

ΣΗΜΕΙΩΣΕΙΣ ΑΡΙΘΜΗΤΙΚΗΣ ΑΝΑΛΥΣΗΣ

Computer – Ανάλυση



Πίνακας Περιεχομένων

Κεφάλαιο 1 - Σύντομη επανάληψη Γραμμικής Άλγεβρας και ΕΥ.....	1
1.1 Υπολογισμοί και Σφάλματα	1
1.1.1 Σφάλμα Στρογγυλοποίησης.....	1
1.1.2 Καταστροφική ακύρωση (απαλοιφή) σημαντικών ψηφίων.....	1
1.1.3 Δεύτερο παράδειγμα καταστροφικής απαλοιφής	3
1.1.4 Άσκηση με ακρίβεια σημαντικών ψηφίων.....	3
1.1.5 Άσκηση με σφάλματα	4
1.2 Αναφορά στοιχειωδών εννοιών για μητρώα και διανύσματα.....	4
Κεφάλαιο 2 – Άμεσες και επαναληπτικές μέθοδοι	10
2.1 Επίλυση γραμμικού συστήματος.....	10
2.2 Σύγκριση άμεσων και επαναληπτικών μεθόδων	11
2.3. Άμεσες μέθοδοι – LU παραγοντοποίηση με μερική οδήγηση	12
2.3.1 Κόστος και μειονεκτήματα οδήγησης – πότε είναι απαραίτητη – περιγραφή μερικής οδήγησης	12
2.3.2 Οδήγηση στην LU παραγοντοποίηση	13
2.3.3 Πρώτο παράδειγμα μερικής οδήγησης	16
2.3.4 Δεύτερο παράδειγμα μερικής οδήγησης	19
2.3.5 Τρίτο παράδειγμα μερικής οδήγησης	19
2.3.6 Τέταρτο παράδειγμα μερικής οδήγησης	20
2.3.7 Πέμπτο παράδειγμα μερικής οδήγησης – Θέμα εξετάσεων	20
2.4. LDUΠαραγοντοποίηση	21
2.4.1 Άσκηση με μερική οδήγηση και LDU παραγοντοποίηση	22
2.5. Παραγοντοποίηση Cholesky.....	24
2.5.1 Πρώτο παράδειγμα εφαρμογής παραγοντοποίησης Cholesky	26
2.5.2 Δεύτερο παράδειγμα εφαρμογής παραγοντοποίησης Cholesky	27
2.5.3. Τρίτο παράδειγμα εφαρμογής παραγοντοποίησης Cholesky	28
2.5.4. Τέταρτο παράδειγμα εφαρμογής παραγοντοποίησης Cholesky	29
2.5.5. Διατήρηση ιδιοτήτων ΑΔΚ και ΣΘΟ μετά από ένα βήμα εφαρμογής απαλοιφής Gauss και Cholesky	30
2.5.6 Άσκηση με παραγοντοποίηση Cholesky (παλιό θέμα)	31
2.5.7 Άσκηση με LU παραγοντοποίηση με μερική οδήγηση (παλιό θέμα).....	32
2.5.8 Άσκηση με LU παραγοντοποίηση με μερική οδήγηση	33
2.5.9 Άσκηση με LU παραγοντοποίηση με μερική οδήγηση και παραγοντοποίηση Cholesky	34
2.5.10 Άσκηση με παραγοντοποίηση Cholesky	37

2.5.11 Ασκήσεις με παραγοντοποίηση Cholesky και θεώρημα Sylvester	39
2.5.12 Άσκηση με παραγοντοποίηση Cholesky και θεώρημα κύκλων Gerschgorin	41
2.5.13 Άσκηση με θεώρημα κύκλων Gerschgorin	41
2.5.14 Άσκηση με παραγοντοποίηση Cholesky και απαλοιφή Gauss (παλιό θέμα).....	42
2.5.15 Σύγκριση αντιστρεψιμότητας μητρώου και παραγοντοποίησης Cholesky	43
2.5.16 Σύγκριση παραγοντοποίησης LU και παραγοντοποίησης Cholesky αναφορικά με το χρόνο εκτέλεσης	44
2.6. Παραγοντοποίηση LDL^T και υπολογισμός αδράνειας συμμετρικού μητρώου	44
2.7. Δείκτης κατάστασης μητρώου – Απώλεια δεκαδικών ψηφίων λόγω δ.κ.	47
2.7.1 Δείκτες κατάστασης προβλήματος $\text{cond}(f; x)$ και αλγόριθμου $\text{cond}(\text{fprog})$	48
2.7.2 Άσκηση υπολογισμού δείκτη κατάστασης μητρώου	49
2.7.3 Άσκηση με δείκτη κατάστασης προβλήματος (παλιό θέμα)	50
2.7.4 Άσκηση με δείκτη κατάστασης μητρώου	50
2.7.5 Επαναληπτικές ασκήσεις με άμεσες μεθόδους – Άσκηση 1	51
2.7.6 Επαναληπτικές ασκήσεις με άμεσες μεθόδους – Άσκηση 2	51
2.8 Επαναληπτικές Μέθοδοι - Είδη μητρώων επανάληψης - κριτήρια σύγκλισης – κριτήρια τερματισμού.....	52
2.8.1 Κριτήρια σύγκλισης επαναληπτικών μεθόδων	53
2.8.2 Άσκηση με επαναληπτικές μεθόδους (Θέμα Φεβ. 2017)	56
2.8.3 Άσκηση με επαναληπτικές μεθόδους (παλιό θέμα)	56
2.8.4 Άσκηση με αναγωγησιμότητα μητρώου	57
2.8.5 Άσκηση με αναγωγησιμότητα μητρώου	57
2.8.6 Άσκηση με διαγώνια κυριαρχία	58
2.8.7 Κριτήρια τερματισμού επαναληπτικών μεθόδων	58
2.9 Μέθοδος Jacobi	59
2.9.1 Άσκηση εφαρμογής μεθόδου Jacobi	59
2.9.2 Άσκηση εφαρμογής μεθόδου Jacobi (Β' Τρόπος)	61
2.10 Μέθοδος Gauss - Seidel	62
2.11 Ανακεφαλαίωση – επανάληψη των επαναληπτικών μεθόδων	62
2.11.1 Άσκηση με Gauss - Seidel	62
2.11.2 Άσκηση με Jacobi και Gauss-Seidel	64
2.11.3 Άσκηση με Gauss-Seidel (θέμα από εξεταστική)	66
2.12 Επαναληπτική μέθοδος Richardson	66
2.12.1 Πρώτη άσκηση με Jacobi, Gauss – Seidel και Richardson	67
2.12.2 Δεύτερη άσκηση με Jacobi, Gauss – Seidel και Richardson	68
2.13 Ταχύτητα σύγκλισης επαναληπτικών μεθόδων	69

2.14 Άσκηση με επαναληπτικές μεθόδους.....	70
2.15 Άσκηση με επαναληπτικές μεθόδους.....	72
2.16 Άσκηση με επαναληπτικές μεθόδους.....	73
2.17 Άσκηση με επαναληπτικές μεθόδους.....	77
2.18 Άσκηση με επαναληπτικές μεθόδους (Παλιό θέμα)	78
2.19 Άσκηση με επαναληπτικές μεθόδους (παλιό θέμα)	79
2.20 Άσκηση με Jacobi (Θέμα Σεπτεμβρίου 2017).....	81
2.21 Μέθοδος συζυγών κλίσεων (CG) και συζυγών κλίσεων με προρυθμιστή (PCG)	82
2.22 Άσκηση με CG	85
2.23 Άσκηση MATLAB	87
2.24 Συμπεράσματα σχετικά με τις επαναληπτικές μεθόδους επίλυσης	88
Κεφάλαιο 3 – Μέθοδος δυνάμεων και κύκλοι Gershgorin.....	89
3.1 Ιδιοτιμές – Μέθοδος Δυνάμεων	89
3.1.1 Πρώτο παράδειγμα με την μέθοδο δυνάμεων	90
3.1.2 Δεύτερο παράδειγμα με τη μέθοδο δυνάμεων (Παλιό θέμα)	91
3.1.3 Τρίτο παράδειγμα με την μέθοδο δυνάμεων	91
3.1.4 Τέταρτο παράδειγμα με την μέθοδο δυνάμεων (παλιό θέμα)	92
3.1.5 Ταχύτητα σύγκλισης μεθόδου δυνάμεων	93
3.1.6 Σύγκλιση και κώδικας Matlab για τη μέθοδο δυνάμεων	94
3.1.7 Παράδειγμα αντίστροφης δύναμης με μετατόπιση	96
3.2 Κύκλοι Gershgorin	99
3.2.1 Πρώτο θέμα με κύκλους Gershgorin	101
3.2.2 Δεύτερο θέμα με κύκλους Gershgorin (παλιό θέμα)	102
3.2.3 Άσκηση με κύκλους Gershgorin.....	103
3.3 Θεώρημα Perron – Frobenius	104
3.4 Ιδιοτιμές αντίστροφου μητρώου	105
3.5 Σύγκριση συναρτήσεων <i>eig</i> και <i>eigs</i>	105
Κεφάλαιο 4 – Παρεμβολή και πολυωνυμα παρεμβολής	106
4.1 Εισαγωγή	106
4.2 Θεώρημα Weierstrass - πολυωνυμα Bernstein και καμπύλες Bezier	107
4.3 Παρεμβολή και πολυωνυμική παρεμβολή	108
4.4 Παρεμβολή του Lagrange	110
4.4 Εκτίμηση σφάλματος της πολυωνυμικής παρεμβολής	113
4.4.1 Άσκηση με άνω φράγμα σφάλματος	114
4.5 Ασκήσεις πολυωνύμων παρεμβολής στη μορφή Lagrange	114
Άσκηση 1	114

Άσκηση 2	115
Άσκηση 3 (παλιό θέμα)	116
Άσκηση 4 (παλιό θέμα)	117
4.6 Ασκήσεις πολυωνύμων παρεμβολής στη μορφή Newton με χρήση πίνακα Δ.Δ.	117
4.6.1 Σύγκριση πολυωνύμων παρεμβολής	118
Άσκηση 1 – Παρεμβολή Newton	119
Άσκηση 2 – Παρεμβολή Newton	119
Άσκηση 3 – Παρεμβολή Newton	120
Άσκηση 4 – Παρεμβολή Newton	121
Άσκηση 5 - Παρεμβολή Newton	121
Άσκηση 6 – Παρεμβολή Newton - ελαχιστοποίηση τετραγωνικού σφάλματος με γραμμικό σύστημα	122
Θεωρία ελαχίστων τετραγώνων – Πολυώνυμα παρεμβολής με τη μέθοδο ελαχίστων τετραγώνων	124
Άσκηση 7 – Ελάχιστα τετράγωνα (Θέμα Σεπτεμβρίου 2016)	125
Άσκηση 8 – Ελάχιστα τετράγωνα (Θέμα Φεβρουαρίου 2017)	126
Άσκηση 9 – Παρεμβολή Newton (Θέμα Ιουνίου 2016)	127
4.6.3 Σύγκριση μεθόδων Lagrange – Newton	127
4.7 Μοναδικότητα πολυωνύμου παρεμβολής	128
4.8 Βαρυκεντρική παρεμβολή	129
4.8.1 Πρώτη άσκηση με παρεμβολή Newton και βαρυκεντρική παρεμβολή	129
4.8.2 Δεύτερη άσκηση με παρεμβολή Newton και βαρυκεντρική παρεμβολή	131
4.9 Παρεμβολή Hermite	132
Είδη πολυωνυμικής παρεμβολής	133
4.10 Τμηματική παρεμβολή – Τμηματικά πολυώνυμα	133
4.10.1 Άσκηση εφαρμογής της κυβικής spline	134
4.10.2 Άσκηση με τμηματική γραμμική παρεμβολή	134
4.10.3 Άσκηση με τμηματική γραμμική παρεμβολή	136
4.10.4 Άσκηση με πολυώνυμα παρεμβολής (παλιό θέμα)	137
4.10.5 Άσκηση με πολυωνυμική παρεμβολή (Φεβρουάριος 2017)	138
4.10.6 Παράδειγμα πολυωνυμικής και γραμμικής παρεμβολής (παλιό θέμα)	138
4.10.7 Άσκηση με πολυώνυμα και νόρμες (παλιό θέμα)	140
4.11 Κόμβοι Chebyshev	141
4.11.1 Άσκηση κόμβων Chebyshev	141
4.12 Φαινόμενο Runge (Αστάθεια πολυωνυμικής παρεμβολής)	143
4.12.1 Αντιμετώπιση του φαινομένου Runge	144

4.13 Επαναληπτικές ασκήσεις με πολυωνύμα – 1^η Άσκηση (παλιό θέμα)	145
4.13 Επαναληπτικές ασκήσεις με πολυωνύμα – 2^η Άσκηση.....	145
4.14 Επαναληπτικές ασκήσεις με πολυωνύμα – 3^η Άσκηση: σύγκριση πολυωνύμων παρεμβολής.....	147
Κεφάλαιο 5 – Επίλυση μη γραμμικών εξισώσεων πρώτου βαθμού	150
5.1 Μη γραμμικά συστήματα εξισώσεων και ασκήσεις.....	150
1^ο παράδειγμα επίλυσης μη – γραμμικού συστήματος.....	150
2^ο παράδειγμα επίλυσης μη – γραμμικού συστήματος.....	151
3^ο παράδειγμα επίλυσης μη – γραμμικού συστήματος.....	153
5.2 Σύγκριση διχοτόμησης, Newton – Raphson, τέμνουσας, εσφαλμένης θέσης (regula falsi).....	153
5.2.1 Συμπεράσματα σύγκρισης μεθόδων διχοτόμησης, Newton – Raphson, τέμνουσας	155
5.2.2. Πλεονεκτήματα/Μειονεκτήματα μεθόδου τέμνουσας (παλιό θέμα).....	157
5.3 Ασκήσεις με διχοτόμηση και Newton – Raphson και τέμνουσα	158
Άσκηση 1	158
Άσκηση 2	159
Άσκηση 3	159
Άσκηση 4 (παλιό θέμα).....	160
Άσκηση 5	160
Άσκηση 6	162
Άσκηση 7	162
Άσκηση 8	162
Άσκηση 9	163
Άσκηση 10	163
5.4 Ρίζες πολυωνύμου – Συνοδευτικό μητρώο	164
Κεφάλαιο 6 – Ολοκλήρωση μέσω προσεγγιστικών τύπων.....	166
6.1 Απλοί κανόνες παραλληλογράμμου, τραπεζίου, Simpson, μέσου σημείου.....	166
6.2 Σύνθετοι κανόνες παραλληλογράμμου, τραπεζίου, Simpson, μέσου σημείου.....	167
6.2.1. Σύνθετη μέθοδος παραλληλογράμμου	167
6.2.2. Σύνθετη μέθοδος τραπεζίου	167
6.2.3 Σύνθετη μέθοδος Simpson	167
6.2.4 Σύνθετη μέθοδος μέσου σημείου.....	167
6.2.5 Σύνθετες μέθοδοι: Σύνοψη	167
6.3 Μελέτη Σφάλματος	168
6.4 Παράδειγμα ολοκλήρωσης με χρήση κανόνων παραλληλογράμμου, τραπεζίου, Simpson	169

6.5 Παράδειγμα ολοκλήρωσης με χρήση κανόνων παραλληλογράμμου, τραπεζίου, Simpson	170
6.6 Παράδειγμα ολοκλήρωσης με χρήση κανόνων παραλληλογράμμου, τραπεζίου, Simpson	172
6.7 Παράδειγμα ολοκλήρωσης με χρήση σύνθετου κανόνα τραπεζίου	172
6.8 Θέμα με αριθμητική ολοκλήρωση (παλιό θέμα).....	173
6.9 Θέμα με αριθμητική ολοκλήρωση (παλιό θέμα).....	175
6.10 Θέμα με αριθμητική ολοκλήρωση	176
Κεφάλαιο 7 – MATLAB – Σφάλματα – Αριθμητική κινητής υποδιαστολής - QR - SVD	178
7.1 Χρήσιμες συναρτήσεις MATLAB αναφορικά με πολυώνυμα.....	178
7.2 Αριθμοί κινητής υποδιαστολής.....	179
7.2.1. Εισαγωγή στους α.κ.υ. και τα χαρακτηριστικά του συνόλου F – Χαρακτηριστικά μεγέθη του F	179
7.2.2. Άσκηση με χρήση em και συνάρτησης roots	181
7.2.3. Αναδρομικός κώδικας με MATLAB	182
7.2.4. Άσκηση με σύστημα α.κ.υ. διαφορετικού πλήθους σημαντικών ψηφίων	182
7.2.5. Άσκηση υπολογισμού διωνυμικού συντελεστή με χρήση MATLAB	183
7.2.6. Άσκηση υπολογισμού σειράς Fibonacci με χρήση MATLAB.....	183
7.3 Εμπρός και πίσω σφάλμα στην επίλυση γραμμικού συστήματος.....	184
7.4 Προς τα πίσω ανάλυση – προς τα πίσω ευστάθεια.....	185
7.5 Χαρακτηριστικά μεγέθη και πράξεις μεταξύ α.κ.υ. – Πρότυπο IEEE 754	185
7.5.1 Καταστροφική απαλοιφή	189
7.5.2 Αρχή ακριβούς στρογγύλευσης.....	189
7.5.3 Επιπτώσεις της αριθμητικής πεπερασμένης ακρίβειας	190
7.6 Παραγοντοποίηση QR	191
7.6.1 Παράδειγμα υπολογισμού QR παραγοντοποίησης.....	193
7.6.2 Παράδειγμα υπολογισμού προσεγγιστικής λύσης γραμμικού συστήματος μέσω QR παραγοντοποίησης	195
7.6.3. Σχέση QR παραγοντοποίησης και παραγοντοποίησης Cholesky	198
7.6.4. Δεύτερη νόρμα αμετάβλητη κάτω από ορθογώνιους μετασχηματισμούς.....	198
7.6.5. Άσκηση (Θέμα Φεβρουαρίου 2017) σχετική με QR παραγοντοποίηση.....	199
7.6.6. Άσκηση με αντίστροφη QR παραγοντοποίηση	199
7.6.7. Άσκηση επίλυσης ενός γραμμικού συστήματος $A *x = b$ με QR παραγοντοποίηση....	200
7.7 Μητρώα ορθογώνιας προβολής (ΟΠ)	201
7.8 SVD Παραγοντοποίηση	202
Κεφάλαιο 8 – Άσκηση επανάληψης εφ' όλης της ύλης – Θέματα	204
Άσκηση 1	204

Ασκηση 2	204
Ασκηση 3	205
Ασκηση 4	205
Ασκηση 5	206
Ασκηση 6	207
Ασκηση 7	207
Ασκηση 8	208
Ασκηση 9	208
Ασκηση 10	209
Ασκηση 11	210
Ασκηση 12	210
Ασκηση 13	210
Ασκηση 14	211
Ασκηση 15	211
Ασκηση 16	212
Ασκηση 17 (Σεπτέμβριος 2017)	212
Ασκηση 18 (Σεπτέμβριος 2017)	212
Ασκηση 19 (Θέμα Σεπτεμβρίου 2017)	213
Ασκηση 20 (Θέμα Σεπτεμβρίου 2017)	214
Ασκηση 21 (Σεπτέμβριος 2017)	215
Ασκηση 22 (Ιούνιος 2018)	215
Ασκηση 23 (Ιανουάριος 2018)	216
Ασκηση 24 (Ιανουάριος 2018)	217
Ασκηση 25 (Ιανουάριος 2018)	217
Ασκηση 26 (Ιούνιος 2018)	218
Ασκηση 27 (Σεπτέμβριος 2019)	219
Ασκηση 28 (Φεβρουάριος 2020)	225
Ασκηση 29 (Θέμα Φροντιστηρίου)	227
Ασκηση 30 (Θέμα Φροντιστηρίου)	228
Ασκηση 31 (Θέμα Φροντιστηρίου)	229
Ασκηση 32 (Θέμα Φροντιστηρίου)	230
Ασκηση 33 (Θέμα Φροντιστηρίου)	230
Ασκηση 34 (Θέμα Φροντιστηρίου)	230
Ασκηση 34 (Θέμα Φροντιστηρίου)	231
Ασκηση 35 (Θέμα Φροντιστηρίου)	233
Ασκηση 36 (Θέμα Φροντιστηρίου)	234

Κεφάλαιο 1 - Σύντομη επανάληψη Γραμμικής Άλγεβρας και ΕΥ

1.1 Υπολογισμοί και Σφάλματα

1.1.1 Σφάλμα Στρογγυλοποίησης

Έστω $x \in R$ (βαθμωτός) και το σύνολο R είναι απειροσύνολο και διαθέτει άπειρη ακρίβεια και μπορεί να απεικονίσει οποιονδήποτε αριθμό. Έστω x^* ή \hat{x} ή $x_{prog} \in F$ (σύνολο α.κ.υ.) και το σύνολο F είναι πεπερασμένο σύνολο, με πεπερασμένη ακρίβεια και οι αριθμοί που μπορούν να παρασταθούν στο σύνολο F είναι εντός κάποιων ορίων (realmax και realmin). Επειδή δεν γίνεται να αναπαρασταθεί ένα απειροσύνολο, όπως το R , σε ένα πεπερασμένο σύνολο, όπως το F , δημιουργούνται σφάλματα αναπαράστασης (στρογγύλευσης), τα οποία είναι δύο ειδών: απόλυτα και σχετικά: $\epsilon = |x^* - x| = |x - x^*|$ απόλυτο σφάλμα και $\frac{|x^* - x|}{x}$ ή $\frac{|x - x^*|}{x}$ σχετικό σφάλμα. Αν το $x \in R^n$ (διάνυσμα), τότε $\epsilon = \|x^* - x\|$ ή $\|x - x^*\|$ και ομοίως και το σχετικό σφάλμα. Τα σφάλματα στρογγύλευσης οδηγούν και σε σφάλματα λόγω πράξεων μεταξύ των α.κ.υ.

Αν ο προσεγγιστικός αριθμός x^* είναι ακριβής σε **κ δεκαδικά ψηφία**, τότε για το απόλυτο σφάλμα του ισχύει $|\epsilon| \leq 0.5 * 10^{-k}$, ενώ αν είναι ακριβής σε **κ σημαντικά ψηφία**, τότε για το απόλυτο σφάλμα του ισχύει $|\epsilon| \leq 5 * 10^{-k} = 0.5 * 10^{1-k}$, όπου $k =$ πλήθος σημαντικών ψηφίων. Υπάρχει **διαφορά** ανάμεσα στους όρους δεκαδικά και σημαντικά ψηφία και πιο συγκεκριμένα τα σημαντικά ψηφία είναι ο αριθμός των ψηφίων ενός αριθμού που γνωρίζουμε με απόλυτη βεβαιότητα και ο τρόπος που είναι γραμμένος ένας αριθμός δηλώνει και τα σημαντικά του ψηφία, π.χ. ο αριθμός **23.21** έχει τέσσερα σημαντικά ψηφία και δύο δεκαδικά, ενώ ο αριθμός **0.062** έχει δύο σημαντικά ψηφία (τα μηδενικά μέχρι την υποδιαστολή δεν μετράνε) και τρία δεκαδικά. Ο αριθμός **60.256** έχει πέντε σημαντικά ψηφία και τρία δεκαδικά.

1.1.2 Καταστροφική ακύρωση (απαλοιφή) σημαντικών ψηφίων

Ένα σημαντικό επακόλουθο της αριθμητικής πεπερασμένης ακρίβειας είναι το φαινόμενο της καταστροφικής ακύρωσης (απαλοιφής) σημαντικών ψηφίων (catastrophic cancellation of significant digits). **Το φαινόμενο αυτό σχετίζεται με την απώλεια σωστών σημαντικών ψηφίων μικρών αριθμών, οι οποίοι απορρέουν από πράξεις μεγάλων αριθμών.**

Παράδειγμα (Θέμα Ιουνίου 2014). Ως γνωστό, για την εύρεση των ριζών $x_{1,2}$ της δευτεροβάθμιας εξίσωσης $ax^2 + bx + c = 0$, όπου $a \neq 0$, χρησιμοποιείται ο τύπος:

$$x_{1,2} = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

Όταν η τιμή του b είναι κατά πολύ μεγαλύτερη της τιμής του όρου $4ac$, τότε για την εύρεση μιας ρίζας οδηγούμαστε στην περίπτωση της αφαίρεσης δύο περίπου **ίσων** αριθμών του b και του $\sqrt{b^2 - 4ac}$. Δώστε εναλλακτικό τύπο για την εύρεση της ρίζας αυτής, ο οποίος να αποφεύγει την παραπάνω αφαίρεση.

Λύση

Προφανώς το πρόβλημα της αφαίρεσης των δυο περίπου ίσων αριθμών βρίσκεται στην περίπτωση του υπολογισμού της παρακάτω ρίζας: $x_1 = \frac{-b + \sqrt{b^2 - 4ac}}{2a}$. Ο αριθμός $\sqrt{b^2 - 4ac}$ είναι πολύ κοντά στο b και όταν αφαιρέσουμε το b, τα σφάλματα παραμένουν και όταν διαιρέσουμε με το 2a (που είναι μικρό) μεγεθύνονται και επηρεάζουν το αποτέλεσμα. Για **εξάλειψη του φαινομένου της καταστροφικής απαλοιφής**, πρέπει να καταργήσουμε τις πράξεις στον αριθμητή που δημιουργούν τα σφάλματα και για να γίνει αυτό, πολλαπλασιάζουμε και τους δύο όρους του κλάσματος με το συζυγή του αριθμητή.

Για να δώσουμε λοιπόν ένα εναλλακτικό τύπο για την εύρεση της ρίζας αυτής, ο οποίος να αποφεύγει την αφαίρεση των δύο περίπου ίσων αριθμών, εργαζόμαστε ως εξής:

$$\frac{-b + \sqrt{b^2 - 4ac}}{2a} = \left(\frac{-b + \sqrt{b^2 - 4ac}}{2a} \right) * \left(\frac{b + \sqrt{b^2 - 4ac}}{b + \sqrt{b^2 - 4ac}} \right) = \frac{-4ac}{2a(b + \sqrt{b^2 - 4ac})} = \frac{-2c}{b + \sqrt{b^2 - 4ac}}$$

Επομένως, η αντίστοιχη ρίζα μπορεί να βρεθεί από τον παρακάτω εναλλακτικό τύπο: $x_1 = \frac{-2c}{b + \sqrt{b^2 - 4ac}}$.

Παρατήρηση: Μια παραλλαγή της προηγούμενης άσκησης είναι η εξής: Να χρησιμοποιηθούν οι προηγούμενοι τύποι υπολογισμού των ριζών της δευτεροβάθμιας εξίσωσης, όταν δοθούν ως συντελεστές οι αριθμοί **a = 1.**, **b = -100000000** και **c = 1**. Να συγκριθούν τα υπολογισμένα αποτελέσματα με τη συνάρτηση **roots([a b c])**. Τι συμβαίνει όταν υπολογίζονται οι ρίζες με ένα calculator; Θα πρέπει να βρείτε ότι οι κλασικοί τύποι είναι καλοί για τον υπολογισμό μιας ρίζας, αλλά όχι και της άλλης. Για το λόγο αυτό, για τον υπολογισμό της δεύτερης ρίζας θα πρέπει να χρησιμοποιηθεί ο τύπος $x_1x_2 = \frac{c}{a}$.

Λύση

Λύση. Πρόκειται για ένα τριώνυμο που, ενώ έχει θετική διακρίνουσα, είναι πρακτικά αδύνατο να βρούμε τις ακριβείς ρίζες! Εδώ παραλείπουμε τους υπολογισμούς καθώς είναι τετριμένοι. Σχολιάζουμε μόνο τις περιπτώσεις ριζών του τριωνύμου. Έστω οι ρίζες του τριωνύμου όπως υπολογίζονται κανονικά

$$x_1 = \frac{-b - \sqrt{\Delta}}{2a}, \quad x_2 = \frac{-b + \sqrt{\Delta}}{2a}$$

Από τις ιδιότητες των τόπων *Viete* ξέρουμε πως, θεωρητικά, και οι αριθμοί $\frac{c}{a \cdot x_1}$ και $\frac{c}{a \cdot x_2}$ πρέπει να είναι ρίζες. Αν θέσουμε x_5, x_6 τις ρίζες που δίνει η συνάρτηση **roots** της MATLAB, τότε έχουμε τις εξής περιπτώσεις ριζών με αντίστοιχες τιμές:

$[x_1, x_2]$	$[x_1, \frac{c}{a \cdot x_1}]$	$[x_2, \frac{c}{a \cdot x_2}]$	$[x_5, x_6]$
[1.0e+08, 7.4506e-09]	[1.0e+08, 1.0e-08]	[7.4506e-09, 134217728]	[1.0e+08, 1.0e-08]

Παρατηρήσεις:

- Οι καλύτερη περίπτωση, ως προς την απόκλιση από το 0, είναι η 2η και η 4η. Επισημαίνεται πως είναι λάθος να πει κανείς πως οι ρίζες είναι οι αριθμοί 7.4506e-09 και 1.0e-08.
- Γενικά, για τη συγκεκριμένη περίπτωση, δεν υπάρχει βέλτιστος τρόπος να βρεθεί η “μεγάλη” ρίζα, αφού και για $x = 10^8$ το αποτέλεσμα είναι 1.
- Ο κλασικός τύπος αποτυγχάνει για τη “μικρή” ρίζα αφού έχουμε καταστροφική απαλοιφή. Κατ’ επέκταση και ο τρίτος τρόπος είναι ανούσιος.

Συμπέρασμα: Διαπιστώνουμε ότι οι ακριβείς λύσεις του προηγούμενου τριώνυμου είναι όσες δεν περιέχουν τη λύση x_2 , η οποία αποτυγχάνει λόγω καταστροφικής απαλοιφής. Πράγματι στην περίπτωση υπολογισμού της ρίζας x_2 με τον κλασσικό τρόπο έχουμε αφαίρεση δύο αριθμών ($-b$ και $\sqrt{Δ}$) που είναι πολύ κοντά ο ένας στον άλλο και στη συνέχεια διαίρεση με ένα πολύ μικρό αριθμό (2α).

Άσκηση: Η καταστροφική απαλοιφή αφορά το πρόβλημα που δημιουργείται:

α) όταν αφαιρούνται δύο α.κ.υ. καθένας από τους οποίους είναι αποτέλεσμα υπολογισμών και στρογγυλεύσεων και το αποτέλεσμα της αφαίρεσης υπόκειται σε στρογγύλευση.

β) όταν αφαιρούνται δύο αριθμοί που διαφέρουν τόσο πολύ σε μέγεθος που το αποτέλεσμα μετά τη στρογγύλευση είναι ίδιο με το μεγαλύτερο από αυτούς.

γ) όταν αφαιρούνται δύο σχεδόν ίσοι αριθμοί που αναπαρίστανται χωρίς σφάλμα στο σύστημα αριθμητικής κινητής υποδιαστολής και το αποτέλεσμα της αφαίρεσης υπόκειται σε στρογγύλευση.

δ) Κανένα από τα υπόλοιπα

Λύση

Η σωστή απάντηση είναι η (δ), διότι οι αριθμοί που πολλαπλασιάζονται και αφαιρούνται μεταξύ τους είναι αριθμοί α.κ.υ. και έχουν εξ' ορισμού σφάλμα στρογγύλευσης, επομένως από τις πράξεις προκύπτουν και σφάλματα υπολογισμών, τα οποία στη συνέχεια μεγαλώνουν. Επίσης, οι αριθμοί που αφαιρούνται πρέπει να είναι κοντά ο ένας στον άλλο.

1.1.3 Δεύτερο παράδειγμα καταστροφικής απαλοιφής

Όταν $x \approx y$ τότε ο υπολογισμός της διαφοράς $\log x - \log y$ μπορεί να οδηγήσει σε καταστροφική απαλοιφή. Αντί αυτού, θα μπορούσαμε να χρησιμοποιήσουμε την ισοδύναμη μαθηματικά έκφραση $\log(x/y)$. Είναι αληθές ότι ο δεύτερος τρόπος δίνει πάντα πιο ακριβή αποτελέσματα; Με βάση τις “ενασθησίες” τις λογαριθμικής συνάρτησης, να βρείτε περιπτώσεις για τα x, y για τις οποίες η πρώτη μέθοδος δίνει ορθά αποτελέσματα (και όχι η δεύτερη) και αντίστροφα.

Λύση. Το σκεπτικό πίσω από αυτήν την άσκηση ήταν η κατανόηση των υπολογισμών σε α.κ.υ.. Μας ενδιαφέρει το πως θα έπρεπε να συμπεριφέρεται ο υπολογισμός ενός λογαρίθμου για διάφορα ορίσματα.

- Έχουμε εμφάνιση απώλειας σημαντικών ψηφίων όταν οι x, y είναι πολύ μικροί αριθμοί, πχ της τάξης του $1.0e-18$ και περίπου ίσοι. Σε τέτοιες περιπτώσεις βολεύει ο υπολογισμός με πηλίκο.
- Αν ο x είναι πολύ μεγάλος (τάξης $1.0e+150$), ενώ ο y πάρα πολύ μικρός, στον υπολογισμό του x/y μπορεί να προκληθεί υπερχείλιση. Σε αυτήν την περίπτωση θα ήταν προτιμότερος ο 1ος τρόπος.
- Ομοίως, όταν ο y είναι πολύ μεγάλος και ο x πολύ μικρός, με τον 2ο τρόπο θα έπρεπε, να υπολογιστεί ο λογάριθμος ενός αριθμού πολύ κοντά στο 0.

1.1.4 Άσκηση με ακριβεία σημαντικών ψηφίων

Να υπολογιστεί η συνάρτηση $f(x) = 1 - \cos(x)$ με αριθμητική 5 σημαντικών ψηφίων για τιμές κοντά στο 0. Προσοχή! Χρησιμοποιήστε μια ταυτότητα τριγωνομετρίας για να αποφύγετε την απώλεια σ.ψ. στο αποτέλεσμα. Εξετάστε τα αποτελέσματα με και χωρίς μετατροπή.

Λύση. Στον παρακάτω πίνακα παρουσιάζουμε τα αποτελέσματα που λαμβάνουμε για τις τρεις διαφορετικές εκφράσεις μέσω MATLAB.

x	0.1	10^{-4}	10^{-8}
$1 - \cos(x)$	0.00499	5.0000e-09	0
$2\sin^2(x/2)$	0.00499	5.0000e-09	5.0000e-17
$\frac{\sin^2(x)}{1+\cos(x)}$	0.00499	5.0000e-09	5.0000e-17

Πρακτικά δεν εμφανίζονται ουσιαστικές διαφορές λόγω της πολύ καλής εκτίμησης που ούτως ή άλλως κάνει η MATLAB. Δοκιμάζοντας όμως με κάποια αριθμομηχανή χειρός ή κινητό παρουσιάζονται πιο σημαντικές αποκλίσεις.

1.1.5 Άσκηση με σφάλματα

Δίνονται οι αριθμοί $\alpha = 0,731$ και $b = 9,12$ στρογγυλοποιημένοι σε δεκαδικά ψηφία, να βρεθεί άνω φράγμα (α) για το απόλυτο σφάλμα του αθροίσματός τους $\alpha + b$ καθώς και (β) να δείξετε ότι η τιμή $0,0015$ είναι κατά προσέγγιση ένα άνω φράγμα για το απόλυτο σφάλμα του πηλίκου α/b .

Λύση

Εάν ε_1 και ε_2 είναι τα σφάλματα στρογγυλοποίησης (στρογγύλευσης) των αριθμών $\alpha = 0,731$ και $b = 9,12$ αντίστοιχα, τότε θα ισχύει: $|\varepsilon_1| \leq \frac{1}{2}10^{-3}$ και $|\varepsilon_2| \leq \frac{1}{2}10^{-2}$. Για την περίπτωση (α) αν υποθέσουμε ότι ε είναι το σφάλμα του αθροίσματος $\alpha + b$, τότε μπορούμε να πάρουμε ότι: $|\delta| \cong |\delta_1| + |\delta_2| \cong \frac{|\varepsilon_1|}{|\alpha|} + \frac{|\varepsilon_2|}{|b|} \leq 0.5 * \frac{10^{-3}}{0.731} + 0.5 * \frac{10^{-2}}{9.12} \cong 0.0006839 + 0.0005482 \cong 0.0012321 \leq 0.0015$, το οποίο είναι και το ζητούμενο άνω φράγμα.

Για την περίπτωση (β) αν υποθέσουμε ότι δ_1 , δ_2 και δ είναι αντίστοιχα τα σχετικά σφάλματα των αριθμών $\alpha = 0.731$, $b = 9.12$ και του πηλίκου α/b , τότε μπορούμε να πάρουμε ότι: $|\delta| \cong |\delta_1| + |\delta_2| \cong \frac{|\varepsilon_1|}{|\alpha|} + \frac{|\varepsilon_2|}{|b|} \leq 0.5 * \frac{10^{-3}}{0.731} + 0.5 * \frac{10^{-2}}{9.12} \cong 0.0006839 + 0.0005482 \cong 0.0012321 \leq 0.0015$. Επομένως η τιμή 0.0015 είναι κατά προσέγγιση ένα άνω φράγμα για το απόλυτο σχετικό σφάλμα του πηλίκου α/b .

1.2 Αναφορά στοιχειωδών εννοιών για μητρώα και διανύσματα

Επειδή για κάθε μητρώο δεν υπάρχει το αντίστροφο μητρώο A^{-1} , αν βρεθεί, τότε το μητρώο A ονομάζεται **αντιστρέψιμο** ή αλλιώς ομαλό ή μη ιδιάζον (non-singular), διαφορετικά λέγεται μη ομαλό ή **ιδιάζον** (singular) ή **μη-αντιστρέψιμο**. Ισχύει ότι αν \exists το A^{-1} , τότε $A * A^{-1} = A^{-1} * A = I$. Υπάρχει μια ειδική περίπτωση, όπου το $A^T = A^{-1}$, οπότε τότε ισχύει ότι: $A * A^T = A^T * A = I$. Στην περίπτωση αυτή το μητρώο A ονομάζεται **ορθογώνιο** (orthogonal) μητρώο. Επιπλέον, ένα **ορθογώνιο** μητρώο έχει στήλες **είναι κάθετες μεταξύ τους**. Αν τετραγωνικό μητρώο $A \in C^{n \times n}$, υπάρχει το **αντίστροφο μητρώο A^{-1}** , όταν ισχύει **ένα** από τα παρακάτω:

- $\det(A) \neq 0$
- $\text{rank}(A) = n$. Η τάξη ενός μητρώου προσδιορίζει το πλήθος των **μη – μηδενικών οδηγών** του μητρώου (τα στοιχεία στην κύρια διαγώνιο του U) ή εναλλακτικά προσδιορίζει το πλήθος των γραμμικά ανεξάρτητων γραμμών ή στηλών του μητρώου. Διευκρινιστικά, ακολουθούν τα εξής παραδείγματα:

```
B =
2 -2 1
4 3 31
9 4 9
>> det(B)
ans =-691
>> rank(B)
ans = 3
```

```

>> inv(B)
ans =
  0.1404 -0.0318  0.0941
 -0.3517 -0.0130  0.0839
  0.0159  0.0376 -0.0203
C =
  1   2   3
  2   4   6
 11  12  16
>> det(C)
ans = 0
>> rank(C)
ans = 2
>> inv(C)
Warning: Matrix is singular to working precision.
ans =
    Inf  Inf  Inf
    Inf  Inf  Inf
    Inf  Inf  Inf

```

- Το “0” **δεν** είναι ιδιοτιμή του A. Αν χρησιμοποιήσουμε τα μητρώα B (αντιστρέψιμο) και C (μη – αντιστρέψιμο) από το προηγούμενο παράδειγμα και εφαρμόσουμε σε καθένα από αυτά την ενδογενή συνάρτηση eig της MATLAB, για υπολογισμό των ιδιοτιμών των μητρώων, θα πάρουμε τα ακόλουθα αποτελέσματα:

```
>> eig(B)
```

```
ans =
```

```
 15.9219
  5.6966
 -7.6185
```

```
>> eig(C)
```

```
ans =
```

```
 22.1297
 -1.1297
 -0.0000
```

- τα διανύσματα γραμμές/στήλες του A είναι γραμμικά ανεξάρτητα. Αυτό υποδεικνύεται από την τάξη του μητρώου (rank) που αναφέρθηκε προηγουμένως. Av $\text{rank}(A) = n$, όπου $A \in \mathbb{R}^{n \times n}$ τότε είναι γρ. ανεξάρτητες.
- το ομογενές γραμμικό σύστημα $Ax = 0$ έχει μοναδική λύση $x = 0$. Αν έχει και άλλες λύσεις εκτός αυτής, τότε το μητρώο δεν αντιστρέφεται.
- το μη – ομογενές γραμμικό σύστημα $Ax = b$ έχει μοναδική λύση $x = A^{-1} * b$.

Μια **ιδιοτιμή (eigenvalue) λ** και το αντίστοιχο **ιδιοδιάνυσμα (eigenvector) x** ενός μητρώου ικανοποιούν την εξής ιδιότητα: $A * x = \lambda * x$. Οι n ιδιοτιμές ενός μητρώου $A \in \mathbb{R}^{n \times n}$ μπορούν να βρεθούν από την επίλυση

της εξίσωσης: $\det(\lambda I - A) = 0 \rightarrow \lambda^n + c_{n-1}\lambda^{n-1} + c_{n-2}\lambda^{n-2} + \dots + c_1\lambda + c_0 = 0$ που ονομάζεται **χαρακτηριστικό πολυώνυμο** και οι ρίζες του $\lambda_1, \lambda_2, \dots, \lambda_n$ που δεν είναι κατ' ανάγκη διακριτές, μας παρέχουν τις **ιδιοτιμές** του μητρώου A. Το σύνολο των ιδιοτιμών ενός μητρώου A, ονομάζεται **φάσμα (spectrum)** του A και συμβολίζεται ως: $\lambda(A) = \{\lambda_1, \lambda_2, \dots, \lambda_n\}$. Σε κάθε ιδιοτιμή αντιστοιχεί και ένα **ιδιοδιάνυσμα** και αυτά υπολογίζονται από τον τύπο $A * x = \lambda * x \Rightarrow (\lambda * I - A) * x = 0$. Ο τελευταίος τύπος εφαρμόζεται για κάθε ιδιοτιμή λ_i του μητρώου και μπορεί εναλλακτικά να γραφτεί στη μορφή: $(\lambda_i * I - A) * x_i = 0, i = 1, \dots, n$. Στη MATLAB οι ιδιοτιμές ενός μητρώου A υπολογίζονται από την ενδογενή συνάρτηση eig, δηλαδή γράφουμε eig(A). Η απολύτως μεγαλύτερη ιδιοτιμή του μητρώου A ονομάζεται **φασματική ακτίνα** (spectral radius) ή **κυρίαρχη ιδιοτιμή** και συμβολίζεται ως: $\rho(A) = \max_{i=1 \dots n} |\lambda_i|$. Μπορεί εύκολα να αποδειχθεί ότι:

α) Οι ιδιοτιμές ενός διαγώνιου ή τριγωνικού μητρώου είναι τα διαγώνια στοιχεία του μητρώου.

```
>> B =
2 -2 1
4 3 31
9 4 9
>> B1=triu(B): επιστρέφει το άνω τριγωνικό μέρος του μητρώου B
B1 = (άνω τριγωνικό)
2 -2 1
0 3 31
0 0 9
>> B2=tril(B): επιστρέφει το κάτω τριγωνικό μέρος του μητρώου B
B2 =
2 0 0
4 3 0
9 4 9
>> B3=diag(diag(B)): δημιουργεί διαγώνιο μητρώο
B3 =
2 0 0
0 3 0
0 0 9
>> eig(B)
ans =
15.9219
5.6966
-7.6185
>> eig(B1)
ans =
2
3
9
>> eig(B2)
```

```

ans =
9
3
2
>> eig(B3)
ans =
2
3
9

```

β) Αν λ είναι ιδιοτιμή του μητρώου A , τότε η ιδιοτιμή λ^{-1} θα είναι ιδιοτιμή του μητρώου A^{-1} (αν υπάρχει αυτό).

γ) **Αν το A είναι μη – αντιστρέψιμο, τότε έχει τουλάχιστον μια μηδενική ιδιοτιμή.** Εναλλάσσοντας δύο γραμμές/στήλες του A , οι ιδιοτιμές παραμένουν ίδιες.

δ) Τα μητρώα A και A^T έχουν τις ίδιες ιδιοτιμές.

ε) Η ορίζουσα $\det(A) = \lambda_1 * \lambda_2 * \dots * \lambda_n$ και το ίχνος $\text{tr}(A) = a_{11} + a_{22} + \dots + a_{nn} = \lambda_1 + \lambda_2 + \dots + \lambda_n$.

στ) Στα ορθογώνια μητρώα οι ιδιοτιμές έχουν απόλυτη τιμή ίση με το «1».

ζ) Επίσης, τα **θετικά ορισμένα μητρώα έχουν θετικές ιδιοτιμές, ενώ τα θετικά ημιορισμένα έχουν μη αρνητικές ιδιοτιμές.** Στο επόμενο παράδειγμα, το μητρώο A είναι θετικά ορισμένο, διότι έχει A.D.K. κατά γραμμές (ή στήλες) και επίσης έχει θετικά διαγώνια στοιχεία. Παρατηρούμε ότι, όλες οι ιδιοτιμές του είναι θετικές. Επίσης, στο δεύτερο παράδειγμα, το μητρώο A είναι θετικά ημιορισμένο, διότι έχει Δ.Κ. Παρατηρούμε ότι και στην περίπτωση αυτή, οι ιδιοτιμές είναι μη αρνητικές.

```

A =
12   3   -1
-1   8    4
 1   2    6
>> eig(A)
ans =
11.4641
4.5359
10.0000
A =
12   3    2
 1   8    4
 4   2    6
>> eig(A)
ans =
14.6489
6.8973
4.453

```

Οι **νόρμες διανυσμάτων** είναι οι ακόλουθες:

- $\|x\|_1 = \sum_{i=1}^n |x_i|$
- $\|x\|_2 = [\sum_{i=1}^n |x_i|^2]^{1/2}$. Αυτή είναι η **προεπιλεγμένη νόρμα του διανύσματος**.

Η **δεύτερη νόρμα** ενός διανύσματος μπορεί μέσω κώδικα MATLAB να υπολογιστεί και ως εξής:

```
function [s]=norm2_naive(x);
n = length(x);
s = 0;
for i = 1:n, s = s+x(i)^2; end
s = sqrt(s);
% faster version
% s = sqrt(sum(x.^2));
```

- $\|x\|_\infty = \max_i |x_i|$

και οι **ιδιότητές** τους είναι οι εξής:

- $\|x\| > 0$, εκτός και αν $x = 0$, οπότε $\|0\| = 0$.
- $\|cx\| = |c|\|x\|$, όπου c σταθερά
- $\|x + y\| \leq \|x\| + \|y\|$, όπου $x, y \in C^n$, δηλαδή ισχύει στις νόρμες η τριγωνική ανισότητα που ισχύει στα Μαθηματικά μεταξύ αριθμών.

Οι **νόρμες μητρώων** είναι οι ακόλουθες:

- $\|A\|_1 = \max_j \sum_{i=1}^m |a_{ij}|$, δηλ. το μέγιστο κατ' απόλυτη τιμή άθροισμα από όλες τις στήλες.
- $\|A\|_2 = [\rho(A^T A)]^{1/2}$. Αυτή είναι η **προεπιλεγμένη νόρμα του μητρώου**.
- $\|A\|_\infty = \max_i \sum_{j=1}^n |a_{ij}|$ δηλ. το μέγιστο κατ' απόλυτη τιμή άθροισμα από όλες τις γραμμές.

και οι ιδιότητές τους είναι οι εξής:

- $\|A\| > 0$, εκτός και αν $A = 0$, οπότε $\|0\| = 0$
- $\|cA\| = |c|\|A\|$, όπου c σταθερά
- $\|A + B\| \leq \|A\| + \|B\|$, όπου $A, B \in C^{m \times n}$
- $\|AB\| \leq \|A\|\|B\|$, όπου $B \in C^{n \times p}$

Επίσης υπάρχει και η **νόρμα** (στάθμη) **Frobenius** (Frobenius norm) που δίνεται ως εξής: $\|A\|_F = \left[\sum_{i=1}^m \sum_{j=1}^n |a_{ij}|^2 \right]^{1/2} = \sqrt{\text{sum}(\text{diag}(A^T * A))}$. Στη MATLAB οι νόρμες των μητρώων υπολογίζονται με τις συναρτήσεις: **norm(A, 1)**, **norm(A, 2)** ή απλά **norm(A)**, **norm(A, Inf)**, **norm(A, 'fro')**. Η πρώτη νόρμα ενός μητρώου A - για την περίπτωση που αυτό έχει **μόνο θετικά στοιχεία** - μπορεί εναλλακτικά να υπολογιστεί και μέσω των ενδογενών συναρτήσεων **max**, **sum** της MATLAB. Δηλαδή γράφουμε: **max(sum(A, 1))**. Για τη νόρμα μεγίστου ή απείρου ενός μητρώου A γράφουμε: **max(sum(A, 2))**.

Παρατήρηση: Η συνάρτηση **sum(A, 1)** υπολογίζει το άθροισμα από κάθε **στήλη** του μητρώου A και η **sum(A, 2)** υπολογίζει το άθροισμα από κάθε **γραμμή** του μητρώου A .

Στη συνέχεια παρουσιάζονται ορισμένα παραδείγματα υπολογισμού νορμών διανυσμάτων και μητρώων:

```
>> x = [1 -1 2 4]
x = 1 -1 2 4
>> norm(x, 1)
ans = 8
>> norm(x, 2)
ans = 4.6904
>> norm(x, Inf)
ans = 4
>> A = [1 2 3; 4 5 6; 7 8 9]
A =
1 2 3
4 5 6
7 8 9
>> norm(A, 1)
ans = 18
>> norm(A, 2)
ans = 16.8481
>> norm(A, Inf)
ans = 24
>> norm(A, 'fro')
ans = 16.8819
```

Ο αριθμός κατάστασης ή συντελεστής κατάστασης ή **δείκτης κατάστασης (condition number)** του **τετραγωνικού και αντιστρέψιμου μητρώου A** ($\det(A) \neq 0$) είναι το μέγεθος: $\kappa \equiv \kappa(A) = \|A\| * \|A^{-1}\|$. Στη MATLAB οι δείκτες κατάστασης των μητρώων υπολογίζονται με τις συναρτήσεις: $\text{cond}(A, 1)$, $\text{cond}(A, 2)$, $\text{cond}(A, \text{Inf})$.

Διαγωνοποίηση μητρώου:

Αν ένα μητρώο $A \in \mathbb{R}^{n \times n}$ περιέχει **διακριτές ιδιοτιμές** (δηλ. ιδιοτιμές με αλγεβρική πολλαπλότητα = 1, δηλ. ιδιοτιμές που δεν επαναλαμβάνονται, είναι μοναδικές), τότε **διαγωνοποιείται**, δηλ. γράφεται στη μορφή $A = P * D * P^{-1}$, όπου D = διαγώνιο μητρώο που περιέχει στην κύρια διαγώνιο του τις ιδιοτιμές του A και P = μητρώο που περιέχει στις στήλες του τα ιδιοδιανύσματα του A . **Αν ένα μητρώο είναι τριγωνικό (άνω/κάτω) ή διαγώνιο, τότε οι ιδιοτιμές του είναι τα στοιχεία της κύριας διαγωνίου.** Έστω μητρώο $A = \begin{bmatrix} 1 & 1 \\ 0 & 2 \end{bmatrix}$. Τότε αυτό είναι **διαγωνοποιήσιμο** (διότι έχει δύο ιδιοτιμές διακριτές που είναι το 1 και 2, δηλαδή τα στοιχεία της κύριας διαγωνίου). Στη συνέχεια η διαγωνοποίηση του μητρώου A είναι: $A = P * D * P^{-1} = \begin{bmatrix} -1 & 1 \\ 1 & -2 \end{bmatrix} * \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} * \begin{bmatrix} -1 & 1 \\ 1 & -2 \end{bmatrix}^{-1}$. Για εύρεση των ιδιοδιανύσμάτων, χρησιμοποιείται ο τύπος $A * v = \lambda * v \Rightarrow \lambda * v - A * v = 0 \Rightarrow (\lambda * I - A) * v = 0$. Από αυτήν την εξίσωση υπολογίζουμε για κάθε ιδιοτιμή το αντίστοιχο ιδιοδιάνυσμα. Πιο συγκεκριμένα:

$$\lambda_1 = 1 \rightarrow (\lambda_1 * I - A) * v_1 = 0 \Rightarrow v_1 = [-1 \ 1]^T.$$

$$\lambda_2 = 2 \rightarrow (\lambda_2 * I - A) * v_2 = 0 \Rightarrow v_2 = [1 \ -2]^T.$$

Κεφάλαιο 2 – Άμεσες και επαναληπτικές μέθοδοι

2.1 Επίλυση γραμμικού συστήματος

Θα δοθούν αριθμητικές μέθοδοι επίλυσης του προβλήματος: $A * x = b$, όπου $A \in R^{n \times n}$, $\det(A) \neq 0$, $b \in R^n$, $b \neq 0$. Το πρόβλημα αυτό μπορεί να αντιμετωπιστεί με το γνωστό τύπο του Cramer. Ο τύπος αυτός δίνει τη λύση με έναν εύχρηστο τρόπο, όμως η εφαρμογή του για μεγάλα n είναι πρακτικά ανέφικτη. Ένας εναλλακτικός τρόπος για την επίλυση του συστήματος που περιγράψαμε, που μπορεί να μας δώσει την λύση σε κλειστή μορφή, μπορεί να πραγματοποιηθεί με τη χρήση του αντιστρόφου A^{-1} του μητρώου A .

Γενικά, το πρόβλημα της επίλυσης γραμμικών συστημάτων η εξισώσεων με οι αγνώστους, αντιμετωπίζεται με αριθμητικές μεθόδους, μερικές από τις οποίες θα αναπτύξουμε στη συνέχεια. Οι μέθοδοι που θα αναπτύξουμε, μπορούν να διακριθούν σε δύο μεγάλες κατηγορίες μεθόδων αριθμητικής επίλυσης γραμμικών συστημάτων. Αυτές είναι:

a) Άμεσες μέθοδοι (direct methods)

b) Έμμεσες ή επαναληπτικές μέθοδοι (iterative methods)

Για την επίλυση του γραμμικού συστήματος $A * x = b$ χρησιμοποιούμε τόσο τις **άμεσες** μεθόδους όσο και τις **επαναληπτικές** μεθόδους. Αν το αρχικό πρόβλημα επίλυσης ενός συστήματος είναι **μη γραμμικό**, οι επαναληπτικές μέθοδοι (π.χ. Newton που αναλύεται παρακάτω) είναι η μόνη διέξοδος. Στη συνέχεια περιγράφονται οι άμεσες και οι επαναληπτικές μέθοδοι και τα κυριότερα χαρακτηριστικά τους:

A. Άμεσες Μέθοδοι: η λύση βρίσκεται, χρησιμοποιώντας **ακριβή** αριθμητική, μετά από ένα **πεπερασμένο** πλήθος πράξεων. Οι **άμεσες** μέθοδοι είναι οι εξής:

- LU με μερική οδήγηση
- Απαλοιφή Gauss (απλή LU)
- Παραγοντοποίηση Cholesky
- QR παραγοντοποίηση

Χαρακτηριστικά – Ιδιότητες άμεσων μεθόδων

1. Οι άμεσες μέθοδοι (επιλυτές) μετασχηματίζονται με **προκαθορισμένες διαδικασίες** που συνήθως είναι ανεξάρτητες των δεδομένων και με **αριθμητική άπειρης ακρίβειας** (θεϊκή αριθμητική), η ακριβής λύση υπολογίζεται μετά από ένα **προβλέψιμο και πεπερασμένο πλήθος πράξεων**.
2. **Δεν** ενδείκνυνται για επίλυση **μεγάλων γραμμικών συστημάτων**, γιατί είναι πολύ **χρονοβόρες** και απαιτούν **μεγάλο όγκο πράξεων**.
3. Λόγω της εύθραυστης τεχνολογίας των άμεσων μεθόδων σε διάφορες εφαρμογές, οι επαναληπτικές μέθοδοι χρησιμοποιούνται μόνο όταν **δεν** υπάρχει εναλλακτική λύση.

B. Επαναληπτικές μέθοδοι: προσεγγίζουν τη λύση του συστήματος, αρχίζοντας με ένα αρχικό διάνυσμα, το οποίο δίνεται από την εκφώνηση και είναι το $x^{(0)}$ και εκφράζει την αρχική προσέγγιση λύσης και στη συνέχεια κατασκευάζουν μια ακολουθία διανυσμάτων, η οποία σύμφωνα με ορισμένες προϋποθέσεις και με ακριβή αριθμητική, συγκλίνει οριακά στην ακριβή λύση. Οι επαναληπτικές μέθοδοι είναι οι εξής:

- Jacobi
- Gauss – Seidel
- Richardson
- Newton (για επίλυση μη γραμμικών συστημάτων)
- CG παραγοντοποίηση (μέθοδος συζυγών κλίσεων)

Χαρακτηριστικά – Ιδιότητες επαναληπτικών μεθόδων

1. Στις επαναληπτικές μεθόδους κατασκευάζεται μια ακολουθία διαδοχικών προσεγγίσεων (βημάτων) λύσης και η διαδικασία σταματά, όταν η προσέγγιση θεωρηθεί αρκετά ακριβής.
2. Εφαρμόζονται για πολύ μεγάλα γραμμικά συστήματα, δηλαδή για πολύ μεγάλα μητρώα, αλλά και για μητρώα ειδικής δομής. Δηλαδή η συμπεριφορά τους εξαρτάται από τις δομικές και μαθηματικές ιδιότητες των μητρώων.
3. Η προσεγγιστική λύση είναι της μορφής: $x^{(k+1)} = \Phi(k, A, b, x^{(k)}, x^{(k-1)}, \dots)$ όπου Φ είναι συνάρτηση επανάληψης, k είναι το βήμα επανάληψης, A και b το μητρώο των συντελεστών και το διάνυσμα αντίστοιχα του γραμμικού συστήματος $A * x = b$, $x^{(k)}$ είναι η προσέγγιση της λύσης στο k -οστό βήμα και $x^{(k-1)}$ είναι η προσέγγιση της λύσης στο $(k-1)$ -οστό βήμα.
4. Προσφέρουν χαμηλό κόστος και ταχύ ρυθμό προσέγγισης λύσης.

2.2 Σύγκριση άμεσων και επαναληπτικών μεθόδων

Για την επίλυση ενός γραμμικού συστήματος μπορεί να χρησιμοποιηθούν τόσο άμεσες μέθοδοι επίλυσης όσο και επαναληπτικές μέθοδοι επίλυσης. Στην πρώτη περίπτωση και με **άπειρη ακρίβεια**, η ακριβής λύση του γραμμικού συστήματος υπολογίζεται μετά από ένα **προβλέψιμο και πεπερασμένο πλήθος πράξεων**. Στην κατηγορία αυτή ανήκουν η απαλοιφή Gauss, η LU παραγοντοποίηση, η Cholesky, η QR κλπ.

Στη δεύτερη περίπτωση κατασκευάζεται μια ακολουθία διαδοχικών προσεγγίσεων της λύσης και η διαδικασία σταματά, όταν η προσέγγιση θεωρηθεί αρκετά ακριβής. Στην κατηγορία αυτή ανήκουν οι μέθοδοι Jacobi, Gauss – Seidel, Richardson, Newton κ.λ.π. **Σε πολύ μεγάλα μητρώα καθώς και σε μητρώα ειδικής δομής, οι άμεσες μέθοδοι είναι απαγορευτικές**. Το ζητούμενο χρήσης μιας επαναληπτικής μεθόδου για την επίλυση ενός γραμμικού συστήματος είναι να έχει χαμηλό κόστος και ταχύ ρυθμό προσέγγισης λύσης.

2.3. Άμεσες μέθοδοι – LU παραγοντοποίηση με μερική οδήγηση

2.3.1 Κόστος και μειονεκτήματα οδήγησης – πότε είναι απαραίτητη – περιγραφή μερικής οδήγησης

Η παραγοντοποίηση LU έχει πολλές παραλλαγές. Προϋπόθεση ύπαρξης LU παραγοντοποίησης ενός μητρώου: Αυστηρή διαγώνια κυριαρχία κατά στήλες. Δηλαδή αν το μητρώο $A \in \mathbb{R}^{n \times n}$ είναι αυστηρά ΔΚ κατά στήλες, τότε υπάρχει η παραγοντοποίηση $A = LU$. **Εναλλακτικά**, αν αντιστρέψιμο μητρώο $A \in \mathbb{R}^{n \times n}$ υπάρχει η LU παραγοντοποίηση αν και μόνο αν όλα τα κυρίαρχα υπομητρώα $1:i, 1:i$ του A είναι όλα αντιστρέψιμα. Η LU παραγοντοποίηση μπορεί να πραγματοποιηθεί με/χωρίς οδήγηση. Στη συνέχεια παρατίθενται δύο παραδείγματα. **Στο πρώτο από αυτά, όπου όλα τα κυρίαρχα υπομητρώα του A είναι αντιστρέψιμα, αλλά το μητρώο A δεν έχει A.D.K. κατά στήλες, επιστρέφεται ένα μητρώο L που δεν είναι κάτω τριγωνικό, αλλά είναι «ψυχολογικά κάτω τριγωνικό».** Στο δεύτερο από αυτά, όπου το μητρώο A εκτός από τα αντιστρέψιμα κυρίαρχα υπομητρώα έχει επιπλέον και A.D.K. κατά στήλες, επιστρέφεται ένα μητρώο L που είναι κάτω τριγωνικό, με «1» στην κύρια διαγώνιο.

>> A = [2 -1 3; 4 4 1; -6 -1 2]

A =

$$\begin{bmatrix} 2 & -1 & 3 \\ 4 & 4 & 1 \\ -6 & -1 & 2 \end{bmatrix}$$

>> [L, U] = lu(A)

L =

$$\begin{matrix} -0.3333 & -0.4000 & 1.0000 \\ -0.6667 & 1.0000 & 0 \\ 1.0000 & 0 & 0 \end{matrix}$$

U =

$$\begin{matrix} -6.0000 & -1.0000 & 2.0000 \\ 0 & 3.3333 & 2.3333 \\ 0 & 0 & 4.6000 \end{matrix}$$

Διευκρίνηση: στην παραπάνω περίπτωση επιστρέφεται στο U ένα άνω τριγωνικό μητρώο και στο L ένα «ψυχολογικά κάτω τριγωνικό μητρώο», δηλαδή το γινόμενο ενός κάτω τριγωνικού μητρώου και ενός μητρώου αντιμετάθεσης (εναλλαγής).

>> A = [12 2 -1; 2 11 -2; 3 4 19]

A =

$$\begin{matrix} 12 & 2 & -1 \\ 2 & 11 & -2 \\ 3 & 4 & 19 \end{matrix}$$

>> [L, U] = lu(A)

L =

$$\begin{matrix} 1.0000 & 0 & 0 \\ 0.1667 & 1.0000 & 0 \end{matrix}$$

$$0.2500 \quad 0.3281 \quad 1.0000$$

U =

$$\begin{matrix} 12.0000 & 2.0000 & -1.0000 \\ 0 & 10.6667 & -1.8333 \\ 0 & 0 & 19.8516 \end{matrix}$$

Διευκρίνηση: στην παραπάνω περίπτωση επιστρέφεται στο U ένα άνω τριγωνικό μητρώο και στο L ένα κάτω τριγωνικό μητρώο.

Συμπέρασμα: όταν ελέγχουμε την ύπαρξη της LU παραγοντοποίησης ενός μητρώου, **το βασικό κριτήριο που χρησιμοποιούμε είναι η αντιστρεψιμότητα όλων των κύριων (κυρίαρχων) υπομητρώων του A.** Θα πρέπει όμως να έχουμε υπόψη ότι στην περίπτωση αυτή, επιστρέφεται στο L ένα «ψυχολογικά κάτω τριγωνικό μητρώο L» και όχι ένα καθαρό κάτω τριγωνικό μητρώο L. Το τελευταίο συμβαίνει **μόνο στην περίπτωση που το μητρώο A διαθέτει επιπλέον και A.Δ.Κ. κατά στήλες.** Επίσης, η κανονική LU χωρίς οδήγηση (δηλ. χρησιμοποιείται πάντα η τιμή στη διαγώνιο ως οδηγός) που ονομάζεται αλλιώς και απαλοιφή Gauss, **δεν είναι ευσταθής αλγόριθμος**, εκτός από μερικές ειδικές περιπτώσεις.

2.3.2 Οδήγηση στην LU παραγοντοποίηση

Η οδήγηση εφαρμόζεται ώστε να αποφευχθούν αστάθειες της απλής παραγοντοποίησης και να αυξηθεί η αξιοπιστία της λύσης. **Η οδήγηση στην LU παραγοντοποίηση δεν είναι απαραίτητη:** α) αν το μητρώο είναι κατά στήλες αυστηρά διαγώνια κυρίαρχο, β) αν το μητρώο είναι συμμετρικά θετικά ορισμένο (Σ.Θ.Ο.). Αν θελήσουμε να διαπιστώσουμε ποιο είναι το **κόστος της οδήγησης**, όταν χρησιμοποιείται στην LU παραγοντοποίηση, τότε μπορούμε να αναφέρουμε τα εξής:

- α) επιβάρυνση σε αριθμητικές πράξεις
- β) επιβάρυνση σε μετακινήσεις και πολυπλοκότητα κώδικα
- $O(n^2)$ πράξεις για μερική οδήγηση σε πυκνό μητρώο.
- $O(n^3)$ πράξεις για πλήρη οδήγηση σε πυκνό μητρώο.
- Επιπλέον μεταφορές και εντολές LOAD, STORE από και προς μνήμες λόγω αναζήτησης και εναλλαγών.
- γ) αλλαγή δομής μητρώου (π.χ. μπορεί ένα αραιό μητρώο να μετατραπεί σε πυκνό).

Στη συνέχεια θα αναφερθούμε σε δύο τεχνικές, οι οποίες πρέπει να λαμβάνονται υπόψη στην περίπτωση όταν κατά τη διαδικασία της απαλοιφής παρουσιάζονται σφάλματα στρογγυλοποίησης. Η πρώτη τεχνική ονομάζεται **μερική οδήγηση** (partial pivoting), ενώ η δεύτερη ονομάζεται **πλήρης ή ολική οδήγηση** (complete pivoting). **Οι δύο αυτές τεχνικές αντιμετωπίζουν τη συσσώρευση των σφαλμάτων στρογγυλοποίησης.** Σύμφωνα με τη **μερική οδήγηση**, απαλείφονται οι άγνωστοι με τη σειρά, αρχίζοντας με τον άγνωστο x_1 , αλλά σε κάθε

στάδιο της απαλοιφής επιλέγεται ως οδηγός εξίσωση εκείνη η εξίσωση που έχει το συντελεστή με τη μεγαλύτερη απόλυτη τιμή μεταξύ όλων των συντελεστών του αγνώστου που απαλείφεται. Η μέθοδος που επιλέγει σε κάθε στάδιο της απαλοιφής ως οδηγό στοιχείο τον απολύτως μεγαλύτερο συντελεστή σε **όλο** το μητρώο ονομάζεται **πλήρης οδήγηση**.

Στις άμεσες μεθόδους περιγράφουμε **μόνο την απαλοιφή Gauss με μερική οδήγηση**. Η απαλοιφή Gauss (η οποία ονομάζεται αλλιώς και LU παραγοντοποίηση) μπορεί να γίνει με/χωρίς οδήγηση. **Μερικά μειονεκτήματα οδήγησης** είναι τα εξής:

α) το επιπλέον κόστος λόγω συγκρίσεων και

β) η καταστροφή χρήσιμων δομικών χαρακτηριστικών του μητρώου, όπως τριγωνικότητα, τριδιαγωνιότητα κ.λ.π. Αν χρησιμοποιηθεί οδήγηση, αυτή μπορεί να είναι **μερική ή πλήρης**. **Πιο συνήθης μέθοδος οδήγησης είναι η πρώτη**, διότι:

α) αφενός είναι **σχεδόν πάντα πίσω ευσταθής**,

β) αφετέρου έχει μικρότερο πλήθος συγκρίσεων σε σχέση με την πλήρη οδήγηση και συγκεκριμένα το κόστος των συγκρίσεων είναι **$O(n^2)$** ενώ της πλήρους οδήγησης είναι **$O(n^3)$** .

Η **μερική οδήγηση είναι (σχεδόν πάντα) αριθμητικά ευσταθής και έχει μικρότερο κόστος από την πλήρη οδήγηση**. Υπενθυμίζεται ότι δοθέντων τετραγωνικού μητρώου και διανύσματος A και b αντίστοιχα, η επίλυση του γραμμικού συστήματος $A * x = b$ μπορεί να επιτευχθεί με την αναγωγή του επαυξημένου $C = [A | b]$ σε $[U | \hat{b}]$, όπου το U είναι άνω τριγωνικό και έχει προέλθει από πολλαπλασιασμούς του C από τα αριστερά με μητρώα εναλλαγής (επιλεγμένα να πραγματοποιούν μερική οδήγηση) και με στοιχειώδη μητρώα Gauss για το μηδενισμό των στοιχείων κάθε στήλης που βρίσκονται κάτω από την κύρια διαγώνιο. Επίσης το \hat{b} είναι το τροποποιημένο διάνυσμα b που προκύπτει από τις πράξεις που κάνουμε στο μητρώο A για να πάρουμε το U .

Στη συνέχεια επιτυγχάνεται η λύση με πίσω αντικατάσταση ή αν υπάρχουν πολλά δεξιά μέλη ή δεν ζητάται η επίλυση αλλά η παραγοντοποίηση LU, τότε υπολογίζονται οι παράγοντες P , L , U ώστε P μητρώο μετάθεσης, L κάτω τριγωνικό με **μονάδες στη διαγώνιο** και U άνω τριγωνικό τέτοια ώστε $P * A = L * U$. **Για το λόγο αυτό η LU με μερική οδήγηση ονομάζεται αλλιώς και PLU παραγοντοποίηση**. Σε κάθε περίπτωση, είναι σημαντικό να επιχειρείται η επίλυση με συστηματικό τρόπο, ειδάλλως είναι σχεδόν σίγουρο ότι θα προκύψουν σφάλματα που θα οφείλονται στην κακή οργάνωση των υπολογισμών.

Στη μεθοδολογία της **μερικής οδήγησης**, βρίσκουμε σε κάθε βήμα, το μεγαλύτερο κατ' απόλυτη τιμή στοιχείο κάθε στήλης και το βάζουμε –αν απαιτείται– στη θέση του οδηγού (δηλ. στην κύρια διαγώνιο). Για το σκοπό αυτό χρησιμοποιούμε το μητρώο εναλλαγής (αντιμετάθεσης) P , το οποίο προκύπτει από το ταυτοτικό μητρώο I με κατάλληλη εναλλαγή στηλών. Για παράδειγμα, αν το αρχικό μητρώο είναι 3×3 και έχει τη μορφή $P = I = [e_1, e_2, e_3]$, για να κάνουμε εναλλαγή 2^{nd} και 3^{rd} γραμμής του μητρώου A , θα γράψουμε $P * A = [e_1, e_3, e_2]$.

e₂] * A. Θα πρέπει να τονίσουμε το γεγονός ότι ένα μητρώο αντιμετάθεσης P έχει δύο ιδιότητες: α) είναι συμμετρικό, οπότε $P^T = P$ και β) είναι ορθογώνιο, οπότε $P^T = P^{-1}$. Αφού κάνουμε την εναλλαγή γραμμών, μηδενίζουμε στη συνέχεια τα στοιχεία κάτω από τον οδηγό κάθε στήλης, χρησιμοποιώντας το **μητρώο απαλοιφής Gauss**, το οποίο και πάλι προκύπτει από το ταυτοτικό μητρώο I και ανάλογα με τα στοιχεία που θέλουμε κάθε φορά να μηδενίσουμε, περιέχει στις κατάλληλες θέσεις το πηλίκο των στοιχείων αυτών με τον οδηγό της αντίστοιχης στήλης και με αντίθετο πρόσημο.

Σημείωση 1: Στην περίπτωση της LU παραγοντοποίησης πρώτα διασπάμε το μητρώο A σε μητρώα L και U και στη **συνέχεια** εφαρμόζουμε εμπρός αντικατάσταση $L * y = b$ και πίσω αντικατάσταση $U * x = y \Rightarrow x = U \setminus y = U \setminus (L \setminus b)$. Η LU παραγοντοποίηση εφαρμόζεται **μόνο** σε ένα μητρώο **nxn**, στο οποίο όλες οι κύριαρχες ορίζουσες είναι μη μηδενικές.

Σημείωση 2: Για την PLU παραγοντοποίηση, η εμπρός αντικατάσταση είναι της μορφής $L * y = P * b = \hat{b}$ και πίσω αντικατάσταση είναι της μορφής $U * x = y \Rightarrow x = U^{-1} * y = U \setminus y = U \setminus (L \setminus \hat{b})$. Στη MATLAB υπάρχουν δύο τελεστές αναφορικά με διαίρεση. Το slash (/) και το backslash (\). Για παράδειγμα το $4/2 = 2$ και το $4\backslash 2 = \frac{2}{4} = 0.5$. Όσον αφορά τα μητρώα, επειδή δεν γίνεται διαίρεση με μητρώο, ο τελεστής \ χρησιμοποιείται για το γρήγορο υπολογισμό ενός αντίστροφου μητρώου. Για παράδειγμα στην πράξη $L * y = b \Rightarrow y = L^{-1} * b$ κ.ο.κ. Επειδή όμως είναι δύσκολος και χρονοβόρος ο υπολογισμός του αντίστροφου μητρώου, θα πρέπει να αποφεύγεται και στον υπολογισμό του θα πρέπει να χρησιμοποιείται ο τελεστής \.

Σημείωση 3: Η **A.Δ.Κ.** ενός μητρώου (είτε κατά γραμμές είτε κατά στήλες) συνεπάγεται την **αντιστρεψιμότητά** του. Αν το μητρώο είναι συμμετρικό, δεν έχει σημασία αν μιλάμε για A.Δ.Κ. κατά γραμμές ή στήλες.

Σημείωση 4: Ένα μητρώο μπορεί να **μην** παραγοντοποιείται πάντα κατά LU (π.χ. μπορεί κάποιο κυρίαρχο υπομητρώο του να μην είναι αντιστρέψιμο), αλλά η **μερική οδήγηση** $P * A = L * U$ είναι **πάντα εφικτή**.

Σημείωση 5: Μέσω της LU παραγοντοποίησης ενός μητρώου A μπορούμε να υπολογίσουμε το **αντίστροφο** μητρώο A^{-1} . Πιο συγκεκριμένα, επιλύουμε γραμμικά συστήματα της μορφής $L * U * x_i = P * e_i$, όπου e_i είναι τα στοιχειώδη διανύσματα για $i = 1, 2, 3, \dots, n$. Κάθε x_i είναι και μια **στήλη** του αντίστροφου μητρώου A^{-1} . Επιλύοντας αυτήν την εξίσωση για $i = 1$, υπολογίζουμε το x_1 , για $i = 2$, υπολογίζουμε το x_2 κ.ο.κ.

Σημείωση 6: Όταν εκτελούμε την LU παραγοντοποίηση στη MATLAB, γράφουμε την εντολή: $[L, U] = lu(A)$ και τότε -υπό την προϋπόθεση ότι υπάρχει η παραγοντοποίηση αυτή- επιστρέφεται στο μητρώο L ο κάτω τριγωνικός παράγοντας και στο μητρώο U ο άνω τριγωνικός παράγοντας της LU. Υπάρχει όμως και η **εναλλακτική** περίπτωση πληκτρολόγησης της εντολής >> $lu(A)$, χωρίς να καταχωρηθεί κάπου το αποτέλεσμα, οπότε στην περίπτωση αυτή επιστρέφεται ως αποτέλεσμα **ένα μητρώο** που περιέχει πληροφορίες (στοιχεία) **τόσο** για το

μητρώο L όσο και για το μητρώο U . Το μητρώο αυτό έχει τη μορφή: $\begin{bmatrix} \eta_{11} & \eta_{12} \dots & \eta_{1n} \\ \lambda_2^{(1)} & \eta_{22} \dots & \eta_{2n} \\ \lambda_n^{(1)} & \lambda_n^{(2)} & \eta_{n,n} \end{bmatrix}$. Τα στοιχεία του μητρώου αυτού που συμβολίζονται με το «η» ανήκουν στο άνω τριγωνικό μητρώο U και τα στοιχεία του μητρώου αυτού που συμβολίζονται με το «λ» ανήκουν στο κάτω τριγωνικό μητρώο L . Ο λόγος για μια τέτοια παραγοντοποίηση είναι η **εξοικονόμηση μνήμης**, αφού αντί για δύο επιστρέφεται μόνο ένα μητρώο. Για παράδειγμα:

```
A =
12   2   4
  1  -7   2
  1   1   4
>> [L,U] = lu(A)
L =
1.0000   0   0
0.0833 1.0000   0
0.0833 -0.1163 1.0000
U =
12.0000  2.0000  4.0000
  0  -7.1667  1.6667
  0      0
3.8605
>> lu(A)
ans =
12.0000  2.0000  4.0000
0.0833 -7.1667  1.6667
0.0833 -0.1163  3.8605
```

Σημείωση 7: Αν γράψουμε όλη αυτή τη διαδικασία σε κώδικα MATLAB και ονομάσουμε B το μητρώο που επιστρέφεται από την LU παραγοντοποίηση, δηλαδή $B = lu(A)$, τότε θα έχουμε: $B = lu(A)$, $L = \text{tril}(B, -1) + \text{eye}(n)^1$, $U = \text{triu}(B)$. Στην περίπτωση που χρησιμοποιούμε **μερική οδήγηση**, θα έχουμε τον εξής κώδικα: $[L, U, P] = lu(A)$, $B = lu(A)$, $L_1 = \text{tril}(B, -1) + \text{eye}(n)$, $U_1 = \text{triu}(B)$. Το γινόμενο $L_1 * U_1 \approx P * A$.

2.3.3 Πρώτο παράδειγμα μερικής οδήγησης

Έστω το μητρώο $A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 6 \\ 1 & 3 & 0 \end{bmatrix}$ για το οποίο ζητάμε να υπολογίσουμε την **ορίζουσά** του με τρεις τρόπους: α) με ανάπτυγμα Laplace και β) μέσω LU παραγοντοποίησης με **μερική οδήγηση** και γ) μέσω **block LU διάσπασης**

Λύση

¹ Με τη βοήθεια της συνάρτησης $\text{eye}(n) = I$, αυτό που παίρνουμε είναι ουσιαστικά οι «1» της κύριας διαγωνίου, που προστίθενται στα στοιχεία «λ», για να δημιουργήσουν το κάτω τριγωνικό μητρώο L .

α) Με **ανάπτυγμα Laplace** και ως προς την 3^η γραμμή του μητρώου A έχουμε ότι: $\det(A) = |A| = (-1)^{(3+1)} * 1 * \begin{vmatrix} 2 & 3 \\ 3 & 6 \end{vmatrix} + (-1)^{(3+2)} * 3 * \begin{vmatrix} 1 & 3 \\ 2 & 6 \end{vmatrix} = (12 - 9) - 3(6 - 6) = 3$. Αν υπολογίζαμε την ορίζουσα ως προς οποιαδήποτε άλλη γραμμή ή στήλη, θα βρίσκαμε το ίδιο αποτέλεσμα. Για παράδειγμα, αν την υπολογίζαμε ως προς την 3^η στήλη, θα είχαμε: $\det(A) = |A| = (-1)^{(1+3)} * 3 * \begin{vmatrix} 2 & 3 \\ 1 & 3 \end{vmatrix} + (-1)^{(2+3)} * 6 * \begin{vmatrix} 1 & 2 \\ 1 & 3 \end{vmatrix} = 3 * (6 - 3) - 6 * (3 - 2) = 9 - 6 = 3$.

β) Με **μερική** οδήγηση, κοιτάμε σε κάθε στήλη το μεγαλύτερο κατ' απόλυτη τιμή στοιχείο (από τη θέση του οδηγού και κάτω) να βρίσκεται στη θέση του οδηγού (δηλαδή στην κύρια διαγώνιο), οπότε έχουμε ότι: $P_1 * A = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} * \begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 6 \\ 1 & 3 & 0 \end{bmatrix} = \begin{bmatrix} 2 & 3 & 6 \\ 1 & 2 & 3 \\ 1 & 3 & 0 \end{bmatrix}$ και $L_1 * P_1 * A = \begin{bmatrix} 1 & 0 & 0 \\ -1/2 & 1 & 0 \\ -1/2 & 0 & 1 \end{bmatrix} * \begin{bmatrix} 2 & 3 & 6 \\ 1 & 2 & 3 \\ 1 & 3 & 0 \end{bmatrix} = \begin{bmatrix} 2 & 3 & 6 \\ 0 & 1/2 & 0 \\ 0 & 3/2 & -3 \end{bmatrix}$. Στη συνέχεια,

κοιτάμε στο αποτέλεσμα του προηγούμενου βήματος να βρούμε το μέγιστο κατ' απόλυτη τιμή στοιχείο στη δεύτερη στήλη, από τη θέση του οδηγού κάτω, οπότε έχουμε ότι: $P_2 * L_1 * P_1 * A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} * \begin{bmatrix} 2 & 3 & 6 \\ 0 & 1/2 & 0 \\ 0 & 3/2 & -3 \end{bmatrix} = \begin{bmatrix} 2 & 3 & 6 \\ 0 & 3/2 & -3 \\ 0 & 0 & 1 \end{bmatrix}$ και στη συνέχεια χρησιμοποιούμε το μητρώο απαλοιφής Gauss L_2 για να μηδενίσουμε το στοιχείο $\frac{1}{2}$, οπότε έχουμε ότι το $L_2 * (P_2 * L_1 * P_1 * A) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -1/3 & 1 \end{bmatrix} * \begin{bmatrix} 2 & 3 & 6 \\ 0 & 3/2 & -3 \\ 0 & 1/2 & 0 \end{bmatrix} = \begin{bmatrix} 2 & 3 & 6 \\ 0 & 3/2 & -3 \\ 0 & 0 & 1 \end{bmatrix}$

= U. Άρα έχουμε ότι $L_2 * P_2 * L_1 * P_1 * A = U$ και στη συνέχεια για να μετατρέψουμε την εξίσωση αυτή σε μορφή μερικής οδήγησης $P * A = L * U$, παρεμβάλουμε το μητρώο αντιμετάθεσης P_2 ακριβώς αριστερά από το P_1 (αυτό το κάνουμε διότι πρέπει στη μορφή μερικής οδήγησης το $P = \dots * P_5 * P_4 * P_3 * P_2 * P_1$), αξιοποιώντας την ιδιότητα ότι οποιοδήποτε μητρώο αντιμετάθεσης είναι **օρθογώνιο**, οπότε $P_2^T * P_2 = I$. Έτσι θα έχουμε: $L_2 * P_2 * L_1 * P_2^T * P_2 * P_1 * A = U \Rightarrow P_2 * P_1 * A = (L_2 * P_2 * L_1 * P_2^T)^{-1} * U = (P_2^T)^{-1} * (L_1)^{-1} * (P_2)^{-1} * (L_2)^{-1} * U = (P_2^{-1})^{-1} * (L_1)^{-1} * (P_2)^{-1} * (L_2)^{-1} * U = P_2 * (L_1)^{-1} * P_2^T * (L_2)^{-1} * U = P_2 * (L_1)^{-1} * P_2 * (L_2)^{-1} * U = L * U$, οπότε το κάτω τριγωνικό μητρώο

L θίνεται: $L = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} * \begin{bmatrix} 1 & 0 & 0 \\ 1/2 & 1 & 0 \\ 1/2 & 0 & 1 \end{bmatrix} * \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} * \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1/3 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 1/2 & 1 & 0 \\ 1/2 & 1/3 & 1 \end{bmatrix}$. Το μητρώο $P = P_2 * P_1 =$

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} * \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}.$$

Σημείωση 1: Ισχύει ότι το **αντίστροφο** μητρώο ενός μητρώου **απαλοιφής Gauss**, όπως π.χ. το αντίστροφο του μητρώου L_1 ή το αντίστροφο του μητρώου L_2 , περιέχει το **αντίθετο πρόσημο** των στοιχείων απαλοιφής.

Επειδή $P * A = L * U \Rightarrow \det(P * A) = \det(L * U) \Rightarrow \det(P) * \det(A) = \det(L) * \det(U) \Rightarrow (-1)^2 * \det(A) = 1 * (2 * 3/2 * 1) \Rightarrow \det(A) = 3$. Παρατηρούμε ότι καταλήξαμε στο **ίδιο αποτέλεσμα** με αυτό του αναπτύγματος Laplace.

Σημείωση 2: Ισχύει ότι η ορίζουσα ενός μητρώου τριγωνικού ή διαγώνιου ισούται με το γινόμενο των στοιχείων της κύριας διαγωνίου, δηλαδή $\det(A) = |A| = \alpha_{11} * \alpha_{22} * \dots * \alpha_{nn}$. Επίσης, ισχύει ότι η **ορίζουσα** ενός μητρώου **αντιμετάθεσης** είναι **-1** και επειδή το μητρώο $P = P_2 * P_1$, ισχύει ότι $\det(P) = (-1) * (-1) = (-1)^2$.

γ) Αν το μητρώο A το γράψουμε σε **block μορφή** θα έχουμε: $A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$, όπου τα επιμέρους υπομητρώα θα είναι: $A_{11} = \begin{bmatrix} 1 & 2 \\ 2 & 3 \end{bmatrix}$, $A_{21} = [1 \ 3]$, $A_{12} = \begin{bmatrix} 3 \\ 6 \end{bmatrix}$ και $A_{22} = 0$. Διαιρούμε το αρχικό μητρώο σε υπομητρώα με τέτοιο τρόπο **έτσι ώστε το block μητρώο που θα προκύψει να είναι άνω ή κάτω τριγωνικό**. Ισχύει ότι αν εφαρμόσουμε **block LU διάσπαση** σε ένα block μητρώο, τότε αυτό θα το διασπάσουμε σε δύο μητρώα L_B και U_B που θα είναι της εξής μορφής: $A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} = L_B * U_B = \begin{bmatrix} I & 0 \\ L_{21} & I \end{bmatrix} * \begin{bmatrix} U_{11} & U_{12} \\ 0 & U_{22} \end{bmatrix} \Rightarrow \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} = \begin{bmatrix} U_{11} & U_{12} \\ L_{21}U_{11} & L_{21}U_{12} + U_{22} \end{bmatrix}$, όπου το S = συμπλήρωμα Schur $= L_{21}U_{12} + U_{22}$. Εξισώνοντας τα αντίστοιχα στοιχεία των δύο μητρώων μεταξύ τους, έχουμε ότι: $A_{11} = U_{11} \Rightarrow U_{11} = A_{11}$, $A_{12} = U_{12} \Rightarrow U_{12} = A_{12}$, $A_{21} = L_{21}U_{11} \Rightarrow L_{21}U_{11} = A_{21} \Rightarrow L_{21} = A_{21} * U_{11}^{-1} = A_{21} * A_{11}^{-1}$, $A_{22} = L_{21}U_{12} + U_{22} \Rightarrow U_{22} = A_{22} - L_{21}U_{12} = A_{22} - L_{21}A_{12} \Rightarrow U_{22} = A_{22} - A_{21}A_{11}^{-1}A_{12}$. Επομένως: $A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} = L_B * U_B = \begin{bmatrix} I & 0 \\ A_{21} * A_{11}^{-1} & I \end{bmatrix} * \begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} - A_{21} * A_{11}^{-1} * A_{12} \end{bmatrix}$, όπου ο όρος $S = A_{22} - A_{21} * A_{11}^{-1} * A_{12}$ ονομάζεται **συμπλήρωμα του Schur ως προς A_{11}** . Αν ένα μητρώο A διασπαστεί σε επιμέρους μητρώα, δηλ. $A = L_B * U_B$, τότε $\det(A) = \det(L_B * U_B) \Rightarrow \det(A) = \det(L_B) * \det(U_B)$. Επίσης ισχύει ότι **$\det(I) = 1$** , όπου I = ταυτοτικό μητρώο και επίσης –όπως προαναφέραμε- ισχύει ότι αν ένα μητρώο είναι διαγώνιο ή τριγωνικό (άνω ή κάτω), τότε **η ορίζουσά του θα είναι ίση με το γινόμενο των στοιχείων της κύριας διαγωνίου ή αν είναι σε block μορφή, τότε θα ισούται με το γινόμενο των οριζουσών των μητρώων της κύριας διαγωνίου του block μητρώου**. Δηλαδή αν έχουμε ένα block μητρώο που είναι άνω ή κάτω τριγωνικό, της μορφής $\begin{bmatrix} A & B \\ 0 & C \end{bmatrix}$ ή της μορφής $\begin{bmatrix} A & 0 \\ B & C \end{bmatrix}$, τότε ισχύει ότι **η ορίζουσα του θα είναι και στις δύο περιπτώσεις το γινόμενο των οριζουσών των υπομητρώων της κύριας διαγωνίου, δηλαδή $\det(A) * \det(C)$** . Για το λόγο αυτό, επειδή τα μητρώα L_B και U_B είναι κάτω και άνω τριγωνικά μητρώα αντίστοιχα, τότε $\det(L_B) = \det(I) * \det(I) = 1 * 1 = 1$ και το $\det(U_B) = \det(A_{11}) * \det(S) = \det(A_{11}) * \det(A_{22} - A_{21} * A_{11}^{-1} * A_{12}) = (-1) * \det(A_{22} - A_{21} * A_{11}^{-1} * A_{12}) = (-1) * \det(0 - [1 \ 3] * \begin{bmatrix} 1 & 2 \\ 2 & 3 \end{bmatrix}^{-1} * \begin{bmatrix} 3 \\ 6 \end{bmatrix}) = (-1) * \det(0 - [1 \ 3] * \begin{bmatrix} -3 & 2 \\ 2 & -1 \end{bmatrix} * \begin{bmatrix} 3 \\ 6 \end{bmatrix}) = (-1) * \det(0 - 3) = (-1) * \det(-3) = (-1) * (-3) = 3$. Άρα το $\det(A) = \det(L_B) * \det(U_B) = 3$. Παρατηρούμε ότι καταλήξαμε στο **ίδιο ακριβώς αποτέλεσμα** με τους δύο προηγούμενους τρόπους.

Σημείωση: επειδή το μητρώο A_{11} είναι 2×2 , το **αντίστροφό** του μπορεί να υπολογιστεί **απευθείας** από τον τύπο: $A^{-1} = \frac{1}{\det(A)} * \begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix}$.

2.3.4 Δεύτερο παράδειγμα μερικής οδήγησης

Έστω το μητρώο $A = \begin{bmatrix} 1 & 2 & 2 \\ 0 & 5 & 1 \\ 3 & 4 & 3 \end{bmatrix}$ και το μητρώο $A^{-1} = \frac{1}{13} \begin{bmatrix} -11 & -2 & 8 \\ -3 & 3 & 1 \\ 15 & -2 & -5 \end{bmatrix}$. Να υπολογίσετε τους όρους P, L, U της παραγοντοποίησης $PA = LU$ (χρησιμοποιώντας μερική οδήγηση).

Λύση

Με μερική οδήγηση, κοιτάμε σε κάθε στήλη το μεγαλύτερο κατ' απόλυτη τιμή στοιχείο (από τη θέση του οδηγού και κάτω) να βρίσκεται στη θέση του οδηγού (δηλαδή στην κύρια διαγώνιο), οπότε έχουμε ότι: $P_1 * A =$

$$\begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} * \begin{bmatrix} 1 & 2 & 2 \\ 0 & 5 & 1 \\ 3 & 4 & 3 \end{bmatrix} = \begin{bmatrix} 3 & 4 & 3 \\ 0 & 5 & 1 \\ \textcolor{red}{1} & 2 & 2 \end{bmatrix} \text{ και } L_1 * P_1 * A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -1/3 & 0 & 1 \end{bmatrix} * \begin{bmatrix} 3 & 4 & 3 \\ 0 & 5 & 1 \\ 1 & 2 & 2 \end{bmatrix} = \begin{bmatrix} 3 & 4 & 3 \\ 0 & 5 & 1 \\ 0 & 2/3 & 1 \end{bmatrix}. \text{ Στη συνέχεια}$$

$$\text{έχουμε: } L_2 * L_1 * P_1 * A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -2/15 & 1 \end{bmatrix} * \begin{bmatrix} 3 & 4 & 3 \\ 0 & 5 & 1 \\ 0 & 2/3 & 1 \end{bmatrix} = \begin{bmatrix} 3 & 4 & 3 \\ 0 & 5 & 1 \\ 0 & 0 & 13/15 \end{bmatrix} = U. \text{ Στη συνέχεια έχουμε ότι: } L_2 * L_1 *$$

$$P_1 * A = U \Rightarrow P_1 * A = (L_2 * L_1)^{-1} * U = (L_1)^{-1} * (L_2)^{-1} * U = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \textcolor{red}{1/3} & 0 & 1 \end{bmatrix} * \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 2/15 & 1 \end{bmatrix} * U = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \textcolor{red}{1/3} & \textcolor{green}{2/15} & 1 \end{bmatrix} * U.$$

$$\text{Άρα το μητρώο } L = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1/3 & 2/15 & 1 \end{bmatrix} \text{ και το μητρώο } P = P_1 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}.$$

2.3.5 Τρίτο παράδειγμα μερικής οδήγησης

Έστω το μητρώο $A = \begin{bmatrix} 10 & -7 & 0 \\ -3 & 2 & 6 \\ 5 & -1 & 5 \end{bmatrix}$. Να εφαρμόσετε την LU παραγοντοποίηση με μερική οδήγηση. Εναλλακτικά θα μπορούσε να αναφέρει το εξής: να κάνετε PLU παραγοντοποίηση.

Λύση

Στην 1^η στήλη έχουμε το μέγιστο κατ' απόλυτη τιμή στοιχείο στη θέση του οδηγού, οπότε δεν χρειάζεται εναλλαγή γραμμών. Άρα απαλείφουμε τα στοιχεία της πρώτης στήλης κάτω από τον οδηγό: $L_1 * A =$

$$\begin{bmatrix} 1 & 0 & 0 \\ \frac{3}{10} & 1 & 0 \\ -\frac{5}{10} & 0 & 1 \end{bmatrix} \begin{bmatrix} 10 & -7 & 0 \\ -3 & 2 & 6 \\ 5 & -1 & 5 \end{bmatrix} = \begin{bmatrix} 10 & -7 & 0 \\ 0 & -1/\textcolor{red}{10} & 6 \\ 0 & 5/2 & 5 \end{bmatrix}. \text{ Πρέπει να εναλλάξουμε 2^η γραμμή με 3^η γραμμή με το μητρώο}$$

$$P_1 = [e_1 \ e_3 \ e_2] = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}, \text{ οπότε } P_1 * L_1 * A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} * \begin{bmatrix} 10 & -7 & 0 \\ 0 & -\frac{1}{10} & 6 \\ 0 & \frac{5}{2} & 5 \end{bmatrix} = \begin{bmatrix} 10 & -7 & 0 \\ 0 & 5/2 & 5 \\ 0 & -1/10 & 6 \end{bmatrix} \text{ και } L_2 * P_1 *$$

$$L_1 * A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & \frac{1}{25} & 1 \end{bmatrix} \begin{bmatrix} 10 & -7 & 0 \\ 0 & \frac{5}{2} & 5 \\ 0 & -\frac{1}{10} & 6 \end{bmatrix} = \begin{bmatrix} 10 & -7 & 0 \\ 0 & \frac{5}{2} & 5 \\ 0 & 0 & \frac{31}{5} \end{bmatrix} = U. \text{ Επειδή } L_2 * P_1 * L_1 * A = U \Rightarrow L_2 * P_1 * L_1 * \textcolor{red}{P_1^T} * \textcolor{red}{P_1} *$$

$$A \Rightarrow P_1 * A = (L_2 * P_1 * L_1 * \textcolor{blue}{P_1^T})^{-1} * U \Rightarrow P_1 * A = \textcolor{blue}{P_1} * \textcolor{blue}{L_1^{-1}} * \textcolor{blue}{P_1^T} * \textcolor{blue}{L_2^{-1}} * U \Rightarrow \textcolor{blue}{L} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} * \begin{bmatrix} 1 & 0 & 0 \\ -\frac{3}{10} & 1 & 0 \\ \frac{5}{10} & 0 & 1 \end{bmatrix} *$$

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} * \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -\frac{1}{25} & 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{5} & 0 & 0 \\ \frac{1}{10} & 0 & 1 \\ -\frac{3}{10} & 1 & 0 \end{bmatrix} * \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} * \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -\frac{1}{25} & 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{10} & 0 & 0 \\ \frac{1}{10} & 1 & 0 \\ -\frac{3}{10} & 0 & 1 \end{bmatrix} * \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -\frac{1}{25} & 1 \end{bmatrix}$$

$$\begin{bmatrix} \frac{1}{10} & 0 & 0 \\ \frac{1}{10} & 1 & 0 \\ -\frac{3}{10} & 0 & 1 \end{bmatrix}.$$

2.3.6 Τέταρτο παράδειγμα μερικής οδήγησης

Να υπολογιστούν οι παράγοντες L , U του μητρώου $A = \begin{bmatrix} 1 & 1 + 0.5 * 10^{-15} & 3 \\ 2 & 2 & 20 \\ 3 & 6 & 4 \end{bmatrix}$ και να επαληθευτεί ότι η LU παραγοντοποίηση δεν είναι ακριβής.

Λύση

$$L_1 * A = \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ -3 & 0 & 1 \end{bmatrix} * \begin{bmatrix} 1 & 1 + 0.5 * 10^{-15} & 3 \\ 2 & 2 & 20 \\ 3 & 6 & 4 \end{bmatrix} = \begin{bmatrix} 1 & 1 + 0.5 * 10^{-15} & 3 \\ 0 & -1 * 10^{-15} & 14 \\ 0 & 3 - 1.5 * 10^{-15} & -5 \end{bmatrix}. \quad \text{Έπειτα} \quad L_2 * L_1 * A =$$

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & \frac{3-1.5*10^{-15}}{1*10^{-15}} & 1 \end{bmatrix} * \begin{bmatrix} 1 & 1 + 0.5 * 10^{-15} & 3 \\ 0 & -1 * 10^{-15} & 14 \\ 0 & 3 - 1.5 * 10^{-15} & -5 \end{bmatrix} = \begin{bmatrix} 1 & 1 + 0.5 * 10^{-15} & 3 \\ 0 & -10^{-15} & 14 \\ 0 & 0 & 14 * \left(\frac{3-1.5*10^{-15}}{1*10^{-15}}\right) - 5 \end{bmatrix} = U \quad \text{Άρα} \quad L_2 *$$

$$L_1 * A = U \Rightarrow A = (L_2 * L_1)^{-1} * U \Rightarrow A = L_1^{-1} * L_2^{-1} * U = L * U, \quad \text{όπου} \quad L = L_1^{-1} * L_2^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 3 & 0 & 1 \end{bmatrix} *$$

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -\frac{3-1.5*10^{-15}}{1*10^{-15}} & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 3 & -\frac{3-1.5*10^{-15}}{1*10^{-15}} & 1 \end{bmatrix} \text{ και αν η LU παραγοντοποίηση είναι ακριβής, πρέπει } A - LU = 0.$$

$$\text{Το γινόμενο} \quad L * U = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 3 & -\frac{3-1.5*10^{-15}}{1*10^{-15}} & 1 \end{bmatrix} * \begin{bmatrix} 1 & 1 + 0.5 * 10^{-15} & 3 \\ 0 & -10^{-15} & 14 \\ 0 & 0 & 14 * \left(\frac{3-1.5*10^{-15}}{1*10^{-15}}\right) - 5 \end{bmatrix} =$$

$$\begin{bmatrix} 1 & 1 + 0.5 * 10^{-15} & 3 \\ 2 & 2 & 20 \\ 3 & 6 & 4 \end{bmatrix} = A. \text{ Αν όμως τρέξουμε τον κώδικα της LU παραγοντοποίησης για το συγκεκριμένο μητρώο } A \text{ σε H/Y τότε λόγω της πεπερασμένης ακρίβειας θα διαπιστώσουμε ότι } A \neq LU.$$

2.3.7 Πέμπτο παράδειγμα μερικής οδήγησης – Θέμα εξετάσεων

Στην εντολή $[L, U, P] = lu(A)$ όπου $A = \begin{pmatrix} 0 & 0 & 0 & -1 \\ -2 & 0 & -2 & 0 \\ 3 & 0 & 2 & -2 \\ 3 & 3 & 0 & -2 \end{pmatrix}$, θα είναι ο παράγοντας P ίσος με το ταυτοτικό μητρώο; Να απαντήσετε Ναι ή Όχι και να τον υπολογίσετε αν δεν είναι. Σε κάθε περίπτωση, να υπολογίσετε και τους παράγοντες L , U .

Λύση

Η MATLAB βασίζεται σε αριθμητικές βιβλιοθήκες που για την επίλυση τετραγωνικών συστημάτων χρησιμοποιούν μερική οδήγηση. **Επομένως, από το «0» στη θέση (1, 1) του A, μπορούμε αμέσως να συμπεράνουμε ότι το P δεν θα είναι το ταυτοτικό μητρώο και αυτό διότι θα πρέπει από το πρώτο βήμα να γίνει εναλλαγή γραμμών.**

Το σημαντικό εδώ όμως είναι να υπολογίσουμε τους παράγοντες. Στη γενική περίπτωση ισχύει ότι τα στοιχειώδη μητρώα από τα οποία θα προκύψουν τα μητρώα U, P υπολογίζονται άμεσα ως εξής: $L_3 * P_3 * L_2 * P_2 * L_1 * P_1 * A = U$, όπου $P = P_3 * P_2 * P_1$. Το τελικό L χρειάζεται λίγη περισσότερη δουλειά. Όπως θα δούμε, για τα συγκεκριμένα δεδομένα, ορισμένα από τα στοιχειώδη μητρώα ταυτοτικά και η εύρεση του L να απλοποιείται. Στην περίπτωσή μας, η μερική οδήγηση καθορίζει την επιλογή των P ώστε στο βήμα $k = 1, 2, 3$ η εναλλαγή να φέρνει το μέγιστο στοιχείο στη διαγώνια θέση (k, k) . Αξίζει να σημειωθεί ότι αν οι πράξεις είναι τότε στο τέλος, όλα τα στοιχεία του L θα είναι φραγμένα σε απόλυτη τιμή από το 1. Στην περίπτωσή μας επιλέγουμε: $P_1 = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$

$$\Rightarrow P_1 * A = \begin{pmatrix} 3 & 0 & 2 & -2 \\ -2 & 0 & -2 & 0 \\ 0 & 0 & 0 & -1 \\ 3 & 3 & 0 & -2 \end{pmatrix} \Rightarrow L_1 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ \frac{2}{3} & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -1 & 0 & 0 & 1 \end{pmatrix} \Rightarrow A^{(1)} = L_1 P_1 A = \begin{pmatrix} 3 & 0 & 2 & -2 \\ 0 & \textcolor{red}{0} & -\frac{2}{3} & -\frac{4}{3} \\ 0 & 0 & 0 & -1 \\ 0 & 3 & -2 & 0 \end{pmatrix}$$

Παρατηρούμε ότι έχουμε σχεδόν τελειώσει λόγω των πολλών μηδενικών που έχουν δημιουργηθεί (και που οφείλονται στη δομή του αρχικού μητρώου).

$$\text{Επιλέγουμε τώρα } P_2 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix} \Rightarrow P_2 * A^{(1)} = \begin{pmatrix} 3 & 0 & 2 & -2 \\ 0 & 3 & -2 & 0 \\ 0 & 0 & \textcolor{red}{0} & -1 \\ 0 & 0 & -\frac{2}{3} & -\frac{4}{3} \end{pmatrix}, \text{ επομένως } L_2 = I. \text{ Τέλος } P_3 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}.$$

$$\text{Επομένως το γινόμενο } P_3 * (P_2 * L_1 * P_1 * A) = U = \begin{pmatrix} 3 & 0 & 2 & -2 \\ 0 & 3 & -2 & 0 \\ -0 & 0 & -\frac{2}{3} & -\frac{4}{3} \\ 0 & 0 & 0 & -1 \end{pmatrix}. \text{ Τέλος υπολογίζουμε το κάτω τριγωνικό μη-$$

τρώο L ως εξής: Από τις ιδιότητες των μητρώων L_j, P_j μπορούμε να γράψουμε: $P_3 * P_2 * L_1 * P_1 * A = U \Rightarrow P_3 * P_2 * L_1 * P_2^T * P_3^T * P_3 * P_2 * P_1 * A = U \Rightarrow [P_3 * P_2 * L_1 * P_2^T * P_3^T] * P_3 * P_2 * P_1 * A = [P_3 * P_2 * L_1 * P_2^T * P_3^T]^{-1} * U \Rightarrow P * A = L * U$, όπου το μητρώο L = $(P_3 * P_2 * L_1 * P_2 * P_3)^{-1} = P_3^{-1} * P_2^{-1} * L_1^{-1} * P_2^{-1} * P_3^{-1} = (P_3^T)^{-1} * (P_2^T)^{-1} * L_1^{-1} * (P_2^T)^{-1} * (P_3^T)^{-1} = P_3 * P_2 * L_1^{-1} * P_2 * P_3$.

$$\text{Επομένως το } L_1^{-1} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ -\frac{2}{3} & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix} \text{ και μετά από πράξεις το τελικό μητρώο } L = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ -\frac{2}{3} & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \text{ Το}$$

$$\text{μητρώο εναλλαγής } P = P_3 * P_2 * P_1 = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}.$$

2.4. LD \hat{U} Παραγοντοποίηση

Έστω το μητρώο $A = \begin{bmatrix} 2 & 4 & 8 \\ 0 & 3 & 9 \\ 0 & 0 & 7 \end{bmatrix}$ και ζητάμε για αυτό να υπολογίσουμε τόσο την LU όσο και την LD \hat{U} παραγοντοποίησή του.

Λύση

Όσον αφορά την LU παραγοντοποίηση, έχουμε να παρατηρήσουμε ότι το μητρώο A έχει ήδη δοθεί σε LU μορφή, αφού το αρχικό μητρώο A είναι ήδη σε άνω τριγωνική μορφή. Αν δεν ήταν τότε θα έπρεπε να κάνουμε LU παραγοντοποίηση με μερική οδήγηση. Επομένως το μητρώο L = $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ και το μητρώο U = $\begin{bmatrix} 2 & 4 & 8 \\ 0 & 3 & 9 \\ 0 & 0 & 7 \end{bmatrix}$.

Όσον αφορά την $LD\widehat{U}$ παραγοντοποίηση, που αποτελεί μια παραλλαγή της LU, έχουμε να παρατηρήσουμε ότι στην περίπτωση αυτή το διαγώνιο μητρώο D περιέχει στην κύρια διαγώνιο του τους οδηγούς του A, δηλ. τα

στοιχεία της κύριας διαγωνίου του μητρώου U, δηλ. $D = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 7 \end{bmatrix}$. Επίσης το μητρώο \widehat{U} περιέχει στην κύρια διαγώνιο του "1", ενώ για τα υπόλοιπα στοιχεία του, λύνουμε το σύστημα: $D * \widehat{U} = U \Rightarrow \begin{bmatrix} 2 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 7 \end{bmatrix} * \begin{bmatrix} 1 & x_1 & x_2 \\ 0 & 1 & x_3 \\ 0 & 0 & 1 \end{bmatrix}$

$= \begin{bmatrix} 2 & 4 & 8 \\ 0 & 3 & 9 \\ 0 & 0 & 7 \end{bmatrix} \Rightarrow \begin{bmatrix} 2 & 2 * x_1 & 2 * x_2 \\ 0 & 3 & 3 * x_3 \\ 0 & 0 & 7 \end{bmatrix} = \begin{bmatrix} 2 & 4 & 8 \\ 0 & 3 & 9 \\ 0 & 0 & 7 \end{bmatrix}$. Εξισώνοντας τα αντίστοιχα των δύο μητρώων μεταξύ τους, έχουμε ότι: $2 * x_1 = 4 \Rightarrow x_1 = 2$ και $2 * x_2 = 8 \Rightarrow x_2 = 4$ και τέλος $3 * x_3 = 9 \Rightarrow x_3 = 3$. Άρα το μητρώο $\widehat{U} = \begin{bmatrix} 1 & 2 & 4 \\ 0 & 1 & 3 \\ 0 & 0 & 1 \end{bmatrix}$.

Επαλήθευση: $D * \widehat{U} = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 7 \end{bmatrix} * \begin{bmatrix} 1 & 2 & 4 \\ 0 & 1 & 3 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 4 & 8 \\ 0 & 3 & 9 \\ 0 & 0 & 7 \end{bmatrix} = U$. Διευκρινίζεται ότι αν το αρχικό μητρώο A είχε δοθεί σε τυχαία μορφή, θα έπρεπε να παραγοντοποιηθεί πρώτα σε L και U, έτσι ώστε να πάρουμε τα στοιχεία της κύριας διαγωνίου του U και να τα τοποθετήσουμε στην κύρια διαγώνιο του D.

2.4.1 Άσκηση με μερική οδήγηση και $LD\widehat{U}$ παραγοντοποίηση

α) Έστω το μητρώο $A = \begin{bmatrix} 0 & 1 & 2 & 3 \\ 3 & 0 & 1 & 2 \\ 2 & 3 & 0 & 1 \\ 1 & 2 & 3 & 0 \end{bmatrix}$. Ζητάμε να υπολογίσουμε τους παράγοντες P, L, U.

β) Ζητάμε διαγώνιο μητρώο D τέτοιο ώστε το $PA = LD\widehat{U}$.

Λύση

Επειδή το μητρώο A είναι **κυκλοτερές** (κάθε γραμμή προκύπτει από την προηγούμενη της με δεξιά κυκλική ολίσθηση), μπορούμε να χρησιμοποιήσουμε **μόνο ένα μητρώο εναλλαγής P**, έτσι ώστε να φέρουμε στην κύρια διαγώνιο **το μέγιστο στοιχείο κάθε στήλης** και να μην χρειαστεί να γίνει άλλη εναλλαγή γραμμών. Ένα τέτοιο μητρώο P είναι το εξής: $P = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix}$ και $P * A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 & 2 & 3 \\ 3 & 0 & 1 & 2 \\ 2 & 3 & 0 & 1 \\ 1 & 2 & 3 & 0 \end{bmatrix} = \begin{bmatrix} 3 & 0 & 1 & 2 \\ 2 & 3 & 0 & 1 \\ 1 & 2 & 3 & 0 \\ 0 & 1 & 2 & 3 \end{bmatrix}$. Στη συνέχεια

εφαρμόζουμε μερική οδήγηση για να μετατρέψουμε το μητρώο P * A \rightarrow U. Συγκεκριμένα, χρησιμοποιούμε το

μητρώο απαλοιφής Gauss L_1 , που είναι: $L_1 * P * A = \begin{bmatrix} 1 & 0 & 0 & 0 \\ -2/3 & 1 & 0 & 0 \\ -1/3 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 3 & 0 & 1 & 2 \\ 2 & 3 & 0 & 1 \\ 1 & 2 & 3 & 0 \\ 0 & 1 & 2 & 3 \end{bmatrix} = \begin{bmatrix} 3 & 0 & 1 & 2 \\ 0 & 3 & -2/3 & -1/3 \\ 0 & 2 & 8/3 & -2/3 \\ 0 & 1 & 2 & 3 \end{bmatrix}$. Στη

συνέχεια χρησιμοποιούμε το μητρώο απαλοιφής Gauss L_2 , που είναι: $L_2 * L_1 * P * A =$

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & -2/3 & 1 & 0 \\ 0 & -1/3 & 0 & 1 \end{bmatrix} \begin{bmatrix} 3 & 0 & 1 & 2 \\ 0 & 3 & -2/3 & -1/3 \\ 0 & 2 & 8/3 & -2/3 \\ 0 & 1 & 2 & 3 \end{bmatrix} = \begin{bmatrix} 3 & 0 & 1 & 2 \\ 0 & 3 & -2/3 & -1/3 \\ 0 & 0 & 28/9 & -4/9 \\ 0 & 0 & 20/9 & 28/9 \end{bmatrix}. \text{ Παρατηρούμε ότι σε κάθε βήμα έχουμε συνεχώς το}$$

μέγιστο στοιχείο κάθε στήλης στην κύρια διαγώνιο. Τέλος το μητρώο L_3 είναι: $L_3 * L_2 * L_1 * P * A =$

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & -20/28 & 1 \end{bmatrix} \begin{bmatrix} 3 & 0 & 1 & 2 \\ 0 & 3 & -2/3 & -1/3 \\ 0 & 0 & 28/9 & -4/9 \\ 0 & 0 & 20/9 & 28/9 \end{bmatrix} = \begin{bmatrix} 3 & 0 & 1 & 2 \\ 0 & 3 & -2/3 & -1/3 \\ 0 & 0 & 28/9 & -4/9 \\ 0 & 0 & \frac{80}{28*9} + \frac{28}{9} & 28/9 \end{bmatrix} = \begin{bmatrix} 3 & 0 & 1 & 2 \\ 0 & 3 & -2/3 & -1/3 \\ 0 & 0 & 28/9 & -4/9 \\ 0 & 0 & 0 & 3.428 \end{bmatrix} = U. \text{ Άρα ισχύει ότι } L_3 * L_2 *$$

$$L_1 * P * A = U \Rightarrow P * A = (L_3 * L_2 * L_1)^{-1} * U \Rightarrow P * A = L_1^{-1} * L_2^{-1} * L_3^{-1} * U \text{ όπου } L = L_1^{-1} * L_2^{-1} * L_3^{-1} \Rightarrow L =$$

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 2/3 & 1 & 0 & 0 \\ 1/3 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} * \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 2/3 & 1 & 0 \\ 0 & 1/3 & 0 & 1 \end{bmatrix} * \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 20/28 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 2/3 & 1 & 0 & 0 \\ 1/3 & 2/3 & 1 & 0 \\ 0 & 1/3 & 20/28 & 1 \end{bmatrix} = L.$$

Παρατήρηση: Για αποφυγή πράξεων στον υπολογισμό του L , μπορούμε απλώς να τοποθετήσουμε όλα τα στοιχεία απαλοιφής από τα μητρώα $L_1^{-1}, L_2^{-1}, L_3^{-1}$ στις αντίστοιχες θέσεις στο τελικό μητρώο L χωρίς πράξεις. Το μητρώο L που προέκυψε, περιέχει "1" στην κύρια διαγώνιο, όπως είναι το ζητούμενο.

β) Ζητάμε διαγώνιο μητρώο D τέτοιο ώστε το $P * A = L * D * \hat{U}$. Το μητρώο **D θα περιέχει στην κύρια διαγώνιο**

$$\text{του τα στοιχεία της κύριας διαγωνίου του μητρώου } U, \text{ δηλαδή θα είναι της μορφής } D = \begin{bmatrix} 3 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 \\ 0 & 0 & 28/9 & 0 \\ 0 & 0 & 0 & 3.428 \end{bmatrix}.$$

Για να έχει το \hat{U} "1" στην κύρια διαγώνιο, θα πρέπει να πολλαπλασιάσουμε το μητρώο U με τα αντίστροφα

$$\text{στοιχεία του } D. \text{ Δηλαδή θα έχουμε: } \begin{bmatrix} 3 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 \\ 0 & 0 & 28/9 & 0 \\ 0 & 0 & 0 & 3.428 \end{bmatrix} * \hat{U} = U \Rightarrow \hat{U} = D^{-1} * U = \begin{bmatrix} 1/3 & 0 & 0 & 0 \\ 0 & 1/3 & 0 & 0 \\ 0 & 0 & 9/28 & 0 \\ 0 & 0 & 0 & 1/3.428 \end{bmatrix} *$$

$$\begin{bmatrix} 3 & 0 & 1 & 2 \\ 0 & 3 & -2/3 & -1/3 \\ 0 & 0 & 28/9 & -4/9 \\ 0 & 0 & 0 & 3.428 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 1/3 & 2/3 \\ 0 & 1 & -2/9 & -1/9 \\ 0 & 0 & 1 & -36/252 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \text{ Εναλλακτικά για τον υπολογισμό του μητρώου } \hat{U} \text{ έχουμε ότι}$$

$$D * \hat{U} = U \Rightarrow \begin{bmatrix} 3 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 \\ 0 & 0 & 28/9 & 0 \\ 0 & 0 & 0 & \frac{80}{28*9} + \frac{28}{9} \end{bmatrix} * \begin{bmatrix} 1 & x_1 & x_2 & x_3 \\ 0 & 1 & x_4 & x_5 \\ 0 & 0 & 1 & x_6 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 3 & 0 & 1 & 2 \\ 0 & 3 & -2/3 & -1/3 \\ 0 & 0 & 28/9 & -4/9 \\ 0 & 0 & 0 & 3.428 \end{bmatrix} \Rightarrow \begin{bmatrix} 3 & 3x_1 & 3x_2 & 3x_3 \\ 0 & 3 & 3x_4 & 3x_5 \\ 0 & 0 & 28/9 & 28/9x_6 \\ 0 & 0 & 0 & \frac{80}{28*9} + \frac{28}{9} \end{bmatrix} =$$

$$\begin{bmatrix} 3 & 0 & 1 & 2 \\ 0 & 3 & -2/3 & -1/3 \\ 0 & 0 & 28/9 & -4/9 \\ 0 & 0 & 0 & 3.428 \end{bmatrix}. \text{ Από την εξίσωση των αντίστοιχων στοιχείων των δύο μητρώων έχουμε ότι } 3x_1 = 0 \Rightarrow x_1 = 0, 3x_2 = 1 \Rightarrow x_2 = 1/3, 3x_3 = 2 \Rightarrow x_3 = \frac{2}{3} \kappa. o. \kappa.$$

Παρατήρηση: Από τη στιγμή που έχουμε ένα μητρώο διαγώνιο -όπως το D - που έχει όλα τα στοιχεία της κύριας διαγωνίου μη-μηδενικά, σημαίνει ότι αντιστρέφεται και το αντίστροφο μητρώο υπολογίζεται απευθείας αντιστρέφοντας τα στοιχεία της κύριας διαγωνίου του D .

2.5. Παραγοντοποίηση Cholesky

Η παραγοντοποίηση Cholesky είναι η πιο **οικονομική** παραγοντοποίηση για την επίλυση του γραμμικού συστήματος $A * x = b$, διότι απαιτεί για την εκτέλεσή της μόνο $\Omega = \frac{n^3}{3} + 2n^2$ πράξεις, ενώ η LU παραγοντοποίηση απαιτεί $\Omega = \frac{2 * n^3}{3} + 2n^2$ πράξεις. Στην περίπτωση της παραγοντοποίησης Cholesky, το μητρώο των συντελεστών Α διασπάται σε δύο μητρώα L (κάτω τριγωνικό) και L^T (άνω τριγωνικό), δηλαδή $A = L * L^T$. Στη συνέχεια εφαρμόζονται τα βήματα της εμπρός ($L * y = b$) και της πίσω αντικατάστασης ($L^T * x = y$) για να υπολογιστεί η λύση x του γραμμικού συστήματος $A * x = b$. Και στις δύο παραγοντοποίησεις, ο δεύτερος (κοινός) όρος των πράξεων Ω , που είναι το $2n^2$, αναφέρεται στην **εμπρός** και την **πίσω αντικατάσταση**. **Κάθε άμεση μέθοδος παραγοντοποίησης, όπως η LU, η PLU, η Cholesky και η QR απαιτεί τρία βήματα για να λύσει το γραμμικό σύστημα $A * x = b$.** Στο πρώτο βήμα γίνεται η διάσπαση του μητρώου A (π.χ. σε L, U ή σε P, L, U ή σε L, L^T ή σε Q, R) και στη συνέχεια, στα επόμενα δύο βήματα γίνεται πάντα η εμπρός και η πίσω αντικατάσταση, που έχουν διαφορετικές μορφές, ανάλογα με την παραγοντοποίηση που πραγματοποιείται. **Προτιμάται γενικά μεταξύ των άμεσων μεθόδων η χρήση της παραγοντοποίησης Cholesky**, επειδή -όπως αναφέρθηκε- είναι η πιο **οικονομική** σε σχέση με τις υπόλοιπες ($\Omega = \frac{n^3}{3} + 2n^2$ έναντι του κόστους $\Omega = \frac{2 * n^3}{3} + 2n^2$ στην LU και PLU και του κόστους $\Omega = \frac{4 * n^3}{3} + 2n^2$ στην QR), όμως υπάρχει ο περιορισμός ότι δεν μπορεί να εφαρμοστεί σε οποιοδήποτε μητρώο.

Ως **προϋπόθεση εφαρμογής** της παραγοντοποίησης Cholesky είναι το μητρώο A να είναι **Σ.Θ.Ο.** Η συμμετρία είναι μια ιδιότητα που ελέγχεται από παρατήρηση του μητρώου A (ισχύει ότι $A^T = A$ ή εναλλακτικά $a_{ij} = a_{ji}$), ενώ για τη **θετική ορισμότητα** (Θ.Ο.) του μητρώου αρκεί να ισχύει **ένα** από τα εξής: α) Το μητρώο A να έχει A.D.K. **και** θετικά διαγώνια στοιχεία β) όλες οι **ιδιοτιμές** του μητρώου A να είναι **θετικές** (επειδή πολλές φορές είναι δύσκολος ο υπολογισμός των ιδιοτιμών ενός μητρώου ηχη, διότι οδηγεί σε ένα πολυώνυμο η βαθμού, χρησιμοποιούμε **συχνά το θεώρημα των κύκλων Gershgorin** για να βρούμε διαστήματα ιδιοτιμών, χωρίς να τις υπολογίσουμε) γ) **όλοι** οι **οδηγοί** του μητρώου A να είναι **θετικοί** (το κριτήριο αυτό δεν είναι ιδιαίτερα εύχρηστο, διότι για να βρεθούν οι οδηγοί ενός μητρώου θα πρέπει να πρώτα υπολογίσουμε το άνω τριγωνικό μητρώο U και από εκεί να πάρουμε τα στοιχεία της κύριας διαγωνίου του) και δ) $\forall x \in \mathbb{R}^n$ με $x \neq 0$, να ισχύει ότι: $x^T * A * x > 0$ και ε) Όλες οι κύριες (κυρίαρχες) ορίζουσες να είναι **θετικές**. Όλες αυτές οι συνθήκες είναι **ικανές**, δηλαδή αν δεν ισχύει μια από αυτές, εξετάζουμε την επόμενη, για να ελέγχουμε αν το μητρώο είναι Θ.Ο.

Παράδειγμα (Θέμα Ιουνίου 2015): Ποια σειρά βημάτων είναι η καλύτερη επιλογή για την αριθμητική προσέγγιση της λύσης x του συστήματος $A * x = b$, όπου $x \in \mathbb{R}^n$ και $A = A^T$ σε περιβάλλοντα όπως η MATLAB;

- (i) Υπολογισμός του $C = A^{-1}$ και στη συνέχεια υπολογισμός του $C * b$.
- (ii) Αν το μητρώο είναι **συμμετρικό**, εκκίνηση της παραγοντοποίησης Cholesky και στη συνέχεια αν η Cholesky είναι **επιτυχής**, υπολογισμός του παράγοντα L, έτσι ώστε $A = L * L^T$ και στη συνέχεια επίλυση του $L * y = b$

και του $L^T * x = y$. Αν η Cholesky δεν είναι **επιτυχής**, υπολογισμός της παραγοντοποίησης LU με έξοδο τα μητρώα P, L, U (μερική οδήγηση) και στη συνέχεια υπολογισμός του $\hat{b} = P^T * b$ και επίλυση του $L * y = \hat{b}$ και του $U * x = y$.

(iii) Εφαρμογή παραγοντοποίησης LU με μερική οδήγηση στο A με έξοδο τα μητρώα P, L, U και επίλυση του $U * y = b$ και του $L * x = y$.

Λύση

Η σωστή απάντηση είναι η (ii), διότι η MATLAB πάντα ξεκινά να λύνει το γραμμικό σύστημα $A * x = b$ με την παραγοντοποίηση Cholesky, αν αυτό είναι εφικτό και στη συνέχεια αν δεν είναι εφικτό, χρησιμοποιεί LU παραγοντοποίηση με μερική οδήγηση. Δεν είναι όμως σωστό να εφαρμόσουμε από την αρχή LU παραγοντοποίηση, χωρίς πρώτα να ελέγξουμε αν μπορεί να εφαρμοστεί η παραγοντοποίηση Cholesky. Για το λόγο αυτό είναι λάθος το (iii). Όσον αφορά το (i) είναι λάθος, διότι **δεν** ενδείκνυται ο **απευθείας** υπολογισμός του αντίστροφου μητρώου A^{-1} , διότι αποτελεί μια χρονοβόρα διαδικασία που παράλληλα απαιτεί και μεγάλο όγκο πράξεων.

Παράδειγμα (Θέμα Φεβρουαρίου 2017) α) Δίνεται το **τριδιαγώνιο** και **συμμετρικό** μητρώο μεγέθους 100 με τις εξής προδιαγραφές: Η κύρια διαγώνιος περιέχει τις τιμές από 1 έως 100, δηλ. στη θέση (i, i) υπάρχει το i για $i = 1, 2, \dots, 100$ και η **υπερδιαγώνιος** είναι 1. Ποιο/ά από τα παρακάτω μπορείτε να συμπεράνετε για τις ιδιοτιμές;

- α) οι ιδιοτιμές είναι θετικές β) οι ιδιοτιμές κείνται στο μοναδιαίο δίσκο $\{z \in C, |z| \leq 1\}$ γ) Δεν υπάρχει αρνητική ιδιοτιμή δ) οι ιδιοτιμές είναι πραγματικές ε) Η μεγαλύτερη ιδιοτιμή είναι το 100.

Λύση

Από τη στιγμή που το μητρώο είναι **συμμετρικό**, **έχει εξ' ορισμού πραγματικές ιδιοτιμές**. Επομένως μια από τις σωστές απαντήσεις είναι η (δ). Μια δεύτερη σωστή απάντηση είναι και η (γ), διότι ισχύει ότι ένα μητρώο **με διαγώνια κυριαρχία (Δ.Κ.)** **έχει μη αρνητικές ιδιοτιμές**. **Υπενθυμίζεται ότι ένα μητρώο με αυστηρή διαγώνια κυριαρχία (Α.Δ.Κ.)** **έχει θετικές ιδιοτιμές**. Το συγκεκριμένο όμως μητρώο έχει διαγώνια κυριαρχία, όπως φαίνεται στη συνέχεια, παίρνοντας ένα μικρό δείγμα του (με κόκκινο χρώμα συμβολίζονται τα στοιχεία της υπερδιαγωνίου και με μπλε χρώμα τα στοιχεία της υποδιαγωνίου, τα οποία έχουν επίσης τιμή «1», λόγω

της συμμετρίας του μητρώου):
$$\begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 2 & 1 & 0 \\ 0 & 1 & 3 & 1 \\ 0 & 0 & 1 & 4 \end{bmatrix}$$
 Η (α) δεν είναι σωστή απάντηση, διότι δεν έχει A.Δ.Κ. Επιπλέον,

αν κάνουμε τη σχεδίαση των κύκλων Gershgorin, διαπιστώνουμε ότι οι ιδιοτιμές του ορίζονται σε **διαστήματα που τα δύο από αυτά ξεκινούν από το 0**. Εναλλακτικά, μπορούμε να κάνουμε συναλήθευση των ιδιοτιμών, κάνοντας χρήση των ακόλουθων ανισοτήτων:

$$|z - 1| \leq 1 \Rightarrow -1 \leq z - 1 \leq 1 \Rightarrow 0 \leq z \leq 2.$$

$$|z - 2| \leq 2 \Rightarrow -2 \leq z - 2 \leq 2 \Rightarrow 0 \leq z \leq 4.$$

Από τη συναλήθευση διαπιστώνουμε ότι τα διαστήματα στα οποία ορίζονται οι ιδιοτιμές δεν είναι αρνητικά, για το λόγο που αναφέραμε.

$$|z - 3| \leq 2 \Rightarrow -2 \leq z - 3 \leq 2 \Rightarrow 1 \leq z \leq 5.$$

$$|z - 4| \leq 1 \Rightarrow -1 \leq z - 4 \leq 1 \Rightarrow 3 \leq z \leq 5.$$

Σημείωση: Για να ήταν σωστή η απάντηση (ε) θα έπρεπε το μητρώο που δίνεται, να είναι **διαγώνιο ή τριγωνικό** (άνω ή κάτω), διότι σε οποιαδήποτε από αυτές τις δύο περιπτώσεις, οι ιδιοτιμές είναι τα στοιχεία της κύριας διαγωνίου. Άρα τότε η μεγαλύτερη ιδιοτιμή θα ήταν το 100.

Παρατήρηση: Αν και οι ιδιοτιμές ενός μητρώου μπορούν να βρεθούν τόσο από την εξίσωση $\det(\lambda * I - A) = 0$ όσο και από την εξίσωση $\det(A - \lambda * I) = 0$, **συνίσταται η χρήση της πρώτης εξίσωσης**, διότι σύμφωνα με σχετική σημείωση που υπάρχει στη Γραμμική Άλγεβρα, χαρακτηριστικό πολυώνυμο ενός μητρώου $A \in R^{n \times n}$, ονομάζεται το πολυώνυμο βαθμού n που συμβολίζεται με $p(\lambda) = \det(\lambda * I - A)$, ενώ το $\hat{p}(\lambda) = \det(A - \lambda * I)$. Επίσης, θα πρέπει να έχουμε υπόψη ότι όταν υπολογίζουμε με το χαρακτηριστικό πολυώνυμο τις ιδιοτιμές ενός μητρώου **ην**, αυτές θα πρέπει να είναι **n** στο πλήθος. Επίσης, θα προκύψει ένα χαρακτηριστικό πολυώνυμο $n -$ βαθμού, το οποίο είναι δύσκολο πολλές φορές να επιλυθεί, για να υπολογίσουμε τις ρίζες του. Για το λόγο αυτό προτιμάται η χρήση των κύκλων Gershgorin.

2.5.1 Πρώτο παράδειγμα εφαρμογής παραγοντοποίησης Cholesky

Έστω το μητρώο $A = \begin{bmatrix} 25 & 15 & -5 \\ 15 & 18 & 0 \\ -5 & 0 & 11 \end{bmatrix}$ για το οποίο θέλουμε να εκτελέσουμε –αν είναι δυνατό– την παραγοντοποίηση Cholesky.

Λύση

Το δοθέν μητρώο A είναι συμμετρικό ως προς την κύρια διαγώνιο (αφού $A^T = A$) και επίσης είναι και **Θετικά ορισμένο (Θ.Ο.)**, διότι έχει αυστηρή διαγώνια κυριαρχία (Α.Δ.Κ.) και θετικά διαγώνια στοιχεία. Άρα μπορεί να διασπαστεί –σύμφωνα με την παραγοντοποίηση Cholesky– σε μητρώα L και L^T της μορφής:

$$A = L * L^T \Rightarrow \begin{bmatrix} 25 & 15 & -5 \\ 15 & 18 & 0 \\ -5 & 0 & 11 \end{bmatrix} = \begin{bmatrix} l_{11} & 0 & 0 \\ l_{21} & l_{22} & 0 \\ l_{31} & l_{32} & l_{33} \end{bmatrix} * \begin{bmatrix} l_{11} & l_{21} & l_{31} \\ 0 & l_{22} & l_{32} \\ 0 & 0 & l_{33} \end{bmatrix} = \begin{bmatrix} l_{11}^2 & l_{11} * l_{21} & l_{31} * l_{11} \\ l_{21} * l_{11} & l_{21}^2 + l_{22}^2 & l_{31} * l_{21} + l_{32} * l_{22} \\ l_{31} * l_{11} & l_{21} * l_{31} + l_{22} * l_{32} & l_{31}^2 + l_{32}^2 + l_{33}^2 \end{bmatrix}.$$

Στη συνέχεια εξισώνουμε τα αντίστοιχα στοιχεία των δύο μητρώων (πρώτου και τελευταίου) μεταξύ τους, οπότε έχουμε ότι:

$$l_{11}^2 = 25 \Rightarrow l_{11} = 5 \text{ (παίρνουμε πάντα τη Θετική ρίζα)}$$

$$l_{11} * l_{21} = 15 \Rightarrow l_{21} = 15/5 = 3$$

$$l_{31} * l_{11} = -5 \Rightarrow l_{31} = -5/5 = -1$$

$$l_{21}^2 + l_{22}^2 = 18 \Rightarrow l_{22}^2 = 18 - 9 = 9 \Rightarrow l_{22} = 3 \text{ (παίρνουμε πάντα τη Θετική ρίζα)}$$

$$l_{31} * l_{21} + l_{32} * l_{22} = 0 \Rightarrow (-1) * 3 + l_{32} * 3 = 0 \Rightarrow l_{32} = 3/3 = 1$$

$$l_{31}^2 + l_{32}^2 + l_{33}^2 = 11 \Rightarrow l_{33}^2 = 11 - (-1)^2 - 1^2 = 9 \Rightarrow l_{33} = 3. \text{ Παράγοντας Cholesky: } L = \begin{bmatrix} l_{11} & 0 & 0 \\ l_{21} & l_{22} & 0 \\ l_{31} & l_{32} & l_{33} \end{bmatrix} = \begin{bmatrix} 5 & 0 & 0 \\ 3 & 3 & 0 \\ -1 & 1 & 3 \end{bmatrix}.$$

2.5.2 Δεύτερο παράδειγμα εφαρμογής παραγοντοποίησης Cholesky

Έστω το μητρώο $A = \begin{bmatrix} 4 & 0 & -1 & -1 \\ 0 & 4 & 0 & -1 \\ -1 & 0 & 4 & 0 \\ -1 & -1 & 0 & 4 \end{bmatrix}$ για το οποίο α) θέλουμε να εκτελέσουμε την παραγοντοποίηση Cholesky και β) Να υπολογίσουμε την παράσταση $\hat{L} * D * \hat{L}^T$.

Λύση

α) Το δοθέν μητρώο A είναι –και πάλι– συμμετρικό ως προς την κύρια διαγώνιο (αφού $A^T = A$) και επίσης είναι και θετικά ορισμένο (Θ.Ο.), διότι έχει αυστηρή διαγώνια κυριαρχία (κάθε μητρώο που έχει ΑΔΚ είναι και αντιστρέψιμο, δηλ. υπολογίζεται το A^{-1}) **και** θετικά διαγώνια στοιχεία. Άρα μπορεί να διασπαστεί –σύμφωνα με την παραγοντοποίηση Cholesky – σε μητρώα L και L^T της μορφής:

$$A = L * L^T \Rightarrow \begin{bmatrix} 4 & 0 & -1 & -1 \\ 0 & 4 & 0 & -1 \\ -1 & 0 & 4 & 0 \\ -1 & -1 & 0 & 4 \end{bmatrix} = \begin{bmatrix} l_{11} & 0 & 0 & 0 \\ l_{21} & l_{22} & 0 & 0 \\ l_{31} & l_{32} & l_{33} & 0 \\ l_{41} & l_{42} & l_{43} & l_{44} \end{bmatrix} * \begin{bmatrix} l_{11} & l_{21} & l_{31} & l_{41} \\ 0 & l_{22} & l_{32} & l_{42} \\ 0 & 0 & l_{33} & l_{43} \\ 0 & 0 & 0 & l_{44} \end{bmatrix} =$$

$$\begin{bmatrix} l_{11}^2 & l_{11} * l_{21} & l_{11} * l_{31} & l_{11} * l_{41} \\ l_{21} * l_{11} & l_{21}^2 + l_{22}^2 & l_{21} * l_{31} + l_{22} * l_{32} & l_{21} * l_{41} + l_{22} * l_{42} \\ l_{31} * l_{11} & l_{31} * l_{21} + l_{32} * l_{22} & l_{31}^2 + l_{32}^2 + l_{33}^2 & l_{31} * l_{41} + l_{32} * l_{42} + l_{33} * l_{43} \\ l_{41} * l_{11} & l_{41} * l_{21} + l_{42} * l_{22} & l_{41} * l_{31} + l_{42} * l_{32} + l_{43} * l_{33} & l_{41}^2 + l_{42}^2 + l_{43}^2 + l_{44}^2 \end{bmatrix}$$

Από την επίλυση της προηγούμενης εξίσωσης βρίσκουμε τα στοιχεία του μητρώου L . Πιο συγκεκριμένα:

$$l_{11}^2 = 4 \Rightarrow l_{11} = 2 \text{ (παίρνουμε πάντα τη θετική ρίζα)}$$

$$l_{11} * l_{21} = 0 \Rightarrow l_{21} = 0. \text{ Επίσης, } l_{11} * l_{31} = -1 \Rightarrow l_{31} = -1/2 \text{ και } l_{11} * l_{41} = -1 \Rightarrow l_{41} = -1/2 \text{ και } l_{21}^2 + l_{22}^2 = 4 \Rightarrow l_{22} = 2.$$

$$l_{31} * l_{21} + l_{32} * l_{22} = 0 \Rightarrow -1/2 * 0 + l_{32} * 2 = 0 \Rightarrow l_{32} = 0. \text{ Επίσης έχουμε ότι: } l_{31}^2 + l_{32}^2 + l_{33}^2 = 4 \Rightarrow (-1/2)^2 + 0^2$$

$$+ l_{33}^2 = 4 \Rightarrow l_{33} = \frac{\sqrt{15}}{2}. \text{ Το } l_{41} * l_{11} = -1 \Rightarrow l_{41} = -1/2 \text{ και το } l_{41} * l_{21} + l_{42} * l_{22} = -1 \Rightarrow (-1/2) * 0 + l_{42} * 2 = -1 \Rightarrow l_{42} = -1/2. \text{ Επίσης, } l_{41} * l_{31} + l_{42} * l_{32} + l_{43} * l_{33} = 0 \Rightarrow l_{43} = -\frac{\sqrt{15}}{30} \text{ και } l_{41}^2 + l_{42}^2 + l_{43}^2 + l_{44}^2 = 4 \Rightarrow l_{44}^2 = 4 - 2/4 - \frac{15}{30^2} \Rightarrow l_{44} =$$

$$\sqrt{\frac{7}{2} - \frac{15}{900}} = \frac{\sqrt{15}}{30} - \frac{\sqrt{20}}{30} = \frac{\sqrt{15} - \sqrt{209}}{30}. \text{ Άρα το μητρώο (παράγοντας Cholesky) } L = \begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ -1/2 & 0 & \frac{\sqrt{15}}{2} & 0 \\ -1/2 & -1/2 & -\frac{\sqrt{15}}{30} & \frac{\sqrt{15} - \sqrt{209}}{30} \end{bmatrix}$$

β) Για τον υπολογισμό της παράστασης: $\hat{L} * D * \hat{L}^T = \hat{L} * D^{1/2} * (D^{1/2})^T * \hat{L}^T = \hat{L} * D^{1/2} * (\hat{L} * D^{1/2})^T = L * L^T$, όπου

$$L = \hat{L} * D^{1/2} \text{ και } L^T = (\hat{L} * D^{1/2})^T. \text{ Κάθε διαγώνιο μητρώο είναι εξ' ορισμού και συμμετρικό, οπότε ισχύει ότι: } D^{1/2} = (D^{1/2})^T. \text{ Για τη συγκεκριμένη παραγοντοποίηση, θα πρέπει να υπολογίσουμε τα μητρώα θα πρέπει να υπολο-$$

γίσουμε τα μητρώα D και \hat{L} . Ξεκινάμε με τον υπολογισμό του διαγώνιου μητρώου D , το οποίο επειδή είναι

$$\text{υψωμένο στο } \%, \text{ θα πρέπει πρώτα να το φέρω στη μορφή: } D = \begin{bmatrix} 4 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 \\ 0 & 0 & \frac{15}{4} & 0 \\ 0 & 0 & 0 & \frac{(\sqrt{15} - \sqrt{209})^2}{30^2} \end{bmatrix}, \text{ δηλαδή να περιέχει}$$

στην κύρια διαγώνιό του τα στοιχεία της κύριας διαγωνίου του μητρώου L , υψωμένα στο **τετράγωνο**, επειδή

στη συνέχεια στην έκφραση $L = \hat{L} * D^{1/2}$ υψώνεται στο $1/2$. Επίσης το μητρώο \hat{L} θα περιέχει "1" στην κύρια διαγώνιο του και "0" σε εκείνες τις θέσεις που και το μητρώο L έχει "0", δηλαδή θα είναι της μορφής:

$$\hat{L} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ x_1 & 0 & 1 & 0 \\ x_2 & x_3 & x_4 & 1 \end{bmatrix}, \text{ έτσι ώστε το γινόμενο } \hat{L} * D^{1/2} = L \Rightarrow \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ x_1 & 0 & 1 & 0 \\ x_2 & x_3 & x_4 & 1 \end{bmatrix} * \begin{bmatrix} 4 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 \\ 0 & 0 & \frac{15}{4} & 0 \\ 0 & 0 & 0 & \frac{(\sqrt{15}-\sqrt{209})^2}{30^2} \end{bmatrix}^{1/2} =$$

$$\begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ -1/2 & 0 & \frac{\sqrt{15}}{2} & 0 \\ -1/2 & -1/2 & -\frac{\sqrt{15}}{30} & \frac{\sqrt{15}-\sqrt{209}}{30} \end{bmatrix} \Rightarrow \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ x_1 & 0 & 1 & 0 \\ x_2 & x_3 & x_4 & 1 \end{bmatrix} * \begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & \frac{\sqrt{15}}{2} & 0 \\ 0 & 0 & 0 & \frac{\sqrt{15}-\sqrt{209}}{30} \end{bmatrix} = \begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ -1/2 & 0 & \frac{\sqrt{15}}{2} & 0 \\ -1/2 & -1/2 & -\frac{\sqrt{15}}{30} & \frac{\sqrt{15}-\sqrt{209}}{30} \end{bmatrix}$$

$$\Rightarrow \begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 2 * x_1 & 0 & \frac{\sqrt{15}}{2} & 0 \\ 2 * x_2 & 2 * x_3 & \frac{\sqrt{15}}{2} * x_4 & \frac{\sqrt{15}-\sqrt{209}}{30} \end{bmatrix} = \begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ -1/2 & 0 & \frac{\sqrt{15}}{2} & 0 \\ -1/2 & -1/2 & -\frac{\sqrt{15}}{30} & \frac{\sqrt{15}-\sqrt{209}}{30} \end{bmatrix} \Rightarrow 2 * x_1 = -1/2 \Rightarrow x_1 = -1/4. \text{ Με ανάλογο}$$

τρόπο προκύπτει ότι $2 * x_2 = -1/2 \Rightarrow x_2 = -1/4$ και $2 * x_3 = -1/2 \Rightarrow x_3 = -1/4$ και $\frac{\sqrt{15}}{2} * x_4 = -\frac{\sqrt{15}}{30} \Rightarrow x_4 = -1/15$. Στη συνέχεια έχοντας υπολογίσει το μητρώο \hat{L} , υπολογίζουμε το ζητούμενο $\hat{L} * D * \hat{L}^T$. Άρα το μητρώο $\hat{L} =$

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ -1/4 & 0 & 1 & 0 \\ -1/4 & -1/4 & -1/15 & 1 \end{bmatrix} \text{ και το μητρώο } D^{1/2} = \begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & \frac{\sqrt{15}}{2} & 0 \\ 0 & 0 & 0 & \frac{\sqrt{15}-\sqrt{209}}{30} \end{bmatrix}.$$

2.5.3. Τρίτο παράδειγμα εφαρμογής παραγοντοποίησης Cholesky

Να γράψετε τον κάτω τριγωνικό παράγοντα Cholesky για το μητρώο $A = \begin{bmatrix} 4 & 0 & -2 & 0 \\ 0 & 1 & -2 & -2 \\ -2 & -2 & 6 & 4 \\ 0 & -2 & 4 & 5 \end{bmatrix}$, ειδάλλως να εξηγήσετε αν δεν υπάρχει.

