



ΕΘΝΙΚΟ ΚΑΙ ΚΑΠΟΔΙΣΤΡΙΑΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ
ΑΘΗΝΩΝ

Διπλωματική εργασία για την απόκτηση του
Μεταπτυχιακού Διπλώματος Ειδίκευσης στη:
Στατιστική και Επιχειρησιακή Έρευνα

Υπολογιστικές μέθοδοι σε Μαρκοβιανές Διαδικασίες
Αποφάσεων υπό ελλιπή πληροφόρηση.

Για το Τμήμα Μαθηματικών

Από τον Κοκαλιάρη Χρήστο

Απρίλιος 2013

Τριμερής Επιτροπή:

Α. Μπουρνέτας – επιβλέπων

Λ. Μελιγκοτσίδου

Κ. Μηλολιδάκης

Περιεχόμενα

Περιεχόμενα.....	i
1 Εισαγωγή.....	1
2 Περιγραφή μοντέλου και ιδιότητες.....	5
2.1 Ιδιότητες Βέλτιστης Πολιτικής.....	5
3 Αλγόριθμος του Sondik.....	15
3.1 Περιγραφή του αλγορίθμου.....	15
3.2 Συζήτηση - Πλεονεκτήματα.....	18
3.2.1 Κριτική - Διορθώσεις.....	18
3.3 Παραλλαγές και δυσκολίες στις μεθόδους ακριβείας.....	19
4 Προσεγγιστικές μέθοδοι για MOMΔΑ	23
4.1 Προσεγγιστικές μέθοδοι πεπερασμένης μνήμης.....	23
4.2 Προσεγγιστικές μέθοδοι πεπερασμένου πλέγματος.....	26
4.2.1 Μέθοδοι σταθερού πλέγματος.....	27
4.2.2 Μέθοδοι μεταβλητού πλέγματος	29
4.3 Συζήτηση.....	30
5 Αλγόριθμος Perseus.....	35
5.1 Περιγραφή του αλγορίθμου.....	35
5.2 Συζήτηση.....	38
6 Εφαρμογή σε ουρά Geo/Geo/1/M.....	41
6.1 Περιγραφή του προβλήματος.....	41
6.1.1 Το μοντέλο Μαρκοβιανής διαδικασίας αποφάσεων με μερική πληροφόρηση	

.....	42
6.1.2 Το πρόβλημα με πλήρη πληροφόρηση.....	44
6.1.3 Το πρόβλημα με μερική πληροφόρηση.....	60
6.2 Αριθμητικά αποτελέσματα από εφαρμογή μεθόδου Perseus.....	62
6.2.1 Διακρίνουσα ανάλυση πάνω στα αποτελέσματα.....	66
6.3 Συμπεράσματα.....	73

1 Εισαγωγή

Η στοχαστική μοντελοποίηση έχει ως στόχο την παροχή λύσεων σε μια ευρεία κατηγορία προβλημάτων όπου το κύριο χαρακτηριστικό τους είναι ο τυχαίος τρόπος με τον οποίο αλλάζουν οι καταστάσεις του υπό εξέταση φαινομένου. Μια ειδική κατηγορία προβλημάτων είναι και οι Μαρκοβιανές Διαδικασίες Αποφάσεων, όπου οι κατάσταση του συστήματος αλλάζει πιθανοθεωρητικά διατηρώντας την Μαρκοβιανή ιδιότητα και ο στόχος μας είναι, σε κάθε κατάσταση να επιλέξουμε μια απόφαση έτσι ώστε να μεγιστοποιηθεί μια έκφραση αναμενόμενου κέρδους, όπου π.χ. το συνολικό αναμενόμενο κέρδος σε πεπερασμένο ορίζοντα, η παρούσα αξία του αναμενόμενου κέρδους σε άπειρο ορίζοντα κλπ.

Μια βασική υπόθεση του μοντέλου είναι το γεγονός ότι στην αρχή της κάθε περιόδου γνωρίζουμε ακριβώς σε ποια κατάσταση βρίσκεται το σύστημα. Προβλήματα αυτού του τύπου έχουν μελετηθεί εκτενώς στην βιβλιογραφία [1], [2].

Τι συμβαίνει όμως όταν έχουμε ελλιπή πληροφόρηση όσον αφορά την αρχική κατάσταση; Αν το φαινόμενο που εξετάζουμε δεν μας επιτρέπει να δούμε την κατάσταση του συστήματος στην αρχή της κάθε περιόδου τι μπορούμε να κάνουμε ώστε να συνεχίσουμε να “απαντάμε βέλτιστα” σε κάθε κατάσταση και ποια θα είναι σε αυτή τη περίπτωση η δομή των βέλτιστων πολιτικών; Στα προβλήματα αυτά καλείται να δώσει λύση η θεωρία των Μερικώς Ορατών Μαρκοβιανών Διαδικασιών Αποφάσεων.

Στην εργασία αυτή γίνεται μια εισαγωγή στο μαθηματικό μοντέλο των μερικώς ορατών Μαρκοβιανών διαδικασιών αποφάσεων. Στο πρώτο μέρος της εργασίας αναφέρονται

τα βασικά θεωρητικά αποτελέσματα από την βιβλιογραφία και αναζητούνται δύο κατηγορίες μεθόδων για την εύρεση βέλτιστων πολιτικών. Η πρώτη κατηγορία είναι ακριβείς μέθοδοι που όμως πολύ μεγάλη υπολογιστική πολυπλοκότητα. Η δεύτερη κατηγορία μεθόδων είναι προσεγγιστικές και υπολογίζονται σε έξυπνη προσαρμογή μεθόδων προσομοίωσης.

Στο δεύτερο μέρος της εργασίας γίνεται εφαρμογή των μεθόδων σε ένα πρόβλημα ελέγχου ουράς αναμονής με μερικώς παρατηρήσιμη κατάσταση. Συγκεκριμένα το πρόβλημα που μελετάμε αφορά μια Μαρκοβιανή ουρά εξυπηρέτησης σε διακριτό χρόνο όπου οι χρόνοι μεταξύ αφίξεων ακολουθούν Γεωμετρική κατανομή με παράμετρο p . Υπάρχει ένας υπηρέτης που εξυπηρετεί με βάση πειθαρχία FCFS. Οι χρόνοι εξυπηρέτησης ακολουθούν Γεωμετρική κατανομή με παράμετρο q . Ο χώρος αναμονής έχει μέγεθος M . Στην είσοδο της ουράς υποθέτουμε ότι υπάρχει αυτόματος μηχανισμός ελέγχου που αποδέχεται ή απορρίπτει τους πελάτες που φτάνουν στο σύστημα.

Το πρόβλημα είναι να βρεθεί μια πολιτική αποδοχής/απόρριψης πελατών που μεγιστοποιεί μια έκφραση, ένα κατάλληλα ορισμένο αναμενόμενο κέρδος σε πεπερασμένο ορίζοντα.

Θα μελετήσουμε το πρόβλημα και στις δυο εκδοχές του, αυτή της πλήρους πληροφόρησης και αυτή της μερικής, αποδεικνύοντας κάποια θεωρητικά αποτελέσματα τα οποία μπορούν να φανούν χρήσιμα και σε άλλες εφαρμογές.

Στο τέλος θα παρουσιάσουμε και έναν απλοποιημένο κανόνα λήψης απόφασης για την περίπτωση της μερικής πληροφόρησης εφαρμόζοντας τον αλγόριθμο προσομοίωσης για

μερικά ορατές Μαρκοβιανές διαδικασίες αποφάσεων και κάνοντας χρήση στατιστικής ανάλυσης δεδομένων.

2 Περιγραφή μοντέλου και ιδιότητες

Υποθέτουμε ότι το σύστημα κινείται σύμφωνα με μια Μαρκοβιανή Διαδικασία από κατάσταση σε κατάσταση αλλά δεν μπορούμε να ξέρουμε ποια ακριβώς είναι η κατάσταση την εκάστοτε χρονική στιγμή, έχουμε δηλαδή ελλιπή πληροφόρηση. Στην αρχή της περιόδου αποφάσεων το σύστημα βρίσκεται στην κατάσταση i , αν επιλέξουμε την απόφαση a τότε έχουμε μετάβαση στην κατάσταση j με πιθανότητα p_{ij}^a που είναι η δεσμευμένη πιθανότητα το σύστημα να βρίσκεται στην κατάσταση j δεδομένου ότι βρισκόταν στην κατάσταση i και επιλέξαμε a . Ο “εσωτερικός” μηχανισμός της μετάβασης καταστάσεων διατηρεί την Μαρκοβιανή ιδιότητα.

Παρόλο που η πραγματική κατάσταση δεν είναι ορατή σε ένα τέτοιο μοντέλο έχουμε υποθέσει και ένα μηχανισμό από σήματα θ . Κάθε φορά που επιλέγουμε την απόφαση a και το σύστημα μεταβαίνει στην κατάσταση j παρατηρούμε και μια τιμή θ με πιθανότητα $r_{j\theta}^a$ που είναι η δεσμευμένη πιθανότητα να παρατηρήσουμε σήμα θ δεδομένου ότι επιλέξαμε a και η κατάσταση του συστήματος έγινε j . Η μετάβαση από την κατάσταση i όταν επιλέγουμε a σε μια νέα κατάσταση j και παρατηρούμε σήμα θ συνοδεύεται και με μια αμοιβή $w_{ij\theta}^a$.

2.1 Ιδιότητες Βέλτιστης Πολιτικής

Ο χώρος καταστάσεων, τα σύνολα αποφάσεων και το σύνολο των δυνατών σημάτων μπορεί να είναι γενικά σύνολα. Στην εργασία αυτή θεωρούμε την περίπτωση όπου όλα τα παραπάνω σύνολα είναι πεπερασμένα.

Συγκεκριμένα ο χώρος καταστάσεων είναι το σύνολο $\{1, 2, \dots, N\}$, ο χώρος αποφάσεων $\{1, 2, \dots, v\}$ και το σύνολο των δυνατών σημάτων $\{1, 2, \dots, \rho\}$.

Σημειώνουμε ότι για την πεπερασμένη περίπτωση μπορούμε χωρίς παραβίαση της γενικότητας να υποθέσουμε ότι ο χώρος αποφάσεων είναι ίδιος για όλες τις δυνατές καταστάσεις και σήματα.

Με βάση τα παραπάνω η εξέλιξη του συστήματος περιγράφεται ως εξής: Στην αρχή της περιόδου t (τέλος της περιόδου $t - 1$) το σύστημα βρίσκεται στην κατάσταση $s(t) = i$. Ο αποφασίζων δεν γνωρίζει την τιμή του $s(t-1)$, αλλά μόνο το σήμα $z(t-1)$ που έχει δοθεί από την προηγούμενη μετάβαση.

Έστω ότι ο αποφασίζων παίρνει την απόφαση $a(t) = a$. Το σύστημα πραγματοποιεί μια μετάβαση από την κατάσταση $s(t) = i$ στην κατάσταση $s(t+1) = j$ με πιθανότητα p_{ij}^a , επίσης παράγεται το σήμα $z(t) = \theta$ με πιθανότητα $r_{j\theta}^a$, το οποίο γίνεται γνωστό στον αποφασίζοντα. Στην αρχή της περιόδου $t + 1$ η διαδικασία επαναλαμβάνεται με την νέα πληροφορία $z(t)$ κ.ο.κ.

Έστω $h(t) = (z(t), a(t), z(t-1), a(t-1), \dots, z(1), a(1))$ το διάνυσμα της συνολικής ιστορίας του συστήματος που είναι διαθέσιμη στον αποφασίζοντα στο τέλος της περιόδου t (αρχή της $t + 1$).

Με βάση τα παραπάνω η απόφαση $a(t+1)$ μπορεί να εξαρτάται μόνο από το διάνυσμα $h(t)$.

Λόγω της έλλειψης πληροφόρησης για την παρούσα κατάσταση θεωρούμε ως την κατάσταση του συστήματος στο τέλος της περιόδου t το διάνυσμα πιθανότητας

$\pi(t) = (\pi_1(t), \pi_2(t), \dots, \pi_N(t))$ όπου $\pi_i(t)$ είναι η πιθανότητα το σύστημα να βρίσκεται στην κατάσταση i , δεδομένης της ιστορίας έως αυτή τη στιγμή. Δηλαδή έχουμε $\pi_j(t) = P(s(t) = j | h(t))$. Το διάνυσμα πιθανότητας $\pi_j(t)$ θα το ονομάζουμε στο εξής διάνυσμα πληροφοριών.

Πρόταση [3]: Το π διάνυσμα κατανομής της παρούσας κατάστασης του συστήματος, είναι επαρκής στατιστική συνάρτηση για την ιστορία των παρατηρήσεων.

Απόδειξη: Έστω η ιστορία $h(t)$ όπως ορίστηκε προηγουμένως, με $z(t) = \theta$. Από τον ορισμό του $\pi_i(t)$ παρατηρούμε ότι:

$$\pi_j(t) = P(s(t) = j | z(t) = \theta, a(t), h(t-1)) = \frac{P(s(t) = j, z(t) = \theta | a(t), h(t-1))}{P(z(t) = \theta | a(t), h(t-1))}. \quad (1)$$

Στην προηγούμενη έκφραση ο αριθμητής γράφεται αναλυτικά ως εξής:

$$\begin{aligned} & P(s(t) = j, z(t) = \theta | a(t), h(t-1)) = \\ & = \sum_i P(s(t) = j, z(t) = \theta | s(t-1) = i, a(t), h(t-1)) \\ & \cdot P(s(t) = j, z(t) = \theta | a(t), h(t-1)) = \\ & = \sum_i P(s(t) = j | s(t-1) = i, a(t), h(t-1)) \cdot P(z(t) = \theta | s(t-1) = i, s(t) = j, a(t), h(t-1)) \\ & \cdot P(s(t-1) = i | a(t), h(t-1)) = \\ & = \sum_i \pi_i(t-1) \cdot p_{ij}^{a(t)} \cdot r_{ij\theta}^{a(t)}. \end{aligned}$$

όπου οι δύο πρώτοι όροι στο γινόμενο οφείλονται στη Μαρκοβιανή ιδιότητα και στον ορισμό του μηχανισμού σημάτων αντίστοιχα, ενώ στον τρίτο όρο παρατηρούμε ότι η απόφαση

$a(t)$ είναι συνάρτηση του $h(t-1)$, και επομένως δε δίνει καμιά επιπλέον πληροφορία σχετικά με την κατάσταση $s(t-1)$.

Τέλος ο παρανομαστής στην (1) είναι ίσος με:

$$\begin{aligned} P(z(t)=\theta|a(t), h(t-1)) &= \\ &= \sum_i P(s(t)=j, z(t)=\theta|a(t), h(t)) \\ &= \sum_{i,j} \pi_i(t-1) \cdot p_{ij}^{a(t)} \cdot r_{ij\theta}^{a(t)}, \end{aligned}$$

και επομένως προκύπτει ότι

$$\pi_j(t) = \frac{\sum_i \pi_i(t-1) p_{ij}^{a(t)} r_{ij\theta}^{a(t)}}{\sum_{i,j} \pi_i(t-1) p_{ij}^{a(t)} r_{ij\theta}^{a(t)}}.$$

Καταλήγουμε στο συμπέρασμα ότι για τον υπολογισμό του διανύσματος πληροφόρησης μετά την περίοδο απόφασης t χρειαζόμαστε μόνο το $\pi(t-1)$, το διάνυσμα πληροφόρησης μετά την περίοδο απόφασης $t-1$. Επομένως το $\pi(t-1)$ περιέχει όλη την πληροφορία που χρειαζόμαστε πριν την περίοδο απόφασης t και αποτελεί μια επαρκή στατιστική συνάρτηση για όλη την ιστορία του συστήματος $h(t-1)$.

□

Επομένως αν η εκ των προτέρων κατανομή πιθανότητας της κατάστασης του συστήματος είναι π , λαμβάνεται απόφαση a και δίνεται το σήμα θ , τότε από το Θεώρημα του Bayes η εκ των υστέρων κατανομή π' δίνεται από:

$$\pi'_j = \left[\sum_i \pi_i p_{ij}^a r_{j\theta}^a \right] / \left[\sum_{i,j} \pi_i p_{ij}^a r_{j\theta}^a \right], \quad j=1, \dots, N. \quad (2)$$

Για ευκολία συμβολίζουμε με:

$$\pi' = T \langle \pi | a, \theta \rangle, \quad (3)$$

όπου T είναι τελεστής από το π στο π' που ορίζεται από την (2).

Το πρόβλημα βελτιστοποίησης συνιστά στο να βρεθεί μια πολιτική λήψης αποφάσεων που εξαρτάται μόνο από το διάλυμα πληροφοριών π σε κάθε περίοδο και μεγιστοποιεί το αναμενόμενο συνολικό όφελος για ορίζοντα N περιόδων.

Για το πρόβλημα αυτό ορίζουμε ως $V_n(\pi)$ την μέγιστη αναμενόμενη συνολική ωφέλεια αν η παρούσα κατάσταση είναι π και απομένουν n στάδια πριν το τέλος. Η $V_n(\pi)$ ικανοποιεί την αναδρομική σχέση:

$$V_n(\pi) = \max_a \left[\sum_i \pi_i \sum_j p_{ij}^a \sum_{\theta} r_{j\theta}^a [w_{ij\theta}^a + V_{n-1}(T \langle \pi | a, \theta \rangle)] \right]. \quad (4)$$

Το πρόβλημα μπορεί να λάβει τη μορφή προβλήματος μαθηματικού δυναμικού προγραμματισμού [2] αφού μπορούμε να απλοποιήσουμε την σχέση (4) θεωρώντας το αναμενόμενο άμεσο όφελος αν το σύστημα βρίσκεται στην κατάσταση i και επιλέγουμε την απόφαση a :

$$q_i^a = \sum_{j,\theta} p_{ij}^a r_{j\theta}^a w_{ij\theta}^a. \quad (5)$$

Τότε από την (5) με αντικατάσταση στην (4) έχουμε την:

$$V_n(\pi) = \max_a \left[\sum_i \pi_i q_i^a + \sum_{i,j,\theta} \pi_i p_{ij}^a r_{j\theta}^a V_{n-1}[T\langle \pi | a, \theta \rangle] \right]. \quad (6)$$

Επιπλέον αν q_i^0 αντιστοιχεί στην αναμενόμενη αμοιβή στο τερματισμό στην κατάσταση i τότε η συνάρτηση αναμενόμενης τερματικής ωφέλειας με τελική κατάσταση π είναι:

$$V_0(\pi) = \sum_i \pi_i q_i^0. \quad (7)$$

Παρόλη την πολυπλοκότητα της συνάρτησης ωφέλειας για n στάδια το πρόβλημα μπορεί να απλοποιηθεί αρκετά αξιοποιώντας το παρακάτω θεώρημα:

Θεώρημα [3] : Η αντικειμενική συνάρτηση $V_n(\pi)$ είναι κυρτή και κατά τμήματα γραμμική ως προς π , και επομένως μπορεί να γραφτεί στην μορφή:

$$V_n(\pi) = \max_{k=1,2,\dots,m(n)} \left[\sum_i \pi_i u_i^k(n) \right], \quad (8)$$

όπου $u^k(n) = (u_1^k(n), \dots, u_N^k(n))'$, $k=1,2,\dots,m(n)$ είναι διανύσματα τα οποία στη συνέχεια θα ονομάσουμε α -διανύσματα ή διανύσματα βελτιστοποίησης ή βέλτιστα διανύσματα για το στάδιο n .

Επομένως αν έχουμε υπολογίσει όλα τα βέλτιστα διανύσματα για το στάδιο n τότε για οποιαδήποτε κατάσταση π η αντίστοιχη τιμή της αντικειμενικής συνάρτησης προκύπτει από μια απλή μεγιστοποίησης πάνω σε όλα τα βέλτιστα διανύσματα του σταδίου. Τα διανύσματα αυτά επομένως χαρακτηρίζουν περιοχές και χωρίζουν το χώρο καταστάσεων σε

κυρτά πολύεδρα πάνω στα οποία μοναδικά διανύσματα μεγιστοποιούν την αντικειμενική συνάρτηση.

