ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ
Εθνικόν και Καποδιστριακόν
Πανεπιστήμιον Αθηνών

Αριθμητική Ανάλυση

Ενότητα 1

Σφάλματα στους Αριθμητικούς Υπολογισμούς

Ν. Μ. Μισυρλής

Τμήμα Πληροφορικής και Τηλεπικοινωνιών,

Αριθμητική Ανάλυση (ή Υπολογιστικά Μαθηματικά)

- Είναι η κοινή περιοχή των Μαθηματικών και της Πληροφορικής.
- Η Αριθμητική Ανάλυση έχει σαν βασικό σκοπό την σχεδίαση και ανάλυση αριθμητικών αλγορίθμων για την επίλυση επιστημονικών προβλημάτων.
- Τα προβλήματα αυτά προκύπτουν από την μαθηματική μοντελοποίηση αντιστοίχων προβλημάτων του πραγματικού κόσμου και ενδιαφέρουν τις Επιστήμες (Φυσική, Οικονομία, Κοινωνία, Ιατρική, Βιολογία, Επιχειρησιακή Έρευνα, κ.α.) και την Τεχνολογία.

- Όσο αφορά την Πληροφορική η Αριθμητική Ανάλυση αποτελεί ένα αυτοδύναμο κλάδο της.
- Το πρόβλημα της εύρεσης του βαθμού μιας σελίδας στο διαδίκτυο (page rank problem) καταλήγει στον υπολογισμό ενός ιδιοδιανύσματος με τη χρήση της μεθόδου των δυνάμεων.
- Η εύρεση του χρωματικού αριθμού ενός γραφήματος συνδέεται, όπως αποδείχτηκε πρόσφατα, με τη φασματική ακτίνα του πίνακα γειτνίασης, η οποία επίσης υπολογίζεται με τη μέθοδο των δυνάμεων.
- Η συνεκτικότητα ενός γραφήματος αποτελεί σημαντική ιδιότητα, η οποία είναι ισοδύναμη με το βαθμό (rank) του πίνακα πρόσπτωσης (incident matrix).

Πολλά προβλήματα της Πληροφορικής μοντελοποιούνται σαν προβλήματα γραμμικής ή μη γραμμικής βελτιστοποίησης.

- Όσο αφορά τη γραμμική βελτιστοποίηση η χρήση της Simplex είναι μια παραλλαγή της αμέσου μεθόδου του Jordan για την επίλυση ενός γραμμικού συστήματος.
  - Πιο πρόσφατες έρευνες προτείνουν τη μέθοδο της LU παραγοντοποίησης στην περίπτωση που ο πίνακας του συστήματος είναι αραιός.
- Αρκετά προβλήματα της μη γραμμικής βελτιστοποίησης ανάγονται σε ένα μη γραμμικό σύστημα, το οποίο επιλύεται με παραλλαγές της μεθόδου του Newton ή της μεθόδου Συζυγών Διευθύνσεων (Conjugate Gradient).
- Επίσης, τα Γραφικά βασίζονται σε μεθόδους παρεμβολής, όπως οι splines.
- Η τριγωνομετρική παρεμβολή με τη χρήση του Διακριτού Μετασχηματισμού Fourier (FFT) έχει πολλές εφαρμογές στην επεξεργασία σήματος, στην κβαντική μηχανική, στην οπτική καθώς και σε πολλές άλλες περιοχές.
- Οι αλυσίδες Markov (Markov Chains) που προκύπτουν κατά την προσομοίωση των δικτύων ουρών (queueing network) καταλήγουν σε γραμμικά συστήματα, τα οποία επιλύονται με τις αμέσους ή επαναληπτικές μεθόδους της Αριθμητικής Ανάλυσης.

Μια χαρακτηριστική εφαρμογή είναι το πρόβλημα εξισορρόπησης φορτίου που προκύπτει κατά την παράλληλη επεξεργασία.

- Για την επίλυση του ανωτέρω προβλήματος χρησιμοποιείται η μέθοδος της Διάχυσης (Diffusion), η μελέτη της οποίας ακολουθεί την ίδια προσέγγιση με αυτή των επαναληπτικών μεθόδων της Αριθμητικής Ανάλυσης.

Υπολογιστική Επιστήμη

- Τα τελευταία χρόνια δημιουργήθηκε η περιοχή της Υπολογιστικής Επιστήμης ή των Επιστημονικών Υπολογισμών, η οποία χρησιμοποιεί την αριθμητική προσομοίωση με ό,τι πιο σύγχρονο αφορά στο περιβάλλον (γραφικά) και στην υλοποίησή τους (αντικειμενοστρεφής προγραμματισμός). Μάλιστα, ήδη έχουν δημιουργηθεί και ολόκληρα τμήματα σε Πανεπιστήμια στην εν λόγω περιοχή.

Σκοπός του μαθήματος

- Ανάπτυξη και μελέτη των βασικών αριθμητικών μεθόδων για την επίλυση επιστημονικών προβλημάτων
- Μελέτη σφάλματος (ακρίβεια λύσεων)

Στόχοι του μαθήματος

Με την ολοκλήρωση του μαθήματος οι φοιτητές να γνωρίζουν

- την ανάπτυξη και υλοποίηση αριθμητικών αλγορίθμων για την επίλυση επιστημονικών προβλημάτων
- την σύγχρονη μεθοδολογία αξιολόγησης και σύγκρισης επίδοσης αριθμητικών αλγορίθμων
- τις σύγχρονες τάσεις στην περιοχή των Επιστημονικών Υπολογισμών
- την σύγχρονη ανάπτυξη επιστημονικού λογισμικού για την προσομοίωση προβλημάτων του φυσικού μας κόσμου.

Περιεχόμενα του μαθήματος

1. Σφάλματα στους Αριθμητικούς Υπολογισμούς
2. Αριθμητική Επίλυση μη Γραμμικών Εξισώσεων.
3. Άμεσοι μέθοδοι για την Επίλυση Γραμμικών Συστημάτων.
4. Επαναληπτικές μέθοδοι για την Επίλυση Γραμμικών Συστημάτων.
5. Αριθμητικός υπολογισμός Ιδιοτιμών και Ιδιοδιανυσμάτων.
6. Παρεμβολή
7. Μέθοδος των Ελαχίστων Τετραγώνων
8. Αριθμητική Παραγώγιση
9. Αριθμητική Ολοκλήρωση
10. Αριθμητική Επίλυση Συνήθων Διαφορικών Εξισώσεων

Σφάλματα στους Αριθμητικούς Υπολογισμούς

- Εισαγωγή
- Αριθμοί Μηχανής
- Ανάλυση σφάλματος των αριθμών κινητής υποδιαστολής
- Ανάλυση σφάλματος στο άθροισμα όρων
- Διαδιδόμενο σφάλμα

