi_1.$$

Επειδή

$$\chi_2 = h_1^{-1}(\chi_1) = g^{-1}(p_0),\ \chi_3 = h_2^{-1}(\chi_2) = f^{-1}(\chi_2),\ p_0 = \chi_1 = h_3^{-1}(\chi_3) = f^{-1}(\chi_3),$$

το χ_2 είναι άμεσος περιοδικός g-απόγονος του p_0, ενώ τα χ_3, p_0 είναι άμεσοι περιοδικοί f-απόγονοι των χ_2, χ_3 αντίστοιχα.

Επειδή

$$p_0 > f(p_0) = \chi_3, \chi_2 < g(0),\ , \chi_3 < g(0)\ \ (g(0) = 0.4286),$$

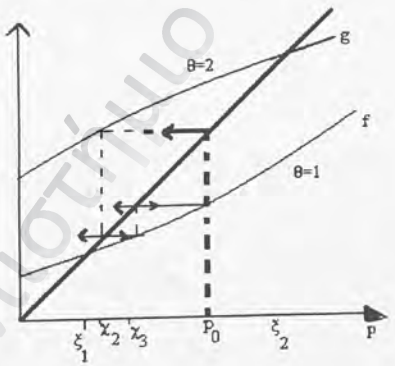
συμπεραίνουμε ότι το p_0 δεν έχει άμεσο (μη περιοδικό) f-απόγονο, και τα χ_2, χ_3 δεν έχουν άμεσο (μη περιοδικό) g-απόγονο.

$D^n = \{p_0\}, n \equiv 0(\mathrm{mod}\,3)$

$D^n = \{\chi_2\}, n = 1(\mathrm{mod}\,3)$

$D^n = \{\chi_3\}, n = 2(\mathrm{mod}\,3)$

Η πολιτική δ^∞ με κρίσιμη ποσότητα p_0 είναι περιοδική.



Σχήμα 7.14 : Διάγραμμα παραδείγματος 7.2.3

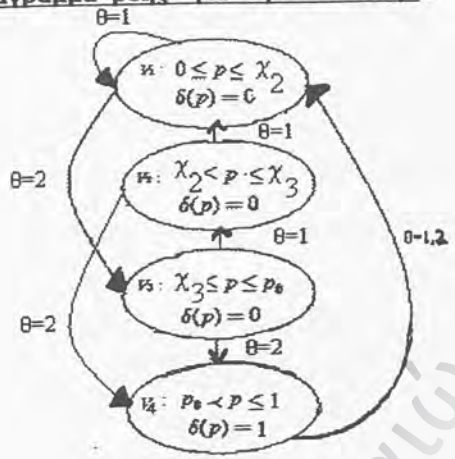
Μαρκοβιανή διαμέριση του διαστήματος [0,1]:

V_1=[0, χ_2], V_2 =(χ_2 , χ_3] ,V_3 =(χ_3 , p_0], V_4 =(p_0,1]

Μαρκοβιανή απεικόνιση ν(j,θ)

j \ θ	1	2
1	1	3
2	1	4
3	2	4
4	1	1

Διάγραμμα ροής για την πολιτική δ^{∞}



Παράδειγμα 7.2.4:

Για $h_1 = g, h_2 = g, h_3 = f, h_4 = f$, παίρνουμε:

$$\sigma_1(p) = g_2(f_2(p)) = \frac{67746 - 15735.p}{93724 - 25090.p},\ 0 \le p \le 1, \uparrow, \text{κυρτή},$$

με πεδίο τιμών [0.7228,0.7578]

και σταθερό σημείο $\chi_1 = 0.746730107$.

$$\sigma_2(p) = g(f_2(g(p)) = \frac{33246 + 18765.p}{59224 + 26660.p},\ 0 \le p \le 1, \uparrow, \text{κοίλη},$$

με πεδίο τιμών [0.5614,0.6056]

και σταθερό σημείο $\chi_2 = 0.591318139$.

$$\sigma_3(p) = f_2(g_2(p)) = \frac{14386 + 17615.p}{60374 + 58285.p},\ 0 \le p \le 1, \uparrow, \text{κοίλη},$$

με πεδίο τιμών [0.2383,0.2697] και σταθερό σημείο $\chi_3 = 0.250742757$.

$$\sigma_4(p) = f\,(g_2(f(p)) = \frac{28886 + 3115.p}{74874 + 285.p},\ 0 \le p \le 1, \uparrow, \text{κοίλη},$$

με πεδίο τιμών [0.3858,0.4258] και σταθερό σημείο $\chi_4 = 0.401900328$

$p_0 = \max\{\chi_1, \chi_2, \chi_3, \chi_4\} = \chi_1$.

Επειδή

$\chi_2 = h_1^{-1}(\chi_1) = g^{-1}(p_0), \chi_3 = h_2^{-1}(\chi_2) = g^{-1}(\chi_2), \chi_4 = h_3^{-1}(\chi_3) = f^{-1}(\chi_3),$

$p_0 = \chi_1 = h_4^{-1}(\chi_4) = f^{-1}(\chi_4),$

τα χ_2, χ_3 είναι άμεσοι περιοδικοί g-απόγονοι των p_0, χ_2 αντίστοιχα ενώ τα χ_4, p_0 είναι άμεσοι περιοδικοί f-απόγονοι των χ_3, χ_4 αντίστοιχα.

Επειδή

$$p_0 \succ f(p_0) = \chi_4, \chi_2 \succ f(p_0) = \chi_4, \ , \chi_3 \prec g(0) \ , \chi_4 \prec g(0)$$

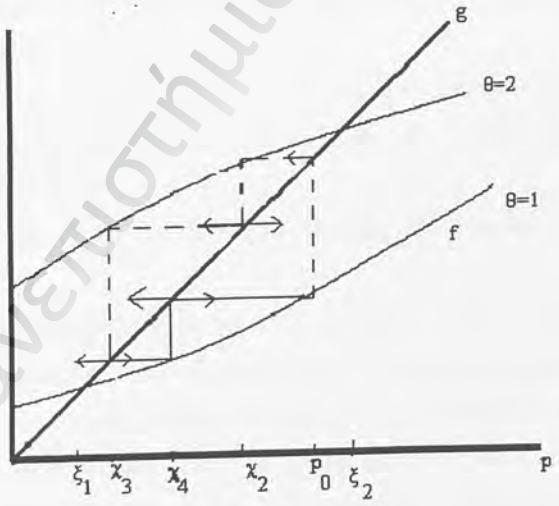
(g(0)=0.4286),συμπεραίνουμε ότι τα p_0, χ_2 δεν έχουν άμεσο (μη περιοδικό) f-απόγονο καθώς επίσης τα χ_3, χ_4 δεν έχουν άμεσο (μη περιοδικό) g-απόγονο.

$$D^n = \{p_0\}, n \equiv 0(\text{mod}\, 4)$$
$$D^n = \{\chi_2\}, n \equiv 1(\text{mod}\, 4)$$
$$D^n = \{\chi_3\}, n \equiv 2(\text{mod}\, 4)$$
$$D^n = \{\chi_4\}, n \equiv 3(\text{mod}\, 4)$$

Η πολιτική δ^∞ με κρίσιμη ποσότητα p_0 είναι περιοδική.

Μαρκοβιανή διαμέριση του διαστήματος [0,1]:

V_1=[0, χ_3], V_2 =(χ_3 , χ_4], V_3 =(χ_4 , χ_2], V_4 =(χ_2 , p_0], V_5=(p_0,1]

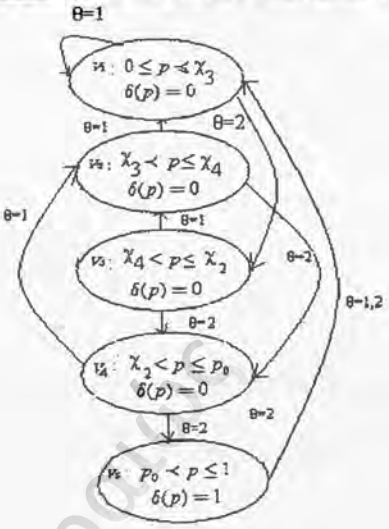


Σχήμα 7.15 : Διάγραμμα παραδείγματος 7.2.3

Μαρκοβιανή απεικόνιση ν(j,θ)

j \ θ	1	2
1	1	3
2	1	4
3	2	4
4	2	5
5	1	1

Διάγραμμα ροής για την πολιτική δ^∞



ΣΥΜΠΕΡΑΣΜΑΤΑ

Στο κεφάλαιο αυτό μελετήσαμε την κλάση των control limit (λογικών) πολιτικών (στην οποία ανήκει και η άριστη πολιτική με το κριτήριο βελτιστοποίησης για άπειρο χρονικό ορίζοντα)σε ένα πρόβλημα αντικατάστασης συστήματος με δύο καταστάσεις, δύο μηνύματα και δύο αποφάσεις .Ειδικότερα

- Διαπιστώσαμε ότι οι συναρτήσεις μεταφοράς που αντιστοιχούν στα δύο μηνύματα είναι ομογραφικές, δώσαμε τις ιδιότητές τους και δείξαμε ότι έχουν μοναδικά σταθερά σημεία.
- Εξετάσαμε τους "απογόνους" της κρίσιμης ποσότητας μιας control-limit πολιτικής.
- Παρουσιάσαμε συνθήκες κάτω από τις οποίες μια control limit πολιτική είναι πεπερασμένα μεταβατική καθώς επίσης και διαγράμματα για κάθε περίπτωση.
- Διαπιστώσαμε ότι πεπερασμένες συνθέσεις των συναρτήσεων μεταφοράς είναι επίσης ομογραφικές συναρτήσεις και έχουν μοναδικά σταθερά σημεία.
- Μελετήσαμε κάτω από ποιες συνθήκες το μέγιστο των σταθερών σημείων, που αντιστοιχούν σε κυκλικές εναλλαγές μιας πεπερασμένης σύνθεσης συναρτήσεων μεταφοράς αποτελεί κρίσιμη ποσότητα μιας control- limit πολιτικής που ικανοποιεί τη συνθήκη **(Α)** της ενότητας 5.3 καθώς και τις ειδικές συνθήκες κάτω από τις οποίες η πολιτική αυτή είναι περιοδική .
- Παρουσιάσαμε αριθμητικά παραδείγματα περιοδικών control- limit πολιτικών .

ΚΕΦΑΛΑΙΟ 8

Το πρόβλημα της αντικατάστασης συστήματος με το κριτήριο του μέσου κόστους ανά μονάδα χρόνου.

Περίληψη

Στο κεφάλαιο αυτό θα ασχοληθούμε με προβλήματα συντήρησης /αντικατάστασης συστήματος,όπου θα χρησιμοποιήσουμε σαν κριτήριο για την βελτιστοποίηση, το κριτήριο του μέσου κόστους ανά μονάδα χρόνου.Το κεφάλαιο οργανώνεται ως εξής:

Στην ενότητα 8.1 ορίζουμε το κριτήριο του μέσου κόστους για προβλήματα POMDP και παρουσιάζουμε την αντίστοιχη εξίσωση αριστοποίησης (average-cost-optimality equation,ACOE). Εξετάζουμε πως συνδέεται το κριτήριο του μέσου κόστους με το κριτήριο του αναμενόμενου ολικού εκπίπτοντος κόστους για άπειρο χρονικό ορίζοντα.

Στην ενότητα 8.2 μελετάμε το γενικό πρόβλημα αντικατάστασης συστήματος των Ohnishi-Ibaraki, που περιγράψαμε στο κεφάλαιο 6, με το κριτήριο του μέσου κόστους και αποδεικνύουμε την ύπαρξη ΜΚ-άριστης πολιτικής αντικατάστασης.

Τέλος στην ενότητα 8.3 εξετάζουμε το ειδικό πρόβλημα αντικατάστασης συστήματος με δύο καταστάσεις και δύο μηνύματα που περιγράψαμε στο κεφάλαιο 7 και μελετάμε τη δομή της ΜΚ-άριστης πολιτικής.

8.1. Το κριτήριο του μέσου κόστους ανά μονάδα χρόνου.

Στην ενότητα αυτή εξετάζουμε προβλήματα POMDP με το κριτήριο του μέσου κόστους ανά μονάδα χρόνου (average cost per unit time).

Έστω D η κλάση όλων των πολιτικών (βλέπε ενότητα 1.5).

Ορισμός 8.1.1: Ορίζουμε μέσο κόστος $J(\delta,\pi_o)$ ανά μονάδα χρόνου του συστήματος για την πολιτική $\delta \in D$, αν το αρχικό δ.π είναι $\pi(\pi(0)=\pi)$, την ποσότητα

$$J(\delta,\pi)= \lim_{n\to\infty} \sup \frac{1}{n} E_\delta [\sum_{n=0}^{n-1} C(X_t,Y_t)/\pi(o)=\pi] \qquad \underline{8.1.1}$$

Ορισμός 8.1.2: Μια πολιτική δ^*, καλείται άριστη ως προς το κριτήριο του μέσου κόστους, σύντομα MK-άριστη, αν

$$J(\delta^*,\pi)= J(\pi)\ ,\ \forall \pi\in\Pi.$$

Ελάχιστο μέσο κόστος: $J(\pi) \equiv \inf_{\delta\in D} J(\delta,\pi)\ ,\pi\in\Pi.$

Έστω $V_\beta(\pi),\pi\in\Pi$ η βέλτιστη (ελάχιστη) συνάρτηση του αναμενόμενου ολικού εκπίπτοντος κόστους σε άπειρο χρονικό ορίζοντα (βέλτιστη συνάρτηση τιμών για άπειρο χρονικό ορίζοντα), όπου $\beta\in(0,1)$, είναι ο συντελεστής έκπτωσης. Η εξίσωση αριστοποίησης του Blackwell γράφεται

$$V_\beta(\pi) = \min_\alpha \{\pi.C^\alpha + \beta.\sum_\theta \{\theta/\pi,\alpha\} V_\beta(T(\pi,\theta,\alpha))\}, \forall \pi\in\Pi\ . \qquad \underline{8.1.2}$$

Θα αναφερόμαστε στην άριστη πολιτική με το κριτήριο του ολικού εκπίπτοντος κόστους για άπειρο χρονικό ορίζοντα ως β-άριστη (β-optimal), δηλώνοντας την εξάρτηση από τον συντελεστή εκπτώσεως β.

Θεώρημα 8.1.1: Αν υπάρχουν μια φραγμένη πραγματική συνάρτηση $h(\pi), \pi\in\Pi$ και μια σταθερά g, έτσι ώστε:

$$g + h(\pi)= \min_\alpha \{\pi.C^\alpha + \sum_\theta \{\theta/\pi,\alpha\} h(T(\pi,\theta,\alpha))\}\ ,\ \forall \pi\in\Pi, \qquad \underline{8.1.3}$$

τότε η στάσιμη πολιτική $(\delta^*)^\infty$ με συνάρτηση ελέγχου που ελαχιστοποιεί το δεύτερο σκέλος της (8.1.3), είναι MK -άριστη και

$$g = J(\delta^*,\pi) = J(\pi) \qquad \forall \pi \in \Pi \ .$$

(Ross)[108]

□

Η εξίσωση (8.1.3) καλείται εξίσωση αριστοποίησης για το μέσο κόστος ανά μονάδα χρόνου (average-cost optimality equation, ACOE).

Συνοψίζοντας, σύμφωνα με το θεώρημα 8.1.1, αν η ACOE έχει φραγμένη λύση, τότε το άριστο (ελάχιστο) κόστος ανά μονάδα χρόνου δεν εξαρτάται από το αρχικό δ.π.- είναι η σταθερή ποσότητα g- και υπάρχει στάσιμη MK-άριστη πολιτική.

Εστω

$$V_n(\pi) := \min_a \{\pi.C^a + \sum_\theta \{\theta/\pi,\alpha\} V_{n-1}(T(\pi,\theta,\alpha))\} \ ,\pi \in \Pi \quad n=1,2,3\ldots$$

$$V_0(\pi)=0$$

Η συνάρτηση $V_n(\pi)$, $\pi \in \Pi$ είναι η συνάρτηση του ελάχιστου αναμενόμενου ολικού κόστους του συστήματος για χρονικό ορίζοντα μήκους n, όταν το αρχικό δ.π είναι το π ,(με συντελεστή έκπτωσης β=1).

Πρόταση 8.1.1: Αν η ACOE έχει φραγμένη λύση (g,h),τότε υπάρχει σταθερά $K \prec \infty$,έτσι ώστε :

$$|V_n(\pi) - ng| \prec K \qquad \forall n=1,2,\ldots, \forall \pi \in \Pi$$

και

$$\lim_{n\to\infty} \frac{V_n(\pi)}{n} = g \qquad \forall \pi \in \Pi \ .$$

(Ross)[108]

□

Θεωρούμε τώρα $\pi^* \in \Pi$

$$h_\beta(\pi) := V_\beta(\pi) - V_\beta(\pi^*) \quad \text{και} \quad g_{\beta=}(1-\beta).V_\beta(\pi^*), \ 0 \prec \beta \prec 1,$$

8.1.4

τότε η εξίσωση αριστοποίησης του Blackwell (8.1.2) για το ολικό κόστος γράφεται ως εξής:

$$g_\beta + h_\beta(\pi) = \min_a \{\pi.C^a + \beta.\sum_\theta \{\theta/\pi,\alpha\} h_\beta(T(\pi,\theta,\alpha))\} \quad \forall \pi \in \Pi . \qquad \underline{8.1.5}$$

Εξετάζουμε τώρα τις κατάλληλες συνθήκες ώστε:

$$g_\beta \to g \quad \text{και} \quad h_\beta(\pi) \to h(\pi) \quad \text{όταν το} \quad \beta \to 1^- .$$

Μπορεί να δειχθεί, ότι αναγκαία συνθήκη για την ύπαρξη φραγμένης λύσης της ACOE είναι ή συνθήκη (UB) (uniform-boundeness) (βλέπε και Ross [108]).

(UB) Υπάρχει μια σταθερά $K \succ 0$ έτσι ώστε:

$$|h_\beta(\pi)| \equiv |V_\beta(\pi) - V_\beta(\pi^*)| \le K, \ \forall \pi \in \Pi \quad \text{και} \quad 0 \prec \beta \prec 1,$$

Περιορίζουμε τον χώρο Π σε ένα κατάλληλο υποσύνολο. Προς τούτο έστω $\pi_0 \in \Pi$. Θεωρούμε το σύνολο :

$$S(\pi_0) = \bigcup_{t=0}^{\infty} S_t(\pi_0) \quad \text{μέ} \quad S_0(\pi_0) = \{\pi_0\}$$

$$S_t(\pi_0) = \{T(\pi,\theta,\alpha): \pi \in S_{t-1}(\pi_0), \theta \in \Theta, \alpha \in A\}, \ t \ge 1 \qquad \underline{8.1.6}$$

Το $S_t(\pi_0)$ εκφράζει το σύνολο των δ.π που είναι δυνατόν να προκύψουν στον χρόνο t,αν το δ.π στον χρόνο t=0 είναι π_0 και είναι προφανώς πεπερασμένο.

Το $S(\pi_0)$ εκφράζει το σύνολο των δ.π που είναι δυνατόν να προκύψουν στους χρόνους t=0,1,2,...,αν στον χρόνο t=0 το δ.π είναι π_0.

Το σύνολο $S(\pi_0)$ είναι αριθμήσιμο σαν αριθμήσιμη ένωση πεπερασμένων συνόλων.

Πρόταση 8.1.2: Έστω $\pi_0 \in \Pi$. Αν η συνάρτηση

$$h_\beta(\pi) = V_\beta(\pi) - V_\beta(\pi^*), \pi \in S(\pi_0),$$

είναι ομαλά φραγμένη **(UB) :** Υπάρχει μια σταθερά $K \succ 0$ έτσι ώστε:

$$|h_\beta(\pi)| \equiv |V_\beta(\pi) - V_\beta(\pi^*)| \le K \quad \forall \pi \in S(\pi_0), \ \forall \beta \in (0,1)$$

τότε,

i)Υπάρχει ακολουθία $\{\beta_n\}$ με $\beta_n \in (0,1)$και $\beta_n \to 1$ αν $n \to \infty$,μια φραγμένη συνάρτηση $h(\pi)$, $\pi \in S(\pi_0)$,και μια σταθερά g_{π_0},έτσι ώστε:

$$h_{\beta_n}(\pi) \underset{n\to\infty}{\to} h(\pi) \qquad \forall \pi \in S(\pi_0),$$

$$g_{\beta_n} \xrightarrow[n\to\infty]{} g_{\pi_0}$$

ii)Η σταθερά g_{π_0} και η συνάρτηση $h(\pi), \pi \in S(\pi_0)$ ικανοποιεί την εξίσωση αριστοποίησης για το μέσο κόστος.

$$g_{\pi_0} + h(\pi) = \min_{a} \{\pi.q^a + \sum_{\theta} \{\theta / \pi, \alpha\} h(T(\pi, \theta, \alpha))\} \qquad \forall \pi \in S(\pi_0)$$

iii) $g_{\pi_0} = J(\pi) \quad \forall \pi \in S(\pi_0)$.

