Κεφάλαιο 6

Κυρτή βελτιστοποίηση - Συνθήκες βελτιστότητας

Σε αυτό το κεφάλαιο, θα εξάγουμε αναγκαίες και ικανές συνθήκες για τον χαρακτηρισμό βέλτιστων λύσεων προβλημάτων κυρτής βελτιστοποίησης.

Αρχικά, θα αποδείξουμε το Λήμμα του Farkas και θα το χρησιμοποιήσουμε για να αποδείξουμε τις συνθήκες Fritz John.

Κατόπιν, κάνοντας μία επιπλέον υπόθεση, θα αποδείξουμε τις συνθήκες Karush-Kuhn-Tucker.

Η προσέγγισή μας, η οποία είναι κυρίως γεωμετρική, είναι διαφορετική από αυτή των Boyd και Vandenberghe,[1] και βασίζεται σε υλικό από το βιβλίο των Bazaraa, Sherali, Shetty και σημειώσεις της M. Epelman (διαθέσιμες στο διαδίκτυο).

Δείτε επίσης τα Κεφάλαια 10 και 11 του βιβλίου του A. Beck.

Οι συνθήκες βελτιστότητας (FJ, KKT) είναι εξαιρετικά σημαντικές διότι

1. προσφέρουν γεωμετρική ερμηνεία και εμβάθυνση στο πρόβλημα,
2. σημαντικοί αλγόριθμοι βελτιστοποίησης με περιορισμούς, ουσιαστικά, αναζητούν

[1] Η οποία μπορεί να θεωρηθεί ως συμπληρωματική της γεωμετρικής προσέγγισης και αξίζει ανεξάρτητης μελέτης.

το βέλτιστο σημείο μέσω αναζήτησης του σημείου το οποίο ικανοποιεί τις συνθήκες βελτιστότητας.

6.1 Αναγκαίες συνθήκες βελτιστότητας

Έστω το πρόβλημα κυρτής βελτιστοποίησης

$$\begin{array}{ll} \text{minimize} & f_0(\mathbf{x}) \\ \text{subject to} & f_i(\mathbf{x}) \leq 0, \quad i = 1, \ldots, m \\ & h_i(\mathbf{x}) = 0, \quad i = 1, \ldots, p, \end{array} \tag{6.1}$$

με $h_i(\mathbf{x}) = \mathbf{a}_i^T \mathbf{x} - b_i$, για $i = 1, \ldots, p$.

Υποθέτουμε ότι οι συναρτήσεις f_i, για $i = 0, \ldots, m$, είναι διαφορίσιμες, με $f_i : \mathbf{dom} f_i \subseteq \mathbb{R}^n \to \mathbb{R}$, με $\mathbf{dom} f_i$ ανοιχτό κυρτό σύνολο.

Οι περιορισμοί ισοτήτων μπορούν να εκφραστούν στη μορφή

$$\mathbf{A}\mathbf{x} = \mathbf{b}, \tag{6.2}$$

με

$$\mathbf{A} := \begin{bmatrix} \mathbf{a}_1^T \\ \vdots \\ \mathbf{a}_p^T \end{bmatrix}, \quad \mathbf{b} := \begin{bmatrix} b_1 \\ \vdots \\ b_p \end{bmatrix}. \tag{6.3}$$

Έστω $\mathbb{D} := \bigcap_{i=0}^m \mathbf{dom} f_i$.

Το εφικτό σύνολο του προβλήματος (6.1) είναι το

$$\mathbb{X} := \{ \mathbf{x} \in \mathbb{D} \,|\, f_i(\mathbf{x}) \leq 0, \; i = 1, \ldots, m, \; \mathbf{A}\mathbf{x} = \mathbf{b} \}. \tag{6.4}$$

Ορισμός 6.1.1. Έστω $\mathbf{x} \in \mathbb{X}$.

1. Το σύνολο $\mathbb{F}_0 := \{\mathbf{d} \in \mathbb{R}^n \,|\, \nabla f_0(\mathbf{x})^T \mathbf{d} < 0\}$ καλείται κώνος των κατευθύνσεων καθόδου της f_0 στο σημείο $\mathbf{x}$.

2. Το σύνολο $\mathbb{I} := \{i \in \{1, \ldots, m\} \,|\, f_i(\mathbf{x}) = 0\}$ καλείται σύνολο των δεικτών των ενεργών περιορισμών ανισοτήτων του προβλήματος (6.1) στο σημείο $\mathbf{x}$.

3. Το σύνολο $\mathbb{G}_0 := \{\mathbf{d} \in \mathbb{R}^n \,|\, \nabla f_i(\mathbf{x})^T \mathbf{d} < 0, \text{ για } i \in \mathbb{I}\}$ καλείται κώνος των "εσωτερικών" κατευθύνσεων των ενεργών περιορισμών ανισοτήτων του προβλήματος (6.1) στο σημείο $\mathbf{x}$.

4. Το σύνολο $\mathbb{H}_0 := \{\mathbf{d} \in \mathbb{R}^n \,|\, \mathbf{A}\mathbf{d} = \mathbf{0}\}$ καλείται σύνολο των επιτρεπόμενων κατευθύνσεων των περιορισμών ισότητας του προβλήματος (6.1) στο σημείο $\mathbf{x}$.

Το πρώτο σημαντικό αποτέλεσμα για το χαρακτηρισμό βέλτιστων λύσεων του προβλήματος (6.1) έχει ως εξής.

Θεώρημα 6.1.1. Αν $\mathbf{x} \in \mathbb{X}$ είναι βέλτιστο σημείο του προβλήματος (6.1), τότε

$$\mathbb{F}_0 \cap \mathbb{G}_0 \cap \mathbb{H}_0 = \emptyset. \tag{6.5}$$

Απόδειξη. Η απόδειξη θα γίνει με εις άτοπο απαγωγή. Έστω ότι $\mathbf{x} \in \mathbb{X}$ είναι βέλτιστο σημείο για το πρόβλημα (6.1) και $\mathbb{F}_0 \cap \mathbb{G}_0 \cap \mathbb{H}_0 \neq \emptyset$.

Έστω $\mathbf{d} \in \mathbb{F}_0 \cap \mathbb{G}_0 \cap \mathbb{H}_0$ και $\theta > 0$.

Τότε, για όλα τα αρκούντως μικρά θ και $i \in \mathbb{I}$, έχουμε ότι

$$f_i(\mathbf{x} + \theta\mathbf{d}) < f_i(\mathbf{x}) = 0 \text{ και } \mathbf{A}(\mathbf{x} + \theta\mathbf{d}) = \mathbf{b}.$$

Επιπλέον, αφού για $i \notin \mathbb{I}$ έχουμε ότι $f_i(\mathbf{x}) < 0$, τότε, για $i \notin \mathbb{I}$ και αρκετά μικρό θ, εξαιτίας της συνέχειας των f_i, θα έχουμε ότι $f_i(\mathbf{x} + \theta\mathbf{d}) < 0$.

Δηλαδή, για αρκετά μικρό θ, τα σημεία $\mathbf{x} + \theta\mathbf{d}$ είναι εφικτά σημεία του προβλήματος (6.1).

Ταυτόχρονα, όμως, για αρκετά μικρό θ, έχουμε ότι $f_0(\mathbf{x} + \theta\mathbf{d}) < f_0(\mathbf{x})$, το οποίο είναι άτοπο, διότι έχουμε υποθέσει ότι το σημείο $\mathbf{x}$ είναι βέλτιστο. □

Το Θεώρημα 6.1.1 υποδεικνύει ότι αν το σημείο $\mathbf{x} \in \mathbb{X}$ είναι βέλτιστο σημείο του προβλήματος (6.1), τότε δεν υπάρχει αυστηρά εφικτή κατεύθυνση κίνησης $\mathbf{d}$ η οποία να είναι, ταυτόχρονα, και κατεύθυνση μείωσης της f_0.

6.2 Λήμμα του Farkas

Στη συνέχεια, θα μεταφράσουμε τη γεωμετρική συνθήκη (6.5) σε αλγεβρική σχέση ανάμεσα στις βαθμίδες των συναρτήσεων κόστους και περιορισμών.

Αρχικά, αποδεικνύουμε το πιο γνωστό αποτέλεσμα από μία κλάση αποτελεσμάτων που καλούνται **Θεωρήματα των Εναλλακτικών** (Theorems of Alternatives).

Λήμμα 6.2.1. *Farkas' Lemma.* Έστω $\mathbf{A} \in \mathbb{R}^{m\times n}$ και $\mathbf{c} \in \mathbb{R}^n$. Τότε, ακριβώς ένα από τα παρακάτω συστήματα έχει λύση:

$$\begin{aligned} &(1) \quad \mathbf{A}\mathbf{x} \leq \mathbf{0},\ \mathbf{c}^T\mathbf{x} > 0, \\ &(2) \quad \mathbf{A}^T\mathbf{y} = \mathbf{c},\ \mathbf{y} \geq \mathbf{0}. \end{aligned} \tag{6.6}$$

Απόδειξη. Ας υποθέσουμε ότι το σύστημα (2) έχει λύση, δηλαδή, υπάρχει $\mathbf{y} \geq \mathbf{0}$ τέτοιο ώστε $\mathbf{A}^T\mathbf{y} = \mathbf{c}$.

