Ονοματεπώνυμο: .... Α.Μ. ....· Έτος ...

ΕΠΙΣΤΗΜΟΝΙΚΟΣ ΥΠΟΛΟΓΙΣΜΟΣ Ι *Εξέταση Προόδου (5/12/06)*

ΚΑΛΗ ΕΠΙΤΥΧΙΑ!!! Κλειστές σημειώσεις, βιβλία, κινητά, υπολογιστές και βοηθήματα. Οι εκφωνήσεις επιστρέφονται με τις απαντήσεις σας. Αν διαπιστωθεί συνεργασία, «βαθμολογείστε» ανάλογα. Διαβάστε προσεκτικά τις εκφωνήσεις. Έχετε **2** *ώρες. Εκτός αν υπο/δηλώνεται διαφορετικά, τα διανύσματα είναι στήλες.*

ΠΡΟΣΟΧΗ: Να απαντήσετε περιεκτικά και συνοπτικά, στα παρακάτω ερωτήματα. Για πλήρη αξιολόγηση του γραπτού σας πρέπει να παρουσιάσετε όλο σας τον συλλογισμό και όλα τα ενδιάμεσα αποτελέσματα. Σε κάθε περίπτωση χρειάζεται επαρκής εξήγηση. Αν τα αποτελέσματα δεν προκύπτουν από τους υπολογισμούς στο γραπτό (πρόχειρους και μη), δεν βαθμολογείστε για την ερώτηση. Επίσης δεν βαθμολογείστε για τις ερωτήσεις Σωστό/Λάθος που απαντάτε απλά ΝΑΙ ή ΟΧΙ.

1. Σωστό ή Λάθος: Γενικά, αν ένας αλγόριθμος είναι πίσω σταθερός, τότε μικρές αλλαγές στα στοιχεία εισόδου του αλγορίθμου οδηγούν κατ' ανάγκη σε μικρές αλλαγές στο υπολογισμένο αποτέλεσμα.

   Απάντηση. ΛΑΘΟΣ: Η πίσω ευστάθεια εξασφαλίζει μόνον ότι το υπολογισμένο αποτέλεσμα του αλγορίθμου είναι ακριβώς ίδιο με το αποτέλεσμα που θα είχαμε αν χρησιμοποιούσαμε αριθμητική άπειρης ακρίβειας με στοιχεία εισόδου λίγο παραλλαγμένα (ή και ίδια). Για να εξασφαλιστεί το ζητούμενο της ερώτησης χρειάζεται το πρόβλημα (όχι ο αλγόριθμος) να έχει μικρό δείκτη κατάστασης. □

2. Σωστό ή Λάθος: Γενικά, οι πράξεις της κατηγορίας BLAS-2 έχουν πολύ μεγαλύτερη τοπικότητα από τις πράξεις των υπόλοιπων κατηγοριών.

   Απάντηση. ΛΑΘΟΣ: τη μεγαλύτερη τοπικότητα ($\mu_{\min} = O(1/n)$ για πρόβλημα με κυρίαρχη διάσταση n) την έχουν οι πράξεις BLAS-3 (πολλαπλασιασμός μητρώων, ανανεώσεις τάξης $k >> 1$, λύση τριγωνικών συστημάτων με πολλά δεξιά μέλη). □

3. Σωστό ή Λάθος: Γενικά, αν στο μοντέλο IEEE στρογγυλεύσουμε έναν αριθμό $x \in \mathbb{R}$ προς τον πλησιέστερο αριθμό κινητής υποδιαστολής και μ είναι η μονάδα στρογγύλευσης, ισχύει ότι $\mathrm{fl}(x) = x(1+\delta)$ όπου $\delta \leq \mu$.

   Απάντηση. ΛΑΘΟΣ, το σωστό είναι $\mathrm{fl}(x) = x(1+\delta)$ όπου $|\delta| \leq \mu$, ειδάλλως το δ θα μπορούσε να είναι πάρα πολύ μεγάλο σε απόλυτη τιμή. □

4. Αναφέρετε συνοπτικά 2 διαφορετικά επιχειρήματα που θα μπορούσαν να χρησιμοποιηθούν από τη Mathworks για να δικαιολογήσουν την αφαίρεση της συνάρτησης `flops` από πρόσφατες εκδοχές MATLAB.

