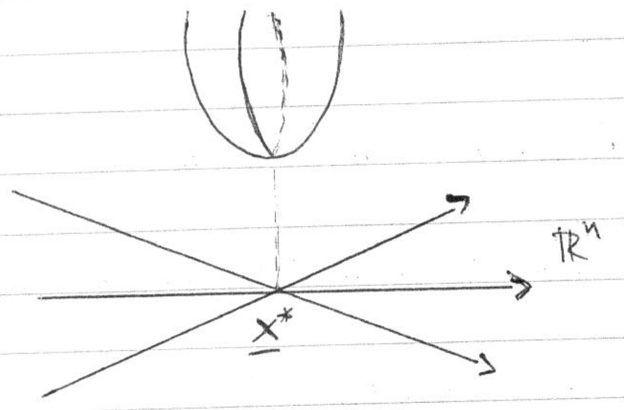


Επιλέγοντας $\underline{\alpha} = \pm e_i$, $i=1,..,n$ (e_i = μοναδιαία διανύσματα)
βρίσκουμε ότι $\nabla f(\underline{x}^*) = 0$

Επίσης αφού $g'_{\underline{\alpha}}(0)=0$, έχουμε:

$$g_{\underline{\alpha}}(t) - g_{\underline{\alpha}}(0) = 0 + \frac{1}{2} g''_{\underline{\alpha}}(0) \cdot t^2 + o(t^2) \quad , t \to 0.$$

ομοίως, $g''_{\underline{\alpha}}(0) = \underline{\alpha}^T \cdot Hf(\underline{x}^*) \cdot \underline{\alpha} \geq 0$, $\forall \underline{\alpha} \in \mathbb{R}^n$.

Άρα ο $Hf(x^*)$ είναι θετικά ημιορισμένος.

Τετάρτη 21/11/2012

Θεώρημα

1. Έστω $f \in C^2(\mathbb{R}^n)$ με τοπικό ελάχιστο x^*.
   Τότε: i) $\nabla f(x^*) = 0$
   ii) ο πίνακας $Hf(x^*)$ είναι θετικά ημιορισμένος.
   ($\underline{x}^T Hf(x^*) \underline{x} \geq 0$.)

2. Έστω $f \in C^2(\mathbb{R}^n)$ κ' ισχύουν τα εξής:
   i) $\nabla f(x^*) = 0$
   ii) ο πίνακας $Hf(x^*)$ είναι θετικά ορισμένος

Τότε το $\underline{x}^*$ είναι τοπικό ελάχιστο της f.

Απόδειξη

2. Έχαμε ότι $\forall \underline{\alpha}$

$$f(\underline{x}^*+\underline{\alpha}) = f(\underline{x}^*) + \underline{\alpha}\, \overbrace{\nabla f(\underline{x}^*)}^{(i)\to 0} + \frac{1}{2}\underline{\alpha}^T \underbrace{Hf(\underline{x}^*)}_{(ii)}\underline{\alpha} + o(\|\alpha\|^2)$$

$$\underset{\text{Rayleigh}}{\overset{\text{θεώρημα}}{\geq}} \lambda_1 \underline{\alpha}^T\underline{\alpha} + o(\|\alpha\|^2) + f(x^*)$$

Θεώρημα Rayleigh

Αν ο A είναι θετικά ορισμένος, τότε $\lambda_1 \leq \dfrac{\underline{\alpha}^T A\underline{\alpha}}{\underline{\alpha}^T\underline{\alpha}} \leq \lambda_n$

όπου $0 \leq \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$

οπότε για $\underline{\alpha}$ με $\|\underline{\alpha}\| < \varepsilon$, για κάποιο ε έχουμε

$f(\underline{x}^*+\underline{\alpha}) \geq f(\underline{x}^*)$

! Τα τοπικά μέγιστα / ελάχιστα είναι κι ολικά αν η f είναι κοίλη/κυρτή.

Θεώρημα

Έστω η f κυρτή σ' ένα σύνολο $F \subset \mathbb{R}^n$.

Τότε κάθε τοπικό ελάχιστο είναι κι ολικό.