Λύση

Από μια πρώτη παρατήρηση φαίνεται ότι το μητρώο που έχει δοθεί είναι συμμετρικό. Στη συνέχεια γίνεται διερεύνηση για το αν είναι Θ.Ο. Έχει θετικά διαγώνια στοιχεία, αλλά δεν έχει Α.Δ.Κ. **Αυτή η συνθήκη όμως είναι ικανή και όχι αναγκαία.** Για το λόγο αυτό εξετάζουμε **άλλες συνθήκες**, όπως το να είναι όλες οι κύριες ορίζουσες του μητρώου θετικές, να είναι όλες οι ιδιοτιμές του μητρώου θετικές ή όλοι οι οδηγοί του μητρώου θετικοί ή τέλος $\forall x \in \mathbb{R}^n$ να ισχύει ότι: $x^T * A * x > 0$. Αν εξετάζουμε το κριτήριο των κύριων οριζουσών, μπορούμε να διαπιστώσουμε ότι $\det(A(1:1, 1:1)) = 4 > 0$, $\det(A(1:2, 1:2)) = 4 > 0$, $\det(A(1:3, 1:3)) = 4 > 0$, $\det(A(1:4, 1:4)) = 4 > 0$. Δηλαδή παρατηρούμε ότι όλες οι κύριες ορίζουσες είναι θετικές και επομένως το μητρώο που έχει δοθεί, είναι Σ.Θ.Ο. Στη συνέχεια προχωράμε στον υπολογισμό του κάτω τριγωνικού παράγοντα Cholesky, επιλύοντας

$$\text{το σύστημα } A = L * L^T \Rightarrow \begin{bmatrix} 4 & 0 & -2 & 0 \\ 0 & 1 & -2 & -2 \\ -2 & -2 & 6 & 4 \\ 0 & -2 & 4 & 5 \end{bmatrix} = \begin{bmatrix} l_{11} & 0 & 0 & 0 \\ l_{21} & l_{22} & 0 & 0 \\ l_{31} & l_{32} & l_{33} & 0 \\ l_{41} & l_{42} & l_{43} & l_{44} \end{bmatrix} * \begin{bmatrix} l_{11} & l_{21} & l_{31} & l_{41} \\ 0 & l_{22} & l_{32} & l_{42} \\ 0 & 0 & l_{33} & l_{43} \\ 0 & 0 & 0 & l_{44} \end{bmatrix} \Rightarrow \begin{bmatrix} 4 & 0 & -2 & 0 \\ 0 & 1 & -2 & -2 \\ -2 & -2 & 6 & 4 \\ 0 & -2 & 4 & 5 \end{bmatrix} =$$

$$\begin{bmatrix} l_{11}^2 & l_{11} * l_{21} & l_{11} * l_{31} & l_{11} * l_{41} \\ l_{21} * l_{11} & l_{21}^2 + l_{22}^2 & l_{21} * l_{31} + l_{22} * l_{23} & l_{21} * l_{41} + l_{22} * l_{24} \\ l_{31} * l_{11} & l_{31} * l_{21} + l_{32} * l_{22} & l_{31}^2 + l_{32}^2 + l_{33}^2 & l_{31} * l_{41} + l_{32} * l_{42} + l_{33} * l_{43} \\ l_{41} * l_{11} & l_{41} * l_{21} + l_{42} * l_{22} & l_{41} * l_{31} + l_{42} * l_{32} + l_{43} * l_{33} & l_{41}^2 + l_{42}^2 + l_{43}^2 + l_{44}^2 \end{bmatrix} \Rightarrow l_{11}^2 = 4 \Rightarrow l_{11} = 2$$

(παίρνουμε πάντα τη θετική ρίζα) και $l_{11} * l_{22} = 0 \Rightarrow l_{22} = 0$ και $l_{11} * l_{31} = -2 \Rightarrow l_{31} = -1$ και $l_{11} * l_{41} = 0 \Rightarrow l_{41} = 0$ και $l_{21}^2 + l_{22}^2 = 1 \Rightarrow l_{22} = 2$. (παίρνουμε πάντα τη θετική ρίζα) κ.ο.κ. Μετά την ολοκλήρωση των πράξεων

$$\text{βρίσκουμε ότι το μητρώο } L = \begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ -1 & -2 & 1 & 0 \\ 0 & -2 & 0 & 1 \end{bmatrix} \text{ (παράγοντας Cholesky).}$$

Παρατήρηση: ένας πιο σύντομος τρόπος εύρεσης διαστημάτων ιδιοτιμών – χωρίς να απαιτείται ο υπολογισμός τους - είναι η χρήση των κύκλων (ή δίσκων) Gershgorin που αναφέρονται στη συνέχεια. Το θεώρημα αυτό βρίσκει διαστήματα ιδιοτιμών χωρίς να τις υπολογίζει. Μια πρώτη προσέγγιση της μεθόδου αυτής, απαιτεί την ικανοποίηση των συνθηκών: $\Delta_i^{(r)} = \left\{ z \in C : |z - a_{ii}| \leq \sum_{j=1, j \neq i}^n |a_{ij}| \right\}$, όπου $\Delta_i^{(r)}$ οι δίσκοι (κύκλοι) Gershgorin. Πιο συγκεκριμένα, από την εφαρμογή αυτής της ανισότητας στην προκειμένη περίπτωση έχουμε ότι:

$$|z - 4| \leq 2 \Rightarrow -2 \leq z - 4 \leq 2 \Rightarrow 2 \leq z \leq 6$$

Σημείωση: Για να κάνουμε συναλήθευση διαστημάτων

$$|z - 1| \leq 4 \Rightarrow -4 \leq z - 1 \leq 4 \Rightarrow -3 \leq z \leq 5$$

ιδιοτιμών, παίρνουμε από τις μικρότερες τιμές των α-

$$|z - 6| \leq 8 \Rightarrow -8 \leq z - 6 \leq 8 \Rightarrow -2 \leq z \leq 14$$

νισοτήτων τη μεγαλύτερη και αντίστοιχα από τις μεγα-

$$|z - 5| \leq 6 \Rightarrow -6 \leq z - 5 \leq 6 \Rightarrow -1 \leq z \leq 11$$

λύτερες τιμές των ανισοτήτων, τη μικρότερη.

Σχεδιάζουμε στη συνέχεια τα τέσσερα διαστήματα των ιδιοτιμών και διαπιστώνουμε ότι αυτά συναληθεύουν στο διάστημα $[2, 5]$, το οποίο περιέχει μόνο θετικές ιδιοτιμές. Έτσι αναφορικά με τα κριτήρια για τα Σ.Θ.Ο. μητρώα, ικανοποιείται **το κριτήριο των θετικών ιδιοτιμών**.

Παρατήρηση: Αν το μητρώο είναι **πραγματικό** και **συμμετρικό** ή **μιγαδικό** και **ερμιτιανό** και η ένωση των δίσκων Gershgorin έστω $\text{Union_i}(D_i)$ βρίσκεται εξ' ολοκλήρου στο θετικό ημιεπίπεδο (δηλ. $z \in \text{Union_i}(D_i)$ ή με άλλα λόγια το πραγματικό μέρος του z είναι θετικό $\text{Re}(z) > 0$), τότε το μητρώο είναι Σ.Θ.Ο. Υπενθυμίζεται ότι επειδή το μητρώο είναι πραγματικό και συμμετρικό ή μιγαδικό και ερμιτιανό, όλες οι ιδιοτιμές του είναι **πραγματικές** και επομένως εξετάζεται μόνο η τομή των δίσκων Gershgorin με τον πραγματικό άξονα του μιγαδικού επιπέδου. Στην περίπτωση αυτή ισχύει ότι $|z - a_{ii}| \leq r_i$, οπότε αν ισχύει ότι $r_i < |a_{ii}|$ έπειται ότι $z > 0$.

2.5.4. Τέταρτο παράδειγμα εφαρμογής παραγοντοποίησης Cholesky

Να γράψτε τον κάτω τριγωνικό παράγοντα Cholesky για το μητρώο $\begin{bmatrix} 3 & -2 & 1 \\ -2 & 2 & 0 \\ 1 & 0 & 2 \end{bmatrix}$, ειδάλλως να εξηγήσετε αν δεν υπάρχει.

Λύση

Καταρχήν το μητρώο ελέγχεται αν είναι Σ.Θ.Ο. **Είναι συμμετρικό, αλλά δεν έχει ΑΔΚ.** Αυτή η συνθήκη όμως -όπως έχει ήδη αναφερθεί- είναι ικανή και όχι αναγκαία. Για το λόγο αυτό εξετάζουμε άλλα κριτήρια για τη θετική ορισμότητα. Αν χρησιμοποιήσουμε το κριτήριο των κύριων οριζουσών, τότε θα έχουμε ότι: $\det(A(1:1, 1:1)) = 3 > 0$, $\det(A(1:2, 1:2)) = 6 - 4 = 2 > 0$, $\det(A(1:3, 1:3)) = (-1)^4 * 1 * (-2) + (-1)^6 * 2 * (6 - 4) = -2 + 4 = 2 > 0$.

Παρατηρούμε ότι όλες οι κύριες ορίζουσες του μητρώου είναι θετικές και συνεπώς το μητρώο **είναι θετικά ορισμένο.**

Εναλλακτικά μπορούμε να χρησιμοποιήσουμε και το θεώρημα των κύκλων (ή δίσκων) Gershgorin. Επειδή έχει δοθεί ένα μητρώο που είναι πραγματικό και συμμετρικό, οι ιδιοτιμές του -σύμφωνα με την παρατήρηση της προηγούμενης άσκησης- είναι πραγματικές και από το θεώρημα κύκλων Gershgorin ισχύει ότι $|z - a_{ii}| \leq r_i$, οπότε αν ισχύει ότι $r_i < |a_{ii}|$ τότε έπεται ότι $z > 0$. Πράγματι στην πρώτη γραμμή έχουμε ότι $-1 < 3$, στη δεύτερη γραμμή έχουμε ότι $-2 < 2$ και στην τρίτη γραμμή έχουμε ότι $1 < 2$. **Άρα ισχύει ότι όλες οι ιδιοτιμές είναι θετικές.** Εναλλακτικά μπορούμε να χρησιμοποιήσουμε και πάλι το θεώρημα των κύκλων (ή δίσκων) Gershgorin και να κάνουμε συναλήθευση ιδιοτιμών. Πιο συγκεκριμένα, ισχύει ότι:

$$|z - 3| \leq 3 \Rightarrow -3 \leq z - 3 \leq 3 \Rightarrow 0 \leq z \leq 6$$

$$|z - 2| \leq 2 \Rightarrow -2 \leq z - 2 \leq 2 \Rightarrow 0 \leq z \leq 4$$

$$|z - 1| \leq 1 \Rightarrow -1 \leq z - 1 \leq 1 \Rightarrow 1 \leq z \leq 3$$

Παρατηρούμε ότι οι ιδιοτιμές συναληθεύουν στο διάστημα $[1, 3]$ και επομένως **είναι όλες θετικές** και το μητρώο **είναι θετικά ορισμένο.** Στη συνέχεια εφαρμόζουμε τα βήματα της παραγοντοποίησης Cholesky:

$$A = L * L^T \Rightarrow \begin{bmatrix} 3 & -2 & 1 \\ -2 & 2 & 0 \\ 1 & 0 & 2 \end{bmatrix} = \begin{bmatrix} l_{11} & 0 & 0 \\ l_{21} & l_{22} & 0 \\ l_{31} & l_{32} & l_{33} \end{bmatrix} * \begin{bmatrix} l_{11} & l_{21} & l_{31} \\ 0 & l_{22} & l_{32} \\ 0 & 0 & l_{33} \end{bmatrix} = \begin{bmatrix} l_{11}^2 & l_{11} * l_{21} & l_{11} * l_{31} \\ l_{21} * l_{11} & l_{21}^2 + l_{22}^2 & l_{21} * l_{31} + l_{22} * l_{32} \\ l_{31} * l_{11} & l_{31} * l_{21} + l_{32} * l_{22} & l_{31}^2 + l_{32}^2 + l_{33}^2 \end{bmatrix}$$

$$\Rightarrow \text{ο παράγοντας Cholesky είναι το μητρώο } L = \begin{bmatrix} \sqrt{3} & 0 & 0 \\ \frac{-2}{\sqrt{3}} & \sqrt{\frac{2}{3}} & 0 \\ \frac{1}{\sqrt{3}} & \sqrt{\frac{2}{3}} & 1 \end{bmatrix}.$$

2.5.5. Διατήρηση ιδιοτήτων ΑΔΚ και ΣΘΟ μετά από ένα βήμα εφαρμογής απαλοιφής Gauss και Cholesky

α) Να δείξετε ότι στο μητρώο $A = \begin{bmatrix} -4 & 1 & 1 & 0 \\ 1 & 3 & -2 & 1 \\ 0 & 0 & 5 & 1 \\ 2 & 1 & 1 & -3 \end{bmatrix}$ που έχει **ΑΔΚ κατά στήλες**, αυτή διατηρείται στο συμπλήρωμα του Schur, μετά από ένα βήμα της απαλοιφής Gauss.

β) Να δείξετε ότι στο μητρώο $A = \begin{bmatrix} 4 & 2 & 0 & 0 \\ 2 & 5 & -1 & 0 \\ 0 & -1 & 3 & 1 \\ 0 & 0 & 1 & 2 \end{bmatrix}$ που είναι **ΣΘΟ**, αυτή η ιδιότητα διατηρείται στο συμπλήρωμα του Schur, μετά από ένα βήμα της παραγοντοποίησης Cholesky.

Λύση

α) Ισχύει το εξής μαθηματικό σχήμα, αναφορικά με την περίπτωση αυτή (S είναι το συμπλήρωμα Schur):

$$\begin{pmatrix} \alpha_{1,1} & v^T \\ u & A_{2:n,2:n} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ u/\alpha_{1,1} & I \end{pmatrix} \begin{pmatrix} \alpha_{1,1} & v^T \\ 0 & \underbrace{A_{2:n,2:n} - uv^T/\alpha_{1,1}}_S \end{pmatrix}$$

Πρέπει να εφαρμόσουμε το σχήμα αυτό στο μητρώο A (που έχει ΑΔΚ κατά στήλες), οπότε μετά από ένα βήμα της απαλοιφής Gauss έχουμε:

$$\left[\begin{array}{ccc|ccccc} -4 & 1 & 1 & 0 & & & & \\ 1 & 3 & -2 & 1 & & & & \\ 0 & 0 & 5 & 1 & & & & \\ 2 & 1 & 1 & -3 & & & & \end{array} \right] = \left[\begin{array}{ccc|ccccc} 1 & 0 & 0 & 0 & & & & \\ -1/4 & 1 & 0 & 0 & & & & \\ 0 & 0 & 1 & 0 & & & & \\ -2/4 & 0 & 0 & 1 & & & & \end{array} \right] * \left[\begin{array}{ccc|ccccc} -4 & 1 & 1 & 0 & & & & \\ 0 & 0 & 0 & 0 & & & & \\ 0 & 0 & S & & & & & \\ 0 & 0 & 0 & & & & & \end{array} \right]. \text{ Πρέπει να υπολογιστεί το συμπλήρωμα Schur (S).}$$

$$\text{Το μητρώο } S = A_{2:4,2:4} - uv^T/\alpha_{11} = \begin{bmatrix} 3 & -2 & 1 \\ 0 & 5 & 1 \\ 1 & 1 & -3 \end{bmatrix} - \left(-\frac{1}{4}\right)^* \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix} * \begin{bmatrix} 1 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 3 & -2 & 1 \\ 0 & 5 & 1 \\ 1 & 1 & -3 \end{bmatrix} + \frac{1}{4} * \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 0 \\ 2 & 2 & 0 \end{bmatrix} =$$

$$\begin{bmatrix} 13/4 & -7/4 & 1 \\ 0 & 5 & 1 \\ 3/2 & 3/2 & -3 \end{bmatrix}. \text{ Παρατηρούμε ότι στο συμπλήρωμα Schur διατηρείται η ΑΔΚ κατά στήλες.}$$

β) Ισχύει το εξής μαθηματικό σχήμα, αναφορικά με την περίπτωση αυτή (S είναι το συμπλήρωμα Schur):

$$\begin{pmatrix} \alpha_{1,1} & v^T \\ v & A_{2:n,2:n} \end{pmatrix} = \begin{pmatrix} \lambda_{1,1} & 0 \\ v/\lambda_{1,1} & I \end{pmatrix} \begin{pmatrix} \lambda_{1,1} & v^T/\lambda_{1,1} \\ 0 & \underbrace{A_{2:n,2:n} - vv^T/\alpha_{1,1}}_S \end{pmatrix}$$

όπου $\lambda_{1,1} = \sqrt{\alpha_{11}}$

Πρέπει να εφαρμόσουμε το σχήμα αυτό στο μητρώο A (που είναι ΣΘΟ), οπότε μετά από ένα βήμα της παραγοντοποίησης Cholesky έχουμε:

$$\left[\begin{array}{ccc|ccccc} 4 & 2 & 0 & 0 & & & & \\ 2 & 5 & -1 & 0 & & & & \\ 0 & -1 & 3 & 1 & & & & \\ 0 & 0 & 1 & 2 & & & & \end{array} \right] = \left[\begin{array}{ccc|ccccc} \sqrt{4} & 0 & 0 & 0 & & & & \\ 1 & 1 & 0 & 0 & & & & \\ 0 & 0 & 1 & 0 & & & & \\ 0 & 0 & 0 & 1 & & & & \end{array} \right] * \left[\begin{array}{ccc|ccccc} \sqrt{4} & 1 & 0 & 0 & & & & \\ 0 & 0 & 0 & 0 & & & & \\ 0 & 0 & S & & & & & \\ 0 & 0 & 0 & & & & & \end{array} \right]. \text{ Πρέπει να υπολογιστεί το συμπλήρωμα Schur (S). Το}$$

$$\text{μητρώο } S = A_{2:4,2:4} - vv^T/\alpha_{11} = \begin{bmatrix} 5 & -1 & 0 \\ -1 & 3 & 1 \\ 0 & 1 & 2 \end{bmatrix} - \frac{1}{4} * \begin{bmatrix} 2 \\ 0 \\ 0 \end{bmatrix} * \begin{bmatrix} 2 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 5 & -1 & 0 \\ -1 & 3 & 1 \\ 0 & 1 & 2 \end{bmatrix} - \frac{1}{4} * \begin{bmatrix} 4 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 5 & -1 & 0 \\ -1 & 3 & 1 \\ 0 & 1 & 2 \end{bmatrix} -$$

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 4 & -1 & 0 \\ -1 & 3 & 1 \\ 0 & 1 & 2 \end{bmatrix}. \text{ Παρατηρούμε ότι στο συμπλήρωμα Schur διατηρείται η ιδιότητα ΣΘΟ μετά από ένα βήμα της παραγοντοποίησης Cholesky.}$$

2.5.6 Άσκηση με παραγοντοποίηση Cholesky (παλιό θέμα)

Έστω ότι δοκιμάζετε την παραγοντοποίηση Cholesky σε πραγματικό συμμετρικό μητρώο B που είναι **αντιστρέψιμο** και ότι σε κάποιο βήμα του αλγόριθμου έχει προκύψει η ανάγκη να υπολογίσετε την τετραγωνική ρίζα αρνητικού αριθμού. Ποιο/ά από τα παρακάτω είναι σωστό/ά; α) Υπάρχει διάνυσμα x τέτοιο ώστε $x^T * B * x < 0$ β) Υπάρχει διάνυσμα x τέτοιο ώστε $B * x = 0$, γ) Οι ιδιοτιμές του B είναι όλες αρνητικές δ) Όλα τα προηγούμενα ε) μόνο τα (α) και (γ) στ) μόνο τα (β) και (γ).

Λύση

Επειδή σε κάποιο βήμα του αλγόριθμου προκύπτει η ανάγκη υπολογισμού της τετραγωνικής ρίζας αρνητικού αριθμού, αυτό σημαίνει ότι το αρχικό μητρώο B δεν είναι Θ.Ο., δηλαδή ισχύει ότι: $x^T * B * x < 0$. **Άρα σωστή απάντηση είναι η (α)**. Το (β) δεν ισχύει, διότι από τη στιγμή που το μητρώο B είναι αντιστρέψιμο, **ισχύει ότι έχει μια λύση που δεν είναι η μηδενική**. Επίσης, το (γ) **δεν ισχύει, διότι από τη στιγμή που το μητρώο B είναι αντιστρέψιμο, οι ιδιοτιμές του είναι όλες μη μηδενικές**, αλλά όχι κατ' ανάγκη αρνητικές. Δηλαδή μπορεί να είναι και θετικές.

2.5.7 Άσκηση με LU παραγοντοποίηση με μερική οδήγηση (παλιό θέμα)

Δίνεται το μητρώο: $A = \begin{bmatrix} 1 & 6 & 2 \\ 2 & 12 & 5 \\ -1 & -3 & -5 \end{bmatrix}$. Μπορεί αυτό το μητρώο να παραγοντοποιηθεί κατά LU? Αν όχι, υπάρχει κάποιος τρόπος να αλλάξουμε το μητρώο αυτό ώστε η LU παραγοντοποίηση να είναι εφικτή; Στην περίπτωση αυτήν, βρείτε τους σχετικούς όρους της παραγοντοποίησης.

Λύση

Προϋπόθεση ύπαρξης LU παραγοντοποίησης ενός μητρώου: Αυστηρά διαγώνια κυριαρχία: Αν $A \in \mathbb{R}^{n \times n}$ είναι αυστηρά ΔΚ κατά στήλες, τότε υπάρχει η παραγοντοποίηση $A = LU$. **Εναλλακτικά**, αναφέρουμε ότι για να μπορεί να παραγοντοποιηθεί ένα μητρώο κατά LU θα πρέπει **κάθε κύριο υπομητρώο να είναι αντιστρέψιμο**. Στη συγκεκριμένη περίπτωση που η διάσταση του μητρώου είναι μικρή (διότι γενικά είναι ασύμφορος ο υπολογισμός μιας ορίζουσας) εξετάζουμε αν οι αντίστοιχες ορίζουσες είναι μη μηδενικές. Αρκεί και μια κύρια ορίζουσα να είναι μηδενική, για να πούμε πως το μητρώο αυτό δεν μπορεί να παραγοντοποιηθεί κατά LU.

Αυτό ισχύει στην προκειμένη περίπτωση για την ορίζουσα $\det(A(1:2, 1:2)) = \begin{vmatrix} 1 & 6 \\ 2 & 12 \end{vmatrix} = 0$. Επομένως, **δεν είναι εφικτή** η διαδικασία. Θα πρέπει να κάνουμε κατάλληλες εναλλαγές στις γραμμές του A , δηλ. να εφαρμόσουμε την PLU παραγοντοποίηση ή αλλιώς την LU με μερική οδήγηση. Έτσι, κάνουμε αρχικά αντιμετάθεση 1^{ης} - 2^{ης}

γραμμής. Έχουμε ότι $P_1 * A = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} * \begin{bmatrix} 1 & 6 & 2 \\ 2 & 12 & 5 \\ -1 & -3 & -5 \end{bmatrix} = \begin{bmatrix} 2 & 12 & 5 \\ 1 & 6 & 2 \\ -1 & -3 & -5 \end{bmatrix}$. Στη συνέχεια έχουμε $L_1 * P_1 * A =$

$\begin{bmatrix} 1 & 0 & 0 \\ -1/2 & 1 & 0 \\ 1/2 & 0 & 1 \end{bmatrix} * \begin{bmatrix} 2 & 12 & 5 \\ 1 & 6 & 2 \\ -1 & -3 & -5 \end{bmatrix} = \begin{bmatrix} 2 & 12 & 5 \\ 0 & 0 & -1/2 \\ 0 & 3 & -5/2 \end{bmatrix}$. Στη συνέχεια κάνουμε εναλλαγή 2^{ης} και 3^{ης} γραμμής του

προηγούμενου μητρώου και έχουμε: $P_2 * L_1 * P_1 * A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} * \begin{bmatrix} 2 & 12 & 5 \\ 0 & 0 & -1/2 \\ 0 & 3 & -5/2 \end{bmatrix} = \begin{bmatrix} 2 & 12 & 5 \\ 0 & 3 & -5/2 \\ 0 & 0 & -1/2 \end{bmatrix} = U$.

Παρατηρούμε ότι το μητρώο που προέκυψε είναι ήδη άνω τριγωνικό, οπότε έχουμε ότι $P_2 * L_1 * P_1 * A = U \Rightarrow$ πρέπει να εφαρμόσουμε στη συνέχεια τη μεθοδολογία της μερικής οδήγησης, δηλ. να μετατρέψουμε την προηγούμενη έκφραση στη μορφή $P * A = L * U$, στην οποία αριστερά από το A έχουμε όλα τα μητρώα εναλλαγής, αφού ως γνωστό το $P_n * P_{n-1} * \dots * P_2 * P_1 = P$, οπότε: $P_2 * L_1 * P_2^T * P_2 * P_1 * A = U \Rightarrow P_2 * P_1 * A = (P_2 * L_1 * P_2^T)^{-1} * U \Rightarrow P * A = L *$

U, όπου $P = P_2 * P_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} * \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$ και στη συνέχεια το κάτω τριγωνικό μητρώο $L =$

$$(P_2 * L_1 * P_2^T)^{-1} = (P_2^T)^{-1} * L_1^{-1} * P_2^{-1} = P_2 * L_1^{-1} * P_2^T = P_2 * L_1^{-1} * P_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} * \begin{bmatrix} 1 & 0 & 0 \\ 1/2 & 1 & 0 \\ -1/2 & 0 & 1 \end{bmatrix} * \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 0 & 0 \\ -1/2 & 1 & 0 \\ 1/2 & 0 & 1 \end{bmatrix} * \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ -1/2 & 0 & 1 \\ 1/2 & 1 & 0 \end{bmatrix}. \text{ Θα πρέπει να σημειωθεί ότι στον τελευταίο πολλαπλασιασμό}$$

μητρώων κάναμε εναλλαγή 2^{ης} και 3^{ης} στήλης του μητρώου $P_2 * L_1^{-1}$. Επομένως, οι σχετικοί όροι της PLU πα-

$$\text{ραγοντοποίησης είναι οι παράγοντες } P, L, U \text{ όπου: } P = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}, L = \begin{bmatrix} 1 & 0 & 0 \\ -1/2 & 0 & 1 \\ 1/2 & 1 & 0 \end{bmatrix}, U = \begin{bmatrix} 2 & 12 & 5 \\ 0 & 3 & -5/2 \\ 0 & 0 & -1/2 \end{bmatrix}.$$

2.5.8 Άσκηση με LU παραγοντοποίηση με μερική οδήγηση

Δίνεται το μητρώο: $A = \begin{bmatrix} 0 & -3 & 1 \\ -3 & -1 & 1 \\ -2 & -1 & -1 \end{bmatrix}$. Να υπολογίσετε και να γράψετε παρακάτω: 1) τα μητρώα που προκύ-

πτουν από τη γραμμή εντολών MATLAB [L, U, P] = lu(A); det(A). 2) Την επιπλέον τιμή που εκτυπώνεται.

Λύση

1) Από τη μορφή των εντολών που δίνονται στη MATLAB, καταλαβαίνουμε ότι πρόκειται για μερική οδήγηση. Επομένως, στο πρώτο βήμα θα χρησιμοποιήσουμε αρχικά το μητρώο εναλλαγής $P_1 = [e_2, e_1, e_3]$ για να

$$\text{εναλλάξουμε } 1^{\text{η}} \text{ και } 2^{\text{η}} \text{ γραμμή του μητρώου } A \text{ και θα έχουμε } P_1 * A = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} * \begin{bmatrix} 0 & -3 & 1 \\ -3 & -1 & 1 \\ -2 & -1 & -1 \end{bmatrix} =$$

$$\begin{bmatrix} -3 & -1 & 1 \\ 0 & -3 & 1 \\ -2 & -1 & -1 \end{bmatrix} \text{ και στη συνέχεια το γινόμενο } L_1 * P_1 * A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -2/3 & 0 & 1 \end{bmatrix} * \begin{bmatrix} -3 & -1 & 1 \\ 0 & -3 & 1 \\ -2 & -1 & -1 \end{bmatrix} = \begin{bmatrix} -3 & -1 & 1 \\ 0 & -3 & 1 \\ 0 & -1/3 & -5/3 \end{bmatrix}.$$

Στο δεύτερο βήμα δεν χρειάζεται να κάνουμε εναλλαγή γραμμών, απλώς να μηδενίσουμε τα στοιχεία της 2^{ης}

$$\text{στήλης κάτω από τον οδηγό, επομένως έχουμε ότι: } L_2 * (L_1 * P_1 * A) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -1/9 & 1 \end{bmatrix} * \begin{bmatrix} -3 & -1 & 1 \\ 0 & -3 & 1 \\ 0 & -1/3 & -5/3 \end{bmatrix} =$$

$$\begin{bmatrix} -3 & -1 & 1 \\ 0 & -3 & 1 \\ 0 & 0 & -16/9 \end{bmatrix} = U. \text{ Επομένως έχουμε ότι } L_2 * L_1 * P_1 * A = U \text{ και θα πρέπει να μεταφέρουμε αυτό που}$$

βρήκαμε σε μορφή μερικής οδήγησης, οπότε $P_1 * A = (L_2 * L_1)^{-1} * U \Rightarrow P_1 * A = L * U = L_1^{-1} * L_2^{-1} * U =$

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 2/3 & 0 & 1 \end{bmatrix} * \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1/9 & 1 \end{bmatrix} * U \Rightarrow P_1 * A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 2/3 & 1/9 & 1 \end{bmatrix} * U \Rightarrow P_1 * A = L * U. \text{ Άρα } L = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 2/3 & 1/9 & 1 \end{bmatrix} \text{ και το } U$$

$$= \begin{bmatrix} -3 & -1 & 1 \\ 0 & -3 & 1 \\ 0 & 0 & -16/9 \end{bmatrix} \text{ και το } P = P_1 = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}. \text{ Επειδή στο συγκεκριμένο υπολογισμό του κάτω τριγωνικού}$$

μητρώου L υπήρχαν μόνο μητρώα απαλοιφής Gauss, μπορούμε να μην κάνουμε την πράξη, αλλά απλώς να τοποθετήσουμε τα στοιχεία απαλοιφής από κάθε επιμέρους μητρώο στο τελικό μητρώο L στις ίδιες θέσεις.

Επειδή $P_1 * A = L * U \Rightarrow \det(P_1 * A) = \det(L * U) \Rightarrow \det(P_1) * \det(A) = \det(L) * \det(U) \Rightarrow (-1) * \det(A) = 1 * (-16)$

$\Rightarrow \det(A) = 16$. Αν επαληθεύσουμε τα μητρώα που υπολογίσαμε στη MATLAB, θα έχουμε ότι:

```
>> A = [0 -3 1;-3 -1 1;-2 -1 -1]
```

A =

$$\begin{bmatrix} 0 & -3 & 1 \\ -3 & -1 & 1 \\ -2 & -1 & -1 \end{bmatrix}$$

```
>> [L, U, P] = lu(A)
```

L =

$$\begin{bmatrix} 1.0000 & 0 & 0 \\ 0 & 1.0000 & 0 \\ 0.6667 & 0.1111 & 1.0000 \end{bmatrix}$$

U =

$$\begin{bmatrix} -3.0000 & -1.0000 & 1.0000 \\ 0 & -3.0000 & 1.0000 \\ 0 & 0 & -1.7778 \end{bmatrix}$$

P =

$$\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

```
>> det(A)
```

ans =

$$16$$

2) Η επιπλέον τιμή που εκτυπώνεται, όταν πληκτρολογήσουμε στη γραμμή εντολών της MATLAB τις εντολές $[L, U, P] = lu(A)$; $\det(A)$, είναι η τιμή της ορίζουσας που είναι ίση με 16. Λόγω του «;» που υπάρχει στο τέλος της εντολής $[L, U, P] = lu(A)$; θα εμφανιστεί **μόνο** η ορίζουσα του μητρώου A.

2.5.9 Άσκηση με LU παραγοντοποίηση με μερική οδήγηση και παραγοντοποίηση Cholesky

Δίνονται τα παρακάτω μητρώα:

$$A = \begin{bmatrix} 2 & -1 & 3 \\ 4 & 4 & 1 \\ -6 & -1 & 2 \end{bmatrix} B = \begin{bmatrix} 1 & c & c \\ c & 1 & c \\ c & c & 1 \end{bmatrix} C = \begin{bmatrix} 4 & 1 & 0 & 0 \\ 1 & 5 & -1 & 0 \\ 0 & -1 & 3 & 1 \\ 0 & 0 & 1 & 2 \end{bmatrix}$$

i) Να δοθεί η LU παραγοντοποίηση του A **με μερική οδήγηση**.

ii) Μέσω της LU που βρήκατε στο (i) να λυθεί το σύστημα $A * x = b$ όπου $b = [4, 7, -5]^T$.

iii) Μπορείτε με βάση την LU παραγοντοποίηση του A (και μόνο) να υπολογίσετε την ορίζουσα αυτού ή τον αντίστροφό του;

iv) Για το μητρώο B, πότε είναι εφικτή η παραγοντοποίηση Cholesky; Να βρεθεί αυτή όταν $c = 0.25$.

v) Αν ενδιαφερόσασταν να επιλύσετε ένα γραμμικό σύστημα με μητρώο συντελεστών το C που δίνεται παραπάνω, ποια **άμεση μέθοδο επίλυσης** θα προτιμούσατε και γιατί; Κάντε τους σχετικούς υπολογισμούς.

Λύση

i) Για τον υπολογισμό της PLU παραγοντοποίησης, πρέπει αρχικά να γίνει εναλλαγή της 1^{ης} με την 3^η γραμμή ώστε το μεγαλύτερο κατ' απόλυτη τιμή στοιχείο της 1^{ης} στήλης να βρίσκεται στην κύρια διαγώνιο. Το μητρώο εναλλαγής είναι το: $P_1 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$ και έχουμε: $P_1 * A = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} * \begin{bmatrix} 2 & -1 & 3 \\ 4 & 4 & 1 \\ -6 & -1 & 2 \end{bmatrix} = \begin{bmatrix} -6 & -1 & 2 \\ 4 & 4 & 1 \\ 2 & -1 & 3 \end{bmatrix}$. Στη συνέ-

χεια έχουμε ότι: $L_1 * P_1 * A = \begin{bmatrix} 1 & 0 & 0 \\ \frac{2}{3} & 1 & 0 \\ \frac{1}{3} & 0 & 1 \end{bmatrix} * \begin{bmatrix} -6 & -1 & 2 \\ 4 & 4 & 1 \\ 2 & -1 & 3 \end{bmatrix} = \begin{bmatrix} -6 & -1 & 2 \\ 0 & \frac{10}{3} & \frac{7}{3} \\ 0 & -\frac{4}{3} & \frac{11}{3} \end{bmatrix}$. Έπειτα, δεν θα πρέπει να γίνει κάποια εναλλαγή, αφού το 10/3 είναι ήδη στη θέση του οδηγού και θα πρέπει να μηδενίσουμε το τελευταίο

στοιχείο της 2^{ης} στήλης: Άρα $L_2 * L_1 * P_1 * A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & \frac{4}{10} & 1 \end{bmatrix} * \begin{bmatrix} -6 & -1 & 2 \\ 0 & \frac{10}{3} & \frac{7}{3} \\ 0 & -\frac{4}{3} & \frac{11}{3} \end{bmatrix} = \begin{bmatrix} -6 & -1 & 2 \\ 0 & \frac{10}{3} & \frac{7}{3} \\ 0 & 0 & \frac{69}{15} \end{bmatrix} = U$. Επομένως, α-

πομένει να βρούμε μόνο το κάτω τριγωνικό μητρώο L, ως εξής: $L_2 * L_1 * P_1 * A = U \Rightarrow P_1 * A = (L_2 * L_1)^{-1} *$

$U \Rightarrow P_1 * A = L_1^{-1} * L_2^{-1} * U$ και το $L = L_1^{-1} * L_2^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ -\frac{2}{3} & 1 & 0 \\ -\frac{1}{3} & 0 & 1 \end{bmatrix} * \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -\frac{4}{10} & 1 \end{bmatrix} \Rightarrow L = \begin{bmatrix} 1 & 0 & 0 \\ -\frac{2}{3} & 1 & 0 \\ -\frac{1}{3} & -\frac{4}{10} & 1 \end{bmatrix}$.

Παρατήρηση: Αν εκτελέσουμε στη MATLAB την PLU παραγοντοποίηση, θα επαληθεύσουμε τα αποτελέσματα:

```
>> [L, U, P] = lu(A)
```

```
L =
 1.0000   0   0
 -0.6667  1.0000   0
 -0.3333 -0.4000  1.0000
U =
 -6.0000 -1.0000  2.0000
  0  3.3333  2.3333
  0   0  4.6000
P =
 0   0   1
 0   1   0
 1   0   0
```

ii) Κάθε άμεση μέθοδος, όταν εφαρμοστεί σε πλήρη ανάπτυξη (δηλαδή με όλα της τα βήματα) για την εύρεση της λύσης x του γραμμικού συστήματος A * x = b, αποτελείται από τρία βήματα, όπως φαίνεται παρακάτω:

a) LU παραγοντοποίηση

$A = L * U$ (διάσπαση ή παραγοντοποίηση)

$L * y = b \Rightarrow y = L^{-1} * b = L \setminus b$ (εμπρός αντικατάσταση)

$U * x = y \Rightarrow x = U^{-1} * y = U \setminus y = U \setminus (L \setminus b)$ (πίσω αντικατάσταση)

b) PLU παραγοντοποίηση

$P * A = L * U$ (διάσπαση ή παραγοντοποίηση)

$L * y = \hat{b} = P * b \Rightarrow y = L^{-1} * \hat{b} = L \setminus (P * b)$ (εμπρός αντικατάσταση)

$U * x = y \Rightarrow x = U^{-1} * y = U \setminus y = U \setminus (L \setminus \hat{b})$ (πίσω αντικατάσταση)

c) Cholesky παραγοντοποίηση

$A = L * L^T$ (διάσπαση ή παραγοντοποίηση)

$$L * y = b \Rightarrow y = L^{-1} * b = L \setminus b \text{ (εμπρός αντικατάσταση)}$$

$$L^T * x = y \Rightarrow x = (L^T)^{-1} * y = L^T \setminus y = L^T \setminus (L \setminus b) \text{ (πίσω αντικατάσταση)}$$

d) QR παραγοντοποίηση

$A = Q * R$ (διάσπαση ή παραγοντοποίηση)

$x = \dots$

$$\text{Στην προκειμένη περίπτωση το διάνυσμα } b = \begin{bmatrix} 4 \\ 7 \\ -5 \end{bmatrix} = [4, 7, -5]^T \text{ και το } \hat{b} = P * b = [e_3, e_2, e_1] * b = \begin{bmatrix} -5 \\ 7 \\ 4 \end{bmatrix} \text{ και για}$$

να λυθεί το σύστημα $A * x = b$ εφαρμόζουμε καταρχήν εμπρός αντικατάσταση για τον υπολογισμό του διανύ-

$$\text{σματος } y: L * y = \hat{b} \Rightarrow \begin{bmatrix} 1 & 0 & 0 \\ -\frac{2}{3} & 1 & 0 \\ -\frac{1}{3} & -\frac{4}{10} & 1 \end{bmatrix} * \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} -5 \\ 7 \\ 4 \end{bmatrix} \Rightarrow 1 * y_1 + 0 * y_2 + 0 * y_3 = -5 \Rightarrow y_1 = -5, y_2 = 11/3 \text{ και } y_3 =$$

$$57/15. \text{ Έπειτα, εφαρμόζουμε πίσω αντικατάσταση: } U * x = y \Rightarrow \begin{bmatrix} -6 & -1 & 2 \\ 0 & \frac{10}{3} & \frac{7}{3} \\ 0 & 0 & \frac{69}{15} \end{bmatrix} * \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -5 \\ \frac{11}{3} \\ \frac{57}{15} \end{bmatrix}, \text{ οπότε υπολογί-}$$

ζουμε από πράξεις το διάνυσμα x . Στην περίπτωση ξεκινάμε από την τελευταία γραμμή του μητρώου (επειδή έχει τα περισσότερα μηδενικά) και πηγαίνουμε προς την πρώτη και έχουμε $\frac{69}{15} * x_3 = \frac{57}{15} \Rightarrow x_3 = \frac{57}{69}$, στη συνέχεια έχουμε $\frac{10}{3} * x_2 + \frac{7}{3} * x_3 = \frac{11}{3} \Rightarrow x_2 = \frac{11}{3} - \frac{10}{3} * \frac{57}{69} = \dots$ και ομοίως υπολογίζουμε και το x_1 .

iii) Από την PLU παραγοντοποίηση μπορούμε να υπολογίσουμε τόσο την **ορίζουσα** του μητρώου A , όσο και το **αντίστροφο** μητρώο A^{-1} . Από το ερώτημα (i) έχουμε $P_1 * A = L * U \Rightarrow \det(P_1 * A) = \det(L * U) \Rightarrow \det(P_1) * \det(A) = \det(L) * \det(U) \Rightarrow (-1) * \det(A) = 1 * (-6) * \frac{10}{3} * \frac{69}{15} \Rightarrow \det(A) = \frac{60}{3} * \frac{69}{15} \Rightarrow \det(A) = 92$.

Στη συνέχεια θα υπολογίσουμε το **αντίστροφο** του μητρώου A . Υποθέτουμε ότι x_i είναι η i -οστή στήλη του A^{-1} , οπότε ισχύει ότι: $A * x_i = e_i$, όπου το τελευταίο είναι το **i - στο** διάνυσμα της ορθοκανονικής βάσης. Άρα, για το πρόβλημά μας, αρκεί να λύσουμε τα τρία συστήματα $L * U * x_i = P * e_i$, $i = 1, 2, 3$. Για παράδειγμα, για την **πρώτη στήλη** του αντίστροφου μητρώου A^{-1} έχουμε τον εξής υπολογισμό:

$$L * U * x_1 = P * e_1 \Rightarrow \begin{bmatrix} 1 & 0 & 0 \\ -\frac{2}{3} & 1 & 0 \\ -\frac{1}{3} & -\frac{4}{10} & 1 \end{bmatrix} * \begin{bmatrix} -6 & -1 & 2 \\ 0 & \frac{10}{3} & \frac{7}{3} \\ 0 & 0 & \frac{69}{15} \end{bmatrix} * \begin{bmatrix} x_{11} \\ x_{12} \\ x_{13} \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} * \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \Rightarrow \begin{bmatrix} -6 & -1 & 2 \\ 4 & 4 & 1 \\ 2 & -1 & 3 \end{bmatrix} * \begin{bmatrix} x_{11} \\ x_{12} \\ x_{13} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

$\Rightarrow -6 * x_{11} - x_{12} + 2 * x_{13} = 0 \quad (1) \text{ και } 4 * x_{11} + 4 * x_{12} + * x_{13} = 0 \quad (2) \text{ και } 2 * x_{11} - x_{12} + 3 * x_{13} = 1 \quad (3)$. Από την επίλυση του συστήματος των εξισώσεων (1), (2), (3) βρίσκουμε την πρώτη στήλη του αντίστροφου μητρώου A^{-1} , που είναι $x_{11} = 9/92, x_{12} = -7/46, x_{13} = 5/23$. Τελικά, επιλύοντας και τις άλλες δύο εξισώσεις με ανάλογο τρόπο, υπολογίζουμε

$$\text{το μητρώο: } A^{-1} = \begin{bmatrix} 9/92 & -1/92 & -13/92 \\ -7/46 & 11/46 & 5/46 \\ 5/23 & 2/23 & 3/232 \end{bmatrix}$$

iv) Για να είναι εφικτή η παραγοντοποίηση Cholesky πρέπει το μητρώο B να είναι συμμετρικό (κάτι που ισχύει), να έχει θετικά διαγώνια στοιχεία (κάτι που ισχύει) και επίσης Α.Δ.Κ. (για να ισχύει το τελευταίο πρέπει

να ισχύει ότι το $1 > |c| + |c|$. Έχουμε ότι το μητρώο $B = \begin{bmatrix} 1 & 0.25 & 0.25 \\ 0.25 & 1 & 0.25 \\ 0.25 & 0.25 & 1 \end{bmatrix}$. Προκειμένου να είναι εφικτή η παραγοντοποίηση Cholesky, θα πρέπει το μητρώο B να είναι Συμμετρικά Θετικά Ορισμένο (ΣΘΟ). Δηλαδή πρέπει να είναι συμμετρικό, να έχει Αυστηρή Διαγώνια Κυριαρχία (ΑΔΚ) και θετικά διαγώνια στοιχεία. Όλες οι

παραπάνω απαιτήσεις ικανοποιούνται από το μητρώο B , άρα ισχύει ότι: $B = L * L^T \Rightarrow \begin{bmatrix} 1 & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & 1 & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} & 1 \end{bmatrix} =$

$$\begin{bmatrix} l_{11} & 0 & 0 \\ l_{21} & l_{22} & 0 \\ l_{31} & l_{32} & l_{33} \end{bmatrix} * \begin{bmatrix} l_{11} & l_{21} & l_{31} \\ 0 & l_{22} & l_{32} \\ 0 & 0 & l_{33} \end{bmatrix} = \begin{bmatrix} l_{11}^2 & l_{11} * l_{21} & l_{11} * l_{31} \\ l_{21} * l_{11} & l_{21}^2 + l_{22}^2 & l_{21} * l_{13} + l_{22} * l_{23} \\ l_{31} * l_{11} & l_{31} * l_{12} + l_{32} * l_{22} & l_{21}^2 + l_{22}^2 + l_{33}^2 \end{bmatrix}$$

$$\begin{aligned} l_{11}^2 = 1 \rightarrow l_{11} = 1 & \text{ (κρατάμε τη θετική ρίζα) και } l_{11} * l_{21} = \frac{1}{4} \rightarrow l_{21} = \frac{1}{4} \text{ και } l_{11} * l_{31} = \frac{1}{4} \rightarrow l_{31} = \frac{1}{4} \text{ και } l_{11} * l_{21} = \frac{1}{4} \rightarrow \\ l_{21} = \frac{1}{4} \text{ και } l_{11} * l_{31} = \frac{1}{4} \rightarrow l_{31} = \frac{1}{4} & \text{ και } l_{21} * l_{12} + l_{22}^2 = 1 \rightarrow l_{22}^2 = 1 - \frac{1}{16} \rightarrow l_{22} = \frac{\sqrt{15}}{4} \text{ και } l_{21} * l_{13} + l_{22} * l_{23} = \frac{1}{4} \rightarrow \\ l_{23} = \frac{3}{4*\sqrt{5}} \text{ και } l_{31} * l_{12} + l_{32} * l_{22} = \frac{1}{4} \rightarrow l_{32} = \frac{3}{4*\sqrt{5}} & \text{ και } l_{13} * l_{31} + l_{32} * l_{23} + l_{33}^2 = 1 \rightarrow l_{33}^2 = 1 - \frac{1}{16} - \frac{9}{16*\sqrt{15}} \rightarrow l_{33}^2 = \frac{15}{16} - \frac{9}{16*\sqrt{15}} \rightarrow l_{33} = \frac{15}{16} - \frac{9}{16*\sqrt{15}} \rightarrow l_{33} = 0.948683. \text{ Επομένως ο παράγοντας Cholesky είναι το μητρώο } L = \begin{bmatrix} 1 & 0 & 0 \\ \frac{1}{4} & \frac{\sqrt{15}}{4} & 0 \\ \frac{1}{4} & \frac{3}{4*\sqrt{5}} & 0.948683 \end{bmatrix} \end{aligned}$$

Παρατήρηση 1: Εναλλακτικά μπορούμε να αναφέρουμε ότι το μητρώο B είναι συμμετρικό. Για να μπορεί να παραγοντοποιηθεί κατά Cholesky θα πρέπει λοιπόν να είναι και **Θετικά ορισμένο**. Εξετάζουμε τις ιδιοτιμές. Με έναν απλό υπολογισμό βρίσκουμε ότι αυτές είναι $1-c$, $1-c$, $1+2c$. Θα πρέπει να είναι όλες θετικές, επομένως πρέπει $c \in (-0.5, 1)$. (Αυτό επαληθεύεται και με δοκιμές για διάφορες τιμές του c).

Παρατήρηση 2: Εναλλακτικά μπορούμε να βρούμε φράγματα για το c από το Θεώρημα Δίσκων Gershgorin. Όμως έτσι βρίσκουμε μόνον ότι $c > -0.5$ (...). Επίσης, από τον ορισμό των θετικά ορισμένων μητρώων, αν θέσουμε $x = [1, 0, -1]^T$, παίρνουμε $x^T * B * x = ... = 2(1 - c) \Rightarrow c < 1$.

v) **Το μητρώο C είναι συμμετρικό.** Επίσης είναι τριδιαγώνιο, με θετικά στοιχεία στη διαγώνιο, και έχει Α.Δ.Κ. Έτσι μπορεί να επαληθευτεί ότι είναι Σ.Θ.Ο., επομένως η προτιμότερη παραγοντοποίηση είναι η Cholesky $C = L * L^T$ της οποίας το κόστος $\Omega = \frac{n^3}{3} + 2n^2$.

2.5.10 Άσκηση με παραγοντοποίηση Cholesky

Έστω $A = I + uu^T$, όπου $A \in R^{n \times n}$ και u ένα διάνυσμα ώστε $\|u\|_2 = 1$.

- i) Να αποδείξετε ότι το μητρώο A είναι **συμμετρικό** και **θετικά ορισμένο**
- ii) Στην περίπτωση όπου $n = 3$, βρείτε την παραγοντοποίηση Cholesky του A .

Λύση

i) Γνωρίζουμε ότι $\|u\|_2 = 1$ από το οποίο μπορούμε να συμπεράνουμε ότι το διάνυσμα u έχει ένα στοιχείο του ίσο με το «1» και όλα τα υπόλοιπα είναι «0». Έστω λοιπόν ότι ο “1” είναι το 2^o στοιχείο του διανύσματος u,

$$\text{δηλαδή} \quad \text{έχουμε} \quad \text{ότι} \quad u = \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}. \quad \text{Άρα} \quad A = I + uu^T \Rightarrow A = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} * [0 \ 1 \ 0 \ \cdots \ 0] =$$

$$\begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix} \Rightarrow A = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 2 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix}$$

το οποίο είναι συμμετρικό και θετικά ορισμένο.

$$\text{ii) Για } n = 3 \text{ έχουμε: } A = I + uu^T \Rightarrow A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} * [0 \ 1 \ 0] = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \Rightarrow A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

$$\text{Επομένως εφαρμόζοντας παραγοντοποίηση Cholesky έχουμε ότι: } A = L * L^T \Rightarrow \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} l_{11} & 0 & 0 \\ l_{21} & l_{22} & 0 \\ l_{31} & l_{32} & l_{33} \end{bmatrix} *$$

$$\begin{bmatrix} l_{11} & l_{21} & l_{31} \\ 0 & l_{22} & l_{32} \\ 0 & 0 & l_{33} \end{bmatrix} \Rightarrow \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} l_{11}^2 & l_{11} * l_{12} & l_{11} * l_{13} \\ l_{21} * l_{11} & l_{21} * l_{12} + l_{22}^2 & l_{21} * l_{13} + l_{22} * l_{23} \\ l_{31} * l_{11} & l_{31} * l_{12} + l_{32} * l_{22} & l_{31} * l_{13} + l_{32} * l_{23} + l_{33}^2 \end{bmatrix}$$

- $l_{11}^2 = 1 \rightarrow l_{11} = \pm 1, l_{11} * l_{12} = 0 \rightarrow l_{12} = 0, l_{11} * l_{13} = 0 \rightarrow l_{13} = 0, l_{11} * l_{21} = 0 \rightarrow l_{21} = 0, l_{11} * l_{31} = 0 \rightarrow l_{31} = 0$
- $l_{21} * l_{12} + l_{22}^2 = 2 \rightarrow l_{22}^2 = 2 \rightarrow l_{22} = \pm\sqrt{2}, l_{21} * l_{13} + l_{22} * l_{23} = 0 \rightarrow l_{23} = 0, l_{31} * l_{12} + l_{32} * l_{22} = 0 \rightarrow l_{32} = 0$
- $l_{13} * l_{31} + l_{32} * l_{23} + l_{33}^2 = 1 \rightarrow l_{33}^2 = 1 \rightarrow l_{33} = \pm 1$. Επομένως το $L = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \sqrt{2} & 0 \\ 0 & 0 & 1 \end{bmatrix}$

Παρατήρηση: Ισχύει ότι το εσωτερικό γινόμενο ενός διανύσματος με τον εαυτό του ισούται με το μήκος του διανύσματος (ή αλλιώς τη δεύτερη νόρμα του διανύσματος) υψωμένο στο τετράγωνο, δηλαδή $x^T * x = x_1^2 + x_2^2 + \cdots + x_n^2 = \|x\|_2^2$, ενώ κανονικά ισχύει ότι $\|x\|_2 = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2} = \sqrt{x^T * x}$.

Εναλλακτικός τρόπος επίλυσης

i) Από τις ιδιότητες του συμμετρικού μητρώου έχουμε ότι $A^T = A$. Πράγματι ισχύει ότι $A = (I + u^* u^T)$ οπότε το $A^T = (I + u^* u^T)^T = I^T + (u^* u^T)^T = I + (u^T)^T * u^T = I + u * u^T = A$. Αποδεικνύουμε τώρα ότι, με βάση τον ορισμό, το A είναι και **θετικά ορισμένο**. Έστω x ένα τυχαίο, μη μηδενικό, διάνυσμα-στήλη διάστασης n, δηλ. $x \in R^n$ με $x \neq 0$. Τότε έχουμε ότι: $x^T * A * x = x^T * (I + u * u^T) * x = x^T * I * x + x^T * u * u^T * x = x^T * x + (u^T * x)^T * u^T * x = \|x\|_2^2 + \|u^T * x\|_2^2 > 0$. Η τελευταία ανίσωση **ισχύει**, διότι το x πρέπει να είναι μη-μηδενικό, διότι από τον ορισμό του Σ.Θ.Ο. ισχύει ότι $\forall x \in R^n$, με $x \neq 0$, $x^T * A * x > 0$. Συνεπώς, το μητρώο A είναι Σ.Θ.Ο. Η συνθήκη $\|u\|_2$ είναι **πλεονάζουσα** για αυτό το ερώτημα. Χρησιμοποιείται όμως στο επόμενο.

(ii) Έστω ότι $u = [a, b, c]^T$ και **L** το κάτω τριγωνικό μητρώο που προκύπτει από την παραγοντοποίηση Cholesky

$$\text{του A. Τότε: } A = I + u * u^T = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} + \begin{bmatrix} a \\ b \\ c \end{bmatrix} * [a \ b \ c] = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} + \begin{bmatrix} a^2 & ab & ac \\ ba & b^2 & bc \\ ca & cb & c^2 \end{bmatrix} = \begin{bmatrix} 1 + a^2 & ab & ac \\ ba & 1 + b^2 & bc \\ ca & cb & 1 + c^2 \end{bmatrix}$$

συνέχεια, αφού αποδείξαμε ότι το μητρώο A είναι ΣΘΟ, κάνουμε παραγοντοποίηση Cholesky, άρα $A = L * L^T$

$$\Rightarrow \begin{bmatrix} 1+a^2 & ab & ac \\ ba & 1+b^2 & bc \\ ca & cb & 1+c^2 \end{bmatrix} = \begin{bmatrix} l_{11} & 0 & 0 \\ l_{21} & l_{22} & 0 \\ l_{31} & l_{32} & l_{33} \end{bmatrix} * \begin{bmatrix} l_{11} & l_{21} & l_{31} \\ 0 & l_{22} & l_{32} \\ 0 & 0 & l_{33} \end{bmatrix} = \begin{bmatrix} l_{11}^2 & l_{11} * l_{12} & l_{11} * l_{13} \\ l_{21} * l_{11} & l_{21} * l_{12} + l_{22}^2 & l_{21} * l_{13} + l_{22} * l_{23} \\ l_{31} * l_{11} & l_{31} * l_{12} + l_{32} * l_{22} & l_{31} * l_{13} + l_{32} * l_{23} + l_{33}^2 \end{bmatrix}.$$

Λύνοντας διαδοχικά τις αντιστοιχίες και χρησιμοποιώντας τη σχέση $a^2 + b^2 + c^2 = 1$, βρίσκουμε ότι: $r_1 =$

$$\sqrt{1+a^2}, r_2 = \frac{ab}{\sqrt{1+a^2}}, r_3 = \sqrt{\frac{2-c^2}{1+a^2}}, r_4 = \frac{ac}{\sqrt{1+a^2}}, r_5 = \frac{bc}{\sqrt{1+a^2}\sqrt{2-c^2}}, r_6 = \sqrt{\frac{2}{2-c^2}}.$$

2.5.11 Ασκήσεις με παραγοντοποίηση Cholesky και θεώρημα Sylvester

Ένα άλλο ενδιαφέρον και σημαντικό ζήτημα είναι ότι σε ορισμένες περιπτώσεις μπορούμε να αποφύγουμε την οδήγηση, ενώ μπορεί να υπάρχουν πιο εξειδικευμένες και οικονομικότερες παραγοντοποίησεις, όπως η Cholesky. Όταν το μητρώο είναι αντιστρέψιμο, η ύπαρξη της παραγοντοποίησης Cholesky: $A = L * L^T$ όπου L κάτω τριγωνικό με θετικά στοιχεία στη διαγώνιο (σε αριθμητική άπειρης ακρίβειας) ισοδυναμεί με την ιδιότητα Σ.Θ.Ο.

Άσκηση 1: Δίνεται το μητρώο $A = \mu * I - u * u^T$ όπου u ήταν διάνυσμα (στήλη), π.χ. $u = [1, -2, 1]^T$ και μ είναι κάποια σταθερά, π.χ. $\mu = 7$. Υπάρχει κάτω τριγωνικό μητρώο L τέτοιο ώστε $A = L * L^T$; Να υπολογίσετε το κατάλληλο L αν Ναι, αν Όχι να εξηγήσετε γιατί δεν υπάρχει.

Σημειώνουμε κατ' αρχήν ότι είναι εύκολο να ελεγχθεί αν τα (συμμετρικά) μητρώα αυτού του τύπου είναι αντιστρέψιμα – κατ' εξαίρεση η ορίζουσα υπολογίζεται εύκολα από το θεώρημα Sylvester- και είναι $\det(A) = \mu^3 * (1 - \frac{1}{\mu} * u^T * u)$. Το θεώρημα αυτό εφαρμόζεται μόνο σε μητρώα συγκεκριμένης μορφής, όπως το παράπάνω μητρώο A . Εφόσον η τιμή της δεν είναι 0, το μητρώο είναι αντιστρέψιμο, και αν υπάρχει το μητρώο L θα είναι ο παράγοντας Cholesky του μητρώου A και η ύπαρξή του θα ισοδυναμεί με τη θετική ορισμότητα του A .

Ο κατάλληλος τρόπος είναι να δοκιμάσουμε εφαρμογή της παραγοντοποίησης Cholesky και αν μπορούμε να την υπολογίσουμε τότε το μητρώο είναι Σ.Θ.Ο, αν όχι τότε δεν είναι. Επομένως, δεν χρειάζεται να διερευνήσουμε χωριστά τη ΣΘΟ. Βέβαια, υπάρχουν περιπτώσεις που το μητρώο φαίνεται άμεσα ότι δεν είναι Σ.Θ.Ο. π.χ. αν έχει μηδενικά ή αρνητικά στοιχεία στη διαγώνιο, ή αν μπορούμε εύκολα να μαντέψουμε κάποιο x τέτοιο ώστε $x^T * A * x \leq 0$ ή αν (όπως στην περίπτωσή μας) η ορίζουσα υπολογίζεται με εύκολο τρόπο και είναι αρνητική. Από την άλλη, γενικά δεν συνίσταται να βασιστούμε στον υπολογισμό των οριζουσών όλων των κυρίαρχων $1x1, 2x2, 3x3$ κ.λ.π. υπομητρώων, εκτός βέβαια αν το μητρώο και τα υπομητρώα έχουν ειδική μορφή που να γνωρίζουμε πως να την αξιοποιήσουμε. Ο λόγος είναι ότι ακόμα και αν αποδείξουμε ότι όλες οι ορίζουσες είναι θετικές, δεν θα γλυτώσουμε τον υπολογισμό του L . Επομένως, ο αναμενόμενος κόπος θα είναι μεγαλύτερος από το να εφαρμόσουμε απευθείας την παραγοντοποίηση Cholesky.

ΜΕΘΟΔΟΛΟΓΙΑ ΕΦΑΡΜΟΓΗΣ ΠΑΡΑΓΟΝΤΟΠΟΙΗΣΗΣ CHOLESKY ΓΙΑ ΤΗΝ ΕΙΔΙΚΗ ΠΕΡΙΠΤΩΣΗ ΤΟΥ Α

Για την **ειδική περίπτωση που εξετάζουμε**, εύκολος υπολογισμός ορίζουσας με το θεώρημα Sylvester, η ενδεδειγμένη μεθοδολογία είναι η εξής: 1) Υπολογισμός **ορίζουσας**, αξιοποιώντας τη μορφή του μητρώου (δηλ. με θεώρημα Sylvester). 2) Αν η ορίζουσα είναι **θετική**, τότε εφαρμόζουμε παραγοντοποίηση Cholesky και **αν ευστοχήσει** (δηλαδή αν δεν οδηγηθούμε στην ανάγκη εύρεσης ρίζας αρνητικού αριθμού), το αποτέλεσμα θα είναι ο ζητούμενος παράγοντας L. Αν αποτύχει, τότε το μητρώο **δεν** είναι ΣΘΟ και δεν υπάρχει το ζητούμενο L. 3) Αν η ορίζουσα είναι **αρνητική**, τότε το μητρώο **δεν** είναι ΣΘΟ και σταματάμε αμέσως καθώς δεν υπάρχει ο ζητούμενος παράγοντας L. 4) Αν η ορίζουσα είναι **μηδέν**, το μητρώο είναι **μη αντιστρέψιμο**. Η περίπτωση αυτή είναι πιο περίπλοκη γιατί ενδέχεται να υπάρχει το **L αν το μητρώο είναι ημιορισμένο** (δηλ. όλες οι ιδιοτιμές μη αρνητικές αλλά τουλάχιστον μία μηδενική) ή να μην υπάρχει (αόριστο μητρώο οπότε μία ή περισσότερες ιδιοτιμές είναι αρνητικές, μία ή περισσότερες είναι θετικές και μία ή περισσότερες είναι μηδενικές). Αν είναι **ημιορισμένο**, τότε υπάρχει παραγοντοποίηση $A = L * L^T$ αλλά το L θα περιέχει τουλάχιστον **μία μηδενική τιμή** στη διαγώνιο.

Λύση

Το μητρώο είναι **συμμετρικό και αντιστρέψιμο**. Επομένως, αρκεί να δοκιμάσουμε την παραγοντοποίηση Cholesky, και αν δεν αστοχήσει, τότε το μητρώο είναι ΣΘΟ και ο παράγοντας προκύπτει από τον υπολογισμό. Αν υπάρχει αστοχία και το μητρώο είναι αντιστρέψιμο, έπειτα ότι δεν υπάρχει τέτοιο L.

Αρχικά υπολογίζουμε την **ορίζουσα** του συγκεκριμένου μητρώου, το οποίο είναι της μορφής $A = \mu * I - u * u^T$ με $\mu = 7$, και για το λόγο αυτό γνωρίζουμε α) **ότι είναι συμμετρικό**, αφού $A^T = A$ και β) η ορίζουσά του υπολογίζεται με εύκολο τρόπο από τον τύπο Sylvester και είναι $\det(A) = \mu^3 * (1 - \frac{1}{\mu} * u^T * u) = 7^3 * (1 - \frac{1}{7} * [1, -2, 1] * \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}) = 7^3 * (1 - \frac{6}{7}) = 7^2 = 49 > 0$. Επειδή η ορίζουσα είναι **θετική**, εφαρμόζουμε παραγοντοποίηση Cholesky,

χωρίς να είμαστε βέβαιοι ότι θα πετύχει. Απλώς, **αν** η ορίζουσα έβγαινε **αρνητική** τότε **ήταν σίγουρο** ότι η παραγοντοποίηση Cholesky θα αποτύγχανε. Για το συγκεκριμένο μητρώο, η απάντηση είναι Ναι, καθώς από την

$$\text{εφαρμογή της Cholesky προκύπτει ότι με μητρώο } A = \mu * I - u * u^T = 7 * I - u * u^T = 7 * \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} - \begin{bmatrix} 1 & -2 & 1 \\ -2 & 4 & -2 \\ 1 & -2 & 1 \end{bmatrix}$$

$$= 7 * \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} - \begin{bmatrix} 6 & 2 & -1 \\ 2 & 3 & 2 \\ -1 & 2 & 6 \end{bmatrix} = L * L^T \Rightarrow \text{μετά από πράξεις το } L = \begin{bmatrix} \sqrt{6} & 0 & 0 \\ \frac{\sqrt{2}\sqrt{3}}{3} & \frac{\sqrt{3}\sqrt{7}}{3} & 0 \\ -\frac{\sqrt{6}}{6} & \frac{\sqrt{3}\sqrt{7}}{3} & \frac{\sqrt{2}\sqrt{7}}{2} \end{bmatrix} \text{ Από τη}$$

στιγμή που υπολογίστηκε ο παράγοντας Cholesky, σημαίνει ότι το μητρώο είναι Θ.Ο.

Άσκηση 2: Δίνεται το μητρώο $A = 2 * I - u * u^T$ όπου $u = [1, -1, 1]^T$. Μπορείτε να υπολογίσετε κάτω τριγωνικό μητρώο L, τέτοιο ώστε $A = L * L^T$;

Λύση

Το μητρώο είναι της γενικής μορφής: $A = \mu * I - u * u^T$ που αναφέρθηκε προηγουμένως και είναι προφανώς συμμετρικό, αφού το $A^T = (2 * I - u * u^T)^T = (2 * I)^T - (u * u^T)^T = (2 * I) - (u^T)^T * u^T = (2 * I) - u * u^T = A = 2 * I - u * u^T$. Είναι από το θεώρημα Sylvester και είναι: $\det(A) = \mu^3 * (1 - \frac{1}{\mu} u^T * u) = 2^3 * (1 - \frac{1}{2} u^T * u) = -4 < 0$. Εφόσον η ορίζουσα είναι **αρνητική**, το μητρώο **δεν** είναι **Σ.Θ.Ο** και **δεν** υπάρχει ο παράγοντας **L**. Είναι η τρίτη προηγούμενη περίπτωση.

Συμπέρασμα: Όταν δίνεται ένα μητρώο της μορφής $A = \mu * I - u * u^T$ τότε αυτό έχει **δύο** χαρακτηριστικά - ιδιότητες: α) είναι **συμμετρικό** και β) η ορίζουσα του υπολογίζεται **απευθείας** από τον τύπο του θεωρήματος Sylvester και είναι $\det(A) = \mu^3 * (1 - \frac{1}{\mu} u^T * u)$.

2.5.12 Άσκηση με παραγοντοποίηση Cholesky και θεώρημα κύκλων Gershgorin

Έστω ότι δινόταν ένα μικρό μητρώο και έπρεπε να διερευνήσουμε χρησιμοποιώντας το θεώρημα των κύκλων **Gershgorin και μόνον**, το κατά πόσο είναι δυνατόν να αποκλείσουμε τη δυνατότητα να υπάρχει ιδιοτιμή με μια συγκεκριμένη τιμή ή αν υπάρχει περίπτωση να υπάρχει ιδιοτιμή με τη συγκεκριμένη τιμή κ.λ.π. Μια άλλη παραλλαγή ήταν να διερευνήσουμε αν είναι δυνατόν (πάντα μέσω του θεωρήματος και μόνον) ότι το μητρώο είναι Θ.Ο. **Προσοχή:** Στο ερώτημα αυτό δεν μπορούμε να χρησιμοποιήσουμε τον υπολογισμό του προσήμου των κύριων υποοριζουσών του μητρώου μέχρις ότου προκύψει 0 ή αρνητική τιμή (οπότε αποκλείεται η ΣΘΟ) ή προκύψουν όλα θετικά. Αν το μητρώο δεν είναι (πραγματικό) συμμετρικό, θα πρέπει να υπολογίσουμε το χωρίο που προκύπτει από την **τομή** των δίσκων που κατασκευάζονται από τις στήλες του μητρώου με την ένωση των δίσκων που κατασκευάζονται από τις γραμμές. **Αν το μητρώο είναι συμμετρικό, δεν υπάρχει κάποια διαφορά**, δηλαδή οι κύκλοι **υπολογίζονται μόνο μια φορά**. Στην περίπτωση αυτή όμως, έχουμε το **επιπλέον στοιχείο** ότι οι **ιδιοτιμές είναι οπωσδήποτε πραγματικές**, οπότε μπορούμε να λύσουμε τις ανισώσεις και να απαλλαγούμε από τις απόλυτες τιμές. Στην περίπτωση αυτή (και μόνο) οι ανισώσεις θα αφορούν υποδιαστήματα στον άξονα των πραγματικών.

2.5.13 Άσκηση με θεώρημα κύκλων Gershgorin

Δίνεται το συμμετρικό μητρώο $A = \begin{bmatrix} 1.2 & -1 & 0 \\ -1 & 2.2 & -1 \\ 0 & -1 & 2 \end{bmatrix}$. Με βάση **μόνο** το θεώρημα των κύκλων Gershgorin, μπορεί να αποδειχθεί ότι το μητρώο είναι θετικά ορισμένο (χρειάζεται σύντομη εξήγηση).

Λύση

Από το θεώρημα των κύκλων Gershgorin έπειται ότι οι ιδιοτιμές του μητρώου A (ή αλλιώς το φάσμα του μητρώου A), δηλαδή το $\Lambda(A)$ θα είναι: $\Lambda(A) \subseteq (\mathbf{D}_1 \cup \mathbf{D}_2 \cup \mathbf{D}_3) \cap \mathbb{R}$ όπου $\mathbf{D}_1 = \{z, \text{όπου } |z - 1.2| \leq 1\}$, $\mathbf{D}_2 = \{z, \text{όπου } |z - 2.2| \leq 2\}$ και $\mathbf{D}_3 = \{z, \text{όπου } |z - 2| \leq 1\}$ είναι οι δίσκοι (κύκλοι) Gershgorin. Επιπλέον, οι ιδιοτιμές είναι πραγματικές λόγω συμμετρίας του μητρώου. Από τη συναλήθευση των σχέσεων αυτών προκύπτει ότι $0.2 \leq z \leq 2.2$ και $0.2 \leq z \leq 4.2$ και $1 \leq z \leq 3$, επομένως για κάθε ιδιοτιμή ισχύει ότι $z > 0.2$ (συναληθεύουν στο διάστημα $[1, 2.2]$) και έτσι όλες οι ιδιοτιμές είναι θετικές. **Τέλος γνωρίζουμε από τη θεωρία ότι ένα συμμετρικό μητρώο με θετικές ιδιοτιμές είναι Σ.Θ.Ο.** Αυτό προκύπτει από το δεύτερο κριτήριο θετικής ορισμότητας που αναφέρθηκε προηγουμένως.

2.5.14 Άσκηση με παραγοντοποίηση Cholesky και απαλοιφή Gauss (παλιό θέμα)

α) Έστω ότι χρησιμοποιείτε απαλοιφή Gauss σε κάποιο συμμετρικό μητρώο $B \in \mathbb{R}^{n \times n}$ όπου $n > 2$. Γνωρίζετε ότι είναι αντιστρέψιμο και ότι το στοιχείο στη θέση $(1, 1)$, δηλαδή το $\beta_{11} \neq 0$. Γράφουμε το μητρώο που προκύπτει μετά από το πρώτο βήμα της απαλοιφής ως $\begin{bmatrix} \beta_{11} & b_1^T \\ 0_{(n-1) \times 1} & B_2 \end{bmatrix}$, όπου B_2 είναι τετραγωνικό, μεγέθους $n-1$ και το $0_{(n-1) \times 1}$ είναι διάνυσμα με μηδενικά στοιχεία. **Να δείξετε ότι το B_2 είναι συμμετρικό.**

β) Αν $A = \begin{bmatrix} 1 & 2 & 0 & 1 \\ 2 & 5 & 1 & 2 \\ 0 & 1 & 2 & -1 \\ 1 & 2 & -1 & 3 \end{bmatrix}$ να υπολογίσετε τον κάτω τριγωνικό παράγοντα Cholesky L , τέτοιο ώστε $A = L^* L^T$.

Λύση

α) Έστω ότι το $n = 3$. Τότε το (συμμετρικό) μητρώο $B = \begin{bmatrix} \beta_{11} & \beta_{12} & \beta_{13} \\ \beta_{21} & \beta_{22} & \beta_{23} \\ \beta_{31} & \beta_{32} & \beta_{33} \end{bmatrix}$ μετά από την απαλοιφή των στοιχείων

κάτω από τον πρώτο οδηγό με απαλοιφή Gauss γίνεται μετά από πράξη $L^* B = \begin{bmatrix} 1 & 0 & 0 \\ -\frac{\beta_{21}}{\beta_{11}} & 1 & 0 \\ -\frac{\beta_{31}}{\beta_{11}} & 0 & 1 \end{bmatrix} * \begin{bmatrix} \beta_{11} & \beta_{12} & \beta_{13} \\ \beta_{21} & \beta_{22} & \beta_{23} \\ \beta_{31} & \beta_{32} & \beta_{33} \end{bmatrix} =$

$\begin{bmatrix} \beta_{11} & \beta_{12} & \beta_{13} \\ -\frac{\beta_{21}}{\beta_{11}} * \beta_{11} + \beta_{21} & -\frac{\beta_{21}}{\beta_{11}} * \beta_{12} + \beta_{22} & -\frac{\beta_{21}}{\beta_{11}} * \beta_{13} + \beta_{23} \\ -\frac{\beta_{31}}{\beta_{11}} * \beta_{11} + \beta_{31} & -\frac{\beta_{31}}{\beta_{11}} * \beta_{12} + \beta_{32} & -\frac{\beta_{31}}{\beta_{11}} * \beta_{13} + \beta_{33} \end{bmatrix} = \begin{bmatrix} \beta_{11} & \beta_{12} & \beta_{13} \\ 0 & \frac{-\beta_{21}}{\beta_{11}} * \beta_{12} + \beta_{22} & \frac{-\beta_{21}}{\beta_{11}} * \beta_{13} + \beta_{23} \\ 0 & \frac{-\beta_{31}}{\beta_{11}} * \beta_{12} + \beta_{32} & \frac{-\beta_{31}}{\beta_{11}} * \beta_{13} + \beta_{33} \end{bmatrix}$. Δηλαδή ουσιαστικά

πολλαπλασιάσαμε το αρχικό μητρώο με το μητρώο απαλοιφής Gauss $L = \begin{bmatrix} 1 & 0 & 0 \\ -\frac{\beta_{21}}{\beta_{11}} & 1 & 0 \\ -\frac{\beta_{31}}{\beta_{11}} & 0 & 1 \end{bmatrix}$. Όμως $\beta_{12} = \beta_{21} = \delta$ και $\beta_{13} = \beta_{31} = \varepsilon$ και $\beta_{23} = \beta_{32} = \zeta$, αφού το αρχικό μητρώο B είναι συμμετρικό. Άρα έχουμε:

$\begin{bmatrix} \beta_{11} & \delta & \varepsilon \\ 0 & \beta_{22} - \frac{\delta^2}{\beta_{11}} & \zeta - \frac{\delta\varepsilon}{\beta_{11}} \\ 0 & \zeta - \frac{\delta\varepsilon}{\beta_{11}} & \beta_{33} - \frac{\varepsilon^2}{\beta_{11}} \end{bmatrix}$. Παρατηρούμε ότι το μητρώο B_2 είναι συμμετρικό. Με όμοιο τρόπο, θα είναι συμμετρικό και για οποιοδήποτε n .

β) Παρατηρούμε ότι το μητρώο είναι **συμμετρικό** αλλά **δεν** έχει αυστηρή διαγώνια κυριαρχία. Στη συνέχεια θα πρέπει να εξετάσουμε αν οι ιδιοτιμές του είναι θετικές, τις οποίες όμως δεν θα υπολογίσουμε για να κάνουμε εξοικονόμηση χρόνου. **Αν το μητρώο ήταν θετικό (δηλαδή είχε όλα τα στοιχεία του θετικά), θα μπορούσαμε να συμπεράνουμε ότι έχει θετικές ιδιοτιμές, σύμφωνα με το θεώρημα Perron - Frobenius.** Συνεπώς θα εξετάσουμε αν οι οδηγοί του είναι θετικοί, μέσω της απαλοιφής Gauss, που ουσιαστικά είναι η LU χωρίς οδήγηση, και αυτό που κάνουμε είναι να μηδενίζουμε τα στοιχεία κάθε στήλης που είναι κάτω από τον οδηγό της στήλης, χωρίς καμία εναλλαγή. Αυτό σημαίνει ότι στην απαλοιφή Gauss χρησιμοποιούμε μόνο μητρώα απαλοιφής L_i . Επομένως

$$L_1 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ -2 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -1 & 0 & 0 & 1 \end{bmatrix}, L_1 A = \begin{bmatrix} 1 & 2 & 0 & 1 \\ \textcolor{red}{0} & 1 & 1 & 0 \\ \textcolor{red}{0} & \textcolor{blue}{1} & 2 & -1 \\ \textcolor{red}{0} & \textcolor{blue}{0} & -1 & 2 \end{bmatrix}, L_2 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, L_2(L_1 A) = \begin{bmatrix} 1 & 2 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & \textcolor{red}{0} & 1 & -1 \\ 0 & \textcolor{red}{0} & -1 & 2 \end{bmatrix}$$

$$L_3 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix}, L_3(L_2 L_1 A) = \begin{bmatrix} 1 & 2 & 0 & 1 \\ \textcolor{red}{0} & \textcolor{blue}{1} & 1 & 0 \\ \textcolor{red}{0} & \textcolor{red}{0} & 1 & -1 \\ \textcolor{red}{0} & \textcolor{red}{0} & 0 & 3 \end{bmatrix} = U$$

Παρατηρούμε επομένως ότι το μητρώο U στο οποίο καταλήξαμε έχει **θετικούς οδηγούς** (στην κύρια διαγώνιο), άρα είναι δυνατή η παραγοντοποίηση Cholesky, την οποία εφαρμόζουμε: $A = L * L^T \Rightarrow \begin{bmatrix} 1 & 2 & 0 & 1 \\ 2 & 5 & 1 & 2 \\ 0 & 1 & 2 & -1 \\ 1 & 2 & -1 & 3 \end{bmatrix} =$

$$\begin{bmatrix} l_{11} & 0 & 0 & 0 \\ l_{21} & l_{22} & 0 & 0 \\ l_{31} & l_{32} & l_{33} & 0 \\ l_{41} & l_{42} & l_{43} & l_{44} \end{bmatrix} \begin{bmatrix} l_{11} & l_{21} & l_{31} & l_{41} \\ 0 & l_{22} & l_{32} & l_{42} \\ 0 & 0 & l_{33} & l_{43} \\ 0 & 0 & 0 & l_{44} \end{bmatrix} = \begin{bmatrix} l_{11}^2 & l_{11} * l_{21} & l_{11} * l_{31} & l_{11} * l_{41} \\ l_{21} * l_{11} & l_{21}^2 + l_{22}^2 & l_{21} * l_{31} + l_{22} * l_{32} & l_{21} * l_{41} + l_{22} * l_{42} \\ l_{31} * l_{11} & l_{31} * l_{21} + l_{32} * l_{22} & l_{31}^2 + l_{32}^2 + l_{33}^2 & l_{31} * l_{41} + l_{32} * l_{42} + l_{33} * l_{43} \\ l_{41} * l_{11} & l_{41} * l_{21} + l_{42} * l_{22} & l_{41} * l_{31} + l_{42} * l_{32} + l_{43} * l_{33} & l_{41}^2 + l_{42}^2 + l_{43}^2 + l_{44}^2 \end{bmatrix}.$$

Από την εξίσωση των αντίστοιχων στοιχείων των δύο μητρώων έχουμε:

$$\begin{aligned} l_{11}^2 = 1 &\Rightarrow l_{11} = 1 & l_{21}^2 + l_{22}^2 = 5 &\Rightarrow 4 + l_{22}^2 = 5 \Rightarrow l_{22} = \sqrt{5-4} \Rightarrow l_{22} = 1 \\ l_{11}l_{21} = 2 &\Rightarrow l_{21} = 2 & l_{21}l_{31} + l_{22}l_{32} = 1 &\Rightarrow l_{32} = 1 & l_{31}l_{41} + l_{32}l_{42} + l_{33}l_{43} = -1 &\Rightarrow l_{43} = -1 \\ l_{11}l_{31} = 0 &\Rightarrow l_{31} = 0 & l_{21}l_{41} + l_{22}l_{42} = 2 &\Rightarrow 2 + l_{42} = 2 \Rightarrow l_{42} = 0 & l_{41}^2 + l_{42}^2 + l_{43}^2 + l_{44}^2 = 3 &\Rightarrow 1 + 1 + l_{44}^2 = 3 \Rightarrow l_{44} = 1 \\ l_{11}l_{41} = 1 &\Rightarrow l_{41} = 1 & l_{31}^2 + l_{32}^2 + l_{33}^2 = 2 &\Rightarrow l_{33} = 1 \end{aligned}$$

Άρα ο παράγοντας Cholesky του μητρώου A είναι ο $L = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 2 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 1 & 0 & -1 & 1 \end{bmatrix}$.

2.5.15 Σύγκριση αντιστρεψιμότητας μητρώου και παραγοντοποίησης Cholesky

Στον πίνακα που ακολουθεί, συγκρίνονται τα χαρακτηριστικά των μητρώων που είναι αντιστρέψιμα και θετικά ορισμένα.

Αντιστρέψιμο μητρώο	Θ.Ο. μητρώο
Ιδιοτιμές του $\neq 0$	Ιδιοτιμές του > 0
Οδηγοί του $\neq 0$	Οδηγοί του > 0
Α.Δ.Κ.	Α.Δ.Κ. και θετικά διαγώνια στοιχεία
-	$x^T * A * x > 0, \forall x \neq 0, x \in R^n$

2.5.16 Σύγκριση παραγοντοποίησης LU και παραγοντοποίησης Cholesky αναφορικά με το χρόνο εκτέλεσης

Αναφορικά με το χρόνο εκτέλεσης των παραγοντοποιήσεων Cholesky και LU θα πρέπει να αναφέρουμε ότι η πρώτη έχει μικρότερο χρόνο εκτέλεσης.

2.6. Παραγοντοποίηση LDL^T και υπολογισμός αδράνειας συμμετρικού μητρώου

Όταν εκτελείται η πράξη $x = A^{-1} * b = A \backslash b$ για να λυθεί το γραμμικό σύστημα $A * x = b$, η MATLAB ελέγχει αν το μητρώο A είναι πραγματικό συμμετρικό μητρώο, δηλαδή εφαρμόζει παραγοντοποίηση LDL^T (με τη συνάρτηση Idl). Η συγκεκριμένη παραγοντοποίηση υπάρχει **πάντα** και χρειάζεται -όπως και η Cholesky- το μισό κόστος της LU. Επιπλέον, αξιοποιεί τη δομή του μητρώου (συμμετρία και ενδεχόμενη αραιότητα). **Άρα, μιλάμε για παραγοντοποίηση LDL^T γενικών συμμετρικών μητρώων. Επομένως, αν ένα συμμετρικό μητρώο δεν είναι θετικά ορισμένο ή ημιορισμένο, η απλή Cholesky αστοχεί, υπάρχει όμως παρόμοια παραγοντοποίηση που αξιοποιεί τη συμμετρία και αυτή είναι η LDL^T .**

Κανόνας: Για κάθε συμμετρικό μητρώο ($A = A^T$), υπάρχει παραγοντοποίηση $PAP^T = LDL^T$, όπου P = μητρώο μετάθεσης, L = κάτω τριγωνικό μητρώο με "1" στη διαγώνιο και D = μπλοκ διαγώνιο μητρώο με μπλοκ 1×1 (βαθμωτοί) ή 2×2 (μητρώα συμμετρικά).

Μορφή: Η γενική μορφή της LDL^T παραγοντοποίησης συμμετρικών μη – ορισμένων μητρώων είναι:

$$A = \begin{bmatrix} 0 & * & * \\ * & 0 & * \\ * & * & 0 \end{bmatrix} = \begin{bmatrix} D & u \\ u^T & 0 \end{bmatrix} = L * D * L^T = \begin{bmatrix} I & 0 \\ u^T D^{-1} & 1 \end{bmatrix} * \begin{bmatrix} D & 0 \\ 0 & -u^T D^{-1} u \end{bmatrix} * \begin{bmatrix} I & u^T D^{-1} \\ 0 & 1 \end{bmatrix}$$

Βασική ιδέα για την ύπαρξη της LDL^T : Αν $a_{11} = 0$ διερευνούμε αν υπάρχει μη – μηδενικό στοιχείο στην κύρια διαγώνιο του A . Αν υπάρχει, με χρήση μητρώων μετάθεσης το φέρνουμε στη θέση (1, 1). Αν δεν υπάρχει, τότε ισχυρίζόμαστε ότι υπάρχει τουλάχιστον ένα μη – μηδενικό στοιχείο στην πρώτη στήλη που μπορούμε να φέρουμε στη θέση (1, 1) ή (2, 1) έτσι ώστε το μητρώο 2×2 στις θέσεις (1:2, 1:2) να είναι αντιστρέψιμα.

Παράδειγμα: Έστω $A = \begin{bmatrix} 0 & 1 & 2 \\ 1 & 0 & -1 \\ 2 & -1 & 0 \end{bmatrix}$ μητρώο συμμετρικό αλλά όχι Θ.Ο. αφού όπως δεν διαθέτει Α.Δ.Κ.

>> chol(A) ↴

Matrix must be positive definite

>> [L, U] = lu(A) ↴

$$L = \begin{bmatrix} 0 & 1.0 & 0 \\ 0.5 & 0.5 & 1.0 \\ 1.0 & 0 & 0 \end{bmatrix} U = \begin{bmatrix} 2 & -1 & 0 \\ 0 & 1 & 2 \\ 0 & 0 & -2 \end{bmatrix}$$

>> [L1, D1, P1] = Idl(A) ↴

$$P_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \text{ μητρώο εναλλαγής}$$

$L_1 = \begin{bmatrix} 1.0 & 0 & 0 \\ 0 & 1.0 & 0 \\ -0.5 & 0.5 & 1.0 \end{bmatrix}$ μητρώο κάτω τριγωνικό με "1" στην κύρια διαγώνιο

$D_1 = \begin{bmatrix} 0 & 2 & 0 \\ 2 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ μπλοκ διαγώνιο μητρώο με μπλοκ 1×1 (βαθμωτοί), στην προκειμένη περίπτωση το 1 ή 2×2 (μητρώα συμμετρικά) που στην προκείμενη περίπτωση είναι το $\begin{bmatrix} 0 & 2 \\ 2 & 0 \end{bmatrix}$

Επαλήθευση παραδείγματος

Με δεδομένο το μητρώο $A = \begin{bmatrix} 0 & 2 & 2 \\ 2 & 0 & -1 \\ 2 & -1 & 0 \end{bmatrix}$, η LDL^T παραγοντοποίηση στη MATLAB θα δώσει:

`>> [L1, D1, P1] = ldl(A)` ↴

$$L_1 = \begin{bmatrix} 1.0 & 0 & 0 \\ 0 & 1.0 & 0 \\ -0.5 & 0.5 & 1.0 \end{bmatrix}, D_1 = \begin{bmatrix} 0 & 2 & 0 \\ 2 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, P_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}.$$

Επειδή το μητρώο A είχε "0" στην κύρια διαγώνιο, σύμφωνα με τη βασική ιδέα για την απόδειξη ύπαρξης της LDL^T παραγοντοποίησης, κάνουμε εναλλαγή με το μητρώο P_1 της 1^{ης} και της 3^{ης} στήλης του A , για να φέρουμε ένα μη – μηδενικό στοιχείο στη πρώτη στήλη, δηλαδή: $A * P_1 = \begin{bmatrix} 2 & 2 & 0 \\ -1 & 0 & 2 \\ 0 & -1 & 2 \end{bmatrix}$. Από τη γενική μορφή της LDL^T

παραγοντοποίησης: $A = \begin{bmatrix} D & u \\ u^T & 0 \end{bmatrix} \leftrightarrow \begin{bmatrix} 0 & 2 & 2 \\ 2 & 0 & -1 \\ 2 & -1 & 0 \end{bmatrix}$ έπειτα ότι: $D = \begin{bmatrix} 0 & 2 \\ 2 & 0 \end{bmatrix}$, $u = \begin{bmatrix} 2 \\ -1 \end{bmatrix}$, $u^T = [2 \quad -1]$ και $u^T * D^{-1} * u = [2 \quad -1] * \begin{bmatrix} 2 \\ -1 \end{bmatrix} = -1 - 1 = -2$.

Με τον όρο **αδράνεια συμμετρικού μητρώου** εννοούμε την **τριπλέτα** μη – αρνητικών ακεραίων (ζ, θ, v) που μετρούν το πλήθος των θετικών, αρνητικών και μηδενικών ιδιοτιμών. **Ισχύει ότι η τάξη του μητρώου $rank(A) = \theta + v$** . Η αδράνεια ενός συμμετρικού μητρώου **υπολογίζεται απευθείας από το μητρώο D της LDL^T** , χωρίς να χρειάζεται εύρεση των ιδιοτιμών. Πιο συγκεκριμένα, αν υποθέσουμε ότι στη διαγώνιο του μητρώου D υπάρχουν s **μπλοκ 1×1 (σύνολο S)** και $q = (\theta - s)/2$ **μπλοκ 2×2 (σύνολο B)**, τότε καθένα από αυτά τα q μπλοκ θα έχει 0 διαγώνιο και επειδή είναι συμμετρικό, θα έχει αρνητική ορίζουσα και θα συνεισφέρει 1 αρνητική και 1 θετική ιδιοτιμή. **Επομένως, η παράμετρος ζ προσδιορίζει πόσα στοιχεία του συνόλου S είναι μηδέν, η παράμετρος θ προσδιορίζει πόσα στοιχεία του συνόλου S είναι θετικά + q και η παράμετρος v προσδιορίζει πόσα στοιχεία του συνόλου S είναι αρνητικά + q .**

Παράδειγμα υπολογισμού αδράνειας συμμετρικού μητρώου

Έστω ότι για τον υπολογισμό της αδράνειας ενός συμμετρικού μητρώου, εκτελούμε τις ακόλουθες εντολές:

`>> A=randn(5)`

`A =`

```
0.5377 -1.3077 -1.3499 -0.2050 0.6715
1.8339 -0.4336 3.0349 -0.1241 -1.2075
-2.2588 0.3426 0.7254 1.4897 0.7172
0.8622 3.5784 -0.0631 1.4090 1.6302
0.3188 2.7694 0.7147 1.4172 0.4889
```

>> A=A+A' % δημιουργία συμμετρικού μητρώου

A =

```
1.0753 0.5262 -3.6087 0.6572 0.9903
0.5262 -0.8672 3.3775 3.4543 1.5620
-3.6087 3.3775 1.4508 1.4266 1.4320
0.6572 3.4543 1.4266 2.8181 3.0474
0.9903 1.5620 1.4320 3.0474 0.9778
```

>> [L,D,P]=ldl(A)

L =

```
1.0000 0 0 0 0
0 1.0000 0 0 0
-0.5323 -0.3407 1.0000 0 0
-1.1299 -0.4825 1.3369 1.0000 0
-0.5761 -0.4461 1.1118 0.3980 1.0000
```

D =

1.0753 -3.6087	0	0	0
-3.6087 1.4508	0	0	0
0 0	3.6540	0	0
0 0	0	-5.1742	0
0 0	0	0	-1.5097

P =

```
1 0 0 0 0
0 0 0 1 0
0 1 0 0 0
0 0 1 0 0
0 0 0 0 1
```

>> eig(A)

ans =

```
-4.8683
-1.6771
-1.0406
4.5189
8.5220
```

>> disp([eig(A),D])

$$\begin{array}{r}
 \boxed{-4.8683} & 1.0753 & -3.6087 & 0 & 0 & 0 \\
 \boxed{-1.6771} & -3.6087 & 1.4508 & 0 & 0 & 0 \\
 \boxed{-1.0406} & 0 & 0 & 3.6540 & 0 & 0 \\
 \boxed{4.5189} & 0 & 0 & 0 & -5.1742 & 0 \\
 \boxed{8.5220} & 0 & 0 & 0 & 0 & -1.5097
 \end{array}$$

Στη διαγώνιο του μητρώου D υπάρχουν s μπλοκ 1×1 (σύνολο S) και $q = (n - s) / 2$ μπλοκ 2×2 (σύνολο B), δηλαδή το $s = 1$ και το $q = 2 = (5 - 1)/2$. Για τον υπολογισμό της αδράνειας του συμμετρικού μητρώου έχουμε:

- Η παράμετρος ζ προσδιορίζει πόσα στοιχεία του συνόλου S είναι μηδέν $\rightarrow 0$ στοιχεία.
- Η παράμετρος θ προσδιορίζει πόσα στοιχεία του συνόλου S είναι θετικά $+ q \rightarrow 0 + 2 = 2$ στοιχεία.
- Η παράμετρος ν προσδιορίζει πόσα στοιχεία του συνόλου S είναι αρνητικά $+ q \rightarrow 1 + 2 = 3$ στοιχεία.

Επομένως η αδράνεια του συμμετρικού μητρώου $= (\zeta, \theta, \nu) = (0, 2, 3)$.

Ισχύει ότι η τάξη του μητρώου $\text{rank}(A) = \theta + \nu = 2 + 3 = 5$.

2.7. Δείκτης κατάστασης μητρώου – Απώλεια δεκαδικών ψηφίων λόγω δ.κ.

Με τον όρο **δείκτης κατάστασης μητρώου** εννοούμε το μέγεθος $\kappa(A) = \|A\| * \|A^{-1}\|$ για οποιαδήποτε νόρμα του μητρώου A . **Η ιδανική (μικρότερη) τιμή του δείκτη κατάστασης είναι 1** (τότε λέμε ότι το μητρώο έχει ιδανικό δείκτη κατάστασης) **και όσο μεγαλύτερη είναι η τιμή του, τόσο μεγαλύτερο σφάλμα προκαλείται από την επίλυση του γραμμικού συστήματος $A^*x = b$, λόγω απώλειας δεκαδικών ψηφίων** (αν $\kappa(A) >> 1$ τότε λέμε ότι το μητρώο έχει **κακή κατάσταση – ill conditioned**). Υπάρχει μάλιστα ο **τύπος $\log_{10}(\kappa(A))$** που δείχνει

την απώλεια των ψηφίων αυτών. Για παράδειγμα, αν έχουμε τα μητρώα $A = \begin{bmatrix} 1 & 2 & 2 \\ 0 & 5 & 1 \\ 3 & 4 & 3 \end{bmatrix}$ και $A^{-1} = \frac{1}{13} * \begin{bmatrix} -11 & -2 & 8 \\ -3 & 3 & 1 \\ 15 & -2 & -5 \end{bmatrix}$

και θέλουμε να υπολογίσουμε: α) το δείκτη κατάστασης (δ.κ.) του μητρώου και β) τα ψηφία που χάνονται λόγω δ.κ. Θα έχουμε τα εξής:

Λύση

α) **Αν χρησιμοποιήσουμε νόρμα μεγίστου για την εύρεση του δ.κ. του μητρώου A ,** θα έχουμε ότι: $\|A\|_\infty = 10$ (μεγαλύτερο κατ' απόλυτη τιμή άθροισμα από όλες τις γραμμές του μητρώου) και $\|A^{-1}\|_\infty = \frac{22}{13}$ τότε ο δ.κ. του μητρώου A είναι: $\kappa(A) = \|A\|_\infty * \|A^{-1}\|_\infty = 10 * \frac{22}{13} = \frac{220}{13}$. **Αν χρησιμοποιούσαμε την 1^η νόρμα για την εύρεση του δ.κ. του μητρώου A** (μεγαλύτερο κατ' απόλυτη τιμή άθροισμα από όλες τις στήλες του μητρώου), θα είχαμε ότι: $\kappa(A) = \|A\|_1 * \|A^{-1}\|_1 = 11 * \frac{29}{13} = \frac{319}{13}$, δηλαδή θα υπολογίζαμε **διαφορετική τιμή για το δ.κ.**

β) **Ψηφία που χάνονται λόγω δ.κ. = $\log_{10}(\kappa(A)) = \log_{10}\frac{220}{13} = 1.228$.** Άρα **χάνεται περίπου 1 ψηφίο**. Ακόμη και στην περίπτωση της 1^{ης} νόρμας χάνεται περίπου 1 ψηφίο, διότι τότε θα είχαμε ότι τα ψηφία που χάνονται λόγω δ.κ. = $\log_{10}(\kappa(A)) = \log_{10}\frac{319}{13} = 1.38$.

Παρατίρηση: Αν χρησιμοποιήσουμε ενδογενείς συναρτήσεις της MATLAB για τον υπολογισμό του δείκτη κατάστασης του μητρώου A , για παράδειγμα ως προς τη νόρμα απείρου, **αυτό μπορεί να γίνει με δύο τρόπους:** $\text{norm}(A, \text{Inf}) * \text{norm}(\text{inv}(A), \text{Inf})$ ή $\text{cond}(A, \text{Inf})$.

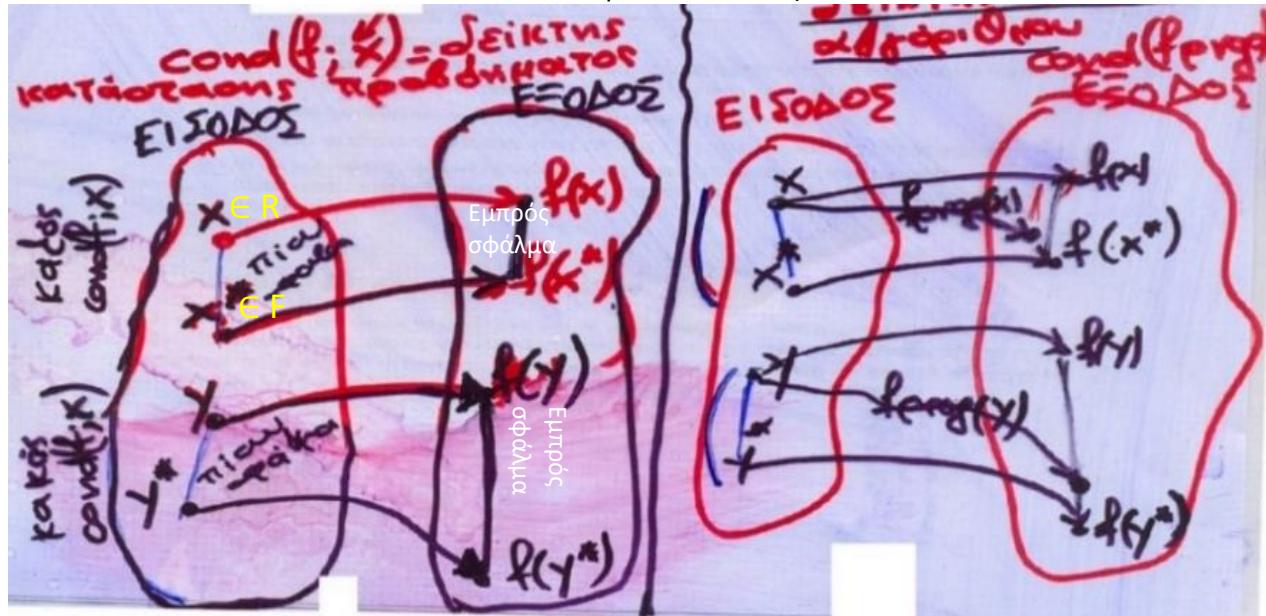
2.7.1 Δείκτες κατάστασης προβλήματος $\text{cond}(f; x)$ και αλγόριθμου $\text{cond}(f_{\text{prog}})$

Η **ευστάθεια ενός αλγόριθμου** αναφέρεται στο κατά πόσο είναι δυνατό η υπολογισμένη **από τον αλγόριθμο και με πράξεις α.κ.υ. λύση**, να μπορεί να θεωρηθεί ως ακριβής λύση (άπειρη ακρίβεια) του ίδιου προβλήματος με τον ίδιο αλγόριθμο, ενδεχομένως με λίγο τροποποιημένα δεδομένα εισόδου.

Ο **δείκτης κατάστασης του αλγόριθμου $\text{cond}(f_{\text{prog}})$** , είναι η **απόσταση** των τροποποιημένων στοιχείων εισόδου (x^*) από τα πραγματικά στοιχεία (x). Αν αυτή η απόσταση είναι σχετικά μικρή, τότε ο αλγόριθμος χαρακτηρίζεται ως **πίσω ευσταθής**. Θα πρέπει να αναφερθεί ότι η πίσω ευστάθεια σχετίζεται με το πίσω σφάλμα ενός αλγόριθμου. **Μεγάλη πίσω ευστάθεια σημαίνει μικρό πίσω σφάλμα και το αντίστροφο.**

Στο **δείκτη κατάστασης προβλήματος $\text{cond}(f; x)$** ενδιαφέρει η **ευαισθησία** της λύσης (δηλαδή των τιμών εξόδου) του προβλήματος ως προς τα δεδομένα (τιμές εισόδου). Αν για σχετικά μικρές αλλαγές στα δεδομένα (θεωρητική λύση) του προβλήματος, μπορεί να υποστεί σχετικά μεγάλη αλλαγή η έξοδος, το πρόβλημα χαρακτηρίζεται συχνά ως **ασταθές Τότε δηλαδή υπάρχει μεγάλη ευαισθησία της εξόδου σε αλλαγές της εισόδου**. Ενδιαφέρει η μέτρηση αυτής της αστάθειας και αυτό είναι δυνατό με ένα δείκτη που ονομάζεται **δείκτης κατάστασης του προβλήματος**. Αν ένα πρόβλημα **έχει μεγάλο δείκτη κατάστασης προβλήματος**, λέμε ότι έχει **κακό δείκτη κατάστασης** ή αλλιώς ότι είναι **κακώς τοποθετημένο (ill – posed)**, σε διαφορετική κατάσταση είναι **καλώς τοποθετημένο (well – posed)**. Στην πρώτη περίπτωση συνήθως βρίσκεται κοντά σε προβλήματα που είναι κακώς τοποθετημένα.

Με άλλα λόγια, εκτός από το **δείκτη κατάστασης μητρώου $\kappa(A)$** που είναι υποπερίπτωση του $\text{cond}(f; x)$, χρησιμοποιούνται οι **δείκτες κατάστασης προβλήματος $\text{cond}(f; x)$ και αλγόριθμου $\text{cond}(f_{\text{prog}})$** . Ο πρώτος από αυτούς αναφέρεται στην **ευαισθησία της εξόδου**, αν συμβούν μικρές αλλαγές στα δεδομένα εισόδου. Πιο συγκεκριμένα, όσο μικρότερος είναι ο δ.κ. προβλήματος $\text{cond}(f; x)$, τόσο λιγότερο μεταβάλλεται η έξοδος, όταν αλλάζει η είσοδος. **Ο δεύτερος από αυτούς, δηλαδή ο δείκτης κατάστασης αλγόριθμου $\text{cond}(f_{\text{prog}})$ δείχνει το κατά πόσο επηρεάζουν την έξοδο τα σφάλματα λόγω πράξεων που εκτελούνται μεταξύ των α.κ.υ. Η ιδινική τιμή του $\text{cond}(f_{\text{prog}})$ είναι «1», διότι τότε έχουμε μεγάλη πίσω ευστάθεια, άρα μικρό πίσω σφάλμα. Με άλλα λόγια, ο δεύτερος από αυτούς, δηλαδή το **$\text{cond}(f_{\text{prog}})$** , αναφέρεται στα σφάλματα που συσσωρεύονται κατά την εκτέλεση του αλγόριθμου (δηλ. στα σφάλματα λόγω πράξεων) σε μικρές αλλαγές των στοιχείων εισόδου (δεδομένων). Η σημασία τους εξηγείται με το σχήμα που ακολουθεί στη συνέχεια:**



Εικόνα 1: Εμπρός και πίσω σφάλμα – δείκτες κατάστασης προβλήματος και αλγόριθμου

Στο παραπάνω αριστερό σχήμα, φαίνεται στην **πρώτη περίπτωση** ότι μια μικρή αλλαγή στα δεδομένα εισόδου (x^* αντί για x , όπου $x \in R$ που σημαίνει ότι έχει **άπειρη** ακρίβεια, ενώ το $x^* \in F$ που σημαίνει ότι έχει κάποιο σφάλμα στρογγύλευσης) προκαλεί μια μικρή αλλαγή στην έξοδο δηλ. το $f(x^*)$ είναι κοντά στο $f(x)$, όπου $f(x)$ είναι η έξοδος του αλγόριθμου χωρίς σφάλμα και $f(x^*)$ είναι η έξοδος του αλγόριθμου με σφάλμα. Στη **δεύτερη** περίπτωση μια μικρή αλλαγή στα δεδομένα εισόδου (y^* αντί για y) προκαλεί μια μεγάλη αλλαγή στην έξοδο. Αυτό σημαίνει ότι στη δεύτερη περίπτωση ο δείκτης κατάστασης προβλήματος $cond(f; x)$ είναι μεγάλος (κακός). Η διαφορά ανάμεσα στα θεωρητικά και τα πραγματικά δεδομένα εισόδου ονομάζεται **πίσω σφάλμα** (αφορά πάντα την είσοδο, δηλαδή τα δεδομένα που εισάγουμε σε ένα αλγόριθμο), ενώ η διαφορά ανάμεσα στα θεωρητικά και τα πραγματικά δεδομένα εξόδου ονομάζεται **εμπρός σφάλμα** (αφορά πάντα την έξοδο, δηλαδή τα αποτελέσματα που παίρνουμε από ένα αλγόριθμο)

Στο **παραπάνω δεξιό σχήμα**, φαίνεται στην πρώτη περίπτωση ότι το σφάλμα των πράξεων προκαλεί μια μικρή αλλαγή στην έξοδο, δηλ. το $f_{prog}(x)$ που είναι η υλοποίηση του αλγόριθμου με σφάλμα, είναι κοντά στο $f(x)$, ενώ στη δεύτερη περίπτωση το σφάλμα των πράξεων προκαλεί μια μεγάλη αλλαγή στην έξοδο, δηλαδή το $f_{prog}(y)$ είναι μακριά από το $f(y)$. Επομένως στη δεύτερη περίπτωση ο δείκτης κατάστασης αλγόριθμου είναι μεγάλος (κακός).

2.7.2 Άσκηση υπολογισμού δείκτη κατάστασης μητρώου

Έστω το μητρώο $A = \begin{bmatrix} 1 & 0 & 0 \\ 6 & 1 & 0 \\ -7 & 0 & 1 \end{bmatrix}$. Ζητάμε το **δείκτη κατάστασης** του μητρώου, δηλ. το $\kappa(A)$ ως προς την 1^η νόρμα και ως προς τη νόρμα μεγίστου.

Λύση

Αρχικά θα υπολογίσουμε το αντίστροφο μητρώο A^{-1} . Επειδή το μητρώο A είναι **κάτω τριγωνικό**, το αντί-

στροφο θα είναι **επίσης** κάτω τριγωνικό της μορφής: $A^{-1} = \begin{bmatrix} 1/1 & 0 & 0 \\ x1 & 1/1 & 0 \\ x2 & x3 & 1/1 \end{bmatrix}$ και μάλιστα τα στοιχεία της κύριας

διαγωνίου του είναι τα **αντίστροφα στοιχεία της κύριας διαγωνίου του αρχικού** και ισχύει ότι $A^{-1} * A = I \Rightarrow$

$$\begin{bmatrix} 1 & 0 & 0 \\ x1 & 1 & 0 \\ x2 & x3 & 1 \end{bmatrix} * \begin{bmatrix} 1 & 0 & 0 \\ 6 & 1 & 0 \\ -7 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \Rightarrow x1 = -6 \text{ και } x2 = 7, \text{ οπότε το } A^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ -6 & 1 & 0 \\ 7 & 0 & 1 \end{bmatrix} \text{ και επομένως το } \kappa(A) =$$

$$\|A\|_1 * \|A^{-1}\|_1 = 14 * 14 = 196, \text{ ενώ ως προς τη νόρμα μεγίστου έχουμε: } \kappa(A) = \|A\|_\infty * \|A^{-1}\|_\infty = 8 * 8 = 64.$$

2.7.3 Άσκηση με δείκτη κατάστασης προβλήματος (παλιό θέμα)

Ο δείκτης κατάστασης **προβλήματος** αφορά: α) στην εύρεση φράγματος των αποτελεσμάτων σε σχέση με τα δεδομένα εισόδου β) στο μέγεθος των αλλαγών που μπορεί να υπάρξουν στα αποτελέσματα, αν τα δεδομένα εισόδου υποστούν μεγάλες αλλαγές γ) στο μέγεθος των αλλαγών στα αποτελέσματα, αν τα δεδομένα εισόδου υποστούν μικρές αλλαγές δ) τη μέτρηση του μεγαλύτερου σφάλματος στρογγύλευσης που προκύπτει κατά τη διάρκεια των πράξεων.

Λύση

Η σωστή απάντηση είναι η (γ) γιατί ικανοποιεί τον ορισμό του δείκτη κατάστασης προβλήματος. Αν επιλέγαμε την απάντηση (δ) τότε θα αναφερόμασταν στο δείκτη κατάστασης αλγόριθμου **cond(f_{prog})**, ο οποίος δείχνει την ευαισθησία της εξόδου σε σφάλματα στρογγυλοποίησης που προκαλούνται λόγω πράξεων μεταξύ των α.κ.υ.

Η απάντηση (β) είναι λάθος, διότι αναφέρει μεγάλες αλλαγές στα δεδομένα εισόδου.

2.7.4 Άσκηση με δείκτη κατάστασης μητρώου

Αν $A \in \mathbb{R}^{10 \times 10}$, τα στοιχεία του είναι μικρότερα του 1 και ο δείκτης κατάστασης είναι $\kappa_2(A) = 10^9$ τότε είναι σίγουρο ότι: α) Τίποτα από τα αναφερόμενα β) το μητρώο είναι αντιστρέψιμο γ) η μέγιστη ιδιοτιμή του είναι 10^9 και δ) το μητρώο δεν είναι αντιστρέψιμο.

Λύση

Η σωστή απάντηση είναι η (δ), διότι μητρώα **με πολύ μεγάλο δείκτη κατάστασης** είναι κοντά σε μη αντιστρέψιμα μητρώα. Ο δείκτης κατάστασης μητρώου υποδηλώνει τη σχετική απόσταση του από τα μη αντιστρέψιμα μητρώα. Για παράδειγμα αν δώσουμε στη γραμμή εντολών της MATLAB το ακόλουθο μητρώο:

```
>> A =
 1  2  3
 -4 -5 -6
 7  8  9
```

και θελήσουμε να υπολογίσουμε το δείκτη κατάστασης του μητρώου ως προς την πρώτη νόρμα, θα έχουμε

```
>> cond(A, 1)
```

Warning: Matrix is close to singular or badly scaled. Results may be inaccurate. RCOND = 1.541976e-18.

ans = 6.4852e+17

Αυτό σημαίνει ότι επειδή έχουμε έναν τεράστιο δείκτη κατάστασης, όπως φαίνεται παραπάνω, το μητρώο γίνεται ιδιάζον (singular) που σημαίνει μη – αντιστρέψιμο. Ανάλογο αποτέλεσμα παίρνουμε και στην περίπτωση που θέλουμε να υπολογίσουμε το δείκτη κατάστασης του μητρώου ως προς την νόρμα απείρου:

>> cond(A, Inf)

Warning: Matrix is close to singular or badly scaled. Results may be inaccurate.

2.7.5 Επαναληπτικές ασκήσεις με άμεσες μεθόδους – Άσκηση 1

Έστω ένα μητρώο A το οποίο θα χρησιμοποιούμε ως μητρώο συντελεστών κάποιου γραμμικού συστήματος $Ax = b$. Πως θα λύνατε το γραμμικό σύστημα, αν γνωρίζατε ότι το μητρώο είναι **τετραγωνικό**; Αν εκτός από τετραγωνικό ήταν και **συμμετρικό** και **θετικά ορισμένο** (ΣΘΟ);

Λύση

Αν το μητρώο ήταν **τετραγωνικό** και δε γνωρίζαμε τίποτα άλλο για αυτό, αναγκαστικά, ο μόνος τρόπος επίλυσης θα ήταν η παραγοντοποίηση LU με την προϋπόθεση ότι το μητρώο έχει A.D.K. κατά στήλες ή εναλλακτικά **όλα τα κύρια υπομητρώα είναι αντιστρέψιμα**. Τότε η λύση θα ήταν: υποθέτοντας ότι έχουμε την απαλοιφή $[L, U] = lu(A)$, $y = L \setminus b$, $x = U \setminus (L \setminus b)$ και αυτό γιατί $A * x = b \Rightarrow L * U * x = b$, θέτω $U * x = y$ και λύνω με "εμπρός αντικατάσταση", το κάτω τριγωνικό σύστημα $L * y = b$. Στη συνέχεια λύνω με "πίσω αντικατάσταση" το άνω τριγωνικό σύστημα $U * x = y$, βρίσκοντας τη λύση. **Διαφορετικά**, αν δεν μπορεί να εφαρμοστεί η LU παραγοντοποίηση, θα εφαρμογή η **PLU παραγοντοποίηση**, δηλαδή η LU παραγοντοποίηση με μερική οδήγηση, η οποία μπορεί **πάντα** να εφαρμοστεί. Στην περίπτωση αυτή η εμπρός αντικατάσταση δίνει $L * y = P^T * b$ και η πίσω αντικατάσταση δίνει $U * x = y \Rightarrow x = U \setminus y = U \setminus (L \setminus (P^T * b))$. Αν το μητρώο ήταν **ΣΘΟ** τότε αντί να πληρώσουμε το κόστος $O\left(\frac{2}{3}n^3\right)$ κάνουμε απαλοιφή Cholesky. Τότε, υποθέτοντας ότι έχουμε υπολογίσει την απαλοιφή Cholesky του A , η λύση θα είναι $y = L \setminus b$, $x = L \setminus y$ και αυτό γιατί $A * x = b \Rightarrow L * L^T * x = b$, θέτω $L^T * x = y$ και λύνω με "εμπρός αντικατάσταση" το κάτω τριγωνικό σύστημα $L * y = b$. Στη συνέχεια λύνω με "πίσω αντικατάσταση" το άνω τριγωνικό σύστημα $L^T * x = y$ βρίσκοντας την λύση.

2.7.6 Επαναληπτικές ασκήσεις με άμεσες μεθόδους – Άσκηση 2

Έστω ότι διαθέτουμε την LU παραγοντοποίηση του $A \in \mathbb{R}^{n \times n}$ και ότι $u, v \in \mathbb{R}^n$.

1. Να δείξετε ότι διαθέτουμε την LU παραγοντοποίηση του $B = A^{-1} * (I - uv^T)$.
2. Να υπολογίσετε τον κυρίαρχο συντελεστή (τύπου C_n^k) για τον αριθμό πράξεων που χρειάζονται για τον υπολογισμό του B . (Στόχος είναι να χρησιμοποιήσετε το μικρότερο πλήθος πράξεων α.κ.υ.).
3. Να περιγράψετε πως θα υπολογίζατε το $B * z$ για $z \in \mathbb{R}^n$.

Λύση

1. Από την υπόθεση γνωρίζουμε την παραγοντοποίηση $A = L * U$. Ζητούμενο είναι ο υπολογισμός του:

$$B = A^{-1}(I - uv^T) = A^{-1} - A^{-1}u v^T$$

2. Παρακάτω δίνεται ο ζητούμενος αλγόριθμος:

- (i) Επίλυση του συστήματος $A * x = u \Rightarrow x = A^{-1} * u \Rightarrow \Omega = n(2n-1)$
- (ii) Υπολογισμός του $A^{-1} \rightarrow \Omega = 2n^3$
- (iii) Υπολογισμός του $C = x * v^T \rightarrow \Omega = n^2$
- (iv) Υπολογισμός του $B = A^{-1} - C \rightarrow \Omega = n^2$

Παρακάτω δίνεται το κόστος του αλγόριθμου για κάθε ένα από τα παραπάνω βήματα:

Συνεπώς απαιτούνται συνολικά $\Omega = 2n^3 + 4n^2 - n$ πράξεις. Αυτό μπορεί να μειωθεί σε $\Omega = \frac{4}{3}n^3 + \frac{7}{2}n^2 + O(n)$.

3. Ο υπολογισμός του $B * z$, μπορεί να γίνει με τον εξής τρόπο:

$$B * z = A^{-1}(I - uv^T)z = A^{-1} * z - A^{-1} * u * v^T * z$$

Συνεπώς:

1. Επίλυση του συστήματος $A * x = z$
2. Υπολογισμός του $\tau = u^T z$
3. Υπολογισμός του $x - \tau y$

Το κυρίαρχο κόστος του παραπάνω αλγόριθμου προέρχεται από τη λύση των δύο συστημάτων στα βήματα 1 και 2, που απαιτούν συνολικά $4n^2$ πράξεις.

2.8 Επαναληπτικές Μέθοδοι - Είδη μητρώων επανάληψης - κριτήρια σύγκλισης – κριτήρια τερματισμού

Στις επαναληπτικές μεθόδους στοχεύουμε στη συστηματική **κατασκευή** μιας ακολουθίας διανυσμάτων $\{x^{(k)}\}$ έτσι ώστε να συγκλίνει στην ακριβή λύση \hat{x} του γραμμικού συστήματος $A * x = b$, χρησιμοποιώντας **το μικρότερο δυνατό πλήθος επαναλήψεων (βημάτων)** και ταυτόχρονα το **μικρότερο συνολικά αποδεκτό κόστος**. Η **συμπεριφορά** των επαναληπτικών μεθόδων **εξαρτάται από τις δομικές και μαθηματικές ιδιότητες του μητρώου των συντελεστών A**, π.χ. αν αυτό είναι Σ.Θ.Ο. κ.λ.π. Στη συνέχεια υποθέτουμε ότι το μητρώο A είναι **μεγάλο και αραιό** (η πλειοψηφία των στοιχείων του είναι μηδενικά), άλλωστε τα μεγάλα αραιά μητρώα εμφανίζονται πολύ πιο συχνά από τα πυκνά.

Γενική μορφή επαναληπτικού σχήματος: $P * x^{(k+1)} = (P - A) * x^{(k)} + b$ με δοθέν $x^{(0)}$ ή $x^{(k+1)} = T * x^{(k)} + g$. Με βάση το πρώτο γενικό σχήμα, $P * x^{(k+1)} = (P - A) * x^{(k)} + b$, **έχουμε τα εξής:**

- **Jacobi:** $P = D$, $P - A = L + U$. Άρα γενική μορφή: $D * x^{(k+1)} = (L + U) * x^{(k)} + b$
- **Gauss – Seidel:** $P = (D - L) = \text{diag}(\text{diag}(A)) + \text{tril}(A, -1) = \text{tril}(A)$, $P - A = U$. Άρα γενική μορφή: $(D - L) * x^{(k+1)} = U * x^{(k)} + b = \text{tril}(A) * x^{(k+1)} + b$
- **Richardson:** $P = \omega I$, $P - A = \omega I - A$. Άρα γενική μορφή: $\omega * I * x^{(k+1)} = (\omega I - A) * x^{(k)} + b$

Σε κάθε μια από τις τρείς επαναληπτικές μεθόδους, υπολογίζονται **αρχικά** τα μητρώα: $D = \text{diag}(\text{diag}(A))$, $L = -\text{tril}(A, -1)$, $U = -\text{triu}(A, 1)$. Η συνάρτηση **diag** όταν παίρνει ως όρισμα ένα **μητρώο**, επιστρέφει τα στοιχεία της κύριας διαγωνίου του μητρώου αυτού ως διάνυσμα, ενώ αντίθετα όταν παίρνει ως όρισμα ένα **διάνυσμα**,

το τοποθετεί στην κύρια διαγώνιο ενός κατά τα άλλα διαγώνιου μητρώου. Στην προκειμένη περίπτωση, **πρώτα** εκτελείται η εσωτερική $\text{diag}(A)$ και επιστρέφει ως διάνυσμα την κύρια διαγώνιο του A και στη **συνέχεια** εκτελείται η εξωτερική $\text{diag}(A)$, η οποία παίρνει ως όρισμα το προηγούμενο διάνυσμα και το τοποθετεί στην κύρια διαγώνιο ενός διαγώνιου μητρώου. Η συνάρτηση tril επιστρέφει το κάτω τριγωνικό τμήμα ενός μητρώου και το -1 σημαίνει ότι παίρνουμε τα στοιχεία ξεκινώντας από την **υποδιαγώνιο και κάτω**, ενώ η συνάρτηση triu επιστρέφει το άνω τριγωνικό τμήμα ενός μητρώου και το $+1$ σημαίνει ότι παίρνουμε τα στοιχεία ξεκινώντας από την **υπερδιαγώνιο και πάνω** π.χ. αν έχουμε το μητρώο συντελεστών $A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$ τότε το $\text{diag}(A) = \begin{bmatrix} 1 \\ 5 \\ 9 \end{bmatrix}$

και το διαγώνιο μητρώο $D = \text{diag}(\text{diag}(A)) = \text{diag}\left(\begin{bmatrix} 1 \\ 5 \\ 9 \end{bmatrix}\right) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 9 \end{bmatrix}$, το κάτω τριγωνικό μητρώο $L = -\text{tril}(A, -1) = \begin{bmatrix} 0 & 0 & 0 \\ -4 & 0 & 0 \\ -7 & -8 & 0 \end{bmatrix}$ και το άνω τριγωνικό μητρώο $U = -\text{triu}(A, 1) = \begin{bmatrix} 0 & -2 & -3 \\ 0 & 0 & -6 \\ 0 & 0 & 0 \end{bmatrix}$.

Με βάση το **δεύτερο γενικό σχήμα**, $x^{(k+1)} = T * x^{(k)} + g$ έχουμε τα εξής:

- **Jacobi:** $T = D^{-1} * (L + U)$, $g = D^{-1} * b$ οπότε $x^{(k+1)} = T * x^{(k)} + g = D^{-1} * (L + U) * x^{(k)} + D^{-1} * b$
- **Gauss–Seidel:** $T = (D - L)^{-1} * U$, $g = (D - L)^{-1} * b$ οπότε $x^{(k+1)} = T * x^{(k)} + g = (D - L)^{-1} * U * x^{(k)} + (D - L)^{-1} * b$
- **Richardson:** $T = I - A$ όπου $0 < \omega < 1$, $x^{(k+1)} = (I - A) * x^{(k)} + g = (I - A) * x^{(k)} + b$

Αν **συνδυάσουμε** τις δύο γενικές μορφές $P * x^{(k+1)} = (P - A) * x^{(k)} + b$ και $x^{(k+1)} = T * x^{(k)} + g$ θα έχουμε ότι:

- **Jacobi:** $P = D$, $P - A = L + U = Q$ και το μητρώο επανάληψης $T = D^{-1} * (L + U) = P^{-1} * (L + U) = P^{-1} * (P - A) = P^{-1} * Q = P \setminus Q$. Επίσης το διάνυσμα $g = D^{-1} * b = P^{-1} * b$.
- **Gauss – Seidel:** $P = D - L$, $P - A = U = Q$ και το μητρώο επανάληψης $T = (D - L)^{-1} * U = P^{-1} * (P - A) = P^{-1} * Q = P \setminus Q$
- **Richardson:** $P = \omega * I$ και το μητρώο επανάληψης $T = I - A$

2.8.1 Κριτήρια σύγκλισης επαναληπτικών μεθόδων

Σύγκλιση: Συνήθως η λύση ενός αριθμητικού προβλήματος εξαρτάται από κάποια «παράμετρο διακριτοποίησης» h . Λέμε ότι μια αριθμητική μέθοδος συγκλίνει αν καθώς $h \rightarrow 0$, η αριθμητική διαδικασία επιστρέφει λύση που τείνει προς τη λύση του μαθηματικού προβλήματος.

Τάξη σύγκλισης: Αν μπορούμε να φράξουμε το (απόλυτο ή σχετικό) σφάλμα συναρτήσει του h ως: $e^c \leq Ch^p$ για σταθερό C ανεξάρτητο του h και όπου p θετικός αριθμός, λέγεται ότι η μέθοδος είναι συγκλίνουσα, τάξης p .

Αναφορικά με τα κριτήρια σύγκλισης, πριν ξεκινήσουμε μια επαναληπτική μέθοδο, θα πρέπει να βεβαιωθούμε ότι συγκλίνει σε κάποια λύση. Υπάρχουν διαφορετικά **κριτήρια σύγκλισης, τα οποία εφαρμόζονται με την εξής σειρά:**

- Αν το μητρώο A στο σύστημα $A * x = b$ έχει **A.Δ.Κ.** (είτε κατά γραμμές είτε κατά στήλες), τότε συγκλίνουν οι μέθοδοι **Jacobi** και **Gauss – Seidel**. Με τον όρο **A.Δ.Κ.** (αυστηρή διαγώνια κυριαρχία) εννοούμε ότι τα στοιχεία

της κύριας διαγωνίου κατ' απόλυτη τιμή είναι μεγαλύτερα από το άθροισμα των απολύτων τιμών των υπολοίπων στοιχείων της γραμμής ή της στήλης που ανήκουν. Αν αναφερόμαστε στις γραμμές, τότε μιλάμε για A.D.K. κατά γραμμές, ενώ αν αναφερόμαστε στις στήλες τότε μιλάμε για A.D.K. κατά στήλες. **Για να έχουμε σύγκλιση της μεθόδου Jacobi ή της μεθόδου Gauss – Seidel αρκεί να έχουμε A.D.K. μόνο κατά ένα από τα δύο είδη A.D.K.**

είτε κατά γραμμές είτε κατά στήλες. Ένα μητρώο όπως το $A = \begin{bmatrix} 6 & -1 & 2 & 2 \\ -1 & 5 & -1 & 2 \\ 1 & 1 & -8 & 5 \\ -1 & 0 & 0 & 3 \end{bmatrix}$ που έχει A. Δ. K. κατά γραμμές,

ικανοποιεί την σχέση $|a_{ii}| > \sum_{j \neq i} |a_{ij}|$ και για αυτό μπορεί να χρησιμοποιηθεί τόσο η Jacobi όσο και η Gauss – Seidel **και επιπλέον** έχει την ιδιότητα ότι είναι **πάντα αντιστρέψιμο**.

(β) Αν το μητρώο A είναι μη – αναγωγήσιμο και διαγώνια κυρίαρχο (ΜΑΔΚ), τότε συγκλίνουν τόσο η Jacobi όσο και η Gauss – Seidel. Ένα μητρώο ονομάζεται μη – αναγωγήσιμο/αναγωγήσιμο, αν αφού το υποδιαιρέσουμε σε πλοκάδες στη μορφή $\begin{bmatrix} A_1 & B \\ 0 & A_2 \end{bmatrix}$ στη συνέχεια διαπιστώνουμε μια από τις ακόλουθες μορφές:

Μη – Αναγωγήσιμο

$$\left[\begin{array}{ccc|cc} A_1 & & & 0 & 0 \\ * & * & & 0 & 0 \\ * & * & * & 0 & 0 \\ 0 & * & * & * & * \\ \hline 0 & 0 & * & * & * \\ & & & & A_2 \end{array} \right] \rightarrow \begin{bmatrix} A_1 & B \\ 0 & A_2 \end{bmatrix} \text{ άνω τριγωνική μπλοκ μορφή}$$

Παρατηρούμε ότι τα **κύρια υπομητρώα A_1 και A_2 (πλοκάδες)** είναι μη – τετραγωνικά.

Αναγωγήσιμο

$$\left[\begin{array}{ccc|cc} A_1 & & & 0 & 0 \\ * & * & 0 & 0 & 0 \\ * & * & * & 0 & 0 \\ 0 & * & * & * & * \\ \hline 0 & 0 & 0 & * & * \\ & & & & A_2 \end{array} \right] \rightarrow \begin{bmatrix} A_1 & B \\ 0 & A_2 \end{bmatrix} \text{ άνω τριγωνική μπλοκ μορφή}$$

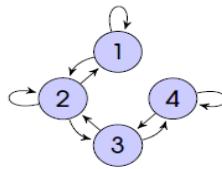
Παρατηρούμε ότι στη δεύτερη περύπτωση τα κύρια υπομητρώα A_1 και A_2 (πλοκάδες) είναι **τετραγωνικά**.

Ένα μητρώο είναι μη – αναγωγήσιμο, διαγώνια κυρίαρχο (ΜΑΔΚ) όταν είναι: i) μη – αναγωγήσιμο και ii) διαγώνια κυρίαρχο με αυστηρή ΔΚ για τουλάχιστον μια γραμμή ή μια στήλη. Η αναγωγησιμότητα/μη αναγωγησιμότητα ενός μητρώου έχει **ισοδύναμη διατύπωση** μέσω του **γραφήματος γειτνίασης** του μητρώου, δηλαδή η αναγωγησιμότητα ελέγχεται **και από το γράφημα γειτνίασης του μητρώου, αν αυτό δίνεται**. Σε μια τέτοια περύπτωση **δεν** χρειάζεται να αναχθεί σε άνω τριγωνική μπλοκ μορφή. Πιο συγκεκριμένα, ένα μητρώο είναι **μη αναγωγήσιμο** αν και μόνο αν το γράφημα γειτνίασης του είναι **ισχυρά συνεκτικό**, δηλ. για κάθε ζεύγος διαφορετικών κόμβων (v_i, v_j) υπάρχουν δύο διαφορετικά κατευθυνόμενα μονοπάτια που τους συνδέουν μεταξύ τους. Αυτό διευκρινίζεται και στην Εικόνα 1 που ακολουθεί.

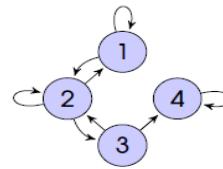
Γενιά αν ένα μητρώο περιέχει **αποκλειστικά μη μηδενικά στοιχεία τότε το αντίστοιχο γράφημα γειτνίασης**

Θα περιέχει όλες τις πιθανές ακμές, συνεπώς θα είναι ισχυρά συνεκτικό. Επίσης, αν το μητρώο παρουσιάζει αυστηρά διαγώνια κυριαρχία για κάποια γραμμή/στήλη θα είναι Θ.Ο.

$$\text{μη αναγωγήσιμο} \quad \left(\begin{array}{ccc|cc} * & * & 0 & 0 \\ * & * & * & 0 \\ 0 & * & * & * \\ \hline 0 & 0 & * & * \end{array} \right)$$

ισχυρά συνεκτικό

$$\text{αναγωγήσιμο} \quad \left(\begin{array}{ccc|cc} * & * & 0 & 0 \\ * & * & * & 0 \\ 0 & * & * & * \\ \hline 0 & 0 & 0 & * \end{array} \right)$$

μη ισχυρά συνεκτικό

Εικόνα 1: Σχέση αναγωγησιμότητας και ισχυρά/μη ισχυρά συνεκτικού γραφήματος

Παρατήρηση: Εναλλακτικά ένα μητρώο ονομάζεται **αναγωγήσιμο** αν υπάρχει μεταθετικό μητρώο P τέτοιο ώστε το $P^* A P^T$ να είναι **κατά πλοκάδες άνω τριγωνικό**, διαφορετικά καλείται **μη αναγωγήσιμο**. Δηλαδή ανν (αν και μόνο αν) ισχύει ότι: $P^* A P^T = \begin{pmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{pmatrix}$. Αν ένα μητρώο είναι **μη αναγωγήσιμο**, κάθε γραμμή και στήλη θα έχει ένα **τουλάχιστον μη μηδενικό στοιχείο πέραν της διαγωνίου**.

(ν) Αν το μητρώο A είναι **Σ.Θ.Ο.** τότε συγκλίνει η **Gauss – Seidel** και η **Richardson**, ενώ για να συγκλίνει **και** η **Jacobi** τότε πρέπει να είναι **Σ.Θ.Ο. και** το μητρώο $2 * D - A$.

(δ) Αν μια οποιαδήποτε νόρμα του **μητρώου επανάληψης T** (κατά προτίμηση όμως η $1^{\text{η}}$ ή η νόρμα απείρου), είναι μικρότερη από 1, δηλ. $\|T\| < 1$, τότε συγκλίνει η **Jacobi** και η **Gauss – Seidel**.

Σημείωση: Όλες οι παραπάνω συνθήκες (κριτήρια) είναι **μόνο** ικανές. Αυτό σημαίνει ότι αν δεν ισχύει μια από αυτές, απλώς εξετάζουμε την επόμενη για να δούμε αν αυτή ικανοποιείται και αν αυτή ισχύει, τότε συγκλίνει η επαναληπτική μέθοδος που εξετάζουμε και προχωράμε στην επίλυση του γραμμικού συστήματος, δηλαδή εφαρμόζουμε την επαναληπτική μέθοδο.

(ε) Το $\rho(T) < 1$, δηλαδή η **φασματική ακτίνα του μητρώου επανάληψης T** (μεγαλύτερη κατ' απόλυτη τιμή ιδιοτιμή) θα πρέπει να είναι μικρότερη από «1» **για να συγκλίνουν και οι τρεις μέθοδοι**. Η τελευταία συνθήκη είναι **ικανή και αναγκαία**. Η φασματική ακτίνα του μητρώου T υπολογίζεται από το **χαρακτηριστικό πολυώνυμο $\det(\lambda * I - T) = 0$** . Από αυτό υπολογίζουμε **όλες** τις ιδιοτιμές του μητρώου επανάληψης T και στη συνέχεια από αυτές υπολογίζουμε τη φασματική ακτίνα του μητρώου.

Παρατήρηση 1: Η διαφορά ανάμεσα στις έννοιες **ικανή συνθήκη** και **ικανή και αναγκαία συνθήκη** είναι ότι αν έχουμε μια ικανή συνθήκη και δεν ισχύει τότε εξετάζουμε την επόμενη συνθήκη για να δούμε αν συγκλίνει κάποια από τις επαναληπτικές μεθόδους. Αν όμως έχουμε μια ικανή και αναγκαία συνθήκη και δεν συγκλίνει, τότε ξέρουμε με σιγουριά ότι δεν συγκλίνει καμία επαναληπτική μέθοδος.

Παρατήρηση 2: Για **πραγματικά συμμετρικά ή μιγαδικά ερμιτιανά μητρώα** η σύγκλιση της μεθόδου **Richardson** εξασφαλίζεται αν $|1 - \lambda(A)| < 1 \Rightarrow 0 < \lambda(A) < 2$ για κάθε ιδιοτιμή. Για **γενικά μητρώα**, η σύγκλιση της

Richardson εξασφαλίζεται αν οι ιδιοτιμές του A περιέχονται στο εσωτερικό του μοναδιαίου δίσκου ή κύκλου, δηλαδή αυτού του κύκλου που έχει κέντρο στο 1 και ακτίνα 1.

2.8.2 Άσκηση με επαναληπτικές μεθόδους (Θέμα Φεβ. 2017)

Για να λύσετε το γραμμικό σύστημα $A * x = b$ όπου το A είναι **μεγάλο, αραιό και αυστηρά διαγώνιο κυρί-αρχο**, ποια από τις παρακάτω μεθόδους θα προτιμούσατε να χρησιμοποιήστε;

- (α) Συζυγών κλίσεων (β) απαλοιφή Gauss (γ) Jacobi (δ) Gauss - Seidel

Λύση

Οι σωστές απαντήσεις είναι οι **δύο τελευταίες**, δηλ. (γ) και (δ), διότι από τη στιγμή που το μητρώο A είναι **μεγάλο, αραιό και αυστηρά διαγώνιο κυρίαρχο δεν ενδείκνυται η χρήση άμεσων μεθόδων** (όπως η απαλοιφή Gauss) και επίσης το γεγονός ότι το μητρώο A έχει **ΑΔΚ** σημαίνει ότι -σύμφωνα με το πρώτο κριτήριο σύγκλισης των επαναληπτικών μεθόδων που αναφέρθηκε νωρίτερα- συγκλίνουν τόσο η Jacobi όσο και η Gauss - Seidel. Η μέθοδος συζυγών κλίσεων έχει την ίδια προϋπόθεση με τη μέθοδο Cholesky, δηλ. απαιτεί από το μητρώο που παραγοντοποιείται να είναι Σ.Θ.Ο. και ανήκει στις **επαναληπτικές μεθόδους**.

