ΣΗΜΕΙΩΣΕΙΣ
ΑΡΙΘΜΗΤΙΚΗΣ ΓΡΑΜΜΙΚΗΣ ΑΛΓΕΒΡΑΣ

Β. ΔΟΥΓΑΛΗΣ, Δ. ΝΟΥΤΣΟΣ, Α. ΧΑΤΖΗΔΗΜΟΣ

Βόλος, Αύγουστος 2007

Περιεχόμενα

Πρόλογος

Η Αριθμητική Γραμμική Αλγεβρα καλύπτει έναν από τους σημαντικότερους και μεγαλύτερους κλάδους της Αριθμητικής Ανάλυσης τόσο από την άποψη της θεωρίας όσο και από την άποψη των εφαρμογών της. Ασχολείται κυρίως με την αριθμητική επίλυση γραμμικών συστημάτων αλγεβρικών εξισώσεων καθώς και με την αριθμητική εύρεση των ιδιοτιμών πινάκων.

Η σημασία και η σπουδαιότητα της Αριθμητικής Γραμμικής Αλγεβρας στην πράξη δεν είναι τυχαία. Αυτό δε γιατί έχει εκτιμηθεί ότι το 70% των προβλημάτων της Επιστήμης και της Τεχνολογίας που καταλήγουν για επίλυση στον Ηλεκτρονικό Υπολογιστή είναι γραμμικά συστήματα. Το τελευταίο δικαιολογείται από το γεγονός ότι, εκτός από τα προβλήματα που είναι από τη φύση τους γραμμικά, τα απλούστερα μοντέλα για την περιγραφή των διαφόρων φαινομένων είναι τα γραμμικά. Οπως είναι γνωστό, ένα φαινόμενο, στο οποίο υπάρχουν μεταβολές φυσικών ποσοτήτων που εξαρτιούνται από άλλες, περιγράφεται συνήθως στα μαθηματικά από μια διαφορική εξίσωση. Για τη μελέτη της και την εν συνεχεία επίλυσή της στον Υπολογιστή απαιτείται προηγουμένως η διακριτοποίησή της. Αν η διαφορική εξίσωση είναι γραμμική είναι επόμενο το διακριτό ανάλογό της να είναι γραμμικό (σύστημα). Αν δεν είναι γραμμική μια καλή πρώτη προσέγγισή της αποτελεί μια γραμμικοποίησή της που οδηγεί τελικά, με διακριτοποίηση, σε γραμμικό σύστημα.

Η Αριθμητική Γραμμική Αλγεβρα, όπως εξάλλου και η Αριθμητική Ανάλυση, ασχολείται γενικά με την ανάπτυξη και μελέτη (αριθμητικών) αλγόριθμων, δηλαδή αλγόριθμων που περιέχουν τις τέσσερις βασικές πράξεις της Αριθμητικής και μόνο, για την εύρεση αριθμητικών αποτελεσμάτων από αριθμητικά δεδομένα στα γραμμικά προβλήματα που αναφέρθηκαν προηγουμένως. Οι αριθμητικοί αλγόριθμοι για να είναι αποτελεσματικοί στην πράξη θα πρέπει να καταλήγουν στο επιδιωκόμενο αποτέλεσμα με τις λιγότερες δυνατές πράξεις ή, με όρους Υπολογιστή, στο μικρότερο δυνατό χρόνο. Ακόμη, το όποιο αριθμητικό αποτέλεσμα, που παίρνεται από τον Υπολογιστή, θα πρέπει να είναι όσο το δυνατόν πιο ακριβές για να είναι αποδεκτό και άρα χρήσιμο. Η χρήση όμως του Υπολογιστή, πέρα από οποιαδήποτε άλλα σφάλματα, που είναι δυνατόν να ενυπάρχουν στα αριθμητικά δεδομένα ενός προβλήματος, όπως π.χ. τα γνωστά σφάλματα αποκοπής, εισάγει πάντα και τα αναπόφευκτα σφάλματα στρογγύλευσης κατά την αποθήκευση των δεδομένων του προβλήματος καθώς και αυτά που δημιουργούνται κατά την εκτέλεση των ενδιάμεσων πράξεων και την αποθήκευση των αποτελεσμάτων τους. Η παρουσία όλων αυτών των σφαλμάτων, που συσσωρεύονται και μεταδίδονται κατά τη διάρκεια των υπολογισμών, έχει ως άμεση συνέπεια τα τελικά αποτελέσματα που παίρνονται από τον Υπολογιστή να μην είναι ακριβή. Αποτελούν δηλαδή μια προσέγγιση της λύσης του προβλήματος. Επόμενο, λοιπόν, είναι ένα μεγάλο μέρος της θεωρίας της Αριθμητικής Γραμμικής Αλγεβρας, όπως και της Αριθμητικής Ανάλυσης, να αφιερώνεται στην ανάπτυξη και κατασκευή μεθόδων, αλγόριθμων, που θα περιορίζουν κατά το δυνατόν τα σφάλματα στα τελικά αποτελέσματα, έτσι ώστε τα αποτελέσματα αυτά να γίνονται αποδεκτά. Η εύρεση της απόκλισης των τελευταίων από τα αντίστοιχα άγνωστα, που αποτελούν τη θεωρητική λύση του προβλήματος, δηλαδή η εύρεση των τελικών σφαλμάτων, καθώς επίσης και η μελέτη και ο περιορισμός τους, αποτελεί επίσης μία από τις πιο σημαντικές συνιστώσες των παρουσών Σημειώσεων.

Από τη φύση των προβλημάτων που μελετιούνται στην Αριθμητική Γραμμική Αλγεβρα, είναι επόμενο πέρα από αριθμούς (βαθμωτά μεγέθη) να εμπλέκονται διανύσματα και πίνακες. Ως εκ τούτου οι όποιες ιδιότητες των πινάκων θα πρέπει να μελετιούνται σε βάθος αφού η δυνατότητα εκμετάλλευσής τους σε μια συγκεκριμένη περίπτωση είναι δυνατόν να αποτελεί αποφασιστικό παράγοντα αφενός μεν στην ανάπτυξη αποτελεσματικών αλγόριθμων αφετέρου δε στην επιλογή του καταλληλότερου μεταξύ περισσότερων του ενός διαθέσιμων αλγόριθμων. Εχοντας εξάλλου υπόψη όσα ήδη αναφέρθηκαν, σχετικά με την ανάπτυξη αποτελεσματικών αλγόριθμων καθώς επίσης και με τη μελέτη και μέτρηση των όποιων σφαλμάτων στα τελικά αποτελέσματα, είναι επόμενο οι αντίστοιχες μετρήσεις να απαιτούν τη χρησιμοποίηση, όχι μόνο απόλυτων σφαλμάτων αριθμών αλλά και, απόλυτων σφαλμάτων διανυσμάτων και πινάκων. Ετσι, η εισαγωγή και η εν συνεχεία χρήση της έννοιας της norm καθίσταται εκ των ων ουκ άνευ.

Ως εκ τούτου στα επόμενα Κεφάλαια θα επιδιωχτεί πρώτα μια σύντομη εισαγωγή σε στοιχεία που έχουν σχέση με διανύσματα και πίνακες πριν γίνει οποιαδήποτε αναφορά στην ανάπτυξη και μελέτη αλγόριθμων, για την επίλυση γραμμικών συστημάτων, η οποία κυρίως θα καλυφτεί. Η αριθμητική εύρεση των ιδιοτιμών τετραγωνικού πίνακα, αν και η εισαγωγή των τελευταίων με ό,τι αυτή συνεπάγεται θεωρητικά θα μας απασχολεί συνεχώς από το Κεφάλαιο 1, θα καλυφτεί συνοπτικά στο Κεφάλαιο 8. Οσες νέες ή γνωστές και χρήσιμες έννοιες καθώς και προτάσεις από τη Γραμμική Αλγεβρα απαιτούνται θα εισάγονται ή θα αναφέρονται όταν και όπου αυτές θα χρειάζονται για πρώτη φορά.

Οι ανά χείρας Σημειώσεις αποτελούν μια ένωση των υλών των μαθημάτων της Αριθμητικής Γραμμικής Αλγεβρας, που διδάχτηκαν κατά τα τέσσερα τελευταία ακαδημαϊκά έτη στο Πανεπιστήμιο Κρήτης σε προπτυχιακό και μεταπτυχιακό επίπεδο και στο Πανεπιστήμιο Ιωαννίνων. Βασίστηκαν κυρίως στις Σημειώσεις του πρώτου συγγραφέα, που αποτέλεσαν κύριο βοήθημα σε αντίστοιχα μαθήματα τα τελευταία 16 χρόνια στο Πανεπιστήμιο Κρήτης, στο Εθνικό Μετσόβιο Πολυτεχνείο και στο Πανεπιστήμιο Αθηνών, σε Σημειώσεις του δεύτερου συγγραφέα, που χρησιμοποιήθηκαν τα τελευταία χρόνια στό Πανεπιστήμιο Ιωαννίνων καθώς και σε Σημειώσεις του τρίτου συγγραφέα, που χρησιμοποιήθηκαν τα τελευταία τέσσερα χρόνια στο Πανεπιστήμιο Κρήτης, τα προηγούμενα 10 στο Πανεπιστήμιο Purdue των Η.Π.Α. και τα πριν από αυτά 15 χρόνια στο Πανεπιστήμιο Ιωαννίνων.

Τέλος, θα θέλαμε να ευχαριστήσουμε ειλικρινά τους κατά καιρούς μαθητές μας, που με τις εύστοχες ερωτήσεις, απορίες και παρατηρήσεις τους, μας βοήθησαν να καταλήξουμε καταρχάς στις Σημειώσεις αυτές, που υποβάλλονται στην κρίση των χρηστών και αναγνωστών μας. Θα ήταν μεγάλη παράλειψη από μέρους μας αν δεν ευχαριστούσαμε ολόθερμα και τον αγαπητό φίλο και συνάδελφο Γιώργο Ακρίβη, Καθηγητή του Πανεπιστημίου Ιωαννίνων, που είχε την καλοσύνη να χρησιμοποιήσει και να διεξέρθει μεγάλο μέρος των Σημειώσεων και να κάνει εύστοχες παρατηρήσεις, όχι μόνο σε σφάλματα και παραλείψεις τους, αλλά και σε σφάλματα ουσίας. Οποιεσδήποτε παρατηρήσεις από μέρους των αναγνωστών μας, που θα βοηθήσουν στη βελτίωση των παρουσών Σημειώσεων, θα είναι με χαρά μας ευπρόσδεκτες.

Ηράκλειο, Ιούνιος 2000
Βασίλειος Δουγαλής, Δημήτριος Νούτσος, Απόστολος Χατζηδήμος

Πρόλογος 2ης Εκδοσης

Η παρούσα έκδοση διαφέρει από την προηγούμενη στο ότι πολλά από τα λάθη, σφάλματα και παραλείψεις της εντοπίστηκαν και διορθώθηκαν. Αυτό έγινε κυρίως κατά τη διάρκεια της διδασκαλίας των μαθημάτων "Αριθμητική Γραμμική Αλγεβρα" και "Θέματα Αριθμητικής Ανάλυσης" κατά τα δύο εξάμηνα του λήγοντος ακαδημαϊκού έτους στα Πανεπιστήμια Αθηνών, Ιωαννίνων, Κρήτης και Κύπρου. Από τη θέση αυτή θεωρούμε υποχρέωσή μας να ευχαριστήσουμε θερμότατα τους φοιτητές που παρακολούθησαν τα αντίστοιχα μαθήματα και οι οποίοι με τις ερωτήσεις και τις υποδείξεις τους συνέβαλαν αποφασιστικά στην όλη παρουσίαση του νέου κειμένου.

Επειδή εξακολουθούμε να πιστεύουμε ότι και η νέα έκδοση δεν μπορεί να είναι απαλλαγμένη λαθών, σφαλμάτων και παραλείψεων, παρακαλούμε τους αναγνώστες της να μας βοηθήσουν με την οποιοδήποτε τρόπο συμβολή τους για μια ακόμη καλύτερη μελλοντική παρουσίαση των ανά χείρας Σημειώσεων.

Ηράκλειο, Ιούνιος 2001
Βασίλειος Δουγαλής, Δημήτριος Νούτσος, Απόστολος Χατζηδήμος

Πρόλογος 3ης Εκδοσης

Στην παρούσα έκδοση, πέρα από διάφορες διορθώσεις που αποτελούν συνήθη πρακτική μιας οποιασδήποτε νεώτερης έκδοσης, έχει γίνει μια μικρή αναδιάταξη της ύλης με διάφορες συμπληρώσεις για την καλύτερη κατανόηση του κειμένου. Η νέα έκδοση έχει επίσης εμπλουτιστεί με μια ποικιλία επιλεγμένων ασκήσεων, που δίνονται στο τέλος κάθε κεφαλαίου, για την ακόμη καλύτερη εμπέδωση της νεοεισαχθείσας ύλης και των περιλαμβανόμενων σ΄ αυτήν εννοιών καθώς και της σύνδεσής τους με τις αντίστοιχες προηγούμενες.

Θα ήταν μεγάλη παράλειψή μας αν δεν ευχαριστούσαμε θερμότατα όλους που μας συμπαραστάθηκαν με τον οποιοδήποτε τρόπο και συνέβαλαν έμμεσα στη διαμόρφωση της νέας έκδοσης.

Ηράκλειο, Οκτώβρης 2003
Βασίλειος Δουγαλής, Δημήτριος Νούτσος, Απόστολος Χατζηδήμος

Πρόλογος 4ης Εκδοσης

Η παρούσα έκδοση αποτελεί μια ακόμη βελτιωμένη εκδοχή της προηγούμενης, όπου, για μια ακόμη φορά, νέα εισχωρήσαντα σφάλματα και παραλείψεις διορθώθηκαν, ιδιαίτερα σε ό,τι αφορά τις Ασκήσεις που προστέθηκαν στην προηγούμενη έκδοση.

Θερμότατες ευχαριστίες απευθύνουμε σε συνάδελφους και φοιτητές που συνέβαλαν στην νέα μας προσπάθεια.

Ηράκλειο, Μάης 2004
Βασίλειος Δουγαλής, Δημήτριος Νούτσος, Απόστολος Χατζηδήμος

Πρόλογος 5ης Εκδοσης

Η νέα έκδοση είναι μια ακόμη βελτιωμένη εκδοχή της προηγούμενης. Διάφορα σφάλματα εντοπίστηκαν και μέρη του κειμένου διορθώθηκαν για την καλύτερη κατανόησή του.

Για μια ακόμη φορά ευχαριστούμε θερμότατα συνάδελφους και φοιτητές που συνέβαλαν με οποιοδήποτε τρόπο στη νέα, και ίσως όχι τελευταία, προσπάθεια.

Ηράκλειο, Ιούνιος 2006
Βασίλειος Δουγαλής, Δημήτριος Νούτσος, Απόστολος Χατζηδήμος

Πρόλογος 6ης Εκδοσης

Η νέα έκδοση διαφέρει από την προηγούμενη στο ότι κάποιες υποδείξεις από τον αγαπητό φίλο και συνάδελφο Γιώργο Ακρίβη, Καθηγητή του Πανεπιστημίου Ιωαννίνων, τον οποίο και ευχαριστούμε θερμά για ακόμη μια φορά, λήφτηκαν υπόψη στις ελάχιστες νέες διορθώσεις.

Βόλος, Αύγουστος 2007
Βασίλειος Δουγαλής, Δημήτριος Νούτσος, Απόστολος Χατζηδήμος

1 Βασική Θεωρία

1.1 Διανύσματα

Από τη Γραμμική Αλγεβρα θεωρούμε γνωστές μερικές από τις βασικότερες έννοιές της. Ιδιαίτερα θεωρούνται γνωστά τα περί: "Διανυσματικών Χώρων", "Γραμμικών Διανυσματικών Χώρων", "Γραμμικής Ανεξαρτησίας (Εξάρτησης) Διανυσμάτων", "Γραμμικού Συνδυασμού (Διανυσμάτων)", "Διανυσμάτων που παράγουν ένα Γραμμικό Διανυσματικό Χώρο", "Βάσης Γραμμικού Διανυσματικού Χώρου", και "Διάστασης Γραμμικού Διανυσματικού Χώρου". Ειδικότερα θα μας απασχολήσουν οι Γραμμικοί Διανυσματικοί Χώροι $\mathbb{C}^n$ και $\mathbb{R}^n$ για τα διανύσματα των οποίων θα χρησιμοποιούνται οι παρακάτω συμβολισμοί:

$$x := \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \equiv \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \quad x_i \in \mathbb{C}\ (\mathbb{R}),\ i = 1(1)n.$$

$x^T = [x_1\ x_2\ \cdots\ x_n]$, ανάστροφο του x. Διαβάζεται και $x-$ανάστροφο.
$x^H = [\overline{x}_1\ \overline{x}_2\ \cdots\ \overline{x}_n]$, συζυγές μιγαδικό ανάστροφο του x, όπου $\overline{x}_i$ είναι ο συζυγής μιγαδικός του $x_i,\ i = 1(1)n$. Διαβάζεται και $x-$Ερμιτιανό.

1.1.1 Ευκλείδειο Εσωτερικό Γινόμενο

Ορισμός 1.1 *Το Ευκλείδειο εσωτερικό γινόμενο δύο διανυσμάτων $x, y \in \mathbb{C}^n$ με τη σειρά αυτή συμβολίζεται με $(x, y)_2$ και ορίζεται ως*

$$(x, y)_2 = \sum_{i=1}^{n} \overline{x}_i y_i.$$

Σημείωση: Είναι φανερό από τον ορισμό ότι $(y, x)_2 = \overline{(x, y)_2}$.

1.1.2 Norms (Νόρμες, Στάθμες) Διανυσμάτων

Είναι γνωστό ότι το μέγεθος ενός πραγματικού αριθμού δίνεται από την απόλυτη τιμή του ενώ ενός μιγαδικού επίσης από την απόλυτη τιμή (ή μέτρο) του. Το μέγεθος ενός διανύσματος δίνεται από τη norm (νόρμα ή στάθμη) του.

Η απεικόνιση $||\cdot|| : \mathbb{C}^n \to \mathbb{R}^{+,0}$ ορίζει μια norm αν και μόνο αν ($\equiv$ ανν) ικανοποιεί τις παρακάτω τρεις ιδιότητες:

i) $||x|| \geq 0$ με $||x|| = 0$ ανν $x = 0$, $\forall\, x \in \mathbb{C}^n$
ii) $||cx|| = |c|\ ||x||$, $\forall\, c \in \mathbb{C}$ και $\forall\, x \in \mathbb{C}^n$
iii) $||x + y|| \leq ||x|| + ||y||$, $\forall\, x, y \in \mathbb{C}^n$

Αμεση συνέπεια των ιδιοτήτων (iii) και (ii) είναι και η $\mid\ ||x|| - ||y||\ \mid \leq ||x \pm y||$.

Στην Αριθμητική Ανάλυση χρησιμοποιούνται κυρίως οι παρακάτω τρεις ορισμοί για norms:

$$\begin{aligned} ||x||_1 &= \textstyle\sum_{i=1}^n |x_i| \ (\ell_1-\text{norm}) \\ ||x||_2 &= (\textstyle\sum_{i=1}^n |x_i|^2)^{\frac{1}{2}} \ \ (= (x,x)_2^{\frac{1}{2}}\ ) \ \ (\ell_2-\text{norm ή Ευκλείδεια norm}) \\ ||x||_\infty &= \max_{i=1(1)n} |x_i| \ (\ell_\infty-\text{norm ή maximum norm}) \end{aligned}$$

Οι τρεις αυτοί ορισμοί αποτελούν μερικές περιπτώσεις του γενικότερου

$$||x||_p = \left(\sum_{i=1}^n |x_i|^p \right)^{\frac{1}{p}},\ p \geq 1,$$

όπου για $p = \infty$ θεωρείται ως ορισμός ο $||x||_\infty = \lim_{p\to\infty} (\sum_{i=1}^n |x_i|^p)^{\frac{1}{p}}$.
Σημείωση: Για $x \in \mathbb{R}^2$, τα γραφήματα των εξισώσεων i) $||x||_1 = 1$, ii) $||x||_2 = 1$ και iii) $||x||_\infty = 1$, που είναι ισοδύναμες με i) $|x_1|+|x_2| = 1$, ii) $x_1^2+x_2^2 = 1$ και iii) $\max\{|x_1|, |x_2|\} = 1$, παριστάνονται, στο επίπεδο των αξόνων Ox_1x_2, i) από ρόμβο (τετράγωνο) με κορυφές $(1,0)$, $(0,1)$, $(-1,0)$, $(0,-1)$, ii) από κύκλο κέντρου $(0,0)$ και ακτίνας 1, και iii) από τετράγωνο με κορυφές $(1,1)$, $(-1,1)$, $(-1,-1)$, $(1,-1)$.

Το γεγονός ότι οι τρεις ορισμοί που δόθηκαν παραπάνω αποτελούν ορισμούς norms είναι εύκολο να αποδειχτεί στοιχειωδώς και συγκεκριμένα ότι ικανοποιούν τις ιδιότητες (i)-(iii) του ορισμού της norm. Η μόνη δυσκολία ίσως βρίσκεται στην περίπτωση της (iii), για την Ευκλείδεια norm, η απόδειξη της οποίας βασίζεται στην ανισότητα των Cauchy-Schwarz. Η τελευταία αποδείχνεται στη συνέχεια.

Θεώρημα 1.1 *Για κάθε $x, y \in \mathbb{C}^n$ ισχύει:*

$$|(x,y)_2| \leq ||x||_2 ||y||_2.$$

Απόδειξη: Για $x = 0$ ή $y = 0$ η δοθείσα ανισότητα ισχύει ως ισότητα $(0 = 0)$. Για κάθε $x, y \in \mathbb{C}^n \setminus \{0\}$ και για κάθε $\theta \in \mathbb{R}$ έχουμε

$$0 \leq \sum_{i=1}^n (\theta|x_i| + |y_i|)^2 = \theta^2 \sum_{i=1}^n |x_i|^2 + 2\theta \sum_{i=1}^n |x_i||y_i| + \sum_{i=1}^n |y_i|^2.$$

Το δεξιό μέλος των παραπάνω σχέσεων είναι ένα τριώνυμο δεύτερου βαθμού ως προς θ, με συντελεστή δευτεροβάθμιου όρου θετικό ($\sum_{i=1}^{n} |x_i|^2 > 0$) και είναι μή αρνητικό για όλες τις πραγματικές τιμές του θ. Αρα η διακρίνουσα του τριωνύμου, D, θα είναι μή θετική. Επομένως

$$0 \geq \frac{D}{4} = \left(\sum_{i=1}^{n} |x_i||y_i|\right)^2 - \left(\sum_{i=1}^{n} |x_i|^2\right)\left(\sum_{i=1}^{n} |y_i|^2\right) \iff$$
$$\left(\sum_{i=1}^{n} |\overline{x}_i y_i|\right)^2 \leq ||x||_2^2 ||y||_2^2 \implies \left|\sum_{i=1}^{n} \overline{x}_i y_i\right|^2 \leq ||x||_2^2 ||y||_2^2 \iff$$
$$|(x,y)_2| \leq ||x||_2 ||y||_2.$$

□

Με βάση την ανισότητα των Cauchy-Schwarz είναι εύκολο να δειχτεί ότι η Ευκλείδεια norm ικανοποιεί την ιδιότητα (iii) του ορισμού της norm ως εξής: Για την απόδειξη της $||x+y||_2 \leq ||x||_2 + ||y||_2$ αρκεί να δειχτεί η ισοδύναμή της $||x+y||_2^2 \leq (||x||_2 + ||y||_2)^2$ ή η

$$(x+y, x+y)_2 \leq ||x||_2^2 + 2||x||_2||y||_2 + ||y||_2^2.$$

Μετά την ανάπτυξη του εσωτερικού γινομένου του πρώτου μέλους και απλοποιώντας, έχοντας υπόψη ότι $(x,x)_2 = ||x||_2^2$ κ.λπ., βρίσκουμε $2Re(x,y)_2 \leq 2||x||_2||y||_2$ και άρα αρκεί να δειχτεί ότι $|(x,y)_2| \leq ||x||_2||y||_2$, η οποία δεν είναι άλλη παρά η ανισότητα των Cauchy-Schwarz, που ισχύει.

Ορισμός 1.2 *Δυο διανύσματα $x, y \in \mathbb{C}^n$ είναι ορθογώνια ανν $(x,y)_2 = 0$.*

Ειδικά για $x, y \in \mathbb{R}^n \setminus \{0\}$ μπορεί να οριστεί γωνία δύο διανυσμάτων από την έκφραση $\cos\theta = \frac{(x,y)_2}{||x||_2||y||_2} \in [-1,1]$, όπου η δεξιά σχέση ισχύει λόγω της ανισότητας των Cauchy-Schwarz. (Σημείωση: Είναι γνωστό από την Αναλυτική Γεωμετρία ότι ο παραπάνω τύπος για το $\cos\theta$, με $\theta \in [0,\pi]$, δίνει τη γωνία δύο διανυσμάτων.)

Μια διανυσματική norm είναι δυνατόν να θεωρηθεί ότι ορίζει μία απόσταση $d(x,y) = ||x-y||$ στο $\mathbb{C}^n$ αφού προφανώς ικανοποιεί τις:

$$\text{i) } d(x,x) = 0, \quad \forall\, x \in \mathbb{C}^n \text{ και } d(x,y) = 0 \text{ ανν } x = y, \quad \forall\, x, y \in \mathbb{C}^n$$
$$\text{ii) } d(x,y) = d(y,x), \quad \forall\, x, y \in \mathbb{C}^n$$
$$\text{iii) } d(x,y) \leq d(x,z) + d(z,y), \quad \forall\, x, y, z \in \mathbb{C}^n$$

Αρα ο $(\mathbb{C}^n,\ d(x,y) = ||x-y||)$ είναι μετρικός χώρος.

Ορισμός 1.3 *Μία ακολουθία διανυσμάτων $x^0,\ x^1,\ x^2,\ \cdots \in \mathbb{C}^n$ (ή $\{x^k\}_{k=0}^{\infty}$) συγκλίνει στο διάνυσμα $x \in \mathbb{C}^n$ (ή έχει όριο το διάνυσμα x) και γράφουμε $x^k_{k\to\infty} \to x$ ή $(x^k - x)_{k\to\infty} \to 0$ (ή αντίστοιχα $\lim_{k\to\infty} x^k = x$ ή $\lim_{k\to\infty}(x^k - x) = 0$) ανν $x^k_{i\,k\to\infty} \to x_i,\ \forall\, i = 1(1)n$.*

Ορισμός 1.4 *Δύο* norms $||\cdot||_\alpha$ *και* $||\cdot||_\beta$ *λέγονται ισοδύναμες (ή συγκρίσιμες) ανν υπάρχουν σταθερές* $c_1, c_2 > 0$ *τέτοιες ώστε (≡ τ.ω.)*

$$c_1||x||_\alpha \leq ||x||_\beta \leq c_2||x||_\alpha.$$

Σημείωση: Η παραπάνω ακολουθία ανισοτήτων είναι ισοδύναμη και με την $c_2^{-1}||x||_\beta \leq ||x||_\alpha \leq c_1^{-1}||x||_\beta$ και άρα δεν έχει σημασία για ποια από τις δύο norms θα θεωρείται ότι πολλαπλάσιά της περιέχουν την άλλη.

Θεώρημα 1.2 *Οποιεσδήποτε δύο* norms *στο* $\mathbb{C}^n$ *είναι συγκρίσιμες.*

Απόδειξη: Χωρίς περιορισμό της γενικότητας αρκεί να δειχτεί για την ℓ_2-norm. Αυτό γιατί αν $d_1||x||_\alpha \leq ||x||_2 \leq d_2||x||_\alpha$ και $d_1'||x||_\beta \leq ||x||_2 \leq d_2'||x||_\beta$ τότε μπορεί αμέσως να βρεθεί ότι και $d_2^{-1}d_1'||x||_\beta \leq ||x||_\alpha \leq d_1^{-1}d_2'||x||_\beta$, που ικανοποιεί τον ορισμό της ισοδυναμίας.

Εστω ότι $e^i,\ i = 1(1)n$, είναι τα διανύσματα της ορθοκανονικής βάσης του $\mathbb{C}^n$. Δηλαδή, $e^i_j = \delta_{ij},\ \ \forall\ i = 1(1)n$ και $j = 1(1)n$, όπου δ_{ij} είναι το δέλτα του Kronecker. Τότε αν $x_i,\ i = 1(1)n$, είναι οι συνιστώσες του x, $\forall\ x \in \mathbb{C}^n$, ως προς την ορθοκανονική βάση, θα έχουμε

$$f(x) := ||x||_\alpha \leq \sum_{i=1}^n |x_i|||e^i||_\alpha \leq \left(\sum_{i=1}^n |x_i|^2\right)^{\frac{1}{2}} \left(\sum_{i=1}^n ||e^i||_\alpha^2\right)^{\frac{1}{2}} = c_2||x||_2, \tag{1.1}$$

όπου η πρώτη από τα αριστερά ανισότητα ισχύει λόγω των ιδιότητων (ii) και (iii) της διανυσματικής norm, η δεύτερη λόγω της ανισότητας των Cauchy-Schwarz και προφανώς $c_2 := (\sum_{i=1}^n ||e^i||_\alpha^2)^{\frac{1}{2}}$ (> 0) είναι σταθερά ανεξάρτητη του x. Για τη συνάρτηση $f : \mathbb{C}^n \to \mathbb{R}^{+,0}$, χρησιμοποιώντας τις ιδιότητες των norms καθώς και το συμπέρασμα που μόλις προέκυψε, έχουμε

$$|f(x) - f(y)| \ = \ \big|\ ||x||_\alpha - ||y||_\alpha\ \big| \ \leq \ ||x - y||_\alpha \ \leq \ c_2||x - y||_2.$$

Αρα η f είναι συνεχής στο $\mathbb{C}^n$. Θεωρούμε τώρα τη μοναδιαία σφαίρα S στο $\mathbb{C}^n$, όπου οι αποστάσεις μετριούνται με την ℓ_2-norm, δηλαδή $S := \{x \in \mathbb{C}^n : ||x||_2 = 1\}$. Είναι φανερό ότι το σύνολο S είναι φραγμένο. Με βάση τον ορισμό της ακολουθίας διανυσμάτων και αυτόν της $l_2 - \text{norm}$, μπορεί να αποδειχτεί ότι είναι και κλειστό. Επομένως είναι συμπαγές και άρα η $f(x)$ ως συνεχής σ΄ αυτό παίρνει την ελάχιστη τιμή της για κάποιο $y \in S$. Δηλαδή, υπάρχει $y \in S :\ \ ||x||_\alpha \geq ||y||_\alpha = c_1 > 0,\ \ \forall\, x \in S$. $(c_1 > 0$, γιατί αν $c_1 = 0$ τότε $y = 0$ και $||y||_2 = 0 \neq 1.)$ Αρα $\forall\ \ x \in \mathbb{C}^n \setminus \{0\}$, $\frac{x}{||x||_2} \in S$, θα έχουμε

$$f(x) := ||x||_\alpha = ||x||_2||\frac{x}{||x||_2}||_\alpha \geq c_1||x||_2. \tag{1.2}$$

Οι (1.1) και (1.2) αποδείχνουν το θεώρημα. □

Με βάση τους ορισμούς της ακολουθίας και της ισοδυναμίας διανυσμάτων καθώς και του θεωρήματος που αποδείχτηκε έχουμε την παρακάτω πρόταση.

Θεώρημα 1.3 *Αν θεωρήσουμε μία ακολουθία διανυσμάτων* $\{x^k\}_{k=0}^{\infty}$ *με* $x^0, x^1, x^2, \cdots \in \mathbb{C}^n$, *τότε για κάθε διανυσματική* norm $||\cdot||_\alpha$ *ισχύει*

$$x^k_{k\to\infty} \to x \iff \lim_{k\to\infty} ||x^k - x||_\alpha = 0.$$

Απόδειξη: Η απόδειξη είναι προφανής. □

ΑΣΚΗΣΕΙΣ

1.: Να αποδειχτεί ότι οι απεικονίσεις στους ορισμούς των $||\cdot||_1$, $||\cdot||_2$ και $||\cdot||_\infty$ ικανοποιούν τις ιδιότητες (i), (ii) και (iii) των διανυσματικών norms.

2.: Να βρεθούν οι $||x||_1$, $||x||_2$ και $||x||_\infty$, όταν

$$x \;=\; [1,\; -i,\; 1+i,\; 1-i]^T.$$

3.: Αν $x, y \in \mathbb{R}^n$ είναι γραμμικά ανεξάρτητα με $||x||_2 = ||y||_2$, να αποδειχτεί ότι τα $x+y$ και $x-y$ είναι ορθογώνια μεταξύ τους.

4.: Εστω $x, y \in \mathbb{R}^n \backslash \{0\}$. Να δειχτεί ότι

$$||x+y||_2 \;=\; ||x||_2 + ||y||_2$$

ανν $y = \lambda x$ με $\lambda \in (0, +\infty)$.

5.: Για κάθε διανυσματική norm ισχύει

$$|||x|| - ||y||| \;\le\; ||x-y||, \quad \forall\, x, y \in \mathbb{C}^n.$$

6.: Να δειχτεί ότι η απεικόνιση, που ορίζεται από την

$$f(x) \;:=\; \sum_{i=1}^{n} i|x_i|, \quad \forall\, x \in \mathbb{C}^n,$$

αποτελεί τον ορισμό μιας διανυσματικής norm ενώ η απεικόνιση

$$g(x) \;:=\; \sum_{i=1}^{n} (i-1)|x_i|, \quad \forall\, x \in \mathbb{C}^n$$

δεν ορίζει norm.

7.: α) Να αποδειχτούν οι παρακάτω σχέσεις ισοδυναμίας των norms, για κάθε $x \in \mathbb{C}^n$.

$$\frac{1}{\sqrt{n}}||x||_1 \leq ||x||_2 \leq ||x||_1,\ \frac{1}{\sqrt{n}}||x||_2 \leq ||x||_\infty \leq ||x||_2,\ ||x||_\infty \leq ||x||_1 \leq n||x||_\infty.$$

και

β) Να βρεθούν έξι διανύσματα $x \in \mathbb{C}^n \backslash \{0\}$ τα οποία ικανοποιούν κάθε μία από τις παραπάνω έξι ισότητες.

1.2 Πίνακες

Θεωρούμε τους πίνακες $A \in \mathbb{C}^{m,n}$ ($\mathbb{R}^{m,n}$), όπου έχουν οριστεί οι πράξεις της πρόσθεσης και του πολλαπλασιασμού επί μιγαδικό (πραγματικό) αριθμό. Οπως είναι γνωστό ο $\mathbb{C}^{m,n}$ αποτελεί ένα γραμμικό διανυσματικό χώρο διάστασης mn.
Οι συμβολισμοί για τους παραπάνω πίνακες $A \in \mathbb{C}^{m,n}$ είναι οι γνωστοί

$$A := \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \equiv \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \equiv [a_{ij}],\ i = 1(1)m,\ j = 1(1)n.$$

Ακόμη, χρησιμοποιούνται συμβολισμοί αντίστοιχοι των διανυσμάτων για τον ανάστροφο πίνακα A^T καθώς και για το συζυγή μιγαδικό ανάστροφο A^H, που διαβάζεται και $A-$Ερμιτιανός.

Το γινόμενο δύο πινάκων $A \in \mathbb{C}^{m,n}$ και $B \in \mathbb{C}^{n,p}$ στη σειρά αυτή ορίζεται ως $C := AB \in \mathbb{C}^{m,p}$, όπου $c_{ij} = \sum_{k=1}^{n} a_{ik}b_{kj},\ i = 1(1)m,\ j = 1(1)p$. Είναι γνωστό ότι η προσεταιριστική και η επιμεριστική ιδιότητα του πολλαπλασιασμού ισχύουν για πίνακες με την προϋπόθεση ότι οι πράξεις των πινάκων που σημειώνονται ορίζονται. Το ίδιο **δεν** ισχύει για την αντιμεταθετική ιδιότητα, ακόμη κι αν τα γινόμενα AB και BA που σημειώνονται ορίζονται. Είναι δηλαδή γενικά $AB \neq BA$.
Υπενθυμίζεται ότι ισχύουν οι σχέσεις $(A^T)^T = A$, $(A^H)^H = A$ και ότι αν το γινόμενο AB ορίζεται ισχύουν και οι $(AB)^T = B^T A^T$ και $(AB)^H = B^H A^H$. Ακόμη, $(x, y)_2 = x^H y$ και $(x, x)_2 = x^H x = ||x||_2^2$.

Εφεξής θα αναφερόμαστε μόνο σε πίνακες τετραγωνικούς $A \in \mathbb{C}^{n,n}$, εκτός και αν σαφώς ορίζεται αλλιώς.

Ο ταυτοτικός πίνακας (ή μοναδιαίος) θα συμβολίζεται με I εκτός κι αν υπάρχει περίπτωση σύγχυσης οπότε θα συμβολίζεται με I_n.

Πίνακας $A \in \mathbb{C}^{n,n}$ με την ιδιότητα $A^H = A$ καλείται Ερμιτιανός, ενώ αν ικανοποιεί τη σχέση $A^T = A$ συμμετρικός. Σημειώνεται ότι ένας πραγματικός Ερμιτιανός πίνακας είναι προφανώς συμμετρικός.

Οπως είναι γνωστό, ο αντίστροφος δοθέντος πίνακα A δεν ορίζεται πάντοτε. Γενικά, για δοθέντα $A \in \mathbb{C}^{n,n}$ αν υπάρχουν πίνακες $X, Y \in \mathbb{C}^{n,n}$ τ.ω. $XA = I = AY$, τότε $X = Y$, ο αντίστροφος $X\ (= Y)$ είναι μοναδικός και συμβολίζεται με A^{-1}. Δεν έχουν, λοιπόν, όλοι οι πίνακες αντίστροφο. Αν όμως οι $A, B \in \mathbb{C}^{n,n}$ είναι αντιστρέψιμοι τότε μπορεί εύκολα να αποδειχτεί ότι ισχύουν τα παρακάτω:

$$(A^{-1})^{-1} = A,\ (A^T)^{-1} = (A^{-1})^T,\ (A^H)^{-1} = (A^{-1})^H \text{ και } (AB)^{-1} = B^{-1}A^{-1}.$$

Στη συνέχεια δίνεται μια πρόταση, χωρίς απόδειξη, ευρεία χρήση της οποίας θα γίνει στα επόμενα.

Θεώρημα 1.4 *Αν $A \in \mathbb{C}^{n,n}$, τότε οι παρακάτω προτάσεις είναι ισοδύναμες:*
i) Υπάρχει ο A^{-1}
ii) Τα διανύσματα-στήλες του A είναι γραμμικά ανεξάρτητα
iii) Το ίδιο ισχύει και για τα διανύσματα-γραμμές του A
iv) $\det(A) \neq 0$
v) Το γραμμικό σύστημα $Ax = 0$ έχει τη μοναδική λύση $x = 0$
vi) Το γραμμικό σύστημα $Ax = b$ έχει τη μοναδική λύση $x = A^{-1}b$

1.2.1 Ιδιοτιμές και Ιδιοδιανύσματα

Είναι γνωστό ότι για κάθε $A \in \mathbb{C}^{n,n}$ υπάρχουν $\lambda_i \in \mathbb{C}$ και $x^i \in \mathbb{C}^n \setminus \{0\},\ i = 1(1)n$, τ.ω. $Ax^i = \lambda_i x^i,\ i = 1(1)n$. Οι αριθμοί λ_i καλούνται ιδιοτιμές και τα x^i (αντίστοιχα) ιδιοδιανύσματα του πίνακα A. Οι n ιδιοτιμές βρίσκονται από την επίλυση της εξίσωσης $\det(A - \lambda I) = 0$ ενώ τα αντίστοιχα n ιδιοδιανύσματα από την επίλυση των n γραμμικών συστημάτων $(A - \lambda_i I)x^i = 0,\ i = 1(1)n$. (Σημείωση: Υπενθυμίζεται ότι τα n γραμμικά συστήματα έχουν πίνακες συντελεστών αγνώστων μή αντιστρέψιμους και ότι οι λύσεις που αναζητιούνται είναι μή μηδενικές. Ακόμη, αν x είναι ένα ιδιοδιάνυσμα του A τότε και το cx, όπου $c \in \mathbb{C} \setminus \{0\}$, είναι επίσης ιδιοδιάνυσμα που δε θεωρείται όμως διαφορετικό από το x. Επίσης, είναι δυνατόν τα n ιδιοδιανύσματα να είναι γραμμικά ανεξάρτητα, οπότε αποτελούν και μια βάση του $\mathbb{C}^n$, ή ο A να μην έχει n γραμμικά ανεξάρτητα ιδιοδιανύσματα.) Θα συμβολίζουμε με

$$\sigma(A) := \{\lambda_1, \lambda_2, \cdots, \lambda_n\} \text{ το φάσμα των ιδιοτιμών του } A$$

και με

$$\rho(A) := \max_{i=1(1)n} |\lambda_i| \text{ τη φασματική ακτίνα του } A.$$

1.2.2 Norms **Πινάκων**

Οι norms πινάκων ορίζονται με παραπλήσιο τρόπο με αυτόν των διανυσμάτων. Χαρακτηρίζονται από τρεις ιδιότητες που είναι αντίστοιχες των διανυσματικών norms καθώς κι από μια τέταρτη που αναφέρεται στην πολλαπλασιαστική ιδιότητα. Συγκεκριμένα έχουμε:

Ορισμός 1.5 *Η απεικόνιση* $||\cdot|| : \mathbb{C}^{n,n} \rightarrow \mathbb{R}^{+,0}$ *ορίζει μια* norm *αν*
i) $||A|| \geq 0,\ ||A|| = 0$ *αν* $A = 0,\ \forall\ A \in \mathbb{C}^{n,n}$
ii) $||cA|| = |c|\ ||A||,\ \forall\ c \in \mathbb{C}$ *και* $\forall\ A \in \mathbb{C}^{n,n}$
iii) $||A + B|| \leq ||A|| + ||B||,\ \forall\ A, B \in \mathbb{C}^{n,n}$
iv) $||AB|| \leq ||A||||B||,\ \forall\ A, B \in \mathbb{C}^{n,n}$

Οι περισσότερο χρησιμοποιούμενες norms πινάκων είναι αυτές που παράγονται από τις διανυσματικές norms με τον ακόλουθο τρόπο:
Για ένα δοθέντα $A \in \mathbb{C}^{n,n}$ και $\forall\ x \in \mathbb{C}^n \setminus \{0\}$ θεωρούμε το σύνολο των πηλίκων $\frac{||Ax||}{||x||}$. Τα εν λόγω πηλίκα μπορεί να δειχτεί ότι είναι φραγμένα. Προς τούτο αρκεί να φράξουμε από πάνω και κάτω τους δυο όρους του κλάσματος, αντίστοιχα, χρησιμοποιώντας την ισοδυναμία της οποιασδήποτε διανυσματικής norm με μια γνωστή, π.χ. την $\ell_\infty-$ norm. Στη συνέχεια μπορεί να δειχτεί ότι η παρακάτω απεικόνιση ορίζει μια norm πίνακα η οποία εφεξής θα καλείται φυσική norm.

Θεώρημα 1.5 *Η απεικόνιση* $||A|| := \sup_{x \in \mathbb{C}^n \setminus \{0\}} \frac{||Ax||}{||x||},\ \forall\ A \in \mathbb{C}^{n,n}$, *ορίζει μια* norm *πίνακα.*

Απόδειξη: Για να δειχτεί ότι η απεικόνιση της πρότασης ορίζει μια norm θα πρέπει να δειχτεί ότι αυτή ικανοποιεί τις ιδιότητες (i)-(iv) του ορισμού της norm. Οι αποδείξεις για τις ιδιότητες (i)-(iii) είναι απλές και παραλείπονται. Η απόδειξη της (iv) έχει ως εξής: Από τον ορισμό έχουμε ότι $\forall\ x \in \mathbb{C}^n \setminus \{0\},\ \frac{||Ax||}{||x||} \leq ||A|| \iff ||Ax|| \leq ||A||||x||$. (Σημείωση: Παρατηρούμε ότι η τελευταία ανισότητα ισχύει και για $x = 0$.) Εφαρμόζοντας την τελευταία ανισότητα διαδοχικά δυο φορές, $\forall\ x \in \mathbb{C}^n \setminus \{0\}$, παίρνουμε ότι $\frac{||ABx||}{||x||} \leq \frac{||A||||Bx||}{||x||} \leq \frac{||A||||B||||x||}{||x||} = ||A||||B||$, πράγμα που αποδείχνει την (iv). □

Ενας ισοδύναμος ορισμός για τη norm πίνακα $A \in \mathbb{C}^{n,n}$ δίνεται στην επόμενη πρόταση.

Θεώρημα 1.6 *Ο ορισμός του Θεωρήματος 1.5 είναι ισοδύναμος με τον ακόλουθο* $||A|| := \sup_{x \in \mathbb{C}^n,\ ||x||=1} ||Ax||$.

Απόδειξη: Με βάση την ιδιότητα (ii) των norms διανυσμάτων το πηλίκο στον ορισμό του Θεωρήματος 1.5 μπορεί να γραφτεί ισοδύναμα ως

$$\frac{||Ax||}{||x||} \;=\; ||Ay||, \;\; \text{όπου} \;\; y = \frac{x}{||x||}.$$

Είναι φανερό ότι χρησιμοποιώντας την ιδιότητα των norms διανυσμάτων που προαναφέρθηκε έχουμε $||y|| = ||\frac{x}{||x||}|| = \frac{1}{||x||}||x|| = 1$. Αντίστροφα, κάθε $y \in \mathbb{C}^n$ με $||y|| = 1$ μπορεί να γραφτεί ως $y = \frac{cy}{||cy||}$, $\forall\, c > 0$, οπότε για $x = cy$, ο ορισμός του παρόντος θεωρήματος δίνει τον ορισμό του Θεωρήματος 1.5. □

Σημείωση: Στηριζόμενοι στον ισοδύναμο ορισμό της φυσικής norm του Θεωρήματος 1.6 είναι δυνατόν να αποδειχτεί ότι η συνάρτηση $f(x) := ||Ax||$, με $||x|| = 1$, είναι συνεχής. Γι' αυτό το σκοπό και για $||x|| = ||y|| = 1$ έχουμε:

$$|f(x) - f(y)| = \big|\; ||Ax|| - ||Ay|| \;\big| \leq ||Ax - Ay|| = ||A(x-y)|| \leq ||A||||x-y||.$$

Επειδή είναι γνωστό ότι η $||A||$ είναι φραγμένη, η ανισότητα του αριστερά και δεξιά μέλους στην παραπάνω ακολουθία σχέσεων αποδείχνει τη συνέχεια της $f(x)$ στο πεδίο ορισμού της, $S := \{x \in \mathbb{C}^n : \; ||x|| = 1\}$. Επειδή δε το S είναι συμπαγές (κλειστό και φραγμένο) και η $f(x)$ είναι συνεχής σ' αυτό θα υπάρχει $x \in S$ στο οποίο η $f(x)$ θα παίρνει τη μέγιστη δυνατή τιμή της. Αυτό σημαίνει ότι το supremum στα Θεωρήματα 1.5 και 1.6 θα μπορεί να αντικατασταθεί από το maximum.

Βασιζόμενοι στην ισοδυναμία (συγκρισιμότητα) δυο οποιωνδήποτε διανυσματικών norms στο $\mathbb{C}^n$ καθώς και στον ορισμό της φυσικής norm πίνακα μπορεί να αποδειχτεί η παρακάτω πρόταση.

Θεώρημα 1.7 *Δυο οποιεσδήποτε φυσικές* norms *πινάκων στο $\mathbb{C}^{n,n}$ είναι ισοδύναμες (συγκρίσιμες).*

1.2.3 Ακολουθίες Πινάκων και Σύγκλιση

Εστω μία ακολουθία πινάκων $A^{(0)}, A^{(1)}, A^{(2)}, \cdots \in \mathbb{C}^{n,n}$, που γράφεται συμβολικά και ως $\{A^{(k)}\}_{k=0}^{\infty}$. Η υπόψη ακολουθία συγκλίνει στον πίνακα $A \in \mathbb{C}^{n,n}$ ($A^{(k)}{}_{k\to\infty} \;\to\; A \iff \lim_{k\to\infty} A^{(k)} = A$) ανν $\lim_{k\to\infty} a_{ij}^{(k)} = a_{ij}, \;\; i,j = 1(1)n$. Συνέπεια του ορισμού αυτού και των αμέσως προηγούμενων θεωρημάτων είναι ότι για μια οποιαδήποτε φυσική norm πίνακα $||\cdot||$ ισχύει ότι $\lim_{k\to\infty} A^{(k)} = A$ ανν $\lim_{k\to\infty} ||A^{(k)} - A|| \;=\; 0$.

1.2.4 Προτάσεις για τις Φυσικές Norms Πινάκων

Στη συνέχεια δίνουμε τέσσερις προτάσεις πάνω στις norms πινάκων η τρίτη από τις οποίες δίνεται χωρίς απόδειξη.

Θεώρημα 1.8 *Για κάθε φυσική* norm *ισχύει:* $||I|| \;=\; 1$.

Απόδειξη: Η απόδειξη με βάση τον ορισμό της norm, για $A = I$, είναι άμεση. □

Θεώρημα 1.9 *Για κάθε φυσική* norm *και* $\forall\, A \in \mathbb{C}^{n,n}$ *ισχύει:* $\rho(A) \leq ||A||$.

Απόδειξη: Εστω $\lambda \in \mathbb{C}$ μια ιδιοτιμή του A και $x \in \mathbb{C}^n \setminus \{0\}$ το αντίστοιχο ιδιοδιάνυσμα. Θα έχουμε: $Ax = \lambda x$, οπότε παίρνοντας norms και εφαρμόζοντας την ιδιότητα (ii) καθώς και την αποδειχθείσα ανισότητα $||Ax|| \leq ||A||||x||$ προκύπτει αμέσως ότι $||A||||x|| \geq |\lambda|||x||$. Διαιρώντας και τα δύο μέλη της τελευταίας σχέσης δια $||x||(> 0)$ έχουμε $|\lambda| \leq ||A||$ και το θεώρημα αποδείχτηκε αφού η τελευταία ανισότητα αληθεύει $\forall\, \lambda \in \sigma(A)$. □

Θεώρημα 1.10 $\forall\, A \in \mathbb{C}^{n,n}$ *και* $\forall\, \epsilon > 0$ *υπάρχει φυσική* norm $||\cdot||$ *τ.ω.* $||A|| \leq \rho(A) + \epsilon$.

(Σημείωση: Η απόδειξη θα δοθεί στο Κεφάλαιο 3 μετά την εισαγωγή της κανονικής μορφής του Jordan.)

Σημείωση: Θα πρέπει να τονιστεί ότι ενώ η φασματική ακτίνα πίνακα $A \in \mathbb{C}^{n,n}$ **δεν** αποτελεί norm, με βαση τα Θεωρήματα 1.9 και 1.10 διαπιστώνεται, ότι αυτή μπορεί να προσεγγιστεί από **κάποια** φυσική norm όσο καλά αυτό είναι επιθυμητό.

Θεώρημα 1.11 *(*Neumann*). Αν* $A \in \mathbb{C}^{n,n}$ *και για μια φυσική* norm *στο* $\mathbb{C}^{n,n}$ *ισχύει* $||A|| < 1$, *τότε*
i*) Ο* $I - A$ *είναι αντιστρέψιμος, και*
ii) $\frac{1}{1+||A||} \leq ||(I-A)^{-1}|| \leq \frac{1}{1-||A||}$.

Απόδειξη: i) Εστω ότι ο $I - A$ είναι μή αντιστρέψιμος. Τότε θα υπάρχει $x \in \mathbb{C}^n \setminus \{0\}$ τ.ω. $(I - A)x = 0$. Από την τελευταία σχέση παίρνουμε διαδοχικά $x = Ax \implies ||x|| \leq ||A||||x||$ και επειδή $||x|| > 0$ έχουμε ισοδύναμα ότι $1 \leq ||A||$, που είναι άτοπο.
ii) Με βάση την αντιστρεψιμότητα του $I - A$ έχουμε ότι $I = (I-A)^{-1}(I-A) = (I-A)^{-1} - (I - A)^{-1}A$. Παίρνοντας norms των ακραίων μελών και εφαρμόζοντας απλές ιδιότητές έχουμε διαδοχικά: $1 = ||I|| \leq ||(I-A)^{-1}||+||(I-A)^{-1}||||A|| = ||(I-A)^{-1}||(1+||A||) \iff \frac{1}{1+||A||} \leq ||(I-A)^{-1}||$.
Με αντίστοιχο τρόπο μπορούμε να πάρουμε ότι:
$1 = ||I|| \geq ||(I-A)^{-1}||-||(I-A)^{-1}||||A|| = ||(I-A)^{-1}||(1-||A||) \iff \frac{1}{1-||A||} \geq ||(I-A)^{-1}||$.
□
Σημείωση: Το Θεώρημα του Neumann είναι δυνατόν να διατυπωθεί με τον ίδιο ακριβώς τρόπο και για τον πίνακα $I + A$ αρκεί τόσο στην υπόθεση του θεωρήματος όσο και στα συμπεράσματά του να τεθεί όπου A το $-A$.

Οπως και στις norms διανυσμάτων έτσι και στις norms πινάκων οι συνηθέστερα χρησιμοποιούμενες είναι αυτές που προκύπτουν, με βάση τον ορισμό, από τις ℓ_1-norm, ℓ_2-norm και $\ell_\infty-$norm των διανυσμάτων. Συγκεκριμένα μπορεί να αποδειχτεί ότι $\forall\, A \in \mathbb{C}^{n,n}$ ισχύουν:

$$\begin{array}{c} ||A||_1 \ = \ \max_{j=1(1)n} \sum_{i=1}^{n} |a_{ij}| \\ ||A||_2 \ = \ \rho^{\frac{1}{2}}(A^H A), \text{ (φασματική norm)} \\ ||A||_\infty \ = \ \max_{i=1(1)n} \sum_{j=1}^{n} |a_{ij}| \end{array}$$

Στη συνέχεια θα δοθούν οι αποδείξεις των τύπων για τη $||A||_\infty$, η απόδειξη για τη $||A||_1$ είναι σε γενικές γραμμές της ίδιας μορφής αλλά απλούστερη, και της $||A||_2$.

Για δοσμένο $A \in \mathbb{C}^{n,n}$ και $\forall\ x \in \mathbb{C}^n$ με $||x||_\infty = 1$ παίρνουμε διαδοχικά

$$\begin{array}{rcl} ||Ax||_\infty & = & \max_{i=1(1)n} |\sum_{j=1}^{n} a_{ij}x_j| \leq \ \max_{i=1(1)n} \sum_{j=1}^{n} (|a_{ij}||x_j|) \\ & \leq & \max_{i=1(1)n} \sum_{j=1}^{n} (|a_{ij}|||x||_\infty) = \max_{i=1(1)n} \sum_{j=1}^{n} |a_{ij}|, \end{array}$$

αφού $||x||_\infty = 1$. Από το αριστερά και δεξιά μέλος των παραπάνω σχέσεων προκύπτει αμέσως ότι

$$||Ax||_\infty \ \leq \ \max_{i=1(1)n} \sum_{j=1}^{n} |a_{ij}|$$

και άρα

$$||A||_\infty \ \leq \ \max_{i=1(1)n} \sum_{j=1}^{n} |a_{ij}|. \tag{1.3}$$

Εστω $k \in \{1,\ 2,\ \cdots,\ n\}$ ο δείκτης της γραμμής για την οποία το άθροισμα $\sum_{j=1}^{n} |a_{kj}|$ είναι μέγιστο. Ορίζουμε το διάνυσμα $y \in \mathbb{C}^n$ με συνιστώσες

$$y_j \ = \ \begin{cases} \frac{\overline{a}_{kj}}{|a_{kj}|}, & \alpha\nu\ a_{kj} \neq 0 \\ 1, & \alpha\nu\ a_{kj} = 0 \end{cases}, \quad j = 1(1)n.$$

Από τον ορισμό των y_j, για το συγκεκριμένο y, συμπεραίνεται ότι $||y||_\infty = 1$ και ακόμη ότι $||Ay||_\infty \ = \ \max_{i=1(1)n} |\sum_{j=1}^{n} a_{ij}y_j| \ \geq \ |\sum_{j=1}^{n} a_{kj}y_j| \ = \ \sum_{j=1}^{n} |a_{kj}| \ = \ \max_{i=1(1)n} \sum_{j=1}^{n} |a_{ij}|$, από τον ορισμό του δείκτη k. Αρα

$$||A||_\infty \ \geq \ ||Ax||_\infty \ \geq ||Ay||_\infty \ \geq \ \max_{i=1(1)n} \sum_{j=1}^{n} |a_{ij}|, \tag{1.4}$$

για το οποιοδήποτε x, με $||x||_\infty = 1$. Οι (1.3) και (1.4) αποδείχνουν την ισχύ του ζητούμενου τύπου για τη $||A||_\infty$, δηλαδή ότι $||A||_\infty \ = \ \max_{i=1(1)n} \sum_{j=1}^{n} |a_{ij}|$.

Για τη $||A||_2$, εφόσον έχουμε την αποδεικτέα έκφραση, θεωρούμε τον πίνακα $A^H A$. Ο πίνακας $A^H A$ είναι Ερμιτιανός διότι $(A^H A)^H = A^H (A^H)^H = A^H A$, έχει δε μή αρνητικές ιδιοτιμές. Ο

ισχυρισμός αυτός αποδείχνεται ως εξής. Εστω λ μία ιδιοτιμή του A^HA με αντίστοιχο ιδιοδιάνυσμα $x \in \mathbb{C}^n \setminus \{0\}$. Θα είναι $A^HAx = \lambda x$, οπότε και $x^HA^HAx = \lambda x^Hx$ ή $||Ax||_2^2 = \lambda||x||_2^2$. Επειδή $||Ax||_2^2 \geq 0$ και $||x||_2^2 > 0$, έπεται ότι $\lambda \geq 0$. Από τη Γραμμική Αλγεβρα γνωρίζουμε ότι κάθε Ερμιτιανός πίνακας έχει n γραμμικά ανεξάρτητα ιδιοδιανύσματα που μπορούν να παρθούν ορθοκανονικά. Αν $x^i,\ i = 1(1)n$, είναι τα ιδιοδιανύσματα του A^HA, για τα οποία ισχύει ότι $(x^i, x^j)_2 = \delta_{ij},\ i,j = 1(1)n$, με αντίστοιχες ιδιοτιμές $\lambda_i,\ i = 1(1)n$, όπου $0 \leq \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$, τότε για κάθε $x \in \mathbb{C}^n$ με $||x||_2 = 1$ το εν λόγω διάνυσμα γράφεται και ως $x = \sum_{i=1}^n c_i x^i$, με $c_i \in \mathbb{C}$ και $\sum_{i=1}^n |c_i|^2 = 1$. Θα έχουμε διαδοχικά τις σχέσεις που προκύπτουν από απλούς μετασχηματισμούς

$$\begin{aligned} &||A||_2^2 = \max||Ax||_2^2 = \max(Ax, Ax)_2 = \max(x, A^HAx)_2 \\ &= \max\left(\sum_{i=1}^n c_i x^i, A^HA\sum_{j=1}^n c_j x^j\right)_2 = \max\left(\sum_{i=1}^n c_i x^i, \sum_{j=1}^n c_j\lambda_j x^j\right)_2 \\ &= \max\sum_{i=1}^n \lambda_i|c_i|^2 \leq \lambda_n \max\sum_{i=1}^n |c_i|^2 = \lambda_n = \rho(A^HA). \end{aligned} \tag{1.5}$$

Αν τώρα επιλέξουμε $x = x^n$, δηλαδή το ιδιοδιάνυσμα που αντιστοιχεί στη μεγαλύτερη ιδιοτιμή, η μοναδική ανισότητα στις σχέσεις (1.5) γίνεται ισότητα πράγμα που αποδείχνει τον τύπο για τη $||A||_2$.

1.2.5 Αριθμός (ή Δείκτης) Κατάστασης Πίνακα

Ορισμός 1.6 *Αριθμός (ή δείκτης) κατάστασης αντιστρέψιμου πίνακα $A \in \mathbb{C}^{n,n}$ ως προς φυσική* norm $||\cdot||$ *καλείται ο αριθμός*

$$\kappa(A) \ = \ ||A||||A^{-1}||.$$

Γι΄ αυτόν ισχύει προφανώς πάντοτε ότι $\kappa(A) \ \geq \ 1$.

Σημειώσεις: α) Οταν αναφερόμαστε σε μια συγκεκριμένη norm, $||\cdot||_\alpha$, τότε ο δείκτης κατάστασης συμβολίζεται με $\kappa_\alpha(A)$. β) Για μή αντιστρέψιμο πίνακα A ο δείκτης κατάστασης ως προς μια οποιαδήποτε φυσική norm θεωρείται ίσος με ∞.
Στη συνέχεια δίνουμε μια πρόταση που αφορά στο δείκτη κατάστασης.

Θεώρημα 1.12 *Εστω $A \in \mathbb{C}^{n,n}$ αντιστρέψιμος. Για το δείκτη κατάστασης του A ισχύει*

$$\frac{1}{\kappa(A)} \ \leq \ \inf_{B \in \mathbb{C}^{n,n},\ \det(B)=0} \left\{\frac{||A-B||}{||A||}\right\}$$

Απόδειξη: Για την απόδειξη της πρότασης αρκεί να αποδειχτεί ότι $\frac{1}{\kappa(A)} \ \leq \ \frac{||A-B||}{||A||}$. Αφού ο τυχόν B είναι μή αντιστρέψιμος υπάρχει $x \ \in \ \mathbb{C}^n \setminus \{0\} \ : \ Bx = 0$ ή ισοδύναμα $(A-B)x = Ax$ ή $A^{-1}(A-B)x = x$. Παίρνοντας norms και εφαρμόζοντας απλές ιδιότητες έχουμε αμέσως ότι

$||A^{-1}||||A-B|| \geq 1$, αφού $||x|| > 0$. Διαιρώντας τώρα και τα δυο μέλη δια του δείκτη κατάστασης έχουμε την προς απόδειξη σχέση. Ο τελευταίος ισχυρισμός είναι αληθής αφού η αποδειχθείσα ανισότητα ($\leq$) ισχύει για τον τυχόντα μή αντιστρέψιμο πίνακα B η ίδια ανισότητα θα ισχύει και για το infimum του δεύτερου μέλους. □

Παρατηρήσεις: α) Αν θεωρήσουμε το κλάσμα του δεύτερου μέλους της πρότασης του θεωρήματος σαν το σχετικό απόλυτο σφάλμα του πίνακα A ως προς το σύνολο των μή αντιστρέψιμων πινάκων B, τότε ο δείκτης κατάστασης εκφράζει το πόσο σχετικά κοντά (ή μακρυά) είναι ο A από του να είναι μή αντιστρέψιμος. β) Υπάρχουν φυσικές norms για τις οποίες η δοθείσα ανισότητα ισχύει ως ισότητα.

ΑΣΚΗΣΕΙΣ

1.: Να δειχτεί ότι κάθε πίνακας $A \in \mathbb{C}^{n,n}$ μπορεί να γραφτεί κατά ένα και μόνο τρόπο ως άθροισμα $A = B + C$, όπου B Ερμιτιανός ($B^H = B$) και C αντι–Ερμιτιανός ($C^H = -C$).

2.: Εστω ότι ο πίνακας $A \in \mathbb{C}^{n,n}$ είναι μή αντιστρέψιμος. Δίνεται η διάσπαση $A = B - C$, όπου ο B είναι αντιστρέψιμος. Να δειχτεί ότι $\rho(B^{-1}C) \geq 1$ και, για οποιαδήποτε φυσική norm, $||B^{-1}|| \geq \frac{1}{||C||}$.

3.: Αν $A \in \mathbb{C}^{n,n}$ να δειχτεί η ισχύς των σχέσεων

$$\alpha) \quad \sum_{i=1}^{n} \lambda_i \;=\; \sum_{i=1}^{n} a_{ii} \quad \kappa\alpha\iota \quad \beta) \quad \prod_{i=1}^{n} \lambda_i \;=\; \det(A),$$

όπου $\lambda_i \in \sigma(A),\ i = 1(1)n$.

4.: Να δειχτεί ότι

$$||A||_1 \;=\; \max_{j=1(1)n} \sum_{i=1}^{n} |a_{ij}| \quad \forall\ A \in \mathbb{C}^{n,n}.$$

5.: Να βρεθούν οι norms $||\cdot||_1,\ ||\cdot||_2$ και $||\cdot||_\infty$ για τους πίνακες

$$\begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix}, \quad \begin{bmatrix} 2 & -1 & 0 & 0 \\ -1 & 2 & -1 & 0 \\ 0 & -1 & 2 & -1 \\ 0 & 0 & -1 & 2 \end{bmatrix}.$$

6.: Να δειχτεί ότι η απεικόνιση $||\cdot||_F : \mathbb{C}^{n,n} \to [0, +\infty)$, που ορίζεται από τη

$$||A||_F \ := \left(\sum_{i,j=1}^{n} |a_{ij}|^2 \right)^{\frac{1}{2}} \quad \forall\ A \in \mathbb{C}^{n,n},$$

αποτελεί τον ορισμό μιας norm (norm του Frobenius). Είναι η εν λόγω norm φυσική;

7.: Να αποδειχτεί ότι η φασματική ακτίνα πίνακα ως απεικόνιση

$$\rho(A) \ := \max_{i=1(1)n} |\lambda_i|, \quad \lambda_i \in \sigma(A),\ i = 1(1)n, \quad \forall\ A \in \mathbb{C}^{n,n}$$

δεν μπορεί να αποτελεί τον ορισμό norm πίνακα.

8.: Αν θεωρήσουμε τους πίνακες $A \in \mathbb{C}^{n,n}$ ως διανύσματα του χώρου $\mathbb{C}^{n^2}$, τότε το ανάλογο της διανυσματικής norm $||\cdot||$ είναι η "norm" πινάκων

$$||A||_{MAX} \ = \max_{i,j=1(1)n} |a_{ij}|.$$

Να αποδειχτεί ότι η ορισθείσα norm ικανοποιεί τις ιδιότητες (i), (ii) και (iii) του ορισμού της norm πίνακα ενώ **δεν** ικανοποιεί την (iv). (Σημείωση: Ενα αντιπαράδειγμα για την ιδιότητα (iv) αρκεί.)

9.: Αν $\kappa_1(A),\ \kappa_2(A)$ και $\kappa_\infty(A)$ είναι οι δείκτες κατάστασης του πίνακα $A \in \mathbb{C}^{n,n}$, που αντιστοιχούν στις norms $||A||_1,\ ||A||_2$, και $|||A||_\infty$, να αποδειχτεί ότι

$$\kappa_2^2(A) \ \leq \kappa_1(A)\kappa_\infty(A).$$

10.: Να αποδειχτεί ότι για κάθε *ορθογώνιο* πίνακα A ($A \in \mathbb{R}^{n,n},\ A^{-1} = A^T$) ισχύουν οι σχέσεις:

$$||A||_2 = 1, \quad ||A||_1 \geq 1, \quad ||A||_\infty \geq 1.$$

11.: Εστω $A \in \mathbb{R}^{n,n}$ και $Q \in \mathbb{R}^{n,n}$ ένας ορθογώνιος πίνακας. Να αποδειχτεί ότι:
α) $||QA||_2 \ = \ ||A||_2 \ = \ ||AQ||_2$. και
β) Αν $||A||_2 < 1$, τότε ο πίνακας $Q - A$ είναι αντιστρέψιμος και ισχύει η σχέση

$$||(Q - A)^{-1}||_2 \ \leq \ \frac{1}{1 - ||A||_2}.$$

12.: Αν $A \in \mathbb{C}^{n,n}$ είναι αντιστρέψιμος και για κάποια διανυσματική norm ισχύει $||Ax|| = ||x||$, $\forall\, x \in \mathbb{C}^n$, να αποδειχτεί ότι για το δείκτη κατάστασης, που συνδέεται με την αντίστοιχη φυσικη norm, ισχύει $\kappa(A) = 1$.

13.: Να αποδειχτούν οι παρακάτω σχέσεις, που συνδέουν τις φασματικές ακτίνες:
α) Δύο πραγματικών συμμετρικών πινάκων $A, B \in \mathbb{R}^{n,n}$,

$$\rho(A+B) \leq \rho(A) + \rho(B), \quad \rho(AB) \leq \rho(A)\rho(B).$$

και,
β) Δύο πραγματικών ορθογώνιων πινάκων $P, Q \in \mathbb{R}^{n,n}$,

$$\rho(P+Q) \leq 2, \quad \rho(PQ) \leq 1.$$

14.: Αν $A, B \in \mathbb{C}^{n,n}$, υπάρχει ο A^{-1} και ο B ικανοποιεί τη $||A-B|| < \frac{1}{||A^{-1}||}$ για κάποια φυσική norm, τότε και ο B είναι αντιστρέψιμος.

