ΑΣΟΕΕ

ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ ΚΑΙ ΠΛΗΡΟΦΟΡΙΚΗΣ

ΣΗΜΕΙΩΣΕΙΣ

ΒΕΛΤΙΣΤΟΠΟΙΗΣΗΣ

ΒΑΓΓΕΛΗΣ Φ. ΜΑΓΕΙΡΟΥ

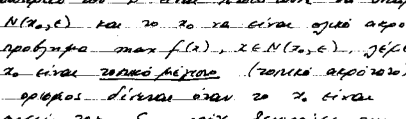
ΕΠΙΚΟΥΡΟΣ ΚΑΘΗΓΗΤΗΣ

ΑΘΗΝΑ 1986

Κεφάλαιο 1 Συναρτήσεις μιάς μεταβλητής

1. Εισαγωγή.

Θεωρούμε συναρτήσεις $f: S \to R$ με $S \subset R$. Αν $x_0 \in S$ τέτοιο ώστε $f(x_0) \geq f(x)$ $\forall x \in S$ λέμε ότι στο x_0 έχουμε ολικό μέγιστο (ολικό ελάχιστο αν $f(x_0) \leq f(x)$) ή γενικά ολικό ακρότατο. Αν επίσης x_0 στο εσωτερικό του S είναι τέτοιο ώστε να υπάρχει περιοχή $N(x_0, \epsilon)$ και το x_0 να είναι ολικό ακρότατο για το πρόβλημα $\max f(x)$, $x \in N(x_0, \epsilon)$, λέμε ότι το x_0 είναι τοπικό μέγιστο (τοπικό ακρότατο). Ανάλογος ορισμός δίνεται όταν το x_0 είναι ακραίο σημείο του S, οπότε θεωρούμε την περιοχή $N = N(x_0, \epsilon) \cap S$. Βλέπε διάγραμμα!



Διάγραμμα 1.

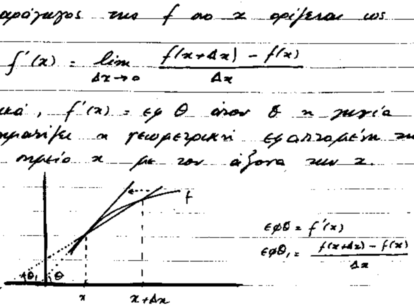
Φυσικά κάθε ολικό ακρότατο είναι και τοπικό. Συνήθως οι συνθήκες που βρίσκουμε αφορούν τοπικά ακρότατα. Η εύρεση ολικών ακροτάτων είναι πολύ πιό δύσκολη.

2. Αναγκαίες συνθήκες γιά τοπικά ακρότατα (Γ' Λυκείου κεφ. 7.13 κ.ε.)

Η παράγωγος της f στο x ορίζεται ως

$$f'(x) = \lim_{\Delta x \to 0} \frac{f(x+\Delta x) - f(x)}{\Delta x}$$

Γεωμετρικά, $f'(x) = \varepsilon\varphi\,\theta$ όπου θ η γωνία που σχηματίζει η γεωμετρική εφαπτομένη της f στό σημείο x με τον άξονα των x.



Θα υποθέσουμε ότι τό όριο $f'(x)$ υπάρχει γιά κάθε $x \in S$. Προφανώς, άν $f'(x) > 0$ περιμένουμε ότι η συνάρτηση αυξάνεται δεξιά του x και μειώνεται αριστερά του. Συγκεκριμένα έχουμε το

Λήμμα 1 Έστω $f'(x) > 0$. Υπάρχει $\epsilon > 0$ έτσι ώστε αν $x < y < x+\epsilon$, $f(x) < f(y)$. Επίσης αν $x - \epsilon < y < x$, $f(y) < f(x)$.

Απόδ. Εφόσον τό όριο $\lim_{\Delta x \to 0} \frac{f(x+\Delta x) - f(x)}{\Delta x} = f'(x)$ είναι θετικό, γιά $|\Delta x|$ αρκετά μικρό, έστω $|\Delta x| < \epsilon$, η τιμή του λόγου $\frac{\Delta f}{\Delta x} = \frac{f(x+\Delta x) - f(x)}{\Delta x}$

θα είναι επίσης θετικό, άρα $\frac{f(x+\Delta x)-f(x)}{\Delta x}>0$.

Αν $\Delta x>0$, αυτό σημαίνει ότι $f(x+\Delta x)>f(x)$. Αν όμως $\Delta x<0$ (και βέβαια $|\Delta x|<\epsilon$), η φορά της ανισότητας αντιστρέφεται εφόσον

$$\Delta x\,\frac{f(x+\Delta x)-f(x)}{\Delta x} \underbrace{<}_{\text{προσοχή}} 0\cdot\Delta x = 0$$

και άρα $f(x+\Delta x) < f(x)$.

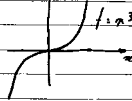
Ο.Ε.Δ.

<u>Πόρισμα 1</u> Αν $f'(x)<0$ υπάρχει $\epsilon>0$ τέτοιο ώστε $f(x)>f(y)$ για $x<y<x+\epsilon$ και $f(x)<f(y)$ για $x-\epsilon<y<x$.

<u>Θεώρημα 1.</u> Αν η f έχει τοπικό ακρότατο σε εσωτερικό σημείο του S, θα είναι $f'(x)=0$, όπου x το σημείο του ακροτάτου.

<u>Απόδ.</u> Αν το ακρότατο ήταν μέγιστο και $f'>0$ τότε για $\Delta x>0$ αρκετά μικρό θα είναι $f(x)<f(x+\Delta x)$ (Λήμμα 1). Αν ήταν $f'<0$ τότε για $|\Delta x|$ αρκετά μικρό και $\Delta x<0$ θα είναι $f(x+\Delta x)>f(x)$. Άρα θα πρέπει $f'(x)=0$, εφόσον δεν μπορεί να είναι ούτε $f'>0$ ούτε $f'<0$. Ανάλογη είναι η απόδειξη αν το ακρότατο είναι ελάχιστο.

Η συνθήκη f' είναι αναγκαία αλλά όχι και ικανή για ακρότατο. Παράδειγμα η $f(x)=x^3$ στο $x=0$, όπου $f'(0)=3x^2|_{x=0}=0$ αλλά δεν υπάρχει ούτε μέγιστο ούτε ελάχιστο στο $x=0$ (βλ. Διαγρ. 2)



Διαγρ. 2.

Για να αποκτήσουμε πιο ισχυρές αναγκαίες συνθήκες καθώς και άλλες που είναι ικανές ή αναγκαίες & ικανές χρειάζεται ορισμένη προεργασία.

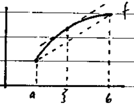
<u>Θεώρημα 2 (Rolle)</u> Έστω $f(a)=f(b)$. Υπάρχει κάποιο $x\in(a,b)$ με $f'(x)=0$

<u>Απόδ.</u> Αν η f είναι σταθερά ($f(x)=f(a)=f(b)$), θα ήταν $f'(x)=0$ $\forall x\in(a,b)$ και το θεώρημα αληθεύει. Αν όχι έστω ότι υπάρχει x με $f(x)>f(a)$. Τότε θεωρούμε το x_0 για το οποίο $f(x_0)\geq f(x)$ $\forall x\in[a,b]$. Τέτοιο x_0 υπάρχει εφόσον η f είναι συνεχής και ισχύει το βασικό Θεώρημα της σελ. 4 (Κεφ. 2), και θα είναι $x_0\neq a,b$. Στο x_0 θα είναι φυσικά $f'(x_0)=0$ βάσει του Θεωρ. 1, πράγμα που θέλαμε να αποδείξουμε. Αν τώρα υπάρχουν μόνο x με $f(x)\leq f(a)$, ισχύει και πάλι η ίδια απόδειξη μόνο που θεωρούμε

το x_0 στο οποίο υπάρχει το ελάχιστο της f στο $[a,b]$.
οεδ

Θεώρημα 3 (Μέσης Τιμής) Υπάρχει $\xi \in (a,b)$ έτσι ώστε $f(b)-f(a) = (b-a)\, f'(\xi)$

(Βλ. Διάγραμμα 3)



Διαγρ. 3

Αποδ. Θεωρούμε την συνάρτηση

$$m(x) = f(x) - \frac{f(b)-f(a)}{b-a}(x-a) - f(a).$$

Είναι $m(a)=0$, $m(b)=0$. Εφαρμόζοντας το Θ.1 στην m, έχουμε ότι $m'(\xi)=0$ για κάποιο $\xi \in (a,b)$. Αλλά $m'(x) = f'(x) - \frac{f(b)-f(a)}{b-a}$ και $m'(\xi)=0$ συνεπαίνει ότι $f'(\xi) = \frac{f(b)-f(a)}{b-a}$.

οεδ.

Το θεώρημα της μέσης τιμής έχει πολύ σημαντικές επιπτώσεις καθώς μας επιτρέπει να μεταφέρουμε τις γνώσεις μας για την f' σε στοιχεία για την ίδια την f:

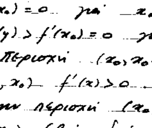
Πόρισμα 2 Αν $f'(x) \geq 0$ $\forall x \in S$ τότε η f είναι αύξουσα στο S.
Αποδ. Έστω $x<y$. Τότε $f(y)-f(x) = f'(\xi)(y-x) \geq 0$, ή $f(y) \geq f(x)$. οεδ.

Πόρισμα 3 Αν $f'(x) \leq 0$ $\forall x \in S$, η f είναι φθίνουσα στο S.

Με βάση τα δύο πορίσματα μπορούμε να δώσουμε μια ικανή συνθήκη για ακρότατα:

Θεώρημα 4 Έστω $f'(x_0)=0$ και $f''(x_0)<0$. Το x_0 είναι τοπικό μέγιστο. Αν $f''(x_0)>0$, το x_0 είναι τοπικό ελάχιστο.

Αποδ. Εφόσον $f'(x_0)=0$ και $f''(x_0)<0$, δεξιά τού x_0 και για κάποια απόσταση ε θα είναι $f'(y) < f'(x_0)=0$ για $x_0<y<x_0+\varepsilon$, ενώ θα είναι $f'(y) > f'(x_0)=0$ για $x_0-\varepsilon<y<x_0$. Άρα στην περιοχή $(x_0, x_0+\varepsilon)$ $f'(x)<0$ και στην $(x_0-\varepsilon, x_0)$ $f'(x)>0$. Επομένως η f είναι φθίνουσα στην περιοχή $(x_0, x_0+\varepsilon)$ και αύξουσα στην $(x_0-\varepsilon, x_0)$ (Βλέπε διαγρ. 4), βάσει των πορισμάτων 2 και 3. Από το διαγρ. 4 φαίνεται ότι στο x_0 έχουμε τοπικό μέγιστο. Η απόδειξη για $f''>0$ είναι παρόμοια.
οεδ.



Η συνθήκη του Θ.4 δεν είναι αναγκαία. Παράδειγμα η $f(x)=x^4$ που έχει ελάχιστο (ολικό) στο $x=0$, αλλά $f''(0)=12x^2=0 \not> 0$. Όμως με βάση το Θ.4 μπορούμε να δώσουμε μια πολύ ισχυρότερη αναγκαία συνθήκη απ' αυτήν του Θ.1

<u>Θεώρημα 5</u> Έστω ότι η f έχει τοπικό μέγιστο στο x_0. Τότε θα είναι $f'(x_0)=0$ και $f''(x_0)\leq 0$. (Αν x_0 τοπ. ελάχιστο, θα είναι $f'(x_0)=0$, $f''(x_0)\geq 0$).

<u>Απόδ.</u> Το ότι $f'=0$ έπεται από το Θ1. Αν τώρα ήταν $f''(x_0)>0$, σύμφωνα με το Θ4 θα είχαμε τοπικό ελάχιστο, και όχι μέγιστο όπως υποθέσαμε. Επομένως λοιπόν δεν μπορεί να είναι $f''>0$, θα πρέπει $f''\leq 0$. οεδ.

Διαγραμματικά, οι συνθήκες των Θ4 και Θ5 παριστάνονται ως εξής

$$\left(\Sigma 1:\ \begin{matrix} f'(x_0)=0 \\ f''(x_0)>0 \end{matrix}\right) \xrightarrow{\text{έπεται}} \left(\begin{matrix}\text{Στο } x_0 \\ \text{έχουμε τοπ.} \\ \text{ελάχιστο}\end{matrix}\right) \xrightarrow{\text{έπεται}} \left(\Sigma 2:\ \begin{matrix} f'(x_0)=0 \\ f''(x_0)\geq 0 \end{matrix}\right)$$

Οι συνθήκες Σ1 & Σ2 είναι σχεδόν οι ίδιες, και διαφέρουν μόνο στην περίπτωση που $f'(x_0)=f''(x_0)=0$. Γι' αυτές τις περιπτώσεις ισχύει το επόμενο θεώρημα, που δίνει αναγκαίες και ικανές συνθήκες για τοπικά ακρότατα

<u>Θεώρημα 6</u> Έστω $f^{(k)}(x_0)$ η παράγωγος k-τάξεως στο x_0. Έστω ότι $f^{(k)}(x_0)=0$ για $k=1,2,..,m$ και $f^{(m)}(x_0)\neq 0$. Αναγκαία και ικανή συνθήκη για μέγιστο (ελάχιστο) είναι (α) Το m να είναι άρτιος και (β) $f^{(m)}(x_0)<0$ ($f^{(m)}(x_0)>0$ για ελάχιστο).

Η απόδειξη του Θ.6 απαιτεί την χρησιμοποίηση των γενικευμένων μορφών του θεωρ. της μέσης τιμής που είναι γνωστά σαν Θεωρήματα Taylor. Προτού αναφερθούμε σ' αυτά, δίνουμε ορισμένα παραδείγματα εφαρμογής του Θ.6.

<u>Παράδειγμα 1.</u> $f(x)=x$, $f'(x)=1\neq 0$. Επομένως $m=1$, περιττός, η $f(x)=x$ <u>δεν</u> έχει ακρότατα.

<u>Παραδ. 2</u> Η $f(x)=x^3$ έχει $f'(0)=0$, $f''(0)=0$ $f^{(3)}(0)=6\neq 0$. Άρα στο $x=0$ δεν έχει τοπικό ακρότατο επειδή m: περιττός

<u>Παραδ. 3</u> Η $f(x)=x^4$ έχει $f'(0)=f''(0)=f'''(0)=0$ αλλά $f^{(4)}(0)=4!>0$. Άρα έχει τοπικό ελάχιστο στο $x=0$

<u>Άσκηση</u> Διερευνήσατε αν η συνάρτηση $f(x)=x^n$ έχει ακρότατο στο $x=0$ για $n=$ περιττός ακέραιος και $n=$ άρτιος ακέραιος.

