ΕΠΙΣΤΗΜΟΝΙΚΟΣ ΥΠΟΛΟΓΙΣΜΟΣ Ι (24 Φεβρ. 2008, 12-3μμ) ΕΠΙΛΕΓΜΕΝΕΣ ΑΠΑΝΤΗΣΕΙΣ

1. α) Σ - Λ : Οι εντολές BLAS-2 μπορούν να υλοποιηθούν να έχουν καλύτερη επίδοση από τις BLAS-3.

   Απάντηση. Λάθος : Οι εντολές BLAS-3 έχουν μικρότερο ελάχιστο αριθμό μεταφορών ανά πράξη α.κ.υ. από τις πράξεις BLAS «μικρότερων κατηγοριών». Επομένως, υπό την προϋπόθεση ότι αναφερόμαστε σε υλοποιήσεις που έχουν γίνει με στόχο την επίτευξη του μικρότερου λόγου μεταφορών προς πράξεις για κάθε κατηγορία, οι πράξεις BLAS-3 θα έχουν καλύτερη επίδοση (μετρούμενου με βάση τα Mflops). □

   β) Σ - Λ : Ξεδίπλωμα βρόχου γενικά χρησιμοποιείται για να μειώσει το πλήθος πράξεων α.κ.υ.

   Απάντηση. ΛΑΘΟΣ το ξεδίπλωμα δεν επιφέρει αλλαγή του Ω, μόνον ο βρόχος εκτελείται λιγότερες φορές αλλά με περισσότερες εντολές σε κάθε επανάληψη. □

   γ) Έστω στη MATLAB οι εκφράσεις $M + 20 - 10 - M$, $M + 20 - M - 10$, $M - 10 - M + 20$. Να εξηγήσετε τις τιμές που υπολογίζονται αν το M αρχικοποιηθεί ως `realmax`.

   Απάντηση. Το `realmax` της α.κ.υ. διπλής ακρίβειας είναι της μορφής $1.* \times 2^{1023}$ επομένως η προσθαφαίρεση αριθμών σαν το 10 και 20 με αυτό δεν επιφέρει καμια αλλαγή λόγω της απαιτούμενης κανονικοποίησης και επακόλουθου μηδενισμού τους κατά την πρώτη φάση της διαδικασίας. Επομένως τα αποτελέσματα θα είναι $((M + 20) - 10) - M = (M - 10) - M = M - M = 0$, $((M + 20) - M) - 10 = (M - M) - 10 = -10$, $((M - 10) - M) + 20 = (M - M) + 20 = 20$. □

   δ) Έστω αντιστρέψιμο $A \in \mathbb{R}^{n\times n}$ με μικρό δείκτη κατάστασης, $b \in \mathbb{R}^n$ και ο υπολογισμός $[L, U] = \mathtt{lu}(A); x = U\backslash(L\backslash b)$ (η MATLAB χρησιμοποιεί LAPACK). Ισχύει ή όχι ότι το εμπρός σφάλμα στο υπολογισμένο x δεν θα είναι μεγάλο ;

   Απάντηση. Για την LU γενικού μητρώου δεν μπορεί να αποδειχτεί μικρή πίσω ευστάθεια, που είναι απαραίτητη για να εγγυηθούμε μικρό εμπρός σφάλμα λόγω μικρού δείκτη κατάστασης, επομένως ΔΕΝ ΙΣΧΥΕΙ. Βασιζόμαστε και στον γνωστό τύπο (εμπρός σφ.) $<$ (πίσω σφ.) $\times$ (δείκτης κατ. A). □

2. Μας δίδονται α.κ.υ. και ένας αλγόριθμος για να τους αθροίσουμε. Να εξηγήσετε ποιοί από τους παρακάτω ισχυρισμούς είναι σωστοί και ποιοί λάθος :

   α) Αν αλλάξουμε τον αλγόριθμο άθροισης, μπορεί να αλλάξουν το πίσω σφάλμα και το εμπρός σφάλμα.

   β) Αν γνωρίζουμε τους α.κ.υ. και τον αλγόριθμο άθροισης, μπορούμε να υπολογίσουμε το *ακριβές* εμπρός σφάλμα.

   γ) Αν οι αριθμοί είναι ομόσημοι, ενας καλός τρόπος άθροισης είναι από το μικρότερο προς το μεγαλύτερο.