(απόδειξη Ross [108])

□

Πρόταση 8.1.3: Αν η συνάρτηση $h_\beta(\pi)$, $\pi \in \Pi$ είναι ομαλά φραγμένη **(UB)**, τότε για δεδομένο δ.π. π' η ΜΚ-άριστη απόφαση είναι α', αν υπάρχει ακολουθία $\{\beta_n\}$ με $\beta_n \in (0,1)$ $\forall n \in N$ και $\beta_n \xrightarrow[n\to\infty]{} 1$, έτσι ώστε η απόφαση α' να είναι β_n –άριστη στο π'.

(Απόδειξη Arapostathis and Fernandez) [4]

□

8.2. Το πρόβλημα της αντικατάστασης συστήματος στα πλαίσια της μερικής διάταξης $\leq_L$ με το κριτήριο του μέσου κόστους ανά μονάδα χρόνου.

Στην παρούσα ενότητα εξετάζουμε το πρόβλημα αντικατάστασης συστήματος των Ohnishi-Ibaraki ,που περιγράψαμε στο κεφάλαιο 6,με το κριτήριο του μέσου κόστους ανά μονάδα χρόνου, κάτω από τις ίδιες υποθέσεις Y1-Y5 (βλέπε ενότητα 6.1).

Η άριστη (ελάχιστη) συνάρτηση τιμών για άπειρο χρονικό ορίζοντα $\mathbf{V_\beta}$ ικανοποιεί την εξίσωση αριστοποίησης (βλέπε σχέση 6.1.5)

$$V_\beta(\pi) = \min\{\pi.C^k + \beta.\sum_{\theta} \{\theta / \pi\} V_\beta(T(\pi, \theta)), \pi.C^R + \beta V_\beta(e_1)\}, \qquad \underline{8.2.1}$$

όπου $e_1 = (1, 0, 0, \ldots, 0)$.

Λήμμα 8.2.1: Η συνάρτηση $h_\beta(\pi) = \mathbf{V_\beta(\pi) - V_\beta(e_1)}$ είναι ομαλά φραγμένη.

Απόδειξη

Από την εξίσωση αριστοποίησης (8.2.1) παίρνουμε:

$$V_\beta(\pi) \le \pi.C^R + \beta.V_\beta(e_1), \ \forall \pi \in \Pi. \qquad \underline{8.2.2}$$

Επειδή σύμφωνα με την πρόταση 6.2.2 η συνάρτηση V_β είναι $\le_L$ αύξουσα και $e_1 \le_L \pi$, $\forall \pi \in \Pi$, παίρνουμε:

$$V_\beta(e_1) \le V_\beta(\pi) \quad \forall \pi \in \Pi. \qquad \underline{8.2.3}$$

Επειδή $\pi \le_L e_N \equiv (0,\ldots,0,1) \ \forall \pi \in \Pi$ και $C^R \in F^N$ (υπόθεση Υ3, ενότητα 6.1) προκύπτει ότι : $\pi.C^R \le e_N.C^R = C^R_N$

Λαμβάνοντας υπόψη τις (8.2.2) και (8.2.3) παίρνουμε:

$$V_\beta(e_1) \le V_\beta(\pi) \le C^R_N + V_\beta(e_1) \qquad \forall \pi \in \Pi, \ \forall \beta \in (0,1)$$

η $$0 \le V_\beta(\pi) - V_\beta(e_1) \le C^R_N \qquad \forall \pi \in \Pi, \ \forall \beta \in (0,1)$$

και επομένως η συνάρτηση :

$h_\beta(\pi) = V_\beta(\pi) - V_\beta(e_1)$ είναι ομαλά φραγμένη. □

Τα σύνολα $S_t(\pi_0)$, $t \ge 0$, $S(\pi_0)$, $\pi_0 \in \Pi$ που ορίσθηκαν με την (8.1.6) γράφονται:

$S_0(\pi_0) = \{\pi_0\}$

$S_t(\pi_0) = \{T(\pi,\theta) : \pi \in S_{t-1}(\pi_0), \theta \in \Theta\} \cup \{e_1\} \quad t \ge 1.$

$S(\pi_0) = \bigcup_{t=0}^{\infty} S_t(\pi_0)$.

Προφανώς $S(e_1) \subseteq S(\pi_0) \quad \forall \pi_0 \in \Pi.$

Πρόταση 8.2.2: Με βάση τις υποθέσεις Υ1-Υ5, (βλέπε ενότητα 6.1) υπάρχει μια φραγμένη συνάρτηση $h(\pi), \pi \in \Pi$, και μια σταθερά g, έτσι ώστε να ικανοποιούν την εξίσωση αριστοποίησης για το μέσο κόστος, δηλαδή

$$g + h(\pi) = \min_{\alpha} \{\pi.C^{\kappa} + \sum_{\theta} \{\theta / \pi\} h(T(\pi,\theta)\}, \pi.C^R\} \qquad \underline{8.2.4}$$

$$h(e_1) = 0$$

και

$$g = J(\pi) \ \forall \pi \in \Pi.$$

Απόδειξη

Άμεση συνέπεια της πρότασης (8.1.2) και του λήμματος (8.2.1).Επειδή μάλιστα $S(e_1) \subseteq S(\pi)$ καί $g_{\pi_0} = J(e_1)$ $\forall\ \pi_0 \in \Pi$. Αρα g_{π_0} είναι ανεξάρτητο του αρχικού δ.π. π_0.

□

8.3.Το πρόβλημα της αντικατάστασης με δυο καταστάσεις και δύο μηνύματα με βάση το κριτήριο του μέσου κόστους ανά μονάδα χρόνου.

Στην παράγραφο αυτή θα εξετάσουμε τη δομή της άριστης πολιτικής με βάση το κριτήριο του μέσου κόστους ανά μονάδα χρόνου, στο πρόβλημα αντικατάστασης ενός συστήματος με δύο καταστάσεις ,S={1,2}(N=2), δύο μηνύματα, Θ={1,2}(M=2) και δύο αποφάσεις, A={0,1},το οποίο μελετήσαμε στο κεφάλαιο 7 και αποτελεί ειδική περίπτωση του προβλήματος αντικατάστασης που εξετάσαμε στο κεφάλαιο 6..

Με 1 και 2 κωδικοποιούμε την καλή (λειτουργική) και την κακή (μη λειτουργική) κατάσταση αντίστοιχα.

Μήνυμα μπορεί να είναι ένα αποτέλεσμα της λειτουργίας του συστήματος, που αντανακλά την άγνωστη σε μας κατάστασή του (π.χ ποσοστό ελαττωματικών αντικειμένων, αριθμός παραγόμενων τεμαχίων ανά ώρα, κ.λ.π). Ετσι, ως μήνυμα 1 και 2 μπορούμε να κωδικοποιήσουμε χαμηλά ή υψηλά ποσοστά ελαττωματικών μονάδων πού αντανακλούν τις καταστάσεις 1 και 2 αντίστοιχα.

Σε κάθε χρονική περίοδο επιλέγεται μια απόφαση από το σύνολο A={0,1},όπου οι κωδικοποιήσεις 0,1 είναι:

0:συνέχιση της λειτουργίας /συντήρηση του συστήματος

1: αντικατάσταση του συστήματος

Αν i είναι η κατάσταση του συστήματος και α η απόφαση, τα άμεσα κόστη c(i,α), i =1,2, α=0,1 είναι: $C(1,0) = c_1$, $C(2,0) = c_2$, $C(1,1)=C(2,1)=R$.

Τα c_1,c_2 είναι κόστη λειτουργίας-συντήρησης και το R είναι κόστος αντικατάστασης του συστήματος.

Στην απόφαση α=0 αντιστοιχούν ο πίνακας μετάβασης καταστάσεων

$$P=\begin{pmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{pmatrix}$$

και ο πίνακας μηνυμάτων

$$R=\begin{pmatrix} r_{11} & r_{12} \\ r_{21} & r_{22} \end{pmatrix}$$

Υποθέσεις

1) Οι πίνακες P,R είναι ολικά θετικοί τάξεως 2 (TP_2) και $p_{12}>0$.

2) $c_1 < c_2 < R$

Σημειώνουμε ότι οι υποθέσεις (1),(2) συνεπάγονται τις υποθέσεις (Y1)-(Y5) του γενικότερου προβλήματος αντικατάστασης του κεφαλαίου 6.Επομένως τα συμπεράσματα για το γενικότερο πρόβλημα ισχύουν και για το ειδικό πρόβλημα που μελετάμε. Όπως και στο κεφάλαιο 7, θα εργαζόμαστε με τη δεύτερη συνιστώσα του δ.π, π=(1-p,p) όπου,p,δηλώνει την α-priori πιθανότητα το σύστημα να βρίσκεται στην κατάσταση 2. Παρόμοια εργαζόμαστε με τη δεύτερη συνιστώσα του α-posteriori δ.π.

$$T(\pi,\theta)=(T_1(\pi,\theta),T_2(\pi,\theta)),$$

Δηλαδή: $$T(p,\theta)\equiv T_2(\pi,\theta)$$

Έχουμε (βλέπε σχέσεις (7.1.1), (7.1.2)):

$$T(p,\theta)=\frac{a_\theta+\beta_\theta.p}{\gamma_\theta+\delta_\theta.p} \quad ,0\le p\le 1$$

$$\{\theta/p\}=\gamma_\theta+\delta_\theta.p \quad ,0\le p\le 1$$

όπου οι ποσότητες α_θ, β_θ, γ_θ, δ_θ, θ=1,2 ορίζονται στην (7.1.3).

Οι ιδιότητες των συναρτήσεων $T(p,1),T(p,2)$, $0\le p\le 1$ παρέχονται από τα λήμματα 7.1.1, 7.1.2 και την πρόταση 7.1.1 της παραγράφου 7.1.

Εστω $V_\beta(p), 0 \le p \le 1$ η συνάρτηση του ελάχιστου αναμενόμενου ολικού εκπίπτοντος κόστους για άπειρο χρονικό ορίζοντα, με συντελεστή έκπτωσης (discount-factor) $\beta \in (0,1)$.

Η συνάρτηση V_β ικανοποιεί την εξίσωση αριστοποίησης

$$V_\beta(p)=\min\{c_1+c.p+\beta.\sum_{\theta=1}^{2}\{\theta/p\}V_\beta(T(p,\theta)),\ R+\beta.V_\beta(0)\}, \qquad \underline{\textbf{8.3.1}}$$

$$0 \le p \le 1$$

όπου $$c = c_2 - c_1 (>0).$$

Η επόμενη πρόταση παρέχει ιδιότητες της άριστης συνάρτησης κόστους V_β καθώς και τη δομή της β-άριστης πολιτικής (δηλαδή της άριστης πολιτικής αναφορικά με το κριτήριο του ολικού εκπίπτοντος κόστους για άπειρο χρονικό ορίζοντα, με συντελεστή έκπτωσης $\beta \in (0,1)$).Η πρόταση αυτή αποτελεί πόρισμα αποτελεσμάτων του γενικού προβλήματος αντικατάστασης συστήματος που περιγράψαμε στις παραγράφους 6.1,6.2,6.3.

Πρόταση 8.3.1:

i) Η συνάρτηση $V_\beta(p), 0 \le p \le 1$ είναι αύξουσα και κοίλη.

ii) $$0 < V_\beta(p) \le \frac{R}{1-\beta}, 0 \le p \le 1$$

iii) Η συνάρτηση ελέγχου δ* της β-άριστης πολιτικής $(\delta^*)^\infty$ έχει την ακόλουθη δομή:

$$\delta^*(p) = \begin{cases} 0 & (\text{συνέχιση λειτουργίας}) \quad \text{αν} \quad 0 \le p \le p_0 \\ 1 & (\text{αντικατάσταση του συστήματος}) \quad \text{αν} \quad p_0 < p \le 1 \end{cases}$$

όπου $p_0 \in (0,1]$ κατάλληλη κρίσιμη ποσότητα.

Απόδειξη

i) Προκύπτει άμεσα από το γεγονός ότι η συνάρτηση $V_\beta(\pi), \pi \in \Pi$ είναι $\le_L$ αύξουσα και κοίλη (παράγραφος 6.2) και το γεγονός ότι για $\pi=(1-p,p), \pi'=(1-p',p') \in \Pi$, έχουμε:

$$\pi \le_L \pi' \Leftrightarrow p \le p'$$

ii) Από την εξίσωση αριστοποίησης (8.3.1) έχουμε:

$$V_\beta(p) \le R + \beta . V_\beta(0) , 0 \le p \le 1$$

Για p=0 παίρνουμε : $V_\beta(0) \le R + \beta . V_\beta(0)$

Από την οποία προκύπτει: $V_\beta(0) \le \frac{R}{1-\beta}$

Επομένως ,

$$V_\beta(p) \le R + \beta . \frac{R}{1-\beta} = \frac{R}{1-\beta}, \qquad 0 \le p \le 1$$

iii) Άμεση συνέπεια της πρότασης 6.3.1(βλέπε και παρατήρηση στην ενότητα 6.3) . □

Εστω $W_\beta(p), 0 \le p \le 1$,η συνάρτηση του αναμενόμενου ολικού εκπίπτοντος κόστους για άπειρο χρονικό ορίζοντα, όταν εφαρμόζουμε την πολιτική δ^∞ με συνάρτηση ελέγχου δ(p)=0 , $0 \le p \le 1$ (δεν αντικαθιστούμε ποτέ το σύστημα).

Προφανώς η συνάρτηση W_β ικανοποιεί την εξίσωση:

$$W_\beta(p) = c_1 + p.c + \beta . \sum_\theta \{\theta / p\} W_\beta(T(p,\theta)), \ 0 \le p \le 1 \qquad \underline{8.3.2}$$

όπου $c = c_2 - c_1 \ (> 0)$.

Για τον υπολογισμό της συνάρτησης W_β θεωρούμε την ακολουθία των συναρτήσεων $\{W_n(p), 0 \le p \le 1\}$που ικανοποιούν την αναγωγική σχέση:Για n=1,2,3…

$$W_n(p) = c_1 + p.c + \beta . \sum_{\theta=1}^{2} \{\theta / p\} W_{n-1}(T(p,\theta)), \ 0 \le p \le 1 \qquad \underline{8.3.3}$$

$$W_0(p) = 0 , \ 0 \le p \le 1.$$

□

Λήμμα 8.3.1:Εστω $\{W_n\}$ η ακολουθία των συναρτήσεων που ορίζονται με την επαναληπτική σχέση (8.3.3).Τότε

$$W_n(p) = (A_n.p + B_n).c + \frac{1-\beta^n}{1-\beta} c_1 \ , n = 0,1,2,\ldots \qquad \underline{8.3.4}$$

όπου

$$A_n = 1 + \beta . |P| . A_{n-1} \quad , \quad B_n = \beta . (p_{12} . A_{n-1} + B_{n-1}) \ , n = 1,2\ldots \qquad \underline{8.3.5}$$

$$A_0 = B_0 = 0$$

Απόδειξη

Το λήμμα ισχύει για n=1, επειδή

$$W_1(p)= c_1 + p.c= c_1+ (A_1.p +B_1).c,$$

Όπου $A_1=1$, $B_1=0$ ικανοποιούν τη σχέση (8.3.5).

Υποθέτουμε ότι το λήμμα ισχύει για κάποιο $n \geq 1$.Θα δείξουμε ότι ισχύει για n+1.Για $0 \leq p \leq 1$ έχουμε:

$$W_{n+1}(p) = c_1 + p.c + \beta. \sum_{\theta} \{\theta / p\} . W_n(T(p,\theta))$$

$$=c_1 + p.c + \beta. \sum_{\theta} \{\theta / p\}[(A_n.T(p,\theta) + B_n).c + \frac{1-\beta^n}{1-\beta}.c_1]$$

$$=c_1 + p.c + \beta. (A_n \sum_{\theta} \{\theta / p\}.T(p,\theta) + B_n).c + \beta.\frac{1-\beta^n}{1-\beta}.c_1$$

$$=\frac{1-\beta^{n+1}}{1-\beta}.c_1 + c.p + \beta.(A_n.\sum_{\theta}(\alpha_\theta + \beta_\theta.p) + B_n).c$$

$$=\frac{1-\beta^{n+1}}{1-\beta}.c_1 + c.p + \beta.[(A_n.((a_1 + a_2) + (\beta_1 + \beta_2).p) + B_n].c$$

Από τις σχέσεις (7.1.3) της ενότητας 7.1, παίρνουμε:

$$\alpha_1+\alpha_2 = p_{12}.(r_{21} + r_{22}) = p_{12}$$

$$\beta_1+ \beta_2 = |P| .(r_{21}+ r_{22}) = |P|$$

Επομένως,

$$W_{n+1}(p)= \frac{1-\beta^{n+1}}{1-\beta}.c_1 + c.p + \beta.[(A_n.(p_{12} + |P|.p) + B_n].c$$

$$=\frac{1-\beta^{n+1}}{1-\beta}.c_1 + (1 + \beta.|P|.A_n).c.p + \beta.(p_{12}.A_n + B_n).c$$

$$=\frac{1-\beta^{n+1}}{1-\beta}.c_1 + (A_{n+1}.p + B_{n+1}).c$$

Αρα το λήμμα ισχύει για n+1 και η απόδειξη της επαγωγής ολοκληρώνεται. □

Πρόταση 8.3.2: Η συνάρτηση $W_\beta(p)$, $0 \le p \le 1$ δίνεται από την σχέση

$$W_\beta(p)=\frac{1}{1-\beta}c_1+\frac{(1-\beta)p+\beta p_{12}}{(1-\beta).(1-\beta.|P|)}c,\ 0 \le p \le 1 \qquad \underline{8.3.6}$$

Απόδειξη

Η ακολουθία των συναρτήσεων $\{W_n\}$ συγκλίνει ομαλά στη συνάρτηση W_β: $W_n \to W_\beta$ όταν $n \to \infty$.

Επομένως από τη σχέση (8.3.4), παίρνουμε:

$$W_\beta(p)=(A.p+B).c+\frac{1}{1-\beta}c_1,\ 0 \le p \le 1 \qquad \underline{8.3.7}$$

όπου $\quad A=\lim_{n\to\infty} A_n\ ,\qquad B=\lim_{n\to\infty} B_n$

Τα όρια Α καί Β των ακολουθιών $\{A_n\}$και $\{B_n\}$που παράγονται από τις αναγωγικές σχέσεις (8.3.5), ικανοποιούν τις σχέσεις:

$$A=1+\beta.|P|.A\ ,\qquad B=\beta.(p_{12}.A+B)$$

Λύνοντας το σύστημα των δύο παραπάνω εξισώσεων παίρνουμε:

$$A=\frac{1}{1-\beta.|P|}\ ,\quad B=\frac{\beta.p_{12}}{(1-\beta).(1-\beta.|P|)}$$

και αντικαθιστώντας στην (8.3.7) παίρνουμε την (8.3.6). □

Η επόμενη πρόταση παρέχει αναγκαία και ικανή συνθήκη, ώστε η πολιτική ΄΄ μην αντικαθιστάς ποτέ το σύστημα΄΄ να είναι β-άριστη, για δοσμένο συντελεστή έκπτωσης $\beta \in (0,1)$.

Πρόταση 8.3.3: Η πολιτική δ^∞ με συνάρτηση ελέγχου $\delta(p)=0$, $0 \le p \le 1$ είναι β-άριστη, τότε και μόνο τότε αν:

$$c_1+\frac{1+\beta.p_{12}}{1-\beta.|P|}c \le R. \qquad \underline{8.3.8}$$

Απόδειξη

Η συνάρτηση W_β είναι άριστη αν ικανοποιεί την εξίσωση αριστοποίησης (8.3.1).Επειδή η W_β ικανοποιεί την εξίσωση (8.3.2) συνάγεται ότι η W_β ικανοποιεί την (8.3.1) τότε και μόνο τότε αν :

$$W_\beta(p) \le R + \beta W_\beta(0),\ 0 \le p \le 1 \qquad \underline{8.3.9}$$

Αντικαθιστώντας την (8.3.6) στην (8.3.9) και μετά από ορισμένες πράξεις βρίσκουμε ότι η (8.3.9) ικανοποιείται τότε και μόνο τότε αν ισχύει η (8.3.10):

$$c_1 + \frac{p + \beta . p_{12}}{1 - \beta . |P|} . c \le R \quad ,\ 0 \le p \le 1 \qquad \underline{8.3.10}$$

Όμως η (8.3.10) ισχύει, αν και μόνο αν ισχύει για p=1.

Επομένως συμπεραίνουμε ότι η συνάρτηση κόστους W_β είναι άριστη (ισοδύναμα η πολιτική δ^∞ :δ(p)=0, $0 \le p \le 1$ είναι β-άριστη),τότε και μόνον τότε αν ισχύει η σχέση (8.3.8). □

Παρατήρηση

Λαμβάνοντας υπόψη την πρόταση 8.3.1 (iii), η πρόταση 8.3.3 μπορεί να διατυπωθεί ως εξής: Η συνθήκη (8.3.8) είναι αναγκαία και ικανή συνθήκη, ώστε η κρίσιμη ποσότητα της β-άριστης πολιτικής να ισούται με τη μονάδα (p_0=1).

