Επιχειρησιακή Έρευνα Ι

Μ. Ζαζάνης

Κεφάλαιο 1

Τετραγωνικές μορφές στον $\mathbb{R}^n$ και το θεώρημα του Taylor

Ορισμός 1. *Έστω*

$$A = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix}$$

συμμετρικός πίνακας $n \times n$ με στοιχεία στους πραγματικούς αριθμούς. Η συνάρτηση $\psi : \mathbb{R}^n \to \mathbb{R}$ που ορίζεται ως

$$\psi(x) = x^T A x = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_i x_j \tag{1.1}$$

ονομάζεται τετραγωνική μορφή.

Για παράδειγμα, η τετραγωνική μορφή που αντιστοιχεί στον πίνακα $\begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix}$ είναι η $\psi(x_1, x_2) = x_1^2 - 2x_1x_2 + 2x_2^2$.

Ορισμός 2. *Ένας συμμετρικός πίνακας A, $n \times n$ ονομάζεται*

Θετικά ορισμένος *αν $x^T A x > 0$ για κάθε $x \in \mathbb{R}^n$, $x \neq 0$,*

Θετικά ημιορισμένος *αν $x^T A x \geq 0$ για κάθε $x \in \mathbb{R}^n$ με την ισότητα να ισχύει για κάποιο $x \neq 0$,*

αρνητικά ορισμένος *αν $x^T A x < 0$ για κάθε $x \in \mathbb{R}^n$, $x \neq 0$,*

αρνητικά ημιορισμένος *αν $x^T A x \leq 0$ για κάθε $x \in \mathbb{R}^n$ με την ισότητα να ισχύει για κάποιο $x \neq 0$.*

Θεώρημα 1. *Έστω συμμετρικός πίνακας A, $n \times n$ με πραγματικά στοιχεία. Οι εξής τέσσερεις προτάσεις είναι ισοδύναμες*

(1) Ο πίνακας A είναι θετικά ορισμένος.

(2) *Όλες οι ιδιοτιμές του A είναι θετικές.*

(3) *Οι ορίζουσες όλων των άνω αριστερά υποπινάκων του A είναι θετικές.*

(4) *Όλοι οι οδηγοί d_i, $i = 1, \dots, n$ στην απαλοιφή κατά Gauss χωρίς εναλλαγές γραμμών είναι θετικοί*

Από το παραπάνω θεώρημα προκύπτει ότι η τετραγωνική μορφή (1.1) παίρνει αυστηρά θετικές τιμές για κάθε $x \neq 0$ αν και μόνο αν ο αντίστοιχος πίνακας A είναι θετικά ορισμένος. Αντίστοιχα η τετραγωνική μορφή παίρνει αυστηρά αρνητικές τιμές για κάθε $x \neq 0$ αν και μόνο αν ο πίνακας A είναι αρνητικά ορισμένος.

Ένα αντίστοιχο θεώρημα ισχύει για αρνητικά ορισμένους πίνακες. Συγκεκριμένα, όλες οι ιδιοτιμές του A στην περίπτωση αυτή είναι αρνητικές. Σε ότι αφορά τις ορίζουσες των άνω αριστερά υποπινάκων τα πρόσημα εναλλάσσονται ως εξής:

$$(-1)^k \begin{vmatrix} a_{11} & \cdots & a_{1k} \\ \vdots & & \vdots \\ a_{k1} & \cdots & a_{kk} \end{vmatrix} > 0 \quad \text{για } k = 1, 2, \dots, n.$$

Θεώρημα 2. *(Θεώρημα Taylor) Έστω $f : \mathbb{R}^n \to \mathbb{R}$ μια συνάρτηση με συνεχείς μερικές παραγώγους δεύτερης τάξης. Τότε*

$$f(x+h) = f(x) + \sum_{i=1}^{n} \frac{\partial f(x)}{\partial x_i} h_i + \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\partial^2 f(x)}{\partial x_i \partial x_j} h_i h_j + o(\|h\|^2)$$

Σημείωση: Γενικά θα χρησιμοποιούμε τον συμβολισμό $x = (x_1, \dots, x_n)$ και $\|x\| = \sqrt{x_1^2 + \cdots + x_n^2}$. Ο όρος $o(\|h\|^2)$, το λεγόμενο υπόλοιπο του Taylor, τείνει στο μηδέν γρηγορότερα από το $\|h\|^2$ όταν $h \to 0$. Το θεώρημα του Taylor γράφεται σε διανυσματική μορφή ως

$$f(x+h) = f(x) + \nabla f(x) \cdot h + \frac{1}{2} h^T H(x) h + o(\|h\|^2)$$

όπου ο Εσσιανός πίνακας (Hessian)

$$H(x) := \begin{bmatrix} \frac{\partial^2 f(x)}{\partial x_1^2} & \cdots & \frac{\partial^2 f(x)}{\partial x_1 \partial x_n} \\ \vdots & & \vdots \\ \frac{\partial^2 f(x)}{\partial x_n \partial x_1} & \cdots & \frac{\partial^2 f(x)}{\partial x_n^2} \end{bmatrix}$$

είναι συμμετρικός δεδομένου ότι, όταν οι δεύτερες παράγωγοι είναι συνεχείς (σε ένα ανοικτό χωρίο), $\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial^2 f}{\partial x_j \partial x_i}$.

Ορισμός 3. *Έστω $f : \mathbb{R}^n \to \mathbb{R}$ δύο φορές συνεχώς παραγωγίσιμη. Το σημείο x^0 θα ονομάζεται* κρίσιμο *αν*

$$\nabla f(x^0) = 0.$$

Θεώρημα 3. *Έστω x^0 κρίσιμο σημείο της f. Αν ο πίνακας $H(x^0)$ είναι θετικά ορισμένος τότε το σημείο αυτό θα είναι σημείο τοπικού ελαχίστου ενώ αν είναι αρνητικά ορισμένος θα είναι σημείο τοπικού μεγίστου.*

(Η απόδειξη του θεωρήματος αυτού είναι άμεση συνέπεια του θεωρήματος του Taylor και μας επιτρέπει να βρούμε τα σημεία τοπικού μεγίστου και τοπικού ελαχίστου.)

Πρόβλημα 1. Να βρείτε τα τοπικά ακρότατα (μέγιστα ή ελάχιστα) της συνάρτησης $f(x, y) = x^2y^2e^{-(x+y)^2}$ για $(x, y) \in \mathbb{R}^2$.

1.1 Κυρτότητα

Ορισμός 4. *Ένα υποσύνολο C του $\mathbb{R}^n$ ονομάζεται κυρτό αν για κάθε $x, y \in C$ και $\theta \in [0, 1]$ $\theta x+(1-\theta)y \in C$.*

Ορισμός 5. *Έστω $B \subset \mathbb{R}^n$. Η* κυρτή θήκη *του B είναι το μικρότερο κυρτό σύνολο το οποίο περιέχει το B, $conv(B) := \{\sum_{i=1}^{k} \theta_k x^k : k \in \mathbb{N}, x^i \in B, \theta_i \geq 0, \sum_{i=1}^{k} \theta_i = 1\}$*

Ορισμός 6. *Μια συνάρτηση $f : C \to \mathbb{R}$ ορισμένη πάνω σ' ένα κυρτό σύνολο $C \subset \mathbb{R}^n$ ονομάζεται* κυρτή *(convex) αν, για κάθε $x, y \in C$ και κάθε $\theta \in [0, 1]$*

$$f(\theta x + (1 - \theta)y) \leq \theta f(x) + (1 - \theta)f(y).$$

Θα ονομάζεται κοίλη *(concave) αν*

$$f(\theta x + (1 - \theta)y) \geq \theta f(x) + (1 - \theta)f(y).$$

Από τον παραπάνω ορισμό είναι σαφές ότι αν η συνάρτηση f είναι κυρτή στο C τότε η $-f$ είναι κοίλη. Παρατηρήστε επίσης ότι είναι σκόπιμο να ορίσουμε μια κυρτή (ή κοίλη) συνάρτηση είτε σε ολόκληρο το $\mathbb{R}^n$ ή σε ένα κυρτό υποσύνολό του, μια και στον ορισμό, αν τα x και y είναι στο πεδίο ορισμού της συνάρτησης, τότε θα πρέπει και το $\theta x + (1 - \theta)y$ να ανήκει επίσης στο πεδίο ορισμού της συνάρτησης. Μπορεί να αποδειχθεί ότι αν η f είναι κυρτή (ή κοίλη) συνάρτηση τότε δεν μπορεί να είναι ασυνεχής στο εσωτερικό του πεδίου ορισμού της. Τυχόν ασυνέχειες, αν υπάρχουν, πρέπει να βρίσκονται στο σύνορο του πεδίου ορισμού.

