Προσεγγιστικές Μέθοδοι τύπου Newton-Raphson για προβλήματα βελτιστοποίησης χωρίς περιορισμούς

Κ. Χρυσαφίνος, Σ.Ε.Μ.Φ.Ε.

Newton-Raphson ως Μέθοδος 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 $f \in C^2(\mathbb{R}^n)$. 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).
- Επομένως μπορούμε να εφαρμόσουμε την κλασική μέθοδο Newton-Raphson για την επίλυση του μη-γραμμικού συστήματος $\nabla f(\bar{x}) = 0$.
- 1. Given $x_0 \in \mathbb{R}^n$

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

  2. $H(x_k)p_k = -\nabla f(x_k)$ (Solve linear system)
  3. $x_{k+1} = x_k + p_k$
- Προφανώς πρέπει ο πίνακας $H(x_k)$, για $\kappa = 1, ..,$ να είναι αντιστρέψιμος. Υπενθύμιση: Επειδή ο Εσσιανός πίνακας $H(x_k)$, για $\kappa = 1, ..,$ είναι συμμετρικός, αν είναι και θετικά ορισμένος τότε εξασφαλίζεται η αντιστρεψιμότητα.

- Δηλαδή, σε περίπτωση που ο Εσσιανός πίνακας είναι θετικά ορισμένος σε μία κατάλληλη περιοχή $B(\bar{x}, \delta)$, για κατάλληλο $\delta > 0$, και το γραμμικό σύστημα $H(y)w = d$, λύνεται εύκολα για κάθε $y \in B(\bar{x}, \delta)$, τότε η διαδικασία μπορεί να εφαρμοστεί και οδηγήσει σε "τετραγωνική" τάξη ακρίβειας.

- Στην μέθοδο Newton-Raphson επιλέγεται ως κατεύθυνση $p_k = -H^{-1}(x_k)\nabla f(x_k)$, όπου $H(x_k)$ συμβολίζει τον Εσσιανό πίνακα της f στο x_k

- Υπενθυμίζουμε πως οι Μέθοδοι Καθόδου περιγράφονται ως εξής: 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.

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

- Εύκολα παρατηρούμε ότι αν $H(x_k)$, για $\kappa = 1,..,$ είναι θετικά ορισμένος τότε η κατεύθυνση $p_k = -H^{-1}(x_k)\nabla f(x_k)$ είναι κατεύθυνση καθόδου.

- Κάτω από παρόμοιες υποθέσεις (βλ. Μάθημα Αριθμητικής Γραμμική Άλγεβρας), αποδεικνύεται ότι η μέθοδος έχει τοπικά τετραγωνική τάξη ακρίβειας, δηλαδή, υπάρχει κατάλληλη περιοχή $B(\bar{x}, \delta)$, για κατάλληλο $\delta > 0$ ώστε $\{f(x_k)\}_{k=1}^{\infty} \in \bar{B}(\bar{x}, \delta)$ και σταθερά $C > 0$ ώστε $\|x_{k+1} - \bar{x}\| \leq C\|x_k - \bar{x}\|^2$.

- Η μέθοδος δεν χρησιμοποιείται από τον αρχικό υπολογισμό, έκτος αν το $x_0 \in \mathbb{R}^n$ είναι κατάλληλα κοντά στο ελάχιστο $\bar{x}$ καθώς αν δεν είναι κατάλληλα κοντά, τότε δεν μπορεί να εξασφαλιστεί η αντιστρεψιμότητα του Εσσιανού πίνακα και επομένως δεν θα παράγεται κατεύθυνση καθόδου.

- Αν ο Εσσιανός πίνακας δεν είναι θετικά ορισμένος, υπάρχει ο κίνδυνος η μέθοδος να συγκλίνει σε σαγματικό σημείο (saddle point) ή να συγκλίνει σε μέγιστο.

- Συνήθως επιλέγεται μία πιο αργή μέθοδος στην αρχή. Στην συνέχεια όταν μετά από κάποια βήματα βελτιωθεί η προσέγγιση, επιλέγεται η Netwon-Raphson για την ολοκλήρωση των υπολογισμών.

