

# Επιστημονικός Υπολογισμός Εργαστηριακή Άσκηση 2019-20

Ημερομηνία κατάθεσης για πλήρη βαθμό ως τις 23:59 στις 19/1/2020<sup>1</sup>

Στη συγγραφή της αναφοράς, παρακαλείστε να θυμηθείτε τις "καλές πρακτικές" που έχετε μάθει στο "Συγγραφή και Παρουσίαση Τεχνικών Κειμένων". Εννοείται ότι θα είναι πολύ καλύτερα αν υποβάλετε την αναφορά σε LaTeX.

## 1 Εισαγωγικά [156]

Να περιγράψετε τα παρακάτω στοιχεία για την άσκηση: Πότε ξεκινήσατε να ασχολείστε με αυτήν, πότε τελειώσατε, και τα χαρακτηριστικά του υπολογιστικού συστήματος το οποίο χρησιμοποιήσατε για την υλοποίηση της παρούσας εργαστηριακής άσκησης και που βρίσκεται αυτό (π.χ. Υ/Κ, λαπτοπ, desktop). Πρέπει να αναφέρετε που ή πως βρήκατε τα στοιχεία.

- (i) Να συμπληρώσετε τα στοιχεία για το σύστημα στο οποίο θα τρέξετε τα πειράματά σας όπως αναφέρονται στον Πίνακα 1. Είναι απαραίτητο για τα στοιχεία που θα δώσετε, να αναφέρετε που ή πως τα βρήκατε. Για χρήστες Windows μπορείτε να κατεβάσετε ειδικά προγράμματα όπως το `cruz` από τη διεύθυνση <http://www.crudd.com/> το οποίο θα σας δώσει τις πληροφορίες που είναι ζητούμενες. Για χρήστες Linux μπορείτε να βρείτε τις ζητούμενες πληροφορίες μέσω των εντολών `cat /proc/meminfo` και `cat /proc/cpuinfo`.
- (ii) Έκδοση του MATLAB που χρησιμοποιήσατε.
- (iii) Επιπλέον πρέπει να δείξετε τον πίνακα που προκύπτει όταν εκτελείτε την εντολή `bench`. Για παράδειγμα, δείτε τον πίνακα 1 στο Παράρτημα.

Πίνακας 1: Στοιχεία για τα πειράματα

Χαρακτηριστικό	ενδεικτική απάντηση
Έναρξη/λήξη εργασίας	15/12/19 - 14/1/20
model	MacBook Pro <sup>2</sup>
O/S	macOS Catalina 10.15.1
processor name	8-Core Intel Core i9 (i9-9880H) <sup>3</sup>
processor speed	2.3 GHz (base)
number of processors	1
total # cores	8
total # theads	16
FMA instruction	yes
L1 cache	256KB Instruction, 256 KB Data write-back <sup>4</sup>
L2 cache	(per core) 256KB, write-back
L3 cache	(shared) 16MB, write-back
Gflops/s	423.2 <sup>5</sup>
Memory	16GB
Memory Bandwidth	41.8 GB/s
MATLAB version	8.4.0.150421 (R2014b) <sup>6</sup>

<sup>1</sup> ΔΕΝ ΘΑ ΔΟΘΕΙ ΠΑΡΑΤΑΣΗ.

**Παραδοτέο**

Πίνακας όπως ο 1 καθώς και οι σχετικές αναφορές.

**1.1 Κατασκευή εργαλείων****1.2 Μητρώο μάσκα: Συνάρτηση mask\_band**

Θα πρέπει φτιάξετε συνάρτηση `mask_band`, για την παραγωγή μητρώων "μάσκα" δηλ. μητρώων με Boolean τιμές 0/1 και που θα μπορεί να χρησιμοποιηθεί σε επόμενες ερωτήσεις για την εύκολη δημιουργία μητρώων ζώνης και μπλοκ τριδιαγώνιων μητρώων. Η συνάρτηση θα καλείται ως

$$P = \text{mask\_band}(n, \text{type}, p) \text{ ή } P = \text{mask\_band}(n, \text{type}, p, q)$$

και θα λειτουργεί ως εξής:

Αν `type == 'band'` τότε επιστρέφει μητρώο ζώνης, διάστασης  $n \times n$ , με άνω ημιεύρος  $q$  και κάτω ημιεύρος  $p$ . Επιπλέον, αν κληθεί ως `mask_band(n, 'band', p)`, τότε θεωρείται ότι το αμφότερα τα άνω και κάτω ημιεύρη είναι  $p$ . Οι μη μηδενικές τιμές του  $P$  είναι όλες 1.

Αν `type == 'btdr'`, η συνάρτηση καλείται με μόνον μια επιπλέον τιμή εισόδου, δηλ. `mask_band(n, 'btdr', ... p)` όπου  $p \times p$  είναι το μέγεθος των διαγώνιων μπλοκ. Σε περίπτωση που το  $p$  δεν διαιρεί ακριβώς το  $n$ , δηλ.  $n = kp + r$ , τότε το  $P$  θα έχει  $k$  διαγώνια μπλοκ και ένα επιπλέον μπλοκ  $r \times r$ . Προσέξτε ότι αυτό εξυπηρετεί και την περίπτωση  $n < p$ . Όλα τα στοιχεία του  $P$  εντός των διαγώνιων μπλοκ είναι 1 και τα υπόλοιπα 0.