3. Θεωρήματα προσέγγισης (Taylor)

(Yamane-Κιτσή Τ. 2, κεφ. 2)

Το θεώρημα της μέσης τιμής μάς δείχνει ότι σε πρώτη προσέγγιση, το $f(x+\Delta x)$ προσεγγίζεται από το $f(x)$, εφόσον $f(x+\Delta x)-f(x)=f'(\xi)\Delta x$ και άρα για μικρό Δx, $|f(x+\Delta x)-f(x)| \le M\cdot|\Delta x|$ όπου $M=\sup_{y\in(x,x+\Delta x)}|f'(y)|$ ($\approx$ περίπου $|f'(x)|$ για $|\Delta x|$ μικρό)

Με συνεχείς εφαρμογές του ίδιου θεωρήματος, ο βαθμός προσέγγισης μπορεί να βελτιωθεί έτσι ώστε το σφάλμα της προσέγγισης αντί να είναι ανάλογο του $|\Delta x|$ να είναι ανάλογο του $|\Delta x|^n$ $n=2,3,\dots$ Φυσικά όσο μεγαλύτερο το n, τόσο το $|\Delta x|^n$ γίνεται μικρότερο (για $|\Delta x|<1$) και έτσι το σφάλμα της προσέγγισης επίσης γίνεται μικρότερο. Στις προσεγγίσεις που δίνονται παρακάτω για το $f(x+\Delta x)$ χρησιμοποιούνται οι παράγωγοι στο x, $f'(x), f''(x), \dots, f^{(k)}(x)$.

Θεώρημα 1. Για κάποιο ξ από $(0,x_0)$ έχουμε

$$f(x_0)=f(0)+x_0 f'(0)+\tfrac{1}{2}x_0^2 f''(\xi)$$

Απόδειξη. Έστω $g(x)=f(x)-f(0)-x f'(0)-\frac{x^2}{2}k$ όπου k κάποιος αυθαίρετος αριθμός. Προφανώς $g(0)=0$, ενώ $g(x_0)\neq 0$. Μπορούμε όμως να διαλέξουμε το k έτσι ώστε $g(x_0)=0$ (το k διαλέγεται έτσι ώστε $0=f(x_0)-f(0)-x_0 f(0)-\frac{x_0^2}{2}k$)

Έστω k_0 αυτό το k για το οποίο $g(x_0)=0$. Εφαρμό-

ζοντας το Θ. του Rolle, υπάρχει $\psi\in(0,x_0)$ ώστε $g'(\psi)=0=f'(\psi)-f'(0)-\psi\hat{k}$, η $\hat{k}=\frac{f'(\psi)-f'(0)}{\psi}$

Εφαρμόζοντας το Θ. μέσης τιμής στο δεξί μέρος της εξίσωσης, έχουμε ότι $\exists\,\xi\in(0,\psi)$ έτσι ώστε $\hat{k}=\frac{f'(\psi)-f'(0)}{\psi}=(f')'(\xi)=f''(\xi)$. Άρα

η αρχική σχέση $g(x_0)=0$ γράφεται

$$0=f(x_0)-f(0)-x_0 f'(0)-\tfrac{1}{2}x_0^2 f''(\xi) \quad \text{ή}$$

$$f(x_0)=f(0)+x_0 f'(0)+\tfrac{1}{2}x_0^2 f''(\xi)$$

ο.ε.δ.

Θεώρημα 2 (Taylor) Έστω ότι υπάρχουν οι παράγωγοι $f^{(k)}(x)$ για $x\in(0,x_0)$ και κάθε k. Τότε για κάποιο $\xi\in(0,x_0)$ έχουμε

$$f(x_0)=f(0)+x_0 f'(0)+\dots+\frac{x^k}{k!}f^{(k)}(0)+\frac{x^{k+1}}{(k+1)!}f^{(k+1)}(\xi).$$

Απόδ. (Προαιρετική) Γράφουμε $g(x)=\sum_{i=0}^{k}\frac{x^i f^{(i)}(0)}{i!}+\frac{x^{k+1}}{(k+1)!}R$

όπου R επιλέγεται έτσι ώστε $g(x_0)=f(x_0)$. Προφανώς $g(0)=f(0)$. Άρα η $m(x)=f(x)-g(x)$ είναι 0 στο $x=0$, $x=x_0$. Άρα υπάρχει ψ ώστε $m'(\xi)=0$. Αυτό σημαίνει ότι

$$R=\frac{f'(\psi)-\left[f'(0)+\psi f''(0)+\dots+\frac{\psi^{k-1}}{(k-1)!}f^{(k)}(0)\right]}{\psi^k}\,k!$$

Αν δείξουμε επαγωγικά ότι το Θ. Taylor ισχύει για $k-1$, τότε ο αριθμητής είναι $\frac{(f')^{(k)}(\xi)}{k!}$ και άρα υπάρχει ξ ώστε
$f^{(k+1)}(\xi) = R$

Εφόσον $g(x_0) = f(x_0)$, θα είναι ακριβώς

$$f(x_0) = \sum_{i=0}^{k} \frac{x^i f^{(i)}(0)}{i!} + \frac{x^{k+1}}{(k+1)!} f^{(k+1)}(\xi) .$$

σεα.

Τα παραπάνω θεωρήματα μπορούν να γραφούν και με την μορφή

$$f(x+\Delta x) = f(x) + f'(x)\cdot\Delta x + f''(x)\,\frac{\Delta x^2}{2} + \dots + \frac{f^{(k)}(x)\,\Delta x^k}{k!}$$

αν εφαρμόσουμε τα θεωρήματα 1 και 2 σε μία συνάρτηση $g(\Delta x) = f(x+\Delta x)$ (όπου Δx η ανεξάρτητη μεταβλητή, με το x να παίζει τον ρόλο μιας παραμέτρου).

<u>Εφαρμογή</u> Να εκτιμηθεί η τιμή του $e^{0.1}$
Εφόσον η e^x έχει όλες τις παραγώγους ίσες με e^x, και $e^0 = 1$, θα είναι

$$e^{0.1} = e^0 + 0.1\, e^0 + \frac{0.1^2}{2} e^0 + \dots + \frac{0.1^k}{k!} e^0 + \frac{0.1^{k+1}}{(k+1)!} e^{\xi}$$

με $\xi \in (0, 0.1)$. Εφόσον $e^{0.1} < e^1 < 3$, το σφάλμα της προσέγγισης αν πάρουμε k όρους στο ανάπτυγμα είναι το πολύ $\frac{0.1^{k+1}}{(k+1)}\cdot 3$. Για $k=3$, το σφάλμα είναι μικρότερο από $\frac{0.1^4}{4!}\cdot 3 = 1.25\times 10^{-5}$

Η προσέγγιση του $e^{0.1}$ είναι

$$e^{0.1} \approx 1 + 0.1 + \frac{0.1^2}{2} + \frac{0.1^3}{3!} = 1.10516666\ldots$$

ενώ η ακριβής τιμή είναι 1.105170918.
(με σφάλμα προσέγγισης -4.25×10^{-6}).

Μπορούμε τώρα να δούμε γιατί ισχύει η γενική αναγκαία και ικανή συνθήκη για ακρότατο.
Έστω ότι $f'(x_0) = \dots = f^{(k)}(x_0) = \dots = f^{(m-1)}(x_0) = 0$
Τότε

$$f(y) = f(x_0) + \frac{(y-x_0)^m}{m!} f^{(m)}(x_0) + \frac{(y-x_0)^{m+1}}{(m+1)!} f^{(m+1)}(\xi)$$

ή, διαιρώντας δια $(y-x_0)^m$,

$$\frac{f(y)-f(x_0)}{(y-x_0)^m} = \frac{f^{(m)}(x_0)}{m!} + \frac{(y-x_0)}{m} f^{(m+1)}(\xi)$$

Αυτό σημαίνει ότι $\lim_{y\to x_0} \frac{f(y)-f(x_0)}{(y-x_0)^m} = \frac{f^{(m)}(x_0)}{m!}$

Αν τώρα $f^{(m)}(x_0) > 0$ και m άρτιος, υπάρχει μία περιοχή $N(x_0,\epsilon)$ έτσι ώστε για κάθε y σ' αυτήν την περιοχή θα είναι

$$\frac{f(y)-f(x_0)}{(y-x_0)^m} = \frac{f^{(m)}(x_0)}{m!} > 0 \quad \text{ή εφόσον } (y-x_0)^m > 0$$

$$f(y) - f(x_0) > 0\cdot(y-x_0)^m = 0$$

άρα το x_0 είναι τοπικό ελάχιστο. Αν $f^{(m)}(x_0) < 0$, θα είχαμε τοπικό μέγιστο, με πάντα m = άρτιος. Αν όμως το m δεν ήταν άρτιο, τότε $(y-x_0)^m > 0$ για $y > x_0$ ενώ $(y-x_0)^m < 0$ για $y < x_0$. Αυτό σημαίνει ότι το $f(y)$ θα μπορούσε να γίνει είτε μικρότερο είτε μεγαλύτερο του $f(x_0)$ για y απείρως κοντά στο x, πράγμα που δείχνει ότι το x_0 δεν μπορεί να είναι ακρότατο.

4. Κυρτές και κοίλες Συναρτήσεις

(Yamane-Kinnis Τομ. Α., Κεφ. 6.2)

Μέχρι τώρα είδαμε συνθήκες γιά τοπικά ακρότατα. Γενικά η αναζήτηση ολικών ακροτάτων είναι δύσκολη και δεν υπάρχει κάποια ενιαία μέθοδος γιά αυτά. Θα εξετάσουμε εδώ μιά ειδική περίπτωση συναρτήσεων όπου τα τοπικά ακρότατα είναι και ολικά.

<u>Ορισμός</u> Μιά συνάρτηση $f(x)$ λέγεται <u>κυρτή</u> αν $f''(x) \ge 0$ γιά κάθε x στο πεδίο ορισμού της. Λέγεται <u>κοίλη</u> αν $f''(x) \le 0$ γιά κάθε x.

<u>Παρατήρηση</u> Οι συναρτήσεις $f(x)=ax+b$ (γραμμικές συναρτήσεις) είναι και κυρτές και κοίλες. Γιατί;

<u>Θεώρημα 1</u> Έστω f κυρτή στο S και $f'(x_0)=0$. Το x_0 είναι ολικό ελάχιστο. Αν η f είναι κοίλη, το x_0 είναι ολικό μέγιστο.

<u>Απόδ.</u> Γιά κάθε άλλο $y \in S$ έχουμε (Θεωρ. Taylor)

$$f(y) = f(x_0) + (y-x_0) f'(x_0) + \frac{1}{2}(y-x_0)^2 f''(\xi)$$

γιά κάποιο ξ μεταξύ του x_0 και y. Εφόσον όμως $f'(x_0)=0$ εξ υποθέσεως, έχουμε

$$f(y) - f(x_0) = \tfrac{1}{2}(y-x_0)^2 f''(\xi)$$

Γιά κυρτές συναρτήσεις $f''(\xi) \ge 0$, άρα

$f(y) - f(x_0) \ge 0$, άρα το x_0 είναι ολικό ελάχιστο

ενώ γιά κοίλες $f''(\xi) \le 0$, άρα

$f(y) - f(x_0) \le 0$ άρα το x_0 είναι ολικό μέγιστο.

<u>Πόρισμα 1</u> Έστω f κυρτή στο $[a,b]$. Το μέγιστο της f βρίσκεται είτε στο a είτε στο b.

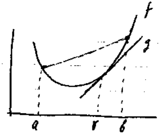
<u>Απόδ.</u> Εφόσον η f έχει κάποιο μέγιστο στο $[a,b]$, υποθέτουμε ότι το μέγιστο x_0 <u>δέν</u> είναι ούτε στο a ούτε στο b, δηλαδή $x_0 \in (a,b)$. Όμως από την αναγκαία συνθήκη θα πρέπει $f'(x_0)=0$. Σύμφωνα τότε με το Θεωρ. 1 στο x_0 η f έχει ολικό ελάχιστο, που αντιβαίνει την παραδοχή ότι στο x_0 έχουμε μέγιστο. Άρα $x_0 \notin (a,b)$ και είτε $x_0 = a$ είτε $x_0 = b$.

<u>Άσκηση 1</u> Έστω f κοίλη στο $[a,b]$. Αποδείξατε ότι το ελάχιστό της είναι είτε στο a είτε στο b.

<u>Άσκηση 2</u> Θεωρούμε την συνάρτηση $f(x)=x^2+x$ στο $[1/2, 3]$. Πού βρίσκεται το μέγιστό της; Πού βρίσκεται το ελάχιστό της;

<u>Άσκηση 3</u> Πού βρίσκεται το μέγιστο της $f(x) = x^3 + 2x^2$ γιά $x \in [-3/2, 1]$

Αν σχεδιάσουμε μία κυρτή συνάρτηση, παρατηρούμε τα εξής αξιοσημείωτα γεωμετρικά της στοιχεία (βλέπε Διάγρ. 4)

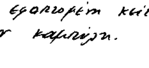


Διαγρ. 4

(α) Η τέμνουσα μεταξύ $f(a)$ και $f(b)$ κείται εντελώς μέσα (βρίσκεται πάνω) στην καμπύλη

(β) Η εφαπτομένη κείται παντού κάτω από την καμπύλη.

Το (β) είναι συνέπεια του εξής υπολογισμού: Η εξίσωση της εφαπτομένης στο γ είναι $g(x)=f(\gamma)+(y-\gamma)f'(\gamma)$. Επομένως $f(x)-g(x)=f(x)-f(\gamma)-(y-\gamma)f'(\gamma)=\frac{1}{2}(y-\gamma)^2 f''(\xi)$ που είναι μη αρνητικό (Θεωρ. Taylor). Άρα $f(x)\geq g(x)$.
Το (α) είναι συνέπεια του (β): (βλέπε Διαγρ. 5)



Διάγραμμα 5

Έχουμε ότι $A \sim f(a)$ $B \sim f(b)$. Αν $x=\lambda a+(1-\lambda)b$ ($\lambda\in(0,1)$) θα είναι $\Gamma=\lambda f(a)+(1-\lambda)f(b)$ (όμοια τρίγωνα..). Θα είναι επίσης $\lambda A'+(1-\lambda)B'=\Gamma'$ (όμοια τρίγωνα). Αλλά $A'\leq A$ $B'\leq B$ και άρα, εφόσον $\Gamma'=f(x)$, θα είναι

$$f(x)=f(\lambda a+(1-\lambda)b)\leq \lambda f(a)+(1-\lambda)f(b).$$

Η τελευταία σχέση

$$f(\lambda a+(1-\lambda)b)\leq \lambda f(a)+(1-\lambda)f(b)$$

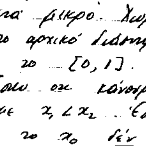
χρησιμοποιείται από πολλά συγγράμματα σαν ορισμός της κυρτότητος μιας συναρτήσεως.

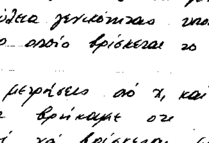
5. Αναζήτηση σε μία διάσταση (One dimensional Search)

Σε πολλές πρακτικές εφαρμογές της βελτιστοποιήσεως προκύπτει το παρακάτω πρόβλημα: Θέλουμε να μεγιστοποιήσουμε ένα μέγεθος, έστω M, που εξαρτάται από κάποια παράμετρο x. Όμως η συγκεκριμένη μορφή της εξάρτησης του M από το x <u>δεν</u> είναι γνωστή. Αυτό μπορεί να συμβεί όταν π.χ. το M συμβολίζει την απόδοση ενός μηχανήματος ενώ το x συμβολίζει την ρύθμιση κάποιας βαλβίδας του μηχανήματος. Η σχέση των M και x δεν είναι γνωστή εκ των προτέρων αλλά μπορεί κανείς να μετρήσει για δεδομένα $x_1, x_2, .., x_N$ τα αντίστοιχα $M(x_1), M(x_2), .., M(x_N)$ με πειραματικό τρόπο. Το πρόβλημα λοιπόν είναι να βρούμε το $\max M(x)$ χρησιμοποιώντας "λίγες" μετρήσεις, εφόσον κάθε μέτρηση $[x, M(x)]$ συνεπάγεται κάποιο κόστος.

