



Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης
Σχολή Θετικών Επιστημών
Τμήμα Μαθηματικών
Θεωρητική Πληροφορική και Θεωρία Συστημάτων και Ελέγχου

Εργασία για το μάθημα

Κυρτή Βελτιστοποίηση

Νίκος Β. Βασίλας

Θεσσαλονίκη , Ιανουάριος 2017

Περιεχόμενα

1	Εισαγωγή	5
2	Γραμμικός Προγραμματισμός	7
2.1	Επίλυση	9
2.1.1	Παράδειγμα 1	10
2.1.2	Παράδειγμα 2	11
2.1.3	Παράδειγμα 3	13
3	Τετραγωνικός Προγραμματισμός	16
3.1	Τετραγωνικό πρόγραμμα με περιορισμούς ισότητας	17
3.2	Συνθήκες βέλτιστης λύσης	18
3.3	Μέθοδος ενεργών περιορισμών	19
3.4	Επίλυση	23
3.4.1	Παράδειγμα	24
4	Κυρτά Σύνολα	27
4.1	Αφινικά και κυρτά σύνολα	27
4.1.1	Γραμμή και τμήμα γραμμής	27
4.1.2	Αφινικό σύνολο	27
4.1.3	Κυρτό Σύνολο	28
4.1.4	Κυρτός κώνος	28
4.2	Σημαντικά Παραδείγματα	29
4.2.1	Υπερεπίπεδα	29
4.2.2	Ημιχώρος	30
4.2.3	(Ευκλείδειες) Μπάλες	30
4.2.4	Ελλειψοειδές	30

4.2.5	Νόρμα	30
4.2.6	Νορμική μπάλα και νορμικός κώνος	31
4.2.7	Πολύεδρα	31
4.2.8	Θετικά ημιορισμένος κώνος	32
4.3	Πράξεις που διατηρούν την κυρτότητα	32
4.3.1	Τομή	32
4.3.2	Αφινικές συναρτήσεις	32
4.3.3	Γραμμική-κλασματική και προοπτική συνάρτηση	33
4.4	Γενικευμένες ανισότητες	34
4.4.1	Γνήσιοι κώνοι και γενικευμένες ανισότητες	34
4.4.2	Ελάχιστο και ελαχιστοτικό στοιχείο	35
4.5	Υπερεπίπεδο διαχωριστικό και στήριξης	35
4.5.1	Θεώρημα διαχωριστικού υπερεπιπέδου	35
4.5.2	Θεώρημα υπερεπιπέδου στήριξης	36
4.6	Δυϊκός κώνος και γενικευμένες ανισότητες	37
4.6.1	Δυϊκός κώνος	37
4.6.2	Δυϊκές γενικευμένες ανισότητες	37
4.6.3	Ελάχιστα και ελαχιστοτικά στοιχεία μέσω δυϊκών ανισοτήτων	38
5	Μέθοδοι Εσωτερικού Σημείου	39
5.1	Ιστορικά στοιχεία	39
5.2	Μέθοδος affine scaling	39
5.3	Αλγόριθμος path-following	43
5.3.1	Λογαριθμική συνάρτηση φράγματος και κεντρική διαδρομή	43
5.3.2	Κεντρική διαδρομή	45
5.3.3	Διατύπωση του αλγορίθμου path-following	47
5.4	Primal-dual μέθοδοι εσωτερικού σημείου	49
5.4.1	Primal-dual διεύθυνση εύρεσης	49
5.4.2	Διατύπωση του αλγορίθμου primal-dual	50
5.5	Ο αλγόριθμος του Karmarkar	51
5.5.1	Κανονική μορφή του Karmarkar	51
5.5.2	Το πρόβλημα του Karmarkar υπό περιορισμούς	52
5.5.3	Μετασχηματισμός στην κανονική μορφή του Karmarkar	53
5.5.4	Διατύπωση του αλγορίθμου του Karmarkar	54
6	Γραμμικές ανισότητες πινάκων	56

6.1	Θεωρητικά στοιχεία	57
6.2	Κατηγορίες γραμμικών ανισοτήτων πινάκων	58
6.2.1	Παράδειγμα 1: ευστάθεια	58
6.2.2	Παράδειγμα 2: εκτίμηση τετραγωνικού κόστους	59
6.3	Επίλυση	59
6.3.1	Παράδειγμα 1	62
6.3.2	Παράδειγμα 2	63
7	Συμπεράσματα	66
	Βιβλιογραφία	68

Κατάλογος Σχημάτων

2.1	Γραφική λύση του παραδείγματος 2 του γραμμικού προγραμματισμού	13
3.1	Αλγόριθμος ενεργών περιορισμών	22
3.2	Λύση παραδείγματος του τετραγωνικού προγραμματισμού	25
4.1	Κυρτά Σύνολα	28
4.2	Κυρτός Κώνος	29
4.3	Θεώρημα διαχωριστικού υπερεπιπέδου	36
4.4	Θεώρημα υπερεπιπέδου στήριξης	37
5.1	Αλγόριθμος path-following	47
5.2	Αλγόριθμος primal-dual εσωτερικού σημείου	50
5.3	Αλγόριθμος εσωτερικού σημείου του Karmarkar	54

Κεφάλαιο 1

Εισαγωγή

Επιστήμονες από όλους τους κλάδους όπως μηχανικοί, μαθηματικοί, οικονομολόγοι και άλλοι, έρχονται αντιμέτωποι με προβλήματα που πρέπει να λύσουν. Τα προβλήματα πολλές φορές απαιτούν ένα βέλτιστο σχεδιασμό, όπως κατανομή σπάνιων πόρων, σχεδιασμό βιομηχανικών δραστηριοτήτων, ή την εύρεση της τροχιάς ενός πυραύλου. Στο παρελθόν, ένα μεγάλο εύρος λύσεων θεωρούταν αποδεκτό. Στον μηχανολογικό σχεδιασμό για παράδειγμα, ήταν συνηθισμένη η χρήση μεγάλων παραγόντων ασφαλείας. Όμως, λόγω του συνεχούς αυξανόμενου ανταγωνισμού, δεν είναι πλέον αρκετός ο σχεδιασμός απλά αποδεκτών λύσεων. Σε άλλες περιπτώσεις, όπως για παράδειγμα ο σχεδιασμός ενός διαστημικού οχήματος, οι αποδεκτές λύσεις είναι περιορισμένες. Έτσι, γεννήθηκε η ανάγκη για απάντηση στα ερωτήματα όπως, αν κάνουμε αποδοτική χρήση των σπάνιων πόρων, ή αν μπορούμε να κάνουμε ένα οικονομικότερο σχεδιασμό. Ως απάντηση στο όλο και αυξανόμενο πεδίο τέτοιων ερωτήσεων, υπήρχε μία πολύ γρήγορη ανάπτυξη μοντέλων και τεχνικών βελτιστοποίησης. Επίσης, η παράλληλη ραγδαία εξέλιξη των υπολογιστικών δυνατοτήτων βοήθησε σημαντικά στη χρήση των τεχνικών που αναπτύχθηκαν.

Η ιδέα της βελτιστοποίησης είναι βαθιά ριζωμένη ως η βασική αρχή στην ανάλυση πολλών πολύπλοκων αποφάσεων. Χρησιμοποιώντας την ιδέα της βελτιστοποίησης, προσεγγίζουμε ένα πολύπλοκο πρόβλημα, που περιλαμβάνει την επιλογή τιμών για έναν αριθμό συγγενικών μεταβλητών, εστιάζοντας σε μία αντικειμενική συνάρτηση που σχεδιάστηκε για να ποσοτικοποιήσει την επίδοση και να μετρήσει την ποιότητα της απόφασης. Αυτή η αντικειμενική μεγιστοποιείται ή ελαχιστοποιείται και υπόκειται σε περιορισμούς που ενδέχεται να περιορίσουν την επιλογή των μεταβλητών απόφασης. Αν μία κατάλληλη πτυχή τους προβλήματος

μπορεί να απομονωθεί και να χαρακτηριστεί από μία αντικειμενική, είτε αυτό είναι κέρδος ή ζημία στον επιχειρηματικό τομέα, είτε ταχύτητα ή απόσταση σε ένα φυσικό πρόβλημα και τα λοιπά, η βελτιστοποίηση είναι το κατάλληλο εργαλείο για την ανάλυση των προβλημάτων αυτών.

Είναι φυσικά εξαιρετικά σπάνια η περίπτωση όπου είναι εφικτό να περιγράψουμε πλήρως όλες τις περίπλοκες αλληλεπιδράσεις των μεταβλητών, των περιορισμών και να σχεδιάσουμε αντικειμενική συνάρτηση που αντιπροσωπεύει ένα πολύπλοκο πρόβλημα απόφασης. Έτσι, η διατύπωση ενός προβλήματος βελτιστοποίησης μπορεί να θεωρηθεί μόνο ως προσέγγιση του πραγματικού προβλήματος. Η ικανότητα στη μοντελοποίηση, δηλαδή να συμπεριλάβουμε όλα τα σημαντικά στοιχεία ενός προβλήματος, και η σωστή κρίση στην ερμηνεία των αποτελεσμάτων είναι απαραίτητα για την εξαγωγή ορθών συμπερασμάτων. Η βελτιστοποίηση συνεπώς, πρέπει να θεωρείται ως ένα εργαλείο για την καλύτερη κατανόηση και ανάλυση του προβλήματος.

Η ικανότητα και η ορθή κρίση, όσον αφορά τη διατύπωση του προβλήματος και την ερμηνεία των αποτελεσμάτων, ενισχύεται μέσω της πρακτικής εμπειρίας και τη μελέτη της θεωρίας. Η διατύπωση του προβλήματος πάντα περιλαμβάνει κάποιο συμβιβασμό μεταξύ της αύξησης της μαθηματικής πολυπλοκότητας του προβλήματος και την απόκλιση του μοντέλου από το πραγματικό πρόβλημα. Για να μπορέσει κάποιος να επιλέξει τη χρυσή τομή πρέπει να είναι σε θέση να αναγνωρίσει την ιδιαιτερότητα του κάθε προβλήματος, κάτι που είναι εφικτό μόνο με τη βαθιά μελέτη του τομέα.

Στην παρούσα εργασία παρουσιάζουμε κάποιους βασικούς τομείς της βελτιστοποίησης. Σε κάθε κεφάλαιο αναλύουμε τη θεωρία που διέπει την κάθε κατηγορία προβλημάτων, αναλύουμε κάποιον αλγόριθμο για την επίλυση του προβλήματος και δίνουμε κάποια τυπικά παραδείγματα στο MATLAB. Η εισαγωγή βασίστηκε ως επί τον πλείστον στα βιβλία [1] και [9], και ο ενδιαφερόμενος αναγνώστης παραπέμπεται εκεί για περαιτέρω πληροφορίες.

Κεφάλαιο 2

Γραμμικός Προγραμματισμός

Ο στόχος του γραμμικού προγραμματισμού, σύμφωνα με [17], είναι να προσδιορίσει τις τιμές των μεταβλητών που μεγιστοποιούν ή ελαχιστοποιούν μια γραμμική αντικειμενική συνάρτηση, όπου οι μεταβλητές υπόκεινται σε κάποιους γραμμικούς περιορισμούς. Το πρόβλημα του γραμμικού προγραμματισμού είναι μία ειδική κατηγορία ενός γενικού προβλήματος βελτιστοποίησης με περιορισμούς. Στη γενική περίπτωση, ο στόχος είναι να βρούμε ένα σημείο που ελαχιστοποιεί την αντικειμενική συνάρτηση και ταυτόχρονα ικανοποιεί τους περιορισμούς. Οποιοδήποτε σημείο ικανοποιεί τους περιορισμούς ονομάζεται εφικτό σημείο. Στο πρόβλημα του γραμμικού προγραμματισμού, η αντικειμενική συνάρτηση είναι γραμμική και το σύνολο των εφικτών σημείων καθορίζεται από γραμμικές σχέσεις ισότητας και/ή ανισότητας.

Οι μέθοδοι για τη λύση προβλημάτων γραμμικού προγραμματισμού παρέχουν τη δυνατότητα της επιλογής του καλύτερου εφικτού σημείου ανάμεσα στα πολλά εφικτά σημεία. Γενικά, τα προβλήματα γραμμικού προγραμματισμού έχουν πολύ μεγάλο μέγεθος, επομένως είναι απαραίτητοι αποδοτικοί μέθοδοι επίλυσης αυτών, που πρωτοεμφανίστηκαν τέλη της δεκαετίας του 1930. Συγκεκριμένα το 1939, ο Kantorovich [5], παρουσίασε διάφορες λύσεις σε προβλήματα σχετικά με το σχεδιασμό παραγωγής και μεταφορών. Κατά τη διάρκεια του Δεύτερου Παγκοσμίου Πολέμου, ο Koopmans [8] συνεισέφερε σημαντικά στη λύση προβλημάτων μεταφοράς. Μάλιστα, και στους δύο προαναφερθέντες απονεμήθηκε το Νόμπελ Οικονομίας το 1975 για τη συνεισφορά τους στη βέλτιστη κατανομή πόρων. Το 1947, ο Dantzig [3] ανέπτυξε μία καινούργια μέθοδο λύσης γραμμικών προγραμμάτων, γνωστή ως μέθοδος simplex. Η μέθοδος αυτή είναι αποδοτική, παρέχει μία κομψή λύση για το πρόβλημα και χρησιμοποιείται ευρέως ως και σήμερα για τα προβλήματα του γραμμικού προγραμματι-

σμού. Παρόλο αυτά, ο αριθμός των βημάτων, και ως εκ τούτου ο συνολικός χρόνος, για την εύρεση λύσης αυξάνει εκθετικά με το μέγεθος των μεταβλητών. Αυτό οδήγησε στο ενδιαφέρον για τη σχεδίαση αλγορίθμων που θα λύνουν γραμμικά προγράμματα με πολυωνυμική πολυπλοκότητα, δηλαδή αλγόριθμοι που θα βρίσκουν τη βέλτιστη λύση σε χρόνο που θα εξαρτάται πολυωνυμικά με τον αριθμό των μεταβλητών. Ο Khachiyan το 1979 [7], ήταν ο πρώτος που ανέπτυξε έναν τέτοιο αλγόριθμο, ωστόσο έλαβε περισσότερο θεωρητικό παρά πρακτικό ενδιαφέρον. Το 1984, ο Karmarkar [6] πρότεινε ένα νέο αλγόριθμο γραμμικού προγραμματισμού με πολυωνυμική πολυπλοκότητα, που φαίνεται να λύνει προβλήματα αρκετά πιο αποδοτικά από ότι η μέθοδος simplex. Η δουλειά του Karmarkar οδήγησε σε άλλες μεθόδους, που δε βασίζονται στη μέθοδο simplex, γνωστές ως μέθοδοι εσωτερικού σημείου (interior-point).

Ένα τυπικό γραμμικό πρόγραμμα βελτιστοποίησης είναι της μορφής

$$\begin{aligned} & \text{minimize} && \mathbf{c}^T \mathbf{x} \\ & \text{subject to} && \mathbf{A}\mathbf{x} = \mathbf{b} \\ & && \mathbf{x} \geq \mathbf{0}, \end{aligned} \tag{2.1}$$

όπου $\mathbf{c} \in \mathbb{R}^n$, $\mathbf{b} \in \mathbb{R}^m$, και $\mathbf{A} \in \mathbb{R}^{m \times n}$. Ο περιορισμός ανισότητας $\mathbf{x} \geq \mathbf{0}$ για τις μεταβλητές σημαίνει ότι κάθε συνιστώσα του διανύσματος \mathbf{x} είναι μη αρνητική. Πολλές παραλλαγές του παραπάνω προβλήματος είναι δυνατές, όπως για παράδειγμα, αντί για ελαχιστοποίηση της αντικειμενικής συνάρτησης να ζητούμε τη μεγιστοποίηση αυτής. Επίσης, οι περιορισμοί θα μπορούσαν αντί ισότητας, να είχαν τη μορφή ανισότητας, όπως $\mathbf{A}\mathbf{x} \geq \mathbf{b}$, ή $\mathbf{A}\mathbf{x} \leq \mathbf{b}$. Αναφερόμαστε σε αυτές τις παραλλαγές ως γραμμικό πρόγραμμα, καθώς είναι δυνατόν να γραφτούν στην παραπάνω μορφή.

Κάθε γραμμικό πρόγραμμα ικανοποιεί έμμεσα κάποιες υποθέσεις, όπως περιγράφονται στο [18]. Η **αναλογικότητα** (proportionality) αναφέρεται σε ξεχωριστές δραστηριότητες, που εξετάζονται ανεξάρτητα από τις υπόλοιπες. Ας υποθέσουμε ότι από τις n δραστηριότητες μόνο μία υλοποιείται και ας την ονομάσουμε k , έτσι ώστε, $x_j = 0$, για $j = 1, \dots, n$, εκτός από $j = k$. Σύμφωνα με αυτή την υπόθεση το μέτρο αποτελεσματικότητας είναι ίσο με $c_k x_k$ και η χρησιμοποίηση κάθε πόρου i είναι ίση με $a_{ik} x_k$. Αυτό σημαίνει ότι οι δύο ποσότητες είναι ευθέως ανάλογες προς το επίπεδο της k δραστηριότητας. Ακόμη σημαίνει ότι δεν υπάρχουν αρχικές επιβαρύνσεις με την έναρξη της δραστηριότητας και ότι η αναλογικότητα ισχύει για όλα τα επίπεδα τιμών της δραστηριότητας.

Με την υπόθεση της αναλογικότητας δεν εξασφαλίζεται ότι η αντικειμενική συνάρτηση και

οι περιορισμοί είναι γραμμικές συναρτήσεις. Η **προσθετικότητα** (additivity) προϋποθέτει ότι δεν υπάρχουν αλληλεπιδράσεις μεταξύ των δραστηριοτήτων. Για κάθε επίπεδο δραστηριοτήτων (x_1, \dots, x_n) η συνολική χρησιμοποίηση κάθε πόρου καθώς και το συνολικό μέτρο αποτελεσματικότητας είναι ίσα με το άθροισμα των αντίστοιχων ποσοτήτων κάθε δραστηριότητας.

Μερικές φορές οι μεταβλητές αποφάσεων έχουν έννοια μόνο όταν παίρνουν ακέραιες τιμές. Η λύση όμως που παίρνουμε από το γραμμικό προγραμματισμό συχνά έχει μη ακέραιες τιμές. Με την υπόθεση της **διαιρετότητας** (divisibility) οι μονάδες δραστηριότητας μπορούν να διαιρεθούν σε οποιοδήποτε κλασματικό επίπεδο, έτσι ώστε οι μη ακέραιες τιμές για τις μεταβλητές αποφάσεων είναι επιτρεπτές.

2.1 Επίλυση

Στη συνέχεια θα παρουσιάσουμε κάποια παραδείγματα γραμμικού προγραμματισμού. Η επίλυση αυτών έγινε με το λογισμικό MATLAB και τη χρήση της βιβλιοθήκης βελτιστοποίησης (Optimization Toolbox) αυτού. Το MATLAB δέχεται ένα γραμμικό πρόγραμμα στην παρακάτω μορφή

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & \mathbf{f}^T \mathbf{x} \\ \text{subject to} \quad & \mathbf{A}\mathbf{x} \leq \mathbf{b} \\ & \mathbf{A}_{eq}\mathbf{x} = \mathbf{b}_{eq} \\ & \mathbf{lb} \leq \mathbf{x} \leq \mathbf{ub}. \end{aligned} \tag{2.2}$$

Στην παραπάνω σχέση, \mathbf{f} είναι το διάνυσμα των συντελεστών της αντικειμενικής συνάρτησης, \mathbf{A} είναι το μητρώο των συντελεστών των ανισοτήτων και \mathbf{b} το αντίστοιχο διάνυσμα. Ακόμη, οι σχέσεις ισότητας περιγράφονται από το μητρώο \mathbf{A}_{eq} και το διάνυσμα \mathbf{b}_{eq} . Τέλος, \mathbf{lb} και \mathbf{ub} είναι τα διανύσματα του κάτω και άνω ορίου του διανύσματος \mathbf{x} .

Η λύση γραμμικών προγραμμάτων γίνεται με τη συνάρτηση **linprog**. Η σύνταξή της είναι

$$[\mathbf{x}, \mathbf{fval}, \text{exitflag}, \text{output}, \text{lambda}] = \text{linprog}(\mathbf{f}, \mathbf{A}, \mathbf{b}, \mathbf{Aeq}, \mathbf{beq}, \mathbf{lb}, \mathbf{ub}, \text{options})$$

με απαραίτητα τα τρία πρώτα ορίσματα της συνάρτησης. Επιστρέφει τη βέλτιστη λύση \mathbf{x} , την τιμή της αντικειμενικής συνάρτησης \mathbf{fval} , τη συνθήκη εξόδου exitflag , τη δομή (structure) output , σχετικά με το τέλος του προγράμματος και τους πολλαπλασιαστές Lagrange lambda . Επίσης, από τη δομή options , μπορούμε να ρυθμίσουμε διάφορες παραμέτρους της διαδι-

κασίας επίλυσης, όπως για παράδειγμα να επιλέξουμε αλγόριθμο επίλυσης (interior-point, dual-simplex), να θέσουμε ανοχές, δηλαδή το κατά πόσο θα ικανοποιείται η αντικειμενική συνάρτηση, οι περιορισμοί, και άλλα πολλά. Ο ενδιαφερόμενος παραπέμπεται στη βοήθεια του MATLAB.

