ΚΕΦΑΛΑΙΟ 4

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

Περίληψη

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

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

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

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

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

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

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

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

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

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

$w \Xi B(P)$, $Hw=w$ ή $w=V^*$.

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

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

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

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

Θεωρούμε $V_0 \in B(\pi)$ και

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

Η παραπάνω σχέση γράφεται:

$$V_n = H_n V_0 \text{ , } n \ge 1$$

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

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

$$\left\| V_n - V^* \right\| £ \frac{b^n . L}{1 - b} \quad " \quad n \ge 1 \text{ ,} \qquad \textbf{4.1.2}$$

όπου $$L = \max_{i,a} \left| q(i,a) \right| .$$

Πράγματι,

$$V_1(p) = HV_0(p) = \max_a \{p.q^a\} £ \max_{i,a} \left| q(i,a) \right| = L \text{ , } \quad " p \Xi P$$

Επομένως $\|V_1\| \leq L$.

Εφαρμόζοντας το λήμμα 4.1.1 (iii) παίρνουμε**:**

$$\|V_n - V^*\| = \|H_n V_0 - V^*\| \leq \frac{\beta^n}{1-\beta} \| HV_0 - V_0 \| = \frac{\beta^n}{1-\beta} \|V_1\| \leq \frac{b^n}{1-b} L \text{ , } n \geq 1$$

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

$$\|V_n - V^*\| \leq h$$

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

$$\frac{b^n}{1-b} L \leq h$$

δηλ. $$n \geq \frac{\ln(\frac{(1-b).h}{L})}{\ln b}.$$

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

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

$$\|\tilde{V}_n - V^*\| \leq \frac{1-b^n}{1-b}.e + \frac{b^n.L}{1-b} \qquad \textbf{4.1.3}$$

Απόδειξη

Από την τριγωνική ιδιότητα της νόρμας του supremum παίρνουμε:

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

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

Επομένως

$$\left\| \tilde{V}_n - V_n \right\| \leq s_n = \frac{1 - b^n}{1 - b} . e$$

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

W

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

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

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

$$\frac{1 - b^n}{1 - b} . e + \frac{b^n . L}{1 - b} \leq h$$

Μπορούμε π.χ να επιλέξουμε τα ε , η έτσι ώστε: $\frac{e}{1 - b} \leq \frac{h}{2}$, $\frac{b^n . L}{1 - b} \leq \frac{h}{2}$,

δηλαδή: $e \leq \frac{(1 - b).h}{2}$, $n \geq \frac{\ln(\frac{(1 - b).h}{2L})}{\ln b}$

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

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

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

$$H_d V(./d) = V(./d).$$

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

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

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

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

$$\left\| V(./d) - V^* \right\| £ h$$

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

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

(όπου V_0=0, μηδενική συνάρτηση)

δηλαδή : $\delta^n(\pi)=\arg\max_a\{p.q^a + b.\varepsilon_q\{q/p,a\}.V_{n-1}(T(p,q,a))\}, p \Xi P$.

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

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

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

όπου $L = \max_{i,a}|q(i,a)|$. ***Bertsekas*** [10]

□

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

$$\frac{2.\beta^n}{1-\beta}.\Lambda \text{ £ } h$$

δηλαδή : $n^3 \frac{\ln(\frac{(1-b).h}{2L})}{\ln b}$

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

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

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

(βλέπε επίσης παρατήρηση 2 στην ενότητα 3.2).

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

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

$$f(e,n) := \frac{1+b-2b^n}{(1-b)^2}.e + \frac{2b^n.L}{1-b}\ , \qquad \textbf{4.1.6}$$

δηλαδή**:**

$$\left\| V(./\tilde{d}^n) - V^* \right\| \le f(e,n).$$

Απόδειξη

Από την τριγωνική ιδιότητα της νόρμας του supremum προκύπτει ότι:

$$\left\| V(./\tilde{d}^n) - V^* \right\| \pounds \left\| V(./\tilde{d}^n) - \tilde{V}_n \right\| + \left\| \tilde{V}_n - V^* \right\|. \qquad \textbf{4.1.7}$$

Εφαρμόζοντας πάλι την τριγωνική ιδιότητα παίρνουμε

$$\left\| V(./\tilde{d}^n) - \tilde{V}_n \right\| \pounds \left\| V(./\tilde{d}^n) - H_{\tilde{d}^n} \tilde{V}_n \right\| + \left\| H_{\tilde{d}^n} \tilde{V}_n - \tilde{V}_n \right\|. \qquad \textbf{4.1.8}$$

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

$$\left\| V(./\tilde{d}^n) - H_{\tilde{d}^n} \tilde{V}_n \right\| = \left\| H_{\tilde{d}^n} V(./\tilde{d}^n) - H_{\tilde{d}^n} \tilde{V}_n \right\| \pounds \beta . \left\| V(./\tilde{d}^n) - \tilde{V}_n \right\|. \qquad \textbf{4.1.9}$$

Από τις (4.1.8) και (4.1.9) παίρνουμε

: $$\left\| V(./\tilde{d}^n) - \tilde{V}_n \right\| \pounds \frac{1}{1- b} . \left\| H_{\tilde{d}^n} \tilde{V}_n - \tilde{V}_n \right\|. \qquad \textbf{4.1.10}$$

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

$$\left\| H_{\tilde{d}^n} \tilde{V}_n - \tilde{V}_n \right\| = \left\| H_{\tilde{d}^n} \tilde{V}_n - H_{\tilde{d}^n} \tilde{V}_{n-1} \right\| \pounds \beta . \left\| \tilde{V}_n - \tilde{V}_{n-1} \right\|. \qquad \textbf{4.1.11}$$

Από την τριγωνική ιδιότητα της νόρμας supremum έχουμε:

$$\left\| \tilde{V}_n - \tilde{V}_{n-1} \right\| \pounds \left\| \tilde{V}_n - V_n \right\| + \left\| V_n - V_{n-1} \right\| + \left\| V_{n-1} - \tilde{V}_{n-1} \right\| \pounds s_n + s_{n-1} + \left\| V_n - V_{n-1} \right\|, \textbf{4.1.12}$$