Παρατήρηση: μια **εναλλακτική διατύπωση** του προηγούμενου είναι η εξής:

Για ένα **συμμετρικό θετικά ορισμένο μητρώο A , μεγάλο και αραιό**, ποια από τις παρακάτω μεθόδους θα προτιμούσατε να χρησιμοποιήσετε και γιατί; (α) Jacobi, (b) Gauss-Seidel, (c) LU με μερική οδήγηση και (d) Cholesky

Λύση

Η σωστή απάντηση είναι η (b) διότι για μεγάλα και αραιά μητρώα **οι ενδεικνυόμενες μέθοδοι επίλυσης του γραμμικού συστήματος $A * x = b$ είναι οι επαναληπτικές μέθοδοι, στις οποίες εντάσσονται οι Jacobi και Gauss-Seidel**. Από αυτές τις δύο επιλέγουμε τη Gauss-Seidel, διότι το μητρώο είναι ΣΘΟ. Η παραγοντοποίηση Cholesky, αν και μπορεί να εφαρμοστεί, εντούτοις δεν ενδείκνυται για μεγάλα και αραιά μητρώα, διότι είναι άμεση μέθοδος.

2.8.3 Άσκηση με επαναληπτικές μεθόδους (παλιό Θέμα)

Αν συγκλίνει η επαναληπτική μέθοδος $x^{(k+1)} = T^*x^{(k)} + g$ για τη λύση του γενικού γραμμικού συστήματος $A * x = b$, όπου T είναι το μητρώο επανάληψης. α) Το T είναι αυστηρά διαγώνια κυρίαρχο β) Ισχύει ότι το A είναι αυστηρά διαγώνια κυρίαρχο γ) Ισχύει ότι $\|T\|_2 < 1$ δ) κανένα από τα υπόλοιπα.

Λύση

Η σωστή απάντηση είναι η (δ), διότι: το (α) **δεν** ισχύει, διότι η Α.Δ.Κ. εξετάζεται στο μητρώο A των συντελεστών και όχι στο μητρώο επανάληψης T . Επίσης το (β) δεν ισχύει, διότι σύμφωνα με το 1^ο κριτήριο σύγκλισης από πριν, ισχύει το αντίστροφο, δηλαδή αν το μητρώο A έχει Α.Δ.Κ., συγκλίνει μια επαναληπτική μέθοδος,

όπως η Jacobi και η Gauss-Seidel. Για τον ίδιο λόγο δεν ισχύει και το (γ), διότι και πάλι έχουμε ότι αν μια οποιαδήποτε νόρμα του μητρώου επανάληψης T είναι μικρότερη από 1 (δηλ. $\|T\| < 1$), συγκλίνει μια επαναληπτική μέθοδος, όπως η Jacobi και η Gauss-Seidel.

Παρατήρηση: Για να ήταν μια τέτοια πρόταση από τις προηγούμενες αληθής, θα έπρεπε από την εκφώνηση της άσκησης να έχει διθεί ότι μια από τις επιλογές ήταν η φασματική ακτίνα του μητρώου επανάληψης T (δηλ. $\rho(T) < 1$).

2.8.4 Άσκηση με αναγωγησιμότητα μητρώου

Αν το μητρώο A είναι **αναγωγήσιμο και αντιστρέψιμο**, τότε συνήθως: α) μπορεί να παραγοντοποιηθεί με Cholesky b) οι ιδιοτιμές του είναι θετικές γ) η επίλυση ενός γραμμικού συστήματος $A^*x = b$ μπορεί να επιτευχθεί μέσω επίλυσης μικρότερων συστημάτων και δ) κανένα από τα υπόλοιπα.

Λύση

Η σωστή απάντηση είναι η (γ). Από τη στιγμή που ένα μητρώο είναι **αναγωγήσιμο**, η ιδιότητα αυτή (αναγωγησιμότητα) επιτρέπει την αναγωγή (διάσπαση) του μητρώου σε μικρότερα. Συγκεκριμένα, από τη στιγμή που ένα μητρώο είναι αναγωγήσιμο, σημαίνει ότι υπάρχουν μητρώα εναλλαγής P και P^T , τέτοια ώστε $P^*A^*P^T = \begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix}$ δηλαδή πρέπει πρώτα το μητρώο να έρθει σε άνω τριγωνική μπλοκ μορφή. Πράγματι, αν αυτό συμβεί, τότε η λύση του γραμμικού συστήματος $A^*x = b$ με μητρώο **A αναγωγήσιμο**, μας δίνει το εξής: $\begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix} * \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} \Rightarrow \boxed{A_{22} * x_2 = b_2} \text{ και } \boxed{A_{11} * x_1 = b_1 - A_{12} * x_2}$. Εδώ έχουμε μικρότερα συστήματα, διότι αντί για ολόκληρο το μητρώο A έχουμε τα υπομητρώα A_{11} και A_{22} , που είναι **απλούστερα**.

Σημείωση: Επίσης, σε ένα **αναγωγήσιμο μητρώο** ισχύει ότι οι ιδιοτιμές του προκύπτουν από τον τύπο $\lambda(A) = \lambda(A_{11}) \cup \lambda(A_{22})$, όπου A_{11} και A_{22} είναι οι πλοκάδες (υπομητρώα) της κύριας διαγωνίου, οι οποίες είναι τετραγωνικές. Δηλαδή, ο τελευταίος τύπος αναφέρει ότι οι ιδιοτιμές ενός αναγωγήσιμου μητρώου προκύπτουν από τις ιδιοτιμές των υπομητρώων της κύριας διαγωνίου του. Αυτό συμβαίνει διότι το αναγωγήσιμο μητρώο είναι σε μπλοκ άνω τριγωνική μορφή κάτι που σημαίνει ότι οι ιδιοτιμές του είναι οι ιδιοτιμές των υπομητρώων της κύριας διαγωνίου. Αυτό ισχύει και για διαγώνια μητρώα.

2.8.5 Άσκηση με αναγωγησιμότητα μητρώου

Αν το μητρώο A είναι **αναγωγήσιμο**, τότε: α) οι ιδιοτιμές του είναι θετικές β) υπάρχει μητρώο μετάθεσης P τέτοιο ώστε PAP^T είναι κατά πλοκάδες άνω τριγωνικό γ) δεν μπορεί να διαγωνοποιηθεί και δ) κανένα από τα υπόλοιπα.

Λύση

Για να είναι σωστή απάντηση η (α) πρέπει το μητρώο να είναι ΣΘΟ, διότι τότε οι ιδιοτιμές του είναι θετικές. Η σωστή απάντηση είναι η (β). Από τη στιγμή που ένα μητρώο είναι **αναγωγήσιμο**, τότε εξ' ορισμού υπάρχει μητρώο P τέτοιο ώστε το μητρώο PAP^T να είναι κατά πλοκάδες άνω τριγωνικό.

$$\text{Π.χ. } \text{Αν } A = \begin{bmatrix} 1 & 2 & 0 \\ 4 & 5 & 6 \\ 7 & 8 & 3 \end{bmatrix} \text{ τότε } PAP^T = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} * \begin{bmatrix} 1 & 2 & 0 \\ 4 & 5 & 6 \\ 7 & 8 & 3 \end{bmatrix} * \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}^T = \begin{bmatrix} 3 & 8 & 7 \\ 6 & 5 & 4 \\ 0 & 2 & 1 \end{bmatrix}.$$

διότι **για να μπορεί να διαγωνοποιηθεί ένα μητρώο**, θα πρέπει **είτε να έχει διακριτές (διαφορετικές) ιδιοτιμές**, κάτι που δεν προκύπτει από την εκφώνηση, **είτε να είναι συμμετρικό**, γιατί εξ' ορισμού όλα τα συμμετρικά μητρώα διαγωνοποιούνται. **Δεν** έχουμε επαρκή πληροφόρηση για το αν ισχύουν ή όχι τα κριτήρια διαγωνοποίησης. **Αν ένα μητρώο μπορεί να διαγωνοποιηθεί**, γράφεται στη μορφή: $A = P * D * P^{-1}$, όπου το D = διαγώνιο μητρώο με τις ιδιοτιμές του A στην κύρια διαγώνιο και P = μητρώο που έχει στις στήλες του τα ιδιοδιανύσματα του A . Ένα ιδιοδιάνυσμα υπολογίζεται με βάση μια ιδιοτιμή από τον τύπο $A * x = \lambda * x$, όπου λ = ιδιοτιμή και x = αντίστοιχο ιδιοδιάνυσμα. **Αν ένα μητρώο είναι συμμετρικό είναι εξ' ορισμού διαγωνοποιήσιμο**.

2.8.6 Άσκηση με διαγώνια κυριαρχία

Αν το μητρώο A είναι **διαγώνια κυριαρχο** τότε:

α) Οι ιδιοτιμές του είναι θετικές b) η επίλυση ενός γραμμικού συστήματος $A * x = b$ μπορεί να επιτευχθεί μέσω επίλυσης μικρότερων συστημάτων c) υπάρχει μητρώο μετάθεσης P τέτοιο ώστε το PAP^T είναι κατά πλοκάδες άνω τριγωνικό, d) τίποτα από τα υπόλοιπα, δεν έχει θετικές ιδιοτιμές.

Λύση

Η σωστή απάντηση είναι η (d). **Από τη στιγμή που ένα μητρώο είναι διαγώνια κυριαρχο, τότε εξ' ορισμού δεν υπάρχει μητρώο P τέτοιο ώστε το μητρώο PAP^T να είναι κατά πλοκάδες άνω τριγωνικό, διότι αυτό ισχύει μόνο για αναγωγήσιμα μητρώα.** Επίσης οι ιδιοτιμές του **δεν** είναι θετικές, διότι αυτό ισχύει για θετικά ορισμένα μητρώα. Επίσης η επίλυση ενός γραμμικού συστήματος $A * x = b$ **δεν** μπορεί να επιτευχθεί μέσω επίλυσης μικρότερων συστημάτων, διότι αυτό ισχύει για **αναγωγήσιμα μητρώα**.

2.8.7 Κριτήρια τερματισμού επαναληπτικών μεθόδων

Αναφορικά με τα κριτήρια τερματισμού, από τη στιγμή που έχει ξεκινήσει η εφαρμογή μιας επαναληπτικής μεθόδου, θα πρέπει σε κάθε βήμα να ελέγχουμε **αν πρέπει να τερματιστεί η μέθοδος**, χρησιμοποιώντας ένα από τα ακόλουθα κριτήρια τερματισμού:

- **Σχετικού καταλοίπου**: $r = b - Ax$, δηλ. $\frac{\|r^{(k)}\|}{\|b\|} \leq \frac{\varepsilon}{k(A)}$ όπου ε μια δοθείσα ακρίβεια, $k(A)$ είναι ο δείκτης κατάστασης του μητρώου A .
- **Μεταβολής**: $\|x^{(k)} - x^{(k-1)}\| \leq \frac{1 - \|T\|}{\|T\|} * \varepsilon$, όπου ε είναι και πάλι μια δοθείσα ακρίβεια. Ειδικά το τελευταίο κριτήριο μπορεί **εναλλακτικά** να γραφτεί και με τη μορφή απόλυτου ή σχετικού σφάλματος ως εξής: **Απόλυτο σφάλμα**: $\|x^{(k)} - x^{(k-1)}\| \leq 10^{-m}$, όπου m είναι τα σημαντικά **Ψηφία** και το **σχετικό σφάλμα** δίνεται από τον

τύπο: $\left\| \frac{x^{(k)} - x^{(k-1)}}{x^{(k)}} \right\| \leq \frac{10^{-m}}{2}$, όπου και πάλι m είναι τα σημαντικά ψηφία και τα οποία δίνονται από την εκφώνηση της άσκησης. Συνήθως εφαρμόζεται το **δεύτερο κριτήριο** και μάλιστα με τους τύπους που περιγράφονται στην συνέχεια:

$$b1. \|x^{(k+1)} - x^{(k)}\| \leq \eta_1 \text{ και } b2. \frac{\|x^{(k+1)} - x^{(k)}\|}{\|x^{(k+1)}\|} \leq \eta_2, \text{ όπου } \eta_1 \text{ και } \eta_2 \text{ μικροί θετικοί αριθμοί.}$$

Το κριτήριο b1 συσχετίζεται με το **απόλυτο σφάλμα**, αν θεωρηθεί ότι το $x^{(k+1)}$ είναι η ακριβής τιμή και ότι το $x^{(k)}$ είναι η προσεγγιστική τιμή. Με αυτή την έννοια το κριτήριο b2, το οποίο συνήθως εφαρμόζεται στην πράξη, συσχετίζεται με το **σχετικό σφάλμα**. Έτσι, στην πράξη, αν επιθυμούμε την προσέγγιση λύσης με **τη σημαντικά ψηφία** και θεωρώντας το δεύτερο κριτήριο, χρησιμοποιώντας την νόρμα απείρου, τότε μπορούμε να πάρουμε: $\eta_2 = \frac{1}{2} 10^{-m}$, $\eta_1 = 10^{-m}$. Αφού βεβαιωθούμε ότι συγκλίνει μια επαναληπτική μέθοδος, προχωράμε στη συνέχεια στη χρήση της, για την επίλυση του γραμμικού συστήματος. Για να τερματιστεί μια επαναληπτική μέθοδος, θα πρέπει σε κάθε βήμα της να εξετάζουμε αν ικανοποιείται κάποιο κριτήριο τερματισμού.

Σημείωση: Με τον όρο σημαντικά ψηφία εννοούμε τον αριθμό των ψηφίων ενός αριθμού που γνωρίζουμε με απόλυτη βεβαιότητα, π.χ. ο αριθμός 23,21 cm έχει τέσσερα σημαντικά ψηφία, ενώ ο αριθμός 0,062 cm έχει δύο σημαντικά ψηφία (τα μηδενικά μέχρι την υποδιαστολή δεν μετράνε).

2.9 Μέθοδος Jacobi

Στη μέθοδο Jacobi το μητρώο ρυθμιστής είναι το $M = D$. Έτσι αν το αντίστροφο μητρώο D^{-1} υπάρχει, το επαναληπτικό σχήμα της μεθόδου Jacobi γράφεται ως εξής: $x^{(k+1)} = D^{-1}(L + U) * x^{(k)} + D^{-1} * b, k = 0, 1, 2, \dots$ με οποιοδήποτε $x^{(0)} \in C^n$. Όσον αφορά τη σύγκλιση, η ικανή συνθήκη για την σύγκλιση του επαναληπτικού σχήματος είναι το μητρώο A να έχει αυστηρά διαγώνια κυριαρχία κατά γραμμές ή κατά στήλες ή να ισχύει ότι: $\|T\| = \|D^{-1}(L + U)\| < 1$, ενώ ικανή και αναγκαία συνθήκη είναι η εξής: $\rho(T) = \rho(D^{-1}(L + U)) < 1$.

2.9.1 Άσκηση εφαρμογής μεθόδου Jacobi

Θα εφαρμόσουμε τη μέθοδο του **Jacobi** για επίλυση του παρακάτω συστήματος των γραμμικών εξισώσεων:

$$x_1 + 2x_2 - 2x_3 = 1$$

$$x_1 + x_2 + x_3 = 1$$

$$2x_1 + 2x_2 + x_3 = 1$$

χρησιμοποιώντας ως αρχική εκτίμηση λύσης την $x^{(0)} = [1 \quad 1 \quad 1]^T$. Πάντα, όταν χρησιμοποιείται επαναληπτική μέθοδος, θα πρέπει να δίνεται η αρχική εκτίμηση της λύσης. Το $m = 2$.

Λύση

Το μητρώο των συντελεστών των αγνώστων και το σταθερό διάνυσμα είναι αντίστοιχα: $A = \begin{bmatrix} 1 & 2 & -2 \\ 1 & 1 & 1 \\ 2 & 2 & 1 \end{bmatrix}$, $b = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$. Το $D = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$, $L = \begin{bmatrix} 0 & 0 & 0 \\ -1 & 0 & 0 \\ -2 & -2 & 0 \end{bmatrix}$, $U = \begin{bmatrix} 0 & -2 & 2 \\ 0 & 0 & -1 \\ 0 & 0 & 0 \end{bmatrix}$, οπότε το μητρώο T είναι: $T = D^{-1} * (L + U) =$

$$\begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 0 & 0 & 0 \\ -1 & 0 & 0 \\ -2 & -2 & 0 \end{bmatrix} = \begin{bmatrix} 0 & -2 & 2 \\ 1 & -1 & 0 \\ 1 & -2 & 0 \end{bmatrix}$$

$\begin{bmatrix} 0 & -2 & 2 \\ -1 & 0 & -1 \\ -2 & -2 & 0 \end{bmatrix}$. Το D^{-1} υπολογίζεται αντιστρέφοντας τα στοιχεία της κύριας διαγωνίου του. Για να δούμε αν η

μέθοδος Jacobi συγκλίνει, εξετάζουμε αρχικά αν το μητρώο **A** έχει αυστηρή διαγώνια κυριαρχία κατά γραμμές ή κατά στήλες. Είναι φανερό ότι το **A** δεν έχει αυστηρά διαγώνια κυριαρχία κατά γραμμές ή κατά στήλες, αφού για την πρώτη γραμμή έχουμε ότι $1 < 4$ και για την πρώτη στήλη έχουμε ότι $1 < 3$. Στη συνέχεια εξετάζουμε αν το μητρώο **A** των συντελεστών είναι **ΜΑΔΚ**. Θα πρέπει να εφαρμόσουμε απαλοιφή Gauss για να μετατρέψουμε το **μητρώο A σε άνω τριγωνική μπλοκ μορφή**. Από τη στιγμή που το μητρώο των συντελεστών **A δεν περιέχει μηδενικά στοιχεία, δεν μπορεί να μετασχηματιστεί σε άνω τριγωνική μπλοκ μορφή με τη βοήθεια κάποιου μητρώου P, οπότε δεν ισχύει το 2^o κριτήριο σύγκλισης**. Στη συνέχεια εξετάζουμε αν το μητρώο **A** των συντελεστών είναι **Σ.Θ.Ο.** (3^o κριτήριο σύγκλισης). Παρατηρούμε ότι **δεν είναι καν συμμετρικό, οπότε δεν χρειάζεται να ελεγχθεί για θετική ορισμότητα**. **Το κοινό χαρακτηριστικό των τριών πρώτων κριτηρίων σύγκλισης είναι ότι εξετάζουν το μητρώο A των συντελεστών, ενώ τα τελευταία δύο εξετάζουν το μητρώο επαναληψης T**. Το επόμενο βήμα είναι να εξετάσουμε αν για κάποια στάθμη (νόρμα) του επαναληπτικού μητρώου **T** ισχύει ότι $\|T\| < 1$ (4^o κριτήριο σύγκλισης). Έχουμε ότι: $\|T\|_1 = \|T\|_\infty = 4 > 1$ και επομένως **δεν μπορούμε να συμπεράνουμε ότι συγκλίνει**. Απομένει να εξετάσουμε τη **φασματική ακτίνα του επαναληπτικού μητρώου T** (5^o κριτήριο σύγκλισης). Θα πρέπει ως γνωστό να αποδείξουμε ότι το $\rho(T) < 1$, για να μπορεί να εφαρμοστεί μια επαναληπτική μέθοδος. Γι' αυτό θα υπολογίσουμε τις ιδιοτιμές του μητρώου **T** από το χαρακτηριστικό πολυώνυμο του **T**, που προκύπτει από την εξίσωση $\det(T - \lambda * I) = 0$. Από αυτή την εξίσωση έχουμε

ότι: $\det(T - \lambda * I) = \begin{vmatrix} 0 - \lambda & -2 & 2 \\ -1 & 0 - \lambda & -1 \\ -2 & -2 & 0 - \lambda \end{vmatrix} = 0 \Rightarrow \begin{vmatrix} -\lambda & -2 & 2 \\ -1 & -\lambda & -1 \\ -2 & -2 & -\lambda \end{vmatrix} = 0$. Η παραπάνω εξίσωση με ανάπτυγμα Laplace, γράφεται ως: $-\lambda(\lambda^2 - 2) + (2\lambda + 4) - 2(2 + 2\lambda) = 0 \Rightarrow \lambda^3 = 0$. Επομένως, οι ιδιοτιμές είναι $\lambda_1 = \lambda_2 = \lambda_3 = 0$, οπότε: $\rho(T) = \max\{|\lambda_1|, |\lambda_2|, |\lambda_3|\} = \max\{0, 0, 0\} = 0$. Δηλαδή το $\rho(T) = 0 < 1$. Επομένως, η μέθοδος Jacobi συγκλίνει. Στη συνέχεια θα εφαρμόσουμε τη μέθοδο **Jacobi**.

Έτσι, ξεκινώντας από το αρχικό σύστημα και λύνοντας την πρώτη εξίσωση ως προς x_1 , τη δεύτερη ως προς x_2 και την τρίτη ως προς x_3 , θα έχουμε:

$$\begin{aligned} x_1 &= 1 - 2x_2 + 2x_3 \\ x_2 &= 1 - x_1 - x_3 \\ x_3 &= 1 - 2x_1 - 2x_2 \end{aligned}$$

Έτσι, αν θεωρήσουμε τις επαναλήψεις, έχουμε την παρακάτω επαναληπτική μορφή της μεθόδου Jacobi:

$$\begin{aligned} x_1^{(k+1)} &= 1 - 2x_2^{(k)} + 2x_3^{(k)} \\ x_2^{(k+1)} &= 1 - x_1^{(k)} - x_3^{(k)} \\ x_3^{(k+1)} &= 1 - 2x_1^{(k)} - 2x_2^{(k)} \end{aligned}$$

Αν στις παραπάνω σχέσεις θέσουμε $k = 0$, θα πάρουμε το $x^{(0)} = [x_1^{(0)} \quad x_2^{(0)} \quad x_3^{(0)}]^T = [1 \quad 1 \quad 1]^T$, και στη συνέχεια θα μπορούμε να πάρουμε διαδοχικά τις παρακάτω προσεγγίσεις λύσης: $x^{(1)} = [x_1^{(1)} \quad x_2^{(1)} \quad x_3^{(1)}]^T = [1 \quad -1 \quad -3]^T$, επομένως $\|x_1 - x_0\| = \|[1 - 1 - 3]_\infty^T - [1 \quad 1 \quad 1]_\infty^T\| = [0 \quad -2 \quad -4]\|_\infty = 4 > 10^{-m} = 10^{-2}$. Άρα συνεχίζουμε με τις επαναλήψεις και υπολογίζουμε για $k = 1$ το $x^{(2)} = [x_1^{(2)} \quad x_2^{(2)} \quad x_3^{(2)}]^T = [-3 \quad 3 \quad 1]^T$

και $x^{(3)} = [x_1^{(3)} \quad x_2^{(3)} \quad x_3^{(3)}]^T = [-3 \quad 3 \quad 1]^T$. Παρατηρούμε ότι οι δύο τελευταίες προσεγγίσεις ταυτίζονται, πράγμα το οποίο σημαίνει ότι ικανοποιούν το κριτήριο τερματισμού: $\frac{\|x^{(3)} - x^{(2)}\|_\infty}{\|x^{(3)}\|_\infty} \leq \frac{1}{2} 10^{-m}$ ή $\|x^{(3)} - x^{(2)}\| \leq 10^{-m}$ για την προσέγγιση της λύσης με τη σημαντικά ψηφία. Εύκολα μπορούμε να επαληθεύσουμε ότι η $x = [-3 \quad 3 \quad 1]^T$ είναι η ζητούμενη λύση, αφού ικανοποιεί όλες τις εξισώσεις του συστήματος.

2.9.2 Άσκηση εφαρμογής μεθόδου Jacobi (Β' Τρόπος)

Εφαρμόστε το επαναληπτικό σχήμα της μεθόδου Jacobi:

$$x^{(k+1)} = D^{-1}(L + U) * x^{(k)} + D^{-1}b, \quad k = 0, 1, 2, \dots$$

στο σύστημα των γραμμικών εξισώσεων

$$\begin{aligned} x_1^{(k+1)} &= 1 - 2x_2^{(k)} + 2x_3^{(k)} \\ x_2^{(k+1)} &= 1 - x_1^{(k)} - x_3^{(k)} \\ x_3^{(k+1)} &= 1 - 2x_1^{(k)} - 2x_2^{(k)} \end{aligned}$$

χρησιμοποιώντας ως αρχική εκτίμηση λύσης την $x^{(0)} = [1 \quad 1 \quad 1]^T$.

Λύση

Το επαναληπτικό μητρώο T είναι: $T = D^{-1} * (L + U) = \begin{bmatrix} 0 & -2 & 2 \\ -1 & 0 & -1 \\ -2 & -2 & 0 \end{bmatrix}$. Επίσης, το μητρώο D^{-1} ισούται με το

μητρώο D , αφού όλα τα διαγώνια στοιχεία του μητρώου D είναι μονάδες. Επίσης, έχουμε ότι το διάνυσμα $g =$

$$D^{-1} * b = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} * \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \text{ οπότε το επαναληπτικό σχήμα της μεθόδου Jacobi για το πρόβλημα που μας}$$

$$\text{δόθηκε είναι το εξής: } \begin{bmatrix} x_1^{(k+1)} \\ x_2^{(k+1)} \\ x_3^{(k+1)} \end{bmatrix} = D^{-1} * (L + U) * x^{(k)} + D^{-1} * b = \begin{bmatrix} 0 & -2 & 2 \\ -1 & 0 & -1 \\ -2 & -2 & 0 \end{bmatrix} \begin{bmatrix} x_1^{(k)} \\ x_2^{(k)} \\ x_3^{(k)} \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}. \text{ Αν στην παρα-}$$

πάνω επαναληπτική σχέση θέσουμε $x^{(0)} = [x_1^{(0)} \quad x_2^{(0)} \quad x_3^{(0)}]^T = [1 \quad 1 \quad 1]^T$, δηλαδή την αρχική εκτίμηση της λύσης που πάντα δίνεται από την εκφώνηση, μπορούμε να πάρουμε διαδοχικά τις παρακάτω προσεγγίσεις

$$\text{λύσης: Πιο συγκεκριμένα, για } k = 0 \text{ έχουμε: } x^{(1)} = \begin{bmatrix} x_1^{(1)} \\ x_2^{(1)} \\ x_3^{(1)} \end{bmatrix} = \begin{bmatrix} 0 & -2 & 2 \\ -1 & 0 & -1 \\ -2 & -2 & 0 \end{bmatrix} * \begin{bmatrix} x_1^{(0)} \\ x_2^{(0)} \\ x_3^{(0)} \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 & -2 & 2 \\ -1 & 0 & -1 \\ -2 & -2 & 0 \end{bmatrix} * \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} +$$

$$\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ -2 \\ -4 \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \\ -3 \end{bmatrix} \Rightarrow x^{(1)} = [x_1^{(1)} \quad x_2^{(1)} \quad x_3^{(1)}]^T = [1 \quad -1 \quad -3]^T. \text{ Με ανάλογο τρόπο υπολογίζουμε στη συ-}$$

νέχεια για $k = 1$, το $x^{(2)} = [x_1^{(2)} \quad x_2^{(2)} \quad x_3^{(2)}]^T = [-3 \quad 3 \quad 1]^T$ και για $k = 2$ το $x^{(3)} = [x_1^{(3)} \quad x_2^{(3)} \quad x_3^{(3)}]^T = [-3 \quad 3 \quad 1]^T$. Παρατηρούμε ότι η ζητούμενη λύση είναι η $x = [-3 \quad 3 \quad 1]^T$.

2.10 Μέθοδος Gauss - Seidel

Στη μέθοδο Gauss - Seidel το μητρώο ρυθμιστής είναι το $M = D - L$. Έτσι αν το αντίστροφο μητρώο $(D - L)^{-1}$ υπάρχει, το επαναληπτικό σχήμα της μεθόδου **Gauss - Seidel** είναι: $x^{(k+1)} = (D - L)^{-1} * U * x^{(k)} + (D - L)^{-1} * b$, όπου $k = 0, 1, 2, \dots$ με οποιοδήποτε $x^{(0)} \in C^n$. Όσον αφορά τη σύγκλιση, η ικανή συνθήκη για την σύγκλιση του επαναληπτικού σχήματος είναι το μητρώο A να έχει αυστηρή διαγώνια κυριαρχία κατά γραμμές ή κατά στήλες ή να ισχύει ότι: $\|T\| = \|(D - L)^{-1}U\| < 1$, ενώ η **αναγκαία και ικανή συνθήκη** για την σύγκλιση είναι η εξής: $\rho(T) = \rho((D - L)^{-1}U) < 1$. Επίσης έχουμε:

$$x^{(k+1)} = \frac{1}{a_{ii}} [b_i - \sum_{j=1}^{i-1} a_{ij} x_j^{(k+1)} - \sum_{j=i+1}^n a_{ij} x_j^{(k)}], \quad i = 1, \dots, n, \quad k = 0, 1, 2, \dots \text{ όπου } x_i^{(0)}, i = 1, \dots, n$$

2.11 Ανακεφαλαίωση – επανάληψη των επαναληπτικών μεθόδων

Υπενθυμίζουμε τα παρακάτω: Η επίλυση **πολύ μεγάλων γραμμικών συστημάτων** απαιτεί την ανάπτυξη **επαναληπτικών** μεθόδων που – με σωστή οργάνωση – επιβαρύνουν λιγότερο το υπολογιστικό και αποθηκευτικό σύστημα από τις άμεσες μεθόδους. Επιπλέον, πολλές φορές τα μητρώα που εμπλέκονται είναι **αραιά**, κάτι που μπορεί να **αξιοποιηθεί από τις επαναληπτικές μεθόδους**. Επίσης, αν τα μητρώα είναι **ειδικής μορφής**, όπως π.χ. τριγωνικά, τριδιαγώνια, διαγώνια, διδιαγώνια κ.λ.π. ενδείκνυται και πάλι η χρήση επαναληπτικών μεθόδων για την επίλυση του γραμμικού συστήματος $A * x = b$.

Ένας τυπικός τρόπος για την κατασκευή ορισμένων από αυτές (όπως τις κλασσικές μεθόδους όχι όμως στη CG) είναι η **διάσπαση του μητρώου και η γραφή του ως διαφορά δύο μερών**, ενός εύκολα αντιστρέψιμου και των υπολογίπων στοιχείων, π.χ. ως $Q = P - A$. Τότε μια επαναληπτική μέθοδος σύμφωνα με τον πρώτο γενικό τύπο, θα έχει τη μορφή $P * x^{(k+1)} = (P - A) * x^{(k)} + b \Rightarrow P * x^{(k+1)} = Q * x^{(k)} + b$ με εκκίνηση από κάποιο βολικό $x^{(0)}$. Κάθε βήμα της διαδικασίας εμπλέκει πολλαπλασιασμό με Q και επίλυση με P , και γι' αυτό θέλουμε το P να είναι τέτοιο ώστε η επίλυση συστημάτων με αυτό να είναι ταχεία (κάτι που συμβαίνει στην Jacobi στην οποία αντιστοιχεί διαγώνιο και στην Gauss - Seidel που αντιστοιχεί κάτω τριγωνικό P). Για παράδειγμα, αν γράψουμε $A = D - L - U$ όπου D είναι το διαγώνιο τμήμα του A , και $-L$ και $-U$ τα αυστηρά κάτω και άνω τριγωνικά τμήματα του A , τότε στη Jacobi, $P = D$ και $Q = L + U$. Στη Gauss-Seidel, $P = D - L$ και $Q = U$.

2.11.1 Άσκηση με Gauss - Seidel

Δίνεται το γραμμικό σύστημα:

$$\begin{aligned} 3x_1 + 2x_2 &= 5 \\ 2x_2 + x_3 &= 3 \\ x_1 + 2x_3 &= 3 \end{aligned}$$

Να δειχθεί ότι η μέθοδος των **Gauss - Seidel** συγκλίνει και να βρεθεί η λύση με προσέγγιση δύο σημαντικών ψηφίων ($m = 2$), χρησιμοποιώντας ως αρχική εκτίμηση την $x^{(0)} = [0 \quad 0 \quad 0]^T$.

Λύση

Το μητρώο των συντελεστών των αγνώστων και το σταθερό διάνυσμα είναι: $A = \begin{bmatrix} 3 & 2 & 0 \\ 0 & 2 & 1 \\ 1 & 0 & 2 \end{bmatrix}$, $b = \begin{bmatrix} 5 \\ 3 \\ 3 \end{bmatrix}$ αντί-

στοιχα. Για να ελέγξουμε αν η μέθοδος Gauss – Seidel συγκλίνει, εξετάζουμε αρχικά αν το μητρώο A των συντελεστών του γραμμικού συστήματος έχει **αυστηρή διαγώνια κυριαρχία κατά γραμμές ή κατά στήλες**. Είναι φανερό ότι το A έχει αυστηρή διαγώνια κυριαρχία κατά γραμμές (δεν μας πειράζει που δεν έχει κατά στήλες) και επομένως η μέθοδος **Gauss – Seidel συγκλίνει**. Στη συνέχεια θα εφαρμόσουμε τη μέθοδο Gauss – Seidel. Έτσι ξεκινώντας από το αρχικό σύστημα και λύνοντας την πρώτη εξίσωση ως προς x_1 , τη δεύτερη ως προς x_2 και την τρίτη ως προς x_3 , έχουμε:

$$\begin{aligned} x_1 &= \frac{1}{3} * [5 - 2x_2] \\ x_2 &= \frac{1}{2} [3 - x_3] \\ x_3 &= \frac{1}{2} [3 - x_1] \end{aligned}$$

Έτσι, αν θεωρήσουμε τις επαναλήψεις **τοποθετώντας στο αριστερό μέλος το $k+1$ και στο δεξιό μέλος το k** ή $k+1$, ανάλογα με το ποια είναι η **πιο πρόσφατη (ενημερωμένη) τιμή που έχει υπολογιστεί**, έχουμε την επαναληπτική μορφή της μεθόδου Gauss – Seidel για το συγκεκριμένο σύστημα:

$$\begin{aligned} x_1^{(k+1)} &= \frac{1}{3} [5 - 2x_2^{(k)}] \\ x_2^{(k+1)} &= \frac{1}{2} [3 - x_3^{(k)}] \\ x_3^{(k+1)} &= \frac{1}{2} [3 - x_1^{(k+1)}] \end{aligned}$$

Αντικαθιστώντας τα $x^{(k+1)}$ του δευτέρου μέλους με τα ίσα τους, οι παραπάνω σχέσεις μπορούν να γραφούν όπως παρακάτω: $x_1^{(k+1)} = \frac{1}{3} [5 - 2x_2^{(k)}]$, $x_2^{(k+1)} = \frac{1}{2} [3 - x_3^{(k)}]$, $x_3^{(k+1)} = \frac{1}{2} [3 - x_1^{(k+1)}]$, από τις οποίες τελικά έχουμε: $x_1^{(k+1)} = \frac{1}{3} [5 - 2x_2^{(k)}]$, $x_2^{(k+1)} = \frac{1}{2} [3 - x_3^{(k)}]$, $x_3^{(k+1)} = \frac{3}{2} - \frac{1}{2} x_1^{(k+1)}$. Αν στις παραπάνω σχέσεις θέσουμε $x^{(0)} = [x_1^{(0)} \quad x_2^{(0)} \quad x_3^{(0)}]^T = [0 \quad 0 \quad 0]^T$, μπορούμε να πάρουμε διαδοχικά τις παρακάτω προσεγγίσεις λύσης: $x^{(1)} = [1.667, 1.5, 0.66]^T$ στη συνέχεια υπολογίζουμε το απόλυτο σφάλμα $\|x^{(1)} - x^{(0)}\|_\infty = \| [1.667, 1.5, 0.66]^T - [0, 0, 0]^T \| = \| [1.667, 1.5, 0.667]^T \| = 2.3396$ και στη συνέχεια ελέγχουμε αν αυτό το αποτέλεσμα είναι μικρότερο από το $10^{-m} = 10^{-2} = 0.01$, κάτι που προφανώς δεν ισχύει. Άρα στη συνέχεια υπολογίζουμε το διάνυσμα $x^{(2)} =$

$$x^{(2)} = \begin{bmatrix} 0,667 \\ 1,167 \\ 1,167 \end{bmatrix}, x^{(3)} = \begin{bmatrix} 0,889 \\ 0,917 \\ 1,056 \end{bmatrix}, x^{(4)} = \begin{bmatrix} 1,055 \\ 0,972 \\ 0,973 \end{bmatrix}, x^{(5)} = \begin{bmatrix} 1,019 \\ 1,014 \\ 0,991 \end{bmatrix}, x^{(6)} = \begin{bmatrix} 0,991 \\ 1,004 \\ 1,004 \end{bmatrix}, x^{(7)} = \begin{bmatrix} 0,997 \\ 0,998 \\ 1,002 \end{bmatrix}, x^{(8)} = \begin{bmatrix} 1,002 \\ 0,999 \\ 0,999 \end{bmatrix}. \text{ Παρατηρούμε ότι εκτελέσαμε οκτώ επαναλήψεις της μεθόδου Gauss – Seidel μέχρι να πετύχουμε σύγκλιση, δηλ. μέχρι να βρεθεί λύση. Αυτό συμβαίνει διότι } \|x^{(8)} - x^{(7)}\|_\infty = \| [1.002, 0.999, 0.999]^T - [0.997, 0.998, 1.002]^T \| = \| [0.005, 0.001, -0.003]^T \|_\infty = 0.005 \leq 10^{-m} = 10^{-2} = 0.01 \text{ (ισχύει). Εναλλακτικά, αντί για το απόλυτο σφάλμα που ελέγχαμε προηγουμένως, θα μπορούσαμε να χρησιμοποιήσουμε το **σχετικό σφάλμα**: } \frac{\|x^{(8)} - x^{(7)}\|_\infty}{\|x^{(8)}\|_\infty} \leq \frac{1}{2} 10^{-m}. \text{ Αν θέλαμε το **απόλυτο σφάλμα** τότε: } \|x^{(8)} - x^{(7)}\|_\infty \leq 10^{-m} \text{ για την προσέγγιση λύσης με δύο σημαντικά ψηφία. Επομένως, η αριθμητική λύση του συστήματος που μας δόθηκε, με προσέγγιση δύο δεκαδικών ψηφίων, είναι } \eta x \approx [1.00 \quad 1.00 \quad 1.00]^T$$

Σημείωση: Η μέθοδος **Gauss – Seidel** συγκλίνει πιο γρήγορα από τη Jacobi (δηλαδή φθάνει ταχύτερα στη λύση με λιγότερα βήματα), διότι χρησιμοποιεί στοιχεία της ίδιας επανάληψης που είναι ήδη υπολογισμένα, ενώ στη Jacobi χρησιμοποιούμε στοιχεία της προηγούμενης επανάληψης που είναι πιο παλιά και συνεπώς πιο μακριά από τη λύση.

2.11.2 Άσκηση με Jacobi και Gauss-Seidel

Δίνεται ο πίνακας $A = \begin{bmatrix} 2 & 1 & 0 \\ 0 & 2 & 1 \\ 1 & 0 & 2 \end{bmatrix}$ και τα διανύσματα $b = \begin{bmatrix} 3 \\ 3 \\ 3 \end{bmatrix}$ και $x^{(0)} = \begin{bmatrix} 0.8 \\ 1.2 \\ 1.1 \end{bmatrix}$. α) Να γράψτε την επαναληπτική μέθοδο Jacobi για το σύστημα αυτό. β) Για το συγκεκριμένο πρόβλημα, η μέθοδος Jacobi συγκλίνει; γ) Να εφαρμόστε τη μέθοδο Jacobi για τον υπολογισμό των $x^{(1)}$ και $x^{(2)}$. δ) Να βρείτε την Ευκλείδεια νόρμα του σφάλματος για $x^{(0)}$ και $x^{(1)}$ ε) Να ξαναγίνουν τα ερωτήματα (γ) και (δ) για τη μέθοδο Gauss – Seidel.

Λύση

α) Το επαναληπτικό σχήμα της **Jacobi** είναι της μορφής: $x^{(\kappa+1)} = D^{-1}(L + U) * x^{(\kappa)} + D^{-1} * b$, για $\kappa = 0, 1, 2$. Το επαναληπτικό σχήμα της **Gauss – Seidel** είναι: $x^{(\kappa+1)} = (D - L)^{-1} * U * x^{(\kappa)} + (D - L)^{-1} * b$, για $\kappa = 0, 1, 2$.

β) Επειδή το μητρώο **A** έχει ΑΔΚ, η μέθοδος **Jacobi** συγκλίνει.

γ) Αρχικά υπολογίζουμε: $D = \text{diag}(\text{diag}(A)) = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix}$, $L = -\text{tril}(A, -1) = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{bmatrix}$, $U = -\text{triu}(A, 1) = \begin{bmatrix} 0 & -1 & 0 \\ 0 & 0 & -1 \\ 0 & 0 & 0 \end{bmatrix}$. Επίσης το άθροισμα $L + U = \begin{bmatrix} 0 & -1 & 0 \\ 0 & 0 & -1 \\ -1 & 0 & 0 \end{bmatrix}$, $D^{-1} = \begin{bmatrix} \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{2} \end{bmatrix}$ και τέλος το μητρώο επανάληψης $T = D^{-1} * (L + U) = \begin{bmatrix} \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{2} \end{bmatrix} * \begin{bmatrix} 0 & -1 & 0 \\ 0 & 0 & -1 \\ -1 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & -\frac{1}{2} & 0 \\ 0 & 0 & -\frac{1}{2} \\ -\frac{1}{2} & 0 & 0 \end{bmatrix}$ και το διάνυσμα $g = D^{-1} * b = \begin{bmatrix} \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{2} \end{bmatrix} * \begin{bmatrix} 3 \\ 3 \\ 3 \end{bmatrix} = \begin{bmatrix} \frac{3}{2} \\ \frac{3}{2} \\ \frac{3}{2} \end{bmatrix}$.

Γνωρίζοντας πλέον όλες τις τιμές, το επαναληπτικό σχήμα της Jacobi είναι: $x^{(k+1)} = T * x^k + g = \begin{bmatrix} 0 & -1/2 & 0 \\ 0 & 0 & -1/2 \\ -1/2 & 0 & 0 \end{bmatrix} * x^k + \begin{bmatrix} 3/2 \\ 3/2 \\ 3/2 \end{bmatrix} = \begin{bmatrix} 0 & -1/2 & 0 \\ 0 & 0 & -1/2 \\ -1/2 & 0 & 0 \end{bmatrix} * \begin{bmatrix} x_1^{(k)} \\ x_2^{(k)} \\ x_3^{(k)} \end{bmatrix} + \begin{bmatrix} 3/2 \\ 3/2 \\ 3/2 \end{bmatrix} = \begin{bmatrix} -\frac{1}{2}x_2^{(k)} \\ -\frac{1}{2}x_3^{(k)} \\ -\frac{1}{2}x_1^{(k)} \end{bmatrix} + \begin{bmatrix} 3/2 \\ 3/2 \\ 3/2 \end{bmatrix} \Rightarrow \begin{bmatrix} x_1^{(k+1)} \\ x_2^{(k+1)} \\ x_3^{(k+1)} \end{bmatrix} =$

$$\begin{bmatrix} -\frac{1}{2}x_2^{(k)} + \frac{3}{2} \\ -\frac{1}{2}x_3^{(k)} + \frac{3}{2} \\ -\frac{1}{2}x_1^{(k)} + \frac{3}{2} \end{bmatrix} \Rightarrow x_1^{(k+1)} = -\frac{1}{2}x_2^{(k)} + \frac{3}{2}, x_2^{(k+1)} = -\frac{1}{2}x_3^{(k)} + \frac{3}{2}, x_3^{(k+1)} = -\frac{1}{2}x_1^{(k)} + \frac{3}{2}.$$

κ = 0
 $x_1^{(1)} = -\frac{1}{2}x_2^{(0)} + \frac{3}{2} = -\frac{1}{2} * 1.2 + \frac{3}{2} = 0.9$, $x_2^{(1)} = -\frac{1}{2}x_3^{(0)} + \frac{3}{2} = -\frac{1}{2} * 1.1 + \frac{3}{2} = 0.95$, $x_3^{(1)} = -\frac{1}{2}x_1^{(0)} + \frac{3}{2} = -\frac{1}{2} * 0.8 + \frac{3}{2} = 1.1$. Άρα το $x^{(1)} = [0.9, 0.95, 1.1]^T$.

κ = 1
 $x_1^{(2)} = -\frac{1}{2}x_2^{(1)} + \frac{3}{2} = -\frac{1}{2} * 0.95 + \frac{3}{2} = 1.025$, $x_2^{(2)} = -\frac{1}{2}x_3^{(1)} + \frac{3}{2} = -\frac{1}{2} * 0.95 + \frac{3}{2} = 0.95$, $x_3^{(2)} = -\frac{1}{2}x_1^{(1)} + \frac{3}{2} = -\frac{1}{2} * 0.9 + \frac{3}{2} = 1.05$. Άρα το $x^{(2)} = [1.025, 0.95, 1.05]^T$.

δ) Η χρήση της Ευκλείδειας νόρμας μπορεί να γίνει τόσο για το απόλυτο όσο και για το σχετικό σφάλμα.

Μπορούμε να υπολογίσουμε οποιοδήποτε από τα δύο. Έστω ότι υπολογίζουμε το απόλυτο. Τότε έχουμε:

$$\|x^{(1)} - x^{(0)}\| = \|[0.9 \ 0.95 \ 1.1]^T - [0.8 \ 1.2 \ 1.1]^T\|_2 = \|[0.1 \ -0.25 \ 0]^T\|_2 = \sqrt{0.1^2 + (-0.25)^2 + 0^2} = 0.2692. \text{ Άν} \text{ μας} \text{ ζητούσε} \text{ η} \text{ εκφώνηση} \text{ την} \text{ Ευκλείδεια} \text{ νόρμα} \|x^{(2)} - x^{(1)}\|_2 = \|[1.025, 0.95, 1.05]^T - [0.9, 0.95, 1.1]^T\|_2 = \|[0.125, 0, -0.05]\|_2 \text{ και} \text{ στη} \text{ συνέχεια} \text{ θα} \text{ υπολογίσουμε} \text{ την} \text{ Ευκλείδεια} \text{ νόρμα} \text{ του} \text{ νέου} \text{ διανύσματος}, \text{ δηλαδή} \sqrt{0.125^2 + 0^2 + (-0.05)^2}.$$

ε) Έστω ότι θέλουμε να υπολογίσουμε τα $x^{(1)}, x^{(2)}$ με χρήση της μεθόδου **Gauss – Seidel**. Αρχικά υπολογίζουμε

το $D - L = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 1 & 0 & 2 \end{bmatrix}$ που είναι μητρώο **κάτω τριγωνικό** και ισχύει ότι το αντίστροφο μητρώο ενός κάτω τριγωνικού **είναι επίσης κάτω τριγωνικό** (διατηρείται η μορφή) και επιπλέον τα στοιχεία στην κύρια διαγώνιο του αντίστροφου μητρώου είναι τα αντίστροφα στοιχεία της κύριας διαγώνιου του κανονικού. Με άλλα λόγια

το μητρώο $(D - L)^{-1} = \begin{bmatrix} 1/2 & 0 & 0 \\ 0 & 1/2 & 0 \\ x_1 & 0 & 1/2 \end{bmatrix}$ το οποίο προέκυψε ως εξής: Ισχύει γενικά ότι $A * A^{-1} = A^{-1} * A = I$.

$$\text{Επομένως: } (D - L) * (D - L)^{-1} = I \Rightarrow \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 1 & 0 & 2 \end{bmatrix} * \begin{bmatrix} \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{2} & 0 \\ x_1 & 0 & \frac{1}{2} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}. \text{ Άρα το μητρώο } (D - L)^{-1} =$$

$$\begin{bmatrix} 1/2 & 0 & 0 \\ 0 & 1/2 & 0 \\ -1/4 & 0 & 1/2 \end{bmatrix}. \text{ Γνωρίζοντας πλέον τις τιμές των μητρώων, η γενική μορφή είναι: } x^{(k+1)} = \begin{bmatrix} x_1^{(k+1)} \\ x_2^{(k+1)} \\ x_3^{(k+1)} \end{bmatrix} =$$

$$T * x^{(k)} + g = (D - L)^{-1} * U * x^{(k)} + (D - L)^{-1} * b = \begin{bmatrix} \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{2} & 0 \\ -\frac{1}{4} & 0 & \frac{1}{2} \end{bmatrix} * \begin{bmatrix} 0 & -1 & 0 \\ 0 & 0 & -1 \\ 0 & 0 & 0 \end{bmatrix} * x^{(k)} + \begin{bmatrix} \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{2} & 0 \\ -\frac{1}{4} & 0 & \frac{1}{2} \end{bmatrix} *$$

$$\begin{bmatrix} 3 \\ 3 \\ 3 \end{bmatrix} = \begin{bmatrix} 0 & -\frac{1}{2} & 0 \\ 0 & 0 & -\frac{1}{2} \\ 0 & \frac{1}{4} & 0 \end{bmatrix} * \begin{bmatrix} x_1^{(k)} \\ x_2^{(k)} \\ x_3^{(k)} \end{bmatrix} + \begin{bmatrix} 1.5 \\ 1.5 \\ 0.75 \end{bmatrix} \Rightarrow \begin{bmatrix} x_1^{(k+1)} \\ x_2^{(k+1)} \\ x_3^{(k+1)} \end{bmatrix} = \begin{bmatrix} -\frac{1}{2} x_2^{(k)} \\ -\frac{1}{2} x_3^{(k)} \\ \frac{1}{4} x_2^{(k)} \end{bmatrix} + \begin{bmatrix} 1.5 \\ 1.5 \\ 0.75 \end{bmatrix} \Rightarrow \begin{bmatrix} x_1^{(k+1)} \\ x_2^{(k+1)} \\ x_3^{(k+1)} \end{bmatrix} = \begin{bmatrix} -\frac{1}{2} x_2^{(k)} + 1.5 \\ -\frac{1}{2} x_3^{(k)} + 1.5 \\ \frac{1}{4} x_2^{(k)} + 0.75 \end{bmatrix}. \text{ Στη}$$

$$x_1^{(k+1)} = -\frac{1}{2} x_2^{(k)} + 1.5 \\ x_2^{(k+1)} = -\frac{1}{2} x_3^{(k)} + 1.5 \\ x_3^{(k+1)} = \frac{1}{4} x_2^{(k+1)} + 0.75$$

συνέχεια εξισώνουμε τα αντίστοιχα στοιχεία των δύο διανυσμάτων

ποιούμε τώρα τη μέθοδο Gauss – Seidel, θα πρέπει στους όρους που εμφανίζονται δεξιά από το $=$, να λαμβάνουμε υπόψη τις **ενημερωμένες τιμές** τους και όχι τις αρχικές.

$$\mathbf{k=0:} x_1^{(1)} = -\frac{1}{2} x_2^{(0)} + 1.5 = -\frac{1}{2} * 1.2 + 1.5 = 0.9, x_2^{(1)} = -\frac{1}{2} x_3^{(0)} + 1.5 = -\frac{1}{2} * 1.1 + 1.5 = 0.95, x_3^{(1)} = \frac{1}{4} x_2^{(1)} + 0.75 = \frac{1}{4} * 0.95 + 0.75 = 0.275. \text{ Για } \mathbf{k=1:} x_1^{(2)} = -\frac{1}{2} x_2^{(1)} + 1.5 = -\frac{1}{2} * 0.95 + 1.5 = 1.025, x_2^{(2)} = -\frac{1}{2} x_3^{(1)} + 1.5 = -\frac{1}{2} * 0.275 + 1.5 = 1.3625, x_3^{(2)} = \frac{1}{4} x_2^{(2)} + 0.75 = \frac{1}{4} * 0.95 + 0.75 = 1.09. \text{ Η Ευκλείδεια (δεύτερη) νόρμα του σφάλματος είναι:} \|x^{(1)} - x^{(0)}\| = \|[0.9 \ 0.95 \ 1.05]^T - [1.025 \ 0.975 \ 0.9875]^T\| = 0.1420.$$

2.11.3 Άσκηση με Gauss-Seidel (θέμα από εξεταστική)

Αν χρησιμοποιήσουμε τη μέθοδο **Gauss – Seidel** για την προσέγγιση της λύσης του γραμμικού συστήματος A

$$* \mathbf{x} = \mathbf{b} \text{ όπου: } \mathbf{A} = \begin{pmatrix} 6 & 2 & 0 & 0 \\ -1 & 8 & 0 & 2 \\ 0 & -1 & 7 & 0 \\ 0 & 1 & 2 & 7 \end{pmatrix}, \text{ τότε με αρχική προσέγγιση } \mathbf{x}^{(0)} = \mathbf{0}, \text{ η προσέγγιση } \mathbf{x}^{(2)} \text{ στο δεύτερο βήμα}$$

υπολογίζεται βάσει του τύπου:

a) κανένα από τα υπόλοιπα

$$\text{b) } \mathbf{P}\mathbf{x}^{(2)} = \mathbf{Q}\mathbf{x}^{(1)} + \mathbf{b} \text{ όπου Q είναι κατάλληλα επιλεγμένο μητρώο και } \mathbf{P} = \begin{pmatrix} 6 & 0 & 0 & 0 \\ -1 & 8 & 0 & 0 \\ 0 & -1 & 7 & 0 \\ 0 & 1 & 2 & 7 \end{pmatrix}$$

$$\text{c) } \mathbf{P}\mathbf{x}^{(2)} = \mathbf{Q}\mathbf{x}^{(1)} + \mathbf{b} \text{ όπου Q είναι κατάλληλα επιλεγμένο μητρώο και } \mathbf{P} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1/3 & 1 & 0 & 0 \\ 1/7 & 1/7 & 1 & 0 \\ 1/4 & -1/8 & 0 & 1 \end{pmatrix}$$

$$\text{d) } \mathbf{P}\mathbf{x}^{(2)} = \mathbf{Q}\mathbf{x}^{(1)} + \mathbf{b} \text{ όπου Q είναι κατάλληλα επιλεγμένο μητρώο και } \mathbf{P} = \begin{pmatrix} 6 & 0 & 0 & 0 \\ 0 & 8 & 0 & 0 \\ 0 & 0 & 7 & 0 \\ 0 & 0 & 0 & 7 \end{pmatrix}$$

Λύση

Γνωρίζουμε ότι στη μέθοδο **Gauss - Seidel** –όπως άλλωστε σε όλες τις επαναληπτικές μεθόδους – ισχύει ο γενικός τύπος (ή αλλιώς επαναληπτικό βήμα): $\mathbf{P} * \mathbf{x}^{(k+1)} = (\mathbf{P} - \mathbf{A}) * \mathbf{x}^{(k)} + \mathbf{b}$, όπου $\mathbf{P} = \mathbf{D} - \mathbf{L}$. Στην προκειμένη περί-

$$\text{πτωση έχουμε ότι το μητρώο } \mathbf{D} = \text{diag}(\text{diag}(\mathbf{A})) = \begin{bmatrix} 6 & 0 & 0 & 0 \\ 0 & 8 & 0 & 0 \\ 0 & 0 & 7 & 0 \\ 0 & 0 & 0 & 7 \end{bmatrix}, \text{ U} = -\text{triu}(\mathbf{A}, 1) = \begin{bmatrix} 0 & -2 & 0 & 0 \\ 0 & 0 & 0 & -2 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \text{ L} = -\text{tril}(\mathbf{A}, -1) =$$

$$\begin{bmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & -1 & -2 & 0 \end{bmatrix} \text{ και το μητρώο } \mathbf{P} = \mathbf{D} - \mathbf{L} = \begin{bmatrix} 6 & 0 & 0 & 0 \\ 0 & 8 & 0 & 0 \\ 0 & 0 & 7 & 0 \\ 0 & 0 & 0 & 7 \end{bmatrix} - \begin{bmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & -1 & -2 & 0 \end{bmatrix} = \begin{bmatrix} 6 & 0 & 0 & 0 \\ -1 & 8 & 0 & 0 \\ 0 & -1 & 7 & 0 \\ 0 & 1 & 2 & 7 \end{bmatrix}. \text{ Αυτό το μητρώο}$$

παραμένει σταθερό, ανεξάρτητα από τον αριθμό των επαναλήψεων. Άρα η σωστή απάντηση είναι η (β).

Παρατήρηση: $\mathbf{A} * \mathbf{x} = \mathbf{b} \Rightarrow \mathbf{x} = \text{ακριβής λύση}, \text{ τότε } \mathbf{\delta} = \mathbf{A} * \mathbf{x} - \mathbf{b} = \mathbf{0}$. Αν όμως το \mathbf{x} είναι αποτέλεσμα μιας επαναληπτικής μεθόδου, όπως η Jacobi ή η Gauss – Seidel, και η Richardson, τότε η διαφορά $\mathbf{A} * \mathbf{x} - \mathbf{b} \neq \mathbf{0}$, με αποτέλεσμα να υπάρχει η διαφορά που ονομάζεται residual και είναι το μέγεθος $r = \mathbf{A} * \mathbf{x} - \mathbf{b}$.

2.12 Επαναληπτική μέθοδος Richardson

Η παράμετρος α_k σε κάθε βήμα υπολογίζεται ως

$$\alpha_k = \frac{(z^{(k)})^\top r^{(k)}}{(z^{(k)})^\top A z^{(k)}}, \text{ όπου } z^{(k)} = P^{-1}r^{(k)}$$

αποκαλείται προρυθμισμένο κατάλοιπο.

```
for k = 0, 1, ...
    Pz^(k) = r^(k),
    alpha_k = (z^(k))^\top r^(k)
              / (z^(k))^\top A z^(k),
    x^(k+1) = x^(k) + alpha_k z^(k),
    r^(k+1) = r^(k) - alpha_k A z^(k)
```

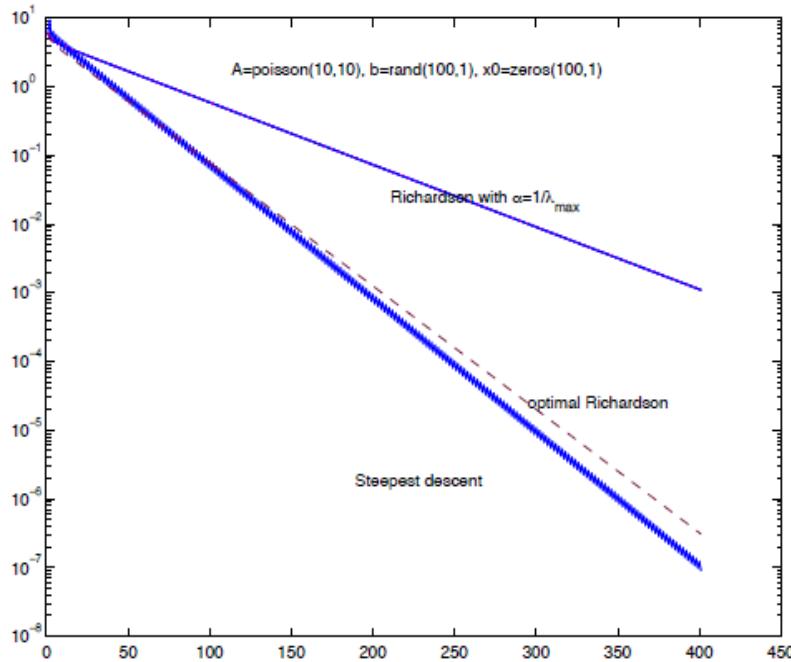
■ Η μέθοδος συγκλίνει αν A ΣΘΟ και ισχύει ότι

$$\|e^{(k)}\|_A \leq \left(\frac{\kappa(P^{-1}A) - 1}{\kappa(P^{-1}A) + 1} \right)^k \|e^{(0)}\|_A$$

όπου $\|x\|_A = \sqrt{x^\top Ax}$ είναι η «Α-νόμα» ή «νόμα ενέργειας».

■ Η ειδική περίπτωση $P = I$ λέγεται και μέθοδος απότομης καθόδου.

Για πραγματικά συμμετρικά ή μιγαδικά ερμιτιανά μητρώα η σύγκλιση της μεθόδου Richardson εξασφαλίζεται αν $|1 - \lambda(A)| < 1 \Rightarrow 0 < \lambda(A) < 2$ για κάθε ιδιοτιμή. Για γενικά μητρώα, η σύγκλιση της Richardson εξασφαλίζεται αν οι ιδιοτιμές του A περιέχονται στο εσωτερικό του μοναδιαίου δίσκου ή κύκλου, δηλαδή αυτού του κύκλου που έχει κέντρο στο σημείο 1 και ακτίνα 1. Στην Εικόνα 2 που ακολουθεί, φαίνονται οι γραφικές παραστάσεις για διάφορες παραλλαγές της μεθόδου Richardson για ένα μητρώο A που είναι ΣΘΟ:



Εικόνα 2: Γραφικές παραστάσεις για παραλλαγές της Richardson όσον αφορά την επιλογή της παραμέτρου α αναφορικά με τη σύγκλιση της μεθόδου

- 1) Richardson με τυπικό α
- 2) Richardson με βέλτιστο α
- 3) Δυναμική Richardson με μεταβαλλόμενο α (steepest descent)

Η ταχύτερη σύγκλιση επιτυγχάνεται στην τρίτη περίπτωση, δηλαδή με τη δυναμική Richardson με μεταβαλλόμενο α .

2.12.1 Πρώτη άσκηση με Jacobi, Gauss – Seidel και Richardson

Να υπολογιστεί η **πρώτη** επανάληψη των μεθόδων Jacobi, Gauss – Seidel και Gradient με προρυθμιστή (εδώ ανήκει η μέθοδος Richardson) για τη λύση του συστήματος: $2x_1 + x_2 = 1$ και $x_1 + 3x_2 = 0$, όταν $x^{(0)} = \begin{bmatrix} 1 & \frac{1}{2} \end{bmatrix}^T$.

Υπόδειξη: Ο προρυθμιστής δίνεται από τη διαγώνιο του μητρώου A.

Λύση

$$A * x = b \Rightarrow \begin{bmatrix} 2 & 1 \\ 1 & 3 \end{bmatrix} * \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

α) Jacobi:

$$\left. \begin{array}{l} x_1^{(1)} = \frac{1}{2}(1 - x_2^{(0)}) \\ x_2^{(1)} = -\frac{1}{3}x_1^{(0)} \end{array} \right\} \rightarrow \begin{array}{l} x_1^{(1)} = \frac{1}{2}\left(1 - \frac{1}{2}\right) = \frac{1}{4} \\ x_2^{(1)} = -\frac{1}{3} * 1 = -\frac{1}{3} \end{array}$$

Το ένα βήμα επίλυσης που ζητά η εκφώνηση, είναι το $x^{(1)} = \begin{bmatrix} x_1^{(1)} \\ x_2^{(1)} \end{bmatrix} = \begin{bmatrix} \frac{1}{4} \\ -\frac{1}{3} \end{bmatrix}$.

Άρα μετά από ένα βήμα της μεθόδου Jacobi έχουμε ότι: $x_1^{(1)} = \frac{1}{4}$ και $x_2^{(1)} = -\frac{1}{3}$

β) Gauss – Seidel:

$$\begin{aligned} x_1^{(1)} &= \frac{1}{2}(1 - x_2^{(0)}) = \frac{1}{2}\left(1 - \frac{1}{2}\right) = \frac{1}{4} \\ x_2^{(1)} &= -\frac{1}{3}x_1^{(1)} = -\frac{1}{3} * \frac{1}{4} = -\frac{1}{12} \end{aligned}$$

Άρα μετά από ένα βήμα της μεθόδου Gauss – Seidel έχουμε ότι: $x_1^{(1)} = \frac{1}{4}$ και $x_2^{(1)} = -\frac{1}{12}$, $x^{(1)} = \begin{bmatrix} x_1^{(1)} \\ x_2^{(1)} \end{bmatrix} = \begin{bmatrix} \frac{1}{4} \\ -\frac{1}{12} \end{bmatrix}$.

γ) Gradient (κατηγορία μεθόδων στην οποία ανήκει και η Richardson): Όσον αφορά τη μέθοδο **Gradient**, πρώτα υπολογίζουμε το (αρχικό) υπόλοιπο ή κατάλοιπο (residual): $r^{(0)} = b - Ax^{(0)} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} - \begin{bmatrix} 2 & 1 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} -\frac{3}{2} \\ -\frac{5}{2} \end{bmatrix}$.

Άρα ο προρυθμιστής P που ισούται με την κύρια διαγώνιο του A (από την εκφώνηση) είναι: $P = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix} \Rightarrow P^{-1} = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{3} \end{bmatrix}$, **επειδή το μητρώο P είναι διαγώνιο.** Άρα το $z^{(0)} = P^{-1} * r^{(0)} = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{3} \end{bmatrix} * \begin{bmatrix} -\frac{3}{2} \\ -\frac{5}{2} \end{bmatrix} = \begin{bmatrix} -\frac{3}{4} \\ -\frac{5}{6} \end{bmatrix}$ και ο συντελεστής

$$\alpha_0 = \frac{(z^{(0)})^T * r^{(0)}}{(z^{(0)})^T * A * z^{(0)}} = \frac{\begin{bmatrix} -\frac{3}{4} & -\frac{5}{6} \end{bmatrix} * \begin{bmatrix} -\frac{3}{2} \\ -\frac{5}{2} \end{bmatrix}}{\begin{bmatrix} -\frac{3}{4} & -\frac{5}{6} \end{bmatrix} * \begin{bmatrix} 2 & 1 \\ 1 & 3 \end{bmatrix} * \begin{bmatrix} -\frac{3}{4} \\ -\frac{5}{6} \end{bmatrix}} = \frac{77}{107}. \text{ Άρα το } x^{(1)} = x^{(0)} + \alpha_0 z^{(0)} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} + \frac{77}{107} * \begin{bmatrix} -\frac{3}{4} \\ -\frac{5}{6} \end{bmatrix} = \begin{bmatrix} \frac{197}{428} \\ -\frac{32}{321} \end{bmatrix}.$$

2.12.2 Δεύτερη άσκηση με Jacobi, Gauss – Seidel και Richardson

Να υπολογιστεί η **πρώτη επανάληψη** των μεθόδων Jacobi, Gauss – Seidel και Gradient με προρυθμιστή (εδώ ανήκει η μέθοδος Richardson) για τη λύση του συστήματος: $3x_1 + 2x_2 = 5$ και $-x_1 + 4x_2 = 2$, όταν $x^{(0)} = [1 \ 1]^T$. **Υπόδειξη:** Ο προρυθμιστής δίνεται από τη διαγώνιο του μητρώου A .

Λύση

$$A * x = b \Rightarrow \begin{bmatrix} 3 & 2 \\ -1 & 4 \end{bmatrix} * \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 5 \\ 2 \end{bmatrix}$$

α) Jacobi:

$$\begin{cases} x_1^{(1)} = -\frac{2}{3}x_2^{(0)} + \frac{5}{3} \\ x_2^{(1)} = \frac{1}{4}x_1^{(0)} + \frac{1}{2} \end{cases} \rightarrow \begin{cases} x_1^{(1)} = -\frac{2}{3} + \frac{5}{3} = 1 \\ x_2^{(1)} = \frac{1}{4} + \frac{1}{2} = \frac{3}{4} \end{cases},$$

Άρα μετά από ένα βήμα της μεθόδου Jacobi έχουμε ότι: $x_1^{(1)} = 1$ και $x_2^{(1)} = \frac{3}{4}$

β) Gauss – Seidel:

$$\begin{aligned} x_1^{(1)} &= -\frac{2}{3}x_2^{(0)} + \frac{5}{3} = -\frac{2}{3} + \frac{5}{3} = 1 \\ x_2^{(1)} &= \frac{1}{4}x_1^{(1)} + \frac{1}{2} = \frac{1}{4} + \frac{1}{2} = \frac{3}{4} \end{aligned}$$

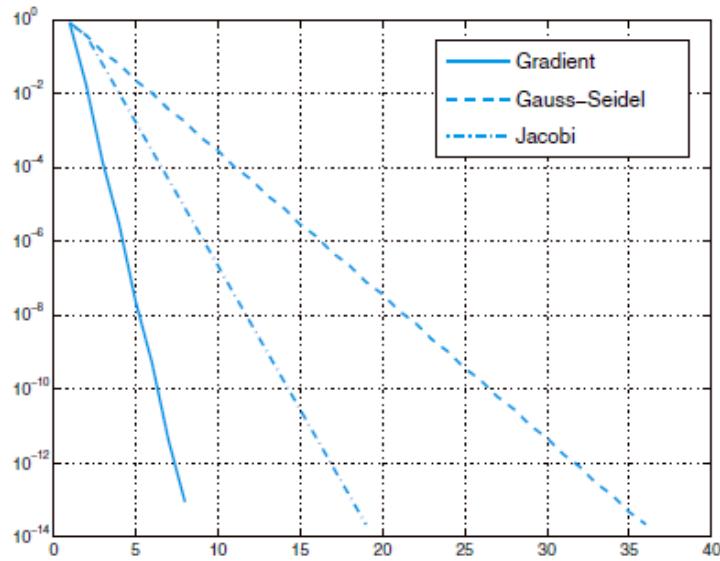
Άρα μετά από ένα βήμα της μεθόδου Gauss – Seidel έχουμε ότι: $x_1^{(1)} = 1$ και $x_2^{(1)} = \frac{3}{4}$

γ) Gradient (κατηγορία μεθόδων στην οποία ανήκει και η Richardson): Όσον αφορά τη μέθοδο **Gradient**, πρώτα υπολογίζουμε το αρχικό υπόλοιπο ή κατάλοιπο (residual): $r^{(0)} = b - Ax^{(0)} = \begin{bmatrix} 5 \\ 2 \end{bmatrix} - \begin{bmatrix} 3 & 2 \\ -1 & 4 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 5 \\ 2 \end{bmatrix} - \begin{bmatrix} 5 \\ 3 \end{bmatrix} = \begin{bmatrix} 0 \\ -1 \end{bmatrix}$. Άρα ο προρυθμιστής P που ισούται με την κύρια διαγώνιο του A (από την εκφώνηση) είναι: $P =$

$$\begin{bmatrix} 3 & 0 \\ 0 & 4 \end{bmatrix} \Rightarrow P^{-1} = \begin{bmatrix} \frac{1}{3} & 0 \\ 0 & \frac{1}{4} \end{bmatrix}. \text{ Άρα το } z^{(0)} = P^{-1} * r^{(0)} = \begin{bmatrix} \frac{1}{3} & 0 \\ 0 & \frac{1}{4} \end{bmatrix} * \begin{bmatrix} 0 \\ -1 \end{bmatrix} = \begin{bmatrix} 0 \\ -\frac{1}{4} \end{bmatrix} \text{ και ο συντελεστής } \alpha_0 = \frac{(z^{(0)})^T r^{(0)}}{(z^{(0)})^T A z^{(0)}} = \frac{\left[0 \quad -\frac{1}{4}\right] \left[\begin{array}{c} 0 \\ -1 \end{array}\right]}{\left[0 \quad -\frac{1}{4}\right] \left[\begin{array}{cc} 3 & 2 \\ -1 & 4 \end{array}\right] \left[\begin{array}{c} 0 \\ -\frac{1}{4} \end{array}\right]} = \frac{\frac{1}{4}}{\frac{1}{4}} = 1. \text{ Άρα το } x^{(1)} = x^{(0)} + \alpha_0 z^{(0)} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} + 1 * \begin{bmatrix} 0 \\ -\frac{1}{4} \end{bmatrix} = \begin{bmatrix} 1 \\ \frac{3}{4} \end{bmatrix}.$$

2.13 Ταχύτητα σύγκλισης επαναληπτικών μεθόδων

Με δεδομένο ότι το μητρώο $A = \begin{bmatrix} 2 & 1 \\ 1 & 3 \end{bmatrix}$ έχει Α.Δ.Κ. –άρα συγκλίνει τόσο η Jacobi όσο και η Gauss – Seidel – και επίσης είναι Σ.Θ.Ο. –άρα συγκλίνει η μέθοδος Richardson-, να συγκριθούν οι τρεις μέθοδοι μεταξύ τους από θέμα ταχύτητας σύγκλισης. Για τη σύγκλιση αυτή, εξετάζουμε την παρακάτω γραφική παράσταση:



Εικόνα 3: Σύγκριση ταχύτητας σύγκλισης επαναληπτικών μεθόδων

Από τη σύγκριση της ταχύτητας σύγκλισης των επαναληπτικών μεθόδων, παρατηρούμε ότι **τη γρηγορότερη ταχύτητα σύγκλισης τη διαθέτει η κατηγορία Gradient (εδώ ανήκει η μέθοδος Richardson)**. Ακολουθεί η Gauss – Seidel και τελευταία είναι η μέθοδος Jacobi. **Όσο μεγαλύτερη είναι η ταχύτητα σύγκλισης, τόσο γρηγορότερα φτάνει μια επαναληπτική μέθοδος στη λύση**. Η Gauss – Seidel συγκλίνει συνήθως πιο γρήγορα από τη μέθοδο Jacobi, διότι **χρησιμοποιεί στοιχεία της ίδιας επανάληψης που είναι ήδη υπολογισμένα**, ενώ στη Jacobi χρησιμοποιούνται στοιχεία της προηγούμενης επανάληψης που είναι πιο παλιά και συνεπώς πιο μακριά από τη λύση. Στη μέθοδο Gauss – Seidel χρησιμοποιούνται **όλες οι πιο πρόσφατες διαθέσιμες συνιστώσες**. Το τελευταίο χαρακτηριστικό της, δημιουργεί τη λανθασμένη εντύπωση ότι η μέθοδος αυτή πλεονεκτεί της μεθόδου Jacobi. **Αυτό όμως δεν είναι πάντα αληθές: Συγκεκριμένα, υπάρχουν περιπτώσεις όπου η μέθοδος Jacobi συγκλίνει, ενώ η μέθοδος Gauss – Seidel αποκλίνει. Επίσης, στη μέθοδο Gauss – Seidel δεν είναι δυνατό να βρεθούν όλες οι συνιστώσες της νέας επανάληψης σε ένα χρονικό βήμα, όπως στην περίπτωση της μεθόδου Jacobi.**

Παρατήρηση: Η φασματική ακτίνα $\rho(T)$ και ο αριθμός των επαναλήψεων που απαιτούνται για τη σύγκλιση μιας επαναληπτικής μεθόδου είναι **μεγέθη ανάλογα μεταξύ τους**. Όσο μικρότερη είναι η φασματική ακτίνα $\rho(T)$, τόσο λιγότερα βήματα επανάληψης έχουμε για τον υπολογισμό της λύσης, άρα τόσο ταχύτερη σύγκλιση πετυχαίνεται.

2.14 Άσκηση με επαναληπτικές μεθόδους

$$\text{Έστω το μητρώο } A = \begin{bmatrix} 0 & 1 & 2 & 3 \\ 3 & 0 & 1 & 2 \\ 2 & 3 & 0 & 1 \\ 1 & 2 & 3 & 0 \end{bmatrix}.$$

- a) Να εξηγήσετε αν εφαρμόζεται η μέθοδος **Jacobi** για την επίλυση του γραμμικού συστήματος $A * x = b$.
- b) Με αρχική προσέγγιση $x^{(0)} = [0 \ 0 \ 0 \ 0]^T$ να εφαρμόσετε τη μέθοδο Gauss-Seidel μια φορά, έτσι ώστε να προκύψει η νέα προσέγγιση $x^{(1)}$ για το γραμμικό σύστημα $B * x = b$, $B = 10 * I - A$ και $b = 4e$, όπου $e = [1 \ 1 \ 1]^T$.
- c) Γράφουμε το παραπάνω $B = D - L - U$, όπου $D = 10 * I$, $-L$ το αυστηρά κάτω τριγωνικό τμήμα του B (δηλαδή $\text{tril}(B, -1)$) και $-U$ το αυστηρά άνω τριγωνικό τμήμα του B (δηλ. $\text{triu}(B, 1)$). Τότε υπολογίζεται ότι: $(D - L)^{-1} = \begin{bmatrix} 1/10 & 0 & 0 & 0 \\ 3/100 & 1/10 & 0 & 0 \\ 29/1000 & 3/100 & 1/10 & 0 \\ 0.0247 & 29/1000 & 3/100 & 1/10 \end{bmatrix}$. Να ορίσετε την περιοχή του μιγαδικού επιπέδου όπου εξασφαλίζεται ότι βρίσκονται οι ιδιοτιμές του μητρώου $(D - L)^{-1} * U$. Να επαναλάβετε το ίδιο για το μητρώο $D^{-1} * (L + U)$.
- d) Να αιτιολογήσετε γιατί αν εφαρμόσετε τη μέθοδο Gauss – Seidel στο παραπάνω σύστημα $B * x = b$, η ακολουθία $x^{(1)}, x^{(2)}, x^{(3)}, \dots$ θα συγκλίνει στη λύση. Στη συνέχεια να αναφέρετε αν θα συγκλίνει η Jacobi για το ίδιο πρόβλημα.

Λύση

- a) Το πρώτο κριτήριο σύγκλισης **δεν** ισχύει, αφού δεν έχουμε ΑΔΚ κατά γραμμές ή στήλες. Αναφορικά με το δεύτερο κριτήριο σύγκλισης, αν και με τον πολλαπλασιασμό $P * A * P^T$ μπορούμε να φέρουμε το μητρώο σε

$$\text{μια μορφή block άνω τριγωνική } \begin{bmatrix} 0 & 1 & 2 & 3 \\ 3 & 0 & 1 & 2 \\ 2 & 3 & 0 & 1 \\ 1 & 2 & 3 & 0 \end{bmatrix} \rightarrow \begin{bmatrix} 3 & 1 & 2 & 0 \\ 2 & 0 & 1 & 3 \\ 1 & 3 & 0 & 2 \\ 0 & 2 & 3 & 1 \end{bmatrix}, \text{ στην οποία τα υπομητρώα } A_1 \text{ και } A_2 \text{ είναι μη -}$$

τετραγωνικά, με αποτέλεσμα το μητρώο **A** να είναι μη – αναγωγήσιμο (**MA**), αλλά όμως δεν έχουμε ΑΔΚ - όπως προβλέπεται- για μια τουλάχιστον γραμμή ή στήλη. Επομένως, **δεν** ισχύει ούτε το δεύτερο κριτήριο σύγκλισης που απαιτεί **όχι** μόνο μη – αναγωγησιμότητα, αλλά και ΑΔΚ μιας τουλάχιστον γραμμής ή στήλης (**MADK**). Επίσης, το μητρώο **A** δεν είναι ΣΘΟ. Επειδή **δεν** ισχύουν τα κριτήρια σύγκλισης αναφορικά με το μητρώο των συντελεστών **A**, θα υπολογίσουμε το μητρώο επανάληψης **T**, το οποίο δίνεται από τον τύπο: $T =$

$$D^{-1} * (L + U). \text{ Επειδή το μητρώο } D = \text{diag}(\text{diag}(A)) = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \text{ δεν αντιστρέφεται, επομένως } \text{δεν} \text{ μπορεί να}$$

υπολογιστεί το μητρώο επανάληψης **T**. Άρα **δεν** εφαρμόζεται η μέθοδος Jacobi.

Computer - Ανάλυση

$$b) \text{ Αρχικά υπολογίζουμε το μητρώο } B = 10 * I - A = \begin{bmatrix} 10 & 0 & 0 & 0 \\ 0 & 10 & 0 & 0 \\ 0 & 0 & 10 & 0 \\ 0 & 0 & 0 & 10 \end{bmatrix} - \begin{bmatrix} 0 & 1 & 2 & 3 \\ 3 & 0 & 1 & 2 \\ 2 & 3 & 0 & 1 \\ 1 & 2 & 3 & 0 \end{bmatrix} = \begin{bmatrix} 10 & -1 & -2 & -3 \\ -3 & 10 & -1 & -2 \\ -2 & -3 & 10 & -1 \\ -1 & -2 & -3 & 10 \end{bmatrix} \text{ και το}$$

διάνυσμα $b = 4 * e = [4 \ 4 \ 4 \ 4]^T$. Επειδή το μητρώο B **έχει Α.Δ.Κ. συγκλίνει η Gauss – Seidel**. Το μητρώο

$$\text{επανάληψης } T \text{ δίνεται από τον τύπο: } T = (D - L)^{-1} * U, \text{ το μητρώο } D = \text{diag}(\text{diag}(B)) = \begin{bmatrix} 10 & 0 & 0 & 0 \\ 0 & 10 & 0 & 0 \\ 0 & 0 & 10 & 0 \\ 0 & 0 & 0 & 10 \end{bmatrix} \text{ και το}$$

$$\text{μητρώο } L = -\text{tril}(B, -1) = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 3 & 0 & 0 & 0 \\ 2 & 3 & 0 & 0 \\ 1 & 2 & 3 & 0 \end{bmatrix} \text{ και το μητρώο } U = -\text{triu}(B, 1) = \begin{bmatrix} 0 & 1 & 2 & 3 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \text{ η διαφορά } D - L =$$

$$\begin{bmatrix} 10 & 0 & 0 & 0 \\ -3 & 10 & 0 & 0 \\ -2 & -3 & 10 & 0 \\ -1 & -2 & -3 & 10 \end{bmatrix}. \text{ Επειδή το μητρώο αυτό είναι **κάτω τριγωνικό**, για να υπολογίζουμε το αντίστροφό του (που}$$

είναι και αυτό κάτω τριγωνικό), θα χρησιμοποιήσουμε την εξίσωση: $(D - L)^{-1} * (D - L) = I \Rightarrow$

$$\begin{bmatrix} 1/10 & 0 & 0 & 0 \\ x_1 & 1/10 & 0 & 0 \\ x_2 & x_3 & 1/10 & 0 \\ x_4 & x_5 & x_6 & 1/10 \end{bmatrix} * \begin{bmatrix} 10 & 0 & 0 & 0 \\ -3 & 10 & 0 & 0 \\ -2 & -3 & 10 & 0 \\ -1 & -2 & -3 & 10 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \Rightarrow (D - L)^{-1} = \begin{bmatrix} 1/10 & 0 & 0 & 0 \\ 3/100 & 1/10 & 0 & 0 \\ 29/1000 & 3/100 & 1/10 & 0 \\ 0.0247 & 29/1000 & 3/100 & 1/10 \end{bmatrix}. \text{ Στη συνέχεια για}$$

τον υπολογισμό του μητρώου επανάληψης T θα βρούμε το γινόμενο $T = (D - L)^{-1} * U =$

$$\begin{bmatrix} 1/10 & 0 & 0 & 0 \\ 3/100 & 1/10 & 0 & 0 \\ 29/1000 & 3/100 & 1/10 & 0 \\ 0.0247 & 29/1000 & 3/100 & 1/10 \end{bmatrix} * \begin{bmatrix} 0 & 1 & 2 & 3 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 1/10 & 2/10 & 3/10 \\ 0 & 3/100 & 0.16 & 0.29 \\ 0 & 29/1000 & 0.088 & 0.247 \\ 0 & 0.0247 & 0.0784 & 0.1621 \end{bmatrix} \text{. Στη συνέχεια υπολογίζουμε το διάνυσμα}$$

$$g = (D - L)^{-1} * b = \begin{bmatrix} 1/10 & 0 & 0 & 0 \\ 3/100 & 1/10 & 0 & 0 \\ 29/1000 & 3/100 & 1/10 & 0 \\ 0.0247 & 29/1000 & 3/100 & 1/10 \end{bmatrix} * \begin{bmatrix} 4 \\ 4 \\ 4 \\ 4 \end{bmatrix} = \begin{bmatrix} \frac{4}{10} \\ \frac{52}{100} \\ \frac{0.636}{100} \\ 0.7348 \end{bmatrix}. \text{ Το επαναληπτικό σχήμα της επαναληπτικής μεθό-$$

δου **Gauss – Seidel** είναι το εξής: $x^{(k+1)} = T * x^{(k)} + g = (D - L)^{-1} * U * x^{(k)} + (D - L)^{-1} * b$, για $k = 0, 1, 2, \dots$

Άρα, μετά την αντικατάσταση των υπολογισθέντων μητρώων και διανυσμάτων έχουμε:

$$x^{(k+1)} = \begin{bmatrix} x_1^{k+1} \\ x_2^{k+1} \\ x_3^{k+1} \\ x_4^{k+1} \end{bmatrix} = \begin{bmatrix} 0 & 1/10 & 2/10 & 3/10 \\ 0 & 3/100 & 0.16 & 0.29 \\ 0 & 29/1000 & 0.088 & 0.247 \\ 0 & 0.0247 & 0.0784 & 0.1621 \end{bmatrix} * x^{(k)} + \begin{bmatrix} \frac{4}{10} \\ \frac{52}{100} \\ \frac{0.636}{100} \\ 0.7348 \end{bmatrix} = \begin{bmatrix} 0 & 1/10 & 2/10 & 3/10 \\ 0 & 3/100 & 0.16 & 0.29 \\ 0 & 29/1000 & 0.088 & 0.247 \\ 0 & 0.0247 & 0.0784 & 0.1621 \end{bmatrix} * \begin{bmatrix} x_1^k \\ x_2^k \\ x_3^k \\ x_4^k \end{bmatrix} + \begin{bmatrix} \frac{4}{10} \\ \frac{52}{100} \\ \frac{0.636}{100} \\ 0.7348 \end{bmatrix}. \text{ Στη}$$

συνέχεια θέτουμε όπου $k = 0$ για την πρώτη επανάληψη και έχουμε ότι μετά από ένα βήμα η λύση του γραμ-

$$\text{μικού συστήματος } B * x = b \text{ είναι } x^{(1)} = \begin{bmatrix} x_1^1 \\ x_2^1 \\ x_3^1 \\ x_4^1 \end{bmatrix} = \begin{bmatrix} 0 & 1/10 & 2/10 & 3/10 \\ 0 & 3/100 & 0.16 & 0.29 \\ 0 & 29/1000 & 0.088 & 0.247 \\ 0 & 0.0247 & 0.0784 & 0.1621 \end{bmatrix} * \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} \frac{4}{10} \\ \frac{52}{100} \\ \frac{0.636}{100} \\ 0.7348 \end{bmatrix} = \begin{bmatrix} \frac{4}{10} \\ \frac{52}{100} \\ \frac{0.636}{100} \\ 0.7348 \end{bmatrix}.$$

c) Το μητρώο $(D - L)^{-1} * U$ έχει υπολογιστεί από το προηγούμενο ερώτημα και είναι το μητρώο επανάληψης T

$$= \begin{bmatrix} 1/10 & 0 & 0 & 0 \\ 3/100 & 1/10 & 0 & 0 \\ 29/1000 & 3/100 & 1/10 & 0 \\ 0.0247 & 29/1000 & 3/100 & 1/10 \end{bmatrix} * \begin{bmatrix} 0 & 1 & 2 & 3 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 1/10 & 2/10 & 3/10 \\ 0 & 3/100 & 0.16 & 0.29 \\ 0 & 29/1000 & 0.088 & 0.247 \\ 0 & 0.0247 & 0.0784 & 0.1621 \end{bmatrix}.$$

Υπενθυμίζεται ότι για την εύρεση του διαστήματος σύγκλι-

σης των ιδιοτιμών, παίρνουμε από τις αριστερές (μικρότερες) τιμές τη μεγαλύτερη Ανίστοιχα, παίρνουμε από τις δε-

$$|z - 0| \leq \frac{1}{10} + \frac{2}{10} + \frac{3}{10} \Rightarrow -6/10 \leq z \leq 6/10$$

$$\left|z - \frac{3}{100}\right| \leq 0.16 + 0.29 \Rightarrow -0.42 \leq z \leq 0.48$$

$$|z - 0.088| \leq \frac{29}{100} + 0.247 \Rightarrow -0.188 \leq z \leq 0.364$$

$$|z - 0.1621| \leq 0.0247 + 0.0784 \Rightarrow 0.059 \leq z \leq 0.2652.$$

Η περιοχή του μιγαδικού επιπέδου όπου εξασφαλίζεται ότι βρίσκονται οι ιδιοτιμές του μητρώου $(D - L)^{-1} * U$ είναι $[0.059, 0.2652]$. **Εναλλακτικά** και πιο ολοκληρωμένα, μπορεί να χρησιμοποιηθεί η επόμενη διατύπωση: Από το θεώρημα των κύκλων (δίσκων) Gershgorin έπεται ότι οι ιδιοτιμές του μητρώου A (ή αλλιώς το φάσμα του μητρώου A) $\Lambda(A)$ θα είναι: $\Lambda(A) \subseteq (D_1 \cup D_2 \cup D_3 \cup D_4) \cap R$ όπου $D_1 = \{z, \text{όπου } |z - 0| \leq \frac{1}{10} + \frac{2}{10} + \frac{3}{10}\}$, $D_2 = \{z, \text{όπου } \left|z - \frac{3}{100}\right| \leq 0.16 + 0.29\}$, $D_3 = \{z, \text{όπου } |z - 0.088| \leq \frac{29}{100} + 0.247\}$ και $D_4 = \{z, \text{όπου } |z - 0.1621| \leq 0.0247 + 0.0784\}$ είναι οι δίσκοι (κύκλοι) Gershgorin και R είναι η ακτίνα τους. Η ακτίνα του πρώτου κύκλου είναι $6/10$, του δεύτερου είναι 0.45 , του τρίτου κύκλου είναι 0.276 και του τέταρτου κύκλου είναι 0.3254 . Από τη συναλήθευση των σχέσεων αυτών προκύπτει ότι $-6/10 \leq z \leq 6/10$ και $-0.42 \leq z \leq 0.48$ και $-0.188 \leq z \leq 0.364$, και $0.059 \leq z \leq 0.2652$, συνεπώς ισχύει ότι $z \geq 0.059$ και $z \leq 0.2652$. Από τη στιγμή που δεν έχουμε ομόκεντρους κύκλους Gershgorin, σημαίνει ότι το μητρώο T δεν έχει μιγαδικές ιδιοτιμές.

d) Επειδή το μητρώο B έχει A.D.K. **συγκλίνουν** η Gauss – Seidel και η Jacobi.

2.15 Άσκηση με επαναληπτικές μεθόδους

Ποια από τις παρακάτω συνθήκες ικανοποιεί την πρόταση «Αν ισχύει η συνθήκη τότε είναι **εξασφαλισμένο** ότι θα συγκλίνει η μέθοδος **Jacobi** αν επιχειρήσουμε να την εφαρμόσουμε για να επιλύσουμε το γραμμικό σύστημα $A * x = b$, όπου το A είναι αντιστρέψιμο».

α) $A = \text{diag}(\text{diag}(A))$, $L = -\text{tril}(A, -1)$, $U = -\text{triu}(A, 1)$ και το $D^{-1}(L + U)$ είναι διαγώνια κυρίαρχο. β) $A = \text{diag}(\text{diag}(A))$, $L = -\text{tril}(A, -1)$, $U = -\text{triu}(A, 1)$ και η μέγιστη **ιδιάζουσα** τιμή του $D^{-1}(L + U)$ είναι μικρότερη του 1. γ) $A = \text{diag}(\text{diag}(A))$, $L = -\text{tril}(A, -1)$, $U = -\text{triu}(A, 1)$ και το $\|(D - U)^{-1}L\|_1 < 1$. δ) Κανένα από τα υπόλοιπα.

Λύση

Ο όρος «**εξασφαλισμένο**» που διατυπώνεται στην εκφώνηση, σημαίνει ότι έχουμε να κάνουμε με **ικανή και αναγκαία συνθήκη** και στη μέθοδο Jacobi το μητρώο επανάληψης $T = D^{-1} * (L + U)$ και θα πρέπει η φασματική ακτίνα του, δηλαδή το **$\rho(T) < 1$** για να υπάρχει σύγκλιση της μεθόδου, δηλαδή για να βρίσκεται πάντα μια λύση για το γραμμικό σύστημα $A * x = b$. Άλλωστε αυτή η συνθήκη είναι ικανή και αναγκαία. Αυτό **δεν** επαληθεύεται με καμία από τις παραπάνω προτάσεις. Επομένως η σωστή απάντηση είναι η (δ). Η απάντηση (α) είναι λάθος διότι η A.D.K αναφέρεται αποκλειστικά στο μητρώο A των συντελεστών και όχι σε αυτό του ερωτήματος. Επίσης, το (β) δεν είναι σωστό, διότι ενώ είναι σωστό το μητρώο επανάληψης $D^{-1}(L + U)$ που δίνεται, θα έπρεπε να λέει

μέγιστη ιδιοτιμή ή αλλιώς φασματική ακτίνα και όχι μέγιστη ιδιάζουσα τιμή. Οι ιδιάζουσες τιμές ενός μητρώου είναι στη γενική περίπτωση διαφορετικές από τις ιδιοτιμές του μητρώου. Όπως έχει ήδη αναφερθεί, οι ιδιοτιμές ενός μητρώου υπολογίζονται από τις ρίζες του χαρακτηριστικού πολυωνύμου του μητρώου, δηλαδή από το $p(\lambda) = \det(\lambda * I - A) = 0$. Αντίστοιχα, **οι ιδιάζουσες τιμές ενός μητρώου (που στην περίπτωση αυτή δεν είναι κατ' ανάγκη τετραγωνικό, όπως στην περίπτωση των ιδιοτιμών)** υπολογίζονται από την λεγόμενη **ιδιάζουσα παραγοντοποίηση** ενός μητρώου, που είναι $A = U * \Sigma * V^T$. Μόνο αν το μητρώο είναι συμμετρικό, οι ιδιάζουσες τιμές είναι ίδιες με τις ιδιοτιμές, διαφορετικά ισχύει ότι το $\sigma_i = \sqrt{\lambda_i}$. Τέλος, το (γ) δεν είναι σωστό, **διότι δεν είναι αυτό το μητρώο επανάληψης T** της, μεθόδου Gauss – Seidel που αναφέρεται μέσα στη νόρμα. **Ακόμη και αν το μητρώο T είχε δοθεί στη σωστή μορφή του, δηλ. $T = D^{-1}(L + U)$, οπότε $\|D^{-1}(L + U)\|_1 < 1$, η απάντηση (γ) θα ήταν λάθος, διότι αποτελεί μόνο ικανή και όχι αναγκαία συνθήκη.** Όμως ο όρος **εξασφαλισμένο** παραπέμπει σε ικανή και αναγκαία συνθήκη, που είναι μόνο η τελευταία ($\rho(T) < 1$).

2.16 Άσκηση με επαναληπτικές μεθόδους

Δίνεται το παρακάτω γραμμικό σύστημα: $\begin{bmatrix} 1 & -2 & 2 \\ -1 & 1 & -1 \\ -2 & -2 & 1 \end{bmatrix} * \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -3 \\ -1 \\ -3 \end{bmatrix}$

- Μπορείτε να το επιλύσετε χρησιμοποιώντας την επαναληπτική μέθοδο **Gauss - Seidel**;
- Μήπως γίνεται με τη μέθοδο **Jacobi**; Εφαρμόστε αυτή τη μέθοδο με το χέρι, ξεκινώντας από το $x_0 = [0, 0, 0]'$.
- Γράψτε δύο δικές σας συναρτήσεις σε **MATLAB** που να υλοποιούν τις δύο παραπάνω μεθόδους και πειραματιστείτε με δικά σας γραμμικά συστήματα. Προσοχή, σε κάθε επανάληψη θα πρέπει να υπολογίζετε και τα αντίστοιχα σφάλματα. Μπορείτε να αποφανθείτε αν κάποια μέθοδος είναι καλύτερη;

Λύση

- Αρχικά θα πρέπει να δούμε αν η μέθοδος των Gauss – Seidel συγκλίνει. Ένα **πρώτο κριτήριο σύγκλισης** είναι να δείξουμε ότι το μητρώο των συντελεστών του γραμμικού συστήματος $\begin{bmatrix} 1 & -2 & 2 \\ -1 & 1 & -1 \\ -2 & -2 & 1 \end{bmatrix}$ έχει αυστηρή διαγώνια κυριαρχία κατά γραμμές ή κατά στήλες, κάτι που όπως παρατηρούμε **δεν ισχύει**.

Ένα 2^{o} κριτήριο σύγκλισης είναι το μητρώο να είναι **μη – αναγωγήσιμο και διαγώνια κυρίαρχο με αυστηρή διαγώνια κυριαρχία για τουλάχιστον μια γραμμή ή στήλη (ΜΑΔΚ)**. Ωστόσο ο δεύτερος ισχυρισμός (αυστηρή διαγώνια κυριαρχία για τουλάχιστον μια γραμμή ή στήλη) **δεν ικανοποιείται επειδή δεν μπορεί να βρεθεί μητρώο εναλλαγής P** τέτοιο ώστε $P * A * P^T = \begin{bmatrix} A_1 & A_2 \\ 0 & A_3 \end{bmatrix}$, αφού το μητρώο A δεν έχει μηδενικά στοιχεία. Πρέπει τα υπομητρώα A_1 και A_3 να είναι μη τετραγωνικά. Επίσης το μητρώο **δεν είναι** Συμμετρικά Θετικά Ορισμένο (ΣΘΟ), οπότε δεν ισχύει ούτε αυτό το κριτήριο σύγκλισης.

Το επόμενο κριτήριο σύγκλισης που θα δοκιμάσουμε είναι μια οποιαδήποτε νόρμα του μητρώου επανάληψης T να είναι μικρότερη της μονάδας. Γνωρίζουμε ότι το μητρώο $T = (D - L)^{-1} * U$, όπου $D = \text{diag}(\text{diag}(A)) =$

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad L = -\text{tril}(A, -1) = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 2 & 2 & 0 \end{bmatrix}, \quad U = -\text{triu}(A, 1) = \begin{bmatrix} 0 & 2 & -2 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}. \quad \text{Επομένως: } D - L = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} - \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 2 & 2 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ -2 & -2 & 1 \end{bmatrix}, \quad \text{Ισχύει ότι το μητρώο } (D - L)^{-1} * (D - L) = I \Rightarrow \begin{bmatrix} 1 & 0 & 0 \\ x & 1 & 0 \\ y & z & 1 \end{bmatrix} * \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ -2 & -2 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}. \quad \text{Άρα: } x - 1 = 0 \Rightarrow x = 1, \quad z - 2 = 0 \Rightarrow z = 2 \quad \text{και} \quad y - z - 2 = 0 \Rightarrow y = 4. \quad \text{Επιστρέφοντας στη σχέση (1) έχουμε ότι το μητρώο επανάληψης } T = (D - L)^{-1} * U = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 4 & 2 & 1 \end{bmatrix} * \begin{bmatrix} 0 & 2 & -2 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} = T = \begin{bmatrix} 0 & 2 & -2 \\ 0 & 2 & -1 \\ 0 & 8 & -6 \end{bmatrix}.$$

αυτό το κριτήριο σύγκλισης². Το τελευταίο κριτήριο σύγκλισης (ικανό και αναγκαίο) που απομένει είναι εξετάζουμε αν η φασματική ακτίνα του μητρώου επανάληψης T είναι μικρότερη της μονάδας. Για το σκοπό αυτό θα πρέπει **πρώτα** να υπολογίσουμε το **χαρακτηριστικό πολυώνυμο**, μέσω του οποίου θα βρούμε τη **φασματική ακτίνα του μητρώου T** . Έχουμε: $\det(\lambda * I - T) = 0 \Rightarrow \det \left(\begin{bmatrix} \lambda & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \lambda \end{bmatrix} - \begin{bmatrix} 0 & 2 & -2 \\ 0 & 2 & -1 \\ 0 & 8 & -6 \end{bmatrix} \right) = 0 \Rightarrow \lambda * ((\lambda - 2)(\lambda + 6) + 8) = 0 \Rightarrow \lambda(\lambda^2 + 6\lambda - 2\lambda - 12 + 8) = 0 \Rightarrow \lambda * (\lambda^2 + 4\lambda - 4) = 0$. Επιλύοντας το σύστημα έχουμε

ότι: $\begin{cases} \lambda_1 = 0 \\ \lambda_2 = -4,8 \\ \lambda_3 = 0,82 \end{cases}$. Επομένως η φασματική του μητρώου επανάληψης T -που είναι η μεγαλύτερη κατ' απόλυτη

τιμή ιδιοτιμή του μητρώου- είναι $\rho(T) = 4,8 > 1$, επομένως η Gauss – Seidel **δεν συγκλίνει**.

(ii) Θα πρέπει **πρώτα** να εξετάσουμε **αν** συγκλίνει η μέθοδος Jacobi. Από τα τρία πρώτα κριτήρια που χρησιμοποιήσαμε και πριν, προκύπτει ότι **δεν συγκλίνει η μέθοδος Jacobi καθώς αυτά αναφέρονται στο μητρώο A των συντελεστών που δίνεται από την εκφώνηση, το οποίο παραμένει το ίδιο**. Ωστόσο, στη μέθοδο Jacobi το μητρώο επανάληψης T υπολογίζεται διαφορετικά, επομένως θα εξετάσουμε τα δύο τελευταία κριτήρια σύγκλισης. Για την επαναληπτική μέθοδο **Jacobi** ισχύει ότι:

$$T = D^{-1} * (L + U) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} * \left(\begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 2 & 2 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 2 & -2 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \right) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} * \begin{bmatrix} 0 & 2 & -2 \\ 1 & 0 & 1 \\ 2 & 2 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 2 & -2 \\ 1 & 0 & 1 \\ 2 & 2 & 0 \end{bmatrix}$$

Η πρώτη νόρμα και η νόρμα μεγίστου του μητρώου επανάληψης T είναι: $\|T\|_1 = \|T\|_\infty = 4 > 1$, άρα θα πρέπει να εξετάσουμε και το κριτήριο σύγκλισης για τη φασματική ακτίνα. Έτσι έχουμε ότι: $\det(T - \lambda * I) = 0 \Rightarrow \det \left(\begin{bmatrix} 0 & 2 & -2 \\ 1 & 0 & 1 \\ 2 & 2 & 0 \end{bmatrix} - \begin{bmatrix} \lambda & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \lambda \end{bmatrix} \right) = (-\lambda) * (\lambda^2 - 2) - (-2\lambda + 4) + 2(2 - 2\lambda) = 0 \Rightarrow -\lambda(\lambda^2 - 2 + 2) = 0 \Rightarrow \lambda^3 = 0$. Από την επίλυση της εξίσωσης αυτής προκύπτει ότι **$\lambda = 0$ (τριπλή ρίζα)**. Άρα $\rho(T) = 0 < 1$. Επομένως μπορούμε να χρησιμοποιήσουμε τη μέθοδο Jacobi. Για τη μέθοδο **Jacobi** έχουμε τον αναδρομικό τύπο της:

² Αν έστω η μία από τις δύο νόρμες που εξετάσαμε ήταν μικρότερη από «1», θα υπήρχε σύγκλιση της μεθόδου.

$$x^{(k+1)} = Tx^{(k)} + g \Rightarrow x^{(k+1)} = Tx^{(k)} + D^{-1}b \Rightarrow \begin{bmatrix} x_1^{(k+1)} \\ x_2^{(k+1)} \\ x_3^{(k+1)} \end{bmatrix} = \begin{bmatrix} 0 & 2 & -2 \\ 1 & 0 & 1 \\ 2 & 2 & 0 \end{bmatrix} * \begin{bmatrix} x_1^{(k)} \\ x_2^{(k)} \\ x_3^{(k)} \end{bmatrix} + \begin{bmatrix} -3 \\ -1 \\ -3 \end{bmatrix}. \text{ Αντικαθιστώντας για}$$

$\kappa=0$, το αρχικό διάνυσμα $x_0 = [0, 0, 0]^T$ που δίνεται από την άσκηση, έχουμε ότι: $\begin{bmatrix} x_1^{(1)} \\ x_2^{(1)} \\ x_3^{(1)} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} -3 \\ -1 \\ -3 \end{bmatrix} = \begin{bmatrix} -3 \\ -1 \\ -3 \end{bmatrix}$.

(iii) Γενικά η μέθοδος Gauss – Seidel υπερτερεί έναντι της μεθόδου Jacobi καθώς έχει μεγαλύτερη ταχύτητα

σύγκλισης. Ακολουθούν δύο συναρτήσεις σε Matlab για τις μεθόδους Gauss – Seidel και Jacobi. Κάθε συνάρτηση στη MATLAB, καλείται από τη γραμμή εντολών της MATLAB με το όνομά της. Θα πρέπει επίσης, πριν την καλέσουμε, να φροντίσουμε να έχουμε ορίσει όλα τα ορίσματα εισόδου της συνάρτησης.

```
function [x, error, iter, flag] = jacobi(A, x, b, max_it, tol)
% [x, error, iter, flag] = jacobi(A, x, b, max_it, tol)
% jacobi.m solves the linear system Ax = b using the Jacobi Method.
% Input A      REAL matrix
%       x      REAL initial guess vector
%       b      REAL right hand side vector (διάνυσμα με τους σταθερούς όρους)
%       max_it INTEGER maximum number of iterations (όριο επαναλήψεων για ασφάλεια)
%       tol    REAL error tolerance (ανοχή σφάλματος δηλ. το μέγιστο άνω φράγμα σφ.)
% Output x      REAL solution vector
%       error  REAL error norm (νόρμα σφάλματος)
%       iter   INTEGER number of iterations performed
%       flag   INTEGER: 0 = solution found to tolerance (βρέθηκε λύση)
%                  1 = no convergence given max_it (δεν βρέθηκε λύση)
iter = 0; flag = 0;% initialization
bnrm2 = norm(b); %δεύτερη νόρμα
if (bnrm2 == 0.0), bnrm2 = 1.0; end
r = b - A * x; %r = residual = υπόλοιπο ή κατάλοιπο, το οποίο δείχνει την απόκλιση της λύσης από την ακριβή λύση (λύση με άπειρη ακρίβεια). Όσο μικρότερη θτιμή έχει το residual, τόσο πιο ακριβής είναι η λύση που υπολογίζουμε.
error = norm(r)/bnrm2; %υπολογισμός σφάλματος
if (error < tol) return, end % αν το σφάλμα της λύσης < προσδιοριζόμενη ανοχή σφάλματος, %τότε επιστροφή από τη συνάρτηση (ολοκλήρωση συνάρτησης).
[m, n]=size(A); %m = γραμμές του A, n = στήλες του A
D = diag(diag(A)); N = D-A; M = A; % matrix splitting
for iter = 1:max_it, % begin iteration
    x_1 = x; %update approximation (υπολογισμός νέας λύσης)
    x = M \ (N*x + b);
    error = norm(x - x_1)/norm(x); % compute error
    if (error <= tol), break, end % check convergence (έλεγχος σύγκλισης)
end
if (error > tol) flag = 1; end % no convergence (δεν υπάρχει σύγκλιση)
% END jacobi.m
```

Στη συνέχεια, καλούμε τη function αυτή από τη γραμμή εντολών της MATLAB, αφού πρώτα δώσουμε τιμές (περιεχόμενο) στα ορίσματα εισόδου της function.

```
>> A=[1 -2 2;-1 1 -1;-2 -2 1]
A =
  1   -2   2
 -1   1   -1
 -2   -2   1
>> b=[-3;-1;-3]
b =
```

```

-3
-1
-3
>> x=[0;0;0]
x =
    0
    0
    0
>> max_it=50
max_it = 50
>> tol=1e-6
tol = 1.0000e-06

```

Στη συνέχεια, καλούμε τη function με τον τρόπο που περιγράφεται παρακάτω. Μας επιστρέφονται όλα τα αποτελέσματα που περικλείονται μέσα σε [].

```

>> [x, error, iter, flag] = jacobi(A, x, b, max_it, tol)
x =
    21.0000
   -29.0000
   -35.0000
error = 6.2263e-16
iter = 4
flag = 0

```

Στη συνέχεια ακολουθεί ο κώδικας για τη μέθοδο Gauss – Seidel, με τη μορφή μιας function που ονομάζεται «*mygs*». Θα πρέπει να καλέσουμε τη συγκεκριμένη συνάρτηση με παρόμοιο τρόπο από τη γραμμή εντολών της MATLAB, φροντίζοντας στην αρχή να προσδιορίσουμε τις τιμές όλων των παραμέτρων εισόδου, οι οποίες είναι ακριβώς ίδιες με αυτές της προηγούμενης συνάρτησης Jacobi. Θα πρέπει η αποθήκευση κάθε συνάρτησης ως αρχείο να γίνεται με το όνομα της συνάρτησης.

```

function [x,XALL, error, iter] = mygs(A, x, b, max_it, tol)
% [x, XALL, error, iter] = mygs(A, x, b, max_it, tol)
% mygs.m solves the linear system Ax=b using the Gauss-Seidel Method.
%
% input   A      REAL matrix
%         x      REAL initial guess vector
%         b      REAL right hand side vector
% max_it  INTEGER maximum number of iterations
% tol     REAL error tolerance
%
% output  x      REAL solution vector
%         XALL  REAL matrix containing the approximations of each step
%         error REAL error norm
%         iter  INTEGER number of iterations performed
%
iter = 0;                                     % initialization
[n,m]=size(A);
bnrm2 = norm( b );
if ( bnrm2 == 0.0 ), bnrm2 = 1.0; end

r = b - A*x; error = norm( r ) / bnrm2;
if ( error < tol ) return, end                %check if initial guess is correct

M=diag(diag(A))-tril(A,-1); N=M\A;           % matrix splitting

for iter= 1:max_it,
    Xprev = x;                                % save current values to calculate error later
    for i = 1:n
        j = 1:n; j(i) = []; %get all the coefficients except A(i,i)
        Xtemp = x; Xtemp(i) = []; %same as before for the variables
        x(i) = (b(i) - sum(A(i,j) * Xtemp)) / A(i,i);
    end
    XALL(:,iter) = x;
    error = norm( x - Xprev ) / norm( Xprev );% compute error
    if ( error <= tol ), break, end            % check convergence
end
%END mygs.m

```

2.17 Άσκηση με επαναληπτικές μεθόδους

Να χρησιμοποιηθεί η μέθοδος **Gauss – Seidel** για την επίλυση του γραμμικού συστήματος

$$10x_1 - x_2 + 2x_3 = 6$$

$$-x_1 + 11x_2 - x_3 + 3x_4 = 25$$

$$2x_1 - x_2 + 10x_3 - x_4 = -11.$$

$$3x_2 - x_3 + 8x_4 = 15 \text{ με αρχική προσέγγιση } x = [0 \ 0 \ 0 \ 0]^T \text{ και επαναλήψεις 3 σ.ψ. για το σχετικό σφάλμα.}$$

Λύση

k	0	1	2	3	4	5
$x_1^{(k)}$	0.0000	0.6000	1.030	1.0065	1.0009	1.0001
$x_2^{(k)}$	0.0000	2.3272	2.037	2.0036	2.0003	2.0000
$x_3^{(k)}$	0.0000	-0.9873	-1.014	-1.0025	-1.0003	-1.0000
$x_4^{(k)}$	0.0000	0.8789	0.984	0.9983	0.9999	1.0000

2.18 Άσκηση με επαναληπτικές μεθόδους (Παλιό Θέμα)

Αν χρησιμοποιήσουμε τη μέθοδο **Jacobi** για την προσέγγιση της λύσης του γραμμικού συστήματος $R * x = b$, είναι γνωστό ότι η προσέγγιση $x^{(k)}$ σε κάθε βήμα μπορεί να υπολογιστεί βάσει του τύπου $Px^{(k)} = Qx^{(k-1)} + b$, για

κατάλληλα επιλεγμένα P, Q . Έστω ότι το $R = \begin{bmatrix} 6 & 2 & -1 & 2 \\ 2 & 6 & 2 & 2 \\ 1 & 1 & 7 & 1 \\ 2 & -1 & 0 & 8 \end{bmatrix}$ και $x^{(0)} = [0 \ 0 \ 0 \ 0]^T$. a) Συμπληρώστε τα μητρώα

P, Q με τις ακριβείς τιμές για τη συγκεκριμένη μέθοδο, b) να γράψτε συνάρτηση MATLAB [`iconv`] = `test_conv(A)` που επιστρέφει `iconv = 1`, αν η παραπάνω μέθοδος για την επίλυση του γραμμικού συστήματος $A * x = b$ **συγκλίνει**, διαφορετικά επιστρέφει `iconv = -1`. Ενδείκνυται να χρησιμοποιήσετε τη συνάρτηση `eig`. c) Να ελέγξετε αν η μέθοδος θα συγκλίνει στην επίλυση του γραμμικού συστήματος. Αν δεν συγκλίνει να εξηγήσετε γιατί.

Λύση

a) Επειδή χρησιμοποιείται η μέθοδος **Jacobi** για την προσέγγιση της λύσης του γραμμικού συστήματος $R * x = b$, είναι γνωστό ότι η προσέγγιση $x^{(k)}$ σε κάθε βήμα της διαδικασίας επίλυσης μπορεί να υπολογιστεί βάσει του δοθέντος τύπου $P * x^{(k)} = Q * x^{(k-1)} + b$ ή της παραλλαγής του: $P * x^{(k+1)} = (P - R) * x^{(k)} + b$, όπου το μητρώο

$P = D$ και το μητρώο $Q = P - R = L + U$. Επομένως το μητρώο $P = D = \text{diag}(\text{diag}(R)) = \begin{bmatrix} 6 & 0 & 0 & 0 \\ 0 & 6 & 0 & 0 \\ 0 & 0 & 7 & 0 \\ 0 & 0 & 0 & 8 \end{bmatrix}$ και το μητρώο

$Q = P - R = L + U = \begin{bmatrix} 0 & -2 & 1 & -2 \\ -2 & 0 & -2 & -2 \\ -1 & -1 & 0 & -1 \\ -2 & 1 & 0 & 0 \end{bmatrix}$, όπου $L = -\text{tril}(R, -1) = \begin{bmatrix} 0 & 0 & 0 & 0 \\ -2 & 0 & 0 & 0 \\ -1 & -1 & 0 & 0 \\ -2 & -1 & 0 & 0 \end{bmatrix}$ και $U = -\text{triu}(R, 1) = \begin{bmatrix} 0 & -2 & 1 & -2 \\ 0 & 0 & -2 & -2 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$

b) Η ζητούμενη συνάρτηση MATLAB είναι η εξής:

```
function [iconv] = test_conv(R) % το μητρώο R δίνεται κατά την κλίση της function από τη γραμμή εντολών
P = diag(diag(R));
L = -tril(R, -1);
U = -triu(R, 1);
Q = L + U;
T = P \ (L + U); % υπολογισμός του μητρώου επανάληψης T = D-1 * (L + U) = P-1 * (L + U) = P \ (L + U)
x = eig(T); % στο διάνυσμα x επιστρέφονται οι ιδιοτιμές του μητρώου T.
max_eig = max(abs(x)) % η μέγιστη ιδιοτιμή είναι κατ' απόλυτη τιμή = ρ(T) = φασματική ακτίνα του T
if (max_eig < 1), iconv = 1; end; % υπάρχει σύγκλιση της μεθόδου Jacobi και για αυτό επιστρέφεται η τιμή 1
if (max_eig >= 1), iconv = -1; end; % δεν υπάρχει σύγκλιση της μεθόδου και για αυτό επιστρέφεται η τιμή -1
end
```

Σημείωση: η συνάρτηση `eig(A)` επιστρέφει τις ιδιοτιμές του μητρώου $A \in R^{n \times n}$ και το πλήθος αυτών των ιδιοτιμών είναι ίσο με n . Η συνάρτηση `abs(x)` επιστρέφει τις απόλυτες τιμές του διανύσματος x . Η συνάρτηση `max` επιστρέφει το μέγιστο στοιχείο του διανύσματος x . Πάντα σε μια function της MATLAB πρέπει να καταχωρούμε τιμές στις μεταβλητές που επιστρέφει η function. Εδώ, η μεταβλητή αυτή είναι `iconv`. Η συνάρτηση `diag` έχει

σημασία που εφαρμόζεται, δηλαδή αν εφαρμόζεται σε μητρώο ή σε διάνυσμα. Αν εφαρμόζεται σε ένα **μητρώο**, τότε επιστρέφει ως αποτέλεσμα ένα διάνυσμα που περιέχει τα στοιχεία της κύριας διαγωνίου του μητρώου. Αν όμως εφαρμοστεί σε ένα **διάνυσμα**, τότε δημιουργεί ένα διαγώνιο μητρώο, τοποθετώντας το διάνυσμα στο οποίο εφαρμόζεται, στην κύρια διαγώνιο του μητρώου. Ο συνδυασμός δύο εντολών diag, για τη δημιουργία του μητρώου D δίνει: $\text{diag}(\text{diag}(A)) = \text{διαγώνιο μητρώο}$. Η σύνταξη μιας function στη MATLAB είναι της μορφής: **function[αποτ1, αποτ. 2, ..., αποτ. n] = όνομα_function(όρισμα 1, ..., ορισμα n)**

c) Για να ελέγξουμε αν συγκλίνει η μέθοδος Jacobi με βάση τις δοθείσες τιμές των μητρώων, θα πρέπει αρχικά

να υπολογίσουμε το μητρώο επανάληψης $T = D^{-1} * (L + U) = \begin{bmatrix} 0 & -1/3 & 1/6 & -1/3 \\ -1/3 & 0 & -1/3 & -1/3 \\ -1/7 & -1/7 & 0 & -1/7 \\ -1/4 & 1/8 & 0 & 0 \end{bmatrix}$. Παρατηρούμε ότι η

$\|T\|_1 < 1$, επομένως υπάρχει **σύγκλιση** της μεθόδου Jacobi. Έστω και μια νόρμα του μητρώου επανάληψης T να βρεθεί κάτω από «1» σταματάμε. **Εναλλακτικά** μπορούμε να θεωρήσουμε ότι το μητρώο R των συντελεστών έχει **A.Δ.Κ. κατά στήλες**, οπότε η μέθοδος Jacobi συγκλίνει (δεν μας πειράζει που δεν έχει A.Δ.Κ. κατά γραμμές, αρκεί να έχει A.Δ.Κ. είτε κατά γραμμές είτε κατά στήλες).

Διευκρίνηση: Στο μέτρο του δυνατού, προσπαθούμε να αποφύγουμε τον υπολογισμό του μητρώου επανάληψης T, για λόγους εξοικονόμησης πράξεων. Αν όμως **δεν** συγκλίνουν τα τρία πρώτα κριτήρια σύγκλισης ή αν θέλουμε να υπολογίσουμε μια ή περισσότερες λύσεις του γραμμικού συστήματος, π.χ. $x^{(1)}, x^{(2)}$ κ.λ.π. τότε αναγκαστικά πρέπει να καταφύγουμε στον υπολογισμό του μητρώου T, διότι μέσα στα επαναληπτικά σχήματα απαιτείται ο προγενέστερος υπολογισμός του μητρώου αυτού.

2.19 Άσκηση με επαναληπτικές μεθόδους (παλιό θέμα)

Αν χρησιμοποιήσουμε τη μέθοδο **Gauss-Seidel** για την προσέγγιση της λύσης του γραμμικού συστήματος $A * x = b$, είναι γνωστό ότι η προσέγγιση $x^{(k)}$, σε κάθε βήμα μπορεί να υπολογιστεί βάσει του τύπου $P * x^{(k)} = Q * x^{(k-1)} + b$ για κατάλληλα επιλεγμένα P, Q. Έστω ότι $A = \begin{bmatrix} 6 & 2 & -1 & 2 \\ 2 & 4 & 1 & 1 \\ 1 & 1 & 7 & 1 \\ 2 & -1 & 0 & 8 \end{bmatrix}$ και $x^{(0)} = [-1 \ 1 \ 1 \ -1]^T$.

a) Συμπληρώστε τα μητρώα με τις ακριβείς τιμές για τη συγκεκριμένη μέθοδο:

b) Να κυκλώσετε **Ναι** αν μέθοδος συγκλίνει ή **Όχι** αν δεν συγκλίνει στη λύση γραμμικού συστήματος. Σε κάθε περίπτωση να εξηγήσετε γιατί.

c) Να γράψετε συνάρτηση MATLAB **[iconv] = test_conv(A)** που να επιστρέφει iconv=0, αν η παραπάνω επαναληπτική μέθοδος για την επίλυση γενικού συστήματος $A * x = b$ συγκλίνει, διαφορετικά επιστρέφει 1. Ενδείκνυται να χρησιμοποιήσετε τη συνάρτηση eig.

Λύση

$$a) \text{Στην Gauss-Seidel ο διαχωρισμός είναι } P = D - L = \begin{bmatrix} 6 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 \\ 0 & 0 & 7 & 0 \\ 0 & 0 & 0 & 8 \end{bmatrix} - \begin{bmatrix} 0 & 0 & 0 & 0 \\ -2 & 0 & 0 & 0 \\ -1 & -1 & 0 & 0 \\ -2 & 1 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 6 & 0 & 0 & 0 \\ 2 & 4 & 0 & 0 \\ 1 & 1 & 7 & 0 \\ 2 & -1 & 0 & 8 \end{bmatrix} = \text{tril}(A), Q = P$$

$$- A = U = -\text{triu}(A, 1) \text{ οπότε } P = \begin{bmatrix} 6 & 0 & 0 & 0 \\ 2 & 4 & 0 & 0 \\ 1 & 1 & 7 & 0 \\ 2 & -1 & 0 & 8 \end{bmatrix} Q = \begin{bmatrix} 0 & -2 & 1 & -2 \\ 0 & 0 & -1 & -1 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

b) **Το μητρώο A είναι (μη αυστηρά) διαγώνια κυρίαρχο και προφανώς μη αναγωγήσιμο (δεν περιέχει παρά ένα "0", επομένως είναι αδύνατον να υπάρξει μετάθεση P τέτοιο ώστε $P^T A P$ να είναι κατά πλοκάδες άνω τριγωνικό). Είναι όμως αυστηρά Δ.Κ. σε τουλάχιστον 1 γραμμή (π.χ. 1^η), επομένως είναι μη αναγωγήσιμα διαγώνια κυρίαρχο (ΜΑΔΚ) και από τη θεωρία η μέθοδος συγκλίνει.**

$$A = \begin{bmatrix} 6 & 2 & -1 & 2 \\ 2 & 4 & 1 & 1 \\ 1 & 1 & 7 & 1 \\ 2 & -1 & 0 & 8 \end{bmatrix} \rightarrow P * A * P^T = P * \begin{bmatrix} 6 & 2 & -1 & 2 \\ 2 & 4 & 1 & 1 \\ 1 & 1 & 7 & 1 \\ 2 & -1 & 0 & 8 \end{bmatrix} * P^T = \begin{array}{c|cccc} A1 & & & & \\ \hline & -1 & 2 & 6 & 2 \\ & 1 & 4 & 2 & 1 \\ & 7 & 1 & 1 & 1 \\ \hline & 0 & -1 & 2 & 8 \end{array} A2$$

Τα υπομητρώα $A1$ και $A2$ που σχηματίζονται αφού πρώτα έρθει το μητρώο σε άνω τριγωνική μπλοκ μορφή είναι μη – τετραγωνικά. Επιπλέον, υπάρχει Α.Δ.Κ. **για μια τουλάχιστον γραμμή ή στήλη** και πιο συγκεκριμένα εδώ ισχύει για την 4^η γραμμή ή η 4^η στήλη. Επομένως, η μέθοδος Gauss – Seidel συγκλίνει. **Εναλλακτικά**, θα πρέπει να εξεταστούν τα κριτήρια σύγκλισης και χρειάζεται να επισημανθεί η διαφορά μιας ικανής συνθήκης σύγκλισης από μια αναγκαία (πιο δύσκολο να ελεγχθεί, αποτελεί όμως εγγύηση). Για παράδειγμα, κριτήρια νόρμας όπως $\|T\| < 1$ για κάποια γνωστή νόρμα του $T = M^{-1} * N$ είναι ικανά αλλά **όχι αναγκαία**³ για σύγκλιση (μπορεί για παράδειγμα το μητρώο επανάληψης T να έχει κάποια νόρμα μεγαλύτερη του 1, αλλά παρόλα αυτά η μέθοδος να συγκλίνει). Για σύγκλιση όμως είναι **απαραίτητο** να ισχύει ότι $\rho(T) < 1$ δηλαδή η φασματική ακτίνα που είναι η μέγιστη σε απόλυτη – και όχι αλγεβρική – τιμή ιδιοτιμή, να είναι κάτω από 1. Το μητρώο είναι διαστάσεων 4×4 με μηδενικά σε μία με δύο το πολύ θέσεις. **Η σύγκλιση των μεθόδων έπεται αν το μητρώο A είναι αυστηρά διαγώνια κυρίαρχο** (δηλαδή για κάθε γραμμή ή κάθε στήλη το διαγώνιο στοιχείο είναι σε απόλυτη τιμή μεγαλύτερο από το άθροισμα των απολύτων τιμών των υπολοίπων στοιχείων). Στα περισσότερα μητρώα υπάρχει διαγώνια κυριαρχία αλλά **όχι αυστηρή** για όλες τις γραμμές ή στήλες. Επομένως **δεν** μπορούσε να εφαρμοστεί αυτό το κριτήριο. Από την άλλη ίσχυε η αυστηρή δ.κ. για τουλάχιστον μία γραμμή ή στήλη και επιπλέον το μητρώο (λόγω της πυκνότητάς του) ήταν **μη αναγωγήσιμο**. Επομένως το μητρώο είναι **Μη Αναγωγήσιμα Διαγώνιο Κυρίαρχο (ΜΑΔΚ)**, οπότε εξασφαλίζοταν η σύγκλιση της **Jacobi** και **Gauss-Seidel**. Εναλλακτικά, θα μπορούσαμε να εξετάσουμε κατά πόσον κάποια γνωστή νόρμα του μητρώου επανάληψης $T = P^{-1} * Q$ είναι μικρότερη από «1» ή κατά πόσο η φασματική ακτίνα του T είναι μικρότερη του «1». Δυστυχώς, και τα δύο κριτήρια απαιτούν τον υπολογισμό του T . Επίσης, το πρώτο κριτήριο μπορεί να μην ικανοποιείται (και παρόλα αυτά να υπάρχει σύγκλιση), ενώ το δεύτερο είναι αρκετά περίπλοκο. Επίσης, για να

³ Μια εξαίρεση είναι όταν $\rho(T) = \|T\|$ πράγμα που ισχύει όταν το T είναι πραγματικό συμμετρικό και η νόρμα που έχει επιλεγεί είναι η δεύτερη.

εφαρμοστεί το δεύτερο κριτήριο σύγκλισης θα πρέπει να υπάρχει ένα τουλάχιστον «0» στο αρχικό μητρώο, διότι αν δεν υπάρχει κανένα «0» τότε δεν μπορούμε να το φέρουμε -με χρήση μητρώων εναλλαγής P και P^T σε μορφή άνω τριγωνική, μέσω της πράξης P^TAP .

c) function [iconv] = test_conv(A)

```
P = triu(A); Q = -triu(A, 1); //To ; στο τέλος μιας εντολής σημαίνει απόκρυψη εντολής κατά την εκτέλεση της
T = P\Q;
ae = abs(eig(T));
rho = max(ae); % Στη μεταβλητή rho καταχωρείται η φασματική ακτίνα του T
if (rho < 1), iconv = 0, else iconv = 1; end
end
```

2.20 Άσκηση με Jacobi (Θέμα Σεπτεμβρίου 2017)

Δίνεται το μητρώο $A = \text{diag}([5 \ 5 \ 4 \ 4]) - \text{ones}(4)$ και θέλουμε να εφαρμόσουμε τη μέθοδο Jacobi για να λύσουμε το γραμμικό σύστημα με αυτό το μητρώο και δεξιό μέλος το διάνυσμα $b = [4; -4; 3; -3]$. Να υπολογίσετε τις τιμές που αντιστοιχούν στο μητρώο T και το διάνυσμα c ώστε η μέθοδος να μπορεί να γραφτεί ως $x^{(k+1)} = T * x^{(k)} + c$. Να βρεθούν οι τιμές του μητρώου T και του διανύσματος c.

Λύση

$$\text{Το μητρώο } A = \text{diag}([5 \ 5 \ 4 \ 4]) - \text{ones}(4) = \begin{bmatrix} 5 & 0 & 0 & 0 \\ 0 & 5 & 0 & 0 \\ 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 4 \end{bmatrix} - \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 4 & -1 & -1 & -1 \\ -1 & 4 & -1 & -1 \\ -1 & -1 & 3 & -1 \\ -1 & -1 & -1 & 3 \end{bmatrix}.$$

$$\text{Το μητρώο } T = D^{-1} * (L + U) = \begin{bmatrix} 4 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 3 \end{bmatrix}^{-1} * \left(\begin{bmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} \right) = \begin{bmatrix} 1/4 & 0 & 0 & 0 \\ 0 & 1/4 & 0 & 0 \\ 0 & 0 & 1/3 & 0 \\ 0 & 0 & 0 & 1/3 \end{bmatrix} *$$

$$\begin{bmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 1/4 & 1/4 & 1/4 \\ 1/4 & 0 & 1/4 & 1/4 \\ 1/3 & 1/3 & 0 & 1/3 \\ 1/3 & 1/3 & 1/3 & 0 \end{bmatrix} \text{ Το διάνυσμα } c = D^{-1} * b = \begin{bmatrix} 1/4 & 0 & 0 & 0 \\ 0 & 1/4 & 0 & 0 \\ 0 & 0 & 1/3 & 0 \\ 0 & 0 & 0 & 1/3 \end{bmatrix} * \begin{bmatrix} 4 \\ -4 \\ 3 \\ -3 \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \\ 1 \\ -1 \end{bmatrix}.$$

Όσον αφορά τη σύγκλιση, παρατηρούμε ότι το μητρώο των συντελεστών A δεν έχει ΑΔΚ, επομένως **δεν** ικανοποιείται το πρώτο κριτήριο σύγκλισης. Επειδή έχουμε υπολογίσει το μητρώο επανάληψης T, μπορούμε να βρούμε μια νόρμα του. Πιο συγκεκριμένα, αν υπολογίσουμε την πρώτη νόρμα του μητρώου παρατηρούμε ότι αυτή είναι $\|T\|_1 = \frac{11}{12} < 1$, επομένως η μέθοδος Jacobi **συγκλίνει** για οποιοδήποτε αρχικό διάνυσμα $x^{(0)}$.

2.21 Μέθοδος συζυγών κλίσεων (CG) και συζυγών κλίσεων με προρυθμιστή (PCG)

Αποτελεί μια ακόμη επαναληπτική μέθοδο επίλυσης του γραμμικού συστήματος $A * x = b$, όπου υπό την προϋπόθεση ότι το μητρώο A είναι Σ.Θ.Ο. έχουμε ότι: $A * x = b \Leftrightarrow x = \arg \min \frac{1}{2} y^T A y - b^T y$, δηλαδή επιζητείται εκείνο το διάνυσμα x που επιτυγχάνει την **ελαχιστοποίηση** της τετραγωνικής μορφής $\phi(y) = \frac{1}{2}(Ay, y) - (b, y)$. Η $\phi(y)$ είναι κυρτή συνάρτηση και το **τοπικό** ελάχιστο είναι και **ολικό** ελάχιστο. Στη μέθοδο CG και για $k = 1, 2, \dots$ διορθώνουμε την προσέγγιση της λύσης με τον αναδρομικό τύπο: $x^{(k+1)} \leftarrow x^{(k)} + d^{(k)}$ ώστε $\phi(x^{(k+1)}) < \phi(x^{(k)})$. Όταν **δεν** υπάρχει τέτοια διόρθωση έχει υπολογιστεί η λύση. **Εναλλακτικά** όταν $\phi(x^{(k+1)})$ είναι αρκετά **μικρό** τότε τερματίζει. Το $d^{(k)} = x^{(k)} + \alpha_k * p^{(k)}$, όπου το διάνυσμα κατεύθυνσης $p^{(k)}$ και ο συντελεστής α_k επιλέγονται με ειδικό τρόπο. **Η CG είναι η σημαντικότερη μέθοδος επίλυσης μεγάλων γραμμικών συστημάτων με ΣΘΟ μητρώα.** Στη συνέχεια παρουσιάζεται η συνάρτηση **cg** της MATLAB.

```
function [x, error, iter, flag] = cg(A, x, b, M, max_it, tol)
```

```
% [x, error, iter, flag] = cg(A, x, b, M, max_it, tol) Ένας τρόπος κλήσης της συνάρτησης cg από τη γραμμή %εντολών της MATLAB cg.m solves the symmetric positive definite linear system A * x = b %using the Conjugate Gradient method with preconditioning.
```

```
% input
```

```
% A REAL symmetric positive definite matrix
% x REAL initial guess vector, b REAL right handside vector
% M REAL preconditioner matrix
% max_it INTEGER maximum number of iterations (μέγιστος αριθμός επαναλήψεων)
% tol REAL error tolerance. Πρόκειται για την ανοχή του σφάλματος, δηλ. %ένα άνω φράγμα σφάλματος, το οποίο προσδιορίζει την επιτρεπτή τιμή του σφάλματος. error % = norm(x - x_1)/norm(x) <= tol. Επειδή, ενδέχεται να κάνουμε πολλές επαναλήψεις, έχουμε %φροντίσει να θέσουμε εκ' των προτέρων ένα μέγιστο αριθμό επαναλήψεων με την παράμετρο %max_it, έτσι ωστε αν ξεπεράσουμε τον αριθμό των επαναλήψεων που προσδιορίζονται με %την παράμετρο αυτή, να τερματιστεί η επανάληψη, ακόμη και αν δεν έχει προκύψει το %επιθυμητό tolerance
```

```
% output
```

```
% x REAL solution vector
% error REAL error norm (δεύτερη νόρμα σφάλματος)
% iter INTEGER number of iterations performed
% flag INTEGER: 0 = solution found to tolerance (υπάρχει σύγκλιση)
% 1 = no convergence given max_it (δεν υπάρχει σύγκλιση)
% initialization
flag = 0; % προσδιορίζει αν υπάρχει σύγκλιση ή όχι
iter = 0;
bnrm2 = norm(b); % δεύτερη νόρμα του διανύσματος b
if (bnrm2 == 0.0), bnrm2 = 1.0; end
```

```
r = b - A * x; error = norm(r)/bnrm2; %r = residual = υπόλοιπο/κατάλοιπο, είναι η %διαφορά b - A * x. Αν είχαμε βρει ακριβή λύση τότε το r = 0. Όσο η διαφορά αυτή είναι κοντά στο «0», τόσο μικρότερο είναι το r, άρα καλύτερη λύση.
```

```
if (error < tol) return; end % αν το σφάλμα της λύσης < από το tolerance, τέλος
for iter = 1: max_it % Άλλιώς, begin iteration
    z = M\r;
    rho= (r' * z) ;
    if (iter > 1),
        beta = rho/rho_1 ;
        p = z + beta * p ;
    else
        p = z;
```

```

end
q = A * p ;
alpha = rho/(p' * q) ; % η μεταβλητή alpha αντιστοιχεί στην παράμετρο  $\alpha_k$ 
% της εξίσωσης  $d^{(k)} = x^{(k)} + \alpha_k * p^{(k)}$ 
x = x + alpha * p; % update approximation vector, υπολογισμός  $x^{(k)} = x^{(k)} + \alpha_k * p^{(k)}$ 
r = r - alpha * q; % compute residual
error = norm(r)/bnrm2; % check convergence, από τη στιγμή που έχει αλλάξει το r.
if (error <= tol), break; end % αν προκύψει η επιθυμητή ακρίβεια, ολοκληρώνεται.
rho_1 = rho;
end
if (error > tol) flag = 1; end % no convergence
% END cg.m

```

Παραλλαγή της cg → $A * x = b$ ($A = \text{symmetric positive definite}$): Για μεγαλύτερη αποτελεσματικότητα συνδυάζεται με προρρύθμιση (preconditioning), δηλαδή χρησιμοποιείται η **pcg** έκδοσή της, δηλ. εφαρμογή της στο ισοδύναμο σύστημα $PAP^{-T} * (P^{-T} x) = Pb$ για ειδικά επιλεγμένο μητρώο μετάθεσης P ώστε $\kappa(PAP^{-T}) \ll \kappa(A)$. Δηλαδή με τη χρήση του μητρώου **P** πετυχαίνεται η βελτίωση (μείωση) του δείκτη κατάστασης του μητρώου **A**, αφού πρώτα το μετατρέψουμε σε μητρώο PAP^T . Τότε αποκαλείται **preconditioned CG** (**pcg**) και υλοποιείται με τη συνάρτηση **pcg** της MATLAB. Η εύρεση του **P** που πετυχαίνει συνολική μείωση του κόστους είναι ένα σημαντικό ερευνητικό ζήτημα (preconditioning). Στη συνέχεια παρουσιάζεται η συνάρτηση **pcg** της MATLAB.

```
function [x, flag, relres, iter, resvec] = pcg(A, b, tol, MAXIT, M1, M2, x0, varargin)
```

%PCG = Preconditioned Conjugate Gradients Method.

% X = PCG(A, B) attempts to solve system A * X = B. A must be symmetric and positive definite

% X = PCG (AFUN, B) accepts function handle AFUN instead.

% AFUN(X) accepts vector X and returns A * X. In all of the following one can replace A by AFUN

% tol specifies the tolerance of the method. If tol is [] then PCG uses the default 1e - 6 .

% MAXIT specifies the max iterations. If MAXIT is [] then PCG uses the default, min(N, 20) .

% X0 specifies initial guess. If X0 = [] then PCG uses the default, an all zero vector.

% [X, FLAG] = PCG(A, B, . . .) also returns a convergence FLAG:

% 0: PCG converged to the desired tolerance TOL within MAXIT iterations → δηλαδή τότε υπάρχει σύγκλιση της μεθόδου

% 1: PCG iterated MAXIT times **but did not converge** → δηλαδή τότε δεν υπάρχει σύγκλιση της μεθόδου.

% 2: preconditioner M was **ill - conditioned** → Ο πίνακας προρυθμιστής M είναι κακής κατάστασης, δηλαδή έχει μεγάλο $\kappa(M)$.

% 3 PCG stagnated → two consecutive iterations were the same.

% 4 one of the scalar quantities calculated during PCG became too small or too large to continue computing.

Ελεγχος τερματισμού pcg: if (normr <= tolb || stag >= max stag steps || more steps), όπου:

- **normr <= tolb:** αν το σχετικό κατάλοιπο γίνει αρκετά μικρό. Αν το $\kappa(A)$ είναι μεγάλο, δεν εξασφαλίζεται μικρό σχετικό σφάλμα.
- **Moresteps:** αν ξεπεραστεί το μέγιστο επιτρεπτό πλήθος επαναλήψεων.
- **Stagnation:** αν παρατηρηθεί ότι υπάρχει στασιμότητα (stagnation), δηλαδή ότι η (μη ικανοποιητική) προσέγγιση στη λύση παραμένει ίδια για 2 διαδοχικά βήματα.

