ΤΕΙ ΚΕΝΤΡΙΚΗΣ ΜΑΚΕΔΟΝΙΑΣ

ΑΝΟΙΚΤΑ ΑΚΑΔΗΜΑΙΚΑ ΜΑΘΗΜΑΤΑ

ΤΕΧΝΟΛΟΓΙΑ ΛΟΓΙΣΜΙΚΟΥ Ι

κ. ΠΕΤΑΛΙΔΗΣ

***ΤΜΗΜΑ* ΜΗΧΑΝΙΚΩΝ ΠΛΗΡΟΦΟΡΙΚΗΣ ΤΕ**

Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο
ΕΠΙΧΕΙΡΗΣΙΑΚΟ ΠΡΟΓΡΑΜΜΑ
ΕΚΠΑΙΔΕΥΣΗ ΚΑΙ ΔΙΑ ΒΙΟΥ ΜΑΘΗΣΗ
επένδυση στην κοινωνία της γνώσης
ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ & ΘΡΗΣΚΕΥΜΑΤΩΝ, ΠΟΛΙΤΙΣΜΟΥ & ΑΘΛΗΤΙΣΜΟΥ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ
ΕΣΠΑ 2007-2013
πρόγραμμα για την ανάπτυξη
ΕΥΡΩΠΑΪΚΟ ΚΟΙΝΩΝΙΚΟ ΤΑΜΕΙΟ
Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης

Άδειες Χρήσης

- Το παρόν εκπαιδευτικό υλικό υπόκειται σε άδειες χρήσης Creative Commons.
- Για εκπαιδευτικό υλικό, όπως εικόνες, που υπόκειται σε άλλου τύπου άδειας χρήσης, η άδεια χρήσης αναφέρεται ρητώς.

Χρηματοδότηση

- Το παρόν εκπαιδευτικό υλικό έχει αναπτυχθεί στα πλαίσια του εκπαιδευτικού έργου του διδάσκοντα.
- Το έργο «Ανοικτά Ακαδημαϊκά Μαθήματα στο ΤΕΙ Κεντρικής Μακεδονίας» έχει χρηματοδοτήσει μόνο τη αναδιαμόρφωση του εκπαιδευτικού υλικού.
- Το έργο υλοποιείται στο πλαίσιο του Επιχειρησιακού Προγράμματος «Εκπαίδευση και Δια Βίου Μάθηση» και συγχρηματοδοτείται από την Ευρωπαϊκή Ένωση (Ευρωπαϊκό Κοινωνικό Ταμείο) και από εθνικούς πόρους.

Τεχνολογικό Εκπαιδευτικό Ίδρυμα
Κεντρικής Μακεδονίας - Σέρρες
Τμήμα Μηχανικών Πληροφορικής

Αριθμητικές Μέθοδοι σε Προγραμματιστικό Περιβάλλον

Δρ. Δημήτρης Βαρσάμης
Επίκουρος Καθηγητής

Αριθμητικές Μέθοδοι σε Προγραμματιστικό Περιβάλλον
Πρώτη Σειρά Διαφανειών

Διδακτικά εγχειρίδια - Εύδοξος

Έντυπα εγχειρίδια (Εύδοξος)

1. Εισαγωγή στην Αριθμητική Ανάλυση,
   Λεωνίδας Πιτσούλης
2. Αριθμητική ανάλυση με εφαρμογές σε Matlab & Mathematica,
   Γεώργιος Σ. Παπαγεωργίου, Χαράλαμπος Γ. Τσίτουρας
3. Αριθμητικές Μέθοδοι και Εφαρμογές για Μηχανικούς,
   Ι. Σαρρής, Θ. Καρακασίδης
4. Αριθμητική Ανάλυση: Εισαγωγή,
   Μιχαήλ Ν. Βραχάτης

Ηλεκτρονικά εγχειρίδια

- Προσωπική Ιστοσελίδα
  1. Διαφάνειες
  2. Συμπληρωματικές Σημειώσεις
  3. E-book

