Μέθοδοι πολυδιάστατης ελαχιστοποίησης

Δ. Γ. Παπαγεωργίου
Τμήμα Μηχανικών Επιστήμης Υλικών
Πανεπιστήμιο Ιωαννίνων

dpapageo@uoi.gr
http://pc164.materials.uoi.gr/dpapageo

Μέθοδοι πολυδιάστατης ελαχιστοποίησης με παραγώγους

- Οι μέθοδοι πολυδιάστατης ελαχιστοποίησης είναι επαναληπτικές.
- Ξεκινούν από ένα αρχικό σημείο $\boldsymbol{x}^{(0)}$ και κατασκευάζουν μια σειρά σημείων $\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}, \boldsymbol{x}^{(3)}, \cdots \boldsymbol{x}^{(k)}$ τέτοια ώστε:
$$f(\boldsymbol{x}^{(0)}) > f(\boldsymbol{x}^{(1)}) > f(\boldsymbol{x}^{(2)}) > \cdots > f(\boldsymbol{x}^{(k)})$$
- Η σειρά των σημείων συγκλίνει στο ελάχιστο $\boldsymbol{x}^*$

Βασικός επαναληπτικός αλγόριθμος

1. Δίνεται το αρχικό σημείο $\boldsymbol{x}^{(0)}$
2. Επανάληψη για $k = 0,1,2,\cdots$
   a. Ελέγχουμε τα κριτήρια τερματισμού
   b. Προσδιορίζουμε ένα βήμα $\boldsymbol{h}^{(k)}$ τέτοιο ώστε:
   $$f(\boldsymbol{x}^{(k)} + \boldsymbol{h}^{(k)}) < f(\boldsymbol{x}^{(k)})$$
   c. Θέτουμε $\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} + \boldsymbol{h}^{(k)}$

Πως προσδιορίζουμε το βήμα $h^{(k)}$

- Μέθοδοι γραμμικής αναζήτησης
  Εύρεση ενός σημείου με "επαρκή μείωση" της τιμής της συνάρτησης πάνω σε μια ευθεία που προσδιορίζεται από ένα διάνυσμα $\boldsymbol{s}$.
- Μέθοδοι διαστημάτων εμπιστοσύνης
  Εύρεση του ελαχίστου του τετραγωνικού μοντέλου μέσα σε μια "υπερσφαίρα" ακτίνας Δ, μέσα στην οποία το τετραγωνικό μοντέλο προσεγγίζει ικανοποιητικά τη συνάρτηση.

Κριτήρια τερματισμού

Ο αλγόριθμος τερματίζεται όταν ικανοποιείται ένα από τα παρακάτω κριτήρια:

$$\text{relgrad} = \frac{\text{ρυθμός αλλαγής στην } f(x)}{\text{ρυθμός αλλαγής στο } x} = \frac{\nabla f(\boldsymbol{x})\boldsymbol{x}}{f(\boldsymbol{x})} < \varepsilon_g$$
Η παράγωγος είναι πολύ μικρή

$$\frac{\|\boldsymbol{x}^{(k+1)} - \boldsymbol{x}^{(k)}\|}{\|\boldsymbol{x}^{(k)}\|} < \epsilon_x$$
Οι μεταβλητές συγκλίνουν αργά

$$f(\boldsymbol{x}^{(k)}) - f(\boldsymbol{x}^{(k+1)}) < \varepsilon_f$$
Η τιμή της συνάρτησης μειώνεται αργά

Πλήθος υπολογισμών της συνάρτησης $> n_{max}$
Περιορισμός στο χρόνο που θα δαπανηθεί

Φθίνουσες διευθύνσεις

Οι μέθοδοι ελαχιστοποίησης που χρησιμοποιούν γραμμική αναζήτηση παράγουν φθίνουσες διευθύνσεις, δηλαδή διευθύνσεις $\boldsymbol{s}$ για τις οποίες ισχύει:

$$f(\boldsymbol{x} + a\boldsymbol{s}) < f(\boldsymbol{x}) \qquad a > 0$$

Με άλλα λόγια η τιμή της συνάρτησης μειώνεται κατά μήκος μιας φθίνουσας διεύθυνσης $\boldsymbol{s}$

Πως βρίσκουμε αν μια διεύθυνση s είναι φθίνουσα;

Χρησιμοποιούμε το ανάπτυγμα Taylor:

$$f(\boldsymbol{x} + \boldsymbol{h}) = f(\boldsymbol{x}) + \boldsymbol{g}^T\boldsymbol{h} + \frac{1}{2}\boldsymbol{h}^T\boldsymbol{G}(\boldsymbol{x})\boldsymbol{h} + \cdots$$

$$f(\boldsymbol{x} + a\boldsymbol{s}) \approx f(\boldsymbol{x}) + \boldsymbol{g}^T a\boldsymbol{s} \ \Rightarrow \ f(\boldsymbol{x} + a\boldsymbol{s}) - f(\boldsymbol{x}) = a\boldsymbol{g}^T\boldsymbol{s}$$

Επειδή $f(\boldsymbol{x} + a\boldsymbol{s}) < f(\boldsymbol{x})$ θα πρέπει $f(\boldsymbol{x} + a\boldsymbol{s}) - f(\boldsymbol{x}) < 0$

Άρα $a\boldsymbol{g}^T\boldsymbol{s} < 0$ και επειδή $a > 0$

$$\boxed{\boldsymbol{g}^T\boldsymbol{s} < 0}$$

Φθίνουσες διευθύνσεις – Παράδειγμα

Δίνεται η συνάρτηση $f(x, y) = x^2 + y^2$ και το σημείο $\boldsymbol{x} = (1,1)$.

Βρείτε αν η διεύθυνση $\boldsymbol{s} = (1, -2)$ είναι φθίνουσα .

Οι πρώτες παράγωγοι είναι:

$$\frac{\partial f}{\partial x} = 2x$$

$$\frac{\partial f}{\partial y} = 2y$$

Στο σημείο $\boldsymbol{x} = (1,1)$ το διάνυσμα των πρώτων παραγώγων είναι:

$$\boldsymbol{g} = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$$

$$\boldsymbol{g}^T\boldsymbol{s} = [2 \quad 2]\begin{bmatrix} 1 \\ -2 \end{bmatrix} = -2 < 0$$ Φθίνουσα διεύθυνση

Η διεύθυνση της πιο απότομης καθόδου

Από όλες τις διευθύνσεις s που είναι φθίνουσες και ικανοποιούν το κριτήριο

$$\boldsymbol{g}^T\boldsymbol{s} < 0$$

ποια είναι αυτή που δίνει τη μεγαλύτερη μείωση στην τιμή της συνάρτησης ;

Η μείωση στην τιμή της συνάρτησης είναι:

$f(\boldsymbol{x}) - f(\boldsymbol{x} + \alpha\boldsymbol{s}) =$

$f(\boldsymbol{x}) - [f(\boldsymbol{x}) + \alpha\boldsymbol{s}^T\boldsymbol{g}] =$ Χρησιμοποιούμε ανάπτυγμα Taylor

$-\alpha\boldsymbol{s}^T\boldsymbol{g} =$

$-\alpha\|\boldsymbol{s}\|\|\boldsymbol{g}\|\cos\theta$ θ είναι η γωνία μεταξύ $\boldsymbol{s}$ και $\boldsymbol{g}$

Η μείωση γίνεται μέγιστη όταν $\cos\theta = -1$ δηλαδή όταν $\theta = \pi$

Αυτό σημαίνει ότι τα $\boldsymbol{s}$ και $\boldsymbol{g}$ είναι συγγραμμικά και αντίθετης φοράς, δηλαδή

$$\boxed{\boldsymbol{s} = -\boldsymbol{g}}$$ Διεύθυνση της πιο απότομης καθόδου

Η διεύθυνση αυτή είναι φθίνουσα αφού:

$$\boldsymbol{s}^T\boldsymbol{g} = -\boldsymbol{g}^T\boldsymbol{g} = -\|\boldsymbol{g}\|^2 < 0$$

Μέθοδοι με γραμμική αναζήτηση

Μέθοδοι με γραμμική αναζήτηση

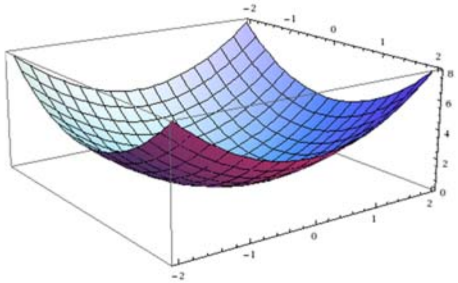
1. Δίνεται αρχικό σημείο $\boldsymbol{x}^{(0)}$
2. Επανάληψη για $k = 0,1,2, \cdots$
   a. Έλεγχος των κριτηρίων τερματισμού
   b. Προσδιορίζεται μια φθίνουσα διεύθυνση $\boldsymbol{s}^{(k)}$
   c. Γραμμική αναζήτηση στη διεύθυνση $\boldsymbol{s}^{(k)}$, δηλαδή ελαχιστοποίηση ως προς a της συνάρτησης
   $$\varphi(\alpha) = f\left(\boldsymbol{x}^{(k)} + a\boldsymbol{s}^{(k)}\right)$$
   d. Θέτουμε $\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} + a^*\boldsymbol{s}^{(k)}$

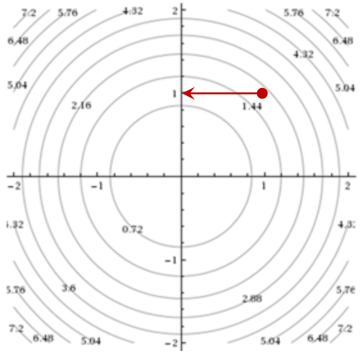
Παρατηρήσεις:

- Στο βήμα 2c πρέπει να βρούμε το ελάχιστο μιας μονοδιάστατης συνάρτησης
- Διαφορετικοί τρόποι επιλογής της διεύθυνσης $\boldsymbol{s}^{(k)}$ δίνουν διαφορετικούς αλγορίθμους ελαχιστοποίησης

Γραμμική αναζήτηση – Παράδειγμα #1

Θεωρείστε τη συνάρτηση $f(x,y) = x^2 + y^2$ και το αρχικό σημείο $\boldsymbol{x} = (1,1)$.
Βρείτε ένα νέο σημείο $\tilde{\boldsymbol{x}}$ με χαμηλότερη τιμή συνάρτησης πραγματοποιώντας γραμμική αναζήτηση στη διεύθυνση $\boldsymbol{s} = (-1,0)$.





Γραμμική αναζήτηση – Παράδειγμα #1

Θα εξετάσουμε πρώτα αν η διεύθυνση $\boldsymbol{s} = (-1,0)$ είναι φθίνουσα.

Οι πρώτες παράγωγοι είναι:

$$\frac{\partial f}{\partial x} = 2x$$
$$\frac{\partial f}{\partial y} = 2y$$

Στο σημείο $\boldsymbol{x} = (1,1)$ το διάνυσμα των πρώτων παραγώγων είναι:

$$\boldsymbol{g} = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$$

Εξετάζουμε τη συνθήκη:

$$\boldsymbol{g}^T \boldsymbol{s} = \begin{bmatrix} 2 & 2 \end{bmatrix} \begin{bmatrix} -1 \\ 0 \end{bmatrix} = -2 < 0$$

Άρα η $\boldsymbol{s}$ είναι φθίνουσα διεύθυνση.

Κατασκευάζουμε τη συνάρτηση $\varphi(\alpha)$

$$\varphi(\alpha) = f(\boldsymbol{x} + \alpha\boldsymbol{s}) =$$
$$f\left(\begin{bmatrix} 1 \\ 1 \end{bmatrix} + \alpha \begin{bmatrix} -1 \\ 0 \end{bmatrix}\right) =$$
$$f\left(\begin{bmatrix} 1-\alpha \\ 1 \end{bmatrix}\right) \Rightarrow$$
$$\varphi(\alpha) = (1-\alpha)^2 + 1$$

Βρίσκουμε που έχει ελάχιστο ως προς α

$$\varphi'(\alpha) = -2(1-\alpha)$$
$$\varphi'(\alpha) = 0 \Rightarrow$$
$$-2(1-\alpha) = 0 \Rightarrow$$
$$\alpha = 1$$

Γραμμική αναζήτηση – Παράδειγμα #1

Το αρχικό σημείο είναι $\boldsymbol{x} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$

Η τιμή της συνάρτησης στο αρχικό σημείο είναι:

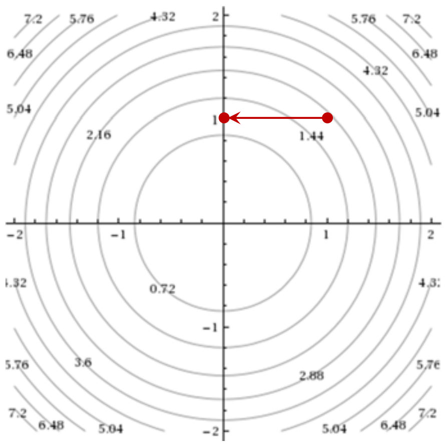
$$f(1,1) = 1^2 + 1^2 = 2$$

Το νέο σημείο είναι:

$$\tilde{\boldsymbol{x}} = \boldsymbol{x} + \alpha\boldsymbol{s} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} + 1 \cdot \begin{bmatrix} -1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

Η νέα τιμή της συνάρτησης είναι:

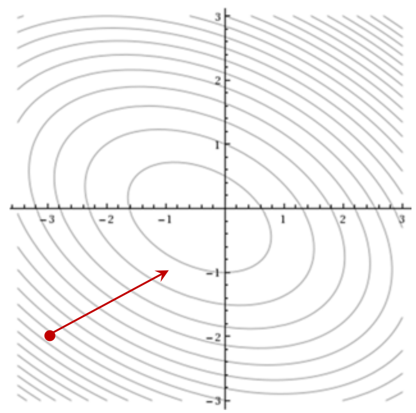
$$f(0,1) = 0^2 + 1^1 = 1$$



Γραμμική αναζήτηση – Παράδειγμα #2

Δίνεται η συνάρτηση $f(x,y) = x^2 + 2y^2 + xy + x + y + 1$ και το αρχικό σημείο $\boldsymbol{x} = (-3,-2)$.