Στη συνέχεια, θεωρούμε το σύστημα (1) και πιο συγκεκριμένα την ανισότητα $\mathbf{A}\mathbf{x} \leq \mathbf{0}$.

Η ανισότητα ικανοποιείται για $\mathbf{x} = \mathbf{0}$. Όμως, αυτή η λύση δεν ικανοποιεί την ανισότητα $\mathbf{c}^T\mathbf{x} > 0$.

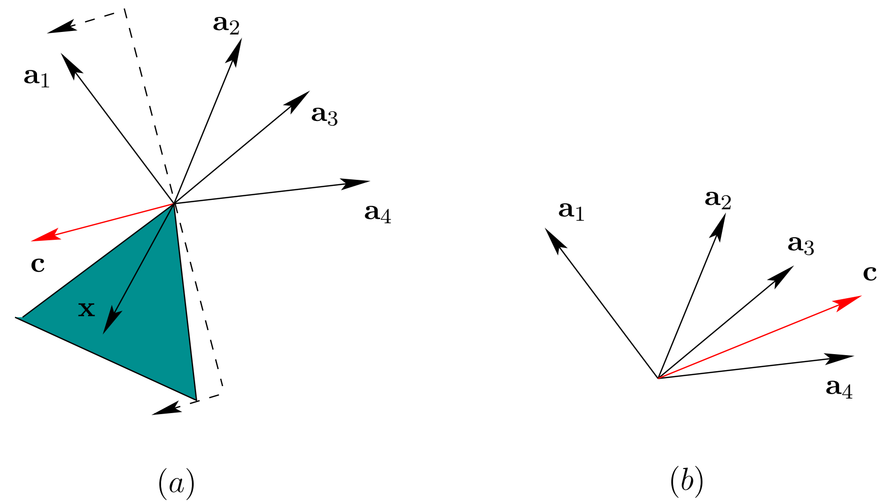
Αν δεν υπάρχει $\mathbf{x} \neq \mathbf{0}$ που να ικανοποιεί αυτή την ανισότητα, τότε το σύστημα (1) δεν έχει λύση.

Αν υπάρχει $\mathbf{x} \neq \mathbf{0}$ το οποίο ικανοποιεί αυτή την ανισότητα, τότε, αφού $\mathbf{y} \geq \mathbf{0}$ και $\mathbf{A}\mathbf{x} \leq \mathbf{0}$, λαμβάνουμε ότι $\mathbf{c}^T\mathbf{x} = \mathbf{y}^T\mathbf{A}\mathbf{x} \leq 0$.

Συνεπώς, και σε αυτή την περίπτωση, το σύστημα (1) δεν έχει λύση.

Στη συνέχεια, υποθέτουμε ότι το σύστημα (2) δεν έχει λύση.

Ορίζουμε το σύνολο $\mathbb{M} := \{\mathbf{x} \in \mathbb{R}^n \,|\, \mathbf{x} = \mathbf{A}^T\mathbf{y}, \mathbf{y} \geq \mathbf{0}\}$.



Σχήμα 6.1: Λήμμα του Farkas (a) το σύστημα (1) έχει λύση (b) το σύστημα (2) έχει λύση - με έντονο χρώμα ο κώνος των δυνατών λύσεων.

Γνωρίζουμε ότι το σύνολο $\mathbb{M}$ είναι κλειστός κυρτός κώνος, διότι αν

$$\mathbf{A} := \begin{bmatrix} \mathbf{a}_1^T \\ \vdots \\ \mathbf{a}_m^T \end{bmatrix}, \tag{6.7}$$

τότε $\mathbb{M} = \{\sum_{i=1}^m y_i \mathbf{a}_i \,|\, y_i \geq 0, \text{ για } i = 1, \ldots, m\}$.

Δηλαδή, το $\mathbb{M}$ είναι η κωνική θήκη των $\mathbf{a}_1, \ldots, \mathbf{a}_m$.

Επιπλέον, από την υπόθεση μη ύπαρξης λύσης για το σύστημα (2), γνωρίζουμε ότι $\mathbf{c} \notin \mathbb{M}$.

Συνεπώς, υπάρχει υπερεπίπεδο το οποίο διαχωρίζει το σημείο $\mathbf{c}$ και το κυρτό σύνολο $\mathbb{M}$.

Δηλαδή, υπάρχουν $\mathbf{p} \in \mathbb{R}^n$ και $\alpha \in \mathbb{R}$ τέτοια ώστε $\mathbf{p}^T\mathbf{c} > \alpha$ και $\mathbf{p}^T\mathbf{x} \leq \alpha$ για κάθε $\mathbf{x} \in \mathbb{M}$. Τότε

1. Αφού $\mathbf{0} \in \mathbb{M}$, έχουμε ότι $0 = \mathbf{p}^T\mathbf{0} \leq \alpha$ και $\mathbf{p}^T\mathbf{c} > \alpha \geq 0$, άρα, $\mathbf{p}^T\mathbf{c} > 0$.

2. Θυμίζουμε ότι για κάθε $\mathbf{x} \in \mathbb{M}$, έχουμε ότι $\alpha \geq \mathbf{p}^T\mathbf{x}$.

Επιπλέον, $\mathbf{x} \in \mathbb{M}$ σημαίνει ότι $\mathbf{x} = \mathbf{A}^T\mathbf{y}$, με $\mathbf{y} \geq \mathbf{0}$. Συνεπώς, για κάθε $\mathbf{y} \geq \mathbf{0}$, έχουμε

$$\alpha \geq \mathbf{p}^T\mathbf{x} = \mathbf{p}^T\mathbf{A}^T\mathbf{y} = \mathbf{y}^T\mathbf{A}\mathbf{p}.$$

Όμως, επειδή το $\mathbf{y} \geq \mathbf{0}$ μπορεί να γίνει αυθαίρετα μεγάλο, η ανισότητα $\alpha \geq \mathbf{y}^T\mathbf{A}\mathbf{p}$, για κάθε $\mathbf{y} \geq \mathbf{0}$, συνεπάγεται ότι $\mathbf{A}\mathbf{p} \leq \mathbf{0}$.

Αυτό συμβαίνει διότι αν κάποιο στοιχείο του $\mathbf{A}\mathbf{p}$ είναι μεγαλύτερο του μηδενός, έστω το $(\mathbf{A}\mathbf{p})_i$, τότε μπορούμε να επιλέξουμε ένα $\mathbf{y}$ το οποίο έχει παντού μηδενικά εκτός από την i-οστή θέση όπου έχει αυθαίρετα μεγάλο y_i, έτσι ώστε να μην ικανοποιείται η ανισότητα $\alpha \geq \mathbf{y}^T\mathbf{A}\mathbf{p}$.

Συνεπώς, κατασκευάσαμε διάνυσμα $\mathbf{p}$ τέτοιο ώστε $\mathbf{A}\mathbf{p} \leq \mathbf{0}$ και $\mathbf{c}^T\mathbf{p} > 0$.

Άρα, βρήκαμε μία λύση για το σύστημα (1), ολοκληρώνοντας την απόδειξη. □

Το παρακάτω αποτέλεσμα είναι μία σημαντική εφαρμογή του Λήμματος Farkas.

Λήμμα 6.2.2. *(Βασικό Λήμμα)* Ακριβώς ένα από τα παρακάτω δύο συστήματα έχει λύση.

$$\begin{aligned} &(1) \quad \mathbf{A}\mathbf{x} < \mathbf{0},\ \mathbf{B}\mathbf{x} \leq \mathbf{0},\ \mathbf{H}\mathbf{x} = \mathbf{0} \\ &(2) \quad \mathbf{A}^T\mathbf{u} + \mathbf{B}^T\mathbf{w} + \mathbf{H}^T\mathbf{v} = \mathbf{0},\ \mathbf{u} \geq \mathbf{0},\ \mathbf{w} \geq \mathbf{0},\ \mathbf{1}^T\mathbf{u} = 1, \end{aligned} \tag{6.8}$$

όπου $\mathbf{1} := (1, \ldots, 1)$.

Απόδειξη. Το σύστημα (1) ισοδύναμα εκφράζεται ως

$$\begin{gathered} \mathbf{A}\mathbf{x} + e\mathbf{1} \leq \mathbf{0},\ e > 0 \\ \mathbf{B}\mathbf{x} \leq \mathbf{0} \\ \mathbf{H}\mathbf{x} \leq \mathbf{0} \\ -\mathbf{H}\mathbf{x} \leq \mathbf{0}. \end{gathered} \tag{6.9}$$

Μία πιο συνεπτυγμένη μορφή του έχει ως εξής:

$$\begin{bmatrix} \mathbf{A} & \mathbf{1} \\ \mathbf{B} & \mathbf{0} \\ \mathbf{H} & \mathbf{0} \\ -\mathbf{H} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ e \end{bmatrix} \leq \mathbf{0}, \quad \begin{bmatrix} 0 & \cdots & 0 & 1 \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ e \end{bmatrix} > 0. \tag{6.10}$$

Το σύστημα (6.10) έχει ακριβώς τη μορφή του συστήματος (1) του Λήμματος 6.2.1.

