

ΠΑΝΕΠΙΣΤΗΜΙΟΝ ΙΩΑΝΝΙΝΩΝ
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ

ΤΕΧΝΙΚΕΣ ΕΛΑΧΙΣΤΟΠΟΙΗΣΗΣ

ΣΗΜΕΙΩΣΕΙΣ

Ι. Η. ΛΑΓΑΡΗΣ
ΣΕΠΤΕΜΒΡΙΟΣ 1995

ΜΑΘΗΜΑΤΙΚΟ ΥΠΟΒΑΘΡΟ ΚΑΙ ΟΡΟΛΟΓΙΑ

Οι μέθοδοι της βελτιστοποίησης χρησιμοποιούν εκτεταμένα τεχνικές της γραμμικής άλγεβρας. Μια σύντομη αναδρομή στις πιο συχνά χρησιμοποιούμενες μεθόδους κρίνεται απαραίτητη.

Τα διανύσματα θα συμβολίζονται με μικρά έντονα γράμματα, οι δε πίνακες θα συμβολίζονται με κεφαλαία έντονα γράμματα.

Παράδειγμα:

$$\mathbf{a} = \begin{pmatrix} a_1 \\ a_2 \\ \cdot \\ \cdot \\ \cdot \\ a_n \end{pmatrix} ; \quad \mathbf{B} = \begin{bmatrix} B_{11} & B_{12} & \dots & B_{1n} \\ B_{21} & B_{22} & \dots & B_{2n} \\ & & \dots & \\ & & & \dots \\ & & & \\ B_{n1} & B_{n2} & \dots & B_{nn} \end{bmatrix}$$

Με τον πάνω δείκτη (\mathbb{T}) συμβολίζουμε τον ανάστροφο πίνακα ή διάνυσμα.

Παράδειγμα:

$$\mathbf{a}^T = (a_1, a_2, \dots, a_n) \text{ και } [\mathbf{B}^T]_{ij} = B_{ji}.$$

Το εσωτερικό γινόμενο δύο διανυσμάτων \mathbf{a} και \mathbf{b} δίνεται ως:

$$\mathbf{a}^T \mathbf{b} = \mathbf{b}^T \mathbf{a} = \sum_{i=1, n} a_i b_i .$$

Ένα σημείο \mathbf{x} στον n -διάστατο χώρο αναπαρίσταται με το διάνυσμα θέσεως $\mathbf{x}^T = (x_1, x_2, \dots, x_n)$.

Σημειώσεις: Τεχνικές βελτιστοποίησης

Η εξίσωση της ευθείας στον n -διάστατο χώρο ορίζεται ως το σύνολο των σημείων $\mathbf{x}(\lambda) = \mathbf{x}' + \lambda \mathbf{s}$, όπου \mathbf{x}' είναι ένα σταθερό σημείο από το οποίο διέρχεται η ευθεία, και \mathbf{s} ένα σταθερό διάνυσμα που καθορίζει την διεύθυνση της ευθείας.

Ο λογισμός συναρτήσεων πολλών μεταβλητών είναι επίσης σημαντικό εργαλείο για την ανάπτυξη και κατανόηση των μεθόδων βελτιστοποίησης.

Η κλίση μιας συνάρτησης $f(\mathbf{x})$ είναι το διάνυσμα $\nabla f(\mathbf{x})$ που έχει ως συνιστώσες τις μερικές παραγώγους της $f(\mathbf{x})$.

Ο Εσσιανός πίνακας (Hessian), όπως ονομάζεται ο πίνακας των δευτέρων παραγώγων, συμβολίζεται ως:

$$\nabla^2 f(\mathbf{x}), \text{ με στοιχεία: } \frac{\partial^2 f(\mathbf{x})}{\partial x_i \partial x_j}.$$

Αναλυτικά για την κλίση και τον Εσσιανό πίνακα έχουμε τις παρακάτω αναπαραστάσεις:

$$\nabla f(\mathbf{x}) = \begin{pmatrix} \frac{df(\mathbf{x})}{dx_1} \\ \cdot \\ \cdot \\ \cdot \\ \frac{df(\mathbf{x})}{dx_n} \end{pmatrix} \quad \text{και} \quad \nabla^2 f(\mathbf{x}) = \begin{bmatrix} \frac{\partial^2 f(\mathbf{x})}{\partial x_1^2} & \frac{\partial^2 f(\mathbf{x})}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f(\mathbf{x})}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f(\mathbf{x})}{\partial x_2 \partial x_1} & \frac{\partial^2 f(\mathbf{x})}{\partial x_2^2} & \cdots & \frac{\partial^2 f(\mathbf{x})}{\partial x_2 \partial x_n} \\ \cdot & \cdot & \cdot & \cdot \\ \frac{\partial^2 f(\mathbf{x})}{\partial x_n \partial x_1} & \frac{\partial^2 f(\mathbf{x})}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f(\mathbf{x})}{\partial x_n^2} \end{bmatrix}$$

Εάν η συνάρτηση $f(\mathbf{x})$ είναι συνεχώς διαφορίσιμη δύο φορές, τότε ο Εσσιανός πίνακας υπάρχει και είναι συμμετρικός. Ισχύει δηλαδή: $\frac{\partial^2 f(\mathbf{x})}{\partial x_i \partial x_j} = \frac{\partial^2 f(\mathbf{x})}{\partial x_j \partial x_i}$.

Η κλίση της $f(\mathbf{x})$, κατά μήκος της ευθείας γραμμής $\mathbf{x}(\lambda) = \mathbf{x}' + \lambda \mathbf{s}$ δίδεται από την σχέση:

$$\frac{d}{d\lambda} f(\mathbf{x}) = \sum_{i=1,n} \frac{d}{d\lambda} x_i(\lambda) \frac{\partial f(\mathbf{x})}{\partial x_i} = \sum_{i=1,n} s_i \frac{\partial f(\mathbf{x})}{\partial x_i} = \mathbf{s}^T \nabla f(\mathbf{x}) = [\nabla f(\mathbf{x})]^T \mathbf{s}$$

Η καμπυλότητα της $f(\mathbf{x})$, κατά μήκος της ευθείας δίδεται ως:

$$\frac{d^2}{d\lambda^2} f(\mathbf{x}) = \frac{d}{d\lambda} \frac{d}{d\lambda} f(\mathbf{x}) = \mathbf{s}^T \nabla ([\nabla f(\mathbf{x})]^T \mathbf{s}) = \mathbf{s}^T [\nabla^2 f(\mathbf{x})] \mathbf{s}$$

Η κλίση συμβολίζεται συνήθως ως $\mathbf{g} = \nabla f$.

Ο Εσσιανός πίνακας συμβολίζεται συνήθως ως $\mathbf{G} = \nabla^2 f$.

Οι παραπάνω ορισμοί για την κλίση και την καμπυλότητα εξαρτώνται από το μέτρο του \mathbf{s} . Για να μην υπάρχει σύγχυση οι ορισμοί συνοδεύονται από την συνθήκη $\|\mathbf{s}\| = 1$, όπου συνήθως το μέτρο που χρησιμοποιείται είναι το L_2 που ορίζεται ως $\|\mathbf{s}\|_2 \equiv \sqrt{\mathbf{s}^T \mathbf{s}}$.

Ορίζουμε το μοναδιαίο διάνυσμα $\mathbf{e}_i^T \equiv (0, 0, \dots, 0, 1, 0, 0, \dots, 0)$, όπου όλες οι συνιστώσες είναι μηδενικές εκτός της συνιστώσας i που είναι μονάδα.

Η γενική μορφή μίας τετραγωνικής συνάρτησης έχει ως:

$$q(\mathbf{x}) = \frac{1}{2} \mathbf{x}^T \mathbf{G} \mathbf{x} + \mathbf{b}^T \mathbf{x} + c$$

όπου \mathbf{G} είναι ένας συμμετρικός πίνακας. Η παραπάνω σχέση μπορεί πάντα να γραφτεί ως:

$$q(\mathbf{x}) = \frac{1}{2}(\mathbf{x} - \mathbf{x}')^T \mathbf{G}(\mathbf{x} - \mathbf{x}') + c', \quad \text{όπου } \mathbf{G}\mathbf{x}' = -\mathbf{b}, \quad c' = c - \frac{1}{2}\mathbf{x}'^T \mathbf{G}\mathbf{x}'.$$

Επίσης:

$$\nabla q(\mathbf{x}) = \frac{1}{2}(\mathbf{G} + \mathbf{G}^T)\mathbf{x} + \mathbf{b} = \mathbf{G}\mathbf{x} + \mathbf{b}.$$

Ο κανόνας διαφόρισης γινομένου είναι:

$$\nabla(\mathbf{u}^T \mathbf{v}) = (\nabla \mathbf{u}^T)\mathbf{v} + (\nabla \mathbf{v}^T)\mathbf{u}$$

Σε αναλυτική μορφή όπου φαίνονται καθαρά οι συνιστώσες, έχουμε:

$$\begin{aligned} q(\mathbf{x}) &= \frac{1}{2} \sum_{i,j} x_i G_{ij} x_j + \sum_i b_i x_i + c \\ \frac{\partial}{\partial x_k} q(\mathbf{x}) &= \frac{1}{2} \sum_{i,j} \{ \delta_{ik} G_{ij} x_j + x_i G_{ij} \delta_{jk} \} + \sum_i b_i \delta_{ik} = \frac{1}{2} \sum_j G_{kj} x_j + \frac{1}{2} \sum_i x_i G_{ik} + b_k \\ &= \frac{1}{2} \sum_j \{ G_{kj} x_j + G_{kj}^T x_j \} + b_k = \sum_j G_{kj} x_j + b_k \end{aligned}$$

Η ποσότητα $\delta_{ij} \equiv \begin{cases} 1, & i=j \\ 0, & i \neq j \end{cases}$ ονομάζεται "Δέλτα του Kronecker".

Ομοίως $\nabla^2 q(\mathbf{x}) = \mathbf{G}$, δηλαδή: $\frac{\partial^2}{\partial x_i \partial x_j} q(\mathbf{x}) = G_{ij}$.

Εάν \mathbf{x}' και \mathbf{x}'' είναι δύο δεδομένα σημεία και εάν $\mathbf{g}' = \nabla q(\mathbf{x}')$ και $\mathbf{g}'' = \nabla q(\mathbf{x}'')$, τότε έχουμε:

$$\mathbf{g}'' - \mathbf{g}' = \nabla(q(\mathbf{x}'') - q(\mathbf{x}')) = \mathbf{G}(\mathbf{x}'' - \mathbf{x}').$$

Βλέπουμε δηλαδή ότι για τετραγωνικές συναρτήσεις, η διαφορά στις θέσεις απεικονίζεται μέσω του Εσσιανού πίνακα σε διαφορά των κλίσεων.

Ένα απαραίτητο εργαλείο για προσεγγίσεις είναι η ανάπτυξη κατά Taylor. Για συναρτήσεις μιάς μεταβλητής έχουμε:

$$f(x + \delta) = f(x) + \delta \frac{d}{dx} f(x) + \frac{1}{2!} \delta^2 \frac{d^2}{dx^2} f(x) + \dots + \frac{1}{n!} \delta^n \frac{d^n}{dx^n} f(x) + \dots$$

Για συναρτήσεις πολλών μεταβλητών έχουμε αντιστοίχως:

$$f(\mathbf{x} + \mathbf{h}) = f(\mathbf{x}) + \sum_{i=1,n} h_i \frac{\partial}{\partial x_i} f(\mathbf{x}) + \frac{1}{2!} \sum_{\substack{i=1,n \\ j=1,n}} h_i h_j \frac{\partial^2}{\partial x_i \partial x_j} f(\mathbf{x}) + \dots$$

Ισοδύναμα σε διανυσματική μορφή γράφουμε:

$$f(\mathbf{x} + \mathbf{h}) = f(\mathbf{x}) + \mathbf{h}^T \nabla f(\mathbf{x}) + \frac{1}{2!} \mathbf{h}^T [\nabla^2 f(\mathbf{x})] \mathbf{h} + \dots$$

Εάν θεωρήσουμε αμελητέους τους όρους υψηλότερου βαθμού στο $\|\mathbf{h}\|$, τότε βλέπουμε ότι για $\|\mathbf{h}\| \rightarrow 0$, η $f(\mathbf{x})$ προσεγγίζεται από μία τετραγωνική συνάρτηση στην περιοχή του \mathbf{x} :

$$f(\mathbf{x} + \mathbf{h}) \approx q(\mathbf{h}) = f(\mathbf{x}) + \mathbf{h}^T \nabla f(\mathbf{x}) + \frac{1}{2!} \mathbf{h}^T [\nabla^2 f(\mathbf{x})] \mathbf{h}$$

Οι τετραγωνικές συναρτήσεις θα δούμε ότι παίζουν σημαντικό ρόλο στην θεωρία της βελτιστοποίησης και ορισμένοι από τους πιο επιτυχείς αλγόριθμους βασίζονται σε τετραγωνικά πρότυπα.

ΠΕΡΙΓΡΑΦΗ ΤΟΥ ΠΡΟΒΛΗΜΑΤΟΣ

Η συνήθης διατύπωση του προβλήματος είναι:

$$\text{Ελαχιστοποιήστε την } f(x), \quad x \in R^n$$

Στο εξής θα αναφερόμαστε στον όρο ελαχιστοποίηση μόνον. Σημειώστε ότι: $\max\{f(x)\} = -\min\{-f(x)\}$ και συνεπώς τα προβλήματα της "μεγιστοποίησης" ανάγονται άμεσα σε προβλήματα ελαχιστοποίησης.

Η υπό ελαχιστοποίηση συνάρτηση $f(x)$ καλείται "**αντικειμενική**" συνάρτηση, το σημείο δε που καθορίζει την θέση του ελαχίστου συμβολίζεται ως x^* και καλείται "**ελαχιστοποιητής**".

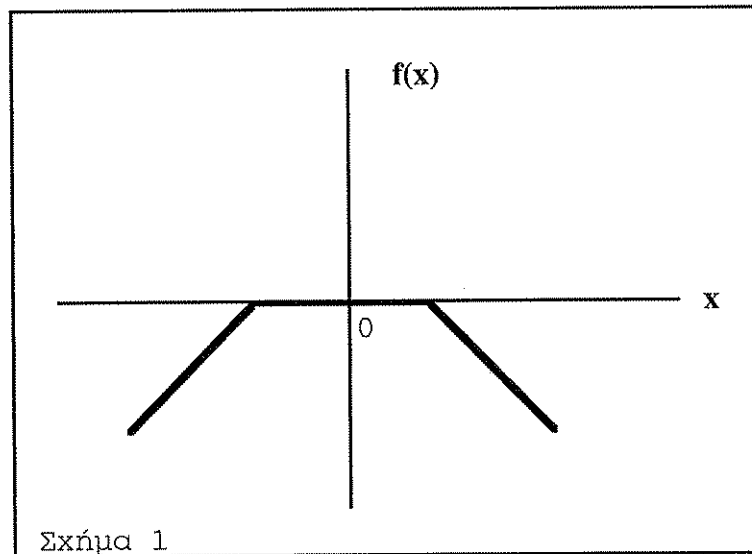
Μια συνάρτηση είναι δυνατόν να έχει περισσότερα του ενός "**τοπικά ελάχιστα**". Το τοπικό ελάχιστο που αντιστοιχεί στην μικρότερη τιμή της αντικειμενικής συνάρτησης, λέγεται "**γενικό ή απόλυτο ελάχιστο**". Οι τεχνικές που θα περιγράψουμε αφορούν την εύρεση τοπικών ελαχίστων. Μέθοδοι εύρεσης του γενικού ελαχίστου είναι πολύπλοκες, η δε αντίστοιχη μαθηματική θεωρία δεν είναι επαρκής για την υποστήριξη αλγορίθμων που θα εγγυώνται την εντόπισή του και παραμένει ένα πολύ ενδιαφέρον ανοικτό ερευνητικό θέμα.

Ο ορισμός του ελαχίστου παρουσιάζει ορισμένες δυσκολίες. Εάν χρησιμοποιηθεί ο κανόνας:

$$f(x) \geq f(x^*) \quad \forall x \text{ στην περιοχή του } x^*$$

τότε επιτρέπονται καταστάσεις σαν αυτή που φαίνεται στο σχήμα 1 παρακάτω.