Βρείτε ένα νέο σημείο $\tilde{\boldsymbol{x}}$ με χαμηλότερη τιμή συνάρτησης πραγματοποιώντας γραμμική αναζήτηση στη διεύθυνση $\boldsymbol{s} = (2,1)$.



Γραμμική αναζήτηση – Παράδειγμα #2

Θα εξετάσουμε πρώτα αν η διεύθυνση $\boldsymbol{s} = (2,1)$ είναι φθίνουσα, δηλαδή θα εξετάσουμε τη συνθήκη $\boldsymbol{g}^T\boldsymbol{s} < 0$

$f(x,y) = x^2 + 2y^2 + xy + x + y + 1$

Οι πρώτες παράγωγοι είναι:

$$\frac{\partial f}{\partial x} = 2x + y + 1$$
$$\frac{\partial f}{\partial y} = 4y + x + 1$$

Στο σημείο $\boldsymbol{x} = (-3,-2)$ το διάνυσμα των πρώτων παραγώγων είναι:

$$\boldsymbol{g} = \begin{bmatrix} -7 \\ -10 \end{bmatrix}$$

Εξετάζουμε τη συνθήκη:

$$\boldsymbol{g}^T\boldsymbol{s} = [-7 \quad -10]\begin{bmatrix} 2 \\ 1 \end{bmatrix} = -24 < 0$$

Άρα η $\boldsymbol{s}$ είναι φθίνουσα διεύθυνση.

Κατασκευάζουμε τη συνάρτηση $\varphi(\alpha)$

$$\varphi(\alpha) = f(\boldsymbol{x} + a\boldsymbol{s}) =$$
$$f\left(\begin{bmatrix} -3 \\ -2 \end{bmatrix} + a\begin{bmatrix} 2 \\ 1 \end{bmatrix}\right) =$$
$$f\left(\begin{bmatrix} 2a-3 \\ a-2 \end{bmatrix}\right) \Rightarrow$$
$$\varphi(\alpha) = (2a-3)^2 + 2(a-2)^2 + (2a-3)(a-2) + (2a-3) + (a-2) + 1$$
$$= 8a^2 - 24a + 19$$

Βρίσκουμε που έχει ελάχιστο ως προς a

$$\varphi'(\alpha) = 16a - 24$$
$$\varphi'(\alpha) = 0 \Rightarrow$$
$$16a - 24 = 0 \ \Rightarrow \ \alpha = \frac{3}{2}$$

Γραμμική αναζήτηση – Παράδειγμα #2

$f(x,y) = x^2 + 2y^2 + xy + x + y + 1$

Το αρχικό σημείο είναι $\boldsymbol{x} = \begin{bmatrix} -3 \\ -2 \end{bmatrix}$

Η τιμή της συνάρτησης στο αρχικό σημείο είναι:

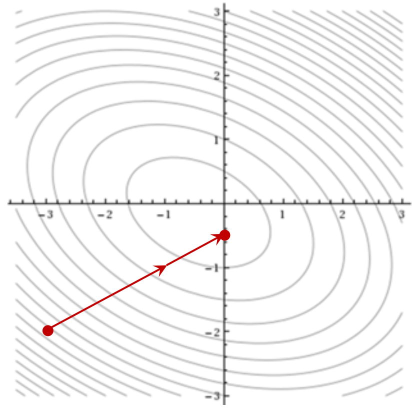
$$f(-3,-2) = (-3)^2 + 2(-2)^2 + (-3)(-2) - 3 - 2 + 1 = 19$$

Το νέο σημείο είναι:

$$\tilde{\boldsymbol{x}} = \boldsymbol{x} + a\boldsymbol{s} = \begin{bmatrix} -3 \\ -2 \end{bmatrix} + \frac{3}{2}\cdot\begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ -\frac{1}{2} \end{bmatrix}$$

Η νέα τιμή της συνάρτησης είναι:

$$f\left(0,-\frac{1}{2}\right) = 0^2 + 2\left(-\frac{1}{2}\right)^2 - 0\frac{1}{2}$$
$$+0 - \frac{1}{2} + 1 = 1$$



Το ελάχιστο μιας τετραγωνικής συνάρτησης στη διεύθυνση s

Δίνεται η τετραγωνική συνάρτηση $q(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}^T\boldsymbol{x} + c$ και μια φθίνουσα διεύθυνση $\boldsymbol{s}$. Ποιό είναι το ελάχιστο της $q(\boldsymbol{x})$ κατά μήκος της διεύθυνσης $\boldsymbol{s}$;

Ορίζουμε τη συνάρτηση:

$$\varphi(\lambda) = q(\boldsymbol{x} + \lambda\boldsymbol{s}) =$$
$$\frac{1}{2}(\boldsymbol{x} + \lambda\boldsymbol{s})^T\boldsymbol{A}(\boldsymbol{x} + \lambda\boldsymbol{s}) + \boldsymbol{b}^T(\boldsymbol{x} + \lambda\boldsymbol{s}) + c =$$
$$\frac{1}{2}(\boldsymbol{x} + \lambda\boldsymbol{s})^T(\boldsymbol{A}\boldsymbol{x} + \lambda\boldsymbol{A}\boldsymbol{s}) + \boldsymbol{b}^T\boldsymbol{x} + \lambda\boldsymbol{b}^T\boldsymbol{s} + c =$$
$$\frac{1}{2}(\boldsymbol{x}^T + \lambda\boldsymbol{s}^T)(\boldsymbol{A}\boldsymbol{x} + \lambda\boldsymbol{A}\boldsymbol{s}) + \boldsymbol{b}^T\boldsymbol{x} + \lambda\boldsymbol{b}^T\boldsymbol{s} + c =$$
$$\frac{1}{2}(\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x} + \lambda\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{s} + \lambda\boldsymbol{s}^T\boldsymbol{A}\boldsymbol{x} + \lambda^2\boldsymbol{s}^T\boldsymbol{A}\boldsymbol{s}) + \boldsymbol{b}^T\boldsymbol{x} + \lambda\boldsymbol{b}^T\boldsymbol{s} + c =$$
$$\frac{1}{2}(\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x} + 2\lambda\boldsymbol{s}^T\boldsymbol{A}\boldsymbol{x} + \lambda^2\boldsymbol{s}^T\boldsymbol{A}\boldsymbol{s}) + \boldsymbol{b}^T\boldsymbol{x} + \lambda\boldsymbol{b}^T\boldsymbol{s} + c$$

