Κεφάλαιο 8

Δυϊκότητα

Για κάθε πρόβλημα βελτιστοποίησης υπάρχει ένα άλλο πρόβλημα βελτιστοποίησης το οποίο είναι στενά συνδεδεμένο με το αρχικό.

Το αρχικό καλείται **πρωτογενές** (primal) πρόβλημα ενώ το συνδεδεμένο καλείται **δυϊκό, κατά Lagrange** (Lagrange dual), πρόβλημα.

Υπό κάποιες συνθήκες, συνήθως, κυρτότητας και "καταλληλότητας" των περιορισμών, το πρωταρχικό και το δυϊκό πρόβλημα έχουν την ίδια βέλτιστη τιμή κόστους και, επιπλέον, είναι δυνατό να επιλύσουμε το πρωταρχικό πρόβλημα, μέσω της λύσης του δυϊκού.

8.1 Πρωτογενές και Δυϊκό πρόβλημα

Έστω το πρόβλημα ελαχιστοποίησης

$$\begin{array}{ll} \text{minimize} & f_0(\mathbf{x}) \\ \text{subject to} & f_i(\mathbf{x}) \leq 0, \quad i = 1, ..., m, \\ & h_i(\mathbf{x}) = 0, \quad i = 1, ..., p, \end{array} \tag{8.1}$$

με $\mathbf{x} \in \mathbb{R}^n$.

Έστω $\mathbb{D} := \bigcap_{i=0}^{m} \mathbf{dom}\, f_i \cap \bigcap_{i=1}^{p} \mathbf{dom}\, h_i$.

Το εφικτό σύνολο ορίζεται ως εξής

$$\mathbb{X} := \{\mathbf{x} \in \mathbb{D} \,|\, f_i(\mathbf{x}) \leq 0,\ i = 1, \ldots, m,\ h_i(\mathbf{x}) = 0,\ i = 1, \ldots, p\}. \tag{8.2}$$

Προς το παρόν, δεν θεωρούμε ότι το πρόβλημα (8.1) είναι πρόβλημα κυρτής βελτιστοποίησης.

Υποθέτουμε ότι υπάρχει λύση του προβλήματος, έστω $\mathbf{x}_* \in \mathbb{X}$, με $f_0(\mathbf{x}_*) = p_*$.

Θέτουμε $\boldsymbol{\lambda} := [\,\lambda_1 \,\cdots\, \lambda_m\,]^T$ και $\mathbf{v} := [\,v_1 \,\cdots\, v_p\,]^T$ και ορίζουμε τη **συνάρτηση Lagrange** (Lagrangian) $L : \mathbb{D} \times \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}$, με

$$L(\mathbf{x}, \boldsymbol{\lambda}, \mathbf{v}) = f_0(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i f_i(\mathbf{x}) + \sum_{i=1}^{p} v_i h_i(\mathbf{x}). \tag{8.3}$$

Τα διανύσματα $\boldsymbol{\lambda}$ και $\mathbf{v}$ καλούνται **δυϊκές μεταβλητές** (dual variables) ή **διανύσματα πολλαπλασιαστών Lagrange** (Lagrange multiplier vectors).

Αρχικά, παρατηρούμε ότι

$$\max_{\boldsymbol{\lambda} \geq \mathbf{0}, \mathbf{v}} L(\mathbf{x}, \boldsymbol{\lambda}, \mathbf{v}) = \begin{cases} f_0(\mathbf{x}), & \text{αν } \mathbf{x} \in \mathbb{X}, \\ +\infty, & \text{αν } \mathbf{x} \notin \mathbb{X}. \end{cases} \tag{8.4}$$

Συνεπώς, το πρόβλημα (8.1) είναι ισοδύναμο με το πρόβλημα

$$\min_{\mathbf{x} \in \mathbb{D}} \max_{\boldsymbol{\lambda} \geq \mathbf{0}, \mathbf{v}} L(\mathbf{x}, \boldsymbol{\lambda}, \mathbf{v}). \tag{8.5}$$

Το δυϊκό πρόβλημα είναι το εξής:

$$\max_{\boldsymbol{\lambda} \geq \mathbf{0}, \mathbf{v}} \min_{\mathbf{x} \in \mathbb{D}} L(\mathbf{x}, \boldsymbol{\lambda}, \mathbf{v}). \tag{8.6}$$

Στη συνέχεια, θα μελετήσουμε το δυϊκό πρόβλημα και τη σχέση του με το πρωταρχικό πρόβλημα.

8.2 Η δυϊκή συνάρτηση κατά Lagrange

Ορίζουμε τη **δυϊκή, κατά Lagrange** (Langrange dual), συνάρτηση, ή, απλά, δυϊκή συνάρτηση, $g : \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}$, ως εξής

$$\begin{aligned} g(\boldsymbol{\lambda}, \mathbf{v}) &:= \inf_{\mathbf{x}\in\mathbb{D}} L(\mathbf{x}, \boldsymbol{\lambda}, \mathbf{v}) \\ &= \inf_{\mathbf{x}\in\mathbb{D}} \left(f_0(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i f_i(\mathbf{x}) + \sum_{i=1}^{p} v_i h_i(\mathbf{x}) \right). \end{aligned} \tag{8.7}$$

Όταν σε ένα σημείο $(\boldsymbol{\lambda}, \mathbf{v})$ η συνάρτηση Lagrange δεν είναι φραγμένη από κάτω, τότε θέτουμε $g(\boldsymbol{\lambda}, \mathbf{v}) = -\infty$.