Το σύστημα το οποίο αντιστοιχεί στο σύστημα (2) του Λήμματος 6.2.1 έχει τη μορφή

$$\begin{bmatrix} \mathbf{A} & \mathbf{1} \\ \mathbf{B} & \mathbf{0} \\ \mathbf{H} & \mathbf{0} \\ -\mathbf{H} & \mathbf{0} \end{bmatrix}^T \begin{bmatrix} \mathbf{u} \\ \mathbf{w} \\ \mathbf{v}_1 \\ \mathbf{v}_2 \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}, \qquad (\mathbf{u}, \mathbf{w}, \mathbf{v}_1, \mathbf{v}_2) \geq \mathbf{0}, \tag{6.11}$$

ή, ισοδύναμα,

$$\mathbf{A}^T\mathbf{u} + \mathbf{B}^T\mathbf{w} + \mathbf{H}^T(\mathbf{v}_1 - \mathbf{v}_2) = \mathbf{0}, \quad \mathbf{1}^T\mathbf{u} = 1, \quad (\mathbf{u}, \mathbf{w}, \mathbf{v}_1, \mathbf{v}_2) \geq \mathbf{0}. \tag{6.12}$$

Θέτοντας $\mathbf{v} := \mathbf{v}_1 - \mathbf{v}_2$, ολοκληρώνουμε την απόδειξη. □

Παρατήρηση: Η σχέση $\mathbf{1}^T\mathbf{u} = 1$, ουσιαστικά, συνεπάγεται ότι $\mathbf{u} \neq 0$. Συνεπώς, στην εκφώνηση του θεωρήματος, η σχέση $\mathbf{1}^T\mathbf{u} = 1$ μπορεί να αντικατασταθεί από τη σχέση $\mathbf{u} \neq 0$.

6.3 Συνθήκες βελτιστότητας πρώτης τάξης

Στη συνέχεια, χρησιμοποιούμε το Λήμμα 6.2.2 και αποδεικνύουμε σημαντικότατες αλγεβρικές συνθήκες βελτιστότητας για το πρόβλημα βελτιστοποίησης (6.1).

Θεώρημα 6.3.1. *(Αναγκαίες συνθήκες Fritz John)* Έστω $\mathbf{x} \in \mathbb{X}$ βέλτιστο σημείο για το πρόβλημα (6.1). Τότε, υπάρχει διάνυσμα $(u_0, \mathbf{u}, \mathbf{v})$ τέτοιο ώστε

$$\begin{aligned} & u_0 \nabla f_0(\mathbf{x}) + \sum_{i=1}^{m} u_i \nabla f_i(\mathbf{x}) + \mathbf{A}^T\mathbf{v} = \mathbf{0}, \\ & (u_0, \mathbf{u}) \geq (0, \mathbf{0}), \ (u_0, \mathbf{u}) \neq (0, \mathbf{0}), \\ & u_i f_i(\mathbf{x}) = 0, \ i = 1, \ldots, m. \end{aligned} \tag{6.13}$$

Απόδειξη. Χωρίς βλάβη της γενικότητας, υποθέτουμε ότι $\mathbb{I} = \{1, \ldots, l\}$ και ορίζουμε

$$\mathcal{A} := \begin{bmatrix} \nabla f_0(\mathbf{x})^T \\ \nabla f_1(\mathbf{x})^T \\ \vdots \\ \nabla f_l(\mathbf{x})^T \end{bmatrix}. \tag{6.14}$$

Αφού το $\mathbf{x} \in \mathbb{X}$ είναι βέλτιστο σημείο για το πρόβλημα (6.1), από το Θεώρημα 6.1.1, συμπεραίνουμε ότι $\mathbb{F}_0 \cap \mathbb{G}_0 \cap \mathbb{H}_0 = \emptyset$.

Δηλαδή, δεν υπάρχει $\mathbf{d}$ το οποίο ικανοποιεί τις σχέσεις $\mathcal{A}\mathbf{d} < \mathbf{0}$ και $\mathbf{Ad} = \mathbf{0}$.

Συνεπώς, από το Λήμμα 6.2.2 για $\mathbf{B} = \mathbf{0}$, συμπεραίνουμε ότι υπάρχουν διανύσματα $(u_0, u_1, \ldots, u_l)$ και $\mathbf{v}$ τέτοια ώστε

$$u_0 \nabla f_0(\mathbf{x}) + \sum_{i=1}^{l} u_i \nabla f_i(\mathbf{x}) + \mathbf{A}^T \mathbf{v} = \mathbf{0}, \tag{6.15}$$

με $u_0 + u_1 + \cdots + u_l = 1$ και $(u_0, u_1, \ldots, u_l) \geq \mathbf{0}$.

Θέτουμε $u_{l+1} = \cdots = u_m = 0$.

Συνεπώς, λαμβάνουμε ότι $(u_0, \mathbf{u}) \geq (0, \mathbf{0})$, $(u_0, \mathbf{u}) \neq (0, \mathbf{0})$ και $u_i f_i(\mathbf{x}) = 0$, για $i = 1, \ldots, m$. Η απόδειξη ολοκληρώθηκε. □

Στη συνέχεια, χρησιμοποιούμε μία επιπλέον υπόθεση και αποδεικνύουμε τις αναγκαίες συνθήκες KKT.

Θεώρημα 6.3.2. *(Αναγκαίες συνθήκες KKT)* Έστω $\mathbf{x} \in \mathbb{X}$ βέλτιστο σημείο του προβλήματος βελτιστοποίησης (6.1). Επιπλέον, έστω ότι οι βαθμίδες όλων των ενεργών περιορισμών του προβλήματος (6.1)[2] στο σημείο $\mathbf{x}$ είναι γραμμικά ανεξάρτητες. Τότε, υπάρχουν διανύσματα $\mathbf{u}$ και $\mathbf{v}$ τέτοια ώστε

$$\begin{gathered} \nabla f_0(\mathbf{x}) + \sum_{i=1}^{m} u_i \nabla f_i(\mathbf{x}) + \mathbf{A}^T \mathbf{v} = \mathbf{0}, \\ \mathbf{u} \geq \mathbf{0}, \quad u_i f_i(\mathbf{x}) = 0, \quad \text{για } i = 1, \ldots, m. \end{gathered} \tag{6.16}$$

[2]Ως ενεργοί περιορισμοί του προβλήματος ορίζονται οι ενεργοί περιορισμοί ανισοτήτων και **όλοι** οι περιορισμοί ισοτήτων.

Απόδειξη. Έχουμε υποθέσει ότι το $\mathbf{x}$ είναι βέλτιστο σημείο του προβλήματος (6.1).

Συνεπώς, θα πρέπει να ικανοποιεί τις συνθήκες Fritz John της (6.13).

Αν $u_0 > 0$, τότε μπορούμε να ξαναορίσουμε τις παραμέτρους των συνθηκών Fritz John ως εξής: $u_i := \frac{u_i}{u_0}$, για $i = 0, \ldots, m$, και $v_i := \frac{v_i}{u_0}$, για $i = 1, \ldots, p$, λαμβάνοντας τις (6.16).

Αν $u_0 = 0$, τότε υπάρχουν $\mathbf{u} \geq \mathbf{0}$, με $\mathbf{u} \neq \mathbf{0}$, και $\mathbf{v}$ τέτοια ώστε

$$\sum_{i \in \mathbb{I}} u_i \nabla f_i(\mathbf{x}) + \sum_{i=1}^{p} v_i \mathbf{a}_i = \mathbf{0}. \tag{6.17}$$

Αυτό, όμως, είναι αδύνατο να συμβεί, διότι έχουμε υποθέσει ότι οι βαθμίδες των ενεργών περιορισμών στο σημείο $\mathbf{x}$ είναι γραμμικά ανεξάρτητες. Η απόδειξη ολοκληρώθηκε. □

Στη συνέχεια, παραθέτουμε ικανές συνθήκες βελτιστότητας.

Θεώρημα 6.3.3. *(Ικανές συνθήκες KKT)* Έστω ότι το σημείο $\mathbf{x} \in \mathbb{X}$ ικανοποιεί τις συνθήκες Karush-Kuhn-Tucker, δηλαδή, για κάποια $\mathbf{u} \in \mathbb{R}^m$ και $\mathbf{v} \in \mathbb{R}^p$, ισχύουν οι σχέσεις

$$\begin{gathered} \nabla f_0(\mathbf{x}) + \sum_{i=1}^{m} u_i \nabla f_i(\mathbf{x}) + \mathbf{A}^T \mathbf{v} = \mathbf{0}, \\ \mathbf{u} \geq \mathbf{0}, \quad u_i f_i(\mathbf{x}) = 0, \quad \text{για } i = 1, \ldots, m. \end{gathered} \tag{6.18}$$

Τότε, το $\mathbf{x}$ είναι βέλτιστο σημείο για το πρόβλημα (6.1).