Απόδειξη: Η απόδειξη θα γίνει με επαγωγή:

Για $n=0$ η (7) ισχύει.

Έστω ότι ισχύει για $n-1$, δηλαδή $V_{n-1}(\pi) = \max_k [\pi \cdot u^k(n-1)]$, όπου $\pi \cdot u^k(n-1)$

συμβολίζει το εσωτερικό γινόμενο.

Επομένως ο χώρος καταστάσεων $\Omega = \{\pi = (\pi_1, \dots, \pi_N), \pi_j \geq 0, \sum_{j=1}^N \pi_j = 1\}$ αποτελείται

από την ένωση ενός πεπερασμένου συνόλου περιοχών (όχι ξένων μεταξύ τους) σε κάθε μια

από τις οποίες ισχύει $V_{n-1}[\pi] = \pi \cdot u^k(n-1)$ για κάποιο k .

Θεωρούμε την ποσότητα

$$V_{n-1}(T\langle \pi | a, \theta \rangle) = \max_k [T\langle \pi | a, \theta \rangle \cdot u^k(n-1)], \quad (9)$$

Που εμφανίζεται στην εξίσωση βελτιστότητας (6). Επειδή η $T\langle \pi | a, \theta \rangle$ είναι συνεχής και φραγμένη συνάρτηση του π , προκύπτει ότι το μέγιστο στην (9) είναι καλά ορισμένο.

Επομένως ορίζεται μια (όχι απαραίτητα μονοσήμαντη) απεικόνιση $l(\pi, a, \theta) = k$ που αντιστοιχεί σε κάποιον από τους δείκτες των διανυσμάτων βελτιστοποίησης του σταδίου $n-1$, έτσι ώστε

$$V_{n-1}(\pi) = \pi \cdot u^{l(\pi, a, \theta)}(n-1). \quad (10)$$

Υπό αυτή την έννοια τα διανύσματα αυτά χαρακτηρίζουν περιοχές του χώρου καταστάσεων. Με αντικατάσταση της (2) στην (9) έχουμε ότι

$$V_{n-1}(T\langle\pi|a, \theta\rangle) = \max_k [\sum_j \pi'_j u_j^k(n-1)]$$

Συνεπώς με αντικατάσταση στην (6) :

$$V_n(\pi) = \max_a [\sum_i \pi_i q_i^a + \sum_{i,j,\theta} \pi_i p_{ij}^a r_{j\theta}^a V_{n-1}[\max_k [\sum_j \pi'_j u_j^k(n-1)]]].$$

Από την (10) προκύπτει διαμέριση $\Delta = B_1 \cup B_2 \cup \dots \cup B_{m(n-1)}$, όπου ισχύει $l(\pi, a, \theta) = k, \forall \pi \in B_k$. Επομένως καταλήγουμε στην:

$$V_n(\pi) = \max_a [\sum_i \pi_i [q_i^a + \sum_{j,\theta} p_{ij}^a r_{j\theta}^a u_j^{l(\pi,a,\theta)}(n-1)]] \quad (11)$$

Για κάθε a και για κάθε i , έχουμε ότι $A_i^a := q_i^a + \sum_{j,\theta} p_{ij}^a r_{j\theta}^a u_j^{l(\pi,a,\theta)}(n-1)$ είναι σταθερή για κάθε π στο B_k για κάθε k .

Άρα $\sum_i \pi_i A_i^a$ είναι γραμμική στο B_k για κάθε k και κυρτή, λόγω της μεγιστοποίησης στην (9) άρα είναι κατά τμήματα γραμμική και κυρτή σε όλο το Δ . Άρα και το μέγιστο αυτής ως προς a , το ζητούμενο έπεται. ■

Αν για το $n-1$ στάδιο έχουμε υπολογίσει τα διανύσματα $u^k(n-1), k=1,2,\dots$ τότε από την (10) και (11) μπορούμε να υπολογίσουμε το αντίστοιχο βέλτιστο διάνυσμα για κάθε κατάσταση π για ορίζοντα n . Επιπλέον από την (11) μαζί με το βέλτιστο διάνυσμα προκύπτει και η αντίστοιχη βέλτιστη απόφαση a . Επομένως για των υπολογισμό της βέλτιστης πολιτικής δεν χρειάζεται να καταγράψουμε τις περιοχές αλλά αρκεί να βρούμε όλα τα βέλτιστα διανύσματα $u^k(n)$ μαζί με τις αντίστοιχες βέλτιστες αποφάσεις a . Τότε για τον υπολογισμό της βέλτιστης απόφασης για τυχόν κατάσταση π αρκεί να εκτελέσουμε την με-

γιστοποίηση από την (8) και αυτό θα μας δώσει το βέλτιστο διάνυσμα και την αντίστοιχη βέλτιστη απόφαση.

3 Αλγόριθμος του Sondik

Ο Αλγόριθμος του Sondik (Sondik's One Pass) [3] είναι η πρώτη ακριβής μέθοδος για προβλήματα πεπερασμένου ορίζοντα που διατυπώθηκε στη διδακτορική του διατριβή το 1971 και υπολογίζει τα βέλτιστα διανύσματα σε ένα πεπερασμένο σύνολο από σημεία του χώρου καταστάσεων. Ο αλγόριθμος θεωρητικά εξασφαλίζει την εύρεση όλων των διανυσμάτων $u^k(n)$ που ορίστηκαν στο κεφάλαιο 2 και ο αριθμός των σημείων που απαιτείται για τους υπολογισμούς ισούται με τον αριθμό των διανυσμάτων.

3.1 Περιγραφή του αλγορίθμου

Έστω ότι γνωρίζουμε όλα τα διανύσματα $u^k(n-1)$ για το $n-1$ στάδιο μαζί με τις αντίστοιχες βέλτιστες επιλογές a . Ξεκινάμε από τυχόν σημείο π^0 του χώρου καταστάσεων και υπολογίζουμε χρησιμοποιώντας την (11) το βέλτιστο διάνυσμα $u^0(n)$ και την αντίστοιχη βέλτιστη απόφαση a^0 για αυτό το σημείο. Η ιδέα είναι ότι για $\pi \neq \pi^0$ το διάνυσμα αυτό παύει να παραμένει βέλτιστο είτε αν η ποσότητα $l(\pi, a^0, \theta)$ αλλάξει από την $l(\pi^0, a^0, \theta)$ για κάποιο σήμα θ , είτε αν η βέλτιστη απάντηση αλλάξει.

Η πρώτη περίπτωση θα συμβεί όταν το π απομακρύνεται από το π^0 μέχρι ένα από τα σημεία $T\langle \pi | a^0, \theta \rangle$ να ξεπεράσει το σύνορο της περιοχής σταθερού $l(\pi, a^0, \theta)$ στο στάδιο $n-1$ για κάποιο θ . Για κάθε θ η συνθήκη ώστε το $T\langle \pi | a^0, \theta \rangle$ να παραμείνει στο χωρίο που ορίζει το $l(\pi^0, a^0, \theta)$ είναι:

$$T\langle \pi | a^0, \theta \rangle a^{l(\pi^0, a^0, \theta)}(n-1) \geq T\langle \pi | a, \theta \rangle a^k(n-1), \quad \forall k=1,2,\dots \quad (12)$$

Αντικαθιστώντας έχουμε ότι:

$$\sum_{i,j} \pi_i P_{ij}^{a^0} r_{j\theta}^{a^0} [u_j^{l(\pi^0, a^0, \theta)}(n-1) - u_j^k(n-1)] \geq 0, \quad \forall k. \quad (13)$$

Η δεύτερη περίπτωση συμβαίνει όταν αλλάξει η βέλτιστη απόφαση. Για να συνεχίσει το a^0 να είναι η βέλτιστη απόφαση πρέπει και αρκεί:

$$\pi \cdot u^0(n) \geq \pi \cdot u_a(n), \quad \forall a, \quad (14)$$

δηλαδή,

$$\sum_i \pi_i [u_i^0(n) - u_{a,i}(n)] \geq 0, \quad \forall a, \quad (15)$$

Όπου $u_a(n)$ είναι το διάνυσμα βελτιστοποίησης που υπολογίζεται από την (11) για την απόφαση a .

Επιπλέον έχουμε τις συνθήκες κανονικότητας:

$$\pi_i \geq 0, \quad 1 \leq i \leq N \quad \text{και} \quad \sum_{i=1}^N \pi_i = 1. \quad (16)$$

Οι σχέσεις (13), (15) και (16) χαρακτηρίζουν την περιοχή στο χώρο καταστάσεων πάνω στην οποία το $u^0(n)$ ορίζει την συνάρτηση ωφέλειας. Επίσης η βέλτιστη απάντηση σε αυτή την περιοχή είναι a^0 που προκύπτει από την απόφαση εκείνη που μεγιστοποιεί την ποσότητα στην (11). Με χρήση γραμμικού προγραμματισμού μπορούμε να εντοπίσουμε το ελάχιστο αριθμό συνθηκών τύπου (13), (15), (16) που απαιτούνται για τον καθορισμό της περιοχής.

Ο αλγόριθμος συνοψίζεται στα εξής βήματα:

1. Επιλέγουμε αρχική κατάσταση π^0 και υπολογίζουμε με χρήση της (11) την βέλτιστη απόφαση a^0 και το αντίστοιχο βέλτιστο διάνυσμα $u^0(n)$.
2. κατασκευάζουμε την πλήρη λίστα συνθηκών (11), (13), (15), (16).
3. Με χρήση γραμμικού προγραμματισμού καθορίζουμε τις συντομικές συνθήκες του $u^0(n)$. Από κάθε σύνορο με χρήση της (11) βρίσκουμε νέο α -διάνυσμα και αποθηκεύουμε σε μια λίστα το κάθε ένα από αυτά: το α -διάνυσμα, την βέλτιστη απόφαση και το σημείο προέλευσης.
4. Αν υπάρχουν στην λίστα α -διανύσματα των οποίων οι περιοχές δεν έχουν υπολογιστεί επιλέγουμε ένα από αυτά και επιστρέφουμε στο 2. Διαφορετικά έχουμε υπολογίσει όλα τα α -διανύσματα για το n στάδιο.

Ο παραπάνω αλγόριθμος μπορεί να χρησιμοποιηθεί διαδοχικά για να υπολογίσουμε τα βέλτιστα ή α -διανύσματα για κάθε στάδιο.

Με λίγα λόγια ο αλγόριθμος ξεκινάει από ένα αρχικό σημείο π^0 χρησιμοποιώντας την (11) υπολογίζουμε το βέλτιστο διάνυσμα $u^0(n)$ στο σημείο αυτό και την βέλτιστη απόφαση a . Καταχωρούμε σε μια λίστα το σημείο, το διάνυσμα και την απόφαση και τρέχουμε το αντίστοιχο π.γ.π ελαχιστοποίησης με συνθήκες (13), (15), (16) και αντικειμενική συνάρτηση κάθε φορά έναν από τους περιορισμούς. Οι βέλτιστες λύσεις που μπορεί να προκύψουν αποτελούν νέα συντομικά σημεία τα οποία τα εισάγουμε στην λίστα για να υπολογίσουμε τα νέα διανύσματα και βέλτιστες αποφάσεις. Ο αλγόριθμος τρέχει μέχρι να

εξαντληθούν όλα τα σημεία προς εξέταση και μετά προχωράει με τον ίδιο τρόπο στο επόμενο στάδιο.

3.2 Συζήτηση - Πλεονεκτήματα

Σε προβλήματα με μεγάλο πλήθος καταστάσεων, μέθοδοι που χρησιμοποιούν διακριτό πλέγμα για να περιγράψουν ή να προσεγγίσουν το χώρο καταστάσεων γενικά είναι πρακτικά ασύμφωρες, λόγω του μεγάλου πλήθους των υπολογισμών που απαιτούνται κάθε φορά, αργότερα όμως θα εξετάσουμε και μια τέτοια μέθοδο η οποία όμως δεν έχει αυτό το μειονέκτημα. Η μέθοδος του Sondik [3] δεν απαιτεί διακριτοποίηση του χώρου καταστάσεων αλλά εκτελεί τους υπολογισμούς σε ένα υποσύνολο από πεπερασμένα σημεία, των οποίων το πλήθος είναι το ελάχιστο δυνατό ώστε να υπολογίσουμε όλα τα διανύσματα.

Στην περίπτωση που υποθέτουμε συντελεστή αποπληθωρισμού ο αλγόριθμος δουλεύει ακριβώς με τον ίδιο τρόπο μετά την απαραίτητη τροποποίηση της συνάρτησης ωφέλειας από την σχέση (11).

Ο αλγόριθμος εφαρμόζεται με κάποιες αλλαγές στις πράξεις, ακόμη και όταν η δομή του προβλήματος αλλάζει από (απόφαση, μετάβαση, παρατήρηση) σε (απόφαση, παρατήρηση, μετάβαση).

3.2.1 Κριτική - Διορθώσεις

Ο Lovejoy [4] προτείνει τις εξής συνθήκες:

$$\sum_{i,j} \pi_i p_{ij}^{a^0} r_{j\theta}^{a^0} [u_j^{l(\pi^0, a^0, \theta)}(n-1) - u_j^k(n-1)] \geq 0, \quad \forall \theta, \quad (17)$$

$$\forall k=1,2,\dots, k \neq l(\pi^0, a^0, \theta),$$

$$\sum_{i,j} \pi_i p_{ij}^a r_{j\theta}^a [u_j^{l(\pi^0, a, \theta)}(n-1) - u_j^k(n-1)] \geq 0, \quad \forall \theta, \quad (18)$$

$$\forall a \neq a^0, \quad \forall k=1,2,\dots, \quad k \neq l(\pi^0, a^0, \theta).$$

Η συνθήκη (13) του Sondik [3] για κάθε θ που είναι ίδια με την (17) από πάνω εξασφαλίζει ότι αρκεί να εξετάσουμε για αλλαγές στο δείκτη $l(\pi, a, \theta)$ μόνο για $a^0 = a$ δηλαδή μόνο όταν η βέλτιστη απάντηση παραμένει ίδια. Ο ισχυρισμός αυτός όμως είναι λάθος διότι ισχύει μονάχα στην περίπτωση όπου οι δείκτες $l(\pi, a, \theta) = l(\pi_0, a, \theta)$ για όλα τα a το οποίο για να ισχύει απαιτεί τις (17) και (18) από παραπάνω και δεν αρκεί μονάχα η (17). Αν αξιοποιήσουμε μόνο την (17) υπάρχει κίνδυνος να υπολογίσουμε περιοχές μεγαλύτερες από τις πρέπουσες. Δηλαδή κάποια από τα σημεία τα οποία αντιστοιχούν σε άλλα διανύσματα να βρίσκονται κατά λάθος εντός αυτών των περιοχών, όπως προκύπτει με το παράδειγμα του Lovejoy [4]. Άρα οι συνθήκες όπως ορίζονται από τον Lovejoy είναι απαραίτητες διότι χαλαρώνοντας τις συνθήκες όπως στην περίπτωση του Sondik [3] υπάρχει κίνδυνος να παράγουμε περιοχές πολύ μεγάλες για τις οποίες κάποια από τα χρήσιμα σύνορα μπορεί να χαθούν.

3.3 Παραλλαγές και δυσκολίες στις μεθόδους ακριβείας

Στην εργασία του Lovejoy [4] γίνεται αναφορά και σε άλλους αλγόριθμους ακριβείας. Αναφορικά στο άρθρο υπάρχει ο αλγόριθμος του Cheng [5] όπου βασίζεται σε μια εναλλακτική χαλάρωση των συνθηκών (16), (17), (18) διαφορετική από αυτή του Sondik. Σε αυτή την περίπτωση κάθε κορυφή που προκύπτει από την αντίστοιχη περιοχή με τις χαλαρές συνθήκες, εξετάζεται για νέα α -διανύσματα και νέα περιοχή. Δηλαδή χαλαρώνοντας τις συνθη-

κές συνθήκες ο Cheng κατασκευάζει μια περιοχή την χαλαρή-περιοχή, κάθε κορυφή αυτής εξετάζεται στη συνέχεια αν μας δίνει νέα α -διανύσματα και άρα και μια νέα χαλαρή-περιοχή. Συστηματικά ο αλγόριθμος επιδιορθώνει αυτές τις περιοχές μέχρι να κατασκευάσει και την σωστή διαμέριση για το χώρο καταστάσεων Δ . Παρόμοιους αλγορίθμους έχουν προτείνει ο Eagle [6] και ο Monahan [7].

Ο Cheng ξεκινάει και υπολογίζει τα α -διανύσματα στα άκρα του Δ και τα καταχωρεί σε ένα σύνολο. Για κάθε α -διάνυσμα υπολογίζει την αντίστοιχη κυρτή περιοχή και την εκτίμηση της βέλτιστης τιμής. Εξετάζοντας για κάθε σημείο το σφάλμα της εκτίμησης εξετάζει το ενδεχόμενο οι περιοχές που υπολόγισε να είναι οι σωστές ή να είναι μεγαλύτερες. Σε αυτή την περίπτωση για το σημείο στο οποίο το σφάλμα είναι διάφορο του μηδενός υπολογίζει στο νέο α -διάνυσμα και την αντίστοιχη περιοχή. Ο αλγόριθμος τρέχει μέχρι να εξαντλήσει κάθε σημείο που παρουσιάζει σφάλμα.

Ο Monahan [7] προτείνει έναν παρόμοιο αλγόριθμο όμως εδώ το σύνολο με τα αρχικά διανύσματα είναι μεγαλύτερο κατά πολύ και ακολουθεί μια αντίστροφη διαδικασία απαλοιφής κάποιων από τα α -διανύσματα που δεν θα έπρεπε να θεωρηθούν υποψήφια. Ομοίως ο Eagle [6] προτείνει αλγόριθμο παρόμοιο με αυτό του Monahan μόνο που εμπεριέχει μια συνθήκη λιγότερη για τον καθορισμό των περιοχών στο κομμάτι του γραμμικού προγραμματισμού και προτείνει μια εναλλακτική συνθήκη για να εξετάσουμε αν τα α -διανύσματα είναι κατάλληλα ή όχι.

Γενικά αλγόριθμοι όπως του Sondik, Cheng, Monahan κ.α, που χρησιμοποιούν γραμμικό προγραμματισμό για τον εντοπισμό των συνόρων(συνοριακών σημείων) για τις περιο-

χές βελτιστοποίησης παρουσιάζουν μια βασική υπολογιστική δυσκολία. Θεωρητικά μπορούμε να βρούμε όλες τις άριστες κορυφές σε ένα πρόβλημα γραμμικού προγραμματισμού με την μέθοδο Simplex. Στην περίπτωση εκφυλισμένης λύσης με μια ακολουθία από γραμμοπράξεις μπορούμε να βρούμε όλες τις άριστες κορυφές. Στην πράξη όμως η κατασκευή αριθμητικών μεθόδων για των εντοπισμό όλων των άριστων κορυφών είναι ένα αυτόνομο και εξαιρετικά δύσκολο πρόβλημα. Αυτό το πρόβλημα εμπεριέχεται στους προαναφερθέντες αλγορίθμους. Επίσης η αλληπάλλληλη εφαρμογή της Simplex στους αλγορίθμους αυτούς δυσχεραίνει το πρόβλημα όσο αυξάνει ο αριθμός των καταστάσεων και των σταδίων.

Συνήθως υπάρχουν δυο παράμετροι που εξετάζει κανείς όταν συγκρίνει ανταγωνιστικούς αλγορίθμους: η υπολογιστική αποτελεσματικότητα και η διαφάνεια. Με τον όρο διαφάνεια εννοούμε την ευκολία με την οποία ένας ερευνητής εκπαιδευμένος στις βασικές τεχνικές της Επιχειρησιακής Έρευνας μπορεί να κατανοήσει και να εφαρμόσει τον αλγόριθμο. Όσο για την υπολογιστική αποτελεσματικότητα, αυτή εξαρτάται και από το μηχανημα υπολογιστή αλλά και τον κώδικα και την γλώσσα προγραμματισμού.