Σημειώσεις: Τεχνικές βελτιστοποίησης



Παρατηρείστε ότι το σημείο $x=0$, είναι ταυτόχρονα τοπικό ελάχιστο αλλά και γενικό μέγιστο. !!!

Ένας πιο αυστηρός ορισμός που αποφεύγει την παραπάνω δυσκολία είναι:

$$f(x) > f(x^*) \quad \forall x \text{ στην περιοχή του } x^*$$

Ένας ακόμη πιο αυστηρός ορισμός είναι αυτός του απομονωμένου ελαχίστου, ο οποίος επιπλέον απαιτεί για το x^* να είναι ο μοναδικός ελαχιστοποιητής στην περιοχή του.

Δυσκολίες επίσης παρουσιάζονται όταν η αντικειμενική συνάρτηση $f(x)$ δεν είναι ομαλή. (π.χ. όταν η παράγωγός της είναι ασυνεχής). Στην συνέχεια θα ασχοληθούμε με μεθόδους που υποθέτουν ότι η $f(x)$ είναι συνεχής και ότι έχει συνεχείς πρώτες και δεύτερες παραγώγους.

Υιοθετούνται οι παρακάτω συντομίες:

$$f^* = f(x^*), \quad g^* = g(x^*), \quad G^* = G(x^*)$$

Οι μέθοδοι που θα παρουσιαστούν είναι επαναληπτικές. Δηλαδή δημιουργούν μια ακολουθία σημείων $\mathbf{x}^{(k)}$ η οποία συγκλίνει στο \mathbf{x}^* . Οι συντομογραφίες που υιοθετούνται είναι αντίστοιχως:

$$f^{(k)} = f(\mathbf{x}^{(k)}), \mathbf{g}^{(k)} = \mathbf{g}(\mathbf{x}^{(k)}), \mathbf{G}^{(k)} = \mathbf{G}(\mathbf{x}^{(k)})$$

ΣΥΝΘΗΚΕΣ ΓΙΑ ΤΗΝ ΥΠΑΡΞΗ ΕΛΑΧΙΣΤΟΥ

Αναγκαίες συνθήκες:

- a) Πρώτης τάξεως: $\mathbf{g}^* = 0$
b) Δευτέρας τάξεως: $\mathbf{s}^T \mathbf{G}^* \mathbf{s} \geq 0, \forall \mathbf{s}$

Ικανές συνθήκες:

- a) Πρώτης τάξεως: $\mathbf{g}^* = 0$
b) Δευτέρας τάξεως: $\mathbf{s}^T \mathbf{G}^* \mathbf{s} > 0, \forall \mathbf{s} \neq 0$

Η δεύτερη συνθήκη διατυπώνεται εναλλακτικά και ισοδύναμα με τους εξής τρόπους:

- i) Ο Εσσιανός πίνακας \mathbf{G}^* είναι θετικά ορισμένος
- ii) Όλες οι ιδιοτιμές του \mathbf{G}^* είναι θετικές
- iii) Οι παράγοντες $\mathbf{L}\mathbf{L}^T$ (Choleski) υπάρχουν με $L_{ii} > 0$
- iv) Οι παράγοντες $\mathbf{L}\mathbf{D}\mathbf{L}^T$ υπάρχουν με $L_{ii} = 1$ και $D_{ii} > 0$

Οι δύο παραπάνω ικανές συνθήκες εγγυώνται την ύπαρξη αυστηρού και απομονωμένου ελαχίστου, όπως αποδύκνεται στην συνέχεια. Έχουμε:

$$f(\mathbf{x}^* + \delta) = f^* + \frac{1}{2} \delta^T \mathbf{G}^* \delta + o(\delta^T \delta)$$

Η δεύτερη συνθήκη εγγυάται ότι: $\delta^T \mathbf{G}^* \delta \geq \alpha \delta^T \delta, \alpha > 0$

(όπου $\alpha > 0$, η ελαχίστη των ιδιοτιμών του \mathbf{G}^*).

Στο όριο $\delta \rightarrow 0$: $f(\mathbf{x}^* + \delta) \geq f^* + \frac{1}{2}\alpha\delta^T\delta > f^*$ και άρα το \mathbf{x}^* είναι αυστηρός ελαχιστοποιητής. Η σειρά Taylor για το \mathbf{g} δίνει:

$$\delta^T \mathbf{g}(\mathbf{x}^* + \delta) = \delta^T \mathbf{g}^* + \delta^T \mathbf{G}^* \delta > 0$$

Συνεπάγεται ότι δεν υπάρχει $\delta \neq 0$ τέτοιο ώστε $f(\mathbf{x}^* + \delta) = 0$ και άρα το \mathbf{x}^* είναι απομονωμένος ελαχιστοποιητής.

Η συνθήκη $\mathbf{g}(\mathbf{x}^*) = 0$ δεν εξασφαλίζει ύπαρξη ελαχίστου. Το σημείο που την ικανοποιεί καλείται "στάσιμο σημείο". Εάν δε ταυτόχρονα ο Εσσιανός πίνακας \mathbf{G}^* είναι:

- 1) Θετικά ορισμένος, τότε το σημείο είναι ελάχιστο
- 11) Αρνητικά ορισμένος, τότε το σημείο είναι μέγιστο
- 111) Αόριστος¹, τότε το σημείο λέγεται σαγμοειδές.

Το γενικότερο πρόβλημα,

Ελαχιστοποιήστε την $f(\mathbf{x})$, $\mathbf{x} \in \mathbb{R}^n$

υπό τις συνθήκες: $\varphi_i(\mathbf{x}) = 0, i = 1, 2, \dots, k$
 $\psi_i(\mathbf{x}) \geq 0, i = 1, 2, \dots, m$

ονομάζεται ελαχιστοποίηση με περιορισμούς (ή υπό συνθήκες).

Αποτελεί ένα ανοικτό ερευνητικό θέμα και έχει πολλές και σημαντικές εφαρμογές.

Εδώ θα ασχοληθούμε μόνον με μεθόδους ελαχιστοποίησης άνευ περιορισμών.

¹ Έχει ιδιοτιμές αμφοτέρων προσήμων

ΣΤΟΙΧΕΙΑ ΑΠΟ ΤΗΝ ΑΡΙΘΜΗΤΙΚΗ ΓΡΑΜΜΙΚΗ ΑΛΓΕΒΡΑ

Επίλυση γραμμικών συστημάτων.

Εστω ότι θέλουμε να λύσουμε το πρόβλημα: $\mathbf{Ax}=\mathbf{b}$, όπου \mathbf{A} είναι ένας τετραγωνικός πίνακας $n \times n$, και \mathbf{x} , \mathbf{b} ανύσματα στον n -διάστατο χώρο.

1) Διαγώνια συστήματα.

Εάν ο πίνακας \mathbf{A} είναι διαγώνιος το σύστημα γράφεται ως:

$$\begin{bmatrix} \alpha_{11} & 0 & \dots & 0 \\ 0 & \alpha_{22} & \dots & 0 \\ \dots & \dots & \dots & 0 \\ 0 & 0 & \dots & \alpha_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \dots \\ b_n \end{bmatrix}$$

η λύση του οποίου είναι ²: $x_i = \frac{b_i}{\alpha_{ii}}$, $i=1,2,\dots,n$.

ii) Κάτω τριγωνικά συστήματα.

Εάν ο πίνακας είναι κάτω τριγωνικός, το σύστημα γράφεται ως:

$$\begin{bmatrix} \alpha_{11} & 0 & \dots & 0 \\ \alpha_{21} & \alpha_{22} & \dots & 0 \\ \dots & \dots & \dots & 0 \\ \alpha_{n1} & \alpha_{n2} & \dots & \alpha_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \dots \\ b_n \end{bmatrix}$$

Υποθέτουμε ότι όλα τα $\alpha_{ii} \neq 0$, ώστε το σύστημα να έχει λύση. Η πρώτη εξίσωση δίνει: $x_1 = \frac{b_1}{\alpha_{11}}$. Η επόμενη λύνεται άμεσα αφού πρώτα αντικατασταθεί η τιμή για το x_1 , κοκ. Η διαδικασία

² Τι γίνεται όταν κάποιο από τα διαγώνια στοιχεία είναι μηδέν ;

αυτή ονομάζεται "εμπρόσθια αντικατάσταση" και ο αλγόριθμος που ακολουθείται έχει ως εξής:

Γιά κάθε $i=1,2,\dots,n$

$$\text{θέστε: } x_i = \frac{b_i - \sum_{j=1}^{i-1} \alpha_{ij}x_j}{\alpha_{ii}}$$

Τέλος

όπου το άθροισμα $\sum_{j=k}^m y_j$ θεωρείται μηδενικό, όταν $m < k$.

iii) Άνω τριγωνικά συστήματα.

Αντίστοιχα παρουσιάζουμε τον αλγόριθμο για την επίλυση άνω τριγωνικών συστημάτων.

Γιά κάθε $i=n,n-1,\dots,1$

$$\text{θέστε: } x_i = \frac{b_i - \sum_{j=i+1}^n \alpha_{ij}x_j}{\alpha_{ii}}$$

Τέλος

Η ως άνω διαδικασία ονομάζεται "οπίσθια αντικατάσταση".

Σημειώσεις: Τεχνικές βελτιστοποίησης

Παραγοντοποίηση πινάκων

Έστω ότι ο πίνακας \mathbf{A} μπορεί να παραγοντοποιηθεί ως $\mathbf{A}=\mathbf{LU}$, όπου \mathbf{L} είναι κάτω τριγωνικός και \mathbf{U} είναι άνω τριγωνικός. Τότε το σύστημα $\mathbf{Ax}=\mathbf{b}$ λύνεται σε δύο φάσεις ως εξής:

1) Το σύστημα: $\mathbf{Lz}=\mathbf{b}$, λύνεται και υπολογίζεται το \mathbf{z} .

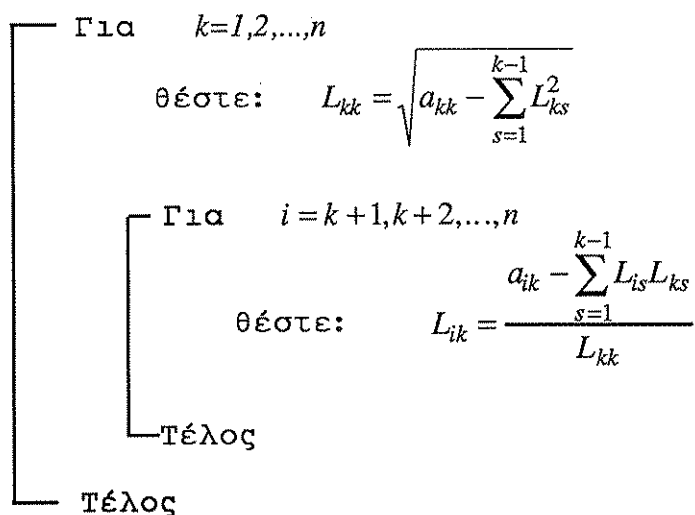
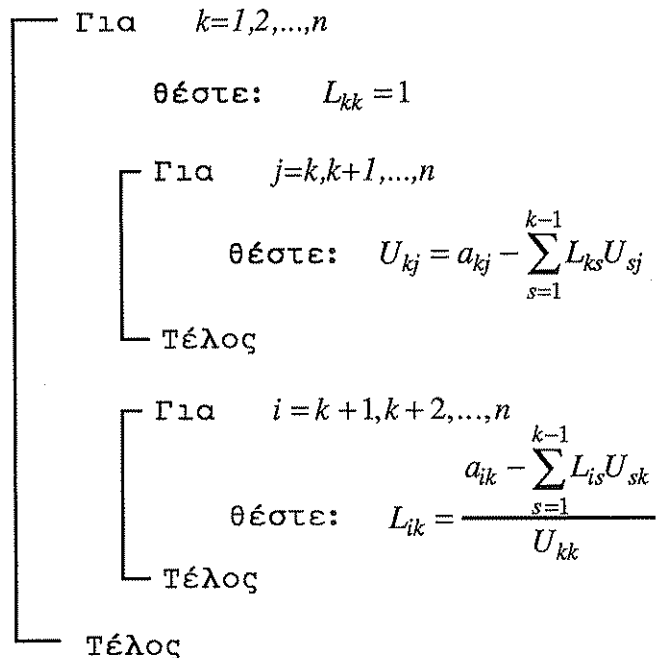
11) Το σύστημα: $\mathbf{Ux}=\mathbf{z}$, λύνεται και υπολογίζεται το \mathbf{x} .

Το πλεονέκτημα είναι ότι τα τριγωνικά συστήματα λύνονται εύκολα. Αποδुकνεύεται ότι ένας πίνακας \mathbf{A} μπορεί να γραφτεί ως $\mathbf{A}=\mathbf{LU}$, εφόσον όλοι οι "κυρίως ελλάσσωνες" (principal minors) του \mathbf{A} έχουν μη-μηδενική ορίζουσα.

Ο k -κυρίως ελλάσσων πίνακας του \mathbf{A} ορίζεται ως ο $k \times k$ πίνακας:

$$\mathbf{A}_k = \begin{bmatrix} \alpha_{11} & \alpha_{12} & \dots & \alpha_{1k} \\ \alpha_{21} & \alpha_{22} & \dots & \alpha_{2k} \\ \dots & \dots & \dots & \dots \\ \alpha_{k1} & \alpha_{k2} & \dots & \alpha_{kk} \end{bmatrix}, \text{ δηλαδή } \mathbf{A}_1 = \alpha_{11}, \mathbf{A}_2 = \begin{bmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{bmatrix} \text{ κλπ}$$

Επίσης αποδुकνεύεται ότι εάν ο πίνακας \mathbf{A} είναι πραγματικός, συμμετρικός και θετικά ορισμένος, τότε παραγοντοποιείται ως: $\mathbf{A}=\mathbf{L}^T\mathbf{L}$, όπου ο \mathbf{L} είναι κάτω τριγωνικός με θετικά διαγώνια στοιχεία. Η παραγοντοποίηση αυτή ονομάζεται Choleski. Στην γενική περίπτωση που η παραγοντοποίηση $\mathbf{A}=\mathbf{LU}$ είναι δυνατή, μπορούμε να διαλέξουμε τις τιμές των διαγωνίων στοιχείων του \mathbf{L} ή του \mathbf{U} εκ των προτέρων. Οι συνήθειες προτιμήσεις είναι $L_{ii}=1$ (Doolittle) ή $U_{ii}=1$ (Crout). Παραθέτουμε τους αλγόριθμους για τις παραγοντοποιήσεις Doolittle και Choleski.



Εάν ο \mathbf{A} μπορεί να γραφτεί ως $\mathbf{A}=\mathbf{LU}$ και επιπλέον είναι συμμετρικός, τότε μπορεί να γραφτεί ως $\mathbf{A}=\mathbf{LDL}^T$, όπου \mathbf{L} κατω τριγωνικός με μονάδες στην διαγώνιο και \mathbf{D} διαγώνιος.

ΜΕΘΟΔΟΙ ΜΟΝΟΔΙΑΣΤΑΤΗΣ ΕΛΑΧΙΣΤΟΠΟΙΗΣΗΣ

Η μονοδιάστατη ελαχιστοποίηση απαντάται συνήθως ως επιμέρους πρόβλημα μεθόδων που αντιμετωπίζουν το γενικότερο πολυδιάστατο πρόβλημα.

Σκιαγραφούμε παρακάτω έναν επαναληπτικό αλγόριθμο πάνω στον οποίο βασίζεται μία πλειάδα μεθόδων πολυδιάστατης ελαχιστοποίησης για δείξουμε πώς εμφανίζεται η ανάγκη επίλυσης μονοδιάστατων προβλημάτων.

Κατά την επανάληψη (k) είναι δοθέν ένα σημείο $x^{(k)}$.

A) Υπολογίζεται μια διεύθυνση $p^{(k)}$ ¹

B) Υπολογίζεται το ελάχιστο της συνάρτησης κατά μήκος της ευθείας που περνά από το σημείο $x^{(k)}$ και είναι παράλληλη προς την διεύθυνση $p^{(k)}$. Δηλαδή επιλύεται το πρόβλημα:

Ελαχιστοποίηση ως προς λ της: $f(x^{(k)} + \lambda p^{(k)})$ (εύρεση του λ^*)

Γ) Αντικατάσταση: $x^{(k+1)} = x^{(k)} + \lambda p^{(k)}$ και η διαδικασία επαναλαμβάνεται έως ότου ικανοποιηθούν κάποια κριτήρια τερματισμού.

