Quasi-Newton

Έχουμε δει ότι αν η Εσσιανή $\nabla^2 f(x_k)$ είναι θετικά ορισμένη και χρησιμοποιήσουμε τη μέθοδος Νεύτωνα για να υπολογίσουμε την επόμενη προσέγγιση

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

όπου λύνουμε το $\nabla^2 f(x_k) p_k = -\nabla f(x_k)$ προκύπτει μια κατεύθυνση καθόδου p_k, διότι $p_k^T \nabla f(x_k) < 0$

Αν όμως είναι “δύσκολο” να υπολογιστεί η εσσιανή σε κάθε βήμα τότε καταφεύγουμε σε μια προσέγγιση $B_k \approx \nabla^2 f(x_k)$, και λύνουμε το

$$B_k p_k = -\nabla f(x_k)$$

Αν ο B_k είναι θετικά ορισμένος η λύση p_k θα είναι μια κατεύθυνση καθόδου

Μέθοδος της τέμνουσας (1-διάσταση)

Για την εύρεση της ρίζας μιας πραγματικής συνάντησης στη 1 διάσταση μπορούμε να χρησιμοποιήσουμε τη μέθοδο της τέμνουσας αντί της μεθόδου του Νεύτωνα.

Λόγω της προσέγγισης

$$f'(x) \approx \frac{f(x) - f(x-h)}{h}$$

Τροποποιούμε τη μέθοδο του Νεύτωνα ως εξής: κατασκευάζουμε μια ακολουθία τιμών x_k, τέτοια ώστε

$$x_{k+1} = x_k - \frac{f(x_k)}{\dfrac{f(x_k) - f(x_{k-1})}{x_k - x_{k-1}}}, \quad k \geq 1, \quad \text{με δοσμένα } x_0, x_1$$

Ταχύτητα σύγκλισης

Η ταχύτητα σύγκλισης της μεθόδου της τέμνουσας είναι μικρότερη από αυτή της μεθόδου του Νεύτωνα και μπορεί να αποδειχθεί ότι είναι

$$p = \frac{1}{2}(1 + \sqrt{5}) \approx 1.618$$

Η γενίκευση της μεθόδου της τέμνουσας στις πολλές διαστάσεις καλείται συχνά και μέθοδος Broyden.

Παρατηρούμε λοιπόν ότι

$$\nabla^2 f(x_k)(x_k - x_{k-1}) \approx \nabla f(x_k) - \nabla f(x_{k-1})$$

Αν θελήσουμε να αντικαταστήσουμε τον $\nabla^2 f(x_k)$ με έναν πίνακα B_k και

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

λέμε ότι ο B_k ικανοποιεί τη ***συνθήκη της τέμνουσας***.

Παρατήρηση:

Αν $f(x) = \frac{1}{2}x^T A x - b^T x$, τότε $\nabla^2 f(x) = A$ και εύκολα μπορούμε να δούμε ότι

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

Δηλαδή ο A ικανοποιεί τη συνθήκη της τέμνουσας.

Συνθήκη τέμνουσας

Θα διατυπώσουμε τη συνθήκη της τέμνουσας διαφορετικά.

Θέτουμε

$$s_k = x_{k+1} - x_k \quad \text{και} \quad y_k = \nabla f(x_{k+1}) - \nabla f(x_k)$$

Η συνθήκη της τέμνουσας διατυπώνεται ως εξής

$$B_{k+1} s_k = y_k$$

Quasi - Newton

Οι μέθοδοι που προκύπτουν αν χρησιμοποιούμε πίνακες B_k αντί για την εσσιανη που ικανοποιούν την συνθήκη της τέμνουσας

Ένα παράδειγμα δίνεται από τον τύπο

$$B_{k+1} = B_k + \frac{(y_k - B_k s_k)(y_k - B_k s_k)^T}{(y_k - B_k s_k)^T s_k}$$

Quasi - Newton

Η συνθήκη της τέμνουσας ικανοποιείται γιατί

$$B_{k+1}s_k = B_k s_k + \frac{(y_k - B_k s_k)(y_k - B_k s_k)^T}{(y_k - B_k s_k)^T s_k} s_k$$

$$= B_k s_k + \frac{(y_k - B_k s_k)((y_k - B_k s_k)^T s_k)}{(y_k - B_k s_k)^T s_k}$$

$$= B_k s_k + (y_k - B_k s_k) = y_k$$

Quasi-Newton

Μπορούμε να χρησιμοποιήσουμε μια μέθοδο Quasi-Newton αρχίζοντας με $B_0 = I$ και εφαρμόζοντας τον προηγούμενο τύπο για να παίρνουμε την νέα προσέγγιση B_{k+1} από τον B_k.