Κανένας από τους παραπάνω αλγορίθμους δεν μπορεί να βρει πρακτική εφαρμογή, μιας και βασίζονται σε κάθε χρονική περίοδο σε ένα πλήρη καθορισμό ενός συνόλου διανυσμάτων που είναι απαραίτητα για την κατασκευή της βέλτιστης πολιτικής. Το σύνολο αυτό είναι πεπερασμένο για προβλήματα πεπερασμένου ορίζοντα, αλλά ο πληθάρθμος του αυξάνει εκθετικά με το χρόνο. Η συνάρτηση βελτιστοποίησης θα συγκλίνει καθώς ο χρόνος t τείνει στο άπειρο σε μια κυρτή συνάρτηση αλλά όχι απαραίτητα κατά τμήματα γραμμική.

Γεωμετρικά μπορούμε να φανταστούμε τον αριθμό των ευθυγράμμων τμημάτων που χρειάζονται για την αναπαράσταση της συνάρτησης καθώς πάμε στο άπειρο αλλά δυστυχώς ένας απαγορευτικά μεγάλος αριθμός από διανύσματα μπορεί να προκύψει μόλις μετά από λίγες επαναλήψεις του αλγορίθμου.

Επιπλέον τα προγράμματα γραμμικού προγραμματισμού που έχουν σχεδιαστεί για να υπολογίσουν τα σύνορα ή για να μειώσουν τον αριθμό των υποψηφίων διανυσμάτων βασίζονται σε διαφορές της μορφής $\pi' \cdot u_0 - \pi' \cdot u$ όπου u_0, u υποψήφια διανύσματα. Καθώς αυτά τα διανύσματα πλησιάζουν το ένα το άλλο, οι διαφορές γίνονται πιο δύσκολα αντιληπτές από το μηδέν, καταλήγοντας έτσι σε προβλήματα ακρίβειας σε πρακτικές εφαρμογές λόγω της πεπερασμένης ακρίβειας των ηλεκτρονικών υπολογιστών. Τα α -διανύσματα θα πλησιάζουν το ένα το άλλο, χωρίς να είναι ίσα, για τυχόν υποσύνολο του χώρου καταστάσεων πάνω στο οποίο η ακολουθία της συνάρτησης βελτιστοποίηση συγκλίνει σε ένα λείο αλλά όχι γραμμικό όριο. Τα προβλήματα ακρίβειας θα προκύψουν πολύ πριν ένα τέτοιο όριο βρεθεί. Επομένως τα προβλήματα της μεγάλης διάστασης αλλά και της ακρίβειας είναι πολύ σημαντικά προβλήματα ώστε να καθιστούν ακόμη και τις πιο ρεαλιστικές ΜΟΜΔΑ αδύνατες να λυθούν με μεθόδους ακριβείας.

4 Προσεγγιστικές μέθοδοι για MOMΔΑ

Το βασικό υπολογιστικό πρόβλημα στους αλγορίθμους ακριβείας είναι ο υπεραριθμήσιμος χώρος καταστάσεων Δ , μιας και μπορούμε να το αναπαραστήσουμε στους αλγορίθμους μόνο με πεπερασμένο πλήθος από σημεία. Στο κεφάλαιο αυτό θα παρουσιάσουμε, κάποιες προσεγγιστικές μεθόδους. Υπάρχουν δύο ευρείς κατηγορίες μεθόδων προσέγγισης: *μέθοδοι πεπερασμένης μνήμης* και *μέθοδοι πεπερασμένου πλέγματος*. Για τις μεθόδους πεπερασμένης μνήμης μόνο μια πεπερασμένη ακολουθία από προηγούμενες αποφάσεις και σήματα διατηρείται στην μνήμη. Στην περίπτωση πεπερασμένου πλέγματος, γίνεται αναπαράσταση του Δ με ένα πεπερασμένο πλήθος από δειγματικά σημεία.

4.1 Προσεγγιστικές μέθοδοι πεπερασμένης μνήμης

Υπάρχει μια περιγραφή του Sondik [8] βασισμένος στον αλγόριθμο βελτιστοποίησης πολιτικής του Howard [9]. Έστω V_δ η τιμή μιας στάσιμης πολιτικής δ στο πρόβλημα πεπερασμένου ορίζοντα. Ο Howard [9] αποδεικνύει ότι οι αποφάσεις a που προκύπτουν για κάθε π κατά την μεγιστοποίηση της συνάρτησης $h(\pi, a, V_\delta)$, για κάθε περίοδο απόφασης καθορίζουν μια στάσιμη πολιτική, έστω δ' τέτοια ώστε $V_{\delta'} > V_\delta$ αν ισχύει $h(\pi, \delta'(\pi), V_\delta) > V_\delta(\pi)$ για κάποιο π . Το γεγονός ότι Δ είναι υπεραριθμήσιμο μας παρεμποδίζει σε αυτή τη διαδικασία μιας και η V_δ δεν είναι απαραίτητα κατά τμήματα γραμμική και κυρτή γι αυτό το λόγω ο Sondik [8] ορίζει μια κλάση από στάσιμες πολιτικές τις “πεπερασμένα μεταβατικές” (finitely transient), για τις οποίες η V_δ είναι κατά τμήματα γραμμική. Αν μια τέτοια πολιτική τύχει να είναι βέλτιστη τότε η $V^* = V_\delta$ θα είναι κατά

τιμήματα γραμμική και κυρτή. Όμως οι βέλτιστες πολιτικές δεν είναι απαραίτητα αυτής τις μορφής “πεπερασμένα μεταβατικές” και οι συνθήκες για να το εξετάσουμε εκ των προτέρων είναι δύσκολες.

Ο Sondik προτείνει έναν αλγόριθμο για τον εντοπισμό τέτοιων πολιτικών. Έστω τυχούσα στάσιμη πολιτική δ και ακέραιος $k < \infty$. Μπορούμε να προσεγγίσουμε την V_δ από μια κατά τμήματα γραμμική συνάρτηση που σε γενικές γραμμές μιμείται την “πεπερασμένα μεταβατική” συμπεριφορά για τις πρώτες k περιόδους. Όσο το k αυξάνει οι προσέγγιση βελτιώνεται, αλλά ο αριθμός των ευθυγράμμων τμημάτων αυξάνει επίσης. Έστω θ_j το διάνυσμα που αφορά το j -οστό ευθύγραμμο τμήμα που αφορά το k -οστό βήμα της προσέγγισης για την V_δ . Ο Sondik θέτει $V'_\delta(\pi) := \max[\pi' \theta_j : \forall j]$ ως μια δευτερεύουσα προσέγγιση συνάρτησης η οποία είναι κατά τμήματα γραμμική και κυρτή. Με αυτήν την συνάρτηση εκτελεί την μεγιστοποίηση και ολοκληρώνει το επαναληπτικό βήμα. Με αυτό κατασκευάζει μια νέα στάσιμη πολιτική και η διαδικασία επαναλαμβάνεται. Τα φράγματα για το σφάλμα προσέγγισης υπολογίζονται δυναμικά και αν αυτά ξεπερνούν ένα προκαθορισμένο όριο η διαδικασία επαναλαμβάνεται με ένα νέο μεγαλύτερο k . Ο Sondik αναφέρει και κάποια παραδείγματα στην διδακτορική του διατριβή, αλλά δεν επεκτείνεται σε υπολογιστικές εφαρμογές του αλγορίθμου.

Ο Platzman [10] προτείνει μια προσεγγιστική μέθοδο στην οποία οι αποφάσεις βασίζονται σε μια πεπερασμένη ακολουθία των πιο πρόσφατων αποφάσεων και παρατηρήσεων, αλλά τονίζει ότι για να υπάρχουν αξιόπιστα αποτελέσματα θα χρειαστεί μεγάλη μνήμη. Ο

Platzman αναφέρεται στις “καταστάσεις μνήμης” που είναι υποσύνολο του χώρου που περιέχει την εφικτή ιστορία. Για παράδειγμα μια κατάσταση μνήμης μπορεί να είναι μια συγκεκριμένη ακολουθία από πρόσφατα σήματα και αποφάσεις ή θα μπορούσε να είναι μια συσσωμάτωση από τέτοια ενδεχόμενα. Η μετάβαση της κατάστασης μνήμης συμβαίνει ως αποτέλεσμα της λήψης μιας απόφασης και παρατήρησης κάποιου σήματος, όπως γίνεται και η εσωτερική μετάβαση. Ο Platzman παρουσιάζει τεχνικές που εξετάζουν πολιτικές βασισμένες σε αυτό το πεπερασμένο χώρο αποφάσεων, φράσσοντας την απόδοση του αλγορίθμου σε σχέση με την βέλτιστη τιμή και για την κατασκευή νέων πεπερασμένων προσεγγίσεων. Επιπλέον επιτρέπει τυχαιοποιημένες πολιτικές, οι οποίες θεωρητικά δεν είναι απαραίτητες για να έχουμε βελτιστοποίηση, παρατηρεί όμως ότι με το να επιτρέπουμε και τυχαιοποιημένες πολιτικές μπορεί να βελτιστοποιηθεί η απόδοση του αλγορίθμου όταν έχουμε προσέγγιση με αποκοπή μνήμης. Η βασική ιδέα είναι ότι ο τυχαίος παράγοντας αποτελεί το κομμάτι της ιστορίας που είναι άγνωστο στον ερευνητή. Παρουσιάζονται κάποια αριθμητικά αποτελέσματα βασισμένα στις εργασίες του Sondik, αλλά ο συγγραφέας δεν αναφέρει επιπλέον εμπειρία και παραδείγματα σε υπολογιστικές εφαρμογές.

Οι White και Scherer [11] εξετάζουν και αυτοί με τη σειρά τους τις προσεγγιστικές μεθόδους βασισμένες στην αποκοπή της μνήμης στις ω πιο πρόσφατες αποφάσεις και παρατηρήσεις. Αν έχουμε m σήματα τότε υπάρχουν $(m|A|)^\omega$ δυνατά διανύσματα μνήμης με αποκοπή, είναι εφικτό να ξαναγράψουμε το πρόβλημα ως πρόβλημα δυναμικού προγραμματισμού με $(m|A|)^\omega$ καταστάσεις. Αν η κατανομή $\pi_{t-\omega}$ είναι γνωστή τότε αυτή μαζί με τις ω αποφάσεις και παρατηρήσεις μπορούν να καθορίσουν πλήρως την εκ των υστέρων

π_t κατανομή. Όμως η $\pi_{t-\omega}$ εξαρτάται από ολόκληρη την προηγούμενη ιστορία $H_{t-\omega}$ που δεν είναι γνωστή στον ερευνητή μιας και αυτός βασίζει τις αποφάσεις του μόνο στις ω πιο πρόσφατες περιόδους. Οι White και Scherer υπολογίζουν τα φράγματα για τις χειρότερες περιπτώσεις πάνω σε όλες τις $\pi_{t-\omega}$ ώστε να έχουν άνω και κάτω φράγμα για την συνάρτηση βελτιστοποίησης. Από τους υπολογισμούς προκύπτει ότι όπως και στην μέθοδο του Platzman [10] χρειάζεται γενικά μεγάλη ποσότητα μνήμης ώστε να έχουμε αξιόπιστα αποτελέσματα, αλλά σε προβλήματα όπου η πληροφορία του παρελθόντος σβήνει γρήγορα δηλαδή είναι λιγότερο σημαντική από την πιο πρόσφατη πληροφορία τα αποτελέσματα είναι εξαιρετικά.

Οι White και Scherer [11] ορίζουν έναν εργοδικό συντελεστή για Μπείζιανές μεταβάσεις ως την μέγιστη δυνατή απόσταση d ανάμεσα στα posterior διανύσματα π_t δοσμένης της ω και μιας τυχούσας $\pi_{t-\omega}$. Τα φράγματα για το σφάλμα προσέγγισης είναι σε αναλογία με αυτό το συντελεστή. Αυτό το αποτέλεσμα μας παρέχει ένα μέσο για την ανίχνευση των προβλημάτων για τα οποία η μέθοδος αποκοπής μνήμης θα ήταν εφικτή.

4.2 Προσεγγιστικές μέθοδοι πεπερασμένου πλέγματος

Οι διακριτές προσεγγίσεις για υπεραριθμήσιμους χώρους καταστάσεων αποτελούν μια συνήθη μέθοδο προσέγγισης για προβλήματα δυναμικού προγραμματισμού που πηγάει με φυσικό τρόπο. Με την εφαρμογή τους στις MOMΔΑ, αυτή η κλάση προσεγγίσεων παράγει μια ακολουθία από προσεγγίσεις των συναρτήσεων ωφέλειας σε ένα πεπερασμένο πλήθος σημείων $\pi \in \Delta$. Με αυτό το τρόπο εξομαλύνεται το στοιχειώδες πρόβλημα του υπεραριθ-

μήσιμου χώρου καταστάσεων. Ο Sondik [3] απέδειξε ότι για πεπερασμένο ορίζοντα και για κάθε κατά τμήματα γραμμική και κυρτή συνάρτηση ωφέλειας V_n , υπάρχει πεπερασμένο σύνολο από σημεία $\pi \in \mathcal{A}$ το πλήθος των οποίων αρκεί για να περιγραφεί πλήρως η V_n . Δυστυχώς η ανεύρεση αυτών των σημείων είναι ισοδύναμη με μια πλήρη ανανέωση για την συνάρτηση ωφέλειας, που έχει ως αποτέλεσμα όλα τα προαναφερθέντα προβλήματα. Οι προσεγγιστικές μέθοδοι πεπερασμένου πλέγματος χρησιμοποιούν μια πιο απλουστευμένη μέθοδο κατασκευής αυτού του πεπερασμένου συνόλου σημείων, αλλά ως αποτέλεσμα έχουμε προσεγγίσεις για την συνάρτηση ωφέλειας. Αυτή η κατηγορία μεθόδων χωρίζεται σε δυο υποκατηγορίες: *σταθερού πλέγματος* και *μεταβλητού πλέγματος*, όπου με τον όρο *πλέγμα* εννοούμε το πεπερασμένο σύνολο αυτό, από σημεία του χώρου καταστάσεων. Στην πρώτη περίπτωση το πλέγμα καθορίζεται μια φορά μόνο στην αρχή των υπολογισμών και παραμένει σταθερό καθ' όλη την διάρκεια της διαδικασίας. Στην δεύτερη περίπτωση η πληροφορία που παίρνουμε κατά την διάρκεια των υπολογισμών χρησιμοποιείται για την ανανέωση του πλέγματος. Με αυτό το τρόπο αυξάνεται το υπολογιστικό βάρος με κάθε ανανέωση, αλλά ως αποτέλεσμα έχουμε μικρότερο σφάλμα προσέγγισης αν οι ανανεώσεις γίνουν με το σωστό τρόπο.

4.2.1 Μέθοδοι σταθερού πλέγματος

Η κατασκευή ενός πεπερασμένου συνόλου από σταθερά σημεία για την αναπαράσταση ενός υπεραριθμήσιμου συνόλου είναι συνήθης τεχνική σε προβλήματα δυναμικού προγραμματισμού, και είναι μια από τις πρώτες μεθόδους προσέγγισης που σκέπτεται κανείς

για τις MOMΔΑ. Οι Kakalik [12] και Eckles [13] στην υπολογιστική τους εργασία χρησιμοποιούν μια προσέγγιση της συνάρτησης ωφέλειας βασισμένη σε μια γραμμική παρεμβολή ανάμεσα σε σταθερά, διακριτά σημεία του χώρου καταστάσεων, αλλά δεν διερευνούν την ποιότητα της προσέγγισης. Ο Sondik [14] αναπτύσσει μια φραγμένη προσέγγιση χρησιμοποιώντας την διακριτή προσέγγιση του Eckles [13].

Ο Lovejoy [15] χρησιμοποιεί ένα διακριτό πλέγμα για να κατασκευάσει άνω και κάτω φράγμα για την συνάρτηση ωφέλειας. Το κάτω φράγμα μπορεί να κατασκευαστεί από οποιοδήποτε πεπερασμένο σύνολο $P \subset \mathcal{A}$ του χώρου καταστάσεων. Ο αλγόριθμος χρησιμοποιεί το σύνολο L_n από a -διανύσματα του n σταδίου για να καθορίσει την συνάρτηση ωφέλειας $VL_n := \max[\pi' \cdot a^s(n) : a^s(n) \in L_n]$ όπως ακριβώς και στις μεθόδους ακριβείας. Όμως, το L_n έχει κατασκευαστεί από το L_{n-1} αξιοποιώντας την (11) στα σημεία του P μονάχα. Επομένως με το να κατασκευάσουμε a -διανύσματα μόνο σε αυτά τα σημεία, έχουμε μονάχα ένα υποσύνολο από τα συνολικά a -διανύσματα, και άρα το μέγιστο πλήθος που μπορούμε να κατασκευάσουμε είναι $|P|$. Τότε θα έχουμε ότι το μέγιστο πάνω σε ένα υποσύνολο είναι μικρότερο ή ίσο του μεγίστου σε όλο το σύνολο άρα ισχύει

$VL_n \leq V_n^*$, για κάθε n . Ο υπολογισμός της VL_n είναι το ίδιο εύκολος με τους αντίστοιχους υπολογισμούς σε ένα σύνθετο πρόβλημα δυναμικού προγραμματισμού με χώρο καταστάσεων P . Οι προσεγγιστικές πολιτικές μπορούν να παραχθούν αξιοποιώντας την VL_n αντί της V_n^* στην (11).

Για να κατασκευάσει άνω φράγμα για την συνάρτηση ωφέλειας ο Lovejoy, χρησιμοποιεί ένα ειδικό πλέγμα P σχεδιασμένο να τριγωνοποιήσει το χώρο καταστάσεων όπως ο Eaves [16]. Ομοίως και εδώ ο αλγόριθμος εφαρμόζεται μονάχα στα σημεία του P , και με αυτό το τρόπο κατασκευάζεται μια μοναδική, κατά τμήματα γραμμική συνάρτηση,

VU_n για την περίοδο απόφασης n . Από την ανισότητα Jensen και το γεγονός ότι η V_n^* είναι κυρτή έχουμε ότι $VU_n \geq V_n^*$ για κάθε n . Ο Lovejoy χρησιμοποιεί τα VL_n και VU_n για να καθορίσει σύνορα για την απώλεια της τιμής σε σχέση με την βέλτιστη τιμή που προκύπτει από τις πολιτικές της VL_n από την (11) για την περίπτωση πεπερασμένου αλλά και άπειρου ορίζοντα.

Ο πληθάριθμος του P είναι $\frac{(M+n-1)!}{M!(n-1)!}$, όπου M είναι το πλήθος των διαστημάτων στο οποίο διαιρείται η κάθε ακμή του εν λόγω simplex κατά την διάρκεια της διακριτοποίησης του χώρου καταστάσεων. Το $|P|$ γίνεται απαγορευτικά μεγάλο για υψηλές τιμές M και n , επομένως η τεχνική μπορεί να χειριστεί μονάχα αραιές περιοχές όταν το n , είναι μεγάλο.

4.2.2 Μέθοδοι μεταβλητού πλέγματος

Οι μέθοδοι μεταβλητού πλέγματος επιτρέπουν στα σημεία του πλέγματος που χρησιμοποιούμε να διαφέρουν σε κάθε επαναληπτικό στάδιο. Αν ειδική πληροφορία του προβλήματος χρησιμοποιείται για τον ανακαθορισμό του πλέγματος, τότε αυτές οι μέθοδοι παράγουν πιο ακριβείς προσεγγίσεις από αυτές του σταθερού πλέγματος.

Οι Sondik και Mendelson [17] προτείνουν μια διαδικασία βελτίωσης πολιτικής που εξετάζει μονάχα εκείνα τα σημεία του χώρου καταστάσεων που μπορούν να προκύψουν αν ακολουθήσουμε μια συγκεκριμένη πολιτική. Αυτό το πλέγμα μας βοηθάει να προσεγγίσουμε την πολιτική υπό εξέταση, επομένως το πλέγμα θα αλλάξει καθώς η πολιτική αλλάζει για κάθε επαναληπτικό βήμα.

