13/02/2018

Το περιεχόμενο του μαθήματος (Συνοπτικά)

1) • Αριθμητική κινητής υποδιαστολής
• Σφάλματα στρογγύλευσης
• Ευαισθησία αλγορίθμων και προβλημάτων σε σφάλματα στρογγύλευσης (ή σε διαταραχές)

Γενικά : Απλή ακρίβεια → 6 δεκαδικά ψηφία
Διπλή ακρίβεια → 13 δεκαδικά ψηφία

2) • Μη γραμμικές εξισώσεις,

Δεδομένο : $f:[a,b] \to \mathbb{R}$

Ζητούμενο : ρίζες της f.

Παρατήρηση : Για πολυωνυμικές συναρτήσεις βαθμού >4 δεν υπάρχουν τύποι εύρεσης των ριζών τους.

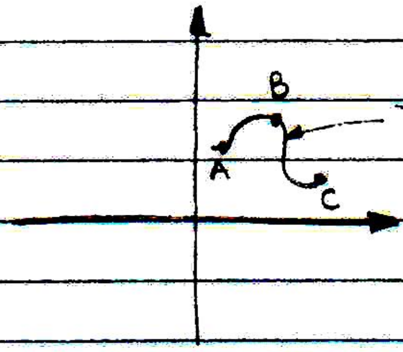
3) • Γραμμικά Συστήματα

Δεδομένα : A $n\times n$ πίνακας, αντιστρέψιμος
$b \in \mathbb{R}^n$, διάνυσμα

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

4) • Παρεμβολή :

- Τρόπος προσέγγισης συναρτήσεων.
- Στην απλούστερη μορφή της παρεμβολής ζητείται μία συνάρτηση, από ένα δεδομένο σύνολο, το γράφημα της οποίας διέρχεται από δεδομένα σημεία (x_i, y_i), $i=1,\dots,n$ πρέπει $x_i \neq x_j$ για $i \neq j$.



ζητείται συνάρτηση από ένα πεπερασμένο σύνολο συναρτήσεων που να περνάει από τα A, B, C.

5) • Αριθμητική ολοκλήρωση :

Δεδομένο : $f:[a,b] \longrightarrow \mathbb{R}$ συνεχής

Ζητούμενο : $\int_a^b f(x)dx$

→ Αν $F' = f$, τότε $\int_a^b f(x)dx = F(b) - F(a)$

τότε προσεγγίζουμε το $\int_a^b f(x)dx$ με αθροίσματα της μορφής $w_0 f(x_0) + w_1 f(x_1) + \dots + w_n f(x_n)$

με κατάλληλα x_i και w_i.

1ο Παράδειγμα :

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

$e^{x-1} \to \leq 1$ για $x \in [0,1]$

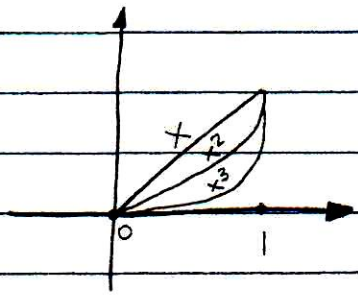
- $minVal = \frac{1}{e}$
- $maxVal = 1$

– **Πρόβλημα:** Υπολογισμός του I_n για αρκετά μεγάλο n.

Ιδιότητες:

– Ολοκληρώνω θετική συνάρτηση $\longrightarrow I_{n+1} > 0$ αφού ολοκληρώνω στο $[0,1]$ όσο αυξάνει ο εκθέτης του x ($n \longrightarrow n+1$) η τιμή του ολοκληρώματος μικραίνει.

Σχηματικά:



$x^{n+1} < x^n \longrightarrow x < 1$

για $0 < x < 1$

Επίσης :

$$0 < I_{n+1} < I_n \leq \int_0^1 x^n dx = \frac{1}{n+1}$$

$\longrightarrow$ φθίνουσα $n \longrightarrow \infty$

Γενικά: $a_n \leq b_n \leq c_n$

$a_n \downarrow 0$, $c_n \downarrow 0$, $b_n \longrightarrow 0$, $n \longrightarrow \infty$

<u>Συμπέρασμα</u>: Η $(I_n)_{n \in \mathbb{N}}$ είναι φθίνουσα μηδενική ακολουθία.
↳ έχει όριο το 0.

$$\begin{cases} I_n = 1 - n I_{n-1}, & n \geq 2 \\ I_1 = 1/e \end{cases}$$

↳ άρρητος αριθμός ⟶ πρόβλημα.
- άρα έχω αρκετό σφάλμα στρογγύλευσης το οποίο θα διογκωθεί στην συνέχεια.

1. Αριθμητική Κινητής Υποδιαστολής

– Σφάλματα στρογγύλευσης

– Υπάρχουν "καλές" και "κακές" αριθμητικές μέθοδοι.

– Η ποιότητα μιας αριθμητικής μεθόδου εξαρτάται από 3 παράγοντες:

• Απαιτούμενη μνήμη } Σχετική αξία
• Απαιτούμενος χρόνος }
• Ακρίβεια των αποτελεσμάτων. → ΑΠΑΡΑΙΤΗΤΟ

– Οι πράξεις στους υπολογιστές γίνονται με πεπερασμένη ακρίβεια, και αυτό έχει ως συνέπεια σφάλματα στρογγύλευσης.

• Ευστάθεια αλγορίθμων:

– Αλγόριθμοι ευαίσθητοι σε σφάλματα στρογγύλευσης λέγονται ασταθείς και οδηγούν σε λανθασμένα αποτελέσματα.

– Χρησιμοποιούμε ευσταθείς αλγορίθμους για να οδηγηθούμε σε ακριβή αποτελέσματα.

27/02/2018

1º Παράδειγμα:

Θέλουμε να υπολογίσουμε το $I_n = \int_0^1 x^n e^{x-1} dx$, $n \in \mathbb{N}$, για αρκετά μεγάλο n.

- **Ιδιότητες των I_n:** η $h_n(x) = x^n e^{x-1}$ είναι θετική συνάρτηση για $x \in (0,1)$ άρα και το ολοκλήρωμά της είναι θετικό.

$$I_n = \int_0^1 h_n(x) dx$$

- $I_{n+1} > 0$
- $x^{n+1} < x^n$ για $x \in (0,1)$

$$\Longrightarrow I_{n+1} < I_n$$

Άρα η ακολουθία $(I_n)_{n\in\mathbb{N}}$ είναι γνησίως φθίνουσα.

