14 Feb. α.
Βελτιστοποίηση

Στο project φάκελο
λύσεις στη
βιβλιογραφία.

- Εισαγωγή -

Βιβλίο

Πολυδάκης
Τεχνικές Βελτιστοποίησης

## Εισαγωγή 1.1

① Αντικειμενική συνάρτηση (objective function)

$$\boxed{\begin{aligned} &\text{minimize } f(x_1, x_2, \ldots, x_n) \\ &\text{subject to } g_1(x_1, x_2, \ldots, x_n) = 0 \\ &\qquad\qquad\quad g_2(x_1, x_2, \ldots, x_n) \geq 0 \end{aligned}}$$

$f : X \subseteq \mathbb{R}^n \to \mathbb{R}$

δεν γίνεται να έχω διανύσματα;
όχι → δεν υπάρχει διάταξη στα διανύσματα
(ομοίως ισχύει για μιγαδικούς)

$$\left\{ \begin{aligned} &\cancel{\vec{f} : X \subseteq \mathbb{R}^n \to \mathbb{R}^m} \ \text{X} \\ &\|\vec{f}\| : X \subseteq \mathbb{R}^n \to \mathbb{R}^+ \ \checkmark \end{aligned} \right\}$$

Σχήμα 1.1
$x^*$ → λύση.

[illegible]
→ minimize $f(x_1, x_2) = (x_1 - 2)^2 + (x_2 - 2)^2$
subject to $x_1 + x_2 = 2$ (ισοτικοί περιορισμοί)
// $g_1(x_1, x_2) = x_1 + x_2 - 2 = 0$

Αν $x_2 = 2 - x_1$

φτιάχνω την $\phi(x_1) = f(x_1, 2 - x_1) = \dots$

άρα έχω minimize $\phi(x_1) = 2(x_1 - 1)^2 + 2$

παρατηρώ ότι έχει ελάχιστο στο σημείο $\underline{x_1 = 1}$ (2)

άρα $x_2 = 1$.

άρα $\vec{x}^* = [1, 1]^T$.

---

Μη γραμμική } βελτιστοποίηση.
Κυρτή

Γραφική Μέθοδος Επίλυσης.

Παράδειγμα 1.2

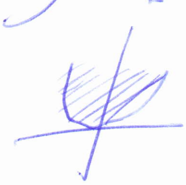
minimize $f(x_1, x_2) = x_1$

subject to $x_1^2 \le x_2$ and $x_1^2 + x_2^2 \le 2$

Τα σχεδιάζουμε.

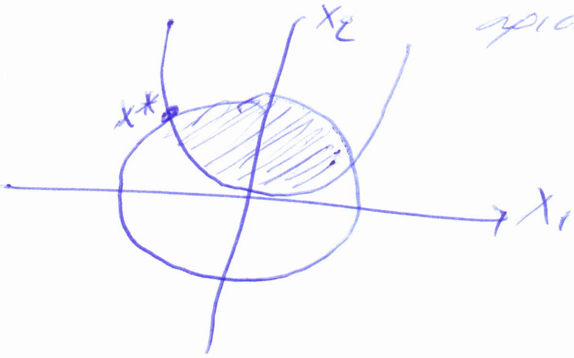
ως ανισότητα είναι μια παραβολή

η είναι εσωτερικός δίσκος



(πάνω διάστημα για το $x_1^2 \le x_2$)

Θέλουμε το ελάχιστο $x_1$, οπότε λέμε στο γράφημα, όσο πιο αριστερά γίνεται, και έτσι βγάζω το $x^*$



14 Feb. b
Βελτιστοποίηση

## 1.2 Εφαρμογές Βελτιστοποίησης

π.χ. SVM

---

## 1.3 Μαθηματική διατύπωση του γενικού προβλήματος βελτιστοποίησης

$f, g_1, g_2, \dots, g_m : X \subseteq \mathbb{R}^n \to \mathbb{R}$ και $\vec{x} = (x_1, x_2, \dots, x_n)$

$\mathcal{E}$ και $\mathcal{I}$ δύο ξένα μεταξύ τους υποσύνολα του $[1, 2, \dots, n]$

Equalities Inequalities

**Το πρόβλημα** (γενικό πρόβλημα βελτιστοποίησης)

$$\boxed{\begin{aligned} &\text{minimize } f(\vec{x}) \\ &\vec{x} \in X \\ &\text{subject to } g_i(\vec{x}) = 0, \ i \in \mathcal{E} \\ &\qquad g_i(\vec{x}) \geq 0, \ i \in \mathcal{I} \end{aligned}} \quad (1)$$

πλήθος ισοτικών δεσμεύσεων

πλήθος ανισοτικών δεσμεύσεων.

**Παράδειγμα**

$m = 5$ και $\mathcal{E} = \{1, 2\}$ και $\mathcal{I} = \{3, 4, 5\}$

άρα γράφουμε

$$\begin{aligned} &\text{minimize } f(\vec{x}) \\ &x \in X \\ &\text{subject to } g_1(\vec{x}) = 0, \ g_2(\vec{x}) = 0 \\ &g_3(\vec{x}) \geq 0, \ g_4(\vec{x}) \geq 0, \ g_5(\vec{x}) \geq 0. \end{aligned}$$

για maximization $\boxed{\max f(\vec{x}) = -\{\min(-f(\vec{x}))\}}$

Ισοδύναμο της (1)

Σύνολο $\Omega = \{\vec{x} \in X : g_i(\vec{x}) = 0,\ i \in \mathcal{E} \text{ and } g_i(\vec{x}) \geq 0,\ i \in \mathcal{I}\}$

$\downarrow$

feasible set.

άρα $\boxed{\begin{array}{c}\text{minimize } f(\vec{x}) \\ \vec{x} \in \Omega\end{array}}$

Παράδειγμα. γραφ. επίλυση..

minimize $f(x_1, x_2) = (x_1 - 2)^2 + (x_2 - 1)^2$

Subject to $x_1^2 - x_2 \leq 0$

$x_1 + x_2 \leq 2$

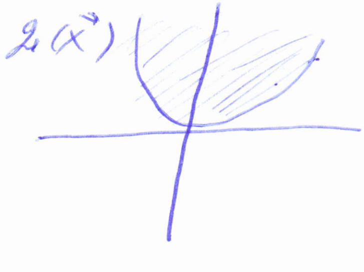
Έχουμε $n = 2$, $\mathcal{E} = \emptyset$

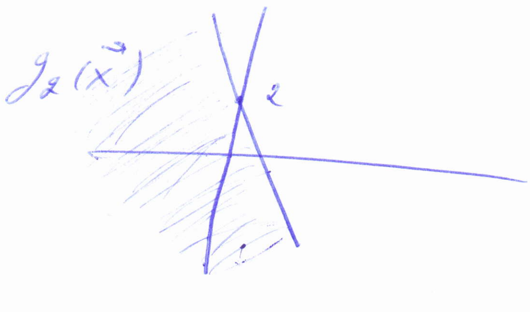
$\mathcal{I}$ : $m = 2$ άρα $\mathcal{I} = \{1, 2\}$

$g_1(\vec{x}) = -x_1^2 + x_2$ //θέλω θετικό πρόσημο.

$g_2(\vec{x}) = 2 - x_1 - x_2$

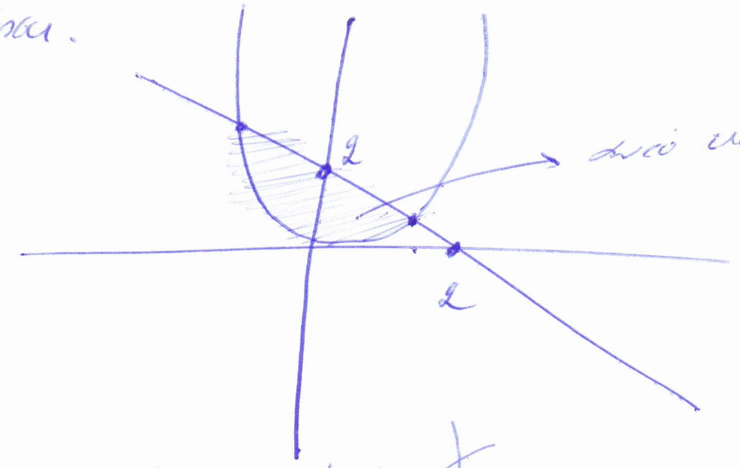
Γραφική Επίλυση...



14 Feb c.

Βελτιστοποίηση.




ποιό είναι το $\Omega$

$$f(x_1, x_2) = (x_1 - 2)^2 + (x_2 - 1)^2 = c$$

κύκλος ακτίνας ρίζας c.
με κέντρο το (2, 1)

τις ισοϋψείς έχουμε κύκλους
με ακτίνα 0.5, 1, 1.5, 2
αντίστοιχα.

η πρώτη φορά που ακουμπάει ο κύκλος στο $\Omega$
θα βρούμε το $x^*$
μετά κύκλοι με μεγαλύτερη ακτίνα από αυτόν

θα έχει και μεγαλύτερη ακτίνα $\Rightarrow$ μεγαλύτερη τιμή της
αντικειμενικής συνάρτησης.

Βελτιστοποίηση 21-Φεβ

κεφ 2

## Βελτιστοποίηση χωρίς δεσμεύσεις.

Παράδειγμα 2.1.

Μοντελοποίηση προβλήματος.

$y_j$ → μετρήσεις

$-\phi(t_j ; x)$ σαν σφάλμα.

Προαπαιτούμενα απο Μαθ. Ανάλυση.

$$x = (x_1, x_2, \dots, x_n) = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \text{ ή } [\dots]^T$$

διάνυσμα κλίσεως ανάδελτα $\nabla f(x)$ $n\times 1$ πίνακας

$\boxed{\nabla f(p) = 0}$ → διάνυσμα 0.

$$\frac{\partial^2 f}{\partial x_i \partial x_j} = \underbrace{\frac{\partial}{\partial x_i}\left(\frac{\partial f}{\partial x_j}\right)}$$

πρώτα παρ. εδώ

μετά εδώ.

Hessian Matrix.

είναι ο πίνακας που εμπεριέχει όλες τις 2ης μερικές παραγώγους.

(πρέπει να ξέρουμε αν το άνω τρίγωνο, ή το κάτω) συμμετρικός πίνακας.

$f \in C^2$ στο $X \subseteq \mathbb{R}^n$

τότε $$\frac{\partial^2 f(\vec{x})}{\partial x_i \partial x_j} = \frac{\partial^2 f(\vec{x})}{\partial x_j \partial x_i}$$

---

Τετραγωνική μορφή πίνακα (βλέπε πλαίσιο 2.3)

$$\boxed{Q f(\vec{p})}(\vec{x}) = x^T \cdot \nabla^2 f(\vec{p}) \cdot \vec{x}$$

αυτό είναι σαν σύμβολο

Hessian Matrix

$1 \times n = n \times n = n \times 1$

$1 \times n \quad n \times 1$

$1 \times 1$ → είναι βαθμωτό.

αν είναι $>0 \Rightarrow$ Hessian Matrix θετικά ορισμένος

αν είναι $<0 \Rightarrow$ Hessian matrix αρνητικά ορισμένος

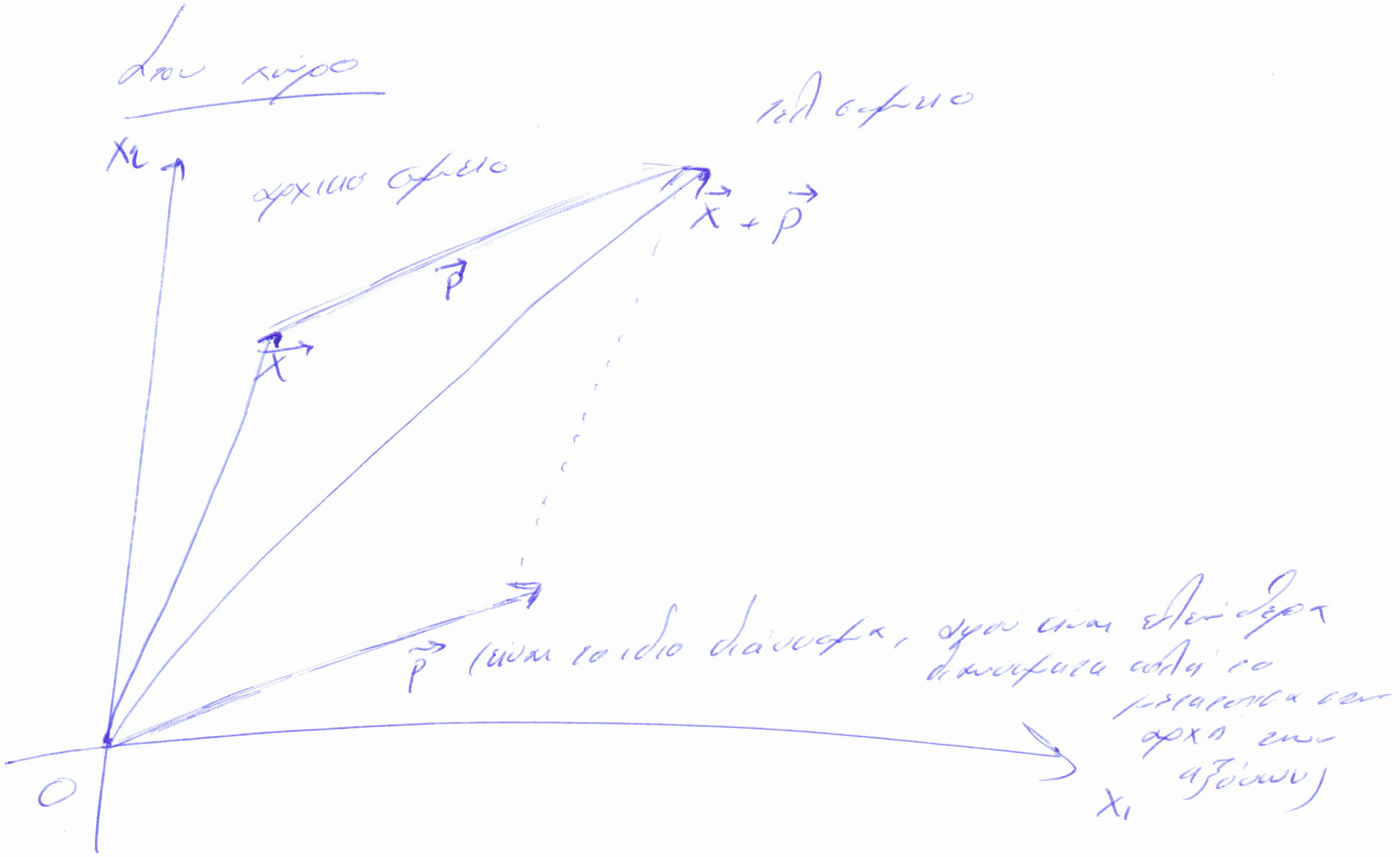
κ.ο.κ.