15.: Είναι γνωστό ότι: " Αν $A \in \mathbb{C}^{n,n}$ και για μια φυσική norm ικανοποιεί τη $||A|| < 1$ τότε οι πίνακες $I \pm A$ είναι αντιστρέψιμοι". Να χρησιμοποιηθεί η συγκεκριμένη ιδιότητα για να εξεταστεί αν ο πίνακας $B = \begin{bmatrix} 1 & -\frac{i}{2} & 0 \\ -\frac{1}{2} & 1 & -\frac{i}{2} \\ 0 & -\frac{1}{2} & 1 \end{bmatrix}$ είναι αντιστρέψιμος.

16.: Να αποδειχτεί ότι:
α) Αν ο πίνακας $A \in \mathbb{C}^{n,n}$ είναι Ερμιτιανός ($A^H = A$) οι ιδιοτιμές του είναι πραγματικές. Επιπλέον ισχύει ότι $x^H A x \in \mathbb{R}\ \forall x \in \mathbb{C}^n$.
β) Αν ο πίνακας $A \in \mathbb{C}^{n,n}$ είναι αντι–Ερμιτιανός ($A^H = -A$) οι ιδιοτιμές του είναι καθαρά φανταστικές. Επιπλέον ισχύει ότι $x^H A x \in \imath\mathbb{R}\ \forall x \in \mathbb{C}^n$, όπου $\imath$ η φανταστική μονάδα, και ακόμη ότι για $A \in \mathbb{R}^{n,n}$ και $\forall x \in \mathbb{R}^n$, $x^T A x = 0$.
γ) Αν ο πίνακας $A \in \mathbb{C}^{n,n}$ είναι *ορθομοναδιαίος* (unitary) ($A^H A = I$) οι ιδιοτιμές του έχουν μέτρο μονάδα.

17.: Χωρίς να βρεθεί ο A^{-1} να αποδειχτεί ότι για τον πίνακα

$$A = \begin{bmatrix} 1.01 & .99 \\ .99 & 1.01 \end{bmatrix}$$

ισχύει $\kappa_\infty(A) \geq 100$. (Υπόδειξη: Ο πίνακας $\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$ **δεν** είναι αντιστρέψιμος.)

18.: Αν οι πίνακες $A, B \in \mathbb{C}^{n,n}$ είναι αντιστρέψιμοι να δειχτεί ότι

$$\frac{||B^{-1} - A^{-1}||}{||B^{-1}||} \leq \kappa(A) \frac{||A - B||}{||A||}.$$

19.: Αν ο Ερμιτιανός πίνακας $A \in \mathbb{C}^{n,n}$ είναι αντιστρέψιμος να δειχτεί ότι ο δείκτης κατάστασής του $\kappa_2(A)$ δίνεται από τον τύπο

$$\kappa_2(A) = \frac{\max_i |\lambda_i|}{\min_i |\lambda_i|},$$

όπου $\lambda_i,\ i = 1(1)n$, οι ιδιοτιμές του A.

20.: Να δειχτεί ότι για οποιαδήποτε φυσική norm πίνακα στο $\mathbb{C}^{n,n}$ ισχύουν ότι:

$$\alpha) \quad \kappa(\lambda A) = \kappa(A), \ \ \forall\, \lambda \in \mathbb{C}\backslash\{0\}.$$

και

$$\beta) \quad \kappa_1(D) = \kappa_2(D) = \kappa_\infty(D) = \frac{\max_{i=1(1)n} |d_{ii}|}{\min_{i=1(1)n} |d_{ii}|},$$

όπου $D \in \mathbb{C}^{n,n}$ αντιστρέψιμος διαγώνιος πίνακας.

21.: Δίνεται ο αντιστρέψιμος πίνακας $A \in \mathbb{C}^{n,n}$ και $B \in \mathbb{C}^{n,n}$ μια προσέγγισή του τ.ω. $||B^{-1}C|| \leq \frac{1}{9}$ για κάποια φυσική norm, με $C = B - A$. Να αποδειχτεί ότι για τον αντίστοιχο δείκτη κατάστασης ισχύει:

$$1 \leq \kappa(B^{-1}A) = \kappa(A^{-1}B) \leq 1.25.$$

22.: Σε μερικές εφαρμογές παρουσιάζονται πίνακες της μορφής $A = \begin{bmatrix} O & B \\ C & O \end{bmatrix}$, με $O, B, C \in \mathbb{C}^{n,n}$, O το μηδενικό πίνακα και $\det(BC) \neq 0$. Να δειχτεί ότι: Αν $\lambda \in \sigma(A)$ τότε $\lambda \neq 0$, και αν $\lambda \in \sigma(A)$ τότε και $-\lambda \in \sigma(A)$.

23.: Το πηλίκο $\frac{(x,Ax)_2}{(x,x)_2} \ \ \forall\ x \in \mathbb{C}^n\backslash\{0\}$, για δοθέντα πίνακα A είναι γνωστό ως πηλίκο του Rayleigh. Να αποδειχτεί ότι αν $A \in \mathbb{C}^{n,n}$, με $A^H = A$, το αντίστοιχο πηλίκο του Rayleigh βρίσκεται μεταξύ των δύο ακραίων (πραγματικών) ιδιοτιμών του πίνακα A.

24.: Δίνεται ο πραγματικός συμμετρικός πίνακας $A = \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix}$. Να δειχτεί ότι

$$2 - \sqrt{2} \leq \frac{x^T A x}{x^T x} \leq 2 + \sqrt{2} \quad \forall\, x \in \mathbb{R}^3\backslash\{0\}.$$

25.: α) Αν $A \in \mathbb{C}^{n,n}$ να αποδειχτεί ότι $\lambda_i \in \sigma(A + A^H) \subset \mathbb{R},\ i = 1(1)n$, και $\mu_j \in \sigma(A - A^H) \subset \imath\mathbb{R},\ j = 1(1)n$, όπου $\imath = \sqrt{-1}$. και
β) Υποτίθεται ότι οι άγνωστες ιδιοτιμές ενός πίνακα $A \in \mathbb{C}^{n,n}$ είναι της μορφής $\mu_i + \imath\xi_i$ με $\mu_i, \xi_i \in \mathbb{R}$, $i = 1(1)n$, και ότι είναι δυνατόν να βρεθούν οι ακραίες ιδιοτιμές ενός οποιουδήποτε Ερμιτιανού πίνακα. Τότε να βρεθεί με βάση το (α) το μικρότερο δυνατό ορθογώνιο R στο μιγαδικό επίπεδο με πλευρές παράλληλες προς τον πραγματικό και φανταστικό άξονα, αντίστοιχα, τέτοιο ώστε $\sigma(A) \subset \overline{R}$. (Υπόδειξη: Να χρησιμοποιηθεί το πηλίκο του Rayleigh.)

26.: α) Αν ο πίνακας $A \in \mathbb{C}^{n,n}$ είναι ορθογώνιος τότε $\det(A) = +1$ ή $= -1$.
β) Αν οι πινακες $A, B \in \mathbb{R}^{n,n}$ είναι ορθογώνιοι και $\det(A) + \det(B) = 0$, τότε και $\det(A + B) = 0$.

27.: Αν οι $A, B \in \mathbb{C}^{n,n}$, με A αντιστρέψιμο πίνακα, συνδέονται με τη σχέση $B = A - uv^H$, όπου $u, v \in \mathbb{C}^n$ και $v^H A^{-1} u \neq 1$, να δειχτεί ότι ο πίνακας C, που δίνεται από την έκφραση

$$C = A^{-1} + \alpha(A^{-1}u)(v^H A^{-1}), \quad \alpha = \frac{1}{1 - v^H A^{-1} u},$$

αποτελεί τον αντίστροφο του B. (Σημείωση: Η παραπάνω έκφραση για τον αντίστροφο του B είναι γνωστή ως τύπος των Sherman-Morrison.)

28.: Αν $A \in \mathbb{C}^{n,n} \backslash \{O\}$ να βρεθούν:
α) Οι σταθερές σύγκρισης των τριών ζευγών των norms $||A||_1$, $||A||_2$ και $||A||_\infty$. (**Υποδείξεις**: α) Για τη σύγκριση της l_2-norm με τις άλλες δύο να χρησιμοποιηθεί το *ίχνος* (*trace*) του πίνακα $A^H A$ και συγκεκριμένα, μεταξύ άλλων, ότι $\rho(A^H A) \leq \text{trace}(A^H A) \equiv \sum_{i=1}^{n} (A^H A)_{ii}$.
β) Να θεωρηθεί γνωστό ότι για δυο norms $||A||_a$ και $||A||_b$ πίνακα $A \in \mathbb{C}^{n,n} \backslash \{0\}$, που παράγονται από δυο διανυσματικές norms $||x||_a$ και $||x||_b$, $x \in \mathbb{C}^n$, αντίστοιχα, ισχύει ότι

$$\sup_{A \in \mathbb{C}^{n,n} \backslash \{0\}} \frac{||A||_a}{||A||_b} \cdot \sup_{A \in \mathbb{C}^{n,n} \backslash \{0\}} \frac{||A||_b}{||A||_a} \ = \ 1.)$$

και
β) Εξι πίνακες $A \in \mathbb{C}^{n,n} \backslash \{O\}$ κάθε ένας από τους οποίους να ικανοποιεί και μία από τις αντίστοιχες έξι ισότητες.

2 Αμεσες Μέθοδοι για την Αριθμητική Επίλυση Γραμμικών Συστημάτων

2.1 Εισαγωγή

Θεωρούμε ένα πραγματικό ομαλό γραμμικό σύστημα n εξισώσεων με n αγνώστους. Συγκεκριμένα, το

$$Ax = b, \quad A \in I\!R^{n,n}, \ \det(A) \neq 0, \quad b \in I\!R^n. \tag{2.1}$$

Από τη Γραμμική Αλγεβρα γνωρίζουμε ότι οι τύποι του Cramer δίνουν τη λύση του από τις εκφράσεις:

$$x_i \ = \ \frac{\det(A_i)}{\det(A)}, \ i = 1(1)n, \tag{2.2}$$

όπου A_i είναι ο ίδιος πίνακας με τον A με τη μόνη διαφορά ότι η i-οστή στήλη του έχει αντικατασταθεί από τη στήλη των συνιστωσών του b. Ομως αν θελήσουμε να βρούμε τους αγνώστους x_i θα πρέπει να βρούμε τις τιμές των $n+1$ οριζουσών των τύπων (2.2). Το ανάπτυγμα όμως κάθε μιας από τις ορίζουσες αυτές αποτελείται από $n!$ όρους κάθε ένας από τούς οποίους είναι γινόμενο n παραγόντων. Ετσι, οι απαιτούμενοι πολλαπλασιασμοί είναι πλήθους $(n+1)\times(n-1)\times n! = (n-1)\times(n+1)!$, για την εκτέλεση μόνο των οποίων αν χρησιμοποιήσουμε έναν Υπολογιστή ικανό να εκτελεί 10^6 πολλαπλασιασμούς/δευτερόλεπτο θα απαιτούνταν π.χ. για $n = 20$ περίπου 3×10^5 αιώνες! Αυτό δείχνει και το πρακτικά ανέφικτο της χρησιμοποίησης του κανόνα του Cramer για την εύρεση της λύσης του (2.1).
Ακόμη είναι γνωστό ότι λύση του (2.1) δίνεται από την κλειστή μορφή $x = A^{-1}b$. Επομένως θα μπορούσε να ισχυριστεί κανείς ότι είναι ίσως δυνατόν να βρεθεί η λύση βρίσκοντας πρώτα κατά κάποιον τρόπο τον αντίστροφο του A. Για την εύρεση του A^{-1} ο πιο προφανής τρόπος θα ήταν να εργαστεί κανείς ως εξής:
Εστω $X = A^{-1}$ ή ισοδύναμα $AX = I$. Αν x^i ειναι τα διανύσματα-στήλες του άγνωστου πίνακα X και $e^i, \ i = 1(1)n$, τα αντίστοιχα διανύσματα-στήλες του I, τότε θα ίσχυε $Ax^i = e^i, \ i = 1(1)n$, και επομένως θα καταλήγαμε στην αντίφαση ότι ενώ το αρχικό μας πρόβλημα ήταν η επίλυση ενός γραμμικού συστήματος (του $Ax = b$) για την εύρεση της λύσης του (2.1) τώρα έχουμε να επιλύσουμε n γραμμικά συστήματα (τα $Ax^i = e^i, \ i = 1(1)n$) για την εύρεση του A^{-1} και στη συνέχεια να βρούμε τη λύση του (2.1) από την έκφραση $x = A^{-1}b$

Ως εκ τούτου σκοπός μας θα είναι η ανάπτυξη αποτελεσματικά ρεαλιστικών μεθόδων για την επίλυση του (2.1). Οι μέθοδοι που θα αναπτύξουμε διακρίνονται χοντρικά σε τρεις μεγάλες κατηγορίες. Στις άμεσες, όπου η λύση (με ακριβή αριθμητική) βρίσκεται μετά από ένα πεπερασμένο πλήθος πράξεων, στις έμμεσες ή επαναληπτικές, όπου κατασκευάζεται μία ακολουθία διανυσμάτων που κάτω από ορισμένες προϋποθέσεις, και με ακριβή αριθμητική, συγκλίνει οριακά στην ακριβή λύση, και τέλος στις μικτές όπου κατασκευάζεται πάλι μία ακολουθία διανυσμάτων η οποία όμως

(με ακριβή αριθμητική) συγκλίνει στην ακριβή λύση μετά από n επαναλήψεις. Οι τελευταίες μέθοδοι που έχουν γνωρίσει μεγάλη ανάπτυξη τις δυο τελευταίες δεκαετίες θεωρούνται (ή/και εφαρμόζονται) από ερευνητές και χρήστες ως επαναληπτικές. Το ποια ακριβώς μέθοδος θα εφαρμοστεί σε μια συγκεκριμένη περίπτωση εξαρτιέται από τις ιδιότητες που μπορεί να έχει ο πίνακας των συντελεστών των αγνώστων A, από το υπάρχον λογισμικό, καθώς και από την προϋπάρχουσα εμπειρία.

Στο μεγαλύτερο μέρος των παρουσών Σημειώσεων θα προσπαθήσουμε να περιγράψουμε τις βασικές αρχές από τις πιο αντιπροσωπευτικές μεθόδους επίλυσης γραμμικών συστημάτων. Για περισσότερες και πιο εξειδικευμένες μεθόδους ο αναγνώστης παραπέμπεται στην παρατιθέμενη βιβλιογραφία.

2.2 Μέθοδοι Απαλοιφής Gauss και LU Παραγοντοποίησης

Ας υποθέσουμε ότι δίνεται για επίλυση το παρακάτω πραγματικό γραμμικό σύστημα n αλγεβρικών εξισώσεων με n αγνώστους

$$Ax = b,\ A \in I\!R^{n,n},\ b \in I\!R^n. \tag{2.3}$$

Από τη Γραμμική Αλγεβρα είναι γνωστό ότι αν:

1. Πολλαπλασιάσουμε μια εξίσωση του (2.3) επί έναν αριθμό $\lambda \in I\!R \setminus \{0\}$
2. Εναλλάξουμε τη σειρά δύο εξισώσεων του (2.3)
3. Αντικαταστήσουμε μια εξίσωση του (2.3) με το άθροισμα αυτής και μιας άλλης πολλαπλασιασμένης επί έναν αριθμό $\lambda \in I\!R \setminus \{0\}$

προκύπτει σύστημα

$$A'x = b' \tag{2.4}$$

ισοδύναμο με το αρχικό (2.3).

Η κλασική μέθοδος απαλοιφής του Gauss συνίσταται σε μία συστηματική εφαρμογή των παραπάνω περιγραφεισών ιδιοτήτων των γραμμικών συστημάτων έτσι ώστε το (2.3) να μετατρέπεται στο (2.4), όπου ο A' είναι άνω (ή κάτω) τριγωνικός πίνακας. Δηλαδή, τ.ω. $a'_{ij} = 0,\ \forall\ i > j$ ($a'_{ij} = 0,\ \forall\ i < j$). Στην περίπτωση που ο A' είναι άνω τριγωνικός το (2.4) θα έχει τη μορφή

$$\begin{bmatrix} a'_{11} & a'_{12} & \cdots & \cdots & a'_{1n} \\ & \ddots & & & \\ & & a'_{kk} & \cdots & a'_{kn} \\ & & & \ddots & \\ & & & & a'_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_k \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} b'_1 \\ \vdots \\ b'_k \\ \vdots \\ b'_n \end{bmatrix}, \tag{2.5}$$

οπότε οι τιμές των αγνώστων θα δίνονται από τις εκφράσεις

$$x_k = (b'_k - \sum_{j=k+1}^{n} a'_{kj}x_j)/a'_{kk}, \quad k = n(-1)1. \tag{2.6}$$

Η διαδικασία για την εύρεση των (2.4)-(2.5) από το (2.3) καλείται απαλοιφή ενώ αυτή της εύρεσης των x_k από τις (2.6) προς τα πίσω αντικατάσταση.
Σημείωση: Προφανώς οι τύποι (2.6) ισχύουν $\forall\ k$ ανν $a'_{kk} \neq 0$. Αν για κάποιο k, $a'_{kk} = 0$ τότε αν η αντίστοιχη εξίσωση είναι της μορφής $0 = 0$ ο αντίστοιχος x_k ορίζεται αυθαίρετα και το σύστημα είναι συμβιβαστό (έχει άπειρες λύσεις). Αν έστω και μία από τις εξισώσεις αυτές δεν ικανοποιεί τη $0 = 0$ τότε το σύστημα δεν είναι συμβιβαστό (δεν έχει λύση).

Στη συνέχεια διατυπώνεται μία πρόταση, η απόδειξη της οποίας θα δοθεί αφού πρώτα ξεκαθαριστούν δύο διαφορετικές προσεγγίσεις για την επίλυση ενός γραμμικού συστήματος. Η μία δεν είναι άλλη από την κλασική μέθοδο απαλοιφής του Gauss, παρουσιασμένη όμως με μορφή πινάκων, ενώ η δεύτερη, που αποτελεί μια παραλλαγή της μεθόδου του Gauss, βασίζεται στην παραγοντοποίηση του πίνακα A κατά ένα συγκεκριμένο τρόπο.

Θεώρημα 2.1 *Εστω $A \in I\!R^{n,n}$ και ότι όλοι οι κύριοι $p \times p$, $p = 1(1)n$, υποπίνακες $A_{p\times p}$ (της άνω αριστερής γωνίας) του A είναι αντιστρέψιμοι. Τότε υπάρχει μοναδικός κάτω τριγωνικός πίνακας M με μοναδιαία διαγώνια στοιχεία ($m_{ii} = 1$, $i = 1(1)n$), και μοναδικός αντιστρέψιμος άνω τριγωνικός πίνακας U, τ. ω.*

$$U = MA. \tag{2.7}$$

Αν δε τεθεί $L = M^{-1}$ τότε

$$LU = A, \tag{2.8}$$

όπου ο L είναι κάτω τριγωνικός με $l_{ii} = 1$, $i = 1(1)n$, η δε συγκεκριμένη παραγοντοποίηση (2.8) του A είναι μοναδική.

Κάτω από τις προϋποθέσεις του Θεωρήματος 2.1 η απαλοιφή του Gauss συνίσταται καταρχάς στον πολλαπλασιασμό από τα αριστερά της εξίσωσης (2.3) επί M, οπότε με βάση το πρώτο μέρος του θεωρήματος έχουμε να λύσουμε το ισοδύναμο σύστημα $MAx = Mb$ ή το $Ux = Mb$. Το τελευταίο όμως είναι της μορφής (2.5) και άρα λύνεται με προς τα πίσω αντικατάσταση χρησιμοποιώντας τους τύπους (2.6).
Η επίλυση του δοθέντος συστήματος με τη μέθοδο της παραγοντοποίησης είναι παρόμοια με αυτήν της απαλοιφής του Gauss. Η βασική διαφορά έγκειται στο ότι πρώτα παραγοντοποιούμε τον πίνακα A στο γινόμενο LU, οπότε το αρχικό σύστημα μετασχηματίζεται στο ισοδύναμο $LUx = b$, και στη συνέχεια επιλύνουμε πρώτα το σύστημα $Ly = b$ με προς τα μπρός αντικατάσταση (λόγω της κάτω τριγωνικής μορφής του L) και μετά το $Ux = y$ με προς τα πίσω αντικατάσταση.
Σημείωση: Οι βασικές διαφορές μεταξύ της (κλασικής) μεθόδου απαλοιφής του Gauss και αυτής

της παραγοντοποίησης (ή τριγωνοποίησης) είναι οι εξής: α) Στη μέθοδο απαλοιφής του Gauss ο πίνακας M **δε** βρίσκεται αναλυτικά. Απλά, η διαδικασία της απαλοιφής (μετασχηματισμοί) εφαρμόζεται συγχρόνως στο πρώτο και στο δεύτερο μέλος για να καταλήξουμε σε σύστημα της μορφής (2.5) ισοδύναμο με το αρχικό (2.3). β) Στη μέθοδο της παραγοντοποίησης οι πίνακες L και U βρίσκονται αναλυτικά ακολουθώντας μια διαδικασία απαλοιφής Gauss **μόνο** στον πίνακα A και ακολουθεί η εύρεση του βοηθητικού διανύσματος y που δεν είναι άλλο παρά το διάνυσμα Mb της απαλοιφής του Gauss. γ) Τέλος μπορούμε να παρατηρήσουμε ότι και στις δυο μεθόδους οι προς τα πίσω αντικαταστάσεις για την εύρεση του x είναι ταυτόσημες.

Απόδειξη του Θεωρήματος 2.1: Η απόδειξη θα γίνει με επαγωγή. Θέτουμε $A^{(1)} = A$ και παρατηρούμε ότι $a_{11}^{(1)} \neq 0$. Αυτό διότι από την υπόθεση του θεωρήματος ο 1×1 κύριος υποπίνακας του A είναι αντιστρέψιμος. Το στοιχείο $a_{11}^{(1)}$, που θα εμφανίζεται, όπως θα δούμε, σε παρονομαστές, θα καλείται οδηγός και η αντίστοιχη γραμμή οδηγός γραμμή. Βρίσκουμε στη συνέχεια τους αριθμούς $m_{j1} = a_{j1}^{(1)}/a_{11}^{(1)},\ j = 2(1)n$, που καλούνται πολλαπλασιαστές, πολλαπλασιάζουμε τα στοιχεία της πρώτης γραμμής επί τον πολλαπλασιαστή m_{j1} και αφαιρούμε τα γινόμενα που βρίσκουμε από τα αντίστοιχα στοιχεία της j-οστής γραμμής. Ετσι τα νέα στοιχεία της j-οστής γραμμής θα είναι $a_{jl}^{(2)} = a_{jl}^{(1)} - m_{j1}a_{1l}^{(1)},\ l = 1(1)n$. (Σημείωση: Είναι φανερό από τον ορισμό των πολλαπλασιαστών ότι $a_{j1}^{(2)} = 0,\ j = 2(1)n$.) Η διαδικασία που μόλις εκτελέσαμε είναι εύκολο να διαπιστωθεί ότι ισοδυναμεί με τον πολλαπλασιασμό του πίνακα $A^{(1)}$ από τα αριστερά επί τον πίνακα M_1, όπου

$$M_1 = \begin{bmatrix} 1 & & & & \\ -m_{21} & 1 & & & \\ -m_{31} & & 1 & & \\ \vdots & & & \ddots & \\ -m_{n1} & & & & 1 \end{bmatrix}.$$

Εχουμε, λοιπόν, ότι:

$$A^{(2)} = M_1 A^{(1)}.$$

Υποθέτουμε ότι η προηγουμένως περιγραφείσα και εφαρμοσθείσα διαδικασία απαλοιφής έχει εκτελεστεί με ανάλογο τρόπο $k-1$ $(1 < k < n)$ φορές και βρισκόμαστε στην αρχή της εφαρμογής της για k-οστή φορά, οπότε η κατάσταση παρουσιάζεται ως εξής:

$$A^{(k)} = M_{k-1} \cdots M_2 M_1 A^{(1)},$$

όπου

$$A^{(k)} = \begin{bmatrix} a_{11}^{(1)} & a_{12}^{(1)} & \cdots & a_{1k}^{(1)} & \cdots & a_{1n}^{(1)} \\ & a_{22}^{(2)} & \cdots & a_{2k}^{(2)} & \cdots & a_{2n}^{(2)} \\ & & \ddots & & & \\ & & & a_{kk}^{(k)} & \cdots & a_{kn}^{(k)} \\ & & & a_{k+1,k}^{(k)} & \cdots & a_{k+1,n}^{(k)} \\ & & & & & \\ & & & a_{nk}^{(k)} & \cdots & a_{nn}^{(k)} \end{bmatrix}.$$

Καταρχάς αποδείχνουμε ότι το στοιχείο $a_{kk}^{(k)}$ που θα χρησιμεύσει ως οδηγός στο k βήμα της απαλοιφής δεν είναι μηδέν. Επικεντρώνουμε την προσοχή μας στον $k \times k$ κύριο υποπίνακα $A_{k\times k}^{(k)}$ του $A^{(k)}$. Αυτός έχει προκύψει από τον $k \times k$ κύριο υποπίνακα $A_{k\times k}$ του A μετά από προσθέσεις πολλαπλάσιων γραμμών του σε άλλες γραμμές του. (Διότι ο πίνακας $A^{(k)}$ έχει προκύψει από τον A με τον ίδιο τρόπο.) Λόγω της αντιστρεψιμότητας του $A_{k\times k}$, από την υπόθεση του θεωρήματος, θα έχουμε διαδοχικά $0 \neq \det(A_{k\times k}) = \det(A_{k\times k}^{(k)}) = a_{11}^{(1)} a_{22}^{(2)} \cdots a_{k-1,k-1}^{(k-1)} a_{kk}^{(k)}$, από τις οποίες συμπεραίνεται ότι $a_{kk}^{(k)} \neq 0$. Στη συνέχεια ορίζουμε πολλαπλασιαστές $m_{jk} = a_{jk}^{(k)}/a_{kk}^{(k)},\ j = k+1(1)n$, πολλαπλασιάζουμε τα στοιχεία της k-οστής γραμμής επί m_{jk} και τα γινόμενα αφαιρούμε από τα αντίστοιχα στοιχεία της j-οστής γραμμής. Τα προκύπτοντα νέα στοιχεία δίνονται από τις σχέσεις $a_{jl}^{(k+1)} = a_{jl}^{(k)} - m_{jk} a_{kl}^{(k)},\ j = k+1(1)n,\ l = k+1(1)n$, ενώ $a_{jk}^{(k+1)} = 0,\ j = k+1(1)n$, από τον ορισμό των αντίστοιχων πολλαπλασιαστών. Είναι δυνατόν να διαπιστωθεί εύκολα ότι η διαδικασία του k βήματος της απαλοιφής μπορεί να διατυπωθεί με μορφή πινάκων ως εξής

$$A^{(k+1)} = M_k A^{(k)},$$

όπου

$$M_k = \begin{bmatrix} 1 & & & & & & \\ & 1 & & & & & \\ & & \ddots & & & & \\ & & & 1 & & & \\ & & & -m_{k+1,k} & 1 & & \\ & & & -m_{k+2,k} & & 1 & \\ & & & \vdots & & & \ddots \\ & & & -m_{nk} & & & 1 \end{bmatrix}$$

και άρα

$$A^{(k+1)} = M_k M_{k-1} \cdots M_2 M_1 A^{(1)}.$$

Με βάση την αρχή της τέλειας επαγωγής ό,τι αποδείχτηκε θα ισχύει και για κάθε τιμή του $k < n$. Θα έχουμε λοιπόν ότι

$$A^{(n)} = M_{n-1} M_{n-2} \cdots M_2 M_1 A \tag{2.9}$$

ή αν θέσουμε

$$U = A^{(n)} \ \kappa\alpha\iota \ M = M_{n-1}M_{n-2}\cdots M_2M_1 \tag{2.10}$$

τότε από τις (2.9) και (2.10) προκύπτει η (2.7). Προφανώς ο πίνακας $U = A^{(n)}$ που προέκυψε μετά από τα $n-1$ βήματα της απαλοιφής θα είναι άνω τριγωνικός και αντιστρέψιμος, αφού $\det(U) = \det(A^{(n)}) = \det(A) \neq 0$, ο δε M, από την (2.10), ως γινόμενο $n-1$ κάτω τριγωνικών πινάκων με διαγώνια στοιχεία μονάδες, θα έχει κι αυτός την ίδια μορφή. Τέλος αν τεθεί $L = M^{-1}$, ο L θα είναι κάτω τριγωνικός με διαγώνια στοιχεία μονάδες και θα ισχύει η (2.8). Η μοναδικότητα των L (και άρα και του M) και U αποδείχνεται με άτοπη απαγωγή. Εστω ότι υπάρχει και άλλο ζεύγος κάτω τριγωνικού πίνακα L_1, με διαγώνια στοιχεία μονάδες, και αντιστρέψιμου άνω τριγωνικού πίνακα U_1 τ.ω. $L_1U_1 = A$. Αρα θα ισχύει $L_1U_1 = LU$. Πολλαπλασιάζοντας την τελευταία ισότητα από αριστερά επί L^{-1} και από δεξιά επί U_1^{-1} έχουμε $L^{-1}L_1 = UU_1^{-1}$. Προφανώς το αριστερά μέλος είναι κάτω τριγωνικός πίνακας και το δεξιά άνω τριγωνικός. Επειδή οι πίνακες είναι ίσοι πρέπει να είναι διαγώνιοι. Ομως ο διαγώνιος $L^{-1}L_1$ είναι γινόμενο κάτω τριγωνικών με διαγώνια στοιχεία μονάδες άρα θα έχει κι αυτός διαγώνια στοιχεία μονάδες και συνεπώς θα είναι ο μοναδιαίος I. Από την $L^{-1}L_1 = I$ και στη συνέχεια από τη $UU_1^{-1} = I$ έπεται ότι $L_1 = L$ και $U_1 = U$.

Παρατήρηση: Από τη (2.2) μπορεί εύκολα να δειχτεί ότι

$$M_k^{-1} = \begin{bmatrix} 1 & & & & & & \\ & 1 & & & & & \\ & & \ddots & & & & \\ & & & 1 & & & \\ & & & m_{k+1,k} & 1 & & \\ & & & m_{k+2,k} & & 1 & \\ & & & \vdots & & & \ddots \\ & & & m_{nk} & & & & 1 \end{bmatrix}$$

και από τη (2.10) ότι

$$L = M_1^{-1}M_2^{-1}\cdots M_{n-1}^{-1} = \begin{bmatrix} 1 & & & & & & \\ m_{21} & 1 & & & & & \\ m_{31} & m_{32} & 1 & & & & \\ m_{41} & m_{42} & m_{43} & 1 & & & \\ \vdots & \vdots & \vdots & \vdots & \ddots & & \\ m_{n-1,1} & m_{n-1,2} & m_{n-1,3} & m_{n-1,4} & \cdots & 1 & \\ m_{n1} & m_{n2} & m_{n3} & m_{n4} & \cdots & m_{n,n-1} & 1 \end{bmatrix}$$

Πόρισμα 2.1 *(Παραγοντοποίηση (Ανάλυση)* Crout*) Κάτω από τις προϋποθέσεις του Θεωρήματος 2.1 ο πίνακας A μπορεί να γραφτεί και ως*

$$A \ = \ LDU, \tag{2.11}$$

όπου L είναι κάτω τριγωνικός με $l_{ii} = 1,\ i = 1(1)n$, D διαγώνιος και U άνω τριγωνικός με $u_{ii} = 1, i = 1(1)n$. Η ανάλυση (2.11) είναι μοναδική.

Σημείωση: Προφανώς ο πίνακας L στη (2.11) είναι ο πίνακας L του Θεωρήματος 2.1, ο πίνακας D έχει στοιχεία τα αντίστοιχα διαγώνια στοιχεία του U του Θεωρήματος 2.1 και ο U στη (2.11) έχει στοιχεία τα αντίστοιχα στοιχεία των γραμμών του U του Θεωρήματος 2.1 διαιρεμένα δια του αντίστοιχου οδηγού.

2.2.1 Αλγόριθμος Απαλοιφής του Gauss

Για την επίλυση του γραμμικού συστήματος (2.1) κάτω από τις υποθέσεις του Θεωρήματος 2.1 είναι δυνατόν να δοθεί σε μορφή ψευδοκώδικα ο (κλασικός) αλγόριθμος απαλοιφής του Gauss. Οπως ήδη τονίστηκε η μόνη διαφορά του από την παραγοντοποίηση του πίνακα A έγκειται στο γεγονός ότι οι μετασχηματισμοί των δεύτερων μελών (συνιστωσών του διανύσματος b) γίνονται ταυτόχρονα με το μετασχηματισμό του πίνακα A κάνοντας χρήση των πολλαπλασιαστών m_{jk}. Ετσι έχουμε:

Αλγόριθμος Απαλοιφής Gauss:

Δεδομένα: $A \in I\!R^{n,n}$, $\det(A) \neq 0$, $b \in I\!R^n$ και όλοι οι κύριοι υποπίνακες της άνω αριστερής γωνίας του A είναι αντιστρέψιμοι.
Για $k = 1(1)n-1$ (βήματα απαλοιφής)
  Για $j = k+1(1)n$
    $m_{jk} = a_{jk}^{(k)} / a_{kk}^{(k)}$
    Για $l = k+1(1)n$
      $a_{jl}^{(k+1)} = a_{jl}^{(k)} - m_{jk} a_{kl}^{(k)}$
    Τέλος 'Για'
    $b_j^{(k+1)} = b_j^{(k)} - m_{jk} b_k^{(k)}$
  Τέλος 'Για'
Τέλος 'Για'

Μετά την απαλοιφή του Gauss όπως δόθηκε στον παραπάνω αλγόριθμο η εύρεση των αγνώστων γίνεται με προς τα πίσω αντικατάσταση χρησιμοποιώντας τους τύπους (2.6). Ο αντίστοιχος αλγόριθμος σε ψευδοκώδικα δίνεται στη συνέχεια:

Αλγόριθμος Προς τα Πίσω Αντικατάστασης:

Για $k = n(-1)1$
  $s_k = b_k^{(k)}$
  Για $j = k+1(1)n$

$s_k = s_k - a_{kj}^{(k)} x_j$

Τέλος 'Για'

$x_k = s_k / a_{kk}^{(k)}$

Τέλος 'Για'

Παρατηρήσεις: α) Οι πάνω δείκτες στους δυο αλγόριθμους που δόθηκαν είναι μόνο ενδεικτικοί και μπορούν να παραλειφτούν. Ετσι τα $a_{jl}^{(k)}$ μπορούν να αποθηκευτούν στις αντίστοιχες θέσεις των αρχικών a_{jl}. Αν ο πίνακας A χρειάζεται για παραπέρα χρήση τότε μπορεί να αποθηκευτεί από την αρχή και σε κάποιες άλλες θέσεις μνήμης. β) Οι πολλαπλασιαστές m_{jk} χρησιμοποιούνται άπαξ. Αν δε χρειάζονται για παραπέρα χρήση οι κάτω δείκτες μπορούν να παραλειφτούν. Τέλος, αν οι πολλαπλασιαστές χρειάζονται στη συνέχεια ενώ ο πίνακας A δε χρειάζεται τότε οι m_{jk} μπορούν να αποθηκευτούν στις αντίστοιχες θέσεις των στοιχείων a_{jk}. γ) Αντίστοιχες παρατηρήσεις ισχύουν και για το διάνυσμα b. Δηλαδή, αν δε χρειάζεται για παραπέρα χρήση τα $b_j^{(k)}$ μπορούν να αποθηκευτούν στις αντίστοιχες θέσεις των b_j και τέλος, ο δείκτης k στο s_k δε χρειάζεται ενώ οι ευρισκόμενες τιμές των αγνώστων x_j μπορούν να αποθηκευτούν στις θέσεις των b_j. δ) Τονίζεται ότι εφεξής, στους Αλγόριθμους που θα παρουσιάζονται, θα παραλείπονται οι άνω δείκτες. Αν και υπάρχουν πίνακες A για τους οποίους ισχύουν οι υποθέσεις του Θεωρήματος 2.1 είναι πιο φυσικό να γνωρίζουμε μόνο ότι ο A αντιστρέφεται. Στην περίπτωση αυτή ισχύει η παρακάτω παραλλαγή του προαναφερθέντος θεωρήματος.

Θεώρημα 2.2 *Εστω το γραμμικό σύστημα (2.3) για το οποίο ισχύει ότι* $\det(A) \neq 0$. *Τότε υπάρχει μεταθετικός πίνακας* $P \in I\!R^{n,n}$, *αντιστρέψιμος άνω τριγωνικός πίνακας* $U \in I\!R^{n,n}$ *και κάτω τριγωνικος πίνακας* $L \in I\!R^{n,n}$ *με* $l_{ii} = 1,\ i = 1(1)n$, *τ.ω.*

$$LU = PA. \tag{2.12}$$

Εκτός από την πιθανή δυνατότητα επιλογής διαφορετικών μεταθετικών πινάκων η ανάλυση (παραγοντοποίηση) (2.12) είναι μοναδική.

Απόδειξη: Η απόδειξη ακολουθεί τα βήματα της απόδειξης του Θεωρήματος 2.1. Συγκεκριμένα, ακολουθιέται η διαδικασία της απαλοιφής στον πίνακα A, όπως αυτή έχει ήδη περιγραφτεί. Αν για όλους τους οδηγούς ισχύει $a_{kk}^{(k)} \neq 0,\ k = 1(1)n-1$, τότε $a_{nn}^{(n)} \neq 0$, αφού $\det(A) \neq 0$, και το θεώρημα ισχύει με $P = I$. Αν για κάποιο $k \in \{1, 2, \cdots, n-1\}$ ισχύει $a_{kk}^{(k)} = 0$, τότε θα υπάρχει τουλάχιστον ένας δείκτης $j \in \{k+1, k+2, \cdots, n\}$ για τον οποίο $a_{jk}^{(k)} \neq 0$. Αυτό, γιατί σε ενάντια περίπτωση θα είχαμε

$$\det(A) = \det(A^{(k)}) = a_{11}^{(1)} a_{22}^{(2)} \cdots a_{k-1,k-1}^{(k-1)} \det \left(\begin{bmatrix} a_{kk}^{(k)} & \cdots & a_{kn}^{(k)} \\ \vdots & & \vdots \\ a_{nk}^{(k)} & \cdots & a_{nn}^{(k)} \end{bmatrix} \right) = 0,$$

αφού $a_{jk}^{(k)} = 0,\ j = k(1)n$, που οδηγεί σε άτοπο. Στην περίπτωση αυτή εναλλάσσουμε την k-οστή γραμμή με μία από τις γραμμές για την οποία $a_{jk}^{(k)} \neq 0$, έστω με αυτήν που αντιστοιχεί στη μικρότερη τιμή του j. Ομως η εναλλαγή των γραμμών k και $j \in \{k+1, k+2, \cdots, n\}$ πετυχαίνεται με έναν από αριστερά πολλαπλασιασμό του πίνακα A επί ένα μεταθετικό πίνακα, συγκεκριμένα του μοναδιαίου στον οποίο έχουν αντιμετατεθεί οι γραμμές k και j. Στη συνέχεια η απαλοιφή συνεχίζεται με τον ίδιο τρόπο και με το νέο οδηγό, που δεν είναι τώρα μηδέν, κ.λπ. Αν, όμως, συνέβαινε να γνωρίζουμε πρίν από την έναρξη της διαδικασίας της απαλοιφής σε ποιες γραμμές θα συναντούσαμε μηδενικούς οδηγούς και με ποιες γραμμές θα έπρεπε να εναλλάξουμε αυτές ώστε να εξασφαλιστεί η μή μηδενικότητα των νέων οδηγών τότε θα μπορούσαμε εξαρχής να πολλαπλασιάσουμε από τα αριστερά τον αρχικό πίνακα A με κατάλληλο μεταθετικό πίνακα, έστω P, ώστε να εξασφαλίζεται η μή μηδενικότητα των οδηγών του πίνακα PA. Τότε ο πίνακας PA, σύμφωνα με το Θεώρημα 2.1, θα επιδεχόταν μονοσήμαντη ανάλυση σε γινόμενο LU. Αυτό αποδείχνει το θεώρημα. □

Σημειώσεις: α) Υπενθυμίζεται ότι ένας μεταθετικός πίνακας P αποτελεί μια μετάθεση των στηλών (ή των γραμμών) του μοναδιαίου πίνακα I. Π.χ., αν $e^i,\ i = 1(1)n$, με $e^i_j = \delta_{ij},\ j = 1(1)n$, το δέλτα του Kronecker, και I είναι ο μοναδιαίος πίνακας, τότε ο I θα γράφεται

$$I = [e^1\ e^2\ \cdots\ e^n] = \begin{bmatrix} e^{1T} \\ e^{2T} \\ \vdots \\ e^{nT} \end{bmatrix}.$$

Επομένως αν $(i_1\ i_2\ \cdots\ i_n)$ αποτελεί μια μετάθεση των $(1\ 2\ \cdots\ n)$ τότε η αντίστοιχη μετάθεση των γραμμών του I δίνει $P_1 = \begin{bmatrix} e^{i_1 T} \\ e^{i_2 T} \\ \vdots \\ e^{i_n T} \end{bmatrix}$ ενώ η αντίστοιχη μετάθεση των στηλών του I δίνει $P_2 = [e^{i_1}\ e^{i_2}\ \cdots\ e^{i_n}]$. Να σημειωθεί ότι γενικά $P_1 \neq P_2$. β) Υπενθυμίζεται ότι ο πολλαπλασιασμός από τα αριστερά ενός πίνακα A επί τον παραπάνω μεταθετικό πίνακα P_1 μεταθέτει τις γραμμές του A έτσι ώστε οι γραμμές του υπ' αριθμόν 1, 2, ..., n, αντικαθιστούνται στη σειρά από τις γραμμές υπ' αριθμόν i_1, i_2, ..., i_n, ενώ ο από τα δεξιά πολλαπλασιασμός του A επί P_2 αντικαθιστά τις στήλες του A, 1, 2, ..., n, από τις στήλες του i_1, i_2, ..., i_n. γ) Ακόμη, ότι $P^{-1} = P^T$ και ότι το γινόμενο μεταθετικών πινάκων είναι μεταθετικός πίνακας.

Ο αλγόριθμος της απαλοιφής του Gauss για την επίλυση του γραμμικού συστήματος (2.3) με την επιπλέον υπόθεση ότι $\det(A) \neq 0$ έτσι ώστε να αντιμετωπίζεται και η περίπτωση πιθανών εναλλαγών γραμμών (εξισώσεων) θα δοθεί μετά την εισαγωγή της Μερικής Οδήγησης στην Παράγραφο 2.3. Εδώ απλά τονίζεται ότι δε χρειάζεται στην πραγματικότητα να εναλλάσσουμε τις αντίστοιχες γραμμές. Αρκεί να γίνεται καταγραφή των αντίστοιχων εναλλαγών τους. Για το σκοπό αυτό θα χρησιμοποιηθεί βοηθητικό διάνυσμα που θα καταγράφει τις εναλλαγές, όπως θα δούμε αργότερα.

2.2.2 Λύση Γραμμικών Συστημάτων με τον Ιδιο Πίνακα Συντελεστών Αγνώστων

Εστω ότι έχουμε να επιλύσουμε m (> 1) γραμμικά συστήματα με τον ίδιο πίνακα συντελεστών αγνώστων A. Συγκεκριμένα τα

$$Ax_l = b_l, \quad A \in I\!R^{n,n},\ b_l \in I\!R^n,\ l = 1(1)m,$$

ή, ισοδύναμα, την εξίσωση πινάκων

$$AX = B, \quad X = [x_1\ x_2\ \cdots\ x_m],\ B = [b_1\ b_2\ \cdots\ b_m]. \tag{2.13}$$

Είναι προφανές ότι μπορούμε πάλι να εργαστούμε με δυο διαφορετικούς τρόπους. Είτε παραγοντοποιώντας τον A και λύνοντας στη συνέχεια m γραμμικά συστήματα με διαδοχικές προς τα μπρός και προς τα πίσω αντικαταστάσεις είτε εφαρμόζοντας την απαλοιφή του Gauss όχι μόνο στον A αλλά συγχρόνως και σε όλα τα δεύτερα μέλη. Ενα κλασικό πρόβλημα της παραπάνω μορφής είναι εκείνο όπου ζητιέται αναλυτικά ο A^{-1}. Στην περίπτωση αυτή θα έχουμε στη (2.13) $B = I$ και $X = A^{-1}$.

2.2.3 Πλήθος και Είδος Απαιτούμενων Πράξεων για την Επίλυση του $Ax = b$

Για την απλούστευση του προβλήματος υποθέτουμε ότι $a_{kk}^{(k)} \neq 0,\ k = 1(1)n$, έτσι ώστε το σύστημα να έχει μοναδική λύση και να μην απαιτούνται εναλλαγές γραμμών κατά τη διαδικασία της απαλοιφής. Διακρίνουμε τέσσερις διαφορετικές φάσεις στη μέθοδο της επίλυσης είτε αυτή γίνεται με την κλασική μέθοδο απαλοιφής του Gauss είτε πραγματοποιείται πρώτα η παραγοντοποίηση του A. Στην πρώτη φάση υπολογίζουμε τις πράξεις που απαιτούνται για την εύρεση όλων των πολλαπλασιαστών $m_{jk} = a_{jk}^{(k)}/a_{kk}^{(k)},\ j = k+1(1)n,\ k = 1(1)n-1$. Στη δεύτερη, βρίσκουμε τις πράξεις που απαιτούνται στη διαδικασία της απαλοιφής όταν εργαζόμαστε μόνο στον πίνακα A, όπου $a_{jl}^{(k+1)} = a_{jl}^{(k)} - m_{jk}a_{kl}^{(k)},\ j,l = k+1(1)n,\ k = 1(1)n-1$. Στην τρίτη, εκτελούμε τις πράξεις της απαλοιφής στο δεύτερο μέλος b, όπου $b_j^{(k+1)} = b_j^{(k)} - m_{jk}b_k^{(k)},\ j = k+1(1)n,\ k = 1(1)n-1$ (ή τις ισοδύναμες πράξεις κατά την προς τα μπρός αντικατάσταση όταν παραγοντοποιείται πρώτα ο A, $y_k = b_k - \sum_{j=1}^{k-1} m_{kj}y_j,\ k = 1(1)n$) και τέλος τις πράξεις που απαιτεί η προς τα πίσω αντικατάσταση, όπου $x_k = (b_k^{(k)} - \sum_{j=k+1}^{n} a_{kj}^{(k)}x_j)/a_{kk}^{(k)},\ k = n(-1)1$, ή $x_k = (y_k - \sum_{j=k+1}^{n} a_{kj}^{(k)}x_j)/a_{kk}^{(k)},\ k = n(-1)1$, όταν έχει προηγηθεί προς τα μπρός αντικατάσταση.

α) Εύρεση πολλαπλασιαστών: Για την εύρεση ενός πολλαπλασιαστή απαιτείται μόνο μία διαίρεση. Αρα για την εύρεση όλων των πολλαπλασιαστών απαιτούνται $\sum_{k=1}^{n-1}(n-k) = \sum_{k=1}^{n-1} k = n(n-1)/2$ διαιρέσεις. Επειδή ο παρονομαστής $a_{kk}^{(k)},\ k = 1(1)n-1$, είναι κοινός σε $n-k$ πολλαπλασιαστές είναι προτιμότερο να βρίσκεται πρώτα ο $1/a_{kk}^{(k)}$ μια φορά και μετά να ακολουθιέται από $n-k$ πολλαπλασιασμούς για την εύρεση των m_{jk} $\left(= \frac{1}{a_{kk}^{(k)}}.a_{jk}^{(k)}\right),\ j = k+1(1)n$. Ετσι, για την εύρεση όλων των πολλαπλασιαστών απαιτούνται $n-1$ διαιρέσεις και $n(n-1)/2$ πολλαπλασιασμοί.

Απαλοιφή Gauss	Παραγοντοποίηση LU	Πολλ/σμοί	Προσθέσεις	Διαιρέσεις
Πολλαπλασιαστές		$\frac{n(n-1)}{2}$		$n-1$
Απαλοιφή σε A		$\frac{n(n-1)(2n-1)}{6}$	$\frac{n(n-1)(2n-1)}{6}$	
Απαλοιφή σε b	Προς τα Μπρός Αντικαταστάσεις	$\frac{n(n-1)}{2}$	$\frac{n(n-1)}{2}$	
Προς τα Πίσω Αντικαταστάσεις		$\frac{n(n-1)}{2}$	$\frac{n(n-1)}{2}$	n
Σύνολο		$\frac{n(n-1)(n+4)}{3}$	$\frac{n(n-1)(2n+5)}{6}$	$2n-1$

Πίνακας 1: Είδος και Πλήθος Πράξεων για τη Λύση του $Ax = b$

β) Εύρεση νέων στοιχείων των πινάκων $A^{(k+1)},\ k = 1(1)n-1$, κατα την απαλοιφή: Για την εύρεση ενός νέου στοιχείου $a_{jl}^{(k+1)}$ απαιτείται ένας πολλαπλασιασμός και μία αφαίρεση (πρόσθεση). Για την εύρεση όλων των στοιχείων όλων των πινάκων απαιτούνται $\sum_{k=1}^{n-1}(n-k)^2 = \sum_{k=1}^{n-1} k^2 = n(n-1)(2n-1)/6$ πολλαπλασιασμοί και άλλες τόσες προσθέσεις.

γ) Εύρεση νέων στοιχείων δεύτερων μελών $b^{(k+1)}$: Για την εύρεση ενός νέου στοιχείου $b_j^{(k+1)}$ απαιτείται ένας πολλαπλασιασμός και μία αφαίρεση (πρόσθεση). Για την εύρεση όλων των νέων στοιχείων όλων των δεύτερων μελών απαιτούνται $\sum_{k=1}^{n-1}(n-k) = \sum_{k=1}^{n-1} k = n(n-1)/2$ πολλαπλασιασμοί και άλλες τόσες προσθέσεις. Αν ακολουθιέται η διαδικασία της προς τα μπρός αντικατάστασης, τότε για την εύρεση ενός από τους βοηθητικούς αγνώστους $y_k,\ k = 1(1)n$, απαιτούνται $k-1$ πολλαπλασιασμοί και άλλες τόσες προσθέσεις. Για την εύρεση όλων των y_k απαιτούνται $\sum_{k=1}^{n}(k-1) = n(n-1)/2$ πολλαπλασιασμοί και άλλες τόσες προσθέσεις. Οπως διαπιστώνει κανείς οι πράξεις για την εύρεση του $b^{(n)}$ ή του y είναι ακριβώς οι ίδιες. (Με μια προσεκτικότερη ματιά διαπιστώνεται εύκολα ότι όχι μόνο $y = b^{(n)}$, πράγμα που αναμένεται, αλλά και ότι οι πράξεις για την εύρεσή τους είναι ακριβώς οι ίδιες με τη μόνη διαφορά ότι εκτελούνται με διαφορετική σειρά.)

δ) Εύρεση τιμών αγνώστων: Η προς τα πίσω αντικατάσταση απαιτεί n βήματα και σε κάθε ένα από αυτά η τιμή ενός αγνώστου x_k προσδιορίζεται από τις τιμές των αγνώστων που έχουν ήδη βρεθεί. Για την εύρεση ενός αγνώστου απαιτούνται $n-k$ πολλαπλασιασμοί και άλλες τόσες προσθέσεις καθώς και μία διαίρεση. Για τις τιμές όλων των αγνώστων απαιτούνται $\sum_{k=1}^{n}(n-k) = \sum_{k=1}^{n-1} k = n(n-1)/2$ πολλαπλασιασμοί και άλλες τόσες προσθέσεις καθώς και n διαιρέσεις. Οι επί μέρους πράξεις φαίνονται αναλυτικά στον Πίνακα 1.

Από τον Πίνακα 1 καθίσταται φανερό ότι οι περισσότερο χρονοβόρες πράξεις κατά την επίλυση ενός γραμμικού συστήματος είναι οι πολλαπλασιασμοί που απαιτούνται για την απαλοιφή του Gauss ή αντίστοιχα για την παραγοντοποίηση του πίνακα A. Αν υποθέσουμε ότι $n \to \infty$, ή στην πράξη ότι το n είναι πολύ μεγάλο, τότε μπορούμε να πούμε ότι το πλήθος των απαιτούμενων πράξεων

(πολλαπλασιασμών) είναι περίπου $\frac{1}{3}n^3$ ή ότι είναι της τάξης $\mathcal{O}(n^3)$ με ασυμπτωτικό συντελεστή $\frac{1}{3}$. Σε αντιπαραβολή με την παραγοντοποίηση του A, για τη διαδικασία της απαλοιφής στο δεύτερο μέλος (ή για τις προς τα μπρός αντικαταστάσεις) καθώς και για τις προς τα πίσω αντικαταστάσεις απαιτούνται για την καθεμία διαδικασία αντίστοιχες πράξεις πλήθους περίπου $\frac{1}{2}n^2$ ή τάξης $\mathcal{O}(n^2)$ με ασυμπτωτικό συντελεστή $\frac{1}{2}$. Με βάση τα παραπάνω μπορεί να βρεί κανείς αμέσως ότι το πλήθος των πράξεων (πολλαπλασιασμών) για την εύρεση του αντίστροφου πίνακα με τη μέθοδο της παραγοντοποίησης του πίνακα A είναι περίπου $\frac{1}{3}n^3+2n\frac{1}{2}n^2 = \frac{4}{3}n^3$ ή ότι είναι τάξης $\mathcal{O}(n^3)$ με ασυμπτωτικό συντελεστή $\frac{4}{3}$.

2.2.4 Αναγώγιμοι και Μή-Αναγώγιμοι Πίνακες

Ορισμός 2.1 *Ενας πίνακας $A \in \mathbb{C}^{n,n}$ καλείται αναγώγιμος ανν υπάρχει μεταθετικός πίνακας $P \in \mathbb{R}^{n,n}$ τ.ω. ο PAP^T να έχει τη μορφή*

$$PAP^T \;=\; \begin{bmatrix} B & C \\ O & D \end{bmatrix}, \tag{2.14}$$

όπου $B \in \mathbb{C}^{r,r},\ 1 \le r \le n-1$ και $O \in \mathbb{C}^{n-r,r}$ ο μηδενικός πίνακας.

Ορισμός 2.2 *Ενας πίνακας $A \in \mathbb{C}^{n,n}$ καλείται μή-αναγώγιμος ανν* **δεν** *είναι αναγώγιμος.*

Οπως μπορεί αμέσως να διαπιστωθεί, ένα γραμμικό σύστημα $Ax = b$, $A \in \mathbb{R}^{n,n}$, $b \in \mathbb{R}^n$, με πίνακα συντελεστών αγνώστων A αναγώγιμο της μορφής (2.14) είναι δυνατόν να γραφτεί ως

$$PAP^T(Px) \;=\; \begin{bmatrix} B & C \\ O & D \end{bmatrix} \begin{bmatrix} (Px)_1 \\ (Px)_2 \end{bmatrix} \;=\; \begin{bmatrix} (Pb)_1 \\ (Pb)_2 \end{bmatrix},$$

όπου $(Px)_1, (Pb)_1 \in \mathbb{R}^{r,r}$ και $(Px)_2, (Pb)_2 \in \mathbb{R}^{n-r,n-r}$. Το τελευταίο σύστημα είναι ισοδύναμο με τα δυο μικρότερων διαστάσεων συστήματα

$$B(Px)_1 + C(Px)_2 \;=\; (Pb)_1, \quad D(Px)_2 \;=\; (Pb)_2.$$

Από τα δύο τελευταία αυτά συστήματα, βρίσκεται πρώτα ο άγνωστος $(Px)_2$ από το δεύτερο σύστημα και με αντικατάσταση στο πρώτο ο $(Px)_1$. Από τους $(Px)_1$ και $(Px)_2$ βρίσκεται τελικά ο x. Μπορεί να αποδειχτεί ότι για την επίλυση των δύο συστημάτων, όπως περιγράφτηκε, απαιτείται μικρότερο πλήθος πράξεων από αυτό της επίλυσης του δοθέντος αρχικού συστήματος.
Σημείωση: Στο εξής θα υποθέτουμε ότι ο πίνακας των συντελεστών των αγνώστων είναι μή-αναγώγιμος εκτός και αν ορίζεται διαφορετικά.

2.2.5 Πυκνοί και Αραιοί Πίνακες

Πέρα από την ακρίβεια και την οικονομία των πράξεων που επιζητούμε στους αλγόριθμους της Αριθμητικής Ανάλυσης ένα τρίτο στοιχείο είναι και αυτό της οικονομίας στην αποθήκευση στον Υπολογιστή των δεδομένων καθώς και των ενδιάμεσων και τελικών αποτελεσμάτων του προς επίλυση προβλήματος. Στην περίπτωση ενός πίνακα $A \in I\!R^{n,n}$ οι απαιτούμενες θέσεις μνήμης είναι προφανώς n^2. Οταν όμως ο A έχει κάποια ειδική μορφή, π.χ. είναι συμμετρικός, έχει τα περισσότερα στοιχεία του 0, κ.λπ., θα πρέπει οι συγκεκριμένες ιδιότητες του A να λαβαίνονται σοβαρά υπόψη στην κατασκευή του αλγόριθμου ώστε να χρησιμοποιούνται πιθανόν πολύ λιγότερες θέσεις μνήμης για την αποθήκευση του πίνακα. Ενας πίνακας A με πλήθος μή μηδενικών στοιχείων της τάξης $\mathcal{O}(n^2)$ καλείται πυκνός και γενικά απαιτεί n^2 θέσεις μνήμης για την αποθήκευσή του. Ενας πίνακας με πλήθος μή μηδενικών στοιχείων της τάξης $\mathcal{O}(n)$ καλείται αραιός. Αραιοί πίνακες προκύπτουν στην πράξη, μεταξύ άλλων, από προβλήματα διακριτοποίησης (Μερικών) Διαφορικών Εξισώσεων. Σε πολλές τέτοιες περιπτώσεις οι αραιοί πίνακες έχουν μια συγκεκριμένη δομή. Μια από τις περιπτώσεις δομημένης αραιότητας αποτελούν και οι καλούμενοι πίνακες ζώνης (ή ταινιωτοί). Ενας πίνακας $A \in I\!R^{n,n}$ (ή και $A \in \mathbb{C}^{n,n}$) καλείται πίνακας ζώνης ανν $a_{ij} = 0,\ \forall\ i,j = 1(1)n$, για τα οποία $i - j > n_a$ ή $j - i > n_b$, όπου $n_a,\ n_b\ <\ n$ σταθεροί φυσικοί ακέραιοι (με $n_a + n_b + 1 < n$). Π.χ. για $n_a = 2$ και $n_b = 1$ ένας τέτοιος πίνακας θα έχει την παρακάτω μορφή

$$\begin{bmatrix} a_{11} & a_{12} & & & & & & \\ a_{21} & a_{22} & a_{23} & & & & & \\ a_{31} & a_{32} & a_{33} & a_{34} & & & & \\ & a_{42} & a_{43} & a_{44} & a_{45} & & & \\ & & \ddots & \ddots & \ddots & \ddots & & \\ & & & a_{n-2,n-4} & a_{n-2,n-3} & a_{n-2,n-2} & a_{n-2,n-1} & \\ & & & & a_{n-1,n-3} & a_{n-1,n-2} & a_{n-1,n-1} & a_{n-1,n} \\ & & & & & a_{n,n-2} & a_{n,n-1} & a_{nn} \end{bmatrix}. \tag{2.15}$$

Ο αριθμός $n_a + n_b + 1$ καλείται εύρος της ζώνης. Πιο συγκεκριμένα, για $n_a = n_b = 1$ ο πίνακας καλείται τριδιαγώνιος, για $n_a = n_b = 2$ πενταδιαγώνιος κ.λπ. Είναι φανερό ότι για την αποθήκευση ενός πίνακα ζώνης αρκούν το πολύ $n(n_a + n_b + 1)$ θέσεις μνήμης. Π.χ. ένας τρόπος αποθήκευσης του πίνακα που δόθηκε στη (2.15) είναι ο ακόλουθος:

$$\begin{bmatrix} & a_{12} & a_{23} & \cdots & a_{n-3,n-2} & a_{n-2,n-1} & a_{n-1,n} \\ a_{11} & a_{22} & a_{33} & \cdots & a_{n-2,n-2} & a_{n-1,n-1} & a_{nn} \\ a_{21} & a_{32} & a_{43} & \cdots & a_{n-1,n-2} & a_{n,n-1} & \\ a_{31} & a_{42} & a_{53} & \cdots & a_{n,n-2} & & \end{bmatrix}.$$

ΑΣΚΗΣΕΙΣ

1.: Να λυθεί το παρακάτω σύστημα με απλή απαλοιφή, όπως δηλαδή περιγράφεται στις (2.3)–(2.5) και (2.6)

$$\begin{array}{rcrcrcrcl} x_1 & + & x_2 & + & x_3 & + & x_4 & = & 4 \\ 2x_1 & + & 2x_2 & +2 & x_3 & + & 2x_4 & = & 8 \\ x_1 & + & x_2 & + & 2x_3 & + & 3x_4 & = & 7 \\ 2x_1 & + & 2x_2 & + & x_3 & + & 2x_4 & = & 7 \end{array}$$

2.: Να δοθούν οι πίνακες, οι οποίοι, πολλαπλασιάζοντας τα μέλη του γραμμικού συστήματος $Ax = b$, $A \in \mathbb{C}^{n,n}$, $b \in \mathbb{C}^n$, από αριστερά, παράγουν ισοδύναμο σύστημα $A'x = b'$, και εκφράζουν έκαστος καθεμιά από τις πράξεις (1), (2) και (3), που δίνανται αμέσως μετά το αρχικό σύστημα (2.1).

3.: Αν

$$A_1 = \begin{bmatrix} 1 & & & & \\ a_{21} & 1 & & & \\ a_{31} & & 1 & & \\ \vdots & & & \ddots & \\ a_{n1} & & & & 1 \end{bmatrix} \quad \kappa\alpha\iota \quad A_2 = \begin{bmatrix} 1 & & & & \\ & 1 & & & \\ & a_{32} & 1 & & \\ & \vdots & & \ddots & \\ & a_{n2} & & & 1 \end{bmatrix}, \quad A_1, A_2 \in \mathbb{C}^{n,n},$$

με $n > 2$, να βρεθούν τα γινόμενα A_1A_2 και A_2A_1.

4.: Να δειχτεί ότι ο αντίστροφος ενός αντιστρέψιμου κάτω τριγωνικού πίνακα στο $\mathbb{C}^{n,n}$ είναι (αντιστρέψιμος) κάτω τριγωνικός πίνακας με διαγώνια στοιχεία τα αντίστροφα των αντίστοιχων διαγώνιων στοιχείων του αρχικού.

5.: Να βρεθεί ο αντίστροφος του

$$A = \begin{bmatrix} 1 & & & & & \\ & 1 & & & & \\ & & \ddots & & & \\ & & & 1 & & \\ & & & a_{i+1,i} & 1 & \\ & & & \vdots & & \ddots \\ & & & a_{ni} & & & 1 \end{bmatrix} \in \mathbb{C}^{n,n}, \ n > 2.$$

6.: Δίνεται ο πίνακας A, όπου

$$A = \begin{bmatrix} 10 & -3 & 2 & 2 \\ 4 & 0 & -1 & 0 \\ 3 & 1 & 5 & 0 \\ 0 & 1 & -1 & 3 \end{bmatrix}.$$

Να δειχτεί ότι ο πίνακας A είναι αντιστρέψιμος, ότι δε χρειάζεται εναλλαγή γραμμών για την εύρεση των παραγόντων L και U της LU παραγοντοποίησης και να βρεθούν οι δυο αυτοί παράγοντες.

7.: Να βρεθούν οι L και U παράγοντες των πινάκων

$$\alpha)\ A_1 = \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix}, \quad \beta)\ A_2 = \begin{bmatrix} 4 & -1 & -1 & 0 \\ -1 & 4 & 0 & -1 \\ -1 & 0 & 4 & -1 \\ 0 & -1 & -1 & 4 \end{bmatrix}, \quad \gamma)\ A_3 = \begin{bmatrix} 1 & 0 & 2 & 1 \\ 0 & 4 & 8 & 10 \\ 2 & 8 & 29 & 22 \\ 1 & 10 & 22 & 42 \end{bmatrix},$$

χρησιμοποιώντας ακριβή αριθμητική και στη συνέχεια να λυθούν τα συστήματα

$$A_1 x = b_1, \quad A_2 x = b_2, \quad A_3 x = b_3,$$

α) Με LU παραγοντοποίηση και β) Με κλασική απαλοιφή Gauss, όπου

$$b_1^T = [2\ 1\ 0], \quad b_2^T = [6\ 1\ 1\ 1], \quad b_3^T = [4\ 22\ 61\ 75].$$

8.: Ενας πίνακας $A \in \mathbb{C}^{n,n}$ καλείται *αυστηρά διαγώνια υπέρτερος κατά γραμμές (κατά στήλες)* ανν

$$|a_{ii}| > \sum_{j=1,\ j\neq i}^{n} |a_{ij}|,\ i = 1(1)n \quad \left(|a_{jj}| > \sum_{i=1,\ i\neq j}^{n} |a_{ij}|,\ j = 1(1)n \right).$$

Να δειχτεί ότι κατά την απαλοιφή Gauss **δεν** απαιτείται εναλλαγή γραμμών και πιο συγκεκριμένα ότι ο εκάστοτε ελάσσονας υποπίνακας της κάτω δεξιά γωνίας είναι αυστηρά διαγώνια υπέρτερος κατά γραμμές (κατά στήλες). (**Προσοχή**: Να αποδειχτεί το ζητούμενο **μόνο** για το πρώτο βήμα απαλοιφής του Gauss.)

9.: Δίνεται ο πίνακας $A = \begin{bmatrix} 0 & 1 & 2 \\ 2 & -2 & 1 \\ 5 & 3 & 1 \end{bmatrix}$.

α) Να χρησιμοποιηθεί μεταθετικός πίνακας $P \in \mathbb{R}^{3,3}$ τέτοιος ώστε να κάνει εφικτή την LU παραγοντοποίηση για τον πίνακα $A' = PA$.
β) Να παραγοντοποιηθεί ο A' σε γινόμενο LU.
γ) Να χρησιμοποιηθεί η παραπάνω παραγοντοποίηση για να βρεθεί ο αντίστροφος του A'. και
δ) Τέλος, να βρεθεί ο A^{-1} από τον αντίστροφο του A'.
(**Περιορισμός**: Ολες οι πράξεις να γίνουν με ακριβή αριθμητική.)

10.: Δίνεται προς λύση το γραμμικό σύστημα $Ax = b$ με $A = \begin{bmatrix} 1 & 1 & 1 & -1 \\ 1 & 1 & 1 & 1 \\ 2 & 1 & 1 & 1 \\ 1 & 1 & 2 & 1 \end{bmatrix}$ και $b = [2\ 4\ 5\ 5]^T$. Να βρεθεί μεταθετικός πίνακας P τ.ω. να πραγματοποιείται η LU παραγοντοποίηση

του πίνακα PA. Στη συνέχεια να λυθεί το σύστημα χρησιμοποιώντας την LU παραγοντοποίηση και τον πίνακα P.

11.: Να υποδειχτεί "οικονομικός" τρόπος υπολογισμού, από την άποψη των πράξεων και της μνήμης του Υπολογιστή που απαιτείται, των διανυσμάτων $A^{-4}b$ και $A^{-1}BA^{-1}b$, όπου $A, B \in \mathbb{R}^{n,n}$ δοσμένοι πίνακες, με A αντιστρέψιμο, και $b \in \mathbb{R}^n$ δοσμένο διάνυσμα.

12.: Να βρεθεί αναλυτικά το είδος και το πλήθος των πράξεων που απαιτούνται για την επίλυση τριδιαγώνιου πραγματικού συστήματος στο οποίο ο πίνακας των συντελεστών των αγνώστων $A \in \mathbb{R}^{n,n}$ ικανοποιεί τις υποθέσεις του Θεωρήματος 2.1.

13.: Δίνεται το σύστημα $Ax = b$, όπου $A \in \mathbb{C}^{n,n}$, $b \in \mathbb{C}^n$. Εστω ότι $A = B + iC$ και $b = c + id$ με $B, C \in \mathbb{R}^{n,n}$ και $c, d \in \mathbb{R}^n$. Να βρεθεί πραγματικό σύστημα, διπλάσιων διαστάσεων, ισοδύναμο του δοθέντος μιγαδικού.

14.: Αφού βρεθεί το είδος και το πλήθος των πράξεων για την επίλυση του συστήματος $Ax = b$, $A \in \mathbb{R}^{n,n}$, $b \in \mathbb{R}^n$, ο πίνακας A του οποίου ικανοποιεί τις υποθέσεις του Θεωρήματος 2.1, με τη μέθοδο απαλοιφής των Gauss-Jordan, να γίνει σύγκριση με το πλήθος των αντίστοιχων πράξεων που απαιτούνται με την κλασική μέθοδο απαλοιφής του Gauss.