- $e^{x-1} \leq 1 \quad \forall x \in [0,1]$

$$\Rightarrow I_n \leq \int_0^1 x^n \cdot 1 dx = \frac{1}{n+1}$$

Άρα $0 < I_{n+1} < I_n < \frac{1}{n+1}$

Ιδιαίτερα: $\lim_{n\to\infty} I_n = 0$, άρα η ακολουθία $(I_n)_{n\in\mathbb{N}}$ είναι μηδενική.

(!!!)

Έχουμε,

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

$\Rightarrow$ (Ολοκλήρωση κατά παράγοντες) $\Rightarrow$

$$= [x^n e^{x-1}]_0^1 - \int_0^1 n x^{n-1} e^{x-1} dx$$

• $n=1$: $I_1 = 1 - 0 - \int_0^1 x^{1-1} e^{x-1} dx =$

$$= 1 - \int_0^1 e^{x-1} dx = 1 - [e^{x-1}]_0^1 =$$

$$= 1 - [e^0 - 1/e] = 1/e$$

• $n \geq 2$:

$$I_n = [x^n e^{x-1}]_0^1 - \int_0^1 n x^{n-1} e^{x-1} dx =$$

$$= [e^0 - 0] - n \int_0^1 x^{n-1} e^{x-1} dx =$$

$$= 1 - n I_{n-1}$$

Συμπέρασμα:

$$\text{(I)} \begin{cases} I_n = 1 - n I_{n-1}, & n = 2, 3, \ldots \\ I_1 = 1/e \end{cases}$$

Ευστάθεια της μεθόδου (I);

- Ξεκινάμε με μία προσέγγιση $\hat{I}_1$ του I_1 και <u>υποθέτουμε</u> ότι όλες οι πράξεις στην συνέχεια γίνονται ακριβώς.
Παίρνουμε έτσι τις προσεγγίσεις,

$$\hat{I}_n = 1 - n\hat{I}_{n-1} \quad , \quad n = 2, 3, \ldots$$

Αφαιρώντας κατά μέλη έχουμε,

$$\text{(1)} \quad \boxed{I_n - \hat{I}_n = -n\left(I_{n-1} - \hat{I}_{n-1}\right)}$$

Το σφάλμα σε κάθε βήμα πολλαπλασιάζεται με το n.

Επαγωγικά παίρνουμε:

$$\text{(2)} \quad I_n - \hat{I}_n = (-1)^{n-1} n! \left(I_1 - \hat{I}_1\right)$$

- Για $n=1$ η (2) γράφεται ως:

$$I_1 - \hat{I}_1 = I_1 - \hat{I}_1 \quad , \text{ σωστό}$$

- <u>$n \to n+1$</u>: Σύμφωνα με την (1) έχουμε

$$I_{n+1} - \hat{I}_{n+1} \overset{(1)}{=} -(n+1)\left(I_n - \hat{I}_n\right) =$$

$$\overset{(2)}{=} -(n+1)(-1)^{n-1} n! \left(I_1 - \hat{I}_1\right) =$$

$$= (-1)^n (n+1)! \left(I_1 - \hat{I}_1\right)$$

δηλαδή η (2) ισχύει και για $n+1$.

Από την (2) παίρνουμε

$$|I_n - \hat{I}_n| = n! \, |I_1 - \hat{I}_1| \, , \quad n \geq 1$$

→ αυξάνει "ταχύτατα" (εκθετικά) με το n.

ο αλγόριθμος (μέθοδος) (I) είναι ασταθής, ο παράγοντας "n" στον τύπο προκαλεί την αστάθεια.

2ος Αλγόριθμος

- Γράφουμε τη $I_n = 1 - nI_{n-1}$ στη μορφή

$$I_{n-1} = \frac{1 - I_n}{n} \quad (3)$$

- Ξεκινάμε με μία προσέγγιση $\hat{I}_m$ του I_m και υπολογίζουμε τα $\hat{I}_{m-1}, \hat{I}_{m-2}, \ldots, \hat{I}_l$ (υποθέτοντας ότι οι πράξεις γίνονται ακριβώς)

ως εξής,

$$(4) \quad \hat{I}_{n-1} = \frac{1 - \hat{I}_n}{n} \, , \quad n = m, \ldots, l+1$$

Είναι ευσταθής ο 2ος Αλγόριθμος?

Αφαιρώντας κατά μέλη τις (3) και (4), παίρνουμε

$$(5) \quad I_{n-1} - \hat{I}_{n-1} = -\frac{I_n - \hat{I}_n}{n}$$

- Το σφάλμα σε κάθε βήμα διαιρείται με το n.
  Επαγωγικά, η (5) δίνει

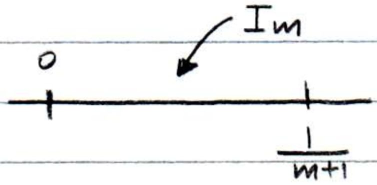
$$I_l - \tilde{I}_l = (-1)^{m-l} \frac{1}{(l+1)\cdots m} \left| I_m - \tilde{I}_m \right|$$

Το σφάλμα καταστέλλεται σε κάθε βήμα.

→ <u>Αλγόριθμος ευσταθής.</u>

- Με ποια τιμή $\tilde{I}_m$ ξεκινάμε;

Γνωρίζω ότι $I_m \in (0, 1/m+1)$



Καλύτερη επιλογή είναι το μέσο του διαστήματος άρα

$$\tilde{I}_m = \frac{1}{2(m+1)}$$

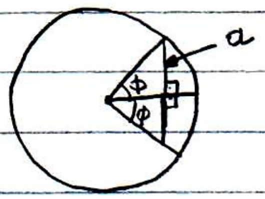
2^{o} Παράδειγμα :

Προσέγγιση του π με τη μέθοδο του Αρχιμήδη.

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

$$y_1 = 2\underbrace{\sin\left(\frac{\pi}{2}\right)}_{=1} = 2$$

- Τι είναι το y_n για $n \geq 2$?