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

$$\left\| V_n - V_{n-1} \right\| = \left\| H_{n-1} V_1 - H_{n-1} V_0 \right\| \pounds \beta^{n-1} . \left\| V_1 - V_0 \right\| \pounds \beta^{n-1} . \Lambda . \qquad \textbf{4.1.13}$$

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

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

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

$$\|V(./\tilde{d}^n) - \tilde{V}_n\| \leq \frac{b.(s_n + s_{n-1})}{1-b} + \frac{b^n}{1-b}.\Lambda = \frac{b.(2-b^n-b^{n-1})}{(1-b)^2}.e + \frac{b^n}{1-b}.\Lambda . \qquad \underline{\mathbf{4.1.14}}$$

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

$$\|V(./\tilde{d}^n) - V^*\| \leq \frac{b.(2-b^n-b^{n-1})}{(1-b)^2}.e + \frac{1-b^n}{1-b}.e + \frac{2b^n}{1-b}.\Lambda =$$

$$= \frac{2b^n}{1-b}.\Lambda + \frac{1+b-2b^n}{(1-b)^2}.e = f(e,n)$$

Επομένως η στάσιμη πολιτική $(\tilde{d}^n)^{\infty}$ είναι $f(e,n)$- $\acute{a}ristha$. W

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

Αν επιθυμούμε η πολιτική $(\tilde{d}^n)^{\infty}$ να είναι h- $\acute{a}ristha$ για δοσμένο $h>0$,τότε μπορούμε να επιλέξουμε προκαθορισμένο σφάλμα e και χρονικό ορίζοντα n έτσι ώστε:

$$f(e,n) \leq h .$$

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

$$\frac{1+b}{(1-b)^2}.e \leq \frac{h}{2} \quad , \quad \frac{2b^n}{1-b}.\Lambda \leq \frac{h}{2}$$

δηλαδή: $e \pounds \frac{(1- b)^2.h}{2.(1+ b)}$, $n \geq \frac{\ln(\frac{(1- b).h}{4L})}{\ln b}$.

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

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

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

και $\boldsymbol{\delta^n}$ είναι η άριστη συνάρτηση ελέγχου στον χρονικό ορίζοντα n ,σύμφωνα με τη σχέση (4.1.4). Τότε η στάσιμη πολιτική $(\delta^n)^\infty$ είναι $\frac{2b.h}{1- b}$- $\acute{a}ristah$,δηλαδή

$$\| V(./d^n) - V^* \| \leq \frac{2b.h}{1- b} \qquad \textbf{4.1.15}$$

Puterman [98]

W

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

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

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

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

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

$$\| \tilde{V}_n - \tilde{V}_{n-1} \| £ h \quad (\text{όπου } h f 0) \qquad \mathbf{4.1.16}$$

και $\tilde{d}^n$ είναι η συνάρτηση ελέγχου στον χρονικό ορίζοντα n σύμφωνα με τη σχέση (4.1.5). Τότε η στάσιμη πολιτική $(\tilde{d}^n)^¥$ είναι $f(e,h) - arist h$,όπου

$$f(e,h) := \frac{1 + b - 2b^n}{(1-b)^2}.e + \frac{2b.h}{1-b}\ , \qquad \mathbf{4.1.17}$$

δηλαδή**:** $\| V(./\tilde{d}^n) - V^* \| \le f(e,h)$

Απόδειξη

Από την τριγωνική ιδιότητα της νόρμας του supremum παίρνουμε

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

$$\leq \left\| V(./\tilde{d}^n) - \tilde{V}_n \right\| + s_n + \left\| V_n - V^* \right\| \qquad \textbf{4.1.18}$$

Εφαρμόζοντας πάλι την τριγωνική ιδιότητα έχουμε

$$\left\| V(./\tilde{d}^n) - \tilde{V}_n \right\| \leq \left\| V(./\tilde{d}^n) - H_{\tilde{d}^n} \tilde{V}_n \right\| + \left\| H_{\tilde{d}^n} \tilde{V}_n - \tilde{V}_n \right\| \qquad \textbf{4.1.19}$$

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

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

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

$$\leq \beta . \left\| V(./\tilde{d}^n) - \tilde{V}_n \right\| + b.h$$

Επομένως

$$\left\| V(./\tilde{d}^n) - \tilde{V}_n \right\| \leq \frac{b.h}{1-b} \qquad \textbf{4.1.20}$$

Από το λήμμα 4.1.1 (iii) παίρνουμε:

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

Άρα

$$\| V_n - V^* \| \leq \frac{b}{1-b} \| V_n - V_{n-1} \| \qquad \textbf{4.1.21}$$

Από την τριγωνική ιδιότητα και τη σχέση (4.1.16) έχουμε

$$\| V_n - V_{n-1} \| \leq \| V_n - \tilde{V}_n \| + \| \tilde{V}_n - \tilde{V}_{n-1} \| + \| \tilde{V}_{n-1} - V_{n-1} \| \leq s_n + h + s_{n-1} \qquad \textbf{4.1.22}$$

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

$$\| V_n - V^* \| \leq \frac{b}{1-b} \| V_n - V_{n-1} \| \leq \frac{b}{1-b} .(s_n + h + s_{n-1}) \qquad \textbf{4.1.23}$$

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

$$\| V(./\tilde{d}^n) - V^* \| \leq \frac{b.h}{1-b} + s_n + \frac{b.(s_n + s_{n-1} + h)}{1-b} =$$

$$= \frac{s_n + b.s_{n-1}}{1-b} + \frac{2bh}{1-b} = \frac{1+b-2b^n}{(1-b)^2}.e + \frac{2b.h}{1-b} = f(e,h)$$

Επομένως η στάσιμη πολιτική $(\tilde{d}^n)^{\infty}$ είναι $f(e,h)$**-** $\acute{a}risth$**.** W

