Εισαγωγή στην Αριθμητική Ανάλυση

Σταμάτης Σταματιάδης
stamatis@materials.uoc.gr

Τμήμα Επιστήμης και Τεχνολογίας Υλικών,
Πανεπιστήμιο Κρήτης

ΠΡΩΤΗ ΔΙΑΛΕΞΗ

Η Αριθμητική Ανάλυση είναι κλάδος των Μαθηματικών που ασχολείται με την επίλυση προβλημάτων που ανακύπτουν συχνά στους επιστημονικούς υπολογισμούς.

Η επίλυση είναι *αριθμητική*, δηλαδή καταλήγουμε σε αριθμό (και όχι τύπο με παραμέτρους) ως απάντηση στο πρόβλημα. Συχνά αυτό μας αρκεί και συνήθως είναι το μόνο εφικτό.

Υπολογισμός ρίζας συνάρτησης

Υπολογισμός σημείου x_0 στο οποίο μια συνάρτηση $f(x)$ μηδενίζεται.

Γνωρίζουμε την απάντηση αναλυτικά για κάποιες συναρτήσεις, π.χ.

$$\cos x = 0 \qquad \Rightarrow x = k\pi + \frac{\pi}{2} \,, \quad k = 0, \pm 1, \pm 2, \ldots \,,$$

ή

$$ax^2 + bx + c = 0 \quad \Rightarrow x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \,, \quad \text{αν } a \neq 0 \,.$$

Τι κάνουμε για πιο πολύπλοκες συναρτήσεις;

Ενδεικτικά προβλήματα (2/5)

Επίλυση συστήματος n γραμμικών εξισώσεων με n αγνώστους

$$
\begin{array}{rcl}
a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n & = & b_1 \\
a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n & = & b_2 \\
\vdots \qquad\quad \vdots \qquad\qquad\quad \vdots & & \vdots \\
a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n & = & b_n \; .
\end{array}
$$

Ενδεικτικά προβλήματα (2/5)

Επίλυση συστήματος n γραμμικών εξισώσεων με n αγνώστους

$$\begin{aligned}
a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\
a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\
\vdots \qquad \vdots \qquad\qquad \vdots \quad &\quad \vdots \\
a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n &= b_n \,.
\end{aligned}$$

Επίλυση συστήματος n μη γραμμικών εξισώσεων με n αγνώστους

$$\begin{aligned}
f_1(x_1, x_2, \ldots, x_n) &= 0 \\
f_2(x_1, x_2, \ldots, x_n) &= 0 \\
\vdots \quad &\quad \vdots \\
f_n(x_1, x_2, \ldots, x_n) &= 0
\end{aligned}$$

Ολοκλήρωση συνάρτησης

Υπολογισμός ολοκληρώματος μιας συνάρτησης $f(x)$ με συγκεκριμένα όρια,

$$\int_a^b f(x)\,\mathrm{d}x\ .$$

Ολοκλήρωση συνάρτησης

Υπολογισμός ολοκληρώματος μιας συνάρτησης $f(x)$ με συγκεκριμένα όρια,

$$\int_a^b f(x)\,\mathrm{d}x \,.$$

Προσέγγιση συνάρτησης

Μια άγνωστη συνάρτηση $f(x)$ περνά από n σημεία $(x_i, f(x_i))$, $i = 1, \ldots, n$.

- Ποια είναι η τιμή της σε κάποιο άλλο σημείο $\bar{x}$;
- Ποια η παράγωγός της στο $\bar{x}$;
- Ποιο είναι το ολοκλήρωμά της σε κάποιο διάστημα;
- Ποια είναι η μέγιστη ή η ελάχιστη τιμή της;

Ολοκλήρωση συνάρτησης

Υπολογισμός ολοκληρώματος μιας συνάρτησης $f(x)$ με συγκεκριμένα όρια,

$$\int_a^b f(x)\,\mathrm{d}x \; .$$

Προσέγγιση συνάρτησης

Μια άγνωστη συνάρτηση $f(x)$ περνά από n σημεία $(x_i, f(x_i))$, $i = 1, \ldots, n$.