Η πρόταση που ακολουθεί παρέχει αναγκαία και ικανή συνθήκη, ώστε η πολιτική ''μην αντικαθιστάς ποτέ το σύστημα'' (ισοδύναμα η πολιτική με κρίσιμη ποσότητα p_0=1 να είναι β-άριστη, για κάθε συντελεστή έκπτωσης $\beta \in (0,1)$.

Πρόταση 8.3.4: Η πολιτική δ^∞ με συνάρτηση ελέγχου δ(p)=0, $0 \le p \le 1$ είναι β-άριστη $\forall \beta \in (0,1)$, τότε και μόνο τότε αν:

$$c_1 + \frac{1 + p_{12}}{1 - |P|} . c \le R \qquad \underline{8.3.11}$$

Απόδειξη

Η συνάρτηση

$$f(\beta) := \frac{1+\beta.p_{12}}{1-\beta.|P|}.c \quad , \quad 0 \prec \beta \prec 1$$

είναι προφανώς γνήσια αύξουσα και

$$\lim_{\beta \to 1^-} f(\beta) = \frac{1+p_{12}}{1-|P|}.c$$

Σημειώνουμε ότι:

$$|P| = p_{22} - p_{12} < 1$$

επειδή $p_{12} > 0$ (υπόθεση (1)).Επομένως η συνθήκη (8.3.11) συνάγεται από την πρόταση 8.3.3, θέτοντας β=1 στη σχέση (8.3.8) . □

Από την πρόταση 8.2.2 συνάγεται, ότι υπάρχει μια φραγμένη συνάρτηση $h(p), 0 \le p \le 1$ και μια σταθερά g έτσι ώστε να ικανοποιείται η εξίσωση αριστοποίησης για το μέσο κόστος ανά μονάδα χρόνου:

$$g + h(p) = \min\{c_1 + c.p + \sum_{\theta} \{\theta / p\} h(T(p,\theta), R\}, 0 \le p \le 1 \qquad \underline{8.3.12}$$

$$h(0)=0.$$

Η ποσότητα

$$g = \lim_{\beta \to 1^-} (1-\beta).V_\beta(0) \qquad \underline{8.3.13}$$

είναι το ελάχιστο μέσο κόστος ανά μονάδα χρόνου ($g = J(p), 0 \le p \le 1$).

Η επόμενη πρόταση παρέχει αναγκαία και ικανή συνθήκη, ώστε η πολιτική ΄΄μην αντικαθιστάς ποτέ το σύστημα΄΄ να είναι ΜΚ-άριστη .

Πρόταση 8.3.5:Η πολιτική δ^∞ με συνάρτηση ελέγχου $\delta(p)=0, 0 \le p \le 1$ είναι ΜΚ-άριστη ,και το ελάχιστο μέσο κόστος ανά μονάδα χρόνου δίνεται από τη σχέση

$$g = c_1 + \frac{p_{12}}{1-|P|}.c \qquad \underline{8.3.14}$$

αν και μόνον αν ισχύει η συνθήκη (8.3.11):

$$c_1 + \frac{1+p_{12}}{1-|P|}.c \le R \ .$$

Απόδειξη

As υποθέσουμε ότι ισχύει η συνθήκη (8.3.11).Τότε από την πρόταση 8.3.4 η πολιτική δ^{∞} με δ(p)=0, $0 \le p \le 1$ είναι β-άριστη για κάθε συντελεστή έκπτωσης $\beta \in (0,1)$ και

$$V_\beta(p)=W_\beta(p)\ ,\ 0 \le p \le 1 \qquad \underline{8.3.15}$$

Από τις σχέσεις (8.3.6), (8.3.13) και (8.3.15) παίρνουμε :

$$g = \lim_{\beta \to 1^-}(1-\beta).V_\beta(0) = \lim_{\beta \to 1^-}(1-\beta).W_\beta(0)$$

$$= \lim_{\beta \to 1^-}(c_1 + \frac{\beta.p_{12}}{1-\beta.|P|}.c) = c_1 + \frac{p_{12}}{1-|P|}.c \qquad \underline{8.3.16}$$

Επίσης για κάθε $\beta \in (0,1)$ έχουμε :

$$h_\beta(p)=V_\beta(p)-V_\beta(0)= W_\beta(p)- W_\beta(0)=\frac{c}{1-\beta.|P|}.p\ ,0 \le p \le 1$$

Για $\beta \to 1^-$ παίρνουμε:

$$h(p) = \lim_{\beta \to 1^-} h_\beta(p) = \frac{c}{1-|P|}.p\ , 0 \le p \le 1.$$

Θα δείξουμε ότι η σταθερά g και η συνάρτηση $h(p), 0 \le p \le 1$ ικανοποιούν την εξίσωση αριστοποίησης (8.3.12). Έχουμε

$$c_1 + c.p + \sum_\theta \{\theta / p\} h(T(p,\theta))$$

$$= c_1 + c.p + \frac{c}{1-|P|}\sum_\theta \{\theta / p\} T(p,\theta))$$

$$= c_1 + c.p + \frac{c}{1-|P|}\sum_\theta (\alpha_\theta + \beta_\theta.p)$$

$$= c_1 + c.p + \frac{c}{1-|P|}(a_1 + a_2 + (\beta_1 + \beta_2).p)$$

$$= c_1 + c.p + \frac{c}{1-|P|}(p_{12} + |P|.p)$$

$$= c_1 + \frac{c}{1-|P|}(p_{12} + p) = g + h(p) \qquad \underline{8.3.17}$$

Λόγω της συνθήκης (8.3.11) έχουμε

$$c_1 + \frac{c}{1-|P|}(p_{12} + p) \le R, 0 \le p \le 1 \qquad \underline{8.3.18}$$

Από τις (8.3.17) και (8.3.18)συνάγεται ότι η σταθερά g και η συνάρτηση $h(p), 0 \le p \le 1$ ικανοποιούν την εξίσωση αριστοποίησης (8.3.12) και η ποσότητα g πού δίνεται από τη σχέση (8.3.16) εκφράζει το ελάχιστο μέσο κόστος ανά μονάδα χρόνου. Επιπλέον η πολιτική δ^{∞} είναι ΜΚ-άριστη επειδή για κάθε $p \in [0,1]$ η απόφαση δ(p)=0 ελαχιστοποιεί το δεύτερο σκέλος της (8.3.12).

Αντιστρόφως ας υποθέσουμε ότι η πολιτική δ^{∞} με δ(p)=0, $0 \le p \le 1$ είναι ΜΚ-άριστη. Τότε σε συνδυασμό με την εξίσωση αριστοποίησης (8.3.12) παίρνουμε την εξίσωση:

$$g + h(p) = c_1 + c.p + \sum_{\theta} \{\theta / p\} h(T(p,\theta)\}, \ 0 \le p \le 1$$

η οποία όπως διαπιστώσαμε από τη σχέση (8.3.17), ικανοποιείται για

$$g = c_1 + \frac{c}{1-|P|}.p_{12} \ , \ h(p) = \frac{c}{1-|P|}.p, 0 \le p \le 1$$

Επιπλέον έχουμε:

$$c_1 + c_2.\sum_{\theta} \{\theta / p\} h(T(p,\theta)\} = c_1 + \frac{c}{1-|P|}(p_{12} + p) \le R, 0 \le p \le 1$$

Η τελευταία σχέση συνεπάγεται από τη συνθήκη (8.3.11). □

Παρατήρηση

Επειδή $1-|P| = 1-(p_{22}-p_{12}) = p_{21}+p_{12}$,

η σχέση (8.3.14) γράφεται:

$$g = c_1 + \frac{p_{12}}{p_{12}+p_{21}}.c = c_1 + \frac{p_{12}}{p_{12}+p_{21}}.(c_2 - c_1) = \frac{p_{21}}{p_{12}+p_{21}}.c_1 + \frac{p_{12}}{p_{12}+p_{21}}.c_2 \qquad \underline{\mathbf{8.3.19}}$$

Η κατανομή $\underset{\sim}{x} = (\frac{p_{21}}{p_{21}+p_{12}}, \frac{p_{12}}{p_{12}+p_{21}})$ είναι η στάσιμη κατανομή του συστήματος (Μαρκοβιανής αλυσίδας) με πίνακα μετάβασης P πού αντιστοιχεί στην απόφαση α=0 και ικανοποιεί τη σχέση

$$\underset{\sim}{x}.\mathrm{P} = \underset{\sim}{x}$$

Οι πιθανότητες $x_1 = \frac{p_{21}}{p_{21}+p_{12}}$ και $x_2 = \frac{p_{12}}{p_{21}+p_{12}}$ εκφράζουν την μακροπρόθεσμη αναλογία των χρονικών περιόδων όπου το σύστημα βρίσκεται στις καταστάσεις 1 και 2 αντίστοιχα. Με την επισήμανση αυτή και λόγω της σχέσης (8.3.19) η ποσότητα g πράγματι ερμηνεύεται ως το μέσο κόστος ανά μονάδα χρόνου που αντιστοιχεί στην πολιτική δ^{∞} με δ(p)=0, $0 \le p \le 1$. □

As θεωρήσουμε τώρα την *παράδοξη* πολιτική δ^{∞} όπου δ(p)=0 αν p=0 και δ(p)=1 αν $p \neq 0$.

Εφαρμόζοντας την πολιτική αυτή, τις μισές χρονικές περιόδους το σύστημα λειτουργεί/συντηρείται και τις υπόλοιπες περιόδους το σύστημα αντικαθίσταται. Πράγματι, αν το σύστημα βρίσκεται αρχικά σε μια «κατάσταση» $p \neq 0$ (πού δηλώνει την a-priori πιθανότητα το σύστημα να βρίσκεται στην κατάσταση 2),τότε εφαρμόζοντας την παράδοξη πολιτική, το σύστημα στις επόμενες χρονικές περιόδους μεταβαίνει διαδοχικά στις «καταστάσεις»:

$$0, T(0,1) \acute{\eta} \mathrm{T}(0,2), 0, \mathrm{T}(0,1) \acute{\eta} \mathrm{T}(0,2), 0 \quad \text{κ.ο.κ.}$$

στις οποίες το σύστημα λειτουργεί /συντηρείται και αντικαθίσταται εναλλάξ. Επομένως το μέσο κόστος ανά μονάδα χρόνου πού αντιστοιχεί στην παράδοξη πολιτική είναι:

$$g = \frac{c_1 + R}{2}$$

Στην πρόταση που ακολουθεί, αποδεικνύουμε ότι αποκλείεται η παράδοξη πολιτική να είναι ΜΚ-άριστη.

Πρόταση 8.3.6:

Η πολιτική δ^{∞} με συνάρτηση ελέγχου

$$\delta(p)=\begin{cases}0 & (\text{συνέχιση λειτουργίας}) \quad \text{άν} \quad p=0\\ 1 & (\text{αντικατάσταση του συστήματος}) \quad \text{αν} \quad 0<p\le 1\end{cases}$$

δεν είναι ΜΚ-άριστη.

Απόδειξη

Θα δείξουμε ότι η εξίσωση αριστοποίησης (8.3.12) ικανοποιείται για

$$g=\frac{c_1+R}{2},\quad h(p)=\frac{R-c_1}{2}, 0<p\le 1,\ \text{h(o)=0}$$

και η πολιτική δ^{∞} είναι ΜΚ-άριστη αν και μόνο αν ισχύει η συνθήκη:

$$R=c_1 \qquad \underline{8.3.20}$$

Πράγματι, έχουμε

$$g+h(p)=\frac{c_1+R}{2}+\frac{R-c_1}{2}=R,\ 0<p\le 1$$

$$g+h(0)=g=\frac{c_1+R}{2}$$

$$c_1+c.p+\sum_{\theta}\{\theta/p\}.h(T(p,\theta))=c_1+c.p+\frac{R-c_1}{2}\sum_{\theta}\{\theta/p\}=$$

$$c_1+c.p+\frac{R-c_1}{2}=\frac{c_1+R}{2}+c.p,\ 0\le p\le 1.$$

Παρατηρούμε ότι για p=0,

$$c_1+c.0+\sum_{\theta}\{\theta/0\}.h(T(0,\theta))=\frac{c_1+R}{2}=g+h(o).$$

Επομένως η εξίσωση αριστοποίησης (8.3.12) ισχύει για p=0,αν και μόνον αν:

$$g+h(0)=\frac{c_1+R}{2}\le R \qquad \underline{8.3.21}$$

Για $p \neq 0$ η εξίσωση αριστοποίησης (8.3.12) ικανοποιείται αν και μόνον αν:

$$g + h(p) = R \leq c_1 + c.p + \sum_{\theta} \{\theta / p\} h(T(p,\theta)), 0 \leq p \leq 1,$$

δηλαδή,

$$R \leq \frac{c_1 + R}{2} + c.p,\ 0 < p \leq 1 \qquad \underline{8.3.22}$$

Όμως η σχέση (8.3.22) ισοδυναμεί με τη σχέση

$$R \leq \frac{c_1 + R}{2} \qquad \underline{8.3.23}$$

Από τις σχέσεις (8.3.21) και (8.3.23) συνάγεται ότι η εξίσωση αριστοποίησης (8.3.12) ικανοποιείται τότε και μόνο τότε, αν ισχύει η ισότητα

$$\frac{c_1 + R}{2} = R$$

η οποία, όπως διαπιστώνουμε άμεσα, ισοδυναμεί με τη σχέση (8.3.20).Συνοψίζοντας, η ποσότητα

$$\frac{c_1 + R}{2} = g$$

εκφράζει το ελάχιστο μέσο κόστος ανά μονάδα χρόνου και η πολιτική δ^{∞} είναι ΜΚ-άριστη, τότε και μόνον τότε ,αν ισχύει η συνθήκη (8.3.20).Όμως λόγω της Υπόθεσης (2): $c_1 < c_2 < R$ η συνθήκη (8.3.20) δεν ισχύει και επομένως αποκλείεται η πολιτική δ^{∞} να είναι ΜΚ-άριστη. □

Πρόταση 8.3.7:

Αν $$(c_2 <) R < c_1 + \frac{1 + p_{12}}{1 - |P|} c \qquad \underline{8.3.24}$$

Τότε η συνάρτηση ελέγχου δ της (ΜΚ-άριστης) πολιτικής δ^{∞} έχει την ακόλουθη δομή :

$$\delta(p) = \begin{cases} 0 & (\text{συνέχιση λειτουργίας}) \quad \text{άν} \quad 0 \leq p \leq p^* \\ 1 & (\text{αντικατάσταση του συστήματος}) \quad \text{αν} \quad p^* < p \leq 1 \end{cases}$$

όπου $p^* \in (0,1)$ είναι κατάλληλη κρίσιμη ποσότητα.

Απόδειξη

Από την πρόταση 8.3.3 και την παρατήρηση πού ακολουθεί συνάγεται ότι για κάθε $\beta \in (0,1)$ η κρίσιμη ποσότητα της β-άριστης πολιτικής $p_0(\beta) \in (0,1)$, αν και μόνο αν ισχύει:

$$R < c_1 + \frac{(1+\beta.p_{12})}{1-\beta.|P|}.c$$

Επειδή η συνάρτηση

$$f(\beta) := \frac{1+\beta.p_{12}}{1-\beta.|P|}.c, 0 \prec \beta \prec 1$$

είναι γνήσια αύξουσα και

$$\lim_{\beta \to 1^-} f(\beta) = \frac{1+p_{12}}{1-|P|}.c,$$

από τη συνθήκη (8.3.24) συνάγεται ότι υπάρχει $0 < \varepsilon < 1$ έτσι ώστε:

$$R \prec c_1 + \frac{(\beta.p_{12}+1)}{1-\beta.|P|}.c \qquad \forall \beta \in (1-\varepsilon,1)$$

Επομένως $0 < p_0(\beta) < 1 \qquad \forall \beta \in (1-\varepsilon,1)$

Θεωρούμε την ακολουθία $\{\beta_n\} \subseteq (1-\varepsilon,1)$ με $\lim_{n\to\infty} \beta_n = 1$.

Τότε $0 \prec p_0(\beta_n) \prec 1 \qquad \forall n = 1,2,...$

Σύμφωνα με το θεώρημα Bolzano-Weirstrass (κάθε φραγμένη ακολουθία περιέχει μια συγκλίνουσα υπακολουθία), υπάρχει μια υπακολουθία $\{\beta_{n_k}\}$ ώστε:

$$p_0(\beta_{n_k}) \to p^* \text{ όταν το } k \to \infty$$

και $p^* \in [0,1]$.

Αν $p^* > 0$, τότε για τυχόν $p \in [0, p^*)$ υπάρχει $\tau \in \mathbb{N}$ ώστε $p < p_0(\beta_{n_k}) \qquad \forall k \geq \tau$.

Επομένως η απόφαση για συνέχιση λειτουργίας / συντήρησης του συστήματος στο p είναι β_{n_k} -άριστη $\forall k \geq \tau$.Σύμφωνα με την πρόταση 8.1.3 η απόφαση ,δ(p)=0,είναι MK-άριστη. Αν $p^* \prec 1$, με όμοιο τρόπο αποδεικνύουμε ότι η απόφαση για αντικατάσταση του συστήματος, δ(p)=1,είναι MK-άριστη για $\forall p \in (p^*, 1]$.

Απομένει να δείξουμε ότι $p^* \neq 0$, $p^* \neq 1$.

Το γεγονός ότι $p^* \neq 0$, προκύπτει άμεσα από την πρόταση 8.3.6.As θεωρήσουμε ότι $p^* = 1$.Τότε για $\forall p \in [0,1)$ η απόφαση δ(p)=0 είναι MK-άριστη.Αυτό όμως αποκλείεται από την πρόταση 8.3.5 λόγω της συνθήκης (8.3.24).Επομένως $p^* \neq 1$.Συνεπώς για την κρίσιμη ποσότητα p^* της MK-άριστης πολιτικής έχουμε $0 < p^* < 1$ και η απόδειξη ολοκληρώθηκε. □

ΣΥΜΠΕΡΑΣΜΑΤΑ

Στο κεφάλαιο αυτό εφαρμόζουμε το κριτήριο του μέσου κόστους ανά μονάδα χρόνου σε προβλήματα POMDP αντικατάστασης συστήματος.

- Σχετικά με το γενικό πρόβλημα αντικατάστασης συστήματος των Ohnishi-Ibaraki δείχνουμε ότι η εξίσωση αριστοποίησης για το μέσο κόστος (ACOE) έχει λύση και ότι υπάρχει MK-άριστη πολιτική.
- Σχετικά με το ειδικό πρόβλημα αντικατάστασης συστήματος με δύο καταστάσεις και δύο μηνύματα δίνουμε:α)αναγκαία και ικανή συνθήκη ώστε η πολιτική να μην αντικαθιστούμε ποτέ το σύστημα να είναι MK-άριστη β) αναγκαία και ικανή συνθήκη ώστε η MK-άριστη πολιτική να είναι control-limit:αντικαθιστούμε το σύστημα αν η πιθανότητα μη λειτουργικής κατάστασης υπερβαίνει μία κρίσιμη ποσότητα $p^* \in (0,1)$.

ΚΕΦΑΛΑΙΟ 9

Εφαρμογές των POMDPs σε βέλτιστες πολιτικές επιλογής διδακτικών μεθόδων

Περίληψη

Οι POMDPs είναι μοντέλα σχεδιασμού και λήψης αποφάσεων σε συστήματα όπου κυρίαρχο στοιχείο είναι η αβεβαιότητα όσον αφορά την κατάσταση ενός συστήματος. Μια ενδιαφέρουσα περίπτωση εφαρμογής έχουμε σε ένα πρόβλημα σχεδιασμού και λήψης αποφάσεων σε ένα εκπαιδευτικό σύστημα βλέπε και Goulionis [37].

Μελετάμε ένα πρόβλημα επιλογής ανάμεσα σε δύο διδακτικές μεθόδους, μία συμβατική και φθηνή και μία εξειδικευμένη και δαπανηρή (π.χ.ενισχυτική,εξατομικευμένη, υποστηριζόμενη από υπολογιστές διδασκαλία κ.λ.π).Το πρόβλημα αυτό τίθεται στη μορφή μερικά παρατηρήσιμης Μαρκοβιανής διαδικασίας αποφάσεων (POMDP) με δύο δυνατές καταστάσεις ,αναφορικά με το βαθμό αφομοίωσης της διδασκόμενης ύλης από την τάξη και δύο μηνύματα (π.χ. επιτυχία/αποτυχία σε test).Υπολογίζεται αναλυτικά η συνάρτηση του ελάχιστου αναμενόμενου ολικού εκπίπτοντος κόστους για άπειρο χρονικό ορίζοντα,και προσδιορίζεται η αντίστοιχη άριστη πολιτική επιλογής διδακτικών μεθόδων σε δύο περιπτώσεις:**α)**περίπτωση πλήρους αβεβαιότητας, όπου το μήνυμα (π.χ. αποτέλεσμα ενός test) είναι ανεξάρτητο από την κατάσταση της τάξης είτε επιλέγεται

η συμβατική, είτε η εξειδικευμένη μέθοδος διδασκαλίας και **β)** περίπτωση πλήρους αβεβαιότητας όταν επιλέγεται η συμβατική μέθοδος ,και μερικής πληροφόρησης όταν επιλέγεται η εξειδικευμένη μέθοδος διδασκαλίας.

