ΕΠΙΣΤΗΜΟΝΙΚΟΣ ΥΠΟΛΟΓΙΣΜΟΣ I *Σεπτέμβριος 2003*

Τα παρακάτω θέματα και οι απαντήσεις τους θα είναι χρήσιμα για όσους ήταν παρόντες στην εξέταση Σεπτεμβρίου

I. Να απαντήσετε σύντομα και περιεκτικά στα παρακάτω ερωτήματα (στις ερωτήσεις Σωστό/Λάθος πρέπει να δικαιολογήσετε τις απαντήσεις για να βαθμολογηθείτε).

1α) Έστω το σύστημα αριθμών κινητής υποδιαστολής με βάση β και t δυαδικά ψηφία στην ουρά. Να ορίσετε τη μονάδα στρογγύλευσης και να την υπολογίσετε ως συνάρτηση των β, t.

Επίλυση. (Κεφ. 3.2): Το μέγιστο σχετικό σφάλμα στρογγύλευσης για μη μηδενικούς αριθμούς. Το μέγιστο σχετικό σφάλμα για μη μηδενικό $z \in G$ είναι

$$\begin{aligned} \left|\frac{z - \mathrm{fl}(z)}{z}\right| &\leq \frac{\beta^{e-t}}{2}/\beta^{-1}\beta^{e} \\ &= \frac{\beta^{1-t}}{2} = u. \end{aligned}$$

□

1β) Έστω το σύστημα αριθμών κινητής υποδιαστολής IEEE με βάση $\beta = 2$ και t δυαδικά ψηφία στην ουρά, ένα από τα οποία είναι κρυμμένο. Να ορίσετε το έψιλον της μηχανής και να το υπολογίσετε ως συνάρτηση των β και t.

Επίλυση. (Κεφ. 3.2): Το έψιλον της μηχανής είναι η απόσταση της μονάδας από τον αμέσως επόμενο μεγαλύτερο α.κ.υ., έστω 1^+ Για το συγκεκριμένο σύστημα αυτός είναι ο $1^+ = 1 + 2^{1-t}$, επομένως $\epsilon_M = 2^{1-t}$. □

2α) Σωστό ή Λάθος: Αν δυο μητρώα είναι άνω Hessenberg τότε το γινόμενό τους θα είναι επίσης άνω Hessenberg.

Επίλυση. ΛΑΘΟΣ όπως μπορείτε να δείξετε με τυχαία γενικά μητρώα άνω Hessenberg μεγέθους 3. □

2β) Σωστό ή Λάθος: Αν δυο μητρώα είναι Toeplitz τότε το άθροισμά τους θα είναι επίσης Toeplitz.

Επίλυση. ΣΩΣΤΟ: Στα μητρώα Toeplitz τα στοιχεία κάθε διαγωνίου (κύριας και μη) είναι ίσα. Επομένως αν τα μητρώα A, B είναι Toeplitz και $C = A + B$ τότε $\gamma_{i,j} = \alpha_{i,j} + \beta_{i,j}$ αλλά $\alpha_{i,j} = \hat{\alpha}_{j-i}$ και $\beta_{i,j} = \hat{\beta}_{j-i}$ για στοιχεία $[\hat{\alpha}_{1-n}, \hat{\alpha}_{2-n}, ..., \hat{\alpha}_0, \hat{\alpha}_1, ..., \hat{\alpha}_{n-1}]$ και ομοίως $[\hat{\beta}_k]_{k=1-n}^{n-1}$. Επομένως $\gamma_{i,j} = \hat{\alpha}_{j-i} + \hat{\beta}_{j-i}$ και τα στοιχεία του C που βρίσκονται στις διαγωνίους είναι ίδια καθώς $\gamma_{i+s,j+s} = \hat{\alpha}_{j-i} + \hat{\beta}_{j-i}$. □

3α) Έστω ότι το υπολογιστικό σας σύστημα εκτελεί την ανανέωση πρώτης τάξης με μεγάλη ταχύτητα και ότι μπορείτε να δώσετε την εντολή για την ανανέωση μέσω μιας συνάρτησης MATLAB , έστω r1upd, ώστε εκτελώντας $B = \mathrm{r1upd}(A, x, y)$ να επιστρέφεται $B = A + xy^\top$. Να γράψετε αλγόριθμο που δοθέντων $A \in \mathbb{R}^{m\times n}$, $B \in \mathbb{R}^{n\times k}$, $C \in \mathbb{R}^{m\times k}$ να υπολογίζει το $C \leftarrow C + AB$ που να αξιοποιεί αυτή τη συνάρτηση για την ταχύτερη εκτέλεση του πολλαπλασιασμού.