Προφανώς αν η $M(x)$ είναι μια τελείως αυθαίρετη συνάρτηση, η θεωρία δεν μπορεί να βοηθήσει. Συχνά όμως είναι γνωστό ότι η $M(x)$ είναι <u>κοίλη</u> συνάρτηση, και αυτό γιατί κοίλες ή κυρτές συναρτήσεις απαντώνται συχνά σε διάφορα φυσικά φαινόμενα. Το πρόβλημα είναι πού θα γίνουν οι μετρήσεις $x_1, x_2, \dots, x_N$ έτσι ώστε το διάστημα αβεβαιότητας, δηλαδή το διάστημα $[L,R]$ μέσα στο οποίο βρίσκεται το μέγιστο x, έστω x_0 να γίνει αρκετά μικρό. Χωρίς απώλεια γενικότητας υποθέτουμε ότι το αρχικό διάστημα στο οποίο βρίσκεται το x_0 είναι το $[0,1]$.

Έστω ότι κάνουμε δύο μετρήσεις στο x_1 και στο x_2 με $x_1 < x_2$. Έστω ότι βρήκαμε ότι
Τότε το x_0 δέν μπορεί νά βρίσκεται μεταξύ 0 καί x_1. Γιατί τότε θά ήταν $x_0 < x_1 < x_2$ και $M(x_0) > M(x_1) < M(x_2)$, πράγμα που αντιβαίνει τήν υπόθεση ότι η M είναι κοίλη. Επίσης, αν είχαμε βρεί ότι $M(x_1) > M(x_2)$ συμπεραίνουμε με τόν ίδιο τρόπο ότι το x_0 πρέπει νά βρίσκεται μεταξύ 0 καί x_2. Βλέπε Διάγραμμα 6. Αν τέλος βρούμε ότι $M(x_1) = M(x_2)$ το x_0 θα πρέπει να είναι μεταξύ x_1 και x_2.





Διάγραμμα 6.

Με βάση τις παραπάνω παρατηρήσεις μπορεί κανείς με $N = 2k$ μετρήσεις να περιορίσει

το αρχικό διάστημα αβεβαιότητας από $(R-L)$ σε $(R-L)\cdot 2^{-k}$. Δηλαδή αν $R=1$, $L=0$ και κάνουμε $N=20$ μετρήσεις, το διάστημα που μπορεί να βρίσκεται το x_0 περιορίζεται από 1 σε $2^{-10} \approx 0{,}001$. Η διαδικασία περιγράφεται στον παρακάτω αλγόριθμο.

Αλγόριθμος Αναζήτησης με Διχοτόμηση

Βήμα 0 Έστω $L<R$ με L,R τα άκρα του διαστήματος όπου βρίσκεται το μέγιστο.

Βήμα 1. Υπολόγισε τα $x_1 = \frac{L+R}{2} - \epsilon$ $x_2 = \frac{L+R}{2} + \epsilon$ όπου ϵ πολύ μικρός αριθμός.

Βήμα 2 Μέτρησε τα $M_1 = M(x_1)$ και $M_2 = M(x_2)$

Βήμα 3 (α) Αν $M_1 < M_2$, το διάστημα αβεβαιότητας γίνεται $[x_1, R]$. Άρα θέσε $L \leftarrow x_1$, και πήγαινε στο Βήμα 4.
(β) Αν $M_1 > M_2$, το διάστημα αβεβαιότητας γίνεται $[L, x_2]$. Άρα θέσε $R \leftarrow x_2$ και πήγαινε στο 4
(γ) Αν $M_1 = M_2$ θέσε $L \leftarrow x_1$, $R \leftarrow x_2$

Βήμα 4 Αν $R-L < \epsilon$ [Επιθυμητή Ακρίβεια] τότε σταμάτα. Διαφορετικά πήγαινε στο Βήμα 1.

Παράδειγμα. Έστω ότι ήταν $M(x) = -x^2 + x$ και θέλουμε να βρούμε το μέγιστο με ακρίβεια στο $x \sim 0.1$ (Διάστημα αβεβαιότητας < 0.1). Το M θεωρούμε ότι δεν είναι γνωστό. Ο παραπάνω αλγόριθμος προχωρεί ως εξής, με αρχικό διάστημα $L=0$, $R=0,75$ και $\epsilon = 0.001$

	L	R	x_1	x_2	M_1	M_2
1	0	0.75	0.374	0.376	0.2341	0.2346
2	0.374	0.75	0.561	0.562	0.2463	0.2462
3	0.374	0.562	0.467	0.469	0.2489	0.2490
4	0.467	0.562	0.514	0.516	0.2498	0.2497
5	0.467	0.516				

Διάστημα αβεβαιότητας $0.049 < 0.1$

Στην πρώτη φάση $L=0$ $R=0.75$, άρα $x_1=0.374$ $x_2=0.376$. Εφόσον $M_1<M_2$, το διάστημα αβεβαιότητας γίνεται $[x_1, R]$ ή $[0.374, 0.75]$. Με το νέο αυτό διάστημα προχωρούμε στην φάση 2. Εφόσον $M_2<M_1$ το νέο διάστημα γίνεται $[L, x_2]$ ή $[0.374, 0.562]$. Τελικά στην 4 φάση το διάστημα αβεβαιότητας (αβεβαιότητας) γίνεται περίπου 0.1, ενώ στην 5η φάση γίνεται $0.049 < 0.1$

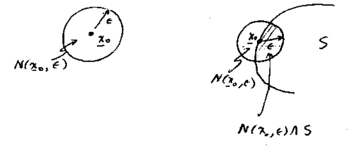
Άσκηση Γράψτε ένα πρόγραμμα σε PASCAL που να υλοποιεί τον παραπάνω αλγόριθμο.

Κεφάλαιο 2 Ακρότατα Συναρτήσεων Πολλών Μεταβλητών I. Ελεύθερα Ακρότατα.

1. Συναρτήσεις Πολλών Μεταβλητών

Μιά πραγματική συνάρτηση πολλών μεταβλητών συμβολίζεται σαν $f: S \to R$ όπου $S \subset R^n$. Εφόσον $S \subset R^n$, η f έχει n ανεξάρτητες μεταβλητές. Σε διανυσματικό συμβολισμό, γράφουμε $\underline{x} = (x_1, x_2, .., x_n)$ και $f(\underline{x}) = f(x_1, x_2, .., x_n)$. Σε περίπτωση που οι μεταβλητές είναι δύο, γράφουμε συνήθως $f(x,y)$ ενώ αν είναι τρείς γράφουμε $f(x,y,z)$.

Ένα σημείο $\underline{x}_0$ λέμε ότι είναι τοπικό μέγιστο αν υπάρχει περιοχή γύρω από το x_0, έστω $N(\underline{x}_0,\epsilon)$ έτσι ώστε $f(\underline{x}_0) \geq f(\underline{x})$ $\forall\, \underline{x} \in N(\underline{x}_0,\epsilon) \cap S$. Αντίστοιχα ορίζεται το τοπικό ελάχιστο. Μιλάμε για τοπικά ακρότατα όταν θέλουμε να αναφερθούμε είτε σε μέγιστα είτε σε ελάχιστα. Υπενθυμίζεται ότι μία περιοχή $N(\underline{x}_0,\epsilon)$ είναι όλα τα σημεία $\underline{x}$ που η απόστασή τους από το $\underline{x}_0$ είναι μικρότερη από ϵ. Βλ. Διαγρ. 1



Θα θεωρήσουμε στο μάθημα αυτό ότι οι f έχουν παραγώγους και μάλιστα είναι συνεχείς (Βλέπε Υ§Κ, Κεφ. 5, Γ.Α.) Υπενθυμίζουμε ότι η μερική παράγωγος της $f(x_1,\dots,x_n)$ ως προς την πρώτη μεταβλητή στο σημείο $x_1=a, x_2=b, x_3=\gamma$ κλπ. είναι

$$\frac{\partial f}{\partial x_1}\Big|_{(a,b,\gamma,\dots)} = \lim_{\Delta x_1\to 0} \frac{f(a+\Delta x_1, b,\gamma,\dots) - f(a,b,\gamma,\dots)}{\Delta x_1}$$

Ανάλογα ορίζονται και οι $\partial f/\partial x_2,\dots,\partial f/\partial x_n$.

Υπενθυμίζουμε την έννοια της ολικής παραγώγου (Υ§Κ Γ.Α. Κεφ. 5.6). Αν $x_1 = x_1(t)$, $x_2=x_2(t)$, $x_n=x_n(t)$ δηλαδή όλες οι "ανεξάρτητες" μεταβλητές εξαρτώνται από μία παράμετρο t, τότε η συνάρτηση $f(x_1(t), x_2(t),\dots,x_n(t))$ είναι συνάρτηση μόνο του t. Γράφουμε $\hat f(t) = f(x_1(t),\dots,x_n(t))$. Η ολική παράγωγος της $\hat f$ ως προς t είναι $d\hat f/dt$ και εκφράζεται ως

$$\frac{d\hat f}{dt} = \frac{\partial f}{\partial x_1}\frac{dx_1}{dt}+\dots+\frac{\partial f}{\partial x_n}\frac{dx_n}{dt}$$

<u>Παράδειγμα 1</u> Έστω $f(x,y)=x^2+y^2$ και $x=x(t)=t$, $y=y(t)=t^3$. Είναι τότε $\hat f(t) = f(x(t),y(t)) = f(t,t^3) = t^2+t^6$ και $\frac{d\hat f}{dt} = \frac{d}{dt}(t^2+t^6) = 2t+6t^5$. Επίσης είναι

$\frac{\partial f}{\partial x}=2x$, $\frac{\partial f}{\partial y}=2y$ και βέβαια $\frac{\partial f}{\partial x} = 2(t)=2t$

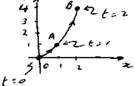
$\frac{\partial f}{\partial y} = 2y = 2(t^3)=2t^3$. Επίσης $\frac{dx}{dt} = \frac{dt}{dt}=1$ $\frac{dy}{dt} = \frac{dt^3}{dt}=3t^2$

Τέλος $\frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt} = (2t)(1)+(2t^3)(3t^2)$

$= 2t+6t^5$ που ισούται ακριβώς με την προηγούμενη έκφραση που βρήκαμε για την $d\hat f/dt$.

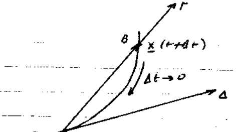
Τώρα η εξάρτηση των $x_1,x_2,\dots,x_n$ από το t ερμηνεύεται γεωμετρικά σαν μία καμπύλη στο R^n.

<u>Παράδειγμα 2</u> Έστω $x(t)=t$ $y(t)=t^2$. Για $t=0$ $(x(t),y(t))=0$, για $t=1$ $(x(1),y(1))=1$ για $t=2$ $(x(2),y(2))=(2,4)$ κλπ. Αν σχεδιάσουμε τα σημεία αυτά έχουμε το παρακάτω διάγραμμα.



καθώς το t μεταβάλλεται από 0 ως 2, το (x,y) κινείται από το 0 στο A στο B.

Θεωρούμε τώρα μια καμπύλη στο R^2 $\underline{x}(t) = (x(t),y(t))$, και θεωρούμε τα σημεία $\underline{x}(t)$ και $\underline{x}(t+\Delta t)$. Γεωμετρικά, το διάνυσμα που συνδέει τα $\underline{x}(t+\Delta t)$ και $\underline{x}(t)$ είναι η διαφορά $\underline{x}(t+\Delta t) - \underline{x}(t) = (x(t+\Delta t)-x(t), y(t+\Delta t)-y(t))$. Άρα ο λόγος $\frac{1}{\Delta t}(\underline{x}(t+\Delta t) - \underline{x}(t))$ είναι ένα διάνυσμα που είναι παράλληλο με το $\underline{x}(t+\Delta t) - \underline{x}(t)$. Βλέπε διάγραμμα 2



$\vec{AB} = \underline{x}(t+\Delta t) - \underline{x}(t)$

$\vec{A\Gamma} = \frac{1}{\Delta t}[\underline{x}(t+\Delta t) - \underline{x}(t)]$

Καθώς $\Delta t \to 0$ το $\vec{A\Gamma}$ τείνει στο $\vec{A\Delta}$

Διαγρ. 2.

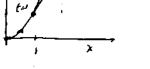
Από το διάγραμμα αυτό φαίνεται ότι καθώς $\Delta t \to 0$ το διάνυσμα $\frac{1}{\Delta t}[\underline{x}(t+\Delta t) - \underline{x}(t)]$

προσεγγίζει την εφαπτομένη της καμπύλης $\underline{x}(t)$ στο σημείο $A = \underline{x}(t)$. Δεδομένου ότι

$$\lim_{\Delta t \to 0} \frac{\underline{x}(t+\Delta t) - \underline{x}(t)}{\Delta t} = \left(\frac{dx}{dt}, \frac{dy}{dt}\right)$$

βλέπουμε ότι η διανυσματική παράγωγος $\left(\frac{dx}{dt}, \frac{dy}{dt}\right) \in R^2$ ερμηνεύεται σαν ένα εφαπτόμενο διάνυσμα της αρχικής καμπύλης, στο σημείο $\underline{x}(t) = (x(t), y(t))$.

<u>Παράδειγμα 3</u> Για $\underline{x}(t) = (t, t^2)$ $\underline{x}'(t) = (1, 2t)$. Για $t=1$ $\underline{x}'(1) = (1,2)$, που όντως είναι το εφαπτόμενο διάνυσμα στο $\underline{x}(t) = (1,1)$. ($\underline{x}'(t)$: παράγωγος)



Με βάση τα παραπάνω, ο τύπος $\frac{d\hat{f}}{dt} = \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt}$ ερμηνεύεται ως εξής: Θεωρούμε το διάνυσμα $\left(\frac{dx}{dt}, \frac{dy}{dt}\right)$ που εφάπτεται στην $\underline{x}(t)$. Θεωρούμε επίσης το διάνυσμα $\left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}\right) \in R^2$. Τότε η ολική παράγωγος γράφεται σαν εσωτερικό γινόμενο

$$\frac{d\hat{f}}{dt} = \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}\right) \cdot \left(\frac{dx}{dt}, \frac{dy}{dt}\right)$$

<u>Παράδειγμα 4</u> Αν $f = x^2 + y^2$, $\left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}\right) = (2x, 2y)$.

Αν $x(t) = t$ $y(t) = t^3$ $\frac{d\underline{x}}{dt} = (1,3)$ στο $t=1$ και $\underline{x}(1) = (1,1)$

Στο $\underline{x}(1) = (1,1)$ έχουμε $\left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}\right) = (2,2)$ και $\underline{x}' = (1,3)$.

οπότε $\frac{d\hat{f}}{dt} = (2,2)\cdot(1,3) = 2 + 2\cdot 3 = 8$, που είναι ακριβώς η παράγωγος που υπολογίστηκε στο παράδειγμα 1.

