Επιστημονικός Υπολογισμός Ι
ΗΥ 343: ΔΙΑΛΕΞΗ 10

Ε. Γαλλόπουλος
Τμήμα Η/Υ & Πληροφορικής
Πανεπιστήμιο Πατρών

Εμπρός σφάλμα

Επομένως η πίσω ανάλυση δεν μπορεί να μας βοηθήσει. Προσέξτε όμως ότι το εμπρός σφάλμα μπορεί να εκτιμηθεί άμεσα!

$$|\mathrm{fl}(\gamma_{ij}) - \alpha_i\beta_j| = |\alpha_i\beta_j\delta_{ij}| \le |\alpha_i\beta_j|u$$

άρα

$$\frac{|\mathrm{fl}(\gamma_{ij}) - \alpha_i\beta_j|}{|\alpha_i\beta_j|} \le u$$

ΥΠΕΝΘΥΜΙΣΗ

Κλασική θεωρία κατάστασης - δείκτης κατάστασης προβλήματος

Ορισμός 3 (Rice' 66). *Έστω γραμμικοί νορμισμένοι[1] χώροι X, Y που διαθέτουν νόρμα και έστω η απεικόνιση $f : \Omega \subset X \to Y$, όπου Ω είναι ανοικτό χωρίο. Έστω x^* σταθερό και η τιμή $y^* := f(x^*)$. Υποθέτουμε ότι τα x^*, y^* δεν είναι τα μηδενικά στοιχεία των X, Y. Ο ασυμπτωτικός σχετικός δείκτης κατάστασης της απεικόνισης f στο x^* ως προς μικρές αλλαγές του x^* ορίζεται ως*

$$\mathrm{cond}(f; x^*) := \lim_{\delta \to 0} \sup_{\|h\|=\delta} \left\{ \frac{\frac{\|f(x^*+h)-f(x^*)\|}{\|f(x^*)\|}}{\frac{\|h\|}{\|x^*\|}} \right\}$$

εφόσον το όριο υπάρχει.

προσοχή στις νόρμες

Εμείς «απλοποιήσαμε» ως

$$\mathrm{cond}(f; x) := \frac{\|f'(x)\|}{\|f(x)\|} \|x\|$$

προσοχή στην παράγωγο (Ιακωβιανό μητρώο της f στο x)

[1] Λέγονται και σταθμισμένοι.

18

Όταν η συνάρτηση είναι γραμμική

- **Προσέξτε ότι αν η συνάρτηση είναι γραμμική, δηλ. f(x+h) = f(x)+f(h) τότε ο αριθμητής (εδώ χρησιμοποιούμε x αντί x*, απλοποιείται αρκετά:**

$$\frac{\|f(h)\|}{\|f(x)\|}$$

- **«υπενθυμίζουμε» ότι η νόρμα συνάρτησης (όπως και του μητρώου) ορίζεται ως**

$$\|f\| = \sup_{h \neq 0} \frac{\|f(h)\|}{\|h\|}$$

Οπότε για γραμμικές συναρτήσεις ο δ.κ. μπορεί να υπολογιστεί ως

$$\mathrm{cond}(f; x) = \|f\| \frac{\|x\|}{\|f(x)\|}$$

ΤΜΗΥΠ
ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ ΗΛΕΚΤΡΟΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ & ΠΛΗΡΟΦΟΡΙΚΗΣ

Πανεπιστήμιο Πατρών

$$\text{cond}(f;x) = \|f\|\frac{\|x\|}{\|f(x)\|}$$
$$f(x) := Ax \Rightarrow$$
$$\text{cond}(f;x) = \|A\|\frac{\|x\|}{\|Ax\|}$$

Επομένως αν υπάρχει A^{-1}

$$\sup_{\|x\|\neq 0} \text{cond}(f;x) = \sup_{\|y\|\neq 0} \|A\|\frac{\|A^{-1}y\|}{\|y\|} = \|A\|\|A^{-1}\|$$