Εισαγωγή

- Αριθμητική Ανάλυση (Numerical Analysis)
  - Μετατροπή μαθηματικών προβλημάτων σε ισοδύναμα προβλήματα που επιλύονται αριθμητικά με την βοήθεια υπολογιστή.
- Προβλήματα εφαρμοσμένων μαθηματικών (Applied Mathematics Problems)
  - Επίλυση μη γραμμικών εξισώσεων
  - Προσέγγιση συναρτήσεων
  - Παραγώγιση
  - Ολοκλήρωση
  - Επίλυση διαφορικών εξισώσεων
  - Βελτιστοποίηση συναρτήσεων

Αριθμητική Ανάλυση και Εφαρμοσμένα Μαθηματικά

- Εφαρμογή σε επιστημονικά πεδία
  - Επιστήμη των Η/Υ
  - Θεωρία Ελέγχου
  - Υπολογιστική Νοημοσύνη
  - Επιχειρησιακή Έρευνα
  - Κρυπτογραφία
  - Εξόρυξη Δεδομένων
  - Στατιστική κ.α.

Εισαγωγή

- Αριθμητική Ανάλυση
  - Δημιουργία κατάλληλης μεθόδου (Αλγόριθμος)
  - Υλοποίηση της μεθόδου σε υπολογιστή

- Μια μέθοδος είναι κατάλληλη όταν προσεγγίζει «*αρκετά καλά*» το αποτέλεσμα, με το μικρότερο υπολογιστικό κόστος, αλλά και την μικρότερη δέσμευση μνήμης.

Εισαγωγή

- Αριθμητικά συστήματα
- Αριθμητική κινητής υποδιαστολής
- Σφάλματα
- Κατάσταση προβλημάτων

Αριθμητικά συστήματα

Κάθε αριθμός μπορεί να παρασταθεί ως εξής

$$\begin{aligned} x &= \pm a_n b^n + a_{n-1} b^{n-1} + \ldots + a_0 b^0 + a_{-1} b^{-1} + \ldots \\ &= \pm \sum_{i=n}^{-\infty} a_i b^i \end{aligned}$$

με $0 \leq a_i < b$.
όπου a_i είναι τα ψηφία του αριθμού x και b είναι η βάση του.

Αριθμητικά συστήματα

- Το ακέραιο μέρος του αριθμού x είναι

$$\begin{aligned} [x] &= \pm a_n b^n + a_{n-1} b^{n-1} + \ldots + a_0 b^0 \\ &= \pm \sum_{i=n}^{0} a_i b^i \end{aligned} \tag{1}$$

- Το κλασματικό μέρος του αριθμού x είναι

$$\begin{aligned} x - [x] &= \pm a_{-1} b^{-1} + a_{-2} b^{-2} + \ldots \\ &= \pm \sum_{i=-1}^{-\infty} a_i b^i \end{aligned} \tag{2}$$

Αριθμητικά συστήματα

- Η αριθμητική παράσταση του αριθμού x είναι

$$x = \pm \left(a_n a_{n-1} \cdots a_0 . a_{-1} \cdots \right)_b$$

  το σύμβολο (.) είναι η υποδιαστολή του αριθμού που διαχωρίζει το ακέραιο με το κλασματικό μέρος ενός αριθμού

- Ανάλογα με την τιμή του b, δηλαδή, της βάσης ονομάζουμε και το αριθμητικό σύστημα.
  Π.χ. $b = 2$, Δυαδικό αριθμητικό σύστημα
  $b = 10$, Δεκαδικό αριθμητικό σύστημα

Αριθμητικά συστήματα

- Το μήκος ενός ακέραιου αριθμού x σε δυαδικά ψηφία δίνεται από τον τύπο

$$L_i = \lceil \log_2(x) \rceil$$

ενώ σε δεκαδικά ψηφία δίνεται από τον τύπο

$$L_i = \lceil \log_{10}(x) \rceil$$

- Για παράδειγμα, ο αριθμός $x = 236$ έχει μήκος

$$L_i = \lceil \log_2(236]) \rceil = \lceil 7.88264 \rceil = 8$$

δυαδικά ψηφία, και

$$L_i = \lceil \log_{10}(236) \rceil = \lceil 2.37291 \rceil = 3$$

δεκαδικά ψηφία.