Το ακόλουθο θεώρημα δείχνει την σχέση ανάμεσα στην κυρτότητα και τις μερικές παραγώγους των κυρτών και κοίλων συναρτήσεων.

Θεώρημα 4. *Έστω $f : C \to \mathbb{R}^n$ κυρτή συνάρτηση με συνεχείς μερικές παραγώγους στο C. Η f είναι κυρτή αν και μόνο αν, για κάθε x_0 στο εσωτερικό του C*

$$f(x) - f(x^0) \;\geq\; \nabla f(x^0) \cdot (x - x^0) \;=\; \sum_{i=1}^{n} \frac{\partial f(x^0)}{\partial x_i}(x_i - x_i^0). \tag{1.2}$$

Ομοίως, η f είναι κοίλη αν και μόνο αν,

$$f(x) - f(x^0) \;\leq\; \nabla f(x^0) \cdot (x - x^0) \;=\; \sum_{i=1}^{n} \frac{\partial f(x^0)}{\partial x_i}(x_i - x_i^0). \tag{1.3}$$

Κεφάλαιο 2

Βελτιστοποίηση υπό περιορισμούς ισότητας

Θέλουμε να μεγιστοποιήσουμε την συνάρτηση $f : \mathbb{R}^n \to \mathbb{R}$ κάτω από τους περιορισμούς $g(\mathbf{x}) = 0$ όπου $g : \mathbb{R}^n \to \mathbb{R}$ με $m < n$. Έστω $m, n \in \mathbb{N}$ όπου $m > n$ και f, g_i, $i = 1, 2, \ldots, m$ συναρτήσεις $\mathbb{R}^n \to \mathbb{R}$. Θεωρούμε το πρόβλημα βελτιστοποίησης

$$\begin{aligned} &\min f(\mathbf{x}) \\ &\text{υπό τους περιορισμούς} \\ &g_i(\mathbf{x}) = 0, \quad i = 1, 2, \ldots, m. \end{aligned} \tag{2.1}$$

Έστω $L : \mathbb{R}^{n+m} \to \mathbb{R}$ η συνάρτηση που ορίζεται ως

$$L(\mathbf{x}, \boldsymbol{\lambda}) := f(\mathbf{x}) - \sum_{i=1}^{m} \lambda_i g_i(\mathbf{x}).$$

Η συνάρτηση αυτή ονομάζεται Λαγκρανζιανή. Βρίσκουμε τα κρίσιμα σημεία υπό περιοριμούς ως λύσεις των εξισώσεων

$$\begin{aligned} &\frac{\partial L}{\partial x_j} = 0, \qquad j = 1, 2, \ldots, n \\ &\frac{\partial L}{\partial \lambda_i} = g_i(\mathbf{x}) = 0, \qquad i = 1, 2, \ldots, m. \end{aligned}$$

Τα σημεία μεγίστου βρίσκονται ανάμεσα σ' αυτά τα σημεία.

Παράδειγμα.

$$\begin{aligned} &\min x_1^2 + x_2^2 + x_3^2 \\ &\text{υπό τους περιορισμούς} \\ &x_1 + x_2 + x_3 = 0 \\ &x_1 + 2x_2 + 3x_3 = 1. \end{aligned}$$

Η Λαγκρανζιανή είναι

$$L = x_1^2 + x_2^2 + x_3^2 - \lambda_1(x_1 + x_2 + x_3) - \lambda_2(x_1 + 2x_2 + 3x_3 - 1)$$

και τα στάσιμα σημεία δίνονται από το σύστημα $\partial L/\partial x_j = 0$, $\partial L/\partial \lambda_i = 0$, $j = 1,2,3$, $i = 1,2$.

$$\begin{array}{rrrrrcl}
-2x_1 & & & -\lambda_1 & -\lambda_2 & = & 0 \\
& -2x_2 & & -\lambda_1 & -2\lambda_2 & = & 0 \\
& & -2x_3 & -\lambda_1 & -3\lambda_2 & = & 0 \\
x_1 & +x_2 & +x_3 & & & = & 0 \\
x_1 & +2x_2 & +3x_3 & & & = & 1
\end{array}$$

Το σύστημα αυτό έχει τη μοναδική λύση $x_1 = -1/2, x_2 = 0, x_3 = 1/2$ και $\lambda_1 = 2, \lambda_2 = -1$.

Η γεωμετρική σημασία του θεωρήματος του Lagrange είναι ότι, προκειμένου να είναι το σημείο $x^* = (x_1^*, \ldots, x_n^*)$ κρίσιμο σημείο θα πρέπει η κλίση της f στο σημείο αυτό, $\nabla f(x^*)$ να είναι είναι κάθετη στην υπερεπιφάνεια που ορίζουν οι περιορισμοί g στο ίδιο σημείο. Ο κάθετος αυτός υπόχωρος είναι είναι ο χώρος που δημιουργείται από τα διανύσματα $\nabla g_i(x^*)$, $i = 1, \ldots, m$. Οι γραμμικοί αυτοί συνδιασμοί είναι ο υπόχωρος $\{\sum_{i=1}^m \lambda_i \nabla g_i(x^*) : \lambda_i \in \mathbb{R}\}$.

2.1 Ένα παράδειγμα που εξηγεί τη φυσική σημασία των συνθηκών

Έστω $f, g : \mathbb{R}^2 \to \mathbb{R}$ συναρτήσεις συνεχώς παραγωγίσιμες. Ας υποθέσουμε ότι θέλουμε να μεγιστοποιήσουμε την $f(x,y)$ κάτω από τον περιορισμό $g(x,y) = c$. Μια απλή σκέψη είναι η εξής: η σχέση $g(x,y) = c$ επιτρέπει, τουλάχιστον 'τοπικά', να λύσουμε ως προς y και να εκφράσουμε το y ως συνάρτηση του x και της σταθεράς c. Αυτό δικαιολογείται από το θεώρημα της 'πεπλεγμένης συναρτήσεως'. Υποθέτοντας ότι υπάρχει συνάρτηση ϕ (η οποία εξαρτάται από την τιμή του c) τέτοια ώστε $g(x, \phi(x)) = c$ και παραγωγίζοντας ως προς x έχουμε

$$\frac{\partial g}{\partial x} + \frac{\partial g}{\partial y}\phi'(x) = 0.$$

Η αναγκαία συνθήκη για να έχει το f τοπικό ακρότατο στο σημείο x είναι

$$\frac{df(x,\phi(x))}{dx} = \frac{\partial f}{\partial x} + \frac{\partial f}{\partial y}\phi'(x) = 0.$$

Από τις δύο αυτές σχέσεις, υπό την προϋπόθεση ότι $\frac{\partial f(x,y)}{\partial y} \neq 0$ και $\frac{\partial g}{\partial y} \neq 0$, έχουμε

$$\frac{\frac{\partial g}{\partial x}}{\frac{\partial g}{\partial y}} = \frac{\frac{\partial g}{\partial x}}{\frac{\partial g}{\partial y}} = \phi'(x)$$

ή, ισοδύναμα, υπάρχει $\lambda \in \mathbb{R}$ τέτοιο ώστε

$$\left(\frac{\partial g}{\partial x}, \frac{\partial g}{\partial y}\right) = \lambda\left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}\right), \qquad \text{ή ισοδύναμα}, \qquad \nabla g(x) = \lambda \nabla f(x).$$

Η παραπάνω σχέση εννοείται ότι ισχύει για τα σημεία εκείνα που είναι δεσμευμένα ακρότατα.

