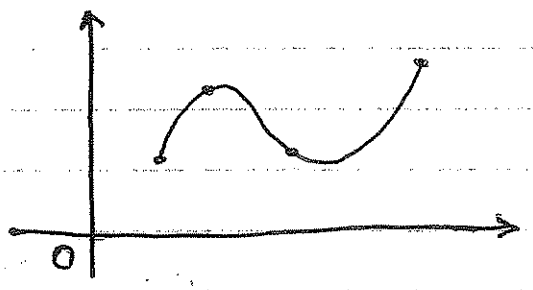
Εισαγωγή στην Αριθμητική Ανάλυση

1. Αριθμητική κινητής υποδιαστολής
   Σφάλματα στρογγύλευσης

2. Μη γραμμικές εξισώσεις
   $f: \mathbb{R} \to \mathbb{R}$
   Να βρούμε ρίζες της f, δηλαδή x τέτοιο ώστε $f(x) = 0$

3. Γραμμικά συστήματα
   $A \in \mathbb{R}^{n,n}$, $b \in \mathbb{R}^n$
   Ζητείται $x \in \mathbb{R}^n$ τέτοιο ώστε $Ax = b$

4. Παρεμβολή



5. Αριθμητική Ολοκλήρωση

$$\int_a^b f(x)dx = F(b) - F(a)$$

$$F' = f$$

$$\int_a^b f(x)dx \approx \sum_{i=1}^{n} w_i f(x_i)$$

1. Αριθμητική κινητής υποδιαστολής
Σφάλματα στρογγύλευσης

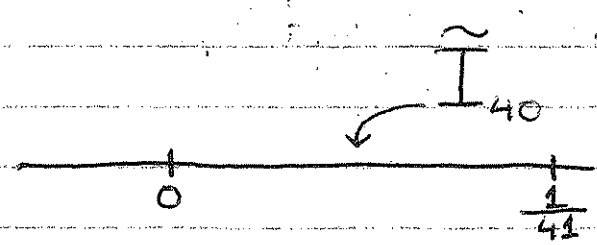
★ **1ο Παράδειγμα**

$$I_n = \int_0^1 x^n e^{x-1}\,dx, \quad n \in \mathbb{N}$$

- $I_n > 0$
- $I_{n+1} < I_n$, άρα η ακολουθία είναι γνήσια φθίνουσα
- $I_n \to 0$, $n \to +\infty$, άρα μηδενική ακολουθία
- $0 < I_n < \frac{1}{n+1}$

$$\begin{cases} I_1 = \frac{1}{e} \\ I_n = 1 - n I_{n-1}, \quad n \geq 2 \end{cases}$$ (ασταθής)

$$\begin{cases} I_{n-1} = \frac{1 - I_n}{n}, \quad n = 40, 39, \dots \\ \tilde{I}_{40} = 0 \text{ ή } \frac{1}{2 \cdot 41} \end{cases}$$ (ευσταθής)



★ 2ο Παράδειγμα

Προσέγγιση του π με τη μέθοδο του Αρχιμήδη

$$\begin{cases} y_n = 2^n \sin\frac{\pi}{2^n}, & n \in \mathbb{N} \\ y_1 = 2 \end{cases}$$

$(y_n)_{n\in\mathbb{N}}$ γνήσια αύξουσα

$y_n \longrightarrow \pi$, $n \to +\infty$

$$\begin{cases} y_{n+1} = 2^{n+1}\sqrt{\frac{1}{2}\left(1-\sqrt{1-(2^{-n}y_n)^2}\right)}, & n = 1,2,\ldots \\ y_1 = 2 \end{cases}$$

(αναδρομής)

$$\begin{cases} y_{n+1} = \sqrt{\frac{2}{1+\sqrt{1-(2^n y_n)^2}}}\, y_n, & n = 1,2,\ldots \\ y_1 = 2 \end{cases}$$

(ευσταθής)

Ποιότητα αριθμητικής μεθόδου

1) Απαιτούμενη μνήμη

2) Απαιτούμενο κόστος

3) Ακρίβεια

(4)

★ Ανάλυση των δύο προηγούμενων παραδειγμάτων

1ο Παράδειγμα

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

• $I_n > 0$, προφανές

• $x^{n+1} < x^n$ στο $(0,1) \Rightarrow \int_0^1 x^{n+1} e^{x-1} dx < \int_0^1 x^n e^{x-1} dx$

δηλαδή $I_{n+1} < I_n$

• $x - 1 \leq 0 \Rightarrow e^{x-1} \leq 1 \Rightarrow I_n = \int_0^1 x^n e^{x-1} dx < \int_0^1 x^n \cdot 1 \, dx = \frac{1}{n+1}$

$\Rightarrow 0 < I_n < \frac{1}{n+1}$

• $I_n = \int_0^1 x^n e^{x-1} dx = \int_0^1 x^n (e^{x-1})' dx =$

$= x^n e^{x-1} \Big|_{x=0}^{x=1} - \int_0^1 (x^n)' e^{x-1} dx =$

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

• $n = 1 \Rightarrow I_1 = 1 - \int_0^1 e^{x-1} dx =$

$= 1 - e^{x-1} \Big|_{x=0}^{x=1} = 1 - (e^0 - e^{-1}) =$

$= \frac{1}{e}$

$\tilde{I}_1$

$\tilde{I}_n = 1 - n\tilde{I}_{n-1}$, $n \geq 2$

Θέτουμε $\varepsilon_n := I_n - \tilde{I}_n$

Έχουμε:

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

$$\Rightarrow \boxed{\varepsilon_n = -n\varepsilon_{n-1}}$$

$\varepsilon_2 = -2\varepsilon_1$

$\varepsilon_3 = (-3)\varepsilon_2 = (-3)(-2)\varepsilon_1 = (-1)^2 3!\,\varepsilon_1$

...

$$\boxed{\varepsilon_n = (-1)^{n-1} n!\,\varepsilon_1}$$

Επαγωγή: $n = 2$

$n \to n+1$: $\varepsilon_{n+1} = -(n+1)\varepsilon_n = -(n+1)(-1)^{n-1} n!\,\varepsilon_1 =$

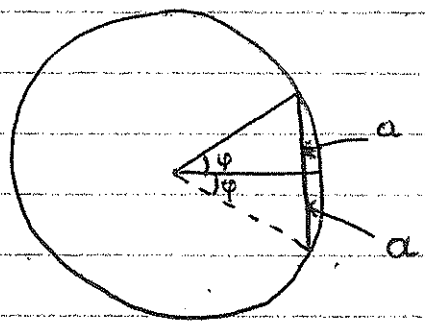
$$= (-1)^n (n+1)!\,\varepsilon$$

2ο Παράδειγμα (υπολογισμός του π με την μέθοδο του Αρχιμήδη)

• $y_n = 2^n \sin\frac{\pi}{2^n}$, $n = 1, 2, \ldots$

• $y_1 = 2\sin\frac{\pi}{2} = 2$

• Για $n \geq 2$:



$\varphi = \frac{\pi}{2^n}$, $a = \sin\frac{\pi}{2^n} \iff 2\varphi = \frac{2\pi}{2^n}$, $2a = 2\sin\frac{\pi}{2^n}$

2a: πλευρά του κανονικού πολυγώνου με 2^n πλευρές, εγγεγραμμένη στον μοναδιαίο κύκλο.

