

Μέθοδος της μέγιστης καθόδου

Η μέθοδος της μέγιστης καθόδου (*steepest descent*) προκύπτει επιλέγοντας

$$\Delta_k = I, \quad k = 1, 2, 3 \dots$$

όπου I ο μοναδιαίος $n \times n$ πίνακας. Η (5.2.6) λοιπόν γίνεται

$$x_{k+1} = x_k - \gamma_k \nabla f(x_k) \quad (5.2.7)$$

Στη μέθοδο της μέγιστης καθόδου, το διάνυσμα κατεύθυνσης $d_k = -\nabla f(x_k)$ ισούται με την αρνητική κλίση της f υπολογισμένης στο σημείο x_k . Αν κανονικοποιήσουμε το d_k βρίσκουμε $d_k = -\frac{\nabla f(x_k)}{|\nabla f(x_k)|}$. Για να προχωρήσουμε, παρατηρήστε:

$$-\nabla f^T(x_k) d_k \leq |\nabla f(x_k)| |d_k|$$

ή

$$\nabla f^T(x_k) d_k \geq -|\nabla f(x_k)| |d_k| \quad (5.2.8)$$

Για μοναδιαία διανύσματα κατεύθυνσης, $|d_k| = 1$ και η (5.2.8) γίνεται:

$$\nabla f^T(x_k) d_k \geq -|\nabla f(x_k)| \quad (5.2.9)$$

Από την (5.2.9) προκύπτει ότι το εσωτερικό γινόμενο $\nabla f^T(x_k) d_k$ ελαχιστοποιείται όταν ισχύει η ισότητα. Άρα όταν:

$$d_k = -\frac{\nabla f(x_k)}{|\nabla f(x_k)|} \quad (5.2.10)$$

Παρατηρούμε λοιπόν ότι απ' όλα τα μοναδιαία διανύσματα κατεύθυνσης d_k εκείνο που ελαχιστοποιεί την κλίση $\nabla f^T(x_k) d_k$ της $f(x_k + \gamma_k d_k)$ κατά την κατεύθυνση d_k με $\gamma_k = 0$, δίνεται από την (5.2.10). Εξαιτίας της παραπάνω ιδιότητας πήρε η μέθοδος την ονομασία της. Η μέθοδος της μέγιστης καθόδου συνοψίζεται ως εξής:

Αλγόριθμος 5.2.1 (Μέθοδος της Μέγιστης Καθόδου)

Αρχικοποίηση Ορίστε $\varepsilon > 0$ τη σταθερά τερματισμού. Επιλέξτε το αρχικό σημείο x_1 και θέστε $k = 1$.

Κύριο Βήμα Αν $|\nabla f(x_k)| < \varepsilon$ ο αλγόριθμος τερματίζει. Διαφορετικά, θέστε $d_k = -\nabla f(x_k)$ και ορίστε ως γ_k την τιμή που ελαχιστοποιεί την $f(x_k - \gamma_k \nabla f(x_k))$ ως προς γ_k με $\gamma_k \geq 0$. Θέστε $x_{k+1} = x_k - \gamma_k \nabla f(x_k)$, $k = k + 1$ και επαναλάβετε το κυρίως βήμα.

Το θεώρημα που ακολουθεί, συντελεί στην καλύτερη κατανόηση της γεωμετρίας της μεθόδου.

Θεώρημα 5.2.1 Αν $\{x_k\}$ είναι μια ακολουθία σημείων παραγόμενων από τη μέθοδο της μέγιστης κλίσης για μια συνάρτηση $f(x)$. Τότε για κάθε φυσικό αριθμό k , το διάνυσμα που ενώνει το x_k με το x_{k+1} είναι κάθετο στο διάνυσμα που ενώνει το x_{k+1} με το x_{k+2} .