2.2 Η οικονομική σημασία των συντελεστών Lagrange

Ας γράψουμε το πρόβλημα μεγιστοποίησης (2.1) στη μορφή

$$\begin{aligned}
&\min f(\mathbf{x}) \\
&\text{υπό τους περιορισμούς} \\
&g_i(\mathbf{x}) = b_i, \quad i = 1, 2, \ldots, m.
\end{aligned} \tag{2.2}$$

Η Λαγκρανζιανή είναι

$$L(\mathbf{x}, \boldsymbol{\lambda}) = f(\mathbf{x}, \boldsymbol{\lambda}) - \sum_{i=1}^{m} \lambda_i (g_i(\mathbf{x}) - b_i)$$

και οι εξισώσεις για τα κρίσιμα σημεία είναι

$$\frac{\partial L}{\partial x_j} = \frac{\partial f}{\partial x_j} - \sum_{i=1}^{m} \lambda_i \frac{\partial g_i}{\partial x_j} = 0, \qquad j = 1, \ldots, n, \tag{2.3}$$

$$g_i(\mathbf{x}) = b_i, \qquad i = 1, \ldots, m. \tag{2.4}$$

Έστω $\mathbf{x}^*$, $\boldsymbol{\lambda}^*$, οι τιμές που μεγιστοποιούν το κριτήριο f και $f(\mathbf{x}^*)$ η βέλτιστη τιμή του κριτηρίου. Όλες αυτές οι ποσήτητες εξαρτώνται βέβαια από τα b_i, $i = 1, \ldots, m$. Το ερώτημα που θέτουμε εδώ είναι πώς αλλάζει η βέλτιστη τιμή του κριτηρίου όταν μεταβληθούν λίγο τα b_i. Για το σκοπό αυτό αυτό υπολογίζουμε τις παραγώγους

$$\frac{\partial f(\mathbf{x}^*)}{\partial b_i} = \sum_{k=1}^{n} \frac{\partial f(\mathbf{x}^*)}{\partial x_k} \frac{\partial x_k^*}{\partial b_i} \tag{2.5}$$

$$\frac{\partial g_k(\mathbf{x}^*)}{\partial b_i} = \sum_{j=1}^{n} \frac{\partial g_k(\mathbf{x}^*)}{\partial x_j} \frac{\partial x_j^*}{\partial b_i} = \delta_{ki} \tag{2.6}$$

όπου

$$\delta_{ki} = \begin{cases} 1 & \text{αν } k = i, \\ 0 & \text{διαφορετικά.} \end{cases}$$

Τέλος, από την (2.3), στις βέλτιστες τιμές $(\mathbf{x}^*, \lambda^*)$,

$$\frac{\partial f(\mathbf{x}^*)}{\partial x_k} = \sum_{l=1}^{m} \lambda_l^* \frac{\partial g_l(\mathbf{x}^*)}{\partial x_k}. \tag{2.7}$$

Από τις (2.5), (2.7), έχουμε

$$\frac{\partial f(\mathbf{x}^*)}{\partial b_i} = \sum_{k=1}^{n} \left(\sum_{l=1}^{m} \lambda_l^* \frac{\partial g_l(\mathbf{x}^*)}{\partial x_k} \right) \frac{\partial x_k^*}{\partial b_i} = \sum_{l=1}^{m} \lambda_l^* \sum_{k=1}^{m} \frac{\partial g_l(\mathbf{x}^*)}{\partial x_k} \frac{\partial x_k^*}{\partial b_i} = \sum_{l=1}^{m} \lambda_l^* \delta_{li} = = \lambda_i^*. \tag{2.8}$$

Συνεπώς, οι τιμές των πολλαπλασιαστών του Lagrange στο βέλτιστο σημείο δείχνουν πόσο θα μεταβαλλόταν οριακά η βέλτιστη τιμή του κριτηρίου αν μεταβαλλόταν οριακά το δεξί μέλος του αντίστοιχου περιορισμού. Αν υποθέσουμε ότι αυξάνουμε το b_i σε $b_i + h$ όπου h μια μικρή ποσότητα, η νέα βέλτιση τιμή του κριτηρίου θα είναι $f(\mathbf{x}^*) + \frac{\partial f(\mathbf{x}^*)}{\partial b_i} h = f(\mathbf{x}^*) + \lambda_i^* h$. Για το λόγο αυτό οι συντελεστές Lagrange υπολογισμένοι στο βέλτιστο ονομάζονται και 'σκιώδεις τιμες' (shadow prices).

Κεφάλαιο 3

Προβλήματα βελτιστοποίησης με ανισοτικούς περιορισμούς

3.1 Συνθήκες Συμπληρωματικής Χαλαρότητας

Θεωρούμε το πρόβλημα βελτιστοποίησης υπό ανισοτικούς περιορισμούς

$$
\begin{aligned}
&\max f(\mathbf{x}) \\
&\text{υπό τους περιορισμούς} \\
&g_i(\mathbf{x}) \leq 0, \quad i = 1, \ldots, m.
\end{aligned}
\tag{3.1}
$$

Η Λαγκρανζιανή είναι

$$L(\mathbf{x}, \lambda) = f(\mathbf{x}) - \sum_{i=1}^{m} \lambda_i g_i(\mathbf{x})$$

και τα κρίσιμα σημεία δίνονται από λύση του συστήματος

$$\frac{\partial L}{\partial x_j} = \frac{\partial f}{\partial x_j} - \sum_{i=1}^{m} \lambda_i \frac{\partial g_i}{\partial x_j} = 0, \qquad j = 1, \ldots, n, \tag{3.2}$$

$$\left\{ \begin{matrix} \lambda_i \geq 0 \\ g_i(\mathbf{x}) = 0 \end{matrix} \right\} \text{ ή } \left\{ \begin{matrix} \lambda_i = 0 \\ g_i(\mathbf{x}) < 0 \end{matrix} \right\}, \qquad i = 1, \ldots, m. \tag{3.3}$$

Οι συνθήκες (3.3) ονομάζονται συνθήκες *συμπληρωματικής χαλαρότητας*. Η έκφραση αυτή σημαίνει ότι, για να είναι ένα σημείο $(\mathbf{x}^*, \boldsymbol{\lambda}^*)$ κρίσιμο σημείο, θα πρέπει (πέραν του να μηδενίζεται η κλίση της Λαγκρανζιανής) για κάθε περιορισμός είτε να είναι χαλαρός, δηλαδή $g_i(\mathbf{x}^*) < 0$, και ο αντίστοιχος πολλαπλασιαστής Lagrange να είναι μηδέν ($\lambda_i^* = 0$) είτε να είναι δεσμευτικός, δηλαδή $g_i(\mathbf{x}^*) = 0$, και $\lambda_i^* \geq 0$.

Παράδειγμα 1: Έστω το πρόβλημα μεγιστοποίησης υπό περιορισμούς

$$\begin{aligned}
&\max f(x_1, x_2) = -x_1^2 - 2x_2^2 + 2x_1 + x_2 \\
&\text{υπό τους περιορισμούς} \\
&x_1 + 3x_2 - 6 \leq 0 \\
&2x_1 + x_2 - 4 \leq 0.
\end{aligned}$$

Η Λαγκρανζιανή είναι

$$L = -x_1^2 - 2x_2^2 + 2x_1 + x_2 - \lambda_1(x_1 + 3x_2 - 6) - \lambda_2(2x_1 + x_2 - 4)$$

και οι συνθήκες για το μέγιστο

$$\begin{aligned}
&\frac{\partial L}{\partial x_1} = -2x_1 + 2 - \lambda_1 - 2\lambda_2 = 0 \\
&\frac{\partial L}{\partial x_2} = -4x_2 + 1 - 3\lambda_1 - \lambda_2 = 0 \\
&\left\{\begin{array}{c} \lambda_1 \geq 0 \\ x_1 + 3x_2 - 6 = 0 \end{array}\right\} \text{ ή } \left\{\begin{array}{c} \lambda_1 = 0 \\ x_1 + 3x_2 - 6 < 0 \end{array}\right\} \\
&\left\{\begin{array}{c} \lambda_2 \geq 0 \\ x_1 + 3x_2 - 6 = 0 \end{array}\right\} \text{ ή } \left\{\begin{array}{c} \lambda_2 = 0 \\ x_1 + 3x_2 - 6 < 0 \end{array}\right\}
\end{aligned}$$