15.: Να δειχτεί ότι με κατάλληλη εκμετάλλευση της παρουσίας των μηδενικών στα διανύσματα-στήλες του μοναδιαίου πίνακα I_n είναι δυνατόν ο ασυμπτωτικός συντελεστής $\frac{4}{3}$, που εμφανίζεται στο απαιτούμενο πλήθος πράξεων για την εύρεση του αντίστροφου αντιστρέψιμου πίνακα $A \in \mathbb{R}^{n,n}$, να καταστεί 1. (Σημείωση: Θα υποτεθεί ότι ο πίνακας A ικανοποιεί τις υποθέσεις του Θεωρήματος 2.1.)

16.: Να βρεθεί το είδος και το πλήθος των πράξεων για την επίλυση ενός γραμμικού συστήματος $Ax = b$, του οποίου ο αντιστρέψιμος πίνακας των συντελεστών των αγνώστων $A \in \mathbb{R}^{n,n}$ είναι αναγώγιμος της μορφής (2.14) σα συνάρτηση του r. Για ποια τιμή του r το πλήθος των απαιτούμενων πολλαπλασιασμών γίνεται ελάχιστο;

17.: Δίνεται πίνακας $A \in \mathbb{R}^{n,n}$, που ικανοποιεί τις υποθέσεις της Θεωρήματος 2.1. Να βρεθεί αναλυτικά το είδος και το πλήθος των πράξεων που απαιτούνται για την εύρεση του A^{-1} εκμεταλλευόμενοι τις θέσεις των μηδενικών του μοναδιαίου πίνακα I με την απαλοιφή των Gauss-Jordan.

18.: Είναι γνωστό ότι στη Γραμμική Αλγεβρα για την εύρεση του αντίστροφου αντιστρέψιμου πίνακα $A \in \mathbb{C}^{n,n}$ χρησιμοποιείται ο επαυξημμένος πίνακας $[A|I]$, όπου $I \in \mathbb{C}^{n,n}$ ο μοναδιαίος, και εφαρμόζοντας επανειλημμένα τις πράξεις (1), (2) και (3), που περιγράφονται αμέσως μετά το αρχικό σύστημα (2.1), δηλαδή πολλαπλασιάζοντας από τα αριστερά επί κατάλληλο πίνακα $Q \in \mathbb{C}^{n,n}$, προσπαθούμε να καταλήξουμε στον $Q[A|I] = [QA|Q]$ με $QA = I$, οπότε

$Q = A^{-1}$. (Σημείωση: Ο ανωτέρω τρόπος εύρεσης του A^{-1} καλείται μέθοδος των Gauss-Jordan.) Προφανώς, αντίστοιχες πράξεις θα μπορούσαν να γίνουν επί των στηλών του A χρησιμοποιώντας πίνακα $R \in \mathbb{C}^{n,n}$ έτσι ώστε $[A|I]R = [AR|R]$, με $AR = I$, οπότε $R = A^{-1}$. Αν χρησιμοποιηθούν κατάλληλοι πίνακες Q και R έτσι ώστε $Q[A|I]R = [QAR|QR]$, με $QAR = I$, είναι ο $QR = A^{-1}$; Αν όχι, πότε μπορεί να είναι και γιατί;

19.: Δίνεται το πραγματικό ομαλό γραμμικό σύστημα $Ax = b$ διάστασης $3n$ διαχωρισμένο σε blocks όπως παρακάτω:

$$\left[\begin{array}{ccc} B_1 & 0 & D_1 \\ 0 & B_2 & D_2 \\ C_1 & C_2 & B_3 \end{array}\right] \left[\begin{array}{c} x_1 \\ \hline x_2 \\ \hline x_3 \end{array}\right] = \left[\begin{array}{c} b_1 \\ \hline b_2 \\ \hline b_3 \end{array}\right].$$

Στο παραπάνω σύστημα οι block υποπίνακες είναι διάστασης n, με $\det(B_1 B_2) \neq 0$, και όλα τα υποδιανύσματα έχουν n συνιστώσες το καθένα. Για να λυθεί το εν λόγω σύστημα υποδείχνεται η ακόλουθη block παραγοντοποίηση του A

$$A = \left[\begin{array}{ccc} B_1 & 0 & D_1 \\ 0 & B_2 & D_2 \\ C_1 & C_2 & B_3 \end{array}\right] = \left[\begin{array}{ccc} L_{11} & & \\ L_{21} & L_{22} & \\ L_{31} & L_{32} & L_{33} \end{array}\right] \left[\begin{array}{ccc} I & U_{12} & U_{13} \\ & I & U_{23} \\ & & I \end{array}\right], \tag{2.16}$$

όπου I είναι ο μοναδιαίος πίνακας διάστασης n.
α) Να προσδιοριστούν αναλυτικά όλοι οι υποπίνακες L_{ij} και U_{ij} στους δυο παράγοντες της (2.16).
β) Να χρησιμοποιηθεί η παραπάνω παραγοντοποίηση και να υποδειχτεί ένας αποτελεσματικός τρόπος για την επίλυση του αρχικού συστήματος, όπου **δε** θα υπολογίζονται αντίστροφοι πινάκων.

20.: Δίνεται το πραγματικό ομαλό γραμμικό σύστημα $Ax = b$ διάστασης $2n$ του οποίου ο πίνακας A έχει την παρακάτω μορφή

$$A = \left[\begin{array}{cccccccccc} \times & \times & \times & \times & \cdots & \cdots & \times & \times & \times & \times \\ 0 & \times & \times & \times & \cdots & \cdots & \times & \times & \times & 0 \\ 0 & 0 & \times & \times & \cdots & \cdots & \times & \times & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \vdots & \vdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & 0 & \times & \times & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & \times & \times & 0 & \cdots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \vdots & \vdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \times & \times & \cdots & \cdots & \times & \times & 0 & 0 \\ 0 & \times & \times & \times & \cdots & \cdots & \times & \times & \times & 0 \\ \times & \times & \times & \times & \cdots & \cdots & \times & \times & \times & \times \end{array}\right],$$

όπου $\times$ συμβολίζει ένα μη μηδενικό (γενικά) στοιχείο.
α) Να προταθεί και να περιγραφτεί ένας αποτελεσματικός αλγόριθμος n βημάτων, για την επίλυση του δοθέντος συστήματος, σε κάθε βήμα του οποίου θα επιλύεται ένα 2×2 γραμμικό σύστημα που θα βρίσκει δύο από τις συνιστώσες του άγνωστου διανύσματος x. (Υπόδειξη: Θα υποτεθεί ότι σε κάθε βήμα μπορεί να εφαρμοστεί χωρίς προβλήματα η μέθοδος απαλοιφής του Gauss.)
β) Να βρεθεί και να δοθεί αναλυτικά (και χωριστά) το πλήθος των απαιτούμενων πράξεων, δηλαδή προσθαφαιρέσεων (Π_ρ), πολλαπλασιασμών ($\Pi_{o\lambda}$) και διαιρέσεων (Δ), για την πλήρη επίλυση του παραπάνω συστήματος.

21.: Δίνεται πίνακας ζώνης $A \in I\!R^{n,n}$ με ημιεύρος ζώνης n_a $(1 \leq n_a < \frac{n}{2} - 1)$. Με την προϋπόθεση ότι κατά την LU παραγοντοποίηση **δεν** απαιτούνται εναλλαγές γραμμών, να βρεθεί το **ακριβές** πλήθος των πολλαπλασιασμών ($\Pi_{o\lambda}$), προσθαφαιρέσεων (Π_ρ) και διαιρέσεων (Δ), που απαιτούνται για την εύρεση των παραγόντων L και U, εκμεταλλευόμενοι πλήρως την παρουσία των εκτός της ζώνης μηδενικών.

2.3 Στρατηγικές Οδήγησης και Κατάσταση Συστήματος

Η εφαρμογή ενός αλγόριθμου για την εύρεση αριθμητικών αποτελεσμάτων από αριθμητικά δεδομένα θα πρέπει να τυχαίνει κάποιας διερεύνησης προτού ο αλγόριθμος εφαρμοστεί. Γενικά, ένας αλγόριθμος λέγεται ευσταθής αν "μικρά" σφάλματα στα δεδομένα του προβλήματος ή/και στους υπολογισμούς επιφέρουν "μικρά" σφάλματα στα αποτελέσματα. Αλλιώς λέγεται ασταθής. Σε περιπτώσεις αστάθειας είναι δυνατόν το όλο πρόβλημα να ξεπερνιέται με αντικατάσταση του αλγόριθμου με κάποιον άλλον ευσταθή. Τότε λέμε ότι το αρχικό πρόβλημα που δόθηκε είναι καλής κατάστασης. Υπάρχουν όμως παθολογικές περιπτώσεις αστάθειας όπου η οποιαδήποτε αντικατάσταση του αλγόριθμου με άλλον δε βελτιώνει την κατάσταση. Στην περίπτωση αυτή μιλάμε για κακή κατάσταση του προβλήματος. Είναι φυσικό, λοιπόν, η ευστάθεια ή όχι του αλγόριθμου ή/και η κατάσταση του προβλήματος να μας απασχολεί και στη λύση των γραμμικών συστημάτων για τους αλγόριθμους που αναπτύχτηκαν μέχρι τώρα. Καταρχάς η θεωρία που αναπτύχτηκε σ΄ ό,τι αφορά την επίλυση ενός γραμμικού συστήματος έγινε κάτω από τη (θεωρητική) παραδοχή της απουσίας σφαλμάτων οποιασδήποτε φύσης. Στην πράξη όμως κι εφόσον είμαστε αναγκασμένοι να εργαζόμαστε με πεπερασμένο πλήθος σημαντικών ψηφίων έχουμε την αναπόφευκτη παρουσία των σφαλμάτων στρογγύλευσης. Ετσι και κατά την επίλυση ενός γραμμικού συστήματος με τις μέχρι τώρα περιγραφείσες μεθόδους δημιουργούνται προβλήματα όχι μόνο από την παρουσία μηδενικών οδηγών, που κατά κάποιον τρόπο μπορούν να ξεπεραστούν με εναλλαγές γραμμών, αλλά και όταν ακόμη οι οδηγοί είναι σχετικά "μικροί". Στη συνέχεια δίνουμε ένα από τα κλασικά παραδείγματα των συνεπειών της εργασίας μας με πεπερασμένο πλήθος σημαντικών ψηφίων. Για το σκοπό αυτό ας υποθέσουμε ότι έχουμε προς επίλυση το παρακάτω σύστημα τα δεδομένα του οποίου δίνονται με

τρία σημαντικά ψηφία

$$\begin{bmatrix} 0.000100 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}.$$

Η ακριβής λύση του δοθέντος συστήματος είναι η $[x_1\ x_2] = [1.00010001\cdots\ 0.99989998\cdots]$, η οποία με προσέγγιση τριών σημαντικών ψηφίων είναι $[x_1\ x_2] \approx [1.00\ 1.00]$. Αν εφαρμόσουμε τη μέθοδο απαλοιφής του Gauss με οδηγό στοιχείο $a_{11} = 0.000100 \neq 0$ βρίσκουμε ως πολλαπλασιαστή $m_{21} = \frac{a_{21}}{a_{11}} = 10000$ και άρα το ισοδύναμο σύστημα

$$\begin{bmatrix} 0.000100 & 1 \\ & -10000 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ -10000 \end{bmatrix},$$

του οποίου η λύση με τρία σημαντικά ψηφία (ή όχι) είναι $[x_1\ x_2] = [0\ 1]$. Δηλαδή, τελείως ανακριβής ως προς την τιμή του αγνώστου x_1. Αν τώρα κάνουμε μία εναλλαγή των γραμμών, τότε έχουμε για επίλυση το ισοδύναμο σύστημα

$$\begin{bmatrix} 1 & 1 \\ 0.000100 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}.$$

Εργαζόμενοι, όπως προηγουμένως, βρίσκουμε ότι $m_{21} = \frac{a_{21}}{a_{11}} = 0.000100$, στο νέο σύστημα, και άρα

$$\begin{bmatrix} 1 & 1 \\ & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \end{bmatrix},$$

του οποίου η λύση είναι $[x_1\ x_2] = [1\ 1]$, που είναι και η ακριβής στα τρία σημαντικά ψηφία.
Οπως είδαμε, λοιπόν, ο αρχικός αλγόριθμος ήταν ασταθής ενώ αυτός που προέκυψε με μια απλή εναλλαγή γραμμών ευσταθής. Αυτό σημαίνει ότι το αρχικό πρόβλημα που δόθηκε ήταν καλής κατάστασης. Αν όμως το αρχικό πρόβλημα δεν είναι καλής κατάστασης τότε μια απλή εναλλαγή γραμμών, όπως αυτή που υποδείχτηκε, δεν είναι δυνατόν να επιφέρει βελτίωση στη λύση που παίρνουμε με αποτέλεσμα αυτή να μην μπορεί να θεωρηθεί αξιόπιστη. Ενα δεύτερο παράδειγμα δείχνει πολύ περισσότερες καταστάσεις από ό,τι έδειξε το προηγούμενο. Ας θεωρήσουμε το παρακάτω πρόβλημα

$$\begin{bmatrix} 0.780 & 0.563 \\ 0.913 & 0.659 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0.218 \\ 0.253 \end{bmatrix},$$

η ακριβής λύση του οποίου είναι $[x_1\ x_2] = [1223\ -1694]$. Αν εργαστούμε με τρία σημαντικά ψηφία βρίσκουμε $m_{21} = 1.17$, οπότε έχουμε το ισοδύναμο σύστημα

$$\begin{bmatrix} 0.780 & 0.563 \\ & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0.218 \\ -0.002 \end{bmatrix}.$$

Προφανώς το τελευταίο σύστημα **δεν** έχει λύση. Εναλλάσσουμε τις εξισώσεις του συστήματος και εργαζόμαστε πάλι με τρία σημαντικά ψηφία. Τότε θα έχουμε διαδοχικά

$$\begin{bmatrix} 0.913 & 0.659 \\ 0.780 & 0.563 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0.253 \\ 0.218 \end{bmatrix},$$

$m_{21} = 0.854$ και

$$\begin{bmatrix} 0.913 & 0.659 \\ & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0.253 \\ 0.002 \end{bmatrix}$$

δηλαδή καταλήγουμε πάλι σε ένα σύστημα που **δεν** έχει λύση και άρα η εναλλαγή γραμμών δε βελτιώνει την κατάσταση. Υποθέτουμε τώρα ότι εργαζόμαστε από την αρχή με αριθμητική διπλής ακρίβειας. Χρησιμοποιώντας έξι σημαντικά ψηφία στους υπολογισμούς μας έχουμε για το αρχικό σύστημα $m_{21} = 1.17051$ και άρα

$$\begin{bmatrix} 0.780 & 0.563 \\ & 0.000003 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0.218 \\ -0.002171 \end{bmatrix}$$

του οποίου η λύση είναι $[x_1\ x_2] = [522.619\ \ -723.667]$. Παρατηρούμε ότι παρά το γεγονός ότι βρισκόμαστε αρκετά μακρυά από την ακριβή λύση δεν είμαστε πια στην περίπτωση του συστήματος που έχει μή συμβιβαστές εξισώσεις. Τέλος, αν εναλλάξουμε τις εξισώσεις στο τελευταίο σύστημα και εργαστούμε πάλι με έξι σημαντικά ψηφία παίρνουμε διαδοχικά $m_{21} = 0.854326$ και άρα

$$\begin{bmatrix} 0.913 & 0.659 \\ & -0.000001 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0.253 \\ 0.001856 \end{bmatrix}.$$

Η λύση του τελευταίου με έξι σημαντικά ψηφία είναι $[x_1\ x_2] = [1339.92\ \ -1856]$, που απέχει ακόμη αρκετά από την ακριβή αλλά είναι οπωσδήποτε πολύ καλύτερη από αυτήν που βρήκαμε προηγουμένως.

Σημείωση: Γενικότερα η βελτίωση στην αριθμητική λύση ενός προβλήματος, που παρουσιάζεται συνήθως, όταν μεταβαίνουμε από αριθμητική κάποιας ακρίβειας σε αριθμητική μεγαλύτερης ακρίβειας ίσως να είναι αναμενόμενη. Αυτό, γιατί αν τελικά κατορθώσουμε και μεταβούμε σε αριθμητική "άπειρης" ακρίβειας, τότε φυσικά θα βρούμε και την ακριβή λύση του προβλήματος.

Με το παράδειγμα που εξαντλητικά περιγράψαμε είδαμε ότι υπάρχουν "παθολογικές" καταστάσεις όπου η αλλαγή του αλγόριθμου είναι δυνατόν να μη βελτιώνει διόλου την όλη κατάσταση ή απλά να τη βελτιώνει αλλά όχι τόσο σημαντικά. Θα μπορούσε να πεί κανείς πως αν δε βρισκόμαστε σε περιπτώσεις παθολογικών καταστάσεων τότε καλό θα ήταν κατά τη λύση συστημάτων να χρησιμοποιούμε κατά το δυνατόν απόλυτα μικρούς πολλαπλασιαστές ή, με άλλα λόγια, απόλυτα μεγαλύτερους οδηγούς. Ετσι για την αποφυγή σχετικά "μικρών" οδηγών δυνάμενων να οδηγήσουν σε μή αξιόπιστα αποτελέσματα ακολουθούμε μία στρατηγική, όπως είναι και αυτή που περιγράφτηκε παραπάνω. Υπάρχουν, βέβαια, αρκετές στρατηγικές που αποτελούν παραλλαγές της περιγραφείσας. Δημοφιλέστερες είναι οι παρακάτω δύο.

Μερική Οδήγηση: Στην αρχή του k-οστού βήματος της απαλοιφής θεωρούμε ως οδηγό $a_{pk}^{(k)}$, (και άρα αντίστοιχη οδηγό γραμμή και εξίσωση) το στοιχείο που αντιστοιχεί στο δείκτη p (συνήθως το μικρότερο) και είναι τ.ω. $|a_{pk}^{(k)}| = \max_{j=k(1)n} |a_{jk}^{(k)}|$.

Ολική Οδήγηση: Στην αρχή του k-οστού βήματος της απαλοιφής θεωρούμε ως οδηγό στοιχείο $a_{pq}^{(k)}$, αυτό που αντιστοιχεί στους δείκτες $p,\ q$ και που είναι τ.ω. $|a_{pq}^{(k)}| = \max_{j,l=k(1)n} |a_{jl}^{(k)}|$.

Παρατηρούμε δηλαδή ότι στην ολική οδήγηση δεν αλλάζουμε ουσιαστικά μόνο γραμμές αλλά και στήλες!

Η μερική οδήγηση απαιτεί για την εύρεση του απόλυτα μεγαλύτερου στοιχείου πράξεις της τάξης $\mathcal{O}(n^2)$ ενώ αυτή της ολικής πράξεις της τάξης $\mathcal{O}(n^3)$. Από την άποψη των πράξεων η μερική οδήγηση πλεονεκτεί έναντι της ολικής αφού γίνεται με λιγότερο κόστος. Από την άλλη πλευρά η ολική φαίνεται να είναι ακριβέστερη κι άρα ασφαλέστερη της μερικής. Σε παθολογικές όμως καταστάσεις και οι δυο, όπως θα δούμε στη συνέχεια, δεν είναι αξιόπιστες. Επειδή αυτές οι παθολογικές καταστάσεις δεν είναι συνήθεις, στις περισσότερες εφαρμογές η μερική οδήγηση, ως οικονομικότερη, αποτελεί την πιο δημοφιλή από τις δυο.

Πολλές φορές στην πράξη ακολουθιέται μια παραλλαγή της μερικής οδήγησης, που καλείται *μερική οδήγηση με στάθμιση*. Σ′ αυτήν, και σε ένα αρχικό στάδιο, βρίσκονται οι μεγαλύτερες απόλυτες τιμές των στοιχείων κάθε γραμμής, έστω $s_i = \max_{j=1(1)n} |a_{ij}|,\ i = 1(1)n$, κατά δε το k-οστό στάδιο της απαλοιφής με μερική οδήγηση, αντί να επιλέγεται ως οδηγό στοιχείο το $a_{pk}^{(k)}$ με $|a_{pk}^{(k)}| = \max_{j=k(1)n} |a_{jk}^{(k)}|$, επιλέγεται το $a_{pk}^{(k)}$ τ.ω. $\frac{|a_{pk}^{(k)}|}{s_p} = \max_{j=k(1)n} \left(\frac{|a_{jk}^{(k)}|}{s_j} \right)$. Το κόστος της απαλοιφής με απλή μερική οδήγηση δεν αυξάνεται κατά πολύ αφού οι επιπλέον πράξεις είναι της τάξης του $\mathcal{O}(n^2)$. Κάτω από ορισμένες παραδοχές μπορεί να δικαιολογηθεί γιατί η μερική οδήγηση με στάθμιση δίνει πιο αποδεκτά αριθμητικά αποτελέσματα.

Σημειώσεις: α) Οταν η μερική οδήγηση (χωρίς ή με στάθμιση) εφαρμόζεται στον πίνακα των συντελεστών A ενός συστήματος $Ax = b$, ουσιαστικά παλλαπλασιάζονται τα μέλη του συστήματος $Ax = b$ επί ένα μεταθετικό πίνακα P κι έτσι επιλύεται το ισοδύναμο σύστημα $(PA)x = Pb$, όπως δηλαδή αυτό γίνεται στην περίπτωση απαλοιφής του Gauss όταν κάποιο από τα οδηγά στοιχεία είναι μηδέν. β) Στην ολική οδήγηση όμως, μεταθετικοί πίνακες P και Q πολλαπλασιάζουν τόσο τις γραμμές όσο και τις στήλες του πίνακα, αντίστοιχα, και άρα τελικά επιλύεται ένα ισοδύναμο σύστημα της μορφής $(PAQ)(Q^T x) = Pb$. Συνεπώς το άγνωστο διάνυσμα που θα βρίσκεται θα είναι το $Q^T x$ και επομένως θα πρέπει να χρησιμοποιηθεί ο Q για να ληφτούν οι άγνωστες συνιστώσες του x. Οπως και στην περίπτωση της μερικής οδήγησης εδώ θα πρέπει να χρησιμοποιούνται δύο διανύσματα που θα καταγράφουν αντίστοιχα τις μεταθέσεις των γραμμών (σειρά των εξισώσεων) και τις μεταθέσεις των στηλών (σειρά των αγνώστων).

Στο σημείο αυτό θεωρούμε σκόπιμο να δώσουμε τον αλγόριθμο της Μερικής Οδήγησης μαζί με την αντίστοιχη προς τα πίσω αντικατάσταση για την επίλυση γραμμικού συστήματος με τη μόνη υπόθεση ότι $\det(A) \neq 0$. Θα ακολουθήσει ένα πολύ απλό αριθμητικό παράδειγμα για την καλύτερη κατανόηση της μεθόδου και κυρίως του τρόπου της καταγραφής των όποιων πιθανών εναλλαγών γραμμών.

Αλγόριθμος Απαλοιφής του Gauss **με Μερική Οδήγηση**:

Δεδομένα: Η διάσταση n, ο πίνακας A, $\det(A) \neq 0$, και το διάνυσμα b.
Για $k = 1(1)n$
$i_k = k$
Τέλος 'Για'
Για $k = 1(1)n-1$ (βήματα απαλοιφής)
$max = |a_{i_k k}|; \quad l = k$
Για $j = k+1(1)n$
Αν $|a_{i_j k}| > max$
$max = |a_{i_j k}|; \quad l = j$ (σημείωση της θέσης του απόλυτα μεγαλύτερου στοιχείου)
Τέλος 'Αν'
Τέλος 'Για'
$c = i_l; \quad i_l = i_k; \quad i_k = c$ (εναλλαγή των δεικτών i_l και i_k)
Για $j = k+1(1)n$
$m_{i_j k} = a_{i_j k}/a_{i_k k}$
Για $l = k+1(1)n$
$a_{i_j l} = a_{i_j l} - m_{i_j k} a_{i_k l}$
Τέλος 'Για'
$b_{i_j} = b_{i_j} - m_{i_j k} b_{i_k}$
Τέλος 'Για'
Τέλος 'Για'
Για $k = n(-1)1$ (Προς τα πίσω αντικατάσταση)
$s_k = b_{i_k}$
Για $j = k+1(1)n$
$s_k = s_k - a_{i_k j} x_j$
Τέλος 'Για'
$x_k = s_k/a_{i_k k}$
Τέλος 'Για'
Αποτέλεσμα: Η λύση του συστήματος είναι το διάνυσμα x.

Παρατηρήσεις: α) Ο αλγόριθμος Απαλοιφής του Gauss με μερική οδήγηση αποτελεί μια απλή σκιαγράφηση της θεωρίας που αναπτύχτηκε και ακολουθεί βήμα προς βήμα τους προηγούμενους δύο αλγόριθμους της Απαλοιφής του Gauss και της Προς τα Πίσω Αντικατάστασης του Θεωρήματος 2.1. Οι μόνες διαφορές είναι ότι στην αρχή του πρώτου αλγόριθμου εισάγεται το διάνυσμα i με $i_k = k$, $k = 1(1)n$, που θα καταγράφει τη σειρά των εξισώσεων του συστήματος μετά από κάθε πιθανή εναλλαγή γραμμών. Ακόμη, τόσο στα βήματα της απαλοιφής όσο και στα βήματα της προς τα πίσω αντικατάστασης ο (πρώτος) δείκτης της εκάστοτε οδηγού γραμμής j των αλγόριθμων του Θεωρήματος 2.1 αντικαθίσταται από το δείκτη i_j στον παραπάνω αλγόριθμο. β) Μερικές από τις παρατηρήσεις που έγιναν αμέσως μετά τους δύο αλγόριθμους του Θεωρήματος 2.1 ισχύουν και στην προκείμενη περίπτωση.

ΠΑΡΑΔΕΙΓΜΑ: Δίνεται προς λύση το παρακάτω σύστημα

$$Ax \;=\; b$$

και συγκεκριμένα το

$$\begin{bmatrix} 2 & 1 & 3 \\ 4 & 2 & 1 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 13 \\ 11 \\ 6 \end{bmatrix}.$$

Καταρχάς μια παρατήρηση σ΄ ό,τι αφορά τους κύριους υποπίνακες του A. Παρατηρούμε ότι

$$\det(A_{1\times 1}) = \det([2]) = 2 \neq 0,\ \det(A_{2\times 2}) = \det\left(\begin{bmatrix} 2 & 1 \\ 4 & 2 \end{bmatrix}\right) = 0$$

και

$$\det(A_{3\times 3}) = \det\left(\begin{bmatrix} 2 & 1 & 3 \\ 4 & 2 & 1 \\ 1 & 1 & 1 \end{bmatrix}\right) = 5 \neq 0.$$

Επομένως, λόγω της μή αντιστρεψιμότητας του 2×2 κύριου υποπίνακα του A, ο οποίος A έχει αντίστροφο, δεν είναι δυνατόν να έχει εφαρμογή το Θεώρημα 2.1 και άρα δεν μπορεί να εφαρμοστεί ο Αλγόριθμος απαλοιφής του Gauss που βασίζεται σ΄ αυτό. Οπως θα δούμε η εφαρμογή του αλγόριθμου που μόλις παρουσιάστηκε δεν παρουσιάζει κανένα πρόβλημα.
Με την έναρξη της εφαρμογής του αλγόριθμου απαλοιφής του Gauss με μερική οδήγηση δημιουργείται το διάνυσμα i, με $n = 3$ συνιστώσες (θέσεις), όπου $i_1 = 1,\ i_2 = 2$ και $i_3 = 3$, τα περιεχόμενα των θέσεων του οποίου δείχνουν την παρούσα φυσική σειρά των γραμμών του πίνακα A (και των εξισώσεων του δοθέντος συστήματος), όπως φαίνεται παρακάτω.

A			b	i
2	1	3	13	1
4	2	1	11	2
1	1	1	6	3

Πριν από την έναρξη του πρώτου βήματος της απαλοιφής βρίσκεται το απόλυτα μέγιστο στοιχείο της πρώτης στήλης του A που θα χρησιμεύσει ως οδηγός. Αυτό συμβαίνει να είναι το στοιχείο της δεύτερης γραμμής, οπότε $l = 2$. Μετά τη διαπίστωση αυτή ακολουθεί η εναλλαγή των περιεχομένων των θέσεων i_k και i_l, όπου $l = 2$ και $k = 1$. Στη συνέχεια βρίσκονται οι πολλαπλασιαστές που αντιστοιχούν στις γραμμές $i_j,\ j = k+1(1)n$, με $j = 2$ και 3. Αρα θα βρεθούν οι πολλαπλασιαστές που αντιστοιχούν στις γραμμές $i_2 = 1$ και $i_3 = 3$, όπως φαίνεται στην παρακάτω διάταξη.

m_{i_jk}	$A^{(1)}$			$b^{(1)}$	$i^{(1)}$
$\frac{2}{4} = 0.5$	2	1	3	13	2
	$\underline{4}$	2	1	11	1
$\frac{1}{4} = 0.25$	1	1	1	6	3

Η κατάσταση αμέσως μετά το πρώτο βήμα της απαλοιφής δίνεται στη συνέχεια.

	$A^{(2)}$		$b^{(2)}$	$i^{(2)}$
0	0	2.5	7.5	2
$\underline{4}$	2	1	11	1
0	0.5	0.75	3.25	3

Ξεκινώντας τη δεύτερη ανακύκλωση βρίσκεται το απόλυτα μεγαλύτερο στοιχείο που αντιστοιχεί στη δεύτερη στήλη και στις γραμμές $i_k^{(2)} = i_2^{(2)} = 1$ και $i_3^{(2)} = 3$. Διαπιστώνεται ότι το ζητούμενο στοιχείο είναι το $a^{(2)}_{i_3^{(2)}2} = 0.5 \neq 0$, το οποίο και θα χρησιμοποιηθεί ως οδηγός. Ακολουθεί η αντιμετάθεση των περιεχομένων των θέσεων των $i_2^{(2)}$ και $i_3^{(2)}$, οπότε έχουμε $i_2^{(3)} = 3$ και $i_3^{(3)} = 1$. Τέλος βρίσκεται ο πολλαπλασιαστής που αντιστοιχεί στην πρώτη γραμμή, οπότε έχουμε

m_{ijk}	$A^{(2)}$			$b^{(2)}$	$i^{(3)}$
$\frac{0}{0.5} = 0$	0	0	2.5	7.5	2
	$\underline{4}$	2	1	11	3
	0	$\underline{0.5}$	0.75	3.25	1

Προφανώς το δεύτερο και τελευταίο βήμα της απαλοιφής θα αφήσει όλα τα στοιχεία της γραμμής που χαρακτηρίζεται από τον $i_2^{(3)} = 1$ αμετάβλητα, αφού $m_{i^{(3)}_3 2} = 0$, το δε στοιχείο $a^{(3)}_{i_3^{(3)}3} = 2.5$ θα είναι διάφορο από το μηδέν κι αυτό γιατί $\det(A^{(3)}) = \det(A) \neq 0$. Ετσι θα έχουμε τελικά ότι

	$A^{(3)}$		$b^{(3)}$	$i^{(3)}$
0	0	$\underline{2.5}$	7.5	2
$\underline{4}$	2	1	11	3
0	$\underline{0.5}$	0.75	3.25	1

Για την επίλυση του αρχικού συστήματος βρίσκουμε τους αγνώστους στη σειρά x_3, x_2 και x_1 χρησιμοποιώντας το τελευταίο τμήμα του αλγόριθμου, δηλαδή την προς τα πίσω αντικατάσταση. Συγκεκριμένα θα έχουμε:

$$x_3 = b^{(3)}_{i_3^{(3)}} / a^{(3)}_{i_3^{(3)}3} = 7.5/2.5 = 3,$$

$$x_2 = (b^{(3)}_{i_2^{(3)}} - a^{(3)}_{i_2^{(3)}3} x_3)/a^{(3)}_{i_2^{(3)}2} = (3.25 - 0.75 \times 3)/0.5 = 2,$$

$$x_1 = (b^{(3)}_{i_1^{(3)}} - a^{(3)}_{i_1^{(3)}2} x_2 - a^{(3)}_{i_1^{(3)}3} x_3)/a^{(3)}_{i_1^{(3)}1} = (11 - 2 \times 2 - 1 \times 3)/4 = 1.$$

Τέλος δίνουμε την LU παραγοντοποίηση του πίνακα PA του Θεωρήματος 2.2 για το παρόν Παράδειγμα. Για το σκοπό αυτό διαπιστώνουμε από τα περιεχόμενα των θέσεων του διανύσματος $i^{(3)}$ ότι αν στον αρχικό πίνακα A ήταν εναλλαγμένες η πρώτη και η δεύτερη γραμμή τότε θα συνέβαιναν τα εξής: Οι πολλαπλασιαστές στο πρώτο βήμα της απαλοιφής θα ήταν 0.5 για την πρώτη γραμμή, που θα

είχε γίνει δεύτερη, και 0.25 για την τρίτη. Στο δεύτερο βήμα της απαλοιφής **δε** θα είχαμε οδηγό στοιχείο 0 αλλά το (απόλυτα μεγαλύτερο) στοιχείο 0.5, το οποίο στο παράδειγμά μας εμφανίστηκε στην τρίτη γραμμή. Ετσι, το δεύτερο βήμα της απαλοιφής θα προχωρούσε χωρίς ιδιαίτερο πρόσκομμα, ο πολλαπλασιαστής 0 θα αντιστοιχούσε τότε στη νέα δεύτερη γραμμή, που προήρθε από την αρχική πρώτη, και θα έπρεπε να εναλλαχτεί με την τρίτη. Ακόμη, θα πρέπει να εναλλαχτούν οι πολλαπλασιαστές 0.5 και 0.25 της πρώτης στήλης αφού οι γραμμές στις οποίες βρίσκονταν θα έχουν εναλλαχτεί . Ο νέος πίνακας $A^{(3)}$ που θα προέκυπτε τοτε θα ήταν βέβαια άνω τριγωνικός. Οι εναλλαγές, όμως, της πρώτης με τη δεύτερη και της νέας δεύτερης με την τρίτη γραμμή του αρχικού πίνακα A πετυχαίνονται με πολλαπλασιασμό από τα αριστερά επί το μεταθετικό πίνακα $P = \begin{bmatrix} e^{2^T} \\ e^{3^T} \\ e^{1^T} \end{bmatrix}$, όπου $e^i,\ i = 1(1)3$, είναι τα διανύσματα στήλες του μοναδιαίου πίνακα. Αρα στην περίπτωση του συγκεκριμένου Παραδείγματος θα είναι

$$\begin{bmatrix} 1 & & \\ 0.25 & 1 & \\ 0.5 & 0 & 1 \end{bmatrix} \begin{bmatrix} 4 & 2 & 1 \\ & 0.5 & 0.75 \\ & & 2.5 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 2 & 1 & 3 \\ 4 & 2 & 1 \\ 1 & 1 & 1 \end{bmatrix} \left(= \begin{bmatrix} 4 & 2 & 1 \\ 1 & 1 & 1 \\ 2 & 1 & 3 \end{bmatrix} \right).$$

Ιδιαίτερα τονίζεται το γεγονός ότι καθώς ο αλγόριθμος της παρούσας απολοιφής εξελίσσεται μετά από κάθε "υποτιθέμενη" εναλλαγή γραμμών του πίνακα $A^{(k)}$ θα πρέπει να εναλλάσσονται, αντίστοιχα, και ΟΛΟΙ οι ευρισκόμενοι πολλαπλασιαστές των γραμμών i_k και i_l, που αντιστοιχούν στη στήλη k, ο δε τελικός πίνακας P θα είναι στη γενική περίπτωση ο

$$P = \begin{bmatrix} e^{i_1^{(n)}{}^T} \\ e^{i_2^{(n)}{}^T} \\ e^{i_3^{(n)}{}^T} \\ \vdots \\ e^{i_{n-1}^{(n)}{}^T} \\ e^{i_n^{(n)}{}^T} \end{bmatrix}.$$

Η απαλοιφή και η επίλυση του γραμμικού συστήματος (2.3), όταν $\det(A) = 0$, ακολουθεί σε γενικές γραμμές τη μέθοδο απαλοιφής του Gauss, όπως αυτή περιγράφτηκε στην περίπτωση του Θεωρήματος 2.2 και στον Αλγόριθμο της Μερικής Οδήγησης που δόθηκε προηγουμένως. Αν $a_{i_j k} = 0$ για όλα τα $j \in \{k, \cdots, n\}$ τότε, αν $k < n$ δεχόμαστε ως οδηγό τον $a_{i_k k}^{(k)} = 0$, και προχωράμε στο επόμενο βήμα της απαλοιφής. Για την προς τα πίσω αντικατάσταση προχωράμε όπως υποδείχτηκε στην περίπτωση της παρατήρησης που δόθηκε αμέσως μετά τις εκφράσεις (2.6).

Θα πρέπει να τονιστεί ότι σε ακραίες περιπτώσεις είναι δυνατόν στην απαλοιφή του Gauss με οποιαδήποτε από τις δυο προαναφερθείσες οδηγήσεις να καταλήγουμε σε τιμή του $a_{nn}^{(n)}$, που να

περιέχει απόλυτο σφάλμα μέχρι και 2^{n-1} φορές το απόλυτο σφάλμα που ίσως υπάρχει στα στοιχεία του αρχικού πίνακα $A^{(1)}$ που αποθηκεύεται. Μία απλουστευμένη ανάλυση που ακολουθεί μπορεί να καταδείξει του λόγου το αληθές.

Ας υποθέσουμε ότι τα στοιχεία $a_{jl},\ j,l = 1(1)n$, του πίνακα A, αποθηκεύονται ως $\overline{a}_{jl}$, στοιχεία του πίνακα $A^{(1)}$ και έχουν σφάλματα $e_{jl}^{(1)}$, τα οποία μπορεί να οφείλονται αποκλειστικά και μόνο στα σφάλματα στρογγύλευσης. Εστω ότι τα σφάλματα αυτά έχουν απόλυτο άνω φράγμα e, δηλαδή $|e_{jl}^{(1)}| \leq e$. Ας υποθέσουμε ακόμη ότι κατά τη διάρκεια των υπολογισμών δεν εισχωρούν νέα σφάλματα και ότι οι πολλαπλασιαστές βρίσκονται ακριβώς σα να μην υπήρχαν διόλου σφάλματα ούτε καν τα αρχικά. Τότε, μετά το πρώτο βήμα της απαλοιφής τα νέα στοιχεία του πίνακα $A^{(2)}$, που παίρνουμε, θα δίνονται από τις σχέσεις

$$\overline{a}_{jl}^{(2)} = \overline{a}_{jl}^{(1)} - m_{j1}\overline{a}_{1l}^{(1)},\ j,l = 2(1)n.$$

Εχοντας υπόψη ότι οι αντίστοιχες ακριβείς σχέσεις της απαλοιφής είναι οι

$$a_{jl}^{(2)} = a_{jl}^{(1)} - m_{j1}a_{1l}^{(1)},\ j,l = 2(1)n,$$

έχουμε αμέσως με αφαίρεση κατά μέλη των δύο παραπάνω σχέσεων ότι τα σφάλματα στα νέα υπολογιζόμενα στοιχεία του πίνακα $A^{(2)}$ θα δίνονται από τις

$$e_{jl}^{(2)} = \overline{a}_{jl}^{(2)} - a_{jl}^{(2)} = e_{j1}^{(1)} - m_{j1}e_{1l}^{(1)},\ j,l = 2(1)n.$$

Παίρνοντας απόλυτες τιμές στις παραπάνω ισότητες και λαβαίνοντας υπόψη ότι ακολουθούμε μια στρατηγική οδήγησης, και επομένως $|m_{j1}| \leq 1,\ j = 2(1)n$, βρίσκουμε ότι

$$|e_{jl}^{(2)}| \leq 2|e_{jl}^{(1)}| \leq 2e,\ j = 2(1)n.$$

Μια απλή επαγωγή αποδείχνει ότι μετά από $n-1$ βήματα απαλοιφής θα έχουμε ότι $|e_{nn}^{(n)}| \leq 2^{n-1}e$. Π.χ., ένα απλό αριθμητικό παράδειγμα στο τελευταίο δείχνει τι μπορεί να συμβεί σε ακραίες περιπτώσεις. Αν είναι $n = 100$, $|a_{lj}| \in (0.1, 1),\ l,j = 1(1)100$, και εργαζόμαστε σε Υπολογιστή με αριθμητική κινητής υποδιαστολής με δέκα σημαντικά ψηφία, που σημαίνει ότι $|e_{jl}^{(1)}| \leq 0.5 \times 10^{-10}$, το μέγιστο απόλυτο σφάλμα που μπορεί να υπάρξει στο στοιχείο $a_{nn}^{(n)}$ είναι ίσο με $2^{99} \times 0.5 \times 10^{-10} \approx 3 \times 10^{19}$!

Οπως είδαμε στα δυο αριθμητικά παραδείγματα των γραμμικών συστημάτων, μερικές φορές η απλή εναλλαγή γραμμών, όπως στο πρώτο παράδειγμα, που ουσιαστικά ήταν η εφαρμογή της μερικής (και συγχρόνως ολικής) οδήγησης (χωρίς ή/και με στάθμιση), έδωσε αποτελέσματα αποδεκτά. Αυτό σημαίνει καταρχάς ότι το πρόβλημα δεν ήταν κακής κατάστασης και το μή ακριβές αποτέλεσμα οφειλόταν στην αστάθεια του αλγόριθμου απαλοιφής χωρίς οδήγηση. Στο δεύτερο παράδειγμα

είδαμε ότι παρά το γεγονός της χρησιμοποίησης της μερικής (και συγχρόνως ολικής) οδήγησης δεν έδωσε αποδεκτά αποτελέσματα. Η κατάσταση άρχισε να βελτιώνεται όταν χρησιμοποιήσαμε αριθμητική διπλής ακρίβειας χωρίς οδήγηση καταρχάς και στη συνέχεια να βελτιώνεται ακόμη περισσότερο με τη χρησιμοποίηση της μερικής (και συγχρόνως ολικής) οδήγησης. Το δεύτερο παράδειγμα είναι χαρακτηριστικό μιας περίπτωσης η οποία οφείλεται στην κακή κατάσταση του προβλήματος. Στη συγκεκριμένη περίπτωση η κακή κατάσταση οφείλεται στο γεγονός ότι ο πίνακας των συντελεστών των αγνώστων βρίσκεται πολύ "κοντά" σε ένα μή αντιστρέψιμο πίνακα ή, όπως θα δούμε, στο γεγονός ότι ο δείκτης κατάστασης του πίνακα είναι μεγάλος. Η ανάλυση που αφορά στο μέγεθος του δείκτη κατάστασης και στις συνέπειες που αυτό μπορεί να έχει στη λύση των γραμμικών συστημάτων θα δοθεί στη συνέχεια.

Υποθέτουμε πως έχουμε για επίλυση το πραγματικό ομαλό γραμμικό σύστημα

$$Ax = b, \tag{2.17}$$

ότι ο πίνακας A αποθηκεύεται ακριβώς ενώ οι συνιστώσες του διανύσματος b αποθηκεύονται με σφάλματα που οφείλονται αποκλειστικά και μόνο στα σφάλματα στρογγύλευσης. Ετσι το διάνυσμα b αποθηκεύεται ως $b^* = b + \delta b$, όπου δb είναι το διάνυσμα-σφάλμα. Εστω ακόμη ότι όλοι οι υπολογισμοί εκτελούνται ακριβώς και επομένως η λύση x^*, που βρίσκεται, περιέχει σφάλματα που οφείλονται αποκλειστικά και μόνο στη μετάδοση των αρχικών σφαλμάτων. Εστω ότι $x^* = x + \delta x$, με δx το διάνυσμα-σφάλμα της λύσης. Θα προσπαθήσουμε να βρούμε ένα άνω φράγμα για το σχετικό απόλυτο σφάλμα της λύσης. Είναι φανερό ότι το σύστημα που επιλύεται αντί του αρχικού είναι το

$$Ax^* = b^*, \tag{2.18}$$

οπότε με αφαίρεση κατά μέλη της (2.17) από τη (2.18) παίρνουμε

$$A\delta x = \delta b.$$

Λύνοντας ως προς δx και εφαρμόζοντας απλές ιδιότητες των norms έχουμε

$$||\delta x|| \ \leq \ ||A^{-1}||\ ||\delta b||. \tag{2.19}$$

Εξάλλου από τη (2.17) παίρνοντας πρώτα norms και ύστερα λύνοντας ως προς $||x||$ βρίσκουμε

$$||x|| \ \geq \ \frac{||b||}{||A||}, \tag{2.20}$$

και τέλος σχηματίζοντας το λόγο $\frac{||\delta x||}{||x||}$ από τις (2.19) και (2.20) έχουμε τελικά ότι

$$\frac{||\delta x||}{||x||} \ \leq \ ||A||\ ||A^{-1}||\frac{||\delta b||}{||b||}$$

ή, ισοδύναμα,

$$\frac{||\delta x||}{||x||} \leq \kappa(A)\frac{||\delta b||}{||b||}. \tag{2.21}$$

Με άλλα λόγια, κάτω από τις προϋποθέσεις της ανάλυσης που έγινε, το μέγιστο σχετικό απόλυτο σφάλμα στη λύση του συστήματος είναι ίσο με το δείκτη κατάστασης του πίνακα A επί το σχετικό απόλυτο σφάλμα του διανύσματος του δεύτερου μέλους. Συνεπώς, πίνακες με μεγάλο δείκτη κατάστασης είναι δυνατόν να οδηγούν σε πρόβλημα κακής κατάστασης και επομένως σε αποτελέσματα τα οποία να μην είναι αξιόπιστα.

Γενικεύοντας τώρα την προηγούμενη ανάλυση υποθέτουμε ότι έχουμε εκτός από τα σφάλματα στρογγύλευσης κατά την αποθήκευση του δεύτερου μέλους και σφάλματα στρογγύλευσης κατά την αποθήκευση των συντελεστών του πίνακα A έτσι ώστε αυτός να αποθηκεύεται ως $A^* = A + \delta A$, όπου δA είναι ο πίνακας των σφαλμάτων στρογγύλευσης του A. Για την απλούστευση της ανάλυσης υποθέτουμε ακόμη ότι ισχύει $||A^{-1}||\ ||\delta A|| < 1$ για κάποια φυσική norm. Η αντίστοιχη εξίσωση της (2.18) είναι τώρα η

$$A^*x^* = b^*. \tag{2.22}$$

Αφαιρώντας τη (2.17) από τη (2.22) και αναδιατάσσοντας παίρνουμε

$$(A + \delta A)\delta x = -\delta A x + \delta b,$$

ή πολλαπλασιάζοντας απο τα αριστερά επί A^{-1} έχουμε

$$(I + A^{-1}\delta A)\delta x = A^{-1}(-\delta A x + \delta b).$$

Παρατηρούμε ότι $||A^{-1}\delta A|| \leq ||A^{-1}||\ ||\delta A|| < 1$, οπότε από το θεώρημα του Neumann έχουμε ότι ο $(I + A^{-1}\delta A)$ είναι αντιστρέψιμος κι άρα από την τελευταία ισότητα παίρνουμε

$$\delta x = (I + A^{-1}\delta A)^{-1}A^{-1}(-\delta A x + \delta b). \tag{2.23}$$

Παίρνοντας norms των δύο μελών της (2.23), εφαρμόζοντας απλές ιδιότητές τους και χρησιμοποιώντας το δεύτερο σκέλος του θεωρήματος του Neumann ($||(I + A^{-1}\delta A)^{-1}|| \leq \frac{1}{1-||A^{-1}\delta A||}$) βρίσκουμε ότι

$$||\delta x|| \leq \frac{1}{1 - ||A^{-1}\delta A||}||A^{-1}||(||\delta A||\ ||x|| + ||\delta b||). \tag{2.24}$$

Τέλος, διαιρώντας και τα δύο μέλη της (2.24) δια $||x||$, λαβαίνοντας υπόψη ότι $||x|| \geq \frac{||b||}{||A||}$ και $||A^{-1}\delta A|| \leq ||A^{-1}||\ ||\delta A|| < 1$, χρησιμοποιώντας αυτά στο δεύτερο μέλος και θέτοντας $||A||\ ||A^{-1}|| = \kappa(A)$, βρίσκουμε ότι

$$\frac{||\delta x||}{||x||} \leq \frac{\kappa(A)}{1 - \kappa(A)\frac{||\delta A||}{||A||}}\left(\frac{||\delta A||}{||A||} + \frac{||\delta b||}{||b||}\right). \tag{2.25}$$

Από τη (2.25) είναι απλό να διαπιστωθεί ότι για $\frac{||\delta A||}{||A||}$ και $\frac{||\delta b||}{||b||}$ σταθερά και υπό την προϋπόθεση ότι $||A^{-1}||\ ||\delta A|| < 1$ το μέγιστο σχετικό απόλυτο σφάλμα στη λύση του συστήματος είναι αύξουσα συνάρτηση του δείκτη κατάστασης $\kappa(A)$. Καταλήγουμε έτσι στο συμπέρασμα ότι για μεγάλο δείκτη κατάστασης η λύση που παίρνουμε από τον Υπολογιστή μπορεί να μην είναι αξιόπιστη.
Σημειώσεις: α) Για $\delta A = 0$, όπως υποθέσαμε στην περίπτωση της ανάλυσης που προηγήθηκε της τελευταίας, βρίσκεται η σχέση (2.21). β) Αν υποθέσουμε ότι ϵ είναι η μονάδα στρογγύλευσης του Υπολογιστή, τότε έχουμε αφενός $||\delta A||_\infty \leq \epsilon||A||_\infty$ και αφετέρου $||\delta b||_\infty \leq \epsilon||b||_\infty$. Κάτω δε από την προϋπόθεση ότι ισχύει $r := \epsilon\kappa_\infty(A) < 1$ προκύπτει αμέσως από τη (2.25) ότι

$$\frac{||\delta x||_\infty}{||x||_\infty} \ \leq \ \frac{2r}{1-r}.$$

Από τη μέχρι τώρα ανάλυση στο παρόν κεφάλαιο έχει διαπιστωθεί πως είναι σχεδόν αδύνατον να βρεθεί η ακριβής λύση ενός γραμμικού συστήματος με τη χρήση Υπολογιστή. Υπάρχουν περιπτώσεις, όμως, όπου η όποια λύση παίρνεται από τον Υπολογιστή μπορεί να τύχει παραπέρα βελτίωσης. Στη συνέχεια δίνουμε έναν αλγόριθμο, που είναι γνωστός ως "Αλγόριθμος Επαναληπτικής Βελτίωσης". Διάφορες επεξηγήσεις αναφορικά μ΄ αυτόν θα δοθούν είτε μέσα στον αλγόριθμο, με μορφή σχολίων, είτε αμέσως μετά την παρουσίασή του.

Αλγόριθμος Επαναληπτικής Βελτίωσης Λύσης Γραμμικού Συστήματος:

Δεδομένα: $A \in I\!R^{n,n},\ \det(A) \neq 0,\ b \in I\!R^n$, t το πλήθος σημαντικών ψηφίων.
LU παραγοντοποίηση του A. (Εστω $LU \approx A$)
    (Σχόλιο: Οι L και U δεν είναι οι ακριβείς παράγοντες του A λόγω χρησιμοποίησης Υπολογιστή.)
Επίλυση των $Lu = b$ και $Ux = u$
$y = x$; $k = 0$
Εφόσον $\frac{||y||_\infty}{||x||_\infty} > 0.5 \times 10^{-t}$ και $k < k_{\max}$ ($k_{\max} << n$)
    $k = k + 1$
    Υπολογισμός $r = b - Ax$ (με διπλή ακρίβεια)
        (Σχόλιο: Στοιχεία των r, b, A, x με απλή ακρίβεια και υπολογισμός στοιχείων του $b - Ax$ με διπλή.)
    Επίλυση των $Lu = r$ και $Uy = u$
    $x = x + y$
Τέλος 'Εφόσον'
Αν $k \leq k_{\max}$ τότε
    Αποτέλεσμα: x η προσέγγιση της λύσης με t σημαντικά ψηφία.
    αλλιώς Αποτέλεσμα: x η προσέγγιση της λύσης, χωρίς να επιτευχθεί η επιζητούμενη ακρίβεια.
Τέλος 'Αν'

Σημειώσεις: α) Ο αλγόριθμος που παρουσιάστηκε δίνει ικανοποιητικά αποτελέσματα σε αρκετές περιπτώσεις όταν οι παράγοντες L και U του A είναι σχετικά ακριβείς. β) Το $k_{\max}$ επιλέγεται από το χρήστη και είναι συνήθως αρκετά μικρότερο από τη διάσταση n του A. γ) Η διαφορά $r = b - Ax$ δεν αναμένεται να είναι το μηδενικό διάνυσμα. Αυτό γιατί το x που βρέθηκε από τον Υπολογιστή δεν είναι συνήθως η ακριβής λύση του συστήματος. Η έννοια, λοιπόν, της διπλής ακρίβειας είναι η ακόλουθη. Τα στοιχεία των A και x δίνονται με απλή ακρίβεια αλλά το γινόμενο Ax, τα στοιχεια του οποίου είναι αθροίσματα γινομένων στοιχείων από τον A και το x, υπολογίζεται με διπλή ακρίβεια. Με διπλή ακρίβεια υπολογίζεται και η διαφορά $b - Ax$, η οποία στη συνέχεια αποθηκεύεται στο r με απλή ακρίβεια. Στη λεπτομέρεια αυτή βασίζεται και η ακόλουθη πρακτική εξήγηση σ΄ ό,τι αφορά την αποτελεσματικότητα της μεθόδου. (Μια θεωρητική απόδειξη, κάτω από ορισμένες προϋποθέσεις, θα δοθεί στο κεφάλαιο των Επαναληπτικών Μεθόδων, όπου είναι δυνατόν να αποδειχτεί ότι η μέθοδος μπορεί να είναι αποτελεσματική ακόμη κι αν δε χρησιμοποιείται διπλή ακρίβεια.) Αν υποτεθεί ότι ένα στοιχείο του b έχει τη μορφή $0.x_1x_2x_3x_4$, με απλή ακρίβεια τεσσάρων σημαντικών ψηφίων, τότε το αντίστοιχο στοιχείο του Ax θα έχει τη μορφή $0.y_1y_2y_3y_4y_5y_6y_7y_8$, με διπλή ακρίβεια οκτώ σημαντικών ψηφίων. Αν οι παράγοντες L και U είναι σχετικά ακριβείς όπως και η λύση x, τότε τα αντίστοιχα στοιχεία των b και Ax θα έχουν μερικά από τα πρώτα σημαντικά ψηφία τους τα ίδια, έστω τα δύο πρώτα $x_1 = y_1$ και $x_2 = y_2$. Συνεπώς, η διαφορά των αντίστοιχων στοιχείων θα είναι της μορφής $0.00z_1z_2z_3z_4z_5z_6$, και αποθηκευόμενη με απλή ακρίβεια στο αντίστοιχο στοιχείο του r, θα είναι έστω $0.00z_1z_2z_3z_4$. Αν τώρα δεν είχε χρησιμοποιηθεί διπλή ακρίβεια, τότε η τελευταία διαφορά θα αποθηκευόταν ως $0.00z_1z_2$. Θα υπήρχε δηλαδή απώλεια δύο σημαντικών ψηφίων πράγμα που, όπως γνωρίζουμε από την Αριθμητική Ανάλυση, μπορεί να αποβεί καταστροφικό σ΄ ό,τι αφορά την ακρίβεια των αποτελεσμάτων. δ) Το κριτήριο σταματήματος των επαναλήψεων στη γενική μορφή είναι $\frac{||y||}{||x||} \leq \epsilon$, όπου $||\cdot||$ είναι μια οποιαδήποτε διανυσματική norm και ϵ ένας "μικρός" θετικός αριθμός. Και τα δύο επιλέγονται από το χρήστη. Η επιλογή της $\ell_\infty-$norm και αυτή του $\epsilon = 0.5 \times 10^{-t}$ εξασφαλίζει, στην περίπτωση σύγκλισης, την ακρίβεια της απόλυτα μεγαλύτερης συνιστώσας της λύσης, που μπορεί να βρεθεί από τον Υπολογιστή, σε t σημαντικά ψηφία. Για την εξασφάλιση ακρίβειας από τον Υπολογιστή t σημαντικών σε **όλες** τις συνιστώσες της λύσης θα πρέπει προφανώς να χρησιμοποιείται σαν κριτήριο σταματήματος των επαναλήψεων το $\max_{i=1(1)n} \left|\frac{y_i}{x_i}\right| \leq 0.5 \times 10^{-t}$.

ΑΣΚΗΣΕΙΣ

1.: Αντί του γραμμικού συστήματος

$$Ax = b, \quad A = \begin{bmatrix} 2 & -1 & 0 \\ -1 & 3 & -1 \\ 0 & -1 & 2 \end{bmatrix}, \quad b = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix},$$

επιλύεται το διαταραγμένο γραμμικό σύστημα $(A+\delta A)(x+\delta x)=b+\delta b$, με

$$A+\delta A \;=\; \begin{bmatrix} 2 & -1 & 0 \\ -1 & 3 & -1.1 \\ 0 & -1.1 & 2 \end{bmatrix} \quad \kappa\alpha\iota \quad b+\delta b \;=\; \begin{bmatrix} 1.1 \\ 1.1 \\ 0.79 \end{bmatrix}.$$

Να βρεθεί ένα άνω φράγμα για το σχετικό απόλυτο σφάλμα και να γίνει η επαλήθευση λύνοντας το διαταραγμένο σύστημα. Δίνεται ότι η λύση του αρχικού είναι $x \;=\; [1\ 1\ 1]^T$ και ότι η απόλυτα μεγαλύτερη και η απόλυτα μικρότερη ιδιοτιμή του A είναι 4 και 1, αντίστοιχα.

2.: Αντί του γραμμικού συστήματος $Ax=b$, επιλύεται το γραμμικό σύστημα $A'x=b$ με

$$A \;=\; \begin{bmatrix} 2 & 0.5 & 0.5 \\ 0.5 & 2 & 0.5 \\ 0.5 & 0.5 & 2 \end{bmatrix}, \quad b \;=\; \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \quad \kappa\alpha\iota \quad A' \;=\; \begin{bmatrix} 2 & 0.5 & 0 \\ 0.5 & 2 & 0.5 \\ 0 & 0.5 & 2 \end{bmatrix}.$$

Να βρεθεί ένα άνω φράγμα για το σχετικό απόλυτο σφάλμα $\frac{||\delta x||_2}{||x||_2}$ χωρίς να λυθεί κανένα από τα δύο συστήματα. (Υπόδειξη: Αν χρειαστεί, να βρεθούν οι ιδιοτιμές του πίνακα A.)

3.: Δίνεται το γραμμικό σύστημα $Ax=b$, όπου $A\in I\!R^{n,n}$, $b\in I\!R^n$ και $A=A^T$ με ιδιοτιμές -10, -5, 1 και 2. Να βρεθεί ένα άνω φράγμα για το σχετικό απόλυτο σφάλμα $\frac{||\delta x||_2}{||x||_2}$ της λύσης x σε σχέση με το σχετικό απόλυτο σφάλμα $\frac{||\delta b||_2}{||b||_2}$ του διανύσματος b, κατά τη λύση του διαταραγμένου γραμμικού συστήματος $A(x+\delta x)=b+\delta b$.

4.: Υποτίθεται ότι για την επίλυση του συστήματος $Ax=b$, όπου $A\in I\!R^{n,n}$, $\det(A)\neq 0$, $b\in I\!R^n\backslash\{0\}$, με μια από τις μεθόδους απαλοιφής Gauss (ή LU παραγοντοποίησης) ο πίνακας A αποθηκεύεται ακριβώς, το διάνυσμα b ως $b^*=b+\delta b$, και όλες οι πράξεις για την εύρεση της λύσης $x^*=x+\delta x$ στον Υπολογιστή εκτελούνται ακριβώς. Να δειχτεί ότι

$$\frac{||\delta x||}{||x||} \;\geq\; \frac{1}{\kappa(A)}\frac{||\delta b||}{||b||}.$$

5.: Χωρίς την εφαρμογή του τύπου (2.25), να βρεθεί ένα άνω φράγμα για το σχετικό απόλυτο σφάλμα $\frac{||\delta x||}{||x||}$ της λύσης του συστήματος της προηγούμενης Ασκησης, όταν ο πίνακας A αποθηκεύεται ως $A^*=A+\delta A$, το διάνυσμα b ακριβώς και όλες οι πράξεις για την εύρεση της λύσης $x^*=x+\delta x$ στον Υπολογιστή εκτελούνται ακριβώς. (Σημείωση: Υποτίθεται ότι $||A^{-1}||||\delta A||<1$.)

6.: Κατά τη λύση του γραμμικού συστήματος $Ax=b$ υπεισέρχονται σφάλματα και τελικά επιλύεται το διαταραγμένο γραμμικό σύστημα $(A+\delta A)(x+\delta x)=b$. Αν για κάποια φυσική norm είναι γνωστό ότι $||A^{-1}\delta A||\leq 0.2$, να αποδειχτεί ότι $||\delta x||\leq 0.25||x||$.

7.: Εστω $a \in I\!R^n \backslash \{0\}$ και $a^* = a + \delta a$ μια προσέγγισή του. Πολλές φορές το πλήθος των "σωστών" σημαντικών ψηφίων του a^* θεωρείται ότι δίνεται από το μεγαλύτερο ακέραιο k που ικανοποιεί τη σχέση

$$\frac{||\delta a||_\infty}{||a||_\infty} \left(\approx \frac{||\delta a||_\infty}{||a^*||_\infty} \right) \leq 0.5 \times 10^{-k}.$$

Για τη λύση του συστήματος $Ax = b$ με $A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$, το διάνυσμα b είναι δυνατόν να δοθεί με ακρίβεια έξι σημαντικών ψηφίων. Πόσα το πολύ σημαντικά ψηφία είναι δυνατόν να αναμένονται στη λύση x του δοθέντος συστήματος αν δεν εισχωρούν νέα σφάλματα στους υπολογισμούς;

2.4 Παραγοντοποίηση Cholesky

Ας υποθέσουμε ότι έχουμε προς επίλυση ένα πραγματικό γραμμικό σύστημα n εξισώσεων με n αγνώστους του οποίου ο πίνακας των συντελεστών των αγνώστων A είναι συμμετρικός, δηλαδή $A^T = A$. Είναι προφανές ότι λόγω της συμμετρίας δε χρειάζεται να αποθηκεύσουμε όλα τα στοιχεία του A. Αρκούν τα στοιχεία του κάτω (ή άνω) τριγωνικού μέρους του, πλήθους $n(n+1)/2$ αντί n^2 που είναι όλα, δηλαδή περίπου τα μισά. Το ερώτημα που γεννιέται τότε είναι μήπως μπορούμε να εκμεταλλευτούμε τη συμμετρία του A και να τον παραγοντοποιήσουμε κατά τέτοιο τρόπο ώστε αντί της συνήθους LU παραγοντοποίησης να έχουμε ένα γινόμενο ενός κάτω κι ενός άνω τριγωνικού πίνακα όπου ο ένας θα αποτελεί τον ανάστροφο του άλλου. Στην περίπτωση αυτή θα αρκεί μόνο η αποθήκευση του ενός από τους δύο παράγοντες. Αυτό φυσικά υπό την προϋπόθεση ότι οι παράγοντες θα έχουν στοιχεία πραγματικά. Η απάντηση είναι καταφατική αν ο A έχει μία επιπλέον ιδιότητα. Αυτή του "θετικά ορισμένου" πίνακα.

Ορισμός 2.3 *Εστω $A \in C\!\!\!\!I^{n,n}$ με $A^H = A$. Ο A είναι θετικά ορισμένος ανν $\forall\ x \in C\!\!\!\!I^n \setminus \{0\}$ ισχύει $(x, Ax)_2\ (= (Ax, x)_2) > 0$.*

Σημείωση: Για $A \in C\!\!\!\!I^{n,n}$, με $A^H = A$, αντίστοιχοι προφανείς ορισμοί χρησιμοποιούνται για να ορίσουν έναν πίνακα ως "μή αρνητικά ορισμένο", "αρνητικά ορισμένο" και "μή θετικά ορισμένο". Αν ο $A \in C\!\!\!\!I^{m,n}$ είναι Ερμιτιανός ($A^H = A$) τότε ισχύουν γι αυτόν μερικές απλές ιδιότητες οι δύο πρώτες των οποίων αποδείχνονται εύκολα: α) Τα διαγώνια στοιχεία του είναι πραγματικά. β) Οι ιδιοτιμές του είναι πραγματικές. και γ) Εχει n γραμμικά ανεξάρτητα ιδιοδιανύσματα x^j τα οποία μπορούν να παρθούν έτσι ώστε να είναι ορθοκανονικά, δηλαδή

$$(x^j, x^k)_2 = \delta_{jk} = \begin{cases} 1, & \alpha\nu\nu\ j = k \\ 0, & \alpha\nu\nu\ j \neq k \end{cases}, \quad j, k = 1(1)n.$$

Το αντίστοιχο του Ερμιτιανού πίνακα στην περίπτωση $A \in I\!R^{n,n}$ με $A^H\ (= A^T) = A$ είναι ο συμμετρικός πίνακας. Γι΄ αυτή την περίπτωση ο ορισμός του θετικά ορισμένου παραμένει ο ίδιος αλλά μπορεί να δειχτεί ότι είναι ισοδύναμος με τον παρακάτω απλούστερο:

Ορισμός 2.4 *Εστω $A \in I\!R^{n,n}$ με $A^T = A$. Ο A είναι θετικά ορισμένος ανν $\forall\ x \in I\!R^n \setminus \{0\}$ ισχύει $(x, Ax)_2\ (= (Ax, x)_2) > 0$.*

Στη συνέχεια αποδείχνουμε μια πρόταση η οποία είναι πολύ χρήσιμη στην παραγοντοποίηση στην οποία θα αναφερθούμε αμέσως μετά.

Θεώρημα 2.3 *Εστω ότι ο $A \in \mathbb{C}^{n,n}$ είναι Ερμιτιανός και θετικά ορισμένος. Τότε αν διαγράψουμε μια οποιαδήποτε γραμμή του και την αντίστοιχη στήλη του ο $(n-1) \times (n-1)$ πίνακας A', που απομένει, είναι επίσης Ερμιτιανός και θετικά ορισμένος.*

Απόδειξη: Εστω ότι διαγράφουμε την i-οστή γραμμή και την αντίστοιχη στήλη του πίνακα A, που δόθηκε και που είχε γραφτεί σε block μορφή ως ακολούθως

$$A \;=\; \left[\begin{array}{c|c|c} A_{11} & A_{12} & A_{13} \\ \hline A_{21} & a_{ii} & A_{23} \\ \hline A_{31} & A_{32} & A_{33} \end{array}\right],$$

όπου οι διάφοροι υποπίνακες είναι τ.ω. $A_{11} \in \mathbb{C}^{i-1,i-1}$, $A_{12} \in \mathbb{C}^{i-1,1}$, $A_{13} \in \mathbb{C}^{i-1,n-i}$, $A_{21} \in \mathbb{C}^{1,i-1}$, $a_{ii} \in \mathbb{C}^{1,1}$, $A_{23} \in \mathbb{C}^{1,n-i}$, $A_{31} \in \mathbb{C}^{n-i,i-1}$, $A_{32} \in \mathbb{C}^{n-i,1}$ και $A_{33} \in \mathbb{C}^{n-i,n-i}$. Ο πίνακας A' που προκυπτει μετά τη διαγραφή της i-οστης γραμμής και στήλης του A θα έχει την block μορφή

$$A' \;=\; \left[\begin{array}{c|c} A'_{11} & A'_{12} \\ \hline A'_{21} & A'_{22} \end{array}\right] \;=\; \left[\begin{array}{c|c} A_{11} & A_{13} \\ \hline A_{31} & A_{33} \end{array}\right],$$

όπου τα αντίστοιχα blocks των δυο πινάκων είναι ταυτοτικά ίσα. Η απόδειξη ότι ο A' είναι επίσης Ερμιτιανός είναι απλή. Αποτελείται από τέσσερα μέρη και στη συνέχεια θα αποδείξουμε μόνο το ένα από αυτά. Εστω $a'_{kl},\ k = 1(1)i-1,\ l = i(1)n-1$, στοιχείο του A'_{12}. Θα έχουμε διαδοχικά $a'_{kl} = a_{k,l+1} = \overline{a}_{l+1,k} = \overline{a}'_{l,k}$. Το τελευταίο στοιχείο ανήκει προφανώς στον υποπίνακα A'_{21}. Κ.λπ. Αρα ο A' είναι Ερμιτιανός. Για να αποδείξουμε ότι είναι και θετικά ορισμένος θεωρούμε το τυχόν διάνυσμα $x \in \mathbb{C}^n \setminus \{0\}$ τ.ω. η i-οστή συνιστώσα του να είναι μηδέν και το διαχωρίζουμε σε blocks σύμφωνα με το διαχωρισμό του A. Δηλαδή, $x = [x_1^T \mid 0 \mid x_3^T]^T$, όπου $x_1 \in \mathbb{C}^{i-1}$ και $x_3 \in \mathbb{C}^{n-i}$. Αφού ο A είναι θετικά ορισμένος $\forall\ x = [x_1^T \mid 0 \mid x_3^T]^T \in \mathbb{C}^n \setminus \{0\}$ θα έχουμε, αν χρησιμοποιήσουμε τις block μορφές των A και x και παραλείψουμε κάποια ενδιάμεσα αποτελέσματα, ότι

$$0 \;<\; (x, Ax)_2 \;=\; x^H Ax \;=\; x_1^H(A_{11}x_1 + A_{13}x_3) + x_3^H(A_{31}x_1 + A_{33}x_3).$$

Το δεξιά μέλος των παραπάνω σχέσεων μπορεί να γραφτεί διαδοχικά ως εξής

$$[x_1^H \mid x_3^H] \left[\begin{array}{c|c} A_{11} & A_{13} \\ \hline A_{31} & A_{33} \end{array}\right] \left[\begin{array}{c} x_1 \\ \hline x_3 \end{array}\right] = x'^H A' x',$$

όπου $x' = [x_1^T \mid x_3^T]^T \in \mathbb{C}^{n-1} \setminus \{0\}$. Με άλλα λόγια ο πίνακας A' είναι θετικά ορισμένος. □

Δύο άμεσα συμπεράσματα του προηγούμενου θεωρήματος είναι και τα εξής.