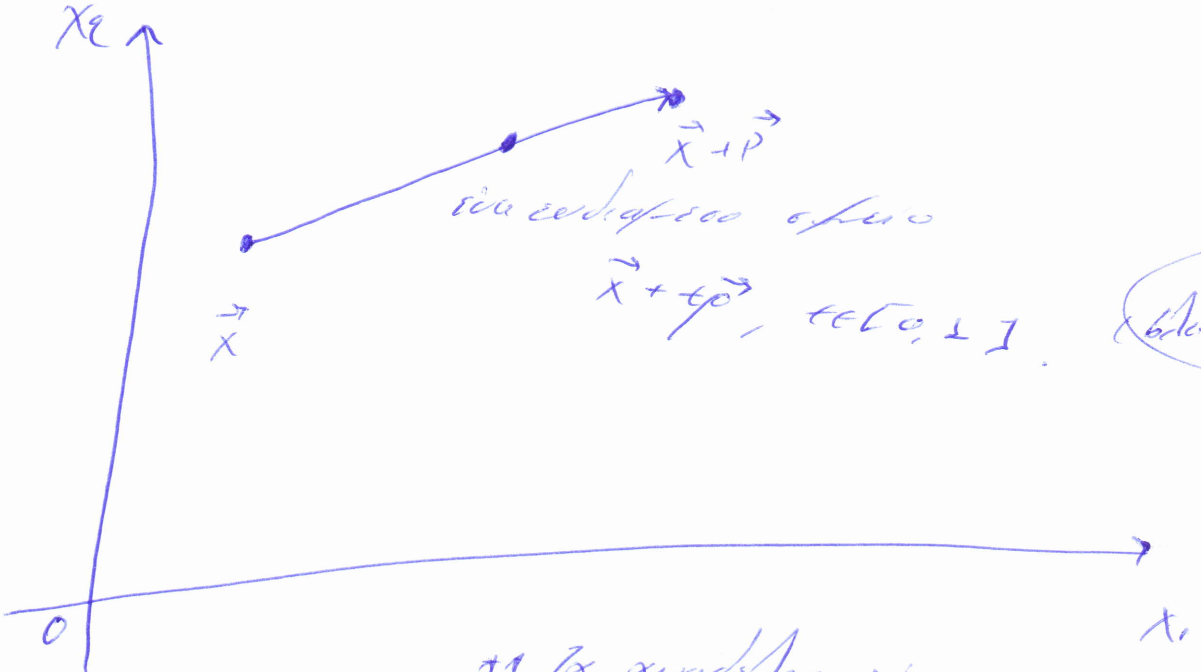
Βελτιστοποίηση

21 Feb 6.

\# Τύπος Taylor → σελ 16.

<u>2ου χώρο</u>




$\vec{x} + t\vec{p}$, $t \in [0, 1]$. (βλέπε τύπο 2.4)

τύπο 2.5

- κοιτά τι τύποι θα μας δώσουν! -

** τα παραδείγματα είναι ~~στο~~ με διορθώσεις.

£ 2.3

Έστω $\vec{x}^*$ είναι τοπικός ελαχιστοποιητής της $f$.

Από

Τύπος Taylor: 1ης τάξης

για $\vec{x} = \vec{x}^*$

$$\boxed{f(\vec{x}^* + p) = f(\vec{x}^*) + \nabla f(\vec{x}^* + \vec{p})^T \cdot \vec{p}.} \quad (1)$$

Διαισθητικά:

αφού $\vec{x}^*$ τοπικός ελαχιστοποιητής,

παίρνοντας από το $\vec{x}^*$ στο $\vec{p}$ (μια μικρή περιοχή του $\vec{x}^*$)

θα πρέπει η συνάρτηση μου να αυξάνει.

άρα η τιμή του θα βγει στο νέο σημείο πρέπει να είναι μεγαλύτερη.

δηλαδή

$$\boxed{f(\vec{x}^* + t\vec{p}) \geq f(\vec{x}^*)} \quad (2)$$

Θέλω η μετατόπιση $\vec{p}$ να είναι τόσο μικρή, ώστε να μην χάνω την έννοια του το ελάχιστο στο $\vec{x}^*$.

$$\vec{x}^* \longrightarrow \vec{x}^* + \vec{p}$$

Από (1) και (2) $\Longrightarrow \nabla f(\vec{x}^* + t\vec{p})^T \cdot \vec{p} \geq 0$

21 Feb e.

Βελτιστοποίηση.

Αφού η μεταβολή θα είναι μηδενική, η τελευταία σχέση, θα ισχύει καθώς $t \to 0$ επομένως

$$\boxed{\nabla f(\vec{x}^*)^T \cdot \vec{p} \geq 0} \quad (3)$$

επειδή η (3) πρέπει να ισχύει για κάθε κατεύθυνση $\vec{p}$

$$\Rightarrow \boxed{\nabla f(\vec{x}^*) = \vec{0}} \quad (4)$$

αναγκαία συνθήκη πρώτης τάξης! (δεν είναι και ικανή συνθήκη)

(θεώρημα 2.2)

↓

φαίνεται ότι δεν μπορεί να πάρω όλα τα $\vec{x}^*$ που θα είναι λύσεις και άρα θα είναι min, max δηλαδή δεν ισχύει το αντίστροφο

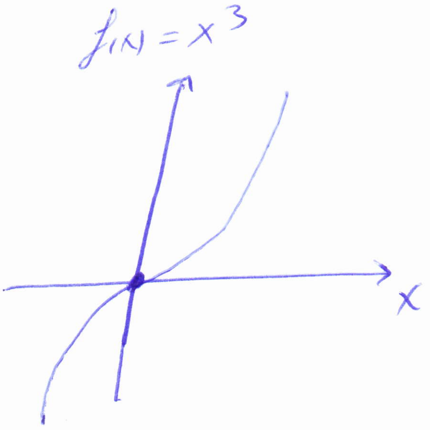
αντιπαράδειγμα

↓

παράδειγμα $f(x) = x^3$

$f'(0) = 0$.

παρατήρηση 2.1




~~σαγματικό~~

→ Έστω $\vec{x}^*$ τοπικός ελαχιστοποιητής της $f$.

Άρα $\boxed{\nabla f(\vec{x}^*) = \vec{0}}$

με εφαρμογή του τύπου Taylor $2^{ης}$ τάξης.

για $\vec{x} = \vec{x}^*$

$$f(\vec{x}^* + \vec{p}) = f(\vec{x}^*) + \underbrace{\nabla f(\vec{x}^*)^T}_{=0} \cdot \vec{p} + \frac{1}{2} \vec{p}^T \nabla^2 f(\vec{x}^* + t\vec{p}) \cdot \vec{p}$$

Άρα

$$\to f(\vec{x}^* + \vec{p}) - f(\vec{x}^*) = \frac{1}{2} \vec{p}^T \nabla^2 f(\vec{x}^* + t\vec{p}) \cdot \vec{p} \quad (1)$$

Επειδή όμως $\vec{x}^*$ τοπικός ελαχιστοποιητής της $f$, ισχύει

$$f(\vec{x}^* + \vec{p}) - f(\vec{x}^*) \geq 0 \quad (2)$$

από (1) και (2)

$$\vec{p}^T \nabla^2 f(\vec{x}^* + t\vec{p}) \cdot \vec{p} \geq 0$$

καθώς ~~[illegible]~~ θα έχουμε ~~[illegible]~~ μικρής μεταβολής

θεωρούμε το όριο καθώς $\epsilon \to 0$

↑ διότι $\vec{p}^T$

$$\boxed{\vec{p}^T \cdot \nabla^2 f(\vec{x}^*) \cdot \vec{p} \geq 0}$$

για κάθε $\vec{p} \in R^n$

δηλαδή δείξαμε ότι ο Hessian matrix $\nabla^2 f(\vec{x}^*)$ θα πρέπει να είναι θετικά ορισμένος

21 Feb cl.

Βελτιστοποίηση

αναγκαία, όχι ικανή

όπως παρατηρήσαμε π.χ.

με την $x^3$.

Διαφορά

$p \to q$

ισχύει $\neg q \to \neg p$

(αναγκαία συνθήκη)

---

• **Ικανές συνθήκες** → θεώρημα 2.5.

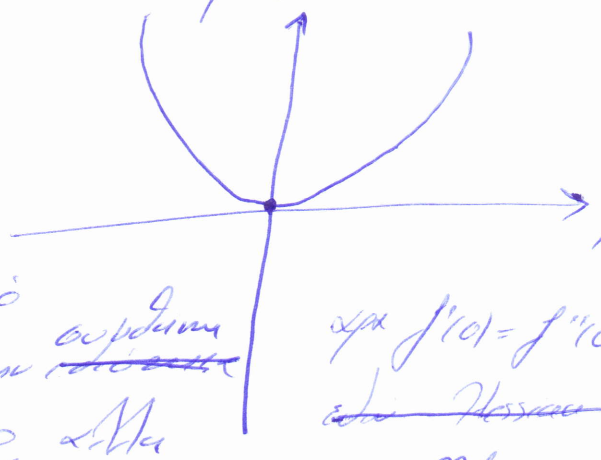
(Από θεώρημα 2.3, είχα ημι ορισμένος)

Χωρίς το "ημι", ~~[illegible]~~ έχω τις ικανές συνθήκες του θεωρήματος 2.5.

Παρατήρηση 2.4

ικανές δεν σημαίνει αναγκαίες.

Αντιπαράδειγμα $f(x)=x^4$




έστω στο $x^* = 0$

η $f(x)=x^4$ έχει αυστηρό ελάχιστο (χάσαμε την ~~[illegible]~~ αυστηρή $f''(0)=0$, αλλά όχι την ιδιότητα τοπ. ελάχιστο)

γιατί $f'(0)=f''(0)=0$.

~~είναι Hessian Matrix~~

Παράδειγμα 2.2

~~Άρα~~ ψάχνω τα σημεία αναγκαίας συνθήκης 1ης τάξης. $\nabla f(\vec{x}^*) = \vec{0}$

που είναι το $\nabla f(x_1, x_2, x_3) = [0 \; 0 \; 0]^T$

το λύνω, και βρίσκω το $\vec{x}^*$

$$\left.\begin{aligned} 3x_1^2 + x_3^2 + 6x_1 &= 0 \\ x_2 &= 0 \\ 2x_1x_3 + 4x_3 &= 0 \end{aligned}\right\} \quad \left.\begin{aligned} 3x_1^2 + x_3^2 + 6x_1 &= 0 \\ x_3(x_1+2) &= 0 \\ x_2 &= 0 \end{aligned}\right\}.$$

~~άρα~~ $x_3(x_1+2) = 0 \rightarrow x_3 = 0$

$\rightarrow x_1 = -2$

ΑΡΑ

άρα

$$\left.\begin{aligned} x_3 &= 0 \\ x_2 &= 0 \\ 3x_1^2 + 6x_1 &= 0 \end{aligned}\right\} \qquad \begin{aligned} x_1 &= -2 \\ x_2 &= 0 \\ x_3^2 &= 0 \end{aligned}$$

$x_1(x_1+2) = 0$

άρα $x_1 = 0$

ή

$x_1 = -2$.

Άρα τα σημεία είναι

~~Άρα~~ $x_a^* = [0, 0, 0]^T$

~~$x_b$~~ $x_b^* = [-2, 0, 0]^T$

άρα για να δω αν η συνάρτηση έχει τοπ. ελαχιστοποιητή ή μεγιστοποιητή, θα εξετάσω τα σημεία χρησιμοποιώντας πλέον τα σημεία των αναγκαίων συνθηκών

21 Feb e.

<u>Βελτιστοποίηση</u>

βλέπω ↓

$Q f (X_a^*) / (X_1, X_2, X_3)$ — πως βρίσκω τις τετραγωνικές μορφές?

$Q f (X_b^*) / (X_1, X_2, X_3)$

1η Πρόοδος → ~~..~~ 21 Μαρτίου !!!

28 Feb α

Βελτιστοποίηση

2.4 σελ 21

Κριτήρια Αναγνώρισης του είδους Τοπικού Βελτιστοποιητή.

2) Προσδιορισμοί ελασσόνων οριζουσών.

Εσσιανός Πίνακας $\nabla^2 f(\vec{x}) = \begin{bmatrix} \cdots \end{bmatrix}$ (ο πίνακας είναι συμμετρικός)

(βρες του!)

$\Delta_1$ = το πάνω αριστερά στοιχείο! $= \dfrac{\partial^2 f}{\partial x_1^2}$

$\Delta_2$ = πάνω αριστερά 2×2 υπομήτρα του Hessian Matrix.

$\Delta_3$ = πάνω αριστερά 3×3 υπομήτρα του Hessian Matrix

$\vdots$

$\Delta_n(\vec{x}) = \det(\nabla^2 f(\vec{x}))$

Εκτελώ τα βήματα εκείνα: → Θεώρημα 2.6.

$\Delta_k(\vec{x}^*)$, $k = 1, \ldots, n$

(1) όλα θετικά

(2) πρόσημα εναλλάξ με το 1ο αρνητικό.

(3) δεν είναι αρνητικό για κάποιο χρόνο κ

(4) δεν το σύνολο έχει και + και − πρόσημα

→ δεν είναι local optimiser.

Αναλυτικά στο του προγραμματισμού μέθοδο, ~~για~~ αυτό μπορώ να την αναπαραστήσω!

<u>Άσκηση</u> (παράδειγμα 2.3)

~~Βρίσκω~~ Βρίσκουμε τις αναλύσεις της $f$, και μετά τις ρίζες των αναλύσεων.

Τα min ή max θα είναι σε ένα από τα δύο σημεία

$X_a^*$ και $X_b^*$ (ικανοποιούν τις συνθήκες
1ης τάξης → ρίζες των αναλύσεων)

Βρίσκουμε τον Hessian Matrix! (Πώς?)