Αριθμητικά συστήματα

- Μετατροπές αριθμών από ένα αριθμητικό σύστημα σε άλλο
  - Μετατροπή ακεραίου x από βάση b σε δεκαδικό σύστημα
  - Μετατροπή κλασματικού x από βάση b σε δεκαδικό σύστημα
  - Μετατροπή ακεραίου x από δεκαδικό σύστημα σε βάση με b
  - Μετατροπή κλασματικού x από δεκαδικό σύστημα σε βάση με b

Αριθμητικά συστήματα

Μετατροπή ακεραίου x από βάση b σε δεκαδικό σύστημα

- Απλή διαδικασία, αν ακολουθήσουμε τον τύπο (1), π.χ.

$$\begin{aligned}(53473)_8 &= 5 \cdot 8^4 + 3 \cdot 8^3 + 4 \cdot 8^2 + 7 \cdot 8^1 + 3 \cdot 8^0 \\ &= (22331)_{10}\end{aligned}$$

- Στη παραπάνω διαδικασία εκτελέστηκαν $5 + 4 + 3 + 2 + 1 = 15$ πολλαπλασιασμοί και 4 προσθέσεις.

Αριθμητικά συστήματα

Algorithm 1 Μετατροπή ακεραίου x από βάση b σε δεκαδικό σύστημα (Απ' ευθείας)

Input: $x \in \mathbb{Z}$, b
  $y \leftarrow 0$
  for $i = 0$ to n **do**
    $y \leftarrow y + a_i * b^i$
  end for
Output: y

a_i είναι τα ψηφία του αριθμού x.

Επομένως, για τον αριθμό $x = (53473)_8$ θα έχουμε

i	y
–	0
0	$0 + 3 \cdot 8^0 = 3$
1	$3 + 7 \cdot 8^1 = 59$
2	$59 + 4 \cdot 8^2 = 315$
3	$315 + 3 \cdot 8^3 = 1851$
4	$1851 + 5 \cdot 8^4 = 22331$

δηλαδή, $y = (22331)_{10}$.

Αριθμητικά συστήματα

Μετατροπή ακεραίου x από βάση b σε δεκαδικό σύστημα (Απ' ευθείας)

- Συνάρτηση σε MATLAB

```
function y=b2dec(x,b)
xc=num2str(x);
n=length(xc);
for i=1:n
    a(i)=str2num(xc(n-i+1));
end
y=0;
for i=1:n
    y=y+a(i)*b^(i-1);
end
```

Αριθμητικά συστήματα

Μετατροπή ακεραίου x από βάση b σε δεκαδικό σύστημα

- Αν ακολουθήσουμε διαφορετική τακτική (Σχήμα Horner) μπορούμε να μειώσουμε τον αριθμό των πράξεων, π.χ.

$$
\begin{aligned}
(53473)_8 &= 5 \cdot 8^4 + 3 \cdot 8^3 + 4 \cdot 8^2 + 7 \cdot 8^1 + 3 \cdot 8^0 \\
&= \left(5 \cdot 8^3 + 3 \cdot 8^2 + 4 \cdot 8^1 + 7\right) \cdot 8 + 3 \\
&= \left(\left(5 \cdot 8^2 + 3 \cdot 8^1 + 4\right) \cdot 8 + 7\right) \cdot 8 + 3 \\
&= \left(\left(\left(5 \cdot 8^1 + 3\right) \cdot 8 + 4\right) \cdot 8 + 7\right) \cdot 8 + 3 \\
&= (22331)_{10}
\end{aligned}
$$

- Στη παραπάνω διαδικασία (στο τελευταίο βήμα) εκτελέστηκαν 4 πολλαπλασιασμοί και 4 προσθέσεις.