Το επιμέρους πρόβλημα στο βήμα B), γνωστό με τον τίτλο "Γραμμική αναζήτηση", είναι ένα πρόβλημα μονοδιάστατης ελαχιστοποίησης.

Ένα διάστημα $[a,b]$ που φράσει την θέση του ελαχίστου x^* ονομάζεται διάστημα αβεβαιότητας. Το εύρος του διαστήματος

¹ Ο τρόπος υπολογισμού της διεύθυνσης αυτής ποικίλει.

ορίζεται ως: $d \equiv b - a$. Εάν το εύρος του διαστήματος αβεβαιότητας είναι μικρό, τότε ήδη έχουμε μια καλή εκτίμηση για την θέση του ελαχίστου $x^* \approx \frac{b+a}{2} \pm \frac{d}{2}$. Πάνω σε αυτήν την

απλή παρατήρηση βασίζεται ένα σύνολο μεθόδων οι οποίες έχουν ως στόχο να συρρικνώνουν διαδοχικά το διάστημα αβεβαιότητας.

Θα θεωρήσουμε στην συνέχεια μεθόδους που λειτουργούν υποθέτοντας ότι η αντικειμενική συνάρτηση είναι μονότροπη (unimodal) στο διάστημα $[a, b]$.

Μιά συνάρτηση $f(x)$ λέγεται ότι είναι μονότροπη στο διάστημα $[a, b]$, εάν υπάρχει ένα μοναδικό σημείο $x^* \in [a, b]$ τέτοιο ώστε η $f(x)$ να είναι γνησίως αύξουσα αυξανομένης της απόστασης $|x - x^*|$ από το x^* .

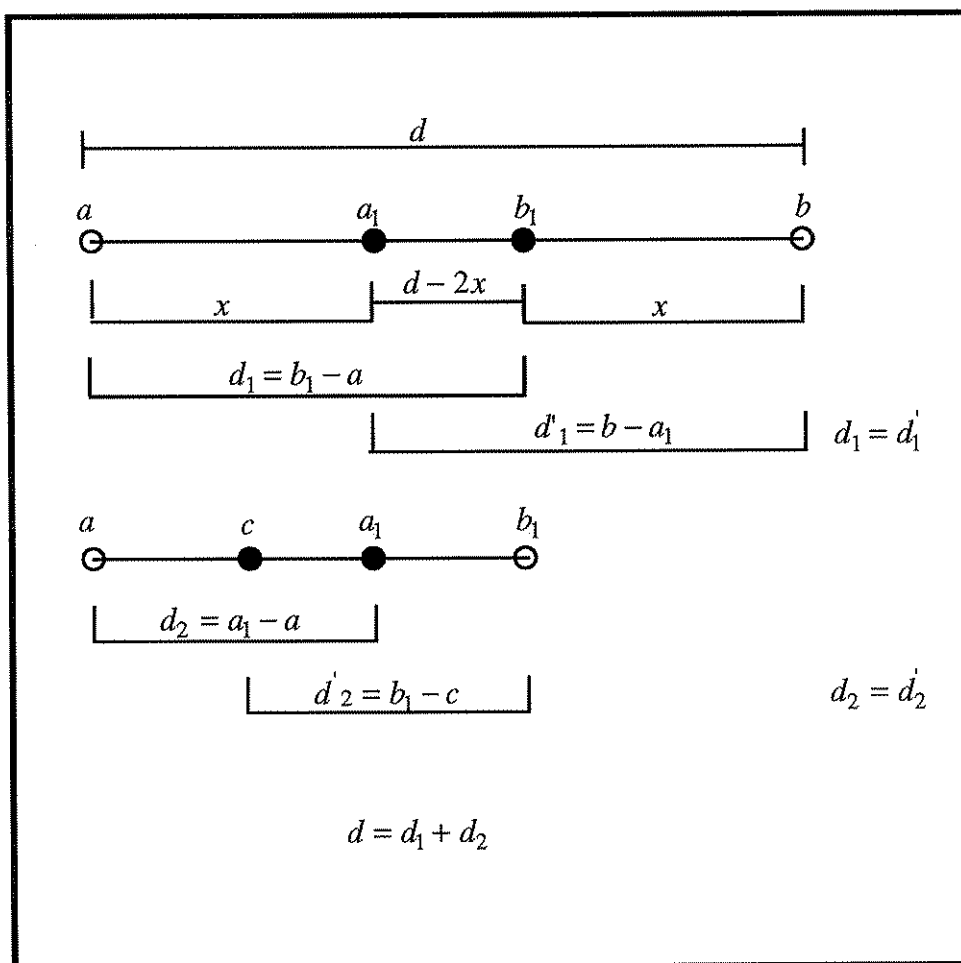
Η απόδοση μιάς μεθόδου μετράται από τον λόγο $r_N \equiv \frac{b_N - a_N}{b - a} \equiv \frac{d_N}{d}$ του τελικού προς το αρχικό εύρος, και N είναι οι φορές που χρειάστηκε να υπολογισθεί η $f(x)$.

Θα μελετήσουμε αρχικά μεθόδους που χρησιμοποιούν μόνον τιμές της συνάρτησης (και όχι των παραγώγων της).

Δεδομένου ενός διαστήματος αβεβαιότητας $[a, b]$, χρειάζονται τουλάχιστον δύο σημεία εντός του διαστήματος αυτού στα οποία πρέπει να υπολογιστεί η συνάρτηση για να είναι δυνατή η αναγωγή του διαστήματος. Έστω $[a_1, b_1] \in [a, b]$, τότε εάν $f(a_1) < f(b_1)$ το ελάχιστο βρίσκεται στο διάστημα $[a, b_1]$, αλλιώς εάν $f(a_1) > f(b_1)$ το ελάχιστο βρίσκεται στο διάστημα $[a_1, b]$. Στην περίπτωση δε που $f(a_1) = f(b_1)$ τότε το διάστημα αβεβαιότητας είναι το $[a_1, b_1]$.

Από το παρακάτω σχήμα φαίνεται πως πρέπει να διαλέξουμε τα ενδιάμεσα σημεία έτσι ώστε τα δύο υποψήφια νέα διαστήματα

αβεβαιότητας να έχουν ίσα εύρη και επιπλέον να ισχύει $d = d_1 + d_2$, δηλαδή ότι το άθροισμα των ευρών δύο διαδοχικών διαστημάτων να ισούται με το εύρος του προηγούμενου διαστήματος. Η απόσταση $x = a_1 - a = b - b_1$ πρέπει να ικανοποιεί



την διπλή ανισότητα: $\frac{d}{3} < x < \frac{d}{2}$

Μεθοδος Fibonacci

Ας υποθέσουμε ότι το προτελευταίο διάστημα εύρους d_{m-1} κωρίζεται σε δύο ίσα διαστήματα, δηλαδή το εύρος του τελευταίου διαστήματος θα είναι αναγκαστικά $d_m = \frac{1}{2}d_{m-1}$ οπότε $d_{m-1} = 2d_m$. Χρησιμοποιώντας το προηγούμενο αποτέλεσμα έχουμε διαδοχικά:

$$d_{m-2} = d_{m-1} + d_m = 3d_m$$

$$d_{m-3} = d_{m-2} + d_{m-1} = 5d_m$$

$$d_{m-4} = d_{m-3} + d_{m-2} = 8d_m$$

και εν γένει: $d_k = \mathcal{F}_{m-k+1}d_m$, όπου $\mathcal{F}_0 = \mathcal{F}_1 = 1$ και για $i = 2, 3, \dots, \infty$ έχουμε: $\mathcal{F}_i = \mathcal{F}_{i-1} + \mathcal{F}_{i-2}$. Η ακολουθία που παράγεται με την παραπάνω αναδρομική σχέση ονομάζεται ακολουθία Fibonacci και μερικοί από τους πρώτους όρους της είναι: 1, 1, 2, 3, 5, 8, 13, 21, ...

Την τελευταία φορά επειδή διχοτομείται το διάστημα δεν γνωρίζουμε ποιό από τα δύο υποδιαστήματα περιέχει το ελάχιστο και πρέπει να υπολογίσουμε την συνάρτηση σε ένα ακόμη σημείο που το παίρνουμε να είναι πολύ κοντά στο μέσο του διαστήματος. Ο λόγος των ευρών δύο διαδοχικών διαστημάτων είναι: $\frac{d_{k+1}}{d_k} = \frac{\mathcal{F}_{m-k}}{\mathcal{F}_{m-k+1}}$ και η απόδοση μετά από m κλήσεις ισούται με $r_m = \frac{1}{\mathcal{F}_m}$.

Εάν λοιπόν απαιτείται αναγωγή του εύρους του διαστήματος αβεβαιότητας τουλάχιστον στο 1% του αρχικού, τότε αφού $\mathcal{F}_{10} = 89$ και $\mathcal{F}_{11} = 144$ πρέπει να διαλέξουμε $m = 11$. Σκιαγραφείται παρακάτω ο αλγόριθμος που ακολουθείται για την υλοποίηση της μεθόδου.

Σημειώσεις: Τεχνικές βελτιστοποίησης

Αλγόριθμος αναζήτησης κατά Fibonacci

Λίδονται ως είσοδος:

Αρχικό διάστημα αβεβαιότητας: a_1, b_1

Συνολικός αριθμός κλήσεων: m

Ενας μικρός αριθμός για την

επιπλέον κλήση την τελευταία φορά: δ

$$c_1 = b_1 - (F_{m-1}/F_m)(b_1 - a_1)$$

$$f_c = f(c_1)$$

$$d_1 = a_1 + (F_{m-1}/F_m)(b_1 - a_1)$$

$$f_d = f(d_1)$$

Επανάληψη για $k=1, 2, \dots, m-2$

Εάν $f_c < f_d$ τότε:

$$a_{k+1} = a_k, b_{k+1} = d_k, d_{k+1} = c_k$$

$$c_{k+1} = b_{k+1} - (F_{m-k-1}/F_{m-k})(b_{k+1} - a_{k+1})$$

Εάν $k = m - 2$ τότε:

$$c_{k+1} = c_{k+1} + \delta$$

Τέλος-Εάν

$$f_d = f_c, f_c = f(c_{k+1})$$

αλλιώς:

$$a_{k+1} = a_k, b_{k+1} = b_k, c_{k+1} = d_k$$

$$d_{k+1} = a_{k+1} - (F_{m-k-1}/F_{m-k})(b_{k+1} - a_{k+1})$$

Εάν $k = m - 2$ τότε:

$$d_{k+1} = d_{k+1} + \delta$$

Τέλος-Εάν

$$f_c = f_d, f_d = f(d_{k+1})$$

Τέλος-Εάν

Τέλος-Επανάληψης

Εάν $f_c < f_d$ τότε:

$$a_m = a_{m-1}, b_m = d_{m-1}$$

αλλιώς:

$$a_m = c_{m-1}, b_m = b_{m-1}$$

Τέλος-Εάν

ΤΕΛΟΣ

Η μέθοδος της "Χρυσής τομής"

Η μέθοδος Fibonacci σε κάθε επανάληψη ανάγει το διάστημα αβεβαιότητας με έναν παράγοντα $\frac{F_i}{F_{i+1}}$.

Οι αριθμοί του Fibonacci δίδονται αναλυτικά από την σχέση:

$$F_i = \frac{1}{\sqrt{5}} \left\{ \left[\frac{1+\sqrt{5}}{2} \right]^{i+1} - \left[\frac{1-\sqrt{5}}{2} \right]^{i+1} \right\}$$

Παρατηρούμε ότι για $i \rightarrow \infty$ έχουμε: $\tau \equiv \lim_{i \rightarrow \infty} \frac{F_i}{F_{i+1}} = \frac{\sqrt{5}-1}{2} \cong 0.618$.

Η μέθοδος της "Χρυσής τομής" ανάγει το διάστημα αβεβαιότητας σε κάθε επανάληψη με τον σταθερό παράγοντα τ (παράγων χρυσής τομής). Παρατηρούμε ότι ισχύει $\tau + \tau^2 = 1$ και άρα έχουμε: $d_{k+1} + d_{k+2} = \tau d_k + \tau d_{k+1} = (\tau + \tau^2) d_k = d_k$, δηλαδή η μέθοδος διατηρεί τον κανόνα για την σχέση μεταξύ των διαδοχικών ευρών όπως και η μέθοδος του Fibonacci. Είναι προφανές ότι $r_m = \tau^{m-1}$. Το πλεονέκτημα της μεθόδου αυτής είναι ότι δεν χρειάζεται να γνωρίζουμε εκ των προτέρων τον ολικό αριθμό των επαναλήψεων (όπως απαιτεί η μέθοδος Fibonacci) και η αναζήτηση μπορεί να διακοπεί σε οποιαδήποτε επανάληψη βάσει οποιουδήποτε επιθυμητού κριτηρίου.

Ασυμπτωτικά είναι δυνατόν να αποδειχθεί ότι η μέθοδος Fibonacci είναι 17% ταχύτερη από την μέθοδο της "Χρυσής τομής" για τον ίδιο (μεγάλο) αριθμό επαναλήψεων.

Ο αλγόριθμος της Χρυσής τομής σκιαγραφείται παρακάτω.

Αλγόριθμος αναζήτησης "Χρυσής τομής"

Δίδονται ως είσοδος:

Αρχικό διάστημα αβεβαιότητας: a_1, b_1

Παράμετρος τερματισμού: tol (μικρός θετικός αριθμός)

$$c_1 = a_1 + (1 - \tau)(b_1 - a_1)$$

$$F_c = F(c_1)$$

$$d_1 = b_1 - (1 - \tau)(b_1 - a_1)$$

$$F_d = F(d_1)$$

Επανάληψη για $k = 1, 2, \dots$

Εάν $F_c < F_d$ τότε:

$$a_{k+1} = a_k, b_{k+1} = d_k, d_{k+1} = c_k$$

$$c_{k+1} = a_{k+1} + (1 - \tau)(b_{k+1} - a_{k+1})$$

$$F_d = F_c, F_c = F(c_{k+1})$$

αλλιώς:

$$a_{k+1} = c_k, b_{k+1} = b_k, c_{k+1} = d_k$$

$$d_{k+1} = b_{k+1} - (1 - \tau)(b_{k+1} - a_{k+1})$$

$$F_c = F_d, F_d = F(d_{k+1})$$

Τέλος-Εάν

Εάν $b_{k+1} - a_{k+1} < tol$ τότε:

Τερματισμός του προγράμματος (stop)

Τέλος-Εάν

Τέλος-Επανάληψης

Η μέθοδος της διχοτόμησης

Μέθοδοι που χρησιμοποιούν τιμές της συνάρτησης όχι μόνον για σύγκριση αλλά και για την δημιουργία μιας τοπικής προσέγγισης (προτύπου) ή που επιπλέον χρησιμοποιούν τιμές της παραγώγου είναι συνήθως περισσότερο αποδοτικές από τις μεθόδους που περιγράψαμε μέχρι τώρα. Εάν η πρώτη παράγωγος χρησιμοποιηθεί και επιπλέον υπάρχει ένα αρχικό διάστημα αβεβαιότητας τότε μιά απλή μέθοδος μπορεί άμεσα να εφαρμοστεί.