Βρίσκω τον Hessian στο $\vec{X}_a^*$

και υπολογίζω $\begin{cases} \Delta_1(\vec{X}_a^*) = -6 < 0 \\ \Delta_2(\vec{X}_a^*) = -12 < 0 \end{cases}$

(ικανοποιεί 3 ή 4)
(το ίδιο είναι)

άρα $X_a^*$ δεν είναι
local optimiser της $f$.

28 Φεβ 4

Βελτιστοποίηση.

Ομοίως δουλεύουμε για το $\vec{X}_6^*$.

και στη πραγμ. μέθοδο υπάρχουν ερωτήματα, αλλά αυτή είναι πιο εύκολη και πιο χρησιμοποιήσιμη.

Τότε $\nabla^2 f(\vec{X}_6^*) = \begin{bmatrix} 6 & 0 \\ 0 & 2 \end{bmatrix}$

$\begin{cases} D_1(\vec{X}_6^*) = 6 > 0 \\ D_2(\vec{X}_6^*) = 12 > 0 \end{cases}$ το $\vec{X}_6^*$ είναι

→ κριτήριο 1 → local minimizer της $f$.

Ερώτηση: Μήπως το $\vec{X}_6^*$ είναι και ολικός ελαχιστοποιητής

(μέχρι τώρα όλα τα κριτήρια που είδαμε είναι μόνο για τοπικές ιδιότητες)

→ Παράδειγμα 2.6

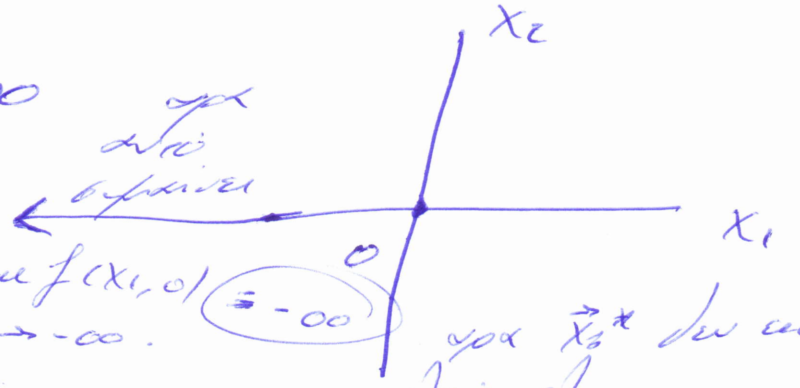
για την $f(x_1, x_2) = x_1^3 - 3x_1^2 + x_2^2$

Εδώ (Με όριο)

Έστω $\phi(x_1) = f(x_1, 0) = x_1^3 - 3x_1^2$

$\lim_{x_1 \to -\infty} \phi(x_1) = -\infty$

(όπου θέλω ελαχιστοποίηση)




$\lim_{x_1 \to -\infty} f(x_1, 0) = -\infty$

άρα $\vec{X}_6^*$ δεν είναι ολικός ελαχιστοποιητής είναι μόνο τοπικός.

Παράδειγμα 2.4 → να το δω στο σπίτι

συνθήκες: θέλω $\begin{cases} x_1^2 + x_1 + 2x_2 = 0 \\ 2x_1 + x_2 - 1 = 0 \end{cases}$

$\begin{cases} x_2 = 1 - 2x_1 \\ x_1^2 + x_1 - 4x_1 + 2 = 0 \end{cases} \Rightarrow$

$\Rightarrow \begin{cases} x_1 = 1 \text{ ή } x_1 = 2 \\ x_2 = 1 - 2x_1 \end{cases}$ → βρίσκω δύο κρίσιμα σημεία

Φτιάχνω τον Hessian Matrix.

Και με $\Delta_1$, $\Delta_2$ βρίσκω ότι έχω είτε local optimisation (όχι global όμως)

Παρατήρηση: εδώ δεν έχω ούτε global minimisation ούτε global maximisation. Άρα η f δεν είναι (άνω ή) κάτω φραγμένη.

2ο Κριτήριο → Θεώρημα 2.7

- έχει να κάνει με τις ιδιοτιμές του Hessian Matrix.

(~~όταν~~ θετικές $\xrightarrow{\text{συνεπάγεται}}$ πίνακας θετικός)

όλες τις ιδιοτιμές θετικές, σημαίνει ότι είναι θετικά ορισμένος → άρα local minimisation

(αυτό είναι ανάλογο με το αντιπαράδειγμα)

18 Δεκ.

βελτιστοποίηση

## Παράδειγμα σελ 25.

έχω διαγώνιο πίνακα → ιδιοτιμές έχει τα στοιχεία της διαγωνίου.
(το ίδιο θα ίσχυε αν είχα και άνω τριγωνικό πίνακα)

γιατί είναι γνωστά

βρίσκω από ΣΑ. $|A-\lambda I|=0$

θα είχα $(\lambda-6)(\lambda-2)(\lambda-8)=0$

---

## Εντολές σε Matlab

diff(f, x1): παράγωγο της f ως προς $x_1$.

subs: substitution

$\left.\begin{matrix} D_2 \\ D_1 \end{matrix}\right\}$ υποορίζουσες.

---

ο 2ος πίνακας είναι με ιδιοτιμές.

---

## Κεφάλαιο 3

- Μέθοδοι αναζήτησης ευθείας.

δηλαδή διανύσματα $\vec{x}_k$ = k=0,1,2...

minimize$(f(\vec{x}))$ | θέλω να παράξω τα διανύσματα $\vec{x}_1, \vec{x}_2$

$\vec{x}\in X\subseteq R^n$ | και αφού υπάρχει κάποιο όριο ως

$\lim_{k\to\infty} \vec{x}_k = \vec{x}^*$

# Αλγόριθμος I

1. $x_0$
2. $\kappa = 0, 1, \ldots$

(α) αν $\vec{x_k}$ ικανοποιεί βέλτιστο-
σημεία όσα πλησιάζει το $x^*$

(β) αλλιώς βρες νέο σημείο $x_{k+1}$
και επανέλαβε το (α)

Πώς κάνω έλεγχο για συνθήκη βελτιστοτ;
Πώς κάνω το update για να βρω το $x_{k+1}$;

# Αλγόριθμος II

Εξειδίκευση αυτής της ερώτησης.

- κατευθύνσεις αναζήτησης -

1. $x_0$
2. $\kappa = 0, 1$

(α) αν $\vec{x_k}$ σ.β. πλησιάζει $x^*$

(β) αλλιώς προσδιόρισε κατεύθυνση $p_k$ ...
ως προς την οποία θα
κινηθούμε για να προσεγγίσ.
το $\vec{x}^*$ (ελαχιστοπ)

βλέπε σχήμα 3.1

όπου στο (3.1) το $\alpha_k$ προσδιορίζει
το πόσο πολύ ως προς την κατεύθυνση
$p_k$ θα κινηθούμε. (αφού το $\alpha_k$
μπορεί να μην υπάρχει και, $\alpha_k = 1$)

Ποιο είναι αυτό το $p_k$?

Άρα πρέπει να βρω
$\begin{cases} p_k \\ \alpha_k \end{cases}$

το $p_k$ δεν είναι το ίδιο σε κάθε βήμα του αλγορίθμου
αλλάζει αφού έχει και εξάρτηση ~~το~~ $k$.

08 Feb d

Βελτιστοποίηση

Τόμος 3.2

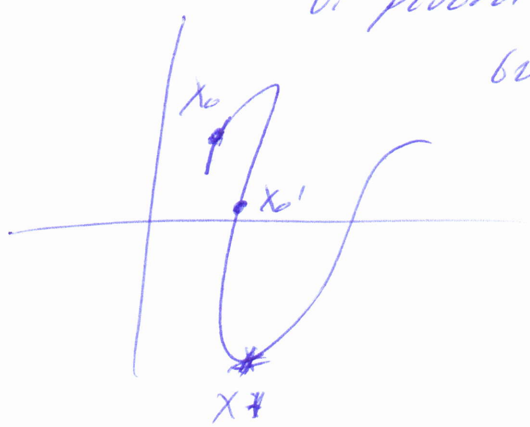
"Ει τρόπος να "βελτιώνεται"" σχετικό σε ...

η συνάρτηση μου να φθίνει. Θέλω μικραίνοντας από $k=1$ σε $k=2$ ...

η τιμή της $f$ να μικραίνει.

οι μέθοδοι που θα μελετήσουμε, είναι πολύ εξαρτώμενες

στην επιλογή του αρχικού σημείου




πχ από το $x_0$ δεν θα βρω το $x^*$

αλλά θα το βρω από το $x_0'$

# Κατεύθυνση Καθόδου ... το $p_k$

Ποιο $p_k$ θα μου δώσει μια μετάβαση

$$\vec{x}_k \longrightarrow \vec{x}_{k+1}$$

πηγαίνω από το $\vec{x}_k$ στο $\vec{x}_{k+1}$ ξέροντας ότι θα πέσει η τιμή μου.

θέλω $f(\vec{x}_{k+1}) < f(\vec{x}_k)$ (λόγω ότι κάνω minimize)

τόμος 3.2

$$\vec{x}_{k+1} = \vec{x}_k + \alpha \vec{p}_k \quad \text{άρα} \quad \boxed{f(\vec{x}_k + \alpha \vec{p}_k) < f(\vec{x}_k)}$$

και η $\vec{p}_k$ που ικανοποιεί αυτή την προϋπόθεση ονομάζεται descent direction.

Παράδειγμα σελ 59

Αφορά τη γραμμική συνάρτηση

$$f(\vec{x}) = \vec{c}^{\,T} \cdot \vec{x} = c_1x_1 + c_2x_2 + \dots + c_nx_n$$

Θέλουμε

$$f(\vec{x_k} + \alpha \vec{p_k}) < f(\vec{x_k}) \Rightarrow$$

$$\vec{c}^{\,T}(\vec{x_k} + \alpha \vec{p_k}) < \vec{c}^{\,T} \cdot \vec{x_k} \Rightarrow$$

$$\Rightarrow \cancel{\vec{c}^{\,T} \cdot \vec{x_k}} + \alpha \vec{c}^{\,T} \cdot \vec{p_k} < \cancel{\vec{c}^{\,T} \cdot \vec{x_k}} \Rightarrow$$

$$\Rightarrow \underset{1 \times n}{\vec{c}^{\,T}} \cdot \underset{n \times 1}{\vec{p_k}} < 0$$

(άρα ως αποτέλεσμα θα έχω κάτι που θα είναι $1 \times 1 < 0$)

χρησιμοποιώ το $\alpha$, αφού το $p_k$ μου δίνει [?] ~~πόσο~~ το σε ποιο κατεύθυνση θα κινηθώ

---

Παράδειγμα 3.2

(3.3) Προσδιορίζοντας το $p_k$ ως 1ο βήμα, προχωράμε στο 2ο βήμα, όπου θα βρούμε το πόσο θα κινηθούμε προς την κατεύθυνση $p_k$, δηλαδή βρίσκω το $\alpha$.

Θέλω minimize $f(x_k + \alpha p_k)$, $\alpha > 0$.

Θέλω $\vec{x_k} \rightarrow \vec{x_{k+1}}$

$$\vec{x_{k+1}} = \underset{\text{ξέρω}}{\vec{x_k}} + \underset{?}{\alpha} \underset{\text{ξέρω}}{\vec{p_k}}$$

~~Θέλω το α να προσδιορίσω ώστε να μειώσει τη f~~

δηλαδή, minimize $f(\vec{x_{k+1}}) = f(\vec{x_k} + \alpha \vec{p_k})$, $\alpha > 0$.

28 Feb 2

Βελτιστοποίηση. Θέλω διαλέξω να ισχύει

Πρώτο το $\alpha$ είναι άγνωστο.

$$\underset{\alpha>0}{\text{minimize}}\ \phi(\alpha)$$

where $\phi(\alpha) = f(\vec{x}_k + \alpha \vec{p}_k)$

} βρίσκω $\phi'(\alpha)$ και $\phi''(\alpha)$

βρίσκω τα σημεία όπου $\phi'(\alpha) = 0$ και μετά ελέγχω το $\phi''$.

Βλέπε ~~Σ~~ Σχήμα 3.3.

καταλαβαίνω γιατί χρειάζεται αυτή η μέθοδος.

μέθοδος αναζήτησης ευθείας.

! η $f = c$ ονομάζεται ισοσταθμική.

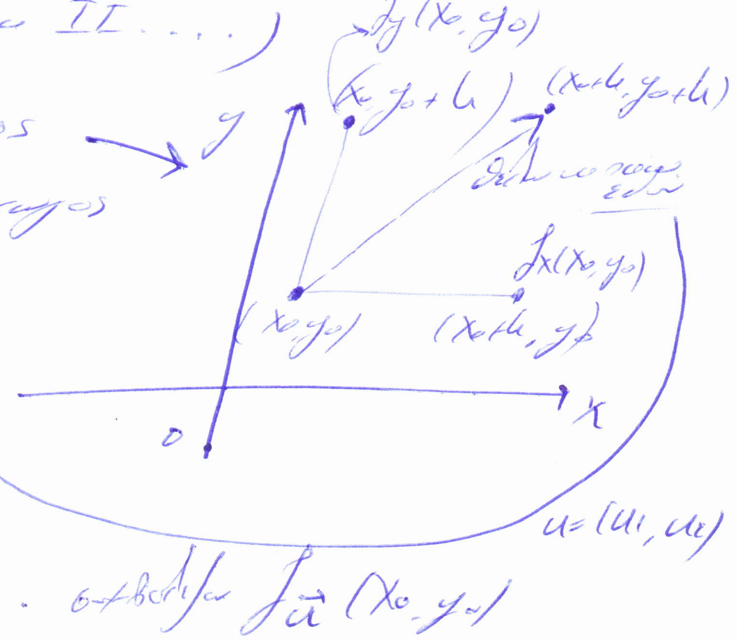
Θέλω βήμα $\alpha$, γιατί η $p_k$ είναι κατεύθυνση καθόδου

(είναι από μια χάλια λύση στη λύση)

3.2 Μαθηματικές Υπενθυμίσεις

(τα έμαθα στην ανάλυση II ...)

