

**ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΑΤΡΩΝ
ΤΜΗΜΑΤΑ ΜΑΘΗΜΑΤΙΚΩΝ - ΜΗΧΑΝΙΚΩΝ Η/Υ & ΠΛΗΡΟΦΟΡΙΚΗΣ**

**ΔΙΑΤΜΗΜΑΤΙΚΟ ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ
«ΜΑΘΗΜΑΤΙΚΑ ΤΩΝ ΥΠΟΛΟΓΙΣΤΩΝ ΚΑΙ ΤΩΝ ΑΠΟΦΑΣΕΩΝ»**

ΑΝΑΣΤΑΣΙΟΣ ΚΑΤΣΙΚΗΣ

**ΑΝΑΛΥΣΗ ΚΑΙ ΥΠΟΛΟΓΙΣΤΙΚΗ ΠΟΛΥΠΛΟΚΟΤΗΤΑ
ΤΕΧΝΙΚΩΝ ΕΠΙΛΥΣΗΣ
ΠΡΟΒΛΗΜΑΤΩΝ ΓΡΑΜΜΙΚΟΥ ΠΡΟΓΡΑΜΜΑΤΙΣΜΟΥ**

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΠΑΤΡΑ 2009

Στον πατέρα μου Χρήστο

Περιεχόμενα

| | |
|---|----|
| Περίληψη | 7 |
| <u>Κεφάλαιο 1</u> | |
| Εισαγωγή - Ιστορική αναδρομή | 11 |
| <u>Κεφάλαιο 2</u> | |
| Θεμελίωση της μεθόδου Simplex | 15 |
| <u>Κεφάλαιο 3</u> | |
| Θεμελίωση της μεθόδου των ελλειψοειδών | 37 |
| <u>Κεφάλαιο 4</u> | |
| Μέθοδοι εσωτερικού σημείου | 45 |
| Ο αλγόριθμος του Karmarkar | 46 |
| Ο πρωτεύων-δυσικός αλγόριθμος εσωτερικού σημείου | 56 |
| <u>Κεφάλαιο 5</u> | |
| Υπολογιστική πολυπλοκότητα του αλγορίθμου Simplex και του αλγορίθμου του Karmarkar | 61 |
| Πολυωνυμική πολυπλοκότητα | 62 |
| Υπολογιστική πολυπλοκότητα του αλγορίθμου Simplex | 68 |
| Ανάλυση της σύγκλισης και της πολυπλοκότητας του αλγορίθμου του Karmarkar. | 74 |
| Βιβλιογραφία. | 82 |

Περίληψη

Το πρώτο κεφάλαιο περιλαμβάνει μια ιστορική αναδρομή σχετικά με τη γέννηση και την ανάπτυξη της Επιχειρησιακής Έρευνας και του Γραμμικού Προγραμματισμού. Επίσης παρουσιάζεται το χρονικό των μεγαλύτερων ανακαλύψεων: ο αλγόριθμος Simplex (Dantzig-1949), ο ελλειψοειδής αλγόριθμος (Khachian-1979) και ο αλγόριθμος εσωτερικών σημείων (Karmarkar-1983). Στη συνέχεια - δεύτερο κεφάλαιο - γίνεται η θεωρητική θεμελίωση της μεθόδου Simplex, συμπεριλαμβάνοντας τόσο την γεωμετρική-εποπτική παρουσίαση της μεθόδου, όσο και την αυστηρή αλγεβρική τεκμηρίωσή της μέσω θεωρημάτων.

Το τρίτο κεφάλαιο αφιερώθηκε στον αλγόριθμο των ελλειψοειδών, στη μέθοδο δηλαδή που ουσιαστικά απέδειξε ότι τα προβλήματα του γραμμικού προγραμματισμού μπορούν να λυθούν σε πολυωνυμικό χρόνο. Στο τέταρτο κεφάλαιο παρουσιάζεται η πιο σύγχρονη τάση στον τομέα επίλυσης προβλημάτων γραμμικού προγραμματισμού: οι μέθοδοι εσωτερικού σημείου. Συγκεκριμένα αναπτύσσεται ο αλγόριθμος του Karmarkar, η κατηγορία των μεθόδων ομοπαράλληλικής αλλαγής κλίμακας και ο πρωτεύοντας-δ्वि-κόσ αλγόριθμος εσωτερικού σημείου.

Τέλος, στο πέμπτο κεφάλαιο περιλαμβάνεται η παρουσίαση της έννοιας της υπολογιστικής πολυπλοκότητας αλγορίθμων, η πλήρης ανάλυση της πολυπλοκότητας των αλγορίθμων Simplex και εσωτερικού σημείου του Karmarkar, καθώς και η σύγκριση των δύο αλγορίθμων.

Abstract

The first chapter includes a historical retrospection in respect of the birth and growth of Operational Research and Linear Programming. Furthermore, the chronicle of the biggest discoveries is presented: the Simplex algorithm (Dantzig-1949), the ellipsoid algorithm (Khachian-1979) and the interior point algorithm (Karmarkar-1983). Thereafter -in the second chapter- the theoretical foundation of Simplex method is presented, including both the geometric-supervisory presentation and the strict algebraic documentation of the method via theorems.

The third chapter refers to the ellipsoid algorithm, namely the method that proved that the problems of linear programming can be solved in polynomial time. In the fourth chapter, the most contemporary tendency in the field of solving problems of linear programming, is presented: the methods of interior point. Particularly, the algorithm of Karmakar and the primal-dual algorithm of interior point are expounded.

Finally, the fifth chapter includes the presentation of the concept of computational complexity of algorithms, the complete analysis of complexity of algorithms Simplex and interior point of Karmakar, as well as the comparison of the two algorithms.

1. ΕΙΣΑΓΩΓΗ - ΙΣΤΟΡΙΚΗ ΑΝΑΔΡΟΜΗ

Η Επιχειρησιακή Έρευνα καρδιά της οποίας είναι ο γραμμικός προγραμματισμός είναι γέννημα, αλλά όχι θρέμμα, του Β' Παγκοσμίου Πολέμου. Τότε εφαρμόστηκαν για πρώτη φορά επιστημονικές μέθοδοι για την καλύτερη (βέλτιστη) κατανομή των συμμαχικών στρατευμάτων και εφοδίων στην Ευρώπη, έτσι ώστε να ελαχιστοποιηθεί ο χρόνος πρόσβασης στις εστίες μαχών και να βελτιωθεί ο εφοδιασμός, ο συντονισμός και η διοίκηση των συμμαχικών δυνάμεων. Την εποχή εκείνη μάλιστα έγινε και η βάπτιση της νεογέννητης επιστήμης. Ονομάστηκε στην αγγλική γλώσσα Operations Research (USA) ή Operational Research (Αγγλία), όρος που σημαίνει έρευνα στρατιωτικών επιχειρήσεων ή πιο ελεύθερα έρευνα πολέμου. Το ελληνικό της όνομα, Επιχειρησιακή Έρευνα, δεν φανερώνει την πραγματική της καταγωγή, αποδίδει όμως καλύτερα το σημερινό της περιεχόμενο, που είναι η μελέτη και επίλυση των προβλημάτων σημερινών επιχειρήσεων, εμπορικών ως επί το πλείστον και όχι στρατιωτικών.

Ας δούμε λίγο πιο αναλυτικά τι διαδραματιζόταν κατά τη διάρκεια του Β' παγκοσμίου πολέμου στις χώρες, οι οποίες ερίζουν για τη μητρότητα της νέας επιστήμης. Μιας επιστήμης βαθιά θεωρητικής από τη μια και απίστευτα εφαρμοσμένης από την άλλη, μιας επιστήμης που έμελλε να απασχολήσει αρκετές φορές την επιστημονική κοινότητα λόγω των σημαντικών θεωρητικών ανακαλύψεων αλλά και την παγκόσμια κοινή γνώμη εξαιτίας της τεράστιας αξίας ορισμένων νέων αποτελεσμάτων.

Η Γηραιά Αλβιώνια διεκδικεί την χρονική πρωτιά στην πρακτική εφαρμογή μεθόδων βελτιστοποίησης σε πολεμικές επιχειρήσεις. Λίγο μετά την έναρξη του πολέμου και προ της απειλούμενης απόβασης των Γερμανών στην Αγγλία, συγκροτήθηκε μια επιστημονική επιτροπή σκοπός της οποίας ήταν ο εντοπισμός των θέσεων στις οποίες θα έπρεπε να εγκατασταθούν μονάδες ραντάρ, ώστε να μεγιστοποιηθεί η πιθανότητα αποτελεσματικής αντιμετώπισης των γερμανικών βομβαρδιστικών.

Στην τότε Σοβιετική Ένωση αθόρυβα αλλά μεθοδικά δούλευε ο εξάιρετος μαθηματικός L.V. Kantorovich. Μεταξύ άλλων αντικείμενο μελέτης του ήταν η θεωρητική επίλυση ενός προβλήματος, το οποίο αργότερα έγινε γνωστό με το

όνομα ενός σημαντικού επιστήμονα της άλλης μεγάλης υπερδύναμης (Η.Π.Α), του F.L. Hitchcock. Οι σημαντικές ανακαλύψεις του Kantorovich [1939] στο μετέπειτα ονομασθέν πρόβλημα μεταφοράς του Hitchcock επηρέασαν σε μεγάλο βαθμό την Βασιλική Ακαδημία Σουηδίας ώστε να του απονεμίσει μαζί με τον T.C. Koopmans, τον άλλον εξίσου σημαντικό μελετητή των εφαρμογών του Γραμμικού Προγραμματισμού, το βραβείο Νόμπελ Οικονομίας το 1975. Ο T.C. Koopmans είναι μάλιστα αυτός που πρώτος πρότεινε το όνομα **Γραμμικός Προγραμματισμός**.

Οι σημαντικότερες όμως ανακαλύψεις έγιναν στην πέραν του Ατλαντικού Υπερδύναμη. Κατά την διάρκεια του Β' Παγκοσμίου Πολέμου και της επικείμενης απόβασης στην Ευρώπη, ζητήθηκε από τις Ένοπλες Δυνάμεις των Η.Π.Α. η βοήθεια των επιστημόνων του τότε RAND CORPORATION. Πολλοί επιστήμονες μεταξύ των οποίων και ο νεαρός τότε διδάκτωρ του μαθηματικού τμήματος του πανεπιστημίου του Σικάγο, **George B. Dantzig**, κλήθηκαν για την επίλυση των κρίσιμων εκείνων προβλημάτων.

Λίγο μετά την λήξη του πολέμου, το 1947, ο George B. Dantzig συνέλαβε την ιδέα της ανάπτυξης μιας μεθόδου επίλυσης του πιο απλού από μαθηματικής απόψεως προβλήματος βελτιστοποίησης, του γραμμικού προβλήματος. Εκείνη τη στιγμή γεννήθηκε ο αλγόριθμος **Simplex**, ίσως το πιο σημαντικό αποτέλεσμα της Επιχειρησιακής Έρευνας. Η πρώτη παρουσίαση του νέου αποτελέσματος δεν έπεισε για την αποτελεσματικότητά του. Η χρήση της γεωμετρίας του χώρου των μεταβλητών για την περιγραφή του ήταν ατυχής επιλογή. Δεν έγινε όμως το ίδιο στην δεύτερη παρουσίαση όταν επιλέχτηκε η γεωμετρία του χώρου των περιορισμών. Ο αλγόριθμος δημοσιεύτηκε για πρώτη φορά το 1949.

Μετά την πρώτη δημοσίευση του σημαντικού αυτού αποτελέσματος οι δρόμοι για την ανάπτυξη μιας νέας επιστήμης άνοιξαν διάπλατα. Η βιβλιογραφία των δεκαετιών του '50 και '60 βρίθει από δημοσιεύσεις σημαντικότερων αποτελεσμάτων συνοδευόμενων πολλές φορές με υπολογιστικά αποτελέσματα, τα οποία αποδείκνυαν την πρακτική αποτελεσματικότητα του αλγορίθμου Simplex. Όμως, ο ρυθμός αύξησης του μεγέθους των πρακτικών προβλημάτων ήταν μεγαλύτερος όχι μόνο του ρυθμού αύξησης της ταχύτητας των ηλεκτρονικών υπολογιστών, αλλά και του ρυθμού παραγωγής νέων σημαντικών αποτελεσμάτων. Ο αλγόριθμος

Simplex άρχισε να αντιμετωπίζει προβλήματα χρόνου στην επίλυση των μεγάλων προβλημάτων. Η πρώτη ευφορία για την πρακτική αποτελεσματικότητα άρχισε σιγά-σιγά να σβήνει. Το σημαντικότερο κτύπημα ήρθε το 1972 όταν οι μαθηματικοί Klee και Minty απάντησαν αρνητικά στο πιο σημαντικό ανοικτό πρόβλημα της εποχής εκείνης, του χαρακτηρισμού της υπολογιστικής πολυπλοκότητας της χειρότερης περίπτωσης (worst case computational complexity). Απέδειξαν ότι ο αλγόριθμος *Simplex* είναι από θεωρητική άποψη το χειρότερο που μπορούσε να συμβεί, ένας εκθετικός και όχι πολυωνυμικός αλγόριθμος.

Το αποτέλεσμα των Klee και Minty έστρεψε την έρευνα σε άλλες κατευθύνσεις. Ιδιαίτερα η ανακάλυψη πολυωνυμικού αλγορίθμου για το γραμμικό πρόβλημα έγινε το νέο διάσημο ανοικτό πρόβλημα, το οποίο ταλαιπώρησε την επιστημονική κοινότητα περίπου 10 χρόνια. Το 1979 ο νεαρός τότε Ρώσος επιστήμονας **L. Khachian** τροποποίησε έναν αλγόριθμο, ο οποίος είχε ανατηχθεί από τον N.Z. Shor και απέδειξε ότι η τροποποίηση του είναι πολυωνυμική. Ήταν ένα από τα πιο σημαντικά αποτελέσματα όχι μόνο του μαθηματικού προγραμματισμού αλλά της θεωρητικής πληροφορικής και των μαθηματικών γενικότερα. Πολλές παγκοσμίου κυκλοφορίας ημερήσιες εφημερίδες γενικού ενδιαφέροντος αναφέρθηκαν στο σημαντικό αυτό γεγονός πολλές φορές μάλιστα υπερβάλλουσες περί της πρακτικής αποτελεσματικότητάς του.

Ο νέος αλγόριθμος, ο **ελλειψοειδής αλγόριθμος** (ellipsoid algorithm) ή ρωσικός αλγόριθμος, ο πρώτος πολυωνυμικός αλγόριθμος για το γραμμικό πρόβλημα, επιφύλαξε μια απίστευτη έκπληξη στους ερευνητές. Παρ' ότι ήταν ότι καλύτερο από θεωρητικής απόψεως έμελλε να μείνει γνωστός στην ιστορία και για την μνημειώδη πρακτική αναποτελεσματικότητά του. Οι συγκρίσεις με τον αλγόριθμο Simplex ήταν απογοητευτικές και η αιτιολόγηση ήταν σχεδόν προφανής. Δούλευε πάντοτε στη χειρότερη περίπτωση του. Παρά την απογοητευτική στην πράξη εφαρμογή του, ο ελλειψοειδής αλγόριθμος παραμένει ένα σημαντικότερο θεωρητικό αποτέλεσμα, λόγω της τεράστιας συνεισφοράς του στην πρόοδο της επιστήμης.

Κάτω από το φως των νέων ανακαλύψεων οι επιστήμονες στράφηκαν στην ανακάλυψη πολυωνυμικών αλγορίθμων, οι οποίοι όμως έπρεπε να είναι και αποτελεσματικοί στην πράξη. Ο αλγόριθμος με αυτά τα πολύ επιθυμητά

χαρακτηριστικά ανακαλύφθηκε από τον επίσης νεαρό επιστήμονα Επιχειρησιακής Έρευνας και Πληροφορικής (Master in Operation Research, Stanford University και Ph.D in Computer Science, University of Berkley) **N. Karmarkar**. Το Δεκέμβριο του 1983 στο συνέδριο της IEEE, ο επιστήμονας των Bell Labs ανακοινώνει ότι ο πολυωνυμικός **αλγόριθμος εσωτερικών σημείων**, που είχε κατασκευάσει, είναι περίπου 50 φορές ταχύτερος του αλγορίθμου Simplex. Ο αλγόριθμος δημοσιεύτηκε ένα χρόνο μετά. Μετέπειτα υπολογιστικές μελέτες απέδειξαν ότι οι ισχυρισμοί του Karmarkar ήταν διογκωμένοι. Όμως άλλοι αλγόριθμοι εσωτερικών σημείων που αναπτύχθηκαν αργότερα, φαίνεται ότι είναι ταχύτεροι του αλγορίθμου Simplex σε προβλήματα μεγάλης διάστασης. Τα Bell Labs υλοποίησαν τον νέο αλγόριθμο στο γνωστό και πανάκριβο για την εποχή του (\$9.999.999 το 1984) λύτη KORBX. Είναι γνωστό ότι ελάχιστα αντίτυπα αυτού του λογισμικού αγοράστηκαν (Us Navy, Us Air Force, Us Delta Airlines).

Ένας άλλος σημαντικός σταθμός στην πορεία του γραμμικού προγραμματισμού είναι η εξήγηση της παράξενης συμπεριφοράς του αλγορίθμου Simplex (αποτελεσματικός στην πράξη, αλλά ότι χειρότερο από θεωρητική άποψη) από τον **Borgwardt**, ο οποίος το 1982 απέδειξε ότι *η μέση υπολογιστική συμπεριφορά των κανόνων περιστροφής του αλγορίθμου Simplex είναι πολυωνυμική.*

Τέλος, πρέπει να μνημονευθεί, ότι ο «πατέρας» του γραμμικού προγραμματισμού, ο G.B. Dantzig δεν τιμήθηκε με βραβείο Νόμπελ. Η Ακαδημία Επιστημών της Σουηδίας θεώρησε πιθανώς ότι το έργο του είναι πολύ μαθηματικό για να τιμηθεί με βραβείο Νόμπελ Οικονομίας και Νόμπελ μαθηματικών δεν απονέμεται. Ένα χρόνο μετά την απονομή του βραβείου Νόμπελ στους T.C. Koopmans και L.V. Kantorovich η Ακαδημία Επιστημών των Η.Π.Α. απένειμε το 1976 το Εθνικό Μετάλλιο Επιστημών στον G.B. Dantzig ως επιβράβευση για την επινόηση του Γραμμικού Προγραμματισμού.

2. ΘΕΩΡΗΤΙΚΗ ΘΕΜΕΛΙΩΣΗ ΤΗΣ ΜΕΘΟΔΟΥ SIMPLEX

Τα μοντέλα γραμμικού προγραμματισμού εμφανίζουν μία ποικιλία μορφών τόσο ως προς το είδος της βελτιστοποίησης (μέγιστο ή ελάχιστο) όσο και ως προς τον τύπο των περιορισμών (\leq ή \geq ή $=$). Ωστόσο με τη βοήθεια στοιχειωδών μετασχηματισμών οι μορφές αυτές είναι μετατρέψιμες (βλ. Μπότσαρης 2002, ενότητα 1.2) σε μια **κανονική**, όπως λέγεται, μορφή, η οποία διευκολύνει στη θεωρητική θεμελίωση της τεχνικής Simplex. Η κανονική μορφή ενός προβλήματος γραμμικού προγραμματισμού συνίσταται στη **μεγιστοποίηση** μιας γραμμικής συνάρτησης **μη αρνητικών** μεταβλητών, οι οποίες απαιτείται να ικανοποιούν ένα **σύστημα γραμμικών ανισώσεων του τύπου \leq** . Έτσι, στα επόμενα μπορούμε να υποθέτουμε ότι το προς επίλυση πρόβλημα είναι

$$\begin{array}{l} \text{να μεγιστοποιηθεί η} \\ \text{με τους περιορισμούς} \end{array} \quad \begin{array}{l} z = \mathbf{c}^t \mathbf{x} \\ \mathbf{A} \mathbf{x} \leq \mathbf{b}, \mathbf{x} \geq \mathbf{0}, \mathbf{A} \in \mathcal{R}^{m \times n}, \mathbf{x} \in \mathcal{R}^n, \mathbf{b} \in \mathcal{R}^n \end{array} \quad (2.1)$$

όπου

$$\mathbf{A} = \begin{bmatrix} \alpha_{11} & \cdots & \alpha_{1j} & \cdots & \alpha_{1n} \\ \vdots & & \vdots & & \vdots \\ \alpha_{i1} & \cdots & \alpha_{ij} & \cdots & \alpha_{in} \\ \vdots & & \vdots & & \vdots \\ \alpha_{m1} & \cdots & \alpha_{mj} & \cdots & \alpha_{mn} \end{bmatrix}, \mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_j \\ \vdots \\ x_n \end{bmatrix}, \mathbf{c} = \begin{bmatrix} c_1 \\ \vdots \\ c_j \\ \vdots \\ c_n \end{bmatrix}, \mathbf{b} = \begin{bmatrix} b_1 \\ \vdots \\ b_i \\ \vdots \\ b_m \end{bmatrix}$$

Στις πιο πάνω σχέσεις με $\mathbf{0}$ παριστάνεται το μηδενικό διάνυσμα, ενώ η διανυσματική ανισότητα $\mathbf{u} \geq \mathbf{w}$, όπου \mathbf{u} και \mathbf{w} είναι δύο διανύσματα του \mathcal{R}^n , υποδηλώνει ότι μεταξύ των αντίστοιχων συνιστωσών τους ισχύει η συγκεκριμένη ανισοτική σχέση.

Η γραφική απεικόνιση του συνόλου των εφικτών λύσεων ενός τυχαίου προβλήματος γραμμικού προγραμματισμού με δύο μεταβλητές (όπως φαίνεται στην εικόνα που ακολουθεί) δείχνει ότι το σύνολο αυτό είναι κυρτό.

Στο επόμενο θεώρημα αποδεικνύεται η κυρτότητα του εφικτού συνόλου ενός γενικού προβλήματος γραμμικού προγραμματισμού.

Θεώρημα 2.1

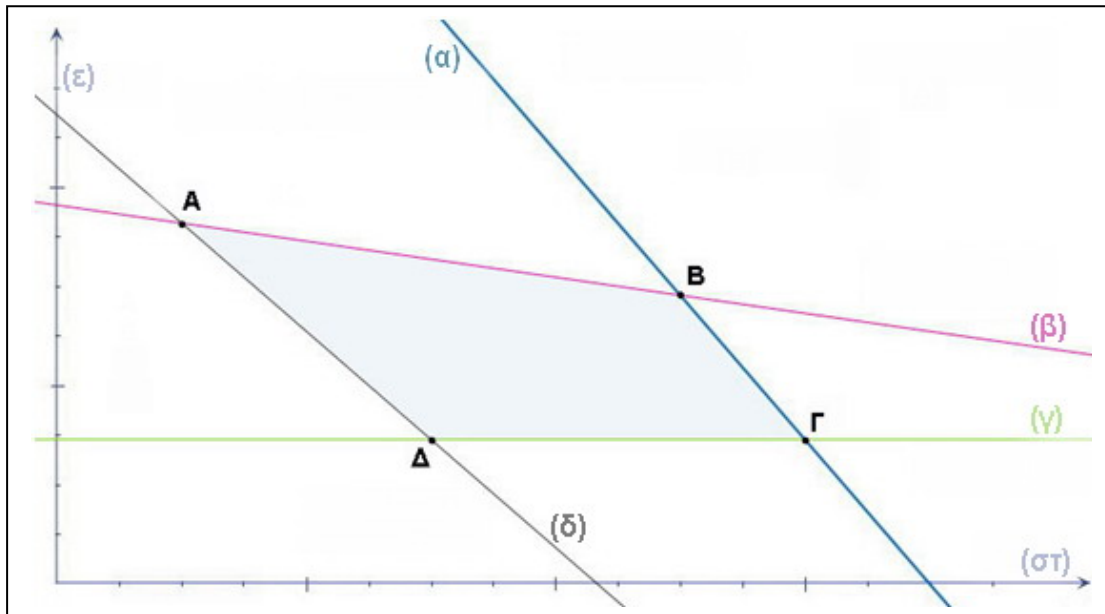
Το σύνολο των εφικτών λύσεων ενός προβλήματος γραμμικού προγραμματισμού, όταν δεν είναι κενό, είναι κυρτό.

Απόδειξη Έστω C το σύνολο των εφικτών λύσεων του προβλήματος (2.1) και \bar{x} και \hat{x} δύο σημεία του συνόλου αυτού. Αφού $\bar{x}, \hat{x} \in C$, θα είναι $A\bar{x} \leq b, \bar{x} \geq 0$ και $A\hat{x} \leq b, \hat{x} \geq 0$. Για οποιονδήποτε κυρτό συνδυασμό x των \bar{x} και \hat{x} , όπου $x = \lambda\bar{x} + (1-\lambda)\hat{x}$, $0 \leq \lambda \leq 1$, έχουμε $Ax = A[\lambda\bar{x} + (1-\lambda)\hat{x}] = \lambda A\bar{x} + (1-\lambda)A\hat{x}$. Δεδομένου ότι $\bar{x} \in C, \hat{x} \in C, \lambda \geq 0, 1-\lambda \geq 0$, προκύπτει $Ax \leq \lambda b + (1-\lambda)b = b$. Εξάλλου, επειδή $\bar{x} \geq 0, \hat{x} \geq 0, \lambda \geq 0$ και $1-\lambda \geq 0$, θα είναι και $x \geq 0$. Έτσι, το σημείο x ικανοποιεί όλους τους περιορισμούς του προβλήματος (2.1) και επομένως ανήκει στο εφικτό του σύνολο C . Άρα το σύνολο C είναι κυρτό. ■

Γνωρίζουμε ότι, κατά τη γεωμετρική απεικόνιση του εφικτού συνόλου ενός προβλήματος γραμμικού προγραμματισμού με δύο μεταβλητές, οι ακμές του συνόλου αυτού είναι κάποιες από τις ευθείες κατά μήκος των οποίων οι ανισοτικοί περιορισμοί του προβλήματος, συμπεριλαμβανομένων των περιορισμών μη αρνητικότητας των μεταβλητών του, ισχύουν οριακά ως εξισώσεις. Κάθε ζεύγος των ευθειών αυτών, του οποίου το σημείο τομής υπάρχει και είναι εφικτό, ορίζει μια κορυφή του εφικτού συνόλου.

Για παράδειγμα, το πρόβλημα του οποίου το εφικτό σύνολο απεικονίζεται στο παρακάτω σχήμα περιλαμβάνει δύο μη αρνητικές μεταβλητές και τέσσερις ανισοτικούς περιορισμούς, εισάγοντας έτσι στο επίπεδο έξι συνολικά ευθείες με εξισώσεις $2x_1 + x_2 = 14$ (α), $x_1 + 4x_2 = 21$ (β), $x_2 = 2$ (γ), $3x_1 + 2x_2 = 13$ (δ), $x_1 = 0$ (ε), $x_2 = 0$ (στ). Οι έξι αυτές ευθείες δίνουν τόσα ζεύγη ευθειών όσοι είναι οι συνδυασμοί των $2 + 4 = 6$ αντικειμένων ανά 2. Επομένως τα δυνατά ζεύγη

$$\text{ευθειών είναι } \binom{4+2}{2} = \frac{6!}{4!2!} = 15.$$



Σχήμα 2.1. Απεικόνιση εφικτού συνόλου

Ωστόσο, οι κορυφές του εφικτού συνόλου είναι μόνο τέσσερις: το σημείο τομής A(1,5) του ζεύγους $(x_1+4x_2=21, 3x_1+2x_2=13)$, το σημείο τομής B(5,4) του ζεύγους $(x_1+4x_2=21, 2x_1+x_2=14)$, το σημείο τομής Γ(6,2) του ζεύγους $(x_2=2, 2x_1+x_2=14)$ καθώς και το σημείο τομής Δ(3,2) του ζεύγους $(3x_1+2x_2=13, x_2=2)$. Τα υπόλοιπα έντεκα ζεύγη δεν ορίζουν κάποια κορυφή, αφού οι ευθείες τους είτε δεν έχουν σημείο τομής είτε τέμνονται εκτός του εφικτού συνόλου. Για παράδειγμα, το ζεύγος των ευθειών $x_1 = 0$ και $x_2 = 0$ μας φέρνει στην αρχή των αξόνων, η οποία βρίσκεται εκτός εφικτού συνόλου, ενώ οι ευθείες $x_2 = 2$ και $x_2 = 0$, ως παράλληλες, δεν τέμνονται πουθενά.

Μπορούμε τώρα να γενικεύσουμε τις έννοιες αυτές σε n διαστάσεις. Έτσι, κάθε εξίσωση της μορφής $a_{i1}x_1 + \dots + a_{ij}x_j + \dots + a_{in}x_n = b_i$ περιγράφει ένα $(n-1)$ -διάστατο επίπεδο στο χώρο \mathbb{R}^n . Η τομή k ανεξάρτητων επιπέδων στο χώρο \mathbb{R}^n είναι ένα $(n-k)$ -διάστατο επίπεδο. Η τομή $n-1$ επιπέδων γίνεται επομένως κατά μήκος μιας ευθείας. Η τομή n ανεξάρτητων επιπέδων είναι ένα σημείο. Έτσι, μια **κορυφή** ορίζεται από την τομή n επιπέδων. Αν παραλείψουμε ένα από τα n τεμνόμενα επίπεδα, τα εναπομένοντα $n-1$ επίπεδα τέμνονται πάνω σε μια **ακμή**. Άρα από κάθε κορυφή διέρχονται n ακμές.

Για ένα πρόβλημα γραμμικού προγραμματισμού της κανονικής μορφής (2.1), οι n περιορισμοί μη αρνητικότητας $x \geq 0$ και οι m ανισοτικοί περιορισμοί $Ax \leq b$ εισάγουν στο χώρο \mathbb{R}^n $n+m$ επίπεδα, τα οποία περιγράφονται από τις

εξισώσεις $x_j = 0$ ($j=1, \dots, n$) και $a_{i1}x_1 + \dots + a_{ij}x_j + \dots + a_{in}x_n = b_i$ ($i=1, \dots, m$). Η απαίτηση μη αρνητικότητας των μεταβλητών απόφασης ορίζει n ημιχώρους, έναν για καθεμιά από τις n συνιστώσες του διανύσματος \mathbf{x} , και μεταφέρει το πρόβλημα στο θετικό μόνιο, αφού μόνο τα σημεία του μορίου αυτού έχουν όλες τις συντεταγμένες τους μη αρνητικές, δηλαδή $x_j \geq 0$ ($j=1, \dots, n$). Οι m ανισοτικοί περιορισμοί $\mathbf{A} \mathbf{x} \leq \mathbf{b}$ ορίζουν m πρόσθετους ημιχώρους. Η τομή αυτών των $n+m$ ημιχώρων αποτελεί το εφικτό σύνολο του προβλήματος, το οποίο, όταν δεν είναι κενό, είναι ένα κυρτό **πολύεδρο** με επίπεδες έδρες. Όταν το πολύεδρο των εφικτών λύσεων είναι φραγμένο, ονομάζεται **πολύτοπο**.

Κάθε κορυφή του εφικτού συνόλου ορίζεται ως η τομή n από το σύνολο των $n + m$ επιπέδων, τα οποία περιγράφονται από τις εξισώσεις $x_j = 0$ ($j = 1, \dots, n$) και $a_{i1}x_1 + \dots + a_{ij}x_j + \dots + a_{in}x_n = b_i$ ($i = 1, \dots, m$). Εδώ έχουμε έμμεσα υποθέσει ότι οποιαδήποτε n από αυτά τα $n+m$ επίπεδα είναι γραμμικώς ανεξάρτητα. Έτσι, οι κορυφές του εφικτού συνόλου προσδιορίζονται αν θεωρήσουμε n από τις $n+m$ ανισώσεις του προβλήματος ως εξισώσεις και λύσουμε ως προς $x_1, \dots, x_j, \dots, x_n$ το $n \times n$ γραμμικό σύστημα που προκύπτει, υποθέτοντας πάντοτε ότι το σύστημα αυτό έχει μοναδική λύση. Εννοείται ότι η λύση αυτή, για να είναι κορυφή, πρέπει να ικανοποιεί και τους υπόλοιπους m περιορισμούς.