2.1.1 Παράδειγμα 1

Το πρώτο παράδειγμα προέρχεται από τη βοήθεια του MATLAB και έχει περιορισμούς ανισότητας, ισότητας καθώς και όρια στις μεταβλητές. Η αντικειμενική συνάρτηση που θέλουμε να ελαχιστοποιήσουμε είναι

$$f(\mathbf{x}) = -x_1 - \frac{1}{3}x_2,$$

και οι σχέσεις ανισότητας που περιγράφουν το πρόβλημα είναι

$$\begin{aligned}x_1 + x_2 &\leq 2 \\x_1 + \frac{1}{4}x_2 &\leq 1 \\x_1 - x_2 &\leq 2 \\-\frac{1}{4}x_1 - x_2 &\leq 1 \\-x_1 - x_2 &\leq -1 \\-x_1 + x_2 &\leq 2.\end{aligned}$$

Επίσης ισχύει ο περιορισμός ισότητας

$$x_1 + \frac{1}{4}x_2 = \frac{1}{2}$$

και τέλος έχουμε τα όρια

$$-1 \leq x_1 \leq 1.5, \quad -0.5 \leq x_2 \leq 1.25.$$

Το παραπάνω πρόβλημα εύκολα γράφεται σε μητρωική μορφή και έτσι από τους περιορισμούς ανισότητας έχουμε

$$\mathbf{A} = \begin{bmatrix} 1 & 1 \\ 1 & \frac{1}{4} \\ 1 & -1 \\ -\frac{1}{4} & -1 \\ -1 & -1 \\ -1 & 1 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 2 \\ 1 \\ 2 \\ 1 \\ -1 \\ 2 \end{bmatrix}.$$

Ακόμη από τον περιορισμό ισότητας $\mathbf{A}_{eq} = \begin{bmatrix} 1 & \frac{1}{4} \end{bmatrix}$, $b_{eq} = \frac{1}{2}$, από τα κάτω και άνω όρια προκύπτει $\mathbf{lb} = [-1 \quad -0.5]^T$, $\mathbf{ub} = [1.5 \quad 1.25]^T$ και τέλος από την αντικειμενική συνάρτηση έχουμε το διάνυσμα $\mathbf{f} = [-1 \quad -\frac{1}{3}]^T$. Λύνοντας το πρόβλημα, βρίσκουμε ότι η βέλτιστη λύση είναι $\mathbf{x}^* = [0.1875 \quad -1.25]^T$ και η αντίστοιχη τιμή της αντικειμενικής $f(\mathbf{x}^*) = -0.6042$. Παρακάτω παρατίθεται ο κώδικας που χρησιμοποιήθηκε για να λυθεί το πρόβλημα.

```
1 function [x, fval] = lp1()
2     A = [1, 1; 1, 1/4; 1, -1
3         -1/4, -1; -1, -1; -1, 1];
4     b = [2, 1, 2, 1, -1, 2];
5
6     Aeq = [1, 1/4];
7     beq = 1/2;
8
9     lb = [-1; -0.5];
10    ub = [1.5; 1.25];
11    f = [-1, -1/3];
12    [x, fval, exitflag, output] = linprog(f, A, b, Aeq, beq, lb, ub);
13 end
```

2.1.2 Παράδειγμα 2

Το επόμενο παράδειγμα πρόκειται για πρόβλημα μεγιστοποίησης και καθότι πιο απλό, είναι ιδιαίτερα χρήσιμο για να δούμε γεωμετρικά τη λύση του γραμμικού προγράμματος. Το

πρόβλημα είναι το εξής,

$$\begin{aligned} & \underset{x}{\text{maximize}} && 4x + 3y \\ & \text{subject to} && 2y \leq 25 - x \\ & && 4y \geq 2x - 8 \\ & && y \leq 2x - 5 \\ & && x \geq 0 \\ & && y \geq 2, \end{aligned}$$

ή στην ισοδύναμη συμπαγή μορφή

$$\begin{aligned} & \underset{x}{\text{minimize}} && \mathbf{f}^T \mathbf{x} \\ & \text{subject to} && \mathbf{A}\mathbf{x} \leq \mathbf{b}, \end{aligned}$$

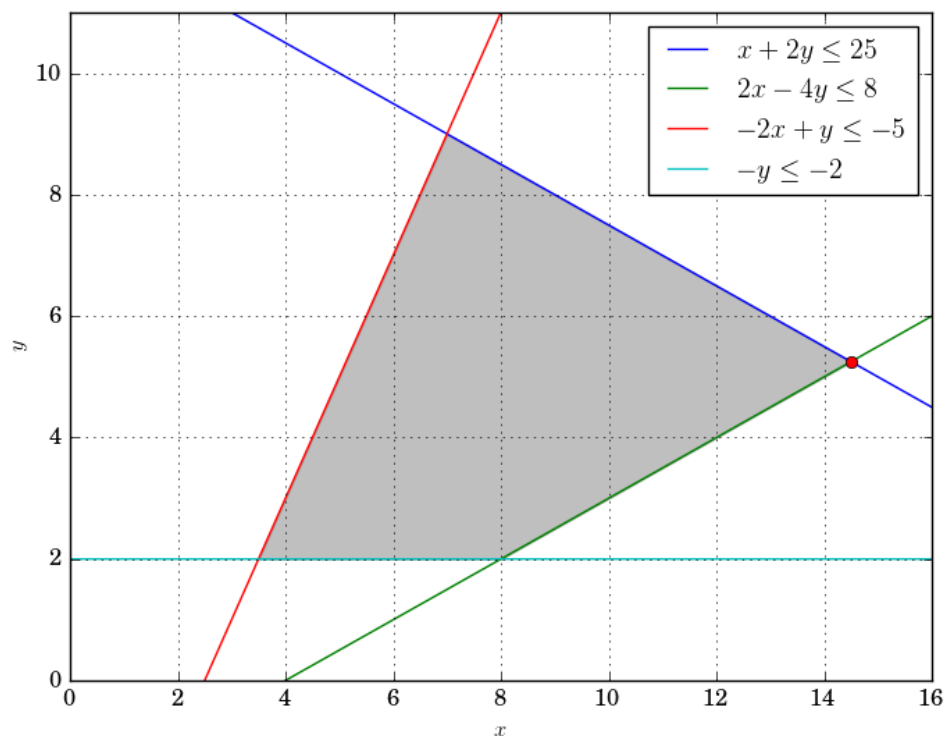
όπου $\mathbf{x} = \begin{bmatrix} x & y \end{bmatrix}^T$, $\mathbf{f} = \begin{bmatrix} -4 & -3 \end{bmatrix}^T$ και

$$\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 2 & -4 \\ -2 & 1 \\ -1 & 0 \\ 0 & -1 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 25 \\ 8 \\ -5 \\ 0 \\ -2 \end{bmatrix}.$$

Είναι προφανές ότι ένα πρόβλημα μεγιστοποίησης είναι το αντίθετο από ένα πρόβλημα ελαχιστοποίησης, για αυτό μπορούμε εύκολα να γράψουμε το γραμμικό πρόγραμμα σε μορφή που αναγνωρίζει το MATLAB. Λύνοντας το πρόβλημα, βρίσκουμε ότι η βέλτιστη λύση είναι $\mathbf{x}^* = \begin{bmatrix} 14.5 & 5.25 \end{bmatrix}^T$ και η αντίστοιχη τιμή της αντικειμενικής $f(\mathbf{x}^*) = 73.75$. Παρακάτω παρατίθεται ο κώδικας που χρησιμοποιήθηκε για να λυθεί το πρόβλημα.

```
1 function [x, fval] = lp2()
2     A = [1, 2; 2, -4; -2, 1; -1, 0; 0, -1];
3     b = [25; 8; -5; 0; -2];
4     f = [-4; -3];
5     [x, fval, exitflag, output] = linprog(f, A, b);
6     fval = -fval;
7 end
```

Παρακάτω παρουσιάζεται η γραφική λύση του γραμμικού προγράμματος. Όπως φαίνεται από το σχήμα 2.1 οι περιορισμοί του προβλήματος δημιουργούν ένα πολύγωνο, γκρι



Σχήμα 2.1: Γραφική λύση του παραδείγματος 2 του γραμμικού προγραμματισμού

σκιαγράφηση, όπου είναι ο χώρος των δυνατών λύσεων. Τελικά από όλα τα σημεία του πολυγώνου το βέλτιστο φαίνεται στο σχήμα με την κόκκινη κουκκίδα και είναι αυτό που μεγιστοποιεί την αντικειμενική συνάρτηση.

2.1.3 Παράδειγμα 3

Τα δεδομένα του επόμενου παραδείγματος προέρχονται από εδώ. Σύμφωνα με το παράδειγμα πρόκειται για ένα πραγματικό πρόβλημα γραμμικού προγραμματισμού. Μία εταιρία που φτιάχνει τροφές για γάτες, επιδιώκει το χαμηλότερο δυνατό κόστος καθώς και την επίτευξη των διατροφικών στόχων που έχουν τεθεί. Οι στόχοι αυτοί για τα 100g προϊόντος, είναι τουλάχιστον 8g πρωτεΐνης και 6g λιπαρών, αλλά όχι παραπάνω από 2g φυτικών ινών και 0.4g αλατιού. Επιδιώκουν δηλαδή να βρουν τη σωστή αναλογία των βασικών συστατικών που μπορούν να χρησιμοποιήσουν(τα βασικά συστατικά είναι κοτόπουλο, βοδινό, αρνί, ρύζι, σιτάρι και τζελ), ούτως ώστε να ικανοποιηθούν οι στόχοι που αναφέραμε. Τα κόστη ανά

γραμμάριο του κοτόπουλου, του βοδινού και του αρνιού είναι 0.013\$, 0.008\$ και 0.01\$ αντίστοιχα, ενώ του ρυζιού, του σιταριού και του τζελ 0.002\$, 0.005\$ και 0.001\$ αντίστοιχα. Κάθε βασικό συστατικό συνεισφέρει στο συνολικό βάρος πρωτεΐνης, λιπαρών, φυτικών ινών και αλατιού στο τελικό προϊόν. Η συνεισφορά του καθενός (σε γραμμάρια) ανά γραμμάριο συστατικού παρουσιάζονται στο παρακάτω πίνακα. Αρχικά θα ορίσουμε ως μεταβλητές, το

	Πρωτεΐνη	Λιπαρά	Φυτικές Ίνες	Αλάτι
Κοτόπουλο	0.1	0.08	0.001	0.002
Χοιρινό	0.2	0.1	0.005	0.005
Αρνί	0.15	0.11	0.003	0.007
Ρύζι	0	0.01	0.1	0.002
Σιτάρι	0.04	0.01	0.15	0.008
Τζελ	0	0	0	0

Πίνακας 2.1: Πίνακας Συστατικών

ποσοστό των συστατικών που μπορεί να περιέχεται στη συσκευασία. Επειδή η συσκευασία είναι 100g, τα ποσοστά αυτά αντιπροσωπεύουν την ποσότητα των γραμμαρίων του κάθε συστατικού. Ορίζουμε ως x_1 το ποσοστό του κοτόπουλου που περιέχεται στη συσκευασία, x_2 το αντίστοιχο ποσοστό βοδινού, x_3 το ποσοστό αρνιού, x_4 ρυζιού, x_5 φυτικών ινών και x_6 το ποσοστό του τζελ. Στόχος είναι η μείωση του κόστους, και ορίζουμε ως αντικειμενική συνάρτηση προς ελαχιστοποίηση

$$0.013x_1 + 0.008x_2 + 0.01x_3 + 0.002x_4 + 0.005x_5 + 0.001x_6,$$

σε δολάρια. Οι περιορισμοί του προβλήματος, έχουν περιγραφεί παραπάνω και μπορούν να γραφτούν

$$0.1x_1 + 0.2x_2 + 0.15x_3 + 0.04x_5 \geq 8$$

$$0.08x_1 + 0.1x_2 + 0.11x_3 + 0.01x_4 + 0.01x_5 \geq 6$$

$$0.001x_1 + 0.005x_2 + 0.003x_3 + 0.1x_4 + 0.15x_5 \leq 2$$

$$0.002x_1 + 0.005x_2 + 0.007x_3 + 0.002x_4 + 0.008x_5 \leq 2.$$

Επίσης, έχουμε έναν προφανή περιορισμό ισότητας, ότι το άθροισμα των συστατικών πρέπει να ισούται με όλη τη συσκευασία, δηλαδή

$$x_1 + x_2 + x_3 + x_4 + x_5 + x_6 = 100.$$

Τέλος, για κάθε ένα συστατικό i , έχουμε ένα κάτω και ένα άνω όριο όπως φαίνεται στην παρακάτω σχέση

$$0 \leq x_i \leq 100.$$

Το παραπάνω πρόβλημα εύκολα γράφεται σε μητρωική μορφή και έτσι από τους περιορισμούς ανισότητας έχουμε

$$\mathbf{A} = \begin{bmatrix} -0.1 & -0.2 & -0.15 & 0 & -0.04 & 0 \\ -0.08 & -0.1 & -0.11 & -0.01 & -0.01 & 0 \\ 0.001 & 0.005 & 0.003 & 0.1 & 0.15 & 0 \\ 0.002 & 0.005 & 0.007 & 0.002 & 0.008 & 0 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} -8 \\ -6 \\ 2 \\ 0.4 \end{bmatrix}.$$

Ακόμη από τον περιορισμό ισότητας $\mathbf{A}_{eq} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}$, $b_{eq} = 100$, από τα κάτω και άνω όρια προκύπτει $\mathbf{lb} = \mathbf{0}_{6 \times 1}^T$, $\mathbf{ub} = 100\mathbf{I}_{6 \times 1}^T$ και τέλος από την αντικειμενική συνάρτηση έχουμε το διάνυσμα $\mathbf{f} = \begin{bmatrix} 0.013 & 0.008 & 0.01 & 0.002 & 0.005 & 0.001 \end{bmatrix}^T$. Λύνοντας το πρόβλημα, βρίσκουμε ότι η βέλτιστη λύση είναι $\mathbf{x}^* = \begin{bmatrix} 0 & 60 & 0 & 0 & 0 & 40 \end{bmatrix}^T$, δηλαδή 60% βοδινό κρέας και 40% τζελ, και η αντίστοιχη τιμή της αντικειμενικής $f(\mathbf{x}^*) = 0.52$ δολάρια ανά συσκευασία. Παρακάτω παρατίθεται ο κώδικας που χρησιμοποιήθηκε για να λυθεί το πρόβλημα.

```

1 function [x, fval] = lp3()
2     A = [-0.1, -0.2, -0.15, 0., -0.04, 0;
3         -0.08, -0.1, -0.11, -0.01, -0.01, 0;
4         0.001, 0.005, 0.003, 0.1, 0.15, 0;
5         0.002, 0.005, 0.007, 0.002, 0.008, 0];
6     b = [-8.; -6.; 2.; 0.4];
7     Aeq = ones(1,6);
8     beq = 100.;
9     lb = zeros(6, 1);
10    ub = 100.*ones(6, 1);
11    f = [0.013, 0.008, 0.01, 0.002, 0.005, 0.001];
12    [x,fval,exitflag,output] = linprog(f,A,b,Aeq,beq,lb,ub);
13 end

```

Κεφάλαιο 3

Τετραγωνικός Προγραμματισμός

Μία ειδική κατηγορία μη-γραμμικού προγραμματισμού παρουσιάζεται όταν η αντικειμενική συνάρτηση f είναι τετραγωνική και οι περιορισμοί είναι γραμμικοί. Τότε το πρόβλημα ονομάζεται **Τετραγωνικός Προγραμματισμός** και είναι της μορφής

$$\begin{aligned} \min_x \quad & q(x) = \frac{1}{2}x^T Gx + x^T c \\ \text{subject to} \quad & a_i^T x = b_i, \quad i \in \mathcal{E}, \\ & a_i^T x \geq b_i, \quad i \in \mathcal{I}, \end{aligned} \tag{3.1}$$

όπου G είναι $n \times n$ συμμετρικό μητρώο, \mathcal{E} και \mathcal{I} είναι πεπερασμένα σύνολα των δεικτών, και c, x , και $\{a_i\}, i \in \mathcal{E} \cup \mathcal{I}$, είναι διανύσματα στο \mathbb{R}^n . Τετραγωνικά προγράμματα μπορούν να λυθούν αριθμητικά (αν έχουν λύση) και η δυσκολία να βρεθεί λύση εξαρτάται από τα χαρακτηριστικά της αντικειμενικής συνάρτησης και τον αριθμό των περιορισμών. Έχουν αναπτυχθεί πολλές μέθοδοι επίλυσης προβλημάτων τετραγωνικού προγραμματισμού που περιέχουν περιορισμούς ισότητας και ανισότητας. Οι μέθοδοι ενεργών περιορισμών (active set) χρησιμοποιούνται ευρέως από το 1970 και είναι αποτελεσματικοί για μικρό και μεσαίου μεγέθους προβλήματα. Επιτρέπουν αποτελεσματική ανίχνευση της μη φραγμένης λύσης και τη μη εφικτότητας και γενικά επιστρέφουν μία ακριβή εκτίμηση της βέλτιστης λύσης. Οι μέθοδοι εσωτερικού σημείου (interior point) είναι πιο πρόσφατοι και έγιναν ιδιαίτερα διάσημοι από το 1990 και μετά. Είναι κατάλληλοι για μεγάλου μεγέθους προβλήματα αλλά μπορεί να μην είναι οι πιο αποδοτικοί όταν θέλουμε να λύσουμε παρόμοια τετραγωνικά προγράμματα, ο ενδιαφερόμενος αναγνώστης παραπέμπεται στο [13]. Μία άλλη μέθοδος είναι η προβολή κλίσης (gradient projection) και πρόκειται για ειδική περίπτωση της μεθόδου

των ενεργών περιορισμών, που είναι περισσότερο αποδοτική όταν οι μόνοι περιορισμοί είναι τα όρια στις μεταβλητές. Στην παρούσα εργασία θα παρουσιάσουμε την μέθοδο των ενεργών περιορισμών σύμφωνα με τη βοήθεια του [14].

3.1 Τετραγωνικό πρόγραμμα με περιορισμούς ισότητας

Ξεκινάμε αναφέροντας την ειδική περίπτωση του γενικού προβλήματος (3.1), καθώς πολλοί αλγόριθμοι λύνουν σε κάθε επανάληψη ένα υποπρόγραμμα με περιορισμούς ισότητας. Το τετραγωνικό πρόγραμμα μπορεί να γραφτεί σε μητρική μορφή

$$\begin{aligned} \min_x \quad & q(x) := \frac{1}{2}x^T Gx + x^T c \\ \text{subject to} \quad & Ax = b, \end{aligned} \quad (3.2)$$

όπου A είναι η $m \times n$ Ιακωβιανή των περιορισμών με γραμμές τα $a_i^T, i \in \mathcal{E}$ και b είναι διάνυσμα στο \mathbb{R}^m με συνιστώσες τα $b_i, i \in \mathcal{E}$.

Οι πρώτης τάξης αναγκαίες συνθήκες για να είναι το x^* λύση του (3.2) τότε υπάρχει διάνυσμα λ^* τέτοιο ώστε να ικανοποιείται το σύστημα

$$\begin{pmatrix} G & -A^T \\ A & 0 \end{pmatrix} \begin{pmatrix} x^* \\ \lambda^* \end{pmatrix} = \begin{pmatrix} -c \\ b \end{pmatrix}.$$

Το διάνυσμα λ^* ονομάζονται πολλαπλασιαστές Lagrange. Το παραπάνω σύστημα μπορεί να γραφτεί σε μορφή βολική για τον υπολογισμό εκφράζοντας το x^* ως $x^* = x + p$, όπου x είναι μία εκτίμηση της λύσης και p είναι το επιθυμητό βήμα. Με αυτόν το συμβολισμό προκύπτει

$$\begin{pmatrix} G & -A^T \\ A & 0 \end{pmatrix} \begin{pmatrix} -p \\ \lambda^* \end{pmatrix} = \begin{pmatrix} g \\ h \end{pmatrix}, \quad (3.3)$$

όπου

$$hAx - b, \quad g = c + Gx, \quad p = x^* - x.$$

Το μητρώο στη (3.3) ονομάζεται Karush-Kuhn-Tucker (KKT) μητρώο. Όταν το μητρώο αυτό είναι μη ιδιάζων, τότε υπάρχει μοναδικό ζεύγος (x^*, λ^*) που ικανοποιεί της πρώτης τάξης αναγκαίες συνθήκες της (3.2). Το (3.3) είναι σύστημα γραμμικών εξισώσεων και έχουν αναπτυχθεί πολλές μέθοδοι για την αποτελεσματική επίλυσή του. Μερικές τις άμεσες μεθόδους είναι η συμμετρική παραγοντοποίηση, η range-space μέθοδος, η null-space μέθοδος. Κάποιες

από τις έμμεσες μεθόδους, δηλαδή επαναληπτικά, είναι οι Krylov μέθοδοι, οι επαναληπτικοί μετασχηματισμοί range-space και οι επαναληπτικοί μετασχηματισμοί null-space.

3.2 Συνθήκες βέλτιστης λύσης

Μπορούμε να ορίσουμε τη Λαγκρανζιανή για το πρόβλημα (3.1)

$$\mathcal{L}(x, \lambda) = \frac{1}{2}x^T Gx + x^T c - \sum_{i \in \mathcal{I} \cup \mathcal{E}} \lambda_i (a_i^T x - b_i).$$

Το ενεργό σύνολο $\mathcal{A}(x^*)$ αποτελείται από όλους τους δείκτες των περιορισμών για τους οποίους ισχύει η ισότητα στο x^* :

$$\mathcal{A}(x^*) = \{i \in \mathcal{E} \cup \mathcal{I} \mid a_i^T x^* = b_i\}.$$

Οι συνθήκες βέλτιστης λύσης Karush-Kuhn-Tucker conditions, (KKT conditions) (3.4)-(3.8) είναι

$$\nabla_x \mathcal{L}(x^*, \lambda^*) = 0 \tag{3.4}$$

$$c_i(x^*) = 0, \quad \text{for all } i \in \mathcal{E} \tag{3.5}$$

$$c_i(x^*) \geq 0, \quad \text{for all } i \in \mathcal{I} \tag{3.6}$$

$$\lambda_i^* \geq 0, \quad \text{for all } i \in \mathcal{I} \tag{3.7}$$

$$\lambda_i^* c_i(x^*) = 0, \quad \text{for all } i \in \mathcal{E} \cup \mathcal{I}. \tag{3.8}$$

Η σχέση (3.4) είναι γνωστή ως συνθήκη στασιμότητας (stationarity condition), οι σχέσεις (3.5) και (3.6) ως συνθήκες εφικτότητας (primal feasibility) με την υπόθεση ότι οι παράγωγοι των ενεργών περιορισμών ανισότητας και ισότητας είναι γραμμικώς ανεξάρτητοι στο σημείο βέλτιστης λύσης (LICQ), η (3.7) ως συνθήκη μη-αρνητικότητας (dual feasibility) και η (3.8) ως συνθήκη συμπληρωματικής χαλαρότητας (complementary slackness) και δηλώνει ότι είτε περιορισμό i είναι ενεργός είτε το αντίστοιχο $\lambda_i^* = 0$.