Το ελάχιστο μιας τετραγωνικής συνάρτησης στη διεύθυνση s

Βρήκαμε ότι:

$$\varphi(\lambda) = \frac{1}{2}(\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x} + 2\lambda\boldsymbol{s}^T\boldsymbol{A}\boldsymbol{x} + \lambda^2\boldsymbol{s}^T\boldsymbol{A}\boldsymbol{s}) + \boldsymbol{b}^T\boldsymbol{x} + \lambda\boldsymbol{b}^T\boldsymbol{s} + c$$

Υπολογίζουμε την παράγωγο της $\varphi(\lambda)$

$$\varphi'(\lambda) = \frac{1}{2}(2\boldsymbol{s}^T\boldsymbol{A}\boldsymbol{x} + 2\lambda\boldsymbol{s}^T\boldsymbol{A}\boldsymbol{s}) + \boldsymbol{b}^T\boldsymbol{s} =$$
$$\boldsymbol{s}^T\boldsymbol{A}\boldsymbol{x} + \lambda\boldsymbol{s}^T\boldsymbol{A}\boldsymbol{s} + \boldsymbol{b}^T\boldsymbol{s} =$$
$$\boldsymbol{s}^T\boldsymbol{A}\boldsymbol{x} + \lambda\boldsymbol{s}^T\boldsymbol{A}\boldsymbol{s} + \boldsymbol{s}^T\boldsymbol{b} =$$
$$\boldsymbol{s}^T(\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}) + \lambda\boldsymbol{s}^T\boldsymbol{A}\boldsymbol{s} =$$
$$\boldsymbol{s}^T\boldsymbol{\nabla}q + \lambda\boldsymbol{s}^T\boldsymbol{A}\boldsymbol{s}$$

Βρίσκουμε που μηδενίζεται η παράγωγος:

$$\varphi'(\lambda) = 0 \ \Rightarrow$$
$$\boldsymbol{s}^T\boldsymbol{\nabla}q + \lambda\boldsymbol{s}^T\boldsymbol{A}\boldsymbol{s} = 0 \ \Rightarrow$$
$$\lambda = -\frac{\boldsymbol{s}^T\boldsymbol{\nabla}q}{\boldsymbol{s}^T\boldsymbol{A}\boldsymbol{s}}$$

Το ελάχιστο μιας τετραγωνικής συνάρτησης στη διεύθυνση s

Η δεύτερη παράγωγος της $\varphi(\lambda)$ είναι:

$$\varphi''(\lambda) = \boldsymbol{s}^T \boldsymbol{A}\boldsymbol{s} \quad > 0$$

αφού ο πίνακας $\boldsymbol{A}$ είναι θετικά όρισμένος.

Άρα η τιμή του λ που βρήκαμε αντιστοιχεί σε ελάχιστο.

Το ελάχιστο μιας τετραγωνικής συνάρτησης στη διεύθυνση s – Παράδειγμα

Θεωρείστε τη συνάρτηση

$$f(x_1, x_2) = x_1^2 + x_2^2$$

και το αρχικό σημείο $\boldsymbol{x} = (1,1)$.

Βρείτε ένα νέο σημείο $\widetilde{\boldsymbol{x}}$ με χαμηλότερη τιμή συνάρτησης πραγματοποιώντας γραμμική αναζήτηση στη διεύθυνση $\boldsymbol{s} = (-1,0)$.

Η συνάρτηση μπορεί να γραφεί στη μορφή

$$f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T \boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}^T\boldsymbol{x} + c$$

με $\boldsymbol{A} = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$ $\quad \boldsymbol{b} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ $\quad c = 0$

$$f(\boldsymbol{x}) = \frac{1}{2}[x_1 \quad x_2]\begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} =$$

$$\frac{1}{2}[x_1 \quad x_2]\begin{bmatrix} 2x_1 \\ 2x_2 \end{bmatrix} =$$

$$\frac{1}{2}(2x_1^2 + 2x_2^2) =$$

$$x_1^2 + x_2^2$$

Το ελάχιστο μιας τετραγωνικής συνάρτησης στη διεύθυνση s – Παράδειγμα

Το διάνυσμα των πρώτων παραγώγων είναι:

$$\nabla f(\boldsymbol{x}) = \begin{bmatrix} \dfrac{\partial f}{\partial x_1} \\ \dfrac{\partial f}{\partial x_2} \end{bmatrix} = \begin{bmatrix} 2x_1 \\ 2x_2 \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$$

Το ελάχιστο κατά μήκος της διεύθυνσης $\boldsymbol{s}$ είναι:

$$\lambda^* = -\frac{\boldsymbol{s}^T \nabla f}{\boldsymbol{s}^T \boldsymbol{A}\boldsymbol{s}} =$$

$$-\frac{[-1 \quad 0]\begin{bmatrix} 2 \\ 2 \end{bmatrix}}{[-1 \quad 0]\begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}\begin{bmatrix} -1 \\ 0 \end{bmatrix}} =$$

$$-\frac{-2}{[-1 \quad 0]\begin{bmatrix} -2 \\ 0 \end{bmatrix}} =$$

$$-\frac{-2}{2} = 1$$

Το νέο σημείο είναι:

$$\widetilde{\boldsymbol{x}} = \boldsymbol{x} + \lambda\boldsymbol{s} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} + 1 \cdot \begin{bmatrix} -1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

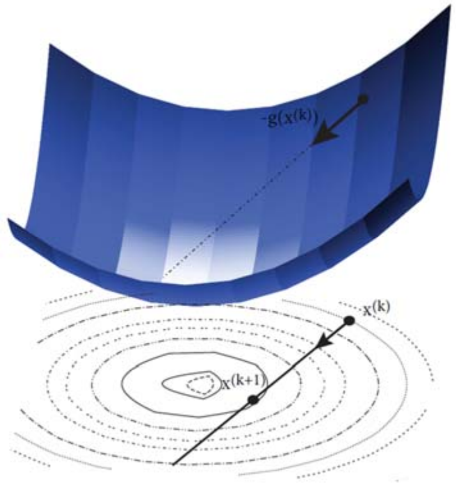
Η μέθοδος της πιο απότομης καθόδου

Η βασική ιδέα:

Χρησιμοποιούμε τη διεύθυνση της πιο απότομης καθόδου:

$$\boldsymbol{s} = -\boldsymbol{g}$$

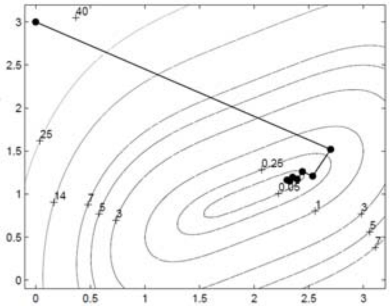
ως διεύθυνση γραμμικής αναζήτησης.