Η παρατήρηση αυτή επιτρέπει να μεταφέρουμε την ανάλυση από το χώρο της γεωμετρίας στο χώρο της γραμμικής άλγεβρας. Πράγματι, ας υποθέσουμε ότι στο αριστερό μέλος της ανίσωσης $a_{i1}x_1 + \dots + a_{ij}x_j + \dots + a_{in}x_n \leq b_i$ ($i = 1, \dots, m$) προστίθεται μια μη αρνητική μεταβλητή s_i η οποία είναι τέτοια ώστε $a_{i1}x_1 + \dots + a_{ij}x_j + \dots + a_{in}x_n + s_i = b_i$. Μετά την εισαγωγή των m μη αρνητικών μεταβλητών περιθωρίου s_i ($i = 1, \dots, m$) στο αριστερό μέλος των ανισώσεων $\mathbf{A} \mathbf{x} \leq \mathbf{b}$, προκύπτει το σύστημα των γραμμικών εξισώσεων

$$\begin{array}{ccccccc}
 a_{11}x_1 & + \dots + & a_{1j}x_j & + \dots + & a_{1n}x_n & + s_1 & = b_1 \\
 \vdots & & \vdots & & \vdots & \ddots & \vdots \\
 a_{i1}x_1 & + \dots + & a_{ij}x_j & + \dots + & a_{in}x_n & + s_i & = b_i \\
 \vdots & & \vdots & & \vdots & \ddots & \vdots \\
 a_{m1}x_1 & + \dots + & a_{mj}x_j & + \dots + & a_{mn}x_n & + s_m & = b_m
 \end{array}$$

όπου

$$\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_j \\ \vdots \\ x_n \end{bmatrix}, \quad \mathbf{s} = \begin{bmatrix} s_1 \\ \vdots \\ s_i \\ \vdots \\ s_m \end{bmatrix} \quad (2.2)$$

Αυτό είναι ένα υποπροσδιορισμένο σύστημα m εξισώσεων με $n + m$ μεταβλητές, το οποίο σε διανυσματική μορφή γράφεται

$$[\mathbf{A} : \mathbf{I}_m] \begin{bmatrix} \mathbf{x} \\ \mathbf{s} \end{bmatrix} = \mathbf{b}$$

όπου \mathbf{I}_m είναι ο ταυτοτικός πίνακας m -οστής τάξης.

Μπορούμε τώρα να δούμε ότι κάθε μη αρνητική λύση

$$x_j \geq 0 \quad (j = 1, \dots, n), \quad s_i \geq 0 \quad (i = 1, \dots, m) \quad (2.3)$$

του επαυξημένου συστήματος (2.2) ορίζει ένα εφικτό διάνυσμα \mathbf{x} για το πρόβλημα (2.1). Πράγματι, σε μια τέτοια περίπτωση το διάνυσμα \mathbf{x} έχει μη αρνητικές συνιστώσες, είναι δηλαδή $\mathbf{x} \geq \mathbf{0}$, ενώ η ανισότητα

$$s_i = b_i - (\alpha_{i1}x_1 + \dots + \alpha_{ij}x_j + \dots + \alpha_{in}x_n) \geq 0$$

συνεπάγεται ότι

$$\alpha_{i1}x_1 + \dots + \alpha_{ij}x_j + \dots + \alpha_{in}x_n \leq b_i, \quad (i = 1, \dots, m).$$

Έτσι, το σημείο \mathbf{x} ικανοποιεί και τους ανισοτικούς περιορισμούς $\mathbf{A} \mathbf{x} \leq \mathbf{b}$, είναι δηλαδή εφικτό για το πρόβλημα (2.1).

Πρέπει να σημειωθεί ότι, αν κάποια από τις μη αρνητικές μεταβλητές s_i , $i \in \{1, \dots, m\}$ έχει τιμή ίση με το μηδέν, τότε η αντίστοιχη εξίσωση

$$\alpha_{i1}x_1 + \dots + \alpha_{ij}x_j + \dots + \alpha_{in}x_n + s_i = b_i$$

γράφεται

$$\alpha_{i1}x_1 + \dots + \alpha_{ij}x_j + \dots + \alpha_{in}x_n = b_i$$

και επομένως το σημείο \mathbf{x} ανήκει στο επίπεδο, στο οποίο ο περιορισμός $\alpha_{i1}x_1 + \dots + \alpha_{ij}x_j + \dots + \alpha_{in}x_n \leq b_i$ ισχύει ως ισότητα. Έτσι η ισότητα $s_i = 0$ και η εξίσωση $\alpha_{i1}x_1 + \dots + \alpha_{ij}x_j + \dots + \alpha_{in}x_n + s_i = b_i$ περιγράφουν το ίδιο επίπεδο.

Μια λύση του συστήματος (2.2) προσδιορίζεται αν δώσουμε αυθαίρετες τιμές σε n από τις $n + m$ μεταβλητές του $x_1, \dots, x_j, \dots, x_n, s_1, \dots, s_i, \dots, s_m$ και λύσουμε το σύστημα ως προς εναπομένουσες m μεταβλητές. Ας υποθέσουμε ότι οι τιμές των n ελεύθερων μεταβλητών τίθενται όλες ίσες με το μηδέν και ότι το σύστημα που προκύπτει έχει μοναδική λύση. Όπως γνωρίζουμε, μια τέτοια λύση, στην οποία n από τις $n + m$ μεταβλητές του συστήματος (2.2) έχουν μηδενική τιμή, ονομάζεται **βασική**. Σε αντιστοιχία, οι m μεταβλητές ως προς τις οποίες λύνεται το σύστημα ονομάζονται βασικές. Στα επόμενα υποθέτουμε ότι, αν για κάποια επιλογή των n ελεύθερων μεταβλητών το σύστημα (2.2) έχει λύση ως προς τις εναπομένουσες m **βασικές** μεταβλητές, τότε η λύση αυτή είναι μοναδική.

Ας θεωρήσουμε τώρα το σημείο \mathbf{x} , το οποίο ορίζεται από μια μη αρνητική βασική λύση του συστήματος (2.2). Έστω $s_{i_k}, i_k \in \{1, \dots, m\}$ ($k = 1, \dots, r$) και $x_{j_l}, j_l \in \{1, \dots, n\}$ ($l = 1, \dots, q$) οι ελεύθερες μεταβλητές της λύσης αυτής, όπου $r \in \{1, \dots, \min(n, m)\}$ και $n = q + r$. Όπως δείξαμε, το σημείο \mathbf{x} είναι εφικτό για το πρόβλημα (2.1). Επιπλέον, αφού $s_k = 0$ ($k = 1, \dots, r$) το σημείο αυτό ανήκει στην τομή των επιπέδων $x_{j_l} = 0$ ($l = 1, \dots, q$) και $\alpha_{i_k 1}x_1 + \dots + \alpha_{i_k j}x_j + \dots + \alpha_{i_k n}x_n = b_{i_k}$ ($k = 1, \dots, r$). Επομένως, το σημείο \mathbf{x} είναι μια κορυφή του εφικτού συνόλου του προβλήματος (2.1), αφού αποτελεί την τομή n από το σύνολο $n + m$ επιπέδων, τα οποία εισάγουν στο χώρο οι m ανισοτικοί περιορισμοί και οι n περιορισμοί μη αρνητικότητας των μεταβλητών απόφασης του προβλήματος αυτού.

Θεώρημα 2.2

Για να είναι ένα σημείο \mathbf{x} κορυφή του συνόλου

$C = \{ \mathbf{x} \in \mathbb{R}^n \mid \mathbf{A} \mathbf{x} \leq \mathbf{b}, \mathbf{x} \geq \mathbf{0} \}$, πρέπει και αρκεί τα διανύσματα \mathbf{x} και \mathbf{s} να ορίζουν μια μη αρνητική βασική λύση του γραμμικού συστήματος

$$[\mathbf{A} \ ; \ \mathbf{I}_m] \begin{bmatrix} \mathbf{x} \\ \mathbf{s} \end{bmatrix} = \mathbf{b}, \mathbf{A} \in \mathbb{R}^{m \times n}, \mathbf{x} \in \mathbb{R}^n, \mathbf{s} \in \mathbb{R}^m, \mathbf{b} \in \mathbb{R}^m \quad (2.4)$$

Απόδειξη Κατ' αρχάς αποδεικνύεται η ικανή συνθήκη. Έστω λοιπόν $\begin{bmatrix} \mathbf{x} \\ \mathbf{s} \end{bmatrix}$ μια μη αρνητική βασική λύση του συστήματος (2.4). Έστω ακόμα ότι $x_{j_l}, j_l \in \{1, \dots, n\}$ ($l=1, \dots, q$) και $s_{i_k}, i_k \in \{1, \dots, m\}$ ($k=1, \dots, r$) οι ελεύθερες μεταβλητές της λύσης αυτής, των οποίων οι τιμές έχουν τεθεί ίσες με το μηδέν, όπου $q + r = n$. Αφού $\begin{bmatrix} \mathbf{x} \\ \mathbf{s} \end{bmatrix}$ είναι μια μη αρνητική λύση του συστήματος

$$(2.4), \text{ θα πρέπει να ισχύει } [\mathbf{A} : \mathbf{I}_m] \begin{bmatrix} \mathbf{x} \\ \mathbf{s} \end{bmatrix} = \mathbf{Ax} + \mathbf{s} = \mathbf{b}, \mathbf{x} \geq \mathbf{0} \text{ και } \mathbf{s} \geq \mathbf{0}.$$

Επομένως $\mathbf{Ax} = \mathbf{b} - \mathbf{s} \leq \mathbf{b}$. Άρα $\mathbf{x} \in C$.

Ας υποθέσουμε τώρα ότι το σημείο \mathbf{x} δεν είναι κορυφή του συνόλου C . Υπενθυμίζουμε ότι το σύνολο C είναι κυρτό. Τότε το σημείο \mathbf{x} θα πρέπει να αποτελεί κυρτό συνδυασμό δύο διαφορετικών σημείων $\bar{\mathbf{x}}$ και $\hat{\mathbf{x}}$ του C , δηλαδή θα πρέπει να υπάρχουν σημεία $\bar{\mathbf{x}}, \hat{\mathbf{x}} \in C$, όπου $\bar{\mathbf{x}} \neq \hat{\mathbf{x}}$, τέτοια ώστε $\mathbf{x} = \lambda \bar{\mathbf{x}} + (1-\lambda) \hat{\mathbf{x}}$ για κάποια τιμή του λ στο ανοικτό διάστημα $(0,1)$. Πιο αναλυτικά έχουμε

$$x_j = \lambda \bar{x}_j + (1-\lambda) \hat{x}_j \quad (j=1, \dots, n) \quad (2.5)$$

Εφόσον τα σημεία $\bar{\mathbf{x}}$ και $\hat{\mathbf{x}}$ ανήκουν στο εφικτό σύνολο C θα πρέπει να ικανοποιούν τους περιορισμούς του προβλήματος (2.1), θα πρέπει δηλαδή να είναι $\mathbf{A}\bar{\mathbf{x}} \leq \mathbf{b}$, $\bar{\mathbf{x}} \geq \mathbf{0}$ και $\mathbf{A}\hat{\mathbf{x}} \leq \mathbf{b}$, $\hat{\mathbf{x}} \geq \mathbf{0}$. Η παραπάνω σχέση (2.5) για

$j = j_1, \dots, j_l, \dots, j_q$ γράφεται

$$x_{j_l} = 0 = \lambda \bar{x}_{j_l} + (1-\lambda) \hat{x}_{j_l} \quad (l=1, \dots, q) \quad (2.6)$$

Επειδή οι δύο όροι του αθροίσματος στο δεξιό μέλος της σχέσης αυτής είναι μη αρνητικοί και έχουν άθροισμα ίσο με το μηδέν, καθένας από αυτούς θα πρέπει να είναι ίσος με το μηδέν, δηλαδή θα πρέπει $\lambda \bar{x}_{j_l} = 0$ και

(1-λ) $\hat{x}_j = 0$. Δεδομένου όμως ότι λ>0 και 1-λ>0, θα πρέπει να είναι $\bar{x}_j = 0$ και $\hat{x}_j = 0$ (j=1, ..., q).

Εξάλλου, επειδή τα σημεία \bar{x} και \hat{x} είναι εφικτά, για i=1, ..., m έχουμε $\alpha_{i1}\bar{x}_1 + \dots + \alpha_{ij}\bar{x}_j + \dots + \alpha_{in}\bar{x}_n \leq b_i$, $\alpha_{i1}\hat{x}_1 + \dots + \alpha_{ij}\hat{x}_j + \dots + \alpha_{in}\hat{x}_n \leq b_i$

Αν λοιπόν εισάγουμε τις μεταβλητές

$$\bar{s}_i = b_i - (\alpha_{i1}\bar{x}_1 + \dots + \alpha_{ij}\bar{x}_j + \dots + \alpha_{in}\bar{x}_n) \geq 0$$

$$\hat{s}_i = b_i - (\alpha_{i1}\hat{x}_1 + \dots + \alpha_{ij}\hat{x}_j + \dots + \alpha_{in}\hat{x}_n) \geq 0$$

Τότε $\begin{bmatrix} \bar{x} \\ \bar{s} \end{bmatrix}$ και $\begin{bmatrix} \hat{x} \\ \hat{s} \end{bmatrix}$ είναι δύο μη αρνητικές λύσεις του συστήματος (2.4) για τις

οποίες

$$\lambda (\alpha_{i1}\bar{x}_1 + \dots + \alpha_{ij}\bar{x}_j + \dots + \alpha_{in}\bar{x}_n) + \lambda \bar{s}_i = \lambda b_i$$

$$(1-\lambda) (\alpha_{i1}\hat{x}_1 + \dots + \alpha_{ij}\hat{x}_j + \dots + \alpha_{in}\hat{x}_n) + (1-\lambda) \hat{s}_i = (1-\lambda) b_i$$

Προσθέτοντας τις δύο αυτές ισότητες κατά μέλη προκύπτει

$$\alpha_{i1}[\lambda\bar{x}_1 + (1-\lambda)\hat{x}_1] + \dots + \alpha_{ij}[\lambda\bar{x}_j + (1-\lambda)\hat{x}_j] + \dots + \alpha_{in}[\lambda\bar{x}_n + (1-\lambda)\hat{x}_n] + \lambda\bar{s}_i + (1-\lambda)\hat{s}_i = b_i$$

Λαμβάνοντας υπόψη και τη σχέση (2.5), έχουμε

$$\alpha_{i1}x_1 + \dots + \alpha_{ij}x_j + \dots + \alpha_{in}x_n + \lambda\bar{s}_i + (1-\lambda)\hat{s}_i = b_i \quad (2.7)$$

Αναγνωρίζοντας επιπλέον ότι $s_k = 0$, $i_k \in \{1, \dots, m\}$ (k=1, ..., r) και επομένως ότι

$$\alpha_{i_k 1}x_1 + \dots + \alpha_{i_k j}x_j + \dots + \alpha_{i_k n}x_n = b_{i_k} \quad (k=1, \dots, r)$$

η σχέση (2.7) για $i = i_1, \dots, i_k, \dots, i_r$ γράφεται $b_{i_k} + \lambda\bar{s}_{i_k} + (1-\lambda)\hat{s}_{i_k} = b_{i_k}$, από όπου

$\lambda \bar{s}_k + (1-\lambda) \hat{s}_k = 0$. Δεδομένου ότι $\bar{\mathbf{s}} \geq \mathbf{0}$, $\hat{\mathbf{s}} \geq \mathbf{0}$, $\lambda > 0$, $1-\lambda > 0$, είναι φανερό ότι για $k = 1, \dots, r$ πρέπει να ισχύει $\bar{s}_k = 0$ και $\hat{s}_k = 0$.

Έτσι λοιπόν οι λύσεις $\begin{bmatrix} \mathbf{x} \\ \mathbf{s} \end{bmatrix}$, $\begin{bmatrix} \bar{\mathbf{x}} \\ \bar{\mathbf{s}} \end{bmatrix}$ και $\begin{bmatrix} \hat{\mathbf{x}} \\ \hat{\mathbf{s}} \end{bmatrix}$ αντιστοιχούν στις ίδιες ελεύθερες

μεταβλητές, αφού $s_i = \bar{s}_i = \hat{s}_i = 0$ ($i = i_1, \dots, i_k, \dots, i_r$) και $x_j = \bar{x}_j = \hat{x}_j = 0$ ($j = j_1, \dots, j_l, \dots, j_q$). Επομένως, και υπό την προϋπόθεση ότι η λύση την οποία επιδέχεται το σύστημα (2.4) για κάποια επιλογή των n ελεύθερων μεταβλητών είναι μοναδική, οι τρεις αυτές λύσεις συμπίπτουν, είναι δηλαδή $\mathbf{x} = \bar{\mathbf{x}} = \hat{\mathbf{x}}$, ισότητα που έρχεται σε αντίφαση με τη σχέση (2.5) και το γεγονός ότι $\bar{\mathbf{x}} \neq \hat{\mathbf{x}}$ και $\lambda \in (0,1)$. Άρα το σημείο \mathbf{x} δεν είναι δυνατόν να εκφραστεί ως κυρτός συνδυασμός δύο διαφορετικών σημείων του C και επομένως αποτελεί μια κορυφή του εφικτού συνόλου.

Στη συνέχεια αποδεικνύεται η αναγκαία συνθήκη. Έστω

$$\mathbf{V} = [\mathbf{A} \mid \mathbf{I}_m], \quad \mathbf{V} \in \mathfrak{R}^{m \times (n+m)}$$

ο πίνακας που προκύπτει μετά την προέκταση του \mathbf{A} με τον ταυτοτικό πίνακα

m -οστής τάξης. Έστω, ακόμα, $\boldsymbol{\omega} = \begin{bmatrix} \mathbf{x} \\ \mathbf{s} \end{bmatrix}$, $\boldsymbol{\omega} \in \mathfrak{R}^{n+m}$, το διάνυσμα που

προκύπτει μετά την προέκταση του \mathbf{x} κατά \mathbf{s} . Αν \mathbf{v}_j είναι οι στήλες του πίνακα \mathbf{V} και ω_j είναι οι συνιστώσες του $\boldsymbol{\omega}$ ($j = 1, \dots, n+m$), τότε το σύστημα (2.4) γράφεται

$$\sum_{j=1}^{n+m} \omega_j \mathbf{v}_j = \mathbf{b} \quad (2.8)$$

Ας θεωρήσουμε τώρα μια κορυφή \mathbf{x} του εφικτού συνόλου C . Το σημείο \mathbf{x} ως εφικτό πρέπει να ικανοποιεί τις ανισώσεις $\mathbf{x} \geq \mathbf{0}$ και $\mathbf{Ax} \leq \mathbf{b}$. Επομένως η σχέση $\mathbf{s} = \mathbf{b} - \mathbf{Ax}$ ορίζει ένα μη αρνητικό διάνυσμα. Έτσι, το διάνυσμα $\boldsymbol{\omega}$ αποτελεί μια μη αρνητική λύση του συστήματος (2.4) ή του ισοδύναμου συστήματος (2.8). Έστω ω_{j_k} , $j_k \in \{1, \dots, n+m\}$ ($k = 1, \dots, r$) οι θετικές συνιστώσες του διανύσματος $\boldsymbol{\omega}$. Τότε το σύστημα (2.8) γράφεται

$$\sum_{k=1}^r \omega_{j_k} \mathbf{v}_{j_k} = \mathbf{b}, \omega_{j_k} > 0 \quad (k = 1, \dots, r) \quad (2.9)$$

αφού όλες οι άλλες συνιστώσες του ω είναι μηδενικές και επομένως δεν υπεισέρχονται στο ανάπτυγμα.

Ας υποθέσουμε τώρα ότι τα διανύσματα-στήλες του πίνακα \mathbf{V} που αντιστοιχούν στα r θετικά στοιχεία του διανύσματος ω είναι γραμμικά εξαρτημένα. Τότε θα πρέπει να υπάρχουν πραγματικοί αριθμοί λ_k ($k = 1, \dots, r$), όχι όλοι ίσοι με το μηδέν, τέτοιοι ώστε

$$\sum_{k=1}^r \lambda_k \mathbf{v}_{j_k} = \mathbf{0}, \quad \sum_{k=1}^r |\lambda_k| \neq 0 \quad (2.10)$$

Από τις σχέσεις (2.9) και (2.10), για οποιονδήποτε θετικό αριθμό μ , προκύπτει

$$\sum_{k=1}^r \omega_{j_k} \mathbf{v}_{j_k} \pm \mu \sum_{k=1}^r \lambda_k \mathbf{v}_{j_k} = \mathbf{b} \Rightarrow \sum_{k=1}^r (\omega_{j_k} \pm \mu \lambda_k) \mathbf{v}_{j_k} = \mathbf{b} \quad (2.11)$$

Στη βάση της τελευταίας αυτής σχέσης θεωρούμε τα διανύσματα $\bar{\omega}$ και $\hat{\omega}$ με συνιστώσες

$$\bar{\omega}_{j_k} = \omega_{j_k} + \mu \lambda_k \quad (k = 1, \dots, r), \quad \bar{\omega}_j = 0 \quad \forall j \notin \{j_1, \dots, j_r\}$$

$$\hat{\omega}_{j_k} = \omega_{j_k} - \mu \lambda_k \quad (k = 1, \dots, r), \quad \hat{\omega}_j = 0 \quad \forall j \notin \{j_1, \dots, j_r\}$$

Δεδομένου ότι $\omega_{j_k} > 0$ και οι αριθμοί λ_k δεν είναι όλοι ίσοι με το μηδέν,

μπορούμε να εξασφαλίσουμε ότι για $k = 1, \dots, r$ θα ισχύει $\bar{\omega}_{j_k} > 0$ και $\hat{\omega}_{j_k} > 0$, επιλέγοντας την τιμή της παραμέτρου μ στο διάστημα

$$0 < \mu < \min_{k=1, \dots, r} \{ \omega_{j_k} / |\lambda_k|, \lambda_k \neq 0 \}$$

Με αυτή την επιλογή της παραμέτρου μ είναι φανερό ότι τα διανύσματα $\bar{\omega}$ και $\hat{\omega}$ αποτελούν δύο διαφορετικές μη αρνητικές λύσεις του συστήματος (2.8).

Συνεπώς, τα διανύσματα $\bar{\mathbf{x}}$ και $\hat{\mathbf{x}}$ που ορίζουν οι n πρώτες συνιστώσες των $\bar{\omega}$ και $\hat{\omega}$, αντιστοίχως, ικανοποιούν τις ανισώσεις $\mathbf{A} \mathbf{x} \leq \mathbf{b}$, $\mathbf{x} \geq \mathbf{0}$ και επομένως ανήκουν στο εφικτό σύνολο C .

Ας παρατηρήσουμε τώρα ότι από τις σχέσεις (2.12) για $j = 1, \dots, n+m$ προκύπτει

$$\omega_j = \frac{1}{2} \bar{\omega}_j + \frac{1}{2} \hat{\omega}_j \Rightarrow \mathbf{x} = \frac{1}{2} \bar{\mathbf{x}} + \frac{1}{2} \hat{\mathbf{x}}$$

σχέση που δείχνει ότι το σημείο \mathbf{x} εκφράζεται ως κυρτός συνδυασμός δύο διαφορετικών σημείων του εφικτού συνόλου. Αυτό έρχεται όμως σε αντίφαση με το γεγονός ότι το σημείο \mathbf{x} είναι κορυφή του συνόλου C και επομένως δεν θα έπρεπε να εκφράζεται ως κυρτός συνδυασμός δύο διαφορετικών σημείων του.

Στην αντίφαση αυτή καταλήξαμε διότι υποθέσαμε ότι τα διανύσματα-στήλες του πίνακα $[\mathbf{A} \mid \mathbf{I}_m]$ που αντιστοιχούν στα θετικά στοιχεία της λύσης $\begin{bmatrix} \mathbf{x} \\ \mathbf{s} \end{bmatrix}$,

όπου \mathbf{x} κορυφή του εφικτού συνόλου, είναι γραμμικά εξαρτημένα. Επομένως, τα διανύσματα αυτά πρέπει να είναι γραμμικά ανεξάρτητα. Ο αριθμός όμως των γραμμικά ανεξάρτητων στηλών του πίνακα $[\mathbf{A} \mid \mathbf{I}_m]$ δεν μπορεί να είναι μεγαλύτερος από m , αφού ο $[\mathbf{A} \mid \mathbf{I}_m]$ είναι ένας $m \times (n+m)$ πίνακας. Αυτό

σημαίνει ότι τα θετικά στοιχεία της λύσης $\begin{bmatrix} \mathbf{x} \\ \mathbf{s} \end{bmatrix}$ είναι το πολύ m . Άρα μια κορυφή του εφικτού συνόλου C αντιστοιχεί σε μια μη αρνητική λύση του συστήματος (2.4), στην οποία n τουλάχιστον από τις $n+m$ μεταβλητές $\begin{bmatrix} \mathbf{x} \\ \mathbf{s} \end{bmatrix}$

έχουν τιμή μηδενική. Έτσι η $\begin{bmatrix} \mathbf{x} \\ \mathbf{s} \end{bmatrix}$ είναι μια μη αρνητική βασική λύση του συστήματος (2.4).

Έτσι, ολοκληρώθηκε το θεώρημα που μας επιτρέπει να μεταφέρουμε το πρόβλημα από τη γλώσσα της γεωμετρίας και των κορυφών στη γλώσσα της γραμμικής άλγεβρας και των βασικών λύσεων (βλ. και Μπότσαρης 2002).

Σε ότι αφορά στην επιλογή των ελεύθερων μεταβλητών για την οποία το σύστημα των γραμμικών εξισώσεων έχει λύση ο $m \times m$ υποπίνακας του πίνακα $\mathbf{V} = [\mathbf{A} \mid \mathbf{I}_m]$, ο οποίος αντιστοιχεί στις βασικές μεταβλητές της επιλογής αυτής, είναι αντιστρέψιμος, με αποτέλεσμα οι τιμές των βασικών μεταβλητών να είναι μονοσήμαντα ορισμένες. Αυτό συμβαίνει όταν οι στήλες του πίνακα $\mathbf{V} = [\mathbf{A} \mid \mathbf{I}_m]$ οι οποίες αντιστοιχούν στις m βασικές μεταβλητές είναι γραμμικά ανεξάρτητες. Είναι όμως ενδεχόμενο, για κάποια επιλογή των ελεύθερων

μεταβλητών, το σύστημα των γραμμικών περιορισμών ενός προβλήματος γραμμικού προγραμματισμού να επιδέχεται απειρία μη αρνητικών λύσεων. Σε μια τέτοια περίπτωση οι στήλες του πίνακα $\mathbf{V} = [\mathbf{A} \mid \mathbf{I}_m]$ οι οποίες αντιστοιχούν στις m βασικές μεταβλητές της επιλογής αυτής είναι γραμμικά εξαρτημένες.

Ας υποθέσουμε τώρα ότι από αυτές τις m γραμμικά εξαρτημένες στήλες οι $r < m$ είναι γραμμικά ανεξάρτητες. Αν \mathbf{v}_{i_k} ($k = 1, 1, \dots, r$) είναι οι r γραμμικά ανεξάρτητες στήλες της συγκεκριμένης επιλογής, τότε το σύστημα (2.8)

εκφυλίζεται στο $\sum_{k=1}^r \omega_{i_k} \mathbf{v}_{i_k} = \mathbf{b}$. Δεδομένου ότι ο πίνακας $\mathbf{V} = [\mathbf{A} \mid \mathbf{I}_m]$ είναι

βαθμού m , αφού περιέχει έναν ταυτοτικό πίνακα $m \times m$, μπορούμε να επιλέξουμε $m-r$ επιπλέον διανύσματα \mathbf{v}_{j_l} ($l = 1, \dots, m-r$) από αυτόν, τέτοια ώστε τα m διανύσματα $\mathbf{v}_{i_1}, \dots, \mathbf{v}_{i_r}, \dots, \mathbf{v}_{i_r}, \mathbf{v}_{j_1}, \dots, \mathbf{v}_{j_l}, \dots, \mathbf{v}_{j_l}, \dots, \mathbf{v}_{j_{m-r}}$ να είναι γραμμικά ανεξάρτητα, όπου βέβαια $j_l \neq i_k$.

Θέτοντας τις τιμές των μεταβλητών ω_{j_l} που αντιστοιχούν στα διανύσματα \mathbf{v}_{j_l} ($l = 1, \dots, m-r$) ίσες με το μηδέν, προκύπτει το γραμμικό σύστημα

$$\sum_{k=1}^r \omega_{i_k} \mathbf{v}_{i_k} + \sum_{l=1}^{m-r} \omega_{j_l} \mathbf{v}_{j_l} = \mathbf{b}$$

το οποίο επιδέχεται μοναδική λύση ως προς τις m βασικές μεταβλητές ω_{i_k} ($k = 1, \dots, r$) και ω_{j_l} ($l = 1, \dots, m-r$). Στη λύση βέβαια αυτή, η οποία ορίζει μια κορυφή του εφικτού συνόλου, οι $m-r$ βασικές μεταβλητές ω_{j_l} έχουν τιμή ίση με το μηδέν. Αν τώρα υπάρχουν δύο ή περισσότεροι διαφορετικοί τρόποι επιλογής των διανυσμάτων \mathbf{v}_{j_l} ($l = 1, \dots, m-r$), τότε **στην ίδια κορυφή αντιστοιχούν περισσότερες από μία μη αρνητικές βασικές λύσεις**.

Μια μη αρνητική βασική λύση, υπό την οποία μία ή περισσότερες από τις βασικές μεταβλητές έχουν τιμή ίση με το μηδέν, ονομάζεται, όπως και η αντίστοιχη κορυφή, **εκφυλισμένη**. Δεδομένου ότι κάθε μεταβλητή του συστήματος (2.8) με μηδενική τιμή ορίζει και ένα επίπεδο, μια εκφυλισμένη κορυφή ικανοποιεί τις εξισώσεις περισσότερων των n επιπέδων, αφού οι μηδενικές μεταβλητές μιας εκφυλισμένης λύσης είναι περισσότερες από n . Έτσι, **περίπτωση εκφυλισμού έχουμε όταν από μια κορυφή διέρχονται περισσότερα από n επίπεδα**.

Το Θεώρημα 2.2 αποδεικνύει ότι ο αριθμός των κορυφών του εφικτού συνόλου, το οποίο ορίζεται από τους ανισοτικούς περιορισμούς $\mathbf{Ax} \leq \mathbf{b}$, $\mathbf{A} \in \mathcal{R}^{m \times n}$, $\mathbf{x} \in \mathcal{R}^n$, $\mathbf{b} \in \mathcal{R}^m$ και τους περιορισμούς μη αρνητικότητας $\mathbf{x} \geq \mathbf{0}$, είναι το πολύ ίσος με τον αριθμό των μη αρνητικών βασικών λύσεων του γραμμικού συστήματος (2.4), δηλαδή του συστήματος

$$[\mathbf{A} \mid \mathbf{I}_m] \begin{bmatrix} \mathbf{x} \\ \mathbf{s} \end{bmatrix} = \mathbf{b}$$

Ο μέγιστος αριθμός των βασικών λύσεων του συστήματος αυτού είναι η ίσος με τον αριθμό των δυνατών επιλογών των n ελεύθερων (ή των m βασικών) μεταβλητών από το σύνολο των $n + m$ μεταβλητών του συστήματος. Η διαπίστωση αυτή οδηγεί στην επομένη πρόταση.