Απόδειξη. Γνωρίζουμε ότι το εφικτό σύνολο του προβλήματος (6.1) είναι κυρτό.

Έστω $\widehat{\mathbf{x}} \in \mathbb{X}$ σημείο διαφορετικό του $\mathbf{x}$.

Τα σημεία $\theta\widehat{\mathbf{x}} + (1-\theta)\mathbf{x}$, με $0 \leq \theta \leq 1$, είναι εφικτά.

Αυτό σημαίνει ότι, για κάθε $i \in \mathbb{I}$ και $0 \leq \theta \leq 1$,

$$f_i(\theta\widehat{\mathbf{x}} + (1-\theta)\mathbf{x}) = f_i(\mathbf{x} + \theta(\widehat{\mathbf{x}} - \mathbf{x})) \leq 0 = f_i(\mathbf{x}). \tag{6.19}$$

Αφού η τιμή της f_i δεν αυξάνεται όταν κινούμαστε από το $\mathbf{x}$ στην κατεύθυνση $\widehat{\mathbf{x}} - \mathbf{x}$, θα πρέπει, για κάθε $i \in \mathbb{I}$, να ισχύει ότι

$$\nabla f_i(\mathbf{x})^T(\widehat{\mathbf{x}} - \mathbf{x}) \leq 0. \tag{6.20}$$

Επίσης, αφού $\widehat{\mathbf{x}}, \mathbf{x} \in \mathbb{X}$, θα πρέπει να έχουμε ότι

$$\mathbf{A}(\widehat{\mathbf{x}} - \mathbf{x}) = \mathbf{0}. \tag{6.21}$$

Χρησιμοποιώντας τις (6.18), (6.20), και (6.21), λαμβάνουμε ότι, $\forall \widehat{\mathbf{x}} \in \mathbb{X}$,

$$\nabla f_0(\mathbf{x})^T(\widehat{\mathbf{x}} - \mathbf{x}) = -\left(\sum_{i=1}^{m} u_i \nabla f_i(\mathbf{x})^T + \mathbf{v}^T\mathbf{A}\right)(\widehat{\mathbf{x}} - \mathbf{x}) \geq 0. \tag{6.22}$$

Θυμίζουμε ότι μία ικανή και αναγκαία συνθήκη για να αποτελεί το σημείο $\mathbf{x}_*$ λύση στο πρόβλημα $\min_{\mathbf{x}\in\mathbb{X}} f_0(\mathbf{x})$ έχει ως εξής:

$$\nabla f_0(\mathbf{x}_*)^T(\mathbf{y} - \mathbf{x}_*) \geq 0, \ \forall \ \mathbf{y} \in \mathbb{X}.$$

Συνεπώς, από την (6.22) συμπεραίνουμε ότι, για κάθε $\widehat{\mathbf{x}} \in \mathbb{X}$, έχουμε ότι $f_0(\widehat{\mathbf{x}}) \geq f_0(\mathbf{x})$.

Δηλαδή, το $\mathbf{x}$ είναι βέλτιστο σημείο του προβλήματος (6.1). □

6.4 Constraint qualification

Στο Θεώρημα 6.3.2, υποθέσαμε ότι

1. το σημείο $\mathbf{x}$ είναι βέλτιστο για το πρόβλημα (6.1),
2. "κάποια απαίτηση" ικανοποιείται από τους περιορισμούς,

και αποδείξαμε τις αναγκαίες συνθήκες KKT.

Η επιπλέον "απαίτηση," η οποία μας επιτρέπει την απόδειξη των συνθηκών KKT, καλείται constraint qualification.

Έχουμε ήδη αποδείξει ότι αν οι βαθμίδες των ενεργών περιορισμών του προβλήματος (6.1) στο βέλτιστο σημείο $\mathbf{x}$ είναι γραμμικά ανεξάρτητες, τότε ισχύουν οι συνθήκες KKT.

Όμως, η απαίτηση αυτή είναι *τοπική*, διότι αφορά μόνο στο βέλτιστο σημείο $\mathbf{x}$, και μπορεί να ελεγχθεί μόνο αν έχουμε προκαταβολικά υπόψιν κάποιο βέλτιστο σημείο.

Μία εξαιρετικά χρήσιμη *ολική* συνθήκη είναι η εξής.

Ορισμός 6.4.1. *(Συνθήκη Slater)* Έστω το πρόβλημα βελτιστοποίησης (6.1). Η **συνθήκη Slater** (Slater condition) ικανοποιείται αν υπάρχει $\bar{\mathbf{x}} \in \mathbb{X}$ τέτοιο ώστε $f_i(\bar{\mathbf{x}}) < 0$, για $i = 1, \ldots, m$.[3]

Θεώρημα 6.4.1. *(Συνθήκη Slater και αναγκαιότητα συνθηκών KKT)* Έστω ότι η συνθήκη Slater ικανοποιείται για το πρόβλημα βελτιστοποίησης (6.1). Τότε, οι συνθήκες KKT είναι αναγκαίες για τον χαρακτηρισμό του βέλτιστου σημείου το προβλήματος, δηλαδή, αν το $\mathbf{x} \in \mathbb{X}$ είναι βέλτιστο σημείο του προβλήματος (6.1), τότε ικανοποιεί τις συνθήκες KKT.

Απόδειξη. Έστω $\mathbf{x} \in \mathbb{X}$ βέλτιστο σημείο του προβλήματος (6.1). Τότε, ισχύουν οι συνθήκες Fritz John, δηλαδή, υπάρχει διάνυσμα $(u_0, \mathbf{u}, \mathbf{v})$, με $(u_0, \mathbf{u}) \geq \mathbf{0}$ και $(u_0, \mathbf{u}) \neq \mathbf{0}$, τέτοιο ώστε

$$u_0 \nabla f_0(\mathbf{x}) + \sum_{i=1}^{m} u_i \nabla f_i(\mathbf{x}) + \mathbf{A}^T \mathbf{v} = \mathbf{0}, \tag{6.23}$$

με $u_i f_i(\mathbf{x}) = 0$, για $i = 1, \ldots, m$. Αν $u_0 \neq 0$, τότε μπορούμε να διαιρέσουμε με το u_0 και να λάβουμε τις συνθήκες KKT.

Ας υποθέσουμε ότι $u_0 = 0$. Αφού έχουμε υποθέσει ότι ικανοποιείται η συνθήκη Slater, υπάρχει $\bar{\mathbf{x}} \in \mathbb{X}$ τέτοιο ώστε, για κάθε $i \in \mathbb{I}$,

$$0 = f_i(\mathbf{x}) > f_i(\bar{\mathbf{x}}). \tag{6.24}$$

Από την κυρτότητα της f_i, έχουμε ότι

$$f_i(\bar{\mathbf{x}}) \geq f_i(\mathbf{x}) + \nabla f_i(\mathbf{x})^T (\bar{\mathbf{x}} - \mathbf{x}). \tag{6.25}$$

Από τις σχέσεις (6.24) και (6.25), συμπεραίνουμε ότι, για κάθε $i \in \mathbb{I}$, θα πρέπει να ισχύει η ανισότητα

$$\nabla f_i(\mathbf{x})^T (\bar{\mathbf{x}} - \mathbf{x}) < 0. \tag{6.26}$$

[3] Προφανώς, αφού $\bar{\mathbf{x}} \in \mathbb{X}$, θα ικανοποιείται και η σχέση $\mathbf{A}\bar{\mathbf{x}} = \mathbf{b}$.

Επιπλέον, επειδή τα $\bar{\mathbf{x}}$ και $\mathbf{x}$ είναι εφικτά σημεία, έχουμε ότι $\mathbf{A}(\bar{\mathbf{x}} - \mathbf{x}) = \mathbf{0}$. Από την (6.23) και την υπόθεση $u_0 = 0$, λαμβάνουμε ότι κάποιο (ή κάποια) από τα u_i, για $i \in \mathbb{I}$, θα πρέπει να είναι θετικό, συνεπώς

$$\begin{aligned} 0 = \mathbf{0}^T(\bar{\mathbf{x}} - \mathbf{x}) &= \left(\sum_{i=1}^{m} u_i \nabla f_i(\mathbf{x})^T + \mathbf{v}^T \mathbf{A} \right) (\bar{\mathbf{x}} - \mathbf{x}) \\ &= \sum_{i \in \mathbb{I}} u_i \nabla f_i(\mathbf{x})^T (\bar{\mathbf{x}} - \mathbf{x}) \\ &< 0, \end{aligned} \tag{6.27}$$

το οποίο είναι άτοπο.

Άρα, αν ισχύει η συνθήκη Slater, τότε θα πρέπει να έχουμε ότι $u_0 > 0$, ολοκληρώνοντας την απόδειξη. □

6.5 Γεωμετρική ερμηνεία για προβλήματα με περιορισμούς ανισοτήτων

Για να δώσουμε μία γεωμετρική ερμηνεία στις συνθήκες KKT, θεωρούμε το πρόβλημα κυρτής βελτιστοποίησης με μόνο περιορισμούς ανισοτήτων.