Απόδειξη: Αρκεί να αποδείξουμε ότι το εσωτερικό γινόμενο $(x_{k+1} - x_k)^T(x_{k+2} - x_{k+1})$ είναι ίσο με το μηδέν. Από την έκφραση της μεθόδου της μέγιστης κλίσης προκύπτει:

$$(x_{k+1} - x_k)^T(x_{k+2} - x_{k+1}) = \gamma_k \gamma_{k+1} \nabla f^T(x_k) \nabla f(x_{k+1})$$

Αρκεί λοιπόν $\nabla f^T(x_k) \nabla f(x_{k+1}) = 0$. Από την έκφραση της μεθόδου

$$x_{k+1} = x_k - \gamma_k \nabla f(x_k)$$

Το βήμα γ_k επιλέγεται έτσι ώστε να ελαχιστοποιείται η $f(x_k - \gamma_k \nabla f(x_k))$ ως προς γ_k . Μ' άλλα λόγια:

$$\frac{df(x_k - \gamma_k \nabla f(x_k))}{d\gamma_k} = 0$$

ή

$$-\nabla f^T(x_k - \gamma_k \nabla f(x_k)) \nabla f(x_k) = 0$$

ή

$$-\nabla f^T(x_{k+1}) \nabla f(x_k) = 0$$

το οποίο και ολοκληρώνει την απόδειξη.

◇

Το Σχ. 5.6 φανερώνει τη γεωμετρία της μεθόδου, για μια συνάρτηση δύο μεταβλητών. Αν και θα περίμενε κανείς η μέθοδος της μέγιστης καθόδου να μας οδηγήσει στο ελάχιστο με τον μικρότερο αριθμό βημάτων (επαναλήψεων), στην πραγματικότητα κάτι τέτοιο δεν ισχύει. Όπως φανερώνει και το Θεώρημα 5.2.1, η πορεία προς το ελάχιστο είναι μια τεθλασμένη γραμμή, με τα διαδοχικά ευθύγραμμα τμήματα να σχηματίζουν γωνία 90° . Αυτή η $z_{ik} - z_{ak}$ συμπεριφορά, καθιστά τη μέθοδο ως επί το πλείστον αργή και κατά συνέπεια υπολογιστικά αναποτελεσματική. Η μόνη

Μέθοδος *Newton*

Η μέθοδος *Newton* προκύπτει αν στην (5.2.6) επιλέξουμε

$$\Delta_k = [\nabla^2 f(x_k)]^{-1}, \quad k = 0, 1, 2, \dots \quad (5.2.15)$$

με την προϋπόθεση ότι ο $\nabla^2 f(x_k)$ είναι θετικά ορισμένος. Αντικαθιστώντας την (5.2.15) στην (5.2.6) βρίσκουμε:

$$x_{k+1} = x_k - \gamma_k [\nabla^2 f(x_k)]^{-1} \nabla f(x_k) \quad (5.2.16)$$

Η μέθοδος του *Newton* λοιπόν αναζητά το σημείο ελαχίστου στην κατεύθυνση:

$$d_k = -[\nabla^2 f(x_k)]^{-1} \nabla f(x_k)$$

Η μέθοδος του *Newton* για την ελαχιστοποίηση συναρτήσεων συνοψίζεται παρακάτω.

Αλγόριθμος 5.2.2 (Μέθοδος *Newton*)

Αρχικοποίηση Ορίστε το σημείο εκκίνησης x_0 , τη σταθερά τερματισμού $\varepsilon \geq 0$ και θέστε $k = 1$.

Βήμα 1 Υπολογίστε την $\nabla f(x_k)$. Αν $|\nabla f(x_k)| < \varepsilon$ ο αλγόριθμος τερματίζει. Διαφορετικά, πηγαίνετε στο Βήμα 2.

Βήμα 2 Υπολογίστε τον $\nabla^2 f(x) = \left[\frac{\partial^2 f}{\partial x_i \partial x_j} \right] \quad i, j = 1, 2, \dots, n$.

Βήμα 3 Υπολογίστε την κατεύθυνση αναζήτησης $d_k = -[\nabla^2 f(x_k)]^{-1} \nabla f(x_k)$.

Βήμα 4 Θέστε $x_{k+1} = x_k + \gamma_k d_k$ όπου το $\gamma_k \geq 0$ υπολογίζεται έτσι ώστε να ελαχιστοποιεί την $f(x_k + \gamma_k d_k)$ ως προς γ_k .

Βήμα 5 Θέστε $k = k + 1$ και πηγαίνετε στο Βήμα 1.