1. τι είναι μερική παράγωγος
2. κατευθυνόμενη παράγωγος




είμαι στο $(x_0,y_0)$ και πάω στο $(x_0+hu_1, y_0+hu_2)$

κατευθυνόμενη παράγωγος της $f$ στο $(x_0,y_0)$ κατά μήκος της κατεύθυνσης $u$. συμβολίζεται $f_{\vec{u}}(x_0,y_0)$

αν πάρω σαν u(0,1) θα έχω $f_y$

για u(1,0) θα έχω $f_x$

Θεώρημα 3.1 → {πολύ σημαντικό}!

$$\boxed{f_{\vec{u}}(\vec{p}) = \nabla f(\vec{p}) \cdot \vec{u}} \quad (3.6)$$

$f_{\vec{u}}(\vec{p})$

δηλαδή τη μεταβολή στο σημείο της συνάρτησης $f$ στο σημείο $p$, κατά μήκος της κατεύθυνσης $\vec{u}$.

(άρα αυτή η έννοια παραγώγου λέγεται πολύ στην βελτιστοποίηση)

άρα λέει θα ποσοτικοποιηθεί το πόσο μπορεί να μειωθεί η συνάρτηση μου.

Δες παράδειγμα 3.3 ...

Θεώρημα 3.2 → {πολύ σημαντικό}!

Είμαι σε ένα σημείο του χώρου.

Προς τα πού να κινηθώ ε.ω. η συνάρτηση μου να μειωθεί ή να αυξηθεί με τον μεγαλύτερο ρυθμό

Απάντηση → Θεώρημα 3.2. (Κλασικό γράφημα)

$\max(f_u(p)) = \|\nabla f(p)\| \quad \theta = 0$

$\min(f_u(p)) = -\|\nabla f(p)\| \quad \theta = \pi$.

7. Μαθ.

Βελτιστοποίηση

Για γρήγορα...

$$\boxed{\text{minimize } \phi(\alpha) = f(\vec{x_k} + \alpha \vec{p_k})}$$

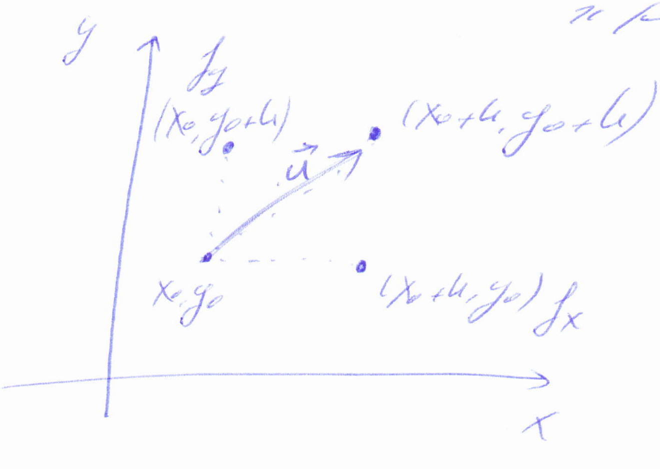
$\alpha > 0$ Ξέρω $\vec{x_k}, \vec{p_k}$ άρα ψάχνω μόνο το $\alpha$!

(Αν το α το γρηγορώ αναζήτηση απ' την κατεύθυνση $\vec{p_k}$, άρα λέγεται δεν κάνω και minimize).

Πώς κάνω minimize?
βρίσκω $\phi'(\alpha)$, $\phi''(\alpha)$

σελ 30 ισοσταθμικές επιφάνειες / καμπύλες.

Μετά είδαμε κατευθυνόμενη παράγωγος...




η μεταβολή έχει η συνάρτηση μου? → διάνυσμα στην

$f_u(p) = \nabla f(p) \cdot u$

μοναδιαίο

διαφορά μερικών παραγώγων.

Θεώρημα 3.2

Αν κοιτάξω σε όλες τις κατευθύνσεις και ως προς τις οποίες μπορώ να κινηθώ

που πρέπει να κινηθώ για να έχω

μέγιστη αύξηση / μείωση.

στην κατεύθυνση $\nabla f(p)$ στην κατεύθυνση $-\nabla f(p)$

New!!!

## Θ3.3 Κατευθύνσεις καθόδου

(κλίση στο $\vec{P}$ κινούμενη ως προς $\vec{u}$) : $f_{\vec{u}}(\vec{p}) = \nabla f(\vec{p}) \cdot \vec{u}$

θέλω ένα γενικό κριτήριο για προβλήματα minimization.

η κατευθυνόμενη παράγωγος εκφράζει τον ρυθμό μεταβολής άρα θέλω η τιμή της συνάρτησης να είναι αρνητική!

**Για minimization**

άρα για $f_{\vec{u}}(\vec{p}) < 0$ θα έχω αρνητικό ρυθμό μεταβολής της τιμής της $f$ τοπικά, όταν ξεκινώ από το $\vec{p}$ και κινούμαι (τοπικά) κατά μήκος της $\vec{u}$.

άρα για την $k$ επανάληψη, είμαι στο $\vec{x_k}$ και θέλω να κινηθώ ως προς $\vec{p_k}$

άρα θέλω $\boxed{\nabla f(\vec{x_k}) \cdot \vec{p_k} < 0}$ $\Rightarrow$

$$\Rightarrow (-\nabla f(\vec{x_k})) \cdot \vec{p_k} > 0$$

έχω εσ. γινόμενο → γεωμ. ερμηνεία → προσανατολισμός
↳ χαρακτ. της γωνίας!

$$\vec{x} \cdot \vec{y} = \|\vec{x}\| \cdot \|\vec{y}\| \cos\theta \qquad 0 \le \theta \le \pi$$

1) $\vec{x} \cdot \vec{y} > 0 \Rightarrow \cos\theta > 0 \Rightarrow 0 \le \theta < \pi/2$
άρα θ οξεία γωνία

6. Μαϊ b
Βελτιστοποίηση

ε) $\vec{x}\cdot\vec{y} < 0 \Rightarrow \cos\theta < 0 \Rightarrow \pi/2 < \theta \leq \pi$
άρα θ αμβλεία γωνία

(για ελάχιστες)
άρα για το $\vec{p_k}$ : σχηματίζει οξεία γωνία ($0 \leq \theta < \pi/2$) με το $(-\nabla f(\vec{x_k}))$
ή
σχηματίζει ($\pi/2 < \theta \leq \pi$) αμβλεία γωνία με το $\nabla f(\vec{x_k})$

σελ 34.
η μέγιστη διανυσματική είναι η ορθογώνια,
και τα $\pm\nabla f(\vec{x_k})$ είναι κάθετα σε αυτήν.
παρατηρώ ότι ισχύουν τα προαναφερθέντα.

σελ ~~35~~ 34 → παρατηρήσεις για την επιλογή του $\alpha_k$.

---

Παράδειγμα 3.6

$$f(x_1, x_2) = (x_1 + x_2^2)^2$$

θεωρώ κατεύθυνση $\vec{p_k} = (-1, 1)$, είναι στο $x_k(1, 0)$

$\vec{p_k} = [-1, 1]^T$
$\vec{x_k} = [1, 0]^T$

i) Αποδείξτε ότι είναι κατεύθυνση καθόδου! (το $\vec{p_k}$)
(descent direction)

βρίσκω $\nabla f(x_1, x_2) = \ldots$
βρίσκω το $\nabla f(\vec{x_k}) = \ldots$

βρίσκω $\nabla f(\vec{x_k})\cdot\vec{p_k} = -2 < 0$
άρα $\vec{p_k}$ descent direction.

ii) βρείτε το $\alpha_k$.

Θέλω να κάνω minimize $f(\vec{x_k} + \alpha \vec{p_k}) = \phi(\alpha)$

με $\alpha > 0$.

άρα $\phi(\alpha) = f((1,0) + \alpha(-1,1)) = f(\overset{x_1}{1-\alpha}, \overset{x_2}{\alpha}) =$

$= (1 - \alpha + \alpha^2)^2$ συνάρτηση μόνο του α.

για την οποία θέλω $\phi'(\alpha) = 0$ και $\phi''(\alpha) > 0$.

$\phi'(\alpha) = 0 \Rightarrow \dots \Rightarrow \alpha = 1/2$

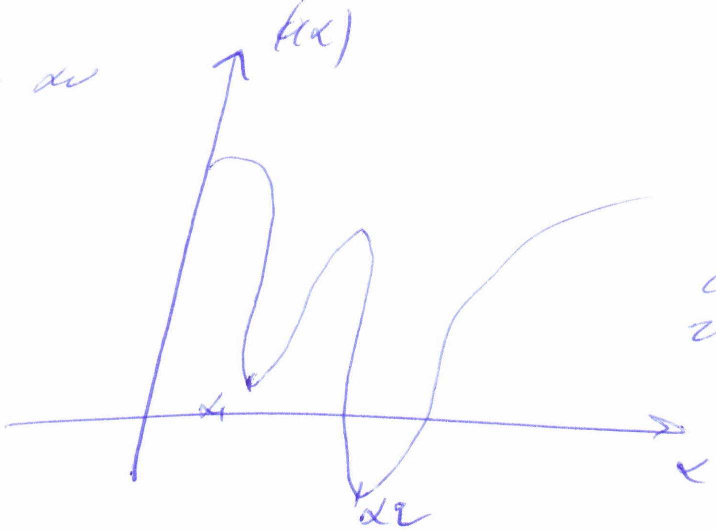
(βρίσκω κρίσιμα σημεία)

ελέγχω δεύτερη παράγωγο στο σημείο αυτό

$\phi''(\frac{1}{2}) = 3 > 0$ ✓

επομένως $x_{k+1} = x_k + \alpha_k p_k = (1,0) + \frac{1}{2}(-1,1) = (\frac{1}{2}, \frac{1}{2})$

παρατηρήσεις: αν




αν είχαμε δύο α $\alpha_1, \alpha_2$, όπως βλέπω στο σχήμα, υπολογίζω την τιμή των $\phi(\alpha_1)$, $\phi(\alpha_2)$ διαλέγω το $\alpha_2$!

6. Μεως
Βελτιστοποίησης

Παράδειγμα 3.5

Θα δείξουμε ότι (αφού είναι κάθετα)

$$\vec{p_k} \cdot \nabla f(\vec{x_{k+1}}) = 0$$

$$\phi(\alpha) = f(\vec{x_k} + \alpha \vec{p_k})$$

→ Για το πρόβλημα minimize $\phi(\alpha)$, $\alpha > 0$

ισχύει $\boxed{\phi'(\alpha) = 0}$ (1)

$$\phi'(\alpha) = \lim_{h \to 0} \frac{\phi(\alpha+h) - \phi(\alpha)}{h} = \lim_{h \to 0} \frac{f(x_k + \alpha p_k + h p_k) - f(x_k + \alpha p_k)}{h}$$

με Taylor δεύτερης τάξης: στο σημείο $f(x_k + \alpha p_k + h p_k)$

κάποιο ενδιάμεσο σημείο

$$= \lim_{h \to 0} \frac{\cancel{f(x_k + \alpha p_k)} + h \nabla f(x_k + \alpha p_k)^T \cdot p_k + \frac{1}{2} h^2 p_k^T \cdot \nabla^2 f(\vec{\xi}) p_k - \cancel{f(x_k + \alpha p_k)}}{h} =$$

(διαιρώ με h)

$$= \lim_{h \to 0} \nabla f(x_k + \alpha p_k)^T p_k + \frac{1}{2} h p_k^T \cdot \nabla^2 f(\vec{\xi}) \cdot p_k$$

$$= \nabla f(x_k + \alpha p_k)^T p_k + \lim_{h \to 0} \frac{1}{2} h p_k^T \nabla^2 f(\vec{\xi}) \cdot p_k$$

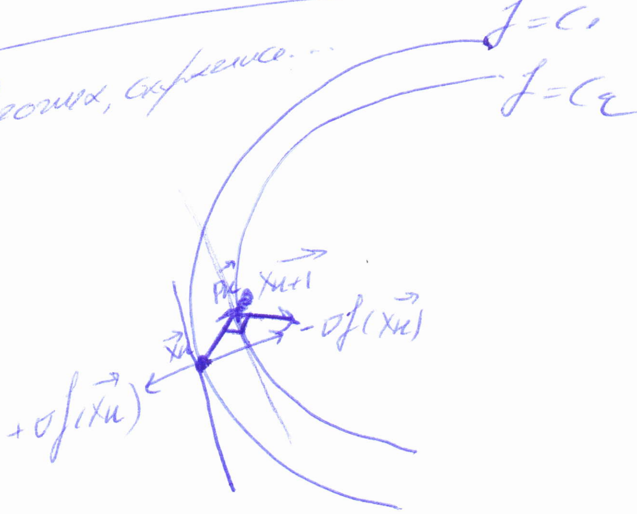
$$= \nabla f(x_k + \alpha p_k)^T \cdot p_k$$

$$= \nabla f(\underbrace{\vec{x_k} + \alpha \vec{p_k}}_{x_{k+1}})^T \cdot \vec{p_k} = \nabla f(\vec{x_{k+1}})^T \cdot \vec{p_k} \ (2)$$

Βρίσκω το $\phi'(\alpha)$ το οποίο θέλω από την (1) να είναι $=0$.

άρα $\boxed{\nabla f(\vec{x}_{k+1})^T \cdot \vec{p}_k = 0}$ ✓

$\vec{x}\cdot\vec{y}=0$ } είναι το ίδιο...
$\vec{x}^T\cdot\vec{y}=0$ }
(όχι μόνο για $=0$).




καθετότητα, ασφαλώς...

$\vec{p_k}$ είναι οποιαδήποτε κατεύθυνση.

σελ 37

~~βήμα... ~~

$\cancel{\vec{x}_{k+1} = \vec{x}_k - a_k B_k (\nabla f(\vec{x}_k))}$

$\vec{x}_{k+1} = \vec{x}_k + \alpha \vec{p}_k$ : linesearch method

$(-\nabla f(\vec{x}_k))\cdot \vec{p}_k > 0$ : descent direction condition

$\Rightarrow$

$\Rightarrow (-\nabla f(x_k))^T \vec{p}_k > 0$

(*)

Για επόμενο θεωρούμε (γενική επιλογή $p_k$)

$\vec{p}_k = -B_k \cdot \nabla f(x_k)$ (**)

όπου $B_k$ συμμετρικός και θετικά ορισμένος πχ πίνακας

6-Μαι d

Βελτιστοποίηση

Αντικατάσταση της (**) στην (*)

$$(-\nabla f(\vec{x_k}))^T \cdot (-\Delta_k \cdot \nabla f(\vec{x_k})) > 0 \Rightarrow$$

$$\Rightarrow \boxed{\nabla f(\vec{x_k})^T \cdot (\Delta_k \cdot \nabla f(\vec{x_k})) > 0}$$ ισχύει;

Ισχύει! αφού $\Delta_k$ είναι θετικά ορισμένος πίνακας.