Σημειώνουμε ότι για ε=0 η πρόταση 4.1.6 δίνει το ίδιο φράγμα όπως και η πρόταση 4.1.5 **:** $f(0,h) = \frac{2b.h}{1-b}$.Αν επιθυμούμε η πολιτική $(\tilde{d}^n)^{\infty}$ να είναι l **-** $\acute{a}risth$ για δοσμένο $l > 0$, τότε μπορούμε να επιλέξουμε προκαθορισμένο σφάλμα e και φράγμα h για το τροποποιημένο κατάλοιπο Bellman σε κάποιο χρονικό ορίζοντα έτσι ώστε:

$$f(e,h) \leq l$$

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

$$\frac{1+b}{(1-b)^2}.e \pounds \frac{l}{2} \quad , \quad \frac{2b.h}{1-b} \pounds \frac{l}{2}$$

δηλαδή: $$e \pounds \frac{(1-b)^2.l}{2.(1+b)} \quad , \quad h \pounds \frac{(1-b).l}{4b}.$$

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

4.2. Κατασκευή φραγμάτων για την βέλτιστη συνάρτηση τιμών.

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

Α) Κατασκευή άνω φραγμάτων για την V*

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

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

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

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

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

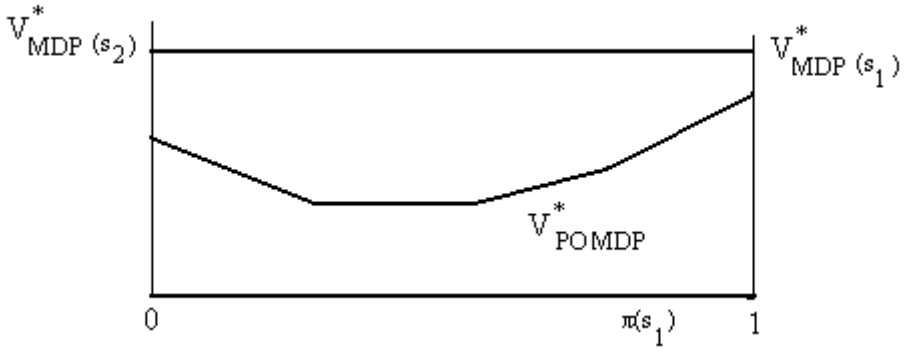
$$V^*(\pi) \le \sum_{i=1}^{N} \pi_i \, . V^*_{MDP}(i). \qquad \textbf{4.2.2}$$

Επειδή

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

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

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



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

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

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

$$V_n = HV_{n-1}, n = 1, 2, \ldots \qquad \textbf{4.2.3}$$

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

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

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

όπου $q_{\max} = \max_{i,a} q(i,a)$.

Για n=0 έχουμε την ειδική περίπτωση

$$V^*(\pi) \le \frac{1}{1-\beta}.q_{\max} \qquad \forall \pi \in \Pi .$$

Β) Κατασκευή κάτω φραγμάτων για την V*

1)Μέσω της βέλτιστης συνάρτησης τιμών της αντίστοιχης UMDP (unobservable MDP), μη παρατηρήσιμη Μαρκοβιανή διαδικασία αποφάσεων, χωρίς μηνύματα (Madani [78]).

Ένα κάτω φράγμα για την βέλτιστη συνάρτηση τιμών V* μιας POMDP είναι η βέλτιστη συνάρτηση τιμών V^*_{UMDP} της αντίστοιχης μη παρατηρήσιμης διαδικασίας αποφάσεων (UMDP),που ικανοποιεί την ακόλουθη εξίσωση αριστοποίησης:

$$V^*_{UMDP}(\pi) = \max_a \{\pi.q(i,a) + \beta.V^*_{UMDP}(T(\pi,\alpha))\}, \pi \in \Pi$$

Όπου $T(\pi,\alpha) = \pi.P^{\alpha}$,$\pi \in \Pi, a \in A$.

Αποδεικνύεται ότι $V^*_{UMDP} \le V^*(\pi) \qquad \forall \pi \in \Pi .$

Παρόλο που το πρόβλημα UMDP είναι απλούστερο από το αντίστοιχο POMDP, ο υπολογισμός της V^*_{UMDP} δεν είναι απλός και σε γενικές γραμμές ακολουθούνται οι ίδιες μέθοδοι προσέγγισης με την V^*.Επομένως το κάτω φράγμα V^*_{UMDP} για την V^* έχει μόνο θεωρητική αξία.

2) Μέσω των συναρτήσεων V_n (Lovejoy [76])

Μια ακολουθία κάτω φραγμάτων για την V^* σχηματίζεται μέσω της ακολουθίας των συναρτήσεων $\{V_n\}$ που ορίζονται από την (9.2.3). Για n=0,1,2,… έχουμε:

$$V_n(\pi) + \frac{\beta^n}{1-\beta}.q_{\min} \le V^*(\pi) \quad \forall \pi \in \Pi . \qquad \textbf{4.2.5}$$

$$q_{\min} = \min_{i,a} q(i,a) .$$

Για n=0 έχουμε την ειδική περίπτωση

$$\frac{1}{1-\beta}.q_{\min} \le V^*(\pi) \qquad \forall \pi \in \Pi .$$

3) Μέσω κάτω φραγμάτων των συναρτήσεων V_n (Lovejoy [76]).

Έστω $\upsilon(\pi), \pi \in \Pi$ μια κατά τμήματα γραμμική και κυρτή συνάρτηση. Όπως είναι γνωστό, ή συνάρτηση Hυ είναι επίσης κατά τμήματα γραμμική και κυρτή συνάρτηση, πού καθορίζεται μέσω ενός πεπερασμένου συνόλου Γ_H από gradients (διανύσματα) του χώρου R^N:

$$H\upsilon(\pi) = \max_{\gamma \in \Gamma_H} \pi.\gamma, \ \pi \in \Pi .$$

Έστω $\gamma(\pi) \in \Gamma_H$ το gradient της Hυ στο π, δηλαδή:

$$H\upsilon(\pi) = \pi.\gamma(\pi).$$

Το gradient γ(π) προσδιορίζεται μέσω του αλγορίθμου του ενός βήματος (βλέπε ενότητα 2.2).

