ΠΟΛΥΤΕΧΝΕΙΟ ΚΡΗΤΗΣ

ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ ΠΑΡΑΓΩΓΗΣ ΚΑΙ ΔΙΟΙΚΗΣΗΣ

ΜΗ ΓΡΑΜΜΙΚΟΣ ΠΡΟΓΡΑΜΜΑΤΙΣΜΟΣ

Καθ. Μάρκος Παπαγεωργίου

ΧΑΝΙΑ, 2020

ΠΙΝΑΚΑΣ ΠΕΡΙΕΧΟΜΕΝΩΝ

1. ΕΙΣΑΓΩΓΗ 1.1
1.1 Γενικά Στοιχεία 1.1
1.2 Μαθηματικό Υπόβαθρο 1.2
Προβλήματα 1.7

ΜΕΡΟΣ ΠΡΩΤΟ:
ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΧΩΡΙΣ ΠΕΡΙΟΡΙΣΜΟΥΣ

2. ΓΕΝΙΚΗ ΜΕΘΟΔΟΛΟΓΙΑ 2.1
2.1 Συνθήκες για Τοπικά Ελάχιστα 2.1
2.2 Ευρετικές Μέθοδοι 2.4
2.3 Χρήσιμες Ιδιότητες Αλγορίθμων 2.6
2.4 Τετραγωνικά Πρότυπα 2.9
2.5 Μέθοδοι Κατάβασης και Ευστάθεια 2.9
2.6 Αλγόριθμοι Αναζήτησης Επί Γραμμής 2.10
Προβλήματα 2.14

3. ΜΕΘΟΔΟΙ ΤΥΠΟΥ NEWTON 3.1
3.1 Η μέθοδος Newton 3.1
3.2 Μέθοδοι Σχεδόν-Newton 3.3
3.3 Αλλαγή Κλίμακας 3.6
3.4 Αριθμητικά Πειράματα 3.7
Προβλήματα 3.10

4. ΜΕΘΟΔΟΙ ΣΥΖΥΓΩΝ ΚΛΙΣΕΩΝ 4.1
4.1 Περιγραφή Μεθόδων 4.1
4.2 Αριθμητικά Πειράματα 4.4
Προβλήματα 4.6

ΜΕΡΟΣ ΔΕΥΤΕΡΟ:
ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΜΕ ΠΕΡΙΟΡΙΣΜΟΥΣ

5. ΕΙΣΑΓΩΓΗ 5.1
5.1 Γενικά Στοιχεία 5.1
5.2 Απαλοιφή και άλλοι Μετασχηματισμοί 5.4
Προβλήματα 5.6

6. ΘΕΩΡΙΑ ΒΕΛΤΙΣΤΟΠΟΙΗΣΗΣ ΜΕ ΠΕΡΙΟΡΙΣΜΟΥΣ 6.1
6.1 Περιορισμοί Ισότητας 6.1
6.2 Περιορισμοί Ισότητας και Ανισότητας 6.5
6.3 Συνθήκες Δευτέρου Βαθμού 6.11
6.4 Κυρτά Προβλήματα Βελτιστοποίησης 6.13
Προβλήματα 6.16

7. ΤΕΤΡΑΓΩΝΙΚΟΣ ΠΡΟΓΡΑΜΜΑΤΙΣΜΟΣ 7.1
7.1 Περιορισμοί Ισότητας 7.1
7.2 Η Μέθοδος Lagrange 7.7
7.3 Μέθοδοι Ενεργού Συνόλου 7.8
Προβλήματα 7.11

8. ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΜΕ ΓΡΑΜΜΙΚΟΥΣ ΠΕΡΙΟΡΙΣΜΟΥΣ 8.1
8.1 Περιορισμοί Ισότητας 8.1
8.2 Περιορισμοί Ανισότητας 8.3
Προβλήματα 8.5

9. ΑΛΓΟΡΙΘΜΟΙ ΜΗ ΓΡΑΜΜΙΚΟΥ ΠΡΟΓΡΑΜΜΑΤΙΣΜΟΥ 9.1
9.1 Μέθοδοι Συνάρτησης Τιμωρίας 9.1
9.2 Συναρτήσεις Τιμωρίας με Πολλαπλασιαστές 9.4
α) Περιορισμοί Ισότητας 9.4
β) Περιορισμοί Ανισότητας 9.8
9.3 Ακολουθητικός Τετραγωνικός Προγραμματισμός (ΑΤΠ) 9.10
α) Περιορισμοί Ισότητας 9.10
β) Περιορισμοί Ανισότητας 9.13
9.4 Μη Γραμμική Απαλοιφή και Μέθοδοι Επιτρεπτής Κατεύθυνσης 9.14
Προβλήματα 9.18

10. ΜΕΘΟΔΟΙ ΑΠΟΛΥΤΗΣ ΒΕΛΤΙΣΤΟΠΟΙΗΣΗΣ 10.1
10.1 Η Μέθοδος Complex 10.1
10.2 Προσομοιωμένη Ανόπτηση 10.2
10.3 Γενετικοί Αλγόριθμοι 10.3
10.4 Συμπεράσματα 10.5

ΠΡΟΤΕΙΝΟΜΕΝΗ ΒΙΒΛΙΟΓΡΑΦΙΑ .....

ΑΣΚΗΣΕΙΣ ΕΞΕΤΑΣΕΩΝ .....

1. ΕΙΣΑΓΩΓΗ

1.1 Γενικά Στοιχεία

Το ευρύτερο πεδίο της **βελτιστοποίησης** ή λήψης βέλτιστων αποφάσεων περιλαμβάνει την αντίστοιχη θεωρία (τμήμα των Εφαρμοσμένων Μαθηματικών) και εφαρμογές σε όλους σχεδόν του τομείς των επιστημών, και ειδικά των τεχνικών και οικονομικών επιστημών. Η βελτιστοποίηση είναι πηγή τεχνητής ευφυΐας με την έννοια του προσδιορισμού βέλτιστων λύσεων σε πολύπλοκα προβλήματα που ξεπερνούν τις δυνατότητες οποιωνδήποτε εμπειρογνωμόνων. Με την χρήση ηλεκτρονικών υπολογιστών η επίλυση των προβλημάτων επιτυγχάνεται ταχύτατα μέσω κατάλληλων αλγορίθμων.

Ο **Μη Γραμμικός Προγραμματισμός** είναι μέρος της θεωρίας Βελτιστοποίησης, εφαρμόσιμος σε μια ειδική κατηγορία προβλημάτων που θα προσδιορισθούν αργότερα. Ο όρος **Προγραμματισμός** χρησιμοποιείται με την έννοια "σχεδιασμός" ή βελτιστοποίηση και όχι με την έννοια προγραμματισμού ηλεκτρονικού υπολογιστή.

Κάθε πρόβλημα μη γραμμικής βελτιστοποίησης εμπεριέχει μια **αντικειμενική συνάρτηση** ή **συνάρτηση κόστους** ή **συνάρτηση ωφέλειας** προς μεγιστοποίηση ή ελαχιστοποίηση. Ανάλογα με την υφή του, μπορεί επίσης το εκάστοτε πρόβλημα να περιλαμβάνει **περιορισμούς ισότητας** ή **ανισότητας.**

Παράδειγμα 1.1.1: Ζητείται το ορθογώνιο τρίγωνο με δεδομένη υποτείνουσα l και μέγιστο εμβαδόν. Η αντικειμενική συνάρτηση προς μεγιστοποίηση είναι

$$f(x_1,x_2) = \frac{1}{2} x_1 \cdot x_2 \qquad (1.1.1)$$

όπου x_1, x_2 οι ζητούμενες κάθετες που πρέπει να ικανοποιούν τους περιορισμούς ισότητας

$$x_1^2 + x_2^2 = l^2 \qquad (1.1.2)$$

και ανισότητας

$$x_1 \geq 0 \qquad (1.1.3)$$
$$x_2 \geq 0.$$

■

1.2 Μαθηματικό Υπόβαθρο

Για τον ευχερέστερο χειρισμό συναρτήσεων πολλών μεταβλητών ορίζουμε ως εξής ένα (κάθετο) **διάνυσμα** πραγματικών αριθμών $a_i \in R$, $i = 1, \ldots n$:

$$\mathbf{a} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} \in R^n$$

και έναν **πίνακα** πραγματικών αριθμών $B_{ij} \in R$, $i = 1, \ldots, n$, $j = 1, \ldots, m$:

$$\mathbf{B} = \begin{bmatrix} B_{11} & B_{12} & \cdots & B_{1m} \\ B_{21} & B_{22} & \cdots & B_{2m} \\ \vdots & \vdots & & \vdots \\ B_{n1} & B_{n2} & \cdots & B_{nm} \end{bmatrix} \in R^{n\times m}.$$

Η **αναστροφή** ενός διανύσματος ή ενός πίνακα ορίζεται

$$\mathbf{a}^T = [a_1\ a_2\ \ldots\ a_n]\ ,\ \mathbf{B}^T = \begin{bmatrix} B_{11} & B_{21} & \cdots & B_{n1} \\ B_{12} & B_{22} & \cdots & B_{n2} \\ \vdots & \vdots & & \vdots \\ B_{1m} & B_{2m} & \cdots & B_{nm} \end{bmatrix}$$

ενώ για το **βαθμωτό γινόμενο** δύο διανυσμάτων ίσων διαστάσεων έχουμε $\mathbf{a}^T\mathbf{z} = \sum_i a_i z_i = \mathbf{z}^T\mathbf{a}$.

Ένα n-διάστατο διάνυσμα μπορεί να θεωρηθεί ως **σημείο** ενός n-διάστατου χώρου, π.χ. του χώρου πραγματικών αριθμών $\mathbf{x} \in R^n$, οπότε x_i είναι η συνιστώσα της διεύθυνσης συντεταγμένων i.

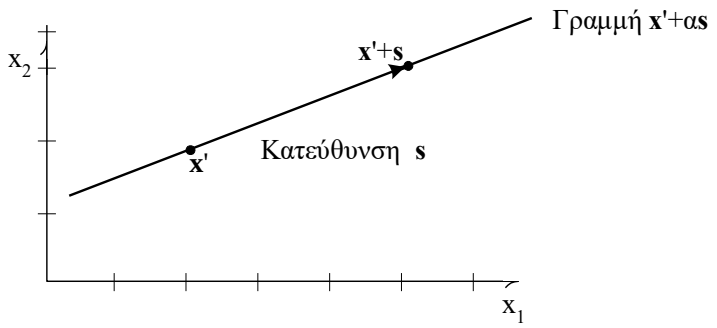
Ονομάζουμε **γραμμή** το σύνολο σημείων

$$\mathbf{x}(\alpha) = \mathbf{x}' + \alpha\mathbf{s}, \quad \alpha \in R \qquad (1.2.1)$$

όπου $\mathbf{x}'$ είναι ένα δεδομένο σημείο της γραμμής (για $\alpha = 0$) και $\mathbf{s}$ είναι η **κατεύθυνση** της γραμμής. Συχνά έχουμε την (1.2.1) $\forall\alpha \geq 0$, δηλαδή μισή γραμμή.

Παράδειγμα 1.2.1: Στο Σχ. 1.2.1 έχουμε γραμμή σε χώρο δύο διαστάσεων με $\mathbf{x}' = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$, $\mathbf{s} = \begin{bmatrix} 3 \\ 1 \end{bmatrix}$.

■



Σχήμα 1.2.1: Γραμμή στο χώρο $\mathbf{x}\in R^2$.

Η κατεύθυνση **s** μπορεί να τυποποιηθεί έτσι ώστε

$$\|\mathbf{s}\|_2 = \sqrt{\mathbf{s}^T\mathbf{s}} = \sqrt{\sum_i s_i^2} = 1$$

πράγμα που δεν αλλάζει τη γραμμή (γιατί;), παρά μόνο τις τιμές α συγκεκριμένων σημείων επ' αυτής. Το μέτρο $\|\cdot\|_2$ αντιστοιχεί στο μέγεθος ενός διανύσματος.

Ορίζουμε, χάριν συντομίας, μια συνάρτηση πολλαπλών μεταβλητών $f(x_1,x_2,...,x_n)$ ως συνάρτηση $f(\mathbf{x})$ του διανύσματος **x**. Ονομάζουμε **περιγράμματα** ή **ισοϋψείς** ή **υψομετρικές επιφάνειες** τα σύνολα των σημείων όπου η τιμή $f(\mathbf{x})$ είναι σταθερή.

Παράδειγμα 1.2.2: Η συνάρτηση μπανάνας του Rosenbrock

$$f(\mathbf{x}) = 100(x_2-x_1^2)^2 + (1-x_1)^2 \qquad (1.2.2)$$

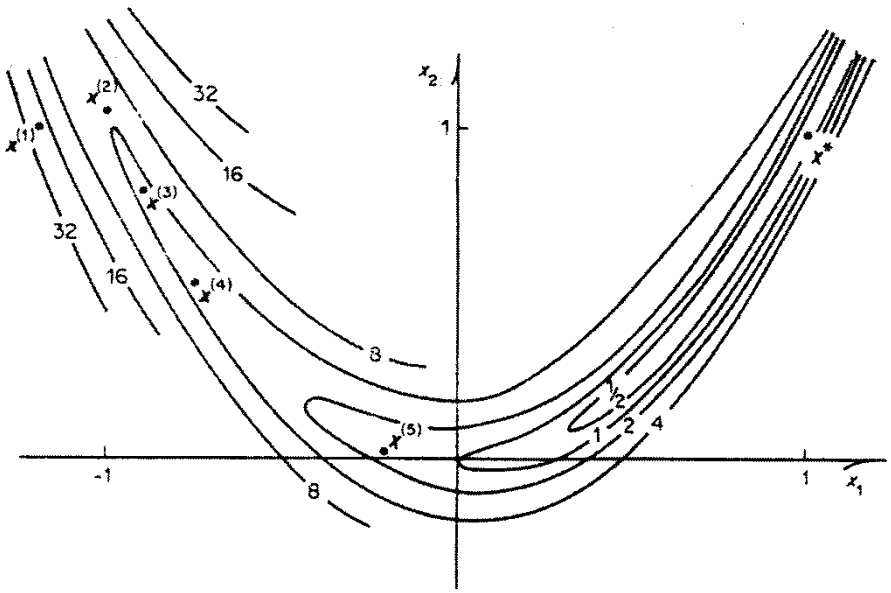
έχει τις υψομετρικές καμπύλες του Σχ. 1.2.2.

■

Πολλοί αριθμητικοί αλγόριθμοι επίλυσης προβλημάτων βελτιστοποίησης είναι **επαναληπτικοί** και παράγουν, σε αντίστοιχες επαναλήψεις, μια ακολουθία σημείων $\mathbf{x}^{(1)}$, $\mathbf{x}^{(2)}$, $\mathbf{x}^{(3)}$,... ή συντομότερα $\{\mathbf{x}^{(k)}\}$ (ο άνω δείκτης υποδεικνύει τον αριθμό επανάληψης) που πρέπει να συγκλίνει σε ένα σταθερό σημείο $\mathbf{x}^*$ ανταποκρινόμενο στη λύση του προβλήματος (βλ. Σχ. 1.2.2). Ονομάζουμε τα $\mathbf{x}^{(k)}$ **επαναληπτικά σημεία** ή απλώς **επαναληπτικά**.

Συναρτήσεις **λείες** (συνεχείς και συνεχώς παραγωγίσιμες) θεωρούνται στοιχεία του συνόλου C^1, δηλαδή $f\in C^1$.Το **διάνυσμα μερικών παραγώγων** ή **διάνυσμα κλίσης** της συνάρτησης $f(\mathbf{x})$, $\mathbf{x}\in R^n$, $f\in C^1$, σε κάθε σημείο **x** ορίζεται και συμβολίζεται ως εξής

$$\begin{bmatrix} \partial f/\partial x_1 \\ \partial f/\partial x_2 \\ \vdots \\ \partial f/\partial x_n \end{bmatrix}_{\mathbf{x}} = \nabla f(\mathbf{x}) = \left.\frac{\partial f}{\partial \mathbf{x}}\right|_{\mathbf{x}}. \qquad (1.2.3)$$



Σχήμα 1.2.2: Περιγράμματα της συνάρτησης Rosenbrock.

Αν $f(\mathbf{x}) \in C^2$ (συνάρτηση συνεχής και δύο φορές συνεχώς παραγωγίσιμη), υπάρχει ένας **πίνακας δεύτερων μερικών παραγώγων** ή **Χεσιανός πίνακας** $\nabla^2 f(\mathbf{x})$ με στοιχεία (i, j) ίσα προς $\partial^2 f/\partial x_i \partial x_j$. Ο Χεσιανός πίνακας είναι προφανώς τετραγωνικός και συμμετρικός (γιατί;). Συχνά θα χρησιμοποιούμε τους συμβολισμούς $\mathbf{g} = \nabla f$ και $\mathbf{G} = \nabla^2 f$.

Παράδειγμα 1.2.3: Για τη συνάρτηση (1.2.2) έχουμε

$$\nabla f(\mathbf{x}) = \begin{bmatrix} -400x_1(x_2 - x_1^2) - 2(1 - x_1) \\ 200(x_2 - x_1^2) \end{bmatrix}$$

$$\nabla^2 f(\mathbf{x}) = \begin{bmatrix} 1200x_1^2 - 400x_2 + 2 & -400x_1 \\ -400x_1 & 200 \end{bmatrix}. \tag{1.2.4}$$

Προφανώς οι ∇f και $\nabla^2 f$ εξαρτώνται από το $\mathbf{x}$, έχουν δηλαδή διαφορετική τιμή σε κάθε σημείο $\mathbf{x}$. Επί παραδείγματι, για $\mathbf{x} = \mathbf{0}$ έχουμε στην (1.2.4)

$$\nabla f(\mathbf{0}) = \begin{bmatrix} -2 \\ 0 \end{bmatrix}, \qquad \nabla^2 f(\mathbf{0}) = \begin{bmatrix} 2 & 0 \\ 0 & 200 \end{bmatrix}.$$

■

Θεωρούμε μια **διανυσματική συνάρτηση**

$$\mathbf{h}(\mathbf{x}) = \begin{bmatrix} h_1(\mathbf{x}) \\ \vdots \\ h_m(\mathbf{x}) \end{bmatrix} \tag{1.2.5}$$

και ορίζουμε τον ορθογώνιο (n×m) **Ιακωβιανό πίνακα** των πρώτων παραγώγων

$$\nabla \mathbf{h}(\mathbf{x}) = \begin{bmatrix} \frac{\partial h_1}{\partial x_1} & \cdots & \frac{\partial h_m}{\partial x_1} \\ \vdots & & \vdots \\ \frac{\partial h_1}{\partial x_n} & \cdots & \frac{\partial h_m}{\partial x_n} \end{bmatrix} \tag{1.2.6}$$

που αποτελεί προφανώς γενίκευση του διανύσματος κλίσης (1.2.3). Επί παραδείγματι, μια διανυσματική συνάρτηση είναι το διάνυσμα κλίσης $\mathbf{g}(\mathbf{x}) = \nabla f(\mathbf{x})$. Ο Ιακωβιανός πίνακας της $\mathbf{g}(\mathbf{x})$ είναι $\nabla^2 f(\mathbf{x})$, δηλαδή ίσος προς το Χεσιανό πίνακα της συνάρτησης $f(\mathbf{x})$. Σημειώνουμε ότι, αν x βαθμωτός, τότε, σύμφωνα με την (1.2.6), $\nabla \mathbf{h}(x)$ είναι **οριζόντιο** διάνυσμα. Σημειώνουμε επίσης ότι ισχύει, βάσει του ορισμού (1.2.6), $\nabla[\mathbf{h}(\mathbf{x})^T] = [\nabla \mathbf{h}(\mathbf{x})]^T$, άρα ο Ιακωβιανός πίνακας μιας **οριζόντιας** διανυσματικής συνάρτησης $\mathbf{h}(\mathbf{x})^T = [h_1(\mathbf{x}) \ \ldots \ h_m(\mathbf{x})]$ είναι ο ανάστροφος του πίνακα της (1.2.6).

Το Πρόβλημα 1.1 παραθέτει χρήσιμους κανόνες παραγωγισμού συναρτήσεων διανυσμάτων $f(\mathbf{x})$ και διανυσματικών συναρτήσεων $\mathbf{f}(\mathbf{x})$.

Ορίζουμε το **διάνυσμα συντεταγμένης**

$$\mathbf{e}_i = [0 \ 0 \ \ldots \ 1(\text{i-τη συνιστώσα}) \ \ldots \ 0]^T. \tag{1.2.7}$$

Για την συνάρτηση $\mathbf{h}(\mathbf{x}) = \mathbf{x}$ έχουμε λόγω της (1.2.6)

$$\nabla \mathbf{x} = [\mathbf{e}_1 \ \mathbf{e}_2 \ldots \mathbf{e}_n] = \mathbf{I} \tag{1.2.8}$$

όπου **I** ο **μοναδιαίος πίνακας**.

Θεωρούμε τώρα την συνάρτηση μιάς μεταβλητής $f(\alpha)$ που προκύπτει από μιά συνάρτηση πολλαπλών μεταβλητών $f(\mathbf{x})$ για $\mathbf{x}$ επί γραμμής (1.2.1), δηλαδή $f(\alpha) = f(\mathbf{x}(\alpha)) = f(\mathbf{x}' + \alpha\mathbf{s})$. Για τον υπολογισμό της πρώτης παραγώγου $f'(\alpha)$ κατά μήκος μιάς γραμμής (1.2.1) έχουμε (κάνοντας χρήση των (Π1.8) και (1.2.6))

$$f'(\alpha) = \frac{df(\mathbf{x}(\alpha))}{d\alpha} = \frac{\partial \mathbf{x}(\alpha)}{\partial \alpha} \frac{\partial f}{\partial \mathbf{x}} = \sum_i \frac{dx_i(\alpha)}{d\alpha} \frac{\partial f}{\partial x_i} = \sum_i s_i \frac{\partial f}{\partial x_i} = \mathbf{s}^T \nabla f\left(\mathbf{x}(\alpha)\right). \tag{1.2.9}$$

Για την **καμπυλότητα** (δεύτερη παράγωγο) $f''(\alpha)$ κατά μήκος της γραμμής $\mathbf{x}(\alpha)$, εφαρμόζουμε τον τύπο (1.2.9) για δεύτερη φορά και έχουμε

$$f''(\alpha) = \frac{d^2 f(\mathbf{x}(\alpha))}{d\alpha^2} = \frac{d}{d\alpha} \frac{df(\mathbf{x}(\alpha))}{d\alpha} = \mathbf{s}^T \nabla(\nabla f^T \mathbf{s}) = \mathbf{s}^T (\nabla^2 f(\mathbf{x}(\alpha)))\, \mathbf{s} \tag{1.2.10}$$

όπου έγινε χρήση της (Π1.1).

Παράδειγμα 1.2.4: Για τη συνάρτηση (1.2.2), την κατεύθυνση γραμμής $\mathbf{s} = [1\ \ 0]^T$ και $\mathbf{x}' = [0\ \ 0]^T$ (δηλαδή, γραμμή = άξονας x_1) έχουμε για το σημείο $\mathbf{x}'$ (δηλαδή για α = 0) $f'(0) = \mathbf{s}^T\nabla f(\mathbf{x}') = -2$ και καμπυλότητα $\mathbf{s}^T\mathbf{G}(\mathbf{x}')\mathbf{s} = 2$.

■

Προφανώς οι κλίση και καμπυλότητα κατά μήκος μιας γραμμής εξαρτώνται και από το μέγεθος του **s**, εκτός αν υπάρχει τυποποίηση του **s**. Αν, για τυχαίο σημείο $\mathbf{x}'$ με διάνυσμα κλίσης $\mathbf{g}'$, θεωρήσουμε όλες τις τυποποιημένες κατευθύνσεις $\|\mathbf{s}\|_2=1$, τότε οι κατευθύνσεις $\pm\mathbf{g}'/\|\mathbf{g}'\|_2$ είναι εκείνες με την μέγιστη και ελάχιστη κλίση, όπως μπορεί εύκολα να αποδειχθεί βάσει της (1.2.9), βλ. Πρόβλημα 1.2. Οι ίδιες κατευθύνσεις είναι ορθογώνιες στο περίγραμμα και το εφαπτόμενο επίπεδο της $f(\mathbf{x})$ στο σημείο $\mathbf{x}'$, βλ. Σχ. 1.2.3 και Πρόβλημα 1.2.

Ορίζουμε τη γενική **γραμμική συνάρτηση** πολλαπλών μεταβλητών ως εξής

$$l(\mathbf{x}) = \sum_i a_i x_i + b = \mathbf{a}^T\mathbf{x} + b \qquad (1.2.11)$$

με δεδομένα **a**, b. Με χρήση της (Π1.4) έχουμε για την (1.2.11) $\nabla l=\mathbf{a}$ και $\nabla^2 l=\mathbf{0}$, όπου **0** ο **μηδενικός πίνακας**.

Η γενική **τετραγωνική συνάρτηση** δίδεται από

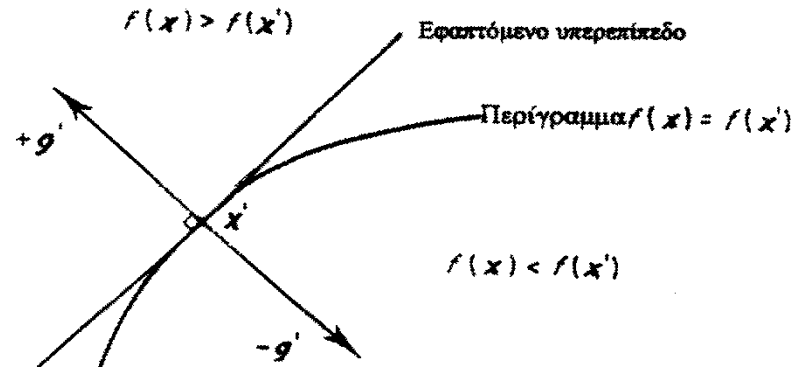
$$q(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T\mathbf{G}\mathbf{x} + \mathbf{b}^T\mathbf{x} + c \qquad (1.2.12)$$

με δεδομένα **G**, **b**, c, όπου ο **G** είναι συμμετρικός ή από

$$q(\mathbf{x}) = \frac{1}{2}(\mathbf{x}-\mathbf{x}')^T\mathbf{G}(\mathbf{x}-\mathbf{x}') + c' \qquad (1.2.13)$$

με δεδομένα **G**, $\mathbf{x}'$, c, όπου $\mathbf{G}\mathbf{x}' = -\mathbf{b}$ και $c' = c - \frac{1}{2}\mathbf{x}'^T\mathbf{G}\mathbf{x}'$. Κάνοντας χρήση του κανόνα (Π1.1) με $\mathbf{u}=\mathbf{x}$ και $\mathbf{v} = \mathbf{G}\mathbf{x}$ καθώς και του τύπου (Π1.3) έχουμε για το διάνυσμα κλίσης της (1.2.12)

$$\nabla q(\mathbf{x}) = \frac{1}{2}(\mathbf{G} + \mathbf{G}^T)\mathbf{x} + \mathbf{b} = \mathbf{G}\mathbf{x} + \mathbf{b} \qquad (1.2.14)$$



Σχήμα 1.2.3: Ιδιότητες του διανύσματος κλίσης.

όπου έγινε χρήση του $\mathbf{G} = \mathbf{G}^T$ αφού ο **G** θεωρήθηκε συμμετρικός. Κατά παρόμοιο τρόπο έχουμε $\nabla^2 q = \mathbf{G}$. Άρα η q έχει σταθερό Χεσιανό πίνακα **G**, ενώ η κλίση της είναι μια γραμμική συνάρτηση του **x**. Κατά συνέπεια, για δεδομένα σημεία **x**', **x**" με κλίσεις **g**', **g**" έχουμε

$$\mathbf{g}'' - \mathbf{g}' = \mathbf{G}(\mathbf{x}'' - \mathbf{x}'). \qquad (1.2.15)$$

Η **σειρά Taylor** μιάς συνάρτησης μιάς μεταβλητής f(x) γύρω από ένα σημείο x' έχει ως γνωστόν ως εξής

$$f(x' + \delta) = f(x') + \delta f'(x') + \frac{1}{2}\delta^2 f''(x') + \ldots \qquad (1.2.16)$$

όπου δ είναι η απόκλιση από το σημείο x'. Θεωρούμε την συνάρτηση μιάς μεταβλητής f(α) που προκύπτει από μία συνάρτηση πολλαπλών μεταβλητών f(**x**) για **x** επί γραμμής (1.2.1), δηλαδή f(α) = f(**x**(α)). Εφαρμόζοντας την (1.2.16) για την f(α) γύρω από το σημείο α' = 0 έχουμε

$$f(\alpha) = f(0) + \alpha f'(0) + \frac{1}{2}\alpha^2 f''(0) + \ldots$$

Κάνοντας χρήση των (1.2.1), (1.2.9), (1.2.10) λαμβάνουμε από την εξίσωση αυτή

$$f(\mathbf{x}' + \alpha\mathbf{s}) = f(\mathbf{x}') + \alpha\mathbf{s}^T\nabla f(\mathbf{x}') + \frac{1}{2}\alpha^2\mathbf{s}^T\nabla^2 f(\mathbf{x}')\mathbf{s} + \ldots \qquad (1.2.17)$$

ή, ορίζοντας την διανυσματική απόκλιση $\boldsymbol{\delta} = \alpha\mathbf{s}$,

$$f(\mathbf{x}' + \boldsymbol{\delta}) = f(\mathbf{x}') + \boldsymbol{\delta}^T\nabla f(\mathbf{x}') + \frac{1}{2}\boldsymbol{\delta}^T\nabla^2 f(\mathbf{x}')\boldsymbol{\delta} + \ldots \qquad (1.2.18)$$

Η (1.2.18) είναι προφανώς (βλ. (1.2.16)) η γενικευμένη σειρά Taylor για συναρτήσεις πολλαπλών μεταβλητών.

Εφαρμόζοντας την (1.2.18) στη διανυσματική συνάρτηση ∇f, στήλη προς στήλη, έχουμε επίσης την σειρά Taylor της κλίσης της f

$$\nabla f(\mathbf{x}' + \boldsymbol{\delta}) = \nabla f(\mathbf{x}') + \nabla^2 f(\mathbf{x}')\,\boldsymbol{\delta} + \ldots \qquad (1.2.19)$$

Παραλείποντας τους όρους άνω του πρώτου βαθμού, βλέπουμε ότι για αρκούντως μικρό **δ** λαμβάνουμε την (1.2.14), πράγμα που σημαίνει ότι κάθε διπλά παραγωγίσιμη συνάρτηση προσεγγίζει μια τετραγωνική συνάρτηση σε μια αρκούντως μικρή περιοχή κάθε σημείου **x**'.

Προβλήματα

1.1. Αποδείξτε τους ακόλουθους τύπους παραγωγισμού συναρτήσεων διανυσμάτων ή διανυσματικών συναρτήσεων:

$$\nabla[\mathbf{u}(\mathbf{x})^T\mathbf{v}(\mathbf{x})] = [\nabla\mathbf{u}(\mathbf{x})]\mathbf{v}(\mathbf{x}) + [\nabla\mathbf{v}(\mathbf{x})]\mathbf{u}(\mathbf{x}) \qquad (\Pi 1.1)$$

$$\nabla [u(\mathbf{x})\mathbf{v}(\mathbf{x})] = [\nabla u(\mathbf{x})]\mathbf{v}(\mathbf{x})^T + [\nabla \mathbf{v}(\mathbf{x})]u(\mathbf{x}) \qquad (\Pi 1.2)$$

$$\nabla (\mathbf{A}\mathbf{x}) = \mathbf{A}^T \qquad (\Pi 1.3)$$

$$\nabla (\mathbf{a}^T\mathbf{x}) = \mathbf{a} \qquad (\Pi 1.4)$$

$$\nabla (\mathbf{x}^T\mathbf{A}) = \mathbf{A}^T \qquad (\Pi 1.5)$$

$$\nabla (\mathbf{x}^T\mathbf{a}) = \mathbf{a} \qquad (\Pi 1.6)$$

$$\nabla (\mathbf{x}^T\mathbf{A}\mathbf{x}) = (\mathbf{A}+\mathbf{A}^T)\mathbf{x} \qquad (\Pi 1.7)$$

$$\nabla f[\mathbf{z}(\mathbf{w}(...(\mathbf{y}(\mathbf{x}))))] = \frac{\partial \mathbf{y}}{\partial \mathbf{x}} \cdots \frac{\partial \mathbf{z}}{\partial \mathbf{w}} \frac{\partial f}{\partial \mathbf{z}} \qquad (\Pi 1.8)$$

$$\nabla \mathbf{f}[\mathbf{z}(\mathbf{w}(...(\mathbf{y}(\mathbf{x}))))] = \frac{\partial \mathbf{y}}{\partial \mathbf{x}} \cdots \frac{\partial \mathbf{z}}{\partial \mathbf{w}} \frac{\partial \mathbf{f}}{\partial \mathbf{z}}\,. \qquad (\Pi 1.9)$$

Λύση

$$\nabla (\mathbf{u}^T\mathbf{v}) = \nabla (u_1v_1 + ... + u_mv_m) \overset{(1.2.3)}{=} \left[\sum_{j=1}^{m} \frac{\partial u_j}{\partial x_i} v_j + u_j \frac{\partial v_j}{\partial x_i}\right]_i \overset{(1.2.6)}{=}$$

$$= (\nabla \mathbf{u})\mathbf{v} + (\nabla \mathbf{v})\mathbf{u}.$$

$$\nabla (u\mathbf{v}) = \nabla [uv_i]_i \overset{(1.2.6)}{=} \left[\frac{\partial u}{\partial x_i} v_j + \frac{\partial v_j}{\partial x_i} u\right]_{ij} = \left[\frac{\partial u}{\partial x_i} v_j\right]_{ij} + u\left[\frac{\partial v_j}{\partial x_i}\right]_{ij} \overset{(1.2.6)}{=}$$

$$= (\nabla u)\mathbf{v}^T + (\nabla \mathbf{v})u.$$

$$\nabla (\mathbf{A}\mathbf{x}) = \nabla \left[\sum_{j=1}^{n} a_{ij}x_j\right]_i \overset{(1.2.6)}{=} [a_{ji}]_{ij} = \mathbf{A}^T.$$

$$\nabla (\mathbf{a}^T\mathbf{x}) \overset{(\Pi 1.1)}{=} \mathbf{a}.$$

$$\nabla (\mathbf{x}^T\mathbf{A}) = \nabla (\mathbf{A}^T\mathbf{x})^T \overset{(\Pi 1.3)}{=} \mathbf{A}^T.$$

$$\nabla (\mathbf{x}^T\mathbf{a}) = \nabla (\mathbf{a}^T\mathbf{x}) \overset{(\Pi 1.4)}{=} \mathbf{a}.$$

$$\nabla (\mathbf{x}^T\mathbf{A}\mathbf{x}) = \nabla [\mathbf{x}^T(\mathbf{A}\mathbf{x})] \overset{(\Pi 1.1),\,(\Pi 1.3)}{=} \mathbf{A}\mathbf{x} + \mathbf{A}^T\mathbf{x} = (\mathbf{A}+\mathbf{A}^T)\mathbf{x}.$$

$$\nabla f[\mathbf{y}(\mathbf{x})] = \left[\frac{\partial f[y_1(\mathbf{x}),\ldots,y_m(\mathbf{x})]}{\partial x_i}\right]_i = \left[\frac{\partial f}{\partial y_1}\frac{\partial y_1}{\partial x_i}+\ldots+\frac{\partial f}{\partial y_m}\frac{\partial y_m}{\partial x_i}\right]_i =$$

$$= \left[\frac{\partial \mathbf{y}^T}{\partial x_i}\frac{\partial f}{\partial \mathbf{y}}\right]_i = \frac{\partial \mathbf{y}}{\partial \mathbf{x}}\frac{\partial f}{\partial \mathbf{y}}. \qquad (\Pi 1.10)$$

$$\nabla f[\mathbf{w}(\mathbf{y}(\mathbf{x}))] \overset{(\Pi 1.10)}{=} \frac{\partial \mathbf{y}}{\partial \mathbf{x}}\frac{\partial f[\mathbf{w}(\mathbf{y})]}{\partial \mathbf{y}} \overset{(\Pi 1.10)}{=} \frac{\partial \mathbf{y}}{\partial \mathbf{x}}\frac{\partial \mathbf{w}}{\partial \mathbf{y}}\frac{\partial f}{\partial \mathbf{w}} \text{ κ.ο.κ}$$

$$\nabla \mathbf{f}[\mathbf{y}(\mathbf{x})] = \left[\frac{\partial f_j[y_1(\mathbf{x}),\ldots,y_m(\mathbf{x})]}{\partial x_i}\right]_{ij} = \left[\frac{\partial f_j}{\partial y_1}\frac{\partial y_1}{\partial x_i}+\ldots+\frac{\partial f_j}{\partial y_m}\frac{\partial y_m}{\partial x_i}\right]_{ij} =$$

$$= \left[\frac{\partial \mathbf{y}^T}{\partial x_i}\frac{\partial f_j}{\partial \mathbf{y}}\right]_{ij} = \frac{\partial \mathbf{y}}{\partial \mathbf{x}}\frac{\partial \mathbf{f}}{\partial \mathbf{y}}.$$

1.2. Αποδείξτε ότι σε κάποιο σημείο $\mathbf{x}'$, όπου $\mathbf{g}' \neq \mathbf{0}$, το διάνυσμα κατεύθυνσης $\mathbf{s} = \mathbf{g}'/\|\mathbf{g}'\|_2$ έχει τη μέγιστη κλίση μεταξύ όλων των κατευθύνσεων για τις οποίες $\mathbf{s}^T\mathbf{s} = 1$ (**διάνυσμα μέγιστης ανάβασης**) ενώ το διάνυσμα κατεύθυνσης $\mathbf{s} = -\mathbf{g}'/\|\mathbf{g}'\|_2$ έχει την ελάχιστη κλίση μεταξύ όλων των κατευθύνσεων για τις οποίες $\mathbf{s}^T\mathbf{s} = 1$ (**διάνυσμα μέγιστης κατάβασης**). Αποδείξτε ακόμη ότι τα $\pm\mathbf{g}'$ είναι ορθογώνια στο περίγραμμα και το εφαπτόμενο επίπεδο στο $\mathbf{x}'$.

Λύση (βλ. Σχ. 1.2.3)

(1.2.9): $\dfrac{df}{d\alpha} = \mathbf{g}'^T\mathbf{s} = \|\mathbf{g}'\|_2 \cdot \|\mathbf{s}\|_2 \cdot \cos\varphi = \|\mathbf{g}'\|_2 \cdot \cos\varphi$, το οποίο είναι μέγιστο αν $\cos\varphi = 1$ $\Rightarrow \varphi = 0 \Rightarrow \mathbf{s}$ έχει κατεύθυνση $\mathbf{g}'$ και μέγεθος 1, άρα $\mathbf{s} = \mathbf{g}'/\|\mathbf{g}'\|_2$.

Ελάχιστη κλίση έχουμε για $\cos\varphi = -1 \Rightarrow \varphi = 180^o \Rightarrow \mathbf{s}$ έχει κατεύθυνση $-\mathbf{g}'$ και μέγεθος 1, άρα $\mathbf{s} = -\mathbf{g}'/\|\mathbf{g}'\|_2$.

Εξ ορισμού κατά μήκος περιγράμματος έχουμε $\dfrac{df}{d\alpha} = 0 = \mathbf{g}'^T\mathbf{s}_\pi$, άρα τα διανύσματα $\mathbf{s}_\pi$ και $\mathbf{g}'$ είναι ορθογώνια.

ΜΕΡΟΣ ΠΡΩΤΟ:

ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΧΩΡΙΣ ΠΕΡΙΟΡΙΣΜΟΥΣ

2. ΓΕΝΙΚΗ ΜΕΘΟΔΟΛΟΓΙΑ

2.1 Συνθήκες για Τοπικά Ελάχιστα

Στο Πρώτο Μέρος θεωρούμε το πρόβλημα εύρεσης μιας τοπικής λύσης του προβλήματος

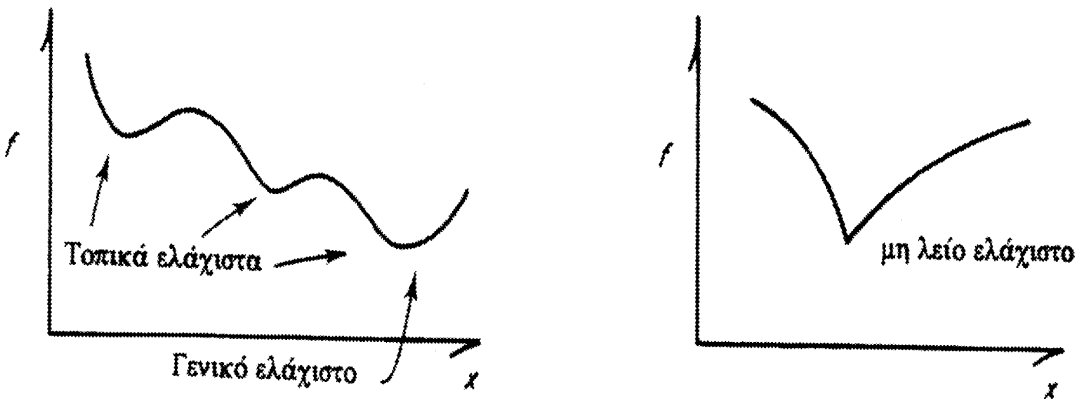
$$\text{Ελαχιστοποίηση } f(\mathbf{x}) \text{ , } \mathbf{x} \in R^n. \qquad (2.1.1)$$

Η f(**x**) είναι η **αντικειμενική συνάρτηση** (ή **αντικειμενικό κριτήριο**) ενώ το σημείο ελάχιστου ή **ελάχιστο** σημειώνεται **x***. **Προβλήματα μεγιστοποίησης** μπορούν να αναχθούν στο (2.1.1) με τον απλό μετασχηματισμό

$$\max_{\mathbf{x}} f(\mathbf{x}) = -\min_{\mathbf{x}} [-f(\mathbf{x})]. \qquad (2.1.2)$$

Συνήθως υποθέτουμε ότι το **x*** υπάρχει, είναι μοναδικό και μπορεί να προσδιορισθεί από μια χρησιμοποιούμενη μέθοδο. Για πολλά όμως προβλήματα οι παραδοχές αυτές μπορεί να μην εκπληρούνται. Το **x*** μπορεί να μην υπάρχει αν η f δεν είναι κάτω φραγμένη (π.χ. $f = x^3$) και, σε κάποιες περιπτώσεις, ακόμη και αν η f είναι κάτω φραγμένη (π.χ. $f=e^{-x}$). Αν το **x*** υπάρχει, μπορεί να μην είναι μοναδικό (π.χ. f = max{–x–1, 0, x–1} ή f = cosx). Μπορούν επίσης να υπάρχουν **τοπικά ελάχιστα** που δεν είναι **γενικά ελάχιστα** (ή **απόλυτα ελάχιστα**), βλ. Σχ. 2.1.1. Τοπικά ελάχιστα μπορεί να υπάρχουν και να παρουσιάζουν ενδιαφέρον ακόμη και αν η f δεν είναι κάτω φραγμένη (π.χ. $f = x^3 - 3x$). Ο προσδιορισμός **γενικών ελάχιστων** δεν είναι θεωρητικά ή πρακτικά εφικτός με βεβαιότητα στην γενική περίπτωση. Ένας πρακτικός τρόπος είναι ο προσδιορισμός πολλών (ή όλων των) τοπικών ελάχιστων και η επιλογή του βέλτιστου εξ αυτών. Προσοχή όμως: Η ύπαρξη τοπικών ελάχιστων δεν σημαίνει ότι υπάρχει αναγκαστικά γενικό ελάχιστο, π.χ. $f(x) = (0.5 + x^2)\, e^{-x^2}$.

Για ένα τοπικό ελάχιστο **x*** της συνάρτησης f(**x**) ισχύει f(**x**) ≥ f(**x***) για κάθε **x** από μια αρκούντως μικρή περιοχή περί το **x***. Ένα **αυστηρό τοπικό ελάχιστο** ικανοποιεί f(**x**) > f(**x***) για κάθε **x**≠**x*** από μια αρκούντως μικρή περιοχή περί το **x***.



Σχήμα 2.1.1: Τύποι ελάχιστων.

Θα θεωρήσουμε μόνο συναρτήσεις όπου οι πρώτες και δεύτερες παράγωγοι υπάρχουν και είναι συνεχείς, αποκλείοντας έτσι μη λεία ελάχιστα όπως αυτό του Σχ. 2.1.1. Θα χρησιμοποιούμε τις παραστάσεις $\mathbf{g}(\mathbf{x}) = \nabla f(\mathbf{x})$, $\mathbf{G}(\mathbf{x}) = \nabla^2 f(\mathbf{x})$, $f^* = f(\mathbf{x}^*)$, $\mathbf{g}^* = \mathbf{g}(\mathbf{x}^*)$, $f^{(k)} = f(\mathbf{x}^{(k)})$, $\mathbf{g}^{(k)} = \mathbf{g}(\mathbf{x}^{(k)})$ κλπ.

Για τον προσδιορισμό συνθηκών ελαχιστοποίησης παρατηρούμε (βλ. Σχ. 2.1.2) ότι για κάθε γραμμή $\mathbf{x}(\alpha) = \mathbf{x}^*+\alpha\mathbf{s}$ πρέπει η συνάρτηση $f(\alpha) = f(\mathbf{x}(\alpha))$ να έχει στο $\mathbf{x}^*$ (δηλαδή για $\alpha = 0$) μηδενική κλίση και μη αρνητική καμπυλότητα, δηλαδή, με τις (1.2.9), (1.2.10), έχουμε $\mathbf{s}^T\mathbf{g}^*=0$ και $\mathbf{s}^T\mathbf{G}^*\mathbf{s}\geq 0$ για κάθε $\mathbf{s}$. Θέτοντας διαδοχικά $\mathbf{s}=\mathbf{e}_1$, $\mathbf{s}=\mathbf{e}_2$ κλπ. στο ως άνω εσωτερικό γινόμενο, έχουμε αντίστοιχα $\partial f / \partial x_1 = 0$, $\partial f / \partial x_2 = 0$ κλπ. Καταλήγουμε έτσι στις ακόλουθες **αναγκαίες συνθήκες τοπικού ελάχιστου x***

$$\mathbf{g}^* = \mathbf{0} \tag{2.1.3}$$

$$\mathbf{s}^T\mathbf{G}^*\mathbf{s} \geq 0 \quad \forall \mathbf{s}. \tag{2.1.4}$$

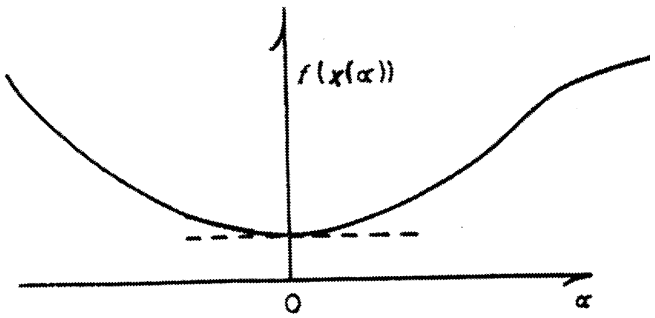
Η (2.1.3) ονομάζεται **αναγκαία συνθήκη πρώτου βαθμού** και η (2.1.4) **αναγκαία συνθήκη δεύτερου βαθμού**. Η τελευταία συνθήκη είναι εξ ορισμού ταυτόσημη με την συνθήκη ότι ο **G*** είναι ένας **θετικά ημιορισμένος πίνακας**.

Θεώρημα 2.1.1: Ικανές συνθήκες για ένα αυστηρό τοπικό ελάχιστο **x*** είναι η (2.1.3) και **G*** θετικά ορισμένος, δηλαδή

$$\mathbf{s}^T\mathbf{G}^*\mathbf{s} > 0 \quad \forall \mathbf{s} \neq \mathbf{0}. \tag{2.1.5}$$

■

Μια ικανή και αναγκαία συνθήκη για να είναι ένας πίνακας **G** θετικά (αρνητικά) ορισμένος είναι να έχει όλες τις ιδιοτιμές πραγματικές και θετικές (αρνητικές), ενώ για θετικά (αρνητικά) ημιορισμένους πίνακες όλες οι ιδιοτιμές είναι πραγματικές και μη αρνητικές (μη θετικές). Μια άλλη ικανή και αναγκαία συνθήκη κάνει χρήση των οριζουσών D_i, $i = 1,\ldots, n$, όλων των κυρίων υποπινάκων του εν λόγω πίνακα **G.** Συγκεκριμένα ισχύουν τα εξής:



Σχήμα 2.1.2: Μηδενική κλίση και μη αρνητική καμπυλότητα για α=0.

$\mathbf{G} > \mathbf{0} \Leftrightarrow D_i > 0 \quad \forall i$

$\mathbf{G} \geq \mathbf{0} \Leftrightarrow D_i \geq 0 \quad \forall i$

$\mathbf{G} < \mathbf{0} \Leftrightarrow (-1)^i D_i > 0 \quad \forall i$

$\mathbf{G} \leq \mathbf{0} \Leftrightarrow (-1)^i D_i \geq 0 \quad \forall i.$

Επί τη ευκαιρία θυμίζουμε ότι ισχύουν οι ισοδυναμίες $\mathbf{G} > \mathbf{0} \Leftrightarrow \mathbf{G}^{-1} > \mathbf{0}$ και $\mathbf{G} < \mathbf{0} \Leftrightarrow \mathbf{G}^{-1} < \mathbf{0}$.

Παράδειγμα 2.1.1: Θεωρούμε την συνάρτηση (1.2.2) με παραγώγους (1.2.4). Από την (2.1.3) έχουμε την μοναδική λύση $\mathbf{x}^* = [1\ 1]^T$, και επειδή ο

$$\mathbf{G}^* = \begin{bmatrix} 802 & -400 \\ -400 & 200 \end{bmatrix}$$

είναι θετικά ορισμένος, συνεπάγεται ότι το ως άνω $\mathbf{x}^*$ είναι το μοναδικό τοπικό ελάχιστο. Καθότι $f^*=0$ και $f(\mathbf{x}) \geq 0\ \forall \mathbf{x}$, συμπεραίνουμε ότι το $\mathbf{x}^*$ είναι γενικό ελάχιστο.

■

Παράδειγμα 2.1.2: Θεωρούμε την ελαχιστοποίηση της γενικής τετραγωνικής συνάρτησης (1.2.12) με θετικά ορισμένο $\mathbf{G}$. Με τις (1.2.14) και (2.1.3) έχουμε

$$\mathbf{Gx} + \mathbf{b} = \mathbf{0} \Rightarrow \mathbf{x}^* = -\mathbf{G}^{-1}\mathbf{b} = \mathbf{x}'$$

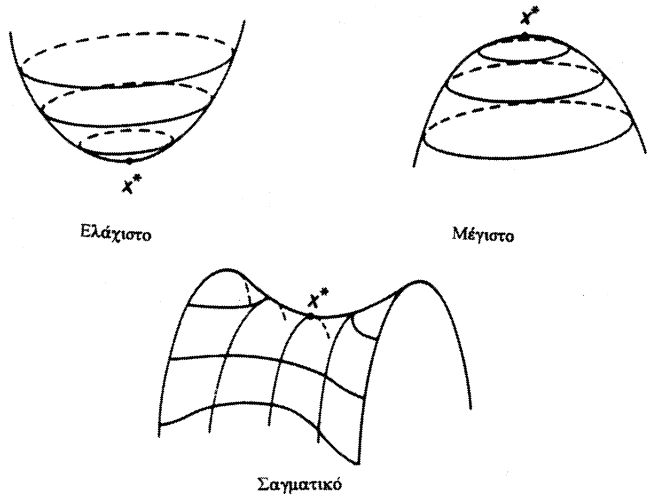
όπου $\mathbf{x}^*$ είναι η μοναδική λύση και επειδή $\nabla^2\mathbf{q} = \mathbf{G} > \mathbf{0}$, πρόκειται για το μοναδικό ελάχιστο. Καθότι $q(\mathbf{x}) \geq q(\mathbf{x}^*)\ \forall \mathbf{x}$ (βλ. (1.2.13)), πρόκειται για απόλυτο ελάχιστο.

■

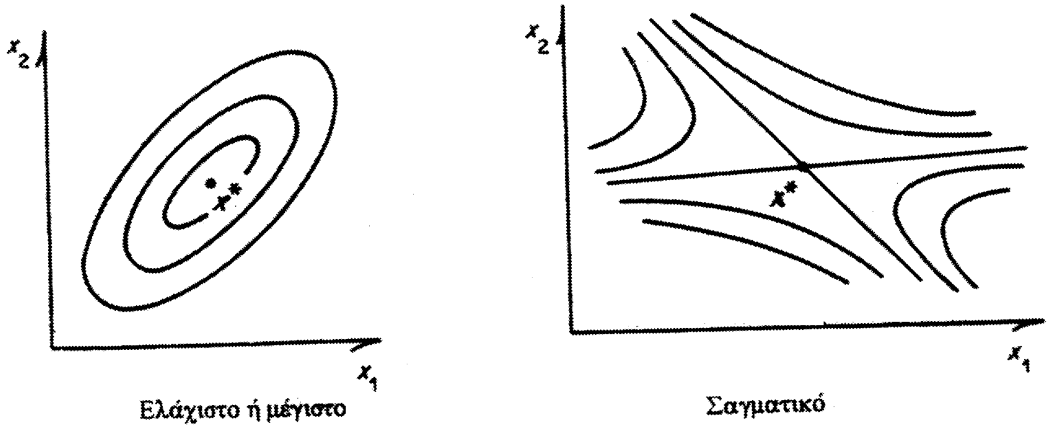
Η διαφορά μεταξύ των αναγκαίων συνθηκών και ικανών συνθηκών ελαχιστοποίησης υφίσταται μόνο σε μερικές περιπτώσεις μηδενικής καμπυλότητας. Επί παραδείγματι η $f(x) = x^4$ έχει ένα απόλυτο ελάχιστο για $x^*=0$, αλλά δεν ικανοποιεί τις ικανές συνθήκες, ενώ η $f(x) = -x^4$ έχει ένα απόλυτο μέγιστο για $x^*=0$, αλλά ικανοποιεί τις αναγκαίες συνθήκες ελαχιστοποίησης.

Τα σημεία που ικανοποιούν την συνθήκη (2.1.3) μόνη, λέγονται **στάσιμα σημεία**. Ας σημειωθεί ότι η (2.1.3) ικανοποιείται από τοπικά ελάχιστα, **τοπικά μέγιστα**, και **σαγματικά σημεία**, βλ. Σχ. 2.1.3 και υψομετρικές καμπύλες στο Σχ. 2.1.4. Παρατηρούμε στο Σχ. 2.1.3 ότι η καμπυλότητα στο $\mathbf{x}^*$ όλων των γραμμών είναι θετική (δηλ. $\mathbf{G}^* > \mathbf{0}$) για τοπικά ελάχιστα, αρνητική ($\mathbf{G}^* < \mathbf{0}$) για τοπικά μέγιστα και μεικτή ($\mathbf{G}^*$ μη ορισμένος) για σαγματικά σημεία.

Η αναγκαία συνθήκη (2.1.3) αποτελεί ένα μη γραμμικό σύστημα n εξισώσεων με n αγνώστους, τις συνιστώσες x_i^*. Το σύστημα μπορεί να έχει μια μοναδική ή καμιά ή πολλαπλές λύσεις. Στην τελευταία περίπτωση έχουμε πολλά υποψήφια ελάχιστα, οπότε οι (2.1.4) ή (2.1.5) μπορούν να χρησιμοποιηθούν για τον διαχωρισμό των τοπικών ελάχιστων από τα άλλα στάσιμα σημεία.



Σχήμα 2.1.3: Τύποι στάσιμων σημείων.



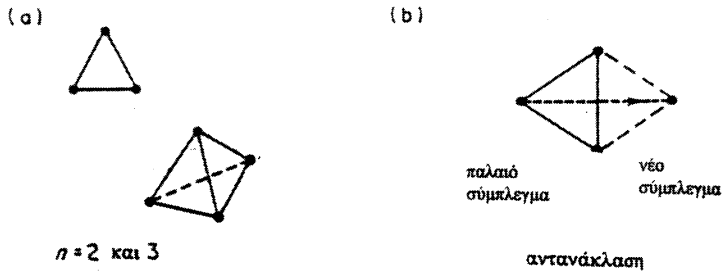
Σχήμα 2.1.4: Περιγράμματα στάσιμων σημείων.

Εννοείται ότι η αναλυτική λύση, μέσω επίλυσης του συστήματος εξισώσεων (2.1.3), είναι εφικτή μόνο για σχετικά απλά προβλήματα βελτιστοποίησης. Για το λόγο αυτό, θα ασχοληθούμε στο υπόλοιπο του Πρώτου Μέρους αποκλειστικά με μεθόδους **αριθμητικής επίλυσης** του προβλήματος (2.1.1).

2.2 Ευρετικές Μέθοδοι

Έχοντας τη δυνατότητα υπολογισμού τιμών της συνάρτησης f(**x**) για δεδομένα σημεία $\mathbf{x} \in R^n$, και ενδεχομένως και αντίστοιχων τιμών του διανύσματος κλίσης **g**(**x**), ποιές συστηματικές και γενικές μέθοδοι μπορούν να χρησιμοποιηθούν για τον αριθμητικό υπολογισμό ενός τοπικού ελάχιστου **x*** κατά τρόπο οικονομικό, δηλαδή με κατά το δυνατό λίγους υπολογισμούς;[1] Εκτός από τις μεθόδους με ισχυρό θεωρητικό υπόβαθρο που θα εξετάσουμε σε επόμενα κεφάλαια, υπάρχουν ευρετικές μέθοδοι, όπως π.χ.:

[1]Παραστατικά θεωρούμε για n=2 ότι βρισκόμαστε σε ορεινή περιοχή σε κατάσταση ομίχλης ή σκότους, εξοπλισμένοι με μετρητή υψομέτρου και έχοντας τη δυνατότητα να δούμε την μέγιστη κλίση στα πόδια μας. Επιθυμούμε να μεταβούμε στο χαμηλότερο σημείο της κοιλάδος (ή στην υψηλότερη κορυφή) με, κατά το δυνατόν, λίγες στάσεις και μετρήσεις.



Σχήμα 2.2.1 Σύμπλεγμα σημείων για n=2 και n=3.

a) Τυχαία επιλογή σημείων.

b) Μέθοδοι συμπλέγματος σημείων (simplex ή complex).

Τα πλεονεκτήματα αυτών των μεθόδων είναι:

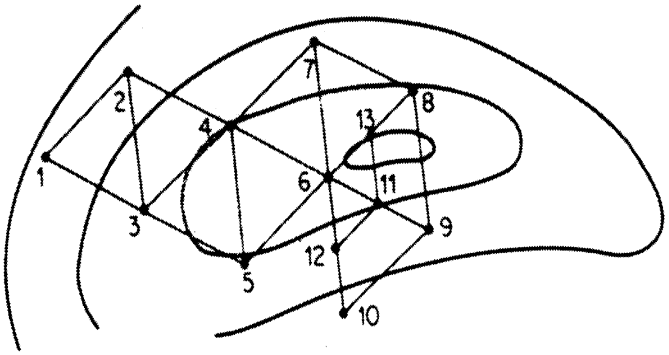
- Δεν χρειάζονται υπολογισμό κλίσης **g**(**x**).
- Αυξημένη πιθανότητα εντοπισμού ενός "καλού" τοπικού ελάχιστου ακόμη και για συναρτήσεις f(**x**) με θόρυβο.

Το σοβαρότερο μειονέκτημα είναι ότι συνήθως απαιτούν σημαντικά υψηλότερο υπολογιστικό φόρτο από άλλες μεθόδους με ισχυρό θεωρητικό υπόβαθρο, σε μερικές μάλιστα ο υπολογιστικός φόρτος αυξάνει εκθετικά (π.χ. αναλογικά με 2^n), πράγμα που τις αχρηστεύει σε προβλήματα με μεγάλο αριθμό μεταβλητών.

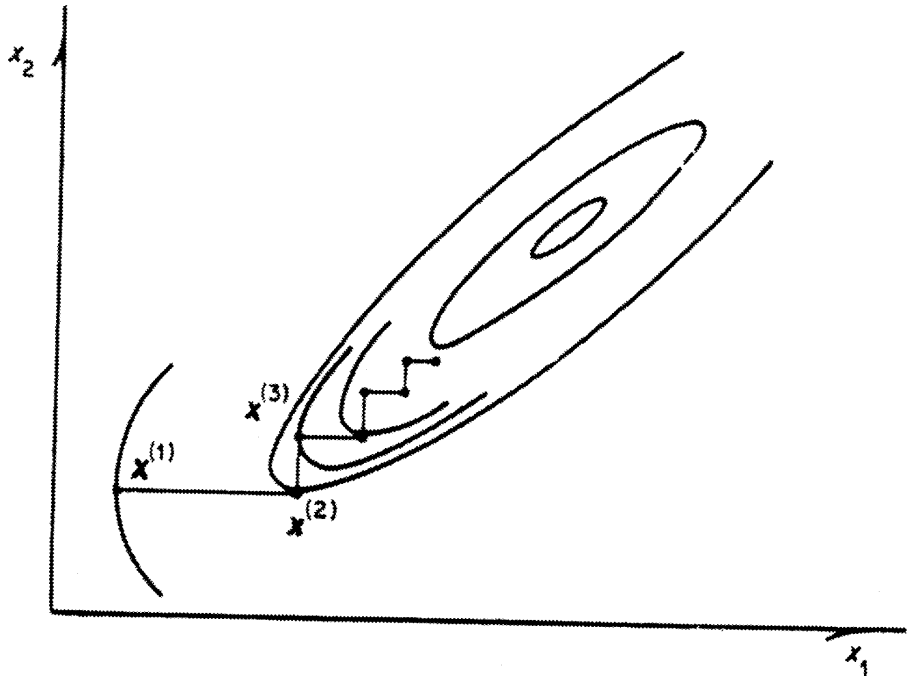
Για την παρουσίαση μιας ευρετικής μεθόδου, της **μεθόδου συμπλέγματος σημείων**, θεωρούμε n+1 σημεία εκκίνησης σε ίσες αποστάσεις μεταξύ τους στο χώρο R^n (βλ. Σχ. 2.2.1a) και υπολογίζουμε τις αντίστοιχες τιμές τις αντικειμενικής συνάρτησης. Στην πρώτη επανάληψη, το σημείο με το μεγαλύτερο κόστος αντικαθίσταται από την αντανάκλασή του στο κέντρο των υπόλοιπων n σημείων (Σχ. 2.2.1b) και έτσι δημιουργείται ένα νέο σύμπλεγμα. Αφού υπολογιστεί το κόστος του νέου σημείου, η διεργασία επαναλαμβάνεται. Αν όμως το νέο σημείο είναι και πάλι το χειρότερο της ομάδος, οδηγούμαστε σε ταλαντώσεις χωρίς πρόοδο, γι' αυτό αντικαθίσταται στην επόμενη επανάληψη το χειρότερο σημείο της ομάδας εξαιρουμένου του νέου σημείου. Μετά από έναν αριθμό επαναλήψεων δεν θα είναι δυνατή περαιτέρω πρόοδος, οπότε πρέπει να εισαχθεί κάποιος νέος κανόνας, π.χ. αν κάποιο σημείο **x**′ ανήκει στο σύμπλεγμα για περισσότερες από M επαναλήψεις, τότε το σύμπλεγμα συστέλλεται με τη μεταφορά των άλλων σημείων στο ήμισυ της απόστασής τους από το **x**′. Μια εμπειρικά κατάλληλη τιμή είναι $M = 1.65n + 0.05n^2$. Το Σχ. 2.2.2 αναπαριστά το αποτέλεσμα μερικών επαναλήψεων της μεθόδου σε ένα υποθετικό πρόβλημα δύο διαστάσεων.

Μιά άλλη ευρετική μέθοδος είναι η **μέθοδος εναλλασσόμενων μεταβλητών** (Σχ. 2.2.3) κατά την οποία επιχειρείται μέγιστη βελτίωση κατά μήκος ενός άξονα συντεταγμένων σε κάθε επανάληψη. Η μέθοδος αυτή είναι συνήθως πολύ λιγότερο αποτελεσματική από τις συναφείς μεθόδους των επόμενων κεφαλαίων.

Άλλες ευρετικές μέθοδοι αναπτύσσονται στο Κεφάλαιο 10.



Σχήμα 2.2.2: Μέθοδος συμπλέγματος σημείων σε δύο διαστάσεις.



Σχήμα 2.2.3: Μέθοδος εναλλασσόμενων μεταβλητών.

2.3 Χρήσιμες Ιδιότητες Αλγορίθμων

Επαναληπτικοί αλγόριθμοι υπολογίζουν επαναληπτικά $\mathbf{x}^{(k)}$ που συγκλίνουν σε τοπικό ελάχιστο $\mathbf{x}^*$. Επειδή ο απόλυτα ακριβής προσδιορισμός του $\mathbf{x}^*$ δεν είναι συνήθως εφικτός, ο αλγόριθμος διακόπτεται όταν ικανοποιείται ένα **κριτήριο σύγκλισης**. Το αντικειμενικό κριτήριο $f(\mathbf{x}^{(k)})$ βελτιώνεται σε κάθε επανάληψη. Αλγόριθμοι που συγκλίνουν στο $\mathbf{x}^*$ από οποιοδήποτε **σημείο εκκίνησης** $\mathbf{x}^{(1)}$ έχουν την ιδιότητα **καθολικής σύγκλισης**. Ιδιότητες **τοπικής σύγκλισης** αναφέρονται στον ρυθμό σύγκλισης του λάθους

$$\mathbf{h}^{(k)} = \mathbf{x}^{(k)} - \mathbf{x}^* \qquad (2.3.1)$$

προς το $\mathbf{0}$ σε μια περιοχή γύρω από το $\mathbf{x}^*$. Αν

$$\left\|\mathbf{h}^{(k+1)}\right\| / \left\|\mathbf{h}^{(k)}\right\|^p \to a, \quad a > 0$$

τότε έχουμε **βαθμό σύγκλισης** p. Συνήθως έχουμε p=1 (**σύγκλιση πρώτου βαθμού** ή **γραμμική σύγκλιση**) ή p=2 (**σύγκλιση δευτέρου βαθμού** ή **τετραγωνική σύγκλιση**).

Γραμμική σύγκλιση είναι ικανοποιητική αν $a \leq 1/4$. Αν όμως $a \to 0$ έχουμε **υπεργραμμική σύγκλιση**. Η πρακτική σημασία των ιδιοτήτων τοπικής σύγκλισης περιορίζεται από το γεγονός ότι ισχύουν σε μια περισσότερο ή λιγότερο μικρή περιοχή περί το ζητούμενο ελάχιστο $\mathbf{x}^*$. Για το λόγο αυτό, πέραν των θεωρητικών αποτελεσμάτων ρυθμού σύγκλισης κάθε μεθόδου, ιδιαίτερα σημαντική, ή και καθοριστική, είναι η αποτελεσματικότητά της σε μια σειρά δοκιμαστικών προβλημάτων που συνήθως χρησιμοποιούνται για την πειραματική σύγκριση εναλλακτικών μεθόδων.

Οι επαναληπτικοί αλγόριθμοι που θα εξετάσουμε στη συνέχεια έχουν την ακόλουθη δομή, εκκινώντας από ένα δεδομένο από τον χρήστη σημείο εκκίνησης $\mathbf{x}^{(1)}$:

(a) Προσδιορισμός μιας κατεύθυνσης αναζήτησης $\mathbf{s}^{(k)}$.

(b) Προσδιορισμός βέλτιστου βήματος $\alpha^{(k)}$ που ελαχιστοποιεί την συνάρτηση $f(\alpha) = f(\mathbf{x}^{(k)}+\alpha\mathbf{s}^{(k)})$ ως προς α. (2.3.2)

(c) Υπολογισμός του νέου επαναληπτικού $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \alpha^{(k)}\mathbf{s}^{(k)}$.

(d) Αν ικανοποιείται το κριτήριο σύγκλισης, ο αλγόριθμος σταματά με $\mathbf{x}^{(k)} \approx \mathbf{x}^*$.

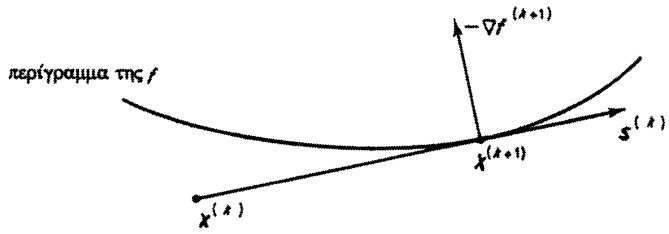
Τα βήματα (a), (b), (c), (d) εκτελούνται σε κάθε επανάληψη του αλγόριθμου. Εναλλακτικές μέθοδοι διαφέρουν στον τρόπο καθορισμού της κατεύθυνσης αναζήτησης $\mathbf{s}^{(k)}$ στο βήμα (a) του αλγορίθμου.

Όλες οι μέθοδοι χρησιμοποιούν στο βήμα (a) μια **κατεύθυνση κατάβασης** που ικανοποιεί την **ιδιότητα κατάβασης**

$$\mathbf{s}^{(k)T}\mathbf{g}^{(k)} < 0 \qquad (2.3.3)$$

που εγγυάται (βλ. (1.2.9)) ότι η κλίση df/dα είναι πάντοτε αρνητική για α=0 (εκτός αν το $\mathbf{x}^{(k)}$ είναι στάσιμο σημείο) και άρα η αντικειμενική συνάρτηση είναι βελτιώσιμη για κάποιο $\alpha^{(k)} > 0$.

Για τον προσδιορισμό κατευθύνσεων κατάβασης θα μπορούσε να χρησιμοποιηθεί η μέθοδος εναλλασσόμενων μεταβλητών που αναφέρθηκε στο Κεφάλαιο 2.2. Η μέθοδος όμως αυτή οδηγεί σε ταλαντώσεις (Σχ. 2.2.3) σε πολλές περιπτώσεις πρακτικών προβλημάτων και δεν είναι ιδιαίτερα αποτελεσματική. Μια άλλη απλή μέθοδος κατάβασης βασίζεται στην κατεύθυνση **μέγιστης κατάβασης** όπου $\mathbf{s}^{(k)} = -\mathbf{g}^{(k)}$. Παρόλο που η επιλογή αυτή φαίνεται καταρχήν εύλογη, στην πράξη μπορεί να οδηγήσει σε ταλαντώσεις παρόμοιες με την μέθοδο εναλλασσόμενων μεταβλητών του Κεφαλαίου 2.2 και πράγματι οδηγεί στα ίδια επαναληπτικά $\{\mathbf{x}^{(k)}\}$ όπως εκείνη στο ειδικό παράδειγμα του Σχ. 2.2.3. Έτσι η μέθοδος μέγιστης κατάβασης δεν ανήκει στις μεθόδους υψηλής αποτελεσματικότητας. Αποτελεσματικότερες μέθοδοι κατάβασης θα παρουσιαστούν στα Κεφάλαια 3, 4.



Σχήμα 2.3.1: Αναζήτηση επί γραμμής.

Στο βήμα (b) έχουμε το υποπρόβλημα **αναζήτησης επί γραμμής**. Η λύση του προβλήματος αυτού επιτυγχάνεται μέσω κατάλληλων αριθμητικών αλγορίθμων που χρησιμοποιούν υπολογισμούς τιμών της συνάρτησης $f(\alpha) = f(\mathbf{x}^{(k)}+\alpha\mathbf{s}^{(k)})$ και της παραγώγου της $f'(\alpha) = df(\mathbf{x}^{(k)}+\alpha\mathbf{s}^{(k)})/d\alpha$ για αντίστοιχες τιμές α, όπως θα δούμε λεπτομερώς στο Κεφάλαιο 2.6. Το **ελάχιστο γραμμής** $\alpha^{(k)}$ ικανοποιεί προφανώς τη συνθήκη $f'(\alpha^{(k)}) = 0$ και άρα με (1.2.9)

$$\nabla f^{(k+1)T}\mathbf{s}^{(k)} = 0 \qquad (2.3.4)$$

(βλ. Σχ. 2.3.1). Με άλλα λόγια, το διάνυσμα κλίσης $\nabla f^{(k+1)}$ στο ελάχιστο γραμμής είναι ορθογώνιο στην κατεύθυνση αναζήτησης $\mathbf{s}^{(k)}$. Στην πράξη, το επιλεγόμενο $\alpha^{(k)}$ δεν χρειάζεται να επιλύει με υψηλή ακρίβεια το υποπρόβλημα του βήματος (b), καθότι αυτό μπορεί να αυξήσει υπέρμετρα τον υπολογιστικό φόρτο χωρίς ουσιαστικό αντίκρισμα (βλ. Κεφάλαιο 2.5).

Το κριτήριο σύγκλισης (βήμα (d) του αλγόριθμου) μπορεί να έχει τη μορφή

$$\left\|\mathbf{g}^{(k)}\right\| \le \varepsilon \qquad (2.3.5)$$

ή εναλλακτικά

$$\left|x_i^{(k)} - x_i^{(k+1)}\right| \le \varepsilon_i \qquad \forall i \qquad (2.3.6)$$

ή ακόμη

$$f^{(k)} - f^{(k+1)} \le \varepsilon \qquad (2.3.7)$$

όπου οι σταθερές ανοχής ε πρέπει να καθοριστούν κατάλληλα από τον χρήστη, ανάλογα και με τις συνήθεις τιμές των μεταβλητών και του κριτηρίου κόστους του αντίστοιχου προβλήματος.

Σε περίπτωση που ο αναλυτικός υπολογισμός του διανύσματος κλίσης δεν είναι δυνατός, υπάρχει η δυνατότητα αριθμητικού υπολογισμού μέσω

$$g_i(\mathbf{x}) \approx [f(\mathbf{x}+\delta\mathbf{e}_i) - f(\mathbf{x})]/\delta \qquad (2.3.8)$$

(**πρόσθια διαφορά**) ή

$$g_i(\mathbf{x}) \approx \frac{1}{2}[f(\mathbf{x}+\delta\mathbf{e}_i) - f(\mathbf{x}-\delta\mathbf{e}_i)]/\delta \qquad (2.3.9)$$

(**κεντρική διαφορά**). Σημειώνουμε ότι οι τύποι (2.3.8) και (2.3.9) απαιτούν n και 2n, αντίστοιχα, υπολογισμούς της συνάρτησης f για κάθε υπολογισμό του διανύσματος κλίσης.

2.4 Τετραγωνικά Πρότυπα

Η κατεύθυνση μέγιστης κατάβασης βασίζεται σε γραμμική προσέγγιση της αντικειμενικής συνάρτησης, πράγμα που οδηγεί σε δυσκολίες, ιδιαίτερα όταν τα επαναληπτικά φθάσουν κοντά στο αναζητούμενο ελάχιστο $\mathbf{x}^*$. Πολύ μεγαλύτερη αποτελεσματικότητα παρουσιάζουν οι τετραγωνικές προσεγγίσεις κατά τον προσδιορισμό κατεύθυνσης αναζήτησης $\mathbf{s}^{(k)}$. Εχουμε τρείς ομάδες μεθόδων που στηρίζονται σε τετραγωνικά πρότυπα:

(a) Μέθοδος Newton (βλ. Κεφάλαιο 3)
(b) Μέθοδοι Σχεδόν-Newton (Quasi-Newton) (βλ. Κεφάλαιο 3)
(c) Μέθοδοι Συζυγών Κλίσεων (Conjugate Gradient) (βλ. Κεφάλαιο 4).