Επίλυση. Αρκεί να εκμεταλλευτούμε τη γνωστή σχέση (βλ. βιβλίο ΕΥ και σημειώσεις παραδόσεων)

$$C = C + \sum_{i=1}^{n} A(1:m, i)B(i, 1:n).$$

Ο αλγόριθμος θα είναι for $i = 1 : n$, C = r1upd (C,A(1:m,i) B(i,1:k)') end □

3β) Σωστό ή Λάθος: Έστω οι ιδιοτιμές ενός συμμετρικού μητρώου, $|\lambda_1| \geq |\lambda_2| \geq \cdots \geq |\lambda_n|$, ταξινομημένες ως προς το μέτρο τους. Τότε ο δείκτης κατάστασης του μητρώου, ως προς την ευκλείδεια νόρμα, είναι ίσος με $\kappa_2(A) = |\lambda_1|/|\lambda_n|$.

Επίλυση. ΣΩΣΤΟ. Επαναλαμβάνουμε τη θεωρία (δεν χρειάζεται πλήρης ανάπτυξη για τη λύση). Ο δείκτης κατάστασης είναι $\kappa_2(A) = \|A\|_2\|A^{-1}\|$. Επειδή το μητρώο είναι συμμετρικό, διαγωνιοποιείται με ορθογώνιο μετασχηματισμό ομοιότητας Q ώστε $\Lambda = Q^\top AQ$. Επίσης, $\Lambda^{-1} = Q^\top A^{-1}Q$. Επίσης $\|A\|_2 = \|Q\Lambda Q^\top\|_2 = \|\Lambda\|_2 = |\lambda_1|$. Ομοίως, $\|A^{-1}\|_2 = \|Q\Lambda^{-1}Q^\top\|_2 = \|\Lambda^{-1}\|_2 = 1/|\lambda_n|$ και το αποτέλεσμα έπεται. □

4)Έστω η πράξη $B = B + (xy^\top)^p$, όπου $B \in \mathbb{R}^{n\times n}$, $x, y \in \mathbb{R}^n$ και p θετικός αριθμός. Συμβολίζουμε 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. Υποθέτοντας ότι η ύψωση ενός βαθμωτού σε δύναμη είναι ισοδύναμη με μια πράξη κινητής υποδιαστολής να υπολογίσετε το $\mu_{\min}$ ως συνάρτηση μόνον του n.

Επίλυση. Επειδή ζητούμε το $\mu_{\min}$ έχουμε αρκετή κρυφή μνήμη και καταχωρητές ώστε α) να φορτώσουμε αρχικά όλα τα δεδομένα από την κύρια μνήμη, β) να εκτελέσουμε τις πράξεις αποθηκεύοντας τα ενδιάμεσα αποτελέσματα στους καταχωρητές και γ) να εγγράψουμε τα αποτελέσματα στην κύρια μνήμη (βλ. Κεφ. 2 βιβλίου). Επομένως, $\Phi_{\min} = 2n^2 + 2n + 1$ (η μονάδα αντιστοιχεί στο p. Σε κάθε περίπτωση, όροι της τάξης $O(1)$ δεν πολυενδιαφέρουν). Για να υπολογίσουμε το Ω, θα αξιοποιήσουμε το γεγονός ότι το $(xy^\top)^p$ μπορεί να γραφτεί ως $x\overbrace{(y^\top x)\cdots(y^\top x)}^{p-1}y^\top$ επομένως εκτελούμε τις παρακάτω πράξεις: $\psi = y^\top x$ δηλ. $2n-1$ πράξεις, το $(\psi^{p-1}x)$, δηλ. $1+n$ πράξεις, και το $B = B + (\psi^{p-1}x)y^\top$, δηλ. $2n^2$ πράξεις. Συνολικά δηλαδή $\Omega = 2n^2 + 3n$ πράξεις (υπολογίζοντας την ύψωση σε δύναμη ως μια πράξη). Προσοχή ότι πρώτα πολλαπλασιάζουμε το ψ^{p-1} με το x και μετά υπολογίζουμε την ανανέωση πρώτης τάξης, αλλοιώς θα είχαμε n^2 επιπλέον πολλαπλασιασμούς. Σχετικά με τις μεταφορές, $\mu_{\min} = (2n^2+3n)/(2n^2+2n+1)$. Προσέξτε ότι χρειάστηκαν $O(n^2)$ θέσεις κρυφής μνήμης. Αξίζει να σημειωθεί ότι αν επιχειρήσετε να εκτελέσετε την πράξη με απλοϊκό τρόπο, π.χ. κόστους $O(n^3)$, δεν θα μπορέσετε να αποφύγετε την εξάρτηση από το p.