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

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

— Άρα το $2a$ είναι η πλευρά του κανονικού πολυγώνου με 2^n πλευρές που είναι εγγεγραμμένο στο μοναδιαίο κύκλο

• Περίμετρος αυτού του πολυγώνου

$$2^n \cdot 2a = 2 \cdot 2^n a = 2 \cdot \underbrace{2^n \sin\frac{\pi}{2^n}}_{= y_n}$$

• Το y_n είναι η ημιπερίμετρος του κανονικού πολυγώνου με 2^n πλευρές, εγγεγραμμένου στον μοναδιαίο κύκλο

• Γεωμετρικά έχουμε :

$$y_n < y_{n+1} \quad \text{(γνησίως αύξουσα ακολουθία)}$$

Οπότε,

$$2y_n \to 2\pi \,, \quad n \to \infty$$

δηλαδή $\quad y_n \to \pi \,, \; n \to \infty$

Αλγεβρική (αναλυτική) απόδειξη:

$\sin(2x) = 2\sin x \cdot \cos x \,, \quad x = \frac{\pi}{2^{n+1}}$

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

<1

$$\Rightarrow y_n < y_{n+1}$$

- $|\sin x| \le |x| \le |\tan x|$ για $x \in (-\pi/2, \pi/2)$

Για $x \neq 0$ οι ανισώτητες αυτές δίνουν,

$$|\cos x| \le \left| \frac{\sin x}{x} \right| \le 1$$

$\Rightarrow \lim\limits_{x \to 0} \frac{\sin x}{x} = 1$, όμως $\forall a \in \mathbb{R}$ με $a \neq 0$ ισχύει,

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

$$y_n = 2^n \sin\frac{\pi}{2^n} = \frac{\sin\frac{\pi}{2^n}}{\frac{1}{2^n}} \longrightarrow \pi \text{ για } n \to \infty$$

Συμπέρασμα:

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

-Πώς υπολογίζω τους όρους της ακολουθίας y_n?

Έχουμε,

$$y_{n+1} = 2^{n+1} \sin\frac{\pi}{2^{n+1}} = \quad \boxed{\sin^2 x = \frac{1-\cos 2x}{2}}, \; x = \frac{\pi}{2^{n+1}}$$

$$= 2^{n+1}\sqrt{\frac{1-\cos\frac{\pi}{2^n}}{2}} \quad \Rightarrow \quad \boxed{\sin^2 x + \cos^2 x = 1}$$

$$\Rightarrow y_{n+1} = 2^{n+1}\sqrt{\frac{1-\sqrt{1-\sin^2(\pi/2^n)}}{2}}$$

Παρατηρώ, $y_n = 2^n \sin\frac{\pi}{2^n} \Leftrightarrow \sin\frac{\pi}{2^n} = y_n \cdot 2^{-n}$

Άρα

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

• Αλγόριθμος:

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

- Λόγω της αφαίρεσης σχεδόν ίσων αριθμών που γίνεται στον παραπάνω αλγόριθμο, αυτός (ο αλγόριθμος) είναι ασταθής!

- Για να καταλήξω σε ευσταθή αλγόριθμο πρέπει να προσπαθήσω να αποφύγω την αφαίρεση (σχεδόν ίσων αριθμών)

Ερώτημα: Πώς μπορούμε να αποφύγουμε την αφαίρεση σχεδόν ίσων αριθμών σε αυτόν τον αλγόριθμο?

- Έχουμε την εξής αφαίρεση,

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

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

Αντικαθιστώντας στην (*) παίρνουμε,

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

Ευσταθής Αλγόριθμος

Παράσταση αριθμών ως προς οποιαδήποτε βάση

- Στην καθημερινή ζωή χρησιμοποιούμε το δεκαδικό σύστημα.

Βάση : 10

Ψηφία : 0, 1, ..., 9

– Παράδειγμα:

$$3.14159 = 3\cdot10^{0} + 1\cdot10^{-1} + 4\cdot10^{-2} + 1\cdot10^{-3} + 5\cdot10^{-4} + 9\cdot10^{-5}$$

Γενικά: Έστω $a_N, a_{N-1}, \dots, a_0, a_{-1}, a_{-2}, \dots$ δεκαδικά ψηφία

Τότε

$$(a_N a_{N-1} \dots a_0 . a_{-1} a_{-2} \dots)_{10} =$$

Όταν χωρίζω το ακέραιο από το κλασματικό μέρος του αριθμού μιλάω για σταθερή υποδιαστολή

$$= a_N\cdot10^{N} + a_{N-1}\cdot10^{N-1} + \dots + a_0\cdot10^{0} + a_{-1}\cdot10^{-1} + a_{-2}\cdot10^{-2} + \dots$$

ακέραιο μέρος πάντα πεπερασμένο

• <u>Ακέραιο μέρος</u> : $a_N a_{N-1} \cdots a_0$, είναι η τιμή του πολυωνύμου $p(x) = a_N x^N + a_{N-1} x^{N-1} + \cdots + a_1 x + a_0$

στο σημείο $x=10$.

• <u>Κλασματικό μέρος</u> : $a_{-1} a_{-2} \ldots$, είναι η τιμή της <u>δυναμοσειράς</u> $\sum_{k=1}^{\infty} a_{-k} x^k$ για $x = \frac{1}{10}$.

– Η σειρά μπορεί να έχει <u>πεπερασμένο</u> ή <u>άπειρο</u> πλήθος όρων.

– <u>Παράδειγμα μη μοναδικότητας της παράστασης:</u>

$$4.130 = 4.12999\ldots = 4.12\overline{9}$$

Γιατί ισχύει?

$$\left(\frac{1}{3} = 0.333\ldots \quad \xrightarrow{\times 3} \quad 1 = 0.9999\ldots \right)$$

→ Γενικά : $\sum_{i=0}^{\infty} \omega^i = \frac{1}{1-\omega}$ για $|\omega| < 1$

$$\left(\sum_{i=k}^{\infty} \omega^i = \frac{\omega^k}{1-\omega} \right)$$

Για το $4.12\overline{9}$ έχω:

$$9 \cdot \sum_{i=1}^{\infty} 10^{-i} = 9 \cdot \frac{10^{-1}}{1-10^{-1}} = 9 \frac{\frac{1}{10}}{\frac{9}{10}} = 1$$

$$\boxed{\text{Γενικά} : (b-1) \sum_{i=1}^{\infty} b^{-i} = 1}$$

- Για να έχουμε μοναδικότητα απαιτούμε:

Για κάθε $k_0 \in \mathbb{N}$ υπάρχει $k \geq k_0$ τ.ω. $a_{-k} \neq 9$

• Σύστημα με βάση $b \geq 2$, $b \in \mathbb{N}$

Βάση : b

Ψηφία : $0, 1, 2, \ldots, b-1$

a_k ψηφία,

$$\pm(a_N \cdots a_0 . a_{-1} a_{-2} \cdots)_b = \pm(a_N b^N + \cdots + a_0 b^0 + a_{-1} b^{-1} + a_{-2} b^{-2} + \cdots)$$

Παράδειγμα :

$$(100110.11)_2 = 1 \cdot 2^5 + 1 \cdot 2^2 + 1 \cdot 2^1 + 1 \cdot 2^{-1} + 1 \cdot 2^{-2} = \cdots = (38.75)_{10}$$

i) Μετατροπή από ένα σύστημα με βάση b στο δεκαδικό.