Πηγές Σφαλμάτων

1. *Σφάλματα που προκύπτουν κατά το σχηματισμό του μαθηματικού μοντέλου*
2. *Σφάλματα στα δεδομένα*
3. *Σφάλματα αποκοπής*
4. *Σφάλματα στρογγύλευσης*

Μέτρηση Σφάλματος

- **απόλυτο σφάλμα**
- **απόλυτο σχετικό σφάλμα**

Ορισμός

Αν $\bar{x}$ είναι μια προσέγγιση του x, το **απόλυτο σφάλμα** είναι η ποσότητα

$$|x - \bar{x}|$$

και το **απόλυτο σχετικό σφάλμα** η ποσότητα

$$\frac{|x - \bar{x}|}{|x|}, \quad x \neq 0.$$

Αριθμοί Μηχανής

- **Ακέραιοι**
- **Πραγματικοί-Κινητή Υποδιαστολή**

Σύστημα Αρίθμησης

Αν $\beta = \mathbf{10} \rightarrow$ Δεκαδικό

$$(432.52)_{10} = 4 \cdot 10^2 + 3 \cdot 10^1 + 2 \cdot 10^0 + 5 \cdot 10^{-1} + 2 \cdot 10^{-2}$$

Αν $\beta = \mathbf{2} \rightarrow$ Δυαδικό

$$(101.11)_2 = 1 \cdot 2^2 + 0 \cdot 2^1 + 1 \cdot 2^0 + 1 \cdot 2^{-1} + 1 \cdot 2^{-2} = 5.75$$

ή

$$(101.11)_2 = (5.75)_{10}$$

όπου ο συμβολισμός $(\cdot)_\beta$ δηλώνει τη βάση β του συστήματος αρίθμησης στο οποίο παριστάνεται ο αριθμός.

Ακέραιοι

Αν υποτεθεί ότι διατίθενται **n** δυαδικά ψηφία (**bits**) για την αποθήκευση των ψηφίων ενός ακεραίου αριθμού, τότε ο μεγαλύτερος ακέραιος αριθμός που μπορεί να αποθηκευθεί στη μνήμη είναι ο:

$$(\overbrace{111\ldots1}^{\mathbf{n}})_{\mathbf{2}} = \mathbf{1}\cdot\mathbf{2}^{\mathbf{n}-\mathbf{1}} + \mathbf{1}\cdot\mathbf{2}^{\mathbf{n}-\mathbf{2}} + \cdots + \mathbf{1}\cdot\mathbf{2}^{\mathbf{0}} = \mathbf{2}^{\mathbf{n}} - \mathbf{1}.$$

Αν $\mathbf{n} = \mathbf{15}$ τότε $\mathbf{2}^{\mathbf{15}} - \mathbf{1} = \mathbf{32.767}$.

Επομένως, όλοι οι ακέραιοι αριθμοί στην περιοχή $[-(\mathbf{2}^{\mathbf{n}} - \mathbf{1}),\ \mathbf{2}^{\mathbf{n}} - \mathbf{1}]$ αποθηκεύονται με την ακριβή τιμή τους στη μνήμη.

Πραγματικοί Αριθμοί - Κινητή Υποδιαστολή

Στο δεκαδικό σύστημα ένας πραγματικός αριθμός μπορεί να παρασταθεί στην **κανονικοποιημένη επιστημονική μορφή**.

Αυτό σημαίνει ότι η δεκαδική τελεία μετατοπίζεται έτσι ώστε όλα τα ψηφία του αριθμού να βρίσκονται στα δεξιά της δεκαδικής τελείας και το πρώτο ψηφίο να είναι διάφορο του μηδενός.

Για παράδειγμα,

$$15.546 = 0.15546 \cdot 10^2$$

Πραγματικοί Αριθμοί - Κινητή Υποδιαστολή

Ένας πραγματικός αριθμός x $(\neq 0)$ μπορεί να παρασταθεί με τη μορφή

$$x = \pm \bar{x} \cdot 10^e$$

όπου e είναι ένας ακέραιος, ο οποίος καλείται **εκθέτης** (exponent) και $\bar{x}$ είναι το **κλασματικό τμήμα** του αριθμού (ή **mantissa**).

Είναι φανερό ότι

$$0.1 \leq \bar{x} < 1$$

Γενικά, ένας αριθμός παριστάνεται σε ένα σύστημα αρίθμησης με βάση β σαν

$$x = \pm \bar{x} \cdot \beta^e$$

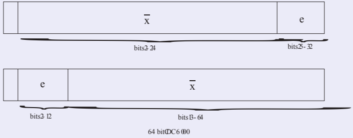
όπου

$$\bar{x} = (0.a_1 a_2 \ldots a_n)_\beta$$

$$0 \leq a_i \leq \beta - 1, \quad a_1 \neq 0$$

Παράσταση των Πραγματικών Αριθμών με Κινητή Υποδιαστολή

Στο ακόλουθο σχήμα παριστάνονται οι πραγματικοί αριθμοί σε υπολογιστές με λέξεις των 32 και 64 δυαδικών ψηφίων (**bits**).



Σχήμα: Παράσταση πραγματικού αριθμού στη μνήμη.

Το πρώτο δυαδικό ψηφίο (bit) είναι 0 αν ο αριθμός είναι θετικός και 1 αν είναι αρνητικός.

Αριθμοί Μηχανής

Η παράσταση

$$x = \pm \bar{x} \cdot \beta^{e}$$

καλείται **κινητής υποδιαστολής** (**floating point**).

Στη συνέχεια προσδιορίζεται το διάστημα $[\mathbf{s},\ \mathbf{L}]$ των πραγματικών αριθμών που μπορούν να αποθηκευθούν *ακριβώς* στη μνήμη.

Υποθέτουμε ότι $\beta = \mathbf{2}$, το **x** είναι αποθηκευμένο σαν μια ακολουθία από **n** δυαδικά ψηφία και

$$|e| \leq M$$

Η ποσότητα $\bar{x}$ φράσσεται ως εξής

$$(0.\overbrace{10\ldots0}^{n})_2 \leq \bar{x} \leq (0.\overbrace{11\ldots1}^{n})_2$$

ή

$$\frac{1}{2} \leq \bar{x} \leq 1 - 2^{-n}$$

οπότε

$$s = \frac{1}{2} \cdot 2^{e} \leq \bar{x} \cdot 2^{e} = |x| \leq \left(1 - 2^{-n}\right) \cdot 2^{e} = L < 2^{e}$$

ή

$$2^{e-1} \leq |x| < 2^{e}$$

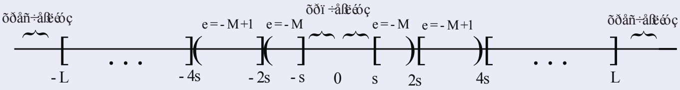
Αριθμοί Μηχανής

Επειδή το $\bar{x}$ είναι της μορφής

$$\bar{x} = (0.1a_2a_3 \ldots a_n)_2$$

για κάθε $e = -M, -M+1, \ldots, M-1, M$, υπάρχουν 2^{n-1} διαφορετικά κανονικοποιημένα $\bar{x}$.