Απόδειξη

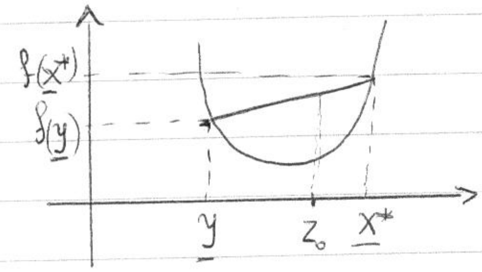
Έστω ότι το τοπικό ελάχιστο x^* δεν είναι ολικό, δηλαδή υπάρχει ένα $\underline{y}$ με $f(\underline{y}) < f(\underline{x}^*)$

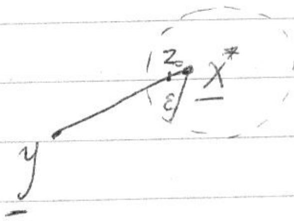
Παίρνουμε ένα σημείο $\underline{z}_0 : \|\underline{z}_0 - \underline{x}^*\| < \varepsilon$, $\underline{z}_0 = \lambda\underline{x}^* + (1-\lambda)\underline{y}$, $\lambda \in (0,1)$.

Αλλά η f είναι κυρτή, οπότε $f(\underline{z}_0) = f(\lambda\underline{x}^* + (1-\lambda)\underline{y}) \underset{\uparrow}{<} \lambda f(\underline{x}^*) + (1-\lambda) f(\underline{y}) = f(\underline{x}^*)$

f κυρτή

κάθε σημείο του ευθυγρ. τμήματος είναι πάνω από την f.





$\rightarrow f(z_0) < f(\underline{x}^*)$ Άτοπο

αφού υποθέσαμε ότι $z_0 \in$ ε-περιοχή του x^* κι' x^* τοπικό ελάχιστο.

Θεώρημα

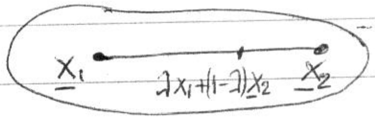
Έστω $g: \mathbb{R}^n \to \mathbb{R}$ κυρτή συνάρτηση

Τότε το σύνολο $F = \{\underline{x} \in \mathbb{R}^n : g(x) \leq c\}$ για κάποιο $c > 0$, είναι κυρτό.

Απόδειξη

Έστω $\underline{x}_1, \underline{x}_2 \in F$, τότε

$$g(\lambda x_1 + (1-\lambda) x_2) \leq \lambda g(\underline{x}_1) + (1-\lambda) g(\underline{x}_2)$$



$$\leq \lambda c + (1-\lambda) c = c.$$

Θεώρημα

Έστω $f \in C^2(\mathbb{R}^n)$

Η f είναι κυρτή ανν οι πίνακες $Hf(\underline{x})$ είναι θετικά ημιορισμένοι $\forall \underline{x} \in \mathbb{R}^n$.

Αλγόριθμοι Βαθμίδας

Θέλουμε να λύσουμε τις εξισώσεις $\nabla f(\underline{x}^*) = \underline{0}$
μέσω επαναληπτικής διαδικασίας που παράγει σημεία $(x_k)_{k=1,2,\ldots}$
τ.ω. $f(x_{k+1}) \leq f(x_k)$, $\forall k$ που ενδεχομένως τείνουν στο τοπικό
ελάχιστο $\underline{x}^*$ με $\|\nabla f(x^*)\| < \varepsilon$. ($\leftarrow$ κριτήριο τερματισμού)

Μια πρώτη ιδέα είναι να πάμε αντίθετα από $\nabla f(\underline{x}_k)$
Δηλαδή $\underline{x}_{k+1} = \underline{x}_k - \nabla f(\underline{x}_k)$
Ισχύει ότι $f(\underline{x}_{k+1}) \leq f(\underline{x}_k)$;
(Δηλαδή είναι μέθοδος καθόδου-descent method)
Από Taylor έχουμε:

$$f(\underline{x}_k + \underline{\xi}) = f(\underline{x}_k) + \underline{\xi}\, \nabla f(\underline{x}_k) + o(\|\xi\|),$$

όπου $\underline{\xi} = -\alpha \nabla f(x_k)$ αν α πολύ μικρό τότε ξ πολύ μικρό κ'

$$f(\underline{x}_k) = f(\underline{x}_k - \alpha \nabla f(\underline{x}_k)) + \alpha \nabla f(\underline{x}_k)^T \nabla f(\underline{x}_k) + \ldots$$
$$\geq f(\underline{x}_k - \alpha \nabla f(\underline{x}_k)), \text{ για } \alpha \text{ αρκετά μικρό}$$

όχι όμως υποχρεωτικά κ' για κάθε άλλο α.