Δεδομένης μιας συναρτήσεως $f: R^n \to R$, ορίζουμε σαν <u>διάνυσμα βαθμίδος</u> της f στο σημείο $\underline{x} = (x_1, x_2, \ldots, x_n)$ το διάνυσμα $\left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \ldots, \frac{\partial f}{\partial x_n}\right) \in R^n$, όπου οι μερικοί παράγωγοι υπολογίζονται στο δεδομένο σημείο $\underline{x}$. Το σύμβολο ∇ ονομάζεται <u>ανάδελτα</u> (<u>ανάποδο</u> Δ) και συχνά ονομάζουμε το ∇f

απλώς – ανάλογα.

Παράδειγμα 5. Έστω $f(x,y,z) = xyz$. Ποιό είναι το $\nabla f \in R^3$ στο $x=1$ $y=2$ $z=3$; Έχουμε
$\nabla f = (yz, xz, xy)|_{x=1,y=2,z=3} = (6,3,2)$.

Παράδειγμα 6. Έστω $f(x_1,..,x_n) = \frac{1}{2}\sum_{i=1}^{n} x_i^2$. Ποιό είναι το $\nabla f (\in R^n)$ στο $x_1 = x_2 = \cdots = x_n = 2$; Έχουμε
$\partial f/\partial x_i = x_i$, άρα $\nabla f = (x_1, x_2,.., x_n) = (2,2,..,2)$.

Πολλά προβλήματα συναρτήσεων πολλών μεταβλητών ανάγονται σε προβλήματα μιάς μεταβλητής, αν χρησιμοποιήσουμε το παρακάτω τέχνασμα: Έστω ότι θέλουμε να συγκρίνουμε την τιμή της $f: R^2 \to R$ στο (x_0, y_0) με την τιμή της στο $(x_0+\Delta x, y_0+\Delta y)$. Θεωρούμε την συνάρτηση $(x(t), y(t)) = (x_0 + t\Delta x, y_0 + t\Delta y)$ με $t \in [-1,1]$. Προφανώς, γιά $t=0$, $(x(0), y(0)) = (x_0, y_0)$, γιά $t=1$ $(x(1), y(1)) = (x_0+\Delta x, y_0+\Delta y)$. Θεωρούμε τώρα την συνάρτηση του t $\hat{f}(t) = f(x(t), y(t)) = f(x_0+t\Delta x, y_0+t\Delta y)$. Έχουμε

$$\frac{d\hat{f}}{dt} = \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt} = \frac{\partial f}{\partial x}\Delta x + \frac{\partial f}{\partial y}\Delta y$$

όπου $\Delta x, \Delta y$ σταθερά. (Αυτό ισχύει εφόσον $\frac{dx(t)}{dt} = \frac{d}{dt}(x_0+t\Delta x) = \Delta x$ και $\frac{dy}{dt} = \frac{d}{dt}(y_0+t\Delta y) = \Delta y$).

Με βάση τώρα το θεώρημα της μέσης τιμής εφαρμοσμένο στην $\hat{f}(1), \hat{f}(0)$, έχουμε

$$f(x_0+\Delta x, y_0+\Delta y) - f(x_0,y_0) = \hat{f}(1) - \hat{f}(0) = \frac{d\hat{f}}{dt}(1-0)$$

$$= \frac{d\hat{f}}{dt}\Big|_{t=t^*} = \frac{\partial f}{\partial x}\Delta x + \frac{\partial f}{\partial y}\Delta y$$

όπου τα $\partial f/\partial x$, $\partial f/\partial y$ έχουν υπολογισθεί σε κάποιο $t^* \in (0,1)$ ή ισοδύναμα σε κάποιο $(x^*, y^*) = (x_0 + t^*\Delta x, y_0 + t^*\Delta y)$.

Με βάση το θεώρημα του Taylor, έχουμε επίσης

$$f(x_0+\Delta x, y_0+\Delta y) - f(x_0,y_0) = \frac{d\hat{f}}{dt}\Big|_{t=0} + \frac{1}{2}\frac{d^2\hat{f}}{dt^2}\Big|_{t=t^*}$$

δηλαδή η $\frac{d^2\hat{f}}{dt^2}$ είναι υπολογισμένη σε κάποιο $t^* \in (0,1)$ ενώ η $d\hat{f}/dt$ στο $t=0$.

Ο υπολογισμός της $d^2\hat{f}/dt^2$ γίνεται ως εξής:

$$\frac{d^2\hat{f}}{dt^2} = \frac{d}{dt}\left(\frac{d\hat{f}}{dt}\right) = \frac{d}{dt}\left(\frac{\partial f}{\partial x}\Delta x + \frac{\partial f}{\partial y}\Delta y\right)$$

$$= \Delta x \frac{d}{dt}\left(\frac{\partial f}{\partial x}\right) + \Delta y \frac{d}{dt}\left(\frac{\partial f}{\partial y}\right)$$

(Προσοχή τα $\Delta x, \Delta y$ είναι σταθερά και άρα δεν συμμετέχουν στην παραγώγιση).

Τώρα η $\frac{\partial f}{\partial x} = \frac{\partial f}{\partial x}(x,y)$ και άρα

$$\frac{d}{dt}\frac{\partial f}{\partial x} = \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial x}\right)\frac{dx}{dt} + \frac{\partial}{\partial y}\left(\frac{\partial f}{\partial x}\right)\frac{dy}{dt} =$$

$$= \frac{\partial^2 f}{\partial x^2}\Delta x + \frac{\partial^2 f}{\partial x \partial y}\Delta y$$ και αντίστοιχα

$$\frac{d}{dt}\frac{\partial f}{\partial y} = \frac{\partial^2 f}{\partial y^2}\Delta y + \frac{\partial^2 f}{\partial x \partial y}\Delta x$$

Τώρα η $d^2 f/dt^2$ γίνεται

$$d^2 f/dt^2 = \Delta x\left(\frac{d}{dt}\frac{\partial f}{\partial x}\right) + \Delta y\left(\frac{d}{dt}\frac{\partial f}{\partial y}\right) = \Delta x\left(\frac{\partial^2 f}{\partial x^2}\Delta x + \frac{\partial^2 f}{\partial x\partial y}\Delta y\right) +$$

$$+ \Delta y\left(\frac{\partial^2 f}{\partial y^2}\Delta y + \frac{\partial^2 f}{\partial x\partial y}\Delta x\right) = \frac{\partial^2 f}{\partial x^2}\Delta x^2 + 2\frac{\partial^2 f}{\partial x\partial y}\Delta x\Delta y + \frac{\partial^2 f}{\partial y^2}\Delta y^2$$

$$= (\Delta x, \Delta y)\begin{pmatrix} \partial^2 f/\partial x^2 & \partial^2 f/\partial x\partial y \\ \partial^2 f/\partial x\partial y & \frac{\partial^2 f}{\partial y^2}\end{pmatrix}\begin{pmatrix}\Delta x \\ \Delta y\end{pmatrix}$$

Τώρα, το θεώρημα του Taylor γιά δύο μεταβλητές γίνεται

$$f(x_0+\Delta x, y_0+\Delta y) = f(x_0,y_0) + \frac{\partial f}{\partial x}\Big|_{x_0,y_0}\Delta x + \frac{\partial f}{\partial y}\Big|_{x_0,y_0}\Delta y +$$

$$+ \frac{1}{2}(\Delta x, \Delta y)\, H(x^*, y^*)\begin{pmatrix}\Delta x \\ \Delta y\end{pmatrix}$$

όπου συμβολίζουμε με $H(x^*, y^*)$ την 2×2 μήτρα $\{h_{ij}\} = \{\partial^2 f/\partial x_i \partial x_j\}$ $(x_1 = x, x_2 = y)$

όπου οι μερικές παράγωγοι έχουν υπολογισθεί στο σημείο $x^* = x_0 + t^*\Delta x$, $y^* = y_0 + t^*\Delta y$, $t^* \in (0,1)$. Η μήτρα H ονομάζεται μήτρα του Έσ (Εσιανή μήτρα). (Hessian).

Γιά συναρτήσεις n μεταβλητών έχουμε

$$f(\underline{x} + \underline{\Delta x}) = f(\underline{x}) + \nabla f\Big|_{\underline{x}} \cdot \underline{\Delta x} + \frac{1}{2}\underline{\Delta x}'\, H_{\underline{x}^*}\, \Delta x$$

όπου το $\nabla f = (\partial f/\partial x_1, \dots, \frac{\partial f}{\partial x_n})$ έχει υπολογισθεί στο $\underline{x}$, ενώ η μήτρα του Έσ

$H = \{h_{ij}\} = \{\partial^2 f/\partial x_i \partial x_j\}$ έχει υπολογισθεί σε κάποιο σημείο $\underline{x}^*$. Ας σημειωθεί ότι εφόσον $\underline{x}^* = (\underline{x} + t^*\underline{\Delta x})$, το $\underline{x}^*$ βρίσκεται επάνω στο ευθύγραμμο τμήμα που ενώνει το $\underline{x}$ με το $\underline{x} + \underline{\Delta x}$.

<u>Παράδειγμα 7.</u> Οι παρακάτω προσεγγίσεις δείχνουν ότι γιά μικρά $\Delta x, \Delta y$, η συνάρτηση $f(x+\Delta x, y+\Delta y)$ προσεγγίζεται από την $f(x,y) + \frac{\partial f}{\partial x}\Delta x + \frac{\partial f}{\partial y}\Delta y$

Έστω $f = xy$, και θέλουμε νά την υπολογίσουμε στο $x = 1.1$ $y = 1.1$. Θά είναι

$$f(1.1, 1.1) \doteq f(1,1) + \frac{\partial f}{\partial x}\, 0.1 + \frac{\partial f}{\partial y}\, 0.1$$

$$= xy\Big|_{1,1} + y\Big|_{y=1} 0.1 + x\Big|_{x=1} 0.1 \doteq 1.20$$

Η πραγματική τιμή της $f(1.1, 1.1) = 1.21$.

<u>Παράδειγμα 8.</u> Γιά $f = xy$, η μήτρα του Έσ είναι η $H = \begin{pmatrix} 0 & 1 \\ 1 & 0\end{pmatrix}$, εφόσον $\frac{\partial^2 (xy)}{\partial x^2} = 0 = \frac{\partial^2(xy)}{\partial y^2}$

ενώ $\frac{\partial^2}{\partial x \partial y}(xy) = \frac{\partial}{\partial x}\left(\frac{\partial}{\partial y}(xy)\right) = \frac{\partial}{\partial x}(x) = 1$

<u>Παράδειγμα 9.</u> Αν $f(x,y) = \frac{1}{2}(x^2 + y^2)$, έχουμε $\frac{\partial^2 f}{\partial x^2} = 1 = \frac{\partial^2 f}{\partial y^2}$ ενώ $\frac{\partial^2 f}{\partial x \partial y} = 0$. Άρα

$$H = \begin{pmatrix} 1 & 0 \\ 0 & 1\end{pmatrix}$$

Μία συνάρτηση $f(x,y)$ μπορεί να παρασταθεί γεωμετρικά με βάση τις ισοϋψείς καμπύλες (σταθερές καμπύλες, CONTOURS), που ορίζονται ως εξής

$$C_\alpha^f = \{ x,y \mid f(x,y) = \alpha \}$$

Παράδειγμα 10 Οι ισοϋψείς της $f(x,y) = \frac{1}{2}(x^2+y^2)$ είναι κύκλοι με κέντρο το $(0,0)$. Π.χ. η C_1^f είναι όλα τα σημεία που έχουν συντεταγμένες (x,y) τέτοιες ώστε $\frac{1}{2}(x^2+y^2)=1$ ή $x^2+y^2=2$, δηλαδή κύκλος με ακτίνα $\sqrt{2}$.

Αν τώρα μία καμπύλη $\underline{x}(t) = (x(t), y(t))$ βρίσκεται επάνω σε μία ισοϋψή C_α^f, θα είναι προφανώς

$$\frac{d\hat{f}}{dt} = \frac{d}{dt} f(x(t), y(t)) = \frac{d}{dt}(\alpha) = 0$$

εφόσον η τιμή της $\hat{f}$ είναι σταθερά για κάθε t. Άρα θα είναι

$$d\hat{f}/dt = \nabla f \cdot \frac{d\underline{x}}{dt} = 0$$

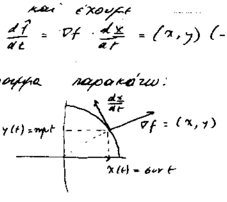
πράγμα που δείχνει ότι το ∇f και το $\frac{d\underline{x}}{dt}$ είναι κάθετα διανύσματα. Αλλά εφόσον το διάνυσμα $\frac{d\underline{x}}{dt}$ είναι εφαπτόμενο στην ισοϋψή, το ∇f θα είναι κάθετο στην ισοϋψή!

Παράδειγμα 11 Έστω $f(x,y) = \frac{1}{2}(x^2+y^2)$ και $\partial f/\partial x = x$, $\partial f/\partial y = y$, οπότε $\nabla f = (x,y)$

Θεωρούμε την ισοϋψή $C^f_{1/2}$, δηλαδή τα (x,y) με $x^2+y^2=1$. Προφανώς η καμπύλη $\underline{x}(t) = (\text{συν}\, t, \text{ημ}\, t)$ βρίσκεται επάνω στην ισοϋψή εφόσον $\text{συν}^2 t + \text{ημ}^2 t = 1$. Η εφαπτομένη της καμπύλης είναι $\frac{d\underline{x}}{dt} = (-\text{ημ}\, t, \text{συν}\, t) = (-y, x)$ και έχουμε

$$\frac{d\hat{f}}{dt} = \nabla f \cdot \frac{d\underline{x}}{dt} = (x,y)(-y,x) = -xy + xy = 0.$$

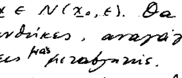
Βλέπε διάγραμμα παρακάτω:



2. Ακρότατα Συναρτήσεων.

Έστω ότι το $\underline{x}_0$ είναι τοπικό ακρότατο (έστω μέγιστο) της $f(\underline{x}) : R^n \to R$. Αυτό σημαίνει ότι υπάρχει μία περιοχή $N(\underline{x}_0, \epsilon)$ έτσι ώστε $f(\underline{x}_0) \ge f(\underline{x})$ για κάθε $\underline{x} \in N(\underline{x}_0, \epsilon)$. Θα βρούμε αναγκαίες και ικανές συνθήκες, ανάγοντας το πρόβλημα σε συναρτήσεις μιας μεταβλητής.

Θεωρούμε κάποιο $\underline{\Delta x}$ καθώς και την γραμμή $\underline{x}(t) = \underline{x}_0 + t\underline{\Delta x}$. Για αρκετά μικρό t θα είναι $\underline{x}(t) \in N(\underline{x}_0, \epsilon)$, εφόσον $\underline{x}(0) = \underline{x}_0$. Βλέπε διάγραμμα:



Για $t \in [-t_1, t_1]$ το $\underline{x}(t) = \underline{x}_0 + t\underline{\Delta x} \in N(\underline{x}_0, \epsilon)$

Αυτό σημαίνει ότι η συνάρτηση $\tilde{f}(t) = f(\underline{x} + t\underline{\Delta x})$ έχει τοπικό μέγιστο στο $t = 0$, εφόσον για όλα τα t που είναι αρκετά μικρά θα είναι $\underline{x}(t) \in N(\underline{x}_0, \epsilon)$ και άρα $\tilde{f}(0) = f(\underline{x}_0) \ge f(\underline{x}(t)) = \tilde{f}(t)$. Αναγκαίες συνθήκες για μέγιστο της $\tilde{f}(t)$ στο $t = 0$ είναι (a) $d\tilde{f}/dt = 0$ και (b) $d^2\tilde{f}/dt^2 \le 0$. Όμως με βάση τα προηγούμενα έχουμε

$$d\tilde{f}/dt = \nabla f \cdot \underline{\Delta x} \qquad d^2\tilde{f}/dt^2 = \underline{\Delta x}' \cdot H \cdot \Delta x$$

όπου $\nabla f = \left(\frac{\partial f}{\partial x_1}, \dots, \frac{\partial f}{\partial x_n}\right)$ $\quad H = \{h_{ij}\} = \{\partial^2 f/\partial x_i \partial x_j\}$

Μπορούμε να διατυπώσουμε την παρακάτω αναγκαία συνθήκη.