- Newton-Raphson (with a shift):
  1. Given $x_0 \in \mathbb{R}^n$

     and for $k = 0,..,$ until convergence
  2. Find suitable shift a_k Compute $\bar{H}(x_k) = H(x_k) + a_k I$
  3. $\bar{H}(x_k)p_k = -\nabla f(x_k)$ (Solve linear system)
  4. $x_{k+1} = x_k + p_k$

Quasi-Newton Methods

- Quasi - Newton-Raphson:
  1. Given $x_0 \in \mathbb{R}^n$

     and for $k = 0,..,$ until convergence
  2. Compute $H(x_k)$ or an approximation $\tilde{H}(x_k)$
  3. Solve $H(x_k)p_k = -\nabla f(x_k)$ or $\tilde{H}(x_k)p_k = -\nabla f(x_k)$
  4. Compute accelaration paremeter a_k
  5. $x_{k+1} = x_k + a_k p_k$

- Προσοχή: Η παράμετρος επιτάχυνσης μπορεί να υπολογιστεί με την βοήθεια τεχνικών γραμμικής αναζήτησης.

- *Θεώρημα 15* (Σύγκλισης): Έστω $f \in \mathbb{R}^n \to \mathbb{R}, f \in C^1(\mathbb{R}^n)$, τέτοια ώστε $\|\nabla f(x) - \nabla f(y)\|_2 \leq C\|x - y\|_2$ για κάποια σταθερά $C > 0$. Έστω ότι για μία μέθοδος καθόδου παράγει ακολουθία επαναλήψεων $\{x_k\}_{k=0}^{\infty}$, όπου οι συντελεστές $a_k \in \mathbb{R}_+$ του όρου διόρθωσης υπολογίζονται από τεχνικές γραμμικής αναζήτησης (line search techniques) μέσω των ισχυρών συνθηκών Wolfe. Τότε , ισχύει *ένα* από τα ακόλουθα ενδεχόμενα:
  1. $\nabla f(x_k) = 0$ για κάποιο δείκτη k
  2. $\lim_{k\to\infty} f(x_k) = -\infty$
  3. $\lim_{k\to\infty} \dfrac{\nabla f(x_k)^T p_k}{\|p_k\|_2^2} = 0$

- Παρατηρήσεις: Έστω ότι η ακολουθία $\{x_k\}_{k=0}^{\infty}$ συγκλίνει σε κάποιο σημείο $\bar{x} \in \mathbb{R}^n$ για κάποια αρχική συνθήκη $x_0 \in \mathbb{R}^n$. Τότε στην γενική περίπτωση δεν εξασφαλίζεται η σύγκλιση σε κάποιο υποψήφιο σημείο ελαχίστου. Το παραπάνω θεώρημα σύγκλισης παρουσιάζει επιπλέον ιδιότητες (Lipschitz συνέχεια, ενδεχόμενο 3 για τις κατευθύνσεις) ώστε να εξασφαλιστεί η σύγκλιση σε υποψήφιο σημείο ελαχίστου.

- Δηλαδή, εκτός από τις (παθολογικές) περιπτώσεις που το διάνυσμα κατεύθυνσης γίνεται πολύ μεγάλο (ως μέγεθος σε σχέση με το διάνυσμα κλίσης) ή ορθογώνιο (ως προς το διάνυσμα κατεύθυνσης) τότε κάθε όριο ακολουθίας $\{x_k\}_{k=0}^{\infty}$, συγκλίνει σε υποψήφιο σημείο ελαχίστου.

- Quasi-Newton Methods (Secant Type Methods): 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 $f \in C^2(\mathbb{R}^n)$. Αναγκαία συνθήκη: $\nabla f(\bar{x}) = 0$.