- Ποια είναι η τιμή της σε κάποιο άλλο σημείο $\bar{x}$;
- Ποια η παράγωγός της στο $\bar{x}$;
- Ποιο είναι το ολοκλήρωμά της σε κάποιο διάστημα;
- Ποια είναι η μέγιστη ή η ελάχιστη τιμή της;

Γραμμική Άλγεβρα

Για πίνακα διαστάσεων $n \times n$ θέλουμε να υπολογίσουμε τις ιδιοτιμές, τα ιδιοδιανύσματα, την ορίζουσα, τον αντίστροφο πίνακα κλπ.

Επίλυση Διαφορικής Εξίσωσης

Μια εξίσωση που περιγράφει μια σχέση μεταξύ μιας ανεξάρτητης μεταβλητής, x, μιας εξαρτημένης συνάρτησης, y, και μίας ή περισσότερων παραγώγων τής y λέγεται *συνήθης Διαφορική Εξίσωση*:

$$y^{(n)}(x) = f\left(x, y(x), y'(x), \ldots, y^{(n-1)}(x)\right) \ .$$

Θα δούμε πώς επιλύεται αν τα $y(x_0), y'(x_0), \ldots, y^{(n-1)}(x_0)$ είναι γνωστά για κάποιο σημείο x_0.

Γρήγορος υπολογισμός των συντελεστών σειράς Fourier

Η σειρά Fourier της $f(x)$ που είναι περιοδική με περίοδο L είναι:

$$f(x) = \sum_{m=-\infty}^{\infty} C_m \exp\left(\mathrm{i}\frac{2m\pi x}{L}\right) .$$

Οι συντελεστές Fourier C_m είναι

$$C_m = \frac{1}{L}\int_0^L \exp\left(-\mathrm{i}\frac{2m\pi x}{L}\right) f(x)\mathrm{d}x , \qquad m = 0, \pm 1, \pm 2, \ldots .$$

Πώς θα τους υπολογίσουμε γρήγορα, έστω και προσεγγιστικά, για οποιαδήποτε $f(x)$;

Προαπαιτούμενες γνώσεις

- Μαθηματικά Α' και Β' έτους.
- Προγραμματισμός (σε οποιαδήποτε γλώσσα).

Σχετικά με το μάθημα

Προαπαιτούμενες γνώσεις

- Μαθηματικά Α' και Β' έτους.
- Προγραμματισμός (σε οποιαδήποτε γλώσσα).

Διεξαγωγή μαθήματος

Διαλέξεις Δευτέρα 11:00-13:00.

Ασκήσεις Τετάρτη 10:00-13:00.

eclass https://teleclass.materials.uoc.gr/courses/SEM4105

Βιβλιογραφία
- Στο eclass διατίθεται το βιβλίο που θα διδαχτεί.
- Παρέχεται επιπλέον βιβλίο μέσω Εύδοξου:
  - «Αριθμητικές Μέθοδοι και Εφαρμογές για Μηχανικούς», Σαρρής Ι.- Καρακασίδης Θ. (ΦΥΣΙΚΟ)
  - «Εισαγωγή στην αριθμητική ανάλυση», Ακρίβης Γ.Δ., Δουγαλής Β.Α. (ΤΕΤΥ)
  - «Εισαγωγή στην Αριθμητική Ανάλυση», Πιτσούλης Λ. (ΦΥΣΙΚΟ και ΤΕΤΥ)

Εξετάσεις
- Πρόοδος (προαιρετική, 40%)
- Τελική εξέταση (60% ή 100%)

Περιλαμβάνουν ασκήσεις, στο χαρτί (33%) και στον υπολογιστή (67%).

Κεφάλαιο Πρώτο

Απόκλιση από την (άγνωστη) πραγματική τιμή, της τιμής που υπολογίζεται με μια μέθοδο της Αριθμητικής Ανάλυσης έχουμε λόγω:

του αλγόριθμου που επιλέγουμε.

