

Συνθήκες αριστεύματος σε προβλήματα χωρίς περιορισμούς.

Εάν x^* είναι η θέση του ελάχιστου μιας συνάρτησης $f(x)$ $x \in \mathbb{R}^n$, που είναι συνεχής και έχει συνεχείς παραγώγους, τότε ικανοποιούνται οι συνθήκες:

$$\nabla f(x^*) = 0 \quad \text{και} \quad \nabla^2 f(x^*) \geq 0$$

ή

$$\frac{\partial f(x^*)}{\partial x_i} = 0 \quad \forall i=1, \dots, n, \quad \frac{\partial^2 f(x^*)}{\partial x_i \partial x_j} \quad \text{θετικά ημιορισμένος}$$

Η πρώτης τάξης συνθήκη συνάγεται ως εξής:

Έστω $F(a) = f(x^* + aS)$ $a \in \mathbb{R}$ και $S \in \mathbb{R}^n$

$$\left. \frac{dF(a)}{da} \right|_{a=0^+} = \lim_{a \rightarrow 0^+} \frac{f(x^* + aS) - f(x^*)}{a} = S^T \nabla f(x^*)$$

Επειδή το x^* είναι η θέση του ελάχιστου και το $a > 0 \Rightarrow S^T \nabla f(x^*) \geq 0$ για οποιοδήποτε S και άρα και για S με αντίθετη φορά $(-S)$ $-S^T \nabla f(x^*) \geq 0$ και άρα $\nabla f(x^*) = 0$.

$$\text{Επίσης: } 0 \leq \frac{f(x^* + aS) - f(x^*)}{a^2} = \frac{1}{2} S^T \nabla^2 f(x^*) S + \frac{O(a^3)}{a^2}$$

Όσο πιο $a \rightarrow 0$ έχουμε: $S^T \nabla^2 f(x^*) S \geq 0$
 που αντιστοιχεί στην συνθήκη δεύτερης τάξης.

Οι συνθήκες για προβλήματα με περιορισμούς είναι πιο πολύπλοκες και θα μελετηθούν αργότερα.

Ανάπτυξη Taylor

Σε μία διαστάση : ($x \in \mathbb{R}$)

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

Σε πολλές διαστάσεις : ($x, h \in \mathbb{R}^n$)

$$\begin{aligned} f(x+h) = f(x) &+ \sum_{i_1=1}^n h_{i_1} \frac{\partial f(x)}{\partial x_{i_1}} + \frac{1}{2!} \sum_{i_1, i_2} h_{i_1} h_{i_2} \frac{\partial^2 f(x)}{\partial x_{i_1} \partial x_{i_2}} + \\ &+ \dots + \frac{1}{n!} \sum_{i_1, i_2, \dots, i_n} h_{i_1} h_{i_2} \dots h_{i_n} \frac{\partial^n f(x)}{\partial x_{i_1} \dots \partial x_{i_n}} + \dots \end{aligned}$$

$$\text{ii) } f(x+h) = f(x) + h^T \nabla f(x) + \frac{1}{2!} h^T \nabla^2 f(x) h + \dots$$

Τετραγωνικές συναρτήσεις

$$Q(x) = \frac{1}{2} x^T B x + d^T x + c \quad x \in \mathbb{R}^n, \quad B \in \mathbb{R}^{n \times n} \text{ σταθερός}$$

Ο πίνακας B συνεκφέρει ως συμμετρικός.

Συμφωνούμε ότι:

$$B = \frac{1}{2} (B + B^T) + \frac{1}{2} (B - B^T)$$

και ότι το πρώτο όρος

είναι συμμετρικός ή ότι δεύτερος αντισυμμετρικός.

Η συνεκφορά του αντισυμμετρικού πίνακα στην $Q(x)$

$$\begin{aligned} \text{είναι μηδενική διότι: } x^T (B - B^T) x &= x^T B x - x^T B^T x = \\ &= x^T (Bx) - (Bx)^T x = 0. \end{aligned}$$

Άρα πλέον θεωρούμε ότι ο B είναι συμμετρικός $B = B^T$

$$\nabla Q(x) = Bx + d, \quad \nabla^2 Q(x) = B$$

Στο ελάχιστο x^* , θα ισχύει: $Bx^* + d = 0$ και $x^* = -B^{-1}d$.

Συμπληρώστε ότι η προσέγγιση:

$$f(x_0+h) \approx q(h) = \frac{1}{2} h^T (\nabla^2 f(x_0)) h + h^T \nabla f(x_0) + f(x_0)$$

είναι μία τετραγωνική συνάρτηση ως προς h .