Η λύση του (4) στην εκδοχή $B = A + (Bxy^\top)^s$ γίνεται ανάλογα. Προσέξτε ότι επειδή η πράξη είναι $B = A + (Bxy^\top)^s$, χρειάζονται n^2 περισσότερες μεταφορές (για το A). Επίσης, στον υπολογισμό, η διάσπαση είναι $Bx(y^\top Bx)\cdots(y^\top Bx)y^\top$ ενώ για οικονομία πρέπει να υπολογίσουμε πρώτα το Bx και να το επαναχρησιμοποιήσουμε και στα εσωτερικά γινόμενα αντί να υπολογίσουμε το $y^\top B$. □

II. (*18 β.*) Ο παρακάτω κώδικας MATLAB μπορεί να υλοποιηθεί έτσι ώστε να παράγει το αποτέλεσμα ($Y \in \mathbb{R}^{n\times m}$) πολύ πιο γρήγορα. Υποθέτουμε ότι $1 < m < n$ και ότι η myvec (n) παράγει διάνυσμα (στήλη) από n άνισους αριθμούς. Να εξηγήσετε τί τροποποιήσεις πρέπει να γίνουν και να γράψετε τον ταχύτερο κώδικα, χρησιμοποιώντας, αν χρειάζεται, και άλλες γνωστές εντολές και συναρτήσεις της MATLAB .

```
function [Y] = vansolve(B);
[n,m]= size(B); x=myvec(n); V=[ ];
for i=1:n
 for j=1:n, V(i,j) = x(i)^(j-1); end
end
for k=1:m, Y(:,k) = V\B(:,k); end
```

Επίλυση. Το πρώτο τμήμα του κώδικα κατασκευάζει ανάστροφο μητρώο Vandermonde μεγέθους n με παράγον διάνυσμα το x. Το τελευταίο τμήμα του κώδικα επιλύει m συστήματα με το ίδιο συμμετρικό και θετικά ορισμένο μητρώο ($V^\top V$) με δεξιά μέλη τις στήλες του B. Για το πρώτο μέρος του κώδικα, μπορούμε να επιταχύνουμε α) προφορτώνοντας το μητρώο V, β) χρησιμοποιώντας αναδρομικό υπολογισμό αντί για ύψωση σε δύναμη, γ) χρησιμοποιώντας πιο έντονα διανύσματα στη διατύπωση του προγράμματος. Για το δεύτερο μέρος του κώδικα, αξιοποιούμε α) το γεγονός ότι πρόκειται για το ίδιο μητρώο, άρα μπορούμε να κάνουμε την παραγοντοποίηση μόνο μια φορά. Το μητρώο είναι αντιστρέψιμο (αλλά με πολύ κακό δείκτη κατάστασης εκτός από ορισμένες περιπτώσεις στοιχείων) γιατί τα στοιχεία του x είναι άνισα μεταξύ τους. Σημειώνουμε ότι υπάρχουν και άλλες πιο συνοπτικές υλοποιήσεις, π.χ.

```
x*ones(1,n).^(ones(n,1)*[0:n-1])
```

```
function [Y] = vansolve(B);
[n,m]= size(B); V=ones(n); x=myvec(n);
for j=2:n
    V(:,j) = V(:,j-1).*x;
end
[L,U]=lu(V);
Y = U\(L\B);
```