Αριθμητικά συστήματα

Algorithm 2 Μετατροπή ακεραίου x από βάση b σε δεκαδικό σύστημα (Σχήμα horner)

Input: a_i, b

  $y \leftarrow a_n$

  for $i = n - 1$ to 0 **do**

    $y \leftarrow a_i + y * b$

  end for

Output: y

---

a_i είναι τα ψηφία του αριθμού x.

Επομένως, για τον αριθμό $x = (53473)_8$ θα έχουμε

i	y
-	5
3	$3 + 5 \cdot 8 = 43$
2	$4 + 43 \cdot 8 = 348$
1	$7 + 348 \cdot 8 = 2791$
0	$3 + 2791 \cdot 8 = 22331$

δηλαδή, $y = (22331)_{10}$.

Αριθμητικά συστήματα

Μετατροπή ακεραίου x από βάση b σε δεκαδικό σύστημα (Σχήμα horner)

- Συνάρτηση σε MATLAB

```
function y=b2dec_h(x,b)
xc=num2str(x);
n=length(xc);
for i=1:n
    a(i)=str2num(xc(n-i+1));
end
y=a(n);
for i=n-1:-1:1
    y=a(i)+b*y;
end
```

Αριθμητικά συστήματα

Μετατροπή κλασματικού x από βάση b σε δεκαδικό σύστημα

- Απλή διαδικασία, αν ακολουθήσουμε τον τύπο (2), π.χ.

$$\begin{aligned}(.53)_8 &= 5 \cdot 8^{-1} + 3 \cdot 8^{-2} \\ &= 5 \cdot \frac{1}{8} + 3 \cdot \frac{1}{8^2} \\ &= (.671875)_{10}\end{aligned}$$

- Ισοδύναμα

$$\begin{aligned}(.53)_8 &= 5 \cdot 8^{-1} + 3 \cdot 8^{-2} \\ &= \left(5 + 3 \cdot \frac{1}{8}\right) \cdot \frac{1}{8} \\ &= (.671875)_{10}\end{aligned}$$

Algorithm 3 Μετατροπή κλασματικού x από βάση b σε δεκαδικό σύστημα

Input: $x \in \mathbb{Z}$, b, $k \in \mathbb{N}$

$y \leftarrow 0$

for $i = -k$ to -1 **do**

$y \leftarrow (a_i + y)/ * b$

end for

Output: y

a_i είναι τα ψηφία του αριθμού x.

Επομένως, για τον αριθμό $x = (.53)_8$ θα έχουμε

i	y
-	0
-2	$(3+0)/8 = 0.375$
-1	$(0.375+5)/8 = 0.671875$

δηλαδή, $y = (0.671875)_{10}$.

Μετατροπή κλασματικού x από βάση b σε δεκαδικό σύστημα

- Συνάρτηση σε MATLAB

```
function y=b2dec_f(x,b)
xc=num2str(x);
n=length(xc)-2;
for i=1:n
     a(i)=str2num(xc(i+2));
end
y=0;
for i=n:-1:1
     y=(a(i)+y)*(1/b);
end
```

Αριθμητικά συστήματα

Algorithm 4 Μετατροπή ακεραίου x από δεκαδικό σύστημα σε βάση b (Αλγόριθμος της Διαίρεσης)

Input: $x \in \mathbb{Z}$, b
  $i \leftarrow 0$
  while $x \neq 0$ **do**
    $a_i \leftarrow x \mod b$
    $x \leftarrow [x/b]$
    $i \leftarrow i + 1$
  end while
Output: a_i

Επομένως, για τον αριθμό $x = (369)_{10}$ σε βάση $b = 8$, θα έχουμε

i	a_i	x
0	1	46
1	6	5
2	5	0

δηλαδή, $a_0 = 1$, $a_1 = 6$, $a_2 = 5$ ή ισοδύναμα $(561)_8$

a_i είναι τα ψηφία του αριθμού που μετατράπηκε.