Αυτά τα $\bar{x}$ αντιστοιχούν σε 2^{n-1} ίσης απόστασης αριθμούς x σε κάθε ένα από τα διαστήματα $[2^{e-1}, 2^e)$ και $(-2^e, -2^{e-1}]$ (βλ. ακόλουθο Σχήμα).



Σχήμα: Παράσταση αριθμών μηχανής.

Αριθμοί Μηχανής-Παρατηρήσεις

- Όταν αυξάνεται ο εκθέτης κατά 1 διπλασιάζεται το μήκος και των δύο διαστημάτων $[2^{e-1}, 2^e)$ και $(-2^e, -2^{e-1}]$. Συνεπώς :

Κατανομή αριθμών μηχανής

Οι παραστάσιμοι αριθμοί είναι πυκνά κατανεμημένοι πλησίον του μηδενός και αραιά κατανεμημένοι μακρυά του μηδενός. Οι αριθμοί αυτοί είναι κινητής υποδιαστολής και καλούνται **αριθμοί μηχανής**.

- Υπάρχει λοιπόν μόνο ένα πεπερασμένο σύνολο πραγματικών αριθμών που μπορούν να αποθηκευθούν στη μνήμη με την *ακριβή* τιμή τους και αυτό βρίσκεται στα δύο διαστήματα $[-L, -s]$ και $[s, L]$.

Overflow-Underflow

Οποιοσδήποτε αριθμός x για τον οποίο ισχύει $|x| > L$ δεν μπορεί να αποθηκευτεί στη μνήμη και το φαινόμενο αυτό είναι γνωστό σαν **υπερχείλιση** (**overflow**). Όμοια αν $|x| < s$ τότε έχουμε το φαινόμενο της **υποχείλισης**

Αριθμοί Μηχανής-Παρατηρήσεις

Μεταβολή στην κατανομή των αριθμών μηχανής

Η χρήση(ή διάθεση) περισσοτέρων δυαδικών ψηφίων για την παράσταση της **mantissa** $\bar{x}$ αυξάνει την πυκνότητα των αριθμών μηχανής, ενώ για την παράσταση του εκθέτη **e** έχει σαν αποτέλεσμα την αύξηση του διαστήματος παράστασής τους.

Παράδειγμα

Αν $\beta = \mathbf{2},\ \mathbf{n} = \mathbf{3},\ \mathbf{m} = -\mathbf{1},\ \mathbf{M} = \mathbf{2}$, να βρεθούν και να παρασταθούν οι αριθμοί μηχανής.

Λύση

Οι αριθμοί μηχανής έχουν τη μορφή

$$x = \pm\bar{x} \cdot \beta^{e}$$

όπου

$$\bar{x} = (0.a_1 a_2 \dots a_n)_\beta$$

ή για τα δεδομένα του παραδείγματος

$$x = \pm\bar{x} \cdot 2^{e}, \quad \bar{x} = (0.1a_2a_3)_2$$

με $0 \le a_i \le 1,\ i = 2,\ 3.$
Ο μικρότερος $\bar{x}$ είναι ο αριθμός $(0.100)_2$, ενώ οι επόμενοι λαμβάνονται αν κάθε φορά προστίθεται ο αριθμός $(0.001)_2$.

.....Παράδειγμα.....

Για δεδομένο e έχουμε τους εξής τέσσερις αριθμούς $\bar{x}$:

$$0.100, \;\; 0.101, \;\; 0.110, \;\; 0.111$$

άρα για

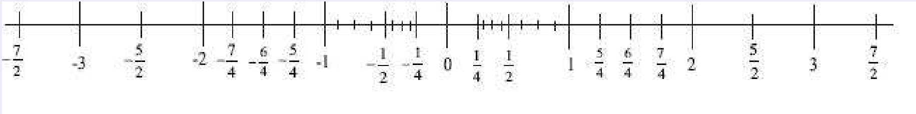
$$e = -1, \; 0, \; 1, \; 2$$

λαμβάνουμε τους ακόλουθους θετικούς αριθμούς μηχανής:

$(0.100)_2 \cdot 2^{-1}$ $(\frac{1}{4})$	$(0.101)_2 \cdot 2^{-1}$ $(\frac{5}{16})$	$(0.110)_2 \cdot 2^{-1}$ $(\frac{6}{16})$	$(0.111)_2 \cdot 2^{-1}$ $(\frac{7}{16})$
$(0.100)_2 \cdot 2^{0}$ $(\frac{1}{2})$	$(0.101)_2 \cdot 2^{0}$ $(\frac{5}{8})$	$(0.110)_2 \cdot 2^{0}$ $(\frac{6}{8})$	$(0.111)_2 \cdot 2^{0}$ $(\frac{7}{8})$
$(0.100)_2 \cdot 2^{1}$ (1)	$(0.101)_2 \cdot 2^{1}$ $(\frac{5}{4})$	$(0.110)_2 \cdot 2^{1}$ $(\frac{6}{4})$	$(0.111)_2 \cdot 2^{1}$ $(\frac{7}{4})$
$(0.100)_2 \cdot 2^{2}$ (2)	$(0.101)_2 \cdot 2^{2}$ $(\frac{5}{2})$	$(0.110)_2 \cdot 2^{2}$ $(\frac{6}{2})$	$(0.111)_2 \cdot 2^{2}$ $(\frac{7}{2})$

.....Παράδειγμα

Η γραφική παράσταση των αριθμών είναι η ακόλουθη:



Σχήμα: Γραφική παράσταση των αριθμών μηχανής

Παρατηρούμε ότι δεν υπάρχουν αριθμοί στα διαστήματα $\left(-\frac{1}{4},\ 0\right)$ και $\left(0,\ \frac{1}{4}\right)$.

Επίσης, οι αριθμοί δεν είναι ομοιόμορφα κατανεμημένοι. Ωστόσο οι αριθμοί που έχουν κοινό εκθέτη απέχουν ίση απόσταση μεταξύ τους.

Ανάλυση σφάλματος των αριθμών κινητής υποδιαστολής

Έστω ο πραγματικός αριθμός κινητής υποδιαστολής

$$x = \pm (0.a_1 a_2 \ldots a_n a_{n+1} \ldots)_\beta \cdot \beta^e, \quad a_1 \neq 0 \tag{1}$$

όπου (β=2, 8, 10, 16). Εάν το μέγιστο πλήθος ψηφίων που μπορούν να αποθηκευτούν είναι n, τότε ο αριθμός αυτός δεν μπορεί να παρασταθεί στη μνήμη.

