

Ανάλυση Σφάλματος στους Αριθμητικούς Υπολογισμούς (floating point)

$$x \rightarrow \underset{\text{αριθμός μηχανής}}{fl(x)}, \quad \left| \frac{x - fl(x)}{x} \right| \leq \frac{1}{2} \beta^{1-k \leftarrow \text{ψηφία στη mantissa}} \quad \text{ή} \quad |\varepsilon| \leq u$$

$$\text{Παρατήρηση: } (1-u)^n \leq 1+E \leq (1+u)^n \Leftrightarrow |E| \leq nu_1, \quad u_1 = 1.01u$$

1. Υπολογισμός του γινομένου n αριθμών x_1, \dots, x_n (σε παράσταση με κινητή υποδιαστολή(floating point))

Ως γνωστό $x \rightarrow fl(x) = x(1 + \varepsilon)$ (ε μονάδα μηχανής).

Υποθέτουμε ότι οι αριθμοί x_1, x_2, \dots, x_n είναι αριθμοί μηχανής τότε το υπολογιζόμενο γινόμενο τους είναι

$$P_n = fl(x_1 x_2 \dots x_n) = x_1 x_2 \dots x_n (1 + E) \stackrel{(1)}{=} x_1 x_2 \dots x_n + \underbrace{x_1 x_2 \dots x_n E}_{E'}$$

Ορίζουμε αναγωγικά τις ποσότητες:

$$P_1 = x_1$$

$$P_r = fl(P_{r-1} x_r) = P_{r-1} x_r (1 + \varepsilon_r), \quad r = 2, 3, \dots, n, \quad \text{όπου } |\varepsilon_r| \leq u$$

Με διαδοχικές αντικαταστάσεις στον ανωτέρω τύπο

για $r = 2, 3, \dots, n$ προκύπτουν:

$$P_2 = fl(P_1 x_2) = P_1 x_2 (1 + \varepsilon_2) = x_1 x_2 (1 + \varepsilon_2)$$

$$P_3 = fl(P_2 x_3) = P_2 x_3 (1 + \varepsilon_3) = x_1 x_2 x_3 (1 + \varepsilon_2)(1 + \varepsilon_3)$$

⋮

$$P_n = fl(P_{n-1} x_n) = P_{n-1} x_n (1 + \varepsilon_n) = x_1 x_2 \dots x_n (1 + \varepsilon_2)(1 + \varepsilon_3) \dots (1 + \varepsilon_n) =$$

$$= x_1 x_2 \dots x_n (1 + E), \quad \text{όπου } 1 + E = (1 + \varepsilon_2)(1 + \varepsilon_3) \dots (1 + \varepsilon_n)$$

$$\text{Αφού } |\varepsilon_r| \leq u \Leftrightarrow 1 - u \leq 1 + \varepsilon_r \leq 1 + u$$

$$\text{Συνεπώς } (1 - u)^{n-1} \leq 1 + E \leq (1 + u)^{n-1} \Rightarrow |E| \leq (n-1)u_1 = 1.01(n-1)u$$

Επομένως το σφάλμα στον υπολογισμό του γινομένου είναι:

$$\text{Σφάλμα} \equiv E' = x_1 x_2 \dots x_n E,$$

$$\text{Απόλυτο σφάλμα} \equiv |E'| = |x_1| |x_2| \dots |x_n| |E| \leq |x_1| |x_2| \dots |x_n| 1.01(n-1)u \text{ και}$$

$$\text{Απόλυτο Σχετικό Σφάλμα} \equiv \frac{|E'|}{|x_1| |x_2| \dots |x_n|} \leq 1.01(n-1)u$$

Άρα στον πολλαπλασιασμό δεν παίζει ρόλο η σειρά των παραγόντων για το σφάλμα.