9.1. Περιγραφή του προβλήματος επιλογής διδακτικών μεθόδων

Στην διαμόρφωση των φυσικών συστημάτων η αρχή της κατάστασης του φυσικού συστήματος, έχει αποδειχθεί ένα πολύτιμο εργαλείο για τον χαρακτηρισμό της λειτουργικότητας του συστήματος. Αυτή η ιδέα είναι επίσης ουσιαστική στην περιγραφή της μαθησιακής διαδικασίας σε μια τάξη. Η κατάσταση μιας τάξης αποτελεί μέτρο της μαθησιακής κατάστασης των μαθητών που την απαρτίζουν.Η εσωτερική κατάσταση κάθε μαθητή συνυφαίνεται με διάφορους παράγοντες, όπως κληρονομικές καταβολές, οικογενειακός και γενικότερα κοινωνικός περίγυρος, προσωπικό μοντέλο σκέψης, υπόβαθρο γνώσεων του μαθητού, συναισθηματική ζωή και γενικότερα ψυχολογικοί παράγοντες.

Η κατάσταση της τάξης εξαρτάται από την επικοινωνία του διδάσκοντος με τους μαθητές, από την διδακτική μέθοδο που χρησιμοποιείται, από το πνεύμα ομαδικότητας των μαθητών, αλληλεπιδράσεις μεταξύ των μαθητών κ.λ.π. Το κυρίαρχο στην διαμόρφωση αρμονικής, και πολυδιάστατης σχέσης ανάμεσα στους μαθητές είναι ο διδάσκων. Ο μαθητής δεν είναι για τον διδάσκοντα ένα σακκούλι, που πρέπει να γεμίσει με γνώσεις, αλλά ένα σπίρτο που ο διδάσκων πρέπει να το ανάψει.

Για τον διδάσκοντα η κατάσταση της τάξης σημαίνει βασικά το βαθμό αφομοίωσης της διδασκόμενης ύλης. Επειδή, όπως επισημάναμε παραπάνω, η κατάσταση της τάξης είναι πολυπαραγοντική θεωρείται άγνωστη. Ωστόσο ο διδάσκων μπορεί να πάρει μία εικόνα αυτής της κατάστασης μέσω κάποιων μηνυμάτων, όπως π.χ επιδόσεις σε tests.Στα υποδείγματα που θα μελετήσουμε θεωρούμε δύο δυνατές καταστάσεις για την τάξη:καλή και κακή, που αντανακλούν υψηλό και χαμηλό βαθμό αφομοίωσης της διδασκόμενης ύλης και κωδικοποιούνται αντίστοιχα ως 1 και 2.

Θεωρούμε επίσης δύο τύπους μηνυμάτων που κωδικοποιούνται ως 1,2. Θεωρούμε ότι το μήνυμα τύπου 1,(θ=1), είναι ευνοϊκό για την κατάσταση 1 (καλή κατάσταση της τάξης), ενώ το μήνυμα τύπου 2,(θ=2), είναι ευνοϊκό για την κατάσταση 2,(κακή κατάσταση της τάξης).

Τα είδη μηνυμάτων που μπορούμε να χρησιμοποιήσουμε είναι ποικίλα. Για παράδειγμα αν το μήνυμα εκφράζεται μέσω των αποτελεσμάτων ενός test,που διενεργεί ο διδάσκων στην τάξη, τότε έχουμε μήνυμα τύπου 1 (θ=1),αν το ποσοστό της επιτυχίας είναι πάνω από ένα κρίσιμο όριο π.χ επιτυχία στο test που υπερβαίνει το 60%,ενώ έχουμε μήνυμα τύπου 2,αν το ποσοστό της επιτυχίας είναι μικρότερο από το κρίσιμο. Άλλα είδη μηνυμάτων είναι ο βαθμός συμμετοχής στην τάξη, η γλώσσα του σώματος, οι τυχόν πρωτοβουλίες που λαμβάνουν οι μαθητές σχετικά με την διδασκαλία του μαθήματος, εφόσον εκφρασθούν ποσοτικά.

Στην αρχή κάθε χρονικής περιόδου επιλέγεται μία απόφαση (μέθοδος διδασκαλίας) από ένα σύνολο A={0,1} όπου οι κωδικοποιήσεις 0,1 είναι:

0: συμβατική μέθοδος διδασκαλίας

1: εξειδικευμένη μέθοδος διδασκαλίας.

Ο περιορισμός σε δύο διδακτικές μεθόδους γίνεται για λόγους απλότητας. Η συμβατική μέθοδος διδασκαλίας θεωρείται φθηνή, ενώ η εξειδικευμένη μέθοδος σχετικά ακριβή και παρέχεται σε ποικίλες μορφές όπως ενισχυτική, εξατομικευμένη, υποστηριζόμενη από υπολογιστές, διδασκαλία που δεν περιορίζεται μόνο στο περιβάλλον του σχολείου αλλά ενισχύεται με δραστηριότητες όπως επισκέψεις σε μουσεία, ιδρύματα ερευνών ,εργαστήρια κ.λ.π.

Αν i είναι η κατάσταση της τάξης και α η απόφαση, τότε τα άμεσα κόστη $C(i,\alpha)$, $i=1,2$, $\alpha=0,1$ είναι:

$$C(1,0)=0,\ C(2,0)=C,\ C(1,1)=C(2,1)=R.$$

με $0<C<R$.

Το κόστος C αντανακλά την έλλειψη προσφοράς ευκαιριών και δυνατοτήτων της συμβατικής διδακτικής μεθόδου, στην οποία ενδέχεται να οφείλεται-μερικά τουλάχιστον- η κακή κατάσταση της τάξης.

Η επίδραση των διδακτικών μεθόδων στην κατάσταση της τάξης εκφράζεται μέσω των πινάκων μετάβασης $P^{\alpha} = (p_{ij}^{\alpha})$ $\alpha = 0,1$. Θεωρούμε ότι :

$$P^0 = \begin{pmatrix} 1-\lambda & \lambda \\ 0 & 1 \end{pmatrix} \qquad P^1 = \begin{pmatrix} 1-\lambda & \lambda \\ 1-\lambda & \lambda \end{pmatrix}$$

όπου $0 < \lambda < 1$.

Οι πίνακες μετάβασης ενσωματώνουν την ποιοτική διαφορά ανάμεσα στις δύο διδακτικές μεθόδους. Έτσι με τη συμβατική μέθοδο διδασκαλίας η πιθανότητα μετάβασης από την κακή στην καλή κατάσταση θεωρείται μηδενική ($p_{21}^0 = 0$), ενώ με την εξειδικευμένη μέθοδο η αντίστοιχη πιθανότητα μετάβασης θεωρείται μη μηδενική ($p_{21}^1 = 1 - \lambda > 0$). Σημειώνουμε ότι η θεώρηση αυτή δεν αφορά κάθε μαθητή μεμονωμένα αλλά την τάξη ως σύνολο.

Οι πίνακες μηνυμάτων $R^{\alpha} = (r_{i\theta}^{\alpha}), \alpha = 0,1$ θεωρούμε ότι έχουν τη μορφή

$$R^{\alpha} = \begin{pmatrix} q^{\alpha} & 1-q^{\alpha} \\ 1-q^{\alpha} & q^{\alpha} \end{pmatrix} , \alpha = 0,1,$$

και υποτίθενται TP_2 ,δηλαδή $q^{\alpha} \in [\frac{1}{2}, 1]$, $\alpha = 0,1$.

Η ποσότητα q^{α} δηλώνει την πιθανότητα να πάρουμε μήνυμα συμβατό με την κατάσταση της τάξης:

$$r_{11}^{\alpha} = r_{22}^{\alpha} = q^{\alpha} \quad , \alpha = 0,1 .$$

Όπως και στα Κεφ 7 και 8, θα εργαζόμαστε με την δεύτερη συνιστώσα του δ.π π =(1-p,p) ,p, που δηλώνει την πιθανότητα η τάξη να βρίσκεται στην κατάσταση 2 (κακή κατάσταση).Παρόμοια εργαζόμαστε με τη δεύτερη συνιστώσα του α *posteriori* δ.π.

$$T(\pi, \theta, \alpha) = (T_1(\pi, \theta, \alpha), T_2(\pi, \theta, \alpha)),$$

την οποία συμβολίζουμε με $T(p, \theta, \alpha)$, δηλαδή

$$T(p, \theta, \alpha) \equiv T_2(\pi, \theta, \alpha)$$

εκφράζει την *α posteriori* πιθανότητα η τάξη στην επόμενη χρονική περίοδο (t+1) να βρεθεί στην κατάσταση 2 δοσμένου ότι στον ίδιο χρόνο (t+1) παίρνουμε μήνυμα θ

και ότι στον τρέχοντα χρόνο (t) η πιθανότητα για την κατάσταση 2 είναι p και επιλέγουμε την διδακτική μέθοδο α.

Η ποσότητα $\{\theta/p,\alpha\}$ εκφράζει την πιθανότητα το μήνυμα που θα ληφθεί στην επόμενη χρονική περίοδο (t+1) να είναι θ, δοσμένου ότι στον τρέχοντα χρόνο (t) η πιθανότητα για την κατάσταση 2 είναι p και επιλέγουμε τη διδακτική μέθοδο α.

Εφαρμόζοντας τις σχέσεις (7.1.1),(7.1.2),(7.1.3) του κεφαλαίου 7 παίρνουμε:

Για $0 \le p \le 1$,

$$T(p,\theta=1,\alpha=0)=\frac{(1-q^0).(\lambda+(1-\lambda).p)}{(1-\lambda).q^0+\lambda(1-q^0)+(1-\lambda).(1-2q^0).p}. \qquad \underline{9.1.1}$$

$$\{\theta=1/p,a=0\}=(1-\lambda).q^0+\lambda.(1-q^0)+(1-\lambda).(1-2q^0).p. \qquad \underline{9.1.2}$$

$$T(p,\theta=2,\alpha=0)=\frac{q^0.(\lambda+(1-\lambda).p)}{(1-\lambda).(1-q^0)+\lambda.q^0+(1-\lambda).(2q^0-1).p}. \qquad \underline{9.1.3}$$

$$\{\theta=2/p,a=0\}=(1-\lambda).(1-q^0)+\lambda.q^0+(1-\lambda).(2q^0-1).p. \qquad \underline{9.1.4}$$

$$T\ (p,\theta=1,\alpha=1)=\frac{\lambda.(1-q^1)}{(1-\lambda).q^1+\lambda.(1-q^1)}\equiv s_1(\text{σταθερό}),\ 0\le p\le 1. \qquad \underline{9.1.5}$$

$$\{\theta=1/p,\ \alpha=1\}=(1-\lambda).q^1+\lambda.(1-q^1)\equiv\gamma_1(\text{σταθερό}),\ 0\le p\le 1. \qquad \underline{9.1.6}$$

$$T\ (p,\theta=2,\alpha=1)=\frac{\lambda.q^1}{(1-\lambda).(1-q^1)+\lambda.q^1}\equiv s_2(\text{σταθερό}),\ 0\le p\le 1. \qquad \underline{9.1.7}$$

$$\{\theta=2/p,\ \alpha=1\}=(1-\lambda).(1-q^1)+\lambda.q^1\equiv\gamma_2(\text{σταθερό}),\ 0\le p\le 1. \qquad \underline{9.1.8}$$

Σχετικά με το κριτήριο ελαχιστοποίησης του αναμενόμενου ολικού εκπίπτοντος κόστους για άπειρο χρονικό ορίζοντα η βέλτιστη συνάρτηση τιμών $V(p), 0 \le p \le 1$, ικανοποιεί την εξίσωση αριστοποίησης:

$$V(p)= min\{C.p + \beta.\sum_{\theta=1}^{\theta=2}\{\theta / p, a=0\}.V(T(p,\theta,\alpha=0)),$$

$$R + \beta.\sum_{\theta=1}^{\theta=2}\{\theta / p, a=1\}.V(T(p,\theta,\alpha=1))\}, 0 \le p \le 1 \qquad \underline{9.1.9}$$

Λαμβάνοντας υπόψη τις (9.1.5)-(9.1.8) η (9.1.9) γράφεται:

$$V(p)= min\{C.p + \beta.\sum_{\theta=1}^{\theta=2}\{\theta / p, a=0\}.V(T(p,\theta,\alpha=0)),$$

$$R + \beta.(\gamma_1.V(s_1) + \gamma_2.V(s_2))\}, 0 \le p \le 1 \qquad \underline{9.1.10}$$

($\beta \in (0,1)$ είναι ο συντελεστής έκπτωσης).

Σχετικά με τους πίνακες μηνυμάτων θα εξετάσουμε δύο περιπτώσεις.

Στην ενότητα 9.2 μελετάμε την περίπτωση όπου το μήνυμα είναι ανεξάρτητο από την κατάσταση της τάξης, είτε επιλέγουμε τη συμβατική είτε την εξειδικευμένη διδακτική μέθοδο (περίπτωση πλήρους αβεβαιότητας για α=0 και α=1).

Στην ενότητα 9.3 μελετάμε την περίπτωση πλήρους αβεβαιότητας όταν επιλέγεται η συμβατική μέθοδος (α=0) και μερικής πληροφόρησης όταν επιλέγεται η εξειδικευμένη μέθοδος (α=1). Και στις δύο περιπτώσεις υπολογίζουμε την άριστη συνάρτηση τιμών και προσδιορίζουμε την άριστη πολιτική επιλογής διδακτικών μεθόδων για άπειρο χρονικό ορίζοντα.

9.2. Περίπτωση πλήρους αβεβαιότητας για οποιαδήποτε επιλογή διδακτικής μεθόδου.

Υποθέτουμε ότι $q^a = \frac{1}{2}$, α=0,1.

Οι πίνακες μηνυμάτων γράφονται:

όπου $p*$ είναι κατάλληλη κρίσιμη ποσότητα.

Στη συνέχεια θα υπολογίσουμε αναλυτικά την κρίσιμη ποσότητα $p*$ και τη βέλτιστη συνάρτηση τιμών $V / [0,1]$.

Με $T_n(p), 0 \le p \le 1$ συμβολίζουμε τη n-στη σύνθεση της συνάρτησης T(p), $0 \le p \le 1$ (βλέπε σχέση 9.2.1)

$$T_n(p) := T(T_{n-1}(p)),\ 0 \le p \le 1,\ n=1,2,\ldots$$

όπου $T_0(p) = p\ ,\ 0 \le p \le 1$. (ταυτοτική συνάρτηση)

Λήμμα 9.2.1: Για n=1,2,3...

$$T_n(p)=1-(1-\lambda)^n.(1-p),\ 0 \le p \le 1$$

Απόδειξη

Η απόδειξη γίνεται με επαγωγή. Το λήμμα αληθεύει για n=1 επειδή

$$T_1(p) = T(p) = \lambda + (1-\lambda).p = 1-(1-\lambda).(1-p), 0 \le p \le 1.$$

As υποθέσουμε ότι το λήμμα αληθεύει για κάποιο $n \ge 1$. Τότε,

$$T_{n+1}(p) = T(T_n(p)) = 1-(1-\lambda).(1-T_n(p)) = 1-(1-\lambda)^{n+1}.(1-p),\ 0 \le p \le 1$$

και η απόδειξη ολοκληρώνεται. □

Θα εξετάσουμε στη συνέχεια αναγκαία και ικανή συνθήκη ώστε η πολιτική δ^∞ με δ(p)=0, $0 \le p \le 1$ (δηλαδή η πολιτική να ακολουθούμε πάντοτε τη συμβατική μέθοδο διδασκαλίας) να είναι άριστη.

Έστω W(p), $0 \le p \le 1$ η συνάρτηση του αναμενόμενου ολικού εκπίπτοντος κόστους, όταν ακολουθούμε την πολιτική δ^∞ με $0 \le p \le 1$ (δηλαδή εφαρμόζουμε συνεχώς την συμβατική μέθοδο). Έχουμε

$$W(p)= C.\sum_{n=0}^{\infty}\beta^n\ .\ T_n(p)\ = C.\sum_{n=0}^{\infty}\beta^n\ .(1-(1-p)\ .(1-\lambda)^n)=$$

$$= C.\left(\frac{1}{1-\beta}\ -\frac{(1-p)}{1-\beta.(1-\lambda)}\right), 0 \le p \le 1. \qquad \underline{9.2.3}$$

$$R^0 = R^1 = \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix}$$

Επομένως το μήνυμα που λαμβάνεται σε οποιαδήποτε χρονική περίοδο είναι ανεξάρτητο από την κατάσταση της τάξης και συνεπώς δεν παρέχει καμιά πληροφορία για την κατάσταση (πλήρης αβεβαιότητα).

Από τις σχέσεις (9.1.1)-(9.1.8) παίρνουμε:

$$T(p,\theta=1,\alpha=0)= T(p,\theta=2,\alpha=0)= p(1-\lambda)+\lambda\ ,\ 0 \le p \le 1\ .$$

και την κοινή τιμή τη συμβολίζουμε με T (p),δηλαδή:

$$T(p) := p(1-\lambda)+\lambda\ ,\ 0 \le p \le 1 \qquad \underline{9.2.1}$$

$$T(p,\theta=1,\alpha=1)= T(p,\theta=2,\alpha=1) =\lambda\ ,\ 0 \le p \le 1\ .$$

$$\{\theta/p,\alpha\}=\frac{1}{2}\ ,\theta=1,2\ ,\alpha=0,1,\ 0 \le p \le 1.$$

Η εξίσωση αριστοποίησης (9.1.9) γράφεται στην περίπτωσή μας:

$$V(p) = \min\{C.p + \beta V(T(p)), R + \beta V(\lambda)\}, 0 \le p \le 1. \qquad \underline{9.2.2}$$

Η βέλτιστη συνάρτηση τιμών $V(p), 0 \le p \le 1$ είναι αύξουσα, κοίλη και συνεχής.

Πράγματι, θεωρούμε την επαναληπτική σχέση

$$V_n(p) = HV_{n-1}(p) = \min\{C.p + \beta.V_{n-1}(T(p)), R + \beta.V_{n-1}(\lambda)\ \},$$

$$0 \le p \le 1, n=1,2,\ldots$$

όπου $V_0(p) = 0\ ,\ 0 \le p \le 1.$

Αποδεικνύεται εύκολα με επαγωγή ότι για κάθε n=1,2,... η συνάρτηση V_n / [0,1] είναι αύξουσα, κοίλη και συνεχής.

Επομένως και το όριο

$$V(p) = \lim_{n\to\infty} V_n(p),\ 0 \le p \le 1.$$

Είναι αύξουσα,κοίλη και συνεχής.(Η συνέχεια προκύπτει από το γεγονός ότι η σύγκλιση είναι ομαλή).

