Προβλήματα Βελτιστοποίησης Χωρίς περιορισμούς: Βασικοί Αλγόριθμοι / Αριθμητική Ανάλυση

Κ. Χρυσαφίνος (ΣΕΜΦΕ, ΕΜΠ).

Μέθοδοι Καθόδου

- Πρόβλημα (ΥΠ): Να υπολογιστούν σημεία τοπικού ή / και ολικού ελάχιστου για το πρόβλημα $\min_{x \in \mathbb{R}^n} f(x) = f(\bar{x})$.

- Υποθέτουμε ότι η $f \in C^1(\mathbb{R}^n)$. Αναγκαία συνθήκη: $\nabla f(\bar{x}) = 0$.

- Οι υπολογιστικές μέθοδοι, για δοσμένη αρχική συνθήκη $x_0 \in \mathbb{R}^n$, κατασκευάζουν ακολουθία $\{x_k\}_{k=0}^{\infty} \in \mathbb{R}^n$ που (στο όριο - με κατάλληλο τρόπο) προσεγγίζει τη λύση ή λύσεις του προβλήματος (ΥΠ).

- Μέθοδοι καθόδου: Για δοσμένη αρχική συνθήκη $x_0 \in \mathbb{R}^n$ αναζητούμε κατεύθυνση (καθόδου) $p_k \in \mathbb{R}^n$, και ένα συντελεστή $a_k \in \mathbb{R}_+$, ώστε η ακολουθία που παράγεται με από την επαναληπτική διαδικασία:
  $$x_{k+1} = x_k + a_k p_k, \quad k = 0,1,2,...$$
  να συγκλίνει σε υποψήφιο σημείο ελαχίστου (ολικού / τοπικού) ελαχίστου.

- Το κύριο κριτήριο για είναι η προσεγγιστική διαδικασία “μέθοδος καθόδου” πρέπει να επιλέγεται ως:
  $$p_k^T \nabla f(x_k) < 0, \qquad \text{αν } \nabla f(x_k) \neq 0$$
  $$p_k = 0, \qquad \text{αν } \nabla f(x_k) = 0$$

- Το κύριο κριτήριο για είναι η προσεγγιστική διαδικασία “μέθοδος καθόδου” πρέπει η κατεύθυνση να επιλέγεται ως:
$$p_k^T \nabla f(x_k) < 0, \quad \text{αν} \quad \nabla f(x_k) \neq 0$$
$$p_k = 0, \quad \text{αν} \quad \nabla f(x_k) = 0$$
Μία κατεύθυνση που επιλέγεται με αυτόν τον τρόπο ονομάζεται κατεύθυνση καθόδου.

- Χρησιμοποιώντας το Θεώρημα Μέσης υπάρχει $\mu \in (0,1)$, ώστε $f(x_k + a_k p_k) = f(x_k) + a_k \nabla f(x_\kappa + \mu a_k p_k)^T p_k$
Επομένως για a_k αρκετά μικρό και επειδή η $f \in C^1(\mathbb{R}^n)$, η από την συνθήκη της κατεύθυνσης καθόδου, συνεπάγεται ότι $f(x_k + a_k p_k) < f(x_k)$.

- Απομένει να υπολογιστεί η σταθερά a_k καθώς και η κατεύθυνση καθόδου p_k.

- Υπολογισμός βήματος a_k: Έστω ότι έχει επιλέγει η κατεύθυνση p_k. Ένας τρόπος (ευθύς) για την εύρεση του βέλτιστου βήματος a_k, είναι ο υπολογισμός ελαχίστου της συνάρτησης $g(a) := f(x_k + ap_k)$.

- Υπολογίζοντας $g'(a) = 0$, εύκολα διαπιστώνουμε ότι το a_k ικανοποιεί την σχέση $\nabla f(x_k + a_k p_k)^T p_k = 0$, ή ισοδύναμα $\nabla f(x_{k+1})^T p_k = 0$.

- <u>*Σημείωση*</u>: Έκτος από συγκεκριμένες περιπτώσεις (π.χ $f(x) = \dfrac{1}{2}x^T A x - b^T x + c$, με $A \in \mathbb{R}^{n\times n}$ συμμετρικό και θετικά ορισμένο), ο υπολογισμός του a_k μέσω της σχέσης $\nabla f(x_{k+1})^T p_k = 0$ δεν είναι εύκολος και πρακτικός. Πολλές φορές είναι μάλιστα αδύνατος.

- Line Search Techniques: Χρησιμοποιούνται ευρύτατα για ένα πρακτικό υπολογισμό του a_k .

Line Search Techniques

- Δεν αναζητείται το βέλτιστο $\alpha \in \mathbb{R}_+$ αλλά εκείνο που επιτυγχάνει ικανοποιητική μείωση της τιμής της συνάρτησης $g(a)$.

- Δηλαδή, ο αλγόριθμος δοκιμάζει μία “ακολουθία” τιμών του $\alpha \in \mathbb{R}_+$, και με βάση κατάλληλο κριτήριο, σταματά όταν έχει πετύχει ικανοποιητική μείωση της $g(a)$.