α) Ακέραιων αριθμών:

Παράδειγμα : $(53473)_8 =$

$$= 5 \cdot 8^4 + 3 \cdot 8^3 + 4 \cdot 8^2 + 7 \cdot 8^1 + 3 \cdot 8^0 = \overset{(*)}{\cdots} =$$

$$= (22331)_{10}$$

- Ο πιο εύχρηστος τρόπος να γίνουν οι πράξεις στο (*) είναι να βγάλω το 8 όσες φορές μπορώ κοινό παράγοντα.

$\to 3+8(7+8(4+8(3+8\cdot 5)))$, Σχήμα Horner

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

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

και κάνουμε τις πράξεις από "μέσα" προς τα "έξω".

$y \leftarrow a_N$

για $i = N-1, \dots, 0$

$y \leftarrow \underbrace{a_i + xy}_{\text{flop.}}$

β) Κλασματικών αριθμών x, $(0<x<1)$

Παράδειγμα $(.11)_2 = 1\cdot 2^{-1} + 1\cdot 2^{-2} = 1/2 + 1/4 = (0.75)_{10}$

ii) Μετατροπή από το δεκαδικό σε ένα σύστημα με βάση β.

α) Ακέραιων αριθμών

- Βασίζεται στον αλγόριθμο της διαίρεσης.

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

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

$$= \underset{\downarrow\, υ}{a_0} + 8\underbrace{(a_1 + 8(a_2 + \cdots))}_{\text{ακέραιος} \to π}$$

Άρα, $369 \div 8 = 8 \cdot π + υ$

→ Οπότε a_0 και $a_1 + 8(\cdots)$ είναι το υπόλοιπο και το πηλίκο της διαίρεσης $369 : 8$.

Παράδειγμα:

$$\begin{array}{r|l} 369 & 8 \\ \hline 49 & 46 \\ 1 & \end{array} \to \begin{cases} a_0 = 1 \\ (46)_{10} = a_1 + 8(\cdots) \end{cases}$$

$$\begin{array}{r|l} 46 & 8 \\ \hline 6 & 5 \end{array} \to \begin{cases} a_1 = 6 \\ a_2 + 8\overbrace{(a_3 + \cdots)}^{=0} = 5 \end{cases}$$

Συμπέρασμα: $a_0 = 1$, $a_1 = 6$, $a_2 = 5$, $a_i = 0$, $i \geq 3$

$$(369)_{10} = (561)_8$$

Επαλήθευση, $(561)_8 = 5\cdot 8^2 + 6\cdot 8 + 1 = \cdots = (369)_{10}$

β) Κλασματικών αριθμών

$0<x<1$ $\quad x$ στο δεκαδικό σύστημα

Μετατροπή του x σε σύστημα με βάση β.

$$x = (.a_{-1}a_{-2}\cdots)_\beta$$

$$= a_{-1}\beta^{-1} + a_{-2}\beta^{-2} + \cdots$$

$$\Rightarrow \beta x = a_{-1} + a_{-2}\beta^{-1} + a_{-3}\beta^{-2} + \cdots$$

πολλαπλασιάζω με β.

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

- Παράδειγμα: Μετατροπή του $x = (.372)_{10}$ στο δυαδικό σύστημα.

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

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 $a_{-4} = 1$, $\gamma_4 := 0.952$

$\vdots$

Επομένως :

$$(0.372)_{10} = (0.0101\ldots)_2$$

• **Παρατήρηση** : Είναι δυνατόν το πλήθος των μη μηδενικών ψηφίων στην παράσταση ενός **κλασματικού** αριθμού να είναι πεπερασμένο σε ένα σύστημα και άπειρο σε άλλο σύστημα. (Για ρητούς)

Ισχυρισμός :

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

Άρα,

$$\frac{1}{10} = \underbrace{2^{-4}+2^{-5}}_{n=1} + \underbrace{2^{-8}+2^{-9}}_{n=2} + \underbrace{2^{-12}+2^{-13}}_{n=3} + \ldots$$

$$(0.1)_{10} = (0.00011001100110011\ldots)_2 = (0.000\overline{1100})_2$$

• Αριθμοί μηχανής

Έστω $x \in \mathbb{R}$, $x \neq 0$

• Σε ένα σύστημα με βάση b, ο x μπορεί να γραφεί στην μορφή,

$$\circledast \quad x = \pm .d_1 d_2 \cdots \cdot b^e$$

με $d_1 \neq 0$,

d_i ψηφία ως προς την βάση b και e κατάλληλος εκθέτης.

• Η μορφή $\circledast$ λέγεται (κανονική) μορφή κινητής υποδιαστολής.

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

- b = βάση του αριθμητικού συστήματος
- t = ακρίβεια = το πλήθος των ψηφίων του κλάσματος του αριθμού.

- L = κάτω φράγμα
- U = άνω φράγμα

$\left.\begin{array}{l} L = \text{κάτω φράγμα} \\ U = \text{άνω φράγμα} \end{array}\right\} \longrightarrow$ του εκθέτη e του b $(L \leq e \leq U)$

όπου L, U ακέραιοι και $L \simeq -U$

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

$$\oplus \quad x = \pm .d_1 \cdots d_t \cdot b^e,$$

με $d_1 \neq 0$ και $L \leq e \leq U$

→ Οι αριθμοί μηχανής είναι το μηδέν και όλοι οι αριθμοί της μορφής ⊕

Το σύνολο M είναι πεπερασμένο (⟹ άρα μπορώ να διατάξω τους αριθμούς του)

• Μέγιστο στοιχείο του M: $d_i = b-1,\ i=1,\dots,t$, $e=u$	• Ελάχιστο στοιχείο του M (με αρνητικό πρόσημο): $d_i = b-1,\ i=1,\dots,t$, $e=u$.
• Μικρότερο <u>θετικό</u> στοιχείο του M: $0.100\cdots0 \cdot b^L$	

— Η απόσταση διαδοχικών αριθμών μηχανής <u>δεν είναι σταθερή</u>.

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

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

$$=\frac{3}{2}\sum_{n=1}^{\infty}\frac{1}{2^{4n}}=\frac{3}{2}\sum_{n=1}^{\infty}\left(\frac{1}{16}\right)^{n}=\frac{3}{2}\,\frac{\frac{1}{16}}{1-\frac{1}{16}}=$$

$$=\frac{3}{2}\,\frac{\frac{1}{16}}{\frac{15}{16}}=\frac{3}{2}\cdot\frac{1}{15}=\frac{1}{10}$$

✱ Συνέχεια, Αριθμοί Μηχανής.

– Το M δεν είναι κλειστό ως προς τον πολλαπλασιασμό, δηλ.