Ο Cheng [5] σημειώνει ότι ο δικός του αλγόριθμος μπορεί να σταματήσει όταν το σφάλμα προσέγγισης πλησιάσει το μηδέν ή όταν ο αριθμός των α -διανυσμάτων φτάσει κάποιο άνω φράγμα. Σε κάθε περίπτωση, υπάρχουν φράγματα για τα αριθμητικά σφάλματα ανάμεσα στην περίπτωση αποκοπής και στην ακριβή διαδικασία. Ο αλγόριθμος του Cheng εισάγει νέα α -διανύσματα βασισμένος στις κορυφές που προκύπτουν από τις περιοχές που καταγράφουν και το μεγαλύτερο σφάλμα στους υπολογισμούς. Επομένως, ουσιαστικά επιλέγει ένα πεπερασμένο πλήθος από σημεία του χώρου καταστάσεων τα οποία τα εξετάζει βασισμένος πάνω σε φράγματα σφαλμάτων που προέκυψαν με δυναμικούς υπολογισμούς. Αν αποθηκεύσουμε το μέγιστο σφάλμα σε κάθε επανάληψη, είναι εύκολη υπόθεση να υπολογίσουμε ένα φράγμα για το συνολικό σφάλμα ανάμεσα στην κατασκευασμένη συνάρτηση ωφέλειας και την βέλτιστη όταν ο αλγόριθμος τερματιστεί.

4.3 Συζήτηση

Οι υλοποιήσιμες τεχνικές προσέγγισης γενικά για MOMΔΑ εξαρτώνται από την πεπερασμένη αναπαράσταση υπεραριθμήσιμων συνόλων. Γενικά όσο αυξάνει ο πληθώραριθμος της πεπερασμένης αναπαράστασης, η ποιότητα της προσέγγισης αλλά και το υπολογιστικό βάρος αυξάνονται. Το βάρος όμως αυξάνει με το πολύ γραμμικό ρυθμό αντί για εκθετικό

σε χρόνο n , επομένως οι προσεγγιστικές μέθοδοι μπορούν να αναλύουν προβλήματα με χρονικούς ορίζοντες πολύ μεγαλύτερους σε σχέση με τις ακριβείς μεθόδους. Η επιλογή του αναλυτή είναι ποιο είδος πεπερασμένης αναπαράστασης να επιλέξει.

Η μέθοδος πεπερασμένης μνήμης των White και Scherer [11] κατασκευάζει φράγματα από τις χειρότερες των περιπτώσεων prior κατανομών μερικές περιόδους πριν, και έτσι δεν θα ταιριάζουν σε προβλήματα όπου η παλιά ιστορία ακόμη έχει επίδραση στο παρόν. Όμως, θα μας έδιναν καλά αποτελέσματα σε προβλήματα όπου η παλιά ιστορία δεν έχει σχεδόν καμία επίδραση στο παρόν. Αναγνωρίζοντας ποια προβλήματα αντιστοιχούν σε κάθε μέθοδο αποτελεί ένα πεδίο για μελέτη. Υπάρχει μια στενή συσχέτιση ανάμεσα στην πεπερασμένη προσέγγιση για MOMΔΑ και την έρευνα για ορίζοντες πρόβλεψης σε πλήρως ορατές Μαρκοβιανές διαδικασίες αποφάσεων.

Οι μέθοδοι πεπερασμένου πλέγματος είναι μια φυσική τεχνική προσέγγισης, άμεσα υιοθετημένη από τους ειδικούς όταν γίνεται προσέγγιση υπεραριθμήσιων συνόλων. Συνήθως, υποθέτουμε την ποιότητα των προσεγγίσεων αντί να την υπολογίσουμε διότι τα φράγματα για τις χειρότερες περιπτώσεις είναι πολύ χαλαρά για να μας δώσουν ουσιαστική πληροφορία. Η εμπειρία στην πράξη είναι πως αυτές οι τεχνικές αποδίδουν καλά, πολύ καλύτερα από ότι προβλέπουν οι εκτιμήσεις των χειρότερων σεναρίων. Αυτή η συμπεριφορά οφείλεται στο ότι συνήθως η συνάρτηση ωφέλειας συμπεριφέρεται αρκούντως καλά ώστε να μπορεί να περιγραφεί με έναν πιο απλοϊκό τρόπο. Αυτό ισχύει στην πράξη σε προβλήματα με πλήρεις παρατηρήσεις. Για τις MOMΔΑ όμως, το σύνολο των προβλημάτων που έχουν

λυθεί με χρήση υψηλής ακρίβειας είναι πολύ μικρό για να αποδεχθούμε αυτή την βολική υπόθεση χωρίς περαιτέρω μελέτη.

Σε αριθμητικές δοκιμές χρησιμοποιώντας το διακριτό πλέγμα του Lovejoy [15], τα απλά heuristics για την κατασκευή πολιτικών από την (11) με χρήση της VL_n για να προσεγγίσουμε την V_n^* έχουν δουλέψει καλά ακόμη και με αραιά σύνολα σε μια σειρά από προβλήματα από την βιβλιογραφία. Στα δικά του δοκιμαστικά προβλήματα, ο Cheng [5] παρατήρησε ότι η δική του προσεγγιστική μέθοδος λειτούργησε καλύτερα από ότι υπεδείκνυαν τα φράγματα για τις χειρότερες περιπτώσεις. Όμως, και οι δύο συγγραφείς χρησιμοποίησαν δοκιμαστικά προβλήματα που ίσως να μην είναι αντιπροσωπευτικά του φάσματος των προβλημάτων που εντοπίζουν οι αναλυτές στην πράξη. Όπως και στην περίπτωση των μεθόδων πεπερασμένης μνήμης, ένα προφανές πεδίο έρευνας είναι η συσχέτιση των δεδομένων του προβλήματος με τις επιμέρους παραμέτρους που οδηγούν σε ποια μέθοδο προσέγγισης θα κατευθυνθούμε.

Επειδή οι διακριτές προσεγγίσεις είναι προσφιλείς για τις πλήρως ορατές ΜΔΑ, οι μέθοδοι πεπερασμένου πλέγματος είναι οι πιο ξεκάθαρες από όλες τις μεθόδους προσέγγισης. Οι μέθοδοι σταθερού πλέγματος είναι από τις πιο εύκολες στην επινόηση και εφαρμογή, μιας και το αρχικό σύνολο δηλαδή το πλέγμα κατασκευάζεται μόνο μια φορά στην αρχή. Οι μέθοδοι μεταβλητού πλέγματος αυξάνουν το υπολογιστικό βάρος μιας και διορθώνουν το πλέγμα σε κάθε βήμα. Όμως, ενδέχεται η ενημέρωση του πλέγματος σε κάθε

σημείο να μας δίνει καλύτερη ακρίβεια της προσέγγισης. Δεν υπάρχουν ακόμη αριθμητικά παραδείγματα για να υπολογίσουν την σχετική απόδοση αυτών των τεχνικών.

5 Αλγόριθμος Perseus

Ο αλγόριθμος Perseus [18] αποτελεί μια επαναληπτική μέθοδο σημειακής προσέγγισης για τον υπολογισμό Μερικώς Ορατών Μαρκοβιανών Διαδικασιών για μεγάλο πλήθος καταστάσεων.

5.1 Περιγραφή του αλγορίθμου

Με χρήση προσομοίωσης [19] κατασκευάζουμε σύνολο B το οποίο αποτελείται από σημεία του χώρου καταστάσεων. Με αυτό το τρόπο κατασκευάζουμε ένα υποσύνολο του χώρου καταστάσεων το οποίο αποτελείται από εφικτά σημεία. Ο στόχος του αλγορίθμου είναι σε κάθε στάδιο n να κατασκευάσει ένα τελεστή $V_{n+1} = \tilde{H}_{Perseus} V_n$ που βελτιώνει την τιμή της αντικειμενικής συνάρτησης πάνω σε αυτά τα σημεία ή τουλάχιστον δεν την ελαττώνει. Ισχύει ότι [18] :

$$V_n(\pi) \leq V_{n+1}(\pi), \quad \forall \pi \in B. \quad (19)$$

Αρχικά αφήνουμε τον αλγόριθμο να ανακαλύψει το χώρο τυχαία και να συλλέξει σημεία για το σύνολο B το οποίο παραμένει ίδιο για όλη την διάρκεια του αλγορίθμου. Αυτό γίνεται με μια μέθοδο προσομοίωσης [19].

Θέτουμε αρχικά, $q_i^0 = \frac{1}{1-\tau} \min_{i,a} q_i^{*a}$, $\forall i=1,2,\dots,N$ όπου q_i^{*a} είναι το τερματικό κόστος στην κατάσταση i όταν επιλέγουμε την απόφαση a και τ κατάλληλος συντελεστής αποπληθωρισμού με $\tau \in (0,1)$. Το Perseus ξεκινάει επομένως να προσεγγίσει την αντι-

κειμενική συνάρτηση σε κάθε στάδιο $n=1,2,\dots$ μέχρι να ικανοποιηθεί κάποιο κριτήριο σύγκλισης.

Έστω ότι είμαστε στο στάδιο n . Ο αλγόριθμος ξεκινάει με $V_n = \emptyset, \tilde{B} = B$, επιλέγει τυχαία σημείο $\pi \in \tilde{B}$ και με χρήση της (11) υπολογίζει το βέλτιστο διάνυσμα $u(n)$ για το σημείο αυτό και την αντίστοιχη βέλτιστη απόφαση a . Αν $\pi \cdot u(n) \geq V_{n-1}(\pi)$ τότε καταχωρούμε το $u(n)$ στην λίστα των διανυσμάτων για το στάδιο n , διαφορετικά επιλέγουμε ένα άλλο διάνυσμα από το $n-1$ στάδιο για το οποίο ισχύει

$$\pi \cdot u^s(n-1) \geq \pi \cdot u^i(n-1), \quad \forall i=1,2,\dots,m(n-1) .$$

Θέτουμε το $\tilde{B} = \{\pi \in B : V_n(\pi) < V_{n-1}(\pi)\}$ και συνεχίζουμε την βελτιστοποίηση πάνω στα $\pi \in \tilde{B}$ μέχρι που να έχουμε εξαντλήσει το \tilde{B} . Έτσι η νέα εκτίμηση της αντικειμενικής συνάρτησης είναι μη φθίνουσα ως προς n για κάθε σημείο του B .

Ο τελεστής $V_{n+1} = \tilde{H}_{Perseus} V_n$ βελτιώνει κάθε φορά την τιμή της αντικειμενικής συνάρτησης πάνω στα σημεία του B και όχι απαραίτητα σε ολόκληρο το Δ . Επιγραμματικά ο αλγόριθμος μπορεί να περιγραφεί ως εξής:

1. Θέτουμε $V_{n+1} = \emptyset, \tilde{B} = B$.
2. Προσομοιώνουμε με ομοιόμορφο τυχαίο τρόπο ένα σημείο b από το \tilde{B} και υπολογίζουμε από την (11) την ποσότητα $u(n+1)$.
3. Αν $b \cdot u(n+1) \geq V_n(b)$ τότε εισάγουμε το $u(n+1)$ στο V_{n+1} , διαφορετικά εισάγουμε το $u^*(n) : b \cdot u^*(n) \geq b \cdot u(n)^i$.
4. Υπολογίζουμε το $\tilde{B} = \{\pi \in B : V_{n+1}(\pi) < V_n(\pi)\}$.

5. Αν $\tilde{B} = \emptyset$ τότε σταματάμε, διαφορετικά επιστροφή στο 2.

Συνήθως ένας μικρός αριθμός από διανύσματα της μορφής $u(n)$ είναι αρκετός για να βελτιωθεί η V_n για κάθε $b \in \tilde{B}$, ειδικά στα πρώτα βήματα του αλγορίθμου. Η ιδέα είναι να κατασκευάσουμε αυτά τα διανύσματα με τυχαίο τρόπο από ένα υποσύνολο του \tilde{B} , το οποίο αποτελεί ένα συνεχώς αυξανόμενο υποσύνολο του B . Παρακολουθούμε τα σημεία του \tilde{B} για τα οποία η τιμή στην αντικειμενική συνάρτηση δεν έχει βελτιωθεί. Στην αρχή του κάθε επαναληπτικού σταδίου, θέτουμε την $V_{n+1} = \emptyset$, δηλαδή το \tilde{B} έχει αρχικοποιηθεί να είναι ίσο με B , το οποίο σημαίνει ότι όλα τα $b \in B$ χρειάζονται βελτίωση σε αυτό το στάδιο. Όσο το \tilde{B} είναι μη κενό, θεωρούμε ένα σημείο b από το \tilde{B} και υπολογίζουμε το αντίστοιχο $u(n+1)$ από την (11). Αν το $u(n+1)$ βελτιώνει την τιμή του b τότε το εισάγουμε στην λίστα με τα νέα α -διανύσματα και ενημερώνουμε την $V_{n+1}(b)$ για κάθε $b \in B$, και όλα αυτά τα σημεία τα βγάζουμε από το σύνολο \tilde{B} . Για όσο το \tilde{B} παραμένει μη κενό θεωρούμε με τυχαίο τρόπο σημεία από αυτό και υπολογίζουμε τα αντίστοιχα α -διανύσματα κάθε φορά.

Για το τερματισμό του κάθε σταδίου θα πρέπει να φροντίσουμε ώστε το \tilde{B} συνεχώς να μικραίνει όταν υπολογίζονται νέα α -διανύσματα, δηλαδή όταν το $u(n+1)$ βελτιώνει τουλάχιστον ένα $b \in B$, διαφορετικά θεωρούμε ένα από τα α -διανύσματα $u^s(n)$ του προηγούμενου σταδίου στη θέση του το οποίο μεγιστοποιεί την τιμή στο b . Θεωρούμε τώρα ότι το b έχει βελτιωθεί και το βγάζουμε από το σύνολο \tilde{B} μαζί με όλα τα υπόλοιπα

σημεία τα οποία βελτιώνονται από το α -διάνυσμα. Αυτή η διαδικασία καθιστά βέβαιο ότι το στάδιο θα λήξει.

5.2 Συζήτηση

Η βασική ιδέα πίσω από τον αλγόριθμο Perseus είναι ότι η συνάρτηση $V_n(\pi)$ δεν βελτιώνει μόνο το σημείο π αλλά και άλλα σημεία του B . Αυτό έχει ως συνέπεια οι αντικειμενικές συναρτήσεις να έχουν ένα σχετικά μικρό αριθμό διανυσμάτων δηλαδή $|V_n| \ll |B|$.

Στην πράξη αυτό μας επιτρέπει να ξεκινήσουμε με σύνολα B τα οποία είναι μεγαλύτερα από ότι σε άλλες σημειακές μεθόδους προσέγγισης και άρα ο βαθμός ακρίβειας στην προσέγγιση είναι καλύτερος. Επιπλέον σε σχέση με άλλες προσεγγιστικές μεθόδους η κατασκευή του B είναι οικονομική μιας και η μοναδική απαίτηση είναι περισυλλογή τυχαίου δείγματος από τον χώρο καταστάσεων ξεκινώντας από κάποιο π^0 . Επιπλέον τυχόν πολλαπλές καταχωρίσεις του ιδίου σημείου π εντός του B σημαίνει ότι το π είναι πιο πιθανό να επιλεγεί για τον σημειακό υπολογισμό του βέλτιστου διανύσματος για κάποιο στάδιο όμως αυτό δεν επηρεάζει το μέγεθος της V_n .

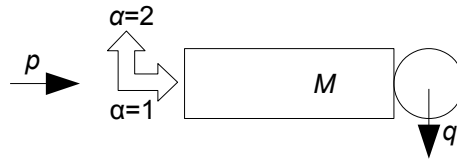
Το κυριότερο πρόβλημα με τις μεθόδους ακριβείας που απαιτούν τον υπολογισμό της βέλτιστης απόφασης για κάθε σημείο π του χώρου καταστάσεων είναι ότι ο αριθμός των διανυσμάτων για τις αντικειμενικές συναρτήσεις δηλαδή το μέγεθος της αντικειμενικής συνάρτησης αυξάνει εκθετικά καθώς ο ορίζοντας του προβλήματος μεγαλώνει. Για το λόγω αυτό και επινοούνται προσεγγιστικές μέθοδοι όπου συνήθως θεωρούμε μονάχα ένα υποσύνολο του χώρου καταστάσεων προς βελτιστοποίηση. Το κίνητρο για να χρησιμοποι-

ήσουμε προσεγγιστικές μεθόδους είναι ότι μας προσφέρουν τη δυνατότητα υπολογισμού πολιτικών πάνω σε μεγάλους χώρους καταστάσεων, πράγμα που τις καθιστά εφικτές παρ όλη την απώλεια σε βελτιστοποίηση. Παρόμοιες μέθοδοι έχουν αναλυθεί στην εργασία του Lovejoy [4].

Η βασική υπόθεση σε αυτές τις προσεγγιστικές σημειακές μεθόδους είναι ότι αν βελτιώνουμε την τιμή της αντικειμενικής συνάρτησης αλλά και το αντίστοιχο διάνυσμα για το συγκεκριμένο σημείο π , τότε η αντίστοιχη πολιτική γενικεύεται και μπορεί να εφαρμοστεί και σε σημεία τα οποία βρίσκονται εκτός του B . Αυτός είναι ένας διαισθητικός κανόνας και κατά πόσο ισχύει ή όχι εξαρτάται από την μορφή του προβλήματος. Πάντως σε πολλά προβλήματα το σύνολο B των εφικτών σημείων του χώρου καταστάσεων όταν ξεκινάμε από το π_0 μπορεί να καλυφθεί αρκετά πυκνά από ένα μικρό αριθμό του Simplex του χώρου καταστάσεων.

6 Εφαρμογή σε ουρά Geo/Geo/1/M

Στο κεφάλαιο αυτό παρουσιάζουμε μια εφαρμογή των αλγορίθμων MOMΔΑ σε ένα πρόβλημα ελέγχου αποδοχής απόρριψης πελατών σε μερικά παρατηρήσιμη ουρά αναμονής.



Σχήμα 1. Η ουρά Geo/Geo/1/M.

6.1 Περιγραφή του προβλήματος

Το πρόβλημα που μελετάμε αφορά μια Μαρκοβιανή ουρά εξυπηρέτησης με έναν υπηρέτη σε διακριτό χρόνο. Οι χρόνοι εξυπηρέτησης ακολουθούν Γεωμετρική κατανομή με παράμετρο q . Η χωρητικότητα της ουράς είναι M και οι πελάτες φτάνουν με Γεωμετρική κατανομή με παράμετρο p . Υποθέτουμε ότι σε κάθε περίοδο μπορεί να συμβεί μια μοναδική άφιξη, μια μοναδική αναχώρηση ή και τα δύο. Η σειρά προτεραιότητας είναι FCFS.

Υποθέτουμε επίσης ότι υπάρχει ένας ελεγκτής στην είσοδο που επιτρέπει ή όχι την είσοδο πελάτη ανάλογα με την κατάσταση του μήκους της ουράς στην αρχή της περιόδου.

Θεωρούμε ότι στην αρχή της κάθε περιόδου υπάρχει λειτουργικό κόστος h για κάθε πελάτη στο σύστημα ενώ για κάθε άφιξη πελάτη και αποδοχή στο σύστημα υπάρχει αμοιβή R .

Το πρόβλημα είναι να βρεθεί μια δυναμική πολιτική αποδοχής /απόρριψης πελατών που μεγιστοποιεί το αναμενόμενο εκπτώτικό όφελος σε άπειρο ορίζοντα.

Θεωρούμε δυο περιπτώσεις πολιτικών αποδοχής. Στην πρώτη περίπτωση (πλήρως παρατηρήσιμη ουρά) ο ελεγκτής κάθε χρονική στιγμή έχει πλήρη πληροφόρηση σχετικά με το μήκος της ουράς και με βάση αυτή αποφασίζει αν θα δεχτεί ή όχι ένα πελάτη.

Στην δεύτερη περίπτωση (μερικώς παρατηρήσιμη ουρά) ο ελεγκτής δε παρατηρεί την ακριβή κατάσταση του συστήματος, αλλά ένα από τα 3 δυνατά σήματα: το πρώτο ($\theta=1$) αντιστοιχεί σε κενό σύστημα και ανενεργό υπηρέτη, το δεύτερο ($\theta=2$) σε ενεργό υπηρέτη, το τρίτο ($\theta=3$) σε πλήρες σύστημα.