Ένα κάτω φράγμα για τη συνάρτηση Hυ κατασκευάζεται ως εξής:Θεωρούμε $G \subseteq \Pi$ ένα πεπερασμένο σύνολο από δ.π (grid points) και Γ_L το σύνολο των gradients της συνάρτησης Hυ στα δ.π του συνόλου G, δηλαδή

$$\Gamma_L = \{\gamma(\pi): \pi \in G\}.$$

Ορίζουμε τη συνάρτηση

$$H_L\upsilon(\pi) := \max_{\gamma \in \Gamma_L} \pi.\gamma, \ \pi \in \Pi .$$

Επειδή $\Gamma_L \subseteq \Gamma_H$ συνάγεται ότι

$$H_L\upsilon(\pi) \le H\upsilon(\pi) \ \ \forall \pi \in \Pi .$$

Ακολούθως κατασκευάζεται η ακολουθία των συναρτήσεων $\{V_{L_n}\}$

$$V_{L_n} = H_L V_{L_{n-1}} \text{ ,n=1,2,3,…}$$

Όπου $V_{L_0} = 0$ (μηδενική συνάρτηση).

Σημειώνουμε ότι το σύνολο G παραμένει το ίδιο για όλα τα n.Με άλλα λόγια, το σύνολο των gradients της συνάρτησης V_{L_n} είναι το σύνολο των gradients της $H_L V_{L_{n-1}}$ στα δ.π του G. Αποδεικνύεται επαγωγικά ότι:

$$V_{L_n}(\pi) \le V_n(\pi) \quad \forall \pi \in \Pi .$$

Μια ακολουθία κάτω φραγμάτων για την V^* σχηματίζεται μέσω της ακολουθίας των συναρτήσεων $\{V_{L_n}\}$.Για n=0,1,2… έχουμε:

$$V_{L_n}(\pi) + \frac{\beta^n}{1-\beta}.q_{\min} \le V^*(\pi) \quad \forall \pi \in \Pi . \qquad \textbf{4.2.6}$$

Προφανώς το κάτω φράγμα της (4.2.6) ως προσέγγιση της V^* υπολείπεται του αντίστοιχου κάτω φράγματος της (4.2.5),όμως ο υπολογισμός του είναι γενικά πολύ απλούστερος.

Απαλοιφή αποφάσεων πού δεν είναι άριστες για κάποιο δ.π

Η εύρεση άνω και κάτω φραγμάτων για την βέλτιστη συνάρτηση τιμών V^* είναι δυνατόν να συμβάλει στην απαλοιφή μη άριστων αποφάσεων για κάποιο δ.π στο πρόβλημα POMDP για άπειρο χρονικό ορίζοντα.

Θεωρούμε τη συνάρτηση $h : \Pi \times A \times B(\Pi) \to \square$

πού ορίζεται ως:

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

(βλέπε ενότητα 1.4).

Έστω V_L, V_U κάτω και άνω φράγμα αντίστοιχα για τήν V^* :

$$V_L(\pi) \le V^*(\pi) \le V_U(\pi) \qquad \forall \ \pi \in \Pi .$$

Η επόμενη πρόταση παρέχει ένα κριτήριο απαλοιφής μη άριστων αποφάσεων για κάποιο δ.π στο πρόβλημα του άπειρου χρονικού ορίζοντα.

Πρόταση 4.2.1 :(MC Queen)[100]

Αν για κάποιο $\pi \in \Pi$ και κάποια απόφαση α ισχύει

$h(\pi, a, V_U) < HV_L(\pi),$

τότε η απόφαση α δεν είναι άριστη για το π στο πρόβλημα του άπειρου χρονικού ορίζοντα. □

4.3. Προσεγγίσεις της άριστης συνάρτησης τιμών για άπειρο χρονικό ορίζοντα και προσδιορισμός σχεδόν άριστων πολιτικών μέσω φραγμάτων.

Στην παρούσα ενότητα θα ασχοληθούμε με προσεγγίσεις της άριστης συνάρτησης τιμών V^* καθώς και με τον προσδιορισμό σχεδόν άριστων πολιτικών, πού βασίζονται σε συναρτήσεις φραγμάτων της V^*.

Ο Hauskrecht [48], λαμβάνοντας δύο οποιαδήποτε άνω και κάτω φράγματα ως αρχικές προσεγγίσεις της V^* και εφαρμόζοντας την επαναληπτική μέθοδο τιμών (value-iteration) κατασκευάζει νέα φράγματα,που αποτελούν αυθαίρετα καλές προσεγγίσεις της V^* και οι προβαλλόμενες από αυτά πολιτικές (lookahead-controllers) είναι σχεδόν άριστες πολιτικές.Το πλήθος των βημάτων (επαναλήψεων) που απαιτούνται εξαρτάται από την «απόσταση» των αρχικών φραγμάτων και την επιθυμητή ακρίβεια της προσέγγισης. Εμείς επεκτείνουμε αυτά τα αποτελέσματα εφαρμόζοντας προσεγγιστικό αλγόριθμο των ακρότατων σημείων. Επιπλέον υπολογίζουμε τον αριθμό των βημάτων καθώς και το προκαθορισμένο σφάλμα του αλγορίθμου, ώστε να επιτύχουμε οποιαδήποτε επιθυμητή ακρίβεια προσέγγισης. Θα περιορισθούμε μόνο σε προσεγγίσεις προβλημάτων POMDP στα πλαίσια του κριτηρίου μεγιστοποίησης των ολικών εσόδων σε άπειρο χρονικό ορίζοντα. Οι αντίστοιχες προσεγγίσεις στα πλαίσια του κριτηρίου ελαχιστοποίησης του ολικού κόστους για άπειρο χρονικό ορίζοντα είναι ανάλογες.