Πρόταση 2.1

Το πλήθος των κορυφών του εφικτού συνόλου του προβλήματος (2.1), είναι πεπερασμένο και το πολύ $n! / m! (n-m)!$

Αν το πρόβλημα δεν είναι εκφυλισμένο, τότε ο αριθμός των μη αρνητικών βασικών λύσεων του συστήματος (2.4) είναι ίσος με τον αριθμό των κορυφών του εφικτού συνόλου. Αν όμως το πρόβλημα έχει μη αρνητικές βασικές λύσεις, στις οποίες μία ή περισσότερες από τις βασικές μεταβλητές είναι ίσες με το μηδέν, τότε σε μια κορυφή του εφικτού συνόλου μπορεί να απεικονίζονται περισσότερες από μία μη αρνητικές βασικές λύσεις του συστήματος.

Όπως έχουμε ήδη αναφέρει στην ενότητα αυτή, το εφικτό σύνολο που ορίζεται από τους ανισοτικούς περιορισμούς $\mathbf{Ax} \leq \mathbf{b}$, $\mathbf{A} \in \mathcal{R}^{m \times n}$, $\mathbf{x} \in \mathcal{R}^n$, $\mathbf{b} \in \mathcal{R}^m$ και τους περιορισμούς μη αρνητικότητας $\mathbf{x} \geq \mathbf{0}$ είναι ένα κυρτό πολυέδρο το οποίο, όταν είναι φραγμένο, αποτελεί ένα πολύτοπο. Σύμφωνα με το Θεώρημα 2.2, οι κορυφές του πολυέδρου αυτού αντιστοιχούν στις μη αρνητικές βασικές λύσεις του γραμμικού συστήματος (2.4). Ένα εύλογο ερώτημα είναι τώρα το κατά πόσο οι κορυφές του κυρτού πολυέδρου των εφικτών λύσεων είναι σε θέση να αναπαράγουν σημεία του εφικτού συνόλου που δεν είναι κορυφές του. Για να απαντήσουμε στο ερώτημα αυτό, πρέπει

πρώτα να εισάγουμε δύο έννοιες: την **εφικτή κατεύθυνση** και τη **χωρίς φράγμα κατεύθυνση**.

Ορισμός 2.1

Ένα μη μηδενικό διάνυσμα \mathbf{d} αποτελεί μια εφικτή κατεύθυνση σε ένα σημείο \mathbf{x} του συνόλου $C = \{ \mathbf{x} \in \mathbb{R}^n \mid \bar{\mathbf{A}}\mathbf{x} \leq \bar{\mathbf{b}}, \hat{\mathbf{A}}\mathbf{x} = \hat{\mathbf{b}}, \mathbf{x} \geq \mathbf{0} \}$, αν υπάρχει $\varepsilon > 0$ τέτοιο ώστε $\mathbf{x} + \rho\mathbf{d} \in C \forall \rho \in [0, \varepsilon]$.

Το σύνολο D των εφικτών κατευθύνσεων στο σημείο $\mathbf{x} \in C$ ονομάζεται **κώνος των εφικτών κατευθύνσεων**.

Ας παρατηρήσουμε κατ' αρχάς ότι για να αποτελεί ένα διάνυσμα \mathbf{d} εφικτή κατεύθυνση στο σημείο $\mathbf{x} \in C$, θα πρέπει να ισχύει

$$\hat{\mathbf{A}}(\mathbf{x} + \rho\mathbf{d}) = \hat{\mathbf{b}} \Rightarrow \hat{\mathbf{A}}\mathbf{x} + \rho\hat{\mathbf{A}}\mathbf{d} = \hat{\mathbf{b}} \Rightarrow \hat{\mathbf{A}}\mathbf{d} = \mathbf{0}$$

Ας παρατηρήσουμε ακόμη ότι στο σημείο \mathbf{x} κάποιοι από τους ανισοτικούς περιορισμούς $\bar{\mathbf{A}}\mathbf{x} \leq \bar{\mathbf{b}}$ θα ικανοποιούνται οριακά ως εξισώσεις και κάποιοι άλλοι από τους περιορισμούς αυτούς θα ικανοποιούνται ως γνήσιες ανισότητες. Επίσης, κάποιες από τις συνιστώσες του διανύσματος \mathbf{x} ενδέχεται να είναι ίσες με το μηδέν. Έστω $\bar{\mathbf{A}}_\varepsilon$ ο υποπίνακας του $\bar{\mathbf{A}}$ που αντιστοιχεί στην πρώτη κατηγορία ανισοτικών περιορισμών και $\bar{\mathbf{A}}_\alpha$ ο υποπίνακας που αντιστοιχεί στη δεύτερη κατηγορία. Τότε έχουμε $\bar{\mathbf{A}}_\varepsilon\mathbf{x} = \bar{\mathbf{b}}_\varepsilon$ και $\bar{\mathbf{A}}_\alpha\mathbf{x} < \bar{\mathbf{b}}_\alpha$ όπου $\bar{\mathbf{b}}_\varepsilon$ και $\bar{\mathbf{b}}_\alpha$ είναι οι αντίστοιχοι σταθεροί όροι των περιορισμών. Έστω \mathbf{d} μια κατεύθυνση στο \mathbf{x} , για την οποία $\hat{\mathbf{A}}\mathbf{d} = \mathbf{0}$. Για μικρές τιμές του μήκους βήματος ρ η μετατόπιση προς την κατεύθυνση του \mathbf{d} δεν παραβιάζει κανέναν από τους ανισοτικούς περιορισμούς που ικανοποιούνται ως γνήσιες ανισότητες στο \mathbf{x} . Πράγματι, αφού $\bar{\mathbf{A}}_\alpha\mathbf{x} < \bar{\mathbf{b}}_\alpha$, τότε για αρκούντως μικρές τιμές του μήκους βήματος ρ μπορούμε να εξασφαλίσουμε ότι θα είναι $\bar{\mathbf{A}}_\alpha(\mathbf{x} + \rho\mathbf{d}) < \bar{\mathbf{b}}_\alpha$. Επιπλέον, με κατάλληλη επιλογή της παραμέτρου ρ μπορούμε να εξασφαλίσουμε ότι κάθε μη μηδενική συνιστώσα του \mathbf{x} θα παραμείνει μη αρνητική κατά τη μετατόπιση προς την κατεύθυνση του \mathbf{d} .

Απομένει λοιπόν η περίπτωση των περιορισμών που ικανοποιούνται ως εξισώσεις στο \mathbf{x} . Για να διατηρείται το σημείο $\mathbf{x} + \rho \mathbf{d}$ εντός του συνόλου C , θα πρέπει να ισχύει

$$\bar{\mathbf{A}}_\varepsilon (\mathbf{x} + \rho \mathbf{d}) \leq \bar{\mathbf{b}}_\varepsilon \Rightarrow \bar{\mathbf{A}}_\varepsilon \mathbf{x} + \rho \bar{\mathbf{A}}_\varepsilon \mathbf{d} \leq \bar{\mathbf{b}}_\varepsilon \Rightarrow \bar{\mathbf{b}}_\varepsilon + \rho \bar{\mathbf{A}}_\varepsilon \mathbf{d} \leq \bar{\mathbf{b}}_\varepsilon \Rightarrow \bar{\mathbf{A}}_\varepsilon \mathbf{d} \leq \mathbf{0}$$

Επιπλέον θα πρέπει να είναι $x_j + \rho d_j \geq 0$. Έτσι, αν $x_j = 0$, τότε $d_j \geq 0$.

Ας υποθέσουμε τώρα ότι το σύνολο C ορίζεται μόνο από ανισοτικούς περιορισμούς και περιορισμούς μη αρνητικότητας, δηλαδή

$$C = \{ \mathbf{x} \in \mathbb{R}^n \mid \mathbf{Ax} \leq \mathbf{b}, \mathbf{x} \geq \mathbf{0} \},$$

και ας θεωρήσουμε το διάνυσμα \mathbf{s} με τη βοήθεια του οποίου οι ανισοτικοί περιορισμοί ανάγονται σε εξισώσεις. Αφού το σημείο \mathbf{x} ανήκει στο σύνολο C , τα διανύσματα \mathbf{x} και \mathbf{s} αποτελούν μια μη αρνητική λύση του συστήματος (2.4), δηλαδή έχουμε

$$[\mathbf{A} \mid \mathbf{I}_m] \begin{bmatrix} \mathbf{x} \\ \mathbf{s} \end{bmatrix} = \mathbf{b}, \quad \mathbf{x} \geq \mathbf{0}, \quad \mathbf{s} \geq \mathbf{0}$$

Έστω ακόμα $\Delta \mathbf{x} = \rho \mathbf{d}$ μια μετατόπιση από το \mathbf{x} προς την κατεύθυνση του διανύσματος \mathbf{d} . Καθώς το \mathbf{x} μεταβάλλεται κατά $\Delta \mathbf{x}$, το \mathbf{s} μεταβάλλεται επίσης κατά κάποιο $\Delta \mathbf{s}$. Έστω ότι $\Delta \mathbf{s} = \rho \mathbf{p}$. Για να διατηρείται το σημείο $\mathbf{x} + \Delta \mathbf{x}$ στο σύνολο C , οι μεταβολές $\Delta \mathbf{x}$ και $\Delta \mathbf{s}$ πρέπει να είναι τέτοιες ώστε να ισχύει

$$[\mathbf{A} \mid \mathbf{I}_m] \begin{bmatrix} \mathbf{x} + \Delta \mathbf{x} \\ \mathbf{s} + \Delta \mathbf{s} \end{bmatrix} = \mathbf{b} \Rightarrow [\mathbf{A} \mid \mathbf{I}_m] \begin{bmatrix} \Delta \mathbf{x} \\ \Delta \mathbf{s} \end{bmatrix} = \mathbf{0} \Rightarrow [\mathbf{A} \mid \mathbf{I}_m] \begin{bmatrix} \mathbf{d} \\ \mathbf{p} \end{bmatrix} = \mathbf{0}$$

Επιπλέον, θα πρέπει να είναι $\mathbf{x} + \Delta \mathbf{x} \geq \mathbf{0}$ και $\mathbf{s} + \Delta \mathbf{s} \geq \mathbf{0}$. Έτσι, για κάθε μηδενική συνιστώσα $x_j = 0$ του \mathbf{x} θα πρέπει να είναι $d_j \geq 0$. Ομοίως, για κάθε μηδενική συνιστώσα $s_i = 0$ του \mathbf{s} , δηλαδή για κάθε ανισοτικό περιορισμό που ισχύει οριακά ως εξίσωση στο \mathbf{x} , θα πρέπει να είναι $p_i \geq 0$.

Είναι φανερό ότι οι περιορισμοί οι οποίοι ισχύουν οριακά ως εξισώσεις σε κάποιο σημείο \mathbf{x} του συνόλου $C = \{ \mathbf{x} \in \mathbb{R}^n \mid \mathbf{Ax} \leq \mathbf{b}, \mathbf{x} \geq \mathbf{0} \}$ ικανοποιούνται σε όλα τα σημεία της ημιευθείας $\mathbf{x} + \rho \mathbf{d}$, με αρχή το σημείο \mathbf{x} και διάνυσμα κατεύθυνσης \mathbf{d} , όπου το \mathbf{d} είναι μια εφικτή κατεύθυνση στο \mathbf{x} . Πράγματι, αν $\alpha_{i1}x_1 + \dots + \alpha_{ij}x_j + \dots + \alpha_{in}x_n = b_i$, τότε $s_i = 0$. Έτσι, σύμφωνα με την

ανάλυση που προηγήθηκε, το διάνυσμα \mathbf{p} με το οποίο προεκτείνεται το διάνυσμα \mathbf{d} στο γραμμικό σύστημα

$$[\mathbf{A} \mid \mathbf{I}_m] \begin{bmatrix} \mathbf{d} \\ \mathbf{p} \end{bmatrix} = \mathbf{0}$$

πρέπει να είναι τέτοιο ώστε $p_i \geq 0$. Επομένως, για κάθε $\rho \geq 0$, έχουμε $\Delta s_i = \rho p_i \geq 0$ και συνεπώς ο i -οστός περιορισμός ικανοποιείται κατά μήκος της ημιευθείας $\mathbf{x} + \rho \mathbf{d}$.

Ομοίως, αν $x_j = 0$, τότε $d_j \geq 0$ και έτσι $\Delta x_j = \rho d_j \geq 0$, δηλαδή για κάθε $\rho \geq 0$ η j -οστή μεταβλητή διατηρείται σε μη αρνητικό επίπεδο.

Μεγάλες όμως τιμές του μήκους βήματος ρ ενδέχεται να παραβιάσουν κάποιον από τους ανισοτικούς περιορισμούς ή τους περιορισμούς μη αρνητικότητας οι οποίοι ισχύουν ως γνήσιες ανισότητες στο \mathbf{x} , οδηγώντας σε σημεία εκτός του εφικτού συνόλου. Για έναν τέτοιο περιορισμό ισχύει βέβαια $x_j > 0$ ή $a_{i1}x_1 + \dots + a_{ij}x_j + \dots + a_{in}x_n < b_i \Rightarrow s_i > 0$. Αν λοιπόν για κάποιο $s_i > 0$ είναι $p_i < 0$, τότε, για να διατηρείται η τιμή $s_i + \rho p_i$ μη αρνητική, θα πρέπει $\rho \leq -s_i / p_i$. Για $\rho = -s_i / p_i$ συναντάμε βέβαια το επίπεδο που περιγράφει η εξίσωση $a_{i1}x_1 + \dots + a_{ij}x_j + \dots + a_{in}x_n = b_i$.

Ομοίως, αν $x_j > 0$ και $d_j < 0$, θα πρέπει $\rho \leq -x_j / d_j$. Για $\rho = -x_j / d_j$ φθάνουμε στο επίπεδο $x_j = 0$.

Είναι τώρα φανερό ότι αν $d_j \geq 0$ για κάθε $x_j > 0$ και $p_i \geq 0$ για κάθε $s_i > 0$, τότε οσοδήποτε μεγάλη τιμή και αν πάρει η παράμετρος ρ το σημείο $\mathbf{x} + \rho \mathbf{d}$ διατηρείται εντός του εφικτού συνόλου. Στην περίπτωση αυτή το σύνολο C είναι μη φραγμένο και το διάνυσμα \mathbf{d} , μαζί με την προέκταση του \mathbf{p} , αποτελεί μια **χωρίς φράγμα κατεύθυνση** στο \mathbf{x} . Λαμβάνοντας υπόψη ότι αν $x_j = 0$ πρέπει $d_j \geq 0$ και αν $s_i = 0$ πρέπει $p_i \geq 0$, καταλήγουμε στο συμπέρασμα ότι μια χωρίς φράγμα κατεύθυνση πρέπει να ικανοποιεί τις σχέσεις

$$[\mathbf{A} \mid \mathbf{I}_m] \begin{bmatrix} \mathbf{d} \\ \mathbf{p} \end{bmatrix} = \mathbf{0}, \mathbf{d} \geq \mathbf{0}, \mathbf{p} \geq \mathbf{0}$$

Στη συνέχεια θα αποδείξουμε ότι, εάν το σύνολο $C = \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{Ax} \leq \mathbf{b}, \mathbf{x} \geq \mathbf{0}\}$ είναι φραγμένο, τότε κάθε σημείο του μπορεί να εκφραστεί ως κυρτός συνδυασμός των κορυφών του. Αν το σύνολο C είναι μη φραγμένο, τότε κάθε

σημείο του αποτελεί γραμμικό συνδυασμό μιας χωρίς φράγμα κατεύθυνσης και ενός κυρτού συνδυασμού των κορυφών του.

Θεώρημα 2.3

Έστω $C = \{ \mathbf{x} \in \mathbb{R}^n \mid \mathbf{Ax} \leq \mathbf{b}, \mathbf{x} \geq \mathbf{0} \}$ το εφικτό σύνολο ενός προβλήματος γραμμικού προγραμματισμού και $V = \{v_1, \dots, v_k, \dots, v_r\}$ το σύνολο των ακραίων σημείων (κορυφών) του. Αν το σύνολο C δεν είναι κενό, τότε ούτε και το σύνολο V είναι κενό και κάθε εφικτό σημείο $\mathbf{x} \in C$ μπορεί να εκφραστεί ως ο γραμμικός συνδυασμός

$$\mathbf{x} = \mathbf{d} + \sum_{k=1}^r \lambda_k \mathbf{v}_k, \quad \sum_{k=1}^r \lambda_k = 1, \quad 0 \leq \lambda_k \leq 1 \quad (k = 1, \dots, r)$$

όπου \mathbf{d} είναι είτε το μηδενικό διάνυσμα είτε μια χωρίς φράγμα κατεύθυνση.

Απόδειξη (βλ. π.χ. Nash and Sogier, 1995) Έστω ότι ο πίνακας $\mathbf{A} \in \mathbb{R}^{m \times n}$ είναι βαθμού m . Η υπόθεση αυτή δεν περιορίζει τη γενικότητα της απόδειξης, αφού αν ο πίνακας \mathbf{A} δεν είναι βαθμού m , μπορεί να αντικατασταθεί από έναν πίνακα με μικρότερο πλήθος γραμμών, αλλά κανονικό ως προς τις γραμμές του. Πράγματι, αν ο πίνακας \mathbf{A} δεν είναι βαθμού m , τότε, ανάλογα με την τιμή του διανύσματος \mathbf{b} , είτε οι περιορισμοί του προβλήματος είναι ασυμβίβαστοι, οπότε το σύνολο C είναι κενό, είτε κάποιοι περιορισμοί πλεονάζουν, οπότε μπορούν να παραλειφθούν χωρίς να μεταβληθεί ούτε το εφικτό σύνολο ούτε η βέλτιστη λύση.

Πρώτα εξετάζουμε την περίπτωση που το σύνολο C είναι φραγμένο, δηλαδή είναι ένα πολύτοπο, οπότε δεν υπάρχουν χωρίς φράγμα κατευθύνσεις και $\mathbf{d} = \mathbf{0}$. Έστω $\mathbf{x} \in C$ ένα εφικτό σημείο. Αν $\mathbf{x} = \mathbf{v}_i$, για κάποιο $i \in \{1, \dots, r\}$, τότε το θεώρημα αληθεύει για $\lambda_i = 1$ και $\lambda_k = 0$, $k \in \{1, \dots, r\}$, $k \neq i$.

Έστω λοιπόν ότι το \mathbf{x} δεν είναι κάποια από τις κορυφές του συνόλου C . Τότε τα διανύσματα \mathbf{x} και \mathbf{s} , όπου κατά τα γνωστά το \mathbf{s} μετατρέπεται τους ανισοτικούς περιορισμούς σε εξισώσεις, αποτελούν μια μη αρνητική λύση του συστήματος (2.4). Η λύση όμως αυτή, σύμφωνα με το Θεώρημα 2.2, δεν είναι βασική, αφού το \mathbf{x} δεν είναι ακραίο σημείο του C . Επομένως το πλήθος των μη μηδενικών συνιστωσών των διανυσμάτων \mathbf{x} και \mathbf{s} είναι από κοινού

μεγαλύτερο του m . Κατά συνέπεια οι στήλες του $m \times (n+m)$ πίνακα $[A \mid I_m]$ που αντιστοιχούν στις μη μηδενικές συνιστώσες των διανυσμάτων αυτών, ως περισσότερες σε πλήθος από m , είναι γραμμικά εξαρτημένες.

Ας θεωρήσουμε τώρα μια εφικτή κατεύθυνση \mathbf{d} στο σημείο \mathbf{x} . Μια τέτοια κατεύθυνση πρέπει να ικανοποιεί τις σχέσεις

$$[A \mid I_m] \begin{bmatrix} \mathbf{d} \\ \mathbf{p} \end{bmatrix} = \mathbf{0}, \quad x_j = 0 \Rightarrow d_j \geq 0, \quad j \in \{1, \dots, n\}$$

$$s_i = 0 \Rightarrow p_i \geq 0, \quad i \in \{1, \dots, m\}$$

Θέτοντας $d_j = 0$ για κάθε μηδενική συνιστώσα $x_j = 0$ του \mathbf{x} και $p_i = 0$ για κάθε μηδενική συνιστώσα $s_i = 0$ του \mathbf{s} , βρίσκουμε μια μη μηδενική λύση του γραμμικού συστήματος που τότε προκύπτει. Η λύση βέβαια αυτή ορίζει μια εφικτή κατεύθυνση στο \mathbf{x} . Επιπλέον, μία τουλάχιστον από τις μη μηδενικές συνιστώσες της λύσης πρέπει να είναι αρνητική, αφού διαφορετικά θα ίσχυε $\mathbf{d} \geq \mathbf{0}$ και $\mathbf{p} \geq \mathbf{0}$ και επομένως το διάνυσμα \mathbf{d} θα αποτελούσε μια χωρίς φράγμα κατεύθυνση, κάτι που από υπόθεση έχει αποκλειστεί. Ας σημειωθεί ότι οι περιορισμοί οι οποίοι ισχύουν οριακά ως εξισώσεις στο σημείο \mathbf{x} ικανοποιούνται ως εξισώσεις σε όλα τα σημεία της ημιευθείας $\mathbf{x} + \rho \mathbf{d}$. Καθώς το μήκος του βήματος αυξάνει, οι τιμές των θετικών συνιστωσών των διανυσμάτων \mathbf{x} και \mathbf{s} , για τις οποίες οι αντίστοιχες συνιστώσες των διανυσμάτων \mathbf{d} και \mathbf{p} είναι αρνητικές, μειώνονται και κάποια από αυτές φθάνει πρώτη στο μηδέν. Αυτό σημαίνει ότι βρήκαμε ένα συνοριακό σημείο $\bar{\mathbf{x}}$ του C , στο οποίο τουλάχιστον μία επιπλέον συνιστώσα είτε του \mathbf{x} είτε του \mathbf{s} είναι μηδενική, πέραν των συνιστωσών τους που έχουν ήδη τιμή ίση με το μηδέν στο σημείο \mathbf{x} . Με άλλα λόγια, στο σημείο \mathbf{x} τουλάχιστον ένας επιπλέον περιορισμός ισχύει οριακά ως εξίσωση. Ας συμβολίσουμε ακόμα με $\hat{\mathbf{x}}$ το συνοριακό σημείο που προκύπτει κατά τη μετατόπιση από το \mathbf{x} προς την κατεύθυνση του διανύσματος $-\mathbf{d}$. Σημειώστε ότι το ζεύγος των διανυσμάτων $(-\mathbf{d}, -\mathbf{p})$ ορίζει επίσης μια εφικτή κατεύθυνση στο σημείο \mathbf{x} , αφού οι συνιστώσες των \mathbf{d} και \mathbf{p} , που αντιστοιχούν στις μηδενικές συνιστώσες των \mathbf{x} και \mathbf{s} , έχουν τεθεί ίσες με το μηδέν. Επιπλέον, το σημείο \mathbf{x} αποτελεί κυρτό συνδυασμό των σημείων $\bar{\mathbf{x}}$ και $\hat{\mathbf{x}}$.

Αν τώρα τα σημεία \bar{x} και \hat{x} είναι και τα δύο κορυφές, η απόδειξη έχει ολοκληρωθεί. Αν κάποιο από τα δύο αυτά σημεία δεν είναι κορυφή, τότε, εφαρμόζοντας την ίδια συλλογιστική, μπορούμε να το εκφράσουμε ως κυρτό συνδυασμό δύο συνοριακών σημείων στα οποία τουλάχιστον μία επιπλέον συνιστώσα είτε του x είτε του s είναι μηδενική. Η διαδικασία αυτή επαναλαμβάνεται μέχρις ότου προκύψουν σημεία τα οποία εκφράζονται ως κυρτοί συνδυασμοί κορυφών. Τότε το σημείο x ως κυρτός συνδυασμός σημείων, που είναι κυρτοί συνδυασμοί κορυφών, είναι ομοίως κυρτός συνδυασμός των κορυφών αυτών. Το σκεπτικό αυτό αποδεικνύει επίσης ότι το σύνολο των κορυφών δεν είναι κενό. Αφού ο αριθμός των μη μηδενικών συνιστωσών των διανυσμάτων x και s μειώνεται σε κάθε επανάληψη της πιο πάνω διαδικασίας τουλάχιστον κατά ένα και ο αριθμός αυτός έχει ως κάτω φράγμα το μηδέν, τελικά τα σημεία που παράγονται από μια τέτοια διαδικασία πρέπει να αποτελούν βασικές εφικτές λύσεις, δηλαδή κορυφές.

Με παρόμοιο τρόπο γίνεται η απόδειξη του θεωρήματος όταν το σύνολο C είναι μη φραγμένο. Έστω x σημείο του C που δεν είναι κορυφή του. Έστω ακόμα ότι το ζεύγος (d, p) ορίζει μια εφικτή κατεύθυνση στο x , όπως στην περίπτωση που το σύνολο C είναι ένα πολύτοπο. Τότε και το ζεύγος $(-d, -p)$ ορίζει μια εφικτή κατεύθυνση. Τώρα όμως μία από τις δύο κατευθύνσεις ενδέχεται να είναι χωρίς φράγμα. Σημειώστε ότι δεν είναι δυνατόν και οι δύο κατευθύνσεις να είναι χωρίς φράγμα, αφού σε μια τέτοια περίπτωση θα ίσχυε ταυτόχρονα $(d \geq 0, -d \geq 0)$ και $(p \geq 0, -p \geq 0)$ και επομένως θα είχαμε $(d = 0, p = 0)$. Έστω (d, p) η χωρίς φράγμα κατεύθυνση. Επομένως η μετατόπιση προς την κατεύθυνση του $-d$ θα μας φέρει τελικά σε ένα συνοριακό σημείο όπου τουλάχιστον ένας επιπλέον περιορισμός ικανοποιείται ως εξίσωση, πέραν των περιορισμών οι οποίοι ήδη ισχύουν ως εξισώσεις στο σημείο x . Η διαδικασία αυτή επαναλαμβάνεται μέχρις ότου προκύψει ένας κυρτός συνδυασμός κορυφών. Δεδομένου ότι ο μη αρνητικός γραμμικός συνδυασμός χωρίς φράγμα κατευθύνσεων αποτελεί ομοίως χωρίς φράγμα κατεύθυνση, τελικά το σημείο x εκφράζεται ως γραμμικός συνδυασμός μιας χωρίς φράγμα κατεύθυνσης και ενός κυρτού συνδυασμού κορυφών.

Το επόμενο θεώρημα ολοκληρώνει τη θεωρητική θεμελίωση της μεθόδου Simplex.

Θεώρημα 2.4

Αν το πρόβλημα $\max \mathbf{c}^t \mathbf{x}$, $\mathbf{Ax} \leq \mathbf{b}$, $\mathbf{x} \geq \mathbf{0}$, $\mathbf{A} \in \mathcal{R}^{m \times n}$, $\mathbf{x} \in \mathcal{R}^n$, $\mathbf{b} \in \mathcal{R}^m$ (2.13) έχει πεπερασμένη βέλτιστη λύση, τότε έχει μία βέλτιστη βασική εφικτή λύση. Αν η βέλτιστη τιμή της αντικειμενικής συνάρτησης αντιστοιχεί σε περισσότερες από μία κορυφές του εφικτού συνόλου C , τότε κάθε κυρτός συνδυασμός των κορυφών αυτών αποτελεί βέλτιστη λύση.

Απόδειξη Έστω \mathbf{x} μια πεπερασμένη βέλτιστη λύση του προβλήματος (2.13). Σύμφωνα με το Θεώρημα 2.3, το σημείο \mathbf{x} μπορεί να τεθεί στη μορφή

$$\mathbf{x} = \mathbf{d} + \sum_{k=1}^r \lambda_k \mathbf{v}_k, \quad \sum_{k=1}^r \lambda_k = 1, \quad 0 \leq \lambda_k \leq 1 \quad (k = 1, \dots, r) \quad (2.14)$$

όπου $V = \{\mathbf{v}_1, \dots, \mathbf{v}_k, \dots, \mathbf{v}_r\}$ είναι το σύνολο των κορυφών του εφικτού συνόλου και \mathbf{d} είναι μια χωρίς φράγμα κατεύθυνση για την οποία ισχύει

$$[\mathbf{A} \mid \mathbf{I}_m] \begin{bmatrix} \mathbf{d} \\ \mathbf{p} \end{bmatrix} = \mathbf{0}, \quad \mathbf{d} \geq \mathbf{0}, \quad \mathbf{p} \geq \mathbf{0} \Leftrightarrow \mathbf{Ad} \leq \mathbf{0}, \quad \mathbf{d} \geq \mathbf{0} \quad (2.15)$$

Η τιμή της αντικειμενικής συνάρτησης στο βέλτιστο σημείο (2.14) είναι

$$z^* = \mathbf{c}^t \mathbf{x} = \mathbf{c}^t \mathbf{d} + \sum_{k=1}^r \lambda_k \mathbf{c}^t \mathbf{v}_k$$

Ας παρατηρήσουμε τώρα ότι το σημείο

$$\mathbf{x}_\lambda = \lambda \mathbf{d} + \sum_{k=1}^r \lambda_k \mathbf{v}_k$$

διατηρείται εντός του εφικτού συνόλου για κάθε $\lambda > 0$. Πράγματι, λαμβάνοντας υπόψη τη σχέση (2.15) καθώς και το γεγονός ότι οι κορυφές \mathbf{v}_k ($k = 1, \dots, r$), ως εφικτά σημεία, ικανοποιούν τις ανισότητες $\mathbf{Av}_k \leq \mathbf{b}$, έχουμε

$$\mathbf{Ax}_\lambda = \lambda \mathbf{Ad} + \sum_{k=1}^r \lambda_k \mathbf{Av}_k \leq \mathbf{b} \sum_{k=1}^r \lambda_k = \mathbf{b}$$

Επιπλέον, αφού $\mathbf{v}_k \geq \mathbf{0}$ ($k = 1, \dots, r$), $\mathbf{d} \geq \mathbf{0}$, $0 \leq \lambda_k \leq 1$, $\lambda > 0$, ($k = 1, \dots, r$) και $\lambda > 0$, είναι $\mathbf{x}_\lambda \geq \mathbf{0}$. Άρα $\mathbf{x}_\lambda \in C$. Η τιμή της αντικειμενικής συνάρτησης στο σημείο \mathbf{x}_λ είναι

$$z(\lambda) = \mathbf{c}^t \mathbf{x}_\lambda = \lambda \mathbf{c}^t \mathbf{d} + \sum_{k=1}^r \lambda_k \mathbf{c}^t \mathbf{v}_k$$

Επομένως, αν $\mathbf{c}^t \mathbf{d} > 0$, τότε καθώς $\lambda \rightarrow \infty$ η τιμή της αντικειμενικής συνάρτησης αποκλίνει στο $+\infty$. Αυτό όμως έρχεται σε αντίφαση με την

υπόθεση ότι το πρόβλημα έχει πεπερασμένη βέλτιστη λύση. Κατά συνέπεια πρέπει να είναι είτε $\mathbf{c}^t \mathbf{d} < 0$ είτε $\mathbf{c}^t \mathbf{d} = 0$. Έστω ότι $\mathbf{c}^t \mathbf{d} < 0$. Τότε όμως προκύπτει η αντίφαση ότι η τιμή της αντικειμενικής συνάρτησης στο σημείο

$$\mathbf{v} = \sum_{k=1}^r \lambda_k \mathbf{v}_k$$

θα είναι μεγαλύτερη από την τιμή της στο βέλτιστο σημείο (2.14), αφού

$$z^* = \mathbf{c}^t \mathbf{x} = \mathbf{c}^t \mathbf{d} + \sum_{k=1}^r \lambda_k \mathbf{c}^t \mathbf{v}_k < \sum_{k=1}^r \lambda_k \mathbf{c}^t \mathbf{v}_k = \mathbf{c}^t \mathbf{v}$$

Επομένως θα πρέπει να είναι $\mathbf{c}^t \mathbf{d} = 0$. Έτσι η βέλτιστη τιμή της αντικειμενικής συνάρτησης είναι

$$z^* = \sum_{k=1}^r \lambda_k \mathbf{c}^t \mathbf{v}_k = \mathbf{c}^t \sum_{k=1}^r \lambda_k \mathbf{v}_k = \mathbf{c}^t \mathbf{v}$$

Έστω τώρα \mathbf{v}^* η κορυφή όπου η αντικειμενική συνάρτηση z παρουσιάζει τη μεγαλύτερη τιμή της μεταξύ των r κορυφών του εφικτού συνόλου, δηλαδή

$$\mathbf{c}^t \mathbf{v}^* = \max_{k=1, \dots, r} \{\mathbf{c}^t \mathbf{v}_k\}$$

Τότε για κάθε κυρτό συνδυασμό \mathbf{v} των κορυφών έχουμε

$$z^* = \mathbf{c}^t \mathbf{v} = \sum_{k=1}^r \lambda_k \mathbf{c}^t \mathbf{v}_k \leq \sum_{k=1}^r \lambda_k \mathbf{c}^t \mathbf{v}^* = \mathbf{c}^t \mathbf{v}^* \sum_{k=1}^r \lambda_k = \mathbf{c}^t \mathbf{v}^*$$

Δεδομένου όμως ότι z^* είναι η βέλτιστη τιμή της αντικειμενικής συνάρτησης, θα πρέπει τελικά να είναι $z^* = \mathbf{c}^t \mathbf{v}^*$. Η σχέση αυτή φανερώνει ότι υπάρχει μια βέλτιστη κορυφή, δηλαδή μια βέλτιστη βασική εφικτή λύση.