Με άλλα λόγια ο ελεγκτής παρατηρεί αν ο υπηρέτης είναι απασχολημένος ή όχι, και αν η ουρά είναι πλήρης, αλλά όχι τον ακριβή αριθμό πελατών στο σύστημα.

Στην περίπτωση που δεν γνωρίζουμε την κατάσταση i του συστήματος στην αρχή της κάθε περιόδου και μοναδική πληροφορία που έχουμε είναι το σήμα θ καλούμαστε να υπολογίσουμε την αντικειμενική συνάρτηση και την βέλτιστη πολιτική του μοντέλου. Σε αυτή τη περίπτωση το πρόβλημα μοντελοποιείται από μια Μερικώς Ορατή Μαρκοβιανή Διαδικασία όπως αυτή περιγράφεται στην ενότητα 1.1 σελίδα 1 στην εισαγωγή.

6.1.1 Το μοντέλο Μαρκοβιανής διαδικασίας αποφάσεων με μερική πληροφόρηση

Ως κατάσταση θεωρούμε το συνολικό αριθμό πελατών στο σύστημα. Επομένως ο χώρος καταστάσεων είναι το σύνολο $\Omega=\{0,1, \dots , M\}$.

Οι δυνατές αποφάσεις του ελεγκτή είναι $a=1$ (αποδοχή) και $a=2$ (απόρριψη).

Υποθέτουμε χωρίς βλάβη της γενικότητας ότι η απόφαση αποδοχής /απόρριψης λαμβάνεται στην αρχή της περιόδου, αφού γίνει γνωστή η κατάσταση του συστήματος και ανεξάρτητα από το αν έρθει ή όχι πελάτης. Με άλλα λόγια η απόφαση ισοδυναμεί με την ανάρτηση πινακίδας στην είσοδο του συστήματος που ενημερώνει τον πελάτη αν έρθει, για το αν γίνεται δεκτός ή όχι.

Με βάση τα παραπάνω, οι πιθανότητες πρώτης μετάβασης είναι:

$$\begin{aligned}
 a=1, \quad & p_{00}^1=1-p, p_{01}^1=p \\
 & p_{M,M-1}^1=q, P_{M,M}^1=1-q \\
 & \left\{ \begin{array}{l} p_{i,i-1}^1=(1-p)q \\ p_{i,i+1}^1=(1-q)p \\ p_{i,i}^1=(1-p)(1-q)+pq \end{array} \right\}, \quad i=1, \dots, M-1
 \end{aligned} \tag{20}$$

$$\begin{aligned}
 a=2, \quad & p_{00}^2=1, p_{0i}^2=0 \\
 & \left\{ \begin{array}{l} p_{i,i-1}^2=q \\ p_{i,i}^2=1-q \end{array} \right\}, \quad i=1, \dots, M
 \end{aligned}$$

Όπως αναφέρθηκε προηγουμένως υποθέτουμε ότι η κατάσταση δεν είναι πλήρως παρατηρήσιμη από τον ελεγκτή. Τα σήματα που λαμβάνει είναι $\theta=1$ όταν $i=0$, $\theta=2$ όταν $0 < i < M$ και $\theta=3$ για $i=M$.

Με βάση τα παραπάνω έχουμε τους πίνακες με τα $r_{j\theta}^a$ τις πιθανότητες των σημάτων και το άμεσο αναμενόμενο κέρδος q_i^a για κάθε κατάσταση i και απόφαση a :

Για $a=1$,

$$R^1 = \begin{matrix} & \theta \\ j & \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ \vdots & \vdots & \vdots \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad q^1 = \begin{pmatrix} p(R-h) \\ p(R-h)+qh-q \\ p(R-h)+qh-2q \\ \vdots \\ p(R-h)+qh-(M-1)q \\ (q-M)h \end{pmatrix} \end{matrix} \quad (21)$$

Για $a=2$,

$$R^2 = \begin{matrix} & \theta \\ j & \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ \vdots & \vdots & \vdots \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad q^2 = \begin{pmatrix} 0 \\ (q-1)h \\ (q-2)h \\ \vdots \\ (q-M+1)h \\ (q-M)h \end{pmatrix} \end{matrix} \quad (22)$$

6.1.2 Το πρόβλημα με πλήρη πληροφόρηση

Αρχικά θα λύσουμε το πρόβλημα πλήρους πληροφόρησης έτσι ώστε η σύγκριση του αναμενόμενου κέρδους ανάμεσα στα δυο προβλήματα να μας δώσει μια εκτίμηση για το κόστος της ελλιπούς πληροφόρησης. Για το πρόβλημα με πλήρη πληροφόρηση έχουμε ότι,

αν $V(i)$ είναι η μέγιστη συνολική αναμενόμενη ωφέλεια με παράγοντα αποπληθωρισμού τ , τότε η $V(i)$ ικανοποιεί την συνάρτηση βελτιστοποίησης:

$$V(i) = C(i) + \max \left[pR + \tau \sum_{j=0}^M p_{ij}^1 V(j), \tau \sum_{j=0}^M p_{ij}^2 V(j) \right], \quad \tau \in (0,1), \quad (23)$$

όπου $C(i) = -hi$ είναι το λειτουργικό κόστος στην αρχή της κάθε περιόδου όταν η ουρά βρίσκεται στην κατάσταση i . Προφανώς η $C(i)$ είναι φθίνουσα στο i .

Εφαρμόζοντας διαδοχικές προσεγγίσεις στο χώρο τιμών έχουμε:

$$\begin{aligned} V_n(0) &= \max [pR + \tau p V_{n-1}(1) + \tau(1-p) V_{n-1}(0), \tau V_{n-1}(0)], \\ V_n(M) &= -hM + \tau q V_{n-1}(M-1) + \tau(1-q) V_{n-1}(M), \\ V_n(i) &= -hi + \max [pR + \tau p(1-q) V_{n-1}(i+1) + \tau(1-p) q V_{n-1}(i-1) \\ &\quad + \tau[(1-p)(1-q) + pq] V_{n-1}(i), \tau q V_{n-1}(i-1) + \tau(1-q) V_{n-1}(i)], \end{aligned} \quad (24)$$

για κάθε $i \in \{1, 2, \dots, M-1\}$, $n = 0, 1, \dots$ με $V_0(i) = 0$, $\forall i$.

Πρόταση 1: Η $V_n(i)$ είναι φθίνουσα στο i , για κάθε n .

Απόδειξη: Η συνάρτηση ωφέλειας για το πεπερασμένο ορίζοντα είναι:

$$\begin{aligned} V_n(0) &= \max [pR + \tau p V_{n-1}(1) + \tau(1-p) V_{n-1}(0), \tau V_{n-1}(0)], \\ V_n(M) &= -hM + \tau q V_{n-1}(M-1) + \tau(1-q) V_{n-1}(M), \\ V_n(i) &= -hi + \max [pR + \tau p(1-q) V_{n-1}(i+1) + \tau(1-p) q V_{n-1}(i-1) \\ &\quad + \tau[(1-p)(1-q) + pq] V_{n-1}(i), \tau q V_{n-1}(i-1) + \tau(1-q) V_{n-1}(i)], \end{aligned} \quad (25)$$

για κάθε $i \in \{1, 2, \dots, M-1\}$.

Η απόδειξη θα γίνει με επαγωγή.

Για $n=0$, η $V_0(i)=0$ είναι τετριμμένα φθίνουσα. Έστω ότι ισχύει για $n-1$, δηλαδή $V_{n-1}(i)$ είναι φθίνουσα. Αφού $V_{n-1}(i)$ είναι φθίνουσα τότε και $V_{n-1}(i-1), V_{n-1}(i+1)$ είναι φθίνουσες και από την εξίσωση της $V_n(i)$ αφού όλοι οι συντελεστές εντός της αγκύλης είναι μη αρνητικοί, έχουμε άθροισμα φθινουσών που είναι φθίνουσα. Το maximum διατηρεί την μονοτονία και επίσης το $C(i)=-hi$ είναι φθίνουσα ως προς i και άρα τελικά ισχύει ότι $V_n(i)$ είναι φθίνουσα ως άθροισμα φθινουσών για κάθε $i \in \{1, 2, \dots, M-1\}$. Για τις οριακές τιμές $i=0, i=M$ έχουμε από την επαγωγική υπόθεση $V_{n-1}(0) \geq V_{n-1}(i) \geq V_{n-1}(M)$. Θα δείξουμε ότι η σχέση ισχύει και για n .

Περίπτωση 1

Έστω ότι $a=1$ είναι η βέλτιστη απόφαση στο στάδιο n για την κατάσταση $i=0$, τότε από την $V_n(0)$ έχουμε:

$$pR + \tau p V_{n-1}(1) + \tau(1-p)V_{n-1}(0) - \tau V_{n-1}(0) \geq 0.$$

Ισοδύναμα ισχύει:

$$pR + \tau p (V_{n-1}(1) - V_{n-1}(0)) \geq 0,$$

το οποίο είναι ισοδύναμο με:

$$0 \geq V_{n-1}(1) - V_{n-1}(0) \geq -\frac{R}{\tau}. \quad (26)$$

Υπάρχουν δύο υποπερίπτωσης για κάθε βέλτιστη απόφαση.

Έστω η πρώτη υποπερίπτωση όπου η βέλτιστη απόφαση στο στάδιο n για την κατάσταση $i=1$ είναι $a=1$. Αντικαθιστώντας:

$$\begin{aligned}
 V_n(0) - V_n(1) &= pR + \tau p V_{n-1}(1) + \tau(1-p)V_{n-1}(0) \\
 &\quad - pR - \tau p(1-q)V_{n-1}(2) - \tau(1-p)qV_{n-1}(0) \\
 &\quad - \tau[(1-p)(1-q) + pq]V_{n-1}(1) + h \\
 &= \tau p[V_{n-1}(1) - V_{n-1}(0)] + \tau q[V_{n-1}(1) - V_{n-1}(0)] \\
 &\quad - \tau[V_{n-1}(1) - V_{n-1}(0)] - \tau p(1-q)[V_{n-1}(2) - V_{n-1}(1)] \\
 &\quad - \tau p q[V_{n-1}(1) - V_{n-1}(0)] + h \\
 &= -\tau(1-p)(1-q)[V_{n-1}(1) - V_{n-1}(0)] - \tau p(1-q)[V_{n-1}(2) - V_{n-1}(1)] + h .
 \end{aligned}$$

Άρα έχουμε:

$$\begin{aligned}
 V_n(0) - V_n(1) &= \tau(1-p)(1-q)[V_{n-1}(0) - V_{n-1}(1)] \\
 &\quad + \tau p(1-q)[V_{n-1}(1) - V_{n-1}(2)] + h .
 \end{aligned} \tag{27}$$

Όμως λόγω της επαγωγικής υπόθεσης η $V_{n-1}(i)$ είναι φθίνουσα επομένως έχουμε ότι

$$V_{n-1}(0) - V_{n-1}(1) \geq 0 \quad \text{και} \quad V_{n-1}(1) - V_{n-1}(2) \geq 0 \quad \text{και αφού } p, q \in (0,1) \text{ προκύπτει ότι}$$

όλοι οι όροι στην (27) είναι θετικοί άρα :

$$V_n(0) - V_n(1) \geq 0. \tag{28}$$

Στη δεύτερη υποπερίπτωση όπου η βέλτιστη απόφαση στο στάδιο n για την κατάσταση $i=1$ είναι $a=2$, αντίστοιχα:

$$\begin{aligned}
 V_n(0) - V_n(1) &= pR + \tau p V_{n-1}(1) + \tau(1-p)V_{n-1}(0) - \tau q V_{n-1}(0) - \tau(1-q)V_{n-1}(1) + h \\
 &= pR + \tau p[V_{n-1}(1) - V_{n-1}(0)] + \tau q[V_{n-1}(1) - V_{n-1}(0)] \\
 &\quad - \tau[V_{n-1}(1) - V_{n-1}(0)] + h \\
 &= pR + \tau(p+q-1)[V_{n-1}(1) - V_{n-1}(0)] + h .
 \end{aligned}$$

Άρα έχουμε:

$$V_n(0) - V_n(1) \geq pR + \tau(p+q-1)[V_{n-1}(1) - V_{n-1}(0)]. \quad (29)$$

Όμως λόγω της (26) έχουμε:

$$pR + \tau(p+q-1)[V_{n-1}(1) - V_{n-1}(0)] \geq pR + \tau(p+q-1)\left(-\frac{R}{\tau}\right) = (1-q)R \geq 0.$$

Άρα:

$$V_n(0) - V_n(1) \geq 0. \quad (30)$$

Επομένως από τις (28) και (30) αν $a=1$ είναι η βέλτιστη απόφαση τότε η συνάρτηση ωφέλειας πεπερασμένου ορίζοντα είναι φθίνουσα δηλαδή για κάθε n ισχύει:

$$V_n(0) \geq V_n(i), i \in \{1, 2, \dots, M-1\} \quad (31)$$

Περίπτωση 2

Έστω ότι $a=2$ είναι η βέλτιστη απόφαση στο στάδιο n για την κατάσταση $i=0$, τότε από την $V_n(0)$ με ανάλογες πράξεις όπως και στην περίπτωση της (26) έχουμε:

$$V_{n-1}(1) - V_{n-1}(0) < -\frac{R}{\tau} \quad (32)$$

Αντίστοιχα με την περίπτωση 1 έστω η πρώτη υποπερίπτωση όπου $a=1$ είναι η βέλτιστη απόφαση στο στάδιο n για την κατάσταση $i=1$. Τότε

$$\begin{aligned}
 V_n(0) - V_n(1) &= \tau V_{n-1}(0) - pR - \tau p(1-q)V_{n-1}(2) - \tau(1-p)qV_{n-1}(0) \\
 &\quad - \tau[(1-p)(1-q) + pq]V_{n-1}(1) + h \\
 &= \tau[V_{n-1}(0) - V_{n-1}(1)] - \tau q[V_{n-1}(0) - V_{n-1}(1)] \\
 &\quad + \tau p(1-q)[V_{n-1}(1) - V_{n-1}(2)] + \tau p q[V_{n-1}(0) - V_{n-1}(1)] - pR + h \\
 &= \tau[1 - (1-p)q][V_{n-1}(0) - V_{n-1}(1)] \\
 &\quad + \tau p(1-q)[V_{n-1}(1) - V_{n-1}(2)] - pR + h .
 \end{aligned}$$

Άρα έχουμε:

$$\begin{aligned}
 V_n(0) - V_n(1) &= \tau[1 - (1-p)q][V_{n-1}(0) - V_{n-1}(1)] \\
 &\quad + \tau p(1-q)[V_{n-1}(1) - V_{n-1}(2)] - pR + h .
 \end{aligned} \tag{33}$$

Όμως τότε από την (32) η παραπάνω σχέση γίνεται:

$$\begin{aligned}
 V_n(0) - V_n(1) &= (1-p)(1-q)[V_{n-1}(0) - V_{n-1}(1)] \\
 &\quad + p(1-q)[V_{n-1}(1) - V_{n-1}(2)] + h .
 \end{aligned} \tag{34}$$

Όμως λόγω της επαγωγικής υπόθεσης η $V_{n-1}(i)$ είναι φθίνουσα επομένως έχουμε ότι

$$V_{n-1}(0) - V_{n-1}(1) \geq 0 \quad \text{και} \quad V_{n-1}(1) - V_{n-1}(2) \geq 0 \quad \text{και αφού } p, q \in (0,1) \text{ προκύπτει ότι}$$

όλοι οι όροι στην (34) είναι θετικοί άρα :

$$V_n(0) - V_n(1) \geq 0 . \tag{35}$$

Έστω η δεύτερη υποπερίπτωση όπου $\alpha=2$ είναι η βέλτιστη απόφαση στο στάδιο n για την κατάσταση $i=1$, τότε:

$$\begin{aligned}
 V_n(0) - V_n(1) &= \tau V_{n-1}(0) - \tau q V_{n-1}(0) - \tau(1-q)V_{n-1}(1) + h \\
 &= \tau[V_{n-1}(0) - V_{n-1}(1)] - \tau q[V_{n-1}(0) - V_{n-1}(1)] + h \\
 &= \tau(1-q)[V_{n-1}(0) - V_{n-1}(1)] + h .
 \end{aligned}$$

Όμως λόγω της επαγωγικής υπόθεσης η $V_{n-1}(i)$ είναι φθίνουσα, δηλαδή έχουμε ότι

$V_{n-1}(0) - V_{n-1}(1) \geq 0$ και αφού όλοι οι όροι είναι θετικοί επομένως:

$$V_n(0) - V_n(1) \geq 0. \quad (36)$$

□

Επομένως από τις (35) και (36) αν $a=2$ είναι η βέλτιστη απόφαση τότε η συνάρτηση ωφέλειας πεπερασμένου ορίζοντα είναι φθίνουσα δηλαδή για $i=0$ ισχύει:

$$V_n(0) \geq V_n(i), i \in \{1, 2, \dots, M\} \quad (37)$$

Η απόδειξη ότι $V_n(M-1) \geq V_n(M)$ είναι εντελώς ανάλογη. Άρα ισχύει ότι η μονοτονία της $V_n(i)$ διατηρείται σε ολόκληρο το $\{0, 1, 2, \dots, M\}$, δηλαδή ότι η $V_n(i)$ είναι φθίνουσα για κάθε n και για κάθε $i \in \{1, 2, \dots, M\}$. Επομένως η επαγωγή έχει ολοκληρωθεί και το συμπέρασμα έπεται.

Στο επόμενο θεώρημα αποδεικνύουμε ότι η συνάρτηση βέλτιστης τιμής είναι κοίλη. Αυτό συνεπάγεται ότι η βέλτιστη πολιτική έχει δομή κατωφλίου.

Έστω $\Delta_n(i) := V_n(i+1) - V_n(i)$. Από την Πρόταση 1 ισχύει ότι $\Delta_n(i) \leq 0 \quad \forall n, i$.

Θεώρημα:

- i. Η $\Delta_n(i)$ είναι φθίνουσα ως προς i για κάθε n , ή ισοδύναμα, η βέλτιστη τιμή $V_n(i)$ είναι κοίλη ως προς i για κάθε n .
- ii. Υπάρχουν ακέραιοι $k_n, n=0,1,\dots, -1 \leq k_n \leq M$ τέτοιοι ώστε για το στάδιο n , η βέλτιστη πολιτική έχει μορφή κατωφλίου:

$$a_n^*(i) = \begin{cases} 1 & , i \leq k_n \\ 2 & , i > k_n \end{cases}.$$

Απόδειξη: Υπενθυμίζουμε ότι οι εξισώσεις για την συνάρτηση ωφέλειας στην περίπτωση πεπερασμένου οριζοντα που δίνονται από την (25) είναι:

$$\begin{aligned} V_n(0) &= \max[pR + \tau p V_{n-1}(1) + \tau(1-p)V_{n-1}(0), \tau V_{n-1}(0)], \\ V_n(M) &= -hM + \tau q V_{n-1}(M-1) + \tau(1-q)V_{n-1}(M), \\ V_n(i) &= -hi + \max[pR + \tau p(1-q)V_{n-1}(i+1) + \tau(1-p)qV_{n-1}(i-1) \\ &\quad + \tau[(1-p)(1-q) + pq]V_{n-1}(i), \tau q V_{n-1}(i-1) + \tau(1-q)V_{n-1}(i)], \end{aligned}$$

για κάθε $i \in \{1, 2, \dots, M-1\}$. Έχουμε ορίσει την $\Delta_n(i) := V_n(i+1) - V_n(i)$. Η απόδειξη θα γίνει με επαγωγή στον αριθμό n των σταδίων.

Έστω $n=0$, τότε $V_0(i)=0$, Επομένως $\Delta_0(i)=0$, που είναι τετριμμένα φθίνουσα, και το i. ισχύει. Επίσης υποθέτοντας συμβατικά ότι σε περίπτωση ισοπαλίας η βέλτιστη απόφαση είναι $a=1$, το ii. ισχύει για $k_0=-1$.

Επαγωγική υπόθεση : Έστω ότι ισχύει το i. για κάποιο $n-1$. Τότε θα έχουμε ότι η $\Delta_{n-1}(i) = V_{n-1}(i+1) - V_{n-1}(i)$, είναι φθίνουσα, δηλαδή ισχύει $\Delta_{n-1}(i) \geq \Delta_{n-1}(i+1)$, για κάθε $i \in \{0, 1, \dots, M-1\}$.