Έστω $V_L(\pi)$, $V_U(\pi)$,$\pi \in \Pi$ συνάρτηση κάτω και άνω φράγματος αντίστοιχα για την συνάρτηση V^*:

$$V_L(\pi) \le V^*(\pi) \le V_U(\pi), \pi \in \Pi .$$

Πρόταση 4.3.1: Milos Hauskrecht [48]

Ας είναι $\eta = \|V_U - V_L\| = \sup_{\pi} |V_U(\pi) - V_L(\pi)|$ και V οποιοδήποτε από τα δύο φράγματα, δηλαδή $V=V_L$ η $V=V_U$.Τότε

i) $\|V - V^*\| \le \eta$

ii) Αν δ είναι η συνάρτηση ελέγχου για την οποία

$$H_{\delta}V = HV$$

δηλαδή

$$\delta(\pi):=\arg \max_{a} \{ \pi q^{\alpha} + \beta . \sum_{\theta=1}^{M} \{\theta/\pi,\alpha\}.V\,(T(\pi,\theta,\alpha))\}, \pi \in \Pi$$

τότε η στάσιμη πολιτική δ^{∞} είναι $\frac{2-\beta}{1-\beta}.\eta - \acute{\alpha}\rho\iota\sigma\tau\eta$,δηλαδή:

$$\| V(./\delta) - V^* \| \leq \frac{2-b}{1-b}.h \qquad W$$

Ο αριθμός η εκφράζει την απόσταση (μέγιστη απόλυτη διαφορά) ανάμεσα στα φράγματα V_L και V_U. Η πρόταση 4.3.1 δηλώνει ότι οποιοδήποτε από τα φράγματα V_L, V_U μπορεί να θεωρηθεί προσέγγιση της βέλτιστης συνάρτησης τιμών V* με μέγιστο σφάλμα το πολύ η .Επίσης οποιοδήποτε από τα δύο φράγματα V_L ,V_U μπορεί να χρησιμοποιηθεί ως εφαλτήριο για την εύρεση $\frac{2-b}{1-b}.h$ -άριστης πολιτικής.

Η δ αντιπροσωπεύει την βέλτιστη συνάρτηση ελέγχου «για ένα βήμα μπροστά»(one-step lookahead controller) αναφορικά με το φράγμα V που επιλέγουμε.

Η συνάρτηση τιμών που αντιστοιχεί στην πολιτική δ^{∞},V(./δ), προσεγγίζει τη βέλτιστη συνάρτηση τιμών V* με μέγιστο σφάλμα το πολύ $\frac{2-b}{1-b}.h$. *Η πολιτική* δ^{∞} *αναφέρεται ως προβαλλόμενη πολιτική από το φράγμα V (lookahead- policy).* Θεωρούμε τις αναγωγικές σχέσεις:

$$V_L^k = HV_L^{k-1},\ V_U^k = HV_U^{k-1}, k = 1,2,3,...$$

$$V_L^o = V_L \quad ,\ V_U^o = V_U$$

Εισάγοντας τον τελεστή H_k πού συμβολίζει την επαναληπτική χρήση του τελεστή Η k φορές, οι παραπάνω σχέσεις γράφονται:

$$V_L^k = H_k V_L \leq H_k V^* = V^* \leq H_k V_U = V_U^k$$

Επομένως οι συναρτήσεις V_L^k, V_U^k αποτελούν αντίστοιχα κάτω και άνω φράγμα για την V*. *Θα αναφερόμαστε σε αυτές ως φράγματα τάξεως k.* Επειδή ο τελεστής H_k είναι συστολή modulus β^k έχουμε:

$$\left\| V_U^k - V_L^k \right\| \le \beta^k \left\| V_L - V_U \right\| = \beta^k . \eta .$$

Συνοψίζοντας, από τα φράγματα V_L, V_U για την συνάρτηση V*,εφαρμόζοντας επαναληπτικά τον τελεστή H k φορές, σχηματίζονται τα φράγματα τάξεως k V_L^k, V_U^k για την V* των οποίων η απόσταση (μέγιστη απόλυτη διαφορά) είναι το πολύ $\beta^k . \eta$.

Από την πρόταση 4.3.1 (i) συνάγεται ότι τα φράγματα k τάξεως μπορούν να θεωρηθούν προσεγγίσεις της συνάρτησης V* με μέγιστο σφάλμα το πολύ $\beta^k . \eta$. Πιο συγκεκριμένα έχουμε:

$$\left\| V^k - V^* \right\| \le \beta^k . \eta \,, \qquad \textbf{4.3.1}$$

όπου V^k είναι οποιοδήποτε από τα δύο φράγματα k τάξεως,

δηλαδή $V^k = V_L^k$ ή $V^k = V_U^k$.

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

$$\left\| V^k - V^* \right\| \le \lambda$$

τότε μπορούμε να επιλέξουμε την τάξη k έτσι ώστε:

$$\beta^k . \eta \le \lambda$$

δηλαδή $$k \ge \frac{\ln(\lambda / \eta)}{\ln \beta}$$

Εστω δ^k η βέλτιστη συνάρτηση ελέγχου για ένα βήμα μπροστά (one-step lookahead conroller) αναφορικά με το φράγμα V^k που επιλέγουμε , δηλαδή:

$$H_{\delta^k} V^k = H V^k \qquad \textbf{4.3.2}$$

Από την πρόταση 4.3.1 (ii) συνάγεται ότι η προβαλλόμενη από το φράγμα V^k πολιτική $(\delta^k)^\infty$ (lookahead policy) είναι $\frac{2-\beta}{1-\beta} . \beta^k . \eta - \acute{\alpha} \rho \iota \sigma \tau \eta .$

Αν επιθυμούμε η πολιτική $(\delta^k)^\infty$ να είναι λ-άριστη για δοσμένο $\lambda \succ 0$, τότε μπορούμε να επιλέξουμε την τάξη k έτσι ώστε:

$$\frac{2-\beta}{1-\beta}.\beta^k.\eta \le \lambda$$

δηλαδή

$$k \ge \frac{\ln(\lambda.(1-\beta)/\eta.(2-\beta))}{\ln\beta}$$

Σημειώνουμε ότι οι προσεγγίσεις πού περιγράψαμε είναι εφικτές μόνο στην περίπτωση όπου εφαρμόζουμε διαδοχικά τον ακριβή αλγόριθμο των ακρότατων σημείων του κεφ. 3 (προκαθορισμένο σφάλμα ε=0), οπότε επιτυγχάνεται ακριβής υπολογισμός των φραγμάτων k τάξεως V_L^k, V_U^k καθώς επίσης ο καθορισμός της συνάρτησης ελέγχου δ^k μέσω της σχέσης (4.3.2) είναι εφικτός.