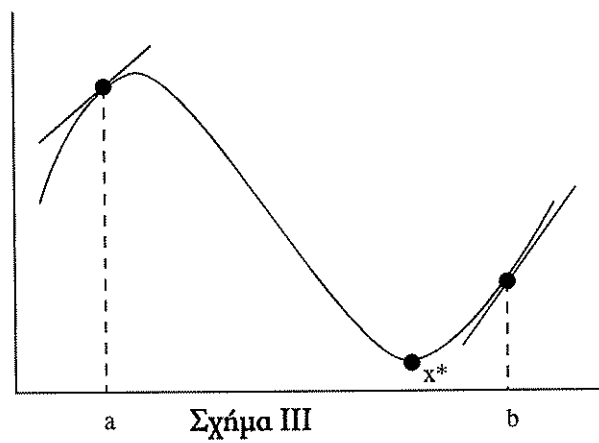
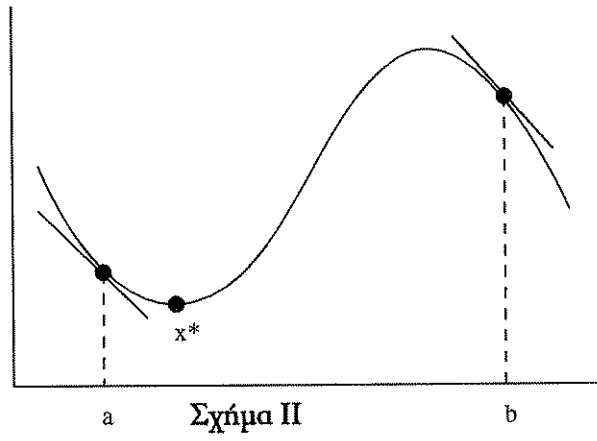
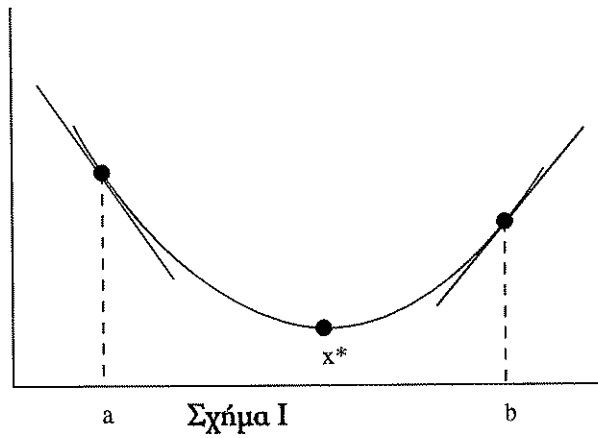
Κατά την k -επανάληψη υπάρχει ένα διάστημα $[a_k, b_k]$ και υπολογίζεται η τιμή της συνάρτησης και της παραγώγου της στο μέσον $c_k = \frac{1}{2}(a_k + b_k)$. Το επόμενο διάστημα $[a_{k+1}, b_{k+1}]$ είναι ένα εκ των $[a_k, c_k]$, $[c_k, b_k]$ το οποίο φράσει το ελάχιστο. Οι ικανές και αναγκαίες συνθήκες ώστε ένα διάστημα $[a, b]$ να φράσει το ελάχιστο δίδονται παρακάτω. (Υποθέτουμε ότι οι τιμές της συνάρτησης και της παραγώγου της στα άκρα του διαστήματος είναι αντίστοιχα F_a, F_b, g_a, g_b).

$$g_a < 0 \text{ και ταυτόχρονα } g_b > 0 \quad (\text{βλέπε σχήμα I})$$

$$\text{ή } g_a < 0 \text{ και ταυτόχρονα } F_b > F_a \quad (\text{βλέπε σχήμα II})$$

$$\text{ή } g_a > 0 \text{ και } g_b > 0 \text{ και } F_b < F_a \quad (\text{βλέπε σχήμα III})$$

Προφανώς ο παράγων αναγωγής του διαστήματος αβεβαιότητας ανά επανάληψη είναι $\frac{1}{2}$. Η μέθοδος της διχοτόμησης δεν είναι απευθείας συγκρίσιμη με τις προηγούμενες, γιατί σε κάθε επανάληψη εκτός από την συνάρτηση υπολογίζεται και η παράγωγός της και συνεπώς το (υπολογιστικό) κόστος ανά επανάληψη είναι διαφορετικό.



2.2 ΜΕΘΟΔΟΙ ΠΟΛΥΩΝΥΜΙΚΗΣ ΠΡΟΣΕΓΓΙΣΗΣ

Οι μέθοδοι αυτού του είδους παρεμβάλουν συνήθως ένα πολυώνυμο στις τιμές της συνάρτησης ή και της παραγώγου της και εν συνεχεία το σημείο που ελαχιστοποιεί το παρεμβάλον πολυώνυμο χρησιμοποιείται για την πρόβλεψη του ελαχίστου της συνάρτησης. Επειδή οι περισσότερες συναρτήσεις με "ομαλή συμπεριφορά" προσεγγίζονται ικανοποιητικά εντός μικρών διαστημάτων με πολυώνυμα, οι μέθοδοι αυτού του τύπου αναμένεται να είναι αποδοτικότερες αυτών που μέχρι τώρα περιγράψαν, παρότι οι περιφερειακές αριθμητικές πράξεις που χρειάζονται είναι περισσότερες.

2.2.1 Τετραγωνική παρεμβολή

Στην περίπτωση αυτή το παρεμβάλον πολυώνυμο είναι δευτέρου βαθμού της μορφής: $\Pi(x) = px^2 + qx + r$. Το ελάχιστο του τριωνύμου (υπάρχει εφόσον $p > 0$) βρίσκεται στην θέση $\hat{x} = -\frac{q}{2p}$. Εστω ότι αρχικά υπάρχει ένα διάστημα αβεβαιότητας $[a, b]$ και ένα ενδιάμεσο σημείο c . Θέτοντας $F_a = F(a)$, $F_b = F(b)$, $F_c = F(c)$ έχουμε για την θέση του ελαχίστου:

$$\hat{x} = \frac{1}{2} \frac{(b^2 - c^2)F_a + (c^2 - a^2)F_b + (a^2 - b^2)F_c}{(b - c)F_a + (c - a)F_b + (a - b)F_c}$$

Παρατηρείστε ότι εάν $F_c < F_a$ και $F_c < F_b$ τότε $\hat{x} \in [a, b]$

Η αντικειμενική συνάρτηση υπολογίζεται στο σημείο \hat{x} και επιλέγονται πάλι τρία σημεία (το \hat{x} και δύο από τα a, b, c) τα οποία φράσουν το ελάχιστο. (Σημειώστε ότι δεν είναι απαραίτητο να απορρίπτεται το σημείο που αντιστοιχεί στην υψηλότερη τιμή της συνάρτησης). Ακολουθεί ο αλγόριθμος.

Σημειώσεις: Τεχνικές βελτιστοποίησης

Αλγόριθμος τετραγωνικής παρεμβολής

Δίδονται ως είσοδος:

Αρχικό διάστημα αβεβαιότητας: a_1, b_1, c_1

Παράμετροι τερματισμού: $ftol, xtol$ (μικροί θετικοί αριθμοί)

$F_a = F(a_1), F_b = F(b_1), F_c = F(c_1)$

Επανάληψη για $k = 1, 2, \dots$

$$\hat{x} = \frac{1}{2} \frac{(b^2 - c^2)F_a + (c^2 - a^2)F_b + (a^2 - b^2)F_c}{(b - c)F_a + (c - a)F_b + (a - b)F_c}$$

$F_x = F(\hat{x})$

Εάν $\hat{x} < c_k$ και $F_x < F_c$ τότε:

$$a_{k+1} = a_k, b_{k+1} = c_k, c_{k+1} = \hat{x}, F_b = F_c, F_c = F_x$$

αλλιώς εάν $\hat{x} > c_k$ και $F_x > F_c$ τότε:

$$a_{k+1} = a_k, b_{k+1} = \hat{x}, c_{k+1} = c_k, F_b = F_x$$

αλλιώς εάν $\hat{x} < c_k$ και $F_x > F_c$ τότε:

$$a_{k+1} = \hat{x}, b_{k+1} = b_k, c_{k+1} = c_k, F_a = F_x$$

αλλιώς:

$$a_{k+1} = c_k, b_{k+1} = b_k, c_{k+1} = \hat{x}, F_a = F_c, F_c = F_x$$

Τέλος-Εάν

Εάν $b_{k+1} - a_{k+1} < xtol$ ή $F(c_k) - F(c_{k+1}) < ftol * F(c_k)$ τότε:

Τερματισμός του προγράμματος (stop)

Τέλος-Εάν

Τέλος-Επανάληψης

2.2.2 Κυβική παρεμβολή

Στην περίπτωση αυτή το παρεμβάλον πολυώνυμο είναι τρίτου βαθμού της μορφής $\Pi(x) = px^3 + qx^2 + rx + s$. Για να υπολογίσουμε τους τέσσερεις συντελεστές χρειάζονται τέσσερεις σχέσεις. Συνήθως (αν και όχι απαραίτητα) χρησιμοποιούνται οι τιμές της συνάρτησης και της παραγώγου της στα άκρα $[a, b]$ του διαστήματος αβεβαιότητας. Η θέση του ελαχίστου αυτού του πολυωνύμου είναι μία από τις ρίζες της παραγώγου $\frac{d}{dx}\Pi(x) = 3px^2 + 2qx + r$, που ταυτόχρονα ικανοποιεί την συνθήκη ότι η δεύτερη παράγωγος $\frac{d^2}{dx^2}\Pi(x) = 6px + 2q$ είναι θετική. Εάν F_a, F_b, g_a, g_b είναι οι τιμές της συνάρτησης και της παραγώγου της στα άκρα, τότε το πολυώνυμο έχει το ελάχιστο στην θέση:

$$\hat{x} = a + (b - a) \left\{ 1 - \frac{g_b + v - w}{g_b - g_a + 2v} \right\}$$

όπου $v = \sqrt{w^2 - g_a g_b}$, $w = \frac{3}{b - a} (F_a - F_b) + g_a + g_b$

Το σημείο αυτό αντικαθιστά ένα από τα δύο άκρα και το εύρος του διαστήματος αβεβαιότητας ελαττώνεται.

Ο αλγόριθμος έχει ως εξής.

Αλγόριθμος κυβικής παρεμβολής

Λίδονται ως είσοδος:

Αρχικό διάστημα αβεβαιότητας: a_1, b_1

Παράμετροι τερματισμού: $gtol, xtol$ (μικροί θετικοί αριθμοί)

$$F_a = F(a_1), F_b = F(b_1), g_a = g(a_1), g_b = g(b_1)$$

Επανάληψη για $k = 1, 2, \dots$

$$w = \frac{3}{b_k - a_k} (F_a - F_b) + g_a + g_b$$

$$v = \sqrt{w^2 - g_a g_b}$$

$$\hat{x} = a_k + (b_k - a_k) \left\{ 1 - \frac{g_b + v - w}{g_b - g_a + 2v} \right\}$$

$$F_x = F(\hat{x}), g_x = g(\hat{x})$$

Εάν $g_a < 0$ και ($g_x > 0$ ή $F_x > F_a$) τότε:

$$a_{k+1} = a_k, b_{k+1} = \hat{x}, F_b = F_x, g_b = g_x$$

αλλιώς:

$$a_{k+1} = \hat{x}, b_{k+1} = b_k, F_a = F_x, g_a = g_x$$

Τέλος-Εάν

Εάν $b_{k+1} - a_{k+1} < xtol$ ή $\min\{|g_a|, |g_b|\} < gtol$ τότε:

Τερματισμός του προγράμματος (stop)

Τέλος-Εάν

Τέλος-Επανάληψης

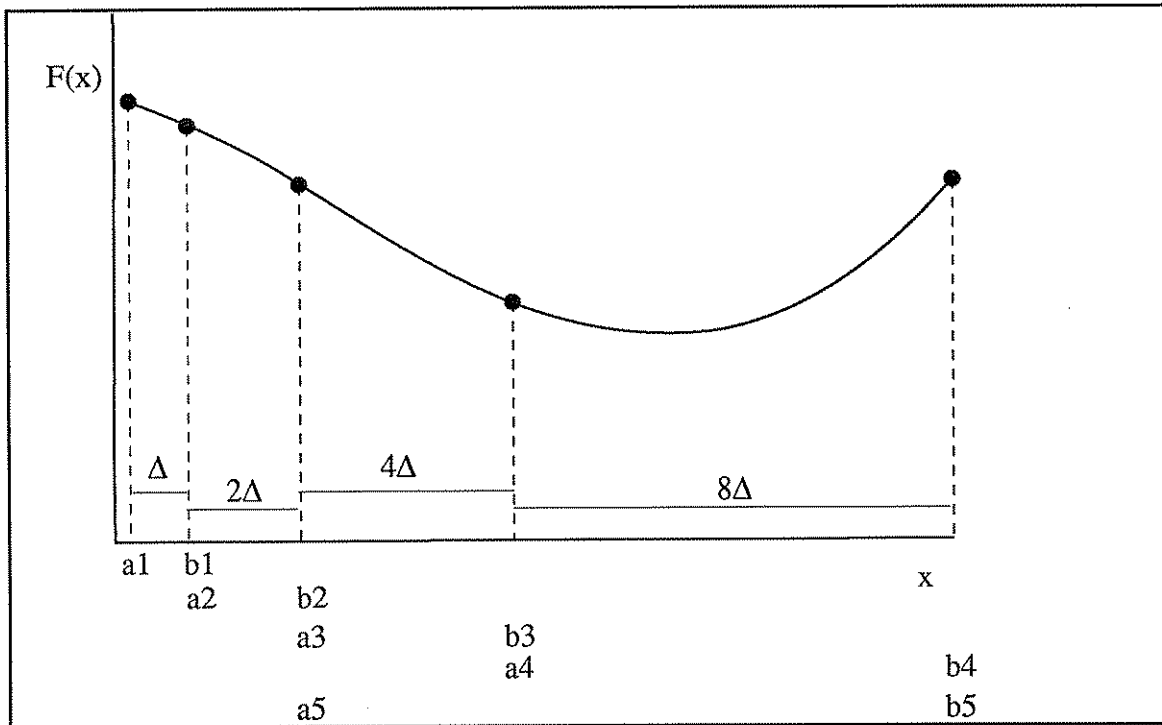
2.3 Εντοπισμός διαστημάτων αβεβαιότητας

Στις μεθόδους που περιγράψαμε, υποθέσαμε ότι υπήρχε αρχικά ένα διάστημα αβεβαιότητας, δηλαδή ένα διάστημα $[a,b]$ που περιείχε την θέση ενός ελαχίστου. Το διάστημα αυτό πρέπει να ευρεθεί πρώτα ώστε στην συνέχεια να μπορούν να εφαρμοστούν οι πιο πάνω τεχνικές. Περιγράφουμε ορισμένες μεθόδους για τον εντοπισμό διαστημάτων αβεβαιότητας.

2.3.1 Μέθοδοι σύγκρισης τιμών

Οι μέθοδοι αυτού του είδους απαιτούν ένα αρχικό σημείο a_1 και ένα βήμα Δ . Υπολογίζεται πρώτα το σημείο $b_1 = a_1 + \Delta$. Εφόσον η τιμή της συνάρτησης στο σημείο αυτό είναι χαμηλότερη το επόμενο δόκιμο σημείο είναι το $b_2 = a_2 + \Delta$ με $a_2 = b_1$.

Η διαδικασία συνεχίζεται όπως φαίνεται και στο παρακάτω σχήμα, έως ότου η τιμή της συνάρτησης αρχίζει να αυξάνει, οπότε το διάστημα αβεβαιότητας δίδεται από την ένωση των δύο τελευταίων διαστημάτων. Δεν θα δώσουμε λεπτομερή αλγοριθμική περιγραφή, απλά σημειώνουμε ότι υπάρχουν πολυάριθμες παραλλαγές που βασίζονται στην παραπάνω μεθοδολογία. Πολλές από τις παραλλαγές αυτές εισάγουν ένα πάνω φράγμα για την τιμή του βήματος, άλλες δεν διπλασιάζουν απλώς το βήμα αλλά το πολλαπλασιάζουν με ένα παράγοντα μεγαλύτερο του δύο κλπ. Εάν η συνάρτηση αυξάνει στο πρώτο βήμα, τότε η διαδικασία αλλάζει φορά. Δηλαδή ψάχνει στην αντίθετη κατεύθυνση.



2.3.2 Μέθοδοι πολυωνυμικής πρόβλεψης

Οι μέθοδοι αυτού του είδους προσπαθούν να προσεγγίσουν την μορφή της αντικειμενικής συνάρτησης με ένα πολυώνυμο. Στην συνέχεια η θέση του ελαχίστου του προσεγγίζοντος πολυωνύμου χρησιμοποιείται για την βελτίωση της προσέγγισης κ.ο.κ.

Θα περιγράψουμε μία μέθοδο που χρησιμοποιεί δευτεροβάθμια τριώνυμα για την τοπική προσέγγιση της συνάρτησης και η οποία έχει αναπτυχθεί από τον M. J. D. Powell*.

* Powell, M. J. D. 'An efficient method of finding the minimum of a function of several variables without calculating derivatives', Computer J. (1964) 7, 155-162.