- Μία λογική επιλογή κριτηρίου είναι: Το $\alpha \in \mathbb{R}_+$ να επιλέγεται κατά τέτοιο τρόπο ώστε να ισχύει, $f(x_{k+1}) < f(x_k)$, (για δοσμένα διανύσματα p_k, x_k). Όμως αν επιλέγει ένας αλγόριθμος που ξεκινάει από μία σχετικά μεγάλη τιμή του $\alpha \in \mathbb{R}_+$, και σταδιακά γίνεται υποδιπλασιασμός του βήματος μέχρι να ικανοποιήσουμε την παραπάνω σχέση, υπάρχει ο κίνδυνος να πάρουμε εντελώς λανθασμένα αποτελέσματα.

- *<u>Δύο Δυσκολίες</u>*:
  1) Χαμηλός ρυθμός καθόδου
  2) Μικρές τιμές του βήματος.

- Ο χαμηλός ρυθμός καθόδου αντιμετωπίζεται αν η επιλογή του βήματος ικανοποιεί τη σχέση:
$$l(a_k) = \frac{1}{a_k}\left(f(x_k) - f(x_k + a_k p_k)\right) \geq -\sigma \nabla f(x_k)^T p_k$$
όπου $\sigma \in (0,1/2)$

- Οι μικρές τιμές του α αποφεύγονται απαιτώντας:
  $|\nabla f(x_k + a_k p_k)^T p_k| \leq \beta |\nabla f(x_k)^T p_k|$
  δηλαδή ο ρυθμός καθόδου (στην κ+1 επανάληψη) δεν είναι μικρότερος από τον ρυθμό καθόδου (στην κ-επανάληψη), όπου $\beta \in (\sigma, 1)$ και σ τέτοιο ώστε να ικανοποιεί και την σχέση $l(a_k) \geq -\sigma \nabla f(x_k)^T p_k$.

- Οι συνθήκες αυτές λέγονται ισχυρές συνθήκες Wolfe.
  $f(x_k) - f(x_k + a_k p_k) \geq -\sigma \alpha_\kappa \nabla f(x_k)^T p_k$
  $|\nabla f(x_k + a_k p_k)^T p_k| \leq \beta |\nabla f(x_k)^T p_k|$

- Τυπικές τιμές των παραμέτρων είναι: $\sigma \in [10^{-5}, 10^{-1}], \beta \in [10^{-1}, 1/2]$.

- Υπάρχουν πάντα κατάλληλα βήματα;

- Έστω $f : \mathbb{R}^n \to \mathbb{R}$ είναι συνεχώς παραγωγίσιμη και κάτω φραγμένη κατά μήκος της ημι-ευθείας $\{x_k + ap_k, a > 0\}$, όπου p_k είναι κατεύθυνση καθόδου. Αν $0 < \sigma < \beta < 1$ τότε υπάρχουν διαστήματα [c,C] που εμπεριέχουν μήκη βήματος a_k, που ικανοποιούν τις ισχυρές συνθήκες Wolfe.

Ο αλγόριθμος καθόδου για τετραγωνικές συναρτήσεις

- Έστω $f : \mathbb{R}^n \to \mathbb{R}, f(x) := \frac{1}{2}x^T A x - b^T x$ όπου $A \in \mathbb{R}^{n\times n}$ συμμετρικός και θετικά ορισμένος.
- Πρόβλημα: Να βρεθεί $\bar{x} \in \mathbb{R}^n$ ώστε $\min_{x\in\mathbb{R}^n} f(x) = f(\bar{x})$. Αναγκαία Συνθήκη: $\nabla f(\bar{x}) = 0$ ή $A\bar{x} = b$.
- Διαλέγουμε $p_k = -\nabla f(x_k)$.
- Παρατηρούμε ότι $p_k = -(Ax_k - b) = r_k$ (υπόλοιπο στο κ βήμα)
- Εύκολα διαπιστώνουμε πως το βέλτιστο βήμα υπολογίζεται αναλυτικά. Πράγματι, αρκεί να ελαχιστοποιήσουμε την ποσότητα

$$
\begin{aligned}
g(a) := f(x_k + ap_k) &= \frac{1}{2}(x_k + ap_k)^T A(x_k + ap_k) - b^T(x_k + ap_k) \\
&= \frac{a^2}{2}p_k^T Ap_k + \frac{1}{2}x_k^T Ax_k + ap_k^T Ax_k - b^T x_k - ap_k^T b && (A^T = A) \\
&= \frac{a^2}{2}p_k^T Ap_k + \frac{1}{2}x_k^T Ax_k + ap_k^T(Ax_k - b) - b^T x_k && (\text{ άλγεβρα }) \\
&= \frac{a^2}{2}p_k^T Ap_k + \frac{1}{2}x_k^T Ax_k - ap_k^T r_k - b^T x_k && (r_k = b - Ax_k) \\
&= \frac{a^2}{2}r_k^T Ar_k - ar_k^T r_k + \frac{1}{2}x_k^T Ax_k - b^T x_k && (p_k = r_k)
\end{aligned}
$$