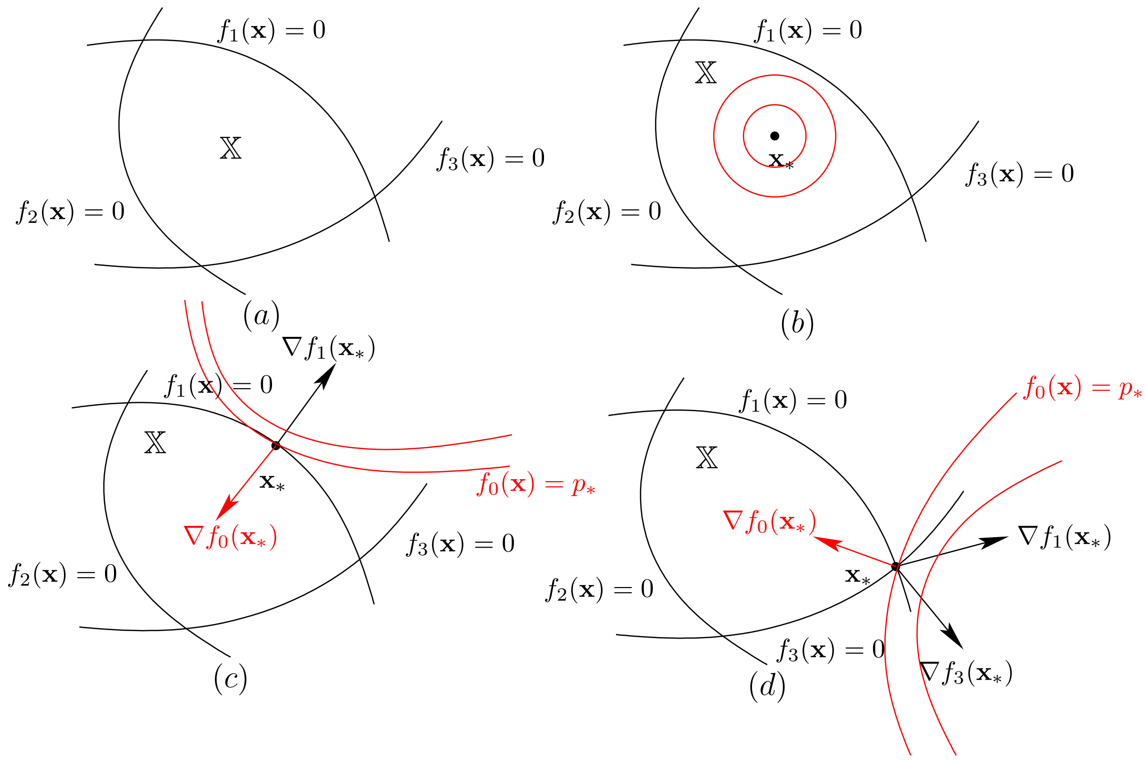
Επιπλέον, υποθέτουμε ότι ισχύει η συνθήκη του Slater (δείτε το Σχήμα 6.2).

Σε αυτή την περίπτωση, οι συνθήκες KKT είναι ικανές και αναγκαίες συνθήκες βελτιστότητας.

Αν $\mathbb{I} = \{1, \ldots, l\}$, τότε το σημείο $\mathbf{x} \in \mathbb{X}$ είναι βέλτιστο αν, και μόνο αν, υπάρχει διάνυσμα $(u_1, \ldots, u_l)$ με $(u_1, \ldots, u_l) \geq \mathbf{0}$, τέτοιο ώστε

$$\sum_{i=1}^{l} u_i \nabla f_i(\mathbf{x}) = -\nabla f_0(\mathbf{x}) \tag{6.28}$$

Δηλαδή, το σημείο $\mathbf{x} \in \mathbb{X}$ είναι βέλτιστο για το πρόβλημα βελτιστοποίησης αν, και μόνο αν, η αρνητική βαθμίδα της f_0 στο σημείο $\mathbf{x}$ ευρίσκεται στον κώνο που παράγουν οι βαθμίδες των ενεργών περιορισμών ανισοτήτων στο σημείο $\mathbf{x}$.



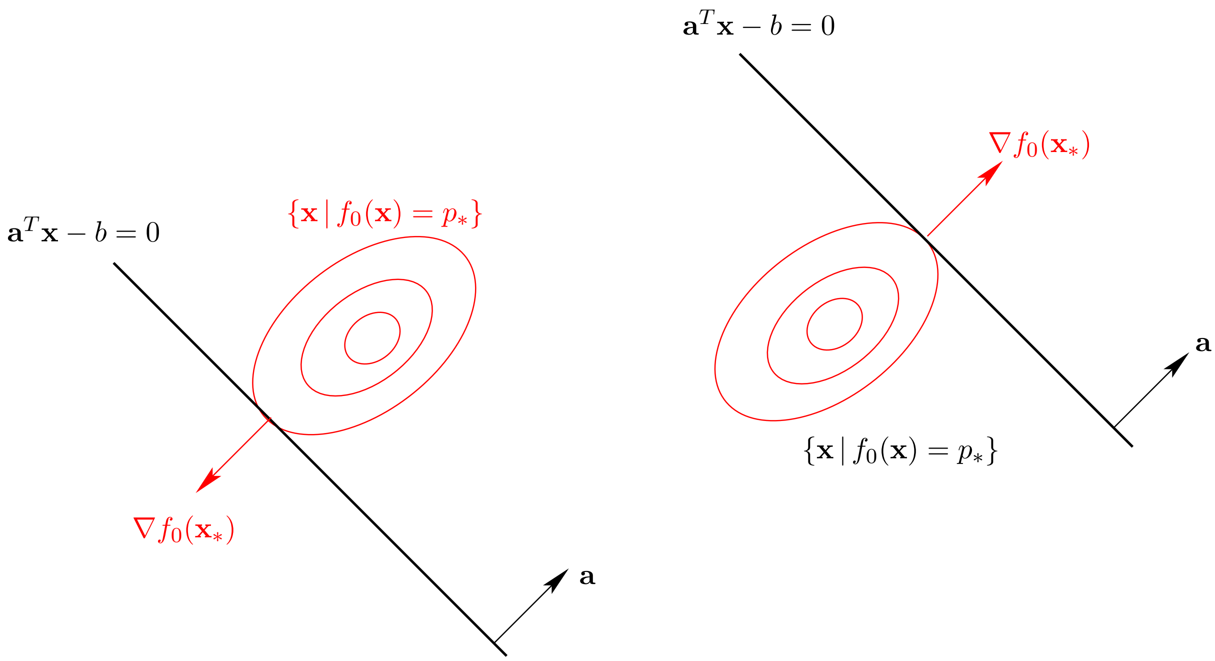
Σχήμα 6.2: Συνθήκες ΚΚΤ. (a) Εφικτό σύνολο περιορισμών ανισοτήτων $\mathbb{X} = \{\mathbf{x} \in \mathbb{R}^n : f_i(\mathbf{x}) \leq 0, i = 1, 2, 3\}$. (b) Βέλτιστο σημείο $\mathbf{x}_*$ στο εσωτερικό του εφικτού συνόλου (c) Βέλτιστο σημείο $\mathbf{x}_*$ με ένα ενεργό περιορισμό (d) Βέλτιστο σημείο $\mathbf{x}_*$ με δύο ενεργούς περιορισμούς.

Στο Σχήμα 6.3, σχεδιάζουμε το πρόβλημα βελτιστοποίησης της κυρτής συνάρτησης $f_0(\mathbf{x})$ υπό τον αφφινικό περιορισμό ισότητας $h_1(\mathbf{x}) = \mathbf{a}^T\mathbf{x} - b = 0$.

Παρατηρήστε ότι η $\nabla f_0(\mathbf{x}_*)$ είναι συγγραμμική με την $\nabla h_1(\mathbf{x}_*)$, αλλά μπορεί να έχει την ίδια ή αντίθετη κατεύθυνση (δηλαδή, $v \in \mathbb{R}$).

6.6 Παραδείγματα

Αρχικά, δίνουμε ένα παράδειγμα όπου οι συνθήκες ΚΚΤ *δεν* ισχύουν.



Σχήμα 6.3: Συνθήκες KKT με ένα γραμμικό περιορισμό για $v_1 > 0$ και $v_1 < 0$.

Παράδειγμα 6.6.1. Έστω το πρόβλημα βελτιστοποίησης

$$\begin{array}{ll} \text{minimize} & x_1 \\ \text{subject to} & (x_1-1)^2 + (x_2-1)^2 \leq 1 \\ & (x_1-1)^2 + (x_2+1)^2 \leq 1 \end{array} \tag{6.29}$$

Το μοναδικό εφικτό σημείο είναι το $\mathbf{x} = (1, 0)$. Να αποδείξετε ότι στο σημείο αυτό ικανοποιούνται οι συνθήκες Fritz John και δεν ικανοποιούνται οι συνθήκες KKT. Προσπαθήστε να καταλάβετε γιατί συμβαίνει αυτό. Ικανοποιείται η συνθήκη Slater; □

Συνεχίζουμε με δύο πολύ σημαντικά παραδείγματα.

