1η Υποχρεωτική Εργασία
Στο Μάθημα της Αριθμητικής Ανάλυσης

Οδηγίες: Η εργασία είναι ατομική. Θα πρέπει να γραφεί παραδοτέο, σε LaTeX (προκειμένου να προσμετρηθεί η προαιρετική εργασία) ή σε οποιονδήποτε άλλο επεξεργαστή κειμένου (δεν προσμετράται η προαιρετική), στο οποίο να περιγράφονται οι λύσεις των ασκήσεων, να δίνονται και να σχολιάζονται τα αποτελέσματα, να περιγράφεται η εκτέλεση του κώδικα. Θα πρέπει το αρχείο του παραδοτέου (σε .pdf) μαζί με τα αρχεία του κώδικα (σε οποιαδήποτε γλώσσα προγραμματισμού επιθυμείτε) να συμπιεστούν σε ένα αρχείο το οποίο θα υποβληθεί μέσω του elearning μέχρι και την ημερομηνία παράδοσης:

22 Δεκεμβρίου 2020, ώρα 24:00

Για κάθε ημέρα αργοπορημένης υποβολής της εργασίας και για 5 ημέρες μειώνεται η βαθμολογία κατά 10%. Μετά από την παράδοση όλων των εργασιών θα ακολουθήσει προφορική εξέταση ή παρουσίαση πάνω στις εργασίες, στην οποία θα περιλαμβάνεται και προφορική εξέταση του κώδικα για τις περιπτώσεις πιθανής αντιγραφής.

Για οποιαδήποτε απορία σχετικά με τις εργασίες μπορείτε να επικοινωνείτε μέσω email στην διεύθυνση : tefas@csd.auth.gr.

Άσκηση 1 (1 μονάδα)
Έστω η συνάρτηση $f(x) = e^{\sin^3 x} + x^6 - 2x^4 - x^3 - 1$ στο διάστημα $[-2, 2]$. Κάντε την γραφική παράσταση της συνάρτησης σε αυτό το διάστημα. Έπειτα υπολογίστε με πρόγραμμα και τις ρίζες της με ακρίβεια 5ου δεκαδικού ψηφίου χρησιμοποιώντας τις εξής μεθόδους: (α) τη μέθοδο της διχοτόμησης, (β) τη μέθοδο Newton-Raphson και (γ) τη μέθοδο της τέμνουσας. Για κάθε ρίζα να συγκρίνετε το πλήθος των επαναλήψεων που έγιναν. Για τη μέθοδο Newton-Raphson να δείξετε για ποιες ρίζες συγκλίνει τετραγωνικά και για ποιες όχι. Ποιο είναι το χαρακτηριστικό των ριζών για τις οποίες η μέθοδος Newton-Raphson δεν συγκλίνει τετραγωνικά (αιτιολογήστε)·

Άσκηση 2 (1 μονάδα)
Υλοποιήστε (α) μια τροποποιημένη μέθοδο Newton-Raphson όπου:

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)} - \frac{1}{2}\frac{f(x_n)^2 f''(x_n)}{f'(x_n)^3}$$

(β) μια τροποποιημένη μέθοδο διχοτόμησης όπου η εκτίμηση για την ρίζα δεν είναι το μέσο του διαστήματος αναζήτησης σε κάθε βήμα αλλά ένα τυχαίο σημείο που επιλέγεται με την χρήση μιας συνάρτησης παραγωγής τυχαίων αριθμών (π.χ. rand()) εντός του διαστήματος αναζήτησης.
(γ) Μια τροποποιημένη μέθοδο της τέμνουσας η οποία χρειάζεται 3 αρχικά σημεία x_n, x_{n+1}, x_{n+2} και υπολογίζει την επόμενη εκτίμηση για την ρίζα με βάση τον τύπο:

$$x_{n+3} = x_{n+2} - \frac{r(r-q)(x_{n+2} - x_{n+1}) + (1-r)s(x_{n+2} - x_n)}{(q-1)(r-1)(s-1)}$$

όπου $q = \frac{f(x_n)}{f(x_{n+1})}$, $r = \frac{f(x_{n+2})}{f(x_{n+1})}$ και $s = \frac{f(x_{n+2})}{f(x_n)}$. Το νέο σημείο x_{n+3} εισάγεται ως τρίτο σημείο και αντικαθιστά το παλαιότερο σημείο των προηγούμενων τριών, δηλαδή το x_n, και ο αλγόριθμος επαναλαμβάνεται μέχρι την σύγκλιση στο επιθυμητό λάθος. Η μέθοδος αυτή πρακτικά βρίσκει μια παραβολή που περνάει από τα τρία σημεία και ακολούθως την τομή της παραβολής με τον άξονα x.