Έστω το στάδιο απόφασης n και τυχόν διαδοχικοί δείκτες $i-1$, i και $i+1$ από το σύνολο $\{1,2,\dots,M\}$.

Τότε $a_n(i)=1$ είναι ισοδύναμο με

$$pR + \tau \sum_j p_{ij}^1 V_{n-1}(j) > \tau \sum_j p_{ij}^2 V_{n-1}(j).$$

Ισοδύναμα έχουμε

$$pR + \tau p(1-q)V_{n-1}(i+1) + \tau(1-p)qV_{n-1}(i-1) + \tau[(1-p)(1-q) + pq]V_{n-1}(i) > \tau qV_{n-1}(i-1) + \tau(1-q)V_{n-1}(i)$$

Ισοδύναμα

$$p(1-q)[V_{n-1}(i+1) - V_{n-1}(i)] + pq[V_{n-1}(i) - V_{n-1}(i-1)] > -\frac{pR}{\tau}$$

δηλαδή

$$p(1-q)\Delta_{n-1}(i) + pq\Delta_{n-1}(i-1) > -\frac{pR}{\tau}. \quad (38)$$

Όμως από την επαγωγική υπόθεση έχουμε ότι $\Delta_{n-1}(i)$ είναι φθίνουσα για κάθε i όπου αυτή ορίζεται και επειδή οι συντελεστές στην παραπάνω ποσότητα είναι όλοι θετικοί το ίδιο ισχύει και για την $F_n(i) := p(1-q)\Delta_{n-1}(i) + pq\Delta_{n-1}(i-1)$. Τότε όμως υπάρχει

$k_n := \max[i : a_n(i)=1]$. Απομένει να δείξουμε ότι $\Delta_n(i)$ είναι φθίνουσα στο i όπου αυτή ορίζεται. Έχουμε τις εξής περιπτώσεις:

- 1) $a_n(i-1)=1$ $a_n(i)=1$ $a_n(i+1)=1$
- 2) $a_n(i-1)=1$ $a_n(i)=1$ $a_n(i+1)=2$

$$3) \quad a_n(i-1)=1 \quad a_n(i)=2 \quad a_n(i+1)=2$$

$$4) \quad a_n(i-1)=2 \quad a_n(i)=2 \quad a_n(i+1)=2$$

Οι οποίες είναι ισοδύναμες με:

$$1) \quad i+1 \leq k_n$$

$$2) \quad i \leq k_n \quad \text{και} \quad i+1 > k_n \quad (39)$$

$$3) \quad i-1 \leq k_n \quad \text{και} \quad i > k_n$$

$$4) \quad i-1 > k_n$$

Σε κάθε περίπτωση θα δείξουμε ότι $\Delta_n(i-1) \geq \Delta_n(i), \forall i$. Ισοδύναμα αρκεί να δείξουμε ότι $\Delta_n(i-1) - \Delta_n(i) \geq 0, \forall i$.

Όπου:

$$\begin{aligned} \Delta_n(i-1) &= V_n(i) - V_n(i-1) \\ \Delta_n(i) &= V_n(i+1) - V_n(i) \end{aligned}, \quad \forall n \quad (40)$$

Για την περίπτωση 1) της σχέσης (39), έχουμε από την (25) τα εξής:

$$\begin{aligned} V_n(i) &= -hi + pR + \tau p(1-q)V_{n-1}(i+1) + \tau(1-p)qV_{n-1}(i-1) \\ &\quad + \tau[(1-p)(1-q) + pq]V_{n-1}(i) \\ V_n(i-1) &= -h(i-1) + pR + \tau p(1-q)V_{n-1}(i) + \tau(1-p)qV_{n-1}(i-2) \\ &\quad + \tau[(1-p)(1-q) + pq]V_{n-1}(i-1) \\ V_n(i+1) &= -h(i+1) + pR + \tau p(1-q)V_{n-1}(i+2) + \tau(1-p)qV_{n-1}(i-1) \\ &\quad + \tau[(1-p)(1-q) + pq]V_{n-1}(i+1) \end{aligned}$$

Τότε από την (40) κάνοντας την αντικατάσταση έχουμε:

$$\begin{aligned}
 \Delta_n(i-1) &= -h + \tau p(1-q)\Delta_{n-1}(i) + \tau(1-p)q\Delta_{n-1}(i-2) \\
 &\quad + \tau[(1-p)(1-q) + pq]\Delta_{n-1}(i-1) \\
 \Delta_n(i) &= -h + \tau p(1-q)\Delta_{n-1}(i+1) + \tau(1-p)q\Delta_{n-1}(i-1) \\
 &\quad + \tau[(1-p)(1-q) + pq]\Delta_{n-1}(i) \\
 \Delta_n(i-1) - \Delta_n(i) &= \tau p(1-q)[\Delta_{n-1}(i) - \Delta_{n-1}(i+1)] \\
 &\quad + \tau(1-p)q[\Delta_{n-1}(i-2) - \Delta_{n-1}(i-1)] \\
 &\quad + \tau[(1-p)(1-q) + pq][\Delta_{n-1}(i-1) - \Delta_{n-1}(i)]
 \end{aligned}$$

Όμως από την επαγωγική υπόθεση έχουμε ότι $\Delta_{n-1}(i)$ είναι φθίνουσα στο i και επομένως όλοι οι όροι της τελευταίας είναι μη αρνητικοί άρα καταλήγουμε στο συμπέρασμα ότι.

$$\Delta_n(i-1) - \Delta_n(i) \geq 0 \tag{41}$$

Για την περίπτωση 2) της σχέσης (39), έχουμε από την (25) τα εξής:

$$\begin{aligned}
 V_n(i) &= -hi + pR + \tau p(1-q)V_{n-1}(i+1) + \tau(1-p)qV_{n-1}(i-1) \\
 &\quad + \tau[(1-p)(1-q) + pq]V_{n-1}(i) \\
 V_n(i-1) &= -h(i-1) + pR + \tau p(1-q)V_{n-1}(i) + \tau(1-p)qV_{n-1}(i-2) \\
 &\quad + \tau[(1-p)(1-q) + pq]V_{n-1}(i-1) \\
 V_n(i+1) &= -h(i+1) + \tau qV_{n-1}(i) + \tau(1-q)V_{n-1}(i+1)
 \end{aligned}$$

Τότε από την (40) κάνοντας την αντικατάσταση έχουμε:

$$\begin{aligned}
 \Delta_n(i-1) &= -h + \tau p(1-q)\Delta_{n-1}(i) + \tau(1-p)q\Delta_{n-1}(i-2) \\
 &\quad + \tau[(1-p)(1-q) + pq]\Delta_{n-1}(i-1) \\
 \Delta_n(i) &= -h - pR + \tau(1-q)\Delta_{n-1}(i) - \tau p(1-q)\Delta_{n-1}(i) \\
 &\quad + \tau q\Delta_{n-1}(i) - \tau p q\Delta_{n-1}(i-1) \\
 &= -h - pR + \tau(1-p)(1-q)\Delta_{n-1}(i) + \tau q(1-p)\Delta_{n-1}(i-1) \\
 \Delta_n(i-1) - \Delta_n(i) &= +\tau(1-p)(1-q)[\Delta_{n-1}(i-1) - \Delta_{n-1}(i)] \\
 &\quad + \tau(1-p)q[\Delta_{n-1}(i-2) - \Delta_{n-1}(i-1)] \\
 &\quad + \tau p[(1-q)\Delta_{n-1}(i) + q\Delta_{n-1}(i-1)] + pR
 \end{aligned}$$

Όμως από την μονοτονία της $\Delta_{n-1}(i)$ λόγω της επαγωγικής υπόθεσης και από την σχέση (38) όλες οι παραπάνω διαφορές και οι συντελεστές τους είναι μη αρνητικές άρα

$$\Delta_n(i-1) - \Delta_n(i) \geq 0 \quad (42)$$

Για την περίπτωση 3) της σχέσης (39), έχουμε από την (25) τα εξής:

$$\begin{aligned}
 V_n(i) &= -hi + \tau q V_{n-1}(i-1) + \tau(1-q)V_{n-1}(i) \\
 V_n(i-1) &= -h(i-1) + pR + \tau p(1-q)V_{n-1}(i) + \tau(1-p)qV_{n-1}(i-2) \\
 &\quad + \tau[(1-p)(1-q) + pq]V_{n-1}(i-1) \\
 V_n(i+1) &= -h(i+1) + \tau q V_{n-1}(i) + \tau(1-q)V_{n-1}(i+1)
 \end{aligned}$$

Τότε από την (40) κάνοντας την αντικατάσταση έχουμε:

$$\begin{aligned}
 \Delta_n(i-1) &= -h - pR + \tau(1-q)\Delta_{n-1}(i-1) - \tau p(1-q)\Delta_{n-1}(i-1) \\
 &\quad + \tau q(1-p)\Delta_{n-1}(i-2) \\
 \Delta_n(i) &= -h + \tau q\Delta_{n-1}(i-1) + \tau(1-q)\Delta_{n-1}(i) \\
 \Delta_n(i-1) - \Delta_n(i) &= +\tau(1-p)q[\Delta_{n-1}(i-2) - \Delta_{n-1}(i-1)] \\
 &\quad + \tau(1-q)[\Delta_{n-1}(i-1) - \Delta_{n-1}(i)] \\
 &\quad - \tau p(1-q)\Delta_{n-1}(i) - \tau p q\Delta_{n-1}(i-1) - pR
 \end{aligned}$$

Όμως αφού $i > k_n$ από την (38) έχουμε:

$$p(1-q)\Delta_{n-1}(i) + pq\Delta_{n-1}(i-1) \leq -\frac{pR}{\tau}. \quad (43)$$

Από την επαγωγική υπόθεση έχουμε ότι η $\Delta_{n-1}(i)$ είναι φθίνουσα, επομένως οι διαφορές στο δεξιό μέλος της $\Delta_n(i-1) - \Delta_n(i)$ είναι θετικές και έχουν θετικούς συντελεστές. Άρα λαμβάνοντας υπόψη και την (43) έχουμε ότι:

$$\Delta_n(i-1) - \Delta_n(i) \geq 0. \quad (44)$$

Για την περίπτωση 4) της σχέσης (39), έχουμε από την (25) τα εξής:

$$\begin{aligned}
 V_n(i) &= -hi + \tau q V_{n-1}(i-1) + \tau(1-q)V_{n-1}(i) \\
 V_n(i-1) &= -h(i-1) + \tau q V_{n-1}(i-2) + \tau(1-q)V_{n-1}(i-1) \\
 V_n(i+1) &= -h(i+1) + \tau q V_{n-1}(i) + \tau(1-q)V_{n-1}(i+1)
 \end{aligned}$$

Τότε από την (40) κάνοντας την αντικατάσταση έχουμε:

$$\begin{aligned} \Delta_n(i-1) &= -h + \tau q \Delta_{n-1}(i-2) + \tau(1-q) \Delta_{n-1}(i-1) \\ \Delta_n(i) &= -h + \tau q \Delta_{n-1}(i-1) + \tau(1-q) \Delta_{n-1}(i) \\ \Delta_n(i-1) - \Delta_n(i) &= +\tau q [\Delta_{n-1}(i-2) - \Delta_{n-1}(i-1)] + \tau(1-q) [\Delta_{n-1}(i-1) - \Delta_{n-1}(i)] \end{aligned}$$

Όμως από την μονοτονία της $\Delta_{n-1}(i)$ λόγω της επαγωγικής υπόθεσης όλες οι παραπάνω διαφορές είναι μη ακριτικές και οι συντελεστές τους είναι μη αρνητικοί άρα

$$\Delta_n(i-1) - \Delta_n(i) \geq 0. \quad (45)$$

Επομένως δείξαμε ότι για κάθε $i \in \{0, 1, \dots, M\}$ υπάρχει πολιτική κατωφλίου που ορίζεται από το $k_n = \max\{i : a_n(i) = 1\}$. Αν βρισκόμαστε στο στάδιο απόφασης n , και η κατάσταση του συστήματος είναι $j \leq k_n$ τότε η βέλτιστη απόφαση είναι $a_n(j) = 1$. Διαφορετικά αν $j > k_n$, τότε $a_n(j) = 2$. Αποδείξαμε ότι $\Delta_n(i)$ είναι φθίνουσα για κάθε $i \in \{1, \dots, M\}$, για να ολοκληρωθεί η απόδειξη πρέπει να αποδείξουμε ότι ισχύει και $\Delta_n(0) - \Delta_n(1) \geq 0$.

Εξετάζουμε αναλυτικά ανάλογα το κατώφλι και έχουμε

- 1) $k_n \geq 2$, τότε $a_n(0) = 1$, $a_n(1) = 1$, $a_n(2) = 1$
- 2) $k_n = 1$, τότε $a_n(0) = 1$, $a_n(1) = 1$, $a_n(2) = 2$
- 3) $k_n = 0$, τότε $a_n(0) = 1$, $a_n(1) = 2$, $a_n(2) = 2$
- 4) $k_n = -1$, τότε $a_n(0) = 2$, $a_n(1) = 2$, $a_n(2) = 2$

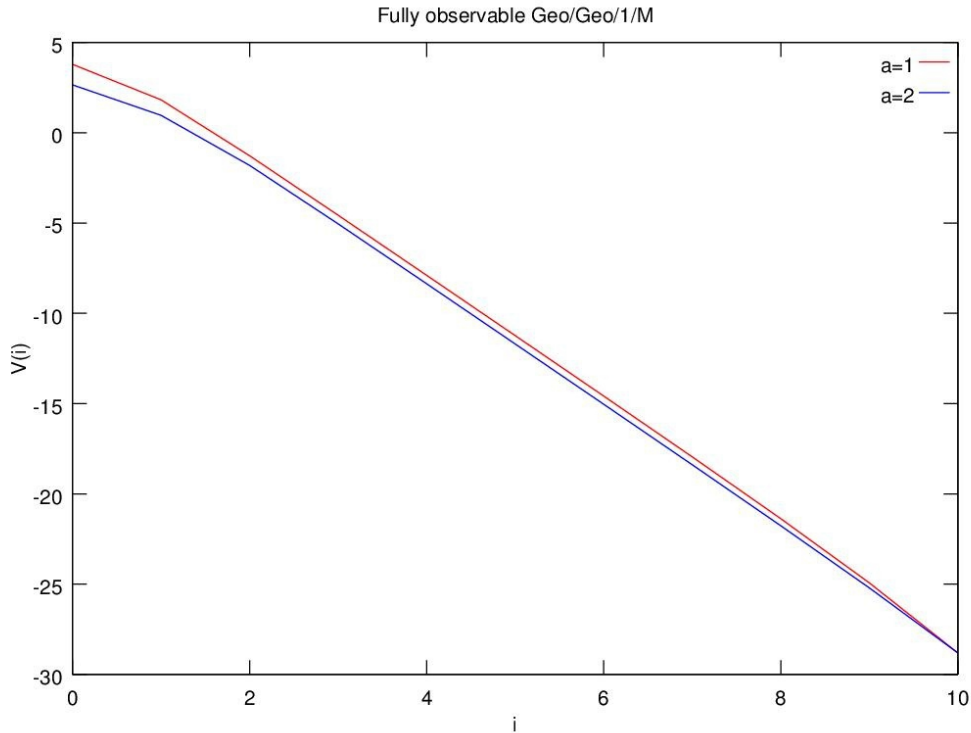
Σε κάθε μια από τις παραπάνω περιπτώσεις η απόδειξη ότι $\Delta_n(0) - \Delta_n(1) \geq 0$, είναι εντελώς ανάλογη με την απόδειξη στις περιπτώσεις 1-4 της (39) και εδώ ολοκληρώνεται και η απόδειξη ότι η $\Delta_n(i)$ είναι φθίνουσα σε όλο το πεδίο ορισμού της. ■

Στην πράξη για μεγάλο n θα υπολογίσουμε το βέλτιστο αναμενόμενο κέρδος $V_n(i)$ για το πρόβλημα πεπερασμένου ορίζοντα. Δηλαδή θα εφαρμόσουμε την μέθοδο των διαδοχικών προσεγγίσεων (Successive approximations) [2]. Λόγω της ομοιόμορφης σύγκλισης της ακολουθίας $V_n(i)$ στην $V(i)$, θα θέσουμε για n μεγάλο $V_n(i) \approx V(i)$, $\forall i$. Από την (23) θα εκτελέσουμε για κάθε κατάσταση i την μεγιστοποίηση ώστε τελικά να υπολογίσουμε το κατώφλι k . Τότε γράφουμε την βέλτιστη πολιτική και έχουμε ολοκληρώσει την λύση του προβλήματος με πλήρη πληροφόρηση.

Πιο συγκεκριμένα έστω το παρακάτω αριθμητικό παράδειγμα με τις ακόλουθες παραμέτρους: $\tau=0.7$, $\rho=0.7$, $q=0.5$, $R=3$, $h=1$, $M=10$, $n=100$. Από τους υπολογισμούς με το λογισμικό Octave (ver: 3.6.4) προκύπτει ότι η τιμή κατωφλίου είναι $k=10$ και η βέλτιστη αναμενόμενη ωφέλεια δίνεται από τον παρακάτω πίνακα:

i	0	1	2	3	4	5	6	7	8	9	10
$V_n(i)$	3,79	1,83	-1,27	-4,56	-7,89	-11,23	-14,57	-17,94	-21,37	-24,92	-28,8

Σχήμα 2. Αναμενόμενη ωφέλεια για κάθε κατάσταση /απόφαση.



Στην εικόνα απεικονίζεται για κάθε επιλογή $a=1,2$ η αντίστοιχη αναμενόμενη ωφέλεια για κάθε κατάσταση. Παρατηρούμε από το γράφημα ότι σε αυτή τη περίπτωση η βέλτιστη απόφαση είναι $a=1$ για κάθε αρχική κατάσταση.

Άρα η βέλτιστη στάσιμη πολιτική είναι: Για κάθε κατάσταση $i \leq k$ επιλέγουμε $a=1$ διαφορετικά αν $i > k$ επιλέγουμε $a=2$. Εδώ όμως $k=10$ επομένως αποφασίζω βέλτιστα κάθε φορά με $a=1$.

6.1.3 Το πρόβλημα με μερική πληροφόρηση

Στην περίπτωση όπου έχουμε ελλιπή πληροφόρηση για την πραγματική κατάσταση του συστήματος υποθέσαμε ότι έχουμε κάποιο σήμα θ το οποίο μας ενημερώνει για την κατάστα-

ση μετά την λήψη της απόφασης a . Όπως αναφερθήκαμε και στην εισαγωγή σε αυτή την περίπτωση ως κατάσταση του συστήματος θεωρούμε την κατανομή πιθανότητας πάνω στο χώρο των καταστάσεων. Σε αυτή την περίπτωση όπως έχουμε δει, ως κατάσταση θεωρούμε την κατανομή του αριθμού πελατών στο σύστημα.

Έστω ότι στην αρχή του σταδίου n ξέρουμε ή έχουμε εκτιμήσει ότι η κατάσταση είναι π , τότε από την (11) επιλέγουμε κατάλληλη απόφαση a που μεγιστοποιεί το δεξιό μέλος. Μετά την επιλογή της βέλτιστης απόφασης έχουμε μετάβαση στην κατάσταση j με πιθανότητα p_{ij}^a και παρατηρούμε σήμα θ . Τότε από την (3) η νέα κατάσταση του συστήματος θα είναι π' . Αν γνωρίζουμε την συνάρτηση βελτιστοποίησης για κάθε στάδιο τότε πορευόμαστε βέλτιστα με αυτό το τρόπο και άρα έχουμε μια βέλτιστη πολιτική [2].

Από τον αλγόριθμο του Sondik [3], τις αναφορές του Lovejoy [4] πάνω σε διορθώσεις και επιπλέον αλγορίθμους αλλά και τον αλγόριθμο Perseus [18] μπορούμε να υπολογίσουμε την συνάρτηση βελτιστοποίησης για πεπερασμένο ορίζοντα για οποιοδήποτε στάδιο και με επαναλαμβανόμενες προσεγγίσεις [2] μπορούμε να υπολογίσουμε την αντικειμενική συνάρτηση με εκπτώτικο παράγοντα $\tau \in (0,1)$ και για την περίπτωση άπειρου ορίζοντα.