Παράδειγμα εφαρμογής pcg

Για το μητρώο $A = \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 1 \end{bmatrix}$ και το διάνυσμα $b = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$ η μέθοδος pcg μέσω της κλήσης της από τη γραμμή

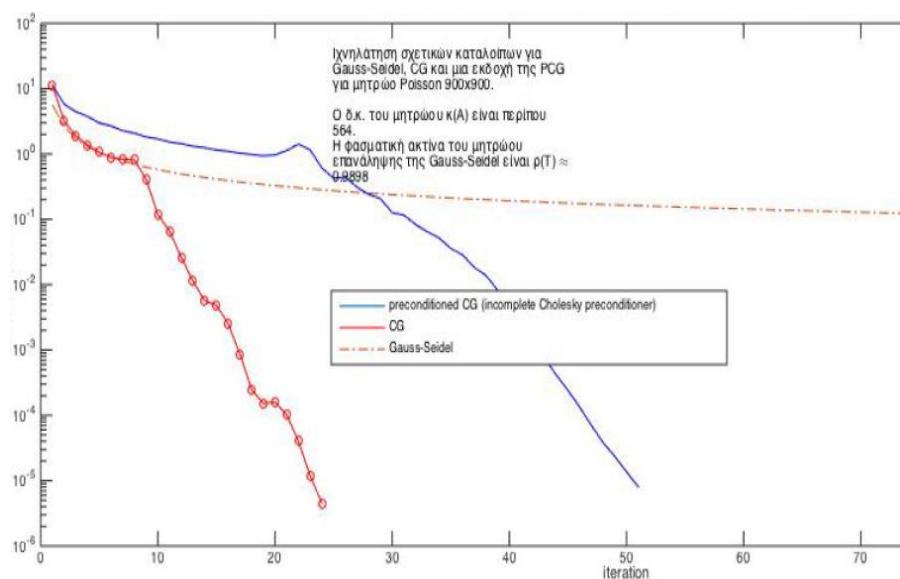
εντολών της MATLAB:

```
>> pcg(A, A * ones(3, 1))

pcg converged at iteration 3 to a solution with relative residual 0 (r = b - A * x)

% επιστρέφεται η ακριβής λύση [1 1 1]
```

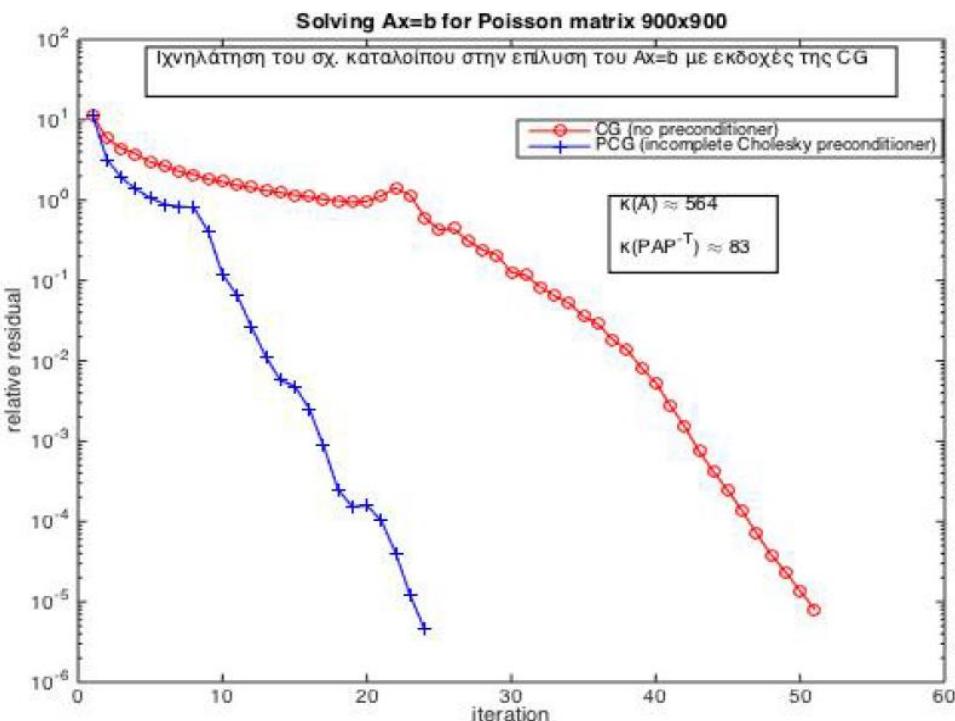
υπολογίζει την ακριβή λύση (**από τη στιγμή που το κατάλοιπο είναι «0»**) σε $n = 3$ βήματα. Αυτό προβλέπεται από τη θεωρία των μεθόδων Krylov. Θεωρητικά, η **ακριβής** λύση του συστήματος $A * x = b$ όπου $A \in \mathbb{R}^{n \times n}$ υπολογίζεται σε $k \leq n$ βήματα. **Στην πράξη, αυτό δεν συμβαίνει λόγω σφαλμάτων στρογγύλευσης.** Στην πράξη δεν ενδιαφέρει, ακόμα και να συνέβαινε, γιατί **η βήματα της cg χρειάζονται περισσότερες πράξεις από την lu.** Στην προκειμένη περίπτωση πράξη και θεωρία συμπίπτουν, γιατί **το μητρώο A που εξετάζεται είναι πολύ μικρό** και τα δεδομένα είναι τέτοια που τα σφάλματα στρογγύλευσης δεν προλαβαίνουν να επηρεάσουν. **Στην Εικόνα 3 που ακολουθεί, φαίνεται μια σύγκριση των αποτελεσμάτων των επαναληπτικών μεθόδων Gauss – Seidel, CG και PCG για την επίλυση του γραμμικού συστήματος $A * x = b$. Είναι εμφανές ότι η μέθοδος Gauss – Seidel συγκλίνει πιο αργά στη λύση, σε σχέση με τις μεθόδους CG και PCG.** Στον κάθετο άξονα του γραφήματος εμφανίζεται το σχετικό κατάλοιπο ή υπόλοιπο (relative residual) και στον οριζόντιο άξονα εμφανίζεται ο αριθμός των επαναλήψεων (iteration).



Εικόνα 3: Σύγκριση των αποτελεσμάτων Gauss – Seidel, CG και PCG για την επίλυση του γραμμικού συστήματος $A * x = b$

Στην Εικόνα 4 που ακολουθεί, φαίνεται μια σύγκριση των αποτελεσμάτων των επαναληπτικών μεθόδων CG και PCG για την επίλυση του γραμμικού συστήματος $A * x = b$. Στον κάθετο άξονα εμφανίζεται το σχετικό κατάλοιπο ή υπόλοιπο (relative residual) και στον οριζόντιο άξονα εμφανίζεται ο αριθμός των επαναλήψεων (iteration).

Είναι εμφανές ότι η μέθοδος PCG συγκλίνει ταχύτερα στη λύση σε σχέση με τη μέθοδο CG (δηλαδή χρησιμοποιεί λιγότερες επαναλήψεις, όπως φαίνεται από το σχετικό γράφημα). Επιπλέον, μειώνεται με ταχύτερο ρυθμό το σχετικό κατάλοιπο της λύσης (relative residual). Η ταχύτερη μείωση του κατάλοιπου αυτού οδηγεί ταχύτερα σε πιο ακριβείς λύσεις. Αυτό συμβαίνει, διότι στη μέθοδο PCG μειώθηκε ο δείκτης κατάστασης του μητρώου (με τη χρήση των μητρώων εναλλαγής) από $\kappa(A) = 564$ σε $\kappa(PAP^{-T}) = 83$. Ο δείκτης κατάστασης ενός μητρώου είναι συνυφασμένος τόσο με την απώλεια των δεκαδικών ψηφίων όσο και με τη ταχύτητα σύγκλισης προς τη τελική λύση του γραμμικού συστήματος $A^*x = b$. Όσο μικρότερος είναι ο $\kappa(A)$ τόσο ταχύτερη σύγκλιση πετυχαίνεται και τόσο λιγότερα ψηφία χάνονται.



Εικόνα 4: Σύγκριση των αποτελεσμάτων CG και PCG για την επίλυση του γραμμικού συστήματος $A^*x = b$

Ερώτηση Σ/Λ: The conjugate gradient method requires to solve a triangular system at each iteration step
Λύση

No. The conjugate gradient method relies on matrix vector multiplications and inner product computations.

2.22 Άσκηση με CG

Μια απλή εκδοχή της μεθόδου CG για την επίλυση του γραμμικού συστήματος $A^*x = b$ δίνεται παρακάτω με μηδενικό αρχικό διάνυσμα. Να δείξτε **μόνο** τα αποτελέσματα που εμφανίζονται στην οθόνη, όταν εκτελέσετε την εντολή: `conjgrad([1 2;-2 1], [1; 1])`; Επίσης να σχολιάσετε την ευστοχία της μεθόδου με βάση τα αποτελέσματα:

```
function [x, r] = conjgrad(A, b)
x=0; r=b; p=r; rsold=r'*r %r ← b = [1; 1] =  $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$  και rsold = r' * r = [1, 1] *  $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$  =
2
for iter =1: length(b) %επιστρέφει το μήκος του b, δηλ. το «2».
```

Computer - Ανάλυση

```

Ap = A * p
%Ap = [1 2] * [1] = [3]
alpha=rsold/(p'*Ap); % alpha = 2/([1 1]*[3]) = 2/2 = 1
iter
x = x + alpha*p
r = r - alpha * Ap;
rsnew = r'*r
if sqrt(rsnew)< 1e-10, break; end
p = r + (rsnew/rsold)*p;
rsold = rsnew;
end

```

Λύση

Τα αποτελέσματα της εκτέλεσης είναι:

>> conjgrad([1 2;-2 1],[1;1])

$$rsold = 2, \text{ διότι } rsold = r' * r = [1 1] * \begin{bmatrix} 1 \\ 1 \end{bmatrix} = 1*1 + 1 * 1 = 2$$

Ap =

3

$$-1, \text{ διότι } Ap = A * p = \begin{bmatrix} 1 & 2 \\ -2 & 1 \end{bmatrix} * \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 3 \\ -1 \end{bmatrix}$$

$$\alpha = rsold/(p' * Ap) = 2/([1 1]*\begin{bmatrix} 3 \\ -1 \end{bmatrix}) = 2/2 = 1 \text{ (δεν εμφανίζεται λόγω του ; στο τέλος της εντολής)}$$

iter = 1 εμφανίζεται στην οθόνη

x =

1

$$1, \text{ διότι } x = x + \alpha * p = 0 + 1 * \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$$r = r - \alpha * Ap = \begin{bmatrix} 1 \\ 1 \end{bmatrix} - 1 * \begin{bmatrix} 3 \\ -1 \end{bmatrix} = \begin{bmatrix} -2 \\ 2 \end{bmatrix} \text{ ως αποτέλεσμα της 1ης εκτέλεσης της επανάληψης}$$

$$rsnew = 8, \text{ διότι } rsnew = r' * r = [-2 2] * \begin{bmatrix} -2 \\ 2 \end{bmatrix} = (-2) * (-2) + 2 * 2 = 8$$

Η συνθήκη **if sqrt(8)< 1e-10** είναι **Ψευδής**, διότι η $\sqrt{8} = 2.82 > 1 * 10^{-10}$, επομένως δεν ολοκληρώνεται το for

$$p = r + (rsnew/rsold) * p = \begin{bmatrix} -2 \\ 2 \end{bmatrix} + \frac{8}{2} * \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} -2 \\ 2 \end{bmatrix} + \begin{bmatrix} 4 \\ 4 \end{bmatrix} = \begin{bmatrix} 2 \\ 6 \end{bmatrix}$$

rsold = rsnew = 8

Ap =

14

$$2 \text{ διότι } Ap = A * p = A * p = \begin{bmatrix} 1 & 2 \\ -2 & 1 \end{bmatrix} * \begin{bmatrix} 2 \\ 6 \end{bmatrix} = \begin{bmatrix} 14 \\ 2 \end{bmatrix}$$

$$\alpha = rsold/(p' * Ap) = 8/([2 6]*\begin{bmatrix} 14 \\ 2 \end{bmatrix}) = 8/40 = 0.2 \text{ (δεν εμφανίζεται λόγω του ; στο τέλος της εντολής).}$$

iter = 2

x =

1.4000

$$2.2000, \text{ διότι } x = x + \alpha * p = \begin{bmatrix} 1 \\ 1 \end{bmatrix} + \frac{8}{40} * \begin{bmatrix} 2 \\ 6 \end{bmatrix} = \begin{bmatrix} 1 + \frac{16}{40} \\ 1 + \frac{48}{40} \end{bmatrix} = \begin{bmatrix} \frac{56}{40} \\ \frac{88}{40} \end{bmatrix} = \begin{bmatrix} 1.4 \\ 2.2 \end{bmatrix}$$

$$r = r - \alpha * Ap = \begin{bmatrix} -2 \\ 2 \end{bmatrix} - \frac{8}{40} * \begin{bmatrix} 14 \\ 2 \end{bmatrix} = \begin{bmatrix} -2 - \frac{112}{40} \\ 2 - \frac{16}{40} \end{bmatrix} = \begin{bmatrix} -\frac{192}{40} \\ \frac{64}{40} \end{bmatrix} = \begin{bmatrix} -4.8 \\ 1.6 \end{bmatrix} \text{ ως αποτέλεσμα της 2ης εκτέλεσης.}$$

$$rsnew = 25.6000, \text{ διότι } rsnew = r' * r = [-4.8, 1.6] * \begin{bmatrix} -4.8 \\ 1.6 \end{bmatrix} = 25.6$$

Η συνθήκη **if sqrt(25.6)< 1e-10** είναι **Ψευδής**, διότι η $\sqrt{25.6} = 5.059 > 1 \cdot 10^{-10}$, επομένως δεν ολοκληρώνεται νωρίτερα το for.

$$p = r + (rsnew/rsold) * p = \begin{bmatrix} -4.8 \\ 1.6 \end{bmatrix} + \frac{25.6}{8} * \begin{bmatrix} 2 \\ 6 \end{bmatrix} = \begin{bmatrix} -4.8 \\ 1.6 \end{bmatrix} + \begin{bmatrix} 6.4 \\ 19.2 \end{bmatrix} = \begin{bmatrix} 1.6 \\ 20.8 \end{bmatrix}$$

$$rsold = rsnew = 25.6$$

$$\begin{aligned} ans = \\ 1.4000 \\ 2.2000 \end{aligned}$$

Εναλλακτικά τα αποτελέσματα είναι τα εξής:

```
>>[x, r]=conjgrad([1 2;-2 1],[1;1])
```

$$\begin{aligned} x = \\ 1.4000 \\ 2.2000 \\ r = \\ -4.8000 \\ 1.6000 \end{aligned}$$

Συμπεραίνουμε ότι όσο μεγαλώνει ο αριθμός των επαναλήψεων τόσο μεγαλώνει και η τιμή της παραμέτρου r που επιστρέφεται, αρχικά είναι το διάνυσμα $r = [-2, 2]^T$ και στη δεύτερη επανάληψη επιστρέφεται το διάνυσμα $r = [-4.8, 1.6]^T$. Παρατηρούμε ότι στη δεύτερη επανάληψη το residual (κατάλοιπο ή υπόλοιπο) μειώνεται, επομένως η μέθοδος συγκλίνει.

2.23 Άσκηση MATLAB

Δίνεται ο παρακάτω κώδικας MATLAB. Να εξηγηθεί η λειτουργία του.

```

1 function [x,res,in,zer] = lu_solve(A,b)
2 in=true;
3 [m,n]=size(A); zer=false;
4
5 if (m ~= n )
6     error('The input matrix is not square');
7 end
8 if (m~=size(b,1))
9     error('Righth side has not proper size ')
10 end
11 [L,U,P]=lu(A); %alternatively you can use gep.m function for LU
12
13
14 mx=max(max(abs(U)));
15 p=length(find(abs(diag(U))<eps(mx)));
16 %p=length(find(diag(U)==0));
17
18 if (p~=0)
19     in=false;
20 end
21 if (P==eye(n))
22     zer=true;
23 end
24 y=L\ P*b; x=U\ y; res=norm(A*x-b);

```

Λύση

Στη γραμμή 11 καλείται η ενδογενής συνάρτηση `Iu()` της MATLAB που υπολογίζει τους παράγοντες L, U, P. Εφόσον έχουμε στη διάθεσή μας τους παράγοντες L, U, P είναι εύκολο να διαπιστωθεί αν το μητρώο A είναι αντιστρέψιμο, ελέγχοντας τους οδηγούς, δηλ. τα στοιχεία στην κύρια διαγώνιο του μητρώου U. Είναι καλύτερο να αποφεύγουμε τη σύγκριση με το «0» και να επιλέγουμε τη σύγκριση με έναν αριθμό κοντά στο «0». Οι εντολές της γραμμής «14» εκτελούν αυτό ακριβώς. Η γραμμή «24» του κώδικα υπολογίζει τη λύση του γραμμικού συστήματος και το κατάλοιπο `res` της λύσης.

2.24 Συμπεράσματα σχετικά με τις επαναληπτικές μεθόδους επίλυσης

- α) Ενδιαφερόμαστε μόνο για μεθόδους που συγκλίνουν.
- β) Μας ενδιαφέρει ο ρυθμός σύγκλισης που εξαρτάται από το είδος του μητρώου.
- γ) Συχνά το κόστος είναι σταθερό σε κάθε βήμα, οπότε το συνολικό κόστος προσεγγίζεται ως το γινόμενο κόστος/βήμα * πλήθος βημάτων.

Κεφάλαιο 3 – Μέθοδος δυνάμεων και κύκλοι Gershgorin

3.1 Ιδιοτιμές – Μέθοδος Δυνάμεων

Αν $C \in R^{n \times n}$ και έχει ως **ιδιοζεύγη** (ή ιδιοποσά) τα μεγέθη λ (ιδιοτιμή) και x (ιδιοδιάνυσμα), τότε ανάμεσά τους ισχύει η σχέση: $A * x = \lambda * x$. Όταν θέλουμε να υπολογίσουμε τις ιδιοτιμές ενός μητρώου και αυτό είναι **μικρής** τάξης, τότε χρησιμοποιείται το χαρακτηριστικό πολυώνυμο, δηλ. το $\det(\lambda * I - A) = 0$. Η χρήση του χαρακτηριστικού πολυωνύμου για την εύρεση ιδιοτιμών ενός μητρώου θεωρείται άμεση μέθοδος. Αν όμως μεγαλώσει πολύ το μέγεθος του μητρώου, τότε θα χρησιμοποιείται η **μέθοδος των δυνάμεων, που θεωρείται επαναληπτική μέθοδος**. Η διαφορά ανάμεσα στη μέθοδο του χαρακτηριστικού πολυωνύμου (άμεση μέθοδος) και τη **μέθοδο των δυνάμεων** (επαναληπτική μέθοδος) είναι ότι η πρώτη υπολογίζει **όλες** τις ιδιοτιμές του μητρώου, ενώ η δεύτερη **μόνο** την κυρίαρχη ιδιοτιμή (ή αλλιώς φασματική ακτίνα) εφόσον υπάρχει. Κάθε μητρώο $A \in R^{n \times n}$ έχει ακριβώς η ιδιοτιμές $\lambda_1, \lambda_2, \dots, \lambda_n$ που υπολογίζονται από το χαρακτηριστικό πολυώνυμο του και το σύνολο των ιδιοτιμών του ονομάζεται **φάσμα μητρώου** και συμβολίζεται με $\Lambda(A) = \{\lambda_1, \lambda_2, \dots, \lambda_n\}$, το οποίο μπορεί να περιέχει πολλαπλές ιδιοτιμές (δηλ. μια ιδιοτιμή να επαναλαμβάνεται περισσότερες από μια φορές), δηλαδή να έχει **αλγεβρική πολλαπλότητα** $AM > 1$. Τις ιδιοτιμές μπορούμε να τις κατατάξουμε σε φθίνουσα σειρά δηλαδή $|\lambda_1| \geq |\lambda_2| \geq \dots \geq |\lambda_n|$. Τότε το $|\lambda_1|$ ονομάζεται **φασματική ακτίνα ή κυρίαρχη ιδιοτιμή**.

Η **πολλαπλότητα** μιας ιδιοτιμής ως ρίζα του χαρακτηριστικού πολυωνύμου ονομάζεται **αλγεβρική πολλαπλότητα**, ενώ η πολλαπλότητα ενός ιδιοδιανύσματος που υπολογίζεται από τον τύπο $null(A - \lambda I)$ ονομάζεται **γεωμετρική πολλαπλότητα**. Αν για κάποια ιδιοτιμή ισχύει ότι η γεωμετρική πολλαπλότητά της είναι **μικρότερη** από την αλγεβρική, ονομάζεται **ελλειμματική** και ένα μητρώο είναι διαγνωνοποιήσιμο αν και μόνο αν δεν έχει ελλειμματικές ιδιοτιμές. Αν ένα μητρώο είναι **διαγωνοποιήσιμο**, γράφεται σε μορφή: $\lambda_1 u_1 v_1^T + \lambda_2 u_2 v_2^T + \dots + \lambda_n u_n v_n^T$, δηλαδή ως άθροισμα εξωτερικών γινομένων.

Η θεωρία της **μεθόδου των δυνάμεων** συνοψίζεται στον επόμενο αλγόριθμο. Το μητρώο A που εμφανίζεται, είναι αυτό για το οποίο αναζητώ την κυρίαρχη ιδιοτιμή (δηλαδή τη μοναδική φασματική ακτίνα εφόσον υπάρχει).

$$\begin{aligned} \text{for } k = 1, 2, \dots \\ \mathbf{x}^{(k)} = A\mathbf{y}^{(k-1)}, \quad \mathbf{y}^{(k)} = \frac{\mathbf{x}^{(k)}}{\|\mathbf{x}^{(k)}\|}, \quad \lambda^{(k)} = (\mathbf{y}^{(k)})^H A \mathbf{y}^{(k)} \end{aligned}$$

$$\text{Κλάσμα ή πηλίκο Rayleigh} = \frac{z^T A z}{z^T z}$$

Περιγραφή αλγόριθμου: Στον αλγόριθμο αυτό το $y^{(k)}$ είναι το **κανονικοποιημένο ιδιοδιάνυσμα** (το διαιρούμε με το μήκος του που είναι η δεύτερη νόρμα, δηλαδή το μέγεθος $\|x^k\|$) και το **$\lambda^{(k)}$ είναι η κυρίαρχη ιδιοτιμή που υπολογίζεται ως έξοδος του αλγόριθμου**. Ο εκθέτης k συμβολίζει το βήμα της επανάληψης και μά-

λιστα ισχύει ότι όσες περισσότερες επαναλήψεις εκτελούμε, τόσο πιο ακριβής είναι η ιδιοτιμή που υπολογίζουμε με τη μέθοδο των δυνάμεων, η οποία είναι μια προσεγγιστική μέθοδος υπολογισμού της κυρίαρχης ιδιοτιμής ενός μητρώου. Τέλος για να χρησιμοποιηθεί η μέθοδος των δυνάμεων (προϋπόθεση), θα πρέπει το μητρώο να περιέχει μόνο μια κυρίαρχη ιδιοτιμή. Αυτό σημαίνει ότι δεν θα πρέπει να περιέχει περισσότερες από μια φορές την κυρίαρχη ιδιοτιμή του ή με άλλα λόγια η κυρίαρχη ιδιοτιμή του μητρώου πρέπει να έχει αλγεβρική πολλαπλότητα ίση με «1», δηλ. να μην επαναλαμβάνεται περισσότερες από μια φορές. Επίσης, θα πρέπει να δίνεται και το αρχικό ιδιοδιάνυσμα, στο βήμα $k = 0$.

3.1.1 Πρώτο παράδειγμα με την μέθοδο δυνάμεων

Με δεδομένο το μητρώο $A = \begin{bmatrix} 1 & 2 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$ και το αρχικό ιδιοδιάνυσμα $x^{(0)} = [1 \ 2 \ 3]^T$, να βρεθεί με τη μέθοδο των δυνάμεων η κυρίαρχη ιδιοτιμή, μετά από ένα βήμα της μεθόδου, δηλ. $\lambda^{(1)} =$;

Λύση

Καταρχήν πρέπει να βεβαιωθούμε ότι υπάρχει **μοναδική κυρίαρχη ιδιοτιμή** στο μητρώο A , για να μπορέσουμε να εφαρμόσουμε τη μέθοδο των δυνάμεων. Αν για παράδειγμα υπάρχουν **δύο ίδιες ιδιοτιμές** –έστω και κατ' απόλυτη τιμή ίσες μεταξύ τους, π.χ. το -2 και το 2 δεν μπορεί να εφαρμοστεί η μέθοδος των δυνάμεων.

Σημείωση: Αν το μητρώο A είχε **όλα τα στοιχεία του θετικά**, τότε σύμφωνα με το θεώρημα **Perron – Frobenius** –που αναλύεται στη συνέχεια– θα περιέχει **κυρίαρχη ιδιοτιμή**, επομένως μπορεί να εφαρμοστεί η μέθοδος των δυνάμεων. Εναλλακτικά –αν και δεν συνίσταται– μπορεί να χρησιμοποιηθεί το **χαρακτηριστικό πολυώνυμο** για να –υπολογιστεί αν υπάρχει– η κυρίαρχη ιδιοτιμή του μητρώου A . Πιο συγκεκριμένα, έχουμε ότι: $\det(A - \lambda * I) = 0 \Rightarrow \det\begin{pmatrix} 1 - \lambda & 2 & 0 \\ 1 & -\lambda & 0 \\ 0 & -1 & -\lambda \end{pmatrix} = 0 \Rightarrow (-1)^{3+3} * (-\lambda)^* \begin{vmatrix} 1 - \lambda & 2 \\ 1 & -\lambda \end{vmatrix} = 0 \Rightarrow -\lambda * (-\lambda^* (1-\lambda) - 2) = 0 \Rightarrow \lambda_1 = 0 \text{ και } \lambda_2 = -2 \text{ και } \lambda_3 = 1.$

Άρα υπάρχει κυρίαρχη ιδιοτιμή και είναι το $\lambda_2 = 2$. Επομένως, μπορεί να εφαρμοστεί η μέθοδος των δυνάμεων:

$$y^{(0)} = \frac{x^{(0)}}{\|x^{(0)}\|} = \frac{\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}}{\sqrt{1^2 + 2^2 + 3^2}} = \frac{1}{\sqrt{14}} * \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \text{ και το ιδιοδιάνυσμα } x^{(1)} = A * y^{(0)} = \begin{bmatrix} 1 & 2 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} * \frac{1}{\sqrt{14}} * \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = \frac{1}{\sqrt{14}} * \begin{bmatrix} 5 \\ 1 \\ 2 \end{bmatrix}. \text{ Το } y^{(1)} =$$

$$\frac{x^{(1)}}{\|x^{(1)}\|} = \frac{\begin{bmatrix} 5 \\ 1 \\ 2 \end{bmatrix}}{\sqrt{\left(\frac{5}{\sqrt{14}}\right)^2 + \left(\frac{1}{\sqrt{14}}\right)^2 + \left(\frac{2}{\sqrt{14}}\right)^2}} = \frac{\begin{bmatrix} 5 \\ 1 \\ 2 \end{bmatrix}}{\sqrt{\frac{30}{14}}} = \frac{\begin{bmatrix} 5 & 1 & 2 \end{bmatrix}^T}{\sqrt{14} * \frac{\sqrt{30}}{\sqrt{14}}} = \frac{\begin{bmatrix} 5 & 1 & 2 \end{bmatrix}^T}{\sqrt{30}}. \text{ Στη συνέχεια υπολογίζουμε την ιδιοτιμή μετά από μια επανάληψη (ένα βήμα) που είναι: } \lambda^{(1)} = (y^{(1)})^T * A * y^{(1)} = \frac{1}{\sqrt{30}} [5 \ 1 \ 2] * \begin{bmatrix} 1 & 2 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} * \frac{1}{\sqrt{30}} \begin{bmatrix} 5 \\ 1 \\ 2 \end{bmatrix} = \frac{1}{30} [6 \ 12 \ 0] \begin{bmatrix} 5 \\ 1 \\ 2 \end{bmatrix} = \frac{1}{30} *$$

$42 = \frac{42}{30} = 1.4$. Στη συνέχεια πρέπει να επαναλάβουμε την ίδια διαδικασία και άλλες φορές για να βρούμε μια καλύτερη προσέγγιση της κυρίαρχης ιδιοτιμής. Γενικά, όσο περισσότερες επαναλήψεις της μεθόδου των δυνάμεων χρησιμοποιούμε, τόσο πιο ακριβές θα είναι το αποτέλεσμα.

3.1.2 Δεύτερο παράδειγμα με τη μέθοδο δυνάμεων (Παλιό Θέμα)

Για το μητρώο $H = \begin{bmatrix} 8 & 1 & 6 \\ 3 & 5 & 7 \\ 4 & 9 & 2 \end{bmatrix}$ να χρησιμοποιήσετε δύο επαναλήψεις της Μεθόδου Δύναμης για να προσεγ-

γίσετε τη μέγιστη ιδιοτιμή του μητρώου, εκκινώντας τη διαδικασία με αρχικό διάνυσμα $x^{(0)} = [1 \ 1 \ 1]^T$.

Λύση

Πρώτα πρέπει να ελέγξουμε ότι το μητρώο H περιέχει κυρίαρχη ιδιοτιμή, για να μπορεί να εφαρμοστεί η μέθοδος των δυνάμεων. **Αυτό ισχύει λόγω του θεωρήματος Perron – Frobenius.**

$$y^{(0)} = \frac{x^{(0)}}{\|x^{(0)}\|} = \frac{\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}}{\sqrt{1^2+1^2+1^2}} = \frac{1}{\sqrt{3}} * \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \text{ και το ιδιοδιάνυσμα } x^{(1)} = H * y^{(0)} = \begin{bmatrix} 8 & 1 & 6 \\ 3 & 5 & 7 \\ 4 & 9 & 2 \end{bmatrix} * \frac{1}{\sqrt{3}} * \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \frac{1}{\sqrt{3}} * \begin{bmatrix} 15 \\ 15 \\ 15 \end{bmatrix}. \text{ Στη συνέχεια}$$

κανονικοποιούμε το διάνυσμα $y^{(1)}$ διαιρώντας κάθε στοιχείο με το μήκος του. Το $y^{(1)} = \frac{x^{(1)}}{\|x^{(1)}\|} =$

$$\frac{\frac{1}{\sqrt{3}} * \begin{bmatrix} 15 \\ 15 \\ 15 \end{bmatrix}}{\sqrt{(\frac{15}{\sqrt{3}})^2 + (\frac{15}{\sqrt{3}})^2 + (\frac{15}{\sqrt{3}})^2}} = \frac{\frac{1}{\sqrt{3}} * \begin{bmatrix} 15 \\ 15 \\ 15 \end{bmatrix}}{15} = \frac{1}{\sqrt{3}} * \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}. \text{ Στη συνέχεια υπολογίζουμε τη μέγιστη (κυρίαρχη) ιδιοτιμή μετά από μια}$$

επανάληψη (ένα βήμα) που είναι: $\lambda^{(1)} = (y^{(1)})^T * H * y^{(1)} = \frac{1}{\sqrt{3}} [1 \ 1 \ 1] * \begin{bmatrix} 8 & 1 & 6 \\ 3 & 5 & 7 \\ 4 & 9 & 2 \end{bmatrix} * \frac{1}{\sqrt{3}} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = 15$. Κατόπιν υπολογί-

ζουμε το ιδιοδιάνυσμα $x^{(2)} = H * y^{(1)} = \begin{bmatrix} 8 & 1 & 6 \\ 3 & 5 & 7 \\ 4 & 9 & 2 \end{bmatrix} * \frac{1}{\sqrt{3}} * \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \frac{1}{\sqrt{3}} * \begin{bmatrix} 15 \\ 15 \\ 15 \end{bmatrix}$ και το κανονικοποιούμε, υπολογίζοντας το

$$y^{(2)} = \frac{x^{(2)}}{\|x^{(2)}\|} = \frac{\frac{1}{\sqrt{3}} * \begin{bmatrix} 15 \\ 15 \\ 15 \end{bmatrix}}{\sqrt{(\frac{15}{\sqrt{3}})^2 + (\frac{15}{\sqrt{3}})^2 + (\frac{15}{\sqrt{3}})^2}} = \frac{\frac{1}{\sqrt{3}} * \begin{bmatrix} 15 \\ 15 \\ 15 \end{bmatrix}}{15} = \frac{1}{\sqrt{3}} * \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}. \text{ Στη συνέχεια υπολογίζουμε τη μέγιστη (κυρίαρχη) ιδιοτιμή}$$

μετά από μια ακόμη επανάληψη της μεθόδου των δυνάμεων που είναι: $\lambda^{(2)} = (y^{(2)})^T * H * y^{(2)} = \frac{1}{\sqrt{3}} [1 \ 1 \ 1] *$

$$\begin{bmatrix} 8 & 1 & 6 \\ 3 & 5 & 7 \\ 4 & 9 & 2 \end{bmatrix} * \frac{1}{\sqrt{3}} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = 15. \text{ Παρατηρούμε ότι καταλήξαμε στο ίδιο αποτέλεσμα με αυτό της πρώτης επανάληψης.}$$

Αυτό σημαίνει ότι ήδη από το πρώτο βήμα έχουμε υπολογίσει την κυρίαρχη ιδιοτιμή.

3.1.3 Τρίτο παράδειγμα με την μέθοδο δυνάμεων

Να επαληθεύσετε ότι **δεν** μπορεί να εφαρμοστεί η μέθοδος των δυνάμεων στο μητρώο A .

$$A = \left[\begin{array}{ccc|cc} 1/3 & 2/3 & & 2 & 3 \\ 1 & 0 & & -1 & 2 \\ 0 & 0 & & -5/3 & -2/3 \\ 0 & 0 & & 1 & 0 \end{array} \right]$$

Λύση

Σε ένα μητρώο **τριγωνικό/διαγώνιο**, οι ιδιοτιμές του βρίσκονται στην κύρια διαγώνιο. Αν είναι σε μπλοκ μορφή, οι ιδιοτιμές του είναι οι ιδιοτιμές των υπομητρώων της κύριας διαγωνίου. Από μια πρώτη ματιά διαπιστώνουμε ότι **δεν** ισχύει το θεώρημα Perron – Frobenius, διότι **δεν** είναι θετικά όλα τα στοιχεία του. Ως

επαλήθευση, αν χωρίσουμε το μητρώο A σε **υπομητρώα (πλοκάδες)**, τότε αυτό δημιουργείται σε άνω τριγωνική μορφή, δηλαδή γίνεται άνω τριγωνικό μπλοκ μητρώο και τότε οι ιδιοτιμές του είναι οι ιδιοτιμές των υπομητρώων A_1 και A_4 που βρίσκονται στην κύρια διαγώνιο.

$$A = \left[\begin{array}{cc|cc} \frac{1}{3} & \frac{2}{3} & 2 & 3 \\ 1 & 0 & -1 & 2 \\ \hline 0 & 0 & -5/3 & -2/3 \\ 0 & 0 & 1 & 0 \end{array} \right] = \begin{bmatrix} A_1 & A_2 \\ 0 & A_4 \end{bmatrix}. \text{ Το } \lambda(A) = \lambda\{A_1\} \cup \lambda\{A_4\} = (\det(\lambda * I - A_1) = 0) \cup (\det(\lambda * I - A_4) = 0) = \{1, -1, -\frac{2}{3}, -\frac{2}{3}\}.$$

Με βάση αυτές τις ιδιοτιμές, διαπιστώνεται ότι **δεν** υπάρχει κυρίαρχη ιδιοτιμή, αφού αν $\lambda_1 = 1, \lambda_2 = -1$ και ισχύει ότι $|\lambda_1| = |\lambda_2|$. Άρα **δεν** μπορεί να εφαρμοστεί η μέθοδος των δυνάμεων, γιατί **έχουμε επανάληψη της κυρίαρχης ιδιοτιμής**.

3.1.4 Τέταρτο παράδειγμα με την μέθοδο δυνάμεων (παλιό θέμα)

Υπολογίστε με τη μέθοδο της δύναμης τη μεγαλύτερη κατ' απόλυτη τιμή ιδιοτιμή (δηλ. την κυρίαρχη ιδιοτιμή) και το **αντίστοιχο ιδιοδιάνυσμα** του πίνακα $A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 3 & 3 \\ 1 & 3 & 6 \end{bmatrix}$ κάνοντας 2 επαναλήψεις. Το $x^{(0)} = [1 \ 1 \ 1]^T$. Συγκρίνετε τα αποτελέσματά σας με την πραγματική τιμή της απόλυτα μεγαλύτερης ιδιοτιμής.

Λύση

Παρατηρούμε ότι όλα τα στοιχεία του μητρώου είναι θετικά, επομένως σύμφωνα με το θεώρημα Perron- Frobenius υπάρχει κυρίαρχη ιδιοτιμή, επομένως μπορεί να εφαρμοστεί η μέθοδος των δυνάμεων για την εύρεσή

$$\text{της. Το διάνυσμα } y^{(0)} = \frac{x^{(0)}}{\|x^{(0)}\|} = \frac{\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}}{\sqrt{1^2+1^2+1^2}} = \frac{1}{\sqrt{3}} * \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \text{ και το ιδιοδιάνυσμα } x^{(1)} = A * y^{(0)} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 3 & 3 \\ 1 & 3 & 6 \end{bmatrix} * \frac{1}{\sqrt{3}} * \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

$$\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \frac{1}{\sqrt{3}} * \begin{bmatrix} 3 \\ 7 \\ 10 \end{bmatrix}. \text{ Στη συνέχεια κανονικοποιούμε το διάνυσμα } x^{(1)} \text{ διαιρώντας κάθε στοιχείο με το μήκος του. Το}$$

$$y^{(1)} = \frac{x^{(1)}}{\|x^{(1)}\|} = \frac{\frac{1}{\sqrt{3}} * \begin{bmatrix} 3 \\ 7 \\ 10 \end{bmatrix}}{\sqrt{(\frac{3}{\sqrt{3}})^2 + (\frac{7}{\sqrt{3}})^2 + (\frac{10}{\sqrt{3}})^2}} = \frac{\frac{1}{\sqrt{3}} * \begin{bmatrix} 3 \\ 7 \\ 10 \end{bmatrix}}{\sqrt{158/3}} = \frac{1}{\sqrt{158}} * \begin{bmatrix} 3 \\ 7 \\ 10 \end{bmatrix}. \text{ Στη συνέχεια υπολογίζουμε τη μέγιστη (κυρίαρχη) ιδιοτιμή μετά από μια επανάληψη (ένα βήμα) που είναι: } \lambda^{(1)} = (y^{(1)})^T * A * y^{(1)} = \frac{1}{\sqrt{158}} [3 \ 7 \ 10] * \begin{bmatrix} 1 & 1 & 1 \\ 1 & 3 & 3 \\ 1 & 3 & 6 \end{bmatrix} *$$

$$\frac{1}{\sqrt{158}} \begin{bmatrix} 3 \\ 7 \\ 10 \end{bmatrix} = \frac{1278}{158} = 8.08. \text{ Κατόπιν υπολογίζουμε το ιδιοδιάνυσμα } x^{(2)} = A * y^{(1)} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 3 & 3 \\ 1 & 3 & 6 \end{bmatrix} * \frac{1}{\sqrt{158}} * \begin{bmatrix} 3 \\ 7 \\ 10 \end{bmatrix} = \frac{1}{\sqrt{158}} * \begin{bmatrix} 3 \\ 7 \\ 10 \end{bmatrix}$$

$$\begin{bmatrix} 20 \\ 54 \\ 84 \end{bmatrix} \text{ και το κανονικοποιούμε, υπολογίζοντας το } y^{(2)} = \frac{x^{(2)}}{\|x^{(2)}\|} = \frac{\frac{1}{\sqrt{158}} * \begin{bmatrix} 20 \\ 54 \\ 84 \end{bmatrix}}{\sqrt{(\frac{20}{\sqrt{158}})^2 + (\frac{54}{\sqrt{158}})^2 + (\frac{84}{\sqrt{158}})^2}} = \frac{\frac{1}{\sqrt{158}} * \begin{bmatrix} 20 \\ 54 \\ 84 \end{bmatrix}}{\sqrt{10372/158}} = \frac{1}{\sqrt{10372}} * \begin{bmatrix} 20 \\ 54 \\ 84 \end{bmatrix}.$$

Στη συνέχεια υπολογίζουμε τη μέγιστη (κυρίαρχη) ιδιοτιμή μετά από μια ακόμη επανάληψη της μεθόδου των

$$\text{δυνάμεων που είναι: } \lambda^{(2)} = (y^{(2)})^T * H * y^{(2)} = \frac{1}{\sqrt{10372}} [20 \ 54 \ 84] * \begin{bmatrix} 1 & 1 & 1 \\ 1 & 3 & 3 \\ 1 & 3 & 6 \end{bmatrix} * \frac{1}{\sqrt{10372}} \begin{bmatrix} 20 \\ 54 \\ 84 \end{bmatrix} = 8.12. \text{ Παρατηρούμε ότι}$$

καταλήξαμε σε παρόμοιο αποτέλεσμα με αυτό της πρώτης επανάληψης. Για να συγκρίνουμε τα αποτελέσματα με την **ακριβή τιμή της απόλυτα μεγαλύτερης ιδιοτιμής**, θα πρέπει στη συνέχεια να υπολογίσουμε τις ιδιοτιμές του μητρώου A από το χαρακτηριστικό πολυώνυμο (ακριβής υπολογισμός), δηλαδή να υπολογίσουμε την

$$\text{ορίζουσα } \det(A - \lambda * I) = 0 \Rightarrow \det \begin{bmatrix} 1 - \lambda & 1 & 1 \\ 1 & 3 - \lambda & 3 \\ 1 & 3 & 6 - \lambda \end{bmatrix} = 0 \Rightarrow (-1)^{1+1} * (1 - \lambda) * \begin{vmatrix} 3 - \lambda & 3 \\ 3 & 6 - \lambda \end{vmatrix} + (-1)^{1+2} * 1 * \begin{vmatrix} 1 & 3 \\ 1 & 6 - \lambda \end{vmatrix} + (-1)^{1+3} * 1 * \begin{vmatrix} 1 & 3 - \lambda \\ 1 & 3 \end{vmatrix} = 0$$

$\Rightarrow [(1 - \lambda) * (3 - \lambda) * (6 - \lambda) - 9] - [(6 - \lambda) * 3] + [3 * (3 - \lambda)] = -\lambda^3 + 10 * \lambda^2 - 16 * \lambda + 6 = 0 \Rightarrow \lambda_1 = 0.56, \lambda_2 = 1.31, \lambda_3 = 8.12$. Παρατηρούμε ότι μετά από δύο βήματα της μεθόδου δυνάμεων υπολογίσαμε την κυρίαρχη ιδιοτιμή του μητρώου A , αφού στο δεύτερο βήμα της μεθόδου βρέθηκε το αποτέλεσμα της πρώτης. Επειδή το μητρώο A που δίνεται στην εκφώνηση έχει όλα τα στοιχεία του θετικά, μπορεί -όπως αναφέρθηκε και προηγουμένως- να εφαρμοστεί η μέθοδος των δυνάμεων, επειδή υπάρχει κυρίαρχη ιδιοτιμή σύμφωνα με το θεώρημα Perron – Frobenius.

3.1.5 Ταχύτητα σύγκλισης μεθόδου δυνάμεων

Θέτοντας την ανοχή ίση με $\epsilon = 10^{-10}$ να χρησιμοποιηθεί η μέθοδος των δυνάμεων για να προσεγγιστεί η μέγιστη ιδιοτιμή κατά μέτρο των μητρώων A_1, A_2, A_3 ξεκινώντας από το αρχικό διάνυσμα $x^{(0)} = [1 \ 2 \ 3]^T$, όπου

$$A_1 = \begin{bmatrix} 1 & 2 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}, A_2 = \begin{bmatrix} 0.1 & 3.8 & 0 \\ 1 & 0 & 0 \\ 1 & 1 & 0 \end{bmatrix}, A_3 = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}.$$

Λύση

Υπολογίζουμε **καταρχήν** τις ιδιοτιμές για το μητρώο A_1 χρησιμοποιώντας το χαρακτηριστικό πολυώνυμο του

$$\text{που είναι το εξής: } \det(A_1 - \lambda * I) = \begin{vmatrix} 1 - \lambda & 2 & 0 \\ 1 & -\lambda & 0 \\ 0 & 1 & -\lambda \end{vmatrix} = 0 \Rightarrow -\lambda^3 + \lambda^2 + 2 * \lambda = 0 \Rightarrow -\lambda * (\lambda + 1) * (\lambda - 2) = 0 \Rightarrow \lambda_3 = 0, \lambda_2 = -1, \lambda_1 = 2.$$

Στη συνέχεια υπολογίζουμε **το πηλίκο των δύο μεγαλύτερων κατ' απόλυτη τιμή ιδιοτιμών** του μητρώου, δηλ. το πηλίκο $\left| \frac{\lambda_2}{\lambda_1} \right| = \frac{1}{2}$. Επειδή θέτουμε στον παρονομαστή του κλάσματος τη μεγαλύτερη κατ' απόλυτη τιμή ιδιοτιμή, το κλάσμα αυτό είναι πάντα μικρότερο της μονάδας και **μάλιστα όσο πιο κοντά στο 1 τείνει το κλάσμα αυτό, τόσο πιο αργή σύγκλιση της μεθόδου των δυνάμεων επιτυγχάνεται**.

Αν εκτελέσουμε τον κώδικα της μεθόδου σε Η/Υ, τότε αυτή θα συγκλίνει μετά από **34** επαναλήψεις στην τιμή $2,00000000004982$. Η αντίστοιχη σύγκλιση για το μητρώο A_2 που έχει κυρίαρχη ιδιοτιμή το 2 (μεταξύ των ιδιοτιμών 0, -1.9, 2) πετυχαίνεται έπειτα από **457** επαναλήψεις. Αυτό συμβαίνει διότι στην περίπτωση αυτή το πηλίκο των δύο μεγαλύτερων κατ' απόλυτη τιμή ιδιοτιμών του μητρώου είναι πιο κοντά στο 1, αφού: $\left| \frac{\lambda_2}{\lambda_1} \right| = \left| \frac{-1.9}{2} \right| = 0.95 \approx 1$. Τέλος, αναφορικά με το μητρώο A_3 , οι ιδιοτιμές του είναι **μιγαδικές** και συγκεκριμένα είναι το i και το $-i$. Επειδή αυτές κατά μέτρο είναι **ίσες** μεταξύ τους, δεν υπάρχει στην περίπτωση αυτή σύγκλιση της μεθόδου των δυνάμεων. Επίσης αναμέναμε να βρεθεί και από το θεώρημα Perron-Frobenius, διότι το μητρώο A_3 περιέχει αρνητικές τιμές.

3.1.6 Σύγκλιση και κώδικας Matlab για τη μέθοδο δυνάμεων

- i) Αποδείξτε πως η Μέθοδος Δύναμης, όταν συγκλίνει, βρίσκει τη μέγιστη κατ' απόλυτη τιμή, ιδιοτιμή ενός μητρώου καθώς και το αντίστοιχο ιδιοδιάνυσμα.
- ii) Υλοποιήστε τη Μέθοδο Δύναμης ως συνάρτηση σε MATLAB. Για το μητρώο $A = [3 \ 2 \ 0; 2 \ 4 \ 2; 0 \ 2 \ 5]$, υπολογίστε ένα-δύο βήματα με το χέρι και στη συνέχεια βρείτε το μέγιστο ιδιοζεύγος μέσω της συνάρτησης που φτιάξατε.

Λύση

- i) Αρχικά μέσω του αλγορίθμου:

for k = 1, 2, ...

$$z^{(k)} = Aq^{(k-1)}$$

$$q^{(k)} = \frac{z^{(k)}}{\|z^{(k)}\|}$$

$$\lambda^{(k)} = (q^{(k)})^T A q^{(k)}$$

end

Σχέση ιδιοτιμών και ιδιοδιανυσμάτων: $A * x = \lambda * x$ και επίσης ισχύει: $A^k * x = \lambda^k * x$

εξετάζουμε αν η μέθοδος των δυνάμεων συγκλίνει. **Με την προϋπόθεση ότι το μητρώο A είναι διαγωνοποιησιμό, υπάρχουν η ανεξάρτητα ιδιοδιανύσματα $x_1, x_2, x_3, \dots, x_n$ για τα οποία ισχύει ότι το αρχικό διάνυσμα**

$q^{(0)} = a_1 * x_1 + a_2 * x_2 + a_3 * x_3 + \dots + a_n * x_n$ με τα $a_1, a_2, a_3, \dots, a_n \in \mathbb{R}$, να αντιπροσωπεύουν τους συντελεστές. Πολλαπλασιάζουμε και τα δύο μέλη της προηγούμενης εξίσωσης από αριστερά με το μητρώο A^k , οπότε:

$$A^k * q^{(0)} = A^k * (a_1 * x_1 + a_2 * x_2 + a_3 * x_3 + \dots + a_n * x_n) = a_1 * A^k * x_1 + a_2 * A^k * x_2 + a_3 * A^k * x_3 + \dots + a_n * A^k * x_n = a_1 * \lambda_1^k * x_1 + a_2 * \lambda_2^k * x_2 + a_3 * \lambda_3^k * x_3 + \dots + a_n * \lambda_n^k * x_n = a_1 * \lambda_1^k * \left(x_1 + \frac{a_2 * \lambda_2^k * x_2}{a_1 * \lambda_1^k} + \frac{a_3 * \lambda_3^k * x_3}{a_1 * \lambda_1^k} + \dots + \frac{a_n * \lambda_n^k * x_n}{a_1 * \lambda_1^k} \right) = a_1 * \lambda_1^k * \left(x_1 + \sum_{j=2}^n \frac{a_j (\lambda_j / \lambda_1)^k}{a_1} x_j \right) = a_1 * \lambda_1^k * \left(x_1 + \sum_{j=2}^n \frac{a_j}{a_1} \left(\frac{\lambda_j}{\lambda_1}\right)^k x_j \right), \text{ αν } \theta \text{ θεωρήσουμε ότι } |\lambda_1| >$$

$|\lambda_2| > |\lambda_3| > \dots > |\lambda_n|$, τότε το λ_1 **χαρακτηρίζεται ως κυρίαρχη ιδιοτιμή** και έχουμε ότι το κλάσμα $\left(\frac{\lambda_j}{\lambda_1}\right)^k$ είναι μικρότερο της «1» και καθώς το k αυξάνεται για $k = 2, \dots, n$, προσεγγίζει το «0». Επομένως, με την προϋπόθεση ότι $a_1 \neq 0$ το $x^{(k)} = A^k * x_0 \approx \lambda_1^k * a_1 * x_1$. Σύμφωνα με τα πηλίκα Rayleigh, αν το (λ, x) είναι **ιδιοζεύγος**, ισχύει ότι: $\frac{x^T A x}{x^T x} = \lambda$. Έτσι τελικά έχουμε ότι: $\frac{(x^{(k)})^T A x^{(k)}}{(x^{(k)})^T x^{(k)}} = \lambda$.

Σημείωση: Όταν υψώνουμε ένα κλάσμα που είναι μικρότερο του «1» σε μια δύναμη, τότε όσο μεγαλώνει ο εκθέτης της δύναμης, το κλάσμα μικραίνει και μπορούμε να θεωρήσουμε ότι για ένα μεγάλο εκθέτη, το κλάσμα \rightarrow «0».

ii) Ο κώδικας Matlab για τη μέθοδο των δυνάμεων περιγράφεται στη συνέχεια. Ο κώδικας αυτός υπολογίζει στη μεταβλητή **lambda** την κυρίαρχη ιδιοτιμή του μητρώου A και στη μεταβλητή **x** το ιδιοδιάνυσμα που αντιστοιχεί στην κυρίαρχη ιδιοτιμή και στη μεταβλητή **iter** (iteration = επανάληψη) τον αριθμό της επανάληψης στην οποία υπολογίστηκε το ιδιοδιάνυσμα **x** του μητρώου A . Ως **ορίσματα εισόδου** στη συνάρτηση **eigpower**

δίνουμε το μητρώο **A** του οποίου θα υπολογίσουμε την κυριαρχη ιδιοτιμή, την ανοχή **tol** (tolerance), το μέγιστο αριθμό επαναλήψεων **nmax** καθώς και το αρχικό ιδιοδιάνυσμα **x0**.

```

function [lambda, x, iter] = eigpower(A, tol ,nmax, x0)
%EIGPOWER Numerically evaluate one eigenvalue of a real matrix.
% LAMBDA = EIGPOWER(A,TOL ,NMAX ,X0) uses an absolute error tolerance TOL (the default is 1.e-6) and a maximum
%number of iterations NMAX (the default is 100), starting from the initial vector X0.
% [LAMBDA, V, ITER] = EIGPOWER(A, TOL, NMAX ,X0) also returns the eigenvector v such that A * v = LAMBDA * v
% and the iteration number at which v was computed.
[n, m] = size(A); % επιστρέφει στις μεταβλητές n και m τις διαστάσεις του A.
if n~=m, error('Only for square matrices'); end %αν το μητρώο δεν είναι τετραγωνικό, τέλος της συνάρτησης
% με την παράλληλη εμφάνιση διευκρινιστικού μηνύματος.

if nargin == 1 % Number of function input arguments (αριθμός ορισμάτων εισόδου). Αν αυτός ο αριθμός που εκφράζεται πάντα με τη μεταβλητή nargin είναι ίσος με 1, τότε η μοναδική παράμετρος εισόδου που χρειάζεται να δοθεί είναι το μητρώο A. Οι υπόλοιπες τρεις παράμετροι εισόδου είναι αυτές που καθορίζονται με τις τρεις αντίστοιχες επόμενες εντολές (και από ότι μπορούμε να παρατηρήσουμε, οι τιμές που έχουν αυτές οι τρεις μεταβλητές είναι οι προεπιλεγμένες). Αν τις δίναμε εκτός της συνάρτησης eigpower, ως εξωτερικά ορίσματα εισόδου, θα είχαμε τη δυνατότητα να δώσουμε άλλες τιμές, διαφορετικές από τις προεπιλεγμένες:
    tol = 1.e-06; % Ακρίβεια, δηλαδή ο αριθμός των δεκαδικών ψηφίων
    x0 = ones(n, 1); % x0 = [1 1 1 .....1]', είναι το αρχικό ιδιοδιάνυσμα
    nmax=100; % Μέγιστος αριθμός επιτρεπτών επαναλήψεων
end
x0=x0/norm(x0); %κανονικοποίηση του αρχικού (δοθέντος) ιδιοδιανύσματος
pro=A * x0; % x^(k) = A * y^(k-1)
lambda=x0'*pro; % λ = x0' * A * x0
err = tol * abs(lambda)+1; % err = 1.e-6 * |λ| + 1
iter = 0; % πλήθος επαναλήψεων
while err > tol * abs(lambda) & abs(lambda) ~= 0 & iter < =nmax % αν η ακρίβεια του υπολογισμού γίνει μικρότερη
%από την ακρίβεια που έχουμε καθορίσει με τη μεταβλητή tol, τότε σταματά η επανάληψη. Εναλλακτικά, καθορίζουμε και ένα μέγιστο (επιτρεπτό) αριθμό επαναλήψεων, έτσι ώστε ακόμη και αν η ακρίβεια των υπολογισμών δεν πέσει κάτω από την επιθυμητή, να σταματήσει ούτως ή άλλως η επανάληψη, όταν ξεπεραστούν οι nmax επαναλήψεις. Στη συνέχεια, μέσα στην επανάληψη, συνεχίζουμε σε κάθε βήμα της να υπολογίζουμε την ιδιοτιμή λ, στο lamdanew
    x=pro; x=x/norm(x);
    pro=A * x; lamdanew = x' * pro;
    err = abs(lamdanew - lambda);
    lambda = lamdanew;
    iter = iter+1;
end
return

```

Αν εκτελέσουμε τον ανωτέρω κώδικα για την περίπτωση του μητρώου $A = \begin{bmatrix} 3 & 2 & 0 \\ 2 & 4 & 2 \\ 0 & 2 & 5 \end{bmatrix}$ και για δεδομένα εισό-

δου που είναι $tol = 1.e-12 = 1 * 10^{-12}$ και $x0 = [1 1 1]'$ και $nmax = 2$, τα οποία πληκτρολογούμε στη γραμμή εντολών της MATLAB, όπως φαίνεται παρακάτω, θα πάρουμε τα ακόλουθα αποτελέσματα:

```

>> tol = 1.e-012;
>> x0 = [1 1 1]';
>> nmax = 2
>> eigpower(A, tol, nmax, x0) % κλήση της function eigpower από τη γραμμή εντολών της MATLAB
% επειδή δεν προβλέπουμε σε ποιες μεταβλητές θα επιστραφούν
% τα αποτελέσματα, η function eigpower επιστρέφει μόνο ένα

```

% αποτέλεσμα και συγκεκριμένα το πρώτο, που είναι το lambda.

% κυρίαρχη ιδιοτιμή του μητρώου A

Αν στη συνέχεια υπολογίζουμε τις ιδιοτιμές του μητρώου A, χρησιμοποιώντας αντί για τη function eigpower, την **ενδογενή συνάρτηση eig** της Matlab (που υπολογίζει ιδιοτιμές ενός μητρώου), θα έχουμε ότι:

```
>> eig(A)
```

```
ans =
```

```
1.0000
```

```
4.0000
```

```
7.0000
```

Παρατηρούμε ότι η function eigpower επέστρεψε ένα αποτέλεσμα **πολύ κοντά στο πραγματικό** (δηλαδή $6.9958 \approx 7.0000$). Ένας εναλλακτικός κώδικας MATLAB για τη **μεθόδου των δυνάμεων**, φαίνεται στη συνέχεια. Στο μητρώο **XALL** (σε κάθε στήλη του) επιστρέφονται τα κανονικοποιημένα ιδιοδιανύσματα από το κάθε βήμα.

```
function [lamda,y,XALL,error ,iter]=mypowmeth(A,x ,tol)
%Power method for the computation of the largest eigenvalue and the
%corresponding eigenvector of matrix A
%
% input   A      REAL matrix
%          x      REAL initial guess vector
%          tol    REAL error tolerance
%
% output  lamda  REAL solution eigenvalue
%          y      REAL solution eigenvector
%          XALL   REAL matrix containing the approximations of each step
%          error   REAL error norm
%          iter   INTEGER number of iterations performed
y=x/norm(x); lamda=y'*A*y;
error=norm(A*y-lamda*y);
if ( error < tol ) return , end

iter=0; XALL=[];
while ( error>tol )
    z=A*y; y=z/norm(z);
    lamda=y'*A*y;
    error=norm(A*y-lamda*y);
    iter=iter+1; XALL(:, iter)=y;
end
%END mypowmeth.m
```

3.1.7 Παράδειγμα αντίστροφης δύναμης με μετατόπιση

Σε αυτή τη μέθοδο -με την οποία υπολογίζουμε και πάλι την κυρίαρχη ιδιοτιμή ενός μητρώου- **αν έχουμε κάποια προσέγγιση λ για μια ιδιοτιμή** ή ψάχνουμε ιδιοζεύγη με ιδιοτιμή κοντά στο δοθέν λ, τότε το **αρχικό ιδιοδιάνυσμα $x^{(0)}$** είναι: $x^{(0)} = \xi_1 u_1 + \dots + \xi_n u_n$, με $\xi_1 \neq 0$ και επειδή ισχύει $A * x = \lambda * x$, αν θέσουμε όπου **A** το μητρώο $(A - \lambda * I)^{-k}$ και όπου λ το $(\lambda_1 - \lambda)^{-k}$ έχουμε ότι η προηγούμενη εξίσωση $A * x = \lambda * x$ μπορεί να γραφτεί (αντικαθιστώντας το διάνυσμα x με το διάνυσμα $x^{(0)}$) στην εξής μορφή: $A * x^{(0)} = \lambda * x^{(0)} \Rightarrow A * x^{(0)} = \lambda * (\xi_1 u_1 + \dots + \xi_n u_n) \Rightarrow A * x^{(0)} = \xi_1 * \lambda * u_1 + \xi_2 * \lambda * u_2 + \dots + \xi_n * \lambda * u_n \Rightarrow$ αντικαθιστούμε το A και το λ → $(A - \lambda * I)^{-k} * x^{(0)} = \xi_1 (\lambda_1 - \lambda)^{-k} * u_1 + \dots + \xi_n (\lambda_n - \lambda)^{-k} * u_n$. Αν για κάποια ιδιοτιμή λ_i ισχύει ότι $|\lambda - \lambda_i| \ll \min |\lambda - \lambda_j|$ με i ≠ j, τότε στην τελευταία εξίσωση κυριαρχεί ο όρος $\xi_i (\lambda_i - \lambda)^{-k} * u_i$, όπως φαίνεται στη συνέχεια: $(A - \lambda * I)^{-k} * x^{(0)} = \xi_1 (\lambda_1 - \lambda)^{-k} * u_1 + \dots + \xi_i (\lambda_i - \lambda)^{-k} * u_i + \dots + \xi_n (\lambda_n - \lambda)^{-k} * u_n$. Ο κώδικας MATLAB για την **αντίστροφη μέθοδο δύναμης**

με μετατόπιση, η οποία υπολογίζει **τόσο την κυρίαρχη ιδιοτιμή ενός μητρώου** (συμβολίζεται με τη μεταβλητή **lambda** στον παρακάτω κώδικα) όσο και το **ιδιοδιάνυσμα που αντιστοιχεί στην ιδιοτιμή αυτή** (συμβολίζεται με τη μεταβλητή **x** στον παρακάτω κώδικα) είναι:

```

function [lambda, x, iter] = INVSHIFT(A, mu, tol , nmax, x0)
% INVSHIFT inverse power method with shift
% LAMBDA = INVSHIFT(A) computes the eigenvalue of A of minimum modulus (εννοεί την κυρίαρχη ιδιοτιμή) with
% the inverse power method (Υπολογισμός κυρίαρχης ιδιοτιμής μητρώου A)
% LAMBDA = INVSHIFT(A, MU) computes the eigenvalue of A closest to the given number (real or complex) MU.
% Δηλαδή η παράμετρος mu είναι η προσέγγιση της κυρίαρχης ιδιοτιμής
% LAMBDA = INVSHIFT(A, MU, TOL , NMAX, X0) uses an absolute error tolerance TOL (the default is 1. e - 6) and a
% maximum number of iterations NMAX (the default is 100), starting from the initial vector x0 .
% [LAMBDA, x, ITER] = INVSHIFT(A, MU, TOL , NMAX, X0 ) also returns the eigenvector x such that A * x = LAMBDA * x
% and the iteration number at which x was computed.

[n, m] = size(A);
if n ~= m, error('Only for square matrices'); end %τερματισμός συνάρτησης, όταν το μητρώο είναι μη τετραγωνικό
if nargin == 1      % Αναγκαστικά δίνουμε τότε μέσα στη function τα υπόλοιπα ορίσματα εισόδου, όπως φαίνεται:
    x0 = rand(n, 1); nmax = 100; tol = 1.e-06; mu = 0;
else if nargin == 2 % Αυτό σημαίνει ότι έχουμε δώσει και τον αρχικό αριθμό μη από την αρχή, ως όρισμα εισόδου
    x0 = rand(n,1); nmax = 100; tol = 1.e-06;
end
[L, U] = lu(A - mu * eye( n)); % Η αντίστροφη μέθοδος δύναμης με μετατόπιση χρησιμοποιεί LU παραγοντοποίηση.
% Το μητρώο A έχει μια μετατόπιση που είναι η διαφορά A - mu * eye( n) και σε αυτή τη μετατόπιση εφαρμόζεται η LU
% παραγοντοποίηση
if norm (x0) == 0, x0 = rand (n, 1); end; % αν δόθηκε ως αρχικό ιδιοδιάνυσμα το μηδενικό τότε δίνουμε εμείς κάποιο
% άλλο με τυχαίες τιμές, μη μηδενικό. Δηλαδή με αυτόν τον έλεγχο εξασφαλίζουμε
% ότι δεν μπορούμε να δώσουμε το μηδενικό ιδιοδιάνυσμα ως αρχικό ιδιοδιάνυσμα
x0 = x0/norm(x0); z0 = L\ x0; pro = U\ z0; % εμπρός και πίσω αντικατάσταση, με εξασφαλισμένο ότι το x0 ≠ 0.
lambda = x0' * pro; err = tol * abs (lambda) + 1; iter= 0 ; %lambda = η αρχική εκτίμηση της κυρίαρχης ιδιοτιμής

while err > tol*abs (lambda) & abs (lambda) ~=0 & iter<= nmax % ίδιες συνθήκες με αυτές της κανονικής μεθόδου των
% δυνάμεων. Στη συνέχεια εκτελούμε ακριβώς τις ίδιες εντολές με αυτές που εκτελούσαμε πριν από την επανάληψη.
x = pro; x = x / norm (x); z = L\ x; pro =U\ z; lambdanew = x' *pro; % Η νέα κυρίαρχη ιδιοτιμή lambdanew υπολογί-
ζεται επί της λύσης x του γραμμικού συστήματος A * x = b.
err = abs (lambdanew - lambda);           % εύρεση σφάλματος
lambda = lambdanew; iter = iter+ 1;        % Παρατηρούμε ότι και οι τρεις μεταβλητές που εμφανίζονται στις τρεις
% αντίστοιχες συνθήκες της while, δηλ. οι μεταβλητές err, lambda, iter ολλάζουν
% τιμή μέσα στην επανάληψη, έτσι ώστε κάποια στιγμή να γίνει ψευδής μια
% τουλάχιστον από τις τρεις συνθήκες που ελέγχονται και λόγω του τελεστή &
% να σταματήσει η επανάληψη.

end
lambda = 1/lambda + mu ; return

```

Όπως μπορούμε να παρατηρήσουμε από τον παραπάνω κώδικα, η αντίστροφη μέθοδος δύναμης με μετατόπιση χρησιμοποιεί **LU παραγοντοποίηση** για να διασπάσει το μητρώο A και στη συνέχεια υπολογίζει τη λύση του γραμμικού συστήματος **A * pro = x**, με εμπρός αντικατάσταση **z= L\ x** και πίσω αντικατάσταση **pro =U\ z = U\ L\ x ()** και στη συνέχεια υπολογίζει από τη λύση αυτή, την ιδιοτιμή με την εντολή **lambdanew = x' * pro**.

Παράδειγμα (Θέμα Ιουνίου 2016)

Έστω ένα μητρώο $A \in R^{n \times n}$ για το οποίο γνωρίζουμε ότι η μεγαλύτερη ιδιοτιμή του είναι 100 και μία άλλη ιδιοτιμή του, είναι πολύ κοντά (αλλά όχι ίση) με -4. Ποια σειρά βημάτων παρακάτω ενδείκνυνται περισσότερο για να υπολογιστεί γρήγορα μία προσέγγιση αυτής της ιδιοτιμής στη μεταβλητή est; (μπορείτε να θεωρήσετε ότι $I = eye(n)$).

a) $[L, U] = lu(A + 4 * I); x = ones(n, 1);$

```
for j = 1: 8, x = x/norm (x); pro = x; x = U\(\(L\|pro); end; est = 1/(x' * pro) -4
```

b) $x = ones(n, 1); x = x / norm(x);$

```
for j = 1:8, x = A * x; x = x/ norm (x); end; est = - (A * x)./x/4;
```

c) $x = eye(n, 1);$

```
for j = 1:8, x = (A + 0. * I) * x; x = x/norm (x); end; est = max ( ( A * x) ./x);
```

d) $x = ones(n, 1);$

```
for j = 1:8, x = x/norm (x); pro = x; x = (A + 4 * I)\pro; end; est = 1/ (x'*pro) -4
```

Λύση

Πρόκειται για ερώτημα πολλαπλής επιλογής στο οποίο ελέγχουμε τη λειτουργία διαφόρων εκδοχών της μεθόδου των δυνάμεων και ειδικότερα **την μέθοδο αντίστροφης δύναμης με μετατόπιση**. Από την εκφώνηση διαπιστώνουμε ότι έχει (μοναδική) κυρίαρχη ιδιοτιμή ίση με 100, επομένως μπορεί να εφαρμοστεί η μέθοδος των δυνάμεων, αλλά και της αντίστροφης δύναμης με μετατόπιση για τον υπολογισμό της (επειδή αυτές οι δύο μέθοδοι έχουν την ίδια προϋπόθεση) και επιπλέον δίνεται και η προσέγγιση της κυρίαρχης ιδιοτιμής, γεγονός που παραπέμπει στη **χρήση της μεθόδου αντίστροφη δύναμη με μετατόπιση**.

Η σωστή απάντηση είναι η (a) που υλοποιεί την μέθοδο της αντίστροφης δύναμης με μετατόπιση με αποτελεσματικό τρόπο, καθότι παραγοντοποιεί το (μετατοπισμένο) μητρώο $A + 4 * I$ μία φορά και εντός του βρόχου και επαναχρησιμοποιεί τους παράγοντες L, U για τη λύση (σε σχέση με τον κώδικα που έχει προηγηθεί, η παράμετρος tu που πολλαπλασιάζει το ταυτοτικό μητρώο είναι ίση με -4). Αξίζει να σημειωθεί ότι με τη χρήση της LU εκτός του βρόχου, υπάρχει μεγάλη μείωση του κόστους από την επιλογή (d). Ειδικότερα, η μεν (d) στοιχίζει περίπου $niter * x (an^3 + bn^2)$ ενώ η (a) περίπου $an^3 + \beta niter * n^2$, όπου $niter$ είναι το πλήθος των επαναλήψεων του βρόχου ως τη σύγκλιση και a, b μικρές σταθερές). Στην περίπτωση (d) υπολογίζεται το αντίστροφο μητρώο, ενώ στην (a) περίπτωση αυτό γίνεται μόνο μια φορά πριν την επανάληψη, όπου εκτελείται η LU παραγοντοποίηση.

3.2 Κύκλοι Gershgorin

Οι κύκλοι (ή δίσκοι) Gershgorin είναι ένας τρόπος υπολογισμού **των διαστημάτων** στα οποία ορίζονται οι ιδιοτιμές ενός μητρώου, χωρίς όμως να γίνει ακριβής υπολογισμός τους. Η βασική σχέση που περιγράφει το θεώρημα κύκλων του Gershgorin είναι: $\Delta_i^{(r)} = \left\{ z \in C : |z - a_{ii}| \leq \sum_{j=1, j \neq i}^n |a_{ij}| \right\}$, όπου $\Delta_i^{(r)}$ οι δίσκοι (κύκλοι)

Gershgorin. Έστω για το μητρώο: $A = \begin{bmatrix} 2 & -\frac{1}{2} & 0 & -\frac{1}{2} \\ 0 & 4 & 0 & 2 \\ -\frac{1}{2} & 0 & 6 & \frac{1}{2} \\ 0 & 0 & 1 & 9 \end{bmatrix}$ τα διαστήματα των ιδιοτιμών του κατά γραμμές είναι:

$$|z - 2| \leq 1 \Rightarrow -1 \leq z - 2 \leq 1 \Rightarrow 1 \leq z \leq 3$$

$$|z - 4| \leq 2 \Rightarrow 2 \leq z \leq 6$$

$$|z - 6| \leq 1 \Rightarrow 5 \leq z \leq 7$$

$$|z - 9| \leq 1 \Rightarrow 8 \leq z \leq 10$$

Διαστήματα ιδιοτιμών μητρώου A κατά στήλες:

$$|z - 2| \leq \frac{1}{2} \Rightarrow \frac{3}{2} \leq z \leq \frac{5}{2}$$

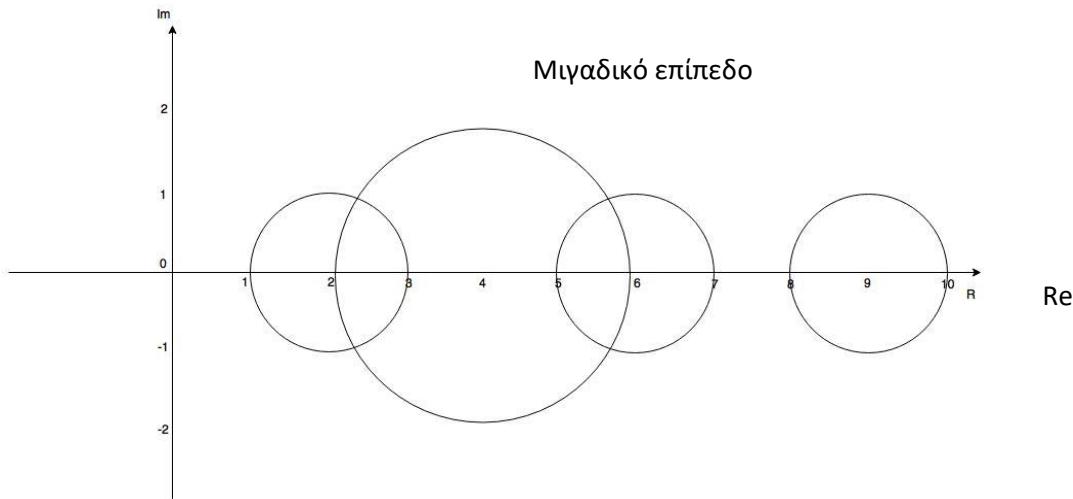
$$|z - 4| \leq \frac{1}{2} \Rightarrow \frac{7}{2} \leq z \leq \frac{9}{2}$$

$$|z - 6| \leq 1 \Rightarrow 5 \leq z \leq 7$$

$$|z - 9| \leq 3 \Rightarrow 6 \leq z \leq 12$$

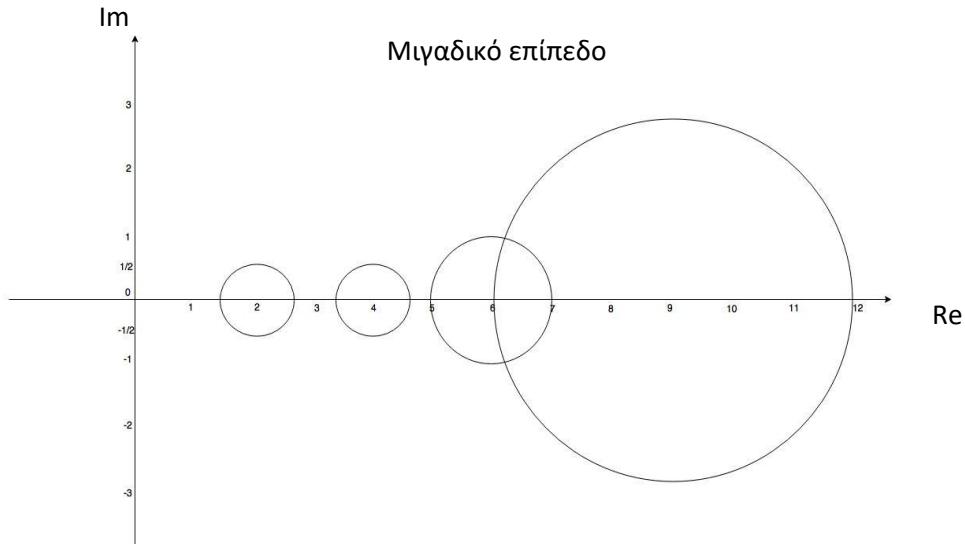
Επιλέγουμε τα **στοιχεία** της κύριας διαγωνίου του μητρώου A ως κέντρα των κύκλων και ως ακτίνα επιλέγουμε το άθροισμα των υπόλοιπων στοιχείων της γραμμής ή της στήλης κατά απόλυτη τιμή. Αν το μητρώο που εξετάζουμε είναι **συμμετρικό** και αν τα διαστήματα στα οποία ορίζονται οι ιδιοτιμές του μητρώου είναι **θετικά**, τότε όλες οι ιδιοτιμές ενός μητρώου είναι θετικές, με αποτέλεσμα το μητρώο να είναι Σ.Θ.Ο. (εφόσον είναι ήδη συμμετρικό). Όταν όλα τα στοιχεία ενός μητρώου είναι **πραγματικά**, οι ιδιοτιμές του εμφανίζονται **σε ζεύγη**, επομένως πρέπει να ανήκουν στον ίδιο κύκλο Gershgorin. Κάθε φορά που εμφανίζονται **απομονωμένοι** κύκλοι, καθένας από αυτούς περιέχει μια **πραγματική** ιδιοτιμή. Αν μετά την **αφαίρεση** των **απομονωμένων κύκλων**, απομένει **περιπτώς** αριθμός κύκλων, τότε θα υπάρχει σίγουρα μια **πραγματική** ιδιοτιμή και οι υπόλοιπες θα είναι είτε **ζεύγη πραγματικών** είτε **ζεύγη συζυγών μιγαδικών**. Αν μετά την **αφαίρεση** των **απομονωμένων κύκλων**, απομένει **άρτιος** αριθμός κύκλων, τότε αυτές θα είναι είτε **ζεύγη πραγματικών** είτε **ζεύγη συζυγών μιγαδικών**.

Αν κάνουμε την απεικόνιση των κύκλων κατά γραμμές/στήλες για το προηγούμενο μητρώο A, θα έχουμε τα σχήματα των Εικόνων 5 και 6 που ακολουθούν: Σε αυτά, ο άξονας x είναι ο πραγματικός (Re) και ο άξονας y είναι ο φανταστικός (Im) και η τομή τους δημιουργεί το **μιγαδικό επίπεδο**. Αν οι ιδιοτιμές είναι πάνω στον άξονα Re, είναι πραγματικές (δηλαδή έχουν **μόνο** πραγματικό μέρος).

Γραμμές

Εικόνα 5: Απεικόνιση ιδιοτιμών κατά γραμμές (κύκλοι Gershgorin κατά γραμμές)

Δεν υπάρχει καμιά εγγύηση ότι ένας κύκλος πρέπει να περιέχει ιδιοτιμή, εκτός και αν είναι απομονωμένος και μάλιστα οι απομονωμένοι κύκλοι έχουν πραγματικές ιδιοτιμές.

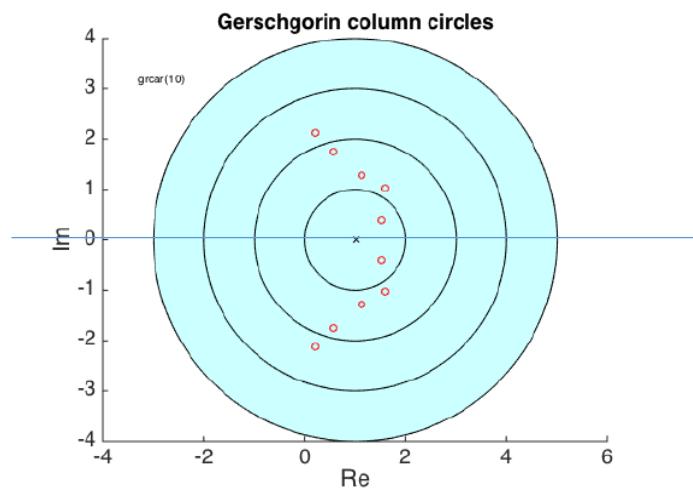
Στήλες

Εικόνα 6: Απεικόνιση ιδιοτιμών κατά στήλες (κύκλοι Gershgorin κατά στήλες)

Όλες οι ιδιοτιμές ενός διθέντος μητρώου $A^{n \times n}$ ανήκουν στην περιοχή του μιγαδικού επιπέδου, η οποία είναι η τομή των δύο περιοχών που σχηματίζονται αντίστοιχα από την ένωση των κύκλων γραμμών και την ένωση των κύκλων στηλών.

Παρατήρηση: Αν σε ένα μητρώο απεικονίσουμε τους κύκλους Gershgorin –κατά στήλες– και πάρουμε το επόμενο σχήμα που αποτελείται από **ομόκεντρους κύκλους**, τότε συμπεραίνουμε ότι το μητρώο αυτό **περιέχει μιγαδικές ιδιοτιμές** (με πραγματικό και φανταστικό μέρος) και μάλιστα **συζυγείς μιγαδικές ιδιοτιμές** (δηλαδή είναι συμμετρικές ως προς τον οριζόντιο άξονα, π.χ. αν μια ιδιοτιμή είναι $3 + 4i$, τότε η άλλη ιδιοτιμή είναι $3 - 4i$).

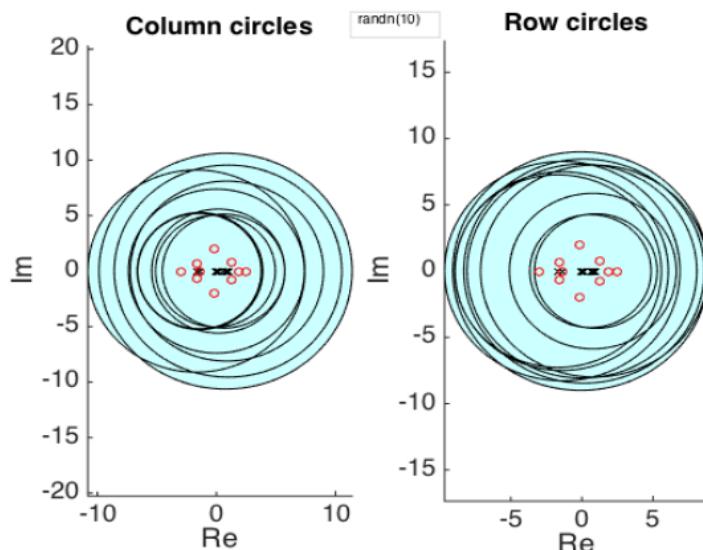
Με 'o' οι ιδιοτιμές, με 'x' τα κέντρα των δίσκων.



Εικόνα 7: Απεικόνιση ιδιοτιμών μητρώου κατά στήλες

Παρατήρηση: Αν σε ένα μητρώο που **δεν** είναι συμμετρικό, απεικονίσουμε τους κύκλους Gershgorin κατά **στήλες** και κατά **γραμμές** θα πάρουμε το επόμενο σχήμα που δείχνει τη διαφορετικότητα των δύο περιπτώσεων. Όμως παρά το γεγονός ότι οι κύκλοι είναι διαφορετικοί στις δύο περιπτώσεις, οι ιδιοτιμές τους είναι **ίδιες**.

Με 'o' οι ιδιοτιμές, με 'x' τα κέντρα των δίσκων.



Εικόνα 8: Απεικόνιση ιδιοτιμών μητρώου κατά γραμμές και κατά στήλες

3.2.1 Πρώτο θέμα με κύκλους Gershgorin

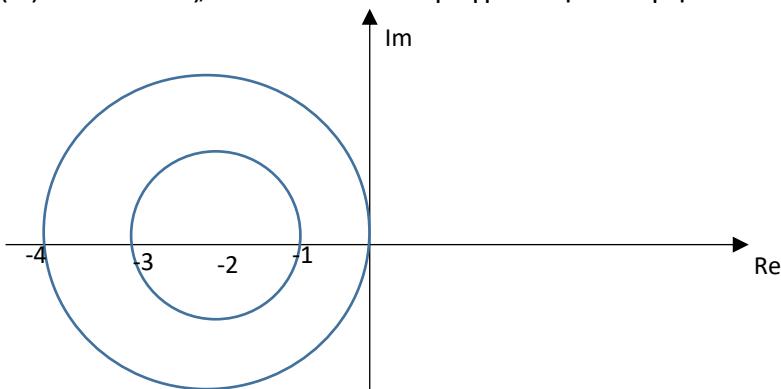
Έστω το τριδιαγώνιο μητρώο $C = \begin{bmatrix} -2 & 1 & & \\ -1 & -2 & 1 & \\ \dots & \dots & \dots & \dots \\ & -1 & -2 & 1 \\ & & -1 & -2 \end{bmatrix}$ για το οποίο ζητάμε να επιλέξουμε τη σωστή/σωστές απαντήσεις, μεταξύ των επόμενων:

- (i) Το μητρώο είναι συμμετρικό και θετικά ορισμένο.
- (ii) Το μητρώο δεν είναι συμμετρικό και θετικά ορισμένο.

- (iii) Το μητρώο έχει τουλάχιστον μια ιδιοτιμή 0.
- (iv) Το μητρώο έχει τουλάχιστον μια μιγαδική ιδιοτιμή.

Λύση

Είναι φανερό από τη μορφή του μητρώου ότι **δεν** είναι συμμετρικό και επίσης **δεν** είναι θετικά ορισμένο, αφού δεν έχει Α.Δ.Κ. και θετικά διαγώνια στοιχεία. Επομένως σωστή απάντηση είναι καταρχήν η (ii). Αν στη συνέχεια χρησιμοποιήσουμε το θεώρημα των κύκλων Gershgorin για να απεικονίσουμε κατά γραμμές ή κατά στήλες (δεν έχει σημασία, αφού το μητρώο είναι τριδιαγώνιο), τότε θα πάρουμε το σχήμα της Εικόνας 7, που ακολουθεί στη συνέχεια. Από την Εικόνα αυτή παρατηρούμε ότι υπάρχουν δύο κύκλοι και μάλιστα **οιμόκεντροι**. Από το γεγονός αυτό συμπεραίνουμε ότι θα υπάρχει σίγουρα **μια τουλάχιστον μιγαδική ιδιοτιμή** στο μητρώο. Επομένως σωστή απάντηση είναι και η (iv). Άλλωστε, σύμφωνα με την Εικόνα 7, **όλες** οι ιδιοτιμές του μητρώου είναι **μιγαδικές**. Επειδή **δεν** υπάρχουν απομονωμένοι κύκλοι, **δεν** υπάρχουν πραγματικές ιδιοτιμές. Επομένως το (iii) είναι λάθος, διότι το 0 είναι πραγματική ιδιοτιμή:



Εικόνα 7: Απεικόνιση των κύκλων Gershgorin μητρώου C

Στους κύκλους Gershgorin ισχύουν δύο βασικές ιδιότητες:

- α)** Όλες οι ιδιοτιμές ενός μητρώου ανήκουν στο χωρίο (διάστημα) του μιγαδικού επιπέδου που αποτελείται από την **τομή** του συνόλου των δίσκων γραμμής και των δίσκων στήλης.
- β)** Αν r δίσκοι γραμμής (ή αντίστοιχα στήλης) όπου $1 \leq r \leq n$, **δεν** έχουν κανένα κοινό σημείο με το σύνολο των υπολοίπων $n - r$ δίσκων (δηλαδή είναι απομονωμένοι), τότε οι r δίσκοι περιέχουν r ακριβώς πραγματικές ιδιοτιμές.

3.2.2 Δεύτερο Θέμα με κύκλους Gershgorin (παλιό θέμα)

Έστω το μητρώο $A = \begin{bmatrix} 1 & 3 & 3 \\ 3 & 2 & 1 \\ 2 & 3 & 2 \end{bmatrix}$ για το οποίο ζητάμε να υπολογίσουμε αν η **μεγαλύτερη** ιδιοτιμή του είναι

- κοντά στο α) 6.0, β) 10.0 γ) 8.0, δ) 4.0

Λύση

Αν εφαρμόσουμε το **θεώρημα των κύκλων Gershgorin** κατά γραμμές είναι:

$$|z - 1| \leq 6 \Rightarrow -6 \leq z - 1 \leq 6 \Rightarrow -5 \leq z \leq 7$$

$$|z - 2| \leq 4 \Rightarrow -4 \leq z - 2 \leq 4 \Rightarrow -2 \leq z \leq 6$$

$$|z - 2| \leq 5 \Rightarrow -5 \leq z - 2 \leq 5 \Rightarrow 1 \Rightarrow -3 \leq z \leq 7$$

Το διάστημα στο οποίο συναληθεύουν και οι τρεις ανισότητες είναι το $[-2, 6]$ και στο διάστημα αυτό ανήκουν οι ιδιοτιμές 4, 6 από τις πιθανές απαντήσεις που έχουν δοθεί στην εκφώνηση. Μεταξύ αυτών των δυο, **μεγαλύτερη ιδιοτιμή είναι το 6.**

Αν πάρουμε τις ιδιοτιμές **κατά στήλες** θα έχουμε:

$$|z - 1| \leq 5 \Rightarrow -5 \leq z - 1 \leq 5 \Rightarrow -4 \leq z \leq 6$$

$$|z - 2| \leq 6 \Rightarrow -6 \leq z - 2 \leq 6 \Rightarrow -4 \leq z \leq 8$$

$$|z - 2| \leq 4 \Rightarrow -4 \leq z - 2 \leq 4 \Rightarrow 1 \Rightarrow -2 \leq z \leq 6$$

Το διάστημα στο οποίο συναληθεύουν και οι τρεις ανισότητες είναι και πάλι το $[-2, 6]$ και ισχύει ότι αναφέρθηκε και πριν. **Αν εναλλακτικά θέλαμε να εφαρμόσουμε τη μέθοδο των δυνάμεων για να υπολογίσουμε την κυρίαρχη ιδιοτιμή του δοθέντος μητρώου, δεν θα μπορούσαμε να την χρησιμοποιήσουμε, διότι δεν δίνεται το αρχικό ιδιοδιάνυσμα $x^{(0)}$.** Επίσης, ως **επαλήθευση** αν χρησιμοποιήσουμε τη συνάρτηση **eig** της MATLAB -ως επαλήθευση- για τον ακριβή υπολογισμό των ιδιοτιμών του μητρώου A, θα πάρουμε τα εξής:

```
>> A=[1 3 3;3 2 1;2 3 2]
```

```
A =
```

$$\begin{bmatrix} 1 & 3 & 3 \\ 3 & 2 & 1 \\ 2 & 3 & 2 \end{bmatrix}$$

```
>> eig(A)
```

```
ans =
```

$$\begin{bmatrix} 6.6056 \\ -1.0000 \\ -0.6056 \end{bmatrix}$$

Από αυτές επαληθεύουμε ότι η **μεγαλύτερη ιδιοτιμή του δοθέντος μητρώου είναι κοντά στο 6.**

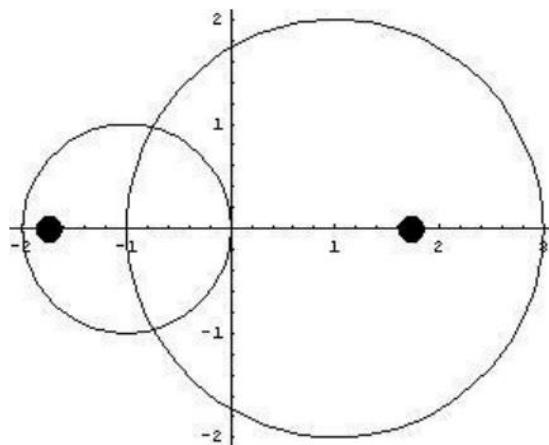
3.2.3 Άσκηση με κύκλους Gershgorin

Έστω το μητρώο $A = \begin{bmatrix} 1 & 2 \\ 1 & -1 \end{bmatrix}$ για το οποίο ζητάμε να υπολογίσουμε τις ιδιοτιμές του με δύο τρόπους: α) με χαρακτηριστικό πολυώνυμο και β) με κύκλους Gershgorin.

Λύση

Αν υπολογίσουμε καταρχήν τις ιδιοτιμές του μητρώου A μέσω **χαρακτηριστικού πολυωνύμου** $\det(\lambda * I - A) = 0 \Rightarrow \det \begin{pmatrix} \lambda - 1 & -2 \\ -1 & \lambda + 1 \end{pmatrix} = 0 \Rightarrow (\lambda - 1) * (\lambda + 1) - 2 = 0 \Rightarrow \lambda^2 - 3 = 0 \Rightarrow \lambda = \pm \sqrt{3}$. Όσον αφορά τον υπολογισμό τους μέσω κύκλων (δίσκων) Gershgorin, έχουμε το ακόλουθο σχήμα (στο οποίο τα κέντρα των κύκλων είναι

τα στοιχεία της κύριας διαγωνίου του μητρώου A και οι ακτίνες των κύκλων είναι τα υπόλοιπα στοιχεία των στηλών (ή των γραμμών). Τα σημεία πάνω στον πραγματικό άξονα συμβολίζουν τις **ακριβείς** ιδιοτιμές που υπολογίστηκαν νωρίτερα από το χαρακτηριστικό πολυώνυμο.



Εικόνα 9: Παράδειγμα κύκλων Gershgorin

Υπάρχουν $n = 2$ δίσκοι στο μιγαδικό επίπεδο, με κέντρα τα προαναφερόμενα σημεία και κάθε ιδιοτιμή πρέπει να κείται εντός ενός από αυτούς τους δίσκους. Επειδή απομένει ζυγός αριθμός κύκλων (μετά την αφαίρεση των απομονωμένων κύκλων, που εδώ δεν υπάρχουν), αυτό σημαίνει ότι θα έχουμε είτε ένα ζεύγος πραγματικών ιδιοτιμών είτε ένα ζεύγος συζυγών μιγαδικών ιδιοτιμών. Επειδή οι κύκλοι **δεν** είναι ομόκεντροι, θα έχουμε **ζεύγος πραγματικών ιδιοτιμών**.

B) Σχηματίζουμε τους κύκλους Gershgorin κατά γραμμές ή στήλες:

$$|z-1| \leq 2 \Rightarrow -2 \leq z-1 \leq 2 \Rightarrow -1 \leq z \leq 3$$

$$|z+1| \leq 1 \Rightarrow -1 \leq z+1 \leq 1 \Rightarrow -2 \leq z \leq 0$$

Άρα το διάστημα συναλήθευσης είναι το $[-1, 0]$

$$|z-1| \leq 1 \Rightarrow -1 \leq z-1 \leq 1 \Rightarrow 0 \leq z \leq 2$$

$$|z+1| \leq 2 \Rightarrow -2 \leq z+1 \leq 2 \Rightarrow -3 \leq z \leq 1$$

Άρα το διάστημα συναλήθευσης είναι το $[0, 1]$

3.3 Θεώρημα Perron – Frobenius

Έστω το μητρώο $A = \begin{bmatrix} 0,4935 & 0,0786 & 0,9612 & 0,24 \\ 0,7036 & 0,03994 & 0,3397 & 0,13 \\ 0,2564 & 0,0765 & 0,2234 & 0,46 \\ 0,0294 & 0,6185 & 0,1943 & 0,77 \end{bmatrix}$ που είναι ένα μητρώο **Θετικό** (δηλαδή **όλα** τα στοιχεία είναι θετικά). Σύμφωνα με το **Θεώρημα Perron – Frobenius** έχουμε μια τουλάχιστον **πραγματική ιδιοτιμή**. Πιο συγκεκριμένα, **αν το μητρώο $A > 0$** και $\rho = \{\lambda\}$, όπου $A * x = \lambda * x$ είναι η τιμή της **φασματικής ακτίνας του**, τότε ισχύουν τα εξής:

- Επειδή η φασματική ακτίνα $\rho > 0$ και επίσης επειδή το $\rho \in \lambda(A)$, υπάρχει ιδιοτιμή του μητρώου ίση με τη φασματική ακτίνα (ή με άλλα λόγια η φασματική ακτίνα ρ ανήκει στο φάσμα $\lambda(A)$ του μητρώου A), επομένως είναι θετική. **Δηλαδή, όταν ένα μητρώο είναι θετικό, τότε και η φασματική ακτίνα του είναι θετική.**
- Η αλγεβρική πολλαπλότητα της φασματικής ακτίνας ρ είναι «1», δηλ. η φασματική ακτίνα δεν επαναλαμβάνεται περισσότερες από μια φορές, επομένως μπορεί να εφαρμοστεί η μέθοδος των δυνάμεων.** Το τελευταίο μπορεί να διατυπωθεί και ως εξής: δεν υπάρχει άλλη ιδιοτιμή με μέτρο $|\rho|$ πέραν της ρ . **Δηλαδή ένα θετικό μητρώο έχει κυρίαρχη ιδιοτιμή** (αφού η αλγεβρική πολλαπλότητα της φασματικής ακτίνας ρ είναι 1) και επομένως **μπορεί να εφαρμοστεί σε αυτό η μέθοδος των δυνάμεων.**
- Για την ιδιοτιμή λ (που είναι ταυτόχρονα και φασματική ακτίνα), υπάρχει μοναδικό θετικό ιδιοδιάνυσμα $\rho > 0$ που είναι κανονικοποιημένο, τέτοιο ώστε $\|\rho\| = 1$. Τότε $A * \rho = \lambda * \rho$ και το ρ ονομάζεται **διάνυσμα Perron** και το λ **ρίζα Perron του A**.
- Υπάρχει ιδιοδιάνυσμα x τέτοιο ώστε $A * x = \rho * x$.

3.4 Ιδιοτιμές αντίστροφου μητρώου

Αν A αντιστρέψιμο μητρώο με ιδιοζεύγος (λ, x) , όπου λ = ιδιοτιμή και x = ιδιοδιάνυσμα που αντιστοιχεί στην ιδιοτιμή, τότε να δειχθεί ότι το αντίστροφο μητρώο A^{-1} έχει την αντίστροφη ιδιοτιμή, δηλ. το $1/\lambda$.

Λύση

Επειδή ισχύει η βασική σχέση $A * x = \lambda * x$ μεταξύ ιδιοτιμών και ιδιοδιανυσμάτων ενός μητρώου A , έπειτα ότι αν πολλαπλασιάσουμε και τα δύο μέλη της προηγούμενης εξίσωσης από αριστερά με το A^{-1} θα έχουμε ότι: $A^{-1} * A * x = A^{-1} * \lambda * x \Rightarrow I * x = \lambda * A^{-1} * x \Rightarrow x * \frac{1}{\lambda} = A^{-1} * x$. Επομένως, αφού $A * x = \lambda * x$ και $A^{-1} * x = \frac{1}{\lambda} * x$, συμπεραίνουμε ότι το **$1/\lambda$ αποτελεί ιδιοτιμή του μητρώου A^{-1}** , δηλ. το αντίστροφο μητρώο περιέχει τις αντίστροφες ιδιοτιμές του μητρώου A . Η ανωτέρω πρόταση ισχύει μόνο όταν το μητρώο A είναι αντιστρέψιμο.

3.5 Σύγκριση συναρτήσεων eig και $eigs$

Η συνάρτηση $E = eig(A)$: παράγει ένα διάνυσμα στήλη E που περιέχει τις **ιδιοτιμές ενός τετραγωνικού μητρώου A** . Η συνάρτηση αυτή έχει διάφορες **παραλλαγές** που φαίνονται στη συνέχεια:

$[V, D] = eig(A)$: παράγει ένα **διαγώνιο μητρώο D με τις ιδιοτιμές** και ένα πυκνό μητρώο V του οποίου οι στήλες είναι τα αντίστοιχα ιδιοδιανύσματα έτσι ώστε $A * V = V * D$

Η συνάρτηση $eigs$ βρίσκει **μερικές** ιδιοτιμές και ιδιοδιανύσματα ενός μητρώου. Πιο συγκεκριμένα:

$D = eigs(A)$: επιστρέφει ένα διάνυσμα των **έξι μεγαλύτερων ιδιοτιμών του μητρώου A** , το οποίο πρέπει να είναι **μεγάλο και αραιό**.

Κεφάλαιο 4 – Παρεμβολή και πολυώνυμα παρεμβολής

Δοθέντων των τιμών (x, y) , ζητάμε τον υπολογισμό του τελεστή T : $T * x = y$, όταν χρησιμοποιούμε πεπερασμένο πλήθος μετρήσεων $x_0, x_1, x_2, \dots, x_n \in [a, b]$. Με την **παρεμβολή** προσπαθούμε να κατασκευάσουμε συνάρτηση με πεδίο ορισμού κάποιο συνεχές διάστημα π.χ. το διάστημα $[x_0, x_1]$. Στην παρεμβολή ζητάμε κατασκευή συνάρτησης που να λαμβάνει ακριβώς τις τιμές (x, y) σε κάποιο διάστημα. Αντίθετα στην προσέγγιση μας ενδιαφέρει να ελαχιστοποιήσουμε κάποια απόσταση μεταξύ του πίνακα τιμών και των τιμών της συνάρτησης.

4.1 Εισαγωγή

Υποθέτουμε ότι η θερμοκρασία ενός ασθενούς μετρήθηκε σε βαθμούς Κελσίου σε χρονικά διαστήματα της μιας ώρας και ότι έχουμε στην διάθεσή μας τα ακόλουθα δεδομένα:

x	3:00	4:00	5:00	6:00	7:00	8:00	9:00	10:00	11:00
y=f(x)	38,5	38,7	40,4	39,2	38,3	37,6	38,5	39,2	38,9

Έστω τώρα ότι μας ζητείται να προσδιορίσουμε σύμφωνα με τα δεδομένα του ασθενή μια ενδιάμεση τιμή της θερμοκρασίας του ασθενή στις 3:15 ή στις 9:30. Ο προσδιορισμός αυτός επιτυγχάνεται με την παρεμβολή. Έτσι γενικά δίνουμε τον παρακάτω ορισμό: Αν έχουμε στην διάθεσή μας τιμές μιας συνάρτησης $f(x)$, που αντιστοιχούν σε διάφορες τιμές $x_0, x_1, x_2, \dots, x_n$ της ανεξάρτητης μεταβλητής x , τότε παρεμβολή είναι η διαδικασία με την οποία αποκτούμε, από τις δοθείσες συναρτησιακές τιμές, προσεγγίσεις της $f(x)$ για τις τιμές της ανεξάρτητης μεταβλητής x που κείνται ενδιάμεσα στις δεδομένες τιμές $x_0, x_1, x_2, \dots, x_n$.

Στη συνέχεια υποθέτουμε ότι μας ζητείται να προσδιορίσουμε σύμφωνα με τα δεδομένα του ασθενή μια τιμή της θερμοκρασίας εκτός του διαστήματος στο οποίο βρίσκονται οι μετρήσεις. Για παράδειγμα, να προσδιορίσουμε την θερμοκρασία του ασθενή στις 11:30 ή στις 2:45. Ο προσδιορισμός αυτός επιτυγχάνεται με την παρεκβολή. Έτσι για την γενική περίπτωση δίνουμε τον παρακάτω ορισμό:

Αν έχουμε στην διάθεσή μας τιμές μιας συνάρτησης $f(x)$, που αντιστοιχούν σε διάφορες τιμές $x_0, x_1, x_2, \dots, x_n$ της ανεξάρτητης μεταβλητής x , τότε παρεκβολή είναι η διαδικασία με την οποία αποκτούμε, από τις δοθείσες συναρτησιακές τιμές, προσεγγίσεις της $f(x)$ για τις τιμές της ανεξάρτητης μεταβλητής x που βρίσκονται εκτός του διαστήματος των δεδομένων τιμών $x_0, x_1, x_2, \dots, x_n$. Ο προσδιορισμός του χρόνου που η θερμοκρασία του ασθενή είχε την κρίσιμη τιμή 40 βαθμών Κελσίου επιτυγχάνεται με την αντίστροφη παρεμβολή. Αν έχουμε στην διάθεσή μας τιμές μιας συνάρτησης $f(x)$, που αντιστοιχούν σε διάφορες τιμές $x_0, x_1, x_2, \dots, x_n$ της ανεξάρτητης μεταβλητής x , τότε αντίστροφη παρεμβολή είναι η διαδικασία με την οποία αποκτούμε, από τις δοθείσες συναρτησιακές τιμές, προσεγγίσεις της ανεξάρτητης μεταβλητής x για τις οποίες η $f(x)$ παίρνει μια συγκεκριμένη τιμή. Η παρεμβολή και η παρεκβολή παίζουν σημαντικό ρόλο στις εφαρμογές και αποτελούν ένα από τα

βασικότερα θέματα της αριθμητικής ανάλυσης. Γενικά, ένας τρόπος για πραγματοποιηθεί η παρεμβολή και η παρεκβολή βασίζεται στην κατασκευή ενός πολυωνύμου που διέρχεται από όλα τα σημεία $(x, f(x))$ που προσδιορίζονται από τις τιμές που έχουμε στην διάθεσή μας αναφορικά με την ανεξάρτητη μεταβλητή x (ισαπέχουσες ή μη ισαπέχουσες) και τις τιμές της συνάρτησης $f(x)$.

Σε αυτό το κεφάλαιο θα δούμε ότι υπάρχει ένα τέτοιο πολυώνυμο και ότι είναι μοναδικό. Όταν έχουμε στην διάθεσή μας το πολυώνυμο αυτό που περνά από όλα τα σημεία της ταύτισης $(x, f(x))$, τότε εύκολα μπορούμε να κάνουμε **παρεμβολή** και **παρεκβολή** υπολογίζοντας την τιμή του πολυωνύμου για οποιαδήποτε τιμή της ανεξάρτητης μεταβλητής x . Με τη διαδικασία αυτή αποκτούμε, από τις διοθείσες συναρτησιακές τιμές, προσεγγίσεις της $f(x)$ για τις τιμές της ανεξάρτητης μεταβλητής x που κείνται ενδιάμεσα στις δεδομένες τιμές, ή βρίσκονται εκτός του διαστήματος των δεδομένων τιμών. Για το λόγο αυτό, χαρακτηριστικά έχει επικρατήσει η άποψη από πολλούς, ότι «παρεμβολή είναι η τεχνική του να μπορεί κανείς να διαβάζει μεταξύ των γραμμών ενός πίνακα».

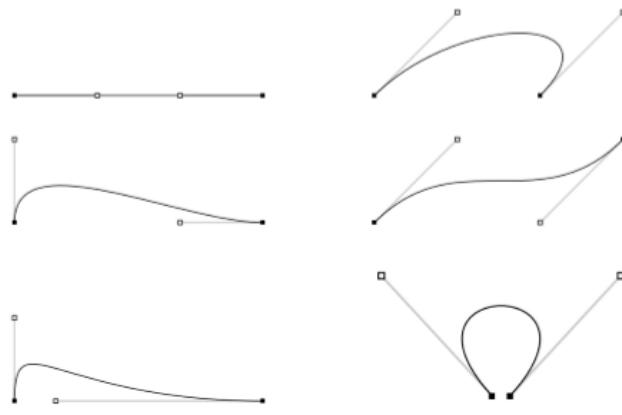
4.2 Θεώρημα Weierstrass - πολυώνυμα Bernstein και καμπύλες Bezier

Το **Θεώρημα Weierstrass** αναφέρει ότι αν μια συνάρτηση f είναι συνεχής σε ένα διάστημα $[a, b]$, δηλαδή $f \in C[a, b]$, $\forall \epsilon > 0$, υπάρχει πολυώνυμο P_n , όπου το n είναι ο βαθμός του πολυωνύμου και έχει βαθμό προσέγγισης ϵ , δηλ. $n = n(\epsilon)$, τέτοιο ώστε $\max_{x \in [a,b]} |f(x) - P_n(x)| \leq \epsilon$. Το θεώρημα λέει ότι οποιαδήποτε συνάρτηση που είναι συνεχής σε κλειστό διάστημα μπορεί να προσεγγιστεί όσο καλά θέλουμε από πολυώνυμο. **Όπως είναι αναμενόμενο, ο βαθμός του πολυωνύμου εξαρτάται από τον επιζητούμενο βαθμό προσέγγισης ϵ** . Χρειάζονται οι τιμές της συνάρτησης στα σημεία $\frac{k}{n}$ (σημεία δειγματοληψίας). Το πολυώνυμο Bernstein (B) ορίζεται σε διάφορα σημεία $0, 1, 2, \dots, \frac{k}{n}$ ως εξής: $B_n(0; f) = f(0), B_n(1; f) = f(1), B_n(2; f) = f(2)$ κ.ο.κ., αλλά στα υπόλοιπα σημεία έχουμε ότι το πολυώνυμο $B_n\left(\frac{k}{n}; f\right) \neq f(k, n)$ και αντικαθιστά τη συνάρτηση f στα σημεία αυτά. **Τα πολυώνυμα Bernstein δεν επιτελούν παρεμβολή**. Με τον όρο **πολυώνυμα Bernstein** ονομάζονται τα πολυώνυμα της μορφής: $b_{n,k}(x) = \binom{n}{k} * x^k * (1 - x)^{n-k}$, όπου οι δείκτες n, k συμβολίζουν το βαθμό του πολυωνύμου και το σημείο παρεμβολής k . Ένα πολυώνυμο Bernstein έχει τις εξής ιδιότητες:

- Ο **βαθμός** του πολυωνύμου είναι: $\deg(b_{n,k}) = n$.
- Το $b_{n,k}(x) \geq 0$ για $0 \leq x \leq 1$.
- Τα πολυώνυμα $\{b_{n,0}(x), b_{n,1}(x), \dots, b_{n,n}(x)\}$ αποτελούν βάση του P_n , όπου P_n είναι το σύνολο όλων των πολυωνύμων μίας μεταβλητής t , βαθμού ως και n και αποτελούν διανυσματικό χώρο. Η διάσταση του διανυσματικού χώρου είναι $n+1$. Μάλιστα μία βάση για το P_n είναι τα μονώνυμα $\{1, t, t^2, \dots, t^n\}$.

Ένα πολυώνυμο σε μορφή Bernstein γράφεται ως γραμμικός συνδυασμός $p(x) = \sum_{k=0}^n \gamma_k * b_{n,k}(x)$ των πολυώνυμων βάσης με συντελεστές $\gamma_0, \dots, \gamma_n$, δηλ. είναι ένας γραμμικός συνδυασμός που αποτελείται από επιμέρους πολυώνυμα Bernstein, το πλήθος των οποίων εξαρτάται από το πλήθος των σημείων παρεμβολής, που είναι από 0 έως n.

Αναφορικά με τις καμπύλες Bezier, (είναι παρόμοιας μορφής με τα πολυώνυμα σε μορφή Bernstein, με τη μόνη διαφορά ότι έχουν ως μεταβλητή το t αντί για το x) αναφέρουμε αρχικά ότι δοθέντων σημείων p_0, p_1, \dots, p_n , η καμπύλη Bezier ορίζεται από τη συνάρτηση: $r(t) = \sum_{k=0}^n p_k * b_{n,k}(t)$, όπου τα σημεία p_j ονομάζονται **σημεία ελέγχου** και το σχήμα $p_0 p_1 \dots p_n$ με κορυφές τα p_j είναι το **πολύγωνο ελέγχου**. Οι κυβικές καμπύλες Bezier ορίζονται από το αρχικό και τελικό σημείο P_0, P_3 και τα σημεία ελέγχου P_1, P_2 που δείχνουν κατεύθυνση, όπως φαίνεται στην επόμενη Εικόνα 11:



Εικόνα 10: Κυβικές καμπύλες Bezier

4.3 Παρεμβολή και πολυωνυμική παρεμβολή

Ένα από τα βασικότερα θέματα της Αριθμητικής Ανάλυσης είναι και το πρόβλημα της προσέγγισης της συνάρτησης $f(x)$ από μια άλλη πιο εύχρηστη συνάρτηση $g(x)$, και η οποία θα είναι κατά μια έννοια αρκετά «κοντά» στην $f(x)$ στ' ένα κλειστό διάστημα $[a, b]$. Στην περίπτωση αυτή ενδιαφερόμαστε για την εύρεση όσο το δυνατόν μικρότερων φραγμάτων ε της μεγαλύτερης δυνατής απόκλισης της $f(x)$ από τη συνάρτηση $g(x)$ που την προσεγγίζει στο διάστημα $[a, b]$, δηλαδή: $|f(x) - g(x)| \leq \varepsilon$. Στην **παρεμβολή** το ζητούμενο είναι η κατασκευή μιας απλής (εύχρηστης) συνάρτησης $g(x)$ που ονομάζεται **συνάρτηση παρεμβολής** και παίρνει τις ίδιες τιμές με τη συνάρτηση $f(x)$ στα σημεία x_i , $i = 0, \dots, n$, τα οποία ονομάζονται **σημεία παρεμβολής**.

Στο πρόβλημα της παρεμβολής το ζητούμενο είναι να βρούμε μια συνάρτηση $g(x)$ που να ανήκει στο σύνολο F_n έτσι ώστε για όλα τα σημεία παρεμβολής x_i , $i = 0, \dots, n$, να ισχύει ότι: $f(x_i) = g(x_i)$. Ορισμός: Όταν η συνάρτηση παρεμβολής $g(x)$ είναι ένα πολυώνυμο, τότε η αντίστοιχη διαδικασία ονομάζεται **πολυωνυμική παρεμβολή**. Στη συνέχεια θα εξετάσουμε την πολυωνυμική παρεμβολή. Στην περίπτωση αυτή η παραπάνω συνάρτηση παρεμβολής $g(x)$ είναι το ακόλουθο πολυώνυμο βαθμού n: $\Pi_n(x) = P_n(x) = a_0 + a_1x + a_2x^2 + \dots +$

$a_n x^n$, όπου οι αντίστοιχες συναρτήσεις $g_i(x)$, $i = 0, \dots, n$ είναι οι συναρτήσεις $g_i(x) = x^i$, για $i = 0, \dots, n$. Όπως και στη γενική περίπτωση, έτσι και εδώ, **το ζητούμενο είναι να βρούμε τους συντελεστές a_i , $i = 0, \dots, n$ του παραπάνω πολυωνύμου έτσι ώστε για όλα τα σημεία παρεμβολής x_i , $i = 0, \dots, n$ να ισχύει ότι: $f(x_i) = P_n(x_i)$** . Είναι φανερό ότι για να ικανοποιούνται οι παραπάνω σχέσεις, πρέπει οι συντελεστές a_i , $i = 0, \dots, n$ να ικανοποιούν το ακόλουθο σύστημα των $(n+1)$ γραμμικών αλγεβρικών εξισώσεων:

$$\begin{aligned} a_0 + x_0 a_1 + x_0^2 a_2 + \cdots + x_0^n a_n &= P_n(x_0) = f(x_0) \\ a_0 + x_1 a_1 + x_1^2 a_2 + \cdots + x_1^n a_n &= P_n(x_1) = f(x_1) \\ &\vdots \\ a_0 + x_n a_1 + x_n^2 a_2 + \cdots + x_n^n a_n &= P_n(x_n) = f(x_n) \end{aligned}$$

Το σύστημα αυτό έχει μια **μοναδική λύση**, και αυτό επειδή η ορίζουσα των συντελεστών των αγνώστων a_i , $i = 0, \dots, n$ του παραπάνω συστήματος είναι η γνωστή ορίζουσα του **Vandermonde**, η οποία δίνεται ως εξής:

$$\det V = \begin{vmatrix} 1 & x_0 & x_0^2 & \cdots & x_0^n \\ 1 & x_1 & x_1^2 & \cdots & x_1^n \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_n & x_n^2 & \cdots & x_n^n \end{vmatrix}$$

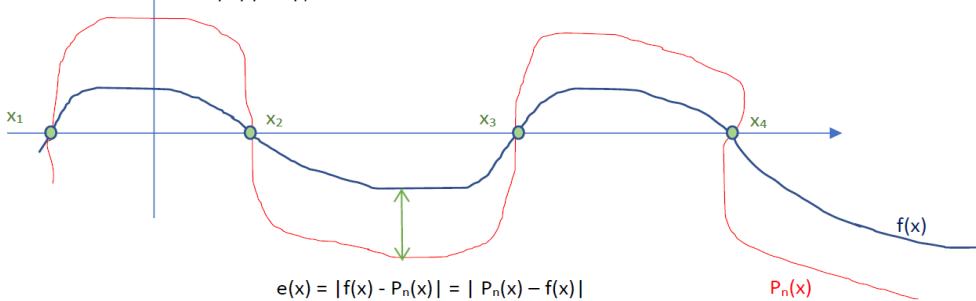
Η τιμή της παραπάνω ορίζουσας ισούται με: $\det V = \prod_{\substack{i,j=0 \\ i>j}}^{n-1} (x_i - x_j) \neq 0$ και είναι διάφορη του μηδενός επειδή

τα x_i , $i = 0, \dots, n$ είναι διακεκριμένα, δηλαδή $x_i \neq x_j$ για $i \neq j$. Έτσι για να προσδιορίσουμε το πολυώνυμο $P_n(x)$ πρέπει να υπολογίσουμε τους συντελεστές a_i , οι οποίοι μπορούν να αποκτηθούν με την επίλυση του παραπάνω γραμμικού συστήματος. Η επίλυση του συστήματος αυτού μπορεί να γίνει με τη χρήση μια αριθμητικής μεθόδου όπως, για παράδειγμα, της απαλοιφής του Gauss. **Το μητρώο Vandermonde έχει πολύ κακό δείκτη κατάστασης όσον αφορά τον πολλαπλασιασμό μητρώων.**

Η παρεμβολή (προσέγγιση) μιας συνάρτησης $f(x)$ από ένα πολυώνυμο $n -$ βαθμού $P_n(x)$ που την παρεμβάλει, φαίνεται στη συνέχεια. Ο λόγος που κάνουμε αυτήν την παρεμβολή (προσέγγιση) της συνάρτησης $f(x)$ από το πολυώνυμο $P_n(x)$ είναι γιατί το πολυώνυμο είναι ένα πολύ καλύτερο μαθηματικό εργαλείο από τη συνάρτηση. Για παράδειγμα μια συνάρτηση μπορεί να έχει ασυνέχεια, δηλαδή να μην μπορεί να οριστεί σε ένα διάστημα $[a, b]$. Επίσης, μπορεί να μην υπολογίζεται η παράγωγός της. Αντίθετα, ένα πολυώνυμο ορίζεται πάντα.

Σημεία παρεμβολής $x_i \in [a, b]$, $i = 1, 2, 3, 4$. Σε αυτά θα πρέπει το $P_n(x) = f(x)$.

Ο βαθμός n του πολυωνύμου είναι πάντα κατά «1» μικρότερος από το πλήθος των σημείων παρεμβολής, δηλ. για $n+1$ σημεία παρεμβολής, το $P_n(x) = a_n * x^n + a_{n-1} * x^{n-1} + \dots + a_1 * x^1 + a_0$. Επομένως, το πλήθος των συντελεστών a_n, a_{n-1}, \dots, a_0 είναι ίσο με το πλήθος των σημείων παρεμβολής.



Εικόνα 11: Πολυωνυμική παρεμβολή συνάρτησης

Θεώρημα πολυωνυμικής παρεμβολής και σφάλμα πολυωνυμικής παρεμβολής: Για κάθε ζεύγος διακριτών τιμών $\{x_i, y_i\}$, $i = 0, \dots, n$ με διακριτούς κόμβους x_i , υπάρχει μοναδικό πολυώνυμο βαθμού έως και n , έστω P_n , τέτοιο ώστε $P_n(x_i) = y_i$ για $i = 0, \dots, n$. Το P_n ονομάζεται πολυώνυμο παρεμβολής και λέγεται ότι παρεμβάλει τη συνάρτηση στους κόμβους x_i . Ο βαθμός του πολυωνύμου P_n είναι το πολύ n . Ισχύει επίσης ότι πλήθος κόμβων (σημείων) παρεμβολής = πλήθος συντελεστών πολυωνύμου > βαθμός πολυωνύμου. Τέλος το πολυωνυμο παρεμβολής είναι μοναδικό, αλλά υπάρχουν διάφοροι τρόποι για να υπολογιστεί. Όσον αφορά το σφάλμα της πολυωνυμικής παρεμβολής έχουμε ότι: Αν $\Omega = [a, b] = [x_0, x_n]$ και η συνάρτηση $f \in C^{(n+1)}(\Omega)$, δηλ. είναι $n+1$ φορές παραγωγίσιμη στο Ω και $x_i \in \Omega$, $i = 0, \dots, n$ είναι $n+1$ διακριτοί κόμβοι παρεμβολής, τότε $\forall x \in \Omega$, $\exists \xi \in \Omega$ τέτοιο ώστε:

$$e_n f(x) = f(x) - P_n(x) = L(x) * \frac{f^{(n+1)}(\xi)}{(n+1)!} = \frac{f^{(n+1)}(\xi)}{(n+1)!} * \prod_{i=0}^n (x - x_i)$$

και το μέγιστο άνω φράγμα του σφάλματος πολυωνυμικής παρεμβολής είναι (ο τύπος αυτός χρησιμοποιείται επίσης όταν δίνονται τα άκρα του διαστήματος στο οποίο υπολογίζεται η πολυωνυμική παρεμβολή):

$$\max |e_n f(x)| \leq \frac{\max_{x \in \Omega} |f^{(n+1)}(x)|}{4(n+1)} \left(\frac{x_n - x_0}{n} \right)^{(n+1)} \Rightarrow \max |e_n f(x)| \leq \frac{\max_{x \in \Omega} |f^{(n+1)}(x)|}{4(n+1)} h^{(n+1)} \text{ όπου } h = \frac{x_n - x_0}{n}.$$

4.4 Παρεμβολή του Lagrange

Το πολυώνυμο παρεμβολής μπορεί να γραφτεί ως εξής: $P_1(x) = L_0(x) * f(x_0) + L_1(x) * f(x_1)$ (βαθμού 1) και

το πολυώνυμο παρεμβολής Lagrange $P_n(x) = L_0(x)f(x_0) + L_1(x)f(x_1) + \dots + L_n(x)f(x_n) = \sum_{i=0}^n \frac{\prod_{j=0}^n (x - x_j)}{\prod_{j=0, j \neq i}^n (x_i - x_j)} * f_i$ (βαθμού n), όπου τα πολυώνυμα $L_i(x)$, $i = 0, \dots, n$ έχουν την παρακάτω ιδιότητα: $L_i(x_j) = \begin{cases} 1 & \text{αν } i = j, \\ 0 & \text{αν } i \neq j, \end{cases}$ για $i, j = 0, \dots, n$. Είναι προφανές ότι αν τα πολυώνυμα $L_i(x)$, $i = 0, \dots, n$ έχουν την παραπάνω ιδιότητα, τότε το πολυωνυμο $P_n(x)$ ικανοποιεί όλες τις συνθήκες ταύτισης. Τα πολυώνυμα Lagrange $L_i(x)$ υπολογίζονται σύμφωνα με τον παρακάτω τύπο (στον αριθμητή αφαιρούμε από το x όλα τα υπόλοιπα σημεία (κόμβους) παρεμβολής εκτός από το τρέχον, που είναι το x_i , ενώ στον παρανομαστή αφαιρούμε από το x_i όλα τα υπόλοιπα σημεία παρεμβολής):

$$L_i(x) = \frac{(x - x_0)(x - x_1)(x - x_2) \dots (x - x_{i-1})(x - x_{i+1}) \dots (x - x_n)}{(x_i - x_0)(x_i - x_1)(x_i - x_2) \dots (x_i - x_{i-1})(x_i - x_{i+1}) \dots (x_i - x_n)}, \quad i = 0, \dots, n \quad \text{ή ισοδύναμα: } L_i(x) = \frac{\prod_{j=0}^n (x - x_j)}{\prod_{j=0, j \neq i}^n (x_i - x_j)}, \quad i = 0, \dots, n$$

Τα παραπάνω πολυώνυμα ονομάζονται συντελεστές Lagrange ή θεμελιώδη πολυώνυμα σημειακής παρεμβολής. Είναι αξιοσημείωτη η ομοιότητα ανάμεσα σε ένα πολυώνυμο παρεμβολής σε μορφή Lagrange και ένα πολυώνυμο παρεμβολής σε μορφή Bernstein, η οποία διατυπώνεται ως εξής:

$$P_n(x) = \sum_{i=0}^n L_i(x) * f_i \text{ (Lagrange)} \leftrightarrow p_n(x) = \sum_{k=0}^n \gamma_k * b_{n,k}(x) \text{ (Bernstein)}$$

Πρόταση: Για τους συντελεστές $L_i(x)$ του Lagrange, ισχύει η παρακάτω σχέση του Cauchy:

$$\sum_{i=0}^n L_i(x) = 1$$

Άσκηση 1: Αποδείξτε τη σχέση του Cauchy για τους συντελεστές $L_i(x)$ του Lagrange, χρησιμοποιώντας την ακόλουθη ιδιότητα των συντελεστών:

$$L_i(x_j) = \begin{cases} 1 & \text{αν } i = j, \\ 0 & \text{αν } i \neq j, \end{cases} \text{ για όλα τα } i, j = 0(1)n$$

Λύση

Σχηματίζουμε το παρακάτω πολυώνυμο: $Q_n(x) = \sum_{i=0}^n L_i(x) - 1$, το οποίο είναι και αυτό το πολύ η βαθμού. Η ιδιότητα των συντελεστών $L_i(x)$ μας λέει ότι για ένα δεδομένο $x=x_k$, $k=0(1)n$ ένας από τους συντελεστές αυτούς θα είναι ένα, ενώ όλοι οι υπόλοιποι θα είναι μηδέν. Δηλαδή, για ένα δεδομένο k , $k=0,..,n$ θα ισχύει ότι: $\sum_{i=0}^n L_i(x_k) = 1$. Επομένως, το σχηματιζόμενο πολυώνυμο $Q_n(x)$ θα μηδενίζεται ταυτοτικά σε όλα τα $(n+1)$ σημεία παρεμβολής x_i , $i=0,...,n$ και, κατά συνέπεια, θα ισχύει ότι: $\sum_{i=0}^n L_i(x) = 1$.

Άσκηση 2: Χρησιμοποιώντας τον τύπο παρεμβολής του Lagrange, να βρεθούν τρία πολυώνυμα το πολύ δευτέρου βαθμού, που να ταυτίζονται στα σημεία παρεμβολής $x = -2, 0, 1$, αντίστοιχα, με τις συναρτήσεις (α) $f(x) = x^3$, (β) $f(x) = x^4$ και (γ) $f(x) = x^5$.

Λύση

Αρχικά ονομάζουμε τα **σημεία παρεμβολής ως εξής**: $x_0 = -2, x_1 = 0, x_2 = 1$. Τότε χρησιμοποιώντας τον τύπο παρεμβολής του Lagrange βρίσκουμε: $P_2(x) = \frac{(x-x_1)(x-x_2)}{(x_0-x_1)(x_0-x_2)} f_0 + \frac{(x-x_0)(x-x_2)}{(x_1-x_0)(x_1-x_2)} f_1 + \frac{(x-x_0)(x-x_1)}{(x_2-x_0)(x_2-x_1)} f_2$. Αντικαθιστώντας τις τιμές των x_i , $i = 0, 1, 2$ στον παραπάνω τύπο έχουμε διαδοχικά: $P_2(x) = \frac{(x-0)*(x-1)}{(-2)*(-3)} f_0 + \frac{(x+2)*(x-1)}{2*(-1)} f_1 + \frac{(x+2)x}{3*1} f_2 = \frac{x(x-1)}{6} f_0 + \frac{(x+2)(x-1)}{-2} f_1 + \frac{(x+2)x}{3} f_2$. Έτσι το πολυώνυμο παρεμβολής έχει τη μορφή: $P_2(x) = \frac{x^2-x}{6} f_0 + \frac{x^2+x-2}{-2} f_1 + \frac{x^2+2x}{3} f_2$. Το πολυώνυμο αυτό θα το χρησιμοποιήσουμε και για τις τρεις περιπτώσεις. Έτσι για να βρούμε τα ζητούμενα πολυώνυμα, θα βρούμε για κάθε περίπτωση συνάρτησης τις τιμές f_0, f_1 και f_2 :

(α) Στην περίπτωση αυτή έχουμε $f_0 = (-2)^3 = -8, f_1 = 0^3 = 0$ και $f_2 = 1^3 = 1$, οπότε αντικαθιστώντας τις τιμές αυτές στο πολυώνυμο έχουμε: $P_2(x) = -\frac{x^2-x}{6} * 8 - \frac{x^2+x-2}{2} * 0 + \frac{x^2+2x}{3} * 1 = -\frac{4x^2-4x}{3} + \frac{x^2+2x}{3} = -x^2 + 2x$. Άρα το ζητούμενο πολυώνυμο βρέθηκε, είναι δευτέρου βαθμού και **επαληθεύει** τις συνθήκες ταύτισης, διότι: $P_2(-2) = f_0 = -8, P_2(0) = f_1 = 0, P_2(1) = f_2 = 1$.

(β) Εδώ έχουμε $f_0 = (-2)^4 = 16, f_1 = 0^4 = 0$ και $f_2 = 1^4 = 1$, οπότε αντικαθιστώντας τις τιμές αυτές στο πολυώνυμο παίρνουμε: $P_2(x) = \frac{x^2-x}{6} * 16 - \frac{x^2+x-2}{2} * 0 + \frac{x^2+2x}{3} * 1 = -\frac{8x^2-8x}{3} + \frac{x^2+2x}{3} = 3x^2 - 2x$. Έτσι, το ζητούμενο πολυώνυμο βρέθηκε, είναι δευτέρου βαθμού και **επαληθεύει** τις συνθήκες ταύτισης, διότι $P_2(-2) = f_0 = 16, P_2(0) = f_1 = 0, P_2(1) = f_2 = 1$.

(γ) Στην περίπτωση αυτή έχουμε $f_0 = (-2)^5 = -32$, $f_1 = 0^5 = 0$ και $f_2 = 1^5 = 1$, οπότε αντικαθιστώντας τις τιμές αυτές στο πολυώνυμο έχουμε: $P_2(x) = -\frac{x^2-x}{6} * 32 - \frac{x^2+x-2}{2} * 0 + \frac{x^2+2x}{3} * 1 = -\frac{16x^2-16x}{3} + \frac{x^2+2x}{3} = -5x^2 + 6x$. Το παραπάνω πολυώνυμο είναι το ζητούμενο, γιατί είναι βαθμού δύο και επαληθεύει τις συνθήκες ταύτισης, αφού: $P_2(-2) = f_0 = -32$, $P_2(0) = f_1 = 0$, $P_2(1) = f_2 = 1$.

Άσκηση 3: Στον πίνακα τιμών που ακολουθεί, δίνονται τα σημεία παρεμβολής x_i , $i = 0, \dots, 3$ και οι αντίστοιχες τιμές f_i , $i = 0, \dots, 3$ της συνάρτησης $f(x)$:

x_i	-2	0	1	3
f_i	3	3	3	1

Για τις παραπάνω τιμές, να βρείτε τα πολυώνυμα $L_i(x)$, για $i = 0, \dots, 3$, δηλαδή τους **συντελεστές του Lagrange**. Στη συνέχεια με βάση τους συντελεστές που βρήκατε, επαληθεύστε τη σχέση του Cauchy. Επίσης χρησιμοποιώντας αυτούς τους συντελεστές, να βρείτε το πολυώνυμο παρεμβολής του Lagrange βαθμού το πολύ τρία. Επαληθεύστε ότι το πολυώνυμο που βρήκατε είναι το ζητούμενο. Τέλος, αφού βρείτε το ζητούμενο πολυώνυμο, να προσεγγίσετε την τιμή της συνάρτησης $f(1.5)$.

Λύση

Αρχικά βρίσκουμε τους **συντελεστές του Lagrange** $L_i(x)$, $i = 0, \dots, 3$ χρησιμοποιώντας τις σχέσεις. Σύμφωνα με τις σχέσεις αυτές έχουμε: $L_0 = \frac{(x-x_1)(x-x_2)(x-x_3)}{(x_0-x_1)(x_0-x_2)(x_0-x_3)}$, $L_1 = \frac{(x-x_0)(x-x_2)(x-x_3)}{(x_1-x_0)(x_1-x_2)(x_1-x_3)}$, $L_2 = \frac{(x-x_0)(x-x_1)(x-x_3)}{(x_2-x_0)(x_2-x_1)(x_2-x_3)}$, $L_3 = \frac{(x-x_0)(x-x_1)(x-x_2)}{(x_3-x_0)(x_3-x_1)(x_3-x_2)}$. Αντικαθιστώντας στους παραπάνω συντελεστές, τις διοθείσες τιμές των σημείων παρεμβολής x_i , $i = 0, \dots, 3$ παίρνουμε: $L_0 = \frac{(x-0)(x-1)(x-3)}{(-2-0)(-2-1)(-2-3)}$, $L_1 = \frac{(x+2)(x-1)(x-3)}{(0+2)(0-1)(0-3)}$, $L_2 = \frac{(x+2)(x-0)(x-3)}{(1+2)(1-0)(1-3)}$, $L_3 = \frac{(x+2)(x-0)(x-1)}{(3+2)(3-0)(3-1)}$.

Από τους παραπάνω συντελεστές μετά από τις αντίστοιχες αλγεβρικές πράξεις έχουμε:

$$L_0 = -\frac{1}{30}(x^3 - 4x^2 + 3x), L_1 = \frac{1}{6}(x^3 - 2x^2 - 5x + 6), L_2 = -\frac{1}{6}(x^3 - x^2 - 6x), L_3 = \frac{1}{30}(x^3 + 2x^2 - 2x). \text{ Στη συνέχεια επαληθεύουμε τη σχέση του Cauchy.}$$

Σύμφωνα με αυτή τη σχέση, πρέπει το άθροισμα των συντελεστών του Lagrange να κάνει "1". Έτσι, προσθέτοντας τους συντελεστές αυτούς έχουμε: $\sum_{i=0}^3 L_i(x) = \frac{1}{30}[(-x^3 + 4x^2 - 3x) + (5x^3 - 10x^2 - 25x + 30) + (-5x^3 + 5x^2 + 30x) + (x^3 + x^2 - 2x)] = \frac{1}{30} * 30 = 1$.

Οπότε η ζητούμενη σχέση επαληθεύτηκε. Στη συνέχεια, θα βρούμε το **ζητούμενο πολυώνυμο παρεμβολής του Lagrange**. Το πολυώνυμο αυτό θα

έχει το πολύ βαθμό τρία και θα είναι της παρακάτω μορφής: $P_3(x) = L_0(x)f_0 + L_1(x)f_1 + L_2(x)f_2 + L_3(x)f_3$,

όπου τα πολυώνυμα $L_i(x)$, $i = 0, \dots, 3$ είναι οι συντελεστές του Lagrange. Επομένως, αντικαθιστώντας στην παραπάνω σχέση τους συντελεστές $L_i(x)$, $i = 0, \dots, 3$ που έχουμε βρει προηγουμένως και αντικαθιστώντας επίσης και τις διοθείσες τιμές της συνάρτησης f_i , για $i = 0, \dots, 3$ παίρνουμε διαδοχικά: $P_3(x) = \frac{1}{30}[(-x^3 + 4x^2 - 3x)f_0 + (5x^3 - 10x^2 - 25x + 30)f_1 + (-5x^3 + 5x^2 + 30x)f_2 + (x^3 + x^2 - 2x)f_3] = \frac{1}{30}[(-x^3 + 4x^2 - 3x)3 + (5x^3 -$

$$10x^2 - 25x + 30)3 + (-5x^3 + 5x^2 + 30x)3 + (x^3 + x^2 - 2x)1] = \frac{1}{30} [(-3x^3 + 12x^2 - 9x) + (15x^3 - 30x^2 - 75x + 90) + (-15x^3 + 15x^2 + 90x) + (x^3 + x^2 - 2x)] = \frac{1}{30} (-2x^3 - 2x^2 + 4x + 90).$$

Από την παραπάνω σχέση τελικά παίρνουμε το ακόλουθο πολυώνυμο: $P_3(x) = -\frac{1}{15}(x^3 + x^2 - 2x - 45)$. Το παραπάνω πολυώνυμο είναι το **ζητούμενο**, γιατί είναι βαθμού τρία και επαληθεύει τις **συνθήκες ταύτισης**, αφού $P_3(-2) = f_0 = 3, P_3(0) = f_1 = 3, P_3(1) = f_2 = 3, P_3(3) = f_3 = 1$. Τώρα, η τιμή $f(1.5)$ μπορεί να προσεγγιστεί από την τιμή $P_3(1.5)$ του παραπάνω πολυωνύμου. Έτσι μπορούμε να πάρουμε: $f(1.5) \approx P_3(1.5) = 2.825$.

Άσκηση 4 Στον πίνακα τιμών της άσκησης, δίνονται τα σημεία παρεμβολής $x_i, i = 0, \dots, 3$ και οι αντίστοιχες τιμές $f_i, i = 0, \dots, 3$ της συνάρτησης $f(x)$.

x_i	-2	0	1	3
f_i	3	3	3	1

Κάνοντας χρήση της σχέσης του Cauchy, να βρεθεί το πολυώνυμο παρεμβολής του Lagrange βαθμού το πολύ 3. Να συγκρίνετε το πολυώνυμο που βρήκατε με το αντίστοιχο πολυώνυμο της προηγούμενης άσκησης.

Λύση

Το ζητούμενο πολυώνυμο παρεμβολής του Lagrange βαθμού το πολύ τρία, θα έχει την παρακάτω μορφή:

$P_3(x) = L_0(x)f_0 + L_1(x)f_1 + L_2(x)f_2 + L_3(x)f_3$, όπου τα πολυώνυμα $L_i(x), i = 0, \dots, 3$, δηλαδή οι συντελεστές του Lagrange δίνονται από τις γνωστές σχέσεις. Με βάση τις δοθείσες τιμές της συνάρτησης f_i , για $i = 0, \dots, 3$ το παραπάνω πολυώνυμο μπορεί να γραφεί ως εξής: $P_3(x) = 3L_0(x) + 3L_1(x) + 3L_2(x) + 1L_3(x) = 3[L_0(x) + L_1(x) + L_2(x)] + L_3(x)$. Όμως από την σχέση Cauchy έχουμε ότι: $L_0(x) + L_1(x) + L_2(x) + L_3(x) = 1$. Συνδυάζοντας τις δύο παραπάνω σχέσεις, μπορούμε να πάρουμε: $P_3(x) = 3[1 - L_3(x)] + L_3(x) = 3 - 2L_3(x)$.

Όμως από τις γνωστές σχέσεις και από τις δοθείσες τιμές των σημείων παρεμβολής $x_i, i = 0, \dots, 3$ έχουμε:

$L_3(x) = \frac{(x-x_0)(x-x_1)(x-x_2)}{(x_3-x_0)(x_3-x_1)(x_3-x_2)} = \frac{(x+2)(x-0)(x-1)}{(3+2)(3-0)(3-1)} = \frac{1}{30}(x^3 + x^2 - 2x)$. Συνδυάζοντας τις δύο παραπάνω σχέσεις, μπορούμε να πάρουμε: $P_3(x) = 3 - \frac{1}{15}(x^3 + x^2 - 2x) = -\frac{1}{15}(x^3 + x^2 - 2x - 45)$. Συγκρίνοντας το παραπάνω πολυώνυμο $P_3(x)$ με το αντίστοιχο πολυώνυμο που βρήκαμε στην προηγούμενη άσκηση, παρατηρούμε ότι αυτά ταυτίζονται.

4.4 Εκτίμηση σφάλματος της πολυωνυμικής παρεμβολής

Με την εύρεση του πολυωνύμου $P_n(x)$ από τα σημεία παρεμβολής $x_i, i = 0, \dots, n$ και τις δοθείσες τιμές $f_i \equiv f(x_i)$, για $i = 0, \dots, n$ της συνάρτησης $f(x)$, τότε η τιμή της συνάρτησης $f(x)$ σε ένα σημείο x που δεν συμπίπτει με κανένα σημείο παρεμβολής $x_i, i = 0, \dots, n$, μπορεί να προσδιοριστεί από την τιμή του πολυωνύμου $P_n(x)$ στο σημείο αυτό. Δηλαδή: $f(x) \approx P_n(x)$, για $x \neq x_i, i = 0, \dots, n$. Η προσέγγιση αυτή δημιουργεί ένα σφάλμα που δίνεται από την παρακάτω σχέση: $\varepsilon(x) = P_n(x) - f(x)$. Για τη διόρθωση του πολυωνύμου παρεμβολής του Lagrange $P_n(x)$ που ορίζεται από τα σημεία παρεμβολής $x_i, i = 0, \dots, n$ και τις συναρτησιακές τιμές $f_i = f(x_i), i = 0, \dots, n$, η προσέγγιση αυτή δημιουργεί ένα σφάλμα που δίνεται από την παραπάνω σχέση: $\varepsilon(x) = P_n(x) - f(x)$.

0,...,n πρέπει να ισχύει ότι στο διάστημα $\Omega = [a, b]$ και για τη συνάρτηση $f \in C^{(n+1)}(\Omega)$ που είναι n+1 φορές συνεχώς παραγωγίσιμη στο Ω και τα $x_j \in \Omega$, j = 0, 1, 2,...n είναι **n + 1 διακριτοί κόμβοι παρεμβολής**, τότε $\forall x \in \Omega$ ισχύει:

Σφάλμα πολυωνυμικής παρεμβολής: $e_n f(x) = f(x) - P_n f(x) = L(x) \frac{f^{(n+1)}(\xi)}{(n+1)!} = \frac{1}{(n+1)!} f^{(n+1)}(\xi) * \prod_{i=0}^n (x - x_i)$, όπου x_i είναι

τα σημεία παρεμβολής και **μέγιστο άνω φράγμα του σφάλματος πολυωνυμικής παρεμβολής**:

$$\max |e_n f(x)| \leq \frac{\max_{x \in \Omega} |f^{(n+1)}(x)|}{4(n+1)} * \left(\frac{x_n - x_0}{n} \right)^{(n+1)} \Rightarrow \max |e_n f(x)| \leq \frac{\max_{x \in \Omega} |f^{(n+1)}(x)|}{4(n+1)} h^{(n+1)} \text{ όπου } h = \frac{x_n - x_0}{n}$$

Εναλλακτικά ο δεύτερος τύπος χρησιμοποιείται αν ζητείται το σφάλμα της πολυωνυμικής παρεμβολής σε **κάποιο συγκεκριμένο διάστημα**, διότι μέσα στο δεύτερο τύπο εμφανίζονται οι τιμές x_0 και x_n που είναι τα **άκρα του διαστήματος $\Omega = [a, b]$** .

4.4.1 Άσκηση με άνω φράγμα σφάλματος

Έστω ότι έχει βρεθεί για 5 ζεύγη σημείων (x_i, y_i) ένα πολυώνυμο παρεμβολής της μορφής: $p(x) = -0.083 * x^4 - 0.33 * x^3 + 1.0833 * x^2 - 0.667 * x + 1$, για το οποίο γνωρίζετε ότι υπάρχει η 5^η παράγωγος της γ σε όλα τα σημεία του κλειστού διαστήματος $J = [-1.0, 3.0]$ και έχει μέγιστη τιμή $M_1 > 0$. Να υπολογίσετε **ένα άνω φράγμα**, όσο το δυνατόν μικρότερο, για το απόλυτο σφάλμα $e(x) = |y(x) - p(x)|$ της προσέγγισης με το προηγούμενο πολυώνυμο παρεμβολής.

