

Μέθοδοι Μεταβλητής Μετρικής (Variable Metric, Quasi-Newton, Secant)

Γνωρίζω και ως μέθοδοι της χορδής, και Quasi-Newton.
Έχω $q(h)$ και ζητάω να βρω το ελάχιστο.

$$q(h) = \frac{1}{2} h^T B h + d^T h + c$$

$$g(h) \equiv \nabla q(h) = B h + d$$

$$\nabla q(h_2) - \nabla q(h_1) = B h_2 + d - (B h_1 + d) = B(h_2 - h_1)$$

$$g(h_2) - g(h_1) = B(h_2 - h_1) \quad \text{ή} \quad \text{εάν } \gamma = g(h_2) - g(h_1) \quad \text{και } \delta = h_2 - h_1$$

Έχουμε $\gamma = B\delta$ εάν δ δώσει

$$\delta = B^{-1}\gamma \equiv H\gamma \quad (H \equiv B^{-1})$$

Η παραπάνω σχέση ισχύει για μεταβλητικές μετρικές.

Στις μεθόδους Newton: έχουμε:

$$\delta^{(k)} = h^{(k+1)} - h^{(k)} \quad (h_1 = 0, h_2 = h^{(k)}) \quad h^{(k+1)} = -H^{(k)} g^{(k)}$$

$$\gamma^{(k)} = g(h^{(k+1)}) - g(h^{(k)}) = g^{(k+1)} - g^{(k)}$$

$$\text{όπου έχουμε} \quad \delta^{(k)} = -H^{(k)} g^{(k)} \neq H^{(k)} \gamma^{(k)}$$

Εάν όμως αναζητούμε να ισχύει: (Συνθήκη Q.N.)

$$\delta^{(k)} = H^{(k+1)} \gamma^{(k)} \quad (\text{ή καθυστέρηση})$$

τότε πρέπει να προσεγγίσουμε με διάφορους τρόπους το $H^{(k+1)}$.

Ενυψίωση 1^{ου} βαθμού (Rank-1 update).

Γράφουμε: $H^{(k+1)} = H^{(k)} + \alpha u u^T \quad u \in \mathbb{R}^N$

$$J^{(k)} = (H^{(k)} + \alpha u u^T) \gamma^{(k)}$$

(καταργώ τον εκδοίτη (k))

$$J = H\gamma + \alpha u(u^T \gamma) \Rightarrow \frac{J - H\gamma}{\alpha(u^T \gamma)} = u$$

Διατίθω το α έτσι ώστε:

$$\alpha(u^T \gamma) = 1 \Rightarrow u = J - H\gamma$$

$$\alpha u^T \gamma = \alpha (J^T - \gamma^T H) \gamma = 1 \Rightarrow \alpha = \frac{1}{J^T J - \gamma^T H \gamma}$$

Οπότε:

$$H^{(k+1)} = H^{(k)} + \frac{1}{J^T (J - H\gamma)} (J - H\gamma)(J - H\gamma)^T$$

Ο τσπος αυτός υποβλήθηκε ενυψίωση 1^{ου} βαθμού.

Όμως ο υποβλήθηκε ελαχιστοποίησης έχει ως:

Αρχικά $H^{(0)} = I$, $x^{(0)}$, $g^{(0)}$. Είναι γνωστά.

Στην k -επαύριση έχουμε:

$$\left\{ \begin{array}{l} h^{(k)} = -H^{(k)} g^{(k)} \\ \min_{\lambda > 0} f(x^{(k)} + \lambda h^{(k)}) \rightarrow \lambda^* \\ x^{(k+1)} = x^{(k)} + \lambda^* h^{(k)} \\ g^{(k+1)} = \nabla f(x^{(k+1)}) \\ H^{(k+1)} = H^{(k)} + \frac{(J^{(k)} - H^{(k)} g^{(k)}) (J^{(k)} - H^{(k)} g^{(k)})^T}{g^{(k)T} (J^{(k)} - H^{(k)} g^{(k)})} \end{array} \right. , \quad \begin{array}{l} g^{(k)} = g^{(k+1)} - g^{(k)} \\ J^{(k)} = \lambda^* h^{(k)} \end{array}$$

Χρησιμοποιώντας SHERMAN - MORRISON:

$$(H + \alpha uu^T)^{-1} = H^{-1} - \alpha \frac{(H^{-1}u)(H^{-1}u)^T}{1 - \alpha u^T H^{-1}u} = B - \alpha \frac{(Bu)(Bu)^T}{1 + \alpha u^T Bu}$$

ψι $u = \delta - H\gamma$ και $\alpha = 1/\gamma^T(\delta - H\gamma)$

Έχουμε: $B^{(k+1)} = B^{(k)} - \frac{(B^{(k)}\delta - \gamma^{(k)})(B^{(k)}\delta - \gamma^{(k)})^T}{\gamma^T (B^{(k)}\delta - \gamma^{(k)})}$

Αυτή είναι η αναδρομική σχέση ενημέρωσης για τον εσθιαρό πίνακα B.

Το φινιέρισμα του μεθόδου αυτής είναι ότι δεν διαρρηί τον εσθιαρό (B ή τον B⁻¹) θετικά ορισμένο.

Επίσης ο παρανοήτως, μπορεί να μπερδεύει όταν το δ έχει κάθεται στο Bδ-γ

Εντερικός Τελερ Βαθμοί
(Rank-2 updates).

$$H^{(k+1)} = H^{(k)} + \alpha u u^T + \beta v v^T$$

$$\delta = H\gamma + \alpha u u^T \gamma + \beta v v^T \gamma$$

$u, v \in \mathbb{R}^n$ δίνονται προσδιορίζονται με κανονικό τρόπο.