Ένας πίνακας A λέγεται θετικά ορισμένος όταν ισχύει

$$\vec{X}^T \cdot A \cdot \vec{X} > 0, \ \forall \vec{x} \in \mathbb{R}^n, \ \vec{x} \neq \vec{0}$$

---

Μέθοδος Steepest descent.

Θέτω $\Delta_k = I$, μοναδιαίος πίνακας. (είναι θετικά ορισμένος)

(αφού $X^T \cdot I \cdot X = \vec{X}^T \vec{X} = \|\vec{X}\|^2 > 0 \ \forall \vec{x} \in \mathbb{R}^n$ με $\vec{x} \neq 0$)

και λαμβάνουμε $\boxed{\vec{p_k} = -\nabla f(\vec{x_k})}$

άρα $\boxed{x_{k+1} = x_k - \alpha_k \nabla f(\vec{x_k})}$

!!! το θεώρημα 3.2 έχει τοπική σύγκλιση

χρειαζόμαστε → μικρή μεταβολή.

~~Στην steepest descent δεν υπάρχει διαφορά χωρίς διατήρηση ρυθμού σύγκλισης.~~

Δεν εφαρμόζεται γιατί είναι πολύ αργή μέθοδος.

1. Αρχικοποίηση αρχικό βήμα $x_0$. $\boxed{\text{Αλγόριθμος}}$

2. Κύριο βήμα

(α) συνθήκη τερματισμού του αλγορίθμου

αν $\|\nabla f(x_k)\| < \varepsilon \rightarrow$ stop

(αυτός ο έλεγχος είναι γενικός, δεν είναι μόνο στην steepest descent)

Αναγκαία συνθήκη πρώτης τάξης.

$$\nabla f(\vec{x}^*) = \vec{0}$$

Κάποια στιγμή θα $\nabla f(\vec{x}_k) \simeq \vec{0}$

τεχνικά για διάφορους λόγους, με αριθμητική μέθοδο δεν θα βρει ποτέ 0 θα βρει κάτι κοντά στο 0, το οποίο θα είναι το $\varepsilon$ το οποίο και καθορίζει την ακρίβεια που θέλω. Το $\varepsilon$ είναι μια αριθμητική τιμή, π.χ. $10^{-3}$ ή $10^{-6}$ (μεγαλύτερη ακρίβεια).

(β) Την πρώτη φορά θέσε $p_0 = -\nabla f(x_0)$ ...

(γ) χρήση της $\phi(\alpha)$ για να προσδιορίσω το $\alpha_0$

(δ) κάνω update...

(ε) k++ ;.

Θα πλησιάσει αλλά σε πολλά βήματα, το οποίο σημαίνει μεγάλο k.

Σχήμα 3.6 εφαρμογή της μεθόδου

6. Μωε

Βελτιστοποίηση

(σελ 39)

Θεώρημα 3.3. (Zig-Zag behaviour of steepest descent method).

το διάνυσμα με το οποίο πάω $\vec{x_k}$ στο $\vec{x_{k+1}}$ είναι κάθετο με το διάνυσμα με το οποίο πάω απ' το $\vec{x_{k+1}}$ στο $\vec{x_{k+2}}$

(όχι απόδειξη)

Ειδική περίπτωση: αν κάνω steepest descent.

όταν $f(x_1, x_2) = x_1^2 + x_2^2$

το min είναι το $(0,0)$

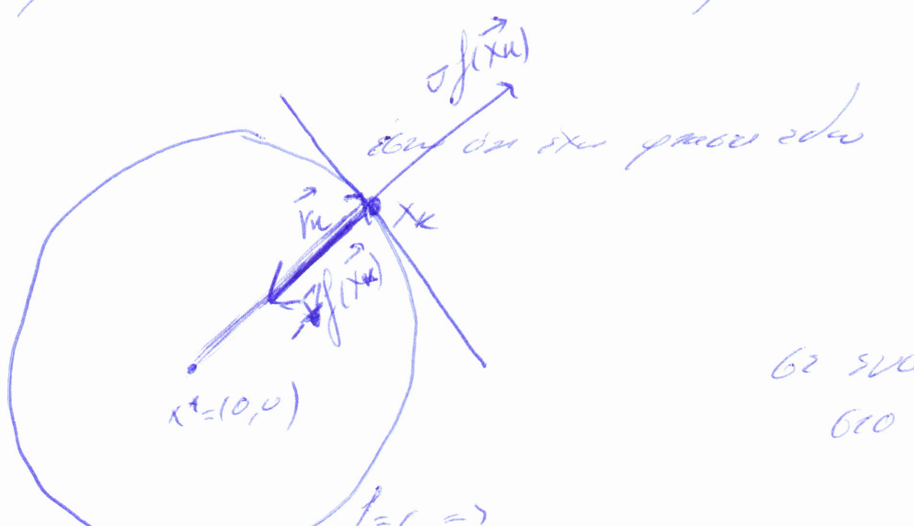
θα βρω το $x^*$ σε ένα βήμα!

οι ισοσταθμικές καμπύλες είναι κύκλοι, όπου $x_1^2 + x_2^2 = c$

ο ελάχιστο λοιπόν είναι το κέντρο του κύκλου.

Υπολογίζω $\nabla f(x_1, x_2) = 2(x_1, x_2) = 2r$. $(r = (x_1, x_2))$

άρα η ~~κατ~~ κατεύθυνση αναζήτησης θα είναι




$p_k = -\nabla f(\vec{x_k})$ ~~...~~ με ε

$\nabla f(\vec{x_k}) = 2\vec{r_k}$

(βλέπε σχήμα 3.8, σελ 40)

σε ένα βήμα λοιπόν, θα πάω στο κέντρο, σε οποίο σημείο και να βρεθώ.

$f = c \Rightarrow$

$\Rightarrow x_1^2 + x^2 = c$. (κύκλος ακτίνας $\sqrt{c}$)

Στο παράδειγμα της έλλειψης, παρατηρώ ότι χρειάζεται περισσότερα από ένα βήμα. γιατί το $-\nabla f(\vec{x}_k)$ αστοχεί!

(Πηγαίνει προς κάποια από τις δύο εστίες?)

~~Είναι αργή~~ (είναι αργή η steepest descent, άρα δεν χρειάζεται να υπολογίσει δεύτερες μερικές παραγώγους.)

Πόσες πράξεις θέλει η steepest descent ανά βήμα

→ υπολογισμός πρώτων μερικών παραγώγων ΜΟΝΟ. άρα υπό αυτή την οπτική είναι μια ελαφριά μέθοδος.

ΠΡΟΟΔΟΣ έως 21 Μαρτίου.

14 Μαΐου

Βελτιστοποίηση.

① Steepest Descent

Επανάληψη

για $B_k = I$

$\vec{p_k} = -\nabla f(\vec{x_k})$

συνήθως μια εγγύηση είναι όταν να λύσω $\nabla f(\vec{x_k}) = 0$

αλλά το προσεγγίζω με $\|\nabla f(\vec{x_k})\| \le \epsilon$

↓

ορίζει την ακρίβεια.

Διαφορά

minimize(α)

$$\boxed{\begin{array}{l} \phi'(\alpha) = 0 \\ \phi''(\alpha) > 0 \end{array}}$$

Αλγόριθμος Ε: Κύριο βήμα.

Το σχήμα 3.7 με τις καθέτους κατευθύνσεις.

στο βήμα Ε). αν δεν έχω αυτό το κριτήριο και αλγόριθμο, και πάρω ένα τυχαίο α με συνθήκες (δηλαδή αν δεν έχω αυτό) τότε δεν θα συνεχίζει αναγκαστικά σε κάθετες κατευθύνσεις.

② Γενική μορφή τετραγωνικής συνάρτησης.

Βελτιστοποίηση σε τετραγωνικές συναρτήσεις.

$$f(\vec{x}) = \underbrace{\frac{1}{2} \underset{1\times n}{x^T} \underset{n\times n}{Q} \underset{n\times 1}{x}}_{1\times 1} - \underbrace{\underset{1\times n}{b^T} \underset{n\times 1}{x}}_{1\times 1} + c$$

άρα το αποτέλεσμα $\in \mathbb{R}$.

Δηλαδή: το $Q$ είναι συμμετρικός πίνακας.

## Παράδειγμα 3.7

$\partial f(\vec{x})$

Ισχύει

$$\frac{\partial f}{\partial x_1} = q_{11}x_1 + q_{12}x_2 - b_1$$

$$\frac{\partial f}{\partial x_2} = q_{12}x_1 + q_{22}x_2 - b_2$$

$$\left.\begin{matrix} \dfrac{\partial f}{\partial x_1} = q_{11}x_1 + q_{12}x_2 - b_1 \\ \dfrac{\partial f}{\partial x_2} = q_{12}x_1 + q_{22}x_2 - b_2 \end{matrix}\right\} \begin{bmatrix} \dfrac{\partial f}{x_1} \\ \dfrac{\partial f}{x_2} \end{bmatrix} =$$

Βγάζω στο γινόμενο διανυσμάτων τα [illegible] σε αναπαράσταση.

$$= \underbrace{\begin{bmatrix} q_{11} & q_{12} \\ q_{12} & q_{22} \end{bmatrix}}_{Q} \underbrace{\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}}_{\vec{x}} - \underbrace{\begin{bmatrix} b_1 \\ b_2 \end{bmatrix}}_{\vec{b}} = Qx - b = \nabla f(\vec{x})$$

Να βρω τον Hessian Matrix $\nabla^2 f(x) = Q$

$$\left.\begin{matrix} \dfrac{\partial^2 f}{\partial x_1^2} = q_{11} & \dfrac{\partial^2 f}{\partial x_1 x_2} = q_{12} \\ \dfrac{\partial^2 f}{\partial x_2^2} = q_{22} & \dfrac{\partial^2 f}{\partial x_2 x_1} = q_{12} \end{matrix}\right\} \begin{bmatrix} \dfrac{\partial^2 f}{\partial x_1^2} & \dfrac{\partial^2 f}{\partial x_2 x_1} \\ \dfrac{\partial^2 f}{\partial x_1 x_2} & \dfrac{\partial^2 f}{\partial x_2^2} \end{bmatrix} =$$

$$= Q = \nabla^2 f(x).$$

14 Μαΐου

Βελτιστοποίηση

Για την τετραγωνική συνάρτηση: $f(\vec{x}) = \frac{1}{2}\vec{x}^T Q \vec{x} - \vec{b}^T \vec{x} + c$

$$\begin{cases} \nabla f(\vec{x}) = Q\vec{x} - \vec{b} \\ \nabla^2 f(\vec{x}) = Q \text{, (δεν έχει x εξάρτηση)} \end{cases}$$

(το 1/2 υπάρχει για να βγει το αποτέλεσμα ίσο Q, και όχι 2Q)

<u>Παράδειγμα 3.8</u>

Ο Q εκτός από συμμετρικός, είναι αντιστρέψιμος $\Rightarrow$
$\det|Q| \neq 0$.

1. Κρίσιμα σημεία = ρίζες της ανάδελτα.

$$\nabla f(\vec{x}^*) = 0.$$

άρα ειδικά για την τετραγωνική συνάρτηση:

$$\Rightarrow Q\vec{x}^* - \vec{b} = \vec{0} \Rightarrow Q\vec{x}^* = \vec{b}$$

πολλαπλασιάζοντας με το $Q^{-1}$ (αφού είναι αντιστρέψιμος)

$$\Rightarrow \vec{x}^* = Q^{-1}\vec{b}.$$

οι λύσεις θα μας δίνουν ένα και <u>μοναδικό</u> σημείο!

Άρα το μοναδικό κρίσιμο σημείο της τετραγωνικής συνάρτησης, είναι το $\boxed{\vec{x}^* = Q^{-1}\vec{b}}$

Βέβαια αυτό δεν σημαίνει ότι θα είναι απαραίτητα βελτιστοποιητής αλλά είναι υποψήφιο.

2

- αν $Q$ θετικά ορισμένος $\rightarrow \vec{x}^*$ local minimizer
- αν $Q$ αρνητικά ορισμένος $\rightarrow \vec{x}^*$ local ~~optimizer~~ maximizer
- αν $Q$ αόριστος $\rightarrow \vec{x}^*$ δεν είναι local optimizer

Άρα τα δύο προαπαιτούμενα μας δίνουν όλη την πληροφορία για τη τετραγωνική συνάρτηση.

---

(1) $\phi'(\alpha) = \vec{p}^T \nabla f(\vec{x} + \alpha \vec{p})$ → από κανόνα παραγώγισης

$\phi(\alpha) = f(\vec{x} + \alpha \vec{p})$

θέλω minimize $\phi(\alpha)$

άρα $\phi'(\alpha) = 0$ και $\phi''(\alpha) > 0$.

αρχικά βρίσκω την παράγωγο ως προς $\alpha$.

$$f(\vec{x} + \alpha \vec{p}) = f(x) + \alpha p^T \nabla f(x) + \frac{1}{2} \alpha^2 p^T \nabla^2 f(x) p$$

δεν έχουμε $\alpha^3$ επειδή μιλάμε για τετραγωνική συνάρτηση άρα το $+\alpha^3 x$ (κάτι) θα είχε 3η παράγωγο το οποίο δεν υπάρχει στις τετρ. συναρτήσεις.

14 Μαΐου

Βελτιστοποίηση.

(άρα $\vec{y}-\vec{x}=\alpha\vec{p}$)

έστω διάνυσμα $\vec{y}=\vec{x}+\alpha\vec{p}$

άρα $f(\vec{y})=f(\vec{x})+(\vec{y}-\vec{x})^T\nabla f(\vec{x})+\frac{1}{2}(\vec{y}-\vec{x})^T\nabla^2 f(\vec{x})(\vec{y}-\vec{x})$

(το ίσον διότι με το ανάπτυγμα).