Λύση

Το μέγιστο άνω φράγμα **σφάλματος πολυωνυμικής παρεμβολής** μιας συνάρτησης από κάποιο πολυώνυμο δίνεται από τον τύπο: $\max |e_n f(x)| \leq \frac{\max_{x \in \Omega} |f^{(n+1)}(x)|}{4(n+1)} h^{(n+1)}$ όπου $h = \frac{x_n - x_0}{n} = \frac{3 - (-1)}{4} = \frac{4}{4} = 1$. Επομένως το $\frac{\max_{x \in \Omega} |f^{(n+1)}(x)|}{4(n+1)} * 1^5 = \frac{M_1}{4(4+1)} * 1^5 = \frac{M_1}{20}$.

4.5 Ασκήσεις πολυωνύμων παρεμβολής στη μορφή Lagrange

Άσκηση 1

- Να βρεθεί το πολυώνυμο παρεμβολής σε μορφή **Lagrange** βαθμού 1, που ορίζεται από τα σημεία παρεμβολής $x_0 = 0$ και $x_1 = 1$ και προσεγγίζει τη συνάρτηση: $f(x) = \sin \frac{\pi x}{2}$. Η $f'(x) = \frac{1}{2} * \cos \frac{\pi x}{2}$ και η $f''(x) = -\frac{1}{4} * \sin \frac{\pi x}{2}$.
- Στη συνέχεια να βρεθούν τα **απόλυτα σφάλματα** με ακρίβεια 7 δεκαδικών ψηφίων στα σημεία $x = 0.5$, $x = 0.1$ και $x = 0.01$ (προσοχή, τα σημεία αυτά **δεν** είναι σημεία παρεμβολής, διαφορετικά το σφάλμα θα ήταν ίσο με «0»).
- Να βρεθούν τα **άνω φράγματα των σφαλμάτων** στα προαναφερόμενα σημεία.
- Για $x = 0.5$, $x = 0.1$ και $x = 0.01$ να βρεθεί η ακριβής τιμή για το μέγιστο άνω φράγμα του σφάλματος.

Λύση

a) Το πολυώνυμο παρεμβολής 1^{ου} βαθμού στη μορφή Lagrange γράφεται στη μορφή: $P_1(x) = L_0(x) * f(x_0) + L_1(x) * f(x_1)$, όπου το $L_0(x) = \frac{x-x_1}{x_0-x_1} = \frac{x-1}{0-1} = 1-x$ και το $L_1(x) = \frac{x-x_0}{x_1-x_0} = \frac{x-0}{1-0} = x$. Άρα το πολυώνυμο παρεμβολής σε μορφή Lagrange είναι το $P_1(x) = (1-x) * f(x_0) + x * f(x_1) = (1-x) * \sin \frac{\pi * 0}{2} + x * \sin \frac{\pi}{2} = x$. Ως εισαγωγή στο επόμενο ερώτημα, αναφέρουμε ενδεικτικά ότι: το $\sin(0) = 0$ και το $\sin \frac{\pi}{2} = 1$. Το $\cos(0) = 1$ και $\cos(\frac{\pi}{2}) = 0$.

b) Με βάση αυτό το πολυώνυμο παρεμβολής, τα απόλυτα σφάλματα για τις προσεγγιστικές τιμές στα σημεία $x = 0.5$, $x = 0.1$ και $x = 0.01$ (που δεν είναι σημεία παρεμβολής, επομένως το σφάλμα που θα βρεθεί θα είναι διάφορο του μηδενός) είναι αντίστοιχα τα εξής: $|\varepsilon_1| = |p_1(0.5) - f(0.5)| = |p_1(0.5) - \sin \frac{\pi}{4}| = |0.5 - 0.707| = 0.207$ και $|\varepsilon_2| = |p_1(0.1) - \sin \frac{0.1\pi}{2}| = |0.1 - 0.15| = 0.05$ και $|\varepsilon_3| = |p_1(0.01) - \sin \frac{0.01\pi}{2}| = |0.01 - 0.015| = 0.005$.

c) Επειδή δεν μας ζητά η εκφώνηση μέγιστο άνω φράγμα σφάλματος και επειδή δεν μας δίνεται συγκεκριμένο διάστημα, δεν μπορούμε να χρησιμοποιήσουμε τον τύπο για το μέγιστο άνω φράγμα του σφάλματος της πολυωνυμικής παρεμβολής. Για το λόγο αυτό, για την εύρεση άνω φραγμάτων για τα σφάλματα αυτά θα χρησιμοποιήσουμε τον τύπο:

$$f(x) - P_n(x) = L(x) * \frac{f^{(n+1)}(\xi)}{(n+1)!} = (x - x_0) * (x - x_1) * \frac{f^{(n+1)}(\xi)}{(n+1)!} = (x - 0) * (x - 1) * \frac{f^{(n+1)}(\xi)}{(n+1)!} = x * (x - 1) * \frac{(-\frac{\pi^2}{4} \sin \frac{\pi \xi}{2})}{2}, \text{όπου το}$$

$$\text{πηλίκο } \frac{f^{(2)}(\xi)}{2!} = \frac{(-\frac{\pi^2}{4} \sin \frac{\pi \xi}{2})}{2}. \text{ Άρα το σφάλμα της πολυωνυμικής παρεμβολής } \varepsilon = P_n(x) - f(x) = -(f(x) - P_n(x)) =$$

$$-x * (x - 1) * \frac{(-\frac{\pi^2}{4} \sin \frac{\pi \xi}{2})}{2} = \frac{\pi^2}{8} * x * (x - 1) * \sin \frac{\pi \xi}{2} \Rightarrow |f(x) - P_n(x)| = \left| \frac{\pi^2}{8} * x * (x - 1) * \sin \frac{\pi \xi}{2} \right| \Rightarrow |f(x) - P_n(x)| = \left| \frac{\pi^2}{8} \right| * |x * (x - 1)| * |\sin \frac{\pi \xi}{2}| \Rightarrow |f(x) - P_n(x)| \leq \frac{\pi^2}{8} * |x^2 - x| = \frac{\pi^2}{8} * |x^2 - x|, \text{ διότι το } \left| \sin \frac{\pi \xi}{2} \right| \leq 1.$$

d) Για $x = 0.5$ το $|f(0.5) - P_1(0.5)| \leq \frac{\pi^2}{8} |0.5^2 - 0.5| = 0.308$, για $x = 0.1$ τότε $|f(0.1) - P_1(0.1)| \leq \frac{\pi^2}{8} |0.1^2 - 0.1| = 0.11$ και για $x = 0.01$ τότε $|f(0.01) - P_1(0.01)| \leq \frac{\pi^2}{8} * |0.01^2 - 0.01| = 0.01$.

Άσκηση 2

Να παρέχετε ένα άνω φράγμα για το σφάλμα παρεμβολής Lagrange των ακόλουθων συναρτήσεων: $f_1(x) = \cos h(x)$, $f_2(x) = \sin h(x)$, $f_3(x) = \cos(x) + \sin(x)$ και τα σημεία παρεμβολής για τις $f_1(x)$, $f_2(x)$ δίνονται από τον τύπο $x_k = -1 + 0.5k$, $k = 0, 1, 2, 3, 4$ και τα σημεία παρεμβολής για την $f_3(x)$ δίνονται από τον τύπο $x_k = -\frac{\pi}{2} + \frac{\pi k}{4}$, $k = 0, 1, 2, 3, 4$.

Λύση

Επειδή σε όλες τις περιπτώσεις το $n = 4$ (βαθμός πολυωνύμου, ένας μικρότερος από τα σημεία παρεμβολής), θα πρέπει να εκτιμήσουμε την 5^η παράγωγο κάθε συνάρτησης στο διοθέν διάστημα. Ισχύει ότι: $(\sin h(x))' = \cos h(x)$ και $(\cos h(x))' = \sin h(x)$. Η $(\cos h(x))''''' = \sin h(x)$ και η $(\sin h(x))''''' = \cos h(x)$, όπου $\sin h(x) = \frac{e^x - e^{-x}}{2}$ και το $\cos h(x) = \frac{e^x + e^{-x}}{2}$ και το $x \in [-1, 1]$. Αν θέσουμε όπου $x = 1$, το $\sin h(x) = \frac{e^1 - e^{-1}}{2} = 1.175 \cong 1.18$. Άρα: $|f_1(x)^{(5)}| < 1.18$ και για να υπολογίσουμε ολοκληρωμένα το άνω φράγμα σφάλματος, χρησιμοποιούμε

τον τύπο: $\max|e_n f(x)| \leq \frac{\max_{x \in \Omega} |f^{(n+1)}(x)|}{4(n+1)} * \left(\frac{x_n - x_0}{n}\right)^{(n+1)}$ οπότε έχουμε ότι: $\max|e_n f(x)| \leq \frac{1.18}{20} * \left(\frac{1}{2}\right)^5$. Επίσης, το

$\cos h(x) = \frac{e^1 + e^{-1}}{2} = 1.543 \cong 1.54$. Άρα $|f_2(x)^{(5)}| < 1.54$ και για να υπολογίσουμε ολοκληρωμένα το άνω φράγμα σφάλματος, χρησιμοποιούμε τον τύπο: $\max|e_n f(x)| \leq \frac{\max_{x \in \Omega} |f^{(n+1)}(x)|}{4(n+1)} * \left(\frac{x_n - x_0}{n}\right)^{(n+1)}$ οπότε έχουμε ότι: $\max|e_n f(x)|$

$\leq \frac{1.54}{20} * \left(\frac{1}{2}\right)^5$. Αναφορικά με την f_3 έχουμε το διάστημα $\left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$ και $f_3(x) = \cos(x) + \sin(x)$ και $f_3'(x) = -\sin(x) + \cos(x)$, $f_3''(x) = -\cos(x) - \sin(x)$, $f_3'''(x) = \sin(x) - \cos(x)$, $f_3''''(x) = \cos(x) + \sin(x)$, $f_3'''''(x) = -\sin(x) + \cos(x)$ και αν θέσουμε ως $x = \frac{\pi}{2}$ τότε $f_3'''''(x) = -\sin\left(\frac{\pi}{2}\right) + \cos\left(\frac{\pi}{2}\right) = -1 + 0$. Αν θέσουμε ως $x = -\frac{\pi}{2}$ τότε $f_3'''''(x) = 1 + 0 = 1$. Άρα $|f_3(x)^{(5)}| < 1$ και αν θέσουμε ως $x = -\frac{\pi}{4}$ έχουμε ότι $f_3^{(5)} = -\sin\left(-\frac{\pi}{4}\right) + \cos\left(-\frac{\pi}{4}\right) = \frac{\sqrt{2}}{2} + \frac{\sqrt{2}}{2} = \sqrt{2}$. Άρα $f_3^{(5)} < \sqrt{2} = 1.41$. Επιλέγουμε την τιμή του x που δίνει το μικρότερο άνω φράγμα και στην προκειμένη περίπτωση $x = -\frac{\pi}{2}$ και για να υπολογίσουμε ολοκληρωμένα το άνω φράγμα σφάλματος, χρησιμοποιούμε τον τύπο: $\max|e_n f(x)| \leq \frac{\max_{x \in \Omega} |f^{(n+1)}(x)|}{4(n+1)} * \left(\frac{x_n - x_0}{n}\right)^{(n+1)}$ οπότε έχουμε ότι: $\max|e_n f(x)| \leq \frac{1}{20} * \left(\frac{\pi}{4}\right)^5$.

Άσκηση 3 (παλιό θέμα)

Θέλουμε να κατασκευάσουμε προσεγγίσεις της συνάρτησης: $f(x) = x^4 + 1$ με βάση τις τιμές της στα σημεία $\pm 1/2$ και 0. Ειδικότερα: (α) Να κατασκευάσετε και να εκφράσετε το πολυώνυμο παρεμβολής για τα σημεία αυτά στη μορφή Lagrange. Προσοχή: πρέπει να γράψετε τη μορφή Lagrange με υπολογισμένη την τιμή του σταθερού συντελεστή του κάθε όρου (που είναι πολυώνυμου κάποιου βαθμού) και (β) με βάση τη θεωρία να υπολογίσετε **καλό φράγμα για το μέγιστο σε απόλυτη τιμή σφάλμα** που μπορεί να υπάρξει στο διάστημα $T = (-1/2, 1/2)$ αν χρησιμοποιηθεί το παραπάνω πολυώνυμο παρεμβολής αντί της $f(x)$. Να υπολογίσετε επίσης και τα σημεία του διαστήματος T , όπου το σφάλμα λαμβάνει τη μέγιστη και την ελάχιστη τιμή του.

Λύση

(α) Επειδή έχουμε **τρία σημεία παρεμβολής**, $x_0 = -\frac{1}{2}$, $x_1 = 0$, $x_2 = \frac{1}{2}$, ο **βαθμός του πολυωνύμου είναι $n = 2$** . Για να υπολογίσουμε το πολυώνυμο παρεμβολής δευτέρου βαθμού στη μορφή Lagrange, θα χρησιμοποιήσουμε τον τύπο: $P_2(x) = L_0 * f(x_0) + L_1 * f(x_1) + L_2 * f(x_2) = \frac{(x-x_1)*(x-x_2)}{(x_0-x_1)*(x_0-x_2)} * f(x_0) + \frac{(x-x_0)*(x-x_2)}{(x_1-x_0)*(x_1-x_2)} * f(x_1) + \frac{(x-x_0)*(x-x_1)}{(x_2-x_0)*(x_2-x_1)} * f(x_2) = \frac{(x-1/2)*(x-0)}{(-1/2-0)*(-1/2-1/2)} * \frac{17}{16} + \frac{(x+1/2)*(x-1/2)}{(0+1/2)*(0-1/2)} * 1 + \frac{(x+1/2)*(x-0)}{(1/2+1/2)*(1/2-0)} * \frac{17}{16} = 0.25 * x^2 + 1$.

(β) Μας δίδονται τα άκρα του διαστήματος T ($x_0 = -\frac{1}{2}$ και $x_n = \frac{1}{2}$) και ο βαθμός του πολυωνύμου είναι $n = 2$ και θα χρησιμοποιήσουμε τον τύπο: $\max|e_n f(x)| \leq \frac{\max_{x \in \Omega} |f^{(n+1)}(x)|}{4(n+1)} h^{(n+1)}$ όπου $h = \frac{x_n - x_0}{2} = \frac{\frac{1}{2} - (-\frac{1}{2})}{2} = 1/2$. Επίσης η τρίτη παράγωγος της $f(x) = x^4 + 1$ είναι $f(x)''' = 24 * x$, οπότε $\max|e_n f(x)| \leq \left|\frac{24*x}{4*(2+1)}\right| * \left(\frac{1}{2}\right)^3 = \left|\frac{24*x}{4*(2+1)}\right| * \frac{1}{2^3} = \left|\frac{x}{4}\right|$. Για να υπολογίσουμε στη συνέχεια τα σημεία του διαστήματος T όπου το σφάλμα λαμβάνει τη μέγιστη και την

ελάχιστη τιμή του, θα χρησιμοποιήσουμε το προηγούμενο αποτέλεσμα, θέτοντας $x = -1/2$ ή $x = 1/2$ και βρίσκουμε $1/8$ (μέγιστο σφάλμα) και για $x = 0$ το σφάλμα μηδενίζεται.

Άσκηση 4 (παλιό θέμα)

Δίνεται ο πίνακας τιμών με τα ζεύγη (x_i, y_i) που φαίνεται παρακάτω:

x_i	-0.5	0.0	0.5	1.0
y_i	1.5	0.0	1.5	1.0

Οι τελευταίοι δύο όροι του πολυνύμου παρεμβολής $p(x)$ που ικανοποιεί $p(x) = y$ είναι: a) $+6x(x-1/2)(x+1/2)(x-1)$, b) $-6x(x+1/2)(x-1) + 4/3(x+1/2)x((x-1/2)$, c) $6x(x-1/2)(x-1) + 4/3(x+1/2)x(x-1/2)$, d) κανένα από τα υπόλοιπα.

Λύση

Επειδή έχουμε τέσσερα σημεία παρεμβολής, $x_0 = -1/2$, $x_1 = 0$, $x_2 = 1/2$, $x_3 = 1$, ο βαθμός του πολυωνύμου είναι $n = 3$. Για να υπολογίσουμε το πολυώνυμο τρίτου βαθμού παρεμβολής στη μορφή Lagrange, θα χρησιμοποιήσουμε τον τύπο: $P_3(x) = L_0 * f(x_0) + L_1 * f(x_1) + L_2 * f(x_2) + L_3 * f(x_3) = \frac{(x-x_1)*(x-x_2)*(x-x_3)}{(x_0-x_1)*(x_0-x_2)*(x_0-x_3)} * f(x_0) +$

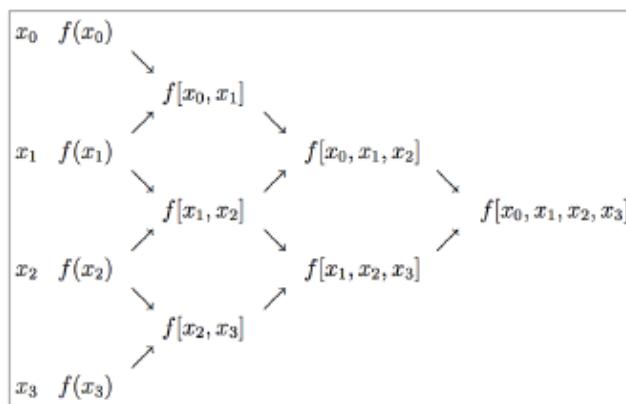
$$\frac{(x-x_0)*(x-x_2)*(x-x_3)}{(x_1-x_0)*(x_1-x_2)*(x_1-x_3)} * f(x_1) + \frac{(x-x_0)*(x-x_1)*(x-x_3)}{(x_2-x_0)*(x_2-x_1)*(x_2-x_3)} * f(x_2) + \frac{(x-x_0)*(x-x_1)*(x-x_2)}{(x_3-x_0)*(x_3-x_1)*(x_3-x_2)} * f(x_3)$$

Οι τελευταίοι δύο όροι του πολωνύμου παρεμβολής είναι:

$$\frac{(x-x_0)*(x-x_1)*(x-x_3)}{(x_2-x_0)*(x_2-x_1)*(x_2-x_3)} * f(x_2) + \frac{(x-x_0)*(x-x_1)*(x-x_2)}{(x_3-x_0)*(x_3-x_1)*(x_3-x_2)} * f(x_3) = \frac{(x+1/2)*(x-0)*(x-1/2)}{(1/2+1/2_0)*(1/2-0)*(1/2-1)} * 3/2 + \frac{(x+1/2)*(x-0)*(x-1/2)}{(1+1/2)*(1-0)*(1-1/2)} * 1 \\ = -6x * (x+1/2) * (x-1) + 4/3 * x * (x+1/2) * (x-1/2). \text{ Άρα η σωστή απάντηση είναι η } (\beta).$$

4.6 Ασκήσεις πολυωνύμων παρεμβολής στη μορφή Newton με χρήση πίνακα Δ.Δ.

Η μορφή Newton είναι ένας εναλλακτικός τρόπος υπολογισμού των συντελεστών ενός πολυωνύμου παρεμβολής, η οποία υλοποιείται συνήθως με τη μορφή ενός πίνακα διαιρεμένων διαφορών (Δ.Δ.). Θέτοντας $f[x_i] = f(x_i) = y_i$ και με βάση τον ορισμό: $f[x_i, x_{i+1}, \dots, x_j] = \frac{f[x_{i+1}, \dots, x_j] - f[x_i, x_{i+1}, \dots, x_{j-1}]}{x_j - x_i}$, έχουμε ότι η γενική μορφή ενός τέτοιου πίνακα είναι αυτή που φαίνεται στη συνέχεια:



Εικόνα 12: Μορφή πίνακα διαιρεμένων διαφορών

Για τη συμπλήρωση των στοιχείων του πίνακα διαιρεμένων διαφορών ισχύουν οι εξής σχέσεις: $f[x_i, x_j] = \frac{f[x_j] - f[x_i]}{x_j - x_i} = \frac{f_j - f_i}{x_j - x_i}$ και $f[x_i, x_j, x_k] = \frac{f[x_j, x_k] - f[x_i, x_j]}{x_k - x_i}$ και $f[x_i, x_j, x_k, x_e] = \frac{f[x_j, x_k, x_e] - f[x_i, x_j, x_k]}{x_e - x_i}$ κ.ο.κ. Η μορφή του πολυωνύμου παρεμβολής σε μορφή Newton μέσω του πίνακα Δ.Δ. είναι: $P_n(x) = f[x_0] + f[x_0, x_1](x - x_0) + f[x_0, x_1, x_2](x - x_0)(x - x_1) + \dots + f[x_0, x_1, x_2, \dots, x_n](x - x_0)(x - x_1) \dots (x - x_{n-1})$. Δηλαδή παρατηρούμε ότι, αφού ολοκληρώσουμε τον πίνακα Δ.Δ., λαμβάνουμε στη συνέχεια υπόψη μας το πρώτο στοιχείο κάθε στήλης του πίνακα αυτού, αρχής γενομένης από τη δεύτερη στήλη, ως συντελεστές του πολυωνύμου παρεμβολής..

4.6.1 Σύγκριση πολυωνύμων παρεμβολής

Δίνονται οι τιμές $f(-1) = 1$, $f(1) = 2$, $f(2) = 5$ και ζητάμε το πολυώνυμο παρεμβολής τόσο στη μορφή Lagrange όσο και στη μορφή Newton.

Λύση

α) Στη μορφή Lagrange το π.π. έχει τη μορφή:

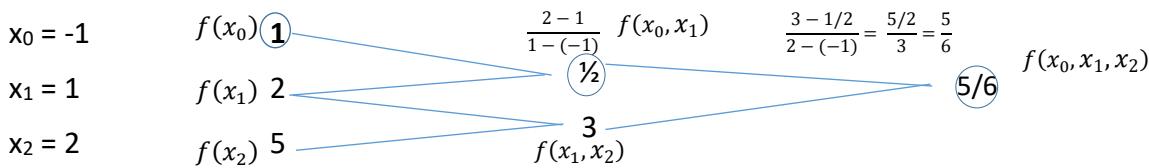
$$P_2(x) = f(x_0)L_0(x) + f(x_1)L_1(x) + f(x_2)L_2(x) = f(-1)L_0(x) + f(1)L_1(x) + f(2)L_2(x) = 1 * L_0(x) + 2 * L_1(x) + 5 * L_2(x).$$

$$\text{Τα πολυώνυμα Lagrange είναι τα εξής: } L_0(x) = \frac{(x-1)(x-2)}{(-1-1)(-1-2)} = \frac{(x-1)(x-2)}{6}, \quad L_1(x) = \frac{(x+1)(x-2)}{(1-(-1))(1-2)} = \frac{(x+1)(x-2)}{-2} \text{ και}$$

$$\text{το } L_2(x) = \frac{(x+1)(x-1)}{(2-(-1))(2-1)} = \frac{(x+1)(x-1)}{3}. \text{ Άρα το πολυώνυμο παρεμβολής 2ου βαθμού στη μορφή Lagrange είναι:}$$

$$P_2(x) = \frac{(x-1)(x-2)}{6} - (x+1)*(x-2) + \frac{5}{3}*(x+1)*(x-1) = \frac{x^2-3x+2}{6} - (x^2-x-2) + \frac{5}{3}*(x^2-1) = \\ \frac{x^2-3x+2-6x^2+6x+12+10x^2-10}{6} = \frac{5}{6}x^2 + \frac{1}{2}x + \frac{4}{6}.$$

β) Αν χρησιμοποιήσουμε τη μορφή Newton σε μορφή πίνακα Δ.Δ. θα έχουμε ότι:



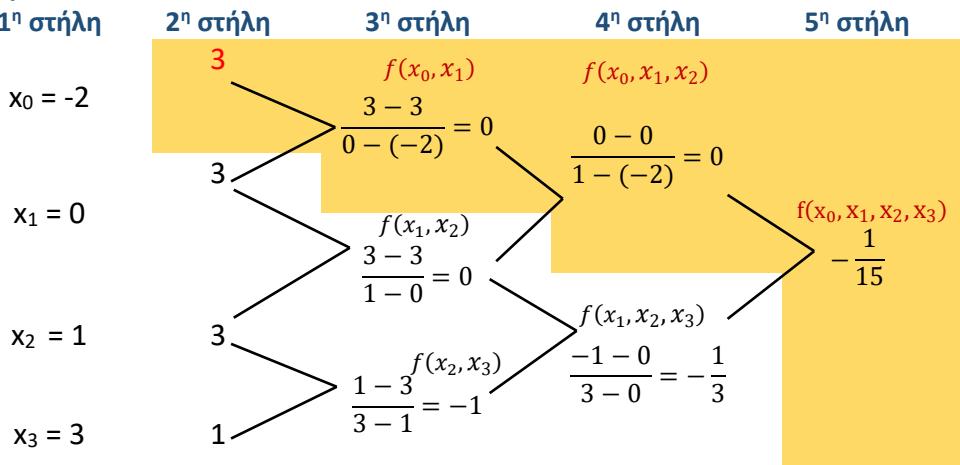
Άρα το πολυώνυμο παρεμβολής 2ου βαθμού που προκύπτει (αφού έχουμε τρία σημεία παρεμβολής και ως γνωστό ο βαθμός του πολυωνύμου παρεμβολής είναι κατά ένα μικρότερος από το πλήθος των σημείων παρεμβολής) είναι το $P_2(x) = f(x_0) + f(x_0, x_1)*(x - x_0) + f(x_0, x_1, x_2)*(x - x_0)*(x - x_1) = 1 + \frac{1}{2}(x + 1) + \frac{5}{6}(x + 1)(x - 1) = 1 + \frac{x}{2} + \frac{1}{2} + \frac{5}{6}(x^2 - 1) = \frac{x}{2} + \frac{3}{2} + \frac{5}{6}x^2 - \frac{5}{6} = \frac{5}{6}x^2 + \frac{4}{6} + \frac{1}{2}x$.

Συμπέρασμα: παρατηρούμε ότι βρήκαμε τους ίδιους συντελεστές, κάτι αναμενόμενο αφού ως γνωστό, το π.π. είναι ένα και μοναδικό, ανεξάρτητα από τον τρόπο που χρησιμοποιείται για τον υπολογισμό του, αν δηλαδή είναι η μορφή Lagrange ή η μορφή Newton ή κάποια άλλη μορφή. Η μορφή Newton -αν και έχει διαφορετικό τρόπο υπολογισμού- θα υπολογίζεται πάντα για τις ανάγκες του μαθήματος με τη βοήθεια του πίνακα Δ.Δ.

Άσκηση 1 – Παρεμβολή Newton

Στον παρακάτω πίνακα τιμών δίνονται τα σημεία παρεμβολής x_i , $i = 0, 1, 2, 3$ της συνάρτησης $f(x)$. Με χρήση πίνακα διαιρεμένων διαφορών (δ. δ.) να βρεθεί το πολυώνυμο παρεμβολής στη μορφή Newton.

x_i	-2	0	1	3
f_i	3	3	3	1

Λύση

$$\pi_3(x) = f(x_0) + f(x_0, x_1)(x - x_0) + f(x_0, x_1, x_2)(x - x_0)(x - x_1) + f(x_0, x_1, x_2, x_3)(x - x_0)(x - x_1)(x - x_2) = 3 + \left(-\frac{1}{15}\right)(x + 2)(x - 0)(x - 1) = -\frac{1}{15}(x^3 + x^2 - 2x + 45), \text{ óπου } \pi_3(-2) = 3 = f_0, \pi_3(0) = 3 = f_1$$

Άσκηση 2 – Παρεμβολή Newton

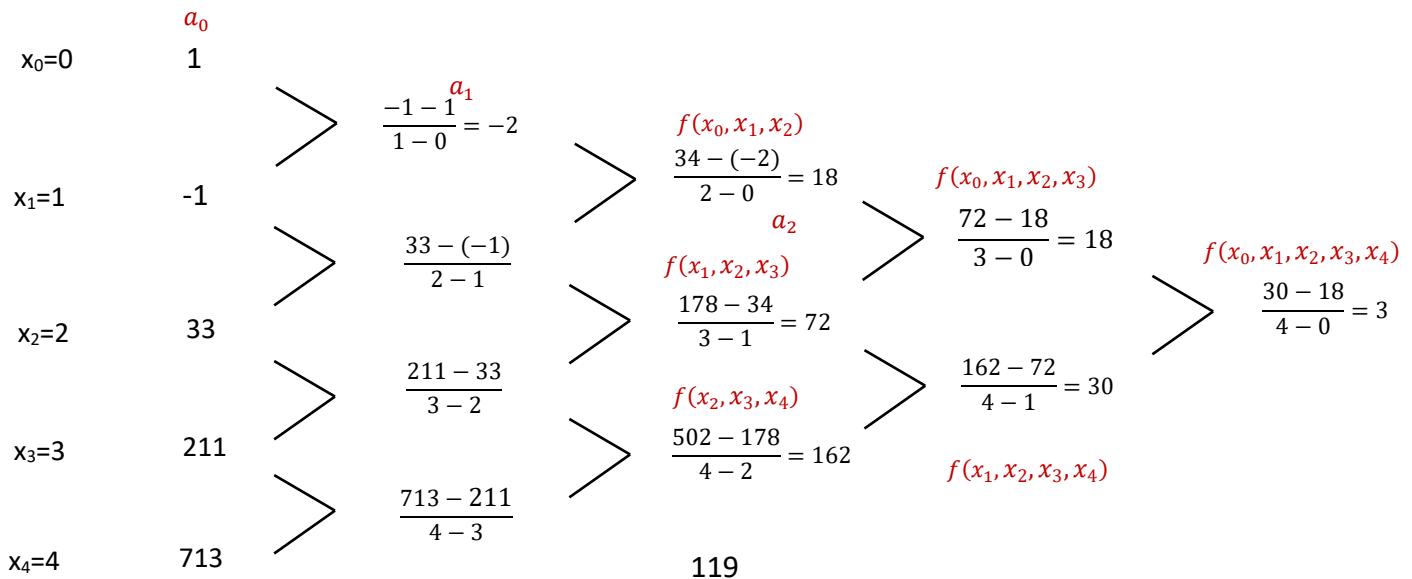
Δίνονται τα ζεύγη σημείων:

x_i	0	1	2	3	4
y_i	1	-1	33	211	713

Να προσεγγιστεί το πολυώνυμο παρεμβολής στη μορφή Newton για τα παραπάνω ζεύγη.

Λύση

Επειδή θέλουμε ένα πολυώνυμο παρεμβολής σε μορφή Newton, θα πρέπει πρώτα να κατεσκευάσουμε τον πίνακα διαιρεμένων διαφορών, για τα σημεία παρεμβολής που δίνονται.



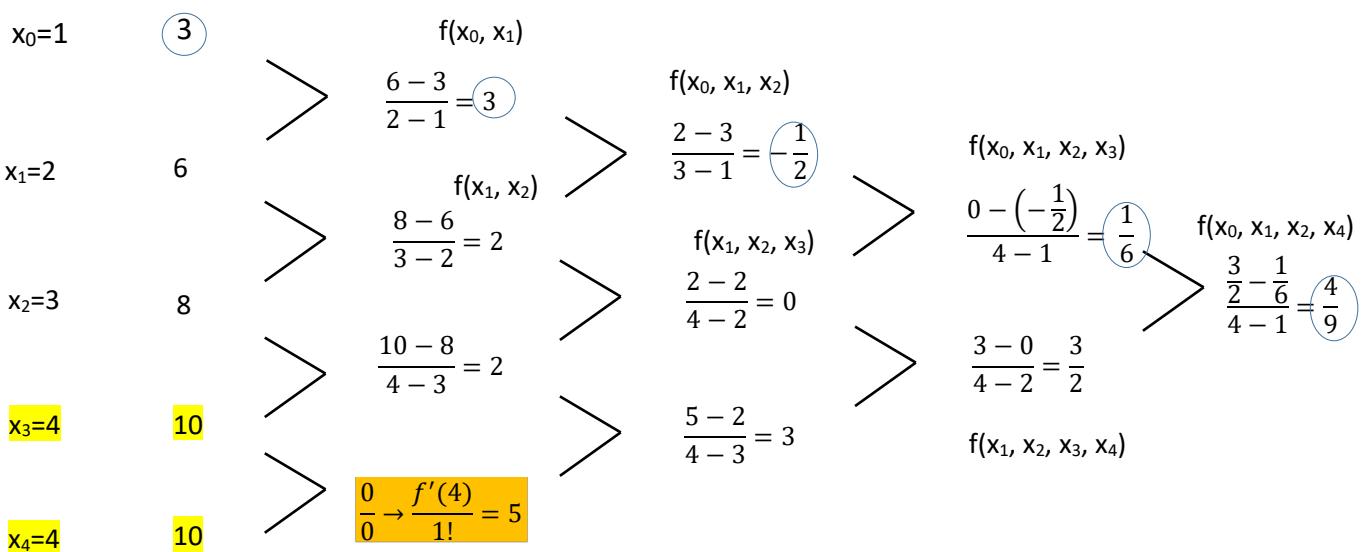
Άρα το π.π. στη μορφή Newton είναι το εξής: $p_4(x) \text{ ή } \pi_4(x) = \alpha_0 + \alpha_1 * (x - x_0) + \alpha_2 * (x - x_0) * (x - x_1) + \alpha_3 * (x - x_0)(x - x_1)(x - x_2) + \alpha_4 * (x - x_0)(x - x_1)(x - x_2)(x - x_3) = 1 + (-2) * (x - 0) + 18 * (x - 0)(x - 1) + 18 * (x - 0)(x - 1)(x - 2) + 3(x - 0)(x - 1)(x - 2)(x - 3) = 3x^4 - 3x^2 - 2x + 1.$

Παρατήρηση 1: Πάντοτε όταν κατασκευάζουμε ένα π.π. σε μορφή Newton, θα πρέπει να έχουμε υπόψη ότι στον τελευταίο όρο του θα πρέπει να εμφανίζονται όλα τα σημεία παρεμβολής εκτός του τελευταίου. Στην προκειμένη περίπτωση που έχουμε 5 σημεία παρεμβολής από $x_0 - x_4$, ο τελευταίος όρος του πολυωνύμου παρεμβολής δεν περιέχει τον όρο x_4 .

Παρατήρηση 2: Πάντοτε όταν κατασκευάζουμε ένα π.π. σε μορφή Newton και στα σημεία παρεμβολής δίνονται και παράγωγοι της συνάρτησης $f(x)$, θα πρέπει να έχουμε υπόψη ότι ανάλογα με την τάξη της παραγώγου που δίνεται/δίνονται, **έχουμε επανάληψη των σημείων παρεμβολής τόσες φορές όσο είναι η τάξη της παραγώγου**. Αυτή η επανάληψη αφορά επίσης και τις τιμές της συνάρτησης $f(x)$. Στην περίπτωση αυτή το πολυώνυμο παρεμβολής ονομάζεται **Hermite**, και ουσιαστικά είναι το **πολυώνυμο Newton με παραγώγους**.

Άσκηση 3 – Παρεμβολή Newton

Να λυθεί το πρόβλημα παρεμβολής με συνθήκες $f(1) = 3$, $f(2) = 6$, $f(3) = 8$, $f(4) = 10$, $f'(4) = 5$

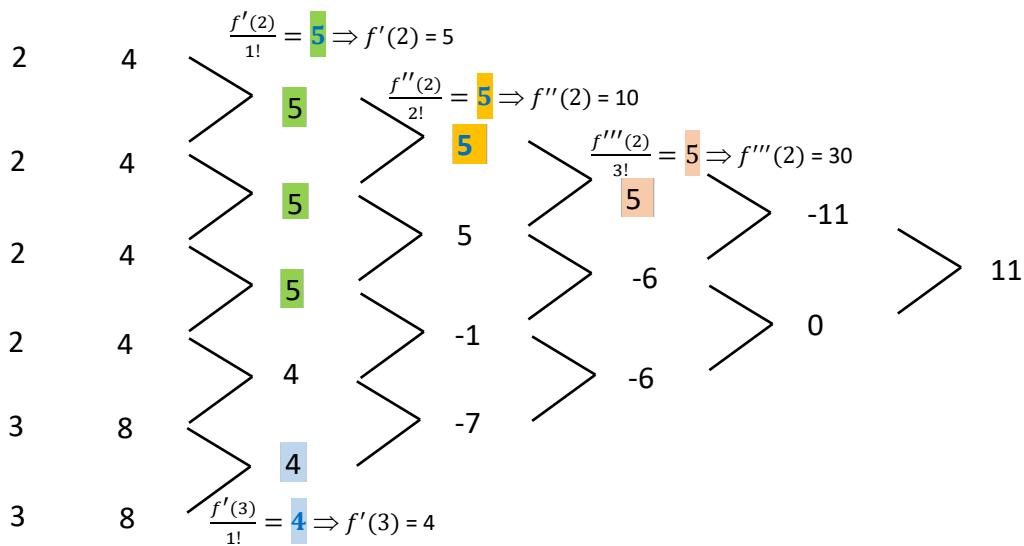


Άρα το π.π. στη μορφή Newton θα είναι: $\pi_4(x) = \alpha_0 + \alpha_1(x - x_0) + \alpha_2(x - x_0)(x - x_1) + \alpha_3(x - x_0)(x - x_1)(x - x_2) + \alpha_4(x - x_0)(x - x_1)(x - x_2)(x - x_3) = 3 + 3(x - 1) - \frac{1}{2}(x - 1)(x - 2) + \frac{1}{6}(x - 1)(x - 2)(x - 3) + \frac{4}{9}(x - 1)(x - 2)(x - 3)(x - 4) = \frac{1}{18}(8x^4 - 77x^3 + 253x^2 - 286x + 156)$

Παρατήρηση: Σε περίπτωση που αφαιρεθούν τα $f(1)$, $f'(4)$, τότε ο πίνακας Δ.Δ. δίνεται ως εξής: $\pi_2(x) = 6 + 2(x - 2) + 0(x - 2)(x - 3)$. Τότε έχουμε ένα πολυώνυμο παρεμβολής δευτέρου βαθμού.

Άσκηση 4 – Παρεμβολή Newton

Έστω ο πίνακας Δ.Δ. που δίνεται παρακάτω:



Να βρεθούν οι αρχικές συνθήκες

Λύση

Όταν έχουμε επανάληψη τιμών, τότε έχουμε παραγώγους και μάλιστα η τιμή της παραγώγου εξαρτάται από τον αριθμό των επαναλήψεων. Πιο συγκεκριμένα, αν στην πρώτη στήλη μια ρίζα επαναλαμβάνεται μόνο μία φορά, τότε υπάρχει η πρώτη παράγωγος, αν επαναλαμβάνεται δύο φορές τότε υπάρχει και η δεύτερη παράγωγος κ.ο.κ. Άρα, στην προκειμένη περίπτωση έχουμε ότι η $f(2) = 4$, $f(3) = 8$, $\frac{f'(2)}{1!} = 5$, $f'(3) = \frac{4}{1!} = 4$, $\frac{f''(2)}{2!} = 5 \Rightarrow f''(2) = 2! * 5 = 10$, $\frac{f'''(2)}{3!} = 5 \Rightarrow f'''(2) = 3! * 5 = 6 * 5 = 30$. Ο υπολογισμός των παραγώγων ξεκινά από την 3^η στήλη και μετά. Δηλαδή στην τρίτη στήλη υπολογίζεται μόνο η 1^η παράγωγος, στην τέταρτη στήλη μόνο η 2η παράγωγος κ.ο.κ.

Παρατήρηση: η διαφορά του π.π. σε μορφή Newton και σε μορφή Lagrange είναι ότι στη μορφή Newton, αν προστεθούν εκ' των υστέρων καινούργια σημεία, **δεν** χρειάζεται να γίνει ο επαναϋπολογισμός όλου του π.π., κάτι που πρέπει να γίνει όταν έχουμε την μορφή Lagrange. Αυτό το γεγονός αναδεικνύει το σημαντικό **πλεονέκτημα που έχει η μορφή Newton έναντι της μορφής Lagrange, που ονομάζεται οικονομική ανανέωση**.

Άσκηση 5 - Παρεμβολή Newton

Να λυθεί το πρόβλημα παρεμβολής με συνθήκες $f(1) = 3$, $f'(1) = 2$, $f''(1) = 2$, $f'''(1) = 12$, $f(2) = 0$, $f'(2) = 1$, δηλαδή να υπολογιστεί το πολυώνυμο παρεμβολής σε μορφή Newton με χρήση πίνακα Δ.Δ.:

Λύση

$x_0 = 1$ **3** $x_1 = 1$

3

$$\frac{3-3}{1-1} \rightarrow \frac{f'(1)}{1!} = 2 \quad \frac{2-2}{1-1} \rightarrow \frac{f''(1)}{2!} = 1$$

 $x_2 = 1$

3

$$\frac{3-3}{1-1} \rightarrow \frac{f'(1)}{1!} = 2 \quad \frac{1-1}{1-1} \rightarrow \frac{f'''(1)}{3!} = 2$$

$$\frac{2-2}{1-1} \rightarrow \frac{f''(1)}{2!} = 1 \quad \frac{-6-2}{2-1} = -8$$

 $x_3 = 1$

3

$$\frac{3-3}{1-1} \rightarrow \frac{f'(1)}{1!} = 2 \quad \frac{-5-1}{2-1} = -6$$

$$\frac{15+8}{2-1} = 23$$

 $x_4 = 2$

0

$$\frac{0-3}{2-1} = -3 \quad \frac{-3-2}{2-1} = -5$$

$$\frac{9+6}{2-1} = 15$$

 $x_5 = 2$

0

$$\frac{0-0}{2-2} \rightarrow \frac{f'(2)}{1!} = 1 \quad \frac{1+3}{2-1} = 4$$

$$\frac{4+5}{2-1} = 9$$

Άρα το π.π. στη μορφή Newton θα είναι: $P_5(x) = 3 + 2 * (x-1) + 1 * (x-1) * (x-1) + 2 * (x-1) * (x-1) * (x-1) * (x-1) - 8 * (x-1) * (x-1) * (x-1) * (x-1) + 23 * (x-1) * (x-1) * (x-1) * (x-2) = 3 + 2 * (x-1) + 1 * (x-1)^2 + 2 * (x-1)^3 - 8 * (x-1)^4 + 23 * (x-1)^4 * (x-2) = 23x^5 - 146x^4 + 356x^3 - 421x^2 + 245x - 54.$

Άσκηση 6 – Παρεμβολή Newton - ελαχιστοποίηση τετραγωνικού σφάλματος με γραμμικό σύστημα

Δίνεται ο επόμενος πίνακας με τα ζεύγη σημείων (x_i, y_i).

x_i	0	0.5	1	1.5
y_i	0	0	1	2.4

α) Να υπολογίσετε το πολυώνυμο με το μικρότερο βαθμό που να παρεμβάλλει τα συγκεκριμένα δεδομένα.

Να απλοποιήστε το αποτέλεσμα τόσο ώστε ο σταθερός συντελεστής κάθε όρου του γινομένου της μορφής να είναι ένας αριθμός και όχι σε μορφή γινομένου.

β) Να γράψετε το γραμμικό σύστημα $M * x = g$, όπου το M τετραγωνικό μητρώο τέτοιο ώστε το x να περιέχει τους συντελεστές α, β της ευθείας $\alpha + \beta x$ που **ελαχιστοποιεί το τετραγωνικό σφάλμα** για τα παραπάνω δεδομένα. Δεν χρειάζεται να λυθεί το σύστημα, απλώς να υπολογιστούν οι τιμές των M και g .

Λύση

α) Κατασκευάζουμε τον πίνακα **διαιρεμένων διαφορών**, ο οποίος φαίνεται παρακάτω:

0	0	0	
1/2	0	2	
1	1	0.8	-0.8
3/2	2.4	2.8	

Άρα το π.π. στη μορφή Newton θα είναι: $\pi_3(x) = \alpha_0 + \alpha_1(x - x_0) + \alpha_2(x - x_0)(x - x_1) + \alpha_3(x - x_0)(x - x_1)(x - x_2) = 0 + 0 + 2 * (x - 0) * \left(x - \frac{1}{2}\right) - 0.8 * (x - 0) * \left(x - \frac{1}{2}\right) * (x - 1) = 2x^2 - x - 0.8 * (x^3 - 3x^2/2 + x/2) = -0.8 * x^3 + 3.2 * x^2 - 1.4 * x.$

Παρεμβολή σημαίνει κατασκευή μιας πολυωνυμικής συνάρτησης που ταιριάζει ακριβώς με τις παρατηρήσεις σε όλα τα σημεία, δηλ. $p(x_i) = y_i$ για $i = 1, \dots, m$. Επομένως με ευθεία μπορεί να γίνει παρεμβολή μόνον αν τα δεδομένα τυχαίνει και είναι ακριβώς συγγραμμικά (οπότε και το σφάλμα - θεωρητικά σε αριθμητική άπειρης ακρίβειας - θα είναι μηδέν). Γενικότερα, για τη διαφορετικά ζεύγη σημείων (x_i, y_i) , είναι πάντα δυνατή η κατασκευή ενός πολυωνύμου βαθμού το πολύ $m-1$, για το οποίο ισχύει ότι $p(x_i) = y_i$ για $i = 1, \dots, m$. Δηλαδή στα σημεία παρεμβολής πρέπει οι τιμές του πολυωνύμου να είναι ίσες με τις τιμές της συνάρτησης που παρεμβάλει.

β) Στη **μέθοδο Newton** κατασκευάζουμε το γραμμικό σύστημα παρεμβολής $g = M * x$, όπου η βάση πολυωνύμων Newton είναι: $\{1, (x - x_0), (x - x_0)(x - x_1), \dots, (x - x_0)(x - x_1)\dots(x - x_{n-1})\}$. Επομένως το π.π. σε μορφή Newton είναι $\Pi_n(x) = \gamma_0 + \gamma_1(x - x_0) + \dots + \gamma_n(x - x_0)(x - x_1)\dots(x - x_{n-1})$, όπου $\gamma_0, \gamma_1, \dots, \gamma_n$ είναι οι συντελεστές που έχουν προκύψει από τον πίνακα Δ.Δ., και από τις συνθήκες παρεμβολής $\Pi_n(x_j) = y_j$, $j = 0, \dots, n$ έχουμε ότι:

$$\Pi_n(x_0) = y_0 = \gamma_0$$

$$\Pi_n(x_1) = y_1 = \gamma_0 + \gamma_1(x_1 - x_0)$$

$$\Pi_n(x_2) = y_2 = \gamma_0 + \gamma_1(x_2 - x_0) + \gamma_2(x_2 - x_0)(x_2 - x_1)$$

.....

$$\Pi_n(x_n) = y_n = \gamma_0 + \gamma_1(x_n - x_0) + \gamma_2(x_n - x_0)(x_n - x_1) + \dots + \gamma_n(x_n - x_0)(x_n - x_1)\dots(x_n - x_{n-1}).$$

Επομένως για το **γραμμικό σύστημα παρεμβολής** $y = A * c$ (ή αλλιώς $g = M * x$) έχουμε

$$\begin{pmatrix} y_0 \\ y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & 0 & & \cdots & \\ 1 & x_1 - x_0 & & 0 & \\ 1 & (x_2 - x_0) & (x_2 - x_0)(x_2 - x_1) & & \ddots \\ \vdots & & & & \ddots \\ 1 & x_n - x_0 & (x_n - x_0)(x_n - x_1) & (x_n - x_0)(x_n - x_1)\dots(x_n - x_{n-1}) & \end{pmatrix} \begin{pmatrix} \gamma_0 \\ \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_n \end{pmatrix}$$

Αν αντικαταστήσουμε στην προκειμένη περίπτωση στην ευθεία $y = \alpha + \beta * x$ τις τιμές των x, y από τον πίνακα που δίνεται, έχουμε ότι:

$$0 = \alpha + \beta * 0$$

$$0 = \alpha + \beta * 0.5$$

$$1 = \alpha + \beta * 1$$

$$2.4 = \alpha + \beta * 1.5$$

Από αυτές τις 4 εξισώσεις μπορούμε να υπολογίσουμε τα στοιχεία του διανύσματος g . Εφαρμόζοντας το δεύτερο προηγούμενο γενικό τύπο στην προκειμένη περίπτωση, έχουμε ότι:

$$\begin{pmatrix} 0 \\ 0 \\ 1 \\ 2.4 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & x_1 - x_0 & 0 & 0 \\ 1 & x_2 - x_0 & (x_2 - x_0) * (x_2 - x_1) & 0 \\ 1 & x_3 - x_0 & (x_3 - x_0) * (x_3 - x_1) & (x_3 - x_0) * (x_3 - x_1) * (x_3 - x_2) \end{pmatrix} * \begin{pmatrix} \end{pmatrix}$$

Σημείωση: στο γραμμικό σύστημα παρεμβολής $g = M * x$ είναι εμφανής η ευκολία προσθήκης νέου στοιχείου (οικονομική ανανέωση):

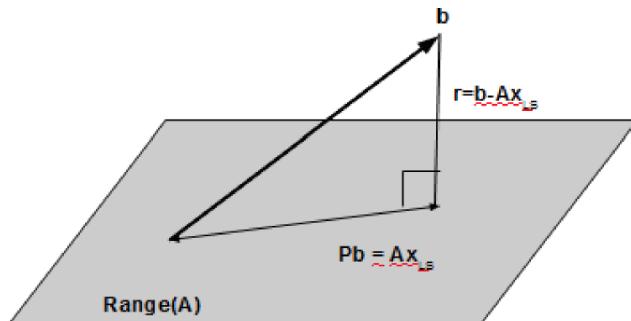
$$\begin{pmatrix} y_0 \\ y_1 \\ y_2 \\ \vdots \\ y_n \\ y_{n+1} \end{pmatrix} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 1 & x_1 - x_0 & & \\ 1 & (x_2 - x_0) & (x_2 - x_0)(x_2 - x_1) & \\ \vdots & & \ddots & \\ 1 & x_n - x_0 & (x_n - x_0)(x_n - x_1) & (x_n - x_0)(x_n - x_1) \cdots (x_n - x_{n-1}) \\ 1 & x_{n+1} - x_0 & (x_{n+1} - x_0)(x_{n+1} - x_1) & (x_{n+1} - x_0)(x_{n+1} - x_1) \cdots (x_{n+1} - x_n) \end{pmatrix} \begin{pmatrix} \gamma_0 \\ \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_n \\ \gamma_{n+1} \end{pmatrix}$$

Θεωρία ελαχίστων τετραγώνων – Πολυώνυμα παρεμβολής με τη μέθοδο ελαχίστων τετραγώνων

Τα ελάχιστα τετράγωνα αποτελούν μια από τις πιο διαδεδομένες μεθόδους προσέγγισης στις εφαρμογές. Αποτελεί τη μέθοδο επιλογής όταν έχουμε **πολλές μετρήσεις** και επιθυμούμε να τις μοντελοποιήσουμε με συνάρτηση που εξαρτάται από **λίγες παραμέτρους**. Η περιγραφή του προβλήματος των «ελαχίστων τετραγώνων» είναι η εξής: Δίνονται μητρώο $A \in \mathbb{R}^{m \times n}$ και διάνυσμα $b \in \mathbb{R}^m$, $m \geq n$ και το γραμμικό πρόβλημα των ελαχίστων τετραγώνων συνίσταται στην εύρεση ενός διανύσματος από το σύνολο $X = \{x \in \mathbb{R}^n : \text{ελαχιστοποιεί το πολυώνυμο παρεμβολής } p(x) = \|Ax - b\|_2 = \|b - Ax\|_2 \text{ ή αλλιώς } p(x) = \|y - Cx_{LS}\|_2\}$. Η τελευταία αναπαράσταση του σφάλματος χρησιμοποιείται όταν αντί για την κλασική εξίσωση $A * x = b$, θεωρήσουμε την εξίσωση $C * x = y$.

Η ελαχιστοποίηση της νόρμας $\|Ax - b\|_2$ ισοδυναμεί με την εύρεση ενός διανύσματος $x_{LS} \in \mathbb{R}^n$ τέτοιου ώστε μεταξύ όλων των διανυσμάτων που παράγονται από γραμμικό συνδυασμό των στηλών του A , το $p = A * x_{LS}$ να είναι το πλησιέστερο στο b (ως προς την ευκλείδεια νόρμα). Για να συμβαίνει αυτό πρέπει το $r = b - A * x_{LS}$ να **είναι κάθετο στον υπόχωρο $\text{range}(A)$ που αποτελεί το χώρο στηλών**, άρα θα πρέπει το $r \in \text{null}(A^T)$ που **αποτελεί τον αριστερό μηδενόχωρο**. Ως γνωστό από την Γ.Α. οι θεμελιώδεις υπόχωροι $\text{range}(A)$ και $\text{null}(A^T)$ είναι κάθετοι μεταξύ τους. Επομένως για οποιοδήποτε $y \in \mathbb{R}^m$ έχουμε: $0 = (A * y)^T * (b - A * x_{LS}) = y^T (A^T * b - A^T * A * x_{LS})$. Επομένως το x_{LS} είναι η λύση του συστήματος $A^T * A * x_{LS} = A^T * b$ που αποκαλούνται **κανονικές εξισώσεις** του προβλήματος των ελαχίστων τετραγώνων.

Ελάχιστα τετράγωνα (στην περίπτωση της γραμμής) σημαίνει επιλογή των α και β που χαρακτηρίζουν την ευθεία, έτσι ώστε η απόσταση $(y_j - \alpha x_j - \beta)^2$ που είναι το τετραγωνικό σφάλμα να ελαχιστοποιείται. Αποδεικνύεται εύκολα (π.χ. με διαφορικό λογισμό) ότι αυτό το πρόβλημα έχει μοναδική λύση.



Εικόνα 13: Πρόβλημα ελαχίστων τετραγώνων και εύρεση του διανύσματος x_{LS}

Όμως, ένας ίσως πιο εύκολος τρόπος είναι να δούμε το πρόβλημα αλγεβρικά, δηλαδή ως γραμμικό σύστημα $C * x = y$ όπου το διάνυσμα $y = (y_1, \dots, y_n)^T$, περιέχει τις τιμές της συνάρτησης στα σημεία παρεμβολής, το διάνυσμα $x = (\alpha, \beta)^T$ περιέχει τους συνετελούστες α και β της ευθείας και το μητρώο $C = [ones(m, 1), *$], όπου η δεύτερη στήλη του είναι το διάνυσμα $* = (x_1, \dots, x_m)^T$ που περιέχει τα σημεία παρεμβολής. Το μητρώο C έχει πάντα σταθερή την πρώτη στήλη του (με «1»), ενώ όσον φορά τη δεύτερη στήλη του, αυτή θα περιέχει πάντα τα σημεία παρεμβολής που δίνονται. Επομένως η βέλτιστη λύση ελαχίστων τετραγώνων (LS = Least Square) θα είναι η $x_{LS} = (C^T C)^{-1} * C^T * y$, ενώ το σφάλμα θα είναι η δεύτερη νόρμα του καταλοίπου $y - A * x_{LS}$. Αν χρησιμοποιηθεί ο τύπος $C * x = y$ το σφάλμα της παρεμβολής ελαχίστων τετραγώνων θα είναι $\|y - C * x_{LS}\|_2$, ενώ αν χρησιμοποιηθεί ο τύπος $A * x = b$ θα είναι: $\|b - A * x_{LS}\|$. Από τη στιγμή που υπολογίσαμε

το $x_{LS} = \begin{pmatrix} x_0 \\ x_1 \\ x_2 \\ x_3 \\ \vdots \end{pmatrix}$, μπορούμε στη συνέχεια να υπολογίσουμε το πολυώνυμο παρεμβολής στην περίπτωση των ελαχίστων τετραγώνων, που είναι της μορφής $\phi(t) = x_0 + x_1 * t + x_2 * t^2 + \dots$. Από αυτό φαίνεται η σχέση μεταξύ πολυωνύμου παρεμβολής και ελαχίστων τετραγώνων.

Άσκηση 7 – Ελάχιστα τετράγωνα (Θέμα Σεπτεμβρίου 2016)

Έστω ότι δίνονται τα ζεύγη $(1, 11)$, $(2, 17)$, $(3, 29)$, $(4, 53)$ που εκφράζουν τιμή μεταβλητής – τιμή μέτρησης και στόχος είναι η δημιουργία συναρτήσεων παρεμβολής και προσέγγισης της άγνωστης συνάρτησης $f(x)$ που παράγει τις δοθείσες τιμές. Ποιο είναι το πολυώνυμο προσέγγισης 1^{ου} βαθμού $g(x)$ σύμφωνα με το κριτήριο των ελαχίστων τετραγώνων;

Λύση

Χρησιμοποιώντας το τυπολόγιο των ελαχίστων τετραγώνων θα έχουμε ότι:

Σύστημα: $C * x = y$, $x = (\alpha, \beta)^T$, $y = (y_1, \dots, y_n)^T = [11, 17, 29, 53]^T$ $C = [\text{ones}(m, 1), x]$, $x = (x_1, \dots, x_m)^T = [1, 2, 3, 4]^T$.

Επομένως η πρώτη στήλη του μητρώου C θα περιέχει «1» και η δεύτερη στήλη θα περιέχει τους συντελεστές ή σημεία παρεμβολής (δηλ. το πρώτο σημείο κάθε ζεύγους της εκφώνησης, δηλ. τις τιμές 1, 2, 3, 4 από τα ζεύγη (1, 11), (2, 17), (3, 29), (4, 53)). Η λύση θα είναι $x_{LS} = (C^T C)^{-1} C^T * y$ και σφάλμα $\|y - Ax_{LS}\|_2$. Πιο

συγκεκριμένα, θα έχουμε ότι το μητρώο των συντελεστών θα είναι το $C = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{bmatrix}$ και η λύση θα είναι: $x_{LS} =$

$$(C^T C)^{-1} C^T \cdot y = \left(\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 \end{bmatrix} * \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{bmatrix} \right)^{-1} * \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 \end{bmatrix} * \begin{bmatrix} 11 \\ 17 \\ 29 \\ 53 \end{bmatrix} = \begin{bmatrix} 4 & 10 \\ 10 & 30 \end{bmatrix}^{-1} * \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 \end{bmatrix} * \begin{bmatrix} 11 \\ 17 \\ 29 \\ 53 \end{bmatrix} =$$

$$\begin{bmatrix} 4 & 10 \\ 10 & 30 \end{bmatrix}^{-1} * \begin{bmatrix} 110 \\ 344 \end{bmatrix} = \frac{1}{20} * \begin{bmatrix} 30 & -10 \\ -10 & 4 \end{bmatrix} * \begin{bmatrix} 110 \\ 344 \end{bmatrix} = \frac{1}{20} * \begin{bmatrix} -140 \\ 276 \end{bmatrix} = \begin{bmatrix} -7 \\ 13.8 \end{bmatrix}.$$

Επομένως, το πολυώνυμο προσέγγισης 1^{ου} βαθμού $g(x)$ σύμφωνα με το κριτήριο των ελαχίστων τετραγώνων θα είναι: $g(x) = x_0 + x_1 * x = -7 + 13.8 * x$.

Σημείωση: επειδή το μητρώο A_{11} είναι 2x2, το αντίστροφό του μπορεί να υπολογιστεί απευθείας από τον τύπο: $A^{-1} = \frac{1}{\det(A)} * \begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix}$.

Άσκηση 8 – Ελάχιστα τετράγωνα (Θέμα Φεβρουαρίου 2017)

Δίνονται τα σημεία $x = \{-1, 1, 2, 3\}$ και οι τιμές τους $y = \{2, 5, 9, 26\}$. Να βρείτε τη βέλτιστη προσέγγιση του $\phi(x) = ax + \beta$ με τη μέθοδο των ελαχίστων τετραγώνων.

Λύση

Από θεωρία ελαχίστων τετραγώνων θα έχουμε ότι: $C * a = y$, $a = (\alpha, \beta)^T$, $y = (y_1, \dots, y_n)^T = (2, 5, 9, 26)^T$ και το $x = (-1, 1, 2, 3)^T$. Η πρώτη στήλη του μητρώου C θα περιέχει πάντα «1» και η δεύτερη στήλη θα περιέχει τα σημεία παρεμβολής, δηλαδή το πρώτο σημείο κάθε ζεύγους που δίνεται: (-1, 2), (1, 5), (2, 9), (3, 26). Επίσης το σφάλμα $\|y - Ax_{LS}\|_2$. Στην προκειμένη περίτωση θα έχουμε ότι το μητρώο των συντελεστών θα είναι το $C = \begin{bmatrix} 1 & -1 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix}$ και η λύση θα είναι: $x_{LS} = (C^T C)^{-1} C^T \cdot y = \left(\begin{bmatrix} 1 & 1 & 1 & 1 \\ -1 & 1 & 2 & 3 \end{bmatrix} * \begin{bmatrix} 1 & -1 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} \right)^{-1} * \begin{bmatrix} 1 & 1 & 1 & 1 \\ -1 & 1 & 2 & 3 \end{bmatrix} * \begin{bmatrix} 2 \\ 5 \\ 9 \\ 26 \end{bmatrix} =$

$[3.8571]$. Άρα η βέλτιστη προσέγγιση του πολυωνύμου $\phi(x)$ είναι: $\phi(x) = ax + \beta = 3.8571 * x + 5.3143$. Σε περίπτωση που χρειαστεί να υπολογίσουμε εκτός από το πολυώνυμο παρεμβολής και το σφάλμα της πολυωνυμικής παρεμβολής, θα έχουμε: $\|y - Cx_{LS}\|_2 = \left\| [2, 5, 9, 26]^T - \begin{bmatrix} 1 & -1 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} * \begin{bmatrix} 3.8571 \\ 5.3143 \end{bmatrix} \right\|_2 = \left\| [2, 5, 9, 26]^T - [-1.46, 9.16, 14.47, 19.78]^T \right\|_2 = \dots$

Άσκηση 9 – Παρεμβολή Newton (Θέμα Ιουνίου 2016)

Έστω η συνάρτηση $f(x) = \log_{10}(x)$.

- Να κατασκευάστε την αναπαράσταση Newton του πολυωνύμου παρεμβολής ελάχιστου βαθμού βάσει των τιμών της f στους κόμβους 1, 10. Να ονομάσετε αυτό το πολυώνυμο $p(x)$
- Να κατασκευάστε την αναπαράσταση Newton του πολυωνύμου παρεμβολής ελάχιστου βαθμού βάσει των τιμών της f στους κόμβους 1, 10, 1000. Να ονομάσετε αυτό το πολυώνυμο $q(x)$. Στη συνέχεια να υπολογίσετε το σφάλμα $e(x) = |f(x) - q(x)|$, όταν $x = 100$.

Λύση

- Κατασκευάζουμε τον πίνακα διαιρεμένων διαφορών, ο οποίος φαίνεται παρακάτω:

$$\begin{array}{cc} 1 & 0 \\ 10 & 1 \end{array} \quad \begin{array}{c} > \\ \square \end{array} \quad \frac{1-0}{10-1} = \frac{1}{9}$$

Άρα το πολυώνυμο παρεμβολής είναι: $p(x) = 0 + \frac{1}{9}(x - 1)$.

- Δεν χρειάζεται να κατασκευάσουμε εκ' νέου τον πίνακα διαιρεμένων διαφορών, ο οποίος ουσιαστικά είναι ο προηγούμενος με την προσθήκη του επιπλέον σημείου παρεμβολής 1000 στο τέλος του. Αυτός φαίνεται παρακάτω.

$$\begin{array}{cc} 1 & 0 \\ 10 & 1 \\ 1000 & 3 \end{array} \quad \begin{array}{c} > \\ > \\ > \end{array} \quad \begin{array}{l} \frac{1-0}{10-1} = \frac{1}{9} \\ \frac{3-1}{1000-10} = \frac{2}{990} \end{array} \quad \begin{array}{c} > \\ > \end{array} \quad \frac{\frac{2}{990} - \frac{1}{9}}{1000-1} = \frac{-\frac{108}{990}}{999}$$

Δεν θα πρέπει να ξεχνάμε ότι ένα από τα πλεονεκτήματα της μορφής Newton είναι η **οικονομική ανανέωση**, με την έννοια ότι είναι εφικτή η προσθήκη νέου στοιχείου (x_{n+1}, y_{n+1}) χωρίς αλλαγή των ήδη υπολογισμένων συντελεστών. Άρα το πολυώνυμο παρεμβολής είναι: $q(x) = 0 + \frac{1}{9}(x - 1) - \frac{108}{989010} * (x - 1) * (x - 10)$. Το σφάλμα $e(x) = |f(x) - q(x)|$, όταν $x = 100$ είναι: $|\log_{10}(100) - [\frac{1}{9}(100 - 1) - \frac{108}{989010} * (100 - 1) * (100 - 10)]| = |2 - 11 + \frac{108}{989010} * 99 * 90|$.

4.6.3 Σύγκριση μεθόδων Lagrange – Newton

Πλεονεκτήματα παρεμβολής Newton – Μειονεκτήματα παρεμβολής Lagrange

- Ένα **πλεονέκτημα της μορφής Newton** είναι ότι αν αυτή υλοποιηθεί με τη μέθοδο των διαιρεμένων διαφορών, οδηγεί στον **ταχύτερο υπολογισμό των συντελεστών του πολυωνύμου παρεμβολής**, σε σχέση με τη

μέθοδο Lagrange, η οποία απαιτεί τον προγενέστερο υπολογισμό των πολυωνύμων Lagrange $L_i(x)$ που αποτελεί μια χρονοβόρα διαδικασία. b) Με άλλα λόγια στη μέθοδο Newton έχουμε την **οικονομική ανανέωση**, με την έννοια ότι είναι εφικτή η προσθήκη ενός νέου στοιχείου (x_{n+1}, y_{n+1}) χωρίς αλλαγή των ήδη υπολογισμένων συντελεστών.

Επίσης, ένα ακόμη **μειονέκτημα της μορφής Lagrange** είναι ότι η παρεμβολή σε ένα επιπλέον σημείο, δηλαδή παρεμβολή στα σημεία x_0, x_1, \dots, x_n , απαιτεί τον υπολογισμό νέων πολυωνύμων Lagrange από την αρχή που σε συνδυασμό με το γεγονός ότι ο υπολογισμός της τιμής του πολυωνύμου σε ένα σημείο στη μορφή Lagrange απαιτεί ούτως ή άλλως περισσότερες πράξεις σε σχέση με τις πράξεις τάξης $O(n)$ που απαιτούνται από τη Newton, καθιστούν τη μέθοδο Lagrange ασύμφορη. Διθέντος του x , ο υπολογισμός κάθε όρου απαιτεί $O(n^2)$ πράξεις για τον υπολογισμό του π.π. $L_n(x)$. Αν το n είναι μεγάλο, υπάρχει περίπτωση ο υπολογισμός της δυναμομορφής για κάθε πολυώνυμο Lagrange $L_i(x)$ **να οδηγεί σε σημαντικά σφάλματα**. Κάθε όρος του πολυωνύμου παρεμβολής σε μορφή Lagrange είναι ένα γινόμενο μιας από τις τιμές με ένα πολυώνυμο Lagrange βαθμού n .

Πλεονέκτημα παρεμβολής Lagrange

α) Ένα **πλεονέκτημα της παρεμβολής Lagrange** είναι ότι η μέθοδος αυτή **δεν** χρειάζεται ομοιόμορφα κατανεμημένες τιμές στο x . β) Τα πολυώνυμα Lagrange εξαρτώνται **μόνο** από τους κόμβους και όχι από τις τιμές.

4.7 Μοναδικότητα πολυωνύμου παρεμβολής

Το πολυώνυμο παρεμβολής –ανεξάρτητα με τον τρόπο με τον οποίο υπολογίζεται- είναι **ένα** και **μοναδικό** και αυτό αποδεικνύεται με το επόμενο παράδειγμα. Έστω τα ζεύγη (x_i, y_i) , σημείων (παρεμβολής) και τιμών μιας συνάρτησης y , του παρακάτω πίνακα:

x_i	0	1	2	3	4	
y_i	1	-1	33	211	713	

Μπορεί να υπάρξει άλλο πολυώνυμο παρεμβολής $4^{\text{ου}}$ βαθμού, το οποίο να παρεμβάλλει όλα τα σημεία του παραπάνω πίνακα;

Λύση

Έστω το πολυώνυμο παρεμβολής $4^{\text{ου}}$ βαθμού $\pi_4(x)$, το οποίο περνά από τα παραπάνω σημεία και είναι το εξής: $\alpha_0 + \alpha_1 * x_i + \alpha_2 * x_i^2 + \alpha_3 * x_i^3 + \alpha_4 * x_i^4 = y_i$. Αν το πολυώνυμο αυτό **δεν** είναι μοναδικό, τότε θα υπάρχει άλλο ένα πολυώνυμο έστω: $\beta_0 + \beta_1 * x_i + \beta_2 * x_i^2 + \beta_3 * x_i^3 + \beta_4 * x_i^4 = y_i$, το οποίο επίσης θα παρεμβάλλει τα σημεία του παραπάνω πίνακα. Τότε $\forall x_i \in \{0, 1, 2, 3, 4\}$, τα δύο πολυώνυμα θα έχουν την **ίδια τιμή** (**ίδιο y_i**), άρα θα ισχύει ότι: $\alpha_0 + \alpha_1 * x_i + \alpha_2 * x_i^2 + \alpha_3 * x_i^3 + \alpha_4 * x_i^4 = \beta_0 + \beta_1 * x_i + \beta_2 * x_i^2 + \beta_3 * x_i^3 + \beta_4 * x_i^4 \Rightarrow (\alpha_0 - \beta_0) + (\alpha_1 - \beta_1) * x_i + (\alpha_2 - \beta_2) * x_i^2 + (\alpha_3 - \beta_3) * x_i^3 + (\alpha_4 - \beta_4) * x_i^4 = 0$.

Παρατηρούμε ότι το πολυώνυμο αυτό παίρνει την τιμή "0" και στα 5 σημεία του πίνακα. Συνεπώς δεν μπορεί παρά να είναι το μηδενικό πολυώνυμο (αφού κάθε μη μηδενικό πολυώνυμο έχει τόσες ρίζες όσες είναι ο μέγιστος βαθμός του), επομένως θα ισχύει ότι οι συντελεστές του είναι "0", που σημαίνει ότι $\alpha_i - \beta_i = 0 \Rightarrow \alpha_i = \beta_i \forall i \in \{0, 1, 2, 3, 4\}$. Άρα τα δύο πολυώνυμα είναι **ίσα**, συνεπώς και το **πολυώνυμο παρεμβολής 4^{ου} βαθμού**, το οποίο παρεμβάλει τα σημεία αυτά, είναι **μοναδικό**.

4.8 Βαρυκεντρική παρεμβολή

Να υπολογίσετε το πολυώνυμο σε **βαρυκεντρική αναπαράσταση**, το οποίο παρεμβάλει όλα τα σημεία του παρακάτω πίνακα:

x_i	0	1	2	3	4	
y_i	1	-1	33	211	713	

Λύση

Η βαρυκεντρική αναπαράσταση (παρεμβολή) είναι μια διαφορετική μορφή (παραλλαγή) της παρεμβολής Lagrange και αποδεικνύεται ότι είναι **πιο ευσταθής και γρήγορη**. Ένας τύπος βαρυκεντρικής αναπαράστασης είναι ο εξής:

$$\Pi_n f(x) = \frac{\sum_{j=0}^n \frac{w_j}{(x-x_j)} f_j}{\sum_{j=0}^n \frac{w_j}{(x-x_j)}}, \text{ όπου } w_j = \frac{1}{\prod_{j \neq k} (x_j - x_k)}.$$

Εφαρμόζοντας τον προηγούμενο τύπο υπολογισμού των βαρών w_j , βρίσκουμε τα βαρυκεντρικά βάρη w_j ως εξής (όπου f_j θέτουμε τα y_i του πίνακα):

$$w_0 = \frac{1}{(x_0 - x_1) * (x_0 - x_2) * (x_0 - x_3) * (x_0 - x_4)} = \frac{1}{24}, \quad w_1 = \frac{1}{(x_1 - x_0) * (x_1 - x_2) * (x_1 - x_3) * (x_1 - x_4)} = -\frac{1}{6}$$

$$w_2 = \frac{1}{(x_2 - x_0) * (x_2 - x_1) * (x_2 - x_3) * (x_2 - x_4)} = \frac{1}{4}, \quad w_3 = \frac{1}{(x_3 - x_0) * (x_3 - x_1) * (x_3 - x_2) * (x_3 - x_4)} = -\frac{1}{6}$$

$w_4 = \frac{1}{(x_4 - x_0) * (x_4 - x_1) * (x_4 - x_2) * (x_4 - x_3)} = \frac{1}{24}$. Το πολυώνυμο $\Pi_n f(x)$ σε **βαρυκεντρική μορφή (αναπαράσταση)** θα είναι το εξής: $\Pi_4 f(x) = \frac{\frac{w_0}{x-x_0} f_0 + \frac{w_1}{x-x_1} f_1 + \frac{w_2}{x-x_2} f_2 + \frac{w_3}{x-x_3} f_3 + \frac{w_4}{x-x_4} f_4}{\frac{w_0}{x-x_0} + \frac{w_1}{x-x_1} + \frac{w_2}{x-x_2} + \frac{w_3}{x-x_3} + \frac{w_4}{x-x_4}} = \frac{\frac{1}{24} * 1 + \frac{-1}{6} * (-1) + \frac{1}{4} * 33 + \frac{-1}{6} * 211 + \frac{1}{24} * 713}{\frac{1}{24} + \frac{-1}{6} + \frac{1}{4} + \frac{-1}{6} + \frac{1}{24}} \Rightarrow$

$$\Pi_4(x) = \frac{\frac{1}{24(x-x_0)} + \frac{1}{6(x-x_1)} + \frac{33}{4(x-x_2)} - \frac{211}{6(x-x_3)} + \frac{713}{24(x-x_4)}}{\frac{1}{24(x-x_0)} - \frac{1}{6(x-x_1)} + \frac{1}{4(x-x_2)} - \frac{1}{6(x-x_3)} + \frac{1}{24(x-x_4)}}$$

Ένα πλεονέκτημα της βαρυκεντρικής παρεμ-

βολής είναι ότι **δεν επιβαρυνόμαστε** με τον υπολογισμό του $\Pi_{n+1}(x)$ κάθε φορά που υπολογίζεται το $L_n(x)$.

4.8.1 Πρώτη άσκηση με παρεμβολή Newton και βαρυκεντρική παρεμβολή

α) Έστω η συνάρτηση $f(x) = \log_2(x)$. Να κατασκευάσετε **την αναπαράσταση Newton** του πολυωνύμου παρεμβολής ελάχιστου βαθμού βάσει των τιμών της f στους κόμβους 16, 64. Να ονομάσετε το πολυώνυμο αυτό $p(x)$.

β) Να κατασκευάσετε την **αναπαράσταση Newton** του πολυωνύμου παρεμβολής ελάχιστου βαθμού βάσει των τιμών της f στους κόμβους 16, 64, 128. Να ονομάσετε το πολυώνυμο $q(x)$. Στη συνέχεια, να υπολογίσετε το απόλυτο σφάλμα $e(x) = |f(x) - q(x)|$ όταν $x = 256$.

γ) Να υπολογίσετε τη **βαρυκεντρική αναπαράσταση** του πολυωνύμου παρεμβολής $B(x)$ ελάχιστου βαθμού βάσει των τιμών της f στους κόμβους 16, 64, 128. Στη συνέχεια, να υπολογίσετε το απόλυτο σφάλμα $e(x) = |f(x) - B(x)|$ όταν $x = 256$.

Λύση

α) Για την κατασκευή της αναπαράστασης Newton ο καλύτερος τρόπος είναι η κατασκευή του πίνακα διαιρεμένων διαφορών, ώστε να υπολογιστούν οι συντελεστές της αναπαράστασης. Σε αρκετές περιπτώσεις το ζητούμενο ήταν το σφάλμα της προσέγγισης σε κόμβο που χρησιμοποιήθηκε στην παρεμβολή, επομένως χωρίς πράξεις μπορούσαμε άμεσα να συμπεράνουμε ότι το σφάλμα θα ήταν 0. Ο πίνακας Δ.Δ:

x	$\log_2(x)$	
16	4	
64	6	$\frac{6-4}{64-16} = \frac{1}{24}$
128	7	$\frac{7-6}{128-64} = \frac{1}{64}$

$$\frac{\frac{1}{64} - \frac{1}{24}}{128-16} = -\frac{5}{21504}$$

Επομένως το πολυώνυμο παρεμβολής είναι το $q(x) = 4 + \frac{1}{24} * (x - 16) - \frac{5}{21504} * (x - 16) * (x - 64)$. Αν λάβουμε υπόψη **μόνο** τους κόμβους 16, 64 τότε το πολυώνυμο παρεμβολής θα είναι: $p(x) = 4 + \frac{1}{24}(x - 16)$.

β) Από τον πίνακα διαιρεμένων διαφορών και μερικούς επιπλέον υπολογισμούς για το $q(x)$, φαίνεται ότι $q(x)$

$$= 4 + \frac{1}{24} * (x - 16) - \frac{5}{21504} * (x-16) * (x-64) \text{ και το σφάλμα για } x = 256 \text{ θα είναι: } e(256) = |f(256) - q(256)| = |8 - \frac{23}{7}| = \frac{33}{7} = 4.7143.$$

γ) Αν θέσουμε $w_i = \frac{1}{(x_i - x_0) \dots (x_i - x_{i-1})(x_i - x_{i+1}) \dots (x_i - x_n)}$, έχουμε ότι η βαρυκεντρική αναπαράσταση μπορεί να γραφεί

$$\text{ως } B(x) = \frac{\sum_{i=0}^n f_i \frac{w_i}{(x-x_i)}}{\sum_{i=0}^n \frac{w_i}{(x-x_i)}} = \frac{\left(\frac{1}{48 \times 112(x-16)} - \frac{1}{48 \times 64(x-64)} + \frac{1}{112 \times 64(x-128)} \right)}{\left(\frac{1}{48 \times 112(x-16)} - \frac{1}{48 \times 64(x-64)} + \frac{1}{112 \times 64(x-128)} \right)} = \frac{\left(\frac{1}{5376(x-16)} - \frac{1}{3072(x-64)} + \frac{1}{7168(x-128)} \right)}{\left(\frac{1}{5376(x-16)} - \frac{1}{3072(x-64)} + \frac{1}{7168(x-128)} \right)} = (x-16)(x-64)(x-128) \left(\frac{1}{1344(x-16)} - \frac{1}{512(x-64)} + \frac{1}{1024(x-128)} \right)$$

Σημείωση: Υπενθυμίζουμε ότι επειδή το πολυώνυμο παρεμβολής είναι **μοναδικό**, το σφάλμα $e(256)$ θα είναι **το ίδιο με εκείνο από την αναπαράσταση Newton** που δείξαμε παραπάνω. Επίσης, δεν απαιτείται ο υπολογισμός του πολυωνύμου παρεμβολής με βαρυκεντρική αναπαράσταση, λόγω της μοναδικότητας που χαρακτηρίζει το πολυώνυμο παρεμβολής. Αυτό σημαίνει ότι είναι το ίδιο με το $q(x)$.

4.8.2 Δεύτερη άσκηση με παρεμβολή Newton και βαρυκεντρική παρεμβολή

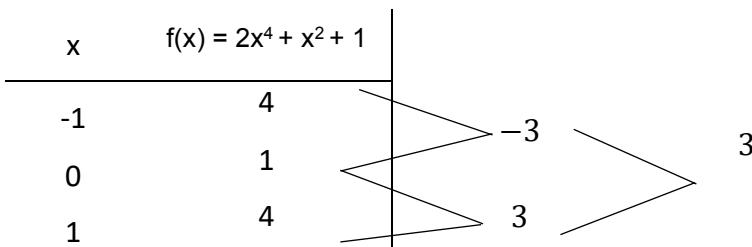
α) Έστω η συνάρτηση $f(x) = 2x^4 + x^2 + 1$. Να κατασκευάσετε την αναπαράσταση Newton του πολυωνύμου παρεμβολής ελάχιστου βαθμού βάσει των τιμών της f στους κόμβους $-1, 0, 1$. Να ονομάσετε το πολυώνυμο αυτό $p(x)$.

β) Να κατασκευάσετε την αναπαράσταση Newton του πολυωνύμου παρεμβολής ελάχιστου βαθμού βάσει των τιμής της f στους κόμβους $-1, 0, 1, 2$. Να ονομάσετε το πολυώνυμο $q(x)$. Στη συνέχεια, να υπολογίσετε το απόλυτο σφάλμα $e(x) = |f(x) - q(x)|$ όταν $x = \frac{1}{2}$.

γ) Να υπολογίσετε τη βαρυκεντρική αναπαράσταση του πολυωνύμου παρεμβολής $B(x)$ ελάχιστου βαθμού βάσει των τιμών της f στους κόμβους $-1, 0, 1$. Στη συνέχεια, να υπολογίσετε το απόλυτο σφάλμα $e(x) = |f(x) - B(x)|$ όταν $x = -1$. **Να υπολογίσετε επίσης το θεωρητικό άνω φράγμα M για το απόλυτο σφάλμα σε οποιοδή-ποτε σημείο στο διάστημα $[-1, 1]$.**

Λύση

α) Για την κατασκευή της αναπαράστασης Newton ο καλύτερος τρόπος είναι η κατασκευή του πίνακα διαιρεμένων διαφορών, ώστε να υπολογιστούν οι συντελεστές της αναπαράστασης. Ο πίνακας διαιρεμένων διαφορών φαίνεται παρακάτω:



Επομένως το πολυώνυμο παρεμβολής είναι το $p(x) = 4 - 3 * (x + 1) + 3 * (x + 1) * x = 3 * x^2 + 1$.

β) Από τον πίνακα διαιρεμένων διαφορών και μερικούς επιπλέον υπολογισμούς για το $q(x)$, φαίνεται ότι

x	$f(x) = 2x^4 + x^2 + 1$			
-1	4		-3	
0	1			3
1	4		3	
2	37		33	15

Επομένως το πολυώνυμο παρεμβολής είναι το $q(x) = 4 - 3 * (x + 1) + 3 * (x + 1) * x + 4 * (x + 1) * x * (x - 1)$. Το σφάλμα $e(x) = |f(x) - q(x)|$ και για $x = \frac{1}{2}$ γίνεται: $e(x) = |f(1/2) - q(1/2)| = |11/8 - 1/4| = |9/8| = 9/8$.

γ) Αρχικά θα υπολογίσουμε τα βάρη $w_0 = \frac{1}{(-1-0)*(-1-1)} = \frac{1}{2}$, $w_1 = \frac{1}{(0+1)*(0-1)} = -1$ και $w_2 = \frac{1}{(1+1)*(1+0)} = \frac{1}{2}$. Άρα το πολυώνυμο παρεμβολής σε βαρυκεντρική αναπαράσταση είναι: $B(x) = \frac{\frac{2}{x+1} - \frac{1}{x} + \frac{2}{x-1}}{\frac{1}{2*(x+1)} - \frac{1}{x} + \frac{1}{2*(x-1)}} = p(x) = 3x^2 + 1$.

Στη συνέχεια θα υπολογίσουμε το απόλυτο σφάλμα $e(x) = |f(x) - B(x)|$ όταν $x = -1$, που είναι $e(x) = 4 - 4 = 0$.

Γενικότερα όταν υπολογίζουμε σφάλματα παρεμβολής ακριβώς στα σημεία παρεμβολής και όχι ενδιάμεσα, τότε το σφάλμα θα είναι ίσο με 0. Για τον υπολογισμό του θεωρητικού άνω φράγματος M χρησιμοποιούμε τον

$$\text{τύπο: } \max|e_n f(x)| \leq \frac{\max_{x \in \Omega} |f^{(n+1)}(x)|}{4(n+1)} \left(\frac{x_n - x_0}{n}\right)^{(n+1)} \Rightarrow \max|e_n f(x)| \leq \frac{\max_{x \in \Omega} |f^{(n+1)}(x)|}{4(n+1)} h^{(n+1)} \text{ όπου } h = \frac{x_n - x_0}{n} = \frac{1 - (-1)}{2} = 1.$$

Άρα $\max|e_n f(x)| \leq \max \frac{|48x|}{12} = \frac{48}{12} * \max|x| = 4$, όπου $x \in [-1, 1]$ και $f^{(3)}(x) = 48x$. Επειδή δίνονται 3 σημεία παρεμβολής, ο βαθμός του πολυωνύμου είναι $n = 2$ και το κλάσμα $\frac{x_n - x_0}{n} = \frac{2}{2} = 1$.

4.9 Παρεμβολή Hermite

Στην παρεμβολή Hermite δίνονται **μαζί** με τα σημεία (κόμβους) παρεμβολής, οι παράγωγοι της συνάρτησης $f(x)$ στα σημεία αυτά. Για παράδειγμα, για τη συνάρτηση $f(x) = x^8 + 1$ που ακολουθεί, ζητάμε να υπολογίσουμε το πολυώνυμο παρεμβολής στη μορφή Hermite, όταν δίνεται ο επόμενος πίνακας:

x	f(x)	f'(x)	f''(x)
-1	2	-8	56
0	1	0	0
1	2	8	56

Λύση

Στην παρεμβολή Newton δεν είναι απαραίτητο να δίνονται για **κάθε σημείο παρεμβολής** οι παράγωγοι της συνάρτησης, κάτι που είναι απαραίτητο στην παρεμβολή Hermite. Κατασκευάζουμε τον **πίνακα διαιρεμένων διαφορών**, όπως φαίνεται παρακάτω, με βάση τη γνωστή μεθοδολογία, όπου κάθε τιμή επαναλαμβάνεται τόσες φορές όσες είναι η τάξη της παραγώγου και γνωρίζουμε εκ' των προτέρων ότι θα προκύψει το πηλίκο 0/0. Στο τέλος το **πολυώνυμο παρεμβολής δημιουργείται** με βάση την παρεμβολή Newton.

$$\begin{array}{ll}
 z_0 = -1 & f[z_0] = 2 \\
 & \frac{f'(z_0)}{1} = -8 \\
 z_1 = -1 & f[z_1] = 2 \quad \frac{f''(z_1)}{2} = 28 \\
 & \frac{f'(z_1)}{1} = -8 \quad f[z_3, z_2, z_1, z_0] = -21 \\
 z_2 = -1 & f[z_2] = 2 \quad f[z_3, z_2, z_1] = 7 \quad 15 \\
 & f[z_3, z_2] = -1 \quad f[z_4, z_3, z_2] = 1 \quad f[z_4, z_3, z_2, z_1] = -6 \quad -10 \\
 z_3 = 0 & f[z_3] = 1 \quad f[z_4, z_3, z_2] = 1 \quad f[z_5, z_4, z_3, z_2] = -1 \quad 5 \quad 4 \\
 & \frac{f'(z_3)}{1} = 0 \quad f''(z_3) = 0 \quad f[z_5, z_4, z_3, z_2] = 1 \quad -2 \quad -1 \\
 z_4 = 0 & f[z_4] = 1 \quad f''(z_4) = 0 \quad f[z_6, z_5, z_4, z_3] = 1 \quad 1 \quad 2 \quad 1 \\
 & \frac{f'(z_4)}{1} = 0 \quad f[z_6, z_5, z_4] = 1 \quad f[z_7, z_6, z_5, z_4] = 6 \quad 2 \quad 1 \\
 z_5 = 0 & f[z_5] = 1 \quad f[z_6, z_5, z_4] = 1 \quad f[z_7, z_6, z_5, z_4] = 6 \quad 5 \quad 4 \quad 10 \\
 & f[z_6, z_5] = 1 \quad f[z_7, z_6, z_5] = 7 \quad f[z_8, z_7, z_6, z_5] = 21 \\
 z_6 = 1 & f[z_6] = 2 \quad \frac{f'(z_6)}{1} = 8 \quad f''(z_6) = 28 \\
 & f[z_7] = 2 \quad \frac{f'(z_7)}{1} = 8 \\
 z_7 = 1 & f[z_7] = 2 \quad \frac{f'(z_7)}{1} = 8 \\
 z_8 = 1 & f[z_8] = 2
 \end{array}$$

Το πολυώνυμο παρεμβολής στη μορφή Hermite είναι το εξής: $P_8(x) = f(z_0) + f(z_0, z_1) * (z - z_0) + f(z_0, z_1, z_2) * (z - z_0) * (z - z_1) + f(z_0, z_1, z_2, z_3) * (z - z_0) * (z - z_1) * (z - z_2) + \dots = 2 - 8 * (x+1) + 28 * (x+1)^2 - 21 * (x+1)^3 + 15 * (x - 0) * (x+1)^3 - 10 * x^2 * (x+1)^3 + 4 * x^3 * (x+1)^3 - 1 * x^3 * (x+1)^3 * (x - 1) + 1 * x^3 * (x+1)^3 * (x-1)^2.$

Είδη πολυωνυμικής παρεμβολής

Ολοκληρώνοντας την **πολυωνυμική παρεμβολή**, αναφέρουμε ότι οι διαφορετικές μορφές (διαφορετικά είδη) των πολυωνύμων παρεμβολής που μπορούμε να έχουμε, είναι τα εξής:

- Πολυώνυμο παρεμβολής στη μορφή **Lagrange**.
- Πολυώνυμο παρεμβολής στη μορφή **Newton**.
- Πολυώνυμο παρεμβολής **ελαχίστων τετραγώνων**.
- Πολυώνυμο παρεμβολής με **βαρυκεντρική αναπαράσταση**.
- Πολυώνυμο παρεμβολής στη μορφή **Hermite**.

4.10 Τμηματική παρεμβολή – Τμηματικά πολυώνυμα

Μέχρι τώρα ασχοληθήκαμε με την κατασκευή ενός πολυωνύμου που παίρνει συγκεκριμένες τιμές στους κόμβους. Στην **πολυωνυμική παρεμβολή** που εξετάσαμε έως τώρα -με τη μορφή Lagrange, Newton, Hermite και βαρυκεντρική- ο βαθμός του πολυωνύμου **εξαρτάται άμεσα από το πλήθος των σημείων (παρεμβολής)** και στη γενική περίπτωση είναι βαθμού **n για n + 1 σημεία**. Μπορούμε να αποφύγουμε τα συγκεκριμένα προβλήματα αν χρησιμοποιήσουμε πιο ευέλικτες συναρτήσεις ή αν αλλάξουμε τους όρους που επιβάλλονται από την παρεμβολή. **Χωρίζουμε το διάστημα παρεμβολής σε υποδιαστήματα (τμήματα)** και σε κάθε **υποδιάστημα (τμήμα)** κατασκευάζουμε ένα διαφορετικό πολυώνυμο. Υπάρχουν τρεις προσεγγίσεις με **τμηματικά πολυώνυμα**:

- a) **Τμηματική γραμμική παρεμβολή:** **μειονέκτημα** ότι υπάρχει **ασυνέχεια** στις παραγώγους.
- b) **Τμηματικά πολυώνυμα τύπου Hermite:** **μειονέκτημα** ότι απαιτείται να είναι **γνωστές** οι **τιμές της παραγώγου της συνάρτησης**, την οποία επιθυμούμε να προσεγγίσουμε.
- c) **Splines:** η συνάρτηση Spline συμβολίζεται με το **S** και επιλέγεται με τέτοιο τρόπο ώστε να είναι **και δύο φορές συνεχώς παραγωγίσιμη** στο διάστημα παρεμβολής. Η κυβική Spline συνάρτηση παρεμβολής για την $f(x)$ έχει την μορφή: $S_i(x) = a_i + b_i(x - x_i) + c_i(x - x_i)^2 + d_i(x - x_i)^3$, για $i = 0, 1, \dots, n-1$, όπου $a_i = f(x_i) = f_i$, $b_i = \frac{1}{h_i}(f_{i+1} - f_i) - \frac{h_i}{3}(2c_i + c_{i+1})$, όπου $h_i = x_{i+1} - x_i$, $d_i = \frac{1}{3h_i}(c_{i+1} - c_i)$ και οι συντελεστές c_i προκύπτουν από την επίλυση του γραμμικού συστήματος $A * c = q$, όπου το μητρώο A και το διάνυσμα q δίνεται από τους τύπους:

Computer - Ανάλυση

$$A = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ h_0 & 2(h_0 + h_1) & h_1 & \dots & 0 \\ 0 & h_1 & 2(h_1 + h_2) & \dots & h_{n-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{bmatrix} \text{ και } q = \begin{bmatrix} 0 \\ \frac{3}{h_1}(f_2 - f_1) - \frac{3}{h_0}(f_1 - f_0) \\ \vdots \\ \frac{3}{h_{n-1}}(f_n - f_{n-1}) - \frac{3}{h_0}(f_{n-1} - f_{n-2}) \\ 0 \end{bmatrix}. \text{ Στο διάνυσμα } q \text{ το πρώτο και το τελευταίο στοιχείο είναι } 0.$$

4.10.1 Άσκηση εφαρμογής της κυβικής spline

Έστω $f(0) = 0, f(1) = 2, f(2) = 1, f(3) = 0$, δηλαδή έχουμε 4 κόμβους (παρεμβολής) 0, 1, 2, 3 και σε κάθε τμήμα μεταξύ των κόμβων κατασκευάζουμε και μια spline. Να υπολογιστούν τα κυβικά πολυώνυμα $S_0(x), S_1(x)$ και $S_2(x)$.

	0	1	2	3	
Λύση	$S_0(x)$	$S_1(x)$	$S_2(x)$		
Ξεκινάμε με την επίλυση του γραμμικού συστήματος, όπου:	$A * c = q$, $A =$				
$\begin{bmatrix} 1 & 0 & 0 & 0 \\ h_0 & 2(h_0 + h_1) & h_1 & 0 \\ 0 & h_1 & 2(h_1 + h_2) & h_2 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 4 & 1 & 0 \\ 0 & 1 & 4 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}$, διότι $h_0 = x_1 - x_0 = 1 - 0 = 1, h_1 = x_2 - x_1 = 2 - 1 = 1,$				
$2 * (h_0 + h_1) = 2 * 2 = 4, h_2 = x_3 - x_2 = 3 - 2 = 1, 2(h_1 + h_2) = 2 * 2 = 4.$	To διάνυσμα $q =$				
$\begin{bmatrix} 0 \\ \frac{3}{h_1}(f_2 - f_1) - \frac{3}{h_0}(f_1 - f_0) \\ 0 \\ \frac{3}{h_2}(f_3 - f_2) - \frac{3}{h_1}(f_2 - f_1) \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ -9 \\ 0 \\ 0 \end{bmatrix} \Rightarrow A * c = q \Rightarrow \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 4 & 1 & 0 \\ 0 & 1 & 4 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix} * \begin{bmatrix} c_0 \\ c_1 \\ c_2 \\ c_3 \end{bmatrix} = \begin{bmatrix} 0 \\ -9 \\ 0 \\ 0 \end{bmatrix} \Rightarrow \begin{bmatrix} c_0 \\ c_0 + 4 * c_1 + c_2 \\ c_1 + 4 * c_2 + c_3 \\ c_3 \end{bmatrix} = \begin{bmatrix} 0 \\ -9 \\ 0 \\ 0 \end{bmatrix}.$					

Εξισώνοντας τα αντίστοιχα στοιχεία των δύο διανυσμάτων μεταξύ τους, έχουμε ότι: $c_0 = 0, c_0 + 4c_1 + c_2 = -9$ (1), $c_1 + 4c_2 + c_3 = 0$ (2) και $c_3 = 0$. Από την επίλυση των εξισώσεων (1) και (2) έχουμε ότι: $c_1 = -\frac{12}{5}$ και $c_2 = \frac{9}{15}$. Για $i = 0$ έχουμε: $S_0(x) = a_0 + b_0(x - x_0) + c_0(x - x_0)^2 + d_0(x - x_0)^3$, όπου οι συντελεστές $a_0 = f_0 = 0$ και $b_0 = \frac{1}{h_0}(f_1 - f_0) - \frac{h_0}{3}(2c_0 + c_1) = 1 * (2 - 0) - \frac{1}{3}(2 * 0 + \frac{12}{5}) = 2 + \frac{12}{15} = \frac{14}{5}$ και $d_0 = \frac{1}{3h_0} * (c_1 - c_0) = \frac{1}{3} * (-\frac{12}{5}) = -\frac{4}{5}$.

Οπότε το κυβικό πολυώνυμο $S_0(x) = \frac{6}{5} * (x - x_0) + \frac{4}{5} * (x - x_0)^3 = \frac{14}{5}x - \frac{4}{5}x^3$. Με ανάλογο τρόπο υπολογίζουμε τα πολυώνυμα S_1 και S_2 .

4.10.2 Άσκηση με τμηματική γραμμική παρεμβολή

Αν $AM_4 AM_3 AM_2 AM_1$ είναι τα τέσσερα τελευταία ψηφιά του αριθμού μητρώου σας, δίνεται η συνάρτηση $f(x) = a_3*x^3 - a_2*x^2 - a_1*x^1 - a_0$, όπου $a_3 = (AM_4 \text{ mod } 4) + 1, a_2 = (AM_3 \text{ mod } 4) + 1, a_1 = (AM_2 \text{ mod } 4) + 1, a_0 = (AM_1 \text{ mod } 4) + 1$. Συμπληρώστε καταρχήν τη συνάρτηση $f(x)$ για το ΑΜ σας. Στη συνέχεια:

- a) Να κατασκευάσετε το πολυώνυμο παρεμβολής **δευτέρου βαθμού** στη μορφή P_x , επιλέγοντας όσους κόμβους παρεμβολής από το σύνολο $\{-1, 0, 2, 3\}$ χρειάζεται και τις αντίστοιχες τιμές της f .

- b) Αν από τους κόμβους που επιλέξατε, ονομάσετε την ελάχιστη τιμή x_1 και τη μέγιστη x_2 , να υπολογίσετε ένα καλό άνω φράγμα για το $\max_{x \in [x_1, x_2]} |f(x) - p(x)|$.
- c) Να βρείτε συνάρτηση $\tau_1(x)$ η οποία να επιτελεί **τμηματική γραμμική παρεμβολή** της συνάρτησης f στο τμήμα $[x_1, x_2]$, με τους κόμβους που χρησιμοποιήσατε στο προηγούμενο ερώτημα.
- d) Να υπολογίσετε την τιμή της συνάρτησης $\tau_1(x)$ στα σημεία $\left\{0, x_1 + \frac{(x_2 - x_1)}{6}\right\}$. Αφού υπολογίσετε τις τιμές της συνάρτησης f σε αυτά τα δύο σημεία, **να βρείτε το αντίστοιχο σφάλμα σε αυτά τα σημεία**.
- e) Να κυκλώσετε τη συνάρτηση για την οποία αν υπολογιστεί πολυώνυμο παρεμβολής P_n στο διάστημα $[-1, 1]$ σε ισαπέχοντες κόμβους για μεγάλες τιμές του n , η προσέγγιση της συνάρτησης με τις τιμές του P_n δε θα είναι ικανοποιητική σε όλο το διάστημα.

- α) $\frac{2}{1+x^2}$
 β) e^{-x^2}
 γ) $(\sin(x))^8 + \cos(x)$

Λύση

a) Έστω $AM = 3543 \Rightarrow f(x) = 4 * x^3 - 2 * x^2 - x - 4$, διότι τα ψηφία είναι $a_3 = 4$, $a_2 = 2$, $a_1 = 1$, $a_0 = 4$. Επειδή το πολυώνυμο παρεμβολής είναι δευτέρου βαθμού, επιλέγουμε 3 σημεία και **συγκεκριμένα (αυθαίρετα) τα τρία πρώτα (-1, 0, 2)**. Τότε το πολυώνυμο παρεμβολής στη μορφή Lagrange θα είναι το εξής: $P_2(x) = L_0 * f(x_0) + L_1 * f(x_1) + L_2 * f(x_2) = L_0 * (-9) + L_1 * (-4) + L_2 * 18 = \frac{(x-0)(x-2)}{(-1-0)(-1-2)} * (-9) + \frac{(x+1)(x-2)}{(0+1)(0-2)} * (-4) + \frac{(x+1)(x-0)}{(2+1)(2-0)} * 18 = \frac{x(x-2)}{3}(-9) + \frac{(x+1)(x-2)}{-2}(-4) + \frac{(x-1)x}{6} * 18 = -3x * (x-2) + 2 * (x+1) * (x-2) + 3 * (x+1) * x = -3x^2 + 6x + 2x^2 - 4x + 2x - 4 + 3x^2 + 3x = 2 * x^2 + 7 * x - 4$.

b) $\max_{x \in [x_1, x_2]} |f(x) - p(x)| \leq \max_{x \in [x_1, x_2]} \frac{|f(x)|}{4(n+1)} * \left(\frac{x_n - x_0}{n}\right)^{(n+1)} = \frac{|f(x)|}{4*3} * \left(\frac{x_2 - x_1}{2}\right)^3 = 2 * \frac{3^3}{2^3} = \frac{3^3}{2^2} = \frac{27}{4}$.

c) Δίνονται τα ζεύγη κόμβων – τιμών $\{(x_i, y_i)\}_{i=0}^n$ όπου τα σημεία παρεμβολής δίνονται με αύξουσα σειρά $x_0 < x_1 < \dots < x_n$ και για την **τμηματική γραμμική παρεμβολή** ορίζεται το **τμηματικό γραμμικό πολυώνυμο παρεμβολής** ως εξής: $\Pi_1^H f(x) = f(x_i) + \frac{f(x_{i+1}) - f(x_i)}{x_{i+1} - x_i} * (x - x_i), \forall x \in [x_i, x_{i+1}]$. Αξίζει να σημειωθεί ότι ο δείκτης του συγκεκριμένου πολυωνύμου - που εκφράζει το βαθμό του - είναι «1», διότι μιλάμε για ευθεία γραμμή. Για το διάστημα $[x_1, x_2] = [-1, 2]$ που προκύπτει από τους κόμβους $[-1, 0, 2]$ του προηγούμενου ερωτήματος, θα έχουμε **ότι αν θέλαμε να υπολογίσουμε για λόγους δοκιμής την τμηματική γραμμική παρεμβολή σε κάθε επιμέρους διάστημα $(-1, 0)$ και $(0, 2)$ χωριστά**, τότε τα τμηματικά γραμμικά πολυώνυμα παρεμβολής θα ήταν τα εξής:

Διάστημα (-1, 0): $f(-1) + \frac{f(0) - f(-1)}{0 - (-1)} * (x + 1) = -9 + \frac{-4 + 9}{1} * (x + 1) = -9 + 5(x + 1) = -9 + 5x + 5 = 5 * x - 4$.

Διάστημα (0, 2): $f(0) + \frac{f(2) - f(0)}{2 - 0} * (x - 0) = -4 + \frac{18 + 4}{2} * x = -4 + 11 * x = 11 * x - 4$.

Επειδή όμως ζητάμε στην προκειμένη περίπτωση το τμηματικό γραμμικό πολυώνυμο παρεμβολής $\tau_1(x)$ που επιτελεί τμηματική γραμμική παρεμβολή στο διάστημα $[x_1, x_2] = [-1, 2]$, θα έχουμε ότι: $\tau_1(x) = f(x_1) + \frac{f(x_2) - f(x_1)}{x_2 - x_1} * (x - x_1)$

$$= f(-1) + \frac{f(2) - f(-1)}{2 - (-1)}(x - (-1)) = -9 + \frac{18 - (-9)}{2 - (-1)} * (x + 1) = -9 + \frac{27}{3} * (x + 1) = -9 + 9 * (x + 1) = -9 + 9 * x + 9 = 9 * x = \tau_1(x).$$

d) Οι τιμές της συνάρτησης $\tau_1(x) = 9 * x$ στα σημεία $\{0, x_1 + \frac{(x_2 - x_1)}{6}\}$ θα είναι οι εξής: Αρχικά στο σημείο 0, θα είναι 0 και μετά στο σημείο $x_1 + \frac{(x_2 - x_1)}{6} = (-1) + \frac{2 - (-1)}{6} = -1 + \frac{3}{6} = -1 + \frac{1}{2} = -\frac{1}{2}$ θα είναι $9 * (-\frac{1}{2}) = -4.5$. Επομένως, οι τιμές του τμηματικού γραμμικού πολυωνύμου παρεμβολής $\tau_1(x)$ είναι $\{0, -4.5\}$. Οι τιμές της συνάρτησης $f(x) = 4 * x^3 - 2 * x^2 - x - 4$ στο σημείο 0 είναι $f(0) = 4 * (0)^3 - 2 * (0)^2 - 0 - 4 = -4$ και στο σημείο $-\frac{1}{2}$ είναι $f(-\frac{1}{2}) = 4(-\frac{1}{2})^3 - 2(-\frac{1}{2})^2 - (-\frac{1}{2}) - 4 = -5$. Άρα, οι τιμές της συνάρτησης $f(x)$ είναι $\{-4, -5\}$.

Ο τύπος που δίνει το **σφάλμα της τμηματικής γραμμικής παρεμβολής** είναι: $\max|f(x) - \Pi_1^H f(x)| \leq \max_{x \in \Omega} \frac{|f^{(2)}(x)|}{8} H^2$ όπου $H = \text{μέγιστο μήκος των υποδιαστημάτων } x_i - x_{i-1}$. Το πλεονέκτημα που προσφέρει η τμηματική γραμμική παρεμβολή είναι η **βελτίωση στην προσέγγιση που επιτυγχάνεται μέσω αύξησης των κόμβων παρεμβολής**, ενώ ο βαθμός του πολυωνύμου σε κάθε υποδιάστημα παραμένει γραμμικός. Άρα, για τα σημεία 0 και $-1/2$, το $H = 0 - (-\frac{1}{2}) = \frac{1}{2}$ και η δεύτερη παράγωγος: $f^{(2)}(x) = 24 * x - 4$, όπου $f^{(2)}(0) = -4$ και $f^{(2)}(-\frac{1}{2}) = 24(-\frac{1}{2}) - 4 = -16$. Άρα, για το σημείο 0 το σφάλμα είναι: $\max|f(0) - \Pi^H f(0)| \leq \max \frac{|f^{(2)}(x)|}{8} * (\frac{1}{2})^2 = \frac{4}{8} * \frac{1}{4} = \frac{1}{8}$ και για το σημείο $-\frac{1}{2}$ είναι $\max|f(-1/2) - \Pi^H f(-1/2)| \leq \max_{x \in \Omega} \frac{|f^{(2)}(x)|}{8} H^2 = \frac{|-16|}{8} * (\frac{1}{2})^2 = \frac{1}{2}$.

e) Η σωστή απάντηση είναι η **πρώτη** και αυτό αιτιολογείται από το **φαινόμενο Runge**, το οποίο μελετά την αστάθεια της πολυωνυμικής παρεμβολής στο διάστημα $[-1, 1]$ για τη συνάρτηση $f(x) = \frac{1}{1+25x^2}$. Η συνάρτηση της πρώτης απάντησης είναι παρόμοιας μορφής με την $f(x)$ που μελετάται από το φαινόμενο Runge.

Επισήμανση 1: Είναι διαφορετικοί οι τύποι για τα **άνω φράγματα σφάλματων** στην **πολυωνυμική παρεμβολή** και στην **τμηματική γραμμική παρεμβολή**. Οι τύποι αυτοί είναι αντίστοιχα οι εξής:

$$\max_{x \in [x_1, x_2]} |f(x) - p(x)| \leq \max_{x \in [x_1, x_2]} \frac{|f^{(n+1)}(x)|}{4(n+1)} * \frac{(x_n - x_0)^{(n+1)}}{n} \quad \text{και} \quad \max|f(x) - \Pi_1^H f(x)| \leq \max_{x \in \Omega} \frac{|f^{(2)}(x)|}{8} H^2$$

Επισήμανση 2: Για τα ζεύγη κόμβων (x_i, y_i) , όπου $x_0 < x_1 < \dots < x_n$ το **τμηματικό γραμμικό πολυώνυμο παρεμβολής** δίνεται από τον τύπο: $\Pi_1^H f(x) = f(x_i) + \frac{f(x_{i+1}) - f(x_i)}{x_{i+1} - x_i} * (x - x_i), \forall x \in [x_i, x_{i+1}]$.

4.10.3 Άσκηση με τμηματική γραμμική παρεμβολή

Έστω ότι έχουμε τη δυνατότητα να υπολογίσουμε την τιμή της $f(x) = \exp(-2x) = e^{-2x}$ σε ένα σημείο $\alpha \in (0, 2)$. Να επιλέξτε τη **μεγαλύτερη δυνατή τιμή του α** έτσι ώστε η **γραμμική παρεμβολή** βασισμένη στο $f(0) = 1$ και $f(\alpha)$ να προσεγγίζει το $f(x)$ για οποιοδήποτε $x \in (0, \alpha)$ με **σφάλμα φραγμένο από 10^{-6}** .

a) $\beta \approx \sqrt{2} * 10^{-3}$ β) σε όλες τις περιπτώσεις που αναφέρονται το σφάλμα της προσέγγισης θα είναι μεγαλύτερο του 10^{-6} c) $\beta \approx x \frac{10^{-3}}{\sqrt{2}}$ και d) $\beta \approx \exp(-1)$.

Λύση

Η συνάρτηση $f(x) = \exp(-2x) = e^{-2x}$ για $x \in (0, \alpha)$, η $f'(x) = -2e^{-2x}$ και η $f''(x) = f^{(2)}(x) = 4e^{-2x}$. Υπενθυμίζεται ότι $(e^x)' = e^x$. Από την εκφώνηση της άσκησης καταλαβαίνουμε ότι θα χρησιμοποιήσουμε **γραμμική παρεμβολή** (προφανώς αναφερόμαστε στην **τμηματική γραμμική παρεμβολή**) που περιγράφεται με τον τύπο: $\Pi_1^H f(x) = f_{(x_i)} +$

$\frac{f(x_{i+1}) - f(x_i)}{x_{i+1} - x_i} * (x - x_i), \forall x \in [x_i, x_{i+1}]$. Το σφάλμα της περιγράφεται από τον τύπο $\max|f(x) - \Pi_1^H f(x)| \leq$

$\max_{x \in \Omega} \frac{|f^{(2)}(x)|}{8} H^2$). Επομένως, χρησιμοποιώντας τον τύπο που δίνει το μέγιστο άνω φράγμα για την **τμηματική γραμμική παρεμβολή**, θα έχουμε ότι: $\max|f(x) - \Pi_1^H f(x)| \leq \max_{x \in \Omega} \frac{|f^{(2)}(x)|}{8} H^2 \leq 10^{-6}$, διότι αυτό δίνεται από

την εκφώνηση ως άνω φράγμα σφάλματος της τμηματικής γραμμικής παρεμβολής, όπου **H = το μέγιστο μήκος του υποδιαστήματος $x_i - x_{i-1} = \alpha - 0 = \alpha$** . Επομένως, αντικαθιστώντας στην προηγούμενη ανισότητα τη δεύτερη

παράγωγο $f^{(2)}(x)$ και το **H** που υπολογίσαμε, θα έχουμε ότι: $\max_{x \in \Omega} \frac{|4e^{-2x}|}{8} (\alpha - 0)^2 \leq 10^{-6} \Rightarrow \max_{x \in \Omega} \frac{e^{-2x}}{2} \alpha^2 \leq 10^{-6} \Rightarrow \frac{e^{-2*\alpha}}{2} \alpha^2 \leq 10^{-6} \Rightarrow 1 * \alpha^2 \leq 2 * 10^{-6} \Rightarrow \alpha \leq \sqrt{2 * 10^{-6}} = \sqrt{2} * \sqrt{10^{-6}} \Rightarrow \alpha \leq \sqrt{2} * 10^{-3}$. Επομένως η σωστή απάντηση είναι η (α) .

Σημείωση: Θέτουμε **στο x την τιμή 0**, διότι ζητάμε τη μέγιστη τιμή της δεύτερης παραγώγου της $f(x)$ και επειδή αυτή υψώνεται σε αρνητική δύναμη, μεγιστοποιείται για τη μικρότερη τιμή του x στο διάστημα $x \in (0, \alpha)$.

4.10.4 Άσκηση με πολυώνυμα παρεμβολής (παλιό θέμα)

Αν δοθούν 5 ζεύγη σημείων (x_i, y_i) , $i = 1, 2, 3, 4, 5$ και **η ακολουθία των πραγματικών αριθμών x_i είναι αύξουσα**: a) υπάρχει ακριβώς ένα πολυώνυμο p βαθμού 5, τέτοιο ώστε $p(x_i) = y_i$, για $i = 1, \dots, 5$ β) υπάρχει ακριβώς ένα πολυώνυμο p βαθμού 3, τέτοιο ώστε $p(x_i) = y_i$, για $i = 1, \dots, 5$ γ) υπάρχει **τμηματικό γραμμικό πολυώνυμο** p τέτοιο ώστε $p(x_i) = y_i$, για $i = 1, \dots, 5$ δ) τίποτα από όσα αναφέρονται.

Λύση

Η σωστή απάντηση είναι η (γ) , επειδή γνωρίζουμε με βεβαιότητα ότι έχουμε **αύξουσα ακολουθία των κόμβων παρεμβολής που δίνονται**, δηλαδή είμαστε βέβαιοι ότι ισχύει η ανισότητα $x_1 < x_2 < x_3 < x_4 < x_5$ κ.λ.π. **Από την εκφώνηση δίνεται ότι η ακολουθία των πραγματικών αριθμών x_i (δηλ. των κόμβων ή σημείων παρεμβολής) είναι αύξουσα**. Προσοχή! Για να οριστεί ένα γραμμικό τμηματικό πολυώνυμο παρεμβολής, θα πρέπει πάντα στα ζεύγη κόμβων – τιμών $\{(x_i, y_i)\}_{i=0}^n$ που δίνονται, τα σημεία παρεμβολής να είναι με αύξουσα σειρά, δηλ. $x_0 < x_1 < \dots < x_n$, κάτι που προφανώς ισχύει στην προκειμένη περίπτωση.

Σημείωση: Αν η εκφώνηση της άσκησης δεν ανέφερε ότι και η ακολουθία των πραγματικών αριθμών x_i είναι αύξουσα, τότε η σωστή απάντηση θα ήταν η (δ) .

4.10.5 Άσκηση με πολυωνυμική παρεμβολή (Φεβρουάριος 2017)

Από τις παρακάτω επιλογές ποια είναι η **μεγαλύτερη απόσταση** που μπορεί να χρησιμοποιηθεί σε ένα πίνακα τιμών για την $f(x) = \cos(x)$ ώστε με **γραμμική παρεμβολή** μεταξύ δύο διαδοχικών κόμβων και αντίστοιχων τιμών, να προσεγγίζεται η τιμή της $f(x)$ στο αντίστοιχο διάστημα με **μέγιστο απόλυτο σφάλμα που δεν υπερβαίνει το 10^{-9}** .

- a) 10^{-1} β) 10^{-3} γ) 10^{-4} δ) 10^{-5}

Λύση

Έχουμε και πάλι τιμηματική γραμμική παρεμβολή, όπου μας δίνεται το μέγιστο άνω φράγμα του σφάλματος. Ο τύπος που δίνει το **μέγιστο άνω φράγμα για την τιμηματική γραμμική παρεμβολή** είναι ο εξής: $\max |f(x) - \Pi_1^H f(x)| \leq \max_{x \in \Omega} \frac{|f''(x)|}{8} H^2 \leq 10^{-9}$. Η συνάρτηση $f(x) = \cos(x)$ οπότε η $f'(x) = -\sin(x)$ και η $f''(x) = -\cos(x)$, οπότε $\max |f(x) - \Pi_1^H f(x)| \leq \max_{x \in \Omega} \frac{|f''(x)|}{8} H^2 \leq 10^{-9} \Rightarrow \max \frac{|\cos(x)|}{8} H^2 \leq 10^{-9} \Rightarrow \frac{1}{8} H^2 \leq 10^{-9} \Rightarrow H^2 \leq 10^{-9} * 8 \Rightarrow H \leq \sqrt{10^{-9} * 8} = \sqrt{10^{-8} * 10^{-1} * 8} = \sqrt{0.8} * 10^{-4} = 0.894 * 10^{-4}$. Άρα η σωστή απάντηση είναι η (γ).

Σημείωση: Θα πρέπει να σημειωθεί ότι στον υπολογισμό του άνω φράγματος της τιμηματικής γραμμικής παρεμβολής έχουμε **θεωρήσει ότι $|\pm \cos(x)| < 1$** . Επίσης, ισχύει ότι **$|\pm \sin(x)| < 1$**

4.10.6 Παράδειγμα πολυωνυμικής και γραμμικής παρεμβολής (παλιό θέμα)

- Να βρεθούν δύο σημεία για την προσέγγιση του $\log_2 100$ με **γραμμική παρεμβολή**.
- Να βρεθούν τρία σημεία για την προσέγγιση του $\log_2 100$ με **πολυωνυμική παρεμβολή**.

Λύση

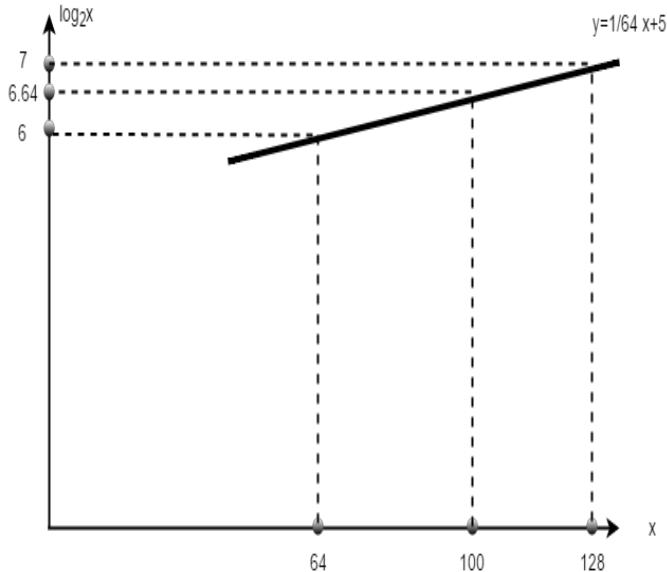
i) **A' Τρόπος:** Για να προσεγγίσουμε το $\log_2 100$ με **γραμμική παρεμβολή**, παίρνουμε την $f(x) = a_0 + a_1 x$ που είναι η εξίσωση της **ευθείας**, και αυτό διότι έχουμε γραμμική παρεμβολή. Παίρνουμε δύο σημεία με τιμές πάνω και κάτω από το 100, π.χ. για $x = 128 = 2^7$ και για $x = 64 = 2^6$ έχουμε τις εξισώσεις:

$$\begin{cases} \log_2 x = 7 \Rightarrow x = 128, \text{ διότι } \log_2 2^7 = 7 * \log_2 2 = 7 \\ \log_2 x = 6 \Rightarrow x = 64, \text{ διότι } \log_2 2^6 = 6 * \log_2 2 = 6 \end{cases}$$

Λύνουμε ένα γραμμικό σύστημα της μορφής $A^* x = b$, δηλαδή $\begin{bmatrix} 1 & 64 \\ 1 & 128 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \end{bmatrix} = \begin{bmatrix} 6 \\ 7 \end{bmatrix}$ προκειμένου να υπολογίσουμε τους συντελεστές a_0 και a_1 της ευθείας $f(x)$ και έχουμε: $\begin{cases} a_0 + 64a_1 = 6 \\ a_0 + 128a_1 = 7 \end{cases} \Rightarrow \begin{cases} a_0 = 5 \\ a_1 = \frac{1}{64} \end{cases}$ Άρα η ευθεία $y = a_0 + a_1 x = \frac{1}{64}x + 5$. Αν θέλουμε, μπορούμε να σχεδιάσουμε την ευθεία αυτή, όπως φαίνεται στη συνέχεια.

B' Τρόπος: **Εναλλακτικά**, αν χρησιμοποιήσουμε τον τύπο της τιμηματικής γραμμικής παρεμβολής θα έχουμε ότι το πολυώνυμο παρεμβολής δίνεται από τον τύπο: $\Pi_1^H f(x) = f(x_i) + \frac{f(x_{i+1}) - f(x_i)}{x_{i+1} - x_i} * (x - x_i), \forall x \in [x_i, x_{i+1}]$ και αν

Θεωρήσουμε ως διάστημα $[x_i, x_{i+1}] = [64, 128]$, θα έχουμε ότι το $\Pi_1^H f(x) = f(64) + \frac{f(128)-f(64)}{128-64} * (x - 64) = 6 + \frac{7-6}{64} * (x - 64) = 6 + \frac{1}{64} * (x - 64) = 6 + \frac{x}{64} - 1 = \frac{1}{64} x + 5$. Δηλαδή καταλήξαμε στο προηγούμενο αποτέλεσμα. Ο λόγος που θεωρήσαμε το συγκεκριμένο διάστημα $[64, 128]$ είναι για ευκολία πράξεων, αφού $f(64) = \log_2 2^6 = 6$ και $f(128) = \log_2 2^7 = 7$. Στη συνέχεια φαίνεται γραφικά η προσέγγιση της τιμής $\log_2 100$ μέσω της ευθείας $y = \frac{1}{64} x + 5$. Αυτή η γραφική παράσταση δεν ζητείται από την εκφώνηση και είναι προαιρετική.



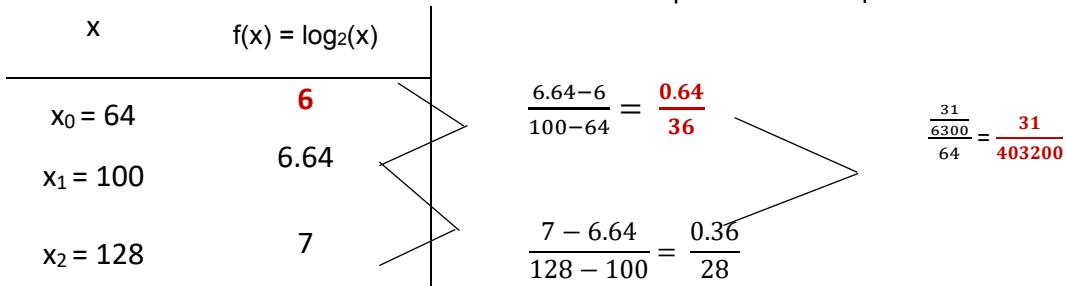
ii) **A' Τρόπος:** Για να προσεγγίσουμε το $\log_2 100$ με πολυωνυμική παρεμβολή, παίρνουμε μια καμπύλη της μορφής $f(x) = \alpha_0 + \alpha_1 x + \alpha_2 x^2$ που είναι ένα πολυώνυμο με τρείς συντελεστές, διότι ζητάμε τρία σημεία.

$$\begin{aligned} x = 128 &\rightarrow \alpha_0 + 128\alpha_1 + 128^2\alpha_2 = 7 \\ x = 64 &\rightarrow \alpha_0 + 64\alpha_1 + 64^2\alpha_2 = 6 \\ x = 32 &\rightarrow \alpha_0 + 32\alpha_1 + 32^2\alpha_2 = 5 \end{aligned} \Rightarrow Ax = b \Rightarrow \begin{bmatrix} 1 & 32 & 32^2 \\ 1 & 64 & 64^2 \\ 1 & 128 & 128^2 \end{bmatrix} \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \alpha_2 \end{bmatrix} = \begin{bmatrix} 5 \\ 6 \\ 7 \end{bmatrix}$$

λύνουμε το σύστημα και βρίσκουμε τους συντελεστές $\alpha_0, \alpha_1, \alpha_2$.

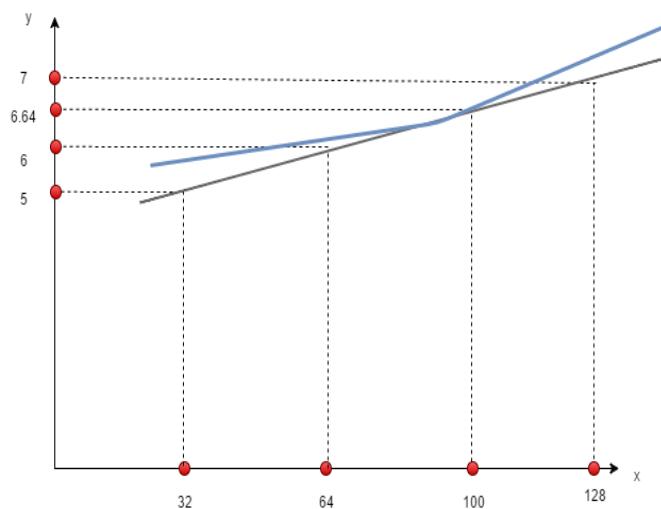
Από την επίλυση του συστήματος των τριών εξισώσεων, θα καταλήξουμε σε μια καμπύλη της μορφής $y = \frac{31x^2}{403200} + 0.00516865x + 5.35429$.

B' Τρόπος: **Εναλλακτικά**, επειδή ζητάμε προσέγγιση της τιμής $\log_2 100$ με πολυωνυμική παρεμβολή, μπορούμε να χρησιμοποιήσουμε **οποιοδήποτε πολυώνυμο παρεμβολής** και στην προκειμένη περίπτωση επιλέξαμε να χρησιμοποιήσουμε ένα πολυώνυμο παρεμβολής στη μορφή Newton. Για το λόγο αυτό, κατασκευάζουμε αρχικά τον πίνακα Δ.Δ. για **τρία** σημεία παρεμβολής (χρησιμοποιούμε τρία σημεία παρεμβολής για να έχουμε πολυωνυμική παρεμβολή, αν χρησιμοποιούσαμε δύο, θα είχαμε γραμμική παρεμβολή), τα οποία επιλέγονται με τυχαίο τρόπο, αλλά με σειρά αύξουσα, έτσι ώστε να είναι εύκολη αν χρειαστεί, η εκ' των υστέρων προσθήκη νέων κόμβων παρεμβολής (αξιοποιώντας το πλεονέκτημα της οικονομικής ανανέωσης).



Επομένως το πολυώνυμο παρεμβολής είναι το $p_2(x) = 6 + \frac{0.64}{36} * (x - 64) + \frac{31}{403200} * (x - 64) * (x - 100) = \frac{31x^2}{403200} +$

0.00516865 x + 5.35429. Στη συνέχεια φαίνεται γραφικά η προσέγγιση της τιμής $\log_2 100$ μέσω της καμπύλης $y = \frac{31x^2}{403200} + 0.00516865 x + 5.35429$. Αυτή η γραφική παράσταση δεν ζητείται από την εκφώνηση.



4.10.7 Άσκηση με πολυώνυμα και νόρμες (παλιό θέμα)

α) Δίνεται το πολυώνυμο $f(\xi) = 4 * \xi^3 - 8 * \xi^2 - 3 * \xi + 1$. Να υπολογίσετε τα στοιχεία του μητρώου V και του διανύσματος g ώστε το γινόμενο $V * g$ να επιστρέφει το διάνυσμα $(f(\xi_1), f(\xi_2), f(\xi_3), f(\xi_4))^T$ με τις τιμές, όπου $\xi_1 = -1, \xi_2 = 0, \xi_3 = 2, \xi_4 = 3$. Για περαιτέρω χρήση, συμβουλεύουμε να υπολογίσετε και τις τιμές αυτές.

β) Δίνεται $A = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}$. Να υπολογίσετε τις νόρμες $\|A\|_1, \|A\|_2, \|A\|_\infty$ και $\|A\|_F$ (νόρμα Frobenius).

Λύση

Με βάση το δοθέν πολυώνυμο, το $f(\xi_1) = -8$, το $f(\xi_2) = 1$, το $f(\xi_3) = -5$ και τέλος το $f(\xi_4) = 28$. Στο γινόμενο $V * g$ το μητρώο V περιέχει σε κάθε γραμμή του τις τιμές του $\xi = -1, 0, 2, 3$ (στα τρία πρώτα στοιχεία, διότι το τέταρτο στοιχείο είναι σταθερό και ίσο με «1»). Το διάνυσμα g περιέχει τους συντελεστές του πολυωνύμου, που είναι οι τιμές 4 (a_3), -8 (a_2), -3 (a_1) και το 1(a_0).

$$\alpha) \begin{bmatrix} (-1)^3 & (-1)^2 & -1 & 1 \\ 0 & 0 & 0 & 1 \\ 2^3 & 2^2 & 2 & 1 \\ 3^3 & 3^2 & 3 & 1 \end{bmatrix} \begin{bmatrix} 4 \\ -8 \\ -3 \\ 1 \end{bmatrix} = \begin{bmatrix} -8 \\ 1 \\ -5 \\ 28 \end{bmatrix} \Rightarrow \begin{bmatrix} -1 & 1 & -1 & 1 \\ 0 & 0 & 0 & 1 \\ 8 & 4 & 2 & 1 \\ 27 & 9 & 3 & 1 \end{bmatrix} \begin{bmatrix} 4 \\ -8 \\ -3 \\ 1 \end{bmatrix} = \begin{bmatrix} -8 \\ 1 \\ -5 \\ 28 \end{bmatrix}$$

Παρατηρούμε ότι το μητρώο V των συντελεστών είναι ένα μητρώο Vandermonde, αφού έχει την κατάλληλη μορφή (η τελευταία στήλη του περιέχει

«1», δηλ. τα στοιχεία του διανύσματος – γεννήτορα υψωμένα στη μηδενική δύναμη, η προτελευταία στήλη του περιέχει τα στοιχεία του διανύσματος – γεννήτορα υψωμένα στην πρώτη δύναμη κ.ο.κ.). Θα πρέπει να υπενθυμίσουμε ότι μητρώα αυτού του είδους **έχουν πολύ κακό δείκτη κατάστασης** ως προς τον πολλαπλασιασμό τος με άλλα μητρώα.

β) Έχουμε ότι: $\|A\|_1 = 3$, (μεγαλύτερο κατ' απόλυτη άθροισμα από όλες τις στήλες) $\|A\|_\infty = 3$ (μεγαλύτερο κατ' απόλυτη άθροισμα από όλες τις γραμμές).

$$\|A\|_2 = \sqrt{\rho(A^T A)} = \sqrt{\rho(\begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}^T * \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix})} = \sqrt{\rho(\begin{bmatrix} 5 & -4 \\ -4 & 5 \end{bmatrix})} = \sqrt{9} = 3 \text{ (φασματική ακτίνα του μητρώου } A^T * A \text{) και τέλος η νόρμα Frobenius } \|A\|_F = \sqrt{\sum(diag(A^T A))} = \sqrt{\sum(diag(\begin{bmatrix} 5 & -4 \\ -4 & 5 \end{bmatrix}))} = \sqrt{5+5} = \sqrt{10}.$$

4.11 Κόμβοι Chebyshev

Οι κόμβοι Chebyshev αποτελούν ένα τρόπο για να εξαλείψουμε **το φαινόμενο του Runge**, δηλ. την αστάθεια της πολυωνυμικής παρεμβολής και ορίζονται ως εξής: Για $j = 0, 1, 2, \dots, n$ στο διάστημα $[a, b]$, υπάρχουν δύο ειδών κόμβων **Chebyshev**:

κόμβοι 1^ο είδους: $x_j = \frac{b+a}{2} + \frac{b-a}{2} \cos\left(\frac{(2j+1)}{(2n)}\pi\right)$, όπου a, b είναι τα άκρα του διαστήματος.

κόμβοι 2^ο είδους: $x_j = \frac{b+a}{2} + \frac{b-a}{2} \cos\left(\frac{j}{n}\pi\right)$, όπου a, b είναι τα άκρα του διαστήματος.

Το **μέγιστο άνω φράγμα σφάλματος του πολυωνύμου παρεμβολής**, όταν έχουμε **ισαπέχοντες** κόμβους δίνεται από τον τύπο: $\max|e_n f(x)| \leq \max \frac{|f^{(n+1)}(x)| h^{n+1}}{4*(n+1)}$ με $x \in \Omega$ και για **μη - ισαπέχοντες** κόμβους – όπως είναι οι **κόμβοι Chebyshev**- δίνεται από τον τύπο: $\max|e_n f(x)| \leq |f^{(n+1)}(\xi)| * \frac{1}{2^n(n+1)!} * \left|\frac{b-a}{2}\right|^{n+1}$ με $x \in \Omega$.

4.11.1 Άσκηση κόμβων Chebyshev

α) Να βρείτε ένα **φράγμα με ισαπέχοντες** κόμβους για το σφάλμα και να βρείτε ποιοι είναι αυτοί. Ορίζουμε το σφάλμα $e(x) = |y(x) - \pi(x)|$ στο διάστημα $[0, 4]$. Επίσης υπάρχει η 5^η παράγωγος στο διάστημα $[0, 4]$ και η μέγιστη τιμή της είναι 1. β) Να υπολογίσετε το απόλυτο σφάλμα αν χρησιμοποιηθούν κόμβοι Chebyshev (δεν είναι ισαπέχοντες) και ποιοι είναι αυτοί; γ) Αν θέλουμε **το ίδιο σφάλμα με ισαπέχοντες** κόμβους, ποια πρέπει να είναι η απόσταση μεταξύ των κόμβων (δηλαδή το h);

Λύση

α) Το **σφάλμα** της παρεμβολής εξαρτάται από τη **συνέχεια της συνάρτησης** και την **κατανομή** των κόμβων.

Από τη στιγμή που υπάρχει η 5^η παράγωγος στο διάστημα $[0, 4]$, σημαίνει ότι το $n = 4$, αφού $n + 1 = 5$. Επειδή ο βαθμός του πολυωνύμου παρεμβολής είναι πάντα κατά «1» μικρότερος από το πλήθος των σημείων παρεμβολής, σημαίνει ότι το πλήθος αυτό είναι ίσο με το 5. Επίσης, για **ισαπέχοντες** κόμβους, **η απόσταση** μεταξύ τους (h) - στην προκειμένη περίπτωση που έχουμε τα σημεία 0, 1, 2, 3, 4 – πρέπει να είναι ίση με 1. Το n

είναι ο δείκτης του τελευταίου σημείου και γενικά το $n+1$ είναι το πλήθος των στοιχείων (σημείων), όταν η αρίθμηση αρχίζει από το μηδέν. Εάν έχουμε η σημεία, τότε η f πρέπει να είναι παραγωγίσιμη στα $(n+1)$ σημεία.

Για **ισαπέχοντες κόμβους** ισχύει ότι το **μέγιστο άνω φράγμα του σφάλματος** είναι: $\max|e_n f(x)| \leq \max \frac{|f^{(n+1)}(x)| h^{n+1}}{4*(n+1)}$ με $x \in \Omega$. Αυτός είναι ο τύπος που αναφέρεται στο μέγιστο άνω φράγμα σφάλματος της πολυωνυμικής παρεμβολής.

Για **μη-ισαπέχοντες κόμβοι** (κόμβοι Chebyshev) ισχύει ότι το άνω φράγμα του σφάλματος είναι:

$$\max|e_n f(x)| \leq |f^{(n+1)}(\xi)| * \frac{1}{2^{n(n+1)!}} * \left|\frac{b-a}{2}\right|^{n+1} \text{ με } x \in \Omega$$

Στην προκειμένη περίπτωση χρησιμοποιούμε **τον πρώτο τύπο υπολογισμού** του άνω φράγματος για το σφάλμα, επειδή έχουμε **ισαπέχοντες κόμβους**, οπότε: $\max|e_n f(x)| \leq \frac{1*1^5}{4*5} = \frac{1}{20} = 0.05$, το οποίο είναι και το **άνω φράγμα σφάλματος**, αν χρησιμοποιηθούν **ισαπέχοντες κόμβοι**. Επίσης το $h = 1$, οπότε το $h^{(n+1)} = h^5 = 1$, διότι για ισαπέχοντες κόμβους η απόσταση μεταξύ τους είναι ίση με 1. Αν τα δοθέντα σημεία ήταν 0, 2, 4, 6, 8 τότε το $h = 2$. Σε κάθε περίπτωση το **h θα είναι σταθερό**, όταν έχουμε ισαπέχοντες κόμβους.

β) Υπενθυμίζουμε ότι έχουμε δύο τύπους υπολογισμού των κόμβων Chebyshev που είναι οι εξής:

1^{ος} τύπος: $x_j = \frac{b+a}{2} + \frac{b-a}{2} * \cos\left(\frac{2j+1}{2n}\pi\right)$ και 2^{ος} τύπος: $x_j = \frac{b+a}{2} + \frac{b-a}{2} * \cos\left(\frac{j}{n}\pi\right)$. Στους τύπους αυτούς το $n = \beta$ αθμός πολυωνύμου = 4, επειδή η εκφώνηση αναφέρει ότι υπάρχει η 5^η παράγωγος, δηλ. $n + 1 = 5 \Rightarrow n = 4$. Ο βαθμός του πολυωνύμου βρίσκεται με δύο τρόπους: α) είτε από το πλήθος των σημείων (κόμβων) παρεμβολής β) είτε από την τάξη της παραγώγου, όπως εδώ. Αν χρησιμοποιήσουμε το **δεύτερο τύπο υπολογισμού των κόμβων Chebyshev** για το διάστημα $[0, 4]$, τότε οι κόμβοι αυτοί θα είναι οι εξής:

$$\text{Για } j = 0, x_0 = \frac{4+0}{2} + \frac{4-0}{2} * \cos\left(\frac{0}{4}\pi\right) = 2 + 2 * \cos\left(\frac{0}{4}\pi\right) = 4$$

$$\text{Για } j = 1, x_1 = 2 + 2 * \cos\left(\frac{1}{4}\pi\right) = 2 + 2 \cos\left(\frac{\pi}{4}\right) = 2 + 2 * \frac{\sqrt{2}}{2} = 2 + \sqrt{2}$$

$$\text{Για } j = 2, x_2 = 2 + 2 \cos\left(\frac{2}{4}\pi\right) = 2 + 2 \cos\left(\frac{\pi}{2}\right) = 2$$

$$\text{Για } j = 3, x_3 = 2 + 2 \cos\left(\frac{3}{4}\pi\right) = 2 - 2 \frac{\sqrt{2}}{2} = 2 - \sqrt{2}$$

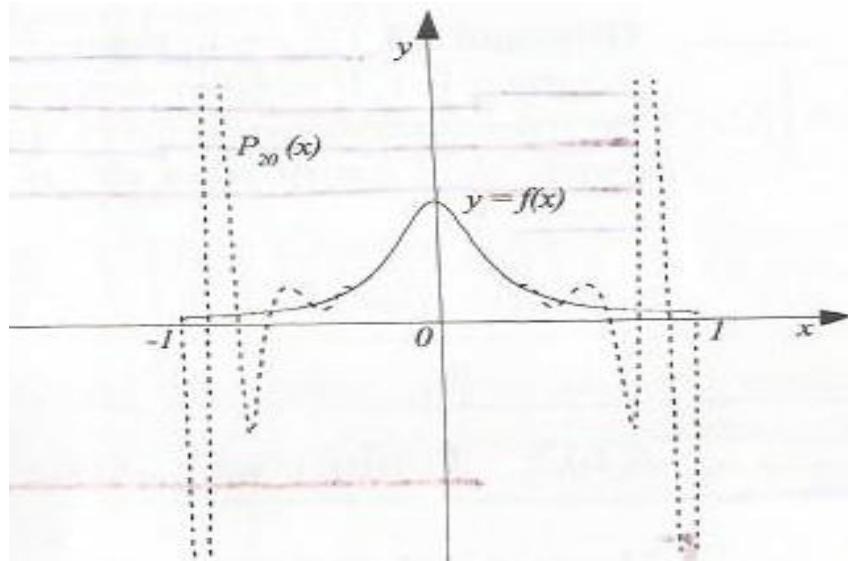
$$\text{Για } j = 4, x_4 = 2 + 2 \cos\left(\frac{4}{4}\pi\right) = 2 - 2 = 0$$

Για τον υπολογισμό του **απόλυτου σφάλματος των κόμβων Chebyshev**, χρησιμοποιούμε το δεύτερο τύπο του απόλυτου σφάλματος που αναφέραμε προηγουμένως, δηλαδή: $\max|e_4 f(x)| \leq \max(1 * \frac{1}{2^{4*5!}} \left(\frac{4-0}{2}\right)^5) = \max \frac{1}{2^4 5!} 2^5 = \frac{2}{5!} = \frac{1}{60} = 0.0167$.

γ) Επειδή μιλάμε για **ισαπέχοντες κόμβους**, θα χρησιμοποιήσουμε τον πρώτο του απόλυτου σφάλματος και από αυτόν θα υπολογίσουμε το h , δηλαδή: $\max|e_4 f(x)| \leq \frac{1*h^5}{20} = 0.0167 \Rightarrow h \leq \sqrt[5]{0.3340} \Rightarrow h \leq 0.8091$.