Η μέθοδος της πιο απότομης καθόδου

Η μέθοδος της πιο απότομης καθόδου

1. Δίνεται αρχικό σημείο $\boldsymbol{x}^{(0)}$
2. Επανάληψη για $k = 0,1,2,\cdots$
   a. Έλεγχος των κριτηρίων τερματισμού
   b. Υπολογίζουμε την παράγωγο $\boldsymbol{g}^{(k)}$ και θέτουμε $\boldsymbol{s}^{(k)} = -\boldsymbol{g}^{(k)}$
   c. Γραμμική αναζήτηση στη διεύθυνση $\boldsymbol{s}^{(k)}$, δηλαδή ελαχιστοποίηση ως προς a της συνάρτησης:
      $$\varphi(a) = f\left(\boldsymbol{x}^{(k)} + a\boldsymbol{s}^{(k)}\right)$$
   d. Θέτουμε $\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} + a^*\boldsymbol{s}^{(k)}$



Πλεονεκτήματα:

- Απλή υλοποίηση
- Απαιτείται μνήμη μόνο $O(N)$

Μειονεκτήματα:

- Εξαιρετικά αργή σύγκλιση.

Η μέθοδος Newton (πολυδιάστατη)

Μονοδιάστατη μέθοδος Newton (υπενθύμιση)

Προσεγγίζουμε τη συνάρτηση με το τετραγωνικό της μοντέλο γύρω από ένα σημείο a

$$q(x) = f(a) + f'(a)(x-a) + f''(a)\frac{(x-a)^2}{2}$$

Η παράγωγος του $q(x)$ είναι:

$$q'(x) = f'(a) + f''(a)(x-a)$$

Βρίσκουμε που μηδενίζεται η παράγωγος:

$$q'(x) = 0 \;\Rightarrow$$

$$f'(a) + f''(a)(x-a) = 0 \;\Rightarrow$$

$$x = a \boxed{-\frac{f'(a)}{f''(a)}}$$

Βήμα Newton

Η μέθοδος Newton (πολυδιάστατη)

Προσεγγίζουμε τη συνάρτηση με το τετραγωνικό της μοντέλο γύρω από ένα σημείο $\boldsymbol{a}$

$$q(\boldsymbol{x}) = f(\boldsymbol{a}) + \boldsymbol{g}(\boldsymbol{x})^T(\boldsymbol{x}-\boldsymbol{a}) + \frac{1}{2}(\boldsymbol{x}-\boldsymbol{a})^T\boldsymbol{G}(\boldsymbol{x})(\boldsymbol{x}-\boldsymbol{a})$$

Γνωρίζουμε ότι το ελάχιστο του $q(\boldsymbol{x})$ είναι:

$$\boldsymbol{x} = \boldsymbol{a} \boxed{-\boldsymbol{G}^{-1}\boldsymbol{g}}$$

$$\boldsymbol{h} = -\boldsymbol{G}^{-1}\boldsymbol{g}$$

Βήμα Newton
για ελαχιστοποίηση

Η μέθοδος Newton (πολυδιάστατη)

Μέθοδος Newton

1. Δίνεται αρχικό σημείο $\boldsymbol{x}^{(0)}$
2. Επανάληψη για $k = 0,1,2,\cdots$
   a. Έλεγχος των κριτηρίων τερματισμού
   b. Υπολογίζουμε τις παραγώγους $\boldsymbol{g}^{(k)}$ και $\boldsymbol{G}^{(k)}$
   c. Υπολογίζουμε το βήμα Newton λύνοντας: $\boldsymbol{G}^{(k)}\boldsymbol{h}^{(k)} = -\boldsymbol{g}^{(k)}$
   d. Θέτουμε $\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} + \boldsymbol{h}^{(k)}$

Πλεονεκτήματα:

- Εξαιρετικά γρήγορη σύγκλιση αν το αρχικό σημείο είναι κοντά στο ελάχιστο.

Μειονεκτήματα:

- Απαιτείται υπολογισμός του πίνακα των δευτέρων παραγώγων και επίλυση γραμμικού συστήματος.
- Απαιτείται μνήμη $O(N^2)$
- Ο πίνακας $\boldsymbol{G}$ μπορεί να μην είναι θετικά ορισμένος.
- Μπορεί να μην συγκλίνει για αρχικά σημεία μακριά από το ελάχιστο.

Η μέθοδος Newton (πολυδιάστατη)

Αν ο πίνακας $\boldsymbol{G}$ δεν είναι θετικά ορισμένος, τότε η διεύθυνση $\boldsymbol{h}$ που προκύπτει στη μέθοδο Newton δεν είναι φθίνουσα διεύθυνση.

Υπενθύμιση:

Θετικά ορισμένος πίνακας $\boldsymbol{G}$ σημαίνει:

$$\boldsymbol{s}^T\boldsymbol{G}\boldsymbol{s} > 0 \qquad \forall \boldsymbol{s} \neq 0$$

Φθίνουσα διεύθυνση $\boldsymbol{s}$ σημαίνει:

$$\boldsymbol{s}^T\boldsymbol{g} < 0$$

Η διεύθυνση $\boldsymbol{h}$ προκύπτει από την επίλυση της

$$\boldsymbol{G}\boldsymbol{h} = -\boldsymbol{g}$$

Πολλαπλασιάζουμε κάθε μέλος από αριστερά με $\boldsymbol{h}^T$

$$\boldsymbol{h}^T\boldsymbol{G}\boldsymbol{h} = -\boldsymbol{h}^T\boldsymbol{g}$$

Αν ο πίνακας G είναι θετικά ορισμένος τότε το αριστερό μέλος είναι θετικό και άρα

$$\boldsymbol{h}^T\boldsymbol{g} < 0$$

Η μέθοδος Newton (πολυδιάστατη)

Τροποποιημένη μέθοδος Newton

1. Δίνεται αρχικό σημείο $\boldsymbol{x}^{(0)}$
2. Επανάληψη για $k = 0,1,2,\cdots$
   a. Έλεγχος των κριτηρίων τερματισμού
   b. Υπολογίζουμε τις παραγώγους $\boldsymbol{g}^{(k)}$ και $\boldsymbol{G}^{(k)}$
   c. Τροποποιούμε τον $\boldsymbol{G}^{(k)}$ ώστε να είναι θετικά ορισμένος
   d. Υπολογίζουμε το βήμα Newton λύνοντας: $\boldsymbol{G}^{(k)}\boldsymbol{h}^{(k)} = -\boldsymbol{g}^{(k)}$
   e. Γραμμική αναζήτηση στη διεύθυνση $\boldsymbol{h}^{(k)}$, δηλαδή ελαχιστοποίηση ως προς a της συνάρτησης
   $$\varphi(a) = f\left(\boldsymbol{x}^{(k)} + a\boldsymbol{h}^{(k)}\right)$$
   f. Θέτουμε $\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} + a^*\boldsymbol{h}^{(k)}$