Περίμετρος: $2^n(2a) = 2^{n+1}\sin\frac{\pi}{2^n} = 2y_n$

• Η ακολουθία $(y_n)_{n\in\mathbb{N}}$ είναι γνησίως αύξουσα και συγκλίνει στο π:

$y_n = 2^n \sin\frac{\pi}{2^n}$, $\sin(2x) = 2\sin x\cos x$

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

$0 < \cos\frac{\pi}{2^{n+1}} < 1 \Rightarrow y_n < y_{n+1}$

• Γνωρίζουμε ότι:

$$\lim_{x\to 0}\frac{\sin x}{x}=1 \Longrightarrow \lim_{x\to 0}\frac{\sin(ax)}{ax}=1,\ a\in\mathbb{R},\ a\neq 0$$

$$\Longrightarrow \lim_{x\to 0}\frac{\sin(ax)}{x}=a$$

Άρα, για $y_n=2^n\sin\frac{\pi}{2^n}=\frac{\sin\frac{\pi}{2^n}}{\frac{1}{2^n}}$, βάσει της παραπάνω σχέσης, όπου $x=\frac{1}{2^n}$ και $a=\pi$, έχουμε:

$$\lim_{n\to+\infty} y_n=\pi$$

• Γνωρίζουμε ότι:

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

$$y_{n+1}=2^{n+1}\sin\frac{\pi}{2^{n+1}} \Leftrightarrow$$

$$y_{n+1}=2^{n+1}\sqrt{\frac{1-\cos\frac{\pi}{2^n}}{2}} \Longleftrightarrow$$

$$y_{n+1}=2^{n+1}\sqrt{\frac{1}{2}\left(1-\sqrt{1-\sin^2\frac{\pi}{2^n}}\right)} \Longleftrightarrow$$

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

(8)

Άρα, καταλήγουμε:

$$\begin{cases} y_1 = 2 \\ y_{n+1} = 2^{n+1}\sqrt{\frac{1}{2}\left(1-\sqrt{1-(2^{-n}y_n)^2}\right)}, & n = 1, 2, \dots \end{cases}$$

Ο αλγόριθμος είναι ασταθής λόγω αφαίρεσης σχεδόν ίσων αριθμών.

Γνωρίζουμε ότι:

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

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

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

$$y_{n+1} = 2^{n+1}\sqrt{\frac{1}{2\left(1+\sqrt{1-(2^{-n}y_n)^2}\right)}}\, 2^{-n} y_n \iff$$

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

Άρα, καταλήγουμε:

$$\begin{cases} y_1 = 2 \\ y_{n+1} = \sqrt{\frac{2}{1+\sqrt{1-(2^{-n}y_n)^2}}}\, y_n, & n = 1, 2, \dots \end{cases}$$

Ο αλγόριθμος είναι ευσταθής.

Με αυτόν τον αλγόριθμο αποφεύγουμε την αφαίρεση σχεδόν ίσων αριθμών.

Παράσταση αριθμών ως προς οποιαδήποτε βάση

• Δεκαδικό σύστημα

Βάση: 10 , Ψηφία: 0, 1, 2, ..., 9

Παράδειγμα: $(3.14159)_{10} = 3\cdot 10^{0} + 1\cdot 10^{-1} + 4\cdot 10^{-2} + 1\cdot 10^{-3} + 5\cdot 10^{-4} + 9\cdot 10^{-5}$

Γενικά: $\pm(a_N a_{N-1} \dots a_1 a_0 a_{-1} a_{-2} \dots)_{10} =$

$= \pm(a_N 10^{N} + a_{N-1}\cdot 10^{N-1} + \dots + a_1\cdot 10^{1} + a_0\cdot 10^{0} + a_{-1}\cdot 10^{-1} + a_{-2}\cdot 10^{-2} + \dots)$

Ακέραιο μέρος: $a_N a_{N-1} \dots a_0$

$p(x) = a_N x^N + a_{N-1} x^{N-1} + \dots + a_1 x + a_0$

Το ακέραιο μέρος είναι η τιμή του p στο $x=10$, $p(10)$.

Κλασματικό μέρος: $a_{-1} a_{-2} \dots$

Το κλασματικό μέρος είναι η τιμή της δυναμοσειράς

$\sum_{k=1}^{+\infty} a_{-k} x^{k}$, για $x = \frac{1}{10}$

Η σειρά μπορεί να έχει άπειρο ή πεπερασμένο πλήθος όρων.

(10)

Μοναδικότητα της παράστασης;

Ισχυρισμός: $4.130 = 4.12\overline{9} = 4.12999...$

$3\cdot10^{-2} = 2\cdot10^{-2} + 9\cdot10^{-3} + 9\cdot10^{-4} + ... \iff$

$3\cdot10^{-2} = 2\cdot10^{-2} + 9\cdot(10^{-3} + 10^{-4} + ...) \iff$

$3\cdot10^{-2} = 2\cdot10^{-2} + 9\cdot10^{-3}\cdot(1 + 10^{-1} + 10^{-2} + ...)$

Γνωρίζουμε ότι:

$$\sum_{\kappa=0}^{+\infty} \omega^{\kappa} = \frac{1}{1-\omega}, \quad |\omega| < 1$$

Άρα, έχουμε:

$3\cdot10^{-2} = 2\cdot10^{-2} + 9\cdot10^{-3}\cdot\left(\frac{1}{1-10^{-1}}\right) \iff$

$3\cdot10^{-2} = 2\cdot10^{-2} + 9\cdot10^{-3}\cdot\left(\frac{1}{9}\cdot10\right) \iff$

$3\cdot10^{-2} = 2\cdot10^{-2} + 10^{-2}$

Για να έχουμε μοναδικότητα αυτό δεν το επιτρέπουμε.

Αν υποθέσουμε ότι:

$\forall \kappa_0 \in \mathbb{N}, \exists \kappa \geq \kappa_0$ τέτοιο ώστε $a_{\kappa} \neq 9$,

τότε η παράσταση είναι μοναδική.

· Σύστημα με βάση β, $\beta \geq 2$, $\beta \in \mathbb{N}$

Βάση: β , Ψηφία: $0, 1, \ldots, \beta-1$

Παράδειγμα: $(100110.11)_2 = 1\cdot 2^5 + 1\cdot 2^2 +$

$+ 1\cdot 2^1 + 1\cdot 2^{-1} + 1\cdot 2^{-2} = (38.75)_{10}$

Γενικά: $\pm(a_N a_{N-1} \ldots a_1 a_0 a_{-1} a_{-2} \ldots)_\beta =$

$= \pm(a_N \cdot \beta^N + a_{N-1}\cdot \beta^{N-1} + \ldots + a_1 \cdot \beta^1 + a_0 \cdot \beta^0 + a_{-1}\cdot \beta^{-1} + a_{-2}\beta^{-2} + \ldots)$

i) Μετατροπή από σύστημα με βάση β στο δεκαδικό

α) Ακέραιων αριθμών

Παράδειγμα: $(53473)_8 = 5\cdot 8^4 + 3\cdot 8^3 + 4\cdot 8^2 +$

$+ 7\cdot 8^1 + 3\cdot 8^0 \Leftrightarrow$

$(53473)_8 = 3 + 8(7 + 8(4 + 8(3 + 5\cdot 8)))$

Κάνουμε τις πράξεις από "μέσα" προς τα "έξω".

Σχήμα του Horner:

$p(x) = a_N x^N + a_{N-1} x^{N-1} + \ldots + a_1 x + a_0 \Leftrightarrow$

$p(x) = a_0 + x(a_1 + x(a_2 + \ldots + x(a_{N-1} + a_N \cdot x)\ldots))$

Διαδικασία υπολογισμού

$y \leftarrow a_N$
για $i = N-1, \ldots, 0$:
  $y \leftarrow a_i + x \cdot y$
Τότε $y = p(x)$

$y \leftarrow a_i + x \cdot y$, flop (FLoating point OPeration)

Το σχήμα του Horner απαιτεί N flop.