**1.3 Έλεγχος Διαγώνιας Κυριαρχίας: Συνάρτηση dd\_check**

Θα πρέπει να κατασκευάσετε συνάρτηση `dd_check` που θα ελέγχει κατά πόσον ένα μητρώο είναι Διαγώνια Κυρίαρχο (ΔΚ). Ειδικότερα, η συνάρτηση θα καλείται ως `[dflag, discrC, discrR] = dd_check(A)`, και στο `dflag` θα επιστρέφεται 1 αν το μητρώο είναι ΔΚ ως προς γραμμές και στήλες, διαφορετικά 0. Επίσης, η τιμή `discrR` θα δείχνει το βαθμό της διαγωνιας κυριαρχίας ως προς τις γραμμές, και το `discrC` ως προς τις στήλες. Ο μαθηματικός ορισμός της `discrR` είναι

$$\text{discrR} = \max_i \{ \min \{ |\alpha_{ii}| - \sum_{j \neq i} |\alpha_{ij}|, 0 \} \}$$

και αντίστοιχα για το `discrC`. Αν το μητρώο είναι μεν ΔΚ αλλά όχι ΑΔΚ, τότε το `dflag` επιστρέφει 1 αλλά ένα ή και τα δύο `discrR`, `discrC` θα είναι 0.

**1.4 Από MATLAB σε LaTeX: Συνάρτηση matrix2latex2**

Θα γράψετε συνάρτηση MATLAB `matrix2latex2` για την κατασκευή πινάκων και μητρώων σε μορφή κατάλληλη για την επεξεργασία σε κείμενο LaTeX (κάτι που μπορείτε να χρησιμοποιήσετε παρακάτω στην εργασία ή και μελλοντικά). Η ιδέα είναι, δοθέντος ενός μητρώου να προκύπτει η επιθυμητή μορφή στο LaTeX. Προτείνεται να ξεκινήσετε με τον κώδικα `matrix2latex.m` του M. Koehler (διαθέσιμος από το Matlab File Exchange) και να τον τροποποιήσετε έτσι ώστε να έχει την δυνατότητα, πέραν της μορφής `tabular`, να επιστρέφει τις δομές `pmatrix`, `bmatrix`, `vmatrix`

<sup>3</sup><https://ark.intel.com/content/www/us/en/ark/products/192987/intel-core-i9-9880h-processor-16m-cache-up-to-4-80-ghz.html>

<sup>4</sup>[https://en.wikichip.org/wiki/intel/core\\_i9/i9-9880h](https://en.wikichip.org/wiki/intel/core_i9/i9-9880h)

<sup>5</sup><https://gadgetversus.com/processor/intel-core-i9-9880h-vs-intel-core-i7-9750h/>

<sup>6</sup>Εντολή `ver` στη MATLAB

$$A = \begin{pmatrix} A_{11} & A_{12} & & & \\ A_{21} & A_{22} & A_{23} & & \\ & & & & \\ & & & A_{n-1,n-2} & A_{n-1,n-1} & A_{n-1,n} \\ & & & & & A_{n,n-1} & A_{n,n} \end{pmatrix}$$

Μπορείτε να αποφασίσετε εσείς για το πως θα διαμορφώσετε τον κώδικα και την είσοδο, αλλά θα πρέπει να εξηγήσετε τις επιλογές σας. Επίσης θα πρέπει να περιγράψετε τις τροποποιήσεις που έχετε κάνει.

Ο έλεγχος του κώδικά σας θα γίνει ως εξής: Με βάση κάποιο αρχικοποιημένο array, έστω A, θα πρέπει να μπορούμε να δώσουμε για είσοδο

```
matrix2latex2(A, Aout, 'Mstyle', 'pmatrix')
```

και στο αρχείο Aout.txt να ο κώδικας LaTeX για την παράγωγή του μητρώου περικλειόμενου σε παρενθέσεις. Για παράδειγμα, αν  $A = \text{diag}([1:3])$ , τότε η κλήση `matrix2latex2(A, Aout, 'Mstyle', 'pmatrix')` να επιστρέφει

```
\begin{pmatrix}
1.00&0.00&0.00&0.00\\
0.00&2.00&0.00&0.00\\
0.00&0.00&3.00&0.00\\
0.00&0.00&0.00&4.00\\
\end{pmatrix}
```

**Προσοχή:** Αν βρείτε και αποφασίσετε να χρησιμοποιήσετε κώδικα που βρήκατε στο διαδίκτυο, θα πρέπει να γράψετε την ακριβή αναφορά (που τον βρήκατε και ποιός είναι ο συγγραφέας) και να μπορείτε να εξηγήσετε όλες τις εντολές και επιλογές που περιλαμβάνει. Αντ' αυτού, καλύτερα να "χτίσετε" επί του αρχικού `matrix2latex` του Koehler.

### Παραδοτέο

i) συνάρτηση `mask_band.m`, ii) συνάρτηση `dd_check.m`, iii) συνάρτηση `matrix2latex2.m`.

## 2 Τανυστές και μπλοκ τριδιαγώνια μητρώα<sup>7</sup> [206]

1. Προτείνετε έναν τρόπο αποθήκευσης ενός μπλοκ τριδιαγώνου μητρώου  $A \in \mathbb{R}^{mn \times mn}$  όπου  $A_{ij} \in \mathbb{R}^{m \times m}$  (όπως παρακάτω) ως 4-mode  $m \times m \times n \times n$  tensor με τρόπο που ο συνολικός χώρος αποθήκευσης του tensor να είναι συγκρίσιμος (δηλ. όχι πολύ μεγαλύτερος) του μητρώου.
2. Αποκλειστικά για μητρώα αυτού του τύπου: Να γράψετε συνάρτηση `b2t` που δοθέντος του μητρώου να το μετατρέπει σε tensor.
3. Να δείξετε την εφαρμογή του `b2t` στο μητρώο

$$P_{(4,3)} = I_3 \otimes T_4 + S_3 \otimes I_4$$

<sup>7</sup>Για το μέρος αυτό, πρέπει να κατεβάσετε το Tensor Toolbox που αναφέραμε στην αντίστοιχη διάλεξη.

