# 1 Μέθοδοι ελαχιστοποίησης

Θεωρούμε το $n \times n$ πραγματικό σύστημα

$$Ax = b, \tag{1.1}$$

με $A$ ένα συμμετρικό και θετικά ορισμένο πίνακα και $b \in \mathbb{R}^n$. Ορίζουμε το συναρτησιακό $\phi : \mathbb{R}^n \to \mathbb{R}$

$$\phi(x) = \frac{1}{2}(Ax, x) - (b, x), \tag{1.2}$$

όπου $(\cdot, \cdot)$ είναι το ευκλείδειο εσωτερικό γινόμενο στον $\mathbb{R}^n$. Αν $z = A^{-1}b$ είναι η λύση του (1.1) τότε για κάθε $y \in \mathbb{R}^n$ έχουμε $\phi(y + z) = \phi(z) + \frac{1}{2}(Ay, y)$. Συμπεραίνουμε ότι $\phi(z + y) > \phi(z)$ για κάθε μη μηδενικό διάνυσμα $y \in \mathbb{R}^n$, δηλαδή το $\phi(x)$ παίρνει την ελάχιστη τιμή του στον $\mathbb{R}^n$ στο σημείο $z = A^{-1}b$. Συνεπώς το πρόβλημα της επίλυσης του (1.1) είναι ισοδύναμο με το πρόβλημα ελαχιστοποίησης χωρίς περιορισμούς

$$\min_{x \in \mathbb{R}^n} \phi(x). \tag{1.3}$$

Επειδή $\nabla\phi(x) = Ax - b$, το $z = A^{-1}b$ είναι το μοναδικό κρίσιμο σημείο στο οποίο το $\phi(x)$ έχει ελάχιστο (είναι συστηρά κυρτό).

Αν ξέρουμε ότι $\nabla\phi(x) \neq 0$ τότε η συνάρτηση $g(\alpha) = \phi(x + au)$, $\alpha \in \mathbb{R}, x, u \in \mathbb{R}^n$, που περιγράφει τη συμπεριφορά του $\phi$ κοντά στο σημείο $x$ κατά την κατεύθυνση του $u$, ελαττώνεται με μέγιστο ρυθμό ελάττωσης όταν $u = -\nabla\phi(x)$. Πράγματι, $g(0) = \phi(x)$ και $g'(0) = (\nabla\phi(x), u)$. Η παρατήρηση αυτή μας οδηγεί σε μια επαναληπτική μέθοδο για τον υπολογισμό των ελαχίστων του $\phi$, τη λεγόμενη *μέθοδο της καθόδου μέγιστης κλίσης* (steepest descent method). Η μέθοδος παράγει μια ακολουθία $\{x^j\}, j \geq 0$ προσεγγίσεων ενός (τοπικού) ελαχίστου $x^*$ του $\phi$ ως εξής: έστω $x^k$ μια προσέγγιση του $x^*$. Η $x^{k+1}$ ορίζεται ως το πλησιέστερο στο $x^k$ ελάχιστο του $\phi(x)$ όταν το $x$ περιορίζεται πάνω στην ακτίνα που διέρχεται από το $x^k$ και έχει τη δεύθυνση του $-\nabla\phi(x^k)$. Βρίσκουμε δηλαδή τον μικρότερο θετικό αριθμό $\alpha^*$ όπου η συνάρτηση $g(\alpha) = \phi(x^k - \alpha\nabla\phi(x^k))$ έχει ελάχιστο και ορίζουμε $x^{k+1} = x^k - \alpha\nabla\phi(x^k)$.

Στην περίπτωση του συναρτησιακού (1.2), στο σημείο $x^k$ το $\phi(x)$ ελαττώνεται ταχύτερα στην κατεύθυνση $-\nabla\phi(x^k) = b - Ax^k$, την οποία ταυτίζουμε με το υπόλοιπο $r^k$ του $x^k$. Αν $r^k \neq 0$ (διαφορετικά $x^k = A^{-1}b$) υπολογίζουμε το $x^{k+1}$ ελαχιστοποιώντας το πολυώνυμο (ως πρός $\alpha$)

$$\phi(x^k + \alpha r^k) = \phi(x^k) - (r^k, r^k)\alpha + \frac{1}{2}(Ar^k, r^k)\alpha^2,$$

του οποίου το ελάχιστο λαμβάνεται για $\alpha = (r^k, r^k)/(Ar^k, r^k)$. Οδηγούμαστε λοιπόν στον εξής αλγόριθμο της μεθόδου της καθόδου μέγιστης κλίσης για την ελαχιστοποίηση του συναρτησιακού (1.2):

$x^0 = 0$
**for** $k = 1, 2, \ldots$ **do**
  $r^{k-1} = b - Ax^{k-1}$
  **if** $r^{k-1} = 0$ **then**
    $x = x^{k-1}$, τέλος
  **else**
    $\alpha_k = (r^{k-1}, r^{k-1})/(Ar^{k-1}, r^{k-1})$
    $x^k = x^{k-1} + \alpha_k r^{k-1}$
  **end**
**end**

Ο αλγόριθμος τερματίζεται όταν ικανοποιηθούν ένα ή περισσότερα από τα συνήθη κριτήρια τερματισμού επαναληπτικών μεθόδων, για παράδειγμα, αν $\|r^k\| \le \delta\|r^0\|$, για κάποια σταθερά $0 < \delta < 1$. Εδώ, $\|\cdot\| = (\cdot,\cdot)^{1/2}$ είναι η νόρμα που παράγεται από το ευκλείδειο εσωτερικό γινόμενο.

Θα μελετήσουμε τη σύγκλιση της μεθόδου στη φυσική νόρμα του προβλήματος δηλαδή τη νόρμα που παράγεται από το εσωτερικό γινόμενο $(Ax, y)$ στον $\mathbb{R}^n$. Θα χρησιμοποιήσουμε το γεγονός ότι ο $A$, σαν συμμετρικός και θετικά ορισμένος πίνακας, έχει $n$ θετικές ιδιοτιμές $\lambda_{\min} = \lambda_1 \le \lambda_2 \le \cdots \le \lambda_n = \lambda_{\max}$ και $n$ αντίστοιχα, ορθοκανονικά ως προς το $(\cdot,\cdot)$, ιδιοδιανύσματα $v^j$, $j = 1, \dots, n$.

**Θεώρημα 1.1.** *Έστω $\{x^j\}_{j\ge 0}$ η ακολουθία που παράγει η μέθοδος της καθόδου μέγιστης κλίσης για οποιοδήποτε $x^0 \in \mathbb{R}^n$ και έστω $x$ η λύση του* (1.1). *Τότε $x^j \to x$, $j \to \infty$. Μάλιστα, αν $e^j = x - x^j$ είναι το σφάλμα της προσέγγισης $x^j$ και $\kappa = \lambda_{\max}/\lambda_{\min}$ ισχύει ότι*

$$(Ae^j, e^j)^{1/2} \le \left(\frac{\kappa - 1}{\kappa + 1}\right)^j (Ae^0, e^0)^{1/2}, \quad j = 0, 1, \dots. \tag{1.4}$$

*Απόδειξη.* Έχουμε $Ae^j = r^j$ και $e^{j+1} = x - x^{j+1} = e^j - \alpha_{j+1} r^j$. Χρησιμοποιώντας τον ορισμό του $\alpha_{j+1}$ εύκολα βλέπουμε ότι $(Ae^{j+1}, r^j) = 0$. Οι σχέσεις αυτές μας επιτρέπουν να γράψουμε, για οποιοδήποτε $a \in \mathbb{R}$,

$$\begin{aligned}(Ae^{j+1}, e^{j+1}) &= (Ae^{j+1}, e^j - \alpha_{j+1} r^j) = (Ae^{j+1}, e^j) - a(Ae^{j+1}, r^j) \\ &= (Ae^{j+1}, e^j - a r^j).\end{aligned}$$

