ΠΕΡΙΕΧΟΜΕΝΟ ΜΑΘΗΜΑΤΟΣ

1. Αριθμοί κινητής υποδιαστολής

→ σφάλματα στρογγύλευσης

Οι πράξεις γίνονται με πεπερασμένη ακρίβεια και <u>αυτό</u> <u>μπορεί</u> να οδηγήσει σε αλλοίωση των αποτελεσμάτων.

Ευστάθεια προβλημάτων σε σφάλματα αλγορίθμων στρογγύλευσης.

2. Μη γραμμικές εξισώσεις

Έστω $f:[a,b]\to\mathbb{R}$ μια συνάρτηση

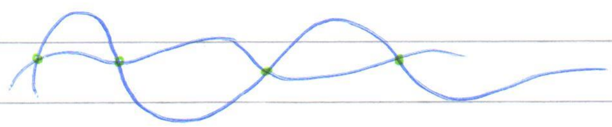
Θα μελετήσουμε αριθμητικές μεθόδους προσέγγισης ριζών της f.

3. Γραμμικά συστήματα

$A\in\mathbb{R}^{n,n}$ $b\in\mathbb{R}^n$

Ζητείται $x\in\mathbb{R}^n$ τ.ω. $Ax=b$

4. Παρεμβολή



5. Αριθμητική ολοκλήρωση

Δεδομένα $f:[a,b]\to\mathbb{R}$

Ζητούμενα $\int_a^b f(x)dx$

12/2/2015

ΑΡΙΘΜΗΤΙΚΗ ΚΙΝΗΤΗΣ ΥΠΟΔΙΑΣΤΟΛΗΣ

"Σφάλματα στρογγύλευσης"

→ Καλές μέθοδοι δίνουν καλά αποτελέσματα
→ Κακές μέθοδοι μπορεί να δώσουν άσχημα αποτελέσματα

Ποιότητα αριθμητικών μεθόδων
- απαιτούμενη μνήμη
- απαιτούμενος χρόνος
- ακρίβεια

Τα σφάλματα στρογγύλευσης μπορεί να οδηγήσουν σε λανθασμένα αποτελέσματα.

Παράδειγμα

$$I_n = \int_0^1 \underbrace{x^n e^{x-1}}_{\geq 0} dx \qquad n \in \mathbb{N}$$

Ιδιότητες των I_n

α) $I_n > 0$

β) $0 < x < 1 \Rightarrow x^{n+1} < x^n \Rightarrow I_{n+1} < I_n$

γ) $(I_n)_{n\in\mathbb{N}}$ είναι γνησίως φθίνουσα

δ) $e^{x-1} \leq 1 \Rightarrow \int_0^1 x^n \cdot e^{x-1} dx \leq \int_0^1 x^n dx = \frac{1}{n-1}$

Συμπέρασμα: $0 < I_n \leq \frac{1}{n-1}$

$\frac{1}{n-1} \longrightarrow 0$ για $n \longrightarrow 0$

Άρα $I_n \to 0$ για $n \to \infty$

Δηλαδή η $(I_n)_{n\in N}$ είναι μοναδική

$$I_n = \int_0^1 x^n \cdot e^{x-1}\,dx = \int_0^1 x^n (e^{x-1})'\,dx = [\underbrace{x^n e^{x-1}}_{1}]_0^1 - \int_0^1 (x^n)' e^{x-1}\,dx$$

Άρα $I_{n-1} = 1 - \int_0^1 n x^{n-1} e^{x-1}\,dx = 1 - nI_{n-1}$

$$I_n = \int_0^1 x(e^{x-1})'\,dx = [xe^{x-1}]_0^1 - \int_0^1 e^{x-1}\,dx = 1 - [e^{x-1}]_0^1 =$$

$$1 - (e^0 - e^{-1}) = e^{-1} = \frac{1}{e}$$

$I_n = 1 - nI_{n-1}$ $n \geq 2$

$I_1 = \frac{1}{e}$

$\tilde{I}_1 \approx I_1$

Τότε $\tilde{I}_1 = 1 - nI_{n-1}$ $n \geq 2$

Ερώτημα: Τι μπορούμε να κάνουμε για τα σφάλματα $I_n - I_{n-1}$

Έχουμε * $I_n - \tilde{I}_n = nI_{n-1} + n\tilde{I}_{n-1} = -n(I_{n-1} - \tilde{I}_{n-1})$

Ισχυρισμός: $I_n - \tilde{I}_n = (-1)^{n-1} n! (I_1 - \tilde{I}_1)$

Επαγωγικά: $n = 2$

$n \longrightarrow n+1$

$$I_{n+1} - I_{n+1} \overset{*}{=} -(n+1)(I_n - \tilde{I}_n) = -(n+1)(-1)^{n-1} n! (I_1 - \tilde{I}_1) =$$
$$(-1)^n (n+1)! (I_1 - \tilde{I}_1) = (-1)^{(n+1)-1} (n+1)! (I_1 - \tilde{I}_1)$$

Άρα $|I_n - \tilde{I}_n| = n! \, |I_1 - \tilde{I}_1|$

↑

αυξάνει πολύ γρήγορα με το n

Συμπέρασμα: Ο αλγόριθμος είναι ασταθής ευαίσθητος σε σφάλματα στρογγύλλευσης

$$I_n = 1 - n \cdot I_{n-1} \Rightarrow I_{n-1} = \frac{1 - I_n}{n} \Rightarrow I_{n-1} - \tilde{I}_{n-1} = \frac{1}{n}(I_n - \tilde{I}_n)$$

Επαγωγικά: $I_n - \tilde{I}_n = (-1)^{m-n} \dfrac{1}{(m+1)(m+2)\ldots m}(I_n - \tilde{I}_m) \Rightarrow$

$$|I_n - \tilde{I}_n| = \frac{1}{(m+1)(m+2)\ldots m} |I_n - \tilde{I}_m|$$

Αν θέλω να υπολογίσω το I_n για ένα συγκεκριμένο "n" εφαρμόζω τον αλγόριθμο:

$$I_{j-1} = \frac{1 - I_j}{j}, \quad j = m, m-1, m-2, \ldots, n+1 \quad m > n$$

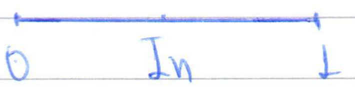
Όσο μεγαλύτερη είναι η διαφορά $m-n$ τόσο μεγαλύτερο είναι το αποτέλεσμα.

Ποια αρχική τιμή $\tilde{I}_n$ θα πάρουμε;
Ξέρουμε ότι $0 < I_n < \dfrac{1}{n+1}$

Η καλύτερη επιλογή αρχικής τιμής είναι: $\tilde{I}_n = \dfrac{1}{2(m+1)}$

Τότε $|\tilde{I}_m - I_n| \le \dfrac{1}{2(m+1)}$

Για $\tilde{I}_m = 0$ έχουμε $|I_m - \tilde{I}_n| \le \dfrac{1}{m+1}$



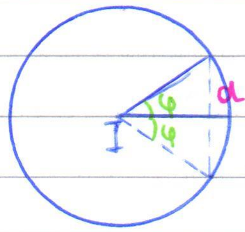
2ο παράδειγμα

Προσέγγιση του "π" με τη μέθοδο του Αρχιμήδη.

Θεωρώ $y_n = 2^n \sin\frac{\pi}{2^n}$ $n \in \mathbb{N}$

$y_1 = 2\sin\frac{\pi}{2} = 2$

Ποιά είναι η γεωμετρική ερμηνεία για το y_n, $n \geq 2$;



$\varphi = \frac{\pi}{2}$ $\alpha = \sin\frac{\pi}{2^n}$

$2\varphi = \frac{2\pi}{2^n}$

Το 2α θα είναι η πλευρά του κανονικού πολυγώνου με 2^n πλευρές που είναι εγγεγραμμένο στο μοναδιαίο κύκλο.

περίμετρος: $2^n \cdot 2\alpha = 2^n \cdot 2 \cdot \sin\frac{\pi}{2} = 2y$

Το y_n είναι η ημιπερίμετρος του πολυγώνου.