Αριθμητικά συστήματα

Algorithm 5 Μετατροπή κλασματικού x από δεκαδικό σύστημα σε βάση b

Input: $x \in \mathbb{Z}$, b
  $y \leftarrow b * x$
  $i \leftarrow -1$
  while $y \neq 0$ **do**
    $a_i \leftarrow \lfloor y \rfloor$
    $y \leftarrow (y - \lfloor y \rfloor) * b$
    $i \leftarrow i - 1$
  end while
Output: a_i

Επομένως, για τον αριθμό $x = (.875)_{10}$ σε βάση $b = 2$, θα έχουμε

i	a_i	y
-	-	1.75
-1	1	1.5
-2	1	1
-3	1	0

δηλαδή, $a_{-1} = 1$, $a_{-2} = 1$, $a_{-3} = 1$ ή ισοδύναμα $(.111)_2$

a_i είναι τα ψηφία του αριθμού που μετατράπηκε.

Μετατροπή κλασματικού από δεκαδικό σύστημα σε βάση b

- Στη μετατροπή **πεπερασμένου κλασματικού δεκαδικού** σε βάση b ένας αριθμός μπορεί να μετατραπεί σε αριθμό με άπειρα ψηφία και το αντίστροφο.
- Στην μετατροπή **μη πεπερασμένου κλασματικού δεκαδικού** σε βάση b ένας αριθμός μπορεί να μετατραπεί σε αριθμό με άπειρα ψηφία.

Αριθμητικά συστήματα - Παραδείγματα

- Να μετατραπούν οι ακόλουθοι αριθμοί σε δεκαδική βάση.

$$(1101)_2 = (13)_{10}$$
$$(.11)_2 = (.75)_{10}$$

- Να μετατραπούν οι ακόλουθοι αριθμοί σε δυαδική βάση.

$$(11)_{10} = (1011)_2$$
$$(.372)_{10} = (.01011....)_2$$

Σημαντικά ψηφία

Ορισμός

Σημαντικά ψηφία ενός αριθμού x ονομάζονται όλα τα ψηφία του αριθμού εκτός των μηδενικών ψηφίων τα οποία δεν επηρεάζουν την απαραίτητη πληροφορία του αριθμού. Δηλαδή,

- στους ακεραίους, δεν είναι σημαντικά ψηφία τα μηδενικά που βρίσκονται δεξιά από ένα μη μηδενικό ψηφίο.
- στους δεκαδικούς, δεν είναι σημαντικά ψηφία τα μηδενικά που βρίσκονται αριστερά από ένα μη μηδενικό ψηφίο.
- τα μηδενικά που βρίσκονται ανάμεσα σε μη μηδενικά ψηφία είναι σημαντικά ψηφία.

Σημαντικά ψηφία

- Παραδείγματα (σημαντικά ψηφία)
  - $x_1 = 0.0997$ (3 σημαντικά ψηφία)
  - $x_2 = 0.099700$ (5 σημαντικά ψηφία)
  - $x_3 = 410.7$ (4 σημαντικά ψηφία)
  - $x_4 = 5.70$ (3 σημαντικά ψηφία)
  - $x_5 = 0.0079$ (2 σημαντικά ψηφία)
  - $x_6 = 1100$ (2 σημαντικά ψηφία)
  - $x_5 = 110001$ (6 σημαντικά ψηφία)

Αριθμητική κινητής υποδιαστολής

- Κάθε μη μηδενικός πραγματικός αριθμός x σε αριθμητικό σύστημα με βάση b μπορεί να γραφεί στην κανονική μορφή κινητής υποδιαστολής.

$$x = \pm\left(.d_1 d_2 \cdots\right) \cdot b^e, \quad \text{με} \quad d_1 \neq 0$$

- Για παράδειγμα
  - $-(.00598)_{10} = -.598 \times 10^{-2}$
  - $(111.001)_2 = .111001 \times 2^3$
  - $(11100)_{10} = .111 \times 10^5$

Αριθμητική κινητής υποδιαστολής

- Οι αριθμοί που αποθηκεύονται στην μνήμη του Η/Υ πρέπει να είναι πεπερασμένοι[1].
- Οι **αριθμοί κινητής υποδιαστολής** καθορίζουν την αριθμητική ακρίβεια των υπολογισμών και κατά συνέπεια την αποτελεσματικότητα των αλγορίθμων.
- Στις περισσότερες γλώσσες προγραμματισμού οι αριθμοί κινητής υποδιαστολής αναφέρονται ως float, double κ.α.