Παράδειγμα 6.6.2. Έστω $f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T\mathbf{P}\mathbf{x} + \mathbf{q}^T\mathbf{x}$ και το τετραγωνικό πρόβλημα με αφφινικούς περιορισμούς ισοτήτων

$$\begin{array}{ll} \text{minimize} & f(\mathbf{x}) \\ \text{subject to} & \mathbf{A}\mathbf{x} = \mathbf{b}. \end{array} \tag{6.30}$$

Οι εξισώσεις KKT έχουν ως εξής

$$\begin{aligned} \nabla f(\mathbf{x}) + \mathbf{A}^T\mathbf{v} &= \mathbf{0} \\ \mathbf{A}\mathbf{x} &= \mathbf{b} \end{aligned} \tag{6.31}$$

και μπορούν να εκφραστούν ως το σύστημα γραμμικών εξισώσεων

$$\begin{bmatrix} \mathbf{P} & \mathbf{A}^T \\ \mathbf{A} & \mathbf{O} \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \mathbf{v} \end{bmatrix} = \begin{bmatrix} -\mathbf{q} \\ \mathbf{b} \end{bmatrix}. \tag{6.32}$$

Αν υπάρχει λύση του συστήματος (6.32), έστω $(\mathbf{x}_*, \mathbf{v}_*)$, τότε το $\mathbf{x}_*$ καλείται **πρωτογενές βέλτιστο** (primal optimal) και το $\mathbf{v}_*$ **δυϊκό βέλτιστο** (dual optimal). □

Πιο πολλά για αυτό το πολύ σημαντικό πρόβλημα θα πούμε στη συνέχεια.

Παράδειγμα 6.6.3. Για δεδομένα $\mathbf{x}_k \in \mathbf{dom}\, f$, με $\nabla f(\mathbf{x}_k) \neq \mathbf{0}$, $\mathbf{A} \succ 0$, και $t > 0$, να υπολογίσετε τη λύση του προβλήματος

$$\begin{aligned} &\text{minimize} && \nabla f(\mathbf{x}_k)^T \mathbf{x} \\ &\text{subject to} && (\mathbf{x} - \mathbf{x}_k)^T \mathbf{A} (\mathbf{x} - \mathbf{x}_k) \leq t^2. \end{aligned} \tag{6.33}$$

Το πρόβλημα (6.33) μπορεί να γραφεί ως

$$\begin{aligned} &\text{minimize} && f_0(\mathbf{x}) := \nabla f(\mathbf{x}_k)^T \mathbf{x} \\ &\text{subject to} && f_1(\mathbf{x}) \leq 0, \end{aligned} \tag{6.34}$$

με $f_1(\mathbf{x}) := (\mathbf{x} - \mathbf{x}_k)^T \mathbf{A} (\mathbf{x} - \mathbf{x}_k) - t^2$.

Οι συνθήκες KKT έχουν ως εξής

$$\begin{aligned} &(a) && \nabla f(\mathbf{x}_k) + 2\lambda\, \mathbf{A}(\mathbf{x}_* - \mathbf{x}_k) = \mathbf{0} \\ &(b) && f_1(\mathbf{x}_*) = (\mathbf{x}_* - \mathbf{x}_k)^T \mathbf{A} (\mathbf{x}_* - \mathbf{x}_k) - t^2 \leq 0 \\ &(c) && \lambda \geq 0,\ \ \lambda\, f_1(\mathbf{x}_*) = 0. \end{aligned} \tag{6.35}$$

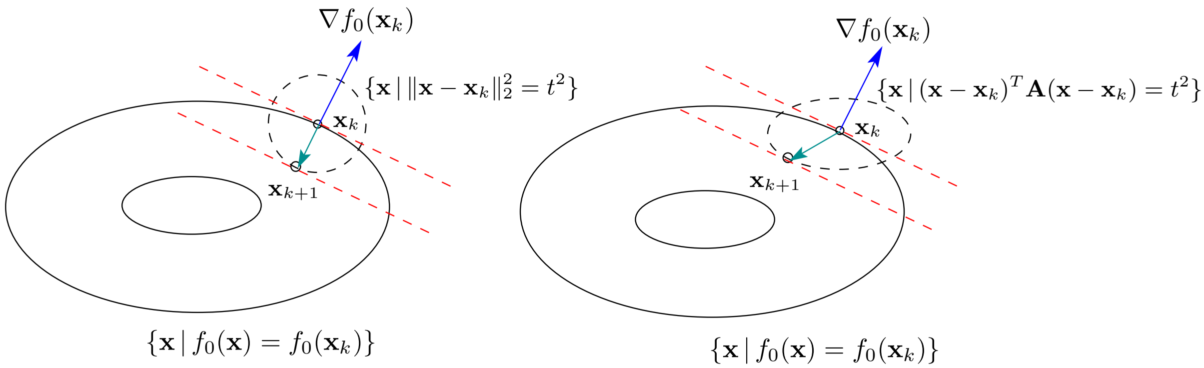
Από την σχέση (a), διαπιστώνουμε ότι, αναγκαστικά, θα πρέπει να έχουμε $\lambda > 0$ (γιατί;). Από αυτό συμπεραίνουμε ότι $f_1(\mathbf{x}_*) = 0$ (γιατί;).

Από την (a), λαμβάνουμε

$$(\mathbf{x}_* - \mathbf{x}_k) = -\frac{1}{2\lambda} \mathbf{A}^{-1} \nabla f(\mathbf{x}_k). \tag{6.36}$$

Θέτοντας αυτή την τιμή στην (b) και λαμβάνοντας υπόψιν ότι $f_1(\mathbf{x}_*) = 0$, έχουμε

$$\begin{aligned} & && \frac{1}{4\lambda^2} \nabla f(\mathbf{x}_k)^T \mathbf{A}^{-1} \nabla f(\mathbf{x}_k) = t^2 \\ &\Longrightarrow && \lambda = \frac{1}{2t} \left(\nabla f(\mathbf{x}_k)^T \mathbf{A}^{-1} \nabla f(\mathbf{x}_k) \right)^{\frac{1}{2}}. \end{aligned} \tag{6.37}$$



Σχήμα 6.4: Λύση προβλήματος (6.33) για $\mathbf{A} = \mathbf{I}$ και $\mathbf{A} = \nabla^2 f_0(\mathbf{x}_k)$.

Αντικαθιστώντας την τιμή του λ στην (6.36), λαμβάνουμε

$$\mathbf{x}_* = \mathbf{x}_k - \frac{t}{(\nabla f(\mathbf{x}_k)^T \mathbf{A}^{-1} \nabla f(\mathbf{x}_k))^{\frac{1}{2}}} \mathbf{A}^{-1} \nabla f(\mathbf{x}_k). \tag{6.38}$$

Παρατηρούμε ότι, όντως, $f_1(\mathbf{x}_*) = 0$.

Επιπλέον, μπορούμε να κάνουμε τις εξής σημαντικές παρατηρήσεις:

1. Θέτοντας $\mathbf{A} = \mathbf{I}$, λαμβάνουμε το βήμα της μεθόδου gradient descent.

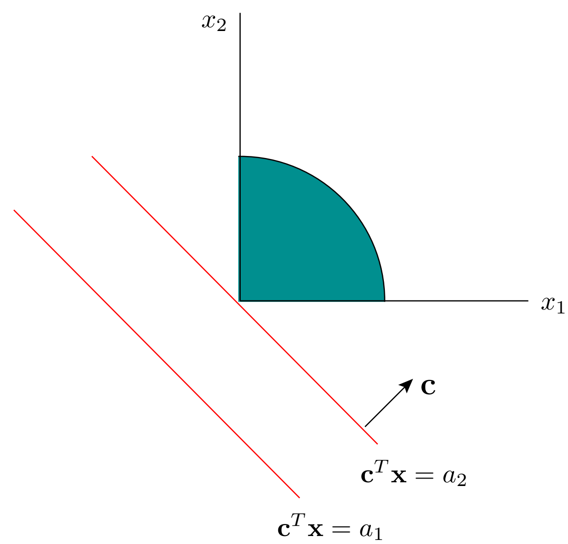
2. Θέτοντας $\mathbf{A} = \nabla^2 f(\mathbf{x}_k)$, λαμβάνουμε το βήμα της μεθόδου Newton.

3. Η λύση θα ήταν ακριβώς η ίδια αν αντί για την $f_0(\mathbf{x}) = \nabla f(\mathbf{x}_k)^T \mathbf{x}$ είχαμε θέσει $f_0(\mathbf{x}) = f(\mathbf{x}_k) + \nabla f(\mathbf{x}_k)^T(\mathbf{x} - \mathbf{x}_k)$ (γιατί;), η οποία είναι η πρώτης τάξης προσέγγιση κατά Taylor της $f(\mathbf{x})$ γύρω από το σημείο $\mathbf{x}_k$.