Θεώρημα 1. Έστω ότι η $f : R^n \to R$ έχει τοπικό μέγιστο στο $\underline{x}_0$. Τότε είναι

(a) $\forall i$, $\partial f/\partial x_i = 0$, όπου οι παράγωγοι υπολογίζονται στο $\underline{x}_0$

(b) $H = \{\partial^2 f/\partial x_i \partial x_j\}$ είναι αρνητικά ημιορισμένη όπου οι $\partial^2 f/\partial x_i \partial x_j$ υπολογίζονται στο $\underline{x}_0$.

Απόδειξη Θα πρέπει $d\tilde{f}/dt|_{t=0} = 0$ ή

$\sum_{i=1}^{n} \partial f/\partial x_i \, \Delta x_i = 0$ για κάθε (αυθαίρετο) Δx_i.

Αυτό σημαίνει ότι $\partial f/\partial x_i = 0$ για κάθε i (Αν ήταν $\partial f/\partial x_j \ne 0$ για κάποιο j, πάρε $\Delta x_i = 0$ και $\Delta x_j \ne 0$, οπότε $\nabla f \cdot \underline{\Delta x} \ne 0 \ne d\tilde{f}/dt$ που αντιβαίνει την υπόθεση $d\tilde{f}/dt = 0$).

Εφόσον τώρα $d^2\tilde{f}/dt^2|_{t=0} \le 0$, θα πρέπει για τυχόν $\underline{\Delta x}$ να είναι

$$\underline{\Delta x}' \, H \, \Delta x \le 0$$

πράγμα που σημαίνει ότι η H είναι αρνητικά ημιορισμένη (βλέπε σημείωση σελ. 43 για ορισμό ημιορισμένης μήτρας). ΟΕΔ.

<u>Πόρισμα 1</u> Αν η $f: R^n \to R$ έχει τοπικό ελάχιστο στο $\underline{x}_0$ τότε $\forall i$ $\partial f/\partial x_i = 0$, και H θετικά ημιορισμένη. Όλα τα παραπάνω είναι υποχρεωτικά στο $\underline{x}_0$.

<u>Παράδειγμα 1</u> $f(x,y) = x^2+y^2$. Στο $x=y=0$ είναι $\nabla f = (2x, 2y) = (0,0)$ και $\partial^2 f/\partial x^2 = \partial^2 f/\partial y^2 = 2$ ενώ $\partial^2 f/\partial x \partial y = 0$ και $H = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$. Άρα στο $(0,0)$ ικανοποιούνται οι συνθήκες του Πορίσματος 1, δηλαδή οι αναγκαίες συνθήκες για τοπικό ελάχιστο.

<u>Παράδειγμα 2</u> $f(x,y) = xy$. Είναι $\nabla f = (y,x)$ και $\nabla f = \underline{0}$ συνεπάγεται $x=y=0$. Έχουμε επίσης $\partial^2 f/\partial x^2 = \partial^2 f/\partial y^2 = 0$ ενώ $\partial^2 f/\partial x \partial y = 1$. Άρα $H = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$. Παρατηρούμε ότι η δεύτερη κύρια ελάσσων είναι αρνητική, εφόσον $\begin{vmatrix} 0 & 1 \\ 1 & 0 \end{vmatrix} = -1$ και επομένως η H

δεν είναι ούτε θετικά ούτε αρνητικά ημι-ορισμένη. Άρα το $(x,y) = (0,0)$ δεν μπορεί να είναι ούτε τοπικό μέγιστο ούτε τοπικό ελάχιστο, γιατί αν ήταν θα παραβίαζε τις αναγκαίες συνθήκες του Θεωρ. 1 ή Πορίσματος 1.

<u>Σημείωση</u> Μια μήτρα $H: n \times n$ που είναι συμμετρική ($H = H'$) λέγεται <u>θετικά ορισμένη</u> αν για κάθε $\underline{x} \in R^n$ είναι $\underline{x}' H x > 0$ εφόσον βέβαια $\underline{x} \neq 0$. Μια μήτρα H είναι <u>θετικά ημιορισμένη</u> αν για

κάθε $\underline{x} \neq \underline{0}$ $(\in R^n)$ έχουμε $\underline{x}' H \underline{x} \geq 0$. Αντίστοιχα ορίζεται η αρνητικά ημιορισμένη μήτρα.

Αναγκαίες και ικανές συνθήκες για να είναι μία μήτρα θετικά ημιορισμένη είναι να είναι όλες οι κύριες ελάσσονες μεγαλύτερες ή ίσες του 0. Παραδείγματος χάριν, η $H = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$ είναι θετικά ημι-ορισμένη, εφόσον $H_1 = 1 > 0$ $H_2 = \begin{vmatrix} 1 & 1 \\ 1 & 1 \end{vmatrix} = 0$

Η διαφορά μιας ορισμένης από μιά ημι-ορισμένη μήτρα είναι ότι η ημιορισμένη μπορεί να έχει $x' H x = 0$ για κάποιο $x \neq 0$. Και τέτοιο δεν μπορεί να συμβαίνει σε ορισμένες μήτρες. Μια ορισμένη μήτρα είναι υποχρεωτικά αντιστρέψιμη, ενώ αυτό δεν ισχύει κατ' ανάγκη για μία ημιορισμένη. Φυσικά, οι ορισμένες μήτρες είναι και ημιορισμένες.

<u>Παράδειγμα 3</u> Έστω $f(x,y) = x^2 + xy + y^2$. Είναι $\nabla f = (2x+y, 2y+x)$. Άρα το <u>κρίσιμο σημείο</u>, όπου $\nabla f = (0,0)$ βρίσκεται στο $x=y=0$. Έχουμε $\partial^2 f/\partial x^2 = 2 = \partial^2 f/\partial y^2$ και $\partial^2 f/\partial x \partial y = 1$ για κάθε x,y. Άρα $H = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$. Έχουμε $H_1 = 2$, $H_2 = 3$

άρα $H > 0$ και ικανοποιούνται οι αναγκαίες συνθήκες για τοπικό ελάχιστο.

Έστω τώρα ότι η μήτρα H είναι (αυστηρά) θετικά ορισμένη σε κάποιο σημείο $\underline{x}_0$, για το οποίο είναι και $\nabla f|_{\underline{x}_0} = (\partial f/\partial x_i) = 0$. Τότε το σημείο $\underline{x}_0$ είναι τοπικό ελάχιστο:

<u>Θεώρημα 2</u> (Ικανές συνθήκες) Έστω $\nabla f = 0$ και $H > 0$ για κάποιο $\underline{x}_0$. Η f έχει τοπικό ελάχιστο στο $\underline{x}_0$.

<u>Απόδειξη</u> Θεωρούμε την γραμμή $\underline{x}(t) = \underline{x}_0 + t\,\Delta\underline{x}$ με $\Delta\underline{x}$ αυθαίρετο. Έστω $\hat{f}(t): R \to R$ με $\hat{f}(t) = f(x_1(t), \ldots, x_n(t))$. Θα είναι

$d\hat{f}/dt = \nabla f \cdot \Delta\underline{x}$ και το $t=0$ $\underline{x}(t) = \underline{x}_0$

και άρα $d\hat{f}/dt|_{t=0} = 0$. Επίσης στο $t=0$

$$d^2\hat{f}/dt^2|_{t=0} = \Delta\underline{x}' H \Delta\underline{x} > 0$$ εφόσον η H είναι θετικά ορισμένη. Αυτό σημαίνει ότι η $\hat{f}(t)$ έχει τοπικό ελάχιστο στο $t=0$, για οποιοδήποτε $\Delta\underline{x}$, και επομένως το $\underline{x}_0$ είναι τοπικό ελάχιστο

<u>Πόρισμα 2</u> Αν $\nabla f = 0$ και $H < 0$ στο $\underline{x}_0$, η f έχει τοπικό μέγιστο στο $\underline{x}_0$.

<u>Παράδειγμα 4</u> Έστω $f(x,y,z) = x^2 + 2y^2 + z^2 + xz + x + y$. Τα κρίσιμα σημεία βρίσκονται από τις συνθήκες $\partial f/\partial x = \partial f/\partial y = \partial f/\partial z = 0$, δηλαδή

$$2x + z + 1 = 0$$
$$4y + 1 = 0$$
$$2z + x = 0$$

και άρα $y = -1/4$, $x = -2/3$ $z = 1/3$.

Έχουμε $\partial^2 f/\partial x^2 = 2$ $\partial^2 f/\partial y^2 = 4$ $\partial^2 f/\partial z^2 = 2$ και $\partial^2 f/\partial x \partial z = 1$, $\partial^2 f/\partial x \partial y = \partial^2 f/\partial y \partial z = 0$. Άρα η H είναι

$$H = \begin{bmatrix} 2 & 0 & 1 \\ 0 & 4 & 0 \\ 1 & 0 & 2 \end{bmatrix}$$

οπότε $H_1 = 2$ $H_2 = \begin{vmatrix} 2 & 0 \\ 0 & 4 \end{vmatrix} = 8$ $H_3 = \begin{vmatrix} 2 & 0 & 1 \\ 0 & 4 & 0 \\ 1 & 0 & 2 \end{vmatrix} = 12 > 0$

Άρα το $(-2/3, -1/4, 1/3)$ είναι τοπικό ελάχιστο.

<u>Παράδειγμα 5</u> Έστω $\underline{x} = (x_1, \ldots, x_n)$ $Q = Q' = n \times n$ μήτρα και $\underline{b} \in R^n$. Έστω $f(\underline{x}) = \frac{1}{2}\underline{x}'Q\underline{x} + \underline{b}'\underline{x}$

Αποδεικνύεται ότι $\nabla f = (\partial f/\partial x_1, \ldots, \partial f/\partial x_n) = \underline{x}'Q + \underline{b}$ δηλαδή $\underline{x} = -Q^{-1}\underline{b}$. Επίσης αποδεικνύεται εύκολα ότι

$$H = Q.$$

Άρα αν $Q > 0$, το $\underline{x} = -Q^{-1}\underline{b}$ είναι τοπικό (και ολικό) ελάχιστο, ενώ αν $Q \le 0$, το $\underline{x} = Q^{-1}\underline{b}$ είναι μέγιστο.

<u>Σημείωση</u>. Αν $f(\underline{x}) = \frac{1}{2}\underline{x}' Q \underline{x} + \underline{b}'\underline{x}$, θα είναι σε μορφή αθροίσματος

$$f(\underline{x}) = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} x_i q_{ij} x_j + \sum_{i=1}^{n} b_i x_i$$

Επομένως $\quad \partial f/\partial x_k = \sum_{j=1}^{n} q_{kj} x_j + b_k$

και $\quad \partial^2 f/\partial x_k \partial x_\ell = q_{k\ell}$. Επομένως

$$\nabla f = (\partial f/\partial x_1, \dots, \partial f/\partial x_n) = \left(\sum_{j=1}^{n} q_{1j} x_j, \dots, \sum_{j=1}^{n} q_{nj} x_j\right) + (b_1, \dots, b_n)$$

$$= (Q\underline{x} + \underline{b})'$$ (Το ()' υποδηλώνει ανάστροφο).

Έχουμε επίσης $H = \{h_{ij}\} = \{\partial^2 f/\partial x_i \partial x_j\} = \{q_{ij}\}$

και άρα $H = Q$.

<u>Εφαρμογή</u>. Έστω $f(x_1, x_2, x_3, x_4) = \frac{1}{2}\underline{x}'\begin{pmatrix} 3 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 4 & 1 \\ 0 & 0 & 1 & 1 \end{pmatrix}\underline{x}' + \underline{x}'\begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}$

Το κρίσιμο σημείο βρίσκεται λύνοντας το σύστημα

$$\begin{pmatrix} 3 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 4 & 1 \\ 0 & 0 & 1 & 1 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}$$

και επειδή η μήτρα είναι θετικά ορισμένη, το κρίσιμο σημείο είναι τοπικό ελάχιστο.

Μια συνάρτηση $f(\underline{x})$ η οποία έχει $H = \{h_{ij}\} = \{\partial^2 f/\partial x_i \partial x_j\}$ θετικά ημιορισμένη για κάθε $\underline{x}$ ονομάζεται κυρτή συνάρτηση. Αντίστοιχα, αν η H είναι αρνητικά ημιορισμένη η f ονομάζεται κοίλη συνάρτηση.

<u>Θεώρημα 3</u>. Έστω $f: R^n \to R$ μία κυρτή συνάρτηση. Αν στο $\underline{x}_0$ είναι $\nabla f = 0$, τότε το $\underline{x}_0$ είναι ολικό ελάχιστο.

<u>Απόδειξη</u>. Αν $\nabla f = 0$, έχουμε για τυχόν $\Delta\underline{x}$ (Taylor)

$$f(\underline{x}_0 + \Delta\underline{x}) - f(\underline{x}_0) = \cancelto{0}{\nabla f}\,\Delta x + \Delta\underline{x}' H(\underline{x}^*)\Delta\underline{x} \geq 0$$

επειδή η H είναι παντού θετικά ημιορισμένη, και άρα η H στο x^* είναι επίσης θετικά ημιορισμένη. Άρα $f(\underline{x}_0 + \Delta\underline{x}) \geq f(\underline{x}_0)$ για τυχόν $\Delta\underline{x}$.

<u>Πόρισμα 3</u>. Αν $\nabla f = 0$ στο $\underline{x}_0$ και η f είναι κοίλη, τότε το $\underline{x}_0$ είναι ολικό μέγιστο.

Για τις κυρτές συναρτήσεις πολλών μεταβλητών ισχύουν οι βασικές ιδιότητες που είδαμε στις συναρτήσεις μιας μεταβλητής και συγκεκριμένα

(α) Το εφαπτόμενο επίπεδο κείται κάτω από την f

(β) Για $\underline{x}, \underline{y}$ και $\lambda \in R \in (0,1)$ ισχύει ότι

$$f(\lambda \underline{x} + (1-\lambda) \underline{y}) \le \lambda f(\underline{x}) + (1-\lambda) f(\underline{y}).$$

Η απόδειξη των παραπάνω γίνεται κατ' αναλογία με τις συναρτήσεις μιάς μεταβλητής.

3. Αναζήτηση σε πολλές διαστάσεις.