Μια μέθοδος έχει την ιδιότητα **τετραγωνικού τερματισμού** αν προσδιορίζει ακριβώς, σε πεπερασμένο και γνωστό αριθμό επαναλήψεων, το ελάχιστο μιας τετραγωνικής συνάρτησης. Οι μέθοδοι αυτές είναι στην πράξη ιδιαίτερα αποτελεσματικές ακόμη και για μη τετραγωνικές αντικειμενικές συναρτήσεις.

Ένας τρόπος που οδηγεί σε τετραγωνικό τερματισμό είναι η σύζευξη ενός συνόλου μη μηδενικών διανυσμάτων $\mathbf{s}^{(1)}$, $\mathbf{s}^{(2)}$,..., $\mathbf{s}^{(n)}$ μέσω ενός δεδομένου, θετικά ορισμένου πίνακα $\mathbf{G}$, δηλαδή

$$\mathbf{s}^{(i)T}\mathbf{G}\,\mathbf{s}^{(j)} = 0 \quad \forall i \neq j. \tag{2.4.1}$$

Κάθε μέθοδος που παράγει τέτοιες κατευθύνσεις όταν εφαρμόζεται σε τετραγωνική συνάρτηση με Χεσιανό $\mathbf{G}$, ονομάζεται **μέθοδος συζυγών κατευθύνσεων**.

Θεώρημα 2.4.1: Κάθε μέθοδος συζυγών κατευθύνσεων με ακριβή αναζήτηση επί γραμμής τερματίζει κατά την ελαχιστοποίηση μιας τετραγωνικής συνάρτησης το πολύ σε n επαναλήψεις, και κάθε $\mathbf{x}^{(k+1)}$ είναι το ελάχιστο του υποχώρου που ορίζεται από $\mathbf{x}^{(1)}$, $\mathbf{s}^{(1)}$, $\mathbf{s}^{(2)}$,..., $\mathbf{s}^{(k)}$, δηλαδή του συνόλου σημείων $\{\mathbf{x} \mid \mathbf{x} = \mathbf{x}^{(1)} + \sum_{j=1}^{k} \alpha_j \mathbf{s}^{(j)} \ \ \forall \alpha_j\}$.

■

Ανάλογα με τον τρόπο που επιτυγχάνεται η σύζευξη (2.4.1) έχουμε διάφορες μεθόδους συζυγών κατευθύνσεων, όπως θα δούμε στα Κεφάλαια 3 και 4.

2.5 Μέθοδοι Κατάβασης και Ευστάθεια

Η ακριβής αναζήτηση επί γραμμής είναι ιδιαίτερα δαπανηρή αλλά και περιττή, αφού το τελικό ζητούμενο είναι το ελάχιστο $\mathbf{x}^*$, ενώ τα ελάχιστα επί γραμμής είναι απλώς ενδιάμεσα βήματα προς τον στόχο αυτό. Η ευστάθεια της όλης αριθμητικής μεθόδου (δηλαδή η αποφυγή ταλαντώσεων) εξαρτάται από την **ουσιαστική** βελτίωση της αντικειμενικής συνάρτησης σε

κάθε επανάληψη, πράγμα που δεν επιτυγχάνεται αν ισχύει απλώς $f^{(k+1)} < f^{(k)}$, καθότι στην περίπτωση αυτή η βελτίωση μπορεί και να είναι μηδαμινή.

Σημειώνουμε $f(\alpha) = f(\mathbf{x}^{(k)}+\alpha\mathbf{s}^{(k)})$ οπότε $f(0)=f^{(k)}$ και η συνθήκη κατάβασης (2.3.3) δίδει $f'(0) < 0$. Συνθήκες που πρέπει να ικανοποιούνται από το υπό προσδιορισμό βήμα $\alpha^{(k)}$, έτσι ώστε να έχουμε ευστάθεια και αποτελεσματικότητα του αριθμητικού αλγόριθμου, είναι

$$f(\alpha^{(k)}) \le f(0) + \alpha^{(k)}\rho\, f'(0) \qquad (2.5.1)$$

$$\left| f'(\alpha^{(k)}) \right| \le -\sigma\, f'(0) \qquad (2.5.2)$$

όπου $\rho\in[0, 0.5]$, $\rho \le \sigma < 1$, με τυπικές τιμές $\rho=0.01$, $\sigma=0.5$. Το Σχ. 2.5.1 δίδει μια παραστατική επεξήγηση των συνθηκών αυτών. Η συνθήκη (2.5.1) δίδει ένα ανώτατο όριο στην αναζήτηση βήματος επί γραμμής, ενώ η (2.5.2) εγγυάται την μη αμελητέα βελτίωση της αντικειμενικής συνάρτησης στην αντίστοιχη επανάληψη. Οι συνθήκες (2.5.1), (2.5.2) χρησιμοποιούνται για τον έλεγχο σύγκλισης των αλγορίθμων αναζήτησης επί γραμμής του επόμενου κεφαλαίου.

2.6 Αλγόριθμοι Αναζήτησης Επί Γραμμής

Η αναζήτηση επί γραμμής (βήμα (b) του αλγορίθμου (2.3.2)) επιχειρεί την αριθμητική ελαχιστοποίηση της συνάρτησης μιας μεταβλητής $f(\alpha) = f(\mathbf{x}^{(k)}+\alpha\mathbf{s}^{(k)})$ ως προς α (τα $\mathbf{x}^{(k)}$ και $\mathbf{s}^{(k)}$ είναι δεδομένα). Προς το σκοπό αυτό μπορούν να υπολογίζονται, για συγκεκριμένες τιμές του βήματος α, οι αντίστοιχες τιμές $f(\alpha)$ και $f'(\alpha)= \mathbf{s}^{(k)^T}\nabla f(\mathbf{x}^{(k)}+\alpha\mathbf{s}^{(k)})$.

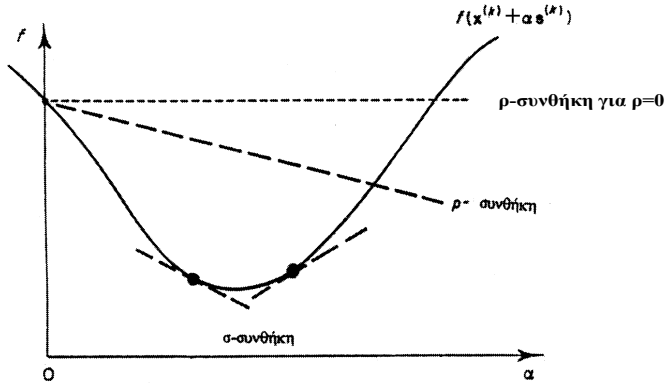
Ένας αλγόριθμος αναζήτησης επί γραμμής υπολογίζει, μέσω αντίστοιχων επαναλήψεων, μια ακολουθία $\{\alpha_i\}$ βημάτων και σταματά όταν μια επανάληψη προσδιορίζει ένα αποδεκτό σημείο, δηλαδή ένα σημείο που ικανοποιεί τις συνθήκες (2.5.1), (2.5.2). Εντός του αλγορίθμου αναζήτησης επί γραμμής εκτελείται καταρχήν μια **φάση εντοπισμού αγκύλης** που προσδιορίζει μια αγκύλη $[a_i,b_i]$ εμπεριέχουσα το υπό αναζήτηση ελάχιστο επί γραμμής. Η φάση αυτή ακολουθείται από την **φάση υποδιαίρεσης**, κατά την οποία η αγκύλη υποδιαιρείται κατ' επανάληψη οδηγώντας σε μια ακολουθία αγκυλών $[a_j,b_j]$ με συνεχώς μειούμενο μήκος.

Θα περιγράψουμε καταρχήν έναν αλγόριθμο εντοπισμού αγκύλης. Εκκινώντας από $\alpha_0=0$ με ένα δεδομένο βήμα α_1 (βλ. παρακάτω), η κάθε επανάληψη αρ. i της φάσης εντοπισμού αγκύλης εκτελεί τα ακόλουθα:

Υπολογισμός $f(\alpha_i)$.
Αν $f(\alpha_i) > f(0) + \alpha_i\, \rho\, f'(0)$ ή $f(\alpha_i) \ge f(\alpha_{i-1})$ τότε $a_i = \alpha_{i-1}$, $b_i = \alpha_i$ δίδει αγκύλη (γιατί;)[1].
Υπολογισμός $f'(\alpha_i)$.
Αν $\left| f'(\alpha_i) \right| \le -\sigma\, f'(0)$, τελείωσε η αναζήτηση επί γραμμής (γιατί;)[2].
Αν $f'(\alpha_i) \ge 0$ τότε $a_i = \alpha_{i-1}$, $b_i = \alpha_i$ δίδει αγκύλη (γιατί;).
$\alpha_{i+1} = \alpha_i + \tau_1 (\alpha_i - \alpha_{i-1})$ (όπου π.χ. $\tau_1 = 2$).

[1] "Δίδει αγκύλη" σημαίνει ότι εντός του προσδιορισθέντος διαστήματος $[a_i,b_i]$ υπάρχει σίγουρα ελάχιστο γραμμής.

[2] "Τελείωσε η αναζήτηση επί γραμμής" σημαίνει ότι για το προσδιορισθέν βήμα α_i ικανοποιούνται οι συνθήκες (2.5.1), (2.5.2) και άρα δεν απαιτείται φάση υποδιαίρεσης.



Σχήμα 2.5.1: Συνθήκες τερματισμού αναζήτησης επί γραμμής.

Από τα παραπάνω είναι φανερό ότι κάθε επανάληψη της φάσης εντοπισμού αγκύλης τελειώνει:

– Είτε με προσδιορισμό αγκύλης $[a_i,b_i]$ που εμπεριέχει ελάχιστο γραμμής, οπότε προχωρούμε στην φάση υποδιαίρεσης,
– Είτε με προσδιορισμό βήματος α_i που ικανοποιεί τις συνθήκες (2.5.1), (2.5.2), οπότε $\alpha^{(k)} = \alpha_i$, δηλαδή τελειώνει η αναζήτηση επί γραμμής χωρίς να χρειασθεί η φάση υποδιαίρεσης,
– Είτε με ένα νέο βήμα α_{i+1} που αποτελεί την αρχή μιας νέας επανάληψης της φάσης εντοπισμού αγκύλης.

Ας σημειωθεί ότι στην τρίτη ως άνω περίπτωση, η επαύξηση, με επιλογή $\tau_1>1$, του βήματος (π.χ. διπλασιασμός, με $\tau_1=2$, του βήματος) αποσκοπεί στην επιταχυμένη προσέγγιση του ελάχιστου γραμμής, πράγμα χρήσιμο, ιδιαίτερα στην περίπτωση που το βήμα εκκίνησης α_1 έτυχε να επιλεγεί πολύ μικρό.

Η φάση υποδιαίρεσης εκτελεί επίσης μια σειρά επαναλήψεων j όπου κάθε επανάληψη τελειώνει με ένα βήμα α_j, που είτε ικανοποιεί τις συνθήκες (2.5.1), (2.5.2), είτε οδηγεί σε μια νέα, βραχύτερη αγκύλη $[a_j,\alpha_j]$ ή $[\alpha_j,b_j]$ για την επόμενη επανάληψη. Οι επαναλήψεις j έχουν ως εξής:

Επιλογή $\alpha_j \in [a_j+\tau_2(b_j-a_j),\ b_j - \tau_3(b_j-a_j)]$ μέσω παρεμβολής (βλ. παρακάτω).
Υπολογισμός $f(\alpha_j)$.
Αν $f(\alpha_j) > f(0) + \alpha_j\rho\, f'(0)$ ή $f(\alpha_j) \geq f(a_j)$ τότε $a_{j+1} = a_j$, $b_{j+1} = \alpha_j$ δίδει νέα αγκύλη (γιατί;).
Υπολογισμός $f'(\alpha_j)$.
Αν $|f'(\alpha_j)| \leq -\sigma f'(0)$, τελείωσε η αναζήτηση επί γραμμής (γιατί;).
Αν $f'(\alpha_j) \geq 0$ τότε $a_{j+1} = a_j$, $b_{j+1} = \alpha_j$ δίδει νέα αγκύλη (γιατί;).
Αν $f'(\alpha_j) < 0$ τότε $a_{j+1} = \alpha_j$, $b_{j+1} = b_j$ δίδει νέα αγκύλη (γιατί;).

Οι τ_2, τ_3 είναι προκαθορισμένες παράμετροι, $0 < \tau_2 < \tau_3 \leq 0.5$, που εγγυώνται ότι θα υπάρξει μη μηδαμινή μείωση της αγκύλης από επανάληψη σε επανάληψη και μάλιστα

$$\left| b_{j+1} - a_{j+1} \right| \le (1-\tau_2) \left| b_j - a_j \right| \qquad (2.6.1)$$

πράγμα που εγγυάται την σύγκλιση του αλγόριθμου. Τυπικές τιμές είναι $\tau_2 = 0.05$, $\tau_3 = 0.1$. Το βήμα α_j που προσδιορίζεται τελικά από τον αλγόριθμο αναζήτησης επί γραμμής, είναι το ζητούμενο βήμα $\alpha^{(k)}$ της αντίστοιχης επανάληψης.

Η χρησιμοποιούμενη παρεμβολή στον αλγόριθμο μπορεί να είναι τετραγωνική ή κυβική. Το ερώτημα στην περίπτωση **τετραγωνικής παρεμβολής** είναι: Δεδομένης της αγκύλης [a,b] και των τιμών [f(a), f(b), f ′(a)] ή [f ′(a), f ′(b)] ή [f(a), f(b), f(c)], όπου c=0.5(a+b), να ευρεθεί το ελάχιστο $\hat{a}$ μιας τετραγωνικής συνάρτησης $\hat{f}(\alpha) = \delta\alpha^2 + \beta\alpha + \gamma$ η οποία ανταποκρίνεται στις δεδομένες τιμές.

Οι αντίστοιχοι εναλλακτικοί τύποι τετραγωνικής παρεμβολής είναι (βλ. Πρόβλημα 2.1):

$$\hat{\alpha}[f(a), f(b), f'(a)] = \frac{1}{2} \frac{2a[f(a) - f(b)] + f'(a)(b^2 - a^2)}{(b-a)f'(a) + f(a) - f(b)} \qquad (2.6.2)$$

$$\hat{\alpha}[f'(a), f'(b)] = \frac{b \cdot f'(a) - af'(b)}{f'(a) - f'(b)} \qquad (2.6.3)$$

$$\hat{\alpha}[f(a), f(b), f(c)] = c - \frac{b-a}{4} \cdot \frac{f(b) - f(a)}{f(b) - 2f(c) + f(a)} . \qquad (2.6.4)$$

Παραδειγματικά θα υπολογίσουμε τον πρώτο από τους τύπους αυτούς. Για την $\hat{f}(\alpha)$ πρέπει να ισχύει

$$\hat{f}(a) = \delta a^2 + \beta a + \gamma = f(a) \qquad (2.6.5)$$

$$\hat{f}(b) = \delta b^2 + \beta b + \gamma = f(b) \qquad (2.6.6)$$

$$\hat{f}'(a) = 2\delta a + \beta = f'(a). \qquad (2.6.7)$$

Η λύση αυτού του συστήματος εξισώσεων μας δίδει τις δ, β, γ σε συνάρτηση των γνωστών τιμών f(a), f(b), $f'(a)$. Το ελάχιστο της $\hat{f}(\alpha)$ δίδεται από $\hat{f}'(\hat{\alpha}) = 0 \Leftrightarrow \hat{\alpha} = -\beta/2\delta$, και αντικαθιστώντας τις β και δ έχουμε τον πρώτο τύπο τετραγωνικής παρεμβολής.

Ανάλογα μπορεί να υπολογισθεί τύπος κυβικής παρεμβολής ως εξής

$$\hat{a}[f(a), f(b), f'(a), f'(b)] = b - \frac{f'(b) + w - z}{f'(b) - f'(a) + 2w}(b-a) \qquad (2.6.8)$$

όπου

$$z = 3\frac{f(a) - f(b)}{b-a} + f'(a) + f'(b) \qquad (2.6.9)$$

και

$$w = \sqrt{z^2 - f'(a)f'(b)}. \qquad (2.6.10)$$

Αν οι παραπάνω τύποι παρεμβολής δώσουν ελάχιστο $\hat{\alpha}$ εκτός της αγκύλης [a,b], τότε επιλέγεται το αντίστοιχο οριακό σημείο της αγκύλης.

Το βήμα εκκίνησης α_1 του αλγόριθμου αναζήτησης επί γραμμής μπορεί και αυτό να προσδιορισθεί μέσω τετραγωνικής παρεμβολής (βλ. Πρόβλημα 2.1)

$$\alpha_1 = -2\ \Delta f/f'(0) \qquad (2.6.11)$$

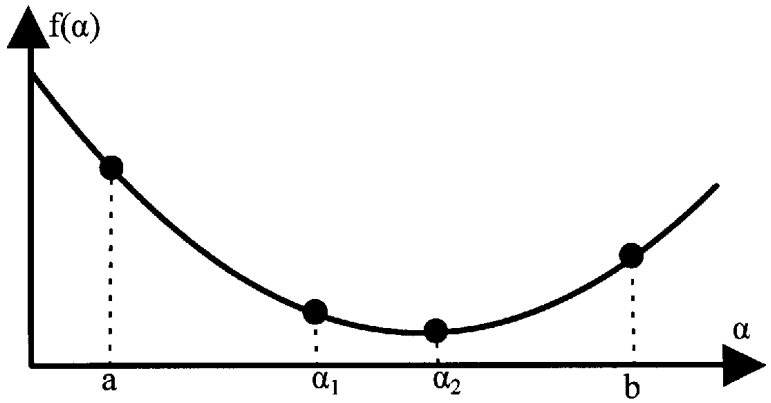
όπου $\Delta f = f^{(k-1)} - f^{(k)}$ είναι η βελτίωση της αντικειμενικής συνάρτησης που είναι όμως άγνωστη στην αρχή της επανάληψης k. Στη θέση της μπορεί να χρησιμοποιηθεί η βελτίωση της προηγούμενης επανάληψης k–1. Στην πρώτη επανάληψη, k=1, πρέπει βέβαια να χρησιμοποιηθεί μια εκτίμηση της αναμενόμενης βελτίωσης.

Σε περίπτωση που δεν είναι δυνατός ο υπολογισμός παραγώγων της f, μπορεί στη φάση υποδιαίρεσης να χρησιμοποιηθεί ο τύπος (2.6.4), ή άλλοι, συνήθως λιγότερο αποτελεσματικοί αλγόριθμοι που δεν κάνουν χρήση παραγώγων, όπως ο **αλγόριθμος χρυσής τομής.** Ο αλγόριθμος χρυσής τομής μειώνει σε μια σειρά επαναλήψεων την αρχική αγκύλη [a,b] κατά ένα σταθερό ποσοστό μέχρι τον εγκλωβισμό του εμπεριεχομένου ελάχιστου με ικανοποιητική ακρίβεια. Για την περιγραφή του αλγόριθμου χρυσής τομής θεωρούμε δύο εσωτερικά σημεία $\alpha_1, \alpha_2 \in [a,b]$ προσδιοριζόμενα ως εξής

$$\alpha_1 = a + (1-c)\,(b-a) \qquad (2.6.12)$$

$$\alpha_2 = a + c\,(b-a) \qquad (2.6.13)$$

με την σταθερά $0.5 < c < 1$. Επειδή $c > 0.5$, έχουμε $\alpha_1 < \alpha_2$, βλ. Σχ. 2.6.1. Μετά από υπολογισμό των $f(\alpha_1)$, $f(\alpha_2)$ έχουμε: Αν $f(\alpha_1) < f(\alpha_2)$, τότε το ζητούμενο ελάχιστο βρίσκεται προφανώς στο διάστημα $[a,\alpha_2]$ ενώ αν $f(\alpha_1) > f(\alpha_2)$, τότε βρίσκεται στο $[\alpha_1,b]$ (γιατί;).



Σχήμα 2.6.1 Υποδιαίρεση με χρυσή τομή.

Οποιαδήποτε από τις δύο περιπτώσεις και αν ισχύει, έχει προσδιορισθεί μια βραχύτερη αγκύλη (μήκους c(b−a)), με την οποία αρχίζει η επόμενη επανάληψη, κ.ο.κ., μέχρι να επιτευχθεί κάποιο προκαθορισμένο όριο εύρους αγκύλης ε. Η συστολή της αγκύλης σε κάθε επανάληψη δίδεται από τον παράγοντα συστολής c και ο απαιτούμενος αριθμός επαναλήψεων είναι (γιατί;)

$$L = \text{entier}\left(\frac{\ln \frac{\varepsilon}{b-a}}{\ln c}\right). \qquad (2.6.14)$$

Ο περιγραφείς αλγόριθμος απαιτεί δύο υπολογισμούς συνάρτησης ανά επανάληψη. Αν όμως $c = (\sqrt{5}-1)/2 \approx 0.618$ (βλ. Πρόβλημα 2.1b), έχουμε

$$\alpha_2 = \alpha_1 + (1-c)\,(b-\alpha_1) \qquad (2.6.15)$$

$$\alpha_1 = a + c\,(\alpha_2-a) \qquad (2.6.16)$$

δηλαδή το σημείο α_2 παίζει το ρόλο του εσωτερικού σημείου α_1 (2.6.12) για την αγκύλη $[\alpha_1,b]$ ενώ το σημείο α_1 παίζει το ρόλο του σημείου α_2 (2.6.13) για την αγκύλη $[a,\alpha_2]$, πράγμα που μειώνει τον αριθμό υπολογισμών συνάρτησης ανά επανάληψη σε έναν (χρυσή τομή!).

Προβλήματα

2.1 a) Αποδείξτε τους τύπους (2.6.2), (2.6.3), (2.6.4), (2.6.11) της τετραγωνικής παρεμβολής.
b) Υπολογίστε τον παράγοντα συστολής c της χρυσής τομής.

Λύση

a) Υπόθεση $\hat{f}(\alpha) = \delta\alpha^2 + \beta\alpha + \gamma \Rightarrow$ Min: $\hat{\alpha} = -\beta/(2\delta)$ (0)

Τύπος (2.6.2): Δεδ.: f_a, f_b, f'_a

$$\left.\begin{array}{l}\text{(i)}\ \delta a^2+\beta a+\gamma=f_a\\ \text{(ii)}\ \delta b^2+\beta b+\gamma=f_b\end{array}\right\}\text{(i)}-\text{(ii)}\Rightarrow\delta(a^2-b^2)+\beta(a-b)=f_a-f_b \quad \text{(iv)}$$

(iii) $2\delta a+\beta=f'_a \Rightarrow \beta = f'_a - 2\delta a$ (v)

(v) σε (iv) $\Rightarrow \delta = [f_b - f_a + f'_a\,(a-b)] / (a-b)^2$ (vi)

(v), (vi) σε (0) $\Rightarrow$ (2.6.2)

Τύπος (2.6.3): Δεδ.: f'_a, f'_b

$$\left.\begin{array}{l}(i)\ 2\delta a+\beta=f'_a\\(ii)\ 2\delta b+\beta=f'_b\end{array}\right\}\ \begin{array}{l}(i)-(ii)\Rightarrow\\ \delta=(f'_a-f'_b)/[2(a-b)]\quad(iii)\end{array}\ \left.\begin{array}{l}\\(iii)\ \sigma\varepsilon\ (ii)\Rightarrow\beta=(af'_b-bf'_a)/(a-b)\quad(iv)\end{array}\right\}\begin{array}{l}(iii),(iv)\ \sigma\varepsilon\ (0)\\ \Rightarrow(2.6.3)\end{array}$$

Τύπος (2.6.4): Δεδ.: f_a, f_b, f_c , $c = (a+b)/2$

$$\left.\begin{array}{l}(i)\quad\delta a^2+\beta a+\gamma=f_a\\(ii)\quad\delta b^2+\beta b+\gamma=f_b\\(iii)\ \delta c^2+\beta c+\gamma=f_c\end{array}\right\}\left.\begin{array}{l}(i)-(iii)\ \Rightarrow(iv)\\(ii)-(iii)\Rightarrow(v)\end{array}\right\}\delta=\ldots\ \text{κλπ. (όπως πάνω)}$$

Τύπος (2.6.11): Δεδ.: f'_a, $\Delta f = f_a - \hat{f}(\hat{a})$

$$(i)\ \delta a^2+\beta a+\gamma-\delta\hat{\alpha}^2-\beta\hat{\alpha}-\gamma=\Delta f$$

$$\left.\begin{array}{l}(ii)\ 2\delta a+\beta=f'_a\\(iii)\ 2\delta\hat{\alpha}+\beta=\hat{f}'(\hat{\alpha})=0\end{array}\right\}\begin{array}{l}(ii)-(iii)\Rightarrow\delta=f'_a/[2(a-\hat{\alpha})]\quad(iv)\\(iv)\ \sigma\varepsilon\ (iii)\Rightarrow\beta=-f'_a\hat{\alpha}/(a-\hat{x})\quad(v)\end{array}$$

$$(iv),(v)\ \text{σέ}\ (i)\Rightarrow\Delta f=\frac{f'_a a^2}{2(a-\hat{\alpha})}-\frac{f'_a a\hat{\alpha}}{a-\hat{\alpha}}+\frac{f'_a\hat{\alpha}^2}{2(a-\hat{\alpha})}=f'_a(a-\hat{\alpha})/2\Rightarrow(2.6.11)\quad(\text{για } a=0)$$

b) Χρυσή Τομή

$$\left.\begin{array}{l}\alpha_1=a+(1-c)(b-a)\Rightarrow(\alpha_1-a)/(b-a)=1-c\\\left.\begin{array}{l}\alpha_2=a+c(b-a)\Rightarrow(\alpha_2-a)/(b-a)=c\\\alpha_1=a+c(\alpha_2-a)\Rightarrow(\alpha_1-a)/(\alpha_2-a)=c\end{array}\right\}\Rightarrow(\alpha_1-a)/(b-a)=c^2\end{array}\right\}\Rightarrow$$

$$\Rightarrow c^2+c-1=0\Rightarrow c=(-1\pm\sqrt{5})/2\approx0.618.$$

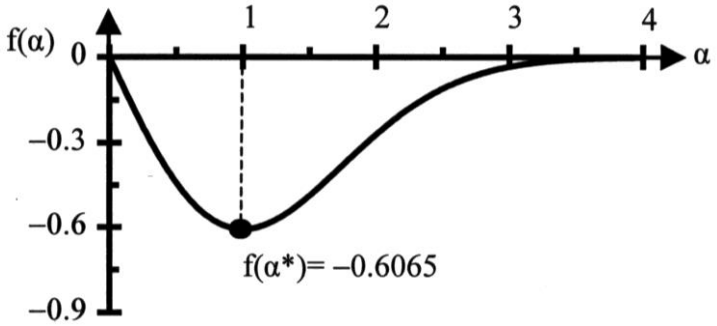
2.2 Ζητείται ελάχιστο της συνάρτησης $f(\alpha) = -\alpha\cdot\exp(-0.5\alpha^2)$, $\alpha\geq0$.

a) Προσδιορίστε το ελάχιστο α* μέσω των αναγκαίων συνθηκών και σχεδιάστε την γραφική παράσταση της συνάρτησης.

b) Επιλύστε το πρόβλημα ελαχιστοποίησης αριθμητικά στην αγκύλη [a=0, b=3] μέσω:

- του τύπου παρεμβολής (2.6.3)
- του αλγόριθμου χρυσής τομής.

Λύση a)



$f'(\alpha) = -\exp(-0.5\alpha^2) \cdot (1-\alpha^2) = 0 \Leftrightarrow \alpha^* = 1$ (αναγκαία συνθήκη)

$f''(\alpha) = \exp(-0.5\alpha^2) \cdot \alpha \cdot (3-\alpha^2) \Leftrightarrow f''(1) > 0$ (ικανή συνθήκη).

b) Αριθμητική λύση

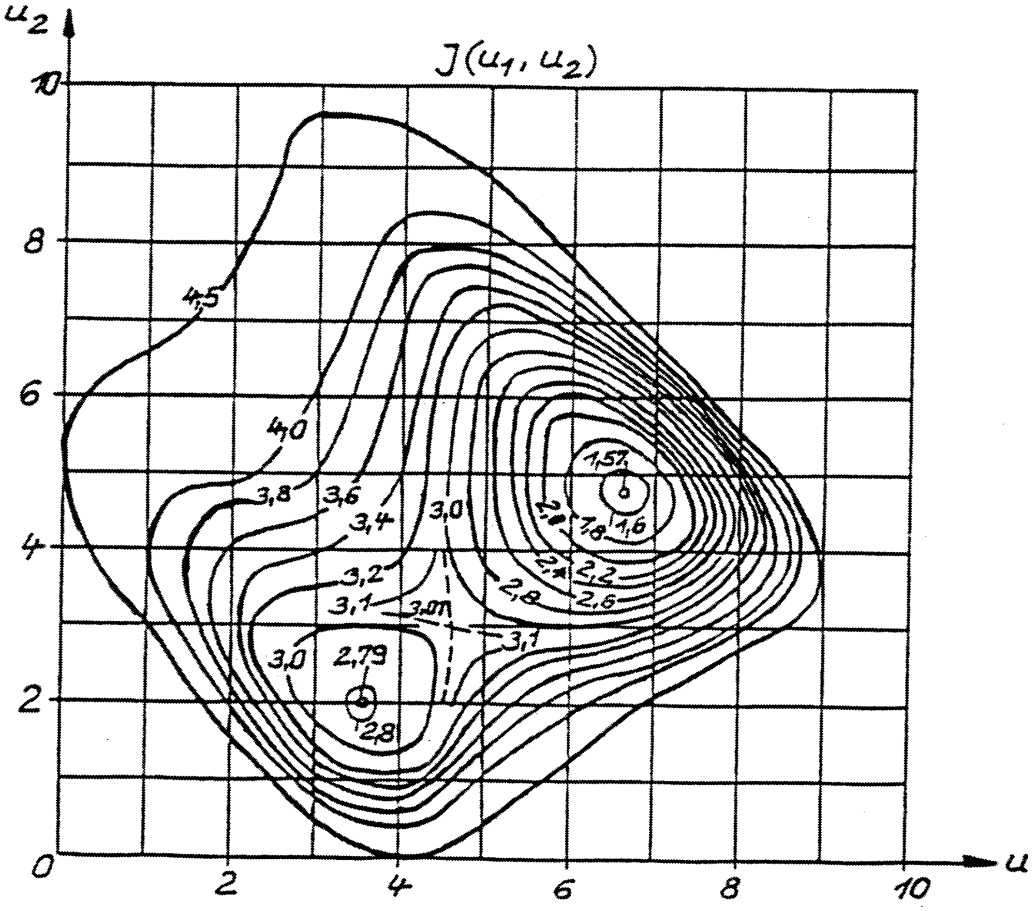
Τετραγωνική Παρεμβολή

Επαν.	a	b	f'_a	f'_b	$\hat{\alpha}$	$\hat{f}$	$\hat{f}'$
1	0	3	–1	0.089	2.75	–0.06	0.15
2	0	2.75	–1	0.15	2.39	–0.14	0.27
3	0	2.39	–1	0.27	1.88	–0.32	0.43
4	0	1.88	–1	0.43	1.31	–0.55	0.30
5	0	1.31	–1	0.30	1.01	–0.61	0.01

Χρυσή Τομή

Επαν.	a	b	α_1	α_2	f_a	f_b	f_1	f_2	b–a
1	0	3	1.146	1.854	0	–0.03	–0.59	–0.33	3
2	0	1.854	0.708	1.146	0	–0.33	–0.55	–0.59	1.854
3	0.708	1.854	1.146	1.416	–0.55	–0.33	–0.59	–0.52	1.146
4	0.708	1.416	0.978	1.146	–0.55	–0.52	–0.61	–0.59	0.708
5	0.708	1.146	0.875	0.978	–0.55	–0.59	–0.60	–0.61	0.438

2.3 Η προς ελαχιστοποίηση αντικειμενική συνάρτηση $J(u_1, u_2)$ κάποιας εφαρμογής έχει τις ισοϋψείς καμπύλες του παρακάτω σχήματος.



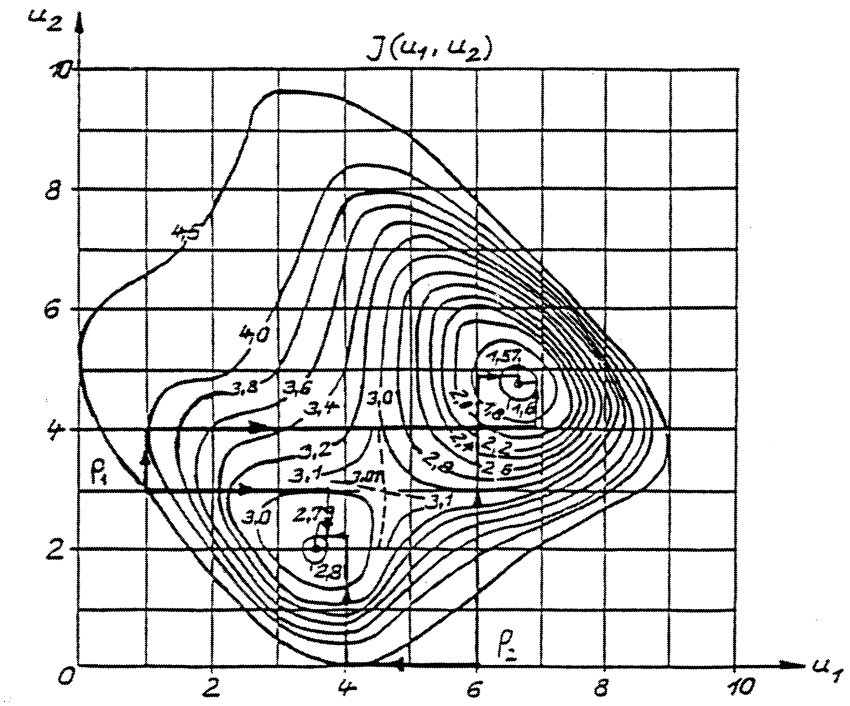
a) Προσδιορίστε γραφικά την πορεία των επαναλήψεων με σημείο εκκίνησης $P_1 = (1,3)$ ή $P_2 = (6,0)$ σε περίπτωση εφαρμογής των:

- Μεθόδου εναλλασσόμενων μεταβλητών με αναζήτηση επί γραμμής
- Μεθόδου μέγιστης κατάβασης με σταθερό βήμα $\|\Delta\mathbf{u}\| = 3$
- Μεθόδου μέγιστης κατάβασης με αναζήτηση επί γραμμής.

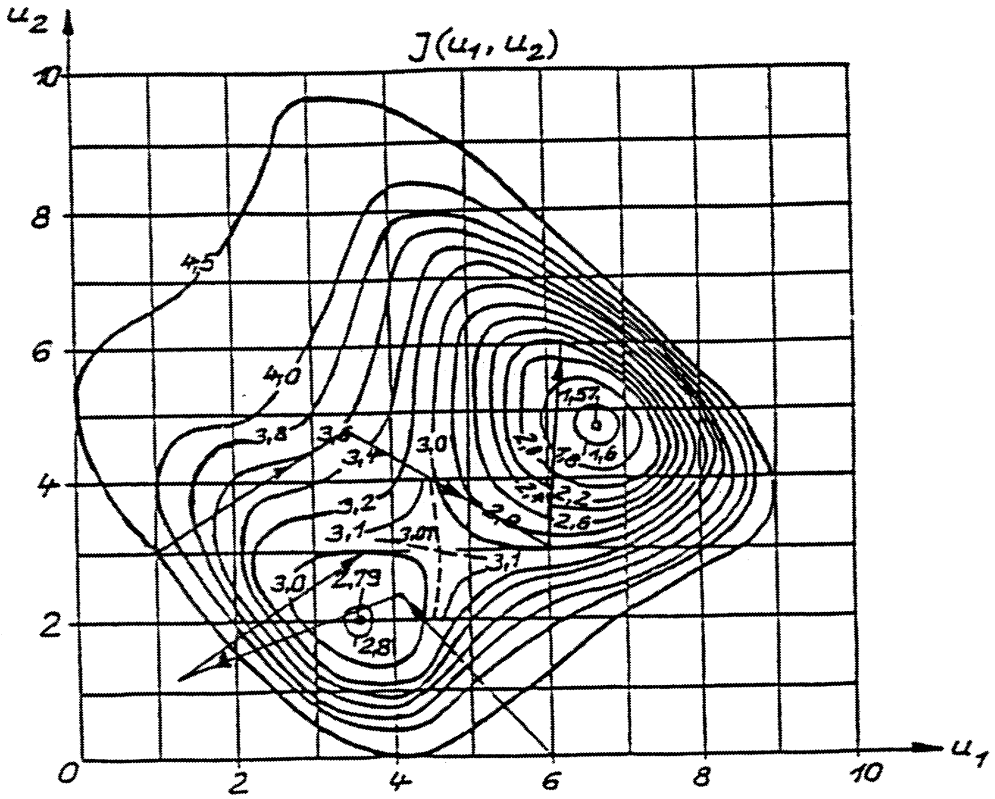
b) Υποθέτουμε τώρα την ύπαρξη περιορισμού $10u_1 + 9u_2 \leq 90$. Προσδιορίστε γραφικά την πορεία των επαναλήψεων με τη μέθοδο μέγιστης κατάβασης και αναζήτηση επί γραμμής εκκινώντας από το σημείο P_1. Τι μπορεί να γίνει ενόψει του περιορισμού;

Λύση

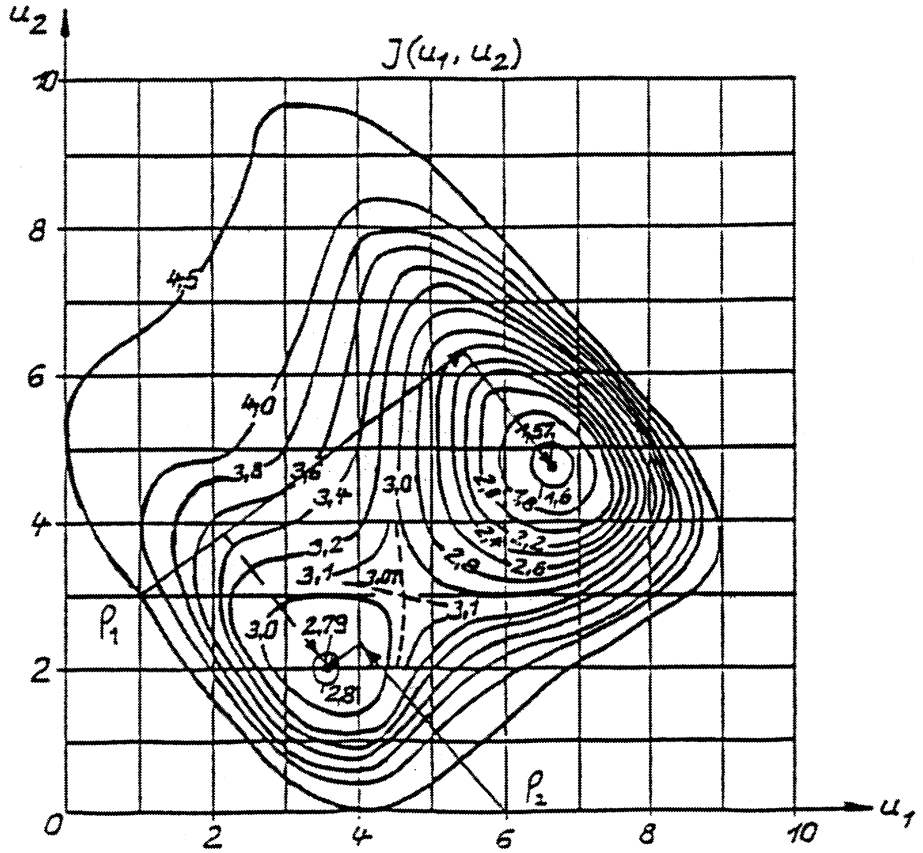
a) Εναλλασσόμενες μεταβλητές: Δύο δυνατότητες για κάθε σημείο εκκίνησης



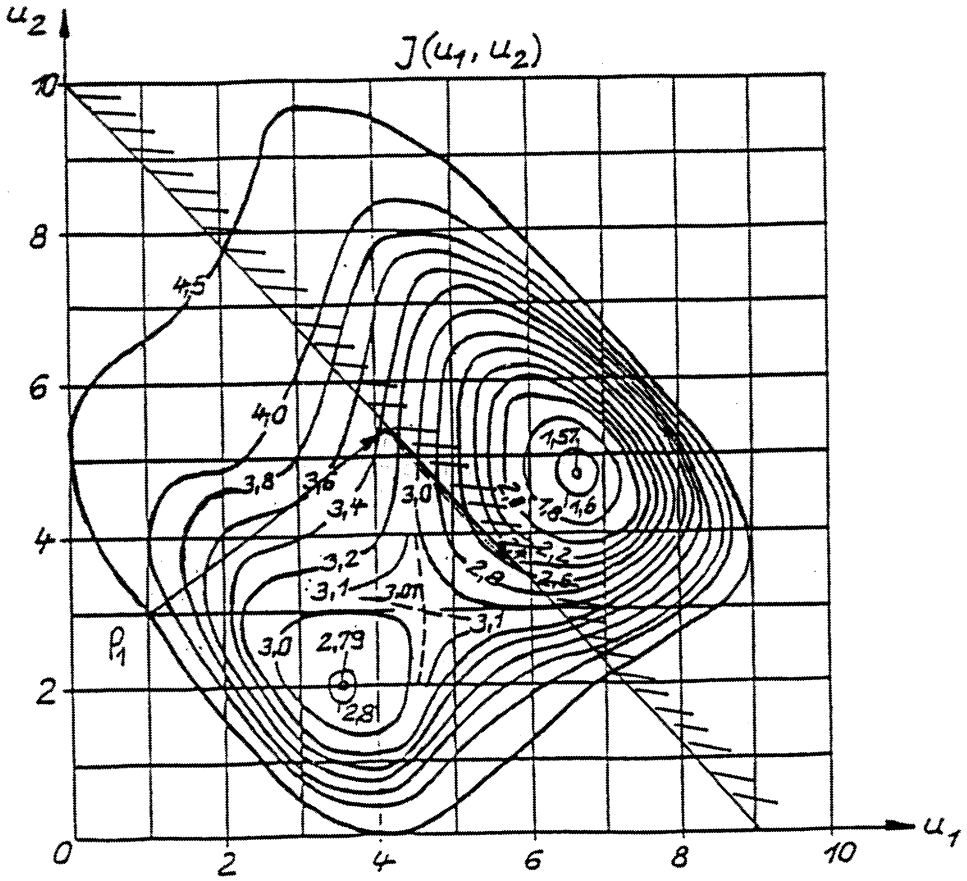
Μέγιστη κατάβαση με $\|\Delta\mathbf{u}\| = 3$: ταλαντώσεις



Μέγιστη κατάβαση με αναζήτηση επί γραμμής



b) Αναζήτηση κατά μήκος του περιορισμού!



2.4 Δίδεται προς ελαχιστοποίηση η συνάρτηση

$$f(\mathbf{x}) = x_1^2 - 4x_1 + 9x_2^2 - 18x_2 - 7.$$

Ζητείται αριθμητική ελαχιστοποίηση με σημείο εκκίνησης $x_1^{(1)} = x_2^{(1)} = 4$.

a) Υπολογίστε αναλυτικά το ζητούμενο ελάχιστο.

b) Μέγιστη κατάβαση με σταθερό βήμα α: Για ποιές τιμές α υπάρχει σύγκλιση; Υπολογίστε μερικά επαναληπτικά.

c) Μέγιστη κατάβαση με αναζήτηση επί γραμμής: Υπολογίστε μερικά επαναληπτικά.

Λύση

a) $\dfrac{\partial f}{\partial x_1} = 2(x_1-2) = 0 \Rightarrow x_1^* = 2$

$\dfrac{\partial f}{\partial x_2} = 18(x_2-1) = 0 \Rightarrow x_2^* = 1$

$\nabla^2 f = \mathbf{diag}(2,18) > \mathbf{0}$ άρα ελάχιστο!

b) $$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} - \alpha \cdot \nabla f^{(k)} = \mathbf{x}^{(k)} - \alpha \cdot \begin{bmatrix} 2x_1^{(k)} - 4 \\ 18x_2^{(k)} - 18 \end{bmatrix} = \underbrace{\begin{bmatrix} 1-2\alpha & 0 \\ 0 & 1-18\alpha \end{bmatrix}}_{\mathbf{A}} \mathbf{x}^{(k)} + \alpha \begin{bmatrix} 4 \\ 18 \end{bmatrix}$$

Ιδιοτιμές του **A**:

$$\begin{vmatrix} \Lambda - 1 + 2\alpha & 0 \\ 0 & \Lambda - 1 + 18\alpha \end{vmatrix} = 0 \Rightarrow \begin{cases} \Lambda_1 = 1 - 2\alpha \\ \Lambda_2 = 1 - 18\alpha \end{cases}.$$

Σύγκλιση αν $|\Lambda_1|, |\Lambda_2| < 1 \Leftrightarrow 0 < \alpha < 1/9$, π.χ. $\alpha = 0.09$ δίδει

k	1	2	3	4	5	...	
$x_1^{(k)}$	4	3.64	3.34	3.1	2.74	...	αργή!
$x_2^{(k)}$	4	–0.86	2.15	0.29	0.73	...	

c) Με αναζήτηση επί γραμμής.

Εδώ αναλυτικός υπολογισμός του βέλτιστου βήματος $\alpha^{(k)}$: $f(\alpha^{(k)}) = f(\mathbf{x}^{(k+1)}) = f(\mathbf{x}^{(k)} - \alpha^{(k)} \nabla f^{(k)})$.

$$f'(\alpha^{(k)}) = \mathbf{s}^{(k)T}\mathbf{g}^{(k+1)} = -\mathbf{g}^{(k)T}\mathbf{g}^{(k+1)} = -\begin{bmatrix} 2x_1^{(k)} - 4 \\ 18x_2^{(k)} - 18 \end{bmatrix}^T \begin{bmatrix} 2x_1^{(k+1)} - 4 \\ 18x_2^{(k+1)} - 18 \end{bmatrix} =$$

$$= -\begin{bmatrix} 2x_1^{(k)} - 4 \\ 18x_2^{(k)} - 18 \end{bmatrix}^T \begin{bmatrix} 2(1-2\alpha^{(k)})x_1^{(k)} + 8\alpha^{(k)} - 4 \\ 18(1-18\alpha^{(k)})x_2^{(k)} + 18^2\alpha^{(k)} - 18 \end{bmatrix} = 0$$

$$\Leftrightarrow \alpha^{(k)} = [4(x_1^{(k)} - 2)^2 + 18^2(x_2^{(k)} - 1)^2]/[8(x_1^{(k)} - 2)^2 + 18^3(x_2^{(k)} - 1)^2].$$

Επαναλήψεις με $\alpha^{(k)}$:

k	1	2	3	4	5	6
$\alpha^{(k)}$	0.0558	0.479	0.0558	0.479	0.0558	0.479
$x_1^{(k)}$	4	3.777	2.074	2.066	2.002	2.002
$x_2^{(k)}$	4	0.985	1.111	0.999	1.004	0.999

2.5 Προσδιορίστε αναλυτικά το βέλτιστο βήμα επί γραμμής $\alpha^{(k)}$ για την αριθμητική ελαχιστοποίηση μιας τετραγωνικής συνάρτησης (1.2.12).

Λύση

$$f'(\alpha^{(k)}) = \mathbf{g}^{(k+1)T}\mathbf{s}^{(k)} = [\mathbf{G}(\mathbf{x}^{(k)} + \alpha^{(k)}\mathbf{s}^{(k)}) + \mathbf{b}]^T\mathbf{s}^{(k)} = \mathbf{g}^{(k)T}\mathbf{s}^{(k)} + \mathbf{s}^{(k)T}\mathbf{G}\mathbf{s}^{(k)}\alpha^{(k)} = 0$$
$$\Rightarrow \alpha^{(k)} = -\mathbf{g}^{(k)T}\mathbf{s}^{(k)}/\mathbf{s}^{(k)T}\mathbf{G}\mathbf{s}^{(k)}.$$

2.6 Να υπολογισθούν τα τοπικά ελάχιστα, τοπικά μέγιστα και σαγματικά σημεία της συνάρτησης

$$f(x_1, x_2) = x_1^2 - 2x_2^2 + x_1x_2.$$

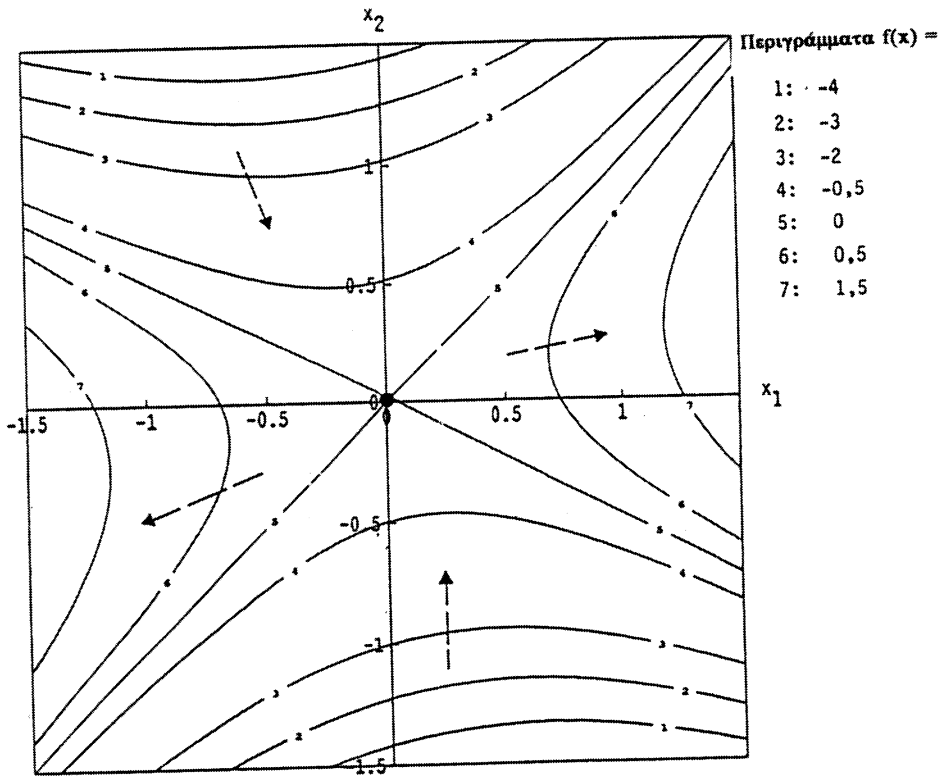
Σχεδιάστε μερικές ισοϋψείς καμπύλες.

Λύση

$$\nabla f = \begin{bmatrix} 2x_1 + x_2 \\ -4x_2 + x_1 \end{bmatrix} = \mathbf{0} \Leftrightarrow x_1^* = x_2^* = 0$$

$$\nabla^2 f = \begin{bmatrix} 2 & 1 \\ 1 & -4 \end{bmatrix} \text{ έχει ιδιοτιμές } \Lambda_{1,2} = -1 \pm \sqrt{10}$$

άρα $\nabla^2 f$ ούτε θετικά ούτε αρνητικά ημιορισμένος
άρα $\mathbf{x}^* = \mathbf{0}$ ούτε ελάχιστο ούτε μέγιστο
άρα $\mathbf{x}^* = \mathbf{0}$ σαγματικό σημείο.



2.7 Για την ελαχιστοποίηση συνάρτησης υπερσφαίρας n διαστάσεων

$$f(\mathbf{x}) = \frac{1}{2}\sum_{i=1}^{n}(x_i - a_i)^2$$

να προσδιορισθεί αναλυτική λύση και αριθμητική λύση με μέγιστη κατάβαση.

Λύση

Αναλυτική: $\nabla f = \mathbf{x}-\mathbf{a} = \mathbf{0} \Leftrightarrow \mathbf{x}^* = \mathbf{a}$

$\nabla^2 f = \mathbf{I} > \mathbf{0} \Rightarrow$ ελάχιστο.

Αριθμητική: Από τη λύση του Προβλ. 2.5: $\alpha^{(k)} = 1 \Rightarrow \mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} - (\mathbf{x}^{(k)}-\mathbf{a}) = \mathbf{a}$

Σύγκλιση σε μία και μόνη επανάληψη από οποιοδήποτε σημείο εκκίνησης!

3. ΜΕΘΟΔΟΙ ΤΥΠΟΥ NEWTON

3.1 Η Μέθοδος Newton

Η **μέθοδος Newton** εφαρμόζει ένα τετραγωνικό πρότυπο προσέγγισης της υπό ελαχιστοποίηση συνάρτησης. Έχουμε δηλαδή από την σειρά Taylor περί το επαναληπτικό $\mathbf{x}^{(k)}$

$$f(\mathbf{x}^{(k)} + \boldsymbol{\delta}) \approx q^{(k)}(\boldsymbol{\delta}) = f^{(k)} + \mathbf{g}^{(k)T}\boldsymbol{\delta} + \frac{1}{2}\boldsymbol{\delta}^T\mathbf{G}^{(k)}\boldsymbol{\delta} \qquad (3.1.1)$$

όπου $\boldsymbol{\delta}=\mathbf{x}-\mathbf{x}^{(k)}$ και $q^{(k)}(\boldsymbol{\delta})$ είναι η τετραγωνική προσέγγιση περί το $\mathbf{x}^{(k)}$ στην επανάληψη k. Το αποτέλεσμα της επανάληψης Newton είναι τότε $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \boldsymbol{\delta}^{(k)}$ όπου η διόρθωση $\boldsymbol{\delta}^{(k)}$ επιλέγεται να ελαχιστοποιεί την $q^{(k)}(\boldsymbol{\delta})$. Επειδή η ελαχιστοποίηση δίδει ένα μοναδικό ελάχιστο αν ο $\mathbf{G}^{(k)}$ είναι θετικά ορισμένος, η όλη μέθοδος λειτουργεί ομαλά αν εκπληρούται αυτή την προϋπόθεση, ειδάλλως μπορεί να παρουσιάσει προβλήματα. Αν $\mathbf{G}^{(k)}>\mathbf{0}$, έχουμε την συνθήκη ελαχιστοποίησης $\nabla q^{(k)}(\boldsymbol{\delta}^{(k)}) = \mathbf{0}$ και άρα, με την (1.2.14), έχουμε για την k-τη επανάληψη της μεθόδου Newton:

(a) Επίλυση της $\mathbf{G}^{(k)}\boldsymbol{\delta} = -\mathbf{g}^{(k)}$ ως προς $\boldsymbol{\delta}=\boldsymbol{\delta}^{(k)}$.

(b) $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \boldsymbol{\delta}^{(k)}$. (3.1.2)

Στο βήμα (a) επιλύεται ένα γραμμικό σύστημα n εξισώσεων.

Παράδειγμα 3.1.1: Θεωρούμε την ελαχιστοποίηση της

$$f(\mathbf{x}) = x_1^4 + x_1x_2 + (1+x_2)^2 \qquad (3.1.3)$$

εκκινώντας από $\mathbf{x}^{(1)} = [0.75\ -1.25]^T$. Ο Πίνακας 3.1.1 δίδει τα αποτελέσματα εφαρμογής της μεθόδου Newton, όπου το λάθος $\mathbf{h}^{(k)}$ ορίζεται στην (2.3.1). Ο πίνακας $\mathbf{G}^{(k)}$ είναι θετικά ορισμένος για κάθε k στο παράδειγμα αυτό. Παρατηρούμε την γρήγορη σύγκλιση, που είναι τυπικό χαρακτηριστικό της μεθόδου όταν το σημείο εκκίνησης $\mathbf{x}^{(1)}$ είναι κοντά στο ελάχιστο $\mathbf{x}^*$. ■

k	1	2	3	4	5
$x_1^{(k)}$	0.75	0.7	0.6959107807	0.6958843872	0.6958843861
$x_2^{(k)}$	−1.25	−1.35	−1.347955390	−1.347942194	−1.347942193
$f^{(k)}$	−0.55859375	−0.5824	−0.5824451726	−0.5824451744	−0.5824451744
$g_1^{(k)}$	0.4375	0.022	0.0001401889	0.0000000058	0
$g_2^{(k)}$	0.25	0	0	0	0
$h_1^{(k)}$	0.0541156139	0.0041156139	0.0000263946	0.0000000011	0
$h_2^{(k)}$	0.097942193	−0.002057807	−0.0000131973	−0.0000000006	0

Πίνακας 3.1.1. Μέθοδος Newton στο Παραδ. 3.1.1.

Παράδειγμα 3.1.2: Θεωρούμε την ελαχιστοποίηση της γενικής τετραγωνικής συνάρτησης (1.2.12) με θετικά ορισμένο **G**. Μια επανάληψη της μεθόδου Newton δίδει

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \boldsymbol{\delta}^{(k)} = \mathbf{x}^{(k)} - (\mathbf{G}^{(k)})^{-1}\,\mathbf{g}^{(k)} = \mathbf{x}^{(k)} - \mathbf{G}^{-1}\,(\mathbf{G}\mathbf{x}^{(k)} + \mathbf{b}) = -\mathbf{G}^{-1}\,\mathbf{b} = \mathbf{x}^* \quad (3.1.4)$$

έχουμε δηλαδή σύγκλιση σε μια και μόνη επανάληψη, ανεξάρτητα από το σημείο εκκίνησης (γιατί;).

■

Θεώρημα 3.1.1: Αν $f \in C^2$ και ο Χεσιανός πίνακας ικανοποιεί την συνθήκη Lipschitz $\|\mathbf{G}(\mathbf{x}) - \mathbf{G}(\mathbf{y})\| \le \lambda \|\mathbf{x} - \mathbf{y}\|$ σε μια περιοχή περί το τοπικό ελάχιστο **x***, αν $\mathbf{x}^{(k)}$ είναι αρκούντως πλησίον του **x*** για κάποιο k και αν ο **G*** είναι θετικά ορισμένος, τότε η μέθοδος Newton είναι καλώς ορισμένη για κάθε k και συγκλίνει σε δεύτερο βαθμό.

■

Η σύγκλιση δευτέρου βαθμού διαφαίνεται στα αποτελέσματα του αριθμητικού Παραδ. 3.1.1. Μια παραλλαγή της μεθόδου Newton χρησιμοποιεί τον ίδιο πίνακα $\mathbf{G}^{(k)}$ για μια σειρά m επαναλήψεων εξοικονομώντας έτσι τον υπολογιστικό χρόνο επίλυσης του βήματος (a), οπότε όμως ο ρυθμός σύγκλισης μειώνεται και γίνεται $(m+1)^{1/m}$ βαθμού.

Η περιγραφείσα βασική μέθοδος Newton δεν είναι εφαρμόσιμη σαν γενική μέθοδος βελτιστοποίησης καθότι ο $\mathbf{G}^{(k)}$ μπορεί να μην είναι θετικά ορισμένος αν το $\mathbf{x}^{(k)}$ ευρίσκεται μακράν του **x***. Αλλά και αν $\mathbf{G}^{(k)} > \mathbf{0}$, η μέθοδος δεν εγγυάται ότι η παραγόμενη ακολουθία $\{f^{(k)}\}$ είναι φθίνουσα. Το τελευταίο μπορεί να επιτευχθεί μέσω της **μεθόδου Newton με αναζήτηση επί γραμμής,** όπου η διόρθωση Newton χρησιμεύει απλώς σαν κατεύθυνση αναζήτησης

$$\mathbf{s}^{(k)} = -\,(\mathbf{G}^{(k)})^{-1}\,\mathbf{g}^{(k)} \qquad (3.1.5)$$

για τον αλγόριθμο (2.3.2). Πράγματι, αν $\mathbf{G}^{(k)} > \mathbf{0}$ τότε η (2.3.3) προφανώς ικανοποιείται και έχουμε την $\mathbf{s}^{(k)}$ στην (3.1.5) ως κατεύθυνση κατάβασης.

Παραμένουμε έτσι με την βασική δυσκολία της μεθόδου Newton, την περίπτωση δηλαδή όπου ο $\mathbf{G}^{(k)}$ δεν είναι θετικά ορισμένος. Μια δυνατότητα αντιμετώπισης της δυσκολίας αυτής είναι η χρήση της κατεύθυνσης μέγιστης κατάβασης $\mathbf{s}^{(k)} = -\mathbf{g}^{(k)}$ για κάθε k όπου ο $\mathbf{G}^{(k)}$ δεν είναι θετικά ορισμένος. Αν όμως αυτό είναι αναγκαίο για πολλές επαναλήψεις, καταλήγουμε στην αργή σύγκλιση της μεθόδου μέγιστης κατάβασης. Ένας άλλος τρόπος αντιμετώπισης της δυσκολίας αυτής είναι ο υπολογισμός κατεύθυνσης από ένα συνδυασμό μεθόδων Newton και μέγιστης κατάβασης

$$(\mathbf{G}^{(k)} + \nu\mathbf{I})\, \mathbf{s}^{(k)} = -\mathbf{g}^{(k)} \qquad (3.1.6)$$

με ν αρκούντως μεγάλο ώστε ο $\mathbf{G}^{(k)}+\nu\mathbf{I}$ να είναι θετικά ορισμένος. Διάφοροι άλλοι τρόποι αντιμετώπισης της δυσκολίας αυτής έχουν προταθεί και εφαρμοσθεί σε αντίστοιχους αλγόριθμους.

Συμπερασματικά, η μέθοδος Newton μπορεί να χρησιμοποιηθεί εντός του αλγορίθμου (2.3.2) για τον υπολογισμό κατεύθυνσης αναζήτησης $\mathbf{s}^{(k)}$ σε κάθε επανάληψη μέσω της (3.1.5) ή (3.1.6). Το βασικό πλεονέκτημα της μεθόδου Newton είναι η γρήγορη σύγκλιση, ιδιαίτερα κοντά στο ζητούμενο ελάχιστο $\mathbf{x}^*$. Τα βασικά μειονεκτήματα της μεθόδου Newton είναι :

- Για την εφαρμογή της μεθόδου πρέπει καταρχήν να υπολογισθεί αναλυτικά ο Χεσιανός πίνακας $\mathbf{G}(\mathbf{x})$ του εκάστοτε προβλήματος, πρέπει δηλαδή να υπολογισθούν αναλυτικά n(n+1)/2 δεύτερες παράγωγοι (δεδομένης της συμμετρίας του πίνακα), πράγμα ενδεχομένως επίπονο για προβλήματα υψηλών διαστάσεων.
- Κατά την εφαρμογή της μεθόδου έχουμε αντιστροφή του Χεσιανού πίνακα για την παραγωγή κατεύθυνσης αναζήτησης σε κάθε επανάληψη, πράγμα που απαιτεί $O(n^3)$ πολλαπλασιασμούς και άρα υπολογιστικό φόρτο που αυξάνει κυβικά με τις διαστάσεις του προβλήματος.

3.2 Μέθοδοι Σχεδόν-Newton

Τα ως άνω μειονεκτήματα της μεθόδου Newton παρακάμπτονται από τις **μεθόδους σχεδόν-Newton** που προσεγγίζουν τον $(\mathbf{G}^{(k)})^{-1}$ με έναν συμμετρικό, θετικά ορισμένο πίνακα $\mathbf{H}^{(k)}$ που μεταβάλεται από επανάληψη σε επανάληψη. Έχουμε έτσι σε κάθε επανάληψη του αλγορίθμου (2.3.2):

(a) $\mathbf{s}^{(k)} = -\mathbf{H}^{(k)}\mathbf{g}^{(k)}$
(b) Αναζήτηση κατά μήκος της κατεύθυνσης $\mathbf{s}^{(k)}$ δίδει $\alpha^{(k)}$.
(c_1) $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \alpha^{(k)}\, \mathbf{s}^{(k)}$. (3.2.1)
(c_2) Μετατροπή του $\mathbf{H}^{(k)}$ σε $\mathbf{H}^{(k+1)}$.

Ο πίνακας εκκίνησης $\mathbf{H}^{(1)}$ μπορεί να είναι οποιοσδήποτε θετικά ορισμένος πίνακας, συνήθως $\mathbf{H}^{(1)} = \mathbf{I}$. Τα πλεονεκτήματα των μεθόδων σχεδόν-Newton έναντι της μεθόδου Newton είναι:

- Απαιτούνται αναλυτικοί υπολογισμοί μόνο πρώτων παραγώγων (διάνυσμα κλίσης).
- Ο $\mathbf{H}^{(k)}$ μπορεί να καθοριστεί έτσι ώστε να είναι πάντα θετικά ορισμένος.
- Απαιτούνται $O(n^2)$ πολλαπλασιασμοί ανά επανάληψη έναντι $O(n^3)$ της μεθόδου Newton.

Αντιθέτως, οι ανάγκες αποθήκευσης είναι παρόμοιες για τις δύο μεθόδους.

Στόχος του **τύπου μετατροπής** του $\mathbf{H}^{(k)}$ σε $\mathbf{H}^{(k+1)}$ είναι να προσεγγισθεί ο αντίστροφος Χεσιανός πίνακας. Ένας τρόπος για να επιτευχθεί αυτό είναι ο εξής. Ορίζοντας τις διαφορές

$$\boldsymbol{\delta}^{(k)} = \alpha^{(k)}\mathbf{s}^{(k)} = \mathbf{x}^{(k+1)} - \mathbf{x}^{(k)} \tag{3.2.2}$$

$$\boldsymbol{\gamma}^{(k)} = \mathbf{g}^{(k+1)} - \mathbf{g}^{(k)} \tag{3.2.3}$$

έχουμε από την σειρά Taylor (1.2.19)

$$\boldsymbol{\gamma}^{(k)} = \mathbf{G}^{(k)}\boldsymbol{\delta}^{(k)} + O(\|\boldsymbol{\delta}^{(k)}\|). \tag{3.2.4}$$

Παραλείποντας τους όρους υψηλών βαθμών, επιθυμούμε την επιλογή του $\mathbf{H}^{(k+1)}$ έτσι ώστε

$$\mathbf{H}^{(k+1)}\boldsymbol{\gamma}^{(k)} = \boldsymbol{\delta}^{(k)} \tag{3.2.5}$$

αφού στην περίπτωση αυτή, όπως γίνεται φανερό με αντικατάσταση της (3.2.4) στην (3.2.5), θα έχουμε $\mathbf{H}^{(k+1)} \approx (\mathbf{G}^{(k)})^{-1}$. Η εξίσωση (3.2.5) φέρει το όνομα **συνθήκη σχεδόν-Newton** και υπάρχουν διάφοροι τρόποι ικανοποίησής της. Ένας απλός τρόπος είναι να θεωρήσουμε

$$\mathbf{H}^{(k+1)} = \mathbf{H}^{(k)} + \mathbf{E}^{(k)} = \mathbf{H}^{(k)} + a\mathbf{u}\mathbf{u}^T \tag{3.2.6}$$

και με την (3.2.5)

$$\mathbf{H}^{(k)}\boldsymbol{\gamma}^{(k)} + a\mathbf{u}\mathbf{u}^T\boldsymbol{\gamma}^{(k)} = \boldsymbol{\delta}^{(k)}.$$

Η ισότητα αυτή ικανοποιείται με $\mathbf{u} = \boldsymbol{\delta}^{(k)} - \mathbf{H}^{(k)}\boldsymbol{\gamma}^{(k)}$ και $a=1/(\mathbf{u}^T\boldsymbol{\gamma}^{(k)})$ οπότε λαμβάνουμε τον **τύπο πρώτης σειράς**

$$\mathbf{H}^{(k+1)} = \mathbf{H} + \frac{(\boldsymbol{\delta} - \mathbf{H}\boldsymbol{\gamma})(\boldsymbol{\delta} - \mathbf{H}\boldsymbol{\gamma})^T}{(\boldsymbol{\delta} - \mathbf{H}\boldsymbol{\gamma})^T\boldsymbol{\gamma}} \tag{3.2.7}$$

όπου παραλείπουμε τον πάνω δείκτη (k) στην δεξιά πλευρά της εξίσωσης, όπως συνηθίζεται. Ο τύπος αυτός που προτάθηκε στο χρονικό διάστημα 1967–1969 ανεξάρτητα από 5 ερευνητές ή ερευνητικές ομάδες, παρουσιάζει μια σειρά πρακτικών μειονεκτημάτων με κυριότερο το ότι δεν εγγυάται ότι οι προκύπτοντες $\mathbf{H}^{(k)}$ είναι θετικά ορισμένοι αν $\mathbf{H}^{(1)} > \mathbf{0}$ (την λεγόμενη **κληρονομική ιδιότητα**).

Ο **τύπος δεύτερης σειράς**, αντιθέτως, είναι περισσότερο ευέλικτος και εξάγεται αν θεωρήσουμε αντί της (3.2.6)

$$\mathbf{H}^{(k+1)} = \mathbf{H}^{(k)} + a\mathbf{u}\mathbf{u}^T + b\mathbf{v}\mathbf{v}^T$$

οπότε με την συνθήκη σχεδόν-Newton (3.2.5) έχουμε

$$\boldsymbol{\delta}^{(k)} = \mathbf{H}^{(k)}\boldsymbol{\gamma}^{(k)} + a\mathbf{u}\mathbf{u}^T\boldsymbol{\gamma}^{(k)} + b\mathbf{v}\mathbf{v}^T\boldsymbol{\gamma}^{(k)}. \tag{3.2.8}$$

Η ισότητα όμως αυτή επιδέχεται πολλές λύσεις $\mathbf{u}$, $\mathbf{v}$, όπως $\mathbf{u}=\boldsymbol{\delta}^{(k)}$ και $\mathbf{v}=\mathbf{H}^{(k)}\boldsymbol{\gamma}^{(k)}$ με $a=1/(\mathbf{u}^T\boldsymbol{\gamma}^{(k)})$ και $b= -1/(\mathbf{v}^T\boldsymbol{\gamma}^{(k)})$ εξών ο τύπος

$$\mathbf{H}_{\mathrm{DFP}}^{(k+1)} = \mathbf{H} + \frac{\boldsymbol{\delta}\boldsymbol{\delta}^T}{\boldsymbol{\delta}^T\boldsymbol{\gamma}} - \frac{\mathbf{H}\boldsymbol{\gamma}\boldsymbol{\gamma}^T\mathbf{H}}{\boldsymbol{\gamma}^T\mathbf{H}\boldsymbol{\gamma}} \tag{3.2.9}$$

που προτάθηκε παραδόξως το 1963, δηλαδή πριν από τον τύπο πρώτης σειράς, από τους Davidon, Fletcher και Powell και είναι γνωστός ως **τύπος DFP**. Η μέθοδος δουλεύει πολύ καλά σε πρακτικά προβλήματα και τυγχάνει ευρείας εφαρμογής. Είναι ασύγκριτα αποτελεσματικότερη από την μέδοδο μέγιστης κατάβασης, αλλά και λίγο αποτελεσματικότερη από τις μεθόδους συζυγών κλίσεων του Κεφαλαίου 4. Στην αρχή η μέθοδος DFP εφαρμόζονταν με σχετικά ακριβή αναζήτηση επί γραμμής. Αργότερα φάνηκε γενικά πλεονεκτικότερο να εφαρμόζεται σχετικά ανακριβής αναζήτηση επί γραμμής, οπότε άλλες μέθοδοι εμφάνισαν πρακτικά πλεονεκτήματα έναντι της DFP. Οι ιδιότητες της DFP συνοψίζονται ως εξής:

- Τερματισμός το πολύ σε n επαναλήψεις με $\mathbf{H}^{(n+1)} = \mathbf{G}^{-1}$ για τετραγωνικές συναρτήσεις με ακριβή αναζήτηση επί γραμμής.
- Παράγει, για τετραγωνικές συναρτήσεις, συζυγείς κατευθύνσεις, και μάλιστα, αν $\mathbf{H}^{(1)} = \mathbf{I}$, συζυγείς κλίσεις.
- Διατηρεί τους πίνακες $\mathbf{H}^{(k)} > \mathbf{0}$ $\forall k > 1$ αν $\mathbf{H}^{(1)} > \mathbf{0}$ (κληρονομική ιδιότητα) και άρα παράγει κατευθύνσεις κατάβασης (υπό την προϋπόθεση ακριβούς αναζήτησης επί γραμμής).
- Χρειάζεται $3n^2 + O(n)$ πολλαπλασιασμούς ανά επανάληψη.
- Υπεργραμμικός βαθμός σύγκλισης.

Ένας άλλος σημαντικός τύπος προτάθηκε το 1970 ανεξάρτητα από τους Broyden, Fletcher, Goldfard και Shanno και είναι γνωστός ως **τύπος BFGS**

$$\mathbf{H}_{\mathrm{BFGS}}^{(k+1)} = \mathbf{H} + \left(1 + \frac{\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{H}\boldsymbol{\gamma}}{\boldsymbol{\delta}^{\mathrm{T}}\boldsymbol{\gamma}}\right)\frac{\boldsymbol{\delta}\boldsymbol{\delta}^{\mathrm{T}}}{\boldsymbol{\delta}^{\mathrm{T}}\boldsymbol{\gamma}} - \frac{\boldsymbol{\delta}\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{H} + \mathbf{H}\boldsymbol{\gamma}\boldsymbol{\delta}^{\mathrm{T}}}{\boldsymbol{\delta}^{\mathrm{T}}\boldsymbol{\gamma}}. \qquad (3.2.10)$$

Ο τύπος BFGS δουλεύει πολύ καλά σε πρακτικές εφαρμογές, συχνά καλύτερα από τον DFP, και χαρακτηρίζεται από τις ίδιες ως άνω ιδιότητες του DFP. Οι δυνατότητες εξαγωγής άλλων τύπων σχεδόν-Newton είναι άπειρες, δεν έχουν όμως προσδιορισθεί μέχρι τώρα άλλοι τύποι που να είναι αποτελεσματικότεροι των DFP και BFGS σε πρακτικές εφαρμογές.

Μια ολόκληρη οικογένεια τύπων, η **οικογένεια Broyden,** δημιουργείται μέσω μιας παραμέτρου φ ως εξής

$$\mathbf{H}_{\varphi}^{(k+1)} = (1-\varphi)\mathbf{H}_{\mathrm{DFP}}^{(k+1)} + \varphi\mathbf{H}_{\mathrm{BFGS}}^{(k+1)} \qquad (3.2.11)$$

όπου για φ=0 έχουμε τον DFP, για φ=1 έχουμε τον BFGS και για άλλες τιμές του φ έχουμε τον τύπο πρώτης σειράς ή άλλους τύπους που έχουν προταθεί κατά καιρούς. Η σημασία της οικογένειας Broyden έγκειται στο γεγονός ότι μια σειρά ιδιοτήτων μπορούν να αποδειχθούν γενικά, για όλα τα μέλη της, π.χ. ο τετραγωνικός τερματισμός και η παραγωγή συζυγών κατευθύνσεων.

3.3 Αλλαγή Κλίμακας

Θεωρούμε τον γραμμικό μετασχηματισμό των μεταβλητών βελτιστοποίησης **x** μέσω ενός κανονικού, τετραγωνικού πίνακα **A**

$$\mathbf{y} = \mathbf{A}\mathbf{x} \qquad (3.3.1)$$

εξού και $\mathbf{x} = \mathbf{A}^{-1}\mathbf{y}$. Η f μπορεί επομένως να υπολογισθεί βάσει του **x** (σημειώνουμε $f_x(\mathbf{x})$) ή βάσει του **y** (σημειώνουμε $f_y(\mathbf{y}) = f_x(\mathbf{A}^{-1}\mathbf{y})$). Παραγωγίζοντας έχουμε,

$$\mathbf{g}_x = \nabla_x f_x = \nabla_x f_y \overset{(\Pi.1.8)}{=} \nabla_x \mathbf{y} \cdot \nabla_y f_y \overset{(\Pi.1.3)}{=} \mathbf{A}^T \mathbf{g}_y \qquad (3.3.2)$$

και ακόμη

$$\mathbf{G}_x = \nabla_x \mathbf{g}_x = \nabla_x(\mathbf{A}^T\mathbf{g}_y) \overset{(\Pi1.9)}{=} \nabla_x \mathbf{y} \cdot \nabla_y (\mathbf{A}^T\mathbf{g}_y) \overset{(\Pi1.3)}{=} \mathbf{A}^T\mathbf{G}_y\mathbf{A}. \qquad (3.3.3)$$

Προφανώς οι (3.3.2) και (3.3.3) προσδιορίζουν τις σχέσεις των παραγώγων στα συστήματα συντεταγμένων **x** και **y** αντίστοιχα.