4.12 Φαινόμενο Runge (Αστάθεια πολυωνυμικής παρεμβολής)

Το φαινόμενο του Runge μελετά την αστάθεια της πολυωνυμικής παρεμβολής. Πιο συγκεκριμένα, το 1901 ο Runge μελέτησε την συμπεριφορά πολυωνύμων παρεμβολής που παρεμβάλλουν με ισαπέχοντα σημεία παρεμβολής για τη συνάρτηση: $f(x) = \frac{1}{1+25x^2}$ στο διάστημα $[-1, 1]$, η οποία τώρα είναι γνωστή ως **συνάρτηση του Runge**. Με τη μελέτη του αυτή ανακάλυψε ότι καθώς ο βαθμός n του πολυωνύμου παρεμβολής $P_n(x)$ αυξάνει, τα πολυώνυμα αυτά δείχνουν μια ασταθή συμπεριφορά, επειδή ισχύει ότι: $\|f(x) - P_n(x)\|_\infty \rightarrow \infty$ καθώς το n τείνει στο άπειρο. Συγκεκριμένα παρατήρησε ότι για μεγάλο n , ενώ το πολυώνυμο παρεμβολής συμπίπτει με τις τιμές της συνάρτησης $f(x)$ στα σημεία παρεμβολής, όμως μεταξύ των σημείων παρεμβολής παρουσιάζονται ταλαντώσεις στα διαστήματα: $-1 < x \leq -0.726$ και $0.726 \leq x < 1$. Επίσης, παρατήρησε ότι το πλάτος των ταλαντώσεων αυξάνει καθώς ο βαθμός n αυξάνει. Αντιθέτως, για το μεσαίο διάστημα: $-0.726 < x < 0.726$ η πολυωνυμική παρεμβολή συμπεριφέρεται ευσταθώς και δίνει καλά αποτελέσματα. Το φαινόμενο αυτό είναι γνωστό ως **φαινόμενο του Runge** και παρουσιάζεται στο σχήμα που ακολουθεί, για ένα πολυώνυμο παρεμβολής εικοστού βαθμού ($n = 20$), το οποίο παρεμβάλλει με ισαπέχοντα σημεία παρεμβολής τη συνάρτηση του Runge.

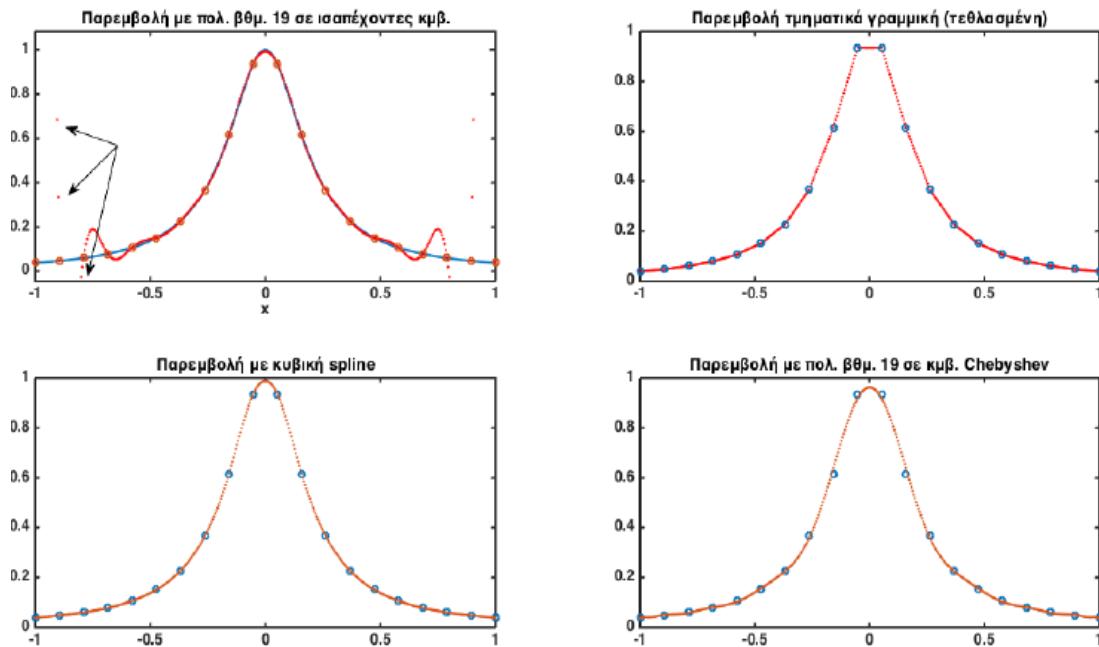


Εικόνα 14: Περιγραφή φαινομένου Runge

Στην επόμενη εικόνα παρουσιάζονται διάφορες παρεμβολές για την προσέγγιση της συνάρτησης $f(x) = \frac{1}{1+25x^2}$. Είναι εύκολο να διαπιστωθεί ότι η καλύτερη παρεμβολή (δηλαδή αυτή με τη μικρότερη απόκλιση) επιτυγχάνεται με τη χρήση της κυβικής spline σε συνδυασμό με τους -μη ισαπέχοντες- κόμβους Chebyshev. Από την πάνω αριστερά παρεμβολή διαπιστώνουμε ότι όταν ο βαθμός του πολυωνύμου είναι μεγάλος (π.χ. 19), αυτό οδηγεί σε σφάλματα κατά την πολυωνυμική παρεμβολή (προσέγγιση). Στην πάνω δεξιά παρεμβολή, λόγω της τμηματικής παρεμβολής που χρησιμοποιείται, έχουμε πολυώνυμα παρεμβολής 1^{ου} βαθμού και χάνουμε κάποια τμήματα της f(x), διότι η τμηματική παρεμβολή πραγματοποιείται αποκλειστικά και μόνο με

ευθείες. Στην κάτω αριστερή παρεμβολή (καλύτερη περίπτωση), η προσέγγιση αυτή πραγματοποιείται με κυβικά πολυώνυμα τμηματικής παρεμβολής και μη – ισαπέχοντες κόμβους παρεμβολής. Στην κάτω δεξιά παρεμβολή, έχουμε και πάλι μη – ισαπέχοντες κόμβους (Chebyshev), αλλά λόγω της αύξησης του βαθμού του πολυωνύμου παρεμβολής, υπάρχει μεγαλύτερη απόκλιση σε σχέση με πριν.

Οι συνθήκες παρεμβολής με splines συχνά οδηγούν σε γραμμικό τριδιαγώνιο σύστημα εξισώσεων για τον υπολογισμό των συντελεστών των πολυωνύμων κάθε υποδιαστήματος.



Εικόνα 15: Διάφορες παρεμβολές για την προσέγγιση της συνάρτησης $f(x) = \frac{1}{1+25x^2}$ με δειγματοληψία από $n = 20$ κόμβους

4.12.1 Αντιμετώπιση του φαινομένου Runge

Η παραπάνω ιδιότητα της καλής συμπεριφοράς των πολυωνύμων παρεμβολής στο κέντρο των δεδομένων μπορεί να αξιοποιηθεί με το να σχηματίζουμε πολυώνυμα παρεμβολής σχετικά υψηλού βαθμού, αλλά α) να τα χρησιμοποιούμε μόνο στην **περιοχή του κέντρου των δεδομένων** με τα οποία αυτά ορίστηκαν και β) μια **εναλλακτική** αντιμετώπιση του φαινομένου του Runge προκύπτει αν δεν χρησιμοποιήσουμε ισαπέχοντα σημεία παρεμβολής, αλλά μη – ισαπέχοντα, τα οποία να πυκνώνουν στα άκρα του διαστήματος παρεμβολής [-1, 1]. Μια τέτοια επιλογή σημείων παρεμβολής είναι τα **σημεία του Chebyshev** που δίνονται από τις ρίζες του πολυωνύμου Chebyshev πρώτου είδους T_{n+1} . Τα σημεία αυτά τα έχουμε συναντήσει σε προηγούμενη ενότητα. Τα σημεία του Chebyshev έχουν όλες τις παραπάνω επιθυμητές ιδιότητες, διότι βρίσκονται όλα στο διάστημα [-1, 1], είναι **διαφορετικά** μεταξύ τους και συσσωρεύονται προς τα άκρα του διαστήματος καθώς το n αυξάνει.

Θεώρημα: Για οποιαδήποτε (τυχαία) επιλογή των σημείων παρεμβολής στο διάστημα [-1, 1] υπάρχει συνεχής συνάρτηση στο διάστημα αυτό για την οποία το πολυώνυμο $P_n(x)$, το οποίο παρεμβάλλει την $f(x)$, στα σημεία αυτά αποτυγχάνει να την προσεγγίσει καθώς το n τείνει στο άπειρο.

4.13 Επαναληπτικές ασκήσεις με πολυώνυμα – 1^η Άσκηση (παλιό θέμα)

Αν έπρεπε να χρησιμοποιήσετε πολυωνυμική παρεμβολή σε μια συνεχή συνάρτηση και είχατε τη δυνατότητα να επιλέξτε την αναπαράσταση του πολυωνύμου παρεμβολής και τους κόμβους, τι θα επιλέγατε; i) αναπαράσταση Newton και ισοκατανεμημένους κόμβους, ii) αναπαράσταση Lagrange και κόμβους Chebyshev, iii) βαρυκεντρική αναπαράσταση και κόμβους Chebyshev και iv) αναπαράσταση Lagrange και ισοκατανεμημένους κόμβους.

Λύση

Η σωστή απάντηση είναι η (iii) διότι αφενός η βαρυκεντρική αναπαράσταση είναι πιο ευσταθής και γρήγορη από την αναπαράσταση Lagrange και αφετέρου ότι δεν επιβαρυνόμαστε με τον υπολογισμό του πολυωνύμου $P_{n+1}(x)$ κάθε φορά που υπολογίζεται το πολυώνυμο Lagrange $L_n(x)$ και αφετέρου οι κόμβοι Chebyshev - ως μη ισοκατανεμημένοι που είναι – εξαλείφουν τις συνέπειες του φαινόμενου Runge. Η απάντηση (i) δεν είναι σωστή διότι μπορεί με την αναπαράσταση Newton να έχουμε οικονομική ανανέωση, με την έννοια ότι είναι εφικτή η προσθήκη ενός νέου στοιχείου (x_{n+1}, y_{n+1}) χωρίς αλλαγή των ήδη υπολογισμένων συντελεστών του πολυωνύμου, αλλά οι ισοκατανεμημένοι κόμβοι συντελούν στην αστάθεια της πολυωνυμικής παρεμβολής.

Το βασικό κριτήριο επιλογής είναι η αποφυγή της αστάθειας της πολυωνυμικής παρεμβολής και στη συνέχεια το δεύτερο κριτήριο επιλογής πρέπει να είναι το πιο οικονομικό πολυώνυμο παρεμβολής. Επίσης, η απάντηση (ii) δεν είναι σωστή διότι μπορεί -όπως προαναφέρθηκε- με τους κόμβους Chebyshev να μειώνεται η αστάθεια της πολυωνυμικής παρεμβολής, **αλλά με την παρεμβολή Lagrange απαιτείται ο προγενέστερος υπολογισμός των πολυωνύμων Lagrange** που αποτελεί μια χρονοβόρα διαδικασία και η παρεμβολή σε ένα επιπλέον σημείο, απαιτεί τον υπολογισμό νέων πολυωνύμων Lagrange από την αρχή, κάτι που σε συνδυασμό με το γεγονός ότι ο υπολογισμός της τιμής του πολυωνύμου σε ένα σημείο στη μορφή Lagrange απαιτεί περισσότερες πράξεις σε σχέση με τις πράξεις τάξης O(n) που απαιτούνται από τη Newton, καθιστούν τη μέθοδο Lagrange ασύμφορη. Τέλος η απάντηση (iv) αποτελεί τη χειρότερη επιλογή με βάση τα όσα προαναφέρθηκαν.

4.13 Επαναληπτικές ασκήσεις με πολυώνυμα – 2^η Άσκηση

Δίνονται τα σημεία $(1, -9), (2, 12), (3, 79), (4, 255)$ που ανήκουν στη γραφική παράσταση μιας συνάρτησης f.

- i) Να βρεθεί το πολυώνυμο που παρεμβάλει την f σε αυτά τα σημεία κατά Lagrange.
- ii) Ποιο θα ήταν το πολυώνυμο παρεμβολής, αν επιπλέον γνωρίζατε ότι $f'(2) = 35$ και $f''(2) = 36$.
- iii) Μαθαίνουμε τώρα πως η συνάρτηση f έχει τη μορφή $ax^4 + b/x$. Τι τιμές θα δίνατε στις παραμέτρους a, b που να ταιριάζουν (μόνο) στα αρχικά δεδομένα? Αφού βρείτε αυτές τις παραμέτρους, συγκρίνετε τις τιμές που δίνει αυτή η προσέγγιση με εκείνες του πολυωνύμου από το (i), σε μερικά σημεία στο $[0, 5]$. [Υπόδειξη: Προσπαθήστε να φτιάξετε ένα γραμμικό σύστημα στο οποίο άγνωστος θα είναι το $(a, b)^T$. Για το δεύτερο σκέλος,

δοκιμάστε για παράδειγμα με $x = [0.5: 0.1: 5]$.

Λύση

Το $n = 3$ διότι έχουμε $3 + 1$ σημεία προς παρεμβολή. Αρχικά, θα υπολογίσουμε τα **πολυώνυμα Lagrange**, τα οποία φαίνονται στη συνέχεια: $L_1(x) = \frac{(x-2)*(x-3)*(x-4)}{(1-2)*(1-3)*(1-4)}$, $L_2(x) = \frac{(x-1)*(x-3)*(x-4)}{(2-1)*(2-3)*(2-4)}$, $L_3(x) = \frac{(x-1)*(x-2)*(x-4)}{(3-1)*(3-2)*(3-4)}$, $L_4(x) = \frac{(x-1)*(x-2)*(x-3)}{(4-1)*(4-2)*(4-3)}$. Επομένως, $\Pi_3 f(x) = \sum_{i=1}^4 L_i(x) * y(i) = L_1(x) * (-9) + L_2(x) * 12 + L_3(x) * 79 + L_4(x) * 255 = \frac{(x-2)*(x-3)*(x-4)}{(1-2)*(1-3)*(1-4)} * (-9) + \frac{(x-1)*(x-3)*(x-4)}{(2-1)*(2-3)*(2-4)} * 12 + \frac{(x-1)*(x-2)*(x-4)}{(3-1)*(3-2)*(3-4)} * 79 + \frac{(x-1)*(x-2)*(x-3)}{(4-1)*(4-2)*(4-3)} * 255 = \frac{21*x^3}{2} - 40 * x^2 + \frac{135*x}{2} - 47$

ii) Πλέον, αφού έχουμε πληροφορία για τιμές των παραγώγων της συνάρτησης, θα κάνουμε παρεμβολή κατά **Hermite**. Κατασκευάζουμε τον σχετικό πίνακα διαφρεμένων διαφορών.

1	-9					
	$f(x_0)$	$f(x_0, x_1)$				
		21				
2	12		$f(x_0, x_1, x_2)$			
			14			
		$f(x_1, x_2)$				
		35		4		
2	12		$f(x_1, x_2, x_3)$			
			18		5	
2	12		$f(x_2, x_3)$			
			35			-2.1250
				14		
					32	-1.3750
3	79				67	11.25
						54.5
4	255				176	

Σημειώνουμε πως το 18 προέκυψε από το $f''(2)/2!$. Υπολογίζουμε τώρα το πολυώνυμο που προκύπτει. Πλέον έχουμε 6 σημεία, άρα θα πρέπει να είναι 5^{ου} βαθμού: $\Pi_5 f(x) = -9 + 21*(x - 1) + 14*(x - 1)*(x - 2) + 4*(x - 1)*(x - 2)*(x - 3) + 5*(x - 1)*(x - 2)^2*(x - 3) - 2.125*(x - 1)*(x - 2)^3*(x - 3) = -2.125x^5 + 26.25x^4 - 113.875x^3 + 241.25x^2 - 233.5x + 73$.

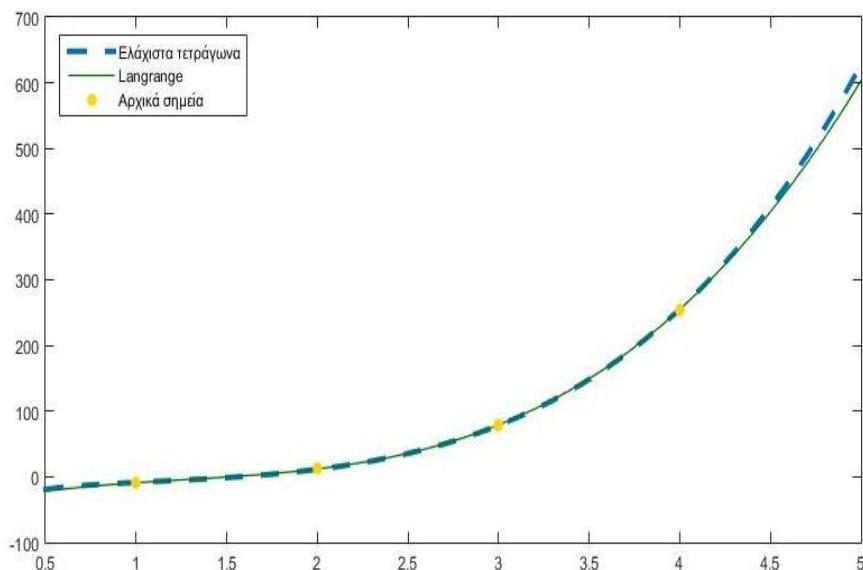
iii) Έστω A , x και b το μητρώο των συντελεστών, το διάνυσμα με τους αγνώστους και το διάνυσμα με τους σταθερούς όρους αντίστοιχα. Το σύστημα $A * x = b$ είναι **υπερκαθορισμένο** διότι έχει περισσότερες εξισώσεις από ότι αγνώστους και μπορούμε να κάνουμε μόνο εκτίμηση των αγνώστων a , b . Αυτό θα γίνει με τη μέθοδο **ελαχίστων τετραγώνων**. Επιλύουμε το σύστημα $A * x = b$, όπου το διάνυσμα $x = [\alpha, \beta]^T$, το διάνυσμα $b = [b_1, b_2, b_3, b_4]^T = [y_1, y_2, y_3, y_4]^T = [-9, 12, 79, 255]^T$ και το μητρώο $A = [\text{ones}(m, 1), x]$, δηλαδή η πρώτη στήλη του θα περιέχει «1» και η δεύτερη στήλη θα περιέχει τους κόμβους παρεμβολής (δηλ. το πρώτο σημείο κάθε ζεύγους, δηλ. τα σημεία παρεμβολής, που είναι οι τιμές 1, 2, 3, 4 από τα ζεύγη (1, -9), (2, 12), (3, 79), (4, 255)). Το μητρώο

των συντελεστών θα είναι το $A = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{bmatrix}$. Η λύση θα είναι το διάνυσμα $x_{LS} = (A^T * A)^{-1} A^T * b$. Επιλύοντας το

σύστημα έχουμε ότι το $x_{LS} = (A^T A)^{-1} A^T b = \begin{bmatrix} 4 & 10 \\ 10 & 30 \end{bmatrix}^{-1} \cdot \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 \end{bmatrix} * \begin{bmatrix} -9 \\ 12 \\ 79 \\ 255 \end{bmatrix} = [1.0064, -9.5197]^T$. Υπενθυμίζουμε ότι το αντίστροφο μητρώο ενός μητρώου 2×2 δίνεται από τον τύπο $A^{-1} = \frac{1}{\det(A)} * \begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix}$.

Ζουμε ότι το πολυώνυμο παρεμβολής στη μέθοδο των ελαχίστων τετράγωνων είναι της μορφής $\phi(t) = x_0 + x_1 * t + x_2 * t^2 + \dots$. Επομένως το πολυώνυμο (συνάρτηση) παρεμβολής θα είναι το $f_{LS}(x) = 1.0064 * x^4 - 9.5197/x$.

Όπως αναμέναμε, έχουμε κάνει μια εκτίμηση, αφού οι τιμές της συνάρτησης είναι **διαφορετικές** από τις τιμές του πολυωνύμου παρεμβολής που υπολογίστηκε με τη μέθοδο των ελαχίστων τετραγώνων. Με άλλα λόγια, **δεν** ικανοποιούνται στην περίπτωση αυτή οι συνθήκες ταύτισης, όπως τις είχαμε συναντήσει σε άλλες περιπτώσεις πολυωνύμων παρεμβολής. Πράγματι: $f_{LS}(1) = -8.5134 \neq f(1) = -9$, $f_{LS}(2) = 11.3419 \neq f(2) = 12$, $f_{LS}(3) = 78.3417 \neq f(3) = 79$ και το $f_{LS}(4) = 255.2475 \neq f(4) = 255$. Συνολικά όμως, η προσέγγιση είναι εξίσου καλή, όπως αποδεικνύεται κι από το ακόλουθο διάγραμμα:



Εικόνα 16: Σύγκριση εκτίμησης πολυωνυμική παρεμβολής με ελάχιστα τετράγωνα και πολυωνυμικής παρεμβολής με Lagrange

4.14 Επαναληπτικές ασκήσεις με πολυώνυμα – 3^η Άσκηση: σύγκριση πολυωνύμων παρεμβολής

	Πλεονεκτήματα	Μειονεκτήματα
Πολυώνυμο παρεμβολής σε μορφή Lagrange	1. Δεν χρειάζεται ομοιόμορφα κατανευμημένες τιμές στο x.	1. Η παρεμβολή σε ένα επιπλέον σημείο απαιτεί τον υπολογισμό νέων πολυωνύμων Lagrange από την αρχή.

	<p>Τα πολυώνυμα Lagrange εξαρτώνται μόνο από τους κόμβους παρεμβολής και όχι από τις τιμές</p>	<p>2. Ο υπολογισμός της τιμής του πολυωνύμου σε ένα σημείο στη μορφή Lagrange απαιτεί περισσότερες πράξεις σε σχέση με τις πράξεις της μορφής Newton.</p> <p>3. Ο υπολογισμός κάθε όρου του πολυωνύμου απαιτεί $O(n^2)$ πράξεις για τον υπολογισμό του π.π. Αν το n είναι μεγάλο, υπάρχει περίπτωση ο υπολογισμός της δυναμομορφής για κάθε πολυώνυμο Lagrange $L_i(x)$ να οδηγεί σε σημαντικά σφάλματα.</p>
Πολυώνυμο παρεμβολής σε μορφή Newton	<ol style="list-style-type: none"> Οικονομική ανανέωση (εφικτή η προσθήκη νέου στοιχείου (x_{n+1}, y_{n+1}), χωρίς αλλαγή των ήδη υπολογισμένων συντελεστών). Αν υλοποιηθεί με τη μέθοδο των διαιρεμένων διαφορών, οδηγεί στον ταχύτερο υπολογισμό των συντελεστών του πολυωνύμου παρεμβολής, σε σχέση με τη μέθοδο Lagrange. Δεν είναι απαραίτητο να δίνονται για κάθε σημείο παρεμβολής, οι παράγωγοι της συνάρτησης, κάτι που είναι απαραίτητο στην παρεμβολή Hermite. 	<ol style="list-style-type: none"> Απαιτείται η υλοποίηση με πίνακα διαιρεμένων διαφορών, για να υπάρχει η οικονομική ανανέωση.
Πολυώνυμο παρεμβολής σε μορφή ελαχίστων τετραγώνων	<ol style="list-style-type: none"> Αποτελεί τη μέθοδο επιλογής όταν έχουμε πολλές μετρήσεις και επιθυμούμε να τις μοντελοποιήσουμε με συνάρτηση που εξαρτάται από λίγες παραμέτρους. Επιλύει υπερπροσδιορισμένα συστήματα, όπου το $m \geq n$. 	<ol style="list-style-type: none"> Αποτελεί προσεγγιστική μέθοδο προσδιορισμού ενός πολυωνύμου παρεμβολής (δηλ. δεν ικανοποιούνται ακριβώς οι συνθήκες ταύτισης).
Πολυώνυμο παρεμβολής σε βαρυκεντρική αναπαράσταση	<ol style="list-style-type: none"> Πιο ευσταθής και γρήγορη από την παρεμβολή Lagrange. 	

	2. Δεν επιβαρυνόμαστε με τον υπολογισμό του $\Pi_{n+1}(x)$ κάθε φορά που υπολογίζεται το $L_n(x)$.	
Πολυώνυμο παρεμβολής σε μορφή Hermite		1. Πρέπει να δίνονται μαζί με τα σημεία παρεμβολής, οι παράγωγοι της συνάρτησης $f(x)$ στα σημεία αυτά.
<p>Κοινό χαρακτηριστικό όλων των πολυωνύμων παρεμβολής είναι ότι ο βαθμός του πολυωνύμου εξαρτάται άμεσα από το πλήθος των σημείων (παρεμβολής) και στη γενική περίπτωση είναι βαθμού $n + 1$ σημεία. Επίσης, απαιτείται μια συνάρτηση να είναι συνεχής σε κάποιο κλειστό διάστημα, για να μπορεί να προσεγγιστεί όσο καλά θέλουμε από ένα πολυώνυμο παρεμβολής.</p>		
Τμηματική κυβική spline	<ol style="list-style-type: none"> Η καλύτερη παρεμβολή (δηλαδή αυτή με τη μικρότερη απόκλιση) επιτυγχάνεται με τη χρήση της κυβικής spline και με τους -μη ισαπέχοντες- κόμβους Chebyshev. Οι συνθήκες παρεμβολής με splines συχνά οδηγούν σε γραμμικό τριδιαγώνιο σύστημα εξισώσεων για τον υπολογισμό των συντελεστών των πολυωνύμων κάθε υποδιαστήματος. 	<ol style="list-style-type: none"> Επλέγεται με τέτοιο τρόπο ώστε να είναι και δύο φορές συνεχώς παραγώγισμη στο διάστημα παρεμβολής.
Τμηματικό γραμμικό πολυώνυμο παρεμβολής	<ol style="list-style-type: none"> Η βελτίωση στην προσέγγιση επιτυγχάνεται μέσω αύξησης των κόμβων παρεμβολής, ενώ ο βαθμός του πολυωνύμου σε κάθε υποδιάστημα παραμένει γραμμικός. 	<ol style="list-style-type: none"> Υπάρχει ασυνέχεια στις παραγώγους
Τμηματικό πολυώνυμα τύπου Hermite		<ol style="list-style-type: none"> Απαιτείται να είναι γνωστές οι τιμές της παραγώγου της συνάρτησης, την οποία επιθυμούμε να προσεγγίσουμε.

Κεφάλαιο 5 – Επίλυση μη γραμμικών εξισώσεων πρώτου βαθμού

5.1 Μη γραμμικά συστήματα εξισώσεων και ασκήσεις

Για την επίλυση ενός μη γραμμικού συστήματος εξισώσεων χρησιμοποιούμε τη μέθοδο **Newton** που αποτελεί γενίκευση της μεθόδου **Newton – Raphson** που χρησιμοποιείται για εύρεση ριζών συναρτήσεων. Το x^* $\in \mathbb{R}^*$ ονομάζεται **ρίζα** της συνάρτησης F_n , αν ισχύει ότι: $F_n(x^*) = \Theta_n$, όπου η F_n γράφεται: $F_n(x) = (f_1(x), f_2(x), \dots, f_n(x))^T$. Δηλαδή η εξισωση $F_n(x^*) = \Theta_n$ μπορεί ισοδύναμα να γραφεί ως ένα **σύστημα εξισώσεων**, όπως φαίνεται παρακάτω:

$$\begin{cases} f_1(x_1, x_2, \dots, x_n) = 0 \\ f_2(x_1, x_2, \dots, x_n) = 0 \\ \dots \\ f_n(x_1, x_2, \dots, x_n) = 0 \end{cases}$$

Για να λύσουμε αυτό το σύστημα εξισώσεων, χρησιμοποιούμε τη **μέθοδο Newton** που είναι της μορφής: $J(x^k) * s^k = -F_n(x^k)$, όπου $k = 0, 1, 2, \dots$ ή εναλλακτικά $x^{k+1} = x^k - J(x^k) * F_n(x^k)$. Το μητρώο J (Jacobi) έχει τη μορφή που φαίνεται παρακάτω:

$$J = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \dots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \dots & \frac{\partial f_2}{\partial x_n} \\ \dots & \dots & \dots & \dots \\ \frac{\partial f_n}{\partial x_1} & \frac{\partial f_n}{\partial x_2} & \dots & \frac{\partial f_n}{\partial x_n} \end{bmatrix}$$

1^o παράδειγμα επίλυσης μη – γραμμικού συστήματος

Να εφαρμοστεί ο **αλγόριθμος Newton** για την προσέγγιση λύσης του παρακάτω συστήματος εξισώσεων (να εκτελέσετε 2 βήματα):

$$\begin{aligned} f_1(x_1, x_2) &= x_1 + x_2 - 3 \\ f_2(x_1, x_2) &= x_1^2 + x_2^2 - 9 \end{aligned}$$

Παρατηρούμε ότι το σύστημα εξισώσεων που πρέπει να λύσουμε, με βάση την αρχική εκτίμηση της λύσης που δίνεται, αποτελείται από δύο εξισώσεις, εκ' των οποίων η δεύτερη είναι μη – γραμμική (δευτέρου βαθμού)

χρησιμοποιώντας ως **αρχική εκτίμηση λύσης** την $x^{(0)} = [1 \ 5]^T$

Λύση

Όταν θέλουμε να λύσουμε σύστημα μη - γραμμικών εξισώσεων, χρησιμοποιούμε τη μέθοδο Newton. Υπολογίζουμε **πρώτα** το Ιακωβιανό μητρώο, $J(x)$, το οποίο περιέχει τις μερικές παραγώγους κάθε διθείσας συνάρτησης του γραμμικού συστήματος ως προς όλες τις μεταβλητές. Πιο συγκεκριμένα, το Ιακωβιανό μητρώο υπολογίζεται ως εξής: $J(x) = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 2x_1 & 2x_2 \end{bmatrix}$. Επειδή ως αρχική εκτίμηση λύσης έχει δοθεί το διάνυσμα

$x^{(0)} = [1 \ 5]^T$, θα προσπαθήσουμε να το βελτιώσουμε, εφαρμόζοντας τη μέθοδο Newton. Αρχικά, εφαρμόζουμε

τον τύπο $J(x^0) * s^0 = -F_n(x^0) \Rightarrow \begin{bmatrix} 1 & 1 \\ 2 * 1 & 2 * 5 \end{bmatrix} * \begin{bmatrix} s_1^{(0)} \\ s_2^{(0)} \end{bmatrix} = -\begin{bmatrix} f_1(x_1^{(0)}, x_2^{(0)}) \\ f_2(x_1^{(0)}, x_2^{(0)}) \end{bmatrix} \Rightarrow \begin{bmatrix} 1 & 1 \\ 2 & 10 \end{bmatrix} * \begin{bmatrix} s_1^{(0)} \\ s_2^{(0)} \end{bmatrix} = -\begin{bmatrix} 3 \\ 17 \end{bmatrix}$ Από την επίλυση

του προηγούμενου συστήματος έχουμε: $s_1^0 + s_2^0 = -3$ και $2 * s_1^0 + 10 * s_2^0 = -17 \Rightarrow s_2^0 = -11/8$ και $s_1^0 = -13/8$ που

είναι η διόρθωση της προσεγγιστικής λύσης. Επομένως, η επόμενη εκτίμηση της λύσης είναι: $x^{(1)} = \begin{bmatrix} x_1^{(1)} \\ x_2^{(1)} \end{bmatrix} =$

$$\begin{bmatrix} x_1^{(0)} \\ x_2^{(0)} \end{bmatrix} + \begin{bmatrix} s_1^{(0)} \\ s_2^{(0)} \end{bmatrix} = \begin{bmatrix} 1 \\ 5 \end{bmatrix} - \begin{bmatrix} \frac{13}{8} \\ \frac{29}{8} \end{bmatrix} = \begin{bmatrix} -\frac{5}{8} \\ \frac{29}{8} \end{bmatrix}. \text{ Οσα περισσότερα βήματα της παραπάνω διαδικασίας εκτελούνται, τόσο}$$

καλύτερη προσέγγιση της ακριβής λύσης επιτυγχάνεται. Με δεδομένο ότι η ακριβής λύση είναι το διάνυσμα $[0, 3]^T$, παρατηρούμε ότι η δεύτερη προσέγγιση λύσης είναι καλύτερη από την πρώτη. Γενικότερα, ισχύει ο

$$\text{τύπος: } \begin{bmatrix} x_1^{(k+1)} \\ x_2^{(k+1)} \\ \vdots \\ x_n^{(k+1)} \end{bmatrix} = \begin{bmatrix} x_1^{(k)} \\ x_2^{(k)} \\ \vdots \\ x_n^{(k)} \end{bmatrix} + \begin{bmatrix} s_1^{(k)} \\ s_2^{(k)} \\ \vdots \\ s_n^{(k)} \end{bmatrix}. \text{ Στη συνέχεια εκτελούμε το δεύτερο βήμα της επανάληψης και επιλύουμε}$$

το γραμμικό σύστημα: $J(x^1) * s^1 = -F_n(x^1) \Rightarrow \begin{bmatrix} 1 & 1 \\ 2 * (-5/8) & 2 * (29/8) \end{bmatrix} * \begin{bmatrix} s_1^{(1)} \\ s_2^{(1)} \end{bmatrix} = -\begin{bmatrix} f_1(x_1^{(1)}, x_2^{(1)}) \\ f_2(x_1^{(1)}, x_2^{(1)}) \end{bmatrix} \Rightarrow$

$$\begin{bmatrix} 1 & 1 \\ -10/8 & 29/4 \end{bmatrix} * \begin{bmatrix} s_1^{(1)} \\ s_2^{(1)} \end{bmatrix} = -\begin{bmatrix} 0 \\ 145/325 \end{bmatrix} \text{ Η λύση του παραπάνω συστήματος για να βρούμε τη διόρθωση είναι:}$$

$$s_1^1 + s_2^1 = 0 \Rightarrow s_1^1 = 145/272 \text{ και από την εξίσωση: } -10/8 * s_1^1 + 29/4 * s_2^1 = -145/32 \Rightarrow s_2^1 = -145/272. \text{ Η επόμενη}$$

$$\text{εκτίμηση λύσης (πιο ακριβής σε σχέση με πριν) είναι } x^{(2)} = \begin{bmatrix} x_1^{(2)} \\ x_2^{(2)} \end{bmatrix} = \begin{bmatrix} x_1^{(1)} \\ x_2^{(1)} \end{bmatrix} + \begin{bmatrix} s_1^{(1)} \\ s_2^{(1)} \end{bmatrix} = \begin{bmatrix} -5/8 \\ -29/8 \end{bmatrix} + \begin{bmatrix} 145/272 \\ -145/272 \end{bmatrix} =$$

$$\begin{bmatrix} -0.092 \\ 3.092 \end{bmatrix}. \text{ Παρατηρούμε ότι η δεύτερη προσέγγιση είναι πιο ακριβής από την πρώτη, άρα όσο πιο πολλά βή-$$

ματα εκτελούμε, τόσο πιο ακριβή προσέγγιση της λύσης του μη - γραμμικού συστήματος των εξισώσεων πε-
τυχαίνουμε.

2º παράδειγμα επίλυσης μη - γραμμικού συστήματος

Βρείτε τα σημεία τομής ενός κύκλου με εξίσωση $x^2 + y^2 = 2$ και μιας υπερβολής με εξίσωση $x^2 - y^2 = 1$ με αρχική προσέγγιση $x_0 = 1, y_0 = 1$ (για $\kappa = 0$).

Λύση

Αν κάνουμε ακριβή επίλυση του μη - γραμμικού συστήματος των δύο εξισώσεων που δίνονται θα έχουμε:

$$x^2 - y^2 = 1 \Rightarrow x^2 - 2 + x^2 = 1 \Rightarrow 2 * x^2 = 3 \Rightarrow x = \pm 1.2247.$$

$$x^2 + y^2 = 2 \Rightarrow y^2 = 2 - x^2 \Rightarrow y^2 = 2 - 1.2247^2 \Rightarrow y = \pm 0.7071.$$

Αν ορίσουμε στη συνέχεια το μη-γραμμικό σύστημα που περιγράφεται από τις επόμενες δύο εξισώσεις: $x^2 + y^2 - 2 = g(x, y)$

$$x^2 - y^2 - 1 = h(x, y)$$

τότε το Ιακωβιανό μητρώο θα είναι: $J = \begin{bmatrix} \frac{\partial g}{\partial x} & \frac{\partial g}{\partial y} \\ \frac{\partial h}{\partial x} & \frac{\partial h}{\partial y} \end{bmatrix} = \begin{bmatrix} 2x & 2y \\ 2x & -2y \end{bmatrix}$ και η αρχική προσέγγιση του μητρώου αυτού θα

είναι $J_0 = J(x^0) = \begin{bmatrix} 2 & 2 \\ 2 & -2 \end{bmatrix}$ και το $b = -F_n(x^0) = \begin{bmatrix} -g(x_0, y_0) \\ -h(x_0, y_0) \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$. Επιλύουμε το σύστημα $J_0 * h_0 = b_0$ με LU παραγοντοποίηση, διότι έτσι δεν αλλάζει το b . Πιο συγκεκριμένα, έχουμε τα τρία βήματα LU παραγοντοποίησης που κατά σειρά είναι τα εξής:

1^o βήμα LU παραγοντοποίησης: διάσπαση του μητρώου J_0 σε μητρώα L και U.

$$L_1 * J_0 = \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix} * \begin{bmatrix} 2 & 2 \\ 2 & -2 \end{bmatrix} = \begin{bmatrix} 2 & 2 \\ 0 & -4 \end{bmatrix} = U \Rightarrow J_0 = L_1^{-1} * U = L * U = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} * \begin{bmatrix} 2 & 2 \\ 0 & -4 \end{bmatrix} = \begin{bmatrix} 2 & 2 \\ 2 & -2 \end{bmatrix}.$$

2^o βήμα: Εμπρός αντικατάσταση: $L * y = b \Rightarrow \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} * \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \Rightarrow y_1 = 0$ και $y_2 = 1$.

3^o βήμα: Πίσω αντικατάσταση: $U * h = y \Rightarrow \begin{bmatrix} 2 & 2 \\ 0 & -4 \end{bmatrix} * \begin{bmatrix} h_1 \\ h_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \Rightarrow h_1 = 1/4$ και $h_2 = -1/4$.

Άρα η διόρθωση είναι: $h_0 = [h_1 \ h_2]^T = [1/4 \ -1/4]^T$ και $x_1 - x_0 = 0.25 = 1/4 \Rightarrow x_1 = 1.25$ και $y_1 - y_0 = -0.25 = -1/4 \Rightarrow y_1 = 0.75$. Αυτή είναι η λύση του γραμμικού συστήματος μετά από ένα βήμα εκτέλεσης της μεθόδου Newton.

Στη συνέχεια επαναλαμβάνουμε τη διαδικασία για άλλη μια φορά. Υπολογίζουμε το μητρώο $J_1 = \begin{bmatrix} 2x_1 & 2y_1 \\ 2x_1 & -2y_1 \end{bmatrix}$

$$= \begin{bmatrix} 2 * 1.25 & 2 * 0.75 \\ 2 * 1.25 & -2 * 0.75 \end{bmatrix} = \begin{bmatrix} 2.5 & 1.5 \\ 2.5 & -1.5 \end{bmatrix} \text{ και το διάνυσμα } b_1 = \begin{bmatrix} -g(x_1, y_1) \\ -h(x_1, y_1) \end{bmatrix} = \begin{bmatrix} -(1.25^2 + 0.75^2 - 2) \\ -(1.25^2 - 0.75^2 - 1) \end{bmatrix} = \begin{bmatrix} -0.125 \\ 0 \end{bmatrix}. \text{ Στη}$$

συνέχεια επιλύουμε το γραμμικό σύστημα $J_1 * h_1 = b_1 \Rightarrow \begin{bmatrix} 2.5 & 1.5 \\ 2.5 & -1.5 \end{bmatrix} * h_1 = \begin{bmatrix} -0.125 \\ 0 \end{bmatrix}$. Με απαλοιφή Gauss

λύνουμε το σύστημα: $L_1 * J_1 = \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix} * \begin{bmatrix} 2.5 & 1.5 \\ 2.5 & -1.5 \end{bmatrix} = \begin{bmatrix} 2.5 & 1.5 \\ 0 & -3 \end{bmatrix} = U \Rightarrow J_1 = L_1^{-1} * U = L * U = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} *$
 $\begin{bmatrix} 2.5 & 1.5 \\ 0 & -3 \end{bmatrix}$ που είναι το πρώτο βήμα της LU παραγοντοποίησης. Στη συνέχεια εκτελούμε το 2^o βήμα που είναι η

εμπρός αντικατάσταση: $L * y = b \Rightarrow \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} * \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} -0.125 \\ 0 \end{bmatrix}$ και $y_1 = -0.125$ και $y_2 = 0.125$ και τέλος εκτελούμε

το 3^o βήμα που είναι η πίσω αντικατάσταση: $U * h = y \Rightarrow \begin{bmatrix} 2.5 & 1.5 \\ 0 & -3 \end{bmatrix} * \begin{bmatrix} h_1 \\ h_2 \end{bmatrix} = \begin{bmatrix} -0.125 \\ 0.125 \end{bmatrix} \Rightarrow h_1 = -0.025$ και $h_2 = -0.0417$.

Άρα η διόρθωση είναι: $h_1 = \begin{bmatrix} h_1 \\ h_2 \end{bmatrix} = \begin{bmatrix} -0.025 \\ -0.0417 \end{bmatrix}$. Επομένως: $x_2 - x_1 = -0.025 \Rightarrow x_2 = 1.225$ και $y_2 - y_1 = -0.0417$

$\Rightarrow y_2 = 0.7083$. Στη συνέχεια επαναλαμβάνουμε τη διαδικασία για άλλη μια φορά και υπολογίζουμε το νέο

μητρώο $J_2 = \begin{bmatrix} 2x_2 & 2y_2 \\ 2x_2 & -2y_2 \end{bmatrix} = \begin{bmatrix} 2 * 1.225 & 2 * 0.7083 \\ 2 * 1.225 & -2 * 0.7083 \end{bmatrix} = \begin{bmatrix} 2.45 & 1.4166 \\ 2.45 & -1.4166 \end{bmatrix}$ και το διάνυσμα $b_2 = \begin{bmatrix} -g(x_2, y_2) \\ -h(x_2, y_2) \end{bmatrix}$

$$= \begin{bmatrix} -(1.225^2 + 0.7083^2 - 2) \\ -(1.225^2 - 0.7083^2 - 1) \end{bmatrix} = \begin{bmatrix} -0.0023 \\ -0.0011 \end{bmatrix}. \text{ Άρα το } L_1 * J_2 = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} * \begin{bmatrix} 2.45 & 1.4166 \\ 2.45 & -1.4166 \end{bmatrix} = \begin{bmatrix} 2.45 & 1.416 \\ 0 & -2.83 \end{bmatrix} = U. \text{ Κάνοντας}$$

και πάλι εμπρός και πίσω αντικατάσταση έχουμε ότι: $h_3 = \begin{bmatrix} -0.0003 \\ -0.0012 \end{bmatrix}$ και $x_3 = 1.2247$ και $y_3 = 0.7071$ που αποτελεί μια καλή προσέγγιση της ακριβής λύσης που υπολογίστηκε στην αρχή.

3^ο παράδειγμα επίλυσης μη – γραμμικού συστήματος

Να εφαρμοστεί ο **αλγόριθμος** Newton για την προσέγγιση λύσης του παρακάτω συστήματος εξισώσεων:

$$\begin{aligned} x^3 + y &= 1 & \text{Παρατηρούμε ότι και πάλι το σύστημα εξισώσεων που πρέπει να λύσουμε, με βάση την αρχική} \\ y^3 - x &= -1 & \text{εκτίμηση της λύσης που δίνεται, είναι μη – γραμμικό, αφού αποτελείται από δύο εξισώσεις} \\ && \text{τρίτου βαθμού.} \end{aligned}$$

χρησιμοποιώντας ως αρχική εκτίμηση λύσης την $x^{(0)} = [0.5, 0.5]^T$

Λύση

Όταν θέλουμε να λύσουμε σύστημα μη - γραμμικών εξισώσεων (στο οποίο μια τουλάχιστον από τις εξισώσεις του είναι δευτέρου βαθμού και πάνω), χρησιμοποιούμε τη μέθοδο Newton. Υπολογίζουμε πρώτα το Ιακωβιανό μητρώο, $J(x)$, το οποίο περιέχει τις μερικές παραγώγους κάθε διθείσας συνάρτησης του γραμμικού συστήματος ως προς όλες τις μεταβλητές. Πιο συγκεκριμένα, το Ιακωβιανό μητρώο υπολογίζεται ως εξής: $J(x) =$

$$\begin{bmatrix} \frac{\partial f_1}{\partial x} & \frac{\partial f_1}{\partial y} \\ \frac{\partial f_2}{\partial x} & \frac{\partial f_2}{\partial y} \end{bmatrix} = \begin{bmatrix} 3x^2 & 1 \\ -1 & 3y^2 \end{bmatrix}. \quad \text{Επειδή η αρχική εκτίμηση λύσης } x^{(0)} = [0.5, 0.5]^T, \text{ το } J(x^0) * S^0 = -F_n(x^0)$$

$$\Rightarrow \begin{bmatrix} 3 * 0.5^2 & 1 \\ -1 & 3 * 0.5^2 \end{bmatrix} * \begin{bmatrix} s_1^{(0)} \\ s_2^{(0)} \end{bmatrix} = - \begin{bmatrix} f_1(x_1^{(0)}, x_2^{(0)}) \\ f_2(x_1^{(0)}, x_2^{(0)}) \end{bmatrix} \Rightarrow \begin{bmatrix} 0.75 & 1 \\ -1 & 0.75 \end{bmatrix} * \begin{bmatrix} s_1^{(0)} \\ s_2^{(0)} \end{bmatrix} = - \begin{bmatrix} -0.375 \\ 0.625 \end{bmatrix}. \quad \text{Από την επίλυση του συστήματος}$$

$$0.75 * s_1^{(0)} + s_2^{(0)} = 0.375 \quad \text{και} \quad -1 * s_1^{(0)} + 0.75 * s_2^{(0)} = -0.625, \quad \text{έχουμε: } s_1^0 = 0.58 \text{ και } s_2^0 = -0.06 \text{ είναι η διόρθωση}$$

$$\text{της προσεγγιστικής λύσης. Η επόμενη εκτίμηση λύσης είναι: } x^{(1)} = \begin{bmatrix} x_1^{(1)} \\ x_2^{(1)} \end{bmatrix} = \begin{bmatrix} x_1^{(0)} \\ x_2^{(0)} \end{bmatrix} + \begin{bmatrix} s_1^{(0)} \\ s_2^{(0)} \end{bmatrix} = \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix} + \begin{bmatrix} 0.58 \\ -0.06 \end{bmatrix} = \begin{bmatrix} 1.08 \\ 0.44 \end{bmatrix}.$$

Όσα περισσότερα βήματα της παραπάνω διαδικασίας εκτελούνται, τόσο καλύτερη προσέγγιση της ακριβής λύσης επιτυγχάνεται. Για το λόγο αυτό, επαναλαμβάνοντας την προηγούμενη διαδικασία άλλη μια φορά, θα

$$\text{έχουμε ότι: το } J(x^1) * S^1 = -F_n(x^1) \Rightarrow \begin{bmatrix} 3 * \textcolor{red}{1.08}^2 & 1 \\ -1 & 3 * \textcolor{red}{0.44}^2 \end{bmatrix} * \begin{bmatrix} s_1^{(1)} \\ s_2^{(1)} \end{bmatrix} = - \begin{bmatrix} f_1(x_1^{(1)}, x_2^{(1)}) \\ f_2(x_1^{(1)}, x_2^{(1)}) \end{bmatrix} \Rightarrow \begin{bmatrix} 3.4992 & 1 \\ -1 & 0.5808 \end{bmatrix} * \begin{bmatrix} s_1^{(1)} \\ s_2^{(1)} \end{bmatrix} = - \begin{bmatrix} 0.6997 \\ 0.0051 \end{bmatrix}. \quad \text{Από την επίλυση του συστήματος } 3 * \textcolor{red}{1.08}^2 * s_1^{(1)} + s_2^{(1)} = -0.6997 \text{ και} \\ -1 * s_1^{(1)} + 3 * \textcolor{red}{0.44}^2 * s_2^{(1)} = -0.0051, \quad \text{έχουμε: } s_1^1 = -0.1337 \text{ και } s_2^1 = -0.2318 \text{ είναι η διόρθωση της προσεγγιστικής λύσης. Η επόμενη εκτίμηση λύσης είναι: } x^{(2)} = \begin{bmatrix} x_1^{(2)} \\ x_2^{(2)} \end{bmatrix} = \begin{bmatrix} x_1^{(1)} \\ x_2^{(1)} \end{bmatrix} + \begin{bmatrix} s_1^{(1)} \\ s_2^{(1)} \end{bmatrix} = \begin{bmatrix} 1.08 \\ 0.44 \end{bmatrix} - \begin{bmatrix} 0.1337 \\ -0.2318 \end{bmatrix} = \begin{bmatrix} 0.9477 \\ 0.2033 \end{bmatrix}. \quad \text{Η ακριβής λύση του συστήματος είναι} \\ x = 1, y = 0.$$

5.2 Σύγκριση διχοτόμησης, Newton – Raphson, τέμνουσας, εσφαλμένης θέσης (regula falsi)

Είναι μέθοδοι που κάνουν εύρεση ριζών (λύσεων) συναρτήσεων (μη γραμμικών εξισώσεων). Πρόκειται ουσιαστικά για **επαναληπτικές μεθόδους προσεγγιστικού υπολογισμού ριζών συναρτήσεων**, κάθε μια από τις οποίες έχει τα πλεονεκτήματα και τα μειονεκτήματά της. **Για την εύρεση μιας ρίζας μιας συνάρτησης, χρησιμοποιούμε το Θεώρημα Bolzano που αναφέρει ότι για να υπάρχει μια τουλάχιστον ρίζα μιας συνάρτησης f στο διάστημα [a b] θα πρέπει να ισχύει ότι $f(a) * f(b) < 0$.**

Αφού βρεθεί ρίζα σε κάποιο διάστημα, στη συνέχεια διχοτομούμε το διάστημα αυτό στη μέση και εφαρμόζουμε το θεώρημα Bolzano στο πάνω και στο κάτω μισό του διαστήματος για να βρούμε με μεγαλύτερη ακρίβεια τη ρίζα της συνάρτησης. Αυτή η διαδικασία επαναλαμβάνεται για όσο υποδεικνύει το **πλήθος των επαναλήψεων της μεθόδου διχοτόμησης** -που διαθέτει **γραμμική σύγκλιση**- και το οποίο δίνεται από τον τύπο:

$k \geq \lceil \log_2((b - a) * \varepsilon^{-1}) \rceil$, όπου **a, b** είναι τα άκρα του διαστήματος και ε είναι η επιθυμητή ακρίβεια της λύσης που προσδιορίζεται από την εκφώνηση του προβλήματος. **Το γεγονός ότι η σύγκλιση της μεθόδου αυτής είναι γραμμική**, σημαίνει ότι η μέθοδος συγκλίνει **αργά** στην προσεγγιστική λύση.

Όσον αφορά τη μέθοδο **Newton – Raphson**, **αναφέρουμε ότι** αν η συνάρτηση είναι γραμμική, η μέθοδος βρίσκει τη λύση **σε μια επανάληψη**. Από τη στιγμή που η προσέγγιση ξ της λύσης (ρίζας) ξ είναι καλή, η **σύγκλιση της μεθόδου γίνεται πολύ γρήγορη**, δηλαδή γίνεται **τετραγωνική** και οι επαναλήψεις της σταματούν, όταν: α) η προσέγγιση θεωρηθεί ικανοποιητική ή β) όταν ξεπεραστεί ένας μέγιστος αριθμός επαναλήψεων ή γ) αν φανεί ότι η διαδικασία αποκλίνει. **Συνήθως**, όταν χρησιμοποιούμε τη μέθοδο αυτή, εφαρμόζουμε **πρώτα** τη μέθοδο της διχοτόμησης για να βρούμε μια όσο το δυνατόν πιο ακριβή προσέγγιση της ρίζας, την οποία στη συνέχεια χρησιμοποιούμε ως αρχική προσέγγιση για τη μέθοδο **Newton - Raphson**, χρησιμοποιώντας τον τύπο: $x_n = x_{n-1} - \frac{f(x_{n-1})}{f'(x_{n-1})}$. Η **συνήθης δυσκολία** της μεθόδου Newton - Raphson είναι ότι **δεν** έχουμε τρόπο να υπολογίσουμε, πότε το σφάλμα της προσέγγισης θα γίνει αποδεκτό, δηλαδή το $|e_k| = |\xi - x_k| < \varepsilon$, για κάποιο ε (ακρίβεια) που έχουμε ορίσει από πριν. Για το λόγο αυτό χρησιμοποιούμε ως κριτήρια τερματισμού της μεθόδου, την απόσταση διαδοχικών τιμών $|x_k - x_{k-1}| < \varepsilon$ ή το μέγεθος του κατάλοιπου: $|r_k| = |f(x_k)| < \varepsilon$.

Εναλλακτικά, αν η ρίζα που ψάχνουμε έχει **μια πολλαπλότητα m**, τότε μια **παραλλαγή** της **Newton – Raphson** είναι η εξής: $x_{n+1} = x_n - m \frac{f(x_n)}{f'(x_n)}$. Η τροποποιημένη μέθοδος Newton – Raphson **έχει υπεργραμμική σύγκλιση** (ανάμεσα στη γραμμική και την τετραγωνική). Ο κώδικας MATLAB της μεθόδου Newton - Raphson είναι:

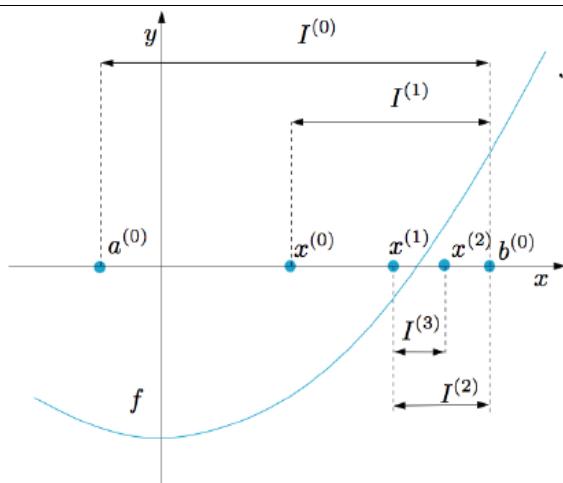
```
function [x, k] = newton(f, fprime, x)
% x = newton(f, fprime, x) tries to find a zero of f(x) near the initial guess x . If successful, newton returns the
% newton approximation to a root of f.
% [x, iter] = newton(f, fprime, x) also returns the number of iterations required to find the root. Both functions
% % f and fprime need to be provided .
k = 0; % αρχικοποίηση μετρητή επαναλήψεων
xprev = realmax; % just to get started
while abs(x - xprev) > eps * abs(x) % για όσο χρόνο δεν έχει επιτευχθεί η επιθυμητή ακρίβεια eps * abs (x)
    xprev = x;
    if fprime(x) ~= 0 % εφόσον η παράγωγος είναι διάφορη του «0»
        x = x - f(x)/fprime(x); % υπολογισμός νέας προσέγγισης της ρίζας
    else
        disp (sprintf( 'Newton iteration failed since derivative is zero .\n' ) )
        return
    end
    k = k + 1;
disp(sprintf('x%d = %16.14 f' , k, x))
```

Όσον αφορά τη μέθοδο της **τέμνουσας (χορδής)** έχουμε ότι: $x_{k+1} = x_k - \frac{(x_k - x_{k-1})}{f(x_k) - f(x_{k-1})} * f(x_k)$. Δηλαδή **χρειάζεται τις δύο προηγούμενες προσεγγίσεις μιας ρίζας σε αντίθεση με τη Newton – Raphson που χρειάζεται μόνο την προηγούμενη προσέγγιση.**

Όσον αφορά τη μέθοδο της **εσφαλμένης θέσης** έχουμε να παρατηρήσουμε ότι **προϋποθέτει τη γνώση για την ύπαρξη ρίζας σε κάποιο διάστημα**. Ο τύπος της μεθόδου αυτής είναι ο εξής: $\tilde{x}_{k+1} = x_k - \frac{(x_k - x_{k-1})}{f(x_k) - f(x_{k-1})} * f(x_k)$. Εναλλακτικά ο τελευταίος τύπος γράφεται και ως εξής: $\tilde{x}_{k+1} = x_k - \frac{(x_k - b)}{f(x_k) - f(b)} * f(x_k), k = 1, 2, \dots$ όπου $x_1 = a$. Δηλαδή το σημείο $x = b$ παραμένει σταθερό, ενώ το άλλο σημείο ανανεώνεται σε κάθε βήμα.

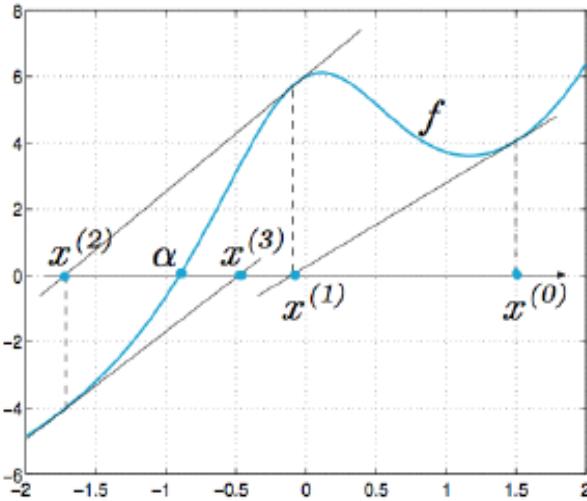
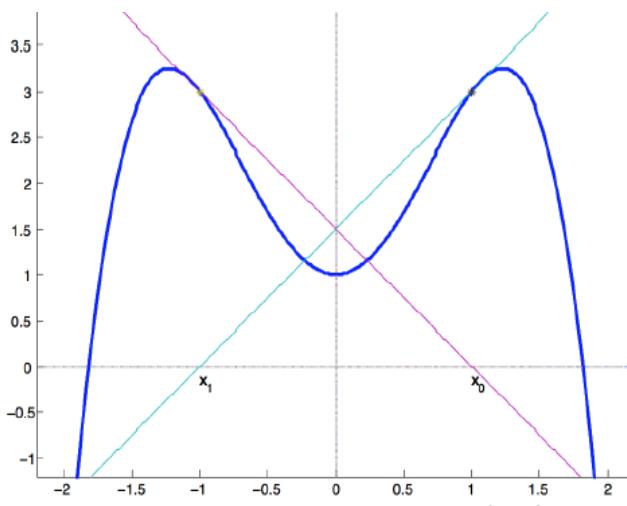
5.2.1 Συμπεράσματα σύγκρισης μεθόδων διχοτόμησης, Newton – Raphson, τέμνουσας

Συμπερασματικά για τη μέθοδο της διχοτόμησης θα πρέπει να αναφέρουμε ότι αν και συγκλίνει αργά (κάνει πολλές επαναλήψεις) σε μια προσεγγιστική λύση, είναι η μόνη μέθοδος στην οποία ξέρουμε **προκαταβολικά** μετά από πόσες επαναλήψεις θα βρούμε τη ζητούμενη προσεγγιστική λύση με μία συγκεκριμένη ακρίβεια που θέλουμε. **Η μέθοδος αυτή είναι σχετικά απλή και δεν αποτυγχάνει ποτέ στο να βρει μια λύση με μια δεδομένη ακρίβεια**, όταν η συνάρτηση f είναι συνεχής στο διάστημα $[a, b]$ και ισχύει το κριτήριο του Bolzano (σύμφωνα με το κριτήριο αυτό για να υπάρχει ρίζα μιας συνάρτησης $f(x)$ στο διάστημα $[a, b]$ πρέπει να ισχύει ότι $f(a) * f(b) < 0$), δηλαδή η μέθοδος αυτή συγκλίνει πάντα στο διάστημα που εφαρμόζεται.



Εικόνα 17: Μέθοδος διχοτόμησης

Συμπερασματικά για τη μέθοδο Newton-Raphson θα πρέπει να αναφέρουμε ότι αν η αρχική τιμή (εκτίμηση) που της δίνουμε δεν είναι κατάλληλη (δηλ. κοντά στη ρίζα) ή η συνάρτηση δεν είναι ομαλή στην περιοχή της ρίζας, τότε αποκλίνει. **Αν όμως συγκλίνει, τότε έχει μεγάλη ταχύτητα σύγκλισης.** Ένα άλλο μειονέκτημα που έχει είναι ότι απαιτεί δύο συναρτησιακούς υπολογισμούς σε κάθε επανάληψη (ένα για τη συνάρτηση και ένα για την παράγωγό της). Επίσης, απαιτεί τον υπολογισμό των παραγώγων της συνάρτησης, που σε μερικές περιπτώσεις αποτελεί πρόβλημα.

Εικόνα 18: Μέθοδος Newton-Raphson για τη συνάρτηση $f(x)=x+e^x+10/(1+x^2)-5$ 

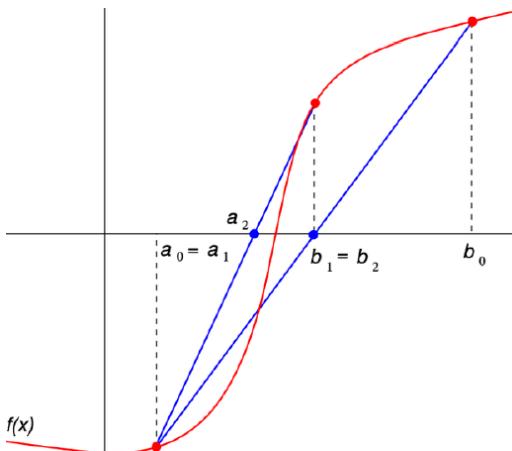
Εικόνα 19: Αστοχία μεθόδου Newton-Raphson

Συμπερασματικά για τη μέθοδο της τέμνουσας, θα πρέπει να αναφέρουμε ότι έχει προβλήματα σύγκλισης ανάλογα με εκείνα της μεθόδου Newton-Raphson. Η ταχύτητά της είναι αρκετά καλή σε σχέση με αυτή της μεθόδου διχοτόμησης, αλλά μικρότερη από αυτή της μεθόδου Newton-Raphson. Αν η f' είναι γνωστή, η μέθοδος της τέμνουσας είναι σε κάθε επανάληψη **οικονομικότερη** από εκείνη της μεθόδου Newton-Raphson, αφού σε κάθε υπολογισμό απαιτεί μόνο τον υπολογισμό της $f(x_k)$, διότι η τιμή $f(x_{k-1})$ έχει ήδη υπολογιστεί στην προηγούμενη επανάληψη. Όμως, επειδή η μέθοδος της τέμνουσας έχει μικρότερη τάξη σύγκλισης, **χρειάζεται περισσότερες επαναλήψεις** για να πετύχει ένα συγκεκριμένο μέγεθος σφάλματος. Επομένως το ποια από τις δύο μεθόδους (**τέμνουσα και Newton-Raphson**) είναι συνολικά **οικονομικότερη**, καθορίζεται από το κόστος υπολογισμού της f' σε σχέση με το κόστος υπολογισμού της f . Αν το κόστος είναι ίδιο τότε η μέθοδος της τέμνουσας είναι συνολικά **πιο οικονομική**. **Η μέθοδος τέμνουσας είναι πιο οικονομική από τη μέθοδο Newton – Raphson καθώς απαιτεί ένα μόνο συναρτησιακό υπολογισμό σε κάθε βήμα.** Όμως είναι πιο αργή στη σύγκλιση.

Μια άλλη μέθοδος που παράγει προσεγγιστικές λύσεις με τον ίδιο περίπου τρόπο όπως και η μέθοδος της τέμνουσας, είναι η μέθοδος της **εσφαλμένης θέσης (regula falsi)**. Η διαφορά της μεθόδου αυτής από εκείνη της τέμνουσας είναι ότι μας παρέχει έναν έλεγχο (χρησιμοποιώντας το θεώρημα του Bolzano) ώστε να εξασφαλίζεται ότι η ρίζα βρίσκεται μεταξύ δύο διαδοχικών προσεγγίσεων. Ο τύπος της **regula-falsi** είναι:

$$x_{k+1} = x_k - \frac{(x_k - b)}{f(x_k) - f(b)} * f(x_k), k = 1, 2, \dots$$

Παρατηρούμε στον τύπο αυτό ότι το σημείο $x = b$ παραμένει **σταθερό**, ενώ το άλλο σημείο ανανεώνεται σε κάθε βήμα. Η μέθοδος εσφαλμένης θέσης προϋποθέτει ότι γνωρίζουμε κάποιο αρχικό διάστημα, έστω (x_0, x_1) όπου $f(x_0) * f(x_1) < 0$, οπότε υπάρχει λύση σύμφωνα το θεώρημα Bolzano (όπως στη διχοτόμηση).



Εικόνα 20: Μέθοδος Regula-Falsi

5.2.2. Πλεονεκτήματα/Μειονεκτήματα μεθόδου τέμνουσας (παλιό θέμα)

Να αναφέρετε 2 βασικά πλεονεκτήματα και 2 βασικά μειονεκτήματα της μεθόδου τέμνουσας (secant) σε σύγκριση με τη μέθοδο διχοτόμησης για την επίλυση γραμμικών εξισώσεων.

Λύση

Πλεονεκτήματα μεθόδου τέμνουσας συγκριτικά με μέθοδο διχοτόμησης:

- 1) Όταν συγκλίνει, η σύγκλιση είναι **υπεργραμμική** (βρίσκεται ανάμεσα στη γραμμική και την τετραγωνική σύγκλιση), ενώ της μεθόδου διχοτόμησης είναι γραμμική.
- 2) Γενικεύεται για την επίλυση συστημάτων μη γραμμικών εξισώσεων, ενώ η διχοτόμηση όχι.

Μειονεκτήματα μεθόδου τέμνουσας συγκριτικά με μέθοδο διχοτόμησης:

- 1) Σε κάθε βήμα χρησιμοποιείται πολύ λίγη πληροφορία για τη συνάρτηση (πρόσημα στα άκρα των διαστημάτων που εγκλείσουν τη λύση).
- 2) Η μέθοδος διχοτόμησης εξασφαλίζει σύγκλιση προς τη λύση, εφόσον έχουμε εκκινήσει από διάστημα στα άκρα του οποίου υπάρχει εναλλαγή προσήμου, ενώ στην μέθοδο τέμνουσας τα σημεία εκκίνησης πρέπει να είναι κοντά σε ρίζα. Είναι πολύ πιο εύκολη η επιλογή ενός διαστήματος (όπου προφανώς ικανοποιείται το θεώρημα Bolzano) από ότι η επιλογή μεμονωμένων σημείων (δύο στην περίπτωση της τέμνουσας) που πρέπει να βρίσκονται κοντά σε ρίζα.

5.3 Ασκήσεις με διχοτόμηση και Newton – Raphson και τέμνουσα

Ασκηση 1

Να βρεθεί με χρήση επαναληπτικής μεθόδου χαρακτηριζόμενης από **τετραγωνική σύγκλιση**, η προσέγγιση της πραγματικής ρίζας της εξίσωσης $f(x) = 2x^3 - 3x^2 + 2x - 9$, με ακρίβεια τριών δεκαδικών ψηφίων.

Λύση

Αναζητάμε πρώτα περιοχή της πραγματικής ρίζας, εφαρμόζοντας το θεώρημα **Bolzano**. Παρατηρούμε ότι $f(2) = -1$ και $f(3) = 24$. Άρα στο διάστημα $[2, 3]$ υπάρχει τουλάχιστον μια πραγματική ρίζα, επειδή ακριβώς το γινόμενο $f(2) * f(3) < 0$. Για να βεβαιωθούμε στη συνέχεια ότι έχουμε **μόνο μια πραγματική ρίζα στο διάστημα αυτό**, βρίσκουμε την πρώτη παράγωγο: $f'(x) = 6x * (x - 1) + 2 > 0$. Επειδή αυτή είναι θετική $\forall x \in [2, 3]$ (δηλαδή γνησίως αύξουσα, το ίδιο όμως θα ίσχυε και για γνησίως φθίνουσα), σημαίνει ότι υπάρχει μόνο μια ρίζα στο διάστημα αυτό.

Παρατήρηση: Περισσότερες από μια ρίζες θα είχαμε αν στο διάστημα $[\alpha, \beta]$ ίσχυε ότι $f'(\alpha) * f(\beta) < 0$

Επειδή θέλουμε **τετραγωνική σύγκλιση**, χρησιμοποιούμε τη μέθοδο **Newton – Raphson**. Για να εφαρμόσουμε τη μέθοδο αυτή, πρέπει **πρώτα** να βρούμε μια **αρχική προσέγγιση της ρίζας** με τη μέθοδο της διχοτόμησης. Με δεδομένο ότι το διάστημα είναι το $[2, 3]$ παίρνουμε ως αρχική προσέγγιση το μέσο του διαστήματος, δηλαδή το σημείο $x = 2.5$. Στη συνέχεια εστιάζουμε στο αν η ρίζα είναι στο διάστημα $[2, 2.5]$ ή στο διάστημα $[2.5, 3]$. Το γινόμενο $f(2) = -1$ και $f(2.5) = 8.5$. Άρα στο διάστημα $[2, 2.5]$ υπάρχει πραγματική ρίζα, επειδή $f(2) * f(2.5) < 0$ και παίρνουμε **ως νέα προσέγγιση της ρίζας το 2.25**. Αυτό το σημείο θα χρησιμοποιηθεί ως αρχική προσέγγιση στη **Newton – Raphson**, της οποίας ο γενικός τύπος: $x_n = x_{n-1} - \frac{f(x_{n-1})}{f'(x_{n-1})}$

Παρατήρηση: Αν θέλουμε να βρούμε μια καλύτερη προσέγγιση της ρίζας, κάνουμε εκ' νέου διχοτόμηση. Γενικότερα, όσο πιο κοντά είναι η αρχική προσέγγιση στη ρίζα, τόσο λιγότερες επαναλήψεις απαιτούνται στη συνέχεια με την **Newton – Raphson**.

Για $k = 1$: $x_1 = x_0 - \frac{f(x_0)}{f'(x_0)} = 2.25 - \frac{3.0938}{18.8750} = 2.0861$. Με σχετικό σφάλμα: $|e_1| = \left| \frac{2.0861 - 2.25}{2.25} \right| = 0.0728$, για να γίνει αποδεκτό το σφάλμα, θα πρέπει να είναι μικρότερο από την επιθυμητή ακρίβεια. Το απόλυτο σφάλμα είναι $|e_1| = |x_k - x_{k-1}| = |x_1 - x_0| = 0.1639$, το οποίο **δεν** ισχύει ότι είναι μικρότερο από το $10^{-3} = 0.001$, άρα θα συνεχίσουμε τις επαναλήψεις.

Για $k = 2$: $x_2 = x_1 - \frac{f(x_1)}{f'(x_1)} = 2.0861 - \frac{0.2734}{15.5943} = 2.0686$ και το απόλυτο σφάλμα $|e_2| = |x_2 - x_1| = 0.0175 > 10^{-3}$

Για $k = 3$: $x_3 = x_2 - \frac{f(x_2)}{f'(x_2)} = 2.0686 - \frac{0.0034}{15.2630} = 2.0684$ και το απόλυτο σφάλμα $|e_3| = |x_3 - x_2| = 0.0002 < 10^{-3}$

Άρα η ρίζα x_3 αποτελεί τη ζητούμενη προσέγγιση.

Παρατήρηση: Αν έπρεπε στη μέθοδο της διχοτόμησης να βρούμε το **ελάχιστο πλήθος των επαναλήψεων** για την εύρεση μιας προσεγγιστικής λύσης με ακρίβεια ϵ , τότε θα χρησιμοποιούσαμε τον τύπο: $k = \lceil \log_2((b - a) * \epsilon^{-1}) \rceil = \log_2 \frac{(3-2)}{10^{-3}} = \log_2 10^3 = \frac{\log_{10}(1*10^3)}{\log_{10} 2} = \frac{3}{0.301} = [9.966] = 10$.

Άσκηση 2

Να βρεθεί η ποσότητα $7^{1/3}$ με διχοτόμηση, με ακρίβεια $\varepsilon = 10^{-3}$

- ο μέγιστος αριθμός επαναλήψεων με διχοτόμηση
- η τελική προσέγγιση της ρίζας
- η τρίτη προσέγγιση της ρίζας με διχοτόμηση

Λύση

a) Η 3^η ρίζα του 7 αποτελεί λύση της εξίσωσης $f(x) = x^3 - 7 = 0$ και ένα διάστημα στο οποίο ικανοποιείται το κριτήριο του Bolzano είναι το $[0, 3]$, όπου $f(0) * f(3) < 0$ και για να βρούμε τον αριθμό των επαναλήψεων που απαιτούνται για να πετύχουμε ακρίβεια $\varepsilon = 10^{-3}$ χρησιμοποιούμε τον τύπο: $k = \lceil \log_2((b-a) * \varepsilon^{-1}) \rceil = \lceil \log_2((3-0) * (10^{-3})^{-1}) \rceil \lceil \log_2((3-0) * 10^3) \rceil = \lceil \frac{\log_{10}(3*10^3)}{\log_{10}2} \rceil = \lceil 11.55 \rceil = 12$ επαναλήψεις.

b) Κάνουμε διαδοχικές διχοτομήσεις, μέχρι να βρούμε το x και μετά από 12 συνολικά επαναλήψεις θα έχουμε ότι η τελευταία διχοτόμηση θα μας δώσει τη ρίζα, που είναι: $x = \frac{a_{12} + b_{12}}{2} = \frac{1,9116 + 1,9130}{2} = 1.9123$.

c) Αν θεωρήσουμε ως διάστημα στο οποίο υπάρχει η ρίζα το διάστημα $[1.5, 2]$ για να έχουμε μεγαλύτερη ακρίβεια στον υπολογισμό της ρίζας, και όχι το διάστημα $[0, 3]$ που είχαμε θεωρήσει νωρίτερα, τότε το γινόμενο $f(1.5) * f(2) < 0$, οπότε στο διάστημα αυτό υπάρχει ρίζα και το νέο μέσο (το οποίο βρίσκεται στο υποδιάστημα $[1.5, 2]$) είναι το σημείο $x_1 = \frac{1.5+2}{2} = 1.75$, αφού το $f(x_1) * f(b_1) = -1.646 < 0$. Η $f'(x) = 3x^2 > 0$, $\forall x \in [1.75, 2]$ και το νέο μέσο (το οποίο βρίσκεται στο υποδιάστημα $[1.75, 2]$) είναι το σημείο $x_2 = \frac{1.75+2}{2} = 1.875$, $f(x_2) * f(b_2) = -0.408 < 0$. Το νέο μέσο (στο υποδιάστημα $[1.875, 2]$) είναι το σημείο $x_3 = \frac{1.875+2}{2} = 1.9375$.

Άσκηση 3

Με εφαρμογή της μεθόδου Newton – Raphson να προσεγγιστεί η ρίζα $17^{1/3}$ με ακρίβεια τεσσάρων δεκαδικών ψηφίων για τη συνάρτηση $f(x) = x^3 - 17 = 0$.

Λύση

Στο διάστημα $[2, 3]$ ισχύει το θεώρημα του Bolzano και επειδή $f(2.5) * f(3) < 0$ και επειδή $f'(x) = 3x^2 > 0$, $\forall x \in [2, 3]$ θα πάρουμε το διάστημα $[2.5, 3]$. Επειδή $f(2.5) * f(2.75) < 0$, η ρίζα θα είναι στο υποδιάστημα $[2.5, 2.75]$ και η νέα προσέγγισή της θα είναι το σημείο 2.6115 και μετά από μια ακόμα διχοτόμηση θα πάρουμε ως $x_0 = 2.6$. Οπότε στη συνέχεια χρησιμοποιούμε τη μέθοδο Newton – Raphson για $x_0 = 2.6$ και $x_1 = x_0 - \frac{f(x_0)}{f'(x_0)} = 2.6 - \frac{f(2.6)}{f'(2.6)} = 2.571$ και το απόλυτο σφάλμα $e_1 = |2.571 - 2.6| = 0.029$ και μετά από μία ακόμη επανάληψη, το $x_2 = x_1 - \frac{f(x_1)}{f'(x_1)} = 2.571 - \frac{f(2.571)}{f'(2.571)} = 2.571$. Άρα με το δεύτερο βήμα της Newton – Raphson παρατηρούμε ότι βρήκαμε την ακριβή ρίζα, αφού το σφάλμα είναι μηδέν, διότι το $e_2 = |2.571 - 2.571| = 0$. Το γεγονός ότι υπολογίσαμε την ακριβή εκτίμηση της ρίζας μόνο με δύο βήματα της Newton – Raphson σημαίνει ότι ήταν πολύ καλή η αρχική εκτίμηση της ρίζας από τη μέθοδο της διχοτόμησης.

Άσκηση 4 (παλιό θέμα)

α) Σας δίνονται τα σημεία $\{-20, 20\}$. Να επιλέξτε ένα από αυτά, έστω d , ώστε να μπορείτε να εκκινήσετε τη μέθοδο της διχοτόμησης για την προσέγγιση μιας πραγματικής ρίζας της f (είναι βέβαιο ότι υπάρχει τουλάχιστον μία) χρησιμοποιώντας το διάστημα $[\min(0, d), \max(0, d)]$. Στη συνέχεια να εφαρμόσετε **τρία** βήματα της μεθόδου διχοτόμησης και να υπολογίσετε το (μικρότερο) διάστημα που προκύπτει και εγκλείει τουλάχιστον μια πραγματική ρίζα της f . Η $f(x) = 4 * x^3 - 2 * x^2 - x - 4$.

β) Χρησιμοποιώντας το διάστημα το οποίο βρήκατε μετά τα τρία βήματα της μεθόδου διχοτόμησης, να εφαρμόσετε ένα βήμα της μεθόδου **τέμνουσας** για να υπολογίσετε νέα προσέγγιση της f .

Λύση

α) Ζητάμε να επιλέξουμε ένα από τα σημεία του διαστήματος $\{-20, 20\}$, έστω d , έτσι ώστε η μέθοδος της διχοτόμησης να μπορεί να εφαρμοστεί, χρησιμοποιώντας το διάστημα $[\min(0, d), \max(0, d)]$. Έστω ότι $d = 20$, οπότε τότε $f(0) * f(20) = -4 * 31176 < 0$. **Άρα ικανοποιείται το θεώρημα του Bolzano** και στο διάστημα $[0, 20]$ υπάρχει τουλάχιστον μια πραγματική ρίζα. **Διχοτομούμε** το διάστημα αυτό στο σημείο 10 (μέσο) και ψάχνουμε αν η ρίζα υπάρχει στο άνω ή στο κάτω μισό του διαστήματος. Παρατηρούμε ότι το γινόμενο $f(0) * f(10) = -4 * 3786 < 0$, επομένως η ρίζα υπάρχει στο διάστημα $[0, 10]$. Άρα ικανοποιείται και πάλι το θεώρημα του Bolzano και στο διάστημα $[0, 10]$ υπάρχει τουλάχιστον μια πραγματική ρίζα. Κάνοντας μια νέα διχοτόμηση στο διάστημα αυτό (στη σημείο 5) και ισχύει ότι το γινόμενο $f(0) * f(5) = -4 * 441 < 0$. Αυτό σημαίνει ότι η ρίζα είναι στο διάστημα $[0, 5]$. Άρα η νέα προσέγγιση της ρίζας μετά από **τρία** βήματα εφαρμογής της μεθόδου διχοτόμησης είναι το **2.5**. Άρα έγιναν τρεις διχοτομήσεις: $10 \rightarrow 5 \rightarrow 2.5$.

β) Από πριν βρήκαμε το διάστημα $[0, 5]$ και με βάση αυτό χρησιμοποιούμε τη μέθοδο της τέμνουσας, για να υπολογίσουμε μια νέα προσέγγιση της ρίζας της f . Ο τύπος της **τέμνουσας** (που απαιτεί δύο προηγούμενες εκτιμήσεις της ρίζας και όχι μόνο μία, όπως απαιτεί η μέθοδος Newton – Raphson) είναι:

$$x_{k+1} = x_k - \frac{(x_k - x_{k-1})}{f(x_k) - f(x_{k-1})} * f(x_k) = 2.5 - \frac{2.5 - 5}{f(2.5) - f(5)} * f(2.5) = 2.5 - \frac{2.5 - 5}{f(2.5) - f(5)} * f(2.5) = 2.5 + \frac{2.5}{43.5 - 441} * 43.5 = 2.2264.$$

Άσκηση 5

α) Ποιο είναι ένα **πλεονέκτημα** της μεθόδου Newton-Raphson σε σχέση με τη μέθοδο διχοτόμησης για την επίλυση μιας μη γραμμικής εξίσωσης.

β) Δίνεται ο τύπος $\xi_n = \frac{1}{2}(\xi_{n-1} + \frac{y}{\xi_{n-1}})$. Να δείξετε ότι υλοποιεί τη μέθοδο Newton – Raphson για τον υπολογισμό του \sqrt{y} .

γ) Να δείξτε ότι αν $e_n = |\sqrt{y} - \xi_n|$, τότε $e_{n+1} = \left| \frac{1}{2\xi_n} \right| * (e_n)^2$

δ) Να εξηγήσετε αν η ακολουθία ξ_n θα συγκλίνει, αν το ξ_0 επιλεγεί θετικό.

Λύση

α) Όταν η μέθοδος Newton-Raphson συγκλίνει, τότε έχει μεγαλύτερη ταχύτητα σύγκλισης από τη μέθοδο της διχοτόμησης. Αυτό συμβαίνει διότι η Newton-Raphson έχει **τετραγωνική σύγκλιση**.

β) Μια εφαρμογή της μεθόδου Newton-Raphson **είναι η χρήση της για τον υπολογισμό της τετραγωνικής ρίζας ενός αριθμού γ**. Πιο συγκεκριμένα, η \sqrt{y} αποτελεί λύση (ρίζα) της εξίσωσης $f(x) = x^2 - y = 0$. Η $f'(x) = 2x$

και η εφαρμογή της μεθόδου Newton-Raphson δίνει ως αποτέλεσμα το εξής: $x_n = x_{n-1} - \frac{f(x_{n-1})}{f'(x_{n-1})} = x_{n-1} - \frac{x_{n-1}^2 - y}{2*x_{n-1}}$ = κάνουμε τα κλάσματα ομώνυμα και έχουμε $\frac{2*x_{n-1}^2 - x_{n-1}^2 + y}{2*x_{n-1}} = \frac{x_{n-1}^2 + y}{2*x_{n-1}} = \frac{1}{2} (x_{n-1} + \frac{y}{x_{n-1}}) = \frac{1}{2} (\xi_{n-1} + \frac{y}{\xi_{n-1}})$, αν

θέσουμε $\xi_{n-1} = x_{n-1}$.

γ) Αν θέσουμε $e_n = |\sqrt{y} - \xi_n| = |\xi_n - \sqrt{y}|$, τότε από τον τύπο Newton – Raphson $x_n = x_{n-1} - \frac{f(x_{n-1})}{f'(x_{n-1})}$ ή εναλλακτικά

$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$ και στη συνέχεια αφαιρούμε από τα δύο μέλη της τελευταίας εξίσωσης το \sqrt{n} , οπότε: $x_{n+1} - \sqrt{n} = x_n - \frac{f(x_n)}{f'(x_n)} - \sqrt{n}$ = κάνουμε τα κλάσματα ομώνυμα και έχουμε: $\frac{f'(x_n)}{f'(x_n)} * x_n - \frac{f(x_n)}{f'(x_n)} - \frac{f'(x_n)}{f'(x_n)} * \sqrt{n} = \beta γάζουμε$ ως

κοινό παράγονται το $f'(x_n)$ και έχουμε: $\frac{f'(x_n)*(x_n - \sqrt{n}) - f(x_n)}{f'(x_n)}$. αλλάζουμε τη σειρά των όρων που αφαιρούνται,

βγάζοντας μπροστά ένα (-) και έχουμε: $\frac{-f'(x_n)*(x_n - \sqrt{n}) - f(x_n)}{f'(x_n)} = \beta γάζουμε$ κοινό παράγοντα το (-) από ολόκληρο

το κλάσμα: $-\frac{1}{f'(x_n)}(f(x_n) + f'(x_n) * (\sqrt{n} - x_n)) \Rightarrow x_{n+1} - \sqrt{n} = -\frac{1}{f'(x_n)}(f(x_n) + f'(x_n) * (\sqrt{n} - x_n))$ (1). Λαμβάνοντας υ-

πόψη ότι το ανάπτυγμα Taylor είναι: $f(\sqrt{n}) = f(x_0) + f'(x_0) * (\sqrt{n} - x_0) + \frac{f''(\xi)}{2!}(\sqrt{n} - x_0)^2 \Rightarrow f(\sqrt{n}) - \frac{f''(\xi)}{2!}(\sqrt{n} -$

$x_0)^2 = f(x_0) + f'(x_0) * (\sqrt{n} - x_0)$ (2), οπότε αντικαθιστώντας τον τύπο (2) στον τύπο (1) θα έχουμε, μετά από α-

ντικατάσταση του δεξιού μέλους της (1) με το αριστερό μέλος της (2): $x_{n+1} - \sqrt{n} = -\frac{1}{f'(x_n)}(f(\sqrt{n}) - \frac{f''(\xi)}{2!}(\sqrt{n} -$

$x_0)^2) \Rightarrow x_{n+1} - \sqrt{n} = \frac{f''(\xi)}{2f'(x_n)}(\sqrt{n} - x_0)^2 \Rightarrow e_{n+1} = x_{n+1} - \sqrt{n} = \left| \frac{f''(\xi)}{2f'(x_n)} \right| e_n^2$, διότι η διαφορά $(\sqrt{n} - x_0)^2 = e_n^2$. Ο

όρος $f(\sqrt{n}) = 0$, διότι το \sqrt{n} αποτελεί ρίζα της εξίσωσης που μελετάμε, δηλ. της εξίσωσης $f(x) = x^2 - y = 0$. Θέτουμε

απόλυτη τιμή για να εξασφαλίσουμε ότι το **κλάσμα είναι πάντα θετικό**, διότι το e_{n+1} είναι σφάλμα και **δεν** μπορεί να πάρει αρνητικές τιμές. Λαμβάνοντας υπόψη τον τελευταίο τύπο, έχουμε ότι το $f(x) = x^2 - y$, $f'(x) = 2x$,

$f''(x) = 2$ και επομένως: $e_{n+1} = \frac{1}{2} \left| \frac{2}{2x_n} \right| e_n^2 = \left| \frac{1}{2x_n} \right| e_n^2 = \left| \frac{1}{2\xi_n} \right| e_n^2$ αν αντικαταστήσουμε το x με ξ_n .

δ) Αν το ξ_0 επιλεγεί **θετικό** τότε από την ακολουθία: $\xi_n = \frac{1}{2}(\xi_{n-1} + \frac{y}{\xi_{n-1}})$, οι όροι συνεχώς θα αυξάνονται με

αποτέλεσμα το σφάλμα $e_n = |\sqrt{y} - \xi_n|$ συνεχώς να αυξάνεται. Έτσι έχουμε **απόκλιση** της ακολουθίας, διότι

δεν πρόκειται ποτέ να γίνει μικρότερη από κάποια συγκεκριμένη ακρίβεια, υπό την προϋπόθεση ότι $\sqrt{y} < \xi_n$.

Στην περίπτωση που το $\sqrt{y} > \xi_n$, όσο αυξάνεται το n , τόσο το ξ_n θα πλησιάζει την επιθυμητή τιμή που είναι

το \sqrt{y} με αποτέλεσμα τόσο μεγαλύτερη σύγκλιση να επιτυγχάνεται.

Άσκηση 6

Το πλήθος επαναλήψεων στη μέθοδο διχοτόμησης για την επίλυση μη γραμμικής εξίσωσης $f(x) = 0$ εξαρτάται κυρίως: α) από τις πολλαπλότητες των ριζών β) από το είδος της συνάρτησης γ) Από το μέγεθος του αρχικού διαστήματος δ) από τη ζητούμενη ακρίβεια.

Λύση

Η μέθοδος διχοτόμησης είναι μέθοδος ολικής σύγκλισης, δηλ. δεν απαιτείται αρχική εκτίμηση της λύσης που να βρίσκεται πλησίον της λύσης και συγκλίνει πάντα στο διάστημα που εφαρμόζεται. Επίσης συγκλίνει αργά (γραμμικά) στην προσεγγιστική λύση. Το πλήθος των επαναλήψεων της εξαρτάται κυρίως από τη ζητούμενη ακρίβεια και λιγότερο από το μέγεθος του αρχικού διαστήματος. Επομένως η σωστή απάντηση είναι η (δ).

Άσκηση 7

Δίνεται η εξίσωση $x^3 + 4x^2 = 10$. Εφαρμόζουμε τη μέθοδο Newton για να βρούμε τη λύση αυτής. Ξεκινώντας από το $x_0 = 2$, στο **δεύτερο** βήμα βρίσκουμε $x_2 =$

- a) καμία από τις υπόλοιπες b) 1.5000 c) 1.3733 d) 1.3653

Λύση

Επειδή ζητάμε λύση εξίσωσης που **δεν** είναι πρώτου βαθμού, θα χρησιμοποιήσουμε τη μέθοδο Newton – Raphson, η οποία για το πρώτο βήμα της επανάληψης δίνει το εξής:

Για $k = 1$: $x_1 = x_0 - \frac{f(x_0)}{f'(x_0)} = 2 - \frac{f(2)}{f'(2)} = 2 - \frac{14}{28} = 1.5$. Για $k = 2$: $x_2 = x_1 - \frac{f(x_1)}{f'(x_1)} = 1.5 - \frac{2.375}{18.75} = 1.3733$. Άρα η σωστή απάντηση είναι η (c).

Άσκηση 8

Δίνεται η εξίσωση $x^3 + 6x = 5$. Εφαρμόζουμε τη μέθοδο Newton για να βρούμε τη λύση αυτής. Ξεκινώντας από $x_0 = 2$, στο **δεύτερο** βήμα βρίσκουμε (χρειάζεται για να δείξετε όλα τα βήματα) $x_2 =$

- a): 0.7602 b): 1.2727 c): καμία από τις υπόλοιπες τιμές d): -0.3124

Λύση

Οι ερωτήσεις στο θέμα αυτό αφορούσαν σε μεθόδους επίλυσης μη γραμμικών εξισώσεων. Για παράδειγμα, στη μέθοδο Newton και τη χρήση της για τον υπολογισμό ρίζας απλών βαθμωτών μη γραμμικών εξισώσεων (και ειδικότερα, πολυωνύμων). Η μέθοδος **Newton (εννοείται Newton – Raphson)** για την εύρεση απλής ρίζας της εξίσωσης $f(x) = 0$ συνιστάται στην εφαρμογή της αναδρομικής σχέσης $x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)}$. Επομένως (εδώ παραλείπουμε τις λεπτομέρειες των ενδιάμεσων πράξεων) $x_{k+1} = x_k - \frac{x_k^3 + 6*x_k - 5}{3*x_k^2 + 6}$, οπότε με αρχική τιμή $x_0 = 2$ έπεται ότι $x_1 = 1.1667$ και $x_2 = 0.8109$. Άρα η σωστή απάντηση είναι η (c).

Άσκηση 9

Να εφαρμόσετε τη μέθοδο Newton – Raphson και να τη συγκρίνετε με τη μέθοδο της τέμνουσας (secant) για την επίλυση της εξίσωσης $x^5 = 18$. Ξεκινήστε από την εκτίμηση $x_0 = 2$ για τη Newton και το επιπρόσθετο στοιχείο $x_1 = 0$ για την τέμνουσα. Σχετικά με την προσέγγιση της λύσης στο **δεύτερο βήμα κάθε μεθόδου**, κυκλώστε μια από τις παρακάτω απαντήσεις: a) $x_2^{(Newton)} = 1.7845$ και $x_2^{(secant)} = 1.59435$, b) $x_2^{(Newton)} = 1.8250$ και $x_2^{(secant)} = 1.1250$ c) $x_2^{(Newton)} = 1.7845$ και $x_2^{(secant)} = 1.7826$.

Λύση

Από τη μέθοδο **Newton – Raphson** $x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)}$ και με δεδομένη την εξίσωση $f(x) = x^5 - 18$ και την αρχική προσέγγιση της ρίζας $x_0 = 2$, έχουμε ότι $x_1 = x_0 - \frac{f(x_0)}{f'(x_0)} = 2 - \frac{f(2)}{f'(2)} = 2 - \frac{14}{80} = 1.825$. Στο δεύτερο βήμα της μεθόδου έχουμε ότι το $x_2 = 1.825 - \frac{f(1.825)}{f'(1.825)} = 1.825 - \frac{2.2447}{55.465} = 1.7846$.

Αναφορικά με τη μέθοδο της **τέμνουσας (secant)** και με δεδομένες τις δύο αρχικές προσεγγίσεις $x_0 = 2$ και $x_1 = 0$ που δίνονται, έχουμε ότι από το γενικό τύπο: $x_{k+1} = x_k - \frac{(x_k - x_{k-1})}{f(x_k) - f(x_{k-1})} * f(x_k)$, η ρίζα στο πρώτο βήμα εφαρμογής της μεθόδου είναι: $x_1 = 2 - \frac{(2-0)}{f(2)-f(0)} * f(2) = 2 - \frac{(2-0)}{32} * 14 = 1.125$. Στο δεύτερο βήμα εφαρμογής της μεθόδου έχουμε ότι: $x_2 = 1.125 - \frac{(1.125-2)}{f(1.125)-f(2)} * f(1.125) = 1.125 - \frac{-0.875}{-16.1979-14} * (-16.1979) = 1.125 - \frac{-0.875}{-30.1979} * (-16.1979) = 1.594$. Άρα η σωστή απάντηση είναι η (α).

Άσκηση 10

Μας ενδιαφέρει ο υπολογισμός της πιο μικρής πραγματικής ρίζας $\hat{\xi}$ της συνάρτησης $f(\xi) = +4\xi^3 - 8\xi^2 - 3\xi + 1$. Μπορείτε να θεωρήσετε ότι η ρίζα βρίσκεται στο διάστημα $[\xi_1, \xi_4]$ που ορίζεται βάσει των τιμών της ερώτησης (1α').

- (1') Αφού εγκλωβίσετε την $\hat{\xi}$ σε κατάλληλο υποδιάστημα, να δώσετε τον **μέγιστο αριθμό επαναλήψεων** της μεθόδου **διχοτόμησης** για την προσέγγισή της με απόλυτο σφάλμα μικρότερο του 10^{-6} .
- (2') Για την προσέγγιση της $\hat{\xi}$ θεωρούμε τώρα την επανάληψη Newton - Raphson. Δώστε τον επαναληπτικό τύπο της και τη συνθήκη τερματισμού με ακρίβεια 8 σημαντικών ψηφίων.
- (3') Έστω η μια ικανοποιητική προσέγγιση της $\hat{\xi}$. Αν k αρκετά μεγάλο, ισχύει $|e_k| = 2 \cdot 10^{-6}$ (e_k : απόλυτο σφάλμα), να δώσετε προσέγγιση του $|e_{k+1}|$ συναρτήσει του η .

Λύση

(1') Σύμφωνα με το θεώρημα Bolzano, υπάρχει μια ρίζα της f στο διάστημα $[0, 1]$ που ορίζεται από την άσκηση 1α', αφού $f(0) = 1 > 0$ και $f(1) = -6 < 0$, και έτσι περιορίζουμε το διάστημα στο οποίο βρίσκεται η $\hat{\xi}$. Η τιμή που δίνεται, δηλ. το 10^{-6} αναφέρεται στην ακρίβεια ε που εμφανίζεται στον τύπο. Για να προσεγγίσουμε τη ρίζα με

τη μέθοδο της διχοτόμησης, θα χρησιμοποιήσουμε τον παρακάτω τύπο: $k \geq \lceil \log((1 - 0) * 10^6) \rceil \Rightarrow k \geq \lceil \log_2(10^6) \rceil \Rightarrow k \geq \lceil \frac{\log 10^6}{\log 2} \rceil \Rightarrow k \geq \lceil \frac{6}{0.301} \rceil \Rightarrow k \geq 20$ δηλαδή θα χρειαστούμε τουλάχιστον 20 επαναλήψεις για να προσεγγίσουμε τη ρίζα.

(2') Ο επαναληπτικός τύπος για τη μέθοδο Newton-Raphson είναι $x_i = x_{i-1} - \frac{f(x_{i-1})}{f'(x_{i-1})}$. Η συνθήκη τερματισμού με ακρίβεια 8 σημαντικών ψηφίων θα είναι $|e| = 5 \cdot 10^{-8}$. Επειδή η ακρίβεια είναι σε **σημαντικά ψηφία**, ο τύπος που δίνει το φράγμα του σφάλματος είναι $|\varepsilon| \leq 5 * 10^{-k} = 0.5 * 10^{1-k}$.

(3') Η τάξη σύγκλισης είναι τετραγωνική, εφόσον η ρίζα η είναι **απλή**. Για να είναι απλή μια ρίζα, θα πρέπει $f'(\eta) \neq 0$ στο διάστημα στο οποίο προσεγγίζεται (εδώ είναι το διάστημα $[0, 1]$), το οποίο σε αυτή την περίπτωση ισχύει. Επομένως η ρίζα είναι **απλή** και η ακολουθία που την προσεγγίζει θα **έχει τουλάχιστον τετραγωνική σύγκλιση**, άρα $|e_{k+1}| \simeq |e_k|^2 * \frac{f'(\eta)}{f''(\eta)} \Rightarrow |e_{k+1}| \simeq 4 \cdot 10^{-12} \frac{12 * \eta^2 - 16 * \eta - 3}{24 * n - 16}$.

5.4 Ρίζες πολυωνύμου – Συνοδευτικό μητρώο

Δίνεται πολυώνυμο βαθμού n της μορφής $p(x) = a_0 + a_1 z + a_2 z^2 + \dots + a_{n-1} z^{n-1} + a_n z^n$ για το οποίο ζητάμε να βρούμε τις ρίζες του. Τα πολυώνυμα αποτελούν **ειδικές περιπτώσεις μη - γραμμικών συναρτήσεων** και οι μέθοδοι εύρεσης ριζών τους μπορούν να χρησιμοποιήσουν τις ειδικές ιδιότητές τους. Ένα πολυώνυμο n - βαθμού έχει ακριβώς n ρίζες (όσες και ο βαθμός του). Για κάθε πολυώνυμο βαθμού n, υπάρχει μητρώο $A \in C^{nxn}$ με **χαρακτηριστικό πολυώνυμο⁴** ίδιο με το p που είναι $p(\lambda) = \det(\lambda * I - A) = 0 \Rightarrow p(\lambda) = a_n * \lambda^n + a_{n-1} * \lambda^{n-1} + \dots + a_1 * \lambda^1 + a_0$. Το μητρώο αυτό είναι το $A = \begin{bmatrix} 0 & 0 & \dots & 0 & -a_0 \\ 1 & 0 & \dots & 0 & -a_1 \\ 0 & 1 & 0 & \dots & -a_2 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 & -a_{n-1} \end{bmatrix}$ και ονομάζεται **συνοδευτικό μητρώο του πολυωνύμου p**.

Ισχύει ότι ο υπολογισμός των ριζών του πολυωνύμου βαθμού n = υπολογισμός ιδιοτιμών συνοδευτικού μητρώου. Για παράδειγμα, αν έχουμε το πρόβλημα υπολογισμού της μεγαλύτερης ρίζας του πολυωνύμου $p(z) = z^5 + 10 * z^4 - 245 * z^3 + 430 * z^2 + 244 * z^1 - 440$, τότε το **συνοδευτικό μητρώο** είναι

$\begin{bmatrix} 0 & 0 & 0 & 0 & 440 \\ 1 & 0 & 0 & 0 & -244 \\ 0 & 1 & 0 & 0 & -430 \\ 0 & 0 & 1 & 0 & 245 \\ 0 & 0 & 0 & 1 & -10 \end{bmatrix}$. Οι ρίζες του χαρακτηριστικού πολυωνύμου του μητρώου A, δηλ. του $\det(\lambda * I - A) = 0$ είναι οι ιδιοτιμές του μητρώου A. Αυτό σημαίνει ότι αν γνωρίζουμε/προσεγγίζουμε τις ιδιοτιμές ενός μητρώου, ουσιαστικά γνωρίζουμε/προσεγγίζουμε τις ρίζες του. Με τον όρο **μονικό πολυώνυμο** εννοούμε κάθε πολυώνυμο με συντελεστή μεγιστοβάθμιου όρου το «1». Επομένως, **κάθε χαρακτηριστικό πολυώνυμο είναι μονικό**.

⁴ Για εύρεση των ιδιοτιμών του μητρώου $A \in C^{nxn}$

Ο υπολογισμός των ρίζών πολυωνύμου με αναγωγή σε πρόβλημα ιδιοτιμών μέσω του συνοδευτικού μητρώου, είναι η μέθοδος που έχει επιλεγεί στη MATLAB. Μάλιστα μπορούμε να επεκτείνουμε την ιδέα όταν το πολυώνυμο δίνεται σε αναπαράσταση Newton. Για παράδειγμα, αν το π.π. είναι το $p(x) = \gamma_0 + \gamma_1(x - p_1) + \gamma_2(x - p_1)(x - p_2) + \gamma_3(x - p_1)(x - p_2)(x - p_3)$, και το συνοδευτικό μητρώο είναι το $C = \begin{bmatrix} p_1 & 1 & 0 \\ 0 & p_2 & 1 \\ -\gamma_0 & -\gamma_1 & -\gamma_2 + p_3\gamma_3 \end{bmatrix}$, τότε το $p(x) = \det(C - x * I)$, δηλαδή το χαρακτηριστικό πολυώνυμο του μητρώου C είναι το $p(x)$.

Παράδειγμα: Δίνεται ο παρακάτω κώδικας MATLAB, για τον οποίο ζητάμε να εξηγήσουμε τι μας δίνει

```
function [r, C] = com(u)
if (u(1)==0)
    error('The first coefficient must be not equal to zero');
end
n = length(u)-1; C = zeros(n); %Αρχικά το μητρώο C περιέχει μηδενικά, διαστάσεων n×n.
for i = 2:n
    C(i, i-1) = 1; %Δημιουργία των «1» στο συνοδευτικό μητρώο C από 1η στήλη έως προτελευταία
end
C(:, n) = -u(n+1:-1:2)/u(1); %Δημιουργία της τελευταίας στήλης του συνοδευτικού μητρώου
r = eig(C); %στο διάνυσμα r επιστρέφονται οι ιδιοτιμές του μητρώου C
```

Τι τιμές αναμένουμε να έχει το διάνυσμα εξόδου r , εάν έχουμε δώσει σαν είσοδο το διάνυσμα $\text{poly}([10 11 12 13 14 15])$;

Λύση

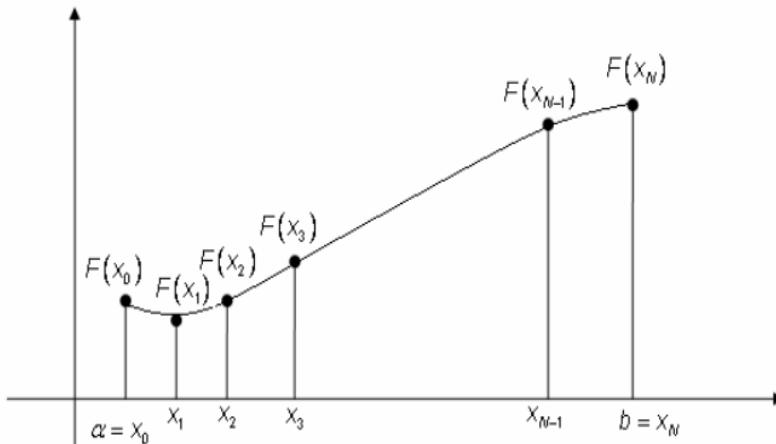
Η συγκεκριμένη συνάρτηση υπολογίζει το συνοδευτικό μητρώο C με βάση το διάνυσμα u , που δίνεται ως είσοδος και επίσης επιστρέφει και τις ιδιοτιμές του συνοδευτικού μητρώου. Επίσης, η συνάρτηση $\text{poly}(u)$ παίρνει ως όρισμα ένα διάνυσμα u που περιέχει τις ρίζες του πολυωνύμου και επιστρέφει τους συντελεστές του πολυωνύμου αυτού.

```
>> u = poly([10 11 12 13 14 15]) %το διάνυσμα u που δίνεται ως είσοδος στη function περιέχει συντελεστές πολυωνύμου
u =
     1      -75     2335    -38625    358024   -1763100    3603600
>> [r, C] = com(u)
r = %παρατηρούμε ότι οι ιδιοτιμές του C που επιστρέφονται στο διάνυσμα r ταυτίζονται με τις ρίζες του πολυωνύμου
10.0000
11.0000
12.0000
13.0000
14.0000
15.0000
C =
     0      0      0      0      0    -3603600
     1      0      0      0      0    1763100
     0      1      0      0      0   -358024
     0      0      1      0      0     38625
     0      0      0      1      0    -2335
     0      0      0      0      1       75
```

Κεφάλαιο 6 – Ολοκλήρωση μέσω προσεγγιστικών τύπων

6.1 Απλοί κανόνες παραλληλογράμμου, τραπεζίου, Simpson, μέσου σημείου

Για να υπολογίσουμε προσεγγιστικά ένα ολοκλήρωμα, επιλέγουμε $n + 1$ κόμβους και διαμερίζουμε το διάστημα $[a, b]$ σε επιμέρους υποδιαστήματα $[x_i, x_{i+1}]$ για $i = 0, \dots, n - 1$ και ονομάζουμε τους κόμβους $\{x_0, x_1, x_2, \dots, x_n\}$ διαμέριση του διαστήματος $[a, b]$.



Εικόνα 21: Διαμέριση του διαστήματος $[a, b]$ σε υποδιαστήματα

Επομένως το διάστημα αυτό αποτελείται από την ένωση όλων των επιμέρους υποδιαστημάτων διαμέρισης, σύμφωνα με τον τύπο: $[a, b] = \bigcup_{i=0}^{n-1} [x_i, x_{i+1}]$. Ονομάζουμε **λεπτότητα** της διαμέρισης το μέγεθος $h = \max_{i=0,1,\dots,n-1} \{x_{i+1} - x_i\}$.

- Ο τύπος που δίνει το ολοκλήρωμα μιας περιοχής σύμφωνα με τον κανόνα (τύπο) του **παραλληλογράμμου** είναι: $P(f) = \int_{x_i}^{x_{i+1}} f(x) dx \approx (x_{i+1} - x_i) * f(x_i)$, όπου $h = \text{απόσταση μεταξύ ισαπεχόντων κόμβων}$ στο διάστημα ολοκλήρωσης που έχουμε θεωρήσει. Συγκεκριμένα, $x_{i+1} - x_i = h = (b - a)/n$, όπου $n = \text{πλήθος υποδιαστημάτων που χωρίζουμε το διάστημα } [a, b]$.

- Ο τύπος που δίνει το ολοκλήρωμα μιας περιοχής σύμφωνα με τον κανόνα του **τραπεζίου** είναι: $T(f) = \frac{x_{i+1} - x_i}{2} * (f(x_i) + f(x_{i+1})) = \frac{h}{2} * (f(x_i) + f(x_{i+1}))$ και το **σφάλμα** είναι: $| \int_{x_i}^{x_{i+1}} f(x) dx - T(f) | \leq \frac{M}{12} (x_{i+1} - x_i)^3$, υπό την προϋπόθεση ότι η f έχει δεύτερη παράγωγο, δηλαδή $f \in C^{(2)}[x_i, x_{i+1}]$, οπότε υπάρχει M ώστε το: $\max_{x \in [x_i, x_{i+1}]} |f''(x)| \leq M$.

- Ο τύπος που δίνει το ολοκλήρωμα μιας περιοχής σύμφωνα με τον κανόνα (τύπο) του **Simpson** είναι:

$$S(f) = \frac{h}{6} (f(x_i) + 4f(\frac{x_i + x_{i+1}}{2}) + f(x_{i+1})) \text{ και το } \text{σφάλμα} \text{ του κανόνα αυτού είναι: } \int_a^b f(x) dx - S(f) = -\frac{f^{(4)}(\eta)}{2880} (x_{i+1} - x_i)^5.$$

- Ο τύπος **μέσου σημείου** είναι ο εξής: $M(f) = f(\frac{x_i + x_{i+1}}{2})(x_{i+1} - x_i)$ και το **σφάλμα** μέσου σημείου είναι: $| \int_{x_i}^{x_{i+1}} f(x) dx - M(f) | \leq \frac{K}{24} (x_{i+1} - x_i)^3$

Παρατήρηση: Στους προηγούμενους τύπους σφαλμάτων τα μεγέθη M , K θεωρούνται σταθερές ποσότητες και τα αφήνουμε με αυτό το συμβολισμό, εκτός και αν προσδιορίζονται με συγκεκριμένες τιμές από την εκφώνηση.

6.2 Σύνθετοι κανόνες παραλληλογράμμου, τραπεζίου, Simpson, μέσου σημείου

Το σφάλμα των τύπων αριθμητικής ολοκλήρωσης στις σύνθετες μεθόδους εξαρτάται από το μέγεθος του M (τάξη της μεθόδου είναι $M = m + 1$) και επιπλέον από το μήκος $b - a$ του διαστήματος $[a; b]$. Για πιο **ακριβή αριθμητική ολοκλήρωση**, μπορούμε να εφαρμόσουμε διαμέριση του αρχικού διαστήματος και να χρησιμοποιήσουμε τον τύπο ολοκλήρωσης σε κάθε υποδιάστημα της διαμέρισης (βασική ιδέα των σύνθετων κανόνων υπολογισμού ολοκληρωμάτων).

6.2.1. Σύνθετη μέθοδος παραλληλογράμμου

$$CP(f) = h \sum_{j=0}^{n-1} f_j$$

6.2.2. Σύνθετη μέθοδος τραπεζίου

Εφαρμόζοντας τον τύπο τραπεζίου (για μη ισαπέχοντες κόμβους) σε κάθε υποδιάστημα μιας διαμέρισης, ο σύνθετος τύπος είναι: $CT(f) = \sum_{i=0}^{n-1} \left(\frac{(x_{i+1} - x_i)}{2} \right) (f(x_i) + f(x_{i+1}))$. Είναι εμφανής η ομοιότητα που υπάρχει ανάμεσα στο σύνθετο και τον απλό κανόνα τραπεζίου, ουσιαστικά ο σύνθετος τύπος είναι ακριβώς ο απλός, υπολογισμένος όμως για ένα άθροισμα υποδιαστημάτων από $i = 0$ έως $i = n - 1$.

Χρησιμοποιώντας **ισαπέχοντες κόμβους** $h = x_{i+1} - x_i$ ($i = 0, \dots, n - 1$), **ο σύνθετος κανόνας του τραπεζιού για ισαπέχοντες κόμβους με λεπτότητα h είναι:** $CT(f) = \frac{h}{2} \left(f(x_0) + f(x_n) + 2 * \sum_{j=1}^{n-1} f_j \right)$. Τότε το **σφάλμα** της σύνθετης μεθόδου τραπεζίου είναι: $\int_a^b f(x) - CT(f) = -\frac{(b-a)f''(\eta)}{12} h^2$.

6.2.3 Σύνθετη μέθοδος Simpson

Εφαρμόζοντας τον τύπο Simpson σε **κάθε υποδιάστημα** μιας διαμέρισης με **ισαπέχοντες κόμβους** και χρησιμοποιώντας $h = (b-a)/s$, δηλ. $x_i = a + 2 * i * h$ για $i = 0, \dots, 2s+1$, κατασκευάζουμε σύνθετο τύπο Simpson. **Ο σύνθετος τύπος του Simpson για ισαπέχοντες κόμβους με λεπτότητα h είναι:**

$$CS(f) = \frac{h}{3} (f(x_0) + 4f(x_1) + 2f(x_2) + 4f(x_3) + 2f(x_4) + \dots + 2f(x_{n-2}) + 4f(x_{n-1}) + f(x_n)).$$

Τότε το **σφάλμα** της σύνθετης μεθόδου Simpson είναι: $\int_a^b f(x) - CS(f) = -\frac{(b-a)f''(\eta)}{180} h^4$.

6.2.4 Σύνθετη μέθοδος μέσου σημείου

$$CM(f) = h \sum_{i=0}^{n-1} f(a + h \frac{2i+1}{2})$$

Εφαρμόζοντας τη σύνθετη μέθοδο μέσου σημείου, έχουμε:

$$\left| \int_a^b f(x) - CM(f) \right| = \frac{(b-a)|f^{(2)}(\eta)|}{24} h^2 \leq \frac{(b-a)M}{24} h^2$$

6.2.5 Σύνθετες μέθοδοι: Σύνοψη

Θεωρούμε πως το διάστημα $[a, b]$ έχει διαμερισθεί με κόμβους:

$$a \leq x_0 < x_1 < \dots < x_{n-1} < x_n \leq b$$

σε υποδιαστήματα $[x_{j-1}, x_j]$, και εφαρμόζουμε Newton-Cotes σε κάθε υποδιάστημα. Στη μεθοδολογία Newton – Cotes υπολογίζουμε με τη βοήθεια πολυωνύμων παρεμβολής το ολοκλήρωμα ενός διαστήματος.

όνομα	Μη ισαπέχοντες κόμβοι	Ισαπέχοντες κόμβοι (σύνθετοι τύποι ολοκλήρωσης)	κόστος
παρ/γρμου	$\sum_{j=0}^{n-1} (x_{j+1} - x_j) * f_j$	$h \sum_{j=0}^{n-1} f_j$	n
Μέσου σημείου		$h \sum_{j=0}^{n-1} f(\alpha + h \frac{2i+1}{2})$	n
Τραπεζίου	$\sum_{j=0}^{n-1} (x_{j+1} - x_j) * \frac{f_j + f_{j+1}}{2}$	$\frac{h}{2} \left(f(a) + f(b) + 2 \sum_{j=1}^{n-1} f_j \right)$	n+1
Simpson	$\frac{1}{6} \sum_{j=0}^{n-1} (x_{j+1} - x_j) * \left[f(x_j) + 4f\left(\frac{x_j + x_{j+1}}{2}\right) + f_{j+1} \right]$	$CS(f) = \frac{h}{3} (f(x_0) + 4f(x_1) + 2f(x_2) + 4f(x_3) + 2f(x_4) + \dots + 2f(x_{n-2}) + 4f(x_{n-1}) + f(x_n))$	2n+1

Η δεύτερη στήλη του παραπάνω πίνακα υπολογίζει τους σύνθετους κανόνες ολοκλήρωσης για κόμβους μη – ισαπέχοντες και για αυτό δεν εμφανίζεται πουθενά το μέγεθος h (λεπτότητα). Η τρίτη στήλη του παραπάνω πίνακα υπολογίζει τους σύνθετους κανόνες ολοκλήρωσης για ισαπέχοντες κόμβους και συνεπώς καθένας από αυτούς περιέχει το μέγεθος h . Το μεγαλύτερο κόστος εφαρμογής το έχει η μέθοδος Simpson.

6.3 Μελέτη Σφάλματος

Έστω $f \in C^{(m+1)}[a, b]$. Δηλαδή αν μια συνάρτηση f είναι $m+1$ φορές παραγωγίσιμη στο διάστημα $[a, b]$ τότε μπορεί να γραφτεί ως ένα άθροισμα πολυωνύμου παρεμβολής $P_m(x)$ και ορισμένου ολοκληρώματος στο διάστημα αυτό, δηλαδή: $f(x) = \underbrace{f(a) + \frac{(x-a)}{1!} f'(a) + \dots + \frac{(x-a)^m}{m!} f^{(m)}(a)}_{P_m(x)} + \underbrace{\frac{1}{m!} \int_a^x f^{(m+1)}(s)(x-s)^m ds}_{\text{υπόλοιπο}}$

Θεώρημα μέσης τιμής για ολοκληρώματα

Έστω $f, g \in C[a, b]$, και $f \geq 0$ για $x \in [a, b]$. Τότε $\int_a^b f(s)g(s)ds = g(\xi) \int_a^b f(s)ds$, για κάποιο $\xi \in [a, b]$

Τάξη ακρίβειας κανόνα αριθμητικής ολοκλήρωσης

Ένας τύπος αριθμητικής ολοκλήρωσης έχει ακρίβεια βαθμού n αν παράγει ακριβή αποτελέσματα για όλα τα πολυώνυμα βαθμού $\leq n$ και δεν είναι ακριβής για πολυώνυμο βαθμού $n+1$. Ένας κανόνας αριθμητικής ολοκλήρωσης λέγεται ότι έχει τάξη ακρίβειας m , αν είναι ακριβής για κάθε πολυώνυμο p , για το οποίο $\deg p \leq m$.

Η τάξη της μεθόδου είναι $M = m + 1$. Εκ' κατασκευής έχουμε ότι:

- Κανόνας παραλληλόγραμμου: $M = 1 = m + 1$.
- Κανόνας της μέσης τιμής: $M = 2 = m + 1$.
- Κανόνας του τραπεζίου: $M = 2 = m + 1$.

Ο βαθμός ακρίβειας ενός ολοκληρώματος $I_n(f)$ είναι $n + 1$, όπου n είναι η ακρίβεια του πολυωνύμου παρεμβολής που χρησιμοποιείται στον υπολογισμό του ολοκληρώματος

$$I(f) = \int_a^b f(x)dx \simeq w_0 f_0 + w_1 f_1 + \dots + w_n f_n$$

6.4 Παράδειγμα ολοκλήρωσης με χρήση κανόνων παραλληλογράμμου, τραπεζίου, Simpson

Να προσεγγίσετε το ορισμένο ολοκλήρωμα της συνάρτησης $f(x) = \sqrt{x}$ στο διάστημα [1.00, 1.30].

- Με τον (απλό) κανόνα του παραλληλογράμμου
- Με τον (απλό) κανόνα του τραπεζίου
- Με τον (απλό) κανόνα Simpson

Λύση

Μεθοδολογία: Αρχικά πρέπει να επιλέξουμε **διαμέριση του χωρίου μας**. Συνήθως αν δεν αναφέρει η εκφώνηση τον ακριβή αριθμό των υποδιαστημάτων που θα χωρίσουμε το διάστημα ολοκλήρωσης, επιλέγουμε **δύο υποδιαστήματα** και αφού υπολογίσουμε το προσεγγιστικό ολοκλήρωμα με τον εκάστοτε κανόνα που ζητείται, στη συνέχεια ελέγχουμε την **ακρίβεια του αποτελέσματος σύμφωνα με μια δοθείσα ακρίβεια** και αν το αποτέλεσμα δεν μας ικανοποιεί, τότε το χωρίζουμε σε περισσότερα υποδιαστήματα και υπολογίζουμε εκ' νέου το νέο προσεγγιστικό ολοκλήρωμα. Στην προκειμένη περίπτωση επιλέγουμε -αυθαίρετα- να διαιρέσουμε το χωρίο μας σε **n = 5** **ισαπέχοντα διαστήματα**. Άρα θα έχουμε ότι η **απόσταση μεταξύ δύο διαδοχικών σημείων** του χωρίου θα είναι: $(x_{i+1} - x_i) = h = \frac{b-a}{n} = \frac{1.30-1.00}{5} = 0.06$.

A. Κανόνας Παραλληλογράμμου: Σε κάθε χωρίο εφαρμόζουμε τον κανόνα **παραλληλογράμμου**:

$P(f) = \int_{x_i}^{x_{i+1}} f(x) d(x) \approx (x_{i+1} - x_i) * f(x_i) = h * f(x_i)$, όπου $f(x_i)$ είναι η τιμή της συνάρτησης στην **αρχή** του κάθε υποδιαστήματος. Οπότε υπολογίζουμε τα επιμέρους ολοκληρώματα σε κάθε διάστημα και παίρνουμε ως αποτέλεσμα ένα εμβαδό:

$$[x_0, x_1] = [1.00, 1.06]: h \times f(x_0) = 0.06 \times \sqrt{1} = 0.06$$

$$[x_1, x_2] = [1.06, 1.12]: h \times f(x_1) = 0.06 \times \sqrt{1.06} = 0.0618$$

$$[x_2, x_3] = [1.12, 1.18]: h \times f(x_2) = 0.06 \times \sqrt{1.12} = 0.0635$$

$$[x_3, x_4] = [1.18, 1.24]: h \times f(x_3) = 0.06 \times \sqrt{1.18} = 0.0652$$

$$[x_4, x_5] = [1.24, 1.30]: h \times f(x_4) = 0.06 \times \sqrt{1.24} = 0.0668$$

Άρα για να υπολογίζουμε **αριθμητικά** το ολοκλήρωμα με τον **κανόνα παραλληλογράμμου**, αρκεί να προσθέσουμε τα επιμέρους εμβαδά που υπολογίσαμε για κάθε διάστημα. Συνεπώς: $\int_{1.00}^{1.30} \sqrt{x} d(x) \approx 0.06 + 0.0618 + 0.0635 + 0.0652 + 0.0668 = 0.3173$.

B. Κανόνας Τραπεζίου: Σε κάθε χωρίο εφαρμόζουμε τον απλό **κανόνα τραπεζίου**:

$$\int_{x_i}^{x_{i+1}} f(x) d(x) \approx \frac{(x_{i+1} - x_i)}{2} \times (f(x_i) + f(x_{i+1})) = \frac{h}{2} \times (f(x_i) + f(x_{i+1}))$$

Οπότε υπολογίζουμε τα επιμέρους ολοκληρώματα σε κάθε διάστημα και παίρνουμε ως αποτέλεσμα ένα εμβαδό:

$$[x_0, x_1] = [1.00, 1.06]: \frac{h}{2} \times (f(x_0) + f(x_1)) = 0.03 \times (\sqrt{1} + \sqrt{1.06}) = 0.0609$$

$$[x_1, x_2] = [1.06, 1.12]: \frac{h}{2} \times (f(x_1) + f(x_2)) = 0.03 \times (\sqrt{1.06} + \sqrt{1.12}) = 0.0626$$

$$\begin{aligned}[x_2, x_3] &= [1.12, 1.18]: \frac{h}{2} \times (f(x_2) + f(x_3)) = 0.03 \times (\sqrt{1.12} + \sqrt{1.18}) = 0.0643 \\[x_3, x_4] &= [1.18, 1.24]: \frac{h}{2} \times (f(x_3) + f(x_4)) = 0.03 \times (\sqrt{1.18} + \sqrt{1.24}) = 0.0660 \\[x_4, x_5] &= [1.24, 1.30]: \frac{h}{2} \times (f(x_4) + f(x_5)) = 0.03 \times (\sqrt{1.24} + \sqrt{1.30}) = 0.0676\end{aligned}$$

Άρα για να υπολογίσουμε αριθμητικά το **ολοκλήρωμα με τον κανόνα τραπεζίου**, αρκεί να προσθέσουμε τα επιμέρους εμβαδά που υπολογίσαμε για κάθε διάστημα. Συνεπώς: $\int_{1.00}^{1.30} \sqrt{x} d(x) \approx 0.0609 + 0.0626 + 0.0643 + 0.0660 + 0.0676 = 0.3214$. Παρατηρούμε ότι υπάρχει απόκλιση μεταξύ του αποτελέσματος που επιστρέφει ο κανόνας τραπεζίου και ο κανόνας του παραλληλογράμμου.

Γ. Κανόνας Simpson: Σε κάθε χωρίο εφαρμόζουμε τον κανόνα **Simpson**:

$$\int_{x_i}^{x_{i+1}} f(x) d(x) \approx \frac{(x_{i+1} - x_i)}{6} \times \left(f(x_i) + 4f\left(\frac{x_i + x_{i+1}}{2}\right) + f(x_{i+1}) \right) = \frac{h}{6} \times \left(f(x_i) + 4f\left(\frac{x_i + x_{i+1}}{2}\right) + f(x_{i+1}) \right).$$

Οπότε υπολογίζουμε τα επιμέρους ολοκληρώματα σε κάθε διάστημα και παίρνουμε ως αποτέλεσμα ένα εμβαδό:

$$\begin{aligned}[x_0, x_1] &= [1.00, 1.06]: \frac{h}{6} \times (f(x_0) + 4f((x_0 + x_1)/2) + f(x_1)) = 0.01 \times (\sqrt{1} + 4 * \sqrt{(1 + 1.06)/2} + \sqrt{1.06}) = 0.0609. \\[x_1, x_2] &= [1.06, 1.12]: \frac{h}{6} \times (f(x_1) + 4f((x_1 + x_2)/2) + f(x_2)) = 0.01 \times (\sqrt{1.06} + 4 * \sqrt{(1.06 + 1.12)/2} + \sqrt{1.12}) = 0.0626. \\[x_2, x_3] &= [1.12, 1.18]: \frac{h}{6} \times (f(x_2) + 4f((x_2 + x_3)/2) + f(x_3)) = 0.01 \times (\sqrt{1.12} + 4 * \sqrt{(1.12 + 1.18)/2} + \sqrt{1.18}) = 0.0646. \\[x_3, x_4] &= [1.18, 1.24]: \frac{h}{6} \times (f(x_3) + 4f((x_3 + x_4)/2) + f(x_4)) = 0.01 \times (\sqrt{1.18} + 4 * \sqrt{(1.18 + 1.24)/2} + \sqrt{1.24}) = 0.0660. \\[x_4, x_5] &= [1.24, 1.30]: \frac{h}{6} \times (f(x_4) + 4f((x_4 + x_5)/2) + f(x_5)) = 0.01 \times (\sqrt{1.24} + 4 * \sqrt{(1.24 + 1.30)/2} + \sqrt{1.30}) = 0.0673.\end{aligned}$$

Άρα για να υπολογίσουμε αριθμητικά το **ολοκλήρωμα με τον κανόνα Simpson**, αρκεί να προσθέσουμε τα επιμέρους εμβαδά που υπολογίσαμε για κάθε διάστημα. Συνεπώς $\int_{1.00}^{1.30} \sqrt{x} d(x) \approx 0.0609 + 0.0626 + 0.0646 + 0.0660 + 0.0673 = 0.3214$. Παρατηρούμε ότι βρέθηκε ίδιο αποτέλεσμα με τον κανόνα τραπεζίου.

Σημείωση: Αν υπολογίζαμε το εμβαδόν στο χαρτί με **πλήρη ακρίβεια** (ακριβή τύπο) τότε αυτό θα ήταν :

$$\int_{1.00}^{1.30} \sqrt{x} d(x) = \int_{1.00}^{1.30} \left(\frac{2}{3} x^{\frac{3}{2}} \right)' d(x) = \left[\frac{2}{3} x^{\frac{3}{2}} \right]_{1.00}^{1.30} = \frac{2}{3} (1.30)^{\frac{3}{2}} - \frac{2}{3} (1.00)^{\frac{3}{2}} = 0.32148$$

Παρατηρούμε, ότι με τις μεθόδους **Simpson** και **τραπεζίου**, έχουμε αρκετά καλή ακρίβεια του λάχιστον **4 δεκαδικών ψηφίων**, ενώ με την μέθοδο του παραλληλογράμμου, όπως ήταν αναμενόμενο, η ακρίβεια υπολογισμού του ολοκληρώματος είναι μικρότερη.

6.5 Παράδειγμα ολοκλήρωσης με χρήση κανόνων παραλληλογράμμου, τραπεζίου, Simpson

Θέλουμε να προσεγγίσουμε το ολοκλήρωμα $I(f) = \int_0^1 \cos(2\pi x^2) dx$ αλλά έχετε ξεχάσει εντελώς τους τύπους ολοκλήρωσης. Γνωρίζετε όμως τους τύπους Simpson, παραλληλογράμμου και τραπεζίου. Με βάση αυτούς, ποια από τις παρακάτω τιμές είναι η καλύτερη προσέγγιση του ολοκληρώματος $I(f)$? Δίνεται ότι $\cos(0) = 1$, $\cos(\pi/2) = 0$, $\cos(\pi/4) = \frac{\sqrt{2}}{2}$

a) $1/3$,	b) 1 ,	c) $\frac{\sqrt{2}}{3}$,	d) 0
------------	----------	---------------------------	--------

Λύση

Ο ακριβής υπολογισμός του ολοκληρώματος δίνει αποτέλεσμα:

$$\int_0^1 \cos(2\pi x^2) dx = \frac{C(2)}{2} \approx 0.244127$$

Αν θεωρήσουμε **δύο υποδιαστήματα**, τότε το $h = \frac{b-a}{n} = \frac{1.00-0.00}{2} = 0.5$. Στη συνέχεια χρησιμοποιώντας **τον κανόνα του παραλληλογράμμου** έχουμε ότι:

$$[x_0, x_1] = [0, 0.5] : h * f(x_0) = 0.5 * \cos(2 * \pi * 0^2) = 0.5$$

$$[x_1, x_2] = [0.5, 1] : h * f(x_1) = 0.5 * \cos\left(2 * \pi * \left(\frac{1}{2}\right)^2\right) = 0.5 * \cos\left(2 * \pi * \frac{1}{4}\right) = 0$$

Άρα το συνολικό εμβαδό είναι το άθροισμα των δύο περιοχών, δηλαδή $0.5 + 0 = 0.5$

Στη συνέχεια χρησιμοποιούμε για τα ίδια υποδιαστήματα τον **κανόνα του τραπεζίου** και έχουμε:

$$[x_0, x_1] = [0, 0.5] : \frac{h}{2} * (f(x_0) + f(x_1)) = 0.25 * (\cos(2 * \pi * 0^2) + \cos(2 * \pi * 0.5^2)) = 0.25 * (1 + 0) = 0.25$$

$$[x_1, x_2] = [0.5, 1] : \frac{h}{2} * (f(x_1) + f(x_2)) = 0.25 * (\cos(2 * \pi * 0.5^2) + \cos(2 * \pi * 1^2)) = 0.25 * (0 + 1) = 0.25$$

Άρα το συνολικό εμβαδό είναι το άθροισμα των δύο περιοχών, δηλαδή $0.25 + 0.25 = 0.5$.

Στη συνέχεια χρησιμοποιούμε για τα ίδια υποδιαστήματα τον **κανόνα του Simpson** και έχουμε:

$$[x_0, x_1] = [0, 0.5] : \frac{h}{6} \left(f(x_0) + 4 * f\left(\frac{x_0 + x_1}{2}\right) + f(x_1) \right) = 1/12 * (1 + 4 * 0.9238 + 0) = 0.3912$$

$$[x_1, x_2] = [0.5, 1] : \frac{h}{6} \left(f(x_1) + 4 * f\left(\frac{x_1 + x_2}{2}\right) + f(x_2) \right) = 1/12 * \left(0 + 4 * \cos\left(9 * \frac{\pi}{8}\right) + 1 \right) = -0.2246$$

Άρα το συνολικό εμβαδό είναι και πάλι το άθροισμα των εμβαδών των δύο περιοχών, δηλαδή $0.39125 - 0.2246$

= 0.1665. Παρατηρούμε ότι καμία από τις τρεις τιμές που υπολογίσαμε δεν προσεγγίζει κάποια από τα διοθέντα αποτελέσματα.

Στη συνέχεια θα πρέπει να αυξήσουμε τον αριθμό των υποδιαστημάτων για να πετύχουμε καλύτερη ακρίβεια στον υπολογισμό των ολοκληρωμάτων. Αν θεωρήσουμε **τρία υποδιαστήματα**, τότε το $h = \frac{1.00-0.00}{3} = 0.33$.

Στη συνέχεια χρησιμοποιώντας **τον κανόνα του παραλληλογράμμου** έχουμε ότι:

$$[x_0, x_1] = [0, 0.33] : h * f(x_0) = 0.33 * \cos(2 * \pi * 0^2) = 0.5$$

$$[x_1, x_2] = [0.33, 0.66] : h * f(x_1) = 0.33 * \cos(2 * \pi * 0.33^2) = 0.3299$$

$$[x_2, x_3] = [0.66, 0.99] : h * f(x_2) = 0.33 * \cos(2 * \pi * 0.66^2) = 0.287$$

Άρα το συνολικό εμβαδό είναι το άθροισμα των τριών περιοχών είναι 1.1169

Στη συνέχεια χρησιμοποιούμε για τα ίδια υποδιαστήματα τον **κανόνα του τραπεζίου** και έχουμε:

$$[x_0, x_1] = [0, 0.33] : \frac{h}{2} * (f(x_0) + f(x_1)) = 0.1666 * (1 + 0.99) = 0.333$$

$$[x_1, x_2] = [0.33, 0.66] : \frac{h}{2} * (f(x_1) + f(x_2)) = 0.166 * (0.99 + 0.997) = 0.3299$$

$$[x_2, x_3] = [0.66, 0.99] : \frac{h}{2} * (f(x_2) + f(x_3)) = 0.166 * (0.997 + 0.997) = 0.33$$

Άρα το συνολικό εμβαδό είναι το άθροισμα των τριών περιοχών είναι 0.9934

Άρα επιλέγουμε τη δεύτερη τιμή, δηλ. το 1.

6.6 Παράδειγμα ολοκλήρωσης με χρήση κανόνων παραλληλογράμμου, τραπεζίου, Simpson

Θέλουμε να προσεγγίσουμε το $I(f) = \int_{-1/2}^{1/2} f(x) dx$, όπου $f(x) = 4 * x^3 - 2 * x^2 - x - 4$ με τον απλό τύπο

Simpson. Επίσης να υπολογιστεί και το σφάλμα της μεθόδου.

Λύση

Θεωρούμε -αυθαίρετα- τέσσερα υποδιαστήματα $(-0.5, -0.25)$, $(-0.25, 0)$, $(0, 0.25)$ και $(0.25, 0.5)$, όπου η λεπτότητα $h = \frac{\frac{1}{2} + \frac{1}{2}}{4} = \frac{1}{4}$ και το $\frac{h}{6} = \frac{1}{24} = 0.04166$.

$$[x_0, x_1]: \frac{h}{6} \left(f(x_0) + 4f\left(\frac{x_0 + x_1}{2}\right) + f(x_1) \right) = 0.04166(-4.5 - 4.11719 - 3.9375) = -1.03776$$

$$[x_1, x_2]: \frac{h}{6} \left(f(x_1) + 4f\left(\frac{x_1 + x_2}{2}\right) + f(x_2) \right) = 0.04166(-3.9375 - 3.91406 - 4) = -1.00391$$

$$[x_2, x_3]: \frac{h}{6} \left(f(x_2) + 4f\left(\frac{x_2 + x_3}{2}\right) + f(x_3) \right) = 0.04166(-4 - 4.14844 - 4.3125) = -1.04297$$

$$[x_3, x_4]: \frac{h}{6} \left(f(x_3) + 4f\left(\frac{x_3 + x_4}{2}\right) + f(x_4) \right) = 0.04166(-4.3125 - 4.44531 - 4.5) = -1.09245$$

Για να προσεγγίσουμε αριθμητικά το ολοκλήρωμα, αρκεί να αθροίσουμε τα επιμέρους εμβαδά που πρέκυψαν από κάθε υποδιάστημα. Έτσι προκύπτει ότι $I = -4.17708$.

Ο ακριβής υπολογισμός του ολοκληρώματος θα ήταν:

$$\begin{aligned} \int_{-1/2}^{1/2} f(x) dx &= \int_{-1/2}^{1/2} (4x^3 - 2x^2 - x - 4) dx = \int_{-1/2}^{1/2} 4x^3 dx - \int_{-1/2}^{1/2} 2x^2 dx - \int_{-1/2}^{1/2} x dx - \int_{-1/2}^{1/2} 4 dx = \\ &= 4 \frac{x^4}{4} \Big|_{-1/2}^{1/2} - 2 \frac{x^3}{3} \Big|_{-1/2}^{1/2} - \frac{x^2}{2} \Big|_{-1/2}^{1/2} - 4x \Big|_{-1/2}^{1/2} = \\ &= \left(\frac{1}{2}\right)^4 - \left(-\frac{1}{2}\right)^4 - \frac{2}{3} \left[\left(\frac{1}{2}\right)^3 - \left(-\frac{1}{2}\right)^3\right] - \frac{1}{2} \left[\left(\frac{1}{2}\right)^2 - \left(-\frac{1}{2}\right)^2\right] - 4 \left(\frac{1}{2} - \left(-\frac{1}{2}\right)\right) = \frac{1}{16} - \frac{1}{16} - \frac{2}{3} \left(\frac{1}{8} + \frac{1}{8}\right) - \frac{1}{2} * 0 - 4 = -\frac{2}{3} * \frac{2}{8} - 4 = -\frac{4}{24} - 4 = -\frac{1}{6} - 4 = -\frac{25}{6} = -4.16666. \end{aligned}$$

Το σφάλμα του απλού κανόνα του Simpson δίνεται από τον τύπο: $\int_a^b f(x) - S(f) = -\frac{f^{(4)}(\eta)}{2880} (x_{i+1} - x_i)^5$ και είναι μηδέν, διότι δεν υπάρχει η τέταρτη παράγωγος.

Παρατήρηση: Συνίσταται να ξεκινήσουμε με δύο υποδιαστήματα και προοδευτικά να τα αυξήσουμε, αν χρειάζεται. Ενδεχομένως η λύση (ακριβής υπολογισμός του $I(f)$) να βρεθεί και με λιγότερα από τέσσερα σημεία.

6.7 Παράδειγμα ολοκλήρωσης με χρήση σύνθετου κανόνα τραπεζίου

- Μας ενδιαφέρει η προσέγγιση της τιμής του ολοκληρώματος $\int_0^1 (x^3 + 10x) dx$ με τη σύνθετη μέθοδο τραπεζίου προκειμένου να επιτευχθεί ακρίβεια 1 δεκαδικού ψηφίου. Δίνεται ότι η ακριβής τιμή είναι 5.25.
- Αν εφαρμοστεί εναλλακτικά ο κανόνας Simpson, ποια είναι η σωστή απάντηση για το σφάλμα της προσέγγισης: α) 0.005 β) περίπου 0.5 γ) 0.

Λύση

Ζητούμενο είναι να χρησιμοποιήσουμε τον **ελάχιστο αριθμό υποδιαστημάτων** που ικανοποιεί το κριτήριο. Επομένως, **δεν** ενδείκνυται να επιλέξουμε από την αρχή αυθαίρετα μεγάλο αριθμό υποδιαστημάτων, γιατί ακόμα και αν το αποτέλεσμα ικανοποιεί τη συνθήκη, δεν έχουμε κάποιο τρόπο να γνωρίζουμε αν αυτό το πλήθος υποδιαστημάτων ήταν το ελάχιστο δυνατό.

Εφαρμόζουμε τον **κανόνα τραπεζίου** αρχίζοντας από την περίπτωση $n = k = 2$ υποδιαστημάτων (αφού η εκφώνηση αναφέρεται στη σύνθετη μέθοδο) οπότε $h = \frac{b-a}{2} = 0.5$, όπου $a = 0$, $b = 1$. Εφαρμόζουμε στη συνέχεια τη **σύνθετη μέθοδο τραπεζίου** θα έχουμε: $I(f) = \frac{h}{2} (f(x_0) + f(x_n) + 2 * \sum_{i=1}^{n-1} f(x_i))$, άρα έπειτα από υπολογισμούς θα έχουμε ότι: $\frac{h}{2} (f(a) + f(b) + 2 * \sum_{i=1}^{n-1} f(x_i)) = \frac{1}{4} * (f(0) + f(1)) + 2 * f(1/2) = \frac{1}{4} * (0 + 11 + 2 * 5.125) = 5.3125$. Ο όρος $2 \sum_{i=1}^{n-1} f_i$ υπολογίζει πάντα τα ενδιάμεσα σημεία του διαστήματος ολοκλήρωσης και στην προκειμένη περίπτωση είναι μόνο ένα, αφού στο άθροισμα που εμφανίζεται μέσα στο προσεγγιστικό τύπο το $i = 1$ έως $n - 1 = 2 - 1 = 1$. Στον όρο αυτό λαμβάνουμε υπόψη το h . Δηλαδή, εδώ έχουμε $h = \frac{1}{2}$.