- Επομένως, $g'(a) = ar_k^T A r_k - r_k^T r_k$ που σημαίνει ότι $\alpha_\kappa = \dfrac{r_k^T r_k}{r_k^T A r_k}$. Υπενθυμίζουμε πως A θετικά ορισμένος, και άρα $r_k^T A r_k > 0,$ όταν $r_k \neq 0$.
- Ορισμός A-νόρμας: Έστω $A \in \mathbb{R}^{n\times n}$, συμμετρικός και θετικά ορισμένος πίνακας, τότε η συνάρτηση $y \in \mathbb{R}^n \to \sqrt{y^T A y} \in \mathbb{R}_+$ ορίζει νόρμα που συμβολίζεται με $\|y\|_A = \sqrt{y^T A y}$. Επομένως, $\alpha_\kappa = \dfrac{\|r_k\|_2^2}{\|r_k\|_A^2}$.
- Εύκολα παρατηρούμε ότι

$$
\begin{aligned}
\frac{1}{2}\|x-\bar{x}\|_A^2 &= \frac{1}{2}(x-\bar{x})^T A(x-\bar{x}) \\
&= \frac{1}{2}\bar{x}^T A\bar{x} + \frac{1}{2}x^T A x - x^T A\bar{x} && (A = A^T) \\
&= \frac{1}{2}\bar{x}^T A\bar{x} + \frac{1}{2}x^T A x - x^T b && (A\bar{x} = b) \\
&= f(x) + \frac{1}{2}\bar{x}^T A\bar{x} \\
&= f(x) + \bar{x}^T A\bar{x} - \frac{1}{2}\bar{x}^T A\bar{x} \\
&= f(x) + \bar{x}^T b - \frac{1}{2}\bar{x}^T A\bar{x} && (A\bar{x} = b) \\
&= f(x) - f(\bar{x})
\end{aligned}
$$

- Επομένως, ο αλγόριθμος λαμβάνει την ακολουθεί μορφή :

1. Given $x_0 \in \mathbb{R}^n$

    Compute $r_0 = b - Ax_0,$

    and for $k = 0,..,$ until convergence

2. $a_k = \dfrac{r_k^T r_k}{r_k^T A r_k}$
3. $x_{k+1} = x_k + a_k r_k$
4. $r_{k+1} = r_k - a_k A r_k$

- *<u>Θεώρημα 9</u>*: Έστω $A \in \mathbb{R}^{n \times n}$ συμμετρικός και θετικά ορισμένος. Τότε η μέθοδος συγκλίνει για κάθε $x_0 \in \mathbb{R}^n$ και ισχύει:

$$\|x_{k+1} - \bar{x}\|_A \leq \frac{k_2(A) - 1}{\kappa_2(A) + 1} \|x_k - \bar{x}\|_A.$$

- Υπενθύμιση: $\kappa_2(A) = |||A|||_2 |||A^{-1}|||_2 = \dfrac{\lambda_{max}}{\lambda_{min}}$ όπου $\lambda_{max}, \lambda_{min}$ η μέγιστη και η ελάχιστη ιδιοτιμή. Αν ο δείκτης κατάστασης είναι μεγάλος, τότε από τον τύπο του σφάλματος παρατηρούμε πως η σύγκλιση μπορεί να είναι αργή.

Η μέθοδος κλίσεων για γενικές μη-γραμμικές συναρτήσεις

- Σε περίπτωση που εφαρμόζουμε την μέθοδο καθόδου σε μία μη γραμμική συνάρτηση $f : \mathbb{R}^n \to \mathbb{R}$, (που δεν είναι τετραγωνική) τότε:
  1. Given $x_0 \in \mathbb{R}^n$

     For $k = 0, \ldots,$ compute until convergence

  2. $p_k = -\nabla f(x_k)$
  3. Compute a_k such that $\min_{a \in \mathbb{R}} f(x_k + a p_k) = f(x_k + a_k p_k)$

     Numerically (e.g. by line search)

  4. $x_{k+1} = x_k + a_k p_k$

- *Θεώρημα 10*: Έστω $f : \mathbb{R}^n \to \mathbb{R}$, $f \in C^2(\mathbb{R}^n)$ και πως τα σημεία που παράγονται σε κάθε επανάληψη της μεθόδου με βάση την γραμμική αναζήτησή συγκλίνουν σε ένα σημείο $\bar{x} \in \mathbb{R}^n$. Αν ο Εσσιανός πίνακας $H(\bar{x})$ είναι θετικά ορισμένος με μέγιστη και ελάχιστη ιδιοτιμή $\lambda_{max}, \lambda_{min}$ τότε ισχύει

$$f(x_{k+1}) - f(\bar{x}) = r^2 \left(f(x_k) - f(\bar{x}) \right), \quad \text{για κάποιο } r \in \left(\frac{\lambda_{max} - \lambda_{min}}{\lambda_{max} + \lambda_{min}}, 1 \right)$$