   δ) Αν η απόλυτη τιμή του υπολογισμένου αθροίσματος είναι πολύ μικρότερη του μέσου όρου των απολύτων τιμών των στοιχείων που αθροίστηκαν, μπορούμε να υποθέσουμε με ασφάλεια ότι το σχετικό εμπρός σφάλμα στο άθροισμα θα είναι και αυτό μικρό.

   Απάντηση. α) ΣΩΣΤΟ, και τα δυο εξαρτώνται από τον αλγόριθμο και επομένως τη σειρά άθροισης (εξάλλου το πίσω σφάλμα μετρά τον «δείκτη κατάστασης του αλγορίθμου».) β) ΛΑΘΟΣ, το ακριβές σφάλμα δεν μπορεί να υπολογιστεί γενικά γιατί χρειαζόμαστε αριθμητική άπειρης ακρίβειας. γ) ΣΩΣΤΟ, γιατί τότε μειώνεται η πιθανότητα σφάλματος από την πρόσθεση αριθμών που διαφέρουν πάρα πολύ σε μέγεθος που θα είχε για συνέπεια μηδενισμό των μικρότερων λόγω κανονικοποίησης των εκθετών. Επίσης τα «δ» που συσσωρεύονται στη διάδοση του σφάλματος επιβαρύνουν περισσότερο τους μικρότερους όρους του αθροίσματος. δ) ΛΑΘΟΣ : Τυπικό παράδειγμα $(1+\delta_1)-(1-\delta_2) = \delta_1+\delta_2$ όπου τα δ_j είναι πολύ μικρά και περιέχουν κυρίως «θόρυβο» από προηγούμενες πράξεις. Κλασικό παράδειγμα που δημιουργείται πρόβλημα από καταστροφικη απαλοιφή. □

3. Δίδονται τα στοιχεία $A \in \mathbb{R}^{10\times m}$ και $b \in \mathbb{R}^m, c \in \mathbb{R}^{10}$ και θέλουμε να υπολογίσουμε το $y \leftarrow c + Ab$. Το n δεν έχει κανέναν περιορισμό. α) Ποιό είναι το $\Phi_{\min}$ για την πράξη ; β) Να δείξετε πώς μπορείτε να υλοποιήσετε τον πολλαπλασιασμό με $\Phi = \Phi_{\min}$ χρησιμοποιώντας κρυφή μνήμη και καταχωρητές $O(1)$ (δηλ. προσωρινή μνήμη άμεσης προσβασης μεγέθους ανεξάρτητου του m).

   Απάντηση. α) Με απλή καταμέτρηση των α.κ.υ. εισόδου/εξόδου που χρησιμοποιούνται στον υπολογισμό, έχουμε $10m$ για φόρτωση του A, $m + 10$ για φόρτωση των c, b, και 10 για την αποθήκευση στο y, συνολικά δηλ. $\Phi_{\min} = 11m + 20$. β) Η σχετική ύλη υπάρχει και στις διαφάνειες. Συνοψίζουμε λέγοντας ότι η υλοποίηση μπορεί να κωδικοποιηθεί ως εξής, εφόσον διατίθεται χώρος για την αποθήκευση σε καταχωρητές και cache της τάξης του $O(1)$. Η μεταβλητή `temp` έχει αναφέρεται σε καταχωρητές μήκους 10.

```
LOAD c
for j = 1 : m
  LOAD b(j)
  for i = 1 : 10
   LOAD A(i, j)
   temp(i) = c(i) + A(i, j) * b(j)
  end
end
STORE y = temp
```

$\square$

4. α) Τι θα εμφανιστεί στην οθόνη αν εκτελέσετε τις παρακάτω εντολές σε περιβάλλον MATLAB και `n=3`:
`for j=1:n, A = kron(ones(j,1),[1:j]), end`

Απάντηση.

$A = 1$

$$A = \begin{pmatrix} 1 & 2 \\ 1 & 2 \end{pmatrix}$$

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \\ 1 & 2 & 3 \end{pmatrix} \quad \square$$

Υπενθυμίζουμε ότι η εντολή `kron(A,B)` επιστρέφει το γινόμενο Kronecker $A \otimes B$.