Παράδειγμα 2: Ένα δεύτερο παράδειγμα είναι το ακόλουθο πρόβλημα γραμμικού προγραμματισμού

$$\begin{aligned}
&\max c_1x_1 + \cdots + c_nx_n \\
&\text{υπό τους περιορισμούς} \\
&a_{11}x_1 + \cdots + a_{1n}x_n - b_1 \leq 0 \\
&\quad\vdots \\
&a_{m1}x_1 + \cdots + a_{mn}x_n - b_m \leq 0
\end{aligned}$$

(εδώ δεν υπάρχουν περιορισμοί μη αρντητικότητας για τα x_j.) Η λαγκρανζιανή είναι

$$L = \sum_{j=1}^{n} c_jx_j - \sum_{i=1}^{m} \lambda_i \left(\sum_{j=1}^{n} a_{ij}x_j - b_i \right)$$

και οι συνθήκες είναι

$$\frac{\partial L}{\partial x_j} = c_j - \sum_{i=1}^{m} \lambda_i a_{ij} = 0 \tag{3.4}$$

$$\left\{\begin{array}{c} \lambda_i = 0 \\ \sum_{j=1}^{n} a_{ij}x_j - b_j < 0 \end{array}\right\} \text{ ή } \left\{\begin{array}{c} \lambda_i \geq 0 \\ \sum_{j=1}^{n} a_{ij}x_j - b_j = 0 \end{array}\right\}, \qquad i = 1, \ldots, m. \tag{3.5}$$

Το δυϊκό του γραμμικού προγράμματος είναι το

$$\begin{aligned}
&\max b_1\lambda_1 + \cdots + b_m\lambda_m \\
&\text{υπό τους περιορισμούς} \\
&a_{11}\lambda_1 + \cdots + a_{m1}\lambda_m - c_1 = 0 \\
&\quad\vdots \\
&a_{1n}\lambda_1 + \cdots + a_{mn}\lambda_m - c_n = 0 \\
&\lambda_i \geq 0, \qquad i = 1, \ldots, m.
\end{aligned}$$

Οι συνθήκες (3.4), (3.5) απαιτούν την ύπαρξη εφικτής λύσης $(x_1, \ldots, x_n)$ για το πρωτεύον και $(\lambda_1, \ldots, \lambda_m)$ για το δυϊκό. Από τις σχέσεις αυτές βλέπουμε ότι $\sum_{j=1}^{n} a_{ij}x_n - b_i \leq 0$ και $\lambda_i \geq 0$ συνεπώς

$$\sum_{i=1}^{m} \lambda_i \left(\sum_{j=1}^{n} a_{ij}x_j - b_i \right) \leq 0.$$

Όμως, λόγω της (3.5), αν $\sum_{j=1}^{n} a_{ij}x_j - b_i < 0$ τότε $\lambda_i = 0$ και συνεπώς

$$\sum_{i=1}^{m} \lambda_i \left(\sum_{j=1}^{n} a_{ij}x_j - b_i \right) = 0.$$

ή

$$\sum_{i=1}^{m} \sum_{j=1}^{n} \lambda_i a_{ij} x_j = \sum_{j=1}^{m} \lambda_i b_i. \tag{3.6}$$

Επίσης, $\sum_{i=1}^{m} \lambda_i a_{ij} - c_j = 0$ και συνεπώς $\sum_{j=1}^{n} x_j \left(\sum_{i=1}^{m} \lambda_i a_{ij} - c_j \right) = 0$ ή

$$\sum_{j=1}^{n} \sum_{i=1}^{m} \lambda_i a_{ij} x_j = \sum_{i=1}^{n} c_j x_j. \tag{3.7}$$

Από τις (3.6), (3.7) προκύπτει ότι, όταν ικανοποιούνται οι συνθήκες (3.4), (3.5),

$$\sum_{j=1}^{n} c_j x_j = \sum_{i=1}^{m} \lambda_i b_i$$

που σημαίνει ότι η τιμή του κριτηρίου προς βελτιστοποίηση του πρωτεύοντος είναι ίση με την τιμή του κριτηρίου του δυϊκού. Αυτό όμως συνεπάγεται ότι η $(x_1, \ldots, x_n)$ είναι βέλτιστη λύση για το πρωτεύον και η $(\lambda_1, \ldots, \lambda_m)$ βέλτιστη λύση για το δυϊκό.

3.2 Οι συνθήκες Kuhn–Tucker

Εξετάζουμε τώρα το πρόβλημα μεγιστοποίησης (3.1) υπό τον επιπλέον περιορισμο της μη αρνητικότητας όλων των μεταβλητών, δηλαδή

$$\begin{aligned}
&\max f(\mathbf{x}) \\
&\text{υπό τους περιορισμούς} \\
&g_i(\mathbf{x}) \leq 0, \quad i = 1, \ldots, m. \\
&x_j \geq 0, \quad j = 1, \ldots, m.
\end{aligned} \tag{3.8}$$

Το πρόβλημα αυτό δεν είναι ουσιαστικά διαφορετικό από το (3.1) αφού μπορεί να τεθεί στην ίδα μορφή, όμως εδώ σκοπεύουμε να χειριστούμε τους περιορισμούς μη αρνητικότητας με διαφορετικό τρόπο. Η λαγκρανζιανή είναι

$$L(\mathbf{x}, \boldsymbol{\lambda}) = f(\mathbf{x}) - \sum_{i=1}^{m} \lambda_i g_i(\mathbf{x}) + \sum_{j=1}^{n} \mu_j x_j$$

(Οι πολλαπλασιαστές Lagrange μ_j στον τελευταίο όρο έχουν αρνητικό πρόσημο επειδή οι περιορισμοί μη αρνητικότητας γράφονται ως $-x_j \leq 0$.) Θέτοντας

$$\tilde{L}(\mathbf{x}, \boldsymbol{\lambda}) = f(\mathbf{x}) - \sum_{i=1}^{m} \lambda_i g_i(\mathbf{x}),$$

οι συνθήκες για το βέλτιστο σημείο γράφονται ως

$$\frac{\partial L}{\partial x_j} = \frac{\partial \tilde{L}}{\partial x_j} + \mu_j = 0, \qquad j = 1, \ldots, n, \tag{3.9}$$

$$\left\{ \begin{matrix} \frac{\partial \tilde{L}}{\partial \lambda_i} = 0 \\ \lambda_i \geq 0 \end{matrix} \right\} \text{ ή } \left\{ \begin{matrix} \frac{\partial \tilde{L}}{\partial \lambda_i} > 0 \\ \lambda_i = 0 \end{matrix} \right\}, \qquad i = 1, \ldots, m, \tag{3.10}$$

(επειδή $\frac{\partial L}{\partial \lambda_i} = \frac{\partial \tilde{L}}{\partial \lambda_i}$) και

$$\left\{ \begin{matrix} \frac{\partial L}{\partial \mu_j} = 0 \\ \mu_j \geq 0 \end{matrix} \right\} \text{ ή } \left\{ \begin{matrix} \frac{\partial L}{\partial \mu_j} > 0 \\ \mu_j = 0 \end{matrix} \right\}, \qquad i = 1, \ldots, m. \tag{3.11}$$

Δεδομένου ότι $\frac{\partial L}{\partial \mu_j} = x_j$ η συνθήκη (3.11) γράφεται ως

$$\left\{ \begin{matrix} x_j = 0 \\ \mu_j \geq 0 \end{matrix} \right\} \text{ ή } \left\{ \begin{matrix} x_j > 0 \\ \mu_j = 0 \end{matrix} \right\}, \qquad j = 1, \ldots, n. \tag{3.12}$$

Λαμβάνοντας υπ΄ όψιν τα παραπάνω μπορούμε να γράψουμε τις συνθήκες για το βέλτιστο σημείο με την ακόλουθη συμμετρική μορφή:

$$\left\{ \begin{matrix} \frac{\partial \tilde{L}}{\partial x_j} = 0 \\ x_j > 0 \end{matrix} \right\} \text{ ή } \left\{ \begin{matrix} \frac{\partial \tilde{L}}{\partial x_j} \leq 0 \\ x_j = 0 \end{matrix} \right\}, \qquad j = 1, \ldots, n, \tag{3.13}$$

και

$$\left\{ \begin{matrix} \frac{\partial \tilde{L}}{\partial \lambda_i} = 0 \\ \lambda_i \geq 0 \end{matrix} \right\} \text{ ή } \left\{ \begin{matrix} \frac{\partial \tilde{L}}{\partial \lambda_i} > 0 \\ \lambda_i = 0 \end{matrix} \right\}, \qquad i = 1, \ldots, m, \tag{3.14}$$

(Η συνθήκη (3.13) προκύπτει από τον συνδιασμό της (3.9) και της (3.12).)