Ας υποθέσουμε τώρα ότι η αντικειμενική συνάρτηση έχει την ίδια μέγιστη τιμή z^* σε $q \in \{2, \dots, r\}$ κορυφές \mathbf{v}_{k_i} ($i = 1, \dots, q$) του εφικτού συνόλου, οπότε $\mathbf{c}^t \mathbf{v}_{k_i} = z^*$. Για κάθε κυρτό συνδυασμό

$$\mathbf{v} = \sum_{i=1}^q \lambda_i \mathbf{v}_{k_i}, \quad \sum_{i=1}^q \lambda_i = 1, \quad 0 \leq \lambda_i \leq 1 \quad (i = 1, \dots, q)$$

των q αυτών κορυφών έχουμε

$$z = \mathbf{c}^t \mathbf{v} = \sum_{i=1}^q \lambda_i \mathbf{c}^t \mathbf{v}_{k_i} = z^* \sum_{i=1}^q \lambda_i = z^*$$

σχέση η οποία αποδεικνύει ότι κάθε κυρτός συνδυασμός των εναλλακτικών βέλτιστων κορυφών ορίζει ένα επίσης βέλτιστο σημείο.

Ο ΑΛΓΟΡΙΘΜΟΣ SIMPLEX

| | |
|--|--|
| <p>Βήμα 1 Προσδιορισμός μιας αρχικής κορυφής</p> | <p>▶ Προσδιόρισε μια αρχική βασική εφικτή λύση $\mathbf{x}_B = \mathbf{B}^{-1}\mathbf{b} \geq 0$ και τα δεικτοσύνολα B και E των βασικών και των ελεύθερων μεταβλητών της λύσης αυτής.</p> <p>▶ Υπολόγισε το διάνυσμα ανηγμένου κέρδους $\mathbf{r}_E^t = \mathbf{c}_E^t - \mathbf{c}_B^t \mathbf{B}^{-1}\mathbf{E}$.</p> |
| <p>Βήμα 2 Κριτήριο τερματισμού</p> | <p>▶ Αν $\mathbf{r}_E^t \leq \mathbf{0}^t$ πήγαινε στο βήμα 7: η βέλτιστη λύση έχει βρεθεί. Διαφορετικά πήγαινε στο βήμα 3.</p> |
| <p>Βήμα 3 Επιλογή εισερχόμενης μεταβλητής</p> | <p>▶ Βάλε στη βάση την ελεύθερη μεταβλητή x_{j_p} για την οποία</p> $r_{j_p} = \max_{j_i \in E} \{r_{j_i}, r_{j_i} > 0\}.$ |
| <p>Βήμα 4 Έλεγχος μη πεπερασμένου της βέλτιστης λύσης</p> | <p>▶ Αν τα στοιχεία της εισερχόμενης στήλης \mathbf{u}_{j_p} είναι μη θετικά, αν δηλαδή $\mathbf{u}_{j_p} = 0$, τερμάτισε: η μέγιστη τιμή του z αποκλίνει στο $+\infty$. Διαφορετικά πήγαινε στο βήμα 5.</p> |
| <p>Βήμα 5 Προσδιορισμός εξερχόμενης μεταβλητής</p> | <p>▶ Βγάλε από τη βάση τη μεταβλητή x_{i_q} για την οποία</p> $\frac{x_{i_q}}{u_{i_q j_p}} = \min_{j_k \in B} \left\{ \frac{x_{i_k}}{u_{i_k j_p}}, u_{i_k j_p} > 0 \right\}.$ |
| <p>Βήμα 6 Οδήγηση</p> | <p>▶ Μετασημάτισε τη στήλη \mathbf{u}_{j_p} της εισερχόμενης μεταβλητής x_{j_p} σε μοναδιαίο διάνυσμα με οδηγό το στοιχείο $u_{i_q j_p}$.</p> <p>▶ Θέσε $B \leftarrow (B - \{i_q\}) \cup \{j_p\}$, $E \leftarrow (E - \{j_p\}) \cup \{i_q\}$.</p> <p>▶ Διαμόρφωσε το νέο ταμπλό.</p> <p>▶ Υπολόγισε το διάνυσμα ανηγμένου κέρδους \mathbf{r}_E^t.</p> <p>▶ Επίστρεψε στο Βήμα 2.</p> |
| <p>Βήμα 7 Προσδιορισμός εναλλακτικών βέλτιστων λύσεων</p> | <p>▶ Βάλε στη βάση κάθε ελεύθερη μεταβλητή x_{j_i} για την οποία $r_{j_i} = 0$, $j_i \in E$ και βρες τις εναλλακτικές βέλτιστες λύσεις.</p> |

3. ΘΕΩΡΗΤΙΚΗ ΘΕΜΕΛΙΩΣΗ ΤΗΣ ΜΕΘΟΔΟΥ ΤΩΝ ΕΛΛΕΙΨΟΕΙΔΩΝ

Σε ότι αφορά στο θέμα της υπολογιστικής πολυπλοκότητας της μεθόδου Simplex, αξίζει να αναφέρουμε ότι επί πολλά χρόνια δεν ήταν γνωστό αν ο γραμμικός προγραμματισμός ανήκει στην επιθυμητή κατηγορία των **P** προβλημάτων (P-problem = polynomial time) που λύνονται σε **πολυωνυμικό χρόνο**.

Λέμε ότι μια κατηγορία προβλημάτων έχει **πολυωνυμική πολυπλοκότητα** αν κάθε πρόβλημα της κατηγορίας αυτής, μεγέθους L , μπορεί να λυθεί με κάποιον αλγόριθμο του οποίου το κόστος στη χειρότερη περίπτωση είναι $O(f(L))$, όπου $f(L)$ είναι ένα πολυώνυμο ως προς L .

Για τα **NP** (nondeterministic polynomial time) προβλήματα πιστεύεται, χωρίς ωστόσο να έχει αποδειχθεί, ότι όλοι οι προσδιοριστικοί αλγόριθμοι απαιτούν στη χειρότερη περίπτωση εκθετικό χρόνο για να βρουν τη βέλτιστη λύση. Πάντως, ο κύβος των Klee-Minty δείχνει ότι, ακόμα και αν ο γραμμικός προγραμματισμός έχει πολυωνυμική πολυπλοκότητα, η μέθοδος Simplex σίγουρα δεν είναι ο αλγόριθμος που δίνει το πολυωνυμικό φράγμα, αφού το 2^m δεν φράσσεται από κάποιο πολυώνυμο.

Ο Khachiyan (1979) ήταν εκείνος ο οποίος, με τη **μέθοδο των ελλειψοειδών** που ανέπτυξε, απέδειξε ότι τα προβλήματα του γραμμικού προγραμματισμού είναι δυνατόν να λυθούν σε πολυωνυμικό χρόνο. Η μέθοδος αυτή, σε αντίθεση με τον αλγόριθμο Simplex που κινείται από κορυφή σε κορυφή, προσεγγίζει τη βέλτιστη λύση μέσα από μια ακολουθία συρρικνούμενων ελλειψοειδών.

Η μέθοδος των ελλειψοειδών έτυχε της έντονης προσοχής των ερευνητών και όχι μόνο. Είναι χαρακτηριστικό ότι οι New York Times, όπως και άλλα έντυπα γενικότερου ενδιαφέροντος δημοσίευσαν άρθρα σχετικά με την ανακάλυψη του Khachiyan, ένδειξη της πρακτικής σπουδαιότητας του γραμμικού προγραμματισμού.

Κατά παράδοξο τρόπο, η μέθοδος του Khachiyan αποδείχθηκε στην πράξη χειρότερη από τη μέθοδο Simplex, η οποία λύνει ένα πρόβλημα γραμμικού προγραμματισμού σε χρόνο που είναι κατά μέσο όρο

πολυωνυμικός. Ο Strang (1995) ισχυρίζεται ότι, για κάποιο λόγο που βρίσκεται κρυμμένος πίσω από τη γεωμετρία των πολυδιάστατων πολυέδρων, τα εφικτά σύνολα τύπου Klee-Minty είναι σπάνια και η μέθοδος Simplex αποδεικνύεται στην πράξη πολύ τυχερή.

Η μέθοδος του Khachiyan αναζητά ένα σημείο το οποίο να ικανοποιεί γνήσια το σύστημα των γραμμικών ανισοεξισώσεων $\mathbf{Ax} \leq \mathbf{b}$. Η μέθοδος δηλαδή προσπαθεί να βρει ένα σημείο τέτοιο ώστε $\mathbf{Ax} < \mathbf{b}$.

Ας θεωρήσουμε τώρα το πρόβλημα:

$$\begin{array}{|l} \max z = \mathbf{c}^t \mathbf{x} \\ \text{με τους } \mathbf{Ax} \leq \mathbf{b}, \mathbf{x} \geq \mathbf{0} \text{ περιορισμούς, όπου } \mathbf{A} \in \mathcal{R}^{m \times n}, \mathbf{b} \in \mathcal{R}^m, \mathbf{c} \in \mathcal{R}^n, \mathbf{x} \in \mathcal{R}^n \end{array}$$

και το δυϊκό του

$$\begin{array}{|l} \min y = \mathbf{b}^t \mathbf{w} \\ \text{με τους } \mathbf{A}^t \mathbf{w} \geq \mathbf{c}, \mathbf{w} \geq \mathbf{0} \text{ ως περιορισμούς, } \mathbf{w} \in \mathcal{R}^m \end{array}$$

Σύμφωνα με το θεώρημα του ισχυρού δυϊσμού, στις βέλτιστες λύσεις οι τιμές των αντικειμενικών συναρτήσεων των δύο προβλημάτων είναι ίσες. Εξάλλου, σύμφωνα με το θεώρημα του ασθενούς δυϊσμού, για κάθε ζεύγος εφικτών λύσεων των δύο προβλημάτων ισχύει $z \leq y$. Έτσι, αν βρούμε ένα ζεύγος εφικτών λύσεων για το οποίο να είναι $z \geq y$, τότε το ζεύγος αυτό πρέπει να είναι βέλτιστο, αφού ικανοποιεί ταυτόχρονα τις ανισότητες $z \leq y$ και $z \geq y$ και, επομένως, $z = y$.

Οι πιο πάνω συνθήκες είναι ισοδύναμες με το σύστημα των γραμμικών ανισώσεων

$$\tilde{\mathbf{A}} \tilde{\mathbf{x}} \leq \tilde{\mathbf{b}}$$

όπου ($\mathbf{0}$ είναι το μηδενικό διάνυσμα και \mathbf{O} είναι ο μηδενικός πίνακας)

$$\tilde{\mathbf{A}} = \begin{bmatrix} -\mathbf{c}^t & \mathbf{b}^t \\ \mathbf{A} & \mathbf{O} \\ \mathbf{O} & -\mathbf{A}^t \\ -\mathbf{I} & \mathbf{O} \\ \mathbf{O} & -\mathbf{I} \end{bmatrix}, \tilde{\mathbf{x}} = \begin{bmatrix} \mathbf{x} \\ \mathbf{w} \end{bmatrix}, \tilde{\mathbf{b}} = \begin{bmatrix} \mathbf{0} \\ \mathbf{b} \\ -\mathbf{c} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix}$$

Με δεδομένη την προηγούμενη ισοδυναμία, στα επόμενα μπορούμε να υποθέτουμε ότι η ανάλυση αναφέρεται σε ένα γραμμικό σύστημα της μορφής $\mathbf{Ax} \leq \mathbf{b}$, όπου \mathbf{A} είναι ένας $m \times n$ πίνακας. Υποθέτοντας ακόμα ότι οι παράμετροι του προβλήματος είναι ακέραιοι αριθμοί, το μέγεθος L του γραμμικού συστήματος $\mathbf{Ax} \leq \mathbf{b}$ είναι

$$L = \sum_{a_{ij} \neq 0} (\lfloor \log_2 |a_{ij}| \rfloor + 2) + \sum_{b_i \neq 0} (\lfloor \log_2 |b_i| \rfloor + 2) + \sum_{c_j \neq 0} (\lfloor \log_2 |c_j| \rfloor + 2) + (\lfloor \log_2 n \rfloor + 1) + (\lfloor \log_2 m \rfloor + 1)$$

Έστω τώρα $\mathbf{e} = [1 \dots 1 \dots 1]^t$. Οι Gacs και Lovasz (1981) απέδειξαν ότι, αν το σύστημα

$$\mathbf{Ax} < \mathbf{b} + 2^{-L} \mathbf{e}$$

έχει λύση, τότε και το σύστημα $\mathbf{Ax} \leq \mathbf{b}$ έχει λύση.

Πριν αρχίσουμε να περιγράψουμε τη μέθοδο των ελλειψοειδών, κρίνουμε σκόπιμο να δώσουμε τον ορισμό του ελλειψοειδούς. Το ελλειψοειδές, λοιπόν, δεν είναι παρά η γενίκευση της έλλειψης στο χώρο των n διαστάσεων. Ορίζεται ως το σύνολο των σημείων

$$E = \{ \mathbf{x} \in \mathbb{R}^n \mid (\mathbf{x} - \bar{\mathbf{x}})^t \mathbf{Q}^{-1} (\mathbf{x} - \bar{\mathbf{x}}) \leq 1 \}$$

όπου $\bar{\mathbf{x}}$ είναι το κέντρο του ελλειψοειδούς και \mathbf{Q} είναι ένας $n \times n$ συμμετρικός και θετικά ορισμένος πίνακας, ο οποίος προσδιορίζει το σχήμα και τον προσανατολισμό του ελλειψοειδούς.

Ο πίνακας \mathbf{Q} επιδέχεται την παραγοντοποίηση $\mathbf{Q} = \mathbf{P}\mathbf{\Lambda}\mathbf{P}^t$, όπου $\mathbf{\Lambda}$ είναι ο διαγώνιος πίνακας $\mathbf{\Lambda} = \text{diag}(\lambda_1 \dots \lambda_i \dots \lambda_n)$ με στοιχεία τις ιδιοτιμές λ_i ($i = 1, \dots, n$) του \mathbf{Q} , και \mathbf{P} είναι ένας ορθογώνιος πίνακας με στήλες τα αντίστοιχα ιδιοδιανύσματα. Στη βάση της ανάπτυξης του \mathbf{Q} στο χώρο των ιδιοδιανυσμάτων του και αναγνωρίζοντας ότι $\mathbf{P}^{-1} = \mathbf{P}^t$ και $\mathbf{Q}^{-1} = \mathbf{P}\mathbf{\Lambda}^{-1}\mathbf{P}^t$, όπου $\mathbf{\Lambda}^{-1} = \text{diag}(1/\lambda_1 \dots 1/\lambda_i \dots 1/\lambda_n)$, μπορούμε τώρα να θέσουμε την έκφραση $(\mathbf{x} - \bar{\mathbf{x}})^t \mathbf{Q}^{-1} (\mathbf{x} - \bar{\mathbf{x}}) \leq 1$ στη μορφή $(\mathbf{x} - \bar{\mathbf{x}})^t \mathbf{P}\mathbf{\Lambda}\mathbf{P}^t (\mathbf{x} - \bar{\mathbf{x}}) \leq 1$. Με την εισαγωγή των μεταβλητών $\boldsymbol{\chi} = \mathbf{x} - \bar{\mathbf{x}}$ και $\boldsymbol{\omega} = \mathbf{P}^t \boldsymbol{\chi}$, το σύνολο E γράφεται

$$E = (\mathbf{x} - \bar{\mathbf{x}})^t \mathbf{P}\mathbf{\Lambda}^{-1}\mathbf{P}^t (\mathbf{x} - \bar{\mathbf{x}}) \leq 1 \Rightarrow (\mathbf{P}^t \boldsymbol{\chi})^t \mathbf{\Lambda}^{-1} (\mathbf{P}^t \boldsymbol{\chi}) \leq 1 \Rightarrow \boldsymbol{\omega}^t \mathbf{\Lambda}^{-1} \boldsymbol{\omega} \leq 1 \Rightarrow$$

$$\Rightarrow \sum_{i=1}^n \frac{\omega_i^2}{\lambda_i} \leq 1$$

Οι αριθμοί $\sqrt{\lambda_i}$ ($i = 1, \dots, n$) είναι τα μήκη των ημιαξόνων του ελλειψοειδούς. Σημειώστε ότι, αφού ο πίνακας \mathbf{Q} είναι θετικά ορισμένος, οι ιδιοτιμές του είναι θετικοί αριθμοί και έτσι είναι οι τετραγωνικές ρίζες υπάρχουν.

Εδώ ας παρατηρήσουμε ότι ο μετασχηματισμός $\boldsymbol{\chi} = \mathbf{x} - \bar{\mathbf{x}}$ ισοδυναμεί με την παράλληλη μεταφορά των αξόνων στο σημείο \mathbf{x} , ενώ ο μετασχηματισμός $\boldsymbol{\omega} = \mathbf{P}^t \boldsymbol{\chi}$ αποτελεί στροφή, αφού $\boldsymbol{\omega}^t \boldsymbol{\omega} = (\mathbf{P}^t \boldsymbol{\chi})^t (\mathbf{P}^t \boldsymbol{\chi}) = \boldsymbol{\chi}^t \mathbf{P} \mathbf{P}^t \boldsymbol{\chi} = \boldsymbol{\chi}^t \boldsymbol{\chi}$, αφού δηλαδή με αυτό το μετασχηματισμό τα μήκη διατηρούνται αναλλοίωτα.

Κατά την εκκίνηση της μεθόδου επιλέγουμε ένα ελλειψοειδές του οποίου το κέντρο $\bar{\mathbf{x}}_0$ είναι η αρχή των αξόνων και το οποίο περιέχει τμήμα \hat{S} του εφικτού συνόλου $S = \{ \mathbf{x} \mid \mathbf{A}\mathbf{x} < \mathbf{b} \}$. Στη συνέχεια υποθέτουμε ότι τα στοιχεία του πίνακα \mathbf{A} και του διανύσματος \mathbf{b} είναι ακέραιοι αριθμοί. Το αρχικό ελλειψοειδές ορίζεται από τον πίνακα \mathbf{Q}_0 , ο οποίος είναι συνήθως ένα πολλαπλάσιο του ταυτοτικού πίνακα \mathbf{I} . Προκειμένου να μειώσουμε το όριο των επαναλήψεων για τη σύγκλιση της μεθόδου, είναι σκόπιμο να επιλέξουμε έναν τέτοιο πίνακα \mathbf{Q}_0 ώστε το αρχικό ελλειψοειδές να είναι όσο το δυνατόν μικρότερο.

Εφόσον δεν υπάρχει άλλη σχετική πληροφορία, θέτουμε $\mathbf{Q}_0 = 2^L \mathbf{I}$, μια επιλογή η οποία εξασφαλίζει ότι το αρχικό ελλειψοειδές E_0 είναι αρκετά μεγάλο ώστε να περιέχει τμήμα του εφικτού συνόλου, αν βέβαια το σύνολο αυτό δεν είναι κενό.

Κατά την $(k+1)$ -οστή επανάληψη ο αλγόριθμος ελέγχει πρώτα μήπως το κέντρο $\bar{\mathbf{x}}_k$ του τρέχοντος ελλειψοειδούς E_k είναι εφικτό, μήπως δηλαδή $\mathbf{A}\bar{\mathbf{x}}_k < \mathbf{b}$. Αν ναι, η μέθοδος τερματίζει. Αν όχι, μία τουλάχιστον από τις ανισότητες $\boldsymbol{\alpha}_i^t \bar{\mathbf{x}}_k \geq b_i$, $i \in \{1, \dots, m\}$, όπου $\boldsymbol{\alpha}_i^t$ είναι οι γραμμές του πίνακα \mathbf{A} , παραβιάζεται στο σημείο $\bar{\mathbf{x}}_k$. Επιλέγουμε οποιαδήποτε από τις ανισότητες για τις οποίες είναι $\boldsymbol{\alpha}_i^t \bar{\mathbf{x}}_k \geq b_i$, $i \in \{1, \dots, m\}$, για να ορίσουμε ένα μικρότερο ελλειψοειδές E_{k+1} , το οποίο όμως να περιέχει το ίδιο τμήμα \hat{S} του εφικτού συνόλου S με το αρχικό ελλειψοειδές E_0 .

Ο προσδιορισμός ενός τέτοιου ελλειψοειδούς γίνεται με την κατάλληλη επιλογή του κέντρου $\bar{\mathbf{x}}_{k+1}$ του ελλειψοειδούς και του πίνακα \mathbf{Q}_{k+1} . Σε κάθε επανάληψη ο όγκος του ελλειψοειδούς συρρικνώνεται κατά σταθερό παράγοντα. Δεδομένου ότι όλες οι παράμετροι του συστήματος $\mathbf{Ax} < \mathbf{b}$ είναι, από υπόθεση, ακέραιοι αριθμοί, μπορούμε να αποδείξουμε ότι, τελικά, είτε θα έχει βρεθεί μια λύση είτε το ελλειψοειδές θα έχει γίνει τόσο μικρό ώστε το εφικτό σύνολο να μην μπορεί παρά να είναι κενό.

Πράγματι, καθένα από τα ελλειψοειδή της παραγόμενης ακολουθίας $\{E_k, k=0,1,2,\dots\}$ ορίζεται έτσι ώστε να περιέχει το ίδιο τμήμα \hat{S} του εφικτού συνόλου S με το αρχικό ελλειψοειδές E_0 . Όπως εξάλλου αποδεικνύεται, αν το εφικτό σύνολο δεν είναι κενό, υπάρχει ένα κάτω φράγμα στον όγκο του \hat{S} . Αν λοιπόν ο όγκος του ελλειψοειδούς γίνει μικρότερος από αυτό το κάτω φράγμα, το εφικτό σύνολο S πρέπει να είναι κενό.

Ας σημειώσουμε τώρα ότι αν E_k είναι το τρέχον ελλειψοειδές και $\alpha_i^t \bar{\mathbf{x}}_k \geq b_i$, $i \in \{1, \dots, m\}$, τότε για κάθε εφικτό σημείο \mathbf{x} πρέπει να ισχύει $\alpha_i^t \mathbf{x} \leq \alpha_i^t \bar{\mathbf{x}}_k$. Έτσι, οι αναδρομικές σχέσεις για την ενημέρωση του κέντρου και του πίνακα του ελλειψοειδούς ορίζουν το ελλειψοειδές ελάχιστου όγκου για το οποίο

$$E_{k+1} \supset E_k \cap \{ \mathbf{x} \mid \alpha_i^t \mathbf{x} \leq \alpha_i^t \bar{\mathbf{x}}_k \}$$

και

$$E_{k+1} \cap \{ \mathbf{x} \mid \alpha_i^t \mathbf{x} = \alpha_i^t \bar{\mathbf{x}}_k \} = E_k \cap \{ \mathbf{x} \mid \alpha_i^t \mathbf{x} = \alpha_i^t \bar{\mathbf{x}}_k \}$$

Οι πιο πάνω σχέσεις φανερώνουν ότι τα ελλειψοειδή E_k και E_{k+1} περιέχουν το ίδιο τμήμα του εφικτού συνόλου.

Εξάλλου, όπως αποδεικνύεται,

$$\text{όγκος}(E_{k+1}) = \gamma(n) \text{όγκος}(E_k)$$

όπου

$$\gamma(n) = \left(\frac{n^2}{n^2-1} \right)^{(n-1)/2} \frac{n}{n+1} < e^{-1/[2(n+1)]}$$

Έτσι, ο όγκος του αρχικού ελλειψοειδούς μειώνεται σε κάθε επανάληψη κατά το σταθερό παράγοντα $\gamma(n)$. Δεδομένου ότι ο όγκος του αρχικού ελλειψοειδούς E_0 φράσσεται από πάνω από τον αριθμό $2^{L(n+1)}$, ο δε όγκος του

τμήματος \hat{S} του εφικτού συνόλου S (όταν το σύνολο S δεν είναι κενό) που περιέχεται στην ακολουθία των ελλειψοειδών $\{E_k, k=0,1,2,\dots\}$, φράσσεται από κάτω από τον αριθμό $2^{-2L(n+1)}$ ευθέως προκύπτει ότι, μετά από $6L(n+1)^2$ το πολύ επαναλήψεις, είτε θα έχει βρεθεί μια εφικτή λύση είτε το εφικτό σύνολο πρέπει να είναι κενό.

Στο σημείο αυτό ας παρατηρήσουμε ότι το κόστος υπολογισμού του $\bar{\mathbf{x}}_{k+1}$, από το $\bar{\mathbf{x}}_k$ είναι $O(mn + n^2)$. Το κόστος αυτό κυριαρχείται από το κόστος για τον εντοπισμό μιας ανισότητας, η οποία παραβιάζεται στο $\bar{\mathbf{x}}_k$, και από το κόστος για τον υπολογισμό του όρου $\mathbf{Q}_k \mathbf{a}_i$. Εξάλλου, ο υπολογισμός του πίνακα \mathbf{Q}_{k+1} , από τον πίνακα \mathbf{Q}_k συνεπάγεται κόστος $O(n^2)$. Έτσι, το κόστος του αλγορίθμου των ελλειψοειδών είναι $O((mn^3 + n^4)L)$. Το κόστος αυτό δείχνει ότι ο αλγόριθμος των ελλειψοειδών έχει πολυωνυμική πολυπλοκότητα, δηλαδή συγκλίνει σε χρόνο πολυωνυμικό ως προς m και n .

Το θεωρητικό πλαίσιο του αλγορίθμου των ελλειψοειδών είναι η μέθοδος η οποία προτάθηκε από τον Shor (1964) για την επίλυση μιας κατηγορίας προβλημάτων κυρτού προγραμματισμού. Η μέθοδος του Shor βελτιώθηκε στη συνέχεια από τους Yudin και Nemirovskii (1976). Ο Khachiyan (1979) ήταν εκείνος ο οποίος απέδειξε ότι αυτή η βελτιωμένη μέθοδος, εξειδικευμένη σε προβλήματα γραμμικού προγραμματισμού, οδηγεί σε έναν αλγόριθμο πολυωνυμικού χρόνου.

Ο ΑΛΓΟΡΙΘΜΟΣ ΤΩΝ ΕΛΛΙΠΣΟΕΙΔΩΝ

| | |
|---|---|
| <p>Βήμα 1 Εκκίνηση του αλγορίθμου</p> | <p>▶ Θέσε $\bar{\mathbf{x}}_0 = \mathbf{0}$, $\mathbf{Q}_0 = 2^L \mathbf{I}$, $k=0$</p> |
| <p>Βήμα 2 Έλεγχος σύγκλισης</p> | <p>▶ Αν $\mathbf{A} \bar{\mathbf{x}}_k < \mathbf{b}$ τερμάτισε: μια εφικτή λύση έχει βρεθεί. Διαφορετικά πήγαινε στο Βήμα 3.</p> |
| <p>Βήμα 3 Έλεγχος ασυμβίβαστου των γραμμικών ανισώσεων</p> | <p>▶ Αν $k > 6L(n+1)^2$, τερμάτισε: το εφικτό σύνολο είναι κενό. Διαφορετικά πήγαινε στο Βήμα 4.</p> |
| <p>Βήμα 4 Προσδιορισμός νέου ελλειψοειδούς</p> | <p>▶ Διάλεξε οποιαδήποτε ανισότητα για την οποία είναι $\alpha_i^t \bar{\mathbf{x}}_k < b_i$, $i \in \{1, \dots, m\}$</p> <p>▶ Θέσε $\bar{\mathbf{x}}_{k+1} = \bar{\mathbf{x}}_k - \frac{1}{n+1} \frac{\mathbf{Q}_k \alpha_i}{\sqrt{\alpha_i^t \mathbf{Q}_k \alpha_i}}$</p> $\mathbf{Q}_{k+1} = \frac{n^2}{n^2-1} \left(\mathbf{Q}_k - \frac{2}{n+1} \frac{\mathbf{Q}_k \alpha_i \alpha_i^t \mathbf{Q}_k}{\alpha_i^t \mathbf{Q}_k \alpha_i} \right)$ |
| <p>Βήμα 5 Επανάληψη της διαδικασίας</p> | <p>▶ Θέσε $k \leftarrow k+1$</p> <p>▶ Επίστρεψε στο Βήμα 2</p> |

4. ΜΕΘΟΔΟΙ ΕΣΩΤΕΡΙΚΟΥ ΣΗΜΕΙΟΥ

Στο κεφάλαιο αυτό παρουσιάζονται οι μέθοδοι εσωτερικού σημείου, οι οποίες στηρίζονται σε μια εντελώς διαφορετική αντίληψη από εκείνη της μεθόδου Simplex. Οι μέθοδοι εσωτερικού σημείου πρωτοεμφανίστηκαν στα μέσα της δεκαετίας του 1980 και έκτοτε έχουν αποτελέσει αντικείμενο έντονης και γόνιμης ερευνητικής δραστηριότητας.

Οι μέθοδοι εσωτερικού σημείου αποτελούν την πιο εντυπωσιακή εξέλιξη στην πρόσφατη ιστορία του γραμμικού προγραμματισμού. Σε αντίθεση με τη μέθοδο Simplex, η οποία μετακινείται στο σύνορο και, πιο συγκεκριμένα, κατά μήκος των ακμών του εφικτού συνόλου, οι μέθοδοι εσωτερικού σημείου παράγουν ακολουθίες σημείων τα οποία προσεγγίζουν τη βέλτιστη λύση από το εσωτερικό του εφικτού συνόλου.

Η ανακοίνωση από τον Karmakar (1984) ενός αλγορίθμου με πολυωνυμικό χρόνο σύγκλισης, του οποίου οι επαναλήψεις διατηρούσαν τα παραγόμενα σημεία στο εσωτερικό του εφικτού συνόλου, άνοιξε νέους και συναρπαστικούς δρόμους έρευνας τόσο στην περιοχή της υπολογιστικής πολυπλοκότητας όσο και στην περιοχή του μαθηματικού προγραμματισμού. Έκτοτε σημειώνεται έντονο ερευνητικό ενδιαφέρον για μεθόδους των οποίων οι επαναλήψεις παράγουν σημεία που να ικανοποιούν ως γνήσιες ανισότητες τουλάχιστον τους περιορισμούς μη αρνητικότητας του προβλήματος.

Η ερευνητική αυτή προσπάθεια οδήγησε στην ανάπτυξη αλγορίθμων οι οποίοι αποδείχθηκαν ανταγωνιστικοί με τη μέθοδο Simplex, τόσο στη θεωρία όσο και στην πράξη. Ειδικότερα μάλιστα για προβλήματα μεγάλης διάστασης, οι αλγόριθμοι εσωτερικού σημείου φαίνεται να υπερέχουν ακόμα και των καλύτερων παραλλαγών της μεθόδου Simplex. Έτσι, αν και αγορά εξακολουθεί ακόμα να κυριαρχείται από κώδικες βασισμένους στη μέθοδο Simplex, έχουν ήδη εμφανιστεί προϊόντα λογισμικού εσωτερικού σημείου, όπως είναι οι κώδικες OB1 και OSL.