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

   and for $k = 0,..,$ until convergence
2. Compute an approximation $\tilde{H}(x_k)$ of Hessian$H(x_k)$
3. Solve $\tilde{H}(x_k)p_k = -\nabla f(x_k)$
4. Compute accelaration paremeter a_k
5. $x_{k+1} = x_k + a_k p_k$

- Πως μπορεί να υπολογιστεί ο $\tilde{H}(x_k)$;

- Υπενθύμιση: Επειδή η $f \in C^2(\mathbb{R}^n)$ ο Εσσιανός πίνακας $H(x_k)$ είναι συμμετρικός επομένως και η "προσέγγιση του" $\tilde{H}(x_k)$ πρέπει να είναι συμμετρικός πίνακας.

- Επιπλέον, αν $\tilde{H}(x_k)$ συμμετρικός τότε προφανώς και ο επόμενος $\tilde{H}(x_{k+1})$ πρέπει να οριστεί κατά τέτοιο τρόπο ώστε να είναι συμμετρικός.

- Ας υποθέσουμε ότι $\{a_k = 1\}_{k=0,\ldots}$ και επομένως η μέθοδος γράφεται (για θεωρητικούς και μόνο σκοπούς)
  $x_{k+1} = x_k - \tilde{H}(x_k)\nabla f(x_k) := x_k + p_k$.

- Εφαρμόζοντας Θεώρημα Taylor:
  $\nabla f(x_{k+1}) \approx \nabla f(x_k) + \tilde{H}(x_{K+1})(x_{k+1} - x_k)$ ή

  $\nabla f(x_k) \approx \nabla f(x_k + p_k) + \tilde{H}(x_{k+1})(x_k - x_{k+1})$

  δηλαδή, έχουμε $\tilde{H}(x_{k+1})p_k \approx \nabla f(x_{k+1}) - \nabla f(x_k)$.

- Ένας επαναληπτικός τρόπος υπολογισμού των πινάκων $\tilde{H}(x_k)$ είναι:

$$\tilde{H}(x_{k+1}) = \tilde{H}(x_k) + \frac{\left(\nabla f(x_{k+1}) - \nabla f(x_k) - \tilde{H}(x_k)p_k\right)\left(\nabla f(x_{k+1}) - \nabla f(x_k) - \tilde{H}(x_k)p_k\right)^T}{\left(\nabla f(x_{k+1}) - \nabla f(x_k) - \tilde{H}(x_k)p_k\right)^T p_k}$$

- Εύκολα αποδεικνύεται πως η επαναληπτική ακολουθία των πινάκων $\tilde{H}(x_k)$, που ορίζεται με τον παραπάνω τύπο, παράγει συμμετρικούς πίνακες.

- Προσοχή: Ο αλγόριθμος μπορεί να παρουσιάσει προβλήματα αριθμητικής αστάθειας σε περίπτωση που $\left(\nabla f(x_{k+1}) - \nabla f(x_k) - \tilde{H}(x_k)p_k\right)^T p_k \approx 0$

- Επισήμανση: Η παραπάνω επαναληπτική διαδικασία δεν εξασφαλίζει ότι η ακολουθία $\tilde{H}(x_k)$ αποτελείται από θετικά ορισμένους πίνακες.

- Υπάρχουν αρκετές διαφορετικές επιλογές διαδικασιών προσέγγισης του Εσσιανού πίνακα.

- <u>*Άσκηση:*</u> Έστω $f : \mathbb{R}^n \to \mathbb{R}, f(x) := \dfrac{1}{2}x^T Ax - b^T x$ όπου $A \in \mathbb{R}^{n \times n}$ συμμετρικός και θετικά ορισμένος. Να αποδειχθεί ότι η f είναι αυστηρά κυρτή.