Ορίζουμε το πεδίο ορισμού της g ως εξής:

$$\mathbf{dom}\, g = \{(\boldsymbol{\lambda}, \mathbf{v}) \in \mathbb{R}^m_+ \times \mathbb{R}^p : g(\boldsymbol{\lambda}, \mathbf{v}) > -\infty\}. \tag{8.8}$$

Θεώρημα 8.2.1. Έστω g η δυϊκή, κατά Lagrange, συνάρτηση της (8.7). Τότε

1. Το $\mathbf{dom}\, g$ είναι κυρτό σύνολο.

2. Η συνάρτηση g είναι κοίλη.

Απόδειξη. Αρχικά, αποδεικνύουμε ότι το πεδίο ορισμού της g είναι κυρτό σύνολο. Έστω $(\boldsymbol{\lambda}_1, \mathbf{v}_1), (\boldsymbol{\lambda}_2, \mathbf{v}_2) \in \mathbf{dom} g$. Τότε

$$\begin{aligned} g(\boldsymbol{\lambda}_1, \mathbf{v}_1) &= \inf_{\mathbf{x}\in\mathbb{D}} L(\mathbf{x}, \boldsymbol{\lambda}_1, \mathbf{v}_1) > -\infty, \\ g(\boldsymbol{\lambda}_2, \mathbf{v}_2) &= \inf_{\mathbf{x}\in\mathbb{D}} L(\mathbf{x}, \boldsymbol{\lambda}_2, \mathbf{v}_2) > -\infty. \end{aligned} \tag{8.9}$$

Ας υποθέσουμε ότι $\alpha \in [0, 1]$. Τότε, λαμβάνουμε ότι

$$\begin{aligned} & g(\alpha\boldsymbol{\lambda}_1 + (1-\alpha)\boldsymbol{\lambda}_2, \alpha\mathbf{v}_1 + (1-\alpha)\mathbf{v}_2) \\ &\quad = \inf_{\mathbf{x}\in\mathbb{D}} L(\mathbf{x}, \alpha\boldsymbol{\lambda}_1 + (1-\alpha)\boldsymbol{\lambda}_2, \alpha\mathbf{v}_1 + (1-\alpha)\mathbf{v}_2) \\ &\quad \overset{!}{=} \inf_{\mathbf{x}\in\mathbb{D}} (\alpha L(\mathbf{x}, \boldsymbol{\lambda}_1, \mathbf{v}_1) + (1-\alpha) L(\mathbf{x}, \boldsymbol{\lambda}_2, \mathbf{v}_2)) \\ &\quad \geq \inf_{\mathbf{x}\in\mathbb{D}} \alpha L(\mathbf{x}, \boldsymbol{\lambda}_1, \mathbf{v}_1) + \inf_{\mathbf{x}\in\mathbb{D}} (1-\alpha)\, L(\mathbf{x}, \boldsymbol{\lambda}_2, \mathbf{v}_2) \\ &\quad = \alpha \inf_{\mathbf{x}\in\mathbb{D}} L(\mathbf{x}, \boldsymbol{\lambda}_1, \mathbf{v}_1) + (1-\alpha) \inf_{\mathbf{x}\in\mathbb{D}} L(\mathbf{x}, \boldsymbol{\lambda}_2, \mathbf{v}_2) \\ &\quad = \alpha g(\boldsymbol{\lambda}_1, \mathbf{v}_1) + (1-\alpha) g(\boldsymbol{\lambda}_2, \mathbf{v}_2) \\ &\quad > -\infty. \end{aligned}$$

Άρα, το $\mathbf{dom}\, g$ είναι κυρτό σύνολο.

Κατά τη διαδικασία της απόδειξης, αποδείξαμε και την κοιλότητα της g. $\square$

Θεώρημα 8.2.2. Έστω g η δυϊκή, κατά Lagrange, συνάρτηση της (8.7). Τότε, για κάθε $\boldsymbol{\lambda} \geq \mathbf{0}$ και $\mathbf{v}$, έχουμε

$$g(\boldsymbol{\lambda}, \mathbf{v}) \leq p_*. \tag{8.10}$$

Απόδειξη. Έστω $\bar{\mathbf{x}}$ εφικτό σημείο του προβλήματος (8.1), δηλαδή, $\bar{\mathbf{x}} \in \mathbb{X}$. Τότε

$$\sum_{i=1}^{m} \lambda_i f_i(\bar{\mathbf{x}}) + \sum_{i=1}^{p} v_i h_i(\bar{\mathbf{x}}) \leq 0,$$

διότι κάθε όρος του πρώτου αθροίσματος είναι μη-θετικός αριθμός και κάθε όρος του δεύτερου αθροίσματος ισούται με μηδέν.