β) Να ενθέσετε (αιτιολογώντας, πάντα) σε επιπλέον κώδικα που να υπολογίζει όσο μπορείτε πιο αξιόπιστα (επιστρέφοντας σε κάποια μεταβλητή) τα Mflop/s των παραπάνω εντολών στο υπολογιστικό σας περιβάλλον. Μπορείτε να υποθέσετε ότι άν $A \in \mathbb{R}^{m_A \times n_A}$, $B \in \mathbb{R}^{m_B \times n_B}$ τότε το κόστος του `kron(A,B)` είναι $\Omega = m_A n_A m_B n_B$.

Απάντηση. Για συντομία συμβολίζουμε με $\mathbf{\Delta}$ τις εντολές `for j=1:n, A = kron(ones(j,1),[1:j]), end`. Προσέξτε ότι το Ω θα είναι $\sum_{j=1}^{n} j^2$. Μπορεί να υπολογιστεί από κλασικούς τύπους αθροισμάτων προόδων ή στο πρόγραμμα, συσσωρεύοντας τις πράξεις κάθε επανάληψης σε μεταβλητή. Τότε

```
· % εκτέλεση για να αποφευχθεί «θόρυβος» από την αρχικοποίηση
tic; for j=1:itmax, Δ; end;
optime = toc/itmax; ops = 0;
for j=1:itmax, ops = ops+j*j;  end; mflops = ops*1e-6/toc;
```

$\square$

5. α) Είναι το μοντέλο διάδοσης του σφάλματος στον πολλαπλασιασμό κινητής υποδιαστολής, $x \tilde{\times} y = x \times y(1+\delta)$ όπου $|\delta| \le \mathbf{u}$, $\mathbf{u}$ η μονάδα στρογγύλευσης και x, y αριθμοί κινητής υποδιαστολής, άμεσο επακόλουθο της «αρχής ακριβούς στρογγύλευσης»; Αν ναι, να το δείξετε, αν όχι να εξηγήσετε γιατί.

Απάντηση. ΕΙΝΑΙ: Η αρχή προσδιορίζει ότι με τις παραπάνω συνθήκες, για τον πολλαπλασιασμό ισχύει ότι η πράξη που εκτελείται στη μηχανή έχει ως αποτέλεσμα την ποσότητα που θα υπολογιζόταν με αριθμητική άπειρης ακρίβειας (δηλ. το $x \times y$) με στρογγύλευση (υποθέτουμε προς το πλησιέστερο) μετά, επομένως το τελικό αποτέλεσμα θα είναι $x \times y(1+\delta)$ όπου $|\delta| \le \mathbf{u}$. $\square$

β) Γνωρίζουμε ότι ο κλασικός δείκτης κατάστασης ενός μητρώου ως προς την επίλυση συστήματος $Ax = b$ ορίζεται ως $\kappa(A) := \|A\|\|A^{-1}\|$ για επιλεγμένη νόρμα. Να δείξετε ένα μητρώο 3×3 για το οποίο το $\kappa(A)$ είναι πάρα πολύ μεγάλο και το υπολογισμένο $\tilde{x}$ να έχει συγκριτικά πολύ μικρό σχετικό σφάλμα.

Απάντηση. Μπορείτε να διαλέξετε ένα διαγώνιο μητρώο A, με διαγώνιο $[1, 1, 1e-10]$, οπότε ο δείκτης κατάστασης είναι $1e10$. Από την άλλη, αν λύσετε το σύστημα $Ax = b$, λόγω της διαγώνιας δομής του A, κάθε στοιχείο της λύσης x υπολογίζεται με μια διαίρεση, επομένως το άνω φράγμα για το σχετικό σφάλμα κάθε στοιχείου της υπολογισμένης λύσης $\tilde{x}$ θα είναι $\mathbf{u}$. $\square$

6. α) Έστω ότι ένα μητρώο $H \in \mathbb{R}^{n\times n}$ έχει μηδενικά στις θέσεις που βρίσκονται κάτω από την πρώτη υποδιαγώνιο, δηλ. $(3:n,1),(4:n,2),...,(n,n-1)$. Να δείξετε ότι (χωρίς οδήγηση και εφόσον υπάρχει) η παραγοντοποίηση LU του H κοστίζει $\Omega = \alpha n^2 + O(n)$. Επίσης να υπολογίσετε τον κυρίαρχο συντελεστή α.