Οπότε δοκιμάζουμε: $\underline{x}_{k+1} = \underline{x}_k - \alpha_k \nabla f(\underline{x}_k)$

ή ακόμη πιο γενικά: $\underline{x}_{k+1} = \underline{x}_k + \alpha_k \underline{d}_k$, για κάποιο $\underline{d}_k$ τ.ω.
$\underline{d}_k \cdot \nabla f(\underline{x}_k) < 0$

Επί παραδείγματι, μπορούμε να κινηθούμε κατά μήκος της ευθείας
από το $\underline{\mathbf{x}}_k$ με κατεύθυνση $\underline{d}_k$.
Έστω $\varphi(\alpha) := f(\underline{x}_k + \alpha \underline{d}_k)$ κ' διαλέγουμε το α τ.ω.
$\alpha_k := \inf\{\alpha > 0 : \varphi'(\alpha) = 0\}$.

Έχουμε: $\varphi(0) = \varphi(\alpha) - \alpha\cancel{\varphi'(\alpha)}^{0} + \frac{\alpha^2}{2}\varphi''(\xi)$, $\xi \in (0,\alpha)$

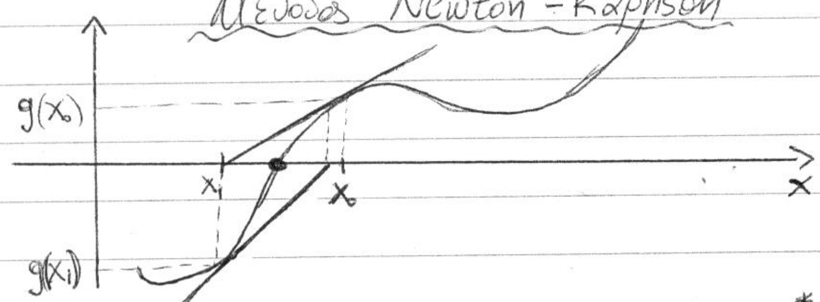
$\geq \varphi(\alpha)$, αν $\varphi''(\xi) \geq 0$, δηλ φ κυρτή

Μπορούμε επίσης να επιλέξουμε $\underline{d}_k = -\nabla f(\underline{x}_k)$
(steepest descent - απότομη κάθοδος)
Αλλά αυτές οι μέθοδοι έχουν αργή σύγκλιση.

<u>φτηνότερες μέθοδοι</u>
- με αργή σύγκλιση
- μέθοδοι αναζήτησης κατά μήκος μιας γραμμής

<u>ακριβότερες υπολογιστικά μέθοδοι</u>
- γρήγορη σύγκλιση
- μέθοδος Newton-Raphson

Μέθοδος Newton-Raphson



Έστω ότι θέλουμε να βρούμε τα μηδενικά x^* μιας συνάρτησης g:

$$g(x^*) \equiv \nabla f(x^*) = 0$$

Έστω μια αρχική τιμή x_k.
Χρησιμοποιούμε τη προσεγγιστική συνάρτηση (1η τάξη του Taylor)

$$\varphi: x \mapsto g(x_k) + g'(x_k)(x - x_k) \quad , x \in \mathbb{R}$$

ξεκινώντας από κάποιο x_k κ' βρίσκουμε πότε μηδενίζεται

$$x_{k+1} = x_k - \frac{g(x_k)}{g'(x_k)} \quad , g'(x_k) \neq 0.$$

Για $g : \mathbb{R}^n \to \mathbb{R}$, C^2, έχουμε την εξίσωση

$$g(\underline{x}_k) + J_g(\underline{x}_k)(\underline{x} - \underline{x}_k) = 0.$$

όπου $J_g(\underline{x}_k)$ είναι η Jacobian της g στο $\underline{x}_k$, οπότε αν είναι αντιστρέψιμος ($\det J_g \neq 0$) έχουμε:

$$\underline{x}_{k+1} = \underline{x}_k - \left(J_g(\underline{x}_k)\right)^{-1} g(\underline{x}_k)$$

ή

$$\underline{x}_{k+1} = \underline{x}_k - \left(Hf(\underline{x}_k)\right)^{-1} \nabla f(\underline{x}_k)$$