Συνεπώς, λαμβάνουμε μία νέα ερμηνεία των μεθόδων gradient descent και Newton, για προβλήματα χωρίς περιορισμούς, σαν ελαχιστοποιήσεις της πρώτης τάξης προσέγγισης της συνάρτησης κόστους στο σημείο $\mathbf{x}_k$, υπό περιορισμό ανισότητας κατάλληλα επιλεγμένων μέτρων στο $\mathbb{R}^n$ (δείτε το Σχήμα 6.4). □

Παράδειγμα 6.6.4. Έστω $\mathbb{S} = \{\mathbf{x} \in \mathbb{R}^2 \,|\, x_1^2 + x_2^2 \leq 1, x_1 \geq 0, x_2 \geq 0\}$. Έστω $f_0(\mathbf{x}) = \mathbf{c}^T\mathbf{x}$, με (1) $\mathbf{c} = [1\ 1]^T$ και (2) $\mathbf{c} = [-2\ \ -1]^T$. Να υπολογίσετε το σημείο

$$\mathbf{x}_* = \arg\min_{\mathbf{x} \in \mathbb{S}} f_0(\mathbf{x}).$$



Σχήμα 6.5: Εφικτό σύνολο και ισοϋψείς της $f_0(\mathbf{x})$, με $\mathbf{c} = [1\ 1]^T$, και $a_1 < a_2$.

Παρατηρούμε ότι το $\mathbf{x}_*$ είναι η λύση του κυρτού προβλήματος (δείτε το Σχήμα 6.5)

$$\begin{aligned} \text{minimize} \quad & f_0(\mathbf{x}) = \mathbf{c}^T\mathbf{x}, \\ \text{subject to} \quad & f_1(\mathbf{x}) = x_1^2 + x_2^2 - 1 \leq 0 \\ & f_2(\mathbf{x}) = -x_1 \leq 0 \\ & f_3(\mathbf{x}) = -x_2 \leq 0. \end{aligned} \tag{6.39}$$

Οι συνθήκες KKT για το πρόβλημα αυτό έχουν ως εξής:

$$\begin{aligned} \nabla f_0(\mathbf{x}_*) + \lambda_1 \nabla f_1(\mathbf{x}_*) + \lambda_2 \nabla f_2(\mathbf{x}_*) + \lambda_3 \nabla f_3(\mathbf{x}_*) = \mathbf{0} & \\ \lambda_i \geq 0, \quad i = 1, 2, 3, & \\ f_i(\mathbf{x}_*) \leq 0, \quad i = 1, 2, 3, & \\ \lambda_i f_i(\mathbf{x}_*) = 0, \quad i = 1, 2, 3, & \end{aligned} \tag{6.40}$$

με

$$\nabla f_0(\mathbf{x}_*) = \mathbf{c},\ \nabla f_1(\mathbf{x}_*) = \begin{bmatrix} 2x_{*,1} \\ 2x_{*,2} \end{bmatrix},\ \nabla f_2(\mathbf{x}_*) = \begin{bmatrix} -1 \\ 0 \end{bmatrix},\ \nabla f_3(\mathbf{x}_*) = \begin{bmatrix} 0 \\ -1 \end{bmatrix}.$$

1. Αρχικά, θεωρούμε την περίπτωση όπου $\mathbf{c} = [1\ 1]^T$. Αντικαθιστώντας στην (6.40),

λαμβάνουμε

$$\begin{aligned}
\begin{bmatrix} 2x_{*,1}\lambda_1 - \lambda_2 \\ 2x_{*,2}\lambda_1 - \lambda_3 \end{bmatrix} &= \begin{bmatrix} -1 \\ -1 \end{bmatrix} \\
\lambda_i \geq 0, &\quad i = 1, 2, 3, \\
f_i(\mathbf{x}_*) \leq 0, &\quad i = 1, 2, 3, \\
\lambda_i f_i(\mathbf{x}_*) = 0, &\quad i = 1, 2, 3.
\end{aligned} \tag{6.41}$$

Αν $\lambda_1 > 0$, τότε

$$x_{*,1} = \frac{-1+\lambda_2}{2\lambda_1}, \quad x_{*,2} = \frac{-1+\lambda_3}{2\lambda_1}. \tag{6.42}$$

Για να είναι το $\mathbf{x}_*$ εφικτό, θα πρέπει $x_{*,1} \geq 0$ και $x_{*,2} \geq 0$. Για να συμβαίνει αυτό θα πρέπει $\lambda_2 \geq 1$ και $\lambda_3 \geq 1$.

Επειδή, όμως, θα πρέπει να ισχύει ότι $\lambda_i f_i(\mathbf{x}_*) = 0$, για $i = 2, 3$, θα πρέπει να έχουμε ότι $x_{*,1} = x_{*,2} = 0$.

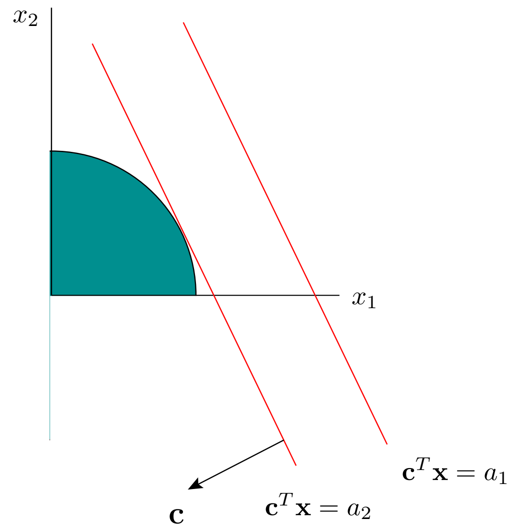
Όμως, σε αυτή την περίπτωση, έχουμε ότι $x_{*,1}^2 + x_{*,2}^2 - 1 < 0$ και δεν ισχύει η σχέση $\lambda_1 f_1(\mathbf{x}_*) = 0$.

Άρα, δεν μπορεί να έχουμε $\lambda_1 > 0$.

Αν $\lambda_1 = 0$, τότε, από την πρώτη εξίσωση της (6.41), λαμβάνουμε ότι $\lambda_2 = 1$ και $\lambda_3 = 1$, το οποίο συνεπάγεται ότι $x_{*,1} = x_{*,2} = 0$. Παρατηρούμε ότι αν θέσουμε αυτές τις τιμές, τότε ικανοποιούμε τις συνθήκες KKT.

Άρα, η βέλτιστη λύση είναι $\mathbf{x}_* = (0, 0)$.

2. Στη συνέχεια, θεωρούμε την περίπτωση όπου $\mathbf{c} = [-2 \ -1]^T$. Από το Σχήμα 6.6, συμπεραίνουμε ότι μόνο ο περιορισμός $f_1(\mathbf{x}) \leq 0$ θα είναι ενεργός στο βέλτιστο



Σχήμα 6.6: Εφικτό σύνολο και ισοϋψείς της $f_0(\mathbf{x})$, με $\mathbf{c} = [-2 \ \ -1]^T$, και $a_1 < a_2$.

σημείο. Συνεπώς, έχουμε $\lambda_2 = \lambda_3 = 0$, και

$$\begin{aligned} \begin{bmatrix} 2x_{*,1}\lambda_1 \\ 2x_{*,2}\lambda_1 \end{bmatrix} &= \begin{bmatrix} 2 \\ 1 \end{bmatrix} \\ \lambda_i &\geq 0, \quad i = 1, 2, 3, \\ f_i(\mathbf{x}_*) &\leq 0, \quad i = 1, 2, 3, \\ \lambda_i f_i(\mathbf{x}_*) &= 0, \quad i = 1, 2, 3. \end{aligned} \tag{6.43}$$

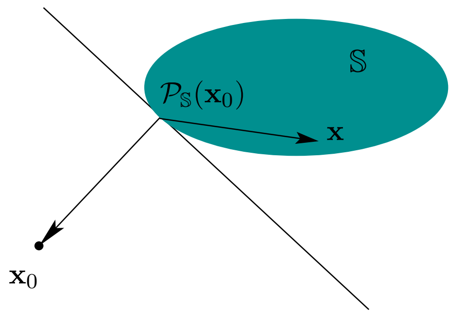
Από την πρώτη εξίσωση, λαμβάνουμε ότι $\lambda_1 = \frac{1}{x_{*,1}} = \frac{1}{2x_{*,2}}$. Συνεπώς, $x_{*,1} = 2x_{*,2}$. Αντικαθιστώντας στη σχέση $f_1(\mathbf{x}_*) = 0$, λαμβάνουμε

$$x_{*,1}^2 + x_{*,2}^2 = 1 \Rightarrow 4x_{*,2}^2 + x_{*,2}^2 = 1 \Rightarrow x_{*,2} = \frac{1}{\sqrt{5}}. \tag{6.44}$$

Άρα, λαμβάνουμε ότι $x_{*,1} = \frac{2}{\sqrt{5}}$, $x_{*,1} = \frac{1}{\sqrt{5}}$, $\lambda_1 = \frac{\sqrt{5}}{2}$, $\lambda_2 = \lambda_3 = 0$. □

6.7 Προβολή σημείου σε κυρτό σύνολο

Στη συνέχεια, θα αναφερθούμε στο εξαιρετικά σημαντικό πρόβλημα της εύρεσης της προβολής σημείου σε κυρτό σύνολο.