Έστω ότι δίνεται μιά συνάρτηση $f: R^2 \to R$ και θέλουμε νά βρούμε το μέγιστό της. Πολλές φορές η λύση των εξισώσεων $\nabla f = 0$ δεν είναι εύκολη και αναγκαζόμαστε νά βρούμε το μέγιστο αναζητώντας το σε διάφορες τιμές του $\underline{x}$. Μια απλή μέθοδος αναζήτησης στηρίζεται στην εξής παρατήρηση: Έστω ότι για μία συνάρτηση 2 μεταβλητών γνωρίζουμε την τιμή της στο (x_0, y_0), καθώς και τις τιμές των $\partial f/\partial x$, $\partial f/\partial y$ στο σημείο αυτό. Για μικρές αλλαγές στα x, y η τιμή της f δίνεται προσεγγιστικά από τον τύπο

$$f(x_0+\Delta x, y_0+\Delta y) \doteq f(x_0, y_0) + \frac{\partial f}{\partial x}\Delta x + \frac{\partial f}{\partial y}\Delta y.$$

Αν διαλέξουμε $\Delta x = \lambda\, \partial f/\partial x$, $\Delta y = \lambda\, \partial f/\partial y$, θα είναι (για $\lambda > 0$) όπου λ "μικρός" αριθμός

$$f(x_0+\Delta x, y_0+\Delta y) - f(x_0, y_0) \doteq \lambda\left(\left(\frac{\partial f}{\partial x}\right)^2 + \left(\frac{\partial f}{\partial y}\right)^2\right) > 0$$

και το νέο σημείο $(x_0+\Delta x, y_0+\Delta y)$ έχει καλύτερη

εφόσον το λ είναι αρκετά μικρό!

τιμή της f απ' ότι τό αρχικό (x_0, y_0). Λέμε ότι μετακινηθήκαμε κατά την διεύθυνση του $\nabla f = (\partial f/\partial x, \partial f/\partial y)$.

Ενδέχεται να μήν ισχύει η παραπάνω ανισότης γιά τό συγκεκριμένο λ που διαλέξαμε, εφόσον η προσέγγιση $f(x_0+\Delta x, y_0+\Delta y) - f(x_0, y_0) = \frac{\partial f}{\partial x}\Delta x + \frac{\partial f}{\partial y}\Delta y$ ισχύει μόνο γιά πολύ μικρά Δx, Δy. Μπορούμε τότε νά μειώσουμε το λ, θέτοντας $\lambda \leftarrow \lambda/2$ και νά ξαναδοκιμάσουμε αν ισχύει η ανισότης $f(x_0+\Delta x, y_0+\Delta y) - f(x_0, y_0)$. Γενικά, ελαττώνουμε το λ έως ότου ισχύσει ότι $f(x_0+\Delta x, y_0+\Delta y) > f(x_0, y_0)$.

Θέτουμε τώρα $x_0 \leftarrow x_0+\Delta x$, $y_0 \leftarrow y_0+\Delta y$ και επαναλαμβάνουμε την ίδια διαδικασία, μετακινούμενοι σε σημεία (x_0, y_0) για τα οποία η f αυξάνει. Όταν η επαναληπτική αυτή διαδικασία δεν επιτυγχάνει μεγάλες αλλαγές στο x_0, y_0 έχουμε προσεγγίσει το μέγιστο.

Ο παρακάτω αλγόριθμος μπορεί να χρησιμοποιηθεί γιά την αναζήτηση του μεγίστου.

<u>Βήμα 0</u> Διάλεξε ένα αρχικό σημείο x_0, y_0 και ένα αριθμό $\epsilon > 0$ (μικρό)

<u>Βήμα 1.</u> Υπολόγισε τα $\partial f/\partial x$, $\partial f/\partial y$ στο x_0, y_0.

Βήμα 2 Θέσε $\lambda = 1$

Βήμα 3 Θέσε $x_1 = x_0 + \lambda\, \partial f/\partial x$, $y_1 = y_0 + \lambda\, \partial f/\partial y$

Βήμα 4 Υπολόγισε την διαφορά $\Delta = f(x_1, y_1) - f(x_0, y_0)$

Βήμα 5 (α) Αν $\Delta > 0$, υπολόγισε την τιμή $E = \lambda \sqrt{\left(\frac{\partial f}{\partial x}\right)^2 + \left(\frac{\partial f}{\partial y}\right)^2}$. Αν $E < \epsilon$ σταμάτα. Διαφορετικά θέσε $x_0 \leftarrow x_1$, $y_0 \leftarrow y_1$ και πήγαινε στο Βήμα 1.
(β) Αν $\Delta \leq 0$, θέσε $\lambda \leftarrow \lambda/2$ και πήγαινε στο Βήμα 3.

Η συμπεριφορά του αλγορίθμου αυτού μπορεί να ελεγχθεί εφαρμόζοντάς τον σε συναρτήσεις για τις οποίες το μέγιστο είναι γνωστό. Δεδομένου ότι ο αλγόριθμος χρησιμοποιεί το ανάδελτα ∇f είναι γνωστός στα βιβλία σαν Αναζήτηση με Ανάδελτα (GRADIENT SEARCH). Είναι ο απλούστερος δυνατός αλγόριθμος αναζήτησης, και η συμπεριφορά του δεν είναι καλή.

Άσκηση Προγραμμάτισε τον αλγόριθμο για διάφορες συναρτήσεις f

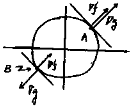
Άσκηση Γράψτε ένα αλγόριθμο αναζήτησης για να βρίσκετε το ελάχιστο μιας συνάρτησης.

Κεφάλαιο 3 Ακρότατα Συναρτήσεων Πολλών Μεταβλητών II
Ισοτικοί περιορισμοί

1. Γεωμετρική εισαγωγή.

Έστω ότι θέλουμε να βρούμε ένα ακρότατο της $f(x,y)$, όπου όμως τα x,y πρέπει να ικανοποιούν και την σχέση $g(x,y) = 0$. Αν είναι $S = \{x,y \mid g(x,y) = 0\}$ τότε το πρόβλημα τίθεται ως $\max f(x,y)$, $x,y \in S$. Θα μπορούσε κανείς να προχωρήσει ως εξής: Η σχέση $g(x,y) = 0$ προσδιορίζει, δεδομένου ενός x_0, κάποιο y_0 έτσι ώστε $g(x_0, y_0) = 0$ με $y_0 = y_0(x_0)$. Το αρχικό πρόβλημα τότε γίνεται $\max f(x_0, y_0(x_0))$ που είναι πάλι ένα πρόβλημα μεγιστοποίησης αλλά χωρίς περιορισμούς, εφόσον η $f(x_0, y_0(x_0))$ εξαρτάται από μία ανεξάρτητη μεταβλητή x_0, και μόνο αυτή. Η μέθοδος αυτή είναι όμως εφικτή για απλές f, g. Αποτυγχάνει σε περιπτώσεις όπου η $g(x,y)$ είναι μία πολύπλοκη συνάρτηση καθώς και σε περιπτώσεις πολλών μεταβλητών, με πολλούς ισοτικούς περιορισμούς. Παρακάτω θα αναπτύξουμε μια μέθοδο που βρίσκει αναγκαίες και ικανές συνθήκες για ακρότατα που δεν απαιτούν την λύση δευτερευουσών εξισώσεων όπως η $g(x,y) = 0$. Παράλληλα η μέθοδος αυτή οδηγεί σε πολύ χρήσιμα συμπεράσματα για την φύση του ελαχίστου – πώς αλλάζει συνάρτηση παραμέτρων, πώς μπορούμε να διατυπώσουμε μεθόδους αναζήτησης κλπ.

Ας θεωρήσουμε το απλό πρόβλημα max $x+y$ έτσι ώστε $x^2+y^2=1$ ($g(x,y) = x^2+y^2-1$, $f(x,y) = x+y$) Αν σχεδιάσουμε το $S = \{x,y \,/\, x^2+y^2=1\}$ βλέπουμε ότι είναι περιφέρεια κύκλου. Το μέγιστο σημείο είναι προφανώς το σημείο όπου ο κύκλος εφάπτεται με την γραμμή $x+y =$ σταθ. Δεδομένου ότι έχουμε 2 σημεία επαφής, προφανώς από την φύση της $f(x,y) = x+y$ επιλέγουμε το άνω δεξιά.



Στο σημείο επαφής και εφόδου τα ∇f και ∇g είναι κάθετα στις αντίστοιχες ισοϋψείς, θα πρέπει να ισχύει για κάποιο αριθμό μ $\nabla f = \mu \nabla g$. Αυτό συμβαίνει διότι στα A, B οι ισοϋψείς εφάπτονται και άρα τα ανάδελτα (κάθετα στις ισοϋψείς) είναι παράλληλα διανύσματα, άρα το ένα είναι πολλαπλάσιο του άλλου. Στο A (μέγιστο) θα είναι $\mu > 0$ (γιατί;) ενώ στο B $\mu < 0$ (γιατί;) Η παραπάνω σχέση $\nabla f = \mu \nabla g$ μπορεί να γραφεί ως $\nabla(f - \mu g) = 0$. Συνήθως θέτουμε $\lambda = -\mu$ και εισάγουμε την συνάρτηση του Lagrange $\mathcal{L}(x,y,\lambda) = f(x,y) + \lambda g(x,y)$. Τότε η αναγκαία συνθήκη για ακρότατο γίνεται

$$\partial\mathcal{L}/\partial x = \partial\mathcal{L}/\partial y = \partial\mathcal{L}/\partial \lambda = 0 \quad (\text{γιατί;})$$

δηλαδή: Για κάποιο αριθμό λ θα ισχύουν οι τρεις εξισώσεις

$$\frac{\partial \mathcal{L}}{\partial x} = \frac{\partial f}{\partial x} + \lambda \frac{\partial g}{\partial x} = 0$$

$$\frac{\partial \mathcal{L}}{\partial y} = \frac{\partial f}{\partial y} + \lambda \frac{\partial g}{\partial y} = 0$$

$$\frac{\partial \mathcal{L}}{\partial \lambda} = g(x,y) = 0.$$

Υπάρχουν 3 εξισώσεις με 3 αγνώστους x, y, λ, η λύση των οποίων μας επιτρέπει να βρούμε διάφορα σημεία που ικανοποιούν τις αναγκαίες συνθήκες.

<u>Παράδειγμα 1.</u> $f(x,y) = x+y$, $g(x,y) = x^2+y^2-1$

$\mathcal{L}(x,y,\lambda) = x+y + \lambda(x^2+y^2-1)$

$\partial\mathcal{L}/\partial x = 1+2\lambda x$ $\quad \partial\mathcal{L}/\partial y = 1+2\lambda y$ $\quad \frac{\partial\mathcal{L}}{\partial\lambda} = x^2+y^2-1 = 0$

Από τις δύο πρώτες έχουμε $x = y = -1/2\lambda$. Αντικαθιστώντας στην τρίτη έχουμε $\frac{1}{4\lambda^2} + \frac{1}{4\lambda^2} = 1$ ή $\lambda^2 = \frac{1}{2}$ και $\lambda = \pm 1/\sqrt{2}$. Από την γεωμετρική ανάλυση θα είναι $\lambda < 0$ (γιατί); Άρα το σημείο $x = y = -\frac{1}{2}(-1/\sqrt{2}) = 1/\sqrt{2}$ ικανοποιεί τις συνθήκες για μέγιστο. Το σημείο που αντιστοιχεί σε $\lambda > 0$, $x = y = -1/\sqrt{2}$ τι αντιπροσωπεύει;

Παράδειγμα 2. min x^2+2y^2 εάν $x+y=1$.

Είναι $f(x,y)=x^2+2y^2$, $g(x,y)=x+y-1$, και

$\mathcal{L}(x,y,\lambda)=x^2+2y^2+\lambda(x+y-1)$, οπότε

$\partial\mathcal{L}/\partial x = 2x+\lambda=0$ $\quad \frac{\partial \mathcal{L}}{\partial y}=4y+\lambda=0$ $\quad \frac{\partial\mathcal{L}}{\partial\lambda}=x+y-1=0$

Από τις δύο πρώτες είναι $x=-\lambda/2$ $\ y=-\lambda/4$ και εφόσον $x+y=1 \to -(\lambda/2+\lambda/4)=1 \to \lambda=-4/3$.

Τέλος $x=2/3$ $\ y=1/3$.

Αν θέλαμε να λύσουμε το ίδιο πρόβλημα υπό την μορφή μεγιστοποίησης, μια μηχανική εφαρμογή της μεθόδου του Lagrange θα μας έδινε πάλι το ίδιο σημείο. Πού είναι το μέγιστο; Υπάρχει; Το παράδειγμα αυτό μας δείχνει ότι πρέπει να είμαστε ιδιαίτερα προσεκτικοί στην εφαρμογή αυτής της μεθόδου.

Η γενική μέθοδος του Lagrange περιγράφεται ως εξής: Έστω $f:R^n\to R$ και k συναρτήσεις $g_1:R^n\to R$, $g_2:R^n\to R$ ... $g_k:R^n\to R$. Θέλουμε να βρούμε τα ακρότατα της $f(\underline{x})$ για $\underline{x}$ τέτοια ώστε $g_i(\underline{x})=0$ $\ i=1,2,\dots,k$.

Παράδειγμα 3 $\quad f(x,y,z)=x^2+y^2+2z^2$ με
$g_1(x,y,z)=x^2+y^2+z^2-1$, $g_2(x,y,z)=x+y-z$

Με αντίστοιχα γεωμετρικά επιχειρήματα στις 3 διαστάσεις, που επεκτείνουμε στις n διαστάσεις, ορίζουμε την συνάρτηση Lagrange

$$\mathcal{L}(x_1,x_2,\dots,x_n;\lambda_1,\lambda_2,\dots,\lambda_k)=f(x_1,\dots,x_n)+\sum_{i=1}^{k}\lambda_i g_i(x_1,\dots,x_n).$$

Οι αναγκαίες συνθήκες είναι οι $n+k$ εξισώσεις

$$\partial\mathcal{L}/\partial x_j = \partial f/\partial x_j+\sum_{i=1}^{k} \lambda_i\, \partial g_i/\partial x_j=0 \qquad j=1,2,\dots,n$$

$$\partial\mathcal{L}/\partial\lambda_i = g_i(x_1,\dots,x_n)=0 \qquad i=1,2,\dots,k$$

δηλαδή $n+k$ εξισώσεις για $n+k$ αγνώστους

Παράδειγμα 3 (συνέχεια) $\mathcal{L}(x,y,z;\lambda_1,\lambda_2)=$

$= x^2+y^2+2z^2+\lambda_1(x^2+y^2+z^2-1)+\lambda_2(x+y-z)$.