Απάντηση. Προσέχουμε ότι σε κάθε βήμα $k = 1, ..., n-1$ της κλασικής απαλοιφής, χρειάζεται να απαλείψουμε μόνον ένα υποδιαγώνιο στοιχείο (στη θέση $(k+1,k)$). Επομένως το κόστος θα είναι $\Omega = \sum_{k=1}^{n-1}(1+\sum_{j=k+1}^{n} 2)$ επομένως $\Omega = n(n-1)+O(n)$ άρα $\alpha = 1$. Ο κώδικας μπορεί να είναι ο εξής (προαιρετικά):

```
for k=1:n-1
  H(k+1,k) = H(k+1,k)/H(k,k)
  for j=k+1:n
    H(k+1,k+1:n) = H(k+1,k+1:n) - H(k+1,k)*H(k+1,k+1:n)
  end
end
```

□

β) Δίδεται $A = \begin{pmatrix} 1 & 1 & 2 & 1 \\ 0 & 2 & 1 & -1 \\ 0 & 3 & -1 & 1 \\ 0 & 4 & 1 & 2 \end{pmatrix}$.

Να υπολογίσετε διάνυσμα Householder ώστε ο (ορθογώνιος) ανακλαστής P που παράγεται από το διάνυσμα, να μηδενίζει τη θέση $(4,2)$ του μητρώου PA καθώς επίσης και του $B = PAP^\top$. Επίσης να υπολογίσετε το B (να φέρετε σε πέρας όλες τις αριθμητικές πράξεις.) Προσοχή: Δεν χρειάζεται (δεν είναι εφικτό) να είναι 0 το στοιχείο στη θέση $(3,2)$.

Απάντηση. Σε MATLAB, $\texttt{u} = [0;0;\texttt{A}(3:4,2)] + [0,0,1,0]' * \texttt{norm}(\texttt{A}(3:4,2))$, επομένως $u = [0,0,8,4]^\top$ και υπολογίζεται ότι

$$B = \begin{pmatrix} 1 & 1 & -2 & -1 \\ 0 & 2 & 0.2 & -1.4 \\ 0 & -5 & 1.88 & -1.16 \\ 0 & 0 & -1.16 & -0.88 \end{pmatrix}$$

□

γ) Για κάθε A, μπορεί να υπολογιστεί (π.χ. η συνάρτηση `hess` στη MATLAB) ορθογώνιο μητρώο Q ως γινόμενο ανακλαστών Householder, ώστε το $QAQ^\top$ να έχει μηδενικά κάτω από την υποδιαγώνιο. Ο υπολογισμός των Q και $QAQ^\top$ κοστίζουν συνολικά περί τις $5n^3$ πράξεις α.κ.υ. Έστω ότι χρειάζεται να υπολογίσετε τις λύσεις $x_j, j = 1, ..., s$ των s συστημάτων $(A-\omega_j I)x_j = b_j$ όπου $A \in \mathbb{R}^{n\times n}$ και τα ω_j είναι πραγματικοί αριθμοί τέτοιοι ώστε τα μητρώα $A-\omega_j I$ να είναι αντιστρέψιμα και I το ταυτοτικό μητρώο. Να περιγράψετε τα βασικά βήματα αλγορίθμου που επιτυγχάνει τη λύση των s συστημάτων με κόστος $\Omega \approx 5n^3 + O(sn^2)$ αντί για $O(sn^3)$ που θα στοίχιζε αν χρησιμοποιούσατε απευθείας LU.

Απάντηση. ΒΙΒΛΙΟ □

7. Δίδεται η διαφορική εξίσωση $u''(x)+10^{-2}(20-u) = 0$ στο διάστημα $[0,10]$ με συνοριακές συνθήκες $u(0) = 40, u(10) = 200$ και θέλουμε να προσεγγίσουμε τη λύση με κεντρισμένες πεπερασμένες διαφορές και ακρίβεια τάξης $O(h^2)$, όπου h είναι η απόσταση μεταξύ των ισαπέχοντων κόμβων του πλέγματος που θα χρησιμοποιήσουμε στη διακριτοποίηση.

α) Να εξηγήσετε σύντομα γιατί συνήθως απαιτούμε από τη συνάρτηση $u(x)$ να έχει παραγώγους μέχρι και 4ης τάξης και αυτές να είναι συνεχείς στο διάστημα $[0,10]$.