Αν γράψουμε τις συνθήκες KKT για το πρόβλημα, τότε κάθε λύση της (3.1) ικανοποιεί τις παρακάτω πρώτης τάξης συνθήκες βελτίστου, για κάποιους πολλαπλασιαστές Lagrange

$\lambda_i^*, i \in \mathcal{A}(x^*)$:

$$Gx^* + c - \sum_{i \in \mathcal{I} \cup \mathcal{E}} \lambda_i a_i = 0 \quad (3.9)$$

$$a_i^T x^* = b_i, \quad \text{for all } i \in \mathcal{A}(x^*) \quad (3.10)$$

$$a_i^T x^* \geq b_i, \quad \text{for all } i \in \mathcal{I} \setminus \mathcal{A}(x^*) \quad (3.11)$$

$$\lambda_i^* \geq 0, \quad \text{for all } i \in \mathcal{I} \cap \mathcal{A}(X^*). \quad (3.12)$$

3.3 Μέθοδος ενεργών περιορισμών

Η μέθοδος των ενεργών περιορισμών για τα τετραγωνικά προγράμματα παρουσιάζει τις παραλλαγές: primal, dual, primal-dual. Θα περιγράψουμε την πρώτη κατηγορία, η οποία μέσω επανάληψεων βρίσκει εφικτές λύσεις του (3.1) ενώ παράλληλα μειώνει την αντικειμενική συνάρτηση $q(x)$.

Η μέθοδος βρίσκει ένα βήμα από τη μία επανάληψη στην άλλη λύνοντας ένα τετραγωνικό υπό πρόβλημα όπου κάποιοι περιορισμοί ανισότητας και οι περιορισμοί ισότητας αντιμετωπίζονται ως ισότητες. Το υποσύνολο αυτό ονομάζεται σύνολο εργασίας (working set) και συμβολίζεται στην k επανάληψη x_k με \mathcal{W}_k . Μία σημαντική απαίτηση είναι τα a_i των περιορισμών να είναι γραμμικώς ανεξάρτητα.

Δοθέντας επανάληψης x_k και συνόλου εργασίας \mathcal{W}_k , αρχικά ελέγχουμε αν το x_k ελαχιστοποιεί τη τετραγωνική q στον υποχώρο που ορίζει το σύνολο εργασίας. Αν όχι, υπολογίζουμε βήμα p λύνοντας τετραγωνικό υποπρόγραμμα όπου οι περιορισμοί που αντιστοιχούν στο σύνολο \mathcal{W}_k θεωρούνται ισότητες και οι υπόλοιποι περιορισμοί δε λαμβάνονται προσωρινά υπόψιν. Για να εκφράσουμε το υποπρόβλημα ως προς το βήμα p , ορίζουμε

$$p = x - x_k, \quad g_k = Gx_k + c.$$

Με αντικατάσταση του x στην αντικειμενική συνάρτηση βρίσκουμε

$$q(x) = q(x_k + p) = \frac{1}{2} p^T G p + g_k^T p + \rho_k,$$

όπου $\rho_k = \frac{1}{2} x_k^T G x_k + c^T x_k$ είναι ανεξάρτητο του p . Έτσι μπορούμε να αγνοήσουμε το ρ_k χωρίς να αλλάξει η λύση του προβλήματος και έτσι το τετραγωνικό πρόγραμμα για να λυθεί

στην k επανάληψη είναι

$$\begin{aligned} \min_p \quad & \frac{1}{2} p^T G p + g_k^T p \\ \text{subject to} \quad & a_i^T p = 0, \quad i \in \mathcal{W}_k. \end{aligned} \quad (3.13)$$

Συμβολίζουμε τη λύση του παραπάνω υποπρογράμματος με p_k . Παρατηρούμε ότι για κάθε $i \in \mathcal{W}_k$, η τιμή του $a_i^T x$ δεν μεταβάλλεται καθώς κινούμαστε στο p_k , αφού $a_i^T (x_k + \alpha p_k) = a_i^T x_k = b_i$ για κάθε α . Αφού οι περιορισμοί στο \mathcal{W}_k ικανοποιούνται στο x_k , θα ικανοποιούνται και στο $x_k + \alpha p_k$, για κάθε α . Η λύση του (3.13) μπορεί να υπολογισθεί με κάποια από τις τεχνικές που αναφέρθηκαν στο κεφάλαιο 3.1.

Έστω ότι η βέλτιστη λύση p_k του (3.13) είναι μη μηδενική, τότε πρέπει να αποφασίσουμε πόσο μακριά να κινηθούμε σε αυτή τη διεύθυνση. Αν $x_k + p_k$ είναι εφικτό ως προς όλους τους περιορισμούς, τότε θέτουμε $x_{k+1} = x_k + p_k$. Διαφορετικά θέτουμε

$$x_{k+1} = x_k + \alpha_k p_k,$$

όπου το μήκος βήματος α_k επιλέγεται να είναι η μεγαλύτερη δυνατή τιμή του διαστήματος $[0, 1]$ για το οποίο ικανοποιούνται όλοι οι περιορισμοί. Μπορούμε να βρούμε έναν ορισμό για το α_k αν εξετάσουμε πως επηρεάζει τους περιορισμούς $i \notin \mathcal{W}_k$, καθώς οι περιορισμοί $i \in \mathcal{W}_k$ θα ικανοποιούνται ανεξαρτήτως της επιλογής του α_k . Αν $a_i^T p_k \geq 0$ για κάποιο $i \notin \mathcal{W}_k$, τότε για όλα τα $\alpha_k \geq 0$ έχουμε $a_i^T (x_k + \alpha_k p_k) \geq a_i^T x_k \geq b_i$. Έτσι, ο περιορισμός i θα ικανοποιείται για όλες τις μη αρνητικές τιμές του μήκους βήματος. Όταν $a_i^T p_k < 0$ για κάποιο $i \notin \mathcal{W}_k$, έχουμε $a_i^T (x_k + \alpha_k p_k) \geq b_i$ μόνο όταν

$$\alpha_k \leq \frac{b_i - a_i^T x_k}{a_i^T p_k}.$$

Για να μεγιστοποιήσουμε τη μείωση στη q , θέλουμε το α_k να είναι όσο το δυνατόν μεγαλύτερο στο $[0, 1]$ υπό τον περιορισμό να διατηρεί την εφικτότητα, έτσι έχουμε τον παρακάτω ορισμό

$$\alpha_k := \min \left(1, \min_{i \notin \mathcal{W}_k, a_i^T p_k < 0} \frac{b_i - a_i^T x_k}{a_i^T p_k} \right). \quad (3.14)$$

Ονομάζουμε τους περιορισμούς i για τους οποίους το ελάχιστο στη (3.14) ικανοποιείται περιορισμούς παρεμπόδισης (blocking constraints). Αν $\alpha_k = 1$ και κανένας νέος περιορισμός είναι ενεργός στο $x_k + \alpha_k p_k$, τότε δεν υπάρχουν καθόλου περιορισμοί παρεμπόδισης. Παρατηρούμε ότι είναι εφικτό το α_k να είναι μηδέν, γιατί θα μπορούσαμε να είχαμε $a_i^T p_k < 0$ για κάποιον περιορισμό i που είναι ενεργός στο x_k αλλά να μην ανήκει στο σύνολο εργασίας

\mathcal{W}_k .

Αν $\alpha_k < 1$, τότε το βήμα κατά μήκος του p_k παρεμποδίζεται από κάποιον περιορισμό που δεν ανήκει στο \mathcal{W}_k και δημιουργείται νέο σύνολο εργασίας \mathcal{W}_{k+1} λαμβάνοντας υπόψιν τον περιορισμό αυτόν.

Συνεχίζουμε τις επαναλήψεις σύμφωνα με την παραπάνω λογική, προσθέτοντας περιορισμούς στο σύνολο εργασίας μέχρι να βρούμε σημείο \hat{x} που ελαχιστοποιεί την τετραγωνική αντικειμενική συνάρτηση ως προς το τρέχων σύνολο εργασίας $\hat{\mathcal{W}}$. Εύκολα αναγνωρίζουμε το σημείο αυτό, καθώς το υποπρόβλημα (3.13) έχει λύση $p = 0$. Εφόσον το $p = 0$ ικανοποιεί τις συνθήκες βέλτιστης λύσης (3.3) για το (3.13) ισχύει

$$\sum_{i \in \hat{\mathcal{W}}} a_i \hat{\lambda}_i = g = G\hat{x} + c, \quad (3.15)$$

για κάποιους πολλαπλασιαστές Lagrange $\hat{\lambda}_i, i \in \hat{\mathcal{W}}$. Τα \hat{x} και $\hat{\lambda}$ ικανοποιούν την πρώτη συνθήκη KKT (3.9), αν ορίσουμε τους πολλαπλασιαστές των ανισοτήτων που δεν ανήκουν στο σύνολο εργασίας να είναι μηδέν. Εξαιτίας του ελέγχου στο μήκος βήματος, το \hat{x} είναι εφικτό ως προς όλους τους περιορισμούς, έτσι οι σχέσεις (3.10) και (3.11) ικανοποιούνται.

Αν οι πολλαπλασιαστές που αντιστοιχούν στους περιορισμούς ανισότητας του συνόλου εργασίας, δηλαδή $i \in \hat{\mathcal{W}} \cap \mathcal{I}$, είναι μη αρνητικοί, τότε η (3.12) ικανοποιείται και έτσι το \hat{x} ικανοποιεί τις συνθήκες βελτίστου και συνεπώς ικανοποιεί το αρχικό πρόβλημα (3.1).

Αν από την άλλη, κάποιος από τους περιορισμούς $\hat{\lambda}_j, j \in \hat{\mathcal{W}} \cap \mathcal{I}$, είναι αρνητικός τότε η συνθήκη (3.12) δεν ικανοποιείται και η αντικειμενική συνάρτηση μπορεί να μειωθεί αν αφαιρέσουμε τον j περιορισμό από το σύνολο εργασίας και λύσουμε ένα νέο υποπρόβλημα για νέο βήμα. Με τον τρόπο αυτόν, δημιουργούμε μία διεύθυνση p στη νέα επανάληψη που είναι εφικτή ως προς τον περιορισμό που αφαιρέθηκε.

Στο σημείο αυτό έχουμε περιγράψει τα βασικά σημεία της μεθόδου ενεργών περιορισμών. Παρακάτω παρουσιάζεται η μέθοδος σε μορφή αλγόριθμου για προβλήματα κυρτού τετραγωνικού προγραμματισμού, όπου υποθέτουμε ότι η αντικειμενική συνάρτηση q είναι φραγμένη στο εφικτό σύνολο των περιορισμών της (3.1).

Όπως βλέπουμε από τον αλγόριθμο, απαιτείται ένα αρχικό εφικτό σημείο για να ξεκινήσει. Διάφορες τεχνικές έχουν αναπτυχθεί για τον υπολογισμό αυτού. Επιγραμματικά θα αναφέρουμε μία από αυτές. Αν \tilde{x} είναι μία αρχική εκτίμηση του διανύσματος x , η οποία μπορεί

Require: Δώσε αρχικό εφικτό σημείο x_0 ;

Require: Το αρχικό σύνολο \mathcal{W}_0 να είναι υποσύνολο των ενεργών περιορισμών στο x_0 ;

```

for  $k = 0, 1, 2, \dots$  do
  Λύσε (3.13) και βρες  $p_k$ ;
  if  $p_k = 0$  then
    Βρες  $\hat{\lambda}_i$  που ικανοποιούν (3.15) με  $\hat{\mathcal{W}} = \mathcal{W}_k$ ;
    if  $\hat{\lambda}_i \geq 0$  για όλα τα  $i \in \mathcal{W}_k \cap \mathcal{I}$  then
      return  $x^* = x_k$ ;
    else
       $j \leftarrow \arg \min_{j \in \mathcal{W}_k \cap \mathcal{I}} \hat{\lambda}_j$ ;
       $x_{k+1} \leftarrow x_k$ ;  $\mathcal{W}_{k+1} \leftarrow \mathcal{W}_k \setminus j$ ;
    end if
  else
    υπολόγισε  $\alpha_k$  από (3.14);
     $x_{k+1} \leftarrow x_k + \alpha_k p_k$ ;
    if υπάρχουν περιορισμοί παρεμπόδισης then
      βρες  $\mathcal{W}_{k+1}$  προσθέτοντας τον περιορισμό παρεμπόδισης του  $\mathcal{W}_k$ ;
    else
       $\mathcal{W}_{k+1} \leftarrow \mathcal{W}_k$ ;
    end if
  end if
end for

```

Σχήμα 3.1: Αλγόριθμος ενεργών περιορισμών

να μην είναι εφικτή, τότε μπορούμε να ορίσουμε το ακόλουθο εφικτό γραμμικό πρόγραμμα

$$\begin{aligned}
 & \min_{(x,z)} e^T z \\
 & \text{subject to } a_i^T x + \gamma_i z_i = b_i, \quad i \in \mathcal{E}, \\
 & \quad a_i^T x + \gamma_i z \geq b_i, \quad i \in \mathcal{I}, \\
 & \quad z \geq 0,
 \end{aligned}$$

όπου $e = \begin{pmatrix} 1 & \dots & 1 \end{pmatrix}^T$, $\gamma_i = -\text{sign}(a_i^T \tilde{x} - b_i)$ για $i \in \mathcal{E}$, και $\gamma_i = 1$ για $i \in \mathcal{I}$. Ένα εφικτό αρχικό σημείο για το πρόβλημα είναι

$$x = \tilde{x}, \quad z_i = |a_i^T \tilde{x} - b_i| \quad (i \in \mathcal{E}), \quad z_i = \max(b_i - a_i^T \tilde{x}, 0) \quad (i \in \mathcal{I}).$$

Αποδεικνύεται ότι αν \tilde{x} είναι εφικτό σημείο για το αρχικό πρόβλημα (3.1) τότε το $(\tilde{x}, 0)$ είναι βέλτιστο για το εφικτό υποπρόβλημα. Γενικά, αν το αρχικό τετραγωνικό πρόγραμμα έχει εφικτά σημεία, τότε βέλτιστη αντικειμενική συνάρτηση του παραπάνω υποπροβλήματος

είναι μηδέν, και κάθε λύση του υποπροβλήματος δίνει εφικτό σημείο στο αρχικό πρόβλημα. Τέλος, το αρχικό σύνολο εργασίας \mathcal{W}_0 μπορεί να βρεθεί παίρνοντας γραμμικώς ανεξάρτητα υποσύνολα των ενεργών περιορισμών στη λύση του υποπροβλήματος.

3.4 Επίλυση

Θα παρουσιάσουμε ένα παράδειγμα τετραγωνικού προγραμματισμού με χρήση του λογισμικού MATLAB και τη χρήση της βιβλιοθήκης βελτιστοποίησης (Optimization Toolbox) αυτού. Το MATLAB δέχεται ένα τετραγωνικό πρόγραμμα στην παρακάτω μορφή

$$\begin{aligned} \min_x \quad & \frac{1}{2}x^T Hx + f^T x \\ \text{subject to} \quad & Ax \leq b \\ & A_{eq}x = b_{eq} \\ & lb \leq x \leq ub. \end{aligned} \tag{3.16}$$

Στην παραπάνω σχέση, H είναι το μητρώο των τετραγωνικών όρων στην αντικειμενική συνάρτηση, f είναι το διάνυσμα των γραμμικών συντελεστών, A είναι το μητρώο των συντελεστών των ανισοτήτων και b το αντίστοιχο διάνυσμα. Ακόμη, οι σχέσεις ισότητας περιγράφονται από το μητρώο A_{eq} και το διάνυσμα b_{eq} . Τέλος, lb και ub είναι τα διανύσματα του κάτω και άνω ορίου του διανύσματος x .

Η λύση τετραγωνικών προγραμμάτων γίνεται με τη συνάρτηση **quadprog**. Η σύνταξή της είναι

```
[x,fval,exitflag,output,lambda]=quadprog(H,f,A,b,Aeq,beq,lb,ub,x0,options)
```

με απαραίτητα τα δύο πρώτα ορίσματα της συνάρτησης. Επιστρέφει τη βέλτιστη λύση x , την τιμή της αντικειμενικής συνάρτησης $fval$, τη συνθήκη εξόδου $exitflag$, τη δομή (structure) $output$, σχετικά με το τέλος του προγράμματος και τους πολλαπλασιαστές Lagrange $lambda$. Επίσης, από τη δομή $options$, μπορούμε να ρυθμίσουμε διάφορες παραμέτρους της διαδικασίας επίλυσης, όπως για παράδειγμα να επιλέξουμε αλγόριθμο επίλυσης (interior-point-convex, trust-region-reflective, ο πρώτος είναι για κυρτά τετραγωνικά και ο δεύτερος αλγόριθμος μπορεί να χρησιμοποιηθεί όταν έχουμε μόνο όρια στις μεταβλητές ή και περιορισμούς ισότητας), να θέσουμε ανοχές, δηλαδή το κατά πόσο θα ικανοποιείται η αντικειμενική συνάρτηση, οι περιορισμοί, να δώσουμε μία αρχική λύση, και άλλα πολλά. Ο ενδιαφερόμενος παραπέμπεται στη βοήθεια του MATLAB.

3.4.1 Παράδειγμα

Το παράδειγμα προέρχεται από τη βοήθεια του MATLAB και έχει περιορισμούς ανισότητας, καθώς και όρια στις μεταβλητές. Η αντικειμενική συνάρτηση που θέλουμε να ελαχιστοποιήσουμε είναι

$$f(x) = \frac{1}{2}x_1^2 + x_2^2 - x_1x_2 - 2x_1 - 6x_2,$$

οι σχέσεις ανισότητας που περιγράφουν το πρόβλημα είναι

$$\begin{aligned}x_1 + x_2 &\leq 2 \\ -x_1 + 2x_2 &\leq 2 \\ 2x_1 + x_2 &\leq 3,\end{aligned}$$

και τέλος έχουμε τα όρια

$$0 \leq x_1, \quad 0 \leq x_2.$$

Το παραπάνω πρόβλημα εύκολα γράφεται σε μητρωνική μορφή

$$f(x) = \frac{1}{2}x^T Hx + f^T x,$$

όπου

$$H = \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix}, \quad f = \begin{pmatrix} -2 \\ -6 \end{pmatrix}, \quad x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}.$$

Επίσης, το μητρώο των περιορισμών A και το αντίστοιχο διάνυσμα b είναι

$$A = \begin{pmatrix} 1 & 1 \\ -1 & 2 \\ 2 & 1 \end{pmatrix}, \quad b = \begin{pmatrix} 2 \\ 2 \\ 3 \end{pmatrix}.$$

Λύνοντας το πρόβλημα, βρίσκουμε ότι η βέλτιστη λύση είναι $x^* = (0.667 \quad 1.333)^T$ και η αντίστοιχη τιμή της αντικειμενικής $f(x^*) = -8.2222$.

Στο σχήμα 3.2 παρουσιάζεται η γραφική λύση του προγράμματος. Σημειώνεται με ροζ αστερίσκο επάνω στην καμπύλη το ελάχιστο του προβλήματος. Στη συνέχεια παρατίθεται ο κώδικας του MATLAB που χρησιμοποιήθηκε για να λυθεί το πρόβλημα.

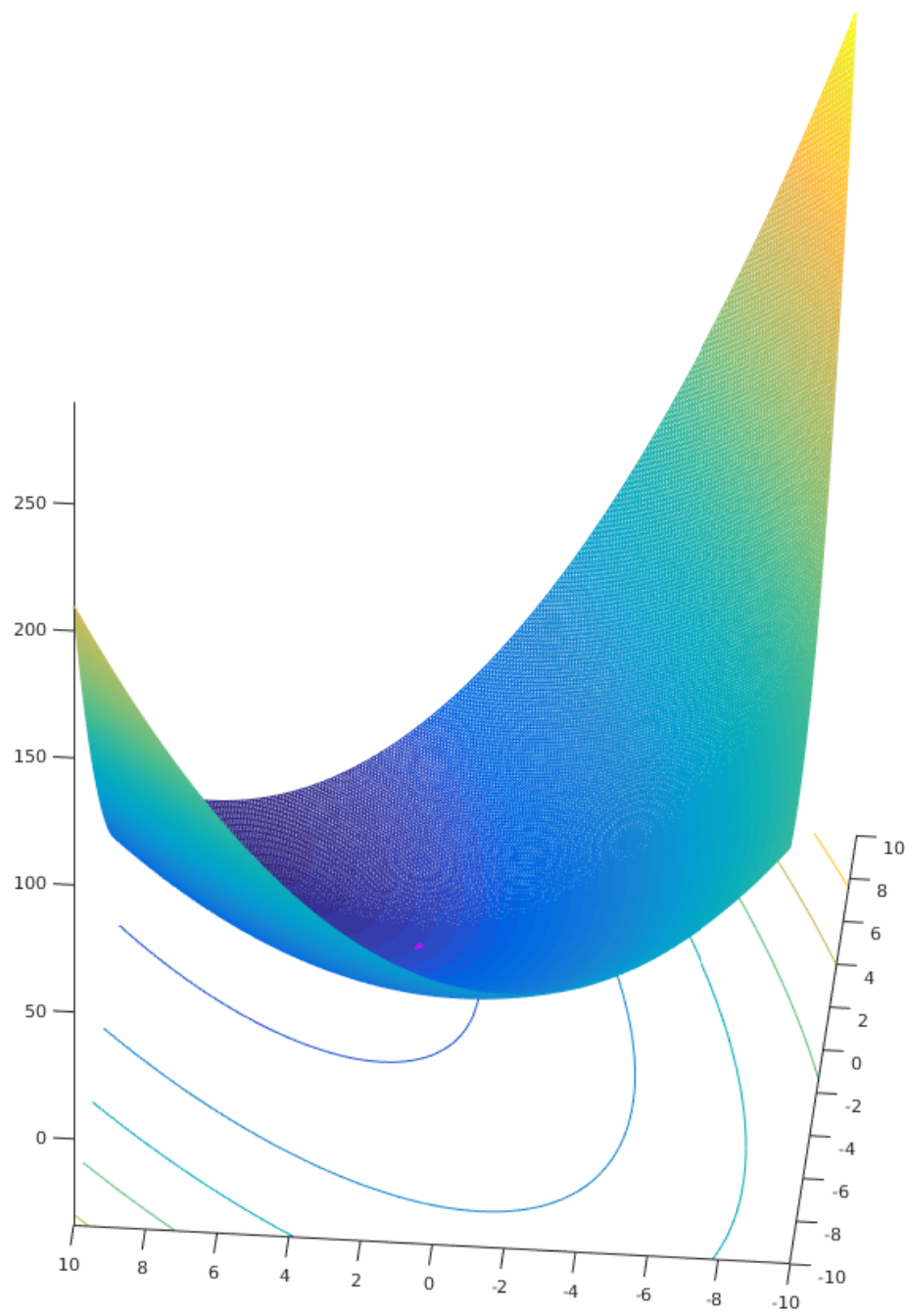


Figure 3.2: Λύση παραδείγματος του τετραγωνικού προγραμματισμού

```
1 function [x, fval] = qp()
```

```
2     H = [1 -1; -1 2];
3     f = [-2; -6];
4     A = [1 1; -1 2; 2 1];
5     b = [2; 2; 3];
6     lb = zeros(2,1);
7     options = optimoptions('quadprog',...
8 'Algorithm','interior-point-convex','Display','off');
9     [x,fval,exitflag,output,lambda] = ...
10 quadprog(H,f,A,b,[],[],lb,[],[],options);
11     x1 = [-10:0.1:10];
12     x2 = [-10:0.1:10];
13     [X, Y] = meshgrid(x1, x2);
14     Z = 0.5*X.^2 + Y.^2 - X.*Y -2*X -6*Y;
15     hold on;
16     meshc(X, Y, Z);
17     scatter3(x(1),x(2),fval, '*', 'm');
18 end
```

Κεφάλαιο 4

Κυρτά Σύνολα

4.1 Αφινικά και κυρτά σύνολα

4.1.1 Γραμμή και τμήμα γραμμής

Όλα τα σημεία που περιγράφονται από τη σχέση

$$x = \theta x_1 + (1 - \theta)x_2,$$

με $x_1, x_2 \in \mathbf{R}^n$ και $\theta \in \mathbf{R}$, αποτελούν μία **γραμμή** που περνάει από τα x_1, x_2 . Για $\theta = 0$ έχουμε $x = x_2$ ενώ $\theta = 1$ αντιστοιχεί σε $x = x_1$. Τιμές της παραμέτρου θ στο μεταξύ 0 και 1 δίνουν **τμήμα γραμμής** ανάμεσα στα x_1 και x_2 .