Εάν (αφαιρετικά) θέλουμε: $u = \delta$, και $\alpha = \frac{1}{\delta^T \delta}$

τότε: $v = H\gamma$ ή $b = \frac{-1}{v^T \gamma}$, τότε:

$$H^{(k+1)} = H^{(k)} + \frac{\delta \delta^T}{\delta^T \delta} - \frac{H \delta \delta^T H}{\delta^T H \delta}$$

και με χρήση της σχέσης Sherman-Morrison έχουμε:

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

Οι δύο παραπάνω εντερικοί είναι γνωστοί

ως: DFP - updates.

(Από τα αρχικά αλγ: Davidson, Fletcher, Powell)

BFGS - εμπειρική

Εάν σ δοστούς (από την αρχή) $u = \gamma$, $\alpha = \frac{1}{\gamma^T \gamma}$

και $\beta = \frac{-1}{\gamma^T B \gamma}$ έχουμε $v = -B \gamma$ και

$$B^{(k+1)} = B^{(k)} + \frac{\gamma \gamma^T}{\gamma^T \gamma} - \frac{B \gamma \gamma^T B}{\gamma^T B \gamma}$$

και για τον H :

$$H^{(k+1)} = H^{(k)} + \left(1 + \frac{\gamma^T H \gamma}{\gamma^T \gamma}\right) \frac{\gamma \gamma^T}{\gamma^T \gamma} - \left(\frac{\gamma \gamma^T H + H \gamma \gamma^T}{\gamma^T \gamma} \right)$$

Οι DFP και BFGS (Broyden, Fletcher, Goldfarb, Shanno)

είναι γ και γ^T σχεδόν (dual) και
μπορούν να εναλλαχθούν ή για από τον γ ή
για τις αντίστοιχες γ^T .

$$B \leftrightarrow H \text{ και } \gamma \leftrightarrow \gamma^T$$

Η εμπειρική του βαρέος είναι αυτοδual.
(self-dual).

Και οι δύο DFP και BFGS διατηρούν τον
εγγεγραμμό (B , ή $H = B^{-1}$) συμμετρικό ορισμένο.

Η ιδιότητα αυτή χάνεται μόνο από αριθμητικό
σφάλμα (round-off) προχρησθέντος.

Υπάρχουν τεχνά για την αποφυγή του φαινομένου αυτού.
Π.χ. διατηρώντας το γινόμενο $B = LDL^T$.

Βελτιστοποιήσεις για συναρτήσεις χωρίς
 Αθροισμα Τετραγώνων.

"Έστω: $f(x) = \sum_{i=1}^M r_i^2(x) \equiv r^T(x) r(x) \quad x \in \mathbb{R}^N$

Εάν $M = N$ τότε έχουμε ένα σύστημα
 N εξισώσεων με N άγνωστους: $r_i(x) = 0, i=1, 2, \dots, M$.

Εάν όμως $M > N$ (όπως είναι προσεγγιστικά δεδομένα)
 το σύστημα είναι υπερπροσδιορισμένο (overdetermined)
 και το σύστημα δεν έχει λύση.

Τα $r_i(x)$ ονομάζονται και "υπόλοιπα" (Residuals).

$$\nabla f(x) \equiv g(x) = 2 A^T r, \quad A = [\nabla r_1, \dots, \nabla r_M]$$

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

$$A_{ki} = \frac{\partial r_i(x)}{\partial x_k} \quad \text{δέχεται Ιακωβιανή,}$$

και είναι πίνακας $N \times M$

$$\nabla^2 f(x) \equiv B(x) = 2 A A^T + 2 \sum_{i=1}^M r_i(x) \nabla^2 r_i(x)$$

Εάν στο minimum x^* έχουμε: $r_i(x^*) \approx 0$
 τότε έχουμε την ίδια προσέγγιση με
 τον εσσιανό πίνακα $B(x)$:

$$B(x) = 2 A(x) A(x)^T \quad \text{θετικά ημιορισμένος.}$$

Η προσέγγιση αυτή ονομάζεται Gauss-Newton.

Η προσέγγιση G-N μπορεί να χρησιμοποιηθεί

ως πλαιγίο της Μεθόδου Newton αντικαθιστώντας
τον Hessiano με την προσέγγιση:

$$B \approx 2AA^T$$

Επίσης ως πλαιγίο των Quasi-Newton θέσεων
την αρχική προσέγγιση

$$B^{(0)} = 2AA^T$$

και στην συνέχεια εκπληρώνεται τον με Q-N έχει.

Όταν $r_f(x^*) < 0$ η Μέθοδος G-N (ως πλαιγίο
της Μεθόδου Newton) είναι ικανοποιητική.

Όταν $r_f(x^*) \gg 0$ τότε το Q-N πλαιγίο είναι
προτιμότερο.

Έχουν αναπτυχθεί υβριδικές μέθοδοι (Fletcher & Xu)
όπου το λέει:

$$f(x^{(k)}) - f^{(k+1)} > \tau f^{(k)} \quad \tau \in (0,1)$$

Διαλέγει G-N ή G-QN, εάν ισχύει ή όχι.

$$\text{Εάν } f^* \neq 0 \text{ τότε } f^{(k+1)} \approx O(f^{(k)}) \Rightarrow \frac{f^{(k)} - f^{(k+1)}}{f^{(k)}} \rightarrow 1$$

οπότε Διαλέγεται G-N.

$$\text{Εάν } f^* = 0 \text{ τότε } \frac{f^{(k)} - f^{(k+1)}}{f^{(k)}} \rightarrow 0 \text{ οπότε Διαλέγεται G-QN.}$$