παραγωγίζω τη τελευταία ως προς $\vec{y}$. (παραγωγίζω ως προς $y$)

$\nabla$

$$\nabla f(\vec{y})=\nabla f(\vec{x})+(\vec{y}-\vec{x})^T\cdot\nabla^2 f(\vec{x})\Rightarrow$$

$$\Rightarrow \nabla f(\vec{y})=\nabla f(\vec{x})+(\vec{y}-\vec{x})^T\cdot Q \Rightarrow$$

$$\Rightarrow \nabla f(\vec{x}+\alpha\vec{p})=\nabla f(\vec{x})+\alpha\vec{p}^T Q \quad (2)$$

~~άρα από (1) και (2)~~ $\Rightarrow \underbrace{\vec{p}^T\nabla f(\vec{x}+\alpha\vec{p})}_{=\phi'(\alpha)}=\vec{p}^T\nabla f(\vec{x})+\alpha\vec{p}^T Q\vec{p}$ (3)

άρα από (1) και (3)

$$\phi'(\alpha)=p^T\cdot\nabla f(\vec{x})+\alpha\vec{p}^T\cdot Q\vec{p}$$

και θέλω $\phi'(\alpha)=0 \Rightarrow \alpha=-\dfrac{\vec{p}^T\cdot\nabla f(\vec{x})}{\vec{p}^T Q\vec{p}}$ που θέλαμε να δείξω.

$\phi''(\alpha)>0$

χρήση πίστης να δείξω ότι $\phi''(\alpha) > 0$.

!! όταν έχω τετραγωνική συνάρτηση, οποιαδήποτε μέθοδο και αν κάνω το $\alpha$ μου θα είναι ίδιο (τύπος 3.14)

!! βήμα $\alpha$ σε μεθόδους αναζήτησης ευθείας (linesearch methods)

$$\alpha = - \frac{\vec{p}^T (Q\vec{x} - \vec{b})}{\vec{p}^T Q \vec{p}}$$

## Παράδειγμα 3.10

Δύο βήματα steepest descent σε τετρ. συνάρτηση.

$\rightarrow x_1 \rightarrow x_2$.

$$Q = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 25 \end{bmatrix} \qquad b = \begin{bmatrix} -1 \\ -1 \\ -1 \end{bmatrix}$$

$$x_0 = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

Θέλω κατεύθυνση καθόδου $p_k = -\nabla f(x_k) = -(Qx_k - b) =$

$$= -Qx_k + b$$

το update $\vec{x}_{k+1} = \vec{x}_k + \alpha_k \vec{p}_k$

SD: $\vec{p}_k = -\nabla f(x_k)$

$$\left.\begin{array}{l} \vec{x}_{k+1} = \vec{x}_k + \alpha_k \vec{p}_k \\ \text{SD: } \vec{p}_k = -\nabla f(x_k) \end{array}\right\} \vec{x}_{k+1} = \vec{x}_k - \alpha_k \cdot \nabla f(\vec{x}_k)$$

14 Μαΐου

Βελτιστοποίηση

βήμα $\alpha_k$ στην steepest Descent.

όπου $\vec{\alpha}_k = -\dfrac{p_k^T \nabla f(x_k)}{p_k^T \cdot Q\, p_k}$ (όπου ισχύει αυτό)

(όπου το βήμα είναι ειδικό για την steepest Descent)

και αφού το $p_k = -\nabla f(x_k)$

$$\alpha_k = \frac{\nabla f(x_k)^T \cdot \nabla f(x_k)}{\nabla f(x_k)^T \cdot Q \cdot \nabla f(x_k)} = \frac{\|\nabla f(x_k)\|^2}{\nabla f(x_k)^T \cdot Q \cdot \nabla f(x_k)}$$

για $k=0$

$\nabla f(x_0) = [1\ 1\ 1]^T, \quad \|\nabla f(x_0)\| = \sqrt{3}$

άρα $\alpha_0 = 0.097$

άρα $x_1 = x_0 - \alpha_0 \nabla f(x_0) = [-0.097\ \ -0.097\ \ -0.097]^T$

για $k=1$

$\nabla f(\vec{x}_1) = Q\vec{x}_1 - \vec{b} = [0.903\ \ 0.516\ \ -1.419]^T$

$\|\nabla f(x_1)\| = 1.760$ άρα

$\alpha_1 = 0.059$

έτσι $x_2 = x_1 - \alpha_1 \nabla f(x_1) = [- \ldots]$

$\|\nabla f(x_0)\| = \sqrt{3}$

όπου $\|\nabla f(x_2)\| = 1.144$. $\|\nabla f(x_1)\| = 1.160$

μέθοδος 1 S.D

λίγη αλλαγή, άρα με αυτά τα αργά βήματα.

αναλ. λύση (όχι Steepest Descent). μέθοδος 2

Υποθέτουμε ότι ο $Q = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 25 \end{bmatrix}$ είναι θετικά ορισμένος.

για ελάχιστο του $Q$: $\nabla f(\vec{x}^*) = \vec{0} \Rightarrow$

$$\Rightarrow Q \cdot \vec{x}^* = \vec{b}$$

άρα $\vec{x}^* = Q^{-1} \cdot \vec{b}$

όπου $Q^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1/5 & 0 \\ 0 & 0 & 1/25 \end{bmatrix}$ και $\vec{b} = \begin{bmatrix} -1 \\ -1 \\ -1 \end{bmatrix}$

λόγω είναι διαγώνιος...
αλλιώς πρέπει να λύσουμε με Gauss.

άρα $\vec{x}^* = [-1, -0.2, -0.04]^T$.

πρέπει να βρω Hessian Matrix και να δω αν είναι θετικά ορισμένος.

όμως όπως προαναφέρθηκε ο Hessian Matrix είναι ο $Q$, ο οποίος βάσει ιδιοτιμών, είναι θετικά ορισμένος. Επομένως $\vec{x}^*$ είναι local minimizer, και είναι ο μοναδικός.

14 Μαΐου

Βελτιστοποίηση

Συγκρίνοντας τις δύο μεθόδους.

$$\vec{x}_2 = [-0.15 \quad -0.127 \quad -0.013]^T$$

και $\vec{x}^* = [-1, -0.2, -0.04]^T$

με την Steepest Descent θα πλησιάσουμε το $\vec{x}^*$ με 216 επαναλήψεις (με $\|\nabla f(\vec{x}_k)\| \leq 10^{-8}$)

---

$$\nabla^2 f(x_1, x_2) = \begin{bmatrix} 2 & 1 \\ 1 & 6 \end{bmatrix}$$

$$\left.\begin{aligned} \Delta_1 &= 2 > 0 \\ \Delta_2 &= |\nabla^2 f(x_1, x_2)| = 11 > 0 \end{aligned}\right\} \text{άρα είναι θετικά ορισμένος.}$$

---

κεφάλαιο 1
κεφάλαιο 2
κεφάλαιο 3

Θεωρία

κεφ 3

→ Ικανοποίηση κελιών βάσει γειτονικά με το $-\sigma f(\vec{x}_n)$

→ Δύο παραδείγματα στο Matlab, να επιλέξουμε μεγάλο $\varphi$, διότι θα βρούμε πολύ γρήγορα λύση.

→ αν έχει ισότιμες τιμές
απο
όταν μου βγάλει δύο τιμές.
θα επιλέξω το μικρότερο.

→ παράδειγμα 3.6
ανοιχτή κατανομή κελιών.

κεφ 2 να διαβάσω κυρίως τις ασκήσεις

→ τα προγράμματα όπως θα μου τα ξαναστείλει θα πρέπει να τα ξέρω.

→ Θεώρημα 2.2 Άσκηση.
ανοιχτοί [illegible] τύποι, 2ης, [illegible]

SOS Θεώρημα 2.6.

παράδειγμα 2.6
παράδειγμα 2.4
SOS

4 Μαΐ

Βελτιστοποίηση.

Κεφ 1

→ γραφικές μέθοδοι (ενδιαφέρον)

σχήμα 1.2 ενδιαφέρον σχήμα.

σχήμα 1.3

→ 1.3 είναι καλό να το δω.

→ Σχήμα 1.4

να περιγράψεις [illegible] 1ου τύπου

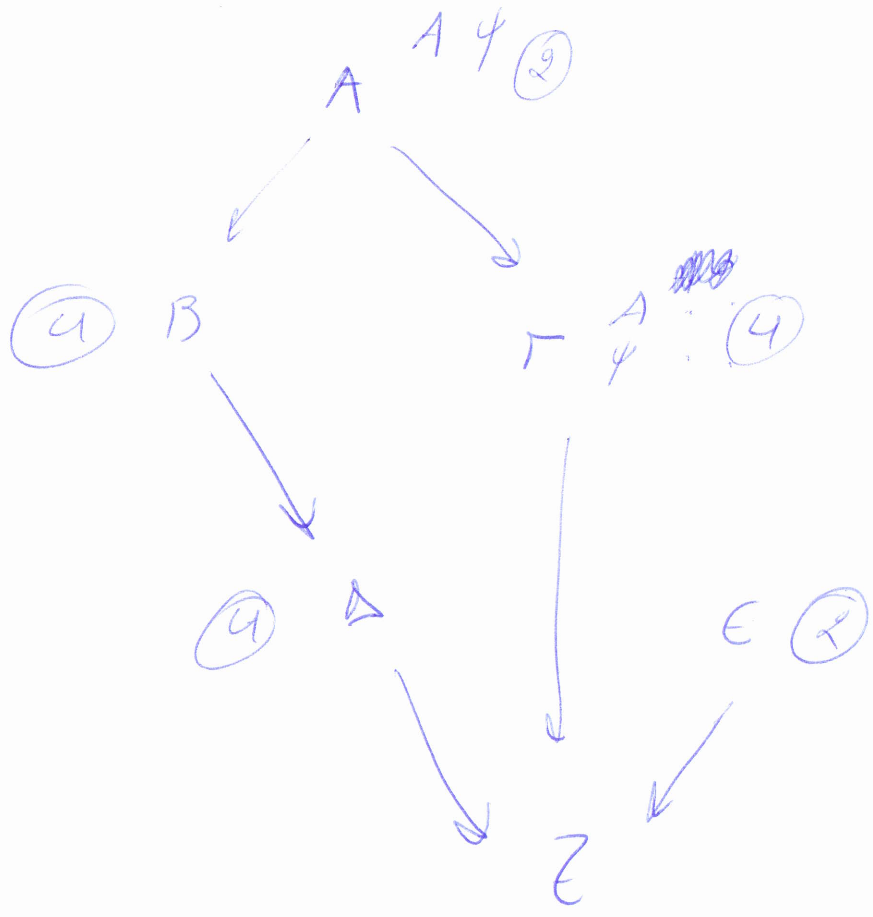
[illegible] στον [illegible].

Σχήμα 1.5 αντίστοιχα.

$\Omega_1$ → local minimizer $x_a^*$

$\Omega_2$ → local min $x_b^*$

το μικρότερο θα είναι το global minimizer.




| | | | | |
|---|---|---|---|---|
| A | A | A | . | . |
| A | A | Ψ | . | . |
| A | Ψ | A | . | . |
| A | Ψ | Ψ | . | . |
| Ψ | A | A | . | . |
| Ψ | A | Ψ | . | . |
| Ψ | Ψ | A | . | . |
| Ψ | Ψ | Ψ | . | . |

$P(\sim\delta, \beta, \sim\varepsilon, \gamma, \sim\zeta)$

$0.999 \times 0.002 \times 0.06 \times 0.05 \times$

$0.99$

$0.999 \times 0.002 \times 0.71 \times 0.05 \times 0.99$

28 Μαΐου

Βελτιστοποίηση

~~[illegible]~~

+ Λύσαμε τη $f(x)=0$ με Newton.

+ Διαλέξεις Online

---

για $f(x)=0$ (Μη γραμμική)

Newton

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}, \quad n=0,1,2,\dots \qquad (1)$$

\# minimize $f(x)$ (ελαχιστοποίηση σε μια διάσταση)

$x \in X \subseteq \mathbb{R}$

(να βρω ολ/τοπικό ελάχιστο)

1) πρέπει να λύσουμε το $f'(x)=0$ να βρω τα κρίσιμα σημεία (αναγκαίες συνθήκες 1ης τάξης)

2) και μετά σε αυτά τα $x$ θα πρέπει $f''(x)>0$.

Αν η $f$ είναι η (1), υπολογίζω αντίστοιχα για την Newton

$$x_{n+1} = x_n - \frac{f'(x_n)}{f''(x_n)}, \quad n=0,1,\dots$$

υπό την προϋπόθεση $f''(x_n) \neq 0$.

Δηλαδή, δεν έχουμε
βάζει άλλα διαφορετικά.

1)η περίπτωση

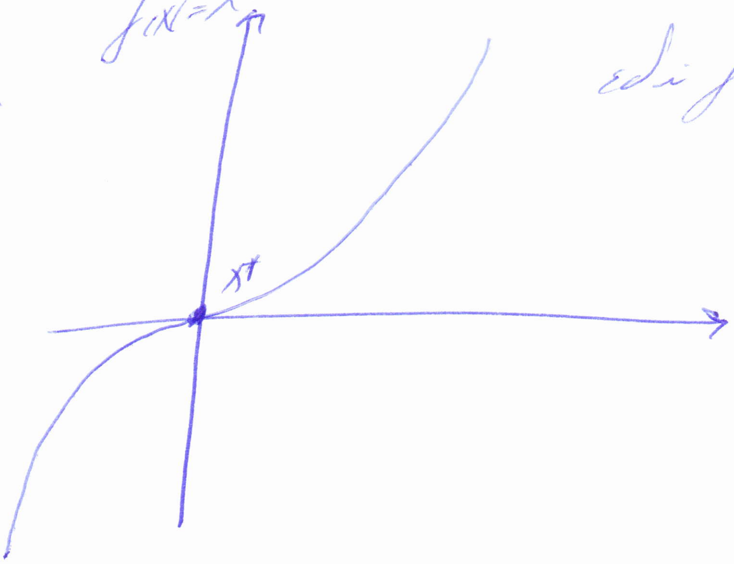
αν βρω ένα $x^*$ για το οποίο

$$f'(x^*) = 0 \quad \text{και} \quad f''(x^*) = 0$$

→ Η Newton θα το βρει, αλλά θα έχει γραμμική σύγκλιση! (τετραγωνική στο θα αργήσει να συγκλίνει)

Το $x^*$ αυτό που θα μηδενίζει τις δύο παραγώγους, θα είναι σημείο καμπής (όχι μίν/μάξ), σε περισσότερες διαστάσεις σαγματικό σημείο.