Η ιδέα της μεθόδου του *Newton* είναι να ελαχιστοποιεί σε κάθε επανάληψη την τετραγωνική προσέγγιση της $f(x)$ γύρω από το x_k που δίνεται από:

$$q_k(x) = f(x_k) + \nabla f^T(x_k)(x - x_k) + \frac{1}{2}(x - x_k)^T \nabla^2 f(x_k)(x - x_k) \quad (5.2.17)$$

Παραγωγίζοντας την $q_k(x)$ και εξισώνοντας το αποτέλεσμα με το μηδέν προκύπτει:

$$\frac{dq_k(x)}{dx} = \nabla f(x_k) + \nabla^2 f(x_k)(x - x_k) = 0 \quad (5.2.18)$$

Η μέθοδος *Newton* μπορεί να παράγει ακολουθία σημείων $\{x_k\}$ με σκοπό την ελαχιστοποίηση της $f(x)$ που να ταλαντώνει ή να αποκλίνει. Το παράδειγμα που ακολουθεί είναι σχετικό.

Παράδειγμα 5.2.1 Θεωρήστε τη συνάρτηση $f(x) = x^{4/3}$. Για τη συνάρτηση αυτή βρίσκουμε ότι $\nabla f(x) = \frac{4}{3}x^{1/3}$ ενώ $\nabla^2 f(x) = \frac{4}{9}x^{-2/3}$. Η μέθοδος *Newton* δίνει:

$$\begin{aligned} x_{k+1} &= x_k - \gamma_k \frac{\frac{4}{3}x_k^{1/3}}{\frac{4}{9}x_k^{-2/3}} \\ &= x_k - 3\gamma_k x_k \\ &= (1 - 3\gamma_k)x_k, \quad \gamma_k > 0 \end{aligned}$$

Παρατηρούμε ότι η ακολουθία $\{x_k\}$ που δημιουργείται, θα συγκλίνει μόνο αν $|1 - 3\gamma_k| < 1$ ή $0 < \gamma_k < \frac{2}{3}$. Για κάθε άλλη επιλογή συντελεστή γ_k η μέθοδος θα αποκλίνει.

Υλοποίηση της μεθόδου *Newton*

Αντί της χρήσης της (5.2.16), υπολογίζουμε τους διαδοχικούς όρους της ακολουθίας της μεθόδου *Newton* λύνοντας το σύστημα των γραμμικών εξισώσεων:

$$\nabla^2 f(x_k)(x_{k+1} - x_k) = -\gamma_k \nabla f(x_k) \quad (5.2.21)$$

δοσμένου του x_k . Ο πίνακας $\nabla^2 f(x_k)$ είναι συμμετρικός αν η $f(x)$ έχει συνεχείς μερικές παραγώγους δεύτερης τάξης. Επομένως, ο υπολογιστικός φόρτος που απαιτείται για την υλοποίηση της (5.2.21) είναι ισοδύναμος με την επίλυση του γραμμικού συστήματος

$$Ax = b \quad (5.2.22)$$

όπου A είναι ένας $n \times n$ συμμετρικός πίνακας και b γνωστό διάνυσμα στο \mathbb{R}^n . Ένας τρόπος για τη λύση της (5.2.22) είναι η χρήση της μεθόδου *Gauss*. Υπάρχουν όμως και πιο αποτελεσματικές μέθοδοι, μια από τις οποίες θα αναφέρουμε παρακάτω.

Ας υποθέσουμε ότι ο πίνακας A μπορεί να παραγοντοποιηθεί ως

$$A = LU$$

όπου L είναι ένας τριγωνικός κάτω πίνακας και U ένας τριγωνικός άνω πίνακας. Τα διαγώνια στοιχεία τόσο του L όσο και του U είναι μη-μηδενικά. Αν $z \in \mathbb{R}^n$ είναι η λύση του συστήματος

$$Lz = b$$

και $y \in \mathbb{R}^n$ η λύση του

$$Uy = z$$

τότε το y αποτελεί και τη λύση του (5.2.22). Πράγματι:

$$Ay = Luy = Lz = b$$

Λόγω του ότι οι πίνακες L, U είναι τριγωνικοί με μη-μηδενικά διαγώνια στοιχεία, τα συστήματα

$$Lz = b, \quad Uy = z$$

λύνονται εξαιρετικά εύκολα. Γράφοντας το $Lz = b$ ως σύστημα γραμμικών εξισώσεων προκύπτει:

$$\begin{aligned} l_{11}z_1 &= b_1 \\ l_{21}z_1 + l_{22}z_2 &= b_2 \\ &\vdots \\ l_{n1}z_1 + l_{n2}z_2 + \dots + l_{nn}z_n &= b_n \end{aligned}$$