4.1.2 Αφινικό σύνολο

Ένα σύνολο $C \subseteq \mathbf{R}^n$ ονομάζεται **αφινικό** αν η γραμμή μεταξύ δύο οποιοδήποτε διαφορετικών σημείων που ανήκουν στο C , βρίσκεται στο C . Για παράδειγμα αν για οποιοδήποτε $x_1, x_2 \in C$ και $\theta \in \mathbf{R}$ τότε

$$x = \theta x_1 + (1 - \theta)x_2 \in C.$$

Με άλλα λόγια, το σύνολο C περιέχει το γραμμικό συνδυασμό δύο σημείων, ο οποίος βρίσκεται στο C δεδομένου ότι οι συντελεστές στο γραμμικό συνδυασμό αθροίζονται στη μονάδα. Η ιδέα αυτή μπορεί να γενικευτεί για παραπάνω από δύο σημεία. Το σημείο της μορφής $\theta_1 x_1 + \dots + \theta_k x_k$, όπου $\theta_1 + \dots + \theta_k = 1$, αναφέρεται ως **αφινικός συνδυασμός** των



Σχήμα 4.1: Κυρτά Σύνολα

σημείων $x_1 + \dots + x_k$. Ένα χαρακτηριστικό παράδειγμα αφινικού συνόλου είναι το σύστημα γραμμικών εξισώσεων της μορφής $Ax = b$. Επίσης, ισχύει ότι κάθε αφινικό σύνολο μπορεί να εκφραστεί ως ένα σύστημα γραμμικών εξισώσεων. Το σύνολο όλων των αφινικών συνδυασμών των σημείων του συνόλου $C \subseteq \mathbf{R}^n$ ονομάζεται **αφινική θήκη** του C και συμβολίζεται με $\text{aff } C$.

4.1.3 Κυρτό Σύνολο

Ένα σύνολο $C \subseteq \mathbf{R}^n$ ονομάζεται **κυρτό** αν το τμήμα γραμμής μεταξύ δύο οποιοδήποτε σημείων που ανήκουν στο C , βρίσκεται στο C . Για παράδειγμα αν για οποιοδήποτε $x_1, x_2 \in C$ και $0 \leq \theta \leq 1$ τότε

$$x = \theta x_1 + (1 - \theta)x_2 \in C.$$

Ένα παράδειγμα κυρτού συνόλου παρουσιάζεται στο 4.1α', ενώ αντιθέτως το σύνολο της 4.1β' δεν είναι κυρτό καθώς μέρος του τμήματος της γραμμής μεταξύ των σημείων x και y δεν περιλαμβάνεται στο σύνολο. Το σημείο της μορφής $\theta_1 x_1 + \dots + \theta_k x_k$, όπου $\theta_1 + \dots + \theta_k = 1$ και $\theta_i \geq 0$ για $i = 1, \dots, k$, αναφέρεται ως **κυρτός συνδυασμός** των σημείων $x_1 + \dots + x_k$. Το σύνολο όλων των κυρτών συνδυασμών των σημείων του συνόλου $C \subseteq \mathbf{R}^n$ ονομάζεται **κυρτή θήκη** του C και συμβολίζεται με $\text{conv } C$.

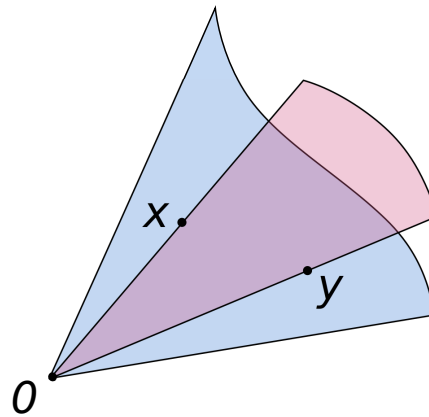
4.1.4 Κυρτός κώνος

Το σύνολο C ονομάζεται **κώνος**, αν για κάθε $x \in C$ και $\theta \geq 0$ ισχύει ότι $\theta x \in C$. Το σύνολο C είναι **κυρτός κώνος** αν είναι κυρτό και κώνος, που σημαίνει ότι για οποιαδήποτε

$x_1, x_2 \in C$ και $\theta_1, \theta_2 \geq 0$ ισχύει

$$x = \theta_1 x_1 + \theta_2 x_2 \in C.$$

Στο σχήμα 4.2 παρουσιάζεται ένας κυρτός κώνος (ανοιχτό μπλε). Μέσα σε αυτό, ο ροζ κυρτός κώνος αποτελείται από όλα τα σημεία $\alpha x + \beta y$ με $\theta_1, \theta_2 \geq 0$.



Σχήμα 4.2: Κυρτός Κώνος

4.2 Σημαντικά Παραδείγματα

4.2.1 Υπερεπίπεδα

Ένα υπερεπίπεδο είναι ένα σύνολο της μορφής

$$\{x \mid a^T x = b\},$$

όπου $a \in \mathbf{R}^n, a \neq 0$ και $b \in \mathbf{R}$. Αναλυτικά είναι το σύνολο λύσεων των γραμμικών εξισώσεων των συνιστωσών x . Γεωμετρικά το υπερεπίπεδο μπορεί να ερμηνευτεί ως το σύνολο των σημείων με σταθερό εσωτερικό γινόμενο ως προς το διάνυσμα a , ή ως το υπερεπίπεδο με κανονικό διάνυσμα a , και η σταθερά b καθορίζει την απόσταση του υπερεπιπέδου από την αρχή. Τα υπερεπίπεδα είναι αφινικά και κυρτά.

4.2.2 Ημιχώρος

Ένα υπερεπίπεδο χωρίζει το \mathbf{R}^n σε δύο **ημιχώρους**. Ένας (κλείστος) ημιχώρος είναι το σύνολο της μορφής

$$\{x \mid a^T x \leq b\},$$

όπου $a \neq 0$, π.χ., το σύνολο λύσης μίας ανισότητας. Οι ημιχώροι είναι κυρτοί αλλά όχι αφινικοί.

4.2.3 (Ευκλείδειες) Μπάλες

Μία (Ευκλείδεια) μπάλα στο \mathbf{R}^n έχει τη μορφή

$$B(x_c, r) = \{x \mid \|x - x_c\|_2 \leq r\} = \{x \mid (x - x_c)^T (x - x_c) \leq r^2\},$$

όπου $r > 0$, και $\|\cdot\|_2$ συμβολίζει την Ευκλείδεια νόρμα. Το διάνυσμα x_c είναι το κέντρο της μπάλας και το βαθμωτό μέγεθος r η ακτίνα της. Η μπάλα $B(x_c, r)$ αποτελείται από όλα τα σημεία που απέχουν το πολύ απόσταση ίση με r από το κέντρο x_c και είναι κυρτό σύνολο.

4.2.4 Ελλειψοειδές

Μία συγγενική κατηγορία κυρτού συνόλου είναι το **ελλειψοειδές**, που περιγράφεται

$$\mathcal{E} = \{x \mid (x - x_c)^T P^{-1} (x - x_c) \leq 1\},$$

όπου $P = P^T > 0$, π.χ., το μητρώο P είναι συμμετρικό και θετικά ορισμένο. Το διάνυσμα $x_c \in \mathbf{R}^n$ είναι το κέντρο του ελλειψοειδούς. Το μητρώο P καθορίζει κατά πόσο το ελλειψοειδές επεκτείνεται σε κάθε διεύθυνση από το x_c . Το μήκος των ημιαξόνων του \mathcal{E} δίνονται από τα $\sqrt{\lambda_i}$, όπου λ_i είναι οι ιδιοτιμές του P .

4.2.5 Νόρμα

Μία συνάρτηση $\|\cdot\|$ που ικανοποιεί

- $\|x\| \geq 0$ και $\|x\| = 0$ αν και μόνο αν $x = 0$,
- $\|tx\| = |t|\|x\|$ για κάθε $t \in \mathbf{R}$,

$$\bullet \|x + y\| \leq \|x\| + \|y\|,$$

είναι **νόρμα**. Ο συμβολισμός $\|\cdot\|$ αποτελεί γενικό συμβολισμό, συνήθως χρησιμοποιούμε κάποιο συμβολισμό για να προσδιορίσουμε τον τύπο της νόρμας.

4.2.6 Νορμική μπάλα και νορμικός κώνος

Θεωρώντας οποιαδήποτε νόρμα $\|\cdot\|$ στο \mathbf{R}^n και από τις ιδιότητες της νόρμας, η **νορμική μπάλα** ακτίνας r και κέντρου x_c που περιγράφεται από $\{x \mid \|x - x_c\| \leq r\}$, είναι κυρτή. Ο **νορμικός κώνος** εφοδιασμένος με τη νόρμα $\|\cdot\|$ είναι το σύνολο

$$C = \{(x, t) \mid \|x\| \leq t\} \subseteq \mathbf{R}^{n+1},$$

και είναι κυρτός κώνος.

4.2.7 Πολύεδρα

Το **πολύεδρο** ορίζεται ως το σύνολο λύσης πεπερασμένων γραμμικών ισοτήτων και ανισοτήτων:

$$\mathcal{P} = \{x \mid a_j^T x \leq b_j, j = 1, \dots, m, c_j^T x = d_j, j = 1, \dots, p\}.$$

Το πολύεδρο είναι ουσιαστικά η τομή πεπερασμένων ημιχώρων και υπερεπιπέδων. Αφινικά σύνολα (π.χ. υποχώροι, υπερεπίπεδα, γραμμές), τμήματα γραμμής και ημιχώροι είναι όλα πολύεδρα. Επίσης, εύκολα αποδεικνύεται ότι τα πολύεδρα είναι κυρτά σύνολα. Ένα φραγμένο πολύεδρο πολλές φορές καλείται **πολύτοπο**. Πολλές φορές είναι βολική η συμπαγής μορφή του πολυέδρου

$$\mathcal{P} = \{x \mid Ax \leq b, Cx = d\}.$$

Simplexes. Τα **simplexes** είναι μία σημαντική κατηγορία πολυέδρων. Έστω τα $k+1$ σημεία $v_0, \dots, v_k \in \mathbf{R}^n$ είναι **αφινικά ανεξάρτητα**, που σημαίνει ότι $v_1 - v_0, \dots, v_k - v_0$ είναι γραμμικώς ανεξάρτητα. Το simplex τότε καθορίζεται από τη σχέση

$$\mathcal{C} = \text{conv} \{v_0, \dots, v_k\} = \{\theta_0 v_0 + \dots + \theta_k v_k \mid \theta \geq 0, \mathbf{1}^T \theta = 1\},$$

όπου το $\mathbf{1}$ συμβολίζει το διάνυσμα με όλα τα στοιχεία μονάδα. Η αφινική διάσταση αυτού του συμπλέγματος είναι k , έτσι πολλές φορές λέγεται k -διάστασης simplex στο \mathbf{R}^n .

Κυρτή θήκη και πολύεδρα. Η κυρτή θήκη του πεπερασμένου συνόλου $\{v_1, \dots, v_k\}$ είναι

$$\text{conv} \{v_1, \dots, v_k\} = \{\theta_1 v_1 + \dots + \theta_k v_k \mid \theta \geq 0, \mathbf{1}^T \theta = 1\}.$$

Το σύνολο αυτό είναι πολύεδρο και φραγμένο.

4.2.8 Θετικά ημιορισμένος κώνος

Με το συμβολισμό \mathbf{S}^n δηλώνουμε το σύνολο των συμμετρικών $n \times n$ μητρώων,

$$\mathbf{S}^n = \{X \in \mathbf{R}^{n \times n} \mid X = X^T\},$$

που είναι διανυσματικός χώρος διάστασης $n(n+1)/2$. Με το συμβολισμό \mathbf{S}_+^n δηλώνουμε το σύνολο των συμμετρικών θετικά ημιορισμένων μητρώων

$$\mathbf{S}_+^n = \{X \in \mathbf{S}^n \mid X \geq 0\},$$

που είναι κυρτός κώνος και με το συμβολισμό \mathbf{S}_{++}^n δηλώνουμε το σύνολο των συμμετρικών θετικά ορισμένων μητρώων

$$\mathbf{S}_{++}^n = \{X \in \mathbf{S}^n \mid X > 0\}.$$

4.3 Πράξεις που διατηρούν την κυρτότητα

4.3.1 Τομή

Η κυρτότητα διατηρείται υπό την τομή συνόλων, αν S_1 και S_2 είναι κυρτά σύνολα, τότε η τομή $S_1 \cap S_2$ είναι κυρτή. Η ιδιότητα αυτή επεκτείνεται για τομή απείρων συνόλων. Ένα απλό παράδειγμα είναι το πολύεδρο που είναι η τομή ημιχώρων και υπερεπιπέδων, που είναι κυρτά, και επομένως είναι κυρτό.

4.3.2 Αφινικές συναρτήσεις

Μία συνάρτηση $f : \mathbf{R}^n \rightarrow \mathbf{R}^m$ είναι **αφινική** αν είναι το άθροισμα γραμμικών συναρτήσεων και σταθεράς. Για παράδειγμα συνάρτηση της μορφής $f(x) = Ax + b$, όπου $A \in \mathbf{R}^{n \times m}$ και $b \in \mathbf{R}^m$. Έστω $S \subseteq \mathbf{R}^n$ είναι κυρτό και $f : \mathbf{R}^n \rightarrow \mathbf{R}^m$ είναι αφινική συνάρτηση. Τότε η

εικόνα του S υπό τον περιορισμό της f ,

$$f(S) = \{f(x) \mid x \in S\},$$

είναι κυρτή. Αντίστοιχα, αν $f : \mathbf{R}^k \rightarrow \mathbf{R}^n$ είναι αφινική συνάρτηση, η αντίστροφη εικόνα του S υπό τον περιορισμό της f ,

$$f^{-1}(S) = \{x \mid f(x) \in S\},$$

είναι κυρτή. Κάποια απλά παραδείγματα αφινικών συναρτήσεων είναι η **κλιμάκωση** και η **μετατόπιση**. Αν $S \subseteq \mathbf{R}^n$ είναι κυρτό, $\alpha \in \mathbf{R}$ και $a \in \mathbf{R}^n$, τότε τα σύνολα αS και $S + a$ είναι κυρτά, όπου

$$\alpha S = \{\alpha x \mid x \in S\}, \quad S + a = \{x + a \mid x \in S\}.$$

Επίσης, το σύνολο λύσης **γραμμικών μητρώων ανισότητας** (linear matrix inequality, LMI), $\{x \mid A(x) \leq B\}$ είναι κυρτό, όπου $B, A_i \in \mathbf{S}^m$.

4.3.3 Γραμμική-κλασματική και προοπτική συνάρτηση

Συνάρτηση προοπτικής. Ορίζουμε τη συνάρτηση προοπτικής $P : \mathbf{R}^n \rightarrow \mathbf{R}^n$, με πεδίο ορισμού $\text{dom } P = \mathbf{R}^n \times \mathbf{R}_{++}$, ως, όπου \mathbf{R}_{++} δηλώνει το σύνολο των πραγματικών θετικών αριθμών, $P(z, t) = z/t$. Η συνάρτηση αυτή κλιμακώνει ή κανονικοποιεί διανύσματα ώστε η τελευταία συνιστώσα να γίνει μονάδα και στη συνέχεια την αποβάλλει. Το σημαντικό είναι ότι αν $C \subseteq \text{dom } P$ είναι κυρτό, τότε και η εικόνα

$$P(C) = \{P(x) \mid x \in C\},$$

υπό τη συνάρτηση προοπτικής είναι κυρτή. Επίσης, η αντίστροφη εικόνα ενός κυρτού συνόλου υπό συνάρτηση προοπτικής είναι κυρτή, δηλαδή αν $C \subseteq \mathbf{R}^n$ είναι κυρτό, τότε

$$P^{-1}(C) = \{(x, t) \in \mathbf{R}^{n+1} \mid x/t \in C, t > 0\},$$

είναι κυρτή.

Γραμμική-κλασματική συνάρτηση. Μία γραμμική-κλασματική συνάρτηση σχηματίζεται από τη σύνθεση συναρτήσεων προοπτικής με αφινικές συναρτήσεις. Έστω $g : \mathbf{R}^n \rightarrow$

\mathbf{R}^{m+1} αφινική, π.χ.,

$$g(x) = \begin{pmatrix} A \\ c^T \end{pmatrix} x + \begin{pmatrix} b \\ d \end{pmatrix},$$

όπου $A \in \mathbf{R}^{m \times n}$, $b \in \mathbf{R}^m$, $c \in \mathbf{R}^n$, και $d \in \mathbf{R}$. Η συνάρτηση $f : \mathbf{R}^n \rightarrow \mathbf{R}^m$, δίνεται από $f = P \circ g$, η σύμφωνα με το παραπάνω παράδειγμα

$$f(x) = (Ax + b)/(c^T x + d), \quad \text{dom } f = \{x \mid c^T x + d > 0\},$$

καλείται **γραμμική-κλασματική** (ή **προβολική**) συνάρτηση. Αν $c = 0$ και $d > 0$, το πεδίο ορισμού της f είναι το \mathbf{R}^n και η f είναι αφινική συνάρτηση. Έτσι μπορούμε να θεωρήσουμε της αφινικές και τις γραμμικές συναρτήσεις σαν ειδικές περιπτώσεις γραμμικών-κλασματικών συναρτήσεων. Όπως και στην προηγούμενη περίπτωση, η εικόνα και η αντίστροφη εικόνα κυρτών συνόλων υπό γραμμικών-κλασματικών συναρτήσεων είναι κυρτά σύνολα.

4.4 Γενικευμένες ανισότητες

4.4.1 Γνήσιοι κώνοι και γενικευμένες ανισότητες

Ένας κώνος $K \subseteq \mathbf{R}^n$ καλείται **γνήσιος κώνος** αν ικανοποιεί τα παρακάτω:

- K είναι κυρτό.
- K είναι κλειστό.
- K είναι στερεό, που σημαίνει ότι έχει μη-κενά εσωτερικά σημεία.
- K είναι σημειακό (pointed), που σημαίνει ότι δεν περιέχει καμία ευθεία (ή ισοδύναμα, $x \in K, -x \in K \Rightarrow x = 0$).

Ένας γνήσιος κώνος K μπορεί να χρησιμοποιηθεί για να ορίσουμε τη **γενικευμένη ανισότητα**, που είναι μερικώς διατεταγμένο στο \mathbf{R}^n και έχει πολλές ιδιότητες της τυπικής διατεταγμένης του \mathbf{R} . Συνδέουμε με το γνήσιο κώνο K το μερικώς διατεταγμένο σύνολο στο \mathbf{R}^n που ορίζεται από

$$x \leq_K y \Leftrightarrow y - x \in K.$$

Επίσης, γράφουμε $x \geq_K y$ για $y \leq_K x$. Αντίστοιχα, ορίζουμε την **αυστηρή γενικευμένη ανισότητα** και συνδέουμε το γνήσιο κώνο με το αυστηρό μερικώς διατεταγμένο σύνολο ως

$$x \leq_K y \Leftrightarrow y - x \in \text{int } K,$$

και γράφουμε $x >_K y$ για $y <_K x$. Κάποια τυπικά παραδείγματα είναι π.χ. όταν $K = \mathbf{R}_+^n$, τότε η γενικευμένη ανισότητα \leq_K αντιστοιχεί σε ανισότητα ως προς την κάθε συνιστώσα μεταξύ διανυσμάτων: $x \leq_K y$ σημαίνει ότι $x_i \leq y_i$, για $i = 1, \dots, n$. Ακόμη, ο θετικά ημιορισμένος κώνος \mathbf{S}_+^n είναι ένας γνήσιος κώνος στο \mathbf{S}^n . Η αντίστοιχη γενικευμένη ανισότητα \leq_K είναι η ανισότητα μητρώων $X \leq_K Y$ που σημαίνει ότι $Y - X$ είναι θετικά ημιορισμένο. Επίσης, οι γενικευμένες ανισότητες \leq_K έχουν όμοιες ιδιότητες με τη γνωστή ανισότητα \leq του \mathbf{R} .