Από την ανισότητα Cauchy–Schwarz έχουμε τώρα

$$(Ae^{j+1}, e^{j+1}) \le (Ae^{j+1}, e^{j+1})^{1/2} (A(e^j - ar^j), e^j - ar^j)^{1/2}, \quad \forall a \in \mathbb{R}.$$

Χρησιμοποιώντας τώρα την σχέση $Ae^j = r^j$ έχουμε $e^j - ar^j = (I - aA)e^j$ οπότε $A(e^j - ar^j) = A(I - aA)e^j = (I - aA)Ae^j$. Συμπεραίνουμε λοιπόν ότι

$$(Ae^{j+1}, e^{j+1}) \le \inf_{a\in\mathbb{R}} ((I - aA)Ae^j, (I - aA)e^j), \quad j \ge 0. \tag{1.5}$$

Αν $u \in \mathbb{R}^n$ μπορούμε να γράψουμε $u = \sum_{j=1}^n (u, v^j) v^j$ και $Au = \sum_{j=1}^n \lambda_j (u, v^j) v^j$. Γενικότερα, για κάθε πραγματικό πολυώνυμο $p$ έχουμε $p(A)u = \sum_{j=1}^n p(\lambda_j)(u, v^j) v^j$, οπότε

$$\begin{aligned}((I - aA)Ae^j, (I - aA)e^j) &= \sum_{i=1}^n (1 - a\lambda_i)^2 \lambda_i (e^j, v^i)^2 \\ &\le \max_{1\le i\le n} (1 - a\lambda_i)^2 \sum_{i=1}^n \lambda_i (e^j, v^i)^2 \\ &= \max_{1\le i\le n} (1 - a\lambda_i)^2 (Ae^j, e^j).\end{aligned}$$

Η σχέση αυτή και η (1.5) δίνουν

$$(Ae^{j+1}, e^{j+1})^{1/2} \le \inf_{a\in\mathbb{R}} \left( \max_{\lambda\in[\lambda_{\min}, \lambda_{\max}]} |1 - a\lambda| \right) (Ae^j, e^j)^{1/2}, \quad j \ge 0. \tag{1.6}$$

Προφανώς,

$$\sigma = \inf_{a\in\mathbb{R}} \left( \max_{\lambda\in[\lambda_{\min}, \lambda_{\max}]} |1 - a\lambda| \right) = \inf_{a\in\mathbb{R}} \left( \max(|1 - a\lambda_{\min}|, |1 - a\lambda_{\max}|) \right),$$

και η συνάρτηση $a \mapsto \max(|1 - a\lambda_{\min}|, |1 - a\lambda_{\max}|)$ παίρνει την ελάχιστη τιμή της όταν $1 - a\lambda_{\min} > 0$, $1 - a\lambda_{\max} < 0$ και $|1 - a\lambda_{\min}| = |1 - a\lambda_{\max}|$, δηλαδή όταν $a = 2/(\lambda_{\min} + \lambda_{\max})$. Τότε $\sigma = (\kappa - 1)/(\kappa + 1)$ και η εκτίμηση της ταχύτητας σύγκλισης της ακολουθίας $\{e^j\}$ προκύπτει από τη σχέση (1.6). $\square$

**Παρατήρηση 1.2.** *Είναι φανερό από τη σχέση* (1.4) *ότι η ταχύτητα σύγκλισης της ακολουθίας* $\{e^j\}$ *μπορεί να είναι πολύ μικρή αν ο λόγος* $\kappa = \lambda_{\max}/\lambda_{\min}$ *είναι μεγάλος, με άλλα λόγια αν ο δείκτης κατάστασης του πίνακα* $A$ *στη νόρμα* $\|\cdot\| = (\cdot,\cdot)^{1/2}$ *είναι μεγάλος.*

Ενδιαφέρον παρουσιάζει η γεωμετρική ερμηνεία της μεθόδου της καθόδου μέγιστης κλίσης. Οι ισοϋψείς επιφάνιες $\phi(x) = \text{σταθ.}$ του συναρτησιακού $\phi$ είναι ελλειψοειδή στον $\mathbb{R}^n$ με κέντρο το σημείο $A^{-1}b$ και άξονες παράλληλους προς τα ιδιοδιανύσματα $v^i$ με μήκη αξόνων ανάλογα των αριθμών $\lambda_i^{-1/2}$. Μπορεί να δειχθεί ακόμα ότι τα διανύσματα $x^{k+1} - x^k$ και $x^k - x^{k-1}$ είναι ορθογώνια. Αν λοιπόν $\lambda_{\max} \gg \lambda_{\min}$ τα ελλειψοειδή είναι υπερβολικά στενόμακρα και η κάθοδος προς το κέντρο μιας βαθιάς χαράδρας με απότομες πλαγιές γίνεται με ζίγκ-ζάγκ στις διαδοχικές ορθογώνιες κατευθύνσεις $r^k = -\nabla\phi(x^k)$ μεταξύ των πλαγιών πράγμα που καθυστερεί πολύ την κάθοδο. Έτσι, η τακτική της ελαχιστοποίησης κατά μήκος των διευθύνσεων μέγιστης κλίσης αποδεικνύεται κακή στρατηγική.

Επιχειρούμε τώρα να ελαχιστοποιήσουμε το συναρτησιακό $\phi$ σε κατευθύνσεις $p^1, p^2, \ldots$, που δεν συμπίπτουν αναγκαστικά με τα υπόλοιπα $r^0, r^1, \ldots$, ελπίζοντας να βελτιώσουμε την ταχύτητα σύγκλισης. Είναι εύκολο να δούμε ότι, για δεδομένα $x^{k-1}$ και μη μηδενικό διάνυσμα $p^k$, η συνάρτηση $\phi(x^{k-1} + \alpha p^k)$ ελαχιστοποιείται όταν

$$a = a_k = (p^k, r^{k-1})/(Ap^k, p^k).$$

Αν λοιπόν θέσουμε $x^k = x^{k-1} + \alpha_k p^k$ θα ισχύει $\phi(x^k) < \phi(x^{k-1})$, εφ' όσον $(p^k, r^{k-1}) \neq 0$. Όπως και πρίν,

$$\phi(x^k) = \phi(x^{k-1}) - \frac{(p^k, r^{k-1})}{2(Ap^k, p^k)}.$$

Γενικεύουμε λοιπόν τον αλγόριθμο της μεθόδου της καθόδου μέγιστης κλίσης ως εξής:

$x^0 = 0$
**for** $k = 1, 2, \ldots$ **do**
  $r^{k-1} = b - Ax^{k-1}$
  **if** $r^{k-1} = 0$ **then**
    $x = x^{k-1}$, τέλος
  **else**
    διάλεξε $p^k$ τέτοιο ώστε $(p^k, r^{k-1}) \neq 0$
    $\alpha_k = (p^k, r^{k-1})/(Ap^k, p^k)$
    $x^k = x^{k-1} + \alpha_k p^k$
  **end**
**end**