2. Υπολογισμός εσωτερικού γινομένου δύο n-διάστατων διανυσμάτων

Έστω δύο n-διάστατα διανύσματα a και b

$$\left. \begin{array}{l} a = (a_1, a_2, \dots, a_n) \\ b = (b_1, b_2, \dots, b_n) \end{array} \right\} \Rightarrow ab \equiv \langle a, b \rangle = a_1 b_1 + \dots + a_n b_n$$

$$\text{Τότε είναι } fl(a_1 b_1 + \dots + a_n b_n) = a_1 b_1 (1 + \varepsilon_1) + \dots + a_n b_n (1 + \varepsilon_n) \quad (1)$$

Ορίζουμε αναγωγικά τις ποσότητες ip_r (internal product), και t_r ως εξής:

$$t_r = fl(a_r b_r)$$

$$ip_1 = t_1$$

$$ip_r = fl(ip_{r-1} + t_r), \quad r = 2, 3, \dots, n$$

οπότε έχουμε:

$$t_r = fl(a_r b_r) = a_r b_r (1 + \xi_r), \quad \text{όπου } |\xi_r| \leq u \quad (\xi_r : \text{το σφάλμα του } r \text{ γινομένου})$$

$$ip_1 = t_1$$

$$ip_r = fl(ip_{r-1} + t_r) = (ip_{r-1} + t_r)(1 + \eta_r), \quad \text{όπου } |\eta_r| \leq u \quad (\eta_r : \text{το σφάλμα του } r \text{ αθροίσματος})$$

Με διαδοχικές αντικαταστάσεις στον ανωτέρω τύπο για $r = 2, 3, \dots, n$ προκύπτουν:

$$\underline{r=1}: \quad t_1 = fl(a_1 b_1) = a_1 b_1 (1 + \xi_1)$$

$$ip_1 = t_1 = a_1 b_1 (1 + \xi_1)$$

$$\underline{r=2}: \quad t_2 = fl(a_2 b_2) = a_2 b_2 (1 + \xi_2)$$

$$ip_2 = fl(ip_1 + t_2) = (ip_1 + t_2)(1 + \eta_2) = [a_1 b_1 (1 + \xi_1) + a_2 b_2 (1 + \xi_2)](1 + \eta_2) + \\ = a_1 b_1 (1 + \xi_1)(1 + \eta_2) + a_2 b_2 (1 + \xi_2)(1 + \eta_2)$$

$$\underline{r=n}: \quad t_n = fl(a_n b_n) = a_n b_n (1 + \xi_n)$$

$$ip_n = fl(ip_{n-1} + t_n) = (ip_{n-1} + t_n)(1 + \eta_n) = a_1 b_1 (1 + \xi_1)(1 + \eta_2) \dots (1 + \eta_n) + a_2 b_2 (1 + \xi_2)(1 + \eta_2) \dots (1 + \eta_n) \\ + \dots + a_n b_n (1 + \xi_n)(1 + \eta_n) \quad (2)$$

Εξισώνοντας τις (1) και (2) προκύπτουν

$$1 + \varepsilon_1 = (1 + \xi_1)(1 + \eta_2) \dots (1 + \eta_n) \quad (3\alpha)$$

$$1 + \varepsilon_r = (1 + \xi_r)(1 + \eta_r) \dots (1 + \eta_n)^{(2)}, \quad r = 2, 3, \dots, n \quad (3\beta)$$

$$\text{Από την (3}\alpha\text{): } (1-u)^n \leq 1 + \varepsilon_1 \leq (1+u)^n \quad (3\alpha')$$

$$\text{Από την (3}\beta\text{): } (1-u)^{n-r+2} \leq 1 + \varepsilon_r \leq (1+u)^{n-r+2}, \quad r = 2, 3, \dots, n \quad (3\beta') \quad \left. \vphantom{\text{Από την (3}\beta\text{):}} \right\} \Rightarrow \text{ενοποίηση των (3}\alpha\text{) και (3}\beta\text{)} (\text{ενιαίο φραγμα})$$

$$\text{Από την (3}\alpha'\text{): } (1-u)^{n-1} \leq (1-u)^n \leq 1 + \varepsilon_1 \leq (1+u)^n \leq (1+u)^{n+1} \quad (\text{διότι } 1-u < 1, 1+u > 1) \quad (3\alpha'')$$

Τελικά από τις (3α'') και (3β') προκύπτει η ενοποιημένη ανισότητα

$$(1-u)^{n-r+2} \leq 1 + \varepsilon_r \leq (1+u)^{n-r+2}, \quad r = 1, 2, 3, \dots, n \quad (4)$$