Πόρισμα 2.2 *Αν $A \in \mathbb{C}^{n,n}$ είναι Ερμιτιανός και θετικά ορισμένος τότε αν διαγράψουμε οσεσδήποτε γραμμές του και τις αντίστοιχες στήλες του ο πίνακας που απομένει είναι Ερμιτιανός και θετικά ορισμένος.*

Πόρισμα 2.3 *Αν $A \in \mathbb{C}^{n,n}$ είναι Ερμιτιανός και θετικά ορισμένος τα διαγώνια στοιχεία του είναι θετικά.*

Με βάση τα προηγούμενα είμαστε σε θέση να διατυπώσουμε και να αποδείξουμε μερικές βασικές προτάσεις, όπως το Θεώρημα του Cholesky, που δίνεται αμέσως μετά. Σημειώνεται ότι όλες οι προτάσεις αφορούν πίνακες A που είναι πραγματικοί και συμμετρικοί. Τονίζεται, όμως, ότι οι ίδιες ακριβώς προτάσεις ισχύουν και στην περίπτωση όπου ο A είναι γενικά Ερμιτιανός πίνακας με κάποιες απλές προφανείς τροποποιήσεις τόσο στις διατυπώσεις των προτάσεων όσο και στις αντίστοιχες αποδείξεις.

Θεώρημα 2.4 (Cholesky) *Εστω $A \in \mathbb{R}^{n,n}$, συμμετρικός και θετικά ορισμένος. Υπάρχει μοναδικός κάτω τριγωνικος πίνακας $L \in \mathbb{R}^{n,n}$ με $l_{ii} > 0,\ i = 1(1)n$, τ.ω. $LL^T = A$.*

Απόδειξη: Η απόδειξη θα γίνει με επαγωγή. Για 1×1 πίνακες έχουμε $A = [a_{11}] = [a_{11}^{\frac{1}{2}}][a_{11}^{\frac{1}{2}}]$. Επειδή δε $a_{11} > 0$ και $a_{11}^{\frac{1}{2}} > 0$, το θεώρημα ισχύει. Εστω ότι το θεώρημα ισχύει για κάθε πραγματικό, συμμετρικό και θετικά ορισμένο $(n-1) \times (n-1),\ n \geq 2$, πίνακα. Θα δειχτεί ότι ισχύει και για πίνακες $n \times n$. Εστω A ένας τέτοιος, πραγματικός συμμετρικός και θετικά ορισμένος, πίνακας για τον οποίο θέλουμε να αποδείξουμε ότι $A = LL^T$, με L να ικανοποιεί τις απαιτήσεις του θεωρήματος. Θεωρούμε το διαχωρισμό των A και L σε block μορφές, οπότε αρκεί να έχουμε

$$\left[\begin{array}{c|c} a_{11} & y^T \\ \hline y & \widetilde{A} \end{array}\right] = \left[\begin{array}{c|c} l_{11} & 0_{n-1}^T \\ \hline z & \widetilde{L} \end{array}\right] \left[\begin{array}{c|c} l_{11} & z^T \\ \hline 0_{n-1} & \widetilde{L}^T \end{array}\right], \tag{2.26}$$

όπου $y \in \mathbb{R}^{n-1}$, $\widetilde{A} \in \mathbb{R}^{n-1,n-1}$, $l_{11} > 0$, $0_{n-1} \in \mathbb{R}^{n-1}$, $z \in \mathbb{R}^{n-1}$ και $\widetilde{L}$ είναι κάτω τριγωνικός με $\widetilde{l}_{ii} > 0,\ i = 1(1)n-1$. Για να ισχύει η ισότητα (2.26) είναι φανερό ότι αρκεί να ισχύουν οι

παρακάτω ισότητες

$$\begin{aligned} a_{11} &= l_{11}^2, \\ y &= l_{11} z, \\ y^T &= l_{11} z^T, \\ zz^T + \widetilde{L}\widetilde{L}^T &= \widetilde{A}. \end{aligned} \tag{2.27}$$

Από την πρώτη των (2.27) προκύπτει ότι μπορεί να οριστεί μονοσήμαντα το $l_{11} = a_{11}^{\frac{1}{2}} > 0$. Από τη δεύτερη μπορεί επίσης να οριστεί μονοσήμαντα το διάνυσμα $z = \frac{1}{a_{11}^{\frac{1}{2}}} y \in I\!R^{n-1}$. Η τρίτη των (2.27) είναι απλά ισοδύναμη με τη δεύτερη. Τέλος για την ισχύ της τέταρτης των (2.27) αρκεί να μπορεί να οριστεί μονοσήμαντα ο πίνακας $\widetilde{L} \in I\!R^{n-1,n-1}$ ώστε να είναι κάτω τριγωνικός και τα διαγώνια στοιχεία του να ικανοποιούν τις $\widetilde{l}_{ii} > 0,\ i = 1(1)n-1$. Για το σκοπό αυτό αρκεί να δείξουμε ότι ο πίνακας $\widetilde{A} - zz^T$ είναι συμμετρικός και θετικά ορισμένος, οπότε σύμφωνα με την υπόθεση της τέλειας επαγωγής θα ισχύει γι' αυτόν το θεώρημα από το οποίο έπονται αμέσως οι απαιτήσεις μας για τον $\widetilde{L}$. Παρατηρούμε ότι ο $\widetilde{A}$ είναι συμμετρικός γιατί προκύπτει από τον A αν διαγράψουμε την πρώτη γραμμή και στήλη του. Επομένως $(\widetilde{A} - zz^T)^T = \widetilde{A} - zz^T$ και άρα ο $\widetilde{A} - zz^T$ είναι επίσης συμμετρικός. Για να αποδείξουμε ότι είναι θετικά ορισμένος θεωρούμε τον αρχικό πίνακα A, που είναι θετικά ορισμένος, και το διάνυσμα $w = [-\frac{1}{a_{11}} x^T y \mid x^T]^T \in I\!R^n$ με $x \in I\!R^{n-1} \setminus \{0\}$. Θα έχουμε

$$0 \ < \ w^T A w = [-\frac{1}{a_{11}} x^T y | x^T] \left[\begin{array}{c|c} a_{11} & y^T \\ \hline y & \widetilde{A} \end{array}\right] \left[\begin{array}{c} -\frac{1}{a_{11}} y^T x \\ \hline x \end{array}\right].$$

Μετά την εκτέλεση όλων των σημειωμένων πράξεων στο δεξιά μέλος των παραπάνω σχέσεων βρίσκεται ότι αυτό είναι ίσο με $x^T(\widetilde{A} - \frac{1}{a_{11}} y y^T)x = x^T(\widetilde{A} - zz^T)x$, που είναι θετικό $\forall\, x \in I\!R^{n-1} \setminus \{0\}$, και άρα ο ισχυρισμός του θεωρήματος αληθεύει και για πίνακες $n \times n$ που ολοκληρώνει την απόδειξη. □

Με βάση κυρίως το προηγούμενο θεώρημα είναι δυνατόν να αποδειχτούν τέσσερις προτάσεις, που δίνονται στη συνέχεια. Η τρίτη από αυτές αποτελεί το αντίστροφο του θεωρήματος του Cholesky ενώ η τέταρτη δίνει έναν ισοδύναμο ορισμό για να είναι θετικά ορισμένος ένας πραγματικός συμμετρικός πίνακας.

Πόρισμα 2.4 *Αν $A \in I\!R^{n,n}$, $A^T = A$ και θετικά ορισμένος τότε* $\det(A) > 0$.

Απόδειξη: Ο A πληροί τις υποθέσεις του θεωρήματος του Cholesky και άρα επιδέχεται ανάλυση κατά Cholesky. Εστω $A = LL^T$, όπου ο L ικανοποιεί τα συμπεράσματα του θεωρήματος του Cholesky. Θα έχουμε διαδοχικά $\det(A) = \det(LL^T) = \det(L)\det(L^T) = l_{11}^2 l_{22}^2 \cdots l_{nn}^2 > 0$. □

Πόρισμα 2.5 *Αν $A \in I\!R^{n,n}$, $A^T = A$ και θετικά ορισμένος πίνακας, τότε το απόλυτα μέγιστο στοιχείο του βρίσκεται επί της διαγωνίου του.*

Απόδειξη: Διαγράφοντας όλες τις γραμμές εκτός της i-οστής και j-οστής ($1 \leq i < j \leq n$) καθώς και τις αντίστοιχες στήλες του πίνακα που δόθηκε απομένει ο 2×2 πίνακας $A_{ij} = \begin{bmatrix} a_{ii} & a_{ij} \\ a_{ji} & a_{jj} \end{bmatrix}$, ο οποίος με βάση το Πόρισμα 2.2 θα είναι συμμετρικός και θετικά ορισμένος. Από το Πόρισμα 2.4 προκύπτει ότι $0 < \det(A_{ij}) = a_{ii}a_{jj} - a_{ij}a_{ji} \leq (\max\{a_{ii}, a_{jj}\})^2 - a_{ij}^2$ και άρα $\max\{a_{ii}, a_{jj}\} > |a_{ij}|$. Η θεώρηση όλων των δυνατών ζευγών $(i, j),\ i, j \in \{1, 2, \cdots, n\},\ i \neq j$, συμπληρώνει την απόδειξη της παρούσας πρότασης. □

Θεώρημα 2.5 *Αν $A \in \mathbb{R}^{n,n}$ επιδέχεται ανάλυση κατά* Cholesky, *τότε ο A είναι συμμετρικός και θετικά ορισμένος.*

Απόδειξη: Σύμφωνα με την υπόθεσή μας έχουμε ότι $A = LL^T$, με $L \in \mathbb{R}^{n,n}$ κάτω τριγωνικό και $l_{ii} > 0,\ i = 1(1)n$. Καταρχάς είναι $A^T = (LL^T)^T = LL^T = A$ και άρα ο A είναι συμμετρικός. Εξάλλου $\forall\ x \in \mathbb{R}^n \setminus \{0\}$ είναι $(x, Ax)_2 = (x, LL^T x)_2 = (L^T x, L^T x)_2 = ||L^T x||_2^2 \geq 0$. Αν $||L^T x||_2 = 0$ τότε $L^T x = 0$ και επειδή $\det(L^T) = l_{11}l_{22}\cdots l_{nn} > 0$ έπεται ότι $x = 0$, που είναι άτοπο. Αρα ο A είναι θετικά ορισμένος. □

Θεώρημα 2.6 *Εστω $A \in \mathbb{R}^{n,n}$ με $A^T = A$. Ο A είναι θετικά ορισμένος ανν όλοι οι κύριοι $p \times p$ υποπίνακές του $A_{p\times p}$, $p = 1(1)n$, (της άνω αριστερής γωνίας) έχουν ορίζουσες θετικές.*

Απόδειξη: Εστω ότι ο A είναι θετικά ορισμένος. Τότε, με βάση το Θεώρημα 2.3, όλοι οι κύριοι υποπίνακες στη διατύπωση της παρούσας πρότασης θα είναι πραγματικοί, συμμετρικοί και θετικά ορισμένοι. Αρα κάθε ένας από αυτούς, σύμφωνα με το Πόρισμα 2.4, θα έχει ορίζουσα θετική. Αντίστροφα, αν οι ορίζουσες των κύριων υποπινάκων του A είναι θετικές, τότε από την απόδειξη του Θεωρήματος 2.1 έχουμε ότι η ορίζουσα του $p\times p$ υποπίνακά του θα είναι ίση με το γινόμενο των οδηγών $a_{11}^{(1)}a_{22}^{(2)}\cdots a_{pp}^{(p)}$ (> 0). Από την τελευταία σχέση, για $p = 1(1)n$, έπεται ότι $a_{pp}^{(p)} > 0,\ \forall\ p = 1(1)n$. Από το συμπέρασμα αυτό και το μονοσήμαντο της παραγοντοποίησης του Crout, στην προκείμενη περίπτωση, λόγω της συμμετρικότητας του A, έπεται ότι $D = \mathrm{diag}(a_{11}^{(1)}, a_{22}^{(2)}, \cdots, a_{nn}^{(n)})$ και $U = L^T$. Οπότε, αν θέσουμε $D^{\frac{1}{2}} = \mathrm{diag}\left((a_{11}^{(1)})^{\frac{1}{2}}, \cdots, (a_{nn}^{(n)})^{\frac{1}{2}}\right)$ θα έχουμε $A = LDL^T = LD^{\frac{1}{2}}D^{\frac{1}{2}}L^T = (LD^{\frac{1}{2}})(LD^{\frac{1}{2}})^T$. Αρα ο A επιδέχεται ανάλυση κατά Cholesky και επομένως είναι θετικά ορισμένος. □

Για την εύρεση αλγόριθμου που θα υπολογίζει τα στοιχεία του πίνακα L στην παραγοντοποίηση (ή ανάλυση) Cholesky χρησιμοποιούμε την ισότητα $LL^T = A$, με A πραγματικό, συμμετρικό και θετικά ορισμένο, εκτελούμε τον πολλαπλασιασμό στο αριστερά μέλος και εξισώνουμε ένα προς ένα τα στοιχεία του κάτω τριγωνικού, έστω, πίνακα του γινομένου που προκύπτει με τα αντίστοιχα στοιχεία του A. Είναι δυνατόν να βρίσκουμε τα στοιχεία του L είτε προχωρώντας από την πρώτη γραμμή προς την n-οστή είτε από την πρώτη στήλη προς την n-οστή. Π.χ. προχωρώντας κατά γραμμές θα έχουμε διαδοχικά:

$l_{11}^2 = a_{11}$,
$l_{21}l_{11} = a_{21},\ l_{21}^2 + l_{22}^2 = a_{22}$,
$l_{31}l_{11} = a_{31},\ l_{31}l_{21} + l_{32}l_{22} = a_{32},\ l_{31}^2 + l_{32}^2 + l_{33}^2 = a_{33}$, κ.λπ.
Είναι φανερό από τις παραπάνω ισότητες ότι τα στοιχεία του L μπορούν να βρεθούν με τη σειρά $l_{11},\ l_{21},\ l_{22},\ l_{31},\ l_{32},\ l_{33}$, κ.λπ. Επομένως ένας αλγόριθμος Cholesky κατά γραμμές (σε συνοπτική μορφή) μπορεί να είναι ο ακόλουθος.

Αλγόριθμος Cholesky **κατά γραμμές:**

Δεδομένα: $A \in I\!R^{n,n},\ A^T = A,\ A$ θετικά ορισμένος.
Για $i = 1(1)n$
  Για $j = 1(1)i-1$
    $l_{ij} = (a_{ij} - \sum_{k=1}^{j-1} l_{ik}l_{jk})/l_{jj}$
  Τέλος 'Για'
  $l_{ii} = (a_{ii} - \sum_{j=1}^{i-1} l_{ij}^2)^{\frac{1}{2}}$
Τέλος 'Για'

Πολλές φορές η παραγοντοποίηση ενός πραγματικού, συμμετρικού και θετικά ορισμένου πίνακα A γράφεται με ένα λίγο διαφορετικό τρόπο ο οποίος είναι ήδη γνωστός και ως παραγοντοποίηση (ή ανάλυση) Crout. Συγκεκριμένα, έχοντας υπόψη την παραγοντοποίηση Cholesky, μπορούμε να γράψουμε αναλυτικά

$$A \;=\; LL^T \;=\; \begin{bmatrix} & & & & \\ l_{11} & & & & \\ l_{21} & l_{22} & & & \\ l_{31} & l_{32} & l_{33} & & \\ \vdots & \vdots & \vdots & \ddots & \\ l_{n1} & l_{n2} & l_{n3} & \cdots & l_{nn} \end{bmatrix} \begin{bmatrix} l_{11} & l_{21} & l_{31} & \cdots & l_{n1} \\ & l_{22} & l_{32} & \cdots & l_{n2} \\ & & l_{33} & \cdots & l_{n3} \\ & & & \ddots & \vdots \\ & & & & l_{nn} \end{bmatrix}$$

Το δεξιά μέλος των παραπάνω ισοτήτων γράφεται και ως εξής

$$A \;=\; \begin{bmatrix} 1 & & & & \\ \frac{l_{21}}{l_{11}} & 1 & & & \\ \frac{l_{31}}{l_{11}} & \frac{l_{32}}{l_{22}} & 1 & & \\ \vdots & \vdots & \vdots & \ddots & \\ \frac{l_{n1}}{l_{11}} & \frac{l_{n2}}{l_{22}} & \frac{l_{n3}}{l_{33}} & \cdots & 1 \end{bmatrix} \begin{bmatrix} l_{11}^2 & & & & \\ & l_{22}^2 & & & \\ & & l_{33}^2 & & \\ & & & \ddots & \\ & & & & l_{nn}^2 \end{bmatrix} \begin{bmatrix} 1 & \frac{l_{21}}{l_{11}} & \frac{l_{31}}{l_{11}} & \cdots & \frac{l_{n1}}{l_{11}} \\ & 1 & \frac{l_{32}}{l_{22}} & \cdots & \frac{l_{n2}}{l_{22}} \\ & & 1 & \cdots & \frac{l_{n3}}{l_{33}} \\ & & & \ddots & \vdots \\ & & & & 1 \end{bmatrix}$$

ή τέλος και ως

$$A \;=\; MDM^T,$$

όπου M κάτω τριγωνικός με

$$m_{ij} \;=\; \begin{cases} \frac{l_{ij}}{l_{jj}} & \alpha\nu\ 1 \;\leq\; j \;<\; i \;\leq\; n \\ 1 & \alpha\nu\ i = j = 1(1)n \\ 0 & \alpha\lambda\lambda\iota\omega\varsigma \end{cases}$$

και D διαγώνιος με διαγώνια στοιχεία $d_{ii} = l_{ii}^2,\ i = 1(1)n$.

ΑΣΚΗΣΕΙΣ

1.: Εστω $A \in \mathbb{C}^{n,n}$ με $A^H = A$ και $(Ax, x)_2 > 0\ \forall\ x \in \mathbb{C}^n \backslash \{0\}$. Να αποδειχτεί, χρησιμοποιώντας τον αντίστοιχο ορισμό και μόνο, ότι:
α) $a_{ii} > 0,\ i = 1(1)n.$ και
β) Αν $\lambda \in \sigma(A)$ τότε $\lambda > 0$.

2.: Να αποδειχτεί ότι αν $A \in \mathbb{R}^{n,n}$, με $A^T = A$, ο A είναι θετικά ορισμένος ανν $(Ax, x)_2 > 0,\ \ \forall\ x \in \mathbb{R}^n \backslash \{0\}$.

3.: Να δειχτεί ότι ο τριδιαγώνιος πίνακας

$$A \;=\; \mathrm{trid}(-1, 2, -1) \;=\; \begin{bmatrix} 2 & -1 & & & \\ -1 & 2 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 2 & -1 \\ & & & -1 & 2 \end{bmatrix} \in \mathbb{R}^{n,n}$$

είναι θετικά ορισμένος.

4.: Να δειχτεί ότι πίνακας $A = \mathrm{trid}(1, a, 1) \in \mathbb{R}^{n,n}$, με $a \geq 2$, είναι θετικά ορισμένος.

5.: Δίνεται ο πίνακας $A \;=\; \begin{bmatrix} 1 & a & a \\ a & 1 & a \\ a & a & 1 \end{bmatrix} \in \mathbb{R}^{3,3}$. Να βρεθούν όλες οι τιμές του a για τις οποίες ο A είναι θετικά ορισμένος.

6.: Δίνεται ο πραγματικός συμμετρικός πίνακας $A = \begin{bmatrix} 9 & 6 & 3 \\ 6 & 8 & 4 \\ 3 & 4 & 3 \end{bmatrix}$.
α) Να δειχτεί ότι ο A είναι θετικά ορισμένος.

β) Να βρεθούν ακριβώς οι παράγοντες L και L^T της παραγοντοποίησης Cholesky του A. και
γ) Να βρεθούν ακριβώς οι L και U παράγοντες της LU παραγοντοποίησης του A.
(Σημείωση: Τα ερωτήματα (α), (β) και (γ) μπορούν να απαντηθούν με οποιαδήποτε σειρά.)

7.: Δίνεται ο πραγματικός συμμετρικός πίνακας $A = \begin{bmatrix} 1 & 0 & 2 & 1 \\ 0 & 4 & 8 & 10 \\ 2 & 8 & 29 & 22 \\ 1 & 10 & 22 & 42 \end{bmatrix}$. Να γίνει η παραγοντοποίηση του A σε γινόμενο LDU, όπου L κάτω τριγωνικός με μονάδες στη διαγώνιο, D διαγώνιος και U άνω τριγωνικός με μονάδες στη διαγώνιο. Χρησιμοποιώντας τη συγκεκριμένη παραγοντοποίηση να αποδειχτεί ότι ο A είναι θετικά ορισμένος και στη συνέχεια να δοθεί η ανάλυση Cholesky του A.

8.: Δίνεται ο πραγματικός συμμετρικός πίνακας $A = \begin{bmatrix} 9 & 6 & 3 \\ 6 & 8 & 4 \\ 3 & 4 & 3 \end{bmatrix}$. Αφού διαπιστωθεί ότι ο A είναι θετικά ορισμένος να γίνει η ανάλυση Cholesky και στη συνέχεια να βρεθεί ο αντίστροφος του A χρησιμοποιώντας την ανάλυση Cholesky.

9.: Δίνεται ο πίνακας $A = \begin{bmatrix} 3 & c & 1 \\ c & 2 & c \\ 1 & c & 3 \end{bmatrix}$. Να βρεθεί η περιοχή για την πραγματική παράμετρο c για την οποία ο A είναι θετικά ορισμένος. Στη συνέχεια να γίνει η ανάλυση Cholesky για $c = 0$.

10.: Δίνεται ο πίνακας $A = \begin{bmatrix} 1 & -1 & 0 \\ -1 & a & -1 \\ 0 & -1 & a \end{bmatrix}$. Να βρεθεί η περιοχή για την πραγματική παράμετρο a για την οποία ο A είναι θετικά ορισμένος. Στη συνέχεια να βρεθεί ο αντίστροφος του A για $a = 2$ χρησιμοποιώντας ανάλυση Cholesky.

11.: Δίνεται το μιγαδικό γραμμικό σύστημα

$$\begin{array}{rcrcl} 25x & + & (20-9i)y & = & 36-4i \\ (20+9i)x & + & 25y & = & 36+4i \end{array}$$

Αφού μετατραπεί σε ισοδύναμο πραγματικό και διαπιστωθεί ότι ο πίνακας των συντελεστών του πραγματικού συστήματος είναι συμμετρικός και θετικά ορισμένος, να λυθεί, το αντίστοιχο πραγματικό, με τη μέθοδο Cholesky και να δοθούν οι τιμές των x και y.

12.: Να αποδειχτεί ότι ο πίνακας

$$A = \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & 2 & 2 & \cdots & 2 \\ 1 & 2 & 3 & \cdots & 3 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 2 & 3 & \cdots & n \end{bmatrix}$$

είναι θετικά ορισμένος και να γίνει η ανάλυση Cholesky αυτού.

13.: Δίνεται ότι ο πίνακας $A \in I\!R^{n,n}$ είναι συμμετρικός και θετικά ορισμένος. Να αποδειχτεί ότι τότε και ο πίνακας $\alpha A + \beta y y^T$ είναι συμμετρικός και θετικά ορισμένος για κάθε $\alpha > 0$, $\beta \geq 0$ και $y \in I\!R^n$. Με βάση ό,τι αποδειχτεί, να αποδειχτεί στη συνέχεια ότι ο πίνακας

$$\begin{bmatrix} \alpha+\beta & \beta & \cdots & \beta \\ \beta & \alpha+\beta & \cdots & \beta \\ \vdots & \vdots & \ddots & \vdots \\ \beta & \beta & \cdots & \alpha+\beta \end{bmatrix}$$

είναι συμμετρικός και θετικά ορισμένος. Χρησιμοποιώντας όλα τα παραπάνω να αποδειχτεί ότι και ο $\begin{bmatrix} 3 & 1 & 1 \\ 1 & 3 & 1 \\ 1 & 1 & 3 \end{bmatrix}$ είναι θετικά ορισμένος και να βρεθούν οι παράγοντες Cholesky, διατηρώντας κλάσματα και ριζικά στους υπολογισμούς.

14.: Δίνεται ο πίνακας

$$A = \begin{bmatrix} 1 & 0 & \cdots & 0 & \alpha \\ 0 & 1 & \cdots & 0 & \alpha \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & \alpha \\ \alpha & \alpha & \cdots & \alpha & 1 \end{bmatrix} \in I\!R^{n,n}.$$

α) Να βρεθούν όλες οι τιμές του α ώστε ο πραγματικός, συμμετρικός πίνακας A να είναι θετικά ορισμένος. και
β) Με την προϋπόθεση ότι ο A είναι θετικά ορισμένος να γίνει η παραγοντοποίηση Cholesky του πίνακα αυτού και να λυθεί το γραμμικό σύστημα $Ax = b$, με $b = [1\ 1\ 1\ \cdots\ 1\ (n-1)\alpha]^T$, χρησιμοποιώντας την υπόψη παραγοντοποίηση.

15.: Να αποδειχτεί ότι κατά τα διαδοχικά βήματα, $k = 1(1)n-1$, της απαλοιφής του Gauss χωρίς οδήγηση, ενός πραγματικού, συμμετρικού, θετικά ορισμένου πίνακα $A \in I\!R^{n,n}$, ο

εκάστοτε κύριος υποπίνακας της κάτω δεξιά γωνίας του $A^{(k+1)}$, $A^{(k+1)}_{n-k,n-k}$, εξακολουθεί να έχει την ιδιότητα του (πραγματικού) συμμετρικού και θετικά ορισμένου πίνακα. (Σημείωση: Να αποδειχτεί το ζητούμενο για τον πίνακα $A^{(2)}_{n-1,n-1}$ και **μόνο**.)

16.: Να αποδειχτεί ότι ένας πίνακας $A \in I\!R^{n,n}$, με $A^T = A$ και $a_{ii} > 0,\ i = 1(1)n$, που είναι *αυστηρά διαγώνια υπέρτερος κατά γραμμές (κατά στήλες)* είναι **και** θετικά ορισμένος.

17: Να γραφτεί συνοπτικά με μορφή ψευδοκώδικα, ο αλγόριθμος εύρεσης των στοιχείων του κάτω τριγωνικού πίνακα L στην παραγοντοποίηση Cholesky, κατά στήλες, ενός πραγματικού συμμετρικού και θετικά ορισμένου πίνακα $A \in I\!R^{n,n}$.

18: Ο πίνακας $A \in I\!R^{n,n}$ με στοιχεία

$$a_{ij} = a^{(1)}_{ij} = \frac{1}{i+j-1},\ i,j = 1(1)n,$$

είναι γνωστός ως πίνακας του Hilbert.
α) Να γίνουν δύο διαδοχικά βήματα απαλοιφής του Gauss, χωρίς οδήγηση, και να βρεθεί η έκφραση του στοιχείου $a^{(3)}_{ij},\ i,j = 3(1)n$, σα συνάρτηση των i και j.
β) Με βάση τις γενικές εκφράσεις των στοιχείων $a^{(2)}_{ij},\ i,j = 2(1)n$, και $a^{(3)}_{ij},\ i,j = 3(1)n$, στο (α), να συναχτεί η έκφραση του γενικού στοιχείου $a^{(k)}_{ij},\ i,j = k(1)n$, σα συνάρτηση των i και j, μετά από $k-1$ διαδοχικά βήματα απαλοιφής, και να αποδειχτεί με επαγωγή, ότι η αντίστοιχη έκφραση θα ισχύει και για το στοιχείο $a^{(k+1)}_{ij},\ i,j = k+1(1)n$, μετά από k βήματα απαλοιφής. (Σημείωση: Υποτίθεται ότι $k < n$.)
γ) Με βάση τις γενικές εκφράσεις των χρησιμοποιηθέντων οδηγών στοιχείων κατά τα διαδοχικά βήματα της απαλοιφής να αποδειχτεί ότι ο πίνακας A του Hilbert, εκτός από συμμετρικός, είναι και θετικά ορισμένος.

3 Επαναληπτικές Μέθοδοι

3.1 Εισαγωγή

Οι μέθοδοι για την επίλυση γραμμικών συστημάτων, που μελετήσαμε ως τώρα, καλούνται άμεσες. Τα κύρια χαρακτηριστικά τους είναι ότι για την εύρεση της λύσης απαιτείται ένας πεπερασμένος αριθμός πράξεων κι ακόμη ότι η λύση βρίσκεται ακριβώς αν εργαστούμε με ακριβή αριθμητική. Το τελευταίο, όμως, δεν είναι δυνατόν να υλοποιηθεί στην πράξη, εφόσον χρησιμοποιείται Υπολογιστής για την εύρεση της λύσης. Επομένως, τα αρχικά σφάλματα στρογγύλευσης καθώς και αυτά κατά τη διάρκεια των υπολογισμών με όλες τις συνέπειές τους δίνουν αποτελέσματα που είναι κατ΄ ανάγκην προσεγγιστικά. Αυτή η παρατήρηση μας οδηγεί στη σκέψη ότι αφού δεν είναι δυνατόν να βρεθεί η ακριβής λύση ίσως θα ήταν δυνατόν να χρησιμοποιηθεί μέθοδος η οποία θα αναζητά, αντί της ακριβούς λύσης, μια πολύ καλή προσέγγισή της. Πάνω στη λογική αυτή έχουν προταθεί κατά καιρούς, αρχής γενομένης από τα τέλη του 19ου αιώνα, διάφορες μέθοδοι οι οποίες ανήκουν στην κατηγορία των λεγόμενων έμμεσων ή επαναληπτικών. Στην πράξη αποδείχνονται πάρα πολύ αποτελεσματικές ιδιαίτερα όταν ο πίνακας των συντελεστών των αγνώστων του προς επίλυση γραμμικού συστήματος είναι μεγάλος και αραιός. Τελευταία, οι επαναληπτικές μέθοδοι έχουν γίνει πολύ δημοφιλείς διότι είναι καταλληλότερες για την επίλυση προβλημάτων όταν χρησιμοποιείται Υπολογιστής Παράλληλης Αρχιτεκτονικής.

Η βασική ιδέα της κλασικής επαναληπτικής μεθόδου, που θα αναπτυχτεί στη συνέχεια, είναι η εξής. Ξεκινάμε καταρχάς από μία αυθαίρετη προσέγγιση της λύσης, έστω $x^{(0)}$, και με βάση κάποιον επαναληπτικό αλγόριθμο κατασκευάζονται διαδοχικά οι όροι μιας ακολουθίας $\{x^{(k)}\}_{k=0}^{\infty}$, η οποία, κάτω από ορισμένες προϋποθέσεις, συγκλίνει στη λύση του προς επίλυση συστήματος. Πιο συγκεκριμένα, έστω

$$Ax = b,\ A \in \mathbb{C}^{n,n},\ \det(A) \neq 0,\ b \in \mathbb{C}^n, \tag{3.1}$$

το προς επίλυση σύστημα, μιγαδικό γενικά. Θεωρούμε μια διάσπαση του πίνακα A

$$A = M - N \tag{3.2}$$

με μόνους, προς το παρόν, περιορισμούς:
α) Ο πίνακας M να είναι αντιστρέψιμος, και
β) Ενα γραμμικό σύστημα με πίνακα συντελεστών αγνώστων M να λύνεται με πολύ λιγότερες πράξεις από ένα άλλο με πίνακα συντελεστών αγνώστων A.
(Σημείωση: Ο πίνακας M στη διάσπαση (3.2) είναι γνωστός ως (προρ)ρυθμιστής πίνακας.) Χρησιμοποιούμε την (3.2) στην (3.1) και αναδιατάσσοντας έχουμε

$$Mx = Nx + b. \tag{3.3}$$

Η (3.3) είναι ισοδύναμη με τη

$$x = Tx + c,\ \ T := M^{-1}N,\ c := M^{-1}b. \tag{3.4}$$

Η νέα εξίσωση (3.4), που είναι μια εξίσωση σταθερού σημείου και είναι ισοδύναμη με την αρχική (3.1), υποδείχνει την κατασκευή του αλγόριθμου

$$x^{(k+1)} = Tx^{(k)} + c,\ k = 0, 1, 2, \cdots, \tag{3.5}$$

με $x^{(0)} \in \mathcal{C}^n$ οποιοδήποτε. (Σημείωση: Ο πίνακας T στον αλγόριθμο (3.5) είναι γνωστός ως επαναληπτικός πίνακας του αλγόριθμου ή της επαναληπτικής μεθόδου.)
Παρατήρηση: Στην πράξη, και για ευνόητους λόγους, για την εύρεση του $x^{(k+1)}$ από το $x^{(k)}$, **δεν** εφαρμόζεται ο αλγόριθμος (3.5), όπου πρέπει να βρεθεί προηγουμένως ο πίνακας $T := M^{-1}N$, αλλά ο ισοδύναμός του

$$Mx^{(k+1)} = Nx^{(k)} + b,\ k = 0, 1, 2, \cdots. \tag{3.6}$$

Αυτό εξηγεί και την απαίτηση (β) για το ρυθμιστή πίνακα M.

Ο αλγόριθμος (3.5) παράγει ακολουθία διανυσμάτων $\{x^{(k)}\}_{k=0}^{\infty}$, η οποία, κάτω από ορισμένες προϋποθέσεις, συγκλίνει στη λύση $x = A^{-1}b$ του (3.1). Μπορεί να διαπιστωθεί ότι αν ο αλγόριθμος (3.5) παράγει ακολουθία συγκλίνουσα αυτή θα συγκλίνει στη λύση $x = A^{-1}b$. Πράγματι, έστω ότι $\lim_{k\to\infty} x^{(k)} = y$. Τότε, παίρνοντας τα όρια των μελών της (3.5) για $k \to \infty$, αντικαθιστώντας τις εκφράσεις για τα T και c από την (3.4) και ακολουθώντας αντίστροφη πορεία απ΄ αυτήν η οποία μας οδήγησε στην (3.4) καταλήγουμε στην $Ay = b$, οπότε $y = A^{-1}b$. Η αναγκαία και ικανή συνθήκη για τη σύγκλιση της ακολουθίας $\{x^{(k)}\}_{k=0}^{\infty}$ δίνεται στο επόμενο θεώρημα αμέσως μετά την απόδειξη του παρακάτω λήμματος.

Λήμμα 3.1 *Για τις διαδοχικές δυνάμεις του πίνακα $T \in \mathcal{C}^{n,n}$ ισχύει $\lim_{k\to\infty} T^k = 0$ ανν $\rho(T) < 1$.*

Απόδειξη: Ως γνωστόν η $\lim_{k\to\infty} T^k = 0$ συνεπάγεται τη $\lim_{k\to\infty} ||T^k|| = 0$ για κάθε φυσική norm. Είναι επίσης γνωστό ότι $\rho^k(T) = \rho(T^k) \le ||T^k||$ για κάθε φυσική norm. Αφού $\lim_{k\to\infty} ||T^k|| = 0$ θα υπάρχει φυσικός αριθμός k_0 τ.ω. για κάθε $k \ge k_0$ να ισχύει $||T^k|| < 1$. Από το τελευταίο συμπέρασμα και τις αμέσως προηγούμενες σχέσεις έπεται ότι $\rho^k(T) < 1$ για κάθε $k \ge k_0$, άρα και για κάθε k, πράγμα που συνεπάγεται $\rho(T) < 1$. Αντίστροφα, αν $\rho(T) < 1$ τότε, σύμφωνα με το Θεώρημα 1.10, για κάθε $\epsilon > 0$ θα υπάρχει φυσική norm τ.ω. $||T|| \le \rho(T) + \epsilon$. Αν επιλεχτεί το ϵ έτσι ώστε $0 < \epsilon < 1 - \rho(T)$ τότε θα έχουμε $\rho(T) + \epsilon < 1$ άρα και $||T|| < 1$ για τη συγκεκριμένη φυσική norm. Τότε όμως θα είναι $||T^k|| \le ||T||^k < 1$ για κάθε k άρα $\lim_{k\to\infty} ||T^k|| = 0$ για την υπόψη φυσική norm. Από την τελευταία σχέση έπεται αμέσως ότι $\lim_{k\to\infty} T^k = 0$, που αποδείχνει το παρόν λήμμα. □

Θεώρημα 3.1 *Αναγκαία και ικανή συνθήκη για τη σύγκλιση της ακολουθίας των παραγόμενων από τον αλγόριθμο (3.5) διανυσμάτων στη λύση $x = A^{-1}b$ του συστήματος (3.1) είναι η*

$$\rho(T) < 1.$$

Απόδειξη: Καταρχάς εισάγουμε το διάνυσμα-σφάλμα στην k επανάληψη ορίζοντας

$$e^{(k)} = x^{(k)} - x. \tag{3.7}$$

Αφαιρώντας κατά μέλη τη σχέση (3.4) από την (3.5) και χρησιμοποιώντας την (3.7) παίρνουμε

$$e^{(k+1)} = Te^{(k)},\ k = 0, 1, 2, \cdots,$$

με $e^{(0)} \in \mathcal{C}^n$ οποιοδήποτε, και με απλή επαγωγή βρίσκουμε ότι

$$e^{(k)} = T^k e^{(0)},\ k = 0, 1, 2, \cdots, \tag{3.8}$$

με $e^{(0)} \in \mathcal{C}^n$ οποιοδήποτε. Εφόσον επιζητούμε να έχουμε σύγκλιση, $\lim_{k\to\infty} x^{(k)} = A^{-1}b$, για οποιοδήποτε $x^{(0)}$, ή, ισοδύναμα, $\lim_{k\to\infty} e^{(k)} = 0$ για οποιοδήποτε $e^{(0)}$, παρατηρούμε ότι αν στη θέση του $e^{(0)}$ στην (3.8) θέσουμε διαδοχικά τα διανύσματα στήλες $e^j,\ j = 1(1)n$, του μοναδιαίου πίνακα I, παίρνουμε ως $e^{(k)}$ τις αντίστοιχες στήλες του πίνακα T^k. Επειδή δε θέλουμε να έχουμε $\lim_{k\to\infty} e^{(k)} = 0$ έπεται ότι οριακά οι στήλες του T^k θα τείνουν στο μηδενικό διάνυσμα ή ότι $\lim_{k\to\infty} T^k = 0$. Από το προηγούμενο, όμως, Λήμμα 3.1 γνωρίζουμε ότι η τελευταία ισότητα ισχύει ανν $\rho(T) < 1$, που αποτελεί την αναγκαία και ικανή συνθήκη για να συγκλίνει ο αλγόριθμος (3.5) στη λύση του συστήματος (3.1). □

Πόρισμα 3.1 *Μια ικανή συνθήκη για τη σύγκλιση του αλγόριθμου (3.5) στη λύση του συστήματος (3.1) είναι η*

$$||T|| < 1,$$

όπου $||\cdot||$ *μια οποιαδήποτε φυσική* norm.

Απόδειξη: Η απόδειξη είναι προφανής αφού λόγω της γνωστής σχέσης $\rho(T) \leq ||T||$ ισχύει το προηγούμενο θεώρημα. □

Πόρισμα 3.2 *Αν για κάποια φυσική* norm *και για κάποιο θετικό ακέραιο* k *για τον επαναληπτικό πίνακα* T *του αλγόριθμου (3.5) ισχύει ότι* $||T^k|| < 1$, *τότε ο αλγόριθμος συγκλίνει.*

Απόδειξη: Επειδή $\rho^k(T) = \rho(T^k) \leq ||T^k|| < 1$ έπεται ότι $\rho(T) < 1$. Αρα ο αλγόριθμος (3.5) συγκλίνει. □

Θα πρέπει να τονιστεί ότι στην πράξη κι εφόσον οι διαδοχικά παραγόμενοι όροι της ακολουθίας των διανυσμάτων από τον αλγόριθμο (3.5), στην περίπτωση σύγκλισης, τείνουν στη λύση στο όριο θεσπίζονται κριτήρια σταματήματος των επαναλήψεων. Συνηθέστερα κριτήρια είναι τα εξής δύο

$$||x^{(k+1)} - x^{(k)}|| \leq \epsilon \text{ και } \frac{||x^{(k+1)} - x^{(k)}||}{||x^{(k+1)}||} \leq \eta,$$

με επικρατέστερο το δεύτερο, όπου ϵ και η είναι μικροί θετικοί αριθμοί προκαθορισμένοι από το χρήστη. Το πρώτο κριτήριο παίζει το ρόλο του απόλυτου σφάλματος, αν κανείς θεωρήσει ότι το $x^{(k+1)}$ είναι η ακριβής λύση και το $x^{(k)}$ η προσεγγιστική, ενώ το δεύτερο παίζει το ρόλο του σχετικού απόλυτου σφάλματος. Συγκεκριμένα, αν ζητάμε προσέγγιση της λύσης με m σημαντικά ψηφία ίσως είναι λογικό να θεωρούμε το δεύτερο κριτήριο με norm την $\ell_\infty-$norm και $\eta = 0.5\times 10^{-m}$ ή, έχοντας υπόψη τη Σημείωση (δ) του Αλγόριθμου Επαναληπτικής Βελτίωσης στο τέλος του Κεφαλαίου 2, το κριτήριο

$$\frac{|x_i^{(k+1)} - x_i^{(k)}|}{|x_i^{(k+1)}|} \leq 0.5 \times 10^{-m},\ x_i^{(k+1)} \neq 0, \quad |x_i^{(k)}| \leq 0.5 \times 10^{-m},\ x_i^{(k+1)} = 0, \quad i = 1(1)n.$$

Σαν ένα παράδειγμα βασικής επαναληπτικής μεθόδου δίνουμε αυτό της επαναληπτικής βελτίωσης της λύσης πραγματικού ομαλού γραμμικού συστήματος $Ax = b$, που έχει βρεθεί, με την άμεση μέθοδο της παραγοντοποίησης του A σε γινόμενο LU, και δοθεί σε προηγούμενο κεφάλαιο. Υποθέτουμε ότι οι παράγοντες L και U είναι αρκετά ακριβείς και ότι κατά την εύρεσή τους δεν απαιτούνται εναλλαγές γραμμών του A. Εστω $x^{(0)}$ η λύση που παίρνουμε από την επίλυση των δυο αρχικών εξισώσεων (με προς τα μπρός και προς τα πίσω αντικαταστάσεις) στη μέθοδο επαναληπτικής βελτίωσης και έστω $x^{(k)}$ η εκάστοτε χρησιμοποιούμενη προσεγγιστική τιμή του x στην k ανακύκλωση. Εστω $r^{(k)}$, $y^{(k)}$ και $x^{(k+1)}$, οι ευρισκόμενες τιμές των διανυσμάτων r, y και της νέας τιμής του x, αντίστοιχα. Εχουμε διαδοχικά ότι

$$\begin{aligned} x^{(k+1)} &= x^{(k)} + y^{(k)} = x^{(k)} + (LU)^{-1}r^{(k)} \\ &= x^{(k)} + (LU)^{-1}(b - Ax^{(k)}) = (I - (LU)^{-1}A)x^{(k)} + (LU)^{-1}b. \end{aligned} \tag{3.9}$$

Από το παραπάνω επαναληπτικό σχήμα, και παίρνοντας όρια για $k \to \infty$, υπό την προϋπόθεση ότι $\lim_{k\to\infty} x^{(k)} = z$, μπορούμε να βρούμε αμέσως ότι $Az = b$ και άρα η ακολουθία των $x^{(k)}$ θα συγκλίνει στην ακριβή λύση του συστήματος που δόθηκε. Για να εξετάσουμε αν η ακολουθία συγκλίνει θεωρούμε τον επαναληπτικό πίνακα $T = I - (LU)^{-1}A$ του αλγόριθμου (3.9). Εφόσον οι πίνακες L και U είναι αρκετά ακριβείς θα πρέπει $A \approx LU$, οπότε $(LU)^{-1}A \approx I$ και $T = I - (LU)^{-1}A \approx 0$. Αφού τα στοιχεία του T είναι στην περιοχή του μηδενός είναι λογικό να περιμένουμε ότι για κάποια φυσική norm θα ισχύει $||T|| \approx 0 < 1$. Η τελευταία σχέση εξασφαλίζει (θεωρητικά) την ικανή συνθήκη για τη σύγκλιση της μεθόδου επαναληπτικής βελτίωσης στη λύση του συστήματος, ακόμη κι αν **δε** γίνουν κάποιοι ενδιάμεσοι υπολογισμοί με διπλή ακρίβεια, όπως είχε θεωρηθεί απαραίτητο, όταν παρουσιάστηκε η συγκεκριμένη μέθοδος.

Από τον ορισμό της φυσικής norm και την (3.8) συνάγεται ότι υπάρχει ένα $e^{(0)} \in \mathbb{C}^n \setminus \{0\}$, για κάθε συγκεκριμένο k, τ.ω. να ισχύει ότι $||e^{(k)}|| = ||T^k e^{(0)}|| = ||T^k||||e^{(0)}||$. Από τις τελευταίες ισότητες μπορούμε να πάρουμε ότι

$$\sup_{e^{(0)}\in\mathbb{C}^n\setminus\{0\}} \left(\frac{||e^{(k)}||}{||e^{(0)}||}\right) = ||T^k||. \tag{3.10}$$

Εχοντας υπόψη το Πόρισμα 3.2 και τη σχέση (3.10) δίνουμε τον ακόλουθο ορισμό.

Ορισμός 3.1 *Εστω $A, B \in \mathcal{C}^{n,n}$. Αν για κάποιο θετικό ακέραιο k και για μια φυσική* norm *είναι $||A^k|| < 1$, τότε η ποσότητα*

$$\mathcal{R}(A^k) = -\ln(||A^k||^{\frac{1}{k}}) = \frac{-\ln||A^k||}{k} \tag{3.11}$$

καλείται μέση ταχύτητα σύγκλισης για k επαναλήψεις. Αν $||B^k|| < 1$ και $\mathcal{R}(A^k) < \mathcal{R}(B^k)$, τότε ο B είναι επαναληπτικά ταχύτερος του A για k επαναλήψεις.

Ο παραπάνω ορισμός έχει ιδιαίτερη σημασία στο όριο, όταν δηλαδή $k \to \infty$. Τότε, και κάτω από τις προϋποθέσεις του ορισμού, μπορεί να αποδειχτεί, χρησιμοποιώντας την κανονική μορφή του Jordan, ότι

$$\mathcal{R}_\infty(A) = \lim_{k\to\infty} \mathcal{R}(A^k) = -\ln \rho(A).$$

Το τελευταίο συμπέρασμα, που είναι ανεξάρτητο της χρησιμοποιούμενης φυσικής norm λόγω της ισχύουσας ισοδυναμίας τους, στην περίπτωση συγκλινουσών επαναληπτικών μεθόδων (3.5), ερμηνεύεται ως εξής. "Οσο μικρότερη είναι η φασματική ακτίνα του επαναληπτικού πίνακα T τόσο ταχύτερα (ασυμπτωτικά) η ακολουθία $\{x^{(k)}\}_{k=0}^{\infty}$ συγκλίνει στη λύση του (3.1)". Στην πράξη το "ασυμπτωτικά" σημαίνει για "μεγάλες" τιμές του k.

Μετά την ανάλυση που προηγήθηκε το συμπέρασμα στο οποίο καταλήγει κανείς σ΄ ό,τι αφορά την επιλογή του ρυθμιστή πίνακα M στην (3.2) είναι ότι πέρα από τους δύο περιορισμούς που πρέπει να πληρούνται πρέπει ακόμη ο M να επιλέγεται έτσι ώστε αφενός $\rho(T) \equiv \rho(M^{-1}N) < 1$ και αφετέρου η $\rho(T)$ να είναι όσο το δυνατόν μικρότερη. Οι παρατηρήσεις, που μόλις έγιναν σχετικά με την επιλογή του M καθιστούν ίσως φανερό ότι αυτή δεν είναι πάντα τόσο εύκολη. Στη συνέχεια θα ασχοληθούμε με κλασικές περιπτώσεις επιλογής του M και παραπέρα ανάπτυξης των αντίστοιχων επαναληπτικών μεθόδων.

3.2 Κλασικές Επαναληπτικές Μέθοδοι

Οι κλασικές επαναληπτικές μέθοδοι βασίζονται στην ακόλουθη διάσπαση του πίνακα των συντελεστών των αγνώστων A

$$A = D - L - U, \tag{3.12}$$

όπου D =diag(A), δηλαδή, διαγώνιος πίνακας με διαγώνια στοιχεία τα αντίστοιχα του A, L αυστηρά κάτω τριγωνικός και U αυστηρά άνω τριγωνικός. Οπως είναι φανερό η διάσπαση (3.12) ορίζεται μονοσήμαντα.

3.2.1 **Μέθοδος** Jacobi

Στη μέθοδο του Jacobi ο ρυθμιστής πίνακας είναι ο $M = D$. Σ′ ό,τι αφορά την ικανοποίηση των δυο βασικών περιορισμών που αφορούν στο ρυθμιστή πίνακα M διαπιστώνουμε τα ακόλουθα. Καταρχάς ο M είναι αντιστρέψιμος ανν $\det(M) = \det(D) = a_{11}a_{22}\cdots a_{nn} \neq 0$. Επομένως η μέθοδος του Jacobi μπορεί να οριστεί ανν $a_{ii} \neq 0,\ i = 1(1)n$. Ακόμη είναι προφανές ότι ένα γραμμικό σύστημα με πίνακα συντελεστών αγνώστων $M = D$ της μορφής (3.6) και δεύτερο μέλος $(L+U)x^{(k)}+b$ είναι πολύ οικονομικότερο σε πράξεις για να λυθεί (απαιτεί πράξεις της τάξης του $\mathcal{O}(n^2)$) από ό,τι είναι ένα σύστημα με πίνακα συντελεστών αγνώστων A (απαιτεί $\mathcal{O}(n^3)$ πράξεις με τη μέθοδο απαλοιφής Gauss). Αρα ικανοποιείται και ο δεύτερος περιορισμός εφόσον ικανοποιείται ο πρώτος. Σ′ ό,τι αφορά τη σύγκλιση της μεθόδου τα πάντα εξαρτιούνται από το αν $\rho(T) = \rho(M^{-1}N) = \rho(D^{-1}(L+U)) < 1$, οπότε η μέθοδος συγκλίνει αλλιώς δε συγκλίνει. Σε μορφή πινάκων η μέθοδος του Jacobi είναι η ακόλουθη

$$x^{(k+1)} = D^{-1}(L+U)x^{(k)} + D^{-1}b,\ k = 0, 1, 2, \cdots, \tag{3.13}$$

με $x^{(0)} \in \mathbb{C}^n$ οποιοδήποτε. Αν θελήσουμε να βρούμε την οποιαδήποτε συνιστώσα του νέου διανύσματος $x^{(k+1)}$ στην k επανάληψη σα συνάρτηση γνωστών συνιστωσών του παλιού $x^{(k)}$ ή/και του νέου διανύσματος πολλαπλασιάζουμε τα μέλη της (3.13) από τα αριστερά επί D, οπότε προκύπτει

$$Dx^{(k+1)} = (L+U)x^{(k)} + b.$$

Αν τώρα εκτελέσουμε τις πράξεις και στα δύο μέλη και εξισώσουμε τις i-οστές συνιστώσες των διανυσμάτων των δύο μελών και λύσουμε ως προς την i-οστή συνιστώσα $x_i^{(k+1)}$ του $x^{(k+1)}$ παίρνουμε αμέσως ότι

$$x_i^{(k+1)} = (b_i - \sum_{j=1,\ j\neq i}^{n} a_{ij}x_j^{(k)})/a_{ii},\ i = 1(1)n. \tag{3.14}$$

3.2.2 **Μέθοδος** Gauss-Seidel

Στην περίπτωση της μεθόδου των Gauss-Seidel και με βάση πάντα τη διάσπαση (3.12) επιλέγεται $M = D - L$. Για να υπάρχει η μέθοδος θα πρέπει να υπάρχει ο αντίστροφος του $D - L$, που είναι κάτω τριγωνικός πίνακας. Αρα αυτός είναι αντιστρέψιμος ανν $\det(D - L) \neq 0$, που ισοδυναμεί με $a_{ii} \neq 0,\ i = 1(1)n$, όπως ακριβώς και στη μέθοδο του Jacobi. Ακόμη, ένα σύστημα με πίνακα συντελεστών αγνώστων $D - L$, λύνεται με προς τα μπρός αντικαταστάσεις (κι απαιτεί συνολικά πράξεις πλήθους $\mathcal{O}(n^2)$) και επομένως είναι οικονομικότερο στην επίλυσή του από ένα σύστημα με πίνακα συντελεστών A. Ικανοποιείται κι ο δεύτερος περιορισμός σ′ ό,τι αφορά τον M με την προϋπόθεση ότι ικανοποιείται ο πρώτος. Η μέθοδος των Gauss-Seidel είναι επομένως η ακόλουθη

$$x^{(k+1)} = (D-L)^{-1}Ux^{(k)} + (D-L)^{-1}b,\ k = 0, 1, 2, \cdots, \tag{3.15}$$

με $x^{(0)} \in \mathbb{C}^n$ οποιοδήποτε. Για τη σύγκλιση της μεθόδου θα πρέπει $\rho((D-L)^{-1}U) < 1$. Για την εύρεση αναλυτικών εκφράσεων για τις συνιστώσες του νέου διανύσματος εργαζόμαστε όπως και πριν. Δηλαδή, πολλαπλασιάζουμε από τα αριστερά τα μέλη της (3.15) επί $D-L$, εκτελούμε τις πράξεις και στα δύο μέλη, εξισώνουμε τις i-οστές συνιστώσες των δύο μελών και τέλος λύνουμε ως προς $x_i^{(k+1)}$, οπότε έχουμε

$$x_i^{(k+1)} = (b_i - \sum_{j=1}^{i-1} a_{ij}x_j^{(k+1)} - \sum_{j=i+1}^{n} a_{ij}x_j^{(k)})/a_{ii},\ i = 1(1)n. \tag{3.16}$$

3.2.3 Jacobi και Gauss-Seidel Μέθοδοι

Αν παρατηρήσει κανείς προσεκτικά τις μεθόδους των Jacobi και Gauss-Seidel θα δεί ότι έχουν πάρα πολλές ομοιότητες και συγχρόνως κάποιες σημαντικές διαφορές. Ιδιαίτερα, αυτό μπορεί να φανεί, αν η (3.14) γραφτεί έτσι ώστε το άθροισμα μέσα στην παρένθεση να διασπαστεί σε δύο αθροίσματα

$$x_i^{(k+1)} = (b_i - \sum_{j=1}^{i-1} a_{ij}x_j^{(k)} - \sum_{j=i+1}^{n} a_{ij}x_j^{(k)})/a_{ii},\ i = 1(1)n.$$

Τότε μπορεί να διαπιστωθεί το εξής. Για την εύρεση μιας συνιστώσας $x_i^{(k+1)}$ της νέας επανάληψης στη μέθοδο του Jacobi χρησιμοποιούνται όλες οι άλλες συνιστώσες της προηγούμενης επανάληψης ενώ στη μέθοδο των Gauss-Seidel χρησιμοποιούνται όλες οι επόμενες συνιστώσες της προηγούμενης επανάληψης και όλες οι προηγούμενες συνιστώσες της τρέχουσας επανάληψης. Πρακτικά, στην Gauss-Seidel χρησιμοποιούνται όλες οι πιο πρόσφατες διαθέσιμες συνιστώσες. Το τελευταίο στοιχείο ίσως κάνει τη μέθοδο των Gauss-Seidel ελκυστικότερη από αυτήν του Jacobi γιατί "φαίνεται" να πλεονεκτεί στην περίπτωση σύγκλισης. Αυτό δεν είναι πάντοτε αληθές. Υπάρχουν περιπτώσεις όπου η μέθοδος του Jacobi συγκλίνει ενώ η μέθοδος των Gauss-Seidel αποκλίνει! Ακόμη, αν πάψει κανείς να σκέφτεται σειριακά (ακολουθιακά), τότε, αν έχουμε στη διάθεσή μας έναν Υπολογιστή με παράλληλη αρχιτεκτονική, στη μεν μέθοδο του Jacobi είναι δυνατόν όλες οι συνιστώσες της νέας επανάληψης να βρεθούν σε ένα χρονικό βήμα από όλες της προηγούμενης επανάληψης ενώ στη μέθοδο των Gauss-Seidel αυτό δεν είναι δυνατόν. Απ΄ ό,τι μπορεί να διαπιστώσει κανείς από τους τύπους (3.16) για να βρεθεί η συνιστώσα $x_2^{(k+1)}$ θα πρέπει να έχει βρεθεί προηγουμένως η συνιστώσα $x_1^{(k+1)}$, για να βρεθεί η συνιστώσα $x_3^{(k+1)}$ θα πρέπει να έχουν βρεθεί προηγουμένως οι συνιστώσες $x_2^{(k+1)}$ και $x_1^{(k+1)}$ κ.ο.κ. Αρα, δεν είναι δυνατόν να βρεθούν όλες οι συνιστώσες της νέας επανάληψης σε ένα χρονικό βήμα, όπως στην περίπτωση της μεθόδου του Jacobi.

Στη συνέχεια παρουσιάζουμε μια κλασική περίπτωση όπου και οι δύο μέθοδοι συγκλίνουν και αφορούν πίνακες της ειδικής κατηγορίας των αυστηρά διαγώνια υπέρτερων κατά γραμμές ή κατά

στήλες. Ακολουθεί η διατύπωση και απόδειξη μιας σειράς προτάσεων μία από τις οποίες, που είναι πολύ σπουδαιότερης και γενικότερης σημασίας, παρουσιάζεται ως θεώρημα και δίνεται χωρίς απόδειξη.

Λήμμα 3.2 *Εστω ότι ο πίνακας* $A \in \mathbb{C}^{n,n}$ *είναι αυστηρά διαγώνια υπέρτερος κατά γραμμές* $(|a_{ii}| > \sum_{j=1,\ j\neq i}^{n} |a_{ij}|,\ i = 1(1)n)$ *(ή κατά στήλες* $(|a_{jj}| > \sum_{i=1,\ i\neq j}^{n} |a_{ij}|,\ j = 1(1)n)$*). Τότε είναι αντιστρέψιμος.*

Απόδειξη: Εστω ότι ο A, που είναι αυστηρά διαγώνια υπέρτερος κατά γραμμές, δεν είναι αντιστρέψιμος. Τότε υπάρχει $x \in \mathbb{C}^n \setminus \{0\}$ τ.ω. $Ax = 0$. Εστω $i \in \{1, 2, \cdots, n\}$ ο δείκτης για τον οποίο $|x_i| = ||x||_\infty$. Από την $Ax = 0$ θα έχουμε για την i-οστή συνιστώσα του Ax ότι

$$\sum_{j=1}^{n} a_{ij}x_j = 0 \iff a_{ii}x_i = -\sum_{j=1,\ j\neq i}^{n} a_{ij}x_j.$$

Διαιρώντας τα δύο μέλη δια $|x_i|$, παίρνοντας απόλυτες τιμές και εφαρμόζοντας απλές ιδιότητες έχουμε ότι

$$|a_{ii}| \leq \sum_{j=1,\ j\neq i}^{n} |a_{ij}|\frac{|x_j|}{|x_i|} \leq \sum_{j=1,\ j\neq i}^{n} |a_{ij}|,$$

αφού $|x_i| = ||x||_\infty \geq |x_j|,\ \forall\ j = 1(1)n$. Αρα ο A δεν μπορεί να είναι αυστηρά διαγώνια υπέρτερος κατά γραμμές, πράγμα που είναι άτοπο.
Αν τώρα ο A είναι αυστηρά διαγώνια υπέρτερος κατά στήλες θεωρούμε τον A^T, που είναι αυστηρά διαγώνια υπέρτερος κατά γραμμές, οπότε ισχύει η προηγούμενη απόδειξη. □

Σημείωση: Τονίζεται ότι το αντιστρέψιμο ενός πίνακα της κατηγορίας που εξετάστηκε αποδείχνεται και με την απαλοιφή Gauss, όπως αυτό γίνεται σε Ασκηση του προηγούμενου Κεφαλαίου.

Θεώρημα 3.2 *(Κανονική μορφή* Jordan*) Για κάθε πίνακα* $A \in \mathbb{C}^{n,n}$ *υπάρχει αντιστρέψιμος πίνακας* $S \in \mathbb{C}^{n,n}$ *τ.ω. ο* $S^{-1}AS$ *έχει τη μορφή*

$$S^{-1}AS \ = \ J \ = \ \mathrm{diag}(J_1,\ J_2,\ \cdots,\ J_p), \tag{3.17}$$

όπου

$$J_i \ = \ \begin{bmatrix} \lambda_i & 1 & 0 & \cdots & 0 & 0 \\ 0 & \lambda_i & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda_i & 1 \\ 0 & 0 & 0 & \cdots & 0 & \lambda_i \end{bmatrix} \in \mathbb{C}^{n_i,n_i},\ i = 1(1)p,\ \sum_{i=1}^{p} n_i = n. \tag{3.18}$$

(Σημείωση: Εκτός από τυχόν μεταθέσεις των διαγώνιων blocks *του* J*, η κανονική μορφή* Jordan *(3.17)-(3.18) είναι μοναδική.)*

Παρατηρήσεις: α) Προφανώς τα λ_i στις διαγωνίους των J_i blocks δίνουν τις ιδιοτιμές του A. β) Είναι δυνατόν οι ίδιες ιδιοτιμές λ_i να παρουσιάζονται σε περισσότερα από ένα blocks. γ) Το πλήθος των γραμμικά ανεξάρτητων ιδιοδιανυσμάτων του πίνακα A είναι ίσο με p και μάλιστα κάθε ένα από αυτά συνδέεται με την πρώτη στήλη ενός και μόνον block του J. δ) Ενας πίνακας $A \in \mathbb{C}^{n,n}$ έχει n γραμμικά ανεξάρτητα ιδιοδιανύσματα, που είναι οι στήλες του πίνακα S, ανν ο J είναι διαγώνιος (δηλαδή αν όλα τα blocks είναι 1×1 πίνακες ή, ισοδύναμα, αν $p = n$). ε) *Αλγεβρική πολλαπλότητα* της ιδιοτιμής $\lambda \in \sigma(A)$ καλείται το συνολικό πλήθος των ιδιοτιμών του A (ή του J) που είναι ίσο με λ, ενώ *γεωμετρική πολλαπλότητα* καλείται η διάσταση του γραμμικού διανυσματικού χώρου που παράγεται από τα γραμμικά ανεξάρτητα ιδιοδιανύσματα του A (είναι το ίδιο με το πλήθος των blocks του J), που συνδέονται με αυτήν.

Με βάση την κανονική μορφή του Jordan είναι τώρα δυνατόν να δοθεί η απόδειξη του Θεωρήματος 1.10, το οποίο διατυπώνουμε εδώ για μια ακόμη φορά.

Θεώρημα 3.3 $\forall\ A \in \mathbb{C}^{n,n}$ *και* $\forall\ \epsilon > 0$ *υπάρχει φυσική* norm $||\cdot||$ *τ.ω.* $||A|| \leq \rho(A) + \epsilon$.

Απόδειξη: Για την απόδειξη, ορίζουμε τον πίνακα

$$\widetilde{J} = D^{-1} J D, \quad D = \mathrm{diag}(1, \epsilon, \epsilon^2, \cdots, \epsilon^{n-1}), \tag{3.19}$$

όπου J ο πίνακας που δίνεται στις (3.17)–(3.18). Είναι φανερό ότι ο πίνακας (3.19) θα έχει μια ανάλογη μορφή αυτής της (3.17) με τα blocks του αντίστοιχα αυτών στην (3.18) με τη μόνη διαφορά ότι οι μονάδες των J_i θα έχουν αντικατασταθεί από ϵ στους $\widetilde{J}_i$. Συγκεκριμένα,

$$\widetilde{J}_i \ = \ \begin{bmatrix} \lambda_i & \epsilon & 0 & \cdots & 0 & 0 \\ 0 & \lambda_i & \epsilon & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda_i & \epsilon \\ 0 & 0 & 0 & \cdots & 0 & \lambda_i \end{bmatrix} \in \mathbb{C}^{n_i, n_i},\ i = 1(1)p.$$

Επομένως $||\widetilde{J}_i||_\infty \leq |\lambda_i| + \epsilon$, με αυστηρή ανισότητα να ισχύει όταν ο $\widetilde{J}_i$, είναι 1×1 block, που συνεπάγεται $||\widetilde{J}||_\infty \leq \rho(\widetilde{J}) + \epsilon$. Από τους ορισμούς των $\widetilde{J}$ και J έχουμε διαδοχικά ότι

$$\widetilde{J} = D^{-1} J D = D^{-1} S^{-1} A S D = (SD)^{-1} A (SD) = T^{-1} A T, \tag{3.20}$$

όπου τέθηκε $T = SD$. Προφανώς ο T θα έχει στήλες τις στήλες του S πολλαπλασιασμένες, αντίστοιχα, επί 1, ϵ, ϵ^2, $\cdots$, ϵ^{n-1}.
Η απεικόνιση, όμως, $||\cdot|| := ||T^{-1}\ .||_\infty : \mathbb{C}^n \to \mathbb{R}^{+,0}$ είναι εύκολο να αποδειχτεί ότι ορίζει μια διανυσματική norm στο $\mathbb{C}^n$. Εστω $||A||$ η φυσική norm στο $\mathbb{C}^{n,n}$ που παράγεται από τη διανυσματική

norm που μόλις ορίστηκε. Για τη φυσική αυτή norm θα έχουμε διαδοχικά

$$\begin{aligned} ||A|| &= \max_{x\in\mathbb{C}^n,\ ||x||=1} ||Ax|| = \max_{x\in\mathbb{C}^n,\ ||T^{-1}x||_\infty=1} ||T^{-1}Ax||_\infty \\ &= \max_{T^{-1}x\in\mathbb{C}^n,\ ||T^{-1}x||_\infty=1} ||T^{-1}AT(T^{-1}x)||_\infty = \max_{T^{-1}x\in\mathbb{C}^n,\ ||T^{-1}x||_\infty=1} ||\widetilde{J}(T^{-1}x)||_\infty \\ &= ||\widetilde{J}||_\infty \leq \rho(A) + \epsilon, \end{aligned}$$

που αποδείχνει το θεώρημα. □

Ορισμός 3.2 *Δυο πίνακες $A, B \in \mathbb{C}^{n,n}$ λέγονται όμοιοι αν υπάρχει αντιστρέψιμος πίνακας $C \in \mathbb{C}^{n,n}$ τέτοιος ώστε*

$$A = CBC^{-1}.$$

Λήμμα 3.3 *Αν $A, B \in \mathbb{C}^{n,n}$ είναι όμοιοι πίνακες τότε $\sigma(A) = \sigma(B)$.*

Απόδειξη: Με βάση τον ορισμό της ομοιότητας μπορεί να προκύψει αμέσως ως πόρισμα του Θεωρήματος 3.2. □

Λήμμα 3.4 *Εστω $A,\ B\ \in\ \mathbb{C}^{n,n}$ και έστω ότι ο ένας από τους δυο, π.χ. ο A, είναι αντιστρέψιμος. Τότε ισχύει ότι $\sigma(AB) = \sigma(BA)$.*

Απόδειξη: Μπορεί να γίνει χρήση της ομοιότητας δύο πινάκων. Πραγματικά, τότε ο BA και ο $A(BA)A^{-1} = AB$ είναι όμοιοι και άρα, με βάση το προηγούμενο λήμμα, $\sigma(AB) = \sigma(BA)$. □

Σημείωση: Η ισχύς του προηγούμενου λήμματος μπορεί να δειχτεί και στην περίπτωση όπου αμφότεροι οι A και B είναι μή αντιστρέψιμοι πίνακες. Αυτό γίνεται με εφαρμογή θεωρίας διαταράξεων και με βάση το γεγονός ότι οι ιδιοτιμές, που είναι ρίζες ενός πολυωνύμου, είναι συνεχείς συναρτήσεις των στοιχείων του πίνακα, δηλαδή των συντελεστών του αντίστοιχου χαρακτηριστικού πολυωνύμου.