$\partial\mathcal{L}/\partial x = 2x+2x\lambda_1+\lambda_2=0$

$\partial\mathcal{L}/\partial y = 2y+2y\lambda_1+\lambda_2=0$

$\partial\mathcal{L}/\partial z = 4z+2z\lambda_1-\lambda_2=0$

Από τις δύο πρώτες είναι $x=y=-\frac{\lambda_2}{2(1+\lambda_1)}$ ενώ

$z=\lambda_2/2(2+\lambda_1)$ $\quad$ (αν $\lambda_1\neq -1,-2$)

Εφόσον $x+y=z$ θα είναι $-\frac{2\lambda_2}{2(1+\lambda_1)}=\frac{\lambda_2}{2(2+\lambda_1)}$

και αν $\lambda_2\neq 0$, $\lambda_1=-5/3$. Από την $x^2+y^2+z^2=1$ έχουμε τέλος

$$2\left(\frac{3}{4}\lambda_2\right)^2+\left(\frac{3}{2}\lambda_2\right)^2=1 \quad \text{ή} \quad \lambda_2=\pm 4/\sqrt{54}$$

και $x=y=\pm\sqrt{\frac{1}{6}}=\frac{\pm 1}{\sqrt{6}}$ $\quad z=\pm\sqrt{\frac{2}{3}}$ ή $(1/\sqrt{6},1/\sqrt{6},\sqrt{2/3}),(-1/\sqrt{6},-1/\sqrt{6}$

Από τα παραπάνω παραδείγματα φαίνεται ότι δεν είναι σαφές από την αρχή αν τα σημεία που βρήκαμε $(1/\sqrt{6}, 1/\sqrt{6}, \sqrt{2/3})$ $-(1/\sqrt{6}, 1/\sqrt{6}, \sqrt{2/3})$ είναι μέγιστα, ελάχιστα ή ενδεχομένως σαγματικά σημεία. Είναι λοιπόν απαραίτητο να δοθούν πρόσθετα κριτήρια για τη μέθοδο του Lagrange, κριτήρια που να εμπλέκουν τις δεύτερες παραγώγους.

2. Αλγεβρική παρουσίαση συνθηκών πρώτης τάξεως για ακρότατα.

Θα αποδείξουμε το παρακάτω θεώρημα

<u>Θεώρημα 1.</u> Έστω η συνάρτηση $f(x_1, x_2, \ldots, x_n)$ $f: R^n \to R$ καθώς και οι συναρτήσεις $g_i(x_1, \ldots, x_n)$ για $i = 1, 2, \ldots, k$ με $g_i : R^n \to R$. Έστω $S = \{ \underline{x} = (x_1, \ldots, x_n) \,/\, g_i(\underline{x}) = 0 \text{ για κάθε } i \}$.

Αν τώρα το $\underline{x}^\circ \in S$ είναι ακρότατο, δηλαδή είναι είτε $f(\underline{x}^\circ) \geq f(\underline{x})$ είτε $f(\underline{x}^\circ) \leq f(\underline{x})$ για κάθε $\underline{x} \in S$, υπάρχουν αριθμοί $\lambda_1, \lambda_2, \ldots, \lambda_k$ τέτοιοι ώστε

$$\frac{\partial}{\partial x_i} \left[f(x_1, \ldots, x_n) + \sum_{i=1}^{k} \lambda_i g_i(x_1, \ldots, x_n) \right] = 0$$

στο σημείο $\underline{x}^\circ$.

<u>Απόδειξη</u> (Προαιρετική) Υποθέτουμε ότι το $\underline{x}^\circ$ είναι μέγιστο. Θεωρούμε τώρα μία καμπύλη $\underline{x}(t)$ ως προς μία παράμετρο t έτσι ώστε $\underline{x}(0) = \underline{x}^\circ$ και $\underline{x}(t) \in S$ για $t \in N(0, \epsilon)$ — δηλαδή για t "κοντά" στο 0. Εφόσον $\underline{x}(t) \in S$, θα είναι $g_i(\underline{x}(t)) = 0$ για κάθε $t \in N$ και άρα $\frac{d}{dt} g_i(\underline{x}(t)) = 0$ στο $t = 0$ ή

$$\sum_{j=1}^{n} \partial g_i / \partial x_j \, \frac{dx_j}{dt} = 0 \qquad \forall\, i \qquad (1)$$

Εφόσον $f(\underline{x}^\circ) \geq f(\underline{x})$ $\forall\, \underline{x} \in S$, θα είναι και $f(\underline{x}(0)) \geq f(\underline{x}(t))$ για $t \in N$, και άρα

$$\frac{df(\underline{x}(t))}{dt} = \sum_{j=1}^{n} \partial f / \partial x_j \, \frac{dx_j}{dt} = 0 \quad \text{για } t = 0. \quad (2)$$

Οι (1) και (2) σημαίνουν ότι για κάθε διάνυσμα $y = \left(\frac{dx_1}{dt}, \ldots, \frac{dx_n}{dt}\right) \in R^n$ που ικανοποιεί το σύστημα

$$Gy = 0$$

με $G : k \times n$ μήτρα, $g_{ij} = \partial g_i / \partial x_j \,|_{\underline{x}^\circ}$

θα ικανοποιεί και την εξίσωση $Fy = 0$ όπου $F = (F_1, \ldots, F_n)$ το διάνυσμα, $F_i = \partial f / \partial x_i \,|_{\underline{x}^\circ}$.

Από συνεπάγεται ότι τα ομογενή συστήματα $Gy = 0$ και $\begin{bmatrix} G \\ F \end{bmatrix} y = 0$ έχουν τους <u>ίδιους</u> μηδενικούς χώρους. Το σύστημα $Gy = 0$ έχει διαστάσεις $k \times n$, το $\begin{bmatrix} G \\ F \end{bmatrix} y = 0$

έχει διαστάσεις $(k+1)\times n$. Αν n_1, r_1 η διάσταση του μηδενικού και του χώρου γραμμών του πρώτου και n_2, r_2 του δευτέρου, θα είναι

$$n_1 + r_1 = n_2 + r_2 = n \quad \text{(Γραμμική Άλγεβρα!)}.$$

Εφόσον όμως $n_1 = n_2$ θα έχουμε και $r_1 = r_2$ πράγμα που σημαίνει ότι η διάσταση των χώρων που παράγονται από τις γραμμές της G και της $\begin{bmatrix} G \\ F \end{bmatrix}$ είναι οι ίδιες. Ισοδύναμα, αυτό δείχνει ότι το F είναι γραμμικός συνδυασμός των γραμμών της G, ή τελικά $(\mu_1,\dots,\mu_k)G = F$ ή $\partial f/\partial x_i = \sum_{j=1}^{k} \mu_j \,\partial g_j/\partial x_i$. Θέτοντας $\lambda_j = -\mu_j$ έχουμε ότι

$$\frac{\partial}{\partial x_i}\left[f(x_1,\dots,x_n) + \sum_{j=1}^{k} \lambda_j g_j(x_1,\dots,x_n)\right] = 0$$

για $i = 1,2,\dots,n$. ο.ε.δ.

Παρατηρήσεις: (1) Μπορεί να αποδειχθεί ότι αν $Gy = 0$ για κάποιο y, υπάρχει καμπύλη $\underline{x}(t) \in S'$ έτσι ώστε $\frac{d}{dt}\underline{x}(t) = y$. Η απόδειξη αυτή παραλείπεται.

(2) Η μήτρα $G = \{g_{ij}\} = \{\partial g_i/\partial x_j\}$ που είναι διαστάσεως $k\times n$ μπορεί να θεωρηθεί ως η "παράγωγος" της συνάρτησης $\hat{g}: R^n \to R^k$ που ορίζεται ως $\{g_1(x_1,\dots,x_n), g_2(x_1,\dots,x_n),\dots,g_k(x_1,\dots,x_n)\} = \hat{g}(x_1,\dots,x_n) \in R^k$. Η G ονομάζεται η μήτρα του Jacobi (Γιακόμπι).

3. Συνθήκες δευτέρας τάξεως

Για να αποφανθεί αν κάποιο σημείο που προκύπτει από την μέθοδο του Lagrange είναι ακρότατο και το είδος του, θα χρειαστεί να χρησιμοποιηθούν οι δεύτερες παράγωγοι. Ο βασικός υπολογισμός που απαιτείται εδώ είναι η δεύτερη παράγωγος πάνω σε μία καμπύλη: Έστω $\underline{x}(t)$ μία καμπύλη ($\underline{x}\in R^n$) και $f: R^n \to R$. Είναι

$$\frac{df(\underline{x}(t))}{dt} = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}\frac{dx_i}{dt} \quad \text{και}$$

$$\frac{d^2 f(\underline{x}(t))}{dt^2} = \frac{d}{dt}\left(\frac{d}{dt}f(\underline{x}(t))\right) = \frac{d}{dt}\left[\sum_{i=1}^{n}\frac{\partial f}{\partial x_i}\frac{dx_i}{dt}\right] = \left(\begin{array}{l}\text{παράγωγος}\\ \text{γινομένου}\\ \text{- αθροίσματος}\end{array}\right)$$

$$= \sum_{i=1}^{n}\left[\frac{d}{dt}\left(\frac{\partial f}{\partial x_i}\right)\cdot\frac{dx_i}{dt} + \frac{\partial f}{\partial x_i}\frac{d^2 x_i}{dt^2}\right] =$$

$$= \sum_{i=1}^{n}\left[\sum_{j=1}^{n}\frac{\partial^2 f}{\partial x_i \partial x_j}\frac{dx_j}{dt}\cdot\frac{dx_i}{dt} + \frac{\partial f}{\partial x_i}\frac{d^2 x_i}{dt^2}\right]$$

$$= \left[\frac{dx_1}{dt},\dots,\frac{dx_n}{dt}\right]\cdot H\cdot\begin{bmatrix} dx_1/dt \\ \vdots \\ dx_n/dt\end{bmatrix} + \nabla f\begin{bmatrix} d^2x_1/dt^2 \\ \vdots \\ d^2x_n/dt^2\end{bmatrix}$$

όπου $H = \{h_{ij}\} = \{\partial^2 f/\partial x_i \partial x_j\}$, η μήτρα του Έσσε και

$$\nabla f = [\partial f/\partial x_1, \dots \partial f/\partial x_n] \quad \text{το διάνυσμα βαθμίδος.}$$

Θεωρούμε πάλι ένα σημείο $\underline{x}^{\circ}$ που όπως και στο προηγούμενο εδάφιο είναι ακρότατο ως προς το $S = \{\underline{x} \,/\, g_i(\underline{x}) = 0\}$, με $\underline{x}^{\circ} \in S$.

Αν $\underline{x}(t)$ είναι μία καμπύλη που διέρχεται από το $\underline{x}^\circ$ στο $t=0$, θα είναι $f(\underline{x}^\circ) \le f(\underline{x})$ αν το $\underline{x}^\circ$ είναι ελάχιστο και άρα στο $t=0$ η $f(\underline{x}(t))$ έχει ελάχιστο. Θα είναι λοιπόν $\frac{df(\underline{x}(t))}{dt}=0$ και $\frac{d^2f(\underline{x}(t))}{dt^2} \ge 0$ στο $t=0$

Από την συνθήκη του Lagrange θα έχουμε ότι $\nabla(f+\sum_{j=1}^{k} g_j)=0$ ή $\partial f/\partial x_i + \sum_{j=1}^{k} \lambda_j \partial g_j/\partial x_i = 0$.

Τώρα, θα είναι $d^2g_j/dt^2 = 0$ ή

$$0 = d^2g_j/dt^2 = \left[\frac{dx_1}{dt}, \dots\right] H^{g} \begin{bmatrix} dx_1/dt \\ \vdots \end{bmatrix} + \left[\frac{\partial g_j}{\partial x_1}, \dots\right] \begin{bmatrix} d^2x_1/dt^2 \\ \vdots \end{bmatrix} \quad (3)$$

όπου $H^{g_j} = \{h^{g_j}_{\mu\nu}\}$ η μήτρα του Es που αντιστοιχεί στην g_j. Θα είναι επίσης

$$0 \le d^2f/dt^2 = \left[\frac{dx_1}{dt}, \dots\right] H^{f} \begin{bmatrix} dx_1/dt \\ \vdots \end{bmatrix} + \left[\frac{\partial f}{\partial x_1}, \dots\right] \begin{bmatrix} d^2x_1/dt^2 \\ \vdots \\ d^2x_n/dt^2 \end{bmatrix} \quad (4)$$

Αν τώρα προσθέσουμε στην d^2f/dt^2 τα $\lambda_j\, d^2g_j/dt^2$, το αποτέλεσμα δεν αλλάζει εφόσον $d^2g_j/dt^2 = 0$. Άρα

$$d^2f/dt^2 = d^2f/dt^2 + \sum_{j=1}^{k} \lambda_j\, d^2g_j/dt^2 =$$
$$\left[\frac{dx_1}{dt}, \dots\right] \left[H^{f} + \sum_{j=1}^{k} \lambda_j H^{g_j}\right] \begin{bmatrix} dx_1/dt \\ \vdots \end{bmatrix} + \left[\frac{\partial f}{\partial x_1} + \sum \lambda_j \frac{\partial g_j}{\partial x_1}, \dots\right] \begin{bmatrix} dx_1/dt \\ \vdots \end{bmatrix}$$
$$\ge 0$$

Ο τελευταίος όρος μηδενίζεται εφόσον $\frac{\partial f}{\partial x_i} + \sum_j \lambda_j \frac{\partial g_j}{\partial x_i} = 0$ και άρα η αναγκαία συνθήκη $d^2f/dt^2 \ge 0$ μετασχηματίζεται αλγεβρικά στην εξής πρόταση

- Για κάθε $\left[\frac{dx_1}{dt}, \dots\right] = y \in R^n$ που ικανοποιεί τις σχέσεις
$$0 = \sum_{i=1}^{n} \frac{\partial g_j}{\partial x_i} \frac{dx_i}{dt} = \nabla g_j \cdot y = 0 \quad j=1,\dots,k$$
θα πρέπει
$$Q(y) = y'\left[H^{f} + \sum_{j=1}^{k} \lambda_j H^{g_j}\right] y \ge 0$$

Αν συμβολίσουμε με $L(x_1,\dots,x_n)$ το $f + \sum_{j=1}^{k} \lambda_j g_j$ θα είναι

$$Q(y) = (y_1, \dots, y_n) \begin{pmatrix} \partial^2 L/\partial x_1^2 & \frac{\partial^2 L}{\partial x_1 \partial x_2} & \cdots \\ \vdots & & \\ \frac{\partial^2 L}{\partial x_n \partial x_1} & & \frac{\partial^2 L}{\partial x_n^2} \end{pmatrix} \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$$

με $Q(y) \ge 0$ όταν $\nabla g_j y = 0$ ή $Gy = 0$.
Στην Γραμμική Άλγεβρα μπορεί να αποδειχθεί το παρακάτω Λήμμα.