Συνεπώς,

$$L(\bar{\mathbf{x}}, \boldsymbol{\lambda}, \mathbf{v}) = f_0(\bar{\mathbf{x}}) + \sum_{i=1}^{m} \lambda_i f_i(\bar{\mathbf{x}}) + \sum_{i=1}^{p} v_i h_i(\bar{\mathbf{x}}) \leq f_0(\bar{\mathbf{x}}) \tag{8.11}$$

και

$$g(\boldsymbol{\lambda}, \mathbf{v}) = \inf_{\mathbf{x} \in \mathbb{D}} L(\mathbf{x}, \boldsymbol{\lambda}, \mathbf{v}) \leq L(\bar{\mathbf{x}}, \boldsymbol{\lambda}, \mathbf{v}) \leq f_0(\bar{\mathbf{x}}). \tag{8.12}$$

Η παραπάνω σχέση ισχύει για κάθε $\bar{\mathbf{x}} \in \mathbb{X}$, άρα ισχύει και για το βέλτιστο σημείο $\mathbf{x}_*$. Συνεπώς

$$g(\boldsymbol{\lambda}, \mathbf{v}) \leq p_*. \tag{8.13}$$

αποδεικνύοντας το θεώρημα. $\square$

Μέσω της δυϊκής συνάρτησης, εξάγουμε ένα μη τετριμμένο κάτω φράγμα για το p_* μόνο όταν $\boldsymbol{\lambda} \geq 0$ και $(\boldsymbol{\lambda}, \mathbf{v}) \in \mathbf{dom}\, g$, δηλαδή, $g(\boldsymbol{\lambda}, \mathbf{v}) > -\infty$.

Τα ζεύγη $(\boldsymbol{\lambda}, \mathbf{v})$ με $\boldsymbol{\lambda} \geq 0$ και $(\boldsymbol{\lambda}, \mathbf{v}) \in \mathbf{dom}\, g$ καλούνται **δυϊκά εφικτά** (dual feasible) σημεία.

Παράδειγμα 8.2.1. Έστω το πρόβλημα βελτιστοποίησης

$$\text{minimize } \mathbf{x}^T\mathbf{x}, \quad \text{subject to } \mathbf{A}\mathbf{x} = \mathbf{b}. \tag{8.14}$$

με $\mathbf{A} \in \mathbb{R}^{p\times n}$. Η συνάρτηση Lagrange είναι η

$$L(\mathbf{x}, \mathbf{v}) = \mathbf{x}^T\mathbf{x} + \mathbf{v}^T(\mathbf{A}\mathbf{x} - \mathbf{b}), \tag{8.15}$$

με $\mathbf{dom}\, L = \mathbb{R}^n \times \mathbb{R}^p$. Η δυϊκή συνάρτηση ορίζεται ως

$$g(\mathbf{v}) := \inf_{\mathbf{x}\in\mathbb{R}^n} L(\mathbf{x}, \mathbf{v}). \tag{8.16}$$

Η $L(\mathbf{x}, \mathbf{v})$ είναι τετραγωνική *κυρτή* συνάρτηση του $\mathbf{x}$.

Συνεπώς, το σημείο $\mathbf{x}$ το οποίο ελαχιστοποιεί την συνάρτηση Lagrange δίδεται από τη λύση της εξίσωσης

$$\nabla_{\mathbf{x}} L(\mathbf{x}, \mathbf{v}) = \mathbf{0} \Rightarrow 2\mathbf{x} + \mathbf{A}^T\mathbf{v} = \mathbf{0} \Rightarrow \mathbf{x} = -\frac{1}{2}\mathbf{A}^T\mathbf{v}. \tag{8.17}$$

Άρα,

$$g(\mathbf{v}) = L\left(-\frac{1}{2}\mathbf{A}^T\mathbf{v}, \mathbf{v}\right) = -\frac{1}{4}\mathbf{v}^T\mathbf{A}\mathbf{A}^T\mathbf{v} - \mathbf{b}^T\mathbf{v}, \tag{8.18}$$

η οποία είναι τετραγωνική *κοίλη* συνάρτηση του $\mathbf{v} \in \mathbb{R}^p$.

Από την ασθενή δυϊκότητα, έχουμε ότι, για κάθε $\mathbf{v} \in \mathbb{R}^p$,

$$-\frac{1}{4}\mathbf{v}^T\mathbf{A}\mathbf{A}^T\mathbf{v} - \mathbf{b}^T\mathbf{v} \leq \inf_{\mathbf{x}\in\mathbb{R}^n} \{\mathbf{x}^T\mathbf{x} \,|\, \mathbf{A}\mathbf{x} = \mathbf{b}\}. \tag{8.19}$$

□

Παράδειγμα 8.2.2. *(Γραμμικό πρόβλημα σε κανονική μορφή)* Έστω το πρόβλημα γραμμικού προγραμματισμού σε κανονική μορφή

$$\begin{aligned} \text{minimize} \quad & \mathbf{c}^T\mathbf{x} \\ \text{subject to} \quad & \mathbf{A}\mathbf{x} = \mathbf{b}, \\ & \mathbf{x} \geq \mathbf{0}. \end{aligned} \tag{8.20}$$

Η ανισότητα $\mathbf{x} \geq \mathbf{0}$ είναι ισοδύναμη με την ανισότητα $-\mathbf{x} \leq \mathbf{0}$, δηλαδή, $-x_i \leq 0$, για $i = 1, \ldots, n$.