Το ερώτημα που τίθεται είναι το εξής: Ποιός είναι ο πλησιέστερος αριθμός μηχανής προς τον x;

Αν θεωρήσουμε δύο διαδοχικούς αριθμούς μηχανής μεταξύ των οποίων βρίσκεται ο x, τότε συμβαίνει μία από τις ακόλουθες δύο περιπτώσεις :

Σχήμα: Δύο πιθανές θέσεις του x (x', x'' αριθμοί μηχανής).

....Ανάλυση σφάλματος....

Σχήμα: Δύο πιθανές θέσεις του x (x', x'' αριθμοί μηχανής).

Στη περίπτωση (a), ο πλησιέστερος αριθμός μηχανής προς τον x είναι ο x' και βρίσκεται αν αποκοπούν τα ψηφία $a_{n+1}\ldots$ από το κλασματικό τμήμα του, δηλαδή είναι ο

$$x' = (0.a_1 a_2 \ldots a_n)_\beta \cdot \beta^e. \tag{2}$$

Ενώ στην περίπτωση (b), ο πλησιέστερος προς τον x είναι ο x'' και βρίσκεται αν στον x' προστεθεί η ποσότητα

$$(0.00\ldots01)_\beta = \beta^{-n}$$

, δηλαδή είναι ο

$$x'' = \left((0.a_1 a_2 \ldots a_n)_\beta + \beta^{-n}\right) \cdot \beta^e. \tag{3}$$

....Ανάλυση σφάλματος....

Η τεχνική αυτή καλείται **στρογγύλευση** (rounding up). Με την **στρογγύλευση** ένας πραγματικός αριθμός αντιστοιχεί σε ένα αριθμό μηχανής. Έτσι λοιπόν ένας πραγματικός αριθμός παριστάνεται (προσεγγίζεται) στη μνήμη με τον αντίστοιχο αριθμό μηχανής.

Αν εφαρμόσουμε την στρογγύλευση για την (a) περίπτωση στο σχήμα, έχουμε ότι το απόλυτο σφάλμα φράσσεται ως εξής

$$\left|x - x'\right| \leq \frac{1}{2}\left|x'' - x'\right| = \left(\frac{1}{2}\beta^{-n}\right) \cdot \beta^{e} \tag{4}$$

ενώ για το απόλυτο σχετικό σφάλμα θα έχουμε

$$\frac{|x - x'|}{|x|} \leq \frac{\frac{1}{2}\beta^{-n} \cdot \beta^{e}}{\overline{x} \cdot \beta^{e}} = \frac{\frac{1}{2}\beta^{-n}}{\overline{x}} \leq \frac{\frac{1}{2}\beta^{-n}}{\beta^{-1}} = \frac{1}{2}\beta^{-n+1}. \tag{5}$$

Εύκολα διαπιστώνεται ότι τα ίδια ισχύουν και για την περίπτωση (b) του σχήματος.

....Ανάλυση σφάλματος....

Συνήθως ο αριθμός μηχανής κινητής υποδιαστολής που είναι πλησιέστερος προς τον x συμβολίζεται με $fl\left(x\right)$. Με το συμβολισμό αυτό οι ανωτέρω τύποι γράφονται

$$\left|x - fl\left(x\right)\right| \leq \frac{1}{2}\beta^{-n} \cdot \beta^{e} \tag{6}$$

και

$$\frac{\left|x - fl\left(x\right)\right|}{\left|x\right|} \leq \frac{1}{2}\beta^{-n+1}, \tag{7}$$

αντίστοιχα.

Οι ανωτέρω τύποι (6) και (7) μας δίνουν **άνω φράγματα** του σφάλματος στρογγύλευσης.

Ανάλυση σφάλματος στρογγύλευσης

Αν τεθεί

$$\varepsilon = \frac{fl\left(x\right) - x}{x},$$

ή

$$fl\left(x\right) = \left(1 + \varepsilon\right)x \tag{8}$$

τότε έχουμε

$$\left|\varepsilon\right| \leq \frac{1}{2}\beta^{-n+1}. \tag{9}$$

Το φράγμα $u = \frac{1}{2}\beta^{-n+1}$ της ποσότητας $\left|\varepsilon\right|$ καλείται μονάδα σφάλματος στρογγύλευσης (ή μονάδα μηχανής).

Η ε είναι *΄μικρή΄* και η (8) δηλώνει ότι ο $fl\left(x\right)$ είναι μια ελαφρά διατάραξη του x.

Ανάλυση σφάλματος με αποκοπή

Μία άλλη τεχνική απεικόνισης των πραγματικών αριθμών στο σύνολο των αριθμών μηχανής είναι αυτή της **αποκοπής** (chopping).

Με την **αποκοπή** ο πραγματικός αριθμός x προσεγγίζεται πάντα με τον πλησιέστερο από τα αριστερά του αριθμού μηχανής, δηλαδή με τον x' (βλ. προηγ. Σχήμα).

Τότε με ανάλογο τρόπο εύκολα βρίσκεται ότι

$$\left|x - fl\left(x\right)\right| \leq \beta^{e-n} \tag{10}$$

και

$$\frac{\left|x - fl\left(x\right)\right|}{|x|} \leq \beta^{-n+1}. \tag{11}$$

Σύγκριση σφάλματος στρογγύλευσης και αποκοπής

1. *Το χειρότερο σφάλμα στρογγύλευσης είναι το μισό εκείνου της αποκοπής.*
2. *Το σφάλμα στη στρογγύλευση είναι αρνητικό στις μισές περίπου περιπτώσεις και θετικό στις άλλες μισές με αποτέλεσμα την απαλοιφή του, ενώ στην αποκοπή έχει συνέχεια το ίδιο πρόσημο.*

Η μελέτη του σφάλματος στρογγύλευσης είναι ένα σημαντικό τμήμα της Αριθμητικής Ανάλυσης και αναπτύχθηκε από τον Wilkinson.

Η αξία της στην μελέτη μιας αριθμητικής μεθόδου είναι αναγκαία και σημαντική, όπως φαίνεται στο ακόλουθο παράδειγμα.

Ανάλυση σφάλματος στρογγύλευσης στο άθροισμα όρων

Έστω ότι επιθυμούμε τον υπολογισμό του αθροίσματος

$$S = \sum_{i=1}^{n} x_i \tag{12}$$

όπου x_i είναι αριθμοί κινητής υποδιαστολής που ήδη έχουν αποθηκευτεί στη μνήμη.