Χρησιμοποιούνται οι τιμές της συνάρτησης F_a, F_b, F_c σε τρία σημεία a, b, c και υπολογίζεται η θέση του ακρότατου του τριωνύμου $\Pi(x) = px^2 + qx + r$ που προσεγγίζει τοπικά την συνάρτηση. Ελέγχεται το είδος του ακρότατου (ελάχιστο ή μέγιστο) υπολογίζοντας το σημείο της δεύτερης παραγώγου του τριωνύμου δηλαδή το σημείο του συντελεστού p . Εκτελώντας τις απαραίτητες πράξεις βρίσκουμε την παρακάτω έκφραση:

$$p = \frac{(c-b)F_a + (a-c)F_b + (b-a)F_c}{(b-c)(c-a)(a-b)}$$

Ο αλγόριθμος χρησιμοποιεί ένα αρχικό σημείο a_1 , ένα αρχικό βήμα Δ και ένα πάνω φράγμα για την μέγιστη δυνατή τιμή του βήματος Δ_{\max} . Δεν γίνεται καμμία υπόθεση για την κατεύθυνση κατά την φορά της οποίας η συνάρτηση είναι φθίνουσα. Τα σημεία διαλέγονται έτσι ώστε πάντα να ισχύει $a_k < c_k < b_k$. Στην περίπτωση που το τριώνυμο δεν έχει ελάχιστο τότε το σημείο με την χαμηλότερη τιμή αντικαθίσταται με ένα άλλο που απέχει από αυτό κατά Δ_{\max} στην κατεύθυνση που φθίνει η συνάρτηση. Επίσης εάν το ελάχιστο του τριωνύμου απέχει περισσότερο από Δ_{\max} από το σημείο με την χαμηλότερη τιμή, τότε το νέο σημείο που χρησιμοποιείται είναι ένα σημείο που απέχει κατά Δ_{\max} από το σημείο με την χαμηλότερη τιμή και καταργείται το σημείο με την υψηλότερη τιμή. Οι επαναλήψεις σταματούν όταν στο ενδιάμεσο σημείο (c_k) η αντικειμενική συνάρτηση έχει τιμή χαμηλότερη από ότι στα άλλα δύο ακραία σημεία (a_k, b_k).

Ο αλγόριθμος διατυπώνεται στην συνέχεια.

Αλγόριθμος Powell εντοπισμού διαστήματος αβεβαιότητας

Δίδονται ως είσοδος: $a_1, \Delta, \Delta_{\max}$

$$c_1 = a_1 + \Delta, F_a = F(a_1), F_c = F(c_1)$$

Εάν $F_a > F_c$ τότε:

$$b_1 = a_1 + 2\Delta, F_b = F(b_1), t_0 = +1$$

αλλιώς:

$$b_1 = c_1, c_1 = a_1, a_1 = a_1 - \Delta$$

$$F_b = F_c, F_c = F_a, F_a = F(a_1), t_0 = -1$$

Τέλος-Εάν

Επανάληψη για $k = 1, 2, \dots$

$$p = \frac{(c_k - b_k)F_a + (a_k - c_k)F_b + (b_k - a_k)F_c}{(b_k - c_k)(c_k - a_k)(a_k - b_k)}$$

Εάν $p > 0$ τότε:

$$\hat{x} = \frac{1}{2} \frac{(b_k^2 - c_k^2)F_a + (c_k^2 - a_k^2)F_b + (a_k^2 - b_k^2)F_c}{(b_k - c_k)F_a + (c_k - a_k)F_b + (a_k - b_k)F_c}$$

Τέλος-Εάν

Εάν $t_0 = +1$ τότε:

Εάν $p \leq 0$ τότε:

$$a_{k+1} = a_k, b_{k+1} = b_k + \Delta_{\max}, c_{k+1} = c_k$$

$$F_b = F(b_{k+1})$$

αλλιώς:

Εάν $\hat{x} - b_k > \Delta_{\max}$ τότε:

$$b_{k+1} = b_k + \Delta_{\max}$$

αλλιώς:

$$b_{k+1} = \hat{x}$$

Τέλος-Εάν

$$a_{k+1} = c_k, \quad c_{k+1} = b_k$$

$$F_a = F_c, \quad F_c = F_b, \quad F_b = F(b_{k+1})$$

Τέλος-Εάν

αλλοιώς:

Εάν $p \leq 0$ τότε:

$$a_{k+1} = a_k - \Delta_{\max}, \quad b_{k+1} = b_k, \quad c_{k+1} = c_k$$

$$F_a = F(a_{k+1})$$

αλλοιώς:

Εάν $a_k - \hat{x} > \Delta_{\max}$ τότε:

$$a_{k+1} = a_k - \Delta_{\max}$$

αλλοιώς:

$$a_{k+1} = \hat{x}$$

Τέλος-Εάν

$$b_{k+1} = c_k, \quad c_{k+1} = a_k$$

$$F_b = F_c, \quad F_c = F_a, \quad F_a = F(a_{k+1})$$

Τέλος-Εάν

Τέλος-Εάν

Τέλος επανάληψης εάν: $F_c < F_a$ και $F_c < F_b$

Μια μέθοδος που αναπτύχθηκε από τον Davidon* χρησιμοποιεί ένα σημείο a και τις τιμές F_a , g_a της συνάρτησης και της παραγώγου της στο σημείο αυτό. Επίσης χρειάζεται ένα κάτω φράγμα F_x για την τιμή του ελαχίστου. Στην συνέχεια προσαρμόζεται ένα δευτεροβάθμιο τριώνυμο έτσι ώστε η τιμή του ελαχίστου του να

* Davidon, W. " Variable metric method for minimization ", Argonne National Laboratory Report ANL-5990, (1959) Argonne, Illinois

είναι ίσο προς F_x και ταυτόχρονα στο σημείο $x = a$ να αναπαράγονται οι παραπάνω τιμές της συνάρτησης και της παραγώγου της. Επίσης το βήμα ελέγχεται σε κάθε επανάληψη, έτσι ώστε να μην πάρει ποτέ κάποια υπερβολικά μεγάλη τιμή που θα μπορούσε να δημιουργήσει αθέμιτα αριθμητικά προβλήματα. Ο αλγόριθμος έχει ως εξής:

Αλγόριθμος Davidson εντοπισμού διαστήματος αβεβαιότητας

Δίδονται ως είσοδος: a_1, Δ_{\max}, F_x

$F_a = F_b = F(a_1), g_a = g_b = g(a_1), a_0 = b_0 = a_1$

Εάν $g_a < 0$ τότε:

Επανάληψη για $k = 1, 2, \dots$

$a_k = b_{k-1}, F_a = F_b, g_a = g_b$

$$\hat{x} = a_k + \frac{2(F_x - F_a)}{g_a}$$

$b_k = \min\{\hat{x}, a_k + \Delta_{\max}\}, F_b = F(b_k), g_b = g(b_k)$

Τέλος επανάληψης εάν $g_b > 0$ ή $F_a < F_b$

αλλιώς:

Επανάληψη για $k = 1, 2, \dots$

$b_k = a_{k-1}, F_b = F_a, g_b = g_a$

$$\hat{x} = b_k + \frac{2(F_x - F_b)}{g_b}$$

$a_k = \max\{\hat{x}, b_k - \Delta_{\max}\}, F_a = F(a_k), g_a = g(a_k)$

Τέλος επανάληψης εάν $g_a < 0$ ή $F_a > F_b$

Τέλος-Εάν

3. ΑΠΛΕΣ ΜΕΘΟΔΟΙ ΠΟΛΥΔΙΑΣΤΑΤΗΣ ΕΛΑΧΙΣΤΟΠΟΙΗΣΗΣ

Όλες οι μέθοδοι ελαχιστοποίησης μή γραμμικών συναρτήσεων είναι επαναληπτικές. Λοθέντος ενός αρχικού σημείου τίθεται σε ενέργεια μιά διαδικασία που παράγει νέα σημεία που προσεγγίζουν το ελάχιστο όλο και περισσότερο. Οι λεγόμενες "απλές" μέθοδοι χρησιμοποιούν μόνον τιμές της αντικειμενικής συνάρτησης και όχι των παραγώγων της και βασίζονται στην σύγκριση των διαφόρων αυτών τιμών. Αν και οι περισσότερες από αυτές αναπτύχθηκαν διαισθητικά με βάση την τακτική "δοκιμή και λάθος" και όχι την λεπτομερή μαθηματική ανάλυση, υπάρχουν μερικές οι οποίες είναι εντυπωσιακά αποτελεσματικές ιδίως σε περιπτώσεις που η αντικειμενική συνάρτηση δεν είναι διαφορίσιμη ή έχει ασυνεχείς παραγώγους ή υποφέρει από τυχαίο θόρυβο. Στην συνέχεια περιγράφονται μερικές από τις πιο χρήσιμες "απλές" μεθόδους.

3.1 Η ΜΕΘΟΔΟΣ SIMPLEX

Η μέθοδος αυτή αναπτύχθηκε αρχικά από τον Spendley* και βελτιώθηκε στην συνέχεια από τους Nelder και Mead†. Η αντικειμενική συνάρτηση υπολογίζεται στις κορυφές ενός πολυτόπου (simplex), δηλαδή στα $(n+1)$ σημεία που ορίζουν ένα

*Spendley W., Hext G. R. and Himsworth F. R. "Sequential application of simplex designs in optimization and evolutionary operation". Technometrics (1962) 4, 441

† Nelder J. A. and Mead R. "A Simplex method for function minimization" Computer Journal (1965) 7, 308

στοιχείο όγκου στο n-διάστατο χώρο. (Το πολύτοπο π.χ. στο χώρο των δύο διαστάσεων είναι ένα τρίγωνο, στο χώρο των τριών διαστάσεων είναι ένα τετράεδρο κλπ). Αριθμούμε αυτά τα (n+1) σημεία κατάλληλα έτσι ώστε εάν $f_i = f(\mathbf{x}_i)$ να ισχύει: $f_0 < f_1 < f_2 < \dots < f_n$. Το "χειρότερο" σημείο \mathbf{x}_n αντικαθίσταται στην συνέχεια από ένα καινούριο που επιλέγεται σύμφωνα με την ακόλουθη στρατηγική.

Υπολογίζεται το κεντροειδές των υπόλοιπων σημείων: $\mathbf{x}_c = \frac{1}{n} \sum_{i=0}^{n-1} \mathbf{x}_i$

Υπολογίζεται το κατοπτρικό σημείο του \mathbf{x}_n ως προς το \mathbf{x}_c ως: $\mathbf{x}_r = \mathbf{x}_c + \alpha(\mathbf{x}_c - \mathbf{x}_n)$ όπου α μία παράμετρος που ονομάζεται συντελεστής ανάκλασης. (Συνήθως $\alpha=1$).

Στην συνέχεια μελετούνται οι τρεις παρακάτω περιπτώσεις:

I) $f_0 \leq f_r \leq f_{n-1}$

II) $f_r < f_0$ (Το \mathbf{x}_r είναι το νέο "καλύτερο" σημείο)

III) $f_r > f_{n-1}$ (Το \mathbf{x}_r είναι το νέο "χειρότερο" σημείο)

I) Το σημείο \mathbf{x}_r γίνεται αποδεκτό ως ένα νέο σημείο και στην συνέχεια επαναδιατάσσονται τα σημεία κατ' αύξουσα τιμή της αντικειμενικής συνάρτησης και επαναλαμβάνεται η διαδικασία.

II) Εφαρμόζεται η διάταξη: $\mathbf{x}_e = \mathbf{x}_c + \gamma(\mathbf{x}_r - \mathbf{x}_c)$

όπου γ μία παράμετρος που ονομάζεται συντελεστής διάταξης. (Συνήθως $\gamma=2$). Εάν $f_e < f_r$, η διάταξη είναι επιτυχής και το νέο σημείο \mathbf{x}_e αντικαθιστά το μέχρι τώρα χειρότερο (\mathbf{x}_n) ειδάλλως το \mathbf{x}_n αντικαθίσταται από το \mathbf{x}_r . Αναδιατάσσονται τα σημεία και επαναλαμβάνεται η διαδικασία.

III) Εφαρμόζεται η συστολή: $\mathbf{x}_k = \mathbf{x}_c + \beta(\mathbf{x}_n - \mathbf{x}_c)$ εάν $f_r \geq f_n$, ή $\mathbf{x}_k = \mathbf{x}_c + \beta(\mathbf{x}_r - \mathbf{x}_c)$ εάν $f_r < f_n$, όπου β μία παράμετρος που ονομάζεται συντελεστής συστολής. (Συνήθως $\beta=0.5$)

Εάν $f_k < \min(f_r, f_n)$ τότε το \mathbf{x}_k αντικαθιστά το \mathbf{x}_n , ειδικά ολό το πολύτοπο συστέλεται ως: $\mathbf{x}_i = \frac{1}{2}(\mathbf{x}_0 + \mathbf{x}_i)$, για $i=1, 2, \dots, n$.

Αναδιατάσσονται τα σημεία και επαναλαμβάνεται η διαδικασία.

Η μέθοδος έχει ένα απλό κριτήριο τερματισμού.

Ορίζουμε $\bar{f} \equiv \frac{1}{n+1} \sum_{i=0}^n f_i$, τον μέσο όρο των τιμών της συνάρτησης

στις κορυφές του πολύτοπου.

Τότε η διακύμανση: $\sigma^2 = \frac{1}{n} \sum_{i=0}^n (f_i - \bar{f})^2$ χρησιμοποιείται για τον

τερματισμό της περαιτέρω αναζήτησης εάν $\sigma < \varepsilon$, όπου ε ένας μικρός θετικός αριθμός. Ακολουθεί η περιγραφή του αλγορίθμου

ΑΛΓΟΡΙΘΜΟΣ ΕΛΑΧΙΣΤΟΠΟΙΗΣΗΣ SIMPLEX

Δίδονται ως είσοδος: $\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_n$ με αντίστοιχες τιμές f_0, f_1, \dots, f_n

οι παράμετροι α, β, γ και ένας μικρός θετικός αριθμός ε .

Βήμα 1. $\mathbf{x}_c = \frac{1}{n} \sum_{i=0}^{n-1} \mathbf{x}_i$, $f_c = f(\mathbf{x}_c)$,

$$\bar{f} \equiv \frac{1}{n+1} \sum_{i=0}^n f_i, \quad \sigma^2 = \frac{1}{n} \sum_{i=0}^n (f_i - \bar{f})^2$$

Εάν $\sigma < \varepsilon$ τότε:

τερματισμός του προγράμματος

Τέλος - Εάν

Βήμα 2. $\mathbf{x}_r = \mathbf{x}_c + \alpha(\mathbf{x}_c - \mathbf{x}_n)$, $f_r = f(\mathbf{x}_r)$

Βήμα 3. Εάν $f_0 \leq f_r \leq f_{n-1}$ τότε:

$\mathbf{x}_{n+1} = \mathbf{x}_r$, ανακατάταξη και επανάληψη από το Βήμα 1.

Τέλος - Εάν

Βήμα 4. Εάν $f_r < f_0$ τότε:

$$\mathbf{x}_e = \mathbf{x}_c + \gamma(\mathbf{x}_r - \mathbf{x}_c), f_e = f(\mathbf{x}_e)$$

Εάν $f_e < f_r$ τότε:

$\mathbf{x}_{n+1} = \mathbf{x}_e$, ανακατάταξη και επανάληψη από το Βήμα 1.

αλλιώς:

$\mathbf{x}_{n+1} = \mathbf{x}_r$, ανακατάταξη και επανάληψη από το Βήμα 1.

Τέλος - Εάν

Τέλος - Εάν

Βήμα 5. Εάν $f_r > f_{n-1}$ τότε:

Εάν $f_r \geq f_n$ τότε:

$$\mathbf{x}_k = \mathbf{x}_c + \beta(\mathbf{x}_n - \mathbf{x}_c), f_k = f(\mathbf{x}_k)$$

αλλιώς:

$$\mathbf{x}_k = \mathbf{x}_c + \beta(\mathbf{x}_r - \mathbf{x}_c), f_k = f(\mathbf{x}_k)$$

Τέλος - Εάν

Εάν $f_k < \min(f_r, f_n)$ τότε:

$$\mathbf{x}_{n+1} = \mathbf{x}_k$$

αλλιώς:

$$\mathbf{x}_i = (\mathbf{x}_0 + \mathbf{x}_i) / 2 \text{ για } i = 1, 2, \dots, n$$

Τέλος - Εάν

ανακατάταξη και επανάληψη από το Βήμα 1.