Μέθοδοι Συζύγων κατευθύνσεων

- Ένα σύνολο μη μηδενικών διανυσμάτων $\{p_i\}_{i=0}^{n-1}$ καλείται συζυγές ως προς τον πίνακα $A \in \mathbb{R}^{n\times n}$, συμμετρικό και θετικά ορισμένο αν ισχύει $p_i^T A p_j = 0 \quad i, j = 0,\ldots,n-1, \quad i \neq j.$

- Η παραπάνω σχέση περιγράφει "ορθογωνιότητα" ορισμένη από το Α-εσωτερικό γινόμενο.

- *<u>Θεώρημα 11</u>*: Έστω διανύσματα $\{p_i\}_{i=0}^{n-1}$ συζυγή ως προς $A \in \mathbb{R}^{n\times n}$ συμμετρικό και θετικά ορισμένο. Τότε είναι γραμμικά ανεξάρτητα.

- *<u>Απόδειξη</u>*: Πράγματι, αν υποθέσουμε πως είναι συζυγή ως προς τον πίνακα συμμετρικό και θετικά ορισμένο πίνακα $A \in \mathbb{R}^{n\times n}$ και γραμμικώς εξαρτημένα. Τότε υπάρχουν

$c_i \in \mathbb{R}, \quad i = 0,..,n-1$ όχι όλα μηδεν, τέτοια ώστε $\displaystyle\sum_{i=0}^{n-1} c_i p_i = 0.$

Τότε έχουμε $A\left(\displaystyle\sum_{i=0}^{n-1} c_i p_i\right) = \displaystyle\sum_{i=0}^{n-1} c_i A p_i = 0.$ Επομένως, παίρνοντας το εσωτερικό γινόμενο με p_k και

χρησιμοποιώντας το γεγονός ότι τα διανύσματα είναι Α - συζυγή, $p_k^T\left(\displaystyle\sum_{i=0}^{n-1} c_i A p_i\right) = c_k p_k^T A p_k = 0.$

Άτοπο, καθώς το $p_k \neq 0$, και Α θετικά ορισμένος.

- Η μέθοδος των συζύγων κλίσεων για το πρόβλημα: Να βρεθεί $\bar{x} \in \mathbb{R}^n$ ώστε $\min_{x \in \mathbb{R}^n} f(x) = f(\bar{x})$ όπου $f : \mathbb{R}^n \to \mathbb{R}, f(x) := \frac{1}{2}x^T Ax - b^T x$ με$A \in \mathbb{R}^{n \times n}$ συμμετρικός και θετικά ορισμένος.

- Έστω διανύσματα $\{p_i\}_{i=0}^{n-1}$ συζυγή ως προς τον πίνακα $A \in \mathbb{R}^{n \times n}$, και $x_0 \in \mathbb{R}^n$ αρχικό διάνυσμα. Τότε μία αρχική μορφή του αλγορίθμου είναι:

  1. Given $x_0 \in \mathbb{R}^n$

     Compute $r_0 = b - Ax_0$,

     and for $k = 0,..,$ until convergence

  2. $a_k = \dfrac{p_k^T r_k}{p_k^T A p_k}$
  3. $x_{k+1} = x_k + a_k p_k$
  4. $r_{k+1} = r_k - a_k A p_k$

  Παρατηρούμε πως όταν $p_k = r_k$, λαμβάνουμε τον κλασικό αλγόριθμο των κλίσεων. Ο βέλτιστος συντελεστής a_k, προκύπτει ελαχιστοποιώντας την συνάρτηση $g(a) := f(x_k + ap_k)$ (Άσκηση: Να κάνετε αναλυτικά τις πράξεις).

- *Θεώρημα 12*: Για κάθε $x_0 \in \mathbb{R}^n$, ο αλγόριθμος των συζυγών κατευθύνσεων συγκλίνει στην λύση του γραμμικού συστήματος σε το πολύ n-βήματα.

- *Απόδειξη*: Υπενθυμίζουμε πως οι κατευθύνσεις $\{p_i\}_{i=0}^{n-1}$ ορίζουν σύνολο γραμμικά ανεξάρτητων διανυσμάτων, επομένως παράγουν τον $\mathbb{R}^n$. Επομένως, υπάρχουν $c_i \in \mathbb{R}, x_0 - \bar{x} = \sum_{i=0}^{n-1} c_i p_i$, που σημαίνει $A(x_0 - \bar{x}) = A\left(\sum_{i=0}^{n-1} c_i p_i\right)$. Επομένως έχουμε,

$$\begin{aligned} p_k^T A(x_0 - \bar{x}) &= p_k^T \sum_{i=0}^{n-1} c_i A p_i \\ &= \sum_{i=0}^{n-1} c_i p_k^T A p_i \qquad \text{(γραμ. εσωτ. γινομένου)} \\ &= c_k p_k^T A p_k \qquad \text{(A-συζυγή)} \end{aligned}$$

Άρα, επειδή ο A είναι θετικά ορισμένος και $p_k \neq 0$, λαμβάνουμε από την παραπάνω σχέση,

$$c_k = \frac{p_k^T A(x_0 - \bar{x})}{p_k^T A p_k}.$$

Όμως $x_k = x_0 + \sum_{i=0}^{\kappa-1} a_i p_i$ από τον αλγόριθμο.