Επομένως η συνάρτηση ελέγχου της άριστης πολιτικής δ^∞ έχει την ακόλουθη δομή:

$$\delta(p) = \begin{cases} 0\ (\textit{συμβατική μέθοδος})\ \alpha\nu\ 0 \le p \le p* \\ 1\ (\textit{εξειδικευμένη μέθοδος})\ \alpha\nu\ p^* < p \le 1 \end{cases}$$

Η πολιτική δ^{∞} με δ(p)=0, $\forall\ 0 \le p \le 1$ είναι άριστη, και η W/[0,1] είναι άριστη συνάρτηση τιμών, αν η συνάρτηση W ικανοποιεί την εξίσωση αριστοποίησης (9.2.2),δηλαδή αν

$$W(p) = \min\{C.p + \beta.W(T(p)), R + \beta.W(\lambda)\} \quad , 0 \le p \le 1 , \qquad \underline{9.2.4}$$

Επειδή $W(p)=C.p+\beta.W(T(p) \quad , 0 \le p \le 1,$

η (9.2.4) ισχύει τότε και μόνο τότε αν:

$$W(p) \le R + \beta W(\lambda), 0 \le p \le 1 \qquad \underline{9.2.5}$$

Η (9.2.5) γράφεται:

$$\frac{C}{1-\beta} - \frac{C.(1-p)}{1-\beta.(1-\lambda)} \le R + \beta.(\frac{C}{1-\beta} - \frac{C.(1-\lambda)}{1-\beta.(1-\lambda)}), 0 \le p \le 1.$$

Επειδή το αριστερό μέλος είναι αύξουσα συνάρτηση του p, η (9.2.5) ισχύει αν και μόνον αν ισχύει για p=1.Τελικά, αναγκαία και ικανή συνθήκη ώστε η συνάρτηση W/[0,1] να ικανοποιεί την εξίσωση αριστοποίησης (9.2.4),και η πολιτική δ^{∞} με δ(p)=0, $0 \le p \le 1$ να είναι άριστη ,είναι η **(Σ)**

$$\frac{C}{1-\beta.(1-\lambda)} \le R. \qquad \textbf{(Σ)}$$

Στη συνέχεια της ενότητας αυτής υποθέτουμε ότι:

$$(C <)R < \frac{C}{1-\beta.(1-\lambda)}. \qquad \underline{9.2.6}$$

Τότε η κρίσιμη ποσότητα που συνδέεται με την άριστη πολιτική δ^{∞} είναι $p^* < 1$, και η άριστη συνάρτηση κόστους $V(p), 0 \le p \le 1$ ικανοποιεί την εξίσωση:

$$V(p)=C.p+\beta.V(T(p)), \qquad 0 \le p \le p^*$$

$$V(p)=V(1) \equiv R+\beta.V(\lambda) \qquad p^* < p \le 1. \qquad \underline{9.2.7}$$

Επειδή η συνάρτηση $V(p), 0 \le p \le 1$ είναι συνεχής, από την (9.2.7) έχουμε:

$$V(1)=V(p^*)=C.p^* + \beta.V(T(p^*))$$

Επειδή $T(p^*) \equiv \lambda + (1-\lambda).p* = p^* + (1-\lambda).p^* > p^*$, έχουμε $: V(T(p^*)) = V(1)$,

και επομένως,

$$V(1) = C.p^* + \beta V(1) \,. \qquad \underline{9.2.8}$$

Πρόταση 9.2.2: Για την κρίσιμη ποσότητα p^* που συνδέεται με την άριστη πολιτική δ^∞ ισχύει $p^* \geq \lambda$.

Απόδειξη

Θεωρούμε ότι $p^* < \lambda$. Θα εφαρμόσουμε εις άτοπον απαγωγή. Από την (9.2.7) έχουμε:

$$V(\lambda) = V(1) = R + \beta V(\lambda) \,.$$

Επομένως $$V(1) = V(\lambda) = \frac{R}{1-\beta} \,.$$

Από την (9.2.8) προκύπτει

$$\rho^* = \frac{(1-\beta)V(1)}{C} = \frac{R}{C}$$

και επειδή $\rho^* < 1$ συνάγεται ότι $R < C$, που αντιβαίνει στην βασική μας υπόθεση $C < R$.

Επομένως $p^* \geq \lambda$. □

Επειδή $p^* \geq \lambda. = 1-(1-\lambda)$, έπεται ότι

$$\left\{ n \in \mathbb{N} : p^* \geq 1-(1-\lambda)^n \right\} \neq \varnothing \,.$$

Έστω $$N := \max\{ n \in \mathbb{N} : p^* \geq 1-(1-\lambda)^n \} \,. \qquad \underline{9.2.9}$$

Πρόταση 9.2.3: Έστω ρ^* η κρίσιμη ποσότητα που συνδέεται με την άριστη πολιτική δ^∞ και **N** ο φυσικός αριθμός που ορίζεται από τη σχέση (9.2.9). Αν $p \in [0, p^*]$, τότε υπάρχει (μοναδικό) $m \in \{1, 2, ..., N+1\}$ έτσι ώστε:

$$T_{m-1}(p) \leq \rho^* < T_m(p) \qquad \underline{9.2.10}$$

$$(T_0(p) \equiv p)$$

Επιπλέον αν $m \neq N+1$, τότε η (9.2.10) ισοδυναμεί με τη σχέση

$$1-\frac{1-p^*}{(1-\lambda)^m} < p \le 1-\frac{1-p^*}{(1-\lambda)^{m-1}}.$$

ενώ αν $m = N+1$ τότε η (9.2.10) ισοδυναμεί με τη σχέση

$$0 \le p \le 1-\frac{1-p^*}{(1-\lambda)^N}.$$

Απόδειξη

Από το λήμμα 9.2.1 συνάγεται ότι

$$T_n(p) = 1-(1-\lambda)^n.(1-p) \uparrow 1 \text{ όταν το } n \to \infty$$

Επομένως, επειδή $p^* < 1$ (λόγω της υπόθεσης (9.2.6)),

$$\{n \in \mathbb{N} : T_n(p) > p^*\} \neq \varnothing.$$

Η σχέση (9.2.10) ικανοποιείται προφανώς για

$$m := \min\{n \in \mathbb{N} : T_n(p) > p^*\}$$

και γράφεται

$$1-(1-\lambda)^{m-1}.(1-p) \le p^* < 1-(1-\lambda)^m.(1-p)$$

ή ισοδύναμα

$$1-\frac{1-p^*}{(1-\lambda)^m} < p \le 1-\frac{1-p^*}{(1-\lambda)^{m-1}} \qquad \underline{9.2.11}$$

Θα δείξουμε ότι $m \le N+1$. Θεωρούμε ότι $m > N+1$. Τότε $m-1 > N$ και από τον ορισμό του N (σχέση (9.2.9)) έχουμε $(1-\lambda)^{m-1} < 1-p^*$, δηλαδή

$$1-\frac{1-p^*}{(1-\lambda)^{m-1}} < 0.$$

Λαμβάνοντας υπόψη την (9.2.11) προκύπτει $p < 0$, πράγμα άτοπο. Άρα $m \le N+1$. Αν $m = N+1$, τότε επειδή $m > N$ έχουμε

$$1-\frac{1-p^*}{(1-\lambda)^{N+1}} < 0 \le p \le 1-\frac{1-p^*}{(1-\lambda)^N}.$$

□

Με τη βοήθεια της πρότασης 9.2.3 θα δείξουμε ότι η άριστη πολιτική δ^{∞} επάγει μία πεπερασμένη ***Μαρκοβιανή διαμέριση***.

Θεωρούμε τα σύνολα $A_0=(p^*,1]$,

$$A_n=\{\ p\in[0,p^*]:\ T_{n-1}(p)\le p^*\ <T_n(p)\}=(1-\frac{1-p^*}{(1-\lambda)^n},1-\frac{1-p^*}{(1-\lambda)^{n-1}}],\ n=1,2,3\ldots,N$$

$$A_{N+1}=\{\ p\in[0,p^*]:T_N(p)\le p^*<T_{N+1}(p)\}=[0,\ 1-\frac{1-p^*}{(1-\lambda)^N}].$$

Τα σύνολα $A_0,A_1,A_2,\ldots,A_{N+1}$ αποτελούν διαμέριση του διαστήματος $[0,1]$.

Το σύνολο A_n $(n\ge 1)$ είναι το σύνολο των *a priori* πιθανοτήτων p αναφορικά με την κατάσταση 2 για τις οποίες εφαρμόζουμε n-1 συνεχόμενες φορές την συμβατική μέθοδο διδασκαλίας ($\alpha=0$) και την n-στη φορά την εξειδικευμένη μέθοδο διδασκαλίας ($\alpha=1$), χρησιμοποιώντας την άριστη πολιτική.

Από τον ορισμό του N (βλέπε (9.2.9)) συνάγεται ότι:

$$\lambda\in A_N. \qquad \underline{\textbf{9.2.12}}$$

Πράγματι έχουμε

$$T_{N-1}(\lambda)=1-(1-\lambda)^N\le p^*$$

και

$$T_N(\lambda)=1-(1-\lambda)^{N+1}>p^*,$$

Δηλαδή:

$$T_{N-1}(\lambda)\le p^*<T_N(\lambda).$$

Παρατηρούμε ότι για n=1,2,…,N+1,

$$p\in A_n\Rightarrow T(p,\theta=1,\alpha=0)=T(p,\theta=2,\alpha=0)=T(p)\in A_{n-1},$$

$$p\in A_0\Rightarrow T(p,\theta=1,\alpha=1)=T(p,\theta=2,\alpha=1)=\lambda\in A_N.$$

Επομένως η παραπάνω διαμέριση του διαστήματος $[0,1]$,είναι **Μαρκοβιανή** που επάγεται από την άριστη πολιτική δ^{∞}.Συμπεραίνουμε ότι η άριστη συνάρτηση τιμών $V(p),0\le p\le 1$ είναι κατά τμήματα γραμμική (βλέπε Κεφ.5).

Έστω $p\in A_n$ για κάποιο $n\in\{1,2,3,\ldots,N+1\}$.

Τότε:

$$V(p)=C.\sum_{i=0}^{n-1}\beta^i.T_i(p)+\beta^n V(T_n(p))$$

$$= C.\sum_{i=0}^{n-1}\beta^i.(1-(1-p).(1-\lambda)^i) + \beta^n.V(1)$$

$$= C.\left\{ \sum_{i=0}^{n-1}\beta^i - (1-p).\sum_{i=0}^{n-1}\beta^i.(1-\lambda)^i \right\} + \beta^n.V(1)$$

$$= C.\left(\frac{1-\beta^n}{1-\beta} - (1-p).\frac{1-\beta^n.(1-\lambda)^n}{1-\beta.(1-\lambda)} \right) + \beta^n.V(1). \qquad \underline{9.2.13}$$

Στην έκφραση (9.2.13) μπορούμε να ενσωματώσουμε και την περίπτωση n=0,δεδομένου ότι για $p \in A_0 = (p^*, 1]$ έχουμε $V(p) = V(1)$.Επομένως η άριστη συνάρτηση τιμών V(p), $0 \le p \le 1$ γράφεται:

$$V(p) = B_n.p + \Gamma_n \quad , \forall p \in A_n \qquad \underline{9.2.14}$$

όπου $$B_n = \frac{1-\beta^n.(1-\lambda)^n}{1-\beta.(1-\lambda)} \cdot C,$$

$$\Gamma_n = C.\left(\frac{1-\beta^n}{1-\beta} \right) + \beta^n.V(1) - B_n \text{ , n=0,1,2...,N+1} \qquad \underline{9.2.15}$$

Σημειώνουμε ότι: B_0=0, Γ_0=V(1) .

Η εξάρτηση της συνάρτησης V(p), $0 \le p \le 1$ από τις άγνωστες παραμέτρους p^* και V(1) (που συνδέονται με τη σχέση (9.2.8)) μας παροτρύνει να τις μελετήσουμε διεξοδικά με σκοπό να τις υπολογίσουμε.

Εφαρμόζοντας τη σχέση (9.2.13) για $p = \lambda$ και λαμβάνοντας υπόψη την (9.2.12) έχουμε:

$$V(\lambda) = C.\left(\frac{1-\beta^N}{1-\beta} - (1-\lambda).\frac{1-\beta^N.(1-\lambda)^N}{1-\beta.(1-\lambda)} \right) + \beta^N.V(1)$$

και αντικαθιστώντας στη σχέση: V(1)= $R + \beta V(\lambda)$

παίρνουμε τελικά

$$V(1) = \frac{1}{1-\beta^{N+1}} \left\{ R + \beta.C.\left(\frac{1-\beta^N}{1-\beta} - (1-\lambda).\frac{1-\beta^N.(1-\lambda)^N}{1-\beta(1-\lambda)} \right) \right\} \qquad \underline{9.2.16}$$

και από την (9.2.8),

$$p^* = \frac{(1-\beta)V(1)}{C} \qquad \underline{9.2.17}$$

Άρα ο υπολογισμός των p^* και V(1) ανάγεται στον προσδιορισμό του φυσικού αριθμού N, από τον οποίο αυτές οι ποσότητες αποκλειστικά εξαρτώνται. Από τις (9.2.9),(9.2.16),(9.2.17),το πρόβλημα ανάγεται στο να βρεθεί ο μέγιστος φυσικός αριθμός N έτσι ώστε:

$$p^* = \frac{1-\beta}{(1-\beta^{N+1}).C}\left\{R+\beta.C.\left(\frac{1-\beta^N}{1-\beta}-(1-\lambda).\frac{1-\beta^N.(1-\lambda)^N}{1-\beta(1-\lambda)}\right)\right\} \geq 1-(1-\lambda)^N$$

9.2.18

Μετά από πράξεις αποδεικνύεται ότι η (9.2.18) είναι ισοδύναμη με την

$$(1\text{-}\lambda)^N.(1\text{-}\frac{\lambda.\beta^{N+1}}{1-\beta(1-\lambda)}) \geq \frac{1-\beta}{C}.(\frac{C}{1-\beta.(1-\lambda)}-R). \qquad \underline{9.2.19}$$

Επομένως το πρόβλημα ανάγεται στην εύρεση του μέγιστου φυσικού αριθμού N που ικανοποιεί την σχέση (9.2.19).

Αλγόριθμος A_5

ΒΗΜΑ 0:Σημειώνουμε τις παραμέτρους του προβλήματος λ, β,C,R.

Αν ισχύει η συνθήκη (Σ), δηλαδή $:\frac{C}{1-\beta.(1-\lambda)} \leq R$,τότε η άριστη συνάρτηση τιμών υπολογίζεται από τη σχέση (9.2.3) και η πολιτική δ^∞ με δ(p)=0, $0 \leq p \leq 1$ είναι άριστη και η διαδικασία τερματίζεται . Αν η συνθήκη (Σ) δεν ισχύει προχωρούμε στα επόμενα βήματα.

ΒΗΜΑ 1:Προσδιορίζουμε τον μέγιστο φυσικό αριθμό N που ικανοποιεί τη σχέση (9.2.19). Κατόπιν υπολογίζουμε τις ποσότητες V(1) και ρ* από τις (9.2.16) και (9.2.17).

ΒΗΜΑ 2: Προσδιορίζουμε τα διαστήματα

$$A_0=(p^*,1],\ A_n=(\alpha_n,\alpha_{n-1}],\ n=1,2,3\ldots N,\ A_{N+1}=[0,\alpha_N]$$

Όπου :
$$\alpha_n=1-\frac{1-p^*}{(1-\lambda)^n},n=0,1,...,N$$

$$(0<a_N<a_{N-1}<...<a_1<a_0=p^*<1)$$

ΒΗΜΑ 3: Υπολογίζουμε την άριστη συνάρτηση κόστους $V(p), 0\le p\le 1$ με βάση τις σχέσεις (9.2.14) και (9.2.15).

Εφαρμογή 9.1: Ας θεωρήσουμε τώρα μια περίπτωση που εμπίπτει στην πλήρη αβεβαιότητα με β=0.9, λ=0.1, C= 4.0, R=10.0.

Στο βήμα 1 η ανίσωση (9.2.19) γράφεται:

$$0.9^N.(1-\frac{0.9^{N+1}}{0.19})\ge 0.276315775$$

Ο μέγιστος ακέραιος που πληροί την παραπάνω ανίσωση : N=10.

Κατόπιν με βάση την σχέση (9.2.16) βρίσκουμε V(1)=26.9140,και με βάση τη σχέση (9.2.17), η κρίσιμη ποσότητα είναι $p^*=0.6728$.

Εφαρμόζοντας το βήμα 2 τα διαιρετικά σημεία της Μαρκοβιανής διαμέρισης είναι: $\alpha_0=0.6728,\alpha_1=0.6364,\alpha_2=0.5960,\alpha_3=0.5511,\alpha_4=0.4961,\alpha_5=0.4458,\alpha_6=0.3843,\alpha_7=0.3157,$ $\alpha_8=0.2399,\ \alpha_9=0.1549,\ \alpha_{10}=0.0616$. Τέλος εφαρμόζοντας το βήμα 3 προκύπτει η βέλτιστη συνάρτηση κόστους για άπειρο χρονικό ορίζοντα, V(p)/[0,1]:

$$V(p)=\begin{cases}18.9794.p+16.9141 & p\in[0.00000,0.062]\\ 18.4931.p+16.9441 & p\in(0.062,\ 0.1549]\\ 17.8927.p+17.0375 & p\in(0.1549,0.2399]\\ 17.1515.p+17.2154 & p\in(0.2399,0.3157]\\ 16.2365.p+17.5046 & p\in(0.3157,0.3843]\\ 15.1067.p+17.9388 & p\in(0.3843,0.4458]\\ 13.7120.p+18.5608 & p\in(0.4458,0.4961]\\ 11.9902.p+19.4241 & p\in(0.4961,\ 0.5511]\\ 9.8644.p+20.5954 & p\in(0.5511,0.5960]\\ 7.2400.p+22.1604 & p\in(0.5960,0.6364]\\ 4.0000.p+24.2227 & p\in(0.6364,0.6728]\\ 26.9140 & p\in(0.6728,1.000]\end{cases}$$

9.3. Περίπτωση πλήρους αβεβαιότητας στην συμβατική μέθοδο και μερικής πληροφόρησης στην εξειδικευμένη μέθοδο διδασκαλίας.

Υποθέτουμε ότι $q^0 = \frac{1}{2}$, $q^1 > \frac{1}{2}$.

Οι πίνακες μηνυμάτων γράφονται:

$$R^0 = \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix}, \quad R^1 = \begin{pmatrix} q^1 & 1-q^1 \\ 1-q^1 & q^1 \end{pmatrix}.$$

Επομένως αν επιλέξουμε τη συμβατική μέθοδο διδασκαλίας (α=0) το μήνυμα που λαμβάνεται σε οποιαδήποτε χρονική περίοδο είναι ανεξάρτητο από την κατάσταση της τάξης και συνεπώς δεν παρέχει καμιά πληροφορία για την κατάσταση της τάξης(πλήρης αβεβαιότητα για $a = 0$). Από τις σχέσεις (9.1.1)- (9.1.4) παίρνουμε:

$$T(p,\theta=1,\alpha=0) = T(p,\theta=2,\alpha=0) = T(p) = p(1-\lambda)+\lambda \ , 0 \le p \le 1 .$$

$$\{\theta/p,\alpha=0\} = \frac{1}{2} \ ,\theta=1,2 \ ,\alpha=0,1, \ 0 \le p \le 1.$$

Τα $T(p,\theta=1,\alpha=1)$, $\{\theta=1/p,\alpha=1\}$,$T(p,\theta=2,\alpha=1)$, $\{\theta=2/p,\alpha=1\}$ δίνονται από τις σχέσεις (9.1.5)-(9.1.8).

Λήμμα 9.3.1:

i) $s_1 < \lambda < s_2$

ii) $\lambda = \gamma_1 . s_1 + \gamma_2 . s_2$

Απόδειξη

Το (i) είναι συνέπεια της υπόθεσης $q^1 > 1/2$ και το (ii) προκύπτει άμεσα από τους ορισμούς των $\gamma_1, \gamma_2, s_1, s_2$, (βλέπε (9.1.5)-(9.1.8)). □

Η εξίσωση αριστοποίησης (9.1.10) γράφεται στην περίπτωσή μας

$$V(p) = \min\{C.p + \beta V(T(p)), R + \beta.\gamma_1 V(s_1) + \beta.\gamma_2 V(s_2)\}, 0 \le p \le 1 \qquad \underline{\mathbf{9.3.1}}$$

Η συνάρτηση $V(p), 0 \le p \le 1$ είναι αύξουσα, κοίλη και συνεχής. Πράγματι θεωρούμε την επαναληπτική σχέση:

$$V_n(p) = HV_{n-1}(p) = \min\{C.p + \beta V_{n-1}(T(p)), R + \beta.\gamma_1 V_{n-1}(s_1) + \beta.\gamma_2 V_{n-1}(s_2)\}$$

οπου $V_0(p) = 0, 0 \le p \le 1.$

Αποδεικνύεται εύκολα με επαγωγή ότι για κάθε n=1,2,3... η συνάρτηση $V_n(p), 0 \le p \le 1$ είναι αύξουσα, κοίλη και συνεχής. Επομένως και το όριο: $V(p) = \lim_{n \to \infty} V_n(p), 0 \le p \le 1$ είναι αύξουσα, κοίλη και συνεχής συνάρτηση. (Η συνέχεια προκύπτει από το γεγονός ότι η σύγκλιση είναι ομαλή).

Επομένως η συνάρτηση ελέγχου της άριστης πολιτικής δ^∞ έχει την μορφή:

$$\delta(p) = \begin{cases} 0 \ (\textit{συμβατική μέθοδος}), \alpha\nu \ 0 \le p \le p^* \\ 1 \ (\textit{εξειδικευμένη μέθοδος}), \alpha\nu \ p^* < p \le 1. \end{cases}$$

όπου p^* είναι κατάλληλη κρίσιμη ποσότητα.

Στην ενότητα αυτή θα εξετάσουμε μια αναλυτική μέθοδο υπολογισμού της κρίσιμης ποσότητας p^* και της βέλτιστης συνάρτησης κόστους, $V(p), 0 \le p \le 1$. Θα εξετάσουμε όμως πρώτα υπό ποίες συνθήκες ισχύει $p^*=1$, δηλαδή η πολιτική δ^∞ με συνάρτηση ελέγχου $\delta(p)=0, 0 \le p \le 1$ (να χρησιμοποιούμε πάντοτε την συμβατική μέθοδο) είναι άριστη.