Το $\min q(h)$ επιτυγχάνεται για $h = h^* = -(\nabla^2 f(x_0))^{-1} \nabla f(x_0)$

Συμπληρώστε ότι πρέπει να ισχύει: $\nabla^2 f(x_0)$ θετικά ορισμένο

· Εάν έχω και μία ιδιοτιμή είναι αρνητική τότε το σημείο h^* δεν είναι ελάχιστο αλλά σταθμονόμο σημείο. Εάν όλες οι ιδιοτιμές είναι αρνητικές τότε το σημείο είναι μέγιστο.

· Εάν έχω και μία ιδιοτιμή είναι μηδενική τότε ο εββισμός δεν μπορεί να αντιστραφεί.

Εύθεια στον $\mathbb{R}^n$

Έστω $x_0 \in \mathbb{R}^n$ και $S \in \mathbb{R}^n$. Τότε η εύθεια που περνά από το x_0 και είναι παράλληλη προς το S δίδεται από την εξίσωση:

$$X(\lambda) = x_0 + \lambda S$$

Επίσης η εύθεια που περνά από δύο σημεία $x_0, x_1 \in \mathbb{R}^n$ δίδεται από:

$$X(\lambda) = x_0 + \lambda (x_1 - x_0)$$

$$f(x_0 + \lambda S) = f(x_0) + \lambda S^T \nabla f(x_0) + \frac{1}{2} \lambda^2 S^T \nabla^2 f(x_0) S + O(\lambda^3)$$

$$\frac{d}{d\lambda} f(x_0 + \lambda S) = S^T \nabla f(x_0) + \lambda S^T \nabla^2 f(x_0) S + O(\lambda^2)$$

$$\left. \frac{d}{d\lambda} f(x_0 + \lambda S) \right|_{\lambda=0} = S^T \nabla f(x_0)$$

SHERMAN MORRISON

Έστω $A \in \mathbb{R}^{n \times n}$ και έστω ότι υπάρχει ο A^{-1}

Έστω επίσης $u, v \in \mathbb{R}^n$.

$$\text{Τότε: } (A - uv^T)^{-1} = A^{-1} + \frac{(A^{-1}u)(v^T A^{-1})}{1 - v^T A^{-1}u}$$

Σύγκλιση

Μια ακολουθία $x_1, x_2, \dots, x_k, \dots$
λέγεται ότι συγκλίνει στο x^*
όταν:

$$\lim_{k \rightarrow \infty} (x_k - x^*) = 0$$

Έστω $h_k = x_k - x^*$ τότε $h_k \rightarrow 0$

Η ταχύτητα σύγκλισης ορίζεται ως:

$$\text{Εάν } \frac{\|h_{k+1}\|}{\|h_k\|} \rightarrow \alpha > 0 \quad \alpha \in (0, 1)$$

γραμμική σύγκλιση

$$\text{Έστω } \frac{\|h_{k+1}\|}{\|h_k\|^2} \rightarrow \alpha \quad \text{τετραγωνική}$$

$$\text{Έστω } \frac{\|h_{k+1}\|}{\|h_k\|^p} \rightarrow \alpha$$

$p > 2$ p -τάξης.

$$\text{Εάν } \frac{\|h_{k+1}\|}{\|h_k\|} \rightarrow \alpha_k \quad \text{όπου } \alpha_k \rightarrow 0$$

υπεργραμμική σύγκλιση

Παράδειγμα

$$1) \quad x_k = 1 + 2^{-k} \quad x_0 = 2, x_1 = 3/2, x_2 = 5/4, \dots$$

Είναι προφανές ότι: $\lim_{k \rightarrow \infty} x_k = 1$

$$\frac{|x_{k+1} - 1|}{|x_k - 1|} = \frac{2^{-(k+1)}}{2^{-k}} = \frac{1}{2} \quad (\text{Γραφική σύγκλιση})$$

$$2) \quad x_k = 1 + 2^{-2^k}, \quad x_k \rightarrow 1$$

$$\frac{|x_{k+1} - 1|}{|x_k - 1|^2} = \frac{2^{-2^{k+1}}}{(2^{-2^k})^2} = 1 \quad (\text{Τετραγωνική})$$

$$3) \quad x_k = \frac{x_{k-1}}{k-1} + 1 - \frac{1}{k-1} \quad x_k \rightarrow 1 \quad (k=2, 3, \dots)$$

$$\frac{|x_{k+1} - 1|}{|x_k - 1|} = \frac{\left| \frac{x_k}{k} - \frac{1}{k} \right|}{|x_k - 1|} = \frac{1}{k} \rightarrow 0 \quad (\text{Υπεργραφική})$$

• Η γραφική σύγκλιση θεωρείται "αρχή".

• Η "επιθυμητή" σύγκλιση είναι ή τετραγωνική (ή ακόμη υψηλότερης τάξης)

Οι περισσότεροι αλγόριθμοι ελαχιστοποίησης επιτυγχάνουν υπεργραφική σύγκλιση σε μία γραμμική (ή και τετραγωνική) προβλήματα.