- *(συνέχεια απόδειξης):* Επομένως, $A(x_k - x_0) = A\sum_{i=0}^{\kappa-1} a_i p_i$. Άρα, δουλεύοντας με παρόμοιο τρόπο έχουμε, $p_k^T A(x_k - x_0) = \sum_{i=0}^{k-1} a_i p_k^T A p_i = 0$.

  Άρα έχουμε

$$
\begin{aligned}
p_k^T A(x_0 - \bar{x}) &= p_k^T A(x_0 - x_k + x_k - \bar{x}) \\
&= p_k^T A(x_k - \bar{x}) && (p_k^T A(x_0 - x_k) = 0) \\
&= p_k^T (Ax_x - b) && (A\bar{x} = b) \\
&= p_k^T r_k
\end{aligned}
$$

  Επομένως $c_k = \dfrac{p_k^T A(x_0 - \bar{x})}{p_k^T A p_k} = \dfrac{p_k^T r_k}{p_k^T A p_k} := a_k$.

- Στην παραπάνω σχέση αποδείξαμε ότι οι συντελεστές $c_k = a_k$, και άρα ο αλγόριθμος σταματά σε n- το πολύ βήματα.

- *Θεώρημα 13*: Για κάθε $x_0 \in \mathbb{R}^n$, ο αλγόριθμος των συζυγών κατευθύνσεων παράγει ακολουθία προσεγγίσεων $\{x_k\}_{i=0,\ldots}$ ικανοποιεί τις σχέσεις:$r_k^T p_i = 0,$ για $i = 0,..,k-1.$ Επιπλέον το $x_k \in \mathbb{R}^n$ είναι το σημείο ελαχίστου της $f : \mathbb{R}^n \to \mathbb{R}, f(x) := \frac{1}{2}x^T Ax - b^T x$, πάνω στον υπόχωρο $S = \{x \in \mathbb{R}^n \,|\, x = x_0 + \sum_{i=0}^{k-1} c_i p_i,$ με $c_i \in \mathbb{R}\}$.

- *Απόδειξη*: Έστω $y \in S$. Θα δείξουμε ότι το $y \in S$ ελαχιστοποιεί την f στο S αν και μόνο αν $r^T p_i = 0,\ i = 0,..,k-1,$ όπου $r := b - Ay$. Έστω ότι το $y \in S$ ελαχιστοποιεί την f στο S. Κάθε $x \in S$, γράφεται ως $x = x_0 + \sum_{i=0}^{k-1} c_i p_i$, για κάποια $c_i \in \mathbb{R}$. Επομένως, υπάρχουν κάποιες σταθερές $\bar{c}_i$ ώστε $y = x_0 + \sum_{i=0}^{k-1} \bar{c}_i p_i$, για κάποια $\bar{c}_i \in \mathbb{R}$. Ορίζουμε την συνάρτηση $\Phi(c_0, \ldots, c_{k-1}) := f(x_0 + c_0 p_0 + \ldots + c_{k-1} p_{k-1})$. Εύκολα αποδεικνύεται πως υπάρχει μοναδικό σημείο ελαχίστου της Φ, το οποίο υπολογίζεται από τις σχέσεις $\frac{\partial \Phi}{\partial c_i}(\bar{c}) = 0,\ i = 0,\ldots,k-1$, όπου $\bar{c} = (\bar{c}_0, \ldots, \bar{c}_{k-1})^T$, ή ισοδύναμα $\nabla \Phi(x_0 + \bar{c}_0 p_0 + \ldots + \bar{c}_{k-1} p_{k-1})^T p_i = 0\ i = 0,\ldots,k-1$, δηλαδή, $\left(A(x_0 + \bar{c}_0 p_0 + \ldots + \bar{c}_{k-1} p_{k-1}) - b\right)^T p_i = r^T p_i = 0,\ i = 1,\ldots,k-1$. Αντιστρόφως έστω τώρα ότι για το $y \in S$, ισχύει $r^T p_i = 0,\ i = 0,..,k-1,$ όπου $r := b - Ay$. Εύκολα διαπιστώνουμε πως είναι το ελάχιστο της f στο S.