Ακόμα και αν A δεν υπάρχει (π.χ. είναι m $\times$ n) ορίζουμε το δ.κ. με νόρμα-2 ως

$$\sup_{\|x\|\neq 0} \text{cond}(f;x) = \sup_{\|y\|\neq 0} \|A\|\frac{\|A^{\dagger}y\|}{\|y\|} = \|A\|\|A^{\dagger}\|$$

Σχόλια

Συμπίπτει με το δ.κ. του αντίστροφου προβλήματος (λύση συστήματος)

- **για τυχαία μητρώα, ο δ.κ. είναι μέτριος**
- **ειδικά μητρώα που ενδιαφέρουν στις εφαρμογές μπορεί να έχουν πολύ μεγάλο δείκτη**
  - **κατηγορίες μητρώων που ο δ.κ. αυξάνει ως n^p για σχετικά μεγάλο p**
    - **Vandermonde, Hilbert, Cauchy, …**
    - **μεγάλος δ.κ. υποδεικνύει ότι το μητρώο είναι «κοντά» σε μη αντιστρέψιμο μητρώο (μη αντιστρέψιμη απεικόνιση)**
    - **ο δείκτης κατάστασης για νόρμα-2 προκύπτει από το SVD**

$$\|A\|_2 = \sigma_{\max}(A), \|A^{-1}\|_2 = \frac{1}{\sigma_{\min}(A)}$$

Έστω $f([A; y]) := A^{-1}y$ όπου το A είναι ακριβές. Τότε

$$\sup_{\|h\|\neq 0} \frac{\|f(y+h) - f(y)\|}{\|h\|} = \sup_{\|h\|\neq 0} \frac{\|A^{-1}h\|}{\|h\|} = \|A^{-1}\|$$

Άρα

$$\begin{aligned} \mathrm{cond}(f; y) &\leq \frac{\|y\|}{\|A^{-1}y\|}\|A^{-1}\| \\ &\leq \frac{\|Ax\|}{\|x\|}\|A^{-1}\| \leq \|A\|\|A^{-1}\| \end{aligned}$$

Υπάρχουν A, y ώστε $\mathrm{cond}(f; y) = \|A\|\|A^{-1}\|$.

Παραδείγματα

Αν ο δείκτης κατάστασης είναι μεγάλος για το πρόβλημα τότε μεγαλώνει το άνω φράγμα για το εμπρός σφάλμα που σημαίνει ότι

μπορεί να υπάρξουν δεδομένα εισόδου για τα οποία το εμπρός σφάλμα να είναι μεγάλο

Έστω ότι για ένα πρόβλημα χρησιμοποιούμε

- «πίσω σταθερό» αλγόριθμο, π.χ. το πίσω σφάλμα είναι φραγμένο από $\phi\mathbf{u}$,
- αλλά το πρόβλημα έχει μεγάλο δείκτης κατάστασης, π.χ. 10^q

τότε για κάποια δεδομένα, το μπρος σφάλμα μπορεί να είναι της τάξης $O(\phi 10^q \mathbf{u})$.

- Μητρώα Hilbert έχουν κακό δείκτη κατάστασης.
- Μητρώα Vandermonde έχουν *συνήθως* κακό δείκτη κατάστασης.
- Αν ο δείκτης είναι μεγάλος, αυτό θα είναι εμφανές οποιαδήποτε νόρμα και να χρησιμοποιήσουμε.

μέγεθος n	$\log_{10} \kappa_2$ (rand(n))	$\log_{10} \kappa_2(\text{hilb}(n))$	$\log_{10} \kappa_2(\text{vander}([1:n]))$
4	1.5	4	3
8	2	10	8
16	2	18	23
32	3	19	55
100	3	19	(∞)

Δυσκολίες

Από τον προηγούμενο πίνακα φαίνεται ότι αν χρησιμοποιήσουμε έναν πίσω σταθερό αλγόριθμο για τη λύση συστήματος ή για πολλαπλασιασμό,

- με Hilbert το αποτέλεσμα μπορεί να είναι λάθος σε περισσότερα από 10 ψηφία αν $n \geq 10$
- με Vandermonde το αποτέλεσμα μπορεί να είναι λάθος σε περισσότερα από 8 ψηφία αν $n \geq 8$
- και με τα δύο μητρώα χάνουμε όλα τα ψηφία αν $n \geq 16$.