Στη συνέχεια εφαρμόζουμε **τον κανόνα τραπεζίου για $n = k = 3$ υποδιαστήματα** (επειδή από την εκφώνηση θα έχουμε ακρίβεια ενός δεκαδικού ψηφίου, θα πρέπει να αυξήσουμε κατά ένα τον προηγούμενο αριθμό διαστημάτων) οπότε $h = \frac{b-a}{3} = \frac{1}{3} = 0.33$, όπου $a = 0$, $b = 1$. Εφαρμόζουμε και πάλι στη συνέχεια τη **σύνθετη μέθοδο τραπεζίου** θα έχουμε: $I(f) = \frac{h}{2} (f(x_0) + f(x_n) + 2 * \sum_{i=1}^{n-1} f(x_i)) = \frac{h}{2} (f(a) + f(b) + 2 * \sum_{i=1}^{n-1} f_i) = \frac{1}{6} * (f(0) + f(1) + 2 * (f(1/3) + f(2/3))) = \frac{1}{6} * (0 + 1 + 2 * (3.3359 + 6.8874)) = 5.24114$. Ο όρος $2 \sum_{i=1}^{n-1} f_i$ υπολογίζει πάντα τα ενδιάμεσα σημεία του διαστήματος ολοκλήρωσης και στην προκειμένη περίπτωση είναι δύο, αφού στο άθροισμα που εμφανίζεται μέσα στο προσεγγιστικό τύπο το $i = 1$ έως $n - 1$ δηλαδή το $i = 1$ έως $3 - 1 = 2$ και τα σημεία αυτά είναι το $1/3$ και το $2/3$, αφού το διάστημα ολοκλήρωσης είναι το $(0, 1)$. Άρα σταματάμε, διότι με βάση τη δοθείσα ακρίβεια του ενός δεκαδικού ψηφίου έχουμε βρει τη σωστή απάντηση. Δηλαδή σταματάμε όταν βρούμε ως πρώτο δεκαδικό ψηφίο το «2», διότι αυτή η τιμή υπάρχει στην ακριβή απάντηση του ολοκληρώματος που δίνεται από την εκφώνηση. Πριν, με τα δύο διαστήματα, βρήκαμε την τιμή 5.3125 που είχε ως πρώτο δεκαδικό ψηφίο το «3» και για το λόγο αυτό δεν σταματήσαμε εκεί, αλλά διαιρέσαμε περαιτέρω το διάστημα ολοκλήρωσης $[a, b]$ σε τρία υποδιαστήματα.

b) Το σφάλμα της προσέγγισης με τη μέθοδο Simpson είναι $\int_a^b f(x) - S(f) = -\frac{f^{(4)}(\eta)}{2880} (x - x_0)^5$. Επειδή η συνάρτηση $f(x)$ είναι 3^{ου} βαθμού, η παράγωγος $f^{(4)}(\eta) = 0$. Άρα το σφάλμα είναι 0 και η σωστή απάντηση είναι η (γ).

6.8 Θέμα με αριθμητική ολοκλήρωση (παλιό θέμα)

α) Μας ενδιαφέρει η προσέγγιση της τιμής του ολοκληρώματος $I = \int_0^1 (x^3 + 5x + 6) dx$. Εφαρμόστε τη **σύνθετη μέθοδο τραπεζίου**, χρησιμοποιώντας όσο γίνεται λιγότερα υποδιαστήματα, ώστε να πετύχετε ακρίβεια 1 δεκαδικού ψηφίου. Δίνεται ότι η ακριβής τιμή είναι **8.75**

β) Επίσης να απαντήσετε στο εξής: Αν εφαρμοστεί εναλλακτικά ο κανόνας **Simpson**, να κυκλώσετε την πιο κατάλληλη απάντηση για το **σφάλμα** της προσέγγισης.

- a): περίπου 0.005. b): 0 c): περίπου 0.5.

Λύση

Δεδομένου ότι η εκφώνηση ζητούσε την εφαρμογή της σύνθετης μεθόδου, μπορούμε να ξεκινήσουμε με δύο μόνο υποδιαστήματα, προσεγγίζοντας και ελέγχοντας το αποτέλεσμα, **και σε περίπτωση που δεν έχει επιτευχθεί η ζητούμενη ακρίβεια, διαμερίζουμε με ένα επιπλέον υποδιάστημα.** Επομένως και πάλι το ζητούμενο είναι να χρησιμοποιήσουμε τον **ελάχιστο αριθμό υποδιαστημάτων** που ικανοποιεί το κριτήριο και δεν ενδείκνυται να επιλέξουμε από την αρχή αυθαίρετα μεγάλο αριθμό υποδιαστημάτων γιατί ακόμα και αν το αποτέλεσμα ικανοποιεί τη συνθήκη, δεν υπάρχει τρόπος να γνωρίζουμε αν αυτό το πλήθος υποδιαστημάτων ήταν το ελάχιστο δυνατό.

Εφαρμόζουμε **τον κανόνα τραπεζίου** αρχίζοντας από την περίπτωση $n = k = 2$ υποδιαστημάτων (αφού η εκφώνηση αναφέρεται στη σύνθετη μέθοδο) οπότε λεπτότητα $h = \frac{b-a}{2} = \frac{1}{2} = 0.5$, όπου $a = 0$, $b = 1$. Εφαρμόζουμε στη συνέχεια τη **σύνθετη μέθοδο τραπεζίου** θα έχουμε $I(f) = \frac{h}{2} (f_0 + f_k + 2 \sum_{j=1}^{k-1} f_j) = \frac{\frac{1}{2}}{2} (f(a) + f(b) + 2 \sum_{j=1}^{k-1} f_j) = \frac{1}{4} * (f(0) + f(1) + 2 * f(1/2)) = \frac{1}{4} * (6 + 12 + 17.25) = \frac{1}{4} * 35.25 = 8.8125$. Παρατηρούμε ότι στο αποτέλεσμα που υπολογίσαμε, **δεν** μας ικανοποιεί η ακρίβεια του ενός δεκαδικού ψηφίου, διότι βρήκαμε αποτέλεσμα το 8.8125, ενώ έπρεπε να βρούμε 8.75. Για το λόγο αυτό, εφαρμόζουμε στη συνέχεια **τον κανόνα τραπεζίου για $n = k = 3$ υποδιαστήματα** (επειδή από την εκφώνηση θέλουμε να έχουμε ακρίβεια ενός δεκαδικού ψηφίου, θα πρέπει αυξήσουμε κατά «1» τον προηγούμενο αριθμό διαστημάτων), οπότε η λεπτότητα $h = \frac{b-a}{3} = \frac{1}{3} = 0.33$, όπου $a = 0$, $b = 1$. Εφαρμόζουμε και πάλι τη **σύνθετη μέθοδο τραπεζίου**: $I(f) = \frac{h}{2} (f_0 + f_k + 2 \sum_{j=1}^{n-1} f_j) = \frac{\frac{1}{3}}{2} (f(a) + f(b) + 2 \sum_{j=1}^{3-1} f_j) = \frac{1}{6} * (f(0) + f(1) + 2 * (f(1/3) + f(2/3))) = \frac{1}{6} * (6 + 12 + 2 * (\frac{1}{27} + \frac{5}{3})) = 8.777$. Άρα σταματάμε, διότι με βάση τη δοθείσα ακρίβεια του ενός δεκαδικού ψηφίου έχουμε βρει τη σωστή απάντηση. Αν χρειαζόταν να χρησιμοποιήσουμε και τέταρτο υποδιάστημα ($n = 4$), τότε και πάλι θα υπολογίζαμε αρχικά το $h = \frac{b-a}{4} = \frac{1}{4}$ και στο άθροισμα $2 \sum_{j=1}^{n-1} f_j = 2 * (f(1/4) + f(2/4) + f(3/4))$ κ.ο.κ.

β) **Το σφάλμα της (απλής) μεθόδου Simpson είναι πολλαπλάσιο της 4^{ης} παραγώγου της συνάρτησης σε κάποιο σημείο στο διάστημα ολοκλήρωσης και αυτό διότι ο ακριβής τύπος για το σφάλμα σε ένα διάστημα (a, b) είναι $\int_a^b f(x) - S(f) = -\frac{f^{(4)}(\eta)}{2880} (x - x_0)^5$.** Επειδή η συνάρτηση $f(x)$ είναι 3^{ου} βαθμού, η παράγωγος $f^{(4)}(\eta) = 0$. Επομένως, στην περίπτωση που η προς ολοκλήρωση συνάρτηση είναι πολυώνυμο ως και 3^{ου} βαθμού (όπως στο συγκεκριμένο ερώτημα), το σφάλμα θα είναι 0. Επομένως, δεν χρειάζονται πράξεις υπολογισμού του σφάλματος και η σωστή απάντηση είναι η (b).

6.9 Θέμα με αριθμητική ολοκλήρωση (παλιό θέμα)

Προσεγγίστε την τιμή του ολοκληρώματος

$$I = \int_{5}^{8} (8x^3 + 2x) dx$$

με τους παρακάτω δύο τρόπους. Εφαρμόστε τη **σύνθετη μέθοδο τραπεζίου**, ξεκινώντας με δύο υποδιαστήματα, και **συνεχίστε μέχρι το σχετικό σφάλμα να είναι μικρότερο του 0.02**. Έπειτα εφαρμόστε τη μέθοδο **Simpson**. Δίνεται ότι η ακριβής τιμή του ολοκληρώματος είναι **6981**. Κυκλώστε μια από τις παρακάτω απαντήσεις και εξηγήστε αναλυτικά.

α) **Η προσέγγιση με τη μέθοδο Simpson ορθά βρίσκει την πραγματική τιμή** και β) **Η προσέγγιση με τη σύνθετη μέθοδο τραπεζίου** είναι πιο κοντά στην πραγματική τιμή.

Λύση

α) Σύνθετη μέθοδος τραπεζίου:

Με 2 υποδιαστήματα: [5, 6.5] και [6.5, 8], έχουμε με λεπτότητα $h = \frac{8-5}{2} = \frac{3}{2} = 1.5$. Εφαρμόζουμε αρχικά **σύνθετη μέθοδο τραπεζίου** και έχουμε ότι το $CT(f) = \frac{h}{2}(f_0 + f_k + 2 \sum_{j=1}^{k-1} f_j) = \frac{h}{2}(f(a) + f(b) + 2 * \sum_{j=1}^{2-1} f_j) = \frac{1.5}{2}(f(5) + f(8) + 2 * f(6.5)) = 0.75 * (1010 + 4112 + 2 * 2210) = 0.75 * (9542) = 7156.5$. Στη συνέχεια ελέγχουμε το σχετικό σφάλμα του υπολογισθέντος ολοκληρώματος. Το σχετικό σφάλμα είναι $\left| \frac{7156.5 - 6981}{7156.5} \right| = 0.02452 > 0.02$. Το σχετικό σφάλμα είναι μεγαλύτερο της τιμής 0.02, οπότε θα πρέπει να χωρίσουμε το αρχικό διάστημα σε περισσότερα υποδιαστήματα. Με $n = 3$ υποδιαστήματα, το $h = \frac{8-5}{3} = \frac{3}{3} = 1$ και τα υποδιαστήματα αυτά είναι [5, 6], [6, 7], [7, 8], και έχουμε ότι ο σύνθετος κανόνας τραπεζίου δίνει το εξής αποτέλεσμα:

$$CT(f) = \frac{h}{2} \left(f(5) + f(8) + 2 * \sum_{j=1}^2 f(j) \right) = 0.5 * [1010 + 4112 + 2 * (f(6) + f(7))] = 0.5 * [5122 + 2(1740 + 2758)] = 7059.$$

Το νέο σχετικό σφάλμα είναι: $\left| \frac{7059 - 6981}{7059} \right| = 0.011 < 0.02$. Το σχετικό σφάλμα είναι στο ζητούμενο εύρος, οπότε έχουμε ικανοποιητική προσέγγιση και σταματάμε εκεί.

β) Μέθοδος Simpson

Χρησιμοποιούμε την **απλή μέθοδο Simpson** $S(f) = \frac{h}{6} * \left[f(a) + 4 * f\left(\frac{a+b}{2}\right) + f(b) \right]$ για καθένα από δύο υποδιαστήματα (όπου $h = \frac{b-a}{n} = \frac{8-5}{2} = \frac{3}{2}$), και έχουμε το εξής:

$$[5, 6.5]: S(f) = \frac{3/2}{6} * \left[f(5) + 4 * f\left(\frac{5+6.5}{2}\right) + f(6.5) \right] = \frac{1}{4} * [1010 + 6129.5 + 2210] = 2337.375$$

$$[6.5, 8]: S(f) = \frac{3/2}{6} * \left[f(6.5) + 4 * f\left(\frac{6.5+8}{2}\right) + f(8) \right] = \frac{1}{4} * [2210 + 12252.5 + 4112] = 4643.625$$

Άρα **συνολικά** το προσεγγιστικό ολοκλήρωμα είναι το άθροισμα των δύο υπολογισθέντων εμβαδών και είναι $2337.375 + 4643.625 = 6981$. Άρα βρήκαμε **την ακριβή τιμή**. Άρα η σωστή απάντηση είναι η **(α)**.

6.10 Θέμα με αριθμητική ολοκλήρωση

Έστω η συνάρτηση $\varphi(x) = \frac{\cos(x\pi)}{1+x}$

- (α) Να χρησιμοποιήσετε **τον απλό (δηλ. μη σύνθετο) κανόνα τραπεζίου** για να προσεγγίσετε το $I = \int_1^2 \varphi(x)dx$.
- (β) Αν το βασικό κόστος στην αριθμητική ολοκλήρωση είναι ο υπολογισμός των τιμών της συνάρτησης σε διαφορετικά σημεία (π.χ. ο απλός κανόνας τραπεζίου χρειάζεται 2 τιμές), να υπολογίσετε καλύτερη προσέγγιση του ως άνω ολοκληρώματος με **σύνθετο κανόνα τραπεζίου και την ελάχιστη δυνατή επιβάρυνση κόστους**.
- (γ) Να αποδείξετε ότι αν για μια συνάρτηση f , οι παράγωγοι f' , f'' είναι συνεχείς στο $[a, b]$ και η σταθερά M ικανοποιεί τη σχέση $\max_{x \in [a,b]} |f''(x)| \leq M$, τότε το $M \frac{(b-a)^3}{12}$ είναι ένα άνω φράγμα για το απόλυτο σφάλμα της προσέγγισης του $\int_a^b f(x)dx$.
- (δ) Να αιτιολογήσετε γιατί ο **απλός κανόνας Simpson** γενικά αναμένεται να είναι πιο **ακριβής** από τον απλό κανόνα του **τραπεζίου**.

Λύση

$$(α) T(f) = \frac{h}{2} [f(a) + f(b)] = \frac{(2-1)/2}{2} [f(1) + f(2)] = \frac{1}{4} (\cos \frac{\pi}{2} + \cos \frac{2\pi}{3}) = \frac{1}{4} (-\frac{1}{2} + \frac{1}{3}) = \frac{1}{4} (-\frac{1}{6}) = -\frac{1}{24}.$$

(β) Αφού απαιτείται η **χαμηλότερη επιβάρυνση κόστους**, θα χρησιμοποιήσουμε το **μικρότερο πλήθος υποδιαστημάτων**, δηλαδή $n = 2$, το $[1, 1.5]$ και $[1.5, 2]$ και η σύνθετη μέθοδος τραπεζίου δίνει: $CT(f) = \frac{2^{-1}}{2} (f(1) + f(2) + 2 * f(1.5)) = \frac{1}{4} * (-\frac{1}{2} + \frac{1}{3} + 0) = \frac{1}{4} (-\frac{1}{6}) = -\frac{1}{24}$, διότι $f(3/2) = \frac{\cos(\frac{3\pi}{2})}{1+3} = 0$.

(γ) Από τη **τμηματική γραμμική παρεμβολή** που αναπτύξαμε σε προηγούμενο κεφάλαιο, γνωρίζουμε ότι το

$$\Pi_1(x) = f(x_i) + \frac{f(x_{i+1}) - f(x_i)}{x_{i+1} - x_i} * (x - x_i), \forall x \in [x_i, x_{i+1}]$$

και στη συνέχεια ολοκληρώνουμε την $f(x)$, δηλ. υπολογίζουμε

το $\int_{x_i}^{x_{i+1}} f(x)dx$ και στη συνέχεια χρησιμοποιούμε αντί για την $f(x)$, το τμηματικό γραμμικό πολυώνυμο που την παρεμβάλει: $\int_{x_i}^{x_{i+1}} f(x)dx \approx \int_{x_i}^{x_{i+1}} f(x_i) + \frac{f(x_{i+1}) - f(x_i)}{x_{i+1} - x_i} * (x - x_i) \approx f(x_i) * (x_{i+1} - x_i) + \frac{f(x_{i+1}) - f(x_i)}{x_{i+1} - x_i} * \frac{(x_{i+1} - x_i)^2}{2}$.

Ο απλός κανόνας τραπεζίου δίνει το $T(f) = \frac{(x_{i+1} - x_i)}{2} \times (f(x_i) + f(x_{i+1})) = \frac{h}{2} \times (f(x_i) + f(x_{i+1}))$. Γνωρίζουμε ότι

αν υπάρχει η δεύτερη παράγωγος της συνάρτησης f , δηλαδή $f \in C^{(2)}[x_i, x_{i+1}]$ τότε υπάρχει ένα M :

$$\max_{x \in [x_i, x_{i+1}]} |f''(x)| \leq M. \text{ Η διαφορά } f(x) - \Pi_1(x) = f(x) - \left(f(x_i) + \frac{f(x_{i+1}) - f(x_i)}{x_{i+1} - x_i} * (x - x_i) \right) = \frac{f''(\xi_x)}{2} (x - x_i)(x - x_{i+1})$$

Η διαφορά ανάμεσα στο ακριβές ολοκλήρωμα και τον προσεγγιστικό τύπο του τραπεζίου είναι:

$$\int_{x_i}^{x_{i+1}} f(x)dx - T(f) = \int_{x_i}^{x_{i+1}} (f(x) - \Pi_1(x)) dx = \frac{1}{2} \int_{x_i}^{x_{i+1}} f''(\xi_x) (x - x_i)(x - x_{i+1}) dx = \frac{f''(\eta)}{2} \int_{x_i}^{x_{i+1}} (x - x_i)(x - x_{i+1}) dx = \frac{f''(\eta)}{12} (x_{i+1} - x_i)^3$$

Άρα αποδείξαμε το ζητούμενο, ότι $|\int_{x_i}^{x_{i+1}} f(x)dx - T(f)| \leq \frac{M}{12} * (x_{i+1} - x_i)^3$.

(δ) Στον κανόνα του **τραπεζίου** υπολογίζουμε προσεγγιστικά την τιμή του $\int_{x_0}^{x_1} f(x)dx$ ολοκληρώνοντας το πολυώνυμο προσέγγισης που διέρχεται από τα σημεία $(x_0, f(x_0))$, $(x_1, f(x_1))$ που είναι πρωτοβάθμιο, δηλαδή ευθεία. Στην περίπτωση του κανόνα του **Simpson** προσεγγίζουμε το ίδιο ολοκλήρωμα **χρησιμοποιώντας πολυώνυμο δευτέρου βαθμού που παρεμβάλλεται σε τουλάχιστον τρία σημεία**. Συνεπώς, αναμένεται ο απλός κανόνας Simpson να είναι πιο **ακριβής** από αυτόν του τραπεζίου.

Κεφάλαιο 7 – MATLAB – Σφάλματα – Αριθμητική κινητής υποδιαστολής - QR - SVD

7.1 Χρήσιμες συναρτήσεις MATLAB αναφορικά με πολυώνυμα

Κάθε **πολυώνυμο βαθμού n** έχει δύο βασικά στοιχεία: **τους συντελεστές του** (που τους συμβολίζουμε ως α_i , όπου i η δύναμη της μεταβλητής που συνοδεύει το συντελεστή και μάλιστα ισχύει ότι ο αριθμός των συντελεστών = $n + 1$) **και τις ρίζες του** (που συνήθως τις συμβολίζουμε ως ρ_i , και μάλιστα ισχύει ότι ο αριθμός των ριζών = n). Σε περίπτωση που κάποιες ρίζες είναι μιγαδικές, τότε αυτές εμφανίζονται σε συζυγή ζεύγη.

Με δεδομένο ότι ένα πολυώνυμο γράφεται στη μορφή: $p(x) = \alpha_n x^n + \alpha_{n-1} x^{n-1} + \dots + \alpha_2 x^2 + \alpha_1 x + \alpha_0 = \sum_{i=0}^n \alpha_i x^i$, που ονομάζεται **μορφή αθροίσματος**, υπάρχει και η **εναλλακτική μορφή** (σε περίπτωση που είναι γνωστές οι ρίζες του πολυωνύμου): $p(x) = \alpha_n (x - p_1)(x - p_2) \dots (x - p_n)$ που ονομάζεται **μορφή γινομένου**.

Χρήσιμες συναρτήσεις MATLAB σχετικές με πολυώνυμα είναι οι εξής:

- i. $a = poly(r)$: επιστρέφει ένα διάνυσμα με τους **συντελεστές του πολυωνύμου**, όταν δίνονται οι **ρίζες του**, π.χ. αν γνωρίζουμε ότι οι ρίζες ενός πολυωνύμου είναι οι τιμές 1, 2, 3, 4, τότε έχουμε το πολυώνυμο $p(x) = \alpha_n(x - 1)(x - 2)(x - 3)(x - 4) = (x^2 - 2x - x + 2)(x^2 - 7x + 12) = (x^2 - 3x + 2)(x^2 - 7x + 12) = x^4 - 7x^3 + 12x^2 - 3x^3 + 21x^2 - 36x + 2x^2 - 14x + 24 = x^4 - 10x^3 + 45x^2 - 50x + 24$. Αν πληκτρολογήσουμε στη γραμμή εντολών της MATLAB τις ακόλουθες εντολές:

```
>> r = [1 2 3 4]
```

```
>> a = poly(r)
```

τότε θα πάρουμε τα αποτελέσματα: $a = 1 -10 45 -50 24$, δηλαδή επιστρέφονται οι **συντελεστές του πολυωνύμου** ως διάνυσμα $a = [1 -10 45 -50 24]$. Παρατηρούμε ότι οι συντελεστές είναι κατά «1» περισσότεροι από τις ρίζες του πολυωνύμου.

- ii. $r = roots(a)$: επιστρέφει τις ρίζες του πολυωνύμου, δηλ. το $r = 4.000, 3.000, 2.000, 1.000$, όταν δίνεται ως όρισμα το διάνυσμα a με τους **συντελεστές του πολυωνύμου**, που υπολογίστηκαν προηγουμένως.

- iii. $y = polyval(p, x)$ αν το x είναι σημείο, διάνυσμα ή μητρώο, τότε επιστρέφεται η τιμή του πολυωνύμου p στο y, το οποίο είναι σημείο, διάνυσμα ή μητρώο αντίστοιχα, δηλαδή: $y = p(1) * x^N + p(2) * x^{N-1} + \dots + p(N) * x + p(N+1)$, π.χ. αν το $>>p = [1 2 3 4]$, τότε η εντολή: $>> y = polyval(p, 1)$ επιστρέφει το αποτέλεσμα: $>>y = 10$, διότι $y = p(1)x^3 + p(2)x^2 + p(3)x + p(4) = 1 * 1^3 + 2 * 1^2 + 3 * 1^1 + 4 * 1 = 10$, δηλαδή επιστρέφεται η τιμή του πολυωνύμου για x = 1.

- iv. $p = polyfit(X, Y, n)$ υπολογίζει τους συντελεστές του πολυωνύμου p(x) βαθμού n, το οποίο προσαρμόζει τα δεδομένα Y στο διάνυσμα γραμμής p μήκους n + 1, που περιέχει **τους συντελεστές του πολυωνύμου με φθίνουσα σειρά**: $p(1) * x^n + p(2) * x^{n-1} + \dots + p(n) * x + p(n+1)$.

7.2 Αριθμοί κινητής υποδιαστολής

7.2.1. Εισαγωγή στους α.κ.υ. και τα χαρακτηριστικά του συνόλου F – Χαρακτηριστικά μεγέθη του F

Έστω F το σύνολο των α.κ.υ. που αναπαρίστανται στον Η/Υ. Σε αντίθεση με τους πραγματικούς, οι α.κ.υ. έχουν αρχή, μέση και τέλος. Στους α.κ.υ. η κωδικοποίησή τους γίνεται σε τρία μέρη ή πεδία (για βάση 2):

- 1 bit για το πρόσημο s (0 αν θετικός, 1 αν αρνητικός)
- Εκθέτης e που λαμβάνει τιμές από την ελάχιστη L (αρνητική) ως τη μέγιστη U (θετική).
- ουρά $(b_0 b_1 b_2 \dots b_{t-1})_2$ (συνήθως κανονικοποιημένη, δηλ. το hidden bit $b_0 = 1$). Είναι τα ψηφία του αριθμού μετά την υποδιαστολή.

Στο υλικό χρησιμοποιείται πολωμένη αναπαράσταση για να γίνεται **αριθμητική αποκλειστικά με θετικούς**.

Σημαντικό ρόλο παίζει η συνάρτηση στρογγύλευσης $f_l: R \rightarrow F$ που για κάθε $z \in R$ επιστρέφει κάποια τιμή $f_l(z) \in F$, σύμφωνα με κάποια προσυμφωνημένη μέθοδο στρογγύλευσης, για την οποία πρέπει να ισχύουν τα εξής:

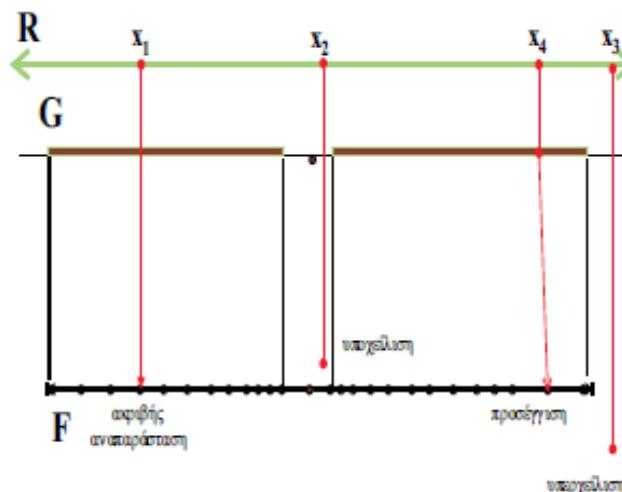
- $\text{Av } z \in F \Rightarrow f_l(z) = z$ (ακριβής αναπαράσταση).
- $\text{Av } z_1 \leq z_2 \text{ τότε } f_l(z_1) \leq f_l(z_2)$. Δηλαδή η συνάρτηση στρογγύλευσης διατηρεί την τάξη των αριθμών.

Το z στις προηγούμενες περιπτώσεις είναι κάποιο δεδομένο που θα αποτελέσει **είσοδο** στο πρόγραμμα ή είναι το **αποτέλεσμα** πράξης μεταξύ δύο α.κ.υ. **Όλοι** οι α.κ.υ. στο πρότυπο IEEE - 754 είναι **ρητοί**. Θέτουμε ως

F το σύνολο των α.κ.υ. για την κατηγορία της αναπαράστασης (μονή ή διπλή ακρίβεια). Ονομάζουμε m και M τον ελάχιστο και μέγιστο α.κ.υ. για την κατηγορία της αναπαράστασης (μονή ή διπλή ακρίβεια). Τότε $z = 0$ ή $|z| \in [m, M]$. Αν το αποτέλεσμα μίας πράξης είναι z , και το $z \in F$ τότε μπορεί να ισχύει ένα από τα εξής:

- $f_l(z) = z \rightarrow$ ακριβής αναπαράσταση.
- $z \notin F$ και $m \leq |z| \leq M$ τότε το z **στρογγυλεύεται σε $f_l(z)$** . Η απόκλιση $|f_l(z) - z|$ ονομάζεται **απόλυτο σφάλμα στρογγύλευσης**.
- $z \neq 0, z \notin (m, M) \rightarrow$ υπερχείλιση ή υποχείλιση.

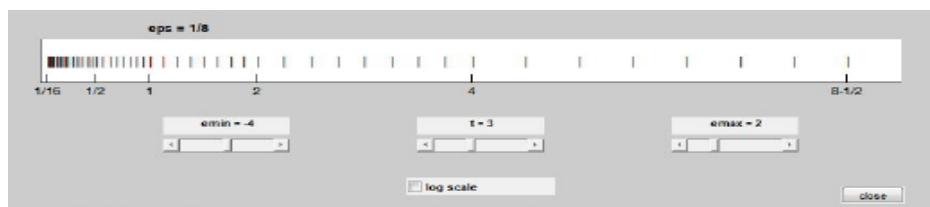
Η επόμενη εικόνα δείχνει τη συνάρτηση στρογγύλευσης $f_l(z)$ να αναπαριστά αριθμούς από το σύνολο R στο σύνολο F . Η πιο κλασσική περίπτωση είναι αυτή της προσέγγισης (στρογγύλευσης).



Εικόνα 16: Συνάρτηση στρογγυλοποίησης f_l

Σε όλες τις στρατηγικές στρογγύλευσης ενός $z \notin F$, έχουν σημασία ο **πλησιέστερος α.κ.υ.** που είναι μικρότερος του z και ο πλησιέστερος α.κ.υ. μεγαλύτερος του z . Έστω ότι τους συμβολίζουμε ως $z-$, $z+ \in F$ αντίστοιχα. Στη στρογγύλευση προς το πλησιέστερο άρτιο, θέτουμε το $\text{fl}(z)$ ίσο με τον πλησιέστερο των $z-$, $z+ \in F$ στο z . Σε περίπτωση που ισαπέχει, θέτουμε $\text{fl}(z)$ εκείνο από τα $z-$, $z+$ που έχει 0 ως τελευταίο bit στην ουρά. Για παράδειγμα, έστω στο σύστημα α.κ.υ. $F(10, 4, -9, 9)$ και οι α.κ.υ. $x_1 = 0.10005$ και $x_2 = 0.10015$. Τότε το $\text{fl}(x_1) = 0.1000$ (θεωρούμε ότι ο αριθμός αυτός ανήκει στο μέσο του διαστήματος $[0.1000 \quad 0.1001]$) και το $\text{fl}(x_2) = 0.1002$ (θεωρούμε ότι ο αριθμός αυτός ανήκει στο μέσο του διαστήματος $[0.1001 \quad 0.1002]$) και παίρνουμε το δεξί άκρο με τέσσερα δεκαδικά ψηφία ακρίβεια, επειδή τελειώνει σε ζυγό ψηφίο που είναι το «2»).

Η MATLAB χρησιμοποιεί CPU που ακολουθεί το πρότυπο IEEE-754. **Η προεπιλογή είναι α.κ.υ. IEEE διπλής ακρίβειας**, με παραμέτρους $F(2, 52 + 1, -1022, 1023)$. Με την εντολή **single** επιτρέπεται ο ορισμός και πράξεις με α.κ.υ. **μονής ακρίβειας** $F(2, 23 + 1, -126, 127)$ και με την εντολή **double** επιστρέφει η **διπλή ακρίβεια**. Οι α.κ.υ. **δεν** είναι ισοκατανεμημένοι στον άξονα των πραγματικών, αλλά είναι πιο πυκνοί προς το κέντρο (δηλαδή κοντά στο 0) και πιο αραιοί όταν απομακρύνομαστε από αυτό. Η απόσταση διαδοχικών α.κ.υ. **με ίδιο εκθέτη είναι σταθερή**, π.χ. $d(e) = (m+1) * 2^e - m * 2^e = 2^e$, ενώ η απόσταση μεταξύ διαδοχικών α.κ.υ. **διπλασιάζεται** κάθε φορά που ο εκθέτης **αυξάνεται κατά 1**: $(m+1) * 2^{e+1} - m * 2^{e+1} = d(e+1) = 2 * d(e) = 2 * 2^e$.



Το **έψιλον** της μηχανής είναι η απόσταση του 1.0 από τον αμέσως μεγαλύτερο α.κ.υ. Στην IEEE-754 στην απλή ακρίβεια (single precision) το $\epsilon_M \approx 1.1921e-007 = 1.192 * 10^{-7}$ και στην double precision το $\epsilon_M \approx 2.2204e-016 = 2.2204 * 10^{-16}$. Γενικεύοντας στη MATLAB, η εντολή **eps(x)** επιστρέφει την απόσταση του x από τον αμέσως επόμενο διαδοχικό του α.κ.υ. x^+ . Η απάντηση διαμορφώνεται ανάλογα με τον τύπο του ορίσματος (single ή double). Στη MATLAB για οποιοδήποτε α.κ.υ. x , η εντολή **eps(x)** επιστρέφει την **απόσταση του x από τον αμέσως επόμενο διαδοχικό του x^+** , δηλαδή $\text{eps}(x) = \delta(x, x^+)$.

Στη βιβλιογραφία αυτό αναφέρεται και ως **ulp(x)** (units in the last place). Η μέγιστη σχετική απόσταση είναι $\text{eps} = \epsilon_M = \frac{2^{-(t-1)}2^e}{2^e} = 2^{1-t}$. Για τις μετρήσεις σφαλμάτων βολεύει να τα εκφράζουμε ως πολλαπλάσια κάποιας μονάδας μέτρησης του σφάλματος στρογγύλευσης. Δηλαδή **δεν υπάρχει ακριβής υπολογισμός σφαλμάτων**, αλλά υπολογισμός (αποδεκτών) άνω φραγμάτων σφαλμάτων. Συνήθως ως **αποδεκτά** άνω φράγματα σφαλμάτων έχουμε το u και το $|θ_n| \leq \gamma_n = \frac{n\mu}{1-n\mu}$. **Η μονάδα στρογγύλευσης u είναι το μέγιστο σχετικό σφάλμα για το συγκεκριμένο τρόπο στρογγύλευσης.** Όταν χρησιμοποιείται στρογγύλευση προς το πλησιέστερο, τότε το $u = \max_{z \neq 0} \frac{|z - \text{fl}(z)|}{|z|} = \frac{2^{1-t}}{2} = 2^{-t} = \frac{\text{eps}}{2}$ ή $\text{eps} = 2 * u$.

7.2.2. Άσκηση με χρήση εμ και συνάρτησης roots

- i) Πόσοι αριθμοί ανήκουν στο σύνολο $F(2, 2, -2, 2)$; Ποια είναι η τιμή του ϵ_M για κάθε σύνολο.
- ii) Για οποιοδήποτε δοθέν διάνυσμα $v \in \mathbb{R}^n$, η εντολή $c = \text{poly}(v)$ κατασκευάζει τους $n+1$ συντελεστές του πολυωνύμου $p(x) = \sum_{k=1}^{n+1} c(k)x^{n+1-k}$ που είναι ίσο με το $\prod_{k=1}^n (x - v(k))$. Με ακριβή αριθμητική (δηλαδή με άπειρη ακρίβεια), πρέπει να βρεθεί ότι $v = \text{roots}(\text{poly}(c))$. Στην πραγματικότητα όμως αυτό **δεν** ισχύει λόγω σφαλμάτων στρογγύλευσης. Πως επηρεάζεται η **ακρίβεια των υπολογισμένων ριζών από το βαθμό του πολυωνύμου**;

Λύση

- i) Η γενική μορφή του συνόλου των α.κ.υ είναι το $F(\beta, t, e_{\min}, e_{\max})$, όπου β = βάση του αριθμητικού συστήματος, t = ακρίβεια δεκαδικών ψηφίων, δηλαδή το πλήθος των ψηφίων της ουράς, e_{\min} = είναι η ελάχιστη τιμή του εκθέτη και e_{\max} = είναι η μέγιστη τιμή του εκθέτη. Αυτές οι τιμές αποτελούν τα **όρια** του εκθέτη. Ένας H/Y αποθηκεύει έναν πραγματικό αριθμό με τον ακόλουθο τρόπο: $x = (-1)^s * (0.a_1a_2 \dots a_t) * \beta^e$, όπου s = πρόσημο, όπου $s = 0$ συμβολίζει ότι έχουμε θετικούς α.κ.υ. και $s = 1$ συμβολίζει ότι έχουμε αρνητικούς α.κ.υ. και β = βάση αριθμητικού συστήματος και a_1, a_2, \dots, a_t είναι τα ψηφία της ουράς. Στη συγκεκριμένη περίπτωση, επειδή έχουμε ακρίβεια δεκαδικών ψηφίων ίση με 2, διότι το $t = 2$, θα έχουμε αριθμούς της μορφής: $\pm 0.1a_2 * 2^e$, όπου το $a_2 = 0, 1$ (θεωρούμε ότι το **a₁ = 1** που είναι το hidden bit) και το $e = \pm 2, \pm 1, 0$.

Επομένως οι αριθμοί που ανήκουν στο παραπάνω σύνολο είναι οι εξής:

$$0,1a_2 * 2^e \left\{ \begin{array}{l} 0,10 * 2^0 \\ 0,10 * 2^1 \\ 0,10 * 2^2 \\ 0,10 * 2^{-1} \\ 0,10 * 2^{-2} \\ \\ 0,11 * 2^0 \\ 0,11 * 2^1 \\ 0,11 * 2^2 \\ 0,11 * 2^{-1} \\ 0,11 * 2^{-2} \end{array} \right.$$

Άρα στο σύνολο $F(2, 2, -2, 2)$ ανήκουν συνολικά **20 αριθμοί** (10 θετικοί και 10 αρνητικοί). Γνωρίζουμε ότι το ϵ_M (ή eps) = β^{1-t} και θέτοντας $\beta = 2$ (βάση δυαδικού συστήματος) και $t = 2$ (ακρίβεια δεκαδικών ψηφίων), λόγω του συνόλου F που έχει δοθεί, έχουμε ότι: $\epsilon_M = \beta^{1-t} = 2^{1-2} = 2^{-1} = \frac{1}{2}$.

- ii) Γνωρίζουμε ότι η ενδογενής συνάρτηση $\text{poly}(v)$ παίρνει ως όρισμα ένα διάνυσμα (v στην προκειμένη περίπτωση) που περιέχει τις ρίζες του πολυωνύμου και επιστρέφει ένα άλλο διάνυσμα που περιέχει τους συντελεστές του πολυωνύμου, δηλ. $c = \text{poly}(v)$. Η ενδογενής συνάρτηση $\text{roots}(c)$ παίρνει ως όρισμα ένα διάνυσμα (c στην προκειμένη περίπτωση) που περιέχει τους συντελεστές του πολυωνύμου και επιστρέφει ένα άλλο διάνυσμα που περιέχει τις ρίζες του πολυωνύμου, δηλ. $v = \text{roots}(c)$. Κανονικά, αν είχαμε **άπειρη ακρίβεια**, θα ίσχυε $v = \text{roots}(c) = \text{roots}(\text{poly}(c))$. Όμως στους α.κ.υ. κάτι τέτοιο δεν ισχύει. **Η ακρίβεια των υπολογισμένων ριζών υποβαθμίζεται (μειώνεται), καθώς αυξάνεται ο βαθμός του πολυωνύμου.** Αυτό έχει σαν αποτέλεσμα ο ακριβής υπολογισμός των ριζών πολυωνύμου υψηλού βαθμού, να καθίσταται προβληματικός.

7.2.3. Αναδρομικός κώδικας με MATLAB

Να γραφτεί πρόγραμμα για τον υπολογισμό της ακόλουθης σειράς: $I_0 = \frac{1}{e}(e - 1)$
 $I_{n+1} = 1 - (n + 1)I_n, \text{ για } n = 0, 1, 2, \dots$

Λύση

```
function I = sequence(n)
```

```
I = zeros(n+2, 1) % διάνυσμα με μηδενικά n+2 στοιχείων επειδή η ακολουθία που δίνεται  

% πηγαίνει από 0 έως n+1.  

I(1) = (exp(1) - 1)/exp(1); % I(1) = (e-1)/e, υπολογισμός του πρώτου όρου  $I_0$  της ακολουθίας. Ο λόγος  

% για τον οποίο ξεκινάμε την ακολουθία από I(1) και όχι I(0) είναι ότι στη  

% MATLAB οι δείκτες των στοιχείων ενός διανύσματος/μητρώου δεν μπορούν  

% να ξεκινούν από το «0». Ξεκινούν υποχρεωτικά από το «1».
```

```
for i = 0: n  

    I(i + 2) = 1 - (i + 1) * I(i+1); % Υλοποίηση του  $I_{n+1}$ 
```

```
end  

end
```

Η ακολουθία (σειρά) που υπολογίζεται από το πρόγραμμα αυτό δεν τείνει στο μηδέν (καθώς το n αυξάνεται) αλλά αποκλίνει με εναλλασσόμενο πρόσημο.

Παρατήρηση: Οι δείκτες διανυσμάτων/μητρώων στη MATLAB πρέπει να είναι θετικοί και ακέραιοι:

```
>> x=[1 2 3 4]
```

```
x =
```

```
1 2 3 4
```

```
>> x(0)=5
```

Subscript indices must either be real positive integers or logicals.

```
>> x(1)=5
```

```
x =
```

```
5 2 3 4
```

7.2.4. Άσκηση με σύστημα α.κ.υ. διαφορετικού πλήθους σημαντικών ψηφίων

Έστω $a = 1/3$ και $b = 0.342678$ (οι ακρίβεις τιμές). Σε σύστημα α.κ.υ με 4 σημαντικά ψηφία και στρογγύλευση, ποια θα είναι τα αποτελέσματα των παρακάτω παραστάσεων,

- a) $fl(a)$ b) $fl(a + b)$ c) $fl(a \cdot b)$ d) $fl(a/b)$

Τι συμβαίνει όταν η ακρίβεια είναι 8 σ.ψ.;

Λύση.

Πράξη	Ακρίβεια 4 σ.ψ	Ακρίβεια 8 σ.ψ.
$fl(a)$	0.3333	0.33333333
$fl(a + b)$	0.6760	0.67601133
$fl(a \cdot b)$	0.1142	0.11422600
$fl(a/b)$	0.9727	0.97273047

Προσέξτε ότι θέλουμε τις τιμές $fl(a \otimes b)$ και όχι $fl(a) \otimes fl(b)$. Μπορεί να επαληθευτεί ότι σε κάθε περίπτωση ισχύει ότι

$$\frac{|fl(x) - x|}{|x|} \leq 0.5 \cdot 10^{1-n} \text{ για } n = 4, 8.$$

Σημείωση: Στην ανωτέρω ανισότητα το δεξιό μέλος είναι το άνω φράγμα του εμπρός σχετικού σφάλματος, όταν λαμβάνονται υπόψη τα **σημαντικά** (και όχι τα δεκαδικά) ψηφία. Το n εκφράζει το πλήθος των σημαντικών ψηφίων και για n = 4 έχουμε ότι το άνω φράγμα είναι $0.5 * 10^{1-n} = 5 * 10^{-1} * 10^{-3} = 5 * 10^{-4}$. Αντίστοιχα διαμορφώνεται το άνω φράγμα του εμπρός σφάλματος για τη διπλή ακρίβεια.

7.2.5. Άσκηση υπολογισμού διωνυμικού συντελεστή με χρήση MATLAB

Να γραφεί ένα πρόγραμμα σε MATLAB (script ή function) που, για φυσικούς αριθμούς n, k με $n \geq k$, να υπολογίζει το διωνυμικό συντελεστή

Λύση

```
function [res] = binomcoef(n, k)
if (n ≥ k)
    res = factorial(n)/factorial(k);           % υπολογισμός του n!/k!
    res = res/factorial(n - k);                % υπολογισμός τελικής παράστασης
else
    error ('Wrong input! n must be >= k');    % τερματισμός της function με εμφάνιση μηνύματος
end
```

7.2.6. Άσκηση υπολογισμού σειράς Fibonacci με χρήση MATLAB

Να γραφούν τρεις συναρτήσεις σε MATLAB που να υπολογίζουν τους όρους της ακολουθίας Fibonacci. Μία συνάρτηση με αναδρομή (με χρήση της δομής if), μια δεύτερη με επαναληπτικό υπολογισμό (μέσω μιας λίστας/διανύσματος) και μια με χρήση της πινακωτής μορφής: $\begin{bmatrix} f_{n+1} \\ f_n \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} f_n \\ f_{n-1} \end{bmatrix}$, όπου χρειάζεται η δύναμη ενός μητρώου. Να συγκριθούν οι αντίστοιχοι χρόνοι εκτέλεσης για $n = 5, 25, 35$. Είναι εύκολο σε κάθε περίπτωση να τροποποιήσετε τον κώδικα σας ώστε να υπολογίζεται η ακολουθία $3, 4, 7, 11, 18, \dots$ ή μια γενικευμένη ακολουθία Fibonacci, όπου οι δύο πρώτοι όροι να δίνονται από το χρήστη. Στη σειρά Fibonacci ισχύει ότι μετά τους δύο πρώτους όρους της ακολουθίας, καθένας από τους υπόλοιπους όρους προκύπτει από το άθροισμα των δύο προηγούμενων όρων.

Λύση

Παραθέτουμε παρακάτω τις τρεις υλοποιήσεις για υπολογισμό αναδρομικά, επαναληπτικά μέσω λίστας και πολλαπλασιασμό μητρώου αντίστοιχα.

A. Αναδρομικός υπολογισμός

```
function f = fiborec (n, a, b)
if n == 1
    f = a;
elseif n ==2
    f = fiborec(n - 1) + fiborec(n - 2);
end
```

B. Επαναληπτικός υπολογισμός μέσω λίστας

```
function fibo = fibolist (n, a, b)
f = zeros(n, 1);
f(1) = a;
```

```

f(2) = b;
for k = 3: n
    f(k) = f(k - 1) + f(k - 2);
end
fibo = f(n); % θα πρέπει το τελικό αποτέλεσμα της σειράς για νόρους να καταχωρηθεί στη μεταβλητή fibo
% που επιστρέφει η function

```

Γ. Επαναληπτικός υπολογισμός μέσω υπολογισμού μητρώου

```

function f = fibomat (n, a, b)
    f = [a; b];
    A = [1 1 ; 1 0];
    f = A ^ n * f;
    f = f(1);

```

7.3 Εμπρός και πίσω σφάλμα στην επίλυση γραμμικού συστήματος

Αν θέλουμε να υπολογίσουμε το **εμπρός και πίσω σφάλμα** στην επίλυση του γραμμικού συστήματος $A * x = b$, τότε θα πρέπει να υπολογίσουμε **πρώτα** το πίσω σφάλμα και **μετά** το εμπρός σφάλμα. Πιο συγκεκριμένα, το πίσω σφάλμα $\eta = \frac{\|b - A\hat{x}\|}{\|A\| * \|\hat{x}\| + \|b\|}$ και το εμπρός σχετικό σφάλμα: $\frac{\|x - \hat{x}\|}{\|x\|} \leq \frac{2*\kappa(A)*\eta}{1-\kappa(A)*\eta}$. Στον τύπο αυτό το **$\kappa(A)$ είναι**

ο δείκτης κατάστασης μητρώου που υποδηλώνει τη σχετική απόστασή του από τα μη αντιστρέψιμα μητρώα και ισούται με το γινόμενο $\|A\| * \|A^{-1}\|$ για **οποιαδήποτε** νόρμα του μητρώου A . **Η ιδανική (μικρότερη) τιμή του δείκτη κατάστασης είναι 1** και όσο μεγαλύτερη είναι η τιμή του, τόσο μεγαλύτερο σφάλμα προκαλείται από την επίλυση του γραμμικού συστήματος $A * x = b$, λόγω απώλειας δεκαδικών ψηφίων. **Εδώ θα πρέπει να σημειωθεί** ότι το πίσω σφάλμα στην περίπτωση αυτή ονομάζεται και **δείκτης κατάστασης αλγόριθμου**.

Παράδειγμα: Η επίλυση ενός γραμμικού συστήματος $A * x = b$ υλοποιημένη με αριθμητική IEEE – 754, διπλής ακρίβειας (64 bits), επιστρέφει ένα διάνυσμα \tilde{x} ή \hat{x} ως **προσέγγιση** της ακριβούς λύσης x . Γνωρίζουμε τα ακόλουθα στοιχεία για τις **δεύτερες νόρμες** των εμπλεκομένων ποσοτήτων: $\|A\| = 10^2$, $\|\tilde{x}\| = 10^{-1}$, $\|r\| = 10^{-10}$, $\|A^{-1}\| = 10^9$, $\|b\| = 1$. Με το συμβολισμό r (residual, υπόλοιπο ή κατάλοιπο) εννοούμε τη διαφορά $r = b - A * \tilde{x}$ Αν είχαμε ακριβή λύση x τότε η διαφορά $b - A * \tilde{x}$ θα ήταν ίση με «0». Να υπολογιστεί ο **δείκτης κατάστασης του μητρώου, τα ψηφία που χάνονται εξαιτίας του** και τέλος να βρεθεί το **εμπρός σχετικό σφάλμα** της λύσης.

Λύση

Ο δείκτης κατάστασης του μητρώου A είναι: $\kappa(A) = \|A\| * \|A^{-1}\| = 10^2 * 10^9 = 10^{11}$. Ανάλογα με τη νόρμα που χρησιμοποιούμε, βρίσκουμε κάθε φορά και διαφορετική τιμή για το δείκτη κατάστασης του μητρώου. Τα ψηφία που χάνονται λόγω του δείκτη κατάστασης είναι $\log_{10}(\kappa(A))$ και στην προκειμένη περίπτωση έχουμε $\log_{10}10^{11} = 11$. Αρχικά υπολογίζουμε το **πίσω σφάλμα**: $\eta = \frac{\|b - A\tilde{x}\|}{\|A\| * \|\tilde{x}\| + \|b\|} = \frac{\|r\|}{\|A\| * \|\tilde{x}\| + \|b\|} = \frac{10^{-10}}{10^2 * 10^{-1} + 1} = \frac{10^{-10}}{11}$.

Στη συνέχεια το **εμπρός σχετικό σφάλμα** είναι: $\frac{\|x - \hat{x}\|}{\|x\|} = \frac{\|\hat{x} - x\|}{\|x\|} \leq \frac{2*\kappa(A)*\eta}{1-\kappa(A)*\eta} = \frac{2*\frac{10}{11}}{1-\frac{10}{11}} = 20$.

Παρατήρηση: Ισχύει ότι το εμπρός σφάλμα \leq δείκτης κατάστασης προβλήματος * πίσω σφάλμα \leq $\text{cond}(f; x)$ * πίσω σφάλμα \leq $\text{cond}(f; x)$ * $\text{cond}(f_{\text{prog}})$ * u. Για να βρούμε ένα όσο το δυνατό μικρότερο άνω φράγμα για το εμπρός σφάλμα, θα πρέπει τόσο ο δείκτης κατάστασης προβλήματος όσο και το πίσω σφάλμα να είναι μικρά (ειδικά το πίσω σφάλμα να είναι φραγμένο από ένα όσο το δυνατόν μικρότερο άνω φράγμα). Γενικά το άνω φράγμα τόσο για το εμπρός όσο και για το πίσω σφάλμα πρέπει να είναι όσο το δυνατόν πιο μικρά, για να έχουμε μεγαλύτερη ακρίβεια υπολογισμών. **Επίσης, όσο μικρότερος είναι ο δείκτης κατάστασης ενός μητρώου, τόσο λιγότερα δεκαδικά ψηφία χάνονται.** Στο παραπάνω άνω φράγμα του εμπρός σφάλματος, το u είναι η μονάδα στρογγύλευσης του συστήματος των α.κ.υ. και χρησιμοποιείται ως άνω φράγμα σφαλμάτων.

7.4 Προς τα πίσω ανάλυση – προς τα πίσω ευστάθεια

Ένας αλγόριθμος χαρακτηρίζεται προς τα πίσω ευσταθής στο x, αν ισχύει ότι: $f_{\text{prog}}(x) = f(x_{\text{prog}}) = f(x^*)$. Με το συμβολισμό $f_{\text{prog}}(x)$ εννοούμε υλοποίηση (program) με σφάλμα σε δεδομένα χωρίς σφάλμα και με το συμβολισμό $f(x_{\text{prog}}) = f(x^*)$ εννοούμε υλοποίηση (program) χωρίς σφάλμα (f) σε δεδομένα με σφάλμα (x_{prog}), δηλαδή σε δεδομένα που ανήκουν στο σύνολο F. Η f_{prog} αποτελείται από όλες τις στοιχειώδεις πράξεις και με τη σειρά που εκτελούνται από το πρόγραμμα που υλοποιεί την f. Αν συμβαίνει αυτό για κάθε x στο πεδίο ορισμού της f, τότε ο αλγόριθμος ονομάζεται προς τα πίσω ευσταθής.

Η ευστάθεια του αλγόριθμου αναφέρεται στο κατά πόσο είναι δυνατό η υπολογισμένη από τον αλγόριθμο και με πράξεις α.κ.υ. λύση (προσεγγιστική λύση $\rightarrow f_{\text{prog}}(x)$), να θεωρηθεί ως ακριβής λύση $\rightarrow f(x_{\text{prog}})$ του ίδιου προβλήματος με τον ίδιο αλγόριθμο, ενδεχομένως όμως με λίγο τροποποιημένα δεδομένα εισόδου, δηλ. αντί για το x να χρησιμοποιηθεί το x_{prog} . Αν αυτό είναι εφικτό, τότε ο αλγόριθμος είναι (χαρακτηρίζεται) πίσω ευσταθής.

Αν κατασκευάσουμε $x_{\text{prog}} = x^* \in F$ κοντά στο $x \in R$ τέτοιο ώστε αν το αποτέλεσμα που υπολογίστηκε με πράξεις α.κ.υ. είναι $z_{\text{prog}} = f_{\text{prog}}(x)$, τότε σε περίπτωση ευστάθειας αλγόριθμου ισχύει ότι το $z_{\text{prog}} = f(x_{\text{prog}})$ και τότε το εμπρός σφάλμα $\|z_{\text{prog}} - z\| = \|f_{\text{prog}}(x) - f(x)\| = \|f(x_{\text{prog}}) - f(x)\|$. Αντί να εκτιμούμε το προς τα εμπρός σφάλμα απευθείας, εκτιμούμε πρώτα τη διαφορά $\|f(x_{\text{prog}}) - f(x)\|$ που αφορά την ευαισθησία της f στις αλλαγές της εισόδου. Αν υπάρχει x_{prog} τέτοιο ώστε το $\|x_{\text{prog}} - x\|$ να είναι μικρό, τότε το πρόβλημα ανάγεται στο μαθηματικό πρόβλημα εύρεσης της f σε μικρές διαταραχές των στοιχείων εισόδου (δηλ. αντί για την είσοδο x παίρνουμε την είσοδο x_{prog}).

7.5 Χαρακτηριστικά μεγέθη και πράξεις μεταξύ α.κ.υ. – Πρότυπο IEEE 754

Ένας κανονικοποιημένος α.κ.υ. που ακολουθεί το συγκεκριμένο πρότυπο έχει την εξής γενική μορφή: $(-1)^s \times 2^e \times (1 + \frac{d_1}{2^1} + \frac{d_2}{2^2} + \frac{d_3}{2^3} + \dots + \frac{d_t}{2^t}) = (-1)^s \times 2^e \times (1 + d_1 \times 2^{-1} + d_2 \times 2^{-2} + \dots + d_t \times 2^{-t})$, όπου s = πρόσημο, e = bits

εκθέτη, d_i = bits ουράς. Ο «1» πριν από τα ψηφία της ουράς δηλώνει ότι ο αριθμός είναι κανονικοποιημένος. Αναφορικά με το κρυμμένο (hidden) bit, οι κανονικοποιημένοι αριθμοί έχουν **μοναδική αναπαράσταση** σε αντίθεση με τους υποκανονικοποιημένους. **Η MATLAB έχει σαν προεπιλεγμένη επιλογή τη διπλή ακρίβεια.** Στη συγκεκριμένη αναπαράσταση ο εκθέτης λαμβάνει όλες τις τιμές στο διάστημα $[L_d, U_d] = [-1022, 1023]$. Στην πράξη όμως του εκθέτη είναι **πολωμένη** ως εξής $[L_d + 1023, U_d + 1023] = [-1022 + 1023, 1023 + 1023] = [1, 2046]$, προκειμένου να χρησιμοποιούνται μόνο θετικοί εκθέτες. Στην απλή ακρίβεια έχουμε $[L_s, U_s] = [-126, 127] \rightarrow [L_s + 127, U_s + 127] = [1, 254]$. Στον πίνακα που ακολουθεί, φαίνονται οι δύο μορφές αναπαράστασης:

Αναπαράσταση	Μέγεθος α.κ.υ.	Πρόσημο s	Εκθέτης e	Ουρά di
Διπλή	64	1	11	52 + 1
Απλή	32	1	8	23 + 1

τότε ο πολωμένος εκθέτης $e = 1023 - 0111111111 = 1023 - 1023 = 0$ (η πόλωση πρέπει να αφαιρεθεί) και τελικά προκύπτει ότι ο αριθμός σε **δεκαδική** μορφή είναι $N = (-1)^{\pi} * 2^{\text{εκθέτης} - \text{πόλωση}} * 1.d_i = (-1)^0 * 2^0 * (1 + 2^{-1} + 2^{-2}) = 1 + 0.75 = 1.75$.

Παρακάτω αναφέρονται ορισμένα χαρακτηριστικά μεγέθη του πρότυπου IEEE – 754 για την αναπαράσταση α.κ.υ. Το πρότυπο αυτό χρησιμοποιεί **δύο μορφές (ακρίβειες)** για την αναπαράσταση των α.κ.υ.

Απλή ακρίβεια (32 bit):	Πρόσημο (s) 1 - bit	Εκθέτης (e) 8 - bit	συντελεστής (ουρά ή mantissa) 23 - bit
Διπλή ακρίβεια(64 bit):	Πρόσημο (s) 1 - bit	Εκθέτης (e) 11 - bit	συντελεστής (ουρά ή mantissa) 52 - bit

Μορφή κανονικοποιημένου α.κ.υ. $f = (-1)^s \times 2^e \times (1 + b_12^{-1} + b_22^{-2} + \dots + b_{t-1}2^{t-1})$. Το **πρόσημο s** όταν είναι 0/1 δηλώνει ότι ο αριθμός είναι θετικός/αρνητικός αντίστοιχα και ο **συντελεστής** δηλώνει το πλήθος των δεκαδικών ψηφίων (δηλαδή των ψηφίων της ουράς) και τα ψηφία b_1, b_2, \dots, b_{t-1} είναι τα ψηφία της ουράς. Επειδή πρόκειται για δεκαδικά ψηφία, είναι υψωμένα σε αρνητικές δυνάμεις του «2». Ο εκθέτης δηλώνει το **μέγεθος (εύρος)** του αριθμού. Το «1» στην προηγούμενη αναπαράσταση δηλώνει το κρυμμένο bit (hidden bit) και για το λόγο αυτό δεν αποθηκεύεται, για να κερδίσουμε έτσι μια θέση στην αποθήκευση.

Η συνάρτηση απεικόνισης f_l αναπαριστά (απεικονίζει) αριθμούς από το σύνολο των πραγματικών αριθμών R (άπειρη ακρίβεια) στο σύνολο F των α.κ.υ. (πεπερασμένη ακρίβεια), δηλ. $f_l: R \rightarrow F$. Εξαιτίας αυτής της απεικόνισης δημιουργούνται σφάλματα στρογγύλευσης, διότι απεικονίζουμε ένα απειροσύνολο (και συνεχές) που είναι το R σε ένα πεπερασμένο σύνολο (και διακριτό) που είναι το F . Το σύνολο F των α.κ.υ. αναπαριστάνεται με τη γενική μορφή: $F(\beta, t, e_{\min}, e_{\max})$, όπου $\beta =$ βάση αριθμητικού συστήματος (συνήθως το 2), $t =$ ακρίβεια δεκαδικών Ψηφίων της ουράς, $e_{\min} =$ ελάχιστη τιμή εκθέτη και $e_{\max} =$ μέγιστη τιμή εκθέτη.

Η απλή ακρίβεια αναπαριστάνεται και με τη μορφή: **F(2, 23+1, -126, 127)** και ενεργοποιείται με την εντολή **single**. Η εμφάνιση **23 + 1** υποδηλώνει το άθροισμα των ψηφίων της ουράς και του κρυμμένου bit (hidden bit) που είναι το '**1**' που εμφανίζεται **πριν** την υποδιαστολή. Το κρυμμένο bit (hidden bit) δεν αποθηκεύεται και έτσι εξοικονομούμε ένα bit στην αποθήκευση των α.κ.υ., αλλά λαμβάνεται υπόψη στους υπολογισμούς.

Η διπλή ακρίβεια (προεπιλεγμένη ακρίβεια της MATLAB) αναπαριστάνεται εναλλακτικά και με τη μορφή: **F(2, 52+1, -1022, 1023)**. Αντίστοιχα με πριν, η εμφάνιση **52 + 1** υποδηλώνει το άθροισμα των ψηφίων της ουράς και του κρυμμένου bit (hidden bit) που είναι το '**1**' που εμφανίζεται **πριν** την υποδιαστολή. Ισχύουν ανάλογες με πριν παρατηρήσεις αναφορικά με το κρυμμένο bit. Στη συνέχεια αναφέρονται ορισμένα ενδεικτικά μεγέθη της αριθμητικής των α.κ.υ. σε διπλή ακρίβεια.

- **u (μονάδα στρογγύλευσης του συστήματος των α.κ.υ. ή μονάδα μέτρησης σφάλματος, διότι χρησιμοποιείται ως αποδεκτό άνω φράγμα για τα σφάλματα)** = $\frac{2^{1-t}}{2} = \frac{2^{1-53}}{2} = 2^{-53}$ για διπλή ακρίβεια στο δυαδικό σύστημα.
- **realmin (ελάχιστος κανονικοποιημένος α.κ.υ.)** = $2^{-1022} = 2.2251e-308 = 2.2251 \times 10^{-308}$ για διπλή ακρίβεια.
- **realmax (μέγιστος κανονικοποιημένος α.κ.υ.)** = $2^{1023} = 1.7977e+308 = 1.7977 \times 10^{308}$ για διπλή ακρίβεια.
- **EPS (έψιλον της μηχανής) ή eps** = $2 * u = 2 * 2^{-53} = 2^{-52} \rightarrow 2.2204e-16 \rightarrow 2.2204 \times 10^{-16}$.

Αυτό υποδεικνύει την ελάχιστη διακριτότητα (ακρίβεια) του συστήματος των α.κ.υ. Όταν η συνάρτηση **eps** δεν έχει ορίσματα, τότε επιστρέφει την απόσταση από το 1.0000.....00 έως τον αμέσως επόμενο (μεγαλύτερο) αριθμό διπλής ακρίβειας, δηλ. το 1.000000....001 = $1^+ = 2^{-52}$ δηλ. τότε το **eps = eps(1) = δ(1, 1⁺)**, όπου $1^+ =$ ο αμέσως επόμενος α.κ.υ. μετά το 1. Με άλλα λόγια **1 + eps = 1⁺**.

$$\text{eps}$$

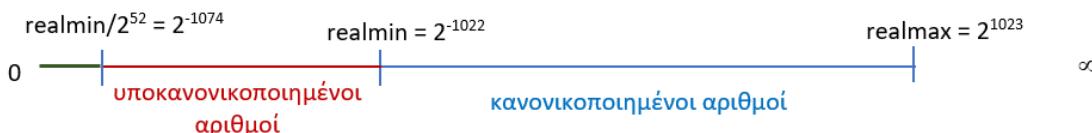
$$1 = 1.000.....00$$

$$1^+ = 1.000.....0001$$

Για οποιονδήποτε α.κ.υ. **x**, η συνάρτηση **eps(x)** επιστρέφει την απόσταση από το **abs(x)** έως τον επόμενο (μεγαλύτερο) α.κ.υ. της ίδιας ακρίβειας, δηλ. τότε **eps(x) = δ(x, x⁺)**. Ενδεικτικά μεγέθη: **eps(1/2) = 2⁻⁵³ = u**, **eps(1) = eps = 2⁻⁵²**, **eps(2) = 2⁻⁵¹** κ.λ.π. Αν στη MATLAB πληκτρολογήσουμε τις παρακάτω εντολές, που είναι σχετικές με το **eps**, θα πάρουμε τα αποτελέσματα που φαίνονται (για να παρατηρήσουμε σωστά τον αριθμό 1^+ , απαιτείται να θέσουμε ως **format** το **hex**):

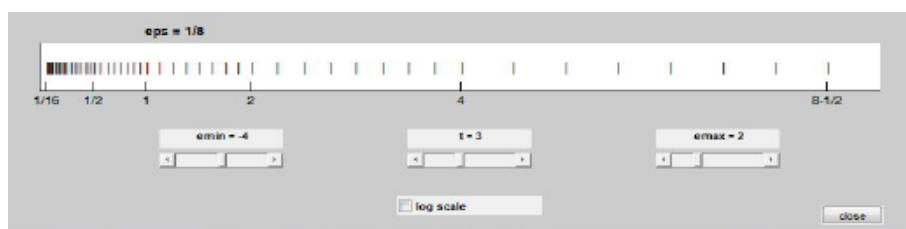
```
>> format short (απλή μορφοποίηση, στην οποία όλες οι τιμές εμφανίζονται με την προεπιλεγμένη ακρίβεια των 4 δ.ψ.)
>> 1+eps
ans = 1.0000
>> format long (μορφοποίηση, στην οποία όλες οι τιμές εμφανίζονται με ακρίβεια 15 δ.ψ.)
>> 1+eps
ans = 1.000000000000000
>> format hex (μορφοποίηση, στην οποία όλες οι τιμές εμφανίζονται με ακρίβεια 64 δ.ψ.)
>> 1+eps
ans = 3ff00000000000001 = 0011 1111 1111 0000 0000 .....0001
>> 1+eps == 1      % έλεγχος συνθήκης και επιστροφής από αυτή τιμής 1/0, αν η συνθήκη είναι αληθής/ψευδής.
ans = 0
>> 1+eps > 1      % έλεγχος συνθήκης και επιστροφής από αυτή τιμής 1/0, αν η συνθήκη είναι αληθής/ψευδής.
ans = 1
```

- Οι **κανονικοποιημένοι αριθμοί** είναι της μορφής: $1.\square\square\square\dots\square \times 2^{e_{min}}$ όπου \square = δυαδικό ψηφίο και $e_{min} = -1022$ (διπλή ακρίβεια συστήματος α.κ.υ.) ή διαφορετικά -126 (απλή ακρίβεια συστήματος α.κ.υ.)
- Οι **υποκανονικοποιημένοι αριθμοί** είναι της μορφής: $0.\square\square\square\dots\square \times 2^{e_{min}}$ όπου \square = δυαδικό ψηφίο και $e_{min} = -1022$ ή -126 και περιέχονται στο διάστημα: **[realmin * eps - realmin]**. Οι υποκανονικοποιημένοι αριθμοί προκύπτουν από τη διαίρεση του realmin διαδοχικά με το $2^1, 2^2, 2^3$ κ.ο.κ. έως το 2^{52} , δηλ. η διαίρεση με το 2^{53} δίνει ως αποτέλεσμα το μηδέν. **Δίνουν τη δυνατότητα αναπαράστασης αριθμών μικρότερων από το realmin** και αυτό το φαινόμενο ονομάζεται **βαθμιαία υποχείλιση** (gradual underflow). Ο μικρότερος υποκανονικοποιημένος α.κ.υ. που μπορεί να παρασταθεί είναι – όπως αναφέρθηκε - ο $realmin/2^{52} = realmin * 2^{-52} = realmin * eps$. Αν διαιρέσουμε τον realmin με το 2^{53} , τότε το αποτέλεσμα δίνει «0», δηλαδή $\frac{realmin}{2^{53}} = 0$. Η βαθμιαία υποχείλιση μας δίνει, όπως φαίνεται και στο επόμενο σχήμα, τη δυνατότητα να αναπαραστήσουμε αριθμούς μικρότερους από τον ελάχιστο κανονικοποιημένο α.κ.υ. (realmin). **Αν δεν υπήρχε η συγκεκριμένη δυνατότητα, τότε όλοι οι αριθμοί που θα ήταν μικρότεροι από τον realmin θα ήταν ίσοι με «0».** Τώρα όμως δεν είναι και μπορούμε να φτάσουμε έτσι έως τον αριθμό **realmin * eps**. Μετά από αυτό το συγκεκριμένο αριθμό, όλοι οι υπόλοιποι είναι ίσοι με «0». Σε αντίθεση με τη βαθμιαία υποχείλιση, δεν υπάρχει αντίστοιχη δυνατότητα βαθμιαίας υπερχείλισης, με αποτέλεσμα όλοι οι αριθμοί πάνω από το μέγιστο κανονικοποιημένο (realmax) να είναι ίσοι με το ∞ .



Εικόνα 22: Αναπαράσταση κανονικοποιημένων και μη - κανονικοποιημένων α.κ.υ.

- Η **απόσταση διαδοχικών α.κ.υ. με ίδιο εκθέτη** είναι σταθερή, δηλ. $d(e) = (m+1)2^e - m2^e = 2^e$, ενώ η **απόσταση μεταξύ διαδοχικών α.κ.υ. διπλασιάζεται** κάθε φορά που ο εκθέτης **αυξάνεται κατά 1**, δηλ. $(m+1)2^{e+1} - m2^{e+1} = m2^{e+1} + 2^{e+1} - m2^{e+1} = 2^{e+1} = 2^e 2^1 = 2 * d(e)$. Αυτό σημαίνει ότι οι α.κ.υ. δεν είναι ισοκατανεμημένοι στον οριζόντιο άξονα, αλλά είναι πιο πυκνοί όσο πλησιάζουμε προς το 0 και πιο αραιοί όσο απομακρυνόμαστε από αυτό.



Εικόνα 17: Κατανομή των α.κ.υ.

Θεώρημα της απορρόφησης: πρέπει οι **εκθέτες** των αριθμών -με βάση το 2- που αθροίζονται μεταξύ τους, να διαφέρουν από 53 και πάνω για να εφαρμοστεί. Αν ισχύει αυτό, τότε ο μεγαλύτερος α.κ.υ. απορροφά το μικρότερο, π.χ. $1 + realmin = 1 + 2^{-1022} = 2^0 + 2^{-1022} = 1$. Επίσης $realmax + realmin = 2^{1023} + 2^{-1022} = 2^{1023} = realmax$.

Αποτελέσματα θεωρήματος απορρόφησης ($10^{20} > 2^{52}$): Σε όσα από τα παρακάτω μεγέθη υπάρχει κόκκινο χρώμα σημαίνει ότι έχει εφαρμοστεί το θεώρημα της απορρόφησης κατά τον υπολογισμό τους. Οι πράξεις γίνονται προεπιλεγμένα από αριστερά προς τα δεξιά. Επίσης, ισχύει ότι $10^{-20} + 1 == 1$

Ορισμένα ενδιαφέροντα φαινόμενα:

Είσοδος	format short.	format hex
$1 + \text{eps}/2$	1	3ff0000000000000000000000000000
$1 + \text{eps}$	1.0000	3ff0000000000000000000000000001
$1 + \text{eps}/2 + \text{eps}$	1.0000	3ff0000000000000000000000000001
$1 + \text{eps} + \text{eps}/2$	1.0000	3ff0000000000000000000000000002
$1 + 2 * \text{eps}$	1.0000	3ff0000000000000000000000000002
$1 + 2 * \text{eps} + \text{eps}/2 + \text{eps}/4$	1.0000	3ff0000000000000000000000000002

$$\begin{aligned} 10^{20} - 10 - 10^{20} + 20 &= 20 \\ 10^{20} + 20 - 10^{20} - 10 &= -10 \\ -10 + 20 - 10^{20} + 10^{20} &= 0 \\ 10^{20} - 10^{20} + 20 - 10 &= 10 \end{aligned}$$

Απόλυτο σφάλμα υπολογισμού: διαφορά μεταξύ ακριβούς λύσης ενός μαθηματικού προβλήματος και της λύσης που υπολογίζεται από την αριθμητική διαδικασία: $e^{\text{abs}} = \|x - \hat{x}\| = \|\hat{x} - x\|$, όπου το $x \in \mathbb{R}^n$.

Σχετικό σφάλμα υπολογισμού: $e^{\text{rel}} = \frac{\|x - \hat{x}\|}{\|x\|} = \frac{\|\hat{x} - x\|}{\|x\|}$, με $x \neq 0$, όπου το $x \in \mathbb{R}^n$.

$$1 + 2 * \text{eps} = 1 + 2 * 2^{-52} = 1 + 2^{-51} = 3ff00000000000000 = 0011\ 1111\ 1111\ 0000\ 0000 \dots \boxed{0010}$$

7.5.1 Καταστροφική απαλοιφή

Όταν **αφαιρούνται** δύο αριθμοί που είναι **σχεδόν ίσοι** μεταξύ τους και περιέχουν σφάλματα τότε αναδεικνύονται «σκουπίδια» ακόμη και αν χρησιμοποιούσαμε αριθμητική άπειρης ακρίβειας. Πιο συγκεκριμένα, αν μια ποσότητα $A = A_1 + \text{θόρυβος}$ (ανεπιθύμητη πληροφορία πολύ μικρού μεγέθους) και $A' = A_2 + \text{θόρυβος}$ με τις τιμές **σχεδόν ίσες** μεταξύ τους, τότε η διαφορά $A_1 - A_2$ υπολογίζεται ουσιαστικά μέσω της διαφοράς $A - A'$ και είναι: $A - A' = A_1 + \text{θόρυβος} - (A_2 + \text{θόρυβος}) = (A_1 - A_2) + \text{θόρυβος}$.

Καταστροφική απαλοιφή: όταν $|A_1 - A_2| = O(\text{θόρυβος})$, όπου $O = \text{πολυπλοκότητα}$. Η μόλυνση από τα «σκουπίδια» επηρεάζει και έχει καταστροφικά αποτελέσματα στην περίπτωση που τα «σκουπίδια» πολλαπλασιαστούν με μεγάλους αριθμούς και χρησιμοποιηθούν στη συνέχεια. **Δηλ. καταστροφική απαλοιφή έχουμε όταν αφαιρούμε δύο αριθμούς που είναι πολύ κοντά ο ένας στον άλλο και στη συνέχεια το αποτέλεσμα πολλαπλασιάζεται με ένα μεγάλο αριθμό (ή αντίστοιχα διαιρείται με ένα μικρό αριθμό).**

7.5.2 Αρχή ακριβούς στρογγύλευσης

Αν $\Theta = \text{πράξη στο } R = \{+, -, *, /, ^\}$ και $\widetilde{\Theta} = \text{πράξη στο } F$ (υλοποίηση πράξης) $= \{\widetilde{+}, \widetilde{-}, \widetilde{*}, \widetilde{/}, \widetilde{^}\}$, τότε $x \widetilde{\Theta} y = f(x \Theta y) \in R$. Δηλαδή θεωρούμε ότι το αποτέλεσμα κάθε πράξης στο σύνολο F των α.κ.υ. (δηλ. στον υπολογιστή) είναι το ίδιο με το να εκτελούσαμε την πράξη αυτή με άπειρη ακρίβεια (στο R) και μετά να στρογγυλεύσουμε το αποτέλεσμα με τη συνάρτηση απεικόνισης f που αναφέρθηκε προηγουμένως. Χρησιμοποιώντας αυτή την

αρχή, μπορούμε να συμπεράνουμε ότι, όταν εκτελούμε πράξεις στον Η/Υ, το υπολογισμένο αποτέλεσμα θα περιέχει σφάλμα που διατυπώνεται στη συνέχεια:

$$a \tilde{+} b = fl(a + b) = (a + b) * (1 + \delta), \text{όπου το } |\delta| \leq u \text{ (μονάδα στρογγύλευσης του συστήματος των α.κ.υ.).}$$

$$a \tilde{-} b = fl(a - b) = (a - b) * (1 + \delta), \text{όπου το } |\delta| \leq u \text{ (μονάδα στρογγύλευσης του συστήματος των α.κ.υ.).}$$

$$a \tilde{*} b = fl(a * b) = (a * b) * (1 + \delta), \text{όπου το } |\delta| \leq u \text{ (μονάδα στρογγύλευσης του συστήματος των α.κ.υ.).}$$

$$a \tilde{/} b = fl(a / b) = (a / b) * (1 + \delta), \text{όπου το } |\delta| \leq u \text{ (μονάδα στρογγύλευσης του συστήματος των α.κ.υ.).}$$

$$a \tilde{\wedge} b = fl(a \wedge b) = (a \wedge b) * (1 + \delta), \text{όπου το } |\delta| \leq u \text{ (μονάδα στρογγύλευσης του συστήματος των α.κ.υ.).}$$

Το γινόμενο πολλών σφαλμάτων **ομαδοποιείται** σύμφωνα με τον παρακάτω τύπο: $\prod_{i=1}^n (1 + \delta_i)^{p_i} = 1 + \theta_n$ όπου $p_i = \pm 1$ και το $|\theta_n| \leq \gamma_n = \frac{nu}{1-nu}$. Επισημαίνεται ότι το γ_n αποτελεί (μαζί με τη μονάδα στρογγύλευσης u) ένα αποδεκτό άνω φράγμα σφαλμάτων (για το εμπρός ή το πίσω σφάλμα). Για παράδειγμα, αν έχουμε το γινόμενο σφαλμάτων $(1 + \delta_1) * (1 + \delta_2) * (1 + \delta_3) = 1 + \theta_3$.

Παράδειγμα: Έστω $\alpha, \beta, \gamma \in F$ και ζητάμε να τους αθροίσουμε και να βρούμε ένα άνω φράγμα για το εμπρός σφάλμα του αθροίσματος.

Απάντηση: Έστω $s = (\alpha + \beta) + \gamma \rightarrow fl(s) = ((\alpha + \beta) * (1 + \delta_1) + \gamma) * (1 + \delta_2) = (\alpha + \beta) * (1 + \delta_1) * (1 + \delta_2) + \gamma * (1 + \delta_2) \Rightarrow fl(s) = (\alpha + \beta) * (1 + \theta_2) + \gamma * (1 + \delta_2)$. Το εμπρός σχετικό σφάλμα του εν' λόγω υπολογισμού είναι: $\frac{|fl(s) - s|}{|s|} = \frac{|a*\theta_2 + \beta*\theta_2 + \gamma*\delta_2|}{|(\alpha+\beta)+\gamma|} = \frac{|\theta_2*(\alpha+\beta+\gamma)|}{|\alpha+\beta+\gamma|} = |\theta_2| * \frac{|\alpha+\beta+\gamma|}{\alpha+\beta+\gamma} = |\theta_2| \leq \gamma_2 = \frac{2u}{1-2u}$

Σημείωση: Όσον αφορά τις πράξεις στο F θα πρέπει να επισημανθεί ότι σε αυτές **δεν** ισχύει η προσεταιριστική ιδιότητα για την άθροιση και τον πολλαπλασιασμό. Επίσης, **δεν** ισχύει η επιμεριστική ιδιότητα.

7.5.3 Επιπτώσεις της αριθμητικής πεπερασμένης ακρίβειας

Αν στη MATLAB πληκτρολογήσουμε τον παρακάτω κώδικα:

```
>> d = 0; while (d ~= 1.0), d = d + 0.1; end;
```

και στη συνέχεια τον εκτελέσουμε, τότε θα πέσουμε σε **ατέρμονο βρόχο** και αυτό γιατί η συνθήκη που εξετάζεται στην αρχή της επανάληψης είναι πάντα αληθής. Αυτό συμβαίνει διότι η μαθηματική τιμή του 0.1 απαιτεί μια άπειρη σειρά για να αναπαρασταθεί που είναι η εξής: $0.1 = \frac{1}{10} = \frac{1}{2^4} + \frac{1}{2^5} + \frac{0}{2^6} + \frac{0}{2^7} + \frac{1}{2^8} + \frac{1}{2^9} + \frac{0}{2^{10}} + \frac{0}{2^{11}}$

$+ \frac{1}{2^{12}} + \dots$. Μετά τον πρώτο όρο ($\frac{1}{2^4}$), η ακολουθία των συντελεστών 1, 0, 0, 1 επαναλαμβάνεται απείρως συχνά. Επομένως το άθροισμα $d + 0.1$ δεν θα φτάσει ποτέ να γίνει ίσο με 1.0. Αν εκτελέσουμε τον κώδικα:

```
>> for j = 1: 100, if (1/j) * j ~= 1, j, end; end
```

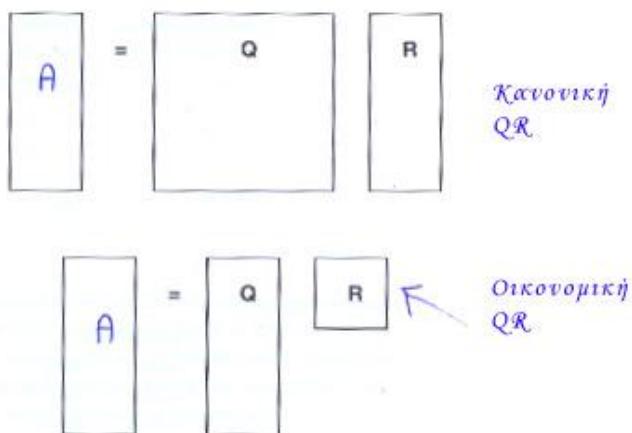
7.6 Παραγοντοποίηση QR

Χρησιμοποιείται σε πολλές εφαρμογές και αποτελεί **υπολογιστικό πυρήνα** για πολλά προβλήματα, όπως το πρόβλημα των ελαχίστων τετραγώνων, εύρεσης ΟΚ βάσεων Δ.Υ. (Διανυσματικών Υποχώρων), λύση υπερκαθορισμένων συστημάτων ($m \geq n$) μέσω ελαχίστων τετραγώνων κ.λ.π. Αποτελεί επίσης **μέθοδο υπολογισμού ιδιοτιμών** - **ιδιοδιανυσμάτων γενικών μητρώων**. Μια άλλη **χρήση** της QR παραγοντοποίησης είναι για την **επίλυση του γραμμικού συστήματος $A * x = b$** (άμεση μέθοδος). Στην περίπτωση αυτή, έχει τις διπλάσιες πράξεις σε σχέση με την LU ($\Omega = \frac{4}{3}n^3 + 2n^2$ έναντι του $\Omega = \frac{2}{3}n^3 + 2n^2$ της LU και του $\Omega = \frac{1}{3}n^3 + 2n^2$ της παραγοντοποίησης Cholesky), αλλά από την άλλη είναι ως μέθοδος **πάντα πίσω ευσταθής (κάτι που σημαίνει ότι το πίσω σφάλμα είναι μικρό, δηλαδή φραγμένο, και ο δείκτης κατάστασης αλγόριθμου $\text{cond}(f_{\text{prog}}) = 1$)**. Θυμίζουμε ότι η LU παραγοντοποίηση με μερική οδήγηση είναι **σχεδόν πάντα πίσω ευσταθής**. Η παραγοντοποίηση Cholesky είναι **πάντα πίσω ευσταθής**, όπως επίσης η εμπρός και η πίσω αντικατάσταση. Το κόστος $\frac{4}{3}n^3$ είναι για τη διάσπαση (παραγοντοποίηση) του μητρώου A σε μητρώα Q και R αντίστοιχα και το κόστος $2n^2$ είναι για την εκτέλεση της εμπρός και της πίσω αντικατάστασης. **Παρατηρούμε ότι όλες οι άμεσες μέθοδοι περιέχουν εμπρός και πίσω αντικατάσταση, προκειμένου να υπολογίσουν τη λύση x του γραμμικού συστήματος $A * x = b$.** Σε αυτό που **διαφέρουν** είναι ο αλγόριθμος με τον οποίο διασπούν το μητρώο των συντελεστών A. Ενδεικτικά αναφέρεται ότι στη MATLAB υπάρχουν τρεις διαφορετικές έτοιμες συναρτήσεις για αυτές τις τρεις παραγοντοποιήσεις:

Σύνοψη άμεσων μεθόδων			
Παραγοντοποίηση	Αλγόριθμος	Κόστος πράξεων (Ω)	Πίσω ευστάθεια
LU $A = L * U$	$[L, U] = \text{lu}(A)$ $L * y = b$ (εμπρός αντικατάσταση) $U * x = y$ (πίσω αντικατάσταση) Αν ενώσουμε εμπρός/πίσω αντικατάσταση: $x = U \setminus y = U \setminus (L \setminus b)$.	$\Omega = \frac{2}{3}n^3 + 2n^2$	Όχι πάντα
LU με μερική οδήγηση PLU $P * A = L * U$	$[P, L, U] = \text{lu}(A)$ $L * y = P * b = \hat{b}$ (εμπρός αντικατάσταση) $U * x = y$ (πίσω αντικατάσταση) Αν ενώσουμε εμπρός/πίσω αντικατάσταση: $x = U \setminus y = U \setminus (L \setminus \hat{b})$.	$\Omega = \frac{2}{3}n^3 + 2n^2$	Σχεδόν πάντα
Cholesky $A = L * L^T$	$L = \text{chol}(A)$, όπου το $L = \text{άνω τριγωνικό}$ $[L, L1] = \text{chol}(A)$ $L * y = b$ (εμπρός αντικατάσταση) $L^T * x = y$ (πίσω αντικατάσταση) Αν ενώσουμε εμπρός και πίσω αντικατάσταση: $x = L^T \setminus y = L^T \setminus (L \setminus b)$, όπου L^T = κάτω τριγωνικό μητρώο και το L = κάτω τριγωνικό. Στη μαθηματική διατύπωση της Cholesky το μητρώο L είναι κάτω τριγωνικό, δηλ. $A = L$ (κάτω τριγωνικό) * L^T (άνω τριγωνικό).	$\Omega = \frac{1}{3}n^3 + 2n^2$	Πάντα

QR $A = Q * R$	$[Q, R] = qr(A)$ $x = R \setminus (Q^T * b)$ (εμπρός και πίσω αντικατάσταση)	$\Omega = \frac{4}{3} n^3 + 2n^2$	Πάντα
--	---	-----------------------------------	-------

Για να διασπαστεί το μητρώο A σε QR παραγοντοποίηση, θα πρέπει το $A \in R^{m \times n}$ όπου $m \geq n$ (δηλαδή θα πρέπει ο αριθμός των γραμμών να είναι τουλάχιστον ίσος με τον αριθμό των στηλών). Τότε το $A = Q * R$, όπου $Q \in R^{m \times n}$ είναι **ορθογώνιο μητρώο** (οι στήλες του είναι ορθογώνιες μεταξύ τους, $Q^T = Q^{-1}$) και $R \in R^{n \times n}$ είναι μητρώο άνω τριγωνικό. **Εναλλακτικά** η QR παραγοντοποίηση μπορεί να διατυπωθεί ως εξής: Υπάρχει παραγοντοποίηση $A = Q * \begin{bmatrix} R_1 \\ 0 \end{bmatrix}$, όπου $Q \in R^{m \times n}$ **ορθογώνιο μητρώο** και $R = \begin{bmatrix} R_1 \\ 0 \end{bmatrix}$ άνω τριγωνικό. Αν οι στήλες του μητρώου A είναι **γραμμικά ανεξάρτητες**, τότε το **μητρώο R_1 αντιστρέφεται**. Αν επιλέξουμε τα διαγώνια στοιχεία θετικά, τότε οι παράγοντες Q και R είναι μοναδικοί. Αυτή είναι η οικονομική (λεπτή) QR παραγοντοποίηση. Η MATLAB επιλύει το πρόβλημα των **ελαχίστων τετραγώνων με την QR παραγοντοποίηση**, η οποία έχει δύο παραλλαγές: Στην **πλήρη (κανονική) παραλλαγή**, το μητρώο R έχει το ίδιο μέγεθος με το A και το μητρώο Q είναι **τετραγωνικό** και έχει τον ίδιο αριθμό γραμμών με το A . Στην **οικονομική (λεπτή) παραλλαγή**, το μητρώο Q έχει το ίδιο μέγεθος με το A και το μητρώο R είναι τετραγωνικό και έχει τον ίδιο αριθμό στηλών με το A . Αυτά φαίνονται στο επόμενο σχήμα:



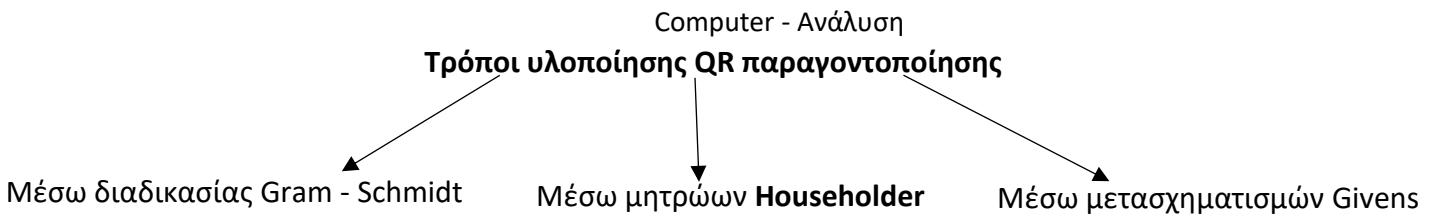
Εικόνα 23: Τύποι παραγοντοποίησης QR

Ιδιότητα QR παραγοντοποίησης αναφορικά με ΟΚ βάσεις θεμελιωδών Δ.Χ.

Αν $A = QR$ είναι η QR παραγοντοποίηση του $A \in R^{m \times n}$ που είναι πλήρους τάξης (δηλ. $\text{rank}(A) = r = m = n$), τότε το $Q = [q_1, \dots, q_n, q_{n+1}, \dots, q_m] = [Q_1, Q_2]$, όπου οι στήλες των Q_1, Q_2 είναι ΟΚ βάσεις⁵ για το **χώρο στηλών**⁶ και για τον **αριστερό μηδενόχωρο**, δηλαδή $\text{range}(A) = \text{span}\{q_1, \dots, q_n\}$ και $\text{null}(A^T) = \text{span}\{q_{n+1}, \dots, q_m\}$.

⁵ Με τον όρο **βάση** ενός Δ.Χ. εννοούμε ένα πλήθος διανυσμάτων που είναι γραμμικά ανεξάρτητα και παράγονται το Δ.Χ. Ένας Δ.Χ. μπορεί να έχει πολλές βάσεις, σίγουρα μεταξύ αυτών θα περιέχεται και η τυπική ή στοιχειώδης βάση (περιέχει τα στοιχειώδη διανύσματα e_1, e_2, \dots, e_n). Μια βάση που είναι ΟΚ (ορθοκανονική) περιέχει γ.α. διανύσματα, τα οποία είναι κάθετα μεταξύ τους.

⁶ Υπάρχουν τέσσερις θεμελιώδεις Δ.Χ., που είναι ο χώρος στηλών $\text{range}(A)$, ο χώρος γραμμών $\text{range}(A^T)$, ο μηδενόχωρος $\text{null}(A)$ και ο αριστερός μηδενόχωρος $\text{null}(A^T)$.



Υπάρχουν **τρεις τρόποι υπολογισμού της QR παραγοντοποίησης**: μέσω ορθογώνιων μετασχηματισμών Householder, μέσω ορθογώνιων μετασχηματισμών Givens, μέσω διαδικασίας Gram - Schmidt. Θα ασχοληθούμε **μόνο** με τη μέθοδο Householder. Ένα **μητρώο Householder** υπολογίζεται από τον τύπο:

$$H(n) = I - 2 * [(u * u^T) / (u^T * u)],$$

όπου **u = διάνυσμα Householder** και **H = μητρώο Householder** που υπολογίζεται από το διάνυσμα **u** και έχει ως ιδιότητες ότι είναι **συμμετρικό** ($H^T = H$) και **ορθογώνιο** ($H^T = H^{-1}$), οπότε ισχύει ότι: $H * H^T = H^T * H = I$, κάτι που εν' γένει ισχύει **μόνο** για τον πολλαπλασιασμό ενός μητρώου με τον αντίστροφό του.

7.6.1 Παράδειγμα υπολογισμού QR παραγοντοποίησης

Με μετασχηματισμούς Householder να βρεθούν οι παράγοντες Q, R όταν $A = \begin{bmatrix} 1 & -1 & 4 \\ 1 & 4 & 2 \\ 1 & 4 & 2 \\ 1 & -1 & 0 \end{bmatrix}$

Λύση

Η γενική μορφή του μητρώου Householder είναι: $H_j = I - 2 * [(u_j * u_j^T) / (u_j^T * u_j)]$ όπου **H = μητρώο Householder** και **u = διάνυσμα Householder** και μάλιστα το τελευταίο υπολογίζεται από τον τύπο: $u_j = A^{(j)} \pm \|A^{(j)}\|_2 * e_j = x \pm \|x\|_2 * e_j$, όπου $A^{(j)}$ είναι η j - οστή στήλη του μητρώου A και e_j είναι το στοιχειώδες διάνυσμα με «1» στη j - οστή θέση. Η **γενική μορφή ενός διανύσματος Householder** είναι η εξής: $u_j = [0, 0, \dots, 0, 1, u_{j+1}, u_{j+2}, \dots, u_n]^T$. Παρατηρούμε ότι στην j - οστή θέση έχει «1», στις προηγούμενες θέσεις έχει «0» και στις θέσεις μετά την j - οστή θέση περιέχει μη - μηδενικές τιμές. Στο συγκεκριμένο παράδειγμα, θα πρέπει αρχικά να υπολογίσουμε το διάνυσμα Householder u_1 , με βάση το οποίο θα ορίσουμε στη συνέχεια το μητρώο Householder H_1 . Πιο

συγκεκριμένα, $u_1 = A^{(1)} \pm \|A^{(1)}\|_2 * e_1 = x \pm \|x\|_2 * e_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} + \sqrt{4} * \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 3 \\ 1 \\ 1 \\ 1 \end{bmatrix}$. Κρατάμε το θετικό πρόσημο του τύπου,

διότι αυτό είναι το **πρόσημο του οδηγού** στην πρώτη θέση της πρώτης στήλης του μητρώου A και το x είναι ένας άλλος συμβολισμός για την πρώτη στήλη του μητρώου A . Στη συνέχεια διαιρούμε το διάνυσμα u_1 με την τιμή 3, έτσι ώστε -σύμφωνα με τη γενική μορφή του διανύσματος Householder που διατυπώθηκε προηγουμένως - το πρώτο στοιχείο του να είναι 1. Άρα το διάνυσμα $u_1 = [1 \ 1/3 \ 1/3 \ 1/3]^T$ και το αντίστοιχο μητρώο $H_1 = I - 2 * [(u_1 * u_1^T) / (u_1^T * u_1)]$. Στη συνέχεια υπολογίζουμε το γινόμενο $H_1 * A$, διότι ο στόχος είναι να μηδενίσουμε τα στοιχεία της 1^{ης} στήλης του μητρώου A που είναι κάτω από τη θέση του οδηγού, προκειμένου στο τέλος να καταλήξουμε σε ένα μητρώο άνω τριγωνικό (R). Πιο συγκεκριμένα, το γινόμενο $H_1 * A = (I - 2 * \frac{u_1 * u_1^T}{u_1^T * u_1}) * A =$

$$\left(\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} - 2 * \frac{\begin{bmatrix} 1 \\ 1/3 \\ 1/3 \\ 1/3 \end{bmatrix} * [1 & 1/3 & 1/3 & 1/3]}{[1 & 1/3 & 1/3 & 1/3] * \begin{bmatrix} 1 \\ 1/3 \\ 1/3 \\ 1/3 \end{bmatrix}} \right) * \begin{bmatrix} 1 & -1 & 4 \\ 1 & 4 & 2 \\ 1 & 4 & 2 \\ 1 & -1 & 0 \end{bmatrix} = \left(\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} - \frac{2}{4/3} * \begin{bmatrix} 1 & 1/3 & 1/3 & 1/3 \\ 1/3 & 1/9 & 1/9 & 1/9 \\ 1/3 & 1/9 & 1/9 & 1/9 \\ 1/3 & 1/9 & 1/9 & 1/9 \end{bmatrix} \right) *$$

$$\begin{bmatrix} 1 & -1 & 4 \\ 1 & 4 & 2 \\ 1 & 4 & 2 \\ 1 & -1 & 0 \end{bmatrix} = \begin{bmatrix} -2 & 3 & 2 \\ 0 & 10 & 4 \\ 0 & 10 & 0 \\ 0 & -5 & 2 \end{bmatrix} = A^{(1)}. \text{ Παρατηρούμε ότι τα στοιχεία της 1^η στήλης κάτω από τον οδηγό έχουν μηδενιστεί.}$$

Στη συνέχεια εκτελώντας το 2^o βήμα της διαδικασίας, υπολογίζουμε το διάνυσμα Householder u_2 , το οποίο θα έχει και αυτό θετικό πρόσημο, λόγω του θετικού προσήμου που έχει το στοιχείο 10 στη θέση 2, 2 του μητρώου $A^{(1)} = H_1 * A$ που είναι οδηγός για τη δεύτερη στήλη. Πιο συγκεκριμένα, το διάνυσμα $u_2 = A^{(2)} \pm \|A^{(2)}\|_2 * e_2 = \begin{bmatrix} 0 \\ 10 \\ 10 \\ -5 \end{bmatrix} + \sqrt{0 + 100 + 100 + 25} * \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 25 \\ 10 \\ -5 \end{bmatrix}$ και εκτελούμε απλοποίηση με 25, οπότε το διάνυσμα $u_2 = [0 \ 1 \ 2/5 \ -1/5]^T$

$$\text{το γινόμενο } H_2 * (H_1 * A) = H_2 * A^{(1)} = (I - 2 * [(u_2 * u_2^T)/(u_2^T * u_2)]) * \begin{bmatrix} -2 & 3 & 2 \\ 0 & 10 & 4 \\ 0 & 10 & 0 \\ 0 & -5 & 2 \end{bmatrix} = \begin{bmatrix} -2 & 3 & 2 \\ 0 & 35 & 10/3 \\ 0 & 0 & -44/15 \\ 0 & 0 & 52/15 \end{bmatrix} = A^{(2)}$$

Παρατηρούμε ότι τα στοιχεία της 2^{ης} στήλης κάτω από τον οδηγό του μητρώου $A^{(2)}$ μηδενιστεί. Τέλος, στο τρίτο βήμα υπολογίζουμε το $u_3 = A^{(3)} \pm \|A^{(3)}\|_2 * e_3 = \begin{bmatrix} 0 \\ 0 \\ -44/15 \\ 52/15 \end{bmatrix} - \sqrt{0 + 0 + (-44/15)^2 + (52/15)^2} * \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 2 \\ -1 \end{bmatrix}$ και το

διαιρούμε με το 2 και γίνεται $u_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ -1/2 \end{bmatrix}$. Με βάση αυτό υπολογίζουμε στη συνέχεια το αντίστοιχο μητρώο Householder H_3 και εκτελούμε τη πράξη $H_3 * (H_2 * (H_1 * A)) = \begin{bmatrix} -2 & -3 & -4 \\ 0 & -5 & 0 \\ 0 & 0 & 2.82 \\ 0 & 0 & 0 \end{bmatrix} = R$.

Υπάρχει άνω τριγωνικό μητρώο R ακόμη και αν το αρχικό μητρώο A είναι μη - τετραγωνικό, όπως συμβαίνει στην προκειμένη περίπτωση που το $A \in \mathbb{R}^{4 \times 3}$. Πρόκειται για την **πλήρη QR**, αφού το R έχει ίδιες διαστάσεις με το A .

Συνοψίζοντας έχουμε ότι: $H_3 * H_2 * H_1 * A = R \Rightarrow A = (H_3 * H_2 * H_1)^{-1} * R = (H_1^{-1} * H_2^{-1} * H_3^{-1}) * R = (H_1^T * H_2^T * H_3^T) * R = (H_1 * H_2 * H_3) * R = Q * R$, διότι το μητρώο Householder είναι συμμετρικό ($H^T = H$) και ορθογώνιο ($H^T = H^{-1}$). Επομένως το ορθογώνιο μητρώο $Q = H_1 * H_2 * H_3$. Γενικεύοντας τον τελευταίο τύπο έχουμε ότι το $Q = H_1 * H_2 * H_3 \dots * H_n$, όπου $n = \text{στήλες του μητρώου } A$.

7.6.2 Παράδειγμα υπολογισμού προσεγγιστικής λύσης γραμμικού συστήματος μέσω QR παραγοντοποίησης

Αν θέλουμε να χρησιμοποιήσουμε τη μέθοδο των ελαχίστων τετραγώνων μέσω QR παραγοντοποίησης για να υπολογίσουμε μια προσεγγιστική λύση στο γραμμικό σύστημα: $A * x = b \Rightarrow A^T * A * x = A^T * b$ (μέθοδος κανονικών εξισώσεων) $\Rightarrow (Q * R)^T * (Q * R) * x = (Q * R)^T * b \Rightarrow (R^T * Q^T) * (Q * R) * x = R^T * Q^T * b \Rightarrow R^T * (Q^T * Q) * R * x = R^T * Q^T * b \Rightarrow R^T * I * R * x = R^T * Q^T * b \Rightarrow x = (R^T * R)^{-1} * R^T * Q^T * b = R^{-1} * (R^T)^{-1} * R^T * Q^T * b \Rightarrow x = R^{-1} * I * Q^T * b \Rightarrow x = R^{-1} * Q^T * b \Rightarrow R * x = R * R^{-1} * Q^T * b \Rightarrow R * x = Q^T * b$ ή $R * x_i = Q^T * b_i$ αν έχουμε σύστημα.

Παρατήρηση 1: Στην εφαρμογή της μεθόδου των κανονικών εξισώσεων $A^T * A * x = A^T * b$, όπου $B = A^T * A$ είναι συμμετρικό, αν οι στήλες του μητρώου A είναι γραμμικά ανεξάρτητες ισχύει: $x^T * A^T * A * x \geq 0$, δηλαδή το μητρώο $A^T * A$ είναι συμμετρικά θετικά ορισμένο (ΣΘΟ).

Παρατήρηση 2: η στοιχειώδης προβολή⁷ είναι απλή περίπτωση του παραπάνω τελεστή. Έπειτα ότι η λύση του ΑΓΑ.2 (πρόβλημα ελαχίστων τετραγώνων) δίνεται από το $x \in \mathbb{R}^n$ τέτοιο ώστε: $A * x = P * b = A(A^T A)^{-1} A^T b$ και θέτουμε: $x = (A^T A)^{-1} A^T b$. Επομένως το x ικανοποιεί το σύστημα των κανονικών εξισώσεων: $A^T A * x = A^T b$. Υποθέσαμε ότι οι στήλες του A είναι γραμμικά ανεξάρτητες, επομένως το $A^T A$ είναι Σ.Θ.Ο. και αντιστρέψιμο. Μπορούμε να ορίσουμε την ακόλουθη μέθοδο επίλυσης του ΑΓΑ.2.

Αλγόριθμος. [Μέθοδος των κανονικών εξισώσεων]

1. Υπολογισμός κάτω τριγωνικού τμήματος του $C = A^T A$ και του $d = A^T b$.
2. Παραγοντοποίηση Cholesky $C = GG^T$.
3. Επίλυση του $Gy = d$ και του $G^T x = y$.

Το αριθμητικό κόστος της μεθόδου είναι: $T_{\alpha\rho\theta} = mn^2 + n^3/3 + O(n^2)$ πράξεις α.κ.υ.

Τα μειονεκτήματα της μεθόδου των κανονικών εξισώσεων είναι τα εξής:

- Ο πολλαπλασιασμός $A^T A$ μπορεί να καταστρέψει όποια ειδική δομή έχει το αρχικό μητρώο A (π.χ. πολλά μηδενικά). Αυτό σημαίνει ότι ενώ το αρχικό μητρώο A είναι αραιό, το γινόμενο $A^T A$ το κάνει πυκνό.
- Ο πολλαπλασιασμός $A^T A$ μπορεί να έχει σαν αποτέλεσμα την απώλεια σημαντικών ψηφίων και αλλοίωση του αποτελέσματος ή την υπερχείλιση. Να σημειωθεί πάντως ότι οι σημερινές αρχιτεκτονικές προσφέρουν μεγάλο φάσμα αναπαράστασης για την α.κ.υ. οπότε το πρόβλημα της υπερχείλισης είναι πιο σπάνιο. Για παράδειγμα, σε σύστημα με επεξεργαστή Pentium III υπερχείλιση δημιουργείται όταν πρέπει να παρασταθεί αποτέλεσμα που είναι μεγαλύτερο του $realmax = 2^{1023} \approx 1.7977 \times 10^{308}$.
- Ίσως το πιο σημαντικό αντικίνητρο για τη χρήση των κανονικών εξισώσεων για το ΑΓΑ.2 είναι η ευαισθησία της μεθόδου σε συσσώρευση αριθμητικών σφαλμάτων. Συγκεκριμένα, μπορεί να αποδειχθεί πως η ακρίβεια

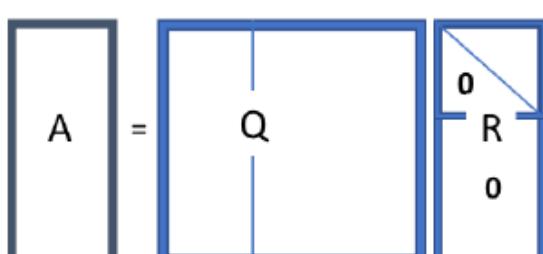
⁷ Όταν θέλουμε να προβάλουμε ένα διάνυσμα σε έναν υπόχωρο, θα πρέπει να πολλαπλασιάσουμε το διάνυσμα αυτό με το μητρώο ορθογώνιας προβολής $P = A * (A^T * A)^{-1} * A^T$. Δηλαδή η προβολή του διανύσματος x είναι το γινόμενο $P * x$.

των αποτελεσμάτων εξαρτάται από το δείκτη κατάστασης $\kappa(A^T A)$. Χρησιμοποιώντας την Ευκλείδεια νόρμα έχουμε $\kappa_2(A^T A) = [\kappa_2(A)]^2$, άρα το σχετικό σφάλμα της λύσης αυξάνει με το τετράγωνο του δείκτη κατάστασης του A και η λύση είναι πιο **ευαίσθητη** σε αριθμητικά σφάλματα.

Παρατήρηση 3: Εκτός από την **πλήρη (κανονική) QR παραγοντοποίηση**, υπάρχει και η **οικονομική (λεπτή) QR παραγοντοποίηση** όπου το $A = [Q_1 \ Q_2] * \begin{bmatrix} R_1 \\ 0 \end{bmatrix} = Q_1 * R_1$, όπου το $Q_1 \in \mathbb{R}^{m \times n}$ έχει ορθοκανονικές στήλες και το $R_1 \in \mathbb{R}^{n \times n}$ είναι άνω τριγωνικό. Οι στήλες του μητρώου Q_1 αποτελούν **ορθοκανονική βάση** του χώρου στηλών του μητρώου A . Επίσης το μητρώο R_1 περιέχει τον άνω παράγοντα Cholesky του $A^T * A$, διότι ισχύει ότι: $A^T A = R_1^T * [Q_1^T * Q_1] * R_1 = R_1^T * R_1$.⁸ Δηλαδή το R_1 είναι ο παράγοντας Cholesky του Σ.Θ.Ο. μητρώου $A^T A$.

Εικόνα 19: Πλήρης (κανονική) και οικονομική (λεπτή) QR παραγοντοποίηση

Σύνοψη QR παραγοντοποίησης



Γιατί;

- Υπολογισμός ιδιοτιμών
- Επίλυση γραμμικού συστήματος
- Δημιουργία ορθοκανονικών βάσεων για θεμελιώδεις υποχώρους
- Παραγοντοποίηση χαμηλής τάξης
- Φθηνότερη εναλλακτική λύση σε σχέση με την SVD

Πως;

- Με μετασχηματισμούς Householder
- Με περιστροφές Givens
- Με ορθοκανονικοποίηση Gram-Schmidt
- QR με οδήγηση

Στη συνέχεια ακολουθεί ένα παράδειγμα εφαρμογής της QR παραγοντοποίησης. Σε αυτό φαίνεται η μορφή του ορθογώνιου μητρώου Q και του άνω τριγωνικού μητρώου R που προκύπτουν από την QR παραγοντοποίηση του μητρώου A . Παρατηρούμε ότι το μητρώο R έχει τις ίδιες διαστάσεις με το $A \in \mathbb{R}^{4 \times 2}$, ενώ το μητρώο Q

⁸ Υπενθυμίζεται ότι η παραγοντοποίηση Cholesky του ΣΘΟ μητρώου A είναι: $A = L * L^T = R^T * R$, όπου το μητρώο R ονομάζεται παράγοντας Cholesky του μητρώου A .

είναι τετραγωνικό με διαστάσεις 4×4 . Από αυτό συμπεραίνουμε ότι πρόκειται για περίπτωση **κανονικής (πλήρους) QR παραγοντοποίησης**.

A =

1.0000	0.5000
0.5000	0.3333
0.3333	0.2500
0.2500	0.2000

>> [Q,R]=qr(A); [Q,R]

ans =

-0.8381	0.5226	0.0876	0.1293	-1.1932	-0.6705
-0.4191	-0.4417	-0.5883	-0.5322	0	-0.1185
-0.2794	-0.5288	0.7763	-0.1992	0	0
-0.2095	-0.5021	-0.2089	0.8126	0	0

Εικόνα 24: Παράδειγμα κανονικής (πλήρους) QR παραγοντοποίησης

Στη συνέχεια ακολουθεί ένα ακόμη παράδειγμα εφαρμογής της QR παραγοντοποίησης. Σε αυτό φαίνονται και πάλι η μορφή του ορθογώνιου μητρώου Q και του άνω τριγωνικού μητρώου R που προκύπτουν από την QR παραγοντοποίηση του μητρώου A. Παρατηρούμε ότι στην περίπτωση αυτή το **μητρώο Q έχει τις ίδιες διαστάσεις με το $A \in \mathbb{R}^{4 \times 2}$** , ενώ το μητρώο R είναι τώρα τετραγωνικό με διαστάσεις 2×2 . Από αυτό συμπεραίνουμε ότι πρόκειται για περίπτωση **οικονομικής QR παραγοντοποίησης** και ο λόγος για τον οποίο προέκυψε έχει να κάνει με την κλήση της συνάρτησης QR που είναι $[Q, R] = qr(A, 0)$. Επίσης υπάρχει και μια δεύτερη περίπτωση κλήσης της μορφής: $X = qr(A, 0)$. Στην περίπτωση αυτή παρατηρούμε ότι το μητρώο X που επιστρέφεται ως αποτέλεσμα **έχει τις ίδιες διαστάσεις με το $A \in \mathbb{R}^{4 \times 2}$ και ως στοιχεία περιέχει τα στοιχεία του μητρώου R στις δύο πρώτες γραμμές του**

A =

1.0000	0.5000
0.5000	0.3300
0.3300	0.2500
0.2500	0.2000

>> [Q,R]=qr(A,0)

Q =

-0.8388	0.5170
-0.4194	-0.4188
-0.2768	-0.5491
-0.2097	-0.5058

R =

-1.1922	-0.6689
0	-0.1181

>> X=qr(A,0)

X =

-1.1922	-0.6689
0.2281	-0.1181
0.1505	0.4079
0.1140	0.3675

Εικόνα 25: Παράδειγμα οικονομικής QR παραγοντοποίησης

7.6.3. Σχέση QR παραγοντοποίησης και παραγοντοποίησης Cholesky

Έστω $A = QR$. Να αποδείξετε ότι το μητρώο R_1 , που είναι το “μειωμένο” R της παραπάνω παραγοντοποίησης, συμπίπτει με το άνω τριγωνικό μητρώο της Cholesky του μητρώου $A^T A$.

Λύση

Έστω η οικονομική QR παραγοντοποίηση του A : $A = Q * R = [Q_1 \ Q_2] * \begin{bmatrix} R_1 \\ 0 \end{bmatrix} = Q_1 * R_1$. Το Q_1 είναι ορθογώνιο και έτσι ισχύει το εξής: $A^T * A = (Q_1 * R_1)^T * (Q_1 * R_1) = R_1^T * Q_1^T * Q_1 * R_1 = R_1^T * I * R_1 = R_1^T * R_1 \Rightarrow A^T * A = R_1^T * R_1$, όπου το R_1 είναι άνω τριγωνικό μητρώο. Έτσι έχουμε βρει την παραγοντοποίηση Cholesky του μητρώου $A^T * A$. Η μοναδικότητα της συγκεκριμένης γραφής προκύπτει από τη μοναδικότητα της παραγοντοποίησης Cholesky. Αυτό συμβαίνει διότι το $A = L * L^T = R^T * R$.

7.6.4. Δεύτερη νόρμα αμετάβλητη κάτω από ορθογώνιους μετασχηματισμούς

Αν η πλήρης QR παραγοντοποίηση είναι της μορφής $A = Q * R$, τότε επειδή το Q είναι μητρώο ορθογώνιο, ισχύει ότι η **2^η νόρμα παραμένει αμετάβλητη κάτω από ορθογώνιους μετασχηματισμούς**. Πιο συγκεκριμένα, αν το x είναι προσεγγιστική λύση του γραμμικού συστήματος $A * x = b$, τότε η νόρμα $\|A * x - b\|_2 = \|b - A * x\|_2 = \|Q^T(b - A * x)\|_2 = \|Q^T * b - Q^T * A * x\|_2 = \|Q^T * b - R * x\|_2$. Στο προηγούμενο διάνυσμα ο πρώτος όρος μπορεί να γραφεί ως γινόμενο $Q^T * b = \begin{bmatrix} Q_1^T * b \\ Q_2^T * b \end{bmatrix}$ και ο δεύτερος όρος μπορεί να γραφεί ως γινόμενο $R * x = \begin{bmatrix} R_1 * x \\ 0 \end{bmatrix}$ τότε αν πάρουμε το τετράγωνο της προηγούμενης νόρμας θα έχουμε: $\|(A * x - b)\|_2^2 = \|Q^T * (A * x - b)\|_2^2 = \|\mathbf{Q}^T * \mathbf{A} * x - Q^T * b\|_2^2 = \|\mathbf{R} * x - Q^T * b\|_2^2 = \left\| \begin{bmatrix} R_1 x \\ Q_2^T * b \end{bmatrix} - \begin{bmatrix} Q_1^T * b \\ Q_2^T * b \end{bmatrix} \right\|_2^2 = \left\| \begin{bmatrix} R_1 x - Q_1^T * b \\ -Q_2^T * b \end{bmatrix} \right\|_2^2 = \|R_1 * x - Q_1^T * b\|_2^2 + \|Q_2^T * b\|_2^2$. Το καλύτερο που μπορεί να γίνει είναι να επιλεγεί το διάνυσμα x_{LS} που προκύπτει από την επίλυση του άνω τριγωνικού συστήματος $R_1 * x = Q_1^T * b$. Στην περίπτωση αυτή μηδενίζεται ο πρώτος όρος του προηγούμενου αθροίσματος. Ο δεύτερος όρος του προηγούμενου αθροίσματος, δηλαδή το $\|Q_2^T * b\|_2^2$ δεν μπορεί να μειωθεί περισσότερο και θα είναι το **σφάλμα** της προσέγγισης, δηλ. $\min_{x \in \mathbb{R}^n} \|A * x_{LS} - b\|_2^2 = \|Q_2^T * b\|_2^2$, όπου $x_{LS} =$ λύση ελαχίστων τετραγώνων.

Παρατήρηση: το ότι η δεύτερη νόρμα παραμένει αμετάβλητη κάτω από ορθογώνιους μετασχηματισμούς αποδεικνύεται και ως εξής: Αφού το μητρώο Q είναι ορθογώνιο, έχει ορθοκανονικές στήλες και ισχύει ότι $Q^T * Q = I$. Με το συνηθισμένο τρόπο για υπολογισμούς με Ευκλείδειες νόρμες έχουμε ότι: $\|Q * x\|_2^2 = (Q * x)^T * (Q * x) = x^T * Q^T * Q * x = x^T * x = \|x\|_2^2 \Rightarrow \|Q * x\|_2 = \|x\|_2$. Άρα αποδείξαμε ότι η δεύτερη νόρμα ενός διανύσματος x διατηρείται όταν αυτό πολλαπλασιαστεί με ένα ορθογώνιο μητρώο Q . Αυτό σημαίνει ότι παραμένει αμετάβλητη κάτω από ορθογώνιους μετασχηματισμούς.

Σημείωση: Ισχύει ότι $\|x\|_2 = \sqrt{x^T * x} \Rightarrow \|x\|_2^2 = x^T * x$

7.6.5. Άσκηση (Θέμα Φεβρουαρίου 2017) σχετική με QR παραγοντοποίηση

Έστω η κλήση $[X] = qr(A, 0)$. Ποια είναι σωστά (αν υπάρχουν);

α) Το X είναι ορθογώνιο.

β) Το X περιέχει τον παράγοντα Q της παραγοντοποίησης QR του A .

γ) Το X έχει ίδιες διαστάσεις με τον A .

δ) Το άνω τριγωνικό τμήμα του X περιέχει τον παράγοντα R της παραγοντοποίησης QR του A .

Λύση

Σωστή απάντηση είναι το (δ) ισχύει. **Επίσης** ισχύει και το (γ) διότι επιστρέφεται το μητρώο X «οικονομικού μεγέθους», που στο άνω τριγωνικό τμήμα του περιέχει τον παράγοντα R . Το (α) **δεν** ισχύει διότι επιστρέφεται το μητρώο R που είναι άνω τριγωνικό. Το (β) **δεν** ισχύει διότι αν το $m > n$ τότε υπολογίζονται **μόνο** οι πρώτες n στήλες του μητρώου Q . Στη συνέχεια παρατίθενται περιπτώσεις εφαρμογής της QR παραγοντοποίησης.

- Η συνάρτηση **$R = qr(A, 0)$** παράγει το μητρώο R «οικονομικού μεγέθους». Αν το $m > n$ τότε υπολογίζονται **μόνο** οι πρώτες n στήλες του μητρώου Q και οι πρώτες n γραμμές του μητρώου R . Αν το $m = n$ είναι το ίδιο με τη διάσπαση $R = qr(A)$

- Η συνάρτηση **$R = qr(A)$** επιστρέφει το άνω τριγωνικό μητρώο R . Θα πρέπει να παρατηρήσουμε ότι $R = chol(A^T * A)$. Από τη στιγμή που το μητρώο Q είναι συχνά πυκνό, προτιμάται η παραγοντοποίηση $[Q, R] = qr(A)$ που περιγράφεται στη συνέχεια.

- Η συνάρτηση **$[Q, R] = qr(A)$** , όπου $A \in \mathbb{R}^{m \times n}$ παράγει ένα άνω τριγωνικό μητρώο $R \in \mathbb{R}^{m \times n}$ και ένα ορθογώνιο μητρώο $Q \in \mathbb{R}^{m \times m}$, τέτοιο ώστε $A = Q * R$.

7.6.6. Άσκηση με αντίστροφη QR παραγοντοποίηση

Με τον όρο **αντίστροφη QR παραγοντοποίηση** εννοούμε την περίπτωση εκείνη στην οποία δεν γνωρίζουμε το αρχικό μητρώο A , αλλά μας δίνεται το αποτέλεσμα της QR παραγοντοποίησης του A , το οποίο και ζητάμε να υπολογίσουμε. Έστω λοιπόν ότι έχουμε εφαρμόσει τον αλγόριθμο παραγοντοποίησης QR στο μητρώο $A \in \mathbb{R}^{3 \times 3}$, π.χ. **$qr(A)$** , και ότι μας επιστρέφεται στη θέση του A ένας πίνακας με στοιχεία. Πιο συγκεκριμένα, έχουμε εκτελέσει την εντολή $qr(A)$ και έχουμε πάρει ως αποτέλεσμα το εξής:

$$\text{ans} = \begin{bmatrix} 1 & 2 & 3 \\ 1 & 1 & 2 \\ 1 & 1 & 1 \end{bmatrix}$$

1. Να υπολογίσετε τα στοιχεία του αρχικού μητρώου του A .

2. Να χρησιμοποιήσετε τα παραπάνω για να λύσετε το σύστημα $A * x = b$ όπου $b = [12, -5, 10]^T$.

Λύση

1) Όταν μας επιστρέφεται το αποτέλεσμα της QR παραγοντοποίησης ενός μητρώου A , τότε το **άνω τμήμα του περιέχει τα στοιχεία του μητρώου R . Το κάτω τριγωνικό τμήμα περιέχει τα απαραίτητα στοιχεία για τη δημιουργία των διανυσμάτων Householder και κατ' επέκταση των μητρώων Householder που χρησιμοποιήθηκαν για να άνω τριγωνοποιήσουν το A .** Δηλαδή: Θέτουμε $u_1 = [1, 1, 1]^T$ και $u_2 = [0, 1, 1]^T$. Γνωρίζουμε ότι με βάση ένα διάνυσμα Householder μπορούμε να υπολογίσουμε το αντίστοιχο $H_j = \left(I - 2 * \frac{u_j u_j^T}{u_j^T u_j} \right)$. Στην προκειμένη περίπτωση, επειδή έχουμε δύο διανύσματα Householder, που είναι τα διανύσματα u_1 και u_2 , θα υπολογίσουμε δύο αντίστοιχα μητρώα Householder H_1 και H_2 , έτσι ώστε: $H_2 * H_1 * A = R \Rightarrow A = (H_2 H_1)^{-1} * R = H_1^{-1} * H_2^{-1} * R = H_1^{-1} * H_2^T * R = [H_1 * H_2] * R = Q * R$, αφού οι ανακλαστές (μητρώα Householder) είναι ορθογώνιοι και συμμετρικοί. Το

$$\frac{u_1 u_1^T}{u_1^T u_1} = \frac{1}{3} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \quad \frac{u_2 u_2^T}{u_2^T u_2} = \frac{1}{2} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix}. \quad \text{Επομένως, κάνοντας τις πράξεις έχουμε: } A = H_1 H_2 R = \frac{1}{3} \begin{bmatrix} -1 & -4 & -9 \\ 2 & 5 & 6 \\ 2 & 2 & 3 \end{bmatrix}.$$

2) Στη συνέχεια καλούμαστε να επιλύσουμε το γραμμικό σύστημα $A * x = b$, με βάση το μητρώο A που υπολογίσθηκε από το προηγούμενο ερώτημα, όταν δίνεται το διάνυσμα $b = [12, -5, 10]^T$. Ξεκινάμε με την QR παραγοντοποίηση του A , οπότε: $A = Q * R = H_1 * H_2 * R \Rightarrow R = (H_1 * H_2)^{-1} * A \Rightarrow R = H_2^{-1} * H_1^{-1} * A \Rightarrow R = H_2^T * H_1^T * A \Rightarrow R = H_2 * H_1 * A \Rightarrow R * \mathbf{x} = H_2 * H_1 * \boxed{A * \mathbf{x}} = H_2 * H_1 * b \Rightarrow \begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{bmatrix} * \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & -1 & 0 \end{bmatrix} * \begin{bmatrix} 12 \\ -5 \\ 10 \end{bmatrix} \Rightarrow \begin{bmatrix} x_1 + 2 * x_2 + 3 * x_3 \\ x_2 + 2 * x_3 \\ x_3 \end{bmatrix} = \begin{bmatrix} 14 \\ 8 \\ 13 \end{bmatrix} \Rightarrow x_1 = 11, x_2 = -18, x_3 = 13.$

7.6.7. Άσκηση επίλυσης ενός γραμμικού συστήματος $A * x = b$ με QR παραγοντοποίηση

Έστω ένα μητρώο A το οποίο χρησιμοποιούμε ως μητρώο συντελεστών κάποιου γραμμικού συστήματος $A * x = b$. **Πως θα λύνατε το γραμμικό σύστημα αν γνωρίζατε ότι το μητρώο A είναι μη τετραγωνικό;**

Λύση

Αν το μητρώο των συντελεστών είναι μη τετραγωνικό, τότε το σύστημα $A * x = b$ θα έχει είτε μία ή καμία λύση. Σε κάθε περίπτωση μπορούμε να βρούμε την προσέγγιση ελαχίστων τετραγώνων και να παρατηρήσουμε αν το σφάλμα $\|A * x - b\|_2$ είναι «0». Αν είναι, τότε το σύστημα έχει λύση, η οποία είναι το διάνυσμα x που βρήκαμε. Αν όμως το σφάλμα είναι μη μηδενικό τότε θα έχουμε βρει το x εκείνο που ελαχιστοποιεί την απόσταση $\|A * x - b\|_2$. Για να λύσουμε το σύστημα θα χρησιμοποιήσουμε **κανονικές εξισώσεις**: $A^T * A * x = A^T * b$. Το μητρώο $A^T A$ ενδέχεται να είναι ΣΘΟ, άρα κάποιος θα μπορούσε να εφαρμόσει απαλοιφή Cholesky. **Ωστόσο αν το μητρώο $A^T A$ έχει πολύ μεγάλο δείκτη κατάστασης (σχεδόν το τετράγωνο του δείκτη κατάστασης του A),**

τότε θα δημιουργήσει πρόβλημα στην επίλυση. Αυτό που μπορούμε να κάνουμε είναι να υπολογίσουμε την QR παραγοντοποίηση του A και στη συνέχεια να λύσουμε το άνω τριγωνικό σύστημα $R^T * x = Q^T * b$ με πίσω αντικατάσταση. Αυτό γιατί, δοθείσης της παραγοντοποίησης QR του A ισχύει: $A^T * A * x = A^T * b \Rightarrow (QR)^T * (QR) * x = (QR)^T * b \Rightarrow R^T * Q^T * Q * R * x = R^T * Q^T * b \Rightarrow R^T * I * R * x = R^T * Q^T * b \Rightarrow (R^T)^{-1} * R^T * R * x = (R^T)^{-1} * R^T * Q^T * b \Rightarrow R * x = Q^T * b$. Για τον υπολογισμό του $Q^T b$ δεν χρειάζεται να κατασκευάσουμε το μητρώο Q, αρκεί να έχουμε τα διανύσματα απαλοιφής Householder u. Τότε μπορούμε να εφαρμόσουμε αναδρομικά τους ανακλαστές στο b ως εξής: Επειδή το ορθογώνιο μητρώο $Q = H_1 H_2 \dots H_n$, το γινόμενο $Q^T * b = (H_1 H_2 \dots H_n)^T * b = (H_n^T * \dots * H_1^T) * b = (H_n H_{n-1} \dots H_1) * b$ και γενικότερα ισχύει ότι το γινόμενο $H_1 H_2 \dots H_n * b = \left(I - \frac{2u_1 u_1^T}{u_1^T u_1}\right) \left(I - \frac{2u_2 u_2^T}{u_2^T u_2}\right) \dots \left(I - \frac{2u_n u_n^T}{u_n^T u_n}\right) * b$, το οποίο θα υπολογιστεί από δεξιά προς τα αριστερά ως εξής: $\left(I - \frac{2u_n u_n^T}{u_n^T u_n}\right) * b = b - \frac{2u_n(u_n^T b)}{u_n^T u_n}$, το οποίο αποτελείται από δύο εσωτερικά γινόμενα και μια αφαίρεση διανυσμάτων. Η διαδικασία αυτή επαναλαμβάνεται η φορές και με τον τρόπο μειώνουμε το **κόστος υπολογισμού⁹** καθώς και το **κόστος αποθήκευσης** του μητρώου Q το οποίο θα είναι πυκνό.

7.7 Μητρώα ορθογώνιας προβολής (ΟΠ)

Παράδειγμα 1: Ένα μητρώο $P \in R^{n \times n}$ ονομάζεται **μητρώο ορθογώνιας προβολής** επί του χώρου στηλών $\text{range}(P)$ αν διαθέτει τις εξής δύο ιδιότητες: $P^2 = P$ και $P = P^T$. Σε κάθε μητρώο προβολής αντιστοιχεί ο διανυσματικός υπόχωρος $V = \text{range}(P)$ επί του οποίου γίνεται η προβολή. Αν η **βάση** για το συγκεκριμένο διανυσματικό υπόχωρο είναι οι στήλες του μητρώου $V = (v_1, v_2, \dots, v_n)$ τότε το **μητρώο $P = V * (V^T * V)^{-1} * V^T$** είναι το **μητρώο ορθογώνιας προβολής** ενός διανύσματος επί του υπόχωρου V .

Παράδειγμα 1: Έστω το διάνοιγμα $\text{span}\{e_1, e_3\}$ στον R^4 , τότε $V = [e_1, e_3]$ και το μητρώο ορθογώνιας προβολής

$$P = V * (V^T * V)^{-1} * V^T = [e_1, e_3] * ([e_1, e_3]^T * [e_1, e_3])^{-1} * [e_1, e_3]^T = \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} * \left(\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} * \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \right)^{-1} *$$

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} * \left(\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right)^{-1} * \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} * \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} * \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} * \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} =$$

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}. \text{ Στη συνέχεια θα πρέπει το μητρώο αυτό να πολλαπλασιαστεί με το διάνυσμα του οποίου θέλουμε να βρούμε την προβολή του πάνω στο υπόχωρο } V.$$

⁹ Ο πολλαπλασιασμός μητρώων – διανυσμάτων ανήκει στην κατηγορία BLAS – 2 με κόστος υπολογισμού πράξεων $\Omega = O(n^2)$, ενώ ο πολλαπλασιασμός μητρώου με μητρώο ανήκει στην κατηγορία BLAS – 3 με κόστος υπολογισμού πράξεων $\Omega = O(n^3)$

Παράδειγμα 2: Έστω ο υπόχωρος $V = \text{span}\{a, b\}$ στο χώρο R^4 , όπου τα διανύσματα $a = [1 \ 1 \ 1 \ 1]^T$, $b = [0, 1, -1, 3]^T$, οπότε το μητρώο $V = [a, b]$ και το μητρώο ορθογώνιας προβολής P υπολογίζεται από τον τύπο:

$$P = V * (V^T * V)^{-1} * V^T = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & -1 \\ 1 & 3 \end{bmatrix} * \begin{bmatrix} 4 & 3 \\ 3 & 11 \end{bmatrix}^{-1} * \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & -1 & 3 \end{bmatrix} = \begin{bmatrix} 11 & 8 & 14 & 2 \\ 35 & 35 & 35 & 35 \\ 8 & 9 & 7 & 11 \\ 35 & 35 & 35 & 35 \\ 14 & 7 & 21 & -7 \\ 35 & 35 & 35 & 35 \\ 2 & 11 & -7 & 29 \\ 35 & 35 & 35 & 35 \end{bmatrix} = \frac{1}{35} * \begin{bmatrix} 11 & 8 & 14 & 2 \\ 8 & 9 & 7 & 11 \\ 14 & 7 & 21 & -7 \\ 2 & 11 & -7 & 29 \\ 11 * \xi_1 + 8 * \xi_2 + 14 * \xi_3 + 2 * \xi_4 \\ 8 * \xi_1 + 9 * \xi_2 + 7 * \xi_3 + 11 * \xi_4 \\ 14 * \xi_1 + 7 * \xi_2 + 21 * \xi_3 - 7 * \xi_4 \\ 2 * \xi_1 + 11 * \xi_2 - 7 * \xi_3 + 29 * \xi_4 \end{bmatrix}. \text{ Αν στη}$$

συνέχεια, με δεδομένο το μητρώο P που υπολογίσαμε, δινόταν ένα τυχαίο διάνυσμα $x = [\xi_1, \xi_2, \xi_3, \xi_4]^T$ τότε η

προβολή του διανύσματος αυτού στον υπόχωρο V θα είναι: $P * x = \frac{1}{35} * \begin{bmatrix} 11 * \xi_1 + 8 * \xi_2 + 14 * \xi_3 + 2 * \xi_4 \\ 8 * \xi_1 + 9 * \xi_2 + 7 * \xi_3 + 11 * \xi_4 \\ 14 * \xi_1 + 7 * \xi_2 + 21 * \xi_3 - 7 * \xi_4 \\ 2 * \xi_1 + 11 * \xi_2 - 7 * \xi_3 + 29 * \xi_4 \end{bmatrix}$. Για να

επαληθεύσουμε ότι $\forall x \in R^4$ η προβολή του x , δηλ. το $P * x \in V$, αρκεί να δείξουμε ότι υπάρχει πάντα διάνυσμα $z \in R^2$ που ικανοποιεί το σύστημα $V * z = P * x$. Για το σκοπό αυτό το επαυξημένο σύστημα $[V \mid P * x] =$

$$\begin{bmatrix} 1 & 0 & 11 * \xi_1 + 8 * \xi_2 + 14 * \xi_3 + 2 * \xi_4 \\ 1 & 1 & 8 * \xi_1 + 9 * \xi_2 + 7 * \xi_3 + 11 * \xi_4 \\ 1 & -1 & 14 * \xi_1 + 7 * \xi_2 + 21 * \xi_3 - 7 * \xi_4 \\ 1 & 3 & 2 * \xi_1 + 11 * \xi_2 - 7 * \xi_3 + 29 * \xi_4 \end{bmatrix} \text{ και με απαλοιφή Gauss το φέρνουμε σε ΑΓΚΜ μορφή:}$$

$$\begin{bmatrix} 1 & 0 & 11 * \xi_1 + 8 * \xi_2 + 14 * \xi_3 + 2 * \xi_4 \\ 0 & 1 & -3 * \xi_1 + \xi_2 - 7 * \xi_3 + 9 * \xi_4 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \text{ για να βρούμε την } \text{τάξη} \text{ του. Παρατηρούμε ότι η τάξη του μητρώου } V$$

δηλαδή το $\text{rank}(V) = \text{rank}([V \mid P * x])$, είναι ίση με την τάξη του επαυξημένου μητρώου $[V \mid P * x]$, κάτι που σημαίνει ότι υπάρχει **μοναδική λύση**.

Παρατήρηση: Το τετραγωνικό μητρώο προβολής P είναι **μοναδικό** διότι **δεν εξαρτάται** από τη βάση του διανυσματικού υπόχωρου $V = \text{range}(P)$.

7.8 SVD Παραγοντοποίηση

Αν $A \in R^{m \times n}$ και $U \in R^{m \times m}$ και $V \in R^{n \times n}$ τότε η **διάσπαση** του μητρώου A που έχει τη μορφή: $A = U * \Sigma * V^T$ ονομάζεται **ιδιάζουσα (SVD) παραγοντοποίηση**. Στη διάσπαση αυτή τα μητρώα U και V είναι **ορθογώνια** (θυμίζουμε ότι η χαρακτηριστική ιδιότητα ενός τέτοιου μητρώου είναι ότι το αντίστροφο μητρώο είναι ίσο με το ανάστροφο) και το Σ είναι **διαγώνιο μητρώο** με τις **ιδιάζουσες τιμές** του A στην **κύρια διαγώνιο** κατά **φθίνουσα σειρά**, δηλ. $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_{\min(m, n)}$. Με την ιδιάζουσα παραγοντοποίηση (SVD) βρίσκουμε τις ιδιάζουσες τιμές και τα ιδιάζοντα διανύσματα για **οποιαδήποτε** μητρώα, όχι όμως κατ' ανάγκη τετραγωνικά, όπως ισχύει στις **ιδιοτιμές**.

Η παραγοντοποίηση αυτή έχει τη μορφή: $A = U * \Sigma * V^T$, το μητρώο A έχει διαστάσεις $m \times n$, δηλ. $A \in R^{m \times n}$ και τα μητρώα U, V, Σ έχουν διαστάσεις $m \times m$, $n \times n$ και $m \times n$ αντίστοιχα, δηλ.: $U \in R^{m \times m}$, $V \in R^{n \times n}$ και το $\Sigma \in R^{m \times n}$

(διαγώνιο μητρώο με ίδιες διαστάσεις με το A). Το ορθογώνιο μητρώο U περιέχει στις στήλες του τα **αριστερά ιδιάζοντα διανύσματα**, το ορθογώνιο μητρώο V περιέχει στις στήλες του τα **δεξιά ιδιάζοντα διανύσματα και το διαγώνιο μητρώο Σ** , περιέχει τις **ιδιάζουσες τιμές** του A σε φθίνουσα σειρά, δηλαδή $\begin{bmatrix} \sigma_{max} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \sigma_{min} \end{bmatrix}$. Μία ιδιοτιμή ή χαρακτηριστική τιμή (eigenvalue) και ένα ιδιοδιάνυσμα (eigenvector) ενός τετραγωνικού μητρώου A , είναι ένα μονότιμο μέγεθος λ και ένα μη-μηδενικό διάνυσμα x τέτοια ώστε $A^* x = \lambda^* x$. Μία **ιδιάζουσα τιμή** (singular value) και ένα ζεύγος **ιδιαζόντων διανυσμάτων** (singular vectors) ενός ορθογώνιου μητρώου A , είναι ένα μη - μηδενικό μονότιμο μέγεθος σ και δύο μη-μηδενικά διανύσματα u και v τέτοια, ώστε $A^* v = \sigma^* u$ ή $A^T * u = \sigma^* v$. Τα ιδιάζοντα διανύσματα σχεδόν πάντοτε κανονικοποιούνται, ώστε να έχουν Ευκλείδειο μήκος ίσο προς την μονάδα, $\|u\|_2 = \|v\|_2 = 1$.

Η σημασία της **SVD παραγοντοποίησης** (διάσπασης) πέραν του ότι υπολογίζει την **τάξη** του μητρώου (πλήθος μη - μηδενικών ιδιαζουσών τιμών), είναι ότι μπορεί να **προσφέρει ορθοκανονικές βάσεις για κάθε θεμελιώδη υπόχωρο του μητρώου A** . Πιο συγκεκριμένα, από το μητρώο $U \in \mathbb{R}^{m \times m}$ προκύπτουν ορθοκανονικές βάσεις τόσο για το χώρο στηλών, δηλ. $\text{basis}(R(A)) = (u_1, u_2, \dots, u_r)$ όσο και για τον αριστερό μηδενόχωρο, δηλ. $\text{basis}(\text{null}(A^T)) = (u_{r+1}, u_{r+2}, \dots, u_m)$, όπου $r = \text{τάξη}$ του μητρώου. Επίσης, από το μητρώο $V \in \mathbb{R}^{n \times n}$ προκύπτουν ορθοκανονικές βάσεις τόσο για το χώρο **γραμμών**, δηλ. $\text{basis}(R(A^T)) = (v_1, v_2, \dots, v_r)$ όσο και για το μηδενόχωρο, δηλ. $\text{basis}(\text{null}(A)) = (v_{r+1}, v_{r+2}, \dots, v_n)$, όπου $r = \text{τάξη}$ του μητρώου. Η σημασία της SVD παραγοντοποίησης φαίνεται στην Εικόνα που ακολουθεί:

$$\begin{array}{c}
 \begin{array}{ccccc}
 A & = & U & \Sigma & V^T \\
 & & \leftarrow r \rightarrow & & \\
 \hline
 \end{array}
 \end{array}$$

$U_1 = U_{:,1:r}$	$U_2 = U_{:,r+1:m}$	$V_1 = V_{:,1:r}$	$V_2 = V_{:,r+1:n}$
$\text{range}(A)$	$\text{null}(A^T)$	$\text{range}(A^T)$	$\text{null}(A)$

$$A = U_1 \hat{\Sigma} V_1^T$$

Εικόνα 23: Σημασία SVD παραγοντοποίησης

Κεφάλαιο 8 – Άσκηση επανάληψης εφ' όλης της ύλης – Θέματα

Άσκηση 1

Τι υπολογίζουν οι παρακάτω εκφράσεις:

- (1) for $i = 1:n$, $x(i) = b(i)/L(i, i)$; $b(i+1:n) = b(i+1:n) - x(i) * L(i+1:n, i)$; end
- (2) $a = A(:, 1)$; $v = a/norm(a) + sign(a(1)) * eye(n, 1)$; $H = eye(n) - 2 * v * v'/norm(v)^2$; $B = H * A$;
- (3) $x = (A' * A) \ (A' * b)$;
- (4) $D = diag(diag(A))$; $L = -tril(A, -1)$; $U = -triu(A, 1)$; for $k = 1:m$, $x = D \backslash b + D \backslash (L + U) * x$; end

Λύση

- (1) Forward substitution for solving the lower triangular system $L * x = b$
- (2) Householder elimination of the subdiagonal entries in the first row of A . Υπολογίζουμε το διάνυσμα Householder για την πρώτη στήλη του μητρώου A και στη συνέχεια με βάση αυτό υπολογίζουμε το αντίστοιχο μητρώο Householder H_1 (ενότητα 7.6.1).
- (3) Solution of the least squares problem $\min \|Ax - b\|_2$ via the normal equation $A^T A * x = A^T * b \Rightarrow x = (A^T A)^{-1} * A^T * b = (A' * A) \ (A' * b)$.
- (4) m steps of the Jacobi iteration for solving the linear system $Ax = b$

Άσκηση 2

- (1) Please provide the $(k + 1)$ st step of a **Newton iteration**.
- (2) Please write a Matlab program performing n **Newton iterations**, starting in $x_0 = 2$ for searching a zero of $f(x) = \sin(x)$. Ο όρος “searching a zero of $f(x)$ ” σημαίνει να υπολογίσουμε μια ρίζα της συνάρτησης $f(x)$.
- (3) Please explain shortly how to use the **QR decomposition** for solving least squares problems.
- (4) Please write a Matlab program solving a least squares problem via the **QR decomposition**.