[1] Να μην έχουν άπειρα ψηφία

Αριθμητική κινητής υποδιαστολής

- Σύστημα αριθμών κινητής υποδιαστολής - Χαρακτηριστικά
  - Η βάση του συστήματος b.
  - Η ακρίβεια t (ή mantissa), δηλαδή το πλήθος των δεκαδικών ψηφίων των αριθμών.
  - Το κάτω φράγμα L και το άνω φράγμα U του εκθέτη e της βάσης b (L, U ακέραιοι με $L \cong -U$).
  - Η μορφή των αριθμών

$$x = \pm\,(.d_1 d_2 \cdots) \cdot b^e, \quad \text{με} \quad d_1 \neq 0$$

Αριθμητική κινητής υποδιαστολής

Τα πρότυπα αριθμών κινητής υποδιαστολής που χρησιμοποιεί η **IEEE** είναι τα παρακάτω:

IEEE	**b**	**t**	**L**	**U**	b^{1-t}
simple	2	24	-125	128	1.2×10^{-7}
double	2	53	-1021	1024	2.2×10^{-16}
extended	2	64	-16381	16384	1.2×10^{-19}

- Στην απλή ακρίβεια, 24 ψηφία είναι για την mastissa, 1 ψηφίο για το πρόσημο και 7 ψηφία για τον εκθέτη.
- Στην διπλή ακρίβεια, 53 ψηφία είναι για την mastissa, 1 ψηφίο για το πρόσημο και 10 ψηφία για τον εκθέτη.

Αριθμητική κινητής υποδιαστολής

- Σύστημα αριθμών κινητής υποδιαστολής με (b, t, L, U)
  - Μέγιστο θετικό στοιχείο

    $$d_i = b - 1, \quad 1 \leq i \leq t, \quad e = U$$

    ή ισοδύναμα

    $$d_1 d_2 d_3 \ldots d_t \times b^U$$

    δηλαδή, σε σύστημα με $(b, t, L, U) = (10, 3, -5, 5)$ θα έχουμε

    $$x_{max} = (.999) \times 10^5$$

  - Ελάχιστο θετικό στοιχείο

    $$.100 \ldots 0 \times b^L$$

    δηλαδή, σε σύστημα με $(b, t, L, U) = (10, 3, -5, 5)$ θα έχουμε

    $$x_{min} = (.100) \times 10^{-5}$$

Αριθμητική κινητής υποδιαστολής

- Σύστημα αριθμών κινητής υποδιαστολής με (b, t, L, U)
  - Υπερχείλιση (overflow)
    Όταν έχουμε θετικό αριθμό μεγαλύτερο από τον μεγαλύτερο θετικό αριθμό του συστήματος (b, t, L, U). Συνήθως οι αριθμοί αυτοί αντικαθίστανται από το άπειρο ή από τον μεγαλύτερο θετικό αριθμό του συστήματος ή προκαλούν πρόβλημα στο λογισμικό.
  - Υπεκχείλιση (underflow)
    Όταν έχουμε θετικό αριθμό μικρότερο μεγαλύτερο από τον μικρότερο θετικό αριθμό του συστήματος (b, t, L, U). Συνήθως οι αριθμοί αυτοί αντικαθίστανται από το μηδέν ή από τον μικρότερο θετικό αριθμό του συστήματος (το μηδέν της μηχανής ή το έψιλον) ή προκαλούν πρόβλημα στο λογισμικό.

Αριθμητική κινητής υποδιαστολής Ι

Ανοχή του συστήματος αριθμών κινητής υποδιαστολής.