Με την εφαρμογή του παράγεται γενικά μια προσεγγιστική τιμή και μια περιοχή τιμών γύρω από αυτή, στην οποία βρίσκεται η πραγματική τιμή. Το εύρος της περιοχής μπορεί να γίνει όσο μικρό επιθυμούμε (αλλά όχι 0, σε πεπερασμένο χρόνο).

Σφάλματα στους υπολογισμούς

Απόκλιση από την (άγνωστη) πραγματική τιμή, της τιμής που υπολογίζεται με μια μέθοδο της Αριθμητικής Ανάλυσης έχουμε λόγω:

του αλγόριθμου που επιλέγουμε.

Με την εφαρμογή του παράγεται γενικά μια προσεγγιστική τιμή και μια περιοχή τιμών γύρω από αυτή, στην οποία βρίσκεται η πραγματική τιμή. Το εύρος της περιοχής μπορεί να γίνει όσο μικρό επιθυμούμε (αλλά όχι 0, σε πεπερασμένο χρόνο).

της αναπαράστασης των πραγματικών αριθμών στον υπολογιστή.

Λόγω πεπερασμένης διαθέσιμης μνήμης, οι υπολογιζόμενες πραγματικές τιμές στρογγυλοποιούνται. Γενικά, δεν μπορούμε να επηρεάσουμε ουσιαστικά αυτό το σφάλμα.

Συστήματα αρίθμησης: αναπαράσταση ακέραιων (1/2)

- Ένας μη αρνητικός ακέραιος αριθμός Κ αναπαρίσταται στο σύστημα με βάση B με μια σειρά ψηφίων

$$d_n d_{n-1} \ldots d_1 d_0 \;, \qquad \text{με } d_n \neq 0 \;.$$

Συστήματα αρίθμησης: αναπαράσταση ακέραιων (1/2)

- Ένας μη αρνητικός ακέραιος αριθμός K αναπαρίσταται στο σύστημα με βάση B με μια σειρά ψηφίων

$$d_n d_{n-1} \ldots d_1 d_0 \ , \qquad \text{με } d_n \neq 0 \ .$$

- Τα ψηφία d_i ικανοποιούν τη σχέση $0 \leq d_i < B$. Αν δεν επαρκούν τα ψηφία 0–9 για τα d_i χρησιμοποιούνται γράμματα του λατινικού αλφάβητου.

Συστήματα αρίθμησης: αναπαράσταση ακέραιων (1/2)

- Ένας μη αρνητικός ακέραιος αριθμός K αναπαρίσταται στο σύστημα με βάση B με μια σειρά ψηφίων

$$d_n d_{n-1} \ldots d_1 d_0 \;, \qquad \text{με } d_n \neq 0 \;.$$

- Τα ψηφία d_i ικανοποιούν τη σχέση $0 \leq d_i < B$. Αν δεν επαρκούν τα ψηφία 0–9 για τα d_i χρησιμοποιούνται γράμματα του λατινικού αλφάβητου.
- Η σειρά ψηφίων κωδικοποιεί ένα άθροισμα δυνάμεων του B:

$$K = d_n \times B^n + \ldots + d_2 \times B^2 + d_1 \times B^1 + d_0 \times B^0 \;.$$

Συστήματα αρίθμησης: αναπαράσταση ακέραιων (1/2)

- Ένας μη αρνητικός ακέραιος αριθμός K αναπαρίσταται στο σύστημα με βάση B με μια σειρά ψηφίων

$$d_n d_{n-1} \ldots d_1 d_0 \ , \qquad \text{με } d_n \neq 0 \ .$$

- Τα ψηφία d_i ικανοποιούν τη σχέση $0 \leq d_i < B$. Αν δεν επαρκούν τα ψηφία 0–9 για τα d_i χρησιμοποιούνται γράμματα του λατινικού αλφάβητου.
- Η σειρά ψηφίων κωδικοποιεί ένα άθροισμα δυνάμεων του B:

$$K = d_n \times B^n + \ldots + d_2 \times B^2 + d_1 \times B^1 + d_0 \times B^0 \ .$$