2º Μάθημα

13/2/2014

<u>Ισχυρισμός:</u> Η $(y_n)_{n\in\mathbb{N}}$ είναι γνησίως αύξουσα και συγκλίνει στο π.

$$y_n = 2^n \sin\frac{\pi}{2^n} \underset{\uparrow}{=} 2^n \cdot 2\sin\frac{\pi}{2^{n+1}} \cdot \cos\frac{\pi}{2^{n+1}} \Rightarrow$$

$\sin(2x) = 2\sin x\cdot\cos x$

$$\Rightarrow y_n = \underbrace{2^{n+1}\sin\frac{\pi}{2^{n+1}}}_{y_{n+1}} \cos\frac{\pi}{2^{n+1}} \quad \text{άρα} \quad y_n = y_{n+1}\cdot\cos\frac{\pi}{2^{n+1}}$$

Επειδή $0 < \cos\frac{\pi}{2^{n+1}} < 1$ συμπεραίνουμε ότι: $y_n < y_{n+1}$

$\lim\limits_{x\to 0}\frac{\sin x}{x} = 1$. Άρα $\lim\limits_{x\to 0}\frac{\sin(ax)}{x} = a$ με $a\in\mathbb{R}$

Επομένως $\lim\limits_{n\to\infty}\underbrace{\frac{\sin\frac{\pi}{2^n}}{\frac{1}{2^n}}}_{=y_n} = \pi$ ή $\lim\limits_{n\to\infty} y_n = \pi$

Αναδρομικός τύπος: $y_{n+1} = 2^{n+1}\sin\frac{\pi}{2^{n+1}} \underset{\uparrow}{=} 2^{n+1}\left\{\frac{1-\cos\frac{\pi}{2^n}}{2}\right\}^{1/2} =$

$\sin^2 x = \frac{1-\cos(2x)}{2}$

$$2^{n+1}\left\{\frac{1-\sqrt{1-\sin^2\frac{\pi}{2^n}}}{2}\right\}^{1/2}$$

Αλλά $y_n = 2^n\sin\frac{\pi}{2^n} \Rightarrow \sin\frac{\pi}{2^n} = 2^{-n}y_n$

<u>Συμπέρασμα:</u>
$$\begin{cases} y_{n+1} = 2^{n+1}\sqrt{\frac{1}{2}\left(1-\underset{\uparrow}{\sqrt{1-(2^{-n}y_n)^2}}\right)} & n=1,2,\ldots \\ y_1 = 2 \end{cases}$$

ασταθής αλγόριθμος λόγω αφαίρεσης

$$1-\sqrt{1-(2^{-n}y_n)^2} = \frac{1-\left[1-(2^{-n}y_n)^2\right]}{1+\sqrt{1-(2^{-n}y_n)^2}} = \frac{2^{-n}y_n}{1+\sqrt{1-(2^{-n}y_n)^2}}$$

Θυμάμαι:

$$\sqrt{a}-\sqrt{\beta}=\frac{(\sqrt{a}-\sqrt{\beta})(\sqrt{a}+\sqrt{\beta})}{\sqrt{a}+\sqrt{\beta}}=\frac{a-b}{\sqrt{a}+\sqrt{\beta}}$$

Επομένως $y_{n+1}=2^{n+1}\sqrt{\frac{(2^{-n}y_n)^2}{2(1+\sqrt{1-(2^{-n}y_n)^2}}}=2^{n+1}\sqrt{\frac{1}{2(1+\sqrt{1-(2^{-n}y_n)^2})}}=$

$\sqrt{\frac{2}{1+\sqrt{1-(2^{-n}y_n)^2}}}$ ευσταθής αλγόριθμος.

17/2/2015

"Παράσταση αριθμών ως προς οποιαδήποτε βάση"

Καθημερινή ζωή: Δεκαδικό σύστημα
βάση: 10
ψηφία: $0, 1, 2, \ldots, 9$
παράδειγμα: $3{,}14159 = 3 \cdot 10^0 + 1 \cdot 10^{-1} + 4 \cdot 10^{-2} + 1 \cdot 10^{-3} + 5 \cdot 10^{-4} + 9 \cdot 10^{-5}$
γενικά: Έστω $\alpha_N, \alpha_{N-1}, \ldots, \alpha_0, \alpha_{-1}, \alpha_{-2} \ldots$ δεκαδικά ψηφία
Τότε $(\alpha_N \alpha_{N-1} \ldots \alpha_0 \alpha_{-1} \alpha_{-2})_{10} = + \alpha_{-1} \cdot 10^{-1} + \alpha_{-2} \cdot 10^{-2} + \ldots$

Ακέραιο μέρος $\alpha_N \alpha_{N-1} \ldots \alpha_0$
$p(x) = \alpha_N x^N + \alpha_{N-1} x^{N-1} + \ldots + \alpha_0$

Η τιμή του p στο σημείο $x = 10$ είναι ο αριθμός
$\alpha_N \alpha_{N-1} \ldots \alpha_0$

Κλασματικό μέρος $\alpha_{-1} \alpha_{-2} \ldots$
Η τιμή της δυναμοσειράς

Η σειρά μπορεί να έχει είτε πεπερασμένο είτε άπειρο πλήθος όρων

παράδειγμα
$4.130 = 4.129\bar{9} = 4{,}12999\ldots$

$9 \sum_{i=1}^{\infty} 10^{-i} = 1$ ή $\frac{1}{3} = 0{,}333\ldots$

Για να έχουμε μοναδικότητα: για κάθε $k_0 \in \mathbb{N}$ απαιτούμε αν υπάρχει $k \geq k_0$ τ.ω. $\alpha_{-k} \neq 9$

Σύστημα με βάση β β>2

Βάση: β
Ψηφία: 1, 2, 3, 4, ..., β-1
a_k ψηφία
$\pm(a_N a_{N-1} \dots a_0 a_{-1} a_{-2} \dots)_\beta =$
$\pm(a_N \cdot \beta^N + a_{N-1} \cdot \beta^{N-1} + \dots + a_0 \beta^0 + a_{-1} \beta^{-1} + a_{-2} \beta^{-2} + \dots)$

παράδειγμα
$(100110,11) = 1 \cdot 2^5 + 1 \cdot 2^2 + 1 \cdot 2^1 + 1 \cdot 2^{-1} + 1 \cdot 2^{-2} = (38,75)_{10}$

1) Μετατροπή απο σύστημα με βάση β στο δεκαδικό.
α. ακέραιων αριθμών
παράδειγμα
$(53473)_8 = 5 \cdot 8^4 + 3 \cdot 8^3 + 4 \cdot 8^2 + 7 \cdot 8^1 + 3 \cdot 8^0 =$
$5+8+(7+8+(4+8+(3+5+8)))$
σχήμα του Horner
Κάνουμε τις πράξεις από μέσα προς τα έξω.
$p(x) = a_N x^N + a_{N-1} x^{N-1} + \dots a_0 =$
$a_0 + x(a_1 + x(a_2 + x(a_3 + \dots + x(a_{N-1} + x a_N)$
$y \leftarrow a_N$ για $i = N-1, \dots, 0$
$y \leftarrow 0 : x + y$
Τότε $y = P(x)$
$y \leftarrow a_i + xy$ flop (floting point operation)
Ο υπολογισμός της τιμής του p σε ένα σημείο απαιτεί (με το σχήμα του Horner) N-flop

β. κλασματικών αριθμών
$x \ (0 < x < 1)$
παράδειγμα
$(.11)_2 = 1 \cdot 2^{-1} + 1 \cdot 2^{-2} = \frac{1}{2} + \frac{1}{4} = (0,75)_{10}$

2) Μετατροπή από το δεκαδικό σε σύστημα με βάση β

α. ακέραιων αριθμών

Βασίζεται στον αλγόριθμο της διαίρεσης

παράδειγμα

Μετατροπή του $(369)_{10}$ στο οκταδικό σύστημα.