1. Να βρείτε όλες τις ρίζες της συνάρτησης $f(x) = 94\cos^3 x - 24\cos x + 177\sin^2 x - 108\sin^4 x - 72\cos^3 x \sin^2 x - 65$ στο διάστημα $[0, 3]$ με ακρίβεια 5ου δεκαδικού ψηφίου χρησιμοποιώντας τις παραπάνω μεθόδους και επιλέγοντας κατάλληλες αρχικοποιήσεις.

2. Να εκτελέσετε τον αλγόριθμο (β) 10 φορές και να διαπιστώσετε αν συγκλίνει πάντα σε ίδιο αριθμό επαναλήψεων.

3. Να συγκρίνετε ως προς την ταχύτητα σύγκλισης πειραματικά τις τροποιημένες μεθόδους σε σχέση με τις κλασικές.

Άσκηση 3 (1 μονάδα)

Έστω ότι έχουμε προς επίλυση το παρακάτω γραμμικό σύστημα:

$$\mathbf{Ax} = \mathbf{b}$$

1. Προγραμματίστε σε οποιαδήποτε γλώσσα προγραμματισμού επιθυμείτε μια συνάρτηση η οποία θα παίρνει σαν είσοδο τον πίνακα $\mathbf{A}$ και το διάνυσμα $\mathbf{b}$ του παραπάνω γραμμικού συστήματος και θα επιστρέφει σαν έξοδο το διάνυσμα των αγνώστων $\mathbf{x}$ (e.g.,`[x] = gauss(A, b)`). Η λύση του συστήματος πρέπει να γίνει με την μέθοδο επίλυσης γραμμικών συστημάτων $\mathbf{PA} = \mathbf{LU}$.

2. Προγραμματίστε σε οποιαδήποτε γλώσσα προγραμματισμού επιθυμείτε μια συνάρτηση η οποία θα παίρνει σαν είσοδο έναν συμμετρικό και θετικά ορισμένο πίνακα $\mathbf{A}$ και θα επιστρέφει έναν κάτω τριγωνικό πίνακα $\mathbf{L}$ που αποτελεί την αποσύνθεση Cholesky του πίνακα $\mathbf{A}$ (e.g., `[L] = cholesky(A)`).

3. Προγραμματίστε σε οποιαδήποτε γλώσσα προγραμματισμού την μέθοδο Gauss-Seidel και χρησιμοποιείστε την για να επιλύσετε με ακρίβεια 4 δεκαδικών ψηφίων (ως προς την άπειρη νόρμα στο σφάλμα της λύσης) το $n \times n$ αραιό σύστημα $\mathbf{Ax} = \mathbf{b}$ για $n = 10$ και $n = 10000$, με $A(i,i) = 5$, $A(i+1,i) = A(i,i+1) = -2$ και $\mathbf{b} = [3, 1, 1, \ldots, 1, 1, 3]^T$.

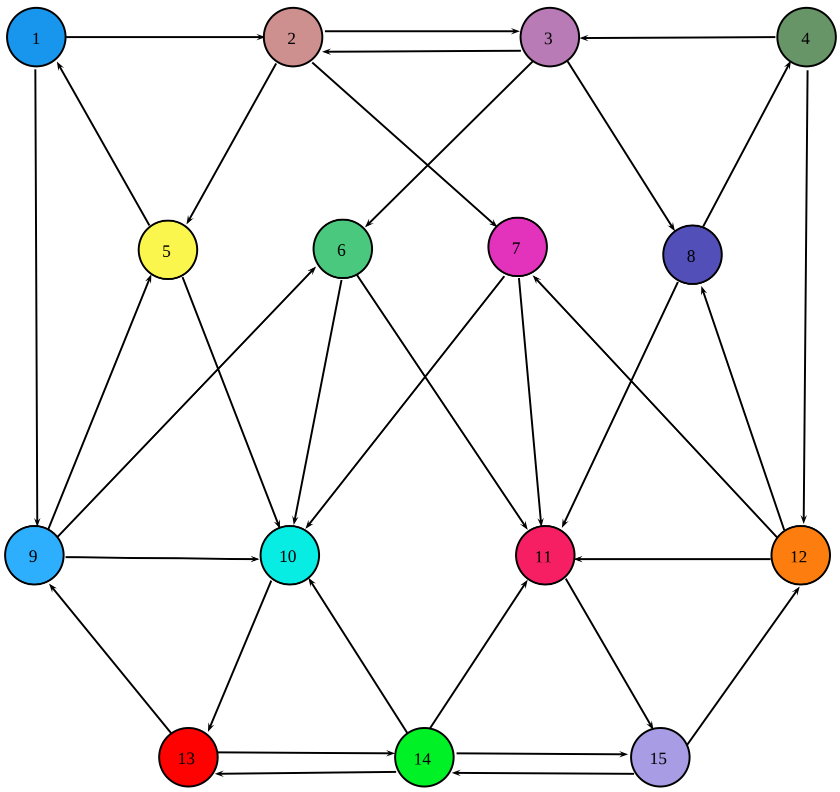
$$\begin{bmatrix} 5 & -2 & & & \\ -2 & 5 & -2 & & \\ & \ddots & \ddots & \ddots & \\ & & -2 & 5 & -2 \\ & & & -2 & 5 \end{bmatrix} \begin{bmatrix} x_1 \\ \\ \vdots \\ \\ x_n \end{bmatrix} = \begin{bmatrix} 3 \\ 1 \\ \vdots \\ 1 \\ 3 \end{bmatrix}$$