Παρατηρούμε ότι, λόγω της αναδρομικής σχέσης $x^k = x^{k-1} + \alpha_k p^k$, μπορούμε να γράψουμε $x^k = a_1 p^1 + \cdots + a_k p^k$, $1 \leq i \leq k$. Θα ήταν ιδανικό, για παράδειγμα, αν διαλέγαμε τα $p^i$ έτσι ώστε να είναι γραμμικά ανεξάρτητα και έτσι ώστε το $x^k$ να λύνει το πρόβλημα ελαχιστοποίησης

$$\min_{x \in \langle p^1, \ldots, p^k \rangle} \phi(x), \tag{1.7}$$

όπου $\langle p^1, \ldots, p^k \rangle$ είναι το σύνολο των διανυσμάτων που είναι γραμμικοί συνδιασμοί των $p^1, \ldots, p^k$. Είναι εύκολο να δειχθεί ότι το πρόβλημα ελαχιστοποίησης (1.7) έχει μοναδική λύση. Είναι φανερό ότι με μια τέτοια κατασκευή ο γενικευμένος αλγόριθμος της μεθόδου της καθόδου μέγιστης κλίσης θα συνέκλινε σε ακριβώς $n$ βήματα στη λύση του συστήματος (1.1). Πράγματι, το $x^n$ θα ελαχιστοποιούσε το $\phi(x)$ για $x \in \langle p^1, \ldots, p^n \rangle = \mathbb{R}^n$, δηλαδή $x^n = A^{-1}b$. Από την άλλη μεριά, όμως, είναι επίσης φανερό ότι ο συγκεκριμένος αλγόριθμος παράγει ένα διάνυσμα $x^k$ που λύνει το *μονοδιάστατο* πρόβλημα ελαχιστοποίησης

$$\min_{\alpha \in \mathbb{R}} \phi(x^{k-1} + \alpha p^k). \tag{1.8}$$

Τίθεται, λοιπόν, το ερώτημα κατά πόσο είναι δυνατόν να διαλέξουμε τις κατευθύνσεις $p^k$ έτσι ώστε το $x^k$, η λύση του (1.8), να είναι επίσης λύση του (1.7). Η ανάλυση που ακολουθεί επιχειρεί να διακιολογήσει τη θετική απάντηση σε αυτό το ερώτημα.

Αν $P_j = \left[p^1, \ldots, p^j\right]$ είναι ο $n \times j$ πίνακας με στήλες τα διανύσματα $p^1, \ldots, p^j$, τότε ένα διάνυσμα $z \in \mathbb{R}^n$ ανήκει στο πεδίο τιμών $\mathcal{R}(P_j)$ του πίνακα $P_j$ αν και μόνο αν $z \in \langle p^1, \ldots, p^j \rangle$. Ειδικώτερα, αν $x \in \mathcal{R}(P_k)$ τότε το $x$ μπορεί να γραφεί στη μορφή $x = P_{k-1}y + ap^k$ για κάποιο $y \in \mathbb{R}^{n-1}$ και $a \in \mathbb{R}$. Τότε έχουμε

$$\phi(P_{k-1}y + ap^k) = \phi(P_{k-1}y) + \frac{a^2}{2}(Ap^k, p^k) - a(p^k, b) + a(P_{k-1}y, Ap^k). \tag{1.9}$$

Για την ελαχιστοποίηση λοιπόν του $\phi(x)$ για $x \in \mathcal{R}(P_k)$ θα μηδενίσουμε, κατ' αρχήν, τον τελευταίο όρο στη σχέση (1.9): με δεδομένα $p^i$, $1 \leq i \leq k-1$ υπολογίζουμε το $p^k$ έτσι ώστε

$$P_{k-1}^T A p^k = 0. \tag{1.10}$$

Μπορούμε ακόμα να μηδενίσουμε τον δεύτερο και τρίτο όρο στο δεξί μέλος της (1.9) επιλέγοντας

$$a = a_k = \frac{(p^k, b)}{(Ap^k, p^k)}. \tag{1.11}$$

Για τον πρώτο όρο στο δεξί μέλος της (1.9), δηλαδή την ελαχιστοποίηση του $\phi(P_{k-1}y)$ για $y \in \mathbb{R}^{k-1}$, αρκεί να ελαχιστοποιήσουμε το $\phi$ πάνω στον υπόχωρο $\langle p^1, \ldots, p^{k-1} \rangle$, πρόβλημα που η λύση του $x^{k-1}$ έχει ήδη βρεθεί. Αν λοιπόν συμβαίνει αυτό και ισχύουν οι σχέσεις (1.10) και (1.11) τότε το πρόβλημα (1.7) λύνεται για $x = x^k = x^{k-1} + a_k p^k$. Επιπλέον, επειδή $x^{k-1} \in \langle p^1, \ldots, p^{k-1} \rangle$, έχουμε $x^{k-1} = P_{k-1}z$ για κάποιο $z \in \mathbb{R}^{k-1}$. Από την (1.10) προκύπτει τότε ότι

$$(p^k, Ax^{k-1}) = (Ap^k, P_{k-1}z) = z^T P_{k-1}^T A p^k = 0,$$

δηλαδή ότι το $a_k$ που δίνεται από τη σχέση (1.11) ικανοποιεί και τη σχέση

$$a_k = \frac{(p^k, b - Ax^{k-1})}{(Ap^k, p^k)} = \frac{(p^k, r^{k-1})}{(Ap^k, p^k)}. \tag{1.12}$$

Αυτό σημαίνει ότι το $a_k$ λύνει το μονοδιάστατο πρόβλημα ελαχιστοποίησης (1.8). Συνοψίζουμε τα συμπεράσματα αυτής της ανάλυσης: αν το $p^k$ επιλεγεί έτσι ώστε να ισχύει η (1.10), δηλαδή να ισχύουν οι σχέσεις

$$(Ap^i, p^k) = 0, \quad i = 1, \ldots, k-1, \tag{1.13}$$

τότε ο γενικευμένος αλγόριθμος της μεθόδου της καθόδου μέγιστης κλίσης κατασκευάζει $a_k$ και $x^k$ που λύνουν τα προβλήματα (1.7) και (1.8). Αν η (1.13) ισχύει, λέμε ότι το $p^k$ είναι $A$-συζυγές ή $A$-ορθογώνιο προς τα $p^1, \ldots, p^{k-1}$.

Απομένει να απαντήσουμε στα εξής ερωτήματα:

1. Είναι τα $p^i$ γραμμικά ανεξάρτητα;
2. Ισχύει ότι $(p^k, r^{k-1}) \neq 0$ έτσι ώστε να έχουμε ελάττωση του $\phi$ στο βήμα $k$ του γενικευμένου αλγορίθμου της μεθόδου της καθόδου μέγιστης κλίσης;
3. Πως υπολογίζουμε στην πράξη τα $p^i$;

Η απάντηση στο πρώτο ερώτημα είναι θετική. Αν τα διανύσματα $p^i$, $1 \leq i \leq m$ του $\mathbb{R}^n$ είναι $A$-συζυγή, δηλαδή αν

$$(Ap^i, p^j) = 0, \quad i \neq j, \tag{1.14}$$

και $\sum_{i=1}^m c_i p^i = 0$ για κάποιες σταθερές $c_1, \ldots, c_m$, παίρνοντας το εσωτερικό γινόμενο με κάποιο $p^j$ και χρησιμοποιώντας την (1.14) έχουμε διαδοχικά

$$0 = \left( \sum_{i=1}^m c_i p^i, p^j \right) = \left( \sum_{i=1}^m c_i A p^i, p^j \right) = \sum_{i=1}^m c_i (Ap^i, p^j) = c_j (Ap^j, p^j),$$

σχέση από την οποία προκύπτει ότι $c_j = 0$. Όπως τονίστηκε παραπάνω, αν κατασκευάσουμε τα $p^i$ έτσι ώστε να ισχύει η (1.13) τότε θα είναι γραμμικά ανεξάρτητα και επομένως ο γενικευμένος αλγόριθμος της μεθόδου της καθόδου μέγιστης κλίσης δίνει την ακριβή λύση του προβλήματος σε $n$ βήματα.