$(369)_{10} = (\ldots \alpha_2 \alpha_1 \alpha_0)_8 = \alpha_0 + 8(\alpha_1 + 8(\alpha_2 \ldots))$

Άρα το α_0 είναι το υπόλοιπο και το $\alpha_1 + 8(\ldots)$ είναι το πηλίκο της διαίρεσης $369:8$

$$\begin{array}{r|l} 369 & 8 \\ \hline 49 & 46 \\ 1 & \end{array}$$

$46 = \alpha_1 + 8(\alpha_2 \ldots)$

Το α_1 είναι το υπόλοιπο και το $\alpha_2 + \ldots$ είναι το πηλίκο της διαίρεσης $46:8$

$$\begin{array}{r|l} 46 & 8 \\ \hline 6 & 5 \end{array}$$

Άρα $\alpha_1 = 6$ και $5 = \alpha_2 + 8(\alpha_3 \ldots)$

Συνεπώς $\alpha_2 = 5$ και $0 = \alpha_3 + \ldots$

Συμπέρασμα $(369)_{10} = (561)_8$

Επαλήθευση: $(561)_8 = 5 \cdot 8^2 + 6 \cdot 8^1 + 1 \cdot 8^0 = 369$

β. κλασματικών αριθμών

$0 < x < 1$ x: στο δεκαδικό σύστημα.

$x = (0{,}\alpha_{-1} \alpha_{-2} \ldots)_\beta = \alpha_{-1} \beta^{-1} + \alpha_{-2} \beta^{-2} + \ldots$

Άρα $\beta x = \alpha_{-1} + \alpha_{-2} \beta^{-1} + \ldots$

Συμπέρασμα: Το α_{-1} είναι το ακέραιο μέρος του βx.

παράδειγμα

Μετατροπή του $x = (0{,}372)_{10}$ στο δεκαδικό σύστημα

$(372)_{10} = (\alpha_{-1} \alpha_{-2} \ldots)_2$

Έχουμε $2x = 0{,}744$ άρα $\alpha_{-1} = 0$ $\gamma_1 = 0{,}744$

$2\gamma_1 = 1{,}488$ άρα $\alpha_{-2} = 1$, $\gamma_2 = 0{,}488$

$2\gamma_2 = 0,976$ άρα $\alpha_{-3} = 0$, $\gamma_3 = 0,976$
$2\gamma_3 = 1,952$ άρα $\alpha_{-4} = 1$, $\gamma_4 = 0,952$

Παρατήρηση: Είναι δυνατόν το πλήθος των ψηφίων να είναι πεπερασμένος σε ένα σύστημα και άπειρο σε ένα άλλο.

Παράδειγμα.

$$\frac{1}{10} = \sum_{n=1}^{\infty}\left(\frac{1}{2^{4n}} + \frac{1}{2^{4n+1}}\right)$$

Για $n=1$: $\underbrace{2^{-4} + 2^{-5}}_{n=1} + \underbrace{2^{-8} + 2^{-9}}_{n=2} + \dots = .00011\,0011\dots = (.0\overline{0011})$

2^4 2^5 2^8

$$\sum_{n=1}^{\infty}\left(\frac{1}{2^{4n}} + \frac{1}{2^{4n+1}}\right) =$$

$$\sum_{n=1}^{\infty}\left(1 + \frac{1}{2}\right)\frac{1}{2^{4n}} =$$

$$\frac{3}{2}\sum_{n=1}^{\infty}\left(\frac{1}{2^4}\right)^n =$$

$$\frac{3}{2}\,\frac{\frac{1}{2^4}}{1-\frac{1}{2^4}} = \frac{3}{2} = \frac{1}{2^4-1} = \frac{3}{2}\,\frac{1}{15} = \frac{1}{10}$$

Θυμάμαι

$$\sum_{n=1}^{\infty} \omega^n = \frac{\omega}{1-\omega} \quad |\omega|<1$$

"Αριθμοί μηχανής"

Έστω $x \in \mathbb{R}$ $x \neq 0$
Σε ένα σύστημα με βάση b, ο x μπορεί να γραφεί ως
$x = \pm .d_1 d_2 \dots b^e$ με $d_1 \neq 0$ (*)
και $d_1, d_2, \dots$ ψηφία ως προς τη βάση b και e κατάλληλος

ακέραιος.
Η μορφή ✱ λέγεται (κανονική μορφή κινητής υποδιαστολής).

ε- Το σύνολο των αριθμών μηχανής $M=(b, t, L, U)$ χαρακτηρίζεται από τις παραμέτρους:
→ b: βάση του αριθμητικού συστήματος
→ t: ακρίβεια = πλήθος των ψηφίων του κλάσματος των αριθμών
→ L: κάτω φράγμα } του εκθέτη e του b
U: άνω φράγμα }
5) $(L \leq e \leq U)$
L, U ακέραιοι $L \approx U$
Κάθε $x \in M$, $x \neq 0$, είναι της μορφής
(+) $x = \pm .d_1 d_2 \ldots d_t b^e$ με $d_1 \neq 0$ και $L \leq e \leq U$.
Το M αποτελείται από όλους τους αριθμούς της μορφής (+) και το 0.
Το M είναι πεπερασμένο.
Μέγιστο στοιχείο του M: $d_i = b-1$, $i = 1, \ldots, M$ κ' $e = U$
Ελάχιστο κατ' απόλυτη τιμή μη μηδενικό στοιχείο του M:
$d_1 = 1$, $d_i = 0$, $i = 2, \ldots, t$
$.100\ldots0 \cdot \beta^L$

26/2/2015

"Αριθμοί μηχανής"

Το σύνολο $M = M(\beta, t, L, U)$ των αριθμών μηχανής αποτελεί-ται από το 0 και τους αριθμούς της μορφής
$x = .d_1 d_2 \ldots d_t \cdot \beta^e$ με $d_1 \neq 0$ και $L \leq e \leq U$

Μέγιστο στοιχείο του M
$x = .d_1 d_2 \ldots d_t \cdot \beta^U$ με $d_1 = d_2 = \ldots = d_t = \mu - 1$.

Ελάχιστο θετικό στοιχείο του M
$.100\ldots0\ \beta^L$

Παρατηρήσεις
1. Η απόσταση μεταξύ δύο διαδοχικών στοιχείων του M δεν είναι σταθερή.

2. Το M δεν είναι κλειστό ως προς τον πολλαπλασιασμό, δλδ γενικά αν έχω 2 αριθμούς $x, x^* \in M \not\Rightarrow x x^* \in M$
(πράγματι: $\underbrace{0.10\ldots0\ \beta^t}_{\in M} \cdot \underbrace{0.100\ldots0\ \beta^t}_{\in M} \notin M$)

3. Το M δεν είναι κλειστό ως προς την πρόσθεση
παράδειγμα
$\beta = 10,\ t = 5$
$1,\ 10^{-5} \in M,\ 1 + 10^{-5} = 1.0001 \notin M$

Μας ενδιαφέρει το M να είναι όσο πιο πυκνό και ευρύ γίνεται δηλαδή να έχει: → μεγάλο t
→ μεγάλο διάστημα $[L, U]$

15 "Προσέγγιση πραγματικού αριθμού με αριθμούς μηχανής"

(α) $|x| > .d_1 \dots d_t \, \beta^u$ με $d_1 = \dots = d_t = \beta - 1$
υπερχείλιση (overflow) σταματούν οι υπολογισμοί.

(β) $0 < |x| < .100\dots0 \, \beta^t$
υπερχείλιση
κατά κανόνα προσεγγίζεται το x με μηδέν και συνεχίζονται οι υπολογισμοί.

(γ) $1.\beta^t < |x| < .d_1 \dots d_t \cdot \beta^u$
με $d_1 = \dots = d_t = \beta \to$
Τότε ο x προσεγγίζεται με έναν αριθμό $fl(x) \in M$

Συνήθως ισχύει:
(+) $|x - fl(x)| \le |x - y| \ \forall \ y \in M$

Ισχυρισμός: Αν ισχύει η (+), τότε έχουμε:
(*) $\left| \frac{fl(x) - x}{x} \right| \le \frac{1}{2} \beta^{1-t} \longrightarrow$ το μέγιστο σφάλμα είναι πάντα ίδιο.