Απάντηση. Από τη θεωρία γνωρίζουμε ότι η διακριτοποίηση βασίζεται στο συνδυασμό τιμών της συνάρτησης σε επιλεγμένους (γειτονικούς) κομβους του πλέγματος και στα σχετικά αναπτύγματα Taylor. Ειδικότερα, υπό την προϋπόθεση ότι η u διαθέτει τουλάχιστον 4 παραγώγους και συμβολίζοντας με u_j την τιμή της συνάρτησης στον κόμβο j ενός φυσικά αριθμημένου πλέγματος, μπορούμε να γράψουμε

$$u_{j\pm1} = u_j \pm hu_j^{(1)} + \frac{h^2}{2}u_j^{(2)} \pm \frac{h^3}{6}u_j^{(3)} + \frac{h^4}{24}u^{(4)}(x_i + \theta_i^{\pm}h)$$

όπου $-1 < \theta_i^- < 0 < \theta_i^+ < 1$. Επομένως

$$u_{j-1} + u_{j+1} - 2u_j \quad = \quad h^2 u_j^{(2)} + \frac{h^4}{24}\left(u^{(4)}(\xi_j + \theta_i^+ h) + u^{(4)}(\xi_j + \theta_i^- h)\right)$$

Επομένως, το σφάλμα διακριτοποίησης της 2ης παραγώγου σε κάθε σημείο εξαρτάται άμεσα από την διακριτοποίηση (δηλ. το h) και τη διακύμανση της τιμής του $|u^{(4)}|$. Το h το επιλέγεται από εμάς, επομένως μπορούμε να το επιλέξουμε όσο μικρό θέλουμε (μόνος περιορισμός είναι το μέγεθος του προκύπτοντος συστήματος) για να πετύχουμε αποδεκτό σφάλμα. Όμως, παράλληλα, θα πρέπει να αποκλείσουμε την περίπτωση να γίνεται το h πολύ μεγάλο. Αυτό εξασφαλίζεται «αυτόματα» όταν η συνάρτηση $u^{(4)}$ είναι συνεχής στο κλειστό διάστημα ορισμού της, καθώς τότε, από γνωστό στοιχειώδες θεώρημα της Μαθηματικής Ανάλυσης, έπεται ότι το $|u^{(4)}|$ θα είναι φραγμένο σε όλο το διάστημα. □

β) Να υπολογίσετε μητρώο $A \in \mathbb{R}^{4\times 4}$ και δεξιό μέλος $b \in \mathbb{R}^4$ τέτοια ώστε το διάνυσμα g που ικανοποιεί το σύστημα $Ag = b$ να προσεγγίζει τη λύση u στους κόμβους.

Απάντηση. Διαμερίζουμε το διάστημα [0,10] σε 4 ισαπέχοντες εσωτερικούς κόμβους επομένως $h = 10/5 = 2$ και οι κόμβοι θα είναι $\xi_j = jh$ για $j = 1, ..., 4$. Χρησιμοποιώντας κεντρισμένες πεπερασμένες διαφορές 2ης τάξης για την προσέγγιση της 2ης παραγώγου θα έχουμε

$$\frac{u(\xi_{j-1}) - 2u(x_j) + u(\xi_{j+1})}{h^2} + 20 \times 10^{-2} - 10^{-2}u(x_j) = 0$$

επομένως οι εξισώσεις σε κάθε σημείο καθορίζονται από τον τύπο

$$\frac{1}{h^2}U_{j-1} - (\frac{2}{h^2} + 10^{-2})U_j + \frac{1}{h^2}U_{j+1} = -20 \times 10^{-2}$$

που ξαναγράφουμε ως

$$-\frac{1}{4}U_{j-1} + (\frac{1}{2} + 10^{-2})U_j - \frac{1}{4}U_{j+1} = 20 \times 10^{-2}$$

Επομένως το σύστημα θα είναι

$$\begin{pmatrix} 0.51 & -0.25 & 0 & 0 \\ -0.25 & 0.51 & -0.25 & 0 \\ 0 & -0.25 & 0.51 & -0.25 \\ 0 & 0 & -0.25 & 0.51 \end{pmatrix} \begin{pmatrix} U_1 \\ U_2 \\ U_3 \\ U_4 \end{pmatrix} = \begin{pmatrix} 10.2 \\ 0.2 \\ 0.2 \\ 50.2 \end{pmatrix}$$