πχ.




εδώ $f'(x^*) = f''(x^*) = 0$.

2)η περίπτωση. $f'(x^*) = 0$ κρίσιμο σημείο με

$f''(x^*) > 0 \longrightarrow x^*$ local min.

28 Μαΐου

Βελτιστοποίηση.

Υπάρχει και η περιγραφή θεώρηση του (ανάλυσης της γραφικής) θεώρηση γραφήματα.

3) Περίπτωση

$f'(x^*) = 0$ κρίσιμο σημείο

$f''(x^*) < 0 \to x^*$ local max.

(εδώ έχω διάνυσμα). minimize σε πολλές διαστάσεις..

minimize $f(\vec{x})$ με $\vec{x} \in X \subseteq \mathbb{R}^n$

γενικεύοντας τον τύπο της μονοδ. σελίδας.

πρέπει να λύσουμε το $\nabla f(\vec{x}) = \vec{0}$

αναγκαίες συνθήκες πρώτης τάξης.

(γενικεύουμε το $f'(x) = 0$)

άρα ~~$\vec{x}_{n+1} = \vec{x}_n - \frac{\nabla f(\vec{x}_n)}{\quad}$~~ δεν μπορώ να διαιρέσω!!! με πίνακα!!!

άρα $\boxed{\vec{x}_{n+1} = \vec{x}_n - [\nabla^2 f(\vec{x}_n)]^{-1} \cdot \nabla f(\vec{x}_n)}$, $n = 0, 1, \ldots$

$\boxed{\text{Κεφάλαιο 4}}$ Γενική μορφή του descent

$\vec{x}_{k+1} = \vec{x}_k + \alpha_k \vec{p}_k$

πρέπει Descent Direction Condition

$(-\nabla f(\vec{x}_k))^T \cdot \vec{p}_k > 0$.

Και αν πάρω ως $p_k = -\Delta_k \nabla f(\vec{x_k})$

θα έχω

$$(-\nabla f(\vec{x_k}))^T (-\Delta_k \nabla f(\vec{x_k})) > 0 \Rightarrow$$

$$\Rightarrow \boxed{\nabla f(x_k)^T \Delta_k \nabla f(x_k) > 0}$$

σημαίνει ότι πρέπει να είναι θετικά ορισμένος ο $\Delta_k$. (nxn πίνακας)

η Steepest descent έχει ως $\Delta_k = I$ (μοναδιαίος) άρα θετικά ορισμένος.

η μέθοδος Newton έχει $\Delta_k = [\nabla^2 f(\vec{x_k})]^{-1}$

και πρέπει να είναι θετικά ορισμένος nxn πίνακας.

αν είναι θετικά ορισμένος ο $A$ θα είναι θετικά ορισμένος και ο $A^{-1}$ (θετικά ορισμένος σημαίνει ότι είναι αντιστρέψιμος).

και άρα καταλήγουμε στην (4.2)

28 Μαΐου

Βελτιστοποίηση: Ο υπολογισμός του $\Delta_k$ στη μέθοδο Newton είναι χρονικά βαρύς, αλλά συνδυάζοντας με την steepest descent που παίρνουμε $\Delta_k = I$ θα έχουμε πολύ πιο γρήγορη σύγκλιση.

Σημείωση για τον τύπο (4.2)
συνήθως το $\Delta_k$ για την μέθοδο Newton είναι $1$, για να μην έχω επιπλέον υπολογιστικό φορτίο.

# Για την κατεύθυνση,

$$p_k = -[\nabla^2 f(\vec{x_k})]^{-1} \cdot \nabla f(\vec{x_k}) \quad (4.3)$$

σελ 47 → η απόδειξη όσων είπαμε.

- Αλγόριθμοι Newton -

Σε σχέση με την steepest descent αλλάζει το βήμα 4).
Και (μπορεί) και να μην υπάρχει το 5)

Βελτιστοποίηση 3

| $n$ | $x_n$ | $\lvert x_n - x_{n-1} \rvert$ | $f(x_n)$ |
|---|---|---|---|
| 0 | $-1.2$ | - | $1.84 \cdot 10^{-1}$ |
| 1 | $-1.23511$ | $3.5 \cdot 10^{-2}$ | $-1.11 \cdot 10^{-3}$ |
| 2 | $-1.23375$ ~~1.36·10^-3~~ | $1.36 \cdot 10^{-3}$ | $1 \cdot 10^{-5}$ |
| 3 | $-1.23375$ | $0 \; (10^{-6})$ | |

άρα πιο προσεγγιστική τιμή της $x^*$ είναι

$$x^* \simeq -1.23375.$$

Απόδειξη.

→ Ο τύπος της μεθόδου Newton μπορεί να εξαχθεί αν θεωρήσουμε την γραμμική προσέγγιση της $f$ από τους δύο όρους του αναπτύγματος Taylor στο σημείο $x_n$.
Η προκύπτουσα εξίσωση λύνεται για να λυθεί η νέα εκτίμηση $x_{n+1}$ της $x^*$

$$\Rightarrow f(x_n + p) \simeq f(x_n) + p f'(x_n)$$

αν $f'(x_n) \neq 0$ τότε $\boxed{f(x^*) \simeq f(x_n) + p \cdot f'(x_n) = 0}$.

και $\boxed{p = -\dfrac{f(x_n)}{f'(x_n)}}$ και άρα η νέα εκτίμηση

$$x_{n+1} = x_n + p \Rightarrow$$

$$\Rightarrow \boxed{x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}}.$$

που είναι ο τύπος που χρειάζεται.

# Ταχύτητα - ρυθμός σύγκλισης

→ ρυθμός σύγκλισης επαναληπτικής μεθόδου η οποία κατασκευάζει επαναληπτική ακολουθία $\vec{x}_k$ που συγκλίνει στη λύση $\vec{x}^*$, καθώς $k \to +\infty$.

Η ακολουθία των λαθών ορίζεται ως εξής:

$$\vec{e}_k = \vec{x}_k - \vec{x}^*$$

με $\lim\limits_{k \to +\infty} \vec{e}_k = \vec{0}$.

Θα λέμε ότι η ακολουθία $\vec{x}_k$ συγκλίνει στο $\vec{x}^*$ με ρυθμό $r$ και σταθερά του ρυθμού $C$ αν:

$$\boxed{\lim_{n \to +\infty} \frac{\|\vec{e}_{k+1}\|}{\|\vec{e}_k\|^r} = C} \quad \text{με } C < +\infty$$

→ ιδανικά, χωρίς να έχουμε το όριο, ισχύει

$$\|\vec{e}_{k+1}\| = C \cdot \|\vec{e}_k\|^r, \text{ για κάθε } k.$$

A) Αν $r = 1$ τότε $\|\vec{e}_{k+1}\| = C \cdot \|\vec{e}_k\|$ και η σύγκλιση ονομάζεται $\boxed{\text{γραμμική}}$ (linear)

i) για $0 < C < 1$ η νόρμα των σφαλμάτων μειώνεται κατά σταθερό παράγοντα σε κάθε επανάληψη.

ii) για $C > 1$, η $\vec{e}_k$ απειρίζεται.

# Βελτιστοποίηση 4

πχ. αν $c = 0.1$ και $\|\vec{e}_0\| = 1$ τότε οι νόρμες των σφαλμάτων είναι ίσες με

$$\|\vec{e}_k\| = 1,\ 10^{-1},\ 10^{-2},\ 10^{-3},\ 10^{-4},\ 10^{-5},\ 10^{-6},\ 10^{-7}$$
$$k = 0,\ 1,\ 2,\ 3,\ 4,\ 5,\ 6,\ 7$$

→ και ακρίβεια 7 ψηφίων, επιτυγχάνεται σε $k=7$ επαναλήψεις.

ενώ αν $c = 0.99$, οι αντίστοιχες νόρμες γίνονται:

$$\|\vec{e}_k\| = 1,\ 0.99,\ 0.9801,\ 0.9703,\ 0.9606,\ 0.9510,\ 0.9415,\ \dots$$
$$k = 0,\ 1,\ 2,\ 3,\ 4,\ 5,\ 6$$

και θα χρειάζονταν $k = 1600$ επαναλήψεις για ακρίβεια επτά σημαντικών ψηφίων.

Β) Αν $r \geq 1$ και $c=0$ τότε η σύγκλιση ονομάζεται Superlinear (υπεργραμμική)

→ Αν $r=2$ έχουμε τετραγωνική σύγκλιση

Γ) πχ. $r=2$, $c=1$, $\|\vec{e}_0\| = 10^{-1}$ λαμβάνουμε

$$\|\vec{e}_k\| = 10^{-1},\ 10^{-2},\ 10^{-4},\ 10^{-8}$$
$$k = 0,\ 1,\ 2,\ 3$$

(όπου $c=1$ το σφάλμα ελαττώνεται στο τετράγωνο σε κάθε επανάληψη)

δηλαδή $k=3$ επαναλήψεις αρκούν για ακρίβεια επτά σημαντικών ψηφίων.

Δ) Αν $1 < r < 2$ τότε υπεργραμμική σύγκλιση.

(εφαρμόζεται σε προβλήματα βελτιστοποίησης όπου υπολογίζουμε μόνο οι πρώτες παράγωγοι).

πχ. με $r = 1.5$, $C = 1$ και $\|\vec{e}_0\| = 10^{-1}$

$$\|\vec{e}_k\| = 10^{-1},\ 3\cdot10^{-2},\ 6\cdot10^{-3},\ 4\cdot10^{-4},\ 9\cdot10^{-6},\ 3\cdot10^{-8}$$
$$k = 0,\ 1,\ 2,\ 3,\ 4,\ 5$$

δηλαδή χρειαζόμαστε $k=5$ επαναλήψεις για ακρίβεια επτά.

## Παραδείγματα

Βρείτε τα $r$ και $C$ για τις

1) $x_k = 1 + 10^{-k}$ και 2) $x_{k+1} = \dfrac{x_k}{2} + \dfrac{2}{x_k}$ με $x_0 = 4$.

**Λύση**

1) $x_k = 1 + 10^{-k}$

με όρους $x_0 = 2$ ?, $x_1 = 1.1$, $x_2 = 1.01$, $x_3 = 1.001$ ...

συγκλίνει στο $x^* = 1$.

και $\boxed{e_k = x_k - x^*} = 10^{-k}$

άρα $\lim\limits_{k\to+\infty} \dfrac{\|e_{k+1}\|}{|e_k|} = \dfrac{1}{10}$ άρα $r = 1$ και $C = \dfrac{1}{10}$.

# Βελτιστοποίηση 5

## Λύση

2) $x_{n+1} = \dfrac{x_n}{2} + \dfrac{2}{x_n}$ με όρους

$x_0 = 4$, $x_1 = 2.5$, $x_2 = 2.05$, $x_3 = 2.00060975, \dots$

άρα $\lim\limits_{n \to +\infty} x_n = x^* = 2$.

$$e_{n+1} = x_{n+1} - x^* = \frac{x_n}{2} + \frac{2}{x_n} - 2 = \frac{1}{2x_n}(\overbrace{x_n - 2}^{e_n})^2 \Rightarrow$$

$$\Rightarrow e_{n+1} = \frac{1}{2x_n}(e_n)^2 \quad \text{άρα} \Rightarrow$$

$$\Rightarrow \lim_{n \to +\infty} \frac{|e_{n+1}|}{|(e_n)|^2} = \lim_{n \to +\infty} \frac{1}{2x_n} = \frac{1}{4} \quad \text{άρα } r = 2 \text{ και } C = \frac{1}{4}.$$

→ Σε πρακτικές υλοποιήσεις η παραπάνω συμπεριφορά ως προς τη ταχύτητα σύγκλισης παρατηρείται μόνο στο όριο για πολύ μεγάλα $k$.

Στις αρχικές επαναλήψεις δεν υπάρχει ακριβής εκτίμηση για τη συμπεριφορά της τάξης του σφάλματος.

**Παράδειγμα** (εφαρμογή της μεθόδου Newton για μονοδιάστατο πρόβλημα).

για την $f(x) = 7x^4 + 3x^3 + 2x^2 + 9x + 4 = 0$

ο αλγόριθμος της μεθόδου Newton είναι

$$x_{n+1} = x_n - \frac{7x_n^4 + 3x_n^3 + 2x_n^2 + 9x_n + 4}{28x_n^3 + 9x_n^2 + 4x_n + 9}$$

για $x_0 = 0$ έχουμε:

| $n$ | $x_n$ | $f(x_n)$ | $\lvert x_n - x^* \rvert$ |
|---|---|---|---|
| 0 | 0 | $4 \cdot 10^{0}$ | $5 \cdot 10^{-1}$ |
| 1 | -0.44444444 | $4 \cdot 10^{-1}$ | $7 \cdot 10^{-2}$ |
| 2 | -0.506325574 | $3 \cdot 10^{-2}$ | $5 \cdot 10^{-3}$ |
| 3 | -0.511009249 | $2 \cdot 10^{-4}$ | $3 \cdot 10^{-5}$ |
| 4 | -0.511041786 | $9 \cdot 10^{-9}$ | $2 \cdot 10^{-9}$ |
| 5 | -0.511041788 | 0 | 0 |

→ Το 0 προκύπτει λόγω της πεπερασμένης ακρίβειας αριθμητικής για τη συγκεκριμένη υλοποίηση.

→ άρα η σύγκλιση Newton είναι πολύ γρήγορη όταν το $x_0$ επιλεγεί "κοντά" στο $x^*$ (κατάλληλα κοντά...)

Το σφάλμα σχεδόν τετραγωνίζεται σε κάθε επανάληψη.