(α) $fl(x) = x$. δηλαδή $x \in M$.
Τότε η (*) ισχύει προφανώς

(β) Αν $x \notin M$, τότε υπάρχουν $x', x'' \in M$ διαδοχικοί τ.ω. $x' < x < x''$
Προφανώς, τότε έχουμε $|fl(x) - x| \le \frac{1}{2} |x' - x''|$

$$\min(x - \alpha, b - x) \le \frac{b - \alpha}{2}$$

χ.β.γ. υποθέτουμε ότι $x > 0$
Τότε $x = .d_1 d_2 \dots d_t, d_{t+1}, \dots \beta^t$

Επομένως, $x' = .d_1 d_2 \dots d_t \beta^k$ και $x'' = (.d_1 d_2 \dots d_t \beta^t) \beta^k$

Συνεπώς $x'' - x' = \beta^{k-t}$

Αυτή είναι η διαφορά διαδοχικών αριθμών, που είναι τόσο πιο μεγάλη όσο πιο μεγάλο είναι το κ.

Άρα $|fl(x) - x| \leq \frac{1}{2} \beta^{k-t}$

Επομένως, $\left| \frac{fl(x) - x}{x} \right| \leq \frac{1}{2x} \beta^{k-t}$

Επίσης έχουμε: $x' \geq .1 \beta^k$, οπότε $\left| \frac{fl(x) - x}{x} \right| \leq \frac{\beta^{k-t}}{0.1 . \beta^k} =$

$\frac{\beta^{-t}}{2\beta^{-1}} = \frac{1}{2} \beta^{1-t}$

(+) $|x - fl(x)| \leq |x - y| \quad \forall y \in M$

Ισχύει όταν από τον x οδηγούμαστε στον $fl(x)$ με στρογγύλευση

Παράδειγμα

$\beta = 10$, $t = 5$

$x = .a_1 a_2 \dots a_5 a_6 \dots 10^k$

Αν $a_6 \geq 5$ τότε $fl(x) = x'' = (.a_1 a_2 \dots a_5 + 10^{-5}) \cdot 10^k$

Αν $a_6 < 5$ τότε $fl(x) = x' = .a_1 a_2 \dots a_5 \cdot 10^k$

Στην περίπτωση $a_6 = 5$, τότε και $a_i = 0$, $i \geq 2$, μπορούμε να επιλέξουμε είτε τον x' είτε τον x''.

Αποκοπή:

Σε αυτή την περίπτωση έχουμε:

$\left| \frac{fl(x) - x}{x} \right| \leq \beta^{1-t}$

Γενικά έχουμε: $\left|\frac{fl(x)-x}{x}\right| \leq u$ με $u = \begin{cases} \frac{1}{2}\beta^{1-t}, & \text{για στρογγύλευση} \\ \beta^{1-t}, & \text{για αποκοπή} \end{cases}$

u: μοναδιαίο σφάλμα στρογγύλευσης

"Πράξεις"

$* \in \{+, -, \times, :\}$

$x * y$

Υπόθεση: $z = fl(\underbrace{fl(x) * fl(y)}_{\text{γίνεται ακριβώς}})$

Παράδοξα: $\beta = 10$, $t = 5$, $U = -L = 10$ στρογγύλευση

$a_1 = 1$, $a_2 = 3 \cdot 10^{-5}$, $a_3 = 3 \cdot 10^{-5}$ $(a_3 = a_2)$

$a_1, a_2, a_3 \in M$

Τότε $a_1 + a_2 = fl(a_1 + a_2) = fl(1.00003) = 1.0000$

(κανονικά $fl(a_1) + fl(a_2)$ αλλά είναι ίδια!)

Το $3<5$ άρα το ξεχνάω. Αν ήταν μεγαλύτερο, θα έβαζα 1 στο προηγούμενο.

$fl(fl(a_1 + a_2) + a_3) = 1$

$fl(a_2 + a_3) = 6 \cdot 10^{-5}$

$fl(a_1 + fl(a_2 + a_3)) = fl(1.00006) = 1.0001$

Έχω διαφορετικά αποτελέσματα

Συμπέρασμα: Έχει σημασία η σειρά με την οποία εκτελούνται οι πράξεις (στον υπολογισμό ενός αθροίσματος)

- Για κάθε $0 < |x| < 5 \cdot 10^{-5}$ έχουμε $fl(1 + fl(x)) = 1$
  (Γενικά το $\frac{1}{2}\beta^{1-t}$ λέγεται έψιλον ή μηδέν της μηχανής)

- Γενικά για $0 < |x| < \frac{1}{2}\beta^{1-t}$ έχουμε $fl(1 + fl(x)) = 1$

Επιρροή των σφαλμάτων στρογγύλευσης στους υπολογισμούς.

Υπόθεση: x, y, $x * y$ μη μηδενικοί αριθμοί στο εύρος των αριθμών μηχανής

Στόχος: Εκτίμηση του σχετικού σφάλματος $\dfrac{fl(fl(x)+fl(y)) - xy}{x*y}$

(*) $\left|\dfrac{fl(x)-x}{x}\right| \le u$

Ισχυρισμός: $fl(x) = x(1+\varepsilon)$ με $\varepsilon = \varepsilon(x)$ τ.ω. $|\varepsilon(x)| \le u$

$\varepsilon = \dfrac{fl(x)-x}{x} \iff fl(x) - x = \varepsilon \cdot x \iff fl(x) = x + \varepsilon \cdot x = x(1+\varepsilon)$

Ισχυρισμός: Αν $\varepsilon_1, \dots, \varepsilon_m$ τ.ω. $|\varepsilon_i| \le u$ $i = 1, \dots, m$ τότε υπάρχει ε τ.ω. $|\varepsilon| \le u$ και $\lambda = \prod_{i=1}^{m} (1+\varepsilon_i) = (1+\varepsilon)^{m}$

Πράγματι, αν θεωρήσουμε τη συνάρτηση $\varphi(x) = (1+x)^m$, $x \in [-u, u]$ τότε θα έχουμε $\underbrace{(1-u)^m}_{\varphi(-u)} \le \lambda \le \underbrace{(1+u)^m}_{\varphi(u)}$

Επομένως, σύμφωνα με το θεώρημα της ενδιάμεσης τιμής, υπάρχει $\varepsilon \in [-u, u]$ τ.ω. $\lambda = \varphi(\varepsilon) = (1+\varepsilon)^m$

- $fl(x) = x(1+\varepsilon_1)$ $\quad |\varepsilon_i| \le u$
  $fl(x) = y(1+\varepsilon_2)$

1. Πολλαπλασιασμός

$z = fl(fl(x) \cdot fl(y)) = fl(x(1+\varepsilon_1) \cdot y(1+\varepsilon_2)) =$
$x \cdot y \underbrace{(1+\varepsilon_1)(1+\varepsilon_2)(1+\varepsilon_3)}_{(1+\varepsilon)^3} = x \cdot y (1+\varepsilon)^3$

Επομένως $\frac{z-xy}{x\cdot y} = \frac{xy(1+\varepsilon)^3 - xy}{xy} = 1+\varepsilon^3 - 1 =$

$1+3\varepsilon+3\varepsilon^2+\varepsilon^3-1 \Rightarrow \left|\frac{z-xy}{xy}\right| = |3\varepsilon+3\varepsilon^2+\varepsilon^3| \le 3u+\underbrace{3u^2+u^3}_{=O(u^2)}$

<u>Συμπέρασμα:</u> Το σχετικό σφάλμα στον πολλαπλασιασμό είναι το πολύ <u>τριπλάσιο</u> του μοναδιαίου σφάλματος στρογγύλευσης

27/2/2015

2. Διαίρεση

$z = fl\left(\frac{fl(x)}{fl(y)}\right) = fl\left(\frac{x(1+\varepsilon_1)}{y(1+\varepsilon_2)}\right) = \frac{x(1+\varepsilon_1)}{y(1+\varepsilon_2)}(1+\varepsilon_3) \quad \left(\begin{matrix}|\varepsilon_i|\le u\\ i=1,2,3\end{matrix}\right)$

Τώρα: $\frac{1}{1+\varepsilon_2} = 1+\delta$ με $\delta = -\frac{\varepsilon_2}{1+\varepsilon_2}$