Εστω $W(p), 0 \le p \le 1$ η συνάρτηση αναμενόμενου ολικού εκπίπτοντος κόστους για άπειρο χρονικό ορίζοντα, όταν εφαρμόζουμε την πολιτική δ^{∞} με δ(p)=0, $0 \le p \le 1$, (δηλαδή όταν χρησιμοποιούμε αποκλειστικά την συμβατική μέθοδο διδασκαλίας για κάθε $0 \le p \le 1$). Αυτή υπολογίζεται όπως ακριβώς στην ενότητα 9.2 και δίνεται από τη σχέση:

$$W(p) = \frac{C}{1-\beta} - \frac{C.(1-p)}{1-\beta.(1-\lambda)}. \qquad \underline{9.3.2}$$

Η πολιτική δ^{∞} με δ(p)=0, $0 \le p \le 1$ είναι άριστη και η W/[0,1] είναι άριστη συνάρτηση τιμών αν η W ικανοποιεί την εξίσωση αριστοποίησης (9.3.1).Επειδή

$$W(p) = C.p + \beta.W(T(p)), 0 \le p \le 1$$

η W ικανοποιεί την (9.3.1) αν και μόνο αν

$$W(p) \le R + \beta.\gamma_1.W(s_1) + \beta.\gamma_2.W(s_2), 0 \le p \le 1 \qquad \underline{9.3.3}$$

Επειδή η συνάρτηση W/[0,1] είναι γραμμική και λαμβάνοντας υπόψη το λήμμα 9.3.1 (ii) και το γεγονός ότι $\gamma_1 + \gamma_2 = 1$,έχουμε:

$$\gamma_1.W(s_1) + \gamma_2.W(s_2) = W(\gamma_1.s_1 + \gamma_2.s_2) = W(\lambda)$$

Επομένως η (9.3.3) γράφεται

$$W(p) \le R + \beta.W(\lambda), 0 \le p \le 1$$

και ικανοποιείται αν και μόνο αν ισχύει η συνθήκη **(Σ)** (βλέπε ενότητα 9.2):

$$\frac{C}{1-\beta.(1-\lambda)} \le R, \qquad \textbf{(Σ)}$$

Συνοψίζοντας, αναγκαία και ικανή συνθήκη ώστε η πολιτική δ^{∞} *με* $\delta(p)=0, 0 \le p \le 1$ *(αποκλειστική χρήση της συμβατικής μεθόδου) να είναι άριστη και η* $W/[0,1]$ *άριστη συνάρτηση τιμών είναι* η (Σ).

Στη συνέχεια της ενότητας αυτής υποθέτουμε ότι :

$$(C <)\mathrm{R} < \frac{C}{1-\beta.(1-\lambda)}. \qquad \underline{\textbf{9.3.4}}$$

Τότε η κρίσιμη ποσότητα που συνδέεται με την άριστη πολιτική δ^{∞} είναι $p^* < 1$ και η άριστη συνάρτηση τιμών, $V(p), 0 \le p \le 1$ ικανοποιεί την εξίσωση:

$$V(p) = C.p + \beta V(T(p)), 0 \le p \le p^*$$

$$\mathrm{V(p)} = \mathrm{V(1)} \equiv \mathrm{R} + \beta.\gamma_1 V(s_1) + \beta.\gamma_2 V(s_2), p^* < p \le 1 \qquad \underline{\textbf{9.3.5}}$$

Επειδή η συνάρτηση $V(p), 0 \le p \le 1$ είναι συνεχής, από την (9.3.5) παίρνουμε :

$$\mathrm{V(1)} = \mathrm{V(p^*)} = \mathrm{C.p^* + \beta.V(T(p^*))}$$

Επειδή μάλιστα $T(p^*) > p^*$, έχουμε: $V(T(p^*)) = V(1)$ και επομένως

$$\mathrm{V\,(1) = C.p^* + \beta.\ V(1)} \qquad \underline{\textbf{9.3.6}}$$

Πρόταση 9.3.1: Αν $0 \le p \le s_1$, τότε η άριστη απόφαση που αντιστοιχεί στο p είναι η συμβατική μέθοδος διδασκαλίας, δηλαδή δ(p)=0.

Απόδειξη

As υποθέσουμε ότι για την κρίσιμη ποσότητα p^* της άριστης πολιτικής ισχύει $p^* \prec s_1$. Επειδή $s_2 \succ s_1$ (λήμμα 9.3.1), έχουμε $s_2 \succ p^*$.

Αρα $V(s_1)= V(s_2)= V(1)$ και επομένως :

$$V(1)=R+\beta.\gamma_1.V(s_1)+\beta.\gamma_2.V(s_2)$$

$$=R+\beta.(\gamma_1+\gamma_2).\ V(1)= R+\beta.V(1)$$

από την οποία προκύπτει : $$V(1)=\frac{R}{1-\beta}$$

Λαμβάνοντας υπόψη την (9.3.6) προκύπτει $R=C.p^*$ ($< C$) ,πού αντιβαίνει στην υπόθεση $C < R$. Επομένως $p^* \geq s_1$ ή ισοδύναμα, $\delta(p)=0,\ 0 \leq p \leq s_1$. □

Πρόταση 9.3.2:Για την κρίσιμη ποσότητα p^* της άριστης πολιτικής ισχύει $p^* \geq \lambda$.

Απόδειξη

Εστω $p^* < \lambda$.Θα εφαρμόσουμε εις άτοπον απαγωγή. Έχουμε:

$$T(p)= p(1-\lambda)+\lambda > \lambda > p^*, 0 \leq p \leq 1.$$

Επομένως $$V(p) = C.p + \beta.V(T(p)) = C.p + \beta.V(1), 0 \leq p \leq p^*.$$

Επειδή $s_1 \leq p^*$ (πρόταση 9.3.1) και $s_2 > \lambda > p^*$ (λήμμα 9.3.1),

$$V(s_1) = C.s_1+\beta.V(T(s_1))= C.s_1+\beta.V(1),\ V(s_2)= V(1).$$

Επομένως $$V(1) = R+\beta.\gamma_1.V(s_1)+\beta.\gamma_2.V(s_2)=$$

$$= R+\beta.\gamma_1.s_1.C+(\beta^2.\gamma_1+\beta.\gamma_2).V(1),$$

από την οποία προκύπτει:

$$V(1) = \frac{R+\beta.\gamma_1.s_1.C}{1-\beta^2.\gamma_1-\beta.\gamma_2}$$

Ο παρανομαστής γράφεται:

$1-\beta^2.\gamma_1-\beta.\gamma_2=1-\beta^2.\gamma_1-\beta.(1-\gamma_1)=(1-\beta).(1+\beta.\gamma_1)$,οπότε λαμβάνοντας υπόψη την (9.3.6),

παίρνουμε: $p^* = \dfrac{(1-\beta).V(1)}{C} = \dfrac{R+\beta.\gamma_1.s_1.C}{(1+\beta.\gamma_1).C}$

Έχουμε : $p^* < \lambda \Leftrightarrow R+\beta.\gamma_1.s_1.C < \lambda.(1+\beta.\gamma_1).C \Leftrightarrow$

$R+\beta\lambda.(1-q^1).C < \lambda.(1+\beta.(1-\lambda).q^1+\beta.\lambda(1-q^1).C \Leftrightarrow R < \lambda.(1+\beta.(1-\lambda)(2.q^1-1)).C$

Επειδή όμως $2.q^1-1 \leq 1$ έχουμε:

$$\lambda.(1+\beta.(1-\lambda)(2.q^1-1)) \leq \lambda+\beta.\lambda.(1-\lambda) \prec \lambda+1-\lambda=1$$

Συμπεραίνουμε ότι $p^* < \lambda \Rightarrow R < C$, που είναι άτοπο επειδή αντιβαίνει στην υπόθεση $C \prec R$. Επομένως $p^* \geq \lambda$. □

Επειδή $p^* \geq \lambda = 1-(1-\lambda)$ έπεται ότι

$$\{n \in \mathbb{N} : p^* \geq 1-(1-\lambda)^n\} \neq \varnothing.$$

Εστω $N=\max\{n \in \mathbb{N} : p^* \geq 1-(1-\lambda)^n\}$. **9.3.7**

Το διάστημα [0,1] διαμερίζεται με τον ίδιο ακριβώς τρόπο όπως στην ενότητα 9.2 (βλ. πρόταση 9.2.3)

Θεωρούμε τα σύνολα $A_0=(p^*,1]$,

$$A_n=\{ p \in [0,p^*]: T_{n-1}(p) \leq p^* < T_n(p)\} = (1-\frac{1-p^*}{(1-\lambda)^n}, 1-\frac{1-p^*}{(1-\lambda)^{n-1}}], \ n=1,2,3\ldots,N$$

$$A_{N+1}=\{ p \in [0,p^*]: T_N(p) \leq p^* < T_{N+1}(p)\} = [0, 1-\frac{1-p^*}{(1-\lambda)^N}],$$

$$(\text{όπου } T_0(p)=p, \ 0 \leq p \leq 1).$$

Τα σύνολα $A_0, A_1, A_2, \ldots, A_{N+1}$ αποτελούν μία διαμέριση του διαστήματος [0,1]. Το σύνολο A_n $(n \geq 1)$ είναι το σύνολο των *a priori* πιθανοτήτων p αναφορικά με την κατάσταση 2, για τις οποίες εφαρμόζουμε n-1 συνεχόμενες φορές την συμβατική μέθοδο διδασκαλίας(α=0)και την n-στη φορά την εξειδικευμένη μέθοδο διδασκαλίας (α=1), χρησιμοποιώντας την άριστη πολιτική.

Έστω $s_1 \in A_k, s_2 \in A_l$.

Παρατηρούμε ότι για n=1,2,…,N+1,

$$p\in A_n \Rightarrow T(p,\theta=1,\alpha=0)=T(p,\theta=2,\alpha=1)=T(p)\in A_{n-1},$$

$$T(p,\theta=1,\alpha=1)=s_1\in A_k .$$

$$p\in A_0 \Rightarrow$$

$$T(p,\theta=2,\alpha=1)=s_2\in A_l .$$

Επομένως η παραπάνω διαμέριση του διαστήματος [0,1] είναι ***Μαρκοβιανή*** που επάγεται από την άριστη πολιτική δ^∞. Συμπεραίνουμε ότι η άριστη συνάρτηση τιμών $V(p), 0\le p\le 1$ είναι κατά τμήματα γραμμική (βλέπε κεφάλαιο 5).

Εστω $p\in A_n$ για κάποιο $n\in\{1,2,3,\ldots,N+1\}$. Τότε

$$V(p)=C.\sum_{i=0}^{n-1}\beta_i.T_i(p)+\beta^n.V(T_n(p))=C.\sum_{i=0}^{n-1}\beta^i.(1-(1-p).(1-\lambda)^i)+\beta^n.V(1)$$

$$=C.(\frac{1-\beta^n}{1-\beta}-(1-p)\frac{1-\beta^n.(1-\lambda)^n}{1-\beta.(1-\lambda)})+\beta^n.V(1). \qquad \underline{9.3.8}$$

Στην έκφραση (9.3.8) μπορούμε να ενσωματώσουμε και την περίπτωση n=0,δεδομένου ότι για $p\in A_0=(p^*,1]$ έχουμε V(p)= V(1).

Επομένως η βέλτιστη συνάρτηση τιμών $V(p), 0\le p\le 1$ γράφεται

$$V(p)=B_n.p+\Gamma_n \quad \forall\ p\in A_n \qquad \underline{9.3.9}$$

όπου

$$B_n=\frac{1-\beta^n.(1-\lambda)^n}{1-\beta.(1-\lambda)}.C ,$$

$$\Gamma_n=\frac{1-\beta^n}{1-\beta}.C-B_n+\beta^n.V(1),\ n=0,1,2,3,\ldots,N+1. \qquad \underline{9.3.10}$$

Σημειώνουμε ότι : $B_0=0$, $\Gamma_0=V(1)$.

Η εξάρτηση της συνάρτησης $V(p), 0\le p\le 1$ από τις άγνωστες παραμέτρους p*, V(1)(που συνδέονται με την σχέση (9.3.6)) μας παροτρύνει να τις μελετήσουμε διεξοδικά με την προσδοκία να βρούμε τρόπο υπολογισμού αυτών.

Εστω $s_1 \in A_k$,$s_2 \in A_l$

Εφαρμόζοντας την σχέση (9.3.9) για p= s_1, s_2 έχουμε:

$$V(s_1)=C.(\frac{1-\beta^k}{1-\beta}-(1-s_1)\frac{1-\beta^k.(1-\lambda)^k}{1-\beta.(1-\lambda)})+\beta^k.V(1)$$

$$V(s_2)=C.(\frac{1-\beta^l}{1-\beta}-(1-s_2)\frac{1-\beta^l.(1-\lambda)^l}{1-\beta.(1-\lambda)})+\beta^l.V(1).$$

Αντικαθιστώντας στην σχέση $V(1)=R+\beta.\gamma_1.V(s_1)+\beta.\gamma_2.V(s_2)$

παίρνουμε τελικά:

$$V(1)=\frac{1}{1-\gamma_1.\beta^{k+1}-\gamma_2.\beta^{k+1}}.[R+\beta.\{\gamma_1.(\frac{1-\beta^k}{1-\beta}-(1-s_1).\frac{1-\beta^k.(1-\lambda)^k}{1-\beta.(1-\lambda)})+\gamma_2.(\frac{1-\beta^l}{1-\beta}-(1-s_2).\frac{1-\beta^l.(1-\lambda)^l}{1-\beta.(1-\lambda)})\}.C],$$

Επίσης από την (9.3.6) παίρνουμε:

$$p^*=\frac{(1-\beta).V(1)}{C}$$

Αρα ο υπολογισμός των p*, V(1) ανάγεται στην εύρεση των κατάλληλων ακεραίων k,l από τους οποίους εξαρτώνται. Η πρόταση που ακολουθεί δίνει σχέσεις των k,l που αξιοποιούνται στην υπολογιστική διαδικασία.

Πρόταση 9.3.3: Εστω $s_1 \in A_k$,$s_2 \in A_l$ και $m=\min\{n \in \mathbb{N}_0 : s_2 < T_n(s_1)\}$.

Τότε ισχύουν:

i) $m \geq 1, k \geq 1, l \leq k$

ii) Αν $l \geq 1$, τότε $k=l+m-1$ ή $k=l+m$.

iii) Αν $l=0$, τότε $1 \leq k \leq m$.

Απόδειξη

i) Επειδή $T_0(s_1) \equiv s_1 < s_2$ έχουμε προφανώς $m \geq 1$. Επειδή $\lambda \leq p^*$ (πρόταση 9.3.2) και $s_1 < \lambda$ (λήμμα 9.3.1) έχουμε ότι $s_1 \notin A_0 = (p^*, 1]$, δηλαδή $k \geq 1$.

Επειδή $s_1 < s_2$ έχουμε

$$T_{k-1}(s_1) \leq p^* < T_k(s_1) < T_k(s_2).$$

Επομένως $l \leq k$. Σημειώνουμε ότι ενδέχεται $l = 0$ $(s_2 \in A_0)$.

ii) Από τον ορισμό του m και επειδή $m \geq 1$, έχουμε: $T_{m-1}(s_1) \leq s_2 < T_m(s_1)$.

Επειδή $s_1 \in A_k$, $s_2 \in A_l$, $k, l \geq 1$ έχουμε:

$$T_{k-1}(s_1) \leq p^* < T_k(s_1), \quad T_{l-1}(s_2) \leq p^* < T_l(s_2)$$

Επομένως,

$$T_{l+m-2}(s_1) = T_{l-1}(T_{m-1}(s_1)) \leq T_{l-1}(s_2) \leq p^* \prec T_l(s_2) \prec T_l(T_m(s_1)) = T_{l+m}(s_1)$$

Αρα

$$T_{l+m-2}(s_1) \leq p^* \prec T_{l+m}(s_1)$$

Υπάρχουν δύο δυνατότητες

α) $T_{l+m-2}(s_1) \leq p^* < T_{l+m-1}(s_1)$ ή

β) $T_{l+m-1}(s_1) \leq p^* < T_{l+m}(s_1)$

Επομένως $k = l + m - 1$ ή $k = l + m$.

iii) Αν $l = 0$, τότε $s_2 \in A_0 = (p^*, 1]$. Επομένως $p^* \prec s_2 \prec T_m(s_1)$ από την οποία συνάγεται ότι $k \leq m$. □

Εστω $k \geq 1$, $l \geq 0$ ακέραιοι αριθμοί,

$$p_{kl} = \frac{(1-\beta)V_{kl}(1)}{C} \qquad \underline{9.3.11}$$

όπου

$$V_{kl}(1) = \frac{1}{1-\gamma_1.\beta^{k+1}-\gamma_2.\beta^{l+1}}.[R+\beta.\{\gamma_1.(\frac{1-\beta^k}{1-\beta}-(1-s_1).\frac{1-\beta^k.(1-\lambda)^k}{1-\beta.(1-\lambda)})+\gamma_2.(\frac{1-\beta^l}{1-\beta}-(1-s_2).$$

$$\frac{1-\beta^l.(1-\lambda)^l}{1-\beta.(1-\lambda)})\}.C] \qquad \underline{9.3.12}$$

Το πρόβλημα ανάγεται στην εύρεση ενός ζεύγους ακεραίων (k,*l*) με $k \geq 1$, $l \geq 0$ έτσι ώστε να συναληθεύουν οι παρακάτω σχέσεις (9.3.13),(9.3.14).

$$T_{k-1}(s_1) \leq p_{kl} \prec T_k(s_1) \qquad \underline{9.3.13}$$

$$T_{l-1}(s_2) \leq p_{kl} \prec T_l(s_2) \quad \text{αν} \quad l \geq 1 \qquad \underline{9.3.14}$$

$$p_{k0} < s_2 \quad \text{αν} \quad l=0.$$

Τότε $p^* = p_{kl}$, $V(1)=V_{kl}(1)$.

Η ακόλουθη πρόταση παρέχει ισοδύναμες συνθήκες για τις σχέσεις (9.3.13), (9.3.14), που είναι εύχρηστες στην υπολογιστική διαδικασία.

Πρόταση 9.3.4: Εστω (k,l) ένα ζεύγος ακεραίων με $k \geq 1, l \geq 0$.

i) Για $l \geq 1, T_{l-1}(s_2) \leq p_{kl} < T_l(s_2) \Leftrightarrow \phi_1(k,l) \prec \frac{C}{1-\beta.(1-\lambda)} - R \leq \phi_2(k,l)$ όπου

$$\phi_1(k,l) \equiv (1-s_2).(1-\lambda)^l.(\frac{\gamma_2.\beta^{l+1}}{1-\beta.(1-\lambda)} + \frac{1-\gamma_1.\beta^{k+1}-\gamma_2.\beta^{l+1}}{1-\beta}).C$$

$$+(1-s_1).(1-\lambda)^k.\frac{\gamma_1.\beta^{k+1}}{1-\beta.(1-\lambda)}.C,$$

$$\phi_2(k,l) \equiv (1-s_2).(1-\lambda)^{l-1}.(\frac{\gamma_2.(1-\lambda)\beta^{l+1}}{1-\beta.(1-\lambda)} + \frac{1-\gamma_1.\beta^{k+1}-\gamma_2.\beta^{l+1}}{1-\beta}).C \; +$$

$$(1-s_1).(1-\lambda)^k.\frac{\gamma_1.\beta^{k+1}}{1-\beta.(1-\lambda)}.C$$

Για $l=0$, $\quad p_{k0} < s_2 \Leftrightarrow \phi_1(k,0) < \dfrac{C}{1-\beta.(1-\lambda)} - R$

ii) $T_{k-1}(s_1) \le p_{kl} < T_k(s_1) \Leftrightarrow f_1(k,l) \prec \dfrac{C}{1-\beta.(1-\lambda)} - R \le f_2(k,l)$

όπου

$$f_1(k,l) \equiv (1-s_1).(1-\lambda)^k.\left(\frac{\gamma_1.\beta^{k+1}}{1-\beta.(1-\lambda)} + \frac{1-\gamma_1.\beta^{k+1}-\gamma_2.\beta^{l+1}}{1-\beta}\right).C$$

$$+ (1-s_2).(1-\lambda)^l.\frac{\gamma_2.\beta^{l+1}}{1-\beta.(1-\lambda)}.C,$$

$$f_2(k,l) \equiv (1-s_1).(1-\lambda)^{k-1}.\left(\frac{\gamma_1.(1-\lambda).\beta^{k+1}}{1-\beta.(1-\lambda)} + \frac{1-\gamma_1.\beta^{k+1}-\gamma_2.\beta^{l+1}}{1-\beta}\right).C$$

$$+(1-s_2).(1-\lambda)^l.\frac{\gamma_2.\beta^{l+1}}{1-\beta.(1-\lambda)}.C. \qquad \square$$

Τα ανωτέρω αποτελέσματα προκύπτουν κατόπιν αρκετών πράξεων που τις παραλείπουμε. Στην ακόλουθη πρόταση παρουσιάζουμε υπολογιστικά χρήσιμες σχέσεις ανάμεσα στις συναρτήσεις ϕ_1, ϕ_2 και ανάμεσα στις f_1, f_2.