Προφανώς, από την πρώτη εξίσωση βρίσκουμε απ' ευθείας το z_1 ως $z_1 = \frac{b_1}{l_{11}}$. Αντικαθιστώντας την τιμή του z_1 στη δεύτερη εξίσωση προσδιορίζουμε το z_2 . Αντικαθιστώντας τα z_1, z_2 στην τρίτη υπολογίζουμε το z_3 κ.ο.κ. Αυτή η διαδικασία επίλυσης ενός συστήματος αναφέρεται ως πρόσω-αντικατάσταση.

Έχοντας υπολογίσει το z , προχωράμε στην λύση του $Uy = z$, το οποίο σε μορφή εξισώσεων γράφεται:

$$\begin{aligned} u_{11}y_1 + u_{12}y_2 + \dots + u_{1n}y_n &= z_1 \\ u_{22}y_2 + \dots + u_{2n}y_n &= z_2 \\ &\vdots \\ u_{nn}y_n &= z_n \end{aligned}$$

Το σύστημα αυτό επιλύεται ακολουθώντας αντίστροφη πορεία από το $Lz = b$, ξεκινώντας τη λύση από την τελευταία εξίσωση και ανεβαίνοντας προς τα πάνω. Αυτή η διαδικασία επίλυσης αναφέρεται ως ανάστροφη-αντικατάσταση. Το γεγονός ότι τα διαγώνια στοιχεία των L, U είναι διάφορα του μηδενός, εξασφαλίζει τη βιωσιμότητα της μεθόδου.

Στα παραπάνω υποθέσαμε ότι ο A μπορεί να παραγοντοποιηθεί ως $A = LU$. Μπορεί ναδειχθεί ότι αν ο A είναι θετικά ορισμένος, είναι δυνατόν να παραγοντοποιηθεί κατά μοναδικό τρόπο στη μορφή:

$$A = \hat{L}D\hat{L}^T \quad (5.2.23)$$

όπου ο \hat{L} είναι ένας κάτω τριγωνικός πίνακας με διαγώνια στοιχεία ίσα με 1 και D ένας διαγώνιος πίνακας με θετικά στοιχεία. Αν $D^{1/2}$ είναι η τετραγωνική ρίζα του D , είναι δηλαδή ένας διαγώνιος πίνακας τα στοιχεία του οποίου είναι η τετραγωνική ρίζα των αντίστοιχων στοιχείων του D , τότε, $D = D^{1/2}D^{1/2}$ και η (5.2.23) γίνεται

$$A = (\hat{L}D^{1/2})(D^{1/2}\hat{L}^T) \quad (5.2.24)$$

Αν ορίσουμε $L = \hat{L}D^{1/2}$ τότε ο L είναι ένας τριγωνικός κάτω πίνακας με θετικά διαγώνια στοιχεία. Ταυτόχρονα, ο $L^T = D^{1/2}\hat{L}^T$ είναι ο αντίστοιχος τριγωνικός άνω πίνακας. Η (5.2.24) γίνεται

$$A = LL^T \quad (5.2.25)$$

Η παραγοντοποίηση (5.2.25) του θετικά ορισμένου πίνακα A ονομάζεται παραγοντοποίηση *Cholesky* του A .

Ο πίνακας L στην παραγοντοποίηση *Cholesky* του A υπολογίζεται απ' ευθείας εξισώνοντας τα αντίστοιχα στοιχεία της εξίσωσης $A = LL^T$.

$$\begin{bmatrix} \alpha_{11} & \alpha_{12} & \dots & \alpha_{1n} \\ \alpha_{21} & \alpha_{22} & \dots & \alpha_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_{n1} & \alpha_{n2} & \dots & \alpha_{nn} \end{bmatrix} = \begin{bmatrix} l_{11} & 0 & \dots & 0 \\ l_{21} & l_{22} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ l_{n1} & l_{n2} & \dots & l_{nn} \end{bmatrix} \begin{bmatrix} l_{11} & l_{12} & \dots & l_{1n} \\ 0 & l_{22} & \dots & l_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & l_{nn} \end{bmatrix}$$