Παρατηρούμε ότι: $|\delta| \le \frac{|\varepsilon_2|}{1+\varepsilon_2} \le \frac{u}{1-u} = u+O(u^2)$

Επομένως: $z = \frac{x}{y}\underbrace{(1+\varepsilon_1)(1+\varepsilon_3)}_{=(1+\varepsilon)^2}(1+\delta)$

οπότε: $\frac{z-\frac{x}{y}}{\frac{x}{y}} = (1+\varepsilon)^2(1+\delta)-1 = 2\varepsilon+\delta+\varepsilon^2+2\varepsilon\delta+\delta\varepsilon^2$

Συνεπώς: $\left|\frac{z-\frac{x}{y}}{\frac{x}{y}}\right| \le 3u+\alpha$ με $\alpha=O(u^2)$

3. Πρόσθεση-Αφαίρεση

$z = fl(fl(x)+fl(y)) = fl(x(1+\varepsilon_1)+y(1+\varepsilon_2)) =$

$x\underbrace{(1+\varepsilon_1)(1+\varepsilon_3)}_{=(1+\varepsilon)^2} + y\underbrace{(1+\varepsilon_2)(1+\varepsilon_3)}_{=(1+\delta)^2}$ με $|\varepsilon_i|\le u$, $|\varepsilon|\le u$, $|\delta|\le u$

Επομένως: $z = x+y+2(\varepsilon x+\delta y)+\underbrace{(x\varepsilon^2+y\delta^2)}_{=O(u^2)}$

Άρα $z \approx x+y+2(\varepsilon x+\delta y) \Rightarrow \dfrac{z-(x+y)}{x+y} \approx 2\,\dfrac{\varepsilon x+\delta y}{x+y} \Rightarrow$

$$\left|\frac{z-(x+y)}{x+y}\right| \le 2\,\frac{|\varepsilon x+\delta y|}{x+y} \le 2\,\frac{|x|+|y|}{|x+y|}\,u$$

1η περίπτωση: x, y ομόσημοι
τότε: $|x|+|y| = |x+y|$ οπότε $\left|\dfrac{z-(x+y)}{x+y}\right| \le 2u$

2η περίπτωση: x, y ετερόσημοι
Στη χειρότερη περίπτωση έχουμε $\varepsilon \approx -\delta$ και $|\varepsilon| \approx u$ δηλ. ε ή $-\varepsilon \approx u$
Τότε: $\left|\dfrac{\varepsilon x+\delta y}{x+y}\right| \approx \dfrac{|x|+|y|}{|x+y|}\,u$

Στην περίπτωση $x \approx -y$ ο παράγοντας $\dfrac{|x|+|y|}{|x+y|}$ μπορεί να γίνει πολύ μεγάλος

Σημείωση: Αν οι x και y είναι αριθμοί μηχανής και $x \approx -y$ τότε η $x+y$ γίνεται χωρίς πρόβλημα.
$z = fl(\underbrace{fl(x)}_{x} + \underbrace{fl(y)}_{y}) = fl(x+y) = (x+y)(1+\varepsilon) \Rightarrow$

$$\frac{z-(x+y)}{x+y} = \varepsilon \Rightarrow \left|\frac{z-(x+y)}{x+y}\right| = |\varepsilon| \le u$$

παράδειγμα
$\beta=10$, $t=5$, $U=-L=10$ στρογγύλευση
$x = .45142708$
$y = .45115944$
$x+y = .26764 \cdot 10^{-3}$

$z = fl(fl(x) + fl(y)) = fl(.45143 - .45116) = .00027 = .27000 \cdot 10^{-3}$

$$\left| \frac{z-(x+y)}{x+y} \right| \approx 88 \cdot 10^{-4}$$

$$u = \frac{1}{2} \cdot \frac{1}{2} \beta^{1-t} = 10^{1-5} = 10^{-4}$$

Πώς μπορούμε να αποφύγουμε την αφαίρεση σχεδόν ίσων αριθμών;

1º παράδειγμα

$\beta = 10, \quad t = 10$

$\sqrt{7298} - \sqrt{7297} = .5628470000 \cdot 10^{-2}$

Όμως: $\sqrt{7298} - \sqrt{7297} = \dfrac{1}{\sqrt{7298} - \sqrt{7297}} = \dots$

$$\sqrt{x} - \sqrt{y} = \frac{x-y}{\sqrt{x}+\sqrt{y}}$$

με οδηγεί στο σωστό αποτ.

2º παράδειγμα

$f(x) = x - \sin x$ όταν $|x|$ μικρή (το $\sin x \approx x$)

Σύμφωνα με το θεώρημα Taylor: $\sin x = x - \dfrac{x^3}{6} + \varepsilon(x)$,

$$|\varepsilon(x)| \le \frac{x^5}{120}$$

Συμπέρασμα: $f(x) \approx \dfrac{x^3}{6}$

3/3/2015

"Σφάλματα στον υπολογισμό αθροισμάτων"

Θα μελετήσουμε την επιρροή σφαλμάτων στρογγύλευσης λόγω της αριθμητικής κινητής υποδιαστολής με πεπερασμένη ακρίβεια στον υπολογισμό αθροισμάτων.

Παράδειγμα

$S_n = 1 + \sum_{1}^{n} \frac{1}{k^2+k}$, $n \in \mathbb{N}$ έχουμε $\frac{1}{k^2+k} = \frac{1}{k} - \frac{1}{k+1}$ οπότε:

$$S_n = 1 + \left(1 - \frac{1}{2}\right) + \left(\frac{1}{2} - \frac{1}{3}\right) + \dots + \left(\frac{1}{n} - \frac{1}{n+1}\right) = 2 - \frac{1}{n+1}$$

$n = 9999$ $\quad ((S_n)_{n\in\mathbb{N}} \uparrow \quad S_n \to 2)$

$S_{9999} = 1{,}9999$

$\beta = 10 \quad t = 10$

- Αθροίζουμε απ' τον μεγαλύτερο προς τον μικρότερο άρα:

$S_0 = 1 \quad S_k = S_{k-1} + \frac{1}{k(k+1)}$, $k = 1, 2, \dots, n$

αποτέλεσμα: $\tilde{S}_{9999} = 1.99989972$

- Αθροίζουμε απ' τον μικρότερο προς τον μεγαλύτερο.

$T_0 = 1 \quad T_k = T_{k-1} + \frac{1}{(n-k)(n-k+1)}$, $k = 1, 2, \dots, n-1$

$T_n = T_{n-1} + 1$ $\qquad (T_n = S_n)$

αποτέλεσμα: $\tilde{T}_{9999} = 1.999900000$

Γιατί στη δεύτερη περίπτωση παίρνουμε καλύτερα αποτελέσματα;

Παρατήρηση: Έστω $\lambda, \mu \in \mathbb{R}$ και $\varepsilon_1, \varepsilon_2 \in [-u, u]$

Τότε $\exists\ \varepsilon_3 \in [-u, u]$: $\lambda\varepsilon_1 + \mu\varepsilon_2 = (|\lambda| + |\mu|)\varepsilon_3$

$\varepsilon_3 = \frac{\lambda\varepsilon_1 + \mu\varepsilon_2}{|\lambda| + |\mu|}$ έχουμε: $|\varepsilon_3| \le \frac{|\lambda|\overbrace{|\varepsilon_1|}^{\le u} + |\mu|\underbrace{|\varepsilon_2|}_{\le u}}{|\lambda| + |\mu|} = \frac{\cancel{|\lambda| + |\mu|}}{\cancel{|\lambda| + |\mu|}}\, u = u$

015 Πρόβλημα

Έστω $a_1, \dots, a_N \in M$. Θέλουμε να υπολογίσουμε το άθροισμα $S_N = \sum_{i=1}^{N} a_i$

Θεωρούμε τον αλγόριθμο $S_1 = a_1$, $S_k = S_{k-1} + a_k$ $k = 2, \dots, N$

Λαμβάνουμε τις προσεγγίσεις:

$\tilde{S}_1 = a_1$, $\tilde{S}_k = fl(\tilde{S}_{k-1} - a_k)$ $k = 2, \dots, N$

Έχουμε: $\tilde{S}_2 = fl(\tilde{S}_1 + a_2) = fl S_2 = S_2(1+\delta) = S_2 + S_2\delta = S_2 + |S_2|\varepsilon$.