Επανερχόμενοι στις επαναληπτικές μεθόδους Jacobi και Gauss-Seidel μπορούμε να αποδείξουμε την παρακάτω πρόταση.

Θεώρημα 3.4 *Αν $A \in \mathbb{C}^{n,n}$ είναι αυστηρά διαγώνια υπέρτερος κατά γραμμές (ή στήλες) τότε*
οι αντίστοιχες μέθοδοι των Jacobi *και* Gauss-Seidel *συγκλίνουν.*

Απόδειξη: Αποδείχνουμε πρώτα την ισχύ της πρότασης για τη μέθοδο του Jacobi και στην περίπτωση που ο A είναι αυστηρά διαγώνια υπέρτερος κατά γραμμές. Τότε θα ισχύει

$$|a_{ii}| > \sum_{j=1,\ j\neq i}^{n} |a_{ij}|,\ i = 1(1)n. \tag{3.21}$$

Από την (3.21) έπεται ότι $|a_{ii}| > 0$ και άρα $a_{ii} \neq 0,\ i = 1(1)n$, πράγμα που συνεπάγεται ότι τόσο η μέθοδος του Jacobi όσο και αυτή των Gauss-Seidel ορίζονται. Ο επαναληπτικός πίνακας του Jacobi $T_J = D^{-1}(L+U)$ θα δίνεται αναλυτικά από τη

$$T_J \ = \ \begin{bmatrix} 0 & -\frac{a_{12}}{a_{11}} & \cdots & -\frac{a_{1n}}{a_{11}} \\ -\frac{a_{21}}{a_{22}} & 0 & \cdots & -\frac{a_{2n}}{a_{22}} \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{a_{n1}}{a_{nn}} & -\frac{a_{n2}}{a_{nn}} & \cdots & 0 \end{bmatrix}.$$

Επομένως θα έχουμε

$$||T_J||_\infty = \max_{i=1(1)n} \sum_{j=1,\ j\neq i}^{n} \left| \ -\tfrac{a_{ij}}{a_{ii}} \ \right| = \max_{i=1(1)n} \frac{\sum_{j=1,\ j\neq i}^{n} |a_{ij}|}{|a_{ii}|} \ < \ 1,$$

όπου η τελευταία ανισότητα ισχύει λόγω της (3.21). Αρα η μέθοδος του Jacobi συγκλίνει.
Οταν ο A είναι αυστηρά διαγώνιος υπέρτερος κατά στήλες, τότε αντί της (3.21) έχουμε την

$$|a_{ii}| > \sum_{j=1, j\neq i}^{n} |a_{ji}|,\ i = 1(1)n. \tag{3.22}$$

Καταρχάς παρατηρούμε ότι λόγω της (3.22), όπως και στην προηγούμενη περίπτωση, θα είναι $a_{ii} \neq 0,\ i = 1(1)n$, και άρα αμφότερες οι μέθοδοι Jacobi και Gauss-Seidel ορίζονται. Για τη μέθοδο του Jacobi έχουμε ότι $T_J = D^{-1}(L+U)$, και επειδή λόγω του Λήμματος 3.4 είναι $\sigma(D^{-1}(L+U)) = \sigma((L+U)D^{-1})$, για τον πίνακα $T'_J = (L+U)D^{-1}$ θα ισχύει

$$||T'_J||_1 = \max_{i=1(1)n} \sum_{j=1,\ j\neq i}^{n} \left| \ -\tfrac{a_{ji}}{a_{ii}} \ \right| = \max_{i=1(1)n} \frac{\sum_{j=1,\ j\neq i}^{n} |a_{ji}|}{|a_{ii}|} \ < \ 1,$$

όπου η τελευταία ανισότητα ισχύει λόγω της (3.22). Επομένως θα είναι $\rho(T_J) = \rho(T'_J) \leq ||T'_J||_1 < 1$ και άρα η μέθοδος του Jacobi συγκλίνει.

Για τη μέθοδο των Gauss-Seidel στην περίπτωση που ο A είναι αυστηρά διαγώνια υπέρτερος κατά γραμμές η μέθοδος της απόδειξης είναι διαφορετική. Εστω ότι $T_{GS} = (D-L)^{-1}U$ είναι ο επαναληπτικός πίνακας της μεθόδου για την οποία υποθέτουμε ότι δε συγκλίνει. Τότε θα υπάρχει $\lambda \in \sigma(T_{GS})$ τ.ω. $|\lambda| \geq 1$. Για την υπόψη ιδιοτιμή θα έχουμε διαδοχικά

$$\begin{gathered} 0 = \det(T_{GS} - \lambda I) = \det((D-L)^{-1}U - \lambda I) \\ \iff \det(U - \lambda(D-L)) = 0 \iff \det(D - L - \tfrac{1}{\lambda}U) = 0. \end{gathered} \tag{3.23}$$

Αποδείξαμε έτσι ότι ο πίνακας $A(\lambda) \equiv D - L - \frac{1}{\lambda}U$ είναι μή αντιστρέψιμος. Για τον $A(\lambda)$ όμως έχουμε

$$\sum_{j=1,\ j\neq i}^{n} |a_{ij}(\lambda)| \ = \ \sum_{j=1}^{i-1} |a_{ij}| + \frac{1}{|\lambda|} \sum_{j=i+1}^{n} |a_{ij}| \ \leq \ \sum_{j=1,\ j\neq i}^{n} |a_{ij}| \ < \ |a_{ii}| \ = \ |a_{ii}(\lambda)|,$$

όπου η τελευταία δεξιά ανισότητα ισχύει λόγω της (3.21). Συνεπώς ο $A(\lambda)$ είναι αυστηρά διαγώνια υπέρτερος κατά γραμμές και άρα με βάση το Λήμμα 3.2 είναι αντιστρέψιμος. Αυτό όμως αντίκειται στην $\det(D - L - \frac{1}{\lambda}U) = 0$, που βρέθηκε προηγουμένως. Επομένως η μέθοδος των Gauss-Seidel συγκλίνει.
Στην περίπτωση που ο A είναι αυστηρά διαγώνια υπέρτερος κατά στήλες η αντίστοιχη απόδειξη είναι παρόμοια και παραλείπεται. □

3.3 Τεχνική της Παρεκβολής (Extrapolation)

Στην εισαγωγική παράγραφο του παρόντος κεφαλαίου περιγράφτηκε η γενική επαναληπτική μέθοδος για τη λύση του γραμμικού συστήματος (3.1). Οπως είδαμε, αν $\rho(T) < 1$, όπου T ο επαναληπτικός πίνακας της μεθόδου, τότε η μέθοδος συγκλίνει αλλιώς ($\rho(T) \geq 1$) η μέθοδος αποκλίνει. Υπάρχουν τεχνικές οι οποίες κάτω από ορισμένες προϋποθέσεις είναι δυνατόν να εφαρμοστούν σε μια συγκλίνουσα μέθοδο και να την κάνουν να συγκλίνει (ασυμπτωτικά) ταχύτερα ή σε μια αποκλίνουσα μέθοδο και να τη μετατρέψουν, μερικές φορές, σε συγκλίνουσα. Μία από τις τεχνικές αυτές είναι και η τεχνική της παρεκβολής (extrapolation).

Υποθέτουμε ότι δίνεται η επαναληπτική μέθοδος (3.5) για την επίλυση του (3.1). Με βάση τη διάσπαση (3.2), από την οποία προήρθε η μέθοδος (3.5), θεωρούμε μια νέα διάσπαση του A. Συγκεκριμένα την

$$A \;=\; M_\omega - N_\omega, \quad M_\omega := \frac{1}{\omega}M \tag{3.24}$$

και $\omega \in \mathbb{C}\backslash\{0\}$ μία παράμετρος που θα καλείται εφεξής παράμετρος της παρεκβολής (extrapolation). Εφόσον στην αρχική μέθοδο ο ρυθμιστής πίνακας M ικανοποιεί τους δυο βασικούς περιορισμούς, δηλαδή είναι αντιστρέψιμος και ένα σύστημα με πίνακα συντελεστών αγνώστων M είναι οικονομικότερο στη λύση του από ένα άλλο με πίνακα συντελεστών αγνώστων A, είναι φανερό από τον ορισμό του M_ω ότι και ο πίνακας αυτός πληροί αυτούς τους περιορισμούς. Με βάση τη διάσπαση (3.24) το νέο επαναληπτικό σχήμα θα είναι το παρακάτω

$$x^{(k+1)} = T_\omega x^{(k)} + c_\omega, \;\; k = 0, 1, 2, \cdots, \tag{3.25}$$

με $x^{(0)} \in \mathbb{C}^n$ οποιοδήποτε, όπου

$$T_\omega := M_\omega^{-1} N_\omega \equiv (1 - \omega)I + \omega T, \;\; c_\omega := \omega c. \tag{3.26}$$

Είναι φανερό ότι για $\omega = 1$, $M_\omega \equiv M$, $T_\omega \equiv T$ και $c_\omega \equiv c$. Με άλλα λόγια το νέο γενικευμένο επαναληπτικό σχήμα, που προτείνεται, περιορίζεται στο αρχικό. Επειδή όμως δεν τέθηκε κανείς περιορισμός στο ω εκτός από το να μην είναι μηδέν συμπεραίνεται ότι ο επαναληπτικός αλγόριθμος (3.25) περιγράφει μία οικογένεια επαναληπτικών αλγόριθμων ένας από τους οποίους είναι και ο αρχικός.

Είναι εύκολο να διαπιστωθεί ότι οι πίνακες T_ω και T έχουν τα ίδια ιδιοδιανύσματα οι δε ιδιοτιμές τους συνδέονται με τη σχέση

$$\mu = 1 - \omega + \omega\lambda, \quad \lambda \in \sigma(T),\ \mu \in \sigma(T_\omega).$$

Π.χ., αν $\lambda \in \sigma(T)$ και x είναι το αντίστοιχο ιδιοδιάνυσμα έχουμε

$$T_\omega x = ((1-\omega)I + \omega T)x = (1-\omega)x + \omega\lambda x = (1 - \omega + \omega\lambda)x,$$

οπότε $\mu = 1 - \omega + \omega\lambda \in \sigma(T_\omega)$ με αντίστοιχο ιδιοδιάνυσμα x. Αντίστροφα, αν $\mu \in \sigma(T_\omega)$ με αντίστοιχο ιδιοδιάνυσμα x, τότε από την (3.26) προκύπτει ότι $T = (1 - \frac{1}{\omega})I + \frac{1}{\omega}T_\omega$, οπότε $Tx = ((1-\frac{1}{\omega})I + \frac{1}{\omega}T_\omega)x = (1 - \frac{1}{\omega} + \frac{1}{\omega}\mu)x$ από την οποία συμπεραίνουμε ότι $\lambda = 1 - \frac{1}{\omega} + \frac{1}{\omega}\mu \in \sigma(T)$ με αντίστοιχο ιδιοδιάνυσμα x. Η τελευταία σχέση των ιδιοτιμών μπορεί να γραφτεί και ως $\mu = 1 - \omega + \omega\lambda$. Αρα για την εύρεση του καλύτερου δυνατού παρεκβαλλόμενου (extrapolated) επαναληπτικού σχήματος είναι φανερό ότι θα πρέπει να βρεθεί η τιμή του ω που ελαχιστοποιεί τη $\rho(T_\omega)$ και επομένως να λυθεί το ακόλουθο πρόβλημα βελτιστοποίησης

$$\min_{\omega \in \mathbb{C}\setminus\{0\}} \rho(T_\omega) \ \equiv \min_{\omega \in \mathbb{C}\setminus\{0\}} \max_{\lambda \in \sigma(T)} |1 - \omega + \omega\lambda|. \tag{3.27}$$

Το πρόβλημα βελτιστοποίησης που παρουσιάζεται στην (3.27) δεν είναι πάντα τόσο εύκολο να λυθεί, πέρα από το γεγονός φυσικά ότι απαιτείται η γνώση του $\sigma(T)$. Γενικά είναι δυσκολότερο και πολυπλοκότερο πρόβλημα από αυτό της επίλυσης του αρχικού συστήματος (3.1). Μπορεί να αποδειχτεί ότι αν είναι γνωστό ένα κυρτό πολύγωνο στο κλειστό εσωτερικό του οποίου βρίσκεται το $\sigma(T)$, και δεν περιέχει το σημείο $(1,0)$ του μιγαδικού επιπέδου, τότε το πρόβλημα (3.27) λύνεται και μάλιστα μονοσήμαντα. Μια τέτοια απλή περίπτωση παρουσιάζουμε στη συνέχεια.

Εστω ότι ο πίνακας A στο σύστημα (3.1) έχει πραγματικές ιδιοτιμές $\lambda_i,\ i = 1(1)n$, διατεταγμένες ως εξής

$$\lambda_1 \ \leq \ \lambda_2 \ \leq \ \cdots \ \leq \ \lambda_n.$$

Εστω ακόμη ότι για τη λύση του δοθέντος συστήματος προτείνεται ως ρυθμιστής ο πίνακας $M = I$. Το επαναληπτικό σχήμα που κατασκευάζεται για το σύστημα που δόθηκε είναι προφανώς το

$$x^{(k+1)} = (I - A)x^{(k)} + b,\ k = 0, 1, 2, \cdots, \tag{3.28}$$

με $x^{(0)} \in \mathbb{C}^n$ οποιοδήποτε, και επαναληπτικό πίνακα $T = I - A$. Για να συγκλίνει ο αλγόριθμος (3.28) θα πρέπει $\rho(I - A) < 1$. Επειδή όμως οι ιδιοτιμές του $I - A$ είναι οι $1 - \lambda_i,\ i = 1(1)n$, η μέθοδος θα συγκλίνει ανν $\max\{|1 - \lambda_1|,\ |1 - \lambda_n|\} \ < 1$, πράγμα που συνεπάγεται ότι θα πρέπει $0 < \lambda_1 \leq \lambda_n < 2$. Οι περιορισμοί αυτοί περιορίζουν και την κλάση των πινάκων A για τους οποίους η μέθοδος (3.28) συγκλίνει. Γεννιέται, λοιπόν, το εύλογο ερώτημα αν και κατά πόσο η τεχνική της extrapolation, με $\omega \in \mathbb{R} \setminus \{0\}$, είναι δυνατόν αφενός μεν στην περίπτωση σύγκλισης να δώσει για κάποια τιμή του ω τη βέλτιστη δυνατή αφετέρου δε στην περίπτωση απόκλισης να δώσει τη

βέλτιστη δυνατή συγκλίνουσα μέθοδο, εφόσον υπάρχει. Για να δοθεί απάντηση στο ερώτημα αυτό θα πρέπει να λυθεί το πρόβλημα (3.27). Το επαναληπτικό σχήμα της extrapolated μεθόδου, που βασίζεται στην (3.28), θα είναι το

$$x^{(k+1)} = ((1-\omega)I + \omega(I-A))x^{(k)} + \omega b \equiv (I - \omega A)x^{(k)} + \omega b,\ k = 0, 1, 2, \cdots,$$

οπότε εύκολα μπορεί να διαπιστωθεί ότι οι ιδιοτιμές του επαναληπτικού πίνακα $T_\omega = I - \omega A$ είναι οι $1 - \omega\lambda_i,\ i = 1(1)n$. Ακόμη μπορεί να συμπεράνει κανείς, από τη διάταξη των λ_i, ότι για $\omega > 0$ η διάταξη των ιδιοτιμών του T_ω είναι η

$$1 - \omega\lambda_1 \geq 1 - \omega\lambda_2 \geq \cdots \geq 1 - \omega\lambda_n$$

ενώ για $\omega < 0$ η αντίστοιχη διάταξη θα είναι

$$1 - \omega\lambda_1 \leq 1 - \omega\lambda_2 \leq \cdots \leq 1 - \omega\lambda_n.$$

Οπως διαπιστώνεται, οι ακραίες ιδιοτιμές είναι πάντα οι $1 - \omega\lambda_1$ και $1 - \omega\lambda_n$ και άρα

$$\rho(T_\omega) = \max\{|1 - \omega\lambda_1|,\ |1 - \omega\lambda_n|\}. \tag{3.29}$$

Συνεπώς οι τιμές του ω για τις οποίες η extrapolated μέθοδος συγκλίνει θα είναι εκείνες, εφόσον υπάρχουν, για τις οποίες συναληθεύουν οι τέσσερις ανισώσεις

$$-1 < 1 - \omega\lambda_1 < 1,\ \ -1 < 1 - \omega\lambda_n < 1.$$

Παρατηρούμε αμέσως από τις δυο δεξιές ανισώσεις, στα δυο ζεύγη των ανισώσεων, ότι για να έχουν αυτές συναληθεύουσα διαφορετική από το κενό σύνολο θα πρέπει οι λ_1 και λ_n να είναι ομόσημες. Αρα θα πρέπει οι πραγματικές ιδιοτιμές του A να είναι ή όλες θετικές ή όλες αρνητικές. Εστω, λοιπόν, ότι $\lambda_1 > 0$. Τότε προκύπτει αμέσως ότι και $\omega > 0$, η δε συναληθεύουσα των τεσσάρων ανισώσεων είναι η

$$0 \ < \ \omega \ < \ \frac{2}{\lambda_n}.$$

Για $\lambda_n < 0$, μπορεί αντίστοιχα να βρεθεί ότι $\omega < 0$ και η συναληθεύουσα είναι η

$$\frac{2}{\lambda_1} \ < \ \omega \ < \ 0.$$

Για να βρεθεί η βέλτιστη τιμή του ω στην περίπτωση $\lambda_1 > 0$ θεωρούμε ότι το $\omega \in (0, \infty)$ μεταβάλλεται συνεχώς από το 0 ως το ∞, και για να βρούμε τη συγκεκριμένη έκφραση για τη φασματική ακτίνα $\rho(T_\omega)$ στη (3.29), ως συνάρτηση του ω, βρίσκουμε το πρόσημο της διαφοράς $|1-\omega\lambda_1|-|1-\omega\lambda_n|$. Εχουμε διαδοχικά

$$\begin{aligned}&\text{sign}(|1 - \omega\lambda_1| - |1 - \omega\lambda_n|) \ = \ \text{sign}((1 - \omega\lambda_1)^2 - (1 - \omega\lambda_n)^2)\\ &= \ \text{sign}((2 - \omega(\lambda_n + \lambda_1))\omega(\lambda_n - \lambda_1)) \ = \text{sign}((2 - \omega(\lambda_n + \lambda_1)),\end{aligned}$$

$\lambda_n > \lambda_1$, οπότε η $\rho(T_\omega)$ δίνεται από τις συγκεκριμένες εκφράσεις

$$\rho(T_\omega) = \begin{cases} 1 - \omega\lambda_1, & \alpha\nu\nu\ \omega \in (0, \frac{2}{\lambda_1+\lambda_n}], \\ \omega\lambda_n - 1, & \alpha\nu\nu\ \omega \in [\frac{2}{\lambda_1+\lambda_n}, \infty). \end{cases} \tag{3.30}$$

(Σημείωση: Η περίπτωση $\lambda_n = \lambda_1$ πρέπει να εξεταστεί χωριστά. Στο τέλος όμως δίνει τα ίδια αποτελέσματα που μπορεί κανείς να βρεί στις τελικές εκφράσεις που θα προκύψουν από την ανάλυση αν θέσουμε $\lambda_n = \lambda_1$.) Ετσι βρίσκεται αμέσως ότι η $\rho(T_\omega)$ στο διάστημα $(0, \frac{2}{\lambda_n+\lambda_1}]$ είναι αυστηρά φθίνουσα ενώ στο $[\frac{2}{\lambda_n+\lambda_1}, \infty)$ αυστηρά αύξουσα κι άρα παρουσιάζει ελάχιστο για $\omega_\beta = \frac{2}{\lambda_n+\lambda_1}$. Συγκεκριμένα έχουμε

$$\min_{\omega \in I\!R\setminus\{0\}} \rho(T_\omega) = \rho(T_{\omega_\beta}) = \frac{\lambda_n - \lambda_1}{\lambda_n + \lambda_1}, \quad \omega_\beta = \frac{2}{\lambda_n + \lambda_1}. \tag{3.31}$$

Στην περίπτωση που είναι $\lambda_n < 0$, μία αντίστοιχη ανάλυση οδηγεί στα εξής συμπεράσματα

$$\min_{\omega \in I\!R\setminus\{0\}} \rho(T_\omega) = \rho(T_{\omega_\beta}) = \frac{\lambda_1 - \lambda_n}{\lambda_n + \lambda_1}, \quad \omega_\beta = \frac{2}{\lambda_n + \lambda_1}. \tag{3.32}$$

Προφανώς οι (3.31) και (3.32) συμπτύσσονται στις

$$\min_{\omega \in I\!R\setminus\{0\}} \rho(T_\omega) = \rho(T_{\omega_\beta}) = \frac{\lambda_n - \lambda_1}{|\lambda_n + \lambda_1|}, \quad \omega_\beta = \frac{2}{\lambda_n + \lambda_1}.$$

Η τεχνική της extrapolation μπορεί να εφαρμοστεί σε κάθε επαναληπτική μέθοδο της γενικής μορφής (3.5), που περιγράψαμε. Ετσι μπορεί να εφαρμοστεί τόσο στη μέθοδο Jacobi όσο και στη μέθοδο Gauss-Seidel. Για την extrapolated Jacobi θα έχουμε σε μορφή πινάκων

$$\begin{aligned} x^{(k+1)} &= [(1-\omega)I + \omega D^{-1}(L+U)]x^{(k)} + \omega D^{-1}b \ = \ D^{-1}(D - \omega A)x^{(k)} + \omega D^{-1}b, \\ & k = 0, 1, 2, \cdots, \end{aligned}$$

με $x^{(0)} \in \mathbb{C}^m$ οποιοδήποτε. Απ' αυτήν, αν εργαστούμε όπως στην περίπτωση της απλής μεθόδου Jacobi, μπορούμε να πάρουμε για την i-οστή συνιστώσα της νέας επανάληψης $x^{(k+1)}$ την έκφραση

$$x_i^{(k+1)} = (1-\omega)x_i^{(k)} + \omega \left(b_i - \sum_{j=1,\ j\neq i}^{n} a_{ij}x_j^{(k)} \right) / a_{ii}, \ i = 1(1)n. \tag{3.33}$$

Για την extrapolated Gauss-Seidel τα αντίστοιχα συμπεράσματα παρατίθενται χωρίς απόδειξη

$$\begin{aligned} & x^{(k+1)} \ = \ [(1-\omega)I + \omega(D-L)^{-1}U]\, x^{(k)} + \omega(D-L)^{-1}b \\ & = \ (D-L)^{-1}[(D-L) - \omega A]x^{(k)} + \omega(D-L)^{-1}b, \\ & k = 0, 1, 2, \cdots, \end{aligned}$$

με $x^{(0)} \in \mathbb{C}^n$ οποιοδήποτε και

$$x_i^{(k+1)} = (1-\omega)x_i^{(k)} + \omega \left(b_i - \sum_{j=1,\ j\neq i}^{n} a_{ij}x_j^{(k)} \right) / a_{ii} + \sum_{j=1}^{i-1} a_{ij}(x_j^{(k)} - x_j^{(k+1)})/a_{ii},\ i = 1(1)n. \quad (3.34)$$

ΑΣΚΗΣΕΙΣ

1.: Δίνεται το γραμμικό σύστημα $Ax = b,\ A \in \mathbb{R}^{n,n},\ \det(A) \neq 0,\ b \in \mathbb{R}^n$. Για τη λύση του συστήματος θεωρείται η διάσπαση $A = M - N$, και προτείνεται ο αλγόριθμος

$$Mx^{(k+1)} = Nx^{(k)} + b,\ k = 0, 1, 2, \ldots,$$

$x^{(0)} \in \mathbb{R}^n$ οποιοδήποτε. Να δειχτεί ότι αν $||A^{-1}N|| < \frac{1}{2}$, για κάποια φυσική norm, ο αλγόριθμος που προτείνεται συγκλίνει στη λύση του δοθέντος συστήματος.

2.: Δίνεται ότι $A \in \mathbb{C}^{n,n}$. Να δειχτεί ότι:
α) $\rho(A) < 1$ ανν ο $I - A$ είναι αντιστρέψιμος και η σειρά $\sum_{i=0}^{k} A^i$ συγκλίνει καθώς $k \to \infty$. (**Υπόδειξη**: Να χρησιμοποιηθεί η ταυτότητα

$$(I - A)\left(I + A + A^2 + \cdots + A^{k-1}\right) \equiv I - A^k.)$$

β) Επιπλέον, αν $\rho(A) < 1$ τότε

$$(I - A)^{-1} = \sum_{i=0}^{\infty} A^i.$$

3.: Δίνεται το γραμμικό σύστημα $Ax = b$, όπου ο πίνακας A είναι ο $A = \begin{bmatrix} 1 & 2 & -2 \\ 1 & 1 & 1 \\ -1 & -1 & 2 \end{bmatrix}$. Τι μπορεί να λεχτεί για τη σύγκλιση των αντίστοιχων μεθόδων Jacobi και Gauss-Seidel για την επίλυση του συστήματος που δόθηκε;

4.: Να εξεταστεί αν συγκλίνουν οι μέθοδοι Jacobi και Gauss-Seidel, που συνδέονται με το γραμμικό σύστημα

$$\begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 3 \\ -5 \\ 5 \end{bmatrix},$$

και να εκτελεστούν δύο επαναλήψεις κάθε μιας με αρχικό διάνυσμα $x^{(0)} = [1\ 1\ 1]^T$ και διατηρώντας κλάσματα στους υπολογισμούς.

5.: Δίνεται ο πίνακας $A = \begin{bmatrix} 1 & \alpha & \alpha \\ \alpha & 1 & 0 \\ \alpha & 0 & 1 \end{bmatrix}$. Να βρεθούν όλες οι δυνατές τιμές του $\alpha \in \mathbb{R}$ ώστε:
α) Η αντίστοιχη μέθοδος Jacobi να συγκλίνει. και
β) Η μέθοδος Gauss-Seidel να συγκλίνει.

6.: Να βρεθούν όλες οι δυνατές τιμές του $\alpha \in \mathbb{R}$ ώστε η επαναληπτική μέθοδος Jacobi, που συνδέεται με τον πίνακα $A = \begin{bmatrix} 1 & \alpha & \alpha \\ \alpha & 1 & \alpha \\ \alpha & \alpha & 1 \end{bmatrix}$, να συγκλίνει.

7.: α) Να αποδειχτεί ότι ο επαναληπτικός πίνακας των Gauss-Seidel έχει τουλάχιστον μια ιδιοτιμή ίση με μηδέν. και
β) Να αποδειχτεί ότι αν ο επαναληπτικός πίνακας των Gauss-Seidel έχει μια ιδιοτιμή ίση με τη μονάδα, τότε ο αρχικός πίνακας A του συστήματος είναι **μη** αντιστρέψιμος.

8.: Αν στο επαναληπτικό σχήμα $x^{(m+1)} = Tx^{(m)} + c$, όπου $T \in \mathbb{C}^{n,n}$, $c \in \mathbb{C}^n$ και $x^{(0)} \in \mathbb{C}^n$ οποιοδήποτε, ισχύει $\rho(T) = 0$, να αποδειχτεί ότι η λύση του συστήματος $(I - T)x = c$ μπορεί να βρεθεί **ακριβώς**! Ποιος είναι ο **μικρότερος** αριθμός επαναλήψεων που απαιτείται ώστε να είναι βέβαιο ότι έχει βρεθεί η ακριβής λύση του συστήματος; (**Υπόδειξη**: Να χρησιμοποιηθεί η κανονική μορφή Jordan του T.)

9.: Για την επίλυση του γραμμικού συστήματος $Ax = b$, με $A \in \mathbb{C}^{n,n}$, $\det(A) \neq 0$, $b \in \mathbb{C}^n$, προτείνεται η επαναληπτική μέθοδος

$$x^{(k+1)} = (I - A)x^{(k)} + b, \ k = 0, 1, 2, \ldots, \tag{3.35}$$

$x^{(0)} \in \mathbb{C}^n$ οποιοδήποτε, καθώς και η ακόλουθη

$$x^{(k+1)} = (I - \omega A)x^{(k)} + \omega b, \ k = 0, 1, 2, \ldots, \tag{3.36}$$

$x^{(0)} \in \mathbb{C}^n$ οποιοδήποτε. Δίνεται, επιπλέον, ότι ο πίνακας A έχει μόνο δύο διακριτές ιδιοτιμές τις: ι) -10, -5, ιι) -5, 5, και ιιι) 5, 10.
α) Να αποδειχτεί ότι η μέθοδος (3.35) αποκλίνει και στις τρεις περιπτώσεις. και
β) Να βρεθούν όλες οι πραγματικές τιμές του ω σε κάθε μία περίπτωση, αν υπάρχουν, ώστε η αντίστοιχη μέθοδος (3.36) να συγκλίνει.

10.: Να αποδειχτούν οι τύποι των μεθόδων της Extrapolated Jacobi (3.33) και της Extrapolated Gauss-Seidel (3.34), αντίστοιχα.

11.: Δίνεται το γραμμικό σύστημα $Ax = b$, όπου $A = \begin{bmatrix} 1 & 4 \\ -4 & 1 \end{bmatrix}$.
α) Να αποδειχτεί ότι η μέθοδος Jacobi αποκλίνει.
β) Να βρεθούν όλες οι δυνατές τιμές $\omega \in I\!R$ ώστε η παρεκβαλλόμενη μέθοδος Jacobi να συγκλίνει. και
γ) Να βρεθεί η βέλτιστη παράμετρος παρεκβολής ω.

12.: Για τη λύση του συστήματος $Ax = b$ με $A = \begin{bmatrix} 1 & 2 \\ -1 & 2 \end{bmatrix}$ προτείνονται οι μέθοδοι Jacobi, Gauss-Seidel και η παρεκβαλλόμενη μέθοδος Gauss-Seidel.
α) Να βρεθεί αν οι μέθοδοι Jacobi και Gauss-Seidel συγκλίνουν. και
β) Να βρεθούν όλες οι τιμές της παραμέτρου παρεκβολής $\omega \in I\!R$ για τις οποίες η παρεκβαλλόμενη μέθοδος Gauss-Seidel συγκλίνει.

13.: Δίνεται το γραμμικό σύστημα $Ax = b$ με $A = \begin{bmatrix} 2 & 1 & 0 & 0 \\ 1 & 2 & 0 & 0 \\ 0 & 0 & 4 & 1 \\ 0 & 0 & 1 & 4 \end{bmatrix}$.
α) Να εξεταστούν ως προς τη σύγκλιση οι μέθοδοι Jacobi και Gauss-Seidel .
β) Να βρεθούν όλες οι τιμές της παραμέτρου $\omega \in I\!R$ για τις οποίες η παρεκβαλλόμενη Gauss-Seidel συγκλίνει. και
γ) Να βρεθεί η βέλτιστη τιμή της παραμέτρου ω.

14.: Δίνεται το γραμμικό σύστημα $Ax = b$ με $A = \begin{bmatrix} 1 & -1 & 0 \\ -1 & 2 & -2 \\ 0 & -2 & -2 \end{bmatrix}$.
α) Να εξεταστούν ως προς τη σύγκλιση οι μέθοδοι Jacobi και Gauss-Seidel.
β) Να βρεθούν όλες οι τιμές της παραμέτρου $\omega \in I\!R$ για τις οποίες η Παρεκβαλλόμενη Gauss-Seidel συγκλίνει. και
γ) Να βρεθεί η βέλτιστη τιμή της παραμέτρου ω.

15.: Δίνεται το γραμμικό σύστημα $Ax = b$ με $A = \begin{bmatrix} 1 & -2 & 2 \\ -1 & 1 & -1 \\ -2 & -2 & 1 \end{bmatrix}$.
α) Να εξεταστούν ως προς τη σύγκλιση οι μέθοδοι Jacobi και Gauss-Seidel.
β) Να γίνουν τέσσερις επαναλήψεις με τη μέθοδο που συγκλίνει. και
γ) Τι μπορεί να παρατηρηθεί σε ό,τι αφορά την ακρίβεια του αποτελέσματος;

16.: Για την επίλυση του γραμμικού συστήματος $(I - T)x = c$, $T \in \mathbb{C}^{n,n}$, $c \in \mathbb{C}^n$, θεωρείται ο αλγόριθμος $x^{(k+1)} = Tx^{(k)} + c$, $k = 0, 1, 2, \ldots$, $x^{(0)} \in \mathbb{C}^n$ οποιοδήποτε, καθώς και ο αντίστοιχος παρεκβαλλόμενος (extrapolated) αλγόριθμος με παράμετρο παρεκβολής

(extrapolation) $\omega \in I\!R\backslash\{0\}$. Αν οι ιδιοτιμές του T είναι $\mu_j = i\alpha_j$, με $\alpha_j \in [-\rho, \rho]$, $j = 1(1)n$, να βρεθούν, με βάση τα δεδομένα του προβλήματος, οι "καλύτερες" δυνατές τιμές του ω και της $\rho(T_\omega)$ (φασματική ακτίνα του επαναληπτικού πίνακα του extrapolated αλγόριθμου).

17.: Για την επίλυση του γραμμικού συστήματος $(I - T)x = c$, $T \in C\!\!\!C^{n,n}$, $c \in C\!\!\!C^n$, προτείνεται το επαναληπτικό σχήμα

$$x^{(k+1)} = Tx^{(k)} + c,\ k = 0, 1, 2, \ldots, \tag{3.37}$$

$x^{(0)} \in C\!\!\!C^n$ οποιοδήποτε. Εστω

$$x^{(k+1)} = [(1 - \omega_1)I + \omega_1 T]x^{(k)} + \omega_1 c,\ k = 0, 1, 2, \ldots, \tag{3.38}$$

$x^{(0)} \in C\!\!\!C^n$ οποιοδήποτε, το παρεκβαλλόμενο σχήμα του (3.37) με παράμετρο παρεκβολής ω_1. Να δοθεί το παρεκβαλλόμενο σχήμα του (3.38) με παράμετρο παρεκβολής ω_2 και να αποδειχτεί ότι το νέο σχήμα είναι ένα παρεκβαλλόμενο σχήμα του (3.37) με παράμετρο παρεκβολής $\omega = \omega_1\omega_2$.

18.: Δίνεται η επαναληπτική μέθοδος $x^{(k+1)} = Tx^{(k)} + c$, $k = 0, 1, 2, \ldots$, $T \in C\!\!\!C^{n,n}$, $c \in C\!\!\!C^n$, $x^{(0)} \in C\!\!\!C^n$ δοσμένο, για την επίλυση του συστήματος $(I - T)x = c$. Εστω ότι οι μόνες διακριτές ιδιοτιμές του T είναι οι αριθμοί 0 και -2.
α) Τι μπορεί να λεχτεί για τη σύγκλιση του δοθέντος επαναληπτικού σχήματος;
β) Να κατασκευαστεί το παρεκβαλλόμενο σχήμα του δοθέντος.
γ) Να βρεθούν οι τιμές της παραμέτρου παρεκβολής $\omega \in I\!R$ ώστε το παρεκβαλλόμενο σχήμα να συγκλίνει. και
δ) Να βρεθεί η βέλτιστη τιμή της παραμέτρου ω και η αντίστοιχη της φασματικής ακτίνας του επαναληπτικού πίνακα του παρεκβαλλόμενου σχήματος.

3.4 Μέθοδος της Διαδοχικής Υπερχαλάρωσης (SOR)

Η μέθοδος της Διαδοχικής Υπερχαλάρωσης, γνωστή διεθνώς ως SOR, από τα αρχικά της αγγλικής ορολογίας Successive Over-Relaxation, αποτελεί μια μονοπαραμετρική γενίκευση της μεθόδου Gauss-Seidel. Συγκεκριμένα, θεωρώντας τη διάσπαση (3.12) με $\det(D) \neq 0$ ορίζουμε το ρυθμιστή πίνακα ως

$$M_\omega = \frac{1}{\omega}(D - \omega L),\ \omega \in C\!\!\!C \setminus \{0\}, \tag{3.39}$$

όπου η παράμετρος ω καλείται παράμετρος της υπερχαλάρωσης (ή overrelaxation παράμετρος ή SOR παράμετρος), οπότε είναι εύκολο να βρεθεί ότι η SOR μέθοδος είναι η ακόλουθη

$$x^{(k+1)} = \mathcal{L}_\omega x^{(k)} + c_\omega,\ k = 0, 1, 2, \cdots,$$

με $x^{(0)} \in \mathbb{C}^n$ οποιοδήποτε, όπου

$$\mathcal{L}_\omega = (D - \omega L)^{-1}[(1-\omega)D + \omega U],\ c_\omega = \omega(D - \omega L)^{-1}b. \tag{3.40}$$

Οπως είναι φανερό από την (3.39), για $\omega = 1$ η SOR μέθοδος δίνει αυτήν των Gauss-Seidel. Θα πρέπει όμως να τονιστεί ότι η SOR είναι διαφορετική από την extrapolated Gauss-Seidel μέθοδο και δε θα πρέπει να συγχέεται με την τελευταία. Αυτό μπορεί να διαπιστωθεί αμέσως αν θεωρήσουμε τις αντίστοιχες εκφράσεις των δύο ρυθμιστών, έστω με διαφορετικούς συμβολισμούς των αντίστοιχων παραμέτρων, και τους εξισώσουμε. Π.χ., αν $\frac{1}{\omega_1}(D - \omega_1 L)$ και $\frac{1}{\omega_2}(D - L)$ είναι οι ρυθμιστές των SOR και extrapolated Gauss-Seidel μεθόδων, αντίστοιχα, η εξίσωσή τους δίνει αμέσως ότι $\omega_1 = \omega_2$ και $(\omega_2 - 1)L = 0$. Οπότε προκύπτει $\omega_1 = \omega_2 = 1$, δηλαδή η περίπτωση της Gauss-Seidel μεθόδου, ή $\omega_1 = \omega_2$ και $L = 0$, δηλαδή ίδια παράμετρος και στις δύο μεθόδους στην τετριμμένη περίπτωση πίνακα A που είναι άνω τριγωνικός.

Η αναλυτική εύρεση της οποιασδήποτε συνιστώσας του διανύσματος $x^{(k+1)}$, μπορεί να επιτευχτεί, με εργασία ανάλογη αυτών στις προηγούμενες κλασικές μεθόδους, ότι δίνεται από την έκφραση

$$x_i^{(k+1)} = (1-\omega)x_i^{(k)} + \omega\left(b_i - \sum_{j=1}^{i-1} a_{ij}x_j^{(k+1)} - \sum_{j=i+1}^{n} a_{ij}x_j^{(k)}\right)/a_{ii},\ i = 1(1)n. \tag{3.41}$$

Μπορούμε να παρατηρήσουμε ότι η οποιαδήποτε συνιστώσα της νέας επανάληψης δίνεται σα βαρυκεντρικός μέσος όρος της ίδιας συνιστώσας της προηγούμενης επανάληψης και της συνιστώσας που θα βρίσκαμε αν εφαρμόζαμε για την εύρεση της αντίστοιχης συνιστώσας τη μέθοδο των Gauss-Seidel.

Για τη σύγκλιση της SOR μεθόδου είναι δυνατόν να βρεθεί μία αναγκαία συνθήκη η οποία είναι ανεξάρτητη των ιδιοτήτων του πίνακα A. Αυτή δίνεται στο παρακάτω θεώρημα.

Θεώρημα 3.5 *(*Kahan*) Αναγκαία συνθήκη για τη σύγκλιση της* SOR *μεθόδου είναι η*

$$|\omega - 1| < 1,\ \ \omega \in \mathbb{C}\ \ \acute{\eta}\ \ \omega \in (0,2),\ \ \omega \in \mathbb{R}.$$

Απόδειξη: Αν $\lambda_i \in \sigma(\mathcal{L}_\omega),\ i = 1(1)n$, είναι οι n ιδιοτιμές του επαναληπτικού πίνακα της SOR θα ισχύει με βάση γνωστή σχέση από τη Γραμμική Αλγεβρα ότι $\prod_{i=1}^n \lambda_i = \det(\mathcal{L}_\omega)$. Αντικαθιστώντας την έκφραση της $\mathcal{L}_\omega$ από την (3.40) μπορούμε να βρούμε αμέσως ότι

$$\begin{aligned}&\textstyle\prod_{i=1}^n \lambda_i = \det\left((D-\omega L)^{-1}[(1-\omega)D + \omega U]\right) = \det((D-\omega L)^{-1})\det\left((1-\omega)D + \omega U\right)\\ &= \tfrac{1}{\det(D)}(1-\omega)^n \det(D) = (1-\omega)^n.\end{aligned} \tag{3.42}$$

Επειδή είναι $|\prod_{i=1}^n \lambda_i| = \prod_{i=1}^n |\lambda_i|$ και για τη σύγκλιση της SOR μεθόδου πρέπει να ισχύει $|\lambda_i| < 1,\ i = 1(1)n$, προκύπτει αμέσως ότι η $\prod_{i=1}^n |\lambda_i| < 1$ αποτελεί μια αναγκαία συνθήκη για τη

σύγκλιση. Χρησιμοποιώντας την (3.42) στην αναγκαία συνθήκη, που μόλις βρέθηκε, παίρνουμε το πρώτο συμπέρασμα του θεωρήματος. Το δεύτερο συμπέρασμα έπεται αμέσως από το πρώτο αν $\omega \in I\!R$. □

Υπάρχουν πίνακες A για τους οποίους η αναγκαία συνθήκη, ιδιαίτερα η $\omega \in (0,2)$, για $\omega \in I\!R$, είναι συγχρόνως και ικανή. Μια τέτοια κατηγορία πινάκων είναι αυτή των Ερμιτιανών και θετικά ορισμένων. Συγκεκριμένα έχουμε το παρακάτω θεώρημα.

Θεώρημα 3.6 *(*Reich-Ostrowski-Varga*) Εστω ότι ο $A \in C\!\!\!C^{n,n}$ είναι Ερμιτιανός και θετικά ορισμένος πίνακας. Τότε για $\omega \in I\!R$, η συνθήκη $\omega \in (0,2)$ είναι αναγκαία και ικανή για τη σύγκλιση της* SOR *μεθόδου.*

Απόδειξη: Το γεγονός ότι η συνθήκη είναι αναγκαία προκύπτει από το προηγούμενο Θεώρημα του Kahan. Για την απόδειξη ότι είναι και ικανή προχωρούμε ως εξής. Θεωρούμε τη διάσπαση (3.12) και έστω ότι $\lambda \in \sigma(\mathcal{L}_\omega)$ με $x \in C\!\!\!C^n \setminus \{0\}$ το αντίστοιχο ιδιοδιάνυσμα. Από την (3.40) θα έχουμε ότι $(D-\omega L)^{-1}[(1-\omega)D+\omega U]x = \lambda x$ ή ισοδύναμα

$$[(1-\omega)D+\omega U]x = \lambda(D-\omega L)x. \tag{3.43}$$

Από την υπόθεση $A^H = A$ προκύπτει ότι $D^H = D \in I\!R^{n,n}$ και μάλιστα $a_{ii} > 0,\ i = 1(1)n$, αφού ο A είναι θετικά ορισμένος. Ακόμη είναι $L^H = U$ και φυσικά $U^H = L$. Με βάση την (3.43) σχηματίζουμε τα Ευκλείδεια εσωτερικά γινόμενα

$$\left(x, [(1-\omega)D+\omega L^H]x\right)_2 = \left(x, \lambda(D-\omega L)x\right)_2. \tag{3.44}$$

Εκτελώντας πράξεις στην (3.44), θέτοντας $a = (x,Ax)_2\ (>0)$, $d = (x,Dx)_2\ (= \sum_{i=1}^n a_{ii}|x_i|^2 > 0)$, $l = (x,Lx)_2$ και $\bar{l} = \overline{(x,Lx)_2} = (x,L^Hx)_2$, έχουμε ότι

$$(1-\omega)d + \omega\bar{l} = \lambda(d-\omega l). \tag{3.45}$$

Θεωρώντας τα τετράγωνα των μέτρων των μελών της (3.45) μπορούμε να πάρουμε

$$(1-\omega)^2d^2 + (1-\omega)\omega d(l+\bar{l}) + \omega^2|l|^2 \ = \ |\lambda|^2\left(d^2 - \omega d(l+\bar{l}) + \omega^2|l|^2\right). \tag{3.46}$$

Παρατηρώντας ότι $l+\bar{l} = \left(x,(L+L^H)x\right)_2 = (x,(D-A)x)_2 = d-a$ η (3.46) γράφεται

$$(1-\omega)^2d^2 + (1-\omega)\omega d(d-a) + \omega^2|l|^2 \ = \ |\lambda|^2\left(d^2 - \omega d(d-a) + \omega^2|l|^2\right)$$

ή ισοδύναμα

$$(1-\omega)d^2 - (1-\omega)\omega da + \omega^2|l|^2 \ = \ |\lambda|^2\left((1-\omega)d^2 + \omega da + \omega^2|l|^2\right). \tag{3.47}$$

Ο δεύτερος παράγοντας του δεύτερου μέλους της (3.47) είναι η έκφραση $|(x,(D-\omega L)x)_2|^2 = |d-\omega l|^2$ δοσμένη αναλυτικά. Αυτή είναι πάντα θετική. Γιατί αν $d-\omega l = 0$, τότε και $(1-\omega)d+\omega\bar{l} = 0$

από την (3.45), οπότε αφαιρώντας τη δεύτερη ισότητα από την πρώτη μπορεί να προκύψει ότι $\omega a = 0$, πράγμα που είναι άτοπο αφού $\omega \neq 0$ και $a > 0$. Λύνοντας τελικά ως προς $|\lambda|^2$, μπορούμε να πάρουμε

$$|\lambda|^2 \;=\; 1 - \frac{\omega(2-\omega)da}{(1-\omega)d^2 + \omega da + \omega^2|l|^2} \;<\; 1. \tag{3.48}$$

Η δεξιά ανισότητα ισχύει από την υπόθεση ότι $\omega \in (0,2)$, πράγμα που αποδείχνει την πρότασή μας. □

Σημείωση: Είναι αξιοσημείωτο ότι η (3.48) πέρα από το ικανό της δοθείσας συνθήκης αποδείχνει συγχρόνως και το αναγκαίο.

Πόρισμα 3.3 *(*Reich*) Κάτω από τις προϋποθέσεις του θεωρήματος η μέθοδος των* Gauss-Seidel *συγκλίνει.*

Πίνακες για τους οποίους προκύπτουν παραπέρα ιδιότητες για την SOR μέθοδο ανήκουν σε πολλές κατηγορίες. Μία από αυτές, η οποία μας επιτρέπει κάτω από ορισμένες προϋποθέσεις να βρίσκουμε την ταχύτερα συγκλίνουσα SOR μέθοδο, είναι και αυτή που δίνουμε και εξετάζουμε στη συνέχεια.

Ορισμός 3.3 *Ενας μή διαγώνιος πίνακας $A \in \mathbb{C}^{n,n}$, με $a_{ii} \neq 0,\ i = 1(1)n$, καλείται δικυκλικός ανν υπάρχει μεταθετικός πίνακας P τ.ω.*

$$PAP^T \;=\; \begin{bmatrix} D_1 & B \\ C & D_2 \end{bmatrix}, \tag{3.49}$$

όπου D_1 και D_2 διαγώνιοι πίνακες, όχι απαραίτητα ίδιων διαστάσεων.

Σημείωση: Η ιδιότητα της ορισθείσας δικυκλικότητας είναι γνωστή και ως "ιδιότητα $\mathcal{A}$ του Young".

Πίνακες δικυκλικοί που απαντιούνται στην πράξη είναι, μεταξύ άλλων, αυτοί που είναι ήδη στη μορφή (3.49) καθώς και οι τριδιαγώνιοι με μή μηδενικά διαγώνια στοιχεία. Το γεγονός ότι οι τελευταίοι πίνακες μπορούν να τεθούν στη μορφή (3.49) αποδείχνεται εύκολα αρκεί να θεωρήσουμε ως P τον πίνακα με $P^T = [e^1\ e^3\ e^5\ \cdots\ e^2\ e^4\ \cdots]$.

Ορισμός 3.4 *Εστω $A \in \mathbb{C}^{n,n}$ δικυκλικός. Θεωρώντας τη διάσπαση (3.12), ο A θα καλείται συνεπώς διατεταγμένος ανν*

$$\sigma\left(D^{-1}(\alpha L + \frac{1}{\alpha}U)\right) \;\equiv\; \sigma\left(D^{-1}(L+U)\right),\quad \forall\ \alpha \in \mathbb{C} \setminus \{0\}. \tag{3.50}$$

Σημείωση: Η σχέση (3.50), που αναφέρεται στο δοθέντα ορισμό, μπορεί να διατυπωθεί και με άλλους τρόπους. Π.χ. "Το φάσμα των ιδιοτιμών του πίνακα $D^{-1}(\alpha L + \frac{1}{\alpha}U)$ ταυτίζεται με το φάσμα των ιδιοτιμών του επαναληπτικού πίνακα του Jacobi, που αντιστοιχεί στον A" ή, ακόμη, "Οι ιδιοτιμές του πίνακα $D^{-1}(\alpha L + \frac{1}{\alpha}U)$ είναι ανεξάρτητες του α".

Μεταξύ των πινάκων που ικανοποιούν τον ορισμό και απαντιούνται συχνά στην πράξη είναι οι τριδιαγώνιοι με μή μηδενικά διαγώνια στοιχεία καθώς και οι πίνακες που έχουν τη μορφή του δεύτερου μέλους της (3.49). Η απόδειξη, ότι οι πίνακες που αναφέρθηκαν, και που είναι ήδη δικυκλικοί, είναι και συνεπώς διατεταγμένοι, στηρίζεται στον ορισμό και έχει ως εξής στην περίπτωση των τριδιαγώνιων πινάκων. Εστω ότι

$$A \;=\; \begin{bmatrix} a_{11} & a_{12} & & & \\ a_{21} & a_{22} & a_{23} & & \\ & \ddots & \ddots & \ddots & \\ & & a_{n-1,n-2} & a_{n-1,n-1} & a_{n-1,n} \\ & & & a_{n,n-1} & a_{nn} \end{bmatrix}$$

είναι ο τριδιαγώνιος πίνακας. Θεωρούμε το διαγώνιο πίνακα

$$E \;=\; \mathrm{diag}\left(1, \frac{1}{\alpha}, \frac{1}{\alpha^2}, \cdots, \frac{1}{\alpha^{n-1}}\right)$$

και σχηματίζουμε τον όμοιο προς τον A πίνακα EAE^{-1}. Για τον πίνακα A, που πρέπει να θεωρήσουμε στον ορισμό (3.50), έχουμε

$$\sigma\left(D^{-1}(\alpha L + \frac{1}{\alpha}U)\right) \;\equiv\; \sigma\left(ED^{-1}(\alpha L + \frac{1}{\alpha}U)E^{-1}\right). \tag{3.51}$$

Ο πίνακας του δεξιά φάσματος στην παραπάνω ισοδυναμία, έστω B, είναι τριδιαγώνιος με στοιχεία της i-οστής γραμμής τα εξής:

$$\begin{aligned} b_{i,i-1} &= \tfrac{1}{\alpha^{i-1}}\left(\alpha(-\tfrac{a_{i,i-1}}{a_{ii}})\right)\alpha^{i-2} \;=\; -\tfrac{a_{i,i-1}}{a_{ii}}, \\ b_{ii} &= 0, \\ b_{i,i+1} &= \tfrac{1}{\alpha^{i-1}}\left(\tfrac{1}{\alpha}(-\tfrac{a_{i,i+1}}{a_{ii}})\right)\alpha^{i} \;=\; -\tfrac{a_{i,i+1}}{a_{ii}}. \end{aligned}$$

Τα στοιχεία όμως που μόλις βρέθηκαν δεν είναι παρά τα στοιχεία του επαναληπτικού πίνακα του Jacobi που αντιστοιχεί στον πίνακα A, όπως μπορεί αμέσως να επαληθευτεί. Αρα $\sigma\left(D^{-1}(\alpha L + \frac{1}{\alpha}U)\right)$ $\equiv \sigma\left(D^{-1}(L+U)\right)$ και ο A είναι δικυκλικός και συνεπώς διατεταγμένος.

Η αντίστοιχη απόδειξη για τον πίνακα (3.49) είναι απλούστερη και παραλείπεται.

Με βάση τον ορισμό του δικυκλικού και συνεπώς διατεταγμένου πίνακα είμαστε σε θέση να αποδείξουμε την εξής απλή πρόταση.

Λήμμα 3.5 *Εστω ότι ο πίνακας $A \in \mathbb{C}^{n,n}$ είναι δικυκλικός και συνεπώς διατεταγμένος. Τότε οι διάφορες από το μηδέν ιδιοτιμές του επαναληπτικού πίνακα του* Jacobi, *που αντιστοιχεί στον A, εμφανίζονται σε αντίθετα ζεύγη.*

Απόδειξη: Εστω $\mu \in \sigma\left(D^{-1}(L+U)\right) \setminus \{0\}$. Από την ιδιότητα του A έχουμε αμέσως τις παρακάτω ισοδυναμίες

$$\begin{array}{l} \mu \in \sigma\left(D^{-1}(L+U)\right) \setminus \{0\} \iff \mu \in \sigma\left(D^{-1}(\alpha L + \frac{1}{\alpha}U)\right) \setminus \{0\},\ \forall\ \alpha \in \mathbb{C} \setminus \{0\}, \\ \iff \mu \in \sigma\left(D^{-1}(-L + \frac{1}{-1}U)\right) \setminus \{0\} \iff -\mu \in \sigma\left(D^{-1}(L+U)\right) \setminus \{0\}. \end{array} \tag{3.52}$$

Η τελευταία σχέση στις (3.52) αποδείχνει την πρόταση. □

Σημείωση: Μπορεί να αποδειχτεί ότι οι μή μηδενικές ιδιοτιμές του επαναληπτικού πίνακα του Jacobi, που αντιστοιχεί σε ένα δικυκλικό και συνεπώς διατεταγμένο πίνακα, εμφανίζονται σε αντίθετα ζεύγη της **ίδιας** πολλαπλότητας. (Σημείωση: Για το τελευταίο ο αναγνώστης παραπέμπεται στο βιβλίο του Varga [43].)

Στη συνέχεια διατυπώνεται και αποδείχνεται μια πρόταση που είναι ίσως η σπουδαιότερη από αυτές που αφορούν τους δικυκλικούς και συνεπώς διατεταγμένους πίνακες.

Θεώρημα 3.7 *Εστω ότι ο $A \in \mathbb{C}^{n,n}$ είναι δικυκλικός και συνεπώς διατεταγμένος. Αν $\mu \in \sigma\left(D^{-1}(L+U)\right)$ και $\lambda \neq 0$ ικανοποιεί τη σχέση*

$$(\lambda + \omega - 1)^2 \ = \ \omega^2 \mu^2 \lambda \tag{3.53}$$

τότε $\lambda \in \sigma(\mathcal{L}_\omega)$. Αντίστροφα, αν $\lambda \in \sigma(\mathcal{L}_\omega) \setminus \{0\}$ και μ ικανοποιεί την (3.53) τότε $\mu \in \sigma\left(D^{-1}(L+U)\right)$.

Απόδειξη: Θεωρούμε την έκφραση $\det(\mathcal{L}_\omega - \lambda I)$ και αντικαθιστούμε σ΄ αυτήν την έκφραση για τον επαναληπτικό πίνακα της SOR από την (3.40), οπότε μπορούμε να λάβουμε διαδοχικά εφαρμόζοντας απλές ιδιότητες οριζουσών

$$\begin{array}{l} \det(\mathcal{L}_\omega - \lambda I) \ = \ \det\left((D-\omega L)^{-1}[(1-\omega)D + \omega U] - \lambda I\right) \\ = \ \frac{1}{\det(D)}(-1)^n \det\left((\lambda+\omega-1)D - \omega\lambda L - \omega U\right) \\ = \ (-\omega\lambda^{\frac{1}{2}})^n \det\left(\frac{\lambda+\omega-1}{\omega\lambda^{\frac{1}{2}}} I - \left(D^{-1}(\lambda^{\frac{1}{2}}L + \frac{1}{\lambda^{\frac{1}{2}}}U)\right)\right). \end{array}$$

Αν τις παραπάνω ίσες οριζουσιακές εκφράσεις εξισώσουμε με το μηδέν τότε από την πρώτη και την τελευταία έκφραση έχουμε αμέσως την ισοδυναμία

$$\lambda \in \sigma\left(\mathcal{L}_\omega\right) \setminus \{0\} \iff \mu \equiv \frac{\lambda+\omega-1}{\omega\lambda^{\frac{1}{2}}} \in \sigma\left(D^{-1}(\lambda^{\frac{1}{2}}L + \frac{1}{\lambda^{\frac{1}{2}}}U)\right) \ \left(\equiv \ \sigma\left(D^{-1}(L+U)\right)\right), \tag{3.54}$$

όπου η τελευταία ισοδυναμία ισχύει λόγω της δικυκλικής και συνεπώς διατεταγμένης ιδιότητας του πίνακα A. Από την ισότητα $\mu \equiv \frac{\lambda+\omega-1}{\omega\lambda^{\frac{1}{2}}}$, αν υψώσουμε στο τετράγωνο, τότε με βάση και το Λήμμα 3.5, προκύπτει η σχέση (3.53), οπότε οι ισοδυναμίες στην (3.54) αποδείχνουν συγχρόνως και τα δύο σκέλη του θεωρήματος. Σημειώνεται πως αν $\lambda = 0 \in \sigma(\mathcal{L}_\omega)$ και η (3.53) ικανοποιείται, τότε θα ικανοποιείται για κάθε μ, άρα και για κάθε $\mu \in \sigma\left(D^{-1}(L+U)\right)$, και επιπλέον θα είναι $\omega = 1$. □

Πόρισμα 3.4 *Εστω ότι ο $A \in \mathbb{C}^{n,n}$ είναι δικυκλικός και συνεπώς διατεταγμένος και έστω ότι T_J και T_{GS} είναι οι επαναληπτικοί πίνακες των μεθόδων* Jacobi *και* Gauss-Seidel, *αντίστοιχα. Αν $\mu \in \sigma(T_J)$ τότε $\lambda = \mu^2 \in \sigma(T_{GS})$ και αν $\lambda \in \sigma(T_{GS}) \setminus \{0\}$ με $\lambda = \mu^2$ τότε $\mu \in \sigma(T_J)$. Τέλος, ισχύει ότι $\rho(T_{GS}) = \rho^2(T_J)$, δηλαδή αν μία από τις δύο μεθόδους* Jacobi *και* Gauss-Seidel *συγκλίνει, τότε θα συγκλίνει και η άλλη και μάλιστα η* Gauss-Seidel *θα συγκλίνει δυο φορές (ασυμπτωτικά) ταχύτερα από την* Jacobi.

Για δικυκλικούς και συνεπώς διατεταγμένους πίνακες $A \in \mathbb{C}^{n,n}$ είναι δυνατόν, σε ορισμένες περιπτώσεις, να βρεθεί η βέλτιστη SOR μέθοδος. Δηλαδή, να βρεθεί η τιμή της παραμέτρου ω για την οποία η αντίστοιχη SOR μέθοδος συγκλίνει (ασυμπτωτικά) ταχύτερα. Για το σκοπό αυτό απαιτείται γνώση είτε του φάσματος των ιδιοτιμών του επαναληπτικού πίνακα του Jacobi είτε μόνο μέρος του φάσματος και κάποιων ιδιοτήτων του. Η απλούστερη περίπτωση παρατίθεται στη συνέχεια με τη μορφή θεωρήματος.

Θεώρημα 3.8 *Εστω ότι ο πίνακας $A \in \mathbb{C}^{n,n}$ είναι δικυκλικός και συνεπώς διατεταγμένος. Εστω ακόμη ότι $\sigma\left(D^{-1}(L+U)\right) \subset (-1,1)$ και ότι $\beta := \rho\left(D^{-1}(L+U)\right)$ (< 1). Η βέλτιστη τιμή ω_β $(\in \mathbb{R})$ της* SOR *παραμέτρου δίνεται από τον τύπο*

$$\omega_\beta = \frac{2}{1+\sqrt{1-\beta^2}}. \tag{3.55}$$

Για την αντίστοιχη φασματική ακτίνα ισχύει ότι

$$\rho(\mathcal{L}_\omega) \;>\; \rho(\mathcal{L}_{\omega_\beta}) \;=\; \omega_\beta - 1, \quad \forall\, \omega \neq \omega_\beta, \tag{3.56}$$

και η μέθοδος συγκλίνει $\forall\, \omega \in (0,2)$.

Απόδειξη: Θεωρούμε για τυχαίο αλλά συγκεκριμένο $\mu \in \sigma\left(D^{-1}(L+U)\right) \setminus \{0\}$, (τότε και $-\mu \in \sigma\left(D^{-1}(L+U)\right) \setminus \{0\}$ και άρα $|\mu| \in \sigma\left(D^{-1}(L+U)\right) \setminus \{0\}$) την εξίσωση (3.53) την οποία γράφουμε αναλυτικά ως προς λ. Δηλαδή,

$$\lambda^2 - (\mu^2\omega^2 - 2\omega + 2)\lambda + (\omega-1)^2 = 0. \tag{3.57}$$

Η διακρίνουσα της δευτεροβάθμιας εξίσωσης (3.57) είναι η

$$\Delta = \mu^2\omega^2(\mu^2\omega^2 - 4\omega + 4),$$

με ρίζες ως προς ω τις 0, $\frac{2}{1+\sqrt{1-\mu^2}}$, $\frac{2}{1-\sqrt{1-\mu^2}}$. Επειδή μας ενδιαφέρουν τιμές του $\omega \in (0,2)$, που αποτελεί την αναγκαία συνθήκη για τη σύγκλιση της SOR, και η μόνη από τις τρεις ρίζες στο διάστημα αυτό είναι η $\frac{2}{1+\sqrt{1-\mu^2}}$ έχουμε αμέσως ότι $\Delta > 0$, $= 0$, < 0, αντίστοιχα με το αν $\omega \in (0, \frac{2}{1+\sqrt{1-\mu^2}})$, $\omega = \frac{2}{1+\sqrt{1-\mu^2}}$, $\omega \in (\frac{2}{1+\sqrt{1-\mu^2}}, 2)$. Στις δυο τελευταίες περιπτώσεις έχουμε $\Delta \leq 0$ για $\omega \in [\frac{2}{1+\sqrt{1-\mu^2}}, 2)$ και επειδή τότε οι δυο ρίζες της (3.57) έχουν το ίδιο μέτρο θα είναι $|\lambda| = \omega - 1$. Αφού μας ενδιαφέρει να έχουμε το μικρότερο δυνατόν μέτρο για το λ αυτό θα συμβαίνει για τη μικρότερη τιμή του ω, δηλαδή για $\omega = \frac{2}{1+\sqrt{1-\mu^2}}$. Για τις τιμές $\omega \in (0, \frac{2}{1+\sqrt{1-\mu^2}})$ είναι $\Delta > 0$, η (3.57) έχει ρίζες πραγματικές μή αρνητικές (γιατί το γινόμενό τους $(\omega-1)^2$ είναι μή αρνητικός αριθμός και το άθροισμά τους $\mu^2\omega^2 - 2\omega + 2 = \frac{1}{2}\mu^2\omega^2 + \frac{1}{2}(\mu^2\omega^2 - 4\omega + 4) = \frac{1}{2}\mu^2\omega^2 + \frac{1}{2}\Delta$ είναι θετικός) και άνισες με μεγαλύτερη τη

$$\lambda \;=\; \frac{1}{2}\left(\mu^2\omega^2 - 2\omega + 2 + |\mu|\omega\sqrt{\mu^2\omega^2 - 4\omega + 4}\right) \;=\; \frac{1}{4}\left(|\mu|\omega + \sqrt{\mu^2\omega^2 - 4\omega + 4}\right)^2.$$

Η παράγωγος αυτής ως προς ω είναι ίση με

$$\frac{d\lambda}{d\omega} \;=\; \frac{1}{2\sqrt{\mu^2\omega^2 - 4\omega + 4}}\left(|\mu|\omega + \sqrt{\mu^2\omega^2 - 4\omega + 4}\right)\left(|\mu|\sqrt{\mu^2\omega^2 - 4\omega + 4} - (2 - \mu^2\omega)\right). \tag{3.58}$$

Παρατηρούμε ότι όλοι οι παράγοντες του γινομένου στην (3.58), εκτός ίσως του τελευταίου, είναι θετικοί. Για τον τελευταίο δεξιά παράγοντα έχουμε ότι ο πρώτος όρος είναι θετικός, όπως είναι και ο δεύτερος. Το τελευταίο ισχύει διότι $2 - \mu^2\omega > 2(1-\mu^2) > 0$. Εξάλλου έχουμε ότι

$$\begin{aligned}\operatorname{sign}\left(|\mu|\sqrt{\mu^2\omega^2 - 4\omega + 4} - (2 - \mu^2\omega)\right) &= \operatorname{sign}\left(\mu^2(\mu^2\omega^2 - 4\omega + 4) - (2 - \mu^2\omega)^2\right) \\ &= \operatorname{sign}(\mu^2 - 1) = -1\end{aligned}$$

και άρα ο τελευταίος δεξιά παράγοντας της (3.58) είναι αρνητικός. Αυτό συνεπάγεται ότι η συνάρτηση $|\lambda| = \lambda$, που μας ενδιαφέρει, είναι αυστηρά φθίνουσα συνάρτηση του $\omega \in (0, \frac{2}{1+\sqrt{1-\mu^2}}]$. Αρα παίρνει την ελάχιστη δυνατή τιμή της για $\omega = \omega_{|\mu|} = \frac{2}{1+\sqrt{1-\mu^2}}$. Αυτή είναι η ίδια τιμή για την οποία η $|\lambda|$ ήταν ελάχιστη στο διάστημα $[\frac{2}{1+\sqrt{1-\mu^2}}, 2)$. Το μέχρι στιγμής συμπέρασμα είναι ότι για ένα σταθερό $|\mu| \neq 0$ η ελάχιστη τιμή της φασματικής ακτίνας (δηλαδή της μή μικρότερης από τα δύο $|\lambda|$) της SOR μεθόδου στο διάστημα $(0,2)$ είναι η $\rho(\mathcal{L}_{\omega_{|\mu|}}) = \omega_{|\mu|} - 1$. Επιπλέον παρατηρούμε τα εξής: α) Η δεύτερη παράγωγος της φασματικής ακτίνας $|\lambda| = \lambda$ στο διάστημα $(0, \omega_{|\mu|}]$ μπορεί να

βρεθεί ότι δίνεται από την έκφραση

$$\frac{d^2\lambda}{d\omega^2} = \frac{\Big(|\mu|(\mu^2\omega^2-4\omega+4)^{\frac{1}{2}}-(2-\mu^2\omega)\Big)\Big[|\mu|\big((\mu^2\omega^2-4\omega+4)+2(2-\omega)\big)+\mu^2\omega(\mu^2\omega^2-4\omega+4)^{\frac{1}{2}}\Big]}{2(\mu^2\omega^2-4\omega+4)^{\frac{3}{2}}},$$

που, όπως είναι φανερό από τη μέχρι τώρα ανάλυση, είναι αρνητική. Αρα το γράφημά της στρέφει τα κοίλα προς τα κάτω κι ακόμη

$$\lim_{\omega\to\omega_{|\mu|}^-}\frac{d\rho(\mathcal{L}_\omega)}{d\omega}=-\infty.$$

β) Για $\omega\in[\omega_{|\mu|},2)$ η φασματική ακτίνα (δηλαδή η $|\lambda|$) $\rho(\mathcal{L}_\omega)=\omega-1$ αυξάνεται γραμμικά στο διάστημα αυτό, η δε κλίση του γραφήματός της είναι ίση με 1.
Ακόμη, για $\mu=0$ έχουμε $\lambda=1-\omega$ και το γράφημα της $|\lambda|$ στο $[0,2]$ είναι ένα ευθύγραμμο τμήμα $|\lambda|=1-\omega$ στο $[0,1]$ και ένα άλλο $|\lambda|=\omega-1$ στο $[1,2]$ με κλίσεις -1 και 1, αντίστοιχα.
Από τη μέχρι τώρα ανάλυση και από την παρατήρηση ότι για οποιαδήποτε δύο μ_1 και μ_2, με $|\mu_1|<|\mu_2|$, έχουμε

$$\rho(\mathcal{L}_{\omega(\mu_1)})\ \begin{cases} < & \rho(\mathcal{L}_{\omega(\mu_2)}) \quad \gamma\iota\alpha\ \omega\in(0,\omega_{|\mu_2|}),\\ = & \rho(\mathcal{L}_{\omega(\mu_2)}) \quad \gamma\iota\alpha\ \omega\in[\omega_{|\mu_2|},2).\end{cases}$$

Από την όλη μέχρι τώρα μελέτη προκύπτει ότι για κάθε $\omega\in(0,2)$ η φασματική ακτίνα της αντίστοιχης SOR μεθόδου θα δίνεται από το μεγαλύτερο δυνατό $|\lambda|$ που αντιστοιχεί, προφανώς, στο μεγαλύτερο δυνατό $|\mu|=\beta\in\sigma\left(D^{-1}(L+U)\right)$. Επομένως οδηγούμαστε στο συμπέρασμα ότι για να βρούμε την "καλύτερη" (βέλτιστη) φασματική ακτίνα της SOR μεθόδου θα πρέπει να επιλέξουμε εκείνη από τις παραπάνω για την οποία το αντίστοιχο ω την καθιστά ελάχιστη. Αυτό όμως μπορεί να επιτευχτεί, όπως αποδείχτηκε, για $\omega=\omega_\beta=\frac{2}{1+\sqrt{1-\beta^2}}$, πράγμα που δίνει τις εκφράσεις και τις σχέσεις στις (3.55) και (3.56).