Παράδειγμα Έστω $x = ones(n,1)$ και $b = Ax$ υπολογισμένο ακριβώς με $A = \text{vander}([1:n])$. Τότε σε σύστημα MATLAB (α.κ.υ. IEEE με $u = 2^{-53} \approx 1.1 \times 10^{-16}$,

n	$\frac{\|x-A\backslash b\|_2}{\|x\|_2}$	κ_2
4	0	1.2ε+3
8	5.2ε-011	9.5ε+8
14	5.5 (μήνυμα: A σχεδόν μη αντιστρέψιμο)	3.3ε+19

Παράδειγμα Έστω $x = \text{ones}(n,1)$ και $b = Ax$ υπολογισμένο ακριβώς με $A = \text{rand}(n)$. Τότε σε σύστημα MATLAB (α.κ.υ. IEEE με $u = 2^{-53} \approx 1.1 \times 10^{-16}$,

n	$\frac{\|x - A\backslash b\|_2}{\|x\|_2}$	κ_2
8	8.2e-16	32
14	1.0e-015	77
100	1.3e-13	6.4e3

Προσοχή Υπάρχουν μητρώα Vandermonde με «τέλειο» δείκτη κατάστασης. Για παράδειγμα αν τα σημεία είναι ισοκατανεμημένα στο μοναδιαίο κύκλο (οπότε είναι το μητρώο για DFT).

Σχετικά με το σφάλμα στην πράξη MM

Όλες οι διατάξεις του κάμβου δίνουν το ίδιο σφάλμα:

$$C = AB \Rightarrow c_i = Ab_i, i = 1:n$$

επομένως

$$\tilde{c}_i = (A + \Delta A_i)b_i, \quad |\Delta A_i| \leq \gamma_n |A|$$

και μπορούμε να πούμε ότι

$$\text{fl}(AB) = (A + \Delta A)B, \quad |\Delta A| \leq \gamma_n |A||B||B^{-1}|$$

Με βάση το παραπάνω φράγμα μπορούμε να πούμε ότι η πίσω ευστάθεια των αλγόριθμων υπολογισμού του MM εξαρτάται από το μέγεθος του παράγοντα $|B||B^{-1}|$.

Επίσης

$$|C - \tilde{C}| \leq \gamma_n |A||B|.$$

Σημειώνουμε πως αυτό το φράγμα δεν μας πληροφορεί για το σχετικό σφάλμα! Όπως και με το εσωτερικό γινόμενο διανυσμάτων, το φράγμα προβλέπει μεγάλο σχετικό σφάλμα αν $|AB| \ll |A||B|$.

Χρήσιμος συμβολισμός

Μερικές φορές γράφουμε για συντομία

$$\prod_{i=1}^{n}(1+\delta_i)^{\rho_i} :=< n >$$

οπότε ισχύουν οι ιδιότητες

$$< n > \times < k > = < n+k >, \quad < n > / < k > = < n+k >$$

Παράδειγμα: Μέθοδος Horner

$s_n = \alpha_n$
for $k = n-1 : -1 : 0$
$s_k = x s_{k+1} + \alpha_k$
end

Χρησιμοποιώντας το γνωστό λήμμα και την προηγούμενη ορολογία:

$$\begin{aligned}
\hat{s}_{n-1} &= (x s_n < 1 > + \alpha_{n-1}) < 1 > = x\alpha_n < 2 > + \alpha_{n-1} < 1 > \\
\hat{s}_{n-2} &= (x s_{n-1} < 1 > + \alpha_{n-2}) < 1 > \\
\hat{s}_0 &= \alpha_0 < 1 > + \alpha_1 x < 3 > + \cdots \alpha_{n-1} x^{n-1} < 2n-1 > + \alpha_n x^n < 2n > \\
&= (1+\theta_1)\alpha_0 + \cdots (1+\theta_{2n})\alpha_n x^n
\end{aligned}$$

$$\hat{s}_0 = f\text{prog}(\alpha_0, ..., \alpha_n, x) = f(\alpha_0(1+\theta_1), ..., \alpha_n(1+\theta_{2n}), x)$$