όπου  $T_k = \text{toeplitz}([30 \ -16 \ 1 \ \text{zeros}(1,k-3)])$ ,  $S_k = \text{toeplitz}([24 \ -12 \ \text{zeros}(1,k-2)])$ , και  $I_k$  είναι το  $k \times k$  ταυτοτικό μητρώο.

4. Σύμφωνα με την αναπαράστασή σας, ποια είναι α) η πρώτη mode-1 ίνα? β) η mode-4 ίνα?
5. Να προτείνετε έναν 3-mode tensor που δεν είναι sparse αλλά αξιοποιεί την μπλοκ τριδιαγώνια δομή και αποθηκεύει το μητρώο σε  $\approx nm^2$  θέσεις. Να δείξετε πώς θα αναφερθείτε στο στοιχείο στη θέση  $(4, 3)$  του μπλοκ  $A_{4,3}$ .

### Παραδοτέο

i) Σύνομη περιγραφή της πρότασής σας για το 1, ii) συνάρτηση b2t.m iii) Τιμές των 4, iv) Σύνομη περιγραφή της πρότασής σας για το 5.

## 3 Πράξεις με μητρώα ειδικής μορφής, πολυώνυμα μητρώων και χρήση διεπαφής mex

Στη συνέχεια θα γράψετε συναρτήσεις και θα πειραματιστείτε με υπολογισμούς με απλές συναρτήσεις μητρώων. Υπενθυμίζουμε ότι ο πολλαπλασιασμός μητρώου με διάνυσμα και η επίλυση ενός γραμμικού συστήματος μπορεί να θεωρηθεί ως η εφαρμογή σε διάνυσμα της συνάρτησης  $f(z) = z$  ή της  $f(z) = z^{-1}$  αντικαθιστώντας τη μεταβλητή  $z$  με το μητρώο.

### 3.1 Μητρώα και δεξιά μέλη [56]

Ως δεξιά μέλος θα χρησιμοποιήσετε ό,τι αναφέρεται στο εκάστοτε ερώτημα. Συνήθως θα πρόκειται για τα διανύσματα  $e$  (όλο μονάδες) ή το διάνυσμα  $Me$ , όπου  $M$  είναι μητρώο ή συνάρτηση μητρώου (δηλ. το διάνυσμα για το οποίο αν χρησιμοποιούσαμε αριθμητική άπειρης ακρίβειας, η λύση θα ήταν το  $e$ .) Θα χρησιμοποιήσετε 6 μητρώα, 1 πυκνό και 5 αραιά. Προσοχή, όταν πρόκειται για αραιά μητρώα, στη MATLAB θα πρέπει να οριστούν ως sparse ώστε να αξιοποιηθεί στο έπακρο η αραιή δομή.

**Μητρώο**  $C_N \in \mathbb{R}^{N \times N}$  με διαγώνια στοιχεία  $(A)_{kk} = 1 + k^\theta$  και  $(A)_{ij} = \frac{1}{|i-j|^2}$  εκτός διαγωνίου, όπου  $\theta = 1/2$ . Το  $N = 1000$ .

**Μητρώα**  $P_{(m,n)} = I_n \otimes T_m + S_n \otimes I_m$ , όπου, όπως στο ερώτημα 2,  $T_k = \text{toeplitz}([30 \ -16 \ 1 \ \text{zeros}(1,k-3)])$ ,  $S_k = \text{toeplitz}([24 \ -12 \ \text{zeros}(1,k-2)])$ , και  $I_k$  είναι το  $k \times k$  ταυτοτικό μητρώο. Τα μητρώα έχουν μέγεθος  $N \times N$  όπου  $N = mn$  και  $(m, n) = \{(100, 10), (10, 100), (100, 100)\}$ .

**Μητρώο bcsstm07** μεγέθους  $420 \times 420$  της ομάδας Harwell-Boing στη συλλογή SuiteSparse. Θα πρέπει να κατεβάσετε το μητρώο και να το φορτώσετε στη MATLAB (γίνεται απευθείας με εντολή από το SuiteSparse.)

**Μητρώο email** Μητρώο γένιασης από ένα email interchange network, Univ. of Rovira i Virgili, Tarragona. Το μητρώο έχει διάσταση  $1133 \times 1133$ .

Για καθένα από τα μητρώα που αναφέρονται στον κατάλογο του μέρους 3 και χρησιμοποιώντας όπου μπορείτε τη συναρτήσεις που αναπτύξατε στο εδάφιο 1.1, να ελέγξετε κατά πόσον είναι διαγώνια κυρίαρχο, ζώνης, αντιστρέψιμο και ποιός είναι ο δείκτης κατάστασης ως προς τη νόρμα 1 (με τη συνάρτηση `cond` ή την `condest`.)

**Παραδοτέο**

i) Πίνακας όπως ο 2 του Παραρτήματος, με μια γραμμή για κάθε μητρώο, όπως τα στοιχεία που παρατίθενται στη 2η γραμμή για το μητρώο `toeplitz`. ii) Σύντομη περιγραφή πως τον κατασκευάσατε.