Σχήμα 6.7: Προβολή σημείου $\mathbf{x}_0$ σε κυρτό σύνολο.

Έστω κυρτό σύνολο $\mathbb{S} \subset \mathbb{R}^n$ και σημείο $\mathbf{x}_0 \in \mathbb{R}^n$. Η λύση του προβλήματος

$$\begin{aligned} &\text{minimize} \quad f_0(\mathbf{x}) := \|\mathbf{x}_0 - \mathbf{x}\|_2^2 \\ &\text{subject to} \quad \mathbf{x} \in \mathbb{S}, \end{aligned} \tag{6.45}$$

καλείται **προβολή** (projection) του σημείου $\mathbf{x}_0$ στο σύνολο $\mathbb{S}$ και συμβολίζεται με $\mathcal{P}_{\mathbb{S}}(\mathbf{x}_0)$.

Από το Θεώρημα (*), γνωρίζουμε ότι το σημείο $\mathbf{x}_*$ ελαχιστοποιεί την κυρτή συνάρτηση $f(\mathbf{x})$ πάνω στο σύνολο $\mathbb{X}$ αν, και μόνο αν,

$$\nabla f(\mathbf{x}_*)^T (\mathbf{x} - \mathbf{x}_*) \geq 0, \quad \forall \mathbf{x} \in \mathbb{X}. \tag{6.46}$$

Συνεπώς, το σημείο $\mathcal{P}_{\mathbb{S}}(\mathbf{x}_0)$ είναι λύση του προβλήματος (6.45) αν, και μόνο αν,

$$(\mathbf{x}_0 - \mathcal{P}_{\mathbb{S}}(\mathbf{x}_0))^T (\mathbf{x} - \mathcal{P}_{\mathbb{S}}(\mathbf{x}_0)) \leq 0, \quad \forall \mathbf{x} \in \mathbb{S}. \tag{6.47}$$

Στη συνέχεια, θα μελετήσουμε ένα εξαιρετικά σημαντικό παράδειγμα, η λύση του οποίου δίδεται σε κλειστή μορφή.

Παράδειγμα 6.7.1. Να βρεθεί η προβολή του σημείου $\mathbf{x}_0 \in \mathbb{R}^n$ στο κυρτό σύνολο $\mathbb{P} := \{\mathbf{x} \in \mathbb{R}^n \,|\, \mathbf{x} \geq 0\}$.

Το πρόβλημα αυτό γράφεται ως εξής:

$$\begin{aligned} &\text{minimize} \quad f_0(\mathbf{x}) := \tfrac{1}{2}\,\|\mathbf{x}_0 - \mathbf{x}\|_2^2 \\ &\text{subject to} \quad f_i(\mathbf{x}) = -x_i \leq 0, \quad i = 1, \ldots, n. \end{aligned} \tag{6.48}$$

Οι εξισώσεις KKT έχουν ως εξής:

$$\begin{aligned} \nabla f_0(\mathbf{x}_*) + \sum_{i=1}^{n} \lambda_i \nabla f_i(\mathbf{x}_*) = \mathbf{0} \\ \lambda_i \geq 0, \quad i = 1, \ldots, n \\ f_i(\mathbf{x}_*) \leq 0, \quad i = 1, \ldots, n \\ \lambda_i f_i(\mathbf{x}_*) = 0, \quad i = 1, \ldots, n. \end{aligned} \tag{6.49}$$

Ορίζοντας $\boldsymbol{\lambda} := (\lambda_1, \ldots, \lambda_n)$, γράφουμε την πρώτη σχέση της (6.49) ως εξής:

$$\begin{aligned} & (\mathbf{x}_* - \mathbf{x}_0) - \boldsymbol{\lambda} = \mathbf{0} \\ \Longrightarrow \qquad & x_{*,i} = x_{0,i} + \lambda_i, \quad i = 1, \ldots, n. \end{aligned} \tag{6.50}$$

Συνεπώς, για $i = 1, \ldots, n$, έχουμε

$$(a)\ \lambda_i \geq 0, \quad (b)\ x_{*,i} = x_{0,i} + \lambda_i \geq 0, \quad (c)\ \lambda_i x_{*,i} = \lambda_i (x_{0,i} + \lambda_i) = 0. \tag{6.51}$$

Θα εξετάσουμε τις παραπάνω σχέσεις ξεχωριστά για κάθε i.

1. Έστω ότι $x_{0,i} \geq 0$. Τότε, εξαιτίας των (a) και (c), έχουμε ότι $\lambda_i = 0$. Συνεπώς, από την (b) λαμβάνουμε $x_{*,i} = x_{0,i}$.

2. Έστω ότι $x_{0,i} < 0$. Τότε, εξαιτίας της (b), έχουμε ότι $\lambda_i > 0$. Συνεπώς, εξαιτίας της (c), έχουμε ότι $x_{*,i} = x_{0,i} + \lambda_i = 0$.

Άρα, $\mathbb{P}(\mathbf{x}_0) = (\mathbf{x}_0)_+ = \max\{\mathbf{0}, \mathbf{x}_0\}$, όπου ο τελεστής $\max\{\cdot, \cdot\}$ εφαρμόζεται στοιχείο-στοιχείο. □

Άσκηση: Να βρεθεί η προβολή του σημείου $\mathbf{x}_0 \in \mathbb{R}^n$ στο σύνολο

$$\mathbf{B}(\mathbf{0}, r) := \{\mathbf{x} \in \mathbb{R}^n \mid \|\mathbf{x}\|_2 \leq r\}.$$

Παράδειγμα 6.7.2. Έστω το υπερεπίπεδο $\mathbb{H} := \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{a}^T\mathbf{x} = b\}$. Να βρεθεί η απόσταση ενός σημείου $\mathbf{x}_0 \in \mathbb{R}^n$ από το $\mathbb{H}$.

Θεωρούμε το πρόβλημα

$$\begin{aligned} \text{minimize} \quad & f_0(\mathbf{x}) := \tfrac{1}{2} \|\mathbf{x}_0 - \mathbf{x}\|_2^2 \\ \text{subject to} \quad & f_1(\mathbf{x}) = \mathbf{a}^T\mathbf{x} - b = 0. \end{aligned} \tag{6.52}$$

Οι συνθήκες KKT έχουν ως εξής:

$$\begin{aligned} &\nabla f_0(\mathbf{x}_*) + \upsilon_1 \nabla f_1(\mathbf{x}_*) = 0 \Longrightarrow \mathbf{x}_* - \mathbf{x}_0 + \upsilon_1 \mathbf{a} = 0, \\ &\mathbf{a}^T \mathbf{x}_* - b = 0. \end{aligned} \tag{6.53}$$

Πολλαπλασιάζοντας την πρώτη εξίσωση από τα αριστερά με το διάνυσμα $\mathbf{a}^T$, λαμβάνουμε

$$\mathbf{a}^T \mathbf{x}_* - \mathbf{a}^T \mathbf{x}_0 + \upsilon_1 \|\mathbf{a}\|_2^2 = 0 \Longrightarrow \upsilon_1 = \frac{1}{\|\mathbf{a}\|_2^2} \left(\mathbf{a}^T \mathbf{x}_0 - b \right). \tag{6.54}$$

Συνεπώς, η προβολή του $\mathbf{x}_0$ στο $\mathbb{H}$ έχει ως εξής:

$$\mathbf{x}_* = \mathbf{x}_0 + \frac{\mathbf{a}^T \mathbf{x}_0 - b}{\|\mathbf{a}\|_2^2} \mathbf{a}, \tag{6.55}$$

και η απόσταση του $\mathbf{x}_0$ από το $\mathbb{H}$ είναι ίση με

$$\|\mathbf{x}_* - \mathbf{x}_0\|_2 = \frac{|\mathbf{a}^T \mathbf{x}_0 - b|}{\|\mathbf{a}\|_2}. \tag{6.56}$$

Άσκηση: Έστω $\mathbf{A} \in \mathbb{R}^{m \times n}$ και $\mathbf{b} \in \mathbb{R}^m$. Να βρεθεί η προβολή του σημείου $\mathbf{x}_0 \in \mathbb{R}^n$ στο σύνολο

$$\mathbb{S} := \{\mathbf{x} \in \mathbb{R}^n \,|\, \mathbf{A}\mathbf{x} = \mathbf{b}\}.$$

Άσκηση: Έστω $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$ με $\mathbf{a} \leq \mathbf{b}$. Να βρεθεί η προβολή του σημείου $\mathbf{x}_0 \in \mathbb{R}^n$ στο σύνολο

$$\mathbb{S} := \{\mathbf{x} \in \mathbb{R}^n \,|\, \mathbf{a} \leq \mathbf{x} \leq \mathbf{b}\}.$$