Η επίλυση ξεκινά από την πρώτη στήλη του A παρατηρώντας:

$$\alpha_{11} = l_{11}^2, \alpha_{21} = l_{11}l_{21}, \dots, \alpha_{n1} = l_{11}l_{n1}$$

Οι παραπάνω εξισώσεις μας επιτρέπουν τον προσδιορισμό των στοιχείων της πρώτης στήλης του L μια και γνωρίζουμε ότι $l_{11} \neq 0$. Στη συνέχεια περνάμε στον υπολογισμό των στοιχείων της δεύτερης γραμμής του L . Τώρα $\alpha_{22} = l_{21}^2 + l_{22}^2$, απ' όπου βρίσκουμε το l_{22} μια και έχουμε ήδη υπολογίσει το l_{21} . Εξισώνοντας και τα υπόλοιπα στοιχεία της δεύτερης στήλης έχουμε:

$$\alpha_{23} = l_{31}l_{21} + l_{32}l_{22}, \dots, \alpha_{2n} = l_{n1}l_{21} + l_{n2}l_{22}$$

Λύνοντας ως προς τα άγνωστα στοιχεία l_{32}, \dots, l_{n2} προσδιορίζουμε πλήρως τα στοιχεία της δεύτερης στήλης του L . Προφανώς, η διαδικασία επαναλαμβάνεται μέχρι τον πλήρη προσδιορισμό του L .

Έχειδειχθεί ότι η χρήση της παραγοντοποίησης *Cholesky* στην επίλυση γραμμικών συστημάτων δίνει αποτελέσματα που επηρεάζονται λιγότερο από τα σφάλματα στρογγυλοποίησης σε σχέση με τη μέθοδο *Gauss*.

Κριτήρια καλής λειτουργίας

Σε προηγούμενη υποενότητα είδαμε ότι αν και η μέθοδος της μέγιστης καθόδου καλύπτεται από ισχυρές εγγυήσεις σύγκλισης, τα υπολογιστικά της μειονεκτήματα είναι τόσο έντονα ώστε να μας οδηγούν σε τροποποιήσεις, εις βάρος κατ' ουσίαν της μαθηματικής αυστηρότητας.

Ένα σοβαρό μειονέκτημα σχετίζεται με τον υπολογισμό του βήματος γ_k . Θυμηθείτε ότι η σταθερά γ_k στην αναδρομική σχέση της μεθόδου της μέγιστης καθόδου:

$$x_{k+1} = x_k - \gamma_k \nabla f(x_k)$$

προσδιορίζεται ελαχιστοποιώντας την

$$\phi_k(\gamma) = f(x_k - \gamma \nabla f(x_k)), \quad \forall \gamma \geq 0$$

Η εν λόγω ελαχιστοποίηση όμως απαιτεί σημαντικό υπολογιστικό χρόνο για να πραγματοποιηθεί και μην ξεχνάτε ότι αυτό δεν είναι παρά ένα μόνο βήμα σ' ένα πραγματικό n -διάστατο πρόβλημα.

Ένα δεύτερο μειονέκτημα της μεθόδου είναι η κίνησή της σε κάθετα βήματα, που καθιστούν τη σύγκλισή της αργή και κατά συνέπεια τη μέθοδο υπολογιστικά ανεπαρκή.

Η εναλλακτική πρόταση της χρήσης της μεθόδου *Newton* εμφανίζει προβλήματα που σχετίζονται με τον $\nabla^2 f(x)$. Συνήθως στην πράξη δεν μπορούμε να εγγυηθούμε ότι ο $\nabla^2 f(x)$ είναι θετικά ορισμένος.

Μια προσέγγιση στην ανάπτυξη πρακτικών αναδρομικών μεθόδων για προβλήματα ελαχιστοποίησης συναρτήσεων $f(x)$ χωρίς περιορισμούς, που να είναι ελεύθερες των προβλημάτων που παρουσιάζονται τόσο στη μέθοδο της μέγιστης καθόδου, όσο και στη μέθοδο *Newton*, είναι να θέσουμε κάποια κριτήρια καλής λειτουργίας και να επιδιώκουμε η επαναληπτική μέθοδος που σχεδιάζουμε να τα ικανοποιεί.