β) Κλασματικών αριθμών

Παράδειγμα: $(.11)_2 = 1 \cdot 2^{-1} + 1 \cdot 2^{-2} = (0.\dot{7}5)_{10}$

★ ii) Μετατροπή από το δεκαδικό σε ένα σύστημα με βάση θ

α) Ακεραίων αριθμών

Βασίζεται στον αλγόριθμο της διαίρεσης!

Παράδειγμα: $(369)_{10} \rightsquigarrow$ στο οκταδικό σύστημα

$(369)_{10} = (\ldots a_2 a_1 a_0)_8 \iff$

$(369)_{10} = \underbrace{a_0}_{\text{ακέραιος}} + 8(a_1 + 8(a_2 + \ldots))$

Συμπέρασμα: Το a_0 είναι το υπόλοιπο της διαίρεσης $369:8$ και το $a_1 + 8(a_2+\ldots)$ είναι το πηλίκο.

369	8
49	46
1	

Άρα, $a_0 = 1$ και $a_1 + 8(a_2 + \ldots) = 46$

46	8
6	5

Άρα, $a_1 = 6$ και $a_2 + 8(a_3 + \ldots) = 5$

Επομένως: $a_2 = 5$, $a_3 + \ldots = 0$

Συμπέρασμα: $(369)_{10} = (561)_8$

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

β) Κλασματικών αριθμών

$0 < x < 1$ στο δεκαδικό σύστημα

$$x = (.a_{-1}a_{-2}\dots)_\beta = a_{-1}\beta^{-1} + a_{-2}\beta^{-2} + \dots$$

$$\Rightarrow \beta x = a_{-1} + a_{-2}\beta^{-1} + \dots$$

Άρα, το a_{-1} είναι το ακέραιο μέρος του βx

Παράδειγμα: $x = (.372)_{10} \rightsquigarrow$ δυαδικό

$$(.372)_{10} = (.a_{-1}a_{-2}\dots)_2$$

Έχουμε:

$2x = 0.744$, άρα $a_{-1} = 0$, $\gamma_1 := 0.744$

$2\gamma_1 = 1.488$, άρα $a_{-2} = 1$, $\gamma_2 := 0.488$

$2\gamma_2 = 0.976$, άρα $a_{-3} = 0$, $\gamma_3 := 0.976$

$2\gamma_3 = 1.952$, άρα $a_{-4} = 1$, $\gamma_4 := 0.952$

...

Συμπέρασμα: $(.372)_{10} = (.0101\dots)_2$

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

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

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

$$=\frac{3}{2}\sum_{n=1}^{+\infty}\frac{1}{2^{4n}}=$$

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

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

$$=\frac{3}{2}\,\frac{\frac{1}{16}}{\frac{15}{16}}=$$

$$=\frac{1}{10}$$

Γνωρίζουμε ότι: $\sum_{n=k}^{+\infty}\omega^n=\frac{\omega^k}{1-\omega}\ ,\ |\omega|<1$

Άρα, έχουμε:

$$(0.1)_{10}=\sum_{n=1}^{+\infty}\left(\frac{1}{2^{4n}}+\frac{1}{2^{4n+1}}\right)\Leftrightarrow$$

$$(0.1)_{10}=2^{-4}+2^{-5}+2^{-8}+2^{-9}+2^{-12}+2^{-13}+\ldots\Leftrightarrow$$

$$(0.1)_{10}=(.000\underline{1100}110011\ldots)_2\Leftrightarrow$$

$$(0.1)_{10}=(.000\overline{1100})_2$$

Στο δεκαδικό σύστημα έχουμε μόνο έναν όρο, ενώ στο δυαδικό σύστημα έχουμε άπειρο πλήθος όρων.

Αριθμοί μηχανής

Έστω $x \in \mathbb{R}$, $x \neq 0$

Σε ένα σύστημα με βάση β, ο x μπορεί να γραφεί ως:

$$(*)\quad x = \pm .d_1 d_2 d_3 \ldots \cdot \beta^e \text{, με } d_1 \neq 0,$$

όπου d_i ψηφία ως προς τη βάση β και e καταλληλος ακέραιος.

Η μορφή $(*)$ λέγεται (κανονική) μορφή κινητής υποδιαστολής.

Το σύνολο των αριθμών μηχανής $M = M(\beta, t, L, U)$ χαρακτηρίζεται από τις παραμέτρους:

- β = βάση του αριθμητικού συστήματος
- t = ακρίβεια = πλήθος των ψηφίων του κλάσματος των αριθμών
- L = κάτω φράγμα του εκθέτη e.
- U = άνω φράγμα του εκθέτη e.

δηλαδή $L \leq e \leq U$
L, U ακέραιοι, συνήθως $L \approx -U$

Κάθε $x \in M$, $x \neq 0$, είναι της μορφής:

(+) $x = \pm . d_1 d_2 \dots d_t \cdot \beta^e$,

με $d_1 \neq 0$ και $L \leq e \leq U$

Το M αποτελείται από όλους τους αριθμούς της μορφής (+) και το μηδέν.

Το M είναι πεπερασμένο σύνολο.

Μέγιστο κατ' απόλυτη τιμή στοιχείο του M:

$d_i = \beta - 1$, $i = 1, \dots, t$ και $e = U$

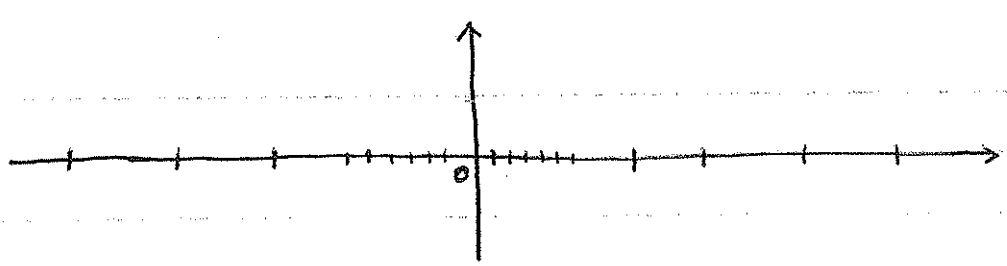
Ελάχιστο κατ' απόλυτη τιμή μη-μηδενικό στοιχείο του M:

$d_1 = 1$, $d_i = 0$, $i = 2, \dots, t$ και $e = L$

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

$\tilde{x} = x + \beta^{-t} \cdot \beta^e \iff$

$\tilde{x} - x = \beta^{e-t}$



Το M δεν είναι κλειστό ως προς τον πολλαπλασιασμό, δηλαδή:

$$x, x^* \in M \not\Rightarrow x \cdot x^* \in M$$

Παράδειγμα: $(\underbrace{100\ldots0 \cdot \beta^L}_{\in M} \times \underbrace{100\ldots0 \cdot \beta^L}_{\in M}) \notin M$

Το M δεν είναι κλειστό ως προς την πρόσθεση, δηλαδή:

$$x, x^* \in M \not\Rightarrow x + x^* \in M$$

Παράδειγμα: $\beta = 10,\ t = 5,\ 1, 10^{-5} \in M$

$$1 + 10^{-5} = 1\,00001 = 0.\underbrace{100001}_{\text{6 ψηφία}} \cdot 10^1 \notin M$$

Μας ενδιαφέρει τα M να είναι όσο πιο πυκνό και όσο πιο ευρύ γίνεται, δηλαδή μεγάλο t και μεγάλο διάστημα $[L, U]$.