- Καθώς

$$K = \left(d_n \times B^{n-i} + \ldots + d_{i+1} \times B + d_i\right) \times B^i + d_{i-1} \times B^{i-1} + \ldots + d_0 \times B^0 \ ,$$

τα ψηφία είναι:

$$d_i = \left(K \operatorname{div} B^i\right) \bmod B \ , \qquad i = 0, 1, \ldots \ .$$

Παραδείγματα

- $$\begin{aligned} 1578_{10} &\equiv 1 \times 10^3 + 5 \times 10^2 + 7 \times 10^1 + 8 \times 10^0 \ , \\ 100100_2 &\equiv 1 \times 2^5 + 0 \times 2^4 + 0 \times 2^3 + 1 \times 2^2 + 0 \times 2^1 + 0 \times 2^0 = 36_{10} \ . \end{aligned}$$
- Το πλήθος των γραμμάτων του αλφαβήτου γράφεται 24 στο δεκαδικό, 11000 στο δυαδικό, 18 στο δεκαεξαδικό σύστημα.
- Ο δεκαδικός 64206 γράφεται *face* στο δεκαεξαδικό.

Συστήματα αρίθμησης: άθροισμα ακέραιων

Το άθροισμα δύο σειρών με ψηφία,

$$a_n a_{n-1} \dots a_1 a_0 \ ,$$

και

$$b_n b_{n-1} \dots b_1 b_0 \ ,$$

στην ίδια βάση B, που αναπαριστούν ακέραιους, είναι μια σειρά με ψηφία c_i για τα οποία ισχύει

$$c_i = (a_i + b_i + e_i) \bmod B \ , \qquad i \geq 0 \ ,$$

όπου e_i το κρατούμενο για το ψηφίο i. Το e_i ικανοποιεί τη σχέση

$$e_i = \begin{cases} 0 \ , & i = 0 \ , \\ (a_{i-1} + b_{i-1} + e_{i-1}) \operatorname{div} B \ , & i > 0 \ . \end{cases}$$

Συστήματα αρίθμησης: αναπαράσταση πραγματικών (1/2)

Ένας μη αρνητικός πραγματικός αριθμός σε κάποια βάση B αναπαρίσταται από μια σειρά ψηφίων που χωρίζονται με τελεία (υποδιαστολή). Π.χ. στο δεκαδικό σύστημα μπορούμε να γράψουμε

$$123.456\ .$$

Αλλά

$$123.456 \equiv 12.3456 \times 10^1 \equiv 1.23456 \times 10^2 \equiv 0.123456 \times 10^3 \text{ κλπ.}$$

και

$$123.456 \equiv 1234.56 \times 10^{-1} \equiv 12345.6 \times 10^{-2} \equiv 123456.0 \times 10^{-3} \text{ κλπ.}$$

Όλες οι μορφές είναι ισοδύναμες.

Επομένως:

- Ένας πραγματικός αριθμός X μπορεί να γραφεί σε βάση B στη μορφή
$$X = \pm d_0.d_1d_2d_3\ldots d_n \times B^e$$
με $d_0 \neq 0$ και με ακέραιο εκθέτη e.
- Τα ψηφία d_i ικανοποιούν τη σχέση $0 \leq d_i < B$.
- Ισχύει
$$X = \pm \left(\sum_{i=0}^{n} d_i B^{-i} \right) \times B^e \ .$$
- Τα ψηφία d_0, d_i, ..., d_n αποτελούν τα *σημαντικά ψηφία (significant digits)* του αριθμού.

Αναπαράσταση ακεραίων στον υπολογιστή (1/2)

- Ο HY χρησιμοποιεί το δυαδικό σύστημα για αποθήκευση ακεραίων. Συνήθως, ο ακέραιος αναπαρίσταται σε 32 bits.
- Οι αρνητικοί αριθμοί αναπαριστώνται συνήθως ως εξής: αν K είναι θετικός αριθμός, ο αριθμός $-K$ είναι αυτός που ικανοποιεί τη σχέση

  $$K + (-K) = 0 \; ,$$

  δηλαδή είναι ο αριθμός που αν προστεθεί στον K δίνει αποτέλεσμα 0. Στην πρόσθεση κρατάμε μόνο τα πρώτα 32 bits.