Η συνάρτηση Lagrange έχει ως εξής:

$$\begin{aligned} L(\mathbf{x}, \boldsymbol{\lambda}, \mathbf{v}) &= \mathbf{c}^T\mathbf{x} - \sum_{i=1}^{n} \lambda_i x_i + \mathbf{v}^T(\mathbf{A}\mathbf{x} - \mathbf{b}) \\ &= -\mathbf{b}^T\mathbf{v} + (\mathbf{c} + \mathbf{A}^T\mathbf{v} - \boldsymbol{\lambda})^T\mathbf{x}. \end{aligned} \tag{8.21}$$

Η δυϊκή συνάρτηση δίδεται από τη σχέση

$$g(\boldsymbol{\lambda}, \mathbf{v}) := \inf_{\mathbf{x}\in\mathbb{R}^n} L(\mathbf{x}, \boldsymbol{\lambda}, \mathbf{v}) = -\mathbf{b}^T\mathbf{v} + \inf_{\mathbf{x}\in\mathbb{R}^n} \left\{(\mathbf{c} + \mathbf{A}^T\mathbf{v} - \boldsymbol{\lambda})^T\mathbf{x}\right\}. \tag{8.22}$$

Αν δεν υπάρχει περιορισμός για το $\mathbf{x}$, τότε μία γραμμική συνάρτηση του $\mathbf{x}$ είναι φραγμένη από κάτω μόνο αν είναι ταυτοτικά μηδενική.

Συνεπώς,

$$g(\boldsymbol{\lambda}, \mathbf{v}) = \begin{cases} -\mathbf{b}^T\mathbf{v}, & \mathbf{c} + \mathbf{A}^T\mathbf{v} - \boldsymbol{\lambda} = \mathbf{0}, \\ -\infty, & \text{διαφορετικά.} \end{cases} \tag{8.23}$$

Η ανισότητα (8.13) είναι μη τετριμμένη μόνο για ζεύγη $(\boldsymbol{\lambda}, \mathbf{v}) \in \mathbb{R}^m \times \mathbb{R}^p$ τέτοια ώστε $\mathbf{c} + \mathbf{A}^T\mathbf{v} - \boldsymbol{\lambda} = \mathbf{0}$.

Τα ζεύγη αυτά είναι η λύση του συστήματος γραμμικών εξισώσεων

$$\begin{bmatrix} -\mathbf{I} & \mathbf{A}^T \end{bmatrix} \begin{bmatrix} \boldsymbol{\lambda} \\ \mathbf{v} \end{bmatrix} = -\mathbf{c}. \tag{8.24}$$

Συνεπώς, ορίζουν ένα αφφινικό σύνολο.

Για αυτά τα ζεύγη σημείων, η συνάρτηση $-\mathbf{b}^T\mathbf{v}$ παρέχει μη τετριμμένα κάτω φράγματα για τη βέλτιστη τιμή του προβλήματος (8.20).

□

Πιο πολλά παραδείγματα στο βιβλίο των B&V.

8.3 Το δυϊκό κατά Lagrange πρόβλημα

Είδαμε ότι, για κάθε ζεύγος $(\boldsymbol{\lambda}, \mathbf{v})$, με $\boldsymbol{\lambda} \geq \mathbf{0}$, η συνάρτηση $g(\boldsymbol{\lambda}, \mathbf{v})$ αποτελεί ένα κάτω φράγμα για τη βέλτιστη τιμή της συνάρτησης κόστους p_*.

Ένα εξαιρετικά σημαντικό ερώτημα είναι το εξής: *"Ποιο είναι το μεγαλύτερο κάτω φράγμα το οποίο μπορούμε να εξάγουμε απο τη δυϊκή συνάρτηση $g(\boldsymbol{\lambda}, \mathbf{v})$;"*

Για να απαντήσουμε σε αυτό το ερώτημα, ορίζουμε το πρόβλημα

$$\begin{aligned} \text{maximize} \quad & g(\boldsymbol{\lambda}, \mathbf{v}) \\ \text{subject to} \quad & \boldsymbol{\lambda} \geq \mathbf{0}. \end{aligned} \tag{8.25}$$

Το πρόβλημα (8.1) καλείται **πρωτογενές** (primal) πρόβλημα, ενώ το (8.25) καλείται **δυϊκό**, **κατά Lagrange** (Lagrange dual), πρόβλημα.

Αν το πρόβλημα (8.25) έχει λύση, τότε το βέλτιστο σημείο, έστω $(\boldsymbol{\lambda}_*, \mathbf{v}_*)$, καλείται **δυϊκό βέλτιστο** (dual optimal) ή **βέλτιστοι πολλαπλασιαστές Lagrange** (optimal Lagrange multipliers).

Το πρόβλημα (8.25) είναι *πάντα* πρόβλημα κυρτής βελτιστοποίησης διότι ο περιορισμός $\boldsymbol{\lambda} \geq \mathbf{0}$ είναι κυρτός και η συνάρτηση $g(\boldsymbol{\lambda}, \mathbf{v})$ είναι κοίλη.