Ανάλογες προσεγγίσεις είναι δυνατές εφαρμόζοντας διαδοχικά τον προσεγγιστικό αλγόριθμο των ακρότατων σημείων. Επιλέγοντας ως αρχικές συναρτήσεις τις συναρτήσεις φραγμάτων V_L, V_U και εφαρμόζοντας k φορές τον αλγόριθμο των ακροτάτων σημείων με προκαθορισμένο σφάλμα $\varepsilon > 0$ υπολογίζονται οι προσεγγίσεις $\tilde{V}_L^k, \tilde{V}_U^k$ των φραγμάτων k τάξης V_L^k, V_U^k αντίστοιχα, με συσσωρευμένο σφάλμα προσέγγισης σ_k σε κάθε περίπτωση (πρβλ. ενότητα 3.4). Συγκεκριμένα,

$$\tilde{V}_L^k = \tilde{\mathrm{H}}\tilde{V}_L^{k-1}, \qquad \tilde{V}_U^k = \tilde{\mathrm{H}}\tilde{V}_U^{k-1}, \text{k=1,2,3…}$$

$$\tilde{V}_L^0 = V_L, \qquad \tilde{V}_U^0 = V_U$$

Για κ =1,2,3,… έχουμε:

$$\tilde{V}_L^k(\pi) \le V_L^k(\pi) \le \tilde{V}_L^k(\pi) + \sigma_k \qquad \forall \pi \in \Pi$$

$$\tilde{V}_U^k(\pi) \le V_U^k(\pi) \le \tilde{V}_U^k(\pi) + \sigma_k \qquad \forall \pi \in \Pi \qquad \textbf{4.3.3}$$

Το συσσωρευμένο σφάλμα προσέγγισης σ_k ικανοποιεί την αναγωγική σχέση (3.4.6) και υπολογίζεται από την σχέση (3.4.7).

Σημειώνουμε ότι η προσέγγιση $\tilde{V}_L^k$ του κάτω φράγματος k τάξεως V_L^k είναι επίσης κάτω φράγμα για την συνάρτηση V^*.Δεν μπορούμε όμως να ισχυριστούμε το ίδιο για την προσέγγιση $\tilde{V}_U^k$ του άνω φράγματος k τάξεως V_U^k πού ενδέχεται να μην είναι άνω φράγμα της V^*.Ωστόσο αμφότερες οι συναρτήσεις $\tilde{V}_L^k, \tilde{V}_U^k$

μπορούν να ληφθούν ως προσεγγίσεις της συνάρτησης V^*. Το σφάλμα αυτών των προσεγγίσεων παρέχεται στην επόμενη πρόταση.

Παρατήρηση

Στα προβλήματα κόστους ισχύει το αντίστροφο. Πράγματι (βλέπε ενότητα 3.5) για k=1,2,3…έχουμε:

$$\tilde{V}_L^k(\pi) - \sigma_k \le V_L^k(\pi) \le \tilde{V}_L^k(\pi)$$

$$\tilde{V}_U^k(\pi) - \sigma_k \le V_U^k(\pi) \le \tilde{V}_U^k(\pi)$$

Επομένως η προσέγγιση $\tilde{V}_U^k(\pi)$ του άνω φράγματος $V_U^k(\pi)$ εξακολουθεί να είναι άνω φράγμα για την συνάρτηση V*,ενώ η προσέγγιση $\tilde{V}_L^k(\pi)$ του κάτω φράγματος $V_L^k(\pi)$,ενδέχεται να μην είναι κάτω φράγμα για την V*.

Ωστόσο και στην περίπτωση αυτή αμφότερες οι συναρτήσεις $\tilde{V}_L^k, \tilde{V}_U^k$,μπορούν να ληφθούν ως προσεγγίσεις της V*.

Πρόταση 4.3.2: Αν $\eta = \|V_U - V_L\| = \sup_\pi |V_U(\pi) - V_L(\pi)|$,τότε για k=1,2,3…

i) $\|\tilde{V}_L^k - \mathbf{V^*}\| \le \dfrac{1-\beta^k}{1-\beta}.\varepsilon + \beta^k.\eta$,

ii) $\|\tilde{V}_U^k - V^*\| \le \dfrac{1-\beta^k}{1-\beta}.\varepsilon + \beta^k.\eta$.

Απόδειξη

i) Από την τριγωνική ιδιότητα της νόρμας supremum και τις σχέσεις (4.3.1),(4.3.3) και (3.4.7) παίρνουμε:

$$\|\tilde{V}_L^k - V^*\| \le \|\tilde{V}_L^k - V_L^k\| + \|V_L^k - V^*\| \le \sigma_k + \beta^k.\eta =$$

$$= \frac{1-\beta^k}{1-\beta}.\varepsilon + \beta^k.\eta$$

ii) Παρόμοια □

Σημειώνουμε ότι είναι δυνατόν να επιτευχθούν προσεγγίσεις της συνάρτησης V* με οποιαδήποτε επιθυμητή ακρίβεια. Συγκεκριμένα αν $\lambda \succ 0$ είναι ένα επιθυμητό άνω φράγμα για το μέγιστο σφάλμα προσέγγισης, δηλαδή

$$\left\|\tilde{V}^k - V^*\right\| \le \lambda$$

όπου $\tilde{V}^k$ είναι οποιαδήποτε από τις δύο προσεγγίσεις $\tilde{V}_L^k, \tilde{V}_U^k$, τότε μπορούμε να επιλέξουμε προκαθορισμένο σφάλμα ε και τάξη k έτσι ώστε

$$\frac{1-\beta^k}{1-\beta}.\varepsilon + \beta^k.\eta \le \lambda$$

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

$$\frac{\varepsilon}{1-\beta} \le \frac{\lambda}{2}, \qquad \beta^k.\eta \le \frac{\lambda}{2}$$