Για το δεύτερο ερώτημα, αν κάποιο $r^{k-1}$ είναι μηδέν τότε ο αλγόριθμος τερματίζει και έχουμε $x^{k-1} = x = A^{-1}b$. Διαφορετικά, δείχνουμε ότι υπάρχει $p^k \in \mathbb{R}^n$ που ικανοποιεί την (1.13) και τέτοιο ώστε $(p^k, r^{k-1}) \neq 0$. Πράγματι, αν για κάθε $p \in \mathbb{R}^n$ που είναι $A$-συζυγές προς τα $p^i$, $1 \leq i \leq k-1$, ισχύει $(p, r^{k-1}) = 0$, ή, ισοδύναμα, $(p, b - Ax^{k-1}) = 0$, τότε $(p, b) = 0$ (θυμηθείτε ότι $x^{k-1} \in \langle p^1, \ldots, p^{k-1} \rangle$). Από τη σχέση αυτή συμπεραίνουμε ότι $p \in \langle Ap^1, \ldots, Ap^{k-1} \rangle^\perp$, ισοδύναμα, $b \in \langle Ap^1, \ldots, Ap^{k-1} \rangle$. Επομένως, $A^{-1}b \in \langle p^1, \ldots, p^{k-1} \rangle$, δηλαδή $x^{k-1} = A^{-1}b$ και $r^{k-1} = 0$, το οποίο είναι άτοπο. Συνεπώς, μπορούμε να βρούμε $p^k$ $A$-συζυγές προς τα $p^i$, $1 \leq i \leq k-1$, τέτιοιο ώστε $(p^k, r^{k-1}) \neq 0$, αν βέβαια $r^{k-1} \neq 0$.

Όπως παρατηρήσαμε παραπάνω, $p$ είναι $A$-συζυγές προς τα $p^i$, $1 \leq i \leq k-1$ αν και μόνο αν $p \in \langle Ap^1, \ldots, Ap^{k-1} \rangle^\perp$. Η λεγόμενη *μέθοδος των συζυγλων κλίσεων* των Hestens και Stiefel (1952) επιλέγει ως $p^k$ το πλησιέστερο προς το $r^{k-1}$ διάνυσμα του υπόχωρου $\langle Ap^1, \ldots, Ap^{k-1} \rangle^\perp$. Μπορούμε λοιπόν να εκφράσουμε τον αλγόριθμο της μεθόδου των συζυγών κλίσεων ως εξής:

$x^0 = 0$
**for** $k = 1, 2, \ldots$ **do**
  $r^{k-1} = b - Ax^{k-1}$
  **if** $r^{k-1} = 0$ **then**
    $x = x^{k-1}$, τέλος
  **else**

$$p^k = \begin{cases} r^0, & \text{αν } k = 1 \\ \text{ορθή προβολή του } r^{k-1} \text{ στον } \langle Ap^1, \ldots, Ap^{k-1} \rangle^\perp, & \text{αν } k > 1 \end{cases}$$

    $\alpha_k = (p^k, r^{k-1})/(Ap^k, p^k)$
    $x^k = x^{k-1} + \alpha_k p^k$
  **end**
**end**

Στην επόμενη ενότητα απαντάμε το τρίτο ερώτημα που θέσαμε παραπάνω, δηλαδή πως υπολογίζουμε αποτελεσματικά τα $p^k$ και, επιπλέον, εκτιμούμε το σφάλμα $e^k = x - x^k$ της μεθόδου των συζυγών κλίσεων.

## Ασκήσεις

**Άσκηση 1.1.** *Βρείτε μια αναδρομική σχέση μεταξύ των υπολοίπων $r^k, r^{k-1}$ που να μας επιτρέπει να γράψουμε τον αλγόριθμο της μεθόδου της καθόδου μέγιστης κλόσης έτσι ώστε να απαιτείται ένας μόνο πολλαπλασιασμός πίνακα επί διάνυσμα για κάθε $k$. Πόσες, συνολικά, πράξεις απαιτεί ο νέος αλγόριθμος;*

**Άσκηση 1.2.** *Δείξτε ότι δύο διαδοχικά υπόλοιπα της μεθόδου της καθόδου της μέγιστης κλίσης είναι ορθογώνια, δηλαδή $(r^j, r^{j-1}) = 0$, $j \geq 1$. Δείξτε επίσης ότι τα διανύσματα $x^{j+1} - x^j$ και $x^j - x^{j-1}$, $j \geq 1$ είναι ορθογώνια.*

**Άσκηση 1.3.** *Δείξτε ότι η παράσταση* $(Ax, y)$, $x, y \in \mathbb{R}^n$, *ορίζει ένα εσωτερικό γινόμενο στον* $\mathbb{R}^n$ *και, ως εκ τούτου, η συνάρτηση* $x \mapsto (Ax, x)^{1/2}$ *είναι νόρμα. Βρείτε τις καλύτερες σταθερές σύγκρισης μεταξύ αυτής της νόρμας και της ευκλείδειας νόρμας.*

**Άσκηση 1.4.** *Χρησιμοποιώντας τον συμβολισμό του Θεωρήματος 1.1 δείξτε ότι για* $j \geq 0$ *έχουμε*

$$(Ae^{j+1}, e^{j+1}) = \left[1 - \frac{(r^j, r^j)^2}{(Ar^j, r^j)(A^{-1}r^j, r^j)}\right](Ae^j, e^j).$$

*Δείξτε επίσης την ανισότητα του Kantorovich*

$$(Ax, x)(A^{-1}x, x) \leq \left[\frac{(\lambda_{\max} + \lambda_{\min})^2}{4\lambda_{\max}\lambda_{\min}}\right]\|x\|^4,$$

*και δώστε έτσι μια δεύτερη απόδειξη του Θεωρήματος 1.1.*

**Άσκηση 1.5.** *Θεωρούμε το* $2 \times 2$ *σύστημα* $Ax = b$ *με*

$$A = \begin{pmatrix} 1 & 0 \\ 0 & \lambda \end{pmatrix}$$

*και* $b = x = 0$. *Αν* $x^k = (x_1, x_2)$ *δείξτε ότι η μέθοδος της καθόδου μέγιστης κλίσης δίνει*

$$x^{k+1} = \frac{x_1 x_2(\lambda - 1)}{x_1^2 + \lambda^3 x_2^2}\begin{pmatrix} \lambda^2 x_2 \\ -x_1 \end{pmatrix}$$

*Σχεδιάστε για* $\lambda = 5$ *τις ισοϋψείς καμπύλες του συναρτησιακού* $\phi(x)$ *και μερικές διαδοχικές προσεγγίσεις* $x^k$, $k = 1, 2, \ldots$, *με* $x^0 = (5, 1)^T$. *Αν* $\lambda = 100$, *πόσα βήματα της μεθόδου θα χρειαστούν έτσι ώστε το σφάλμα* $(Ae^j, e^j)^{1/2}$ *να γίνει μικρότερο από* $\epsilon(Ae^0, e^0)^{1/2}$ *όπου* $\epsilon > 0$ *δεδομένο;*

**Άσκηση 1.6.** *Έστω* $M$ *ένας υπόχωρος του* $\mathbb{R}^n$ *και* $y \in \mathbb{R}^n$ *ένα δεδομένο διάνυσμα. Δείξτε ότι το πρόβλημα ελαχιστοποίησης του* $\phi(x)$ *για* $x \in y + M$ *έχει μοναδική λύση.*