- <u>*Απόδειξη:*</u> Θεωρούμε το πολυώνυμο $p(a) = f(x + a(y - x)) - f(x) - a[f(y) - f(x)]$ Εύκολα διαπιστώνουμε πως το παραπάνω πολυώνυμο είναι δευτέρου βαθμού ως προς α. Αρκεί να δείξουμε ότι $p(a) < 0$, για κάθε $a \in (0,1), x \neq y$. Τότε,
$p(a) = f(x + a(y - x)) - f(x) - a[f(y) - f(x)] < 0$ που σημαίνει ότι
$$f(x + a(y - x)) < f(x) + a[f(y) - f(x)]$$
$$= (1 - a)f(x) + af(y)$$
και επομένως η f είναι αυστηρά κυρτή. Παρατηρούμε ότι, $p(0) = 0$,$p(1) = 0$. Επίσης,
$$p(a) = \frac{1}{2}(x + a(y - x))^T A(x + a(y - x)) - b^T(x + a(y - x))$$
$$-(1 - a)\left(\frac{1}{2}x^T Ax - b^T x\right) - a\left(\frac{1}{2}y^T Ay - b^T y\right)$$
$$= \frac{a^2}{2}(y - x)^T A(y - x) + ax^T A(y - x) + \frac{1}{2}x^T Ax - b^T x - ab^T(x - y)$$
$$-(1 - a)\left(\frac{1}{2}x^T Ax - b^T x\right) - a\left(\frac{1}{2}y^T Ay - b^T y\right)$$

- *(συνέχεια απόδειξης):* επομένως έχουμε,

$$p(a) = \frac{a^2}{2}(y-x)^T A(y-x) + ax^T A(y-x) + \frac{1}{2}x^T Ax - b^T x - ab^T(x-y)$$

$$-(1-a)\left(\frac{1}{2}x^T Ax - b^T x\right) - a\left(\frac{1}{2}y^T Ay - b^T y\right)$$

$$= \frac{a^2}{2}(y-x)^T A(y-x) - ab^T(x-y) + ax^T A(y-x)$$

$$+a\left(\frac{1}{2}x^T Ax - b^T x\right) - a\left(\frac{1}{2}y^T Ay - b^T y\right)$$

$$= \frac{a^2}{2}(y-x)^T A(y-x) - ab^T(x-y) + ax^T A(y-x)$$

$$+a\left(\frac{1}{2}x^T Ax - \frac{1}{2}y^T Ay\right)$$

Παρατηρούμε ότι επειδή, $a \in (0,1), x \neq y$, και $A \in \mathbb{R}^{n\times n}$ είναι θετικά ορισμένος, έχουμε ότι $\frac{1}{2}(y-x)^T A(y-x) > 0$, έχουμε ότι $p(a) < 0$, $a \in (0,1), x \neq y$.

- *Άσκηση:* Έστω $f:\mathbb{R}^n \to \mathbb{R}, f(x) := \dfrac{1}{2}x^T A x - b^T x$ όπου $A \in \mathbb{R}^{n\times n}$ συμμετρικός και θετικά ορισμένος. Να αποδειχθεί ότι η f είναι πιεστική.

- *Απόδειξη:* Θέτουμε $u = \dfrac{x}{\|x\|_\infty}$, για $x \neq 0$. Τότε,

  $f(x) = \|x\|_\infty^2 \left(\dfrac{1}{2}u^T A u - \dfrac{b^T u}{\|x\|_\infty}\right)$. Παρατηρούμε ότι η συνάρτηση

  $u \in S := \{u \in \mathbb{R}^n \,|\, \|u\|_\infty = 1\} \to u^T A u \in \mathbb{R}_+$

  είναι συνεχής και ορισμένη στο συμπαγές (ως κλειστό και φραγμένο) υποσύνολο του $\mathbb{R}^n$. Επομένως υπάρχει τουλάχιστον ένα σημείο ελαχίστου, $\bar{u} \in S$, και επειδή ο Εσσιανός πίνακας είναι θετικά ορισμένος έχουμε $u^T A u \geq \bar{u}^T A \bar{u} = \mu > 0$. Επομένως έχουμε,

  $$\begin{aligned} f(x) &= \|x\|_\infty^2 \left(\frac{1}{2}u^T A u - \frac{b^T u}{\|x\|_\infty}\right) \\ &\geq \|x\|_\infty^2 \left(\frac{1}{2}\bar{u}^T A \bar{u} - \frac{\|b\|_1 \|u\|_\infty}{\|x\|_\infty}\right) \\ &\geq \|x\|_\infty^2 \left(\frac{1}{2}\mu - \frac{\|b\|_1 \|u\|_\infty}{\|x\|_\infty}\right) = \|x\|_\infty^2 \left(\frac{1}{2}\mu - \frac{\|b\|_1}{\|x\|_\infty}\right) \end{aligned}$$

  Παρατηρούμε πως $\left(\dfrac{1}{2}\mu - \dfrac{\|b\|_1}{\|x\|_\infty}\right) > 0$, όταν $\|x\|_\infty \to \infty$. Επομένως, $\lim\limits_{\|x\|_\infty \to \infty} f(x) = \infty$.