**3.2 Επίλυση με απλά πολυώνυμα μητρώου [306]**

Στη συνέχεια θα διερευνήσετε τρόπους επίλυσης συστημάτων της μορφής  $p(A)x = b$  όπου  $p$  είναι πολυώνυμο,  $A \in \mathbb{R}^{N \times N}$ ,  $b \in \mathbb{R}^{N \times 1}$  είναι γνωστά. Το πολυώνυμο που ενδιαφέρει σε αυτό το ερώτημα είναι το

$$p(\zeta) = \zeta^4 + 14 * \zeta^3 + 71 * \zeta^2 + 154 * \zeta + 120$$

Το πρώτο βήμα είναι η κατασκευή ενός δεξιού μέλους για το οποίο η λύση να είναι γνωστή. Όπως συνηθίζεται, επιλέγουμε το  $b = p(A)e$ . Για να το υπολογίσετε, πρέπει να γράψετε συνάρτηση σαν την `polyvalm` η οποία όμως αντί να υπολογίζει πρώτα το  $p(A)$ , θα υπολογίζει το  $p(A)b$  αξιοποιώντας το γεγονός ότι η πράξη μπορεί να εκτελεστεί αποκλειστικά μέσω BLAS-2 (ουσιαστικά με πράξεις πολλαπλασιασμού `_GEMV`) αντί BLAS-3 (δηλ `_GEMM`). Προσοχή: Εδώ έχουμε μία περίπτωση που για λόγους οικονομίας πράξεων και αποθήκευσης, ενδείκνυται περισσότερο ο υπολογισμός με BLAS-2 αντί BLAS-3.

**Παραδοτέο**

Συνάρτηση `polyvalm_MV(p,A,b)` που δοθέντος τετραγωνικού μητρώου  $A$ , διανύσματος  $c$  (π.χ.  $e$ ) και των συντελεστών της δυναμομορφής (π.χ.  $p=[1,14,71,154,120]$ ), θα επιστρέφει το διάνυσμα  $p(A)c$ .

**3.2.1 Διαδικασίες επίλυσης**

**Explicit** Υπολογισμός του  $C = p(A)$  και στη συνέχεια επίλυση του  $Cx = b$  με την ανάποδη κάθετο της MATLAB.

**Serial** Υπολογισμός των ριζών  $\zeta_1, \zeta_2, \dots, \zeta_d$  του  $p(z)$ , όπου  $d$  είναι ο βαθμός του  $p$  και στη συνέχεια επίλυση των συστημάτων

$$(A - \zeta_j I)x^{(j)} = x^{(j-1)}, j = 1, \dots, d, \quad \text{όπου } x^{(0)} = b \text{ και } x = x^{(d)}.$$

με τις μεθόδους που αναφέρονται παρακάτω.

**Parallel (προ-2014)**

Να αξιοποιήσετε την διάσπαση του  $1/p(z)$  σε μερικά κλάσματα. Δηλαδή να αξιοποιήσετε την εναλλακτική γραφή

$$x = p(A)^{-1}b = \gamma_1(A - \zeta_1 I)^{-1}b + \dots + \gamma_d(A - \zeta_d I)^{-1}b$$

όπου οι συντελεστές  $\gamma_j$  είναι οι συντελεστές μερικών κλασμάτων. Χρησιμοποιήστε την συνάρτηση `residue` για να υπολογίσετε τους συντελεστές (επιστρέφει και τις ρίζες που είναι οι πόλοι του ρητού πολυωνύμου  $1/p$ ).<sup>a</sup>

<sup>a</sup> Η μέθοδος προσφέρεται για παράλληλη επεξεργασία καθώς επιλύονται παράλληλα  $d$  συστήματα.

**Επίλυση απλών συστημάτων**  $(A - \zeta I)z = y$

Για τους υπολογισμούς  $(A - \zeta I)^{-1}y$  θα πρέπει να χρησιμοποιήσετε τις παρακάτω μεθόδους που σχετίζονται με τη δομή του  $A$  (οι (ii) και (iii) αστοχούν αν το μητρώο δεν είναι ΣΘΟ).

- i) Την ανάποδη κάθετο της MATLAB.
- ii) Τη μέθοδο PCG<sup>a</sup> που η MATLAB παρέχει ως `pcg.m` α) χωρίς προρρυθμισμό, β) με προρρυθμισμό block Jacobi. Βλ. και [1, κεφ. 11.5.3].
- iii) Επίλυση μέσω παραγοντοποίησης Cholesky για μπλοκ τριδιαγώνιο μητρώο (μ.τ.), αν το μητρώο είναι μ.τ. και ΣΘΟ. Για την παραγοντοποίηση Cholesky θα πρέπει να αξιοποιήσετε όσο μπορείτε την μ.τ. δομή του μητρώου<sup>6</sup>. Να γράψετε μία εκδοχή της υλοποίησης της παραγοντοποίησης και επίλυσης απευθείας σε MATLAB και μία υλοποίηση χρησιμοποιώντας `mex` που θα καλεί BLAS και LAPACK<sup>γ</sup>

<sup>a</sup> Preconditioned Conjugate Gradient method

<sup>6</sup> Η μ.τ. δομή είναι ίσως η γενικότερη που υπάρχει καθώς οποιοδήποτε μητρώο μπορεί να θεωρηθεί, με τριτομμένο τρόπο, ότι είναι τέτοιο. Η πιο ενδιαφέρουσα δομή, για την κατασκευή ειδικών αλγορίθμων, είναι όταν τα μη μηδενικά μπλοκ έχουν διάσταση πολύ μικρότερη του συνολικού μεγέθους του μητρώου, δηλ. στην περίπτωση μας όταν  $m \ll mn$ . Εφαρμογές, από τη δομική μηχανική ως τη μηχανική μάθηση οδηγούν σε μ.τ. μητρώα (π.χ. [4]) και στην ανάγκη για ειδικούς αλγορίθμους, π.χ. Spike [1, Κεφ. 4.5.4].

<sup>γ</sup> Δείτε συναρτήσεις της Intel MKL εδώ.