Επομένως υπολογίζουμε για κάθε στάδιο n την αντικειμενική συνάρτηση $V_n(\pi)$, $\forall \pi$ από τον εκάστοτε αλγόριθμο, όπου για μεγάλο n θεωρούμε ότι $V_n(\pi) \approx V(\pi)$, $\forall \pi$. Από την (8) έχουμε ότι $V_n(\pi) = \max_j [\pi \cdot u^j(n)] = \pi \cdot u^s(n)$, επομένως για την κατάσταση π το βέλτιστο διάνυσμα είναι $u^s(n)$ με βέλτιστη απόφαση

a^s που προέκυψε μαζί με το διάνυσμα από την μεγιστοποίηση του δεξιού μέλους στην (11) κατά την εύρεση της αντικειμενικής συνάρτησης, δηλαδή $V_n(\pi) = \pi \cdot u^s(n)$

$$= \left[\sum_i \pi_i [q_i^{a^s} + \tau \sum_{j,\theta} p_{ij}^{a^s} r_{j\theta}^{a^s} u_j^{l(\pi, a^s, \theta)}(n-1)] \right] = \max_a \left[\sum_i \pi_i [q_i^a + \tau \sum_{j,\theta} p_{ij}^a r_{j\theta}^a u_j^{l(\pi, a, \theta)}(n-1)] \right].$$

Επομένως όταν έχουμε ολοκληρώσει πλέον τον υπολογισμό της συνάρτησης βελτιστοποίησης $V = \{u^1(n), u^2(n), \dots, u^j(n), \dots\}$, όπου $u^j(n)$ είναι το βέλτιστο διάνυσμα από την (11) με βέλτιστη απόφαση a^j , τότε για τον καθορισμό της βέλτιστης πολιτικής έχουμε ότι: για κάθε κατάσταση π η βέλτιστη απόφαση a ισούται με a^s που αντιστοιχεί από την μεγιστοποίηση της (8) στο βέλτιστο διάνυσμα $u^s(n)$.

6.2 Αριθμητικά αποτελέσματα από εφαρμογή μεθόδου Perseus

Ακολουθώντας το προηγούμενο παράδειγμα στην παράγραφο 6.1.2 σελίδα 58 με εφαρμογή του αλγορίθμου Perseus [18] έχουμε τα παρακάτω συμπεράσματα.

Για κατώφλι σφάλματος $e=10^{-10}$ και δείγμα μεγέθους $N=10^5$ ομοιόμορφα κατανεμημένο πάνω στο χώρο καταστάσεων, με την μέθοδο διαδοχικών προσεγγίσεων εφαρμοσμένη κάθε φορά στην συνάρτηση βελτιστοποίησης $V_n(\pi)$ για το πεπερασμένο ορίζοντα που υπολογίζουμε κάθε φορά μέσω Perseus, έχουμε ότι προκύπτει σύγκλιση μετά από $n=103$ προσεγγίσεις. Για τους υπολογισμούς αξιοποιήθηκε τυχαίο, ομοιόμορφα κατανεμημένο δείγμα B , από το χώρο καταστάσεων $\Delta=\{\pi:=(\pi_0, \pi_1, \dots, \pi_M): \pi \cdot \mathbf{1}=1, \pi \geq 0\}$.

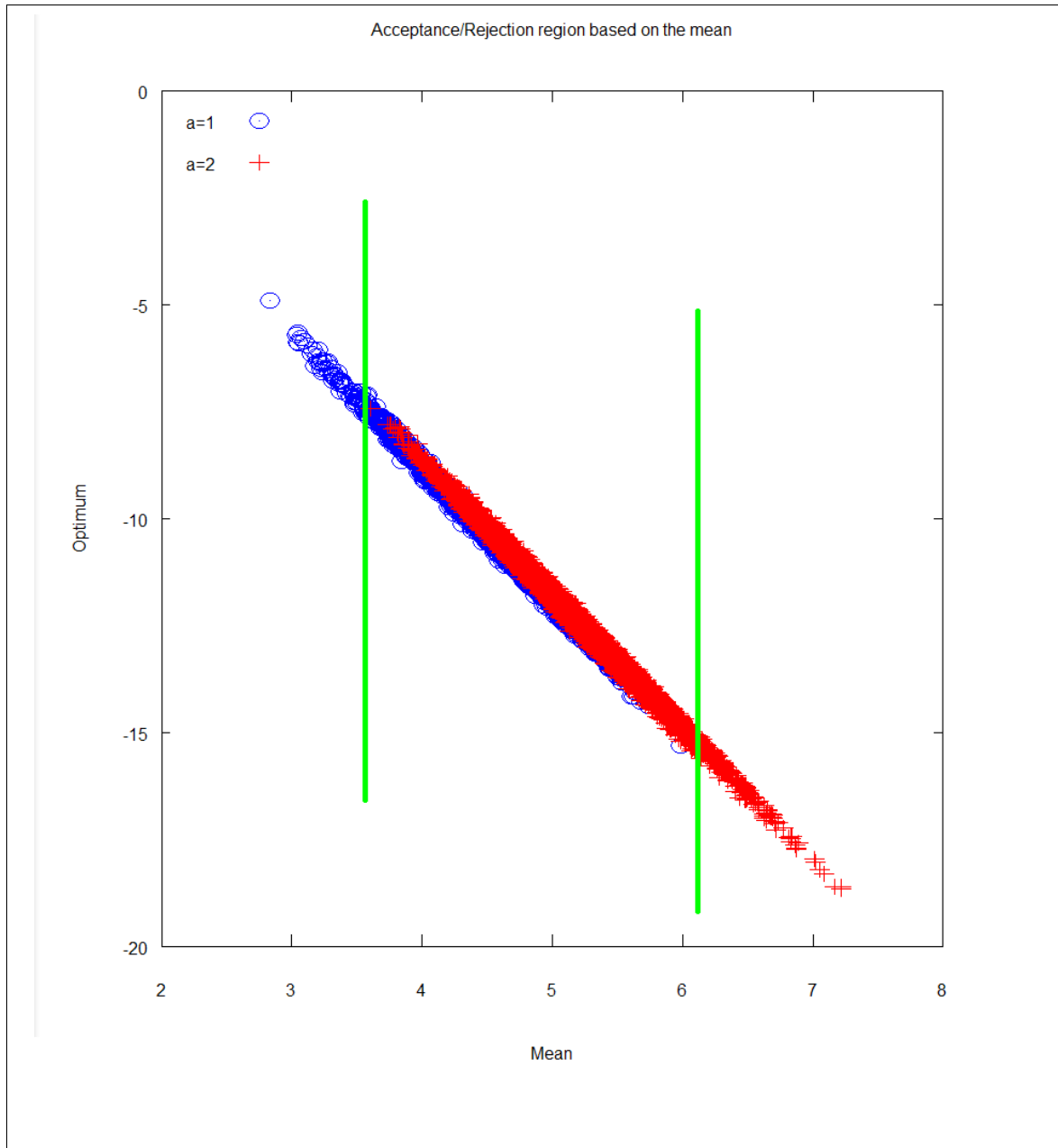
$$V = \begin{matrix} & & & s \\ & & & \begin{pmatrix} 2.14329 & 1.65486 & 2.30100 & 2.36408 \\ 0.82561 & 0.50374 & 0.95434 & 1.00896 \\ -1.97083 & -2.03645 & -1.93179 & -1.91793 \\ -5.01506 & -4.94270 & -5.06943 & -5.11000 \\ -8.19273 & -8.04607 & -8.29738 & -8.37200 \\ -11.44224 & -11.25558 & -11.57398 & -11.66692 \\ -14.73044 & -14.52223 & -14.87676 & -14.97957 \\ -18.03947 & -17.81966 & -18.19364 & -18.30177 \\ -21.35971 & -21.13367 & -21.51811 & -21.61523 \\ -24.68600 & -24.45659 & -24.79003 & -24.84307 \\ -27.78432 & -27.78432 & -27.86461 & -27.90102 \end{pmatrix} \\ j & & & \end{matrix}, \quad a = \begin{pmatrix} 1 \\ 2 \\ 1 \\ 1 \end{pmatrix}$$

Η κάθε στήλη του πίνακα V αποτελεί και ένα από τα βέλτιστα διανύσματα $u^s(n)$ για κάθε s με την αντίστοιχη βέλτιστη απόφαση a^s από το διπλανό πίνακα.

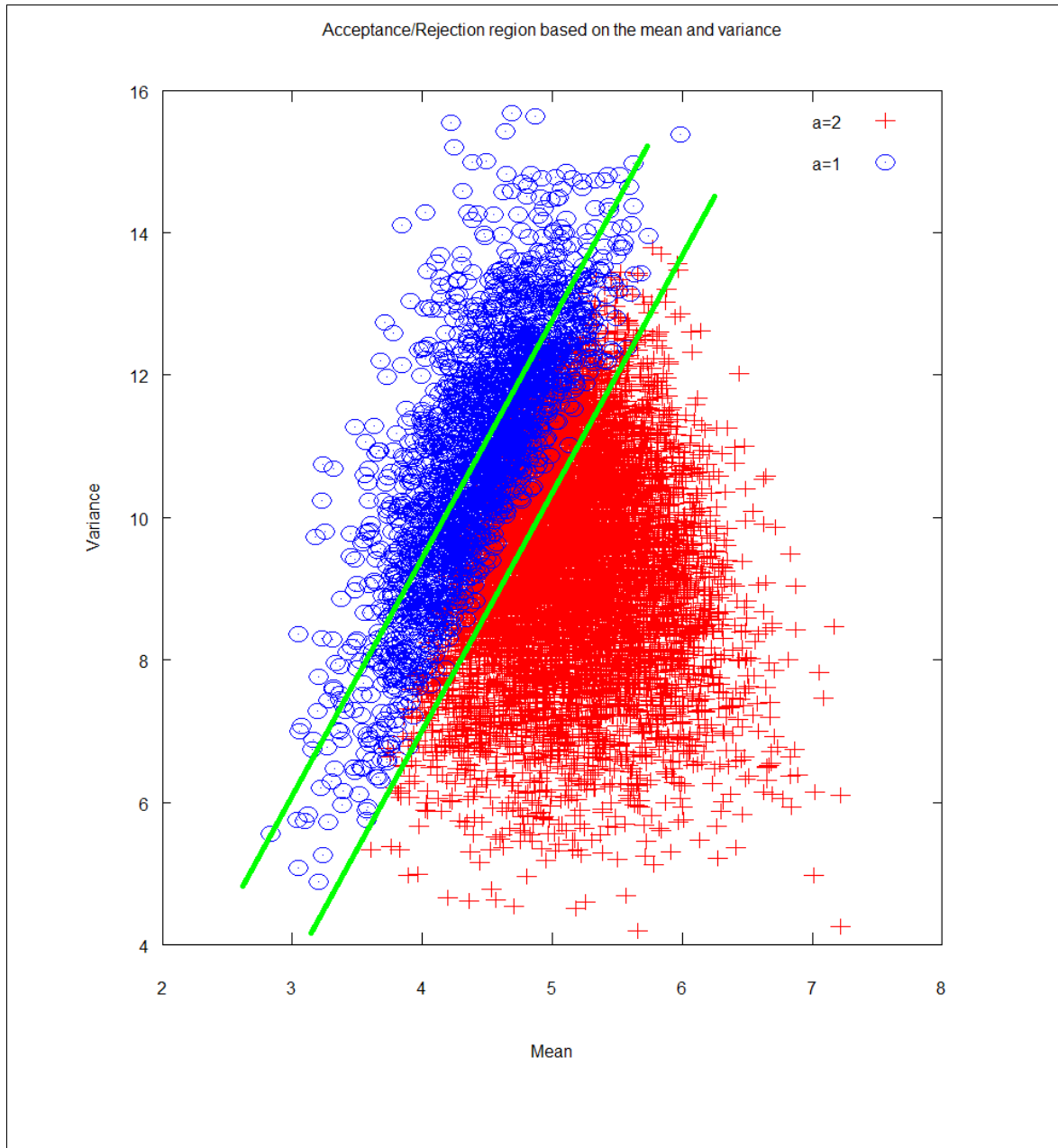
Για κάθε σημείο $\pi \in B$, εκτελούμε την μεγιστοποίηση από την (8) και αποθηκεύουμε σε μια λίστα το σημείο, την μέση του τιμή, διασπορά, μέγιστη αναμενόμενη ωφέλεια και βέλτιστη απόφαση. Με αυτά τα αριθμητικά αποτελέσματα δημιουργούμε διαγράμματα για να αποκτήσουμε μια γεωμετρική άποψη των περιοχών αποδοχής /απόρριψης.

Η ιδέα είναι να απλοποιήσουμε την επιλογή της βέλτιστης πολιτικής και να την μετατρέψουμε σε έναν απλό κανόνα αποδοχής /απόρριψης αξιοποιώντας ως πληροφορία μόνο την μέση τιμή της κατανομής και την αντίστοιχη διασπορά. Έχουμε τα παρακάτω γραφήματα:

Γράφημα 1: Αποδοχή-Απόρριψη με βάση μόνο την μέση τιμή μ .



Γράφημα2:Αποδοχή-Απόρριψη με βάση την μέση τιμή μ , και την διασπορά σ^2 .



Για το γράφημα 1, στον οριζόντιο άξονα έχουμε τις μέσες τιμές ενώ στο κατακόρυφο την μέγιστη αναμενόμενη συνολική ωφέλεια. Παρατηρούμε ότι κοντά στα άκρα του διαστήματος της μέσης τιμής έχουμε πλήρη διαχώριση των αποφάσεων. Εδώ προκύπτει ότι για $\mu \leq 3.6$ έχουμε $a=1$ ενώ για $\mu \geq 6.1$ ισχύει $a=2$. Όμως δεν μπορούμε να μιλήσουμε με βεβαιότητα για τα ενδιάμεσα σημεία.

Για το γράφημα 2, Στον οριζόντιο άξονα έχουμε την μέση τιμή ενώ στον κατακόρυφο την αντίστοιχη διασπορά. Εδώ παρατηρούμε ότι σε αντίθεση με το προηγούμενο γράφημα μπορούμε να μιλήσουμε για όλο το εύρος τιμών για την μέση τιμή. Καθώς η μέση τιμή αυξάνει και η διασπορά ελαττώνεται παρατηρούμε ότι ισχύει $a=2$, ενώ καθώς η μέση τιμή αυξάνει και η διασπορά αυξάνει ισχύει $a=1$. Συνεχίζει και υφίσταται το πρόβλημα της πλήρους διαχώρισης για μια λωρίδα σημείων που έχουν μέση τιμή και διασπορά εντός του χωρίου που περιγράφουν οι δύο ευθείες στο παραπάνω γράφημα. Ακριβέστερα έστω ότι

$\varepsilon_1: \alpha_1 \mu + \beta_1 \sigma^2 = \gamma_1$ και $\varepsilon_2: \alpha_2 \mu + \beta_2 \sigma^2 = \gamma_2$, τότε αν $\pi: \alpha_1 \mu_\pi + \beta_1 \sigma_\pi^2 \geq \gamma_1$ ισχύει $a=1$ ενώ για $\pi: \alpha_2 \mu_\pi + \beta_2 \sigma_\pi^2 \leq \gamma_2$ ισχύει $a=2$. Δεν μπορούμε να πούμε κάτι για την περίπτωση όπου για το π ισχύουν και οι δύο ανισότητες με την αντίθετη φορά, δηλαδή αν το π βρίσκεται εντός του χωρίου που περιγράφουν οι ε_1 και ε_2 .

6.2.1 Διακρίνουσα ανάλυση πάνω στα αποτελέσματα

Από το γράφημα οδηγούμαστε στο συμπέρασμα ότι θα μπορούσε να εφαρμοστεί διακρίνουσα ανάλυση στα δεδομένα μας και να φτιαχτεί έτσι ένας απλός κανόνας απόφαση με χρήση της διακρίνουσας συνάρτησης του Fisher. Έχουμε επομένως τα εξής:

$$\text{Ομάδα 1, } X^{(1)} = \begin{pmatrix} \mu_1^{(1)} & \sigma_1^{2(1)} \\ \mu_2^{(1)} & \sigma_2^{2(1)} \\ \vdots & \vdots \\ \mu_{n_1}^{(1)} & \sigma_{n_1}^{2(1)} \end{pmatrix} = \begin{pmatrix} x_1^{(1)'} \\ x_1^{(2)'} \\ \vdots \\ x_{n_1}^{(1)'} \end{pmatrix}, \quad \bar{x}^{(1)} = (\bar{\mu}^{(1)}, \bar{\sigma}^{2(1)})',$$

$$\text{Ομάδα 2, } X^{(2)} = \begin{pmatrix} \mu_1^{(2)} & \sigma_1^{2(2)} \\ \mu_2^{(2)} & \sigma_2^{2(2)} \\ \vdots & \vdots \\ \mu_{n_2}^{(2)} & \sigma_{n_2}^{2(2)} \end{pmatrix} = \begin{pmatrix} x_1^{(1)'} \\ x_1^{(2)'} \\ \vdots \\ x_{n_2}^{(1)'} \end{pmatrix}, \quad \bar{x}^{(2)} = (\bar{\mu}^{(2)}, \bar{\sigma}^{2(2)})',$$

Όπου $x_i^{(a)} = (\mu_i^{(a)}, \sigma_i^{2(a)})'$, $\forall i=1,2,\dots,n_a$ είναι η i -οστή παρατήρηση της ομάδας a .

Ορίζουμε το κέντρο βάρους κάθε ομάδας a ως $\bar{x}^{(a)} = (\bar{\mu}^{(a)}, \bar{\sigma}^{2(a)})'$, $a=1,2$.

Επίσης ορίζουμε το κέντρο βάρους όλων των παρατηρήσεων:

$$\bar{x} = \frac{n_1 \bar{x}^{(1)} + n_2 \bar{x}^{(2)}}{N}, \quad N = n_1 + n_2.$$

Για τα κεντρικοποιημένα δεδομένα κάθε ομάδας έχουμε:

$$X_1^* = \begin{pmatrix} \mu_1^{(1)} - \bar{\mu}^{(1)} & \sigma_1^{2(1)} - \bar{\sigma}^{2(1)} \\ \mu_2^{(1)} - \bar{\mu}^{(1)} & \sigma_2^{2(1)} - \bar{\sigma}^{2(1)} \\ \vdots & \vdots \\ \mu_{n_1}^{(1)} - \bar{\mu}^{(1)} & \sigma_{n_1}^{2(1)} - \bar{\sigma}^{2(1)} \end{pmatrix} = \begin{pmatrix} x_1^{(1)'} - \bar{x}^{(1)} \\ x_1^{(1)'} - \bar{x}^{(1)} \\ \vdots \\ x_{n_1}^{(1)'} - \bar{x}^{(1)} \end{pmatrix}$$

$$X_2^* = \begin{pmatrix} \mu_1^{(2)} - \bar{\mu}^{(2)} & \sigma_1^{2(2)} - \bar{\sigma}^{2(2)} \\ \mu_2^{(2)} - \bar{\mu}^{(2)} & \sigma_2^{2(2)} - \bar{\sigma}^{2(2)} \\ \vdots & \vdots \\ \mu_{n_2}^{(2)} - \bar{\mu}^{(2)} & \sigma_{n_2}^{2(2)} - \bar{\sigma}^{2(2)} \end{pmatrix} = \begin{pmatrix} x_1^{(2)'} - \bar{x}^{(2)} \\ x_1^{(2)'} - \bar{x}^{(2)} \\ \vdots \\ x_{n_2}^{(2)'} - \bar{x}^{(2)} \end{pmatrix}$$

Τότε για κάθε ομάδα ορίζουμε τους πίνακες συνδιακύμανσης:

$$S_a = \frac{1}{n_a - 1} X_a^{*'} X_a^*, \quad a=1,2.$$

Ο πίνακας με όλα τα κεντρικοποιημένα δεδομένα του δείγματος είναι:

$$X^* = \begin{pmatrix} \mu_1^{(1)} - \bar{\mu} & \sigma_1^{2(1)} - \bar{\sigma}^2 \\ \mu_2^{(1)} - \bar{\mu} & \sigma_2^{2(1)} - \bar{\sigma}^2 \\ \vdots & \vdots \\ \mu_{n_1}^{(1)} - \bar{\mu} & \sigma_{n_1}^{2(1)} - \bar{\sigma}^2 \\ \vdots & \vdots \\ \mu_{n_2}^{(1)} - \bar{\mu} & \sigma_{n_2}^{2(1)} - \bar{\sigma}^2 \end{pmatrix} = \begin{pmatrix} x_1^{(1)'} - \bar{x} \\ x_1^{(1)'} - \bar{x} \\ \vdots \\ x_{n_1}^{(1)'} - \bar{x} \\ \vdots \\ x_{n_2}^{(2)'} - \bar{x} \end{pmatrix}$$

Ο κοινός πίνακας συνδιακύμανσης ορίζεται ως:

$$S = \frac{1}{N-2} X^{*'} X^*.$$

Πρόταση [20]: Ισχύει $T=W+B$, όπου

$T := X^{*'} X^*$, ο πίνακας τετραγώνων αποστάσεων από το συνολικό κέντρο βάρους.