Πρόταση 9.3.5:

i) $\phi_2(k,l) = \phi_1(k,l) + (1-s_2).\lambda(1-\lambda)^{l-1}.\dfrac{1-\gamma_1.\beta^{k+1}-\gamma_2.\beta^{l+1}}{1-\beta}.C$

ii) $\phi_2(k,l+1) = \phi_1(k,l)$

iii) $f_2(k,l) = f_1(k,l) + (1-s_1).\lambda.(1-\lambda)^{k-1}.\dfrac{1-\gamma_1.\beta^{k+1}-\gamma_2.\beta^{l+1}}{1-\beta}.C$

iv) $f_2(k+1,l) = f_1(k,l)$

Απόδειξη

Τα (i) και (iii) προκύπτουν άμεσα από τον ορισμό των ϕ_1, ϕ_2 και f_1, f_2 αντίστοιχα.

ii) Η απόδειξη του (ii) ανάγεται στη σχέση

$$\frac{\gamma_2(1-\lambda)\beta^{l+2}}{1-\beta.(1-\lambda)}+\frac{1-\gamma_1.\beta^{k+1}-\gamma_2.\beta^{l+2}}{1-\beta}=\frac{\gamma_2.\beta^{l+1}}{1-\beta.(1-\lambda)}+\frac{1-\gamma_1.\beta^{k+1}-\gamma_2.\beta^{l+1}}{1-\beta}$$

που αποδεικνύεται εύκολα.

iv) Η απόδειξη είναι παρόμοια με εκείνη του ii) □

Συνοψίζοντας, έχουμε: $p^* = p_{kl}$, $V(1)= V_{kl}(1)$,
αν το ζεύγος των ακεραίων (k,l), $k \geq 1, l \geq 0$, ικανοποιεί τις ακόλουθες σχέσεις.

$$f_1(k,l) < \frac{C}{1-\beta.(1-\lambda)} - R \leq f_2(k,l) \qquad \underline{9.3.15}$$

$$\phi_1(k,l) \leq \frac{C}{1-\beta.(1-\lambda)} - R \leq \phi_2(k,l), \text{αν } l \geq 1 \qquad \underline{9.3.16}$$

$$\phi_1(k,0) \leq \frac{C}{1-\beta.(1-\lambda)} - R, \text{αν } l=0. \qquad \underline{9.3.17}$$

Αλγόριθμος A_6

ΒΗΜΑ 0: Σημειώνουμε τις παραμέτρους του προβλήματος q^1,λ,β,C,R.

ΒΗΜΑ 1: Αν ισχύει η συνθήκη (Σ) τότε η άριστη συνάρτηση τιμών υπολογίζεται από τη σχέση (9.2.3) η πολιτική δ^∞ με δ(p)=0, $0 \leq p \leq 1$ είναι άριστη και η διαδικασία τερματίζεται. Αν η συνθήκη (Σ) δεν ισχύει προχωράμε στα επόμενα βήματα.

ΒΗΜΑ 2: Υπολογίζουμε τα $\gamma_1,\gamma_2,s_1,s_2$ από τις σχέσεις (9.1.5)-(9.1.8) και τον αριθμό m όπως ορίσθηκε στην πρόταση 9.3.3.Παραθέτουμε τα επιτρεπτά ζεύγη (k,*l*) που υπαγορεύονται από την πρόταση 9.3.3 ακολουθώντας την εξής προφανή διάταξη: Ξεκινάμε με *l*=0 και γράφουμε όλα τα επιτρεπτά ζεύγη (k,o) με την φυσική σειρά. Συνεχίζουμε την διαδικασία με *l*=1,κατόπιν με *l*=2 κ.ο.κ. Ετσι παίρνουμε την ακόλουθη διάταξη.

(1,0), (2,0),…,(m,0),(m,1),(m+1,1),(m+1,2),(m+2,2),…,(*l*+m-1,l),(*l*+m,l),…

Ακολουθώντας τη διαδρομή κατά μήκος της διάταξης επιλέγουμε εκείνο το ζεύγος (k,*l*) πού ικανοποιεί τις σχέσεις (9.3.15),(9.3.16),(9.3.17).

ΒΗΜΑ 3: Υπολογίζουμε τα p_{kl},$V_{kl}(1)$ από τις σχέσεις (9.3.11) και (9.3.12) και παίρνουμε : $p^* = p_{kl}$, V(1)= $V_{kl}(1)$.

Προσδιορίζουμε τον αριθμό $N=\max\{n\in\mathbb{N}:(1-\lambda)^n\geq 1-p^*\}$, από τη σχέση (9.3.7), και τα διαστήματα:

$A_0=(p^*,1]$,$A_n=(\alpha_n,\alpha_{n-1}]$, n=1,2,3,…N, $A_{N+1}=[0,\alpha_N]$

$$\text{όπου } a_n = 1-\frac{1-p^*}{(1-\lambda)^n},\ n=0,1,2\ldots,N$$

$$(0 \prec a_N \prec a_{N-1} \prec \ldots \prec \alpha_1 \prec \alpha_0 = p^* \prec 1)$$

ΒΗΜΑ 4: Υπολογίζουμε την άριστη συνάρτηση κόστους $V(p), 0\leq p\leq 1$ από τις σχέσεις (9.3.9),(9.3.10).

Παράδειγμα 9.2: β=0.9 ,λ=0.1 ,q^1=0.8,C=4,R=10.

Εχουμε γ_1=0.74,γ_2=0.26, $s_1=\frac{2}{74}$, $s_2=\frac{8}{26}$

$m \equiv \min\{n: s_2 < T_n(s_1)\} = 4$,

(επειδή $T_3(s_1)=0.2907 < s_2 = 0.30769 < T_4(s_1) = 0.3616$)

Διάταξη των (k,l):(1,0),(2,0),(3,0),(4,0), (4,1),(5,1),(5,2),(6,2), (6,3),…

$$\frac{C}{1-\beta.(1-\lambda)} - R = \frac{4}{0.19} - 10 = 11.05263 \equiv d$$

(k=1,l=0) $\phi_1(1,0) = 19.07417004 > d$

(k=2,l=0) $\phi_1(2,0) = 18.63452696 > d$

(k=3,l=0) $\phi_1(3,0) = 18.42780499 > d$

(k=4,l=0) $\phi_1(4,0) = 18.39481023 > d$

.....

(k=10,l=7) $\phi_1(10,7) = 11.12572387 > d$

(k=11,l=7) $\phi_1(11,7) = 11.11817573 > d$

(k=11,l=8) $\phi_1(11,8) < d$

$\phi_2(11,8) = \phi_1(11,7) = 11.11817573 > d$

$f_1(11,8) = 10.4058309 < d$

$f_2(11,8) = 11.34254408 > d$

Αρα

το ζεύγος (k=11,l=8) ικανοποιεί τις σχέσεις $\phi_1(11,8) < d < \phi_2(11,8)$

$$f_1(11,8) < d < f_2(11,8)$$

Επομένως $V(1)= V_{11,8}(1)=26.84980905$

$p^* = p_{11,8} =0.671245226$

(Επαλήθευση: $T_{10}(s_1) = 0.66074 < p^* < T_{11}(s_1) = 0.69467$

$T_7(s_2) = 0.66887 < p^* < T_8(s_2) = 0.70198$)

$$N=\max\{n \in \mathbb{N} : (0.9)^n \geq 1 - p^* = 0.328754774\} = 10$$

Εφαρμόζοντας το βήμα 3,τα διαιρετικά σημεία της Μαρκοβιανής διαμέρισης είναι:

$$\alpha_0 = p^* = 0.671245226$$

α_1=0.63417 α_2=0.59413 α_3=0.54903 α_4=0.49893 α_5=0.44325

α_6=0.38139 α_7=0.31266 α_8=0.23628 α_9=0.15143 α_{10}=0.05714

Τέλος εφαρμόζοντας το βήμα 4, η άριστη συνάρτηση κόστους για άπειρο χρονικό ορίζοντα είναι:

Αρα

$$V(p)=\begin{cases} 18.9794.p+16.8939, & 0 \le p \le 0.05714 \\ 18.4931.p+16.9217, & 0.05714 \prec p \le 0.15143 \\ 17.8927.p+17.0126, & 0.15143 \prec p \le 0.23628 \\ 17.1515.p+17.1878, & 0.23628 \prec p \le 0.31266 \\ 16.2365.p+17.4738, & 0.31266 \prec p \le 0.38139 \\ 15.1067.p+17.9047, & 0.38139 \prec p \le 0.44325 \\ 13.7120.p+18.5229, & 0.44325 \prec p \le 0.49893 \\ 11.9902.p+19.3820, & 0.49893 \prec p < 0.54903 \\ 9.8644.p+20.5491, & 0.54903 \prec p \le 0.59413 \\ 7.24.p+22.1083, & 0.59413 \prec p \le 0.63417 \\ 4p+24.1648, & 0.63417 \prec p \le 0.67125 \\ 26.8498, & 0.67125 \prec p \le 1 \end{cases}$$

ΣΥΜΠΕΡΑΣΜΑΤΑ

Στο κεφάλαιο αυτό μελετούμε ένα πρόβλημα επιλογής ανάμεσα σε δύο διδακτικές μεθόδους, με δύο μαθησιακές καταστάσεις (βαθμούς αφομοίωσης της διδασκόμενης ύλης από την τάξη) και δύο μηνύματα (επιτυχία /αποτυχία σε test) σε δύο περιπτώσεις:

α) περίπτωση πλήρους αβεβαιότητας, όπου το μήνυμα είναι ανεξάρτητο από την μαθησιακή κατάσταση της τάξης είτε επιλέγεται η συμβατική μέθοδος είτε η εξειδικευμένη μέθοδος διδασκαλίας και

β)περίπτωση πλήρους αβεβαιότητας όταν επιλέγεται η συμβατική μέθοδος διδασκαλίας και μερικής πληροφόρησης όταν επιλέγεται η εξειδικευμένη μέθοδος .

Σε κάθε περίπτωση η δομή της άριστης πολιτικής για άπειρο χρονικό ορίζοντα είναι η ακόλουθη:

- Αν ισχύει η συνθήκη (Σ), τότε η πολιτική να επιλέγεται πάντοτε η συμβατική μέθοδος διδασκαλίας είναι άριστη.
- Αν δεν ισχύει η συνθήκη (Σ),τότε η άριστη πολιτική είναι control-limit. Επιλέγεται η εξειδικευμένη μέθοδος διδασκαλίας αν η a- posteriori πιθανότητα για την ελλιπή μαθησιακή κατάσταση υπερβαίνει μία κρίσιμη ποσότητα και η συμβατική μέθοδος διαφορετικά.Η άριστη πολιτική επάγει Μαρκοβιανή διαμέριση στο χώρο των δ.π. και η άριστη (ελάχιστη) συνάρτηση του αναμενόμενου ολικού εκπίπτοντος κόστους είναι κατά τμήματα γραμμική.

Παρέχουμε αλγόριθμο σε κάθε μια από τις παραπάνω περιπτώσεις για τον ακριβή υπολογισμό της κρίσιμης ποσότητας της άριστης πολιτικής και της άριστης συνάρτησης κόστους σε άπειρο χρονικό ορίζοντα. Η απλότητα αυτών των αλγορίθμων για τις ειδικές αυτές περιπτώσεις είναι ιδιαίτερα ελκυστική.

ΠΑΡΑΡΤΗΜΑ Α

Πρόταση 1: *Εστω $h/[0,1]$συνεχής,κοίλη ή κυρτή πραγματική συνάρτηση,με $h(o)\succ 0$, $h(1)\prec 1$.Τότε η $h(x)$ έχει μοναδικό σταθερό σημείο $\xi\in(0,1)$. Επιπλέον $h(x)\succ x$ $\forall x\in[0,\xi)$ και $h(x)\prec\chi$ $\forall x\in(\xi,1]$.*

Απόδειξη

Θα αποδείξουμε την **ύπαρξη** σταθερού σημείου:Θεωρούμε προς τούτο την συνάρτηση σ(x)=h(x)-x, $0\le x\le 1$.Επειδή σ(0)= h(0)$\succ$0, σ(1)= h(1)-1$\prec$0 και η σ(x) συνεχής στο [0,1] ,από το θεώρημα Bolzano υπάρχει τουλάχιστον ένα $\xi\in(0,1)$ ώστε σ(ξ)=0,δηλαδή h(ξ)=ξ. Αρα η h(x) έχει σταθερό σημείο στο (0,1).

Θα αποδείξουμε την **μοναδικότητα** του σταθερού σημείου:Ας θεωρήσουμε ότι η h έχει περισσότερα από ένα σταθερά σημεία στο διάστημα (0,1).Εστω ξ_1,ξ_2 δύο σταθερά σημεία της h για τα οποία υποθέτουμε χωρίς να χαλάσει η γενικότητα ότι $\xi_1\prec\xi_2$.

Διακρίνουμε δύο περιπτώσεις.

i) Πρώτα υποθέτουμε ότι η h είναι κοίλη συνάρτηση.

Θεωρούμε το ευθύγραμμο τμήμα που συνδέει τα σημεία (0,h(0)) και $(\xi_2,h(\xi_2))$.Τότε έχουμε:

$$h((1-\lambda).\xi_2)\ge\lambda.h(0)+(1-\lambda).h(\xi_2)\ ,0\le\lambda\le 1 \qquad (1)$$

Αν επιλέξουμε $\lambda'=1-\frac{\xi_1}{\xi_2}$, έχουμε $\lambda'\in(0,1)$ και $(1-\lambda').\xi_2=\xi_1$.Επομένως λόγω της σχέσης (1),έχουμε:

$$h(\xi_1)=h((1-\lambda').\xi_2)\ge\lambda'.h(0)+(1-\lambda)'.h(\xi_2).$$

Επειδή ξ_1,ξ_2 είναι σταθερά σημεία για την h , η παραπάνω σχέση γράφεται :

$$\xi_1\ge\lambda'.h(0)+(1-\lambda').\xi_2=\lambda'.h(0)+\xi_1$$

από την οποία προκύπτει f(0)$\le$ 0, πράγμα άτοπο.

ii) Υποθέτουμε τώρα, ότι η $h(\chi)$ είναι κυρτή συνάρτηση.

Θεωρούμε το ευθύγραμμο τμήμα που συνδέει τα σημεία $(\xi_1,h(\xi_1))$ και (1, h(1)).Εχουμε ότι:

$$h(\lambda.\xi_1+(1-\lambda).1)\le\lambda.h(\xi_1)+\lambda.h(1),0\le\lambda\le1 \qquad (2)$$

Αν επιλέξουμε $\lambda''=\frac{1-\xi_2}{1-\xi_1}$ και αφού $\xi_1\prec\xi_2$, $\lambda''\in(0,1)$ ισχύει ότι:

$$\lambda''.\xi_1+(1-\lambda'').1=\xi_2.$$

Επομένως λόγω της (2),

$$h(\xi_2)=h(\lambda''.\xi_1+(1-\lambda'').1)\le\lambda''.h(\xi_1)+(1-\lambda'').h(1).$$

Επειδή ξ_1,ξ_2 είναι σταθερά σημεία για την h,η παραπάνω σχέση γράφεται :

$$\xi_2\le\lambda''.\xi_1+(1-\lambda'').h(1)$$

Από την οποία συνάγεται $h(1)\ge1$,πράγμα άτοπο. Αρα η h έχει μοναδικό σταθερό σημείο $\xi\in(0,1)$.

Θα αποδείξουμε τώρα ότι $h(x)\succ x \quad \forall x\in[0,\xi)$.

Η απόδειξη με την εις άτοπον. Εστω για κάποιο $\chi_0\in[0,\xi)$, με $h(x_0)\prec\chi_0$, τότε επειδή $h(0)\succ0$, σύμφωνα με το θεώρημα τού Bolzano,υπάρχει $\xi'\in(0,\xi)$ έτσι ώστε $h(\xi')=\xi'$..Αυτό όμως είναι άτοπο,επειδή έρχεται σε αντίθεση με την μοναδικότητα του σταθερού σημείου της h.Επίσης η περίπτωση $h(x_0)=x_0$ αποκλείεται για τον ίδιο ακριβώς λόγο. Επομένως $h(x)\succ x \quad \forall x\in[0,\xi)$.

Ανάλογα αποδεικνύεται ότι $h(x)\prec x \quad \forall x\in(\xi,1]$.

Πρόταση 2: *Εστω $h/[0,1]$ συνεχής, γνήσια αύξουσα, κοίλη ή κυρτή πραγματική συνάρτηση, με $h(o)\succ0, h(1)\prec1$ και $\xi\in(0,1)$ τό μοναδικό σταθερό σημείο τής h. Τότε*

$\chi\prec h(\chi)\prec\xi \quad \forall x\in[0,\xi)$,

$\xi\prec h(\chi)\prec\chi \quad \forall x\in(\xi,1]$. □

Πρόταση3: *Εστω $h/[0,1]$ συνεχής,γνήσια αύξουσα,κοίλη ή κυρτή πραγματική συνάρτηση, με $h(o)\succ0$, $h(1)\prec1$ και $\xi\in(0,1)$ τό (μοναδικό) σταθερό σημείο τής h. Τότε για $n=2,3,...$ η n-στη σύνθεση της h με τον εαυτό της, $h_n/[0,1]$,είναι συνεχής, γνήσια αύξουσα και έχει σταθερό σημείο το ξ. Για οποιοδήποτε $\chi\in[0,1]$ η ακολουθία $\{h_n(\chi)\}$* συγκλίνει μονότονα στο σημείο ξ όταν $n\to\infty$.Ειδικότερα

i) Αν $x\in[0,\xi)$, τότε η ακολουθία $\{h_n(\chi)\}$ είναι γνήσια αύξουσα,

$\chi\prec h_n(\chi)\prec\xi$,n=1,2,3…

και $h_n(\chi)\nearrow\xi$, όταν $n\to\infty$.

ii) Αν $x\in[\xi,1)$, τότε η ακολουθία $\{h_n(\chi)\}$ είναι γνήσια φθίνουσα,

$\xi \prec h_n(\chi) \prec \chi$,n=1,2,3...

και $h_n(\chi) \searrow \xi$, όταν $n \to \infty$.

Απόδειξη

Επειδή $h([0,1]\} = [h(0), h(1)] \subset (0,1)$, για n=1,2,3..., η n-στη σύνθεση της h, h_n /[0,1], είναι καλά ορισμένη,

$$0 \prec h_n(\chi) \prec 1 \qquad \forall \chi \in [0,1].$$

Προφανώς η h_n /[0,1] είναι συνεχής, γνήσια αύξουσα και έχει σταθερό σημείο ξ.

i) Έστω $x\in[0,\xi)$. Έχουμε

$$0 \prec h_n(\chi) \prec h_n(\xi) = \xi, \quad ,n=1,2,3...$$

Από την πρόταση 1 έχουμε:

$$h_n(x) = h(h_{n-1}(\chi)) \succ h_{n-1}(\chi), n = 2,3...$$

δηλαδή η ακολουθία $\{h_n(\chi)\}$ είναι γνήσια αύξουσα και επομένως συγκλίνουσα επειδή είναι φραγμένη.Εστω

$$Z=\lim_{n\to\infty} h_n(x).$$

Επειδή η συνάρτηση h είναι συνεχής, έχουμε:

$$Z=\lim_{n\to\infty} h_n(x) = \lim_{n\to\infty} h(h_{n-1}(\chi)) = h(z).$$

Αρα το όριο Z είναι σταθερό σημείο για την h.Επειδή η συνάρτηση h έχει μοναδικό σταθερό σημείο (πρόταση 1), συνάγεται ότι Z=ξ. Επειδή $\chi \prec \xi$, έχουμε

$$\chi \prec h(\chi) \prec \xi \text{ (πρόταση 2).}$$

Επιπλέον, επειδή

$$h(\chi) \prec h_n(\chi) \prec h_n(\xi) = \xi, n = 2,3,...$$

συνάγεται ότι :

$$\chi \prec h_n(\chi) \prec \xi, n = 2,3,...$$

ii) Αποδεικνύεται ανάλογα. □

ΒΙΒΛΙΟΓΡΑΦΙΑ

[1]S.Albright.Structural results for partially observable Markov decision processes, *Operation research* **27**,1041-1053,1979.

[2]R.Andrew and McCallum.*Overcoming incomplete perception with utile distinction memory*. In Proceedings of the Tenth International Conference on Machine Learning, Amherst, Massachusetts,1993.Morgan Kaufmann.

[3]M.Aoki.*Optimization of Stochastic Systems*. Academic Press, New-York, NY, 1967.

[4]A.Arapostathis and Fernandez .Discrete-time controlled Markov Processes with average cost criterion. *Siam Journal of control and optimization, 31 (2)* 282-344,1993.

[5]K.J.Astrom.Optimal control of Markov decision processes with incomplete state estimation. *Journal of Mathematical Analysis and Applications* **10**, 174-205,1965.