### Παραδοτέο

- (i) Συναρτήσεις α) `chol_btr.m` και β) `chol_btr_mex.m` για την επίλυση συστήματος με μ.τ. ΣΘΟ μητρώο χρησιμοποιώντας Cholesky με MATLAB και με διεπαφή σε BLAS και LAPACK χρησιμοποιώντας `mex`. γ) Κώδικα τύπου script για την επίλυση με ανάποδη κάθετο και την `pcg` χωρίς ή με προρρυθμιστή.
- (ii) Χρονομετρήσεις και σφάλματα στην επίλυση του  $p(A)x = b$  όπου  $b = p(A)e$  χρησιμοποιώντας τις κατάλληλες μεθόδους στα μητρώα της παραπάνω συλλογής εκτός από το μητρώο `email`. Τα αποτελέσματα θα τα δείξετε σε πίνακες όπως οι 3 και 4 του Παραρτήματος. Για να θεωρηθεί επιτυχής μία μέθοδος για το πρόβλημα που εξετάζουμε, πρέπει να επιτυγχάνει σφάλμα μικρότερο του  $10^{-7}$  και αν πρόκειται για επαναληπτική μέθοδο, αυτό σε λιγότερες από 50 επαναληψεις.

## 4 Εφαρμογές σε δίκτυα [206]

Το ερώτημα αυτό αφορά σε μία σημαντική εφαρμογή στη σημερινή ανάλυση δικτύων που εμπλέκει την γρήγορη επίλυση γραμμικών συστημάτων με συναρτήσεις μητρώων. Εδώ θα περιοριστούμε στα απλά γραμμικά πολυώνυμα του τύπου  $p(B) = I - \alpha B$ .

**Σύντομη περιγραφή της θεωρίας** Ένα σημαντικό πρόβλημα στην ανάλυση γραφημάτων και δικτύων είναι η ταυτοποίηση των πιο σημαντικών κόμβων σύμφωνα με κάποιον ορισμό της σημαντικότητας. Μία κατηγορία μεθόδων ιεραρχεί τους κόμβους βάσει δεικτών κεντρικότητας (centrality). Ένας κλασικός δείκτης αυτής της κατηγορίας φέρει το όνομα Katz centrality ή Katz status index<sup>8</sup>. Ο δείκτης είναι παραμετροποιημένος, καθώς εξαρτάται από την τιμή μίας παραμέτρου που μπορούμε να επιλέξουμε με βάση κάποια κριτήρια και περιορισμούς. Συγκεκριμένα:

<sup>8</sup>Προτάθηκε από τον L. Katz το 1953 [3] για τη μέτρηση του status των ατόμων σε μια κοινωνία και χρησιμοποιείται μέχρι σήμερα ενώ αποτελεί αναπόσπαστο μέρος για όποιον ενδιαφέρεται να καταλάβει μετρικές όπως το PageRank. Αξίζει να σημειωθεί ότι ο δείκτης Katz γενικεύθηκε σχετικά πρόσφατα για την ανάλυση χρονοεξαρτώμενων δικτύων, δείτε π.χ. [2].

Ο δείκτης Katz του κόμβου  $i$  ενός δικτύου  $N$  κόμβων, με μητρώο γειτνίασης  $A \in \mathbb{R}^{N \times N}$  για την παράμετρο  $\alpha$ , είναι η  $i$ -οστή τιμή του  $(I - \alpha A^\top)^{-1}e$ , όπου  $e$  είναι το διάνυσμα μονάδων.

Θυμηθείτε από τη Γραμμική Άλγεβρα και από τη Θεωρία Γραφημάτων, ότι αν  $A$  είναι μητρώο γειτνίασης ενός γραφήματος, τότε η τιμή στη θέση  $(i, j)$  του  $A^k$  δείχνει το πλήθος των μονοπατιών που οδηγούν από τον κόμβο  $i$  στον  $j$  μήκους ακριβώς  $k$  (δηλ. μέσω  $k - 1$  κόμβων). Επομένως, αν το  $\alpha$  έχει επιλεγεί κατάλληλα, η σειρά  $\sum_{j=0}^{\infty} \alpha^j (A^\top)^j$  θα συγκλίνει στο  $(I - \alpha A^\top)^{-1}$  και η θέση  $(i, j)$  του αποτελέσματος περιέχει ένα σταθμισμένο άθροισμα των διαδρομών όλων των δυνατών μεγεθών (από 1 ως άπειρο). Επομένως, το  $(I - \alpha A^\top)^{-1}e$  δείχνει σε κάθε θέση  $i$ , το σταθμισμένο άθροισμα των μεγεθών όλων των δυνατών διαδρομών που εκκινούν από κάποιον κόμβο  $j$  και καταλήγουν στον  $i$ . Επομένως, όσο μικρότερο είναι το  $\alpha$ , τόσο μικρότερη συνεισφορά έχουν στην κεντρικότητα του κόμβου οι μακρύτερες διαδρομές από το  $j$  στον  $i$ . Προσοχή, αν το γράφημα είναι μη κατευθυνόμενο, τότε  $A = A^\top$  άρα δεν χρειάζεται να αναφερόμαστε στο ανάστροφο.

**Ζητούμενα** Ένα σημαντικό θέμα με τους δείκτες Katz αφορά στην επιλογή του  $\alpha$ . Στο μέρος αυτό της άσκησης θα ασχοληθείτε με αυτές τις επιλογές. Για το ερώτημα αυτό θα χρησιμοποιήσετε ως  $A$  αποκλειστικά το μητρώο `email`.