## 2 Η μέθοδος των συζυγών κλίσεων

Ξεκινάμε την ανάλυση της μεθόδου των συζυγών κλίσεων αποδεικνύοντας μερικές χρήσιμες ιδιότητες των υπολοίπων $r^i$ και κατευθύνσεων $p^i$. Κατ' αρχήν, θα λέμε ότι ο αλγόριθμος της μεθόδου των συζυγών κλίσεων *ολοκληρώνει $k$ βήματα* αν $r^i \neq 0$, $0 \leq i \leq k-1$. Σε αυτή την περίπτωση, τα $x^i$, $1 \leq i \leq k$ έχουν υπολογιστεί επαγωγικά από τις σχέσεις: $x^0 = 0$ και

$$x^i = x^{i-1} + a_i p^i, \quad i \geq 1. \tag{2.1}$$

**Λήμμα 2.1.** *Αν ο αλγόριθμος της μεθόδου των συζυγών κλίσεων ολοκληρώνει $k$ βήματα, τότε για $i = 1, 2, \ldots, k$ έχουμε*

$$r^i = r^{i-1} - a_i A p^i, \tag{2.2}$$

$$P_i^T r^i = 0, \tag{2.3}$$

*όπου $P_i$ είναι ο $n \times i$ πίνακας $[p^1, \ldots, p^i]$ με στήλες τα $p^i$.*

*Απόδειξη.* Η σχέση (2.2) προκύπτει εύκολα από την (2.1). Για την (2.3) τώρα, γνωρίζουμε ότι το πρόβλημα ελαχιστοποίησης

$$\min_{x \in \langle p^1, \ldots, p^i \rangle} \phi(x), \tag{2.4}$$

λύνεται μοναδικά για $x = x^i$. Εξ' άλλου, επειδή $x \in \langle p^1, \ldots, p^i \rangle$, δηλαδή, $x \in \mathcal{R}(P_i)$, το (2.4) είναι ισοδύναμο με το πρόβλημα

$$\min_{y \in \mathbb{R}^i} \phi(P_i y), \tag{2.5}$$

του οποίου η λύση $y^i$ ικανοποιεί $(P_i^T A P_i) y^i = P_i^T b$. Από τη σχέση $x^i = P_i^y$ έπεται αμέσως η $P_i^T A x^i = P_i^T b$, δηλαδή $P_i^T r^i = 0$. □

**Λήμμα 2.2.** *Για $k \geq 2$, τα διανύσματα $p^k$ που παράγονται από τον αλγόριθμο της μεθόδου των συζυγών κλίσεων (εφ' όσον ολοκληρώνει $k$ βήματα) δίνονται από τις σχέσεις*

$$p^k = r^{k-1} - A P_{k-1} z^{k-1}, \tag{2.6}$$

*όπου $z^{k-1} \in \mathbb{R}^{k-1}$ είναι η μοναδική λύση του προβλήματος ελαχιστοποίησης*

$$\min_{z \in \mathbb{R}^{k-1}} \|r^{k-1} - A P_{k-1} z\|. \tag{2.7}$$

*Απόδειξη.* Εξ' ορισμού, το $p^k$ είναι η ορθή προβολή του $r^{k-1}$ στον υπόχωρο $\mathcal{R}(AP_{k-1})^{\perp}$. Συνεπώς, το $r^{k-1} - p^k$ είναι η ορθή προβολή του $r^{k-1}$ στον υπόχωρο $\mathcal{R}(AP_{k-1})$, δηλαδή $r^{k-1} - p^k = AP_{k-1} z^{k-1}$ για $z^{k-1} \in \mathbb{R}^{k-1}$. Η μοναδικότητα του $z^{k-1}$ έπεται από την γραμμική ανεξαρτησία των $p^i$. Το $z^{k-1}$ λύνει το πρόβλημα ελαχιστοποίησης

$$\begin{aligned}
\|r^{k-1} - (r^{k-1} - p^k)\| = \|p^k\| &= \|r^{k-1} A P_{k-1} z^{k-1}\| \\
&= \min_{u \in \mathcal{R}(AP_{k-1})} \|r^{k-1} - u\| = \min_{z \in \mathbb{R}^{k-1}} \|r^{k-1} - A P_{k-1} z\|.
\end{aligned}$$

□

Τα δύο αυτά λήμματα θα τα χρησιμοποιήσουμε για να απόδειξουμε το εξής σημαντικό, θεωρητικό αποτέλεσμα.

**Πρόταση 2.3.** *Αν ο αλγόριθμος της μεθόδου των συζυγών κλίσεων ολοκληρώνει $k$ βήματα, τότε τα υπόλοιπα $r^0, \dots, r^{k-1}$ είναι ορθογώνια μεταξύ τους και για $j = 1, \dots, k$, ισχύει*

$$\langle p^1, \dots, p^j \rangle = \langle r^0, \dots, r^{j-1} \rangle = \langle b, Ab, \dots, A^{j-1}b \rangle. \tag{2.8}$$

*Απόδειξη.* Από την (2.2) για $1 \le i \le k-1$ έχουμε ότι $Ap^i = (r^i - r^{i-1})/a_i$. Άρα $Ap^i \in \langle r^0, \dots, r^i \rangle$. Τότε η (2.6) δίνει ότι $p^j = r^{j-1} - [Ap^1, \dots, Ap^{j-1}]z^{j-1} \in \langle r^0, \dots, r^i \rangle$, για $1 \le j \le k$. Έπεται λοιπόν ότι $p^j = c_{j,0}r^0 + \cdots + c_{j,j-1}r^{j-1}$, όπου $c_{j,j-1} \ne 0$ γιατί τα $p^j$ είναι γραμμικά ανεξάρτητα (ως $A$-συζυγή). Εισάγοντας τον $n \times k$ πίνακα $R_k = [r^0, \dots, r^{k-1}]$ έχουμε $P_k = R_k C$, όπου $C$ ένας $k \times k$ αντιστρέψιμος, άνω τριγωνικός πίνακας. Συνεπώς, $R_k = P_k C^{-1}$ και επειδή ο $C^{-1}$ είναι άνω τριγωνικός συμπεραίνουμε ότι για $1 \le j \le k$, $r^{j-1} \in \langle p^1, \dots, p^j \rangle$. Από την (2.3) όμως έχουμε ότι $(p^j, r^i) = 0$, $1 \le j \le i \le k-1$. Από τις δύο αυτές σχέσεις έχουμε ότι $(r^j, r^i) = 0$ για $0 \le j \le i \le k-1$, το οποίο αποδεικνύει τον πρώτο ισχυρισμό της πρότασης.