□

γ) Έστω ότι η παραπάνω διαφορική εξίσωση τροποποιείται σε $u''(x) + 10^{-2}(20 - u) - (1 + x^2) = 0$. Ποιοί θα είναι τώρα οι νέοι παράγοντες A και b;

Απάντηση. Για να ληφθεί υπόψη ο νέος παράγοντας $1 + x^2$, διαφοροποιείται μόνον το δεξιό μέλος: $b = [15.2, 17.2, 37.2, 115.2]^\top$. □

δ) Στη συνέχεια, αλλάζουμε τη συνοριακή συνθήκη του αρχικού προβλήματος (δηλ. του μέρους (α)) από $u(0) = 40$ σε $u'(0) = -2$. Χρησιμοποιώντας κεντρισμένες πεπερασμένες διαφορές 2ης τάξης να γράψετε το νέο σύστημα που θα προκύψει, έστω $\hat{A}\hat{g} = \hat{b}$. Προσοχή: Τα $\hat{A}, \hat{b}$ μπορεί να έχουν διαφορετικό μέγεθος από πριν.

Απάντηση. Με την αλλαγή αυτή δεν γνωρίζουμε πλέον το $u(0)$ αλλά την παράγωγο την οποία προσεγγίζουμε ως

$$\frac{U_1 - U_{-1}}{2h} \approx u'(0) = 40 \Rightarrow U_{-1} = U_1 - 160$$

θεωρώντας ότι U_{-1} είναι προσέγγιση του u στο -2. Επίσης, γράφουμε την εξίσωση για το σημείο 0, δηλ.

$$-\frac{1}{4}U_{-1} + (\frac{1}{2} + 10^{-2})U_0 - \frac{1}{4}U_1 = 20 \times 10^{-2}$$

οπότε

$$-\frac{1}{4}(U_1 - 160) + (\frac{1}{2} + 10^{-2})U_0 - \frac{1}{4}U_1 = 20 \times 10^{-2}$$

άρα επαυξάνουμε το αρχικό σύστημα ως εξής:

$$\begin{pmatrix} 0.51 & -0.5 & 0 & 0 & 0 \\ -0.25 & 0.51 & -0.25 & 0 & 0 \\ 0 & -0.25 & 0.51 & -0.25 & 0 \\ 0 & 0 & -0.25 & 0.51 & -0.25 \\ 0 & 0 & 0 & -0.25 & 0.51 \end{pmatrix} \begin{pmatrix} U_0 \\ U_1 \\ U_2 \\ U_3 \\ U_4 \end{pmatrix} = \begin{pmatrix} -39.8 \\ 0.2 \\ 0.2 \\ 0.2 \\ 50.2 \end{pmatrix}$$

□

8. Έστω η διαφορική εξίσωση $u'''(t) = -1000u(t) - 300u'(t) - 30u''(t)$ με αρχικές τιμές $u(0) = 1, u'(0) = 0, u''(0) = 1.$ α) Να υπολογίσετε το $u(1.6)$ χρησιμοποιώντας εμπρός Euler και βήμα $h = 0.8.$ (Προσοχή: Η εξίσωση είναι 3ης τάξης και είναι προτιμότερο να την μετατρέψετε σε γραμμικό σύστημα συνήθων διαφορικών εξισώσεων). β) Να εξηγήσετε αν με το παραπάνω βήμα μπορεί να παρουσιαστεί αστάθεια αν συνεχίσετε την προσέγγιση για πολλά βήματα και αν ναι, να υπολογίσετε άνω φράγμα για το βήμα h ώστε να αποφευχθεί η αστάθεια.