- Οι πράξεις με αριθμούς έξω από τα όρια της ανοχής χάνουν πληροφορίες (Σημαντικά Ψηφία).
- Το Matlab ορίζει όλες της αριθμητικές μεταβλητές ως double.
- Κάτω όριο ανοχής του Matlab, το υπολογίζουμε με τον παρακάτω κώδικα

```
e=1;
while e+1>1
    e=e/2;
end
e
```

το οποίο μας επιστρέφει

Αριθμητική κινητής υποδιαστολής II

```
e=
   1.11022302462516e-016
```

- Υπολογίζουμε τα δυαδικά ψηφία του e

```
>> log2(e)

ans=
      -53
```

- Παραδείγματα ανοχής

Αριθμητική κινητής υποδιαστολής III

```
>> e/2

ans =
      5.55111512312578e-017

>> e+1

ans =
      1
```

- Άνω όριο ανοχής του Matlab, το υπολογίζουμε με τον παρακάτω κώδικα

Αριθμητική κινητής υποδιαστολής IV

```
E=1
while E+1>E
    E=E*2;
end
E
```

το οποίο μας επιστρέφει

```
E=
   9.00719925474099e+015
```

- Υπολογίζουμε τα δυαδικά ψηφία του E

Αριθμητική κινητής υποδιαστολής V

```
>> log2(E)
ans=
     53
```

- Παραδείγματα ανοχής

```
>> E+1

ans =
     9.00719925474099e+015

>> 2*E

ans =
       1.8014398509482e+016
```

Αριθμητική κινητής υποδιαστολής VI

Αριθμητική κινητής υποδιαστολής

- Σύστημα αριθμών κινητής υποδιαστολής με (b, t, L, U)
  - Στρογγύλευση $fl(.)$
  - Έστω ο αριθμός

$$x = \pm(.d_1 d_2 \dots d_t d_{t+1} \dots) \cdot b^e \quad \text{με} \quad d_1 \neq 0$$

θα γίνει

$$fl(x) = \pm(.d_1 d_2 .... d'_t) \cdot b^e \quad \text{με} \quad d_1 \neq 0$$

με

$$d'_t = \begin{cases} d_t, & \text{αν } d_{t+1} < 5 \\ d_{t+1}, & \text{αν } d_{t+1} \geq 5 \end{cases}$$

  - Δηλαδή, κάνουμε στρογγυλοποίηση στο t-οστο ψηφίο.

Αριθμητική κινητής υποδιαστολής

- Σύστημα αριθμών κινητής υποδιαστολής με (b, t, L, U)
    - Αποκοπή $fl(.)$
    - Έστω ο αριθμός

      $$x = \pm(.d_1 d_2 \ldots d_t d_{t+1} \ldots) \cdot b^e \quad \text{με} \quad d_1 \neq 0$$

      θα γίνει

      $$fl(x) = \pm(.d_1 d_2 \ldots . d_t) \cdot b^e \quad \text{με} \quad d_1 \neq 0$$

    - Δηλαδή, αποκόπτουμε τα ψηφία μετά το t-οστο ψηφίο.
- Στα πλαίσια του μαθήματος, για μεγαλύτερη ακρίβεια, στον ορισμό των αριθμών στο σύστημα αριθμών κινητής υποδιαστολής χρησιμοποιούμε την στρογγύλευση.

Αριθμητική κινητής υποδιαστολής

- Σύστημα αριθμών κινητής υποδιαστολής με (b, t, L, U)
  - Πράξεις $fl(fl(x) * fl(y))$.
  - Για την εκτέλεση της πράξης $x * y$ μετατρέπουμε τα x και y στο σύστημα αριθμών κινητής υποδιαστολής $fl(x)$ και $fl(y)$ αντίστοιχα και εκτελούμε την πράξη $fl(x) * fl(y)$.
  - Το αποτέλεσμα της παραπάνω πράξης δεν είναι στο σύστημα αριθμών κινητής υποδιαστολής (υποθέτουμε ότι δεν έχουμε υπερχείλιση ή υπεκχείλιση) και το μετατρέπουμε στο σύστημα αριθμών κινητής υποδιαστολής $fl(fl(x) * fl(y))$.