3.3 Οι συνθήκες Kuhn–Tucker όταν υπάρχουν και ισωτικοί περιορισμοί

Η περίπτωση αυτή εμπεριέχεται στην γενική περίπτωση που μελετήσαμε δεδομένου ότι ο περιορισμός $h(\mathbf{x}) = 0$ μπορεί να γραφεί και ως ένα ζεύγος ανισωτικών περιορισμών, $h(\mathbf{x}) \leq 0$ και $-h(\mathbf{x}) \leq 0$. Ας υποθέσουμε λοιπόν ότι έχουμε το πρόβλημα

$$\begin{aligned} &\max f(\mathbf{x}) \\ &\text{υπό τους περιορισμούς} \\ &g_i(\mathbf{x}) \leq 0, \quad i = 1, \ldots, m \\ &h_l(\mathbf{x}) = 0, \quad l = r, \ldots, m \\ &x_j \geq 0, \quad j = 1, \ldots, m. \end{aligned} \tag{3.15}$$

Αυτό μετατρέπεται στο

$$\begin{aligned} &\max f(\mathbf{x}) \\ &\text{υπό τους περιορισμούς} \\ &g_i(\mathbf{x}) \leq 0, \quad i = 1, \ldots, m \\ &h_l(\mathbf{x}) \leq 0, \quad l = r, \ldots, m \\ &-h_l(\mathbf{x}) \leq 0, \quad l = r, \ldots, m \\ &x_j \geq 0, \quad j = 1, \ldots, m. \end{aligned} \tag{3.16}$$

με λαγκρανζιανή

$$\tilde{L}(\mathbf{x}, \boldsymbol{\lambda}) = f(\mathbf{x}) - \sum_{i=1}^{m} \lambda_i g_i(\mathbf{x}) - \sum_{l=1}^{r} \nu_l^+ h_l(\mathbf{x}) + \sum_{l=1}^{r} \nu_l^- h_l(\mathbf{x}).$$

Οι συνθήκες είναι

$$\left\{ \begin{array}{c} \frac{\partial \tilde{L}}{\partial x_j} = 0 \\ x_j > 0 \end{array} \right\} \text{ ή } \left\{ \begin{array}{c} \frac{\partial \tilde{L}}{\partial x_j} \leq 0 \\ x_j = 0 \end{array} \right\}, \qquad j = 1, \ldots, n,$$

και

$$\left\{ \begin{array}{c} \frac{\partial \tilde{L}}{\partial \lambda_i} = 0 \\ \lambda_i \geq 0 \end{array} \right\} \text{ ή } \left\{ \begin{array}{c} \frac{\partial \tilde{L}}{\partial \lambda_i} > 0 \\ \lambda_i = 0 \end{array} \right\}, \qquad i = 1, \ldots, m,$$

και

$$\left\{ \begin{array}{c} \frac{\partial \tilde{L}}{\partial \nu_l^+} = 0 \\ \nu_l^+ \geq 0 \end{array} \right\} \text{ ή } \left\{ \begin{array}{c} \frac{\partial \tilde{L}}{\partial \nu_l^+} > 0 \\ \nu_l^+ = 0 \end{array} \right\}, \qquad l = 1, \ldots, r,$$

και

$$\left\{ \begin{array}{c} \frac{\partial \tilde{L}}{\partial \nu_l^-} = 0 \\ \nu_l^- \geq 0 \end{array} \right\} \text{ ή } \left\{ \begin{array}{c} \frac{\partial \tilde{L}}{\partial \nu_l^-} > 0 \\ \nu_l^- = 0 \end{array} \right\}, \qquad l = 1, \ldots, r,$$

Αν λάβουμε υπ' όψιν μας ότι $\frac{\partial \tilde{L}}{\partial \nu_l^+} = -\frac{\partial \tilde{L}}{\partial \nu_l^-}$ βλέπουμε ότι ζεύγη ανισωτήτων της μορφής $\frac{\partial \tilde{L}}{\partial \nu_l^+} > 0$ και $\frac{\partial \tilde{L}}{\partial \nu_l^-} > 0$ ή $\frac{\partial \tilde{L}}{\partial \nu_l^+} > 0$ και $\frac{\partial \tilde{L}}{\partial \nu_l^-} = 0$ κ.λπ. είναι ανέφικτα. Συνεπώς οι εξισώσεις που προκύπτουν από τους ισοτικούς περιορισμούς παίρνουν την μορφή

$$\left\{\begin{array}{l} \frac{\partial \tilde{L}}{\partial \nu_l^+} = 0 \\ \nu_l^+ \geq 0 \end{array}\right\}, \qquad l = 1, \ldots, r,$$

και

$$\left\{\begin{array}{l} \frac{\partial \tilde{L}}{\partial \nu_l^-} = 0 \\ \nu_l^- \geq 0 \end{array}\right\}, \qquad l = 1, \ldots, r,$$

Συνεπώς μπορούμε να αντικαταστήουμε τις εξισώσεις αυτές με τις $\frac{\partial \tilde{L}}{\partial \nu_l} = 0$ όπου ν_l είναι πολλαπλασιαστές του Lagrange χωρίς περιορισμούς στο πρόσημο.

Η λαγκρανζιανή μπορεί να γραφεί επομένως ως

$$\tilde{L}(\mathbf{x}, \boldsymbol{\lambda}) = f(\mathbf{x}) - \sum_{i=1}^{m} \lambda_i g_i(\mathbf{x}) - \sum_{l=1}^{r} \nu_l h_l(\mathbf{x}).$$

και οι συνθήκες στις οποίες καταλήγουμε παίρνουν την μορφή

$$\left\{\begin{array}{l} \frac{\partial \tilde{L}}{\partial x_j} = 0 \\ x_j > 0 \end{array}\right\} \text{ ή } \left\{\begin{array}{l} \frac{\partial \tilde{L}}{\partial x_j} \leq 0 \\ x_j = 0 \end{array}\right\}, \qquad j = 1, \ldots, n,$$

και

$$\left\{\begin{array}{l} \frac{\partial \tilde{L}}{\partial \lambda_i} = 0 \\ \lambda_i \geq 0 \end{array}\right\} \text{ ή } \left\{\begin{array}{l} \frac{\partial \tilde{L}}{\partial \lambda_i} > 0 \\ \lambda_i = 0 \end{array}\right\}, \qquad i = 1, \ldots, m,$$

και

$$\frac{\partial \tilde{L}}{\partial \nu_l} = 0, \qquad l = 1, \ldots, r.$$

3.4 Πρόβλημα καταμερισμού πόρων.