Η σημασία του μετασχηματισμού (3.3.1) έγκειται στο γεγονός ότι πολλές μέθοδοι αριθμητικής βελτιστοποίησης λειτουργούν αποτελεσματικότερα αν ο Χεσιανός πίνακας της υπό ελαχιστοποίηση συνάρτησης προσεγγίζει τον **I**, δηλαδή αν έχουμε μια συνάρτηση που προσεγγίζει μια υπερσφαίρα (βλ. Πρόβλημα 2.7). Λέμε τότε ότι το πρόβλημα είναι **καλής κατάστασης**. Αντίθετα, προβλήματα όπως αυτό του Σχ. 2.2.3 με επιμήκη περιγράμματα, έχουν **νοσηρή κατάσταση** και δυσκολεύουν τη λειτουργία των αλγόριθμων βελτιστοποίησης. Η κατάσταση ενός προβλήματος μπορεί να εκτιμηθεί μεσω των ιδιοτιμών του **G**: όσο πιο συγκεντρωμένες είναι οι ιδιοτιμές, τόσο καλύτερη η κατάσταση του προβλήματος. Θεωρώντας λοιπόν την (3.3.3), γίνεται φανερό πως με κατάλληλη επιλογή του πίνακα **A**, με κατάλληλο δηλαδή γραμμικό μετασχηματισμό των μεταβλητών του προβλήματος, μπορεί ένα πρόβλημα νοσηρής κατάστασης ως προς **x** να μετασχηματισθεί σε πρόβλημα καλής κατάστασης ως προς **y**. Αρκεί να επιλεγεί ο **A** έτσι ώστε $\mathbf{G}_y = \mathbf{A}^{-T}\,\mathbf{G}_x\,\mathbf{A}^{-1} \approx \mathbf{I}$. Βέβαια ο προσδιορισμός του πίνακα **A** δυσκολεύει από το γεγονός ότι ο Χεσιανός πίνακας γενικών μη γραμμικών συναρτήσεων δεν είναι σταθερός. Εν τούτοις είναι συχνά δυνατή μέσω γραμμικού μετασχηματισμού η κατά μέσο όρο βελτίωση της κατάστασης του προβλήματος.

Παράδειγμα 3.3.1: Θεωρούμε τη συνάρτηση

$$f_x(\mathbf{x}) = \frac{1}{2}\left(\frac{x_1^2}{a^2} + \frac{x_2^2}{b^2}\right), \quad a \neq b$$

με ελλειπτικά περιγράμματα. Έχουμε σταθερό $\mathbf{G}_x = \mathbf{diag}(a^{-2}, b^{-2})$ και με $\mathbf{A} = \mathbf{diag}(a^{-1}, b^{-1})$ λαμβάνουμε $\mathbf{G}_y = \mathbf{I}$. Αντικαθιστώντας $\mathbf{x} = \mathbf{A}^{-1}\mathbf{y} = [ay_1\ by_2]^T$ στην $f_x(\mathbf{x})$ έχουμε

$$f_y(\mathbf{y}) = \frac{1}{2}(y_1^2 + y_2^2)$$

της οποίας τα περιγράμματα είναι πράγματι κύκλοι.

■

Ο γραμμικός μετασχηματισμός (3.3.1) με στόχο την βελτίωση της κατάστασης ενός προβλήματος ονομάζεται **αλλαγή κλίμακας** ή **κλιμακοποίηση**. Ένας εύκολος και συνήθης τρόπος επιλογής του **πίνακα κλιμακοποίησης A** είναι

$$\mathbf{A} = \mathbf{diag}(\sqrt{\partial^2 f / (\partial x_i)^2}). \qquad (3.3.4)$$

Σημειώνουμε ότι η αλλαγή κλίμακας ενός προβλήματος είναι δυνατή χωρίς να αλλάξουν ρητώς οι χρησιμοποιούμενες μεταβλητές, παρά μόνον οι τύποι προσδιορισμού κατεύθυνσης αναζήτησης. Αν η κατεύθυνση αναζήτησης στο σύστημα συντεταγμένων **y** είναι $\mathbf{s}_y$, έχουμε $\mathbf{s}_y = \mathbf{A}\mathbf{s}_x$ και άρα λαμβάνουμε για την μέθοδο μέγιστης κατάβασης

$$\mathbf{s}_y = -\mathbf{g}_y = -\mathbf{A}^{-T}\mathbf{g}_x = \mathbf{A}\mathbf{s}_x.$$

Θεωρώντας πως ο **A** επιλέγεται συμμετρικός, λαμβάνουμε από την εξίσωση αυτή την **κλιμακοποιημένη κατεύθυνση μέγιστης κατάβασης**

$$\mathbf{s}_x = -\mathbf{A}^{-2}\mathbf{g}_x. \qquad (3.3.5)$$

Για τις κλιμακοποιημένες μεθόδους DFP και BFGS είναι εύκολο να αποδειχθεί ότι όλοι οι τύποι παραμένουν αμετάβλητοι και απαιτείται απλώς στην πρώτη επανάληψη η χρήση $\mathbf{H}^{(1)} = \mathbf{A}^{-2}$ αντί $\mathbf{H}^{(1)} = \mathbf{I}$. Στη μέθοδο Newton θα είχαμε στο σύστημα συντεταγμένων **y**

$$\mathbf{s}_y = -\mathbf{G}_y^{-1}\,\mathbf{g}_y = -\mathbf{A}\mathbf{G}_x^{-1}\,\mathbf{A}^T\mathbf{A}^{-T}\mathbf{g}_x = -\mathbf{A}\mathbf{G}_x^{-1}\mathbf{g}_x$$

και επειδή $\mathbf{s}_y = \mathbf{A}\mathbf{s}_x$, λαμβάνουμε $\mathbf{s}_x = -\mathbf{G}_x^{-1}\mathbf{g}_x$, άρα η μέθοδος Newton παραμένει αμετάβλητη από γραμμικούς μετασχηματισμούς (3.3.1).

3.4 Αριθμητικά Πειράματα

Όλα τα θεωρητικά αποδεδειγμένα πλεονεκτήματα (υπεργραμμική ή τετραγωνική σύγκλιση, τετραγωνικός τερματισμός κλπ.) των μεθόδων που στηρίζονται σε τετραγωνικά πρότυπα (έναντι π.χ. της μεθόδου μέγιστης κατάβασης) αφορούν την εφαρμογή των μεθόδων αυτών σε τετραγωνικές συναρτήσεις ή τετραγωνικές προσεγγίσεις της υπό ελαχιστοποίηση συνάρτησης. Για την πρακτικά ενδιαφέρουσα περίπτωση εφαρμογής των μεθόδων σε γενικές μη γραμμικές συναρτήσεις δεν υπάρχουν θεωρήματα που να αναφέρονται στην ταχύτητα σύγκλισης ή τον αριθμό των απαιτούμενων επαναλήψεων. Η πρακτική αποτελεσματικότητα των μεθόδων σε προβλήματα ελαχιστοποίησης μπορεί επομένως να διερευνηθεί κυρίως μέσω πρακτικών εφαρμογών, και έτσι μια σημαντική φάση κατά τον σχεδιασμό αλγορίθμων ελαχιστοποίησης είναι ο πειραματισμός εφαρμογής τους σε επιλεγμένα προβλήματα.

Στο παρόν κεφάλαιο θα περιγράψουμε μερικά τέτοια πειράματα που δείχνουν παραστατικά τα τυπικά πλεονεκτήματα και μειονεκτήματα διάφορων επιλογών, π.χ. χρήση διάφορων τύπων,

περισσότερο ή λιγότερο ακριβής αναζήτηση επί γραμμής κλπ. Τα πειράματα αυτά δείχνουν όμως και μερικές δυσκολίες αξιολόγησης αριθμητικών αποτελεσμάτων. Θα χρησιμοποιηθούν δύο συναρτήσεις, που είναι αρκετά γνωστές και αρκετά αντιπροσωπευτικές για πραγματικές εφαρμογές. Η συνάρτηση Rosenbrock (1.2.2) έχει δύο μόνο μεταβλητές, αλλά είναι δύσκολο να ελαχιστοποιηθεί εξ αιτίας της επιμήκους και κυρτής κοιλάδας που εμπεριέχει. Η συνάρτηση Rosenbrock μπορεί όμως και να γενικευθεί για τυχαίο αριθμό n μεταβλητών ως εξής

$$f(\mathbf{x}) = \sum_{i=1}^{n-1}[(1-x_i)^2 + 100(x_{i+1} - x_i^2)^2]. \qquad (3.4.1)$$

Η δεύτερη συνάρτηση είναι η συνάρτηση Chebyquad η οποία μπορεί επίσης να αναπτυχθεί για οιονδήποτε αριθμό n μεταβλητών.

$$f(\mathbf{x}) = \left(\sum_{i=1}^{n} x_i - 1\right)^2 + \left[\sum_{i=1}^{n}(2x_i - 1)^2 - \frac{2}{3}\right]^2. \qquad (3.4.2)$$

Η γενίκευση των δύο συναρτήσεων για n μεταβλητές θα χρησιμοποιηθεί προκειμένου να διερευνηθεί η αποτελεσματικότητα των μεθόδων σε προβλήματα διαφόρων διαστάσεων. Για τον έλεγχο σύγκλισης των μεθόδων χρησιμοποιήθηκε το κριτήριο (2.3.5) με $\varepsilon=10^{-4}$.

Ο Πίνακας 3.5.1 συνοψίζει τα αποτελέσματα διερεύνησης των μεθόδων μέγιστης κατάβασης, DFP και BFGS με χρήση κυβικής παρεμβολής (2.6.8)-(2.6.10) κατά την διαδικασία αναζήτησης επί γραμμής. Για κάθε διερευνηθείσα μέθοδο και πρόβλημα, ο πίνακας παρουσιάζει τον αριθμό επαναλήψεων (E) και τον υπολογιστικό χρόνο (CPU-seconds) που απαιτήθηκαν μέχρι την σύγκλιση. Ο ίδιος πίνακας περιλαμβάνει αποτελέσματα για σχετικά ακριβή αναζήτηση επί γραμμής (σ=0.1 στην (2.5.2)) καθώς και για ανακριβή αναζήτηση επί γραμμής (σ=0.9). Σημειώνεται ότι στην δεύτερη περίπτωση (σ=0.9) το κριτήριο τερματισμού της αναζήτησης επί γραμμής ικανοποιείται σχεδόν πάντοντε εφόσον επιτυγχάνεται μείωση της συνάρτησης f. Βάσει των αποτελεσμάτων αυτών παρατηρούμε τα εξής.

Για σ=0.1 (ακριβής αναζήτηση):

- Οι δύο μέθοδοι σχεδόν-Newton παρουσιάζουν σχεδόν ταυτόσημους αριθμούς επαναλήψεων, και αυτό θα ίσχυε και για άλλες μεθόδους σχεδόν-Newton. Σε επίπεδο χρόνου όμως, η DFP παρουσιάζει ένα ελαφρύ πλεονέκτημα έναντι της BFGS (κυρίως για μεγαλύτερα προβλήματα) αφού ο τύπος DFP (3.2.9) είναι λιγότερο δαπανηρός υπολογιστικά από τον αντίστοιχο τύπο BFGS (3.2.10).
- Η μέθοδος μέγιστης κατάβασης παρουσιάζει σοβαρό πρόβλημα σύγκλισης για την συνάρτηση Rosenbrock όπου ο αριθμός επαναλήψεων παρουσιάζεται υπερβολικά υψηλός. Αυτό οφείλεται στην συγκεκριμένη μορφή της συνάρτησης αυτής. Αντιθέτως τα αποτελέσματα για την συνάρτηση Chebyquad είναι πολύ ικανοποιητικά.

Πρόβλημα	σ = 0.1						σ = 0.9					
	Μέγιστη κατάβαση		DFP		BFGS		Μέγιστη κατάβαση		DFP		BFGS	
n	E	CPU-s	E	CPU-s	E	CPU-s	E	CPU-s	E	CPU-s	E	CPU-s
Rosenbrock												
2	3239	0.27	16	$<10^{-3}$	14	$<10^{-3}$	7413	0.61	19	$<10^{-3}$	18	$<10^{-3}$
4	8727	1.21	28	$<10^{-3}$	28	0.06	10275	1.37	46	$<10^{-3}$	35	$<10^{-3}$
6	7742	1.43	34	0.06	35	0.06	11456	2.03	83	0.05	41	$<10^{-3}$
8	10419	2.42	50	0.06	49	0.05	12432	2.8	131	0.11	56	0.05
10	7161	2.03	69	0.11	62	0.11	12769	3.52	106	0.11	68	0.11
20	11252	5.99	49	0.16	47	0.22	14629	7.52	222	0.83	55	0.27
30	15085	11.59	71	0.55	71	0.82	17896	13.51	116	0.88	81	0.88
40	19458	19.55	102	1.37	97	1.92	20893	20.77	306	4.01	108	2.09
60	24152	35.92	205	5.98	188	8.02	27930	41.14	1348	38.34	290	12.08
80	31953	62.83	316	16.04	271	20.1	37126	72.61	1944	96.67	314	22.96
Chebyquad												
2	20	$<10^{-3}$	8	$<10^{-3}$	8	$<10^{-3}$						
4	28	0.06	14	$<10^{-3}$	14	$<10^{-3}$						
6	26	0.06	16	$<10^{-3}$	17	$<10^{-3}$						
8	100	0.05	21	0.05	21	$<10^{-3}$						
10	52	$<10^{-3}$	28	0.05	28	0.06						
20	142	0.22	51	0.28	45	0.28						
30	122	0.27	91	0.82	87	1.1						
40	121	0.39	119	1.82	122	2.52						
60	106	0.49	173	5.5	166	7.47						
80	99	0.66	220	11.86	222	17.14						

Πίνακας 3.5.1: Αποτελέσματα μέγιστης κατάβασης, DFP και BFGS για κυβική παρεμβολή.

- Για όλες τις μεθόδους ο αριθμός των επαναλήψεων και ο υπολογιστικός χρόνος αυξάνουν πολυωνυμικά με την διάσταση n.
- Οι μέθοδοι σχεδόν-Newton είναι γενικά πολύ αποτελεσματικότερες από την μέθοδο μέγιστης κατάβασης. Αυτό επιβεβαιώνεται από τα αποτελέσματα στην περίπτωση της συνάρτησης Rosenbrock. Στην συνάρτηση Chebyquad, οι μέθοδοι σχεδόν-Newton παρουσιάζουν ως επί το πλείστον μικρότερο αριθμό επαναλήψεων από την μέθοδο μέγιστης κατάβασης, όμως το υψηλότερο υπολογιστικό κόστος για τον προσδιορισμό κατεύθυνσης αναζήτησης σε κάθε επανάληψη οδηγεί σε υψηλότερους υπολογιστικούς χρόνους (ιδιαίτερα σε προβλήματα μεγαλύτερων διαστάσεων).

Για σ=0.9 (ανακριβής αναζήτηση):

- Τα προβλήματα σύγκλισης της μέγιστης κατάβασης για την συνάρτηση Rosenbrock οξύνονται ακόμη περισσότερο.
- Όλες οι μέθοδοι χρειάζονται μεγαλύτερο αριθμό επαναλήψεων μέχρι την σύγκλιση.
- Ο αριθμός επαναλήψεων της DFP είναι τώρα αισθητά μεγαλύτερος από της BFGS, ιδιαίτερα για μεγάλα προβλήματα ($n \geq 40$) πράγμα που οδηγεί σε αισθητά μειωμένους χρόνους της BFGS ως προς την DFP.
- Σε κάποιες περιπτώσεις (ιδιαίτερα για την BFGS) ο αυξημένος αριθμός επαναλήψεων για σ=0.9 συγκριτικά με σ=0.1 αντισταθμίζεται από τους μικρότερους χρόνους ανά επανάληψη (γιατί;) με αποτέλεσμα την μείωση του συνολικού υπολογιστικού χρόνου.

Είναι δύσκολο να δοθεί ένας γενικός κανόνας για τον βέλτιστο βαθμό ακρίβειας αναζήτησης επί γραμμής, θεωρείται όμως γενικά ότι η BFGS είναι αποτελεσματικότερη με σχετικά χαλαρή αναζήτηση επί γραμμής (π.χ. σ = 0.5).

Πρόβλημα n	DFP E	DFP CPU-s	BFGS E	BFGS CPU-s
Rosenbrock				
2	17	$<10^{-3}$	16	$<10^{-3}$
4	29	$<10^{-3}$	28	$<10^{-3}$
6	37	$<10^{-3}$	36	$<10^{-3}$
8	55	0.06	49	0.06
10	70	0.05	63	0.11
20	59	0.22	45	0.28
30	88	0.77	74	0.88
40	143	2.03	115	2.31
60	268	8.02	196	8.35
80	416	21.36	284	21.5

Πίνακας 3.5.2: Αποτελέσματα με τετραγωνική παρεμβολή (σ=0.1).

Ο Πίνακας 3.5.2 παρουσιάζει αντίστοιχα αποτελέσματα για τις δύο μεθόδους σχεδόν-Newton με σ=0.1, χρησιμοποιώντας όμως τώρα τετραγωνική (αντί της κυβικής) παρεμβολή κατά την αναζήτηση επί γραμμής. Συγκρίνοντας τον Πίνακα 3.5.2 με τις αντίστοιχες στήλες του Πίνακα 3.5.1 διαπιστώνουμε ένα ελαφρύ πλεονέκτημα της κυβικής παρεμβολής, τόσο σε αριθμό επαναλήψεων όσο και σε υπολογιστικό χρόνο.

Προβλήματα

3.1 Επιλύστε το Πρόβλημα 2.4c) εφαρμόζοντας την μέθοδο Newton.

Λύση

$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} - (\mathbf{G}^{(k)})^{-1}\mathbf{g}^{(k)}$

$\mathbf{G}^{(k)} = \mathbf{G} = \begin{bmatrix} 2 & 0 \\ 0 & 18 \end{bmatrix} \Rightarrow \mathbf{G}^{-1} = \begin{bmatrix} 1/2 & 0 \\ 0 & 1/18 \end{bmatrix}$

$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} - \begin{bmatrix} 1/2 & 0 \\ 0 & 1/18 \end{bmatrix} \begin{bmatrix} 2x_1^{(k)} - 4 \\ 18x_2^{(k)} - 18 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$: Σύγκλιση σε μια επανάληψη καθότι η f είναι τετραγωνική συνάρτηση.

3.2 Θεωρούμε τη συνάρτηση

$f(\mathbf{x}) = 2x_1^2 + x_2^2 - 2x_1x_2 + 2x_1^3 + x_1^4.$

a) Να προσδιορισθούν τα απόλυτα ελάχιστα της συνάρτησης.

b) Να προσδιορισθεί η περιοχή του R^2 όπου $\nabla^2 f > \mathbf{0}$. Να προσδιορισθούν κύκλοι με μέγιστη ακτίνα περί τα απόλυτα ελάχιστα, έτσι ώστε $\nabla^2 f > \mathbf{0}$ εντός του κύκλου και άρα είναι εφαρμόσιμη χωρίς δυσκολία η μέθοδος Newton.

Λύση

a) $\nabla f = \begin{bmatrix} 4x_1 - 2x_2 + 6x_1^2 + 4x_1^3 \\ 2x_2 - 2x_1 \end{bmatrix} = \mathbf{0} \Rightarrow x_1^* = x_2^*$

και άρα $4x_1^3 + 6x_1^2 + 2x_1 = 0 = 2x_1(2x_1^2 + 3x_1 + 1)$

$\Rightarrow x_1^* = 0$ ή $x_1^* = -0.5$ ή $x_1^* = -1$

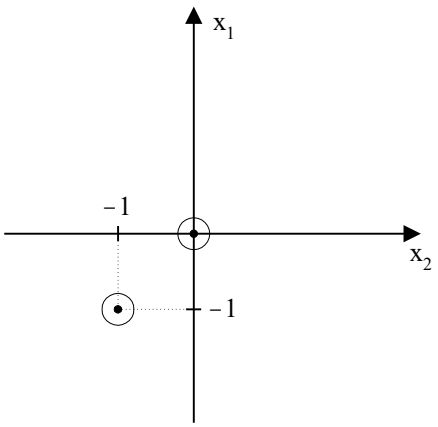
$$\nabla^2 f = \begin{bmatrix} 4 + 12x_1 + 12x_1^2 & -2 \\ -2 & 2 \end{bmatrix}$$

Ορίζουσες κυρίων υποπινάκων	$x_1 = 0$	$x_1 = -0.5$	$x_1 = -1$
$4+12x_1+12x_1^2$	$4 > 0$	$1 > 0$	$4 > 0$
$4+24x_1+24x_1^2$	$4 > 0$	$-2 < 0$	$4 > 0$
Τύπος στάσιμου σημείου	ελάχιστο	σαγματικό	ελάχιστο

$f(0,0) = f(-1, -1) = 0 \rightarrow$ απόλυτα ελάχιστα!

b) $\nabla^2 f > \mathbf{0}$: $12x_1^2 + 12x_1 + 4 > 0$ (ισχύει πάντα) και $24x_1^2 + 24x_1 + 4 > 0$

$\Leftrightarrow x_1 > -0.2113$ ή $x_1 < -0.7887$



μέγιστη ακτίνα: r = 0.2113 (η ίδια για αμφότερα γενικά ελάχιστα)

3.3 Θεωρούμε την τετραγωνική συνάρτηση

$f(\mathbf{x}) = 10x_1^2 + x_2^2$.

Να εφαρμοσθούν οι μέθοδοι σχεδόν-Newton πρώτης σειράς, DFP και BFGS με ακριβή αναζήτηση επί γραμμής προς ελαχιστοποίηση της συνάρτησης, εκκινώντας από $\mathbf{x}^{(1)} = [0.1 \quad 1]^T$ και $\mathbf{H}^{(1)} = \mathbf{I}$.

Λύση

Γενικό ελάχιστο προφανώς για $\mathbf{x}^* = \mathbf{0}$ με $\mathbf{G} = \begin{bmatrix} 20 & 0 \\ 0 & 2 \end{bmatrix} > \mathbf{0}$.

Εκκίνηση με $\mathbf{x}^{(1)} = \begin{bmatrix} 0.1 \\ 1 \end{bmatrix}$ και $\mathbf{H}^{(1)} = \mathbf{I} \Rightarrow \mathbf{g}^{(1)} = \begin{bmatrix} 20x_1^{(1)} \\ 2x_2^{(1)} \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$.

Και σε κάθε επανάληψη k:

$\mathbf{s}^{(k)} = -\mathbf{H}^{(k)}\mathbf{g}^{(k)}$

$\alpha^{(k)} = -\mathbf{g}^{(k)T}\mathbf{s}^{(k)}/\mathbf{s}^{(k)T}\mathbf{G}\mathbf{s}^{(k)}$ (βλ. λύση Προβλ. 2.5)

$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \alpha^{(k)}\mathbf{s}^{(k)}$

$\mathbf{g}^{(k+1)} = \begin{bmatrix} 20x_1^{(k+1)} \\ 2x_2^{(k+1)} \end{bmatrix}$

$\boldsymbol{\delta}^{(k)} = \mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}$, $\boldsymbol{\gamma}^{(k)} = \mathbf{g}^{(k+1)} - \mathbf{g}^{(k)}$

για μέθοδο πρώτης σειράς: $\mathbf{H}^{(k+1)}$ κατά (3.2.7)

για DFP: $\mathbf{H}^{(k+1)}$ κατά (3.2.9)

για BFGS: $\mathbf{H}^{(k+1)}$ κατά (3.2.10)

Αποτελέσματα μεθόδου πρώτης σειράς:

k	$\mathbf{x}^{(k)}$	$\mathbf{g}^{(k)}$	$\boldsymbol{\delta}^{(k-1)}$	$\boldsymbol{\gamma}^{(k-1)}$	$\mathbf{u}^{(k-1)}$	$\mathbf{H}^{(k)}$		$\mathbf{s}^{(k)}$	$\alpha^{(k)}$
1	$\frac{1}{10}$	2	–	–	–	1	0	-2	$\frac{1}{11}$
	1	2	–	–	–	0	1	-2	
2	$-\frac{9}{110}$	$-\frac{18}{11}$	$-\frac{2}{11}$	$-\frac{40}{11}$	$\frac{38}{11}$	$\frac{21}{382}$	$-\frac{19}{382}$	$\frac{360}{2101}$	$\frac{191}{400}$
	$\frac{9}{11}$	$\frac{18}{11}$	$-\frac{2}{11}$	$-\frac{4}{11}$	$\frac{2}{11}$	$-\frac{19}{382}$	$\frac{381}{382}$	$-\frac{3600}{2100}$	
3	0	0	$\frac{9}{110}$	$\frac{18}{11}$	$-\frac{1881}{21010}$	$\frac{1}{20}$	0		
	0	0	$-\frac{9}{11}$	$\frac{18}{11}$	$\frac{1881}{2101}$	0	$\frac{1}{2}$		

για DFP

$$\mathbf{H}^{(2)} = \begin{bmatrix} \frac{123}{2222} & -\frac{119}{2222} \\ -\frac{119}{2222} & \frac{2301}{2222} \end{bmatrix}, \quad \mathbf{s}^{(2)} = \begin{bmatrix} \frac{18}{101} \\ -\frac{180}{101} \end{bmatrix}, \quad \alpha^{(2)} = \frac{101}{220}.$$

για BFGS

$$\mathbf{H}^{(2)} = \begin{bmatrix} \frac{15}{242} & -\frac{19}{242} \\ -\frac{29}{242} & \frac{411}{242} \end{bmatrix}, \quad \mathbf{s}^{(2)} = \begin{bmatrix} \frac{36}{121} \\ -\frac{360}{121} \end{bmatrix}, \quad \alpha^{(2)} = \frac{11}{40}.$$

Κατά τα άλλα, όμοια αποτελέσματα όπως στον πίνακα.

3.4 Επιλύστε το Πρόβλημα 2.4 c) εφαρμόζοντας τις μεθόδους DFP και BFGS.

Λύση

Εκκίνηση με $\mathbf{x}^{(1)} = \begin{bmatrix} 4 \\ 4 \end{bmatrix}$, $\mathbf{H}^{(1)} = \mathbf{I} \Rightarrow \mathbf{g}^{(1)} = \begin{bmatrix} 2x_1^{(1)} - 4 \\ 18x_2^{(1)} - 18 \end{bmatrix} = \begin{bmatrix} 4 \\ 54 \end{bmatrix}$.

Και σε κάθε επανάληψη k:

$\mathbf{s}^{(k)} = -\mathbf{H}^{(k)}\mathbf{g}^{(k)}$

$\alpha^{(k)} = -\mathbf{g}^{(k)T}\mathbf{s}^{(k)}/\mathbf{s}^{(k)T}\mathbf{G}\mathbf{s}^{(k)}$ (βλ. λύση Προβλ. 2.5)

$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \alpha^{(k)}\,\mathbf{s}^{(k)}$

$$\mathbf{g}^{(k+1)} = \begin{bmatrix} 2x_1^{(k+1)} - 4 \\ 18x_2^{(k+1)} - 18 \end{bmatrix}$$

$\boldsymbol{\delta}^{(k)} = \mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}$, $\boldsymbol{\gamma}^{(k)} = \mathbf{g}^{(k+1)} - \mathbf{g}^{(k)}$

για DFP: $\mathbf{H}^{(k+1)}$ κατά (3.2.9)

για BFGS: $\mathbf{H}^{(k+1)}$ κατά (3.2.10).

Η πρώτη επανάληψη είναι όμοια με εκείνη της μέγιστης κατάβασης στη λύση του Προβλ. 2.4 c), καθότι $\mathbf{H}^{(1)} = \mathbf{I}$, και δίδει

$$\mathbf{x}^{(2)} = \begin{bmatrix} 3.777 \\ 0.985 \end{bmatrix} \Rightarrow \mathbf{g}^{(2)} = \begin{bmatrix} 3.554 \\ -0.27 \end{bmatrix}$$

άρα $\boldsymbol{\delta}^{(1)} = \begin{bmatrix} -0.223 \\ -3.015 \end{bmatrix}$, $\boldsymbol{\gamma}^{(1)} = \begin{bmatrix} -0.446 \\ -54.27 \end{bmatrix}$.

Η DFP δίδει όπως και η BFGS: $\mathbf{H}^{(2)} = \begin{bmatrix} 1 & -0.004 \\ -0.004 & 0.0556 \end{bmatrix}$

και στη 2η επανάληψη:

$\mathbf{s}^{(2)} = \begin{bmatrix} -3.555 \\ 0.0296 \end{bmatrix}$, $\alpha^{(2)} = 0.5$, $\mathbf{x}^{(3)} = \begin{bmatrix} 2 \\ 1 \end{bmatrix} = \mathbf{x}^*$.

Σύγκλιση DFP, BFGS σε 2 επαναλήψεις (με ακριβή αναζήτηση επί γραμμής).

4. ΜΕΘΟΔΟΙ ΣΥΖΥΓΩΝ ΚΛΙΣΕΩΝ

4.1 Περιγραφή Μεθόδων

Ενόψει του Θεωρήματος 2.4.1 που εγγυάται τετραγωνικό τερματισμό για μεθόδους συζυγών κατευθύνσεων, φαίνεται ενδιαφέρον να επιχειρηθεί ένας συνδυασμός ιδιοτήτων συζυγίας και μέγιστης κατάβασης, έτσι ώστε να συνδυαστούν τα πλεονεκτήματα αποτελεσματικότητας και απλότητας. Αυτός είναι ο στόχος των **μεθόδων συζυγών κλίσεων**. Μια μέθοδος παραγωγής συζυγών κλίσεων είναι η **μέθοδος Fletcher-Reeves** (1964) που έχει ως εξής:

$$\mathbf{s}^{(k)} = -\mathbf{g}^{(k)} + \beta^{(k-1)}\,\mathbf{s}^{(k-1)} \tag{4.1.1}$$

όπου $\beta^{(0)} = 0$, ενώ για $k > 1$

$$\beta^{(k)} = \frac{\mathbf{g}^{(k+1)T}\mathbf{g}^{(k+1)}}{\mathbf{g}^{(k)T}\mathbf{g}^{(k)}}. \tag{4.1.2}$$

Παρατηρούμε ότι η (4.1.1) δίδει, για k=1, $\mathbf{s}^{(1)} = -\mathbf{g}^{(1)}$, δηλαδή ότι η μέθοδος εκκινεί με την κατεύθυνση μέγιστης κατάβασης.

Θεώρημα 4.1.1: Για τετραγωνικές συναρτήσεις (1.2.12) η μέθοδος Fletcher-Reeves με ακριβή αναζήτηση επί γραμμής τερματίζει σε ένα στάσιμο σημείο $\mathbf{x}^{(m+1)}$ μετά από $m \le n$ επαναλήψεις και ισχύουν τα ακόλουθα για κάθε i, $1 \le i \le m$:

$$\mathbf{s}^{(i)T}\mathbf{G}\mathbf{s}^{(j)} = 0\ ,\ j=1, 2, \dots , i-1 \tag{4.1.3}$$

$$\mathbf{g}^{(i)T}\,\mathbf{g}^{(j)} = 0\ ,\ j=1, 2, \dots , i-1 \tag{4.1.4}$$

$$\mathbf{s}^{(i)T}\,\mathbf{g}^{(i)} = -\mathbf{g}^{(i)T}\mathbf{g}^{(i)}. \tag{4.1.5}$$

■

Η (4.1.3) είναι η συνθήκη συζυγίας (2.4.1), ενώ η συνθήκη (4.1.4) ονομάζεται συνθήκη ορθογώνιων κλίσεων. Η (4.1.5) εγγυάται ότι η $\mathbf{s}^{(i)}$ είναι κατεύθυνση κατάβασης αφού ικανοποιεί την (2.3.3). Η (4.1.5) είναι άμεση απόρροια των (2.3.4) (ακριβής αναζήτηση επί γραμμής) και (4.1.1) και άρα ισχύει και για μη τετραγωνικές συναρτήσεις, εφόσον εφαρμόζεται ακριβής αναζήτηση επί γραμμής.

Η εφαρμογή μεθόδων συζυγών κλίσεων σε γενικές, μη τετραγωνικές συναρτήσεις εμφανίζει δύο ερωτήματα. Πρώτον, πόσο ακριβής πρέπει να είναι η αναζήτηση επί γραμμής; Δεν υπάρχει γενική απάντηση στο ερώτημα αυτό που είναι καλύτερα να διερευνάται για κάθε συγκεκριμένη

εφαρμογή, θεωρείται όμως ότι η μέθοδος συνήθως συμπεριφέρεται καλύτερα για σχετικά ακριβή αναζήτηση επί γραμμής (π.χ. σ = 0.1). Δεύτερον, καθότι η μέθοδος δεν τερματίζει μετά από n επαναλήψεις για μη τετραγωνικές συναρτήσεις, εμφανίζεται το ερώτημα αν μπορεί η μέθοδος να συνεχισθεί, ως έχει, και μετά τις πρώτες n επαναλήψεις. Πρακτικές εφαρμογές και θεωρητικά αποτελέσματα δείχνουν ότι η μέθοδος είναι αποτελεσματικότερη αν η κατεύθυνση αναζήτησης επανατοποθετείται (reset) στην μέγιστη κατάβαση κάθε M επαναλήψεις, με την επιλογή M≤n, δηλαδή $\mathbf{s}^{(cM+1)} = -\mathbf{g}^{(cM+1)}$ για c = 1, 2, ...

Λόγω της μη ακριβούς αναζήτησης επί γραμμής, οι παραγόμενες κατευθύνσεις των συζυγών μεθόδων των Κεφαλαίων 3 και 4 μπορούν να εκφυλισθούν και να καταστούν σχεδόν ορθογώνιες προς το διάνυσμα κλίσης (δηλαδή $\mathbf{s}^{(k)T}\mathbf{g}^{(k)} \approx 0$), οπότε η βελτίωση της αντικειμενικής συνάρτησης είναι μηδαμινή, ή και ανοδικές (δηλαδή $\mathbf{s}^{(k)T}\mathbf{g}^{(k)} > 0$). Ο εκφυλισμός αυτού του είδους μπορεί να διαγνωσθεί μέσω της **ακόλουθης συνθήκης ικανής αρνητικότητας**

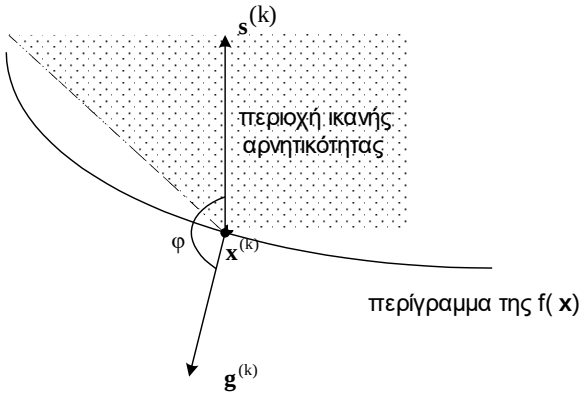
$$\mathbf{s}^{(k)T}\mathbf{g}^{(k)} < -\sigma_2 |\mathbf{s}^{(k)}| \cdot |\mathbf{g}^{(k)}|, \quad \sigma_2 > 0.$$

Όταν η ως άνω συνθήκη παραβιάζεται, συνιστάται **επανατοποθέτηση** της κατεύθυνσης αναζήτησης στην μέγιστη κατάβαση. Επειδή $\mathbf{s}^{(k)T}\mathbf{g}^{(k)} = |\mathbf{s}^{(k)}| \cdot |\mathbf{g}^{(k)}| \cdot \cos\varphi$, η ως άνω συνθήκη είναι ισοδύναμη με $\cos\varphi < -\sigma_2$ και άρα η επιλογή της παραμέτρου σ_2 καθορίζει το εύρος ενός επιτρεπτού τομέα για την κατεύθυνση αναζήτησης $\mathbf{s}^{(k)}$ (Σχ. 4.1.1).

Εκτός της μεθόδου Fletcher-Reeves υπάρχουν και άλλες δυνατότητες παραγωγής συζυγών κλίσεων. Μια απ'αυτές δίδεται από τον **τύπο συζυγούς κατάβασης** όπου η (4.1.1) παραμένει, αλλά η (4.1.2) αντικαθίσταται από

$$\boldsymbol{\beta}^{(k)} = -\frac{\mathbf{g}^{(k+1)T}\mathbf{g}^{(k+1)}}{\mathbf{g}^{(k)T}\mathbf{s}^{(k)}}. \qquad (4.1.6)$$

Μια ακόμη δυνατότητα δίδεται από την **μέθοδο Polak-Ribiere** (1969) όπου



Σχήμα 4.1.1: Συνθήκη ικανής αρνητικότητας.

$$\beta^{(k)} = \frac{(\mathbf{g}^{(k+1)} - \mathbf{g}^{(k)})^T \mathbf{g}^{(k+1)}}{\mathbf{g}^{(k)T} \mathbf{g}^{(k)}}. \qquad (4.1.7)$$

Οι μέθοδοι (4.1.6), (4.1.7) έχουν και αυτές τις ιδιότητες του Θεωρήματος 4.1.1 και άρα είναι, λόγω της (4.1.4), ισοδύναμες με την μέθοδο Fletcher-Reeves για τετραγωνικές συναρτήσεις. Σε μη τετραγωνικές συναρτήσεις όμως η αποτελεσματικότητα κάθε τύπου μπορεί να είναι διαφορετική.

Σημειώνουμε εδώ το προτέρημα "αυτόματης επανατοποθέτησης" της μεθόδου Polak-Ribiere όταν η παραγόμενη κατεύθυνση αναζήτησης $\mathbf{s}^{(k)}$ γίνεται σχεδόν ορθογώνια στο διάνυσμα κλίσης $\mathbf{g}^{(k)}$. Συγκεκριμένα, αν τα $\mathbf{s}^{(k)}$ και $\mathbf{g}^{(k)}$ είναι σχεδόν ορθογώνια, τότε (βλ. Σχ. 4.1.1) $\mathbf{x}^{(k+1)} \approx \mathbf{x}^{(k)}$ (αφού η βελτίωση κατά μήκος του περιγράμματος είναι μηδαμινή) και άρα $\mathbf{g}^{(k+1)} \approx \mathbf{g}^{(k)}$, οπότε από την (4.1.7) έχουμε $\beta^{(k)} \approx 0$, δηλαδή αυτόματη επανατοποθέτηση της κατεύθυνσης αναζήτησης $\mathbf{s}^{(k+1)}$ στην κατεύθυνση μέγιστης κατάβασης.

Σημειώνουμε ότι οι απαιτούμενοι υπολογισμοί για τον προσδιορισμό κατεύθυνσης αναζήτησης με μεθόδους συζυγών κλίσεων, δηλαδή οι (4.1.1) και (4.1.2) ή (4.1.6) ή (4.1.7), περιλαμβάνουν μόνον πολλαπλασιασμούς διανυσμάτων και άρα απαιτούν υπολογιστό φόρτο ανάλογο της διάστασης n του προβλήματος και όχι ανάλογο του n^2 όπως στις μεθόδους σχεδόν-Newton (λόγω πολλαπλασιασμών πινάκων) ή ανάλογο του n^3 όπως στη μέθοδο Newton (λόγω αντιστροφής πίνακος).

Η αποτελεσματικότητα των μεθόδων συζυγών κλίσεων μπορεί συχνά να βελτιωθεί αισθητά με αλλαγή κλίμακας (βλ. Κεφάλαιο 3.3). Αντικαθιστώντας $\mathbf{s}_y = \mathbf{A}\mathbf{s}_x$, $\mathbf{g}_y = \mathbf{A}^{-1}\mathbf{g}_x$, όπου $\mathbf{A}$ συμμετρικός πίνακας κλιμακοποίησης, στους τύπους (4.1.1), (4.1.2), (4.1.6), (4.1.7), έχουμε τους ακόλουθους τύπους **κλιμακοποιημένων συζυγών κλίσεων** (όπου οι δείκτες FR, ΣΚ, PR υποδηλώνουν τους τύπους Fletcher-Reeves, συζυγούς κατάβασης και Polak-Ribiere αντίστοιχα).

$$s_x^{(k)} = -\mathbf{A}^{-2}\mathbf{g}_x^{(k)} + \beta^{(k-1)}\mathbf{s}_x^{(k-1)} \qquad (4.1.8)$$

$$\beta_{FR}^{(k)} = \frac{\mathbf{g}_x^{(k+1)T}\mathbf{A}^{-2}\mathbf{g}_x^{(k+1)}}{\mathbf{g}_x^{(k)T}\mathbf{A}^{-2}\mathbf{g}_x^{(k)}} \qquad (4.1.9)$$

$$\beta_{\Sigma K}^{(k)} = \frac{\mathbf{g}_x^{(k+1)T}\mathbf{A}^{-2}\mathbf{g}_x^{(k+1)}}{\mathbf{g}_x^{(k)T}\mathbf{s}_x^{(k)}} \qquad (4.1.10)$$

$$\beta_{PR}^{(k)} = \frac{(\mathbf{g}_x^{(k+1)} - \mathbf{g}_x^{(k)})^T\mathbf{A}^{-2}\mathbf{g}_x^{(k+1)}}{\mathbf{g}_x^{(k)T}\mathbf{A}^{-2}\mathbf{g}_x^{(k)}}. \qquad (4.1.11)$$

Ο πίνακας κλιμακοποίησης $\mathbf{A}$ μπορεί να υπολογίζεται, π.χ. από την (3.3.4), στην αρχή κάθε επανατοποθέτησης (reset) και να διατηρείται σταθερός για όλες τις επαναλήψεις μέχρι την επόμενη επανατοποθέτηση. Με τον τρόπο αυτό, ο πίνακας κλιμακοποίησης, που για μη

γραμμικές αντικειμενικές συναρτήσεις εξαρτάται κατά την (3.3.4) από το **x**, μπορεί να μεταβάλεται κατά τη διάρκεια της αριθμητικής επίλυσης του αντίστοιχου προβλήματος.

4.2 Αριθμητικά Πειράματα

Ο Πίνακας 4.2.1 συνοψίζει τα αποτελέσματα αριθμητικών πειραμάτων για τα ίδια προβλήματα του Κεφαλαίου 3.4 με το ίδιο κριτήριο σύγκλισης και λοιπά χαρακτηριστικά του αριθμητικού αλγορίθμου και κυβική παρεμβολή κατά την αναζήτηση επί γραμμής.

Τα συμπεράσματα από αυτά τα περιορισμένης έκτασης αριθμητικά πειράματα συνοψίζονται ως εξής:

- Η μέθοδος Fletcher-Reeves εμφανίζει σημαντική δυσκολία στο πρόβλημα Rosenbrock με αριθμό μεταβλητών $n \geq 6$ λόγω εκφυλισμού των κατευθύνσεων αναζήτησης και έλλειψης μηχανισμού επανατοποθέτησης. Αντιθέτως η Polak-Ribiere δεν αντιμετωπίζει δυσκολίες λόγω του ενσωματωμένου μηχανισμού αυτόματης επανατοποθέτησης.
- Στο πρόβλημα Chebyquad, η Polak-Ribiere εμφανίζεται σημαντικά αποτελεσματικότερη της Fletcher-Reeves τόσο σε αριθμό επαναλήψεων όσο και σε υπολογιστικό φόρτο για παρόμοιους λόγους όπως παραπάνω.
- Η σύγκριση της Polak-Ribiere με τις μεθόδους σχεδόν-Newton (Πίνακας 3.5.1) για $\sigma=0.1$ αναδεικνύει το εξής γνωστό φαινόμενο: Οι μέθοδοι σχεδόν-Newton είναι πιο αποτελεσματικές σε αριθμό επαναλήψεων για κάθε n. Λόγω όμως της εμπλοκής πινάκων στους υπολογισμούς της κατεύθυνσης αναζήτησης, ο απαιτούμενος υπολογιστικός χρόνος αυξάνει γρηγορότερα με την διάσταση n του προβλήματος σε σύγκριση με την Polak-Ribiere, με αποτέλεσμα η τελευταία να γίνεται αποτελεσματικότερη σε υπολογιστικό χρόνο για αρκούντως υψηλό n.
- Η εφαρμογή ανακριβούς αναζήτησης επί γραμμής ($\sigma = 0.9$) φαίνεται να χειροτερεύει την αποτελεσματικότητα του αλγορίθμου Polak-Ribiere.

Ο Πίνακας 4.2.2 παρουσιάζει τα αποτελέσματα εφαρμογής των δύο μεθόδων συζυγούς κλίσης με επανατοποθέτηση (reset) κάθε $M=n$ επαναλήψεις (κυβική παρεμβολή, $\sigma=0.1$). Παρατηρούμε τα εξής:

- Η επανατοποθέτηση βοήθησε την μέθοδο Fletcher-Reeves να παρακάμψει το πρόβλημα σύγκλισης που παρουσιάζεται στη συνάρτηση Rosenbrock (Πίνακας 4.2.1) και να προσεγγίσει την αποτελεσματικότητα της Polak-Ribiere για όλα τα προβλήματα.
- Η επανατοποθέτηση φαίνεται να βελτιώνει ως επί το πλείστον την αποτελεσματικότητα των δύο μεθόδων.

Γενικώς, οι μέθοδοι σχεδόν-Newton έχουν συνήθως μεγαλύτερη αποτελεσματικότητα και ευρωστία που τις αναδεικνύουν σε πρώτη επιλογή υπό κανονικές συνθήκες. Όπως είδαμε όμως, υπάρχει ένα ιδιαίτερο πλεονέκτημα των μεθόδων συζυγών κλίσεων που ανάγεται στην απλή μορφή των (4.1.1), (4.1.2), (4.1.7) οι οποίες δεν απαιτούν αριθμητικές πράξεις πινάκων

για τον υπολογισμό της κατεύθυνσης αναζήτησης $\mathbf{s}^{(k)}$ σε κάθε επανάληψη. Επιπλέον, οι μέθοδοι Fletcher-Reeves και συζυγούς κατάβασης απαιτούν την αποθήκευση μόνο τριών και η μέθοδος Polak-Ribiere μόνο τεσσάρων n-διανυσμάτων. Για τους λόγους αυτούς οι μέθοδοι συζυγών κλίσεων είναι οι μόνες μέθοδοι που είναι εφαρμόσιμες σε πολύ μεγάλα προβλήματα εκατοντάδων χιλιάδων μεταβλητών. Έτσι, μέθοδοι συζυγών κλίσεων εφαρμόζονται με ιδιαίτερη επιτυχία σε πολύ μεγάλα προβλήματα, συγκλίνοντας μάλιστα σε αριθμό επαναλήψεων πολύ μικρότερο της διάστασης n του προβλήματος. Στην περίπτωση μεγάλων προβλημάτων συνιστάται επιλογή M κατά πολύ μικρότερου του n.

Πρόβλημα	σ = 0.1				σ = 0.9			
	Fletcher-Reeves		Polak-Ribiere		Fletcher-Reeves		Polak-Ribiere	
n	E	CPU-s	E	CPU-s	E	CPU-s	E	CPU-s
Rosenbrock								
2	50	$<10^{-3}$	16	$<10^{-3}$	76	$<10^{-3}$	283	0.05
4	369	0.05	46	$<10^{-3}$	367	0.17	1153	0.43
6	13431	3.25	151	0.05	426	0.22	1585	0.82
8	17974	5.44	158	0.05	17474	23.73	1523	1.05
10	12489	4.78	268	0.16	25538	40.31	1968	1.76
20	65731	41.63	283	0.28	66664	99.86	2031	3.52
30			147	0.22			1843	4.56
40	***δυσκολία***		260	0.5	***δυσκολία***		2519	8.68
60	***σύγκλισης***		110	0.33	***σύγκλισης***		2149	10.82
80			316	1.1			4127	27.95
Chebyquad								
2	21	$<10^{-3}$	8	$<10^{-3}$				
4	567	0.17	17	$<10^{-3}$				
6	248	0.11	18	$<10^{-3}$				
8	475	0.22	30	0.05				
10	412	0.28	30	0.06				
20	569	0.77	45	0.06				
30	813	1.54	38	0.05				
40	2311	5.38	43	0.11				
60	3729	12.58	47	0.28				
80	3529	15.82	47	0.33				

Πίνακας 4.2.1: Αποτελέσματα Fletcher-Reeves και Polak-Ribiere.

Πρόβλημα	Fletcher-Reeves		Polak-Ribiere	
n	E	CPU-s	E	CPU-s
Rosenbrock				
2	575	0.11	575	0.06
4	197	0,05	297	0.05
6	165	0,06	213	0,11
8	137	0,05	146	0,11
10	143	0,05	106	0,1
20	106	0,11	113	0,11
30	136	0,22	108	0,17
40	178	0,33	138	0,27
60	391	1,04	151	0,44
80	538	1,75	191	0,71
Chebyquad				
2	20	$<10^{-3}$	20	$<10^{-3}$
4	21	$<10^{-3}$	19	0.06
6	21	0,06	17	0.06
8	38	0,11	30	0,11
10	47	0,16	30	0,06
20	72	0.38	43	0,21
30	74	0,55	41	0,33
40	101	1,05	47	0,55
60	124	1,87	62	1,05
80	243	4,67	59	1,32

Πίνακας 4.2.2: Αποτελέσματα Fletcher-Reeves και Polak-Ribiere με επανατοποθέτηση (reset) κάθε M=n επαναλήψεις.

Προβλήματα

4.1 Να εφαρμοσθούν οι τρεις περιγραφείσες μέθοδοι συζυγών κλίσεων με ακριβή αναζήτηση επί γραμμής στη συνάρτηση του Προβλήματος 3.3 εκκινώντας από το ίδιο σημείο $\mathbf{x}^{(1)}$.

Λύση

Εκκίνηση με $\mathbf{x}^{(1)} = \begin{bmatrix} 0.1 \\ 1 \end{bmatrix} \Rightarrow \mathbf{g}^{(1)} = \begin{bmatrix} 20x_1^{(l)} \\ 2x_2^{(l)} \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$, επίσης $\beta^{(0)} = 0$

και σε κάθε επανάληψη k:

$\mathbf{s}^{(k)} = -\mathbf{g}^{(k)} + \beta^{(k-1)}\, \mathbf{s}^{(k-1)}$

$\alpha^{(k)} = -\mathbf{g}^{(k)T}\mathbf{s}^{(k)}/\mathbf{s}^{(k)T}\mathbf{G}\mathbf{s}^{(k)}$ (βλ. λύση Προβλ. 2.5)

$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \alpha^{(k)}\, \mathbf{s}^{(k)}$

$\mathbf{g}^{(k+1)} = \begin{bmatrix} 20x_1^{(k+l)} \\ 2x_2^{(k+l)} \end{bmatrix}$

$\beta^{(k)}$ σύμφωνα με (4.1.2) ή (4.1.6) ή (4.1.7).

Αποτελέσματα μεθόδου Fletcher-Reeves:

k	$\mathbf{x}^{(k)}$	$\mathbf{g}^{(k)}$	$\beta^{(k-1)}$	$\mathbf{s}^{(k)}$	$\alpha^{(k)}$
1	$\frac{1}{10}$ 1	2 2	(0)	-2 -2	$\frac{1}{11}$
2	$-\frac{9}{110}$ $\frac{9}{11}$	$-\frac{18}{11}$ $\frac{18}{11}$	$\frac{81}{121}$	$\frac{36}{121}$ $-\frac{360}{121}$	$\frac{11}{40}$
3	0 0	0 0			

Σύγκλιση σε 2 επαναλήψεις. Αποτελέσματα των άλλων δύο μεθόδων πανομοιότυπα.

4.2 Επιλύστε το Πρόβλημα 2.4 c) εφαρμόζοντας την μέθοδο Fletcher-Reeves.

Λύση

Εκκίνηση με $\mathbf{x}^{(1)} = \begin{bmatrix} 4 \\ 4 \end{bmatrix} \Rightarrow \mathbf{g}^{(1)} = \begin{bmatrix} 2x_1^{(1)} - 4 \\ 18x_2^{(1)} - 18 \end{bmatrix} = \begin{bmatrix} 4 \\ 54 \end{bmatrix}$

$\beta^{(0)} = 0$

$\mathbf{s}^{(1)} = -\mathbf{g}^{(1)} = -\begin{bmatrix} 4 \\ 54 \end{bmatrix}$

άρα όμοια αποτελέσματα με εκείνα της μέγιστης κατάβασης στη λύση του Προβλ. 2.4 c) για την πρώτη επανάληψη

$$\mathbf{x}^{(2)} = \begin{bmatrix} 3.77667 \\ 0.9854 \end{bmatrix} \Rightarrow \mathbf{g}^{(2)} = \begin{bmatrix} 3.55334 \\ -0.2628 \end{bmatrix}$$

$\beta^{(1)} = \mathbf{g}^{(2)T}\mathbf{g}^{(2)}/\mathbf{g}^{(1)T}\mathbf{g}^{(1)} = 12.6953/2932 = 0.0043299$

$$s^{(2)} = -\begin{bmatrix} 3.55334 \\ -0.2628 \end{bmatrix} - 0.0043299 \begin{bmatrix} 4 \\ 54 \end{bmatrix} = \begin{bmatrix} -3.57066 \\ 0.028985 \end{bmatrix}$$

$$\alpha^{(2)} = 0.5, \quad \mathbf{x}^{(3)} = \begin{bmatrix} 2 \\ 1 \end{bmatrix} = \mathbf{x}^*.$$

Σύγκλιση σε 2 επαναλήψεις (με ακριβή αναζήτηση επί γραμμής).

ΜΕΡΟΣ ΔΕΥΤΕΡΟ:

ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΜΕ ΠΕΡΙΟΡΙΣΜΟΥΣ

5. ΕΙΣΑΓΩΓΗ ΣΤΗ ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΜΕ ΠΕΡΙΟΡΙΣΜΟΥΣ

5.1 Γενικά Στοιχεία

Οι αλγόριθμοι βελτιστοποίησης χωρίς περιορισμούς αποτελούν συχνά τη βάση για την ανάπτυξη αλγορίθμων βελτιστοποίησης με περιορισμούς. Θα πρέπει όμως να τονισθεί ότι οι αλγόριθμοι βελτιστοποίησης με περιορισμούς είναι πιο πολύπλοκοι, πράγμα που μπορεί να δυσκολέψει την αριθμητική σύγκριση αποτελεσματικότητας εναλλακτικών αλγορίθμων για γενικά προβλήματα βελτιστοποίησης. Για κάποια συγκεκριμένη εφαρμογή βελτιστοποίησης με περιορισμούς, μπορεί να μην είναι εύκολο να προσδιορισθεί εκ των προτέρων ο αποτελεσματικότερος αλγόριθμος, και να απαιτηθούν πειραματικές συγκρίσεις για την επιλογή του καταλληλοτέρου.

Η γενική μορφή των προβλημάτων βελτιστοποίησης με περιορισμούς που θα εξετάσουμε είναι:

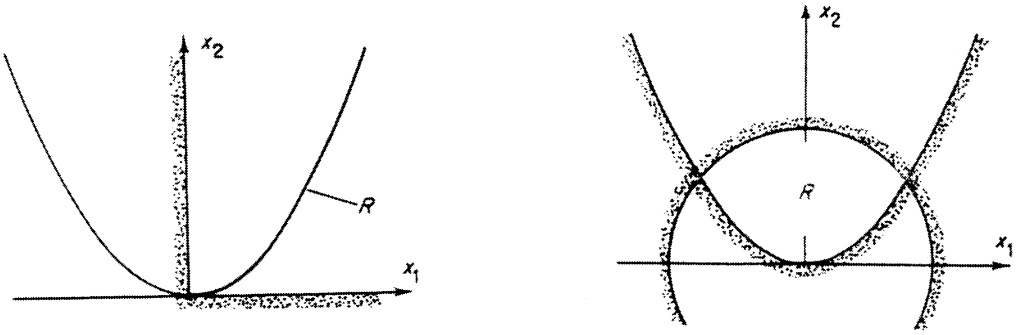
$$\begin{array}{lll} \text{Ελαχιστοποίηση} & f(\mathbf{x}), \quad \mathbf{x}\in R^n & \\ \text{λαμβάνοντας υπόψη} & c_i(\mathbf{x}) = 0, \quad i\in E & \qquad (5.1.1) \\ & c_i(\mathbf{x}) \geq 0, \quad i\in I & \end{array}$$

όπου $f(\mathbf{x})$ είναι η **αντικειμενική συνάρτηση** και $c_i(\mathbf{x})$, $i=1,...,p$, (όπου $p=|E\cup I|$) είναι **συναρτήσεις περιορισμών**. E και I είναι τα σύνολα δεικτών των περιορισμών ισότητας και ανισότητας, αντίστοιχα. Ο τρόπος διατύπωσης των περιορισμών είναι αρκετά γενικός, π.χ. ένας περιορισμός $d_i(\mathbf{x})\leq b$ μετατρέπεται σε $b-d_i(\mathbf{x})\geq 0$ κλπ. Προφανώς δεν μπορούμε να έχουμε περισσότερους από n ανεξάρτητους περιορισμούς ισότητας, δηλαδή $|E| \leq n$ (γιατί;).

Ένα σημείο $\mathbf{x}'$ που ικανοποιεί όλους τους περιορισμούς ονομάζεται **επιτρεπτό σημείο** και το σύνολο όλων των επιτρεπτών σημείων είναι η **επιτρεπτή περιοχή** $\mathcal{R}$. Συμβολίζοντας με $\mathcal{R}_i$ την **επιτρεπτή περιοχή του περιορισμού** i, έχουμε προφανώς $\mathcal{R}=\mathcal{R}_1\cap\mathcal{R}_2\cap...\cap\mathcal{R}_p$. Ένα τοπικό ελάχιστο $\mathbf{x}^*$ ορίζεται με $f(\mathbf{x}^*)\leq f(\mathbf{x})$ για κάθε **επιτρεπτό x** από μιά αρκούντως μικρή περιοχή περί το $\mathbf{x}^*$. Ένα γενικό ελάχιστο ικανοποιεί $f(\mathbf{x}^*)\leq f(\mathbf{x})$ $\forall\mathbf{x}\in\mathcal{R}$.

Για n=2 μπορούμε να αναπαραστήσουμε γραφικά έναν περιορισμό μέσω της γραφικής παράστασης του μηδενικού του περιγράμματος. Για περιορισμούς ισότητας, τα σημεία του μηδενικού περιγράμματος είναι τα επιτρεπτά σημεία. Για περιορισμούς ανισότητας, το μηδενικό περίγραμμα είναι το σύνορο της επιτρεπτής περιοχής, και αυτό αναπαριστάται συνήθως σκιάζοντας τη μη επιτρεπτή πλευρά του περιγράμματος, βλ. Σχ. 5.1.1. Στην περίπτωση (a) του Σχ. 5.1.1 έχουμε τους περιορισμούς $x_2 = x_1^2$, $\mathbf{x}\geq\mathbf{0}$, δηλαδή $c_1(\mathbf{x}) = x_2 - x_1^2$, $c_2(\mathbf{x}) = x_1$, $c_3(\mathbf{x}) = x_2$ και $E = \{1\}$, $I = \{2,3\}$. Στην περίπτωση (b) του Σχ. 5.1.1 έχουμε τους περιορισμούς $x_2\geq x_1^2$, $x_1^2+x_2^2\leq 1$, δηλαδή $c_1(\mathbf{x})=x_2-x_1^2$, $c_2(\mathbf{x}) = 1-x_1^2-x_2^2$, $I = \{1,2\}$ και $E=\{\}$.

Η διατύπωση (5.1.1) καλύπτει πολλά προβλήματα βελτιστοποίησης συμπεριλαμβανομένων και διακριτών τιμών για το **x**. Επί παραδείγματι, ο διακριτός περιορισμός $x\in\{X_1, X_2,..., X_\rho\}$ μπορεί να εκφρασθεί ως περιορισμός ισότητας



a) b)

Σχήμα 5.1.1. Παραδείγματα επιτρεπτών περιοχών.

$$c(x) = (x-X_1)(x-X_2) \dots (x-X_\rho) = 0. \qquad (5.1.2)$$

Εν τούτοις, οι αλγόριθμοι που θα αναπτυχθούν παρακάτω δεν είναι κατάλληλοι για την επεξεργασία αυτού του είδους προβλημάτων που ανήκουν στον τομέα **ακέραιου** ή **διακριτού προγραμματισμού**. Σημειώνεται επίσης, ότι η διατύπωση (5.1.1) δεν καλύπτει περιορισμούς τύπου $c_i(\mathbf{x}) > 0$.

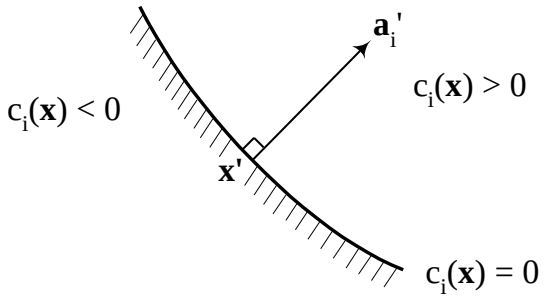
Θα κάνουμε τις γενικές παραδοχές ότι οι συναρτήσεις $c_i(\mathbf{x})$ είναι συνεχείς, εξού συνεπάγεται ότι η $\mathcal{R}$ είναι κλειστή, και ότι η $f(\mathbf{x})$ είναι συνεχής για κάθε $\mathbf{x}\in\mathcal{R}$. Αν, επί πλέον, η $\mathcal{R}$ δεν είναι κενή και είναι φραγμένη (δηλαδή $\exists a>0$: $\|\mathbf{x}\|<a$ $\forall\mathbf{x}\in\mathcal{R}$), τότε συνεπάγεται ότι υπάρχει λύση (δηλαδή απόλυτο ελάχιστο) $\mathbf{x}^*$. Άλλως το πρόβλημα μπορεί να είναι **μη φραγμένο** (δηλαδή $\exists\mathbf{x}\in\mathcal{R}$: $f(\mathbf{x})\to -\infty$) και, γενικότερα, να μην έχει λύση. Το τελευταίο ισχύει προφανώς αν $\mathcal{R}=\{\}$, δηλαδή αν οι περιορισμοί δεν είναι συμβατοί, αλλά και σε άλλες περιπτώσεις που αναφέρθηκαν στο Κεφάλαιο 2.1. Οι περισσότερες πρακτικές μέθοδοι επίλυσης απαιτούν πάντως την ισχυρότερη παραδοχή f, $c_i \in C^1$ ή C^2.

Το διάνυσμα ∇c_i συχνά σημειώνεται $\mathbf{a}_i$ και ονομάζεται **κάθετο διάνυσμα** του περιορισμού c_i. Το $\mathbf{a}_i$ είναι προφανώς η i–τη στήλη του Ιακωβιανού πίνακα (βλ. (1.2.6)) $\nabla\mathbf{c}(\mathbf{x})=\mathbf{A}$ της διανυσματικής συνάρτησης περιορισμών $\mathbf{c}(\mathbf{x})=[c_1(\mathbf{x})\ c_2(\mathbf{x}) \dots c_p(\mathbf{x})]^T$. Το διάνυσμα $\mathbf{a}_i' = \mathbf{a}_i(\mathbf{x}')$ είναι η κατεύθυνση μέγιστης αύξησης της $c_i(\mathbf{x})$ στο $\mathbf{x}'$, και αν $c_i(\mathbf{x}')=0$ και $i\in I$ τότε η κατεύθυνση αυτή δείχνει προς την επιτρεπτή πλευρά του περιορισμού (γιατί;) και είναι κάθετη προς το μηδενικό περίγραμμα (βλ. Σχήμα 5.1.2). Αν οι συνεχείς συναρτήσεις f, c_i δεν έχουν συνεχείς παραγώγους, το πρόβλημα ανάγεται στην **μη λεία** ή **μη παραγωγίσιμη βελτιστοποίηση** και απαιτούνται ειδικές μέθοδοι επίλυσης.

Οι **ενεργοί περιορισμοί** σε κάποιο $\mathbf{x}'$ ορίζονται από το σύνολο δεικτών

$$\mathcal{A}' = \mathcal{A}(\mathbf{x}') = \{i \mid c_i(\mathbf{x}')=0\}. \qquad (5.1.3)$$

Άρα ένας περιορισμός είναι ενεργός στο $\mathbf{x}'$ αν το $\mathbf{x}'$ βρίσκεται στο όριο της επιτρεπτής περιοχής του συγκεκριμένου περιορισμού. Προφανώς αν $\mathbf{x}'\in\mathcal{R}$ τότε $E\subset\mathcal{A}'$. Ορίζουμε $\mathcal{A}^*$ =



Σχήμα 5.1.2. Το κάθετο διάνυσμα.

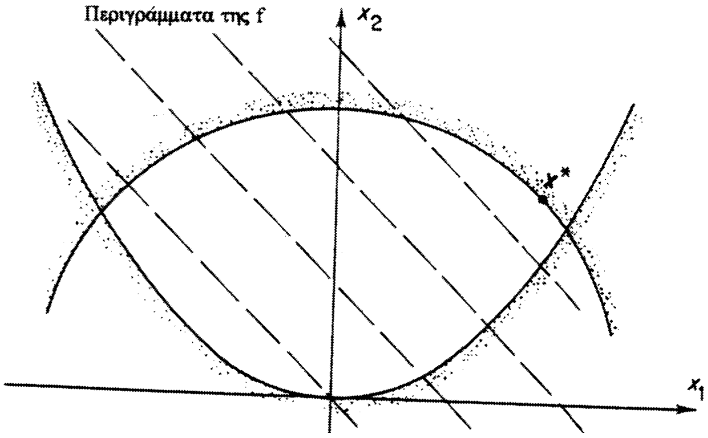
$\mathcal{A}(\mathbf{x}^*)$. Αν το $\mathcal{A}^*$ είναι γνωστό, τότε οι περιορισμοί εκτός $\mathcal{A}^*$ μπορούν να αγνοηθούν (τοπικά) και το πρόβλημα μπορεί να επιλυθεί ως πρόβλημα περιορισμών ισότητας με $E=\mathcal{A}^*$. Αυτό είναι δυνατό επειδή περιορισμοί $i\notin\mathcal{A}^*$ μπορούν να μεταβληθούν κατά λίγο χωρίς να επηρεάσουν την τοπική λύση, πράγμα που γενικά δεν ισχύει για τους ενεργούς περιορισμούς.

Παράδειγμα 5.1.1: Θεωρούμε ελαχιστοποίηση της $f(\mathbf{x})=-x_1-x_2$ λαμβάνοντας υπόψη $x_2 \geq x_1^2$ και $x_1^2 + x_2^2 \leq 1$ (Σχ. 5.1.3). Η λύση (γραφικά) είναι $\mathbf{x}^* = [1/\sqrt{2} \quad 1/\sqrt{2}]^T$, όπου το περίγραμμα της $f(\mathbf{x})$ εφάπτεται του μοναδιαίου κύκλου. Έχουμε άρα $\mathcal{A}^*=\{2\}$, δηλαδή ο περιορισμός κύκλου είναι ενεργός, ενώ ο περιορισμός παραβολής είναι ανενεργός και, όντως, μπορεί να μεταβληθεί κατά λίγο ή να αγνοηθεί, χωρίς αυτό να επηρεάζει την λύση $\mathbf{x}^*$.

■

Αλγόριθμοι επίλυσης του προβλήματος (5.1.1) είναι συνήθως επαναληπτικοί, εκκινούν δηλαδή από ένα δεδομένο σημείο $\mathbf{x}^{(1)}$ και παράγουν μια σειρά σημείων $\mathbf{x}^{(1)}$, $\mathbf{x}^{(2)}$, $\mathbf{x}^{(3)}$,... που πρέπει να συγκλίνει στο $\mathbf{x}^*$. Αν το $\mathbf{x}^*$ είναι μέλος της σειράς, λέμε ότι η μέθοδος **τερματίζει**.

Μπορούμε να διακρίνουμε στην βελτιστοποίηση υπό περιορισμούς δύο κατηγορίες, **προγραμματισμό με γραμμικούς περιορισμούς** και **μη γραμμικό προγραμματισμό**, με αρκετές διαφορετικές ιδιότητες για κάθε μια. Στον προγραμματισμό με γραμμικούς περιορισμούς έχουμε αποκλειστικά γραμμικές συναρτήσεις περιορισμών, δηλαδή



Σχήμα 5.1.3. Ενεργοί και ανενεργοί περιορισμοί.

$c_i(\mathbf{x}) = \mathbf{a}_i^T\mathbf{x} - b_i$, των οποίων τα σύνορα είναι υπερεπίπεδα και τα κάθετα διανύσματα $\mathbf{a}_i$ είναι σταθερά. Αν, επιπλέον, η αντικειμενική συνάρτηση είναι γραμμική ή τετραγωνική, έχουμε τον **γραμμικό** ή **τετραγωνικό προγραμματισμό**, αντίστοιχα. Μια ειδική περίπτωση μη γραμμικού προγραμματισμού είναι ο **γεωμετρικός προγραμματισμός** όπου οι συναρτήσεις f, c_i έχουν πολυωνυμική μορφή. Για την επίλυση προβλημάτων προγραμματισμού με γραμμικούς περιορισμούς, υπάρχουν αποτελεσματικοί αλγόριθμοι, που μάλιστα στην περίπτωση γραμμικού ή τετραγωνικού προγραμματισμού τερματίζουν (δηλαδή προσδιορίζουν με ακρίβεια το ζητούμενο $\mathbf{x}^*$). Για τα γενικότερα προβλήματα μη γραμμικού προγραμματισμού υπάρχουν μεν αρκετοί αλγόριθμοι και αρκετή συσσωρευμένη εμπειρία εφαρμογής των, δεν υπάρχει όμως μια γενική μέθοδος που να είναι αποδεδειγμένα η καλύτερη για κάθε είδους προβλήματα.

Γραμμικοποίηση μιας συνάρτησης περιορισμών περί κάποιο $\mathbf{x}^{(k)}$ επιτυγχάνεται μέσω της σειράς Taylor

$$c_i(\mathbf{x}^{(k)} + \boldsymbol{\delta}) \approx l_i^{(k)}(\boldsymbol{\delta}) = c_i^{(k)} + \mathbf{a}_i^{(k)T}\boldsymbol{\delta} \qquad (5.1.4)$$

όπου $l_i^{(k)}(\boldsymbol{\delta})$ είναι η γραμμικοποιημένη συνάρτηση περί το επαναληπτικό $\mathbf{x}^{(k)}$.

5.2 Απαλοιφή και άλλοι Μετασχηματισμοί

Υπάρχουν πολλές δυνατότητες μετασχηματισμών ενός προβλήματος βελτιστοποίησης με περιορισμούς, που δίδουν στο πρόβλημα μια ευκολότερα επιλυόμενη μορφή. Είναι όμως σημαντικό να τονισθεί ότι αυτές οι διαδικασίες εμπεριέχουν συχνά κινδύνους, π.χ. επειδή οι λύσεις του αρχικού και του μετασχηματισμένου προβλήματος μπορεί να μην είναι ακριβώς ταυτόσημες ή επειδή μερικές μέθοδοι επίλυσης μπορεί να μη λειτουργούν αποτελεσματικά στο μετασχηματισμένο πρόβλημα. Θα αναφερθούν στο παρόν κεφάλαιο, αλλά και αργότερα, μια σειρά τέτοιων μετασχηματισμών.

Η απλούστερη δυνατότητα για προβλήματα με περιορισμούς ισότητας είναι να χρησιμοποιηθούν οι εξισώσεις με στόχο την **απαλοιφή** μερικών μεταβλητών του προβλήματος. Αν έχουμε m περιορισμούς ισότητας $\mathbf{c}(\mathbf{x})=\mathbf{0}$ που μπορούν να μετασχηματισθούν έτσι ώστε

$$\mathbf{x}_1 = \boldsymbol{\varphi}(\mathbf{x}_2) \qquad (5.2.1)$$

όπου $\mathbf{x}_1 \in R^m$ και $\mathbf{x}_2 \in R^{n-m}$ είναι τμήματα του $\mathbf{x}$, τότε μπορούμε να αντικαταστήσουμε την αρχική αντικειμενική συνάρτηση $f(\mathbf{x}_1,\mathbf{x}_2)$ με

$$\psi(\mathbf{x}_2)=f(\boldsymbol{\varphi}(\mathbf{x}_2),\mathbf{x}_2) \qquad (5.2.2)$$

και να ελαχιστοποιήσουμε την $\psi(\mathbf{x}_2)$ ως προς $\mathbf{x}_2$ χωρίς περιορισμούς. Οι παράγωγοι της ψ μπορούν εύκολα να υπολογισθούν βάσει των παραγώγων των f, **c** (βλ. Πρόβλημα 5.4). Ας σημειωθεί όμως ότι η μέθοδος απαλοιφής μπορεί σε μερικές περιπτώσεις να αποτύχει (βλ. Πρόβλημα 5.2b).

Μέθοδοι απαλοιφής δεν μπορούν να εφαρμοσθούν άμεσα σε περιορισμούς ανισότητας, εκτός αν είναι γνωστό εκ των προτέρων το σύνολο $\mathcal{A}^*$. Μπορεί όμως κανείς να πειραματισθεί κάνοντας υποθέσεις ως προς το $\mathcal{A}^*$ (δηλαδή θεωρώντας κάποια εκτίμηση $\hat{\mathcal{A}}^*$), επιλύοντας το πρόβλημα με περιορισμούς ισότητας μόνον, και ελέγχοντας εκ των υστέρων αν η επιτευχθείσα λύση είναι επιτρεπτή ως προς τους περιορισμούς που αγνοήθηκαν. Αν όχι, μπορεί ένας από τους παραβιασθέντες περιορισμούς να συμπεριληφθεί στο $\hat{\mathcal{A}}^*$ και να λυθεί το προκύπτον πρόβλημα εκ νέου. Αν ναι, πρέπει να ελεγχθεί αν η εγκατάληψη (προς επιτρεπτή κατεύθυνση) ενός ενεργού περιορισμού, που περιέχεται στο $\hat{\mathcal{A}}^*$, αυξάνει ή μειώνει την αντικειμενική συνάρτηση. Αν η αντικειμενική συνάρτηση αυξάνει για κάθε εγκατάληψη περιορισμού, τότε το πρόβλημα έχει λυθεί. Άλλως πρέπει να αφαιρεθεί από το $\hat{\mathcal{A}}^*$ ένας περιορισμός, του οποίου η εγκατάληψη μειώνει την αντικειμενική συνάρτηση, και το πρόβλημα να επιλυθεί εκ νέου κ.ο.κ. Η περιγραφείσα διαδικασία μπορεί να εκτελεσθεί από τον επιλύοντα σε μικρά προβλήματα. Θα γνωρίσουμε όμως σε παρακάτω κεφάλαια μεθόδους που εφαρμόζουν με συστηματικό και αυτόματο τρόπο την παραπάνω διαδικασία.

Μερικοί άλλοι μετασχηματισμοί συσχετίζουν περιορισμένα προβλήματα ισοτήτων και ανισοτήτων. Επί παραδείγματι, ο περιορισμός ισότητας $c_i(\mathbf{x}) = 0$ είναι ισοδύναμος με τους περιορισμούς ανισότητας $c_i(\mathbf{x}) \geq 0$ και $c_i(\mathbf{x}) \leq 0$. Αν και θεωρητικά ο μετασχηματισμός αυτός είναι χρήσιμος, πρακτικά έχει μικρή σημασία, αφού οι αριθμητικοί αλγόριθμοι επίλυσης αντιμετωπίζουν ευκολότερα περιορισμούς ισότητας παρά ανισότητας. Ένας άλλος μετασχηματισμός αντικαθιστά $c_i(\mathbf{x}) \geq 0$ με $c_i(\mathbf{x}) - z_i = 0$ **και** $z_i \geq 0$, εισάγοντας στο πρόβλημα τη νέα **μεταβλητή χαλαρότητας** z. Εναλλακτικά μπορεί να χρησιμοποιηθεί μια **τετραγωνική μεταβλητή χαλαρότητας** $c_i(\mathbf{x}) - y_i^2 = 0$ οπότε περιττεύει μεν ο περιορισμός μη αρνητικότητας, μπορούν όμως να εμφανισθούν άλλες δυσκολίες όπως θα δούμε παρακάτω.