Υπόδειξη: Δεν επιτρέπεται να γίνει χρήση built in συναρτήσεων (e.g., chol(), lu(), left division operator, etc.).

Άσκηση 4 (3 μονάδες)

Οι μηχανές αναζήτησης, όπως το `http://www.google.com/` χαρακτηρίζονται από τα εξαιρετικά αποτελέσματα που επιστρέφουν σε ερωτήματα. Θα δώσουμε μία χονδρική εξήγηση του πως γίνεται αυτή η ταξινόμηση χρησιμοποιώντας την πληροφορία σύνδεσης μεταξύ ιστοσελίδων στο WWW.Το google.com δίνει έναν θετικό πραγματικό αριθμό, την τάξη σελίδας (page rank), σε κάθε σελίδα που επιστρέφει. Η τάξη σελίδας υπολογίζεται από τη google.com σε έναν από τους μεγαλύτερους όπως λέγεται υπολογισμούς ιδιοδιανυσμάτων με τη μέθοδο των δυνάμεων στον κόσμο. Στο γράφημα του Σχήματος 1 κάθε κόμβος n αναπαριστά μία ιστοσελίδα και μία ακμή από το i στο j αναπαριστά ένα σύνδεσμο από την ιστοσελίδα του i στην ιστοσελίδα του j.

Έστω $\mathbf{A} \in \mathbb{R}^{n \times n}$, $\quad n = 15$, ο πίνακας γειτνίασης του γραφήματος που απεικονίζεται στο Σχήμα 1.



Σχήμα 1

$$\mathbf{A} = \begin{bmatrix}
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0
\end{bmatrix}$$

Για τον $\mathbf{A}$ ισχύει $\mathbf{A}_{(i,j)} = 1$ όταν υπάρχει σύνδεσμος από τον κόμβο i στον κόμβο j και 0 διαφορετικά.

Οι κατασκευαστές του google.com φαντάστηκαν έναν χρήστη σε αυτό το δίκτυο που είναι στην σελίδα i με πιθανότητα p_i. Έπειτα, ο χρήστης είτε μετακινείται σε μία τυχαία σελίδα (με συγκεκριμένη πιθανότητα q που είναι ίση συνήθως με 0.15) ή με πιθανότητα $1-q$ επιλέγει τυχαία έναν από τους συνδέσμους που βρίσκονται στη σελίδα i. Η πιθανότητα ότι ο χρήστης μετακινείται από την σελίδα i στη σελίδα j δίνεται από τη σχέση:

$$p_{\{i \to j\}} = \frac{q}{n} + (1-q) * \frac{\mathbf{A}_{(i,j)}}{n_i}$$

όπου n_i είναι το άθροισμα της i-οστής γραμμής του $\mathbf{A}$ (στην ουσία το πλήθος των συνδέσεων της σελίδα i). Η πιθανότητα ο χρήστης σε οποιαδήποτε χρονική στιγμή να είναι στην σελίδα j είναι το άθροισμα της $p_{\{i \to j\}}$ στην i-οστή γραμμή του πίνακα $\mathbf{A}$.

$$p_j = \sum_{i=1}^{n} \left(\frac{qp_i}{n} + (1-q) \frac{p_i}{n_i} \mathbf{A}_{(i,j)} \right) \tag{1}$$

Η Σχέση (1) μπορεί να γραφεί και με μορφή πινάκων ως εξής:

$$\mathbf{p} = \mathbf{G}\mathbf{p} \tag{2}$$

όπου $\mathbf{p}$ είναι το διάνυσμα με τις n πιθανότητες του να βρίσκεται ο χρήστης στις n ιστοσελίδες και $\mathbf{G}$ είναι ο πίνακας του οποίου το κελί $\mathbf{G}_{(i,j)} = \frac{q}{n} + \frac{\mathbf{A}_{(j,i)}(1-q)}{n_j}$. Θα ονομάσουμε τον πίνακα $\mathbf{G}$, πίνακα Google. Κάθε στήλη του πίνακα Google δίνει άθροισμα 1 και επομένως είναι στοχαστικός πίνακας που σημαίνει ότι η μέγιστη ιδιοτιμή του είναι 1. Το ιδιοδιάνυσμα που αντιστοιχεί στην ιδιοτιμή 1 είναι το σύνολο των μη μεταβλητών πιθανοτήτων, που είναι επί της ουσίας η τάξη της σελίδας.