Στην πρώτη ενότητα του κεφαλαίου παρουσιάζουμε την πρώτη ιστορικά μέθοδο εσωτερικού, τον αλγόριθμο του Karmarkar. Στη δεύτερη ενότητα σκιαγραφούμε την θεωρητικά και πρακτικά πιο αποτελεσματική μέθοδο της κατηγορίας αυτής, τον πρωτεύοντα-δουικό αλγόριθμο εσωτερικού σημείου.

■ Ο αλγόριθμος του Karmarkar

Η μέθοδος Simplex, έχοντας υποστεί σημαντικό ραφινάρισμα και εκλέπτυνση στην εφαρμογή της, δεν είχε σοβαρό ανταγωνισμό μέχρι το 1984 όταν ο N. Karmarkar πρότεινε τον πολυωνυμικού-χρόνου αλγόριθμο για επίλυση προβλημάτων γραμμικού προγραμματισμού. Τα προβλήματα αυτά είναι της μορφής

Minimize $\mathbf{c}\mathbf{x}$

Με τους περιορισμούς $\mathbf{A}\mathbf{x} = \mathbf{0}$ (4.1)

$$\mathbf{1}\mathbf{x} = 1$$

$$\mathbf{x} \geq \mathbf{0}$$

όπου \mathbf{A} είναι $m \times n$ τάξης m , $n \geq 2$, \mathbf{c} και \mathbf{A} είναι όλοι ακέραιοι, $\mathbf{1}$ είναι ένα διάνυσμα γραμμή με n μονάδες και επιπλέον ισχύουν οι ακόλουθες προϋποθέσεις (A1) και (A2):

(A1) Το σημείο $\mathbf{x}_0 = \left(\frac{1}{n}, \dots, \frac{1}{n} \right)$ είναι εφικτό στο πρόβλημα (4.1).

(A2) Η βέλτιστη αντικειμενική τιμή του προβλήματος (4.1) είναι μηδέν.

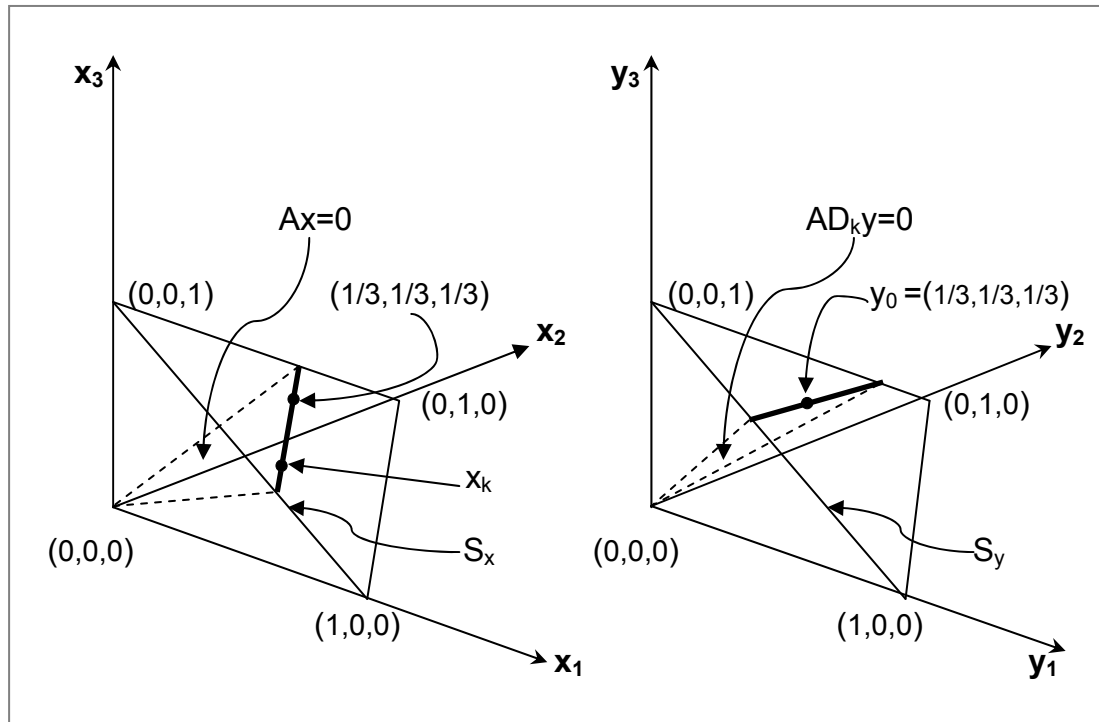
Με μία πρώτη ματιά, η μορφή του γραμμικού προγράμματος (4.1) και οι προϋποθέσεις (A1) και (A2) που το συνοδεύουν μπορεί να φανεί ότι είναι αδικαιολόγητα περιοριστικές. Εντούτοις, οποιοδήποτε γενικό πρόβλημα γραμμικού προγραμματισμού μπορεί να μετατραπεί σε αυτή την μορφή μέσω της χρήσης των τεχνητών μεταβλητών, ενός τεχνητού περιορισμού που θέτει το φράγμα, και μέσω ποικίλων επαναπροσδιορισμών, εάν είναι απαραίτητο. Επιπλέον, η ανάλυση υπό αυτές τις προϋποθέσεις μπορεί να επεκταθεί εύκολα σε πολυωνυμική επίλυση του γενικού προβλήματος γραμμικού προγραμματισμού. Ας προχωρήσουμε στην παραγωγή ενός πολυωνυμικού-χρόνου αλγόριθμου για να λύσουμε το πρόβλημα (4.1) υπό τους περιορισμούς (A1) και (A2). Σημειώστε ότι υπό αυτές τις προϋποθέσεις, το πρόβλημα (4.1) είναι εφικτό και φραγμένο, και ως εκ τούτου έχει βέλτιστη λύση.

Αρχίζοντας από το $x_0 = (1/n, \dots, 1/n)$ και $k = 0$, ο αλγόριθμος εκτελεί το ακόλουθο επαναληπτικό βήμα, λαμβάνοντας ως δεδομένο οποιαδήποτε εφικτή λύση $x_k > 0$. Ορίζουμε τη διαγώνια μήτρα $D_k = \text{diag} \{x_{k1}, \dots, x_{kn}\}$, όπου $x_k = (x_{k1}, \dots, x_{kn})$ και θεωρούμε τον ακόλουθο μετασχηματισμό.

$$y = \frac{D_k^{-1}x}{\mathbf{1}D_k^{-1}x}$$

ο οποίος είναι

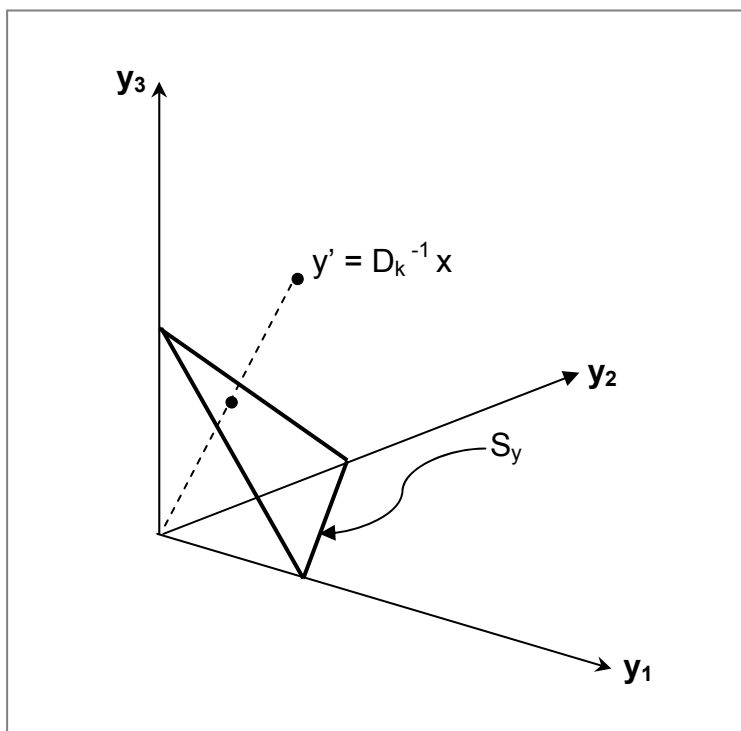
$$y_i = \frac{x_i/x_{ki}}{\sum_{j=1}^n x_j/x_{kj}}, \text{ για } i = 1, \dots, n. \quad (4.2\alpha)$$



Σχήμα 4.1 Προβολικός μετασχηματισμός της εφικτής περιοχής

Μπορούμε να παρατηρήσουμε ότι η εφικτή περιοχή του προβλήματος (4.1) περιγράφεται από τη διατομή του διάστασης $(n-m)$ γραμμικού υποχώρου ορισμένου από το ομοιογενές σύνολο ισοτήτων $Ax=0$, με το διάστασης $(n-1)$ Simplex $S_x = \{x : \mathbf{1}x = 1, x \geq 0\}$. Η διατομή είναι μια περιοχή διάστασης $(n-m-1)$. Το σχήμα 4.1 απεικονίζει μια κατάσταση σε $n = 3$ διαστάσεις με ένα μόνο $(m=1)$ ομοιογενή περιορισμό ισότητας. Υπό τον μετασχηματισμό (4.2α) οποιοδήποτε σημείο στο S_x που μετασχηματίζεται σε ένα σημείο στο $(n-1)$

διαστατικό Simplex $S_y = \{ \mathbf{y} : \mathbf{1}\mathbf{y}=1, \mathbf{y}\geq\mathbf{0} \}$. Συγκεκριμένα, (όπως φαίνεται στο σχήμα 4.1) το τρέχον σημείο \mathbf{x}_k μετασχηματίζεται στο σημείο $\mathbf{y}_0 = (1/n, \dots, 1/n)$, το κέντρο του Simplex S_y . Σημειώνουμε ότι εάν κάποιος ήταν να χρησιμοποιήσει τον ομοπαράλληλο μετασχηματισμό $\mathbf{y}' = \mathbf{D}_k^{-1}\mathbf{x}$, ο οποίος είναι απλά μια κλιμακωτή πράξη, τότε ένα σημείο $\mathbf{x} \in S_x$, θα μετασχηματίζονταν, παραδείγματος χάριν, σε ένα σημείο \mathbf{y}' όπως φαίνεται στο σχήμα 4.2.



Σχήμα 4.2 Προβολικός μετασχηματισμός

Εάν αυτό το σημείο \mathbf{y}' προβάλλεται επάνω στο S_y κατά μήκος της ακτίνας που ορίζεται από την αρχή και το \mathbf{y}' , τότε θα λάβουμε το σημείο \mathbf{y} που δίνεται από το μετασχηματισμό (4.2α), όπως φαίνεται στο σχήμα 4.2. Αυτό ισοδυναμεί με διαίρεση του \mathbf{y}' από το άθροισμα των συνιστωσών του όπως στο μετασχηματισμό (4.2α), έτσι ώστε οι συνιστώσες του \mathbf{y} να αθροίζονται στη μονάδα. Για αυτόν τον λόγο ο μετασχηματισμός (4.2α) είναι γνωστός σαν *προβολικός μετασχηματισμός*. Εκτελώντας αυτές τις πράξεις στην εφικτή περιοχή στο x -διάστημα καταλήγουμε στην εφικτή περιοχή στο y -διάστημα όπως απεικονίζεται στο σχήμα 4.1.

Για να περιγράψουμε αλγεβρικά αυτήν την εφικτή περιοχή στο y -διάστημα, θεωρούμε τον αντίστροφο προβολικό μετασχηματισμό που λαμβάνεται με την

επίλυση ως προς \mathbf{x} στο S_x από το μετασχηματισμό (4.2α). Αυτό δίνει $\mathbf{x} = (\mathbf{D}_k \mathbf{y})(\mathbf{1D}_k^{-1} \mathbf{x})$. Αφού $\mathbf{1x} = 1$, έχουμε $(\mathbf{1D}_k \mathbf{y})(\mathbf{1D}_k^{-1} \mathbf{x}) = 1$. Επομένως θα έχουμε $(\mathbf{1D}_k^{-1} \mathbf{x}) = 1/(\mathbf{1D}_k \mathbf{y})$. Αντικατάσταση αυτού στην προηγούμενη έκφραση για το \mathbf{x} δίνει

$$\mathbf{x} = \frac{\mathbf{D}_k \mathbf{y}}{\mathbf{1D}_k \mathbf{y}} \quad (4.2\beta)$$

Επομένως ο μετασχηματισμός (4.2) χαρτογραφεί σημεία από το Simplex S_x επάνω στο Simplex S_y και αντίστροφα.

Υπό το μετασχηματισμό (4.2), το πρόβλημα (4.1) μετασχηματίζεται στο ακόλουθο πρόβλημα στο y -διάστημα ως προς ένα υποσύνολο του Simplex S_y .

$$\text{Minimize } \left\{ \frac{\mathbf{cD}_k \mathbf{y}}{\mathbf{1D}_k \mathbf{y}} : \mathbf{AD}_k \mathbf{y} = \mathbf{0}, \mathbf{1y} = 1, \mathbf{y} \geq \mathbf{0} \right\} \quad (4.3)$$

Σημειώστε ότι αν και οι περιορισμοί παραμένουν γραμμικοί επειδή το $\mathbf{Ax} = \mathbf{0}$ είναι ομογενές, η αντικειμενική συνάρτηση έχει γίνει ένα πηλίκο γραμμικών συναρτήσεων. Αυτός ο τύπος προβλήματος είναι γνωστός ως ένα *κλασματικό πρόβλημα γραμμικού προγραμματισμού*. Εντούτοις από την υπόθεση (A2), η βέλτιστη αντικειμενική τιμή στην εξίσωση (4.3) είναι μηδέν. Ως εκ τούτου μπορούμε ισοδύναμα να ελαχιστοποιήσουμε τον αριθμητή στο πρόβλημα, δεδομένου ότι ο παρονομαστής είναι θετικός και φραγμένος μακριά από το μηδέν για κάθε $\mathbf{y} \in S_y$. Αυτό οδηγεί στο ακόλουθο πρόβλημα.

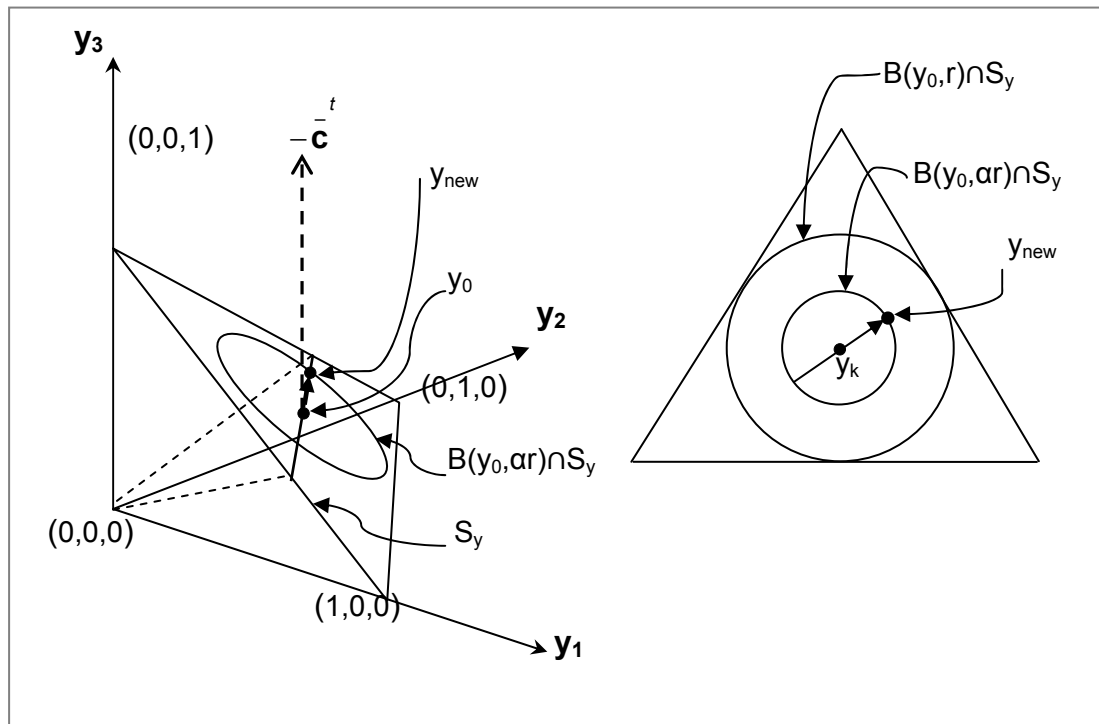
$$\text{Minimize } \{ \bar{\mathbf{c}} \mathbf{y} : \mathbf{Py} = \mathbf{P}_0, \mathbf{y} \geq \mathbf{0} \} \quad (4.4)$$

όπου

$$\bar{\mathbf{c}} = \mathbf{cD}_k, \quad \mathbf{P} = \begin{bmatrix} \mathbf{AD}_k \\ \mathbf{1} \end{bmatrix} \quad \text{και} \quad \mathbf{P}_0 = \begin{bmatrix} \mathbf{0} \\ 1 \end{bmatrix}$$

Αντί να λυθεί το πρόβλημα (4.4), που είναι ισοδύναμο με την επίλυση του αρχικού προβλήματος, θα βελτιστοποιήσουμε ένα συγκεκριμένο, απλούστερο περιορισμό αυτού του προβλήματος. Αυτό θα παράγει μια νέα εφικτή λύση. Επαναλαμβάνοντας το βήμα επαναληπτικά, θα είμαστε σε θέση να λάβουμε πολυωνυμικά μια βέλτιστη λύση στο αρχικό πρόβλημα.

Για να ορίσουμε αυτόν τον περιορισμό, ας θεωρήσουμε την n-διάστασης σφαίρα ή τη σφαίρα $B(y_0, r)$ με κέντρο στο $y_0 = (1/n, \dots, 1/n)$ και με την κατάλληλη ακτίνα r έτσι ώστε η διατομή αυτής της σφαίρας με το Simplex περιορισμό $\{ \mathbf{y} : \mathbf{1y} = 1 \}$ να είναι μια σφαίρα διάστασης $(n - 1)$ με το ίδιο κέντρο και ακτίνα, τα οποία είναι εγγραμμένα στο S_y (σχήμα 4.3). Σημειώστε ότι αυτή η ακτίνα r είναι η απόσταση από το κέντρο $(1/n, \dots, 1/n)$ του Simplex μέχρι το κέντρο μιας από τις πλευρές του. Ωστόσο οι πλευρές του S_y είναι απλά μιας χαμηλότερης διάστασης Simplicies. Ως εκ τούτου εξετάζοντας την πλευρά στην οποία η συνιστώσα y_1 είναι μηδέν, για παράδειγμα, λαμβάνουμε το r ως την απόσταση από το $(1/n, \dots, 1/n)$ στο $(0, 1/(n-1), \dots, 1/(n-1))$. Αυτό δίνει $r = 1 / \sqrt{n(n-1)}$.



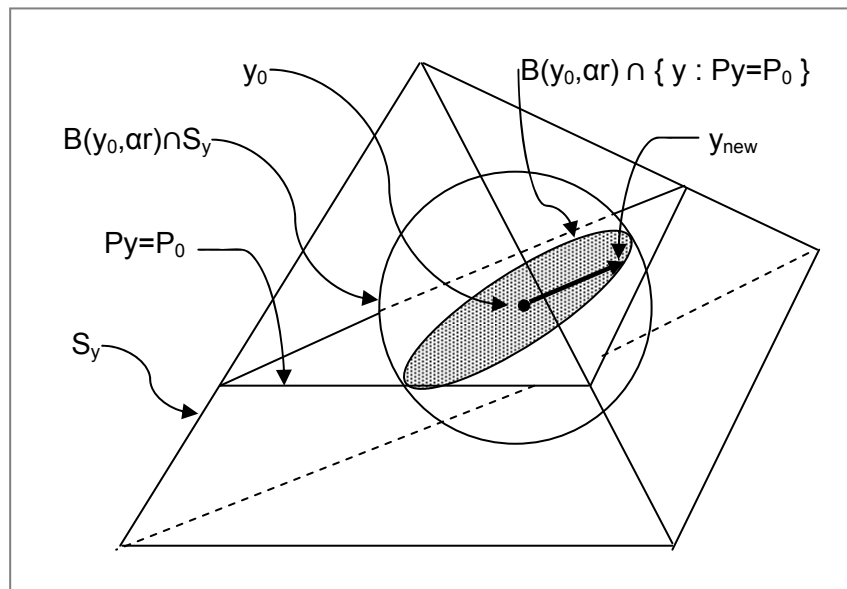
Σχήμα 4.3 Βασικό βήμα του αλγορίθμου του Karmarkar

Θεωρήστε έναν περιορισμό της εξίσωσης (4.4) στον οποίο ψάχνουμε να ελαχιστοποιήσουμε το $\bar{\mathbf{c}}\mathbf{y}$ υπό τον όρο ότι $\mathbf{Py} = \mathbf{P}_0$, $\mathbf{y} \geq \mathbf{0}$ και $\mathbf{y} \in B(\mathbf{y}_0, \alpha r)$, όπου $0 < \alpha < 1$ είναι μια σταθερά. (Τη συγκεκριμένη επιλογή του α θα την εξετάσουμε αργότερα.) Σημειώστε ότι έχουμε συρρικνώσει την εγγεγραμμένη σφαίρα κατά ένα παράγοντα α , και έτσι η ακτίνα της είναι αr . Επιπλέον παραμένοντας εφικτό σε αυτό τον περιορισμό, η επανάληψη παραμένει

αυστηρά θετική. Αφού $\mathbf{y} \geq \mathbf{0}$ συνεπάγεται από την τομή της σφαίρας και του Simplex περιορισμού $\{\mathbf{y} : \mathbf{1}\mathbf{y} = 1\}$, αυτός ο περιορισμός είναι ισοδύναμος με το πρόβλημα

$$\text{Minimize } \left\{ \bar{\mathbf{c}}\mathbf{y} : \mathbf{P}\mathbf{y} = \mathbf{P}_0, (\mathbf{y}-\mathbf{y}_0)^t(\mathbf{y}-\mathbf{y}_0) \leq \alpha^2 r^2 \right\} \quad (4.5)$$

όπου $\{\mathbf{y} : (\mathbf{y}-\mathbf{y}_0)^t(\mathbf{y}-\mathbf{y}_0) \leq \alpha^2 r^2\}$ περιγράφει τη σφαίρα $B(\mathbf{y}_0, \alpha r)$. Το σχήμα 4.3 δείχνει την εφικτή περιοχή του πάνω από δύο διαστάσεις Simplex.



Σχήμα 4.4 Εικονογράφηση του αλγορίθμου του Karmarkar

Παρατηρήστε ότι το σύστημα $\mathbf{P}\mathbf{y} = \mathbf{P}_0$ ορίζει ένα ομοπαράλληλο υποχώρο διάστασης $(n-m-1)$, που περνά από το κέντρο της σφαίρας $B(\mathbf{y}_0, \alpha r)$. Ως εκ τούτου, η εφικτή περιοχή στο πρόβλημα (4.5) είναι μια σφαίρα διάστασης $(n-m-1)$ που κεντροθετείται στο \mathbf{y}_0 . Το σχήμα 4.4 απεικονίζει μια κατάσταση για $n=4$, $m=1$ για $n-1 = 3$ διαστάσεων Simplex. Πρέπει να είναι εμφανές ότι η βέλτιστη λύση στο πρόβλημα (4.5) λαμβάνεται με απλή προβολή της αρνητικής κλίσης της αντικειμενικής συνάρτησης $-\bar{\mathbf{c}}^t$, που κεντροθετείται στο \mathbf{y}_0 , επάνω στο μηδενικό διάστημα ή την επιφάνεια περιορισμού του $\mathbf{P}\mathbf{y} = \mathbf{P}_0$ και μετακινούμενοι από \mathbf{y}_0 κατά μήκος αυτής της προβαλλόμενης κατεύθυνσης

προς το όριο της σφαίρας $B(y_0, ar)$. (Σχήματα 4.3 και 4.4). Παρατηρήστε ότι η απλότητα αυτής της διαδικασίας επίλυσης, είναι συνέπεια του γεγονότος ότι βελτιστοποιούμε μια γραμμική συνάρτηση πάνω σε μια σφαίρα και ότι η τρέχουσα θέση είναι στο κέντρο αυτής της σφαίρας. Δηλώνοντας την προβολή της κλίσης του διανύσματος $\bar{\mathbf{c}}^t$ ως \mathbf{c}_p και το βέλτιστο στο πρόβλημα (4.5) σαν \mathbf{y}_{new} , παίρνουμε

$$\mathbf{y}_{new} = \mathbf{y}_0 - ar \frac{\mathbf{c}_p}{\|\mathbf{c}_p\|} \quad (4.6a)$$

[Σημειώνουμε ότι αν $\mathbf{c}_p = \mathbf{0}$, τότε οποιαδήποτε εφικτή λύση είναι βέλτιστη και μπορούμε να τερματίσουμε με x_k σαν βέλτιστο στο πρόβλημα (4.1)].

Προκειμένου να υπολογιστεί το \mathbf{c}_p , παρατηρήστε ότι αφού το \mathbf{c}_p βρίσκεται στην επιφάνεια του περιορισμού $\mathbf{P}\mathbf{y} = \mathbf{P}_0$, το διάνυσμα $(\bar{\mathbf{c}}^t - \mathbf{c}_p)$ περιλαμβάνεται στο διάστημα που καλύπτεται από τις κλίσεις στο $\mathbf{P}\mathbf{y} = \mathbf{P}_0$. Ως εκ τούτου υπάρχει ένα διάνυσμα $\bar{\mathbf{w}}$ τέτοιο ώστε $\mathbf{P}^t \bar{\mathbf{w}} = \bar{\mathbf{c}}^t - \mathbf{c}_p$. Πολλαπλασιάζοντας και τις δύο πλευρές με το \mathbf{P} , παίρνουμε $\mathbf{P}\mathbf{P}^t \bar{\mathbf{w}} = \mathbf{P}\bar{\mathbf{c}}^t$, αφού εξ ορισμού $\mathbf{P}\mathbf{c}_p = \mathbf{0}$. Δεδομένου ότι το \mathbf{A} είναι πλήρους γραμμής τάξη και $x_k > 0$, ο τετραγωνικός πίνακας $\mathbf{P}\mathbf{P}^t$ είναι αντιστρεπτός, το οποίο δίνει $\bar{\mathbf{w}} = (\mathbf{P}\mathbf{P}^t)^{-1} \mathbf{P}\bar{\mathbf{c}}^t$. Επομένως έχουμε $\mathbf{c}_p = \bar{\mathbf{c}}^t - \mathbf{P}^t \bar{\mathbf{w}}$ που γράφεται

$$\mathbf{c}_p = [\mathbf{I} - \mathbf{P}^t (\mathbf{P}\mathbf{P}^t)^{-1} \mathbf{P}] \bar{\mathbf{c}}^t \quad (4.6\beta)$$

Σημειώνουμε ότι υπάρχει και εναλλακτική προσέγγιση για να δείξουμε ότι η βέλτιστη λύση για το πρόβλημα (4.5) δίνεται από τις εξισώσεις (4.6).

Επισημαίνουμε εδώ ότι η λύση $\bar{\mathbf{w}}$ είναι η λύση για το πρόβλημα ελάχιστων τετραγώνων ελαχιστοποίησης του $\left\| \bar{\mathbf{c}}^t - \mathbf{P}^t \mathbf{w} \right\|^2$ πάνω στο $\mathbf{w} \in E^{(m+1)}$. Αντί του να υπολογίσουμε ρητά το \mathbf{c}_p χρησιμοποιώντας την εξίσωση (4.6β), μπορούμε να το «αποκτήσουμε» διαιρώντας μια λύση $\bar{\mathbf{w}}$ σε αυτό το πρόβλημα με οποιαδήποτε μέθοδο και να θέσουμε $\mathbf{c}_p = \bar{\mathbf{c}}^t - \mathbf{P}^t \bar{\mathbf{w}}$. Σε κάθε περίπτωση, ο τύπος (4.6β) δίνει το \mathbf{c}_p .

Δοθέντος του \mathbf{y}_{new} στην εξίσωση (4.6α), το αντίστοιχο αναθεωρημένο

διάνυσμα x_{k+1} στο x -διάστημα λαμβάνεται μέσω του αντίστροφου μετασχηματισμού (4.2β) ως

$$x_{k+1} = \frac{D_k y_{new}}{1 D_k y_{new}} \quad (4.7)$$

Ας παρατηρήσουμε ότι $y_{new} > 0$ δεδομένου ότι βρίσκεται στο εσωτερικό της σφαίρας διάστασης $(n - 1)$ που εγγράφεται στο S_y . Ως εκ τούτου στην εξίσωση (4.7) θα είναι $x_{k+1} > 0$. Στο σημείο αυτό ολοκληρώνεται μια επανάληψη και το βήμα μπορεί να επαναληφθεί αφού αυξήσουμε το k κατά ένα. Όπως θα δούμε στην επόμενη παράγραφο, η αντικειμενική τιμή στο πρόβλημα (4.1) που παράγεται με την ακολουθία των επαναλήψεων προσεγγίζει τη βέλτιστη αντικειμενική τιμή του μηδενός οριακά καθώς $k \rightarrow \infty$, αν και όχι απαραίτητα με μονότονο τρόπο. Ως εκ τούτου για πρακτικούς λόγους, αυτή η διαδικασία μπορεί να τερματισθεί όταν η αντικειμενική τιμή φτάσει αρκετά κοντά στο μηδέν.

Για θεωρητικούς λόγους, θα εξετάσουμε το ακόλουθο κατώτερο φράγμα L στο μήκος των δεδομένων εισόδου για το πρόβλημα (4.1):

$$L = \left\lceil 1 + \log(1 + |c_{j_{max}}|) + \log(|\det_{max}|) \right\rceil$$

όπου $|c_{j_{max}}|$ είναι η μεγαλύτερη αριθμητική τιμή οποιουδήποτε συντελεστή κόστους c_j και $|\det_{max}|$ είναι η μεγαλύτερη αριθμητική τιμή της ορίζουσας οποιασδήποτε βάσης για το πρόβλημα (4.1). Όπως προκύπτει από μια ιδιότητα των οριζουσών το L είναι όντως κατώτερο φράγμα στο μήκος εισόδου για το πρόβλημα (4.1) αφού $\log(|\det_{max}|) \leq \log(1+n) + \sum_i \sum_j \log(1+|a_{ij}|)$.

(Σημειώνουμε ότι μια εναλλακτική, αν και μεγαλύτερη, τιμή του L δίνεται με την αντικατάσταση του $\log(|\det_{max}|)$ με τον τελευταίο όρο). Στο επόμενο κομμάτι θα δείξουμε ότι χρησιμοποιώντας μια τιμή $\alpha = (n-1)/3n$, ο αλγόριθμος παράγει μια λύση της αντικειμενικής τιμής μικρότερης από 2^{-L} σε $10nL$ επαναλήψεις με μια «προσπάθεια» πολυωνυμικής πολυπλοκότητας $O(n^{3.5}L)$. Έτσι, η ακόλουθη πολυωνυμική διαδικασία μπορεί να μας δώσει μια ακριβή βέλτιστη λύση ακραίου σημείου του προβλήματος (4.1).