Ένας άλλος μετασχηματισμός απαλείφει περιορισμούς $x_i \geq 0$ εισάγοντας στο πρόβλημα μια νέα μεταβλητή $y_i^2 = x_i$ που αντικαθιστά την x_i. Παρομοίως, ο περιορισμός $l_i \leq x_i \leq u_i$ μπορεί να απαλειφθεί αντικαθιστώντας $x_i = l_i + (u_i - l_i)\sin^2 y_i$ στην διατύπωση του προβλήματος. Αυστηροί περιορισμοί $x_i > 0$ μπορούν να ληφθούν υπόψη (και να απαλειφθούν) με την αντικατάσταση $x_i = e^{y_i}$. Και οι μετασχηματισμοί αυτοί όμως δεν είναι πάντα χωρίς προβλήματα. Επί παραδείγματι, το πρόβλημα $\min(x^2)$ λαμβάνοντας υπόψη $x \geq 0$ μετασχηματίζεται με $y^2 = x$ σε $\min(y^4)$, όπου όμως το μετασχηματισθέν πρόβλημα έχει μονήρη Χεσιανό πίνακα στη λύση, πράγμα που μειώνει την αποτελεσματικότητα των αλγορίθμων επίλυσης.

Προβλήματα

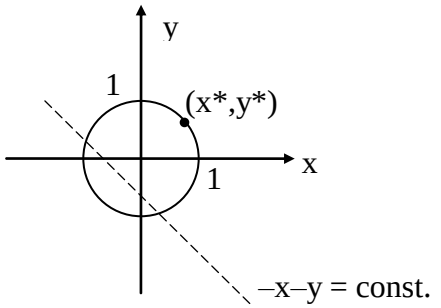
5.1 Να επιλυθεί το πρόβλημα ελαχιστοποίησης –x–y λαμβάνοντας υπόψη $x^2+y^2 = 1$

(a) γραφικά

(b) μέσω απαλοιφής του x.

Λύση

a)



$$x^* = \frac{\sqrt{2}}{2} = y^*$$

b) $x^2 + y^2 = 1 \Rightarrow x = \pm\sqrt{1-y^2}$

αν $x = \sqrt{1-y^2} \Rightarrow \psi(y) = -\sqrt{1-y^2} - y$

$$\psi'(y) = \frac{y}{\sqrt{1-y^2}} - 1 = \frac{y-\sqrt{1-y^2}}{\sqrt{1-y^2}} = 0 \Rightarrow y^2 = 1-y^2 \text{ και } y \geq 0 \Rightarrow y = \frac{\sqrt{2}}{2}$$

$$\psi''(y) = \frac{1-y^2+y^2}{(1-y^2)^{3/2}} > 0 \text{ για } y = \frac{\sqrt{2}}{2} \Rightarrow \text{ελάχιστο}$$

άρα $y^* = \frac{\sqrt{2}}{2}$ και $x^* = \sqrt{1-y^2} = \frac{\sqrt{2}}{2}$.

Αν $x = -\sqrt{1-y^2} \Rightarrow \psi(y) = \sqrt{1-y^2} - y$

$$\psi'(y) = -\frac{y}{\sqrt{1-y^2}} - 1 = \frac{y+\sqrt{1-y^2}}{\sqrt{1-y^2}} = 0 \Rightarrow y^2 = 1-y^2 \text{ και } y \leq 0 \Rightarrow y = -\frac{\sqrt{2}}{2}$$

$$\psi''(y) = -\frac{1}{(1-y^2)^{3/2}} < 0 \text{ για } y = -\frac{\sqrt{2}}{2} \Rightarrow \text{μέγιστο.}$$

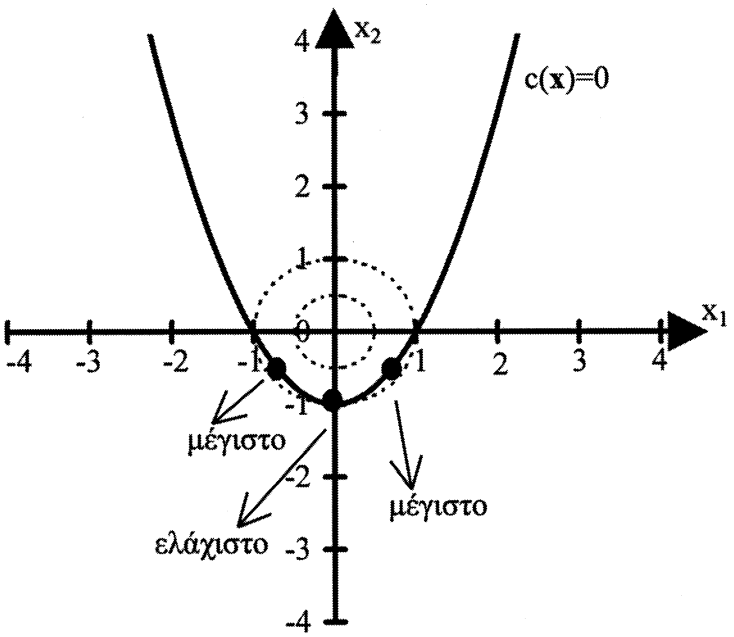
5.2. Να ευρεθούν μέγιστα και ελάχιστα της αντικειμενικής συνάρτησης $f(\mathbf{x}) = 4 - x_1^2 - x_2^2$ λαμβάνοντας υπόψη $c(\mathbf{x}) = 1 - x_1^2 + x_2 = 0$

(a) γραφικά

(b) μέσω απαλοιφής του x_1.

Λύση

a)



Μέγιστα:
$\mathbf{x}^* = [\sqrt{2}/2 \;\; -0.5]^T$
$\mathbf{x}^* = [-\sqrt{2}/2 \;\; -0.5]^T$
Ελάχιστο:
$\mathbf{x}^* = [0 \;\; -1]^T$
(Απόλυτο ελάχιστο: $x_1 \to \pm\infty$)

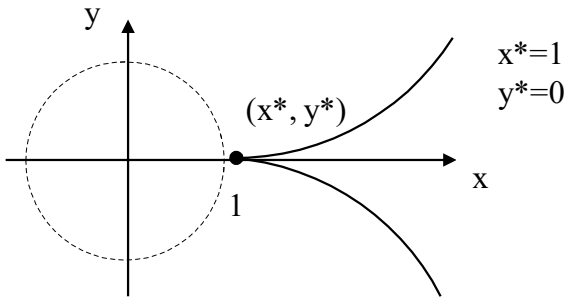
b) $1 - x_1^2 + x_2 = 0 \Rightarrow x_1^2 = 1 + x_2$ (απαλοιφή)
$f(\mathbf{x}) = \psi(x_2) = 3 - x_2 - x_2^2$
$\psi'(x_2) = -1 - 2x_2 = 0 \Rightarrow x_2^* = -0.5$
$\psi''(x_2) = -2 < 0 \Rightarrow$ μέγιστο!
$x_1^* = \pm\sqrt{2}/2$, άρα δύο μέγιστα $\mathbf{x}^* = [\pm\sqrt{2}/2 \;\; -0.5]^T$.

Το τοπικό ελάχιστο $[0 \;\; -1]^T$ "χάθηκε" στην απαλοιφή. Γιατί; Ο τύπος απαλοιφής $x_1^2 = 1 + x_2$ προφανώς ισχύει μόνο για $x_2 \geq -1$ και επομένως η ελαχιστοποίηση της $\psi(x_2)$ πρέπει να γίνει λαμβάνοντας υπόψη τον περιορισμό $x_2 \geq -1$, οπότε το $x_2^* = -1$ δίδει όντως τοπικό ελάχιστο!

5.3 Να επιλυθεί το πρόβλημα ελαχιστοποίησης $x^2 + y^2$ λαμβάνοντας υπόψη $(x-1)^3 = y^2$
(a) γραφικά
(b) μέσω απαλοιφής του y.

Λύση

a)



b) $y^2 = (x-1)^3 \Rightarrow \psi(x) = x^2+(x-1)^3$

$\psi'(x) = 2x+3(x-1)^2 = 3x^2-4x+3 = 0$ δεν έχει λύση! $\Rightarrow$ δεν υπάρχει ελάχιστο της $\psi(x)$. Τι συνέβει;

Λόγω $y^2 = (x-1)^3$, πρέπει $x\geq 1$, άρα η $\psi(x)$ πρέπει να ελαχιστοποιηθεί λαμβάνοντας υπόψη $x\geq 1$, οπότε έχουμε πράγματι το ελάχιστο $x^* = 1$.

5.4 Να αποδειχθούν οι τύποι

$$\nabla_2\boldsymbol{\varphi} = -\mathbf{A}_2\,\mathbf{A}_1^{-1}$$
$$\nabla_2\psi = \mathbf{g}_2 - \mathbf{A}_2\,\mathbf{A}_1^{-1}\mathbf{g}_1$$

όπου $\boldsymbol{\varphi}(\mathbf{x}_2)$, $\psi(\mathbf{x}_2)$ κατά (5.2.1), (5.2.2) και

$$\nabla = \begin{bmatrix}\nabla_1\\ \nabla_2\end{bmatrix},\ \mathbf{g} = \begin{bmatrix}\mathbf{g}_1\\ \mathbf{g}_2\end{bmatrix},\ \mathbf{A}=\begin{bmatrix}\mathbf{A}_1\\ \mathbf{A}_2\end{bmatrix}.$$

Λύση

Παραγωγίζοντας την $\mathbf{c}(\boldsymbol{\varphi}(\mathbf{x}_2), \mathbf{x}_2) = \mathbf{0}$ ως προς $\mathbf{x}_2$ έχουμε

$$\nabla_2\boldsymbol{\varphi}\cdot\nabla_1\mathbf{c} + \nabla_2\mathbf{c} = \nabla_2\boldsymbol{\varphi}\,\mathbf{A}_1 + \mathbf{A}_2 = \mathbf{0} \Rightarrow \nabla_2\boldsymbol{\varphi} = -\mathbf{A}_2\,\mathbf{A}_1^{-1}.$$

Παραγωγίζοντας την $\psi(\mathbf{x}_2) = f(\boldsymbol{\varphi}(\mathbf{x}_2),\mathbf{x}_2)$ ως προς $\mathbf{x}_2$ έχουμε
$\nabla_2\psi = \nabla_2\boldsymbol{\varphi}\cdot\nabla_1 f + \nabla_2 f = \mathbf{g}_2 - \mathbf{A}_2\,\mathbf{A}_1^{-1}\mathbf{g}_1$.

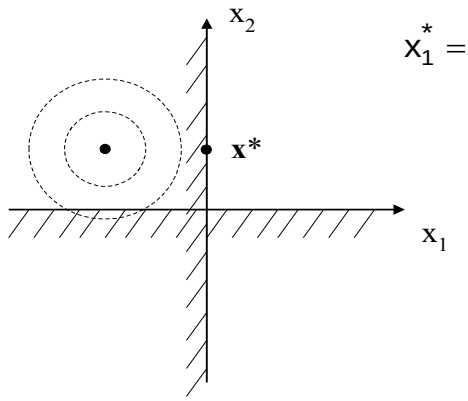
5.5 Να επιλυθεί το πρόβλημα ελαχιστοποίησης $(x_1+2)^2 + (x_2-1)^2$ λαμβάνοντας υπόψη $x_1 \geq 0$, $x_2\geq 0$

(a) γραφικά

(b) μέσω του μετασχηματισμού $y_i^2 = x_i$, $i=1,2$.

Λύση

a)



$x_1^* = 0,\ x_2^* = 1$

b) $\varphi(\mathbf{y}) = (y_1^2+2)^2 + (y_2^2-1)^2$

$$\frac{\partial\varphi}{\partial y_1} = 4(y_1^2+2)y_1 = 0 \Rightarrow y_1 = 0$$

$$\frac{\partial\varphi}{\partial y_2} = 4(y_2^2-1)y_2 = 0 \Rightarrow y_2 = 0 \text{ ή } 1 \text{ ή } -1$$

$$\nabla_y^2\varphi = \begin{bmatrix} 12y_1^2+8 & 0 \\ 0 & 12y_2^2-4 \end{bmatrix} > 0 \text{ μόνο για } (0, \pm 1)$$

$$\Rightarrow \mathbf{y}^* = \begin{bmatrix} 0 \\ \pm 1 \end{bmatrix} \Rightarrow \mathbf{x}^* = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

6. ΘΕΩΡΙΑ ΒΕΛΤΙΣΤΟΠΟΙΗΣΗΣ ΜΕ ΠΕΡΙΟΡΙΣΜΟΥΣ

6.1 Περιορισμοί Ισότητας

Θεωρούμε καταρχήν προβλήματα βελτιστοποίησης με περιορισμούς ισότητας μόνον. Υποθέτοντας ένα ελάχιστο $\mathbf{x}^*$ και μια επιτρεπτή απόκλιση $\boldsymbol{\delta}$, έχουμε μέσω της σειράς Taylor

$$c_i(\mathbf{x}^*+\boldsymbol{\delta}) = c_i^* + \boldsymbol{\delta}^T\mathbf{a}_i^* + o(\|\boldsymbol{\delta}\|)$$

όπου $\mathbf{a}_i = \nabla c_i$, ενώ $o(\|\boldsymbol{\delta}\|)$ αφορά όρους που μπορούν να αγνοηθούν για αρκούντως μικρό $\|\boldsymbol{\delta}\|$. Αφού η απόκλιση είναι επιτρεπτή, έχουμε $c_i(\mathbf{x}^*+\boldsymbol{\delta}) = c_i^* = 0$, άρα κάθε επιτρεπτή απειροστική απόκλιση βρίσκεται σε μια **επιτρεπτή κατεύθυνση s** που ικανοποιεί

$$\mathbf{s}^T\mathbf{a}_i^* = 0 \qquad \forall i\in E. \tag{6.1.1}$$

Σε μια **κανονική** περίπτωση, δηλαδή αν τα διανύσματα $\mathbf{a}_i^*$, $i\in E$, είναι ανεξάρτητα, κάθε κατεύθυνση **s** που ικανοποιεί την (6.1.1) είναι επιτρεπτή και άρα είναι δυνατόν να προσδιορισθεί μια επιτρεπτή απόκλιση $\boldsymbol{\delta}$ για κάθε τέτοια **s**. Αν, επιπλέον, η $f(\mathbf{x})$ έχει αρνητική κλίση κατά μήκος της **s**, δηλαδή

$$\mathbf{s}^T\mathbf{g}^* < 0 \tag{6.1.2}$$

τότε η **s** είναι κατεύθυνση κατάβασης και η επιτρεπτή απόκλιση κατά μήκος της **s** μειώνει την $f(\mathbf{x})$. Αυτό όμως δεν μπορεί να συμβεί εφ' όσον το $\mathbf{x}^*$ είναι τοπικό ελάχιστο. Άρα για τοπικά ελάχιστα δεν μπορεί να υπάρχει **s** που να εκπληροί την (6.1.1) **και** την (6.1.2) συγχρόνως.

Αυτό ισχύει όντως (ικανή συνθήκη), αν το $\mathbf{g}^*$ είναι γραμμικός συνδυασμός των διανυσμάτων $\mathbf{a}_i^*$, $i\in E$, δηλαδή **αν**

$$\mathbf{g}^* = \sum_{i\in E}\mathbf{a}_i^*\lambda_i^* = \mathbf{A}^*\boldsymbol{\lambda}^* \tag{6.1.3}$$

όπως γίνεται προφανές αντικαθιστώντας την (6.1.3) στην (6.1.2). Μπορούμε όμως να αποδείξουμε ότι η (6.1.3) είναι όχι μόνο ικανή, αλλά και αναγκαία συνθήκη για να μην ισχύουν συγχρόνως οι (6.1.1), (6.1.2). Η απόδειξη αναγκαιότητας της (6.1.3) γίνεται διά της εις άτοπον απαγωγής. Αν η (6.1.3) δεν ισχύει, τότε μπορούμε να εκφράσουμε το $\mathbf{g}^*$ ως εξής

$$\mathbf{g}^* = \mathbf{A}^*\boldsymbol{\lambda} + \boldsymbol{\mu} \tag{6.1.4}$$

όπου $\boldsymbol{\mu}\neq\mathbf{0}$ είναι η συνιστώσα του **g** που είναι ορθογώνια στα διανύσματα $\mathbf{a}_i^*$, επομένως έχουμε $\mathbf{A}^{*T}\boldsymbol{\mu} = \mathbf{0}$. Τότε όμως η κατεύθυνση $\mathbf{s} = -\boldsymbol{\mu}$ ικανοποιεί τόσο την (6.1.1) όσο και την (6.1.2), άρα υπάρχει επιτρεπτή απόκλιση $\boldsymbol{\delta}$ κατά μήκος της **s** που μειώνει την $f(\mathbf{x})$, πράγμα που αντιφάσκει με την παραδοχή τοπικού ελάχιστου στο $\mathbf{x}^*$. Επομένως η (6.1.3) είναι ικανή και αναγκαία συνθήκη για να μην ισχύουν συγχρόνως οι (6.1.1) και (6.1.2).

Εν κατακλείδι λοιπόν έχουμε:

$$\mathbf{x}^* \text{ τοπικό ελάχιστο} \Rightarrow \overline{(6.1.1)} \vee \overline{(6.1.2)} \ \forall \mathbf{s} \Leftrightarrow (6.1.3)$$

και επομένως η (6.1.3) είναι **αναγκαία συνθήκη για τοπικό ελάχιστο**. Οι πολλαπλασιαστές λ_i^* αυτού του γραμμικού συνδυασμού φέρουν το όνομα **πολλαπλασιαστές Lagrange** και ο δείκτης * υποδηλώνει ότι αναφέρονται στο σημείο $\mathbf{x}^*$. Αν ο ορθογώνιος πίνακας $\mathbf{A}^*$ είναι πλήρους βαθμού, τότε το $\boldsymbol{\lambda}^*$ έχει, μέσω της (6.1.3), μοναδική τιμή

$$\boldsymbol{\lambda}^* = \mathbf{A}^{*+}\mathbf{g}^*$$

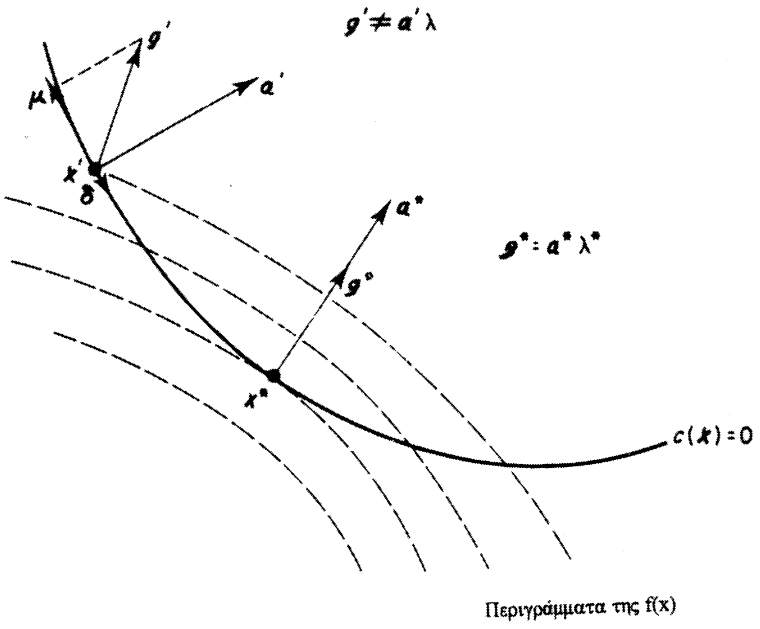
όπου $\mathbf{A}^+ = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T$ είναι ο γενικευμένος αντίστροφος του $\mathbf{A}$. Βέβαια, αν δεν υπάρχουν περιορισμοί ισότητας, η (6.1.3) ανάγεται στην γνωστή συνθήκη στάσιμων σημείων $\mathbf{g}^*=\mathbf{0}$.

Το Σχ. 6.1.1 δείχνει παραστατικά τις περιγραφείσες συνθήκες για n=2. Στο $\mathbf{x}'$ που δεν είναι τοπικό ελάχιστο, έχουμε $\mathbf{g}' \neq \mathbf{a}'\lambda$ και έτσι υπάρχει μη μηδενικό διάνυσμα $\boldsymbol{\mu}$ ορθογώνιο στο $\mathbf{a}'$ και η απόκλιση $\boldsymbol{\delta}$ κατά μήκος της επιτρεπτής κατεύθυνσης κατάβασης $-\boldsymbol{\mu}$ μειώνει την $f(\mathbf{x})$. Αντιθέτως στο $\mathbf{x}^*$ έχουμε $\mathbf{g}^* = \mathbf{a}^*\lambda$ και δεν υπάρχει επιτρεπτή κατεύθυνση κατάβασης.

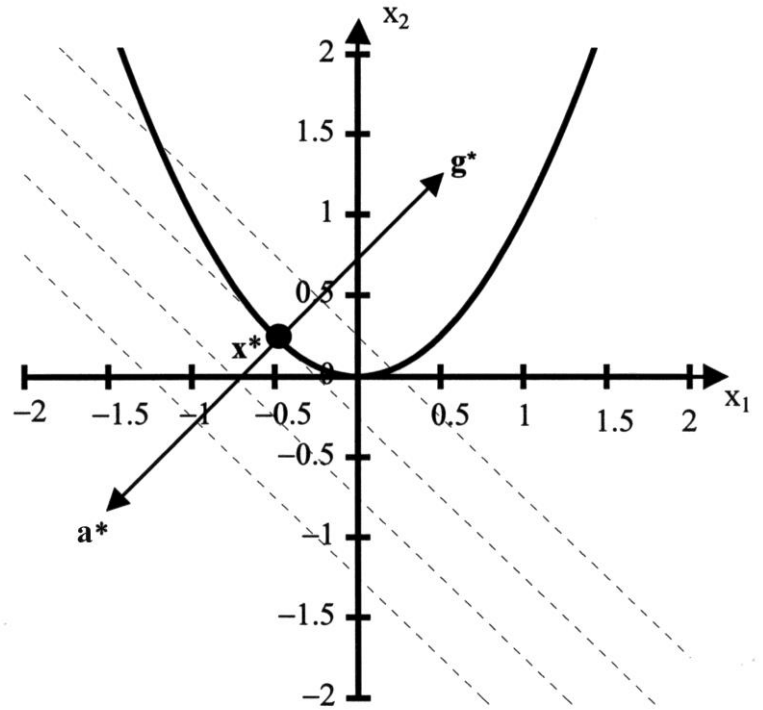
Παράδειγμα 6.1.1: Θεωρούμε την ελαχιστοποίηση της $f(\mathbf{x}) = x_1+x_2$ υπό τον περιορισμό $c(\mathbf{x}) = x_1^2 - x_2 = 0$. Το Σχ. 6.1.2 δίδει γραφικά τη λύση $\mathbf{x}^* = [-0.5\ 0.25]^T$ εξού $\mathbf{g}^* = [1\ 1]^T$ και $\mathbf{a}^* = [-1\ -1]^T$, δηλαδή η (6.1.3) ικανοποιείται με $\lambda^*=-1$. Παρατηρούμε επίσης ότι η παραδοχή κανονικότητας ισχύει, αφού το $\mathbf{a}^*$ είναι ανεξάρτητο, αφού δηλαδή $\mathbf{a}^*\neq\mathbf{0}$.

■

Οι παραπάνω συνθήκες μπορούν να εκφρασθούν μέσω της κλασσικής **μεθόδου πολλαπλασιαστών Lagrange** για την επίλυση προβλημάτων ελαχιστοποίησης με περιορισμούς ισότητας. Η μέθοδος συνίσταται στον προσδιορισμό των διανυσμάτων $\mathbf{x}^*$ και $\boldsymbol{\lambda}^*$ που επιλύουν τις εξισώσεις



Σχήμα 6.1.1: Ο ρόλος των πολλαπλασιαστών Lagrange.



Σχήμα 6.1.2: Γραφική επίλυση του Παραδ. 6.1.1.

$$\mathbf{g}(\mathbf{x}) = \sum_{i\in E} \mathbf{a}_i(\mathbf{x})\lambda_i$$

$$c_i(\mathbf{x}) = 0 \quad i\in E \tag{6.1.5}$$

που απορρέουν από την (6.1.3) και τους περιορισμούς ισότητας. Αν έχουμε σε κάποιο πρόβλημα m περιορισμούς ισότητας, οι (6.1.5) δίδουν n+m εξισώσεις για n+m αγνώστους **x** και **λ** και έτσι μπορούν να προσδιορισθούν οι υποψήφιες λύσεις. Βέβαια, επειδή οι εξισώσεις είναι μη γραμμικές ως προς **x** και **λ**, η αναλυτική επίλυση μπορεί να είναι δύσκολη ή αδύνατη. Επίσης, οι (6.1.5) δεν εμπεριέχουν πληροφορία δευτέρου βαθμού και άρα η λύση τους μπορεί να αντιστοιχεί σε περιορισμένο μέγιστο ή περιορισμένο σαγματικό σημείο.

Είναι σύνηθες, για λόγους απλούστερης χρήσης, να παρουσιάζονται οι αναγκαίες συνθήκες βελτιστοποίησης μέσω της **συνάρτησης Lagrange** (ή **Λαγκρανζιανής συνάρτησης**)

$$L(\mathbf{x},\boldsymbol{\lambda}) = f(\mathbf{x}) - \sum_i \lambda_i c_i(\mathbf{x}) \tag{6.1.6}$$

οπότε οι (6.1.5) είναι ισοδύναμες με

$$\blacktriangledown L(\mathbf{x}^*, \boldsymbol{\lambda}^*) = \mathbf{0} \tag{6.1.7}$$

όπου

$$\blacktriangledown = \begin{bmatrix} \nabla_x \\ \nabla_\lambda \end{bmatrix} \tag{6.1.8}$$

είναι ο τελεστής πρώτης παραγώγου στο χώρο των n+m μεταβλητών. Άρα η αναγκαία συνθήκη τοπικού ελάχιστου είναι ότι τα $\mathbf{x}^*$, $\boldsymbol{\lambda}^*$ προσδιορίζουν ένα στάσιμο σημείο της Λαγκρανζιανής συνάρτησης.

Παράδειγμα 6.1.2: Θεωρούμε το πρόβλημα του Παραδ. 6.1.1 και έχουμε με την συνάρτηση Lagrange (6.1.6)

$$L(\mathbf{x},\lambda) = x_1+x_2-\lambda(x_1^2-x_2).$$

Από την (6.1.7) λαμβάνουμε

$$\frac{\partial L}{\partial x_1} = 1 - 2x_1\lambda = 0$$

$$\frac{\partial L}{\partial x_2} = 1 + \lambda = 0$$

$$\frac{\partial L}{\partial \lambda} = x_1^2 - x_2 = 0$$

που μπορούν να επιλυθούν ως προς x_1^*, x_2^*, λ^* δίδοντας την γνωστή λύση $\lambda^* = -1$, $x_1^* = -0.5$, $x_2^* = 0.25$, $f^* = -0.25$.

■

Για την βαθύτερη κατανόηση του ρόλου των πολλαπλασιαστών Lagrange, θα εξετάσουμε τί συμβαίνει αν μεταβάλουμε τους περιορισμούς ισότητας κατά ε_i

$$c_i(\mathbf{x}) = \varepsilon_i \ , \quad i \in E. \tag{6.1.9}$$

Σημειώνουμε $\mathbf{x}(\boldsymbol{\varepsilon})$, $\boldsymbol{\lambda}(\boldsymbol{\varepsilon})$ τη λύση του προβλήματος σε συνάρτηση του $\boldsymbol{\varepsilon}$. Η Λαγκρανζιανή του παραλλαχθέντος προβλήματος είναι

$$L(\mathbf{x},\boldsymbol{\lambda},\boldsymbol{\varepsilon}) = f(\mathbf{x}) - \sum_{i\in E} \lambda_i[c_i(\mathbf{x}) - \varepsilon_i].$$

Από την παραπάνω εξίσωση, την $f(\mathbf{x}(\boldsymbol{\varepsilon})) = L(\mathbf{x}(\boldsymbol{\varepsilon}),\boldsymbol{\lambda}(\boldsymbol{\varepsilon}),\boldsymbol{\varepsilon})$ (γιατί;), την (Π1.8) και τις $\nabla_x L = \nabla_\lambda L = \mathbf{0}$ (που προκύπτουν από την (6.1.7)) έχουμε στο σημείο της λύσης $\mathbf{x}(\boldsymbol{\varepsilon})$, $\boldsymbol{\lambda}(\boldsymbol{\varepsilon})$, $\boldsymbol{\varepsilon}$

$$\frac{df(\varepsilon_i)}{d\varepsilon_i} = \frac{dL}{d\varepsilon_i} = \frac{\partial x^T}{\partial \varepsilon_i}\nabla_x L + \frac{\partial \lambda^T}{\partial \varepsilon_i}\nabla_\lambda L + \frac{\partial L}{\partial \varepsilon_i} = \frac{\partial L}{\partial \varepsilon_i} = \lambda_i(\varepsilon_i) \tag{6.1.10}$$

Άρα ο πολλαπλασιαστής Lagrange κάθε περιορισμού μετρά τον βαθμό μεταβολής της ελάχιστης τιμής της αντικειμενικής συνάρτησης που απορρέει από ενδεχόμενη μεταβολή του περιορισμού αυτού. Αυτή η πληροφορία είναι συχνά χρήσιμη καθότι δείχνει την ευαισθησία της ελάχιστης τιμής της αντικειμενικής συνάρτησης σε μεταβολές διάφορων περιορισμών.

Παράδειγμα 6.1.3: Θεωρούμε το πρόβλημα των Παραδ. 6.1.1, 6.1.2, με περιορισμό ισότητας $x_1^2 - x_2 = \varepsilon$ αντί του $x_1^2 - x_2 = 0$. Με την νέα Λαγκρανζιανή $L(\mathbf{x},\lambda) = x_1 + x_2 - \lambda(x_1^2 - x_2 - \varepsilon)$ έχουμε τις αναγκαίες συνθήκες βελτιστοποίησης

$$\frac{\partial L}{\partial x_1} = 1 - 2x_1\lambda = 0$$

$$\frac{\partial L}{\partial x_2} = 1 + \lambda = 0$$

$$\frac{\partial L}{\partial \lambda} = x_1^2 - x_2 - \varepsilon = 0$$

εξών η λύση $\lambda^* = -1$, $x_1^* = -0.5$, $x_2^* = 0.25-\varepsilon$, $f^* = -0.25-\varepsilon$. Προφανώς έχουμε $df^*/d\varepsilon = -1$, που επαληθεύει την (6.1.10).

■

6.2 Περιορισμοί Ισότητας και Ανισότητας

Προχωρούμε τώρα σε προβλήματα που περιέχουν, επιπλέον των περιορισμών ισότητας, και περιορισμούς ανισότητας. Θυμίζουμε ότι μόνον **ενεργοί περιορισμοί** $\mathcal{A}^*$ στο $\mathbf{x}^*$ παίζουν κάποιο ρόλο. Σημειώνουμε το σύνολο των ενεργών περιορισμών ανισότητας στο $\mathbf{x}^*$ με I^* ($=\mathcal{A}^* \cap I$). Επειδή $c_i^* = 0$ και $c_i(\mathbf{x}^* + \boldsymbol{\delta}) \geq 0$ για $i \in I^*$, κάθε **επιτρεπτή απόκλιση δ** βρίσκεται σε μια επιτρεπτή κατεύθυνση **s** που ικανοποιεί

$$\mathbf{s}^T \mathbf{a}_i^* \geq 0 \quad \forall i \in I^* \tag{6.2.1}$$

επιπλέον της (6.1.1). Όπως στο Κεφάλαιο 6.1, είναι και εδώ φανερό (δι' αντικαταστάσεως της (6.2.2)) ότι δεν υπάρχει κατεύθυνση **s** που να ικανοποιεί συγχρόνως τις (6.1.1), (6.2.1) και (6.1.2), **αν** (ικανή συνθήκη) ισχύουν συγχρόνως

$$\mathbf{g}^* = \sum_{i \in \mathcal{A}^*} \mathbf{a}_i^* \lambda_i^* \tag{6.2.2}$$

και

$$\lambda_i^* \geq 0 \quad i \in I^*. \tag{6.2.3}$$

Μπορούμε όμως να αποδείξουμε ότι οι (6.2.2), (6.2.3) είναι όχι μόνο ικανές αλλά και αναγκαίες συνθήκες για να μην ισχύουν συγχρόνως οι (6.1.1) και (6.2.1) από τη μια πλευρά και η (6.1.2) από την άλλη. Η αναγκαιότητα της (6.2.2), καταρχήν, αποδεικνύεται ακριβώς όπως και της (6.1.3). Η απόδειξη αναγκαιότητας της (6.2.3) γίνεται και πάλι δια της εις άτοπον απαγωγής υπό την παραδοχή κανονικότητας, δηλαδή γραμμικής ανεξαρτησίας των κάθετων διανυσμάτων $\mathbf{a}_i^*$, $i \in \mathcal{A}^*$. Υποθέτουμε λοιπόν ότι η (6.2.3) δεν ισχύει, ότι υπάρχει δηλαδή κάποιος $\lambda_q^* < 0$, $q \in I^*$. Τότε όμως είναι δυνατό να βρεθεί μια κατεύθυνση **s** για την οποία $\mathbf{s}^T \mathbf{a}_i^* = 0$, $i \in \mathcal{A}^* - \{q\}$ και $\mathbf{s}^T \mathbf{a}_q^* = 1$. Μια τέτοια κατεύθυνση είναι $\mathbf{s} = \mathbf{A}^{*+T} \mathbf{e}_q$, για την οποία συνεπάγεται όντως $\mathbf{A}^{*T}\mathbf{s} = \mathbf{e}_q$, και άρα η **s** είναι επιτρεπτή σύμφωνα με τις (6.1.1) και (6.2.1), ενώ από την (6.2.2) συνεπάγεται

$$\mathbf{s}^T\mathbf{g}^* = \mathbf{s}^T\mathbf{a}_q^*\lambda_q = \lambda_q < 0. \qquad (6.2.4)$$

Άρα η **s** είναι καθοδική, και λόγω κανονικότητας υπάρχει επιτρεπτή απόκλιση **δ** επί της **s** που μειώνει την f(**x**), πράγμα που αντιφάσκει με την παραδοχή τοπικού ελάχιστου στο $\mathbf{x}^*$. Αποδεικνύεται έτσι ότι οι (6.2.2), (6.2.3) είναι ικανές και αναγκαίες για να μην ισχύουν συγχρόνως οι [(6.1.1), (6.2.1)] και (6.1.2). Σημειώνουμε ότι αυτή η απόδειξη χρησιμοποιεί την γραμμική ανεξαρτησία των $\mathbf{a}_i^*$, $i\in\mathcal{A}^*$, κατά τη χρήση του γενικευμένου αντίστροφου $\mathbf{A}^{*+}$. Υπάρχουν γενικότερες αποδείξεις που απαιτούν ηπιότερες παραδοχές, όπως θα δούμε παρακάτω.

Εν κατακλείδι λοιπόν έχουμε

$$\mathbf{x}^* \text{ τοπικό ελάχιστο} \Rightarrow [\overline{(6.1.1)\wedge(6.2.1)}]\vee\overline{(6.1.2)}\ \ \forall\mathbf{s} \Leftrightarrow (6.2.2)\wedge(6.2.3)$$

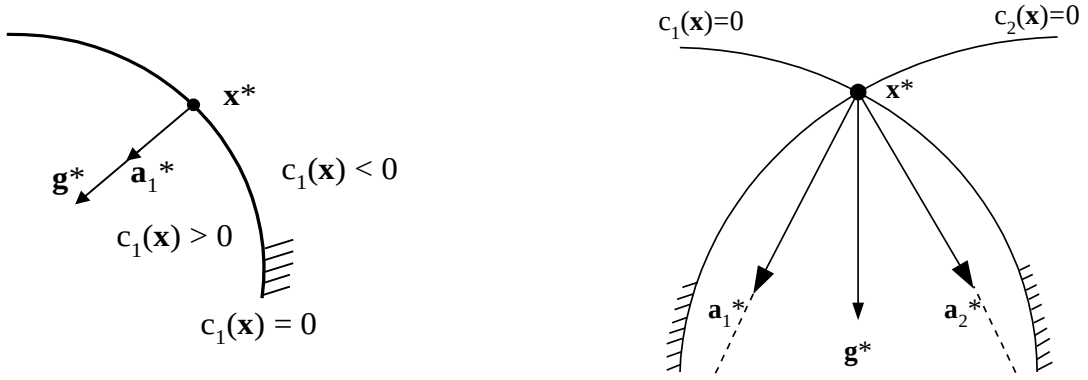
και επομένως οι (6.2.2), (6.2.3) είναι **αναγκαίες συνθήκες για τοπικό ελάχιστο**. Η παρουσία (ενεργών) περιορισμών ανισότητας οδηγεί λοιπόν στην επί πλέον συνθήκη μη αρνητικότητας των αντίστοιχων πολλαπλασιαστών Lagrange.

Ας σημειωθεί ότι η αναγκαιότητα της (6.2.3) μπορεί να εξαχθεί απλούστερα και από τις (6.1.9) και (6.1.10). Μεταβάλλοντας έναν ενεργό περιορισμό ανισότητας από $c_i(\mathbf{x}) \geq 0$ σε $c_i(\mathbf{x}) \geq \varepsilon_i > 0$, $i\in I^*$, περιορίζουμε το πρόβλημα περισσότερο. Επομένως, η λύση $\mathbf{x}(\varepsilon)$ του μεταβληθέντος προβλήματος δεν μπορεί να μειώνει την ελάχιστη τιμή της $f(\mathbf{x}(\varepsilon))$, πρέπει δηλαδή να ισχύει $df^*/d\varepsilon_i \geq 0$ και άρα $\lambda_i^* \geq 0$. Έτσι, η αναγκαιότητα της (6.2.3) έχει μια προφανή ερμηνεία.

Για την γραφική επεξήγηση της (6.2.3) στην περίπτωση n=2, E={}, $\mathcal{A}^*=\{1\}$, παρατηρούμε το Σχ. 6.2.1a). Προφανώς η $\lambda_1^*\geq 0$ δεν απαιτεί μόνον να είναι τα διανύσματα $\mathbf{g}^*$ και $\mathbf{a}_1^*$ παράλληλα (όπως στην περίπτωση περιορισμών ισότητας), αλλά και να έχουν την ίδια κατεύθυνση. Αν αυτό δεν ισχύει, τότε η αντικειμενική συνάρτηση μπορεί προφανώς να μειωθεί σε επιτρεπτή κατεύθυνση ($c_1(\mathbf{x})>0$), οπότε το $\mathbf{x}^*$ δεν μπορεί να είναι τοπικό ελάχιστο. Εξάλλου, το Σχ. 6.2.1b) επεξηγεί την περίπτωση n=2, E={}, $\mathcal{A}^*=\{1,2\}$. Και εδώ είναι εύκολο να παρατηρηθεί ότι η τομή των περιορισμών $c_1(\mathbf{x})=0$ και $c_2(\mathbf{x})=0$ δεν μπορεί να είναι τοπικό ελάχιστο παρά μόνον αν το $\mathbf{g}^*$ βρίσκεται μεταξύ των $\mathbf{a}_1^*$ και $\mathbf{a}_2^*$, δηλαδή μόνο αν $\lambda_1^* \geq 0$ και $\lambda_2^* \geq 0$, άλλως μπορεί η αντικειμενική συνάρτηση να μειωθεί σε επιτρεπτή κατεύθυνση.

Παράδειγμα 6.2.1: Θεωρούμε ελαχιστοποίηση της $f(\mathbf{x})=-x_1-x_2$ λαμβάνοντας υπόψη $c_1(\mathbf{x})=x_2-x_1^2\geq 0$ και $c_2(\mathbf{x})=1-x_1^2-x_2^2\geq 0$ (Σχ. 5.1.3). Η λύση είναι $\mathbf{x}^*= [1/\sqrt{2}\quad 1/\sqrt{2}]^T$, άρα ο c_1 δεν είναι ενεργός και έχουμε $\mathcal{A}^*=\{2\}$. Έχουμε επίσης $\mathbf{g}^*=[-1\ \ -1]^T$ και $\mathbf{a}_2^* = [-\sqrt{2}\quad -\sqrt{2}]^T$ που ικανοποιούν τις (6.2.2), (6.2.3) με $\lambda_2^* = 1/\sqrt{2} > 0$. ■

Το γεγονός ότι μπορεί να προσδιορισθεί επιτρεπτή κατεύθυνση κατάβασης αν κάποιος $\lambda_q^* < 0$, δείχνει μια ακόμη ιδιότητα των πολλαπλασιαστών Lagrange σε σχέση με τους περιορισμούς ανισότητας. Οι συνθήκες $\mathbf{s}^T\mathbf{a}_i^*=0$, $i\neq q$, και $\mathbf{s}^T\mathbf{a}_q^*=1$ σημαίνουν ότι η αντίστοιχη επιτρεπτή απόκλιση κατάβασης ικανοποιεί $c_i(\mathbf{x}+\boldsymbol{\delta})=0$ για $i\neq q$ και $c_q(\mathbf{x}^*+\boldsymbol{\delta}) > 0$, ότι, δηλαδή, η f(**x**) μπορεί να μειωθεί σε επιτρεπτή κατεύθυνση **αν εγκαταληφθεί (απενεργοποιηθεί) το σύνορο του περιορισμού** q, πράγμα που υποδηλώνει εξάλλου και η (6.1.10). Το γεγονός αυτό έχει



a) b)

Σχήμα 6.2.1: Επεξήγηση των αναγκαίων συνθηκών τοπικού ελάχιστου στην περίπτωση n=2.

καθοριστική σημασία για τις μεθόδους ενεργού συνόλου που θα εξετάσουμε αργότερα. Έτσι, αν οι συνθήκες (6.2.3) δεν ικανοποιούνται σε κάποιο επιτρεπτό σημείο **x**', τότε κάποιος περιορισμός q με $\lambda_q'<0$ μπορεί να απομακρυνθεί από το αντίστοιχο ενεργό σύνολο.

Παράδειγμα 6.2.2: Θεωρούμε στο πρόβλημα του Παραδ. 6.2.1 το επιτρεπτό σημείο $\mathbf{x}'=[0.786\ 0.618]^T$ όπου είναι ενεργοί και οι δύο περιορισμοί (βλ. Σχ. 5.1.3). Καθότι $\mathbf{g}'=[-1\ -1]^T$, $\mathbf{a}_1' = [-1.572\ 1]^T$, $\mathbf{a}_2' = [-1.572\ -1.236]^T$, ικανοποιείται η (6.2.2) με $\boldsymbol{\lambda}' = [-0.096\ 0.732]^T$, όχι όμως και η (6.2.3), και άρα το **x**' δεν είναι τοπικό ελάχιστο. Επειδή $\lambda_1'<0$, η αντικειμενική συνάρτηση μπορεί να μειωθεί με απομάκρυνση από το σύνορο του περιορισμού 1, π.χ. κατά μήκος της κατεύθυνσης που ικανοποιεί $\mathbf{s}^T\mathbf{a}_1'=1$ και $\mathbf{s}^T\mathbf{a}_2'=0$, δηλαδή $\mathbf{s}=[-0.352\ 0.447]^T$ που είναι εφαπτόμενη στον κύκλο στο σημείο **x**'. Κινούμενοι λοιπόν πάνω στο τόξο του κύκλου οδηγούμεθα στο γνωστό τοπικό ελάχιστο $\mathbf{x}^*$ όπου είναι ενεργός μόνον ο περιορισμός 2.

■

Οι αναγκαίες συνθήκες (6.2.2), (6.2.3) μπορούν και να διατυπωθούν σε συνάρτηση όλων των περιορισμών (όχι μόνο των ενεργών), αν θεωρήσουμε απλώς ότι οι πολλαπλασιαστές Lagrange των μη ενεργών περιορισμών είναι μηδέν, αφού η ευαισθησία της βέλτιστης τιμής της αντικειμενικής συνάρτησης ως προς τους μη ενεργούς περιορισμούς είναι όντως μηδενική. Στην εκδοχή αυτή, οι αναγκαίες συνθήκες (6.2.2), (6.2.3) και η αναγκαιότητα επιτρεπτής λύσης μπορούν να συνδυαστούν κατά το ακόλουθο θεώρημα, όπου $L(\mathbf{x},\boldsymbol{\lambda})$ είναι η Λαγκρανζιανή συνάρτηση (6.1.6) που περιλαμβάνει τώρα τόσο τους περιορισμούς ισότητας, όσο και όλους τους περιορισμούς ανισότητας.

Θεώρημα 6.2.1 (Αναγκαίες συνθήκες πρώτου βαθμού): Αν το $\mathbf{x}^*$ είναι τοπικό ελάχιστο του προβλήματος (5.1.1) και αν ικανοποιείται μια συνθήκη καταλληλότητας (βλ. παρακάτω), τότε υπάρχουν πολλαπλασιαστές Lagrange $\boldsymbol{\lambda}^*$ που εκπληρούν τις ακόλουθες συνθήκες

$$\begin{aligned} \nabla_{\mathbf{x}} L(\mathbf{x}^*, \boldsymbol{\lambda}^*) &= \mathbf{0} \\ c_i(\mathbf{x}^*) &= 0 \quad i\in E \\ c_i(\mathbf{x}^*) &\geq 0 \quad i\in I \\ \lambda_i^* &\geq 0 \quad i\in I \\ \lambda_i^*\, c_i(\mathbf{x}^*) &= 0 \quad \forall i. \end{aligned} \tag{6.2.5}$$

■

Οι συνθήκες αυτές ονομάζονται **συνθήκες Kuhn-Tucker**, αφού οι Kuhn και Tucker τις δημοσίευσαν πρώτοι το 1951, και το $\mathbf{x}^*$ συχνά αναφέρεται ως **σημείο KT**.[1]

Η συνθήκη $\lambda_i^* c_i^* = 0$ φέρει το όνομα **συνθήκη συμπληρωματικότητας** και απαγορεύει τα λ_i^* και $c_i(\mathbf{x}^*)$ να είναι συγχρόνως διάφορα του μηδενός. Για τους ενεργούς περιορισμούς ανισότητας έχουμε $c_i(\mathbf{x}^*)=0$ και άρα ισχύει προφανώς η συνθήκη. Για τους μη ενεργούς περιορισμούς θεωρούμε $\lambda_i^*=0$, όπως σημειώθηκε και παραπάνω, και άρα ισχύει επίσης η συνθήκη συμπληρωματικότητας. Αν δεν υπάρχει i για το οποίο $\lambda_i^* = c_i^* = 0$, τότε έχουμε **αυστηρή συμπληρωματικότητα**. Η περίπτωση $\lambda_i^* = c_i^* = 0$ είναι μια οριακή κατάσταση μεταξύ ισχυρής ενεργοποίησης και μη ενεργοποίησης ενός περιορισμού, όπως επεξηγείται παραστατικά στο Σχ. 6.2.2.

Η συνθήκη καταλληλότητας, που αναφέρεται στο Θεώρημα 6.2.1, ικανοποιείται σίγουρα, αν το $\mathbf{x}^*$ είναι κανονικό σημείο, δηλαδή αν τα $\mathbf{a}_i^*$ είναι ανεξάρτητα. Η συνθήκη καταλληλότητας είναι ασθενέστερη από την συνθήκη κανονικότητας, η πλήρης παρουσίασή της θα απαιτούσε όμως αρκετό χώρο. Επειδή τα περισσότερα πρακτικά προβλήματα εκπληρούν την συνθήκη καταλληλότητας, παραιτούμεθα από την λεπτομερή παρουσίασή της και περιοριζόμαστε εδώ στην παράθεση ενός ειδικού παραδείγματος που δεν την εκπληροί.

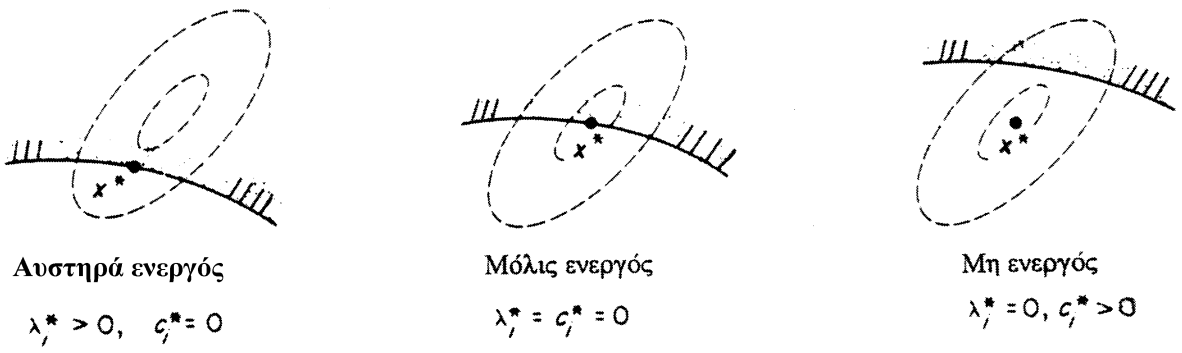
Παράδειγμα 6.2.3: Θεωρούμε ελαχιστοποίηση της $f(\mathbf{x})=x_1$ υπό τους περιορισμούς ανισότητας $c_1(\mathbf{x})=x_1^3-x_2\geq 0$ και $c_2(\mathbf{x})=x_2\geq 0$. Το Σχ. 6.2.3 δείχνει το ελάχιστο $\mathbf{x}^*=\mathbf{0}$, όπου όμως το $\mathbf{g}^*$ δεν είναι γραμμικός συνδυασμός των $\mathbf{a}_1^*$ και $\mathbf{a}_2^*$. Πράγματι δεν ικανοποιείται η συνθήκη καταλληλότητας, αλλά ούτε και η συνθήκη κανονικότητας αφού τα $\mathbf{a}_1^*=[0\ \ -1]^T$ και $\mathbf{a}_2^*=[0\ \ 1]^T$ δεν είναι ανεξάρτητα.

■

Παράδειγμα 6.2.4: Θεωρούμε ελαχιστοποίηση της αντικειμενικής συνάρτησης

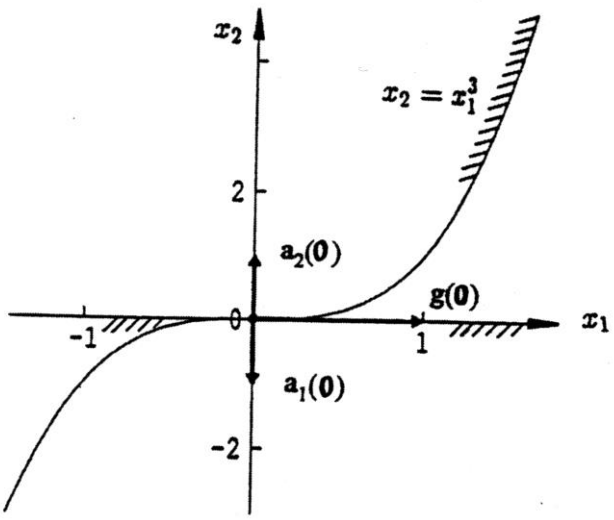
$$f(\mathbf{x}) = x_1^2+x_2^2+x_1x_2+2x_1 \qquad (6.2.6)$$

λαμβάνοντας υπόψη τους περιορισμούς ανισότητας



Σχήμα 6.2.2: Συνθήκες συμπληρωματικότητας.

[1] Ας σημειωθεί ότι, όπως διαπιστώθηκε πολύ αργότερα, το ως άνω Θεώρημα είχε διατυπωθεί ήδη το 1939 στην μεταπτυχιακή εργασία του **W. Karush**, ο οποίος όμως δεν είχε δημοσιεύσει τα αποτελέσματά του. Για το λόγο αυτό οι (6.2.5) αναφέρονται συχνά και ως **συνθήκες Karush-Kuhn-Tucker** (**KKT**).



Σχήμα 6.2.3: Γραφική επεξήγηση για το Παράδ. 6.2.3.

$$c_1(\mathbf{x}) = 1.5 - x_1^2 - x_2^2 \geq 0 \qquad (6.2.7)$$
$$c_2(\mathbf{x}) = -x_1 \geq 0 \qquad (6.2.8)$$
$$c_3(\mathbf{x}) = x_2 \geq 0. \qquad (6.2.9)$$

Η Λαγκρανζιανή είναι

$$L(\mathbf{x},\boldsymbol{\lambda}) = x_1^2 + x_2^2 + x_1x_2 + 2x_1 - \lambda_1(1.5 - x_1^2 - x_2^2) + \lambda_2 x_1 - \lambda_3 x_2.$$

Το Σχ. 6.2.4 δείχνει την επιτρεπτή περιοχή και μερικά περιγράμματα της αντικειμενικής συνάρτησης. Για την αναλυτική επίλυση του προβλήματος πρέπει να εξετάσουμε μια σειρά από ενδεχόμενες περιπτώσεις ενεργοποίησης (ή όχι) ενός συνδυασμού περιορισμών ανισότητας. Για κάθε τέτοια περίπτωση εξετάζεται αν η αντίστοιχη λύση:

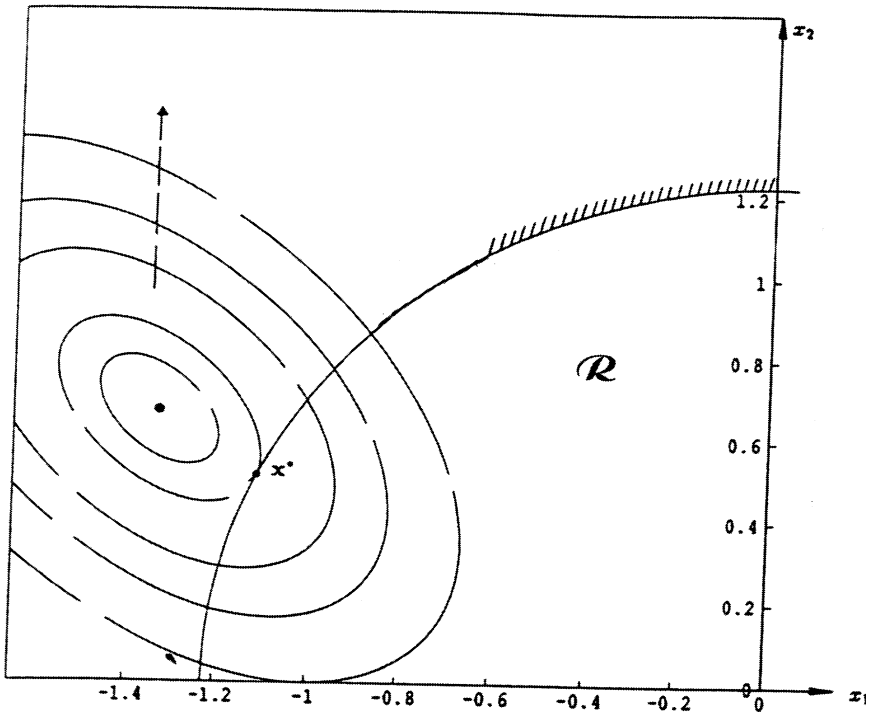
- παραβιάζει κάποιον από τους μη ενεργοποιημένους περιορισμούς ανισότητας
- κάποιος από τους προσδιορισθέντες πολλαπλασιαστές Lagrange παραβιάζει τη συνθήκη $\lambda_i^* \geq 0 \quad i \in I$.

Αν η απάντηση στα δύο αυτά ερωτήματα είναι αρνητική, τότε η αντίστοιχη εξεταζόμενη περίπτωση ικανοποιεί όλες τις συνθήκες KKT και άρα αποδίδει ένα υποψήφιο τοπικό ελάχιστο. Αλλιώς, η εξεταζόμενη περίπτωση αποβαίνει άκαρπη.
Αφού εξεταστούν όλες οι περιπτώσεις, έχουν προσδιορισθεί όλα τα υποψήφια τοπικά ελάχιστα.

Περίπτωση 1: Έστω ότι δεν ενεργοποιείται κανείς περιορισμός στο $\mathbf{x}^*$, άρα $\lambda_1^* = \lambda_2^* = \lambda_3^* = 0$. Οι αναγκαίες συνθήκες δίδουν

$$\partial L/\partial x_1 = 2x_1^* + x_2^* + 2 = 0$$
$$\partial L/\partial x_2 = 2x_2^* + x_1^* = 0$$

εξών $x_1^* = -4/3$, $x_2^* = 2/3$. Το σημείο αυτό βρίσκεται εκτός επιτρεπτής περιοχής αφού παραβιάζει τον περιορισμό c_1.



Σχήμα 6.2.4: Γραφική παράσταση του Παραδ. 6.2.4.

Περίπτωση 2: Έστω ότι ενεργοποιείται ο c_1, όχι όμως οι c_2, c_3, οπότε $\lambda_1^* \geq 0$, $\lambda_2^* = \lambda_3^* = 0$, και από τις αναγκαίες συνθήκες έχουμε

$$\partial L/\partial x_1 = 2x_1^* + x_2^* + 2 + 2\lambda_1^* x_1^* = 0$$
$$\partial L/\partial x_2 = 2x_2^* + x_1^* + 2\lambda_1^* x_2^* = 0$$

εξών

$$x_1^* = \frac{4(1+\lambda_1^*)}{N} \qquad x_2^* = -\frac{2}{N} \tag{6.2.10}$$

όπου $N=1-4(1+\lambda_1^*)^2$. Αφού ο c_1 είναι ενεργός, έχουμε

$$x_1^{*2} + x_2^{*2} = 1.5 \tag{6.2.11}$$

και δι' αντικαταστάσεως της (6.2.10) στην (6.2.11), και με $\alpha = 4(1+\lambda_1^*)^2$, καταλήγουμε στην δευτεροβάθμια εξίσωση $1.5\alpha^2-7\alpha-2.5 = 0$ με την θετική λύση $\alpha=5$, και άρα $\lambda_1^* = -1+\sqrt{1.25}$ και $x_1^* = -\sqrt{5}/2$, $x_2^* = 0.5$. Αφού $\lambda_1^* \geq 0$, το προσδιορισθέν $\mathbf{x}^*$ ικανοποιεί όλες τις αναγκαίες συνθήκες πρώτου βαθμού για τοπικό ελάχιστο.

Περίπτωση 3: Έστω ότι ενεργοποιείται ο c_2, όχι όμως οι c_1, c_3, οπότε $\lambda_2^* \geq 0$, $\lambda_1^* = \lambda_3^* = 0$, και από τις αναγκαίες συνθήκες έχουμε

$$\partial L/\partial x_1 = 2x_1^* + x_2^* + 2 + \lambda_2^* = 0$$
$$\partial L/\partial x_2 = 2x_2^* + x_1^* = 0. \tag{6.2.12}$$

Αφού ο c_2 είναι ενεργός, έχουμε $x_1^*=0$ και άρα $x_2^*=0$ και $\boldsymbol{\lambda}_2^*=-2$. Επειδή $\lambda_2^*<0$, παραβιάζονται οι αναγκαίες συνθήκες και άρα το προσδιορισθέν $\mathbf{x}^*$ δεν είναι τοπικό ελάχιστο.

Κατά παρόμοιο τρόπο μπορεί να αποκλεισθεί η παρουσία τοπικού ελάχιστου και σε κάθε άλλη περίπτωση συνδυασμού ενεργοποίησης περιορισμών και άρα υπάρχει ένα μόνο υποψήφιο τοπικό ελάχιστο, αυτό της Περίπτωσης 2.

■

Το παράδειγμα αυτό δείχνει ότι η αναλυτική αξιοποίηση των αναγκαίων συνθηκών βελτιστοποίησης για τον προσδιορισμό της λύσης του προβλήματος είναι περισσότερο επίπονη και περίπλοκη όταν υπάρχουν περιορισμοί ανισότητας.

6.3 Συνθήκες Δευτέρου Βαθμού

Για την εξαγωγή αναγκαίων συνθηκών δευτέρου βαθμού, που συμπεριλαμβάνουν πληροφορίες καμπυλότητας της αντικειμενικής συνάρτησης και των περιορισμών, θα υποθέσουμε ότι f, $c_i \in C^2$.

Αν το πρόβλημα περιλαμβάνει μόνο περιορισμούς ισότητας, και αν το $\mathbf{x}^*$, που ικανοποιεί τις συνθήκες (6.2.5), είναι κανονικό σημείο, δηλαδή αν τα $\mathbf{a}_i^*$, $i\in E$, είναι ανεξάρτητα, τότε υπάρχει, ως γνωστόν, ένα μοναδικό $\boldsymbol{\lambda}^*=\mathbf{A}^{*+}\,\mathbf{g}^*$, και, επίσης, υπάρχει επιτρεπτή απόκλιση $\boldsymbol{\delta}$ για κάθε επιτρεπτή κατεύθυνση $\mathbf{s}$ στο $\mathbf{x}^*$. Για επιτρεπτές αποκλίσεις ισχύει $L(\mathbf{x}^*+\boldsymbol{\delta}, \boldsymbol{\lambda}) = f(\mathbf{x}^*+\boldsymbol{\delta})$, και με τις (6.2.5) έχουμε για την σειρά Taylor

$$\begin{aligned} f(\mathbf{x}^*+\boldsymbol{\delta}) &= L(\mathbf{x}^*+\boldsymbol{\delta}, \boldsymbol{\lambda}^*) \\ &= L(\mathbf{x}^*, \boldsymbol{\lambda}^*) + \boldsymbol{\delta}^T\nabla_x L(\mathbf{x}^*,\boldsymbol{\lambda}^*)+\frac{1}{2}\boldsymbol{\delta}^T\mathbf{W}^*\boldsymbol{\delta}+o(\|\boldsymbol{\delta}\|^3) \\ &= f^*+\frac{1}{2}\boldsymbol{\delta}^T\mathbf{W}^*\boldsymbol{\delta}+o(\|\boldsymbol{\delta}\|^3) \qquad (6.3.1) \end{aligned}$$

όπου ο $\mathbf{W}^*=\nabla_x^2 L(\mathbf{x}^*,\boldsymbol{\lambda}^*) = \nabla^2 f(\mathbf{x}^*)-\sum_i \lambda_i^* \nabla^2 c_i(\mathbf{x}^*)$ είναι ο Χεσιανός πίνακας της συνάρτησης Lagrange ως προς $\mathbf{x}$. Αν η f^* είναι η ελάχιστη τιμή, πρέπει να έχουμε για $\boldsymbol{\delta}\to\mathbf{0}$

$$\mathbf{s}^T\mathbf{W}^*\mathbf{s} \geq 0 \qquad (6.3.2)$$

για όλες τις επιτρεπτές κατευθύνσεις, δηλαδή για όλες τις κατευθύνσεις που ικανοποιούν την (6.1.1), δηλαδή την

$$\mathbf{A}^{*T}\mathbf{s} = \mathbf{0}. \qquad (6.3.3)$$

Έτσι, η **αναγκαία συνθήκη δευτέρου βαθμού για τοπικό ελάχιστο** είναι η ικανοποίηση της (6.3.2) για κάθε $\mathbf{s}$ που ικανοποιεί την (6.3.3), δηλαδή η Λαγκρανζιανή πρέπει να έχει μη αρνητική καμπυλότητα για κάθε επιτρεπτή κατεύθυνση στο $\mathbf{x}^*$. Προφανώς, αν δεν υπάρχουν περιορισμοί ισότητας, οι (6.3.2), (6.3.3) ανάγονται στην απεριόριστη αναγκαία συνθήκη δευτέρου βαθμού $\mathbf{G}^*\geq \mathbf{0}$.

Υπογραμμίζουμε εδώ ότι, για προβλήματα βελτιστοποίησης με περιορισμούς, ο Χεσιανός πίνακας που πρέπει να ελεγχθεί στις συνθήκες δεύτερου βαθμού είναι ο **W**, δηλαδή ο Χεσιανός πίνακας της Λαγκρανζιανής, και όχι ο **G**. Αυτό οφείλεται στο γεγονός ότι στην προσέγγιση δευτέρου βαθμού πρέπει να ληφθεί υπόψη η καμπυλότητα όχι μόνο της προς ελαχιστοποίηση συνάρτησης, αλλά και των (εν γένει μη γραμμικών) περιορισμών, επί των οποίων βρίσκεται η επιτρεπτή απόκλιση **δ**. Πράγματι, υπάρχουν προβλήματα (βλ. Πρόβλημα 6.9) στα οποία $\mathbf{G}^*>\mathbf{0}$ σε κάποιο σημείο ΚΤ $\mathbf{x}^*$, το οποίο όμως δεν είναι τοπικό ελάχιστο.

Οι **ικανές συνθήκες για αυστηρό τοπικό ελάχιστο** δίδονται από την (6.1.7) και

$$\mathbf{s}^T\mathbf{W}^*\mathbf{s} > 0 \qquad (6.3.4)$$

για κάθε $\mathbf{s}\neq\mathbf{0}$ που εκπληροί την (6.3.3).

Πως μπορεί, για κάποιο $\mathbf{x}^*$, να ελεγχθεί αν ισχύει η (6.3.2) ή η (6.3.4) για κάθε **s** που ικανοποιεί την (6.3.3); Μια τετραγωνική μορφή $\mathbf{s}^T\mathbf{Bs}$, όπου $\mathbf{B}\in R^{n\times n}$ συμμετρικός, ονομάζεται **θετικά ορισμένη υπό M**, όπου $\mathbf{M}\in R^{p\times n}$, p<n, κανονικός πίνακας, αν $\mathbf{s}^T\mathbf{Bs}>0$ για κάθε $\mathbf{s}\neq\mathbf{0}$ που ικανοποιεί $\mathbf{Ms}=\mathbf{0}$. Η ιδιότητα αυτή υφίσταται, αν και μόνον αν οι ρίζες του ακόλουθου πολυωνύμου (n–p) βαθμού p(Λ) είναι πραγματικές και θετικές

$$p(\Lambda) = \begin{vmatrix} \Lambda\mathbf{I} - \mathbf{B} & \mathbf{M}^T \\ \mathbf{M} & \mathbf{0} \end{vmatrix} = 0. \qquad (6.3.5)$$

Αν οι ρίζες είναι πραγματικές και μη αρνητικές ή αρνητικές ή μη θετικές, τότε $\mathbf{s}^T\mathbf{Bs}\geq 0$ ή $\mathbf{s}^T\mathbf{Bs} < 0$ ή $\mathbf{s}^T\mathbf{Bs} \leq 0$, αντίστοιχα, για κάθε $\mathbf{s}\neq\mathbf{0}$ που ικανοποιεί $\mathbf{Ms}=\mathbf{0}$**,** και η τετραγωνική μορφή ονομάζεται, αντίστοιχα, θετικά ημιορισμένη ή αρνητικά ορισμένη ή αρνητικά ημιορισμένη υπό **M**.

Για την συγκεκριμένη εφαρμογή των συνθηκών δευτέρου βαθμού (6.3.2)-(6.3.4) έχουμε το πολυώνυμο (6.3.5)

$$p(\Lambda) = \begin{vmatrix} \Lambda I - \mathbf{W}^* & \mathbf{A}^* \\ \mathbf{A}^{*T} & \mathbf{0} \end{vmatrix} = 0. \qquad (6.3.6)$$

Εννοείται πως, αν $\mathbf{W}^*>\mathbf{0}$ ή $\mathbf{W}^*\geq\mathbf{0}$, δηλαδή αν η (6.3.2) ή (6.3.4), αντίστοιχα, ισχύει **για κάθε s** $\neq\mathbf{0}$, τότε προφανώς ισχύει και για τις ειδικές $\mathbf{s}\neq\mathbf{0}$ που εκπληρούν την (6.3.3) και άρα εκπληρούνται οι αναγκαίες ή ικανές συνθήκες δευτέρου βαθμού, αντίστοιχα, και δεν χρειάζεται να γίνει διερεύνηση ριζών του πολυωνύμου (6.3.6).

Παράδειγμα 6.3.1: Θεωρούμε το πρόβλημα των Παραδ. 6.1.1, 6.1.2 και εξετάζουμε την ισχύ των συνθηκών δευτέρου βαθμού στο προσδιορισθέν $\mathbf{x}^*$. Εχουμε

$$\nabla_x^2 L = \mathbf{W} = \begin{bmatrix} -2\lambda & 0 \\ 0 & 0 \end{bmatrix}, \quad \mathbf{W}^* = \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix}.$$

Επειδή $\mathbf{W}^*\geq\mathbf{0}$, ικανοποιείται σίγουρα η αναγκαία συνθήκη δευτέρου βαθμού, όχι όμως αναγκαστικά και οι ικανές συνθήκες. Με $\mathbf{A}^{*T} = [2\,x_1^*\ -1] = [-1\ -1]$, έχουμε από την (6.3.6)

$$p(\Lambda) = \begin{vmatrix} \Lambda-2 & 0 & -1 \\ 0 & \Lambda & -1 \\ -1 & -1 & 0 \end{vmatrix} = -(\Lambda-2)-\Lambda = -2\Lambda+2 = 0.$$

Η ρίζα αυτού του πολυωνύμου n–p=1 βαθμού είναι Λ=1>0, άρα η ικανή συνθήκη (6.3.4) ικανοποιείται για κάθε $\mathbf{s}\neq\mathbf{0}$ στην (6.3.3), και το $\mathbf{x}^*$ είναι σίγουρα τοπικό ελάχιστο.

■

Προχωρούμε τώρα στην εξαγωγή συνθηκών δευτέρου βαθμού για προβλήματα ελαχιστοποίησης με περιορισμούς ισότητας **και** ανισότητας. Αναπτύσσοντας και πάλι τη σειρά Taylor (6.3.1) και ορίζοντας $\mathbf{W}^*=\nabla_x^2 L(\mathbf{x}^*, \boldsymbol{\lambda}^*)$, όπου τώρα η Λαγκρανζιανή εμπεριέχει βέβαια και τους περιορισμούς ανισότητας, καταλήγουμε και πάλι στην συνθήκη (6.3.2) που πρέπει να ικανοποιείται για κάθε επιτρεπτή κατεύθυνση **s**. Για τους περιορισμούς ισότητας, οι επιτρεπτές **s** ικανοποιούν την (6.1.1), ενώ για τους περιορισμούς ανισότητας την (6.2.1). Εδώ κάνουμε μια διαφοροποίηση ανάμεσα στους αυστηρά ενεργούς περιορισμούς ανισότητας, όπου δηλαδή $\lambda_i^* > 0$, και τους μόλις ενεργούς περιορισμούς ανισότητας με $\lambda_i^* = 0$. Για τους πρώτους είναι σίγουρο, λόγω $\lambda_i^* > 0$, ότι κάθε επιτρεπτή κατεύθυνση προς το εσωτερικό της επιτρεπτής περιοχής αυξάνει την αντικειμενική συνάρτηση και άρα αρκεί να θεωρήσουμε τις κατευθύνσεις **κατά μήκος** του περιορισμού, δηλαδή $\mathbf{s}^T\mathbf{a}_i^* = 0$, όπως ακριβώς και στους περιορισμούς ισότητας. Έτσι, ορίζοντας το **σύνολο των αυστηρά ενεργών περιορισμών**

$$\mathcal{A}_+^* = \{i \mid i\in E \text{ ή } \lambda_i^* > 0\} \qquad (6.3.7)$$

η **αναγκαία συνθήκη δευτέρου βαθμού** έγκειται στην ικανοποίηση της (6.3.2) για κάθε **s** που εκπληροί

$$\mathbf{a}_i^{*T}\mathbf{s} = 0 \quad \forall i\in\mathcal{A}^* \qquad (6.3.8)$$

$$\mathbf{a}_i^{*T}\mathbf{s} \geq 0 \quad \forall i\in\mathcal{A}^*-\mathcal{A}_+^*. \qquad (6.3.9)$$

Οι **ικανές συνθήκες για αυστηρό τοπικό ελάχιστο** δίδονται από τις (6.2.5) και την (6.3.4) για κάθε $\mathbf{s}\neq\mathbf{0}$ που εκπληροί τις (6.3.8), (6.3.9).

Συμπερασματικά, αν όλοι οι ενεργοί περιορισμοί ανισότητας είναι αυστηροί, ο έλεγχος των συνθηκών δευτέρου βαθμού γίνεται ακριβώς όπως και στα προβλήματα με περιορισμούς ισότητας μόνον, λαμβάνοντας όμως υπόψη και τους ενεργούς περιορισμούς ανισότητας. Αν όμως υπάρχουν και μόλις-ενεργοί περιορισμοί ανισότητας, πρέπει ενδεχομένως να ελεγχθεί αναλυτικά κατά περίπτωση αν οι (6.3.2) ή (6.3.4) μπορούν να αποδειχθούν περιορίζοντας τις **s** μέσω των (6.3.8), (6.3.9).

Παράδειγμα 6.3.2: Θεωρούμε το πρόβλημα του Παραδ. 6.2.4 και εξετάζουμε την ισχύ των συνθηκών δευτέρου βαθμού στο προσδιορισθέν $\mathbf{x}^*$. Έχουμε

$$\nabla_x^2 L = \mathbf{W} = \begin{bmatrix} 2+2\lambda_1 & 1 \\ 1 & 2+2\lambda_1 \end{bmatrix}, \ \mathbf{W}^* = \begin{bmatrix} 2\sqrt{1.25} & 1 \\ 1 & 2\sqrt{1.25} \end{bmatrix}$$

όπου ισχύει $\mathbf{W}^*>\mathbf{0}$ για κάθε **s**, χωρίς δηλαδή να χρειασθεί να περιορίσουμε τις **s** μέσω (6.3.8), (6.3.9), και άρα οι ικανές συνθήκες ικανοποιούνται και το $\mathbf{x}^*$ είναι σίγουρα τοπικό ελάχιστο.

■

6.4 Κυρτά Προβλήματα Βελτιστοποίησης

Κυρτότητα είναι μια μαθηματική έννοια με ιδιαίτερη σημασία για την βελτιστοποίηση, αφού οδηγεί στον προσδιορισμό μιας κατηγορίας προβλημάτων, για τα οποία οι προσδιορισθείσες αναγκαίες συνθήκες πρώτου βαθμού είναι **αναγκαίες και ικανές** συνθήκες, και μάλιστα για **απόλυτο ελάχιστο**.

Κυρτό σύνολο Τομή

a) b)

Σχήμα 6.4.1: Κυρτά σύνολα.

Ένα **σύνολο** $K\subset R^n$ ονομάζεται **κυρτό**, αν για κάθε $\mathbf{x}_0$, $\mathbf{x}_1\in K$ συνεπάγεται $\mathbf{x}_\theta \in K$, όπου (βλ. Σχ. 6.4.1a)

$$\mathbf{x}_\theta= (1-\theta)\mathbf{x}_0 + \theta\mathbf{x}_1 \quad \forall\theta\in[0,1]. \qquad (6.4.1)$$

Από τον ορισμό συνεπάγεται ότι για κυρτό K και για κάθε $\mathbf{x}_0, \mathbf{x}_1, \ldots , \mathbf{x}_m\in K$ ισχύει $\mathbf{x}_\theta \in K$ όπου

$$\mathbf{x}_\theta = \sum_{i=0}^{m}\theta_i\mathbf{x}_i \ , \quad \sum_{i=0}^{m}\theta_i = 1 \ , \quad \theta_i\geq 0. \qquad (6.4.2)$$

Το $\mathbf{x}_\theta$ στις (6.4.1) ή (6.4.2) αναφέρεται ως **κυρτός συνδυασμός** των $\mathbf{x}_0, \mathbf{x}_1, \ldots , \mathbf{x}_m$. Αν τα $\mathbf{x}_0, \mathbf{x}_1,\ldots, \mathbf{x}_m$ είναι δεδομένα σημεία, τότε το σύνολο των σημείων $\mathbf{x}_\theta$ που ορίζονται από την (6.4.2) είναι κυρτό και αναφέρεται ως **κυρτό κάλυμμα** του συνόλου των σημείων.