Η λεπτομερής ανάλυση που έγινε στην όλη απόδειξη αποδείχνει άμεσα και το γεγονός ότι υπό τις προϋποθέσεις του θεωρήματος η SOR μέθοδος συγκλίνει για κάθε $\omega\in(0,2)$. □

ΑΣΚΗΣΕΙΣ

1.: Αν $1\in\sigma(\mathcal{L}_\omega)$, όπου $\mathcal{L}_\omega$ ο επαναληπτικός πίνακας της SOR μεθόδου, τότε ο πίνακας A με τον οποίο συνδέεται ο $\mathcal{L}_\omega$ **δεν** είναι αντιστρέψιμος.

2.: Να αποδειχτεί ο τύπος (3.41) της SOR μεθόδου.

3.: Για την επίλυση του γραμμικού συστήματος $Ax = b$, με $A \in \mathbb{C}^{n,n}$, $\det(A) \neq 0$, $a_{ii} \neq 0$, $i = 1(1)n$, $A = D - L - U$, όπου D, L, U οι γνωστοί πίνακες, $b \in \mathbb{C}^n$, προτείνεται το επαναληπτικό σχήμα, με $\omega \in \mathbb{C}\backslash\{0\}$,

$$x^{(k+1)} = \mathcal{L}_{r,\omega} x^{(k)} + c_{r,\omega},$$
$$\mathcal{L}_{r,\omega} := (D - rL)^{-1}[(1-\omega)D + (\omega - r)L + \omega U],\ c_{r,\omega} := \omega(D - rL)^{-1}b$$

α) Να δειχτεί ότι το προτεινόμενο επαναληπτικό σχήμα επιλύει το δοθέν σύστημα, εφόσον συγκλίνει.
β) Για ποια ζεύγη τιμών (r, ω) είναι δυνατόν να ληφτούν οι μέθοδοι: i) Jacobi, ii) Gauss-Seidel, iii) SOR, iv) Extrapolated Jacobi, v) Extrapolated Gauss-Seidel και vi) Extrapolated SOR; και
γ) Ποιες είναι σε κάθε περίπτωση, εφόσον υπάρχουν, οι τιμές των παραμέτρων παρεκβολής και ποιες αυτές των παραμέτρων υπερχαλάρωσης;

4.: Να αποδειχτεί, με οποιοδήποτε τρόπο, ότι η μέθοδος SOR, που αντιστοιχεί στον πίνακα $A = \begin{bmatrix} 4 & -1 & -1 & 0 \\ -1 & 4 & 0 & -1 \\ -1 & 0 & 4 & -1 \\ 0 & -1 & -1 & 4 \end{bmatrix}$ συγκλίνει $\forall\ \omega \in (0, 2)$.

5.: Να γίνει (βήμα προς βήμα) η αντίστοιχη απόδειξη του Θεωρήματος των Reich-Ostrowski-Varga για τη μέθοδο Gauss-Seidel. Δηλαδή, αν $A \in \mathbb{C}^{n,n}$ είναι Ερμιτιανός και θετικά ορισμένος η αντίστοιχη μέθοδος των Gauss-Seidel συγκλίνει.

6.: Στο γραμμικό σύστημα $Ax = b$ ο πίνακας των συντελεστών των αγνώστων είναι $A = \text{trid}(1, 2, 1) \in \mathbb{R}^{4,4}$. α) Να βρεθούν οι ιδιοτιμές του επαναληπτικού πίνακα Jacobi που συνδέεται με τον A.
β) Να δικαιολογηθεί, με οποιοδήποτε τρόπο, ότι η επαναληπτική μέθοδος SOR του συστήματος που δόθηκε συγκλίνει για κάθε $\omega \in (0, 2)$. και
γ) Να δικαιολογηθεί γιατί μπορεί να βρεθεί η βέλτιστη SOR μέθοδος και στη συνέχεια να βρεθούν τόσο η βέλτιστη SOR παράμετρος όσο και βέλτιστη φασματική ακτίνα του επαναληπτικού πίνακα της SOR.

7.: Δίνεται το γραμμικό σύστημα

$$\begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}.$$

α) Χρησιμοποιώντας τους αντίστοιχους ορισμούς να δειχτεί ότι ο συντελεστής πίνακας A του συστήματος είναι δικυκλικός και συνεπώς διατεταγμένος.

β) Χωρίς να βρεθούν οι ιδιοτιμές του επαναληπτικού πίνακα του Jacobi, που αντιστοιχεί στον A, να δειχτεί ότι η μέθοδος Jacobi συγκλίνει. και
γ) Να εκτελεστεί μια επανάληψη της SOR μεθόδου για την επίλυση του συστήματος, που δόθηκε, χρησιμοποιώντας ακριβή αριθμητική, με $\omega = 1.25$ και $x^{(0)} = [1\ 1\ 1]^T$.

8.: Δίνεται ο πίνακας

$$A = \begin{bmatrix} I_{n_1} & B \\ C & I_{n_2} \end{bmatrix}, \quad B \in \mathbb{C}^{n_1,n_2},\ C \in \mathbb{C}^{n_2,n_1}.$$

Να αποδειχτεί ότι **όλες** οι ιδιοτιμές $\lambda \in \sigma(A)\backslash\{1\}$ μπορούν να χωριστούν σε ζεύγη με άθροισμα καθενός ίσο με 2.

9.: Δίνεται το γραμμικό σύστημα $Ax = b$ με $A = \begin{bmatrix} 2 & 1 & 0 & 0 \\ 1 & 4 & 1 & 0 \\ 0 & 1 & 4 & 1 \\ 0 & 0 & 1 & 2 \end{bmatrix}$.

α) Να εξεταστούν ως προς τη σύγκλιση οι μέθοδοι Jacobi και Gauss-Seidel .
β) Να βρεθεί η βέλτιστη τιμή της παραμέτρου $\omega \in \mathbb{R}$ για την SOR μέθοδο. και
γ) Να γίνει σύγκριση των τριών μεθόδων ως προς την ταχύτητα σύγκλισης.
(**Περιορισμός**: Να γίνουν ακριβείς πράξεις διατηρώντας ριζικά και κλάσματα στους υπολογισμούς.)

10.: Δίνεται το γραμμικό σύστημα

$$\begin{array}{rrrcr} x_1 & -x_2 & & = & 2 \\ -x_1 & +2x_2 & -x_3 & = & -4 \\ & -x_2 & +4x_3 & = & 5 \end{array}$$

α) Να εξεταστούν ως προς τη σύγκλιση οι μέθοδοι Jacobi και Gauss-Seidel και να βρεθεί ο λόγος των μέσων ασυμπτωτικών ταχυτήτων σύγκλισής τους. και
β) Να γίνουν δυο επαναλήψεις με τη μέθοδο που συγκλίνει ταχύτερα και δυο επαναλήψεις με την SOR μέθοδο με $\omega = 1.25$.

11.: Δίνεται το γραμμικό σύστημα $Ax = b$ με $A = \begin{bmatrix} 2 & 0 & -1 & 0 \\ 0 & 2 & 0 & -1 \\ -1 & 0 & 2 & 0 \\ 0 & -1 & 0 & 2 \end{bmatrix}$ και $b = [1\ 1\ 1\ 1]^T$.

α) Να εξεταστούν ως προς τη σύγκλιση και να συγκριθούν μεταξύ τους οι μέθοδοι Jacobi, Gauss-Seidel και βέλτιστης SOR με $\omega \in \mathbb{R}$. και
β) Να γίνουν δυο επαναλήψεις της SOR με $\omega = 1.1$ και αρχικό διάνυσμα $x^{(0)} = 0$, διατηρώντας τρία δεκαδικά ψηφία στους υπολογισμούς.

12: Αν ο πίνακας $A \in \mathbb{C}^{n,n}$ είναι αυστηρά διαγώνια υπέρτερος κατά γραμμές (κατά στήλες) να αποδειχτεί ότι η SOR μέθοδος, που είναι συνδεδεμένη μ΄ αυτόν, συγκλίνει για κάθε $\omega \in (0, 1]$.

3.5 Συμμετρική SOR (SSOR) Επαναληπτική Μέθοδος

Η Συμμετρική (Symmetric) Μέθοδος της Διαδοχικής Υπερχαλάρωσης (SSOR) είναι μία επαναληπτική μέθοδος στην οποία κάθε επανάληψη αποτελείται από δύο ημι-επαναλήψεις. Συγκεκριμένα η πρώτη ημι-επανάληψη είναι μία συνήθης SOR μέθοδος ενώ η δεύτερη είναι μία SOR, όπου οι ρόλοι των πινάκων L και U έχουν εναλλαχτεί. Συγκεκριμένα,

$$\begin{array}{rcl} x^{(k+\frac{1}{2})} & = & (D-\omega L)^{-1}[(1-\omega)D+\omega U]x^{(k)}+\omega(D-\omega L)^{-1}b, \\ x^{(k+1)} & = & (D-\omega U)^{-1}[(1-\omega)D+\omega L]x^{(k+\frac{1}{2})}+\omega(D-\omega U)^{-1}b. \end{array} \tag{3.59}$$

Για να γραφτεί η (3.59) σα μία κλασική επαναληπτική μέθοδος θα πρέπει η έκφραση για το $x^{(k+\frac{1}{2})}$ από την πρώτη ημι-επανάληψη να αντικατασταθεί στη δεύτερη, να εκτελεστούν όλες οι πράξεις και τέλος να γραφτεί αυτή στη μορφή

$$x^{(k+1)} = \mathcal{S}_\omega x^{(k)} + c_\omega, \ k = 0, 1, 2, \cdots, \tag{3.60}$$

με $x^{(0)} \in \mathbb{C}^n$ οποιοδήποτε. Μετά τη διαδικασία που μόλις υποδείχτηκε μπορούμε να καταλήξουμε από τις (3.59) στην (3.60), όπου

$$\begin{array}{rcl} \mathcal{S}_\omega & = & (D-\omega U)^{-1}[(1-\omega)D+\omega L](D-\omega L)^{-1}[(1-\omega)D+\omega U], \\ c_\omega & = & \omega(2-\omega)(D-\omega U)^{-1}D(D-\omega L)^{-1}b. \end{array} \tag{3.61}$$

Σημείωση: Για $\omega = 1$, η SSOR μέθοδος είναι γνωστή ως μέθοδος του Aitken. Αποτελείται τότε από δυό ημι-επαναλήψεις η πρώτη από τις οποίες είναι η Gauss-Seidel μέθοδος και η δεύτερη μία Gauss-Seidel με εναλλαγμένους τους ρόλους των L και U.

Για την SSOR μέθοδο ισχύουν αρκετές προτάσεις που είναι ανάλογες αντίστοιχων της SOR μεδόδου. Μερικές από αυτές δίνονται και αποδείχνονται στη συνέχεια.

Θεώρημα 3.9 *Εστω $A \in \mathbb{C}^{n,n}$. Για $\omega \in \mathbb{C}$ αναγκαία συνθήκη για τη σύγκλιση της* SSOR *μεθόδου είναι η $|\omega - 1| < 1$. Για $\omega \in \mathbb{R}$ η αναγκαία συνθήκη γίνεται $\omega \in (0, 2)$.*

Απόδειξη: Αρχίζουμε την απόδειξη, όπως και στην περίπτωση του Θεωρήματος του Kahan για την SOR μέθοδο, θεωρώντας το γινόμενο των ιδιοτιμών $\lambda_i,\ i = 1(1)n$, του επαναληπτικού πίνακα SSOR από τις (3.61). Από την πρώτη των (3.61) θα είναι

$$\begin{array}{l} \prod_{i=1}^n \lambda_i \ = \ \det(S_\omega) \ = \ \det\left((D-\omega U)^{-1}[(1-\omega)D+\omega L](D-\omega L)^{-1}[(1-\omega)D+\omega U]\right) \\ = \ \det\left((D-\omega U)^{-1}\right)\det\left([(1-\omega)D+\omega L]\right)\det\left((D-\omega L)^{-1}\right)\det\left([(1-\omega)D+\omega U]\right) \\ = \ \frac{1}{\det(D)}(1-\omega)^n\det(D)\frac{1}{\det(D)}(1-\omega)^n\det(D) \ = \ (1-\omega)^{2n}. \end{array} \tag{3.62}$$

Από την πρώτη και την τελευταία έκφραση των παραπάνω ισοτήτων έχουμε ότι $|\prod_{i=1}^{n}\lambda_i| = \prod_{i=1}^{n}|\lambda_i| = |1-\omega|^{2n}$. Οπως γίνεται αμέσως φανερό για να έχουμε σύγκλιση της SSOR μεθόδου θα πρέπει $|\lambda_i| < 1,\ i = 1(1)n$, και άρα από τις προηγούμενες ισότητες παίρνουμε αμέσως τις αναγκαίες συνθήκες, που δίνονται στη διατύπωση της παρούσας πρότασης. □

Λόγω της συμμετρικής μορφής που έχει ο επαναληπτικός πίνακας $\mathcal{S}_\omega$ της SSOR μεθόδου πολλές φορές μπορούν να εξαχτούν συμπεράσματα για τη μέθοδο αυτή πιο γενικά από τα αντίστοιχα της SOR μεθόδου. Οπως θα δούμε και στη συνέχεια είναι σκοπιμότερο αντί του επαναληπτικού πίνακα $\mathcal{S}_\omega$ να χρησιμοποιούνται κάποιοι άλλοι πίνακες που έχουν τις ίδιες ιδιοτιμές μ΄ αυτόν. Ενας από αυτούς είναι ο όμοιος προς τον $\mathcal{S}_\omega$,

$$\mathcal{S}'_\omega = (D-\omega U)\mathcal{S}_\omega(D-\omega U)^{-1} = [(1-\omega)D+\omega L](D-\omega L)^{-1}[(1-\omega)D+\omega U](D-\omega U)^{-1}. \quad (3.63)$$

Ο νέος πίνακας $\mathcal{S}'_\omega$ μπορεί να γραφτεί με διάφορους τρόπους χρησιμοποιώντας διαδοχικούς μετασχηματισμούς, όπως φαίνεται παρακάτω. Μερικές από τις ισοδύναμες εκφράσεις του θα χρησιμοποιηθούν στη συνέχεια.

$$\begin{aligned}
\mathcal{S}'_\omega &= [(1-\omega)D+\omega L](D-\omega L)^{-1}[(1-\omega)D+\omega U](D-\omega U)^{-1}\\
&= [(2-\omega)D-(D-\omega L)](D-\omega L)^{-1}[(2-\omega)D-(D-\omega U)](D-\omega U)^{-1}\\
&= [(2-\omega)D(D-\omega L)^{-1}-I][(2-\omega)D(D-\omega U)^{-1}-I]\\
&= I-(2-\omega)D[(D-\omega L)^{-1}+(D-\omega U)^{-1}]+(2-\omega)^2D(D-\omega L)^{-1}D(D-\omega U)^{-1}\\
&= I-(2-\omega)D(D-\omega L)^{-1}[(D-\omega U)+(D-\omega L)-(2-\omega)D](D-\omega U)^{-1}\\
&= I-\omega(2-\omega)D(D-\omega L)^{-1}A(D-\omega U)^{-1}.
\end{aligned} \quad (3.64)$$

Μια πρόταση, που είναι αντίστοιχη αλλά κάπως γενικότερη αυτής των Reich-Ostrowski-Varga της SOR, διατυπώνεται και αποδείχνεται στη συνέχεια αφού πρώτα διατυπωθεί και αποδειχτεί τυπικά ένα λήμμα που αναφέρεται στο γνωστό πηλίκο του Rayleigh. Σημειώνεται ότι μέρος της απόδειξης του λήμματος έχει ήδη δοθεί στη Βασική Θεωρία και το ίδιο λήμμα έχει δοθεί ως Ασκηση σε προηγούμενο Κεφάλαιο.

Λήμμα 3.6 *Εστω ότι ο Ερμιτιανός πίνακας $A \in \mathcal{C}^{n,n}$ έχει τις πραγματικές ιδιοτιμές του $\lambda_i,\ i = 1(1)n$, διατεταγμένες έτσι ώστε $\lambda_1 \le \lambda_2 \le \cdots \le \lambda_n$. Τότε για κάθε $x \in \mathcal{C}^n \setminus \{0\}$ ισχύει*

$$\lambda_1 \le \frac{(x,Ax)_2}{(x,x)_2} \le \lambda_n. \quad (3.65)$$

Απόδειξη: Εστω $x^i,\ i = 1(1)n$, τα αντίστοιχα ιδιοδιανύσματα του A, τα οποία μπορούν να ληφτούν έτσι ώστε να αποτελούν μια ορθοκανονική βάση. Τότε θα είναι $x = \sum_{i=1}^{n}\alpha_i x^i$, όπου $\alpha_i \in \mathcal{C},\ i = 1(1)n$. Αντικαθιστώντας στην έκφραση του πηλίκου που δόθηκε θα έχουμε

$$\begin{aligned}
\frac{(x,Ax)_2}{(x,x)_2} &= \frac{(\sum_{i=1}^{n}\overline{\alpha}_i x^{iH})A(\sum_{i=1}^{n}\alpha_i x^i)}{||x||_2^2} = \frac{(\sum_{i=1}^{n}\overline{\alpha}_i x^{iH})(\sum_{i=1}^{n}\alpha_i\lambda_i x^i)}{||x||_2^2}\\
&= \frac{\sum_{i=1}^{n}\lambda_i|\alpha_i|^2}{||x||_2^2} \le \lambda_n\frac{\sum_{i=1}^{n}|\alpha_i|^2}{||x||_2^2} = \lambda_n\frac{||x||_2^2}{||x||_2^2} = \lambda_n.
\end{aligned} \quad (3.66)$$

Με παρόμοιο τρόπο αποδείχνεται και η αριστερή ανισότητα στην (3.65). □

Σημείωση: Η δεξιά ισότητα στις (3.65) πετυχαίνεται για $x = x^n$ ενώ η αριστερή για $x = x^1$.

Θεώρημα 3.10 *Εστω $A \in \mathbb{C}^{n,n}$ Ερμιτιανός με θετικά διαγώνια στοιχεία. Τότε για οποιοδήποτε $\omega \in (0,2)$ ο πίνακας $\mathcal{S}_\omega$ έχει πραγματικές μή αρνητικές ιδιοτιμές. Επιπλέον, αν ο A είναι θετικά ορισμένος τότε η* SSOR *μέθοδος συγκλίνει. Αντίστροφα, αν η* SSOR *μέθοδος συγκλίνει και $\omega \in \mathbb{R}$, τότε $\omega \in (0,2)$, που είναι προφανές, και ο A είναι θετικά ορισμένος.*

Απόδειξη: Από την υπόθεση ότι ο A είναι Ερμιτιανός θα ισχύει $U = L^H$ ($L = U^H$). Αν $D^{\frac{1}{2}} = \text{diag}\left(a_{11}^{\frac{1}{2}}, a_{22}^{\frac{1}{2}}, \cdots, a_{nn}^{\frac{1}{2}}\right)$, όπου $\text{diag}(.,.,\cdots,.)$ συμβολίζει διαγώνιο πίνακα με αντίστοιχα διαγώνια στοιχεία τα εντός των παρενθέσεων, τότε, από τις εκφράσεις (3.64), παρατηρούμε ότι ο πίνακας

$$\mathcal{S}'_\omega \;=\; [(2-\omega)D(D-\omega L)^{-1} - I][(2-\omega)D(D-\omega U)^{-1} - I]$$

έχει τις ίδιες ιδιοτιμές με τον πίνακα

$$\mathcal{S}'_\omega{}' \;=\; D^{-\frac{1}{2}}\mathcal{S}'_\omega D^{\frac{1}{2}} \;=\; D^{\frac{1}{2}}[(2-\omega)(D-\omega L)^{-1} - D^{-1}]D^{\frac{1}{2}}D^{\frac{1}{2}}[(2-\omega)(D-\omega L^H)^{-1} - D^{-1}]D^{\frac{1}{2}}.$$

Η έκφραση όμως που βρέθηκε, $\forall\, x \in \mathbb{C}^n \setminus \{0\}$ δίνει ότι για τον Ερμιτιανό πίνακα $\mathcal{S}'_\omega{}'$ το εσωτερικό γινόμενο $(x, \mathcal{S}'_\omega{}' x)_2$ είναι διαδοχικά ίσο με τις εκφράσεις

$$\begin{aligned}\Big(D^{\frac{1}{2}}[(2-\omega)(D-\omega L^H)^{-1} - D^{-1}]D^{\frac{1}{2}}x,\, D^{\frac{1}{2}}[(2-\omega)(D-\omega L^H)^{-1} - D^{-1}]D^{\frac{1}{2}}x\Big)_2 \\ =\; \big|\big|\; [(2-\omega)D^{\frac{1}{2}}(D-\omega L^H)^{-1}D^{\frac{1}{2}} - I]x\;\big|\big|_2^2 \;\geq\; 0.\end{aligned} \tag{3.67}$$

Από την παραπάνω τελευταία σχέση έπεται ότι ο $\mathcal{S}'_\omega{}'$ είναι μή αρνητικά ορισμένος και επομένως θα έχει, όπως και ο όμοιός του $\mathcal{S}_\omega$, πραγματικές μή αρνητικές ιδιοτιμές.
Εστω ότι ο A είναι θετικά ορισμένος. Εχοντας υπόψη τις (3.64), η έκφραση $\mathcal{S}'_\omega{}' = D^{-\frac{1}{2}}\mathcal{S}'_\omega D^{\frac{1}{2}}$ γράφεται και ως εξής

$$\mathcal{S}'_\omega{}' = I - \omega(2-\omega)D^{\frac{1}{2}}(D-\omega L)^{-1}A(D-\omega L^H)^{-1}D^{\frac{1}{2}}.$$

Ομως, $\omega(2-\omega) > 0$ και

$$\forall\; x \in \mathbb{C}^n \setminus \{0\} \iff \forall\; y = (D-\omega L^H)^{-1}D^{\frac{1}{2}}x \;\in \mathbb{C}^n \setminus \{0\}$$

ισχύει ότι $y^H A y > 0$, αφού ο A είναι θετικά ορισμένος, άρα ο πίνακας $\omega(2-\omega)D^{\frac{1}{2}}(D-\omega L)^{-1}A(D-\omega L^H)^{-1}D^{\frac{1}{2}}$ θα είναι κι αυτός θετικά ορισμένος και επομένως θα έχει θετικές ιδιοτιμές. Το τελευταίο συμπέρασμα συνεπάγεται ότι ο $\mathcal{S}'_\omega{}'$, άρα και ο $\mathcal{S}_\omega$, θα έχει πραγματικές ιδιοτιμές αυστηρά μικρότερες από τη μονάδα. Επειδή, όπως αποδείχτηκε, έχει και πραγματικές μή αρνητικές ιδιοτιμές τα δύο αυτά

συμπεράσματα οδηγούν στο τελικό συμπέρασμα ότι η SSOR συγκλίνει.
Αντίστροφα, αν η SSOR συγκλίνει τότε ασφαλώς $\omega \in (0,2)$ από την αναγκαία συνθήκη. Αν ο Ερμιτιανός πίνακας A δεν είναι θετικά ορισμένος τότε θα έχει μία τουλάχιστον ιδιοτιμή $\lambda \leq 0$. Εστω x το αντίστοιχο ιδιοδιάνυσμα, οπότε $Ax = \lambda x,\ x \in \mathcal{C}^n \setminus \{0\}$. Ομως, από την τελευταία δεξιά έκφραση στις (3.64), για τον Ερμιτιανό πίνακα $\mathcal{S}_\omega' {}' = D^{-\frac{1}{2}} \mathcal{S}_\omega' D^{\frac{1}{2}}$, έχουμε αν $y = D^{-\frac{1}{2}}(D - \omega L^H)x$ ($\in \mathcal{C}^n \setminus \{0\}$)

$$\frac{(y, \mathcal{S}_\omega' {}' y)_2}{(y,y)_2} \ = \ 1 - \frac{\lambda\omega(2-\omega)(x,x)_2}{(y,y)_2} \ \geq \ 1.$$

Αλλά η τελευταία ανισότητα οδηγεί σε άτοπο διότι από το πηλίκο του Rayleigh (Λήμμα 3.6) γνωρίζουμε ότι για τον Ερμιτιανό πίνακα $\mathcal{S}_\omega' {}'$ ο λόγος $\frac{(y,\mathcal{S}_\omega' {}' y)_2}{(y,y)_2}$ βρίσκεται μεταξύ της μικρότερης και μεγαλύτερης ιδιοτιμής του. Αρα η μεγαλύτερη ιδιοτιμή του θα ήταν τότε μεγαλύτερη ή ίση από τη μονάδα και επομένως η SSOR δε θα συνέκλινε. Συνεπώς ο πίνακας A είναι θετικά ορισμένος. □

Κλείνουμε το παρόν κεφάλαιο με την περίπτωση όπου ο $A \in \mathcal{C}^{n,n}$ συμβαίνει να είναι δικυκλικός και συνεπώς διατεταγμένος της ειδικής μορφής

$$A \ = \ \begin{bmatrix} D_1 & B \\ C & D_2 \end{bmatrix}, \tag{3.68}$$

όπου οι $D_1 \in \mathcal{C}^{n_1,n_1},\ D_2 \in \mathcal{C}^{n_2,n_2}$, με $n_1 + n_2 = n,\ \det(D_1)\det(D_2) \neq 0$, είναι διαγώνιοι. Προφανώς στη συγκεκριμένη περίπτωση έχουμε

$$D \ = \ \begin{bmatrix} D_1 & 0 \\ 0 & D_2 \end{bmatrix}, \ L \ = \ \begin{bmatrix} 0 & 0 \\ -C & 0 \end{bmatrix}, \ U \ = \ \begin{bmatrix} 0 & -B \\ 0 & 0 \end{bmatrix}. \tag{3.69}$$

Ο επαναληπτικός πίνακας $\mathcal{S}_\omega$ της SSOR μεθόδου θα δίνεται και πάλι από την έκφραση στην (3.61). Χρησιμοποιώντας τον όμοιό του $\mathcal{S}_\omega'$ της (3.63) και μετά τον όμοιο προς τον $\mathcal{S}_\omega'$, $\widetilde{\mathcal{S}}_\omega' = D^{-1}\mathcal{S}_\omega' D$ μπορούμε να μετασχηματίσουμε τον τελευταίο και ως εξής:

$$\widetilde{\mathcal{S}}_\omega' \ = \ [(1-\omega)I + \omega\widetilde{L}](I - \omega\widetilde{L})^{-1}[(1-\omega)I + \omega\widetilde{U}](I - \omega\widetilde{U})^{-1},$$

όπου

$$\widetilde{L} \ = \ \begin{bmatrix} 0 & 0 \\ -D_2^{-1}C & 0 \end{bmatrix}, \ \widetilde{U} \ = \ \begin{bmatrix} 0 & -D_1^{-1}B \\ 0 & 0 \end{bmatrix}.$$

Θέτοντας $\overline{L} = -D_2^{-1}C$ και $\overline{U} = -D_1^{-1}B$, στις παραπάνω εκφράσεις για τον $\widetilde{\mathcal{S}}_\omega'$, και εκτελώντας όλες τις δυνατές πράξεις καταλήγουμε σε μια νέα έκφραση για τον $\widetilde{\mathcal{S}}_\omega'$, που είναι η ακόλουθη

$$\widetilde{\mathcal{S}}_\omega' \ = \ \begin{bmatrix} (1-\omega)^2 I_{n_1} & (1-\omega)\omega(2-\omega)\overline{U} \\ (1-\omega)\omega(2-\omega)\overline{L} & (1-\omega)^2 I_{n_2} + \omega^2(2-\omega)^2\overline{LU} \end{bmatrix}, \tag{3.70}$$

όπου θα πρέπει να σημειωθεί πως ενδιάμεσα χρησιμοποιήθηκε το γεγονός ότι $(I - \omega\widetilde{L})^{-1} = I + \omega\widetilde{L}$ και $(I - \omega\widetilde{U})^{-1} = I + \omega\widetilde{U}$, αφού $\widetilde{L}^2 = 0$ και $\widetilde{U}^2 = 0$.

Η σχέση (3.70) μπορεί να γραφτεί ισοδύναμα και ως εξής:

$$\frac{1}{\omega(2-\omega)}(\widetilde{\mathcal{S}}'_\omega - (\omega-1)^2 I) \;=\; \begin{bmatrix} 0 & (1-\omega)\overline{U} \\ (1-\omega)\overline{L} & \omega(2-\omega)\overline{LU} \end{bmatrix}. \tag{3.71}$$

Με βάση την (3.71) προσπαθούμε τώρα να βρούμε σχέση που να συνδέει τις ιδιοτιμές λ του πίνακα $\mathcal{S}_\omega$, ή του ομοίου του $\widetilde{\mathcal{S}}'_\omega$, με τις ιδιοτιμές μ του επαναληπτικού πίνακα του Jacobi του πίνακα A.

Εστω ότι μ είναι μία μή μηδενική ιδιοτιμή του επαναληπτικού πίνακα του Jacobi και $x = [x_1^T \; x_2^T]^T \in \mathbb{C}^n \setminus \{0\}$, $x_1 \in \mathbb{C}^{n_1}$, $x_2 \in \mathbb{C}^{n_2}$, το αντίστοιχο ιδιοδιάνυσμα. Θα ισχύει

$$\begin{bmatrix} 0 & \overline{U} \\ \overline{L} & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \;=\; \mu \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

ή ισοδύναμα

$$\overline{U}x_2 \;=\; \mu x_1 \;\; \kappa\alpha\iota \;\; \overline{L}x_1 \;=\; \mu x_2, \tag{3.72}$$

από τις οποίες προκύπτει

$$\overline{LU}x_2 = \mu^2 x_2. \tag{3.73}$$

Είναι όμως, $x_2 \neq 0$. Γιατί αν $x_2 = 0$, τότε επειδή $\mu \neq 0$ από τήν πρώτη των (3.72) θα προέκυπτε $x_1 = 0$. Αυτό όμως αντίκειται στην υπόθεσή μας ότι το x είναι ιδιοδιάνυσμα του επαναληπτικού πίνακα του Jacobi. Αρα, από την (3.73) προκύπτει ότι το x_2 είναι ιδιοδιάνυσμα του πίνακα $\overline{LU}$ με αντίστοιχη ιδιοτιμή μ^2. Επανερχόμενοι στη σχέση (3.71) παρατηρούμε αμέσως πως αν λ είναι μια ιδιοτιμή του πίνακα $\mathcal{S}_\omega$ ή ισοδύναμα του $\widetilde{\mathcal{S}}'_\omega$ τότε μια ιδιοτιμή του πίνακα του πρώτου μέλους της (3.71) θα είναι η $\nu \;=\; \frac{\lambda-(\omega-1)^2}{\omega(2-\omega)}$. Η ν θα είναι και ιδιοτιμή του πίνακα του δεύτερου μέλους της (3.71). Εστω $y = [y_1^T \; y_2^T]^T \in \mathbb{C}^n \setminus \{0\}$, $y_1 \in \mathbb{C}^{n_1}$, $y_2 \in \mathbb{C}^{n_2}$, το αντίστοιχο ιδιοδιάνυσμα. Αντικαθιστώντας στη σχέση

$$\begin{bmatrix} 0 & (1-\omega)\overline{U} \\ (1-\omega)\overline{L} & \omega(2-\omega)\overline{LU} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \;=\; \nu \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$$

έχουμε ισοδύναμα ότι

$$(1-\omega)\overline{U}y_2 \;=\; \nu y_1 \;\; \kappa\alpha\iota \;\; (1-\omega)\overline{L}y_1 + \omega(2-\omega)\overline{LU}y_2 \;=\; \nu y_2, \tag{3.74}$$

από τις οποίες για $\nu \neq 0$ ($\Longleftrightarrow \;\lambda \neq (1-\omega)^2$) παίρνουμε $y_1 \;=\; \frac{(1-\omega)}{\nu}\overline{U}y_2$. Αντικαθιστώντας στη δεύτερη των (3.74) έχουμε τελικά ότι

$$\left[(1-\omega)^2 + \omega(2-\omega)\nu\right]\overline{LU}y_2 \;=\; \nu^2 y_2. \tag{3.75}$$

Η τελευταία εξίσωση, επειδή $y_2 \neq 0$, με το ίδιο όπως και πριν σκεπτικό για τα x_2 και x_1, δίνει ότι η έκφραση $\frac{\nu^2}{(1-\omega)^2+\omega(2-\omega)\nu}$ είναι μια μή μηδενική ιδιοτιμή του $\overline{LU}$ και άρα ίση με μ^2. Από την τελευταία παρατήρηση προκύπτει η ισότητα

$$\left(\lambda - (\omega-1)^2\right)^2 \;=\; \omega^2(2-\omega)^2\lambda\mu^2, \tag{3.76}$$

που είναι τελικά η σχέση που συνδέει τις ιδιοτιμές των επαναληπτικών πινάκων Jacobi και SSOR στην περίπτωση που εξετάσαμε. Συγκεκριμένα έχουμε:

Θεώρημα 3.11 *Εστω ότι ο* $A \in \mathbb{C}^{n,n}$ *είναι της μορφής (3.68). Αν* $\lambda \in \sigma(\mathcal{S}_\omega) \setminus \{(\omega-1)^2\}$ *και* $\mu\ (\neq 0)$ *ικανοποιεί την (3.76) τότε η* μ *είναι ιδιοτιμή του αντίστοιχου επαναληπτικού πίνακα του* Jacobi. *Αντίστροφα, αν* $\mu\ (\neq 0)$ *είναι ιδιοτιμή του επαναληπτικού πίνακα του* Jacobi, *που αντιστοιχεί στον* A, *και* $\lambda(\neq (\omega-1)^2)$ *ικανοποιεί την (3.76) τότε* $\lambda \in \sigma(\mathcal{S}_\omega)$.

Αν οι ιδιοτιμές του επαναληπτικού πίνακα του Jacobi, που αντιστοιχεί στον πίνακα A της προηγούμενης περίπτωσης που εξετάστηκε, είναι πραγματικές και η επαναληπτική μέθοδος του Jacobi συγκλίνει, όπως π.χ. στην περίπτωση όπου ο A είναι Ερμιτιανός και θετικά ορισμένος, τότε είναι δυνατόν να βρεθεί μια βέλτιστη τιμή του $\omega \in \mathbb{R}$, για την οποία η SSOR μέθοδος συγκλίνει ασυμπτωτικά με τη μεγαλύτερη δυνατή ταχύτητα. Σχετικά διατυπώνουμε και αποδείχνουμε το παρακάτω θεώρημα.

Θεώρημα 3.12 *Κάτω από τις προϋποθέσεις του προηγούμενου θεωρήματος και με τις επιπλέον υποθέσεις ότι* $\sigma\left(D^{-1}(L+U)\right) \subset \mathbb{R}$ *και* $\rho\left(D^{-1}(L+U)\right) < 1$ *η βέλτιστη* SSOR *μέθοδος για* $\omega \in \mathbb{R}$, *σ΄ ό,τι αφορά την ασυμπτωτική ταχύτητα σύγκλισής της, είναι αυτή πού αντιστοιχεί σε* $\omega = \omega_\beta$, *όπου*

$$\omega_\beta \;=\; 1, \;\; \rho(\mathcal{S}_\omega) \;>\; \rho(\mathcal{S}_{\omega_\beta}) \;=\; \rho(\mathcal{L}_1) \;=\; \rho^2\left(D^{-1}(L+U)\right), \;\; \forall\, \omega \neq \omega_\beta,$$

και η SSOR *θα συγκλίνει για κάθε* $\omega \in (0,2)$.

Απόδειξη: Αν θέσουμε $\widetilde{\omega} = \omega(2-\omega)$ στην (3.76), αυτή γράφεται ως εξής

$$(\lambda + \widetilde{\omega} - 1)^2 \;=\; \lambda \widetilde{\omega}^2 \mu^2. \tag{3.77}$$

Η παραπάνω εξίσωση όμως δεν είναι παρά η εξίσωση που συνδέει τις ιδιοτιμές λ μιας SOR μεθόδου, με SOR παράμετρο $\widetilde{\omega}$, με τις ιδιοτιμές μ του αντίστοιχου επαναληπτικού πίνακα του Jacobi στην περίπτωση ενός δικυκλικού και συνεπώς διατεταγμένου πίνακα A. Επειδή δε οι ιδιοτιμές του αντίστοιχου επαναληπτικού πίνακα του Jacobi είναι πραγματικές και η μέθοδος Jacobi συγκλίνει τότε θα ισχύει τόσο το αντίστοιχο θεώρημα που δίνει τη βέλτιστη SOR όσο και η ανάλυση που έγινε για να βρεθεί η βέλτιστη τιμή του ω της SOR. Στην παρούσα περίπτωση η αναγκαία συνθήκη $\omega \in (0,2)$ για τη σύγκλιση της SSOR δίνει $\widetilde{\omega} = \omega(2-\omega) \in (0,1]$. Επειδή από την ανάλυση της αντίστοιχης SOR περίπτωσης, βρέθηκε ότι η συνάρτηση $\rho(\mathcal{L}_{\widetilde{\omega}})$ είναι γνήσια φθίνουσα στο διάστημα $(0,1]$, έπεται ότι $\widetilde{\omega}_\beta = 1 \iff \omega_\beta = 1$. Για $\omega_\beta = 1$ η SSOR μέθοδος είναι η μέθοδος του Aitken που, όπως μόλις αποδείχτηκε, έχει βέλτιστη φασματική ακτίνα αυτήν του αντίστοιχου επαναληπτικού πίνακα των Gauss-Seidel ή το τετράγωνο της φασματικής ακτίνας του επαναληπτικού πίνακα του Jacobi. Το γεγονός ότι η SSOR μέθοδος συγκλίνει για κάθε $\omega \in (0,2)$ προκύπτει άμεσα από τα παραπάνω. □

3.6 Block **Επαναληπτικές Μέθοδοι**

Στο παρόν κεφάλαιο πραγματοποιείται η επέκταση των κλασικών επαναληπτικών μεθόδων που αναπτύχτηκαν στα τρία προηγούμενα κεφάλαια. Για το σκοπό αυτό θεωρούμε πάλι για επίλυση το γραμμικό σύστημα (3.1), όπου τώρα διαχωρίζουμε τον πίνακα A σε μιά $p \times p$ block μορφή. Βασική προϋπόθεση είναι ότι τα διαγώνια blocks (υποπίνακες) πρέπει να είναι τετραγωνικοί πίνακες. Είναι φανερό ότι αν $A_{ii},\ i = 1(1)p$, είναι οι διαγώνιοι υποπίνακες διαστάσεων $n_i \times n_i,\ i = 1(1)p$, θα πρέπει να ισχύει ότι $\sum_{i=1}^{p} n_i = n$. Φυσικά, η επιλογή $p = n$ μας οδηγεί τελικά στις κλασικές επαναληπτικές μεθόδους, όπως αυτές αναπτύχτηκαν, γι΄ αυτό και σε αντιδιαστολή με τις block μεθόδους, που θα αναπτυχτούν στη συνέχεια, οι προηγούμενες καλούνται point (σημειακές) επαναληπτικές μέθοδοι.

3.6.1 Block Jacobi **Επαναληπτική Μέθοδος**

Καταρχάς θεωρούμε το γραμμικό σύστημα (3.1), όπου ο πίνακας των συντελεστών των αγνώστων A είναι διαχωρισμένος σύμφωνα με τα εκτεθέντα στην προηγούμενη παράγραφο. Στη συνέχεια θεωρούμε μια διάσπαση της ίδιας μορφής με αυτήν της (3.12) με τη βασική διαφορά ότι ο πίνακας D δεν είναι ο γνωστός $D = \text{diag}(A)$ αλλά ο block διαγώνιος πίνακας $D = \text{diag}(A_{11}, A_{22}, \cdots, A_{pp})$. Οι πίνακες L και U ορίζονται πάλι ως αυστηρά κάτω τριγωνικός και αυστηρά άνω τριγωνικός, αντίστοιχα, έτσι ώστε η διάσπαση (3.12) να είναι μονοσήμαντα ορισμένη και να εξαρτιέται μόνο από το διαχωρισμό σε blocks του πίνακα A. Για να μπορεί να οριστεί καταρχάς η αντίστοιχη block μέθοδος του Jacobi θα πρέπει ο D να είναι αντιστρέψιμος. Είναι φανερό πως το τελευταίο συμβαίνει ανν οι block υποπίνακες $A_{ii},\ i = 1(1)p$, είναι αντιστρέψιμοι, οπότε θα ισχύει ότι $D^{-1} = \text{diag}(A_{11}^{-1}, A_{22}^{-1}, \cdots, A_{pp}^{-1})$. Το γεγονός ότι ο πίνακας D πληροί και τη δεύτερη προϋπόθεση του (3.2) είναι επίσης φανερό αφού ένα γραμμικό σύστημα με πίνακα συντελεστών αγνώστων D, στην πραγματικότητα p συστήματα με πίνακες συντελεστών αγνώστων $A_{ii},\ i = 1(1)p$, αντίστοιχα, λύνονται οικονομικότερα από ένα σύστημα με πίνακα συντελεστών αγνώστων A. Για την εφαρμογή της block Jacobi επαναληπτικής μεθόδου γίνεται και ένας αντίστοιχος block διαχωρισμός τόσο στο άγνωστο διάνυσμα x όσο και στο γνωστό διάνυσμα b. Συγκεκριμένα, κάθε ένα από τα δύο διανύσματα διαχωρίζεται σε p blocks συνιστωσών έτσι ώστε

$$x \ = \ [x_1^T \ x_2^T \ \cdots \ x_p^T]^T, \quad b \ = \ [b_1^T \ b_2^T \ \cdots \ b_p^T]^T \quad \mu\epsilon \quad x_i, b_i \in \mathbb{C}^{n_i},\ i = 1(1)p.$$

Η block Jacobi μέθοδος θα έχει την ίδια ακριβώς γενική μορφή, όπως η (3.13), με τη μόνη βασική διαφορά ότι οι πίνακες $D,\ L$ και U θα ορίζονται όπως αναφέρθηκε προηγουμένως. Η block Jacobi μέθοδος θα συγκλίνει ανν $\rho\left(D^{-1}(L+U)\right) < 1$. Η εύρεση των block συνιστωσών $x_i^{(k+1)},\ i = 1(1)p$, της νέας επανάληψης $x^{(k+1)}$ θα βρίσκονται από τις block συνιστώσες της προηγούμενης επανάληψης. Για να βρούμε την αντίστοιχη σχέση θα πρέπει να πολλαπλασιάσουμε και τα δύο μέλη της (3.13)

επί D, οπότε θα έχουμε

$$\begin{bmatrix} A_{11} & & & & & \\ & A_{22} & & & & \\ & & \ddots & & & \\ & & & A_{ii} & & \\ & & & & \ddots & \\ & & & & & A_{pp} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_i \\ \vdots \\ x_p \end{bmatrix}^{(k+1)}$$

$$= \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_i \\ \vdots \\ b_p \end{bmatrix} - \begin{bmatrix} 0_{n_1} & A_{12} & \cdots & A_{1i} & \cdots & A_{1p} \\ A_{21} & 0_{n_2} & \cdots & A_{2i} & \cdots & A_{2p} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ A_{i1} & A_{i2} & \cdots & 0_{n_i} & \cdots & A_{ip} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ A_{p1} & A_{p2} & \cdots & A_{pi} & \cdots & 0_{n_p} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_i \\ \vdots \\ x_p \end{bmatrix}^{(k)}, \quad k = 0, 1, 2, \cdots, \tag{3.78}$$

με $x^{(0)} \in \mathbb{C}^n$ οποιοδήποτε. Αν τώρα εξισώσουμε το i-οστό block διάνυσμα του πρώτου μέλους με το αντίστοιχο του δεύτερου θα έχουμε

$$A_{ii} x_i^{(k+1)} = b_i - \sum_{j=1,\ j\neq i}^{p} A_{ij} x_j^{(k)},\ i = 1(1)p,\ k = 0, 1, 2, \cdots,$$

ή

$$x_i^{(k+1)} = A_{ii}^{-1} \left(b_i - \sum_{j=1,\ j\neq i}^{p} A_{ij} x_j^{(k)} \right),\ i = 1(1)p,\ k = 0, 1, 2, \cdots.$$

Για την καλύτερη κατανόηση της block Jacobi μεθόδου και για τη σύγκρισή της, ομοιότητες και διαφορές, με την αντίστοιχη point Jacobi θα παραθέσουμε ένα απλό παράδειγμα που θα εξετάσουμε εξαντλητικά. Ετσι θα γίνουν κατανοητές και οι άλλες block επαναληπτικές μέθοδοι στις οποίες θα αναφερθούμε αμέσως μετά χωρίς όμως να επεκταθούμε.

ΠΑΡΑΔΕΙΓΜΑ: Το γραμμικό σύστημα $Ax = b$ που δίνεται στη συνέχεια

$$Ax \equiv A_1 x \equiv \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 3 \\ -1 \\ 0 \end{bmatrix} =: b$$

έχει λύση το διάνυσμα $x = [2\ 1\ 1]^T$. Θα θεωρήσουμε την point Jacobi μέθοδο, που αντιστοιχεί στο σύστημα που δόθηκε, καθώς και τις δύο block Jacobi μεθόδους, που αντιστοιχούν στους

παρακάτω δυο διαχωρισμούς του πίνακα των συντελεστών των αγνώστων A

$$A_2 = \left[\begin{array}{cc|c} 2 & -1 & 0 \\ -1 & 2 & -1 \\ \hline 0 & -1 & 1 \end{array}\right], \quad A_3 = \left[\begin{array}{c|cc} 2 & -1 & 0 \\ \hline -1 & 2 & -1 \\ 0 & -1 & 1 \end{array}\right].$$

α) Θα βρούμε τις φασματικές ακτίνες των τριών επαναληπτικών πινάκων (point και block) Jacobi που αντιστοιχούν στον $A = A_1$ καθώς και στους διαχωρισμούς των A_2 και A_3. β) Θα εφαρμόσουμε από μία επανάληψη των τριών επαναληπτικών μεθόδων Jacobi με αρχικό διάνυσμα $x^{(0)} = [1\ 1\ 1]^T$.
ΛΥΣΗ: α) Εύρεση φασματικών ακτίνων των επαναληπτικών μεθόδων του Jacobi:
1)

$$T_{J_1} = D_1^{-1}(L_1 + U_1) = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$
$$= \begin{bmatrix} \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 0 & \frac{1}{2} & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 1 & 0 \end{bmatrix}.$$

Οι ιδιοτιμές θα βρίσκονται ως οι ρίζες της εξίσωσης $\det(T_{J_1} - \lambda I) = 0$. Επομένως

$$\det\left(\begin{bmatrix} -\lambda & \frac{1}{2} & 0 \\ \frac{1}{2} & -\lambda & \frac{1}{2} \\ 0 & 1 & -\lambda \end{bmatrix}\right) = -\lambda^3 + \frac{3}{4}\lambda = 0,$$

από την οποία προκύπτει ότι $\lambda_1 = \frac{\sqrt{3}}{2}$, $\lambda_2 = -\frac{\sqrt{3}}{2}$, και $\lambda_3 = 0$, οπότε $\rho(T_{J_1}) = \frac{\sqrt{3}}{2}$.
2) Για το διαχωρισμό του A_2 έχουμε

$$T_{J_2} = D_2^{-1}(L_2 + U_2) = \left[\begin{array}{cc|c} 2 & -1 & 0 \\ -1 & 2 & 0 \\ \hline 0 & 0 & 1 \end{array}\right]^{-1} \left[\begin{array}{cc|c} 0 & 0 & 0 \\ 0 & 0 & 1 \\ \hline 0 & 1 & 0 \end{array}\right]$$
$$= \left[\begin{array}{cc|c} \frac{2}{3} & \frac{1}{3} & 0 \\ \frac{1}{3} & \frac{2}{3} & 0 \\ \hline 0 & 0 & 1 \end{array}\right] \left[\begin{array}{cc|c} 0 & 0 & 0 \\ 0 & 0 & 1 \\ \hline 0 & 1 & 0 \end{array}\right] = \begin{bmatrix} 0 & 0 & \frac{1}{3} \\ 0 & 0 & \frac{2}{3} \\ 0 & 1 & 0 \end{bmatrix}. \tag{3.79}$$

Οι ιδιοτιμές του T_{J_2} βρίσκονται από την $\det(T_{J_2} - \lambda I) = 0$ και άρα

$$\det\left(\begin{bmatrix} -\lambda & 0 & \frac{1}{3} \\ 0 & -\lambda & \frac{2}{3} \\ 0 & 1 & -\lambda \end{bmatrix}\right) = -\lambda^3 + \frac{2}{3}\lambda = 0. \tag{3.80}$$

Από την (3.80) έχουμε ότι $\lambda_1 = \frac{\sqrt{6}}{3}$, $\lambda_2 = -\frac{\sqrt{6}}{3}$, και $\lambda_3 = 0$, οπότε $\rho(T_{J_2}) = \frac{\sqrt{6}}{3}$.
3) Τέλος για το δεύτερο block διαχωρισμό A_2 έχουμε

$$T_{J_3} \;=\; D_3^{-1}(L_3+U_3) \;=\; \left[\begin{array}{c|cc} 2 & 0 & 0 \\ \hline 0 & 2 & -1 \\ 0 & -1 & 1 \end{array}\right]^{-1} \left[\begin{array}{c|cc} 0 & 1 & 0 \\ \hline 1 & 0 & 0 \\ 0 & 0 & 0 \end{array}\right]$$
$$= \left[\begin{array}{c|cc} \frac{1}{2} & 0 & 0 \\ \hline 0 & 1 & 1 \\ 0 & 1 & 2 \end{array}\right] \left[\begin{array}{c|cc} 0 & 1 & 0 \\ \hline 1 & 0 & 0 \\ 0 & 0 & 0 \end{array}\right] \;=\; \left[\begin{array}{ccc} 0 & \frac{1}{2} & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{array}\right].$$

Οπότε οι ιδιοτιμές του T_{J_3} θα είναι οι ρίζες της εξίσωσης $\det(T_{J_3}-\lambda I)=0$ από την οποία παίρνουμε ότι

$$\det\left(\left[\begin{array}{ccc} -\lambda & \frac{1}{2} & 0 \\ 1 & -\lambda & 0 \\ 1 & 0 & -\lambda \end{array}\right]\right) \;=\; -\lambda^3+\frac{1}{2}\lambda \;=\; 0.$$

Από την τελευταία εξίσωση προκύπτει τελικά ότι $\lambda_1=\frac{\sqrt{2}}{2}$, $\lambda_2=-\frac{\sqrt{2}}{2}$ και $\lambda_3=0$, οπότε $\rho(T_{J_3}) \;=\; \frac{\sqrt{2}}{2}$.
Από τη σύγκριση των τριών φασματικών ακτίνων έχουμε αμέσως ότι

$$\rho(T_{J_3}) \;<\; \rho(T_{J_2}) \;<\; \rho(T_{J_1})$$

και επομένως ο block διαχωρισμός στον $A=A_3$ δίνει επαναληπτική μέθοδο Jacobi (ασυμπτωτικά) ταχύτερη κι από τις τρεις ενώ η κλασική point Jacobi αντιστοιχεί σε (ασυμπτωτικά) βραδύτερη επαναληπτική μέθοδο.
β) Εύρεση της πρώτης επανάληψης $x^{(1)}$ των τριών επαναληπτικών μεθόδων Jacobi:
Για την εύρεση της $x^{(1)}$ θα εφαρμοστούν οι σχέσεις που δίνονται στην (3.78) για $k=0$. Ετσι θα πάρουμε για κάθε μία από τις τρεις μεθόδους τα παρακάτω.
1)

$$\left[\begin{array}{ccc} 2 & & \\ & 2 & \\ & & 1 \end{array}\right] \left[\begin{array}{c} x_1 \\ x_2 \\ x_3 \end{array}\right]^{(1)} \;=\; \left[\begin{array}{c} 3 \\ -1 \\ 0 \end{array}\right] \;-\; \left[\begin{array}{ccc} 0 & -1 & 0 \\ -1 & 0 & -1 \\ 0 & -1 & 0 \end{array}\right] \left[\begin{array}{c} 1 \\ 1 \\ 1 \end{array}\right]$$

Από την οποία παίρνουμε

$$2x_1^{(1)}=4,\; 2x_2^{(1)}=1,\; x_3^{(1)}=1$$

και επομένως

$$x^{(1)}=[2\;0.5\;1]^T.$$

2)

$$\left[\begin{array}{cc|c} 2 & -1 & 0 \\ -1 & 2 & 0 \\ \hline 0 & 0 & 1 \end{array}\right] \left[\begin{array}{c} x_1 \\ x_2 \\ \hline x_3 \end{array}\right]^{(1)} \;=\; \left[\begin{array}{c} 3 \\ -1 \\ \hline 0 \end{array}\right] \;-\; \left[\begin{array}{cc|c} 0 & 0 & 0 \\ 0 & 0 & -1 \\ \hline 0 & -1 & 0 \end{array}\right] \left[\begin{array}{c} 1 \\ 1 \\ \hline 1 \end{array}\right]$$

από την οποία παίρνουμε

$$\begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}^{(1)} = \begin{bmatrix} 3 \\ 0 \end{bmatrix}, \quad x_3^{(1)} = 1$$

και επομένως

$$x^{(1)} = [2\ 1\ 1]^T.$$

3)

$$\left[\begin{array}{c|cc} 2 & 0 & 0 \\ \hline 0 & 2 & -1 \\ 0 & -1 & 1 \end{array}\right] \left[\begin{array}{c} x_1 \\ \hline x_2 \\ x_3 \end{array}\right]^{(1)} = \left[\begin{array}{c} 3 \\ \hline -1 \\ 0 \end{array}\right] - \left[\begin{array}{c|cc} 0 & -1 & 0 \\ \hline -1 & 0 & 0 \\ 0 & 0 & 0 \end{array}\right] \left[\begin{array}{c} 1 \\ \hline 1 \\ 1 \end{array}\right]$$

από την οποία παίρνουμε

$$2x_1^{(1)} = 4, \quad \begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} x_2 \\ x_3 \end{bmatrix}^{(1)} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

και επομένως

$$x^{(1)} = [2\ 0\ 0]^T.$$

Σημείωση: Το γεγονός ότι στη δεύτερη περίπτωση βρέθηκε η ακριβής λύση μετά μία επανάληψη είναι τελείως συμπτωματικό και οφείλεται στην επιλογή της αρχικής προσέγγισης $x^{(0)}$ και μόνο.

3.6.2 Οι Αλλες Block Επαναληπτικές Μέθοδοι

Οι block επαναληπτικές μέθοδοι των Gauss-Seidel και της SOR κατασκευάζονται με ανάλογο τρόπο. Οι περιορισμοί για την ύπαρξή τους είναι ακριβώς οι ίδιοι με αυτούς της block Jacobi. Δηλαδή, οι $A_{ii},\ i = 1(1)p$, πρέπει να είναι τετραγωνικοί αντιστρέψιμοι πίνακες. Στη συνέχεια δίνουμε τη μορφή των μεθόδων αυτών όταν επιλύσουμε ως προς την i-οστή block γραμμή. Συγκεκριμένα η block Gauss-Seidel μέθοδος θα είναι

$$A_{ii}x_i^{(k+1)} = b_i - \sum_{j=1}^{i-1} A_{ij}x_j^{(k+1)} - \sum_{j=i+1}^{p} A_{ij}x_j^{(k)},\ i = 1(1)p,\ k = 0, 1, 2, \cdots,$$

ή

$$x_i^{(k+1)} = A_{ii}^{-1}\left(b_i - \sum_{j=1}^{i-1} A_{ij}x_j^{(k+1)} - \sum_{j=i+1}^{p} A_{ij}x_j^{(k)}\right),\ i = 1(1)p,\ k = 0, 1, 2, \cdots.$$

Για την block SOR μέθοδο ισχύει κάτι ανάλογο. Συγκεκριμένα

$$A_{ii}x_i^{(k+1)} = (1-\omega)A_{ii}x_i^{(k)} + \omega\left(b_i - \sum_{j=1}^{i-1} A_{ij}x_j^{(k+1)} - \sum_{j=i+1}^{p} A_{ij}x_j^{(k)}\right),\ i = 1(1)p,\ k = 0,1,2,\cdots,$$

ή

$$x_i^{(k+1)} = (1-\omega)x_i^{(k)} + \omega A_{ii}^{-1}\left(b_i - \sum_{j=1}^{i-1} A_{ij}x_j^{(k+1)} - \sum_{j=i+1}^{p} A_{ij}x_j^{(k)}\right),\ i = 1(1)p,\ k = 0,1,2,\cdots.$$

Η θεωρία που αναπτύχτηκε τόσο για τις point Jacobi και Gauss-Seidel όσο και για τις point SOR και SSOR επαναληπτικές μεθόδους ισχύει με ελάχιστες ίσως τροποποιήσεις και στην περίπτωση των αντίστοιχων block επαναληπτικών μεθόδων.
Π.χ. για να αναφέρουμε μερικά από τα βασικά συμπεράσματα που αφορούν στην block SOR μέθοδο:
1) Για $\omega \in \mathbb{C}$ η συνθήκη $|\omega - 1| < 1$ αποτελεί την αναγκαία συνθήκη για τη σύγκλιση της SOR μεθόδου, η οποία για $\omega \in \mathbb{R}$ γίνεται $\omega \in (0,2)$. Η απόδειξη παραμένει η ίδια με αυτή που δόθηκε στην απόδειξη του Θεωρήματος 3.5 του Kahan. Η μόνη διαφορά είναι ότι όπου παρουσιάζεται το γινόμενο $a_{11}a_{22}\cdots a_{nn} = \det(D)$ αυτό αντικαθίσταται τώρα από το γινόμενο $\det(A_{11})\det(A_{22})\cdots\det(A_{pp}) = \det(D)$.
2) Για πίνακες Ερμιτιανούς και θετικά ορισμένους με $\omega \in \mathbb{R}$, η $\omega \in (0,2)$ αποτελεί αναγκαία και ικανή συνθήκη για τη σύγκλιση της SOR. Επίσης, η απόδειξη παραμένει η ίδια με αυτήν του Θεωρήματος 3.6 των Reich-Ostrowski-Varga. Η μόνη παρατήρηση που μπορεί να γίνει είναι ότι όπου παρουσιάζονται τα στοιχεία $a_{ii} > 0,\ i = 1(1)n$, τώρα έχουμε τους διαγώνιους υποπίνακες $A_{ii},\ i = 1(1)p$, που είναι Ερμιτιανοί και θετικά ορισμένοι σύμφωνα με το Θεώρημα 2.3 και άρα $d = (x, Dx)_2 > 0,\ \forall\ x \in \mathbb{C}^n \setminus \{0\}$.
3) Οι ορισμοί οι σχετικοί με το "δικυκλικό" και με το "συνεπώς διατεταγμένο" πίνακα $A \in \mathbb{C}^{m,n}$ με $\det(D) = \det(A_{11})\det(A_{22})\cdots\det(A_{pp}) \neq 0$, παραμένουν οι ίδιοι στη γενική τους μορφή. Η μόνη διαφορά έγκειται στο γεγονός ότι τώρα αναφερόμαστε σε blocks και άρα μιλάμε για "block δικυκλικό" και "block συνεπώς διατεταγμένο" πίνακα. Μεταξύ των πινάκων που ικανοποιούν τους ορισμούς είναι και οι block τριδιαγώνιοι πίνακες με διαγώνια blocks αντιστρέψιμους πίνακες. Π.χ.

$$\begin{bmatrix} A_{11} & A_{12} & & & \\ A_{21} & A_{22} & A_{23} & & \\ & \ddots & \ddots & \ddots & \\ & & A_{p-1,p} & A_{p-1,p-1} & A_{p-1,p} \\ & & & A_{p,p-1} & A_{pp} \end{bmatrix}.$$

Οι αποδείξεις των ιδιοτήτων, στην περίπτωση αυτή, παραμένουν οι ίδιες με τις αντίστοιχες της περίπτωσης του point τριδιαγώνιου πίνακα. Οι διαφορές έγκεινται στο ότι ο πίνακας P θα είναι τώρα ο

πίνακας με

$$P^T = [e^1\ e^2\ \cdots\ e^{n_1}\ e^{n_1+n_2+1}\ \cdots\ e^{n_1+n_2+n_3}\ \cdots\ e^{n_1+1}\ e^{n_1+2}\ \cdots\ e^{n_1+n_2}\ \cdots],$$

δηλαδή ο P^T θα είναι ένας block πίνακας με μή-μηδενικά τετραγωνικά blocks τα $(1,1)$, $(3,2)$, $(5,3)$, $\ldots$, $(2, \left[\frac{p+1}{2}\right]+1)$, $(4, \left[\frac{p+1}{2}\right]+2)$, $\ldots$, που θα είναι αντίστοιχα I_{n_1}, I_{n_3}, I_{n_5}, $\ldots$, I_{n_2}, I_{n_4}, $\ldots$, και ο E θα είναι ο $E = \mathrm{diag}(I_{n_1},\ \frac{1}{\alpha}I_{n_2}, \frac{1}{a^2}I_{n_3},\ \cdots,\ \frac{1}{a^{p-1}}I_{n_p})$, όπου $I_{n_i},\ i = 1(1)p$, ο $n_i \times n_i$, μοναδιαίος πίνακας ίδιων διαστάσεων με αυτές του A_{ii}.
4) Τέλος, όλες οι υπόλοιπες προτάσεις που αφορούν σε δικυκλικούς και συνεπώς διατεταγμένους πίνακες στην περίπτωση της SOR μεθόδου μεταφέρονται αυτούσια και στην περίπτωση των block δικυκλικών και συνεπώς διατεταγμένων πινάκων στην block SOR μέθοδο, με τις προφανείς τροποποιήσεις όταν και όπου αυτές χρειάζονται.

Στην περίπτωση της block SSOR μεθόδου ισχύουν και πάλι ορισμοί και προτάσεις ανάλογες με αυτές της point SSOR και επομένως μάλλον περιττεύει η επανάληψή τους. Δύο σημεία όμως θα πρέπει να τονιστούν ιδιαίτερα.
1) Σε κάποιες από τις αποδείξεις στην point SSOR μέθοδο χρησιμοποιήθηκε ο διαγώνιος πίνακας $D^{\frac{1}{2}} = \mathrm{diag}(a_{11}^{\frac{1}{2}}, a_{22}^{\frac{1}{2}}, \cdots, a_{nn}^{\frac{1}{2}})$, με την προϋπόθεση βέβαια ότι $a_{ii} > 0,\ i = 1(1)n$. Στην περίπτωση όμως της block SSOR ο D είναι block και όχι point διαγώνιος πίνακας. Γεννιέται, λοιπόν, το εύλογο ερώτημα αν και στην πρώτη περίπτωση είναι δυνατόν να οριστεί κάτι παρόμοιο ως "τετραγωνική ρίζα" πίνακα. Για το σκοπό αυτό διατυπώνουμε το παρακάτω θεώρημα που αφορά σε ύπαρξη και μονοσήμαντο, και όπου αποδείχνουμε μόνο την ύπαρξη.

Θεώρημα 3.13 *Εστω ότι ο $A \in \mathbb{C}^{n,n}$, με $A^H = A$, είναι θετικά ορισμένος. Τότε υπάρχει μοναδικός πίνακας $B \in \mathbb{C}^{n,n}$, με $B^H = B$, και θετικά ορισμένος, που ικανοποιεί την $B^2 = A$.*

Απόδειξη: Εστω ότι οι ιδιοτιμές του A, $\lambda_i,\ i = 1(1)n$, που είναι πραγματικές και θετικές, έχουν αντίστοιχα ιδιοδιανύσματα $x^i,\ i = 1(1)n$, τα οποία έχουν ληφτεί έτσι ώστε να αποτελούν μία ορθοκανονική βάση. Αν $X = [x^1\ x^2\ \cdots\ x^n]$ είναι ο πίνακας που έχει στήλες τα ιδιοδιανύσματα x^i, και $L = \mathrm{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n)$, τότε λόγω της ορθοκανονικότητας των x^i ($X^{-1} = X^H$) θα ισχύει ότι $A = X\Lambda X^H$. Η σχέση αυτή μπορεί να γραφτεί και ως $A = X\Lambda^{\frac{1}{2}}X^H X\Lambda^{\frac{1}{2}}X^H$, όπου $\Lambda^{\frac{1}{2}} = \mathrm{diag}(\lambda_1^{\frac{1}{2}}, \lambda_2^{\frac{1}{2}}, \cdots, \lambda_n^{\frac{1}{2}})$. Αν θέσουμε $B = X\Lambda^{\frac{1}{2}}X^H$, τότε έχουμε ότι $A = B^2$, όπου είναι πολύ εύκολο να δειχτεί ότι ο B ικανοποιεί τις απαιτήσεις της εκφώνησης, δηλαδή, ο B είναι Ερμιτιανός και θετικά ορισμένος. (Σημείωση: Ο πίνακας B καλείται και (θετική) τετραγωνική ρίζα του A και συμβολίζεται ως $B = A^{\frac{1}{2}}$.) Για το μονοσήμαντο η απόδειξη δεν είναι τετριμμένη και ο αναγνώστης παραπέμπεται στο βιβλίο του Young [48]. □

Με βάση το προηγούμενο θεώρημα είναι δυνατόν για έναν Ερμιτιανό και θετικά ορισμένο πίνακα A, ο οποίος έχει διαχωριστεί σε μια block μορφή, να οριστεί ο block διαγώνιος πίνακας $D = \mathrm{diag}(A_{11}, A_{22}, \cdots, A_{pp})$, που θα είναι κι αυτός, σύμφωνα με γνωστή πρόταση, Ερμιτιανός

και θετικά ορισμένος, οπότε θα μπορεί να οριστεί η (θετική) τετραγωνική ρίζα του, δηλαδή ο $D^{\frac{1}{2}} = \mathrm{diag}(A_{11}^{\frac{1}{2}}, A_{22}^{\frac{1}{2}}, \cdots, A_{pp}^{\frac{1}{2}})$. Ο $D^{\frac{1}{2}}$ θα είναι κι αυτός Ερμιτιανός και θετικά ορισμένος. Ετσι το Θεώρημα 3.10 θα ισχύει ως προς το πρώτο μέρος του και ως προς το ευθύ του δεύτερου μέρους του. Σ΄ ό,τι αφορά το αντίστροφο του δεύτερου μέρους θα πρέπει η υπόθεση να τροποποιηθεί έτσι ώστε να ισχύει ότι ο block διαγώνιος πίνακας D είναι θετικά ορισμένος. Εφόσον αυτό το τελευταίο ισχύει τότε και τα συμπεράσματα του αντίστοιχου Θεωρήματος 3.10 ισχύουν.
2) Σ΄ ό,τι αφορά τους ορισμούς και τις προτάσεις που αναφέρονται στους block δικυκλικούς και συνεπώς διατεταγμένους πίνακες A και που αφορούν στην ειδική μορφή (3.68) παραμένουν οι ίδιες με τη μόνη διαφορά ότι οι πίνακες D_1 και D_2, στην περίπτωση της block SSOR μεθόδου, είναι απλά τετραγωνικοί και αντιστρέψιμοι. Τονίζεται απλά ότι ο πίνακας A, όπως δόθηκε στην (3.68) είναι συγχρόνως point και block δικυκλικός και συνεπώς διατεταγμένος, ενώ όταν οι πίνακες D_1 και D_2 είναι γενικά τετραγωνικοί και αντιστρέψιμοι τότε ο A είναι μόνο block δικυκλικός και συνεπώς διατεταγμένος.

ΑΣΚΗΣΕΙΣ

1.: Να θεωρηθούν οι δυο block διαχωρισμοί του πραγματικού πίνακα A,

$$A := A_1 := \left[\begin{array}{c|cc} a & 0 & b \\ \hline c & d & 0 \\ 0 & e & f \end{array}\right] = \left[\begin{array}{cc|c} a & 0 & b \\ c & d & 0 \\ \hline 0 & e & f \end{array}\right] =: A_2,$$

με $adf \neq 0$. Να προσδιοριστούν όλες οι ιδιοτιμές των δύο block επαναληπτικών πινάκων Jacobi, J_1 και J_2, που αντιστοιχούν στους παραπάνω block διαχωρισμούς, και να συγκριθούν οι φασματικές ακτίνες τους.