□

III. (*26 β.*) Να γράψετε αλγόριθμο που δοθέντος ενός μητρώου κάτω Hessenberg, $A \in \mathbb{R}^{n\times n}$, να εκτελεί παραγοντοποίηση LU χωρίς οδήγηση (επομένως, ο αλγόριθμος μπορεί και να αποτύχει αν τα δεδομένα δεν είναι κατάλληλα) αξιοποιώντας τη δομή του A και επιστρέφοντας τους παράγοντες L, U στις θέσεις του A. β) Να τροποποιήσετε τον αλγόριθμο με βάση τις εξής επιπλέον προδιαγραφές: Ο αλγόριθμος θα πρέπει να υπολογίζει σε κάθε ένα από τα $k = 1, ..., n-1$ βήματα το συντελεστή $\frac{\max\{\gamma_0,...,\gamma_{k-1}\}}{\gamma_0}$, όπου $\gamma_k := \max_{i,j} |\alpha_{ij}^{(k)}|$ είναι τα στοιχεία με μέγιστη απόλυτη τιμή στις θέσεις $(k+1:n, k+1:n)$ του μητρείου $L_k \cdots L_1 A$ που έχει προκύψει μετά από k βήματα της LU. Αν ο συντελεστής γίνει μεγαλύτερος από κάποιο προκαθορισμένο μέγεθος, έστω M, ο αλγόριθμος πρέπει να σταματά και να επιστρέφει στο χρήστη μήνυμα ότι πρέπει να χρησιμοποιηθεί οδήγηση. Ειδάλλως, ο αλγόριθμος πρέπει να επιστρέφει τους παράγοντες LU στις θέσεις του A. Μπορείτε να χρησιμοποιήσετε τις συναρτήσεις MATLAB error('MSG') που τυπώνει στην οθόνη το μήνυμα MSG και σταματά την εκτέλεση της συνάρτησης και max(X) που επιστρέφει το μέγιστο στοιχείο ενός διανύσματος X ή το διάνυσμα με στοιχεία το μέγιστο κάθε στήλης του μητρώου X.

Επίλυση. α) Ο αλγόριθμος παραγοντοποίησης - χωρίς οδήγηση - είναι ο ακόλουθος (προκύπτει άμεσα από την κλασική LU χωρίς οδήγηση εφόσον ληφθεί υπόψη η μηδενική δομή του μητρώου):

for $j = 1 : n-1$
  $A(j+1, j) = A(j+1, j)/A(j, j)$
  $A(j+1, j+1:n) = A(j+1, j+1:n) - A(j+1, j)A(j, j+1:n)$
end

β) Τροποποιούμε ως εξής:

$r = 1;$
for $j = 1 : n-1$
  if $r < M$
    $A(j+1, j) = A(j+1, j)/A(j, j)$
    $A(j+1, j+1:n) = A(j+1, j+1:n) - A(j+1, j)A(j, j+1:n)$
    $r = \mathtt{max}\left(\mathtt{max}(\mathtt{max}(\mathtt{abs}(A(j+1:n, j+1:n))))/s, r\right)$ □
  else
    `error`('Must use pivoting')
  end
end

IV. *(20 β.)* Έστω οι θετικοί αριθμοί κινητής υποδιαστολής $\alpha \geq \xi_1 \geq ..., \xi_n > 0$ και ο υπολογισμός

$y = 1;$
for $j = 1 : n$
  $y = y \cdot (\alpha + \xi_j)$
end

α) Να δείξετε ότι ο αλγόριθμος υπολογισμού είναι πίσω σταθερός. β) Να δείξετε ότι ο δείκτης κατάστασης του $y(\xi_1, ..., \xi_n, \alpha)$ ως προς τη νόρμα μεγίστου είναι φραγμένος από $P = \alpha \sum_{j=1}^{n} \frac{1}{\alpha + \xi_j}$.