Αναπαράσταση ακεραίων στον υπολογιστή (1/2)

- Ο ΗΥ χρησιμοποιεί το δυαδικό σύστημα για αποθήκευση ακεραίων. Συνήθως, ο ακέραιος αναπαρίσταται σε 32 bits.
- Οι αρνητικοί αριθμοί αναπαριστώνται συνήθως ως εξής: αν K είναι θετικός αριθμός, ο αριθμός $-K$ είναι αυτός που ικανοποιεί τη σχέση

$$K + (-K) = 0 \ ,$$

δηλαδή είναι ο αριθμός που αν προστεθεί στον K δίνει αποτέλεσμα 0. Στην πρόσθεση κρατάμε μόνο τα πρώτα 32 bits.

Παράδειγματα

- Ο ακέραιος 1569 αντιπροσωπεύεται από τη σειρά

$$00000000\ 00000000\ 00000110\ 00100001 \ .$$

- Ο αριθμός -1569 είναι αυτός που αν προστεθεί στο 1569 δίνει 0, η σειρά, δηλαδή,

$$11111111\ 11111111\ 11111001\ 11011111 \ .$$

Αναπαράσταση ακεραίων στον υπολογιστή (2/2)

Παρατηρήσεις

- Ο μεγαλύτερος (εμπρόσημος) ακέραιος σε 32 bits είναι ο

  $$01111111\ 11111111\ 11111111\ 11111111$$

  δηλαδή, ο 2147483647 του δεκαδικού.
- Ο μικρότερος ακέραιος σε 32 bits είναι ο

  $$10000000\ 00000000\ 00000000\ 00000000$$

  δηλαδή ο -2147483648. Ο αντίθετος του συγκεκριμένου αριθμού αναπαρίσταται με την ίδια σειρά.
- Ο αριθμός

  $$11111111\ 11111111\ 11111111\ 11111111$$

  δεν είναι ο μέγιστος που μπορεί να αναπαρασταθεί. Έχει αντίθετο τον

  $$00000000\ 00000000\ 00000000\ 00000001$$

  δηλαδή, ο αρχικός είναι ο -1.

Στο δυαδικό σύστημα ο πραγματικός αριθμός έχει τη μορφή

$$\pm 1.d_1 d_2 d_3 \ldots d_n \times 2^e \ .$$

Αποθηκεύεται σε 32 bits (για απλή ακρίβεια) ή σε 64 bits (για διπλή ακρίβεια), ως εξής:

- Το πρώτο bit αναπαριστά το πρόσημο του αριθμού: είναι 0/1 αν το πρόσημο είναι $+/-$.
- Τα επόμενα 8 (σε απλή ακρίβεια) ή 11 bits (σε διπλή ακρίβεια) αποθηκεύουν το *e αφού προστεθεί το* $2^{8-1}-1=127$ *(σε απλή ακρίβεια) ή το* $2^{11-1}-1=1023$ *(σε διπλή ακρίβεια).*
- Στα τελευταία 23 (σε απλή ακρίβεια) ή 52 bits (σε διπλή ακρίβεια) αποθηκεύονται ισάριθμα δυαδικά ψηφία d_1, d_2,....
  Το d_0, που είναι πάντα 1, δεν αποθηκεύεται.

Αναπαράσταση πραγματικών στον υπολογιστή (2/4)

Όρια πραγματικών αριθμών

- Συγκεκριμένες σειρές των 32 ή 64 bits αντιστοιχούν στο $\pm$infinity και στο NaN (Not A Number). Σε αυτές τα 8 (σε απλή ακρίβεια) ή 11 bits (σε διπλή ακρίβεια) για τον εκθέτη είναι όλα 1.
- Ο μεγαλύτερος εκθέτης που μπορεί να αναπαρασταθεί είναι ο 127 (σε απλή ακρίβεια) ή ο 1023 (σε διπλή ακρίβεια).
- Ο μικρότερος εκθέτης είναι ο -126 (σε απλή ακρίβεια) ή ο -1022 (σε διπλή ακρίβεια).