Αρχίζοντας από την τελική επανάληψη x_k που λαμβάνεται με την αντικειμενική τιμή $cx_k < 2^{-L}$, αυτή η διαδικασία βρίσκει μια λύση ακραίου

σημείου με τουλάχιστον μια καλή αντικειμενική τιμή χρησιμοποιώντας την ακόλουθη μέθοδο που είναι γνωστή ως *purification scheme*. Αν n γραμμικά ανεξάρτητοι περιορισμοί είναι δεσμευμένοι στο \mathbf{x}_k , τότε είναι ήδη μια βασική εφικτή λύση. Διαφορετικά υπάρχει μια κατεύθυνση $\mathbf{d} \neq \mathbf{0}$ που βρίσκεται στο μηδενικό διάστημα των δεσμευμένων περιορισμών, που ικανοποιεί το ομογενές σύστημα ισότητας, που αντιστοιχεί στους δεσμευτικούς περιορισμούς. Η μέθοδος κινεί τώρα την τρέχουσα επανάληψη κατά μήκος της κατεύθυνσης \mathbf{d} αν $\mathbf{c}\mathbf{d} < 0$ και κατά μήκος της κατεύθυνσης $-\mathbf{d}$ ειδάλλως, μέχρι κάποιος περιορισμός να μπλοκάρει περαιτέρω κίνηση. Αυτό πρέπει να συμβεί δεδομένου ότι η εφικτή περιοχή είναι φραγμένη. Σημειώστε ότι στη νέα λύση η αντικειμενική τιμή δεν είναι μεγαλύτερη από $\mathbf{c}\mathbf{x}_k < 2^{-L}$ και τουλάχιστον ένας πρόσθετος γραμμικά ανεξάρτητος περιορισμός είναι δεσμευμένος.

Προχωρώντας με αυτό τον τρόπο, μια βασική εφικτή λύση $\bar{\mathbf{x}}$ στο πρόβλημα (4.1) μπορεί να ληφθεί με την αντικειμενική τιμή αυστηρά μικρότερη από 2^{-L} .

Σημειώνουμε ότι αυτή η διαδικασία περιλαμβάνει το ανώτερο $n - (m+1)$ τέτοια βήματα, δεδομένου ότι αρχίζει με $(m+1)$ γραμμικά ανεξάρτητους περιορισμούς ισότητας στο πρόβλημα (4.1) και μετά προσθέτει τουλάχιστον έναν επιπλέον γραμμικά ανεξάρτητο δεσμευμένο περιορισμό σε αυτό το σύνολο σε κάθε βήμα. Επιπλέον, κάθε βήμα είναι πολυωνυμικής πολυπλοκότητας. Ως εκ τούτου το $\bar{\mathbf{x}}$ καθορίζεται από την τελική επανάληψη \mathbf{x}_k σε πολυωνυμικό χρόνο (με ένα πολυωνυμικό φράγμα τάξης $O(m^2n)$).

Τώρα θα δείξουμε ότι η προκύπτουσα λύση ακραίου σημείου $\bar{\mathbf{x}}$ είναι πράγματι βέλτιστη για το πρόβλημα (4.1). Η διαδικασία που ακολουθούμε για αυτό το σκοπό, είναι γνωστή ως ρουτίνα βέλτιστης στρογγυλοποίησης. Δεδομένου ότι τα δεδομένα είναι όλα ακέραιοι αριθμοί, οποιαδήποτε βασική εφικτή λύση για το πρόβλημα (4.1) με τη βάση \mathbf{B} έχει αντικειμενική τιμή " $\mathbf{c}_B \mathbf{B}^{-1} \mathbf{b}$ " που δίνεται από το N/D , όπου N και το D είναι ακέραιοι αριθμοί, και όπου $D \equiv |\det \mathbf{B}|$. Αλλά $D < 2^L$, όπου το L ορίζεται όπως προηγουμένως. Ως εκ τούτου από την υπόθεση (A2), οποιαδήποτε μη βέλτιστη λύση ακραίου σημείου για το πρόβλημα (4.1) πρέπει να έχει μια θετική αντικειμενική τιμή $N/D \geq 1/D > 2^{-L}$. Αφού $\mathbf{c}\bar{\mathbf{x}} < 2^{-L}$, πρέπει επομένως $\bar{\mathbf{x}}$ να είναι βέλτιστο για το πρόβλημα (4.1) με $\mathbf{c}\bar{\mathbf{x}} = 0$.

Ο ΑΛΓΟΡΙΘΜΟΣ ΤΟΥ ΚΑΡΜΑΡΚΑΡ

ΕΙΣΑΓΩΓΗ

Υπολογίστε $r=1/\sqrt{n(n-1)}$, $L = \left\lceil 1 + \log(1+|c_{j_{\max}}|) + \log(|\det_{\max}|) \right\rceil$

και επιλέξτε $\alpha = (n-1)/3n$.

Έστω $\mathbf{x}_0 = (1/n, \dots, 1/n)$,

θέστε $k=0$

ΚΥΡΙΟ ΒΗΜΑ

Αν το $\mathbf{c}\mathbf{x}_k < 2^{-L}$ χρησιμοποιείτε τη ρουτίνα βέλτιστης στρογγυλοποίησης για να καθορίσετε μια βέλτιστη λύση, και σταματήστε. (Ουσιαστικά, δεδομένου ότι 2^{-L} μπορεί να είναι πολύ μικρό, κάποιος μπορεί να σταματήσει όταν το $\mathbf{c}\mathbf{x}_k$ είναι μικρότερο από κάποια άλλη επιθυμητή ανοχή.)

Διαφορετικά ορίστε $\mathbf{D}_k = \text{diag} \{ x_{k1}, \dots, x_{kn} \}$, $\mathbf{y}_0 = \left(\frac{1}{n}, \dots, \frac{1}{n} \right)^t$

$$\mathbf{P} = \begin{bmatrix} \mathbf{A}\mathbf{D}_k \\ \mathbf{1} \end{bmatrix} \quad \text{και} \quad \bar{\mathbf{c}} = \mathbf{c}\mathbf{D}_k$$

και υπολογίστε

$$\mathbf{y}_{\text{new}} = \mathbf{y}_0 - \alpha \frac{\mathbf{c}_p}{\|\mathbf{c}_p\|}, \quad \text{όπου } \mathbf{c}_p = [\mathbf{I} - \mathbf{P}^t (\mathbf{P}\mathbf{P}^t)^{-1} \mathbf{P}] \bar{\mathbf{c}}^t$$

Έτσι, θα λάβετε $\mathbf{x}_{k+1} = (\mathbf{D}_k \mathbf{y}_{\text{new}}) / (\mathbf{1}\mathbf{D}_k \mathbf{y}_{\text{new}})$.

Αυξήστε το k κατά ένα και επαναλάβετε το κύριο βήμα.

ΡΟΥΤΙΝΑ ΒΕΛΤΙΣΤΗΣ ΣΤΡΟΓΓΥΛΟΠΟΙΗΣΗΣ

Αρχίζοντας από \mathbf{x}_k καθορίστε μια ακραίου σημείου λύση $\bar{\mathbf{x}}$ για το πρόβλημα (4.1) με $\mathbf{c}\bar{\mathbf{x}} \leq \mathbf{c}\mathbf{x}_k < 2^{-L}$ χρησιμοποιώντας το purification scheme.

Ολοκληρώστε με $\bar{\mathbf{x}}$ ως βέλτιστη λύση στο πρόβλημα (4.1).

■ Ο πρωτεύων-δυσικός αλγόριθμος εσωτερικού σημείου

Ας θεωρήσουμε το πρωτεύον πρόβλημα γραμμικού προγραμματισμού

$$\begin{array}{ll} \max & z = \mathbf{c}^t \mathbf{x} \\ \text{με τους περιορισμούς} & \mathbf{Ax} = \mathbf{b}, \mathbf{x} \geq \mathbf{0} \end{array} \quad (\Pi)$$

και το δυσικό του

$$\begin{array}{ll} \min & y = \mathbf{b}^t \mathbf{w} \\ \text{με τους περιορισμούς} & \mathbf{A}^t \mathbf{w} \geq \mathbf{c} \end{array} \quad (\Delta)$$

Μετά την αναγωγή του δυσικού σε τυπική μορφή με τη βοήθεια των μεταβλητών περιθωρίου \mathbf{s} , το πρόβλημα αυτό γράφεται

$$\begin{array}{ll} \min & y = \mathbf{b}^t \mathbf{w} \\ \text{με τους περιορισμούς} & \mathbf{A}^t \mathbf{w} - \mathbf{s} = \mathbf{c}, \mathbf{s} \geq \mathbf{0} \end{array}$$

Έστω τώρα \mathbf{x}_k , μια εφικτή λύση του πρωτεύοντος προβλήματος (Π) και $(\mathbf{w}_k, \mathbf{s}_k)$ μια εφικτή λύση του δυσικού του (Δ). Οι λύσεις αυτές είναι βέλτιστες εφόσον ικανοποιούν τις συμπληρωματικές συνθήκες χαλαρότητας

$$(\mathbf{A}^t \mathbf{w}_k)_j > c_j \Rightarrow (\mathbf{x}_k)_j = 0 \quad (j = 1, \dots, n)$$

ή ισοδύναμα

$$(\mathbf{x}_k)_j (\mathbf{s}_k)_j = 0 \quad (j = 1, \dots, n) \quad (4.8)$$

Η κεντρική ιδέα πίσω από τον πρωτεύοντα-δυσικό αλγόριθμο εσωτερικού σημείου είναι η παραγωγή μιας ακολουθίας $\{\mathbf{x}_k, (\mathbf{w}_k, \mathbf{s}_k)\}$ ζευγών λύσεων οι οποίες να κινούνται στο εσωτερικό του εφικτού συνόλου του κάθε προβλήματος και να ικανοποιούν σε ολοένα και μεγαλύτερο βαθμό τις συμπληρωματικές συνθήκες χαλαρότητας. Ειδικότερα, κάθε επανάληψη του αλγορίθμου κατασκευάζει διανύσματα $\mathbf{x}_k(\theta_k)$, $\mathbf{w}_k(\theta_k)$ και $\mathbf{s}_k(\theta_k)$ τα οποία, για κάποια θετική τιμή της παραμέτρου θ_k , ικανοποιούν τις σχέσεις

$$\left. \begin{aligned} \mathbf{Ax}_k &= \mathbf{b}, \mathbf{x}_k \geq \mathbf{0} \\ \mathbf{A}^t \mathbf{w}_k - \mathbf{s}_k &= \mathbf{c}, \mathbf{s}_k \geq \mathbf{0} \\ (\mathbf{x}_k)_j (\mathbf{s}_k)_j &= \theta_k \quad (j = 1, \dots, n) \end{aligned} \right\} \quad (\text{ΠΔ})$$

Η τιμή της παραμέτρου θ_k από επανάληψη σε επανάληψη ελαττώνεται, μέχρις ότου επιτευχθεί ικανοποιητικός βαθμός σύγκλισης. Είναι φανερό ότι για $\theta_k > 0$, η συνθήκη $(\mathbf{x}_k)_j (\mathbf{s}_k)_j = \theta_k$ εγγυάται ότι $(\mathbf{x}_k)_j > 0$ και $(\mathbf{s}_k)_j > 0$ ($j = 1, \dots, n$), δηλαδή ότι τα σημεία $\mathbf{x}_k(\theta_k)$ και $\mathbf{w}_k(\theta_k)$ παραμένουν στο εσωτερικό του εφικτού συνόλου του κάθε προβλήματος. Σημειώστε ότι εισάγοντας τους διαγώνιους πίνακες

$$\mathbf{X}_k = \begin{bmatrix} (\mathbf{x}_k)_1 & & & & \\ & \ddots & & & \\ & & (\mathbf{x}_k)_j & & \\ & & & \ddots & \\ & & & & (\mathbf{x}_k)_n \end{bmatrix}, \quad \mathbf{S}_k = \begin{bmatrix} (\mathbf{s}_k)_1 & & & & \\ & \ddots & & & \\ & & (\mathbf{s}_k)_j & & \\ & & & \ddots & \\ & & & & (\mathbf{s}_k)_n \end{bmatrix}$$

μπορούμε να θέσουμε τις συμπληρωματικές συνθήκες χαλαρότητας στη διανυσματική μορφή

$$\mathbf{X}_k \mathbf{S}_k \mathbf{e} = \theta_k \mathbf{e} \quad (4.9)$$

Έστω τώρα $\mathbf{x}_k > \mathbf{0}$, \mathbf{w}_k και $\mathbf{s}_k > \mathbf{0}$ οι τρέχουσες εκτιμήσεις των βέλτιστων λύσεων. Οι λύσεις αυτές, ως εφικτές, ικανοποιούν τις περιοριστικές εξισώσεις $\mathbf{Ax}_k = \mathbf{b}$ και $\mathbf{A}^t \mathbf{w}_k - \mathbf{s}_k = \mathbf{c}$ των δύο προβλημάτων. Ο πρωτεύων-δουϊκός αλγόριθμος εσωτερικού σημείου βρίσκει μετατοπίσεις $\Delta \mathbf{x}_k$, $\Delta \mathbf{w}_k$ και $\Delta \mathbf{s}_k$ τέτοιες ώστε τα διανύσματα $\mathbf{x}_k + \Delta \mathbf{x}_k$, $\mathbf{w}_k + \Delta \mathbf{w}_k$ και $\mathbf{s}_k + \Delta \mathbf{s}_k$ να ικανοποιούν με μεγαλύτερη ακρίβεια τις συμπληρωματικές συνθήκες χαλαρότητας.

Εδώ ας παρατηρήσουμε ότι τα διανύσματα $\mathbf{x}_k + \Delta \mathbf{x}_k$, $\mathbf{w}_k + \Delta \mathbf{w}_k$ και $\mathbf{s}_k + \Delta \mathbf{s}_k$ πρέπει κατ' αρχάς να είναι εφικτά για τα δύο προβλήματα. Πρέπει δηλαδή να ικανοποιούν τις σχέσεις

$$\mathbf{A}(\mathbf{x}_k + \Delta \mathbf{x}_k) = \mathbf{b} \Rightarrow \mathbf{Ax}_k + \mathbf{A}\Delta \mathbf{x}_k = \mathbf{b} \Rightarrow \mathbf{A}\Delta \mathbf{x}_k = \mathbf{0}$$

$$\mathbf{A}^t(\mathbf{w}_k + \Delta \mathbf{w}_k) - (\mathbf{s}_k + \Delta \mathbf{s}_k) = \mathbf{c} \Rightarrow \mathbf{A}^t \mathbf{w}_k - \mathbf{s}_k + \mathbf{A}^t \Delta \mathbf{w}_k - \Delta \mathbf{s}_k = \mathbf{c} \Rightarrow \mathbf{A}^t \Delta \mathbf{w}_k - \Delta \mathbf{s}_k = \mathbf{0}$$

Αυτά είναι δύο συστήματα γραμμικά ως προς $\Delta \mathbf{x}_k$, $\Delta \mathbf{w}_k$ και $\Delta \mathbf{s}_k$.

Τα διανύσματα $\mathbf{x}_k + \Delta\mathbf{x}_k$ και $\mathbf{s}_k + \Delta\mathbf{s}_k$ πρέπει επιπλέον να ικανοποιούν με μεγαλύτερη ακρίβεια και τις συμπληρωματικές συνθήκες χαλαρότητας, δηλαδή τις σχέσεις

$$\begin{aligned} & [(\mathbf{x}_k)_j + (\Delta\mathbf{x}_k)_j] [(\mathbf{s}_k)_j + (\Delta\mathbf{s}_k)_j] = \theta_{k+1} \Rightarrow \\ \Rightarrow & (\mathbf{s}_k)_j (\Delta\mathbf{x}_k)_j + (\mathbf{x}_k)_j (\Delta\mathbf{s}_k)_j + (\Delta\mathbf{x}_k)_j (\Delta\mathbf{s}_k)_j = \theta_{k+1} - (\mathbf{x}_k)_j (\mathbf{s}_k)_j \quad (j = 1, \dots, n) \end{aligned}$$

όπου $\theta_{k+1} < \theta_k$. Από τις πιο πάνω σχέσεις μπορούμε να παραλείψουμε τον όρο $(\Delta\mathbf{x}_k)_j (\Delta\mathbf{s}_k)_j$ με το σκεπτικό ότι οι μετατοπίσεις $\Delta\mathbf{x}_k$ και $\Delta\mathbf{s}_k$ είναι πολύ μικρές και συνεπώς ο όρος $(\Delta\mathbf{x}_k)_j (\Delta\mathbf{s}_k)_j$ θα πρέπει να είναι ακόμα πιο μικρός, σε κάθε περίπτωση μάλιστα μικρότερος συγκριτικά με τους όρους $(\mathbf{s}_k)_j (\Delta\mathbf{x}_k)_j$ και $(\mathbf{x}_k)_j (\Delta\mathbf{s}_k)_j$. Έτσι έχουμε

$$(\mathbf{s}_k)_j (\Delta\mathbf{x}_k)_j + (\mathbf{x}_k)_j (\Delta\mathbf{s}_k)_j = \theta_{k+1} - (\mathbf{x}_k)_j (\mathbf{s}_k)_j \quad (j = 1, \dots, n)$$

και σε διανυσματική μορφή

$$\mathbf{S}_k \Delta\mathbf{x}_k + \mathbf{X}_k \Delta\mathbf{s}_k = \theta_{k+1} \mathbf{e} - \mathbf{X}_k \mathbf{S}_k \mathbf{e}$$

Επομένως, ο προσδιορισμός μετατοπίσεων $\Delta\mathbf{x}_k$, $\Delta\mathbf{w}_k$ και $\Delta\mathbf{s}_k$ τέτοιων ώστε τα σημεία $\mathbf{x}_k + \Delta\mathbf{x}_k$, $\mathbf{w}_k + \Delta\mathbf{w}_k$ και $\mathbf{s}_k + \Delta\mathbf{s}_k$ να ικανοποιούν τις πρωτεύουσες δεικές συνθήκες (ΠΔ) ανάγεται στην επίλυση των γραμμικών συστημάτων

$$\mathbf{A} \Delta\mathbf{x}_k = \mathbf{0}, \quad \mathbf{A}^t \Delta\mathbf{w}_k - \Delta\mathbf{s}_k = \mathbf{0}, \quad \mathbf{S}_k \Delta\mathbf{x}_k + \mathbf{X}_k \Delta\mathbf{s}_k = \theta_{k+1} \mathbf{e} - \mathbf{X}_k \mathbf{S}_k \mathbf{e} \quad (4.10)$$

Από τη δεύτερη των σχέσεων (4.10) προκύπτει ότι $\Delta\mathbf{s}_k = \mathbf{A}^t \Delta\mathbf{w}_k$. Αντικαθιστώντας την έκφραση αυτή για το $\Delta\mathbf{s}_k$, στην τρίτη των σχέσεων (4.10) λαμβάνουμε

$$\mathbf{S}_k \Delta\mathbf{x}_k + \mathbf{X}_k \mathbf{A}^t \Delta\mathbf{w}_k = \theta_{k+1} \mathbf{e} - \mathbf{X}_k \mathbf{S}_k \mathbf{e} \quad (4.11)$$

Πολλαπλασιάζοντας τη σχέση (4.11) από τα αριστερά με τον πίνακα $\mathbf{A} \mathbf{S}_k^{-1}$ και αναγνωρίζοντας ότι $\mathbf{A} \Delta\mathbf{x}_k = \mathbf{0}$, έχουμε

$$\mathbf{A} \Delta\mathbf{x}_k + \mathbf{A} \mathbf{S}_k^{-1} \mathbf{X}_k \mathbf{A}^t \Delta\mathbf{w}_k = \mathbf{A} \mathbf{S}_k^{-1} (\theta_{k+1} \mathbf{e} - \mathbf{X}_k \mathbf{S}_k \mathbf{e}) \Rightarrow$$

$$\Rightarrow \mathbf{A} \mathbf{S}_k^{-1} \mathbf{X}_k \mathbf{A}^t \Delta \mathbf{w}_k = \mathbf{A} \mathbf{S}_k^{-1} (\theta_{k+1} \mathbf{e} - \mathbf{X}_k \mathbf{S}_k \mathbf{e})$$

Ορίζοντας τώρα το διαγώνιο πίνακα \mathbf{D}_k και το διάνυσμα $\mathbf{v}_k(\theta_{k+1})$, όπου

$$\mathbf{D}_k = \mathbf{S}_k^{-1} \mathbf{X}_k, \mathbf{v}_k(\theta_{k+1}) = \theta_{k+1} \mathbf{e} - \mathbf{X}_k \mathbf{S}_k \mathbf{e}$$

προκύπτει

$$(\mathbf{A} \mathbf{D}_k \mathbf{A}^t) \Delta \mathbf{w}_k = \mathbf{A} \mathbf{S}_k^{-1} \mathbf{v}_k(\theta_{k+1}) \Rightarrow \Delta \mathbf{w}_k = (\mathbf{A} \mathbf{D}_k \mathbf{A}^t)^{-1} \mathbf{A} \mathbf{S}_k^{-1} \mathbf{v}_k(\theta_{k+1})$$

$$\Delta \mathbf{s}_k = \mathbf{A}^t \Delta \mathbf{w}_k \Rightarrow \Delta \mathbf{s}_k = \mathbf{A}^t (\mathbf{A} \mathbf{D}_k \mathbf{A}^t)^{-1} \mathbf{A} \mathbf{S}_k^{-1} \mathbf{v}_k(\theta_{k+1})$$

$$\mathbf{S}_k \Delta \mathbf{x}_k + \mathbf{X}_k \Delta \mathbf{s}_k = \mathbf{v}_k(\theta_{k+1}) \Rightarrow \Delta \mathbf{x}_k = \mathbf{S}_k^{-1} \mathbf{v}_k(\theta_{k+1}) - \mathbf{S}_k^{-1} \mathbf{X}_k \Delta \mathbf{s}_k \Rightarrow$$

$$\Delta \mathbf{x}_k = \mathbf{S}_k^{-1} \mathbf{v}_k(\theta_{k+1}) - \mathbf{D}_k \Delta \mathbf{s}_k$$

Όπως αποδεικνύεται, αν η παράμετρος θ από επανάληψη σε επανάληψη ενημερώνεται σύμφωνα με τη σχέση $\theta_{k+1} = \alpha \theta_k$, $0 < \alpha < 1$ τότε κάτω από ορισμένες προϋποθέσεις για τον πολλαπλασιαστικό συντελεστή α ο αλγόριθμος συγκλίνει σε χρόνο πολυωνυμικό. Οι **Monteiro** και **Adler** (1989), στους οποίους οφείλεται ο αλγόριθμος που περιγράψαμε στην ενότητα αυτή, απέδειξαν ότι θέτοντας

$$\alpha = 1 - \frac{3.5}{\sqrt{n}}$$

ο αλγόριθμος βρίσκει τη βέλτιστη λύση μετά από $O(\sqrt{n} L)$ το πολύ επαναλήψεις. Το εύρημα αυτό αποτελεί ένα πολύ καλό θεωρητικό φράγμα. Στην πράξη όμως η συγκεκριμένη τιμή του α δεν επιτρέπει τη γρήγορη μείωση της παραμέτρου θ και, έτσι, ο αλγόριθμος χρειάζεται πολλές επαναλήψεις μέχρις ότου φθάσει σε μια τόσο μικρή τιμή για το θ ώστε η διαφορά της τιμής της αντικειμενικής συνάρτησης του πρωτεύοντος από την τιμή της αντικειμενικής συνάρτησης του δυϊκού του να είναι πολύ κοντά στο μηδέν. Για να πάρουμε έναν πρακτικά πιο αποτελεσματικό αλγόριθμο, θα έπρεπε να μειώνουμε ταχύτερα την τιμή της παραμέτρου θ . Μια μεγάλη όμως μείωση της τιμής του θ θα είχε ως αποτέλεσμα οι νέες εκτιμήσεις των βέλτιστων λύσεων να μην είναι γνήσια θετικές. Γι' αυτό το λόγο θέτουμε

$$\mathbf{x}_{k+1}(\theta_{k+1}) = \mathbf{x}_k + \sigma_k \Delta \mathbf{x}_k, \quad \mathbf{w}_{k+1}(\theta_{k+1}) = \mathbf{w}_k + \sigma_k \Delta \mathbf{w}_k, \quad \mathbf{s}_{k+1}(\theta_{k+1}) = \mathbf{s}_k + \sigma_k \Delta \mathbf{s}_k$$

επιλέγοντας το μήκος του βήματος σ_k έτσι ώστε $\mathbf{x}_{k+1} > 0$ και $\mathbf{s}_{k+1} > 0$.

5. ΥΠΟΛΟΓΙΣΤΙΚΗ ΠΟΛΥΠΛΟΚΟΤΗΤΑ ΤΟΥ ΑΛΓΟΡΙΘΜΟΥ SIMPLEX ΚΑΙ ΤΟΥ ΑΛΓΟΡΙΘΜΟΥ ΤΟΥ ΚΑΡΜΑΚΑΡ.

Σε αυτό το κεφάλαιο θα αναφερθούμε σε κάποια θεμελιώδη θέματα της υπολογιστικής πολυπλοκότητας των αλγορίθμων που χρησιμοποιούμε για να λύσουμε προβλήματα γραμμικού προγραμματισμού. Συγκεκριμένα θα αναλύσουμε τη χειρότερη περίπτωση εκτέλεσης του κάθε αλγορίθμου, προκειμένου να τον «χαρακτηρίσουμε» ως αλγόριθμο είτε πολυωνυμικού είτε εκθετικού χρόνου.

Σύμφωνα με αυτή την ταξινόμηση, ο αλγόριθμος Simplex φαίνεται να είναι «κακός» επειδή, ο χρόνος επίλυσης συγκεκριμένων κατηγοριών προβλημάτων, αυξάνεται εκθετικά καθώς το μέγεθος του προβλήματος μεγαλώνει. Ωστόσο, υπάρχουν μερικά θεωρητικά αποτελέσματα, σε κάποιες «μέσου όρου» περιπτώσεις που εξηγούν τη διαφορά ανάμεσα στα θεωρητικά αποτελέσματα της χειρότερης περίπτωσης και την παρατηρηθείσα πρακτική επάρκεια του αλγορίθμου Simplex.

Στον αντίποδα υπάρχουν αλγόριθμοι που κατά την επίλυση του προβλήματος επιδεικνύουν χρόνους υπολογισμού, πολυωνυμικά εξαρτώμενους από τα εισαγόμενα μεγέθη. Οι δύο πιο διαδεδομένοι αλγόριθμοι που εμπίπτουν σε αυτή τη κατηγορία, είναι οι αλγόριθμοι του Khachian και Karmarkar, με τον πρώτο από αυτούς να μην έχει πρακτική υπολογιστική αξία και το δεύτερο να είναι σαφώς πιο «ελπιδοφόρος». Αξίζει να αναφέρουμε πάντως, ότι και οι δύο αλγόριθμοι είναι μη γραμμικές προσεγγίσεις σε προβλήματα γραμμικού προγραμματισμού.

■ Πολυωνυμική πολυπλοκότητα

Όπως έχουμε αναφέρει η γέννηση του γραμμικού προγραμματισμού είχε σαν αποτέλεσμα την ραγδαία ανάπτυξη της επιχειρησιακής έρευνας. Έτσι μαζί με την εμφάνιση μιας πλούσιας συλλογής ποικίλων τύπων προβλημάτων, εμφανίστηκε συνακολούθως και ένα μεγάλο πλήθος ανταγωνιστικών αλγόριθμων επίλυσής τους. Συνεπώς, εμφανίστηκε η ανάγκη μιας «επίσημης» ταξινόμησης και σύγκρισης τόσο των προβλημάτων όσο και των αλγόριθμων επίλυσης με βάση την υπολογιστική τους «υπακοή». Και ήταν αυτή ακριβώς η ανάγκη, που μέσα στη δεκαετία του 1970 οδήγησε τους επιστήμονες των υπολογιστών και της επιχειρησιακής έρευνας να ερευνήσουν το θέμα της υπολογιστικής πολυπλοκότητας των προβλημάτων και των αλγορίθμων. Η άποψη που υιοθετήθηκε ήταν να γίνει μια εκτίμηση του προβλήματος ή του αλγόριθμου, η οποία να βασίζεται στην απόδοση που θα επιδείξει στην κατάσταση της χειρότερης περίπτωσης. Η ιδέα είναι να εξακριβώσουμε πόσο δύσκολο είναι να λυθεί ένα πρόβλημα στη χειρότερη περίπτωση, ή ποια είναι η αύξηση της υπολογιστικής προσπάθειας ενός αλγόριθμου σαν συνάρτηση του μεγέθους ενός προβλήματος στη χειρότερη περίπτωση. Και είναι αυτή ακριβώς η ανάλυση που παρέχει μια εγγύηση απόδοσης ενός αλγόριθμου, ο οποίος λύνει μια δοθείσα κατηγορία προβλημάτων.

Ας θεωρήσουμε το παρακάτω πρόβλημα γραμμικού προγραμματισμού

$$\max \mathbf{c}'\mathbf{x} : \mathbf{Ax} = \mathbf{b}, \mathbf{x} \geq \mathbf{0}$$

όπου \mathbf{A} είναι $m \times n$ και $m, n \geq 2$. Υποθέτουμε εδώ (και σε όλο το υπόλοιπο κεφάλαιο) ότι τα δεδομένα είναι όλα ακέραια. Ως *περίπτωση* αυτού του προβλήματος θεωρούμε ένα συγκεκριμένο μέλος αυτής της κατηγορίας προβλημάτων με συγκεκριμένες τιμές παραμέτρων ή δεδομένα. Έτσι, ο προσδιορισμός των τιμών $m, n, \mathbf{c}, \mathbf{A}$ και \mathbf{b} καθορίζει μια περίπτωση του παραπάνω προβλήματος γραμμικού προγραμματισμού. Το μέγεθος μιας περίπτωσης από αυτό το πρόβλημα αντιπροσωπεύεται από τις οντότητες (m, n, L) , όπου L είναι το νούμερο των δυαδικών ψηφίων που απαιτούνται για

να για να καταγράψουμε όλα τα δεδομένα του προβλήματος και είναι γνωστό ως το *μήκος εισόδου* μιας περίπτωσης του προβλήματος. Σημειώνουμε ότι ο αριθμός των δυαδικών ψηφίων που απαιτούνται για να παραστήσουν έναν θετικό ακέραιο Δ είναι $\lceil \log(1+\Delta) \rceil$, όπου $\lceil \square \rceil$ δηλώνει την προς τα πάνω στρογγυλοποιημένη τιμή και $\log(\square)$ δηλώνει λογάριθμο με βάση το 2. Για παράδειγμα, προκειμένου να αναπαραστήσουμε οποιοδήποτε ακέραιο $\Delta \in [2^r, 2^{r+1}]$ για έναν ακέραιο $r \geq 1$, απαιτούμε $(r + 1)$ δυαδικά ψηφία. Αν Δ είναι οποιοδήποτε πρόσημο, μπορούμε να χρησιμοποιήσουμε ένα επιπλέον ψηφίο για να καταγράψουμε το πρόσημό του και επομένως μπορούμε να αναπαραστήσουμε τον Δ χρησιμοποιώντας $1 + \lceil \log(1+|\Delta|) \rceil$ ψηφία. Συνεπώς, τα δεδομένα από το προαναφερθέν γραμμικό πρόβλημα μπορούν να καταγραφούν από ένα δυαδικό σχήμα κωδικοποίησης χρησιμοποιώντας τον ακόλουθο αριθμό ψηφίων:

$$L = \lceil 1 + \log(1+m) \rceil + \lceil 1 + \log(1+n) \rceil + \sum_j \{1 + \lceil \log(1+|c_j|) \rceil\} + \\ + \sum_i \sum_j \{1 + \lceil \log(1+|a_{ij}|) \rceil\} + \sum_i \{1 + \lceil \log(1+|b_i|) \rceil\}$$

Ας θεωρήσουμε έναν αλγόριθμο που λύνει οριστικά ένα πρόβλημα γραμμικού προγραμματισμού. Προκειμένου να αναλύσουμε την υπολογιστική του πολυπλοκότητα, απαιτείται να ορίσουμε ένα άνω φράγμα στην προσπάθεια επίλυσης οποιοδήποτε περίπτωσης αυτού του προβλήματος. Αυτή η προσπάθεια μπορεί να μετρηθεί με βάση τον αριθμό των στοιχειωδών πράξεων (όπως προσθέσεων, πολλαπλασιασμών και συγκρίσεων που απαιτούνται για να λυθεί το πρόβλημα) ως συνάρτηση του μεγέθους του προβλήματος. Ευτυχώς πάντως, μια ακριβής μέτρηση του συνολικού αριθμού τέτοιων πράξεων δεν είναι απαραίτητη. Το μόνο που απαιτείται είναι να ορίσουμε μια συνάρτηση $g(m,n,L)$ βάσει του μεγέθους του προβλήματος, τέτοια ώστε για κάποια επαρκώς μεγάλη σταθερά $\tau > 0$ ο συνολικός αριθμός των στοιχειωδών πράξεων που απαιτούνται από τον αλγόριθμο για να λυθεί το πρόβλημα να μην είναι παραπάνω από $\tau g(m,n,L)$. Σε αυτή την περίπτωση, λέμε ότι ο αλγόριθμος έχει πολυπλοκότητα τάξης $O(g(m,n,L))$. Συγκεκριμένα

όταν η συνάρτηση $g(m,n,L)$ είναι ένα πολυώνυμο ως προς m , n και L , λέμε ότι ο αλγόριθμος είναι πολυωνυμικής πολυπλοκότητας ή ότι είναι πολυωνυμικά φραγμένος. Αν για παράδειγμα, ένας αλγόριθμος εκτελεί το μέγιστο $6m^2n + 16mn + 26m$ στοιχειώδεις πράξεις για την επίλυση ενός προβλήματος, του οποίου το μέγεθος καθορίζεται από τα m και n , μπορούμε απλά να πούμε ότι είναι πολυπλοκότητας $O(m^2n)$ (καθώς οι όροι $16mn$ και $26m$ είναι συγκριτικά αμελητέοι και η αύξηση στην υπολογιστική προσπάθεια υπαγορεύεται κατά κύριο λόγο από την συνάρτηση $g(m,n) = m^2n$).