- *Συνέχεια Απόδειξης* : Απομένει να δείξουμε πως $r := r_k$. Θα χρησιμοποιήσουμε επαγωγή. Για κ=1, παρατηρούμε πως,

$x_1 = x_0 + a_0 p_0 = x_0 + \dfrac{r_0^T p_0}{p_0^T A p_0} p_0$, και

$$
\begin{aligned}
r_1^T p_0 &= (b - Ax_1)^T p_0 \\
&= \left(b - A(x_0 + \frac{r_0^T p_0}{p_0^T A p_0} p_0) \right)^T p_0 && \text{(αντικαθ. το } x_1\text{)} \\
&= \left(b - Ax_0 - \frac{r_0^T p_0}{p_0^T A p_0} A p_0 \right)^T p_0 && \text{(αλγεβρα)} \\
&= r_0^T p_0 - \frac{r_0^T p_0}{p_0^T A p_0} p_0^T A p_0 && (A = A^T) \\
&= r_0^T p_0 - r_0^T p_0 = 0
\end{aligned}
$$

Έστω ότι ισχύει για $\kappa - 1$, δηλαδή $r_{k-1}^T p_i = 0$, για $i = 0,..,k-2$. Θα δείξουμε ότι ισχύει και για κ, δηλαδή ότι $r_k^T p_i = 0$, για $i = 0,..,k-1$. Υπενθυμίζουμε πως $r_k = r_{k-1} - a_{k-1} A p_{k-1}$.

- *Συνέχεια απόδειξης* : Επομένως έχουμε, για δείκτη $k-1$,

$$
\begin{aligned}
r_k^T p_{k-1} &= (r_{k-1} - a_{k-1} A p_{k-1})^T p_{k-1} \\
&= r_k^T p_{k-1} - a_{k-1} p_{k-1}^T A^T p_{k-1} \\
&= r_k^T p_{k-1} - \frac{r_{k-1}^T p_{k-1}}{p_{k-1}^T A p_{k-1}} p_{k-1}^T A p_{k-1} \\
&= r_k^T p_{k-1} - r_{k-1}^T p_{k-1} = 0
\end{aligned}
$$

ενώ για όλους του υπόλοιπους δείκτες $\iota = 0,..,\kappa - 2$,

$$
\begin{aligned}
r_k^T p_i &= (r_{k-1} - a_{k-1} A p_{k-1})^T p_i \\
&= r_{k-1}^T p_i - a_{k-1} p_{k-1}^T A^T p_i \\
&= 0 - \frac{r_{k-1}^T p_{k-1}}{p_{k-1}^T A p_{k-1}} p_{k-1}^T A p_i = 0
\end{aligned}
$$

όπου χρησιμοποιήσαμε την επαγωγική υπόθεση $r_{k-1}^T p_i = 0,\ i = 0,..k-2$ και (Α- συζυγή) $p_{k-1}^T A p_i = 0$.

- *Υπολογισμός Α-συζυγών κατευθύνσεων*:
  Αναζητούμε κατευθύνσεις $\{p_i\}_{i=0}^{n-1}$, ώστε να είναι Α-ορθογώνιες μεταξύ τους, δηλαδή να ικανοποιούν $p_i^T A p_j = 0 \quad i, j = 0,\ldots,n-1, \quad i \neq j$.

- *Θεώρημα 14*: Υποθέτουμε ότι οι κατευθύνσεις αυτές επιλέγονται από το επαναληπτικό σχήμα:
  $p_k = r_k - \beta_{k-1} p_{k-1}$ για κατάλληλα επιλεγμένα $\beta_k \in \mathbb{R},\ k = 1,..$
  $p_0 = r_0$
  Θα δείξουμε ότι οι κατευθύνσεις που κατασκευάζονται μ' αυτό τον τρόπο είναι Α-ορθογώνιες.

- *Απόδειξη*: Οι συντελεστές β_k υπολογίζονται εύκολα χρησιμοποιώντας τις σχέσεις της Α-ορθογωνιότητας. Πράγματι,

$$
\begin{aligned}
0 &= p_k^T A p_{k-1} \\
&= (r_k - \beta_{k-1} p_{k-1} p_{k-1})^T A p_{k-1} \\
&= r_k^T A p_{k-1} - \beta_{k-1} p_{k-1}^T A p_{k-1}
\end{aligned}
$$

δηλαδή, $\beta_{k-1} = \dfrac{r_k^T A p_{k-1}}{p_{k-1}^T A p_{k-1}},\ k = 1,2,\ldots$

Επαγωγή: Για κ=1, έχουμε $p_1^T A p_0 = (r_1 - \beta_0 p_0)^T A p_0 = = r_1^T A p_0 - \dfrac{r_1^T A p_0}{p_0^T A p_0} p_0^T A p_0 = 0$.