Όρια πραγματικών αριθμών

- Συγκεκριμένες σειρές των 32 ή 64 bits αντιστοιχούν στο $\pm$infinity και στο NaN (Not A Number). Σε αυτές τα 8 (σε απλή ακρίβεια) ή 11 bits (σε διπλή ακρίβεια) για τον εκθέτη είναι όλα 1.
- Ο μεγαλύτερος εκθέτης που μπορεί να αναπαρασταθεί είναι ο 127 (σε απλή ακρίβεια) ή ο 1023 (σε διπλή ακρίβεια).
- Ο μικρότερος εκθέτης είναι ο -126 (σε απλή ακρίβεια) ή ο -1022 (σε διπλή ακρίβεια).

Παρατηρήσεις

- **Σε απλή ακρίβεια**, ο μεγαλύτερος πραγματικός αριθμός που μπορεί να αποθηκευτεί είναι της τάξης του $2^{127} \approx 10^{38}$ ενώ ο κατά μέτρο μικρότερος είναι της τάξης του $2^{-126} \approx 10^{-38}$.
- **Σε διπλή ακρίβεια**, ο μεγαλύτερος πραγματικός αριθμός είναι της τάξης του $2^{1023} \approx 10^{308}$ ενώ ο κατά μέτρο μικρότερος είναι της τάξης του $2^{-1022} \approx 10^{-308}$.

Αναπαράσταση πραγματικών στον υπολογιστή (3/4)

Σφάλμα αναπαράστασης

- *Η αναπαράσταση πραγματικών αριθμών δεν είναι πάντα δυνατή με απόλυτη ακρίβεια* λόγω του πεπερασμένου αριθμού bits.
- Αποκόπτονται ή στρογγυλεύονται τα bits μετά το 23ο (απλή ακρίβεια) ή 52ο (διπλή ακρίβεια).
- Το σφάλμα έχει μέγιστη τιμή $\varepsilon = 2^{-23} \approx 1.19 \times 10^{-7}$ (για απλή ακρίβεια) ή $\varepsilon = 2^{-52} \approx 2.22 \times 10^{-16}$ (για διπλή ακρίβεια).
- Η μέγιστη τιμή σφάλματος αναπαράστασης αποκαλείται *έψιλον της μηχανής*.

Αναπαράσταση πραγματικών στον υπολογιστή (4/4)

Παρατηρήσεις

Λόγω της πεπερασμένης ακρίβειας αναπαράστασης:

- Σε υπολογιστή με έψιλον της μηχανής ε, για κάθε πραγματικό x με $|x| < \varepsilon$ ισχύει $1 + x = 1$. Δηλαδή:

  υπάρχει ένα όριο κάτω από το οποίο οι πραγματικοί αριθμοί συμπεριφέρονται ως μηδέν σε προσθέσεις ή αφαιρέσεις με αριθμούς της τάξης του 1.

- το αποτέλεσμα της πράξης μεταξύ πραγματικών $x + (y + z)$ μπορεί να είναι διαφορετικό από το $(x + y) + z$. Επομένως, δεν ισχύει η προσεταιριστική ιδιότητα της πρόσθεσης.
- η τιμή $x + y + z$ μπορεί να είναι διαφορετική από την $x + z + y$. Δεν ισχύει η αντιμεταθετική ιδιότητα της πρόσθεσης.
- Η σύγκριση για ισότητα δύο πραγματικών αριθμών που προέκυψαν από πράξεις πρέπει να αποφεύγεται. Λόγω των σφαλμάτων στρογγύλευσης, δυο πραγματικοί που θα έπρεπε μαθηματικά να είναι ίσοι, πιθανότατα δεν είναι.