   Απάντηση. α) Επειδή στους σύγχρονους επεξεργαστές ο χρόνος που χρειάζεται για την εκτέλεση μιας πράξης α.κ.υ. έχει γίνει πολύ μικρότερος από το χρόνο μεταφοράς των αντίστοιχων δεδομένων, θεωρείται ότι το πλήθος των πράξεων δεν είναι αρκετά ενδεικτικό του χρόνου υπολογισμού. β) Η τοποθέτηση μετρητών και υλοποίηση των μετρήσεων του πλήθους των πράξεων επιβαρύνει το συνολικό χρόνο εκτέλεσης. (Γνώμη του διδάσκοντα είναι ότι η απόφαση αφαίρεσης της εντολής δεν ήταν ορθή και ότι επιχειρήματα σαν τα παραπάνω μπορούν να καταρριφθούν). □

5. Έστω ότι χρησιμοποιούμε αριθμητική κινητής υποδιαστολής διπλής ακρίβειας (δηλ. $*.*\cdots*$ όπου υπάρχουν 52 bits μετά το δεκαδικό «.» ενώ το πρώτο ψηφίο είναι 1 αν ο αριθμός είναι κανονικοποιημένος, ειδάλλως 0 (αυτό μπορεί να υποδειχθεί από κάποιο status flag bit σε ειδικό καταχωρητή ή διαφορετικά και δεν αφορά στην άσκηση). Να μετρήσετε πόσοι *υποκανονικοποιημένοι θετικοί* αριθμοί κινητής υποδιαστολής μπορούν να αναπαρασταθούν υπό τις παραπάνω προϋποθέσεις.

   Απάντηση. Οι θετικοί υποκανονικοποιημένοι αριθμοί έχουν τη μορφή $0.*...*$ όπου τα bits δεξιά της υποδιαστολής δεν είναι όλα 0. Επομένως μπορούμε να αναπαραστήσουμε όλους τους δυαδικούς από το $0\ldots01$ ως το $1\ldots11$, δηλ. $2^{52}-1$ αριθμούς. □

6. Έστω ότι χρησιμοποιούμε αριθμητική IEEE και ότι στη μεταβλητή `realmin` έχουμε αποθηκεύσει τον ελάχιστο κανονικοποιημένο αριθμό κινητής υποδιαστολής. Να εξηγήσετε αν ο παρακάτω κώδικας επιστρέφει 1 ή 0 στη μεταβλητή `boole` (ένα μόνον είναι σωστό).

   ```
   temp2 = realmin/2; temp = 2*temp2; boole = (temp == realmin);
   ```

Απάντηση. Επιστρέφει 1: Η πρώτη διαίρεση έχει ως αποτέλεσμα τον πρώτο (μεγαλύτερο) υποκανονικοποιημένο αριθμό (συγκεκριμένα τον $0.100000 \times 2^{e_{\min}}$, όπου $e_{\min}$ είναι ο ελάχιστος εκθέτης) και ο επόμενος διπλασιασμός επαναφέρει τον `realmin`. □

7. Σε κάποιο πρόγραμμα, υπάρχουν οι εντολές: `s=0; for j = 1:4, s = s+ x(j)*x(j); end;` Το `x` περιέχει αποκλειστικά κοινούς αριθμούς κινητής υποδιαστολής (όχι ειδικούς αριθμούς). Να υπολογίσετε καλό άνω φράγμα για το εμπρός σχετικό σφάλμα στο τελικό αποτέλεσμα των παραπάνω εντολών υπό τον όρο ότι η υλοποίηση χρησιμοποιεί εντολές `FMA` και ότι δεν δημιουργείται υποχείλιση ή υπερχείλιση. Είστε ελεύθεροι να χρησιμοποιήσετε όποιον τρόπο ανάλυσης του σφάλματος προτιμάτε.

   Απάντηση. Θέτουμε $s = \sum_{j=1}^{4} x(j)^2$ και $\mathbf{u}$ τη μονάδα στρογγύλευσης (σε προηγούμενη ερώτηση συμβολίστηκε με μ). Θα έχουμε την εξής σειρά υπολογισμών σε α.κ.υ.

$$\hat{s} = \mathrm{fl}(\mathrm{fl}(\mathrm{fl}(\mathrm{fl}(0 + x(1) * x(1)) + x(2) * x(2)) + x(3) * x(3)) + x(4) * x(4))$$

   που με τη χρήση της `FMA` για κάθε υπολογισμό (χρησιμοποιώντας ως στοιχεία εισόδου τα τρέχοντα `s`, `x(j)` θα έχουμε:

$$\begin{aligned} \hat{s} &= (((x(1)^2(1+\delta_1) + x(2)^2)(1+\delta_2) + x(3)^2)(1+\delta_3) + x(4)^2)(1+\delta_4) \\ &= x(1)^2(1+\theta_4) + x(2)^2(1+\theta_3) + x(3)^2(1+\theta_2) + x(4)^2(1+\delta_4) \end{aligned}$$

   επομένως το σχετικό εμπρός σφάλμα φράσσεται ως εξής:

$$\begin{aligned} \frac{|\hat{s}-s|}{|s|} &= \frac{|x(1)^2\theta_4 + x(2)^2\theta_3 + x(3)^2\theta_2 + x(4)^2\delta_4|}{|s|} \\ &\leq \frac{|x(1)^2||\theta_4| + |x(2)^2||\theta_3| + |x(3)^2||\theta_2| + |x(4)^2||\delta_4|}{|s|} \\ &\leq \gamma_4 \frac{|x(1)^2| + |x(2)^2| + |x(3)^2| + |x(4)^2|}{|s|} \end{aligned}$$

   αλλά λόγω της θετικότητας των όρων $|x(j)^2|$ το άθροισμα στον αριθμητή είναι επίσης s, επομένως το εμπρός σφάλμα φράσσεται από το $\gamma_4 := \frac{4\mathbf{u}}{1-4\mathbf{u}}$. □

8. Να εξηγήσετε με σαφήνεια πώς μπορείτε να αξιοποιήσετε τα στοιχεία του x κατά την εκτέλεση του παραπάνω προγράμματος, ενδεχομένως προσθέτοντας επιπλέον εντολές ώστε να μειώσετε ακόμα παραπάνω το αναμενόμενο εμπρός σφάλμα στον παραπάνω υπολογισμό.