επομένως αν θέσουμε

$$\begin{aligned}
X &= [\alpha_0, ..., \alpha_n, x] \\
X_{\text{prog}} &= [\alpha_0(1+\theta_1), ..., \alpha_n(1+\theta_{2n}), x]
\end{aligned}$$

$$|X_{\text{prog}} - X| \leq \gamma_{2n}|X| \Rightarrow \|X_{\text{prog}} - X\|_\infty \leq \gamma_{2n}\|X\|_\infty.$$

Το προς τα πίσω σφάλμα του αλγόριθμου Horner είναι μικρό και ο δείκτης κατάστασης του αλγορίθμου είναι $\gamma_{2n}/\mathbf{u}$.

Ειδικότερα:

> Αφού θ_{2n+1} είναι ο παράγοντας που χαρακτηρίζει τη μέγιστη ασάφεια σους συντελεστές, είναι ασφαλές να συμπεράνουμε ότι αν κάθε συντελεστής δεν είναι ακριβώς γνωστός αλλά γνωρίζουμε ότι $|\alpha_j - \hat{\alpha}_j| \leq \theta_{2n+1}|\alpha_j|$ για τον καθένα, τότε το σφάλμα που προκύπτει από την ασάφεια αυτή θα είναι σίγουρα μεγαλύτερο από οποιοδήποτε σφάλμα οφείλεται στις πράξεις α.κ.υ. που εκτελούνται από τον αλγόριθμο.

Αυτό από μόνο του δεν είναι αρκετό για να εγγυηθεί το σφάλμα των αποτελεσμάτων αλλά μας επιτρέπει να εξετάσουμε το προς τα εμπρός σφάλμα ανεξάρτητα από τις λεπτομέρειες του αλγορίθμου.

Εύκολα αποδεικνύεται η σχέση

$$\frac{|p(x) - \hat{s}_0|}{|p(x)|} \leq \gamma_{2n+1} \frac{\sum_{k=0}^{n} |\alpha_k||x|^k}{|p(x)|},$$

από την οποία προκύπτει ότι δεν μπορούμε να εγγυηθούμε μικρό προς τα εμπρός σφάλμα.

Δείκτης κατάστασης αναδρομής

- Αναδρομικές σχέσεις «φυτρώνουν» παντού στην Επιστήμη των Υπολογιστών και στα Μαθηματικά,
- Π.χ. Ανάλυση αλγορίθμων, υπολογισμοί συναρτήσεων, επίλυση διαφορικών εξισώσεων.
- Πολλές φορές η απλή εφαρμογή κάποιου τύπου μπορεί να οδηγήσει σε εντελώς λάθος αποτελέσματα!

Απλή αναδρομή - «απρόβλεπτη»!

$x_k=f(x_{k-1})$: f(x)=(n+1)x-1 ⇒ f(1/n)=1/n

x=1/n;for k=1:10, y(k)=(n+1)*x-1;x=y(k); end

Δοκιμάζουμε σε MATLAB τα παραπάνω

n=2; y(k)=0.5
n=3; y(k) = ??