Παράδειγμα 8.3.1. *(Γραμμικό πρόβλημα σε κανονική μορφή)* Έστω το πρόβλημα γραμμικού προγραμματισμού σε κανονική μορφή

$$\begin{aligned} \text{minimize} \quad & \mathbf{c}^T\mathbf{x} \\ \text{subject to} \quad & \mathbf{A}\mathbf{x} = \mathbf{b}, \\ & \mathbf{x} \geq \mathbf{0}. \end{aligned} \tag{8.26}$$

Όπως είδαμε, η δυϊκή συνάρτηση δίδεται από τη σχέση

$$g(\boldsymbol{\lambda}, \mathbf{v}) = \begin{cases} -\mathbf{b}^T\mathbf{v}, & \mathbf{c} + \mathbf{A}^T\mathbf{v} - \boldsymbol{\lambda} = \mathbf{0}, \\ -\infty, & \text{διαφορετικά.} \end{cases} \tag{8.27}$$

Το δυϊκό πρόβλημα έχει ως εξής

$$\begin{aligned} \text{maximize} \quad & g(\boldsymbol{\lambda}, \mathbf{v}) \\ \text{subject to} \quad & \boldsymbol{\lambda} \geq \mathbf{0}. \end{aligned} \tag{8.28}$$

Η συνάρτηση g παίρνει τιμές διαφορετικές του $-\infty$ μόνο αν $\mathbf{c} + \mathbf{A}^T\mathbf{v} - \boldsymbol{\lambda} = \mathbf{0}$.

Συνεπώς, μία ισοδύναμη έκφραση για το δυϊκό πρόβλημα έχει ως εξής

$$\begin{aligned} \text{maximize} \quad & -\mathbf{b}^T\mathbf{v} \\ \text{subject to} \quad & \mathbf{c} + \mathbf{A}^T\mathbf{v} - \boldsymbol{\lambda} = \mathbf{0}, \\ & \boldsymbol{\lambda} \geq \mathbf{0}. \end{aligned} \tag{8.29}$$

ή, ισοδύναμα,

$$\begin{aligned} \text{maximize} \quad & -\mathbf{b}^T\mathbf{v} \\ \text{subject to} \quad & \mathbf{c} + \mathbf{A}^T\mathbf{v} \geq \mathbf{0}, \end{aligned} \tag{8.30}$$

το οποίο είναι ένα πρόβλημα γραμμικού προγραμματισμού σε μορφή ανισότητας. □

Πιο πολλά παραδείγματα στο βιβλίο των B&V.

8.4 Ασθενής δυϊκότητα

Γνωρίζουμε ότι $g(\boldsymbol{\lambda}, \mathbf{v}) \leq p_*$, για κάθε $\boldsymbol{\lambda} \geq \mathbf{0}$ και $\mathbf{v}$. Αν ορίσουμε

$$d_* := \max_{\boldsymbol{\lambda} \geq \mathbf{0}, \mathbf{v}} g(\boldsymbol{\lambda}, \mathbf{v}), \tag{8.31}$$

τότε λαμβάνουμε ότι

$$d_* \leq p_*. \tag{8.32}$$

Η ανισότητα (8.32) καλείται **ασθενής δυϊκότητα** (weak duality).

Η σχέση (8.32) ισχύει ακόμα και όταν οι τιμές των d_* και p_* είναι $\pm\infty$.

Για παράδειγμα, αν $p_* = -\infty$, δηλαδή, αν το πρωταρχικό πρόβλημα δεν είναι φραγμένο από κάτω, τότε $d_* = -\infty$, δηλαδή, το δυϊκό πρόβλημα δεν είναι εφικτό, και αντίστροφα.

Η ποσότητα $p_* - d_*$ καλείται **βέλτιστο χάσμα δυϊκότητας** (optimal duality gap) και είναι πάντα μη αρνητική.

8.5 Ισχυρή δυϊκότητα

Αν $d_* = p_*$, τότε λέμε ότι έχουμε **ισχυρή δυϊκότητα** (strong duality).

Η ισχυρή δυϊκότητα δεν ισχύει πάντα.

Όμως, αν το πρωτογενές πρόβλημα είναι κυρτό, τότε πολύ συχνά ισχύει.

Για παράδειγμα, αν υπάρχει $\mathbf{x} \in \mathbb{X}$ το οποίο ικανοποιεί τη συνθήκη Slater, τότε μπορεί να αποδειχθεί ότι ισχύει η ισχυρή δυϊκότητα (δείτε το εδάφιο 5.3.2 του βιβλίου των B&V).

8.6 Συνθήκες βελτιστότητας

Στη συνέχεια, υποθέτουμε ότι ισχύει η ισχυρή δυϊκότητα και εξάγουμε συνθήκες βελτιστότητας για το πρωτογενές πρόβλημα.[1]

8.6.1 Συμπληρωματική χαλαρότητα

Έστω $\mathbf{x}_*$ το πρωτογενές βέλτιστο σημείο και $(\boldsymbol{\lambda}_*, \mathbf{v}_*)$, με $\boldsymbol{\lambda}_* \geq \mathbf{0}$, το δυϊκό βέλτιστο σημείο. Τότε

$$\begin{aligned} f_0(\mathbf{x}_*) = g(\boldsymbol{\lambda}_*, \mathbf{v}_*) &= \inf_{\mathbf{x}\in\mathbb{D}} \underbrace{\left(f_0(\mathbf{x}) + \sum_{i=1}^{m} \lambda_{*,i} f_i(\mathbf{x}) + \sum_{i=1}^{p} v_{*,i} h_i(\mathbf{x}) \right)}_{L(\mathbf{x},\boldsymbol{\lambda}_*,\mathbf{v}_*)} \\ &\overset{(a)}{\leq} f_0(\mathbf{x}_*) + \sum_{i=1}^{m} \lambda_{*,i} f_i(\mathbf{x}_*) + \sum_{i=1}^{p} v_{*,i} h_i(\mathbf{x}_*) \\ &\overset{(b)}{\leq} f_0(\mathbf{x}_*), \end{aligned} \tag{8.33}$$