Εκτός από τον προηγούμενο τύπο υπάρχουν και άλλες υλοποιήσεις μεθόδων Quasi-Newton. Όλες έχουν την ίδια μορφή, υπολογίζουμε το

$$B_{k+1} = B_k + [\text{τύπος ανανέωσης}]$$

Αλγόριθμος Quasi - Newton

1. Θεωρούμε μια αρχική προσέγγιση $x_0 \in \mathbb{R}^n$, μια αρχική προσέγγιση της εσσιανής B_0 (π.χ. $B_0 = I$ και ένα $\epsilon > 0$

2. Για $k = 0,1,2,\ldots$

   (i) Αν το $\|\nabla f(x_k)\| < \epsilon$ έχουμε βρει "σχεδόν" το ελάχιστο και σταματάμε

   (ii) Βρείτε μια κατεύθυνση καθόδου p_k λύνοντας το $B_k p_k = -\nabla f(x_k)$

   (iii) Βρείτε ένα βήμα a_k, ώστε $x_{k+1} = x_k + a_k p_k$ να είναι καλύτερη προσέγγιση από το x_k

   (iv) Υπολογίζουμε τα $s_k = x_{k+1} - x_k$, $y_k = \nabla f(x_{k+1}) - \nabla f(x_k)$

   (v) Υπολογίζουμε τον B_{k+1}, χρησιμοποιώντας ένα τύπο όπως π.χ. $B_{k+1} = B_k + \dfrac{(y_k - B_k s_k)(y_k - B_k s_k)^T}{(y_k - B_k s_k)^T s_k}$

Quasi-Newton

Το όφελος από τη χρήση μεθόδων Quasi-Newton είναι κυρίως ο περιορισμός του απαιτούμενου αριθμού πράξεων για την επίλυση του γραμμικού συστήματος $B_k p_k = -\nabla f(x_k)$.

Αν υπάρχει ένας τύπος ανανέωσης όπως των πινάκων B_k για μια ανάλυση του B_k, όπως π.χ. η ανάλυση Cholesky, και απαιτεί $O(n^2)$ πράξεις, τότε το κόστος της λύσης του $B_k p_k = -\nabla f(x_k)$ είναι πολύ μικρότερο από $O(n^3)$

Quasi-Newton

Μπορούμε να δούμε ότι ο τύπος ανανέωσης $B_{k+1} = B_k + \dfrac{(y_k - B_k s_k)(y_k - B_k s_k)^T}{(y_k - B_k s_k)^T s_k}$ διατηρεί τη συμμετρικότητα.

Καλείται και συμμετρικός τύπος ανανέωσης τάξεως ένα, διότι ο $\dfrac{(y_k - B_k s_k)(y_k - B_k s_k)^T}{(y_k - B_k s_k)^T s_k}$ είναι πίνακας τάξεως ένα.

Quasi-Newton (Πίνακας ανανέωσης 1ης τάξεως)

Λήμμα: Έστω B_k συμμετρικός πίνακας. Έστω $B_{k+1} = B_k + C$, και $C \neq 0$, πίνακας τάξεως ένα. Αν ο B_{k+1} είναι συμμετρικός, $B_{k+1}s_k = y_k$ και $(y_k - B_k s_k)^T s_k \neq 0$, τότε

$$C = \frac{(y_k - B_k s_k)(y_k - B_k s_k)^T}{(y_k - B_k s_k)^T s_k}$$

Απόδειξη

Απόδειξη: Αν $B_{k+1} = B_k + C$ και B_{k+1}, B_k είναι συμμετρικοί τότε και ο C θα είναι συμμετρικός.

Quasi-Newton (Θετικα ορισμένος πίνακας)

Μια άλλη επιθυμητή ιδιότητα που θα θέλαμε να έχει ο B_k είναι να είναι θετικά ορισμένος. Αν ισχύει αυτό τότε θα είμαστε σίγουροι ότι οι κατευθύνσεις p_k είναι κατευθύνσεις καθόδου.

Παρατήρηση:

Μπορούμε να δείξουμε ότι δεν υπάρχει τύπος ανανέωσης τάξεως ένα ο οποίος να διατηρεί τη συμμετρικότητα και τη θετική ορισιμότητα.

Quasi-Newton (Θετικα ορισμένος πίνακας)

Υπάρχουν πολλοί τύποι ανανέωσης με πίνακα τάξεως δύο οι οποίοι διατηρούν τη συμμετρικότητα και τη θετική ορισιμότητα.

Ένα παράδειγμα είναι ο τύπος BFGS

$$B_{k+1} = B_k - \frac{(B_k s_k)(B_k s_k)^T}{s_k^T B_k s_k} + \frac{y_k y_k^T}{y_k^T s_k}$$

Quasi-Newton (BFGS - Θετικά ορισμένος)

Λήμμα: Έστω B_k συμμετρικός και θετικά ορισμένος και παίρνουμε τον B_{k+1} χρησιμοποιώντας τον τύπο BFGS. Τότε ο B_{k+1} είναι θετικά ορισμένος αν και μόνο αν $y_k^T s_k > 0$

Σύγκλιση Quasi-Newton

Θεώρημα: Έστω f μια πραγματική συνάρτηση, $x_0 \in \mathbb{R}^n$ αρχική προσέγγιση και $\{x_k\}$, με $x_{k+1} = x_k + a_k p_k$. Υποθέτουμε ότι

(i) $S = \{x : f(x) \leq f(x_0)\}$

(ii) $f, \ \nabla f, \ \nabla^2 f$ είναι συνεχείς για $x \in S$

(iii) $\nabla^2 f(x)$ είναι θετικά ορισμένος για όλα τα x

(iv) Τα $\{p_k\}$ υπολογίζονται ως $B_k p_k = -\nabla f(x_k)$, όπου $B_0 = I$ και B_k δίνονται από το τύπο ανανέωσης BFGS

(v)Τα $\{a_k\}$ ικανοποιούν

$$f(x_k + a_k p_k) \leq f(x_k) + \mu a_k p_k^T \nabla f(x_k)$$

$$p_k^T f(x_k + a_k p_k) \geq \eta p_k^T \nabla f(x_k)$$

με $0 < \mu < \eta < 1$ και $a_k = 1$ όποτε είναι δυνατόν