Επίλυση. α) Εύκολα φαίνεται ότι

$$\mathrm{fl}(y) = \prod_{j=1}^{n} (\alpha + \xi_j) \prod_{j=1}^{n} (1 + \delta_j) \prod_{j=1}^{n-1} (1 + \delta_j^{'})$$

όπου δ_j είναι το σφάλμα στρογγύλευσης που προκύπτει από την πρόσθεση $\alpha + \xi_j$ και $\delta_j^{'}$ είναι το σφάλμα στρογγύλευσης που προκύπτει κατά τον πολλαπλασιασμό $y = \mathrm{fl}(y) \cdot \mathrm{fl}(\alpha + \xi_j)$. Μπορούμε να θεωρήσουμε ότι δεν προκύπτει σφάλμα κατά τον πρώτο πολλαπλασιασμό με το 1. Και στις δυο περιπτώσεις, τα σφάλματα είναι φραγμένα από τη μονάδα στρογγύλευσης $\mathbf{u}$. Από τη θεωρία (λήμμα) έχουμε ότι

$$\prod_{j=1}^{n} (1 + \delta_j) \prod_{j=1}^{n-1} (1 + \delta_j^{'}) = 1 + \theta_{2n-1}, \quad |\theta_{2n-1}| \leq \gamma_{2n-1} = \frac{(2n-1)\mathbf{u}}{1 - (2n-1)\mathbf{u}}$$

Επομένως,

$$\mathrm{fl}(y) \quad = \quad (1 + \theta_{2n-1}) \prod_{j=1}^{n} (\alpha + \xi_j)$$

$$
\begin{aligned}
&= (\alpha+\xi_1)(1+\theta_{2n-1})\prod_{j=2}^{n}(\alpha+\xi_j)\\
&= (\alpha'+\xi_1^{'})\prod_{j=2}^{n}(\alpha+\xi_j),
\end{aligned}
$$

όπου $\alpha' = \alpha(1+\theta_{2n-1})$ και $\xi_1^{'} = \xi_1(1+\theta_{2n-1})$. Επομένως, το υπολογισμένο αποτέλεσμα μπορεί να θεωρηθεί ως το ακριβές γινόμενο αλλά με στοιχεία $\xi_1^{'}, \xi_2, , ..., \xi_n, \alpha'$. Ο αλγόριθμος είναι πίσω ευσταθής γιατί $|\xi_1^{'} - \xi_1|/|\xi_1| \leq \gamma_{2n-1}$ και παρόμοια για το α.

β) Σύμφωνα με τη θεωρία (Κεφ. 3.5) ο δείκτης κατάστασης είναι

$$
\kappa(y) = \frac{\|[\xi_1, ..., \xi_n, \alpha]\|}{|y|}\|J\|
$$

όπου J είναι η Ιακωβιανή $J \in \mathbb{R}^{1\times(n+1)}$

$$
J = [\frac{\partial y}{\partial \xi_1}, \cdots, \frac{\partial y}{\partial \xi_n}, \frac{\partial y}{\partial \alpha}].
$$

Θα χρησιμοποιήσουμε τη νόρμα μεγίστου. Αλλά

$$
\frac{\partial y}{\partial \xi_j} = \prod_{i\neq j}^{n}(\alpha+\xi_i) := p_j
$$

ενώ

$$
\frac{\partial y}{\partial \alpha} = \sum_{j=1}^{n} p_j.
$$

Επομένως, η νόρμα μεγίστου $\|J\| = \max\{|p_1|, ..., |p_n|, |\sum_{j=1}^{n} p_j|\}$ και καθώς όλα τα στοιχεία είναι θετικά, $\|J\| = \sum_{j=1}^{n} p_j$. Επίσης η νόρμα μεγίστου $\|[\xi_1, ..., \xi_n, \alpha]\| = \alpha$ από αυτά που γνωρίζουμε για τα στοιχεία. Επομένως ο δείκτης κατάστασης είναι

$$
\begin{aligned}
\kappa(y) &= \frac{\alpha}{\prod_{j=1}^{n}(\alpha+\xi_j)}\sum_{j=1}^{n} p_j\\
&= \alpha\sum_{j=1}^{n}\frac{1}{\alpha+\xi_j}
\end{aligned}
$$

□