Για δοσμένη συνάρτηση $f(x)$ με συνεχείς μερικές παραγώγους πρώτης τάξης στο \mathbb{R}^n και δοσμένο σημείο εκκίνησης x_0 , θέλουμε να σχεδιάσουμε μια μέθοδο που να παράγει μια ακολουθία $\{x_k\}$ που προσδιορίζεται από την αναδρομική σχέση:

$$x_{k+1} = x_k + \gamma_k d_k, \quad \gamma_k > 0$$

Οι ποσότητες που υπόκεινται σε σχεδιασμό είναι το βήμα γ_k και το διάνυσμα d_k στην κατεύθυνση του οποίου αναζητούμε το ελάχιστο. Για κάθε k λοιπόν, απαιτούμε από τα γ_k, d_k να ικανοποιούν τα παρακάτω κριτήρια:

Κριτήριο 1 $f(x_{k+1}) < f(x_k)$, οποτεδήποτε $\nabla f(x_k) \neq 0$.

Κριτήριο 2 $d_k^T \nabla f(x_k) < 0$.

Πριν προχωρήσουμε, καλό είναι να σημειωθεί ότι τα κριτήρια 1, 2 έχουν ήδη αναφερθεί σε προηγούμενες υποενότητες και για το λόγο αυτό δεν θα τα αναλύσουμε περαιτέρω.

Τα επόμενα δύο κριτήρια σχετίζονται με το μέγεθος του βήματος γ_k . Αν το γ_k επιλεγεί αρκετά μικρό και θετικό, τότε είναι μάλλον απίθανο τα x_k να κινηθούν γρήγορα προς το ελάχιστο x^* , οδηγώντας τη μέθοδο σε μεγάλο αριθμό επαναλήψεων. Το κριτήριο που ακολουθεί διασφαλίζει ότι το γ_k δεν θα είναι πολύ μικρό.

Κριτήριο 3 Υπάρχει β με $0 < \beta < 1$ τέτοιο ώστε

$$d_k^T \nabla f(x_{k+1}) > \beta d_k^T \nabla f(x_k)$$

Ας δούμε τώρα πως η συνύπαρξη των Κριτηρίων 2 και 3 εμποδίζει το γ_k από το να πάρει πολύ μικρές τιμές. Έστω ότι $0 < \beta < 1$ και ότι επιλέγουμε το γ_k έτσι ώστε να ικανοποιείται το Κριτήριο 3. Επειδή $\beta < 1$ και $d_k^T \nabla f(x_k) < 0$ (από το Κριτήριο 2) έχουμε:

$$d_k^T \nabla f(x_{k+1}) = d_k^T \nabla f(x_k + \gamma_k d_k) > \beta d_k^T \nabla f(x_k) > d_k^T \nabla f(x_k) \quad (5.2.26)$$

Αφαιρώντας την ποσότητα $d_k^T \nabla f(x_k)$ και στα δύο μέρη της (5.2.26) βρίσκουμε:

$$d_k^T \nabla f(x_k + \gamma_k d_k) - d_k^T \nabla f(x_k) > (\beta - 1) d_k^T \nabla f(x_k) > 0 \quad (5.2.27)$$

Η (5.2.27) μας λέει ότι το γ_k δε μπορεί να επιλεγεί οσοδήποτε μικρό. Πράγματι, αν θεωρήσουμε ότι το γ_k παίρνει τιμές κοντά στο μηδέν, το αριστερό μέρος της (5.2.27)

θα τείνει στο μηδέν $\lim_{\gamma_k \rightarrow 0} [d_k^T \nabla f(x_k + \gamma_k d_k) - d_k^T \nabla f(x_k)] = 0$ ενώ ταυτόχρονα, το δεξί μέρος παραμένει σταθερό (ανεξάρτητο του γ_k), το οποίο είναι άτοπο.