Εσείς πρέπει να υπολογίσετε με αποτελεσματικό τρόπο τους δείκτες Katz για τιμές της παραμέτρου  $\alpha$  από το διάνυσμα  $[0.01:0.01:1]$  για τις οποίες έχει νόημα ο υπολογισμός τους. Προφανώς, θα πρέπει το μητρώο να είναι αντιστρέψιμο και το διάνυσμα  $(I - \alpha A)^{-1}e \geq 0$  (αρνητική κεντρικότητα δεν έχει νόημα). Ειδικότερα: Να βρείτε και να δικαιολογήσετε ποιά είναι η μέγιστη τιμή του  $\alpha$  για την οποία έχει νόημα ο υπολογισμός. Για την επίλυση να χρησιμοποιήσετε τις μεθόδους `explicit` (δηλ. με χρήση ανάποδης καθέτου) και την `pcg` χωρίς `preconditioning` και με `incomplete Cholesky preconditioning`. Αφού βρήτε τη μέγιστη τιμή για την οποία έχει νόημα ο υπολογισμός, έστω  $\alpha_{\max}$ , να συμπληρώσετε τις τιμές που υπολογίσατε σε πίνακα όπως ο 5 του Παραρτήματος.

**Παραδοτέο:** Συνάρτηση `multiKatz`, που δοθέντος του μητρώου γειτνίασης και ενός διανύσματος  $\alpha$  μήκους  $s \geq 1$ , η κλήση

$$[X] = \text{multiKatz}(A, \alpha, \text{mth}, \text{pcg\_parms}\{1:\text{end}\})$$

επιστρέφει στις στήλες του  $X$  τις τιμές των δεικτών Katz για το αντίστοιχο  $\alpha = \alpha(i)$ .

Τα στοιχεία που καθορίζουν τη μέθοδο επίλυσης αποθηκεύονται στη μεταβλητή `mth` τύπου `string` και στη μεταβλητή `pcg_parms` που είναι τύπου `cell`. Αν `mth='direct'`, το σύστημα λύνεται με την ανάποδη κάθετο. Αν `mth='pcg'`, τότε το σύστημα θα λυθεί με PCG<sup>9</sup> και στο `cell pcg_parms` περνάτε όλες τις τιμές των μεταβλητών που θέλετε να ορίσετε για την `pcg` και ειδικότερα: Στο `pcg_parms{1}` το κατώφλι για το σφάλμα, στο `pcg_parms{2}` το μέγιστο αριθμό επαναλήψεων, και στα `pcg_parms{3}` και `pcg_parms{4}` τον ή τους `preconditioners`.

**Παραδοτέο:**

- i) Για κάθε μέθοδο υπολογισμού, να ετοιμάσετε πίνακα χρόνων, επαναλήψεων και καταλοίπων (βλ. Πίνακα 5 στο Παράρτημα).
- ii) Να δείξετε τα αποτελέσματα σε γραφικές παραστάσεις (βλ. παράδειγμα στο Σχήμα 2 Παραρτήματος) των καταλοίπων  $\|b - Ax^{(k)}\|_2$  ως προς το βήμα  $k$  για τις τιμές του  $\alpha$  για τις οποίες η κάθε μέθοδος συγκλίνει.
- iii) Για κάθε μέθοδο επίλυσης και τιμή του  $\alpha$ , να δείξετε ποιοί κόμβοι του γραφήματος που

<sup>9</sup> Preconditioned Conjugate Gradient method



αντιστοιχεί στο email υπολογίζονται ως οι 5 σημαντικότεροι. Δηλαδή να συμπληρώσετε την τελευταία στήλη του Πίνακα 5 στο Παράρτημα.

## 5 Οδηγίες

**Για τις χρονομετρήσεις** να χρησιμοποιήσετε τις συναρτήσεις `tic`, `toc` όπως αυτό υποδείχθηκε στις διαλέξεις, με αξιόπιστο τρόπο εξηγώντας καθαρά την επιλογή σας.

**Παρουσίαση γραφικών παραστάσεων:** Σε κάθε σχήμα, τα αποτελέσματα για κάθε είδος επιλυτή πρέπει να είναι color-blind, δηλ. η διαφορά τους να φαίνεται με τις επιλογές στίγματος και είδους γραμμής και όχι μόνον με το χρώμα. Επιπλέον θα πρέπει να δώσετε τίτλο στη γραφική παράσταση, στους άξονες καθώς και σε κάθε καμπύλη που παρουσιάζεται, όπως στο Σχήμα 2 του Παραρτήματος.

**Αξιολόγηση** Τα παραδοτέο σας θα κριθούν βάσει των εξής στοιχείων:

1. Της ορθότητας των αποτελεσμάτων (προφανώς οι κώδικες πρέπει να παράγουν τα αναμενόμενα αποτελέσματα).
2. Της ταχύτητας και αποδοτικότητας των αλγορίθμων σας. Θα πρέπει να αποφεύγονται περιττές πράξεις και να γίνεται αξιοποίηση της δομής (structure) μητρώων και διανυσμάτων στο έπακρο. Η βαθμολόγηση του μέρους αυτού της άσκησης εξαρτάται καίρια από τις επιλογές σας που να δείχνουν την κατανόηση του προβλήματος!
3. Της ποιότητας της αναφοράς (συνοπτική παρουσίαση και ευπαρουσίαστα αποτελέσματα).

Το πιο σημαντικό και απαραίτητο είναι η προσπάθεια να έχει γίνει από εσάς (δεν πειράζει να συζητήσετε και να συνεργαστείτε με συναδέλφους σας, όμως αυτό δεν πρέπει να είναι αντιληπτό αν πχ. σας ρωτήσουμε σε προφορική εξέταση σχετικά με τις απαντήσεις σας και την άσκηση).