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

Παράδειγμα : $.100\ldots0 \cdot b^{L} \cdot .100\ldots0\cdot b^{L} \notin M$

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

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

→ όμοια σε αφαίρεση και διαίρεση.

Παράδειγμα : $b=10$, $t=5$

1, $10^{-5}=.00001 \in M$

$1+10^{-5}=1.00001=.\underbrace{100001}_{6>t}$ άρα $\notin M$

• Μας ενδιαφέρει το M να είναι <u>όσο πιο πυκνό</u> και <u>όσο πιο ευρύ</u> γίνεται, δηλαδή να έχει:

• μεγάλο t
• μεγάλο διάστημα $[L, U]$

• <u>Προσέγγιση αριθμών με αριθμούς μηχανής</u>:

<u>i.</u> Αν $|x| > .d_1d_2\cdots d_t \cdot b^U$, $d_i = b-1$ για $i=1,\dots,t$

τότε παίρνουμε υπερχείλιση (overflow): σταματούν οι υπολογισμοί.

→ Τέτοιοι αριθμοί δεν προσεγγίζονται.

<u>ii.</u> Αν $0 < |x| < .100\cdots 0 \cdot b^L$, τότε έχουμε <u>υπεκχείλιση</u>

→ Τότε κατά κανόνα ο x προσεγγίζεται με το μηδέν και οι υπολογισμοί συνεχίζουν.

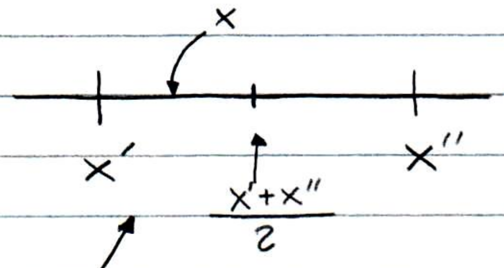
iii. Αν $0.100\cdots 0 \cdot b^L \le |x| \le$ μέγιστο στοιχείο του M, τότε ο x προσεγγίζεται με τον αριθμό $fl(x)$,

- <u>Συνήθως</u> ισχύει ότι,

① $|x - fl(x)| \le |x - y| \quad \forall y \in M$,

→ δεν υπάρχει κανένας αριθμός μηχανής που να απέχει λιγότερο από όσο απέχει ο $fl(x)$ του x.

Π.χ. (Σχηματικά):



(Η διαδικασία αυτή ονομάζεται στρογγύλευση).

$x', x'' \in M$ διαδοχικοί

- Αν $x < \frac{x'+x''}{2}$ τότε προσεγγίζω τον x με τον x'.
- Αν $x > \frac{x'+x''}{2}$ τότε προσεγγίζω το x με x''.
- Αν $x = \frac{x'+x''}{2}$ τότε επιλέγω x' ή x''.

- Ισχυρισμός: Αν ισχύει η 1, τότε,

σχετικό σφάλμα $\left\{ \left| \frac{\overbrace{fl(x)-x}^{\text{απόλυτο σφάλμα}}}{x} \right| \le \frac{1}{2} b^{1-t} \right.$ (για $x \neq 0$) (2)

Απόδειξη του 2:

a) Αν $fl(x)=x$ (δηλαδή $x=x'$ ή $x=x''$) τότε η 2 είναι προφανής

b) Αν $x \notin M$, τότε υπάρχουν διαδοχικοί αριθμοί $x', x'' \in M$ τ.ω. $x' < x < x''$. Προφανώς

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

οπότε, $\left| \frac{fl(x)-x}{x} \right| \le \frac{|x''-x'|}{2|x|}$

οι απόλυτες τιμές δεν παίζουν κάποιο ρόλο εδώ.

Έστω ότι $x>0$ (όμοια για $x<0$),

$$x = \underbrace{.d_1 d_2 \cdots d_t d_{t+1} \cdots}_{x \notin M} \cdot \beta^k \quad , \quad L \leq k \leq U$$

→ Τότε

$$x' = .d_1 d_2 \cdots d_t \cdot \beta^k \quad , \quad d_1 \neq 0$$

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

Επομένως, $\boxed{x'' - x' = \beta^{k-t}}$ ← Η διαφορά διαδοχικών αριθμών μηχανής δεν είναι σταθερή.

Παράδειγμα:

5 δεκαδικά ψηφία,

$x = 0.34572|1383$

$x' = 0.34572$

$x'' = 0.34573$

$x' < x < x''$

Άρα, $\dfrac{|fl(x) - x|}{|x|} \leq \dfrac{\beta^{k-t}}{2x}$ → βιβλίο σελ. 8

Τώρα, $\boxed{x \geq .1 \cdot \beta^k}$ άρα, $\dfrac{|fl(x) - x|}{|x|} \leq \dfrac{\beta^{k-t}}{2 \cdot 0.1 \cdot \beta^k} =$

$= \dfrac{\beta^{k-t}}{2 \cdot \beta^{k-1}} = \dfrac{1}{2} \beta^{1-t} \Rightarrow \left| \dfrac{fl(x) - x}{x} \right| \leq \dfrac{1}{2} \beta^{1-t}$

ως προς την βάση $\beta \geq 2$.

Άρα αν,

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

τότε

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

– Στον $fl(x)$ οδηγούμαστε από τον x είτε με <u>στρογγύλευση</u> είτε με <u>αποκοπή</u>. Όταν κάνουμε στρογγύλευση ισχύει η (1).

• <u>Στρογγύλευση</u>: π.χ. $b=10$, $t=5$

$x = .a_1a_2 \cdots a_5a_6 \cdots \cdot 10^k$

$b/2 = 10/2 = 5.$

• Αν $a_6 \geq 5$, τότε $fl(x) = x'' = (.a_1a_2 \cdots a_5 + 10^{-5}) \cdot 10^k$

• Αν $a_6 < 5$, τότε $fl(x) = x' = .a_1a_2 \cdots a_5 \cdot 10^k$

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

— Εναλλακτικός τρόπος υπολογισμού του $fl(x)$: <u>αποκοπή</u>

→ Η στρογγύλευση δίνει καλύτερα αποτελέσματα από την αποκοπή.

π.χ. $x = 1.1\overline{9}$ → με στρογγύλευση, $fl(x) = 1.2$

→ με αποκοπή, $fl(x) = 1.1$

– Αντίστοιχη της ② (στην αποκοπή) είναι τώρα η

③ $$\left| \frac{fl(x)-x}{x} \right| \leq \beta^{1-t}$$ (χωρίς το 1/2)

• Συμπέρασμα και για τις δύο περιπτώσεις (αποκοπή ή στρογγύλευση).

Έχουμε,

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

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

u = μοναδιαίο σφάλμα στρογγύλευσης.