Αφού όπως είπαμε αυτό που απαιτείται είναι απλά να ορίσουμε ένα άνω φράγμα στην προσπάθεια επίλυσης, είναι σαφές ότι δε χρειάζεται να υπολογίσουμε ακριβώς το L για να ορίσουμε το μέγεθος του προβλήματος. Με άλλα λόγια, αν πάρουμε το L ως ένα κάτω φράγμα του πραγματικού αριθμού των ψηφίων που απαιτούνται για να καταγραφούν τα δεδομένα και επιδειχθεί πολυωνυμική αύξηση στην προσπάθεια επίλυσης ως αύξουσα συνάρτηση αυτής της ποσότητας, τότε αυτό θα συνεπάγεται ένα πολυωνυμικό φράγμα στην προσπάθεια επίλυσης, με όρους του πραγματικού μεγέθους του προβλήματος. Επομένως θα μπορούσαμε (στην ανάλυση των προβλημάτων γραμμικού προγραμματισμού) να πάρουμε το L να είναι η επόμενη ποσότητα

$$L = \left[1 + \log m + \log n + \sum_j \{1 + \lceil \log(1 + |c_j|) \rceil\} \right] + \sum_i \sum_j \{1 + \lceil \log(1 + |a_{ij}|) \rceil\} + \sum_i \{1 + \lceil \log(1 + |b_i|) \rceil\} \quad (5.1)$$

Η «εμμονή» μας για έναν αλγόριθμο πολυωνυμικού χρόνου μπορεί να γίνει κατανοητή συγκρίνοντας την αύξηση του n^2 κατά 2^n . Αφήνουμε το n να αντιπροσωπεύει το μέγεθος κάποιου προβλήματος και ονομάζουμε A_1 και A_2 δύο διαφορετικούς αλγόριθμους επίλυσης του προβλήματος αυτού. Επιπλέον υποθέτουμε ότι ο αλγόριθμος A_1 είναι πολυπλοκότητας $O(n^2)$ και ο αλγόριθμος A_2 είναι πολυπλοκότητας $O(2^n)$. Αυτό σημαίνει ότι υπάρχουν σταθερές t_1 και t_2 τέτοιες που η προσπάθεια επίλυσης με τους αλγόριθμους A_1 και A_2 είναι αντιστοίχως φραγμένη από τις εκφράσεις $t_1 n^2$ και $t_2 2^n$. Συνεπώς για ένα πρόβλημα μεγέθους $n=50$, θεωρώντας $t_1 = t_2 = 1$, ενώ ο αλγόριθμος A_1 δε θα εκτελέσει περισσότερες από 2500 πράξεις, ο αριθμός των πράξεων που απαιτούνται από τον αλγόριθμο A_2 μπορεί να φθάσει τις

1.1259×10^{15} ! Υπάρχουν δύο προφανείς «ανωμαλίες» σε αυτό το σύστημα εκτίμησης, τις οποίες πρέπει να σημειώσουμε: Πρώτον, πάντα προτιμούμε πολυώνυμα μικρής τάξης. Ένας αλγόριθμος πολυωνυμικού χρόνου με πολυπλοκότητα $O(n^{50})$ δεν έχει πρακτικά μεγαλύτερη αξία από τον αλγόριθμο A_2 . Δεύτερον, θα θέλαμε οι συντελεστές μεγέθους να είναι σχετικά μικροί. Αν $\tau_1 = 2^{50}$ και $\tau_2 = 1$, τότε παρότι η προσπάθεια επίλυσης με τον αλγόριθμο A_2 καταστρέφεται σημαντικά πιο γρήγορα με μια αύξηση στο n , από ότι με τον αλγόριθμο A_1 , από πρακτικής πλευράς ο αλγόριθμος A_1 είναι ιδιαίτερα δύσχρηστος για κάθε τιμή από $n \geq 1$. Αντίθετα, για επαρκώς μικρές τιμές του n ο αλγόριθμος A_2 είναι τουλάχιστον εύχρηστος.

Αξίζει να τονιστεί ο ρόλος του δυαδικού σχήματος κωδικοποίησης στον προσδιορισμό ενός αλγόριθμου πολυωνυμικού χρόνου. Ας υποθέσουμε ότι για την επίλυση κάποιων προβλημάτων γραμμικού προγραμματισμού προτάθηκε ένας αλγόριθμος, ο οποίος έχει πολυπλοκότητα τάξης $O(mn\Delta)$, όπου Δ είναι η μεγαλύτερη απόλυτη τιμή από τα c , b και A . Τότε αυτός ο αλγόριθμος δεν είναι πολυωνυμικής πολυπλοκότητας. Στην πράξη, ορίζοντας $L_1 = \log(\Delta)$, και αντιπροσωπεύοντας έτσι το μέγεθος του προβλήματος (για παράδειγμα, για καθορισμένο m και n), βλέπουμε ότι ο αλγόριθμος είναι εκθετικής πολυπλοκότητας, τάξης $O(mn2^{L_1})$. Ας υποθέσουμε ότι επιλέξαμε να μετρήσουμε το μέγεθος ενός προβλήματος με ένα που χρησιμοποιεί ένα «χτύπημα» για κάθε μονάδα δεδομένων. Για να αποθηκεύσουμε τον ακέραιο $\Delta > 0$ θα απαιτούνταν Δ χτυπήματα, ώστε το μέγεθος του προβλήματος να καθορίζεται από την τριάδα (m, n, Δ) και όχι από την τριάδα $(m, n, \log \Delta)$. Σε αυτή την περίπτωση, η αλγοριθμική «προσπάθεια» είναι άνω φραγμένη από ένα πολυώνυμο στο (m, n, Δ) . Ένας τέτοιος αλγόριθμος αναφέρεται ως **ψευδοπολυωνυμικός**.

Από την άλλη μεριά, ένας αλγόριθμος του οποίου η πολυπλοκότητα εξαρτάται μόνο από τον αριθμό των παραμέτρων του προβλήματος (όπως ορίστηκαν για τα γραμμικά προγράμματα τα m και n) και είναι ανεξάρτητη από τα μεγέθη αυτών των παραμέτρων, λέγεται ότι είναι **γνήσια** ή **ισχυρά πολυωνυμικός**. Ενώ δεν υπάρχει γνωστός ισχυρός πολυωνυμικός αλγόριθμος για γενικά προβλήματα γραμμικού προγραμματισμού, τέτοιοι αλγόριθμοι υπάρχουν για την ειδική κατηγορία των γραμμικών

προγραμμάτων που είναι δομημένα με δίκτυο.

Ολοκληρώνοντας αυτή την παράγραφο, ας τοποθετήσουμε τα προαναφερθέντα στο γενικό πλαίσιο της προσέγγισης που γίνεται από τους προγραμματιστές υπολογιστών. Στην περιοχή της επιστήμης των υπολογιστών, τα προβλήματα συνήθως τίθενται περισσότερο ως προβλήματα αποφάσεων παρά ως προβλήματα βελτιστοποίησης. Έτσι το πρόβλημα βελτιστοποίησης γραμμικού προγραμματισμού έχει ως «αντίγραφο» του το ακόλουθο πρόβλημα απόφασης γραμμικού προγραμματισμού.

Με δεδομένα τα \mathbf{c} , \mathbf{b} και \mathbf{A} (των κατάλληλων διαστάσεων)
και δοθέντος ενός ρητού αριθμού K ,
υπάρχει ένα ρητό διάνυσμα \mathbf{x} τέτοιο ώστε $\mathbf{Ax}=\mathbf{b}$, $\mathbf{x} \geq \mathbf{0}$ και $\mathbf{cx} \leq K$;

Ένας πολυωνυμικού χρόνου αλγόριθμος για ένα πρόβλημα αποφάσεων μπορεί (ανεπίσημα) να οριστεί ως αυτός που μπορεί να «λύσει» κάθε περίπτωση του προβλήματος χρησιμοποιώντας έναν αριθμό στοιχειωδών πράξεων, άνω φραγμένο από κάποιο πολυώνυμο του μεγέθους του προβλήματος. Με το «λύνουμε», εννοούμε τελείωμα με μια απάντηση ναι, δεδομένης μιας ναι-περίπτωσης του προβλήματος αποφάσεων. Έτσι αν κάποιος έχει έναν αλγόριθμο πολυωνυμικού χρόνου και δίνεται μια όχι-περίπτωση ενός προβλήματος, είναι επαρκές να υπερβούμε το δεδομένο πολυωνυμικό φράγμα χωρίς να αμφιταλαντευόμαστε να συμπεράνουμε ότι η απάντηση είναι όχι. Επομένως το πολυωνυμικό φράγμα ισχύει για όλα τα παραδείγματα. Προβλήματα για τα οποία υπάρχει κάποιος γνωστός, πολυωνυμικού χρόνου αλγόριθμος λέμε πως ανήκουν στη διακεκριμένη P τάξη προβλημάτων.

Είναι χρήσιμο να δούμε τη σχέση ανάμεσα στους πολυωνυμικού χρόνου αλγόριθμους για επίλυση προβλημάτων βελτιστοποίησης και για επίλυση προβλημάτων λήψης αποφάσεων. Ένας αλγόριθμος πολυωνυμικού χρόνου για ένα πρόβλημα βελτιστοποίησης μπορεί ξεκάθαρα να λύσει ένα πρόβλημα αποφάσεων σε πολυωνυμικό χρόνο. Αντιστρόφως, ας θεωρήσουμε ότι έχουμε έναν αλγόριθμο πολυωνυμικού χρόνου για ένα πρόβλημα αποφάσεων γραμμικού προγραμματισμού. Αν επιπλέον θεωρήσουμε ένα συγκεκριμένο

πρόβλημα βελτιστοποίησης που είναι εφικτό και φραγμένο, τότε υπάρχει μια βασική δυνατή λύση με αντικειμενική τιμή $\leq 2^L$ σε μέγεθος. Επομένως η βέλτιστη τιμή βρίσκεται στο διάστημα $[-2^L, 2^L]$. Αν λοιπόν θεωρήσουμε ότι ο αλγόριθμος που διαθέτουμε για να λύσουμε το πρόβλημα αποφάσεων παράγει λύση σε μία ναι-περίπτωση, κάποιος μπορεί να χρησιμοποιήσει αυτόν τον αλγόριθμο για να λύσει το πρόβλημα βελτιστοποίησης με το να εκτελεί επανειλημμένα μια έρευνα διχοτόμησης στο διάστημα αυτό. Αυτό μπορεί να γίνει ως εξής.

Πρώτα, χρησιμοποιώντας $K=0$, που είναι το μέσο του τρέχοντος διαστήματος αβεβαιότητας $[-2^L, 2^L]$ μπορεί να βρεθεί σε πολυωνυμικό χρόνο αν η βέλτιστη τιμή βρίσκεται στο διάστημα $[-2^L, 0]$ ή στο διάστημα $[0, 2^L]$. Στη συνέχεια, χρησιμοποιώντας το διάστημα που περιέχει τη βέλτιστη τιμή, ως το νέο διάστημα αβεβαιότητας μπορούμε προχωρήσουμε ξανά σε διχοτόμηση. Αυτή η διαδικασία μπορεί να συνεχιστεί μέχρι να γνωρίζουμε ότι η βέλτιστη τιμή βρίσκεται στο διάστημα $[u_1, u_2]$, όπου $u_2 - u_1 < 2^{-2L}$. Καθώς το αρχικό διάστημα αβεβαιότητας είναι μήκους 2^{L+1} και υποδιπλασιάζεται σε κάθε διχοτόμηση, μπορούμε να αποκτήσουμε αυτό το ελαττωμένο διάστημα αβεβαιότητας αφού χρησιμοποιήσουμε $3L+2$ εφαρμογές του αλγόριθμου του προβλήματος αποφάσεων. Αυτό σημαίνει ότι θα έχουμε χρησιμοποιήσει έναν αλγόριθμο πολυωνυμικού χρόνου, για ένα πολυωνυμικό αριθμό φορών και μάλιστα αυτό το έχουμε επιτύχει σε πολυωνυμικό χρόνο. Αν δεν υπάρχει διαθέσιμη λύση σε αυτό το διάστημα, μπορούμε να αποκτήσουμε κάποια εφαρμόζοντας το πρόβλημα αποφάσεων με το σκοπό να βρεθεί μια λύση x με $cx \leq u_2$. Όπως θα δείξουμε αργότερα δεδομένης μιας λύσης με αυτή την ακρίβεια στην αντικειμενική τιμή, κάποιος μπορεί να χρησιμοποιήσει ένα πολυωνυμικού χρόνου « στρογγυλοποιημένο σχήμα » με σκοπό να προσδιορίσει μια βέλτιστη λύση στο πρόβλημα βελτιστοποίησης. Συμπερασματικά, κάθε πρόβλημα βελτιστοποίησης μπορεί να λυθεί σε πολυωνυμικό χρόνο, χρησιμοποιώντας ένα αλγόριθμο πολυωνυμικού χρόνου για το πρόβλημα λήψης αποφάσεων.

Προχωράμε τώρα, στο να επιδείξουμε ότι ο αλγόριθμος Simplex είναι εν γένει εκθετικής πολυπλοκότητας, που υπό συνθήκες παρουσιάζεται ως ένας συγκεκριμένος, πολυωνυμικού χρόνου, αλγόριθμος για να λύνουμε προβλήματα γραμμικού προγραμματισμού.

■ Υπολογιστική πολυπλοκότητα του αλγόριθμου Simplex

Όταν ο Dantzig πρωτοπαρουσίασε τον αλγόριθμο Simplex, η «διαισθητική» αντίδραση της ερευνητικής κοινότητας, ήταν ότι αυτός ο αλγόριθμος δε θα ήταν επαρκής. Από τη φύση του, ο αλγόριθμος Simplex «έρπεται» στις ακμές του πολυέδρου και δε κάνει καμία προσπάθεια να πηδήσει στο εσωτερικό του ή πάνω από μεγαλύτερης διάστασης επιφάνειες. Στην πραγματικότητα σκοντάφτει σποραδικά, μέσω μιας αλληλουχίας από εκφυλισμένους μετασχηματισμούς, σε κάποια ακραία σημεία. Παρόλα αυτά οι ερευνητές ξαφνιάστηκαν ευχάριστα όταν διαπίστωσαν ότι στην πράξη η μέθοδος αυτή αποδίδει υπερβολικά καλά. Για τα περισσότερα πρακτικά προβλήματα, η μέθοδος αυτή παρατηρήθηκε εμπειρικά ότι χρειάζεται περίπου $3m/2$ επαναλήψεις, και σπάνια περισσότερες από $3m$ επαναλήψεις, όπου ο πίνακας συντελεστών \mathbf{A} είναι $m \times n$. Μάλιστα, υπάρχει μια εξίσωση παλινδρόμησης της μορφής $Km^{2.5}nd^{0.33}$, για παράδειγμα, όπου d είναι η πυκνότητα του πίνακα \mathbf{A} , που μπορεί να προβλέψει την επίδοση του αλγόριθμου Simplex σε ένα δεδομένο πρόβλημα αρκετά καλά. Παρόλα αυτά, παραμένει το γεγονός ότι ο αλγόριθμος είναι επί της ουσίας παγιδευμένος στην εν δυνάμει συνδυαστική γωνιακή απόσταση των πάνω από $\binom{n}{m}$

κορυφών που ο αλγόριθμος είναι ικανός να επισκεφτεί. Αφού

$$\binom{n}{m} = \frac{n!}{m!(n-m)!} = \binom{n}{m} \binom{n-1}{m-1} \binom{n-2}{m-2} \cdots \left[\frac{n-(m-1)}{m-(m-1)} \right] > \left(\frac{n}{m} \right)^m$$

το οποίο είναι τουλάχιστον 2^m όταν $n \geq 2m$, η πιθανότητα για πολυπλοκότητα εκθετικής τάξης για κάποια προβλήματα είναι αρκετά μεγάλη.

Ο φόβος της εκθετικής πολυπλοκότητας της μεθόδου Simplex επιβεβαιώθηκε το 1971 όταν οι Victor Klee και George Minty παρουσίασαν μια κατηγορία προβλημάτων που καθορίζονταν από $m = n$ περιορισμούς με $2n$ μη αρνητικές μεταβλητές, για τις οποίες ο αλγόριθμος Simplex απαιτεί $2^n - 1$ επαναλήψεις, διασχίζοντας όλες τις κορυφές του εφικτού συνόλου του προβλήματος. Αυτή η κατηγορία προβλημάτων προκύπτει ορίζοντας ως δυνατή περιοχή μια κατάλληλη παραμόρφωση του n διαστάσεων υπερκύβου στην E^n , που έχει 2^n κορυφές. Μαθηματικά, το πρόβλημα στην E^n μπορεί να

διατυπωθεί ως ακολούθως, όπου ϵ είναι κάποιος τυχαίος αριθμός στο διάστημα $(0, \frac{1}{2})$.

$$\begin{aligned} \text{Maximize} \quad & x_n \\ \text{Με τους περιορισμούς} \quad & 0 \leq x_1 \leq 1 \\ & \epsilon x_{j-1} \leq x_j \leq 1 - \epsilon x_{j-1} \quad \text{για } j = 2, \dots, n \\ & x_j \geq 0, \quad j = 1, \dots, n \end{aligned}$$

Αφότου μετατρέψουμε το πρόβλημα καταλλήλως, θα δείξουμε ότι χρησιμοποιώντας τον κανόνα του Dantzig βάσει του οποίου εισάγουμε τη μη βασική μεταβλητή με το μεγαλύτερο ελαττωμένο κόστος και ξεκινάμε από την αρχή, ο αλγόριθμος Simplex μετατρέπει $2^n - 1$ ακμές από αυτό το πολύτοπο και διατρέχει όλες τις 2^n κορυφές του εφικτού συνόλου του προβλήματος.

Προκειμένου να φέρουμε το πρόβλημα στην επιθυμητή μορφή, θεωρούμε την γραμμική μετατροπή:

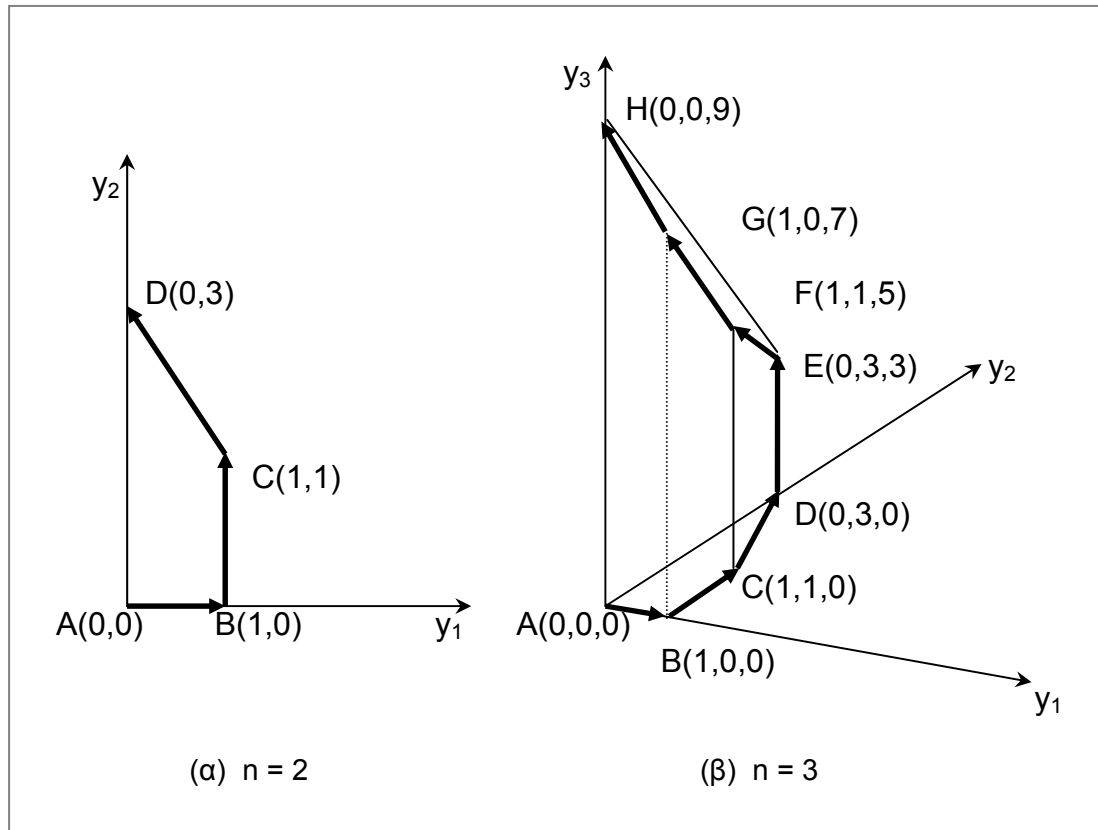
$$y_1 = x_1, \quad y_j = (x_j - \epsilon x_{j-1}) / \epsilon^{j-1} \quad \text{για } j = 2, \dots, n \quad (5.2)$$

Χρησιμοποιώντας αυτή τη μετατροπή και θέτοντας $\theta = 1/\epsilon$, αποκτούμε το ακόλουθο ισοδύναμο πρόβλημα.

$$\begin{aligned} \text{Maximize} \quad & \sum_{j=1}^n y_j \\ \text{Με τους περιορισμούς} \quad & y_1 \leq 1 \\ & y_j + 2 \leq \sum_{k=1}^{j-1} y_k \theta^{j-k} \quad \text{για } j = 2, \dots, n \\ & y_1, \dots, y_n \geq 0 \end{aligned} \quad (5.3)$$

Για ευκολία, ορίζουμε s_1, s_2, \dots, s_n ως τις slack μεταβλητές που αντιστοιχούν στους παραπάνω n περιορισμούς (βλ. σελ.16). Παρατήρησε ότι η βασική λύση με βασικές μεταβλητές (s_1, \dots, s_n, y_n) είναι εφικτή, αποδίδοντας $y = (0, \dots, 0, \theta^{n-1})$ με τιμή αντικειμενικής συνάρτησης θ^{n-1} . Για το πρωτότυπο πρόβλημα αυτό αντιστοιχεί στη λύση $x = (0, \dots, 0, 1)$. Παρατηρούμε

ότι ο τελευταίος περιορισμός και η αντικειμενική συνάρτηση στο αρχικό πρόβλημα, αντιστοιχεί ξεκάθαρα στη βέλτιστη λύση σε αυτό το πρόβλημα, γεγονός που επιβεβαιώνεται υπολογίζοντας τις τιμές $z_j - c_j$ για τις μη βασικές μεταβλητές.



Σχήμα 5.1 Οπτικοποίηση του πολυτόπου του προβλήματος των Klee και Minty

Ας θεωρήσουμε τώρα την περίπτωση $n = 2$. Το πρόβλημα σε αυτή την περίπτωση είναι της μορφής: να μεγιστοποιηθεί το $y_1 + y_2$, υπό τους περιορισμούς $y_1 \leq 1$, $y_2 + 2y_1 \leq \theta$ και $y_1, y_2 \geq 0$. Στο σχήμα 5.1.α εικονίζεται η εφικτή περιοχή χρησιμοποιώντας $\theta = 3$. Σημειώνουμε ότι οι βασικές μεταβλητές είναι οι (s_1, s_2) . Από εδώ και στο εξής, οι $z_j - c_j$ τιμές για y_1 και y_2 είναι και οι δύο -1 . Επιλέγοντας την y_1 ως τη μεταβλητή εισόδου ερχόμαστε στη λύση του γειτονικού ακραίου σημείου (y_1, s_2) . Συνεχίζοντας, ακολουθούμε το Simplex-μονοπάτι A, B, C, D όπως φαίνεται στο σχήμα 8.1α. Με τον τρόπο αυτό, διερχόμαστε από τις 2^2 βασικές εφικτές λύσεις με αντίστοιχες βασικές μεταβλητές (s_1, s_2) , (y_1, s_2) , (y_1, y_2) και τελικά (s_1, y_2) . Εύκολα μπορούμε να επιβεβαιώσουμε ότι $z_j - c_j = -1$ για την μεταβλητή που εισάγεται σε κάθε επανάληψη. Αξίζει να σημειώσουμε ότι αν στην πρώτη επανάληψη είχαμε

επιλέξει την y_2 ως την μεταβλητή που εισάγουμε, θα είχαμε λύσει αυτό το πρόβλημα σε μία επανάληψη.

Στη συνέχεια θεωρούμε την περίπτωση $n = 3$. Το πρόβλημα σε αυτή την περίπτωση είναι της μορφής: να μεγιστοποιηθεί το $y_1 + y_2 + y_3$, υπό τους περιορισμούς $y_1 \leq 1$, $y_2 + 2y_1 \leq \theta$, $y_3 + 2y_1 + 2y_2 \leq \theta^2$ και $y_1, y_2, y_3 \geq 0$. Το σχήμα 5.1β δείχνει την εφικτή περιοχή χρησιμοποιώντας όπως και προηγουμένως $\theta = 3$. Ας κάνουμε τώρα τις ακόλουθες παρατηρήσεις. Πρώτα, σημειώνουμε ότι για $y_3 = 0$, τα ακραία σημεία στο διάστημα (y_1, y_2) που αποκτήθηκε για την περίπτωση $n = 2$ παραμένουν ακραία σημεία και στην παρούσα περίπτωση, στην οποία η s_3 είναι η τρίτη βασική μεταβλητή. Τα αντίστοιχα σύνολα βασικών μεταβλητών είναι (s_1, s_2, s_3) , (y_1, s_2, s_3) , (y_1, y_2, s_3) και τελικά (s_1, y_2, s_3) . Δεύτερον, παρατηρούμε ότι κατά τη μετακίνηση από κάθε μία από αυτές τις βάσεις στην επόμενη, η τιμή $z_j - c_j$ για την εισερχόμενη μεταβλητή είναι -1 , όπως στην περίπτωση $n = 2$. Η τιμή $z_j - c_j$ για την καινούρια μη βασική μεταβλητή y_3 επίσης παραμένει στο -1 παντού, αφού η τελευταία συνιστώσα του Simplex πολλαπλασιαστικού διανύσματος $\mathbf{c}_B \mathbf{B}^{-1}$ είναι μηδέν για όλες αυτές τις βάσεις \mathbf{B} . Γι' αυτό, το Simplex-μονοπάτι A, B, C, D, για την περίπτωση $n = 2$ παραμένει ένα μονοπάτι Simplex για την περίπτωση $n = 3$ υπό τον κανόνα εισαγωγής του Dantzig. Επιπλέον, η y_3 παραμένει συνεχώς ικανή για εισαγωγή. Τρίτον, στην κορυφή D, πρέπει να εισάγουμε y_3 και αφού s_3 και y_3 δε μπορούν να γίνουν και τα δύο βασικά, η s_3 πρέπει να αφήσει τη βάση και για τις εναπομένουσες επαναλήψεις η y_3 παραμένει βασική. Χρησιμοποιώντας την τελευταία εξίσωση για να λύσουμε ως προς y_3 συναρτήσει των y_1, y_2 και s_3 και απαλείφοντας το y_3 από το πρόβλημα παίρνουμε το ίδιο πρόβλημα με αυτό που είχαμε και στην περίπτωση για $n = 2$, εκτός από το ότι οι αντικειμενικοί συντελεστές του y_1 και y_2 είναι τώρα -1 . Με άλλα λόγια, είναι σαν το $n = 2$ πρόβλημα να έχει αλλάξει από πρόβλημα μεγιστοποίησης σε πρόβλημα ελαχιστοποίησης στο τέλος της λύσης του! Συνεπώς, διατηρώντας το κριτήριο εισαγωγής του Dantzig, το μονοπάτι Simplex απλά αντιστρέφεται για τις βάσεις (για την περίπτωση $n=2$), με το y_3 να είναι συνεχώς η επιπρόσθετη βασική μεταβλητή. Κατόπιν επισκεπτόμαστε τις κορυφές E, F, G και H (όπως φαίνεται στο σχήμα 5.1β), με αντίστοιχα σύνολα βασικών μεταβλητών (s_1, y_2, y_3) , (y_1, y_2, y_3) , (y_1, s_2, y_3) και (s_1, s_2, y_3) . Επομένως, στην επίλυση του προβλήματος, έχουμε μετατρέψει $2^3 - 1$

ακμές και έχουμε επισκεφτεί (όλα τα) 2^3 ακραία σημεία σε αυτή τη διαδικασία.

Τα παραπάνω εύκολα γενικεύονται χρησιμοποιώντας ένα επαγωγικό επιχείρημα: Για παράδειγμα για την περίπτωση $n=4$, ξεκινάμε πρώτα πηγαίνουμε με ένα σύνολο από 2^3 βάσεις με βασικές μεταβλητές (s_1, s_2, s_3, s_4) , (y_1, s_2, s_3, s_4) , (y_1, y_2, s_3, s_4) , (s_1, y_2, s_3, s_4) , (s_1, y_2, y_3, s_4) , (y_1, y_2, y_3, s_4) , (y_1, s_2, y_3, s_4) και (s_1, s_2, y_3, s_4) . Αυτή η ακολουθία αποκτιέται με το να χρησιμοποιήσουμε την $n = 3$ περίπτωση με την s_4 σαν επιπρόσθετη βασική μεταβλητή. Μετά εισάγεται η y_4 και η s_4 αφήνει τη βάση, δίνοντας ένα σύνολο από βασικές μεταβλητές (s_1, s_2, y_3, y_4) . Η αλληλουχία των αξόνων αντιστρέφεται με το πρόβλημα μεγιστοποίησης για την $n=3$ περίπτωση να αλλάζει σε πρόβλημα ελαχιστοποίησης όταν η y_4 παραμένει βασική. Ακολουθως παράγουμε τις 2^3 βάσεις με βασικές μεταβλητές (s_1, s_2, y_3, y_4) , (y_1, s_2, y_3, y_4) , (y_1, y_2, y_3, y_4) , (s_1, y_2, y_3, y_4) , (s_1, y_2, s_3, y_4) , (y_1, y_2, s_3, y_4) , (y_1, s_2, s_3, y_4) και (s_1, s_2, s_3, y_4) . Αυτό δίνει ως αποτέλεσμα $2 \times 2^3 = 2^4$ βάσεις πάνω σε $2^4 - 1$ άξονες. Η κατασκευή της επίσημης επαγωγικής απόδειξης είναι τώρα απλή. Είναι ενδιαφέρον να παρατηρήσουμε ότι για την γενική περίπτωση n , για όλες τις πρώτες 2^{n-1} βασικές εφικτές λύσεις η μεταβλητή y_n παραμένει ικανή για εισαγωγή. Επίσης αξίζει να προσέξουμε ότι στο τέλος των 2^{n-1} μετασχηματισμών, οι βασικές μεταβλητές είναι $(s_1, s_2, \dots, s_{n-2}, y_{n-1}, y_n)$. Αν εισαχθεί η s_{n-1} σε αυτό το μετασχηματισμό, τότε προκύπτει το βέλτιστο αποτέλεσμα. (Βλέποντας το σχήμα 5.1β αυτό θα ήταν ανάλογο με το να πηγαίναμε από το E στο H). Παρόλα αυτά, και η y_1 και η s_{n-1} έχουν εδώ $z_j - c_j = -1$, και η επιλογή του να εισάγουμε το y_1 μας φέρνει ως το τέλος της αντίστροφης ακολουθίας μετασχηματισμών.