- *Άσκηση*: Έστω A συμμετρικός και θετικά ορισμένος. Να δείξετε ότι η συνάρτηση $f : \mathbb{R}^n \to \mathbb{R}, f(u) := \frac{1}{2}u^T Au - b^T u + 2\sum_{i=1}^{n} cosh(u_i)$ ικανοποιεί την σχέση: $\forall u, v \in \mathbb{R}^n,\ u \neq v,\ \lambda \in (0,1)$
$$\lambda f(u) + (1-\lambda)f(v) - f(\lambda u + (1-\lambda)v) > \frac{1}{2}\lambda(1-\lambda)\|u - v\|_A^2$$

- *Απόδειξη (υπόδειξη)*: Παρατηρούμε ότι $f(u) := g(u) + h(u)$ όπου $g(u) = \frac{1}{2}u^T Au - b^T u$ και $h(u) = \sum_{i=1}^{n} cosh(u_i)$. Παρατηρούμε πως η συνάρτηση cosh(.) είναι κυρτή συνάρτηση…
Εκτελώντας αλγεβρικές πράξεις έχουμε:
$$\begin{aligned}
&\lambda g(u) + (1-\lambda)g(v) - g(\lambda u + (1-\lambda)v) \\
&= \frac{1}{2}\lambda u^T Au - \lambda b^T u + \frac{1}{2}(1-\lambda)v^T Av - (1-\lambda)b^T v \\
&\quad - \frac{1}{2}(\lambda u + (1-\lambda)v)^T A(\lambda u + (1-\lambda)v) - b^T(\lambda u + (1-\lambda)v) \\
&= \frac{1}{2}\lambda u^T Au + \frac{1}{2}(1-\lambda)v^T Av - \frac{1}{2}\lambda^2 u^T Au - \frac{1}{2}(1-\lambda)^2 v^T Av - \frac{1}{2}\lambda(1-\lambda)u^T Av \\
&= \frac{1}{2}\lambda(1-\lambda)\|u - v\|_A^2
\end{aligned}$$

- *Άσκηση*:Έστω $f : \mathbb{R}^n \to \mathbb{R}, f(x) := \frac{1}{2}x^T A x - b^T x$ όπου $A \in \mathbb{R}^{100\times 100}$ συμμετρικός και θετικά ορισμένος, $A = \begin{bmatrix} 5 & -1 & 0 & \cdots & 0 \\ -1 & 5 & -1 & \ddots & 0 \\ 0 & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & \ddots & \ddots & -1 \\ 0 & \cdots & 0 & -1 & 5 \end{bmatrix}$, $b = \begin{bmatrix} 4 \\ 3 \\ \cdots \\ 3 \\ 4 \end{bmatrix}$.Να εκτελεστεί μία επανάληψη της μεθόδου των κλίσεων (με βέλτιστο βήμα) με αρχικό διάνυσμα $x_0 = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix}$. Να υπολογιστεί το πλήθος των επαναλήψεων που χρειάζεται ώστε το σφάλμα (σε κατάλληλη νόρμα) να είναι μικρότερο από δοσμένη παράμετρο ανοχής Tol.