Παραδείγματα κυρτών συνόλων είναι {}, {**x**'}, R^n, μια γραμμή ή ευθύγραμμο τμήμα, ένα υπερεπίπεδο $\mathbf{a}^T\mathbf{x}=b$, ο ήμισυς χώρος $\mathbf{a}^T\mathbf{x}\geq b$, η σφαίρα $\|\mathbf{x}-\mathbf{x}'\|_2 \leq h$, κυρτοί κώνοι και πολλά άλλα.

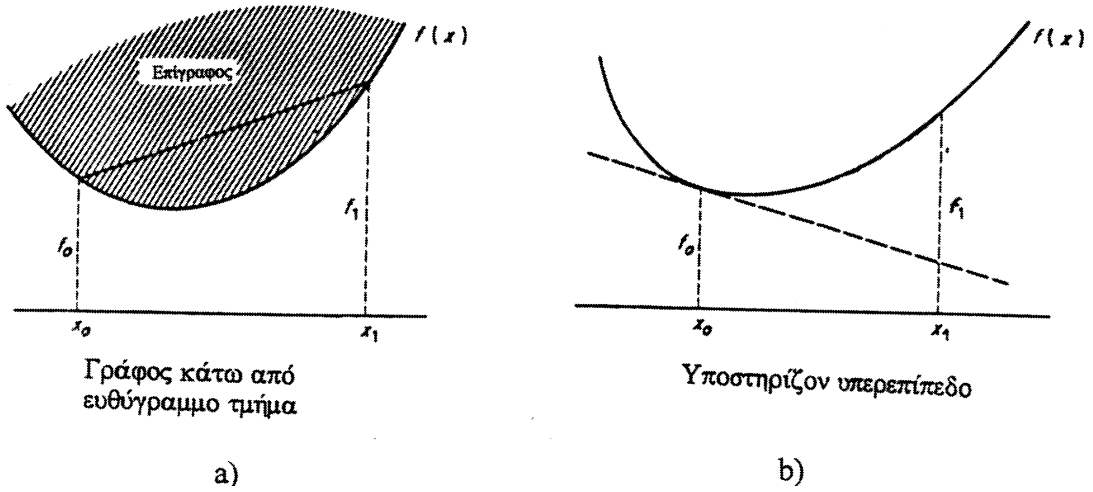
Είναι εύκολο να αποδειχθεί πως, αν K_i, i=1,..., m, είναι κυρτά σύνολα, τότε η τομή $K=K_1\cap K_2\cap\ldots\cap K_m$ είναι κυρτό σύνολο (βλ. Σχ. 6.4.1.b). Άρα, αν κάθε περιορισμός ενός προβλήματος βελτιστοποίησης προσδιορίζει ένα κυρτό σύνολο, τότε η επιτρεπτή περιοχή $\mathcal{R}$ είναι κυρτό σύνολο.

Μια συνεχής **συνάρτηση** f(**x**), $\mathbf{x}\in K$, K κυρτό, ονομάζεται **κυρτή,** αν ο επίγραφός της (Σχ. 6.4.2a), δηλαδή το σύνολο των σημείων του RxR^n που βρίσκονται επί ή υπεράνω του γράφου

της f(**x**), είναι κυρτό σύνολο. Άρα ο ορισμός είναι ισοδύναμος με την ιδιότητα, για κάθε $\mathbf{x}_0$, $\mathbf{x}_1 \in K$, να συνεπάγεται

$$f_\theta \leq (1-\theta)\, f_0 + \theta f_1 \quad \forall \theta \in [0,1] \tag{6.4.3}$$

όπου $f_\theta = f(\mathbf{x}_\theta)$ κλπ. και $\mathbf{x}_\theta$ από την (6.4.1). Αν το K είναι ανοικτό σύνολο και $f(\mathbf{x}) \in C^1$, τότε ένας ισοδύναμος ορισμός κυρτότητας (που μπορεί εύκολα να αποδειχθεί) είναι, για κάθε $\mathbf{x}_0, x_1 \in K$, να συνεπάγεται (βλ. Σχ. 6.4.2b)



Σχήμα 6.4.2: Κυρτές συναρτήσεις.

$$f_1 \geq f_0 + (\mathbf{x}_1 - \mathbf{x}_0)^T \nabla f_0 \tag{6.4.4}$$

δηλαδή ο γράφος της f(**x**) βρίσκεται επί ή υπεράνω της γραμμικοποίησης της f(**x**) περί το $\mathbf{x}_0$, και άρα η γραμμικοποίηση δίδει ένα **υποστηρίζον υπερεπίπεδο** της κυρτής συνάρτησης. Άμεση απόρροια της (6.4.4) είναι

$$(\mathbf{x}_1 - \mathbf{x}_0)^T \nabla f_1 \geq f_1 - f_0 \geq (\mathbf{x}_1 - \mathbf{x}_0)^T \nabla f_0 \tag{6.4.5}$$

που υποδηλώνει ότι η κλίση μιας κυρτής συνάρτησης είναι αύξουσα κατά μήκος κάθε γραμμής.

Αν $f(\mathbf{x}) \in C^2$ και κυρτή, τότε ένας ισοδύναμος ορισμός κυρτότητας είναι

$$\nabla^2 f \geq \mathbf{0} \quad \forall \mathbf{x} \in K \tag{6.4.6}$$

δηλαδή κάθε κυρτή συνάρτηση έχει μη αρνητική καμπυλότητα.

Μια **αυστηρά κυρτή συνάρτηση** ορίζεται από την (6.4.3) με αυστηρή ανισότητα για κάθε ζεύγος διαφορετικών $\mathbf{x}_0$, $\mathbf{x}_1$ και $\theta \in (0,1)$. Για C^1-συναρτήσεις, η (6.4.4) είναι και πάλι ισοδύναμος ορισμός αν η ανισότητα είναι αυστηρή και $\mathbf{x}_0 \neq \mathbf{x}_1$. Όμως, για C^2-συναρτήσεις, η αυστηρή εκδοχή της (6.4.6) δεν είναι ισοδύναμη, καθότι, αν $\nabla^2 f > \mathbf{0}$ $\forall \mathbf{x}_0 \in K$, τότε όντως η f(**x**) είναι αυστηρά κυρτή, δεν ισχύει όμως το αντίθετο αφού, π.χ., η $f(x)=x^4$ είναι αυστηρά κυρτή, όμως $\nabla^2 f(0)=0$.

Μια **συνάρτηση** f(**x**) λέγεται **(αυστηρά) κοίλη** αν η –f(**x**) είναι (αυστηρά) κυρτή. Κοίλες συναρτήσεις συνδέονται με φθίνουσα κλίση και μη θετική καμπυλότητα.

Όλες οι γραμμικές συναρτήσεις είναι κυρτές και κοίλες συγχρόνως (γιατί;). Μια τετραγωνική συνάρτηση είναι κυρτή αν ο Χεσιανός της πίνακας είναι θετικά ημιορισμένος. Επίσης, η $\|\mathbf{x}\|$ είναι πάντα κυρτή για κάθε μέτρο, όχι όμως και η $\|\mathbf{c}(\mathbf{x})\|$, εκτός αν η **c**(**x**) είναι γραμμική συνάρτηση.

Αν $f_i(\mathbf{x})$, i=1, ...,m, είναι κυρτές συναρτήσεις σε κυρτό σύνολο K και $\lambda_i \geq 0$, τότε η $\sum_i \lambda_i f_i(x)$ είναι κυρτή συνάρτηση στο K.

Ένα πρόβλημα **κυρτού προγραμματισμού** έχει κυρτή αντικειμενική συνάρτηση και κυρτή επιτρεπτή περιοχή K. Τέτοια προβλήματα προκύπτουν αν το (5.1.1) μπορεί να διατυπωθεί ως εξής:

Ελαχιστοποίηση f(**x**)
λαμβάνοντας υπόψη $\mathbf{x} \in K = \{\mathbf{x} \mid c_i(\mathbf{x}) \geq 0,\ i=1, \ldots, m\}$ (6.4.7)

όπου η f(**x**) είναι κυρτή στο K και οι $c_i(\mathbf{x})$, i=1, ...,m, είναι κοίλες στο R^n. Το ότι η επιτρεπτή περιοχή στο (6.4.7) είναι κυρτή, προκύπτει από την εξής, εύκολα αποδεικνυόμενη ιδιότητα:

Αν η c(**x**) είναι κοίλη συνάρτηση, τότε το σύνολο $S(k) = \{\mathbf{x} \mid c(\mathbf{x}) \geq k\}$ είναι κυρτό.

Από αυτή την ιδιότητα, και επειδή η τομή κυρτών συνόλων είναι κυρτό σύνολο, συνεπάγεται ότι η επιτρεπτή περιοχή K στο (6.4.7) είναι κυρτό σύνολο. Βέβαια το (6.4.7) δεν επιτρέπει καταρχήν την παρουσία περιορισμών ισότητας. Γραμμικοί όμως περιορισμοί μπορούν να εισαχθούν εμμέσως, αφού κάθε γραμμικός περιορισμός ισότητας c(**x**)=0 μπορεί να αναχθεί σε δύο γραμμικούς (άρα κοίλους) περιορισμούς ανισότητας $c(\mathbf{x}) \geq 0$ και $-c(\mathbf{x}) \geq 0$. Τα προβλήματα γραμμικού προγραμματισμού, επί παραδείγματι, είναι ειδικές περιπτώσεις κυρτού προγραμματισμού.

Η σημασία της κυρτότητας για την βελτιστοποίηση γίνεται φανερή από τα ακόλουθα θεωρήματα.

Θεώρημα 6.4.1: Κάθε τοπικό ελάχιστο $\mathbf{x}^*$ ενός προβλήματος κυρτού προγραμματισμού (6.4.7) είναι και απόλυτο ελάχιστο, και το σύνολο S των απόλυτων ελάχιστων είναι κυρτό. Αν, επιπλέον, η f(**x**) είναι αυστηρά κυρτή στο K, τότε το απόλυτο ελάχιστο είναι μοναδικό.

■

Θεώρημα 6.4.2: Στο πρόβλημα κυρτού προγραμματισμού (6.4.7), αν f(**x**) και $c_i(\mathbf{x})$, i=1,...,m, είναι C^1-συναρτήσεις στο K και αν ικανοποιούνται οι συνθήκες Kuhn-Tucker (6.2.5) στο $\mathbf{x}^*$, τότε το $\mathbf{x}^*$ είναι απόλυτο ελάχιστο, δηλαδή οι συνθήκες Kuhn-Tucker είναι **ικανές και αναγκαίες συνθήκες** για απόλυτα ελάχιστα.

■

Προβλήματα

6.1 Να προσδιορισθούν οι αναγκαίες συνθήκες πρώτου βαθμού που απορρέουν από τις (6.2.5) για το μονοδιάστατο πρόβλημα ελαχιστοποίησης f(x) λαμβάνοντας υπόψη $a \le x \le b$, όπου $a < b$. Ποιά πρακτική διαδικασία προσδιορισμού τοπικών ελαχίστων προκύπτει;

Λύση

Έχουμε $c_1(x) = x–a$ και $c_2(x) = b–x$, άρα $L(x, \boldsymbol{\lambda}) = f(x) – \lambda_1(x–a) – \lambda_2(b–x)$, εξού με τις (6.2.5)

$$\partial L/\partial x = \partial f/\partial x – \lambda_1+\lambda_2 = 0.$$

Οι συνθήκες συμπληρωματικότητας απαιτούν $\lambda_1 = 0$, αν $x>a$, και $\lambda_2 = 0$, αν $x<b$. Άρα, αν $x = a$ ή $x = b$, έχουμε $\lambda_1 = \partial f/\partial x$ ή $\lambda_2 = –\partial f/\partial x$, αντίστοιχα. Επειδή $\lambda_i \ge 0$, $i = 1, 2$, έχουμε τελικά τις ακόλουθες δυνατότητες τοπικού ελάχιστου:

(i) $a<x^*<b$ αν $\partial f^*/\partial x = 0$ (οπότε $\lambda_1^* = \lambda_2^* = 0$).
(ii) $x^* = a$ αν $\partial f(a)/\partial x \ge 0$ (οπότε $\lambda_1^* = \partial f(a)/\partial x$, $\lambda_2^* = 0$)
(iii) $x^* = b$ αν $\partial f(b)/\partial x \le 0$ (οπότε $\lambda_1^* = 0$, $\lambda_2^* = –\partial f(b)/\partial x$).

Άρα προσδιορισμός τοπικών ελαχίστων ως εξής:
(i) Προσδιορίζω τις λύσεις της $\partial f/\partial x = 0$ και επιλέγω εκείνες που ικανοποιούν $a < x^* < b$.
(ii) Αν $\partial f(a)/\partial x \ge 0$ τότε, επιπλέον των (i), $x^* = a$.
(iii) Αν $\partial f(b)/\partial x \le 0$ τότε, επιπλέον των (i) και (ii), $x^* = b$.

6.2 Να επιλυθεί το Πρόβλημα 5.1 μέσω της μεθόδου πολλαπλασιαστών Lagrange.

Λύση

$$L = –x–y–\lambda(x^2+y^2–1)$$

$$\left.\begin{aligned}\frac{\partial L}{\partial x} &= -1 - 2\lambda x = 0\\ \frac{\partial L}{\partial y} &= -1 - 2\lambda y = 0\end{aligned}\right\} \Rightarrow x^* = y^*$$

$$\frac{\partial L}{\partial \lambda} = x^2+y^2–1=0 \Rightarrow 2x^2=1 \Rightarrow x^*=y^*=\pm\frac{\sqrt{2}}{2}, \quad \lambda^*=\mp\frac{\sqrt{2}}{2}$$

$$\nabla_x^2 L = \mathbf{W} = \begin{bmatrix} -2\lambda & 0 \\ 0 & -2\lambda \end{bmatrix} \quad \mathbf{W}^* = \begin{bmatrix} \pm\sqrt{2} & 0 \\ 0 & \pm\sqrt{2} \end{bmatrix}$$

για $\lambda^* = -\frac{\sqrt{2}}{2} \Rightarrow \mathbf{W}^*>\mathbf{0} \Rightarrow x^*=y^* = \frac{\sqrt{2}}{2}$ ελάχιστο

για $\lambda^* = \frac{\sqrt{2}}{2} \Rightarrow \mathbf{W}^*<\mathbf{0} \Rightarrow x^*=y^* = -\frac{\sqrt{2}}{2}$ μέγιστο.

6.3 Να επιλυθεί το Πρόβλημα 5.2 μέσω της μεθόδου πολλαπλασιαστών Lagrange.

Λύση

$$L = 4 - x_1^2 - x_2^2 - \lambda(1 - x_1^2 + x_2)$$

$$\frac{\partial L}{\partial x_1} = -2x_1 + 2x_1\lambda = 0$$
$$\frac{\partial L}{\partial x_2} = -2x_2 - \lambda = 0$$
$$\frac{\partial L}{\partial \lambda} = 1 - x_1^2 + x_2 = 0$$

3 Λύσεις:

Λύση	x_1^*	x_2^*	λ^*
1	0	–1	2
2	$\sqrt{2}/2$	–0.5	1
3	$-\sqrt{2}/2$	–0.5	1

$$\nabla_x^2 L = \mathbf{W} = \begin{bmatrix} -2+2\lambda & 0 \\ 0 & -2 \end{bmatrix} \Rightarrow \mathbf{W}_1^* = \begin{bmatrix} 2 & 0 \\ 0 & -2 \end{bmatrix}, \quad \mathbf{W}_{2/3}^* = \begin{bmatrix} 0 & 0 \\ 0 & -2 \end{bmatrix}.$$

$\mathbf{W}_1^*, \mathbf{W}_2^*, \mathbf{W}_3^*$ μη ορισμένοι για τυχαίες **s**, άρα διερευνούμε το πολυώνυμο $p(\Lambda) = 0$:

$$\begin{bmatrix} \Lambda+2-2\lambda & 0 & -2x_1 \\ 0 & \Lambda+2 & 1 \\ -2x_1 & 1 & 0 \end{bmatrix} = 0 \Rightarrow \Lambda = \frac{-2+2\lambda-8x_1^2}{4x_1^2+1}$$

$\Rightarrow \Lambda_1 = 2 > 0$: ελάχιστο, $\Lambda_{2/3} = -4/3 < 0$: μέγιστα

6.4 Ζητείται το ελάχιστο της $\varphi(\mathbf{x}) = -x_1 - x_2 + 3x_3$ λαμβάνοντας υπόψη τους περιορισμούς

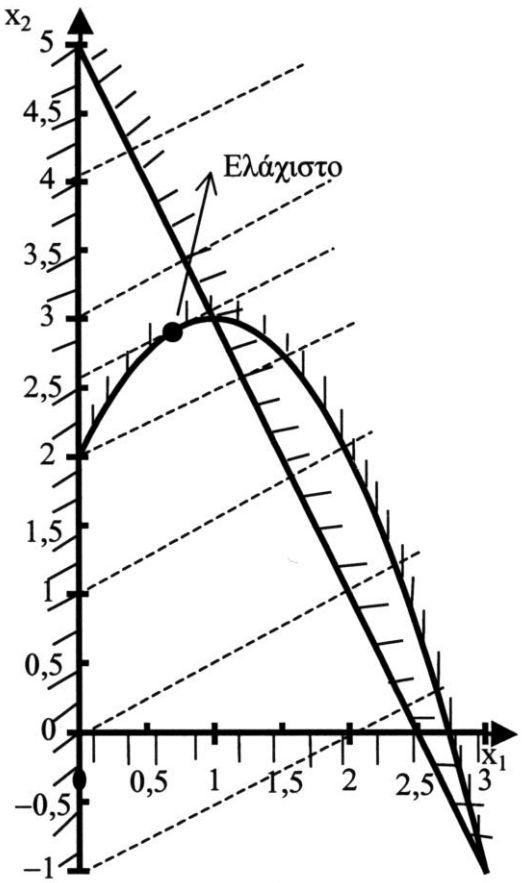
$x_1 - 2x_3 = 0$
$x_1 + x_2 + 2x_3 \le 5$
$(x_1-1)^2 + x_2 \le 3$
$x_1 \ge 0$
$x_2 \ge 0.$

a) Απαλείψτε το x_3 από το πρόβλημα.
b) Προσδιορίστε γραφικά τη λύση.
c) Αποδείξτε ότι η γραφική λύση εκπληροί τις ικανές συνθήκες ελαχιστοποίησης.

Λύση

a) $x_3=x_1/2 \Rightarrow \varphi(\mathbf{x}) = f(x_1,x_2) = 0.5x_1-x_2$.
Περιορισμοί ανισότητας: $5-2x_1-x_2 \geq 0$, $3-(x_1-1)^2-x_2\geq 0$, $x_1\geq 0$, $x_2\geq 0$.

b)



c) Ενεργός ο περιορισμός παραβολής $\Rightarrow L=0.5x_1-x_2-\lambda[3-(x_1-l)^2-x_2]$

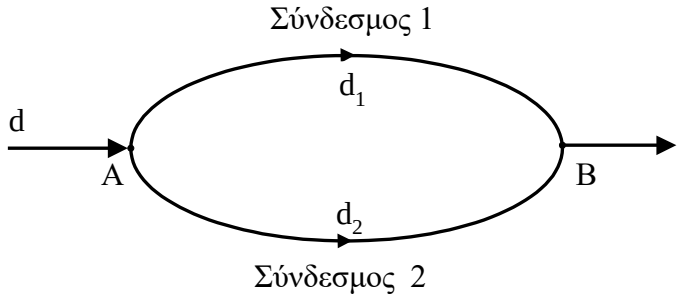
$$\left.\begin{aligned} \frac{\partial L}{\partial x_1} &= 0.5+2\lambda(x_1-1)=0 \\ \frac{\partial L}{\partial x_2} &= -1+\lambda=0 \Rightarrow \lambda^*=1>0 \end{aligned}\right\} \Rightarrow x_1^*=0.75$$

$3-(x_1-1)^2-x_2=0 \Rightarrow x_2^* = 2.94$.

$\mathbf{W} = \begin{bmatrix} 2\lambda & 0 \\ 0 & 0 \end{bmatrix}$, $\mathbf{W}^* = \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix} \geq \mathbf{0}$ εκπληρούνται οι αναγκαίες, όχι όμως αναγκαστικά και οι ικανές συνθήκες

$$p(\Lambda) = \begin{bmatrix} \Lambda-2 & 0 & 0.5 \\ 0 & \Lambda & -1 \\ 0.5 & -1 & 0 \end{bmatrix} = 0 \Rightarrow \Lambda = 1.6 > 0,$$ άρα σίγουρα ελάχιστο!

6.5 Η γνωστή εισροή d [οχήματα/h] στο σημείο Α ενός κυκλοφοριακού δικτύου (βλ. Σχήμα) κατανέμεται σε δύο ροές d_1 (σύνδεσμος 1) και d_2 (σύνδεσμος 2). Οι χρόνοι διαδρομής από Α προς Β είναι $t_1 = 1 + \frac{d_1^2}{3}$ (σύνδεσμος 1) και $t_2 = 2 + \frac{d_2^2}{3}$ (σύνδεσμος 2). Ο συνολικός χρόνος όλων των οχημάτων είναι $T(d_1, d_2) = t_1d_1 + t_2d_2$.



Ζητείται η βέλτιστη κατανομή της εισροής d που ελαχιστοποιεί τον συνολικό χρόνο Τ.

α) Διατυπώστε το πρόβλημα βελτιστοποίησης με μεταβλητές d_1, d_2.
b) Αναγράψτε τις αναγκαίες συνθήκες Kuhn-Tucker.
c) Επιλύστε το πρόβλημα σε συνάρτηση της εισροής $d \geq 0$. Ελέγξτε τις ικανές συνθήκες.
d) Υπολογίστε για d = 0.5 και d = 2 την βέλτιστη κατανομή, τους χρόνους διαδρομής επί των συνδέσμων 1 και 2 και τον ελάχιστο συνολικό χρόνο Τ.

Λύση

a) $T(\mathbf{d}) = d_1(1+\frac{d_1^2}{3}) + d_2(2+\frac{d_2^2}{3}) \rightarrow \text{Min}$

υπό $d_1 + d_2 = d$, $d_1 \geq 0$, $d_2 \geq 0$.

b) $L = d_1(1+\frac{d_1^2}{3}) + d_2(2+\frac{d_2^2}{3}) - \lambda_1(d_1+d_2-d) - \lambda_2 d_1 - \lambda_3 d_2$

$\frac{\partial L}{\partial d_1} = 1 + d_1^2 - \lambda_1 - \lambda_2 = 0$ (1) $\lambda_2, \lambda_3 \geq 0$ (4)

$\frac{\partial L}{\partial d_2} = 2 + d_2^2 - \lambda_1 - \lambda_3 = 0$ (2) $\lambda_2 d_1 = 0, \lambda_3 d_2 = 0$ (5)

$\frac{\partial L}{\partial \lambda_1} = d_1 + d_2 - d = 0$ (3) $d_1, d_2 \geq 0$ (6)

c) **Περίπτωση 1**: $d_1^*, d_2^* > 0$

(5) $\Rightarrow \lambda_2^*, \lambda_3^* = 0$

$$\left.\begin{array}{l}(1),(2)\Rightarrow 1-d_1^2+d_2^2=0\\ \quad(3)\end{array}\right\}\Rightarrow 1+d^2-2dd_1=0\Rightarrow d_1^*=\frac{d^2+1}{2d}>0\ \forall d>0$$

$$(3)\Rightarrow d_2^*=d-d_1^*=\frac{d^2-1}{2d}>0 \text{ αν } d>1.$$

Περίπτωση 2: $d_1^*=0,\ d_2^*>0\Rightarrow\lambda_3^*=0$

$$\left.\begin{array}{l}(2),(3)\Rightarrow \lambda_1=2+d^2\\ \quad(1)\end{array}\right\}\Rightarrow \lambda_2^*=-1-d^2<0 \text{ αδύνατο!}$$

Περίπτωση 3: $d_1^*>0,\ d_2^*=0\Rightarrow\lambda_2^*=0$

(1), (2), (3) $\Rightarrow$ $\lambda_3^*=2-\lambda_1=1-d^2\geq 0$ αν $d\leq 1$.

Άρα γενική λύση:

$$d_1^*=\begin{cases}d & \text{αν } d\leq 1\\ \dfrac{d^2+1}{2d} & \text{αν } d>1\end{cases}\qquad d_2^*=\begin{cases}0 & \text{αν } d\leq 1\\ \dfrac{d^2-1}{2d} & \text{αν } d>1.\end{cases}$$

Ικανές συνθήκες:

Περίπτωση 1: $\nabla_x^2 L=\begin{bmatrix}2d_1 & 0\\ 0 & 2d_2\end{bmatrix}>\mathbf{0}$ αφού $d_1, d_2>0\Rightarrow$ ελάχιστο.

Περίπτωση 3: $\nabla_x^2 L=\begin{bmatrix}2d_1 & 0\\ 0 & 2d_2\end{bmatrix}\geq\mathbf{0}$ αφού $d_2=0$, άρα δεν αποδεικνύεται ότι εκπληρούνται οι ικανές συνθήκες.

$p(\Lambda)$ κατά μήκος $d_1+d_2-d=0$:

$$p(\Lambda)=\begin{vmatrix}\Lambda-2d_1 & 0 & 1\\ 0 & \Lambda-2d_2 & 1\\ 1 & 1 & 0\end{vmatrix}=0\Rightarrow\Lambda=d_1>0\Rightarrow \text{ελάχιστο.}$$

d) $d=0.5$: $d_1=0.5$, $d_2=0$, $t_1=1.08$, $t_2=2$, $T=0.54$
$d=2$: $d_1=1.25$, $d_2=0.75$, $t_1=1.52$, $t_2=2.19$, $T=3.54$.

6.6 Τρεις μονάδες παραγωγής P_1, P_2, P_3 παράγουν το ίδιο προϊόν με ρυθμούς παραγωγής x_1, x_2, x_3. Ο συνολικός ρυθμός παραγωγής είναι $X=x_1+x_2+x_3$. Οι συναρτήσεις κόστους παραγωγής $k_i(x_i)$ των μονάδων είναι γνωστές. Για δεδομένο και σταθερό ρυθμό παραγωγής X, ζητείται η κατανομή **x** που ελαχιστοποιεί το συνολικό κόστος $K=k_1+k_2+k_3$.

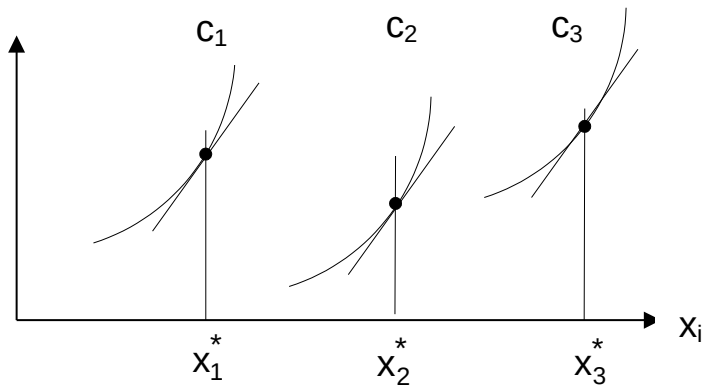
a) Αναγράψτε τις αναγκαίες συνθήκες ελαχιστοποίησης πρώτου βαθμού χρησιμοποιώντας την μέθοδο Lagrange. Δώστε μια γραφική ερμηνεία των συνθηκών.
b) Επιλύστε το πρόβλημα για $k_i = a_i + b_i x_i^2$, όπου a_i, $b_i > 0$, i = 1, 2, 3, και υπολογίστε το ελάχιστο συνολικό κόστος παραγωγής K*.
c) Υπολογίστε την ευαισθησία dK*/dX του ελάχιστου κόστους παραγωγής ως προς το συνολικό ρυθμό παραγωγής.

Λύση

a) $K(\mathbf{x}) = k_1(x_1) + k_2(x_2) + k_3(x_3) \rightarrow \text{Min}$
υπό $x_1 + x_2 + x_3 - X = 0$ (αγνοούμε τους $x_i \geq 0$, i = 1, 2, 3)
$L(\mathbf{x}, \lambda) = k_1(x_1) + k_2(x_2) + k_3(x_3) - \lambda(x_1 + x_2 + x_3 - X)$

$$\frac{\partial L}{\partial x_i} = \frac{\partial k_i}{\partial x_i} - \lambda = 0,\ i = 1, 2, 3 \Rightarrow \frac{\partial k_i}{\partial x_i} \text{ ίσα για } i = 1, 2, 3$$

δηλαδή ίσα "διαφορικά κόστη" (= ίση μεταβολή κόστους)



b) $k_i(x_i) = a_i + b_i x_i^2$

$$\Rightarrow \frac{\partial k_i}{\partial x_i} = 2b_i x_i \Rightarrow b_1x_1 = b_2x_2 = b_3x_3 \text{ και } x_1+x_2+x_3 = X.$$

Με $N = b_1b_2 + b_2b_3 + b_1b_3$ έχουμε

$$x_1^* = \frac{b_2b_3}{N}X,\ x_2^* = \frac{b_1b_3}{N}X,\ x_3^* = \frac{b_1b_2}{N}X.$$

Απόλυτο ελάχιστο, αφού το πρόβλημα είναι κυρτό! (γιατί;)

$$\lambda^* = 2\frac{b_1b_2b_3}{N}X,\ \ K^* = a_1 + a_2 + a_3 + \frac{b_1b_2b_3}{N}X^2.$$

c) $\dfrac{dK^*}{dX} = 2\dfrac{b_1b_2b_3}{N}X = \lambda^*$ σύμφωνα με την (6.1.10).

6.7 a) Δείξτε γραφικά ότι $\mathcal{R} = \{\mathbf{x}\in R^2 \mid x_1 \geq 0,\ x_2 \geq 0,\ 1-x_1-x_2 \geq 0\}$ είναι κυρτό σύνολο.

b) Αποδείξτε ότι $\mathcal{R} = \{\mathbf{x}\in R^n \mid \mathbf{Ax} \leq \mathbf{b}\}$, $\mathbf{A}\in R^{m\times n}$, είναι κυρτό σύνολο.

c) Είναι οι ακόλουθες συναρτήσεις κυρτές;

(1) $f(x) = e^{-x}$, $x\in R$

(2) $f(x) = \sqrt{x}$, $x\in R^+$

(3) $f(x) = e^{-x} - \sqrt{x}$, $x\in R^+$

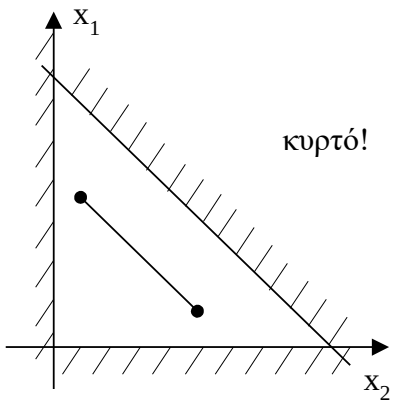
(4) $f(\mathbf{x}) = x_1^2+2x_2$, $\mathbf{x}\in R^2$

(5) $f(\mathbf{x}) = x_1^2-x_2$, $\mathbf{x}\in R^2$

(6) $f(\mathbf{x}) = \mathbf{x}^T\mathbf{Q}\mathbf{x}$, $\mathbf{x}\in R^n$.

Λύση

a)



b) $\mathbf{x}_1\in\mathcal{R}$, $\mathbf{x}_2\in\mathcal{R}$, $\theta\in[0,1]$:

$$\mathbf{x}_\theta = \theta\mathbf{x}_1 + (1-\theta)\mathbf{x}_2 \Rightarrow \mathbf{Ax}_\theta = \theta\underbrace{\mathbf{Ax}_1}_{\leq\mathbf{b}} + (1-\theta)\underbrace{\mathbf{Ax}_2}_{\leq\mathbf{b}} \leq \mathbf{b} \Rightarrow \mathbf{x}_\theta\in\mathcal{R} \text{ ο.ε.δ.}$$

c) (1) $f''(x) = e^{-x} > 0\ \forall x\in R \Rightarrow$ κυρτή

(2) $f''(x) = -\frac{1}{4}x^{-3/2} < 0\ \forall x\in R \Rightarrow$ κοίλη

(3) f(x) κυρτή, αφού θετικό άθροισμα δύο κυρτών συναρτήσεων (η $-\sqrt{x}$ είναι κυρτή, αφού η $\sqrt{x}$ είναι κοίλη)

(4) $\nabla^2 f = \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix} \geq \mathbf{0} \Rightarrow$ κυρτή

(5) $\nabla^2 f = \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix} \geq \mathbf{0} \Rightarrow$ κυρτή

(6) $\nabla^2 f = \mathbf{Q}+\mathbf{Q}^T \Rightarrow$ κυρτή π.χ. αν $\mathbf{Q}$ συμμετρικός και $\mathbf{Q} \geq \mathbf{0}$.

6.8 Θεωρούμε το πρόβλημα ελαχιστοποίησης της $f(\mathbf{x}) = x_1^3 + x_1^2x_2 + 0.5(x_1-6)^2$ υπό $x_1+x_2 = 1$. Επιλύστε το πρόβλημα με εφαρμογή των αναγκαίων και ικανών συνθηκών.

Λύση

$$L(\mathbf{x},\lambda) = x_1^3 + x_1^2x_2 + 0.5(x_1-6)^2 - \lambda(x_1+x_2-1)$$

Αναγκαίες συνθήκες 1ου βαθμού:

$$\left.\begin{aligned}
\frac{\partial L}{\partial x_1} &= 3x_1^2 + 2x_1x_2 + x_1 - 6 - \lambda = 0 \\
\frac{\partial L}{\partial x_2} &= x_1^2 - \lambda = 0 \Rightarrow \lambda = x_1^2 \\
\frac{\partial L}{\partial \lambda} &= x_1 + x_2 - 1 = 0 \Rightarrow x_2 = 1 - x_1
\end{aligned}\right\} \Rightarrow x_1^* = 2,\ x_2^* = -1,\ \lambda^* = 4.$$

Ικανές συνθήκες:

$$\mathbf{W} = \begin{bmatrix} 6x_1 + 2x_2 + 1 & 2x_1 \\ 2x_1 & 0 \end{bmatrix} \Rightarrow \mathbf{W}^* = \begin{bmatrix} 11 & 4 \\ 4 & 0 \end{bmatrix} \text{ μη ορισμένος!}$$

$$p(\Lambda) = \begin{vmatrix} \Lambda - 11 & -4 & 1 \\ -4 & \Lambda & 1 \\ 1 & 1 & 0 \end{vmatrix} = -2\Lambda + 3 = 0 \Rightarrow \Lambda = 1.5 > 0 \text{ άρα ελάχιστο!}$$

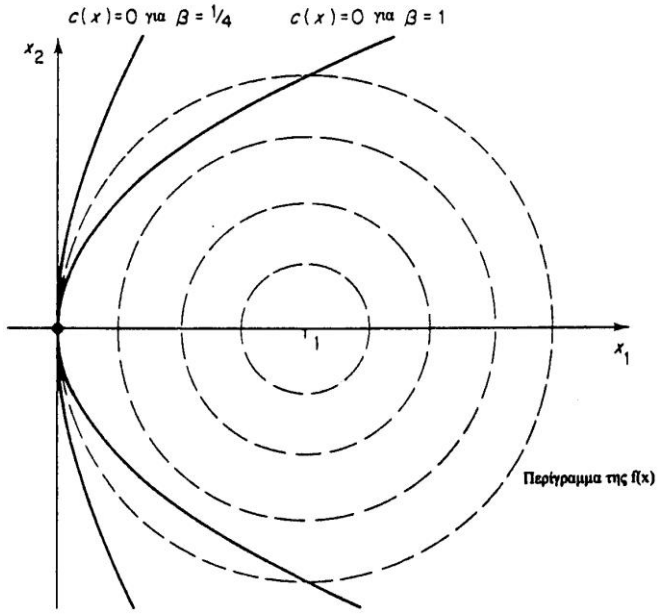
6.9 Θεωρούμε το πρόβλημα ελαχιστοποίησης της $f(\mathbf{x}) = 0.5[(x_1-1)^2 + x_2^2]$ υπό $-x_1 + \beta x_2^2 = 0$, όπου $\beta>0$ δεδομένη παράμετρος. Επιλύστε το πρόβλημα:

(a) Γραφικά

(b) Μέσω αναγκαίων και ικανών συνθηκών.

Λύση

α)



$\mathbf{x}^* = \mathbf{0}$ είναι ελάχιστο για $\beta \le 0.5$, όχι όμως για $\beta > 0.5$.

b) $L(\mathbf{x},\lambda) = 0.5[(x_1-1)^2 + x_2^2] - \lambda(-x_1 + \beta x_2^2)$

Αναγκαίες συνθήκες 1ου βαθμού:

$$\left.\begin{aligned} \frac{\partial L}{\partial x_1} &= x_1 - 1 + \lambda = 0 \\ \frac{\partial L}{\partial x_2} &= x_2 - 2\lambda\beta x_2 = x_2(1 - 2\lambda\beta) = 0 \\ \frac{\partial L}{\partial \lambda} &= x_1 - \beta x_2^2 = 0 \end{aligned}\right\} \Rightarrow$$

1η λύση: $x_1^* = x_2^* = 0$ και $\lambda^* = 1$

2η λύση: $x_1^* = \dfrac{2\beta - 1}{2\beta}$, $x_2^* = \pm\dfrac{1}{\beta}\sqrt{\dfrac{2\beta-1}{2\beta}}$, $\lambda = \dfrac{1}{2\beta}$ (μόνον για $\beta > 0.5$)

Ικανές συνθήκες:

$$\mathbf{W} = \begin{bmatrix} 1 & 0 \\ 0 & 1 - 2\lambda\beta \end{bmatrix} \Rightarrow$$

1η λύση:

$\mathbf{W}^* = \begin{bmatrix} 1 & 0 \\ 0 & 1 - 2\beta \end{bmatrix} > \mathbf{0}$ για $\beta < 0.5$ ενώ για $\beta > 0.5$ ο $\mathbf{W}^*$ είναι μη ορισμένος και άρα το $\mathbf{x}^* = \mathbf{0}$ δεν είναι ελάχιστο παρότι $\mathbf{G} > \mathbf{0}$! Για $\beta = 0.5$ πρέπει να ελεγχθεί το πολυώνυμο $p(\Lambda)$.

2η λύση:

$\mathbf{W}^* = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \ge \mathbf{0}$ (πρέπει να ελεγχθεί το πολυώνυμο $p(\Lambda)$).

$$p(\Lambda) = \begin{bmatrix} \Lambda - 1 & 0 & -1 \\ 0 & \Lambda - 1 + 2\lambda\beta & 2\beta x_2 \\ -1 & 2\beta x_2 & 0 \end{bmatrix} = 0 \Leftrightarrow \Lambda = 1 - \frac{2\lambda\beta}{4\beta^2 x_2^2 + 1}$$

1η λύση: $\Lambda = 1 - 2\beta$, δηλαδή για $\beta = 0.5 \Rightarrow \Lambda = 0$ και άρα δεν ικανοποιούνται οι ικανές συνθήκες. Εν τούτοις, για $\beta = 0.5$ το $\mathbf{x}^* = \mathbf{0}$ είναι ελάχιστο, βλ. γραφική λύση.

2η λύση: $\Lambda = 2\dfrac{2\beta - 1}{5\beta - 2} > 0$ για $\beta > 0.5 \Rightarrow$ ελάχιστο!

Συνοπτικά έχουμε ελάχιστο

$$\mathbf{x}^* = \begin{cases} \mathbf{0} & \text{για } \beta < 0.5 \\ \mathbf{0} & \text{για } \beta = 0.5 \quad \text{(που όμως δεν ικανοποιεί τις ικανές συνθήκες)} \\ [\dfrac{2\beta-1}{2\beta} \quad \pm\dfrac{1}{\beta}\sqrt{\dfrac{2\beta-1}{2\beta}}]^T & \text{για } \beta > 0.5. \end{cases}$$

6.10 Δίνεται προς ελαχιστοποίηση η συνάρτηση

$f(\mathbf{x}) = (x_1 - 3)^2 + (x_2 - 2)^2$

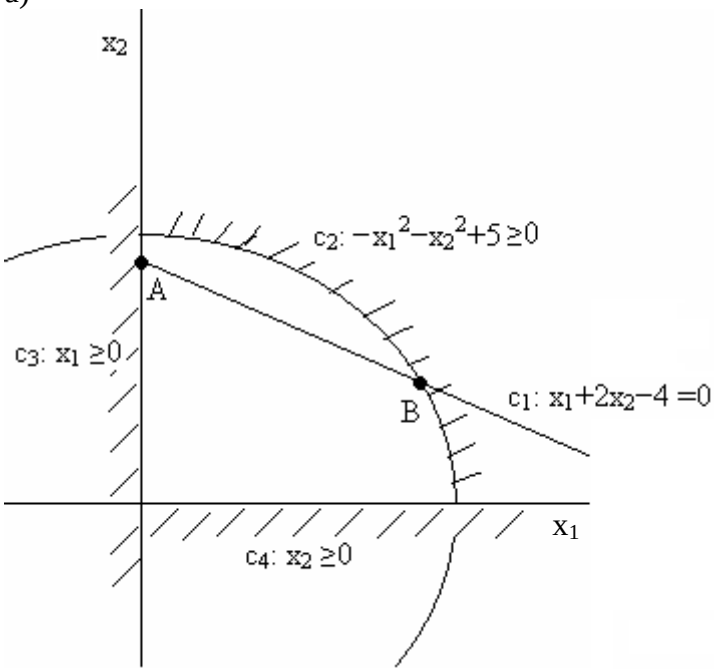
υπό τους περιορισμούς

$x_1 + x_2 = 4$
$x_1^2 + x_2^2 \leq 5$
$x_1 \geq 0$
$x_2 \geq 0.$

a) Σχεδιάστε στο χώρο (x_1,x_2) τους περιορισμούς ισότητας και ανισότητας και καθορίστε την επιτρεπτή περιοχή του προβλήματος. Ποιοι περιορισμοί ανισότητας μπορούν να ενεργοποιηθούν βάσει της προσδιορισθείσας επιτρεπτής περιοχής;
b) Επιλύστε αναλυτικά το πρόβλημα μέσω των αναγκαίων και ικανών συνθηκών βελτιστοποίησης. *(Σημείωση: Κατά την* επίλυση, εξετάστε μόνον περιπτώσεις ενεργοποίησης περιορισμών ανισότητας που είναι επιτρεπτές βάσει του προηγούμενου ερωτήματος.)

Λύση

a)



Η επιτρεπτή περιοχή του προβλήματος είναι το ευθύγραμμο τμήμα AB (δηλαδή το κομμάτι του περιορισμού ισότητας για το οποίο ικανοποιούνται οι περιορισμοί ανισότητας). Οι περιορισμοί ανισότητας που μπορούν να ενεργοποιηθούν (κατά μόνας) είναι είτε ο c_2 $(-x_1^2-x_2^2+5\geq 0)$ είτε ο c_3 $(x_1\geq 0)$, αφού έχουν (διαφορετικά) σημεία που ανήκουν στην επιτρεπτή περιοχή.

b) Συνθήκες Kuhn-Tucker (χωρίς τον περιορισμό $x_2\geq 0$ ο οποίος δεν ενεργοποιείται εντός της επιτρεπτής περιοχής)

$L = (x_1-3)^2 + (x_2-2)^2 - \lambda_1(x_1+2x_2-4) + \lambda_2(x_1^2+x_2^2-5) - \lambda_3 x_1$

$\frac{\partial L}{\partial x_1} = 2(x_1-3) - \lambda_1 + 2\lambda_2 x_1 - \lambda_3 = 0$ (α) $\lambda_2, \lambda_3 \geq 0$ (δ)

$\frac{\partial L}{\partial x_2} = 2(x_2-2) - 2\lambda_1 + 2\lambda_2 x_2 = 0$ (β) $\lambda_2 c_2 = \lambda_3 c_3 = 0$ (ε)

$\frac{\partial L}{\partial \lambda_1} = -x_1 - 2x_2 + 4 = 0$ (γ) $c_2, c_3 \geq 0$ (στ)

<u>Περίπτωση 1:</u> $\lambda_2=\lambda_3=0$ (δεν ενεργοποιείται κανένας περιορισμός)

$$\left.\begin{aligned} \frac{\partial L}{\partial x_1} &= 2(x_1-3) - \lambda_1 = 0 \\ \frac{\partial L}{\partial x_2} &= 2(x_2-2) - 2\lambda_1 = 0 \\ \frac{\partial L}{\partial \lambda_1} &= -x_1 - 2x_2 + 4 = 0 \end{aligned}\right\} \quad x_1^* = \frac{12}{5},\ x_2^* = \frac{4}{5},\ \lambda_1^* = -\frac{6}{5} \text{ (μη αποδεκτή λόγω } c_2)$$

<u>Περίπτωση 2:</u> $\lambda_3=c_2=0$ (ενεργοποιείται μόνον ο c_2)

$$\frac{\partial L}{\partial x_1} = (2x_1-3) - \lambda_1 + 2\lambda_2 x_1 = 0 \quad (1)$$

$$\frac{\partial L}{\partial x_2} = 2(x_2-2) - 2\lambda_1 + 2\lambda_2 x_2 = 0 \quad (2)$$

$$\frac{\partial L}{\partial \lambda_1} = x_1 - 2x_2 + 4 = 0 \quad (3)$$

$$x_1^2 + x_2^2 - 5 = 0 \quad (4)$$

Από (3),(4) $\Rightarrow 5x_2^2 - 16x_2 + 11 = 0 \Rightarrow x_2^* = \frac{11}{5}$ ή $x_2^* = 1$

αν $x_2^* = \frac{11}{5}$ τότε $x_1^* = \frac{-2}{5} < 0$ (μη αποδεκτή λύση λόγω (στ))

άρα $x_2^*=1$, $x_1^*=2$, $\lambda_1^* = -\frac{2}{3}$, $\lambda_2^* = \frac{1}{3} > 0$ (αποδεκτή λύση).

Ικανές συνθήκες δευτέρου βαθμού:

$\mathbf{W}=\begin{bmatrix} 2+2\lambda_2^* & 0 \\ 0 & 2+2\lambda_2^* \end{bmatrix}>\mathbf{0}$ για κάθε $\lambda_2\geq 0$, άρα και για λ_2^*, άρα $\mathbf{x}^*=\begin{bmatrix} 2 \\ 1 \end{bmatrix}$ ελάχιστο.

<u>Περίπτωση 3:</u> $\lambda_2=c_3=0$ (ενεργοποιείται μόνον ο c_3)

Άρα $x_1=0$ και από (α)-(στ) $\Rightarrow$ $x_1^*=0$, $x_2^*=2$, $\lambda_1^*=0$ και $\lambda_3^*=-6<0$ (μη αποδεκτή λύση λόγω (δ)).

<u>Περίπτωση 4:</u> $c_2=c_3=0$, $\lambda_2, \lambda_3>0$ (ενεργοποιούνται και οι 2 περιορισμοί ανισότητας)

Άτοπο: δεν μπορούν να ενεργοποιηθούν οι 2 περιορισμοί ανισότητας ταυτόχρονα, όπως φαίνεται και από το σχήμα.

Άρα $\mathbf{x}^*=\begin{bmatrix} 2 \\ 1 \end{bmatrix}$, $\lambda_1^*=-\frac{2}{3}$, $\lambda_2^*=\frac{1}{3}$ είναι το μοναδικό ελάχιστο.

7. ΤΕΤΡΑΓΩΝΙΚΟΣ ΠΡΟΓΡΑΜΜΑΤΙΣΜΟΣ

Μια ειδική κατηγορία προβλημάτων που μπορούν να επιλυθούν σε πεπερασμένο αριθμό επαναλήψεων είναι εκείνα του **Τετραγωνικού Προγραμματισμού** (**ΤΠ**):

$$\begin{aligned} &\text{Ελαχιστοποίηση} && q(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T\mathbf{G}\mathbf{x}+\mathbf{b}^T\mathbf{x} \\ &\text{λαμβάνοντας υπόψη} && \mathbf{a}_i^T\mathbf{x} = d_i \quad , \quad i\in E \\ & && \mathbf{a}_i^T\mathbf{x} \geq d_i \quad , \quad i\in I \end{aligned} \qquad (7.1.1)$$

όπου ο **G** είναι συμμετρικός. Αν $\mathcal{R}$ = {} ή η λύση δεν είναι φραγμένη, αυτό μπορεί να διαπιστωθεί εύκολα από τους αριθμητικούς αλγόριθμους επίλυσης. Αν $\mathbf{G} \geq \mathbf{0}$ τότε κάθε τοπικό ελάχιστο $\mathbf{x}^*$ είναι απόλυτο ελάχιστο, και αν $\mathbf{G}>\mathbf{0}$ τότε το $\mathbf{x}^*$ είναι και μοναδικό λόγω κυρτότητας και αυστηρής κυρτότητας, αντίστοιχα, του προβλήματος. Αν ο **G** δεν είναι ορισμένος, μπορεί να υπάρχουν τοπικά ελάχιστα που είναι ή δεν είναι απόλυτα ελάχιστα.

7.1 Περιορισμοί Ισότητας

Θεωρούμε καταρχήν προβλήματα ΤΠ που εμπεριέχουν μόνο περιορισμούς ισότητας:

$$\begin{aligned} &\text{Ελαχιστοποίηση} && q(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T\mathbf{G}\mathbf{x}+\mathbf{b}^T\mathbf{x} \\ &\text{λαμβάνοντας υπόψη} && \mathbf{A}^T\mathbf{x} = \mathbf{d} \end{aligned} \qquad (7.1.2)$$

όπου $\mathbf{d}\in R^m$, $m\leq n$, $\mathbf{A}\in R^{n\times m}$ πλήρους βαθμού m. Η παραδοχή ότι ο πίνακας **A** είναι πλήρους βαθμού δεν περιορίζει τη γενικότητα του προβλήματος (γιατί;). Το (7.1.2) μπορεί να επιλυθεί δι' απαλοιφής. Με τους καταμερισμούς

$$\mathbf{x} = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix}, \ \mathbf{A} = \begin{bmatrix} \mathbf{A}_1 \\ \mathbf{A}_2 \end{bmatrix}, \ \mathbf{b} = \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \end{bmatrix}, \ \mathbf{G} = \begin{bmatrix} \mathbf{G}_{11} & \mathbf{G}_{12} \\ \mathbf{G}_{21} & \mathbf{G}_{22} \end{bmatrix}$$

όπου $\mathbf{x}_1\in R^m$, $\mathbf{x}_2\in R^{n-m}$ κλπ., έχουμε $\mathbf{A}_1^T\mathbf{x}_1 + \mathbf{A}_2^T\mathbf{x}_2 = \mathbf{d}$ και άρα

$$\mathbf{x}_1 = \mathbf{A}_1^{-T}(\mathbf{d} - \mathbf{A}_2^T\mathbf{x}_2). \qquad (7.1.3)$$

Βέβαια ο καταμερισμός του **x** πρέπει να γίνει κατά τρόπο ώστε ο τετραγωνικός πίνακας $\mathbf{A}_1\in R^{m\times m}$ να είναι πλήρους βαθμού (και άρα αντιστρέψιμος).

Αντικαθιστώντας στην q(**x**) έχουμε το πρόβλημα απεριόριστης ελαχιστοποίησης της τετραγωνικής συνάρτησης

$$\psi(\mathbf{x}_2) = \frac{1}{2}\,\mathbf{x}_2^{\mathrm{T}}(\mathbf{G}_{22}-\mathbf{G}_{21}\mathbf{A}_1^{-\mathrm{T}}\mathbf{A}_2^{\mathrm{T}}-\mathbf{A}_2\mathbf{A}_1^{-1}\mathbf{G}_{12}+\mathbf{A}_2\mathbf{A}_1^{-1}\mathbf{G}_{11}\mathbf{A}_1^{-\mathrm{T}}\mathbf{A}_2^{\mathrm{T}})\mathbf{x}_2$$
$$+\mathbf{x}_2^{\mathrm{T}}(\mathbf{G}_{21}-\mathbf{A}_2\mathbf{A}_1^{-1}\mathbf{G}_{11})\mathbf{A}_1^{-\mathrm{T}}\mathbf{d}+\frac{1}{2}\mathbf{d}^{\mathrm{T}}\mathbf{A}_1^{-1}\mathbf{G}_{11}\mathbf{A}_1^{-\mathrm{T}}\mathbf{d}$$
$$+\mathbf{x}_2^{\mathrm{T}}(\mathbf{b}_2-\mathbf{A}_2\mathbf{A}_1^{-1}\mathbf{b}_1)+\mathbf{b}_1^{\mathrm{T}}\mathbf{A}_1^{-\mathrm{T}}\mathbf{d} \qquad (7.1.4)$$

που έχει μοναδικό ελάχιστο $\mathbf{x}_2^*$ αν ο Χεσιανός $\nabla^2\psi$ στον τετραγωνικό όρο της (7.1.4) είναι θετικά ορισμένος, οπότε το $\mathbf{x}_2^*$ υπολογίζεται από την επίλυση του γραμμικού συστήματος $\nabla\psi(\mathbf{x}_2) = \mathbf{0}$, ενώ το $\mathbf{x}_1^*$ με αντικατάσταση στην (7.1.3). Ο πολλαπλασιαστής Lagrange εκπληροί την $\mathbf{g}^*=\mathbf{A}\boldsymbol{\lambda}^*$ (βλ. (6.1.3)), όπου $\mathbf{g}^*=\nabla q(\mathbf{x}^*)=\mathbf{b}+\mathbf{G}\mathbf{x}^*$, και υπολογίζεται από τον άνω καταμερισμό της, δηλαδή από $\mathbf{g}_1^*=\mathbf{A}_1\boldsymbol{\lambda}^*$. Έτσι έχουμε

$$\boldsymbol{\lambda}^*=\mathbf{A}_1^{-1}(\mathbf{b}_1+\mathbf{G}_{11}\mathbf{x}_1^*+\mathbf{G}_{12}\mathbf{x}_2^*). \qquad (7.1.5)$$

Παράδειγμα 7.1.1: Θεωρούμε το πρόβλημα ΤΠ

$$q(\mathbf{x}) = x_1^2+x_2^2+x_3^2 \rightarrow \underset{\mathbf{x}}{\mathrm{Min}} \qquad (7.1.6)$$

υπό τους περιορισμούς

$$x_1 + 2x_2 - x_3 = 4$$
$$x_1 - x_2 + x_3 = -2.$$

Με απαλοιφή του x_3 έχουμε

$$x_1 = -\frac{1}{3}x_3, \quad x_2 = 2+\frac{2}{3}x_3 \qquad (7.1.7)$$

που αντιστοιχεί στην (7.1.3) για $\mathbf{x}_1=[x_1\ x_2]^{\mathrm{T}}$, $\mathbf{x}_2=x_3$,

$$\mathbf{A}_1=\begin{bmatrix}1 & 1\\ 2 & -1\end{bmatrix},\ \mathbf{A}_2=[-1\quad 1].$$

Αντικαθιστώντας την (7.1.7) στην (7.1.6) έχουμε

$$\psi(x_3) = \frac{14}{9}x_3^2 + \frac{8}{3}x_3+4 \qquad (7.1.8)$$

που αντιστοιχεί στην (7.1.4) με

$$\mathbf{G}_{11} = \begin{bmatrix}2 & 0\\ 0 & 2\end{bmatrix},\ \mathbf{G}_{12}=\mathbf{G}_{21}^{\mathrm{T}}=\mathbf{0},\ G_{22}=2$$

και **b**=**0**. Ο Χεσιανός της (7.1.8) είναι 28/9 και είναι θετικά ορισμένος, άρα το ελάχιστο υπολογίζεται από $\nabla\psi=0$ και είναι $x_3^* = -6/7$, και από τις (7.1.7) έχουμε $x_1^*=2/7$, $x_2^*=10/7$. Το σύστημα $\mathbf{g}^*=\mathbf{A}\boldsymbol{\lambda}^*$ γίνεται

$$\frac{2}{7}\begin{bmatrix}2\\10\\-6\end{bmatrix}=\begin{bmatrix}1 & 1\\2 & -1\\-1 & 1\end{bmatrix}\begin{bmatrix}\lambda_1^*\\\lambda_2^*\end{bmatrix}$$

και επιλύοντας τις πρώτες δύο γραμμές έχουμε $\lambda_1^* = 8/7$ και $\lambda_2^* = -4/7$ που ικανοποιούν και την τρίτη γραμμή.

■

Η άμεση απαλοιφή μεταβλητών δεν είναι ούτε η μόνη, ούτε αναγκαστικά η καλύτερη μέθοδος επίλυσης του (7.1.2). Μια **γενικευμένη μέθοδος απαλοιφής** επιχειρεί αρχικά έναν γραμμικό μετασχηματισμό μεταβλητών. Έστω $\mathbf{Y}\in R^{n\times m}$ και $\mathbf{Z}\in R^{n\times(n-m)}$ δύο πίνακες, έτσι ώστε ο [**Y Z**] να είναι κανονικός και $\mathbf{A}^T\mathbf{Y} = \mathbf{I}$, $\mathbf{A}^T\mathbf{Z} = \mathbf{0}$. Ο πίνακας **Y** είναι επομένως δεξιός γενικευμένος αντίστροφος του $\mathbf{A}^T$ και άρα μια λύση του $\mathbf{A}^T\mathbf{x}=\mathbf{d}$ είναι $\mathbf{x}=\mathbf{Y}\mathbf{d}$. Βέβαια η λύση αυτή δεν είναι μοναδική, και άλλα επιτρεπτά σημεία δίδονται από $\mathbf{x}=\mathbf{Y}\mathbf{d}+\boldsymbol{\delta}$, όπου το $\boldsymbol{\delta}$ ανήκει στον κενό χώρο στηλών του **A**, δηλαδή στο γραμμικό χώρο που ορίζεται

$$\{\boldsymbol{\delta} \mid \mathbf{A}^T\boldsymbol{\delta}=\mathbf{0}\} \qquad (7.1.9)$$

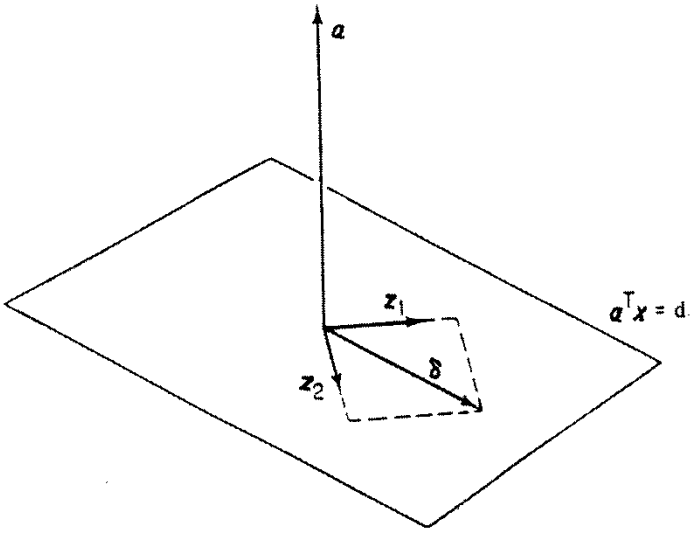
με διάσταση n–m. Ο ρόλος του πίνακα **Z** είναι ότι έχει n–m γραμμικά ανεξάρτητες στήλες $\mathbf{z}_1$, $\mathbf{z}_2$, ..., $\mathbf{z}_{n-m}$ που ανήκουν σ' αυτόν τον κενό χώρο, και άρα ενεργούν ως βασικά διανύσματα (ή κατευθύνσεις μειωθεισών συντεταγμένων) για τον κενό χώρο. Άρα, κάθε ως άνω $\boldsymbol{\delta}$ μπορεί να γραφεί

$$\boldsymbol{\delta}=\mathbf{Z}\mathbf{y} = \sum_{i=1}^{n-m} \mathbf{z}_i y_i \qquad (7.1.10)$$

όπου y_1, y_2, ..., y_{n-m} είναι συνιστώσες (ή μειωθείσες μεταβλητές) για τις αντίστοιχες μειωθείσες συντεταγμένες (βλ. Σχ. 7.1.1 για n=3, m=1). Έτσι, κάθε επιτρεπτό σημείο γράφεται

$$\mathbf{x}=\mathbf{Y}\mathbf{d}+\mathbf{Z}\mathbf{y} \qquad (7.1.11)$$

και αυτό μπορεί να ερμηνευθεί (Σχ. 7.1.2 για n=2, m=1) σαν ένα βήμα από το **0** στο επιτρεπτό σημείο **Yd**, ακολουθούμενο από μια επιτρεπτή διόρθωση **Zy** μέχρι το σημείο **x**. Έτσι, η (7.1.11) απαλείφει τους περιορισμούς $\mathbf{A}^T\mathbf{x} = \mathbf{d}$ ως προς **y**, πράγμα που συνιστά μια γενίκευση της (7.1.3). Αντικαθιστώντας στην q(**x**) έχουμε την μειωθείσα τετραγωνική συνάρτηση



Σχήμα 7.1.1: Μειωθείσες συντεταγμένες για την επιτρεπτή περιοχή.

$$\psi(\mathbf{y}) = \frac{1}{2}\,\mathbf{y}^T\mathbf{Z}^T\mathbf{G}\mathbf{Z}\mathbf{y} + (\mathbf{b}+\mathbf{G}\mathbf{Y}\mathbf{d})^T\mathbf{Z}\mathbf{y} + \frac{1}{2}\,(\mathbf{b}+\mathbf{G}\mathbf{Y}\mathbf{d})^T\mathbf{Y}\mathbf{d}. \qquad (7.1.12)$$

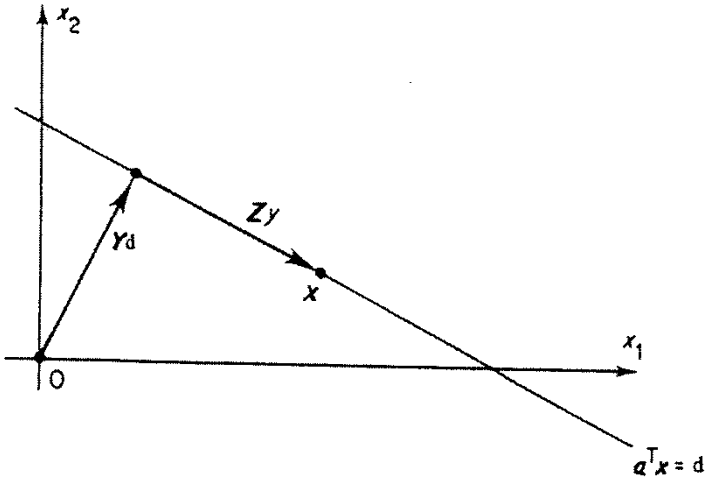
Αν $\mathbf{Z}^T\mathbf{G}\mathbf{Z} > \mathbf{0}$, τότε υπάρχει μοναδικό ελάχιστο $\mathbf{y}^*$ που επιλύει (από $\nabla\psi(\mathbf{y}) = \mathbf{0}$) το γραμμικό σύστημα

$$\mathbf{Z}^T\mathbf{G}\mathbf{Z}\,\mathbf{y} = -\mathbf{Z}^T(\mathbf{b}+\mathbf{G}\mathbf{Y}\mathbf{d}) \qquad (7.1.13)$$

και το $\mathbf{x}^*$ υπολογίζεται δι' αντικαταστάσεως στην (7.1.11). Ο πίνακας $\nabla^2\psi = \mathbf{Z}^T\mathbf{G}\mathbf{Z}$ είναι ο **μειωθείς Χεσιανός** ενώ $\nabla\psi = \mathbf{Z}^T\mathbf{G}\mathbf{Z}\mathbf{y}+\mathbf{Z}^T(\mathbf{b}+\mathbf{G}\mathbf{Y}\mathbf{d})$ είναι το **μειωθέν διάνυσμα κλίσης**. Οι τύποι αυτοί μπορούν να εξαχθούν και με εφαρμογή της (Π1.8) στην (7.1.11) και $\psi(\mathbf{y}) = q(\mathbf{Y}\mathbf{d}+\mathbf{Z}\mathbf{y})$, οπότε έχουμε $\nabla\psi(\mathbf{y}) = \mathbf{Z}^T\nabla q(\mathbf{x}) = \mathbf{Z}^T(\mathbf{G}\mathbf{x}+\mathbf{b})$ και $\nabla^2\psi(\mathbf{y}) = \mathbf{Z}^T\nabla^2 q(\mathbf{x})\mathbf{Z}=\mathbf{Z}^T\mathbf{G}\mathbf{Z}$.

Πολλαπλασιάζοντας την $\mathbf{g}^* = \mathbf{A}\boldsymbol{\lambda}^*$ με $\mathbf{Y}^T$ έχουμε

$$\boldsymbol{\lambda}^* = \mathbf{Y}^T\mathbf{g}^* = \mathbf{Y}^T(\mathbf{G}\mathbf{x}^*+\mathbf{b}) \qquad (7.1.14)$$



Σχήμα 7.1.2: Γενικευμένη απαλοιφή.

εξής υπολογίζονται οι πολλαπλασιαστές Lagrange. Συγκεντρωτικά έχουμε, απαλείφοντας το **y** από τις ως άνω εξισώσεις, την λύση

$$\mathbf{x}^* = \mathbf{Yd} - \mathbf{Z}(\mathbf{Z}^T\mathbf{GZ})^{-1}\,\mathbf{Z}^T(\mathbf{b}+\mathbf{GYd}) \qquad (7.1.15)$$

$$\boldsymbol{\lambda}^* = \mathbf{Y}^T(\mathbf{I}-\mathbf{GZ}(\mathbf{Z}^T\mathbf{GZ})^{-1}\mathbf{Z}^T)\mathbf{b} + \mathbf{Y}^T\mathbf{G}(\mathbf{I}-\mathbf{Z}(\mathbf{Z}^T\mathbf{GZ})^{-1}\mathbf{Z}^T\mathbf{G})\mathbf{Yd}. \qquad (7.1.16)$$

Υπάρχουν μια σειρά μεθόδων για την επιλογή των **Y** και **Z**. Για παράδειγμα, μια QR παραγοντοποίηση του πίνακα **A** δίδει

$$\mathbf{A} = \mathbf{Q}\begin{bmatrix}\mathbf{R}\\ \mathbf{0}\end{bmatrix} = [\mathbf{Q}_1\ \mathbf{Q}_2]\begin{bmatrix}\mathbf{R}\\ \mathbf{0}\end{bmatrix} = \mathbf{Q}_1\mathbf{R} \qquad (7.1.17)$$

όπου ο $\mathbf{Q}\in R^{nxn}$ είναι ορθοκανονικός (δηλαδή $\mathbf{Q}^T\mathbf{Q}=\mathbf{I}$) και ο $\mathbf{R}\in R^{mxm}$ είναι άνω τριγωνικός πλήρους βαθμού. Τότε για τους $\mathbf{Q}_1\in R^{nxm}$, $\mathbf{Q}_2\in R^{nx(n-m)}$ ισχύουν προφανώς $\mathbf{Q}_1^T\mathbf{Q}_1=\mathbf{I}$, $\mathbf{Q}_2^T\mathbf{Q}_2=\mathbf{I}$ και $\mathbf{Q}_1^T\mathbf{Q}_2=\mathbf{0}$, οπότε η επιλογή

$$\mathbf{Y}=\mathbf{A}^{+T} = \mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1} = \mathbf{Q}_1\mathbf{R}(\mathbf{R}^T\mathbf{R})^{-1}= \mathbf{Q}_1\mathbf{R}^{-T}, \quad \mathbf{Z}=\mathbf{Q}_2 \qquad (7.1.18)$$

εκπληροί τις ζητούμενες ιδιότητες των **Y**, **Z**, αφού $\mathbf{Y}^T\mathbf{A} = \mathbf{I}$ και $\mathbf{Z}^T\mathbf{A} = \mathbf{0}$. Στην περίπτωση αυτή το διάνυσμα **Yd** στην (7.1.11) και στο Σχ. 7.1.2 είναι ορθογώνιο στον περιορισμό και οι κατευθύνσεις μειωθεισών συντεταγμένων $\mathbf{z}_i$ είναι αμοιβαίως ορθογώνιες. Αυτή είναι η **μέθοδος ορθογώνιας παραγοντοποίησης** των Gill και Murray (1974).

Παράδειγμα 7.1.2: Θεωρούμε το πρόβλημα του Παραδ. 7.1.1. Η QR παραγοντοποίηση του **A** είναι

$$\mathbf{A} = \begin{bmatrix} 1/\sqrt{6} & 4/\sqrt{21} & 1/\sqrt{14} \\ 2/\sqrt{6} & -1/\sqrt{21} & -2/\sqrt{14} \\ -1/\sqrt{6} & 2/\sqrt{21} & -3/\sqrt{14} \end{bmatrix} \begin{bmatrix} \sqrt{6} & -\sqrt{6}/3 \\ 0 & -\sqrt{6}/3 \\ 0 & 0 \end{bmatrix}$$

εξού

$$\mathbf{Y} = \frac{1}{14}\begin{bmatrix} 5 & 8 \\ 4 & -2 \\ -1 & 4 \end{bmatrix} \quad , \quad \mathbf{Z} = \frac{1}{\sqrt{14}}\begin{bmatrix} 1 \\ -2 \\ -3 \end{bmatrix}. \qquad (7.1.19)$$

Άρα έχουμε στην (7.1.13) $\mathbf{Yd} = \frac{1}{7}[2\ \ 10\ \ -6]^T$ και από $\mathbf{b}=\mathbf{0}$ και $\mathbf{GYd} = 2\mathbf{Yd}$ συνεπάγεται

$\mathbf{Z}^T(\mathbf{b}+\mathbf{GYd}) = \mathbf{0}$, άρα $\mathbf{y}^* = \mathbf{0}$ και $\mathbf{x}^* = \mathbf{Yd} = \frac{1}{7}[2\ 10\ -6]^T$, $\mathbf{g}^*=\mathbf{b}+\mathbf{GYd}=\frac{2}{7}[2\ \ 10\ -6]^T$, $\boldsymbol{\lambda}^* = [8/7\ \ -4/7]^T$, αποτελέσματα ταυτόσημα με εκείνα του Παραδ. 7.1.1. ■

Ένας γενικός τρόπος προσδιορισμού κατάλληλων **Y**, **Z** είναι μέσω επιλογής τυχαίου πίνακα $\mathbf{V}\in R^{n\times(n-m)}$ έτσι ώστε ο τετραγωνικός πίνακας [**A V**] να είναι κανονικός. Παίρνουμε τότε τον αντίστροφο

$$[\mathbf{A}\ \ \mathbf{V}]^{-1} = \begin{bmatrix} \mathbf{Y}^T \\ \mathbf{Z}^T \end{bmatrix} \tag{7.1.20}$$

και έχουμε τους **Y**, **Z** που εκπληρούν τις ζητούμενες ιδιότητες, αφού από την (7.1.20) έχουμε $\mathbf{I}=[\mathbf{Y}\ \mathbf{Z}]^T[\mathbf{A}\ \mathbf{V}]$ και άρα $\mathbf{Y}^T\mathbf{A} = \mathbf{I}$ και $\mathbf{Z}^T\mathbf{A} = \mathbf{0}$. Όλες οι μέθοδοι προσδιορισμού κατάλληλων πινάκων **Y**, **Z** μπορούν να θεωρηθούν ειδικές περιπτώσεις αυτού του γενικού τρόπου, καθότι, όπως είναι προφανές από την (7.1.20), για κάθε επιλογή **Y**, **Z** υπάρχει ένας αντίστοιχος μοναδικός **V**. Αν, επί παραδείγματι, επιλεγεί

$$\mathbf{V} = \begin{bmatrix} \mathbf{0} \\ \mathbf{I} \end{bmatrix} \tag{7.1.21}$$

τότε μπορεί εύκολα να πιστοποιηθεί ότι η (7.1.20) οδηγεί στη μέθοδο άμεσης απαλοιφής των $x_1, \ldots, x_m$, αφού η ταυτότητα

$$\begin{bmatrix} \mathbf{A}_1 & \mathbf{0} \\ \mathbf{A}_2 & \mathbf{I} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{A}_1^{-1} & \mathbf{0} \\ -\mathbf{A}_2\mathbf{A}_1^{-1} & \mathbf{I} \end{bmatrix} = \begin{bmatrix} \mathbf{Y}^T \\ \mathbf{Z}^T \end{bmatrix} \tag{7.1.22}$$

καθορίζει τους **Y**, **Z**, δηλαδή

$$\mathbf{Y} = \begin{bmatrix} \mathbf{A}_1^{-T} \\ \mathbf{0} \end{bmatrix},\ \mathbf{Z} = \begin{bmatrix} -\mathbf{A}_1^{-T}\mathbf{A}_2^T \\ \mathbf{I} \end{bmatrix}$$

που όντως οδηγούν από την (7.1.11) στην άμεση απαλοιφή (7.1.3). Εξ άλλου, με την επιλογή

$$\mathbf{V} = \mathbf{Q}_2 \tag{7.1.23}$$

με $\mathbf{Q}_2$ από την (7.1.17), οδηγούμεθα στη μέθοδο ορθογώνιας παραγοντοποίησης.

7.2 Η Μέθοδος Lagrange

Μια εναλλακτική μέθοδος προσδιορισμού της λύσης $\mathbf{x}^*$ του (7.1.2) και των αντίστοιχων πολλαπλασιαστών Lagrange $\boldsymbol{\lambda}^*$ είναι η **μέθοδος πολλαπλασιαστών Lagrange** (6.1.5). Η Λαγκρανζιανή του προβλήματος (7.1.2) είναι

$$L(\mathbf{x},\boldsymbol{\lambda}) = \frac{1}{2}\mathbf{x}^T\mathbf{G}\mathbf{x} + \mathbf{b}^T\mathbf{x} - \boldsymbol{\lambda}^T(\mathbf{A}^T\mathbf{x}-\mathbf{d}) \qquad (7.2.1)$$

και από την (6.1.7) έχουμε

$$\nabla_{\mathbf{x}}L = \mathbf{0} \Rightarrow \mathbf{G}\mathbf{x} + \mathbf{b} - \mathbf{A}\boldsymbol{\lambda} = \mathbf{0}$$
$$\nabla_{\boldsymbol{\lambda}}L = \mathbf{0} \Rightarrow \mathbf{A}^T\mathbf{x} - \mathbf{d} = \mathbf{0}$$

ή το γραμμικό σύστημα

$$\begin{bmatrix} \mathbf{G} & -\mathbf{A} \\ -\mathbf{A}^T & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \boldsymbol{\lambda} \end{bmatrix} = -\begin{bmatrix} \mathbf{b} \\ \mathbf{d} \end{bmatrix} \qquad (7.2.2)$$

όπου ο παραπάνω πίνακας αναφέρεται ως **Λαγκρανζιανός πίνακας** και είναι συμμετρικός αλλά όχι θετικά ορισμένος. Επειδή η αντιστροφή του Λαγκρανζιανού πίνακα προξενεί σε πολλά προβλήματα αριθμητικές δυσκολίες, αλλά και για να αποφευχθεί ο αντίστοιχος υπολογιστικός φόρτος σε προβλήματα υψηλών διαστάσεων, έχουν αναπτυχθεί πολλές μέθοδοι επίλυσης του (7.2.2) που αποφεύγουν την αντιστροφή του Λαγκρανζιανού πίνακα.