Προσέγγιση πραγματικών αριθμών με αριθμούς μηχανής

i) $|x| >$ μέγιστο στοιχείο του M.

Υπερχείλιση (overflow).

Οι υπολογισμοί σταματούν.

ii) $0 < |x| < .100\ldots0 \cdot \beta^L$

Υπεκχείλιση

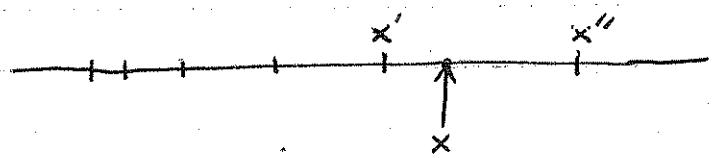
Συνήθως, ο x προσεγγίζεται με το μηδέν και συνεχίζονται οι υπολογισμοί.

iii) $.100\ldots0 \cdot \beta^L \leq |x| \leq .d_1 d_2 \ldots d_t \cdot \beta^U$

με $d_i = \beta - 1$, $i = 1, \ldots, t$

Ο x προσεγγίζεται με ένα στοιχείο $fl(x) \in M$

⊕ Συνήθως ισχύει: $|x - fl(x)| \leq |x - y|$, $\forall y \in M$



⊛ Ισχυρισμός: $\left|\frac{x - fl(x)}{x}\right| \leq \frac{1}{2}\beta^{1-t}$

ανεξάρτητο του x

α) $x \in M$, $fl(x) = x$

Η ⊛ ισχύει.

β) $x \notin M$. Τότε υπάρχουν δύο διαδοχικά στοιχεία $x', x'' \in M$ τέτοια ώστε $x' < x < x''$
Τότε ισχύει:

$$|x - fl(x)| \leq \frac{|x' - x''|}{2}$$

Γνωρίζουμε ότι:

$$\min(x-a, b-x) \leq \frac{b-a}{2}$$

Έστω χ.π.χ $x > 0$:

$$x = .d_1 d_2 \dots d_t \underbrace{d_{t+1} \dots}_{} \cdot \beta^k$$

Τότε:

$$x' = .d_1 d_2 \dots d_t \cdot \beta^k$$

$$x'' = (.d_1 d_2 \dots d_t + \beta^{-t}) \cdot \beta^k$$

$$\Rightarrow x'' - x' = \beta^{-t} \cdot \beta^k = \beta^{k-t}$$

$$x \geq .100\ldots0 \cdot \beta^k$$

$$\frac{|x - fl(x)|}{|x|} \leq \frac{\frac{1}{2}\beta^{k-t}}{.100\ldots0 \cdot \beta^k} \iff$$

$$\frac{|x - fl(x)|}{|x|} \leq \frac{1}{2} \frac{\beta^{-t}}{0.1} \iff$$

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

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

Η διαδικασία για να πετύχουμε την $\oplus$ λέγεται στρογγύλευση.

Παράδειγμα: $\beta = 10$, $t = 5$

$$x = .a_1 a_2 \ldots a_5 | a_6 a_7 \ldots \cdot 10^k$$

α) Αν $a_6 \geq 5$ τότε:

$$fl(x) = x'' = (.a_1 a_2 \ldots a_5 + 10^{-5}) \cdot 10^k$$

β) Αν $a_6 < 5$ τότε:

$$fl(x) = x' = .a_1 a_2 \ldots a_5 \cdot 10^k$$

(Αν $a_6 = 5$ και $a_i = 0$, $i \geq 7$, τότε μπορούμε ως $fl(x)$ να επιλέξουμε είτε τον x' είτε τον x'')

Αποκοπή: $t = 5$

$$x = \cdot a_1 a_2 \cdots a_5 \cdot 10^k$$

Στην περίπτωση της αποκοπής για το σχετικό σφάλμα ισχύει η εκτίμηση:

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

Γενικά:

$$\left| \frac{x - fl(x)}{x} \right| \leq u$$

$$u = \begin{cases} \rightarrow \frac{1}{2}\beta^{1-t}, & \text{για την στρογγύλωση} \\ \rightarrow \beta^{1-t}, & \text{για την αποκοπή} \end{cases}$$

u: μοναδιαίο σφάλμα στρογγύλωσης

Πράξεις

$x * y$, $* \in \{+, -, \cdot, :\}$

$z = fl(\underbrace{fl(x) * fl(y)}_{\text{υποθέσαμε ότι η πράξη αυτή γίνεται ακριβώς.}})$

Κατα κανόνα η πράξη αυτή γίνεται με διπλή ακρίβεια.

Παράδειγμα: $\beta = 10$, $t = 5$, $U = -L = 10$, στρογγύλωση.

$x = 5891.26$, $y = 0.0773414$

$x + y$

$fl(x) = .58913 \cdot 10^4$

$fl(y) = .77341 \cdot 10^{-1} = .0000077341 \cdot 10^4$

$fl(x) + fl(y) = .5891377341 \cdot 10^4$

$z = fl(fl(x) + fl(y)) = .58914 \cdot 10^4$

$z = fl(fl(x) + fl(y)) \neq x + y$

$z = fl(fl(x) + fl(y)) \neq fl(x + y)$

$z = fl(fl(x) + fl(y)) \neq fl(x) + fl(y)$

(24)

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

$a_1 = 1$, $a_2 = a_3 = 3 \cdot 10^5$, $a_1 + a_2 + a_3$, $a_1, a_2, a_3 \in M$

$fl(a_1 + a_2) = fl(1.00003) = 1$

$fl(fl(a_1 + a_2) + a_3) = fl(1.00003) = 1$

$\Rightarrow (a_1 + a_2) + a_3 = 1$

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

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

$\Rightarrow a_1 + (a_2 + a_3) = 1.0001$

Διαφορετικά αποτελέσματα!

Έχει σημασία η σειρά με την οποία γίνονται οι προσθέσεις σε ένα άθροισμα!

Για κάθε $0 < |x| < 5 \cdot 10^{-5}$ έχουμε $fl(1 + fl(x)) = 1$

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

Το $\frac{1}{2}\theta^{1-t}$ λέγεται έψιλον ή μηδέν της μηχανής

$x = .0 \ldots 0 d_t d_{t+1} \ldots$, $d_t < \frac{\theta}{2}$

$1 + x = \underbrace{1.00 \ldots 0}_{t+1 \text{ θέσεις}} d_t d_{t+1} \ldots$

$fl(1 + x) = 1.00 \ldots 0 = 1$

★ Επιρροή των σφαλμάτων στρογγύλευσης στους υπολογισμούς (όταν κάνουμε μία μόνο πράξη)

$$\left| \frac{fl(fl(x) * fl(y)) - x*y}{x*y} \right| = ;$$

Να εκτιμήσουμε το σχετικό σφάλμα στην πράξη $x*y$

$$\left| \frac{x - fl(x)}{x} \right| \le u \iff$$

$$\boxed{fl(x) = x(1+\varepsilon) \quad \text{και} \quad |\varepsilon| \le u}, \quad \text{με } \varepsilon = \varepsilon(x)$$

Ισχυρισμός: $\varepsilon_i \in [-u, u]$, $i = 1, \dots, m$

$$\Longrightarrow \text{υπάρχει } \varepsilon \in [-u,u] : \prod_{i=1}^{m} (1+\varepsilon_i) = (1+\varepsilon)^m$$