Δείκτης κατάστασης αναδρομής

- **Θα εξετάσουμε τον υπολογισμό του**

$$I_n = \int_0^1 \frac{t^n}{t+5} dt$$

- **Λίγη ανάλυση δείχνει ότι ισχύει η αναδρομική σχέση:**

$$I_k \quad = \quad -5I_{k-1} + \frac{1}{k},$$

- **Με «αρχική συνθήκη»** $I_o = ln(6/5)$

```
y(1)=log(6/5); for j=1:28,y(j+1)=-5*y(j)+1/j;end
>> syms t
>> int(t^2/(t+5),0,1)
```

- Ο αναδρομικός υπολογισμός του I_n θεωρείται ως συνάρτηση από την τιμή I_0 (αρχική) στην I_n:

- $f_n (I_0) \rightarrow I_n$
- Αναζητούμε το δείκτη κατάστασης της συνάρτησης
- Προσέξτε ότι $\lim_{n\to\infty} I_n=0$

- Προσοχή, το I_0 δεν είναι αναπαραστήσιμο ακριβώς σε α.κ.υ. άρα ήδη η εκκίνηση γίνεται από παραπλήσιο I_0*
- Στη συνέχεια εξετάζουμε την ευαισθησία της f_n υποθέτοντας αριθμητική άπειρης ακρίβειας

ΤΜΗΥΠ

ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ ΗΛΕΚΤΡΟΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ & ΠΛΗΡΟΦΟΡΙΚΗΣ

Πανεπιστήμιο Πατρών

$$
\begin{aligned}
y_1 &= f_1(y_0) = -5y_0 + 1 \\
y_2 &= f_2(y_0) = -5y_1 + \frac{1}{2} = (-5)^2 y_0 - 5 + \frac{1}{2} \\
\cdots &\vdots \cdots \\
y_n &= f_n(y_0) = (-5)^n y_0 + p_n
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{cond}(f_n; y_0) &= \left|\frac{y_0 f_n'(y_0)}{y_n}\right| = \left|\frac{y_0(-5)^n}{y_n}\right| \\
&> 5^n
\end{aligned}
$$

ΤΜΗΥΠ

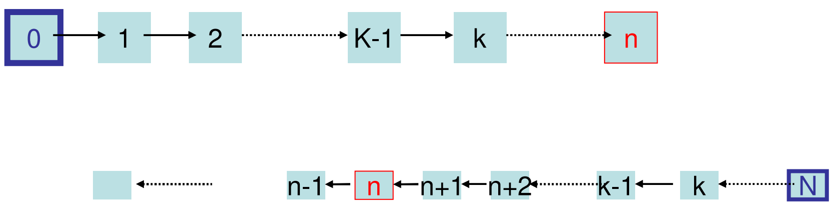
ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ ΗΛΕΚΤΡΟΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ & ΠΛΗΡΟΦΟΡΙΚΗΣ

Πανεπιστήμιο Πατρών

Υλοποίηση αναδρομής «προς τα πίσω»

$$I_k = -5I_{k-1} + \frac{1}{k} \Rightarrow I_{k-1} = -\frac{I_k}{5} + \frac{1}{5k}$$

- **Θα μπορούσαμε να χρησιμοποιήσουμε τη αντίστροφη αναδρομική σχέση**
- **… αλλά πώς ξεκινάμε αφού δεν γνωρίζουμε το I_N?**
- **ΙΔΕΕΣ**
  - **Μήπως το γνωρίζουμε?**
  - **Π.χ αν ξέρουμε το I_∞**
  - **ΑΞΙΟΠΕΡΙΕΡΓΟ: Μπορεί και να μην χρειαζεται!**



Όπως και πριν

$$\mathrm{cond}(g_n; y_N) = \left|\frac{y_N g_n^{'}(y_N)}{y_n}\right| = \left|\frac{y_N(-\frac{1}{5})^{N-n}}{y_n}\right|$$
$$< (\frac{1}{5})^{N-n} \quad \text{(τα } y_n \text{ φθίνουν)}$$

$$\frac{|I_n - g_n(I_N)|}{|I_n|} \leq \mathrm{cond}(g_n; I_N)\frac{|I_N - g_N(I_N)|}{|I_N|}$$

```
N=40;z(N)=0;for j=N:-1:2,z(j-1)=-z(j)/5+1/(5*j);end

>> syms t; int(t^10/(t+5),0,1)
```


Το πρόβλημα της άθροισης [Higham'02]

Δίδεται σύνολο αριθμών $\mathcal{S} = \{\xi_1, \dots, \xi_n\}$ και ζητούμε το άθροισμα $\mathrm{sum}(\mathcal{S})$.