Λύση

- (1) The $(k+1)$ st step of a Newton iteration is defined as $x_{k+1} = x_k - f(x_k)/f'(x_k)$.

- (2) $x = 2$; for $k = 1:n$, $x = x - \sin(x)/\cos(x)$; end

- (3) Let $A \in \mathbb{R}^{m \times n}$, $m \geq n$, and $b \in \mathbb{R}^m$. We rewrite

$$\begin{aligned}\|Ax - b\|_2^2 &= \|QRx - b\|_2^2 = \|Rx - Q^T b\|_2^2 \\ &= \|R(1:n, :)x - (Q^T b)(1:n)\|_2^2 + \|(Q^T b)(n+1:m)\|_2^2\end{aligned}$$

so that the least squares solution is the solution of the upper triangular system $R(1:n, :)x = (Q^T b)(1:n)$.

(Υποκεφάλαιο 7.6.2)

- (4) $[Q, R] = qr(A)$; $b = Q' * b$; $x = R(1:n, :) \backslash b(1:n)$; (Υποκεφάλαιο 7.6)

Σημείωση: Αναφορικά με το (4), πρόκειται για την επίλυση του γραμμικού συστήματος $A * x = b$ με χρήση ελαχίστων τετραγώνων, μέσω QR διάσπασης. Δηλαδή, $A^T * A * x = A^T * b \Rightarrow (QR)^T * (QR) * x = (QR)^T * b \Rightarrow \dots \Rightarrow R * x = Q^T * b \Rightarrow x = R^{-1} * Q^T * b \Rightarrow x = R \backslash (Q^T * b)$ ή ισοδύναμα $x = R(1:n, :) \backslash (Q^T * b)$.

Άσκηση 3

Ένα πρόγραμμα περιέχει την εντολή $y = (sqrt(x + h) - sqrt(x))/h$ όπου οι μεταβλητές περιέχουν αριθμούς κινητής υποδιαστολής και καμία δεν είναι μηδέν. Εξηγήστε τι πρόβλημα μπορεί να προκύψει στον υπολογισμό και προτείνετε βελτίωσή του.

Λύση

Αν τιμή του $|h|/|x|$ είναι πολύ μικρή, τότε η υλοποίηση της αφαίρεσης $fl(\sqrt{x+h}) - fl(\sqrt{x})$ στη CPU μπορεί να οδηγήσει σε καταστροφική απαλοιφή και το αποτέλεσμα της αφαίρεσης να είναι σκουπίδια (καταστροφική ακύρωση σημαντικών ψηφίων). Στην περίπτωση της άσκησης, το πρόβλημα επιτείνεται από τη διαίρεση, καθώς το αποτέλεσμα της αφαίρεσης μεγεθύνεται κατά $1/h$. Σημειώνουμε ότι όποιος γνωρίζει MATLAB ΔΕΝ θα θεωρήσει πρόβλημα (όπως ανέφεραν ορισμένοι) ότι η τιμή του $x + h$ μπορεί να γίνει αρνητική. Για παράδειγμα: $x = 1.0; h = -1.1; (sqrt(x+h) - sqrt(x))/h, ans = 0.9091 - 0.2875i$

Άσκηση 4

Δίνεται συμμετρικό μητρώο $A \in \mathbb{C}^{n \times n}$, 1000×1000 , τέτοιο ώστε $\|A\|_2 = 1$ και $\|A^{-1}\|_2 \approx 10^8$. Θέλετε να λύσετε το γραμμικό σύστημα $A * x = b$ για κάποιο b . Επιλέξτε ποια από τις παρακάτω σειρές εντολών MATLAB θα είναι η ορθότερη επιλογή (χωρίς εξήγηση):

- $[R] = chol(A); x = R \setminus (R' * b)$
 - $[R] = chol(A' * A); x = R \setminus (R' \setminus (A' * b))$
 - $[L, U] = lu(A); x = U \setminus (L \setminus b)$
 - $[x] = pcg(A, b);$

Λύση

Επειδή το μητρώο δεν είναι **θετικά ορισμένο** δεν εφαρμόζονται οι παραγοντοποιήσεις Cholesky και CG, επομένως απορρίπτονται η **πρώτη** και η **τελευταία** επιλογή. Η LU δεν χρησιμοποιείται, επειδή δεν αναφέρεται από την εκφώνηση ότι συντρέχει κάποια από τις προϋποθέσεις εφαρμογής της, δηλ. τα κύρια υπομητρώα να είναι αντιστρέψιμα ή εναλλακτικά το μητρώο A να έχει ΑΔΚ κατά στήλες. Επίσης, επειδή το μητρώο έχει μεγάλο δείκτη κατάστασης $\kappa(A) = \|A\|_2 * \|A\|_2^{-1} = 10^8$, σημαίνει ότι υπάρχει μεγάλη απώλεια δεκαδικών ψηφίων, επομένως πρέπει να χρησιμοποιήσουμε την QR παραγοντοποίηση, στην οποία το $\text{cond}(f_{\text{prog}}) = 1$. Ουσιαστικά η παραγοντοποίηση $\text{chol}(A' * A)$ **ισοδυναμεί με την QR παραγοντοποίηση του άνω τριγωνικού μητρώου R_1** , διότι ισχύει ότι $A^T * A = (Q_1 * R_1)^T * (Q_1 * R_1) = R_1^T * Q_1^T * Q_1 * R_1 = R_1^T * I * R_1 = R_1^T * R_1 \Rightarrow A^T * A = R_1^T * R_1$ και ως γνωστό η QR του A : $A = Q * R = [Q_1 \ Q_2] * \begin{bmatrix} R_1 \\ 0 \end{bmatrix} = Q_1 * R_1$. Επομένως: $A * x = b \Rightarrow A^T * A * x = A^T * b \Rightarrow x = (A^T * A)^{-1} * A^T * b$.

$$* A^{-1} * A^T * b \Rightarrow x = (R_1^T * R_1)^{-1} * A^T * b \Rightarrow x = (R_1)^{-1} * (R_1^T)^{-1} * A^T * b \Rightarrow x = (R_1)^{-1} * (R_1')^{-1} * A' * b \Rightarrow x = (R_1)^{-1} * (R_1') \backslash (A' * b) \Rightarrow x = (R_1) \backslash ((R_1') \backslash (A' * b)).$$

Κανονικά για την παραγοντοποίηση Cholesky ισχύει: $A = L * L^T = R^T * R$

Άσκηση 5

Μια απλή εκδοχή της μεθόδου των συζυγών κλίσεων (CG) για την επίλυση του $A * x = b$ με μηδενικό αρχικό διάνυσμα δίνεται παρακάτω. Να δείξετε μόνον τα αποτελέσματα που εμφανίζονται στην οθόνη όταν εκτελέσετε `conjgrad ([2 1; 1 2], [1; -4])`; Επίσης να σχολιάσετε την ευστοχία της μεθόδου με βάση τα αποτελέσματα. (Υπενθύμιση: Αν μία εντολή MATLAB τελειώνει με ";", το αποτέλεσμα της εντολής δεν εμφανίζεται στην οθόνη.)

```
function [x, r] = conjgrad (A, b)
x=0; r = b; p = r; rsold=r' * r
for iter = 1: length(b)
    Ap = A * p
    alpha=rsold / (p' * Ap);
    x=x + alpha * p
    r=r - alpha * Ap;
    iter
    rsnew=r' * r;
    if sqrt (rsnew) < 1e-10, break; end
    p = r + (rsnew / rsold) * p;
    rsold = rsnew;
end;
```

Λύση

Κατ' αρχήν παρατηρούμε ότι ό,τι και να συμβεί, εκτελούνται το πολύ 2 επαναλήψεις. Εκτυπώνοντας μόνον όπου δεν υπάρχουν ερωτηματικά, τα αποτελέσματα (πιο εύκολα υπολογίζονται σε κλασματική μορφή, αλλά εδώ κρατάμε τη δεκαδική) θα είναι:

```
>>conjgrad ([2, 1 ; 1, 2], [1 ; -4]);
rsold = 17
Ap =
-2
-7
x =
0.6538
-2.6154
iter = 1
Ap =
4.5266
1.1317
x =
2
-3
iter = 2
```

Η λύση που δίνεται ικανοποιεί το σύστημα. Η μέθοδος συζυγών κλίσεων (CG) συγκλίνει καθώς το μητρώο είναι εμφανώς συμμετρικό αλλά και θετικά ορισμένο, π.χ. προκύπτει απευθείας καθώς αν $x = (\xi_1, \xi_2)^T$, τότε $x^T Ax = 2\xi_1^2 + 2\xi_2^2 + 2\xi_1\xi_2 > 0$. Ισοδύναμα, ισχύει ότι είναι ΣΘΟ γιατί οι ιδιοτιμές είναι θετικές (μπορείτε να τις υπολογίσετε απευθείας ή να

χρησιμοποιήσετε το θεώρημα Gershgorin). Επίσης, αν δεν υπάρχουν σφάλματα στρογγύλευσης, η CG βρίσκει την ακριβή λύση ενός γραμμικού συστήματος που είναι ΣΘΟ σε πλήθος βημάτων ίδιο ή μικρότερο με το μέγεθος του συστήματος. Στην περίπτωσή μας έχουμε την ακριβή απάντηση $x = (2, -3)^T$ σε 2 βήματα.

Άσκηση 6

Για το πολυώνυμο $f(x) = 5x^3 - x - 2$, να χρησιμοποιήσετε τις τιμές του στους κόμβους 0, 1, 2, 3 για να υπολογίσετε τη βαρυκεντρική του αναπαράσταση (όποιον τύπο προτιμάτε).

Λύση

Στον παρακάτω πίνακα αναγράφονται οι τιμές του πολυωνύμου και τις αντίστοιχες τιμές των συντελεστών w_i που εμφανίζονται στον δεύτερο τύπο της βαρυκεντρικής αναπαράστασης του πολυωνύμου παρεμβολής:

x_i	$f(x_i)$	$1/w_i$
0	-2	$-6 = (-1)(-2)(-3)$, διότι π.χ. το $w_0 = \frac{1}{(x_0-x_1)*(x_0-x_2)*(x_0-x_3)} = \frac{1}{(0-1)(0-2)(0-3)} = -\frac{1}{6}$ και το $f(x_0) = -2$
1	2	$2 = (1)(-1)(-2)$, $w_1 = \frac{1}{2}$ και το $f(x_1) = 2$
2	36	$-2 = (2)(1)(-1)$, $w_2 = -\frac{1}{2}$ και το $f(x_2) = 36$
3	130	$6 = (3)(2)(1)$, $w_3 = \frac{1}{6}$ και το $f(x_3) = 130$

$$\text{Ο βαθμός του πολυωνύμου θα είναι } n = 3, \text{ επομένως: } p^{(bar)}(x) = \frac{\sum_{i=0}^3 f(x_i) \frac{w_i}{x-x_i}}{\sum_{i=0}^3 \frac{w_i}{x-x_i}} = \frac{\frac{-2}{(-6)x} + \frac{2}{2(x-1)} - \frac{36}{2(x-2)} + \frac{130}{6(x-3)}}{\frac{1}{6x} + \frac{1}{2(x-1)} - \frac{1}{2(x-2)} + \frac{1}{6(x-3)}} =$$

$$\frac{\frac{1}{3x} + \frac{1}{x-1} - \frac{18}{x-2} + \frac{65}{3(x-3)}}{\frac{1}{6x} + \frac{1}{2(x-1)} - \frac{1}{2(x-2)} + \frac{1}{6(x-3)}}$$

Παρατήρηση: Σε όλες τις εκδοχές των ερωτήσεων, η απάντηση θα έπρεπε να δοθεί στη μορφή που ζητούσε η εκφώνηση (εδώ βαρυκεντρική). Θα ήταν λάθος αν την τροποποιούσαμε σε άλλη μορφή (π.χ. αν γραφόταν ως δυναμομορφή ή μορφή Newton ή Lagrange).

Άσκηση 7

Ένας τρόπος να βρούμε το πολυώνυμο παρεμβολής σε **δυναμομορφή** για τα ζεύγη κόμβος - μέτρηση $\{(\xi_i, \psi_i), i = 1, \dots, n+1\}$ είναι να λύσουμε το γραμμικό σύστημα $Ag = g$ όπου το διάνυσμα g θα περιέχει τους συντελεστές της δυναμομορφής, το $y = (\psi_1, \dots, \psi_{n+1})^T$ και στις θέσεις $i, j = 1, 2, \dots, n+1$ του A υπάρχουν τα στοιχεία ξ_i^{j-1} . Να αιτιολογήσετε συνοπτικά γιατί η μέθοδος αυτή αποφεύγεται στην πράξη.

Λύση

Το μητρώο A που προκύπτει είναι τύπου **Vandermonde**, το οποίο γνωρίζουμε ότι στις περισσότερες περιπτώσεις (π.χ. για τυχαία επιλογή κόμβων ή για ισαπέχοντες κόμβους) έχει πολύ μεγάλο δείκτη κατάστασης ακόμα και για μικρές τιμές του n . **Από τη θεωρία γνωρίζουμε ότι η επίλυση συστημάτων με μητρώα με πολύ μεγάλο δείκτη κατάστασης ακόμα και με τον πιο ευσταθή αλγόριθμο επίλυσης, οδηγεί σε απώλεια πολλών δεκαδικών ψηφίων και σε μεγάλο σφάλμα.** Επομένως, η χρήση της παραπάνω μεθόδου υπολογισμού του πολυώνυμου παρεμβολής αποφεύγεται λόγω της αναμενόμενης μεγάλης απόκλισης του διανύσματος που θα υπολογιστεί επιλύοντας αριθμητικά το εν λόγω γραμμικό σύστημα από τις θεωρητικές τιμές των συντελεστών της δυναμομορφής του πολυωνύμου. Το μητρώο A που προκύπτει από το πρόβλημα είναι της μορφής:

$$\begin{pmatrix} 1 \\ 1 \\ 1 \\ \dots & \dots & \dots & \dots \\ 1 \end{pmatrix}$$

Άσκηση 8

Η εξίσωση: $f(x) = 0$ όπου $f(x) = 5x^3 - x - 2$ έχει τουλάχιστον μία λύση στο διάστημα $[0, 1]$. Βρείτε μία εκτίμηση γι' αυτήν που εγγυημένα δεν απέχει περισσότερο από 2^{-4} από την ακριβή λύση και συμπληρώστε τα πρώτα 4 δεκαδικά της ψηφία παρακάτω. Η απάντηση είναι: $x_{\text{approx}} = 0.8125$.

Λύση

Παρατηρούμε ότι $f(0) * f(1) \leq 0$ και η συνάρτηση είναι πολυώνυμο άρα συνεχής, επομένως από Bolzano υπάρχει τουλάχιστον μία λύση στο διάστημα $(0, 1)$. Εφόσον μας ενδιαφέρει να εξασφαλίσουμε ότι η προσέγγιση στη λύση δεν θα απέχει περισσότερο από μία δοθείσα τιμή (στην περίπτωσή μας 2^{-4}) το απλούστερο που μπορούμε να κάνουμε είναι να χρησιμοποιήσουμε τη **Μέθοδο Διχοτόμησης** (οτιδήποτε άλλο περιπλέκει την εξασφάλιση του μέγιστου σφάλματος). Πριν ξεκινήσουμε, παρατηρούμε ότι αφού υπάρχει τουλάχιστον μία ρίζα στο $(0, 1)$ αν θεωρήσουμε ως προσέγγιση το μέσο $x_1 = 1/2$, αυτή δεν θα απέχει περισσότερο από 2^{-1} από την ακριβή λύση, διότι $|1/2 - 0.8125| = 0.3125 < 0.5$. Εφόσον $f(1/2) * f(1) < 0$ τότε υπάρχει τουλάχιστον μία λύση στο $[1/2, 1]$ που δεν απέχει περισσότερο από 2^{-2} από το μέσο του υποδιαστήματος, δηλ. από το $x_2 = 3/4$, διότι $|3/4 - 0.8125| = 0.0625 = 2^{-4} < 2^{-2}$. Επίσης $f(3/4) * f(1) < 0$ επομένως $x_3 = 7/8$ που δεν απέχει περισσότερο από 2^{-3} από τη λύση, διότι η διαφορά $|7/8 - 0.8125| = 0.0625 = 2^{-4}$ και τέλος $f(3/4) * f(\frac{7}{8}) < 0$, επομένως η τιμή $x_4 = \frac{3/4 + 7/8}{2} = 0.8125$. Αν και ΔΕΝ χρειάζεται στην απάντηση, σημειώνουμε ότι $f(x_4) = -0.1306$ και ότι το απόλυτο σφάλμα είναι $|x_4 - x^*| \approx 0.0144 < 2^{-4}$. Ο ακριβής αριθμός βημάτων στη διχοτόμηση είναι:

$$k \geq \lceil \log_2((b - a) * \varepsilon^{-1}) \rceil = \lceil \log_2((1 - 0) * 2^4) \rceil = \lceil \log_2(2^4) \rceil = 4 \text{ βήματα (διχοτομήσεις)}.$$

Σημείωση: Ο λόγος που δεν χρησιμοποιήσαμε τη μέθοδο Newton – Raphson για τον υπολογισμό της ρίζας της συνάρτησης είναι ότι δεν αναφέρει η εκφώνηση την τετραγωνική σύγκλιση.

Άσκηση 9

Να αναφέρετε 2 βασικά **πλεονεκτήματα** και 2 βασικά **μειονεκτήματα** της μεθόδου τέμνουσας (secant) σε σύγκριση με τη μέθοδο διχοτόμησης για την επίλυση γραμμικών εξισώσεων.

Λύση

Πλεονεκτήματα μεθόδου τέμνουσας συγκριτικά με μέθοδο διχοτόμησης:

- 1) Όταν συγκλίνει, η σύγκλιση είναι **υπεργραμμική** ενώ της μεθόδου διχοτόμησης είναι **γραμμική**.
- 2) Γενικεύεται για την επίλυση συστημάτων μη - γραμμικών εξισώσεων ενώ η διχοτόμηση όχι.

Μειονεκτήματα μεθόδου τέμνουσας συγκριτικά με μέθοδο διχοτόμησης:

- 1) Σε κάθε βήμα χρησιμοποιείται πολύ λίγη πληροφορία για τη συνάρτηση (πρόσημα στα άκρα των διαστημάτων που εγκλείσουν τη λύση).
- 2) Η μέθοδος διχοτόμησης εξασφαλίζει σύγκλιση προς τη λύση εφόσον έχουμε εκκινήσει από διάστημα στα άκρα του οποίου υπάρχει εναλλαγή προσήμου, ενώ στην μέθοδο τέμνουσας τα σημεία εκκίνησης πρέπει να είναι κοντά σε ρίζα.

Είναι πολύ πιο εύκολη η επιλογή ενός διαστήματος (όπου προφανώς ικανοποιείται το θεώρημα Bolzano) από ότι η επιλογή μεμονωμένων σημείων (δύο στην περίπτωση της τέμνουσας) που πρέπει να βρίσκονται κοντά σε ρίζα.

Άσκηση 10

Αποδεικνύεται εύκολα ότι οι ρίζες του πολυωνύμου $5 * x^3 - x - 2$ είναι οι τιμές που επιστρέφει η κλήση $eig(A)$ όπου $A = \begin{pmatrix} 0 & 0 & 0.4 \\ 1 & 0 & 0.2 \\ 0 & 1 & 0 \end{pmatrix}$. Επίσης, είναι γνωστό ότι αυτό το πολυώνυμο έχει 1 πραγματική και 2 μιγαδικές ρίζες. Να χρησιμοποιήσετε αυτά τα στοιχεία καθώς και το αποτέλεσμα που λάβατε στην προηγούμενη Άσκηση 8, για να υπολογίσετε μία προσέγγιση του πρώτου δεκαδικού της ρίζας, χρησιμοποιώντας αποκλειστικά πολλαπλασιασμούς του A με διάνυσμα, εσωτερικά γινόμενα και διαιρέσεις. Εξηγήστε τι κάνατε.

Λύση

Το μητρώο αυτό ονομάζεται **συνοδευτικό μητρώο** του πολυωνύμου και οι ιδιοτιμές του είναι οι ρίζες (του πολυωνύμου).

Για κάθε πολυώνυμο βαθμού n, υπάρχει μητρώο $A \in C^{n \times n}$ με χαρακτηριστικό πολυώνυμο ίδιο με το p που είναι της

μορφής $p(\lambda) = a_n * \lambda^n + a_{n-1} * \lambda^{n-1} + \dots + a_1 * \lambda^1 + a_0$. Το μητρώο αυτό είναι το $A = \begin{bmatrix} 0 & 0 & \dots & 0 & -\alpha_0 \\ 1 & 0 & \dots & 0 & -\alpha_1 \\ 0 & 1 & \dots & 0 & -\alpha_2 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 & -\alpha_{n-1} \end{bmatrix}$ και

ονομάζεται **συνοδευτικό μητρώο του πολυωνύμου p**. Στην προκειμένη περίπτωση το $p(\lambda) = \lambda^2 - 0.2 * \lambda^1 - 0.4$. Αυτό είναι ένα μονικό πολυώνυμο, αφού ως γνωστό, το χαρακτηριστικό πολυώνυμο ενός μητρώου είναι πάντα μονικό¹⁰ και έχει προκύψει από το αρχικό δοθέν πολυώνυμο $5 * x^3 - x - 2$, μετατρέποντάς το σε μονικό (δηλαδή διαιρώντας κάθε όρο του με το «5», προκειμένου ο συντελεστής του μεγιστοβάθμιου όρου να γίνει μονάδα). Ισχύει ότι ο υπολογισμός των ριζών του πολυωνύμου βαθμού n ≡ υπολογισμός ιδιοτιμών συνοδευτικού μητρώου. Άρα για τη συνάρτηση, η πραγματική ιδιοτιμή θα είναι η πραγματική ρίζα και αξίζει να σημειωθεί επίσης ότι αυτή θα είναι και η μέγιστη σε απόλυτη τιμή ρίζα (θεώρημα Perron - Frobenius). Επομένως, χρησιμοποιούμε τη μέθοδος δύναμης εκκινώντας, π.χ. από το διάνυσμα $e = (1, 1, 1)^T$. Εκτελώντας ένα βήμα της μεθόδου (χωρίς κανονικοποίηση, για οικονομία) και υπολογίζοντας το κλάσμα Rayleigh, έχουμε ότι $z = Ae$, $z^T Az / z^T z = 0.8$. Πιο συγκεκριμένα, το $z = x^{(1)} = A * y^{(0)} = \begin{pmatrix} 0 & 0 & 0.4 \\ 1 & 0 & 0.2 \\ 0 & 1 & 0 \end{pmatrix} * \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0.4 \\ 1.2 \\ 1 \end{pmatrix}$. Στη

συνέχεια στο κλάσμα Rayleigh, έχουμε ότι $z^T Az = (0.4 \quad 1.2 \quad 1) * \begin{pmatrix} 0 & 0 & 0.4 \\ 1 & 0 & 0.2 \\ 0 & 1 & 0 \end{pmatrix} * \begin{pmatrix} 0.4 \\ 1.2 \\ 1 \end{pmatrix} = 2.08$, ενώ το εσωτερικό γινόμενο $z^T z = (0.4 \quad 1.2 \quad 1) * \begin{pmatrix} 0.4 \\ 1.2 \\ 1 \end{pmatrix} = 2.6$. Επομένως, το πηλίκο Rayleigh $= \frac{z^T Az}{z^T z} = \frac{2.08}{2.6} = 0.8$ και φαίνεται να υπάρχει σύμπτωση

με το πρώτο ψηφίο της προσέγγισης που λάβαμε με διχοτόμηση, επομένως σταματάμε. Το κλάσμα κάθε ιδιοδιανύσματος είναι η ιδιοτιμή του. Στην προκειμένη περίπτωση χρησιμοποιήσαμε ως αρχικό ιδιοδιάνυσμα το $e = (1, 1, 1)^T$

¹⁰ Με τον όρο αυτό εννοούμε ότι ο συντελεστής του μεγιστοβάθμιου όρου του πολυωνύμου είναι «1».

Άσκηση 11

α) Χρησιμοποιώντας τη μέθοδο του τραπεζίου, υπολογίστε μια προσέγγιση για το ολοκλήρωμα $\int_0^1 f(x)dx$ για τη συνάρτηση $f(x) = 5 * x^3 - x - 2$.

β) Επιλέξτε μία από τις παρακάτω ως προτιμότερη από πλευράς ακρίβειας και κόστους για να βελτιωθεί σημαντικά η παραπάνω προσέγγιση (χωρίς εξήγηση).

σύνθετος κανόνας Simpson

σύνθετη μεθ. τραπεζίου

κανόνας Simpson

Λύση

a) $I(f) = \frac{h}{2} (f(0) + f(1)) = 0$.

b) Το σφάλμα στον κανόνα Simpson είναι φραγμένο από έναν παράγοντα που πολλαπλασιάζεται με την τιμή της 4^{ης} παραγώγου της συνάρτησης σε κάποιο σημείο εντός του διαστήματος. Δεδομένου ότι η συνάρτηση προς ολοκλήρωση είναι κυβικό πολυώνυμο, η 4^η παράγωγος μηδενίζεται, η αριθμητική ολοκλήρωση είναι ακριβής.

Άσκηση 12

Αιτιολογήστε κατά πόσον θα συγκλίνει μία επαναληπτική μέθοδος που χρησιμοποιεί τον αναδρομικό τύπο $x_{k+1} = Tx_k + c$, όπου $T = \begin{pmatrix} 0.2 & 0 & 0.7 \\ 0 & -0.3 & 0.6 \\ 0.5 & 0 & 0.1 \end{pmatrix}$, και x_0 και c είναι κάποια διαθέντα διανύσματα. Σε περίπτωση που συγκλίνει, βρείτε επίσης σε ποιο διάνυσμα, συναρτήσει των T, c .

Λύση

Ικανή και αναγκαία συνθήκη είναι το μητρώο επανάληψης T να έχει φασματική ακτίνα μικρότερη του 1. Ικανή συνθήκη είναι κάποια νόρμα του T να είναι μικρότερη του 1. Το μητρώο έχει διάσταση 3×3 επομένως το χαρακτηριστικό πολυώνυμο έχει βαθμό 3 και οπωσδήποτε θέλουμε να αποφύγουμε τον υπολογισμό των ριζών του εκτός αν συντρέχουν προϋποθέσεις για απλοποιήσεις που οδηγούν σε εύκολες παραγοντοποιήσεις (δεν συμβαίνει εδώ). Στο συγκεκριμένο πρόβλημα, παρατηρούμε ότι η νόρμα μεγίστου (μέγιστο των αθροισμάτων των απολύτων τιμών των στοιχείων κάθε γραμμής) είναι $\|T\|_\infty = 0.9 < 1$, επομένως η μέθοδος συγκλίνει. Δεν μας πειράζει που $\|T\|_1 = 1.4 > 1$. Εφόσον συγκλίνει, και το όριο της ακολουθίας των x_k είναι \hat{x} τότε θα ισχύει ότι: $\lim_{k \rightarrow \infty} x_{k+1} = \lim_{k \rightarrow \infty} Tx_k + c$, $\hat{x} = T\hat{x} + c$, επομένως $\hat{x} = (I - T)^{-1}c$.

Σημείωση: Ιδιότητες όπως ΣΘΟ, διαγώνια κυριαρχία κ.λ.π. αφορούσαν το μητρώο από το οποίο προήλθε το T και όχι το μητρώο επανάληψης.

Άσκηση 13

Για κάθε εντολή MATLAB παρακάτω, να γράψετε τι θα τυπωθεί στην οθόνη μετά την εκτέλεσή της

a) 1/0 - 1/0,

b) (1 + eps + eps^2 == 1 + eps^2 + eps)

c) 2 * realmax - realmax == realmax

d) realmin/2^55

e) NaN - NaN

f) a = (10 + eps + eps^2 == 10 + eps^2 + eps); b = (1/0) * (2/Inf); c = (eps * eps^2 == eps^3); d = (NaN==NaN)

Λύση

- a) Είναι ισοδύναμη με Inf - Inf, άρα το αποτέλεσμα που επιστρέφει είναι NaN.
- b) Το αριστερό μέλος στην πρώτη άθροιση επιστρέφει **1 + eps**, ενώ η δεύτερη άθροιση με το eps^2 δεν αλλάζει την τιμή (δεν προκαλεί απορρόφηση) καθώς το eps^2 είναι μικρότερο του $\text{eps}/2 = 2^{-53}$. Για τους ίδιους λόγους το δεξιό μέλος έχει το ίδιο αποτέλεσμα (διότι το $1 + \text{eps}^2 + \text{eps} = 1 + \text{eps} = 1^+$), επομένως το αποτέλεσμα της σύγκρισης είναι αληθές και η τιμή που επιστρέφεται είναι «1».
- c) Το αριστερό μέλος επιστρέφει Inf, διότι οτιδήποτε πάνω από `realmax` είναι Inf, άρα η σύγκριση επιστρέφει λάθος, άρα «0».
- d) Ο ελάχιστος κανονικοποιημένος αριθμός αν διαιρεθεί με το 2 γίνεται **υποκανονικοποιημένος** και περιέχει στην ουρά την τιμή :10....0. Το μήκος της ουράς είναι 52, επομένως μετά από 55 διαιρέσεις με το 2, η ουρά θα περιέχει 0 και η τιμή του αριθμού θα είναι μηδέν.
- e) Επιστρέφει NaN, διότι δεν μπορούμε να κάνουμε πράξεις με το NaN.
- f) Οι απαντήσεις είναι **a=1, b=NaN, c=1, d=0**. Το μόνο ίσως απρόσμενο είναι το τελευταίο που φαίνεται να υποδηλώνει ότι ένας αριθμός δεν είναι ίσος με τον εαυτό του. Όμως το **NaN** είναι **σύμβολο και όχι αριθμός** και μάλιστα δεν έχει μοναδική δυαδική αναπαράσταση (αρκεί ο εκθέτης να περιέχει αποκλειστικά 1 και η ουρά οποιοδήποτε συνδυασμό με τουλάχιστον ένα bit ίσο με 1 για να επιστραφεί το NaN).

Άσκηση 14

Έστω ότι ο δείκτης κατάστασης υπολογισμού της λύσης ενός συστήματος $A * x = b$ ως προς την Ευκλείδεια νόρμα είναι 101×10^6 , όπου ενώ γνωρίζετε ότι η υπολογισμένη λύση \hat{x} ικανοποιεί τη $\|b - A * \hat{x}\|_2 = 10^{-12}$ καθώς επίσης ότι $\|\hat{x}\|_2 = 10$, $\|A\|_2 = 10$, $\|b\|_2 = 1$. Να υπολογίσετε ένα άνω φράγμα για τη νόρμα του εμπρός σχετικού σφάλματος $\frac{\|x - \hat{x}\|_2}{\|x\|_2}$.

Λύση

Πρώτα υπολογίζουμε το πίσω σχετικό σφάλμα ως προς την Ευκλείδεια νόρμα που είναι $\eta = \frac{\|b - A * \hat{x}\|}{\|A\| * \|\hat{x}\| + \|b\|} = \frac{10^{-12}}{10 * 10 + 1} = \frac{10^{-12}}{101}$. Στη συνέχεια το **εμπρός σχετικό σφάλμα**: $\frac{\|x - \hat{x}\|}{\|x\|} = \frac{\|\hat{x} - x\|}{\|x\|} \leq \frac{2 * \kappa(A) * \eta}{1 - \kappa(A) * \eta} = 2 * \frac{10^{-6}}{1 - 10^{-6}}$.

Άσκηση 15

Γενικά, αν ένας αλγόριθμος είναι πίσω σταθερός (ευσταθής), τότε μικρές αλλαγές στα στοιχεία εισόδου του αλγορίθμου οδηγούν κατ' ανάγκη σε μικρές αλλαγές στο υπολογισμένο αποτέλεσμα.

Λύση

Λάθος, διότι η **πίσω ευστάθεια εξασφαλίζει μόνο** 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 `cond(f; x)`. Μεγάλη πίσω ευστάθεια σημαίνει μικρό πίσω σφάλμα. Το υπολογισμένο αποτέλεσμα είναι η έξοδος και επομένως έχει εμπρός σφάλμα. Ισχύει ο τύπος: **εμπρός σφάλμα ≤ δείκτης κατάστασης προβλήματος (δηλαδή το `cond(f; x)`) * πίσω σφάλμα** και επειδή το **πίσω σφάλμα ≤ δείκτης κατάστασης αλγόριθμου (δηλαδή το `cond(fprog)`) * u**, έχουμε ότι το **εμπρός σφάλμα ≤ cond(f; x) * cond(f_{prog}) * u**. Επομένως, **δεν** αρκεί να είναι μικρός μόνο ο δείκτης `cond(fprog)` λόγω της πίσω ευστάθειας, αλλά θα πρέπει να είναι μικρός και ο `cond(f; x)`.

Άσκηση 16

Από τις παρακάτω επιλογές, ποια είναι η **μεγαλύτερη απόσταση διαδοχικών κόμβων** που μπορεί να χρησιμοποιηθεί σε ένα πίνακα τιμών για την $f(x) = \cos(x)$, ώστε με **γραμμική παρεμβολή** μεταξύ δύο διαδοχικών κόμβων και αντίστοιχων τιμών, να προσεγγίζεται η τιμή της $f(x)$ στο αντίστοιχο διάστημα με μέγιστο απόλυτο σφάλμα που δεν υπερβαίνει το 10^{-8} .

 10^{-1} 10^{-2} 10^{-3} 10^{-4} **Λύση**

Ως γνωστό, ο τύπος που δίνει το μέγιστο άνω φράγμα του σφάλματος για την τμηματική γραμμική παρεμβολή είναι ο εξής: $\max |f(x) - P_1^H f(x)| \leq \max_{x \in \Omega} \frac{|f''(x)|}{8} H^2$. Με δεδομένο ότι $f(x) = \cos(x)$, έχουμε ότι $f'(x) = -\sin(x)$ και $f''(x) = -\cos(x)$. Ζητάμε το H , που είναι η απόσταση μεταξύ **διαδοχικών κόμβων**, οπότε έχουμε ότι: $\max_{x \in \Omega} \frac{|f''(x)|}{8} H^2 \leq 10^{-8} \Rightarrow \max_{x \in \Omega} \frac{|-\cos(x)|}{8} H^2 \leq 10^{-8} \Rightarrow \max_{x \in \Omega} \frac{1}{8} H^2 \leq 10^{-8} \Rightarrow H^2 \leq 8 * 10^{-8} \Rightarrow H \leq \sqrt{8 * 10^{-8}} = \sqrt{8} * \sqrt{10^{-8}} = \sqrt{8} * 10^{-4} = 2.8 * 10^{-4}$.

Άσκηση 17 (Σεπτέμβριος 2017)

Δίνεται **συμμετρικό μητρώο A**. Ποιος από τους παρακάτω τρόπους θεωρείτε ότι είναι ο καλύτερος για την επίλυση γραμμικού συστήματος;

- $[Q, R] = qr(A, 0); x = R \setminus (Q' * b)$
- $[R] = chol(A); x = R \setminus (R' \setminus b)$
- $[L, U] = lu(A); x = U \setminus (L \setminus b)$
- $[x] = pcg(A, b)$

Λύση

Η μέθοδος $x = pcg(A, b)$ που ερμηνεύεται ως **Preconditioned conjugate gradients method**, προσπαθεί να επιλύσει το γραμμικό σύστημα $A * x = b$ και το μητρώο $A \in \mathbb{R}^{n \times n}$ πρέπει να είναι **συμμετρικό και θετικά ορισμένο** και επίσης πρέπει επίσης να είναι **μεγάλο και αραιό**. Η σωστή απάντηση είναι η **QR παραγοντοποίηση**, διότι το μητρώο A **δεν** είναι ΣΘΟ παρά μόνο συμμετρικό, οπότε δεν μπορεί να εφαρμοστεί ούτε η παραγοντοποίηση Cholesky, αλλά ούτε και η μέθοδος CG των συζυγών κλίσεων και επίσης το μητρώο **δεν** έχει A.D.K. κατά στήλες για να μπορεί να εφαρμοστεί η LU. Αν υπήρχε ως απάντηση $[P, L, U] = lu(A); x = U \setminus (L \setminus b)$ τότε θα επιλέγαμε αυτή, διότι η PLU παραγοντοποίηση υπάρχει πάντα και έχει μικρότερο κόστος από την LU. Εναλλακτικά θα μπορούσαμε να επιλέξουμε και την **Idl παραγοντοποίηση**, αν δινόταν από την εκφώνηση.

Άσκηση 18 (Σεπτέμβριος 2017)

Ένα πρόγραμμα περιέχει την εντολή $y = (\sqrt{x}) - \sqrt{x-h})/h$ όπου οι μεταβλητές περιέχουν αριθμούς κινητής υποδιαστολής και καμία δεν είναι μηδέν. Εξηγήστε τι πρόβλημα μπορεί να προκύψει στον υπολογισμό και προτείνετε βελτίωσή του.

Λύση

Το πρόβλημα που μπορεί να προκύψει είναι αν το $x >> h$ τότε η διαφορά στον αριθμητή του κλάσματος $\text{sqrt}(x) - \text{sqrt}(x - h)$ θα δώσει ένα αποτέλεσμα πολύ κοντά στο «0», δηλαδή αφαιρούμε δύο περίπου ίσους α.κ.υ. μεταξύ τους και το αποτέλεσμα διαιρείται στη συνέχεια με ένα πολύ μικρό αριθμό που είναι το h , με αποτέλεσμα την κλιμάκωση των όποιων σφαλμάτων υπάρχουν από την αφαίρεση στον αριθμητή. Επομένως μπορεί να προκύψει το φαινόμενο της **καταστροφικής απαλοιφής**. Για να εξαλείψουμε το φαινόμενο αυτό θα πρέπει να πολλαπλασιάσουμε αριθμητή και παρανομάστε το συζυγή του αριθμητή, δηλαδή με την ποσότητα $\text{sqrt}(x) + \text{sqrt}(x - h)$. Με τον τρόπο αυτό θα απαλείψουμε τις πράξεις από τον αριθμητή και το όποιο σφάλμα προκύπτει από αυτές.

Άσκηση 19 (Θέμα Σεπτεμβρίου 2017)

Νέος επεξεργαστής επιτρέπει τον υπολογισμό των κλασικών αριθμητικών πράξεων σε nsec, αλλά η συνάρτηση της τετραγωνικής ρίζας δεν έχει ακόμη βελτιστοποιηθεί και απαιτεί χρόνους τάξης msec. Με αυτές τις συνθήκες χρησιμοποιήστε κατάλληλη μέθοδο για τη γρήγορη προσέγγιση του $\sqrt{17}$. Δείξτε το συλλογισμό και τις πράξεις που χρειάζονται και υπολογίστε την προσέγγιση με ακρίβεια τριών δεκαδικών.

Λύση

Το γεγονός ότι χρησιμοποιούμε μεγαλύτερο χρόνο (msec αντί για nsec) σημαίνει ότι θα χρησιμοποιήσουμε τη **μέθοδο της διχοτόμησης** για τη γρηγορότερη προσέγγιση του $\sqrt{17}$, η οποία ως γνωστό έχει **γραμμική σύγκλιση και επομένως συγκλίνει πιο αργά** (δηλαδή σε μεγαλύτερο χρόνο) προς μια ρίζα. Επειδή λοιπόν η συνάρτηση της τετραγωνικής ρίζας απαιτεί πιο αργούς χρόνους από αυτούς που υποστηρίζει ο επεξεργαστής, θα χρησιμοποιήσουμε τη μέθοδο της διχοτόμησης, η οποία ως γνωστό έχει γραμμική σύγκλιση. Η $\sqrt{17}$ αποτελεί λύση της εξίσωσης $f(x) = x^2 - 17 = 0$ και ένα διάστημα στο οποίο ικανοποιείται το κριτήριο του Bolzano είναι το $[-1, 5]$, όπου $f(-1) * f(5) < 0$ και για να βρούμε τον αριθμό των επαναλήψεων που απαιτούνται για να πετύχουμε ακρίβεια $\varepsilon = 10^{-3}$ (τριών δεκαδικών ψηφίων) χρησιμοποιούμε τον τύπο: $n = \lceil \log_2((b - a) * \varepsilon^{-1}) \rceil = \lceil \log_2((5 - (-1)) * 10^3) \rceil = \lceil \frac{\log_{10}(6 * 10^3)}{\log_{10} 2} \rceil = \lceil 12.55076 \rceil = 13$ επαναλήψεις.

Στη συνέχεια για να υπολογίσουμε μια προσέγγιση της ρίζας με ακρίβεια 3 δ.ψ. διχοτομούμε το διάστημα $[-1, 5]$ και εφαρμόζουμε θεώρημα Bolzano στο πάνω ή στο κάτω μισό υποδιάστημα. Πιο συγκεκριμένα το γινόμενο $f(3) * f(5) < 0$, επομένως μια πρώτη προσέγγιση της ρίζας είναι το 4. Στη συνέχεια εφαρμόζουμε εκ' νέου διχοτόμηση στο διάστημα $[3, 5]$ και το γινόμενο $f(4) * f(5) < 0$, οπότε η νέα προσέγγιση της ρίζας είναι το 4.5 και ισχύει ότι $f(4) * f(4.5) < 0$, επομένως η καινούργια προσέγγιση της ρίζας είναι η 4.25 και ισχύει ότι $f(4) * f(4.25) < 0$, οπότε με μια νέα διχοτόμηση στο $[4, 4.25]$ παίρνουμε τη ζητούμενη προσέγγιση 4.125 με τρία δεκαδικά ψηφία.

Παρατήρηση: Άν η εκφώνηση ανέφερε ότι ένας νέος επεξεργαστής επιτρέπει τον υπολογισμό των κλασικών αριθμητικών πράξεων σε χρόνο nsec τότε θα χρησιμοποιούσαμε μέθοδο **Newton – Raphson**. Αν χρησιμοποιούσαμε το διάστημα $[4, 5]$ στο οποίο επίσης ισχύει το θεώρημα του Bolzano και επειδή $f(4) * f(5) < 0$ και επίσης επειδή $f'(x) = 2x > 0, \forall x \in [4, 5]$ ισχύει ότι υπάρχει **μόνο μια ρίζα** στο συγκεκριμένο διάστημα. Θα πάρουμε το διάστημα $[4, 4.5]$, διότι $f(4) * f(4.5) < 0$ και η νέα προσέγγιση είναι το 4.25 και θα πάρουμε το σημείο αυτό ως αρχική εκτίμηση για τη μέθοδο Newton – Raphson, δηλαδή $x_0 = 4.25$. Σε μια τέτοια περίπτωση θα είχαμε ότι για $x_0 = 4.25$ και $x_1 = x_0 - \frac{f(x_0)}{f'(x_0)} =$

4.125 και το απόλυτο σφάλμα $e_1 = |4.125 - 4.25| = 0.125$ και αυτό δεν είναι μικρότερο από το 10^{-3} . Υπολογίζουμε τη δεύτερη εκτίμηση της ρίζας που είναι το $x_2 = x_1 - \frac{f(x_1)}{f'(x_1)} = 4.123$. Άρα με το δεύτερο βήμα της Newton – Raphson βρήκαμε την ακριβή ρίζα, αφού το σφάλμα είναι μηδέν, διότι το $e_2 = |4.123 - 4.125| = |0.002|$ δεν είναι μικρότερο από το 0.001. Επομένως χρειάζεται άλλη μια επανάληψη της Newton – Raphson.

Άσκηση 20 (Θέμα Σεπτεμβρίου 2017)

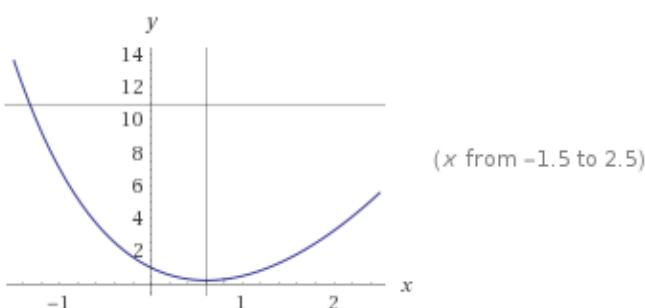
Έστω ότι $f(x) = -x * 2^{-x+1} + 2^x + x^2$. Με λίγες πράξεις επαληθεύεται ότι $f(0.6412) \approx 10^{-10}$ αλλά έστω ότι αυτό δεν είναι γνωστό. Το μόνο που ξέρουμε είναι ότι υπάρχει τουλάχιστον μία ρίζα στο διάστημα [0 1]. Αν θέλουμε να την προσεγγίσουμε με αριθμητική μέθοδο, αφού εξετάστε προσεκτικά τη συνάρτηση, ποια από τα παρακάτω είναι αληθή;

- Η μέθοδος διχοτόμησης μπορεί να εφαρμοστεί και θα έχει γραμμική σύγκλιση.
- Η μέθοδος Newton μπορεί να εφαρμοστεί και θα έχει γραμμική σύγκλιση.
- Η μέθοδος τέμνουσας μπορεί να εφαρμοστεί.
- Η μέθοδος διχοτόμησης μπορεί να εφαρμοστεί και θα έχει τετραγωνική σύγκλιση.
- Η μέθοδος Newton δεν μπορεί να εφαρμοστεί.

Λύση

Σωστή απάντηση είναι η (a). Η μέθοδος διχοτόμησης μπορεί να εφαρμοστεί, αφού όπως δίνεται από την εκφώνηση υπάρχει τουλάχιστον μία ρίζα στο διάστημα [0 1] και εξ' ορισμού θα έχει γραμμική σύγκλιση και όχι τετραγωνική. Η μέθοδος τέμνουσας δεν μπορεί να εφαρμοστεί, διότι απαιτείται η γνώση των δυο προηγούμενων προσεγγίσεων μιας ρίζας, που δεν δίνεται από την εκφώνηση της άσκησης.

Παρατήρηση: Όσον αφορά τη μέθοδο Newton (εννοείται Newton – Raphson) ισχύει ότι η ακολουθία $\{x_0, x_1, x_2, \dots, x_k\}$ ονομάζεται **ακολουθία Newton και για να μην υπάρχει αστοχία θα πρέπει να ισχύει ότι η $f'(x_0) \neq 0$ και ομοίως για όλα τα σημεία της ακολουθίας θα πρέπει να μην μηδενίζεται η παράγωγος**. Η παράγωγος της συνάρτησης $f(x)$ είναι $f'(x) = 2^{-x} * (1 - 2 * x) + x^2$ και παρατηρούμε ότι η $f'(x_0) = 0.6412 * 2^{(-0.6412+1)} + 2^{(-0.6412)} + 0.6412^2 = 0.23 \neq 0$, άρα η μέθοδος Newton (εννοείται Newton – Raphson) δεν αστοχεί στο πρώτο βήμα. **Γενικότερα ισχύει -όπως αναφέρθηκε- ότι ο τύπος Newton (εννοείται Newton – Raphson) μπορεί να αστοχήσει σε κάποια επανάληψη αν $f'(x_k) = 0$.** Αν κάποια πρόταση ανέφερε ότι μπορεί να εφαρμοστεί η μέθοδος Newton η οποία έχει τετραγωνική σύγκλιση θα ήταν σωστή, επειδή από πριν αποδείξαμε ότι η παράγωγος της συνάρτησης $f(x)$ είναι διάφορη του «0» για το συγκεκριμένο x_0 .



Άσκηση 21 (Σεπτέμβριος 2017)

α) Αν $A \in \mathbb{R}^{1000 \times 10}$ και $A^T A$ θετικά ορισμένο, επιλέξτε την πιο κατάλληλη σειρά εντολών για τον υπολογισμό του $x \in \mathbb{R}^{10}$ που ελαχιστοποιεί την απόσταση $\|b - Ax\|_2$.

- $[L, U] = lu(A); y = U \setminus (L \setminus b)$
- $x = (A' * A) \setminus (A' * b)$
- $x = \text{inv}(A) * b;$
- $x = (A * A') \setminus (A' * b)$
- $x = A \setminus b$

β) Για το παραπάνω x , το αναμενόμενο πλήθος πράξεων αριθμών κινητής υποδιαστολής για τον υπολογισμό του είναι πιο κοντά στο 10^2 10^4 10^5 10^6

Λύση

α) Χρησιμοποιούμε μέθοδο **ελαχίστων τετραγώνων** προκειμένου να ελαχιστοποιήσουμε την απόσταση $\|b - Ax\|_2$. Στην περίπτωση αυτή ως γνωστό χρησιμοποιούμε τη μέθοδο των **κανονικών εξισώσεων** για να επιλύσουμε το γραμμικό σύστημα $A^* x = b$ και αφού πολλαπλασιάσουμε και τα δύο μέλη της εξίσωσης με το A^T από αριστερά, στη συνέχεια επιλύουμε την εξίσωση ως προς x και βρίσκουμε ότι $x = (A^T * A)^{-1} * A^T * b = (A^T * A) \setminus (A^T * b) = (A' * A) \setminus (A' * b)$. Επομένως σωστή απάντηση είναι η δεύτερη.

β) Ο παραπάνω υπολογισμός περιέχει τον υπολογισμό του αντίστροφου μητρώου $(A^T * A)^{-1}$ και ως γνωστό για τον υπολογισμό **ενός αντίστροφου μητρώου A^{-1}** (εννοείται $A \in \mathbb{R}^{n \times n}$ όπου στην προκειμένη περίπτωση το μητρώο $A \in \mathbb{R}^{10 \times 10}$) απαιτούνται **$\Omega = 2n^3$ πράξεις** ή αλλιώς **$\Omega = 2 * 10^3$ πράξεις**. Επιπλέον απαιτούνται πράξεις και για τον υπολογισμό του γινομένου $A' * b$ που είναι $n * (2n-1) * 1 = 10 * (20-1) * 1 = 190$ και για τον υπολογισμό του γινομένου $(A^T * A)^{-1} * (A^T * b) = n * (2n-1) * 1 = 10 * (20-1) * 1 = 190$. Συνολικά εκτελούνται $2000 + 190 + 190 = 2380$ πράξεις. Αυτό το αποτέλεσμα ως μέγεθος **είναι πιο κοντά στο $10^2 = 100$ και όχι στο $10^4 = 10000$** , διότι η διαφορά $2380 - 100 = 2280$, ενώ η διαφορά $10000 - 2380 = 7620$. **Άρα η σωστή απάντηση στο β' ερώτημα είναι η πρώτη.**

Άσκηση 22 (Ιούνιος 2018)

α) Έστω το σύστημα $Ax = b$ όπου $A = \begin{pmatrix} 2 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & -1 & 2 \end{pmatrix}$ και $b = [-1, 0, 1]^T$

1. Σωστό ή Λάθος: Αν για το παραπάνω σύστημα κατασκευαστεί επαναληπτικό σχήμα $x(k+1) = T * x(k) + c$, όπου $T \in \mathbb{R}^{n \times n}$ και $x(k), c \in \mathbb{R}^n$ για το οποίο ισχύει ότι $\|T\|_\infty > 1$ τότε η ακολουθία $x^{(0)}, x^{(1)}, \dots$ αποκλείεται να συγκλίνει στο διάνυσμα x για το οποίο ισχύει ότι $x = Tx + c$.

2. Μαυρίστε το κουτάκι για όποιες από τις παρακάτω μεθόδους θα συγκλίνουν, αν εφαρμοστούν για την επίλυσή του και αν δεν συγκλίνει καμία, τότε μαυρίστε την εναλλακτική επιλογή

- Gauss – Seidel Jacobi CG Καμία από τις προτεινόμενες απαντήσεις

Λύση

1. Λάθος, διότι αν δεν ισχύει ότι $\|T\|_\infty < 1$, επειδή ακριβώς αυτή η συνθήκη είναι μόνο ικανή και όχι αναγκαία μπορεί να ισχύει κάποια από τις υπόλοιπες συνθήκες σύγκλισης του επαναληπτικού σχήματος, όπως για παράδειγμα αν το μητρώο A στο σύστημα $A * x = b$ έχει **A.Δ.Κ.** (είτε κατά γραμμές είτε κατά στήλες), τότε συγκλίνουν οι μέθοδοι **Jacobi** και **Gauss – Seidel**. **Επίσης αν** η πρώτη νόρμα του μητρώου **επανάληψης T** είναι μικρότερη από 1, δηλ. $\|T\| < 1$, τότε συγκλίνουν και πάλι η **Gauss – Seidel** και η **Jacobi**. **Μόνο αν δινόταν από την εκφώνηση** ότι το χαρακτηριστικό πολυώνυμο του μητρώου επανάληψης, δηλ. το $\rho(T) > 1$ (**φασματική ακτίνα του μητρώου επανάληψης T** που είναι η μεγαλύτερη κατ' απόλυτη τιμή ιδιοτιμή, είναι μεγαλύτερη από «1») **δεν θα σύγκλιναν και οι τρεις μέθοδοι**.

2. Το μητρώο A δεν έχει **A.Δ.Κ.**, είτε κατά γραμμές είτε κατά στήλες και επίσης η $\|T\|_\infty > 1$, όπως δίνεται από την εκφώνηση της άσκησης. **Επειδή όμως** το μητρώο των συντελεστών A έχει ήδη δοθεί σε μια μορφή που είναι **κατά πλοκάδες άνω τριγωνικό**, δηλαδή $\begin{pmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{pmatrix} = \begin{pmatrix} 2 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & -1 & 2 \end{pmatrix}$ όπου τα κύρια υπομητρώα A_{11} και A_{22} (**πλοκάδες**), είναι μη τετραγωνικά, το μητρώο A είναι **μη αναγωγήσιμο** και επίσης είναι και **διαγώνια κυρίαρχο με αυστηρή ΔΚ** για τουλάχιστον **μια γραμμή ή μια στήλη** (π.χ. 1^η γραμμή), **επομένως** συγκλίνουν τόσο η **Gauss – Seidel** όσο και η **Jacobi**. Η μέθοδος **CG** δεν συγκλίνει, επειδή έχει την ίδια προϋπόθεση εφαρμογής με την Cholesky, δηλ. το μητρώο στο οποίο εφαρμόζεται πρέπει να είναι **Σ.Θ.Ο.** και το συγκεκριμένο μητρώο A δεν είναι καν συμμετρικό.

Άσκηση 23 (Ιανουάριος 2018)

Η εξίσωση $f(x) = 0$ όπου $f(x) = 4x^3 - x - 2$ έχει τουλάχιστον μια λύση στο διάστημα $[0, 1]$.

1. Βρείτε μια εκτίμηση x_b για αυτήν που απέχει λιγότερο από 2^{-3} από την ακριβή λύση και συμπληρώστε τα πρώτα 3 δεκαδικά ψηφία της παρακάτω: $x_b =$
2. Τι τιμή προσέγγισης υπολογίζετε για τη ρίζα αν χρησιμοποιήσετε 2 βήματα Newton εκκινώντας από το $x_N^{(0)} = 1$; Η απάντηση είναι $x_N^{(2)} =$
3. Ποιο από τα δύο αποτελέσματα x_b και $x_N^{(2)}$ θεωρείτε ότι είναι καλύτερη εκτίμηση της ρίζας και γιατί;

Λύση

1. Θα χρησιμοποιήσουμε τη μέθοδο της διχοτόμησης για να βρούμε τη ρίζα στο διθέν διάστημα. Αρχικά διχοτομούμε το διάστημα $[0, 1]$ στη μέση, δηλαδή στο σημείο 0.5 και εφαρμόζουμε το θεώρημα Bolzano για να δούμε που ισχύει. Ισχύει ότι $f(1/2) * f(1) < 0$, επομένως η ρίζα είναι αρχικά το σημείο 0.75. **Η ακριβής ρίζα είναι το σημείο 0.89816** και η διαφορά $|x_k - x| = |0.89816 - 0.75| = 0.14816$ που δεν είναι μικρότερο από $2^{-3} = 1/8 = 0.125$. Επομένως, πρέπει να κάνουμε μια νέα διχοτόμηση στο διάστημα $[0.75, 1]$ και η νέα ρίζα είναι το μέσο του διαστήματος αυτού, δηλαδή το σημείο 0.875. Το νέο απόλυτο σφάλμα είναι η διαφορά $|x_k - x| = |0.89816 - 0.875| = 0.02316 < 2^{-3} = 1/8 = 0.125$. Επομένως, η εκτίμηση της ρίζας είναι $x_b = 0.875$. **Θα πρέπει να σημειωθεί ότι στη μέθοδο της διχοτόμησης το μήκος του εκάστοτε διαστήματος είναι το σφάλμα της μεθόδου, το οποίο το συγκρίνουμε με το σφάλμα της εκφώνησης και για αυτό είναι δύσκολο να βρεθεί η ακριβής εκτίμηση της ρίζας.**

2. Αν χρησιμοποιήσουμε 2 βήματα Newton εκκινώντας από το $x_N^{(0)} = 1$, θα έχουμε ότι $x_1 = x_0 - \frac{f(x_0)}{f'(x_0)} = 1 - \frac{1}{11} = \frac{10}{11}$. Υπολογίζουμε στη συνέχεια τη δεύτερη εκτίμηση της ρίζας που είναι το $x_2 = x_1 - \frac{f(x_1)}{f'(x_1)} = \frac{10}{11} - \frac{f(\frac{10}{11})}{f'(\frac{10}{11})} = \frac{10}{11} - \frac{\frac{128}{1331}}{\frac{1079}{121}} = 0,8983 = x_N^{(2)}$.

3. Επομένως, και με δεδομένο ότι η ακριβής ρίζα είναι το σημείο 0.89816, η καλύτερη εκτίμηση της ρίζας είναι αυτή που υπολογίστηκε με τη μέθοδο Newton – Raphson, διότι έχει μικρότερο σφάλμα. Πράγματι, το απόλυτο σφάλμα του υπολογισμού της συγκεκριμένης μεθόδου είναι: $|0.89816 - 0.8983| = 0.00014$. Αντίθετα, το σφάλμα υπολογισμού που προκύπτει από τη μέθοδο της διχοτόμησης είναι: $|0.89816 - 0.875| = 0.02316$.

Άσκηση 24 (Ιανουάριος 2018)

Έστω το μητρώο $A = \begin{pmatrix} * & 0 & 0 & 0 \\ * & * & * & * \\ * & * & * & 0 \\ 0 & * & 0 & 0 \end{pmatrix}$, όπου τα * δηλώνουν άγνωστες μη μηδενικές τιμές. Από τις επόμενες δύο επιλογές, επιλέξτε μόνο μία και απαντήστε στο χώρο που διατίθεται. Διαγράψτε την άλλη επιλογή.

1. Να βρείτε μητρώα μετάθεσης P και Q (ενδεχομένως ένα από αυτά να είναι ταυτοικό) τέτοια ώστε το PAQ να είναι

κάτω τριγωνικό. Δηλαδή $P = \begin{pmatrix} & & & \\ & & & \\ & & & \\ & & & \end{pmatrix}$ και $Q = \begin{pmatrix} & & & \\ & & & \\ & & & \\ & & & \end{pmatrix}$

2. Αν το μητρώο έχει ανατεθεί στη μεταβλητή A, να δώσετε κατάλληλες τιμές στα διανύσματα p και q έτσι ώστε το A(p, q) να είναι κάτω τριγωνικό. Δηλαδή $p = [\quad]$ και $q = [\quad]$.

Λύση

Η σωστή επιλογή είναι η **πρώτη**. Πράγματι, αφού το μητρώο $A = \begin{pmatrix} * & 0 & 0 & 0 \\ * & * & * & * \\ * & * & * & 0 \\ 0 & * & 0 & 0 \end{pmatrix}$, και θέλουμε το μητρώο PAQ να είναι

κάτω τριγωνικό, δηλαδή της μορφής $\begin{pmatrix} * & 0 & 0 & 0 \\ * & * & 0 & 0 \\ * & * & * & 0 \\ * & * & * & * \end{pmatrix}$, θα πρέπει να κάνουμε μια εναλλαγή 2^{ης} και 4^{ης} γραμμής του A,

θέτοντας το μητρώο εναλλαγής $P = [e_1, e_4, e_3, e_2] = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}$, με αποτέλεσμα το γινόμενο των μητρώων $P * A =$

$\begin{pmatrix} * & 0 & 0 & 0 \\ 0 & * & 0 & 0 \\ * & * & * & 0 \\ * & * & * & * \end{pmatrix}$, δηλαδή παρατηρούμε ότι τελικά το μητρώο $P * A$ έγινε κάτω τριγωνικό (το «0» που υπάρχει στη θέση 2, 2).

1 δεν μας ενοχλεί, διότι δεν επηρεάζει τη μορφή του μητρώου. Με άλλα λόγια σε ένα μητρώο άνω/κάτω τριγωνικό μπορούν να υπάρχουν μηδενικά στοιχεία κάτω/πάνω από την κύρια διαγώνιο αντίστοιχα. **Το μητρώο Q = I**, διότι δεν χρησιμοποιήθηκε κάποιο μητρώο για εναλλαγή στηλών.

Άσκηση 25 (Ιανουάριος 2018)

Έστω το μητρώο A που προκύπτει από τις εντολές $C = rand(1000); A = C + C'$. Να επιλέξτε την πιο κατάλληλη αλληλουχία εντολών για την επίλυση γραμμικού συστήματος με το μητρώο A και δεξιό μέλος κάποιο συμβατό στις διαστάσεις διάνυσμα b.

[x] = pcg(A, b);

[R] = chol(A' * A); x = R\(\backslash(R'\backslash(A'*b))

- $[R] = \text{chol}(A); x = R \backslash (R' \backslash b)$
- $[L, D] = \text{ldl}(A); x = L' \backslash (\text{\\} (D \backslash (L \backslash b)))$

Λύση

Υπενθύμιση: Η μέθοδος $x = \text{pcg}(A, b)$ που ερμηνεύεται ως **Preconditioned conjugate gradients method**, προσπαθεί να επιλύσει το γραμμικό σύστημα $A * x = b$ και το μητρώο $A \in \mathbb{R}^{n \times n}$ πρέπει να είναι **συμμετρικό** και **θετικά ορισμένο** και επίσης πρέπει επίσης να είναι **μεγάλο** και **αραιό**. Η μέθοδος $[L, D] = \text{ldl}(A)$ που ερμηνεύεται ως **Block LDL παραγοντοποίηση** για Ερμιτιανά αόριστα μητρώα. Υποθέτοντας ότι το A είναι ένα Ερμιτιανό μητρώο (δηλαδή $A = A'$), η **συγκεκριμένη παραγοντοποίηση αποθηκεύει** ένα **block διαγώνιο μητρώο D** και ένα **γινόμενο μητρώου μετάθεσης** και κάτω **τριγωνικού μητρώου στο L** , έτσι ώστε να ισχύει $A = L * D * L'$. Αυτή η σύνταξη **δεν** ισχύει για αραιά μητρώα.

Στην προκειμένη περίπτωση το συγκεκριμένο μητρώο που παράγεται από την εντολή `rand` είναι πυκνό, περιέχει τυχαίες τιμές και είναι διαστάσεων 1000×1000 . Άρα δεν μπορεί να εφαρμοστεί η παραγοντοποίηση Cholesky μέσω της **ενδογενούς συνάρτησης `chol()`**, διότι το μητρώο $A = C + C'$ δεν είναι Σ.Θ.Ο., αφού παράγεται ως **άθροισμα τυχαίων πυκνών μητρώων**. Επίσης και η μέθοδος $x = \text{pcg}(A, b)$ προϋποθέτει -όπως προαναφέρθηκε- το μητρώο A να είναι επίσης Σ.Θ.Ο., κάτι που δεν ισχύει. Άρα δια της άτοπου απαγγής, η σωστή απάντηση είναι η **δεύτερη**.

Άσκηση 26 (Ιούνιος 2018)

a) Για το παρακάτω πρόγραμμα να δείξετε ότι ακριβώς εκτυπώνεται αν εκτελέσουμε την εντολή: `mytest([1 1;-1 2;1 3;0 1])`. Για κάθε εντολή που εκτυπώνει τιμές να γράφετε και τον αριθμό της εντολής.

```
function [Q, R] = mytest(A)
[m, n] = size(A)
Q = zeros(m, n)
R = zeros(n)
R(1, 1)= norm(A(:, 1))
Q(:, 1) = A(:, 1)/R(1, 1)
for j = 2:n
    R(1:j-1, j) = Q(:, 1:j-1)' * A(:, j)
    temp = A(:, j) - Q(:, 1:j-1) * R(1:j-1, j)
    R(j, j) = norm(temp)
    Q(:, j) = temp/R(j, j)
end
```

b) Ποια από τις παρακάτω διαδικασίες υλοποιεί το πρόγραμμα;

- Παραγοντοποίηση QR με ανακλαστές Householder
- Παραγοντοποίηση QR με οδήγηση
- Κλασική ορθοκανονικοποίηση Gram - Schmidt

Λύση

a) $m = 4, n = 2$

$Q =$

$$\begin{matrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{matrix}$$

$R =$

$$\begin{matrix} 0 & 0 \\ 0 & 0 \end{matrix}$$

$R =$

```

1.7321      0
      0      0
Q =
  0.5774      0
 -0.5774      0
  0.5774      0
      0      0
R =
  1.7321  1.1547
      0      0
temp =
  0.3333
  2.6667
  2.3333
  1.0000
R =
  1.7321  1.1547
      0  3.6968
Q =
  0.5774  0.0902
 -0.5774  0.7213
  0.5774  0.6312
      0  0.2705

```

Παρατηρούμε ότι στο τέλος προκύπτει ένα **2x2 μητρώο R που είναι άνω τριγωνικό** και ένα **4x2μητρώο Q που είναι ίδιων διαστάσεων με το μητρώο A και ορθογώνιο**, γιατί αν πάρουμε το εσωτερικό γινόμενο των δύο στηλών του, αυτό ισούται με «0».

b) Όσον αφορά την **QR παραγοντοποίηση με μερική οδήγηση (pivoted QR)**, η τάξη του μητρώου A είναι $\text{rank}(A) = r < n < m$ και τότε (i) το παραγόμενο μητρώο R θα περιέχει μηδενικό στοιχείο στην κύρια διαγώνιο και δεν θα είναι αντιστρέψιμο, (ii) η QR μπορεί να καταρρεύσει στην επίλυση του προβλήματος των ελαχίστων τετραγώνων και (iii) οι πρώτες r στήλες του Q μπορεί να μην είναι βάση για το χώρο στηλών $\text{range}(A)$. Τα παραπάνω επιδιορθώνονται αν χρησιμοποιήσουμε **οδήγηση στηλών (column pivoting) κατά την QR**. Δηλαδή για να ξεχωρίσουμε πότε έχουμε την κλασσική QR παραγοντοποίηση με μητρώα Householder και πότε τη (σπανιότερη) QR παραγοντοποίηση με οδήγηση στηλών (column pivoting), θα πρέπει να **αναφέρουμε ότι για να έχουμε τη δεύτερη περίπτωση το παραγόμενο μητρώο R θα περιέχει μηδενικό στοιχείο στην κύρια διαγώνιο**, κάτι που δεν συμβαίνει στην προκειμένη περίπτωση, επομένως η σωστή απάντηση είναι **η πρώτη**.

Άσκηση 27 (Σεπτέμβριος 2019)

a) Έστω το μητρώο $A = [-1, -1, 0; 2, 3, 1; 0, -1, 2]$ και ότι $b = [1; 1; -1]$.

1. Ποιο βασικό αλγόριθμο χρησιμοποιεί η MATLAB αν καλέσουμε $A \setminus b$;
- παραγοντοποίηση LU με πλήρη οδήγηση και σε περίπτωση αστοχίας εφαρμογή μερικής οδήγησης.
 - LU με μερική οδήγηση και σε περίπτωση αστοχίας εφαρμογή πλήρους οδήγησης.
 - παραγοντοποίηση LU με μερική οδήγηση.
 - παραγοντοποίηση QR

Η σωστή απάντηση είναι **η δεύτερη**, και αυτό γιατί η προεπιλεγμένη μορφή παραγοντοποίησης στη MATLAB είναι η **LU με μερική οδήγηση**.

2. Να υπολογίσετε και να συμπληρώσετε τις τιμές που θα επιστραφούν αν κληθεί η συνάρτηση $[L, U] = lu(A); R = A - L * U; r = norm(R, 1); L, U, r$.

Λύση

Επειδή πρόκειται για την απλή LU παραγοντοποίηση, έχουμε $L_1 A = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} * \begin{bmatrix} -1 & -1 & 0 \\ 2 & 3 & 1 \\ 0 & -1 & 2 \end{bmatrix} = \begin{bmatrix} -1 & -1 & 0 \\ 0 & 1 & 1 \\ 0 & -1 & 2 \end{bmatrix}$ και

$$L_2 * L_1 * A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} -1 & -1 & 0 \\ 0 & 1 & 1 \\ 0 & -1 & 2 \end{bmatrix} = \begin{bmatrix} -1 & -1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 3 \end{bmatrix} = U. \text{ Ισχύει από πριν } L_2 * L_1 * A = U \Rightarrow A = (L_2 * L_1)^{-1} *$$

$$U \Rightarrow A = L_1^{-1} * L_2^{-1} * U, L = L_1^{-1} * L_2^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} * \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix}. \text{ Το ζητούμενο μητρώο είναι } R = A -$$

$$L * U = \begin{bmatrix} -1 & 1 & 0 \\ 2 & 3 & 1 \\ 0 & -1 & 2 \end{bmatrix} - \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} -1 & -1 & 0 \\ 0 & 1 & 3 \\ 0 & 0 & 3 \end{bmatrix} = \begin{bmatrix} -1 & 1 & 0 \\ 2 & 3 & 1 \\ 0 & -1 & 2 \end{bmatrix} - \begin{bmatrix} -1 & -1 & 0 \\ 2 & 3 & 3 \\ 0 & -1 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 2 & 0 \\ 0 & 0 & -2 \\ 0 & 0 & 2 \end{bmatrix}. \text{ Τέλος, } \eta$$

$norm(R, 1) = 4$.

Σημείωση: Ισχύει ότι στην LU παραγοντοποίηση το μητρώο $L = L_1^{-1} * L_2^{-1} * L_3^{-1} * \dots * L_n^{-1}$

3. Να βρείτε ορθογώνιο μητρώο Q τέτοιο ώστε η πρώτη στήλη του QA να είναι «0» στις θέσεις 2 και κάτω. Επίσης να υπολογίστε το στοιχείο του QA στη θέση (1, 1)

$$Q = \left(\quad \right), (QA)_{1,1} =$$

Λύση

Καταρχήν υπολογίζουμε το διάνυσμα Householder u_1 : $u_1 = x \pm \|x\|_2 e_1 = \begin{bmatrix} -1 \\ 2 \\ 0 \end{bmatrix} - \sqrt{5} \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} -3.2361 \\ 2 \\ 0 \end{bmatrix}$. Στη συνέχεια υπολογί-

$$\text{ζουμε το μητρώο Householder } H_1: H_1 = I - 2 * \frac{u_1 u_1^T}{u_1^T u_1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} - 2 * \frac{\begin{bmatrix} -3.2361 \\ 2 \\ 0 \end{bmatrix} * \begin{bmatrix} -3.2361 & 2 & 0 \end{bmatrix}}{\begin{bmatrix} -3.2361 & 2 & 0 \end{bmatrix} * \begin{bmatrix} -3.2361 \\ 2 \\ 0 \end{bmatrix}} = \begin{bmatrix} -0.4472 & 0.8944 & 0 \\ 0.8944 & 0.4472 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Το γινόμενο $H_1 * A = \begin{bmatrix} 2.236 & 3.1305 & 0.8944 \\ 0 & 0.4472 & 0.4472 \\ 0 & -1 & 2 \end{bmatrix}$. Στη συνέχεια επικεντρωνόμαστε στη δεξιά στήλη του μητρώου $H_1 * A$. Το

νέο διάνυσμα $x = \begin{bmatrix} 0 \\ 0.4472 \\ -1 \end{bmatrix}$. Το διάνυσμα Householder u_2 : $u_2 = x + \|x\|_2 * e_2 = \begin{bmatrix} 0 \\ 0.4472 \\ -1 \end{bmatrix} + 3.3166 * \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 1.5426 \\ -1 \end{bmatrix}$. Στη συ-

νέχεια υπολογίζουμε το μητρώο Householder H_2 : $H_2 = I - 2 * \frac{u_2 u_2^T}{u_2^T u_2} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} - 2 * \frac{\begin{bmatrix} 0 \\ 1.5426 \\ -1 \end{bmatrix} * \begin{bmatrix} 0 & 1.5426 & -1 \end{bmatrix}}{\begin{bmatrix} 0 & 1.5426 & -1 \end{bmatrix} * \begin{bmatrix} 0 \\ 1.5426 \\ -1 \end{bmatrix}} =$

$\begin{bmatrix} 1 & 0 & 0 \\ 0 & -0.4082 & 0.9129 \\ 0 & 0.9129 & 0.4082 \end{bmatrix}$. Στο τέλος το γινόμενο των μητρώων $H_2 * H_1 * A = \begin{bmatrix} 2.2361 & 3.1305 & 0.8944 \\ 0 & -1.0954 & 1.6432 \\ 0 & 0 & 1.2247 \end{bmatrix}$. Το μητρώο $Q = (H_2$

$$* H_1)^{-1} = H_1^{-1} * H_2^{-1} = H_1^T * H_2^T = H_1 * H_2 = \begin{bmatrix} -0.4472 & 0.8944 & 0 \\ 0.8944 & 0.4472 & 0 \\ 0 & 0 & 1 \end{bmatrix} * \begin{bmatrix} 1 & 0 & 0 \\ 0 & -0.4082 & 0.9129 \\ 0 & 0.9129 & 0.4082 \end{bmatrix}$$

4. Συγκλίνει η μέθοδος Jacobi για την επίλυση του παραπάνω $Ax = b$;

Λύση

Επειδή το μητρώο $A = \begin{bmatrix} -1 & -1 & 0 \\ 2 & 3 & 1 \\ 0 & -1 & 2 \end{bmatrix}$ είναι ΜΑΔΚ, συγκλίνει. Πράγματι, τα κύρια υπομητρώα $A1$ και $A2$ είναι μη - τετραγωνικά, ενώ υπάρχει ΑΔΚ για μια τουλάχιστον γραμμή ή στήλη.

5. Αν στο άνω $Ax = b$ εφαρμόστε ένα βήμα Jacobi με $x^{(0)}$ το $(1, -1, 1)^T$, βρείτε το $x^{(1)}$?

$$x^{(1)} = (\quad , \quad , \quad)^T$$

Λύση

Εφαρμόζουμε το επαναληπτικό σχήμα της μεθόδου Jacobi, $x^{(k+1)} = D^{-1}(L + U) * x^{(k)} + D^{-1}b$, $k = 0, 1, 2, \dots$,

χρησιμοποιώντας ως αρχική εκτίμηση λύσης την $x^{(0)} = [1, -1, 1]^T$. Το μητρώο των συντελεστών των αγνώστων και το

σταθερό διάνυσμα είναι αντίστοιχα: $A = \begin{bmatrix} -1 & -1 & 0 \\ 2 & 3 & 1 \\ 0 & -1 & 2 \end{bmatrix}$, $b = \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}$. Τα μητρώα D , L , U είναι: $D = \begin{bmatrix} -1 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 2 \end{bmatrix}$, $L = \begin{bmatrix} 0 & 0 & 0 \\ -2 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$, $U = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & -1 \\ 0 & 0 & 0 \end{bmatrix}$, οπότε το μητρώο επανάληψης T είναι: $T = D^{-1} * (L + U) = \begin{bmatrix} 0 & -1 & 0 \\ -2/3 & 0 & -1/3 \\ 0 & 1/2 & 0 \end{bmatrix}$ και το

διάνυσμα $D^{-1} * b = \begin{bmatrix} -1/1 & 0 & 0 \\ 0 & 1/3 & 0 \\ 0 & 0 & 1/2 \end{bmatrix} * \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix} = [-1, \frac{1}{3}, -\frac{1}{2}]^T$. Γνωρίζοντας πλέον τις τιμές των μητρώων έχουμε:

$$x^{(k+1)} = \begin{bmatrix} 0 & -1 & 0 \\ -2/3 & 0 & -1/3 \\ 0 & 1/2 & 0 \end{bmatrix} * x^k + \begin{bmatrix} -1 \\ 1/3 \\ -1/2 \end{bmatrix} = \begin{bmatrix} 0 & -1 & 0 \\ -2/3 & 0 & -1/3 \\ 0 & 1/2 & 0 \end{bmatrix} \begin{bmatrix} x_1^{(k)} \\ x_2^{(k)} \\ x_3^{(k)} \end{bmatrix} + \begin{bmatrix} -1 \\ \frac{1}{3} \\ -\frac{1}{2} \end{bmatrix} \Rightarrow \begin{bmatrix} x_1^{(k+1)} \\ x_2^{(k+1)} \\ x_3^{(k+1)} \end{bmatrix} = \begin{bmatrix} -x_2^{(k)} \\ -\frac{2}{3}x_1^{(k)} - \frac{1}{3}x_3^{(k)} \\ \frac{1}{2}x_2^{(k)} - \frac{1}{2} \end{bmatrix} +$$

$$\begin{bmatrix} -1 \\ \frac{1}{3} \\ -\frac{1}{2} \end{bmatrix} = \begin{bmatrix} -x_2^{(k)} - 1 \\ -\frac{2}{3}x_1^{(k)} - \frac{1}{3}x_3^{(k)} + \frac{1}{3} \\ \frac{1}{2}x_2^{(k)} - \frac{1}{2} \end{bmatrix} \Rightarrow \text{για } \kappa = 0 \text{ έχουμε } \begin{bmatrix} x_1^{(1)} \\ x_2^{(1)} \\ x_3^{(1)} \end{bmatrix} = \begin{bmatrix} -1 - 1 \\ -\frac{2}{3} - \frac{1}{3} + \frac{1}{3} \\ -\frac{1}{2} + \frac{1}{2} \end{bmatrix} = \begin{bmatrix} -2 \\ -\frac{2}{3} \\ 0 \end{bmatrix}.$$

Ερώτηση 2

Δίνεται η εξίσωση: $\alpha_1 + \alpha_2 \frac{1}{(1+x^2)} + \alpha_3 x^3 = 0$, όπου $\alpha_1 = 40$, $\alpha_2 = -1$, $\alpha_3 = -1$.

1. Επιλέξτε το καλύτερο από τα παρακάτω διαστήματα για τον εντοπισμό μιας ρίζας για την εκκίνηση της μεθόδου διχοτόμησης (bisection).

(3, 4)

(3, 5)

(-3, 0)

Λύση

Δοκιμάζουμε σε ποιο διάστημα ισχύει το θεώρημα Bolzano. Παρατηρούμε ότι $f(3) * f(4) = \frac{129}{10} * \frac{-409}{17} < 0$. Επομένως, το καλύτερο από τα παρακάτω διαστήματα για τον εντοπισμό μιας ρίζας για τη διχοτόμηση (bisection) είναι το (3, 4).

2. Υπολογίστε προσέγγιση στη ρίζα με ένα βήμα της secant (τέμνουσας) εκκινώντας από τα σημεία $x^{(0)} = L + (R - L)/3$, $x^{(1)} = L + (R - L)/2$ όπου L , R είναι το αριστερό και δεξιό άκρο του διαστήματος που επιλέξατε στο υποερώτημα 1.

Δίνονται ότι $x^{(0)} = L$ και $x^{(1)} = R$.

Λύση

Εφαρμόζουμε τον τύπο της τέμνουσας και έχουμε ότι: $x_{k+1} = x_k - \frac{(x_k - x_{k-1})}{f(x_k) - f(x_{k-1})} * f(x_k) \Rightarrow$ για $k = 1$ έχουμε ότι $x^2 = x^{(1)} - \frac{x^{(1)} - x^{(0)}}{f(x^{(1)}) - f(x^{(0)})} * f(x^{(0)}) = L + \frac{R-L}{3} - \frac{L + \frac{R-L}{3} - (L + \frac{R-L}{2})}{f(L + \frac{R-L}{3}) - f(L + \frac{R-L}{2})} * f(L + \frac{R-L}{2})$

Ερώτηση 3

Αν a , b , c είναι α.κ.ν στη MATLAB στο διάστημα $[1, 2000]$, C η ακριβής τιμή του τύπου $(a + b) * c$ (δηλαδή χωρίς σφάλμα στρογγύλευσης), X η τιμή που υπολογίζει η MATLAB, u η μονάδα στρογγύλευσης, επιλέξτε το καλύτερο από τα παρακάτω ως άνω φράγμα για το συνολικό σφάλμα στρογγύλευσης:

$|C - X| \leq |C|2u^2$

$|C - X| \leq |C| (1 + u)$

$|C - X| \leq |C| u$

$|C - X| \leq |C| (2u + u^2)$

Λύση

Σωστή απάντηση είναι η **τέταρτη**, διότι για σφάλματα της μορφής: $\frac{|(x_1 \otimes x_2) \otimes x_3 - x_1 x_2 x_3|}{|x_1 x_2 x_3|} \leq 2u + u^2$.

Ερώτηση 4

Δίνονται τα παρακάτω ζεύγη κόμβων – τιμών για κάποια συνάρτηση

	i = 0	i = 1	i = 2
x _i	-2.0	0	2.0
f(x _i)	4.5	5.0	0.5

1. Να γράψετε το πολυώνυμο παρεμβολή q για τη συνάρτηση σε μορφή Newton.

Λύση

$$\begin{aligned} x_0 &= -2 & 4.5 \\ x_1 &= 0 & 5 \\ x_2 &= 2 & 0.5 \end{aligned}$$

$$\begin{aligned} \frac{5-4.5}{0-(-2)} &= \frac{1}{4} \\ \frac{0.5-5}{2-0} &= \frac{-4.5}{2} \\ \frac{-4.5-1}{2-4} &= \frac{-10}{16} = -\frac{5}{8} \end{aligned}$$

Επομένως το πολυώνυμο παρεμβολής σε μορφή Newton είναι: $q(x) = 4.5 + \frac{1}{4} * (x+2) - \frac{5}{8} * (x+2)^2 * x$

2. Στη συνέχεια δίνεται επιπροσθέτως η τιμή της παραγώγου $f'(x_0) = 0.25$. Να υπολογίσετε το πολυώνυμο παρεμβολής Hermite της συνάρτησης βάσει αυτών των τιμών.

Λύση

$$\begin{aligned} x_0 &= -2 & 4.5 \\ x_1 &= -2 & 4.5 \\ x_2 &= 0 & 5 \\ x_3 &= 2 & 0.5 \end{aligned}$$

$$\begin{aligned} \frac{0-4.5}{0} &\rightarrow \frac{f'(-2)}{1!} = 0.25 = \frac{1}{4} \\ \frac{5-4.5}{0-(-2)} &= \frac{1}{4} \\ \frac{0.5-5}{2-0} &= -\frac{4.5}{2} = -\frac{9}{4} \\ \frac{1/4-1/4}{5-4.5} &= \frac{0}{4.5} = 0 \\ \frac{-9/4-1/4}{2-(-2)} &= -\frac{5}{8} \\ &- \frac{5}{32} \end{aligned}$$

Επομένως το πολυώνυμο παρεμβολής Hermite είναι: $h(x) = 4.5 + \frac{1}{4} * (x+2) - \frac{5}{32} * (x+2)^2 * x$

3. Για την ίδια συνάρτηση f και το q στο 1^o υποερώτημα, αν γνωρίζουμε ότι η 3^η παράγωγος $\|f^{(3)}\|_{\infty} \leq M$ για κάποια σταθερά M, να υπολογίσετε την τιμή γ ώστε το γM να είναι ένα καλό άνω φράγμα για το σφάλμα $|f(x) - q(x)| \forall x \in (0, 1)$.

Λύση

Επειδή δίνονται τα όρια του διαστήματος (0, 1), για να βρεθεί ένα καλό άνω φράγμα για το εμπρός σφάλμα, θα χρησιμο-

πούσον με τον τύπο: $\max |e_n f(x)| \leq \frac{\max_{x \in \Omega} |f^{(n+1)}(x)|}{4(n+1)} \left(\frac{x_n - x_0}{n}\right)^{(n+1)} \Rightarrow \max |e_n f(x)| \leq \frac{\max_{x \in \Omega} |f^{(n+1)}(x)|}{4(n+1)} h^{(n+1)}$ όπου $h = \frac{x_n - x_0}{n}$. Στην

προκειμένη περίπτωση έχουμε ότι το άνω φράγμα θα είναι: $\max |e_n f(x)| \leq \frac{\max_{x \in \Omega} |f^{(n+1)}(x)|}{4(n+1)} \left(\frac{x_n - x_0}{n}\right)^{(n+1)} \leq \frac{M}{4 \cdot 3} \left(\frac{1-0}{2}\right)^3 = \frac{M}{96}$.

Άρα, θα έχουμε ότι: $|f(x) - q(x)| \leq \gamma M$ όπου $\gamma = 1/96$.

Ερώτηση 5

1. Να υπολογίσετε προσέγγιση I για το ολοκλήρωμα $\int_0^1 \frac{\sin(2\pi x)}{4+x^2} dx$ χρησιμοποιώντας τον σύνθετο κανόνα του μέσου σημείου με 2 ισομεγέθη υποδιαστήματα.

Λύση

Εφαρμόζοντας τη σύνθετη μέθοδο μέσου σημείου, έχουμε:

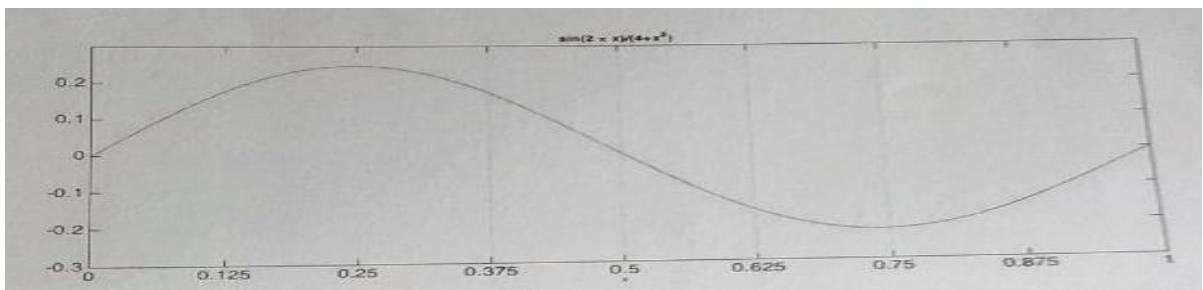
$$CM(f) = h \sum_{i=0}^{n-1} f(a + h \frac{2i+1}{2})$$

Εφαρμόζοντας τη σύνθετη μέθοδο μέσου σημείου, έχουμε: $\left| \int_a^b f(x) - CM(f) \right| = \frac{(b-a)|f''(\eta)|}{24} h^2 \leq \frac{(b-a)M}{24} h^2$. Το σφάλμα της σύνθετης μεθόδου μέσου σημείου, έχουμε:

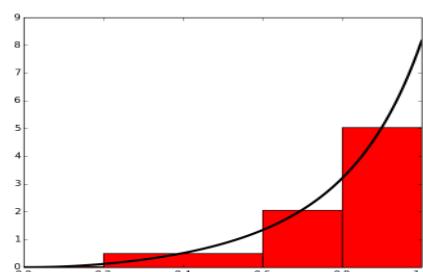
$$\frac{1.00-0.00}{2} = 0.5. \text{ Εφαρμόζουμε στη συνέχεια τη σύνθετη μέθοδο μέσου σημείου και έχουμε, } I(f) = h * \left[\left(f(a + h \frac{2*0+1}{2}) + f(a + h \frac{2*1+1}{2}) \right) \right] = 0.5 * \left[\left(f(0 + 0.5 \frac{2*0+1}{2}) + f(0 + 0.5 \frac{2*1+1}{2}) \right) \right] = 0.5 * \left[\left(f(\frac{0.5}{2}) + f(\frac{1.5}{2}) \right) \right] = 0.5 * \left[\frac{\sin(2\pi \frac{1}{4})}{4+(\frac{1}{4})^2} + \frac{\sin(2\pi \frac{1.5}{2})}{4+(\frac{1.5}{2})^2} \right], h = 1/2$$

Η τιμή της προσέγγισης είναι $I =$

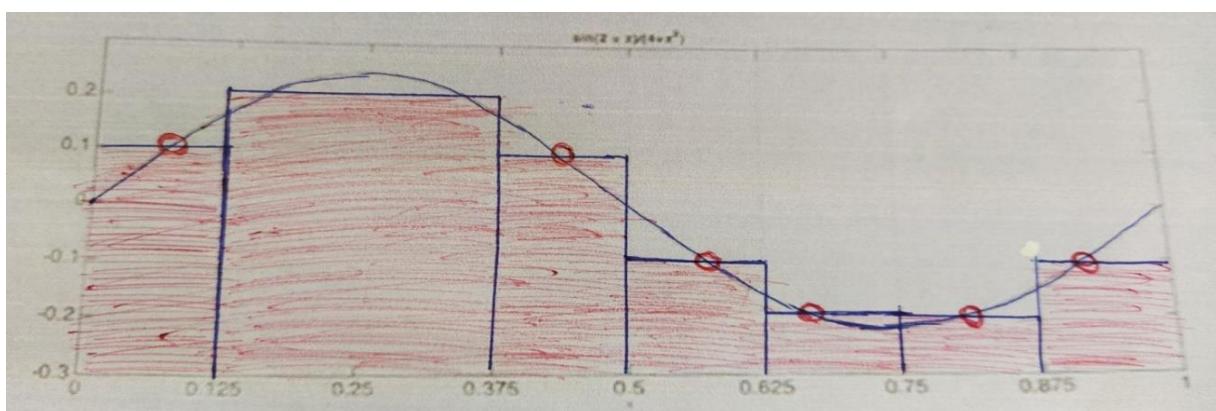
2. Να αναδείξετε ξεκάθαρα στο σχήμα τα πολύγωνα (ορίζοντας τις πλευρές και τις συντεταγμένες των ακμών τους επάνω στο σχήμα και σκιάζοντάς τα με ελαφριά διαγράμμιση) το συνολικό εμβαδόν των οποίων χρησιμοποιείται από την παραπάνω μέθοδο για να προσεγγιστεί το ολοκλήρωμα.

**Λύση**

Φροντίζουμε ώστε η καμπύλη να τέμνει τα πολύγωνα που σχεδιάζουμε στο μέσο της πάνω πλευράς τους, όπως φαίνεται στη συνέχεια.



Στην προκειμένη περίπτωση τα ζητούμενα πολύγωνα φαίνονται στη συνέχεια:



Ερώτηση 6

Αν για κάποια συνάρτηση $f: [a, b] \rightarrow \mathbb{R}$ και $f(a)f(b) < 0$ τότε ισχύει πάντα ότι η μέθοδος διχοτόμησης (bisection) θα συγκλίνει σε ρίζα, δηλαδή f , ζ.

□ ΛΑΘΟΣ

□ ΣΩΣΤΟ

Λύση

Σωστό, όταν ικανοποιείται το Θεώρημα Bolzano, τότε ισχύει πάντα ότι η μέθοδος διχοτόμησης (bisection) θα συγκλίνει σε ρίζα.

Ερώτηση 7

Για το σύστημα μη γραμμικών εξισώσεων: $-3x + y^2 = -2, x + y - \frac{1}{2} \exp(y) = -2$ συμπληρώστε τις τιμές για το **πρώτο βήμα** της Newton αν $(x^{(0)}, y^{(0)})^T = (1, 0)$.

$$\begin{pmatrix} x^{(1)} \\ y^{(1)} \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} - \left(\quad \right)^{-1} \left(\quad \right)$$

Λύση

Έστω $f_1(x, y) = -3x + y^2 + 2$ και $f_2(x, y) = x + y - \frac{1}{2} \exp(y) + 2 = +y - \frac{1}{2} * e^y$. Όταν θέλουμε να λύσουμε σύστημα μη - γραμμικών εξισώσεων, χρησιμοποιούμε τη μέθοδο **Newton**. Υπολογίζουμε πρώτα το Ιακωβιανό μητρώο, $J(x)$, το οποίο περιέχει τις μερικές παραγώγους κάθε διθείσας συνάρτησης του γραμμικού συστήματος ως προς όλες τις μεταβλητές. Ισχύει ότι η $(e^x)' = e^x$

$$J(x) = \begin{bmatrix} \frac{\partial f_1}{\partial x} & \frac{\partial f_1}{\partial y} \\ \frac{\partial f_2}{\partial x} & \frac{\partial f_2}{\partial y} \end{bmatrix} = \begin{bmatrix} -3 & 2y \\ 1 & 1 - \frac{e^y}{2} \end{bmatrix}. \text{ Επειδή η αρχική εκτίμηση λύσης } [x^{(0)}, y^{(0)}]^T = [1, 0]^T, \text{ το } J(x^0) * s^0 = -F_n(x^0) \Rightarrow$$

$$\begin{bmatrix} -3 & 2 * 0 \\ 1 & 1 - \frac{e^0}{2} \end{bmatrix} * \begin{bmatrix} s_1^{(0)} \\ s_2^{(0)} \end{bmatrix} = - \begin{bmatrix} f_1(x_1^{(0)}, x_2^{(0)}) \\ f_2(x_1^{(0)}, x_2^{(0)}) \end{bmatrix} \Rightarrow \begin{bmatrix} -3 & 0 \\ 1 & \frac{1}{2} \end{bmatrix} * \begin{bmatrix} s_1^{(0)} \\ s_2^{(0)} \end{bmatrix} = - \begin{bmatrix} f_1(x_1^{(0)}, x_2^{(0)}) \\ f_2(x_1^{(0)}, x_2^{(0)}) \end{bmatrix} \Rightarrow \begin{bmatrix} -3 & 0 \\ 1 & \frac{1}{2} \end{bmatrix} * \begin{bmatrix} s_1^{(0)} \\ s_2^{(0)} \end{bmatrix} = - \begin{bmatrix} -1 \\ 5/2 \end{bmatrix}$$

$$\Rightarrow \begin{bmatrix} s_1^{(0)} \\ s_2^{(0)} \end{bmatrix} = - \begin{bmatrix} -3 & 0 \\ 1 & \frac{1}{2} \end{bmatrix}^{-1} * \begin{bmatrix} -1 \\ 5/2 \end{bmatrix}. \text{ Η επόμενη εκτίμηση λύσης είναι: } \begin{bmatrix} x^{(1)} \\ y^{(1)} \end{bmatrix} = \begin{bmatrix} x^{(0)} \\ y^{(0)} \end{bmatrix} + \begin{bmatrix} s_1^{(0)} \\ s_2^{(0)} \end{bmatrix} = \begin{bmatrix} x^{(0)} \\ y^{(0)} \end{bmatrix} - \begin{bmatrix} -3 & 0 \\ 1 & \frac{1}{2} \end{bmatrix}^{-1} *$$

$\begin{bmatrix} -1 \\ 5/2 \end{bmatrix}$ και είναι στη μορφή της εκφώνησης.

Σημείωση: Αν θέλαμε να βρούμε το ακριβές διάνυσμα διόρθωσης $\begin{bmatrix} s_1^{(0)} \\ s_2^{(0)} \end{bmatrix}$, θα επιλύαμε την εξίσωση $\begin{bmatrix} -3 & 0 \\ 1 & \frac{1}{2} \end{bmatrix} * \begin{bmatrix} s_1^{(0)} \\ s_2^{(0)} \end{bmatrix} = - \begin{bmatrix} -1 \\ 5/2 \end{bmatrix} \Rightarrow$ από την επίλυση του συγκεκριμένου γραμμικού συστήματος έχουμε: $-3 * s_1^0 + 0 * s_2^0 = 1$ και $1 * s_1^0 + \frac{1}{2} * s_2^0 = -5/2 \Rightarrow s_2^0 = -13/3$ και $s_1^0 = -1/3$ και αυτή είναι ουσιαστικά η διόρθωση της προσεγγιστικής λύσης. Η επόμενη εκτίμηση λύσης είναι: $\begin{bmatrix} x^{(1)} \\ y^{(1)} \end{bmatrix} = \begin{bmatrix} x^{(0)} \\ y^{(0)} \end{bmatrix} + \begin{bmatrix} s_1^{(0)} \\ s_2^{(0)} \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} - \begin{bmatrix} \frac{1}{3} \\ \frac{13}{3} \end{bmatrix} = \begin{bmatrix} \frac{2}{3} \\ -\frac{13}{3} \end{bmatrix}$.

Οσα περισσότερα βήματα της παραπάνω διαδικασίας εκτελούνται, τόσο καλύτερη προσέγγιση της ακριβής λύσης επιτυγχάνεται. Γενικότερα ισχύει ο τύπος: $\begin{bmatrix} x_1^{(k+1)} \\ x_2^{(k+1)} \\ \dots \\ x_n^{(k+1)} \end{bmatrix} = \begin{bmatrix} x_1^{(k)} \\ x_2^{(k)} \\ \dots \\ x_n^{(k)} \end{bmatrix} + \begin{bmatrix} s_1^{(k)} \\ s_2^{(k)} \\ \dots \\ s_n^{(k)} \end{bmatrix}$.

Άσκηση 28 (Φεβρουάριος 2020)

a) Δίνονται τα παρακάτω ζεύγη κόμβων – τιμών για κάποια συνάρτηση

	i = 0	i = 1	i = 2
x _i	0	0.50	1.00
f(x _i)	2.00	0.80	-2.50

1. Να γράψετε το πολυώνυμο παρεμβολής για τη συνάρτηση σε μορφή Lagrange. Προσοχή: Να απλοποιήσετε το αποτέλεσμα έτσι ώστε ο σταθερός συντελεστής κάθε όρου της αναπαράστασης να είναι σε κλασματική μορφή και όχι γινόμενο τιμών.

Το πολυώνυμο σε αναπαράσταση Lagrange είναι: p(x) = ...

$$P_2(x) = \frac{(x-x_1)(x-x_2)}{(x_0-x_1)(x_0-x_2)} f_0 + \frac{(x-x_0)(x-x_2)}{(x_1-x_0)(x_1-x_2)} f_1 + \frac{(x-x_0)(x-x_1)}{(x_2-x_0)(x_2-x_1)} f_2 = \frac{(x-0.50)(x-1)}{(0-0.50)(0-1)} * 2 + \frac{(x-0)(x-1)}{(0.50-0)(0.50-1)} * 0.80 + \frac{(x-0)(x-0.50)}{(1-0)(1-0.50)} * (-2.50) = 0.025 * (78x^2 - 183x + 80)$$

2. Να γράψετε τη βαρυκεντρική αναπαράσταση (επιλέξτε όποια από τις μορφές προτιμάτε) για το παραπάνω πολυώνυμο παρεμβολής.

Πρέπει πρώτα να υπολογίσουμε τα βάρη για κάθε συντελεστή από τον τύπο $w_i = \frac{1}{(x_i-x_0)\dots(x_i-x_{i-1})(x_i-x_{i+1})\dots(x_i-x_n)}$. Για παράδειγμα το $w_0 = \frac{1}{(x_0-x_1)\cdot(x_0-x_2)} = \frac{1}{(0-0.5)\cdot(0-1)} = \frac{1}{0.5}$. Με ανάλογο τρόπο υπολογίζουμε στη συνέχεια και τα βάρη $w_1 = \frac{1}{(0.5-0)\cdot(0.5-1)} = -\frac{1}{0.25}$ και $w_2 = \frac{1}{(1-0)\cdot(1-0.5)} = \frac{1}{0.5}$. Στη συνέχεια η βαρυκεντρική αναπαράσταση μπορεί να γραφεί ως $B(x) = \frac{\sum_{i=0}^{n=2} f_i \frac{w_i}{(x-x_i)}}{\sum_{i=0}^{n=2} \frac{w_i}{(x-x_i)}} = \frac{f_0 \frac{w_0}{(x-x_0)} + f_1 \frac{w_1}{(x-x_1)} + f_2 \frac{w_2}{(x-x_2)}}{\frac{w_0}{(x-x_0)} + \frac{w_1}{(x-x_1)} + \frac{w_2}{(x-x_2)}} = \frac{2.00 * \frac{w_0}{(x-0)} + 0.85 * \frac{w_1}{(x-0.5)} + (-2.50) * \frac{w_2}{(x-1)}}{\frac{w_0}{(x-0)} + \frac{w_1}{(x-0.5)} + \frac{w_2}{(x-1)}} = \frac{2.00 * \frac{1}{0.5} + 0.85 * \frac{-1}{0.25} - 2.50 * \frac{1}{0.5}}{\frac{w_0}{(x-0)} + \frac{w_1}{(x-0.5)} + \frac{w_2}{(x-1)}}$.

Ερώτηση 2

Δίνεται η $f(x) = (x-1)^2 - (x-1)e^{-(x-1)} - 1$

1. Υπολογίστε το καλύτερο από τα παρακάτω διανύσματα για την έναρξη της μεθόδου διχοτόμησης για τον εντοπισμός μιας ρίζας της εξίσωσης $f(x) = 0$.

[0, 1]

[-1, 0]

[-2, -1]

2. Υπολογίστε προσέγγιση στη ρίζα με 1 βήμα της secant (τέμνουσας) εκκινώντας από τα σημεία $x^{(0)} = L$ και $x^{(n)} = R$, όπου L, R είναι το αριστερό και δεξιό άκρο του διαστήματος που βρήκατε στο υποερώτημα 1.

Ερώτηση 3

1. Αν ένα πρόγραμμα περιέχει την εντολή $y = (\text{sqrt}(x) - \text{sqrt}(x-h))/h$, τι πρόβλημα μπορεί να προκύψει για μικρές (αλλά μη μηδενικές) τιμές του h και πως μπορείτε να το αποφύγετε;

2. Έστω σύστημα αριθμητικής κινητής υποδιαστολής όπως της IEEE 754 μόνον που έχει μόνον 3+1 bits στην ουρά (το 1 κρυμμένο) δηλαδή στο δυαδικό σύστημα η ουρά είναι της μορφής $[1\text{-}]\square\square\square$. Κατά τα άλλα είναι όπως το πρότυπο και το προεπιλεγμένο σύστημα στρογγύλευσης είναι (όπως και στην IEEE 754) προς το "πλησιέστερο ζυγό". Για την ερώτηση, δεν ενδιαφέρει το μήκος του εκθέτη (μπορεί να είναι όσο μεγάλο χρειάζεται.)

(α) ποιο θα είναι το έψιλον της μηχανής:

2^{-4}

2^{-2}

2^{-5}

2^{-3}

(β) τι αποτέλεσμα θα προκύψει αν η άθροιση γίνει από τα αριστερά προς τα δεξιά?

i) $f(1 + 2^{-3} + 2^{-4} + 2^{-5})$

ii) $f(2^{-3} + 2^{-4} + 2^{-5} + 1)$

Αύση

2α. Γνωρίζουμε ότι το $\text{eps} = \varepsilon_M = \frac{2^{-(t-1)}2^e}{2^e} = 2^{1-t} = 2^{1-4} = 2^{-3}$. Εναλλακτικά, $\text{eps} = 2 * u = 2 * 2^{-4} = 2^{-3}$. Πρέπει να σημειωθεί ότι στον υπολογισμό του t λαμβάνουμε υπόψη και το hidden bit.

2β. Στο IEEE – 754 και στη διπλή ακρίβεια η απορρόφηση γίνεται όταν οι εκθέτες διαφέρουν από 53 και πάνω, όπου $t = 52$ είναι η ακρίβεια των δεκαδικών ψηφίων. Στην προκειμένη περίπτωση το $t = 3$. Στην πρώτη περίπτωση η άθροιση θα δώσει αποτέλεσμα $1 + \text{eps} = 1^+$. Στη δεύτερη περίπτωση η άθροιση θα δώσει αποτέλεσμα 2^{-2} .

Ερώτηση 4

Έστω το μητρώο $A = [-1, -1, 0; 2, 3, 1; 0, -1, 2]$ και ότι $b = [1; 1; -1]$.

1. Να εκτελέσετε 2 βήματα της επαναληπτικής μεθόδου Jacobi με αρχικό διάνυσμα $x_j^{(0)} = 0$ και να υπολογίσετε την προσέγγιση $x_j^{(2)}$ και τη ζητούμενη νόρμα του αντίστοιχου καταλοίπου:

$$\begin{aligned} x_j^{(2)} &= []^T \\ \|r_j^{(2)}\|_1 &= \end{aligned}$$

Είναι απαραίτητος ο υπολογισμός του μητρώου επανάληψης T , προκειμένου να υπολογιστεί το ζητούμενο.

$T = D^{-1} * (L + U) = []$. Επίσης, το μητρώο D^{-1} ισούται με το μητρώο D , αφού όλα τα διαγώνια στοιχεία του μητρώου D είναι μονάδες. Επίσης, έχουμε ότι το διάνυσμα $g = D^{-1} * b = \begin{bmatrix} -1 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 2 \end{bmatrix} * \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix} = \begin{bmatrix} -1 \\ 3 \\ -2 \end{bmatrix}$, οπότε το **επαναληπτικό σχήμα της μεθόδου Jacobi** για το πρόβλημα που μας δόθηκε είναι το εξής: $\begin{bmatrix} x_1^{(k+1)} \\ x_2^{(k+1)} \\ x_3^{(k+1)} \end{bmatrix} = D^{-1} * (L + U) * x^{(k)} + D^{-1} * b =$

$$\begin{bmatrix} x_1^{(k)} \\ x_2^{(k)} \\ x_3^{(k)} \end{bmatrix} + \begin{bmatrix} -1 \\ 3 \\ 2 \end{bmatrix} = \begin{bmatrix} -1 \\ 3 \\ 2 \end{bmatrix} \text{ για } k = 0$$

2. Να εκτελέσετε 1 βήμα της επαναληπτικής μεθόδου Gauss-Seidel με αρχικό διάνυσμα $x_{??}^{(0)} = 0$ και να υπολογίσετε την αντίστοιχη προσέγγιση $x_{GS}^{(1)}$ και τη ζητούμενη νόρμα του καταλοίπου:

$$\begin{aligned} x_{GS}^{(1)} &= []^T \\ \|r_{GS}^{(1)}\|_1 &= \end{aligned}$$

3. Επιλέξτε και κυκλώστε Σωστό ή Λάθος (ΧΕ): Η μέθοδος Jacobi για το παραπάνω μητρώο θα συγκλίνει.

Σ Λ

Η μέθοδος Jacobi συγκλίνει για το συγκεκριμένο μητρώο συντελεστών A , επειδή ικανοποιείται το δεύτερο κριτήριο σύγκλισης, δηλ. το μητρώο A είναι ΜΑΔΚ, όπως φαίνεται στη συνέχεια:

$$A = \left[\begin{array}{c|cc} -1 & -1 & 0 \\ \hline 2 & 3 & 1 \\ 0 & -1 & 2 \end{array} \right]$$

4. Για το παραπάνω γραμμικό σύστημα, ποιο βασικό αλγόριθμο χρησιμοποιεί η MATLAB αν καλέσουμε $A \setminus b$: (σημειώστε το κουτάκι – ΧΕ):

- παραγοντοποίηση LU με πλήρη οδήγηση
- παραγοντοποίηση LU με πλήρη οδήγηση και σε περίπτωση αστοχίας εφαρμογή μερικής οδήγησης
- παραγοντοποίηση LU με μερική οδήγηση
- παραγοντοποίηση Cholesky και σε περίπτωση αστοχίας παραγοντοποίηση LDL^T .

Η σωστή απάντηση είναι η τρίτη

Αν το μητρώο των συντελεστών είναι το $A = \begin{bmatrix} -1 & -1 & 0 \\ 2 & 3 & 1 \\ 0 & -1 & 2 \end{bmatrix}$, το οποίο δεν έχει κάποιο ιδιαίτερο δομικό χαρακτηριστικό (έναι απλά ένα πυκνό μητρώο), η σωστή απάντηση είναι η τρίτη, διότι η MATLAB προεπιλεγμένα χρησιμοποιεί LU με μερική οδήγηση (PLU) για να λύσει το γραμμικό σύστημα $A * x = b$. Θα πρέπει όμως να αναφερθεί ότι η MATLAB ανιχνεύει πρώτα από όλα αν το μητρώο είναι συμμετρικό και σε περίπτωση που είναι, εκτελεί την Cholesky και αν αυτή επιτύχει, σταματά εκεί. Αν το μητρώο των συντελεστών είναι το $A = \begin{bmatrix} 10 & -1 & -1 \\ -1 & 10 & -1 \\ -1 & -1 & 10 \end{bmatrix}$, το οποίο είναι ήδη συμμετρικό, επειδή έχει και A.D.K..

5. Επιλέξτε και κυκλώστε Σωστό ή Λάθος (XE): Av $C = C^T \in \mathbb{R}^{n \times n}$ και γνωρίζουμε ότι η μέθοδος Jacobi για το $C * x = b$ συγκλίνει, τότε ισχύει ότι $\|e^{(k+1)}\|_2 \leq \|e^{(k)}\|_2$ όπου $e^{(k)} = C^{-1}b - x^{(k)}$.

 $\square \Sigma$ $\square \Lambda$

Είναι **σωστή** η πρόταση, διότι καθώς η μέθοδος Jacobi συγκλίνει, σημαίνει ότι καθώς αυξάνονται τα βήματα η διαφορά της ακριβής λύσης του γραμμικού συστήματος $C * x = b$, δηλ. το $C^{-1}b$ από την προσεγγιστική λύση $x^{(k)}$ μειώνεται.

Άσκηση 29 (Θέμα Φροντιστηρίου)

α) Για το μητρώο $A = \begin{bmatrix} 1 & 2 & 1 \\ -1 & 1 & 1 \\ 1 & 3 & 3 \end{bmatrix}$ να αποφανθείτε περί σύγκλισης της μεθόδου Jacobi, για οποιαδήποτε διάνυσμα αρχικής προσέγγισης και δεξιού μέλους (κατάλληλων διαστάσεων) αντίστοιχα.

β) Να κατασκευαστεί συνάρτηση σε Matlab (jac.m) η οποία θα υλοποιεί την επαναληπτική μέθοδο **Jacobi** και παρουσιάζει τα ακόλουθα στοιχεία εισόδου-εξόδου:

Είσοδος

- $x1$: διάνυσμα αρχικής προσέγγισης
- A : Τετραγωνικό μητρώο συντελεστών
- b : δεξί μέλος
- $maxit$: Μέγιστο πλήθος επαναλήψεων
- tol : Βαθμωτός που δηλώνει την ανεκτικότητα της διαφοράς μεταξύ διαδοχικών προσεγγίσεων (ως προς 2^n νόρμα)

Έξοδος

- it : Το πλήθος των επαναλήψεων
- $x2$: Το διάνυσμα της τελικής προσέγγισης

Λύση

α) Το μητρώο A περιέχει **αποκλειστικά μη μηδενικά στοιχεία. Το αντίστοιχο γράφημα γειτνίασης θα περιέχει όλες τις πιθανές ακμές, συνεπώς θα είναι ισχυρά συνεκτικό**. Επίσης, το A παρουσιάζει αυστηρά διαγώνια κυριαρχία για την τρίτη του στήλη. Άρα, είναι ΜΑΔΚ και η Jacobi θα συγκλίνει.

β) Η ζητούμενη συνάρτηση φαίνεται στη συνέχεια:

```
function [it, x2] = jac(x1, A, b, maxit, tol)
M = diag(diag(A)); N = -triu(A, 1)-tril(A, -1); n = length(x1); %N = -triu(A, 1) - tril(A, -1) = U + L
x2 = ones(n,1) * inf; %αρχική τιμή διανύσματος x2
it = 0;
while ((it < maxit) && (norm(x1-x2) > tol))
    if (it~= 0)
        x1 = x2;
    end
```

```

x2 = M\N*x1+b;           %από αυτή τη γραμμή καταλαβαίνουμε ότι πρόκειται για τη Jacobi, διότι
x2 = M\N*x1+b = M^-1 * (N * x1 + b) = D^-1 * N * x1 + D^-1 * b = D^-1 * (L + U) * x1 + D^-1 * b
Κατ' αντιστοιχία έχουμε:
x^{k+1} = D^-1 * (L + U) * x^k + D^-1 * b
it = it +1;
end

```

Άσκηση 30 (Θέμα Φροντιστηρίου)

$$\text{Για το μητρώο } \begin{bmatrix} 1 & 0 & 0 \\ 2 & -1 & 0 \\ 1 & 2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 1 \\ 0 & -1 & 2 \\ 0 & 0 & 1 \end{bmatrix}$$

α) Να εξεταστεί αν θα συγκλίνει η Gauss – Seidel.

β) Αν διαπιστώσετε ότι συγκλίνει να εφαρμοστεί ένα βήμα της μεθόδου για αρχική προσέγγιση $x^{(0)} = \begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix}$ και δεξί μέλος

$$b = \begin{bmatrix} 4 \\ 7 \\ 7 \end{bmatrix}.$$

γ) Να κατασκευαστεί συνάρτηση σε Matlab, η οποία θα υλοποιεί την επαναληπτική μέθοδο Gauss-Seidel και παρουσιάζει τα ακόλουθα στοιχεία εισόδου - εξόδου:

Έισοδος

- x1: διάνυσμα αρχικής προσέγγισης
- A: Τετραγωνικό μητρώο συντελεστών
- b: δεξί μέλος
- maxit: Μέγιστο πλήθος επαναλήψεων
- tol: Βαθμωτός που δηλώνει την ανεκτικότητα της διαφοράς μεταξύ διαδοχικών προσεγγίσεων.

Έξοδος

- it: Το πλήθος των επαναλήψεων
- x2: Το διάνυσμα της τελικής προσέγγισης

Λύση

α) Το μητρώο $A = \begin{bmatrix} 1 & 0 & 0 \\ 2 & -1 & 0 \\ 1 & 2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 1 \\ 0 & -1 & 2 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 1 \\ 2 & 5 & 0 \\ 1 & 0 & 6 \end{bmatrix}$ προκύπτει από το γινόμενο ενός κάτω τριγωνικού μητρώου

με το ανάστροφό του, επομένως είναι ΣΤΟ.

β) Υπολογίζουμε τα μητρώα διάσπασης $M = D - L = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 5 & 0 \\ 1 & 0 & 6 \end{bmatrix}$ και $N - U = \begin{bmatrix} 0 & -2 & -1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$. Κάθε επόμενη προσέγγιση

προκύπτει από το επαναληπτικό σχήμα $M * x^{(k+1)} = N * x^k + b \Rightarrow \begin{bmatrix} 1 & 0 & 0 \\ 2 & 5 & 0 \\ 1 & 0 & 6 \end{bmatrix} * \begin{bmatrix} x_1^{(1)} \\ x_2^{(1)} \\ x_3^{(1)} \end{bmatrix} = \begin{bmatrix} 4 \\ 7 \\ 7 \end{bmatrix}$. Λύνοντας το γραμμικό σύστημα

$$\text{Θα βρούμε ότι το } x_1^{(1)} = \begin{bmatrix} 4 \\ -1/5 \\ 1/2 \end{bmatrix}.$$

γ) Η συνάρτηση σε Matlab είναι:

```

function [it,x2] = gauss_sei(x1, A, b, maxit, tol)
M = tril(A); N = -triu(A,1); n = length(x1); x2 = ones(n,1) * inf;
it = 0;
while (it < maxit) && (norm(x1-x2) > tol))
    if (it ~= 0)
        x1 = x2;
    end
    x2 = M\N*x1+b;
    it = it + 1;
end

```

```

end
x2 = M \ (N * x1 + b); %από αυτή τη γραμμή καταλαβαίνουμε ότι πρόκειται για τη Gauss - Seidel, διότι
x2 = M \ (N * x1 + b) = M-1 * (N * x1 + b) = M-1 * N * x1 + M-1 * b
Κατ' αντιστοιχία έχουμε:
xk+1 = diag(diag(A)) - (-tril(A, -1)) xk + D-1 * b = diag(diag(A)) + tril(A, -1) + D-1 * b = tril(A) + D-1 * b
it = it+1;
end

```

Άσκηση 31 (Θέμα Φροντιστηρίου)

a) Για το μητρώο $A = \begin{bmatrix} 3 & -0.5 & 0.2 \\ -0.5 & 1 & 0.1 \\ 0.2 & 0.1 & 7 \end{bmatrix}$ να εξηγηθεί αν είναι καλή ιδέα να χρησιμοποιηθεί η (απλή) μέθοδος Richardson, για τη προσέγγιση λύσης γραμμικού συστήματος με μητρώο συντελεστών το A. β) Να κατασκευαστεί συνάρτηση σε Matlab, η οποία θα υλοποιεί την παραμετροποιημένη επαναληπτική μέθοδο Richardson και παρουσιάζει τα ακόλουθα στοιχεία εισόδου-εξόδου:

Είσοδος

- x1: Διάνυσμα αρχικής προσέγγισης
- A: ΣΤΟ μητρώο συντελεστών
- b: Δεξί μέλος
- maxit: Μέγιστο πλήθος επαναλήψεων
- tol: Βαθμωτός που δηλώνει την ανεκτικότητα της διαφοράς μεταξύ διαδοχικών προσεγγίσεων
- w: Παράμετρος Richardson: Αν ο χρήστης δεν δώσει τη συγκεκριμένη τιμή, θα εκτελείται η απλή Richardson

Έξοδος

- it: Το πλήθος των επαναλήψεων
- x2: Το διάνυσμα της τελικής προσέγγισης

Λύση

α) Γνωρίζουμε ότι για να συγκλίνει η απλή Richardson για συμμετρικά μητρώα (όπως το A), θα πρέπει όλες οι ιδιοτιμές του μητρώου να είναι **θετικές και μικρότερες του 2**. Από τους δίσκους Gershgorin και πιο συγκεκριμένα από το δίσκο που προκύπτει από την τρίτη γραμμή του A, προκύπτει ότι υπάρχει ιδιοτιμή λ τέτοια ώστε $|\lambda - 7| < 0.7$. Το A είναι **συμμετρικό, συνεπώς όλες οι ιδιοτιμές του είναι πραγματικές, άρα μπορούμε να γράψουμε $6.7 < \lambda < 7.3$, συνεπώς υπάρχει ιδιοτιμή $\lambda > 2$ και η Richardson θα αποκλίνει.** Εναλλακτικά, μπορούμε να χρησιμοποιήσουμε την **παραμετροποιημένη Richardson**, χρησιμοποιώντας κατάλληλο βαθμωτό w. Η **παραμετροποιημένη Richardson** εφαρμόζει το ακόλουθο επαναληπτικό σχήμα: $x^{(k+1)} = x^{(k)} + (\hat{b} - \hat{A}) * x^{(k)}$ με $\hat{b} = \frac{1}{\omega}b$, $\hat{A} = \frac{1}{\omega}A$.

```

function [it, x2] = rich(x1, A, b, maxit, tol, w)
if nargin < 6
    w = 1;
end
A = A * 1/w; b = b * 1/w; %από αυτή τη γραμμή υπολογίζει το  $\hat{b}$  και  $\hat{A}$ 
n = length(x1);
x2 = ones(n,1) * inf;
it = 0;
while ((it < maxit) && (norm(x1-x2) > tol))
    if (it ~= 0)
        x1 = x2;
    end
    r = b - A * x1; x2 = x1 + r;
    it = it+1;
end

```

Άσκηση 32 (Θέμα Φροντιστηρίου)

Εξηγήστε τη λειτουργία της συνάρτησης fun2.m

```

function est=fun2(A, k)
L= chol(A)          % υπολογισμός άνω τριγωνικού παράγοντα Cholesky
n = size(A,1)       % Στο n επιστρέφεται η πρώτη διάσταση του A, δηλ. ο αριθμός γραμμών
x = ones(n,1)        %
for i = 1:k
    x=x/norm(x);    % κανονικοποίηση διανύσματος x, διαίρεση κάθε στοιχείου του x με το
    % μήκος του, δηλ. με τη δεύτερη νόρμα του
    x = L'\(L\x);   % επίλυση γραμμικού συστήματος με Cholesky
    x =x/norm(x);    % κανονικοποίηση διανύσματος x
end
est = x' * A * x    % υπολογισμός βαθμωτού est και έλεγχος θετικής ορισμότητας

```

Λύση

Ελέγχει εν' τέλει αν το συμμετρικό μητρώο A είναι Θ.Ο. με το 4^o κριτήριο θετικής ορισμότητας.

Άσκηση 33 (Θέμα Φροντιστηρίου)

Έστω ο ακόλουθος κώδικας:

```

A = rand(20), A = A + A',      % Δημιουργία μητρώου 20x20 με τυχαίες τιμές
t = eig(A),                    % Εύρεση ιδιοτιμών μητρώου A, στο διάνυσμα t.
h = length(find(t > 0))      % Εύρεση πλήθους στοιχείων διανύσματος t που είναι θετικά
                                % δηλ. εύρεση θετικών ιδιοτιμών
if(h == 19)                   % αν το πλήθος των θετικών ιδιοτιμών είναι ίσο με 19
[~,D,~] = ldl(A)            % ldl παραγοντοποίηση μητρώου A
end

```

Τι δομή περιμένετε να έχει το μητρώο D;

Λύση

Επειδή είναι αποτέλεσμα ldl παραγοντοποίησης, αναμένεται ότι το μητρώο D θα είναι διαγώνιο. Γνωρίζουμε ότι για κάθε συμμετρικό μητρώο ($A = A^T$), υπάρχει παραγοντοποίηση $PAP^T = LDL^T$, όπου P = μητρώο μετάθεσης, L = κάτω τριγωνικό μητρώο με "1" στη διαγώνιο και D = μπλοκ διαγώνιο μητρώο με μπλοκ 1 x 1 (βαθμωτοί) ή 2 x 2 (μητρώα συμμετρικά) στην κύρια διαγώνιο.

Άσκηση 34 (Θέμα Φροντιστηρίου)

Εξηγήστε τη λειτουργία της συνάρτησης fun1.m:

```

function[d]= fun1(A, b)
[j, i]=size(A)      % j = γραμμές μητρώου A και i = στήλες μητρώου A
k = rank(A)          % k = τάξη μητρώου A (πλήθος μη μηδενικών οδηγών, πλήθος γρ.
                      % ανεξάρτητων γραμμών/στηλών, πλήθος μη μηδενικών ιδιοτιμών, πλήθος μη
                      % μηδενικών ιδιαζουσών τιμών)
if (j>= i) && (k == i) % αν έχει περισσότερες γραμμές ≥ στήλες και τάξη = 3
    [W, H] = qr(A,0)  % οικονομική QR παραγοντοποίηση μητρώου A, στο W επιστρέφεται
                      % μητρώο ορθογώνιο και στο H μητρώο άνω τριγωνικό
    y = H\ (W' * b); % επίλυση γρ. συστήματος A * y = b μέσω QR παραγοντοποίησης
                      % όπου σε ένα βήμα έχουν συγχωνευθεί η εμπρός και η πίσω αντικ.
                      % κατ' αντιστοιχία με τον τύπο x=R\ (Q^T * b)
    f = H\ (H'\ A' * b) % επίλυση γρ. συστήματος A * f = b κατ' αντιστοιχία με τον τύπο
                          % x=R\ (R'\ A' * b)
    d = norm(y - f);
else
    error('stop');
end

```

Λύση

```

>> [d] = fun1(A, b)
j =3
i =3
k =3
W =
-0.9733 0.0149 0.2289
-0.1622 -0.7502 -0.6410
0.1622 -0.6611 0.7326
H =
-6.1644 -2.1089 -1.9467
0 -3.5430 -2.5105
0 0 4.9908
f =
0.0917
-0.2569
0.3211
d =6.9389e-17.

```

Υπολογίζει τη λύση του γραμμικού συστήματος $A * x = b$ μέσω QR παραγοντοποίησης, με δύο τρόπους: Στον πρώτο τρόπο η λύση είναι $y = R \setminus (Q' * b)$ και στο δεύτερο τρόπο η λύση είναι $f = R \setminus (R' \setminus A' * b)$. Επειδή οι δύο λύσεις είναι πανομοιότυπες μεταξύ τους, η διαφορά τους που εκφράζεται με το διάνυσμα d , έχει τιμή κοντά στο «0».

Άσκηση 34 (Θέμα Φροντιστηρίου)

a) Να ερμηνεύσετε τη λειτουργία της συνάρτησης qr_solve. Τι παρατηρείτε για τις επιστρεφόμενες τιμές;

```

function [rqr, rch, r] = qr_solve(A, b)
[m, n] = size(A) % m, n επιστρέφονται οι διαστάσεις του A
bhat = A' * b % bhat = A' * b
[Q, R]= qr(A) %Πλήρης (κανονική) QR παραγοντοποίηση
Q1 = Q(:, 1:n); %Διάσπαση του μητρώου Q σε δύο υπομητρώα Q1 και Q2
Q2 = Q(:, n+1:m) %Το μητρώο Q1 περιέχει τις πρώτες n στήλες του Q και το Q2 τις υπόλοιπες
R1 = R(1:n,:); %Παίρνουμε το μειωμένο μητρώο R1
xqr = R1 \ (Q1' * b) %Λύση του γραμμικού συστήματος μέσω QR παραγοντοποίησης
rqr = norm(A * xqr - b) %Το residual της λύσης του γραμμικού συστήματος με τον α' τρόπο.
xch = R1 \ (R1' \ bhat) %Λύση του γραμμικού συστήματος μέσω παραγοντοποίησης Cholesky
rch = norm(A*xch-b) %Το residual της λύσης του γραμμικού συστήματος με τον β' τρόπο.
r = norm(Q2' * b)

```

Λύση

```
>> A=[1 -2;1 -1;1 0;1 1]
```

```
A =
1 -2
1 -1
1 0
1 1
```

```
>> b=[1 0 -1 2]'
```

```
b =
1
0
-1
2
```

```
>> [rqr, rch, r] = qr_solve(A,b)
```

```
m = 4, n = 2
```

```
bhat =
2
0
```

Q =
 -0.5000 -0.6708 -0.4236 -0.3472
 -0.5000 -0.2236 0.3060 0.7787
 -0.5000 0.2236 0.6588 -0.5157
 -0.5000 0.6708 -0.5412 0.0843

R =
 -2.0000 1.0000
 0 2.2361
 0 0
 0 0

Q1 =
 -0.5000 -0.6708
 -0.5000 -0.2236
 -0.5000 0.2236
 -0.5000 0.6708

Q2 =
 -0.4236 -0.3472
 0.3060 0.7787
 0.6588 -0.5157
 -0.5412 0.0843

R1 =
 -2.0000 1.0000
 0 2.2361

xqr =
 0.6000
 0.2000

rqr = 2.1909

xch =
 0.6000
 0.2000

rch = 2.1909

r = 2.1909

rqr = 2.1909

rch = 2.1909

r = 2.1909.

Από το γεγονός ότι $rqr = rch = 2.1909$ συμπεραίνουμε ότι η επίλυση του γρ. συστήματος μέσω QR παραγοντοποίη-

σης αλλά και μέσω παραγοντοποίησης Cholesky οδήγησε στο ίδιο αποτέλεσμα, αφού ως γνωστό $A = [Q_1 \ Q_2] * \begin{bmatrix} R_1 \\ 0 \end{bmatrix}$

$= Q_1 * R_1$, όπου το $Q_1 \in R^{m \times n}$ έχει ορθοκανονικές στήλες και το $R_1 \in R^{n \times n}$ είναι άνω τριγωνικό. Οι στήλες του μητρώου Q_1

αποτελούν ορθοκανονική βάση του χώρου στηλών του μητρώου A. Επίσης το μητρώο R_1 περιέχει τον άνω παράγοντα

Cholesky του $A^T * A$, διότι ισχύει ότι: $A^T * A = R_1^T * [Q_1 \ Q_2] * R_1 = R_1^T * R_1$.¹¹ Δηλαδή το R_1 είναι ο παράγοντας Cholesky του

Σ.Θ.Ο. μητρώου $A^T * A$, δηλ. $[A^T * A = R_1^T * R_1]$. Με χρήση της οικονομικής QR έχουμε -όπως προαναφέρθηκε- ότι $A = Q * R =$

$[Q_1 \ Q_2] * \begin{bmatrix} R_1 \\ 0 \end{bmatrix} = Q_1 * R_1$ και $A * x = b \Rightarrow A^T * A * x = A^T * b \Rightarrow x = (A^T * A)^{-1} * A^T * b \Rightarrow x = (R_1^T * R_1)^{-1} * A^T * b \Rightarrow x = (R_1)^{-1} *$

$(R_1^T)^{-1} * A^T * b \Rightarrow x = (R_1)^{-1} * (R_1)^{-1} * A^T * b \Rightarrow x = (R_1)^{-1} * (R_1) \backslash (A^T * b) \Rightarrow x = (R_1) \backslash ((R_1) \backslash (A^T * b))$. Κανονικά για την παραγοντοποίηση Cholesky ισχύει: $A = L * L^T = R^T * R$.

¹¹ Υπενθυμίζεται ότι η παραγοντοποίηση Cholesky του ΣΘΟ μητρώου A είναι: $A = L * L^T = R^T * R$, όπου το μητρώο R ονομάζεται παράγοντας Cholesky του μητρώου A.

Άσκηση 35 (Θέμα Φροντιστηρίου)

α) Θεωρώντας τη συνάρτηση $f(x) = x^7 - x^6 + 1$ και τις αρχικές συνθήκες του παρακάτω πίνακα, να βρεθεί το πολυώνυμο παρεμβολής σε μορφή Hermite:

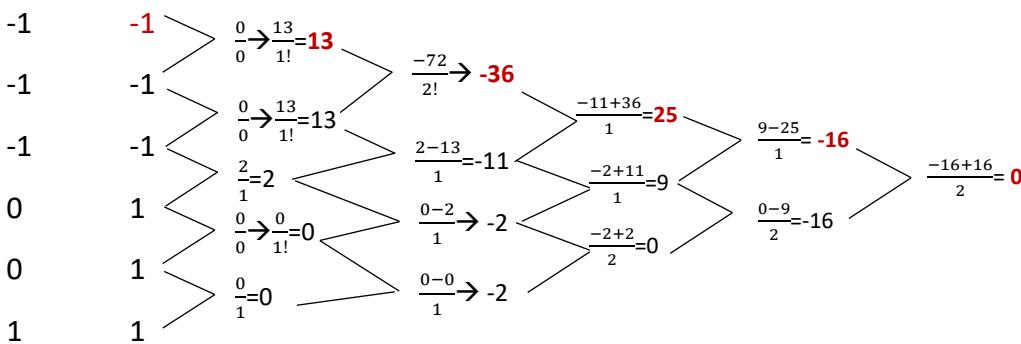
x_i	-1	0	1
$f(x_i)$	-1	1	1
$f'(x_i)$	13	0	-
$f''(x_i)$	-72	-	-

β) Να κατασκευαστεί script σε MATLAB που να υπολογίζει και να εκτυπώνει τις παρακάτω ποσότητες:

- Τις τιμές της συνάρτησης f στα σημεία $h = -1: 0.02: 1$.
- Τις τιμές του πολυωνύμου που προκύπτει από το ερώτημα (α) f στα σημεία h .
- Τις τιμές του πολυωνύμου στα σημεία h που προκύπτει από την polyfit, για 6 ίσα κατανεμημένους κόμβους στο διάστημα $[-1, 1]$.

Λύση

α) Κατασκευάζουμε τον πίνακα Δ.Δ. για τις δοθείσες τιμές:



Άρα το π.π. είναι $P(x) = f(x_0) + f(x_0, x_1) * (x - x_0) + f(x_0, x_1, x_2) * (x - x_0) * (x - x_1) + f(x_0, x_1, x_2, x_3) * (x - x_0) * (x - x_1) * (x - x_2) + f(x_0, x_1, x_2, x_3, x_4) * (x - x_0) * (x - x_1) * (x - x_2) * (x - x_3) = -1 + 13 * (x+1) + (-36) * (x+1)^2 + 25 * (x+1)^3 + (-16) * (x+1)^4 + 0 * (x+1)^5 + 0 * (x+1)^6 * x^2$.

β) Η ζητούμενη συνάρτηση σε MATLAB είναι:

```

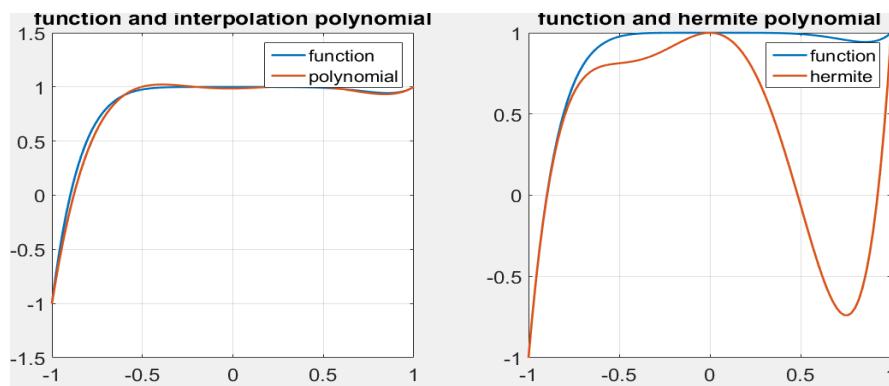
x = -1:1; v = (max(x)-min(x))/5; %v = 0.4
xp = min(x):v:max(x); %xp = -1.0000 -0.6000 -0.2000 0.2000 0.6000 1.0000
h = -1:0.02:1; yf = h.^7-h.^6+1;
%interpolation for 6 equal nodes in [-1,1]
y = xp.^7-xp.^6+1; p = polyfit(xp, y, length(xp)-1); yp = polyval(p,h);
%hermite interpolation
yher = -1 +13.* (h+1)-36.* (h+1).^2+25.* (h+1).^3-16.* (h+1).^4+6.* (h+1).^5.* (h).^2;
set(0, 'DefaultAxesFontSize', 18);
set(0, 'DefaultLineWidth', 2);
figure; %δημιουργία παραθύρου για σχεδίαση
subplot(1,2,1); %σχεδιάζουμε σε 1 γραμμή και 2 στήλες, στο αριστερό παράθυρο
plot(h, yf); %γραφική παράσταση διανυσμάτων h και yf
hold on; %διατήρηση παραθύρου σχεδίασης στην οθόνη
plot (h ,yp); %γραφική παράσταση διανυσμάτων h και yp
grid on; %εμφάνιση πλέγματος
title('function and interpolation polynomial'); %δημιουργία τίτλου γραφήματος
legend('function','polynomial'); %δημιουργία υπομνήματος
subplot(1,2,2); %σχεδιάζουμε σε 1 γραμμή και 2 στήλες, στο δεξιό παράθυρο
plot (h,yf); %γραφική παράσταση διανυσμάτων h και yf
hold on; %διατήρηση παραθύρου σχεδίασης στην οθόνη
plot(h, yher); %γραφική παράσταση διανυσμάτων h και yf

```

```

grid on; %εμφάνιση πλέγματος
title('function and hermite polynomial'); %δημιουργία τίτλου γραφήματος
legend('function', 'hermite'); %δημιουργία υπομνήματος
hold off;

```



Παρατήρηση 1: η συνάρτηση $p = \text{polyfit}(X, Y, N)$ υπολογίζει τους συντελεστές του πολυωνύμου $p(x)$ βαθμού N που ταιριάζει (προσαρμόζει) τα δεδομένα Y στο διάνυσμα γραμμής p μήκους $n + 1$ που περιέχει τους συντελεστές του πολυωνύμου με φθίνουσα σειρά: $p(1) * x^n + p(2) * x^{n-1} + \dots + p(n) * x + p(n+1)$.

Παρατήρηση 2: η συνάρτηση $p = \text{polyval}(p, x)$ ανάλογα με το αν το x είναι σημείο, διάνυσμα ή μητρώο, τότε επιστρέφεται η τιμή του πολυωνύμου p στο y , το οποίο είναι σημείο, διάνυσμα ή μητρώο αντίστοιχα, δηλαδή: $y = p(1) * x^n + p(2) * x^{n-1} + \dots + p(n) * x + p(n + 1)$, π.χ. αν το $>>p = [1 \ 2 \ 3 \ 4]$, τότε η εντολή: $>>y = \text{polyval}(p, 1)$ επιστρέφει το αποτέλεσμα: $>>y = 10$, διότι $y = p(1)x^3 + p(2)x^2 + p(3)x + p(4) = 1 * 1^3 + 2 * 1^2 + 3 * 1^1 + 4 * 1 = 10$, δηλαδή επιστρέφεται η τιμή του πολυωνύμου για $x = 1$.

Άσκηση 36 (Θέμα Φροντιστηρίου)

Έστω μοντέλο αναπαράστασης αριθμών κινητής υποδιαστολής με μήκος ουράς $t = 3$ bits και εύρος εκθέτη $-2 : 3$. Για το συγκεκριμένο μοντέλο να βρεθούν το **realmax**, το **realmin** και το **eps**. Να γραφτεί script σε MATLAB που να εμφανίζει όλους τους κανονικοποιημένους α.κ.υ. Πόσοι είναι οι δυνατοί α.κ.υ. που μπορούν να αναπαρασταθούν;

Λύση

a) Στο συγκεκριμένο μοντέλο οι **κανονικοποιημένοι α.κ.υ. έχουν τη μορφή: $\pm (1.d_1d_2d_3) \times 2^e$** . Συνεπώς το **realmax** = $1.111 \times 2^3 = (1 \times 2^0 + 1 \times 2^{-1} + 1 \times 2^{-2} + 1 \times 2^{-3}) \times 2^3 = (1 + 0.5 + 0.25 + 0.125) \times 2^3 = 1.875 \times 2^3 = 15$. Ο **realmin** = $1.000 \times 2^{-2} = (1 \times 2^0) \times 2^{-2} = 1 \times 2^{-2} = 0.25$. Το **eps** = $1.001 - 1.000 = 2^{-3} = 0.125$.

b) Για τη **συγγραφή script σε MATLAB που να εμφανίζει όλους τους κανονικοποιημένους α.κ.υ.** Θα πρέπει να αναφερθεί ότι πρόκειται για επαναληπτική δομή για την ουρά, που ξεκινά από το $1.000 = 1$ και καταλήγει στο $1.111 = 1.875$ με βήμα 0.125 (όσο είναι το **eps**).

```

clc
x = [] ; %κενό μητρώο ή διάνυσμα
for i = 1: 0.125: 1.8750
    for j=-2:3
        x = [x i*2^j];
    end
end
x=[-x 0 x]; x=sort(x)

```

Οσον αφορά τους δυνατούς α.κ.υ. που μπορούν να παρασταθούν, το πλήθος τους θα προκύψει από το γινόμενο πλήθος

τιμών εκθέτη $x 2^{\text{αριθμός bits ουράς}} \times 2 \text{ (θετικοί - αρνητικοί)} + 1 \text{ (αναπαράσταση 0)}$. Στην προκειμένη περίπτωση ο εκθέτης μπορεί να λάβει 6 τιμές στο διάστημα [-2 3]. Επίσης, έχουμε 3 bits στην ουρά, συνεπώς 8 δυνατές αναπαραστάσεις. Άρα προκύπτουν: 6×8 αναπαραστάσεις για τους θετικούς και άλλες τόσες για τους αρνητικούς και αν επιλέξουμε και μια αναπαράσταση για το «0», θα έχουμε συνολικά 97 αναπαραστάσεις ($48 + 48 + 1$).

Άσκηση 37 (Θέμα Φροντιστηρίου)

Να κατασκευαστεί συνάρτηση σε MATLAB που θα δέχεται ως όρισμα εισόδου ένα πραγματικό μητρώο A και θα επιστρέφει ένα διάνυσμα d, το οποίο σε κάθε θέση του θα περιέχει τις **άνω αριστερές υποορίζουσες του μητρώου $A^T * A$** .

Λύση

Η ζητούμενη συνάρτηση είναι η εξής:

```
function [d] = dets(A)
    n = size(A, 2) %επιστρέφεται στη μεταβλητή n μόνο ο αριθμός στηλών του A
    d = zeros(n, 1)%αρχικό περιεχόμενο διανύσματος d με μηδενικά
    Q = A' * A
    for i = 1:n
        d(i) = det(Q(1:i, 1:i))
    end
```