Έτσι, έχουμε

$$\begin{aligned} S_2 &= fl\left(x_1 + x_2\right) \\ S_3 &= fl\left(x_3 + S_2\right) \\ S_4 &= fl\left(x_4 + S_3\right) \\ &\vdots \\ S_n &= fl\left(x_n + S_{n-1}\right) \end{aligned} \tag{13}$$

όπου S_n είναι το αποτέλεσμα του υπολογισμού του S.

Ανάλυση σφάλματος στρογγύλευσης στο άθροισμα όρων

Λόγω της (8), οι ανωτέρω σχέσεις γράφονται

$$
\begin{aligned}
S_2 &= (x_1 + x_2)(1 + \varepsilon_2) \\
S_3 &= (x_3 + S_2)(1 + \varepsilon_3) \\
&\vdots \\
S_n &= (x_n + S_{n-1})(1 + \varepsilon_n)
\end{aligned}
\tag{14}
$$

όπου

$$
|\varepsilon_i| \leq \frac{1}{2}\beta^{-n+1}, \quad i = 2, 3, \cdots, n.
$$

Ανάλυση σφάλματος στρογγύλευσης στο άθροισμα όρων

Αναπτύσσοντας τις πρώτες ποσότητες της (14) έχουμε

$$\begin{aligned}
S_2 &= (x_1 + x_2) + (x_1 + x_2)\,\varepsilon_2 \\
S_3 &= \left[(x_1 + x_2 + x_3) + (x_1 + x_2)\,\varepsilon_2\right](1 + \varepsilon_3) \\
&= (x_1 + x_2 + x_3) + (x_1 + x_2)\,\varepsilon_2 + (x_1 + x_2 + x_3)\,\varepsilon_3 + (x_1 + x_2)\,\varepsilon_2\varepsilon_3
\end{aligned}$$

ή παραλείποντας τον τελευταίο όρο, επειδή $\varepsilon_2\varepsilon_3 \ll \varepsilon_2, \varepsilon_3$, λαμβάνουμε

$$S_3 \simeq (x_1 + x_2 + x_3) + (x_1 + x_2)\,\varepsilon_2 + (x_1 + x_2 + x_3)\,\varepsilon_3.$$

Αναγωγικά βρίσκουμε τελικά ότι

$$S_n \simeq \sum_{i=1}^{n} x_i + (x_1 + x_2)\,\varepsilon_2 + (x_1 + x_2 + x_3)\,\varepsilon_3 + \ldots + (x_1 + x_2 + \ldots + x_n)\,\varepsilon_n$$

ή

$$\begin{gathered}
S_n - S \simeq x_1\,(\varepsilon_2 + \varepsilon_3 + \ldots + \varepsilon_n) + x_2\,(\varepsilon_2 + \varepsilon_3 + \ldots + \varepsilon_n) \\
+ x_3\,(\varepsilon_3 + \varepsilon_4 + \ldots + \varepsilon_n) + \ldots + x_n\varepsilon_n
\end{gathered}$$

ή

$$\begin{gathered}
|S_n - S| \lesssim |x_1|\,(|\varepsilon_2| + \ldots + |\varepsilon_n|) + |x_2|\,(|\varepsilon_2| + \ldots + |\varepsilon_n|) \\
+ |x_3|\,(|\varepsilon_3| + \ldots + |\varepsilon_n|) + \ldots + |x_n|\,|\varepsilon_n|\,.
\end{gathered} \tag{15}$$

Ανάλυση σφάλματος στρογγύλευσης στο άθροισμα όρων

Παρατηρώντας προσεκτικά την (15) και προσπαθώντας να ελαχιστοποιήσουμε το απόλυτο σφάλμα $|S_n - S|$ καταλήγουμε στο συμπέρασμα ότι: οι όροι θα πρέπει να διαταχθούν, πριν τον υπολογισμό του αθροίσματός τους, έτσι ώστε

$$|x_1| \leq |x_2| \leq |x_3| \leq \ldots |x_n| .$$

Υπό την προυπόθεση αυτή, οι όροι στο δεξί μέλος της (15) με το μεγαλύτερο πλήθος σφαλμάτων ε_i θα πολλαπλασιάζονται με τις μικρότερες τιμές μεταξύ των x_i.

Διαδιδόμενο σφάλμα

Για τον έλεγχο της προσέγγισης του $\bar{x}$ σε σχέση με την τιμή του x χρησιμοποιούμε τα ακόλουθα κριτήρια :

- Αν

$$|\varepsilon_{\bar{x}}| = |x - \bar{x}| \leq \frac{1}{2} \cdot 10^{-d}$$

τότε ο $\bar{x}$ προσεγγίζει τον x σε **d δεκαδικά ψηφία**.

Αυτό σημαίνει ότι ο $\bar{x}$ έχει το πολύ **d** και τουλάχιστον **d − 1** **δεκαδικά ψηφία** ακριβή (ίδια με τον x).

●● Αν $x \neq 0$, τότε μπορεί να χρησιμοποιηθεί το σχετικό σφάλμα προκειμένου να εξασφαλισθεί προσέγγιση σε ένα επιθυμητό πλήθος σημαντικών ψηφίων.

Σημαντικά ψηφία ενός δεκαδικού αριθμού είναι όλα τα ψηφία του αριθμού, από αριστερά προς τα δεξιά, του πρώτου μη μηδενικού ψηφίου (συμπεριλαμβανομένου).

Αν

$$|\varrho_{\bar{x}}| = \left|\frac{\varepsilon_{\bar{x}}}{x}\right| \leq 5 \cdot 10^{-s}$$

τότε ο $\bar{x}$ προσεγγίζει τον x σε **s σημαντικά ψηφία**.

Αυτό σημαίνει ότι ο $\bar{x}$ έχει το πολύ **s** και τουλάχιστον **s − 1** **σημαντικά ψηφία** ακριβή.

Παράδειγμα

Δίνονται οι ακόλουθοι αριθμοί x και $\bar{x}$.

(α) $x = 28.254$, $\bar{x} = 28.271$, (β) $x = 0.028254$, $\bar{x} = 0.028271$

(γ) $x = e$, $\bar{x} = 19/7$ (δ) $x = \sqrt{2}$, $\bar{x} = 1.414$

(ϵ) $x = log2$, $\bar{x} = 0.7$.

Να βρεθεί το πλήθος των δεκαδικών και σημαντικών ψηφίων τα οποία είναι ακριβή.

Λύση

(α)

$$\begin{aligned} |\varepsilon_{\bar{x}}| &= |x - \bar{x}| = |28.254 - 28.271| = |-0.017| \\ &= 0.17 \cdot 10^{-1} < 0.5 \cdot 10^{-1} \end{aligned}$$

Συνεπώς, ο $\bar{x}$ έχει **ένα δεκαδικό** ψηφίο ακριβές. Επίσης

$$|\varrho_{\bar{x}}| = \left|\frac{\varepsilon_{\bar{x}}}{x}\right| = \frac{0.17 \cdot 10^{-1}}{28.254} = \frac{0.17 \cdot 10^{-1}}{0.28254 \cdot 10^{2}} \simeq 0.602 \cdot 10^{-3} < 5 \cdot 10^{-3}$$

που σημαίνει ότι ο $\bar{x}$ έχει **τρία σημαντικά** ψηφία ακριβή.