### Προσοχή:

- Σημασία έχει η ατομική σας προσπάθεια στην επίλυση για την επιτυχημένη ολοκλήρωση του μαθήματος. Θεωρούμε ότι οι ασκήσεις που παραδίδονται είναι αποτέλεσμα προσωπικής προσπάθειας όποιου/ας την υπογράφει, που θα φέρει και την ευθύνη να απαντήσει αν της/του ζητηθεί να αιτιολογήσει και να υποστηρίξει όσα γράφονται. Ένα επιπλέον κέρδος είναι ότι η επίδοσή σας στη τελική εξέταση εξαρτάται από τις γνώσεις που αποκτήσατε στην προετοιμασία της εργασίας.
- Σε περίπτωση που διαπιστωθεί ότι η εργασία εκπονήθηκε από άλλον/ους, η ποινή που μπορεί να επιβληθεί είναι αντίστοιχη με τις ποινές για αντιγραφή με ηλεκτρονικά μέσα κατά τις εξετάσεις. Σημειώνεται επίσης ότι οι απαντήσεις στα ερωτήματα μπορούν να δοθούν με πολλούς διαφορετικούς τρόπους και είναι αρκετά απίθανο δύο διαφορετικές εργασίες να μοιάζουν πολύ! Γενικά, είναι πολύ προτιμότερο να παραδώσετε μία εργασία που είναι το αποτέλεσμα δικής σας προσωπικής προσπάθειας έστω και αν είναι ελλιπής.

## Σχετικά με τα παραδοτέα

Είναι απαραίτητο να ακολουθήσετε όσα αναφέρονται, ειδάλλως δεν θα βαθμολογηθείτε.

Θα τα αναρτήσετε στο e-Class σε ζιπαρισμένο αρχείο (zip) με όνομα



δηλ. έτος εισαγωγής, τα τελευταία 4 ψηφία του ΑΜ σας και το επίθετό σας με λατινικούς χαρακτήρες πρώτο γράμμα κεφαλαίο και τα υπόλοιπα πεζά. Για παράδειγμα αν υπήρχε (μάλλον αιώνιος) φοιτητής "Γαλλόπουλος" με ΑΜ που τελειώνει σε ΑΜ 8696 και έτος εισαγωγής 1996, θα πρέπει να χρησιμοποιήσει 1996\_8696\_Gallopoulos.pdf.

**ΠΡΟΣΟΧΗ: Οποιαδήποτε άλλη ονοματοδοσία θα είναι αιτία μηδενισμού της άσκησης.**

**Αναφορά** Η αναφορά σας πρέπει να είναι σε μορφή pdf με σύνθετο όνομα όπως και του zip αρχείου, μόνον με το σωστό επίθεμα, ΕΕΙΣΓ\_ΑΜ\_ΕΠΙΘΕΤΟ.pdf. Να είστε ιδιαίτερα προσεκτικοί ώστε η αναφορά να είναι αναγνώσιμη χωρίς πρόβλημα συμβατότητας γραμματοσειρών κ.λπ. Ιδιαίτερη σημασία θα δοθεί στον τρόπο και στην οργάνωση της παρουσίασης.

- Μέγεθος γραμματοσειράς 10pt. Σας συνιστούμε να χρησιμοποιήσετε το report style του LaTeX (π.χ. μέσω Overleaf.)
- Η πρώτη σελίδα πρέπει να περιέχει τα στοιχεία σας καθώς και πίνακα περιεχομένων.
- Οι σελίδες πρέπει να είναι αριθμημένες.
- Κάθε σελίδα της αναφοράς πρέπει να περιέχει το όνομα και το ΑΜ σας σε ευδιάκριτο σημείο (π.χ. όπως δεξιά και αριστερά στην άνω μέρος του παρόντος για ένα υποθετικό ΑΜ).

**Κώδικες:** Όλες τις συναρτήσεις, scripts και εντολές που χρησιμοποιήσατε από την εργασία σας. Συνοπτικά, οι αυτοτελείς συναρτήσεις που πρέπει να καταθέσετε είναι οι ακόλουθες: i) mask\_band.m (βλ. 1.1) ii) dd\_check.m (βλ. 1.1) iii) matrix2latex2.m (βλ. 1.1) iv) b2t.m (βλ. 2), v) polyvalm\_MV (βλ. 3), vi) συνάρτηση chol\_btr (βλ. 3), vii) συνάρτηση chol\_btr\_mex (βλ. 3), viii) multiKatz (βλ. 4)

**ΠΡΟΣΟΧΗ:** Για βαθμολόγηση, η 2η γραμμή κάθε συνάρτησης πρέπει να είναι σχόλιο με το ονοματεπώνυμό σας, ΑΜ και ημερομηνία συγγραφής, π.χ.

```
function [T]=b2t(A);
% Author: Ε. ΓΑΛΛΟΠΟΥΛΟΣ, ΑΜ 76848696, Date: 20/12/2019
```

Οι συναρτήσεις πρέπει να είναι εκτελέσιμες άμεσα στο φάκελλο που θα γίνει το unzip από τους βαθμολογητές, αν δοθούν ορθά στοιχεία εισόδου. Αν αυτό δεν συμβαίνει, π.χ. επειδή υπάρχουν εξαρτήσεις από αλλού, δεν θα βαθμολογηθείτε.

## Αναφορές

- [1] G. Golub and C.F. Van Loan. Θεωρία και Υπολογισμοί Μητρώων. Πεδίο, 2015. Μτφρ. 4ης Αμερικάνικης έκδ.
- [2] Peter Grindrod and Desmond J. Higham. A matrix iteration for dynamic network summaries. SIAM Rev., 55(1):118–128, January 2013.
- [3] L. Katz. A new status index derived from sociometric analysis. Psychometrika, 18(1):39–43, 1953.
- [4] Myungjun Kim, Yonghyun Nam, and Hyunjung Shin. An inference method from multi-layered structure of biomedical data. BMC Medical Informatics and Decision Making, 17(S1), May 2017.