Αλλά το γ_k δεν θα πρέπει να επιλεγεί και πολύ μεγάλο. Κάτι τέτοιο αναμένεται μεν να οδηγήσει σε μεγάλο βήμα από το x_k στο x_{k+1} , χωρίς όμως να αποτελεί πάντα και την ενδεδειγμένη κίνηση προς την ελαχιστοποίηση της $f(x)$. Είδαμε ότι οι μέθοδοι κλίσης μπορεί να οδηγήσουν σε ταλαντώσεις ή ακόμα και να αποκλίνουν όταν το βήμα γ_k είναι ακούρντως μεγάλο. Το επόμενο κριτήριο έρχεται να διασφαλίσει ότι το γ_k δεν θα πάρει πολύ μεγάλες τιμές.

Κριτήριο 4 Υπάρχει α με $0 < \alpha < \beta < 1$ τέτοιο ώστε

$$f(x_{k+1}) \leq f(x_k) + \alpha \gamma_k d_k^T \nabla f(x_k)$$

Για να κατανοήσουμε τη λειτουργία του Κριτηρίου 4, θα το γράψουμε στη μορφή:

$$\frac{f(x_k) - f(x_{k+1})}{\gamma_k} \geq \alpha [-d_k^T \nabla f(x_k)] \quad (5.2.28)$$

Το αριστερό μέρος της (5.2.28) παριστάνει τη σχετική μείωση της τιμής της $f(x)$ ως προς την αύξηση στις τιμές του γ μεταξύ των σημείων x_k και x_{k+1} . Ο όρος $-d_k^T \nabla f(x_k)$ που είναι θετικός λόγω του Κριτηρίου 2, αποτελεί πολλαπλάσιο του μέτρου του ρυθμού μείωσης της $f(x)$, στην κατεύθυνση του d_k , στο σημείο x_k . Κατά συνέπεια, το Κριτήριο 4 μας λέει ότι η σχετική μείωση στις τιμές της $f(x)$ ως προς το μέγεθος του γ_k , θα πρέπει να ξεπερνά ένα προαποφασισμένο ποσοστό του μέτρου του ρυθμού μείωσης της $f(x)$ στο x_k στην κατεύθυνση προς το x_{k+1} . Μια ελαφρά διαφορετική ερμηνεία του Κριτηρίου 4 δίνεται αν παρατηρήσουμε ότι:

$$\frac{f(x_k) - f(x_{k+1})}{|x_k - x_{k+1}|} = \frac{f(x_k) - f(x_{k+1})}{\gamma_k |d_k|} \geq M > 0 \quad (5.2.29)$$

όπου

$$M = -\frac{\alpha d_k^T \nabla f(x_k)}{|d_k|}$$

Από την (5.2.29) προκύπτει:

$$f(x_k) - f(x_{k+1}) \geq M |x_k - x_{k+1}| \quad (5.2.30)$$

Συνεπώς, αν το βήμα από το x_k στο x_{k+1} είναι μεγάλο, τότε και η μείωση της τιμής της f θα είναι επίσης μεγάλη με σταθερά αναλογίας M . Από την (5.2.30)

Μπορεί ναδειχθεί ότι το $\bar{\mu}_k$ ισούται με την απόλυτη τιμή της μεγαλύτερης ιδιοτιμής του $\nabla^2 f(x_k)$. Αν επιλέξουμε $\mu_k > \bar{\mu}_k$ και $x \neq 0$ έχουμε:

$$x^T \nabla^2 f(x_k) x + \mu_k |x|^2 = |x|^2 \left[\frac{x^T}{|x|} \nabla^2 f(x_k) \frac{x}{|x|} + \mu_k \right] > |x|^2 (-\bar{\mu}_k + \mu_k) > 0 \quad (5.2.38)$$

μια και ο $\nabla^2 f(x_k)$ δεν είναι θετικά ορισμένος. Από την (5.2.38) προκύπτει ότι ο $\nabla^2 f(x_k) + \mu_k I$ είναι για $\mu_k > \bar{\mu}_k$ θετικά ορισμένος.

Τροποποιημένος αλγόριθμος Newton (μέθοδος Levenberg – Marquardt)

Αν ο $\nabla^2 f(x_k)$ δεν είναι θετικά ορισμένος $\forall x \in \mathbb{R}^n$ τότε μπορούμε να χρησιμοποιήσουμε τον παρακάτω τροποποιημένο αλγόριθμο *Newton* που ακολουθεί τα παρακάτω βήματα.