με $|\delta|, |\varepsilon_2| \le u$

Παρόμοια: $\tilde{S}_3 = fl(\tilde{S}_2 + a_3) = (\tilde{S}_2 + a_3)(1+\delta') = (S_2 + |S_2|\varepsilon_2 + a_3)(1+\delta') =$

$(S_3 + |S_2|\varepsilon_2)(1+\delta') \approx S_3 + |S_2|\varepsilon_2 + S_3\delta' =$

παραλείπω το $|S_2|\varepsilon_2\delta'$, διότι είναι της τάξης u^2

$= S_3 + (|S_2| + |S_3|)\varepsilon_3$ με $|\varepsilon_3|, |\delta'| \le u$

Συνεχίζοντας παίρνουμε

$\tilde{S}_N \approx S_N + (|S_2| + |S_3| + \dots + |S_N|)\varepsilon_N$ με $|\varepsilon_N| \le u$ με σφάλμα τάξης $O(u^2)$

Επομένως $\left|\dfrac{\tilde{S}_N - S_N}{S_N}\right| \approx \dfrac{|S_2| + |S_3| + \dots + |S_N|}{|S_N|} |\varepsilon_N| \le u$

Θέτουμε: $\gamma_N = |S_2| + |S_3| + \dots + |S_N|$ και τότε το $P_N = \dfrac{\gamma_N}{|S_N|}$

είναι ο συντελεστής μετάδοσης του σχετικού σφάλματος στρογγύλευσης στον αλγόριθμό μας

Όταν ο P_N είναι μεγάλος, τότε ο αλγόριθμος είναι ασταθής.

Αν κάποιο από τα ενδιάμεσα αθροίσματα S_N έχει πολύ μεγαλύτερη τιμή από την απόλυτη τιμή του τελικού S_N αθροίσματος τότε το P_N είναι μεγάλο.

Ειδική περίπτωση: $a_i > 0$, $i = 1, \dots, N$

Τότε $\gamma_N = S_2 + S_3 + \dots + S_N = (N-1)a_1 + (N-1)a_2 + (N-2)a_3 + \dots + a_N$

Το γ_N ελαχιστοποιείται αν $a_1, a_2 \le a_3 \le a_4 \le \dots \le a_N$ και

μεγιστοποιείται αν $a_1, a_2 \ge a_3 \ge a_4 \ge \dots \ge a_N$

Παράδειγμα

Προσέγγιση του e^{-x} για $x >> 1$.

$$S_N(x) = 1 - x + \frac{x^2}{2!} - \frac{x^3}{3!} + \dots + (-1)^{N-1} \frac{x^{N-1}}{(N-1)!}$$

Για αρκετά μεγάλο N έχουμε $S_N(x) \approx e^{-x}$

Για $x=100$ έχουμε $e^{-100} \approx 0$, ενώ

$S_1 = 1$, $S_2 = -99$, $S_3 = 4901$, $S_4 = -161766$ κλπ.

παταγωδώς αποτυχία!

Εναλλακτικά, με $e^{-x} = \frac{1}{e^x}$ μπορούμε να προσεγγίσουμε εύκολα το e^x και μετά να διαιρέσουμε.

"Ευστάθεια αλγορίθμων"

Ένας αλγόριθμος λέγεται ασταθής αν είναι ευαίσθητος σε σφάλματα στρογγύλευσης, δηλαδή αν μικρά σφάλματα που γίνονται κατά την παράσταση των αριθμών και στις πράξεις σε έναν υπολογιστή, είναι δυνατόν να επιφέρουν μεγάλες μεταβολές στο τελικό αποτέλεσμα.

Ένας αλγόριθμος λέγεται ευσταθής αν τα τελικά αποτελέσματά του δεν επηρεάζονται πολύ από τα σφάλματα στρογγύλευσης που γίνονται σε κάθε βήμα του.

1ο παράδειγμα

Προσέγγιση του e^{-x} για $x >> 1$ με $S_N(x) = \cdots$ ασταθής

$$e^{-x} = \frac{1}{e^x} \approx \frac{1}{1 + x + \frac{x^2}{2!} + \dots + \dots}$$ ευσταθής

2º παράδειγμα

$$I_n = \int_0^1 x^n e^{x-1}\,dx$$

$$\left\{\begin{array}{l} I_1 = \dfrac{1}{e} \\ \\ I_n = 1 - n I_{n-1}, \quad n = 2, \dots \end{array}\right) \text{ ασταθής}$$

ευσταθής ...

3º παράδειγμα

$$y_n = 2^n \sin \frac{\pi}{2^n} \qquad n \in \mathbb{N}$$

$$\left.\begin{array}{l} y_1 = 2 \\ y_{n+1} = 2^{n+1}\sqrt{\dfrac{1}{2}\left(1 - \sqrt{1 - (2^{-n} y_n)^2}\right)} \qquad n = 1, 2, \dots \end{array}\right) \text{ ασταθής}$$

$$\left.\begin{array}{l} y_1 = 2 \\ y_{n+1} = \sqrt{\dfrac{2}{1 + \sqrt{1 - (2^{-n} y_n)^2}}} \cdot y_n \qquad n = 1, 2, \dots \end{array}\right) \text{ ευσταθής}$$

"Κατάσταση προβλημάτων"

Λέμε ότι ένα πρόβλημα έχει καλή κατάσταση, αν μικρές μεταβολές στα δεδομένα του έχουν ως συνέπεια μικρή μεταβολή στη λύση.

Λέμε ότι ένα πρόβλημα έχει κακή κατάσταση, αν μικρές μεταβολές στα δεδομένα του μπορούν να έχουν ως συνέπεια μεγάλη μεταβολή στη λύση.

Παράδειγμα

$(x-2)^6 = 0 \iff x^* = 2.$

$(x-2)^6 = 10^{-6}$

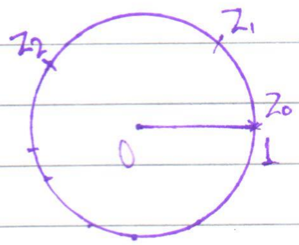
κακή κατάσταση $\rightarrow$

Θυμάμαι:

$z^n = 1$, $n \in \mathbb{N}$

έχει λύσεις:

$$z_k = e^{\frac{2i\pi k}{n}} \qquad k = 0, 1, \dots, n-1$$



Τα z_i είναι οι κορυφές του κανονικού n-γώνου που είναι εγγεγραμμένο στον μοναδιαίο κύκλο και έχει μια κορυφή στο σημείο $(1,0)$

$$(x-2)^6 = 10^{-6} \iff x_k - 2 = 10^{-1} e^{\frac{2i\pi k}{6}} \iff$$

$$|x_k - x^*| = \frac{1}{10} \left| e^{\frac{2\pi i k}{6}} \right| = 1 \iff |x_k - x^*| = \frac{1}{10}$$

καλή κατάσταση

$(x-2)^6 = 0 \iff x-2 = 0$

$x - 2 = 10^{-6} \longrightarrow \tilde{x} - 2 = 10^{-6}$.

$|\tilde{x} - 2| = 10^{-6}$

2ο ΕΠΙΛΥΣΗ ΜΗ ΓΡΑΜΜΙΚΩΝ ΕΞΙΣΩΣΕΩΝ

Έστω f μια συνάρτηση μιας πραγματικής μεταβλητής. Ζητούνται $\boxed{x \in \mathbb{R}}$ τ.ω. $f(x) = 0$, δηλαδή ρίζες της f.

"Αριθμητικές μέθοδοι"

Αριθμητικές μέθοδοι: δίνουν κατά κανόνα μια ακολουθία προσεγγίσεων $x_0, x_1, x_2, \dots$ η οποία, υπό κατάλληλες προϋποθέσεις, συγκλίνει σε μια ρίζα x^* της f

6/3/2015

Άσκηση 1.2 (βιβλίου)

Βρείτε κατάλληλους τρόπους υπολογισμού των παρακάτω παραστάσεων, έτσι ώστε να μην χάνεται ακρίβεια, όταν οι πράξεις γίνονται με αριθμητική κινητής υποδιαστολής και πεπερασμένη ακρίβεια.