Ολοκληρώνοντας πρέπει να σημειώσουμε την υπερβολική προσπάθεια που πρέπει να καταβάλλουμε για την επίλυση του προβλήματος, καθώς το μέγεθός του αυξάνεται. Συγκεκριμένα, για ένα πρόβλημα με n περιορισμούς και $2n$ μη αρνητικές μεταβλητές απαιτούνται $0.25 n^2$ στοιχειώδεις πράξεις ανά μετασχηματισμό και κάθε πράξη χρειάζεται ένα nanosecond για να «εκτελεσθεί». Με βάση αυτό, η υπολογιστική προσπάθεια για $n = 50$ είναι 22.3 χρόνια και για $n = 60$, είναι πάνω από 328 αιώνες! Κάποιος μπορεί να επιχειρηματολογήσει υπέρ του αλγόριθμου Simplex, λέγοντας ότι ο αλγόριθμος είχε την ευκαιρία να λύσει το προαναφερθέν πρόβλημα σε μια

επανάληψη. Πράγματι, ταξινομώντας τις στήλες, για παράδειγμα, χρησιμοποιώντας το μετασχηματισμό $y_i = 2^{-i} y_i$, μπορούμε να απαλείψουμε τα εμπόδια στο κριτήριο εισαγωγής κατά την πρώτη επανάληψη και επομένως να επιλύσουμε το πρόβλημα (αναμφισβήτητα) σε μια επανάληψη. Παρατηρούμε επίσης ότι ο κανόνας του να εισάγεις αυτή τη μεταβλητή που δίνει τη μέγιστη βελτίωση στην τιμή της αντικειμενικής συνάρτησης, επίσης λύνει το πρόβλημα, σαφώς σε μια επανάληψη. Επιπλέον αυτό το κριτήριο είναι αμετάβλητο σε σχέση με την κλιμάκωση και την παρουσίαση του προβλήματος, αφού είναι άρρηκτα συνδεδεμένο με τη δομή του ίδιου του πολύτοπου. Το 1973, παρόλα αυτά, ο Jeroslow έδειξε την ύπαρξη κατηγοριών προβλημάτων τα οποία αποφέρουν μια εκθετική προσπάθεια στην επίλυση με το μέγιστης-βελτίωσης κριτήριο εισαγωγής. Συγκεκριμένα, για κάποιο δεδομένο ακέραιο $r \geq 1$, αν κάποιος ήταν να εκτελέσει όλους r επόμενους μετασχηματισμούς και να διαλέξει εκείνο το Simplex μονοπάτι του r -μετασχηματισμού, που δίνει τη μέγιστη βελτίωση στην αντικειμενική τιμή, τότε ο αλγόριθμος θα εξακολουθούσε να επιδεικνύει εκθετική τάση σε κάποιες κατηγορίες προβλημάτων. Η ρίζα του προβλήματος είναι επομένως στην «μυωπική» οπτική του αλγόριθμου Simplex, στον οποίο οι αποφάσεις βασίζονται στην τοπική κατασκευή-δομή του πολύτόπου, περιορίζοντας την κίνηση σε ένα μονοπάτι ακμής.

■ Ανάλυση της σύγκλισης και της πολυπλοκότητας του αλγορίθμου Karmarkar

Προχωράμε τώρα στην ανάλυση της σύγκλισης και της πολυπλοκότητας του αλγορίθμου του Karmarkar, ο οποίος εφαρμόζεται στο πρόβλημα (4.1) υπό την ισχύ των υποθέσεων (A1) και (A2).

Προκειμένου να αποδείξουμε τη σύγκλιση του αλγόριθμου και να «εδραιώσουμε» την πολυωνυμική πολυπλοκότητά του, υπάρχουν δύο δομές τις οποίες υιοθετούμε. Η πρώτη είναι η μελέτη-εξέταση χαλάρωσης του προβλήματος (4.4), το οποίο συμπληρώνει ο περιορισμός του (4.5). Η δεύτερη είναι η χρήση μιας νέας συνάρτησης γνωστής σαν την δυνητική συνάρτηση, που μετράει την πρόοδο του αλγόριθμου με έναν πολύ έξυπνο τρόπο και βοηθά να επικυρωθεί η πολυωνυμική του πολυπλοκότητα.

Για να κατασκευασθεί η απαραίτητη χαλάρωση του προβλήματος (4.4) θεωρούμε την n -διαστάσεων σφαίρα $B(y_0, R)$ στο y -διάστημα που έχει το κέντρο της στο y_0 και που συμπίπτει με αυτό του Simplex S_y , και που έχει τέτοια ακτίνα R ώστε η $B(y_0, R)$ τεμνόμενη με το $\{y : \mathbf{1}y = 1\}$ είναι μια σφαίρα διάστασης $(n - 1)$ που περιγράφει - καθορίζει τα όρια του S_y (σχήμα 5.2). Ως εκ τούτου το R είναι η απόσταση από το σημείο $(1/n, \dots, 1/n)$ σε οποιοδήποτε κορυφή του S_y , για παράδειγμα την $(1, 0, \dots, 0)$. Από αυτό έχουμε ότι $R = \sqrt{(n-1)/n}$. Τώρα θεωρούμε το πρόβλημα που προκύπτει με την προσθήκη του πλεονάζοντος περιορισμού $y \in B(y_0, R)$ στο πρόβλημα (4.4), αλλά διαγράφοντας τους περιορισμούς μη αρνητικότητας:

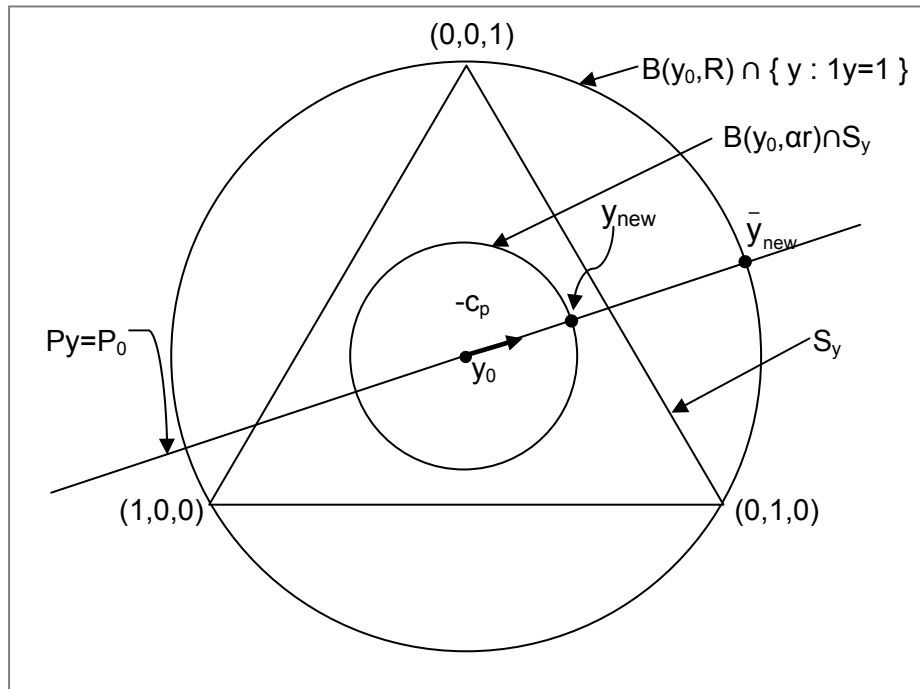
$$\text{Minimize } \{ \bar{\mathbf{c}}y : \mathbf{P}y = \mathbf{P}_0, (\mathbf{y}-\mathbf{y}_0)^t(\mathbf{y}-\mathbf{y}_0) \leq R^2 \} \quad (5.4)$$

Πάλι, η εφικτή περιοχή στην εξίσωση (5.4) είναι μια $(n-m-1)$ διαστάσεων σφαίρα με κέντρο το y_0 και που καθορίζεται από την τομή του $(n-m-1)$ διαστάσεων ομοπαράλληλικού υπόχωρου $\{y : \mathbf{P}y = \mathbf{P}_0\}$ με την $B(y_0, R)$. Το σχήμα 5.2 παρουσιάζει αυτή την τομή για μια περίπτωση με $n = 3$ και $m = 1$. (Για μια περίπτωση με $n = 4$ και $m = 1$, μια παρόμοια τομή σε σχέση με την εγγραμμένη (στενεμένη) σφαίρα παρουσιάζεται στο σχήμα 4.4). Ως εκ τούτου η βέλτιστη λύση \bar{y}_{new} στο πρόβλημα (5.4) λαμβάνεται με παρόμοιο τρόπο με

αυτό που οδηγηθήκαμε στην εξίσωση (4.6), και δίνεται από τη σχέση

$$\bar{\mathbf{y}}_{\text{new}} = \mathbf{y}_0 - \frac{R \mathbf{c}_p}{\|\mathbf{c}_p\|} \quad (5.5)$$

όπου \mathbf{c}_p ορίζεται στην εξίσωση (4.6). (Βλ. σχήμα 5.2)



Σχήμα 5.2 Relaxation για την περιγεγραμμένη σφαίρα

Κάποιος θα μπορούσε τώρα, να κάνει μια εκτίμηση για την πρόοδο που σημειώνεται όσον αφορά την αντικειμενική συνάρτηση $\bar{\mathbf{c}} \mathbf{y}$ στο y -διάστημα ως εξής: Σημειώνουμε καταρχάς ότι από την εξίσωση (4.6) πράγματι $\bar{\mathbf{c}} \mathbf{y}_{\text{new}} < \bar{\mathbf{c}} \mathbf{y}_0$ αν $\mathbf{c}_p \neq \mathbf{0}$, αφού $\bar{\mathbf{c}} (\mathbf{y}_0 - \mathbf{y}_{\text{new}}) = \alpha r \bar{\mathbf{c}} \mathbf{c}_p / \|\mathbf{c}_p\| = \alpha r \|\mathbf{c}_p\| > 0$. Η τελευταία εδώ ισότητα προκύπτει από το γεγονός ότι $\bar{\mathbf{c}}^t = \mathbf{c}_p + \mathbf{P}^t(\mathbf{P}\mathbf{P}^t)^{-1}\mathbf{P} \bar{\mathbf{c}}^t$ από την εξίσωση (4.6β), και έτσι $\bar{\mathbf{c}} \mathbf{c}_p = \|\mathbf{c}_p\|^2 + \bar{\mathbf{c}} \mathbf{P}^t(\mathbf{P}\mathbf{P}^t)^{-1}\mathbf{P} \mathbf{c}_p = \|\mathbf{c}_p\|^2$ αφού εξ ορισμού $\mathbf{P}\mathbf{c}_p = \mathbf{0}$. Στην πραγματικότητα, δηλώνοντας \mathbf{y}^* ως βέλτιστη λύση στο πρόβλημα (4.4), δεδομένου ότι το πρόβλημα (4.5) (που λύνεται από το \mathbf{y}_{new}) είναι ένας περιορισμός αυτού του προβλήματος και το πρόβλημα (5.4) (που λύνεται από

το $\bar{\mathbf{y}}_{\text{new}}$) είναι ένα relaxation αυτού του προβλήματος, έχουμε

$$\bar{\mathbf{c}}\mathbf{y}_{\text{new}} \leq \bar{\mathbf{c}}\mathbf{y}^* \leq \bar{\mathbf{c}}\mathbf{y}_{\text{new}} \leq \bar{\mathbf{c}}\mathbf{y}_0$$

Επιπλέον, χρησιμοποιώντας αυτό μαζί με τις εξισώσεις (5.5) και (4.6β), παίρνουμε

$$0 < \bar{\mathbf{c}}(\mathbf{y}_0 - \mathbf{y}_{\text{new}}) \leq \bar{\mathbf{c}}(\mathbf{y}_0 - \mathbf{y}^*) \leq \bar{\mathbf{c}}(\mathbf{y}_0 - \bar{\mathbf{y}}_{\text{new}}) = \frac{R}{\|\mathbf{c}_p\|} \bar{\mathbf{c}}\mathbf{c}_p = \frac{R}{\alpha r} \bar{\mathbf{c}}(\mathbf{y}_0 - \mathbf{y}_{\text{new}})$$

Έτσι, αυτό βεβαιώνει το

$$\bar{\mathbf{c}}(\mathbf{y}_0 - \mathbf{y}^*) \leq \frac{R}{\alpha r} \bar{\mathbf{c}}(\mathbf{y}_0 - \mathbf{y}_{\text{new}}) = \frac{R}{\alpha r} [\bar{\mathbf{c}}(\mathbf{y}_0 - \mathbf{y}^*) - \bar{\mathbf{c}}(\mathbf{y}_{\text{new}} - \mathbf{y}^*)].$$

Λύνοντας ως προς $\bar{\mathbf{c}}(\mathbf{y}_{\text{new}} - \mathbf{y}^*)$ και λαμβάνοντας υπόψη ότι $\bar{\mathbf{c}}(\mathbf{y}_0 - \mathbf{y}^*) > 0$ παίρνουμε

$$\frac{\bar{\mathbf{c}}(\mathbf{y}_{\text{new}} - \mathbf{y}^*)}{\bar{\mathbf{c}}(\mathbf{y}_0 - \mathbf{y}^*)} \leq 1 - \frac{\alpha r}{R} \equiv 1 - \frac{\alpha}{(n-1)} \quad (5.6)$$

Σημειώνουμε ότι κάτω από την υπόθεση (A2), έχουμε $\bar{\mathbf{c}}\mathbf{y}^* \equiv 0$, έτσι

$$\frac{\bar{\mathbf{c}}\mathbf{y}_{\text{new}}}{\bar{\mathbf{c}}\mathbf{y}_0} \leq 1 - \frac{\alpha}{(n-1)} \quad (5.7)$$

Η εξίσωση (5.6) μας λέει ότι σε κάθε επανάληψη, σε σχέση με την αντικειμενική συνάρτηση $\bar{\mathbf{c}}\mathbf{y}$, το κενό στη βελτιστοποίηση μειώνεται κατά $100\alpha / (n-1)$ τοις εκατό. Εντούτοις το $\bar{\mathbf{c}} = \mathbf{c}\mathbf{D}_k$ εξαρτάται από τη συγκεκριμένη επανάληψη k , και επιπλέον, η εξίσωση (5.6) απλά εγγυάται μια μείωση στον αριθμητή του προβλήματος (4.3). Στην πραγματικότητα, η αντικειμενική τιμή στο πρόβλημα (4.3), και ως εκ τούτου η αρχική αντικειμενική συνάρτηση στο πρόβλημα (4.1), μπορεί να μην πέσει, αλλά στην πράξη να αυξηθεί. Ευτυχώς, υπάρχει μια άλλη συνάρτηση που συντηρεί την γνησία μονοτονία και ως εκ τούτου βεβαιώνει τη σύγκλιση στη βελτιστοποίηση. Στην πραγματικότητα, βεβαιώνει την πολυωνυμικού-χρόνου σύγκλιση. Αυτή η συνάρτηση είναι γνωστή ως δυνητική συνάρτηση, και δίνεται από τη σχέση

$$f(\mathbf{x}) \equiv \sum_{j=1}^n \ln \left[\frac{\mathbf{c}\mathbf{x}}{x_j} \right] = n \ln(\mathbf{c}\mathbf{x}) - \sum_{j=1}^n \ln(x_j) \quad (5.8\alpha)$$

όπου $\ln(\cdot)$ δηλώνουμε το φυσικό λογάριθμο με βάση e . { Παίρνοντας το αντίλογο της εξίσωσης (5.8α) μπορεί κάποιος ισοδύναμα να χρησιμοποιήσει την πολλαπλασιαστική δυνητική συνάρτηση $\mathbf{c}\mathbf{x} / [\prod_{j=1}^n x_j]^{1/n}$ αντί της εξίσωσης (5.8α). Ο λογαριθμικός μετασχηματισμός λαμβάνεται εδώ μόνο για τη μαθηματική ευκολία.}

Παρατηρούμε ότι ενώ ο προβολικός μετασχηματισμός (4.2) δε συντηρεί τη γραμμικότητα μιας συνάρτησης, όπως φαίνεται από τους «στόχους» στις εξισώσεις (4.1) και (4.3), συντηρεί τους λόγους των γραμμικών συναρτήσεων.

Πράγματι, υπό την εξίσωση (4.2) ισχύει, $\mathbf{c}\mathbf{x} / x_j = \mathbf{c}\mathbf{D}_k\mathbf{y} / x_{kj}y_j \equiv \bar{\mathbf{c}}\mathbf{y} / x_{ky}y_j$, έτσι ώστε η δυνητική συνάρτηση (5.8α) να μετασχηματίζεται σε

$$F(\mathbf{y}) \equiv f\left[\frac{\mathbf{D}_k\mathbf{y}}{\mathbf{1D}_k\mathbf{y}}\right] = \sum_{j=1}^n \ln\left[\frac{\bar{\mathbf{c}}\mathbf{y}}{x_{kj}y_j}\right] = n \ln(\bar{\mathbf{c}}\mathbf{y}) - \sum_{j=1}^n \ln(y_j) - \sum_{j=1}^n \ln(x_{kj}) \quad (5.8\beta)$$

Ας μετρήσουμε τώρα τη μείωση στην τιμή της δυνητικής συνάρτησης (5.8α), ή ισοδύναμα στη δυνητική συνάρτηση (5.8β) στο \mathbf{y} -διάστημα στην επανάληψη k . Χρησιμοποιώντας την πιθανή συνάρτηση (5.8β), παίρνουμε

$$F(\mathbf{y}_{\text{new}}) - F(\mathbf{y}_0) = n \ln\left[\frac{\bar{\mathbf{c}}\mathbf{y}_{\text{new}}}{\bar{\mathbf{c}}\mathbf{y}_0}\right] - \sum_{j=1}^n \ln[n y_{(\text{new})_j}]$$

αφού $y_{0j} = 1/n$ για $j = 1, \dots, n$. Επομένως από την εξίσωση (5.7) συνάγουμε ότι

$$\ln\left[\frac{\bar{\mathbf{c}}\mathbf{y}_{\text{new}}}{\bar{\mathbf{c}}\mathbf{y}_0}\right] \leq \ln\left[1 - \frac{\alpha}{(n-1)}\right] \leq -\frac{\alpha}{(n-1)}$$

βάσει ιδιότητας της συνάρτησης $\ln(\cdot)$. Χρησιμοποιώντας αυτή την εξίσωση, εξάγουμε

$$F(\mathbf{y}_{\text{new}}) - F(\mathbf{y}_0) \leq -\frac{n\alpha}{(n-1)} - \sum_{j=1}^n \ln[n y_{(\text{new})_j}] \quad (5.9)$$

Συνεπώς, αν το $-\sum_{j=1}^n \ln[n y_{(\text{new})_j}]$ δεν αυξάνεται τόσο πολύ σε σχέση με το

$\frac{n\alpha}{(n-1)}$, θα αποκτήσουμε μια επαρκή πτώση στη δυνητική συνάρτηση. Και

αυτό θα μας καταστήσει ικανούς να εγγυηθούμε τα επιθυμητά αποτελέσματα σύγκλισης. Στην πραγματικότητα, ορίζοντας $\bar{\alpha} = \frac{n\alpha}{(n-1)}$ και επιλέγοντας $0 < \alpha < 1$

αρκετά μικρό ώστε να $\sqrt{(n-1)/n} \bar{\alpha} = \sqrt{(n-1)/n} \alpha < 1$, θα δείξουμε ότι

$$\|n\mathbf{y} - \mathbf{1}\| \leq \sqrt{\frac{(n-1)}{n}} \bar{\alpha} < 1, \quad \mathbf{1}\mathbf{y} = 1 \quad \text{και} \quad \mathbf{y} > \mathbf{0}$$

συνεπάγεται ότι

$$0 \leq -\sum_{j=1}^n \ln [ny_j] \leq \frac{\bar{\alpha}^2}{2(1-\bar{\alpha})^2} \quad (5.10)$$

Παρατηρούμε ότι \mathbf{y}_{new} ικανοποιεί τις συνθήκες στην εξίσωση (5.10) αφού $\mathbf{1}\mathbf{y}_{\text{new}} = 1$, $\mathbf{y}_{\text{new}} > \mathbf{0}$, και από την εξίσωση (4.6α), $\|n\mathbf{y}_{\text{new}} - \mathbf{1}\| = n \alpha r = \sqrt{(n-1)/n} \bar{\alpha}$. Χρησιμοποιώντας εξίσωση (5.10) στην εξίσωση (5.9), αποκτάμε

$$F(\mathbf{y}_{\text{new}}) - F(\mathbf{y}_0) \leq -\bar{\alpha} + \frac{\bar{\alpha}^2}{2(1-\bar{\alpha})^2} \leq -\frac{1}{5}, \quad \text{όταν} \quad \bar{\alpha} \equiv \frac{n\alpha}{(n-1)} = \frac{1}{3} \quad (5.11)$$

Επομένως όταν $\bar{\alpha} = \frac{1}{3}$, η συνάρτηση $F(\cdot)$ και συνεπώς η δυνητική συνάρτηση

$f(\cdot)$ πέφτει κατά $\frac{1}{5}$ σε κάθε επανάληψη. Αυτό σημαίνει ότι σε περισσότερες

από k επαναλήψεις, $f(\mathbf{x}_k) - f(\mathbf{x}_0) \leq -k/5$. Αλλά

$$f(\mathbf{x}_k) - f(\mathbf{x}_0) = n \ln \left[\frac{\mathbf{c}\mathbf{x}_k}{\mathbf{c}\mathbf{x}_0} \right] - \sum_{j=1}^n \ln (nx_{kj})$$

αφού $\mathbf{x}_0 = \left(\frac{1}{n}, \dots, \frac{1}{n} \right)$. Γι' αυτό

$$n \ln \left[\frac{\mathbf{c}\mathbf{x}_k}{\mathbf{c}\mathbf{x}_0} \right] \leq \sum_{j=1}^n \ln (nx_{kj}) - \frac{k}{5}$$

Αλλά χρησιμοποιώντας το γεγονός ότι ο γεωμετρικός μέσος $\left[\prod_{j=1}^n (x_{kj}) \right]^{1/n}$ δεν

είναι πιο μεγάλος από τον αριθμητικό μέσο $\mathbf{1}\mathbf{x}_k / n$ και ότι $\mathbf{1}\mathbf{x}_k = 1$, έχουμε

$$\sum_{j=1}^n \ln (nx_{kj}) = n \ln \left[n \left(\prod_{j=1}^n x_{kj} \right)^{1/n} \right] \leq n \ln [\mathbf{1x}_k] = 0$$

Αυτό επομένως σημαίνει από την προηγούμενη ανισότητα ότι

$$\ln \left[\frac{\mathbf{cx}_k}{\mathbf{cx}_0} \right] \leq -\frac{k}{5n} \quad \text{για όλα τα } k = 0, 1, 2, \dots \quad (5.12)$$

Επομένως, παρόλο που η τιμή της αντικειμενικής συνάρτησης μπορεί να αυξάνεται από τη μια επανάληψη στην επόμενη, μια επαρκής ολική μείωση από την αρχική αντικειμενική τιμή στο $k=0$ διατηρείται, καθώς οι επαναλήψεις προχωρούν. Στην πραγματικότητα, από την εξίσωση (5.12), θα έχουμε $\mathbf{cx}_k < 2^{-L}$ όταν $\mathbf{cx}_k \leq (\mathbf{cx}_0) e^{-k/5n} < 2^{-L}$. Επομένως όταν $k = 10nL$, θα έχουμε

$$\mathbf{cx}_k \leq (\mathbf{cx}_0) e^{-k/5n} = \left(\sum_{j=1}^n c_j / n \right) (e^{-2L}) < (2^L)(2^{-2L}) = 2^{-L}$$

Επιπλέον κάθε επανάληψη απαιτεί όχι περισσότερο από $O(n^3)$ υπολογισμούς, και επομένως ο αλγόριθμος είναι πολυωνυμικής πολυπλοκότητας $O(n^4L)$.

Σημαντικό είναι να παρατηρήσουμε ότι

$$\mathbf{PP}^t = \begin{bmatrix} \mathbf{AD}_k^2 \mathbf{A}^t & \mathbf{0} \\ \mathbf{0} & n \end{bmatrix}$$

και ότι η μόνη αλλαγή στον πίνακα από τη μια επανάληψη στην επόμενη είναι στα στοιχεία του διαγώνιου πίνακα \mathbf{D}_k . Αναλόγως, ο Karmarkar δείχνει πως μια μικρή τροποποίηση του αλγόριθμου, βασιζόμενη στο updating και όχι στον επαναυπολογισμό ενός κατάλληλου αντίστροφου στην εφαρμογή προβολής κλίσης, μπορεί να γίνει για να τρέξει καταβάλλοντας προσπάθεια τάξης $O(n^{2.5})$ για κάθε επανάληψη. Αυτό έχει ως αποτέλεσμα μια γενικευμένη ελαττωμένη πολυωνυμική πολυπλοκότητα του $O(n^{3.5}L)$ για τον τροποποιημένο αλγόριθμο.

Ολοκληρώνουμε το κεφάλαιο, σκιαγραφώντας το προκύπτον της εξίσωσης (5.10). Θεωρείστε το ακόλουθο πρόβλημα βελτιστοποίησης, το οποίο μας καθιστά ικανούς να φράσουμε το $-\sum_{j=1}^n \ln [ny_j]$ υπό τις συνθήκες στην εξίσωση (5.10):

$$\text{Maximize } \left\{ -\sum_{j=1}^n \ln [ny_j] : (\mathbf{ny} - \mathbf{1})^t (\mathbf{ny} - \mathbf{1}) \leq \left(\frac{n-1}{n} \right) \bar{\alpha}^2, \mathbf{1y} = 1, \mathbf{y} \geq \mathbf{0} \right\} \quad (5.13)$$

Παρατηρούμε από τον πρώτο περιορισμό ότι αφού το δεξί μέλος του είναι μικρότερο από τη μονάδα, πρέπει να έχουμε $\mathbf{y} > \mathbf{0}$. Αφού η αντικειμενική συνάρτηση είναι ισοδύναμη στο να ελαχιστοποιεί το προϊόν $[y_1, y_2, \dots, y_n]$ μπορεί να φανεί ότι στη βελτιστοποίηση στο πρόβλημα (5.13), ο πρώτος περιορισμός είναι δεσμευμένος. Επιπλέον, μπορεί να φανεί ότι κάποιες $1 \leq q \leq (n-1)$ συνιστώσες μιας βέλτιστης λύσης είναι οι ίδιες και είναι ίσες με την τιμή U , ας πούμε, ενώ οι υπόλοιπες $(n-q)$ συνιστώσες είναι ίσες με κάποια κοινή τιμή V , όπου $0 < U < 1/n < V < 1$. Αφού $\mathbf{1y} = 1$ και $(\mathbf{ny} - \mathbf{1})^t (\mathbf{ny} - \mathbf{1}) = [(n-1)/n] \bar{\alpha}^2$ πρέπει να ικανοποιούνται, παίρνουμε $qU + (n-q)V = 1$ και $q(nU-1)^2 + (n-q)(nV-1)^2 = [(n-1)/n] \bar{\alpha}^2$. Αυτό δίνει $nU + (1-\theta_1)$, $nV + (1+\theta_2)$

$$\text{όπου} \quad \theta_1 = \sqrt{\frac{(n-1)(n-q)}{q} \frac{\bar{\alpha}}{n}} \quad \text{και} \quad \theta_2 \equiv \frac{q\theta_1}{(n-q)} \quad (5.14)$$

Σημειώνουμε ότι η βέλτιστη λύση είναι θετική, όπως απαιτείται. Επομένως για κάποια $1 \leq q \leq (n-1)$ η αντικειμενική τιμή στην εξίσωση (5.13) ικανοποιεί

$$-\sum_{j=1}^n \ln [ny_j] \leq -q \ln[1-\theta_1] - (n-q) \ln[1+\theta_2] \quad (5.15)$$

Από το γνωστό δεύτερης τάξης Taylor ή θεώρημα μέσης τιμής, εφαρμοσμένο στη συνάρτηση $\ln(\cdot)$, παίρνουμε για κάθε $0 < \theta < 1$ ότι $\ln[1-\theta] = -\theta - \theta^2/2\Delta^2$, όπου $(1-\theta) < \Delta < 1$. Αυτό ακολούθως σημαίνει $-\ln[1-\theta] \leq \theta + \theta^2/2(1-\theta)^2$. Παρομοίως $-\ln[1+\theta] \leq -\theta + \theta^2/2$ για κάθε $0 < \theta < 1$. Χρησιμοποιώντας αυτές τις ανισώσεις στην σχέση 5.15, μαζί με το ότι $0 < \theta_i < 1$ για $i = 1, 2$ και χρησιμοποιώντας τη σχέση 5.14, συμπεραίνουμε ότι

$$\begin{aligned} -\sum_{j=1}^n \ln [ny_j] &\leq -q \left[-\theta_1 + \frac{\theta_1^2}{2(1-\theta_1)^2} \right] + (n-q) \left[-\theta_2 + \frac{\theta_2^2}{2} \right] \\ &= \frac{q\theta_1^2}{2(1-\theta_1)^2} \left[\frac{n-q\theta_1(2-\theta_1)}{(n-q)} \right] \\ &= \frac{(n-1)}{n} \frac{\bar{\alpha}^2}{2(1-\theta_1)^2} \left[1 - \frac{q\theta_1(2-\theta_1)}{n} \right] \end{aligned}$$

$$\leq \frac{\bar{\alpha}^2}{2(1-\theta_1)^2} \leq \frac{\bar{\alpha}^2}{2(1-\bar{\alpha})^2}$$

αφού $0 < \theta_1 < \bar{\alpha} < 1$. Αυτό αποδεικνύει τη σχέση 5.10.

Βιβλιογραφία

Jiří Matoušek and Bernd Gärtner (2007). *Understanding and using Linear Programming*. Springer.

Mokhtar Bazaraa, John Jarvis and Hanif Sherali (1990). *Linear Programming and Network Flows*. Wiley.

Monteiro R.D.C. and Adler I. (1989). *Interior path following primal dual algorithms: Part I: Linear Programming*. *Mathematical Programming* 44,27-41.

S.G.Nash and A.Sofer (1995). *Linear and Nonlinear Programming*.
Mc Graw-Hill.

Strang (1995). *Γραμμική Άλγεβρα και Εφαρμογές* (απόδοση στα ελληνικά: Π.Παμφίλος). Πανεπιστημιακές Εκδόσεις Κρήτης.

Χαράλαμπος Μπότσαρης (2002). *Επιχειρησιακή Έρευνα-Τόμος Ι, Γραμμικός Προγραμματισμός και Θεωρία Παιγνίων*. Ελληνικά Γράμματα.

Κωνσταντίνος Παπαρρίζος(2005). *Γραμμικός Προγραμματισμός*. Εκδόσεις Πανεπιστημίου Μακεδονίας.