- Μια από τις πιο συχνές πράξεις σε επιστημονικούς και εμπορικούς υπολογισμούς
- π.χ. συνοπτικά στατιστικά χαρακτηριστικά, π.χ. μέσος όρος, απόκλιση, νόρμα, ...
- Αξίζει τον κόπο μια πιο προσεκτική θεώρηση
- ... σε σχέση με την *ταχύτητα* και την *ακρίβεια*

Αναδρομική άθροιση

```
s=0
for i=1:n
  s = s+ ξ_i
end
```

Σχετικά με την αναδρομική άθροιση

- Πολύ απλή υλοποίηση
- Ιδιαίτερα χρήσιμη για ομόσημα στοιχεία μετά από ταξινόμηση σε αύξουσα σειρά κατ' απόλυτη τιμή
- Αποφυγή προβλημάτων «απορρόφησης» από μεγάλα στοιχεία.
- ... πρέπει να είναι γνωστά τα στοιχεία και να ταξινομήσουμε πριν την άθροιση.

Άλλες μέθοδοι

- Αναγωγική άθροιση ανά ζεύγη[4]
- Άθροιση καταχώρησης
- Αντισταθμισμένη άθροιση
- Αλγόριθμοι διύλυσης
- Άθροιση προσήμου: Υπολογισμος όλων των μερικών αθροισμάτων

$$S \to \{\xi_1, \xi_1+\xi_2, \xi_1+\xi_2+\xi_3, \cdots, \sum_{i=1}^{n} \xi_i\}$$

Χρήσιμα χαρακτηριστικά

- Είναι το σύνολο ομόσημο;
- είναι η ακολουθία διαθέσιμη από την αρχή;
- ποιό είναι το εύρος (μέγιστο, ελάχιστο);
- είναι η ακολουθία διατεταγμένη/ταξινομημένη;

[4] pairwise / cascade / fan-in

Αλγόριθμοι

Αναγωγική άθροιση Θεωρώντας ότι $n = 2^k$ όπου $k = \log_2 n$:

y = x;
for $i = 1 : \log_2 n$ for $i = 1 : 2^{i-1} : \frac{n}{2^i}$
$\psi_i = \psi_{2i-1} + \psi_{2i}$
end
end

- παράλληλη υλοποίηση - άθροιση σε δίκτυο διασύνδεσης τύπου δυαδικού δένδρου βάθους $\log_2 n$
- το φράγμα στο σφάλμα ανάλογο με το $\log_2 n$

Άθροιση καταχώρησης

1. ταξινόμηση του $\mathcal{S}$ κατ' απόλυτη τιμή σε αύξουσα σειρά $\mathcal{L} := \xi_1 \leq \xi_2 \leq \xi_3 \leq \cdots \leq \xi_n$
2. Διαγραφή του $\xi_1 + \xi_2$ απο την $\mathcal{L}$ και καταχώρηση στην $\mathcal{L}$ στην κατάλληλη θέση ώστε να παραμείνει μονοτονική. Επαναρίθμηση του $\mathcal{L}$ και αν περιέχει 2 ή περισσότερα στοιχεία, επιστροφή στο (2).

Δομή «πρόσθεσης-αντικατάστασης»:

Οι περισσότεροι αλγόριθμοι μοιάζουν με αλγόριθμοι καταχώρησης χωρίς τη μονοτονικότητα

Παρατηρήσεις

Οι αλγόριθμοι άθροισης μπορούν να αναπαρασταθούν με δυαδικό δένδρο, $\mathcal{T}$, όπου τα φύλλα είναι τα στοιχεία εισόδου και οι ($n-1$ εσωτερικοί) κόμβοι τα μερικά αθροίσματα. Η ρίζα θα περιέχει το τελικό αποτέλεσμα.