Η (2.8) αποδεικνύεται με επαγωγή ως προς $j$. Επειδή $p^1 = r^0 = b$, ισχύει για $j = 1$. Έστω ότι ισχύει για κάποιο $j$, $1 \le j \le k-1$. Από την (2.2) έχουμε ότι $r^j = r^{j-1} - a_j Ap^j \in \langle b, Ab, \dots, A^j b \rangle$, ενώ η (2.6) δίνει $p^{j+1} = r^j - A(z_1^j p^j + \cdots + z_j^j p^j) \in \langle b, Ab, \dots, A^j b \rangle$. Συνεπώς, οι δύο υπόχωροι $\langle r^0, \dots, r^j \rangle$ και $\langle p^1, \dots, p^{j+1} \rangle$ (και οι δύο διαστάσεως $j+1$) περιέχονται στον υπόχωρο $\langle b, Ab, \dots, A^j b \rangle$ ο οποίος έχει επίσης διάσταση $j+1$. Άρα η (2.8) ισχύει για $j+1$. □

**Παρατήρηση 2.4.** *Μπορούμε εύκολα να γενικεύσουμε τη μέθοδο των συζυγών κλίσεων για αρχική τιμή $x^0 \ne 0$. Ορισμένα όμως προφανή πράγματα αλλάζουν στις αποδείξεις, π.χ., το $x^j$ λύνει τώρα το πρόβλημα ελαχιστοποίησης*

$$\min_{y \in x^0 + \langle p^0, \dots p^j \rangle} \phi(x). \tag{2.9}$$

*Οι σχέσεις (2.1)–(2.3) εξακολουθούν να ισχύουν όπως επίσης και το πρώτο συμπέρασμα της Πρότασης 2.3. Ακόμα, η (2.8) θα πρέπει να αντικατασταθεί από την*

$$\langle p^1, \dots, p^j \rangle = \langle r^0, \dots, r^{j-1} \rangle = \langle r^0, Ar^0, \dots, A^{j-1}r^0 \rangle, \quad 1 \le j \le k. \tag{2.10}$$

Είμαστε τώρα σε θέση να εκτιμήσουμε το σφάλμα $x - x^k$ σε κάθε βήμα την μεθόδου.

**Θεώρημα 2.5.** *Έστω $\{x^j\}$, $j \ge 0$ η ακολουθία που παράγει η μέθοδος των συζυγών κλίσεων για οποιαδήποτε αρχική τιμή $x^0 \in \mathbb{R}^n$, $x$ η λύση του (1.1) και $e^j = x - x^j$. Έστω ακόμα $\kappa = \lambda_{\max}/\lambda_{\min}$. Για $j \ge 1$ έχουμε τότε ότι*

$$(Ae^j, e^j)^{1/2} \le 2 \left[ \frac{\sqrt{\kappa} - 1}{\sqrt{\kappa} + 1} \right]^j (Ae^0, e^0)^{1/2}. \tag{2.11}$$

*Απόδειξη.* Κατ' αρχήν σημειώνουμε ότι για οποιοδήποτε $y \in \mathbb{R}^n$ έχουμε $(A(x-y), x-y) = 2\phi(y) + (Ax, x)$. Συνεπώς το πρόβλημα ελαχιστοποίησης $\min_{y \in S} \phi(y)$ για $S \subset \mathbb{R}^n$ είναι ισοδύναμο με το πρόβλημα ελαχιστοποίησης $\min_{y \in S}(A(x-y), x-y)$. Συμπεραίνουμε, επειδή $x^j$ είναι η μοναδική λύση του (2.9), ότι

$$\begin{aligned} (Ae^j, e^j) = (A(x - x^j), x - x^j) &= \min_{y \in x^0 + \langle p^1, \dots, p^j \rangle} (A(x-y), x-y) \\ &= \min_{z \in \langle p^1, \dots, p^j \rangle} (A(e^0 + z), e^0 + z). \end{aligned} \tag{2.12}$$

Λόγω της (2.10) έχουμε ότι $z \in \langle p^1, \dots, p^j \rangle$ αν και μόνο αν $z \in \langle r^0, Ar^0, \dots, A^{j-1}r^0 \rangle$, δηλαδή αν $z = \pi_{j-1}(A)r^0$ για κάποιο πολυώνυμο $\pi_{j-1} \in \mathbb{P}_{j-1}$. Η (2.12) και το γεγονός ότι $r^0 = Ae^0$ δίνουν ότι

$$(Ae^j, e^j) = \min_{\pi \in \mathbb{P}_{j-1}} ((I + A\pi(A))^2 Ae^0, e^0). \tag{2.13}$$

Εδώ υποθέτουμε ότι $j \geq 1$ και ότι ο αλγόριθμος της μεθόδου των συζυγών κλίσεων ολοκληρώνει $j$ βήματα. Η σχέση (2.13) ισχύει και για $j = 0$ αν ορίσουμε $\mathbb{P}_{-1} = \{0\}$. Με πράξεις ανάλογες με εκείνες της απόδειξης του Θεωρήματος 1.1 παίρνουμε

$$(Ae^j, e^j) \leq \min_{\pi \in \mathbb{P}_{j-1}} \max_{\lambda \in [\lambda_{\min}, \lambda_{\max}]} (1 + \lambda\pi(\lambda))^2 \, (Ae^0, e^0).$$

Έχουμε τελικά

$$(Ae^j, e^j)^{1/2} \leq \sigma_j (Ae^0, e^0)^{1/2}, \quad j \geq 0, \tag{2.14}$$

όπου για $j \geq 0$ έχουμε θέσει $\sigma_0 = 1$ και

$$\sigma_j = \min_{\pi \in \mathbb{P}_{j-1}, \pi(0)=1} \max_{\lambda \in [\lambda_{\min}, \lambda_{\max}]} |\pi(\lambda)|. \tag{2.15}$$

Αποδεικνύουμε στο Λήμμα 2.7 παρακάτω ότι

$$\sigma_j \leq 2 \left[ \frac{\sqrt{\kappa} - 1}{\sqrt{\kappa} + 1} \right]^j, \quad j \geq 1. \tag{2.16}$$

Από τις σχέσεις (2.14) και (2.16) έπεται αμέσως η (2.11) και αυτό ολοκληρώνει την απόδειξη του θεωρήματος. □

Η λύση του προβλήματος "min–max" (2.14) γίνεται με χρήση πολυωνύμων Chebyshev. Υπενθυμίζουμε ότι τα πολυώνυμα Chebyshev $T_j(z)$ ορίζονται από τις αναδρομικές σχέσεις

$$\begin{aligned} T_0(z) &= 1, \\ T_1(z) &= z, \\ T_j(z) &= 2zT_{j-1}(z) - T_{j-2}(z), \quad j \geq 2. \end{aligned} \tag{2.17}$$

Από τον ορισμό τους είναι εύκολο να αποδείξει κανείς τις ακόλουθες ιδιότητες των πολυωνύμων Chebyshev $T_m(z)$, $m \geq 0$:

1. $\deg(T_m(z)) = m$, $T_m(1) = 1$ και $T_m(z)$ είνα άρτια, αντίστοιχα, περιττή συνάρτηση αν $m$ άρτιος, αντίστοιχα, περιττός.
2. Έχουμε,
$$T_m(z) = \begin{cases} \cos(m \cos^{-1} z), & -1 \leq z \leq 1 \\ \cosh(m \cosh^{-1} z), & z \geq 1 \end{cases}$$
δηλαδή, $T_m(z) = \cos(m\theta)$, όπου $z = \cos(\theta), \theta \in [0, \pi]$, αν $-1 \leq z \leq 1$, και $T_m(z) = \cosh(mu)$, όπου $z = \cosh u, u \geq 0$, αν $z \geq 1$.
3. $T_m(z) = \frac{1}{2} \left[ \left(z + (z^2 - 1)^{1/2}\right)^m + \left(z - (z^2 - 1)^{1/2}\right)^m \right]$

Τα πολυώνυμα Chebyshev είναι χρήσιμα στην περίπτωσή μας λόγω της εξής σημαντικής ιδιότητας:

**Λήμμα 2.6.** *Έστω $0 < a < b$. Τότε το πρόβλημα* "min–max"

$$\min_{\pi \in \mathbb{P}_m, \pi(0)=1} \left( \max_{a \leq z \leq b} |\pi(z)| \right) \tag{2.18}$$

*λύνεται μοναδικά από το πολυώνυμο*

$$\widetilde{\pi}_m(z) = T_m\left(\frac{b + a - 2z}{b - a}\right) / \, T_m\left(\frac{b + a}{b - a}\right), \tag{2.19}$$