- (<u>*συνέχεια απόδειξης*</u>): Επιπλέον, έστω ότι ισχύει για $\iota = 0{,}1{,}...{,}\kappa - 1$, δηλαδή ότι $p_i^T A p_j = 0, \ i, j = 0{,}..{,}k-1 \ \ i \neq j$ θα δείξουμε ότι ισχύει και για κ.
  Παρατηρούμε ότι, για δείκτες $\iota = 0{,}1{,}...{,}\kappa - 2$,

  $$\begin{aligned} p_i^T A p_k &= p_i^T A(r_k - \beta_{k-1} p_{k-1}) \\ &= p_i^T A r_k - \beta_{k-1} p_i^T A p_{k-1} \\ &= p_i^T A r_k && \text{ο δείκτης } i \neq k-1 \text{ άρα } p_i^T A p_{k-1} = 0 \\ &= p_i^T A r_k \,. \end{aligned}$$

  Εύκολα δείχνουμε (χρησιμοποιώντας τον ορισμό του β_{k-1}), $p_{\kappa-1}^T A p_\kappa = 0$.
  Επομένως, αρκεί να δείξουμε ότι $p_\iota^T A r_k = 0$, για $i = 0{,}1{,}..{,}k-2$, για να ολοκληρώσουμε την απόδειξη. Χρησιμοποιώντας την σχέση $r_{k-1} = r_{k-2} - a_{k-2} A p_{k-2}$ και επειδή $r_0 = p_0$, $p_{k-1} = r_{k-1} - \beta_{k-2} p_{k-2}$ λαμβάνουμε ότι $span(p_0, p_1, \ldots, p_{k-1}) = span(r_0, r_1, \ldots, r_{k-1})$ και επομένως $Ap_{k-2} \in span(r_0, r_1, \ldots, r_{k-1}) = span(p_0 . p_1, \ldots, p_{k-1})$. Λόγω του Θεωρήματος 13, έχουμε ότι $Ap_{k-2} \perp r_k$ που ολοκληρώνει την απόδειξη

- Αλγόριθμος της μεθόδου των συζύγων κλίσεων για ελαχιστοποίηση συνάρτησης $f : \mathbb{R}^n \to \mathbb{R}, f(x) := \frac{1}{2}x^T Ax - b^T x$, με $A \in \mathbb{R}^{n\times n}$ συμμετρικό και θετικά ορισμένο.

  1. Given $x_0 \in \mathbb{R}^n$

     Compute $r_0 = b - Ax_0,\ p_0 = r_0$

     and for $k = 0,..,$ until convergence

     2. $a_k = \dfrac{p_k^T r_k}{p_k^T A p_k}$

     3. $x_{k+1} = x_k + a_k p_k$

     4. $r_{k+1} = r_k - a_k A p_k$

-
     5. $\beta_k = \dfrac{(Ap_k)^T r_{k+1}}{(Ap_k)^T p_k}$

     6. $p_{k+1} = r_{k+1} - \beta_k p_k$

- *Θεώρημα 15*: Η μέθοδος των συζύγων κλίσεων συγκλίνει σε το πολύ n-βήματα. Επιπλέον, το σφάλμα $e_k = x_k - \bar{x},$ για k < n , ικανοποιεί, $\|e_k\|_A \leq \dfrac{2r^k}{1 + r^{2k}}\|e_0\|_A$ με $r = \dfrac{\sqrt{K_2(A)} - 1}{\sqrt{K_2(A)} + 1}$.

- Επισημαίνεται πως ανά βήμα, απαιτείται μόνο ένας πολλαπλασιασμός πίνακα επί διάνυσμα.

- Η μέθοδος αυτή ανήκει στην κατηγορία Matrix - free μεθόδων. Μία εναλλακτική γραφή είναι:

1. Given $x_0 \in \mathbb{R}^n$

   Compute $r_0 = b - Ax_0,\ p_0 = r_0$

   and for $k = 0,..,$ until convergence

   2. $a_k = \dfrac{r_k^T r_k}{p_k^T A p_k}$

   3. $x_{k+1} = x_k + a_k p_k$

   4. $r_{k+1} = r_k - a_k A p_k$

   5. $\beta_k = -\dfrac{r_{k+1}^T r_{k+1}}{r_k^T r_k}$

   6. $p_{k+1} = r_{k+1} - \beta_k p_k$

Η μέθοδος των συζύγων συναρτήσεων για μη-τετραγωνικές συναρτήσεις.

- Η μέθοδος των συζύγων κλίσεων μπορεί να εφαρμοστεί (με διάφορες παραλλαγές) στο γενικό πρόβλημα $\min_{x \in \mathbb{R}^n} f(x) = f(\bar{x})$ (εδώ η συνάρτηση δεν είναι τετραγωνική).
- Τότε όμως οι συντελεστές α_κ, β_k δεν μπορούν να υπολογιστούν με τους προηγούμενους τύπους.
- Για το α_κ μπορεί να χρησιμοποιηθεί μία τεχνική “γραμμικής αναζήτησης” (line search technique), ενώ το β_k δεν προσδιορίζεται με μοναδικό τρόπο. Μία επιλογή που έχει προταθεί από τους Fletcher & Reeves είναι:

$$\beta_0 = 0,\ \beta_k = \frac{\|\nabla f(x_k)\|_2^2}{\|\nabla f(x_{k-1})\|_2^2},\ k = 1, \ldots$$