δηλαδή $\qquad \varepsilon \le \dfrac{\lambda.(1-\beta)}{2} \qquad ,k \ge \dfrac{\ln(\frac{\lambda}{2\eta})}{\ln(\beta)}$

Εστω $\tilde{\delta}^k$ η συνάρτηση ελέγχου που προκύπτει εφαρμόζοντας τον αλγόριθμο των ακρότατων σημείων με προκαθορισμένο σφάλμα $\varepsilon > 0$ στην προσέγγιση $\tilde{V}^k$, δηλαδή

$$H_{\tilde{\delta}^K}\tilde{V}^k = \tilde{H}\tilde{V}^k \qquad \textbf{4.3.4}$$

(βλέπε επίσης παρατήρηση 2 στην ενότητα 3.2)

Η στάσιμη πολιτική $(\tilde{\delta}^k)^\infty$ μπορεί να ληφθεί ως προσέγγιση της άριστης πολιτικής $(\delta^*)^\infty$. Το σφάλμα αυτής της προσέγγισης παρέχεται στην πρόταση 4.3.3

Λήμμα 4.3.1: Αν $\eta = \left\|V_U - V_L\right\| = \sup_\pi \left|V_U(\pi) - V_L(\pi)\right|$, τότε για k=1,2,3…

$\left\|HV^k - V^k\right\| \le \beta^k.\eta$

όπου V^k είναι οποιοδήποτε από τα φράγματα k τάξεως V_L^k, V_U^k.

Απόδειξη

Επειδή

$V_L^K \le V^* \le V_U^k$ και

$HV_L^K \le HV^* = V^* \le HV_U^k$

συνάγεται ότι για κάθε $\pi \in \Pi$,

$$\left|HV^k(\pi) - V^k(\pi)\right| \le \max\{\left|HV^k(\pi) - V^*(\pi)\right|, \left|V^k(\pi) - V^*(\pi)\right|\} \le$$

$$\le \max\{\left\|HV^k - V^*\right\|, \left\|V^k - V^*\right\|\}$$

από την οποία προκύπτει ότι:

$$\left\|HV^k - V^k\right\| \le \max\{\left\|HV^k - V^*\right\|, \left\|V^k - V^*\right\|\} \qquad \textbf{4.3.5}$$

Από τις σχέσεις (4.3.1),(4.3.5) και την ακόλουθη σχέση

$$\left\|HV^k - V*\right\| = \left\|HV^k - HV^*\right\| \le \beta.\left\|V^k - V^*\right\| \le \beta^{k+1}.\eta$$

συνάγεται ότι

$$\left\|HV^k - V^k\right\| \le \max\{\beta^{k+1}.\eta, \beta^k.\eta\} = \beta^k.\eta \qquad \square$$

Πρόταση 4.3.3: Αν $\eta = \left\|V_U - V_L\right\| = \sup_\pi \left|V_U(\pi) - V_L(\pi)\right|$, $\tilde{V}^k$ είναι οποιαδήποτε από τις προσεγγίσεις $\tilde{V}_L^k, \tilde{V}_U^k$ των φραγμάτων k-τάξεως V_L^K, V_U^k και $\tilde{\delta}^k$ η συνάρτηση ελέγχου πού ορίζεται από τη σχέση (4.3.4) τότε η στάσιμη πολιτική $(\tilde{\delta}^k)^\infty$ είναι f (ε,κ)-άριστη, όπου

$$f(\varepsilon,\kappa) := \frac{3-\beta-2\beta^k}{(1-\beta)^2}\cdot\varepsilon + \frac{2-\beta}{1-\beta}\cdot\beta^k\cdot\eta \qquad \textbf{4.3.6}$$

δηλαδή:

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

Απόδειξη

Από την τριγωνική ιδιότητα της νόρμας supremum παίρνουμε

$$\left\|V(./\tilde{\delta}^k) - V^*\right\| \le \left\|V(./\tilde{\delta}^k) - \tilde{V}^k\right\| + \left\|\tilde{V}^k - V^*\right\| \qquad \textbf{4.3.7}$$

Εφαρμόζοντας πάλι την τριγωνική ιδιότητα έχουμε:

$$\left\|V(./\tilde{\delta}^k) - \tilde{V}^k\right\| \le \left\|V(./\tilde{\delta}^k) - H_{\tilde{\delta}^k}\tilde{V}^k\right\| + \left\|\mathrm{H}_{\tilde{\delta}^k}\tilde{V}^k - \tilde{V}^k\right\| \qquad \textbf{4.3.8}$$

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

$$\left\|V(./\tilde{\delta}^k)-H_{\tilde{\delta}^\kappa}\tilde{V}^k\right\|=\left\|\mathrm{H}_{\tilde{\delta}^\kappa}V(./\tilde{\delta}^k)-H_{\tilde{\delta}^k}\tilde{V}^k\right\|\le\beta\cdot\left\|V(\cdot/\tilde{\delta}^k)-\tilde{V}^k\right\| \quad \textbf{4.3.9}$$

Από τις σχέσεις (4.3.8),(4.3.9) παίρνουμε

$$\left\|V(./\tilde{\delta}^k)-\tilde{V}^k\right\|\le\left\|H_{\tilde{\delta}^k}\tilde{V}^k-\tilde{V}^k\right\|\Big/(1-\beta) \quad \textbf{4.3.10}$$