Μέθοδος Gauss–Newton για αθροίσματα τετραγώνων

Κάποιες συναρτήσεις έχουν την ειδική μορφή ενός αθροίσματος τετραγώνων

$$f(\boldsymbol{x}) = \frac{1}{2}\left(r_1^2(\boldsymbol{x}) + r_2^2(\boldsymbol{x}) + \cdots + r_M^2(\boldsymbol{x})\right)$$

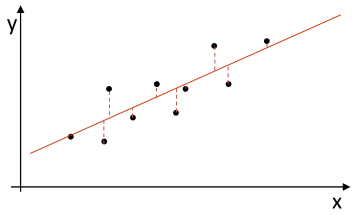
Τέτοιες συναρτήσεις προκύπτουν από προσαρμογή δεδομένων σε αναλυτικά μοντέλα.

Παράδειγμα συνάρτησης αθροίσματος τετραγώνων

Δίνεται ένα σύνολο σημείων

$$(x_i, y_i),\ i = 1 \cdots M$$

Ποια είναι η ευθεία που περνάει με “βέλτιστο τρόπο” από τα σημεία;



Η εξίσωση της ευθείας είναι: $y = ax + b$

Προσδιορίζουμε τα a και b ως εξής:

1. Υπολογίζουμε την απόσταση κάθε σημείου από την ευθεία: $|ax_i + b - y_i|$
2. Κατασκευάζουμε τη συνάρτηση σφάλματος αθροίζοντας τα τετράγωνα όλων των αποστάσεων:
$$f(a,b) = \sum_{i=1}^{M}(ax_i + b - y_i)^2$$
3. Βρίσκουμε τις τιμές a και b που δίνουν την ελάχιστη τιμή της $f(a,b)$

Μέθοδος Gauss–Newton για αθροίσματα τετραγώνων

Για το προγούμενο παράδειγμα:

$$f(a,b) = \sum_{i=1}^{M} (ax_i + b - y_i)^2$$

η συνάρτηση $f(a,b)$ μπορεί να γραφεί στη μορφή:

$$f(\boldsymbol{x}) = \frac{1}{2}\left(r_1^2(\boldsymbol{x}) + r_2^2(\boldsymbol{x}) + \cdots + r_M^2(\boldsymbol{x})\right)$$

με

$$r_i = \sqrt{2}(ax_i + b - y_i)$$

Τα r_i ονομάζονται “υπόλοιπα” και αποτελούν διάνυσμα $\boldsymbol{r}$ με M στοιχεία, οπότε οι συναρτήσεις που είναι αθροίσματα τετραγώνων γράφονται:

$$\boxed{f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{r}^T\boldsymbol{r}}$$

Άθροισμα τετραγώνων

Στις συναρτήσεις που είναι αθροίσματα τετραγώνων οι πρώτες και δεύτερες παράγωγοι έχουν ειδική μορφή.

Μέθοδος Gauss–Newton για αθροίσματα τετραγώνων

$$f(\boldsymbol{x}) = \frac{1}{2}\left(r_1^2(\boldsymbol{x}) + r_2^2(\boldsymbol{x}) + \cdots + r_M^2(\boldsymbol{x})\right) = \frac{1}{2}\sum_{i=1}^{M} r_i^2(\boldsymbol{x})$$

Η παράγωγος ως προς μια από τις μεταβλητές είναι:

$$\frac{\partial f}{\partial x_k} =$$

$$\frac{\partial}{\partial x_k}\left[\frac{1}{2}\sum_{i=1}^{M} r_i^2(\boldsymbol{x})\right] =$$

$$\frac{1}{2}\sum_{i=1}^{M} \frac{\partial}{\partial x_k} r_i^2(\boldsymbol{x}) =$$

$$\frac{1}{2}\sum_{i=1}^{M} 2\frac{\partial r_i}{\partial x_k} r_i =$$

$$\sum_{i=1}^{M} \frac{\partial r_i}{\partial x_k} r_i =$$

$$\frac{\partial r_1}{\partial x_k} r_1 + \frac{\partial r_2}{\partial x_k} r_2 + \cdots + \frac{\partial r_M}{\partial x_k} r_M =$$

$$\begin{bmatrix} \frac{\partial r_1}{\partial x_k} & \frac{\partial r_2}{\partial x_k} & \cdots \end{bmatrix} \begin{bmatrix} r_1 \\ r_2 \\ \vdots \end{bmatrix}$$

Μέθοδος Gauss–Newton για αθροίσματα τετραγώνων

Το συνολικό διάνυσμα των πρώτων παραγώγων είναι:

$$\boldsymbol{g} = \boldsymbol{J}^T\boldsymbol{r}$$

Ο πίνακας J ονομάζεται Ιακωβιανός πίνακας και έχει στοιχεία:

$$J_{kl} = \frac{\partial r_k}{\partial x_l}$$

$$J = \begin{bmatrix} \frac{\partial r_1}{\partial x_1} & \frac{\partial r_1}{\partial x_2} & \cdots & \frac{\partial r_1}{\partial x_N} \\ \frac{\partial r_2}{\partial x_1} & \frac{\partial r_2}{\partial x_2} & & \frac{\partial r_2}{\partial x_N} \\ \vdots & & \ddots & \\ \frac{\partial r_M}{\partial x_1} & \frac{\partial r_M}{\partial x_2} & & \frac{\partial r_M}{\partial x_N} \end{bmatrix}$$

Μπορούμε να υπολογίσουμε το διάνυσμα των πρώτων παραγώγων υπολογίζοντας τις παραγώγους των επιμέρους υπολοίπων r_i

Μέθοδος Gauss–Newton για αθροίσματα τετραγώνων

Παρόμοια υπολογίζεται ο πίνακας των δεύτερων παραγώγων ως:

$$\boldsymbol{G} = \boldsymbol{J}^T\boldsymbol{J} + \sum_{i=1}^{M} r_i \nabla^2 r_i$$

Τα προβλήματα ελαχιστοποίησης αθροισμάτων τετραγώνων κατηγοριοποιούνται ως:

- Προβλήματα μικρών υπολοίπων

  $$r_i(x^*) \approx 0$$

  Τότε η δεύτερες παράγωγοι της συνάρτησης μπορούν να προσεγγιστούν ως:

  $$\boldsymbol{G} \approx \boldsymbol{J}^T\boldsymbol{J}$$

  δηλαδή οι δεύτερες παράγωγοι της συνάρτησης μπορούν να υπολογιστούν αν γνωρίζουμε τις πρώτες παραγώγους των υπολοίπων.
- Προβλήματα μεγάλων υπολοίπων

  $$|r_i(x^*)| \gg 0$$

Μέθοδος Gauss–Newton για αθροίσματα τετραγώνων

Μέθοδος Gauss-Newton για μικρά υπόλοιπα

1. Δίνεται αρχικό σημείο $\boldsymbol{x}^{(0)}$
2. Επανάληψη για $k = 0,1,2,\cdots$
   a. Έλεγχος των κριτηρίων τερματισμού
   b. Υπολογίζουμε τον Ιακωβιανό πίνακα $\boldsymbol{J}^{(k)}$
   c. Υπολογίζουμε τις παραγώγους $\boldsymbol{g}^{(k)} = \boldsymbol{J}^{T(k)}\,\boldsymbol{r}^{(k)}$ και $\boldsymbol{G}^{(k)} = \boldsymbol{J}^{T(k)}\,\boldsymbol{J}^{(k)}$
   d. Τροποποιούμε τον $\boldsymbol{G}^{(k)}$ ώστε να είναι θετικά ορισμένος
   e. Υπολογίζουμε το βήμα Newton λύνοντας: $\boldsymbol{G}^{(k)}\boldsymbol{h}^{(k)} = -\boldsymbol{g}^{(k)}$
   f. Γραμμική αναζήτηση στη διεύθυνση $\boldsymbol{h}^{(k)}$, δηλαδή ελαχιστοποίηση ως προς a της συνάρτησης
   $$\varphi(a) = f\left(\boldsymbol{x}^{(k)} + a\boldsymbol{h}^{(k)}\right)$$
   g. Θέτουμε $\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} + a^{*}\boldsymbol{h}^{(k)}$

Μέθοδοι quasi–Newton

Κάποια από τα μειονεκτήματα της μεθόδου Newton είναι ότι:

- Απαιτείται υπολογισμός του πίνακα των δευτέρων παραγώγων $\boldsymbol{G}$.
- Ο πίνακας $\boldsymbol{G}$ μπορεί να μην είναι θετικά ορισμένος.

Βασική ιδέα των μεθόδων quasi-Newton:

Δεν χρησιμοποιείται ο πίνακας των δευτέρων παραγώγων αλλά χρησιμοποιείται μία προσέγγισή του
$$\boldsymbol{B} \approx \boldsymbol{G}$$

Θα θέλαμε να μην υπολογίζουμε εκ νέου σε κάθε επανάληψη τον πίνακα B, αλλά να τον βρίσκουμε από τον πίνακα B της προηγούμενης επανάληψης.

Υπενθύμιση: Τετραγωνικό μοντέλο πολυδιάστατης συνάρτησης

Ανάπτυγμα Taylor πολυδιάστατης συνάρτησης γύρω από το σημείο $\boldsymbol{a}$ όπου κρατάμε μέχρι τους όρους δεύτερης τάξης.

$$q(\boldsymbol{x}) = f(\boldsymbol{a}) + \boldsymbol{g}(\boldsymbol{x})^{T}(\boldsymbol{x}-\boldsymbol{a}) + \frac{1}{2}(\boldsymbol{x}-\boldsymbol{a})^{T}\boldsymbol{G}(\boldsymbol{x})(\boldsymbol{x}-\boldsymbol{a})$$

ή

$$q(\boldsymbol{p}) = f(\boldsymbol{x}+\boldsymbol{p}) = f(\boldsymbol{x}) + \boldsymbol{g}(\boldsymbol{x})^{T}\boldsymbol{p} + \frac{1}{2}\boldsymbol{p}^{T}\boldsymbol{G}(\boldsymbol{x})\boldsymbol{p}$$

Μέθοδοι quasi–Newton

Φτιάχνουμε ένα τετραγωνικό μοντέλο στο σημείο $\boldsymbol{x}_k$ χρησιμοποιώντας ένα συμμετρικό και θετικά ορισμένο πίνακα $\boldsymbol{B}$.

$$q_k(\boldsymbol{p}) = \frac{1}{2}\boldsymbol{p}^{T}\boldsymbol{B}_k\boldsymbol{p} + \boldsymbol{g}_k^{T}\boldsymbol{p} + f_k$$

Ο πίνακας $\boldsymbol{B}_k$ αποτελεί μια προσέγγιση (όχι απαραίτητα καλή) του πίνακα των δευτέρων παραγώγων $\boldsymbol{G}_k$.

Η παράγωγος του $q_k(\boldsymbol{p})$ είναι:
$$\nabla q_k(\boldsymbol{p}) = \boldsymbol{B}_k\boldsymbol{p} + \boldsymbol{g}_k$$

Το ελάχιστο του $q(\boldsymbol{p})$ είναι:
$$\boldsymbol{p}_k = \boldsymbol{B}_k^{-1}\boldsymbol{g}_k$$

Αν $\boldsymbol{B}_k = \boldsymbol{G}_k$, τότε το $\boldsymbol{p}_k$ είναι το βήμα Newton.

Κάνουμε γραμμική αναζήτηση κατά μήκος της διεύθυνσης $\boldsymbol{p}_k$, δηλαδή βρίσκουμε το ελάχιστο ως προς α της συνάρτησης:
$$\varphi(\alpha) = f(\boldsymbol{x}_k + \alpha\boldsymbol{p}_k)$$

Το νέο σημείο είναι:
$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + a_k\boldsymbol{p}_k$$

Το τετραγωνικό μοντέλο στο νέο σημείο x_{k+1} είναι:
$$q_{k+1}(\boldsymbol{p}) = \frac{1}{2}\boldsymbol{p}^{T}\boldsymbol{B}_{k+1}\boldsymbol{p} + \boldsymbol{g}_{k+1}^{T}\boldsymbol{p} + f_{k+1}$$

Πως θα βρούμε τον $\boldsymbol{B}_{k+1}$ (στο νέο σημείο $\boldsymbol{x}_{k+1}$) ;

Μέθοδοι quasi–Newton

Τετραγωνικό μοντέλο στο "παλαιό" σημείο:

$$q_k(\boldsymbol{p}) = \frac{1}{2}\boldsymbol{p}^T\boldsymbol{B}_k\boldsymbol{p} + \boldsymbol{g}_k^T\boldsymbol{p} + f_k$$

Τετραγωνικό μοντέλο στο "νέο" σημείο:

$$q_{k+1}(\boldsymbol{p}) = \frac{1}{2}\boldsymbol{p}^T\boldsymbol{B}_{k+1}\boldsymbol{p} + \boldsymbol{g}_{k+1}^T\boldsymbol{p} + f_{k+1}$$