$1.$ Given $x_0 \in \mathbb{R}^n$

Compute $p_0 = r_0 = \nabla f(x_0)$

and for $k = 0, \ldots,$ until convergence

$2.$ Approximate a_k such that $f(x_k + a_k p_k) \approx \min_{a \in \mathbb{R}_+} f(x_k + a p_k)\ (e.g.\ linesearch)$

$3.$ $x_{k+1} = x_k + a_k p_k$

$4.$ $r_{k+1} = \nabla f(x_{k+1})$

$5.$ If $k = 0$ then $\beta_0 = 1,$

else $\beta_k = \dfrac{\|\nabla f(x_k)\|_2^2}{\|\nabla f(x_{k-1})\|_2^2}$

$6.$ $p_{k+1} = r_{k+1} - \beta_k p_k$

Η μέθοδος των συζυγών κλίσεων με προρύθμιση (The preconditioned conjugate gradient method).

- Η μέθοδος των συζύγων κλίσεων για ελαχιστοποίηση συνάρτησης $f : \mathbb{R}^n \to \mathbb{R}, f(x) := \dfrac{1}{2}x^T Ax - b^T x$, με $A \in \mathbb{R}^{n\times n}$ συμμετρικό και θετικά ορισμένο

- Υποθέτουμε ότι η $f \in C^1(\mathbb{R}^n)$. Αναγκαία συνθήκη: $\nabla f(\bar{x}) = 0$ ή ισοδύναμα $A\bar{x} = b$.

- Τεχνικές Προρύθμισης: Σε περίπτωση που ο δείκτης κατάστασης του πίνακα $A \in \mathbb{R}^{n\times n}$ δεν είναι καλός, μετατρέπουμε το σύστημα $A\bar{x} = b$ σε σύστημα της μορφής $BA\bar{x} = Bb$, με κατάλληλη επιλογή πίνακα $B \in \mathbb{R}^{n\times n}$ ώστε το καινούριο (προρυθμισμένο) σύστημα να λύνεται γρήγορα και εύκολα, καθώς και να έχει βελτιωμένο δείκτη κατάστασης. Η επιλογή του προρυθμιστή (preconditioner) εξαρτάται από τη δομή του αρχικού συστήματος.

- Έστω $C \in \mathbb{R}^{n\times n}$, συμμετρικός και θετικά ορισμένος με ιδιοτιμές $\lambda_i \geq 0,\ i = 1,\ldots,n$ . . Υπενθυμίζουμε πως από το φασματικό θεώρημα της γραμμικής άλγεβρας υπάρχει ορθογώνιος πίνακας $U \in \mathbb{R}^{n\times n}$, ώστε $U^TCU = D$, όπου ο $D \in \mathbb{R}^{n\times n}$ διαγώνιος με $d_{ii} = \lambda_i,\ i = 1,\ldots,n$. Επομένως μπορούμε να ορίσουμε τον $C^{1/2}$.

- Ορίζουμε $y = C^{1/2}x$, ισοδύναμα το αρχικό σύστημα γράφεται $C^{-1/2}AC^{-1/2}y = C^{-1/2}b$
  Στην πράξη δεν χρειάζεται να υπολογίσουμε τους πίνακες $C^{1/2}$ και $C^{-1/2}$.
  1. Given $x_0 \in \mathbb{R}^n$

     Compute $r_0 = b - Ax_0$, $z_0 = C^{-1}r_0$ $p_0 = z_0$

     and for $k = 0,..,$ until convergence

     2. $a_k = \dfrac{z_k^T r_k}{p_k^T A p_k}$

     3. $x_{k+1} = x_k + a_k p_k$

     4. $r_{k+1} = r_k - a_k A p_k$

     5. $C z_{k+1} = r_{k+1}$

     6. $\beta_k = \dfrac{z_{k+1}^T r_{k+1}}{z_k^T p_k}$

     7. $p_{k+1} = z_{k+1} + \beta_k p_k$

- Το υπολογιστικό κόστος αυξάνεται, καθώς πρέπει να επιλυθεί το γραμμικό σύστημα του βήματος 5. Παρατηρούμε όμως ότι ο προρυθμιστής επιλέγεται κατά τέτοιο τρόπο ώστε το να λύνεται εύκολα και γρήγορα.

Βιβλιογραφία

- Αλ. Μπακόπουλος και Ιων. Χρυσοβέργης, “Εισαγωγή στην Αριθμητική Ανάλυση”, Εκδόσεις Συμμετρία.
- Γ. Ακρίβης και Β. Δουγαλής, “Εισαγωγή στην Αριθμητική Ανάλυση”, Πανεπιστημιακές Εκδόσεις Κρήτης.
- J. Nocedal and S. Wright, “Numerical Optimization”, Springer-Verlag 2006.
- A. Quarteroni, R. Sacco, and F. Saleri, “Numerical Mathematics”, Springer-Verlag 2007.
- E. Polak, “Optimization”, Springer-Verlag 1997