$\varphi(x) = (1+x)^m$, $x \in [-u, u]$

$$(1-u)^m \le \prod_{i=1}^{m} (1+\varepsilon_i) \le (1+u)^m \iff$$

$$\varphi(-u) \le \prod_{i=1}^{m} (1+\varepsilon_i) \le \varphi(u)$$

Σύμφωνα με το θεώρημα της ενδιάμεσης τιμής υπάρχει $\varepsilon \in [-u, u]$ τέτοιο ώστε:

$$\boxed{\prod_{i=1}^{m} (1+\varepsilon_i) = \varphi(\varepsilon) = (1+\varepsilon)^m}$$

(26)

$z = fl(fl(x) * fl(y))$

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

Πολλαπλασιασμός

$z = fl(fl(x) \cdot fl(y)) \iff$

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

$z = x \cdot y(1+\varepsilon_1)(1+\varepsilon_2)(1+\varepsilon_3) \iff$

$z = x \cdot y(1+\varepsilon)^3$

$$\frac{z - x\cdot y}{x\cdot y} = \frac{x\cdot y(1+\varepsilon)^3 - xy}{x\cdot y} = (1+\varepsilon)^3 - 1 = 3\varepsilon + 3\varepsilon^2 + \varepsilon^3$$

$$\Rightarrow \left|\frac{z-xy}{x\cdot y}\right| = |3\varepsilon + 3\varepsilon^2 + \varepsilon^3| \le 3|\varepsilon| + 3|\varepsilon|^2 + |\varepsilon|^3 \iff$$

$$\left|\frac{z-xy}{x\cdot y}\right| \le 3u + 3u^2 + u^3 \iff$$

$$\left|\frac{z-xy}{x\cdot y}\right| \le 3u + 4u^2 \iff$$

$$\left|\frac{z-xy}{x\cdot y}\right| \le 3u + O(u^2)$$

Λέμε ότι το σχετικό σφάλμα στον πολλαπλασιασμό είναι το πολύ τρεις φορές το μοναδιαίο σφάλμα στρογγύλευσης.

Διαίρεση

$$z = fl\left(\frac{fl(x)}{fl(y)}\right) = fl\left(\frac{x(1+\varepsilon_1)}{y(1+\varepsilon_2)}\right) = \frac{x(1+\varepsilon_1)}{y(1+\varepsilon_2)}(1+\varepsilon_3)$$

Γνωρίζουμε ότι:

$$\frac{1}{1+\varepsilon_2} = 1+\delta \iff$$

$$\delta = \frac{1}{1+\varepsilon_2} - 1 = \frac{-\varepsilon_2}{1+\varepsilon_2}$$

$$\Rightarrow |\delta| \le \frac{u}{1-u} = u(1+u+\dots) = u+u^2+u^3+\dots = u+O(u^2) \iff$$

$$\boxed{|\delta| \le u + O(u^2)}$$

Άρα, έχουμε:

$$z = \frac{x}{y}(1+\varepsilon_1)(1+\varepsilon_3)(1+\delta) \iff$$

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

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

$$\frac{z-\frac{x}{y}}{\frac{x}{y}} = 2\varepsilon+\delta+\varepsilon^2+2\varepsilon\delta+\delta\varepsilon^2$$

$$\Rightarrow \left|\frac{z-\frac{x}{y}}{\frac{x}{y}}\right| \le 3u + O(u^2)$$

Πρόσθεση - Αφαίρεση

$$z = fl(fl(x) + fl(y)) \iff$$

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

$$z = x(1+\varepsilon_1)(1+\varepsilon_3) + y(1+\varepsilon_2)(1+\varepsilon_3) \iff$$

$$z = x(1+\varepsilon)^2 + y(1+\delta)^2 \iff$$

$$z = x + 2\varepsilon x + y + 2\delta y + x\varepsilon^2 + y\delta^2 \iff$$

$$z = x + 2\varepsilon x + y + 2\delta y + O(u^2)$$

$$\Rightarrow z \approx (x+y) + 2(\varepsilon x + \delta y)$$

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

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

1η Περίπτωση: x, y ομόσημοι

$|x+y| = |x| + |y|$

Σφάλμα: $\left|\frac{z-(x+y)}{x+y}\right| \leq 2u$

2η Περίπτωση: x, y ετερόσημοι

Στη χειρότερη περίπτωση $\varepsilon \approx -\delta$, $|\varepsilon| \approx u$.

Οπότε θα έχουμε:

$$\left|\frac{z-(x+y)}{x+y}\right| \approx 2 \cdot \frac{|x-y|}{|x+y|} \cdot u$$

Το κλάσμα $\frac{|x-y|}{|x+y|}$ μπορεί να είναι μεγάλο.

Αν $x \approx -y$ τότε ο παράγοντας $\frac{|x-y|}{|x+y|}$ μπορεί να γίνει πολύ μεγάλος!

Τα σφάλματα στρογγύλευσης μπορούν να έχουν καταστροφική επιρροή στην αφαίρεση σχεδόν ίσων αριθμών.

Η αφαίρεση σχεδόν ίσων αριθμών πρέπει να αποφεύγεται (ή να γίνεται με διπλή ακρίβεια)

Σημείωση: Αν x, y αριθμοί μηχανής, τότε η αφαίρεσή τους γίνεται με ακρίβεια ακόμα και αν $x \approx y$

$$z = fl(fl(x) + fl(y)) = fl(x+y) = (x+y)(1+\varepsilon)$$
$$\Longrightarrow \frac{z-(x+y)}{x+y} = \varepsilon \Longrightarrow \left|\frac{z-(x+y)}{x+y}\right| \leq u$$

(30)

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

$x = .45142|708$

$y = -.45115|944$

$x+y = .26764 \cdot 10^{-3}$

$z = fl(fl(x)+fl(y)) \iff$

$z = fl(.45143 - .45116) \iff$

$z = .00027 = .27000 \cdot 10^{-3}$

λάθος!

Πώς μπορούμε να αποφύγουμε την αφαίρεση σχεδόν ίσων αριθμών;

1ο Παράδειγμα: $\beta = 10,\ t = 10$

$\sqrt{7298} = .8883|692926 \cdot 10^2$

$\sqrt{7297} = .8883|130079 \cdot 10^2$

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

↑ στον υπολογιστή

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

$\sqrt{7298} - \sqrt{7297} = \frac{1}{\sqrt{7298}+\sqrt{7297}} \iff$

$\frac{1}{\sqrt{7298}+\sqrt{7297}} = .5628468294 \cdot 10^{-2}$ (πολύ καλή προσέγγιση)

↑ στον υπολογιστή

2ο Παράδειγμα

$f(x) = x - \sin x$

Θέλουμε να υπολογίσουμε τιμές της f για $|x|$ μικρή.

Επειδή: $\lim_{x \to 0} \frac{\sin x}{x} = 1$,

θα αναγκαστούμε να αφαιρέσουμε σχεδόν ίσους αριθμούς.

Ανάπτυγμα Taylor (ως προς το μηδέν)

$\sin x = x - \frac{x^3}{6} + \varepsilon(x)$,

με $|\varepsilon(x)| \leq \frac{|x|^5}{180}$

Επομένως: $f(x) \approx \frac{x^3}{6}$ (ο υπολογισμός γίνεται χωρίς πρόβλημα!)