όπου

1. η ανισότητα (a) ισχύει διότι το infimum της $L(\mathbf{x}, \boldsymbol{\lambda}_*, \mathbf{v}_*)$, πάνω σε όλα τα $\mathbf{x} \in \mathbb{D}$, είναι μικρότερο ή ίσο από την τιμή $L(\mathbf{x}_*, \boldsymbol{\lambda}_*, \mathbf{v}_*)$,
2. η ανισότητα (b) ισχύει διότι $\lambda_{*,i} f_i(\mathbf{x}_*) \leq 0$, για $i = 1, \ldots, m$, και $v_{*,i} h_i(\mathbf{x}_*) = 0$, για $i = 1, \ldots, p$.

Με λίγη σκέψη, διαπιστώνουμε ότι οι δύο τελευταίες ανισο-ισότητες της (8.33) ισχύουν ως ισότητες. Συνεπώς, θα πρέπει να ισχύει ότι

$$\sum_{i=1}^{m} \lambda_{*,i} f_i(\mathbf{x}_*) = 0, \tag{8.34}$$

[1] Η προσέγγιση είναι συμπληρωματική αυτής του προηγούμενου κεφαλαίου.

από το οποίο συνεπάγεται ότι

$$\lambda_{*,i} f_i(\mathbf{x}_*) = 0, \quad \text{για } i = 1, ..., m. \tag{8.35}$$

Η σχέση (8.35) καλείται **συμπληρωματική χαλαρότητα** (complementary slackness).

8.6.2 Συνθήκες KKT

Εξαιτίας του ότι η ανισο-ισότητα (a) της σχέσης (8.33) ισχύει ως ισότητα, συμπεραίνουμε ότι το $\mathbf{x}_*$ ελαχιστοποιεί τη συνάρτηση $L(\mathbf{x}, \boldsymbol{\lambda}_*, \mathbf{v}_*)$.

Αν υποθέσουμε ότι οι εμπλεκόμενες κυρτές συναρτήσεις f_i, για $i = 0, \ldots, m$, και h_i, για $i = 1, \ldots, p$, είναι διαφορίσιμες, τότε τα πεδία ορισμού τους $\mathbf{dom}\, f_i$, για $i = 0, \ldots, m$ και και $\mathbf{dom}\, h_i$, για $i = 1, \ldots, p$ είναι ανοιχτά σύνολα, συνεπώς, και το $\mathbb{D}$ είναι ανοιχτό σύνολο.

Παρατηρήστε ότι η συνάρτηση $L(\mathbf{x}, \boldsymbol{\lambda}_*, \mathbf{v}_*)$ είναι διαφορίσιμη συνάρτηση του $\mathbf{x}$. Συνεπώς, στο σημείο $\mathbf{x}_*$ το οποίο την ελαχιστοποιεί θα πρέπει η βαθμίδα της $L(\mathbf{x}, \boldsymbol{\lambda}_*, \mathbf{v}_*)$ να ισούται με το μηδενικό διάνυσμα, δηλαδή,

$$\nabla f_0(\mathbf{x}_*) + \sum_{i=1}^{m} \lambda_{*,i} \nabla f_i(\mathbf{x}_*) + \sum_{i=1}^{p} v_{*,i} \nabla h_i(\mathbf{x}_*) = \mathbf{0}. \tag{8.36}$$

Αν συγκεντρώσουμε όλες τις σχέσεις οι οποίες θα πρέπει να ικανοποιούνται από τα $\mathbf{x}_*$, $\boldsymbol{\lambda}_*$, και $\mathbf{v}_*$, λαμβάνουμε

$$\begin{aligned}
\nabla f_0(\mathbf{x}_*) + \sum_{i=1}^{m} \lambda_{*,i} \nabla f_i(\mathbf{x}_*) + \sum_{i=1}^{p} v_{*,i} \nabla h_i(\mathbf{x}_*) = \mathbf{0}, \\
f_i(\mathbf{x}_*) \leq 0,\ i = 1, ..., m, \\
h_i(\mathbf{x}_*) = 0,\ i = 1, ..., p, \\
\lambda_{*,i} \geq 0,\ i = 1, ..., m, \\
\lambda_{*,i} f_i(\mathbf{x}_*) = 0,\ i = 1, ..., m.
\end{aligned} \tag{8.37}$$

Δηλαδή, λαμβάνουμε τις συνθήκες KKT.

Συνεπώς, για κάθε πρόβλημα βελτιστοποίησης με διαφορίσιμες συναρτήσεις για το οποίο ισχύει η ισχυρή δυϊκότητα, κάθε ζεύγος πρωτογενών και δυϊκών βέλτιστων σημείων πρέπει να ικανοποιεί τις συνθήκες ΚΚΤ.