Παράδειγμα 7.2.1: Θεωρούμε και πάλι το πρόβλημα του Παραδ. 7.1.1 και έχουμε την Λαγκρανζιανή

$$L(\mathbf{x},\boldsymbol{\lambda}) = x_1^2+x_2^2+x_3^2 - \lambda_1(x_1+2x_2-x_3-4) - \lambda_2(x_1-x_2+x_3+2)$$

και με την (6.1.7)

$$\partial L/\partial x_1 = 2x_1-\lambda_1-\lambda_2 = 0$$
$$\partial L/\partial x_2 = 2x_2 - 2\lambda_1+\lambda_2 = 0$$
$$\partial L/\partial x_3 = 2x_3 + \lambda_1-\lambda_2 = 0$$
$$\partial L/\partial \lambda_1 = x_1+2x_2-x_3-4 = 0$$
$$\partial L/\partial \lambda_2 = x_1-x_2+x_3+2 = 0$$

ή το γραμμικό σύστημα

$$\begin{bmatrix} 2 & 0 & 0 & -1 & -1 \\ 0 & 2 & 0 & -2 & 1 \\ 0 & 0 & 2 & 1 & -1 \\ -1 & -2 & 1 & 0 & 0 \\ -1 & 1 & -1 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \lambda_1 \\ \lambda_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ -4 \\ 2 \end{bmatrix}$$

με μοναδική λύση $x_1^* = 2/7$, $x_2^* = 10/7$, $x_3^* = -6/7$, $\lambda_1^* = 8/7$, $\lambda_2^* = -4/7$, ταυτόσημη με εκείνη του Παραδ. 7.1.1.

■

7.3 Μέθοδοι Ενεργού Συνόλου

Θα εξετάσουμε τώρα την εφαρμογή μεθόδων ΤΠ με περιορισμούς ισότητας, στην επίλυση προβλημάτων ΤΠ με περιορισμούς ανισότητας μέσω της **μεθόδου ενεργού συνόλου**. Για να αποφύγουμε παρουσιαζόμενες δυσκολίες, θα υποθέσουμε $\mathbf{G}>\mathbf{0}$, οπότε κάθε τοπικό ελάχιστο είναι μοναδικό απόλυτο ελάχιστο, η μέθοδος όμως, ενδεχόμενα με μερικές μικρές προσθήκες, μπορεί να εφαρμοσθεί και σε προβλήματα που δεν εκπληρούν αυτήν την παραδοχή.

Η μέθοδος ενεργού συνόλου επιδρά επί των εκάστοτε ενεργών περιορισμών του ενεργού συνόλου $\mathcal{A}$, θεωρώντας τους ως περιορισμούς ισότητας, μεταβάλει όμως από επανάληψη σε επανάληψη το περιεχόμενο του $\mathcal{A}$ ώστε να καταλήξει τελικά στο σωστό ενεργό σύνολο $\mathcal{A}^*$ και την τελική λύση του προβλήματος. Έτσι, στην επανάληψη k, βρισκόμαστε σε κάποιο επιτρεπτό σημείο $\mathbf{x}^{(k)}$ που ικανοποιεί $\mathbf{a}_i^T\mathbf{x}^{(k)}=d_i$, $\forall i\in\mathcal{A}^{(k)}$, και $\mathbf{a}_i^T\mathbf{x}^{(k)}>d_i$, $\forall i\notin\mathcal{A}^{(k)}$. Κάθε επανάληψη k επιχειρεί να προσδιορίσει την λύση ενός ΤΠ προβλήματος ισοτήτων (ΠΙ) όπου περιλαμβάνονται μόνο περιορισμοί ισότητας, δηλαδή μόνον οι περιορισμοί που εμπεριέχονται στο $\mathcal{A}^{(k)}$. Αυτό γίνεται μετασχηματίζοντας το αρχικό πρόβλημα γύρω από το $\mathbf{x}^{(k)}$ μέσω μιας διόρθωσης $\boldsymbol{\delta}$. Το μετασχηματισθέν ΠΙ λαμβάνεται αντικαθιστώντας στο (7.1.1) $\mathbf{x}=\mathbf{x}^{(k)}+\boldsymbol{\delta}$ και είναι

$$\begin{aligned} &\frac{1}{2}\boldsymbol{\delta}^T\mathbf{G}\boldsymbol{\delta} + \boldsymbol{\delta}^T\mathbf{g}^{(k)} \to \min_{\delta} \\ &\text{υπό} \quad \mathbf{a}_i^T\boldsymbol{\delta}=0,\ i\in\mathcal{A}^{(k)} \end{aligned} \tag{7.3.1}$$

όπου $\mathbf{g}^{(k)} = \mathbf{b}+\mathbf{G}\mathbf{x}^{(k)}$. Το (7.3.1) μπορεί προφανώς να επιλυθεί με μια μέθοδο των Κεφαλαίων 7.1, 7.2. Αν η λύση $\boldsymbol{\delta}^{(k)}$ είναι επιτρεπτή ως προς τους περιορισμούς εκτός $\mathcal{A}^{(k)}$, τότε το επόμενο επαναληπτικό είναι $\mathbf{x}^{(k+1)}=\mathbf{x}^{(k)}+\boldsymbol{\delta}^{(k)}$. Αν όχι, επιχειρείται αναζήτηση επί γραμμής στην κατεύθυνση του $\boldsymbol{\delta}^{(k)}$ για να προσδιορισθεί το βέλτιστο επιτρεπτό σημείο επ' αυτής. Αυτό μπορεί να επιτευχθεί ορίζοντας τη λύση του (7.3.1) ως κατεύθυνση αναζήτησης $\mathbf{s}^{(k)}$ και επιλέγοντας το βήμα $\alpha^{(k)}$ ως το μέγιστο δυνατό, μέχρι την ενεργοποίηση κάποιου περιορισμού

$i \notin \mathcal{A}^{(k)}$. Ένας περιορισμός $i \notin \mathcal{A}^{(k)}$, δηλαδή $\mathbf{a}_i^T \mathbf{x}^{(k)} > d_i$, προφανώς ενεργοποιείται στην κατεύθυνση $\mathbf{x}^{(k)} + \alpha \mathbf{s}^{(k)}$ όταν $\mathbf{a}_i^T (\mathbf{x}^{(k)} + \alpha_i \mathbf{s}^{(k)}) = d_i$, δηλαδή για βήμα $\alpha_i = (d_i - \mathbf{a}_i^T \mathbf{x}^{(k)}) / \mathbf{a}_i^T \mathbf{s}^{(k)}$. Άρα το μέγιστο δυνατό βήμα $\alpha^{(k)}$ είναι

$$\alpha^{(k)} = \min \{1, \min_{\substack{i \notin \mathcal{A}^{(k)} \\ \mathbf{a}_i^T \mathbf{s}^{(k)} < 0}} \frac{d_i - \mathbf{a}_i^T \mathbf{x}^{(k)}}{\mathbf{a}_i^T \mathbf{s}^{(k)}}\} \qquad (7.3.2)$$

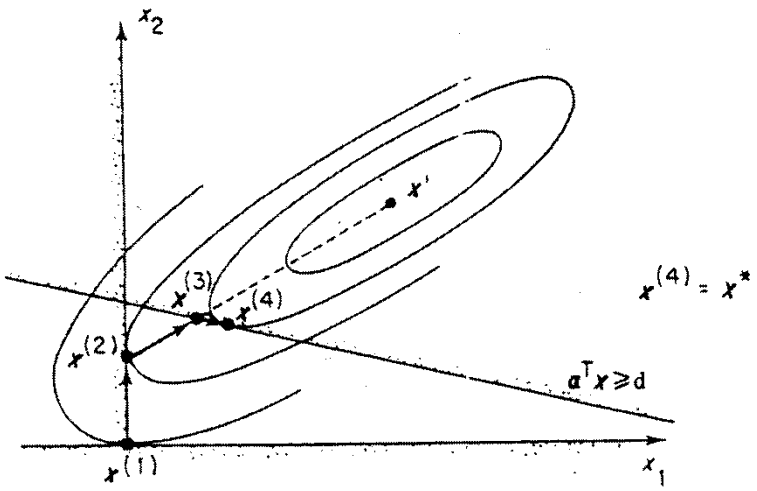
και επομένως $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \alpha^{(k)} \mathbf{s}^{(k)}$. Ο προσδιορισμός του $\alpha^{(k)}$ κατά την (7.3.2) γίνεται μόνο για εκείνα τα i, για τα οποία $\mathbf{a}_i^T \mathbf{s}^{(k)} < 0$· διότι όταν $\mathbf{a}_i^T \mathbf{s}^{(k)} > 0$, και επειδή εξ ορισμού $\mathbf{a}_i^T \mathbf{x}^{(k)} > d_i$, $\forall i \notin \mathcal{A}^{(k)}$, τότε σίγουρα $\mathbf{a}_i^T \mathbf{x}^{(k+1)} > d_i$ για κάθε $\alpha > 0$, δηλαδή δεν πρόκειται να υπάρξει ενεργοποίηση των περιορισμών αυτών στην κατεύθυνση $\mathbf{s}^{(k)}$. Αν η (7.3.2) δώσει $\alpha^{(k)} < 1$, τότε ενεργοποιείται ένας νέος περιορισμός, έστω ο p, που εισάγεται στο $\mathcal{A}$.

Αν το $\mathbf{x}^{(k)}$ (δηλαδή $\boldsymbol{\delta}^{(k)} = \mathbf{0}$) είναι η λύση του τρέχοντος ΠΙ (7.3.1), τότε μπορούν να υπολογισθούν πολλαπλασιαστές Lagrange $\boldsymbol{\lambda}^{(k)}$ κατά τα Κεφάλαια 7.1, 7.2, και τα $\mathbf{x}^{(k)}$, $\boldsymbol{\lambda}^{(k)}$ ικανοποιούν όλες τις αναγκαίες συνθήκες πρώτου βαθμού (6.2.5) για το αρχικό πρόβλημα, εκτός αν υπάρχουν κάποιοι $\lambda_i^{(k)} < 0$, $i \in I$. Ελέγχουμε λοιπόν αν $\lambda_i^{(k)} \geq 0$ $\forall i \in \mathcal{A}^{(k)} \cap I$. Αν ναι, το αρχικό πρόβλημα έχει επιλυθεί. Άλλως, υπάρχει ένας τουλάχιστον δείκτης $q \in \mathcal{A}^{(k)} \cap I$ με $\lambda_q^{(k)} < 0$, οπότε, σύμφωνα με το Κεφάλαιο 6.2, η τιμή της $q(\mathbf{x})$ μπορεί να ελαττωθεί αν ο περιορισμός q απενεργοποιηθεί. Έτσι απομακρύνουμε τον περιορισμό q από το $\mathcal{A}^{(k)}$ και ο αλγόριθμος συνεχίζει στην επόμενη επανάληψη επιχειρώντας τη λύση του νέου διαμορφωθέντος ΠΙ. Αν υπάρχουν περισσότεροι του ενός δείκτες με $\lambda_i^{(k)} < 0$, μπορεί να επιλεγεί ο δείκτης με την μικρότερη τιμή πολλαπλασιαστού, δηλαδή ο επιλεγόμενος δείκτης q ικανοποιεί

$$\min_{i \in \mathcal{A}^{(k)} \cap I} \lambda_i^{(k)}. \qquad (7.3.3)$$

Συμπερασματικά, έχουμε τον ακόλουθο αλγόριθμο που εκκινεί από ένα δεδομένο επιτρεπτό σημείο εκκίνησης $\mathbf{x}^{(1)}$ και το αντίστοιχο $\mathcal{A}^{(1)}$:

a) Δίδεται από τον χρήστη επιτρεπτό σημείο εκκίνησης $\mathbf{x}^{(1)}$ και αντίστοιχο $\mathcal{A}^{(1)}$, k=1.
b) Αν $\boldsymbol{\delta} = \mathbf{0}$ δεν επιλύει το ΠΙ (7.3.1), πήγαινε στο (d).
c) Υπολογισμός $\boldsymbol{\lambda}^{(k)}$ και επιλογή $\lambda_q^{(k)}$ από (7.3.3). Αν $\lambda_q^{(k)} \geq 0$, η λύση είναι $\mathbf{x}^* = \mathbf{x}^{(k)}$, άλλως απομακρύνεται ο περιορισμός q από το $\mathcal{A}^{(k)}$.
d) Επίλυση του ΠΙ (7.3.1) για προσδιορισμό της $\mathbf{s}^{(k)}$.
e) Υπολογισμός $\alpha^{(k)}$ κατά (7.3.2) και $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \alpha^{(k)} \mathbf{s}^{(k)}$.
f) Αν $\alpha^{(k)} < 1$, προσθήκη του ενεργοποιούμενου περιορισμού p στο $\mathcal{A}^{(k)}$.
g) k:=k+1, πήγαινε στο (b).



Σχήμα 7.3.1: Η μέθοδος ενεργού συνόλου για το Παραδ. 7.3.1.

Μπορεί να αποδειχθεί ότι ο αλγόριθμος τερματίζει σε πεπερασμένο αριθμό επαναλήψεων, πλην ορισμένων εκφυλισμένων περιπτώσεων που απαιτούν αντίστοιχες επεκτάσεις.

Παράδειγμα 7.3.1: Θεωρούμε ένα πρόβλημα ΤΠ δύο μεταβλητών, με περιορισμούς $\mathbf{x}\geq\mathbf{0}$ και $\mathbf{a}^T\mathbf{x}\geq d$ κατά το Σχ. 7.3.1. Στο σημείο εκκίνησης $\mathbf{x}^{(1)}=\mathbf{0}$ οι περιορισμοί $x_1\geq 0$ και $x_2\geq 0$ είναι ενεργοί και το $\mathbf{x}^{(1)}$ (δηλαδή $\boldsymbol{\delta}=\mathbf{0}$) επιλύει το αντίστοιχο ΠΙ (7.3.2). Ο υπολογισμός του $\boldsymbol{\lambda}^{(1)}$ δείχνει ότι ο περιορισμός $x_2\geq 0$ έχει αρνητικό πολλαπλασιαστή και έτσι απενεργοποιείται για την επόμενη επανάληψη, ενώ ο $x_1\geq 0$ παραμένει ενεργός, έχουμε δηλαδή $\mathcal{A}^{(1)} = \{1\}$. Το νέο ΠΙ επιλύεται οδηγώντας στο $\mathbf{x}^{(2)}$ (ή στο $\boldsymbol{\delta}^{(1)}=\mathbf{x}^{(2)}-\mathbf{x}^{(1)}$) που είναι επιτρεπτό και άρα γίνεται η αφετηρία της επόμενης (2ης) επανάληψης. Ο υπολογισμός του $\boldsymbol{\lambda}^{(2)}$ δείχνει ότι ο περιορισμός $x_1\geq 0$ έχει αρνητικό πολλαπλασιαστή, άρα απενεργοποιείται και αυτός και έχουμε $\mathcal{A}^{(2)}=\{\}$. Το νέο ΠΙ (7.3.1) επιλύεται, όμως η λύση $\mathbf{x}'$ δεν είναι επιτρεπτή, άρα υπολογίζεται η κατεύθυνση $\mathbf{s}^{(2)} = \mathbf{x}'-\mathbf{x}^{(2)}$ και από την (7.3.2) καταλήγουμε στο βέλτιστο επιτρεπτό $\mathbf{x}^{(3)}$. Ο περιορισμός $\mathbf{a}^T\mathbf{x}\geq d$ ενεργοποιείται στο $\mathbf{x}^{(3)}$ και επομένως εισάγεται στο $\mathcal{A}$. Το $\mathbf{x}^{(3)}$ δεν επιλύει το νέο ΠΙ, επομένως δεν υπολογίζονται πολλαπλασιαστές, αλλά επιλύεται το νέο ΠΙ που δίδει $\mathbf{x}^{(4)}$ στην επόμενη επανάληψη. Ο πολλαπλασιαστής για το σημείο αυτό υπολογίζεται θετικός, άρα $\mathbf{x}^*=\mathbf{x}^{(4)}$ και ο αλγόριθμος τερματίζει.

■

Αν ο χρήστης του αλγορίθμου δεν γνωρίζει ένα επιτρεπτό σημείο εκκίνησης $\mathbf{x}^{(1)}$, μπορεί να επιλύσει καταρχήν ένα βοηθητικό πρόβλημα, του οποίου η λύση είναι ένα επιτρεπτό σημείο του αρχικού προβλήματος. Ένα τέτοιο βοηθητικό πρόβλημα είναι επί παραδείγματι το (9.1.4).

Έχουν αναπτυχθεί και άλλοι αριθμητικοί αλγόριθμοι επίλυσης γενικών προβλημάτων ΤΠ, μεταξύ αυτών και κατάλληλες επεκτάσεις του αλγόριθμου Simplex Γραμμικού Προγραμματισμού, που όμως δεν θα εκθέσουμε εδώ. Επίσης υπάρχουν ειδικοί αλγόριθμοι για

την αποτελεσματικότερη λύση ειδικών προβλημάτων ΤΠ, π.χ. προβλημάτων που εμπεριέχουν μόνο περιορισμούς του τύπου $l_i \le x_i \le u_i$, i=1, ... ,n, κλπ.

Προβλήματα

7.1 Θεωρούμε το ακόλουθο πρόβλημα ΤΠ

$$x_1^2 - x_1x_2 + x_2^2 - x_2x_3 + x_3^2 + 2x_1 - x_2 \to \underset{\mathbf{x}}{\text{Min}}$$

υπό $3x_1 - x_2 + x_3 = 0$

$2x_1 - x_2 - x_3 = 0.$

Επιλέξτε τον πίνακα **V** και υπολογίστε τους πίνακες **Y**, **Z** που οδηγούν στην άμεση απαλοιφή των x_1, x_2 και επιλύστε το προκύπτον πρόβλημα χωρίς περιορισμούς. Υπολογίστε τους πολλαπλασιαστές **λ***.

Λύση

$$\mathbf{A} = \begin{bmatrix} 3 & 2 \\ -1 & -1 \\ 1 & -1 \end{bmatrix}, \mathbf{d} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \mathbf{G} = \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix}, \mathbf{b} = \begin{bmatrix} 2 \\ -1 \\ 0 \end{bmatrix}.$$

Κατά (7.1.21): $\mathbf{V} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$.

Κατά (7.1.20):

$$[\mathbf{A} \quad \mathbf{V}]^{-1} = \begin{bmatrix} 1 & 2 & 0 \\ -1 & -3 & 0 \\ -2 & -5 & 1 \end{bmatrix} = \begin{bmatrix} \mathbf{Y}^T \\ \mathbf{Z}^T \end{bmatrix}, \text{ άρα } \mathbf{Y} = \begin{bmatrix} 1 & -1 \\ 2 & -3 \\ 0 & 0 \end{bmatrix}, \mathbf{Z} = \begin{bmatrix} -2 \\ -5 \\ 1 \end{bmatrix}.$$

Από (7.1.11): $\mathbf{x} = \mathbf{Yd} + \mathbf{Z}y = \begin{bmatrix} -2 \\ -5 \\ 1 \end{bmatrix} y$ ικανοποιεί τους περιορισμούς.

Από (7.1.12) (εδώ: $\mathbf{d} = \mathbf{0}$): $\psi(y) = \frac{1}{2} y^2 \mathbf{Z}^T\mathbf{GZ} + \mathbf{b}^T\mathbf{Z}y$

$\mathbf{Z}^T\mathbf{GZ} = 50 > 0$, $\mathbf{b}^T\mathbf{Z} = 1 \Rightarrow \psi(y) = 25y^2 + y$

(7.1.13): $50y = -1 \Rightarrow y^* = -0.02 \Rightarrow \mathbf{x}^* = \begin{bmatrix} -2 \\ -5 \\ 1 \end{bmatrix}$ $y^* = \begin{bmatrix} 0.04 \\ 0.1 \\ -0.02 \end{bmatrix}$

(7.1.14): $\boldsymbol{\lambda}^* = \begin{bmatrix} 1 & 2 & 0 \\ -1 & -3 & 0 \end{bmatrix}(\mathbf{G}\mathbf{x}^* + \mathbf{b}) = \begin{bmatrix} 0.34 \\ 0.48 \end{bmatrix}$.

7.2 Επιλύστε το Πρόβλημα 7.1 μέσω της μεθόδου Lagrange.

Λύση

$L(\mathbf{x},\boldsymbol{\lambda}) = x_1^2 - x_1x_2 + x_2^2 - x_2x_3 + x_3^2 + 2x_1 - x_2 - \lambda_1(3x_1 - x_2 + x_3) - \lambda_2(2x_1 - x_2 - x_3)$.

Λύση κατά (7.2.2) από

$$\begin{bmatrix} 2 & -1 & 0 & -3 & -2 \\ -1 & 2 & -1 & 1 & 1 \\ 0 & -1 & 2 & -1 & 1 \\ -3 & 1 & -1 & 0 & 0 \\ -2 & 1 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \lambda_1 \\ \lambda_2 \end{bmatrix} = \begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} \Leftrightarrow \begin{bmatrix} x_1^* \\ x_2^* \\ x_3^* \\ \lambda_1^* \\ \lambda_2^* \end{bmatrix} = \begin{bmatrix} 0.04 \\ 0.1 \\ -0.02 \\ 0.34 \\ 0.48 \end{bmatrix}$$

7.3 Ζητείται η ελαχιστοποίηση της αντικειμενικής συνάρτησης του Προβλήματος 7.1 υπό τους ίδιους περιορισμούς, λαμβανομένους όμως τώρα σαν ανισότητες
$3x_1 - x_2 + x_3 \geq 0$
$2x_1 - x_2 - x_3 \geq 0$
και επιπλέον, $x_1 \geq 0$, $x_2 \geq 0$.

Λύση

Αφού στη λύση του Προβλήματος 7.1 $x_1^* = 0.04 > 0, x_2^* = 0.1 > 0,$ $\lambda_1^* = 0.34 > 0$ και $\lambda_2^* = 0.48 > 0,$ ικανοποιούνται όλες οι αναγκαίες συνθήκες (6.2.5) και η λύση είναι η ίδια.

7.4 Εκκινώντας από $\mathbf{x}^{(1)} = \mathbf{0}$, να εφαρμοσθούν τα βήματα του αλγορίθμου ενεργού συνόλου για επίλυση του προβλήματος

$$q(\mathbf{x}) = x_1^2 - x_1x_2 + x_2^2 - 3x_1 \rightarrow \underset{\mathbf{x}}{\text{Min}}$$

υπό $x_1 \geq 0$, $x_2 \geq 0$, $-x_1 - x_2 \geq -2$.

Λύση

$\mathbf{G} = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} > \mathbf{0}$, $\mathbf{a}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ $\mathbf{a}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$, $\mathbf{a}_3 = \begin{bmatrix} -1 \\ -1 \end{bmatrix}$, $d_1 = d_2 = 0$, $d_3 = -2$

(a) $\mathcal{A}^{(1)} = \{1, 2\}$, <u>k = 1</u>.

(b) $\mathbf{x}^{(1)} = \mathbf{0}$ επιλύει το ΠΙ.

(c) $\mathbf{g}^{(1)} = \mathbf{A}\boldsymbol{\lambda}^{(1)} \Leftrightarrow \begin{bmatrix} -3 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \lambda_1^{(1)} \\ \lambda_2^{(1)} \end{bmatrix} \Leftrightarrow \lambda_1^{(1)} = -3,\ \lambda_2^{(1)} = 0 \Rightarrow q = 1$

$\Rightarrow \mathcal{A}^{(1)} := \mathcal{A}^{(1)} - \{1\} = \{2\}$.

(d) ΠΙ: $x_1^2 - x_1x_2 + x_2^2 - 3x_1 \to \underset{\mathbf{x}}{\text{Min}}$ υπό $x_2 = 0$

Λύση: $x_1 = 1.5$, $x_2 = 0$, $\lambda_2 = -1.5$

(e) επιτρεπτή, άρα $\alpha^{(1)} = 1$ και $\mathbf{x}^{(2)} = \begin{bmatrix} 1.5 \\ 0 \end{bmatrix}$.

(g) <u>k = 2</u>

(b) $\boldsymbol{\delta} = \mathbf{0}$ επιλύει το ΠΙ.

(c) $\lambda_2^{(2)} = -1.5 \Rightarrow q = 2 \Rightarrow \mathcal{A}^{(2)} := \mathcal{A}^{(2)} - \{2\} = \{\}$.

(d) ΠΙ: $x_1^2 - x_1x_2 + x_2^2 - 3x_1 \to \underset{\mathbf{x}}{\text{Min}}$ (χωρίς περιορισμούς)

Λύση: $x_1 = 2$, $x_2 = 1 \Rightarrow \mathbf{s}^{(2)} = \begin{bmatrix} 2 \\ 1 \end{bmatrix} - \begin{bmatrix} 1.5 \\ 0 \end{bmatrix} = \begin{bmatrix} 0.5 \\ 1 \end{bmatrix}$

(e) μη επιτρεπτή, άρα

$\alpha^{(2)} = \min\{1, \dfrac{-2+1.5}{-1.5}\} = \dfrac{1}{3} \Rightarrow \mathbf{x}^{(3)} = \begin{bmatrix} 1.5 \\ 0 \end{bmatrix} + \begin{bmatrix} 1/6 \\ 1/3 \end{bmatrix} = \begin{bmatrix} 5/3 \\ 1/3 \end{bmatrix}$.

(f) $\mathcal{A}^{(2)} := \mathcal{A}^{(2)} \cup \{3\} = \{3\}$.

(g) <u>k = 3</u>

(b) $\boldsymbol{\delta} = \mathbf{0}$ δεν επιλύει το ΠΙ.

(d) ΠΙ: $x_1^2 - x_1x_2 + x_2^2 - 3x_1 \to \underset{\mathbf{x}}{\text{Min}}$ υπό $-x_1 - x_2 = -2$

Λύση: $x_1 = 1.5$, $x_2 = 0.5$, $\lambda_3 = 0.5$

(e) επιτρεπτή, άρα $\alpha^{(3)} = 1$ και $\mathbf{x}^{(4)} = \begin{bmatrix} 1.5 \\ 0.5 \end{bmatrix}$.

(g) <u>k = 4</u>

(b) $\boldsymbol{\delta} = \mathbf{0}$ επιλύει το ΠΙ.

(c) $\lambda_3^{(4)} = 0.5 > 0$, άρα τερματισμός με $\mathbf{x}^* = \mathbf{x}^{(4)}$.

8. ΒΕΛΤΙΣΤΟΠΟΙΗΣΗ ΜΕ ΓΡΑΜΜΙΚΟΥΣ ΠΕΡΙΟΡΙΣΜΟΥΣ

Στο κεφάλαιο αυτό θεωρούμε προβλήματα με γενικά μη γραμμική αντικειμενική συνάρτηση και γραμμικούς περιορισμούς:

$$\begin{aligned} &\text{Ελαχιστοποίηση } f(\mathbf{x}) \\ &\text{υπό } \mathbf{a}_i^T\mathbf{x} = d_i, \quad i\in E \\ &\qquad \mathbf{a}_i^T\mathbf{x} \geq d_i, \quad i\in I. \end{aligned} \tag{8.1.1}$$

Οι μέθοδοι χειρισμού των γραμμικών περιορισμών είναι παρόμοιες με εκείνες του ΤΠ του Κεφαλαίου 7, αλλά η μη τετραγωνική αντικειμενική συνάρτηση επιφέρει μερικές επιπλέον δυσκολίες. Μιά απ' αυτές είναι ότι το πρόβλημα δεν μπορεί πλέον να επιλυθεί σε πεπερασμένο αριθμό επαναλήψεων, και έτσι η λύση $\mathbf{x}^*$ είναι το ασυμπτωτικό όριο μιας επαναληπτικής ακολουθίας $\{\mathbf{x}^{(k)}\}$ όπως και στις αριθμητικές μεθόδους του Πρώτου Μέρους.

8.1 Περιορισμοί Ισότητας

Θα θεωρήσουμε καταρχήν το πρόβλημα με περιορισμούς ισότητας μόνον

$$\begin{aligned} &f(\mathbf{x}) \rightarrow \underset{\mathbf{x}}{\text{Min}} \\ &\text{υπό } \mathbf{A}^T\mathbf{x} = \mathbf{d} \end{aligned} \tag{8.1.2}$$

όπου ο πίνακας $\mathbf{A}\in R^{n\text{x}m}$ είναι πλήρους βαθμού m και $\mathbf{d}\in R^m$. Η **μέθοδος γενικευμένης απαλοιφής** του Κεφαλαίου 7.1 μπορεί και εδώ να οδηγήσει σε ένα απεριόριστο πρόβλημα μέσω πινάκων **Y**, **Z** έτσι ώστε $\mathbf{Y}^T\mathbf{A}=\mathbf{I}$, $\mathbf{Z}^T\mathbf{A}=\mathbf{0}$ και [**Y Z**] κανονικός.

Σε προβλήματα ΤΠ (Κεφάλαιο 7), η λύση του μειωθέντος προβλήματος υπολογίζεται (λόγω τετραγωνικής αντικειμενικής συνάρτησης) άμεσα μέσω των (7.1.15), (7.1.16). Αντιθέτως, λόγω της γενικά μη γραμμικής αντικειμενικής συνάρτησης f(**x**) στο (8.1.2), πρέπει εδώ να εφαρμοσθεί ένας επαναληπτικός αλγόριθμος του Πρώτου Μέρους για την αριθμητική επίλυση του μειωθέντος προβλήματος.

Αν $\mathbf{x}^{(k)}\in\mathcal{R}$ είναι το παρόν επιτρεπτό επαναληπτικό, τότε ένα γενικό επιτρεπτό σημείο είναι

$$\mathbf{x} = \mathbf{x}^{(k)} + \boldsymbol{\delta} = \mathbf{x}^{(k)} + \mathbf{Zy} \tag{8.1.3}$$

όπου $\boldsymbol{\delta}=\mathbf{Zy}$ είναι επιτρεπτή διόρθωση στον κενό χώρο του **A** (βλ. (7.1.9), (7.1.10)). Έτσι, μια ισοδύναμη εκδοχή του (8.1.2) είναι το ακόλουθο πρόβλημα χωρίς περιορισμούς

$$\psi(\mathbf{y}) = f(\mathbf{x}^{(k)}+\mathbf{Zy}) \rightarrow \underset{\mathbf{y}}{\text{Min}}\,. \tag{8.1.4}$$

Σημειώνουμε ότι κατά την (8.1.4) έχουμε σε κάθε επανάληψη μια νέα μειωθείσα συνάρτηση $\psi^{(k)}(\mathbf{y})$ αφού αλλάζει το αντίστοιχο επαναληπτικό $\mathbf{x}^{(k)}$, αλλάζει δηλαδή η αρχή των αξόνων

στον χώρο **y**, το οποίο όμως δεν επηρεάζει τα διανύσματα κλίσης της ψ(**y**) ούτε τις κατευθύνσει $\mathbf{s}_y$. Το διάνυσμα κλίσης της ψ(**y**) υπολογίζεται μέσω των (8.1.3) και (Π1.3) εξών

$$\nabla_y = \mathbf{Z}^T\nabla_x \tag{8.1.5}$$

και άρα

$$\nabla_y\, \psi(\mathbf{y}) = \mathbf{Z}^T\mathbf{g}(\mathbf{x}) \tag{8.1.6}$$

είναι το **μειωθέν διάνυσμα κλίσης** και

$$\nabla_y^2\psi(\mathbf{y}) = \mathbf{Z}^T\mathbf{G}(\mathbf{x})\mathbf{Z} \tag{8.1.7}$$

είναι ο **μειωθείς Χεσιανός πίνακας**.

Οι (8.1.3), (8.1.6), (8.1.7) παρέχουν ικανή πληροφορία για εφαρμογή των αριθμητικών μεθόδων βελτιστοποίησης χωρίς περιορισμούς του Πρώτου Μέρους στον χώρο των νέων μεταβλητών **y**. Ικανές συνθήκες για τοπικό ελάχιστο είναι $\mathbf{Z}^T\mathbf{g}^*=\mathbf{0}$ και $\mathbf{Z}^T\mathbf{G}^*\mathbf{Z}>\mathbf{0}$, ενώ αναγκαίες συνθήκες (δευτέρου βαθμού) είναι $\mathbf{Z}^T\mathbf{g}^*=\mathbf{0}$ και $\mathbf{Z}^T\mathbf{G}^*\mathbf{Z}\geq\mathbf{0}$. Σαν επιτρεπτό σημείο εκκίνησης μπορεί να ληφθεί $\mathbf{x}^{(1)}=\mathbf{Y}\mathbf{d}$, που ως γνωστόν είναι το κοντινότερο επιτρεπτό σημείο στο **0** στην περίπτωση ορθογώνιας παραγοντοποίησης (βλ. Κεφάλαιο 7.1), ή γενικότερα $\mathbf{x}^{(1)}=\mathbf{x}'+\mathbf{Y}(\mathbf{d}-\mathbf{A}^T\mathbf{x}')$ που είναι το κοντινότερο επιτρεπτό σημείο σε ένα τυχαίο σημείο $\mathbf{x}'$ (πάντα στην περίπτωση ορθογώνιας παραγοντοποίησης).

Εφαρμογή της **μεθόδου Newton** στο χώρο **y** σημαίνει επιλογή κατεύθυνσης αναζήτησης

$$\mathbf{s}_y^{(k)}=-[\mathbf{Z}^T\mathbf{G}^{(k)}\mathbf{Z}]^{-1}\mathbf{Z}^T\mathbf{g}^{(k)}. \tag{8.1.8}$$

και, βάσει της (8.1.3), αυτό σημαίνει για το χώρο **x** την επανάληψη

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)}-\alpha^{(k)}\mathbf{Z}[\mathbf{Z}^T\mathbf{G}^{(k)}\mathbf{Z}]^{-1}\mathbf{Z}^T\mathbf{g}^{(k)}. \tag{8.1.9}$$

Είναι προφανές ότι, αν το $\mathbf{x}^{(k)}$ είναι επιτρεπτό σημείο, τότε η (8.1.9) οδηγεί για κάθε $\alpha^{(k)}$ σε επιτρεπτό $\mathbf{x}^{(k+1)}$, αφού ουσιαστικά πρόκειται για μετακίνηση στον επιτρεπτό χώρο των **y**. Αυτό ισχύει και για τις παρακάτω μεθόδους (8.1.12) και (8.1.14).

Για την εφαρμογή των **μεθόδων σχεδόν-Newton** θεωρούμε τον πίνακα $\mathbf{H}^{(k)}$ στο χώρο **y** να είναι προσέγγιση του αντίστροφου Χεσιανού

$$\mathbf{H}^{(k)} \approx [\mathbf{Z}^T\mathbf{G}^{(k)}\mathbf{Z}]^{-1} \tag{8.1.10}$$

και επομένως έχουμε στο χώρο **y** την κατεύθυνση αναζήτησης

$$\mathbf{s}_y^{(k)}= -\mathbf{H}^{(k)}\nabla_y\psi^{(k)} \tag{8.1.11}$$

και με τις (8.1.3), (8.1.6) την κατεύθυνση αναζήτησης στο χώρο **x**

$$\mathbf{s}^{(k)} = -\mathbf{Z}\mathbf{H}^{(k)}\mathbf{Z}^{T}\mathbf{g}^{(k)}. \qquad (8.1.12)$$

Για τον μετασχηματισμό του $\mathbf{H}^{(k)}$ κατά το Κεφάλαιο 3.2 (βλ. (3.2.2), (3.2.3)) έχουμε τα διανύσματα $\boldsymbol{\gamma}^{(k)} = \mathbf{Z}^{T}(\mathbf{g}^{(k+1)} - \mathbf{g}^{(k)})$ και $\boldsymbol{\delta}^{(k)} = \mathbf{y}^{(k)} - \mathbf{0} = \mathbf{y}^{(k)}$ στο χώρο των μειωθεισών μεταβλητών **y**. Η επιλογή $\mathbf{H}^{(1)}$ μπορεί να είναι ένας τυχαίος θετικά ορισμένος πίνακας, συνήθως $\mathbf{H}^{(1)}=\mathbf{I}$, και όλες οι ιδιότητες που αναφέρθηκαν στο Κεφάλαιο 3.2 ισχύουν και εδώ.

Παρομοίως, η εφαρμογή κατεύθυνσης συζυγών κλίσεων στο χώρο **y** δίδει

$$\mathbf{s}_{y}^{(k)} = -\nabla_{y}\psi^{(k)} + \beta^{(k-1)}\mathbf{s}_{y}^{(k-1)} \qquad (8.1.13)$$

και με τις (8.1.3), (8.1.6) έχουμε για το χώρο **x**

$$\mathbf{s}^{(k)} = -\mathbf{Z}\mathbf{Z}^{T}\mathbf{g}^{(k)} + \beta^{(k-1)}\mathbf{s}^{(k-1)} \qquad (8.1.14)$$

όπου για την μέθοδο Fletcher-Reeves

$$\beta^{(k)} = \frac{\mathbf{g}^{(k+1)^{T}}\mathbf{Z}\mathbf{Z}^{T}\mathbf{g}^{(k+1)}}{\mathbf{g}^{(k)T}\mathbf{Z}\mathbf{Z}^{T}\mathbf{g}^{(k)}} \qquad (8.1.15)$$

και παρόμοια για τους άλλους τύπους συζυγών κλίσεων.

Ο υπολογισμός πολλαπλασιαστών Lagrange $\boldsymbol{\lambda}^{*}$, που είναι αναγκαίος για την επακόλουθη μέθοδο ενεργού συνόλου, είναι δυνατός καταρχήν μέσω των (7.1.14) ή (7.1.16). Βέβαια τα $\mathbf{x}^{*}$ και $\mathbf{g}^{*}$ δεν υπολογίζονται πλέον επακριβώς και οι αντίστοιχες προσεγγίσεις $\mathbf{x}^{(k)}$ και $\mathbf{g}^{(k)}$ μπορούν να είναι περισσότερο ή λιγότερο ανακριβείς. Ο υπολογισμός (βλ. (7.1.14))

$$\boldsymbol{\lambda}^{(k)} = \mathbf{Y}^{T}\mathbf{g}^{(k)} \qquad (8.1.16)$$

μπορεί να χρησιμοποιηθεί για αντίστοιχη προσέγγιση των ζητούμενων πολλαπλασιαστών.

8.2 Περιορισμοί Ανισότητας

Για την επίλυση προβλημάτων που εμπεριέχουν και γραμμικούς περιορισμούς ανισότητας γίνεται χρήση των μεθόδων του Κεφαλαίου 8.1 και μιας μεθόδου ενεργού συνόλου, παρόμοιας με εκείνη του Κεφαλαίου 7.3. Στην μέθοδο ενεργού συνόλου οι περιορισμοί που εμπεριέχονται στο ενεργό σύνολο $\mathcal{A}$ θεωρούνται ισότητες, ενώ οι υπόλοιποι δεν λαμβάνονται προσωρινά υπόψη, όπως και στο Κεφάλαιο 7.3. Κάθε επαναληπτικό $\mathbf{x}^{(k)}$ είναι ένα επιτρεπτό σημείο και κάθε επανάληψη επιχειρεί να προσδιορίσει τη λύση του ΠΙ

$$f(\mathbf{x}^{(k)}+\boldsymbol{\delta}) \rightarrow \underset{\ddot{a}}{\text{Min}} \qquad (8.2.1)$$

$$\text{υπό } \mathbf{a}_i^T \boldsymbol{\delta} = 0, \ i \in \mathcal{A}^{(k)}$$

εκκινώντας από ένα επιτρεπτό σημείο $\mathbf{x}^{(1)}$. Κατά την επίλυση του (8.2.1) προσδιορίζεται μια κατεύθυνση αναζήτησης $\mathbf{s}^{(k)}$, σύμφωνα με την χρησιμοποιούμενη μέθοδο, και το $\mathbf{x}^{(k+1)}$ είναι (στην ιδανική περίπτωση ακριβούς αναζήτησης επί γραμμής) το βέλτιστο επιτρεπτό σημείο επί της γραμμής $\mathbf{x}^{(k)}+\alpha\mathbf{s}^{(k)}$. Στον ΤΠ υπάρχει σαφής έκβαση κάθε επανάληψης, είτε δηλαδή το $\mathbf{x}^{(k+1)}$ επιλύει το ΠΙ, είτε ενεργοποιεί έναν περιορισμό που βρίσκονταν εκτός $\mathcal{A}^{(k)}$. Αυτό δεν ισχύει πλέον στο γενικότερο πρόβλημα (8.2.1), όπου η λύση του ΠΙ δεν είναι παρά το όριο μιας ακολουθίας επαναλήψεων. Έτσι, σε κάθε επανάληψη, πρέπει καταρχήν να αποφασισθεί αν το $\mathbf{x}^{(k)}$ (δηλαδή $\boldsymbol{\delta}=\mathbf{0}$) είναι μια αποδεκτή λύση του ΠΙ. Αν όχι, εκτελούνται μια ή και περισσότερες επαναλήψεις, κατά τις οποίες δεν πρέπει να παραβιάζονται οι περιορισμοί που βρίσκονται εκτός του τρέχοντος ενεργού συνόλου $\mathcal{A}^{(k)}$. Αν ναι, δηλαδή αν το $\mathbf{x}^{(k)}$ γίνεται αποδεκτό ως λύση του ΠΙ, τότε εξετάζονται οι πολλαπλασιαστές Lagrange $\boldsymbol{\lambda}^{(k)}$ για να διαπιστωθεί αν το $\mathbf{x}^{(k)}$ είναι σημείο ΚΤ. Αν ναι, ο αλγόριθμος σταματά έχοντας προσδιορίσει την λύση του αρχικού προβλήματος, άλλως ο δείκτης q του περιορισμού ανισότητας με το μικρότερο $\lambda_q^{(k)}<0$ απομακρύνεται από το $\mathcal{A}^{(k)}$.

Η αναζήτηση επί γραμμής για τον προσδιορισμό του βέλτιστου επιτρεπτού σημείου σ' αυτήν τη γενικότερη περίπτωση δεν είναι ούτε αυτή άμεση, αφού απαιτεί μια επαναληπτική διαδικασία που είναι πολυπλοκότερη από την (7.3.2). Κατά τα Κεφάλαια 2.5 και 2.6, είναι αναγκαίο να καθορισθεί ένα πλαίσιο αποδεκτών τιμών του βήματος α και να προσδιορισθεί μια βέλτιστη τιμή $\alpha^{(k)}$ μέσω υποδιαίρεσης και παρεμβολής. Επί πλέον όμως, εδώ, το βήμα α δεν πρέπει να υπερβεί μια τιμή $\bar{\alpha}^{(k)}$, όπως στην (7.3.2),

$$\bar{\alpha}^{(k)} = \min_{\substack{i \notin \mathcal{A}^{(k)} \\ \mathbf{a}_i^T \mathbf{s}^{(k)} < 0}} \frac{\mathbf{d}_i - \mathbf{a}_i^T \mathbf{x}^{(k)}}{\mathbf{a}_i^T \mathbf{s}^{(k)}}. \qquad (8.2.2)$$

Αυτός ο περιορισμός απαιτεί κατάλληλες αλλαγές στους αλγορίθμους αναζήτησης επί γραμμής των Κεφαλαίων 2.5 και 2.6. Επί παραδείγματι, αν $f'(\bar{\alpha}^{(k)}) <0$ τότε το ζητούμενο βήμα είναι $\alpha^{(k)} = \bar{\alpha}^{(k)}$ (γιατί;). Αν $f'(\bar{\alpha}^{(k)}) >0$, τότε έχουμε αγκύλη $[0, \bar{\alpha}^{(k)}]$ εντός της οποίας μπορεί να εφαρμοστεί ο αλγόριθμος υποδιαίρεσης.

Με τις παρατηρήσεις αυτές, η μέθοδος ενεργού συνόλου για το (8.1.1) έχει ως εξής:

a) Δίδεται επιτρεπτό $\mathbf{x}^{(1)}$ και αντίστοιχο $\mathcal{A}^{(1)}$, k=1.
b) Αν $\boldsymbol{\delta}=\mathbf{0}$ δεν είναι αποδεκτή λύση του (8.2.1), πήγαινε στο (d).
c) Έστω $\lambda_q^{(k)} = \min_i \{\lambda_i^{(k)}, \ i \in I \cap \mathcal{A}^{(k)}\}$. Αν $\lambda_q^{(k)} \geq 0$, η λύση είναι $\mathbf{x}^*=\mathbf{x}^{(k)}$, άλλως απομακρύνεται το q από το $\mathcal{A}^{(k)}$.
d) Προσδιορίζεται για το (8.2.1) επιτρεπτή κατεύθυνση κατάβασης $\mathbf{s}^{(k)}$.

e) Επιλογή $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)}+\alpha^{(k)}\mathbf{s}^{(k)}$ ως βέλτιστου επιτρεπτού σημείου επί της γραμμής $\mathbf{x}^{(k)}+\alpha\mathbf{s}^{(k)}$ υπό τον περιορισμό $\alpha^{(k)} \le \bar{\alpha}^{(k)}$.

f) Αν $\alpha^{(k)}= \bar{\alpha}^{(k)}$, προστίθεται ο αντίστοιχος ενεργοποιούμενος περιορισμός p στο $\mathcal{A}^{(k)}$.

g) k:=k+1, πήγαινε στο (b).

Μια επιπλέον ιδιαιτερότητα του αλγορίθμου αυτού πηγάζει από το γεγονός ότι ο αριθμός των ελεύθερων μεταβλητών του προβλήματος μπορεί να αυξομειώνεται από επανάληψη σε επανάληψη εξ αιτίας των αλλαγών του ενεργού συνόλου στα βήματα (c) και (f). Έχουν προταθεί διάφοροι τρόποι οικονομικής και αποτελεσματικής αντιμετώπισης της ιδιαιτερότητας αυτής, καθώς και μερικών άλλων δυσκολιών που μπορούν να εμφανισθούν στον παραπάνω αλγόριθμο, π.χ. ταλαντώσεις από επανάληψη σε επανάληψη μεταξύ δύο ενεργών συνόλων, φαινόμενο γνωστό ως **ζιγκ-ζαγκ**.

Προβλήματα

8.1 Θεωρούμε το Πρόβλημα 6.8.

a) Υπολογίστε τους πίνακες **Y**, **Z** που αντιστοιχούν σε άμεση απαλοιφή.

Ζητούνται τώρα οι εξισώσεις αριθμητικής επίλυσης του προβλήματος στο χώρο **x**.

b) Υπολογίστε ένα επιτρεπτό σημείο εκκίνησης $\mathbf{x}^{(1)} = \mathbf{Y}d$.

c) Υπολογίστε την κατεύθυνση αναζήτησης $\mathbf{s}^{(k)}$ στον χώρο **x** που αντιστοιχεί σε μέγιστη κατάβαση στον χώρο y.

d) Προσδιορίστε και επιλύστε αναλυτικά το πρόβλημα αναζήτησης επί γραμμής και αναγράψτε τον τρόπο υπολογισμού του νέου επαναληπτικού $\mathbf{x}^{(k+1)}$. Πόσες επαναλήψεις απαιτούνται για τη λύση; Αιτιολογήστε.

Λύση

a) $\mathbf{A} = \begin{bmatrix}1\\1\end{bmatrix}$. Για άμεση απαλοιφή $\mathbf{V} = \begin{bmatrix}0\\1\end{bmatrix}$

$$\begin{bmatrix}\mathbf{A} & \mathbf{V}\end{bmatrix}^{-1} = \begin{bmatrix}1 & 0\\1 & 1\end{bmatrix}^{-1} = \begin{bmatrix}1 & 0\\-1 & 1\end{bmatrix} = \begin{bmatrix}\mathbf{Y}^T\\\mathbf{Z}^T\end{bmatrix} \Rightarrow \mathbf{Y} = \begin{bmatrix}1\\0\end{bmatrix},\ \mathbf{Z} = \begin{bmatrix}-1\\1\end{bmatrix}$$

$$\mathbf{x} = \mathbf{x}^{(k)} + \mathbf{Z}y \Rightarrow x_1 = x_1^{(k)} - y,\quad x_2 = x_2^{(k)} + y$$

άρα $x_1+x_2=1$ αν $x_1^{(k)} + x_2^{(k)} = 1$.

b) Επιτρεπτό σημείο εκκίνησης $\mathbf{x}^{(1)} = \mathbf{Y}d = \begin{bmatrix}1\\0\end{bmatrix}$.

c) $s_y^{(k)} = -\nabla_y \psi^{(k)}(y)$ και με (8.1.3), (8.1.6)

$\mathbf{s}^{(k)} = -\mathbf{Z}\mathbf{Z}^T\mathbf{g}^{(k)} = \begin{bmatrix} g_2^{(k)} - g_1^{(k)} \\ g_1^{(k)} - g_2^{(k)} \end{bmatrix}$: κατεύθυνση αναζήτησης στο χώρο $\mathbf{x}$

όπου $g_1 = \dfrac{\partial f}{\partial x_1} = 3x_1^2 + 2x_1x_2 + x_1 - 6, \quad g_2 = \dfrac{\partial f}{\partial x_2} = x_1^2$.

d) Αναζήτηση επί γραμμής: $\alpha^{(k)}$ ελαχιστοποιεί $f(\mathbf{x}^{(k)} + \alpha\mathbf{s}^{(k)})$.

Με $\mathbf{s}^{(k)} = -s^{(k)}\begin{bmatrix} 1 \\ -1 \end{bmatrix}$, όπου $s^{(k)} = g_1^{(k)} - g_2^{(k)}$, έχουμε

$f(\mathbf{x}^{(k)} + \alpha\mathbf{s}^{(k)}) = (x_1^{(k)} - \alpha s^{(k)})^3 + (x_1^{(k)} - \alpha s^{(k)})^2 (x_2^{(k)} + \alpha s^{(k)}) + 0.5(x_1^{(k)} - \alpha s^{(k)} - 6)^2$

και από $\dfrac{df}{d\alpha} = 0 \Rightarrow \alpha^{(k)} = (x_1^{(k)} - 2)/s^{(k)}$.

Νέο επαναληπτικό: $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \alpha^{(k)}\mathbf{s}^{(k)}$, δηλαδή

$x_1^{(k+1)} = x_1^{(k)} - \alpha^{(k)}s^{(k)} = 2 = x_1^*$

$x_2^{(k+1)} = x_2^{(k)} + \alpha^{(k)}s^{(k)} = \underbrace{x_2^{(k)} + x_1^{(k)}}_{=1} - 2 = -1 = x_2^*$.

Άρα προσδιορισμός της λύσης σε μια μόνο επανάληψη, ανεξάρτητα από το σημείο εκκίνησης $\mathbf{x}^{(1)}$. Αιτιολογία: Οι παραπάνω επαναλήψεις στο χώρο $\mathbf{x}$ αντιστοιχούν σε επαναλήψεις του χώρου y, ο οποίος όμως είναι μονοδιάστατος και ως εκ τούτου μια και μόνη αναζήτηση επί γραμμής αρκεί για τον προσδιορισμό του ελάχιστου.

9. ΑΛΓΟΡΙΘΜΟΙ ΜΗ ΓΡΑΜΜΙΚΟΥ ΠΡΟΓΡΑΜΜΑΤΙΣΜΟΥ

Η αριθμητική επίλυση γενικών προβλημάτων (5.1.1) Μη Γραμμικού Προγραμματισμού, όπου τόσο η αντικειμενική συνάρτηση όσο και οι συναρτήσεις περιορισμών μπορούν να είναι μη γραμμικές, είναι η δυσκολότερη περίπτωση λείων προβλημάτων βελτιστοποίησης. Δεν υπάρχει αυτή τη στιγμή μέθοδος που να αναγνωρίζεται γενικά ως η αποτελεσματικότερη για κάθε τύπο προβλήματος. Η αντίστοιχη έρευνα συνεχίζεται και είναι πιθανό να οδηγήσει σε νέα ενδιαφέροντα αποτελέσματα στο μέλλον. Στο κεφάλαιο αυτό θα παρουσιασθούν μια σειρά από τις γνωστότερες μεθόδους αριθμητικής επίλυσης που συγκεντρώνουν, κάθε μία, μιά σειρά από πλεονεκτήματα και μειονεκτήματα. Πολύ περισσότερες μέθοδοι ή παραλλαγές μεθόδων έχουν προταθεί στην σχετική επιστημονική βιβλιογραφία, και η αντίστοιχη έρευνα, όπως ειπώθηκε παραπάνω, συνεχίζεται αμείωτη. Για τις περισσότερες από τις παρακάτω μεθόδους, όπως εξ άλλου και για τις μεθόδους προηγούμενων κεφαλαίων, υπάρχουν έτοιμα προγράμματα για γενική χρήση σε πολλές συλλογές προγραμμάτων υπολογιστών όπως IMSL, NAG, MATLAB κλπ.

9.1 Μέθοδοι Συνάρτησης Τιμωρίας

Οι **μέθοδοι συνάρτησης τιμωρίας** επιχειρούν, όπως και άλλες αριθμητικές μέθοδοι, να αναγάγουν το πρόβλημα περιορισμένης βελτιστοποίησης σε ένα πρόβλημα απεριόριστης βελτιστοποίησης που μπορεί να επιλυθεί μέσω των μεθόδων του Πρώτου Μέρους. Προς το σκοπό αυτό, η μέθοδος συνάρτησης τιμωρίας επεκτείνει την αντικειμενική συνάρτηση $f(\mathbf{x})$ με κατάλληλους **όρους τιμωρίας**, κατά τρόπο ώστε να επιβάλεται η ικανοποίηση των περιορισμών ισότητας και ανισότητας

$$\varphi(\mathbf{x}, \sigma) = f(\mathbf{x}) + \frac{1}{2}\sigma\sum_{i\in E} c_i(\mathbf{x})^2 + \frac{1}{2}\sigma\sum_{i\in I}\min\{c_i(\mathbf{x}),0\}^2 \qquad (9.1.1)$$

όπου $\sigma>0$ είναι παράμετρος βάρους. Οι τετραγωνικοί όροι "τιμωρούν" την παραβίαση των περιορισμών, έτσι ώστε, για $\sigma\to\infty$, η απεριόριστη ελαχιστοποίηση της $\varphi(\mathbf{x}, \sigma)$ να οδηγεί όχι μόνο στην ελαχιστοποίηση της $f(\mathbf{x})$, αλλά και σε $c_i(\mathbf{x})\to 0$, $i\in E$, και $c_i(\mathbf{x})\geq 0$, $i\in I$. Υπάρχουν βέβαια, εκτός των παραπάνω τετραγωνικών όρων, και άλλες δυνατότητες τιμωρίας παραβίασης περιορισμών, που οδηγούν όμως σε παρόμοια τελικά αποτελέσματα. Για κάποιο πεπερασμένο σ σημειώνουμε $\mathbf{x}^*(\sigma)$ την λύση του απεριόριστου προβλήματος ελαχιστοποίησης (9.1.1).

Η μέθοδος συνάρτησης τιμωρίας ήταν ιστορικά μια από τις πρώτες προταθείσες μεθόδους αριθμητικής επίλυσης γενικών προβλημάτων Μη Γραμμικού Προγραμματισμού. Η αρχική αισιοδοξία, απόρροια της εξαιρετικής απλότητας της μεθόδου, μετριάσθηκε σύντομα λόγω των παρουσιαζόμενων πρακτικών δυσκολιών. Πράγματι, η σχετικά ακριβής ικανοποίηση των περιορισμών απαιτεί υψηλές τιμές του σ, πράγμα που οδηγεί σε προβλήματα απεριόριστης

ελαχιστοποίησης νοσηρής κατάστασης και δυσκολεύει σημαντικά την εφαρμογή ακόμη και των αποτελεσματικότερων μεθόδων του Πρώτου Μέρους. Για να απαλυνθούν τα αριθμητικά προβλήματα που γεννά η νοσηρή κατάσταση, επιλύεται συνήθως μια ακολουθία k=1,2,... μη περιορισμένων προβλημάτων με αντίστοιχα αυξανόμενα βάρη $\sigma^{(k)}$, π.χ. $\sigma^{(k+1)} = 10\sigma^{(k)}$, όπου το σημείο εκκίνησης του (k+1)-του προβλήματος τίθεται ίσο με την λύση $\mathbf{x}^*(\sigma^{(k)})$ του προηγούμενου προβλήματος της ακολουθίας. Βέβαια ο υπολογιστικός φόρτος επίλυσης του αρχικού περιορισμένου προβλήματος αυξάνει λόγω της επίλυσης όχι ενός, αλλά μιας ακολουθίας μη περιορισμένων προβλημάτων. Η ακολουθία μη περιορισμένων ελαχιστοποιήσεων διακόπτεται όταν ικανοποιούνται κριτήρια ανοχής $|c_i[\mathbf{x}^*(\sigma^{(k)})]| < \varepsilon_i$, $i\in E$, και $|\min\{0, c_i[\mathbf{x}^*(\sigma^{(k)})]\}| < \varepsilon_i$, $i\in I$, για δεδομένες παραμέτρους ανοχής ε_i.

Μπορεί εύκολα να αποδειχθεί ότι, αν τα $\mathbf{x}^*(\sigma)$ είναι απόλυτα ελάχιστα των απεριόριστων προβλημάτων, ισχύουν τα ακόλουθα:

– $\varphi[\mathbf{x}^*(\sigma^{(k)}), \sigma^{(k)}]$ είναι συνεχώς αύξουσα ως προς k
– $|c_i[\mathbf{x}^*(\sigma^{(k)})]|$, $i\in E$, και $|\min\{0, c_i[\mathbf{x}^*(\sigma^{(k)})]\}|$, $i\in I$, είναι συνεχώς φθίνουσες ως προς k
– $f[\mathbf{x}^*(\sigma^{(k)})]$ είναι συνεχώς αύξουσα ως προς k.

Οι ιδιότητες αυτές εγγυώνται τη σύγκλιση της ακολουθίας προς τη λύση $\mathbf{x}^*$ του περιορισμένου προβλήματος βελτιστοποίησης. Εναλλακτικά μπορεί βέβαια να επιχειρηθεί μια και μόνη απεριόριστη ελαχιστοποίηση της (9.1.1). Το βάρος σ θα πρέπει τότε να αντανακλά έναν κατάλληλο συμβιβασμό μεταξύ ταχύτητας σύγκλισης αφ' ενός και ακρίβειας ικανοποίησης των περιορισμών αφ' ετέρου.

Παράδειγμα 9.1.1: Θεωρούμε την ελαχιστοποίηση της αντικειμενικής συνάρτησης $f(x_1,x_2)= -x_1-x_2$ υπό $1-x_1^2-x_2^2 = 0$. Η αναλυτική λύση του προβλήματος μέσω των συνθηκών βελτιστοποίησης του Κεφαλαίου 6 δίδει $x_1^* = x_2^* = 1/\sqrt{2}$. Με εφαρμογή της συνάρτησης τιμωρίας έχουμε

$$\varphi(x_1,x_2,\sigma) = -x_1-x_2+\frac{1}{2}\sigma(1-x_1^2-x_2^2)^2. \qquad (9.1.2)$$

Η ελαχιστοποίηση της φ χωρίς περιορισμούς απαιτεί

$$\partial\varphi/\partial x_1 = -1-2\sigma x_1(1-x_1^2-x_2^2) = 0$$
$$\partial\varphi/\partial x_2 = -1-2\sigma x_2(1-x_1^2-x_2^2) = 0. \qquad (9.1.3)$$

Από τη συμμετρία των δύο εξισώσεων έχουμε καταρχήν $x_1=x_2$, και δι' αντικαταστάσεως στην (9.1.3) $4\sigma x_1^3 - 2\sigma x_1 - 1 = 0$. Η ικανή συνθήκη $\nabla^2\varphi>\mathbf{0}$ ικανοποιείται εφ' όσον $x_1>1/\sqrt{2}$. Η λύση της παραπάνω εξίσωσης τρίτου βαθμού δίδει λύσεις $x_1^*(\sigma)$ που αναφέρονται στον Πίνακα 9.1.1 σε συνάρτηση του σ. Όπως γίνεται φανερό, αύξουσες τιμές του βάρους σ οδηγούν σε βελτίωση ακρίβειας της λύσης ως προς την ικανοποίηση του περιορισμού ισότητας. Το Σχ.

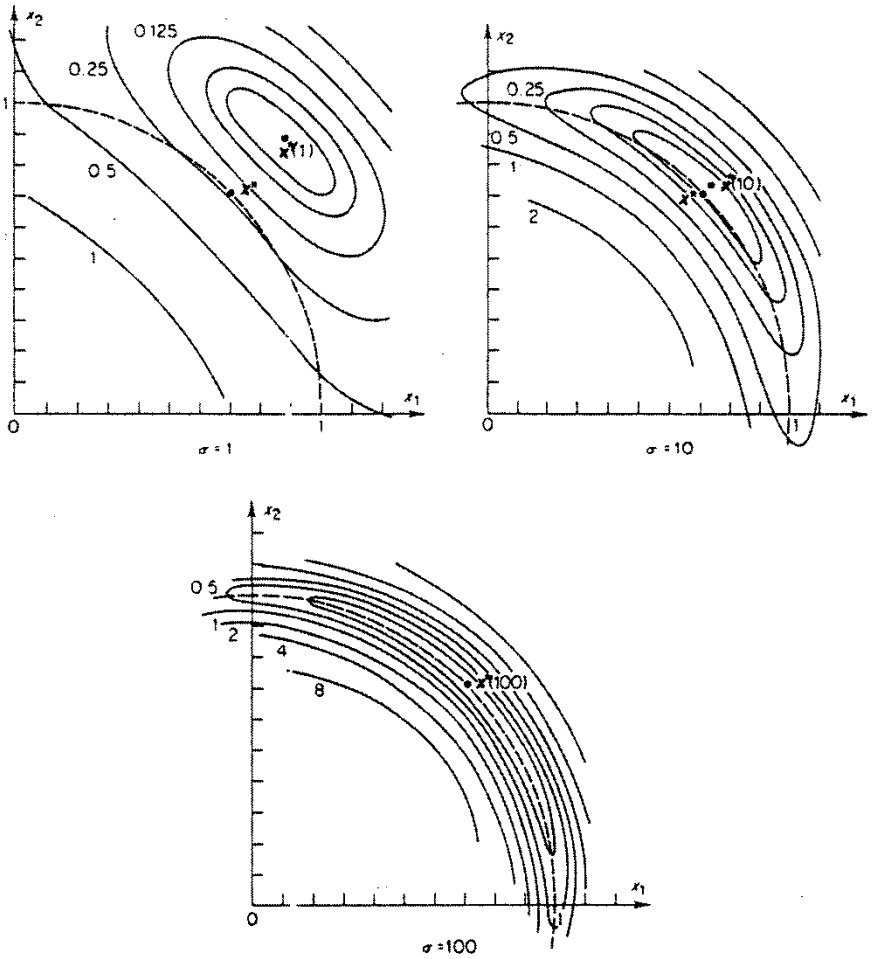
9.1.1 δείχνει τα περιγράμματα της φ για διάφορες τιμές του σ, όπου γίνεται φανερό ότι μεγαλύτερη ακρίβεια οδηγεί σε νοσηρή κατάσταση του προβλήματος.

■

Η μέθοδος συνάρτησης τιμωρίας (όπως και οι μέθοδοι των ακόλουθων κεφαλαίων) μπορεί βέβαια να χρησιμοποιηθεί και για ένα μέρος μόνο των περιορισμών. Για παράδειγμα, γραμμικοί περιορισμοί μπορούν να ληφθούν υπόψη κατά το Κεφάλαιο 8, ενώ όροι τιμωρίας να εισαχθούν μόνο για μη γραμμικούς περιορισμούς κλπ.

σ	**1**	**10**	**100**	**1000**	**10000**	**∞**
$x_1^*(\sigma) = x_2^*(\sigma)$	0.885	0.731	0.71	0.7074	0.70713	$1/\sqrt{2}$
$\lvert c[x^*(\sigma)] \rvert$	0.566	0.0687	0.0082	$8.3\cdot10^{-4}$	$6.6\cdot10^{-5}$	0
$f(x^*(\sigma))$	–1.77	–1.462	–1.42	–1.4148	–1.41426	$-\sqrt{2}$

Πίνακας 9.1.1: Εφαρμογή της μεθόδου συνάρτησης τιμωρίας.



Σχήμα 9.1.1: Περιγράμματα κατά την εφαρμογή της μεθόδου συνάρτησης τιμωρίας.

Μερικές από τις μεθόδους αυτού ή προηγούμενων κεφαλαίων απαιτούν την παρουσία ενός τυχαίου επιτρεπτού σημείου εκκίνησης των επαναλήψεων. Αν ο χρήστης δεν διαθέτει ένα επιτρεπτό σημείο, μπορεί να ελαχιστοποιηθεί χωρίς περιορισμούς η ακόλουθη συνάρτηση, πράγμα που οδηγεί στον προσδιορισμό ενός επιτρεπτού σημείου

$$\varphi(\mathbf{x}) = \frac{1}{2}\sum_{i\in E} c_i(\mathbf{x})^2 + \frac{1}{2}\sum_{i\in I} \min\{0, c_i(\mathbf{x})\}^2. \qquad (9.1.4)$$

Αυτή η συνάρτηση τιμωρίας προκύπτει αν απομακρυνθεί από την (9.1.1) η αντικειμενική συνάρτηση $f(\mathbf{x})$ και άρα και η παράμετρος βάρους σ. Αν $\min\varphi(\mathbf{x}) = 0$, έχει προφανώς προσδιορισθεί ένα επιτρεπτό σημείο, άλλως το προσδιορισθέν ελάχιστο της ως άνω συνάρτησης φ δεν είναι απόλυτο ή η επιτρεπτή περιοχή του αντίστοιχου προβλήματος είναι κενή.

9.2 Συναρτήσεις Τιμωρίας με Πολλαπλασιαστές

α) Περιορισμοί Ισότητας

Οι συναρτήσεις τιμωρίας χρησιμοποιούνται για τον προσδιορισμό ενός τοπικού ελάχιστου $\mathbf{x}^*$ στο όριο $\sigma^{(k)}\to\infty$. Αν όμως μετατεθεί κατάλληλα το μηδενικό σημείο του όρου τιμωρίας, τότε μπορεί το $\mathbf{x}^*$ να προσδιορισθεί για πεπερασμένο σ. Έχουμε, λοιπόν, καταρχήν για προβλήματα με περιορισμούς ισότητας μόνον,

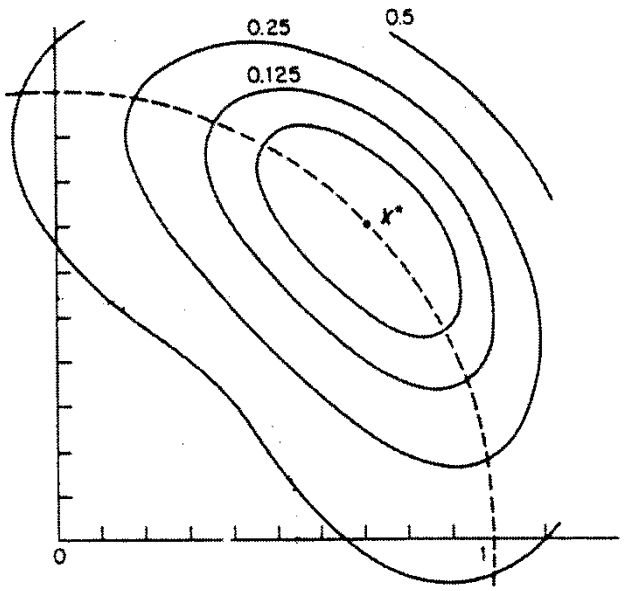
$$\varphi(\mathbf{x}, \boldsymbol{\theta}, \boldsymbol{\sigma}) = f(\mathbf{x}) + \frac{1}{2}\sum \sigma_i\,(c_i(\mathbf{x}) - \theta_i)^2 = f(\mathbf{x}) + \frac{1}{2}\,[\mathbf{c}(\mathbf{x})-\boldsymbol{\theta}]^T\mathbf{S}[\mathbf{c}(\mathbf{x})-\boldsymbol{\theta}] \qquad (9.2.1)$$

όπου $\mathbf{S}=\mathbf{diag}(\sigma_i)$ και οι παράμετροι θ_i αντικατοπτρίζουν την μετάθεση του μηδενικού σημείου. Για κάποια πεπερασμένη τιμή του $\boldsymbol{\sigma}$, υπάρχει μια κατάλληλη τιμή του $\boldsymbol{\theta}$ έτσι ώστε η απεριόριστη ελαχιστοποίηση της φ στην (9.2.1) να οδηγεί στο περιορισμένο ελάχιστο $\mathbf{x}^*$ του αρχικού προβλήματος. Έτσι, στο πρόβλημα του Παραδ. 9.1.1, η κατάλληλη μετάθεση για σ=1 είναι $\theta=1/\sqrt{2}$. Τα περιγράμματα της (9.2.1) για αυτές τις τιμές δίδονται στο Σχ. 9.2.1 και δείχνουν ότι, πρώτον, η απεριόριστη ελαχιστοποίηση της (9.2.1) λύνει το αρχικό πρόβλημα, και, δεύτερον, η (9.2.1) δεν έχει νοσηρή κατάσταση.

Στην εφαρμογή της (9.2.1) είναι προτιμότερο να χρησιμοποιούνται διαφορετικές παράμετροι

$$\lambda_i=\theta_i\sigma_i \qquad (9.2.2)$$

ενώ αγνοείται ο όρος $\frac{1}{2}\sum\sigma_i\theta_i^2$ που είναι ανεξάρτητος του $\mathbf{x}$. Καταλήγουμε έτσι στην



Σχήμα 9.2.1: Περιγράμματα συνάρτησης τιμωρίας με πολλαπλασιαστές για το Παράδ. 9.1.1.

$$\varphi(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\sigma}) = f(\mathbf{x}) - \boldsymbol{\lambda}^T\mathbf{c}(\mathbf{x}) + \frac{1}{2}\,\mathbf{c}(\mathbf{x})^T\mathbf{S}\mathbf{c}(\mathbf{x}). \qquad (9.2.3)$$

Σημειώνουμε, για δεδομένο $\boldsymbol{\sigma}$, με $\mathbf{x}(\boldsymbol{\lambda})$ το ελάχιστο της (9.2.3) ως προς $\mathbf{x}$ σε συνάρτηση του $\boldsymbol{\lambda}$. Σημειώνουμε επίσης με $\mathbf{x}^*$, $\boldsymbol{\lambda}^*$ τη λύση του αρχικού προβλήματος. Έχουμε τότε το ακόλουθο θεώρημα για την (9.2.3).

Θεώρημα 9.2.1: Αν τα $\mathbf{x}^*$, $\boldsymbol{\lambda}^*$ ικανοποιούν τις ικανές συνθήκες του αρχικού προβλήματος, τότε υπάρχει $\boldsymbol{\sigma}'\geq\mathbf{0}$ έτσι ώστε για κάθε $\boldsymbol{\sigma}>\boldsymbol{\sigma}'$ το $\mathbf{x}^*$ να είναι αυστηρό τοπικό ελάχιστο της $\varphi(\mathbf{x}, \boldsymbol{\lambda}^*, \boldsymbol{\sigma})$, δηλαδή $\mathbf{x}^* = \mathbf{x}(\boldsymbol{\lambda}^*)$.

■

Από το Θεώρημα 9.2.1 συμπεραίνουμε ότι για $\boldsymbol{\lambda}=\boldsymbol{\lambda}^*$ η ελαχιστοποίηση της $\varphi(\mathbf{x},\boldsymbol{\lambda},\boldsymbol{\sigma})$ μας οδηγεί στη λύση $\mathbf{x}^*$ του αρχικού προβλήματος, και μάλιστα για οποιαδήποτε πεπερασμένη τιμή $\boldsymbol{\sigma}\geq\boldsymbol{\sigma}'$. Αφού η λύση μπορεί να προσδιορισθεί ανεξάρτητα από την τιμή του $\boldsymbol{\sigma}$, μπορούμε να αγνοήσουμε την εξάρτηση της φ από το $\boldsymbol{\sigma}$ και να γράφουμε $\varphi(\mathbf{x}, \boldsymbol{\lambda})$ στην (9.2.3).

Με τα δεδομένα αυτά μπορούμε να σκιαγραφήσουμε καταρχήν έναν αλγόριθμο ακολουθίας ελαχιστοποιήσεων, όπου ο $\boldsymbol{\lambda}$ γίνεται ο ρυθμιστής της όλης διαδικασίας, ως εξής:

(i) Καθορισμός μιας ακολουθίας $\{\boldsymbol{\lambda}^{(k)}\} \rightarrow \boldsymbol{\lambda}^*$.

(ii) Για κάθε $\boldsymbol{\lambda}^{(k)}$, προσδιορισμός ενός απεριόριστου τοπικού ελάχιστου $\mathbf{x}(\boldsymbol{\lambda}^{(k)})$ της $\varphi(\mathbf{x},\boldsymbol{\lambda}^{(k)})$ (μέσω των αριθμητικών μεθόδων του Πρώτου Μέρους). (9.2.4)

(iii) Τερματισμός, όταν το $\mathbf{c}(\mathbf{x}(\boldsymbol{\lambda}^{(k)}))$ είναι αρκούντως μικρό.

Η κύρια διαφορά μεταξύ αυτής της ακολουθίας ελαχιστοποιήσεων και εκείνης του Κεφαλαίου 9.1 είναι ότι το $\boldsymbol{\lambda}^*$ δεν είναι γνωστό εκ των προτέρων, και άρα η ακολουθία $\{\boldsymbol{\lambda}^{(k)}\}$ δεν μπορεί να προκατασκευασθεί. Επιπλέον, το κάτω όριο $\boldsymbol{\sigma}'$ δεν είναι γνωστό εκ των προτέρων και έτσι η επιλογή του $\boldsymbol{\sigma}$ είναι δυσχερής, αφού, αν $\boldsymbol{\sigma} >> \boldsymbol{\sigma}'$, η κατάσταση του προβλήματος ελαχιστοποίησης της φ χειροτερεύει χωρίς λόγο, ενώ, αν $\boldsymbol{\sigma} < \boldsymbol{\sigma}'$, μπορεί να μην υπάρχει σύγκλιση προς τη λύση του αρχικού προβλήματος. Θα δούμε όμως παρακάτω πώς μπορούν να αντιμετωπισθούν οι δυσκολίες αυτές.

Επειδή η (9.2.3) μπορεί να θεωρηθεί ότι προέρχεται από την (9.1.1) με την προσθήκη του όρου πολλαπλασιαστών $-\boldsymbol{\lambda}^T\mathbf{c}$, η φ ονομάζεται **συνάρτηση τιμωρίας με πολλαπλασιαστές**. Μια εναλλακτική θεώρηση είναι πως η (9.2.3) είναι η Λαγκρανζιανή συνάρτηση (6.1.6), όπου η αντικειμενική συνάρτηση f έχει επαυξηθεί με τον όρο $(1/2)\mathbf{c}^T\mathbf{S}\mathbf{c}$. Για το λόγο αυτό, η φ ονομάζεται συχνά και **επαυξημένη Λαγκρανζιανή συνάρτηση**.