Έστω n συναρτήσεις, $f_j : \mathbb{R}^+ \to \mathbb{R}^+$, $j = 1, \ldots, n$, τις οποίες θεωρούμε αύξουσες και κυρτές με συνεχή δεύτερη παράγωγο και $f_j(0) = 0$. Ο στόχος είναι η μεγιστοποίηση της απόδοσης

$$\max \sum_{j=1}^{n} f_j(x_j)$$

υπό τους περιορισμούς

$$\sum_{j=1}^{n} x_j \leq C$$

$$x_j \geq 0, \quad j = 1, \ldots, n.$$

Έστω

$$\tilde{L} = \sum_{j=1}^{n} f_j(x_j) - \lambda\left(\sum_{j=1}^{n} x_j - C\right).$$

Οι συνθήκες Kuhn-Tucker στη συμμετρική τους μορφή γράφονται ως

$$\left\{\begin{matrix} \frac{\partial\tilde{L}}{\partial x_j} = f_j'(x_j) - \lambda = 0 \\ x_j > 0 \end{matrix}\right\} \text{ ή } \left\{\begin{matrix} \frac{\partial\tilde{L}}{\partial x_j} = f_j'(x_j) - \lambda \leq 0 \\ x_j = 0 \end{matrix}\right\}, \quad j = 1, \ldots, n, \tag{3.17}$$

$$\left\{\begin{matrix} \frac{\partial\tilde{L}}{\partial \lambda} = \sum_{j=1}^{n} x_j - C = 0 \\ \lambda \geq 0 \end{matrix}\right\} \text{ ή } \left\{\begin{matrix} \frac{\partial\tilde{L}}{\partial \lambda} = -\sum_{j=1}^{n} x_j + C \geq 0 \\ \lambda = 0 \end{matrix}\right\}. \tag{3.18}$$

Οι παραπάνω συνθήκες επιτρέπουν την πλήρη διερεύνηση του προβλήματος, εμείς όμως για να απλουστεύσουμε τη συζήτηση θα υποθέσουμε ότι $f_j'(x) > 0$ για κάθε $x \in \mathbb{R}^+$ και $j = 1, \ldots, n$. Με αυτή την υπόθεση, αν $\lambda = 0$ τότε υποχρωτικά, από την (3.17) $x_j = 0$ για όλα τα j και συνεπώς η τιμή του κριτηρίου προς βελτιστοποίηση είναι 0. Αυτή η λύση μας δίνει το προφανές ελάχιστο, εμείς όμως ενδιαφερόμαστε για το μέγιστο.

Ας εξετάσουμε την συγκεκριμένη περίπτωση που $f_j(x_j) = b_j x_j^{\alpha_j}$ όπου $\alpha_j \in (0,1)$ και $b_j > 0$ για κάθε j. Τότε $f_j'(x_j) = \alpha_j b_j x_j^{\alpha_j - 1}$ και οι συνθήκες γίνονται

$$\left\{\begin{matrix} \alpha_j b_j x_j^{\alpha_j - 1} = \lambda \\ x_j > 0 \end{matrix}\right\} \text{ ή } \left\{\begin{matrix} \alpha_j b_j x_j^{\alpha_j - 1} \leq \lambda \\ x_j = 0 \end{matrix}\right\}, \quad j = 1, \ldots, n, \tag{3.19}$$

$$\left\{\begin{matrix} \sum_{j=1}^{n} x_j = C \\ \lambda \geq 0 \end{matrix}\right\} \text{ ή } \left\{\begin{matrix} \sum_{j=1}^{n} x_j \leq C \\ \lambda = 0 \end{matrix}\right\}. \tag{3.20}$$

Από το πρώτο σκέλος της (3.19) έχουμε

$$x_j = \left(\frac{\lambda}{b_j \alpha_j}\lambda\right)^{(1-\alpha_j)^{-1}}, \quad j = 1, \ldots, n, \tag{3.21}$$

ενώ από το πρώτο σκέλος της (3.20), αντικαθιστώντας τις παραπάνω τιμές,

$$\sum_{j=1}^{n} (b_j a_j)^{(1-\alpha_j)^{-1}} \lambda^{-(1-\alpha_j)^{-1}} = C. \tag{3.22}$$

Η παραπάνω εξίσωση επιλύεται αριθμητικά, είναι όμως εύκολο να δει κανείς ότι έχει μοναδική λύση. Δεδομένου ότι

$$\frac{d}{d\lambda}\lambda^{-(1-\alpha_j)^{-1}} = -(1-\alpha_j)^{-1}\lambda^{-(1-\alpha_j)^{-1}-1} < 0$$

η συνάρτηση $\phi(\lambda) := \sum_{j=1}^{n} (b_j a_j)^{(1-\alpha_j)^{-1}} \lambda^{-(1-\alpha_j)^{-1}}$ είναι γνησίως φθίνουσα, συνεχής, και ισχύει ότι $\lim_{\lambda\to 0}\phi(\lambda) = \infty$, $\lim_{\lambda\to\infty}\phi(\lambda) = 0$. Συνεπώς υπάρχει μοναδική τιμή του λ που να ικανοποιεί την (3.22).

Η λύση για την γενική περίπτωση, όπως προκύπτει από τις συνθήκες (3.17), (3.18), είναι παρόμια. Έχουμε $x_j = f_j'^{-1}(\lambda)$ (όπου $f_j'^{-1}$ η αντίστροφη συνάρτηση της f_j' η οποία υπάρχει και είναι παραγωγίσιμη, αφού δεύτερη παράγωγος f_j'' είναι φθίνουσα συνάρτηση λόγω της υπόθεσης της κυρτότητας της f_j. Συνεπώς πρέπει να λύσουμε την εξίσωση $\sum_{j=1}^{n} f_j'^{-1}(\lambda) = C$. Όπως και στη συγκεκριμένη περίπτωση που εξετάσαμε πιο πάνω μπορούμε να δείξουμε ότι η $f_j'^{-1}(\lambda)$ είναι φθίνουσα συνάρτηση εξετάζοντας την παράγωγο $\frac{d}{d\lambda}f_j'^{-1}(\lambda) = \frac{1}{f_j''(f_j'(x_i))} = \frac{1}{f_j''(\lambda)} < 0$.

3.5 Προβλήματα Τετραγωνικού Προγραμματισμού

Έστω το πρόβλημα βελτιστοποίησης

$$\begin{aligned}
&\max \ \sum_{j=1}^{n} c_j x_j - \frac{1}{2}\sum_{j=1}^{n}\sum_{k=1}^{n} q_{jk} x_j x_k \\
&\text{s.t.} \\
&a_{11} + \cdots + a_{1n} x_n \ \le \ b_1 \\
&\vdots \\
&a_{m1} + \cdots + a_{mn} x_n \ \le \ b_m \\
&x_j \ge 0, \qquad j = 1, \ldots, n.
\end{aligned} \tag{3.23}$$

Ο πίνακας $Q = [q_{jk}]$ είναι συμμετρικός και θετικά ορισμένος. Η Λαγκρανζιανή είναι

$$\tilde{L} = \sum_{j=1}^{n} c_j x_j - \frac{1}{2}\sum_{j=1}^{n}\sum_{k=1}^{n} q_{jk} x_j x_k - \sum_{i=1}^{m} \lambda_i \left(\sum_{j=1}^{n} a_{ij} x_j - b_i \right)$$

και

$$\frac{\partial \tilde{L}}{\partial x_j} = c_j - \sum_{k=1}^{n} q_{jk} x_k - \sum_{i=1}^{n} \lambda_i a_{ij},$$

επομένως οι συνθήκες Kuhn–Tucker είναι οι εξής:

$$\left\{ \begin{matrix} c_j - \sum_{k=1}^{n} q_{jk} x_k - \sum_{i=1}^{n} \lambda_i a_{ij} = 0 \\ x_j > 0 \end{matrix} \right\} \text{ ή } \left\{ \begin{matrix} c_j - \sum_{k=1}^{n} q_{jk} x_k - \sum_{i=1}^{n} \lambda_i a_{ij} \le 0 \\ x_j = 0 \end{matrix} \right\}, \quad j = 1, \ldots, n, \tag{3.24}$$

και

$$\left\{ \begin{matrix} -\sum_{j=1}^{n} a_{ij} x_j + b_i = 0 \\ \lambda_i \ge 0 \end{matrix} \right\} \text{ ή } \left\{ \begin{matrix} -\sum_{j=1}^{n} a_{ij} x_j + b_i > 0 \\ \lambda_i = 0 \end{matrix} \right\}, \quad i = 1, \ldots, m, \tag{3.25}$$