4.4.2 Ελάχιστο και ελαχιστοτικό στοιχείο

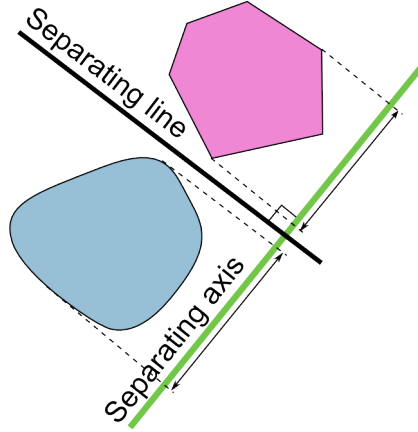
Ενώ πολλές από τις ιδιότητες των γενικευμένων ανισοτήτων βρίσκονται σε αναλογία με αυτές του \mathbf{R} , αυτό δεν συμβαίνει για όλες. Η πιο εμφανής διαφορά είναι ότι η \leq στο \mathbf{R} είναι γραμμικώς διατεταγμένη, δηλαδή κάθε δύο σημεία είναι συγκρίσιμα. Η ιδιότητα αυτή δεν ισχύει για τις γενικευμένες ανισότητες και αυτό ως συνέπεια οι έννοιες του ελάχιστου και μέγιστου να χρήζουν ορισμού. Λέμε ότι $x \in S$ είναι το **ελάχιστο** στοιχείο του S , σε σχέση με τη γενικευμένη ανισότητα \leq_K , αν για κάθε $y \in S$ ισχύει $x \leq_K y$. Ορίζουμε το **μέγιστο** στοιχείο του S , σε σχέση με τη γενικευμένη ανισότητα με τον αντίστοιχο προφανή τρόπο. Αν ένα σύνολο έχει ελάχιστο (μέγιστο) στοιχείο, τότε είναι μοναδικό. Μία παρόμοια έννοια είναι το **ελαχιστοτικό στοιχείο (minimal element)**. Λέμε ότι $x \in S$ είναι το ελαχιστοτικό στοιχείο του S , σε σχέση με τη γενικευμένη ανισότητα \leq_K , αν $y \in S$ ισχύει $y \leq_K x$ μόνο όταν $y = x$. Αντίστοιχα για το **μεγιστοτικό στοιχείο (maximal element)**.

4.5 Υπερεπίπεδο διαχωριστικό και στήριξης

4.5.1 Θεώρημα διαχωριστικού υπερεπιπέδου

Έστω C και D μη-κενά ξένα κυρτά σύνολα, δηλαδή $C \cap D = \emptyset$. Τότε υπάρχει $a \neq 0$ και b τέτοια ώστε $a^T x \leq b$ για όλα τα $x \in C$ και $a^T x \geq b$ για όλα τα $x \in D$. Με άλλα λόγια, η αφινική συνάρτηση $a^T x - b$ είναι μη-θετική στο C και μη-αρνητική στο D . Το υπερεπίπεδο $\{x \mid a^T x = b\}$ ονομάζεται **διαχωριστικό υπερεπίπεδο** για τα σύνολα C και D , ή λέμε ότι

διαχωρίζει το σύνολο C και D . Σχηματικά το θεώρημα παρουσιάζεται στο γράφημα 4.3. Όπως ορίστηκε το παραπάνω θεώρημα, αν ισχύει μόνο η ανισότητα, και στις δύο σχέσεις,



Σχήμα 4.3: Θεώρημα διαχωριστικού υπερεπιπέδου

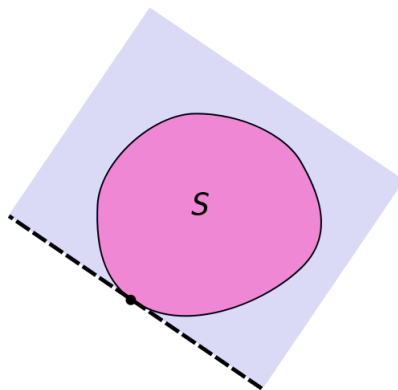
τότε έχουμε το **αυστηρό διαχωριστικό υπερεπίπεδο**.

4.5.2 Θεώρημα υπερεπιπέδου στήριξης

Έστω $C \subseteq \mathbf{R}^n$, και x_0 είναι σημείο του συνόρου

$$x_0 \in \mathbf{bd} C.$$

Αν $a \neq 0$ ικανοποιεί $a^T x \leq a^T x_0$ για όλα τα $x \in C$, τότε το υπερεπίπεδο $\{x \mid a^T x = a^T x_0\}$ ονομάζεται **υπερεπίπεδο στήριξης** του C στο x_0 . Ισοδύναμα μπορούμε πούμε ότι το σημείο x_0 και το σύνολο C διαχωρίζονται από το υπερεπίπεδο $\{x \mid a^T x = a^T x_0\}$. Γεωμετρικά, όπως φαίνεται στο 4.4, το υπερεπίπεδο είναι η εφαπτομένη του C στο x_0 και ο υποχώρος $\{x \mid a^T x \leq a^T x_0\}$ περιέχει το C . Ένα βασικό αποτέλεσμα, που ονομάζεται **θεώρημα υπερεπιπέδου στήριξης**, δηλώνει ότι για κάθε μη-κενό κυρτό σύνολο C , και για οποιοδήποτε $x_0 \in \mathbf{bd} C$, τότε υπάρχει υπερεπίπεδο στήριξης του C στο x_0 .



Σχήμα 4.4: Θεώρημα υπερεπιπέδου στήριξης

4.6 Δυϊκός κώνος και γενικευμένες ανισότητες

4.6.1 Δυϊκός κώνος

Έστω K κώνος. Το σύνολο

$$K^* = \{y \mid x^T y \geq 0 \text{ για όλα τα } x \in K\}$$

ονομάζεται **δυϊκός κώνος** του K . Όπως υποδηλώνει το όνομα, K^* είναι κώνος και είναι πάντα κυρτός, ακόμα και αν ο κώνος K δεν είναι. Για παραδείγματα ο κώνος \mathbf{R}_+^n είναι ο ίδιος δυϊκός

$$x^T y \geq 0 \text{ για όλα τα } x \geq 0 \Leftrightarrow y \geq 0$$

και λέμε τους κώνους αυτούς **αυτοδυϊκούς**.

4.6.2 Δυϊκές γενικευμένες ανισότητες

Αν ο κυρτός κώνος K είναι γνήσιος, άρα επάγεται μία γενικευμένη ανισότητα \leq_K . Τότε ο δυϊκός κώνος είναι επίσης γνήσιος, και επάγει μία γενικευμένη ανισότητα. Αναφερόμαστε στη γενικευμένη ανισότητα \leq_{K^*} ως τη **δυϊκή** της γενικευμένης ανισότητας \leq_K . Μερικές σημαντικές ιδιότητες σχετικά με τη γενικευμένη ανισότητα και την δυϊκή της είναι

- $x \leq_K y$ αν και μόνο αν $\lambda^T x \leq \lambda^T y$ για κάθε $\lambda \geq_{K^*} 0$.
- $x <_K y$ αν και μόνο αν $\lambda^T x < \lambda^T y$ για κάθε $\lambda \geq_{K^*} 0, \lambda \neq 0$.

4.6.3 Ελάχιστα και ελαχιστοτικά στοιχεία μέσω δυϊκών ανισοτήτων

Χρησιμοποιώντας τις δυϊκές γενικευμένες ανισότητες μπορούμε να χαρακτηρίσουμε ελάχιστα και ελαχιστοτικά στοιχεία (κυρτών ή όχι) συνόλων $S \subseteq \mathbf{R}^m$ ως προς τις γενικευμένες ανισότητες που επάγονται από το γνήσιο κώνο K .

Δυϊκός χαρακτηρισμός ελάχιστου στοιχείου. Το x είναι **ελάχιστο** στοιχείο του S , ως προς τη γενικευμένη ανισότητα \leq_K , αν και μόνο αν για κάθε $\lambda >_{K^*} 0$, το x είναι το μοναδικό που ελαχιστοποιεί τη σχέση $\lambda^T z$ για $z \in S$. Γεωμετρικά, αυτό σημαίνει ότι για κάθε $\lambda >_{K^*} 0$, το υπερεπίπεδο

$$\{z \mid \lambda^T(z - x) = 0\}$$

είναι ένα αυστηρό υπερεπίπεδο στήριξης του S στο x . Αυστηρό υπερεπίπεδο στήριξης σημαίνει ότι το υπερεπίπεδο τέμνει το S μόνο στο x . Η κυρτότητα του συνόλου S δεν είναι απαραίτητη σύμφωνα με το παραπάνω.

Δυϊκός χαρακτηρισμός ελαχιστοτικού στοιχείου. Αν το x ελαχιστοποιεί $\lambda^T z$ για $z \in S$ και $\lambda >_{K^*} 0$ τότε το x είναι **ελαχιστοτικό**. Το αντίστροφο δεν ισχύει στη γενική περίπτωση, ένα σημείο x μπορεί να είναι ελαχιστοτικό του S , αλλά να μην ελαχιστοποιεί τη σχέση $\lambda^T z$ για $z \in S$ και οποιοδήποτε λ . Αν το σύνολο S είναι κυρτό τότε για κάθε ελαχιστοτικό στοιχείο x υπάρχει μη μηδενικό $\lambda \leq_{K^*} 0$ τέτοιο ώστε το x να ελαχιστοποιεί τη $\lambda^T z$ για $z \in S$.

Κεφάλαιο 5

Μέθοδοι Εσωτερικού Σημείου

5.1 Ιστορικά στοιχεία

Ο γραμμικός προγραμματισμός σύμφωνα με το βιβλίο [15] είναι ένας από τους πιο πετυχημένους τομείς της βελτιστοποίησης. Από τη διατύπωσή του τη δεκαετία του 1930 μέχρι την ανάπτυξη του αλγορίθμου simplex τη δεκαετία του 1940 από τον Dantzig, γενιές εργαζομένων στα οικονομικά, στη μηχανική και άλλους τομείς εκπαιδεύτηκαν για να λύνουν προβλήματα γραμμικού προγραμματισμού. Ακόμα και όταν τα μοντέλα ήταν μη-γραμμικά, γραμμικές μέθοδοι χρησιμοποιούνταν λόγω της μεγάλης ανάπτυξης που είχαν.

Η δημοσίευση το 1948 του Karmarkar [6] ήταν το πιο σημαντικό γεγονός στο γραμμικό προγραμματισμό μετά την ανακάλυψη της μεθόδου simplex. Ο λόγος ήταν ότι ο αλγόριθμος του Karmarkar υποσχόταν πολυωνυμική πολυπλοκότητα, σε σύγκριση με τη λογαριθμική πολυπλοκότητα της μεθόδου simplex, και εξαιρετική πρακτική επίδοση σε μεγάλου μεγέθους προβλήματα. Οι ισχυρισμοί αυτοί δεν επιβεβαιώθηκαν, αλλά η δημοσίευση έδωσε μία επανάσταση στην έρευνα του γραμμικού προγραμματισμού που οδήγησε σε έναν νέο τομέα αλγορίθμων που ονομάζονται μέθοδοι εσωτερικού σημείου.

5.2 Μέθοδος affine scaling

Σε αυτή την ενότητα θα περιγράψουμε έναν απλό αλγόριθμο, που ονομάζεται affine scaling, για την επίλυση προβλημάτων γραμμικών προγραμματισμού, βασιζόμενοι στο βιβλίο [17].

Η μέθοδος αυτή είναι μία μέθοδος εσωτερικού σημείου και επιτελεί το ρόλο εισαγωγής στον αλγόριθμο του Karmarkar. Γενικά, οι μέθοδοι εσωτερικού σημείου διαφέρουν από τη μέθοδο simplex σε ένα βασικό σημείο: η μέθοδος εσωτερικού σημείου ξεκινάει στο εσωτερικού του εφικτού συνόλου και κινείται σε αυτό για την εύρεση βέλτιστου σημείου. Αντίθετα, η μέθοδος simplex κινείται στις κορυφές του εφικτού συνόλου.

Έστω το γραμμικό πρόβλημα

$$\begin{aligned} &\text{minimize} && c^T x \\ &\text{subject to} && Ax = b \\ &&& x \geq 0. \end{aligned}$$

Έστω ότι έχουμε το εφικτό σημείο $x^{(0)}$ που είναι αυστηρώς εσωτερικό (δηλαδή $x > 0$). Θέλουμε να βρούμε ένα νέο σημείο $x^{(1)}$ ψάχνοντας στη διεύθυνση του $d^{(0)}$ που μειώνει την αντικειμενική συνάρτηση. Δηλαδή

$$x^{(1)} = x^{(0)} + a_0 d^{(0)},$$

με a_0 το μέγεθος του βήματος. Προκειμένου το $x^{(1)}$ να βρίσκεται εντός του εφικτού συνόλου, είναι απαραίτητο το διάνυσμα $d^{(0)}$ να βρίσκεται στο μηδενικό χώρο του A , με άλλα λόγια θέλουμε να ισχύει $Ax^{(1)} = b$. Επειδή, το $x^{(0)}$ είναι εφικτό σημείο από τις παραπάνω εξισώσεις προκύπτει

$$A(x^{(1)} - x^{(0)}) = a_0 A d^{(0)} = 0.$$

Για να επιλέξουμε $d^{(0)}$ που να βρίσκεται στο μηδενικό χώρο του A και να κινείται στην μείωση της αντικειμενικής, δηλαδή κοντά στο $-c$, παίρνουμε την ορθογώνια προβολή του $-c$ στο μηδενικό χώρο του A . Αυτή η προβολή περιγράφεται από τον πίνακα

$$P = I_n - A^T(AA^T)^{-1}A.$$

Έτσι θέτουμε το $d^{(0)}$ να είναι στη διεύθυνση της ορθογώνιας προβολής του $-c$ στο μηδενικό χώρο του A

$$d^{(0)} = -Pc,$$

και έτσι, βρίσκουμε ένα καινούργιο εφικτό σημείο $x^{(1)}$ από τη σχέση

$$x^{(1)} = x^{(0)} - a_0 Pc.$$

Παρατηρώντας τις σχέσεις καταλήγουμε στο συμπέρασμα ότι αν το αρχικό εφικτό σημείο

βρίσκεται στο “κέντρο” του εφικτού συνόλου, τότε μπορούμε να κινηθούμε με μεγαλύτερο βήμα και αυτό σημαίνει μεγαλύτερη μείωση στην τιμή της αντικειμενικής και άρα ταχύτερη σύγκλιση.

Στην περίπτωση που το αρχικό σημείο δεν βρίσκεται στο κέντρο του συνόλου, μπορούμε με έναν αφηνικό μετασχηματισμό να το μετατρέψουμε σε κεντρικό. Για να μετατρέψουμε το $x^{(0)}$ στο κεντρικό σημείο e , χρησιμοποιούμε τον αφηνικό μετασχηματισμό (affine scaling transformation)

$$e = D_0^{-1}x^{(0)},$$

όπου $D_0 = \text{diag}[x_1^{(0)}, \dots, x_n^{(0)}]$. Με την υπόθεση που κάναμε ότι το $x^{(0)}$ είναι αυστηρώς εσωτερικό, ο πίνακας D_0 είναι αντιστρέψιμος. Προφανώς, στην πράξη δεν επιδιώκουμε να βρούμε το κεντρικό σημείο αλλά κάποιο σημείο αρκετά κοντά σε αυτό με μικρή σχετικά απόκλιση. Βασικό βέβαια είναι το αρχικό σημείο να είναι αυστηρώς εσωτερικό.

Με την παραπάνω διαδικασία, μετατρέψαμε το αρχικό εφικτό διάνυσμα $x^{(0)}$ με το μετασχηματισμό D_0^{-1} και ουσιαστικά αλλάξαμε βάση. Έτσι στη νέα βάση το πρόβλημα της ενότητας μετασχηματίζεται στο αντίστοιχο γραμμικό

$$\begin{aligned} &\text{minimize} && \bar{c}_0^T \bar{x} \\ &\text{subject to} && \bar{A}_0 \bar{x} = b \\ &&& \bar{x} \geq 0, \end{aligned}$$

όπου

$$\bar{c}_0 = D_0 c, \quad \bar{A}_0 = A D_0.$$

Εφαρμόζοντας τη διαδικασία που περιγράψαμε παραπάνω στο νέο σύστημα συντεταγμένων, υπολογίζουμε την επόμενη τιμή από τη σχέση

$$\bar{x}^{(1)} = \bar{x}^{(0)} - a_0 \bar{P}_0 \bar{c}_0,$$

με $\bar{x}^{(0)} = D_0^{-1}x^{(0)}$. Το νέο σημείο στις αρχικές συντεταγμένες δίνεται με το μετασχηματισμό $x^{(1)} = D_0 \bar{x}^{(1)}$. Επαναλαμβάνοντας τη διαδικασία για μία ακολουθία σημείων $\{x^{(k)}\}$, όπου

$$x^{(k+1)} = x^{(k)} + a_k d^{(k)},$$

με

$$\begin{aligned} D_k &= \text{diag} [x_1^{(k)}, \dots, x_n^{(k)}] \\ \bar{A}_k &= A D_k \\ \bar{P}_k &= I_n - \bar{A}_k^T (\bar{A}_k \bar{A}_k^T)^{-1} \bar{A}_k \\ d^{(k)} &= -D_k \bar{P}_k D_k c. \end{aligned}$$

Σε κάθε βήμα του αλγορίθμου πρέπει να εξασφαλίζουμε ότι το $x^{(k)}$ είναι αυστηρώς εσωτερικό. Η συνθήκη των περιορισμών ικανοποιείται λόγω του τρόπου που επιλέξαμε το $d^{(k)}$. Όμως, πρέπει ακόμα να εξασφαλίσουμε ότι $x_i^{(k)} > 0$ για $i = 1, \dots, n$, και αυτό γίνεται με την κατάλληλη επιλογή του βήματος, a_k .

Αναφέραμε ήδη ότι επιδιώκουμε μεγάλο βήμα, για αυτό το λόγο πραγματοποιούμε τον αφινικό μετασχηματισμό στο αρχικό σημείο για να το μετατρέψουμε σε κεντρικό. Όμως πρέπει να επιλέξουμε τέτοιο βήμα ούτως ώστε το νέο σημείο να είναι θετικό. Για να το επιτύχουμε αυτό, ορίζουμε

$$r_k = \min_{\{i: d_i^{(k)} < 0\}} - \frac{x_i^{(k)}}{d_i^{(k)}},$$

που αντιπροσωπεύει το μεγαλύτερο βήμα a_k έτσι ώστε οι συνιστώσες του $x^{(k+1)}$ να είναι μη-αρνητικές και υπολογίζουμε το βήμα από τη σχέση $a_k = \alpha r_k$ με $\alpha \in (0, 1)$. Τυπικές τιμές του συντελεστή α είναι 0.9 ή 0.99.

Σε αντίθεση με άλλες μεθόδους, η μέθοδος *affine scaling* μέθοδος δε θα φτάσει σε βέλτιστη λύση σε πεπερασμένο αριθμό βημάτων στη γενική περίπτωση. Επομένως, είναι απαραίτητο κάποιο κριτήριο τερματισμού και συχνά στη βιβλιογραφία συναντάται η σχέση

$$\frac{|cx^{(k+1)} - cx^{(k)}|}{\max 1, |cx^{(k)}|} < \epsilon,$$

με το ϵ να έχει επιλεχθεί ανάλογα με την ακρίβεια που επιθυμούμε. Πολλές φορές στην πράξη δε γνωρίζουμε το αρχικό εφικτό σημείο. Ακόμα η εύρεση του σημείου αυτού δεν είναι προφανής. Για το λόγο αυτό έχουν αναπτυχθεί κάποιες μεθοδολογίες για τον υπολογισμό αυτού και την έναρξη της μεθόδου. Ο ενδιαφερόμενος αναγνώστης παραπέμπεται στο βιβλίο [17] για τη διαδικασία εύρεσης αρχικής εφικτής λύσης.

5.3 Αλγόριθμος path-following

Στο σημείο αυτό θα αναφέρουμε τον αλγόριθμο path-following, γνωστός και ως μέθοδος φράγματος (barrier method) που είναι μέλος της οικογένειας των μεθόδων εσωτερικού σημείου για να λύσουμε προβλήματα κυρτής βελτιστοποίησης που περιλαμβάνουν περιορισμούς ανισότητας και είναι της μορφής,

$$\begin{aligned} & \text{minimize} && f_0(x) \\ & \text{subject to} && f_i(x) \leq 0, \quad i = 1, \dots, m, \\ & && Ax = b, \end{aligned} \tag{5.1}$$

όπου $f_0, \dots, f_m : \mathbb{R}^n \rightarrow \mathbb{R}$ είναι κυρτές και δύο φορές συνεχώς παραγωγίσιμες, και $A \in \mathbb{R}^{p \times n}$ με $\text{rank } A = p < n$. Θεωρούμε ότι το πρόβλημα έχει λύση, δηλαδή η βέλτιστη λύση x^* υπάρχει και θα συμβολίζουμε τη βέλτιστη τιμή $f_0(x^*)$ ως p^* .

Επίσης, υποθέτουμε ότι το πρόβλημα είναι αυστηρώς εφικτό, δηλαδή υπάρχουν $\lambda^* \in \mathbb{R}^m, v^* \in \mathbb{R}^p$ τέτοια ώστε μαζί με το x^* να ικανοποιούνται οι συνθήκες βέλτιστης λύσης Karush–Kuhn–Tucker (KKT)

$$\begin{aligned} Ax^* &= b, & f_i(x^*) &\leq 0, & i &= 1, \dots, m \\ \lambda^* &\geq 0 \\ \nabla f_0(x^*) + \sum_{i=1}^m \lambda_i^* \nabla f_i(x^*) + A^T v^* &= 0 \\ \lambda_i^* f_i(x^*) &= 0, & i &= 1, \dots, m. \end{aligned} \tag{5.2}$$

Οι μέθοδοι εσωτερικού σημείου λύνουν το πρόβλημα (5.1) ή το (5.2), εφαρμόζοντας τη μέθοδο Newton σε μία ακολουθία προβλημάτων με περιορισμούς ισότητας, ή αντίστοιχα σε μία ακολουθία παραλλαγών των συνθηκών KKT. Για την ανάπτυξη της θεωρίας και τη διατύπωση του αλγορίθμου θα βασιστούμε στο βιβλίο [13].