Εξετάζοντας την συνάρτηση

$$\psi(\boldsymbol{\lambda}) = \varphi(\mathbf{x}(\boldsymbol{\lambda}), \boldsymbol{\lambda}) \tag{9.2.5}$$

έχουμε

$$\psi(\boldsymbol{\lambda}) = \varphi(\mathbf{x}(\boldsymbol{\lambda}), \boldsymbol{\lambda}) \le \varphi(\mathbf{x}^*,\boldsymbol{\lambda}) = \varphi(\mathbf{x}^*,\boldsymbol{\lambda}^*) = \psi(\boldsymbol{\lambda}^*) \tag{9.2.6}$$

(όπου έγινε χρήση της $\mathbf{c}^*=\mathbf{0}$). Άρα το $\boldsymbol{\lambda}^*$ είναι απεριόριστο τοπικό μέγιστο της $\psi(\boldsymbol{\lambda})$. Το $\boldsymbol{\lambda}^*$ είναι μάλιστα απόλυτο μέγιστο της $\psi(\boldsymbol{\lambda})$, αν το $\mathbf{x}(\lambda)$ είναι απόλυτο ελάχιστο της $\varphi(\mathbf{x},\boldsymbol{\lambda})$. Τότε όμως μπορεί να αναπτυχθεί μια ακολουθία $\boldsymbol{\lambda}^{(k)} \to \boldsymbol{\lambda}^*$ εφαρμόζοντας μεθόδους μη περιορισμένης ελαχιστοποίησης στην $-\psi(\boldsymbol{\lambda})$. Για να γίνει αυτό, απαιτούνται τύποι υπολογισμού των παραγώγων $\nabla\psi$ και $\nabla^2\psi$ ως προς $\boldsymbol{\lambda}$. Έχουμε από την (9.2.5) με χρήση της (Π1.8)

$$\frac{\mathrm{d}\psi}{\mathrm{d}\boldsymbol{\lambda}} = \frac{\partial\mathbf{x}}{\partial\boldsymbol{\lambda}}\,\frac{\partial\varphi}{\partial\mathbf{x}} + \frac{\partial\varphi}{\partial\boldsymbol{\lambda}}.$$

Επειδή όμως $\partial\varphi/\partial\mathbf{x} = \mathbf{0}$ (αφού το $\mathbf{x}(\boldsymbol{\lambda})$ ελαχιστοποιεί την φ) και $\partial\varphi/\partial\boldsymbol{\lambda} = -\mathbf{c}$ από την (9.2.3), καταλήγουμε στην

$$\nabla\psi(\boldsymbol{\lambda}) = -\mathbf{c}(\mathbf{x}(\boldsymbol{\lambda})). \tag{9.2.7}$$

Παίρνοντας την δεύτερη παράγωγο της $\psi(\boldsymbol{\lambda})$ καταλήγουμε κατά παρόμοιο τρόπο στην

$$\nabla^2\psi(\boldsymbol{\lambda}) = -\mathbf{A}^T\,\mathbf{W}_{\boldsymbol{\sigma}}^{-1}\,\mathbf{A}\,\big|_{\mathbf{x}(\lambda)} \tag{9.2.8}$$

όπου $\mathbf{W}_\sigma = \nabla^2\varphi(\mathbf{x}(\boldsymbol{\lambda}),\boldsymbol{\lambda})$ και $\mathbf{A} = -\partial^2\varphi/\partial\mathbf{x}\partial\boldsymbol{\lambda}$.

Μια δυνατότητα ακολουθίας $\{\boldsymbol{\lambda}^{(k)}\}$ κάνει χρήση της μεθόδου Newton

$$\boldsymbol{\lambda}^{(k+1)} = \boldsymbol{\lambda}^{(k)} - (\mathbf{A}^T\mathbf{W}_\sigma^{-1}\mathbf{A})^{-1}\mathbf{c}\,|_{\mathbf{x}(\boldsymbol{\lambda}^{(k)})}. \qquad (9.2.8)$$

Η (9.2.8) απαιτεί βέβαια τον αναλυτικό υπολογισμό δεύτερων παραγώγων και έχει τα γνωστά μειονεκτήματα της μεθόδου Newton. Επειδή για μεγάλο $\boldsymbol{\sigma}$

$$(\mathbf{A}^T\mathbf{W}_\sigma^{-1}\mathbf{A})^{-1} \approx \mathbf{S} \qquad (9.2.9)$$

οι Powell (1969) και Hestenes (1969) πρότειναν την πολύ απλούστερη ακολουθία

$$\boldsymbol{\lambda}^{(k+1)} = \boldsymbol{\lambda}^{(k)} - \mathbf{S}\mathbf{c}^{(k)}. \qquad (9.2.10)$$

Καταλήγουμε έτσι στον ακόλουθο αλγόριθμο:

(i) Τιμές εκκίνησης $\boldsymbol{\lambda}=\boldsymbol{\lambda}^{(1)}$, $\boldsymbol{\sigma}=\boldsymbol{\sigma}^{(1)}$, $\|\mathbf{c}^{(k)}\|_\infty=\infty$, k=1.

(ii) Προσδιορισμός απεριόριστου ελάχιστου $\mathbf{x}(\boldsymbol{\lambda},\boldsymbol{\sigma})$ της $\varphi(\mathbf{x},\boldsymbol{\lambda},\boldsymbol{\sigma})$ (μέσω των αριθμητικών μεθόδων του Πρώτου Μέρους) και υπολογισμός $\mathbf{c}=\mathbf{c}(\mathbf{x}(\boldsymbol{\lambda},\boldsymbol{\sigma}))$.

(iii) Αν $\|\mathbf{c}\|_\infty > \frac{1}{4}\|\mathbf{c}^{(k)}\|_\infty$, τότε $\sigma_i:=10\sigma_i\ \forall i$: $|c_i| > \frac{1}{4}\|\mathbf{c}^{(k)}\|_\infty$, και πήγαινε στο (ii).

(iv) Αν $\|\mathbf{c}\|_\infty < \varepsilon$, τερματισμός.

(v) k:=k+1, $\boldsymbol{\lambda}^{(k)} = \boldsymbol{\lambda}$, $\boldsymbol{\sigma}^{(k)}=\boldsymbol{\sigma}$, $\mathbf{c}^{(k)}=\mathbf{c}$.

(vi) $\boldsymbol{\lambda}=\boldsymbol{\lambda}^{(k)}-\mathbf{S}^{(k)}\mathbf{c}^{(k)}$, πήγαινε στο (ii).

Μέσω του βήματος (iii) του αλγόριθμου ελέγχεται η τιμή του $\boldsymbol{\sigma}$, αφού το κάτω όριο $\boldsymbol{\sigma}'$ του Θεωρήματος 9.2.1 είναι άγνωστο.

Παράδειγμα 9.2.1: Θεωρούμε το πρόβλημα ελαχιστοποίησης της $f(\mathbf{x})=0.5(x_1-2)^2+0.5(x_2-2)^2$ υπό $c(\mathbf{x})=x_1+x_2-1=0$. Η αναλυτική λύση είναι $x_1^*=x_2^*=0.5$, $\lambda^*=-1.5$. Η συνάρτηση τιμωρίας με πολλαπλασιαστές είναι

$$\varphi(\mathbf{x},\lambda) = 0.5(x_1-2)^2+0.5(x_2-2)^2-\lambda(x_1+x_2-1)+\frac{1}{2}\sigma(x_1+x_2-1)^2 \qquad (9.2.11)$$

και από $\nabla\varphi=\mathbf{0}$ λαμβάνουμε

$$x_1(\lambda) = x_2(\lambda) = \frac{2+\lambda+\sigma}{1+2\sigma} \qquad (9.2.12)$$

ενώ οι ικανές συνθήκες ελαχιστοποίησης της φ ως προς $\mathbf{x}$ εκπληρούνται για $\sigma>-0.5$. Η $\psi(\lambda)$ είναι

$$\psi(\lambda)=\left(\frac{\lambda-3\sigma}{1+2\sigma}\right)^2-\lambda\frac{2\lambda+3}{1+2\sigma}+\frac{1}{2}\sigma\left(\frac{2\lambda+3}{1+2\sigma}\right)^2=\frac{4.5\sigma-3\lambda-\lambda^2}{1+2\sigma}. \qquad (9.2.13)$$

Για την μεγιστοποίηση της ψ(λ) έχουμε

$$\nabla\psi = -\frac{2\lambda+3}{1+2\sigma} = 0 \qquad (9.2.14)$$

εξής $\lambda^* = -1.5$, ανεξάρτητα του σ, ενώ $\nabla^2\psi < 0$ αν $\sigma > -0.5$. Δι' αντικαταστάσεως του λ^* στην (9.2.12) έχουμε τη λύση $x_1^* = x_2^* = 0.5$ ανεξάρτητα του σ. Η λύση του προβλήματος μπορεί επομένως να προσδιορισθεί ανεξάρτητα του $\sigma > -0.5$ από την ελαχιστοποίηση της φ ως προς **x** και την μεγιστοποίηση της ψ ως προς λ.

■

Παράδειγμα 9.2.2: Θεωρούμε την ελαχιστοποίηση της $f(\mathbf{x}) = -x_1 - x_2$ υπό $c(\mathbf{x}) = 1 - x_1^2 - x_2^2 = 0$ (βλ. Παράδ. 9.1.1) μέσω του αλγόριθμου συνάρτησης τιμωρίας με πολλαπλασιαστές. Ο Πίνακας 9.2.1 δίδει τα αποτελέσματα με εκκίνηση από $\lambda^{(1)}=0$ για τις εξής περιπτώσεις:

– Μέθοδος Newton (9.2.8) στο βήμα (vi) και σ=1.
– Μέθοδος Powell-Hestenes (9.2.10) στο βήμα (vi) και σ=1.
– Μέθοδος Powell-Hestenes (9.2.10) στο βήμα (vi) και σ=10.

Σε όλες τις περιπτώσεις έχουμε σύγκλιση $\lambda^{(k)} \to \lambda^* = 1/\sqrt{2}$ και $c^{(k)} \to 0$. Η μέθοδος Powell-Hestenes παρουσιάζει γραμμική σύγκλιση με ρυθμό 0.26 για σ=1 και 0.034 για σ=10. Αυτό υποδηλώνει επιτάχυνση της σύγκλισης του λ κατά δέκα για αύξηση κατά δέκα του βάρους σ.

■

β) Περιορισμοί Ανισότητας

Ο όρος τιμωρίας για περιορισμούς ανισότητας με μετάθεση μηδενικού σημείου είναι $\frac{1}{2}\sigma_i \min\{c_i - \theta_i, 0\}^2$, και με την (9.2.2) έχουμε $\frac{1}{2}\sigma_i \min\left\{c_i - \frac{\lambda_i}{\sigma_i}, 0\right\}^2$. Έτσι, η τιμή του όρου αυτού ελέγχεται από την ανισότητα $c_i \le \lambda_i/\sigma_i$ και η συνάρτηση τιμωρίας με πολλαπλασιαστές πάιρνει, στην περίπτωση περιορισμών ανισότητας, την μορφή

	Μέθοδος Newton (9.2.8), σ = 1		Μέθοδος Powell-Hestenes (9.2.10) σ = 1		σ = 10	
Επανάληψη	$\lambda^{(k)}$	$c^{(k)}$	$\lambda^{(k)}$	$c^{(k)}$	$\lambda^{(k)}$	$c^{(k)}$
1	0	−0.5651977	0	−0.5651977	0	−0.0684095
2	0.6672450	−0.0296174	0.5551977	−0.1068981	0.6840946	−0.0022228
3	0.7068853	−0.0001637	0.6720958	−0.0259956	0.7063222	−0.0000758
4	0.7071068	$-0.149_{10}-7$	0.6980914	−0.0066692	0.7070801	−0.0000026
5			0.7047606	−0.0017339	0.7071058	$-0.894_{10}-7$
6			0.7064945	−0.0004524		
7			0.7069469	−0.0001181		
8			0.7070650	−0.0000308		

Πίνακας 9.2.1: Αποτελέσματα του Παραδ. 9.2.2.

$$\varphi(\mathbf{x},\boldsymbol{\lambda},\boldsymbol{\sigma}) = f(\mathbf{x}) + \sum_i \begin{cases} -\lambda_i c_i + \frac{1}{2}\sigma_i c_i^2 & \text{αν } c_i \le \lambda_i/\sigma_i \\ -\frac{1}{2}\lambda_i^2/\sigma_i & \text{αν } c_i \ge \lambda_i/\sigma_i . \end{cases} \tag{9.2.15}$$

(όπου αφαιρέθηκε ο ανεξάρτητος του x όρος $\frac{1}{2}\sigma_i\theta_i^2 = \frac{1}{2}\lambda_i^2/\sigma_i$ για να υπάρχει αντιστοιχία με την (9.2.3)).

Το Θεώρημα 9.2.1 μπορεί να επεκταθεί στην περίπτωση της (9.2.15). Η συνάρτηση $\psi(\boldsymbol{\lambda})$ ορίζεται και εδώ μέσω της (9.2.5) και έχουμε

$$\psi(\boldsymbol{\lambda}) = \varphi(\mathbf{x}(\boldsymbol{\lambda}),\boldsymbol{\lambda}) \le \varphi(\mathbf{x}^*,\boldsymbol{\lambda})$$

$$= f^* + \sum_i \begin{cases} -c_i^*\lambda_i + \frac{1}{2}\sigma_i c_i^{*2} & c_i^* \le \lambda_i/\sigma_i \\ -\frac{1}{2}\lambda_i^2/\sigma_i & c_i^* \ge \lambda_i/\sigma_i \end{cases}$$

$$\le f^* + \sum_i \begin{cases} -\frac{1}{2}\sigma_i c_i^{*2} \\ -\frac{1}{2}\lambda_i^2/\sigma_i \end{cases} \le f^* = \varphi(\mathbf{x}^*,\boldsymbol{\lambda}^*) = \psi(\boldsymbol{\lambda}^*) \tag{9.2.16}$$

δηλαδή τελικά $\psi(\boldsymbol{\lambda}^*) \ge \psi(\boldsymbol{\lambda})$, και έτσι το $\boldsymbol{\lambda}^*$ είναι απόλυτο απεριόριστο μέγιστο της $\psi(\boldsymbol{\lambda})$ αν το $\mathbf{x}(\boldsymbol{\lambda})$ είναι απόλυτο απεριόριστο ελάχιστο της $\varphi(\mathbf{x}, \boldsymbol{\lambda})$. Ισχύει ακόμη, όπως στις (9.2.7), (9.2.8),

$$\nabla\psi(\boldsymbol{\lambda}) = -\min(c_i, \lambda_i/\sigma_i) \tag{9.2.17}$$

όπου $c_i = c_i(\mathbf{x}(\boldsymbol{\lambda}))$ και

$$\nabla^2\psi(\boldsymbol{\lambda}) = \begin{bmatrix} -\mathbf{A}^T\mathbf{W}_\sigma^{-1}\mathbf{A} & \mathbf{0} \\ \mathbf{0} & -\mathbf{S}^{-1} \end{bmatrix}_{\mathbf{x}(\lambda)} \tag{9.2.18}$$

όπου οι στήλες του $\mathbf{A}$ αντιστοιχούν στους δείκτες i για τους οποίους $c_i < \lambda_i/\sigma_i$ και οι στήλες του $\mathbf{S}^{-1}$ στους δείκτες i για τους οποίους $c_i \ge \lambda_i/\sigma_i$. Ένας τρόπος δημιουργίας ακολουθίας $\boldsymbol{\lambda}^{(k)}$ που απορρέει από την (9.2.17) είναι

$$\lambda_i^{(k+1)} = \lambda_i^{(k)} - \min\{\sigma_i c_i^{(k)},\ \lambda_i^{(k)}\}. \tag{9.2.19}$$

Η (9.2.19) μπορεί να χρησιμοποιηθεί στον αλγόριθμο του Κεφαλαίου 9.2.α) για προβλήματα με περιορισμούς ανισότητας χρησιμοποιώντας συγχρόνως $\|\nabla\psi\|_\infty$ αντί του (ή παράλληλα με το) $\|\mathbf{c}\|_\infty$.

Όπως και η μέθοδος συνάρτησης τιμωρίας, η μέθοδος συνάρτησης τιμωρίας με πολλαπλασιαστές μπορεί να χρησιμοποιηθεί σε συνδυασμό με άλλες μεθόδους, π.χ. για τον ευκολότερο χειρισμό γραμμικών περιορισμών κλπ.

9.3 Ακολουθητικός Τετραγωνικός Προγραμματισμός (ΑΤΠ)

Ένας αμεσότερος τρόπος επαναληπτικής επίλυσης του περιορισμένου προβλήματος είναι η προσέγγιση, σε κάθε επανάληψη, της αντικειμενικής συνάρτησης f(**x**) μέσω μιας τετραγωνικής συνάρτησης, και των περιορισμών **c**(**x**) μέσω γραμμικών περιορισμών. Με τον τρόπο αυτό έχουμε προς επίλυση σε κάθε επανάληψη ένα πρόβλημα ΤΠ που μπορεί να επιλυθεί με τις μεθόδους του Κεφαλαίου 7. Αν η ακρίβεια της προσέγγισης βελτιώνεται από επανάληψη σε επανάληψη, τότε θα υπάρχει σύγκλιση προς τη λύση $\mathbf{x}^*$, $\boldsymbol{\lambda}^*$ του αρχικού προβλήματος.

α) Περιορισμοί Ισότητας

Για την παρουσίαση της μεθόδου ΑΤΠ στην περίπτωση προβλημάτων με περιορισμούς ισότητας μόνον, μπορούμε να θεωρήσουμε καταρχήν την εφαρμογή της μεθόδου Newton για προσδιορισμό του στάσιμου σημείου της συνάρτησης Lagrange μέσω αριθμητικής επίλυσης του συστήματος μη γραμμικών εξισώσεων (6.1.7). Η μέθοδος φέρει για το λόγο αυτό και το όνομα **μέθοδος Newton-Lagrange.** Η ανάπτυξη Taylor του ∇L περί τα $\mathbf{x}^{(k)}$, $\boldsymbol{\lambda}^{(k)}$ δίδει

$$\nabla L(\mathbf{x}^{(k)}+\boldsymbol{\delta}\mathbf{x}, \boldsymbol{\lambda}^{(k)}+\boldsymbol{\delta}\boldsymbol{\lambda}) = \nabla L^{(k)} + \nabla^2 L^{(k)}[\boldsymbol{\delta}\mathbf{x}^T\ \boldsymbol{\delta}\boldsymbol{\lambda}^T]^T + \ldots \qquad (9.3.1)$$

Αγνοώντας τους όρους υψηλών βαθμών και θέτοντας την αριστερή πλευρά της (9.3.1) ίση με το μηδέν, όπως απαιτεί η (6.1.7), έχουμε

$$\nabla^2 L^{(k)} \begin{bmatrix} \boldsymbol{\delta}\mathbf{x} \\ \boldsymbol{\delta}\boldsymbol{\lambda} \end{bmatrix} = -\nabla L^{(k)} \qquad (9.3.2)$$

της οποίας η λύση δίδει τις διορθώσεις **δx**, **δλ**, και δεν είναι παρά η εφαρμογή της κλασσικής μεθόδου Newton στο σύστημα εξισώσεων (6.1.7). Με την (6.1.6) λαμβάνουμε από την (9.3.2)

$$\begin{bmatrix} \mathbf{W}^{(k)} & -\mathbf{A}^{(k)} \\ -\mathbf{A}^{(k)T} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{\delta}\mathbf{x} \\ \boldsymbol{\delta}\boldsymbol{\lambda} \end{bmatrix} = \begin{bmatrix} -\mathbf{g}^{(k)} + \mathbf{A}^{(k)}\boldsymbol{\lambda}^{(k)} \\ \mathbf{c}^{(k)} \end{bmatrix} \qquad (9.3.3)$$

όπου

$$\mathbf{W}^{(k)} = \nabla^2 f(\mathbf{x}^{(k)}) - \sum_i \lambda_i^{(k)} \nabla^2 c_i(\mathbf{x}^{(k)}) \qquad (9.3.4)$$

είναι ο Χεσιανός πίνακας $\nabla_x^2 L(\mathbf{x}^{(k)},\boldsymbol{\lambda}^{(k)})$ και $\mathbf{A}^{(k)}=\nabla\mathbf{c}^{(k)}$ είναι ο Ιακωβιανός πίνακας της συνάρτησης περιορισμών. Από την (9.3.3), ορίζοντας $\boldsymbol{\lambda}^{(k+1)} = \boldsymbol{\lambda}^{(k)} + \boldsymbol{\delta\lambda}$ και $\boldsymbol{\delta}^{(k)} = \boldsymbol{\delta}\mathbf{x}$, έχουμε

$$\begin{bmatrix} \mathbf{W}^{(k)} & -\mathbf{A}^{(k)} \\ -\mathbf{A}^{(k)T} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{\delta}^{(k)} \\ \boldsymbol{\lambda}^{(k+1)} \end{bmatrix} = \begin{bmatrix} -\mathbf{g}^{(k)} \\ \mathbf{c}^{(k)} \end{bmatrix} \tag{9.3.5}$$

από την οποία υπολογίζονται τα $\boldsymbol{\delta}^{(k)}$ και $\boldsymbol{\lambda}^{(k+1)}$, ενώ το $\mathbf{x}^{(k+1)}$ δίδεται από

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \boldsymbol{\delta}^{(k)}. \tag{9.3.6}$$

Η μέθοδος εκκινεί από τυχαία σημεία $\mathbf{x}^{(1)}$, $\boldsymbol{\lambda}^{(1)}$, και χρησιμοποιεί τις (9.3.5), (9.3.6) για τον προσδιορισμό μιας ακολουθίας $\{\mathbf{x}^{(k)}, \boldsymbol{\lambda}^{(k)}\}$.

Όπως και στο Κεφάλαιο 3, η εφαρμογή της μεθόδου Newton μπορεί να θεωρηθεί ως επίλυση ενός βοηθητικού προβλήματος ελαχιστοποίησης μιας τετραγωνικής συνάρτησης. Θεωρούμε το βοηθητικό πρόβλημα τετραγωνικού προγραμματισμού (ΤΠ)

$$\begin{aligned} & q^{(k)}(\boldsymbol{\delta}) \to \underset{\delta}{\mathrm{Min}} \\ & \text{υπό } \mathbf{l}^{(k)}(\boldsymbol{\delta}) = \mathbf{0} \end{aligned} \tag{9.3.7}$$

όπου

$$q^{(k)}(\boldsymbol{\delta}) = \frac{1}{2}\boldsymbol{\delta}^T\mathbf{W}^{(k)}\boldsymbol{\delta}+\mathbf{g}^{(k)T}\boldsymbol{\delta}+f^{(k)} \tag{9.3.8}$$

$$\mathbf{l}^{(k)}(\boldsymbol{\delta}) = \mathbf{A}^{(k)T}\boldsymbol{\delta}+\mathbf{c}^{(k)}. \tag{9.3.9}$$

Από τις (7.2.1), (7.2.2), οι συνθήκες πρώτου βαθμού για το (9.3.7) δίδονται όντως από την (9.3.5), δηλαδή το $\boldsymbol{\delta}^{(k)}$ είναι στάσιμο σημείο του (9.3.7) και μάλιστα, αν το $\mathbf{x}^{(k)}$ είναι αρκετά κοντά στο $\mathbf{x}^*$, το $\boldsymbol{\delta}^{(k)}$ είναι ελάχιστο του (9.3.7). Έχουμε λοιπόν τον ακόλουθο επαναληπτικό αλγόριθμο για δεδομένα σημεία εκκίνησης $\mathbf{x}^{(1)}$, $\boldsymbol{\lambda}^{(1)}$:

Για k=1, 2, ...

(i) Επίλυση του προβλήματος τετραγωνικού προγραμματισμού (9.3.7) (ή (9.3.11)) για προσδιορισμό του $\boldsymbol{\delta}^{(k)}$. Οι αντίστοιχοι πολλαπλασιαστές Lagrange είναι $\boldsymbol{\lambda}^{(k+1)}$. (9.3.10)

(ii) $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \boldsymbol{\delta}^{(k)}$.

Αν υπάρχει ένα μοναδικό ελάχιστο του (9.3.7) για κάθε k, τότε οι παραπάνω επαναλήψεις δίδουν ταυτόσημα αποτελέσματα με την μέθοδο Newton–Lagrange (9.3.5) και (9.3.6). Όμως οι επαναλήψεις του (9.3.7) είναι προτιμότερες, αφού οι (9.3.5), (9.3.6) μπορούν να οδηγήσουν σε κάποιο στάσιμο σημείο που δεν είναι ελάχιστο.

Το βοηθητικό πρόβλημα (9.3.7) έχει μια άμεση σχέση με το αρχικό πρόβλημα, αφού οι περιορισμοί του (9.3.7) μπορούν να θεωρηθούν προερχόμενοι από γραμμικοποίηση περί το $\mathbf{x}^{(k)}$

των μη γραμμικών περιορισμών $\mathbf{c}(\mathbf{x})=\mathbf{0}$, ενώ η τετραγωνική συνάρτηση του (9.3.7) είναι τετραγωνική προσέγγιση περί το $\mathbf{x}^{(k)}$ της μη γραμμικής αντικειμενικής συνάρτησης $f(\mathbf{x})$, που περιλαμβάνει όμως στο Χεσιανό πίνακα $\mathbf{W}^{(k)}$ (9.3.4) και τους περιορισμούς.

Παράδειγμα 9.3.1: Θεωρούμε ελαχιστοποίηση της $f(\mathbf{x}) = -x_1-x_2$ υπό $c(\mathbf{x}) = 1-x_1^2-x_2^2 = 0$ όπως και στα Παραδ. 9.1.1 και 9.2.2. Έχουμε από την (9.3.4)

$$\mathbf{W}^{(k)} = \begin{bmatrix} 2\boldsymbol{\lambda}^{(k)} & 0 \\ 0 & 2\boldsymbol{\lambda}^{(k)} \end{bmatrix}$$

και άρα το πρόβλημα ΤΠ (9.3.7)

$$q^{(k)}(\delta) = \lambda^{(k)}\delta_1^2 + \lambda^{(k)}\delta_2^2 - \delta_1 - \delta_2 \to \underset{\delta}{\mathrm{Min}}$$
$$\text{υπό} \quad -2\,x_1^{(k)}\delta_1 - 2x_2^{(k)}\delta_2 + 1 - x_1^{(k)2} - x_2^{(k)2} = 0.$$

Για τη λύση αυτού του προβλήματος ΤΠ έχουμε

$$L(\boldsymbol{\delta},\lambda^{(k+1)}) = \lambda^{(k)}\delta_1^2 + \lambda^{(k)}\delta_2^2 - \delta_1 - \delta_2 - \lambda^{(k+1)}(-2x_1^{(k)}\delta_1 - 2x_2^{(k)}\delta_2 + 1 - x_1^{(k)2} - x_2^{(k)2})$$

και άρα τις αναγκαίες συνθήκες βελτιστοποίησης

$$\partial L/\partial\delta_1 = 2\lambda^{(k)}\delta_1 - 1 + 2\lambda^{(k+1)}x_1^{(k)} = 0$$
$$\partial L/\partial\delta_2 = 2\lambda^{(k)}\delta_2 - 1 + 2\lambda^{(k+1)}x_2^{(k)} = 0$$
$$\partial L/\partial\lambda^{(k+1)} = -2x_1^{(k)}\delta_1 - 2x_2^{(k)}\delta_2 + 1 - x_1^{(k)2} - x_2^{(k)2} = 0$$

εξών λαμβάνουμε τη λύση

$$\delta_1^{(k)} = \frac{\lambda^{(k)}(1 - x_1^{(k)2} - x_2^{(k)2})x_1^{(k)} - x_2^{(k)}(x_1^{(k)} - x_2^{(k)})}{2\lambda^{(k)}(x_1^{(k)2} + x_2^{(k)2})}$$

$$\delta_2^{(k)} = \frac{\lambda^{(k)}(1 - x_1^{(k)2} - x_2^{(k)2})x_2^{(k)} - x_1^{(k)}(x_2^{(k)} - x_1^{(k)})}{2\lambda^{(k)}(x_1^{(k)2} + x_2^{(k)2})}$$

$$\lambda^{(k+1)} = \frac{x_1^{(k)} + x_2^{(k)} - \lambda^{(k)}(1 - x_1^{(k)2} - x_2^{(k)2})}{2(x_1^{(k)2} + x_2^{(k)2})}$$

και έχουμε

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \boldsymbol{\delta}^{(k)}.$$

Ο Πίνακας 9.3.1 δίδει τα αποτελέσματα των επαναλήψεων για $x_1^{(1)} = 0$, $x_2^{(1)} = \lambda^{(1)} = 0.1$. Η ακριβής λύση προσεγγίζεται κατά τα πρώτα 9 δεκαδικά ψηφία σε 8 επαναλήψεις.

Συγκρίνοντας τον Πίνακα 9.3.1 με τον Πίνακα 9.2.1 της μεθόδου Συνάρτησης Τιμωρίας με Πολλαπλασιαστές θα πρέπει να ληφθεί υπόψη ότι κάθε επανάληψη στον 9.2.1 εμπεριέχει

k	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**ακριβής λύση**
$x_1^{(k)}$	0	5	2.6	1.341	0.903	0.7106	0.7086	0.7071052	0.707106781	$1/\sqrt{2}$
$x_2^{(k)}$	0.1	5.05	2.525	1.418	0.842	0.737	0.7063	0.7071103	0.707106781	$1/\sqrt{2}$
$\lambda^{(k)}$	0.1	5.05	0.124	0.252	0.455	0.6507	0.7055	0.7071051	0.707106781	$1/\sqrt{2}$

Πίνακας 9.3.1: Επαναλήψεις ΑΤΠ στο Παράδ. 9.3.1.

αριθμητική επίλυση ενός απεριόριστου προβλήματος με μεθόδους του Πρώτου Μέρους, γεγονός που απαιτεί αρκετούς υπολογισμούς της αντικειμενικής συνάρτησης και του διανύσματος κλίσης, ενώ κάθε επανάληψη στον Πίνακα 9.3.1 εμπεριέχει επίλυση ενός προβλήματος ΤΠ με μεθόδους του Κεφαλαίου 7, η οποία απαιτεί λιγότερο υπολογιστικό φόρτο.

■

β) Περιορισμοί Ανισότητας

Η διαδικασία (9.3.10) μπορεί εύκολα να γενικευθεί ώστε να συμπεριλάβει μη γραμμικούς περιορισμούς ανισότητας $\mathbf{c}(\mathbf{x}) \geq \mathbf{0}$. Προς το σκοπό αυτό, επαυξάνουμε τον πίνακα **W** στην (9.3.4) ώστε να συμπεριλάβει τώρα και τους περιορισμούς ανισότητας και προσδένουμε στο βοηθητικό πρόβλημα ΤΠ (9.3.7) τους γραμμικοποιημένους περιορισμούς ανισότητας

$$\mathbf{l}_a^{(k)}(\boldsymbol{\delta}) \geq \mathbf{0}. \qquad (9.3.11)$$

Λόγω της επαναληπτικής επίλυσης ενός προβλήματος ΤΠ, η μέθοδος είναι γνωστή ως **Ακολουθητικός Τετραγωνικός Προγραμματισμός** (ΑΤΠ).

Παράδειγμα 9.3.2: Θεωρούμε το πρόβλημα του Παραδ. 6.2.1 (βλ. και Σχ. 5.1.3) και εφαρμόζουμε ΑΤΠ με τιμές εκκίνησης $\mathbf{x}^{(1)} = [0.5 \quad 1]^T$, $\boldsymbol{\lambda}^{(1)}=\mathbf{0}$. Ο Πίνακας 9.3.2 δίδει τα αποτελέσματα των επαναλήψεων. Επειδή $\boldsymbol{\lambda}^{(1)}=\mathbf{0}$, οι περιορισμοί δεν περιλαμβάνονται στον $\mathbf{W}^{(1)}$, και, αφού η αντικειμενική συνάρτηση είναι γραμμική, έχουμε $\mathbf{W}^{(1)} = \mathbf{0}$, και άρα το πρόβλημα της πρώτης επανάληψης είναι το ακόλουθο πρόβλημα γραμμικού προγραμματισμού

$$q^{(1)}(\boldsymbol{\delta}) = -\delta_1 - \delta_2 - 1.5 \rightarrow \underset{\boldsymbol{\delta}}{\text{Min}}$$

υπό

$$-\delta_1 + \delta_2 + 0.75 \geq 0$$

$$-\delta_1 - 2\delta_2 - 0.25 \geq 0.$$

k	$x_1^{(k)}$	$x_2^{(k)}$	$\lambda_1^{(k)}$	$\lambda_2^{(k)}$	$c_1^{(k)}$	$c_2^{(k)}$
1	$\frac{1}{2}$	1	0	0	$\frac{3}{4}$	$-\frac{1}{4}$
2	$\frac{11}{12}$	$\frac{2}{3}$	$\frac{1}{3}$	$\frac{2}{3}$	-0.173611	-0.284722
3	0.747120	0.686252	0	0.730415	0.128064	-0.029130
4	0.708762	0.706789	0	0.706737	0.204445	-0.001893
5	0.707107	0.707108	0	0.707105	0.207108	$-0.28_{10}-5$

Πίνακας 9.3.2: Αποτελέσματα επαναλήψεων του Παραδ. 9.3.2.

Επιλύοντας το πρόβλημα αυτό έχουμε $\delta_1^{(1)}=5/12$, $\delta_2^{(1)}=-1/3$, δηλαδή το νέο επαναληπτικό $\mathbf{x}^{(2)}=\mathbf{x}^{(1)}+\boldsymbol{\delta}^{(1)}$, είναι η τομή των δύο περιορισμών, γραμμικοποιημένων περί το $\mathbf{x}^{(1)}$. Βέβαια υπάρχουν πολλά $\mathbf{x}^{(1)}$ για τα οποία το πρόβλημα γραμμικού προγραμματισμού της πρώτης επανάληψης δεν θα ήταν φραγμένο, πράγμα που υποδηλώνει ότι η μέθοδος ΑΤΠ μπορεί να έχει δυσκολίες για $\mathbf{x}^{(k)}$, $\boldsymbol{\lambda}^{(k)}$ μακράν των $\mathbf{x}^*$, $\boldsymbol{\lambda}^*$. Εδώ όμως το πρόβλημα γραμμικού προγραμματισμού είναι καλώς ορισμένο, και έχουμε $\boldsymbol{\lambda}^{(2)} = [1/3 \;\; 2/3]^T$, πράγμα που υποδηλώνει ότι οι γραμμικοποιημένοι περιορισμοί είναι ενεργοί. Έτσι έχουμε στη δεύτερη επανάληψη

$$\mathbf{W}^{(2)} = \mathbf{0} - \frac{1}{3}\begin{bmatrix} -2 & 0 \\ 0 & 0 \end{bmatrix} - \frac{2}{3}\begin{bmatrix} -2 & 0 \\ 0 & -2 \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 0 & 4/3 \end{bmatrix} > \mathbf{0}.$$

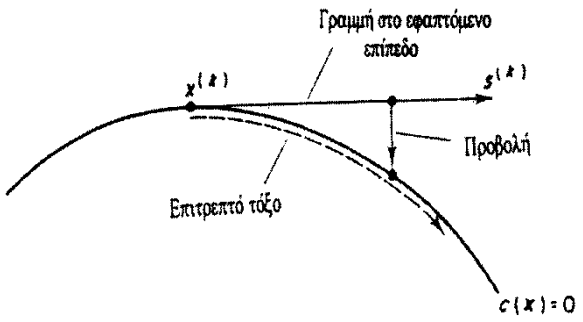
Επίλυση του αντίστοιχου προβλήματος ΤΠ οδηγεί σε απενεργοποίηση του $l_1^{(2)} \geq 0$ αφού $\lambda_1^{(3)} = 0$. Έτσι έχει ήδη προσδιορισθεί το σωστό ενεργό σύνολο και παρατηρούμε την τυπική γρήγορη σύγκλιση της μεθόδου Newton στις υπόλοιπες επαναλήψεις.

■

Η μέθοδος ΑΤΠ, όταν λειτουργεί ομαλά, θεωρείται η αποτελεσματικότερη μέθοδος επίλυσης γενικών προβλημάτων μη γραμμικού προγραμματισμού. Έχουν προταθεί στην βιβλιογραφία και έχουν υλοποιηθεί σε αντίστοιχα προγράμματα πολλές επεκτάσεις και παραλλαγές της βασικής μεθόδου ΑΤΠ, που περιγράφηκε εδώ, με στόχο την αύξηση της ευρωστίας, της αποτελεσματικότητας και του πεδίου εφαρμογής της.

9.4 Μη Γραμμική Απαλοιφή και Μέθοδοι Επιτρεπτής Κατεύθυνσης

Οι **μέθοδοι επιτρεπτής κατεύθυνσης** επιχειρούν τη γενίκευση των αντίστοιχων διαδικασιών που αναπτύχθηκαν στα Κεφάλαια 7 και 8, επιχειρούν δηλαδή τον προσδιορισμό τοπικού ελάχιστου κατά μήκος επιτρεπτών (μη γραμμικών) τόξων. Ο χειρισμός περιορισμών ανισότητας επιτυγχάνεται με την μέθοδο ενεργού συνόλου, όπως στα Κεφάλαια 7.3 και 8.2, και έτσι στρέφουμε την προσοχή μας εδώ κατά κύριο λόγο στους μη γραμμικούς περιορισμούς ισότητας. Η παρουσία μη γραμμικών περιορισμών δεν επιτρέπει επιτρεπτή αναζήτηση επί γραμμής. Έτσι, μια παραστατική επεξήγηση της ακολουθούμενης διαδικασίας είναι ότι υπολογίζεται σε κάθε επιτρεπτό σημείο $\mathbf{x}^{(k)}$ μια κατεύθυνση αναζήτησης $\mathbf{s}^{(k)}$ στο εφαπτόμενο επίπεδο, από την οποία λαμβάνεται ένα επιτρεπτό τόξο μέσω προβολής κάθε σημείου της προκύπτουσας γραμμής σε αντίστοιχο σημείο της επιτρεπτής περιοχής (Σχ. 9.4.1). Ακολουθεί



Σχήμα 9.4.1: Αναζήτηση σε επιτρεπτή κατεύθυνση.

αναζήτηση κατά μήκος του επιτρεπτού τόξου με στόχο την μείωση της f(**x**). Μια άλλη δυνατή προσέγγιση είναι η απαλοιφή περιορισμών μέσω αριθμητικής επίλυσης ενός συστήματος μη γραμμικών εξισώσεων. Μπορεί όμως να αποδειχθεί ότι η προσέγγιση αυτή είναι ισοδύναμη με την μέθοδο επιτρεπτής κατεύθυνσης αφού η προαναφερθείσα προβολή αντιστοιχεί πράγματι στην επίλυση ενός συστήματος μη γραμμικών εξισώσεων.

Είδαμε στο Κεφάλαιο 7.1 ότι οι μέθοδοι γενικευμένης απαλοιφής γραμμικών περιορισμών ανάγονται σε αντίστοιχη επιλογή του **V** στην (7.1.20). Ένας άλλος τρόπος ερμηνείας της γενικευμένης απαλοιφής είναι ότι γίνεται γραμμικός μετασχηματισμός σε νέες μεταβλητές $\mathbf{y}\in R^n$ μέσω

$$\mathbf{y}=\begin{bmatrix}\mathbf{y}_1\\ \mathbf{y}_2\end{bmatrix}=[\mathbf{A}\ \ \mathbf{V}]^T\mathbf{x}-\begin{bmatrix}\mathbf{d}\\ \mathbf{0}\end{bmatrix}=\begin{bmatrix}\mathbf{A}^T\mathbf{x}-\mathbf{d}\\ \mathbf{V}^T\mathbf{x}\end{bmatrix}$$

με τον καταμερισμό $\mathbf{y}_1\in R^m$, $\mathbf{y}_2\in R^{n-m}$. Επειδή ο [**A V**] είναι κανονικός, ο μετασχηματισμός είναι αμφιμονοσήμαντος. Έτσι έχουμε απαλοιφή κρατώντας $\mathbf{y}_1=\mathbf{0}$, ικανοποιώντας δηλαδή τους περιορισμούς, και ελαχιστοποιώντας ως προς $\mathbf{y}_2$ που αντιστοιχεί στο **y** της (7.1.11). (Το τελευταίο γίνεται φανερό αν πολλαπλασιασθεί η (7.1.11) με $\mathbf{V}^T$ και γίνει χρήση των $\mathbf{V}^T\mathbf{Y}=\mathbf{0}$ και $\mathbf{V}^T\mathbf{Z}=\mathbf{I}$ που προκύπτουν από την (7.1.20).) Μια ανάλογη προσέγγιση είναι δυνατή σε μη γραμμικούς περιορισμούς. Επειδή θα χρησιμοποιούμε σε κάθε επανάληψη το αντίστοιχο $\mathbf{x}^{(k)}$ ως αρχή των αξόνων για τις ελεύθερες μεταβλητές $\mathbf{y}_2$, δηλαδή θα χρησιμοποιούμε τον μετασχηματισμό (8.1.3) αντί του (7.1.11), έχουμε $\mathbf{y}_2=\mathbf{V}^T\boldsymbol{\delta}$. Έτσι, υποθέτοντας ότι $\mathbf{x}^{(k)}$ είναι ένα επιτρεπτό σημείο όπου ο $\mathbf{A}^{(k)}$ είναι κανονικός, έχουμε, κατ' αναλογίαν με τον παραπάνω γραμμικό μετασχηματισμό, τον μη γραμμικό μετασχηματισμό

$$\mathbf{y}=\begin{bmatrix}\mathbf{y}_1\\ \mathbf{y}_2\end{bmatrix}=\begin{bmatrix}\mathbf{c}(\mathbf{x}^{(k)}+\boldsymbol{\delta})\\ \mathbf{V}^T\boldsymbol{\delta}\end{bmatrix} \qquad (9.4.1)$$

όπου ο **V** επιλέγεται έτσι ώστε ο [$\mathbf{A}^{(k)}$ **V**] να είναι κανονικός, οπότε ο μετασχηματισμός είναι αμφιμονοσήμαντος σε μια περιοχή περί το $\mathbf{x}^{(k)}$. Πράγματι, όπως δείχνει η (9.4.1), ορίζονται νέες μεταβλητές **y** σε κάθε επανάληψη, έτσι ώστε ο μετασχηματισμός να είναι πάντοτε καλώς ορισμένος.

Για κάθε δεδομένη τιμή $\mathbf{y}_2$ των ελεύθερων μεταβλητών, απαιτείται το αντίστοιχο $\boldsymbol{\delta}\in R^n$ που επιλύει την (9.4.1) με $\mathbf{y}_1=\mathbf{0}$. Αυτό μπορεί να επιτευχθεί μέσω εσωτερικών επαναλήψεων r, π.χ. με την μέθοδο Newton-Raphson. Θεωρούμε λοιπόν αριθμητική επίλυση των εξισώσεων

$$\begin{bmatrix} \mathbf{c}(\mathbf{x}^{(k)}+\boldsymbol{\delta}) \\ \mathbf{V}^T\boldsymbol{\delta}-\mathbf{y}_2 \end{bmatrix} = \mathbf{0}. \qquad (9.4.2)$$

Για την αριθμητική επίλυση της παραπάνω εξίσωσης (με δεδομένο πάντοτε $\mathbf{y}_2$) εκκινούμε από

$$\boldsymbol{\delta}^{(1)} = \mathbf{Z}\mathbf{y}_2$$

εξού (με $\mathbf{V}^T\mathbf{Z} = \mathbf{I}$) συνεπάγεται $\mathbf{y}_2=\mathbf{V}^T\boldsymbol{\delta}^{(1)}$ και άρα η κάτω συνιστώσα της προς επίλυση εξίσωσης ικανοποιείται από το $\boldsymbol{\delta}^{(1)}$. Ο Ιακωβιανός πίνακας της (9.4.2) είναι προφανώς $[\mathbf{A}\ \mathbf{V}]$ και άρα η πρώτη επανάληψη Newton-Raphson δίδει (με χρήση της (7.1.20))

$$\boldsymbol{\delta}^{(2)} = \boldsymbol{\delta}^{(1)} - [\mathbf{A}\quad \mathbf{V}]^{-T}\begin{bmatrix} \mathbf{c}(\mathbf{x}^{(k)}+\boldsymbol{\delta}^{(1)}) \\ \mathbf{0} \end{bmatrix} = \boldsymbol{\delta}^{(1)} - \mathbf{Y}\mathbf{c}(\mathbf{x}^{(k)}+\boldsymbol{\delta}^{(1)}).$$

Τότε όμως ισχύει $\mathbf{V}^T\boldsymbol{\delta}^{(2)} = \mathbf{V}^T\boldsymbol{\delta}^{(1)} = \mathbf{y}_2$ (αφού $\mathbf{V}^T\mathbf{Y}=\mathbf{0}$), και παρομοίως $\mathbf{V}^T\boldsymbol{\delta}^{(r)} = \mathbf{y}_2$ για τα επόμενα επαναληπτικά, δηλαδή τα $\boldsymbol{\delta}^{(r)}$ ικανοποιούν πάντοτε την κάτω συνιστώσα της προς επίλυση εξίσωσης. Συμπερασματικά έχουμε λοιπόν τις ακόλουθες r-επαναλήψεις

$$\boldsymbol{\delta}^{(1)} = \mathbf{Z}\mathbf{y}_2$$

$$\boldsymbol{\delta}^{(r+1)} = \boldsymbol{\delta}^{(r)} - [\mathbf{A}\quad \mathbf{V}]_{(r)}^{-T}\begin{bmatrix} \mathbf{c}^{(r)} \\ \mathbf{0} \end{bmatrix} = \boldsymbol{\delta}^{(r)} - \mathbf{Y}\mathbf{c}^{(r)}, \quad r\geq 1 \qquad (9.4.3)$$

όπου οι $\mathbf{A}$, $\mathbf{Y}$, $\mathbf{c}^{(r)}$ υπολογίζονται στο $\mathbf{x}^{(k)}+\boldsymbol{\delta}^{(r)}$. Μετά τη σύγκλιση των ως άνω r-επαναλήψεων, έχει προσδιορισθεί η λύση $\boldsymbol{\delta}=\boldsymbol{\delta}^{(r+1)}$ της μη γραμμικής εξίσωσης (9.4.2). Τόσο το $\boldsymbol{\delta}$ όσο και το $\mathbf{x}$ ($=\mathbf{x}^{(k)}+\boldsymbol{\delta}$) είναι κατά βάση συναρτήσεις του $\mathbf{y}_2$, αφού προκύπτουν από την επίλυση της (9.4.2) για κάποιο δεδομένο $\mathbf{y}_2$, μέσω των επαναλήψεων (9.4.3). Στην πράξη είναι απλούστερο στην (9.4.3) να υπολογίζεται ο $\mathbf{A}$, και άρα ο $\mathbf{Y}$, στο $\mathbf{x}^{(k)}$, αν και αυτό μεταβάλει την τετραγωνική σύγκλιση των r-επαναλήψεων. Μπορεί να αποδειχθεί ότι οι r-επαναλήψεις συγκλίνουν αν η $\boldsymbol{\delta}^{(1)}$ είναι αρκούντως ακριβής προσέγγιση της λύσης, πράγμα που συμβαίνει για αρκούντως μικρό $\mathbf{y}_2$. Στο Σχ. 9.4.1, το $\boldsymbol{\delta}^{(1)}$ είναι ένα βήμα στο εφαπτόμενο επίπεδο, ενώ οι επαναλήψεις (9.4.3) αντιστοιχούν σε προβολή του αρχικού $\mathbf{x}^{(k)}+\boldsymbol{\delta}^{(1)}$ έτσι ώστε το τελικό $\mathbf{x}^{(k)}+\boldsymbol{\delta}$ να ικανοποιεί τους μη γραμμικούς περιορισμούς. Η κατεύθυνση της προβολής εξαρτάται βέβαια από την επιλογή του $\mathbf{Y}$, δηλαδή του $\mathbf{V}$.

Αυτή η βασική διαδικασία μπορεί να οδηγήσει σε μέθοδο αναζήτησης, ορίζοντας $\mathbf{p}^{(k)}\in R^{n-m}$ ως κατεύθυνση αναζήτησης στο χώρο των ελεύθερων μεταβλητών και παίρνοντας $\mathbf{y}_2=\alpha\cdot\mathbf{p}^{(k)}$ ως βήμα σ' αυτή την κατεύθυνση. Τότε το $\boldsymbol{\delta}$ ($=\boldsymbol{\delta}(\alpha)$) υπολογίζεται από τις επαναλήψεις (9.4.3) και ορίζει ένα επιτρεπτό τόξο $\mathbf{x}(\alpha) = \mathbf{x}^{(k)}+\boldsymbol{\delta}(\alpha)$. Η κατεύθυνση του τόξου στο $\mathbf{x}^{(k)}$ είναι η επιτρεπτή κατεύθυνση $\mathbf{s}^{(k)}=\mathbf{Z}\mathbf{p}^{(k)}$ (βλ. Σχ. 9.4.1). Με τον τρόπο αυτό είναι δυνατό να προσδιορισθεί ένα $\alpha^{(k)}$ που ελαχιστοποιεί την $f(\mathbf{x}^{(k)}+\boldsymbol{\delta})$ μέσω αναζήτησης επί του επιτρεπτού τόξου.

Μια μέθοδος αυτού του τύπου είναι η **μέθοδος γενικευμένης μειωθείσας κλίσης** των Abadie και Carpentier (1969) που βασίζεται στην άμεση απαλοιφή μεταβλητών. Το $\mathbf{x}$ καταμερίζεται σε $\mathbf{x}_1 \in R^m$ και $\mathbf{x}_2 \in R^{n-m}$ και οι μεταβλητές $\mathbf{x}_2$ είναι οι ελεύθερες μεταβλητές $\mathbf{y}_2$. Ο πίνακας $\mathbf{Y}$ είναι $[\mathbf{A}_1^{-1} \;\; \mathbf{0}]^T$ (βλ. (7.1.22)) και ο $\mathbf{Z}$ είναι $[-\mathbf{A}_2\mathbf{A}_1^{-1} \;\; \mathbf{I}]^T$. Ένας τρόπος περιγραφής των επαναλήψεων προβολής (9.4.3) είναι τότε να θεωρηθεί η $\{\mathbf{x}^{(k+1,r)}\}$ μια ακολουθία εκτίμησης του $\mathbf{x}^{(k+1)}$ που ορίζεται από

$$\begin{aligned} \mathbf{x}^{(k+1,1)} &= \mathbf{x}^{(k)} + \alpha^{(k)}\mathbf{Z}\mathbf{p}^{(k)} \\ \mathbf{x}_1^{(k+1,r+1)} &= \mathbf{x}_1^{(k+1,r)} - \mathbf{A}_1^{-T}\mathbf{c}^{(k+1,r)} \;, \quad r=1,2,... \end{aligned} \tag{9.4.4}$$

ενώ το $\mathbf{x}_2^{(k+1)} = \mathbf{x}_2^{(k+1,r)}$ παραμένει σταθερό, πράγμα που δείχνει ότι οι μεταβλητές $\mathbf{x}_1$ επιλέγονται έτσι ώστε να απαλειφθούν οι περιορισμοί.

Στρέφουμε τώρα την προσοχή μας στον υπολογισμό της κατεύθυνσης $\mathbf{s}^{(k)}$ $(=\mathbf{Z}\mathbf{p}^{(k)})$ του επιτρεπτού τόξου. Η αντικειμενική συνάρτηση είναι $f_y(\mathbf{y}) = f_x(\mathbf{x})$ όπου τα $\mathbf{x}$ και $\mathbf{y}$ συσχετίζονται μέσω της (9.4.1). Όπως στο Κεφάλαιο 3.3, έχουμε για τις πρώτες παραγώγους

$$\mathbf{g}_x = [\mathbf{A} \;\; \mathbf{V}] \begin{bmatrix} \mathbf{g}_{y_1} \\ \mathbf{g}_{y_2} \end{bmatrix} \tag{9.4.5}$$

και μέσω της (7.1.20) έχουμε $\mathbf{g}_{y_2} = \mathbf{Z}^T\mathbf{g}_x$, το μειωθέν διάνυσμα κλίσης. Έτσι, αν θέλουμε να εφαρμόσουμε κατεύθυνση μέγιστης κατάβασης, έχουμε για την μειωθείσα κατεύθυνση $\mathbf{y}_2=\alpha\mathbf{p}^{(k)}$ όπου $\mathbf{p}^{(k)} = -\mathbf{Z}^T\mathbf{g}^{(k)}$, και άρα την κατεύθυνση του επιτρεπτού τόξου $\mathbf{s}^{(k)} = -\mathbf{Z}\mathbf{Z}^T\mathbf{g}^{(k)}$.

Για την εφαρμογή της μεθόδου Newton στον προσδιορισμό κατεύθυνσης αναζήτησης χρειάζονται δεύτερες παράγωγοι που λαμβάνονται από παραγώγιση της (9.4.5) (βλ. και (3.3.3))

$$\mathbf{G}_x = \sum_{i=1}^{m} \nabla^2 c_i \partial f_y / \partial y_i + [\mathbf{A} \;\; \mathbf{V}]\, \mathbf{G}_y\, [\mathbf{A} \;\; \mathbf{V}]^T \tag{9.4.6}$$

και άρα ο μειωθείς Χεσιανός στο $\mathbf{x}^{(k)}$ (δηλαδή για $\mathbf{y}_2=\mathbf{0}$) είναι

$$\nabla^2_{y_2} f_y(\mathbf{0}) = \mathbf{Z}^T\mathbf{W}^{(k)}\mathbf{Z} \tag{9.4.7}$$

όπου $\mathbf{W}^{(k)} = \mathbf{G}^{(k)} - \sum \lambda_i^{(k)} \nabla^2 c_i^{(k)}$ με $\boldsymbol{\lambda}^{(k)} = \mathbf{Y}^T\mathbf{g}^{(k)}$. Έτσι, η μέθοδος Newton υπολογίζει το $\mathbf{y}_2^{(k)}$ (χωρίς αναζήτηση επί γραμμής) μέσω επίλυσης του συστήματος

$$\mathbf{Z}^T\mathbf{W}^{(k)}\mathbf{Z}\mathbf{y}_2 = -\mathbf{Z}^T\mathbf{g}^{(k)} \tag{9.4.8}$$

που χρησιμοποιείται με την (9.4.3). Ανάλογη εφαρμογή άλλων μεθόδων (σχεδόν–Newton, συζυγών κλίσεων) είναι δυνατή κατά παρόμοιο τρόπο.

Παράδειγμα 9.4.1: Θεωρούμε το πρόβλημα του Παραδ. 9.1.1 και εφαρμόζουμε τη μέθοδο επιτρεπτής κατεύθυνσης με χρήση της μεθόδου Newton (9.4.8) για τον προσδιορισμό των επαναληπτικών σημείων. Για τον μετασχηματισμό των μεταβλητών γίνεται χρήση ορθογώνιας

παραγοντοποίησης με $\mathbf{Z}^T\mathbf{A}=\mathbf{0}$, $\mathbf{Z}^T\mathbf{Z}=\mathbf{I}$, $\mathbf{Y}=\mathbf{A}^+$. Ο Πίνακας 9.4.1 δίδει τα αποτελέσματα των επαναλήψεων k (από (9.4.8)) και r (από (9.4.3)) εκκινώντας από το επιτρεπτό σημείο $x_1^{(1)}=0.8,\ x_2^{(1)}=0.6$.

■

Το βασικό μειονέκτημα της μεθόδου επιτρεπτής κατεύθυνσης έγκειται στην ύπαρξη μιας επιπλέον, εσωτερικής r-επανάληψης και ειδικότερα στη λανθάνουσα δυνατότητα απόκλισης, πράγμα που δεν συναντάται σε περιπτώσεις γραμμικών περιορισμών (Κεφάλαια 7, 8) για την ίδια μέθοδο. Επιπλέον, δυσκολίες μπορούν να παρουσιασθούν με την χρήση της μεθόδου ενεργού συνόλου (που παρουσιάσθηκε ήδη στα Κεφάλαια 7.3 και 8.2) κατά τον χειρισμό μη γραμμικών περιορισμών ανισότητας. Συγκεκριμένα, οι τύποι (7.3.2) και (8.2.2) για τον προσδιορισμό των ενεργοποιούμενων περιορισμών κατά μήκος μιας κατεύθυνσης αναζήτησης δεν ισχύουν για μη γραμμικούς περιορισμούς, για τους οποίους απαιτούνται επιπλέον εσωτερικές επαναληπτικές διαδικασίες προκειμένου να εντοπισθούν σημεία τομής του τόξου αναζήτησης με μη ενεργούς μη γραμμικούς περιορισμούς.

Εξ αιτίας των δυσκολιών αυτών, η μέθοδος επιτρεπτής κατεύθυνσης βρίσκει μεν εφαρμογή (και είναι αποτελεσματικότατη) σε ειδικά προβλήματα ελαχιστοποίησης. Επί παραδείγματι, η απαλοιφή περιορισμών ισότητας σε προβλήματα βέλτιστου ελέγχου είναι εύκολη και δεν απαιτεί εσωτερικές επαναλήψεις, λόγω της ειδικής μορφής των περιοριμών υπό μορφή καταστατικών εξισώσεων. Σε πολλές άλλες εφαρμογές, μπορεί όμως η μέθοδος να οδηγήσει σε σοβαρές πρακτικές δυσκολίες, πράγμα που μειώνει την αξία της σαν **γενικής** μεθόδου επίλυσης προβλημάτων μη γραμμικού προγραμματισμού.

Προβλήματα

9.1 Να επιλυθεί το Πρόβλημα 6.6 μέσω της μεθόδου συνάρτησης τιμωρίας. Προσδιορίστε τη λύση:

a) Αναλυτικά

b) Με εφαρμογή της μεθόδου Newton.

Λύση

k	r	$x_1^{(k,r)}$	$x_2^{(k,r)}$	$\lambda^{(k)}$	$y_2^{(k)}$	$c^{(r)}$
1	—	0.8	0.6	0.7	−0.142857	
2	1	0.714286	0.714286			−0.020408
	2	0.706122	0.708163			−0.000104
	3	0.706081	0.708132			−0.000001
	4	0.706080	0.708132			$-1_{10}-8$
2	—			0.707106	0.001451	
3	1	0.707108	0.707108			−0.000002
	2	0.707107	0.707107			$-4_{10}-13$
3	—			0.707107	0	

Πίνακας 9.4.1: Αποτελέσματα επαναλήψεων για το Παραδ. 9.4.1.

$$\varphi(\mathbf{x}, \sigma) = \sum_{i=1}^{3}(a_i + b_i x_i^2) + \frac{1}{2}\sigma\,(X - x_1 - x_2 - x_3)^2 \to \underset{\mathbf{x}}{\mathrm{Min}}$$

a) $\partial\varphi/\partial x_i = 2b_i x_i - \sigma(X - x_1 - x_2 - x_3) = 0, \quad i = 1, 2, 3$

$$\Rightarrow \begin{bmatrix} 2b_1+\sigma & \sigma & \sigma \\ \sigma & 2b_2+\sigma & \sigma \\ \sigma & \sigma & 2b_3+\sigma \end{bmatrix} \mathbf{x} = \sigma X \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}.$$

$$\Rightarrow x_1 = \frac{b_2 b_3 X}{\dfrac{b_1 b_2 b_3}{\sigma} + N}, \quad x_2 \text{ και } x_3 \text{ ανάλογα.}$$

$\sigma \to \infty$: x_1^*, x_2^*, x_3^* όπως στο Πρόβλημα 6.6.

b) Μέθοδος Newton: $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} - \nabla^2\varphi(\mathbf{x}^{(k)})^{-1}\,\nabla\varphi(\mathbf{x}^{(k)})$

$$\nabla^2\varphi = \sigma \begin{bmatrix} 2\dfrac{b_1}{\sigma}+1 & 1 & 1 \\ 1 & 2\dfrac{b_2}{\sigma}+1 & 1 \\ 1 & 1 & 2\dfrac{b_3}{\sigma}+1 \end{bmatrix}$$

$$(\nabla^2\varphi)^{-1} = \frac{1}{2\sigma}\frac{1}{\det}\begin{bmatrix} \dfrac{b_2 b_3}{\sigma^2} + \dfrac{b_3}{\sigma} - 1 & -\dfrac{b_3}{\sigma}+1 & -\dfrac{b_2}{\sigma} \\ -\dfrac{b_3}{\sigma}+1 & \dfrac{b_1 b_3}{\sigma^2} + \dfrac{b_2}{\sigma} - 1 & -\dfrac{b_1}{\sigma} \\ -\dfrac{b_2}{\sigma} & -\dfrac{b_1}{\sigma} & \dfrac{b_1 b_2}{\sigma^2} + \dfrac{b_1}{\sigma} + \dfrac{b_2}{\sigma} \end{bmatrix}$$

Μετά από τις αντίστοιχες πράξεις αποδεικνύεται σύγκλιση σε μια επανάληψη (τετραγωνική συνάρτηση).

9.2 Θεωρούμε την ελαχιστοποίηση της $f(\mathbf{x}) = \frac{1}{2}(x_1-1)^2 + \frac{1}{2}(x_2-2)^2$ υπό $2 - x_1 - x_2 \geq 0$.

a) Επιλύστε αναλυτικά το πρόβλημα μέσω των συνθηκών ελαχιστοποίησης.

b) Επιλύστε αναλυτικά το πρόβλημα μέσω της συνάρτησης τιμωρίας. Υπολογίστε τις λύσεις $x_1^*(\sigma)$, $x_2^*(\sigma)$ και τα αντίστοιχα όρια για $\sigma \to \infty$.

c) Επιλύστε αναλυτικά το πρόβλημα μέσω της συνάρτησης τιμωρίας με πολλαπλασιαστές. Υπολογίστε τα $x_1(\lambda,\sigma)$, $x_2(\lambda,\sigma)$, $\psi(\lambda)$, καθώς και την βέλτιστη λύση.

d) Διερευνήστε την σύγκλιση του αλγορίθμου συνάρτησης τιμωρίας με πολλαπλασιαστές.

Λύση

a) $L(\mathbf{x},\lambda) = \frac{1}{2}(x_1-1)^2 + \frac{1}{2}(x_2-2)^2 - \lambda(2-x_1-x_2)$.

Η περίπτωση μη ενεργοποίησης του περιορισμού οδηγεί σε μη επιτρεπτή λύση, άρα θεωρούμε $2-x_1-x_2 = 0 \Rightarrow$

$$\left.\begin{array}{l} x_1 = 2 - x_2 \\ \dfrac{\partial L}{\partial x_1} = x_1 - 1 + \lambda = 0 \\ \dfrac{\partial L}{\partial x_2} = x_2 - 2 + \lambda = 0 \end{array}\right\} \Rightarrow \begin{array}{l} x_1^* = 0.5 \\ x_2^* = 1.5 \\ \lambda^* = 0.5 > 0. \end{array}$$

$\nabla^2 L = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} > \mathbf{0} \Rightarrow$ ελάχιστο (πρόβλημα κυρτού προγραμματισμού)

b) $\varphi(\mathbf{x},\sigma) = \frac{1}{2}(x_1-1)^2 + \frac{1}{2}(x_2-2)^2 + \frac{1}{2}\sigma\cdot\min\{0, 2-x_1-x_2\}^2$

$\frac{\partial\varphi}{\partial x_1} = x_1-1+\sigma\cdot\max\{0, x_1 + x_2 - 2\} = 0$

$\frac{\partial\varphi}{\partial x_2} = x_2-2+\sigma\cdot\max\{0, x_1 + x_2 - 2\} = 0$

1η περίπτωση: $x_1 + x_2 - 2 < 0$

$\Rightarrow x_1^* = 1,\ x_2^* = 2$: αντίφαση αφού $x_1^* + x_2^* - 2 = 1 > 0$

2η περίπτωση: $x_1+x_2-2 > 0$

$$\Rightarrow \left.\begin{array}{l} x_1 - 1 + \sigma(x_1 + x_2 - 2) = 0 \\ x_2 - 2 + \sigma(x_1 + x_2 - 2) = 0 \end{array}\right\} \Rightarrow x_1(\sigma) = \frac{1+\sigma}{1+2\sigma},\ x_2(\sigma) = \frac{2+3\sigma}{1+2\sigma}$$

$\nabla^2\varphi = \begin{bmatrix} 1+\sigma & \sigma \\ \sigma & 1+\sigma \end{bmatrix} > \mathbf{0}$ αν $\sigma > -1$ και $(1+\sigma)^2 - \sigma^2 = 1 + 2\sigma > 0 \Rightarrow \sigma > -0.5$

$x_1(\sigma\to\infty) = x_1^* = 0.5$

$x_2(\sigma\to\infty) = x_2^* = 1.5.$

c) $\varphi(\mathbf{x},\lambda,\sigma) = \frac{1}{2}(x_1-1)^2 + \frac{1}{2}(x_2-1)^2$

$$+ \begin{cases} -\lambda(2-x_1-x_2) + \frac{1}{2}\sigma(2-x_1-x_2)^2 & \text{αν} \quad 2-x_1-x_2 \le \lambda/\sigma \\ -\frac{1}{2}\lambda^2/\sigma & \text{αν} \quad 2-x_1-x_2 \ge \lambda/\sigma \end{cases}$$

$$\left.\begin{aligned} 0 &= \frac{\partial\varphi}{\partial x_1} = x_1 - 1 + \begin{cases} \lambda - \sigma(2 - x_1 - x_2) \\ 0 \end{cases} \\ 0 &= \frac{\partial\varphi}{\partial x_2} = x_2 - 2 + \begin{cases} \lambda - \sigma(2 - x_1 - x_2) \\ 0 \end{cases} \end{aligned}\right\} \Rightarrow x_1(\lambda,\sigma) = \begin{cases} \dfrac{1+\sigma-\lambda}{1+2\sigma} \\ 1 \end{cases},\ x_2(\lambda,\sigma) = \begin{cases} \dfrac{2+3\sigma-\lambda}{1+2\sigma} \\ 2 \end{cases}.$$

Αν ισχύει η κάτω περίπτωση, τότε $-1 \geq \lambda/\sigma \Rightarrow \lambda < 0$ αντίκειται στις αναγκαίες συνθήκες. Άρα ισχύει η πάνω περίπτωση.

$$\nabla^2\varphi = \begin{bmatrix} 1+\sigma & \sigma \\ \sigma & 1+\sigma \end{bmatrix} > 0 \ \text{ αν } \sigma > -0.5$$

$$\psi(\lambda) = \varphi(\mathbf{x}(\lambda,\sigma),\lambda,\sigma) = \left(\frac{\sigma+\lambda}{1+2\sigma}\right)^2 - \lambda\frac{2\lambda-1}{1+2\sigma} + \frac{1}{2}\sigma\left(\frac{2\lambda-1}{1+2\sigma}\right)^2 =$$

$$= \frac{\sigma^2 - \lambda^2 + 2\lambda\sigma - 2\lambda^2\sigma + \lambda + \dfrac{\sigma}{2}}{(1+2\sigma)^2}$$

$$\frac{\partial\psi}{\partial\lambda} = \frac{-2\lambda + 2\sigma - 4\lambda\sigma + 1}{(1+2\sigma)^2} = 0 \Rightarrow \lambda^* = 0.5 \text{ (ανεξαρτήτως της τιμής του } \sigma)$$

$$\frac{\partial^2\psi}{\partial\lambda^2} = \frac{-2-4\sigma}{(1+2\sigma)^2} < 0 \text{ αν } \sigma > -0.5$$

$$\Rightarrow x_1(\lambda^*) = x_1^* = 0.5,\ x_2(\lambda^*) = x_2^* = 1.5.$$

d) $\lambda^{(k)} = \lambda^{(k-1)} - \sigma c^{(k-1)} = \lambda^{(k-1)} - \sigma(2 - x_1^{(k-1)} - x_2^{(k-1)}) = \lambda^{(k-1)} - \sigma\dfrac{2\lambda^{(k-1)} - 1}{1+2\sigma} =$

$$= \frac{1}{1+2\sigma}\lambda^{(k-1)} + \frac{\sigma}{1+2\sigma}$$

Συγκλίνει αν $\left|\dfrac{1}{1+2\sigma}\right| < 1$, π.χ. για $\sigma>0$

Στάσιμη τιμή: $\bar{\lambda} = \dfrac{1}{1+2\sigma}\bar{\lambda} + \dfrac{\sigma}{1+2\sigma} \Rightarrow \bar{\lambda} = \lambda^* = 0.5$.

Ταχύτητα σύγκλισης αυξάνει για αυξάνον $\sigma>0$!

9.3 Θεωρούμε την ελαχιστοποίηση της $f(\mathbf{x}) = x_1^4 + (x_2-1)^2 + x_1x_2$ υπό $c(\mathbf{x}) = x_1+x_2 - 0.5 = 0$.

a) Επιλύστε αριθμητικά το πρόβλημα με ΑΤΠ.

b) Επιβεβαιώστε ότι η αριθμητική λύση ικανοποιεί τις ικανές συνθήκες ελαχιστοποίησης.

Λύση

a) $\mathbf{g}(k) = \begin{bmatrix} 4x_1^{(k)3} + x_2^{(k)} \\ 2x_2^{(k)} - 2 + x_1^{(k)} \end{bmatrix}, \quad \mathbf{W}^{(k)} = \begin{bmatrix} 12x_1^{(k)2} & 1 \\ 1 & 2 \end{bmatrix}.$

Άρα έχουμε το ακόλουθο πρόβλημα ΤΠ (9.3.7):

$$q^{(k)}(\boldsymbol{\delta}) = 6\,x_1^{(k)2}\delta_1^2 + \delta_2^2 + \delta_1\delta_2 + (4x_1^{(k)3} + x_2^{(k)})\delta_1 + (2x_2^{(k)} - 2 + x_1^{(k)})\delta_2 \to \underset{\delta}{\mathrm{Min}}$$

υπό $\delta_1+\delta_2+ x_1^{(k)} + x_2^{(k)} - 0.5 = 0$.

Για τη λύση αυτού του προβλήματος ΤΠ έχουμε την Λαγκρανζιανή

$$L(\boldsymbol{\delta},\lambda^{(k+1)}) = 6\,x_1^{(k)2}\delta_1^2 + \delta_2^2 + \delta_1\delta_2 + (4x_1^{(k)3} + x_2^{(k)})\delta_1 + (2x_2^{(k)} - 2 + x_1^{(k)})\delta_2 - \\ - \lambda^{(k+1)}(\delta_1+\delta_2+ x_1^{(k)} + x_2^{(k)} - 0.5)$$

και άρα τις αναγκαίες συνθήκες:

$$\frac{\partial L}{\partial \delta_1} = 12x_1^{(k)2}\,\delta_1+\delta_2+ 4x_1^{(k)3} + x_2^{(k)} - \lambda^{(k+1)} = 0$$

$$\frac{\partial L}{\partial \delta_2} = 2\delta_2 + \delta_1 + 2x_2^{(k)} - 2 + x_1^{(k)} - \text{‘}^{(k+1)} = 0$$

$$-\frac{\partial L}{\partial \lambda^{(k+1)}} = \delta_1+\delta_2+ x_1^{(k)} + x_2^{(k)} - 0.5 = 0$$

εξών η λύση

$$\delta_1^{(k)} = -\frac{1}{8x_1^{(k)2}} - \frac{1}{3}x_1^{(k)}, \quad \delta_2^{(k)} = \frac{1}{8x_1^{(k)2}} - \frac{2}{3}x_1^{(k)} - x_2^{(k)} + 0.5\,,\ \lambda^{(k+1)} = \frac{1}{8x_1^{(k)2}} - \frac{2}{3}x_1^{(k)} - 1$$

$$x_1^{(k+1)} = x_1^{(k)} + \delta_1^{(k)}, \quad x_2^{(k+1)} = x_2^{(k)} + \delta_2^{(k)}\,.$$

Αυτές οι επαναλήψεις δίδουν για τιμές εκκίνησης $x_1^{(l)} = \lambda^{(1)} = 0.1,\ x_2^{(l)} = 0$:

	$0.1 = \lambda^{(1)}$	$0.1 = x_1^{(1)}$	$0.0 = x_2^{(1)}$
k= 1			
k= 2	l[2]=11.433333333	x[2][1]=-12.433333333	x[2][2]=12.933333333
k= 3	l[3]= 7.289697491	x[3][1]=-8.289697491	x[3][2]= 8.789697491
k= 4	l[4]= 4.528283994	x[4][1]=-5.528283994	x[4][2]= 6.028283994
k= 5	l[5]= 2.689612719	x[5][1]=-3.689612719	x[5][2]= 4.189612719
k= 6	l[6]= 1.468924049	x[6][1]=-2.468924049	x[6][2]= 2.968924049
k= 7	l[7]= 0.666456008	x[7][1]=-1.666456008	x[7][2]= 2.166456008
k= 8	l[8]= 0.155982050	x[8][1]=-1.155982050	x[8][2]= 1.655982050
k= 9	l[9]=-0.135803046	x[9][1]=-0.864196954	x[9][2]= 1.364196954
k= 10	l[10]=-0.256496025	x[10][1]=-0.743503975	x[10][2]= 1.243503975
k= 11	l[11]=-0.278208368	x[11][1]=-0.721791632	x[11][2]= 1.221791632
k= 12	l[12]=-0.278874599	x[12][1]=-0.721125401	x[12][2]= 1.221125401
k= 13	l[13]=-0.278875215	x[13][1]=-0.721124785	x[13][2]= 1.221124785
k= 14	l[14]=-0.278875215	x[14][1]=-0.721124785	x[14][2]= 1.221124785

b) Ικανές συνθήκες του περιορισμένου προβλήματος:

$$L(\mathbf{x}, \lambda) = x_1^4 + (x_2 - 1)^2 + x_1x_2 - \lambda(x_1 + x_2 - 0.5)$$

$$\frac{\partial L}{\partial x_1} = 4x_1^{(k)3} + x_2 - \lambda = 0, \quad \frac{\partial L}{\partial x_2} = 2(x_2 - 1) + x_1 - \lambda = 0, \quad -\frac{\partial L}{\partial \lambda} = x_1 + x_2 - 0.5 = 0$$

ικανοποιούνται όλες για λ^*, $\mathbf{x}^*$.

$$\frac{\partial^2 L}{\partial \mathbf{x}^2} = \begin{bmatrix} 12x_1^2 & 1 \\ 1 & 2 \end{bmatrix} > \mathbf{0} \text{ για } x_1 = x_1^*.$$

9.4 Θεωρούμε ελαχιστοποίηση της αντικειμενικής συνάρτησης

$$f(\mathbf{x}) = 0.5x_1 - x_2$$

λαμβάνοντας υπόψη τους περιορισμούς

$$c_1(\mathbf{x}) = -2x_1 - x_2 + 5.5 \geq 0$$
$$c_2(\mathbf{x}) = 3 - (x_1 - 1)^2 - x_2 \geq 0$$
$$c_3(\mathbf{x}) = x_1 \geq 0$$
$$c_4(\mathbf{x}) = x_2 \geq 0.$$

Επιλύστε το πρόβλημα εφαρμόζοντας την μέθοδο επιτρεπτής κατεύθυνσης και ενεργού συνόλου με κατεύθυνση αναζήτησης μέγιστης κατάβασης και με άμεση απαλοιφή μεταβλητών. Το σημείο εκκίνησης είναι $\mathbf{x}^{(1)} = [0.25\ 0.5]^T$. Αναπαραστήστε γραφικά στο χώρο $\mathbf{x}$ τους περιορισμούς και την πορεία των επαναλήψεων.

Λύση

$$\mathbf{x}^{(1)} = [0.25\ 0.5]^T, \quad \mathcal{A}^{(1)} = \{\}$$

Επανάληψη 1:

$\mathcal{A}^{(1)} = \{\}$, άρα x_1, x_2 ελεύθερες μεταβλητές.

$\mathbf{p}^{(1)} = \mathbf{s}^{(1)} = -\mathbf{g} = [-0.5\ 1]^T$

$x_1^{(2)} = 0.25 - \alpha^{(1)} \cdot 0.5, \quad x_2^{(2)} = 0.5 + \alpha^{(1)}, \quad f(\alpha) = -1.25\alpha - 0.375.$

Περιορισμένο ελάχιστο: $\alpha^{(1)} = 0.5$ με ενεργοποίηση του c_3
$\Rightarrow x_1^{(2)} = 0,\ x_2^{(2)} = 1,\ \mathcal{A}^{(2)} = \{3\}.$

Επανάληψη 2:

x_2 ελεύθερη μεταβλητή, δηλαδή $\mathbf{V} = [0\ 1]^T$. Με $A_1 = 1$, $A_2 = 0$
$\Rightarrow \mathbf{Y} = [1\ 0]^T$, $\mathbf{Z} = [0\ 1]^T$.
Μειωθείσα κλίση $g_{x2} = \mathbf{Z}^T\mathbf{g} = -1$
Μειωθείσα κατεύθυνση αναζήτησης $p^{(2)} = s_2^{(2)} = -g_{x2} = 1$
άρα $x_2^{(3)} = 1+\alpha^{(2)}$, $x_1^{(3)} = 0$ (επί του ενεργού περιορισμού), $f(\alpha) = -1 - \alpha$.
Περιορισμένο ελάχιστο: $\alpha^{(2)} = 1$ με ενεργοποίηση του c_2
$\Rightarrow x_1^{(3)} = 0,\ x_2^{(3)} = 2,\ \mathcal{A}^{(3)} = \{2, 3\}.$

Επανάληψη 3: $\boldsymbol{\delta} = \mathbf{0}$ επιλύει το ΠΙ.