.....Παράδειγμα.....

(β) Επειδή

$$
\begin{aligned}
|\varepsilon_{\bar{x}}| &= |x - \bar{x}| = |0.028254 - 0.028271| = 0.000017 \\
&\simeq 0.17 \cdot 10^{-4} < 0.5 \cdot 10^{-4}
\end{aligned}
$$

και

$$|\varrho_{\bar{x}}| = |\frac{\varepsilon_{\bar{x}}}{x}| = \frac{0.17 \cdot 10^{-4}}{0.28254 \cdot 10^{-1}} \simeq 0.602 \cdot 10^{-3} < 5 \cdot 10^{-3}$$

ο $\bar{x}$ είναι ακριβής σε **τέσσερα δεκαδικά** ψηφία ενώ είναι ακριβής σε **τρία σημαντικά** ψηφία.

Παρατηρήστε ότι ενώ η διαίρεση με δυνάμεις του 10 αυξάνει την ακρίβεια σε δεκαδικά ψηφία, αυτή παραμένει σταθερή στην περίπτωση των σημαντικών ψηφίων (βλ. (α) και (β)).

....Παράδειγμα...)

(γ) Ομοίως, επειδή

$$\begin{aligned} |\varepsilon_{\bar{x}}| &= |x - \bar{x}| = |e - 19/7| = |2.718281 - 2.714286| \simeq 0.003995 \\ &= 0.3995 \cdot 10^{-2} < 0.5 \cdot 10^{-2} \end{aligned}$$

και

$$|\varrho_{\bar{x}}| = \left|\frac{\varepsilon_{\bar{x}}}{x}\right| = \frac{0.3995 \cdot 10^{-2}}{0.2718281 \cdot 10^{1}} \simeq 1.4697 \cdot 10^{-3} < 5 \cdot 10^{-3}$$

Άρα, ο $\bar{x}$ είναι ακριβής σε **δύο δεκαδικά** και **τρία σημαντικά** ψηφία.

(δ) Έχουμε ότι

$$\begin{aligned} |\varepsilon_{\bar{x}}| &= |x - \bar{x}| = |\sqrt{2} - 1.414| = |1.414214 - 1.414| = 0.000214 \\ &= 0.214 \cdot 10^{-3} < 0.5 \cdot 10^{-3} \end{aligned}$$

και

$$|\varrho_{\bar{x}}| = \left|\frac{\varepsilon_{\bar{x}}}{x}\right| = \frac{0.214 \cdot 10^{-3}}{0.1414214 \cdot 10^{1}} \simeq 1.51 \cdot 10^{-4} < 5 \cdot 10^{-4}.$$

Άρα, ο $\bar{x}$ είναι ακριβής σε **τρία δεκαδικά** και **τέσσερα σημαντικά** ψηφία.

Διαδιδόμενο σφάλμα

Ας συμβολίσουμε με

- $\square$ μια αριθμητική πράξη $(+, -, \times, /)$ και με
- $\square^*$ την ίδια πράξη που εκτελείται στον υπολογιστή, η οποία περιέχει το σφάλμα στρογγύλευσης.

Έστω $\overline{x}$ και $\overline{y}$ οι αριθμοί που χρησιμοποιούνται στους υπολογισμούς και είναι οι προσεγγίσεις των τιμών

$$x = \overline{x} + \varepsilon_{\overline{x}}, \quad y = \overline{y} + \varepsilon_{\overline{y}} \tag{16}$$

όπου $\varepsilon_{\overline{x}}$ και $\varepsilon_{\overline{y}}$ τα αντίστοιχα σφάλματα.

Διαδιδόμενο σφάλμα

Τότε με την εκτέλεση της αριθμητικής πράξης $\square$ ο αριθμός που υπολογίζεται είναι ο

$$\overline{x}\square^*\overline{y}$$

και το συνολικό σφάλμα δίνεται από τον τύπο

$$\begin{aligned} \varepsilon_{\overline{x}\square\overline{y}} &= x\square y - \overline{x}\square^*\overline{y} \\ &= \underbrace{\left(x\square y - \overline{x}\square\overline{y}\right)}_{\text{διαδιδόμενο}} + \underbrace{\left(\overline{x}\square\overline{y} - \overline{x}\square^*\overline{y}\right)}_{\text{στρογγύλευσης}} . \end{aligned} \tag{17}$$

Όροι σφάλματος

Ο πρώτος όρος στο δεξί μέλος της (17) είναι το **διαδιδόμενο σφάλμα** και ο δεύτερος όρος είναι το **σφάλμα στρογγύλευσης** κατά τον υπολογισμό του $\overline{x}\square\overline{y}$.

....Διαδιδόμενο σφάλμα.....

Επειδή όμως

$$fl\left(\overline{x}\square\overline{y}\right) = \overline{x}\square^{*}\overline{y} \tag{18}$$

που σημαίνει ότι το $\overline{x}\square\overline{y}$ υπολογίζεται ακριβώς και στη συνέχεια στρογγυλεύεται, λόγω των (7) και (18) έχουμε

$$\frac{\left|\overline{x}\square\overline{y} - \overline{x}\square^{*}\overline{y}\right|}{\left|\overline{x}\square\overline{y}\right|} \leq \frac{1}{2}\beta^{-n+1} \tag{19}$$

με την υπόθεση ότι χρησιμοποιείται στρογγύλευση.

Υποθέτοντας ότι το σφάλμα στρογγύλευσης είναι πολύ μικρό($\simeq$ 0) θα μελετηθεί η συμπεριφορά του σφάλματος διάδοσης.

....Διαδιδόμενο σφάλμα.....

Έστω $\overline{x}$ και $\overline{y}$ οι αριθμοί που χρησιμοποιούνται στους υπολογισμούς και είναι οι προσεγγίσεις των τιμών

$$x = \overline{x} + \varepsilon_{\overline{x}}, \quad y = \overline{y} + \varepsilon_{\overline{y}} \tag{20}$$

όπου $\varepsilon_{\overline{x}}$ και $\varepsilon_{\overline{y}}$ τα αντίστοιχα σφάλματα.