[6]K.J.Astrom. Optimal control of Markov decision processes with incomplete state Information. *Journal of Mathematical Analysis and Applications* **26**,403-406,1969.

[7]K.J.Astrom.Theory and applications of adaptive control, a survey. *Automatica* **19**,471-186,1983.

[8]J.Bean. Conditions for the existence of planning horizons. *Math Opns.Res.* **9**,391-401,1984.

[9]R.Bellman. *Dynamic programming*. Princeton University Press Princeton, New Jersey,1957.

[10]D.P.Bertsekas.Distributed dynamic programming.*IEEE Transactions on Automatic Control*, AC-27, 610-616, 1982.

[11]D.P.Bertsekas.Dynamic Programming and Optimal Control, Vols. 1 and 2 Athena Scientific, Belmont, Massachusetts, 1995.

[12]D.P.Bertsekas and R.G.Gallagher.*Data Networks*. Prentice Hall., Englewood Cliffs, N.J.,1992.

[13]D.P.Bertsekas and John N.Tsitsiklis. *Neuro-Dynamic Programming.* Athena Scientific, Belmont, Massachusetts, 1996.

[14]D.Blackwell.Discrete dynamic programming. *Annals of Mathematical Statistics*, **33**(2), 719-726, June 1962.

[15]D.Blackwell. Discounted-dynamic-programming. *Annals of Mathematical Statistics* **36**, 226-235,1965.

[16]D.Blackwell, *Positive Dynamic Programming*/Proc. Of Fifth Berkeley Symposium on Mathematical Statistics and Probability, Univ. of California Press, Berkeley, California,415- 418(1967).

[17]C.Boutilier and D. Poole.*Computing optimal policies for partially observable decision processes using compact representations.* In Proceedings of the Thirteenth National Conference on Artificial Intelligence, pages 1168-1175, Portland, Oregon, 1996.

[18]R.Brafman. A heuristic variable grid solution method for POMDPs. In *Proceedings of the Fourteenth National Conference on Artificial Intelligence, pages 727-733,Providence, Rhode Island, 1997.*

[19]J.Buhmann, W. Burgard, Cremers A., D. Fox, T. Hofmann, F. Scheider, J. Strikes,and S.Thnm:The mobile robot RHINO.Magazine,**16**(2}:31-37, Summer 1995.

[20]B. Bukiet, Elliotte Rusty Harold, and Jose Luis Palacios. A Markov chain approach to baseball. *Operations Research*, **45**(1), 14—23,1997.

[21]D. Burago, M.Rougemont,and A. Slissenko.On the complexity of partially observed Markov decision processes.
Theoretical Computer Science **157** (2), 161-183,1996.

[22]G.B.Dantzing.*Linear programming and extensions*,Princeton university,New Jersey,1984.

[23]E.V.Denardo.Contraction mapping in the theory underlying dynamic programming. *Siam reviews* **9**,165-177,1967.

[24] C. Derman,Finite State Markovian Decision Processes, Academic Press, New York (1967).

[25] J.L Doob. *Stochastic Processes*, John Wiley & Sons, New York (1953).

[26]A.W. Drake.*Observation of a Markov Process Through a Noisy Channel. Phd.* thesis Electrical Engineering Department, M.I.T.Cambridge, Mass, June 1962.

[27]S.E.Dreyfus. *The art and theory of dynamic programming*,Academic press ,London 1977.

[28]E.B.Dynkin. Controlled random sequences. Theory of Probability and its Applications, **28**, 1-14, 1965.

[29]J. N. Eagle. The optimal search for a moving target when the search path is constrained. *Operations Research*, **32**(5), 1107-1115,1984.

[30]J.E.Eckles. Optimum maintenance with incomplete information. Operations Research , **16**, 1058-1067,1968.

[31]Fernandez-Arapostathis. *Adaptive Markov control processes*. Verlang New-York 1989.

[32]H.Freudenthal, Simplizialzerlegungen von Berstracter,1942 *Ann. Math* **23**,580-582.

[33] S. I.Gass, *Linear programming* ,Mc Graw-Hill 1995.

[34]J.E.Goulionis (2007).Structural properties for a two-state partially observable Markov decision process with an average cost criterion. Journal of Statistics & Management Systems. ***(To appear).***

[35]J.E.Goulionis and V.K Benos.Optimal control limit strategies,using the partially observable Markov decision processes. ***Advances and applications in Statistics.*** **7(3),357-388 (2007).**

[36]J.E.Goulionis.POMDPs with uniformly distributed signal processes. ***Spoydai*** **55, 35-55, 2005.**

[37] J.E.Goulionis. (2006). An optimal policy with uncertain information. ***Journal of operation research society of Japan. (Revised- paper).***

[38]J.E.Goulionis.A Model of learning using POMDPS, ***Mathematical inspection*** **(Μαθηματική επιθεώρηση) 63,36-49,2005.**

[39]J.E.Goulionis.*Extension of a finitely transient strategy for a POMDP*,working paper **14**, 287-320, University of Piraeus .*(Essays in Honour of professor Skoytzos).* (2005).

[40]J.E.Goulionis.*An optimal replacement policy*,working paper **20**,267-287,University of Piraeus, (2005). *(Essays in Honour of professor Sarantides A.Stylianos).*

[41]J.E.Goulionis.*(2006).*Function approximators for POMDPs. Yogoslav *Journal of Operation Research . (Revised paper).*

[42]Γκουλιώνης Ιωάννης.*Σύγχρονες εξελίξεις ηλεκτρικής ενεργειακής τεχνολογίας.*Δεύτερο διεθνές Συνέδριο με θέμα την εφαρμογή των ανανεώσιμων πηγών ενέργειας(RMS).Προτεραιότητες σε συνθήκες απελευθερωμένης αγοράς 19-21 Μάρτη 2001.ΚΕΠΠΠ,Γουδί,Αθήνα.

[43]Γκουλιώνης Ιωάννης.*Η διδασκαλία των θετικών επιστημών*.Εθνικό Συνέδριο από την ένωση για την διδακτική των φυσικών επιστημών.3-3-2001(πολεμικό μουσείο).

[44]Γκουλιώνης Ιωάννης.*Διοίκηση δημόσιων υπηρεσιών*.Εθνικό κέντρο δημόσιας διοίκησης,συνέδριο τομέα δημόσιου Μάνατζμεντ (24 Νοέμβρη-9 Δεκέμβρη 1998).

[45]J.E.Goulionis.(2006). An algorithm to obtain an optimal strategy for the Markov decision processes,with probability distribution for the planning horizon. *Journal of information and optimization theory. (Submitted paper).*

[46] J.E.Goulionis. (2006). Modeling medical treatment using partially observable Markov Decision Processes. *European Journal of operation research. (Submitted paper).*

[47]J.E.Goulionis (2006). A Useful criterion for the search in Policy space. Journal of applied probability. *(Submitted paper).*

[48] M. Hauskrecht . Value function approximations for POMDPs. *Journal of artificial intelligence research 13*,33-94, (2000).

[49]J.E. Hopcroft and J.D.Ullman. *Introduction to Automata Theory, and Computation.*
Addison Wesley Publishing Company, Reading, Massachusetts,1979.

[50]R.A.Howard and J.Matheson. Risk sensitive Markov decision processes. *Management Science*, **18** (7) 356-370, 1972.

[51]R.A.Howard. Dynamic Programming and Markov Processes.The MIT Press,Cambridge,Massachusetts,1960.

[52]J.Hughes.*Optimal internal audit timing.* Accounting Review,LII,56-68, 1977.

[53]T.Jakola, P.Satinder ,and M.I. Jordan. Monte- Carlo reinforcement learning in non-Markovian decision problems. *In Advances in Neural Information Processing Systems* **7**, *1995.*

[54]A. Kachites McCallum.Reinforcement Learning with Selective Perception and Hidden State, PhD thesis, University of Rochester, 1996.

[55]L.P. Kaelbling,T. Dean,J.Kirman, and A.Nicholson. Planning with deadlines in stochastic domains- In Proceedings of the Eleventh National Conference on Artificial Intelligence, Washington, DC, 1993.

[56]J.S.Kakalik.Optimal policies for partially observable Markov systems. Technical Report TR-1S, Massachusetts Institute of Technology, Cambridge, MA, October 1965.

[57]R.E.Kaimai. A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, pages 35-45, March 1960.

[58]R.Kaplan.Optimal investigation strategies with imperfect information. *Journal of Accounting Research*,7, 32-43, 1969.

[59]W.Karush and R.Dear.Optimal strategy for item presentation in learning models. *Management Science*, **13**,773-785, 1967.

[60]S.Karlin. *Total positivity*, Stanford University Press,CA,1968

[61]S.Koenig.*Optimal probabilistic and decision-theoretic planning using Markovian decision theory.* Technical Report UCB/CSD 92/685, Berkeley, May 1992.

[62]P.R.Kumar. A survey of some results in stochastic adaptive control. SIAM *Journal on Control and Optimization*,**23**, 329-380,1985.

[63]N.Kushmeric, S. Hanks, and D.Weld. *An algorithm for probabilistic planning*.Technical Report **93**-06-03, Department of Computer Science, University of Washington, 1993.

[64]N.Kushmerick,S. Hanks,and D.S. Weld.An algorithm for probabilistic planning.*Artificial Intelligence*,**76**(2),239-256,September 1995.

[63]H.J. Kusner, *Introduction to stochastic Control*,Brown University 1971,Holt,Rineart.

[65]H.J.Kushner and A.J.Kleinman.Mathematical programming and the control of Markov chains. *International Journal of Control*,**13** (5), 801-820,1971.

[66]D.E.Lane. A partially observable model of decision making by fishermen, *Operations Research*, **37**,240-256,1989.

[67]J.J.Leonard and H.Durrant. Localization by tracking geometric beacons. *IEEE Transactions on Robotics and Automation*, 7(6),1991.

[68]H.R.Lewis and C.H.Papadimitriou. Elements of the Theory of Computation. Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1981.

[69]M.L.Littman. Memoryless policies:Theoretical limitations and practical results. From Animals to Animate S, Brighton, UK,1994.

[70]M.L.Littman. The witness algorithm for solving partially observable Markov decision processes.Technical Report CS-94-40,Brown University, Providence, Rhode Island,1994.

[71]M.L.Littman. Algorithms for Sequential Decision Making. PhD thesis. Department of Computer Science, Brown University, February 1996. Also Technical Report CS-96-09.

[72]M.L.Littman. Planning and acting in partially observable stochastic domains.*Artificial Intelligence*,(101),pages 99-134 Providence, Rhode Island, 1998.

[73]M.L. Littman.Anthony R. Cassandra, and Leslie Pack Kaelbling: Efficient dynamic-programming updates in partially observable Markov decision processes- Technical Report CS-95-19, Brown University, Providence, Rhode Island, 1995.

[74]W.S.Lovejoy. A survey of algorithmic methods for partially observed Markov decision processes.*Annals of Operations Research*, 28(1), 47-65,1991.

[75]W.S.Lovejoy. Computationally feasible bounds for p.o.m.d.p.s *Operations Research*, **39**,162-175,1991.

[76]W.S Lovejoy. Some monotonicity results of POMDPs *Operations Research*, **35** ,736-743, 1987.

[77]W.S. Lovejoy. A Note on Exact Solution of POMDPs. Research paper 1003. Stanford University , July 1988.

[78]O.N.Madani. Complexity results for infinite-horizon MDP, PhD thesis, University of Washington, 2000.

[79]W.N.Maitra.Discounted dynamic programming on compact metric spaces *Sankya* ser. **30** , 211-216,1968.

[80]A.Manne. Linear programming and sequential decisions.*Management Science*, **6**,259-267,1960.

[81]B.Martos,Nonlinear programming theory and methods,American elsevier 1991.

[82]T.H. Mattheis. An algorithm for determining irrelevant constraints and all verticies in systems of linear inequalities. *Operations Research*, **21**, 247-260, 1973.

[83]T.H. Mattheis and D.S. Rubin. A survey and comparison of methods for finding all vertices of convex polyhedral sets.*Mathematics of Operations Research*, **5**(2),167-185,1980.

[84] L.E.Mangasarian.*Nonlinear programming* ,Mc Graw-hill book company 1996.

[85]G.E.Monahan. A survey of partially observable Markov decision processes: Theory, models, and algorithms. *Management Science*,**28**(1),1-16, 1982.

[86]L.Morgenstern.Knowledge preconditions for actions and plans.In Proceedings of the 10th International Joint Conference on Artificial Intelligence, pages 867-874, 1987.

[87]S.Mukherjee and K. Seth.A corrected and improved computational scheme for partially observable Markov processes. INFOR, **29**(3),206-212, 1991.

[88]M.Mundhenk, J. Goldsmith, and E.Allender. The complexity of policy evaluation for finite-horizon partially-observable markov decision processes. In Proceedings of the 25th Mathematical Foundations of Computer Sciences, pages 129-138- Lecture Notes in Computer Science 1295, Springer-Verlag, 1997.

[89]M.Mimdhenk, J.Goldsmith, C. Lusena,and E. Allender: *Encyclopaedia of complexity results for finite-horizon Markov decision process problems.*Technical Report TR **273**-97,University of Kentucky,Lexington, Kentucky, September 1997.

[90]R. Munos.*A convergent reinforcement learning algorithm in the continuous case:the finite-element reinforcement learning.* In Proceedings of the Thirteenth International Conference on Machine Learning, 1996.

[91]M.Ohnishi and S.Ibaraki. *An optimal inspection and replacement policy under incomplete state information. Stochastic models in reliability theory* 187-197, Berlin (1986),Springer -Verlang.

[92]C.H.Papadimitriu and J.N.Tsitsiklis.The complexity of Markov decision processes.Mathematics of operation research **12**(3),441-450 1987.

[93]R.Parr and S.Russell. Approximating optimal policies for partially observable stochastic domains.In Proceedings of the International Joint Conference on Artificial Intelligence,pages 1088-1094.Morgan Kaufmann, 1995.

[94]M.A. Peot and D.E. Smith.Conditional nonlinear planning. In Proceedings of the First International Conference on Artificial Intelligence Planning Systems, pages 189-197,1992.

[95]W.Pierskalla and J.Voelker. A survey of maintenance models.The control and surveillance of deteriorating systems. *Naval Research Logistics Quarterly*, **23**, 353-388,1976.

[96]L.K. Platzman. Optimal infinite-horizon undiscounted control of finite probabilistc systems.*SIAM Journal of Control and Optimization*,**18**,362-380, 1980.

[97]E.Pollock.A simple model of search for a moving target.*Operations Research*, **18**,883-903, 1970.

[98]M.L.Puterman. Markov Decision Processes Discrete Stochastic Dynamic Programming. John Wiley & Sons, Inc., New York, New-York, 1994.

[99]M.L.Puterman and Moon Chirl Shin.Modified policy iteration algorithms for discounted Markov decision problems.*Management Science*, **24**,1127-1137, 1978.

[100]M.Queen.A test of suboptimal actions in Markovian decision processes *O.R* **15** 559-561 1968.

[101]D.Rosenfield. Markovian deterioration with imperfect information *Journal of Accounting Research*, 7, 32-43, 1969.

[102]R.T.Rockfellar.Augmented lagrangians and applications of the αpproximate point algorithm in convex programming.*Mathematics of operation research* **1**,(2),May 1976 , 97-116.

[103]D.Rosenfield.Markovian deterioration with uncertain information. *Operations Research*, **24** (1),141-155, 1976.

[104]S.M.Ross.*Applied probability models with optimization applications*,San Francisco-California 1980.

[105]S.M.Ross.Quality control under Markovian deterioration. *Management Science*, **17**(9), 587-596,1971.

[106]S.M.Ross.Arbitrary State Markovian decision processes.Ann. of Math.Statistics ,**39,** 2118-2122,1968.

[107]S.M.Ross.*Introduction to probability models*.Academic press, New – York,1988.

[108] S.M.Ross. *Introduction to stochastic dynamic programming.* Academic press ,1983.

[109]K.Sawaki.*Piecewise linear Markov decision processes with an application to POMDPS* Academic press ,New –York,1980

[110]K.Sawaki and A. Ichikawa. Optimal control for partially observable Markov decision processes over an infinite horizon.*Journal of the Operations Research Society of Japan*,**21**(1),1-14, March 1978.

[111]Y.Sawaragi and T.Yoshikawa. Discrete time Markov decision processes with incomplete state information.*Annals of Mathematics and Statistics*,**41**,78-86,1970.

[112]J.Schmidhuber. Reinforcement learning in Markovian and non-Markovtan environments.In Advances in Neural Information Processing Systems **3**,500-506,1991.

[113]A.Segall. Dynamic file assignment in a computer network. IEEE *Transactions on. Automatic Control*, AC-**21**,161-173, 1976.

[114]H. Shatkay and L. Pack Kaelbling. *Learning topological maps with weak local odometric information.* In Proceedings of the Fifteenth International Joint Conference on Artificial Intelligence,Nagoya, Japan, August 1997.

[115]S. E. Shreve and D.Bertsekas . Stochastic Optimal control.Academic press New York 1988.

[116]R. Simmons and S. Koenig. Probabilistic navigation in partially observable environments. In Fourteenth International Joint Conference on Artificial Intelligence, pages 1080-1087,Montreal,Canada,1995.Morgan Kaufinann.

[117]E.Sondik.The optimal control of partially observable Markov processes,Phd.thesis, Department of electrical Engineering,Stanford University 1971.

[118]R.Smallwood,E.Sondik.Toward and integrated methodology for the analysis of health-care systems. *Operations Research*,**19**, 1300-1322, 1971.

[119]R.Smallwood and E. J. Sondik.The optimal control of partially observable Markov processes over a finite horizon. *Operations Research*, **21**,1071-1088, 1973.

[120]E. J.Sondik. The optimal control of partially observable Markov processes over the infinite horizon: Discounted costs. *Operations Research*, **26**(2),282-304, 1978.

[121]C.T. Striebel.Sufficient statistics in the optimal control of stochastic systems. *Journal of Mathematical Analysts and Applications*,**12**,576—592,1965.

[122]R. S. Sutton. Learning to predict by the methods of temporal differences. *Machine Learning*, **3**,29-44, 1988.

[123] C.Tijms Henk. Stochastic modelling and analysis,John Wiley and sons, 1996.

[124]G.J.Tesauro and D.Gammon. A self-teaching backgammon program, achieves master-level play. *Neural Computation*, **6**,215-219, 1994.

[125]R.Washington.*Uncertainty and real-time therapy planning: incremental Markov-model approaches.*AAAI Spring Symposium on Artificial Intelligence in Medicine,1996.

[126]C.H.Watkins and P.Dayan. Q-learning.*Machine Learning*, **3**{3):279-292,1992.

[127]J.Wald .Structural results for POMDPs ,*OR* 27,1030-1050,1969.

[128]R.Whitt.Approximations of dynamic programs.Mathematics of *Operation research* **3** ,231-243, 1978.

[129]C.C.White. Cost equality and inequality results for a partially observed stochastic optimization problem. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-5 (6):576-582, November 1975.

[130]C.C.White.Optimal diagnostic questionaires which allow less than truthful responses. *Information and Control* ,**32**, 61-74,1976.

[131]C.C.White. Procedures for the solution of a finite-horizon partially observed, semi-Markov optimization problem.*Operations Research*, **24**(2),338-358, 1976.

[132]C.C.White. Monotone control laws for noisy, countable-state Markov chains. *European Journal of Operations Resea*rch, **5**,124-132, 1980.

[133]C.C.White.Partially observed Markov decision processes:A survey. *Annals of Operations Research*, **32**, 1991,78-97.

[134]C.C.White and W.T. Scherer. Solution procedures for partially observed Markov decision processes. *Operations Research*, **37**(5), 791-797,1989.

[135]C.C.White,III and William T. Scherer. Finite memory suboptimal design for partially observed Markov decision processes.*Operations Research*, **42**(3),439-455.

[136]Whitt R. Approximation of dynamic programs.*Mathematics of operation research* **3**,231-243, 1978.

[137]E.Whitehead and Dana H. Learning to perceive and act by trial and error. *Machine Learning*,7(1),45—83,1991.

[138]M.Wiering and J. Schmidhuber:HQ-learning:discovering Markovian subgoals for non-Markovian reinforcement learaiag. Technical Report IDSIA-95-96, IDSIA, Switzerland, October 1996.

[139]D.Wilkins,K. Myers,J. Lowrance, and K. Leonard Wesley. Planning and reacting in uncertain and dynamic environments.*Artificial Intelligence*,7,121-152,1995.

[140]W.L.Winston.Introduction to Mathematical Programming: Applications and Algorithms.PWS-KENT, Boston, Massachusetts,1991.

[141]N.L.Zhang.Efficient planning in stochastic domains through exploiting problem characteristics. Technical Report HKUST-CS95-40, Department of Computer Science, Hong Kong University of Science and Technology, August 1995.

[142]N.L.Zhang and W.Liu.Planning in stochastic domains. Problem characteristics and approximation. Technical Report HKUST-CS96-31, Department of Computer Science, Hong Kong University of Science and Technology,1996.