Θεωρούμε $\mathrm{V}^k=V_L^k$ αν $\tilde{V}^k=\tilde{V}_L^k$

$\mathrm{V}^k=V_U^k$ άν $\tilde{V}^k=\tilde{V}_U^k$

Από την τριγωνική ιδιότητα, τις σχέσεις (4.3.3),(4.3.4) και το λήμμα 4.3.1 παίρνουμε:

$$\left\|H_{\tilde{\delta}^\kappa}\tilde{V}^k-\tilde{V}^k\right\|=\left\|\tilde{\mathrm{H}}\tilde{V}^k-\tilde{V}^k\right\|\le\left\|\tilde{H}\tilde{V}^k-H\tilde{V}^k\right\|+\left\|HV^k-V^k\right\|+\left\|V^k-\tilde{V}^k\right\|\le$$

$$\le\sigma_{k+1}+\beta^k.\eta+\sigma_k \quad \textbf{4.3.11}$$

Από τις σχέσεις (4.3.10), (4.3.11) παίρνουμε

$$\left\|V(./\tilde{\delta}^k)-\tilde{V}^k\right\|\le\frac{\sigma_{k+1}+\sigma_\kappa+\beta^k.\eta}{1-\beta} \quad \textbf{4.3.12}$$

Από την πρόταση 4.3.1 έχουμε:

$$\left\|\tilde{V}^k-V^*\right\|\le\sigma_k+\beta^k.\eta \quad \textbf{4.3.13}$$

Από τις σχέσεις (4.3.7),(4.3.13) και λαμβάνοντας υπόψη τη σχέση (3.4.7) συνάγεται ότι:

$$\left\|V(./\tilde{\delta})-V^*\right\|\le\frac{\sigma_{k+1}+\sigma_\kappa+\beta^k.\eta}{1-\beta}+\sigma_k+\beta^k.\eta=\frac{(2-\beta).\sigma_k+\sigma_{\kappa+1}}{1-\beta}+\frac{2-\beta}{1-\beta}.\beta^k.\eta$$

$$=\frac{3-\beta-2\beta^k}{(1-\beta)^2}.\varepsilon+\frac{2-\beta}{1-\beta}\cdot\beta^k\cdot\eta=f(\varepsilon,k)$$

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

Αν επιθυμούμε η πολιτική $(\tilde{\delta}^k)^\infty$ να είναι λ -άριστη για δοσμένο $\lambda > 0$,τότε μπορούμε να επιλέξουμε το προκαθορισμένο σφάλμα ε και την τάξη k έτσι ώστε

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

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

$$\frac{3-\beta}{(1-\beta)^2}.\varepsilon \le \frac{\lambda}{2} \quad , \quad \frac{2-\beta}{1-\beta}\cdot\beta^k\cdot\eta \le \frac{\lambda}{2},$$

δηλαδή

$$\varepsilon \le \frac{\lambda'(1-\beta)^2}{2(3-\beta)} \quad , \quad k \ge \frac{\ln(\frac{\lambda(1-\beta)}{2(2-\beta).\eta})}{\ln\beta}. \qquad \square$$

Παρατηρήσεις

1) Σε όλες τις περιπτώσεις, είτε χρησιμοποιούμε τον ακριβή είτε τον προσεγγιστικό αλγόριθμο των ακρότατων σημείων, ο ελάχιστος αριθμός βημάτων (επαναλήψεων) k πού απαιτείται ώστε να επιτύχουμε οποιαδήποτε επιθυμητή ακρίβεια στην προσέγγιση της V^* ή της άριστης πολιτικής εξαρτάται από την απόσταση η των αρχικών άνω και κάτω φραγμάτων. Επομένως η διαδικασία προσέγγισης μπορεί να επιταχυνθεί σημαντικά αν το η είναι μικρό, δεδομένου ότι το η αποτελεί μέτρο του σφάλματος της αρχικής προσέγγισης της V^* (πρόταση 4.3.1(i)).

2) Επειδή έχουμε τη δυνατότητα ως αρχική προσέγγιση της V^* να επιλέξουμε ένα από τα δύο αρχικά φράγματα (είτε το κάτω είτε το άνω φράγμα),προφανώς είναι λογικό να επιλέξουμε το απλούστερο από αυτά. Σημαντικό πλεονέκτημα ως αρχική προσέγγιση της V^* παρουσιάζουν τα φράγματα πού υπολογίζονται μέσω της V^*_{MDP} (βλέπε παράγραφο 4.2) λόγω της απλότητάς τους.

ΣΥΜΠΕΡΑΣΜΑΤΑ

Προσεγγίσεις της άριστης συνάρτησης τιμών και της άριστης πολιτικής σε άπειρο χρονικό ορίζοντα επιτυγχάνονται με οποιαδήποτε επιθυμητή ακρίβεια μέσω διαδοχικών επαναλήψεων του αλγόριθμου των ακροτάτων σημείων που περιγράψαμε στο κεφάλαιο 3.Καλές προσεγγίσεις επιτυγχάνονται επίσης αν το ‘‘κατάλοιπο Bellman’’ σε κάποιο βήμα (επανάληψη) είναι αρκούντως μικρό.

Η μέθοδος των φραγμάτων αναφέρεται στην επιλογή άνω και κάτω φραγμάτων ως αρχικών προσεγγίσεων της άριστης συνάρτησης τιμών η οποία συνοδεύεται από την επαναληπτική εφαρμογή του αλγόριθμου των ακροτάτων σημείων για την δημιουργία νέων προσεγγίσεων.Από τις προβαλλόμενες συναρτήσεις ελέγχου των νέων προσεγγίσεων (lookahead controllers) κατασκευάζονται σχεδόν άριστες πολιτικές. Επιτάχυνση της διαδικασίας είναι δυνατή αν η απόσταση των αρχικών φραγμάτων είναι μικρή.

Σε κάθε περίπτωση υπολογίζεται ο απαιτούμενος αριθμός επαναλήψεων καθώς και το προκαθορισμένο σφάλμα του αλγορίθμου των ακροτάτων σημείων, έτσι ώστε να επιτυγχάνεται προσέγγιση με οποιαδήποτε επιθυμητή ακρίβεια.