5.3.1 Λογαριθμική συνάρτηση φράγματος και κεντρική διαδρομή

Στόχος είναι να διατυπώσουμε το πρόβλημα με τους περιορισμούς ανισότητας (5.1), σε ισοδύναμο πρόβλημα με περιορισμούς ισότητας στο οποίο μπορούμε να εφαρμόσουμε τη μέθοδο Newton. Ενσωματώνοντας στην αντικειμενική συνάρτηση τους περιορισμούς ανισότητας

προκύπτει

$$\begin{aligned} & \text{minimize} && f_0(x) + \sum_{i=1}^m I_{-}(f_i(x)) \\ & \text{subject to} && Ax = b, \end{aligned} \quad (5.3)$$

όπου $I_{-} : \mathbb{R} \rightarrow \mathbb{R}$ είναι η συνάρτηση

$$I_{-}(u) = \begin{cases} 0 & u \leq 0 \\ \infty & u > 0. \end{cases}$$

Η βασική ιδέα της μεθόδου φράγματος είναι να προσεγγίσει τη συνάρτηση I_{-} με τη συνάρτηση

$$\hat{I}_{-}(u) = (-1/t) \log(-u), \quad \text{dom } \hat{I}_{-} = -\mathbb{R}_{++},$$

όπου $t > 0$ είναι παράμετρος που θέτει την ακρίβεια της προσέγγισης. Όπως η I_{-} , η συνάρτηση \hat{I}_{-} είναι κυρτή και δεν μειώνεται και κατά σύμβαση τείνει στο ∞ όταν $u > 0$. Σε αντίθεση με την I_{-} , η \hat{I}_{-} είναι παραγωγίσιμη και κλειστή. Αντικαθιστώντας τη \hat{I}_{-} στη θέση της I_{-} στην (5.3) προκύπτει η προσέγγιση

$$\begin{aligned} & \text{minimize} && f_0(x) + \sum_{i=1}^m -(1/t) \log(-f_i(x)) \\ & \text{subject to} && Ax = b. \end{aligned} \quad (5.4)$$

Η αντικειμενική στην παραπάνω σχέση είναι κυρτή, και αυξάνεται με το u και διαφορίσιμη. Υποθέτοντας ότι υπάρχει συνθήκη για να είναι κλειστή, τότε μπορούμε να λύσουμε με τη μέθοδο Newton.

Η συνάρτηση

$$\phi(x) = - \sum_{i=1}^m \log(-f_i(x)),$$

με $\text{dom } \phi = \{x \in \mathbb{R}^n \mid f_i(x) < 0, i = 1, \dots, m\}$, ονομάζεται λογαριθμική συνάρτηση φράγματος για το πρόβλημα (5.1).

5.3.2 Κεντρική διαδρομή

Έστω $t > 0$, ορίζουμε $x^*(t)$ ως λύση της

$$\begin{aligned} & \text{minimize} && t f_0(x) + \phi(x) \\ & \text{subject to} && Ax = b. \end{aligned}$$

Η *κεντρική διαδρομή* (central path) που σχετίζεται με το πρόβλημα (5.1) ορίζεται ως το σύνολο των σημείων $x^*(t)$, $t > 0$, τα οποία ονομάζουμε *κεντρικά σημεία* (central points). Σημεία στην κεντρική διαδρομή χαρακτηρίζονται από τις ακόλουθες αναγκαίες και ικανές συνθήκες: $x^*(t)$ είναι αυστηρώς εφικτό, δηλαδή ισχύει

$$Ax^*(t) = b, \quad f_i(x^*(t)) < 0, \quad i = 1, \dots, m,$$

και υπάρχει διάνυσμα $\hat{v} \in \mathbb{R}^p$ τέτοιο ώστε να ισχύει

$$0 = t \nabla f_0(x^*(t)) + \nabla \phi(x^*(t)) + A^T \hat{v}$$

ή ισοδύναμα με αντικατάσταση της κλίσης $\nabla \phi(x^*(t))$

$$0 = t \nabla f_0(x^*(t)) + \sum_{i=1}^m \frac{1}{-f_i(x^*(t))} \nabla f_i(x^*(t)) + A^T \hat{v}. \quad (5.5)$$

Δυϊκά σημεία από την κεντρική διαδρομή. Μία σημαντική ιδιότητα της κεντρικής διαδρομής είναι ότι κάθε κεντρικό σημείο παράγει ένα δυϊκό εφικτό σημείο, και έτσι ένα κάτω όριο στην βέλτιστη τιμή p^* . Πιο συγκεκριμένα αν ορίσουμε

$$\lambda_i^*(t) = -\frac{1}{t f_i(x^*(t))}, \quad i = 1, \dots, m, \quad v^*(t) = \hat{v}/t, \quad (5.6)$$

τότε το ζεύγος $\lambda^*(t), v^*(t)$ είναι δυϊκό εφικτό. Αρχικά, $\lambda^*(t) > 0$ επειδή $f_i(x^*(t)) < 0$. Αντικαθιστώντας στην σχέση (5.5) την παραπάνω προκύπτει

$$t \nabla f_0(x^*(t)) + \sum_{i=1}^m \lambda_i^*(t) \nabla f_i(x^*(t)) + v^*(t) = 0,$$

και βλέπουμε ότι το $x^*(t)$ ελαχιστοποιεί τη Λαγκρανζιανή

$$L(x, \lambda, v) = f_0(x) + \sum_{i=1}^m \lambda_i \nabla f_i(x) + v^T (Ax - b),$$

για $\lambda = \lambda^*(t)$ και $v = v^*(t)$, που σημαίνει ότι τα $\lambda^*(t), v^*(t)$ είναι δυϊκό εφικτό ζεύγος. Συνεπώς, η δυϊκή συνάρτηση είναι πεπερασμένη

$$\begin{aligned} g(\lambda^*(t), v^*(t)) &= f_0(x^*(t)) + \sum_{i=1}^m \lambda_i^*(t) \nabla f_i(x^*(t)) + v^*(t)^T (Ax^*(t) - b) \\ &= f_0(x^*(t)) - m/t, \end{aligned}$$

και συνεπώς το δυϊκό κενό μεταξύ του $x^*(t)$ και του δυϊκού ζεύγους $\lambda^*(t), v^*(t)$ είναι απλά m/t . Ως επακόλουθο αυτού έχουμε

$$f_0(x^*(t)) - p^* \leq m/t,$$

που επιβεβαιώνει τη διαισθητική ιδέα ότι το $x^*(t)$ συγκλίνει σε βέλτιστο σημείο καθώς $t \rightarrow \infty$.

Ερμηνεία μέσω των συνθηκών KKT. Μπορούμε να ερμηνεύσουμε τις συνθήκες κεντρικής διαδρομής (5.5) ως ένα συνεχή μετασχηματισμό των συνθηκών KKT (5.2). Ένα σημείο x ισούται με το $x^*(t)$ αν και μόνο αν υπάρχουν λ, v τέτοια ώστε

$$\begin{aligned} Ax &= b, \quad f_i(x) \leq 0, \quad i = 1, \dots, m \\ \lambda &\geq 0 \\ \nabla f_0(x) + \sum_{i=1}^m \lambda_i \nabla f_i(x) + A^T v &= 0 \\ -\lambda_i f_i(x) &= 1/t, \quad i = 1, \dots, m. \end{aligned}$$

Η μόνη διαφορά των παραπάνω και των συνθηκών (5.2) είναι ότι η συμπληρωματική συνθήκη $-\lambda_i f_i(x) = 0$ αντικαταστήθηκε από τη συνθήκη $-\lambda_i f_i(x) = 1/t$.

Ερμηνεία μέσω δυνάμεων πεδίου. Συσχετίζουμε με κάθε περιορισμό τη δύναμη

$$F_i(x) = -\nabla(-\log(-f_i(x))) = \frac{1}{f_i(x)} \nabla f_i(x)$$

που επιδρά στο “σωματίδιο” στη θέση x . Η συνολική δύναμη που δημιουργείται από τους περιορισμούς είναι η λογαριθμική συνάρτηση φράγματος ϕ . Καθώς το σωματίδιο κινείται προς το όριο του εφικτού συνόλου, απωθείται έντονα από τις δυνάμεις που δημιουργούνται από τους περιορισμούς.

Έστω τώρα μία άλλη δύναμη που επιδρά στο σωματίδιο

$$F_0(x) = -t\nabla f_0(x),$$

στο σωματίδιο στη θέση x . Αυτή η αντικειμενική δύναμη πεδίου επιδρά έλκοντας το σωματίδιο στην αντίθετη κατεύθυνση της κλίσης. Η παράμετρος t κλιμακώνει την αντικειμενική σχετικά με τους περιορισμούς δύναμης.

Το κεντρικό σημείο $x^*(t)$ είναι το σημείο που οι περιορισμοί δύναμης ισορροπούν με τη δύναμη της αντικειμενικής στο σωματίδιο. Καθώς η παράμετρος t μεγαλώνει, το σωματίδιο έλκεται εντονότερα προς το βέλτιστο σημείο, αλλά πάντα μένει εντός του εφικτού συνόλου λόγω της επίδρασης των περιορισμών.

5.3.3 Διατύπωση του αλγορίθμου *path-following*

Η μέθοδος αυτή βασίζεται στην επίλυση μίας ακολουθίας χωρίς περιορισμούς (ή με γραμμικούς περιορισμούς) προβλημάτων ελαχιστοποίησης, χρησιμοποιώντας το τελευταίο σημείο που υπολογίστηκε ως αρχική λύση για το επόμενο πρόβλημα ελαχιστοποίησης χωρίς περιορισμούς. Με άλλα λόγια, υπολογίζουμε το $x^*(t)$ για μία ακολουθία με αυξανόμενη την τιμή του t , μέχρι $t \geq m/\epsilon$, που μας εξασφαλίζει ότι θα έχουμε ένα βέλτιστο υποπρόβλημα του αρχικού προβλήματος με απόκλιση ϵ . Η μέθοδος αρχικά προτάθηκε από τον Fiacco και McCormick τη δεκαετία του 1960. Σήμερα η μέθοδος αυτή είναι γνωστή ως *μέθοδος φράγματος* (barrier method) ή *path-following μέθοδος* και εφαρμόζεται ευρέως για την επίλυση προβλημάτων. Ο αλγόριθμος της μεθόδου παρουσιάζεται στο 5.1.

Require: αυστηρώς εφικτό σημείο $x, t := t^{(0)} > 0, \mu > 1$, ανοχή $\epsilon > 0$.

repeat

1. Κεντρικό βήμα. Υπολόγισε $x^*(t)$ ελαχιστοποιώντας $tf_0 + \phi$ που υπόκειται $Ax = b$, ξεκινώντας από το x .

2. Ανανέωσε. $x := x^*(t)$.

until 3. Κριτήριο τερματισμού $m/t < \epsilon$

4. Αύξησε. $t := \mu t$.

Σχήμα 5.1: Αλγόριθμος *path-following*

Σε κάθε επανάληψη (εκτός της πρώτης) υπολογίζουμε το κεντρικό σημείο $x^*(t)$ ξεκινώντας από το προηγούμενο κεντρικό σημείο που υπολογίστηκε, και στη συνέχεια αυξάνουμε το t κατά συντελεστή $\mu > 1$. Ο αλγόριθμος μπορεί επίσης να επιστρέψει το δυϊκό σημείο

$$\lambda = \lambda^*(t), v = v^*(t).$$

Ονομάζουμε κάθε εκτέλεση του βήματος 1, κεντρικό βήμα ή εξωτερική επανάληψη, καθώς τότε υπολογίζουμε κεντρικά σημεία, και του πρώτου κεντρικού βήματος ως αρχικό κεντρικό βήμα. Μπορούμε να χρησιμοποιήσουμε διάφορες μεθόδους για την επίλυση του προβλήματος ελαχιστοποίησης με γραμμικού περιορισμούς που προκύπτει στην επανάληψη του κεντρικού βήματος, όπως για παράδειγμα τη μέθοδο Newton και αναφερόμαστε στην επαναλήψεις κατά τη μέθοδο Newton ως εσωτερικές επαναλήψεις. Σε κάθε εσωτερική επανάληψη έχουμε ένα βασικό εφικτό σημείο, αλλά δυϊκό εφικτό σημείο έχουμε στο τέλος του κεντρικού βήματος.

Ο υπολογισμός του $x^*(t)$ με μεγάλη ακρίβεια δεν είναι απαραίτητο καθώς η κεντρική διαδρομή θα συγκλίνει σε βέλτιστο σημείο ακόμα και όταν το $x^*(t)$ δεν έχει υπολογισθεί με μεγάλη ακρίβεια. Σε αυτήν την περίπτωση όμως τα δυϊκά σημεία $\lambda^*(t), v^*(t)$ δεν είναι δυϊκά εφικτά. Αυτό μπορεί να διορθωθεί προσθέτοντας ένα διορθωτικό όρο στις σχέσεις (5.6) που υπολογίζονται. Από την άλλη, το κόστος υπολογισμού ελαχιστοποίησης με μεγάλη ακρίβεια της $tf_0 + \phi$ είναι οριακά μεγαλύτερο από το κόστος υπολογισμού του ελάχιστου της συνάρτησης με μία ικανοποιητική ακρίβεια, μερικά Newton βήματα μακριά.

Η επιλογή της παραμέτρου μ επηρεάζει τις εσωτερικές και εξωτερικές επαναλήψεις του αλγορίθμου. Αν το μ είναι μικρό τότε σε κάθε εξωτερική επανάληψη το t αυξάνει κατά ένα μικρό ποσοστό. Ως αποτέλεσμα αυτού, το αρχικό σημείο για κάθε φορά που εφαρμόζουμε τη μέθοδο Newton είναι πολύ καλό και ο αριθμός των επαναλήψεων που απαιτείται για να συγκλίνει η μέθοδος Newton είναι μικρός. Έτσι για μικρό μ αναμένουμε μικρό αριθμό βημάτων Newton στις εξωτερικές επαναλήψεις, αλλά μεγάλο αριθμό εξωτερικών επαναλήψεων καθώς κάθε εξωτερική επανάληψη μειώνει την απόκλιση της απόστασης του σημείου από το βέλτιστο κατά ένα μικρό ποσοστό. Στην περίπτωση αυτή πολλές φορές ο αλγόριθμος συναντάται με την ονομασία *short-step path-following* ή SPF.

Αντίθετα, όταν το μ είναι μεγάλο έχουμε την ακριβώς αντίθετη συμπεριφορά. Μετά από κάθε εξωτερική επανάληψη το t αυξάνεται κατά μεγάλο ποσοστό, έτσι η συγκεκριμένη επανάληψη πιθανόν δεν είναι αρκετά καλή προσέγγιση της επόμενης επανάληψης. Έτσι αναμένουμε πολλές εσωτερικές επαναλήψεις. Αυτή η επιθετική επιλογή της παραμέτρου t οδηγεί σε λιγότερες εξωτερικές επαναλήψεις, αλλά σε περισσότερες εσωτερικές. Στην περίπτωση αυτή πολλές φορές ο αλγόριθμος συναντάται με την ονομασία *long-step path-following* ή LPF.

Μία άλλη επιλογή που έχουμε να κάνουμε είναι η αρχική τιμή του t . Αν το $t^{(0)}$ είναι πολύ μεγάλο, η πρώτη εξωτερική επανάληψη θα χρειαστεί πολλές επαναλήψεις. Αν το $t^{(0)}$ είναι πολύ μικρό, ο αλγόριθμος θα χρειαστεί πολλές εξωτερικές επαναλήψεις και πιθανόν πολλές

εσωτερικές στο πρώτο κεντρικό βήμα.

Για περαιτέρω ανάλυση σχετικά με τον αλγόριθμο path-following και την ανάλυση σύγκλισης, την πολυπλοκότητα κτλ καθώς για διάφορες παραλλαγές του αλγορίθμου όπως ο αλγόριθμος πρόβλεψης-διόρθωσης (predictor-corrector), ο ενδιαφερόμενος αναγνώστης παραπέμπεται στη βιβλιογραφία, ενδεικτικά αναφέρονται τα βιβλία [13], [11], [10], [15].

5.4 Primal-dual μέθοδοι εσωτερικού σημείου

Θα αναφερθούμε επιγραμματικά στη μέθοδο primal-dual εσωτερικού σημείου. Η μέθοδος αυτή έχει αρκετά κοινά με τη μέθοδο path-following αλλά με κάποιες διαφορές. Πρώτη διαφορά είναι ότι υπάρχει μία μόνο επανάληψη, δηλαδή δεν υπάρχει διαφορά μεταξύ της εσωτερικής και της εσωτερικής επανάληψης και σε κάθε επανάληψη ανανεώνονται οι βασικές και οι δυϊκές μεταβλητές. Δεύτερον, η διεύθυνση εύρεσης της λύσης, στη μέθοδο primal-dual λαμβάνεται από τη μέθοδο Newton που εφαρμόζεται σε παραλλαγή των συνθηκών KKT. Η λογική εύρεσης της διεύθυνσης είναι παρόμοια με αυτή της μεθόδου path-following αλλά όχι η ίδια. Τέλος, στη μέθοδο primal-dual, οι βασικές και δυϊκές μεταβλητές κατά τη διάρκεια των επαναλήψεων δεν είναι απαραίτητα εφικτές.

Η primal-dual μέθοδος είναι πολλές φορές περισσότερο αποδοτική από τη μέθοδο path-following, ειδικά όταν απαιτείται μεγάλη ακρίβεια. Για πολλές κατηγορίες προβλημάτων όπως προβλήματα, γραμμικού, τετραγωνικού, ημι-ορισμένου προγραμματισμού κτλ η μέθοδος primal-dual είναι αρκετά αποδοτικότερη από τη μέθοδο path-following. Για γενικά προβλήματα μη-γραμμικού προγραμματισμού η καθιέρωση της μεθόδου είναι αντικείμενο μελέτης αλλά με θετικά σημάδια.

5.4.1 Primal-dual διεύθυνση εύρεσης

Ξεκινάμε με τροποποίηση των συνθηκών KKT εκφραζόμενες ως $r_t(x, \lambda, v) = 0$ και

$$r_t(x, \lambda, v) = \begin{pmatrix} \nabla f_0(x) + Df(x)^T \lambda + A^T v \\ -\text{diag}(\lambda)f(x) - (1/t)\mathbf{1} \\ Ax - b \end{pmatrix},$$

με $t > 0$ και $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$. Συγκεκριμένα, x είναι η βασική μεταβλητή και λ, v είναι οι δυϊκές. Το πρώτο μπλοκ γραμμών στην παραπάνω λέγεται δυϊκό υπόλοιπο, r_{dual} , το τελευταίο βασικό

υπόλοιπο, r_{pri} και το μεσαίο υπόλοιπο κεντρικότητας r_{cent} . Αν συμβολίσουμε $y = (x, \lambda, v)$ τότε το επόμενο βήμα Newton είναι $r_t(y + \Delta y) \approx r_t(y) + Dr_t(y)\Delta y = 0$, και άρα το νέο σημείο είναι $\Delta y = -Dr_t(y)^{-1}r_t(y)$. Δηλαδή, για να υπολογίσουμε το νέο σημείο, primal-dual διεύθυνση λύσης Δy_{pd} , πρέπει να λύσουμε το σύστημα

$$\begin{pmatrix} \nabla^2 f_0(x) + \sum \lambda_i \nabla^2 f_i(x) & Df(x)^T & A^T \\ -\text{diag}(\lambda)Df(x) & -\text{diag}(f(x)) & 0 \\ A & 0 & 0 \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta \lambda \\ \Delta v \end{pmatrix} = - \begin{pmatrix} r_{dual} \\ r_{cent} \\ r_{pri} \end{pmatrix}.$$

5.4.2 Διατύπωση του αλγορίθμου primal-dual

Στο σημείο αυτό θα παρουσιάσουμε τη μέθοδο σε μορφή αλγορίθμου και θα σχολιάσουμε κάποια στοιχεία αυτής. Στο πρώτο βήμα του αλγορίθμου, θέτουμε την παράμετρο $t = \mu m / \hat{\eta}$,

Require: x που ικανοποιεί $f_1(x) < 0, \dots, f_m(x) < 0, \lambda > 0, \mu > 1, \epsilon_{feas} > 0, \epsilon > 0$.

repeat

1. Υπολόγισε t . Όρισε $t := \mu m / \hat{\eta}$.
2. Υπολόγισε primal-dual διεύθυνση λύσης. Δy_{pd} .
3. Line search και ανανέωση.

Υπολόγισε μήκος βήματος $s > 0$ και όρισε $y := y + s\Delta y_{pd}$.

until $\|r_{pri}\|_2 \leq \epsilon_{feas}, \|r_{dual}\|_2 \leq \epsilon_{feas}$ και $\hat{\eta} \leq \epsilon$

Σχήμα 5.2: Αλγόριθμος primal-dual εσωτερικού σημείου

που σχετίζεται με το αντίστοιχο δυϊκό κενό $\hat{\eta}$. Που ορίζεται για το x που ικανοποιεί $f(x) < 0$ και $\lambda \geq 0$ ως $\hat{\eta} = f(x)^T \lambda$ και αν $r_{pri} = 0$ και $r_{dual} = 0$ τότε θα ήταν ίσο με το δυϊκό κενό δηλαδή θα εξέφραζε την απόκλιση της βασικής από τις δυϊκές μεταβλητές, όπως το είδαμε στην προηγούμενη ενότητα. Αν τα x, λ, v ακολουθούσαν κεντρική διαδρομή, όπως στον αλγόριθμο path-following, τότε θα αυξάναμε την παράμετρο t επί την ποσότητα μ . Ο αλγόριθμος 5.2 τερματίζει όταν βασική μεταβλητή x και οι δυϊκές λ, v είναι εφικτές, με απόκλιση όσο η ανοχή ϵ_{feas} , και όταν το κενό $\hat{\eta}$ είναι μικρότερο της ανοχής ϵ . Επειδή η μέθοδος αυτή είναι γρήγορη γραμμική σύγκλιση, είναι συνηθισμένο να επιλέγουμε μικρές ανοχές.