Πράξεις : $* \in \{+,-,\times,:\}$

Υπόθεση : Δεδομένων αριθμών x, y προσεγγίζω με αριθμούς μηχανής άρα $fl(x), fl(y)$.

και, $z = fl(\underbrace{fl(x) * fl(y)})$

→ αυτή η πράξη δεν γίνεται ακριβώς, παρ'όλα αυτά είναι πολύ κοντά στο ακριβές αποτέλεσμα.

- **Παράδοξα :** $b = 10$, $t = 5$, $U = -L = 10$

- Έστω ότι κάνω στρογγύλευση

Αριθμοί : $a_1 = 1 \longrightarrow .1 \cdot 10^1$
$a_2 = 3 \cdot 10^{-5} \longrightarrow .3 \cdot 10^{-4}$
$a_3 = 3 \cdot 10^{-5} \longrightarrow .3 \cdot 10^{-4}$

$a_1, a_2, a_3 \in M$

Τότε, (αφού $a_1 \in M$ $a_1 = fl(a_1)$)

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

Άρα $\longrightarrow fl(fl(a_1 + a_2) + a_3) = 1$

- Αλλά $a_2 + a_3 = 6 \cdot 10^{-5} \in M$

$fl(a_1 + (a_2 + a_3)) =$ (παραλείπω το fl γιατί είναι ήδη αριθμοί μηχανής)

$= fl(1+6\cdot 10^{-5}) = fl(1.00006) = .10001\cdot 10^{1} = 1.0001$

- Διαφορετικά αποτελέσματα!

<u>Συμπέρασμα</u>: Έχει σημασία η σειρά με την οποία γίνονται οι προσθέσεις για τον υπολογισμό ενός αθροίσματος στον υπολογιστή!

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

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

($\frac{1}{2}\beta^{1-t}$ = έψιλον της μηχανής
μηδέν της μηχανής)

→ Βιβλίο σελ. 11.

• <u>Επιρροή των σφαλμάτων στρογγύλευσης στους υπολογισμούς</u>

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

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

<u>08/03/2018</u>

• <u>Επιρροή των σφαλμάτων στρογγύλευσης στους υπολογισμούς</u>

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

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

<u>Ζητούμενο</u> : Εκτίμηση (της απόλυτης τιμής) του <u>σχετικού σφάλματος</u>,

$$\boxed{\frac{fl(fl(x)*fl(y)) - x*y}{x*y}}$$

Γνωρίζω ότι:

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

• <u>Δύο βοηθητικές παρατηρήσεις :</u>

<u>1.</u> Η ⊛ γράφεται στη μορφή

$$fl(x) = x(1+\varepsilon) \quad \text{με} \quad \varepsilon = \varepsilon(x)$$
τ.ω. $|\varepsilon| \leq u$

με $\varepsilon := \dfrac{fl(x) - x}{x}$ έχουμε και τα δύο ζητούμενα.

<u>2.</u> Αν ε_i, $1 \leq i \leq m$, ικανοποιούν τις $|\varepsilon_i| \leq u < 1$, τότε υπάρχει ε, με $|\varepsilon| \leq u$, τ.ω.

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

<u>Απόδειξη</u> : Θέτω $\lambda := \prod_{i=1}^{m} (1+\varepsilon_i)$ και έχω

$$(1-u)^m \leq \lambda \leq (1+u)^m$$

- Θεωρούμε τη συνεχή συνάρτηση $f(x) = (1+x)^m$ και παρατηρούμε ότι,

$$f(-u) \leq \lambda \leq f(u)$$

Άρα σύμφωνα με το θεώρημα της ενδιάμεσης τιμής, υπάρχει $\varepsilon \in [-u, u]$ τ.ω. $\lambda = f(\varepsilon)$.

– Άρα για αριθμούς x, y με

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

α) <u>Πολλαπλασιασμός</u>:

$$z = fl(fl(x) \cdot fl(y)) = fl(xy(1+\varepsilon_1)(1+\varepsilon_2)) =$$
$$= xy(1+\varepsilon_1)(1+\varepsilon_2)(1+\varepsilon_3) = xy(1+\varepsilon)^3$$

με $|\varepsilon_1|, |\varepsilon_2|, |\varepsilon_3|, |\varepsilon| \le u$

Άρα,

$$\left|\frac{z-xy}{xy}\right| = \left|\frac{xy(1+\varepsilon)^3 - xy}{xy}\right| =$$

πολύ μικρός

$$= |(1+\varepsilon)^3 - 1| = |3\varepsilon + 3\varepsilon^2 + \varepsilon^3| \le 3u + \cancel{3}_{4}u^2$$

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

β) <u>Διαίρεση</u>:

$$z = fl\left(\frac{fl(x)}{fl(y)}\right) = fl\left(\frac{x(1+\varepsilon_1)}{y(1+\varepsilon_2)}\right) = \frac{x(1+\varepsilon_1)}{y(1+\varepsilon_2)}(1+\varepsilon_3) =$$
$$= \frac{x}{y}(1+\varepsilon)^2 \cdot \frac{1}{1+\varepsilon_2}$$

Έχουμε,

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

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

→ θέλω την μέγιστη τιμή του

Επομένως, $\left|\frac{z - x/y}{x/y}\right| = \left|(1+\varepsilon)^2(1+\delta) - 1\right| =$

$$= |\underbrace{2\varepsilon + \delta}_{\le 3u + O(u^2)} + \varepsilon^2 + 2\varepsilon\delta + \delta\varepsilon^2|$$

$\varepsilon^2 \to O(u^2)$, $\quad 2\varepsilon\delta \to \le 2(u(u+O(u^2)))$, $\quad \delta\varepsilon^2 \to \le (u+O(u^2))\,u^2$

$\to \le 3u + O(u^2)$

γ) <u>Πρόσθεση - Αφαίρεση</u>

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

$$= x\underbrace{(1+\varepsilon_1)(1+\varepsilon_3)}_{=(1+\varepsilon)^2} + y\underbrace{(1+\varepsilon_2)(1+\varepsilon_3)}_{=(1+\delta)^2}$$

με $|\varepsilon|, |\delta| \le u$

Άρα, $z = x(1+\varepsilon)^2 + y(1+\delta)^2 \simeq (x+y) + 2x\varepsilon + 2y\delta \Rightarrow$

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

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

;

• **1η Περίπτωση** : x, y ομόσημοι,

Τότε $|x+y| = |x| + |y|$, οπότε

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

• **2η Περίπτωση** : x, y ετερόσημοι,

- Στη χειρότερη περίπτωση $\varepsilon \simeq -\delta$ και $|\varepsilon| \simeq u$, οπότε

$$\left| \frac{x\varepsilon + y\delta}{x+y} \right| \simeq \frac{|x-y|}{|x+y|}\, u$$

- Αυτός ο παράγοντας μπορεί να γίνει πολύ μεγάλος όταν αφαιρούμε σχεδόν ίσους αριθμούς.