$W := \sum_a X_a^{*'} X_a^*$, πίνακες τετραγώνων αποστάσεων από τα κέντρα βάρους κάθε ομάδας.

$B := \sum_a n_a (\bar{x}^{(a)} - \bar{x})(\bar{x}^{(a)} - \bar{x})'$, πίνακας τετραγώνων αποστάσεων του κέντρου βάρους

κάθε ομάδας από το συνολικό κέντρο βάρους.

Αν υπολογίσουμε τους B και W , και αν ο W είναι αντιστρέψιμος, τότε η διακρίνουσα συνάρτηση Fisher είναι $F = u' \cdot x$, όπου u είναι το ιδιοδιάνυσμα του $W^{-1}B$ που αντιστοιχεί στην μέγιστη ιδιοτιμή λ και υπάρχει κατάλληλη σταθερά c , τέτοια ώστε μια παρατήρηση θα ταξινομείται στην ομάδα 1 αν $u' \cdot x \leq c$. Διαφορετικά ανήκει στην ομάδα 2 αν ισχύει $u' \cdot x > c$.

Τρέχοντας τον αλγόριθμο στην Octave 3.6.4 έχουμε ότι η μεγαλύτερη ιδιοτιμή είναι $\lambda = 6.4986 \times 10^{-1}$ και το αντίστοιχο ιδιοδιάνυσμα $u = (0.96913, -0.24655)'$. Το διάνυσμα u χρησιμεύει ώστε να γίνει ο καλύτερος δυνατός διαχωρισμός ανάμεσα στις δυο ομάδες. Οι προβολές των δύο ομάδων πάνω στο διάνυσμα u πετυχαίνουν τον μέγιστο διαχωρισμό.

Η ταξινόμηση των παρατηρήσεων γίνεται ως εξής:

Παίρνουμε την προβολή $y = u' \cdot x$ της παρατήρησης $x = (\mu, \sigma^2)$, πάνω στο u . Η παρατήρηση πρέπει να ταξινομηθεί στην ομάδα από την οποία η προβολή της απέχει λιγότερο από την προβολή του αντίστοιχου κέντρου βάρους σε σχέση με όλες τις υπόλοιπες ομάδες.

Στην περίπτωση μας επειδή έχουμε μόνο δύο ομάδες τα πράγματα απλοποιούνται αρκετά και το γενικότερο κριτήριο που εξετάζει την απόσταση για κάθε ιδιοδιάνυσμα u_i , που αντιστοιχεί την ιδιοτιμή $\lambda_i > 0, \forall i = 1, \dots, r$ εδώ γίνεται [20]:

Έστω $d_a(x) = (u' \cdot x - u' \cdot \bar{x}^{(a)})^2$ να είναι η τετραγωνική απόσταση της προβολής του x από την προβολή του κέντρου βάρους της ομάδας $a = 1, 2$.

Κριτήριο: Η παρατήρηση x θα καταταχθεί στην ομάδα a αν και μόνο αν ισχύει

$$d_a(x) = \min[d_i(x), i = 1, 2].$$

Επομένως αν $x = (\mu, \sigma^2)$, είναι ένα νέο σημείο τότε κατατάσσεται στην ομάδα $a = 1$, αν $d_1(x) \leq d_2(x)$.

Ισοδύναμα έχουμε

$$(u' \cdot x - u' \cdot \bar{x}^{(1)})^2 \leq (u' \cdot x - u' \cdot \bar{x}^{(2)})^2.$$

Δηλαδή

$$(u' \cdot x)^2 - 2u' \cdot \bar{x}^{(1)} u' \cdot x + (u' \cdot \bar{x}^{(1)})^2 \leq (u' \cdot x)^2 - 2u' \cdot \bar{x}^{(2)} u' \cdot x + (u' \cdot \bar{x}^{(2)})^2.$$

Όμως $\bar{x}^{(1)}$ και $\bar{x}^{(2)}$ είναι γνωστές σταθερές και επομένως θέτουμε και

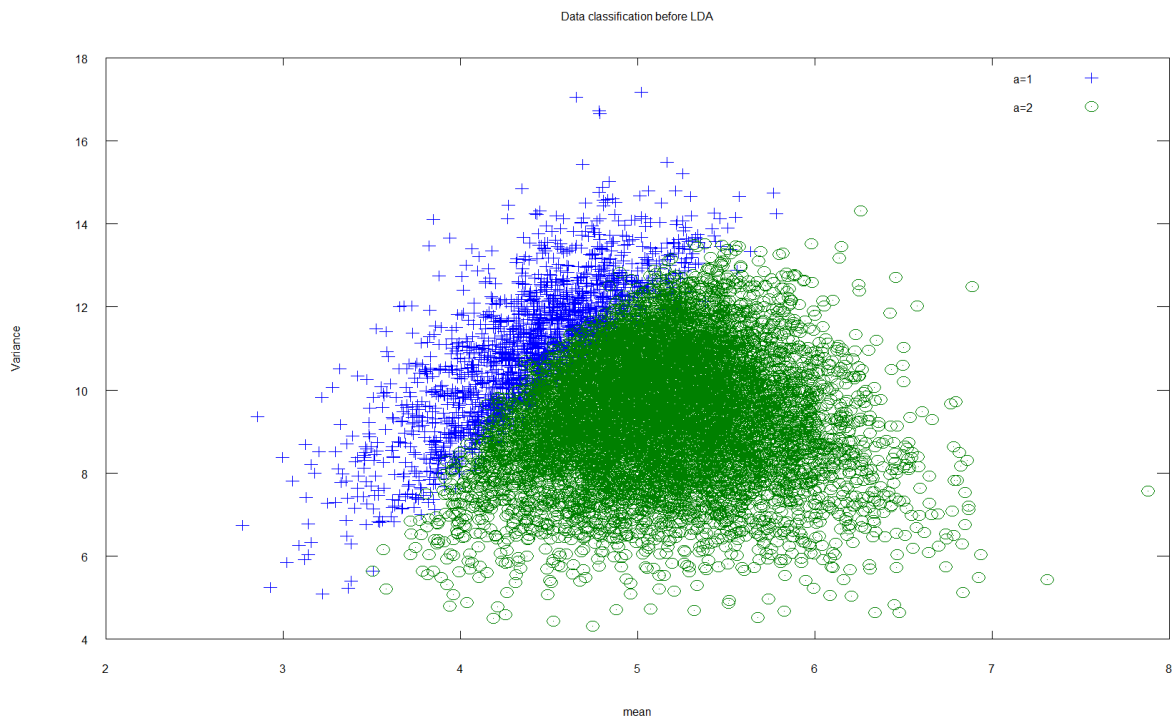
$c_1 = u' \cdot \bar{x}^{(2)}$. Τότε ισοδύναμα θα έχουμε ότι το x κατατάσσεται στην ομάδα $a = 1$, αν

$$(u' \cdot x)^2 - 2c_1 u' \cdot x + c_1^2 \leq (u' \cdot x)^2 - 2c_2 u' \cdot x + c_2^2.$$

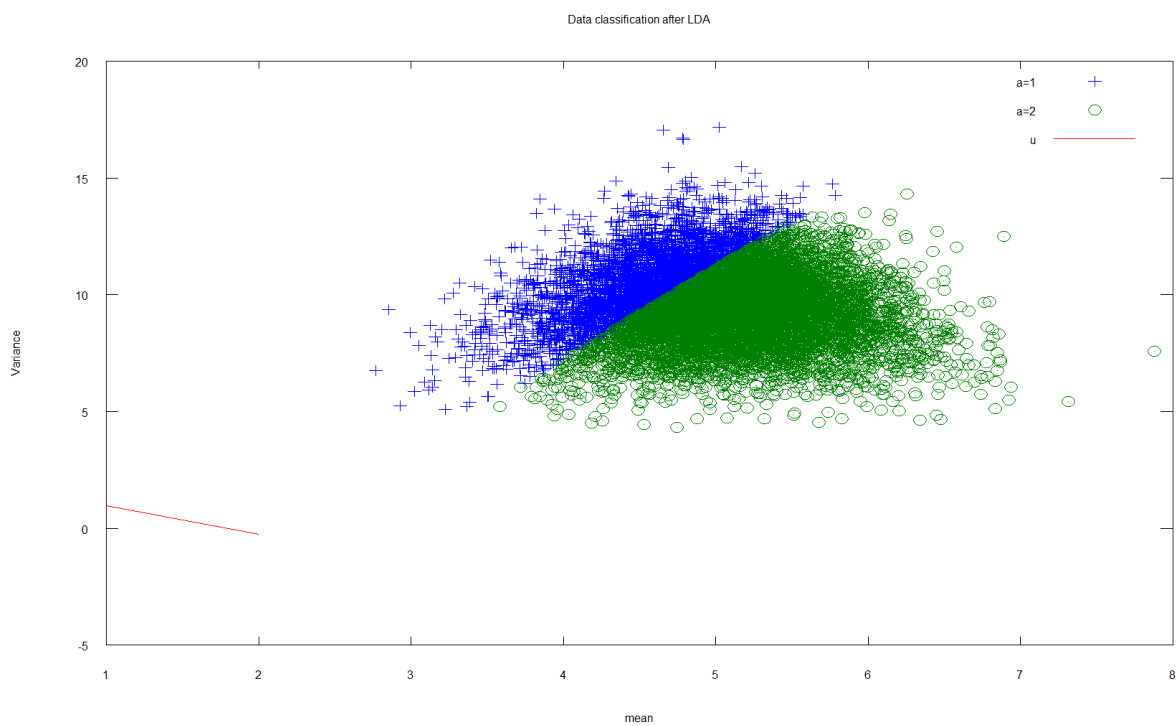
Ισοδύναμα

$$u' \cdot x \leq c, \quad c = \frac{c_1 + c_2}{2}. \quad (46)$$

Επομένως από τα αριθμητικά αποτελέσματα του προβλήματος προκύπτουν $\lambda = 6.4986 \times 10^{-1}$, $u = (0.96913, -0.24655)'$, $c = 2.1050$. Τότε έχουμε τα παρακάτω γραφήματα πριν την ταξινόμηση και μετά την ταξινόμηση των παρατηρήσεων σύμφωνα με την διακρίνουσα συνάρτηση του Fisher.



Στο παραπάνω γράφημα έχουμε την παρατήρηση των δεδομένων στις αντίστοιχες ομάδες όπου αυτές ανήκουν, χωρίς την αξιοποίηση της διακρίνουσας συνάρτησης του Fisher.



Σε αυτό το γράφημα παρουσιάζουμε την ταξινόμηση των δεδομένων σε δύο ομάδες με χρήση της μεθόδου διακρίνουσας ανάλυση του Fisher. Τα δεδομένα έχουν χωριστεί σε δύο κατηγορίες ανάλογα την ομάδα τους όπου με Ομάδα 1 έχουμε θέσει όλα εκείνα τα δειγματικά σημεία για τα οποία το $a=1$ είναι η βέλτιστη απόφαση στο πρόβλημά μας ενώ η Ομάδα 2 αποτελείται από εκείνα τα δειγματικά σημεία για τα οποία το $a=2$ αποτελεί την βέλτιστη απόφαση. Επίσης φαίνεται και το διάνυσμα u πάνω στο οποίο οι ορθές προβολές των σημείων μας δίνουν την καλύτερη διαχώρηση κατά Fisher.

6.3 Συμπεράσματα

Η θεωρία και τα μέχρι τώρα ευρήματα που αφορούν τις MOMΔΑ μπορούν να χρησιμοποιηθούν για την μοντελοποίηση οποιουδήποτε προβλήματος όπου ο πράκτορας καλείται να λάβει την βέλτιστη απόφαση σε ένα περιβάλλον με ελλιπή πληροφόρηση. Μπορούμε να φανταστούμε μια πληθώρα από τέτοια προβλήματα όπως για παράδειγμα την αντικατάσταση χαλασμένου εξαρτήματος ή μηχανής στο σύγγραμμα του Sodnik [3], διάφορα προβλήματα τεχνητής νοημοσύνης [18], αυτόματη εύρεση στόχου, αυτόματη εκμάθηση μιας εργασίας για συγκεκριμένη μηχανή, το παράδειγμα μας σε αυτό το σύγγραμμα με την ουρά εξυπηρέτησης αλλά και πολλά άλλα. Αν αναλογιστούμε ότι στην καθημερινότητα σχεδόν κάθε τι που κάνουμε, διαδραματίζεται σε ένα χώρο όπου δεν έχουμε πάντα πλήρη πληροφόρηση τότε αντιλαμβανόμαστε ότι όρια για το πεδίο εφαρμογών για τις MOMΔΑ βάζει μονάχα η εκάστοτε ανθρώπινη φαντασία.

Όμως παρόλο την καλή συμπεριφορά της αντικειμενικής συνάρτησης σε αυτά τα προβλήματα, που είναι δηλαδή κατά τμήματα γραμμική και κυρτή, ο ακριβής υπολογισμός βέλτιστης στάσιμης πολιτικής είναι γενικά εφικτός μονάχα για τα πιο απλά προβλήματα. Το υπολογιστικό κόστος στις μεθόδους ακριβείας αποτελεί το σημαντικότερο εμπόδιο για την ακριβή επίλυση αυτών των προβλημάτων. Όπως αναφέραμε και κατά διάρκεια του συγγράμματος έχουν αναπτυχθεί κατά τα τελευταία τριάντα χρόνια αρκετές προσεγγιστικές μέθοδοι που ελαφρύνουν κατά πολύ τον υπολογιστικό φόρτο αλλά υπάρχει πάντα κάποιο κόστος που θα πρέπει να πληρώσουμε. Οι μέθοδοι αυτές είναι εξειδικευμένες και η επιλογή για το ποια μέθοδος είναι η κατάλληλη κάθε φορά εξαρτάται κάθε φορά από την

φύση του προβλήματος. Επιπρόσθετα οι προσεγγιστικές μέθοδοι παρουσιάζουν μια ασυνέπεια όσο αφορά την τελική προσέγγιση της συνάρτησης ωφέλειας. Ενδέχεται πολλές φορές στα ενδιάμεσα στάδια να έχουμε περισσότερα α-διανύσματα από ότι στο τελικό στάδιο, πράγμα το οποίο καθιστά ανέφικτο τον υπολογισμό αξιόπιστων φραγμάτων για την προσέγγιση.

Στην πράξη,σε εφαρμογές στην βιομηχανία έχει παρατηρηθεί ότι οι προσεγγιστικές μέθοδοι μας δίνουν προβλέψεις καλύτερες από τους χειρότερους οιωνούς αλλά δεν υπάρχει αρκετό θεωρητικό πλαίσιο ακόμη ώστε τα αποτελέσματα αυτά να θεωρούνται έγκυρα.

Αναλογιζόμενοι όλες αυτές τις δυσκολίες αλλά και το γεγονός ότι με αυτά τα εργαλεία μπορούμε να μοντελοποιήσουμε ακόμη περισσότερες κατηγορίες προβλημάτων τις οποίες συναντάμε στην καθημερινότητά μας, συνυπολογίζοντας και το γεγονός ότι η τεχνολογική άνοδος είναι αναπόφευκτη και οι υπολογιστική ισχύ όλο και μεγαλώνει, μπορούμε να πούμε ότι το κομμάτι αυτό της επιχειρησιακής έρευνας θα αποτελεί ένα εξαιρετικά ενδιαφέρον πεδίο για περαιτέρω επιστημονική έρευνα.

Βιβλιογραφία

- [1]: Kulkarni, V.G. (1995), *Modeling and Analysis of Stochastic Systems*, Chapman & Hall/CRC.
- [2]: Ross, S.M. (1995), *Introduction to Stochastic Dynamic Programming*, Academic Press.
- [3]: Smallwood, R.D. and Sondik, E.J (1973), *The Optimal Control of Partially Observable Markov Processes over a Finite Horizon*, Operations Research, Vol. 21, No. 5 (Sep. - Oct., 1973), pp. 1071-1088.
- [4]: Lovejoy, W.S. (1991), *A survey of algorithmic methods for partially observed Markov decision processes*, Annals of Operations Research 28 (1991) 47-66.
- [5]: Cheng, H. (1988), *Algorithms for partially observed Markov decision processes*, Ph.D. Dissertation, Faculty of Commerce and Business Administration, University of British Columbia.
- [6]: Eagle, J. (1984), *The optimal search for a moving target when the search path is constrained*, Ph.D. Dissertation, Oper.Res. 32 (1984) 1107-1115..
- [7]: Monahan, G. (1982), *A survey of partially observable Markov decision processes*, Manag. Sci. 28(1982) 1-16.
- [8]: Sondik, E.J. (1978), *The optimal control of partially observable Markov processes over the infinite horizon: discounted case*, Oper. Res. 26 (1978) 282-304.
- [9]: Howard, R. (1971), *Dynamic Probabilistic Systems*, Wiley, New York.

- [10]: Platzman, L.K. (1977), *Optimal infinite-horizon undiscounted control of finite probabilistic systems*, Ph.D. Dissertation, SIAM. J. Control Opt. 18 (1980) 362-380.
- [11]: White, C.C. and Sherer, W. (1989), *Finite memory suboptimal design for partially observed Markov decision processes*, Technical report, Department of Systems Engineering, University of Virginia, Charlottesville, VA (1989), submitted to Oper. Res.
- [12]: Kakalik, J. (1965), *Optimum policies for partially observable Markov systems*, Tech. report 18, Operations Research Center, Massachusetts Institute of Technology, Cambridge, MA.
- [13]: Eckles, J. (1966), *Optimum replacement of stochastically failing systems*, Ph.D. Thesis, Department of Engineering-Economic Systems, Stanford University, Stanford CA.
- [14]: Sondik, E.J. (1971), *The optimal control of partially observable Markov processes*, Ph.D. Dissertation, Department of Electrical Engineering, Stanford University, Stanford, CA.
- [15]: Lovejoy, W.S. (1989), *Computationally feasible bounds for partially observed Markov decision processes*, Research paper 1024, Stanford University Graduate School of Business, Stanford, CA (1989), to appear in Oper. Res.
- [16]: Eaves, B.C. (1984), *A Course in Triangulations for Solving Equations with Deformations*, Springer, New York.
- [17]: Sondik, E.J. and Mendelsohn, R. (1979), *Information seeking in Markov decision processes*, Sondik, Southwest Fisheries Center Administrative Report H-79-13, Southwest Fisheries Center, National Marine Fisheries Service, NOAA, Honolulu, HI.

- [18]: Spaan, M.T.J. and Vlassis, N. (2005), *Perseus: Randomized Point-based Value Iteration for POMDPs*, Journal of Artificial Intelligence Research 24 195-220.
- [19]: Ross, S. (2006), *Simulation (4th edition)*, Academic Press.
- [20]: Μπουρνέτας, Α. (2006), *Στατιστική Ανάλυση Δεδομένων*, ΕΘΝΙΚΟ ΚΑΙ ΚΑΠΟΔΙΣΤΡΙΑΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ, ΤΜΗΜΑ ΜΑΘΗΜΑΤΙΚΩΝ.