α) $1-\cos x$ για $|x|$ μικρό

$1-\cos x = 2\sin^2\frac{x}{2}$

β) e^{x-y} για μεγάλα θετικά x, y

$e^{x-y} = \frac{e^x}{e^y}$

γ) $\log x - \log y$ για μεγάλα θετικά x, y

$\log x - \log y = \log\left(\frac{x}{y}\right)$

δ) $\sin(a+x) - \sin a$ για μικρό $|x|$

$\sin(a+x) - \sin a = 2\cdot\sin\frac{x}{2}\cdot\cos\left(a+\frac{x}{2}\right)$

Άσκηση 1.3 (βιβλίου)

Θεωρήστε τη δευτεροβάθμια εξίσωση $x^2-2ax+\beta=0$, όπου $a, b>0$ και $a^2 \gg b$. Δώστε έναν ευσταθή αλγόριθμο για τον υπολογισμό των ριζών της.

Λύση:

$x_1 = a+\sqrt{a^2-b}$ είναι ευσταθής

$x_2 = a-\sqrt{a^2-b}$ δεν είναι ευσταθής διότι έχω αφαίρεση σχεδόν ίσων αριθμών

$x^2-2ax+\beta = \underbrace{(x-x_1)(x-x_2)}_{\shortparallel}$

$x^2-\underbrace{(x_1+x_2)}_{=2a}x+\underbrace{x_1x_2}_{=\beta}$

Έχω $x_1x_2 = b \Rightarrow x_2 = \frac{b}{x_1}$

αυτός είναι ευσταθής

Άσκηση 1.4 (βιβλίο)

Θεωρήστε τη τριτοβάθμια εξίσωση $x^3+3px+2q=0$, όπου p, q πραγματικοί αριθμοί και $p^3+q^2>0$.

α) Αποδείξτε ότι έχει μια μόνο πραγματική ρίζα ρ που μάλιστα δίνεται από τον τύπο του Cardano. $\rho = u - v$ όπου

$$u=\left(\sqrt{p^3+q^2}-q\right)^{1/3} \quad \text{και} \quad v=\left(\sqrt{p^3+q^2}-q\right)^{1/3}$$

Λύση:

Η εξίσωση έχει ακριβώς μια πραγματική ρίζα ρ που δίνεται από τον παραπάνω τύπο.

$f: \mathbb{R}\to\mathbb{R}, \quad f(x)=x^3+3px+2q$

ύπαρξη ρίζας: Έχουμε $\lim\limits_{x\to\infty} f(x)=\infty \to$ η f παίρνει θετικές τιμές

$\lim\limits_{x\to-\infty} f(x)=-\infty \to$ η f παίρνει αρνητικές τιμές

Αφού η f είναι συνεχής και λαμβάνει τόσο θετικές όσο και αρνητικές τιμές σύμφωνα με το θεώρημα ενδιάμεσης τιμής, θα έχει τουλάχιστον μια ρίζα.

μοναδικότητα: $f'(x)=3x^2+3p=3(x^2+p)$

- Αν $p\geq 0$

Τότε η f είναι γνησίως αύξουσα. Άρα η f έχει το πολύ μια ρίζα.

- Αν $p<0$

$f'(x)=0 \Rightarrow x^2+p=0 \Rightarrow x=\pm\sqrt{-p}$

x	$-\infty$		$-\sqrt{-p}$		$+\sqrt{-p}$		$+\infty$
$f'(x)$		$+$	0	$-$	0	$+$	
$f(x)$		$\nearrow$		$\searrow$		$\nearrow$	

Ισχυρισμός: Η f δεν έχει ρίζα στο $[-\sqrt{-p}, \sqrt{-p}]$

πράγματι: $f(-\sqrt{-p}) = p\sqrt{-p}-3p\sqrt{-p}+2q = 2(q-p\sqrt{-p})$

$f(\sqrt{-p} = 2(q+p\sqrt{-p})$

οπότε: $f(-\sqrt{-p})\cdot f(\sqrt{-p}) = 4\underbrace{(q-p\sqrt{-p})(q+p\sqrt{-p})}_{q^2-p^2(-p)=p^3+q^2} = 4(p^3+q^2)$

<u>1η περίπτωση:</u> $f(-\sqrt{-p}) > 0$

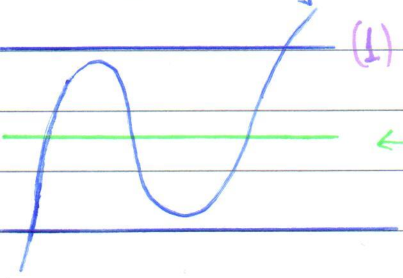
Τότε η f έχει ακριβώς μια ρίζα στο $(-\infty, -\sqrt{-p})$

Επίσης σε αυτή την περίπτωση ισχύει $f(\sqrt{-p}) > 0$, οπότε δεν έχει ρίζα στο $[\sqrt{-p}, +\infty)$

<u>2η περίπτωση:</u> $f(-\sqrt{-p}) < 0$

Τότε η f δεν έχει ρίζα στο $(-\infty, -\sqrt{-p})$

Επίσης $f(\sqrt{-p}) < 0$, οπότε η f έχει ακριβώς μια ρίζα.



← αυτή η περίπτωση αποκλείεται!

Ο άξονας x μπορεί να είναι η περίπτωση (1) ή (2).

<u>Ισχυρισμός:</u> Το p είναι ρίζα της f.

$$f(p) = f(u-v) = (u-v)^3 + 3p(u-v) + 2q = u^3 - 3u^2v + 3uv - v^3 + 3p(u-v) + 2q$$

$$\underbrace{(u^3 - v^3)}_{=2q} - 3uv(u-v) + 3p(u-v) + \cancel{2q} = -3uv(u-v) + 3p(u-v) =$$

$$= 3(\underbrace{p - uv}_{=0})(u-v) = 0$$

$$uv = \left[(\sqrt{p^3+q^2} - q)(\sqrt{p^3+q^2} - q)\right]^{1/3} = (p^3 + \cancel{q^2} - \cancel{q^2})^{1/3} = p$$

13/3/2015

β) Αν $p^3 >> q^2$ τότε ο τύπος αυτός έχει προβλήματα ευστάθειας.

<u>Λύση:</u>

Τότε $u \approx (\sqrt{p^2})^{1/3} = \sqrt{p}$

και $v \approx (\sqrt{p^2})^{1/3} = \sqrt{p}$

οπότε έχουμε αφαίρεση σχεδόν ίσων αριθμών.

γ) Να χρησιμοποιήσετε τον τύπο $u^3 - v^3 = (u-v)(u^2 + uv + v^2)$ για να βρείτε έναν ευσταθή τρόπο.

Λύση:

$$p = u - v = \frac{u^3 - v^3}{u^2 + uv + v^2} \Rightarrow p = \frac{-2q}{u^2 + \underbrace{p}_{\geq 0} + v^2}$$ ευσταθής τρόπος

Άσκηση 1.7 (βιβλίου) ένα μέρος της

α) $n \geq 3$ $\quad y_n = n \cdot \sin\frac{\pi}{n}$ $\quad \Psi_n = n \cdot \tan\frac{\pi}{n}$

Τότε: $y_n = \pi - \frac{\pi^3}{6n^2} + O\left(\frac{1}{n^4}\right)$ $\quad \Psi_n = \pi + \frac{\pi^3}{3n^2} + O\left(\frac{1}{n^4}\right)$

Θέλω να απαλοίψω το n^2 και να βρω n^4.