*για το οποίο ισχύει*

$$\max_{a \leq z \leq b} |\widetilde{\pi}(z)| = 1 / \, T_m\left(\frac{b + a}{b - a}\right). \tag{2.20}$$

*Απόδειξη.* Για $m \geq 0$ έχουμε $\max_{-1\leq y\leq 1}|T_m(y)| = 1$ και η μέγιστη αυτή απόλυτη τιμή λαμβάνεται στα σημεία $y_j = \cos(j\pi/m)$, $j = 0, 1, \ldots, m$, όπου $T_m(y_j) = (-1)^j$. Θεωρούμε τον γραμμικό μετασχηματισμό $z \mapsto y, y = (b+a-2x)/(b-a)$ που απεικονίζει αμφιμονοσήμαντα το διάστημα $a \leq z \leq b$ πάνω στο διάστημα $-1 \leq y \leq 1$. Για το πολυώνυμο $\widetilde{\pi}(z), a \leq z \leq b$ στη σχέση (2.19) έχουμε $\widetilde{\pi}(0) = 1$ και

$$\max_{a\leq z\leq b}|\widetilde{\pi}(z)| = \max_{-1\leq y\leq 1}|T_m(y)|/T_m\left(\frac{b+a}{b-a}\right) = 1/T_m\left(\frac{b+a}{b-a}\right),$$

δηλαδή ισχύσει η (2.20). Η μέγιστη αυτή τιμή του $\widetilde{\pi}(z)$ λαμβάνεται σε $m+1$ διακριτά σημεία $z_j, j = 0, \ldots, m$, όπου το $\widetilde{\pi}$ έχει εναλλασσόμενο πρόσημο. Έστω τώρα ότι υπάρχει άλλο πραγματικό πολυώνυμο $q_m(z)$, βαθμού $\leq m$, τέτοιο ώστε $q_m(0) = 1$ και $\max_{a\leq z\leq b}|q_m(z)| < \max_{a\leq z\leq b}|\widetilde{\pi}(z)|$. Τότε το πολυώνυμο $p_m(z) = q_m(z) - \widetilde{\pi}(z)$ είναι βαθμού $\leq m$ και έχει $m+1$ διακριτές ρίζες. Άρα $p_m(z) = 0$, άτοπο. Συνεπώς το πρόβλημα $\widetilde{\pi}(z)$ λύνει το πρόβλημα "min–max" (2.18). □

Είμαστε τώρα σε θέση να αποδείξουμε τις σχέσεις (2.16).

**Λήμμα 2.7.** *Αν οι αριθμοί $\sigma_j$ ορίζονται από την* (2.15) *τότε ισχύουν οι σχέσεις* (2.16).

*Απόδειξη.* Από το Λήμμα 2.6 και τη σχέση (2.15) συμπεραίνουμε ότι

$$\sigma_j = 1/T_j\left((\lambda_{\max} + \lambda_{\min})/(\lambda_{\max} - \lambda_{\min})\right),$$

αν $\lambda_{\min} \neq \lambda_{\max}$. Συνεπώς για $\kappa \neq 1$, $j \geq 1$ έχουμε

$$\sigma_j = 1/T_j\left((\kappa + 1)/(\kappa - 1)\right).$$

Χρησιμοποιώντας την παράσταση (3) των πολυωνύμων Chebyshev που δώσαμε παραπάνω, για $z = (\kappa + 1)/(\kappa - 1)$, παίρνουμε

$$\sigma_j = f(\kappa^{-1/2})\left[(\kappa^{1/2} - 1)/(\kappa^{1/2} + 1)\right]^j, \quad j \geq 1, \tag{2.21}$$

όπου για $x \in [0, 1]$ η συνάρτηση $f(x)$ ορίζεται ως

$$f(x) = 2(1 + x)^{2j}/\left[(1 + x)^{2j} + (1 - x)^{2j}\right],$$

για την οποία ισχύει $1 = f(0) \leq f(x) \leq f(1) = 2$, $x \in [0, 1]$. Συνεπώς, λόγω της (2.21) και του ότι $\kappa \geq 1$, μια εκτίμηση του $\sigma_j$ δίνεται από την (2.16). □

Ερχόμαστε τώρα στο πρόβλημα της αποτελεσματικής υλοποίησης του αλγορίθμου της μεθόδου των συζυγών κλίσεων που αναπτύξαμε στην προηγούμενη ενότητα. Κατ' αρχήν, από τα Λήμματα 2.1 και 2.2 προκύπτει ότι για $k \geq 2$ το διάνυσμα $p^k$ είναι γραμμικός συνδιασμός των $p^{k-1}$ και $r^{k-1}$. Για $k = 2$ αυτό είναι προφανές λόγω της (2.8). Για $k \geq 2$, υποθέτοντας πάντα ότι $r^{k-1} \neq 0$, θεωρούμε το διάνυσμα $z^{k-1} \in \mathbb{R}^{k-1}$ της (2.6) το οποίο το γράφουμε ως $z^{k-1} = (w, \mu)^T$, με $w \in \mathbb{R}^{k-2}$ και $\mu \in \mathbb{R}, \mu \neq 0$. Χρησιμοποιώντας την (2.6) και την (2.2) με $i = k - 1$ έχουμε

$$p^k = (1 + \frac{\mu}{a_{k-1}})r^{k-1} + s^{k-1}, \tag{2.22}$$

όπου,

$$s^{k-1} = -\frac{\mu}{a_{k-1}}r^{k-2} - AP_{k-2}w. \tag{2.23}$$

Η (2.23) δίνει τώρα ότι $(s^{k-1}, r^{k-1}) = 0$ επειδή τα $r^i$ είναι ορθογώνια και, λόγω της (2.8), $AP_{k-2}w = A(w_1 p^1 + \cdots + w_{k-2}P^{k-2}) \in \langle Ab, A^2b, \ldots, A^{k-2}b \rangle \subset \langle r^0, \ldots, r^{k-2} \rangle$. Συνεπώς, από το Πυθαγόρειο Θεώρημα και την (2.22) έχουμε

$$\|p^k\|^2 = (1 + \frac{\mu}{a_{k-1}})^2 \|r^{k-1}\|^2 + \|s^{k-1}\|^2. \tag{2.24}$$

Το $z^{k-1} = (w, \mu)$ είναι εκείνο το στοιχείο του $\mathbb{R}^{k-1}$ που λύνει το πρόβλημα ελαχίστων τετραγώνων (2.7). Η κατασκευή του είναι ισοδύναμη με τον καθορισμό εκείνων των $w, \mu$ που ελαχιστοποιούν το $\|p^k\| = \|r^{k-1} - AP_{k-1}]z^{k-1}\|$, δηλαδή το δεξί μέλος της (2.24). Επειδή, πό την (2.23), η ποσότητα

$$\|s^{k-1}\|^2 = \left(\frac{\mu}{a_{k-1}}\right)^2 \|r^{k-2} - AP_{k-2}w'\|^2, \quad w' = \frac{a_{k-1}w}{\mu}$$

ελαχιστοποιείται ως προς $w'$ για $r^{k-2} - AP_{k-2}w' = p^{k-1}$, θα πρέπει το $s^{k-1}$ να είναι πολλαπλάσιο του $p^{k-1}$. Η (2.22) αποδεικνύει τον ισχυρισμό.