Για να υπολογίσουμε το μήκος βήματος s , αρχικά βρίσκουμε το μέγιστο δυνατό που δίνει για

το νέο σημείο $\lambda^+ \geq 0$,

$$\begin{aligned} s^{max} &= \sup\{s \in [0, 1] : \lambda + s\Delta\lambda \geq 0\} \\ &= \min\{1, \min\{-\lambda_i/\Delta\lambda_i : \Delta\lambda_i < 0\}\}. \end{aligned}$$

Εφόσον βρούμε το μέγιστο s^{max} βρίσκουμε το μήκος βήματος $s = \beta s^{max}$ με $\beta \in (0, 1)$ μέχρι να ικανοποιείται η ανισότητα

$$\|r_t(x^+, \lambda^+, v^+)\|_2 \leq (1 - \alpha s) \|r_t(x, \lambda, v)\|_2,$$

με τυπικές τιμές για τις παραμέτρους (α, β) να είναι στο εύρος 0.01 με 0.1 για τη πρώτη και 0.3 με 0.8 για τη δεύτερη.

5.5 Ο αλγόριθμος του Karmarkar

Ο αλγόριθμος του Karmarkar ήταν ο πρώτος αλγόριθμος εσωτερικού σημείου που εμφανίστηκε. Παρόλο που σήμερα δεν χρησιμοποιείται, τουλάχιστον όχι στη μορφή που παρουσιάστηκε από τον Karmarkar, καθώς οι αλγόριθμοι εσωτερικού σημείου του παρόντος κεφαλαίου παρουσιάζουν καλύτερες ιδιότητες, αξίζει να αναφερθούμε στη μέθοδο αυτή. Η προσέγγιση που θα ακολουθήσει βασίστηκε στο βιβλίο [17] και αφορά γραμμικά προβλήματα. Για τροποποιήσεις της μεθόδου για μη-γραμμικά προβλήματα προτείνεται το βιβλίο [10].

5.5.1 Κανονική μορφή του Karmarkar

Για να εφαρμοστεί ο αλγόριθμος του Karmarkar πρέπει το γραμμικό πρόβλημα να μετασχηματιστεί σε συγκεκριμένη μορφή, η λεγόμενη κανονική μορφή του Karmarkar της μορφής

$$\begin{aligned} &\text{minimize} && c^T x \\ &\text{subject to} && Ax = 0 \\ &&& \sum_{i=1}^n x_i = 1 \\ &&& x \geq 0, \end{aligned}$$

όπου $x = [x_1, \dots, x_n]^T$. Για τη συνέχεια, χωρίς μείωση της γενικότητας θα θεωρήσουμε ότι A, c είναι ακέραιοι.

Αρχικά, θα δώσουμε κάποιους συμβολισμούς, Έστω $e = [1, \dots, 1]^T$ το μοναδιαίο διάνυσμα στο \mathbb{R}^n . Έστω Ω ο μηδενοχώρος του A , δηλαδή το σύνολο

$$\Omega = \{x \in \mathbb{R}^n : Ax = 0\}.$$

Ορίζουμε το simplex ως

$$\Delta = \{x \in \mathbb{R}^n : e^T x = 1, x \geq 0\},$$

και δηλώνουμε το κέντρο του simplex Δ με

$$a_0 = \frac{e}{n} = \left[\frac{1}{n}, \dots, \frac{1}{n} \right]^T.$$

Έτσι η κανονική μορφή του Karmarkar μπορεί να γραφτεί ως

$$\begin{aligned} &\text{minimize} && c^T x \\ &\text{subject to} && x \in \Omega \cap \Delta, \end{aligned} \tag{5.7}$$

με το σύνολο των περιορισμών, δηλαδή το εφικτό σύνολο να είναι

$$\Omega \cap \Delta = \{x \in \mathbb{R}^n : Ax = 0, e^T x = 1, x \geq 0\}.$$

5.5.2 Το πρόβλημα του Karmarkar υπό περιορισμούς

Ο αλγόριθμος του Karmarkar λύνει γραμμικά προβλήματα που είναι στην κανονική μορφή που περιγράψαμε με τις ακόλουθες υποθέσεις:

1. Το κέντρο a_0 του simplex Δ είναι εφικτό σημείο, δηλαδή $a_0 \in \Omega$.
2. Η ελάχιστη τιμή της αντικειμενικής συνάρτησης στο εφικτό σύνολο είναι μηδέν.
3. Ο $(m+1) \times n$ πίνακας

$$\begin{pmatrix} A \\ e^T \end{pmatrix}$$

έχει βαθμό $m+1$.

4. Μας δίνεται παράμετρος τερματισμού $q > 0$, τέτοια ώστε αν βρούμε εφικτό σημείο που να ικανοποιεί

$$\frac{c^T x}{c^T a_0} \leq 2^{-q},$$

τότε θεωρούμε ότι το πρόβλημα λύθηκε.

Κάθε γραμμικό πρόβλημα που είναι της κανονικής μορφής του Karmarkar και ικανοποιεί τις τέσσερις παραπάνω υποθέσεις καλείται πρόβλημα Karmarkar υπό περιορισμούς. Συγκεκριμένα, η πρώτη υπόθεση δεν είναι περιοριστική, διότι οποιοδήποτε γραμμικό πρόβλημα που έχει βέλτιστη λύση μπορεί να μετασχηματιστεί σε κανονική μορφή που ικανοποιεί την υπόθεση αυτήν. Όσον αφορά την δεύτερη υπόθεση, κάθε κανονικοποιημένο γραμμικό πρόβλημα μπορεί να μετατραπεί ούτως ώστε να ικανοποιείται η δεύτερη υπόθεση με την προϋπόθεση ότι γνωρίζουμε εκ των προτέρων την ελάχιστη τιμή της αντικειμενικής. Η τρίτη υπόθεση σχετίζεται με την εφαρμογή του αλγορίθμου και η τέταρτη έχει να κάνει με τη συνθήκη τερματισμού που είναι χαρακτηριστικό των αριθμητικών μεθόδων.

5.5.3 Μετασχηματισμός στην κανονική μορφή του Karmarkar

Θα δείξουμε πως κάθε γραμμικό πρόβλημα μπορεί να μετασχηματιστεί σε ισοδύναμο πρόβλημα της κανονικής μορφής του Karmarkar, δηλαδή σε η λύση του ενός μπορεί να χρησιμοποιηθεί για να βρούμε λύση για το άλλο. Ένα τυπικό γραμμικό πρόβλημα είναι της μορφής

$$\begin{aligned} &\text{minimize} && c^T x \\ &\text{subject to} && Ax = b \\ &&& x \geq 0, \end{aligned}$$

με $x \in \mathbb{R}^n$. Μία απλή μέθοδος για να μετασχηματίσουμε το παραπάνω είναι ορίζοντας μία νέα μεταβλητή $z \in \mathbb{R}^{n+1}$ ως

$$z = \begin{pmatrix} x \\ 1 \end{pmatrix}.$$

Ακόμη, ορίζουμε $c' = (c^T, 0)^T$ και $A' = (A, -b)$. Με τους συμβολισμούς αυτούς προκύπτει

$$\begin{aligned} &\text{minimize} && c'^T z \\ &\text{subject to} && A'z = 0 \\ &&& z \geq 0. \end{aligned}$$

Ένα ακόμα βήμα χρειάζεται για να μετατρέψουμε το πρόβλημα σε αντίστοιχο με τους περιορισμούς να αθροίζονται την μονάδα. Ορίζουμε $y = (y_1, \dots, y_{n+1})^T \in \mathbb{R}^{n+1}$ όπου

$$y_i = \frac{x_i}{x_1 + \dots + x_n + 1}, \quad i = 1, \dots, n$$

$$y_{n+1} = \frac{1}{x_1 + \dots + x_n + 1}.$$

Ο μετασχηματισμός αυτός από τις μεταβλητές x στις y λέγεται προβολικός μετασχηματισμός. Με τον τρόπο αυτόν μετατρέψαμε το αρχικό γραμμικό πρόβλημα στο αντίστοιχο της κανονικής μορφής Karmarkar

$$\begin{aligned} &\text{minimize} && c^T y \\ &\text{subject to} && A'y = 0 \\ &&& e^T y = 1 \\ &&& y \geq 0. \end{aligned}$$

Η τεχνική αυτή μπορεί να αλλαχθεί ελαφρώς για να εξασφαλίσουμε ότι η πρώτη υπόθεση του προηγούμενου υπο-κεφαλαίου θα ικανοποιείται. Περισσότερες λεπτομέρειες μπορούν να βρεθούν στο βιβλίο του [17].

5.5.4 Διατύπωση του αλγορίθμου του Karmarkar

Στο σημείο αυτό μπορούμε να περιγράψουμε τον αλγόριθμο του Karmarkar. Ο αλγόριθμος αφορά γραμμικό πρόβλημα που είναι της κανονικής μορφής του Karmarkar και ικανοποιούνται οι τέσσερις υποθέσεις που περιγράφηκαν προηγουμένως, έχουμε δηλαδή να κάνουμε με το πρόβλημα της σχέσης (5.7). Ο αλγόριθμος είναι επαναληπτικός και ξεκινάει δοθέντος ενός αρχικού σημείου $x^{(0)}$ και της παραμέτρου q και δίνει μία ακολουθία σημείων $x^{(1)}, x^{(2)}, \dots, x^{(N)}$.

Require: 1. **Όρισε:** $k := 0, x^{(0)} = a_0 = e/n$

repeat

2. **Ανανέωσε:** Θέσε $x^{(k+1)} := \Psi(x^{(k)})$ όπου Ψ είναι ο πίνακας ανανέωσης

3. **Επανάλαβε:** Θέσε $k := k + 1$

until 4. **Κριτήριο τερματισμού:** Να ικανοποιείται $c^T x^{(k)} / c^T x^{(0)} \leq 2^{-q}$

Σχήμα 5.3: Αλγόριθμος εσωτερικού σημείου του Karmarkar

Η διαδικασία για τον υπολογισμό της ανανέωσης $x^{(k+1)} = \Psi(x^{(k)})$ μπορεί να συνοψιστεί στα παρακάτω βήματα.

1. Υπολόγισε τους πίνακες

$$D_k = \text{diag} \left(x_1^{(k)}, \dots, x_n^{(k)} \right), \quad B_k = \begin{pmatrix} AD_k \\ e^T \end{pmatrix}.$$

2. Υπολόγισε την ορθογώνια προβολή στο μηδενικό χώρο του B_k

$$P_k = I_n - B_k^T (B_k B_k^T)^{-1} B_k.$$

3. Υπολόγισε την κανονικοποιημένη ορθογώνια προβολή του c στο μηδενικό χώρο του B_k

$$\hat{c}^{(k)} = \frac{P_k D_k c}{\|P_k D_k c\|}.$$

4. Υπολόγισε το διάνυσμα διεύθυνσης

$$d^{(k)} = -r \hat{c}^{(k)}, \quad \text{όπου } r = \frac{1}{\sqrt{n(n-1)}}.$$

5. Υπολόγισε $\bar{x}^{(k+1)}$ από

$$\bar{x}^{(k+1)} = a_0 + \alpha d^{(k)},$$

όπου α είναι το προκαθορισμένο μέγεθος βήματος, $\alpha \in (0, 1)$.

6. Υπολόγισε $x^{(k+1)}$ εφαρμόζοντας τον αντίστροφο μετασχηματισμό U_k^{-1}

$$x^{(k+1)} = U_k^{-1}(\bar{x}^{(k+1)}) = \frac{D_k \bar{x}^{(k+1)}}{e^T D_k \bar{x}^{(k+1)}}.$$

Κεφάλαιο 6

Γραμμικές ανισότητες πινάκων

Τα τελευταία χρόνια, οι γραμμικές ανισότητες πινάκων (LMI) έχουν αναδυθεί ως ένα ισχυρό εργαλείο προσέγγισης προβλημάτων ελέγχου που φαίνονται δύσκολα, αν όχι αδύνατον να επιλυθούν αναλυτικά. Η ιστορία των γραμμικών ανισοτήτων πινάκων ξεκινά γύρω στο 1890 με τον Lyapunov και τη μελέτη ευστάθειας δυναμικών συστημάτων. Τη δεκαετία του σαράντα, επιστήμονες από τη Σοβιετική Ένωση εφάρμοσαν τις ιδέες του Lyapunov προσπαθώντας να λύσουν προβλήματα ελέγχου. Αν και δεν διατύπωσαν τη μορφή των γραμμικών ανισοτήτων πινάκων, η συνθήκη ευστάθειας είχε τη μορφή LMI. Μεγάλη έξαρση στη θεωρία τους παρουσιάστηκε τη δεκαετία του εξήντα με τη δουλεία των Kalman, Yakubovich, Popov και άλλων, με ιδιαίτερη έμφαση στις εφαρμογές των LMI στη θεωρία ελέγχου. Τις τελευταίες δεκαετίες αναπτύχθηκαν ισχυροί αλγόριθμοι για την αριθμητική επίλυση των γραμμικών ανισοτήτων πινάκων. Η ιδέα των LMI και των εφαρμογών τους βασίζονται στο γεγονός ότι τα LMI μπορούν να μετασχηματιστούν σε ισοδύναμα προβλήματα κυρτού προγραμματισμού τα οποία μπορούν εύκολα να επιλυθούν.

Στο κεφάλαιο αυτό θα παρουσιάσουμε κάποια βασικά θεωρητικά στοιχεία των γραμμικών ανισοτήτων πινάκων με βάση τα βιβλία [2] και [12]. Στη συνέχεια θα παρουσιάσουμε κάποια παραδείγματα LMI στο MATLAB χρησιμοποιώντας το Robust Control Toolbox.

6.1 Θεωρητικά στοιχεία

Μία γραμμική ανισότητα πινάκων είναι της μορφής

$$F(x) = F_0 + \sum_{i=1}^m x_i F_i > 0, \quad (6.1)$$

όπου $x = (x_1 \dots x_m) \in \mathbb{R}^m$ είναι η μεταβλητή βελτίστου και οι συμμετρικοί πίνακες $F_i = F_i^T \in \mathbb{R}^{n \times n}$ για $i = 0, \dots, m$ δίνονται. Το σύμβολο ανισότητας στην παραπάνω σχέση σημαίνει ότι η $F(x)$ είναι θετικά ορισμένη. Πολλές φορές συναντάμε LMI όπου η $F(x) \geq 0$ είναι θετικά ημι-ορισμένη.

Το LMI (6.1) είναι ένας κυρτός περιορισμός στο x , δηλαδή το σύνολο $\{x | F(x) > 0\}$ είναι κυρτό. Αν και φαίνεται ότι LMI έχει ειδική μορφή, μπορεί να συμβολίσει μεγάλο εύρος κυρτών περιορισμών που παρουσιάζονται σε φυσικά προβλήματα. Πιο συγκεκριμένα, γραμμικές ανισότητες, (κυρτές) τετραγωνικές ανισότητες, περιορισμοί που εμφανίζονται σε προβλήματα της θεωρίας ελέγχου, όπως συναρτήσεις Lyapunov και κυρτές τετραγωνικές ανισότητες πινάκων.

Ένα σύστημα LMI $F^{(1)}(x) > 0, \dots, F^{(p)}(x) > 0$ μπορεί να μετασχηματιστεί στην ισοδύναμη μορφή ενός LMI

$$\begin{pmatrix} F^{(1)}(x) & & & \\ & F^{(2)}(x) & & \\ & & \ddots & \\ & & & F^{(p)}(x) \end{pmatrix} > 0.$$

Ιδιαίτερα χρήσιμο είναι το συμπλήρωμα του Schur που μας επιτρέπει να μετατρέψουμε μη-γραμμικές ανισότητες σε LMI. Πιο συγκεκριμένα,

$$F(x) = \begin{pmatrix} F_{11}(x) & F_{12}(x) \\ F_{21}(x) & F_{22}(x) \end{pmatrix}$$

όπου $F_{11}(x)$ είναι τετράγωνος. Τότε $F(x) > 0$ αν και μόνο αν

$$\begin{aligned} F_{22}(x) &> 0 \\ F_{11}(x) - F_{12}(x)F_{22}^{-1}(x)F_{21}(x) &> 0, \end{aligned}$$

ή ισοδύναμα

$$\begin{aligned} F_{11}(x) &> 0 \\ F_{22}(x) - F_{21}(x)F_{11}^{-1}(x)F_{12}(x) &> 0. \end{aligned}$$

Έτσι αν κάθε $F_i(x)$ είναι γραμμική συνάρτηση τότε μπορούμε να μετατρέψουμε τη μη-γραμμική ανισότητα σε LMI.

6.2 Κατηγορίες γραμμικών ανισοτήτων πινάκων

Πολλά προβλήματα βελτιστοποίησης στον έλεγχο, στην επεξεργασία σήματος και τα λοιπά, μπορούν να διατυπωθούν ως γραμμικές ανισότητες πινάκων. Προφανώς, έχει νόημα να εκφράσεις τα προβλήματα αυτά ως LMI αν μπορούν να επιλυθούν με αποδοτικό και αξιόπιστο τρόπο. Κατά τη μελέτη γραμμικών ανισοτήτων πινάκων διακρίνουμε δύο κατηγορίες προβλημάτων:

1. **Εφικτότητα:** Το ερώτημα αν υπάρχει x τέτοιο ώστε $F(x) > 0$ ονομάζεται πρόβλημα εφικτότητας. Το LMI $F(x) > 0$ καλείται εφικτό αν υπάρχει τέτοιο x , αλλιώς καλείται μη-εφικτό.
2. **Βελτιστοποίηση:** Έστω η αντικειμενική συνάρτηση f που υπόκειται σε περιορισμούς της μορφής $\{x | F(x) > 0\}$. Το πρόβλημα έγκειται στην εύρεση του x το οποίο θα ελαχιστοποιεί την αντικειμενική συνάρτηση.

Στη συνέχεια θα δώσουμε κάποια τυπικά απλά παραδείγματα των παραπάνω κατηγοριών.

6.2.1 Παράδειγμα 1: ευστάθεια

Θα εξετάσουμε την εκθετική ευστάθεια του γραμμικού αυτόνομου συστήματος

$$\dot{x} = Ax$$

όπου $A \in \mathbb{R}^{n \times n}$. Με αυτό εννοούμε αν υπάρχει ή δεν υπάρχει θετικός και σταθερός M και $a > 0$ τέτοιοι ώστε για κάθε αρχική συνθήκη x_0 η λύση $x(t)$ με $x(t_0) = x_0$ να ικανοποιεί το όριο

$$\|x(t)\| \leq \|x(t_0)\| M e^{-a(t-t_0)}, \quad \text{για κάθε } t \geq t_0.$$

Από τον Lyapunov γνωρίζουμε ότι το σύστημα είναι εκθετικά ευσταθές αν και μόνο αν υπάρχει $X = X^T$ τέτοιο ώστε $X > 0$ και $A^T X + X A < 0$. Συνεπώς, η εκθετική ευστάθεια του συστήματος είναι ισοδύναμο με το πρόβλημα εφικτότητας LMI της μορφής

$$\begin{pmatrix} -X & 0 \\ 0 & A^T X + X A \end{pmatrix} < 0.$$

6.2.2 Παράδειγμα 2: εκτίμηση τετραγωνικού κόστους

Έστω το γραμμικό αυτόνομο σύστημα

$$\dot{x} = Ax, \quad x(0) = x_0$$

μαζί με το δείκτη επίδοσης $J = \int_0^\infty x^T(t) Q x(t) dt$ όπου $Q = Q^T \geq 0$. Έστω ότι το σύστημα είναι ασυμπτωτικά ευσταθές. Τότε όλες οι λύσεις του x είναι τετραγωνικά ολοκληρώσιμες ώστε $J < \infty$. Εξετάζοντας τη μη-αυστηρή γραμμική ανισότητα πινάκων $X \geq 0$ και $A^T X X A + Q \leq 0$. Για κάθε εφικτή λύση $X = X^T$ μπορούμε να παραγωγίσουμε την ποσότητα $x^T(t) X x(t)$ ως προς το χρόνο

$$\frac{d}{dt} [x^T(t) X x(t)] = x^T(t) [A^T X + X A] x(t) \leq -x^T(t) Q x(t).$$

Ολοκληρώνοντας την τελευταία από $t = 0$ έως ∞ παίρνουμε το πάνω όριο

$$J = \sum_0^\infty x^T(t) Q x(t) dt \leq x_0^T X x_0,$$

όπου χρησιμοποιήσαμε $\lim_{t \rightarrow \infty} x(t) = 0$. Επίσης, το μικρότερο πάνω όριο της J προκύπτει ελαχιστοποιώντας τη συνάρτηση $f(X) = x_0^T X x_0$ ως προς το $X = X^T$ που ικανοποιεί τις ανισώσεις $X \geq 0$ και $A^T X + X A + Q \leq 0$. Τελικά, προκύπτει ένα πρόβλημα ελαχιστοποίησης με LMI περιορισμούς.

6.3 Επίλυση

Προβλήματα LMI μπορούν να λυθούν στο MATLAB με το Robust Control Toolbox. Η διαδικασία για να ορίσουμε το πρόβλημα είναι αρκετά περίπλοκη, για αυτό θα περιγράψουμε τα βήματα που απαιτούνται.