Τέλος - Εάν

3.2 Η ΜΕΘΟΔΟΣ ΕΝΑΛΛΑΣΣΟΜΕΝΩΝ ΔΙΕΥΘΥΝΣΕΩΝ

Κατά αυτήν την μέθοδο εφαρμόζονται γραμμικές αναζητήσεις σε κάθε διεύθυνση ξεχωριστά. Δηλαδή διατηρούνται οι (n-1) μεταβλητές αναλοίωτες και μεταβάλλεται μόνον μία. Την πρώτη φορά μεταβάλλεται η μεταβλητή x_1 την δεύτερη φορά η x_2 κ.ο.κ. Αλγοριθμικά η μέθοδος παρουσιάζεται ως εξής:

ΑΛΓΟΡΙΘΜΟΣ ΕΛΑΧΙΣΤΟΠΟΙΗΣΗΣ ΕΝΑΛΛΑΣΣΟΜΕΝΩΝ ΔΙΕΥΘΥΝΣΕΩΝ

Δεδομένου ενός αρχικού σημείου $\mathbf{x}^{(0)} = (x_1^{(0)}, x_2^{(0)}, \dots, x_n^{(0)})$

Επανάληψη για : $k = 0, 1, 2, \dots$

Επανάληψη για : $i = 1, 2, \dots, n$

Εύρεση ενός λ^* που ελαχιστοποιεί την $f(\mathbf{x}^{(k)} + \lambda \hat{e}_i)$

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \lambda^* \hat{e}_i$$

Τέλος-Επανάληψης

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)}$$

Ελεγχος κριτηρίων τερματισμού

Τέλος-Επανάληψης

Η παραπάνω μέθοδος δεν έχει καλές επιδόσεις, ιδίως όταν υπάρχει συσχέτιση μεταξύ των μεταβλητών της συνάρτησης η οποία δεν λαμβάνεται καθόλου υπόψη. Τότε η πορεία προς το ελάχιστο γίνεται με ολοένα και μικρότερα βήματα με αποτέλεσμα να εκτελούνται πολλές κλήσεις στην αντικειμενική συνάρτηση και ως εκ τούτου η μέθοδος να μην είναι αποδοτική. Παρόλα αυτά, η μέθοδος αυτή έχει αποτελέσει την βάση για την περαιτέρω ανάπτυξη και δημιουργία άλλων "απλών" μεθόδων πιο αποδοτικών όπως αυτών της ικνοαναζήτησης (Pattern Search),

Σημειώσεις: Τεχνικές βελτιστοποίησης

μιά γενική περιγραφή των οποίων δίνεται εν συντομία στην επόμενη παράγραφο.

3.3 ΜΕΘΟΔΟΙ ΙΧΝΟΑΝΑΖΗΤΗΣΗΣ

Η πιο γνωστή από αυτές τις μεθόδους είναι αυτή που βασίζεται στον αλγόριθμο των Hooke and Jeeves*. Στην μέθοδο αυτή απαντούνται δύο στάδια. Το πρώτο είναι ένα διερευνητικό στάδιο όπου γίνεται αντιληπτή η τοπική γεωμετρία. Το δεύτερο είναι το στάδιο της αναζήτησης επί του ίχνους, και είναι το μέρος όπου η μέθοδος διαφέρει ριζικά από την προηγούμενη.

Στο διερευνητικό στάδιο εφαρμόζεται μια επανάληψη γραμμικών αναζητήσεων ως προς όλες τις μεταβλητές μιά προς μιά, όπως και στην μέθοδο των εναλασσομένων διευθύνσεων, με την διαφορά ότι οι γραμμικές αναζητήσεις είναι "χαλαρές", δηλαδή δεν εντοπίζουν κάθε φορά το ακριβές ελάχιστο, απλά απαιτούν κάποια πτώση της τιμής της αντικειμενικής συνάρτησης έτσι ώστε αυτή να μην καλείται πολλές φορές. Το τελικό σημείο που εντοπίζεται στο στάδιο αυτό ονομάζεται *σημείο βάσης*.

Στην συνέχεια, στο δεύτερο στάδιο, δοκιμάζεται μια αυστηρή γραμμική αναζήτηση στην διεύθυνση που ορίζεται από την ευθεία που ενώνει τα δύο πιο πρόσφατα σημεία βάσης.

* Hooke R. and Jeeves T. A. "Direct Search solution of numerical and statistical problems". J. Ass. Comput. Mach. (1961) 8, 212-229

Υπάρχουν διάφορες παραλλαγές αυτής της μεθόδου η οποία είναι απλή και αρκετά αποτελεσματική; όμως δεν μπορεί να συγκριθεί με τις μεθόδους που θα περιγράψουμε στην συνέχεια οι οποίες χρησιμοποιούν και τιμές των παραγώγων της συνάρτησης.

3.4 ΣΤΟΧΑΣΤΙΚΕΣ ΜΕΘΟΔΟΙ

Άλλο ένα είδος μεθόδων με πολλές παραλλαγές και που εμπίπτουν στην κατηγορία των "απλών" μεθόδων. Συνήθως ορίζεται ένα στοιχείο όγκου στον n -διάστατο χώρο, εντός του οποίου διαλέγονται τυχαία σημεία. Υπολογίζεται κάθε φορά η συνάρτηση και εάν υπάρχει "πρόοδος" το σημείο γίνεται αποδεκτό ως το νέο καλύτερο σημείο, άλλως απορρίπτεται.

Οι μέθοδοι αυτού του τύπου δεν είναι αποδοτικές, αλλά η υλοποίησή τους είναι απλούστατη. Για συναρτήσεις που εμπεριέχουν "θόρυβο" στον υπολογισμό τους, μπορούν να είναι χρήσιμες. Επίσης μία τυχαία διαταραχή μπορεί να φανεί χρήσιμη σαν ένα ενδιάμεσο βήμα άλλων μεθόδων που αδυνατούν να επιτύχουν περαιτέρω πρόοδο, επειδή πχ βρίσκονται σε μία περιοχή όπου η κλίση της συνάρτησης είναι ασυνεχής.

Οι πιο πολλές εφαρμογές πλέον των στοχαστικών μεθόδων κείνται στο πεδίο της γενικής και όχι της τοπικής ελαχιστοποίησης. Ένας επιτυχής στοχαστικός αλγόριθμος γενικής ελαχιστοποίησης ονομάζεται "Simulated annealing" και έχει χρησιμοποιηθεί σε σχεδιασμό ολοκληρωμένων κυκλωμάτων.

4.1 ΜΕΘΟΔΟΙ ΜΕ ΠΑΡΑΓΩΓΟΥΣ

Οι μέθοδοι αυτής της κατηγορίας είναι επίσης επαναληπτικοί. Στην αρχή της k -ιοστής επανάληψης, η προσέγγιση για το ελάχιστο δίδεται από το σημείο $\mathbf{x}^{(k)}$. Κατά την επανάληψη υπολογίζεται μία διεύθυνση $\mathbf{s}^{(k)}$ κατά μήκος της οποίας γίνεται μια γραμμική αναζήτηση, δηλαδή ελαχιστοποιείται ως προς λ η συνάρτηση: $f(\mathbf{x}^{(k)} + \lambda \mathbf{s}^{(k)})$. Η νέα προσέγγιση για το ελάχιστο είναι: $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \lambda^* \mathbf{s}^{(k)}$. Ο τρόπος υπολογισμού της διεύθυνσης $\mathbf{s}^{(k)}$ θα μας απασχολήσει διότι είναι αυτός που καθορίζει τις διάφορες μεθόδους.

4.2 ΦΘΙΝΟΥΣΕΣ ΚΑΤΕΥΘΥΝΣΕΙΣ

Μια φθίνουσα κατεύθυνση καθορίζεται από ένα διάνυσμα \mathbf{s} για το οποίο ισχύει: $f(\mathbf{x} + a\mathbf{s}) < f(\mathbf{x})$, $a > 0$. Για $a \rightarrow 0$ έχουμε: $f(\mathbf{x} + a\mathbf{s}) \approx f(\mathbf{x}) + a\mathbf{g} \cdot \mathbf{s}$. Για να είναι λοιπόν φθίνουσα η διεύθυνση \mathbf{s} πρέπει να ισχύει: $\mathbf{g} \cdot \mathbf{s} < 0$. Γενικά μία μέθοδος που παράγει φθίνουσες κατευθύνσεις, θα συγκλίνει στο ελάχιστο με τις κατάλληλες γραμμικές αναζητήσεις. Ένα πρόβλημα που ανακύπτει είναι τα σαγμοειδή σημεία. Σαγμοειδές είναι ένα σημείο για το οποίο ισχύει: $\nabla f(\mathbf{x}) \equiv \mathbf{g} = 0$, αλλά ο Εσσιανός πίνακας $\nabla^2 f(\mathbf{x}) \equiv \mathbf{G}$ δεν είναι θετικά ορισμένος. (Δηλαδή έχει και τουλάχιστον μία αρνητική ιδιοτιμή). Επανεξετάζουμε την ανάπτυξη κατά Taylor γύρω από το σαγμοειδές και έχουμε: $f(\mathbf{x} + a\mathbf{s}) \approx f(\mathbf{x}) + \frac{1}{2} a^2 \mathbf{s} \cdot \mathbf{G} \cdot \mathbf{s}$. Αμέσως βλέπουμε ότι για να είναι φθίνουσα κατεύθυνση η \mathbf{s} πρέπει να ισχύει: $\mathbf{s} \cdot \mathbf{G} \cdot \mathbf{s} < 0$. Τετοιου

είδους φθίνουσες κατευθύνσεις ονομάζονται "φθίνουσες κατευθύνσεις αρνητικής καμπυλότητας".

4.3 ΓΡΑΜΜΙΚΗ ΑΝΑΖΗΤΗΣΗ

Ένα από τα πιο σημαντικά μέρη της διαδικασίας για την πολυδιάστατη ελαχιστοποίηση είναι η γραμμική (μονοδιάστατη) αναζήτηση, κατά την οποία η συνάρτηση $f(\mathbf{x}^{(k)} + \lambda \mathbf{s}^{(k)})$ ελαχιστοποιείται ως προς λ . Η αναγκαία συνθήκη $\frac{\partial f(\mathbf{x}^{(k)} + \lambda \mathbf{s}^{(k)})}{\partial \lambda} = 0$ μεταφράζεται ως: $\frac{\partial f(\mathbf{x}^{(k)} + \lambda \mathbf{s}^{(k)})}{\partial \lambda} = \mathbf{s}^{(k)} \cdot \nabla f(\mathbf{x}^{(k)} + \lambda \mathbf{s}^{(k)}) \equiv \mathbf{s}^{(k)} \cdot \mathbf{g}^{(k+1)} = 0$.

Δηλαδή η διεύθυνση $\mathbf{s}^{(k)}$, είναι κάθετη στην κλίση της συνάρτησης στο σημείο που τερματίζει η γραμμική αναζήτηση. Το αποτέλεσμα αυτό είναι χρήσιμο για την απόδειξη ορισμένων ιδιοτήτων των αλγορίθμων, παρότι ακριβείς γραμμικές αναζητήσεις δεν χρησιμοποιούνται πλέον διότι είναι δαπανηρές.

4.3 ΥΠΟΛΟΓΙΣΜΟΣ ΠΑΡΑΓΩΓΩΝ

Ο υπολογισμός των παραγώγων είναι πολύ σημαντικός στις τεχνικές ελαχιστοποίησης. Είναι βέβαια προτιμητέον να υπολογίζονται αναλυτικά. Πολλές φορές όμως αυτό δεν είναι εύκολο. Σε αυτές τις περιπτώσεις καταφεύγει κανείς σε προσεγγίσεις πεπερασμένων διαφορών. Οι πιο συνήθεις προσεγγίσεις είναι:

$$\frac{df}{dx} \approx \frac{f(x+h) - f(x)}{h}, \quad \frac{df}{dx} \approx \frac{f(x+h) - f(x-h)}{2h}$$

Οι κεντρικές διαφορές είναι μεγαλύτερης ακρίβειας.

Επίσης για τις δεύτερες παραγώγους κατασκευάζονται αντίστοιχες προσεγγίσεις. Πχ για τα διαγώνια στοιχεία του Εσσιανού πίνακα έχουμε:

Σημειώσεις: Τεχνικές βελτιστοποίησης

$$\frac{d^2 f}{dx^2} \approx \frac{f(x+h) + f(x-h) - 2f(x)}{h^2}$$

και για τα μη διαγώνια στοιχεία του Εσσιανού πίνακα:

$$\frac{\partial^2 f}{\partial x \partial y} \approx \frac{f(x+h, y+s) - f(x-h, y+s) - f(x+h, y-s) + f(x-h, y-s)}{4hs}$$

ή εναλλακτικά:

$$\frac{\partial^2 f}{\partial x \partial y} \approx \frac{f(x+h, y+s) - f(x, y+s) - f(x+h, y) + f(x, y)}{hs}$$

Τα βήματα h και s για τις πεπερασμένες διαφορές πρέπει να διαλεχτούν καταλλήλως για να επιτύχουν την μεγίστη δυνατή ακρίβεια.

Για τις μη-κεντρικές διαφορές χρησιμοποιείται συνήθως ένα βημα ως: $h = \sqrt{\eta} \max\{|x|, typ(x)\} sign(x)$,

όπου η η ακρίβεια (διακριτικότητα) της μηχανής (δηλαδή ο μέγιστος αριθμός η για τον οποίο οι ποσότητες $1+\eta$ και 1 , έχουν την ίδια τιμή σαν συνέπεια της πεπερασμένης ακρίβειας της μηχανής) και $typ(x)$ μιά τυπική τιμή για το μέγεθος του x .

Για τις κεντρικές διαφορές η κατάλληλη επιλογή είναι: $h = (\eta)^{1/3} \max\{|x|, typ(x)\} sign(x)$.

Σε μιά γειτονιά περίξ του ελαχίστου, προτιμούνται οι τύποι των κεντρικών διαφορών, διότι εκεί η συνάρτηση δεν είναι ποτέ γραμμική και μάλιστα είναι συνήθως δευτέρου βαθμού.

4.4 Η ΜΕΘΟΔΟΣ ΤΗΣ ΜΕΓΙΣΤΗΣ ΠΤΩΣΗΣ

Η πτώση της τιμής της συνάρτησης γε μιά επανάληψη είναι:

$$f(\mathbf{x}^{(k)} + \lambda \mathbf{s}^{(k)}) - f(\mathbf{x}^{(k)}) \approx \lambda \mathbf{s}^{(k)} \cdot \mathbf{g}^{(k)} = \lambda \|\mathbf{s}^{(k)}\| \|\mathbf{g}^{(k)}\| \cos(\vartheta)$$

όπου ϑ η γωνία μεταξύ $\mathbf{s}^{(k)}$ και $\mathbf{g}^{(k)}$. Η διαφορά αυτή είναι μεγίστη όταν $\vartheta = \pi$, δηλαδή όταν το $\mathbf{s}^{(k)}$ είναι συγγραμμικό με το $\mathbf{g}^{(k)}$ αλλά αντιθέτου φοράς.

Στην μέθοδο της μέγιστης πτώσης θέτουμε: $\mathbf{s}^{(k)} = -\mathbf{g}^{(k)}$.

Προφανώς είναι φθίνουσα διεύθυνση διότι ισχύει:

$$\mathbf{s}^{(k)} \cdot \mathbf{g}^{(k)} = -\mathbf{g}^{(k)} \cdot \mathbf{g}^{(k)} = -\|\mathbf{g}^{(k)}\|^2 < 0 .$$

Ο αλγόριθμος έχει ως εξής:

Αλγόριθμος μέγιστης πτώσης

Αρχικά δεδομένα: $\mathbf{x}^{(0)}$, $gtol$

Επανάληψη για $k=0, 1, 2, \dots$

$$\mathbf{s}^{(k)} = -\mathbf{g}^{(k)}$$

Γραμμική αναζήτηση κατά την διεύθυνση $\mathbf{s}^{(k)}$,

και υπολογισμός του $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \lambda \mathbf{s}^{(k)}$ και του $\mathbf{g}^{(k+1)}$

Εάν $\|\mathbf{g}^{(k+1)}\| \leq gtol$, Τερματισμός. (STOP)

Τέλος-Επανάληψης

Η μέθοδος αυτή αν και έχει καλύτερες επιδόσεις από την μέθοδο των εναλλασσόμενων μεταβλητών, παρουσιάζει και αυτή το πρόβλημα της ταλαντωτικής συμπεριφοράς που έχει σαν αποτέλεσμα την αργή πρόοδο. Επίσης η μέθοδος τερματίζει συχνά μακριά από το ελάχιστο συνεπεία των αριθμητικών σφαλμάτων στρογγύλευσης.