Προκειμένου να βρούμε τα x_j και λ_i που ικανοποιούν αυτές τις συνθήκες μπορούμε να τις διατυπώσουμε ως ένα σύστημα γραμμικών εξισώσεων χρησιμοποιώντας μεταβλητές χαλαρότητας u_i, v_j, ως εξής

$$\begin{aligned}
&\sum_{k=1}^{n} q_{jk} x_k + \sum_{i=1}^{n} \lambda_i a_{ij} - v_j = c_j \\
&\sum_{j=1}^{n} a_{ij} x_j + u_i = b_i \\
&x_j, v_j, \lambda_i, u_i \ge 0, \qquad i = 1, \ldots, m, \ j = 1, \ldots, n.
\end{aligned} \tag{3.26}$$

Προκειμένου να ικανοποιούνται οι συνθήκες (3.24), (3.25), θα πρέπει να ισχύουν οι σχέσεις

$$x_j = 0 \Rightarrow v_j > 0, \quad j = 1, \ldots, n, \tag{3.27}$$

$$u_i > 0 \Rightarrow \lambda_i = 0, \quad i = 1, \ldots, m. \tag{3.28}$$

Το πρόβλημα μπορεί να λυθεί με τη μέθοδο Simplex. Παρατηρείστε ότι στην τελευταία αυτή διατύπωση δεν υπάρχει αντικειμενική συνάρτηση προς βελτιστοποίηση παρά μόνο η αναζήτηση μιας εφικτής λύσης. Είναι ακριβώς το πρόβλημα που αντιμετωπίζει κανείς στη μέθοδο των 'δύο φάσεων' της Simplex όταν αναζητεί μια εφικτή λύση χρησιμοποιώντας τεχνητές μεταβλητές μόνο που κατά την εφαρμογή της μεθόδου πρέπει να λαμβάνονται υπ΄ όψιν και οι συνθήκες (3.27), (3.28).

Κεφάλαιο 4

Εφαρμογές

4.1 Μεγιστοποίηση εντροπίας

Ορισμός 7. *Έστω $\{p_i; i = 0, 1, \ldots, \}$ μια κατανομή πιθανότητας στο $\mathbb{N}$, δηλαδή $\pi_i \geq 0$ και $\sum_{i=0}^{\infty} p_i = 1$. Η* εντροπία *της κατανομής αυτής ορίζεται ως η ποσότητα*

$$H = -\sum_{i=0}^{\infty} p_i \log p_i. \tag{4.1}$$

Η εντροπία είναι ένα μέτρο της αβεβαιότητας που εκφράζει μια κατανομή πιθανότητας. Εναλλακτικά, αν X είναι μια τυχαία μεταβλητή με κατανομή $P(X = i) = p_i$, τότε η εντροπία είναι ένα μέτρο της πληροφορίας που παίρνουμε όταν μάθουμε την τιμή της τυχαίας μεταβλητής X. Για παράδειγμα, αν X είναι μια τυχαία μεταβλητή Bernoulli με $P(X = 0) = 1 - p$ και $P(X = 1) = p$, τότε η εντροπία της X είναι

$$H(X) = -p \log p - (1 - p) \log(1 - p)$$

Παρατηρείστε ότι η εντροπία μεγιστοποιείται για $p = 1/2$. Πράγματι, για αυτή την τιμή του p η αβεβαιότητα είναι μέγιστη πριν μάθουμε την τιμή της X ή, ισοδύναμα, η πληροφορία πού παίρνουμε γνωρίζοντας την τιμή της X μεγιστοποιείται.

Ας υποθέσουμε τώρα ότι έχουμε μια κατανομή $\{p_i\}$, $i = 1, \ldots, n$ στους φυσικούς $\{1, \ldots, n\}$. Η εντροπία της θα είναι

$$H = -\sum_{i=1}^{n} p_i \log p_i. \tag{4.2}$$

Η κατανομή στο σύνολο $\{1, \ldots, n\}$ που μεγιστοποιεί την εντροπία προκύπτει από την μεγιστοποίηση της (4.2) υπό τους περιορισμούς $\sum_{i=1}^{n} p_i = 1$ και $p_i \geq 0$, $p_i \leq 1$, $i = 1, \ldots, n$. Ορίζοντας την λαγκρανζιανή ως $\tilde{L} = -\sum_{i=1}^{n} p_i \log p_i - \lambda \left(\sum_{i=1}^{n} p_i - 1 \right) - \sum_{i=1}^{n} \mu_i (p_i - 1)$ και οι συνθήκες Kuhn–Tucker

Σχήμα 4.1: Η εντροπία μιας κατανομής Bernoulli

δίνουν

$$
\begin{aligned}
&\left\{\begin{matrix} \frac{d\tilde{L}}{dp_i} = 0 \\ p_i > 0 \end{matrix}\right\} \text{ ή } \left\{\begin{matrix} \frac{d\tilde{L}}{dp_i} \leq 0 \\ p_i = 0 \end{matrix}\right\}, \quad i = 1, \ldots, n, \\
&\left\{\begin{matrix} \frac{d\tilde{L}}{d\mu_i} = 0 \\ \mu_i \geq 0 \end{matrix}\right\} \text{ ή } \left\{\begin{matrix} \frac{d\tilde{L}}{d\mu_i} \geq 0 \\ \mu_i = 0 \end{matrix}\right\}, \quad i = 1, \ldots, n, \\
&\frac{d\tilde{L}}{d\lambda} = 0,
\end{aligned}
$$

ή, ισοδύναμα,

$$
\begin{aligned}
&\left\{\begin{matrix} -\log p_i - 1 - \lambda = 0 \\ p_i > 0 \end{matrix}\right\} \text{ ή } \left\{\begin{matrix} -\log p_i - 1 - \lambda \leq 0 \\ p_i = 0 \end{matrix}\right\}, \quad i = 1, \ldots, n, \\
&\left\{\begin{matrix} -p_i + 1 = 0 \\ \mu_i \geq 0 \end{matrix}\right\} \text{ ή } \left\{\begin{matrix} -p_i + 1 \geq 0 \\ \mu_i = 0 \end{matrix}\right\}, \quad i = 1, \ldots, n, \\
&\sum_{i=1}^{n} p_i = 1.
\end{aligned}
$$

Η περίπτωση $p_i = 0$ οδηγεί σε ιδιομορφία (και σε τοπικά ελάχιστα της H) και δεν θα την εξετάσουμε. Για τον ίδιο λόγο αποκλείουμε και την περίπτωση $p_i = 1$ για κάποιο i, πράγμα που σημαίνει ότι όλες οι άλλες πιθανότητες είναι 0. Αν όλα τα p_i είναι αυστηρά θετικά, έχουμε $p_i^{-1} = 1 + \lambda$ και συνεπώς $1 = \sum_{i=1}^{n} p_i = \frac{n}{1+\lambda}$ ή $\lambda = n - 1$. Η περίπτωση αυτή είναι το ολικό μέγιστο της εντροπίας, δηλαδή η εντροπία στην περίπτωση αυτή μεγιστοποιείται οταν η κατανομή είναι ομοιόμορφη με $p_i = \frac{1}{n}$, $i = 1, \ldots, n$.