Σφάλματα στον υπολογισμό αθροισμάτων

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

$(S_n)_{n\in\mathbb{N}}$ γνήσια αύξουσα ακολουθία

$$S_n = 1 + \sum_{k=1}^{n} \frac{1}{k^2+k} \iff$$

$$S_n = 1 + \sum_{k=1}^{n} \frac{1}{k(k+1)} \iff$$

$$S_n = 1 + \sum_{k=1}^{n} \left(\frac{1}{k} - \frac{1}{k+1}\right) \iff$$

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

Αυτό είναι ένα τηλεσκοπικό άθροισμα, οπότε:

$$S_n = 2 - \frac{1}{n+1}$$

$$S_n \longrightarrow 2 \,,\ n \to +\infty$$

$$S_{9999} = 1.9999$$

Αναδρομικός τύπος:

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

(Αθροίζουμε τους όρους από τον μεγαλύτερο προς τον μικρότερο)

Έστω $\beta = 10$, $t = 10$ $\Rightarrow$ Αποτέλεσμα: $\tilde{S}_{9999} = \boxed{1.999}899972$

Αθροίζοντας από τον πιο μικρό προς τον πιο μεγάλο όρο του ίδιου αθροίσματος έχουμε:

$$(*)\begin{cases} T_0 = \frac{1}{n(n+1)} \\ T_k = T_{k-1} + \frac{1}{(n-k)(n-k+1)}, \quad k = 1, \dots, n-1 \\ T_n = T_{n-1} + 1 \end{cases}$$

Προφανώς $T_n = S_n$

Υλοποιώντας τον αλγόριθμο (*) με τον υπολογιστή με $\theta = 10$, $t = 10$, παίρνουμε:

$$\tilde{T}_{9999} = \boxed{1.999900\dots0}$$

Γιατί δίνει ο δεύτερος αλγόριθμος καλύτερο αποτέλεσμα από τον πρώτο;

Παρατήρηση:

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

Τότε, υπάρχει $\varepsilon_3 \in [-u, u]$ τέτοιο ώστε

$$\lambda\varepsilon_1 + \mu\varepsilon_2 = (|\lambda| + |\mu|)\varepsilon_3 \iff$$

$$\varepsilon_3 = \frac{\lambda\varepsilon_1 + \mu\varepsilon_2}{|\lambda| + |\mu|}$$

$$\Longrightarrow |\varepsilon_3| \leq \frac{|\lambda|\overbrace{|\varepsilon_1|}^{\leq u} + |\mu|\overbrace{|\varepsilon_2|}^{\leq u}}{|\lambda| + |\mu|} \leq u$$

Πρόβλημα: Έστω $a_1, a_2, \ldots, a_N \in M$

Θέλουμε να υπολογίσουμε το άθροισμα:

$$S_N = \sum_{i=1}^{N} a_i$$

Αλγόριθμος:

$S_1 = a_1$, $S_2 = a_1 + a_2, \ldots, S_K = S_{K-1} + a_K$, $K = 2, \ldots, N$

Τι μπορώ να πω για το σφάλμα, αν υλοποιήσω αυτόν τον αλγόριθμο σε έναν υπολογιστή;

Προσεγγίσεις:

$$\tilde{S}_1 = a_1$$
$$\tilde{S}_K = fl(\tilde{S}_{K-1} + a_K), \quad K = 2, \ldots, N$$

Έχουμε:

$$\tilde{S}_2 = fl(\tilde{S}_1 + a_2) = fl(a_1 + a_2)$$

Γνωρίζουμε ότι: $|\delta| \le u$, $|\varepsilon_2| \le u$, $fl(x) = x(1+\varepsilon)$, $|\varepsilon| \le u$

$$\tilde{S}_2 = \underbrace{(a_1 + a_2)}_{S_2}(1+\delta) = S_2 + S_2\delta = S_2 + |S_2|\varepsilon_2$$

$$\tilde{s}_3 = fl(\tilde{s}_2 + a_3) = (\tilde{s}_2 + a_3)(1+\delta') \quad , \quad |\delta'| \le u$$

$$\tilde{s}_3 = ((s_2 + |s_2|\varepsilon_2 + a_3))(1+\delta') \iff$$

$$\tilde{s}_3 = (s_3 + |s_2|\varepsilon_2)(1+\delta')$$

$$\Longrightarrow \tilde{s}_3 \approx s_3 + |s_2|\varepsilon_3 + s_3\delta' = s_3 + (|s_2| + |s_3|)\varepsilon_3$$

Παρόμοια, βρίσκουμε:

$$\tilde{s}_N \overset{\approx}{\uparrow} s_N + (|s_2| + |s_3| + \dots + |s_N|)\varepsilon_N$$

Παραλείπουμε όρους της τάξης του u^2

$$\frac{\tilde{s}_N - s_N}{s_N} \approx \frac{|s_2| + |s_3| + \dots + |s_N|}{s_N} \cdot \varepsilon_N$$

Σχετικό σφάλμα:

$$\frac{|\tilde{s}_N - s_N|}{|s_N|} \approx \frac{|s_2| + |s_3| + \dots + |s_N|}{|s_N|} \cdot |\varepsilon_N|$$

Έστω: $|s_2| + |s_3| + \dots + |s_N| = \gamma_N$, τότε:

$$p_N := \frac{\gamma_N}{|s_N|}$$

Συντελεστής μετάδοσης του σχετικού σφάλματος στρογγύλευσης στον αλγόριθμό μας.

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

$$|s_2| + \dots + |s_N| = (N-1)a_1 + (N-1)a_2 + (N-2)a_3 + \\ + (N-3)a_4 + \dots + 2a_{N-1} + a_N$$

Το άθροισμα γίνεται ελάχιστο αν:

$$a_1, a_2 \leq a_3 \leq a_4 \leq \dots \leq a_N$$

Το άθροισμα γίνεται μέγιστο αν:

$$a_1, a_2 \geq a_3 \geq a_4 \geq \dots \geq a_N$$ (πρέπει να το αποφύγουμε!)

Όταν το P_N είναι πολύ μεγάλο ο αλγόριθμος είναι ασταθής.

Το P_N είναι μεγάλο, αν κάποια από τα ενδιάμεσα αθροίσματα S_K έχουν απόλυτη τιμή πολύ μεγαλύτερη από $|s_N|$.

Παράδειγμα: Προσέγγιση του e^{-x} για $x >> 1$.

Γνωρίζουμε ότι: $e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \dots + \frac{x^N}{N!} + \dots$

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

$$S_N(x) \to e^{-x}, \quad N \to +\infty$$

Για αρκετά μεγάλο N το $S_N(x)$ είναι καλή προσέγγιση του e^{-x}.

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

$s_1=1,\ s_2=-99,\ s_3=4901,\ s_4 \approx -161766$ κλπ.

Ο αλγόριθμος είναι ασταθής. Παραγώδης αποτυχία!

$$e^{-x} = \frac{1}{e^x} \approx \frac{1}{1+x+\frac{x^2}{2!}+\ldots+\frac{x^{N-1}}{(N-1)!}}$$

Ο αλγόριθμος αυτός είναι ευσταθής.

Ευστάθεια αλγορίθμων

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

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

Παραδείγματα:

1) e^{-x}, $x \gg 1$, <u>ασταθής</u>

$\frac{1}{e^x}$, <u>ευσταθής</u>

2) $I_n = \int_0^1 x^n e^{-x} dx$, $n = 1, 2, \ldots$

$$\begin{cases} I_1 = \frac{1}{e} \\ I_n = 1 - (n) I_{n-1}, \quad n = 2, 3, \ldots \end{cases}$$, <u>ασταθής</u>

$$\begin{cases} I_{20} = 0 \\ I_{n-1} = \frac{1 - I_n}{n}, \quad n = 2, 3, \ldots \end{cases}$$, <u>ευσταθής</u>

(40)

3) $y_n = 2^n \sin\frac{\pi}{2^n}$, $n \in \mathbb{N}$

$$\begin{cases} y_1 = 2 \\ y_{n+1} = 2^{n+1}\sqrt{\frac{1}{2}\left(1 \ominus \sqrt{1-(2^{-n}\cdot y_n)^2}\right)}, & n = 1, 2, \dots \end{cases}$$

αστάθής

$$\begin{cases} y_1 = 2 \\ y_{n+1} = \sqrt{\frac{2}{1+\sqrt{1-(2^{-n}\cdot y_n)^2}}} \cdot y_n, & n = 1, 2, \dots \end{cases}$$

ευστάθής

Κατάσταση προβλημάτων

Λέμε ότι ένα πρόβλημα έχει καλή κατάσταση, αν μικρές μεταβολές στα δεδομένα του, έχουν ως αποτέλεσμα μικρές μεταβολές της λύσης του.