4.5 Η ΜΕΘΟΔΟΣ NEWTON ΜΕ ΠΑΡΑΛΛΑΓΕΣ

Στην γειτονιά του ελαχίστου η συνάρτηση μπορεί να προσεγγιστεί επιτυχώς με ένα τετραγωνικό μοντέλο με τον Ερσιανό πίνακα θετικά ορισμένο. Στην περίπτωση αυτή η σύγκλιση στο ελάχιστο είναι ταχεία. Αναπτύσσοντας κατά Taylor την κλίση έχουμε: $\mathbf{g}^{(k+1)} = \mathbf{g}(\mathbf{x}^{(k)} + \mathbf{s}^{(k)}) \approx \mathbf{g}^{(k)} + \mathbf{G}^{(k)} \cdot \mathbf{s}^{(k)}$. Εάν το σημείο $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \mathbf{s}^{(k)}$ είναι το ελάχιστο, πρέπει να ισχύει: $\mathbf{g}^{(k+1)} = 0$ δηλαδή $\mathbf{G}^{(k)} \cdot \mathbf{s}^{(k)} = -\mathbf{g}^{(k)}$, οπότε έχουμε: $\mathbf{s}^{(k)} = -\mathbf{G}^{(k)-1} \cdot \mathbf{g}^{(k)}$. Επειδή η προσέγγιση δεν είναι ακριβής παρά μόνο για τετραγωνικές συναρτήσεις, η διαδικασία επαναλαμβάνεται και έτσι έχουμε τον παρακάτω αλγόριθμο.

Αλγόριθμος Newton

Αρχικά δεδομένα: $\mathbf{x}^{(0)}$, $gtol$

Επανάληψη για $k=0, 1, 2, \dots$

Επιλύεται ως προς $\mathbf{s}^{(k)}$ η εξίσωση: $\mathbf{G}^{(k)} \cdot \mathbf{s}^{(k)} = -\mathbf{g}^{(k)}$

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \mathbf{s}^{(k)}$$

Υπολογίζεται το $\mathbf{g}^{(k+1)} = \nabla f(\mathbf{x}^{(k+1)}) = \nabla f(\mathbf{x}^{(k)} + \mathbf{s}^{(k)})$

Εάν $\|\mathbf{g}^{(k+1)}\| \leq gtol$, Τερματισμός. (STOP)

Τέλος-Επανάληψης

Η επίλυση του γραμμικού συστήματος $\mathbf{G}^{(k)} \cdot \mathbf{s}^{(k)} = -\mathbf{g}^{(k)}$ δεν γίνεται ποτέ με αντιστροφή του πίνακα \mathbf{G} , αλλά με άλλους τρόπους, όπως πχ με παραγοντοποίηση Choleski: $\mathbf{G} = \mathbf{LDL}^T$, Goldfarb &

Ιδιότητες: $\mathbf{Z}^T \mathbf{G} \mathbf{Z} = \mathbf{I}$, QR: $\mathbf{G} = \mathbf{Q} \mathbf{R}$ (\mathbf{Q} ορθογώνιος, \mathbf{R} άνω τριγωνικός), με διαγωνοποίηση, δηλαδή με μεθόδους που ανιχνεύουν κατά την διάρκεια της επίλυσης κατά πόσον η λύση είναι εφικτή, και εάν ο Εσσιανός πίνακας είναι θετικά ορισμένος ή ιδιάζων.

4.6 ΠΡΟΒΛΗΜΑΤΑ ΜΕ ΤΗΝ ΜΕΘΟΔΟ NEWTON

Για να είναι φθίνουσες οι κατευθύνσεις $\mathbf{s}^{(k)}$, πρέπει να ισχύει: $-\mathbf{s}^{(k)} \cdot \mathbf{g}^{(k)} = \mathbf{g}^{(k)} \cdot \mathbf{G}^{(k)-1} \cdot \mathbf{g}^{(k)} > 0$. Αυτό συμβαίνει πάντα εφόσον η κλίση είναι μή μηδενική και ο Εσσιανός πίνακας (και άρα και ο αντίστροφός του) είναι θετικά ορισμένος. Στην περιοχή ενός απομονωμένου ελαχίστου, ο πίνακας \mathbf{G} είναι θετικά ορισμένος και το τετραγωνικό μοντέλο είναι επιτυχές. Η μέθοδος του Newton θα συγκλίνει ταχέως εάν το αρχικό σημείο είναι στην περιοχή του ελαχίστου. Τα προβλήματα αρχίζουν όταν η αρχική προσέγγιση δεν είναι στην γειτονιά του ελαχίστου και ο Εσσιανός δεν είναι θετικά ορισμένος ή είναι ιδιάζων. Επίσης το βήμα $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \mathbf{s}^{(k)}$ μπορεί να είναι υπερβολικά μεγάλο όταν η συνάρτηση δεν προσεγγίζεται επαρκώς από το τετραγωνικό μοντέλο. (Αυτό όμως μπορεί να διορθωθεί με την πρόσθεση μιας γραμμικής αναζήτησης κατά την διεύθυνση $\mathbf{s}^{(k)}$). Επίσης όταν συμβαίνει να ισχύει $\mathbf{s}^{(k)} \cdot \mathbf{g}^{(k)} = 0$ ($\mathbf{g}^{(k)} \neq 0$) ότι το σημείο $\mathbf{x}^{(k)}$ είναι ήδη στο ελάχιστο κατά μήκος του $\mathbf{s}^{(k)}$ και άρα δεν μπορεί να υπάρξει περαιτέρω πρόοδος.

Η μέθοδος έχει επίσης πρόβλημα με τα σαγμοειδή σημεία, όπου έχουμε $\mathbf{g}^{(k)} = 0$ αλλά ο \mathbf{G} δεν είναι θετικά ορισμένος, και η μόνη διεύθυνση που επιλύει την $\mathbf{G}^{(k)} \cdot \mathbf{s}^{(k)} = -\mathbf{g}^{(k)}$ είναι η $\mathbf{s}^{(k)} = 0$

Σημειώσεις: Τεχνικές βελτιστοποίησης

που είναι τετριμμένη και άχρηστη. Τα περισσότερα από τα προβλήματα αυτά επιλύονται ή αποφεύγονται με κάποιες αλλαγές του βασικού αλγορίθμου που προσεγγίζουν τον Εσσιανό πίνακα \mathbf{G} με έναν θετικά ορισμένο πίνακα $\bar{\mathbf{G}}$.

4.7 ΠΑΡΑΓΟΝΤΕΣ CHOLESKI. (ΜΕΘΟΔΟΣ GILL & MURRAY*)

Κατά την μέθοδο αυτή ο Εσσιανός πίνακας παραγοντοποιείται ως $\mathbf{G} = \mathbf{LDL}^T$ όπου ο πίνακας \mathbf{D} είναι διαγώνιος και ο \mathbf{L} κάτω τριγωνικός με μονάδες στην διαγώνιο. Η παραγοντοποίηση γίνεται βάσει του παρακάτω αλγορίθμου που είναι παρόμοιος με την \mathbf{LL}^T παραγοντοποίηση που έχει περιγραφεί αρχικά.

$$\begin{array}{l}
 \text{Επανάληψη για } j = 1, 2, \dots, n \\
 \left| \begin{array}{l}
 D_{jj} = G_{jj} - \sum_{q=1}^{j-1} D_{qq} L_{jq}^2 \\
 \text{Επανάληψη για } i = j+1, j+2, \dots, n \\
 \left| \begin{array}{l}
 L_{ij} = \frac{G_{ij} - \sum_{q=1}^{j-1} D_{qq} L_{iq} L_{jq}}{D_{jj}} \\
 \text{Τέλος-Επανάληψης} \\
 \text{Τέλος-Επανάληψης}
 \end{array}
 \right.
 \end{array}
 \right.
 \end{array}$$

Το σύστημα $\mathbf{G}^{(k)} \cdot \mathbf{s}^{(k)} = -\mathbf{g}^{(k)}$ τώρα λύνεται εύκολα ως εξής:

$$\mathbf{L}^{(k)} \mathbf{t}^{(k)} = -\mathbf{g}^{(k)}$$

$$\mathbf{L}^{(k)T} \mathbf{s}^{(k)} = \mathbf{D}^{(k)-1} \mathbf{t}^{(k)}$$

* P.E. Gill and W. Murray "Newton type methods for unconstrained and linearly constrained optimization" Math. Prog. (1974) 7, 311-350

Εάν ο \mathbf{G} είναι θετικά ορισμένος όλα τα διαγώνια στοιχεία του \mathbf{D} είναι θετικά και η παραγοντοποίηση είναι αριθμητικά σταθερή. Όταν ο \mathbf{G} δεν είναι θετικά ορισμένος τότε η παραγοντοποίηση είναι αριθμητικά ασταθής. Έτσι εξετάζοντας κατά την διάρκεια του υπολογισμού τα στοιχεία \mathbf{D}_{qq} , μπορούμε να ελέγχουμε κατά πόσον ο \mathbf{G} είναι θετικά ορισμένος. Εάν λοιπόν κάποιο από τα στοιχεία \mathbf{D}_{qq} είναι μικρότερα από κάποια μικρή θετική τιμή δ , τότε μία θετική ποσότητα \mathbf{R}_{qq} προστίθεται στο \mathbf{G}_{qq} για "επανόρθωση". Έτσι λοιπόν το διαγώνιο στοιχείο \mathbf{D}_{qq} αντικαθίσταται από το $\bar{\mathbf{D}}_{qq} = \mathbf{D}_{qq} + \mathbf{R}_{qq} = \delta$. Στην συνέχεια ο προκύπτων παράγων Choleski συμβολίζεται ως $\bar{\mathbf{L}}$ και ο προσεγγίζων τον Εσσιανό πίνακα $\bar{\mathbf{G}} = \bar{\mathbf{L}}\bar{\mathbf{D}}\bar{\mathbf{L}}^T = \mathbf{G} + \mathbf{R}$, (\mathbf{R} διαγώνιος με στοιχεία \mathbf{R}_{qq}). Εάν ο \mathbf{G} είναι θετικά ορισμένος τότε $\bar{\mathbf{G}} = \mathbf{G}$. Ο $\bar{\mathbf{G}}$ είναι θετικά ορισμένος εκ κατασκευής και ως εκ τούτου εγγυάται την δημιουργία φθίνουσας κατεύθυνσης $\mathbf{s}^{(k)}$. Για την περίπτωση των σαγμοειδών επιλέγεται ως διεύθυνση αναζήτησης μια $\mathbf{s}^{(k)}$ με αρνητική καμπυλότητα ως εξής: $\bar{\mathbf{L}}^T \mathbf{s}^{(k)} = \hat{\mathbf{e}}_m$ όπου m είναι η θέση στην οποία προστέθηκε το μέγιστο στοιχείο επανόρθωσης \mathbf{R}_{mm} . Κατευθύνσεις αρνητικής καμπυλότητας μπορούν να κατασκευαστούν με πολλούς τρόπους, αλλά ο προαναφερθείς είναι "οικονομικός" λόγω της προϋπαρξης του πίνακα $\bar{\mathbf{L}}$. Η χρήση της ως άνω κατεύθυνσης γίνεται όταν $\mathbf{g}^{(k)} = 0$ και $\|\mathbf{R}\| > 0$ (συνθήκες που επικρατούν στα σαγμοειδή). Επειδή όμως αυτές οι συνθήκες μπορεί να ισχύουν και σε κάποιο παθολογικό ελάχιστο πρέπει πάντα να γίνεται έλεγχος.

Σημειώσεις: Τεχνικές βελτιστοποίησης

Αλγόριθμος Gill & Murray

Αρχικά δεδομένα: $\mathbf{x}^{(0)}$, $gtol$, δ

$\tau = 0$

Επανάληψη για $k=0, 1, 2, \dots$

Υπολογισμός των: $\mathbf{G}, \mathbf{R}, \bar{\mathbf{L}}, \bar{\mathbf{D}}$

Εάν $\|\mathbf{g}^{(k)}\| > gtol$ τότε:

Υπολογισμός $\mathbf{s}^{(k)}$ ως:

$$\mathbf{L}^{(k)} \mathbf{t}^{(k)} = -\mathbf{g}^{(k)}, \quad \mathbf{L}^{(k)T} \mathbf{s}^{(k)} = \mathbf{D}^{(k)-1} \mathbf{t}^{(k)}$$

αλλιώς Εάν $\|\mathbf{R}\| > 0$ τότε:

Υπολογισμός του m για το οποίο $R_{mm} = \max_q \{R_{qq}\}$

Υπολογισμός $\mathbf{s}^{(k)}$ ως: $\bar{\mathbf{L}}^T \mathbf{s}^{(k)} = \hat{\mathbf{e}}_m$

Εάν $\|\mathbf{g}^{(k)}\| > 0$: $\mathbf{s}^{(k)} = -\text{sign}(\mathbf{s}^{(k)} \cdot \mathbf{g}^{(k)}) \mathbf{s}^{(k)}$

Εάν $\mathbf{s}^{(k)} \cdot \mathbf{G} \cdot \mathbf{s}^{(k)} > 0$: $\tau = 1$

αλλιώς :

$\tau = 1$

Τέλος-Εάν

Εάν $\tau = 0$ τότε:

Γραμμική αναζήτηση επι της $\mathbf{s}^{(k)}$

υπολογισμός $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \lambda \mathbf{s}^{(k)}$

Τέλος-Εάν

Εάν $\tau = 1$: Τερματισμός (STOP)

Τέλος-Επανάληψης

4.8 Μέθοδοι Quasi-Newton

Οι μέθοδοι αυτού του τύπου δεν χρησιμοποιούν δεύτερες παραγώγους. Διατηρείται μια προσέγγιση του Εσσιανού πίνακα από πληροφορία που προέρχεται από πρώτες παραγώγους και η οποία ενημερώνεται σε κάθε επανάληψη. Εάν συμβολίσουμε με $\mathbf{B}^{(k)}$ την ως άνω προσέγγιση τότε η φθίνουσα κατεύθυνση υποορίζεται από την σχέση: $\mathbf{B}^{(k)}\mathbf{s}^{(k)} = -\mathbf{g}^{(k)}$. Εάν δε συμβολίσουμε με $\mathbf{H}^{(k)}$ την αντίστοιχη προσέγγιση για τον αντίστροφο του Εσσιανού $\mathbf{G}^{(k)-1}$, έχουμε $\mathbf{s}^{(k)} = -\mathbf{H}^{(k)}\mathbf{g}^{(k)}$. Το επόμενο σημείο επιλέγεται με την διαδικασία της γραμμικής αναζήτησης. Στην συνέχεια δημιουργείται η νέα προσέγγιση για τον Εσσιανό ή τον αντίστροφό του ως: $\mathbf{H}^{(k+1)} = \mathbf{H}^{(k)} + \mathbf{Q}_H^{(k)}$ όπου ο πίνακας \mathbf{Q} υπολογίζεται μόνο από τις τιμές των $\mathbf{x}^{(k+1)}$, $\mathbf{x}^{(k)}$, $\mathbf{g}^{(k+1)}$, $\mathbf{g}^{(k)}$, $\mathbf{H}^{(k)}$. Η αρχική προσέγγιση $\mathbf{H}^{(0)}$ μπορεί να είναι ένας οποιοδήποτε θετικά ορισμένος πίνακας και μιά συνήθης επιλογή είναι ο πίνακας μονάδα, $\mathbf{H}^{(0)} = \mathbf{I}$. Απαιτούμε ορισμένες ιδιότητες για τον πίνακα $\mathbf{H}^{(k+1)}$ κατάλληλες ώστε να προσομοιάζει τον $\mathbf{G}^{(k)-1}$. Διαλέγεται ο πίνακας \mathbf{Q} έτσι ώστε: $\mathbf{H}^{(k+1)}\Delta\mathbf{g}^{(k)} = \Delta\mathbf{x}^{(k)}$ όπου $\Delta\mathbf{g}^{(k)} = \mathbf{g}^{(k+1)} - \mathbf{g}^{(k)}$, $\Delta\mathbf{x}^{(k)} = \mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}$. Η συνθήκη αυτή ονομάζεται "Συνθήκη Quasi-Newton". Η παραπάνω διαδικασία ορίζει μιά νέα κατηγορία μεθόδων που ονομάζονται "Μέθοδοι Quasi-Newton" ή "Μέθοδοι μεταβλητής μετρικής".