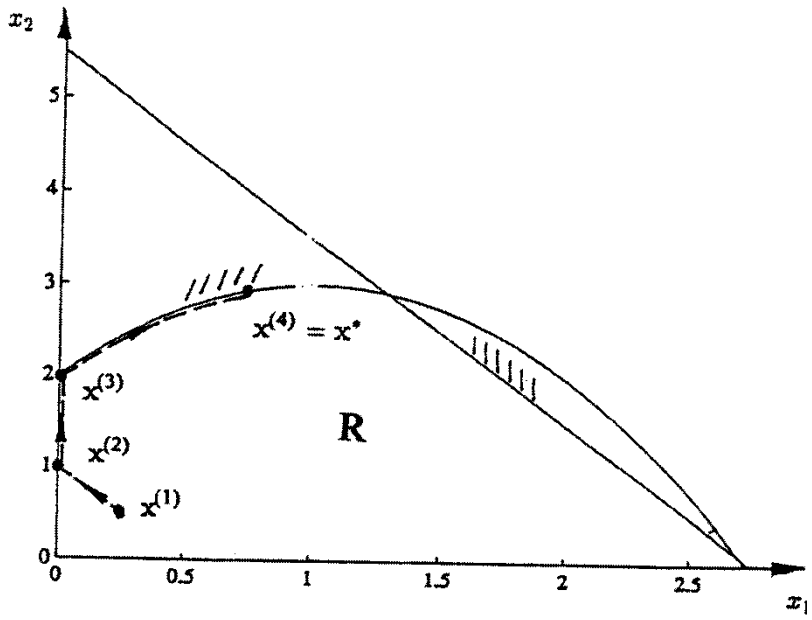
$$\mathbf{g} = \mathbf{A}\boldsymbol{\lambda}^{(3)} \Leftrightarrow \begin{bmatrix} 0.5 \\ -1 \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ -1 & 0 \end{bmatrix} \boldsymbol{\lambda}^{(3)} \Leftrightarrow \begin{matrix} \lambda_2^{(3)} = 1 \\ \lambda_3^{(3)} = -1.5 < 0! \end{matrix}$$

άρα απενεργοποίηση του c_3: $\mathcal{A}^{(3)} = \{2\}$.
x_1 ελεύθερη μεταβλητή, δηλαδή $\mathbf{V} = [1\ 0]^T$ δίδει $\mathbf{Y} = [0\ -1]^T$, $\mathbf{Z} = [1\ 2]^T$.
Μειωθείσα κατεύθυνση: $p^{(3)} = s_1^{(3)} = -g_{x1} = -\mathbf{Z}^T\mathbf{g} = 1.5$
άρα $x_1^{(4)} = 1.5\alpha^{(3)}$, $x_2^{(4)} = 3 - (x_1^{(4)} - 1)^2$ (επί του περιορισμού),
$f(\alpha) = 0.75\alpha - 3 + (1.5\alpha - 1)^2$.
Απεριόριστο ελάχιστο: $\alpha^{(3)} = 0.5 \Rightarrow x_1^{(4)} = 0.75, x_2^{(4)} = 2.9375$

Επανάληψη 4: δ = 0 επιλύει το ΠΙ.

$$\mathbf{g} = \mathbf{A}\lambda^{(4)} \Leftrightarrow \begin{bmatrix} 0.5 \\ -1 \end{bmatrix} = \begin{bmatrix} 0.5 \\ -1 \end{bmatrix} \lambda_2^{(4)} \Rightarrow \lambda_2^{(4)} = 1 > 0$$

άρα $x_1^* = x_1^{(4)} = 0.75, x_2^* = x_2^{(4)} = 2.9375$ είναι η λύση.



10. ΜΕΘΟΔΟΙ ΑΠΟΛΥΤΗΣ ΒΕΛΤΙΣΤΟΠΟΙΗΣΗΣ

Η εφαρμογή των περισσότερων μεθόδων που αναπτύχθηκαν στα προηγούμενα κεφάλαια μπορεί, σε κάποιες κατηγορίες προβλημάτων, να αντιμετωπίσει δυσκολίες που πηγάζουν από:

- παρουσία διακριτών μεταβλητών,
- μη λείες συναρτήσεις,
- πολλά τοπικά ελαχίστα της αντικειμενικής συνάρτησης.

Στις περιπτώσεις αυτές είναι σκόπιμο να χρησιμοποιούνται μέθοδοι **απόλυτης βελτιστοποίησης** που επιχειρούν τον εντοπισμό ενός απόλυτου ελάχιστου ή τουλάχιστον ενός, από πρακτικής άποψης, ικανοποιητικού τοπικού ελάχιστου. Πρέπει βέβαια να σημειωθεί ότι οι μέθοδοι απόλυτης βελτιστοποίησης απαιτούν συνήθως πολύ υψηλότερο υπολογιστικό φόρτο από τις μεθόδους προηγούμενων κεφαλαίων, καθώς και ότι περίπλοκοι περιορισμοί δεν δύνανται συνήθως να ληφθούν υπόψη παρά μόνο μέσω όρων τιμωρίας (βλ. Κεφάλαιο 9.1). Επίσης ο αριθμός μεταβλητών κατά την εφαρμογή των μεθόδων αυτών είναι συνήθως σχετικά περιορισμένος. Από τον μεγάλο αριθμό ευρετικών αλγόριθμων απόλυτης βελτιστοποίησης θα παρουσιάσουμε στο παρόν κεφάλαιο συνοπτικά τρεις κατηγορίες μεθόδων. Θα περιοριστούμε σε στοχαστικούς αλγορίθμους απόλυτης βελτιστοποίησης.

10.1 Η Μέθοδος Complex

Η **μέθοδος Complex** επιχειρεί την ελαχιστοποίηση μιας αντικειμενικής συνάρτησης f(**x**), $\mathbf{x}\in R^n$, υπό τους περιορισμούς ανισότητας $\mathbf{X}_{min} \le \mathbf{X} \le \mathbf{X}_{max}$, $\mathbf{X}\in R^m$, όπου τα **X**, $\mathbf{X}_{min}$, $\mathbf{X}_{max}$ μπορούν να είναι συναρτήσεις του **x**. Πριν από την εκκίνηση του αλγορίθμου προσδιορίζεται καταρχήν μια ομάδα (complex) από $N \ge n+1$ επιτρεπτά σημεία εκκίνησης που μπορούν ενδεχομένως να υπολογισθούν μέσω της ελαχιστοποίησης (9.1.4). Τα σημεία αυτά υπολογίζονται κατά στοχαστικό τρόπο διάσπαρτα εντός της επιτρεπτής περιοχής του προβλήματος. Η μέθοδος επιχειρεί κατόπιν μια σειρά επαναλήψεων, όπου ο στόχος κάθε επανάληψης είναι να αντικατασταθεί το χειρότερο σημείο της τρέχουσας ομάδας (δηλαδή εκείνο που έχει την μεγαλύτερη τιμή της αντικειμενικής συνάρτησης) με ένα νέο, καλύτερο σημείο. Η αναζήτηση του νέου σημείου επιτυγχάνεται μέσω της εκτέλεσης των ακόλουθων βημάτων:

- Καταρχήν επιχειρείται η επέκταση του χώρου που καλύπτεται από τα N σημεία. Προς το σκοπό αυτό διερευνώνται σημεία ευρισκόμενα επί της ευθείας που ενώνει το κέντρο $\mathbf{x}_\kappa^{(k)}$ της ομάδας με το χειρότερο σημείο $\mathbf{x}_\chi^{(k)}$, και δη **εκτός** του ευθύγραμμου τμήματος $(\mathbf{x}_\kappa^{(k)}, \mathbf{x}_\chi^{(k)})$.
- Αν η παραπάνω διερεύνηση παραβιάζει κάποιον περιορισμό, επιλέγεται ένα σημείο επί του αντίστοιχου περιορισμού.
- Αν η παραπάνω διαδικασία δεν οδηγήσει στην ανεύρεση ενός καλύτερου σημείου, τότε διερευνώνται σημεία **εντός** του ευθύγραμμου τμήματος $(\mathbf{x}_\kappa^{(k)}, \mathbf{x}_\chi^{(k)})$.

Οι επαναλήψεις σταματούν αν όλα τα N σημεία έχουν παρόμοιες τιμές (δηλαδή πρακτικά συμπίπτουν), έτσι ώστε να μην είναι εφικτή περαιτέρω βελτίωση.
Η μέθοδος Complex διακρίνεται για την ευρωστία της στην περίπτωση αντικειμενικών συναρτήσεων με πολλά τοπικά ελάχιστα. Φυσικά η μέθοδος δεν εγγυάται την ανεύρεση ενός απόλυτου ελάχιστου, μεγαλώνει όμως την πιθανότητα προσδιορισμού μιας καλής λύσης.

10.2 Προσομοιωμένη Ανόπτηση

Οι μέθοδοι **προσομοιωμένης ανόπτησης** (simulated annealing) βασίζονται σε μια αναλογία με την μεταλλουργική διεργασία της ανόπτησης, κατά την οποία ένα μέταλλο θερμαίνεται καταρχήν μέχρι τήξεως και στη συνέχεια ψύχεται αργά με στόχο την επίτευξη ομοιογενών μοριακών σχηματισμών που ευνοούν τις ιδιότητες σκληρότητας του παρασκευάσματος. Κατά την ψύξη μειώνεται η ελεύθερη ενέργεια του υλικού, και τα συστατικά του σωματίδια δημιουργούν πλέγματα κρυστάλλων. Αν η ψύξη είναι αργή, δημιουργείται στην ιδανική περίπτωση ένας μόνον ομοιογενής κρύσταλλος, και η αντίστοιχη ελεύθερη ενέργεια μειώνεται στο μηδέν (απόλυτο ελάχιστο). Αν η ψύξη γίνει γρήγορα, παρουσιάζονται ανομοιογένειες και ατέλειες στην κρυσταλλοποίηση, που οδηγούν σε ένα τοπικό ελάχιστο της ελεύθερης ενέργειας.

Μια ανάλογη διαδικασία μπορεί να επιχειρηθεί κατά την αριθμητική ελαχιστοποίηση μιας συνάρτησης $f(\mathbf{x})$. Εκκινούμε με ένα σημείο $\mathbf{x}^{(1)}$ και υπολογίζουμε την αντίστοιχη τιμή της αντικειμενικής συνάρτησης $f(\mathbf{x}^{(1)})$. Δίδουμε σε μια παράμετρο θερμοκρασίας T μια αρκούντως υψηλή τιμή $T=T_{max}$. Με δείκτη επαναλήψεων k=1 εκτελούνται εν συνεχεία τα ακόλουθα επαναληπτικά βήματα:
(i) Στοχαστικός προσδιορισμός ενός βήματος $\Delta\mathbf{x}^{(k)}$.
(ii) Υπολογισμός $\Delta f = f(\mathbf{x}^{(k)}+\Delta\mathbf{x}^{(k)})-f(\mathbf{x}^{(k)})$. Αν $\Delta f \le 0$ τότε $\Delta f = 0$.
(iii) Προσδιορισμός του επόμενου επαναληπτικού $\mathbf{x}^{(k+1)}$ με τη βοήθεια μιας γεννήτριας τυχαίων αριθμών, ως εξής:
$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \Delta\mathbf{x}^{(k)}$ με πιθανότητα $P(\Delta f)$
$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)}$ με πιθανότητα $1-P(\Delta f)$
όπου $P(\Delta f) = \exp(-\Delta f/k_B T)$ και k_B είναι μια σταθερή παράμετρος που αντιστοιχεί στην σταθερά Boltzmann του φυσικού φαινομένου. Αντικατάσταση k: = k+1.
(iv) Αποφασίζεται αν θα μειωθεί η θερμοκρασία T. Αν όχι, πήγαινε στο (i).
(v) Μείωση της θερμοκρασίας κατά ΔT. Αν T>0, πήγαινε στο (i), άλλως stop.

Η βασική διαφορά αυτής της διαδικασίας, συγκριτικά με την γενική αλγοριθμική δομή του Κεφαλαίου 2.3, έγκειται στο γεγονός ότι μπορεί στο ως άνω βήμα (iii) να επιλεχθεί με κάποια πιθανότητα ένα νέο επαναληπτικό $\mathbf{x}^{(k+1)}$ που είναι χειρότερο από το προηγούμενο. Η πιθανότητα αυτή είναι τόσο μεγαλύτερη, όσο υψηλότερη η θερμοκρασία T και μικρότερη η χειροτέρευση $\Delta f>0$. (Σημειώνουμε ότι, αν έχουμε βελτίωση της συνάρτησης, τότε τίθεται

$\Delta f = 0$ στο βήμα (ii) και άρα το νέο επαναληπτικό $\mathbf{x}^{(k+1)}$ επιλέγεται με πιθανότητα $P(\Delta f = 0) = 1$.) Επειδή στις πρώτες επαναλήψεις η θερμοκρασία T είναι υψηλή και μειώνεται σταδιακά, η περιγραφείσα διαδικασία εμπεριέχει τη δυνατότητα αποφυγής μιας παγίδευσης του αλγόριθμου σε τοπικά ελάχιστα με υψηλή τιμή της αντικειμενικής συνάρτησης και κατά συνέπεια αυξάνει την πιθανότητα προσδιορισμού ενός καλύτερου τοπικού ελάχιστου ή και του απόλυτου ελάχιστου. Αν ετίθετο εξαρχής T=0 (απότομη ψύξη του τηχθέντος μετάλλου), τότε ο αλγόριθμος θα αποδέχονταν μόνο βήματα $\Delta\mathbf{x}^{(k)}$ που οδηγούν σε βελτίωση της αντικειμενικής συνάρτησης, πράγμα που θα απέκλειε τη δυνατότητα "υπερπήδησης λόφων" γύρω από ένα τοπικό ελάχιστο. Η θερμοκρασία πρέπει λοιπόν να μειώνεται στο βήμα (iv) αρκούντως αργά ώστε να δοθεί στον αλγόριθμο η ευκαιρία να διαφύγει από την περιοχή έλξεως ενός ενδεχομένως κακού τοπικού ελάχιστου. Κατά το τέλος της επαναληπτικής διαδικασίας έχει μειωθεί η θερμοκρασία αρκετά ώστε να προσδιορισθεί πλέον μέσω των βημάτων (i)-(iii) το τοπικό ελάχιστο της αντίστοιχης περιοχής έλξεως που σε πολλές περιπτώσεις μπορεί όντως να είναι το ζητούμενο απόλυτο ελάχιστο.

10.3 Γενετικοί Αλγόριθμοι

Οι **γενετικοί αλγόριθμοι** μιμούνται, κατά τρόπο απλουστευμένο, την βιολογική διεργασία της εξέλιξης με στόχο την απόλυτη μεγιστοποίηση μιας συνάρτησης $f(\mathbf{x})$. Προς το σκοπό αυτό γίνεται κατάλληλη χρήση των βιολογικών εννοιών της μετάλλαξης, της διασταύρωσης και της επιλογής, που μεταφράζονται σε αντίστοιχες μαθηματικές και αλγοριθμικές διεργασίες.

Οι μεταβλητές x_i του προβλήματος μεγιστοποίησης είναι συνήθως δυαδικοί αριθμοί δεδομένου μήκους, των οποίων τα ψηφία (0 ή 1) αντιστοιχούν στα φυσικά χρωμοσώματα. Προβλήματα όπου $\mathbf{x}\in R^n$, μπορούν να μετασχηματισθούν, με πεπερασμένη ακρίβεια, σε αντίστοιχα δυαδικά προβλήματα μετατρέποντας τις πραγματικές μεταβλητές σε αντίστοιχες δυαδικές. Υπάρχουν όμως και παραλλαγές γενετικών αλγορίθμων που εφαρμόζονται απ' ευθείας σε προβλήματα με πραγματικές μεταβλητές $\mathbf{x}\in R^n$, χωρίς να απαιτείται η μετατροπή τους σε δυαδικές.

Η επαναληπτική διεργασία βελτιστοποίησης γενετικών αλγορίθμων βασίζεται σε ένα πληθυσμό $\mathbf{x}_j$, j=1,...,N, σημείων, που, κατ' αναλογία προς την βιολογική διεργασία, ονομάζονται και **άτομα**. Το χρωμόσωμα κάθε ατόμου είναι μια σειρά ψηφίων (0 ή 1) που προκύπτει από την αντίστοιχη σειρά τιμών των μεταβλητών x_i του προβλήματος μεγιστοποίησης.

Το **μέτρο προσαρμοστικότητας** ενός ατόμου $\mathbf{x}_i$ από ορίζεται ως εξής

$$F(\mathbf{x}_i) = f(\mathbf{x}_i)/\sum_{j=1}^{N} f(\mathbf{x}_j)\ . \qquad (10.1)$$

Οι γενετικοί αλγόριθμοι εκκινούν με έναν πληθυσμό εκκίνησης $\mathbf{x}_j^{(1)}$, j=1,...,N, και επιχειρούν, μέσω επαναληπτικής επιλογής και κατάλληλων μεταβολών του πληθυσμού, μια σταδιακή βελτίωση, μέχρι τον προσδιορισμό ενός καλού τοπικού ή και του απόλυτου μέγιστου.

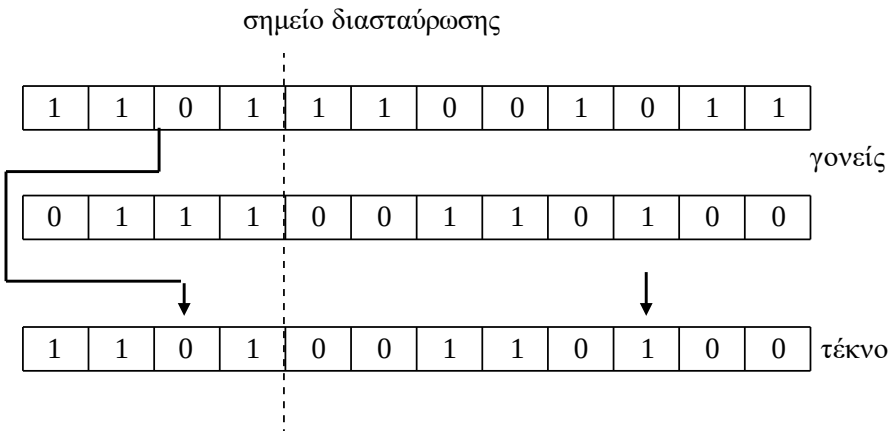
Κατά την διαδικασία **επιλογής** επιλέγονται M (M<N) άτομα από τον παρόντα πληθυσμό N ατόμων με στοχαστικό τρόπο. Ένα άτομο $\mathbf{x}_i$ επιλέγεται ή όχι με μια πιθανότητα ίση προς το μέτρο προσαρμοστικότητάς του, πράγμα που βέβαια δεν αποκλείει και την επιλογή "κακών" ατόμων, τα οποία όμως θα μπορούσαν εν δυνάμει να οδηγήσουν σε καλύτερους απογόνους.

Τα έτσι επιλεχθέντα άτομα υπόκεινται εν συνεχεία σε δύο φάσεις **γενετικών μετατροπών**. Κατά την πρώτη φάση, την **διασταύρωση**, επιλέγονται τυχαία ζεύγη ατόμων, και δημιουργούνται, μέσω μείξης χρωμοσωμάτων, νέα άτομα. Ο πιο συνήθης τρόπος μείξης είναι μέσω του στοχαστικού υπολογισμού ενός **σημείου διασταύρωσης** στην σειρά χρωμοσωμάτων των δύο γονέων, έτσι ώστε το νέο άτομο (τέκνο) να κληρονομεί τα χρωμοσώματα του πρώτου γονέα μέχρι το σημείο διασταύρωσης, και του δεύτερου γονέα μετά το σημείο διασταύρωσης (βλ. Σχ. 10.1). Υπάρχουν διάφορες παραλλαγές γενετικών αλγορίθμων όπου, π.χ., το τέκνο προκύπτει από δύο ή περισσότερα σημεία διασταύρωσης. Μπορούν επίσης να παράγονται περισσότερα του ενός τέκνα (με διαφορετικά σημεία διασταύρωσης) από κάθε ζεύγος γονέων, ή να ζευγαρώνει κάθε γονέας περισσότερες από μια φορά κλπ. Άλλες πάλι παραλλαγές **κατανεμημένων γενετικών αλγορίθμων** εμπεριέχουν την έννοια της γεωγραφικής γειτνίασης και επιτρέπουν μόνο το ζευγάρωμα γειτονικών ατόμων, επιτρέποντας έτσι την παράλληλη επεξεργασία σε πολλαπλούς υπολογιστές. Σε κάθε περίπτωση απαιτείται, στο τέλος της φάσης διασταύρωσης, η παρουσία και πάλι N συνολικά ατόμων.

Η νέα γενεά ατόμων υπόκειται εν συνεχεία στη φάση της **μετάλλαξης**, κατά την οποία μπορούν να αλλαχθούν, με σχετικά μικρή πιθανότητα, οι τιμές μεμονωμένων δυαδικών ψηφίων των χρωμοσωμάτων κάθε τέκνου. Η **μετάλλαξη** προσφέρει τη δυνατότητα διερεύνησης νέων περιοχών τιμών της f(**x**) και έτσι αυξάνει την πιθανότητα προσδιορισμού του απόλυτου μέγιστου.

Οι γενετικές μετατροπές οδηγούν έτσι σε έναν νέο πληθυσμό (αποτελούμενο από γονείς και τέκνα) N ατόμων που αποτελεί την αφετηρία της επόμενης επανάληψης, δηλαδή επιλογής και πάλι M ατόμων, διασταύρωσης, μετάλλαξης κλπ.

Διαφορετικές παραλλαγές γενετικών αλγορίθμων επιτρέπουν διάφορες μείξεις γενεών κατά τον σχηματισμό του νέου πληθυσμού, ενώ οι λεγόμενες **ελιτιστικές** παραλλαγές γενετικών αλγορίθμων διατηρούν οπωσδήποτε στον πληθυσμό τους τα n άτομα (π.χ. n = 1…5) με το καλύτερο μέτρο προσαρμοστικότητας.



Σχήμα 10.1: Παράδειγμα διασταύρωσης.

10.4 Συμπεράσματα

Συμπερασματικά διατυπώνουμε τις ακόλουθες διαπιστώσεις σχετικά με την απόλυτη αριθμητική βελτιστοποίηση:

- Οι περιγραφείσες μέθοδοι απόλυτης βελτιστοποίησης χρησιμοποιούν κάποιες **στοχαστικές** διαδικασίες για να διαφύγουν από την περιοχή έλξης τοπικών ελάχιστων και για να αυξήσουν την πιθανότητα προσδιορισμού ενός απόλυτου ελάχιστου. Ο βέβαιος προσδιορισμός ενός απόλυτου ελάχιστου σε πεπερασμένο χρόνο δεν είναι όμως δυνατός, παρά μόνον αν γίνουν περιοριστικές παραδοχές.

- Οι μέθοδοι στηρίζονται σε απλά πρότυπα, χωρίς ισχυρό θεωρητικό υπόβαθρο.

- Το τίμημα για τον προσδιορισμό ενός καλού ελάχιστου είναι ο υψηλός υπολογιστικός φόρτος, που συνήθως είναι πολλαπλάσιος εκείνου που απαιτούν οι μέθοδοι προηγούμενων κεφαλαίων.

- Η χρήση μεθόδων απόλυτης βελτιστοποίησης είναι αναπόφευκτη κατά την ελαχιστοποίηση συναρτήσεων με υψηλό αριθμό τοπικών ελάχιστων. Είναι δυνατός επίσης ο συνδυασμός απόλυτων μεθοδολογιών με μεθόδους που κάνουν χρήση διανυσμάτων κλίσεως (όπως αυτές των προηγούμενων κεφαλαίων) κατά τρόπο ώστε η απόλυτη μεθοδολογία να εντοπίζει καταρχή την περιοχή έλξεως ενός καλού τοπικού (ή και του απόλυτου) ελάχιστου, ενώ μια μέθοδος κλίσεων να προσδιορίζει εν συνεχεία με αποτελεσματικότητα και ακρίβεια την τιμή του ελαχίστου.

- Οι περισσότερες μέθοδοι απόλυτης βελτιστοποίησης είναι εφαρμόσιμες και σε προβλήματα συνδυαστικής βελτιστοποίησης διαφόρων τύπων.

- Οι περισσότερες μέθοδοι απόλυτης βελτιστοποίησης είναι εφαρμόσιμες μόνο σε προβλήματα με σχετικά μικρό αριθμό μεταβλητών και απλούς περιορισμούς ισότητας ή ανισότητας. Πολύπλοκοι περιορισμοί πρέπει να απαλειφθούν (περιορισμοί ισότητας) ή να ενσωματωθούν ως όροι τιμωρίας στο αντικειμενικό κριτήριο κατά το Κεφάλαιο 9.1.

ΠΡΟΤΕΙΝΟΜΕΝΗ ΒΙΒΛΙΟΓΡΑΦΙΑ

Fletcher, R. (1987) **Practical Methods of Optimization** (2nd Edition). John Wiley & Sons. Chichester.

Gill, P.E., Murray, W., Wright, M.H. (1981). **Practical Optimization**. Academic Press, New York.

Papageorgiou, M., Leibold, M., Buss, M. (2015). **Optimierung - Statische, dynamische, stochastische Verfahren für die Anwendung** (4η Έκδοση). Springer, Berlin.

ΑΣΚΗΣΕΙΣ ΕΞΕΤΑΣΕΩΝ

Πρόβλημα 1

Κατά τις διατάξεις μιάς αεροπορικής εταιρίας επιτρέπεται σε επιβάτες υπερατλαντικών ταξιδίων να αποδίδουν για μεταφορά μια μόνο αποσκευή (σχήματος ορθογώνιου παραλληλεπίπεδου), της οποίας οι διαστάσεις πρέπει να ικανοποιούν τον περιορισμό

$$b+h+l \leq c$$

όπου b το πλάτος, h το ύψος και l το μήκος της αποσκευής, ενώ c>0 είναι μια δεδομένη σταθερά.

1.1 Ζητούνται οι βέλτιστες διαστάσεις b*, l*, h* μιάς αποσκευής που εκπληροί την παραπάνω συνθήκη και έχει μέγιστο όγκο. Προφανώς, η παραπάνω συνθήκη μπορεί να θεωρηθεί κατά την επίλυση του προβλήματος ως περιορισμός ισότητας (γιατί;).

Επιλύστε το προκύπτον πρόβλημα βελτιστοποίησης με τη μέθοδο Lagrange. Ελέγξτε τις ικανές συνθήκες βελτιστοποίησης.

1.2 Επιπλέον της παραπάνω συνθήκης ζητείται τώρα

$$b \leq b_{max},\ l \leq l_{max},\ h \leq h_{max}$$

όπου οι b_{max}, l_{max}, h_{max} είναι δεδομένες σταθερές.

1.2.1 Δώστε τις αναγκαίες συνθήκες πρώτου βαθμού του επεκταθέντος προβλήματος.

1.2.2 Έστω $b_{max} = c/5$, $l_{max} = c/2$, $h_{max} = c/2$. Προσδιορίστε τη λύση του προβλήματος με την παραδοχή $b^* = b_{max}$, $l^* < l_{max}$, $h^* < h_{max}$, έτσι ώστε να εκπληρούνται όλες οι αναγκαίες συνθήκες πρώτου βαθμού.

Πρόβλημα 2

Ζητείται το σημείο $\mathbf{x}^* \in R^2$ της ευθείας

$$x_2 - 3x_1 - 2 = 0$$

που έχει ελάχιστη απόσταση από την αρχή των αξόνων (x_1, x_2).

2.1. Διατυπώστε το πρόβλημα βελτιστοποίησης. Πρόκειται για πρόβλημα κυρτού προγραμματισμού; (αιτιολογήστε).

2.2. Επιλύστε αναλυτικά το πρόβλημα μέσω απαλοιφής. Ελέγξτε τις ικανές συνθήκες βελτιστοποίησης.

2.3. Επιλύστε αναλυτικά το πρόβλημα με τη μέθοδο συνάρτησης τιμωρίας. Προσδιορίστε τις γενικές λύσεις $x_1^*(\sigma)$, $x_2^*(\sigma)$, τις αντίστοιχες ικανές συνθήκες βελτιστοποίησης και το όριο των λύσεων για $\sigma \to \infty$.

2.4. Δώστε τον τύπο μιας επανάληψης Newton για την αριθμητική επίλυση του προβλήματος του 2.3 (να αντιστραφεί ο πίνακας και να γίνουν όλες οι πράξεις). Πόσες επαναλήψεις θα χρειασθεί η μέθοδος Newton;

Πρόβλημα 3

Για την επίλυση του προβλήματος

$$f(\mathbf{x}) = -x_1 e^{-0.5\, x_1^2} + x_1 x_2 + x_2^2 \to \underset{\mathbf{x}}{\text{Min}}$$

υπό $x_1 - x_2 - 1 = 0$

θα εφαρμόσουμε Ακολουθητικό Τετραγωνικό Προγραμματισμό (ΑΤΠ).

3.1 Αναπτύξτε το πρόβλημα Τετραγωνικού Προγραμματισμού που πρέπει να επιλύεται σε κάθε επανάληψη.

3.2 Επιλύστε αναλυτικά το προκύπτον πρόβλημα Τετραγωνικού Προγραμματισμού, και προσδιορίστε τις αναδρομικές εξισώσεις κάθε επανάληψης του αλγόριθμου ΑΤΠ.

3.3 Ποιά συνθήκη πρέπει να ικανοποιεί το επαναληπτικό $\mathbf{x}^{(k)}$, έτσι ώστε ο Χεσιανός πίνακας του προκύπτοντος προβλήματος Τετραγωνικού Προγραμματισμού να είναι θετικά ορισμένος;

Πρόβλημα 4

Ζητείται ελάχιστο της συνάρτησης $f(\alpha) = \frac{1}{3}\alpha^3 - 1.5\alpha^2 + 2\alpha + 5$, $\alpha \geq 0$.

4.1 Προσδιορίστε το ελάχιστο α* μέσω των αναγκαίων και ικανών συνθηκών βελτιστοποίησης. Είναι η συνάρτηση f(α) κυρτή; Ποιά είναι η ελάχιστη τιμή f*;

4.2 Επιλύστε το πρόβλημα ελαχιστοποίησης αριθμητικά στην αγκύλη [1.5, 3] μέσω:

- του τύπου τετραγωνικής παρεμβολής $\hat{a}$ [f'(a), f'(b)]
- του τύπου κυβικής παρεμβολής
- του αλγόριθμου χρυσής τομής.

Εκτελέστε 3 επαναλήψεις για κάθε τύπο. Πόσες επαναλήψεις θα απαιτηθούν από τον αλγόριθμο χρυσής τομής για προσδιορισμό της λύσης με ακρίβεια ε = 0.05; Πόσες επαναλήψεις απαιτεί ο τύπος κυβικής παρεμβολής;

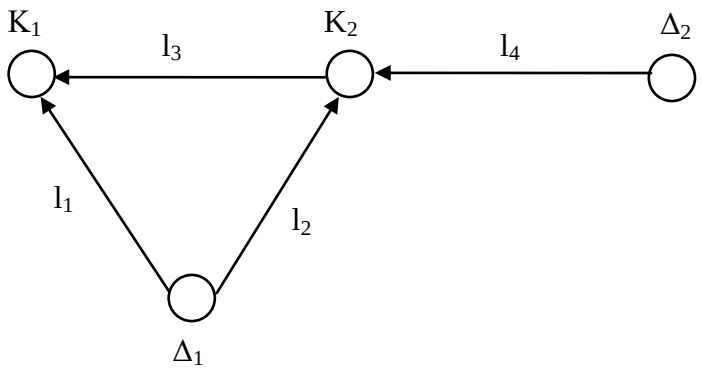
Πρόβλημα 5

Η παραγόμενη βενζίνη δύο διυλιστηρίων Δ_1 και Δ_2 μεταφέρεται μέσω των τεσσάρων αγωγών του σχήματος, με μήκη l_1, l_2, l_3 και l_4, στα σημεία κατανάλωσης K_1, K_2, με σταθερή κατανάλωση k_1, k_2 αντίστοιχα.

Το συνολικό κόστος μεταφοράς δίδεται από

$$J = \sum_{i=1}^{4} l_i x_i^2$$

όπου x_i είναι η ροή του αγωγού i. Ζητούνται οι βέλτιστες ροές $x_i \geq 0$, i=1, 2, 3, 4, που ελαχιστοποιούν το κόστος μεταφοράς καλύπτοντας απολύτως την κατανάλωση χωρίς να υπερβαίνουν τις χωρητικότητες C_i, i=1, 2, 3, 4, των αγωγών, ούτε την μέγιστη παραγωγή D_1, D_2 των διυλιστηρίων.



5.1 Διατυπώστε το πρόβλημα βελτιστοποίησης με το κριτήριο κόστους και όλους τους περιορισμούς. Πρόκειται για πρόβλημα κυρτού προγραμματισμού;

5.2 Απαλείψτε τις μεταβλητές x_3, x_4 συναρτήσει των x_1, x_2 και εκφράστε το μειωθέν πρόβλημα με το κριτήριο κόστους και όλους τους περιορισμούς.

5.3 Δώστε όλες τις αναγκαίες συνθήκες πρώτου βαθμού για το μειωθέν πρόβλημα.

Δίδονται τώρα οι αριθμητικές τιμές $l_1=l_4=0.5$, $l_2=l_3=1$, $k_1=4$, $k_2=3$, $C_1=3$, $C_2=C_3=2$, $C_4=3$, $D_1=5$, $D_2=3$.

5.4 Επιλύστε αριθμητικά το μειωθέν πρόβλημα κάνοντας την υπόθεση ότι δεν ενεργοποιείται κανένας περιορισμός. Αν η λύση παραβιάζει κάποιον περιορισμό, λύστε και πάλι το πρόβλημα θεωρώντας τον παραβιασθέντα περιορισμό ενεργό, κ.ο.κ. μέχρι την τελική λύση του περιορισμένου προβλήματος. Δώστε τις τελικές βέλτιστες τιμές x_1^*, x_2^*, x_3^*, x_4^* και το ελάχιστο κόστος f^*.

Πρόβλημα 6

Θεωρούμε την ελαχιστοποίηση της $f(\mathbf{x}) = x_1 + \frac{1}{2}x_1^2 + \frac{1}{2}(x_2-1)^2$ υπό τον περιορισμό $x_1+x_2=2$.

6.1 Επιλύστε αναλυτικά το πρόβλημα μέσω των αναγκαίων και ικανών συνθηκών ελαχιστοποίησης.

6.2 Επιλύστε αναλυτικά το πρόβλημα μέσω της συνάρτησης τιμωρίας. Υπολογίστε τις λύσεις $x_1^*(\sigma)$, $x_2^*(\sigma)$ και τα αντίστοιχα όρια για $\sigma\to\infty$.

6.3 Θεωρούμε επίλυση του προβλήματος μέσω της συνάρτησης τιμωρίας με πολλαπλασιαστές. Υπολογίστε τα $x_1(\lambda,\sigma)$, $x_2(\lambda,\sigma)$, $\psi(\lambda)$. (Δεν ζητείται προσδιορισμός της τελικής λύσης του προβλήματος.)

Πρόβλημα 7

Θεωρούμε εφαρμογή της μεθόδου συμπλέγματος σημείων σε δισδιάστατα (n=2) προβλήματα ελαχιστοποίησης. Έστω $\mathbf{x}^{(1)}$, $\mathbf{x}^{(2)}$, $\mathbf{x}^{(3)}$ τα τρία σημεία του συμπλέγματος.

7.1 Αναπτύξτε την εξίσωση για τον υπολογισμό ενός νέου σημείου $\mathbf{x}^{(4)}$ που θα αντικαταστήσει το χειρότερο σημείο του συμπλέγματος (π.χ. το $\mathbf{x}^{(1)}$) και θα ευρίσκεται στην αντανάκλαση του χειρότερου σημείου στο κέντρο των υπόλοιπων 2 σημείων.

Θεωρούμε προς ελαχιστοποίηση τη συνάρτηση $f(\mathbf{x}) = x_1^2 + x_2^2 + x_1x_2 - 2x_1 - x_2$.

7.2 Υπολογίστε τον συμμετρικό πίνακα **G** και το διάνυσμα **b** έτσι ώστε

$$f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T\mathbf{G}\mathbf{x} + \mathbf{b}^T\mathbf{x}.$$

7.3 Υπολογίστε το ελάχιστο της f(**x**) μέσω των αναγκαίων και ικανών συνθηκών βελτιστοποίησης.

7.4 Εφαρμόστε τη μέθοδο συμπλέγματος σημείων για αριθμητική ελαχιστοποίηση της f(**x**) εκκινώντας από τα σημεία $\mathbf{x}^{(1)} = [5\ \ -2]^T$, $\mathbf{x}^{(2)} = [5\ \ 0]^T$, $\mathbf{x}^{(3)} = [6.7\ \ -1]^T$ και εκτελώντας 6 επαναλήψεις (χωρίς συστολή του συμπλέγματος). Σχεδιάστε την πορεία των επαναλήψεων σε ένα διάγραμμα (x_1, x_2).

Πρόβλημα 8

Μια βιομηχανία κατασκευής βαρελιών υγρών καυσίμων θέλει να αναπτύξει έναν τύπο βαρελιού κυλινδρικού σχήματος, το οποίο για δεδομένη επιφάνεια C^2 (δηλαδή για δεδομένη ποσότητα υλικού κατασκευής) μεγιστοποιεί τον αποθηκεύσιμο όγκο υγρών.

8.1 Διατυπώστε το πρόβλημα ελαχιστοποίησης με άγνωστες μεταβλητές την ακτίνα βάσης $r \geq 0$ και το ύψος $l \geq 0$ του κυλίνδρου, αναγράφοντας το αντικειμενικό κριτήριο και όλους τους περιορισμούς.

8.2 Επιλύστε το πρόβλημα με τη μέθοδο Lagrange λαμβάνοντας υπόψη μόνο τον περιορισμό ισότητας. Διερευνήστε τις αναγκαίες και τις ικανές συνθήκες ελαχιστοποίησης.

8.3 Για τεχνικούς λόγους πρέπει τώρα να ληφθεί υπόψη ένας επιπλέον περιορισμός: Το σύνολο των γραμμών συγκόλλησης μεταλικών επιφανειών σε κάθε βαρέλι δεν πρέπει να υπερβαίνει το ανώτατο όριο L, δηλαδή

$$4\pi r \leq L.$$

Προσδιορίστε τη βέλτιστη λύση $r^*(C, L)$ και $l^*(C, L)$ σε συνάρτηση των C, L μέσω της μεθόδου Kuhn-Tucker, διερευνώντας τις αναγκαίες συνθήκες βελτιστοποίησης πρώτου βαθμού. (Σημείωση: Οι περιορισμοί $r \geq 0$ και $l \geq 0$ δεν χρειάζεται να ληφθούν υπόψη.)

Πρόβλημα 9

Θεωρούμε το πρόβλημα ελαχιστοποίησης της $f(\mathbf{x}) = 0.5(x_1-1)^2 + 0.5(x_2-3)^2$ υπό $c(\mathbf{x}) = x_1+x_2-1 = 0$.

9.1 Προσδιορίστε την αναλυτική λύση του προβλήματος μέσω απαλοιφής του x_2.

9.2 Επιλύστε αναλυτικά το πρόβλημα (αναγκαίες και ικανές συνθήκες) μέσω της συνάρτησης τιμωρίας. Υπολογίστε τις λύσεις $x_1^*(\sigma)$, $x_2^*(\sigma)$ και τα αντίστοιχα όρια για $\sigma \to \infty$.

9.3 Θεωρούμε επαναληπτική επίλυση του προβλήματος μέσω της συνάρτησης τιμωρίας με πολλαπλασιαστές.

α) Υπολογίστε τα $x_1(\lambda, \sigma)$, $x_2(\lambda, \sigma)$.

β) Χρησιμοποιείστε για το $\lambda^{(k+1)}$ την επαναληπτική εξίσωση $\lambda^{(k+1)} = \lambda^{(k)} - \sigma c^{(k)}$.

γ) Εκτελέστε 3 επαναλήψεις με χρήση των α) και β) για $\sigma = 1$ εκκινώντας από την τιμή $\lambda^{(1)} = 0$.

δ) Προσδιορίστε τις τιμές του σ που οδηγούν σε σύγκλιση του επαναληπτικού αλγόριθμου.

Πρόβλημα 10

Δίδεται η προς ελαχιστοποίηση συνάρτηση

$$f(\mathbf{x}) = x_1^2 + 3x_2^2 + 6x_1 - x_2.$$

10.1 Υπολογίστε αναλυτικά το ζητούμενο ελάχιστο.

10.2 Προσδιορίστε αναλυτικά το βέλτιστο βήμα $\alpha^{(k)}$ κατά μήκος μιάς κατεύθυνσης αναζήτησης $\mathbf{s}^{(k)}$ έτσι ώστε $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \alpha^{(k)}\mathbf{s}^{(k)}$ να είναι το ελάχιστο επί της κατεύθυνσης αναζήτησης.

10.3 Χρησιμοποιώντας το προσδιορισθέν βήμα $\alpha^{(k)}$, εφαρμόστε τη μέθοδο εναλλασσόμενων μεταβλητών για αριθμητική ελαχιστοποίηση της $f(\mathbf{x})$ εκκινώντας από το σημείο $\mathbf{x}^{(1)} = [10 \;\; 4]^T$. Η μέθοδος εναλλασσόμενων μεταβλητών χρησιμοποιεί ως κατευθύνσεις αναζήτησης

$$\mathbf{s}^{(1)} = \begin{bmatrix} -\text{sign}(g_1^{(1)}) \\ 0 \end{bmatrix}, \; \mathbf{s}^{(2)} = \begin{bmatrix} 0 \\ -\text{sign}(g_2^{(2)}) \end{bmatrix}, \; \mathbf{s}^{(3)} = \begin{bmatrix} -\text{sign}(g_1^{(3)}) \\ 0 \end{bmatrix}, \text{ κ.ο.κ.}$$

όπου $\mathbf{g} = \nabla f$ και $\text{sign}(x) = \begin{cases} 1, & \text{αν } x > 0 \\ -1, & \text{αν } x < 0. \end{cases}$

Πόσες επαναλήψεις απαιτούνται για την εύρεση του ελάχιστου; Γιατί ;

Ικανοποιούν οι ως άνω κατευθύνσεις αναζήτησης την ιδιότητα συζυγίας για το παρόν πρόβλημα;

Πρόβλημα 11

Ζητείται η βέλτιστη προσέγγιση της συνάρτησης e^{-x}, για $x \in [0, 1]$, μέσω της γραμμικής συνάρτησης $\alpha_1 + \alpha_2 x$, έτσι ώστε να ελαχιστοποιείται η συνάρτηση σφάλματος

$$f(\boldsymbol{\alpha}) = \int_0^1 (e^{-x} - \alpha_1 - \alpha_2 x)^2 \, dx.$$

11.1 Υπολογίστε την συνάρτηση $f(\boldsymbol{\alpha})$ επιλύοντας αναλυτικά το ως άνω ορισμένο ολοκλήρωμα. (Ισχύει $\int x\, e^{-x}\, dx = -(x + 1)\, e^{-x}$.)

11.2 Υπολογίστε τους συντελεστές α_1, α_2 που ελαχιστοποιούν τη συνάρτηση σφάλματος $f(\boldsymbol{\alpha})$.

11.3 Θεωρείστε επιπλέον ότι η γραμμική συνάρτηση πρέπει να τέμνει την συνάρτηση e^{-x} στο σημείο $x = \frac{1}{2}$. Επιλύστε το επεκταθέν πρόβλημα βελτιστοποίησης με τη μέθοδο Lagrange. Ελέγξτε τις ικανές συνθήκες βελτιστοποίησης.

Πρόβλημα 12

Θεωρούμε την ελαχιστοποίηση της $f(\mathbf{x}) = \frac{1}{2}(x_1 - 3)^2 + x_2^2$ υπό τους περιορισμούς:

$$1 - x_2 + x_1 = 0$$
$$3\, x_1 + x_2 \le 2.$$

12.1 Επιλύστε αναλυτικά το πρόβλημα μέσω των συνθηκών ελαχιστοποίησης.

12.2 Επιλύστε αναλυτικά το πρόβλημα μέσω της συνάρτησης τιμωρίας. Υπολογίστε τις λύσεις $x_1^*(\sigma)$, $x_2^*(\sigma)$ και τα αντίστοιχα όρια για $\sigma \to \infty$.

12.3 Επιλύστε μέσω της μεθόδου Newton το πρόβλημα του 12.2. Πόσες επαναλήψεις θα χρειαστεί η μέθοδος Newton;

Πρόβλημα 13

Σε κυκλικό οικόπεδο διαμέτρου d ζητείται εγγεγραμμένο ορθογώνιο παραλληλόγραμμο πλάτους l_1, και μήκους l_2 μέγιστου εμβαδού.

13.1 Δώστε τη μαθηματική διατύπωση του προβλήματος ελαχιστοποίησης με το αντικειμενικό κριτήριο και όλους τους περιορισμούς.

13.2 Επιλύστε αναλυτικά το πρόβλημα ελαχιστοποίησης με τη μέθοδο Lagrange. (Οι περιορισμοί ανισότητας να μην ληφθούν υπόψη κατά την επίλυση αλλά να ελεγχθούν εκ των υστέρων.)

13.3 Επιλύστε αναλυτικά το πρόβλημα μέσω της συνάρτησης τιμωρίας. Υπολογίστε τις λύσεις $l_1^*(\sigma)$, $l_2^*(\sigma)$ και τα αντίστοιχα όρια για $\sigma \to \infty$.

Πρόβλημα 14

Θεωρούμε την ελαχιστοποίηση της συνάρτησης

$$f(\mathbf{x}) = 2\,x_1^2 + x_2^2 + 3\,x_3^2 - x_1\,x_2 - 4x_2 - x_3$$

υπό τον περιορισμό:

$$-x_1 + 2\,x_2 + 2\,x_3 = 5.$$

14.1 Επιλύστε αναλυτικά το πρόβλημα μέσω των συνθηκών ελαχιστοποίησης.

14.2 Επιλέξτε τον πίνακα **V** και υπολογίστε τους πίνακες **Y** και **Z** που οδηγούν στην άμεση απαλοιφή του x_1. Επιλύστε το προκύπτον απεριόριστο πρόβλημα βελτιστοποίησης και υπολογίστε τον πολλαπλασιαστή Lagrange λ^*.

14.3 Λαμβάνοντας τον ως άνω περιορισμό σαν ανισότητα

$$-x_1 + 2\,x_2 + 2\,x_3 \le 5 \quad (1)$$

και επιπλέον :

$$x_1 \le 0 \quad (2)$$
$$x_2 \ge 0 \quad (3)$$
$$x_3 \ge 0 \quad (4)$$

εφαρμόστε τα βήματα του αλγορίθμου ενεργού συνόλου για την ελαχιστοποίηση της $f(\mathbf{x})$ εκκινώντας από το σημείο $\mathbf{x}^{(1)} = \mathbf{0}$ και θεωρώντας αρχικά ενεργούς τους περιορισμούς (2), (3), και (4).

Πρόβλημα 15

Θεωρούμε την ελαχιστοποίηση της f(**x**) = $3x_1^2 + x_1 x_2 + x_2^2$ υπό τον περιορισμό

$$x_1 - 3x_2^2 = 1.$$

Για την επίλυση του προβλήματος βελτιστοποίησης, εφαρμόστε Ακολουθητικό Τετραγωνικό Προγραμματισμό (ΑΤΠ).

15.1 Αναπτύξτε το πρόβλημα Τετραγωνικού Προγραμματισμού που πρέπει να επιλύεται σε κάθε επανάληψη.

15.2 Επιλύστε αναλυτικά το πρόβλημα και προσδιορίστε τις αναδρομικές εξισώσεις κάθε επανάληψης του αλγορίθμου ΑΤΠ.

15.3 Προσδιορίστε τη συνθήκη που πρέπει να ικανοποιείται σε κάθε επανάληψη, έτσι ώστε να ικανοποιούνται οι ικανές συνθήκες ελαχιστοποίησης για το πρόβλημα Τετραγωνικού Προγραμματισμού.

Πρόβλημα 16

Ζητείται το ελάχιστο της συνάρτησης $f(\alpha) = -\alpha \cdot \exp(0.5\alpha^2 - 4\alpha)$.

16.1 Προσδιορίστε ένα τοπικό ελάχιστο α* μέσω των συνθηκών βελτιστοποίησης. Είναι η συνάρτηση f(α) κυρτή; Ποιά είναι η ελάχιστη τιμή f*;

Επιλύστε το πρόβλημα ελαχιστοποίησης αριθμητικά στην αγκύλη [0, 2].

16.2 Χρησιμοποιείστε τον τύπο της τετραγωνικής παρεμβολής $\hat{a}[f'(a), f'(b)]$ και εκτελέστε 4 επαναλήψεις.

16.3 Χρησιμοποιείστε τον αλγόριθμο χρυσής τομής και προσδιορίστε τη λύση με όριο ακρίβειας ε = 0.15.

Πρόβλημα 17

Σε κυκλικό οικόπεδο ακτίνας r ζητείται εγγεγραμμένο πολύγωνο n πλευρών μέγιστου εμβαδού με δεδομένη τη γωνία α που αντιστοιχεί στην πλευρά n.

17.1 Δώστε τη μαθηματική διατύπωση του προβλήματος ελαχιστοποίησης με το αντικειμενικό κριτήριο και όλους τους περιορισμούς.

17.2 Επιλύστε αναλυτικά το πρόβλημα ελαχιστοποίησης με τη μέθοδο Lagrange. (Οι περιορισμοί ανισότητας να μην ληφθούν υπόψη κατά την επίλυση αλλά να ελεχθούν εκ των υστέρων.)

17.3 Ελέγξτε τις αναγκαίες και ικανές συνθήκες βελτιστοποίησης.

(Ισχύει $\sin 2A = 2\sin A \cdot \cos A$ για δεδομένη γωνία A.)

Πρόβλημα 18

Θεωρούμε την ελαχιστοποίηση της συνάρτησης

$$f(\mathbf{x}) = 2x_1 + x_1 x_2 + 0.5\,(x_1 - 5)^2$$

υπό τον περιορισμό

$$x_1 - 3x_2 = 1.$$

18.1 Επιλύστε αναλυτικά το πρόβλημα μέσω των συνθηκών ελαχιστοποίησης.

18.2 Επιλέξτε τον πίνακα **V** και υπολογίστε τους πίνακες **Y** και **Z** που οδηγούν στην άμεση απαλοιφή του x_1. Επιλύστε το προκύπτον απεριόριστο πρόβλημα βελτιστοποίησης και υπολογίστε τον πολλαπλασιαστή Lagrange λ^*.

18.3 Επιλύστε αναλυτικά το πρόβλημα μέσω της συνάρτησης τιμωρίας. Υπολογίστε τις λύσεις $x_1^*(\sigma)$, $x_2^*(\sigma)$ και τα αντίστοιχα όρια για $\sigma \to \infty$.

Πρόβλημα 19

Δίδεται η προς ελαχιστοποίηση συνάρτηση:

$$f(\mathbf{x}) = x_1^2 + x_2^2 + \beta\, x_1 x_2 + x_1 + 2\, x_2, \quad \beta \in R.$$

19.1 Προσδιορίστε τις τιμές του β για τις οποίες η συνάρτηση f(**x**) έχει αυστηρό τοπικό ελάχιστο.

19.2 Εκκινώντας από το σημείο $\mathbf{x}^{(1)} = [0\ \ 0]^T$ και εφαρμόζοντας την μέθοδο Newton προσδιορίστε το ελάχιστο $x_1^*(\beta), x_2^*(\beta)$. Πόσες επαναλήψεις θα χρειασθεί η μέθοδος Newton και γιατί;

19.3 Θεωρείστε επιπλέον τον περιορισμό

$$-x_1 - 2\beta\, x_2 = 4.$$

Επιλέξτε τον πίνακα **V** και υπολογίστε τους πίνακες **Y** και **Z** που οδηγούν σε άμεση απαλοιφή του x_1. Επιλύστε το προκύπτον απεριόριστο πρόβλημα βελτιστοποίησης και υπολογίστε τον πολλαπλασιαστή Lagrange λ*. Για ποιές τιμές του β έχουμε ελάχιστο;

Πρόβλημα 20

Ζητείται να κατασκευαστεί πισίνα σχήματος ορθογωνίου παραλληλεπιπέδου με πλάτος x_1, μήκος x_2 και ύψος x_3 (σε m) η οποία να έχει δεδομένο όγκο $V = 32\ m^3$ και η επιφάνειά της που θα επενδυθεί με πλακάκια να είναι η ελάχιστη δυνατή.

20.1 Δώστε τη μαθηματική διατύπωση του προβλήματος ελαχιστοποίησης με το αντικειμενικό κριτήριο και όλους τους περιορισμούς.

20.2 Επιλύστε αναλυτικά το πρόβλημα ελαχιστοποίησης με τη μέθοδο Lagrange. (Οι περιορισμοί ανισότητας να μην ληφθούν υπόψη κατά την επίλυση αλλά να ελεγχθούν εκ των υστέρων.)

20.3 Ελέγξτε όλες τις αναγκαίες και ικανές συνθήκες βελτιστοποίησης.

Πρόβλημα 21

Θεωρούμε το πρόβλημα ελαχιστοποίησης της συνάρτησης:

$$f(\mathbf{x}) = \frac{1}{2}(x_1^2 - x_2^2) - 3\,x_2$$

υπό τον περιορισμό

$$-x_1 - 2x_2 = 6.$$

21.1 Επιλύστε αναλυτικά το πρόβλημα μέσω των αναγκαίων και ικανών συνθηκών ελαχιστοποίησης.

21.2 Επιλύστε αναλυτικά το πρόβλημα μέσω της συνάρτησης τιμωρίας. Υπολογίστε τις λύσεις $x_1^*(\sigma)$, $x_2^*(\sigma)$ και τα αντίστοιχα όρια για $\sigma \to \infty$.

21.3 Θεωρούμε επαναληπτική επίλυση του προβλήματος μέσω της συνάρτησης τιμωρίας με πολλαπλασιαστές.

(α) Υπολογίστε τα $x_1(\lambda, \sigma)$ και $x_2(\lambda, \sigma)$

(β) Εκτελέστε 4 επαναλήψεις με χρήση του (α) και της επαναληπτικής εξίσωσης $\lambda^{(k+1)}=\lambda^{(k)}-\sigma c^{(k)}$ για $\sigma = 1$ εκκινώντας από την τιμή $\lambda^{(1)} = 0.5$.

Πρόβλημα 22

Θεωρούμε την ελαχιστοποίηση της συνάρτησης:

$$f(\mathbf{x}) = \frac{1}{2}(1-x_1)^2 + x_1x_2 + 4x_2^2.$$

22.1 Επιλύστε το πρόβλημα αριθμητικά χρησιμοποιώντας την μέθοδο Fletcher-Reeves και εκκινώντας από το σημείο $\mathbf{x}^{(1)} = [0\ 0]^T$. Πόσες επαναλήψεις θα χρειαστούν;

Θεωρούμε τώρα τους περιορισμούς:

$$x_1 + x_2 \le \frac{1}{3} \quad (1)$$

$$x_1 \ge 0 \quad (2)$$

$$x_2 \ge 0 \quad (3)$$

22.2 Επιλύστε αναλυτικά το πρόβλημα μέσω των αναγκαίων και ικανών συνθηκών ελαχιστοποίησης.

22.3 Εφαρμόστε τα βήματα του αλγορίθμου ενεργού συνόλου για την ελαχιστοποίηση της $f(\mathbf{x})$ εκκινώντας από το σημείο $\mathbf{x}^{(1)} = [0\ 0]^T$ και θεωρώντας αρχικά ενεργούς τους περιορισμούς (2) και (3).

Πρόβλημα 23

Ζητείται η ελάχιστη απόσταση του σημείου Ο με συντεταγμένες (x_0, y_0, z_0) από το επίπεδο

$$3x - y + z = 5.$$

23.1 Διατυπώστε το πρόβλημα βελτιστοποίησης. Είναι πρόβλημα κυρτού προγραμματισμού; (Αιτιολογείστε).

23.2 Επιλύστε αναλυτικά το πρόβλημα ελαχιστοποίησης μέσω απαλοιφής του z.

23.3 Επιλύστε αναλυτικά το πρόβλημα με τη μέθοδο συνάρτησης τιμωρίας θεωρώντας ότι το σημείο Ο έχει συντεταγμένες (x_0, y_0, z_0) = (0, 0, 0). Προσδιορίστε τις γενικές λύσεις $x^*(\sigma)$, $y^*(\sigma)$, $z^*(\sigma)$ και τα αντίστοιχα όρια για $\sigma \to \infty$.

Πρόβλημα 24

Θεωρούμε την ελαχιστοποίηση της συνάρτησης:

$$f(\mathbf{x}) = (x_1^2 - 1)^2 + x_2^4 + 5x_1$$

υπό τον περιορισμό

$$x_1 + x_2 = -1.$$

Για την επίλυση του προβλήματος βελτιστοποίησης, εφαρμόστε Ακολουθητικό Τετραγωνικό Προγραμματισμό (ΑΤΠ).

24.1 Αναπτύξτε το πρόβλημα Τετραγωνικού Προγραμματισμού (ΤΠ) που πρέπει να επιλύεται σε κάθε επανάληψη.

24.2 Επιλύστε αναλυτικά το προκύπτον πρόβλημα ΤΠ και προσδιορίστε τις αναδρομικές εξισώσεις κάθε επανάληψης του αλγορίθμου ΑΤΠ.

24.3 Προσδιορίστε τη συνθήκη που πρέπει να ικανοποιείται σε κάθε επανάληψη, έτσι ώστε να ικανοποιούνται οι ικανές συνθήκες ελαχιστοποίησης για το πρόβλημα Τετραγωνικού Προγραμματισμού.

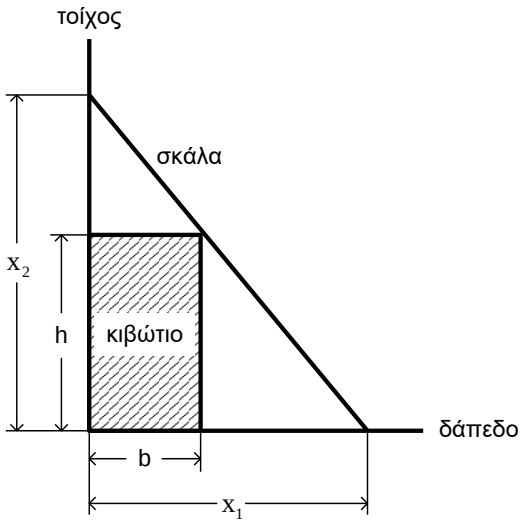
Πρόβλημα 25

Ένα κιβώτιο με δεδομένες διαστάσεις $b \times h$ ακουμπά σε έναν τοίχο όπως φαίνεται στο παρακάτω σχήμα. Μια σκάλα πρέπει να ακουμπήσει, επίσης, στον τοίχο και στο κιβώτιο (βλ. σχήμα). Ζητείται το ελάχιστο δυνατό μήκος της σκάλας.

25.1 Διατυπώστε το αντίστοιχο πρόβλημα βελτιστοποίησης με το αντικειμενικό κριτήριο και όλους τους περιορισμούς βάσει των μεταβλητών x_1, x_2 του σχήματος. Πρόκειται για πρόβλημα κυρτού προγραμματισμού;

25.2 Επιλύστε το πρόβλημα μέσω των συνθηκών βελτιστοποίησης και προσδιορίστε τα $\mathbf{x}^*, \boldsymbol{\lambda}^*, f^*$. (Υπόδειξη: ενδεχόμενοι περιορισμοί ανισότητας μπορούν να μη ληφθούν υπόψη και να ελεγχθούν μετά την αναλυτική επίλυση του προβλήματος.)

25.3 Υπολογίστε τις βέλτιστες τιμές $\mathbf{x}^*, \boldsymbol{\lambda}^*, f^*$ για $b = 1,\ h = 8$.



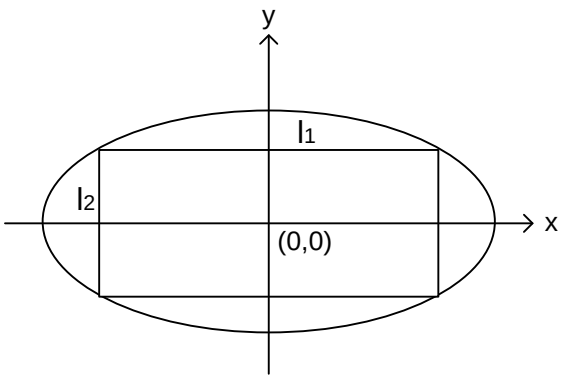
Πρόβλημα 26

Σε ελλειψοειδές οικόπεδο που ορίζεται από την εξίσωση $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$, όπου α, β > 0 γνωστές παράμετροι, ζητείται εγγεγραμμένο ορθογώνιο μήκους l_1 και πλάτους l_2 μέγιστου εμβαδού.

26.1 Δώστε τη μαθηματική διατύπωση του προβλήματος ελαχιστοποίησης με το αντικειμενικό κριτήριο και όλους τους περιορισμούς. (Υπόδειξη για τον καθορισμό του περιορισμού ισότητας: ποιές συντεταγμένες πρέπει να έχει μια τυχαία κορυφή του ορθογώνιου παραλληλόγραμμου;)

26.2 Επιλύστε αναλυτικά το πρόβλημα ελαχιστοποίησης με τη μέθοδο Lagrange. (Οι περιορισμοί ανισότητας μπορούν να μη ληφθούν υπόψη και να ελεγχθούν μετά την αναλυτική επίλυση του προβλήματος.)

26.3 Επιλύστε αναλυτικά το πρόβλημα μέσω της συνάρτησης τιμωρίας. Υπολογίστε τις λύσεις $l_1^*(\sigma)$, $l_2^*(\sigma)$ και τα αντίστοιχα όρια για $\sigma \to \infty$.



Πρόβλημα 27

Ζητούνται τα μήκη των πλευρών ορθογωνίου τριγώνου, δεδομένου εμβαδού Y m^2, που ελαχιστοποιούν την περίμετρο του τριγώνου.

27.1 Δώστε τη μαθηματική διατύπωση του προβλήματος βελτιστοποίησης με το αντικειμενικό κριτήριο και όλους τους περιορισμούς. Πρόκειται για πρόβλημα κυρτού προγραμματισμού;

27.2 Επιλύστε αναλυτικά το πρόβλημα με τη μέθοδο Lagrange και προσδιορίστε τα $x^*(Y)$, $\lambda^*(Y)$, $f^*(Y)$. (Οι περιορισμοί μη αρνητικότητας μπορούν να μη ληφθούν υπόψη και να ελεγχθούν μετά την αναλυτική επίλυση του προβλήματος.) Ελέγξτε για τη λύση που προσδιορίσατε εάν ισχύουν οι ικανές συνθήκες ελαχιστοποίησης.

Πρόβλημα 28

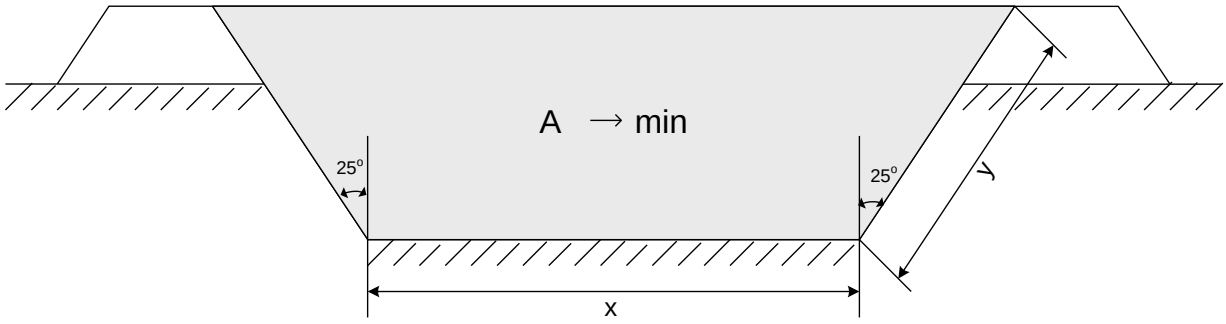
Στο παρακάτω σχήμα φαίνεται η τομή ενός καναλιού που κατασκευάστηκε για τη διέλευση πλοίων. Ζητείται το τραπέζιο με το ελάχιστο δυνατό εμβαδόν τομής του καναλιού, για το οποίο το άθροισμα των πλευρών που βρέχονται είναι ίσο με 500 m.

28.1 Δώστε τη μαθηματική διατύπωση του προβλήματος ελαχιστοποίησης με το αντικειμενικό κριτήριο και όλους τους περιορισμούς. (Υπόδειξη: Οι πλευρές που βρέχονται είναι ο βυθός του καναλιού και οι δύο πλάγιες πλευρές. Θεωρείστε ότι το κανάλι είναι πλήρως γεμάτο.)

28.2 Θεωρούμε επαναληπτική επίλυση του προβλήματος μέσω της συνάρτησης τιμωρίας με πολλλαπλασιαστές. (Υπόδειξη: Ενδεχόμενοι περιορισμοί μη αρνητικότητας μπορούν να μη ληφθούν υπόψη και να ελεγθούν μετά την επίλυση του προβλήματος.)
α) Υπολογίστε τα $x(\lambda,\sigma)$, $y(\lambda,\sigma)$.
β) Χρησιμοποείστε για το $\lambda^{(\kappa+1)}$ την επαναληπτική εξίσωση $\lambda^{(k+1)} = \lambda^{(k)} - \sigma\, c^{(k)}$.

γ) Προσδιορίστε τα x^*, y^* χρησιμοποιώντας τα α) και β) για σ = 1 και εκκινώντας από την τιμή $\lambda^{(1)}$= 5. Εκτελέστε 2 επαναλήψεις.

δ) Προσδιορίστε τις τιμές του σ που οδηγούν σε σύγκλιση του επαναληπτικού αλγορίθμου



Πρόβλημα 29

Μια βιομηχανία σχεδιάζει την παρασκευή δύο ειδών προϊόντων Α και Β έχοντας στη διάθεσή της 100 και 1400 τόνους από τα υλικά Y_1 και Y_2 αντίστοιχα. Για την παραγωγή ενός τόνου προϊόντος Α χρειάζονται 400 κιλά υλικού Y_1 και 600 κιλά υλικού Y_2, ενώ για την παραγωγή ενός τόνου προϊόντος Β χρειάζονται 200 κιλά υλικού Y_1 και 800 κιλά Y_2. Το κόστος παραγωγής για το προϊόν Α είναι 20 €/τόνο ενώ για το προϊόν Β είναι 30 €/τόνο. Οι ζητήσεις x_A του προϊόντος Α και x_B του προϊόντος Β συνδέονται με τις τιμές τους σε ευρώ ανά τόνο p_A και p_B μέσω των σχέσεων $x_A=1000-20\cdot p_A$ και $x_B=400-5\cdot p_B$, αντίστοιχα. Θεωρείστε ότι η ζήτηση ικανοποιείται πλήρως και ότι δεν μένει απόθεμα προϊόντων Α ή Β στην αποθήκη.

Ζητείται να υπολογιστούν οι ποσότητες των προϊόντων Α και Β που πρέπει να παραχθούν ώστε να μεγιστοποιηθεί το κέρδος της εταιρείας.

29.1 Αναπτύξτε ένα πλήρες μαθηματικό μοντέλο με μεταβλητές x_A, x_B για την επίλυση του παραπάνω προβλήματος. (Οι μεταβλητές p_A και p_B να απαλειφθούν με χρήση αντιστοίχων περιορισμών ισότητας.) Πρόκειται για πρόβλημα κυρτού προγραμματισμού;

29.2 Επιλύστε αναλυτικά το πρόβλημα μεγιστοποίησης μέσω των αναγκαίων συνθηκών Kuhn-Tucker. (Οι περιορισμοί μη αρνητικότητας των παραγόμενων ποσοτήτων κάθε προϊόντος μπορούν να μη ληφθούν υπόψη και να ελεχθούν μετά την αναλυτική επίλυση του προβλήματος.)

Πρόβλημα 30

Θεωρούμε προς μεγιστοποίηση τη συνάρτηση

$$f(x)=x\cdot e^{-x},\ x\geq 0.$$

30.1 Υπολογίστε το σημείο μεγιστοποίησης x* μέσω των αναγκαίων και ικανών συνθηκών βελτιστοποίησης. Είναι η συνάρτηση f(x) κοίλη; Ποια η μέγιστη τιμή f*;

30.2 Επιλύστε το πρόβλημα αριθμητικά στην αγκύλη [0.5, 1.5] εκτελώντας 2 επαναλήψεις μέσω του τύπου τετραγωνικής παρεμβολής $\hat{x}[f'(a), f'(b)]$.

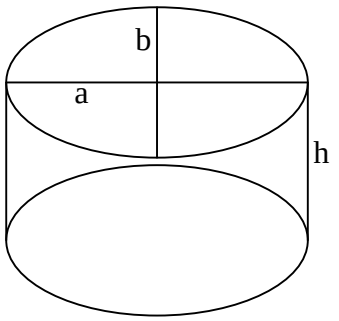
30.3 Επαναλάβατε την αριθμητική επίλυση του προβλήματος στην ίδια αρχική αγκύλη και εκτελώντας τον ίδιο αριθμό επαναλήψεων μέσω του αλγορίθμου χρυσής τομής, τροποποιημένου κατάλληλα για προβλήματα μεγιστοποίησης.

Πρόβλημα 31

Θεωρούμε μια πισίνα ελλειψοειδούς τομής με διαστάσεις a, b και ύψος h, η οποία πρέπει να έχει δεδομένο όγκο V=64π. Ζητούνται οι διαστάσεις a, b, h έτσι ώστε η εσωτερική επιφάνεια της πισίνας που πρόκειται να καλυφθεί με πλακάκια, να είναι η ελάχιστη δυνατή.

31.1 Δώστε την μαθηματική διατύπωση του προβλήματος ελαχιστοποίησης με το αντικειμενικό κριτήριο και όλους τους περιορισμούς.
(Δίδονται: εμβαδόν έλλειψης E=πab και περίμετρος έλλειψης κατά προσσέγγιση $\Pi=\pi\sqrt{2(a^2+b^2)}$.)

31.2 Επιλύστε αναλυτικά το πρόβλημα ελαχιστοποίησης με τη μέθοδο Lagrange, εξετάζοντας μόνο τις αναγκαίες συνθήκες πρώτου βαθμού. (Σημείωση: Να μη ληφθούν υπόψη οι περιορισμοί ανισότητας, αλλά να ελεγχθούν εκ των υστέρων.)

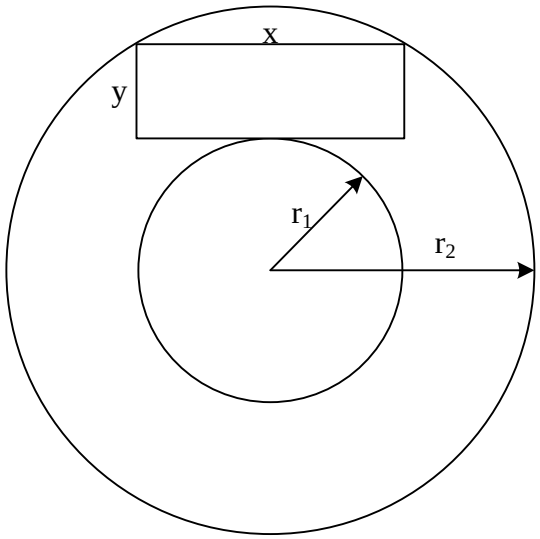


Πρόβλημα 32

Μεταξύ δύο δεδομένων ομόκεντρων κύκλων ζητείται εγγεγραμμένο ορθογώνιο (κατά το παρακάτω σχήμα) παραλληλόγραμμο, μήκους x και πλάτους y, μέγιστου εμβαδού. Οι ακτίνες των δύο κύκλων είναι r_1=1 m και r_2=2 m.

32.1 Διατυπώστε το πρόβλημα βελτιστοποίησης με το κριτήριο κόστους και όλους τους περιορισμούς. Πρόκειται για πρόβλημα κυρτού προγραμματισμού;

32.2 Επιλύστε το πρόβλημα μέσω απαλοιφής αγνοώντας ενδεχόμενους περιορισμούς μη αρνητικότητας. Επιβεβαιώστε ότι η προσδιορισθείσα λύση ικανοποιεί τις αναγκαίες και ικανές συνθήκες του αρχικού προβλήματος βελτιστοποίησης.



Πρόβλημα 33

Σε χώρο δύο διαστάσεων (x_1, x_2) δίνονται 2 σημεία $\mathbf{y_1} = \begin{bmatrix} 0.5 \\ 2.5 \end{bmatrix}$ και $\mathbf{y_2} = \begin{bmatrix} 1.5 \\ 1.5 \end{bmatrix}$ και η ευθεία $x_1 = 2x_2 - 2$. Ζητείται το σημείο $\mathbf{X} = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$ της ευθείας που ελαχιστοποιεί το αντικειμενικό κριτήριο $f(\mathbf{X}) = \frac{1}{2}\sum_{i=1}^{2} \|\mathbf{X} - \mathbf{y_i}\|^2$. (*Σημείωση:* για διάνυσμα $\boldsymbol{\alpha} = [\alpha_1\ \alpha_2 \ldots \alpha_n]^T$ έχουμε $\|\boldsymbol{\alpha}\| = \sqrt{\sum_{i=1}^{n} a_i^2}$.)

33.1 Διατυπώστε το πρόβλημα βελτιστοποίησης.

33.2 Επιλύστε αναλυτικά το πρόβλημα με τη μέθοδο Lagrange.

33.3 Επιλύστε αναλυτικά το πρόβλημα μέσω της μεθόδου συνάρτησης τιμωρίας.

33.4 Επιλέξτε τον πίνακα **V** και υπολογίστε τους πίνακες **Y** και **Z** που οδηγούν στην άμεση απαλοιφή του X_1. Επιλύστε το προκύπτον απεριόριστο πρόβλημα βελτιστοποίησης και υπολογίστε τον πολλαπλασιαστή Lagrange λ^*.

Πρόβλημα 34

Θεωρούμε δύο εξωτερικά εφαπτόμενους κύκλους με ακτίνες r_1 και r_2 αντίστοιχα. Τα κέντρα των κύκλων απέχουν 10 cm μεταξύ τους. Ζητούνται οι ακτίνες r_1, r_2 έτσι ώστε να έχουμε το ελάχιστο συνολικό εμβαδόν για τους δύο κύκλους.

34.1 Δώστε την πλήρη μαθηματική διατύπωση του προβλήματος ελαχιστοποίησης.

34.2 Επιλύστε το πρόβλημα μέσω των αναγκαίων και ικανών συνθηκών βελτιστοποίησης. (Σημείωση: Να μη ληφθούν υπόψη οι περιορισμοί μη αρνητικότητας, αλλά να ελεγχθούν εκ των υστέρων.)

34.3 Χωρίς να λάβετε υπόψη τους περιορισμούς μη αρνητικότητας, επιλέξτε τον πίνακα **V** και υπολογίστε τους πίνακες **Y** και **Z** που οδηγούν σε άμεση απαλοιφή του r_1. Επιλύστε το προκύπτον απεριόριστο πρόβλημα βελτιστοποίησης.