Απαιτούμε ότι:

Στο νέο σημείο $\boldsymbol{x}_{k+1}$: Οι παράγωγοι του νέου μοντέλου q_{k+1} να είναι ίσες με τις παραγώγους τις συνάρτησης, δηλαδή

$$\nabla q_{k+1}(0) = \boldsymbol{g}_{k+1}$$

Αυτό ισχύει εκ κατασκευής του μοντέλου, διότι

$$\nabla q_{k+1}(p) = \boldsymbol{B}_{k+1}\boldsymbol{p} + \boldsymbol{g}_{k+1} \;\Rightarrow$$
$$\nabla q_{k+1}(0) = \boldsymbol{g}_{k+1}$$

Στο παλαιό σημείο $\boldsymbol{x}_k$: Οι παράγωγοι του νέου μοντέλου q_{k+1} να είναι ίσες με τις παραγώγους της συνάρτησης, δηλαδή

$$\nabla q_{k+1}(-\alpha_k\boldsymbol{p}_k) = \boldsymbol{g}_k$$

Όμως η παράγωγος του $q_{k+1}(\boldsymbol{p})$ είναι

$$\nabla q_{k+1}(\boldsymbol{p}) = \boldsymbol{B}_{k+1}\boldsymbol{p} + \boldsymbol{g}_{k+1}$$

Μέθοδοι quasi–Newton

Συνεπώς:

$$\nabla q_{k+1}(-\alpha_k\boldsymbol{p}_k) = \boldsymbol{g}_k \;\Rightarrow$$
$$\boldsymbol{B}_{k+1}(-\alpha_k\boldsymbol{p}_k) + \boldsymbol{g}_{k+1} = \boldsymbol{g}_k \;\Rightarrow$$
$$\boldsymbol{B}_{k+1}(\boldsymbol{x}_{k+1} - \boldsymbol{x}_k) = \boldsymbol{g}_{k+1} - \boldsymbol{g}_k$$

Ονομάζουμε:

$$\boldsymbol{\delta}_k = \boldsymbol{x}_{k+1} - \boldsymbol{x}_k$$
$$\boldsymbol{\gamma}_k = \boldsymbol{g}_{k+1} - \boldsymbol{g}_k$$

$$\boxed{\boldsymbol{B}_{k+1}\boldsymbol{\delta}_k = \boldsymbol{\gamma}_k}$$

Εξίσωση τέμνουσας

Πολλαπλασιάζοντας την εξίσωση τέμνουσας με $\boldsymbol{\delta}_k^T$ από αριστερά

$$\boldsymbol{\delta}_k^T\boldsymbol{B}_{k+1}\boldsymbol{\delta}_k = \boldsymbol{\delta}_k^T\boldsymbol{\gamma}_k$$

Όμως

$$\boldsymbol{\delta}_k^T\boldsymbol{B}_{k+1}\boldsymbol{\delta}_k > 0$$

αφού ο $\boldsymbol{B}_{k+1}$είναι θετικά ορισμένος. Άρα

$$\boxed{\boldsymbol{\delta}_k^T\boldsymbol{\gamma}_k > 0}$$

Συνθήκη κυρτότητας

Μέθοδοι quasi–Newton

Για να βρούμε τον $\boldsymbol{B}_{k+1}$ απαιτούμε να είναι "κοντά" στον $\boldsymbol{B}_k$, δηλαδή

$$\min_B \| \boldsymbol{B} - \boldsymbol{B}_k \|$$

υπό τις συνθήκες

$$\boldsymbol{B} = \boldsymbol{B}^T \quad \text{και} \quad \boldsymbol{B}\boldsymbol{\delta}_k = \boldsymbol{\gamma}_k$$

$$\boldsymbol{B}_{k+1} = \left(\boldsymbol{I} - \frac{\boldsymbol{\gamma}_k\boldsymbol{\delta}_k^T}{\boldsymbol{\gamma}_k^T\boldsymbol{\delta}_k}\right)\boldsymbol{B}_k\left(\boldsymbol{I} - \frac{\boldsymbol{\delta}_k\boldsymbol{\gamma}_k^T}{\boldsymbol{\gamma}_k^T\boldsymbol{\delta}_k}\right) + \frac{\boldsymbol{\gamma}_k\boldsymbol{\gamma}_k^T}{\boldsymbol{\gamma}_k^T\boldsymbol{\delta}_k}$$

Μέθοδος Davidon-Fletcher-Powell

$$\boldsymbol{B}_{k+1} = \boldsymbol{B}_k - \frac{\boldsymbol{B}_k\boldsymbol{\delta}_k\boldsymbol{\delta}_k^T\boldsymbol{B}_k}{\boldsymbol{\delta}_k^T\boldsymbol{B}_k\boldsymbol{\delta}_k} + \frac{\boldsymbol{\gamma}_k\boldsymbol{\gamma}_k^T}{\boldsymbol{\gamma}_k^T\boldsymbol{\delta}_k}$$

Μέθοδος Broyden-Fletcher-Goldfarb-Shanno

Μέθοδοι quasi–Newton

Γενικός αλγόριθμος μεθόδων quasi-Newton

1. Δίνεται ένα αρχικό σημείο $\boldsymbol{x}_0$ και η αρχική προσέγγιση $\boldsymbol{B}_0$
2. Επανάληψη για $k = 0,1,2,\cdots$
   a. Έλεγχος των κριτηρίων τερματισμού
   b. Υπολογίζεται το διάνυσμα των πρώτων παραγώγων $\boldsymbol{g}_k$
   c. Υπολογίζεται η διεύθυνση $\boldsymbol{p}_k = -\boldsymbol{B}_k^{-1}\boldsymbol{g}_k$
   d. Γραμμική αναζήτηση στη διεύθυνση $\boldsymbol{p}_k$, δηλαδή ελαχιστοποίηση ως προς a της συνάρτησης
   $$\varphi(\alpha) = f(\boldsymbol{x}_k + a\boldsymbol{p}_k)$$
   e. Θέτουμε $\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + a_k^*\boldsymbol{p}_k$
   f. Ενημερώνουμε τον $\boldsymbol{B}_k$ και παίρνουμε τον $\boldsymbol{B}_{k+1}$

Μέθοδοι quasi–Newton

Πως βρίσκουμε τον αρχικό πίνακα $\boldsymbol{B}_0$;

- Χρησιμοποιούμε την προσέγγιση
$$\boldsymbol{B}_0 = \boldsymbol{I}$$
- Υπολογίζουμε αριθμητικά τον πίνακα $\boldsymbol{G}$ των δευτέρων παραγώγων και θέτουμε $\boldsymbol{B}_0 = \boldsymbol{G}$ (υπό την προϋπόθεση ότι ο $\boldsymbol{G}$ είναι θετικά ορισμένος).