1. **Αρχικοποίηση του LMI συστήματος** Ένα σύστημα LMI αρχικοποιείται με τη συνάρτηση `setlmiis([])`. Σε περίπτωση που θέλουμε να επεξεργαστούμε ένα υπάρχον LMI αυτό γίνεται δηλώνοντας το όνομα μέσα στις παρενθέσεις. Για παράδειγμα εάν το υπάρχον LMI ονομάζεται `lmi0` τότε με τη `setlmiis(lmi0)` μπορούμε να το επεξεργαστούμε.
2. **Δήλωση μεταβλητών βελτιστοποίησης** Η δήλωση των μεταβλητών βελτιστοποίησης γίνεται με τη συνάρτηση `X = lmiivar(type, struct)` ορίζουμε τη μεταβλητή πίνακα που θέλουμε να βελτιστοποιήσουμε για το LMI σύστημα που μελετάμε. Με την πρώτη είσοδο της συνάρτησης, `type`, καθορίζουμε τον τύπο της μεταβλητής και με το `struct` δίνουμε επιπλέον στοιχεία για τη δομή του πίνακα.
 - (α') **type=1** Τότε έχουμε συμμετρικούς πίνακες με μπλοκ διαγώνια μορφή. Κάθε διαγώνιο μπλοκ είναι είτε φουλ (συμμετρικός πίνακας), είτε βαθμωτό μέγεθος (πολλαπλάσιο μοναδιαίου πίνακα), είτε μηδενικός πίνακας. Αν X έχει R διαγώνια μπλοκ, τότε η δομή `struct` είναι $R \times 2$ πίνακας όπου, `struct(r, 1)` είναι το μέγεθος του r -μπλοκ και `struct(r, 2)` είναι ο τύπος το r -μπλοκ (1 για φουλ, 0 για βαθμωτό και -1 για μηδενικό μπλοκ).
 - (β') **type=2** Για $m \times n$ ορθογώνιο πίνακα. Το `struct = [m, n]` στην περίπτωση αυτή.
 - (γ') **type=3** Για άλλες μορφές πινάκων. Με την τιμή αυτή κάθε στοιχείο του X καθορίζεται μηδενικό ή $\pm x_n$ όπου x_n δηλώνει την n -μεταβλητή βελτιστοποίησης.
3. **Δήλωση των όρων του LMI** Κάθε όρος του LMI δηλώνεται ξεχωριστά, με τη συνάρτηση `lmiterm(termID, A, B, flag)`. Οι όροι του LMI διαχωρίζονται σε εξωτερικούς όρους, σε σταθερούς όρους και μεταβλητούς όρους. Όταν δηλώνουμε τους όρους του LMI μπορούμε να δηλώσουμε μόνο τους όρους πάνω ή κάτω από τη διαγώνιο, το πρόγραμμα θα δηλώσει αυτόματα τον όρο που παραλείπουμε. Ο πρώτος όρος του `termID` μπορεί να είναι

$$\text{termID}(1) = \begin{cases} +p \\ -p \end{cases}$$

όπου $+p$ δηλώνει τους όρους αριστερού μέλους του LMI και $-p$ τους δεξιούς όρους. Επίσης, η τιμή του $\|p\|$ έχει να κάνει με το LMI που περιγράφουμε. Όπως είπαμε, μπορούμε να έχουμε πολλά LMI και έτσι η τιμή του p αντιστοιχείται από τη συνάρτηση `newlmi`.

Οι υπόλοιπες τιμές του termID είναι

$$\text{termID}(2:3) = \begin{cases} [0, 0] & \text{για τους εξωτερικούς όρους,} \\ [i, j] & \text{για τη θέση του μπλοκ του εσωτερικού όρου,} \end{cases}$$

και

$$\text{termID}(4) = \begin{cases} 0 & \text{για τους εξωτερικούς όρους,} \\ x & \text{για τους όρους } AXB \\ -x & \text{για τους όρους } AX^TB, \end{cases}$$

όπου x είναι ότι επιστρέφεται από τη συνάρτηση `lminvar` για τη μεταβλητή πίνακα X , όπως δηλώθηκε δηλαδή στο προηγούμενο βήμα.

Τα στοιχεία A , B είναι τα αριθμητικά δεδομένα του προβλήματος. Αν κάποιος πίνακας δεν εμφανίζεται στον όρο του LMI που δηλώνουμε τότε παίρνει την τιμή 1. Αν είναι σταθερός όρος ή εξωτερικός όρος τότε ο B παραλείπεται.

Τέλος, αν `flag = 's'` τότε προσθέτεται ο ανάστροφος του όρου στον όρο που δηλώσαμε με τη συνάρτηση `lmiterm`. Για παράδειγμα η δήλωση `lmiterm([1 1 1 X], A, 1, 's')` μας δίνει τον όρο $AX + X^T A^T$ του $(1,1)$ -μπλοκ του πρώτου LMI και είναι ισοδύναμο με τις εντολές `lmiterm([1 1 1 X], A, 1)` και `lmiterm([1 1 1 -X], 1, A')`.

4. **Επιβεβαίωση του LMI** Εφόσον έχουμε ολοκληρώσει τις δηλώσεις του συστήματος των LMI που θέλουμε να λύσουμε, με τη συνάρτηση `getlmis` επιβεβαιώνεται ότι δεν έχουμε εισάγει σωστά τα δεδομένα και τελικά επιστρέφεται από τη συνάρτηση η εσωτερική αναπαράσταση του προβλήματος.
5. **Επίλυση του LMI** Τελευταίο βήμα είναι να λύσουμε το σύστημα LMI. Το MATLAB επιλύει προβλήματα εφικτότητας με τη συνάρτηση

$$[\text{tmin}, \text{xfeas}] = \text{feasp}(\text{lmissys}, \text{options}, \text{target})$$

όπου υποχρεωτικό όρισμα της παραπάνω συνάρτησης είναι το `lmissys` που θέλουμε να λύσουμε. Μας επιστρέφει το `xfeas` που είναι διάνυσμα που σχετίζεται με τις μεταβλητές απόφασης και όχι με τον πίνακα βελτιστοποίησης του LMI. Για αυτό απαιτείται η χρήση της συνάρτησης `dec2mat`. Επίσης, αν το `tmin` είναι μη θετικό τότε είναι εφικτό το πρόβλημα και αν είναι αρνητικό τότε είναι αυστηρώς εφικτό.

Αν θέλουμε να λύσουμε πρόβλημα γενικευμένων ιδιοτιμών γίνεται με την παρακάτω συνάρτηση.

$$[\text{lopt}, \text{xopt}] = \text{gevp}(\text{lmsys}, \text{nlf}, \text{options}, \text{linit}, \text{xinit}, \text{target})$$

Εδώ υποχρεωτικό όρισμα είναι και ο όρος `nlf` που δηλώνει το γραμμικό-κλασματικό περιορισμό. Η λύση προκύπτει πάλι με τη συνάρτηση `dec2mat` και η τιμή εξόδου `lopt` είναι το ολικό ελάχιστο.

Τέλος, αν θέλουμε να ελαχιστοποιήσουμε γραμμική αντικειμενική συνάρτηση που υπόκειται σε LMI περιορισμούς γίνεται με τη συνάρτηση

$$[\text{copt}, \text{xopt}] = \text{mincx}(\text{lmsys}, \text{c}, \text{options}, \text{xinit}, \text{target}).$$

Ο όρος `c` είναι διάνυσμα που πολλαπλασιάζει το διάνυσμα που θέλουμε να ελαχιστοποιήσουμε. Ο όρος `copt` είναι το ολικό ελάχιστο της αντικειμενικής συνάρτησης και ο όρος `xopt`, το βέλτιστο διάνυσμα, σχετίζεται με τους αντίστοιχους πίνακες LMI και πάλι με τη συνάρτηση `dec2mat`.

6.3.1 Παράδειγμα 1

Ένα χαρακτηριστικό παράδειγμα LMI που διατυπώνεται στο βιβλίο [16], είναι η ανισότητα Riccati

$$A^T X + X A + X B R^{-1} B^T X + Q < 0,$$

όπου ψάχνουμε τον θετικά ορισμένο πίνακα X . Προφανώς η ανισότητα είναι μη-γραμμική. Όμως κάνοντας χρήση του συμπληρώματος του Schur μπορούμε να διατυπώσουμε την ανισότητα ως LMI καθώς οι πίνακες A, B, Q είναι σταθεροί. Έτσι έχουμε

$$\begin{pmatrix} A^T X + X A + Q & X B \\ B^T X & -R \end{pmatrix} < 0, \quad \text{και } X > 0. \quad (6.2)$$

Χάριν του παραδείγματος έστω

$$A = \begin{pmatrix} -2 & -2 & -1 \\ -3 & -1 & -1 \\ 1 & 0 & -4 \end{pmatrix}, \quad B = \begin{pmatrix} -1 & 0 \\ 0 & -1 \\ -1 & -1 \end{pmatrix}, \quad Q = \begin{pmatrix} -2 & 1 & -2 \\ 1 & -2 & -4 \\ -2 & -4 & -2 \end{pmatrix}, \quad R = I_2.$$

Παρακάτω παρατίθεται ο κώδικας στο MATLAB για το πρόβλημα.

```

1 function X = lmi1()
2     A = [-2,-2,-1; -3,-1,-1; 1,0,-4];
3     B = [-1,0; 0,-1; -1,-1];
4     Q = [-2,1,-2; 1,-2,-4; -2,-4,-2];
5     R = eye(2);
6
7     setlmis([]); %create an LMI framework
8     X = lmivar(1,[3 1]); %declare X as a 3x3 symmetrical matrix
9     lmiterm([1 1 1 X],A',1,'s'); %(1,1) block, 's' means A^TX + XA
10    lmiterm([1 1 1 0],Q); %(1,1) block, plus the constant matrix Q
11    lmiterm([1 1 2 X],1,B); %(1,2) block, meaning XB
12    lmiterm([1 2 2 0],-1); %(2,2) block, meaning -R
13    lmiterm([-2,1,1,X],1,1); %the second inequality meaning X > 0
14    lmis = getlmis; %complete the LMI framework setting
15    [tmin xfeas] = feasp(lmis); %solve the feasible problem
16    X = dec2mat(lmis,xfeas,X); %extract the solution matrix X
17 end

```

Όπως φαίνεται στον κώδικα στη γραμμή 8, ο αριστερός όρος της σχέσης (6.2) είναι το LMI 1 και ο δεξιός το LMI 2, γραμμή 13. Ο πίνακας X είναι 3×3 συμμετρικός. Τελικά, βρίσκουμε ότι $t_{min} = -0.3962$ και ο πίνακας εφικτής λύσης που αντιστοιχεί στη βέλτιστη τιμή επιστρέφεται από τη συνάρτηση `dec2mat` και είναι

$$X = \begin{pmatrix} 1.0329 & 0.4647 & -0.2358 \\ 0.4647 & 0.7790 & -0.0507 \\ -0.2358 & -0.0507 & 1.4336 \end{pmatrix}.$$

6.3.2 Παράδειγμα 2

Το επόμενο παράδειγμα προέρχεται από το LMI Control Toolbox του MATLAB [4]. Έστω το πρόβλημα ελαχιστοποίησης

$$\begin{aligned} &\text{minimize} \quad \text{trace}(X) \\ &\text{subject to} \quad A^T X + X A + X B B^T + Q < 0, \end{aligned}$$

με δεδομένα

$$A = \begin{pmatrix} -1 & -2 & 1 \\ 3 & 2 & 1 \\ 1 & -2 & -1 \end{pmatrix}, \quad B = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \quad Q = \begin{pmatrix} 1 & -1 & 0 \\ -1 & -3 & -12 \\ 0 & -12 & -36 \end{pmatrix}.$$

Αποδεικνύεται ότι η βέλτιστη λύση X^* είναι η λύση της αλγεβρικής εξίσωσης Riccati

$$A^T X + X A + X B B^T X + Q = 0.$$

Η λύση της παραπάνω μπορεί να υπολογιστεί από τη συνάρτηση `care` του MATLAB και να συγκριθεί με το αποτέλεσμα της LMI ρουτίνας `mincx`.

Οι περιορισμοί του προβλήματος μπορούν να μετατραπούν σε LMI μορφή και επειδή το ίχνος είναι γραμμική συνάρτηση, προκύπτει το πρόβλημα ελαχιστοποίησης γραμμικής αντικειμενικής συνάρτησης

$$\begin{aligned} & \text{minimize} \quad \text{trace}(X) \\ & \text{subject to} \quad \begin{pmatrix} A^T X + X A + Q & X B \\ B^T X & -I \end{pmatrix} < 0. \end{aligned}$$

Στις γραμμές 6 με 12 δηλώνουμε το LMI όπως το προηγούμενο παράδειγμα. Στην γραμμή 13 δηλώνουμε την αντικειμενική συνάρτηση, δηλαδή το ίχνος του X . Άρα το διάνυσμα c είναι το I_3 καθώς το MATLAB αναμένει την αντικειμενική στη μορφή $c^T x$. Με την εντολή `mat2den` μετατρέπουμε τις μεταβλητές πινάκων σε μεταβλητές απόφασης για το γραμμικό πρόβλημα. Από τα `options` θέτουμε το σχετικό σφάλμα 10^{-5} . Λύνοντας το LMI πρόβλημα ελαχιστοποίησης βρίσκουμε ότι $\text{trace}(X) = -18.7166$ και ικανοποιείται με σχετική ακρίβεια $9.5 \cdot 10^{-6}$. Η μεταβλητή πίνακα που αντιστοιχεί στη βέλτιστη τιμή επιστρέφεται από τη συνάρτηση `dec2mat` και είναι

$$X = \begin{pmatrix} -6.3542 & -5.8895 & 2.2046 \\ -5.8895 & -6.2855 & 2.2201 \\ 2.2046 & 2.2201 & -6.0771 \end{pmatrix}.$$

Το αποτέλεσμα μπορεί να συγκριθεί με τη λύση της εξίσωσης Riccati που υπολογίζεται από τη συνάρτηση `care`. Τελικά, βρίσκουμε ότι η απόκλιση της βέλτιστης τιμής, όπως υπολογίστηκε από την προσέγγιση με τα LMI, με τη τιμή της εξίσωσης Riccati, όπως υπολογίστηκε από τη συνάρτηση `care`, γραμμή 17, είναι $6.539 \cdot 10^{-5}$.

Παρακάτω παρατίθεται ο κώδικας στο MATLAB για το πρόβλημα.

```
1 function X = lmi2()
2     A = [-1,-2,1; 3,2,1; 1,-2,-1];
3     B = [1;0;1];
4     Q = [1,-1,0; -1,-3,-12; 0,-12,-36];
5
6     setlmis([]); %create a blank LMI framework
7     X = lmivar(1,[3 1]); %declare X as a 3x3 symmetrical matrix
8     lmiterm([1 1 1 X],A',1,'s'); %(1,1) block, 's' means A^TX + XA
9     lmiterm([1 1 1 0],Q); %(1,1) block, plus the constant matrix Q
10    lmiterm([1 2 2 0],-1); %(2,2) block, meaning -I
11    lmiterm([1 2 1 X],B',1); %(2,1) block, meaning B^TX
12    lmis = getlmis; %complete the LMI framework setting
13    c = mat2dec(lmis,eye(3));
14    options = [1e-5,0,0,0,0]; %set relative accuracy 1e-5
15    [copt xopt] = mincx(lmis,c,options); %solve min problem
16    X = dec2mat(lmis,xopt,X); %extract the solution matrix X
17    Xst = care(A,B,Q,-1); %Riccati solution
18    norm(X - Xst) %LMI - Riccati solution
19 end
```

Κεφάλαιο 7

Συμπεράσματα

Στην παρούσα εργασία αναλύσαμε σε θεωρητικό επίπεδο διάφορες τεχνικές βελτιστοποίησης. Τα προβλήματα βελτιστοποίησης εμφανίζονται σε κάθε τομέα της επιστήμης και είναι απαραίτητο να γνωρίζουμε κάποια θεωρητικά κομμάτια των αλγορίθμων. Ακόμα, το μέγεθος των προβλημάτων που εμφανίζονται καθιστούν απαραίτητη τη χρήση εξειδικευμένου λογισμικού, όπως το MATLAB, για την επίλυση αυτών.

Στο πρώτο κεφάλαιο δώσαμε κάποια ιστορικά στοιχεία σχετικά με το γραμμικό προγραμματισμό. Στη συνέχεια πραγματοποιήσαμε μία βασική εισαγωγή στο θεωρητικό κομμάτι του γραμμικού προγραμματισμού, και τέλος παρουσιάσαμε κάποια βασικά παραδείγματα στο MATLAB. Το κεφάλαιο αυτό είναι από τα πιο σημαντικά κομμάτια στον τομέα της αριθμητικής βελτιστοποίησης. Καθώς όχι μόνο έχει μελετηθεί εκτενώς από θεωρητική σκοπιά αλλά έχουν αναπτυχθεί ισχυροί αλγόριθμοι που μας επιτρέπουν τη γρήγορη και αποτελεσματική επίλυση των προβλημάτων. Επίσης, τα γραμμικά προβλήματα βρίσκονται στην καρδιά κάθε προβλήματος και πάντα επιδιώκουμε να εκφράσουμε το πρόβλημα με γραμμική μορφή, αν φυσικά η φύση του προβλήματος το επιτρέπει.

Βασικό μειονέκτημα του γραμμικού προγραμματισμού αποτελεί το γεγονός ότι πολλές το πρόβλημα δε μπορεί να περιγραφεί ικανοποιητικά από γραμμικούς όρους. Για αυτό έχουν αναπτυχθεί άλλες τεχνικές βελτιστοποίησης. Στο δεύτερο κεφάλαιο αναφέραμε κάποια θεωρητικά στοιχεία του τετραγωνικού προγραμματισμού, δηλαδή όταν η συνάρτηση που θέλουμε να βελτιστοποιήσουμε είναι τετραγωνικής μορφής. Αναλύσαμε τη μέθοδο ενεργών περιορισμών που είναι ένας από τους επικρατέστερους αλγορίθμους για την επίλυση προβλημάτων τετραγωνικού προγραμματισμού και παραθέσαμε κάποια παραδείγματα επίλυσης τέτοιων

προβλημάτων.

Στο τρίτο κεφάλαιο παραθέτουμε τα απαραίτητα θεωρητικά κομμάτια της κυρτής θεωρίας. Είναι ένας κλάδος που έχει αναπτυχθεί ευρέως και παρουσιάζει μεγάλο ενδιαφέρον καθώς έχουν αναπτυχθεί αποτελεσματικοί μέθοδοι επίλυσης προβλημάτων κυρτής βελτιστοποίησης. Δεν υπάρχει αναλυτική λύση για τα προβλήματα αυτά, όμως έχει αποδειχθεί ότι σε κάποιες περιπτώσεις δύναται να λύσουν προβλήματα με εξαιρετική ακρίβεια και με πολυπλοκότητα που δε ξεπερνά την πολυωνυμική σε σχέση με το μέγεθος του προβλήματος. Η έρευνα σε αλγορίθμους εσωτερικού σημείου για γενικά μη-γραμμικά κυρτά προβλήματα συνεχίζεται και είναι πολύ ενεργή καθώς τα αποτελέσματα είναι πολύ καλά. Στο τέταρτο κεφάλαιο παρουσιάζουμε τη μέθοδο φράγματος, ενός αλγορίθμου που υπάγεται στην κατηγορία των μεθόδων εσωτερικών σημείων.

Στο τελευταίο κεφάλαιο κάνουμε μία μικρή αναφορά στις γραμμικές ανισότητες πινάκων. Είναι πολύ συνηθισμένο στον έλεγχο οι περιορισμοί να είναι μη-γραμμικοί όμως κάποιες τεχνικές να μετασχηματίζονται σε LMI μορφή, δηλαδή σε γραμμικές ανισότητες πινάκων. Επικεντρωθήκαμε κυρίως στον τρόπο που επιλύονται τα προβλήματα αυτά στο περιβάλλον προγραμματισμού MATLAB.

Βασικό κομμάτι των προβλημάτων βελτιστοποίησης είναι η κατανόηση του προβλήματος και η ευχέρεια να καταλάβουμε που κατατάσσεται, για παράδειγμα αν πρόκειται για γραμμικό πρόβλημα ή αν μπορεί να εκφραστεί ικανοποιητικά ως γραμμικό. Το επόμενο βήμα είναι η μαθηματική διατύπωση αυτού. Φυσικά δεν είναι εύκολο να εκφραστεί ένα φυσικό πρόβλημα μαθηματικά, είτε λόγω μεγέθους είτε λόγω πολυπλοκότητας του προβλήματος και ούτω καθεξής. Τέλος, η υλοποίηση προγράμματος που επιλύει το πρόβλημα δεν είναι εύκολη. Παρόλο αυτά, έχουν αναπτυχθεί διάφορα προγράμματα ανοιχτού και κλειστού κώδικα που διευκολύνουν τη διαδικασία της επίλυσης.

Βιβλιογραφία

- [1] M.S. Bazaraa, H.D. Sherali, and C.M. Shetty. *Nonlinear Programming: Theory and Algorithms*. Wiley, 2013.
- [2] Stephen P Boyd et al. *Linear matrix inequalities in system and control theory*. Vol. 15. SIAM, 1994.
- [3] G. B. Dantzig. *Linear Programming and Extensions*. Princeton landmarks in mathematics and physics. Princeton University Press, 1963.
- [4] PM Gahinet, Alan J Laub, and Mahmoud Chilali. «The LMI control toolbox». In: Citeseer.
- [5] L. V. Kantorovich. «Mathematical Methods of Organizing and Planning Production». In: *Management Science* 6.4 (1960), pp. 366–422.
- [6] N. Karmarkar. «A new polynomial-time algorithm for linear programming». In: *Combinatorica* 4.4 (1984), pp. 373–395.
- [7] L. G. Khachiyan. «A polynomial algorithm in linear programming». In: 20 (1979), pp. 191–194.
- [8] T. C. Koopmans. «Optimum utilization of the transportation system». In: *Econometrica* 17 (1949), pp. 136–146.
- [9] D.G. Luenberger and Y. Ye. *Linear and Nonlinear Programming*. International Series in Operations Research & Management Science. Springer US, 2008.
- [10] Y. Nesterov and A. Nemirovskii. *Interior-point Polynomial Algorithms in Convex Programming*. Studies in Applied Mathematics. Society for Industrial and Applied Mathematics (SIAM, 3600 Market Street, Floor 6, Philadelphia, PA 19104), 1994.
- [11] J. Renegar. *A Mathematical View of Interior-Point Methods in Convex Optimization*. MPS-SIAM Series on Optimization. Society for Industrial and Applied Mathematics, 2001.

- [12] Carsten W Scherer and Siep Weiland. «Linear matrix inequalities in control». In: (2011).
- [13] S. P. Boyd & L. Vandenberghe. *Convex Optimization*. Berichte über verteilte messysteme. Cambridge University Press, 2004.
- [14] J. Nocedal & S. Wright. *Numerical Optimization*. Springer Series in Operations Research and Financial Engineering. Springer New York, 2006.
- [15] S.J. Wright. *Primal-Dual Interior-Point Methods*. Society for Industrial and Applied Mathematics, 1997.
- [16] D. Xue and Y. Chen. *Solving Applied Mathematical Problems with MATLAB*. CRC Press, 2008.
- [17] E. K. P. Chong & S. H. Zak. *An introduction to optimization, 2nd edition*. Wiley India Pvt. Limited, 2010.
- [18] F. S. Hillier & G. J. Lieberman (μετάφραση: Γ. Οικονόμου). *Εισαγωγή στην Επιχειρησιακή Έρευνα*. Εκδόσεις Παπαζήσης, 1985.