Λέμε ότι ένα πρόβλημα έχει κακή κατάσταση, αν είναι δυνατόν μικρές μεταβολές στα δεδομένα του, να οδηγούν σε μεγάλες μεταβολές της λύσης του.

Παράδειγμα

$(x-2)^6=0 \Longrightarrow$ Λύση: $x^*=2$

$(x-2)^6=10^{-6} \Rightarrow$ Λύσεις: $x_\kappa = 2+\frac{1}{10}e^{\frac{2\pi i\kappa}{6}}$,

όπου $\kappa=0,\dots,5$, Μεταβολή στα δεδομένα: 10^{-6}

Γνωρίζουμε ότι: $|e^{ix}|=1$, για $x\in\mathbb{R}$

Μεταβολή στη λύση: $|x_\kappa-2|=\frac{1}{10}=10^{-1}$

δηλαδή η μεταβολή στη λύση είναι 10^5 επί τη μεταβολή στα δεδομένα. Κακή κατάσταση!

$x-2=0 \Rightarrow$ Λύση: $x^*=2$, $x-2=10^{-6} \Rightarrow$ Λύση: $x^*=2+10^{-6}$

Η μεταβολή στη λύση είναι 10^{-6}, όσο είναι και η μεταβολή στα δεδομένα. Καλή κατάσταση!

Ασκήσεις

Άσκηση 1.2

α) $1 - \cos x = 2\sin^2\frac{x}{2}$

$|x|$ μικρή (χωρίς Taylor)

β) $e^{x-y} = \frac{e^x}{e^y}$

γ) $\log x - \log y = \log\frac{x}{y}$

δ) $\sin(a+x) - \sin(a) = 2\sin\frac{x}{2}\cos(a+\frac{x}{2})$

$|x|$ μικρή

$\sin\alpha - \sin\beta = 2\sin\left(\frac{\alpha-\beta}{2}\right)\cos\left(\frac{\alpha+\beta}{2}\right)$

Άσκηση 1.3

$x^2 - 2ax + \beta = 0$, $a, \beta > 0$, $a^2 >> \beta$

$x_1 = a + \sqrt{a^2 - \beta}$, ευσταθής

$x_2 = a - \sqrt{a^2 - \beta}$, αφαίρεση σχεδόν ίσων αριθμών.

$$x_2 = a - \sqrt{a^2 - \beta} = \frac{(a - \sqrt{a^2 - \beta})(a + \sqrt{a^2 - \beta})}{a + \sqrt{a^2 - \beta}} \iff$$

$$x_2 = \frac{a^2 - (a^2 - \beta)}{a + \sqrt{a^2 - \beta}} = \frac{\beta}{a + \sqrt{a^2 - \beta}} = \frac{\beta}{x_1} \iff$$

$$x_2 = \frac{\beta}{x_1}$$

$x^2 - 2ax + \beta = (x - x_1)(x - x_2) \iff$

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

Άσκηση 1.4

$x^3 + 3px + 2q = 0$, $p, q \in \mathbb{R}$ και $p^3 + q^2 > 0$

a) Η εξίσωση έχει ακριβώς μία πραγματική ρίζα ρ, και:

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

$$v = \left(\sqrt{p^3+q^2} + q\right)^{1/3}$$

Τύπος του Cardano

Λύση

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

Ύπαρξη ρίζας

$$\lim_{x \to +\infty} f(x) = +\infty \text{ , } \lim_{x \to -\infty} f(x) = -\infty$$

Επομένως η f παίρνει και θετικές και αρνητικές τιμές. Άρα, σύμφωνα με το θεώρημα της ενδιάμεσης τιμής, (η f είναι συνεχής) η f παίρνει και την τιμή μηδέν.

(4)

Μοναδικότητα

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

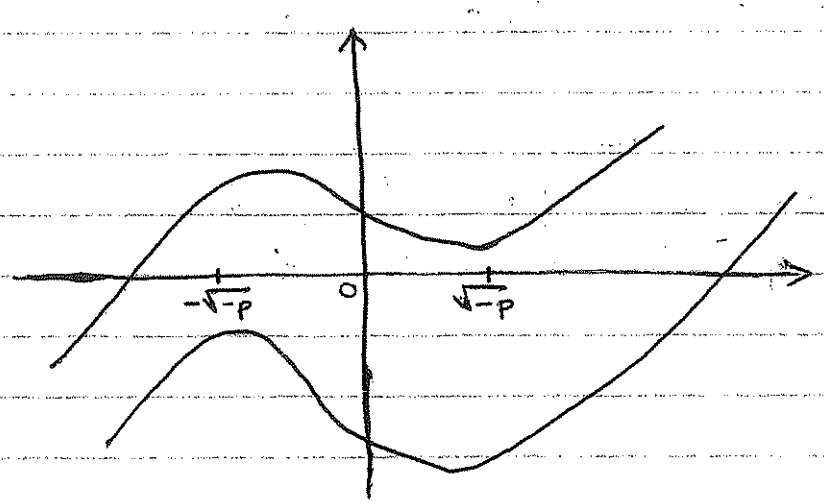
- $p \geq 0$: $f'(x) > 0, \ \forall x \in \mathbb{R}, \ x \neq 0$
  $f'(x) \geq 0, \ \forall x \in \mathbb{R}$

Η f είναι γνήσια αύξουσα, οπότε δεν μπορεί να έχει περισσότερες από μία ρίζες.

- $p < 0$: $f'(x) = 3(x^2 + p)$
  $f'(x) = 0 \iff x = \pm\sqrt{-p}$

f'	$+$	$-$	$+$
f	↗ $-\sqrt{-p}$	↘	$\sqrt{-p}$ ↗

Η f είναι γνήσια αύξουσα στο $(-\infty, -\sqrt{-p})$
Η f είναι γνήσια φθίνουσα στο $(-\sqrt{-p}, \sqrt{-p})$
Η f είναι γνήσια αύξουσα στο $(\sqrt{-p}, +\infty)$



Για να έχουμε μόνο μία ρίζα θα πρέπει οι τιμές $f(-\sqrt{-p})$ και $f(\sqrt{-p})$ να είναι ομόσημες.

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

$$\Rightarrow f(-\sqrt{-p}) \cdot f(\sqrt{-p}) = 4(q^2 + p^2) > 0$$

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

Τότε η f έχει μία ρίζα στο $(-\infty, -\sqrt{-p})$, καμία ρίζα στο $(-\sqrt{-p}, \sqrt{-p})$ και καμία ρίζα στο $(\sqrt{-p}, +\infty)$

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

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

Άσκηση 1.4

$x^3+3px+2q=0,\ p,q\in\mathbb{R},\ p^3+q^2>0$

$f(x)=x^3+3px+2q$

α) Η εξίσωση έχει ακριβώς μία πραγματική ρίζα ρ και ισχύει

$\rho=u-v$, με $u=\left(\sqrt{p^3+q^2}-q\right)^{1/3}$,

$v=\left(\sqrt{p^3+q^2}+q\right)^{1/3}$

Τύπος του Cardano

α) $f(\rho)=f(u-v)=(u-v)^3+3p(u-v)+2q \iff$

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

$f(\rho)=-2q+3(p-uv)(u-v)+2q \iff$

$f(\rho)=3(p-uv)(u-v)=0$

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

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

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

$uv=p$

β) Αν $p^3 >> q^2$, ο τύπος του Cardano έχει προβλήματα ευστάθειας.