Το σφάλμα μπορεί να αναπαρασταθεί με τον ίδιο τρόπο, καθώς η άθροιση είναι γραμμική διαδικασία. Το τελικό σφάλμα έχει τη μορφή

$$|\sum_{j=1}^{n-1} \delta_i T_i| \leq \mathbf{u} \qquad \underbrace{\sum_{j=1}^{n-1} |T_i|}_{C(\mathcal{T}):=\text{άνω φράγμα για το μέγιστο σφάλμα}}$$

όπου T_i είναι οι υπολογισμένες τιμές σε κάθε κόμβο του δένδρου.
Προσοχή: Οι τιμές T_i εξαρτώνται από τον τρόπο που αθροίζουμε, δηλ. το δένδρο που αντιστοιχεί στην άθροιση. Το *βέλτιστο δένδρο άθροισης* *είναι το* $\mathcal{T}$ *που ελαχιστοποιεί το παραπανω* $C(\mathcal{T})$.

Θεώρημα 5. *[KaoWang'00] Ο υπολογισμός ενός βέλτιστου δένδρου άθροισης είναι NP-hard.* □

Περί της δυσκολίας

- **Δίνονται n αριθμοί (θετικοί και αρνητικοί) και έστω ότι κάθε τρόπος άθροισής τους επιφέρει ένα μέγιστο σφάλμα E_j, j=1, …**
- **Ονομάζουμε mE το μικρότερο από αυτά. Αν οι αριθμοί είναι θετικοί και αρνητικοί, τότε είναι NP-hard να βρούμε ποιός τρόπος άθροισης είναι αυτός που δίνει μέγιστο σφάλμα mE.**
- **Υπάρχουν βελτιώσεις (για φανατικούς):Ο τίτλος "Linear time approximation algorithms for computing numerical summation with provably small errors" τα λέει όλα! [Kao&Wang,2000]**

Συστάσεις

Εφόσον το πρόβλημα κατασκευής του βέλτιστου δένδρου άθροισης είναι πολύ δύσκολο:

- *χαλαρώνουμε* τις απαιτήσεις και αναζητούμε κάτι λιγότερο φιλόδοξο,
- *αξιοποιούμε* ό,τι πληροφορία υπάρχει για το $\mathcal{S}$.
- Αν χρειάζεται μεγάλη ακρίβεια, σκεφτείτε τη χρήση αντισταθμισμένης άθροισης ή άθροιση με εκτεταμένη ακρίβεια
- Για τις περισσότερες μεθόδους, το σφάλμα είναι στη χειρότερη περίπτωση ανάλογο με το n. Για πολύ μεγάλο n καθίσταται ενδιαφέρουσα η χρήση της αντισταθμισμένης άθροισης ή της αναγωγικής άθροισης
- Αν το σύνολο ομόσημο, όλες οι μέθοδοι προσφέρουν σχετικό σφάλμα το πολύ $n\mathbf{u}$ με καλύτερη την αντισταθμισμένη άθροιση. Η διατεταγμένη αύξουσα άθροιση
- Αν υπάρχει πιθανότητα σημαντικής (ως καταστροφικής) απαλοιφής λόγω μεικτών προσήμων, η αναδρομική άθροιση με φθίνουσα διάταξη είναι συνήθως καλύτερη.

Σχόλια και αδυναμίες του αριθμητικού μοντέλου και ανάλυσης σφάλματος με τα μοντέλα διάδοσης που χρησιμοποιήσαμε

- **Βάσεις και εργαλεία**
  - **Πρότυπο α.κ.υ. IEEE**

    Άλλες αριθμητικές και επεκτάσεις
  - **Μοντέλο διάδοσης σφάλματος «1+δ» και εμπρός ανάλυση**

    Μπορεί να οδηγήσει σε άστοχες/απαισιόδοξες προβλέψεις, δυσκολία εφαρμογής
  - **Πίσω ανάλυση σφάλματος**

    Δυσκολία υλοποίησης, κατανόησης και αδυναμία εφαρμογής
  - **Δείκτης κατάστασης προβλήματος**

    Προσεγγίσεις 1ης τάξης και ασυμπτωτικά φαινόμενα

Δεν προκύπτει πάντα σφάλμα!