- Αν $\Box = \pm$, τότε από την (17) έχουμε

$$\begin{aligned} \varepsilon_{\overline{x}\pm\overline{y}} &= (x \pm y) - (\overline{x} \pm \overline{y}) \\ &= (x - \overline{x}) \pm (y - \overline{y}) \end{aligned}$$

ή

$$\varepsilon_{\overline{x}\pm\overline{y}} = \varepsilon_{\overline{x}} \pm \varepsilon_{\overline{y}}$$

και τέλος για τα απόλυτα σφάλματα έχουμε τη σχέση

$$\left|\varepsilon_{\overline{x}\pm\overline{y}}\right| \leq \left|\varepsilon_{\overline{x}}\right| + \left|\varepsilon_{\overline{y}}\right|. \tag{21}$$

Ισχύει δηλαδή το ακόλουθο θεώρημα.

...Διαδιδόμενο σφάλμα...

Θεώρημα

Η μέγιστη τιμή του απολύτου σφάλματος του αθροίσματος ή της διαφοράς δύο αριθμών είναι ίση με το άθροισμα των απολύτων αφαλμάτων των αριθμών αυτών.

Λόγω του ανωτέρω θεωρήματος, αν $\overline{x}$ και $\overline{y}$ έχουν ακρίβεια τεσσάρων δεκαδικών ψηφίων (δηλαδή, αν $\left|\varepsilon_{\overline{x}}\right|, \left|\varepsilon_{\overline{y}}\right| \leq \frac{1}{2} \cdot 10^{-4}$), τότε η ποσότητα $\overline{x} \pm \overline{y}$ μπορεί να διαφέρει από την $x \pm y$ το πολύ κατά 10^{-4}.

Επομένως είναι πιθανό η ποσότητα $\overline{x} \pm \overline{y}$ να έχει ένα ψηφίο λάθος στην τέταρτη δεκαδική θέση.

Επιπλέον, αν οι $\overline{x}$ και $\overline{y}$ έχουν διαφορετική ακρίβεια, τότε η ποσότητα $\overline{x} \pm \overline{y}$ στη χειρότερη περίπτωση θα είναι λανθασμένη από εκείνη τη δεκαδική θέση που αντιστοιχεί στο μεγαλύτερο από τα $\left|\varepsilon_{\overline{x}}\right|$ και $\left|\varepsilon_{\overline{y}}\right|$.

Διαδιδόμενο σφάλμα...

●● Αν $\Box = \cdot$, τότε

$$\varepsilon_{\overline{xy}} = xy - (x - \varepsilon_{\overline{x}})(y - \varepsilon_{\overline{y}})$$

ή

$$\varepsilon_{\overline{xy}} = y\varepsilon_{\overline{x}} + x\varepsilon_{\overline{y}} + \varepsilon_{\overline{x}}\varepsilon_{\overline{y}} \tag{22}$$

●●● Όμοια εάν $\Box = /$ και $y, \overline{y} \neq 0$, τότε

$$\varepsilon_{\overline{x}/\overline{y}} = \frac{x}{y} - \frac{\overline{x}}{\overline{y}} = \frac{x}{y} - \frac{x - \varepsilon_{\overline{x}}}{y - \varepsilon_{\overline{y}}} = \frac{y\varepsilon_{\overline{x}} - x\varepsilon_{\overline{y}}}{y^2 - y\varepsilon_{\overline{y}}} \tag{23}$$

....Διαδιδόμενο σφάλμα....

Διατηρώντας τους υπερέχοντες όρους οι (22) και (23) δίνουν

$$\varepsilon_{\overline{xy}} \simeq x\varepsilon_{\overline{y}} + y\varepsilon_{\overline{x}} \tag{24}$$

και

$$\varepsilon_{\overline{x}/\overline{y}} \simeq \frac{y\varepsilon_{\overline{x}} - x\varepsilon_{\overline{y}}}{y^2} \tag{25}$$

αντίστοιχα.

Παρατηρήσεις

Από την (24) παρατηρούμε ότι μεγάλες τιμές του $\overline{x}$ ή του $\overline{y}$ έχουν σαν αποτέλεσμα την αύξηση του σφάλματος στο γινόμενο xy.

Η (25) θα παράγει μεγάλο σφάλμα στη διαίρεση $\overline{x}/\overline{y}$ για μεγάλες τιμές του $\overline{x}$ και/ή μικρές τιμές του $\overline{y}$.

Συμπέρασμα

Ο πολ/σμός με μεγάλους αριθμούς και οι διαιρέσεις όπου ο διαιρετέος είναι ο μεγάλος αριθμός και/ή ο διαιρέτης είναι μικρός θα πρέπει να αποφεύγονται με αναδιάταξη των υπολογισμών

...Διαδιδόμενο σφάλμα...

Θεώρημα

Η μέγιστη τιμή του απόλυτου σχετικού σφάλματος του γινομένου ή του πηλίκου δύο αριθμών είναι κατα προσέγγιση ίση με το άθροισμα των απολύτων σχετικών σφαλμάτων των αριθμών αυτών.

Απόδειξη.

Διαιρώντας την (22) δια xy λαμβάνουμε

$$\varrho_{\overline{xy}} = \frac{\varepsilon_{\overline{xy}}}{xy} = \frac{\varepsilon_{\bar{x}}}{x} + \frac{\varepsilon_{\bar{y}}}{y} - \frac{\varepsilon_{\bar{x}}}{x} \cdot \frac{\varepsilon_{\bar{y}}}{y} \tag{26}$$

ή

$$\varrho_{\overline{xy}} = \varrho_{\bar{x}} + \varrho_{\bar{y}} - \varrho_{\bar{x}}\varrho_{\bar{y}}. \tag{27}$$

□

...Απόδειξη...

Απόδειξη.

Αν $|\varrho_{\bar{x}}|, \quad |\varrho_{\bar{y}}| << 1$, με άλλα λόγια αν

$$|\varepsilon_{\bar{x}}| << |x| \quad \text{ή} \quad |x - \bar{x}| << |x|$$

και

$$|\varepsilon_{\bar{y}}| << |y| \quad \text{ή} \quad |y - \bar{y}| << |y|$$

, δηλαδή οι ποσότητες $\bar{x}$ και $\bar{y}$ είναι καλές προσεγγίσεις, με την έννοια ότι τα σφάλματα είναι μικρά σε σχέση με τις ακριβείς τιμές x και y, τότε η (27) γράφεται

$$\varrho_{\bar{x}\bar{y}} \simeq \varrho_{\bar{x}} + \varrho_{\bar{y}}$$

ή

$$|\varrho_{\bar{x}\bar{y}}| \lesssim |\varrho_{\bar{x}}| + |\varrho_{\bar{y}}| \tag{28}$$

και αποδείχθηκε το θεώρημα για το γινόμενο δύο αριθμών. □

Απόδειξη (συνέχεια...)

Απόδειξη.