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

• **Συμπέρασμα** : Πρέπει να αποφεύγουμε την αφαίρεση σχεδόν ίσων αριθμών (ή να την κάνουμε με μεγαλύτερη ακρίβεια).

- <u>Παρατήρηση</u>: Η αφαίρεση σχεδόν ίσων αριθμών <u>μηχανής</u> γίνεται χωρίς πρόβλημα:

$$z = fl(x+y) = (x+y)(1+\varepsilon)$$

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

<u>Παράδειγμα</u>:

$b=10,\ t=5,\ U=-L=10$, στρογγύλευση

$x = .45142708$
$y = -.45115944$

$x+y = .26764 \cdot 10^{-3}$ (Ακριβές αποτέλεσμα)

$fl(x) = .45143$
$fl(y) = -.45116$

Έχουμε: $z = fl(fl(x)+fl(y)) = fl(.45143 - .45116) =$

$= \underbrace{.000}27 = .27 \cdot 10^{-3}$ → ένδειξη αφαίρεσης σχεδόν ίσων αριθμών.

Έχουμε, $\left| \dfrac{z-(x+y)}{x+y} \right| \simeq 88 \cdot 10^{-4}$ (88 φορές πιο μεγάλο)

Αν ήταν ομόσημοι → $2u = 10^{-4}$, $u = \frac{1}{2} b^{1-t} = \boxed{\frac{1}{2} 10^{-4}}$

- Πως μπορούμε να αποφύγουμε την αφαίρεση σχεδόν ίσων αριθμών,

1º Παράδειγμα: $b=10$, $t=10$

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

από υπολογιστή

$\underbrace{0000}$ λάθος

Πώς να το αποφύγω:

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

Άρα,

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

$= .5628468294 \cdot 10^{-2}$ πολύ καλή προσέγγιση.

2º Παράδειγμα: Θέλουμε να υπολογίσουμε τις τιμές της συνάρτησης $f(x) = x - \sin x$ για $|x|$ μικρό.

- x, $\sin x$ ομόσημα
- $\lim_{x \to 0} \frac{\sin x}{x} = 1$

- Άρα έχουμε πρόβλημα αφαίρεσης σχεδόν ίσων αριθμών.

Πώς το λύνουμε:

• Ανάπτυγμα Taylor,

$$\sin x = x - \frac{x^3}{6} + \varepsilon(x) \quad \text{με} \quad |\varepsilon(x)| \leq \frac{|x|^5}{120}$$

Άρα $f(x) \simeq \frac{x^3}{6}$

• **Σφάλματα στον υπολογισμό αθροισμάτων:**

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

Παράδειγμα:

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

– Έχουμε $\frac{1}{k^2+k} = \frac{1}{k(k+1)} = \frac{1}{k} - \frac{1}{k+1}$, οπότε

$$S_n = 1 + (1 - 1/2) + (1/2 - 1/3) + \cdots + (1/n - 1/{n+1}) =$$
$$= 2 - \frac{1}{n+1} \quad \text{(τηλεσκοπικό άθροισμα)}$$

Η ακολουθία $(S_n)_{n\in\mathbb{N}}$ είναι αύξουσα και τείνει στο 2.

Άρα, $\boxed{S_{9999} = 1.9999}$, ακριβές αποτέλεσμα

<u>$1^{ος}$ Αλγόριθμος</u> : Προσθέτω από τον μεγαλύτερο προς τον μικρότερο όρο (όλοι οι όροι είναι θετικοί)

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

Με $b=10$, $t=10$, παίρνουμε $\hat{S}_{9999} = 1.999899972$

<u>$2^{ος}$ Αλγόριθμος</u> : Προσθέτω από τον μικρότερο προς τον μεγαλύτερο όρο :

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

Προφανώς $T_n = S_n$.

Με $b=10$, $t=10$ παίρνουμε

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

- <u>Ερώτημα</u> : Γιατί με τον δεύτερο αλγόριθμο παίρνουμε καλύτερα αποτελέσματα;

<u>Γενικότερα</u> : Πώς υπολογίζω σωστά αθροίσματα?

- <u>Παρατήρηση</u>: Έστω $\lambda, \mu \in \mathbb{R}$ και $\varepsilon_1, \varepsilon_2 \in [-u, u]$. Τότε υπάρχει $\varepsilon_3 \in [-u, u]$ τ.ω.

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

Απόδειξη του (*):

για $\lambda = \mu = 0$ είναι προφανές.

αλλιώς με $\varepsilon_3 = \dfrac{\lambda\varepsilon_1 + \mu\varepsilon_2}{|\lambda| + |\mu|}$ ικανοποιείται

η (*) και $|\varepsilon_3| = \dfrac{|\lambda\varepsilon_1 + \mu\varepsilon_2|}{|\lambda| + |\mu|} \leq$

$$\leq \frac{|\lambda||\varepsilon_1| + |\mu||\varepsilon_2|}{|\lambda| + |\mu|} \leq \frac{(|\lambda| + |\mu|)u}{(|\lambda| + |\mu|)} = u$$

• Σφάλματα στον υπολογισμό αθροισμάτων

(a)

- **Παρατήρηση**: Έστω $\lambda, \mu \in \mathbb{R}$ και $\varepsilon_1, \varepsilon_2 \in [-u, u]$. Τότε $\exists$ $\varepsilon_3 \in [-u, u]$ τ.ω. $\lambda\varepsilon_1 + \mu\varepsilon_2 = (|\lambda| + |\mu|)\varepsilon_3$

• **Πρόβλημα**: Έστω $a_1, \dots, a_N \in M$

Θέλουμε να υπολογίσουμε το άθροισμα,

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

• **Αλγόριθμος**: $S_1 = a_1$

$S_k = S_{k-1} + a_k \quad , k = 2, 3, \dots, N$

- Παίρνουμε τις προσεγγίσεις:

$$\hat{S}_1 = S_1 = a_1 \quad , \hat{S}_k = fl(\hat{S}_{k-1} + a_k) \quad , k = 2, \dots, N$$

είναι ήδη αριθμοί μηχανής.