Το παρακάτω θεώρημα είναι χρήσιμο σε ορισμένες εφαρμογές.

Θεώρημα 2. *Έστω ότι οι α.κ.υ. x, y ικανοποιούν τη σχέση $\frac{1}{2} \leq x/y \leq 2$ και ότι χρησιμοποιούμε α.κ.υ. με ακριβή στρογγύλευση (δηλ. υπάρχει ψηφίο προστασίας). Τότε, αν δεν προκύψει υποχείλιση, ο αριθμός $x - y$ υπολογίζεται χωρίς σφάλμα στρογγύλευσης και είναι ακριβώς αναπαραστήσιμος στο σύστημα $\mathcal{F}$.* □

Το χρησιμοποιούμε δημιουργώντας τέτοιες πράξεις με x,y δεδομένα εισόδου.

$$\begin{aligned} x^2 - y^2 &= (x+y)(x-y) \\ \sin(x) - \sin(y) &= \sin(\frac{x-y}{2})\cos(\frac{x+y}{2}) \\ e^x - e^y &= 2\sinh(\frac{x-y}{2})\exp(\frac{x+y}{2}) \end{aligned}$$

:

Chaining και **Fused Multiply and Add**

Chaining: 'Σύνδεση' πρόσθεσης και πολλαπλασιασμού
Φιλοσοφία RISC, Cray, κλπ.
IBM RS/6000 Fused Multiply and Add (FMA) :

$$fl(z + x \times y) = (z + x \times y)(1 + \delta)$$

Υποστηρίζεται στο Intel Itanium

Σε πολλές περιπτώσεις μπορούμε να εκφράσουμε υπολογισμούς μέσω FMA ώστε να έχουμε καλύτερο φράγμα για το σφάλμα (θέλει όμως κάποια προσοχή...)

ΤΜΗΥΠ
ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ ΗΛΕΚΤΡΟΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ & ΠΛΗΡΟΦΟΡΙΚΗΣ

Πανεπιστήμιο Πατρών

Σχετικά με την εντολή FMA

- Πλεονεκτήματα εντολής
  - Η πράξη εκτελείται με ένα σφάλμα στρογγύλευσης αντί για δύο
  - Είναι δυνατή η μείωση στην καθυστέρηση και στο υλικό που χρειάζεται για την εκτέλεση της πράξης με ξεχωριστές μονάδες άθροισης και πολλαπλασιασμού
  - Ο πολλαπλασιασμός και η πρόσθεση υλοποιούνται ως ειδικές περιπτώσεις της εντολής
- Προβλήματα ... προς επίλυση
  - Η καθυστέρηση μιας μόνον πράξης (πρόσθεση ή πολλαπλασιασμός) είναι μεγαλύτερη όταν υλοποιείται μέσω FMA
  - Αν υπάρχει μόνο μια μονάδα FMA αποκλείεται η ταυτόχρονη εκτέλεση πολλαπλασιασμού και άθροισης

ΤΜΗΥΠ
ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ ΗΛΕΚΤΡΟΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ & ΠΛΗΡΟΦΟΡΙΚΗΣ

Πανεπιστήμιο Πατρών

Α.Κ.Υ. σε επεξεργαστές Intel

- Οι πράξεις υλοποιούνται σε αριθμητική εκτεταμένης ακρίβειας **80 bits,** τα ενδιάμεσα αποτελέσματα μπορεί και αυτά να μείνουν σ'αυτή τη μορφή. Η στρογγυλοποίηση γίνεται όταν πρέπει να γίνει εγγραφή στη μνήμη.
- Ενδιάμεσα αποτελέσματα στους καταχωρητές παραμένουν σ' αυτή τη μορφή

- Δείτε την περιγραφή στις εργασίες για την α.κ.υ. στο **Itanium** (ημερολόγιο)

ΤΜΗΥΠ

ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ ΗΛΕΚΤΡΟΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ & ΠΛΗΡΟΦΟΡΙΚΗΣ

Πανεπιστήμιο Πατρών