Βελτιστοποίηση 6

Απόδειξη

Θ. Απόδειξη ότι ο Newton έχει τετραγωνικό ρυθμό σύγκλισης.

i) ορίζουμε $e_n = x_n - x^*$

ii) από ανάπτυγμα Taylor της $f$ έχουμε:

$$f(x^*) = f(x_n - e_n) = f(x_n) - e_n f'(x_n) + \frac{1}{2} e_n^2 f''(\xi)$$

Επειδή $f(x^*) = 0$ και εξ υποθέσεως $f'(x_n) \neq 0$

λαμβάνουμε

iii) $$e_n - \frac{f(x_n)}{f'(x_n)} = \frac{1}{2} e_n^2 \frac{f''(\xi)}{f'(x_n)} \Rightarrow$$

$$\Rightarrow \underbrace{x_n - \frac{f(x_n)}{f'(x_n)}}_{x_{n+1}} - x^* = \frac{1}{2} (x_n - x^*)^2 \frac{f''(\xi)}{f'(x_n)} \Rightarrow$$

$$\Rightarrow x_{n+1} - x^* = \frac{1}{2} (x_n - x^*)^2 \frac{f''(\xi)}{f'(x_n)}$$

iv) Όταν η ακολουθία $x_n$ συγκλίνει τότε, $\xi \to x^*$ και έτσι από την τελευταία για κατάλληλα μεγάλο $n$ ώστε

$\lim_{n \to +\infty} x_n = x^*$ γράφεται

$$x_{n+1} - x^* \simeq \left( \frac{1}{2} \frac{f''(x^*)}{f'(x^*)} \right) (x_n - x^*)^2$$

άρα η Newton έχει $r = 2$ και $c = \frac{1}{2} \frac{f''(x^*)}{f'(x^*)}$.

# Θεώρημα

Έστω η $C^2$ συνάρτηση $f(x)$ με $f(x^*)=0$ και $f'(x^*) \neq 0$. Αν $|x_0 - x^*|$ κατάλληλα μικρό τότε η ακολουθία της μεθόδου Newton

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$$

συγκλίνει $\boxed{\text{τετραγωνικά}}$ στο $x^*$ με σταθερά $\boxed{C = \frac{|f''(x^*)|}{2|f'(x^*)|}}$

## - Περιορισμοί της μεθόδου Newton -

Υποθέσαμε ότι $f'(x_n) \neq 0$ και $f'(x^*) \neq 0$

1) Αν $f'(x_n)=0$ για κάποιο $n$, τότε η μέθοδος Newton αποτυγχάνει διότι υπάρχει διαίρεση με 0 στον τύπο της.

→ Γεωμετρικά αυτό σημαίνει ότι η εφαπτομένη στο σημείο $(x_n, f(x_n))$ είναι οριζόντια και έτσι δεν τέμνει τον x-άξονα, άρα δεν μπορεί να οριστεί η επόμενη προσέγγιση $x_{n+1}$.

2) Αν $f'(x_n) \neq 0$ για κάθε $n$, $f''(x^*) \neq 0$ αλλά $f'(x^*)=0$.

ο συντελεστής $C \to +\infty$. Ο Αλγόριθμος Newton χάνει την τετραγωνική σύγκλιση του.

Αν $f$ πολυωνυμική, σημαίνει ότι έχει πολλαπλή ρίζα στο $x^*$.

## Βελτιστοποίηση 7

### Παράδειγμα 1

?

$$f(x) = (x-1)^2(x-2)(x-3) = x^4 - 7x^3 + 17x^2 - 17x + 16 = 0$$

με $f(x^*) = f'(x^*) = 0$ για το $x^* = 1$.

Ο τύπος της μεθόδου Newton είναι:

$$x_{n+1} = x_n - \frac{x_n^4 - 7x_n^3 + 17x_n^2 - 17x_n + 16}{4x_n^3 - 21x_n^2 + 34x_n - 17}$$

για $x_0 = 1.1$, η μέθοδος συγκλίνει στο $x^* = 1$ με γραμμικό ρυθμό.

για $x_0 = 2.1$, συγκλίνει στο $x^*$ με τετραγωνικό ρυθμό (η μέθοδος μπορεί να υπολογίσει αυτό και αν δεν υπάρχει διαίρεση με το 0 στον τύπο της).

### Παράδειγμα 2

$$f(x) = \tanh x = \frac{e^x - e^{-x}}{e^x + e^{-x}} = 0.$$

Τα παραδείγματα δείχνουν την ευαισθησία της μεθόδου Newton ως προς την επιλογή αρχ. σημείου $x_0$.

i)
για $x_0 = 1$ λαμβάνουμε

$x_1 = -0.8134$, $x_2 = 0.4094$, $x_3 = -0.0473$,
$x_4 = 7.06 \cdot 10^{-5}$, $x_5 = -2.35 \cdot 10^{-13}$.

με $f(x_5) = -2.35 \cdot 10^{-13}$ η μέθοδος συγκλίνει στη ρίζα $x^* = 0$.

ii) για $x_0 = 1.1$ έχουμε

$x_1 = -1.1986$, $x_2 = 1.2341$, $x_3 = -1.6952$, $x_4 = 5.7154$,
$x_5 = -2.30 \cdot 10^4$

με $f(x_5) = 1$ και στο $x_6$ γίνεται overflow.

Δηλαδή με $x_0 = 1.1$ η μέθοδος Newton δεν συγκλίνει.

(η μέθοδος Newton μπορεί να αποκλίνει ανεξάρτητα από το αν το $x_0$ επιλέγεται κοντά στο $x^*$ ή όχι)

Παράδειγμα 3

Δείξτε ότι η ακολουθία της Newton για την $f(x) = x^{1/3} = 0$ αποκλίνει για οποιοδήποτε $x_0 \neq 0$.

Λύση

$$f'(x) = \frac{1}{3} x^{-2/3} \text{ άρα } x_{n+1} = x_n - \frac{x_n^{1/3}}{\frac{1}{3} x_n^{-2/3}} =$$

$$= x_n - 3x_n \Rightarrow x_{n+1} = -2x_n.$$

και $\left| \frac{x_{n+1}}{x_n} \right| = 2 > 1$ οπότε η $x_n$ αποκλίνει.

αποδείξτε ότι αποκλίνει ...
υπολογίσω $\left| \frac{x_{n+1}}{x_n} \right|$ ...

Βελτιστοποίηση - 8

- Κεφάλαιο 4 -

• Μέθοδος Newton στην Αδέσμευτη βελτιστοποίηση -

μέθοδος αναζήτησης ευθείας

$$x_{k+1} = x_k - \alpha_k D_k \nabla f(x_k)$$

κατεύθυνση αναζήτησης $p_k = -D_k \nabla f(x_k)$

→ Μέθοδος Newton ⇒ ειδική επιλογή πίνακα $D_k = [\nabla^2 f(x_k)]^{-1}$

άρα προκύπτει ο τύπος επανάληψης μεθόδου Newton.

$$\boxed{x_{k+1} = x_k - \alpha_k \underbrace{[\nabla^2 f(x_k)]^{-1} \cdot \nabla f(x_k)}_{= -p_k}}$$

Είναι πράγματι κατεύθυνση καθόδου αφού $\nabla f(x_k)^T p_k < 0$.

για $\nabla f(x_k) \neq 0$ και $\nabla^2 f(x_k)$ θετικά ορισμένος.

! Η εφαρμογή της μεθόδου Newton, σε κάθε βήμα ελαχιστοποιεί την τετραγωνική συνάρτηση $q_k(p)$, η οποία προσεγγίζει κατά Taylor την αντικειμενική συνάρτηση $f(x)$ στην περιοχή του σημείου $x_k$.

# Αλγόριθμος Newton

1. Αρχικοποίηση:

- επιλογή αρχικής προσέγγισης $\boxed{x_0}$
- σταθεράς τερματισμού $\varepsilon$
- θέτω $k=0$.

2. Κύριο βήμα:

α. Αν $\|\nabla f(x_k)\| < \varepsilon$

→ τερματισμός

→ προσεγγιστική τιμή $= x_k$.

β. Αλλιώς υπολογίζουμε τα Hessian Matrix $\nabla^2 f(x_k)$ και την κατεύθυνση αναζήτησης

$$p_k = -[\nabla^2 f(x_k)]^{-1} \nabla f(x_k)$$

γ. Προσδιορίζουμε ένα βήμα μήκους $a_k$ που ελαχιστοποιεί τη συνάρτηση $\varphi(a_k) = f(x_k + a_k p_k)$ ως προς $a_k > 0$.

δ. Προσδιορίζουμε

$$x_{k+1} = x_k + a_k p_k$$

ε. Θέτω $k = k+1$ και επαναλαμβάνω το κύριο βήμα.

Βελτιστοποίηση 9

## Παράδειγμα 4.1 (Newton για τετραγωνικές συναρτήσεις)

περίπου απόδειξη.

Έστω η τετραγωνική συνάρτηση $f$:

$$f(x) = \frac{1}{2}x^TQx - b^Tx + c$$

Επιβεβαιώστε ότι η ακολουθία $x_{k+1} = x_k - \alpha_k[\nabla^2 f(x_k)]^{-1}\nabla f(x_k)$ της μεθόδου Newton με $\alpha_k = 1$, συγκλίνει για οποιοδήποτε αρχικό $x_0$, σε ένα μόνο βήμα στο $x^*$. (δηλαδή $x_1 = x^*$)

λύση

Ισχύει

$$x^* = Q^{-1}b$$

για $k=0$ και $\alpha_0 = 1$

$$x_1 = x_0 - [\nabla^2 f(x_0)]^{-1}\nabla f(x_0) \quad (1)$$

Ισχύουν

$$\nabla^2 f(x_0) = Q$$

$$\nabla f(x) = Qx - b$$

άρα από $(1) \Rightarrow x_1 = x_0 - Q^{-1}(Qx_0 - b)$

$$\Rightarrow x_1 = x_0 - \frac{\cancel{Q}x_0}{\cancel{Q}} + Q^{-1}b = Q^{-1}b = x^*.$$

4.2

1 Στη μέθοδο Newton, θεωρούμε σταθερό βήμα μήκους $a_k = 1$.
Όταν αυτό όμως δεν προσφέρει ικανοποιητική μείωση, πάμε στο βήμα 2.γ του αλγορίθμου.

Παράδειγμα 4.2 — Δοκιμάστε άλλο για ποιο $a_k$ θα συγκλίνει η μέθ. Newton..

minimize $f(x) = x^{\frac{4}{3}}$

για $a_k = 1$, για $n = 1$ έχουμε:

$$x_{k+1} = x_k - \frac{f'(x_k)}{f''(x_k)} \quad \text{με } f'(x) = \frac{4}{3} x^{\frac{1}{3}}$$

$$f''(x) = \frac{4}{9} x^{-2/3}$$

$\Rightarrow x_{k+1} = -2x_k$ και αφού $\left|\frac{x_{k+1}}{x_k}\right| = 2 > 1$

$\rightarrow$ αποκλίνει για οποιαδήποτε επιλογή $x_0$.

Όμως αν $a_k \neq 1$ ($a_k > 0$)

θα έχουμε $x_{k+1} = (1 - 3a_k) \cdot x_k$.

άρα θέλουμε $\frac{|x_{k+1}|}{|x_k|} = |1 - 3a_k| < 1 \Rightarrow 0 < a_k < \frac{2}{3}$

για να έχουμε σύγκλιση.

# Βελτιστοποίηση 10

(2) Στην περίπτωση που ο Hessian Matrix $\nabla^2 f(x_k)$ στο βήμα $k$ δεν είναι θετικά ορισμένος,
η κατεύθυνση αναζήτησης $p_k$ ενδέχεται να μην ορίζεται.
Αλλά ακόμα και αν ο $\nabla^2 f(x_k)$ είναι αντιστρέψιμος ενδέχεται να μην ικανοποιείται η κατεύθυνση καθόδου $\nabla f(x_k)^T p_k < 0$.

→ Σε αυτές τις περιπτώσεις η ακολουθία $x_k$ $\begin{cases} \text{αποκλίνει} \\ \text{[illegible]} \end{cases}$

(3) Η γενική μέθοδος Newton με $p_k = -[\nabla^2 f(x_k)]^{-1} \nabla f(x_k)$
αναφέρεται να έχει υψηλό, τετραγωνικό, ρυθμό σύγκλισης.
συνήθως

Υπολογίζουμε μεταβλητά σε κάθε βήμα → βρίσκω Hessian Matrix.
Για την αντιμετώπιση έχουν διατυπωθεί quasi-Newton methods που περιέχουν μια εναλλακτική κατεύθυνση, που προσφέρουν σύγκλιση με υπεργραμμική τάξη.

$$p_k = -B_k^{-1} \nabla f(x_k)$$

→ προσέγγιση του Hessian Matrix.

(4) Κλασσική μέθοδος Newton

$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)}$$

, επίλυση της $\nabla f(x) = 0$
ανακτώντας συνθήκη 1ης τάξης

→ και με αυτήν την επίλυση λαμβάνουμε του (4.2)

Παράδειγμα 4.3 SOS

$e_k = x_k - x^*$

minimize $f(x) = x^4 - 1$

προσδιορίστε τις 3 πρώτες επαναλήψεις με $x_0 = 4$

$f'(x) = 4x^3$, $f''(x) = 12x^2$

παίρνω χωρίς τον τύπο αφού n=1

$$x_{k+1} = x_k - \frac{f'(x_k)}{f''(x_k)} = \frac{2}{3} x_k$$

άρα $x_0 = \frac{2}{3} x_k$

$x_0 = 4 \left(\frac{2}{3}\right)^k$

για $k=0$, $x_0 = 4$.

$$\frac{|x_{k+1}|}{|x_k|} = \frac{2}{3} < 1$$ άρα συγκλίνει για οποιαδήποτε αρχ. επιλογή $x_0$

$$x_k = \left(\frac{2}{3}\right)^k$$

$x_0 = 4$, $x_1 = \frac{8}{3}$, $x_2 = \frac{16}{9}$, $x_3 = \frac{32}{27}$.

Υπολογίζουμε τον $x^* = \lim_{k \to +\infty} x_k = 4 \lim_{k \to +\infty} \left(\frac{2}{3}\right)^k = 0$

για ρυθμό σύγκλισης:

$$\lim_{k \to +\infty} \frac{|e_{k+1}|}{|e_k|} = \frac{|x_{k+1} - x^*|}{|x_k - x^*|} = \frac{2}{3}$$

άρα $r = 1$: ~~γραμμ.~~ ρυθμός σύγκλισης, και $c = \frac{2}{3}$

Foulis + verde soie

ο Σωκράτης αγαπάει την Δέσποινα [και τα γράμματά του δεν διαβάζονται]