- *Απόδειξη (υπόδειξη)*: (από κώδικα) Υπολογίζουμε το αρχικό υπόλοιπο, $r_0 = b - Ax_0 = b$. Υπολογίζουμε την σταθερά, $a_0 = \frac{r_0^T r_0}{r_0^T A r_0}$. Ξεκινώντας, $Ar_0 = [17,8,9,...,9,8,17]^T$. $a_0 \approx 0,32925072$. Επομένως, $x_1 = a_0 b = [1.31700288,0.98775216,....,1.31700288]^T$. (Με πολύ μικρό αριθμό επαναλήψεων έχουμε προσέγγιση της λύσης). Πραγματική λύση $\bar{x} = [1,1,...,1]^T$. Από τον τύπο του σφάλματος έχουμε,

- (<u>*συνέχεια απόδειξης*</u>) :

$$\|x_k - \bar{x}\|_A \leq \frac{k_2(A) - 1}{k_2(A) + 1}\|x_{k-1} - \bar{x}\|_A = \frac{\lambda_{max} - \lambda_{min}}{\lambda_{max} + \lambda_{min}}\|x_{k-1} - \bar{x}\|_A$$

όπου $\lambda_{max}, \lambda_{min}$ είναι η μέγιστη και ελάχιστη ιδιοτιμή του πίνακα A αντιστοίχως. Εφαρμόζοντας επαναληπτικά τον παραπάνω τύπο, λαμβάνουμε,

$\|x_k - \bar{x}\|_A \leq \left(\dfrac{\lambda_{max} - \lambda_{min}}{\lambda_{max} + \lambda_{min}}\right)^k \|x_0 - \bar{x}\|_A$. Για να εξασφαλίσουμε ότι το σφάλμα είναι μικρότερο από παράμετρο ανοχής Tol, αρκεί να ισχύει,

$\left(\dfrac{\lambda_{max} - \lambda_{min}}{\lambda_{max} + \lambda_{min}}\right)^k \|x_0 - \bar{x}\|_A \leq Tol$. Μπορούμε εύκολα να επιλύσουμε ως προς k, αρκεί να προσεγγιστούν τα $\lambda_{max}, \lambda_{min}$.

$$A = \begin{bmatrix} 5 & -1 & 0 & \cdots & 0 \\ -1 & 5 & -1 & \ddots & 0 \\ 0 & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & \ddots & \ddots & -1 \\ 0 & \cdots & 0 & -1 & 5 \end{bmatrix} = 5I - 5\begin{bmatrix} 0 & 1/5 & 0 & \cdots & 0 \\ 1/5 & 0 & 1/5 & \ddots & 0 \\ 0 & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & \ddots & \ddots & 1/5 \\ 0 & \cdots & 0 & 1/5 & 0 \end{bmatrix} := 5(I - M)$$

Επομένως, $I - M = (1/5)A$. Παρατηρούμε ότι $\|M\|_\infty = 2/5 < 1$.

- Άρα ο $I - M$ είναι αντιστρέψιμος και ισχύει, $\|(I-M)^{-1}\|_\infty \leq \dfrac{1}{1-\|M\|_\infty} = 5/3$, που σημαίνει ότι $5\|A^{-1}\|_\infty \leq 5/3,$ δηλαδή $\|A^{-1}\|_\infty \leq 1/3.$ Επομένως, $\rho(A^{-1}) = \max\{\,|\lambda|,\ \lambda$ ιδιοτιμή του $A^{-1}\}$

$$= \inf_{\|\cdot\| \text{ φυσ. νόρ.}} \|A^{-1}\| \leq \|A^{-1}\|_\infty \leq 1/3$$

που σημαίνει πως $\lambda_{max}(A^{-1}) = \dfrac{1}{\lambda_{min}(A)} \leq 1/3.$ Εύκολα υπολογίζεται πως $\lambda_{max} \leq \|A\|_\infty \leq 7.$ Άρα $\dfrac{\lambda_{max} - \lambda_{min}}{\lambda_{max} + \lambda_{min}} \leq \dfrac{1-3/7}{1+3/7} = 2/5.$ (Άσκηση: Μπορείτε να υπολογίσετε ένα μικρότερο άνω φράγμα ?)