   Απάντηση. Επειδή κάθε επιμέρους όρος του αθροίσματος s είναι μη αρνητικός, είναι προτιμότερο να τους αθροίσουμε αφού ταξινομήσουμε τα στοιχεία του x κατ' αύξουσα απόλυτη τιμή αντί της δοθείσης διάταξης. Δηλαδή, αν $|x(j)| < |x(k)|$ για $j \neq k$, τότε στην άθροιση να προηγηθεί η πρόσθεση του $x(j)^2$ από το $x(k)^2$. Στο προηγούμενο φράγμα είδαμε ότι

$$\frac{|\hat{s}-s|}{|s|} \leq \frac{|x(1)^2||\theta_4| + |x(2)^2||\theta_3| + |x(3)^2||\theta_2| + |x(4)^2||\delta_4|}{|s|}$$

   επομένως, αν προταξινομήσουμε ώστε να ισχύει ότι $|x(1)^2| \leq |x(2)^2| \leq |x(3)^2| \leq |x(4)^2|$ το μέγιστο σφάλμα θα είναι μικρότερο από το μέγιστο σφάλμα της άθροισης χωρίς ταξινόμηση. □

9. Έστω η πράξη $\Delta = \Delta + x^\top (xy^\top)^p$, όπου $x, y \in \mathbb{R}^n$, p θετικός ακέραιος και Δ αρχικοποιημένος πίνακας (μπορείτε να συμπεράνετε το μέγεθός του από τα συμφραζόμενα). Να περιγράψετε αλγόριθμο βασισμένο αποκλειστικά σε πράξεις τύπου BLAS-1 για τον υπολογισμό του Δ. Για ευκολία, μπορείτε να εκφράσετε τις πράξεις μέσω των γνωστών εντολών MATLAB αλλά για την κάθε μια θα πρέπει να αναφέρετε σε ποια πράξη BLAS-1 αντιστοιχεί. Η απάντησή σας πρέπει να ελαχιστοποιεί το πλήθος των πράξεων αριθμητικής κινητής υποδιαστολής. Για το σκοπό αυτό, είστε ελεύθεροι να χρησιμοποιήσετε τις ιδιότητες της γραμμικής άλγεβρας.

   Απάντηση. Από τις διαστάσεις των δεδομένων, έπεται ότι το Δ θα είναι διάνυσμα γραμμή, μεγέθους n, με πραγματικά στοιχεία. Λόγω της προσεταιριστικής ιδιότητας, ο παραπάνω υπολογισμός μπορεί να ξαναγραφτεί ως

$$\Delta \quad = \quad \Delta + x^\top x (y^\top x)^{p-1} y^\top$$

επομένως ο αλγόριθμος μπορεί να εκτελεστεί ως εξής (γράφοντας για διευκόλυνση της επόμενης ερώτησης και το κόστος σε α.κ.υ.):

$\mathrm{temp}_1 1 \leftarrow x^\top x$	με _DOT	$\Omega_1 = 2n-1$
$\mathrm{temp}_2 \leftarrow y^\top x$	με _DOT	$\Omega_2 = 2n-1$
$\mathrm{temp}_1 \leftarrow \mathrm{temp}_1 \times (\mathrm{temp}_2)^{p-1}$	απλές πράξεις βαθμωτών	$\Omega_3 = 2$
$\Delta \leftarrow \Delta + \mathrm{temp}_1 y^\top$	με _AXPY	$\Omega_4 = 2n$

□

10. Σε συνέχεια του προηγούμενου ερωτήματος: Συμβολίζουμε, ως συνήθως, με $\mu_{\min} = \Phi_{\min}/\Omega$, όπου Ω είναι ο αριθμός πράξεων α.κ.υ. και $\Phi_{\min}$ ο ελάχιστος αριθμός μεταφορών μεταξύ της κύριας μνήμης και της κρυφής μνήμης και των καταχωρητών. Να υπολογίσετε το $\mu_{\min}$ ως συνάρτηση των p, n.

    Απάντηση. Θα έχουμε με βάση τον παραπάνω πίνακα για τα επιμέρους Ω ότι $\Omega = \sum_{j=1}^{4} \Omega_j = 6n$. Επίσης θα έχουμε $\Phi_{\min} = 4n+1$ οπότε $\mu_{\min} = \frac{4n+1}{6n} = \frac{2}{3} + O(\frac{1}{n})$. □