Λήμμα 1.* Έστω Q $n \times n$ συμμετρική μήτρα και A: $k \times n$ μήτρα. Για να είναι $\underline{x}'Q\underline{x} \ge 0$ για κάθε $\underline{x}$ με $A\underline{x} = 0$ αρκεί να ισχύουν οι παρακάτω σχέσεις:

* Οφείλεται στον Finsler (1937)

Αν B_m η μήτρα $(k+m)\times(k+m)$

$$B_m = \begin{bmatrix} 0 & 0 & 0 & a_{11} & \cdots & a_{1m} \\ \vdots & & & & & \\ 0 & 0 & 0 & a_{k1} & \cdots & a_{km} \\ a_{11} & \cdots & a_{k1} & q_{11} & \cdots & q_{mn} \\ \vdots & & & \vdots & & \\ a_{1m} & \cdots & a_{km} & q_{mn} & \cdots & q_{mm} \end{bmatrix} = \begin{bmatrix} O & A_m \\ A_m' & Q_m \end{bmatrix}$$

Θα πρέπει $(-1)^k \operatorname{Det} B_m > 0$ για $m=k+1,k+2,\dots$

Αν $\underline{x}'Q\underline{x} \le 0$ θα είναι

$$(-1)^{m+1} \operatorname{Det} B_m < 0 \quad m=k+1,\dots,n$$

<u>Παράδειγμα 4</u> Ποιό είναι το πρόσημο της $Q(x,y)=(x,y)\begin{pmatrix}1&2\\2&1\end{pmatrix}\begin{pmatrix}x\\y\end{pmatrix}$ όταν $(x,y)\begin{pmatrix}1\\1\end{pmatrix}=0$;

Προφανώς $Q(x,y)=x^2+y^2+4xy$ επεί η συνθήκη $x+y=0$ σημαίνει $x=-y$. Άρα αν ισχύει ο περιορισμός η τετραγωνική μορφή έχει τιμή $Q(x,-x)=x^2+(-x)^2+4x(-x)=x^2+x^2-4x^2$ $=-2x^2\le 0$. Εφαρμόζουμε τώρα τις μήτρες B_m, για να ελέγξουμε και τον άλλο τρόπο υπολογισμού του προσήμου της Q

Εφόσον $k=1$ και $n=2$ μας ενδιαφέρει μόνο η B_2 που είναι

$$B_2 = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 1 & 2 \\ 1 & 2 & 1 \end{bmatrix}$$

Εφόσον $\operatorname{Det} B_2 = 2$, είναι $(-1)^k \operatorname{Det} B_2 = (-1)^1\cdot 2 = -2$ πράγμα που δείχνει ότι <u>δεν</u> τηρείται η συνθήκη $(-1)^k \operatorname{Det} B_m > 0$ για $m=k+1,\dots,n$ (για θετικό πρόσημο της Q). Ισχύει όμως $(-1)^{2+1}\operatorname{Det} B_2 = (-1)\cdot 2 = -2 < 0$ και άρα τηρείται η συνθήκη για αρνητικό πρόσημο της Q, όπως άλλωστε περιμέναμε.

<u>Παράδειγμα 5.</u> Ποιό είναι το πρόσημο της μορφής $Q(x,y,z)=(x,y,z)\begin{pmatrix}2&0&0\\0&-1&-1\\0&-1&-1\end{pmatrix}\begin{pmatrix}x\\y\\z\end{pmatrix}$ όταν $x-y-z=0$;

Είναι $Q(x,y,z)=2x^2-(y^2+z^2+2zy)=2x^2-(y+z)^2$. Αν $x-y-z=0$, θα είναι προφανώς $Q(x,y,z)=$ $=2x^2-x^2=x^2\ge 0$. Περιμένουμε δηλαδή το πρόσημο της τετρ. μορφής να βγει θετικό <u>και</u> από τον έλεγχο των B_m.

Έχουμε $k=1$ και $n=3$. Άρα εξετάζουμε τα $m=k+1,\dots,n$ ή $m=2,3$. Τα B_2, B_3 βρίσκονται αν θεωρήσουμε τις κύριες ελάσσονες τελευταίας και προτελευταίας τάξεως της μήτρας

$$\begin{bmatrix} 0 & 1 & -1 & -1 \\ 1 & 2 & 0 & 0 \\ -1 & 0 & -1 & -1 \\ -1 & 0 & -1 & -1 \end{bmatrix}$$

Είναι προφανώς $\mathrm{Det}\, B_3 = 0$ ενώ $\mathrm{Det}\, B_2 = \begin{vmatrix} 0 & 1 & -1 \\ 1 & 2 & 0 \\ -1 & 0 & -1 \end{vmatrix} = -1$
Άρα είναι $(-1)^k \,\mathrm{Det}\, B_2 = (-1)^1(-1) = 1 > 0$
και $(-1)^k \,\mathrm{Det}\, B_3 = 0 \geq 0$ και επομένως ικανοποιούνται οι συνθήκες για μη αρνητικό πρόσημο.

Εφαρμόζοντας το Λήμμα 1 στην ανάλυση των ακροτάτων, θέτουμε $Q = \{q_{ij}\} = \{\partial^2 L/\partial x_i \partial x_j\}$

και $A = \{a_{ij}\} = \{\partial g_i/\partial x_j\}$. Η $(k+n)\times(k+n)$ μήτρα

$$\bar{H} = \begin{bmatrix} 0 & A \\ A' & Q \end{bmatrix}$$

ονομάζεται "περιφραγμένη μήτρα του Hesse", δηλαδή είναι

$$\bar{H} = \left[\begin{array}{ccc|ccc} 0 & \cdots & 0 & \partial g_1/\partial x_1 & \cdots & \partial g_1/\partial x_n \\ 0 & \cdots & 0 & \partial g_k/\partial x_1 & \cdots & \partial g_k/\partial x_n \\ \hline \partial g_1/\partial x_1 & & \dfrac{\partial g_k}{\partial x_1} & \partial^2 L/\partial x_1^2 & \cdots & \partial^2 L/\partial x_1 \partial x_n \\ \vdots & & \vdots & \vdots & \partial^2 L/\partial x_i \partial x_j & \vdots \\ \partial g_1/\partial x_n & & \partial g_k/\partial x_n & \partial^2 L/\partial x_1 \partial x_n & \cdots & \partial^2 L/\partial x_n^2 \end{array}\right]$$

Με βάση λοιπόν το Λήμμα 1 καταλήγουμε στις παρακάτω ικανές συνθήκες για ακρότατο:

Θεώρημα 2. Έστω ότι στο $\underline{x}^\circ$ είναι $\nabla f + \Sigma \lambda_j \nabla g_j = 0$ και η περιφραγμένη $\bar{H}$ έχει κύριες ελάσσονες τάξεως $k+1, k+2, \ldots, n$ που ικανοποιούν τις σχέσεις $(-1)^k \,\mathrm{Det}\, H_m \geq 0$ $m = k+1, k+2, \ldots n$. Τότε το $\underline{x}^\circ$ είναι τοπικό ελάχιστο. Αν οι ελάσσονες ικανοποιούν τις σχέσεις $(-1)^{m+1} \,\mathrm{Det}\, H_m < 0$ $m = k+1, \ldots, n$ τότε έχουμε τοπικό μέγιστο

Παράδειγμα 3 (Συνέχεια) (Βλέπε σελ. 55).

Είναι $g_1(x,y,z) = (x^2+y^2+z^2-1)$ άρα $\partial g_1/\partial x = 2x$ $\partial g_1/\partial y = 2y$ $\partial g_1/\partial z = 2z$. Επίσης $g_2(x,y,z) = x+y-z$ άρα $\partial g_2/\partial x = \partial g_2/\partial y = 1$ ενώ $\partial g_2/\partial z = -1$.
Έχουμε $\partial^2 L/\partial x \partial y = \partial^2 L/\partial y \partial z = \partial^2 L/\partial z \partial x = 0$
και $\partial^2 L/\partial x^2 = \partial^2 L/\partial y^2 = 2(1+\lambda_1)$, $\partial^2 L/\partial z^2 = 2(2+\lambda_1)$
Άρα η περιφραγμένη μήτρα του Έσ είναι η

$$\bar{H} = \begin{bmatrix} 0 & 0 & 2x & 2y & 2z \\ 0 & 0 & 1 & 1 & -1 \\ 2x & 1 & 2(1+\lambda_1) & 0 & 0 \\ 2y & 1 & 0 & 2(1+\lambda_1) & 0 \\ 2z & -1 & 0 & 0 & 2(2+\lambda_1) \end{bmatrix}$$

Εφόσον $k=2$ και $n=3$ ενδιαφερόμαστε για την $\bar{H}_3$, την τελευταία κύρια ελάσσονα που ταυτίζεται με την $\bar{H}$. Τα κρίσιμα σημεία που βρήκαμε στην σελ. 55 είναι είτε $x = y = +\sqrt{1/6}$ ενώ $z = +2\sqrt{1/6}$ είτε $x=y=-1/6$, $z=2x$

Επίσης είναι $\lambda_1 = -5/3$. Για τις θετικές ρίζες έχουμε

$$H_3 = 4\cdot\frac{1}{6}\cdot\begin{bmatrix} 0 & 0 & 1 & 1 & 2 \\ 0 & 0 & 1 & 1 & -1 \\ 1 & 1 & -4/3 & 0 & 0 \\ 1 & 1 & 0 & -4/3 & 0 \\ 2 & -1 & 0 & 0 & 2/3 \end{bmatrix}$$

και $\operatorname{Det} H_3 = \frac{2}{3}\operatorname{Det}\begin{bmatrix} 0 & \cdots \\ 0 & \cdots \\ 1 & \cdots \\ 1 & \cdots \\ 2 & \cdots \end{bmatrix}$ και

$$\begin{vmatrix} 0 & 0 & 1 & 1 & 2 \\ 0 & 0 & 1 & 1 & -1 \\ 1 & 1 & -4/3 & 0 & 0 \\ 1 & 1 & 0 & -4/3 & 0 \\ 2 & -1 & 0 & 0 & 2/3 \end{vmatrix} = \begin{vmatrix} 0 & 0 & 0 & 0 & 3 \\ 0 & 0 & 1 & 1 & -1 \\ 0 & 1 & -4/3 & 0 & 0 \\ 0 & 1 & 0 & -4/3 & 0 \\ 3 & -1 & 0 & 0 & 2/3 \end{vmatrix} = -\begin{vmatrix} 0 & 1 & 1 \\ 1 & -4/3 & 0 \\ 1 & 0 & -4/3 \end{vmatrix}$$

$$= -\begin{vmatrix} 0 & 1 & 1 \\ 0 & -4/3 & 4/3 \\ 1 & 0 & -4/3 \end{vmatrix} = 4/3 \begin{vmatrix} 1 & 1 \\ 1 & -1 \end{vmatrix} = -8/3 < 0$$

Επειδή το κριτήριο είναι $(-1)^{s+1}\operatorname{Det} H_5 = (-1)^{3+1}\operatorname{Det} H_3 = -8/3 < 0$ έχουμε ότι το υπό εξέταση σημείο πληροί τις ικανές συνθήκες για μέγιστο.

<u>Άσκηση</u> Αποδείξτε ότι και οι αρνητικές ρίζες αντιστοιχούν σε σημείο μεγίστου.

4. Ερμηνεία των πολλαπλασιαστών Lagrange

Έστω ότι στο σημείο $\underline{x}^0$ μεγιστοποιείται η συνάρτηση $f(\underline{x})$ υπό τους k περιορισμούς

$$\bar g_1(x_1,\dots,x_n) = g_1(x_1,\dots,x_n) - a_1 = 0$$
$$\vdots$$
$$\bar g_k(x_1,\dots,x_n) = g_k(x_1,\dots,x_n) - a_k = 0$$

Θέλουμε να δούμε πώς μεταβάλλεται το x^0 και το $f(x^0)$ (η τιμή του μεγίστου) αν αλλάξει ο περιορισμός j και από $g_j(x_1\dots x_n) = a_j$ γίνει $g_j(x_1,\dots,x_n) = a_j + \Delta a_j$.

Έστω ότι το μέγιστο για $a_j + \Delta a_j$ υπάρχει στην τιμή των μεταβλητών $\underline{x}^0 + \underline{\Delta x}$. Αν Δa_j είναι μικρό, περιμένουμε ότι <u>και</u> το $\underline{\Delta x}$ θα είναι μικρό. Η νέα λοιπόν τιμή του μεγίστου γίνεται $f(\underline{x}^0 + \underline{\Delta x}) \doteq f(\underline{x}^0) + \nabla f(\underline{x}^0)\cdot\underline{\Delta x}$ (Taylor)

Έχουμε όμως ότι στο $\underline{x}^0$ ισχύει η σχέση

$$\nabla f(x_0) + \sum_{i=1}^{k} \nabla g_i(x^0)\cdot\lambda_i = 0$$

και άρα

$$\Delta f = f(\underline{x}^0 + \underline{\Delta x}) - f(\underline{x}^0) = \nabla f(\underline{x}^0)\cdot\Delta x$$
$$= -\sum_{j=1}^{k} \lambda_i \nabla g_i(x^0)\cdot\underline{\Delta x}$$

Θα είναι όμως $\Delta g_i(\underline{x}^0) = g_i(\underline{x}^0 + \underline{\Delta x}) - g_i(\underline{x}^0) =$

$$= \begin{cases} 0 & \text{για } i \neq j \\ \Delta a_j & \text{για } i = j \end{cases}$$

Επειδή όμως $\Delta g_i(x^0) = \nabla g_i(x^0)\cdot \Delta \underline{x}$
θα είναι

$$\nabla g_i(x^0)\cdot \Delta \underline{x} = \begin{cases} 0 & \text{για } i\neq j \\ \Delta a_j & \text{για } i=j \end{cases}$$

Τελικά έχουμε

$$\Delta f = -\sum_{i=1}^{k} \lambda_i \nabla g_i(x^0)\cdot \Delta \underline{x} = -\lambda_j \Delta a_j$$

και άρα $\Delta f/\Delta a_j = -\lambda_j$

Αυτό ερμηνεύεται ως εξής: Τα λ_j δείχνουν πόσο αλλάζει η τιμή του μεγίστου αν αλλάξει η τιμή του j-περιορισμού.

<u>Παράδειγμα 1</u>. Στο πρόβλημα max $x+y$ με $g(x,y)=x^2+y^2-1=0$ (σελ. 53) είχαμε $x=y=1/\sqrt2$ και $\lambda=-1/\sqrt2$, ενώ $f(x^0,y^0)=2/\sqrt2=\sqrt2$

Αν ο περιορισμός γίνει $\bar g(x,y)=x^2+y^2-1.1=0$ έχουμε $\Delta a=0.1$ και περιμένουμε $\Delta f/\Delta a=-\lambda$ ή

$$\Delta f = -\left(-\tfrac{1}{\sqrt2}\right)0.1 = 0.1/\sqrt2 = 0.0707.$$

Το νέο μέγιστο είναι το $\mathring{x}=\mathring{y}=0.74162$
και $f(x^0,y^0)=1.48324$ και άρα

$$\Delta f = 1.48324 - \underbrace{1.41421}_{\sqrt2} = 0.0690 \doteq 0.0707$$

που όντως δείχνει ότι η προσέγγιση $\Delta f \doteq -\lambda\Delta a$ είναι ικανοποιητική

<u>Παράδειγμα 2</u>. Στο πρόβλημα min $x+y$ με $g(x,y)=x^2+y^2-1=0$ η λύση είναι $x=y=-1/\sqrt2$ ενώ $f(x^0,y^0)=-\sqrt2$. Αν ο περιορισμός γίνει $\bar g(x,y)=x^2+y^2-1.1$ έχουμε $\lambda=+1/\sqrt2$ και περιμένουμε το Δf να είναι περίπου $\Delta f \doteq -\lambda\Delta a = -0.1/\sqrt2 = -0.0707$. Έχουμε ότι το νέο ελάχιστο είναι $x^0=y^0=-0.74162$ και άρα $f(x^0,y^0)=-1.48324$ οπότε

$$\Delta f = -1.48324-(-1.41421) = -0.0690$$

που πάλι προσεγγίζεται ικανοποιητικά από το $-\lambda\Delta a=-0.0707$.

Τα παραπάνω παραδείγματα δείχνουν το εξής: Τα λ_j δείχνουν πόσο αλλάζει η τιμή του ελαχίστου αν αλλάξει η τιμή του j-περιορισμού η γενικά

<u>Τα λ_j δείχνουν πόσο αλλάζει η τιμή του ακροτάτου αν αλλάξει η τιμή του j-οστού περιορισμού.</u>