Δεν θα επεκταθούμε περαιτέρω διότι έχουμε ήδη δει στο προηγούμενο κεφάλαιο πότε οι συνθήκες ΚΚΤ είναι ικανές και αναγκαίες συνθήκες βελτιστότητας κυρτών προβλημάτων βελτιστοποίησης (δείτε το εδάφιο 5.3.3 του βιβλίου B&V).

8.7 Χρησιμότητα δυϊκότητας

Το δυϊκό πρόβλημα είναι πολύ χρήσιμο για πολλούς λόγους. Στη συνέχεια, παρουσιάζουμε συνοπτικά δύο από αυτούς.

8.7.1 Εγγύηση υποβελτιστότητας και κριτήρια τερματισμού αλγορίθμων

Αν $(\bar{\boldsymbol{\lambda}}, \bar{\mathbf{v}})$ είναι εφικτό σημείο για το δυϊκό πρόβλημα και $\bar{\mathbf{x}}$ εφικτό σημείο για το πρωτογενές πρόβλημα, τότε

$$g(\bar{\boldsymbol{\lambda}}, \bar{\mathbf{v}}) \leq p_* \leq f(\bar{\mathbf{x}}). \tag{8.38}$$

Αν υποθέσουμε ότι

$$f(\bar{\mathbf{x}}) - g(\bar{\boldsymbol{\lambda}}, \bar{\mathbf{v}}) = \bar{\epsilon}, \tag{8.39}$$

τότε

$$f(\bar{\mathbf{x}}) - p_* \leq f(\bar{\mathbf{x}}) - g(\bar{\boldsymbol{\lambda}}, \bar{\mathbf{v}}) = \bar{\epsilon}. \tag{8.40}$$

Δηλαδή, αν γνωρίζουμε εφικτό σημείο $\bar{\mathbf{x}}$ και εφικτό δυϊκό σημείο $(\bar{\boldsymbol{\lambda}}, \bar{\mathbf{v}})$ τέτοια ώστε να ισχύει η σχέση (8.39), τότε μπορούμε να συμπεράνουμε ότι το $\bar{\mathbf{x}}$ είναι (το πολύ) $\bar{\epsilon}$–υποβέλτιστο.

Αυτό το συμπέρασμα μπορεί να χρησιμοποιηθεί ως κριτήριο τερματισμού ορισμένων αλγορίθμων βελτιστοποίησης.

Άλλες περιπτώσεις για τις οποίες αυτό το συμπέρασμα είναι εξαιρετικά χρήσιμο είναι οι περιπτώσεις όπου το πρωτογενές πρόβλημα είναι εξαιρετικά δύσκολο να επιλυθεί, συνεπώς, μπορούμε να αισθανθούμε ικανοποιημένοι με "καλές" υποβέλτιστες λύσεις.

8.7.2 Επίλυση πρωτογενούς προβλήματος μέσω του δυϊκού

Σε κάποιες περιπτώσεις (1) η επίλυση του δυϊκού προβλήματος είναι πιο εύκολη από την επίλυση του πρωτογενούς και (2) μπορούμε να επιλύσουμε το πρωτογενές πρόβλημα χρησιμοποιώντας τη λύση του δυϊκού.

Ας δούμε ένα παράδειγμα.

Παράδειγμα 8.7.1. Έστω το πρόβλημα του παραδείγματος 8.2.1

$$\begin{aligned} &\text{minimize} \quad f(\mathbf{x}) = \mathbf{x}^T\mathbf{x} \\ &\text{subject to} \quad \mathbf{A}\mathbf{x} = \mathbf{b}, \end{aligned} \tag{8.41}$$

με $\mathbf{A} \in \mathbb{R}^{p\times n}$ και $\mathbf{rank}(\mathbf{A}) = p$.

Όπως είδαμε, η δυϊκή συνάρτηση δίδεται από τη σχέση

$$g(\mathbf{v}) = -\frac{1}{4}\mathbf{v}^T\mathbf{A}\mathbf{A}^T\mathbf{v} - \mathbf{b}^T\mathbf{v}. \tag{8.42}$$

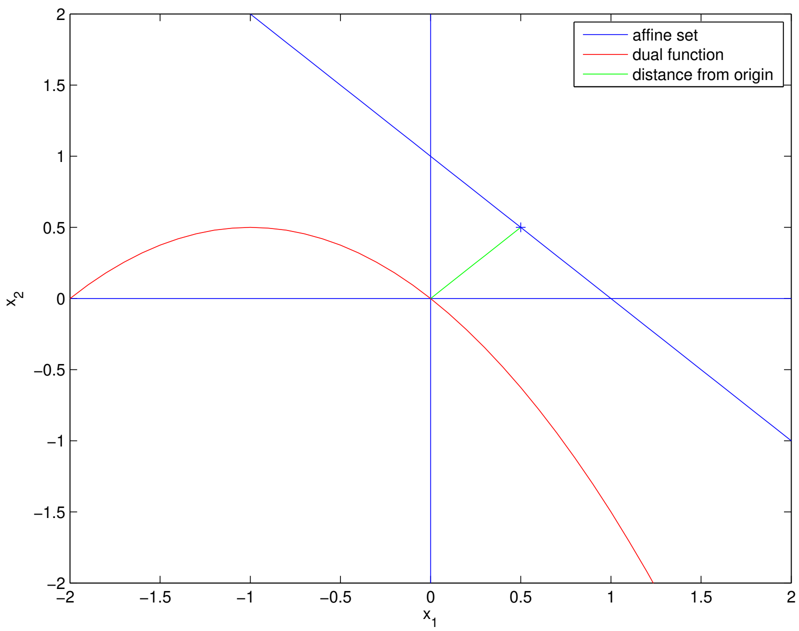
Το δυϊκό πρόβλημα είναι το τετραγωνικό πρόβλημα, χωρίς περιορισμούς,

$$\text{minimize} \quad -g(\mathbf{v}) = \tfrac{1}{4}\mathbf{v}^T\mathbf{A}\mathbf{A}^T\mathbf{v} + \mathbf{b}^T\mathbf{v}. \tag{8.43}$$