Τέλος ας δούμε ποια είναι η κατανομή p_i στους φυσικούς, $i \in \{1, 2, 3, \ldots, \}$ που μεγιστοποιεί την εντροπία, κάτω από την συνθήκη ότι ο μέσος της κατανομής είναι δεδομένος και ίσος με $\theta = \sum_{i=1}^{\infty} i p_i$ όπου θα πρέπει βέβαια να ισχύει $\theta > 1$. Το πρόβλημα αυτό βγαίνει κάπως έξω από το πλαίσιο που συζητάμε μια και ο αριθμός των μεταβλητών δεν είναι πεπερασμένος, όμως μπορούμε να το αντιμετωπίσουμε με τα ίδια εργαλεία. Η λαγκρανζιανή είναι

$$-\sum_{i=1}^{\infty} p_i \log p_i - \lambda\left(\sum_{i=1}^{\infty} p_i - 1\right) - \nu\left(\sum_{i=1}^{\infty} i p_i - \theta\right)$$

και εξετάζοντας μόνο την περίπτωση που $0 < p_i < 1$ για κάθε i έχουμε τις εξισώσεις

$$-\log p_i - 1 - \lambda - i\nu = 0, \qquad i = 1, 2, \ldots, \tag{4.3}$$

$$\sum_{i=1}^{\infty} p_i = 1, \tag{4.4}$$

$$\sum_{i=1}^{\infty} i p_i = \theta. \tag{4.5}$$

Από τις σχέσεις (4.3) προκύπτει ότι

$$p_i = e^{-(1+\lambda+i\nu)}, \;\; i = 1, 2, \ldots, \tag{4.6}$$

και αντικαθιστώντας στις (4.5) και (4.6) παίρνουμε τις

$$1 = \sum_{i=1}^{\infty} e^{-(1+\lambda+i\nu)} = e^{-(1+\lambda)} \sum_{i=1}^{\infty} \left(e^{-\nu}\right)^i = e^{-(1+\lambda)} \frac{e^{-\nu}}{1 - e^{-\nu}} \tag{4.7}$$

(όπου στην τελευταία έκφραση χρησιμοποιήσαμε το άθροισμα της γεωμετρικής σειράς $\sum_{i=0}^{\infty} x^i = \frac{1}{1-x}$ για $|x| < 1$) και

$$\theta = \sum_{i=1}^{\infty} i e^{-(1+\lambda+i\nu)} = e^{-(1+\lambda)} \sum_{i=1}^{\infty} i \left(e^{-\nu}\right)^i = e^{-(1+\lambda)} \frac{e^{-\nu}}{(1 - e^{-\nu})^2} \tag{4.8}$$

(όπου χρησιμοποιήσαμε ότι $\sum_{i=0}^{\infty} i x^{i-1} = \frac{1}{(1-x)^2}$ για $|x| < 1$). Από τις (4.7) και (4.8) προκύπτει ότι

$$\frac{1}{1 - e^{-\nu}} = \theta$$

ή

$$\nu = \log \frac{\theta}{\theta - 1}$$

(ισχύει ότι $\theta > 1$). Αντικαθιστώντας στην (4.4), μετά από στοιχειώδεις πράξεις, έχουμε

$$p_i = \frac{1}{\theta}\left(1 - \frac{1}{\theta}\right)^{i-1}, \;\; i = 1, 2, \ldots$$

δηλαδή την γεωμετρική κατανομή με πιθανότητα επιτυχίας $\frac{1}{\theta}$.

4.2 Προβλήματα πολλαπλών σταδίων

Θεωρούμε το εξής απλοποιημένο πρόβλημα επένδυσης–κατανάλωσης. Ξεκινάμε με δεδομένο αρχικό κεφάλαιο x_0 και σε κάθε στάδιο του χρονικού ορίζοντα που διαρκεί N περιόδους καλούμεθα να αποφασίσουμε τι τμήμα του κεφαλαίου θα καταναλώσουμε και τι τμήμα θα επενδύσουμε. Ας υποθέσουμε ότι στην αρχή της περιόδου n το κεφάλαιό μας είναι x_n, καταναλώνουμε u_n και επενδύουμε το υπόλοιπο τμήμα του κεφαλαίου, $x_n - u_n$. Θα κάνουμε την απλή υπόθεση ότι, υπό αυτές τις συνθήκες, το κεφάλαιο στην αρχή της περιόδου $n+1$ θα είναι

$$x_{n+1} = c(x_n - u_n) \tag{4.9}$$

όπου $c > 1$. Η χρησιμότητα που μας αποφέρει η κατανάλωση u έστω ότι είναι $\phi(u)$ όπου ϕ μια θετική, αύξουσα, κυρτή συνάρτηση. Σκοπός μας είναι να μεγιστοποιήσουμε την συνολική χρησιμότητα $\sum_{n=0}^{N-1} \phi(u_n)$ επιλέγοντας κατάλληλα τα u_n, $n = 0, \ldots, N_1$ και ικανοποιώντας τους περιορισμούς $u_n \geq l_n \geq 0$ και $x_N = K$. Το l_n είναι η ελάχιστη κατανάλωση κατά την περίοδο n (που μπορεί να προκύπτει από ανηλειμμένες υποχρεώσεις) ενώ το K είναι ένας τελικός περιορισμός για το κεφάλαιο που πρέπει να υπάρχει στο τέλος της χρονικού ορίζοντα.

Οι μεταβλητές μας είναι τα x_n, $n = 1, \ldots, N$ και u_n, $n = 0, \ldots, N-1$ και η λαγκρανζιανή είναι

$$\tilde{L} = \sum_{n=0}^{N-1} \phi(u_n) - \sum_{n=0}^{N-1} \lambda_n(x_{n+1} - c(x_n - u_n)) - \sum_{n=0}^{N-1} \mu_n(l_n - u_n) - \nu(x_N - K)$$

Οι συνθήκες Kuhn–Tucker δίνουν

$$\begin{aligned}
&\frac{\partial \tilde{L}}{\partial x_n} = \lambda_{n-1} - c\lambda_n = 0, \qquad n = 1, \ldots, N-1,\\
&\frac{\partial \tilde{L}}{\partial x_N} = \lambda_{N-1} - \nu = 0,\\
&\frac{\partial \tilde{L}}{\partial u_n} = \phi'(u_n) - c\lambda_n + \mu_n = 0, \qquad n = 0, \ldots, N-1,\\
&\frac{\partial \tilde{L}}{\partial \lambda_n} = x_{n+1} - c(x_n - u_n) = 0, \qquad n = 0, \ldots, N-1,\\
&\frac{\partial \tilde{L}}{\partial \nu} = -x_N + K = 0,\\
&\left\{\begin{matrix} \frac{\partial \tilde{L}}{\partial \mu_n} = u_n - l_n = 0 \\ \mu_n \geq 0 \end{matrix}\right\} \text{ ή } \left\{\begin{matrix} \frac{\partial \tilde{L}}{\partial \mu_n} = u_n - l_n > 0 \\ \mu_n = 0 \end{matrix}\right\}, \qquad n = 0, \ldots, N-1,
\end{aligned}$$

ή ισοδύναμα

$$\lambda_{n-1} = c\lambda_n, \qquad n = 1, \ldots, N-1, \tag{4.10}$$

$$\lambda_{N-1} = \nu, \tag{4.11}$$

$$\phi'(u_n) - c\lambda_n + \mu_n = 0, \qquad n = 0, \ldots, N-1, \tag{4.12}$$

$$x_{n+1} - c(x_n - u_n) = 0, \qquad n = 0, \ldots, N-1, \tag{4.13}$$

$$x_N = K, \tag{4.14}$$

$$\left\{\begin{matrix} u_n = l_n \\ \mu_n \geq 0 \end{matrix}\right\} \text{ ή } \left\{\begin{matrix} u_n > l_n \\ \mu_n = 0 \end{matrix}\right\}, \qquad n = 0, \ldots, N-1, \tag{4.15}$$

Οι εξισώσεις (4.11), (4.12), προσδιορίζουν άμεσα τα λ_n:

$$\lambda_n = \nu c^{N-1-n}, \qquad n = 0, \ldots, N-1.$$

Προκειμένου να λύσουμε εύκολα το σύστημα ας υποθέσουμε ότι $l_n = 0$ για κάθε n και ας διερευνήσουμε την ύπαρξη λύσης με $\mu_n = 0$ για όλα τα n. Επίσης, παρ΄ ότι δεν είναι αναγκαίο, ας εξετάσουμε την συγκεκριμένη περίπτωση $\phi(u) = \sqrt{u}$. Τότε η (4.13) δίνει $\frac{1}{2}u_n^{-1/2} = c\nu c^{N-1-n}$ ή

$$u_n = (2c\nu)^{-2} c^{2(n+1-N)}, \quad n = 0, \ldots, N-1$$

και

$$x_{n+1} = cx_n - \frac{1}{4c\nu^2} c^{2(n+1-N)}, \quad n = 0, \ldots, N-1,$$

όπου το x_0 είναι δεδομένο. Η παραπάνω αναδρομή λύνεται εύκολα ενώ η τιμή του ν προσδιορίζεται από την σχέση $x_N = K$.