Φθινοβες κατευθύνσεις

$$\text{Έστω: } x(\lambda) = x^{(0)} + \lambda d$$

$$x(\lambda), x^{(0)}, d \in \mathbb{R}^n$$

$$\lambda \in \mathbb{R}$$

$$\text{Έστω } \lambda \in (0, \epsilon) \Leftrightarrow 0$$

και ισχύει ότι:

$$f(x(\lambda)) < f(x^{(0)})$$

Τότε η κατεύθυνση d είναι φθινοβη.

Πρακτικά αυτό σημαίνει ότι η τιμή της

f μειώνεται κατά μήκος της d .

$$\text{Συμμερίζουμε ότι: } \frac{d}{d\lambda} x(\lambda) = d$$

$$\text{Έτσι: } \left. \frac{d}{d\lambda} f(x(\lambda)) \right|_{\lambda=0} = \sum_i \left. \frac{\partial f}{\partial x_i(\lambda)} \right|_{\lambda=0} \left. \frac{d x_i(\lambda)}{d\lambda} \right|_{\lambda=0}$$

$$\text{Οπότε: } \left. \frac{d}{d\lambda} f(x^{(0)} + \lambda d) \right|_{\lambda=0} = d^T \nabla f(x^{(0)})$$

$$\left. \frac{d}{d\lambda} f(x^{(0)} + \lambda d) \right|_{\lambda=0} < 0 \Rightarrow d^T \nabla f(x^{(0)}) < 0$$

Εάν η d είναι φθινοβη.

Η συνθήκη ότι για κατεύθυνση d είναι φθινοβη κρίνεται ως:

$$d^T \nabla f(x^{(0)}) < 0$$

Παραδείγματα φθίνουσών κατευθύνσεων

Η κατεύθυνση: $d = -\nabla f(x)$ είναι φθίνουσα

$$\text{Διότι: } d^T \nabla f(x) = -(\nabla f(x))^T \nabla f(x) = -\|\nabla f(x)\|^2 < 0$$

Επίσης η κατεύθυνση $d = -H \nabla f(x)$, όπου H

πινάκας θετικά ορισμένος. Διότι:

$$d^T \nabla f = -(\nabla f^T H) \nabla f = -(\nabla f)^T H (\nabla f) < 0$$

Οι φθίνουσες κατευθύνσεις παίζουν

σημαντικό ρόλο στις μεθόδους βελτιστοποίησης.

Γραμμική αναζήτηση (Line-search) (Γ.Α.)

Η Γ.Α. είναι η αναζήτηση ενός ελαχίστου κατά μήκος μιας ευθείας.

$$\text{Διδαδῆ: } \min_{\lambda} f(x^0 + \lambda s)$$

Συνήθως ό εντοπισμός του ελαχίστου x^*

παιν ἀντιστοιχεί σε κάποιο λ^* , διδαδῆ

$$x^* = x^{(0)} + \lambda^* s$$

δίνεται προσεγγιστικό και όχι με μεγάλη ακρίβεια.

Οι αλγόριθμοι ελαχιστοποίησης έχουν συνήθως την παρακάτω δομή:

- Δεδομένου ενός σημείου $x^{(k)}$, με $f^{(k)} = f(x^{(k)})$, $g^{(k)} = \nabla f(x^{(k)})$
- 1) Υπολογίζεται μία φθίνουσα κατεύθυνση $s^{(k)}$
 - 2) Γίνεται μία Γ.Α. κατά μήκος της διεύθυνσης:

$$x(x^{(k)}) = x^{(k)} + \lambda^{(k)} s^{(k)}, \quad \dots$$
 - 3) $x^{(k+1)} = x(\lambda^{(k)})$

και επαναλαμβάνονται τα βήματα έως ότου ο αλγόριθμος συγκλίνει.

Το βήμα 1) δίνει και το όνομα στην μέθοδο π.χ.

Εάν $s^{(k)} = -g^{(k)}$ τότε η μέθοδος που προκύπτει ονομάζεται: "μέθοδος της κλίσης πτώσης" (steepest descent).

Εάν $s^{(k)} = -B_k^{-1} g^{(k)}$ όπου B_k είναι μία προσέγγιση του $\nabla^2 f(x^{(k)})$ με $B_k > 0$ η μέθοδος λέγεται Newton ή παραλλαγή του Newton (modified Newton)

Εάν B_k είναι μία προσέγγιση του $\nabla^2 f$ από επικριτικές πρώτες ή δεύτερες βαθμίδες τότε η μέθοδος λέγεται Quasi-Newton ή "μεταβλητές Μετρικές".

Εάν $s^{(k)} = -g^{(k)} + \beta^{(k)} s^{(k-1)}$ η μέθοδος ($\beta^{(k)}$ είναι αριθμός) λέγεται:

"Συνjugate κλίσεων" (conjugate gradient).