2.: Δίνεται ο πίνακας A στην παρακάτω block μορφή $A = \left[\begin{array}{cc|cc} 4 & -1 & -1 & 0 \\ -1 & 4 & 0 & -1 \\ \hline -1 & 0 & 4 & -1 \\ 0 & -1 & -1 & 4 \end{array}\right]$.

α) Να εξεταστούν ως προς τη σύγκλιση και να συγκριθούν ως προς την ταχύτητα σύγκλισης οι μέθοδοι block Jacobi και Gauss-Seidel, που αντιστοιχούν στον παραπάνω block διαχωρισμό του πίνακα A. και
β) Να βρεθεί επίσης η βέλτιστη SOR παράμετρος $\omega \in I\!R$ της block SOR, που αντιστοιχεί στον ίδιο block διαχωρισμό.

3.: Δίνεται ο (σημειακός) πίνακας A και ένας block διαχωρισμός του. Συγκεκριμένα:

$$A = \left[\begin{array}{ccc} 1 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{array}\right] = \left[\begin{array}{cc|c} 1 & -1 & 0 \\ -1 & 2 & -1 \\ \hline 0 & -1 & 2 \end{array}\right].$$

α) Με βάση τους αντίστοιχους ορισμούς να δειχτεί ότι ο block διαχωρισμένος A είναι block δικυκλικός και συνεπώς διατεταγμένος. και
β) Δεδομένου ότι ο A στη σημειακή μορφή του είναι point δικυκλικός και συνεπώς διατεταγμένος, να βρεθούν και στις δυο περιπτώσεις, αν είναι δυνατόν, οι βέλτιστες SOR παράμετροι καθώς και οι βέλτιστες φασματικές ακτίνες των αντίστοιχων point και block SOR επαναληπτικών πινάκων.

4.: Δίνεται το σύστημα $Ax = b$, $A = \begin{bmatrix} 2 & 0 & -2 \\ -1 & 2 & 0 \\ 0 & -1 & 2 \end{bmatrix}$, $b \in I\!R^3$.
α) Να εξεταστούν ως προς τη σύγκλιση οι μέθοδοι Jacobi και Gauss-Seidel.
β) Επίσης να εξεταστουν ως προς τη σύγκλιση οι μέθοδοι block Jacobi, block Gauss-Seidel και η βελτιστη SOR, που βασίζονται στον block διαχωρισμό $A = \left[\begin{array}{cc|c} 2 & 0 & -2 \\ -1 & 2 & 0 \\ \hline 0 & -1 & 2 \end{array}\right]$, $b \in I\!R^3$. και
γ) Να εξεταστούν όλες οι προηγούμενες μέθοδοι μεταξύ τους ως προς την ταχύτητα σύγκλισης.

5.: Δίνεται το γραμμικό σύστημα $Ax = b$, με $A = \begin{bmatrix} 2 & -1 & & \\ -1 & 2 & -1 & \\ & -1 & 2 & -1 \\ & & -1 & 2 \end{bmatrix}$ και $b = [1\ 0\ 0\ 1]^T$.
Να εξεταστούν ως προς τη σύγκλιση μεταξύ τους οι μέθοδοι block Jacobi, block Gauss-Seidel και η βέλτιστη block SOR, που βασίζονται στον block διαχωρισμό

$$A = \left[\begin{array}{cc|cc} 2 & -1 & & \\ -1 & 2 & -1 & \\ \hline & -1 & 2 & -1 \\ & & -1 & 2 \end{array}\right]$$

και να γίνουν δύο επαναλήψεις κάθε μιας με $x^{(0)} = 0 \in I\!R^4$.

6.: Δίνεται το διαχωρισμένο σε blocks γραμμικό σύστημα

$$Ax := \left[\begin{array}{cc|c} 2 & -1 & 0 \\ -1 & 2 & -1 \\ \hline 0 & -1 & 1 \end{array}\right] \left[\begin{array}{c} x_1 \\ x_2 \\ \hline x_3 \end{array}\right] = \left[\begin{array}{c} 3 \\ -1 \\ \hline 0 \end{array}\right] =: b.$$

Χρησιμοποιώντας ως αρχικό διάνυσμα το $x^{(0)} = [1\ 1\ 1]^T$, να εκτελεστούν:
α) Μία επανάληψη της block επαναληπτικής μεθόδου του Jacobi. και
β) Μία επανάληψη της block επαναληπτικής μεθόδου των Gauss-Seidel.

7.: Δίνεται ο block τριδιαγώνιος πίνακας

$$\begin{bmatrix} I_{n_1} & A_{12} & & & \\ A_{21} & I_{n_2} & A_{23} & & \\ & \ddots & \ddots & \ddots & \\ & & A_{n_{p-1},n_{p-2}} & I_{n_{p-1}} & A_{n_{p-1},n_p} \\ & & & A_{n_p,n_{p-1}} & I_{n_p} \end{bmatrix} \in \mathbb{C}^{n,n} \quad \kappa\alpha\iota \quad \sum_{i=1}^{p} n_i = n,$$

όπου $I_{n_i} \in \mathbb{C}^{n_i n_i},\ i = 1(1)p$, ο μοναδιαίος πίνακας.
α) Να αποδειχτεί ότι είναι συγχρόνως και point αλλά και block δικυκλικός και συνεπώς διατεταγμένος. και
β) Υπάρχει κάποια διαφορά στα φάσματα των ιδιοτιμών των αντίστοιχων point και block Jacobi επαναληπτικών πινάκων και γιατί;

4 Ημι-επαναληπτικές Μέθοδοι

Υποθέτουμε ότι για την επίλυση του γραμμικού συστήματος $Ax = b$ με $A \in \mathbb{C}^{n,n}$, $\det(A) \neq 0$ και $b \in \mathbb{C}^n$, χρησιμοποιώντας κάποιον προρρυθμιστή, έχει προκύψει η (συγκλίνουσα) επαναληπτική μέθοδος

$$x^{(k+1)} \;=\; Tx^{(k)} + c,\;\; k = 0, 1, 2, \cdots, \tag{4.1}$$

με T Ερμιτιανό και $x^{(0)} \in \mathbb{C}^n$ οποιοδήποτε.

Το διάνυσμα-σφάλμα $e^{(k)}$ στην k επανάληψη είναι προφανώς ίσο με

$$e^{(k)} \;=\; x^{(k)} - x,$$

όπου x η ακριβής λύση του συστήματος. Η βασική ιδέα των ημι-επαναληπτικών μεθόδων είναι να κατασκευστεί μία νέα ακολουθία επαναλήψεων κάθε όρος της οποίας θα είναι ένας βαρυκεντρικός μέσος όρος όλων των μέχρι τότε επαναλήψεων της αρχικής μεθόδου (μέθοδος βάσης), έτσι ώστε μετά από k επαναλήψεις, και για κάθε k, η μέση ταχύτητα σύγκλισης της νέας ακολουθίας να είναι η μέγιστη δυνατή. Αν από την αρχική ακολουθία των επαναλήψεων $\{x^{(k)}\}_{k=0}^{\infty}$ δημιουργήσουμε τη νέα ακολουθία επαναλήψεων $\{y^{(k)}\}_{k=0}^{\infty}$, με τον τρόπο που περιγράφτηκε, τότε θα έχουμε

$$y^{(k)} \;=\; \sum_{j=0}^{k} a_{kj} x^{(j)},\;\; k = 0, 1, 2, \cdots, \tag{4.2}$$

όπου

$$\sum_{j=0}^{k} a_{kj} \;=\; 1,\;\; \mu\epsilon\; a_{kj} \in \mathbb{R},\;\; k = 0, 1, 2, \cdots. \tag{4.3}$$

(Σημείωση: Προφανώς για $a_{kk} = 1$ και $a_{kj} = 0,\;\; j = 0(1)k-1$, η ακολουθία (4.2) ταυτίζεται με την αρχική (4.1).)

Αν ορίσουμε τα πολυώνυμα

$$p_k(z) \;=\; \sum_{j=0}^{k} a_{kj} z^j,\;\; k = 0, 1, 2, \cdots,$$

βαθμού το πολύ k, τότε το διάνυσμα-σφάλμα $\widetilde{e}^{(k)}$ της νέας ακολουθίας $\{y^{(k)}\}_{k=0}^{\infty}$ θα δίνεται από την έκφραση

$$\widetilde{e}^{(k)} = y^{(k)} - x = \sum_{j=0}^{k} a_{kj} x^{(j)} - \sum_{j=0}^{k} a_{kj} x = \sum_{j=0}^{k} a_{kj} e^{(j)} = \sum_{j=0}^{k} a_{kj} (T^j e^{(0)}) = p_k(T) e^{(0)}. \tag{4.4}$$

Από την (4.4), χρησιμοποιώντας τη φυσική ℓ_2-norm, και εκμεταλλευόμενοι το γεγονός ότι ο T είναι Ερμιτιανός και οι συντελεστές του $p_k(z)$ πραγματικοί, έχουμε ότι για κάθε k και κάθε $p_k(z)$ υπάρχει $e^{(0)} \in \mathcal{C}^n \backslash \{0\}$ τ.ω. να μπορούμε να έχουμε αμέσως τις σχέσεις

$$\sup_{e^{(0)} \in \mathcal{C}^n \backslash \{0\}} \left(\frac{||\widetilde{e}^{(k)}||_2}{||e^{(0)}||_2} \right) = ||p_k(T)||_2 \ = \ \rho(p_k(T)) \ = \ \max_{\lambda_i \in \sigma(T)} |p_k(\lambda_i)|, \ k = 0, 1, 2, \cdots,$$

και ακόμη ότι

$$\sup_{e^{(0)} \in \mathcal{C}^n \backslash \{0\}} \left(\frac{||\widetilde{e}^{(k)}||_2}{||e^{(0)}||_2} \right) \ = \ \max_{\lambda_i \in \sigma(T)} |p_k(\lambda_i)| \ \leq \ \max_{-1 < \alpha \leq z \leq \beta < 1} |p_k(z)|, \tag{4.5}$$

όπου θεωρήθηκε ότι $-1 < \alpha \leq \lambda_i \leq \beta < 1$, πράγμα που ισχύει αφού ο T είναι Ερμιτιανός με $\rho(T) < 1$.

Από τον ορισμό της μέσης ταχύτητας σύγκλισης μετά k επαναλήψεις της (3.11), στην παρούσα περίπτωση, όπου χρησιμοποιούμε ℓ_2-norms, θα έχουμε κατ΄ αναλογίαν ότι

$$\mathcal{R}(p_k(T)) \ \equiv \ -\frac{\ln ||p_k(T)||_2}{k}. \tag{4.6}$$

Για να πετύχουμε, λοιπόν, τη μέγιστη ταχύτητα σύγκλισης της ακολουθίας $\{y^{(k)}\}_{k=0}^{\infty}$, με βάση την ανάλυση που προηγήθηκε, θα πρέπει να ελαχιστοποιήσουμε το δεύτερο μέλος της (4.5) για κάθε k. Δηλαδή, αν $\mathcal{P}_k$ είναι το σύνολο των πραγματικών πολυωνύμων βαθμού το πολύ k, έχουμε να επιλύσουμε το ακόλουθο min-max πρόβλημα

$$\min_{p_k \in \mathcal{P}_k, \ p_k(1)=1} \left\{ \max_{-1 < \alpha \leq z \leq \beta < 1} |p_k(z)| \right\}. \tag{4.7}$$

Η λύση του min-max προβλήματος (4.7) είναι κλασική και δίνεται από τα πολυώνυμα του Chebyshev πρώτου είδους τα οποία ορίζονται ως εξής

$$\begin{aligned} T_0(z) \ &= \ 1, \\ T_1(z) \ &= \ z, \\ T_m(z) \ &= \ 2zT_{m-1}(z) - T_{m-2}(z), \ m = 2, 3, 4, \cdots. \end{aligned} \tag{4.8}$$

Με βάση τον παραπάνω ορισμό μπορεί να αποδειχτούν επαγωγικά οι ακόλουθες ιδιότητες των πολυωνύμων του Chebyshev.

α) βαθμός$(T_m(z)) = m$, $T_m(1) = 1$, $T_m(z)$ άρτια (αντίστοιχα, περιττή) συνάρτηση του z αν m άρτιος (αντίστοιχα, περιττός).

β) $T_m(z) \ = \ \begin{cases} \cos(m \cos^{-1} z), & \alpha\nu \ -1 \leq z \leq 1, \\ \cosh(m \cosh^{-1} z), & \alpha\nu \ z \geq 1, \end{cases}$

δηλαδή

β1) $T_m(z) \;=\; \cos(m\theta),\;\; z=\cos\theta,\; \theta\in[0,\pi],\; \alpha\nu\nu\; -1\le z\le 1,$
β2) $T_m(z) \;=\; \cosh(mu),\;\; z=\cosh u,\; u\ge 0,\; \alpha\nu\nu\; z\ge 1.$
γ) $T_m(z) \;=\; \frac{1}{2}[(z+(z^2-1)^{\frac{1}{2}})^m+(z-(z^2-1)^{\frac{1}{2}})^m].$
Επιπλέον, για $z\in[-1,1]$ μπορεί να αποδειχτούν αμέσως και οι παρακάτω ιδιότητες
δ) $\max_{-1\le z\le 1}|T_m(z)| \;=\; 1.$
ε) $T_m(z_j) \;=\; (-1)^j,\; \gamma\iota\alpha\; z_j \;=\; \cos\left(\frac{j\pi}{m}\right),\; j=0(1)m.$
στ) $T_m(z_l) \;=\; 0,\; \gamma\iota\alpha\; z_l \;=\; \cos\left(\frac{(2l+1)\pi}{2m}\right),\; l=0(1)m-1.$

Τα πολυώνυμα του Chebyshev χρησιμεύουν για τη λύση του min-max προβλήματος (4.7) λόγω της ακόλουθης ιδιότητάς τους που δίνεται στη συνέχεια με μορφή πρότασης.

Θεώρημα 4.1 *Εστω ότι* $-1<\alpha<\beta<1$. *Τότε το* min-max *πρόβλημα*

$$\min_{p_m\in\mathcal{P}_m,\; p_m(1)=1}\left(\max_{-1<\alpha\le z\le\beta<1}|p_m(z)|\right), \tag{4.9}$$

όπου $\mathcal{P}_m$ *το σύνολο των πολυωνύμων βαθμού μικρότερου ή ίσου του* m *με πραγματικούς συντελεστές, λύνεται μοναδικά από το πολυώνυμο*

$$\widetilde{p}_m(z) \;=\; \frac{T_m\left(\frac{2z-(\beta+\alpha)}{\beta-\alpha}\right)}{T_m\left(\frac{2-(\beta+\alpha)}{\beta-\alpha}\right)} \tag{4.10}$$

για το οποίο ισχύει ότι

$$\max_{\alpha\le z\le\beta}|\widetilde{p}_m(z)| \;=\; \frac{1}{T_m\left(\frac{2-(\beta+\alpha)}{\beta-\alpha}\right)}. \tag{4.11}$$

Απόδειξη: Καταρχάς θα πρέπει να διαπιστωθεί αν το πολυώνυμο $\widetilde{p}_m(z)$, που δίνεται αναλυτικά, ορίζεται, ανήκει στο σύνολο $\mathcal{P}_m$, ικανοποιεί τη συνθήκη $p_m(1)=1$ και τέλος αν η μέγιστη απόλυτα τιμή του στο διάστημα $[\alpha,\beta]$ δίνεται πράγματι από την έκφραση της (4.11). Μπορεί εύκολα να διαπιστωθεί ότι το όρισμα του αριθμητή του δεύτερου μέλους της (4.10) για $\alpha\le z\le\beta$ παίρνει τιμές στο διάστημα $[-1,\ 1]$ και συγκεκριμένα ο γραμμικός μετασχηματισμός

$$z\mapsto y,\; y=\frac{2z-(\beta+\alpha)}{\beta-\alpha} \tag{4.12}$$

είναι ένα προς ένα επί του διαστήματος $[\alpha,\ \beta]$ στο $[-1,\ 1]$. Αρα ο αριθμητής στο κλάσμα της (4.11) ορίζεται και μάλιστα από την ιδιότητα (β1), με την προφανή αντικατάσταση του z στη (β1) από το y της (4.12). Το όρισμα του παρονομαστή του ίδιου κλάσματος μπορεί να διαπιστωθεί αμέσως ότι είναι πάντα μεγαλύτερο του 1. Αρα κι ο παρονομαστής ορίζεται με βάση όμως την

ιδιότητα (β2) με την αντικατάσταση του z της (β2) από τη σταθερά $\frac{2-(\beta+\alpha)}{\beta-\alpha}$. Συνεπώς το $\widetilde{p}_m(z)$ ορίζεται και είναι προφανώς πολυώνυμο βαθμού m ως προς z αφού ο αριθμητής στην (4.10) είναι πολυώνυμο του Chebyshev βαθμού m με όρισμα γραμμική συνάρτηση του z και ο παρονομαστής είναι σταθερά. Αρα $\widetilde{p}_m(z) \in \mathcal{P}_m$. Επιπλέον, αμέσως διαπιστώνεται ότι $\widetilde{p}_m(1) = 1$. Το μόνο που απομένει να διαπιστωθεί είναι η ικανοποίηση της (4.11). Πράγματι, ο παρονομαστής είναι μια θετική σταθερά, αφού το όρισμα του αντίστοιχου πολυωνύμου του Chebyshev είναι μεγαλύτερο από 1 και ο ορισμός του γίνεται από τη (β2). Ακόμη, λόγω της ένα προς ένα επί απεικόνισης του διαστήματος $[\alpha,\ \beta]$ για το z στο διάστημα $[-1,\ 1]$ για το y δίνει από την ιδιότητα (δ) ότι $\max_{\alpha\le z\le\beta}|\widetilde{p}_m(z)| = \frac{\max_{-1\le y\le 1}|T_m(y)|}{|T_m(\frac{2-(\beta+\alpha)}{\beta-\alpha})|} = \frac{1}{T_m\left(\frac{2-(\beta+\alpha)}{\beta-\alpha}\right)}$. Είναι φανερό ότι αυτή η μέγιστη τιμή του $|\widetilde{p}_m(z)|$ λαμβάνεται διαμέσου της μέγιστης τιμής του $|T_m(y)|$. Η μέγιστη τιμή του $|T_m(y)|$, με βάση τις ιδιότητες (δ) και (ε), λαμβάνεται στα σημεία $y_j = \cos\left(\frac{j\pi}{m}\right),\ j = 0(1)m$, όπου οι τιμές του $T_m(y)$ είναι $T_m(y_j) = (-1)^j$, δηλαδή τιμές με απόλυτη τιμή μονάδα και εναλλασσόμενα πρόσημα. Η αντίστοιχη μέγιστη τιμή του $|\widetilde{p}_m(z)|$ θα λαμβάνεται ενόψει της (4.12) στα $m+1$ διακριτά σημεία $z_j = \frac{1}{2}[(\beta-\alpha)y_j + (\beta+\alpha)],\ j = 0(1)m$, του διαστήματος $[\alpha,\beta]$, και επομένως το $\widetilde{p}_m(z)$ θα παίρνει στα σημεία αυτά την απόλυτα μέγιστη τιμή $\frac{1}{T_m\left(\frac{2-(\beta+\alpha)}{\beta-\alpha}\right)}$ με εναλλασσόμενα πρόσημα.

Απομένει τώρα να δειχτεί ότι το πολυώνυμο $\widetilde{p}_m(z)$ είναι το μόνο που λύνει το πρόβλημα του θεωρήματος. Για το σκοπό αυτό θα πρέπει να δειχτεί ότι δεν υπάρχει άλλο πολυώνυμο με πραγματικούς συντελεστές βαθμού το πολύ m και άθροισμα συντελεστών μονάδα που να έχει μέγιστη απόλυτη τιμή στο διάστημα $[\alpha,\ \beta]$ μικρότερη ή ίση από την $\frac{1}{T_m\left(\frac{2-(\beta+\alpha)}{\beta-\alpha}\right)}$. Στη συνέχεια θα αποδειχτεί η περίπτωση του "μικρότερη" μόνον. Η περίπτωση του "ίση" αποδείχνεται με παρόμοια επιχειρηματολογία και παραλείπεται. (Ο ενδιαφερόμενος αναγώστης παραπέμπεται στο βιβλίο του Young [48].) Εστω, λοιπόν, ότι υπάρχει και άλλο πολυώνυμο $q_m(z)$ με πραγματικούς συντελεστές, βαθμού μικρότερου ή ίσου του m, με $q_m(1) = 1$, και τέτοιο ώστε

$$\max_{\alpha\le z\le\beta}|q_m(z)| < \max_{\alpha\le z\le\beta}|\widetilde{p}_m(z)|. \tag{4.13}$$

Θεωρούμε τη διαφορά $r_m(z) = \widetilde{p}_m(z) - q_m(z)$, που είναι πολυώνυμο βαθμού μικρότερου ή ίσου του m. Είναι φανερό ότι, αρχίζοντας από το $z_0 = \beta$ και καταλήγοντας στο $z_m = \alpha$, το $r_m(z)$, λόγω της (4.13), παίρνει στα σημεία $z_j,\ j = 0(1)m$, τιμές με πρόσημα εναλλάξ θετικά και αρνητικά. Επομένως, σύμφωνα με το Θεώρημα του Rolle, σε κάθε ένα από τα διαστήματα $(z_{j+1},\ z_j),\ j = 0(1)m-1$, υπάρχει μία τουλάχιστον ρίζα s_j του $r_m(z)$. Αρα θα υπάρχουν τουλάχιστον m διακριτά σημεία $s_j,\ j = 0(1)m-1$, τ.ω. $-1 < \alpha \le z_{j+1} < s_j < z_j \le \beta < 1$, όπου $r_m(s_j) = 0$. Εχουμε όμως και $r_m(1) = \widetilde{p}_m(1) - q_m(1) = 0$, όπου το 1 είναι εκτός του διαστήματος $[\alpha,\ \beta]$. Δηλαδή, το πολυώνυμο $r_m(z)$ έχει τουλάχιστον $m+1$ ρίζες πράγμα που είναι άτοπο. □

Επανερχόμαστε στην ημι-επαναληπτική μέθοδο που έχουμε θεωρήσει: Με τη χρησιμοποίηση των πολυωνύμων του Chebyshev και με την υπόθεση ότι γνωρίζουμε μόνο τη φασματική ακτίνα, $\rho(T)$, του επαναληπτικού πίνακα T του αρχικού σχήματος, το πρόβλημα που θέσαμε στην αρχή του παρόντος κεφαλαίου έχει ήδη λυθεί. Ομως ένα μειονέκτημα της νέας μεθόδου φαίνεται να είναι το ακόλουθο. Σε κάθε βήμα k η νέα μέθοδος χρησιμοποιεί όλες τις προηγούμενες

επαναλήψεις $x^{(j)},\ j = 0(1)k$, του αρχικού επαναληπτικού σχήματος. Αυτό σημαίνει ότι για να βρεθεί η συγκεκριμένη επανάληψη $y^{(k)}$ της ημι-επαναληπτικής μεθόδου θα πρέπει πρώτα να βρίσκεται η αντίστοιχη επανάληψη $x^{(k)}$ του αρχικού σχήματος και έπειτα, με το βαρυκεντρικό μέσο όρο από τις (4.2)-(4.3) και με ό,τι υποδείχνει η θεωρία των πολυωνύμων του Chebyshev, να βρίσκεται η $y^{(k)}$. Η διαδικασία όμως αυτή αυξάνει σημαντικά (ιδιαίτερα όταν το k αυξάνει) το κόστος ανά επανάληψη σε πλήθος πράξεων της νέας μεθόδου, πράγμα που είναι αποτρεπτικό για τη χρησιμοποίησή της. Αν όμως εξετάσουμε λίγο προσεκτικότερα την αναδρομική σχέση του ορισμού των πολυωνύμων του Chebyshev, είναι δυνατόν να διαπιστώσουμε ότι το μειονέκτημα αυτό ξεπερνιέται.

Συγκεκριμένα, από το Θεώρημα 4.1 με $-1 < -\rho = \alpha \leq z \leq \beta = \rho < 1$, όπου $\rho = \rho(T)$, έχουμε ότι

$$\widetilde{p}_m(z) = \frac{1}{T_m\left(\frac{1}{\rho}\right)} T_m\left(\frac{z}{\rho}\right). \tag{4.14}$$

Χρησιμοποιώντας την έκφραση της (4.14) έχουμε $T_m\left(\frac{z}{\rho}\right) = T_m\left(\frac{1}{\rho}\right)\widetilde{p}_m(z)$, οπότε με βάση την τελευταία γράφοντας το $T_{k+1}\left(\frac{z}{\rho}\right)$ σα συνάρτηση των $T_k\left(\frac{z}{\rho}\right)$ και $T_{k-1}\left(\frac{z}{\rho}\right)$, από την (4.8), καταλήγουμε στην

$$T_{k+1}\left(\frac{1}{\rho}\right)\widetilde{p}_{k+1}(z) \ = \ \frac{2z}{\rho} T_k\left(\frac{1}{\rho}\right)\widetilde{p}_k(z) - T_{k-1}\left(\frac{1}{\rho}\right)\widetilde{p}_{k-1}(z), \quad k \geq 1, \tag{4.15}$$

όπου, από την (4.10), έχουμε $\widetilde{p}_0(z) = 1$ και $\widetilde{p}_1(z) = z$. Αν στην (4.15) θέσουμε αντί z τον πίνακα T θα έχουμε μια αντίστοιχη ισότητα πολυωνυμικών εκφράσεων του πίνακα T. Αν τα μέλη της ισότητας αυτής πολλαπλασιάσουμε από τα δεξιά επί $e^{(0)}$, τότε λόγω της (4.4), με $\widetilde{p}_m(T)$ στη θέση του $p_m(T)$, που δίνει $\widetilde{p}_m(T)e^{(0)} = \widetilde{e}^{(m)}$, έχουμε

$$T_{k+1}\left(\frac{1}{\rho}\right)\widetilde{e}^{(k+1)} \ = \ \frac{2}{\rho} T_k\left(\frac{1}{\rho}\right) T\widetilde{e}^{(k)} - T_{k-1}\left(\frac{1}{\rho}\right)\widetilde{e}^{(k-1)}, \quad k \geq 1.$$

Χρησιμοποιώντας τον ορισμό των $\widetilde{e}^{(k)} = y^{(k)} - x$, από την (4.4), και έχοντας υπόψη ότι $x = Tx + c$, από τη (4.1), είναι δυνατόν να καταλήξουμε στην

$$y^{(k+1)} \ = \ \frac{2T_k\left(\frac{1}{\rho}\right)}{\rho T_{k+1}\left(\frac{1}{\rho}\right)} Ty^{(k)} - \frac{T_{k-1}\left(\frac{1}{\rho}\right)}{T_{k+1}\left(\frac{1}{\rho}\right)} y^{(k-1)} + \frac{2T_k\left(\frac{1}{\rho}\right)}{\rho T_{k+1}\left(\frac{1}{\rho}\right)} c$$

ή στην

$$y^{(k+1)} \ = \ \omega_{k+1} Ty^{(k)} + (1 - \omega_{k+1}) y^{(k-1)} + \omega_{k+1} c, \tag{4.16}$$

αν θέσουμε

$$\omega_{k+1} = \frac{2T_k\left(\frac{1}{\rho}\right)}{\rho T_{k+1}\left(\frac{1}{\rho}\right)} \ \left(= 1 + \frac{T_{k-1}\left(\frac{1}{\rho}\right)}{T_{k+1}\left(\frac{1}{\rho}\right)} \right), \ k \geq 1. \tag{4.17}$$

Σημειώνεται ότι από την $\omega_2 = \frac{2}{2-\rho^2}$, που προκύπτει από την (4.17) για $k = 1$, μπορούν να βρεθούν όλα τα άλλα ω_{k+1} από τη σχέση

$$\omega_{k+1} \;=\; \frac{1}{1-\left(\frac{\rho^2\omega_k}{4}\right)},\ k \geq 2. \tag{4.18}$$

Η (4.18) προκύπτει αν απαλειφτεί το $T_k\left(\frac{1}{\rho}\right)$ από τις $\omega_{k+1} = \frac{2T_k\left(\frac{1}{\rho}\right)}{\rho T_{k+1}\left(\frac{1}{\rho}\right)}$ και $\omega_k = \frac{2T_{k-1}\left(\frac{1}{\rho}\right)}{\rho T_k\left(\frac{1}{\rho}\right)}$ και στη συνέχεια αντικατασταθεί ο λόγος $\frac{T_{k-1}\left(\frac{1}{\rho}\right)}{T_{k+1}\left(\frac{1}{\rho}\right)}$ με βάση την εντός των παρενθέσεων έκφραση της σχέσης (4.17) από το $\omega_{k+1} - 1$. Μ′ αυτόν τον τρόπο είναι περιττό να βρίσκονται και να χρησιμοποιούνται οι αναλυτικές εκφράσεις των πολυωνύμων του Chebyshev $T_k\left(\frac{1}{\rho}\right)$ και $T_{k+1}\left(\frac{1}{\rho}\right)$, πράγμα που ελαττώνει σημαντικά τους ανά επανάληψη υπολογισμούς.

Παρατηρήσεις: Στη συνέχεια παραθέτουμε μερικές βασικές παρατηρήσεις που αφορούν στη συμπεριφορά των παραμέτρων $\omega_{k+1},\ k \geq 1$.

α) Είναι γνωστό ότι η συνάρτηση $\cosh z$ είναι γνήσια αύξουσα συνάρτηση του z στο διάστημα $[1, \infty)$. Αρα ο λόγος $\frac{T_{k-1}\left(\frac{1}{\rho}\right)}{T_{k+1}\left(\frac{1}{\rho}\right)}$ βρίσκεται στο ανοικτό διάστημα $(0, 1)$, αφού, αν θέσουμε $\frac{1}{\rho} = \cosh u$, για $k > 1$ είναι $T_{k-1}\left(\frac{1}{\rho}\right) = \cosh\left((k-1)u\right),\ T_{k+1}\left(\frac{1}{\rho}\right) = \cosh\left((k+1)u\right)$ και $0 < (k-1)u < (k+1)u$, ενώ για $k = 1$ ο αντίστοιχος λόγος είναι $\frac{T_0\left(\frac{1}{\rho}\right)}{T_2\left(\frac{1}{\rho}\right)} = \frac{\rho^2}{2-\rho^2} < 1$. Το συμπέρασμα είναι ότι όλοι οι όροι της ακολουθίας $\{\omega_k\}_{k=2}^{\infty}$ βρίσκονται στο ανοικτό διάστημα $(1, 2)$.

β) Οι λόγοι $\frac{T_{k-1}\left(\frac{1}{\rho}\right)}{T_{k+1}\left(\frac{1}{\rho}\right)}$ αποτελούν μια γνήσια φθίνουσα ακολουθία ως προς k. Για το σκοπό αυτό αρκεί να αποδείξουμε ότι για δυο διαδοχικούς λόγους ισχύει

$$\frac{T_{k-1}\left(\frac{1}{\rho}\right)}{T_{k+1}\left(\frac{1}{\rho}\right)} \;>\; \frac{T_k\left(\frac{1}{\rho}\right)}{T_{k+2}\left(\frac{1}{\rho}\right)}$$

ή ισοδύναμα ότι

$$T_{k-1}\left(\frac{1}{\rho}\right) T_{k+2}\left(\frac{1}{\rho}\right) \;>\; T_k\left(\frac{1}{\rho}\right) T_{k+1}\left(\frac{1}{\rho}\right)$$

ή

$$\cosh\left((k-1)u\right)\cosh\left((k+2)u\right) \;>\; \cosh\left(ku\right)\cosh\left((k+1)u\right)$$

ή

$$\cosh\left((2k+1)u\right) + \cosh\left(3u\right) \;>\; \cosh\left((2k+1)u\right) + \cosh u,$$

που προφανώς ισχύει.

γ) Με βάση τα συμπεράσματα των παρατηρήσεων (α) και (β) έχουμε ότι η ακολουθία $\{\omega_k\}_{k=2}^{\infty}$

είναι γνήσια φθίνουσα και φραγμένη εκ των κάτω από το 1. Αρα θα συγκλίνει. Επομένως, αν $\lim_{k\to\infty}\omega_k = \widetilde{\omega}$ θα εχουμε από την (4.18), αν πάρουμε όρια με k να τείνει στο ∞, ότι ο $\widetilde{\omega}$ ικανοποιεί την εξίσωση

$$\rho^2\widetilde{\omega}^2 - 4\widetilde{\omega} + 4 = 0$$

της οποίας η μόνη ρίζα στο διάστημα $(1,2)$ δίνεται από την έκφραση

$$\widetilde{\omega} = \frac{2}{1+\sqrt{1-\rho^2}}. \tag{4.19}$$

Η έκφραση για το $\widetilde{\omega}$ της (4.19) είναι η ίδια έκφραση που βρίσκεται για τη βέλτιστη παράμετρο της SOR μεθόδου στην περίπτωση όπου το αρχικό προς επίλυση σύστημα είναι το $(I-T)x = c$, και οι εμπλεκόμενοι πίνακες έχουν επιπλέον τις ιδιότητες, ο πίνακας $I-T$ να είναι δικυκλικός και συνεπώς διατεταγμένος και ο πίνακας T να είναι ο επαναληπτικός πίνακας της αντίστοιχης μεθόδου Jacobi.

Το επαναληπτικό σχήμα (4.16), στο οποίο έχουμε καταλήξει, είναι απαλλαγμένο από όλα τα μειονεκτήματα που εμφάνιζε η αρχική μορφή του. Ακόμη, είναι γνωστό από τον ορισμό της ακολουθίας $y^{(k)}$ ότι $y^{(0)} = x^{(0)}$, και επομένως αν τεθεί $\omega_1 = 1$, τότε είναι δυνατόν να διαπιστωθεί ότι $y^{(1)} = Ty^{(0)} + c$. Εξάλλου από την (4.16) είναι φανερό ότι όλες οι επόμενες επαναλήψεις $y^{(k+1)}, k \geq 1$, είναι απλές συναρτήσεις των δύο προηγούμενων επαναλήψεων. Η τελευταία παρατήρηση καθιστά την προτεινόμενη νέα μέθοδο κάτι το τελείως διαφορετικό από τις μέχρι τώρα επαναληπτικές μεθόδους, που μελετήσαμε στα προηγούμενα κεφάλαια. Επιπλέον είναι δυνατόν να παρατηρήσουμε πως αν "δεχτούμε" προς στιγμήν ότι $y^{(k-1)} = y^{(k)}$, τότε το σχήμα που προκύπτει θυμίζει έντονα το extrapolated σχήμα του αρχικού $x^{(k+1)} = Tx^{(k)} + c$, με τη μόνη διαφορά ότι η extrapolation παράμετρος ω_{k+1} μεταβάλλεται από επανάληψη σε επανάληψη. Λόγω των δύο αυτών χαρακτηριστικών, δηλαδή της εξάρτησης της $k+1$ επανάληψης από τις δυο προηγούμενες και της μεταβολής της παραμέτρου ω_{k+1}, το παρόν σχήμα (4.16) ανήκει σε μια γενικότερη κατηγορία σχημάτων που καλούνται "μή στατικά" "δεύτερης τάξης" σε αντίθεση με τα μέχρι τώρα γνωστά που καλούνται για προφανείς λόγους "στατικά" "πρώτης τάξης".

Τέλος, για την εύρεση της ταχύτητας σύγκλισης μετά από (κάθε) k επαναλήψεις κάτω από τις μέχρι τώρα αρχικές προϋποθέσεις, ότι ο πίνακας T είναι Ερμιτιανός και ότι είναι γνωστή μόνο η φασματική ακτίνα $\rho = \rho(T)$ (< 1) του αρχικού επαναληπτικού σχήματος, έχουμε από τη θεωρία που αναπτύχτηκε ότι

$$||\widetilde{p}_k(T)||_2 \;=\; \frac{1}{T_k\left(\frac{1}{\rho}\right)},\; k \geq 0.$$

Επειδή $\frac{1}{\rho} > 1$, από την ιδιότητα (γ) των πολυωνύμων του Chebyshev, αντικαθιστώντας στην παραπάνω σχέση, έχουμε

$$||\widetilde{p}_k(T)||_2 \;=\; 2\left[\left(\frac{\rho}{1+\sqrt{1-\rho^2}}\right)^k + \left(\frac{1+\sqrt{1-\rho^2}}{\rho}\right)^k\right]^{-1}, \tag{4.20}$$

που, όπως αναφέρθηκε, δίνει διαμέσου της (4.6) τη μεγαλύτερη δυνατή ταχύτητα σύγκλισης μετά από (κάθε) k επαναλήψεις.

Κλείνοντας το παρόν κεφάλαιο θα εξετάσουμε την περίπτωση κατά την οποία ο T, πέρα από Ερμιτιανός, είναι ο επαναληπτικός πίνακας του Jacobi, που αντιστοιχεί στον $I-T$, με $\rho(T)<1$, και επιπλέον ο $I-T$ είναι δικυκλικός και συνεπώς διατεταγμένος, όπως δηλαδή μπορεί να συμβαίνει στην περίπτωση που αναφέρθηκε λίγο πριν στην Παρατήρηση (γ) και που αφορούσε το όριο στο οποίο έτεινε η παράμετρος της ημι-επαναληπτικής μεθόδου. Τότε, η έκφραση (4.20) μπορεί να δοθεί συντομογραφικά αν χρησιμοποιηθεί η αντίστοιχη έκφραση για την overrelaxation παράμετρο της βέλτιστης SOR, που είναι η $\omega_\beta = \frac{2}{1+\sqrt{1-\rho^2(T)}}$. Θεωρώντας την έκφραση $\frac{\rho}{1+\sqrt{1-\rho^2}}$, που εμπλέκεται στην (4.20), παίρνουμε διαδοχικά

$$\frac{\rho}{1+\sqrt{1-\rho^2}} = \left(\frac{\rho^2}{(1+\sqrt{1-\rho^2})^2}\right)^{\frac{1}{2}} = \left(\frac{1-\sqrt{1-\rho^2}}{1+\sqrt{1-\rho^2}}\right)^{\frac{1}{2}} = (\omega_\beta - 1)^{\frac{1}{2}}.$$

Με βάση το παραπάνω αποτέλεσμα η σχέση (4.20) μπορεί να γραφτεί ως

$$||\widetilde{p}_k(T)||_2 \;=\; (\omega_\beta - 1)^{\frac{k}{2}} \left[\frac{2}{1+(\omega_\beta-1)^k}\right]. \tag{4.21}$$

Εχοντας υπόψη την (4.21), η μέση **ασυμπτωτική** ταχύτητα σύγκλισης της μεθόδου στην προκείμενη περίπτωση θα βρίσκεται με βάση και την (4.6) ως εξής:

$$\begin{aligned}\mathcal{R}_\infty(\widetilde{p}_k(T)) &= \lim_{k\to\infty}\mathcal{R}(\widetilde{p}_k(T)) = -\lim_{k\to\infty}\ln(||\widetilde{p}_k(T)||_2)^{\frac{1}{k}} \\ &= -\lim_{k\to\infty}\left[\tfrac{1}{2}\ln(\omega_\beta-1) + \tfrac{1}{k}\ln\left(\tfrac{2}{1+(\omega_\beta-1)^k}\right)\right] = -\ln(\omega_\beta-1)^{\frac{1}{2}}.\end{aligned}$$

Σημείωση: Οπως καθίσταται φανερό, στην περίπτωση που μόλις εξετάστηκε, η μέση ασυμπτωτική ταχύτητα της ημι-επαναληπτικής μεθόδου Chebyshev είναι το μισό της αντίστοιχης της βέλτιστης SOR!

ΑΣΚΗΣΕΙΣ:

1.: Να αποδειχτούν οι ιδιότητες (α)–(στ) των πολυωνύμων του Chebyshev, που βασίζονται στον Ορισμό (4.8).

2.: Να γίνουν τρεις επαναλήψεις της Chebyshev μεθόδου για τη λύση του συστήματος $Ax=b$, με $A=\mathrm{trid}(-1,2,-1)\in I\!R^{3,3}$, και $b=[1\ 0\ 1]^T$, χρησιμοποιώντας ως αρχική μέθοδο (μέθοδο βάσης) τη μέθοδο Jacobi και αρχικό διάνυσμα $x^{(0)}=0$.

5 Θεωρητικές Εφαρμογές των Επαναληπτικών Μεθόδων

5.1 Εισαγωγή

Στο παρόν Κεφάλαιο εισάγονται μερικές απλές έννοιες και δίνονται μερικές προτάσεις από τη θεωρία των εξισώσεων διαφορών καθώς και από αυτή των τανυστικών γινομένων. Οπως θα καταστεί φανερό, χρησιμοποιώντας τις έννοιες και τις προτάσεις αυτές, σε ορισμένες απλές περιπτώσεις αριθμητικής επίλυσης με επαναληπτικές μεθόδους προβλημάτων-μοντέλων (ελλειπτικών μερικών) διαφορικών εξισώσεων, που παρουσιάζονται συχνά στην πράξη, είναι δυνατόν να βρίσκονται αναλυτικές εκφράσεις για τις ιδιοτιμές, και μερικές φορές και για τα ιδιοδιανύσματα, πινάκων, που εμπλέκονται σ΄ αυτές, καθώς επίσης και να γίνονται συγκρίσεις σ΄ ό,τι αφορά τη μέση ασυμπτωτική ταχύτητα σύγκλισης των επαναληπτικών αυτών μεθόδων.

5.2 Εξισώσεις Διαφορών

Ορισμός 5.1 *Εξίσωση διαφορών τάξης n καλείται κάθε εξίσωση της μορφής*

$$f(y(k+n), y(k+n-1), \cdots, y(k+1), y(k)) = 0,$$

με άγνωστη μία συνάρτηση $y := y_k \equiv y(k)$ της ακέραιας μεταβλητής k, και η οποία αληθεύει γενικά για κάθε τιμή της $k \in \boldsymbol{Z}$.

Ορισμός 5.2 *Μια εξίσωση διαφορών τάξης n καλείται γραμμική, ομογενής, με σταθερούς συντελεστές, όταν είναι της μορφής*

$$a_n y_{k+n} + a_{n-1} y_{k+n-1} + \cdots + a_1 y_{k+1} + a_0 y_k = 0, \tag{5.1}$$

όπου $a_i \in \mathbb{C}$ σταθερές με $a_n a_0 \neq 0$ και αληθεύει $\forall\ k = 0, \pm 1, \pm 2, \cdots$.

Για την εύρεση των λύσεων της (5.1), εκτός της τετριμμένης $y \equiv 0$, θεωρούμε την πολυωνυμική εξίσωση

$$p(r) := a_n r^n + a_{n-1} r^{n-1} + \cdots + a_1 r + a_0 = 0,$$

η οποία καλείται και "χαρακτηριστική" της (5.1), βρίσκουμε τις ρίζες της, έστω $r_1, r_2, \cdots, r_m$, αντίστοιχων πολλαπλοτήτων $n_1, n_2, \cdots, n_m$ με $n_1 + n_2 + \cdots + n_m = n$, οπότε η καλούμενη "γενική" λύση της (5.1) δίνεται από την έκφραση

$$y(k) = \sum_{i=1}^{m} \left(\sum_{j=0}^{n_i - 1} c_{ij} k^j \right) r_i^k, \tag{5.2}$$

όπου $c_{ij} \in \mathbb{C},\ j = 0(1)n_i - 1, i = 1(1)m$, αυθαίρετες σταθερές. Οι σταθεροί συντελεστές στην (5.2) προσδιορίζονται μονοσήμαντα και μάλιστα εύκολα σε ειδικές περιπτώσεις, π.χ. αν οι ρίζες της χαρακτηριστικής εξίσωσης είναι διακεκριμένες και δίνονται οι τιμές της $y(k)$ για n διαδοχικές τιμές του k, οπότε οδηγούμαστε σε "ειδικές" λύσεις.

Σα μία εφαρμογή των μέχρι τώρα αναφερθέντων θεωρούμε την αριθμητική επίλυση της παρακάτω απλής διαφορικής εξίσωσης με συνοριακές συνθήκες. Εστω $\Omega = (0,1)$ το ανοικτό μοναδιαίο διάστημα της πραγματικής ευθείας και η διαφορική εξίσωση

$$-\frac{d^2u(x)}{dx^2} = f(x), \quad ορισμένη\ στο\ \ \Omega, \quad με\ \ u(0) = \alpha,\ u(1) = \beta, \tag{5.3}$$

όπου $f(x)$ γνωστή συνάρτηση. Για την αριθμητική επίλυση της (5.3) θεωρούμε μια ομοιόμορφη διαμέριση του διαστήματος $[0,1]$ σε $n+1,\ n \geq 3$, ίσα υποδιαστήματα μήκους $h = \frac{1}{n+1}$. Εστω $x_i = ih,\ i = 0(1)n+1$, τα σημεία της διαμέρισης μαζί με τα άκρα. Στη συνέχεια διακριτοποιούμε το συνεχές πρόβλημα (5.3) προσεγγίζοντας τη δεύτερη παράγωγο στα σημεία της διαμέρισης με κεντρικές διαφορές δεύτερης τάξης. Συγκεκριμένα

$$\frac{d^2u(x_i)}{dx^2} \approx \frac{u(x_{i-1}) - 2u(x_i) + u(x_{i+1})}{h^2}, \quad i = 1(1)n, \tag{5.4}$$

όπου το τοπικό σφάλμα αποκοπής σε κάθε μία από τις εξισώσεις της (5.4) είναι $\mathcal{O}(h^2)$. Αν u_i είναι οι προσεγγιστικές τιμές των $u(x_i)$ και $f_i = f(x_i),\ i = 1(1)n$, τότε χρησιμοποιώντας την (5.3) καταλήγουμε στο γραμμικό σύστημα

$$\begin{gathered} Au = c, \\ A = \text{trid}(-1,2,-1) \in \mathbb{R}^{n,n},\ u = [u_1\ u_2\ \cdots\ u_n]^T\ \kappa\alpha\iota\ c = [h^2f_1 + \alpha,\ h^2f_2,\ \cdots,\ h^2f_n + \beta]^T, \end{gathered} \tag{5.5}$$

που αποτελεί το διακριτό ανάλογο του συνεχούς προβλήματος (5.3) και ο συμβολισμός $\text{trid}(a,b,c)$ χρησιμοποιείται για να δηλώσει έναν τριδιαγώνιο πίνακα με διαγώνια στοιχεία ίσα με b, κάτω από τη διαγώνιο στοιχεία ίσα με a και πάνω από τη διαγώνιο στοιχεία ίσα με c.

Είναι εύκολο να διαπιστωθεί ότι ο πίνακας των συντελεστών A του συστήματος (5.5) είναι πραγματικός, συμμετρικός, θετικά ορισμένος και δικυκλικός και συνεπώς διατεταγμένος. Επομένως όλες οι άμεσες και οι επαναληπτικές μέθοδοι επίλυσης γραμμικών συστημάτων μπορούν να εφαρμοστούν. Δηλαδή, οι μέθοδοι απαλοιφής Gauss (ή LU παραγοντοποίησης), και μάλιστα χωρίς οδήγηση, Cholesky, καθώς και οι επαναληπτικές μέθοδοι Jacobi , Gauss-Seidel, SOR, SSOR και η ημι-επαναληπτική μέθοδος Chebyshev.
Ενα-δυο σημεία που αφορούν στη σύγκλιση των επαναληπτικών μεθόδων θα πρέπει να τονιστούν ιδιαίτερα. Καταρχάς, επειδή ο A είναι πραγματικός, συμμετρικός και θετικά ορισμένος, η SOR και η SSOR μέθοδοι συγκλίνουν, για $\omega \in \mathbb{R}$, όπου και θα περιοριοριστούμε στο παρόν Κεφάλαιο, για κάθε $\omega \in (0,2)$. Αφού η SOR συγκλίνει για $\omega \in (0,2)$ θα συγκλίνει και η Gauss-Seidel ($\omega = 1$). Λόγω της δικυκλικής και συνεπώς διατεγμένης ιδιότητας του A κι εφόσον η Gauss-Seidel συγκλίνει

θα συγκλίνει και η Jacobi αφού η φασματική ακτίνα του επαναληπτικού πίνακα της Gauss-Seidel είναι το τετράγωνο της φασματικής ακτίνας του αντίστοιχου πίνακα της Jacobi. Τέλος, η ημιεπαναληπτική μέθοδος Chebyshev αν βασιστεί στην επαναληπτική μέθοδο του Jacobi θα συγκλίνει αφού, όπως μπορεί να διαπιστωθεί, ικανοποιούνται όλες οι προϋποθέσεις που αναφέρθηκαν στο προηγούμενο κεφάλαιο.

Για την εύρεση φασματικών ακτίνων και σύγκριση μέσων (ασυμπτωτικών) ταχυτήτων σύγκλισης των επαναληπτικών μεθόδων απαιτείται η γνώση των ιδιοτιμών του επαναληπτικού πίνακα του Jacobi που αντιστοιχεί στον A. Λόγω των ιδιοτήτων του πίνακα A η εύρεση των ιδιοτιμών του θα οδηγήσει στη συνέχεια στην εύρεση και όλων των άλλων στοιχείων που τυχόν μας ενδιαφέρουν.

Για τον πίνακα A της (5.5), αν λ είναι μια ιδιοτιμή του και $x = [x_1 x_2 \cdots x_n]^T$ το αντίστοιχο ιδιοδιάνυσμα θα ισχύει ότι

$$Ax = \lambda x, \tag{5.6}$$

όπου $\lambda \in \mathbb{C}$ και $x \in \mathbb{C}^n \setminus \{0\}$. Οι ιδιοτιμές λ, όμως, είναι πραγματικές και θετικές, αφού ο A είναι πραγματικός, συμμετρικός και θετικά ορισμένος, το δε ιδιοδιάνυσμα x μπορεί να παρθεί πραγματικό. Επιπλέον ισχύει ότι $\lambda \leq ||A||_\infty = 4$. Αν δε σχηματιστεί ο πίνακας $4I - A$ είναι δυνατόν να διαπιστωθεί ότι είναι κι αυτός θετικά ορισμένος και άρα $0 < \lambda < 4$. Εξισώνοντας τώρα τις i-οστές συνιστώσες των δύο μελών της (5.6) παίρνουμε

$$x_{i+1} - (2 - \lambda)x_i + x_{i-1} = 0, \ i = 1(1)n, \quad \mu\epsilon \ x_0 = x_{n+1} = 0. \tag{5.7}$$

Η (5.7) δεν είναι παρά μία ομογενής, γραμμική εξίσωση διαφορών δεύτερης τάξης με σταθερούς συντελεστές, που μπορεί να επεκταθεί και να ισχύει για κάθε τιμή του i αν ορίσουμε αυθαίρετα το x_{i-1} να επαληθεύει την (5.7) για κάθε $i = 0, -1, -2, \cdots$ και το x_{i+1} να επαληθεύει την ίδια εξίσωση για $i = n+1, n+2, n+3, \cdots$. Για την επίλυση της (5.7), έστω r_1, r_2 οι ρίζες της χαρακτηριστικής της

$$r^2 - (2 - \lambda)r + 1 = 0, \tag{5.8}$$

για τις οποίες ισχύουν

$$r_1 + r_2 = 2 - \lambda \ \ \kappa\alpha\iota \ \ r_1 r_2 = 1. \tag{5.9}$$

Οι r_1, r_2 είναι μιγαδικές συζυγείς (άρα $r_1 \neq r_2$), αφού η διακρίνουσα της (5.8) είναι ίση με $(2 - \lambda)^2 - 4 = \lambda(\lambda - 4) < 0$. Επομένως η γενική λύση της (5.8) δίνεται από τη

$$x_i = c_1 r_1^i + c_2 r_2^i. \tag{5.10}$$

Χρησιμοποιώντας τις συνοριακές συνθήκες $x_0 = x_{n+1} = 0$, παίρνουμε, λόγω της (5.10), ότι

$$c_1 + c_2 = 0, \quad c_1 r_1^{n+1} + c_2 r_2^{n+1} = 0. \tag{5.11}$$

Απαλείφοντας τις σταθερές c_1, c_2 από τις (5.11) και έχοντας υπόψη την $r_1 r_2 = 1$, από τις (5.9), προκύπτει ότι

$$r_1^{2(n+1)} = 1 = \cos(2k\pi) + \mathrm{i}\sin(2k\pi),$$

όπου $k \in \boldsymbol{Z}$ οποιοσδήποτε, οπότε με εφαρμογή του τύπου του De Moivre έχουμε

$$r_1 = \cos(\frac{k\pi}{n+1}) + \mathrm{i}\sin(\frac{k\pi}{n+1}),$$

όπου $k \in \boldsymbol{Z}$ οποιοσδήποτε. Από την $r_1 r_2 = 1$ προκύπτει η αντίστοιχη έκφραση για την r_2, δηλαδή

$$r_2 = \cos(\frac{k\pi}{n+1}) - \mathrm{i}\sin(\frac{k\pi}{n+1}),$$

όπου $k \in \boldsymbol{Z}$ οποιοσδήποτε. Τέλος, από την πρώτη των (5.9) και τις δύο προηγούμες εκφράσεις βρίσκουμε ότι

$$\lambda = 2\left(1 - \cos(\frac{k\pi}{n+1})\right) = 4\sin^2\left(\frac{k\pi}{2(n+1)}\right),$$

όπου $k \in \boldsymbol{Z}$ οποιοσδήποτε. Αν παρατηρήσουμε ότι για $k = 1(1)n$ βρίσκονται n διαφορετικές τιμές του λ συμπεραίνουμε ότι όλες οι ιδιοτιμές του πίνακα A δίνονται από τις εκφράσεις

$$\lambda_k = 4\sin^2\left(\frac{k\pi}{2(n+1)}\right), \quad k = 1(1)n. \tag{5.12}$$

Είναι αξιοσημείωτο ότι μπορούν να βρεθούν και τα αντίστοιχα ιδιοδιανύσματα του A αν χρησιμοποιηθούν οι εκφράσεις του γενικού όρου της εξίσωσης διαφορών (5.10). Συγκεκριμένα,

$$x_i = c_1\left(\cos\left(\frac{k\pi}{n+1}\right) + \mathrm{i}\sin\left(\frac{k\pi}{n+1}\right)\right)^i + c_2\left(\cos\left(\frac{k\pi}{n+1}\right) - \mathrm{i}\sin\left(\frac{k\pi}{n+1}\right)\right)^i$$

και επειδή $c_2 = -c_1$, παίρνουμε ότι

$$x_i = 2\mathrm{i}c_1 \sin\left(\frac{ik\pi}{n+1}\right), \quad i = 1(1)n,\ k = 1(1)n.$$

Ο $2\mathrm{i}c_1$ ως κοινός παράγοντας όλων των συνιστωσών του k-οστού ιδιοδιανύσματος μπορεί να παραλειφτεί, οπότε τελικά, το ιδιοδιάνυσμα που αντιστοιχεί στην ιδιοτιμή $\lambda_k,\ k = 1(1)n$, είναι το

$$x^{(k)} = \left[\sin\left(\frac{k\pi}{n+1}\right)\ \sin\left(\frac{2k\pi}{n+1}\right)\ \cdots\ \sin\left(\frac{nk\pi}{n+1}\right)\right]^T, \quad k = 1(1)n. \tag{5.13}$$

Για την εύρεση των φασματικών ακτίνων των επαναληπτικών πινάκων των (point) επαναληπτικών μεθόδων που έχουμε μελετήσει ως τώρα έχουμε:

Jacobi: Από το δοθέντα πίνακα A βρίσκουμε αμέσως ότι ο επαναληπτικός πίνακας της μεθόδου είναι ο $T_J = \frac{1}{2}\mathrm{trid}(1, 0, 1)$. Οι ιδιοτιμές του T_J βρίσκονται εύκολα γιατί στην προκείμενη περίπτωση έχουμε $T_J = \frac{1}{2}(2I - \mathrm{trid}(-1, 2, -1)) = I - \frac{1}{2}A$. Αρα οι ιδιοτιμές του T_J θα είναι ίσες με $\mu_k =$

$1-\frac{1}{2}\lambda_k = 1-\frac{1}{2}4\sin^2(\frac{k\pi}{2(n+1)}) = \cos(\frac{k\pi}{n+1})$ και η φασματική ακτίνα της μεθόδου Jacobi θα δίνεται από τη

$$\rho(T_J) \;=\; \cos(\frac{\pi}{n+1}).$$

Gauss-Seidel: Οπως αναφέρθηκε, ο A ως τριδιαγώνιος είναι δικυκλικός και συνεπώς διατεταγμένος. Επομένως η φασματική ακτίνα του επαναληπτικού πίνακα, T_{GS}, της μεθόδου θα δίνεται από την έκφραση

$$\rho(T_{GS}) \;=\; \cos^2(\frac{\pi}{n+1}).$$

Βέλτιστη SOR: Επειδή ο A είναι δικυκλικός και συνεπώς διατεταγμένος και ο πίνακας Jacobi που αντιστοιχεί σ΄ αυτόν έχει πραγματικές ιδιοτιμές με $\rho(T_J) < 1$ είναι δυνατόν να βρεθεί μια βέλτιστη τιμή για την overrelaxation παράμετρο ω. Οπως είναι γνωστό αυτή δίνεται από τον τύπο $\omega_\beta = \frac{2}{1+\sqrt{1-\rho^2(T_J)}}$. Για την αντίστοιχη φασματική ακτίνα της SOR θα έχουμε

$$\rho(\mathcal{L}_{\omega_\beta}) = \omega_\beta - 1 = \frac{2}{1+\sqrt{1-\rho^2(T_J)}} - 1 = \frac{2}{1+\sqrt{1-\cos^2(\frac{\pi}{n+1})}} - 1 = \frac{1-\sin(\frac{\pi}{n+1})}{1+\sin(\frac{\pi}{n+1})}. \quad (5.14)$$

Βέλτιστη SSOR: Σύμφωνα με τη θεωρία που αναπτυχτήκε στο αντίστοιχο κεφάλαιο εφόσον ο A είναι δικυκλικός και συνεπώς διατεταγμένος με ιδιοτιμές του επαναληπτικού πίνακα του Jacobi πραγματικές με μέτρο μικρότερο από τη μονάδα υπάρχει βέλτιστη SSOR μέθοδος, αντιστοιχεί στην τιμή $\omega_\beta = 1$, και η φασματική ακτίνα του επαναληπτικού της πίνακα, $\rho(\mathcal{S}_{\omega_\beta})$, είναι ίση με αυτήν του επαναληπτικού πίνακα της Gauss-Seidel. Δηλαδή

$$\rho(\mathcal{S}_{\omega_\beta}) = \rho(T_{GS}) = \cos^2(\frac{\pi}{n+1}).$$

Chebyshev: Για την ημι-επαναληπτική μέθοδο του Chebyshev, που βασίζεται στην επαναληπτική μέθοδο του Jacobi έχουμε ότι ο T_J είναι συμμετρικός με ιδιοτιμές στο κλειστό διάστημα $[-\rho(T_J), \rho(T_J)]$, όπου $\rho(T_J) < 1$, και επιπλέον ο $I - T$ είναι δικυκλικός και συνεπώς διατεταγμένος. Επομένως, σύμφωνα με τη θεωρία που αναπτύχτηκε στο προηγούμενο κεφάλαιο, η μέση ταχύτητα σύγκλισης μετά από κάθε m επαναλήψεις θα είναι

$$-\ln\left((\omega_\beta-1)^{\frac{1}{2}}\left[\frac{2}{1+(\omega_\beta-1)^m}\right]^{\frac{1}{m}}\right),$$

όπου $\omega_\beta - 1$ η έκφραση που δίνεται στην (5.14), η δε μέση ασυμπτωτική ταχύτητα σύγκλισης θα δίνεται από την

$$-\ln(\omega_\beta-1)^{\frac{1}{2}},$$

και θα είναι ίση με το μισό της βέλτιστης SOR μεθόδου.

Η σύγκριση των μέσων ασυμπτωτικών ταχυτήτων σύγκλισης για μερικές από τις παραπάνω μεθόδους, στη συγκεκριμένη περίπτωση που εξετάζεται, είναι γνωστή. Π.χ., αυτή της Gauss-Seidel είναι διπλάσια της αντίστοιχης της Jacobi, της βέλτιστης SOR είναι μεγαλύτερη αυτής της Gauss-Seidel, της βέλτιστης SSOR είναι ίδια με την αντίστοιχη της Gauss-Seidel και της ημιεπαναληπτικής μεθόδου του Chebyshev είναι το μισό αυτής της βέλτιστης SOR. Η μόνη σύγκριση που μπορεί να γίνει παραπέρα είναι η συγκεκριμενοποίηση της σχέσης μεταξύ μέσων ασυμπτωτικών ταχυτήτων σύγκλισης των μεθόδων Gauss-Seidel και βέλτιστης SOR και ό,τι αυτή αμέσως συνεπάγεται. Η σύγκριση αυτή θα γίνει για προβλήματα που ενδιαφέρουν στην πράξη και άρα για "μεγάλα" n ή θεωρητικά για $n \to \infty$, που είναι ισοδύναμο με το $h \to 0^+$.

Καταρχάς βρίσκουμε τις ισοδύναμες προς τις $\mathcal{R}_\infty(\mathcal{L}_{\omega_\beta})$ και $\mathcal{R}_\infty(T_{GS})$ εκφράσεις. Για την πρώτη, έχουμε από την (5.14) ότι

$$\mathcal{R}_\infty(\mathcal{L}_{\omega_\beta}) = -\ln \rho(\mathcal{L}_{\omega_\beta}) = -\ln(\omega_\beta - 1) = -\ln\left(\frac{1-\sin(\pi h)}{1+\sin(\pi h)}\right),$$

αν δε χρησιμοποιήσουμε διαδοχικά τις συγκλίνουσες σειρές $\sin x = x - \frac{x^3}{3!} + \mathcal{O}(x^5)$ καθώς και την $\ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} + \mathcal{O}(x^4)$, για $x \to 0$, μπορεί να προκύψει ότι

$$\mathcal{R}_\infty(\mathcal{L}_{\omega_\beta}) = 2\pi h + \mathcal{O}(h^3). \tag{5.15}$$

Για τη δεύτερη έκφραση, εργαζόμενοι με τον ίδιο τρόπο και χρησιμοποιώντας και τη σειρά $\cos x = 1 - \frac{x^2}{2!} + \mathcal{O}(x^4)$, προκύπτει ότι

$$\mathcal{R}_\infty(T_{GS}) = -\ln(\rho(T_{GS})) = -\ln(\cos^2(\pi h)),$$

οπότε βρίσκουμε

$$\mathcal{R}_\infty(T_{GS}) = \pi^2 h^2 + \mathcal{O}(h^4). \tag{5.16}$$

Από τις (5.15) και (5.16) παρατηρούμε αμέσως ότι η μέση ασυμπτωτική ταχύτητα σύγκλισης της βέλτιστης SOR είναι κατά μία τάξη μεγέθους, ως προς h, μεγαλύτερη από αυτήν της Gauss-Seidel, ο δε λόγος της πρώτης προς τη δεύτερη δίνει επιπλέον και τον ασυμπτωτικό παράγοντα $\frac{2}{\pi}\frac{1}{h}$. Βασιζόμενοι στο τελευταίο αποτέλεσμα είναι δυνατόν να γίνουν αμέσως οι συγκρίσεις των μεθόδων, οι οποίες, στην ανάλυση που προηγήθηκε, δεν έχουν συγκριθεί μεταξύ τους.

5.3 Τανυστικά Γινόμενα

Αρχίζουμε την παρούσα παράγραφο με τον ορισμό του τανυστικού γινομένου δύο πινάκων.

Ορισμός 5.3 *Εστω οι πίνακες* $A \in \mathbb{C}^{m,n}$ *και* $B \in \mathbb{C}^{p,q}$. *Τανυστικό γινόμενο (ή γινόμενο* Kronecker*) των πινάκων* A *επί* B, *με τη σειρά αυτή, καλείται ο πίνακας* $C \in \mathbb{C}^{mp,nq}$, *που, συμβολίζεται ως* $A \otimes B$ *και, είναι τ.ω.*

$$C \;=\; A \otimes B \;=\; \begin{bmatrix} a_{11}B & a_{12}B & \cdots & a_{1n}B \\ a_{21}B & a_{22}B & \cdots & a_{2n}B \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1}B & a_{m2}B & \cdots & a_{mn}B \end{bmatrix}. \tag{5.17}$$

Με βάση τον ορισμό, που μόλις δόθηκε, είναι εύκολο να αποδειχτεί ότι ισχύουν οι παρακάτω ιδιότητες:

α) Αν $A \in \mathbb{C}^{m,n}$ και $B \in \mathbb{C}^{p,q}$, τότε $(A \otimes B)^H = A^H \otimes B^H$.

β) Αν $A \in \mathbb{C}^{m,n}$, $B \in \mathbb{C}^{p,q}$, $C \in \mathbb{C}^{k,l}$ και $D \in \mathbb{C}^{r,s}$ ισχύει ότι

$$(A \otimes B)(C \otimes D) \;=\; (AC) \otimes (BD),$$

με την προϋπόθεση ότι τα γινόμενα AC και BD ορίζονται, δηλαδή ισχύει ότι $n = k$ και $q = r$.

γ) Αν $A \in \mathbb{C}^{m,m}$ και $B \in \mathbb{C}^{n,n}$ είναι αντιστρέψιμοι, τότε και ο $A \otimes B$ είναι αντιστρέψιμος και ισχύει ότι

$$(A \otimes B)^{-1} \;=\; A^{-1} \otimes B^{-1}.$$

Για τις ιδιοτιμές και τα ιδιοδιανύσματα του τανυστικού γινομένου δύο τετραγωνικών πινάκων ισχύουν ακόμη και οι ακόλουθες δύο ιδιότητες, που μπορούν να αποδειχτούν με βάση τον ορισμό του τανυστικού γινομένου και αυτές που ήδη αναφέρθηκαν. Συγκεκριμένα:

δ) Αν $A \in \mathbb{C}^{m,m}$ και $\lambda \in \sigma(A)$ με $x \in \mathbb{C}^m \setminus \{0\}$ το αντίστοιχο ιδιοδιάνυσμα και $B \in \mathbb{C}^{n,n}$ και $\mu \in \sigma(B)$ με $y \in \mathbb{C}^n \setminus \{0\}$ το αντίστοιχο ιδιοδιάνυσμα, τότε $\lambda\mu \in \sigma(A \otimes B)$ με αντίστοιχο ιδιοδιάνυσμα το $x \otimes y \in \mathbb{C}^{mn} \setminus \{0\}$.

ε) Αν λ_i είναι οι ιδιοτιμές του $A \in \mathbb{C}^{m,m}$ με $x^{(i)} \in \mathbb{C}^m \setminus \{0\}$, $i = 1(1)m$, τα αντίστοιχα ιδιοδιανύσματα, και συμβαίνει να είναι γραμμικά ανεξάρτητα, και μ_j είναι οι ιδιοτιμές του $B \in \mathbb{C}^{n,n}$ με $y^{(j)} \in \mathbb{C}^n \setminus \{0\}$, $j = 1(1)n$, και συμβαίνει να είναι επίσης γραμμικά ανεξάρτητα, τότε $\lambda_i\mu_j$ είναι οι ιδιοτιμές του $A \otimes B$ με αντίστοιχα ιδιοδιανύσματα τα $x^{(i)} \otimes y^{(j)} \in \mathbb{C}^{mn} \setminus \{0\}$, $i = 1(1)m$, $j = 1(1)n$, που μπορεί να αποδειχτεί ότι είναι επίσης γραμμικά ανεξάρτητα.

Σαν εφαρμογή των παραπάνω εκτεθέντων και συγχρόνως των επαναληπτικών μεθόδων των προηγούμενων κεφαλαίων δίνουμε την επέκταση της εφαρμογής της προηγούμενης παραγράφου στις δύο διαστάσεις που αποτελεί το γνωστό πρόβλημα-μοντέλο της εξίσωσης του Poisson υπό συνοριακές συνθήκες Dirichlet.

Εστω η εξίσωση

$$-\frac{\partial^2 u}{\partial x^2} - \frac{\partial^2 u}{\partial y^2} \;=\; f(x,y),\ \textit{ορισμένη στο}\ \Omega = (0,1) \times (0,1), \tag{5.18}$$

με

$$u(x,y) \;=\; g(x,y),\ στο\ \partial\Omega, \tag{5.19}$$

όπου $f(x,y)$ και $g(x,y)$ γνωστές συναρτήσεις.