Να αποδείξετε ότι: $Z_n = \frac{1}{3}(2y_n + \Psi_n) \longrightarrow Z_n = \pi + O\left(\frac{1}{n^4}\right)$

προέκταση κατά Richardson (extrapolation)

Λύση:

Έχουμε $2y_n = 2\pi - \frac{2\pi^3}{6n^2} + O\left(\frac{1}{n^4}\right) = 2\pi - \frac{\pi^3}{3n^2} + O\left(\frac{1}{n^4}\right)$

$\Psi_n = \pi + \frac{\pi^3}{3n^2} + O\left(\frac{1}{n^4}\right)$

Άρα $2y_n + \Psi_n = 3\pi + O\left(\frac{1}{n^4}\right) \Rightarrow \frac{2y_n + \Psi_n}{3} = \pi + O\left(\frac{1}{n^4}\right) \Rightarrow$

$Z_n = \pi + O\left(\frac{1}{n^4}\right)$

Άσκηση 1.12 (βιβλίου) SOS

Θεωρούμε $y_n = \int_0^1 \frac{x^n}{x+\alpha}\,dx$, $n=0,1,\ldots$

α) Αν $\alpha>0$ νδο $(y_n)_{n\in\mathbb{N}}$ είναι φθίνουσα και μηδενική

Λύση

$x^{n+1} < x^n \Rightarrow \quad \forall x\in(0,1)$

$$\frac{x^{n+1}}{x+\alpha} < \frac{x^n}{x+\alpha} \Rightarrow \underbrace{\int_0^1 \frac{x^{n+1}}{x+\alpha}\,dx}_{y_{n+1}} < \underbrace{\int_0^1 \frac{x^n}{x+\alpha}\,dx}_{y_n}$$

Άρα αφού $y_{n+1} < y_n$ η $(y_n)_{n\in\mathbb{N}}$ είναι (γνήσια) φθίνουσα.

$$0 \le \frac{x^n}{x+\alpha} \le \frac{x^n}{\alpha} \Rightarrow 0 \le \int_0^1 \frac{x^n}{x+\alpha}\,dx \le \int_0^1 \frac{x^n}{\alpha}\,dx \Rightarrow$$

$$0 \le y_n \le \frac{1}{\alpha}\int_0^1 x^n dx \Rightarrow 0 \le y_n \le \frac{1}{\alpha}\cdot\frac{1}{n+1}$$

Αν $\frac{1}{\alpha}\cdot\frac{1}{n+1} \to 0$ τότε το $n\to\infty$

Άρα $\lim_{n\to\infty} y_n = 0$

β) Αν $\alpha \gg 1$ νδο $y_n = \sum_{k=0}^{n} (-1)^k \binom{n}{k} \alpha^k \frac{(1+\alpha)^{n-k} - \alpha^{n-k}}{n-k} + (-\alpha)^n \log\frac{1+\alpha}{\alpha}$

Ν.δ.ο. ο τύπος έχει προβλήματα ευστάθειας.

εκτός άσκησης:

$$y_n = \int_0^1 \frac{x^n}{x+\alpha}\,dx \underset{y=x+\alpha}{=} \int_\alpha^{1+\alpha} \frac{(y-\alpha)^n}{y}\,dy = \ldots$$

• y_n μικρός (θετικός) αριθμός

• κάποιοι από τους όρους $\binom{n}{k} a^k \frac{(1+a)^{n-k} - a^{n-k}}{n-k}$ έχουν μεγάλη απόλυτη τιμή. Επομένως, κάποια από τα ενδιάμεσα αθροίσματα έχουν μεγάλη απόλυτη τιμή

Συμπέρασμα: Ο τρόπος αυτός είναι ασταθής $\left(\gamma_n = \frac{|S_2|+|S_3|+\dots+|S_N|}{|S_N|} \right)$

γ) $a>0$ και $y_n = \int_0^1 \frac{x^n}{x+a}\,dx$, $n=0,1,2$

$$y_0 = \int_0^1 \frac{x^0}{x+a}\,dx = \int_0^1 \frac{1}{x+a}\,dx = \log(x+a)\Big|_{x=0}^{x=1} =$$

$$\log(1-a) - \log a = \log \frac{1+a}{a}$$

$$y_n = \int_0^1 \frac{x^n}{x+a}\,dx = \int_0^1 \frac{x^n + ax^{n-1} - ax^{n-1}}{x+a}\,dx =$$

$$\int_0^1 \underbrace{\frac{x^n + ax^{n-1}}{x+a}}_{= x^{n-1}}\,dx - \int_0^1 \frac{a x^{n-1}}{x+a}\,dx = \frac{1}{n} - a y_{n-1}$$

$$\begin{cases} y_0 = \log \frac{a+1}{a} \\ y_n = \frac{1}{n} - a y_{n-1}, \quad n = 1, 2, \dots \end{cases}$$

20/3/2015

Άσκηση 1.12 (βιβλίου)

Ευστάθεια;

Έστω $\tilde{y}_0$ μια προσέγγιση του y_0

Θεωρούμε τα $(\tilde{y}_n)_{n\in\mathbb{N}}$,

$$\tilde{y}_n = \frac{1}{n} - a\tilde{y}_{n-1} \quad , n=1,2,\dots$$

Τότε $y_n - \tilde{y}_n = -a(y_{n-1} - \tilde{y}_{n-1})$ οπότε επαγωγικά έχουμε

$y_n - \tilde{y}_n = (-a)^n(y_0 - \tilde{y}_0) \Rightarrow |y_n - \tilde{y}_n| = a^n |y_0 - \tilde{y}_0|$

πολύ μεγάλος αριθμός

Οπότε ο αλγόριθμος είναι ευσταθής.

δ) Ευσταθής αλγόριθμος για τον υπολογισμό του y_{10}.

$$y_n = \frac{1}{n} - a y_{n-1} \Rightarrow \quad y_{n-1} = \frac{1}{a}\left(\frac{1}{n} - y_n\right)$$

Αρχίζουμε με μια προσέγγιση π.χ. $\tilde{y}_{20}$ του y_{20} και υπολογίζουμε τα $\tilde{y}_{19}, \dots, \tilde{y}_{10}$ αναδρομικά από τον τύπο.

$$\tilde{y}_{n-1} = \frac{1}{a}\left(\frac{1}{n} - \tilde{y}_n\right) \quad , \quad n=20,19,\dots,11.$$

Ευστάθεια

$$y_{n-1} - \tilde{y}_{n-1} = -\frac{1}{a}(y_n - \tilde{y}_n) \Rightarrow \dots - y_{10} - \tilde{y}_{10} = \frac{1}{(-a)^{10}}(y_{20} - \tilde{y}_{20})$$

Άρα $|y_{10} - \tilde{y}_{10}| = \frac{1}{a^{10}}|y_{20} - \tilde{y}_{20}|$ πολύ ευσταθής

Ξέρουμε ότι $0 < y_n \le \frac{1}{a} \cdot \frac{1}{n+1}$, οπότε $0 < y_{20} \le \frac{1}{21a}$

Προσεγγίζοντας το y_{20} με $\tilde{y}_{20} = 0$ έχουμε

$$|y_{20} - \tilde{y}_{20}| \le \frac{1}{21a}$$

Άσκηση 1.13 (βιβλίου)

$$\left.\begin{array}{l} x + y = 1 \\ x + (1-a)y = 0 \end{array}\right\}$$ $a \in \mathbb{R}$ κατάσταση; Για $a=0$ δεν έχει λύση. Υποθέτουμε $a \neq 0$.

Αφαιρώ κατά μέλη $$\left.\begin{array}{l} \tilde{x} + y = 1 + \varepsilon_1 \\ \tilde{x} + (1-a)y = \varepsilon_2 \end{array}\right\}$$

Θέτουμε $u = \tilde{x} - x$ και $v = \tilde{y} - y$ και έχουμε $$\left.\begin{array}{l} u + v = \varepsilon_1 \\ u + (1-a)v = \varepsilon_2 \end{array}\right\} \Rightarrow$$

$$\left.\begin{array}{l} u = \varepsilon_1 - \dfrac{\varepsilon_1 - a}{a} \\ v = \dfrac{\varepsilon_1 - \varepsilon_2}{a} \end{array}\right\}$$

Για $|a|$ μικρή, το σύστημα έχει κακή κατάσταση, ενώ για $|a|$ μεγάλη το σύστημα έχει καλή κατάσταση.