Απάντηση. α) Όπως προτείνεται μετατρέπουμε το παραπάνω σε σύστημα με την εισαγωγή βοηθητικών μεταβλητών (δείτε βιβλίο και διαφάνειες): $u_1(t) := u(t)$, $u'(t) := u_2(t)$, και $u''(t) := u_3(t)$ οπότε η διαφορική μετατρέπεται σε σύστημα 3 σύνηθων διαφορικών, ως εξής

$$\frac{d}{dt}\begin{pmatrix} u_1(t) \\ u_2(t) \\ u_3(t) \end{pmatrix} = -\begin{pmatrix} 0 & -1 & 0 \\ 0 & 0 & -1 \\ 1000 & 300 & 30 \end{pmatrix}\begin{pmatrix} u_1(t) \\ u_2(t) \\ u_3(t) \end{pmatrix}$$

ή για συντομία

$$\frac{d}{dt}\mathbf{u} = -A\mathbf{u}$$

όπου $\mathbf{u} := [u_1, u_2, u_3]^\top$ (παραλείπουμε το t το οποίο εννοείται). Εφαρμόζοντας εμπρός Euler με το βήμα $h = 0.8$ και $U(0) = [1, 0, 1]^\top$, για να υπολογίσουμε την τιμή στο $t = 2h$ έχουμε έχουμε ότι

$$U(2h) = (I - hA)((I - hA)U(0)) = [\mathbf{1.64}, -657.6, 17937]^\top.$$

Με παχειά γραφή έχουμε συμβολίσει το ζητούμενο, δηλ. την προσέγγιση στο $u(2h)$ με εμπρός Euler.

β) Προσέξτε ότι από τη διακύμανση των στοιχείων φαίνεται ότι μάλλον υπάρχει αστάθεια! Για να το επιβεβαιώσουμε, εξετάζουμε τη μέγιστη ιδιοτιμή του $I - hA$ για το βήμα h που χρησιμοποιήσαμε. Οι ιδιοτιμές του A είναι οι ρίζες του πολυωνύμου $1000 + 300\lambda + 30\lambda^2 + \lambda^3 = 0$, οπότε $\lambda_1 = \lambda_2 = \lambda_3 = -10$. Επομένως με $h = 0.8$ η φασματική ακτίνα του $I - hA$ θα είναι $7 = |1 - 0.8 \times 10|$ και θα έχουμε αστάθεια. Εδικότερα, το βήμα h πρέπει να επιλέγεται μικρότερο από $2/\max|\lambda_j| = 0.5$. □

γ) Γενικά στην Euler για την επίλυση ενός γραμμικού προβλήματος του τύπου $u' = -Au$, είναι σωστό ή λάθος ότι αν μειωθεί το βήμα στο μισό, τότε το μέγιστο ολικό σφάλμα διακριτοποίησης θα υποτετραπλασιαστεί.

Απάντηση. ΛΑΘΟΣ, το ολικό σφάλμα συμπεριφέρεται όπως το $O(1/h)$ άρα περιμένουμε να υποδιπλασιαστεί. □

δ) Για καθένα από τα παρακάτω σχετικά με τις άμεσες μεθόδους Runge-Kutta τάξης 2 και πάνω για την επίλυση της ΣΔΕ $u'(t) = f(t, u)$, να κυκλώσετε αν είναι σωστό ή λάθος:

(Σ - Λ) Προβλέπουν τη νέα τιμή συνδυάζοντας την προσέγγιση στο t_k με προσεγγίσεις της παραγώγου της u σε μια ή περισσότερες τιμές του t στο διάστημα $[t_k, t_{k+1})$.

Απάντηση. ΣΩΣΤΟ, οι μέθοδοι RK είναι μονοβηματικές και χρησιμοποιούν ως πληροφορία την προσέγγιση στο t_k με εκτιμήσεις της παραγώγου στο t_k και άλλα σημεία στο παραπάνω διάστημα. Ο γενικός τύπος είναι

$$U_{n+1} \quad = \quad U_k + h\sum_{i=1}^{s} b_i K_i$$

όπου

$$K_i \quad = \quad f(t_n + c_i h, U_n + h \sum_{j=1}^{s} a_{ij} K_j)$$

□

(Σ - Λ) Έχουν πιο εκτεταμένο χωρίο ευστάθειας από την πίσω Euler.

Απάντηση. ΣΩΣΤΟ, η πίσω ευστάθεια καθορίζεται από χωρία της μορφής

$$\mathcal{D} \quad := \quad \{z := h\lambda || p_n(z)| \leq 1\}, \quad p_n(z) = \sum_{j=0}^{s} \frac{z^j}{j!}$$

όπου το πολυώνυμο προκύπτει από την εφαρμογή της μεθόδου στο $u' = \lambda u$. □