Η ως άνω περιγραφή είναι βασισμένη στην απευθείας κατασκευή του πίνακα $\mathbf{H}^{(k+1)}$ που έχει το πλεονέκτημα ότι αποφεύγει την

επίλυση του γραμμικού συστήματος $\mathbf{B}^{(k)}\mathbf{s}^{(k)} = -\mathbf{g}^{(k)}$. Αντίστοιχα η περιγραφή θα μπορούσε ισοδύναμα να βασισθεί στον πίνακα $\mathbf{B}^{(k)}$.

4.9 ΜΕΘΟΔΟΙ DFP και BFGS

Εάν χρησιμοποιήσουμε την "Quasi-Newton" συνθήκη και αντικαταστήσουμε την έκφραση για τον πίνακα \mathbf{Q} , καταλήγουμε στην παρακάτω εξίσωση:

$$\mathbf{Q}_H^{(k)} \Delta \mathbf{g}^{(k)} = \Delta \mathbf{x}^{(k)} - \mathbf{H}^{(k)} \Delta \mathbf{g}^{(k)}.$$

Μία απλή λύση της οποίας δίδεται από:

$$\mathbf{Q}_H^{(k)} = \frac{\Delta \mathbf{x}^{(k)} \mathbf{y}^{(k)T}}{\mathbf{y}^{(k)T} \Delta \mathbf{g}^{(k)}} - \frac{\mathbf{H}^{(k)} \Delta \mathbf{g}^{(k)} \mathbf{z}^{(k)T}}{\mathbf{z}^{(k)T} \Delta \mathbf{g}^{(k)}}$$

όπου τα διανύσματα $\mathbf{y}^{(k)}$ και $\mathbf{z}^{(k)}$ είναι αυθαίρετα, αλλά με τον περιορισμό να μην μηδενίζονται οι παρονομαστές. Επειδή ο πίνακας $\mathbf{G}^{(k)-1}$ είναι συμμετρικός επιβάλλουμε και στον πίνακα \mathbf{Q} να είναι συμμετρικός. Ο απλούστερος ίσως τρόπος είναι να επιλέξουμε κατάλληλα τα διανύσματα $\mathbf{y}^{(k)}$ και $\mathbf{z}^{(k)}$.

Θέτουμε: $\mathbf{y}^{(k)} = \Delta \mathbf{x}^{(k)}$ και $\mathbf{z}^{(k)} = \mathbf{H}^{(k)} \Delta \mathbf{g}^{(k)}$. Με αυτή την επιλογή παίρνουμε την μέθοδο DFP* (Davidon, Fletcher, Powell):

* Davidon, W. " Variable metric method for minimization ", Argonne National Laboratory Report ANL-5990, (1959) Argonne, Illinois

Fletcher R. and Powell M.J.D. " A rapidly convergent descent method for minimization" Computer Journal (1963) 6, 163-168

$$\mathbf{H}^{(k+1)} = \mathbf{H}^{(k)} + \frac{\Delta \mathbf{x}^{(k)} \Delta \mathbf{x}^{(k)T}}{\Delta \mathbf{x}^{(k)T} \Delta \mathbf{g}^{(k)}} - \frac{\mathbf{H}^{(k)} \Delta \mathbf{g}^{(k)} \Delta \mathbf{g}^{(k)T} \mathbf{H}^{(k)}}{\Delta \mathbf{g}^{(k)T} \mathbf{H}^{(k)} \Delta \mathbf{g}^{(k)}}$$

Μιά πιο γενική έκφραση έχει διατυπωθεί από τον Broyden[#] και έχει ως εξής:

$$\mathbf{H}^{(k+1)} = \mathbf{H}^{(k)} + \frac{\Delta \mathbf{x}^{(k)} \Delta \mathbf{x}^{(k)T}}{\Delta \mathbf{x}^{(k)T} \Delta \mathbf{g}^{(k)}} - \frac{\mathbf{H}^{(k)} \Delta \mathbf{g}^{(k)} \Delta \mathbf{g}^{(k)T} \mathbf{H}^{(k)}}{\Delta \mathbf{g}^{(k)T} \mathbf{H}^{(k)} \Delta \mathbf{g}^{(k)}} + \pi^{(k)} \Delta \mathbf{g}^{(k)T} \mathbf{H}^{(k)} \Delta \mathbf{g}^{(k)} \mathbf{w}^{(k)} \mathbf{w}^{(k)T}$$

όπου $\pi^{(k)}$ ελεύθερη παράμετρος και $\mathbf{w}^{(k)} = \frac{\Delta \mathbf{x}^{(k)}}{\Delta \mathbf{x}^{(k)T} \Delta \mathbf{g}^{(k)}} - \frac{\mathbf{H}^{(k)} \Delta \mathbf{g}^{(k)}}{\Delta \mathbf{g}^{(k)T} \mathbf{H}^{(k)} \Delta \mathbf{g}^{(k)}}$

Οι παραπάνω σχέσεις ορίζουν την οικογένεια της ενημέρωσης Broyden. Η DFP μέθοδος είναι ειδική περίπτωση της δια $\pi^{(k)}=0$. Η ειδική περίπτωση $\pi^{(k)}=1$ είναι πολύ σημαντική και η αντίστοιχη σχέση ενημέρωσης είναι γνωστή ως BFGS από τα αρχικά των ερευνητών που την ανακάλυψαν ανεξάρτητα (Broyden, Fletcher, Goldfarb & Shanno). Σήμερα η μέθοδος BFGS θεωρείται ως η πιο επιτυχής γενική μέθοδος για την επίλυση προβλημάτων ελαχιστοποίησης. Η BFGS σχέση μπορεί να γραφτεί λίγο διαφορετικά ως εξής:

$$\mathbf{H}^{(k+1)} = \left[\mathbf{I} - \frac{\Delta \mathbf{x}^{(k)} \Delta \mathbf{g}^{(k)T}}{\Delta \mathbf{x}^{(k)T} \Delta \mathbf{g}^{(k)}} \right] \mathbf{H}^{(k)} \left[\mathbf{I} - \frac{\Delta \mathbf{x}^{(k)} \Delta \mathbf{g}^{(k)T}}{\Delta \mathbf{x}^{(k)T} \Delta \mathbf{g}^{(k)}} \right]^T + \frac{\Delta \mathbf{x}^{(k)} \Delta \mathbf{x}^{(k)T}}{\Delta \mathbf{x}^{(k)T} \Delta \mathbf{g}^{(k)}}$$

Η διατύπωση των ως άνω μεθόδων με βάση τον πίνακα $\mathbf{B}^{(k)}$

επιτυγχάνεται με τις εξής αντικαταστάσεις:

$$\mathbf{B}^{(k)} \leftrightarrow \mathbf{H}^{(k)}, \quad \mathbf{B}^{(k+1)} \leftrightarrow \mathbf{H}^{(k+1)} \quad \text{και} \quad \Delta \mathbf{x}^{(k)} \leftrightarrow \Delta \mathbf{g}^{(k)}.$$

Η BFGS σχέση διαμορφώνεται ως:

$$\mathbf{B}^{(k+1)} = \mathbf{B}^{(k)} + \frac{\Delta \mathbf{g}^{(k)} \Delta \mathbf{g}^{(k)T}}{\Delta \mathbf{x}^{(k)T} \Delta \mathbf{g}^{(k)}} - \frac{\mathbf{B}^{(k)} \Delta \mathbf{x}^{(k)} \Delta \mathbf{x}^{(k)T} \mathbf{B}^{(k)}}{\Delta \mathbf{x}^{(k)T} \mathbf{B}^{(k)} \Delta \mathbf{x}^{(k)}}$$

Αντίστοιχα η έκφραση για την DFP έχει ως:

[#] Broyden C. G. "Quasi-Newton methods and their application to function minimization" Math. Comp. (1967) 21, 368-381

$$\mathbf{B}^{(k+1)} = \left[\mathbf{I} - \frac{\Delta \mathbf{g}^{(k)} \Delta \mathbf{x}^{(k)T}}{\Delta \mathbf{x}^{(k)T} \Delta \mathbf{g}^{(k)}} \right] \mathbf{B}^{(k)} \left[\mathbf{I} - \frac{\Delta \mathbf{g}^{(k)} \Delta \mathbf{x}^{(k)T}}{\Delta \mathbf{x}^{(k)T} \Delta \mathbf{g}^{(k)}} \right]^T + \frac{\Delta \mathbf{g}^{(k)} \Delta \mathbf{g}^{(k)T}}{\Delta \mathbf{x}^{(k)T} \Delta \mathbf{g}^{(k)}}$$

Ο αλγόριθμος διατυπώνεται για την περίπτωση που χρησιμοποιείται ο πίνακας $\mathbf{B}^{(k)}$ ως εξής:

Αρχικά δεδομένα: $\mathbf{x}^{(0)}, \mathbf{B}^{(0)}, gtol$

Επανάληψη για $k=0, 1, \dots$

Επίλυση του συστήματος: $\mathbf{B}^{(k)} \mathbf{s}^{(k)} = -\mathbf{g}^{(k)}$

για την φθίνουσα κατεύθυνση $\mathbf{s}^{(k)}$

Υπολογισμός του $\lambda^{*(k)}$ με αναζήτηση επί της $\mathbf{x}^{(k)} + \lambda^{(k)} \mathbf{s}^{(k)}$.

$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \lambda^{*(k)} \mathbf{s}^{(k)}$.

Ενημέρωση του $\mathbf{B}^{(k+1)}$ με μία από τις σχέσεις DFP ή BFGS

Εάν $\|\mathbf{g}^{(k+1)}\| < gtol$ τερματισμός (stop)

Τέλος-Επανάληψης

Αντίστοιχος είναι ο αλγόριθμος και για την περίπτωση που χρησιμοποιείται ο πίνακας $\mathbf{H}^{(k)}$. Σύγχρονες υλοποιήσεις χρησιμοποιούν την μέθοδο με τον πίνακα $\mathbf{B}^{(k)}$ και ενημερώνουν τους παράγοντες Choleski ή τους παράγοντες Goldfarb & Idnani που ορίζονται από την σχέση: $\mathbf{Z}^{(k)T} \mathbf{B}^{(k)} \mathbf{Z}^{(k)} = \mathbf{I}$ που μπορεί να γραφτεί και ως: $\mathbf{Z}^{(k)} \mathbf{Z}^{(k)T} = \mathbf{B}^{(k)-1}$. Οι τεχνικές αυτές δεν θα αναπτυχθούν περαιτέρω στις παρούσες σημειώσεις διότι είναι αρκετά περίπλοκες.

Οι τεχνικές τύπου Newton και Quasi-Newton απαιτούν μνήμη ανάλογη του n^2 , όπου n ο αριθμός των παραμέτρων του προβλήματος. Όταν ο αριθμός n είναι αρκετά μεγάλος ($n \geq 10000$) οι τεχνικές αυτές δεν είναι εφαρμόσιμες.

4.10 Μέθοδοι συζυγών κατευθύνσεων

Οι μέθοδοι αυτές δεν απαιτούν την αποθήκευση του Ερσιανού πίνακα παρά μόνον ορισμένων διανυσμάτων (ένας τυπικός αριθμός είναι τέσσερα διανύσματα) και για αυτόν το λόγο είναι κατάλληλες για προβλήματα με πολλές μεταβλητές.

Δύο κατευθύνσεις $\mathbf{s}^{(1)}, \mathbf{s}^{(2)}$ ονομάζονται συζυγείς ως προς έναν συμμετρικό και θετικά ορισμένο πίνακα \mathbf{G} όταν ισχύουν:

$$\mathbf{s}^{(1)\top} \mathbf{G} \mathbf{s}^{(2)} = 0, \quad \mathbf{s}^{(1)} \neq 0, \quad \mathbf{s}^{(2)} \neq 0$$

Εάν ο πίνακας \mathbf{G} είναι $n \times n$ και οι n κατευθύνσεις: $\mathbf{s}^{(0)}, \mathbf{s}^{(1)}, \mathbf{s}^{(2)}, \dots, \mathbf{s}^{(n-1)}$ είναι ανά δύο συζυγείς, τότε αποδυναμώνεται ότι είναι και γραμμικώς ανεξάρτητες.

Εστω ένα τετραγωνικό μοντέλο: $f(\mathbf{x}) = \frac{1}{2} \mathbf{x}^\top \mathbf{G} \mathbf{x} + \mathbf{b}^\top \mathbf{x} + c$ όπου \mathbf{G} σταθερός $n \times n$ πίνακας (ανεξάρτητος του \mathbf{x}), συμμετρικός και θετικά ορισμένος και \mathbf{b} σταθερό διάνυσμα στον n -διάστατο χώρο. Εάν τα διανύσματα $\mathbf{s}^{(0)}, \mathbf{s}^{(1)}, \mathbf{s}^{(2)}, \dots, \mathbf{s}^{(n-1)}$ είναι ανά δύο συζυγή (ως προς \mathbf{G}), τότε αποδυναμώνεται ότι το ελάχιστο της $f(\mathbf{x})$ μπορεί να εντοπιστεί με n διαδοχικές ακριβείς γραμμικές αναζητήσεις κατά μήκος των $\mathbf{s}^{(0)}, \mathbf{s}^{(1)}, \mathbf{s}^{(2)}, \dots, \mathbf{s}^{(n-1)}$. Υπάρχουν πολλοί τρόποι υπολογισμού τέτοιων διευθύνσεων. Ένας προφανής είναι για παράδειγμα η διαγωνοποίηση του πίνακα \mathbf{G} . Τα προκύπτοντα ιδιανύσματα είναι συζυγή. Το ζητούμενο όμως είναι να ευρεθεί ένα σύνολο ανά δύο συζυγών κατευθύνσεων χωρίς να εμπλέκεται άμεσα ο πίνακας \mathbf{G} , για να παρακαμφθεί έτσι το πρόβλημα της αποθήκευσής του που μπορεί να είναι ανυπέρβλητο.

Οι παρακάτω αλγόριθμοι επιτυγχάνουν αυτό το σκοπό.

Σημειώσεις: Τεχνικές βελτιστοποίησης

Αρχικά δεδομένα: $\mathbf{x}^{(0)}$, $gtol$

$$\mathbf{s}^{(0)} = -\mathbf{g}^{(0)}$$

Επανάληψη για $k=1, 2, \dots$

Υπολογισμός του $\mathbf{x}^{(k)}$ με γραμμική αναζήτηση κατά μήκος της διεύθυνσης $\mathbf{s}^{(k-1)}$.

Υπολογισμός ενός αριθμού $\beta^{(k)} \geq 0$ (βλέπε παρακάτω)

Υπολογισμός της διεύθυνσης $\mathbf{s}^{(k)}$ ως: $\mathbf{s}^{(k)} = -\mathbf{g}^{(k)} + \beta^{(k)}\mathbf{s}^{(k-1)}$

Εάν $\|\mathbf{g}^{(k)}\| \leq gtol$ Τερματισμός (stop)

Τέλος-Επανάληψης

Υπάρχουν διάφορες επιλογές για το $\beta^{(k)}$ που χαρακτηρίζουν τις μεθόδους. Οι πιο γνωστές είναι:

Hestenes & Stiefel:

$$\beta^{(k)} = \frac{\Delta \mathbf{g}^{(k-1)T} \mathbf{g}^{(k)}}{\Delta \mathbf{g}^{(k-1)T} \mathbf{s}^{(k-1)}}$$

Fletcher & Reeves:

$$\beta^{(k)} = \frac{\mathbf{g}^{(k)T} \mathbf{g}^{(k)}}{\mathbf{g}^{(k-1)T} \mathbf{g}^{(k-1)}}$$

Polak & Ribiere:

$$\beta^{(k)} = \frac{\Delta \mathbf{g}^{(k-1)T} \mathbf{g}^{(k)}}{\mathbf{g}^{(k-1)T} \mathbf{g}^{(k-1)}}$$

Εάν συμβεί το $\beta^{(k)}$ να πάρει αρνητική τιμή τότε τίθεται $\beta^{(k)}=0$. Για τετραγωνικές συναρτήσεις οι παραπάνω επιλογές είναι ισοδύναμες. Ο αλγόριθμος των Polak & Ribiere είναι σημαντικός διότι έχει πολύ καλύτερες επιδόσεις σε προβλήματα με μεγάλο αριθμό μεταβλητών.