Σύμφωνα με την παρατήρηση του λήμματος 3 προκύπτει:

$$|\varepsilon_r| \leq (n-r+2)1.01u \quad (5)$$

Από την (1) έχουμε

$$fl(a_1b_1 + \dots + a_nb_n) = a_1b_1(1 + \varepsilon_1) + \dots + a_nb_n(1 + \varepsilon_n) = a_1b_1 + \dots + a_nb_n + \underbrace{a_1b_1\varepsilon_1 + \dots + a_nb_n\varepsilon_n}_{\text{σφάλμα } E}$$

Επομένως το σφάλμα στον υπολογισμό του εσωτερικού γινομένου είναι:

$$\text{Απόλυτο Σφάλμα} \equiv |E| = |a_1b_1\varepsilon_1 + \dots + a_nb_n\varepsilon_n| \leq |a_1||b_1||\varepsilon_1| + \dots + |a_n||b_n||\varepsilon_n| \leq$$

$$\stackrel{(5)}{\leq} 1.01u\{(n+1)|a_1||b_1| + n|a_2||b_2| + \dots + 2|a_n||b_n|\} \leq$$

$$\leq 1.01u(n+1)(|a_1||b_1| + |a_2||b_2| + \dots + |a_n||b_n|) = 1.01u(n+1)|a|^T|b|$$

(συμβολικά $|ab| = |a|^T|b|$)

$$\text{Απόλυτο σχετικό σφάλμα} \equiv \frac{|fl(a^T b) - a^T b|}{|a^T b|} = \frac{|E|}{|a^T b|} \leq \frac{(n+1)1.01u|a|^T|b|}{|a^T b|}$$

Αν $|a^T b| \ll |a|^T|b|$ το απόλυτο σχετικό σφάλμα μπορεί να μην είναι "μικρό".

Ανάλυση Σφάλματος(Βιβλιογραφία)

Είναι αξιοσημείωτο να αναφερθεί ότι ο θεμελιωτής της θεωρίας σφάλματος είναι ο J. Wilkinson: με το βιβλίο του με τίτλο Rounding Error in Algebraic Processes.

Η ανάλυση σφάλματος διακρίνεται σε 1) **Forward**(εκ των προτέρων) και 2) **Backward**(εκ των υστέρων) ανάλυση σφάλματος.

Αν υποτεθεί ότι x είναι η είσοδος σε μία συνάρτηση f . Έστω \hat{y} η προσέγγιση της $y=f(\underset{\text{είσοδος}}{x})$ με ακρίβεια ε .

1) Forward ανάλυση: Εύρεση άνω φραγματος για το forward σφάλμα $|\Delta y| = |y - \hat{y}|$

2) Backward ανάλυση: Εύρεση άνω φραγματος για το backward σφάλμα $\Delta x: f(x+\Delta x) = \hat{y}$

Πρόβλημα καλώς ορισμένο (well posed)

1) Υπάρχει η λύση του

2) Είναι μοναδική

3) Εξαρτάται κατά συνεχή τρόπο από τα δεδομένα (ευστάθεια), δηλαδή μικρές μεταβολές στην είσοδο (δεδομένα) δεν παρουσιάζουν μεγάλες μεταβολές στην έξοδο (αποτελέσματα).

Πρόβλημα well-conditioned (καλής κατάστασης)

1) Καλώς ορισμένο

2) Έχει μικρό μεταδιδόμενο σφάλμα

Σχετικός αριθμός συνθήκης (δείκτης κατάστασης (condition number)):

$$\text{cond}_f(x) = \frac{\left\| \frac{f(\hat{x}) - f(x)}{\hat{x} - x} \right\|}{\left\| \frac{\hat{x} - x}{x} \right\|} = \frac{\frac{\|\Delta y\|}{\|y\|}}{\frac{\|\Delta x\|}{\|x\|}}$$

❖ Προβλήματα

➤ Καλώς Ορισμένα

▪ Well Conditioned ("μικρό" $\frac{\|\Delta y\|}{\|y\|} / \frac{\|\Delta x\|}{\|x\|}$)

▪ Ill conditioned

➤ Κακώς Ορισμένα

❖ Αλγόριθμοι

➤ Backward ευσταθείς ("μικρό" $\frac{\|\Delta x\|}{\|x\|}$)

➤ Ασταθείς

Ακριβής λύση αν : • Well Conditioned πρόβλημα • Backward ευσταθής αλγόριθμος