Αριθμητική κινητής υποδιαστολής - Παράδειγμα 1

- Έστω το σύστημα αριθμών κινητής υποδιαστολής $(b, t, L, U) = (10, 5, -10, 10)$ και οι αριθμοί $x = 5891.26$ και $y = 0.0773414$
    - Να βρεθούν ο μεγαλύτερος και ο μικρότερος θετικός αριθμός του συστήματος
    - Να βρεθεί το άθροισμα $x + y$ στο σύστημα αριθμών κινητής υποδιαστολής με στρογγύλευση και με αποκοπή.

Στο σύστημα $(b, t, L, U) = (10, 5, -10, 10)$ θα έχουμε

$$Max = (.99999) \times 10^{10} = 9999900000$$

και

$$Min = (.10000) \times 10^{-10} = 0.0000000001$$

Αριθμητική κινητής υποδιαστολής - Παράδειγμα 1

Στο σύστημα $(b, t, L, U) = (10, 5, -10, 10)$ με στρογγύλευση, θα έχουμε

$$fl(x) = (.58913) \times 10^4, \quad fl(y) = (.77341) \times 10^{-1}$$

άρα

$$fl(x) + fl(y) = 5891.377341$$

επομένως

$$fl\big(fl(x) + fl(y)\big) = (.58914) \times 10^4 = 5891.4$$

με πραγματικό αποτέλεσμα $x + y = 5891.3373414$.

Αριθμητική κινητής υποδιαστολής - Παράδειγμα 1

Στο σύστημα $(b, t, L, U) = (10, 5, -10, 10)$ με αποκοπή, θα έχουμε

$$fl(x) = (.58912) \times 10^4, \quad fl(y) = (.77341) \times 10^{-1}$$

άρα

$$fl(x) + fl(y) = 5891.277341$$

επομένως

$$fl\big(fl(x) + fl(y)\big) = (.58912) \times 10^4 = 5891.2$$

με πραγματικό αποτέλεσμα $x + y = 5891.3373414$.

Αριθμητική κινητής υποδιαστολής - Παράδειγμα 2

- Έστω το σύστημα αριθμών κινητής υποδιαστολής $(b, t, L, U) = (10, 5, -10, 10)$ και οι αριθμοί $a = 1$, $b = 0.00003$ και $c = 0.00003$. Να γίνουν οι πράξεις:
    - $a + (b + c)$
    - $(a + b) + c$

Στο σύστημα[2] $(b, t, L, U) = (10, 5, -10, 10)$ θα έχουμε

$$fl(a) = (.1) \times 10^{1}, \quad fl(b) = (.3) \times 10^{-4}, \quad fl(c) = (.3) \times 10^{-4}$$

άρα, για το άθροισμα $a + (b + c)$

$$fl(b) + fl(c) = 0.00006 \quad \text{με} \quad fl\big(fl(b) + fl(c)\big) = (.6) \times 10^{-4}$$

[2] Στα πλαίσια του μαθήματος, για μεγαλύτερη ακρίβεια, στον ορισμό των αριθμών στο σύστημα αριθμών κινητής υποδιαστολής χρησιμοποιούμε την στρογγύλευση.

Αριθμητική κινητής υποδιαστολής - Παράδειγμα 2

επομένως

$$fl(a) + fl\big(fl(b) + fl(c)\big) = 1.00006$$

και τελικά,

$$fl\Big(fl(a) + fl\big(fl(b) + fl(c)\big)\Big) = (.10001) \times 10^1$$

Ενώ, για το άθροισμα $(a + b) + c$

$$fl(a) + fl(b) = 1.00003 \quad \text{με} \quad fl\big(fl(a) + fl(b)\big) = (.1) \times 10^1$$

επομένως

$$fl\big(fl(a) + fl(b)\big) + fl(c) = 1.00003$$

και τελικά,

$$fl\Big(fl\big(fl(a) + fl(b)\big) + fl(c)\Big) = (.1) \times 10^1$$