Χωρίς περιορισμό της γενικότητας υποθέτουμε ότι

$$p^k = r^{k-1} + \beta_k p^{k-1}, \quad k \geq 2. \tag{2.25}$$

Επειδή τα $p^k, p^{k-1}$ είναι $A$-συζυγή η (2.25) δίνει

$$\beta_k = -(p^{k-1}, Ar^{k-1})/(Ap^{k-1}, p^{k-1}). \tag{2.26}$$

Οι δύο αυτές σχέσεις ορίζουν πλήρως το $p^k$. Εξ' άλλου, επειδή $(p^{k-1}, r^{k-1}) = 0$, η (2.25) και η (1.12) δίνουν ότι

$$a_k = \|r^{k-1}\|^2/(Ap^k, p^k). \tag{2.27}$$

Οδηγούμαστε λοιπόν στην εξής υλοποίηση του αλγορίθμου της μεθόδου των συζυγών κλίσεων:

```
x^0 = 0
for k = 1, 2, ... do
    r^{k-1} = b - A x^{k-1}
    if r^{k-1} = 0 then
        x = x^{k-1}, τέλος
    else
        if k = 1 then
            p^k = r^0
        else
            β_k = -(p^{k-1}, A r^{k-1})/(A p^{k-1}, p^{k-1})
            p^k = r^{k-1} + β_k p^{k-1}
        end
        α_k = ||r^{k-1}||^2/(A p^k, p^k)
        x^k = x^{k-1} + α_k p^k
    end
end
x = x^n
```

Ο αλγόριθμος απαιτεί δύο πολλαπλασιασμούς του πίνακα $A$ επί διάνυσμα σε κάθε βήμα. Παρατηρώντας όμως ότι τα υπόλοιπα μπορούν να υπολογισθούν αναδρομικά μέσω της σχέσης $r^j = r^{j-1} - a_j A p^j$, έχουμε, χρησιμοποιώντας τη σχέση αυτή για $j = k-1$ και την ορθογωνιότητα των $r^{k-1}, r^{k-2}$, ότι $\|r^{k-1}\|^2 = -a_{k-1}(r^{k-1}, Ap^{k-1})$ και $\|r^{k-2}\|^2 = -a_{k-1}(r^{k-2}, Ap^{k-1}) = a_{k-1}(p^{k-1}, Ap^{k-1})$ (η τελευταία σχέση λόγω του ότι $p^{k-1} = r^{k-2} + \beta_{k-1}p^{k-2}$). Συνεπώς, $\beta_k = -(p^{k-1}, Ar^{k-1})/(Ap^{k-1}, p^{k-1}) = \|r^{k-1}\|^2/\|r^{k-2}\|^2$, αποφεύγοντας έτσι τον πολλαπλασιασμό $Ar^{k-1}$. Η τελική μορφή του αλγορίθμου της μεθόδου των συζυγών κλίσεων, όπως την παρουσίασαν οι Hestens και Stiefel (1952) είναι η ακόλουθη:

```
x^0 = 0
r^0 = b
for k = 1, 2, ... do
    r^{k-1} = b - A x^{k-1}
    if r^{k-1} = 0 then
        x = x^{k-1}, τέλος
    else
        if k = 1 then
            p^k = r^0
        else
            β_k = ||r^{k-1}||^2 / ||r^{k-2}||^2
            p^k = r^{k-1} + β_k p^{k-1}
        end
        α_k = ||r^{k-1}||^2 / (A p^k, p^k)
        x^k = x^{k-1} + α_k p^k
        r^k = r^{k-1} - a_k A p^k
    end
end
x = x^n
```

**Παρατήρηση 2.8.** *Η απόδειξη της βασικής ιδιότητας της μεθόδου των συζυγών κλίσεων, δηλαδή ότι βρίσκει τη λύση σε $n$ βήματα, έγινε με την προϋπόθεση βέβαια ότι όλοι υπόλογισμοί στον παραπάνω αλγόριθμο γίνονται ακριβώς. Στην πράξη όμως τα σφάλματα στρογγύλευσης που οφείλονται στην αριθμητική πεπερασμένης ακρίβειας καταστρέφουν την ορθογωνιότητα των $r^j$ έτσι ώστε τελικά το $r^n$ να μην είναι μηδέν. Παρ' όλα αυτά, η μέθοδος όντως ελαττώνει την τιμή του συναρτησιακού $\phi(x)$ από βήμα σε βήμα οπότε μπορεί να χρησιμοποιηθεί στην πράξη ως επαναληπτική μέθοδος. Η επανάληψη δεν σταματάει στο βήμα $n$ αλλά συνεχίζει μέχρι να ικανοποιηθούν κάποια από τα γνωστά κριτήρια τερματισμού. Σημειώνουμε τέλος ότι μια και η κύρια πράξη στον αγλοριθμο των συζυγλων κλίσεων είναι ο πολλαπλασιασμός πίνακα επί διάνυσμα, η μέθοδος είναι ιδιαίτερα αποτελεσματική για αραιούς πίνακες.*

**Παρατήρηση 2.9.** *Η ταχύτητα της μεθόδου των συζυγών κλίσεων είναι $1 - 2\kappa^{-1/2} + O(\kappa^{-1})$ όταν $\kappa \to \infty$, η οποία είναι αρκετά αργή για μεγάλο $\kappa$. Μια σημαντική τεχνική που χρησιμοποιείται για την επιτάχυνση τηε σύγκλισης είναι η λεγόμενη προρύθμιση (preconditioning), για την υλοποίηση της οποίας ο αλόριθμος απαιτεί τη λύση ενός $n \times n$ γραμμικού συστήματος με συμμετρικό, θετικά ορισμένο πίνακα, τον προρυθμιστή $M$. Η επιλογή λοιπόν του προρυθμιστή είναι σημαντικό πρόβλημα: πρέπει να έχει απλή δομή έτσι ώστε η λύση συστημάτων με πίνακα $M$ να μην απαιτεί πολλές πράξεις και, επιπλέον, ο πίνακας $\widehat{A} = M^{-1}A$ να έχει λόγο ιδιοτιμών μικρότερο του $\kappa$ ώστε πράγματι να επιταχύνεται η σύγκλιση.*

## Ασκήσεις

**Άσκηση 2.1.** *Αναπτύξτε τη μέθοδο των συζυγών κλίσεων για* $x^0 \neq 0$*. Ειδικώτερα, βρείτε τα ανάλογα των Λημμάτων 2.1 και 2.2. Σημειώστε, π.χ. ότι το* $x^i$ *λύνει τώρα το πρόβλημα* (2.9).

**Άσκηση 2.2.** *Χρησιμοποιήστε τον ορισμό των πολυωνύμων Chebyshev για να αποδείξετε τις ιδιότητες τους που χρησιμοποιήθηκαν σε αυτή την ενότητα.*

**Άσκηση 2.3.** *Έστω* $r \in \mathbb{R}^n$ *και* $M$ *υπόχωρος του* $\mathbb{R}^n$*. Λέμε ότι το* $p \in M$ *είναι η* ορθή προβολή *του* $r$ *στον* $M$ *αν* $(r,x) = (p,x)$, $x \in M$*, Δείξτε ότι:*

1. *η προβολή* $p$ *του* $r$ *στον* $M$ *υπάρχει, είναι μοναδική και ικανοποιεί* $\|p\|^2 = \|r\|^2 - \|r-p\|^2$.
2. *το* $r - p$ *είναι προβολή του* $r$ *στον υπόχωρο* $M^{\perp}$.
3. $\|r - p\| = \min_{x \in M} \|r - x\|$ *και ότι η ιδιότητα αυτή χαρακτηρίζει μονοσήμαντα την προβολή* $p$ *του* $r$.
4. *για κάθε* $B \in \mathbb{R}^{m \times n}$, $\mathcal{R}(B)^{\perp} = \mathcal{N}(B^T)$*, όπου* $\mathcal{N}(B^T) = \{x \in \mathbb{R}^m : B^T x = 0\}$.