Αφού έχουμε υποθέσει ότι οι γραμμές του $\mathbf{A}$ είναι γραμμικά ανεξάρτητες, ο $(p \times p)$ πίνακας $\mathbf{A}\mathbf{A}^T$ είναι αντιστρέψιμος.

Η λύση του δυϊκού προβλήματος ισούται με

$$\mathbf{v}_* = -2\,(\mathbf{A}\mathbf{A}^T)^{-1}\mathbf{b}. \tag{8.44}$$



Σχήμα 8.1: Αφφινικό σύνολο, δυϊκή συνάρτηση και σημείο αφφινικού συνόλου ελάχιστου Ευκλείδειου μέτρου.

Οι συνθήκες KKT είναι αναγκαίες και ικανές για το πρόβλημα αυτό (γιατί;) και δίδονται από τη σχέση

$$\begin{aligned} & \nabla f(\mathbf{x}_*) + \mathbf{A}^T \mathbf{v}_* = \mathbf{0} \\ \Longrightarrow \quad & 2\mathbf{x}_* - 2\mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1}\mathbf{b} = \mathbf{0}. \end{aligned} \tag{8.45}$$

Συνεπώς, η βέλτιστη λύση δίδεται από τη σχέση

$$\mathbf{x}_* = \mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1}\mathbf{b}. \tag{8.46}$$

Δηλαδή, λύσαμε το πρωταρχικό πρόβλημα, **με αφφινικούς περιορισμούς**, λύνοντας το δυϊκό πρόβλημα **χωρίς περιορισμούς**.

Το σημείο αυτό είναι το σημείο του αφφινικού συνόλου $\{\mathbf{x} \in \mathbb{R}^n \,|\, \mathbf{A}\mathbf{x} = \mathbf{b}\}$ με το ελάχιστο Ευκλείδειο μέτρο.

Στο Σχήμα 8.1, σχεδιάζουμε

1. το αφφινικό σύνολο $\mathbf{a}^T\mathbf{x} = b$ στο $\mathbb{R}^2$, για $\mathbf{a}^T = [\,1\ 1\,]$ και $b = 1$,

2. το βέλτιστο σημείο $\mathbf{x}_* = (0.5, 0.5)$, με $f(\mathbf{x}_*) = \mathbf{x}_*^T\mathbf{x}_* = 0.5$,

3. τη δυϊκή συνάρτηση $g(v)$, με μέγιστη τιμή $g(v_*) = g(-1) = 0.5$.

□

8.8 Υπολογισμός Πολλαπλασιαστών Lagrange

Το τελευταίο ερώτημα το οποίο θα μας απασχολήσει σε αυτό το κεφάλαιο είναι το εξής: "*Πώς μπορούμε να υπολογίσουμε τους πολλαπλασιαστές Lagrange;*"

1. Ένας προφανής τρόπος είναι μέσω της λύσης του δυϊκού προβλήματος $\max\limits_{\boldsymbol{\lambda} \geq \mathbf{0}, \mathbf{v}} g(\boldsymbol{\lambda}, \mathbf{v})$.

2. Αν έχουμε λύσει το πρωτεύον πρόβλημα, τότε γνωρίζουμε το βέλτιστο σημείο $\mathbf{x}_*$. Συνεπώς, μπορούμε να αποφανθούμε σχετικά με το ποιοι περιορισμοί ανισοτήτων είναι ενεργοί ή όχι (δηλαδή, για ποια i έχουμε ότι $f_i(\mathbf{x}) = 0$ και $f_i(\mathbf{x}) < 0$). Για τους ανενεργούς περιορισμούς, θέτουμε $\lambda_{*,i} = 0$. Αν $\mathbb{I}$ είναι το σύνολο των δεικτών των ενεργών περιορισμών, τότε έχουμε ότι

$$\nabla f_0(\mathbf{x}_*) + \sum_{i \in \mathbb{I}} \lambda_{*,i} \nabla f_i(\mathbf{x}_*) + \sum_i v_{*,i} \nabla h_i(\mathbf{x}_*) = \mathbf{0} \tag{8.47}$$

το οποίο είναι σύστημα γραμμικών εξισώσεων με αγνώστους τα $\lambda_{*,i}$ και $v_{*,i}$ (προφανώς, θα πρέπει να έχουμε $\lambda_{*,i} \geq 0$).

3. Όπως θα δούμε στη συνέχεια, σημαντικοί αλγόριθμοι βελτιστοποίησης υπολογίζουν *ταυτόχρονα* τα πρωτογενή και τα δυϊκά βέλτιστα σημεία.