- Έχουμε,

$$\hat{S}_2 = fl(\hat{S}_1 + a_2) = \underbrace{(a_1 + a_2)}_{S_2}(1 + \delta) = \quad , |\delta| \le u$$

$$= S_2 + S_2\delta = S_2 + |S_2|\varepsilon_2$$

εφαρμογή παρατήρησης (a)

με $|\delta|, |\varepsilon_2| \le u$

Παρόμοια,

$$\hat{S}_3 = fl(\hat{S}_2 + a_3) = (\hat{S}_2 + a_3)(1 + \delta') =$$

$$= (S_2 + |S_2|\varepsilon_2 + a_3)(1 + \delta') =$$

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

$$= S_3 + \underbrace{|S_2|\varepsilon_2 + S_3\delta'}_{= (|S_2| + |S_3|)\varepsilon_3} + |S_2|\varepsilon_2 \cdot \delta'$$

τάξης $O(u^2)$

με $|\delta'|, |\varepsilon_3| \le u$

Άρα,

$$\hat{S}_3 \simeq S_3 + (|S_2| + |S_3|)\varepsilon_3$$

Με σφάλμα της τάξης του u^2.

- Συνεχίζοντας ανάλογα παίρνουμε,

$$\hat{S}_N \simeq S_N + (|S_2| + |S_3| + \cdots + |S_N|)\varepsilon_N \quad \text{με } |\varepsilon_N| \leq u$$

με σφάλμα της τάξης του u^2.

- <u>Με ενδιαφέρει το σχετικό σφάλμα</u>:

$$\frac{\hat{S}_N - S_N}{S_N} \simeq \frac{|S_2| + \cdots + |S_N|}{S_N}\varepsilon_N$$

$$\left|\frac{\hat{S}_N - S_N}{S_N}\right| \simeq \frac{|S_2| + \cdots + |S_N|}{|S_N|}|\varepsilon_N|$$

$$\leq \frac{|S_2| + \cdots + |S_N|}{|S_N|}u$$

Αυτός ο λόγος μου δίνει την πληροφορία

Με

$$\gamma_N := |S_2| + \cdots + |S_N|$$

και $\quad \rho_N := \dfrac{\gamma_N}{|S_N|}$

έχουμε,

$$\left| \frac{\tilde{S}_N - S_N}{S_N} \right| \simeq \frac{\gamma_N}{|S_N|} |\varepsilon_N| \leq \underbrace{\frac{\gamma_N}{|S_N|}}_{\rho_N} u$$

ρ_N : <u>συντελεστής μετάδοσης</u> του σχετικού σφάλματος στρογγύλευσης για τον αλγόριθμό μας.

- Αν ο ρ_N είναι μεγάλος αριθμός, τότε ο αλγόριθμός είναι <u>ασταθής</u>.

- Αν ένα ενδιάμεσο άθροισμα έχει απόλυτη τιμή πολύ μεγαλύτερη από την $|S_N|$, τότε ο ρ_N είναι μεγάλος.

<u>Παράδειγμα</u> : Προσέγγιση του e^{-x} για $x >> 1$

Τα

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

είναι καλές προσεγγίσεις του e^{-x} για μεγάλο N.

Τώρα, για $x = 100$ έχουμε $e^{-100} \simeq 0$ ενώ,

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

→ Άρα ο αλγόριθμος είναι εντελώς ασταθής το ρ_N γίνεται πολύ μεγάλος.

Πώς αντιμετωπίζεται;

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

αυτό υπολογίζεται εύκολα.

• <u>Ειδική περίπτωση</u>: $a_i > 0$, $i=1,\dots,N$

Τότε, $\gamma_N = S_2 + S_3 + \dots + S_N$
$= (N-1)a_1 + (N-1)a_2 +$
$+ (N-2)a_3 + (N-3)a_4 + \dots$
$+ a_N$

δεν χρειάζονται οι απόλυτες τιμές αφού $a_i > 0$ και τα ενδιάμεσα αθροίσματα είναι θετικά

- <u>Παρατηρώ ότι</u>: το γ_N ελαχιστοποιείται άν το $a_1, a_2 \le a_3 \le a_4 \le \dots \le a_N$
- Θέλω τους μεγάλους συντελεστές να πολλαπλασιάζονται με μικρούς αριθμούς a_i.

- Το γ_N μεγιστοποιείται $a_1, a_2 \ge a_3 \ge a_4 \ge \dots \ge a_N$

• Κάθε άθροισμα με θετικούς και αρνητικούς όρους γράφεται ώς διαφορά δύο αθροισμάτων με θετικούς όρους.

13/03/2018

- Ευστάθεια αλγορίθμων

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

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

- Έχουμε ήδη δει πολλά παραδείγματα. Ας θυμηθούμε κάποια:

1. Υπολογισμός το e^{-x} για $x>>1$, από τον τύπο

$$e^{-x} = 1 - x + \frac{x^2}{2!} - \frac{x^3}{3!} + \dots$$

(ασταθής τρόπος)

– Αλλά ο υπολογισμός με τον τύπο,

$$e^{-x} = \frac{1}{e^x} = \frac{1}{1 + \frac{x^2}{2} + \frac{x^3}{3!} + \dots}$$

γίνεται με ευσταθή τρόπο.

2. $I_n = \int_0^1 x^n e^{x-1} dx \quad , n=1,2,\dots$

$$\begin{cases} I_1 = 1/e \\ I_n = 1 - n I_{n-1} \quad , n=2,\dots \end{cases}$$

– Ασταθής αλγόριθμος. Είδαμε επίσης και ευσταθή αλγόριθμο για τον υπολογισμό των I_n.

3. Προσέγγιση του π

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

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

Αφαίρεση σχεδόν ίσων αριθμών για μεγάλο n.

→ Ασταθής αλγόριθμος

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

→ Ευσταθής αλγόριθμος

• Κατάσταση προβλημάτων

• Λέμε ότι ένα πρόβλημα έχει <u>καλή κατάσταση</u>, αν μικρές μεταβολές στα δεδομένα του έχουν ως συνέπεια μικρή μεταβολή της λύσης του.

• Λέμε ότι ένα πρόβλημα έχει <u>κακή κατάσταση</u>, αν είναι δυνατόν μικρές μεταβολές στα δεδομένα του να έχουν ως συνέπεια μεγάλη μεταβολή της λύσης του.

<u>Παράδειγμα</u>:

$(x-2)^6=0$ Λύση: $x^*=2$ (πολλαπλότητας 6).

$(x-2)^6=10^{-6}$ Λύσεις: $x_k = 2+\frac{1}{10}e^{\frac{2i\pi k}{6}}$, $k=0,1,\dots,5$

(Άσκηση 1.15)

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

$$\Rightarrow \boxed{|x_k-x^*| = \frac{1}{10}}$$

<u>Κακή κατάσταση</u>!

– Όμως, $(x-2)^6=0 \Leftrightarrow \underline{x-2=0}$

↳ <u>Πολύ καλή κατάσταση</u>!