Με βάση το παράδειγμα που δώσαμε προηγουμένως, οι τάξεις κάθε σελίδας για $q = 0.15$ δίνονται παρακάτω:

$$\mathbf{p}= [0.0268238,0.0298611,0.0298612,0.0268241,0.0395877,0.0395878, \ldots$$
$$\ldots, 0.039588,0.0395881,0.074565,0.106319,0.10632,0.0745657,0.125089,\ldots, \ldots,0.11633,0.12509]^\top$$

Το συγκεκριμένο ιδιοδιάνυσμα έχει κανονικοποιηθεί ώστε το άθροισμα να είναι ίσο με 1 (όπως θα πρέπει να ισχύει για πιθανότητες). Η τάξη σελίδας είναι υψηλότερη για τις σελίδες 13 και 15 ακολουθούμενες από τις σελίδες 14, 10 και 11. Προσέξτε ότι η τάξη σελίδας δεν εξαρτάται μόνο από το βαθμό εισόδου (πλήθος σελίδων που δείχνουν σε μία σελίδα) αλλά προκύπτει με πιο πολύπλοκο τρόπο. Παρόλο που οι κόμβοι 10 και 11 έχουν το μεγαλύτερο βαθμό εισόδου το γεγονός ότι δείχνουν στους κόμβους 13 και 15 μεταφέρει τη σημαντικότητά τους σε αυτούς. Αυτή είναι η ιδέα πίσω από το φαινόμενο του google-bombing, δηλαδή της τακτικής του να αυξάνεται η σημαντικότητα μίας ιστοσελίδας πείθοντας άλλες ιστοσελίδες υψηλής κίνησης να δείχνουν σε αυτή. Προσέξτε την σύνδεση μεταξύ της τάξης σελίδας και της λέξης σημαντικότητας. Η τάξη σελίδας είναι αυτή τη στιγμή ο καλύτερος τρόπος να αναπαραστήσουμε τη σημαντικότητα μία σελίδας αν και κανείς ακόμα δεν ξέρει τι είναι σημαντικότητα. Επομένως, κάποιος μπορεί να βρει μία καλύτερη μέθοδο για να μοντελοποιεί τη σημαντικότητα.

Ζητούμενα:

1. Αποδείξτε αναλυτικά ότι ο πίνακας $\mathbf{G}$ είναι στοχαστικός.

2. Κατασκευάστε τον πίνακα $\mathbf{G}$ για το συγκεκριμένο παράδειγμα και ελέγξτε ότι το ιδιοδιάνυσμα της μέγιστης ιδιοτιμής είναι αυτό που δίνεται παραπάνω (με τη μέθοδο της δυνάμεως).

3. Προσθέστε 4 συνδέσεις που εσείς θα επιλέξετε και αφαιρέστε 1 από αυτές που ήδη υπάρχουν στο γράφημα και κατασκευάστε τον νέο πίνακα $\mathbf{G}$ έτσι ώστε να βελτιώσετε τον βαθμό σημαντικότητας μιας σελίδας που θα επιλέξετε.

4. Αλλάξτε την πιθανότητα μεταπήδησης q σε (α) $q = 0.02$ και (β) $q = 0.6$ στον νέο γράφο. Περιγράψτε τις αλλαγές στην τάξη σελίδας. Ποιος είναι ο σκοπός της πιθανότητας μεταπήδησης;

5. Έστω ότι στο αρχικό γράφημα του Σχήματος 1 η ιστοσελίδα 11 ήθελε να βελτιώσει την τάξη της σε σχέση με τον ανταγωνιστή της, την σελίδα 10. Έστω ότι αυτό θέλει να το πετύχει πείθοντας τις σελίδες 8 και 12 να δείχνουν τη σύνδεση στην 11 με καλύτερο τρόπο. Έστω ότι μοντελοποιούμε αυτή την κατάσταση αντικαθιστώντας το $\mathbf{A}_{(8,11)}$ και το $\mathbf{A}_{(12,11)}$ με 3 στον πίνακα γειτνίασης. Αυτή η στρατηγική δουλεύει;

6. Μελετήστε την επίδραση της διαγραφής της σελίδας 10 από το γράφημα. Ποιες τάξεις σελίδων αυξάνονται και ποιες μειώνονται;