**Α' Παράρτημα**

Computer Type	LU	FFT	ODE	Sparse	2-D	3-D
This machine	0.0693	0.0472	0.0695	0.0841	0.2684	0.5397
Macintosh (64-bit) 3.5 GHz Intel Core i7	0.0909	0.0593	0.0734	0.0889	0.3833	0.6312
Linux (64-bit) 2.66 GHz Intel Xeon	0.1295	0.0654	0.1933	0.1295	0.5581	0.5806
Windows 7 Enterprise (64-bit) 2.66 GHz Intel Xeon	0.1357	0.0705	0.1678	0.1535	0.4210	0.9613
Windows 7 Professional (64-bit) 3.07 GHz Intel Xeon W3550	0.2462	0.0844	0.1382	0.1476	0.3723	0.8357
Windows 7 Enterprise (64-bit) 2.7 GHz Intel Core i7	0.2926	0.0820	0.1075	0.1311	0.5387	1.0212
Macintosh (64-bit) 2.6 GHz Intel Core i7	0.1880	0.0904	0.0884	0.1297	0.9780	1.0788
Windows 8.1 (64-bit) 2.67 GHz Intel Xeon X5650	0.3146	0.0713	0.2053	0.2084	0.5210	0.8763
Windows XP (32-bit) 2.4 GHz Intel Core 2 Quad	0.6330	0.2110	0.1671	0.2818	0.5823	1.3302

Σχήμα 1: Αποτελέσματα της bench σε Pentium i9.

Πίνακας 2: Χαρακτηριστικά μητρώων ελέγχου

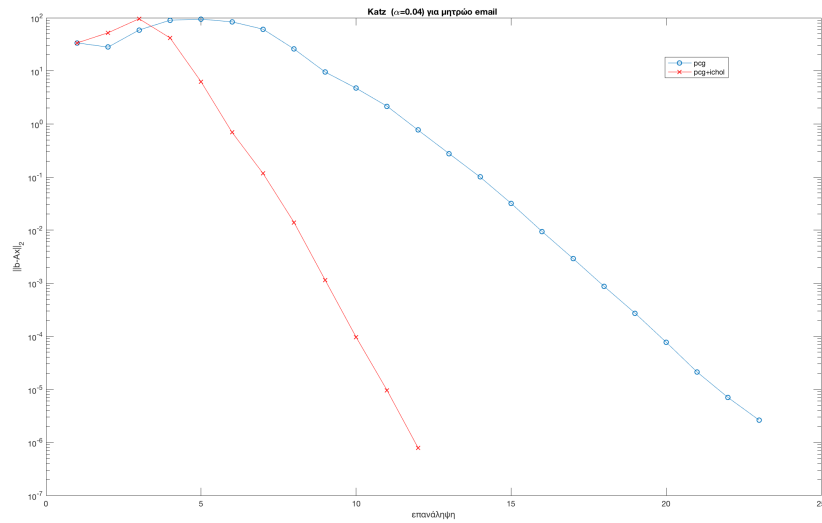
matrix	$N$	$\Delta K$	συμμετρικό	ζώνης	αντιστρέψιμο	δ.κ. $\kappa_1(A)$
toeplitz ([2,-1, zeros(1,8)])	10	ναι	ναι	(1,1)	ναι	60
$C_{1000}$						
$P_{(100,10)}$						
$P_{(10,100)}$						
$P_{(100,100)}$						
bcsstm07						
email						

Πίνακας 3: Χρόνοι εκτέλεσης σε sec (στη μορφή  $xx.xxxx\pm xx$  όπως στην επιλογή format short e

RUNTIMES (sec)	$C_{1000}$	$P_{(10,100)}$	$P_{(100,10)}$	$P_{(100,100)}$	bcsstm07
explicit					
serial+backslash					
serial+Cholesky	X				
serial+PCG	X				
serial+PCG+prec(blockJ)	X				
parallel+backslash (προ-2014)					

Πίνακας 4: Νόρμες-1 σφαλμάτων  $\|e - \tilde{x}\|_2$  (στη μορφή **xx.xxxx**e±**xx** όπως στο format short e.

ERRORS ( $\ \cdot\ _2$ )	$C_{1000}$	$P_{(10,100)}$	$P_{(100,10)}$	$P_{(100,100)}$	bcsstm07
explicit					
serial+backslash					
serial+Cholesky	X				
serial+PCG	X				
serial+PCG+prec(blockJ)	X				
parallel+backslash (προ-2014)					



Σχήμα 2: Τρόπος παρουσίασης γραφικών παραστάσεων

Πίνακας 5: Στη θέση των  $\alpha_{\max}$  θα θέσετε το μέγιστο  $\alpha$  που έχει νόημα για τη μέθοδο.

I-alpha(1)*email	$\alpha$	runtime (sec)	επαναλήψεις	$\ b - A\tilde{x}\ _2$	top-5 (hi to lo)
<b>format</b>	<b>x.xx</b>	<b>xx.xxxx</b>	<b>xxx</b>	<b>x.xxxx</b> e± <b>xx</b>	<b>π.χ. (1133, 1029, 2, 1,15)</b>
explicit	0.01		X		
	$\alpha_{\max}$		X		
serial+PCG	0.01				
	$\alpha_{\max}$				
serial+PCG+prec(ichol)	0.01				
	$\alpha_{\max}$				