Το θεώρημα σύγκλισης του Αλγόριθμου Κλίσεων

- Υπενθύμιση: 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 $k = 0,..,$ compute until convergence

     2. $p_k = -\nabla f(x_k)$

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

        Numerically (e.g. by line search)

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

- Με Χρήση του "μετρητή" $\delta_k := -\|\nabla f(x_k)\|_2^2$

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

     For $k = 0,..,$ compute until convergence

     2. $p_k = -\nabla f(x_k)$

     3.If $\delta_k = 0$ Stop. Else

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

           Numerically (e.g. by line search)

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

- *Θεώρημα:* Έστω $f \in \mathbb{R}^n \to \mathbb{R}, f \in C^1(\mathbb{R}^n)$. Υποθέτουμε ότι το σύνολο $S = \{x \in \mathbb{R}^n | f(x) \leq f(x_0)\}$ είναι φραγμένο. Αν ο αλγόριθμος σταματά για κάποιο κ, τότε $\delta_k := -\|\nabla f(x_k)\|_2^2$ = 0 και το $\bar{x} := x_k$ ικανοποιεί την αναγκαία συνθήκη. Διαφορετικά, η ακολουθία $\{x_k\}_{k=0,\ldots}$ που παράγει ο αλγόριθμος είναι άπειρη και κάθε όριο $\bar{x}$ συγκλίνουσας υπακολουθίας ικανοποιεί $\nabla f(\bar{x}) = 0$. Επιπλέον, ολόκληρη η ακολουθία $\lim_{k\to\infty} \delta_k = 0$.

- *Απόδειξη (υπόδειξη)*: Τα κύρια βήματα της απόδειξης στην περίπτωση που το βέλτιστο βήμα υπολογίζεται ακριβώς, δηλαδή, $g(a_k) = \min_{a\in\mathbb{R}} g(a) := \min_{a\in\mathbb{R}} f(x_k + ap_k) - f(x_k)$. Θα δείξουμε επαγωγικά ότι αν $\delta_k \neq 0$, η ακολουθία $\{x_k\}_{k=0,\ldots}$,είναι καλά ορισμένη, (δηλαδή υπάρχει το $x_k \in S$, για κάθε κ) και ότι η $\{f(x_k)\}_{k=0,\ldots}$ είναι φθίνουσα.

- *συνέχεια (απόδειξης):* Παρατηρούμε πως αναζητούμε ελάχιστο στο σύνολο S , καθώς για όλα τα υπόλοιπα x (που δεν ανήκουν στο S) προφανώς ισχύει $f(x) > f(x_0)$. Το σύνολο $S \subset \mathbb{R}^n$είναι κλειστό και φραγμένο και επομένως συμπαγές. Επομένως υπάρχει τουλάχιστον ένα σημείο ελαχίστου στο S. Έστω $x_0 \in S$. Αν $\delta_0 \neq 0$, τότε για να οριστεί το x_1, πρέπει να δείξουμε ότι το πρόβλημα $\min_{a\in\mathbb{R}} f(x_0 - a\nabla f(x_0)) - f(x_0)$ έχει λύση. Ορίζουμε το σύνολο $S_0 = \{x \in \mathbb{R}^n : x = x_0 - a\nabla f(x_0),\ a \in \mathbb{R}_+\} \cap S$. Παρατηρούμε πως το πρόβλημα $\min_{a\in\mathbb{R}} f(x_0 - a\nabla f(x_0)) - f(x_0)$ είναι ισοδύναμο με το $\min_{x\in S_0} f(x) - f(x_0)$. Προφανώς επειδή το S_0 είναι συμπαγές υπάρχει ελάχιστο. Επομένως υπάρχει α_0. Επομένως το $x_1 = x_0 - a_o\nabla f(x_0)$ είναι καλά ορισμένο. Παρατηρούμε επίσης πως

$$
\begin{aligned}
f(x_1) - f(x_0) &= f(x_0 - a_0\nabla f(x_0)) - f(x_0) \\
&= \min_{a\in\mathbb{R}^n} f(x_0 - a\nabla f(x_0)) - f(x_0) \\
&\leq f(x_0 - 0\nabla f(x_0)) - f(x_0) = 0.
\end{aligned}
$$