Αλγόριθμος 5.2.3 (Μέθοδος Levenberg – Marquardt)

Βήμα 1 Δοσμένου x_0 :

Βήμα 2 Υπολογίστε το μ_k έτσι ώστε ο $\nabla^2 f(x_k) + \mu_k I$ να είναι θετικά ορισμένος.

Βήμα 3 Λύστε το σύστημα $[\nabla^2 f(x_k) + \mu_k I] d_k = -\nabla f(x_k)$.

Βήμα 4 Υπολογίστε το γ_k ώστε να ικανοποιούνται τα Κριτήρια 3, 4 του Θεωρήματος 5.2.6.

Βήμα 5 Θέστε $x_{k+1} = x_k + \gamma_k d_k$.

Βήμα 6 Πηγαίνετε στο Βήμα 1.

Παρατηρήστε πως όταν το μ_k είναι αρκετά μεγάλο, ο παράγοντας $\mu_k I$ κυριαρχεί σε σχέση με τον $\nabla^2 f(x_k)$. Κατά συνέπεια η μέθοδος *Levenberg – Marquardt* θα λειτουργεί σχεδόν όπως η μέθοδος της μέγιστης καθόδου. Αν όμως το μ_k είναι μικρό, τότε ο $\nabla^2 f(x_k)$ υπερισχύει και η μέθοδος συμπεριφέρεται σαν τη μέθοδο *Newton*.

Επιλογή βήματος (συνέχεια)

Το πρόβλημα της επιλογής βήματος γ_k σε κάθε επανάληψη k του αναδρομικού αλγορίθμου για την ελαχιστοποίηση της συνάρτησης $f(x)$ μας έχει ήδη απασχολήσει. Σε προηγούμενες ενότητες είδαμε τόσο μεθόδους που στηρίζονται στην ελαχιστοποίηση της $f(x_k + \gamma_k d_k)$ ως προς $\gamma_k > 0$, όσο και τεχνικές που προσπαθούν να ικανοποιήσουν κάποια κριτήρια καλής λειτουργίας. Στην παρούσα ενότητα θα ασχοληθούμε περαιτέρω με το πρόβλημα, δίνοντας κι άλλους συχνά χρησιμοποιούμενους αλγορίθμους.

Ο κανόνας Armijo

Ο κανόνας Armijo αποτελεί μια μέθοδο διαδοχικής μείωσης του γ_k που αποσκοπεί στην ικανοποίηση του Κριτηρίου 4. Σύμφωνα με τη μέθοδο, το βήμα επιλέγεται ως:

$$\gamma_k = s\beta^{m_k}, \quad 0 < \beta < 1$$

και m_k να είναι ο μικρότερος μη-αρνητικός ακέραιος που ικανοποιεί το Κριτήριο 4 δηλαδή:

$$f(x_k) - f(x_{k+1}) \geq -\alpha\beta^{m_k} s d_k^T \nabla f(x_k) \quad (5.2.39)$$

Μ' άλλα λόγια, δοκιμάζουμε διαδοχικές τιμές του m_k με προεπιλεγμένα α, β, s , μέχρι την ικανοποίηση της (5.2.39). Συνήθως $\alpha \in [10^{-5}, 10^{-1}]$ και $\beta \in [\frac{1}{10}, \frac{1}{2}]$. Ο συντελεστής s δηλώνει το αρχικό βήμα. Στο Σχ. 5.8 δίνεται γραφικά ο κανόνας.

Σταθερό βήμα

Στη μέθοδο αυτή το βήμα διατηρείται σταθερό σ' όλη τη διάρκεια της αναζήτησης του ελαχίστου. Είναι η απλούστερη δυνατή μέθοδος που ελαχιστοποιεί κατ' ουσίαν την υπολογιστική πολυπλοκότητα του αλγορίθμου ελαχιστοποίησης της $f(x)$. Βέβαια, αν το βήμα επιλεγεί πολύ μεγάλο μπορεί να παρατηρηθεί μέχρι και απόκλιση της διαδικασίας αναζήτησης, ενώ αν επιλεγεί πολύ μικρό, ο ρυθμός σύγκλισης μικραίνει, αυξάνοντας σημαντικά το πλήθος των επαναλήψεων. Η μέθοδος αν και απλή, προϋποθέτει μεγάλη εμπειρία για την ορθή εφαρμογή της.