β) $u \approx \sqrt{p} \approx v$

Άρα, έχουμε αφαίρεση σχεδόν ίσων αριθμών, οπότε έχουμε προβλήματα ευστάθειας.

γ) $p^3 >> q^2$

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

γ) $u - v = \frac{u^3 - v^3}{u^2+uv+v^2} \iff$

$u - v = \frac{-2q}{u^2+p+v^2}$, $p > 0$

Ευσταθής τρόπος.

Άσκηση 1.7

α) $n \geq 3$

$y_n = n \sin \frac{\pi}{n}$, ημιπερίμετρος του εγγεγραμμένου στον μοναδιαίο κύκλο κανονικού n-γώνου

$Y_n = n \tan \frac{\pi}{n}$, ημιπερίμετρος του περιγεγραμμένου στον μοναδιαίο κύκλο κανονικού n-γώνου.

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

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

Προέκταση κατά Richardson:

$$\left.\begin{aligned} 2y_n &= 2\pi - \frac{\pi^3}{3n^2} + O\left(\frac{1}{n^4}\right) \\ Y_n &= \pi + \frac{\pi^3}{3n^2} + O\left(\frac{1}{n^4}\right) \end{aligned}\right\} \Longrightarrow$$

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

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

Άσκηση 1.12

$a \in \mathbb{R}$, a μεγάλος αριθμός

$y_n = \int_0^1 \frac{x^n}{x+a}\,dx$, $n = 0, 1, 2, \ldots$

α) $a > 0 \Rightarrow (y_n)$ γνήσια φθίνουσα και μηδενική (δηλαδή $\lim\limits_{n\to+\infty} y_n = 0$)

α) $y_{n+1} = \int_0^1 \frac{x^{n+1}}{x+a}\,dx < \int_0^1 \frac{x^n}{x+a}\,dx = y_n$

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

Άρα, $(y_n)_{n\in\mathbb{N}_0}$ γνήσια φθίνουσα.

$0 \le y_n = \int_0^1 \frac{x^n}{x+a}\,dx \le \int_0^1 \frac{x^n}{a}\,dx = \frac{1}{a}\,\frac{1}{n+1}$

$\Rightarrow 0 \le y_n \le \frac{1}{a}\,\frac{1}{n+1}$

$\lim\limits_{n\to+\infty} \frac{1}{a}\,\frac{1}{n+1} = 0$

Άρα, $\lim\limits_{n\to+\infty} y_n = 0$

β) ...

(10)

γ) $y_{n-1} \longrightarrow y_n$;

γ) $y_n = \int_0^1 \frac{x^n}{x+a}\,dx \iff$

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

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

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

$y_n = \frac{1}{n} - a y_{n-1}\,, \quad n = 1, 2, \dots$

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

$y_0 = \log(a+1) - \log a = \log\frac{a+1}{a}$

Άρα έχουμε:

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

$a >> 1$, Ευστάθεια; $\tilde{y}_0$ $\quad \tilde{y}_n = \frac{1}{n} - a\tilde{y}_{n-1}\,, \quad n = 1, 2, \dots$

Αφαιρώντας κατά μέλη παίρνουμε:

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

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

μεγαλώνει πολύ γρήγορα, ασταθία!

$$\text{β)}\ y_n = \sum_{k=0}^{n-1} (-1)^k \binom{n}{k} a^k \frac{(1+a)^{n-k} - a^{n-k}}{n-k} + (-a)^n \log\frac{a+1}{a}$$

με $n \geq 1$, Ευστάθεια:

$$y_n = \int_0^1 \frac{x^n}{x+a}\,dx = \int_a^{a+1} \frac{(y-a)^n}{y}\,dy \iff$$

$$y_n = \int_a^{a+1} \frac{1}{y} \sum_{k=0}^{n} \binom{n}{k} (-1)^k a^k y^{n-k}\,dy \iff$$

$$y_n = \sum_{k=0}^{n-1} (-1)^k \binom{n}{k} a^k \int_a^{a+1} y^{n-k-1}\,dy + (-1)^n a^n \int_a^{a+1} \frac{1}{y}\,dy \iff$$

$$y_n = \sum_{k=0}^{n-1} (-1)^k \binom{n}{k} a^k \frac{(1+a)^{n-k} - a^{n-k}}{n-k} + (-a)^n \log\frac{a+1}{a}$$

Σημείωση: $(y-a)^n = (-a+y)^n = \sum_{k=0}^{n} \binom{n}{k} (-a)^k y^{n-k}$

Για κατάλληλο k οι όροι $\binom{n}{k} a^k \frac{(1+a)^{n-k} - a^{n-k}}{n-k}$ είναι μεγάλοι, οπότε αναγκαστικά κάποια ενδιάμεσα αθροίσματα είναι μεγάλα (σε απόλυτη τιμή) ενώ το y_n είναι μικρό.

Ο αλγόριθμος είναι αστάθης!

(12)

★★ δ) Ζητείται ευσταθής τρόπος για να υπολογίσουμε το y_{10}

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

Έστω ότι έχουμε μια προσέγγιση $\tilde{y}_n$ του y_n, τότε:

$\tilde{y}_{n-1} = \frac{1}{a}\left(\frac{1}{n} - \tilde{y}_n\right)$

Οπότε:

$y_{n-1} - \tilde{y}_{n-1} = \left(-\frac{1}{a}\right)(y_n - \tilde{y}_n)$

↑ μικρός αριθμός

Γνωρίζουμε ότι $0 \leq y_{20} \leq \frac{1}{21a}$

Με την προσέγγιση $\tilde{y}_{20} = 0$, το σφάλμα είναι το πολύ $\frac{1}{21a}$

Εφαρμόζοντας τον αλγόριθμο (*) βρίσκουμε τις προσεγγίσεις $\tilde{y}_{19}, \tilde{y}_{18}, \ldots, \tilde{y}_{10}$

Άσκηση 1.13

$$\left.\begin{aligned} x + y &= 1 \\ x + (1-a)y &= 0 \end{aligned}\right\}$$

$a \in \mathbb{R}$, Κατάσταση του συστήματος;

$a = 0$:

Το σύστημα δεν έχει λύση.

$a \neq 0$:

$$\left.\begin{aligned} \tilde{x} + \tilde{y} &= 1 + \varepsilon_1 \\ \tilde{x} + (1-a)\tilde{y} &= \varepsilon_2 \end{aligned}\right\}$$

Έστω $u := \tilde{x} - x$ και $v := \tilde{y} - y$

Και έχουμε:

$$\left.\begin{aligned} u + v &= \varepsilon_1 \\ u + (1-a)v &= \varepsilon_2 \end{aligned}\right\}$$

$$\Rightarrow u = \varepsilon_1 - \frac{\varepsilon_1 - \varepsilon_2}{a}, \quad v = \frac{\varepsilon_1 - \varepsilon_2}{a}$$

Για $|a|$ μικρή, η κατάσταση του συστήματος είναι κακή!

Για $|a|$ μεγάλη, η κατάσταση του συστήματος είναι καλή!