Όμοια για το πηλίκο, διαιρώντας την (23) δια x/y λαμβάνουμε

$$\varrho_{\bar{x}/\bar{y}} = \frac{\varepsilon_{\bar{x}/\bar{y}}}{x/y} = \frac{y\varepsilon_{\bar{x}} - x\varepsilon_{\bar{y}}}{x(y - \varepsilon_{\bar{y}})}. \tag{29}$$

Αν $|\varrho_{\bar{y}}| << 1$ ή $|\varepsilon_{\bar{y}}| << |y|$ τότε η (29) δίνει

$$\varrho_{\bar{x}/\bar{y}} \simeq \frac{y\varepsilon_{\bar{x}} - x\varepsilon_{\bar{y}}}{xy} = \frac{\varepsilon_{\bar{x}}}{x} - \frac{\varepsilon_{\bar{y}}}{y}$$

ή

$$\varrho_{\bar{x}/\bar{y}} \simeq \varrho_{\bar{x}} - \varrho_{\bar{y}}$$

ή

$$|\varrho_{\bar{x}/\bar{y}}| \lesssim |\varrho_{\bar{x}}| + |\varrho_{\bar{y}}| \tag{30}$$

. □

Συμπεράσματα

- Από το παραπάνω θεώρημα προκύπτει ότι αν $\bar{x}$ και $\bar{y}$ έχουν ακρίβεια s σημαντικών ψηφίων, τότε οι ποσότητες $\bar{x}\bar{y}$ και $\bar{x}/\bar{y}$ θα έχουν περίπου ακρίβεια s σημαντικών ψηφίων.
- Παρατηρούμε ότι

$$\varrho_{\bar{x}\pm\bar{y}} = \frac{\varepsilon_{\bar{x}\pm\bar{y}}}{x \pm y} = \frac{\varepsilon_{\bar{x}}}{x \pm y} \pm \frac{\varepsilon_{\bar{y}}}{x \pm y} = \left(\frac{x}{x \pm y}\right)\varrho_{\bar{x}} \pm \left(\frac{y}{x \pm y}\right)\varrho_{\bar{y}}. \qquad (31)$$

Η (31) δηλώνει ότι αν $x \pm y$ είναι πολύ μικρότερο από το x ή y, τότε οι παράγοντες $x/(x \pm y)$ και $y/(x \pm y)$ θα λάβουν μεγάλες τιμές με συνέπεια να αυξηθεί η τιμή του $\varrho_{\bar{x}\pm\bar{y}}$.

Για το λόγο αυτό θα πρέπει να αποφεύγεται η πρόσθεση ενός πολύ μεγάλου και ενός πολύ μικρού αριθμού ή η αφαίρεση δύο περίπου ίσων αριθμών.

Παράδειγμα

Δίνονται οι αριθμοί $\bar{x}_1 = 4.54$, $\bar{x}_2 = 3.00$ και $\bar{x}_3 = 15.0$ με ακρίβεια **τριών σημαντικών ψηφίων**. Να υπολογιστεί, κατά προσέγγιση

(α) Το μέγιστο απόλυτο σφάλμα της ποσότητας $\bar{x}_1 - \bar{x}_2 + \bar{x}_3$

(β) Το μέγιστο σχετικό σφάλμα της ποσότητας $\bar{x}_1 \bar{x}_2 / \bar{x}_3$

(γ) Τι συμπεράσματα εξάγετε σχετικά με την ακρίβεια του αποτελέσματος στα (α) και (β) ;

Λύση

(α) Επειδή οι αριθμοί είναι ακριβείς σε τρία σημαντικά ψηφία συνεπάγεται οτι οι $\bar{x}_1,\ \bar{x}_2$ είναι ακριβείς σε δύο δεκαδικά ψηφία και ο $\bar{x}_3$ σε ένα δεκαδικό ψηφίο, δηλαδή

$$|\varepsilon_{\bar{x}_1}|,\ |\varepsilon_{\bar{x}_2}| \leq \frac{1}{2}\cdot 10^{-2}$$

και

$$|\varepsilon_{\bar{x}_3}| \leq \frac{1}{2}\cdot 10^{-1}$$

.

Λόγω του Θεωρήματος 5.1 έχουμε ότι

$$\begin{aligned}
|\varepsilon_{\bar{x}_1-\bar{x}_2+\bar{x}_3}| &\lesssim |\varepsilon_{\bar{x}_1}| + |\varepsilon_{\bar{x}_2}| + |\varepsilon_{\bar{x}_3}| = 2(\frac{1}{2}\cdot 10^{-2}) + \frac{1}{2}\cdot 10^{-1} \\
&= 10^{-2} + 0.5\cdot 10^{-1} = 0.1\cdot 10^{-1} + 0.5\cdot 10^{-1} \\
&= 0.6\cdot 10^{-1} = 0.06\cdot 10^{0} < 0.5\cdot 10^{0}
\end{aligned}$$

Στη χειρότερη περίπτωση το αποτέλεσμα δεν θα έχει κανένα δεκαδικό ψηφίο ακριβές.

Λύση (συνέχεια...)

(β) Επειδή οι αριθμοί είναι ακριβείς σε τρία σημαντικά ψηφία συνεπάγεται

$$|\varrho_{\bar{x}_1}|, |\varrho_{\bar{x}_2}|, |\varrho_{\bar{x}_3}| \leq 5 \cdot 10^{-3}.$$

Λόγω του Θεωρήματος 5.2 έχουμε ότι

$$|\bar{\varrho}_{\bar{x}_1 \bar{x}_2 / \bar{x}_3}| \lesssim |\varrho_{\bar{x}_1}| + |\varrho_{\bar{x}_2}| + |\varrho_{\bar{x}_3}| \leq 3 \cdot 5 \cdot 10^{-3} = 1.5 \cdot 10^{-2} < 5 \cdot 10^{-2}$$

Συνεπώς στη χειρότερη περίπτωση το αποτέλεσμα θα είναι ακριβές σε δύο σημαντικά ψηφία.

Σημειώματα

Σημείωμα Αναφοράς

Σημείωμα Αδειοδότησης

Διατήρηση Σημειωμάτων

Οποιαδήποτε αναπαραγωγή ή διασκευή του υλικού θα πρέπει να συμπεριλαμβάνει:

- το Σημείωμα Αναφοράς
- το Σημείωμα Αδειοδότησης
- τη δήλωση Διατήρησης Σημειωμάτων
- το Σημείωμα Χρήσης Έργων Τρίτων (εφόσον υπάρχει)

μαζί με τους συνοδευόμενους υπερσυνδέσμους.

Σημείωμα Χρήσης Έργων τρίτων

Το Έργο αυτό κάνει χρήση του ακόλουθου έργου:
¨ Εισαγωγή στην Αριθμητική Ανάλυση : Μια αλγοριθμική προσέγγιση, αυτο-έκδοση, Αθήνα, 2009¨, Νικόλαος Μισυρλής.