- *(συνέχεια απόδειξης)* : Επαγωγικά, χρησιμοποιώντας παρόμοια επιχειρήματα, αποδεικνύεται πως η ακολουθία $\{x_k\}_{k=0,...}$,είναι καλά ορισμένη, (δηλαδή υπάρχει το $x_k \in S,$ για κάθε κ) και ότι η $\{f(x_k)\}_{k=0,...}$ είναι φθίνουσα. Προφανώς αν ο αλγόριθμος σταματά για κάποιο $\delta_k := -\|\nabla f(x_k)\|_2^2 = 0$ το $\bar{x} := x_k$ ικανοποιεί την αναγκαία συνθήκη. Διαφορετικά, η ακολουθία $\{x_k\}_{k=0,...}$ που παράγει ο αλγόριθμος είναι άπειρη. Θα δείξουμε ότι $\delta_k \to \delta := 0$. Επειδή το $S \subset \mathbb{R}^n$είναι συμπαγές και ολόκληρη η ακολουθία $x_k \in S,$ για κάθε κ, έχουμε πως υπάρχει συγκλίνουσα υπακολουθία, δηλαδή υπάρχει $\bar{x} \in S$, ώστε $\lim_{k_n \to \infty} x_{k_n} = \bar{x}$. Επειδή, $f \in C^1(\mathbb{R}^n)$ έχουμε $\lim_{k_n \to \infty} -\|\nabla f(x_{k_n})\|_2^2 = -\|\nabla f(\bar{x})\|_2^2 := \delta \leq 0$. Θα δείξουμε ότι $\delta = 0$. Έστω ότι $\delta < 0$.

- *(συνέχεια απόδειξης)*: Θεωρούμε την συνάρτηση $h(a) := f(x_{k_n} - a\nabla f(x_{k_n}))$. Εύκολα υπολογίζουμε πως $h'(a) := -\nabla f(x_{k_n} - a\nabla f(x_{k_n}))^T \nabla f(x_{k_n})$. Από το θεώρημα Μέσης Τιμής, υπάρχει ένα $\mu_a \in (0,a)$ ώστε να ισχύει
$$h(a) - h(0) = -(a-0)\nabla f(x_{k_n} - \mu_a \nabla f(x_{k_n}))^T \nabla f(x_{k_n})$$
και επομένως καθώς
για $a \in (0,a'),\ k_n \geq k_{n_0},$ για κάποια a', k_{n_0} με $\epsilon_{k_n,a} \to 0$ όταν $k_n \to \infty,\ \ a \to 0$
$$\begin{aligned} f(x_{k_n} - a\nabla f(x_{k_n})) - f(x_{k_n}) &= -a\nabla f(x_{k_n} - \mu_a \nabla f(x_{k_n}))^T \nabla f(x_k) \\ &= a\left(-\nabla f(\bar{x})^T \nabla f(\bar{x}) + \epsilon_{k_n,a}\right) \\ &= a(\delta + \epsilon_{k_n,a}) \\ &\leq a\delta\beta, \text{ για κάποιο } \beta \in (0,1) \end{aligned}$$
Από τον ορισμό του βέλτιστου βήματος έχουμε, (παρατηρούμε ότι $(\delta + \epsilon_{k_n,a}) < 0$ καθώς $\epsilon_{k_n,a} \to 0$),
$$\begin{aligned} f(x_{k_n} - a_k \nabla f(x_{k_n})) - f(x_{k_n}) &\leq f(x_{k_n} - a\nabla f(x_{k_n})) - f(x_{k_n}) \\ &\leq a\delta\beta, \text{ για κάποιο } \beta \in (0,1) \end{aligned}$$
Επειδή $\delta < 0$, και η ακολουθία $\{f(x_k)\}_{k=0,\ldots}$ είναι φθίνουσα, η παραπάνω σχέση οδηγεί στο συμπέρασμα πως $\lim_{k_n \to \infty} f(x_{k_n}) = -\infty$, που είναι άτοπο καθώς $\lim_{k_n \to \infty} f(x_{k_n}) = f(\bar{x})$ αλλά το $\bar{x} \in S$ (όπου S φραγμένο).

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