Για την αριθμητική επίλυση της ελλειπτικής μερικής διαφορικής εξίσωσης (5.18)–(5.19) επιθέτουμε στο $\Omega \cup \partial\Omega$ ένα ομοιόμορφο δικτυωτό με n εσωτερικά σημεία διαμέρισης στις $x-$ και $y-$διευθύνσεις, οπότε το βήμα της διαμέρισης είναι $h = \frac{1}{n+1}$. Κάθε κόμβος του δικτυωτού με συντεταγμένες $(x_i, y_j) = (ih, jh),\ i,j = 0(1)n+1$, χαρακτηρίζεται από το ζεύγος (i,j). Προσεγγίζουμε την (5.18)–(5.19) στα εσωτερικά σημεία του δικτυωτού $(i,j),\ i,j = 1(1)n$, με κεντρικές διαφορές δεύτερης τάξης, οπότε προκύπτουν n^2 εξισώσεις, κάθε μία από τις οποίες παρουσίαζει ένα τοπικό σφάλμα αποκοπής της τάξης $\mathcal{O}(h^2)$. Τις εξισώσεις αυτές τις κατατάσσουμε στη λεγόμενη φυσική τους διάταξη, θεωρώντας τα σημεία του δικτυωτού, στα οποία αντιστοιχούν, πρώτα από αριστερά προς τα δεξιά και ύστερα από κάτω προς τα πάνω. Κάθε μία από τις προαναφερθείσες εξισώσεις έχει τη μορφή

$$-u_{i-1,j} + 2u_{ij} - u_{i+1,j} - u_{i,j-1} + 2u_{ij} - u_{i,j+1} \;=\; h^2 f_{ij},\quad i = 1(1)n,\ j = 1(1)n, \tag{5.20}$$

όπου u_{ij} είναι η προσεγγιστική τιμή της $u(x_i, y_j)$ και $f_{ij} = f(x_i, y_j)$, η δε συλλογή όλων των εξισώσεων (5.20), με τη σειρά που προαναφέρθηκε, οδηγεί σε ένα γραμμικό σύστημα της γενικής μορφής

$$Au \;=\; c. \tag{5.21}$$

Τα στοιχεία του συστήματος (5.21), όταν αυτά γραφτούν σε block μορφή, δίνονται αναλυτικά αμέσως παρακάτω.

$$A \;=\; \left[\begin{array}{ccccc|ccccc|ccccc|ccccc}
4 & -1 & & & & -1 & & & & & & & & & & & & & & \\
-1 & 4 & -1 & & & & -1 & & & & & & & & & & & & & \\
 & \ddots & \ddots & \ddots & & & & \ddots & & & & & & & & & & & & \\
 & & -1 & 4 & -1 & & & & -1 & & & & & & & & & & & \\
 & & & -1 & 4 & & & & & -1 & & & & & & & & & & \\
\hline
-1 & & & & & 4 & -1 & & & & -1 & & & & & & & & & \\
 & -1 & & & & -1 & 4 & -1 & & & & -1 & & & & & & & & \\
 & & \ddots & & & & \ddots & \ddots & \ddots & & & & \ddots & & & & & & & \\
 & & & -1 & & & & -1 & 4 & -1 & & & & -1 & & & & & & \\
 & & & & -1 & & & & -1 & 4 & & & & & -1 & & & & & \\
\hline
 & & & & & & & & & & & & \ddots & & & & & & & \\
\hline
 & & & & & & & & & & -1 & & & & & 4 & -1 & & & \\
 & & & & & & & & & & & -1 & & & & -1 & 4 & -1 & & \\
 & & & & & & & & & & & & \ddots & & & & \ddots & \ddots & \ddots & \\
 & & & & & & & & & & & & & -1 & & & & -1 & 4 & -1 \\
 & & & & & & & & & & & & & & -1 & & & & -1 & 4
\end{array}\right] \tag{5.22}$$

και

$$u = [u_{11}\ u_{21}\ \cdots\ u_{n1}|u_{12}\ u_{22}\ \cdots\ u_{n2}|\cdots|u_{1n}\ u_{2n}\ \cdots\ u_{nn}]^T,$$
$$c = [c_{11}\ c_{21}\ \cdots\ c_{n1}|c_{12}\ c_{22}\ \cdots\ c_{n2}|\cdots|c_{1n}\ c_{2n}\ \cdots\ c_{nn}]^T,$$

όπου

$$c_{ij} = \begin{cases} h^2 f_{ij}, & \forall\ i,j = 2(1)n-1, \\ h^2 f_{ij} + g(0,jh), & \forall\ i = 1, j = 2(1)n-1, \\ h^2 f_{ij} + g(1,jh), & \forall\ i = n, j = 2(1)n-1, \\ h^2 f_{ij} + g(ih,0), & \forall\ i = 2(1)n-1, j = 1, \\ h^2 f_{ij} + g(ih,1), & \forall\ i = 2(1)n-1, j = n, \\ h^2 f_{ij} + g(0,h) + g(h,0), & (i,j) = (1,1), \\ h^2 f_{ij} + g(nh,0) + g(1,h), & (i,j) = (n,1), \\ h^2 f_{ij} + g(1,nh) + g(nh,1), & (i,j) = (n,n), \\ h^2 f_{ij} + g(0,nh) + g(h,1), & (i,j) = (1,n). \end{cases}$$

Για την εύρεση των ιδιοτιμών, και ιδιοδιανυσμάτων, του πίνακα A της (5.22), παρατηρούμε ότι αυτός έχει άμεση σχέση με τον πίνακα $A = \text{trid}(-1,2,-1)$ της μονοδιάστατης εφαρμογής στην (5.5). Συγκεκριμένα, αν $T = \text{trid}(-1,2,-1) \in I\!R^{n,n}$, τότε ο πίνακας A στην παρούσα περίπτωση μπορεί να γραφτεί, σε block μορφή, συνοπτικά ως εξής

$$A = \begin{bmatrix} 2I+T & -I & & & \\ -I & 2I+T & -I & & \\ & \ddots & \ddots & \ddots & \\ & & -I & 2I+T & -I \\ & & & -I & 2I+T \end{bmatrix}. \tag{5.23}$$

Με βάση τον ορισμό του τανυστικού γινομένου (5.17), μπορούμε αμέσως να διαπιστώσουμε, από την παραπάνω μορφή του A, ότι

$$A = I \otimes T + T \otimes I. \tag{5.24}$$

Γνωρίζοντας τις ιδιοτιμές και τα ιδιοδιανύσματα του πίνακα $T = \text{trid}(-1,2,-1)$, από τις (5.12) και (5.13), και εκμεταλλευόμενοι τις ιδιότητες του τανυστικού γινομένου, μπορούμε να βρούμε τις βασικές ιδιότητες του πίνακα A καθώς και τα ιδιοζεύγη του, που θα μας επιτρέψουν στη συνέχεια να μελετήσουμε συγκλίσεις και συγκρίσεις των βασικών point και block επαναληπτικών μεθόδων, για την αριθμητική επίλυση του συστήματος (5.21).

Είναι προφανές, από την (5.22), ότι ο $A \in I\!R^{n^2,n^2}$ είναι πραγματικός και συμμετρικός. Το ότι είναι και θετικά ορισμένος μπορεί να δειχτεί αμέσως μετά την εύρεση των ιδιοτιμών του. Λόγω της block τριδιαγώνιας μορφής του A, αυτός είναι block δικυκλικός και συνεπώς διατεταγμένος. Είναι όμως και point δικυκλικός και συνεπώς διατεταγμένος, πράγμα που αποδείχνεται ως εξής. Για το ότι είναι point δικυκλικός χρησιμοποιούμε το μεταθετικό πίνακα P με

$$\begin{aligned} P^T = [&e^1\ e^3\ \ldots\ e^{2[\frac{n+1}{2}]-1}\ e^{n+2}\ e^{n+4}\ \ldots\ e^{n+2[\frac{n}{2}]}\ e^{2n+1}\ e^{2n+3}\ \ldots\ e^{2n+2[\frac{n+1}{2}]-1}\ \ldots \\ &e^2\ e^4\ \ldots\ e^{2[\frac{n}{2}]}\ e^{n+1}\ e^{n+3}\ \ldots\ e^{n+2[\frac{n+1}{2}]-1}\ e^{2n+2}\ e^{2n+4}\ \ldots\ e^{2n+2[\frac{n}{2}]}\ \ldots] \in I\!R^{n^2,n^2}, \end{aligned} \tag{5.25}$$

όπου $[x]$ συμβολίζει το ακέραιο μέρος του αριθμού x. Τότε μπορεί να δειχτεί ότι ο PAP^T είναι της μορφής (3.49). Οτι ο point δικυκλικός πίνακας A είναι και συνεπώς διατεταγμένος αποδείχνεται στη συνέχεια. Ο point Jacobi πίνακας, που αντιστοιχεί στον A, έχει τη μορφή

$$T_J^{(p)} \;=\; \frac{1}{4}(I\otimes G + G\otimes I), \textit{όπου } G = \mathrm{trid}(1,0,1). \tag{5.26}$$

Εχοντας τον $T_J^{(p)}$, σχηματίζουμε τον πίνακα $T_J^{(p)}(\alpha) = D^{-1}(\alpha L + \frac{1}{\alpha}U)$, όπου $D = \mathrm{diag}(A)$, και L και U, αυστηρά κάτω τριγωνικός και αυστηρά άνω τριγωνικός πίνακες, και $\alpha \in \mathbb{C}\setminus\{0\}$. Από τη μορφή του A στην (5.22), καθώς και από τις ισοδύναμες μορφές του (5.23) και (5.24), παρατηρούμε ότι ο πίνακας $T_J^{(p)}(\alpha)$ μπορεί να γραφτεί και ως

$$T_J^{(p)}(\alpha) \;=\; \frac{1}{4}(I\otimes F + F\otimes I),\ \textit{όπου } F = \mathrm{trid}(\alpha,0,\frac{1}{\alpha}). \tag{5.27}$$

Θεωρούμε τον πίνακα $E = \mathrm{diag}(1, \frac{1}{\alpha}, \cdots, \frac{1}{\alpha^{n-1}})$, σχηματίζουμε τον όμοιο προς τον $T_J^{(p)}(\alpha)$ πίνακα $(E\otimes E)T_J^{(p)}(\alpha)(E\otimes E)^{-1}$, χρησιμοποιούμε την έκφραση (5.27), και εφαρμόζοντας απλές ιδιότητες του τανυστικού γινομένου παίρνουμε διαδοχικά

$$\begin{aligned}(E\otimes E)T_J^{(p)}(\alpha)(E\otimes E)^{-1} \;&=\; (E\otimes E)\left(\tfrac{1}{4}(I\otimes F + F\otimes I)\right)(E^{-1}\otimes E^{-1})\\ &=\; \tfrac{1}{4}\left((EI)\otimes(EF)+(EF)\otimes(EI)\right)(E^{-1}\otimes E^{-1})\\ &=\; \tfrac{1}{4}\left((EIE^{-1})\otimes(EFE^{-1})+(EFE^{-1})\otimes(EIE^{-1})\right)\\ &=\; \tfrac{1}{4}\left(I\otimes(EFE^{-1})+(EFE^{-1})\otimes I\right).\end{aligned}$$

Ο πίνακας, όμως, $EFE^{-1} \;=\; G$, όπως μπορεί να διαπιστωθεί αμέσως από τη μορφή του F στην (5.27) και G στην (5.26), είναι ανεξάρτητος του α. Αρα και ο $T_J^{(p)}(\alpha)$ είναι ανεξάρτητος του α, πράγμα που σημαίνει ότι ο A είναι point δικυκλικός και συνεπώς διατεταγμένος.

Για την εύρεση των ιδιοτιμών και ιδιοδιανυσμάτων του A χρησιμοποιούμε τη μορφή (5.24), όπου για τον $T = \mathrm{trid}(-1,2,-1)$ έχουμε ήδη βρεί ιδιοτιμές και ιδιοδιανύσματα στις (5.12) και (5.13), αντίστοιχα. Σημειώνουμε ακόμη ότι ο μοναδιαίος πίνακας I έχει μεν όλες τις ιδιοτιμές του ίσες με τη μονάδα αλλά ως ιδιοδιανύσματά του μπορούν να θεωρηθούν οποιαδήποτε n γραμμικά ανεξάρτητα διανύσματα. Θεωρούμε, λοιπόν, αυτά του T. Αν, συνεπώς, θεωρήσουμε τα διανύσματα $x^{(k)}\otimes x^{(l)},\ k,l = 1(1)n$, όπου $x^{(k)}, x^{(l)}$, τα ιδιοδιανύσματα του T, τότε για κάθε $k,l = 1(1)n$, έχουμε διαδοχικά

$$\begin{aligned}A(x^{(k)}\otimes x^{(l)}) \;&=\; (I\otimes T + T\otimes I)(x^{(k)}\otimes x^{(l)})\\ &=\; (Ix^{(k)})\otimes(Tx^{(l)}) + (Tx^{(k)})\otimes(Ix^{(l)})\\ =\; x^{(k)}\otimes(\lambda_l x^{(l)}) + (\lambda_k x^{(k)})\otimes x^{(l)} \;&=\; (\lambda_k+\lambda_l)(x^{(k)}\otimes x^{(l)}).\end{aligned} \tag{5.28}$$

Η ισότητα του πρώτου και τελευταίου μέλους των ισοτήτων (5.28) αποδείχνει ότι ο πίνακας A έχει ιδιοτιμές τις $\lambda_k+\lambda_l$ με αντίστοιχα ιδιοδιανύσματα τα $x^{(k)}\otimes x^{(l)},\ k,l = 1(1)n$, που είναι γραμμικά

ανεξάρτητα, όπως συμπεραίνεται από τη γραμμική ανεξαρτησία των $x^{(k)},\ k = 1(1)n$. Αναλυτικά, οι εκφράσεις των ιδιοτιμών είναι

$$\lambda_{k,l} \ = \ 4\sin^2\left(\frac{\pi k}{2(n+1)}\right) + 4\sin^2\left(\frac{\pi l}{2(n+1)}\right),\ k,l = 1(1)n, \tag{5.29}$$

των δε αντίστοιχων ιδιοδιανυσμάτων είναι οι

$$x^{(k,l)} \ = \ x^{(k)} \otimes x^{(l)},\ k,l = 1(1)n, \tag{5.30}$$

όπου οι συνιστώσες των $x^{(k)}$, $x^{(l)}$, με βάση την (5.13), είναι αυτές που δίνονται στη συνέχεια

$$\begin{aligned} x^{(k)} \ &= \ \left[\sin(\tfrac{k\pi}{n+1})\ \sin(\tfrac{2k\pi}{n+1})\ ...\ \sin(\tfrac{nk\pi}{n+1})\right]^T,\ k = 1(1)n, \\ x^{(l)} \ &= \ \left[\sin(\tfrac{l\pi}{n+1})\ \sin(\tfrac{2l\pi}{n+1})\ ...\ \sin(\tfrac{nl\pi}{n+1})\right]^T,\ l = 1(1)n. \end{aligned}$$

Σημείωση: Από τις εκφράσεις (5.29) των ιδιοτιμών βλέπουμε αμέσως ότι $\lambda_{k,l} = \lambda_{l,k}, k \neq l = 1(1)n$. Το γεγονός ότι αυτές οι ίσες ιδιοτιμές αποτελούν πράγματι διπλές ιδιοτιμές του πίνακα A, και άρα οι εκφράσεις (5.29) δίνουν όλες τις ιδιοτιμές του, αποδείχνεται από τη γραμμική ανεξαρτησία των ιδιοδιανυσμάτων (5.30). Για την απόδειξη του τελευταίου ισχυρισμού σχηματίζουμε τον πίνακα

$$\begin{aligned} Y \ &= \ [x^{(1)} \otimes x^{(1)}\ x^{(1)} \otimes x^{(2)}\ \cdots\ x^{(1)} \otimes x^{(n)}\ x^{(2)} \otimes x^{(1)}\ \cdots\ x^{(n)} \otimes x^{(n)}] \\ &= \ [x^{(1)}\ x^{(2)}\ \cdots\ x^{(n)}] \otimes [x^{(1)}\ x^{(2)}\ \cdots\ x^{(n)}] \ =: \ X \otimes X. \end{aligned}$$

Ομως ο πίνακας X είναι αντιστρέψιμος αφού στήλες του είναι τα ιδιοδιανύσματα $x^{(k)}, k = 1(1)n$, του πίνακα A της εφαρμογής της μονοδιάστατης περίπτωσης που είναι γραμμικά ανεξάρτητα. Αρα και ο πίνακας Y είναι αντιστρέψιμος πράγμα που συνεπάγεται τη γραμμική ανεξαρτησία των στηλών του.

Από την έκφραση των ιδιοτιμών διαπιστώνεται αμέσως ότι όλες οι n^2 ιδιοτιμές του A είναι θετικές, πράγμα που συνεπάγεται ότι ο A, ως πραγματικός και συμμετρικός, θα είναι και θετικά ορισμένος.

Για την εύρεση των φασματικών ακτίνων των point και block επαναληπτικών μεθόδων Jacobi, Gauss-Seidel και SOR στις οποίες και θα περιοριστούμε, αφού για τις άλλες μεθόδους, που μελετήθηκαν στην εφαρμογή της μίας διάστασης, μπορούμε να καταλήξουμε σε συμπεράσματα διαμέσου των τριών πρώτων μεθόδων, όπως είδαμε, εργαζόμαστε ως εξής.

point Jacobi: Ο αντίστοιχος επαναληπτικός πίνακας $T_J^{(p)}$ είναι ο $T_J^{(p)} = I - \frac{1}{4}A$ και άρα οι ιδιοτιμές του, $\mu_{k,l}$, θα δίνονται από τις εκφράσεις $\mu_{k,l} = 1 - \frac{1}{4}\lambda_{k,l} = \frac{1}{2}\left(\cos\left(\frac{k\pi}{n+1}\right) + \cos\left(\frac{l\pi}{n+1}\right)\right)$. Επομένως η φασματική ακτίνα του επαναληπτικού πίνακα της point Jacobi θα είναι

$$\rho(T_J^{(p)}) \ = \ \cos\left(\frac{\pi}{n+1}\right).$$

Παρατηρούμε ότι η φασματική ακτίνα του point Jacobi επαναληπτικού πίνακα, που βρέθηκε, είναι ταυτόσημη με αυτήν του επαναληπτικού πίνακα του Jacobi στην περίπτωση της εφαρμογής στη μία διάσταση. Λόγω δε και της point δικυκλικής και συνεπώς διατεταγμένης ιδιότητας του A και οι φασματικές ακτίνες της point Gauss-Seidel και της βέλτιστης SOR, στην παρούσα εφαρμογή στις δύο διαστάσεις, θα είναι ταυτόσημες με αυτές των αντίστοιχων μεθόδων στην εφαρμογή στη μία διάσταση. Ομοίως, και τα συμπεράσματα τα αφορώντα σε συγκλίσεις και συγκρίσεις μέσων ασυμπτωτικών ταχυτήτων σύγκλισης των μεθόδων στην περίπτωση που εξετάζουμε θα είναι ακριβώς τα ίδια με αυτά της περίπτωσης στη μία διάσταση και επομένως η εκ νέου παράθεσή τους παραλείπεται.

Για τις αντίστοιχες block επαναληπτικές μεθόδους αρχίζουμε με την εύρεση του block Jacobi επαναληπτικού πίνακα, $T_J^{(b)}$, τον οποίο θα προσπαθήσουμε να εκφράσουμε χρησιμοποιώντας τανυστικά γινόμενα ώστε να μπορέσουμε να εκμεταλλευτούμε τις αναλυτικές εκφράσεις των ιδιοτιμών, και των ιδιοδιανυσμάτων, του πίνακα A. Με τη χρησιμοποίηση των μορφών (5.23) και (5.24) του A μπορούμε να πάρουμε διαδοχικά ότι

$$\begin{aligned} T_J^{(b)} &= (I \otimes (2I+T))^{-1} (I \otimes (2I+T) - A) \\ &= (I \otimes (2I+T)^{-1}) (I \otimes (2I+T) - I \otimes T - T \otimes I) \\ &= (I \otimes (2I+T)^{-1}) (I \otimes 2I - T \otimes I) \\ &= (I\ I) \otimes ((2I+T)^{-1}\ 2I) - (I\ T) \otimes ((2I+T)^{-1}\ I) \\ &= I \otimes 2(2I+T)^{-1} - T \otimes (2I+T)^{-1} = (2I-T) \otimes (2I+T)^{-1}. \end{aligned} \tag{5.31}$$

Είναι γνωστό ότι οι ιδιοτιμές του T, και τα αντίστοιχα ιδιοδιανύσματα, είναι αυτά του πίνακα T της εφαρμογής στη μονοδιάστατη περίπτωση. Ακόμη γνωρίζουμε ότι οι ιδιοτιμές και τα ιδιοδιανύσματα του αντίστροφου ενός πίνακα, εφόσον υπάρχει, είναι οι αντίστροφες ιδιοτιμές, που αντιστοιχούν στα ίδια ιδιοδιανύσματα του αρχικού πίνακα, που είναι και ιδιοδιανύσματα του αντίστροφου πίνακα. Με βάση τα παραπάνω, οι ιδιοτιμές και τα ιδιοδιανύσματα του $2I-T$ στο δεξιά μέλος των ισοτήτων (5.31) θα είναι $2-\lambda_k$ και $x^{(k)},\ k=1(1)n$, αντίστοιχα, ενώ του $(2I+T)^{-1}$ θα είναι $(2+\lambda_k)^{-1}$, και $x^{(k)},\ k=1(1)n$. Λόγω του τανυστικού γινομένου ο πίνακας $T_J^{(b)}$ θα έχει ιδιοτιμές

$$\lambda_{k,l} = \frac{2-\lambda_k}{2+\lambda_l} = \frac{\cos\left(\frac{k\pi}{n+1}\right)}{2-\cos\left(\frac{l\pi}{n+1}\right)},\ k,l=1(1)n,$$

και ιδιοδιανύσματα

$$x^{(k)} \otimes x^{(l)},\ k,l=1(1)n. \tag{5.32}$$

Σημειώνεται ότι τα ιδιοδιανύσματα στην (5.32) είναι τα ίδια ακριβώς με αυτά του πίνακα A στην (5.30) και άρα θα είναι γραμμικά ανεξάρτητα. Η φασματική ακτίνα του block Jacobi επαναληπτικού πίνακα της περίπτωσης που εξετάζουμε είναι

$$\rho(T_J^{(b)}) = \frac{\cos\left(\frac{\pi}{n+1}\right)}{2-\cos\left(\frac{\pi}{n+1}\right)}.$$

Από την έκφραση για τη φασματική ακτίνα του block Jacobi πίνακα και την block δικυκλική και συνεπώς διατεταγμένη ιδιότητα του A βρίσκονται αμέσως οι φασματικές ακτίνες των block Gauss-Seidel και block SOR επαναληπτικών πινάκων, που θα είναι

$$\rho(T_{GS}^{(b)}) = \frac{\cos^2\left(\frac{\pi}{n+1}\right)}{\left(2-\cos\left(\frac{\pi}{n+1}\right)\right)^2}$$

και

$$\rho(\mathcal{L}_{\omega_\beta}^{(b)}) = \frac{2-\cos\left(\frac{\pi}{n+1}\right)-2\sqrt{1-\cos\left(\frac{\pi}{n+1}\right)}}{2-\cos\left(\frac{\pi}{n+1}\right)+2\sqrt{1-\cos\left(\frac{\pi}{n+1}\right)}},$$

αντίστοιχα.

Για την εύρεση των μέσων ασυμπτωτικών ταχυτήτων σύγκλισης των point και block επαναληπτικών μεθόδων Jacobi, Gauss-Seidel και SOR, καθώς και για τις αντίστοιχες συγκρίσεις, θα θεωρήσουμε ότι $h=\frac{1}{n+1}\to 0^+$, και θα χρησιμοποιήσουμε αναπτύγματα, ως προς h, συγκλινουσών σειρών. Για τις αντίστοιχες point μεθόδους απλά αντιγράφουμε από αυτά που ήδη βρέθηκαν για τις ίδιες μεθόδους στην εφαρμογή της μονοδιάστατης περίπτωσης. Συγκεκριμένα,

$$\begin{aligned}\mathcal{R}_\infty(T_J^{(p)}) &= \tfrac{\pi^2}{2}h^2+\mathcal{O}(h^4),\\ \mathcal{R}_\infty(T_{GS}^{(p)}) &= \pi^2h^2+\mathcal{O}(h^4),\\ \mathcal{R}_\infty(\mathcal{L}_{\omega_\beta}^{(p)}) &= 2\pi h+\mathcal{O}(h^3).\end{aligned}$$

Για τις block επαναληπτικές μεθόδους μπορούμε εύκολα να βρούμε ότι

$$\begin{aligned}\mathcal{R}_\infty(T_J^{(b)}) &= \pi^2h^2+\mathcal{O}(h^4),\\ \mathcal{R}_\infty(T_{GS}^{(b)}) &= 2\pi^2h^2+\mathcal{O}(h^4),\\ \mathcal{R}_\infty(\mathcal{L}_{\omega_\beta}^{(b)}) &= 2\sqrt{2}\pi h+\mathcal{O}(h^3).\end{aligned}$$

Οπως μπορεί να διαπιστωθεί, οι block Jacobi και block Gauss-Seidel είναι δυο φορές ταχύτερες από τις αντίστοιχες point μεθόδους, ενώ η block SOR είναι κατά $\sqrt{2}$ φορές ταχύτερη από την point SOR μέθοδο. Και στην περίπτωση των block μεθόδων παρατηρούμε ακόμη ότι η block SOR είναι ταχύτερη από την block Gauss-Seidel κατά μία τάξη μεγέθους, ως προς h, με αντίστοιχο ασυμπτωτικό παράγοντα $\frac{\sqrt{2}}{\pi}\frac{1}{h}$.

ΑΣΚΗΣΕΙΣ

1.: Να βρεθούν οι γενικές λύσεις των εξισώσεων διαφορών.

$$\alpha)\quad y_{k+2}-5y_{k+1}+6y_k=0,\ k=0,1,2,\ldots,$$

$$\beta)\quad y_{k+2}-4y_{k+1}+4y_k=0,\ k=0,1,2,\ldots.$$

2.: Στη γνωστή ακολουθία Fibonacci δίνονται οι δυο πρώτοι όροι της $a_0 = 0,\ a_1 = 1$, ενώ κάθε επόμενος όρος της ορίζεται από την αναδρομική σχέση

$$a_{k+2} = a_{k+1} + a_k,\ k = 0, 1, 2, \ldots .$$

Να βρεθεί ο γενικός όρος της.

3.: Με τη βοήθεια των εξισώσεων διαφορών να βρεθεί η τιμή της $\det(\mathrm{trid}(-1, 2, -1))$, όπου $\mathrm{trid}(-1, 2, -1) \in I\!R^{n,n}$.

4.: Χρησιμοποιώντας εξισώσεις διαφορών να βρεθεί η γενική έκφραση των πολυωνύμων $T_k(z)$ του Chebyshev πρώτου είδους βαθμού k, που ορίζονται, είτε για $z \in [-1, 1]$ είτε για $z \geq 1$, από τη σχέση $T_k(z) = 2zT_{k-1}(z) - T_{k-2}(z),\ k = 2, 3, \ldots$, και τους δυο πρώτους όρους $T_0(z) = 1$ και $T_1(z) = z$.

5.: Να βρεθούν οι ιδιοτιμές και τα ιδιοδιανύσματα του πίνακα $\mathrm{trid}(b, a, c) \in I\!R^{n,n}$, όπου $bc > 0$. (Υπόδειξη: Αν χρειαστεί μπορεί να χρησιμοποιηθεί κατάλληλα διαγώνιος πίνακας ώστε ο δοθείς να καταστεί όμοιος προς συμμετρικό.)

6.: Να βρεθούν οι ιδιοτιμές και τα ιδιοδιανύσματα των $n \times n$ πινάκων

$$\begin{bmatrix} 1 & -1 & & & \\ -1 & 2 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 2 & -1 \\ & & & -1 & 2 \end{bmatrix}, \quad \begin{bmatrix} 2 & -1 & & & \\ -1 & 2 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 2 & -1 \\ & & & -1 & 1 \end{bmatrix}.$$

7.: Να βρεθούν οι ιδιοτιμές και τα ιδιοδιανύσματα του $n \times n$ πίνακα

$$\begin{bmatrix} 2 & -2 & & & \\ -1 & 2 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 2 & -1 \\ & & & -2 & 2 \end{bmatrix}.$$

(**Προσοχή**: Ο πίνακας είναι μή αντιστρέψιμος και **δεν** είναι συμμετρικός. Αν χρειαστεί μπορεί να χρησιμοποιηθεί κατάλληλα διαγώνιος πίνακας ώστε ο δοθείς να καταστεί όμοιος προς συμμετρικό.)

8.: Δίνεται ο $n \times n$ $(n \geq 3)$ πίνακας

$$A = \begin{bmatrix} 1 & 1 & & & & \\ 1 & 2 & 1 & & & \\ & 1 & 2 & 1 & & \\ & & \ddots & \ddots & \ddots & \\ & & & 1 & 2 & 1 \\ & & & & 1 & 1 \end{bmatrix}.$$

α) Να δειχτεί, με οποιοδήποτε τρόπο, ότι ο A έχει μία από τις ιδιοτιμές του ίση με μηδέν, όλες δε τις άλλες πραγματικές και θετικές.
β) Χρησιμοποιώντας εξισώσεις διαφορών να βρεθούν οι αναλυτικές εκφράσεις όλων των ιδιοτιμών του A. και
γ) Να βρεθούν επίσης αναλυτικές εκφράσεις των ιδιοδιανυσμάτων του πίνακα A.

9.: Να βρεθούν οι ιδιοτιμές του πίνακα

$$A = \begin{bmatrix} a & 1 & & & & \\ 1 & -a & 1 & & & \\ & 1 & a & 1 & & \\ & & 1 & -a & 1 & \\ & & & \ddots & \ddots & \ddots \end{bmatrix} \in I\!R^{n,n},\ a \neq 0.$$

(Υπόδειξη: Να χρησιμοποιηθεί μεταθετικός πίνακας P και να θεωρηθεί ο όμοιος προς τον A πίνακας PAP^T, ώστε ο τελευταίος να είναι της μορφής $\left[\begin{array}{c|c} aI_{n_1} & B \\ \hline B^T & -aI_{n_2} \end{array}\right]$. Στη συνέχεια να θεωρηθούν δυο περιπτώσεις n άρτιος και n περιττός.)

10.: Να βρεθούν οι ιδιοτιμές του πίνακα

$$A = \begin{bmatrix} a & c & & & & \\ b & d & c & & & \\ & b & a & c & & \\ & & b & d & c & \\ & & & \ddots & \ddots & \ddots \end{bmatrix} \in I\!R^{n,n},\quad a,b,c,d \in I\!R,\ ad \neq 0,\ bc > 0.$$

(Υπόδειξη: Μπορεί να χρησιμοποιηθεί το αποτέλεσμα της προηγούμενης Ασκησης.)

11.: Να αποδειχτούν οι ιδιότητες (α)-(ε) που βασίζονται στον ορισμό του τανυστικού γινομένου.

6 Μέθοδοι Ελαχιστοποίησης

6.1 Εισαγωγή

Οι μέθοδοι που θα μελετηθούν στο παρόν κεφάλαιο αναπτύχτηκαν κυρίως για την επίλυση πραγματικών γραμμικών συστημάτων $Ax = b$, με $A \in I\!R^{n,n}$, $A^T = A$, θετικά ορισμένο και $b \in I\!R^n$. Εφεξής ο A θα ικανοποιεί τις προαναφερθείσες ιδιότητες εκτός και αν σαφώς ορίζεται αλλιώς. Ολες οι μέθοδοι που θα αναπτυχτούν βασίζονται στην ελαχιστοποίηση ενός συναρτησιακού, δηλαδή πραγματικής συνάρτησης

$$f \;:=\; f(x_1,\; x_2,\; \cdots,\; x_n) \;\equiv\; f(x), \quad \textit{όπου } x \in (\Omega \subseteq)\; I\!R^n.$$

Θα υποτίθεται ότι η f είναι αρκετά ομαλή έτσι ώστε τα σημεία στα οποία έχει τοπικά ελάχιστα να βρίσκονται μεταξύ των κρίσιμων σημείων της, δηλαδή των σημείων x στα οποία η κλίση της είναι μηδέν

$$\nabla f(x) = \left[\frac{\partial f}{\partial x_1}\;\; \frac{\partial f}{\partial x_2}\;\; \cdots\;\; \frac{\partial f}{\partial x_n}\right]^T \;=\; 0. \tag{6.1}$$

Εφεξής θα γράφουμε

$$(\cdot,\cdot) \;\equiv\; (\cdot,\cdot)_2 \;\; \textit{και} \;\; ||\cdot|| \;\equiv\; ||\cdot||_2,$$

εκτός κι αν ορίζεται διαφορετικά.

Η βασική ιδέα μιας μεθόδου ελαχιστοποίησης περιγράφεται στη συνέχεια. Καταρχάς η συνάρτηση f ορίζεται κατά τέτοιο τρόπο ώστε ένα σημείο στο οποίο έχει τοπικό ελάχιστο να είναι η λύση του συστήματος που θεωρούμε, δηλαδή η $x^* = A^{-1}b$. Τότε, για ένα $x \in I\!R^n$ σταθερό και για κάθε $u \in I\!R^n \setminus \{0\}$ με $||u|| = c > 0$ σταθερό θεωρούμε τη συνάρτηση

$$g(\alpha) \;=\; f(x + \alpha u) \;=\; f(x) + \alpha(\nabla f(x), u) + \mathcal{O}(\alpha^2),$$

όπου το δεξιά μέλος προκύπτει αν αναπτύξουμε κατά Taylor την $f(x+\alpha u)$ στο x. Προφανώς έχουμε $g(0) = f(x)$ και $g'(0) = (\nabla f(x), u)$, οπότε για $\alpha \to 0^+$ και $\nabla f(x) \neq 0$ η διαφορά $f(x+\alpha u) - f(x)$ θα έχει το πρόσημο του $(\nabla f(x), u)$. Αρα η $g(a)$ θα είναι γνήσια αύξουσα στην περιοχή του μηδενός ανν $(\nabla f(x), u) > 0$ και, αντίστοιχα, γνήσια φθίνουσα ανν $(\nabla f(x), u) < 0$. Στην περίπτωση της γνήσιας φθίνουσας $g(\alpha)$ η "ταχύτητα" ελάττωσής της θα είναι η μέγιστη δυνατή ανν $(\nabla f(x), u) = ||\nabla f(x)||\;||u|| \cos\theta$, με $\theta \;=\; \widehat{\nabla f(x), u} \;=\; \pi$. Δηλαδή στην κατεύθυνση $u = -\nabla f(x)$. Επομένως ξεκινώντας από κάποιο αυθαίρετο x και κινούμενοι προς την κατεύθυνση $-\nabla f(x)$ κινούμαστε από την τρέχουσα τιμή του συναρτησιακού προς τιμές του, που είναι οι μικρότερες δυνατές στην περιοχή του x. Αρα προς τιμές που βρίσκονται πλησιέστερα προς το ελάχιστο που μας ενδιαφέρει, κι αυτό γίνεται, με τη μεγαλύτερη δυνατή ταχύτητα.

6.2 Μέθοδος της Απότομης Καθόδου (Steepest Descent)

Η μέθοδος της Απότομης Καθόδου είναι μία επαναληπτική μέθοδος, οφείλεται στον Cauchy και ακολουθεί τη βασική ιδέα που ήδη αναπτύχτηκε. Πιο συγκεκριμένα, στην αρχή της $k+1$ επανάληψης είναι γνωστή η προσέγγιση $x^{(k)}$, με τιμή του συναρτησιακού $f(x^{(k)})$. Για να ορίσουμε το $x^{(k+1)}$ κινούμαστε στην κατεύθυνση $u = -\nabla f(x^{(k)})$, οπότε $x^{(k+1)} = x^{(k)} - \alpha \nabla f(x^{(k)})$, το δε α ορίζεται ως η μικρότερη θετική τιμή του $\alpha = \alpha_{k+1}$, που καθιστά το $f(x^{(k+1)}) = f(x^{(k)} - \alpha \nabla f(x^{(k)}))$ ελάχιστο. Το συγκεκριμένο συναρτησιακό που χρησιμοποιείται είναι το

$$f(x) := \frac{1}{2}(Ax, x) - (b, x), \quad x \in I\!R^n. \tag{6.2}$$

Η επιλογή του συναρτησιακού (6.2) με βάση την ανάλυση που προηγήθηκε δικαιολογείται από την παρακάτω ισχύουσα πρόταση.

Θεώρημα 6.1 *Εστω το γραμμικό σύστημα*

$$Ax = b,\ A \in I\!R^{n,n},\ A^T = A,\ A\ \theta\epsilon\tau\iota\kappa\acute{\alpha}\ o\rho\iota\sigma\mu\acute{\epsilon}\nu o\varsigma\ \kappa\alpha\iota\ b \in I\!R^n, \tag{6.3}$$

και το συναρτησιακό (6.2). Τότε η λύση του (6.3) ($x^* = A^{-1}b$) είναι το σημείο στο οποίο η τιμή του $f(x)$ γίνεται ελάχιστη.

Απόδειξη: Εστω $x = x^* + y,\ y \in I\!R^n$, τότε, λαβαίνοντας υπόψη τις ιδιότητες του A, η τιμή του $f(x)$ θα γίνεται διαδοχικά

$$\begin{array}{l} f(x) = f(x^* + y) = \frac{1}{2}(A(x^* + y), x^* + y) - (b, x^* + y) \\ = \frac{1}{2}(Ax^*, x^*) + \frac{1}{2}(Ax^*, y) + \frac{1}{2}(Ay, x^*) + \frac{1}{2}(Ay, y) - (b, x^*) - (b, y) \\ = f(x^*) + (Ax^* - b, y) + \frac{1}{2}(Ay, y) = f(x^*) + \frac{1}{2}(Ay, y) \geq f(x^*) \end{array}, \tag{6.4}$$

με το "=" να ισχύει ανν $y = 0$, δηλαδή, ανν $x = x^*$. (Σημείωση: Η ελάχιστη τιμή του $f(x)$ είναι τότε $\min_{x \in I\!R^n} f(x) = f(x^*) = \frac{1}{2}(Ax^*, x^*) - (b, x^*) = -\frac{1}{2}(b, A^{-1}b)$.) □

Αν προσπαθήσουμε να δείξουμε, με τη μέθοδο που περιγράφτηκε στην εισαγωγή, ότι το x^* είναι το σημείο στο οποίο η $f(x)$ ελαχιστοποιείται, θα πρέπει πρώτα να βρούμε τα κρίσιμα σημεία της και από αυτά να βρούμε εκείνο, αν υπάρχει, που ελαχιστοποιεί την τιμή της. Από την έκφραση της $f(x)$ στην (6.2) βρίσκουμε την κλίση της από την (6.1), για την k-οστή συνιστώσα της οποίας έχουμε

$$\begin{array}{l} (\nabla f(x))_k = \frac{\partial}{\partial x_k} f(x) = \frac{\partial}{\partial x_k}\left[\frac{1}{2}(Ax, x) - (b, x)\right] = \frac{1}{2}\frac{\partial}{\partial x_k}\left[\sum_{i=1}^n \left(x_i \sum_{j=1}^n a_{ij} x_j\right)\right] \\ -\frac{\partial}{\partial x_k}\left(\sum_{i=1}^n b_i x_i\right) = \frac{1}{2}\left(\sum_{j=1}^n a_{kj} x_j + \sum_{i=1}^n x_i a_{ik}\right) - b_k = \sum_{i=1}^n a_{ki} x_i - b_k = (Ax - b)_k, \end{array}$$

οπότε $\nabla f(x) = Ax - b$. Με το μηδενισμό της κλίσης παίρνουμε $x = A^{-1}b$, ως το μόνο κρίσιμο σημείο της $f(x)$. Για το αν το σημείο που βρέθηκε είναι σημείο ελάχιστου της $f(x)$ θεωρούμε τον Εσιανό πίνακά της, H, για τον οποίο ισχύει ότι $h_{ij} = \frac{\partial^2}{\partial x_i x_j}(f(x)) = \frac{\partial}{\partial x_j}[(Ax-b)_i] = a_{ij}$, και άρα $H = A$. Επειδή ο A, εκτός από πραγματικός συμμετρικός, είναι και θετικά ορισμένος το μοναδικό κρίσιμο σημείο που ήδη βρέθηκε είναι σημείο ελάχιστου της $f(x)$.

Επανερχόμενοι στη μέθοδο απότομης καθόδου που περιγράφουμε, ορίζουμε καταρχάς το διάνυσμα-υπόλοιπο (ή απλά υπόλοιπο) στο τέλος της k επανάληψης ως

$$r^{(k)} = b - Ax^{(k)} = -\nabla f(x^{(k)}).$$

Αν $r^{(k)} \neq 0$ ($x^{(k)} \neq A^{-1}b$), αλλιώς $r^{(k)} = 0$ και η λύση θα έχει βρεθεί, τότε η μέθοδος της απότομης καθόδου υπολογίζει την επανάληψη $x^{(k+1)} = x^{(k)} + \alpha r^{(k)}$, για $\alpha > 0$, ελαχιστοποιώντας τη συνάρτηση $f(x^{(k)} + \alpha r^{(k)})$ ως προς α. Είναι εύκολο να πάρουμε διαδοχικά ότι

$$\begin{aligned}
&f(x^{(k)} + \alpha r^{(k)}) = \tfrac{1}{2}\left(A(x^{(k)} + \alpha r^{(k)}), x^{(k)} + \alpha r^{(k)}\right) - (b, x^{(k)} + \alpha r^{(k)}) \\
&= \tfrac{1}{2}(Ax^{(k)}, x^{(k)}) + \tfrac{1}{2}\alpha(Ax^{(k)}, r^{(k)}) + \tfrac{1}{2}\alpha(Ar^{(k)}, x^{(k)}) + \tfrac{1}{2}\alpha^2(Ar^{(k)}, r^{(k)}) - (b, x^{(k)}) - \alpha(b, r^{(k)}) \\
&= f(x^{(k)}) + \tfrac{1}{2}\alpha^2(Ar^{(k)}, r^{(k)}) + \alpha(Ax^{(k)}, r^{(k)}) - \alpha(b, r^{(k)}) \\
&= f(x^{(k)}) + \tfrac{1}{2}\alpha^2(Ar^{(k)}, r^{(k)}) - \alpha(r^{(k)}, r^{(k)}).
\end{aligned}$$

Η έκφραση στο δεξιά μέλος των παραπάνω ισοτήτων είναι τριώνυμο δεύτερου βαθμού ως προς α. Επειδή $\frac{1}{2}(Ar^{(k)}, r^{(k)}) > 0$, αφού ο A είναι πραγματικός, συμμετρικός και θετικά ορισμένος και $r^{(k)} \neq 0$, το τριώνυμο θα ελαχιστοποιείται για

$$\alpha = \alpha_{k+1} = \frac{(r^{(k)}, r^{(k)})}{(Ar^{(k)}, r^{(k)})}.$$

Επιπλέον έχουμε, μετά από απλές πράξεις, ότι

$$f(x^{(k)} + \alpha_{k+1} r^{(k)}) = f(x^{(k)}) - \frac{(r^{(k)}, r^{(k)})^2}{2(Ar^{(k)}, r^{(k)})} < f(x^{(k)}).$$

Ο αλγόριθμος της μεθόδου απότομης καθόδου είναι εύκολο να διατυπωθεί και θα δοθεί στη συνέχεια. Απλά τονίζουμε δύο σημεία. Το πρώτο είναι ότι πρέπει να οριστεί το κριτήριο σύγκλισης της μεθόδου. Οπως και στις κλασικές επαναληπτικές μεθόδους μπορούμε να ορίσουμε ως κριτήριο σύγκλισης το απόλυτο κριτήριο: $\|x^{(k)} - x^{(k-1)}\| \leq \epsilon$ ή το σχετικό απόλυτο κριτήριο: $\frac{\|x^{(k)} - x^{(k-1)}\|}{\|x^{(k)}\|} \leq \eta$, όπου ϵ και η επιθυμητά φράγματα του σφάλματος. Επειδή $r^{(k)} = b - Ax^{(k)} = A(x - x^{(k)}) = -Ae^{(k)}$, είναι φανερό ότι αν $\lim_{k\to\infty} r^{(k)} = 0$ τότε και $\lim_{k\to\infty} x^{(k)} = x$. Για το λόγο αυτό μπορεί να ληφτεί και ως κριτήριο σύγκλισης το: $\|r^{(k)}\| \leq \epsilon$. Το δεύτερο σημείο είναι ότι συνήθως παίρνουμε $x^{(0)} = 0$. (Σημείωση: Στον παρακάτω, όπως και στους άλλους αλγόριθμους του παρόντος κεφαλαίου, παίρνεται ως κριτήριο σύγκλισης το $\|r^{(k)}\| \leq \epsilon$ και ως αρχική προσέγγιση το $x^{(0)} = 0$. Αν παρθεί

άλλο κριτήριο σύγκλισης ή άλλη αρχική προσέγγιση τότε ο αντίστοιχος αλγόριθμος θα πρέπει να υποστεί μικρή τροποποίηση.)

Αλγόριθμος Μεθόδου Απότομης Καθόδου:

Δεδομένα: $A \in I\!R^{n,n},\ A^T = A,\ A$ θετικά ορισμένος, $b \in I\!R^n$, $\epsilon \in I\!R^+$ το επιθυμητό σφάλμα.
$x^{(0)} = 0$
$r^{(0)} = b$
$k = 0$
Εφόσον $\|r^{(k)}\| > \epsilon$
$k = k + 1$
$\alpha_k = \frac{(r^{(k-1)}, r^{(k-1)})}{(Ar^{(k-1)}, r^{(k-1)})}$
$x^{(k)} = x^{(k-1)} + \alpha_k r^{(k-1)}$
$r^{(k)} = b - Ax^{(k)}$
Τέλος 'Εφόσον'
Αποτέλεσμα: $x = x^{(k)}$ η προσέγγιση της λύσης.

Αν μελετήσουμε προσεκτικά τα βήματα του αλγόριθμου που μόλις δόθηκε θα δούμε ότι σε κάθε επανάληψη το μεγαλύτερο κόστος των υπολογισμών σε πλήθος πράξεων προέρχεται από τα δύο γινόμενα πίνακα επί διάνυσμα, $Ax^{(k)}$ και $Ar^{(k-1)}$, που είναι της τάξης $\mathcal{O}(n^2)$. Οι δυο αυτοί πολλαπλασιασμοί είναι δυνατόν να αναχτούν σε έναν αν χρησιμοποιήσουμε την παρακάτω ιδιότητα που προέρχεται από τις μέχρι τώρα χρησιμοποιηθείσες σχέσεις. Συγκεκριμένα

$$\begin{aligned} r^{(k)} &= b - Ax^{(k)} = b - A(x^{(k-1)} + \alpha_k r^{(k-1)}) = r^{(k-1)} - \alpha_k A r^{(k-1)} \\ &\left(= r^{(k-1)} - \frac{(r^{(k-1)}, r^{(k-1)})}{(Ar^{(k-1)}, r^{(k-1)})} A r^{(k-1)} \right). \end{aligned} \tag{6.5}$$

Αυτό σημαίνει ότι η έκφραση $r^{(k)} = b - Ax^{(k)}$, που υπάρχει στο τελευταίο βήμα της ανακύκλωσης του αλγόριθμου, μπορεί να αντικατασταθεί από την $r^{(k)} = r^{(k-1)} - \alpha_k A r^{(k-1)},\ k = 1, 2, 3, \cdots$, όπου το γινόμενο $Ar^{(k-1)}$ θα έχει βρεθεί προηγουμένως κατά τον υπολογισμό του α_k και θα έχει αποθηκευτεί.

Ιδιότητες σαν αυτή που μόλις αναφέρθηκε υπάρχουν και άλλες, μερικές από τις οποίες δίνουμε στη συνέχεια υπό μορφή προτάσεων.

Θεώρημα 6.2 $(r^{(k)}, r^{(k-1)}) = 0.$

Απόδειξη: Πράγματι, αν χρησιμοποιηθεί η τελευταία έκφραση από τις (6.5), που είναι μέσα στις παρενθέσεις, έχουμε

$$\begin{aligned} (r^{(k)}, r^{(k-1)}) &= (r^{(k-1)} - \frac{(r^{(k-1)}, r^{(k-1)})}{(Ar^{(k-1)}, r^{(k-1)})} A r^{(k-1)}, r^{(k-1)}) \\ &= (r^{(k-1)}, r^{(k-1)}) - \frac{(r^{(k-1)}, r^{(k-1)})}{(Ar^{(k-1)}, r^{(k-1)})} (Ar^{(k-1)}, r^{(k-1)}) = 0. \end{aligned}$$

□

Μια άμεση συνέπεια του προηγούμενου θεωρήματος είναι και το παρακάτω πόρισμα.

Πόρισμα 6.1 $(x^{(k+1)} - x^{(k)}, x^{(k)} - x^{(k-1)}) = 0.$

Απόδειξη: Εχουμε ότι $(x^{(k+1)} - x^{(k)}, x^{(k)} - x^{(k-1)}) = (\alpha_{k+1} r^{(k)}, \alpha_k r^{(k-1)}) = 0.$ □

Αν $e^{(k)} = x^{(k)} - x$ είναι το διάνυσμα-σφάλμα στο τέλος της k επανάληψης είναι δυνατόν να βρεθούν απλές σχέσεις που να το συνδέουν με το διάνυσμα-υπόλοιπο και οι οποίες μπορούν να φανούν χρήσιμες στη συνέχεια. Π.χ.,

$$\begin{aligned} &\text{i) } Ae^{(k)} = A(x^{(k)} - x) = Ax^{(k)} - b = -r^{(k)}. \\ &\text{ii) } (Ae^{(k)}, r^{(k-1)}) = 0. \\ &\text{iii) } e^{(k)} = x^{(k)} - x = x^{(k-1)} + \alpha_k r^{(k-1)} - x = e^{(k-1)} + \alpha_k r^{(k-1)}. \end{aligned} \tag{6.6}$$

Για να δοθεί σχέση που να συνδέει δυό διαδοχικά σφάλματα $e^{(k)}$, $e^{(k+1)}$ που να είναι ανεξάρτητη από οποιαδήποτε υπόλοιπα πρέπει να εισαχτεί μία νέα διανυσματική norm. Για το σκοπό αυτό χρησιμοποιούμε τη (θετική) τετραγωνική ρίζα ενός Ερμιτιανού και θετικά ορισμένου πίνακα A, που ορίστηκε σε προηγούμενο κεφάλαιο. Συγκεκριμένα έχουμε:

Για ένα δοσμένο πίνακα $A \in \mathbb{C}^{n,n}$, που είναι Ερμιτιανός και θετικά ορισμένος, και για κάθε $x \in \mathbb{C}^n$ ορίζουμε τη συνάρτηση (απεικόνιση του $\mathbb{C}^n$ στο $\mathbb{R}^{+,0}$)

$$||x||_{A^{\frac{1}{2}}} := (Ax, x)^{\frac{1}{2}}. \tag{6.7}$$

Η συνάρτηση $(Ax, x)^{\frac{1}{2}}$ της (6.7) ορίζει μια διανυσματική norm, η οποία είναι γνωστή και ως norm ενέργειας (energy norm). Για να δειχτεί ότι η (6.7) ορίζει μία norm αρκεί να μετασχηματίσουμε την $(Ax, x)^{\frac{1}{2}}$ ως εξής

$$(Ax, x)^{\frac{1}{2}} = (A^{\frac{1}{2}}x, A^{\frac{1}{2}}x)^{\frac{1}{2}} = ||A^{\frac{1}{2}}x||,$$

οπότε είναι πολύ εύκολο να διαπιστώσουμε ότι επαληθεύονται οι τρεις βασικές ιδιότητες τις οποίες μία συνάρτηση πρέπει να ικανοποιεί για να ορίζει μια διανυσματική norm στο $\mathbb{C}^n$. Οι αποδείξεις ως πολύ απλές παραλείπονται.

Θεώρημα 6.3 *Εστω $x^{(k)}$, $k \geq 0$, η ακολουθία που παράγεται από τον αλγόριθμο της μεθόδου απότομης καθόδου για οποιοδήποτε $x^{(0)} \in \mathbb{R}^n$ και έστω x η λύση του συστήματος $Ax = b$, όπου ο $A \in \mathbb{R}^{n,n}$ είναι συμμετρικός και θετικά ορισμένος. Εστω $\kappa = \kappa_2(A) = \frac{\lambda_{\max}}{\lambda_{\min}}$, όπου $\lambda_{\max}$ και $\lambda_{\min}$ η μέγιστη και η ελάχιστη ιδιοτιμή του A, αντίστοιχα. Αν $e^{(k)} = x^{(k)} - x$ είναι το διάνυσμα-σφάλμα στην k επανάληψη τότε*

$$||e^{(k)}||_{A^{\frac{1}{2}}} \leq \left(\frac{\kappa - 1}{\kappa + 1}\right)^k ||e^{(0)}||_{A^{\frac{1}{2}}}, \; k = 1, 2, 3, \cdots. \tag{6.8}$$

Απόδειξη: Καταρχάς χρησιμοποιώντας τις (6.6) έχουμε ότι για οποιοδήποτε $\alpha \in I\!R$ προκύπτει αμέσως ότι

$$\begin{array}{l} ||e^{(k+1)}||^2_{A^{\frac{1}{2}}} \;=\; (Ae^{(k+1)}, e^{(k+1)}) \\ =\; (Ae^{(k+1)}, e^{(k)} + \alpha_{k+1} r^{(k)}) \;=\; (Ae^{(k+1)}, e^{(k)}) + \alpha_{k+1}(Ae^{(k+1)}, r^{(k)}) \\ =\; (Ae^{(k+1)}, e^{(k)}) + \alpha(Ae^{(k+1)}, r^{(k)}) \;=\; (Ae^{(k+1)}, e^{(k)} + \alpha r^{(k)}), \end{array} \tag{6.9}$$

όπου μπορέσαμε και αντικαταστήσαμε το συγκεκριμένο α_{k+1} με το οποιοδήποτε $\alpha \in I\!R$ αφού ο παράγοντας $(Ae^{(k+1)}, r^{(k)})$ είναι μηδέν.

Χρησιμοποιώντας την ανισότητα των Cauchy-Schwarz στο εσωτερικό γινόμενο (Ax_1, x_2), όπου $x_1, x_2 \in I\!R^n$, έχουμε $|(Ax_1, x_2)| = |(A^{\frac{1}{2}}x_1, A^{\frac{1}{2}}x_2)| \le ||A^{\frac{1}{2}}x_1||||A^{\frac{1}{2}}x_2|| = ||x_1||_{A^{\frac{1}{2}}}||x_2||_{A^{\frac{1}{2}}}$, οπότε, χρησιμοποιώντας το αποτέλεσμα που μόλις προέκυψε στην (6.9), παίρνουμε

$$||e^{(k+1)}||^2_{A^{\frac{1}{2}}} \;\le\; ||e^{(k+1)}||_{A^{\frac{1}{2}}}\, ||e^{(k)} + \alpha r^{(k)}||_{A^{\frac{1}{2}}}, \quad \forall\, \alpha \in I\!R,$$

ή, ισοδύναμα,

$$||e^{(k+1)}||^2_{A^{\frac{1}{2}}} \le ||e^{(k)} + \alpha r^{(k)}||^2_{A^{\frac{1}{2}}} = (A(e^{(k)} + \alpha r^{(k)}), e^{(k)} + \alpha r^{(k)}), \quad \forall\, \alpha \in I\!R.$$

Επειδή $e^{(k)} + \alpha r^{(k)} \;=\; (I - \alpha A)e^{(k)}$ και $A(e^{(k)} + \alpha r^{(k)}) \;=\; (I - \alpha A)Ae^{(k)}$ συμπεραίνουμε τελικά ότι

$$||e^{(k+1)}||^2_{A^{\frac{1}{2}}} \;\le\; \inf_{\alpha \in I\!R} \left((I - \alpha A)Ae^{(k)}, (I - \alpha A)e^{(k)}\right), \quad k \ge 0. \tag{6.10}$$

Εστω ότι $\lambda_i,\ i = 1(1)n$, είναι οι πραγματικές θετικές ιδιοτιμές του πραγματικού, συμμετρικού και θετικά ορισμένου πίνακα A στη διάταξη

$$0 \;<\; \lambda_{\min} \;:=\; \lambda_1 \;\le\; \lambda_2 \;\le\; \cdots \;\le\; \lambda_n \;=:\; \lambda_{\max} \tag{6.11}$$

και $z^i,\ i = 1(1)n$, τα αντίστοιχα ορθοκανονικά ιδιοδιανύσματά του, που μπορούν να παρθούν πραγματικά. Για κάθε $u \in I\!R^n$ προφανώς θα ισχύει $u = \sum_{i=1}^n (u, z^i) z^i$, και γενικότερα για κάθε πραγματικό πολυώνυμο $P(x)$ θα ισχύει $P(A)u \;=\; \sum_{i=1}^n P(\lambda_i)(u, z^i) z^i$. Συνεπώς, αν $e^{(k)} = \sum_{i=1}^n \beta_i z^i$, όπου $\beta_i = (e^{(k)}, z^i),\ i = 1(1)n$, τότε για το εσωτερικό γινόμενο που βρίσκεται στο infimum της (6.10) και για κάθε $\alpha \in I\!R$ μπορούμε να πάρουμε διαδοχικά

$$\begin{array}{l} \left((I - \alpha A)Ae^{(k)}, (I - \alpha A)e^{(k)}\right) \\ =\; ((I - \alpha A)A(\sum_{i=1}^n \beta_i z^i), (I - \alpha A)(\sum_{i=1}^n \beta_i z^i)) \\ =\; (\sum_{i=1}^n (1 - \alpha\lambda_i)\lambda_i \beta_i z^i, \sum_{i=1}^n (1 - \alpha\lambda_i)\beta_i z^i) \\ =\; \sum_{i=1}^n (1 - \alpha\lambda_i)^2 \lambda_i \beta_i^2 \\ \le\; (\max_{i=1(1)n} |1 - \alpha\lambda_i|)^2 \sum_{i=1}^n \lambda_i \beta_i^2 \\ =\; (\max_{i=1(1)n} |1 - \alpha\lambda_i|)^2 (\sum_{i=1}^n \lambda_i \beta_i z^i, \sum_{i=1}^n \beta_i z^i) \\ =\; (\max_{i=1(1)n} |1 - \alpha\lambda_i|)^2 (\sum_{i=1}^n \beta_i A z^i, \sum_{i=1}^n \beta_i z^i) \\ =\; (\max_{i=1(1)n} |1 - \alpha\lambda_i|)^2 (Ae^{(k)}, e^{(k)}) \\ =\; (\max_{i=1(1)n} |1 - \alpha\lambda_i|)^2 ||e^{(k)}||^2_{A^{\frac{1}{2}}}. \end{array}$$

Με βάση τις παραπάνω σχέσεις, από την (6.10) προκύπτει ότι

$$||e^{(k+1)}||_{A^{\frac{1}{2}}} \leq \inf_{\alpha\in I\!R}\left(\max_{\lambda\in[\lambda_{\min},\lambda_{\max}]}|1-\alpha\lambda|\right)||e^{(k)}||_{A^{\frac{1}{2}}}, \quad k\geq 0. \tag{6.12}$$

Το min-max πρόβλημα που έχουμε τώρα για επίλυση δεν είναι παρά το min-max πρόβλημα της τεχνικής της extrapolation, όπως αυτό τέθηκε και λύθηκε στη συγκεκριμένη περίπτωση του αντίστοιχου κεφαλαίου, πράγμα που μπορούμε να διαπιστώσουμε αμέσως από τη διάταξη (6.11) και τις εκφράσεις (3.29)-(3.30) αν θέσουμε $\alpha=\omega$. Το βέλτιστο α (α_β) βρίσκεται τότε, για $\lambda=\lambda_{\min}$ ή $\lambda_{\max}$, από την έκφραση (3.31). Συγκεκριμένα

$$\alpha_\beta = \frac{2}{\lambda_{\max}+\lambda_{\min}} \tag{6.13}$$

και με βάση την (6.13), η (6.12) γίνεται

$$||e^{(k+1)}||_{A^{\frac{1}{2}}} \leq \frac{\lambda_{\max}-\lambda_{\min}}{\lambda_{\max}+\lambda_{\min}}||e^{(k)}||_{A^{\frac{1}{2}}}$$

από την οποία παίρνουμε με απλή επαγωγή ότι

$$||e^{(k)}||_{A^{\frac{1}{2}}} \leq \left(\frac{\lambda_{\max}-\lambda_{\min}}{\lambda_{\max}+\lambda_{\min}}\right)^k||e^{(0)}||_{A^{\frac{1}{2}}}.$$

Από την τελευταία ανισότητα, επειδή στην προκείμενη περίπτωση ισχύει ότι $\kappa=\kappa_2(A)=\frac{\lambda_{\max}}{\lambda_{\min}}$, έχουμε τελικά

$$||e^{(k)}||_{A^{\frac{1}{2}}} \leq \left(\frac{\kappa-1}{\kappa+1}\right)^k||e^{(0)}||_{A^{\frac{1}{2}}}, \tag{6.14}$$

που αποδείχνει το θεώρημα. □

Σημειώσεις: α) Η σχέση (6.14) αποδείχνει ότι $\lim_{k\to\infty}e^{(k)}=0$ και άρα $\lim_{k\to\infty}x^{(k)}=A^{-1}b$. Δηλαδή η μέθοδος της απότομης καθόδου συγκλίνει ασυμπτωτικά στη λύση του θεωρηθέντος συστήματος. β) Από την ίδια σχέση συμπεραίνεται ότι όσο μεγαλύτερος είναι ο δείκτης κατάστασης $\kappa_2(A)$ τόσο βραδύτερη (ασυμπτωτικά) είναι δυνατόν να είναι η σύγκλιση.

6.3 Μέθοδος Γενικών Διευθύνσεων

Στη μέθοδο απότομης καθόδου η επιλογή των διευθύνσεων στην $k+1$ επανάληψη, κατά μήκος των οποίων ελαχιστοποιούμε την τιμή $f(x^{(k+1)})$ του δοθέντος συναρτησιακού, είναι πάντοτε η $r^{(k)}=-\nabla f(x^{(k)}),\ k=0,1,\cdots$. Αν προς στιγμήν θεωρήσουμε ότι αντί της $r^{(k)}$ επιλέγουμε μία "τυχαία" διεύθυνση $p^{(k+1)}\neq 0$ και προσπαθήσουμε να ακολουθήσουμε τη λογική της ελαχιστοποίησης της

τιμής του συναρτησιακού $f(x^{(k+1)})$, $x^{(k+1)} = x^{(k)} + \alpha p^{(k+1)}$, τότε μπορούμε να καταλήξουμε σε κάποια μέθοδο αντίστοιχη με αυτήν της απότομης καθόδου.

Πραγματικά, θεωρώντας το συναρτησιακό $f(x) = \frac{1}{2}(Ax, x) - (b, x)$ και αναπτύσσοντας την τιμή $f(x^{(k+1)})$, όπως προηγούμενα, θα έχουμε

$$\begin{aligned} f(x^{(k+1)}) = f(x^{(k)} + \alpha p^{(k+1)}) &= \tfrac{1}{2}(A(x^{(k)} + \alpha p^{(k+1)}), x^{(k)} + \alpha p^{(k+1)}) - (b, x^{(k)} + \alpha p^{(k+1)}) \\ &= \tfrac{1}{2}(Ax^{(k)}, x^{(k)}) + \alpha(Ax^{(k)}, p^{(k+1)}) + \tfrac{1}{2}\alpha^2(Ap^{(k+1)}, p^{(k+1)}) - (b, x^{(k)}) - \alpha(b, p^{(k+1)}) \\ &= f(x^{(k)}) + \tfrac{1}{2}\alpha^2(Ap^{(k+1)}, p^{(k+1)}) - \alpha(p^{(k+1)}, r^{(k)}). \end{aligned} \tag{6.15}$$

Η ελαχιστοποίηση του $f(x^{(k+1)})$ ως προς α πετυχαίνεται για $\alpha_{k+1} = \frac{(p^{(k+1)}, r^{(k)})}{(Ap^{(k+1)}, p^{(k+1)})}$, και άρα $f(x^{(k+1)}) = f(x^{(k)}) - \frac{(p^{(k+1)}, r^{(k)})^2}{2(Ap^{(k+1)}, p^{(k+1)})} < f(x^{(k)})$ ανν $(p^{(k+1)}, r^{(k)}) \neq 0$. Επομένως αν δεν υπάρξει k τ.ω. $r^{(k)} = 0$, οπότε η λύση $x = x^{(k)}$ θα έχει βρεθεί, τότε, με την προϋπόθεση ότι οι γενικές διευθύνσεις επιλέγονται έτσι ώστε $(p^{(k+1)}, r^{(k)}) \neq 0$, η ακολουθία των $f(x^{(k)})$ θα είναι αυστηρά φθίνουσα και φραγμένη εκ των κάτω από την τιμή $f(A^{-1}b)$ και άρα θα συγκλίνει έτσι ώστε $\lim_{k\to\infty} f(x^{(k)}) \geq f(A^{-1}b)$. Συνεπώς, αν $\lim_{k\to\infty} f(x^{(k)}) = f(A^{-1}b)$ τότε και $\lim_{k\to\infty} x^{(k)} = A^{-1}b$, αφού το σημείο $A^{-1}b$ είναι το μοναδικό σημείο στο οποίο η $f(x)$ παίρνει την ελάχιστη τιμή της, και άρα η ακολουθία των $x^{(k)}$ θα συγκλίνει στη λύση του συστήματος. Αν όμως $\lim_{k\to\infty} f(x^{(k)}) > f(A^{-1}b)$, τότε είτε η ακολουθία των $x^{(k)}$ θα συγκλίνει σε κάποιο σημείο $x' \neq A^{-1}b$ είτε δε θα συγκλίνει διόλου. Παρά το γεγονός αυτό, δίνουμε στη συνέχεια τον αντίστοιχο αλγόριθμο, καθώς και ορισμένες ιδιότητες της μεθόδου, γιατί η βασική ιδέα της παρούσας μεθόδου αποτελεί την αφετηρία σημαντικότερων μεθόδων ελαχιστοποίησης.

Ο αλγόριθμος της μεθόδου των γενικών διευθύνσεων που αναπτύχτηκε δίνεται στη συνέχεια.

Αλγόριθμος Μεθόδου Γενικών Διευθύνσεων:

Δεδομένα: $A \in I\!R^{n,n}$, $A^T = A$, A θετικά ορισμένος, $b \in I\!R^n$, $\epsilon \in I\!R^+$ το επιθυμητό σφάλμα.
$x^{(0)} = 0$
$r^{(0)} = b$
$k = 0$
Εφόσον $\|r^{(k)}\| > \epsilon$
    $k = k + 1$
    Επιλογή διεύθυνσης $p^{(k)}$: $(p^{(k)}, r^{(k-1)}) \neq 0$
    $\alpha_k = \frac{(p^{(k)}, r^{(k-1)})}{(Ap^{(k)}, p^{(k)})}$
    $x^{(k)} = x^{(k-1)} + \alpha_k p^{(k)}$
    $r^{(k)} = b - Ax^{(k)}$
Τέλος 'Εφόσον'
Αποτέλεσμα: $x = x^{(k)}$ η προσέγγιση της λύσης.

Οπως και στον αλγόριθμο της μεθόδου απότομης καθόδου υπάρχουν δύο πολλαπλασιασμοί πίνακα επί διάνυσμα ανά ανακύκλωση, οι $Ax^{(k)}$ και $Ap^{(k)}$, που αποτελούν και το κύριο κόστος του αλγόριθμου. Τελικά και επειδή ισχύει ότι $r^{(k)} = b - Ax^{(k)} = b - A(x^{(k-1)} + \alpha_k p^{(k)}) = r^{(k-1)} - \alpha_k Ap^{(k)}$, οι δυο πολλαπλασιασμοί μπορούν ουσιαστικά να αναχτούν σε έναν. Σημειώνεται ακόμη ότι αρκετές ανάλογες ιδιότητες με αυτές που βρέθηκαν αμέσως μετά τον αλγόριθμο της μεθόδου απότομης καθόδου ισχύουν. Π.χ.,

$$(r^{(k)}, p^{(k)}) \;=\; 0,$$

με απόδειξη αντίστοιχη αυτής για την $(r^{(k)}, r^{(k-1)}) = 0$, για δε το διάνυσμα-σφάλμα ισχύουν ότι $Ae^{(k)} \;=\; -r^{(k)}$, $(Ae^{(k)}, p^{(k)}) \;=\; 0$ και $e^{(k)} \;=\; e^{(k-1)} + \alpha_k p^{(k)}$.

6.4 Μέθοδος Συζυγών Διευθύνσεων (Conjugate Directions)

Και οι δύο μέθοδοι ελαχιστοποίησης που αναπτύχτηκαν ως τώρα έχουν ένα κοινό χαρακτηριστικό. Η οποιαδήποτε επανάληψη $x^{(k)}$ μπορεί να γραφτεί ως ένας γραμμικός συνδυασμός της αρχικής προσέγγισης $x^{(0)}$ και των διευθύνσεων που χρησιμοποιούνται. Συγκεκριμένα, ισχύει

$$x^{(k)} \;=\; x^{(k-1)} + \alpha_k r^{(k-1)} \;=\; x^{(k-2)} + \alpha_{k-1} r^{(k-2)} + \alpha_k r^{(k-1)} \;=\; x^{(0)} + \sum_{i=1}^{k} \alpha_i r^{(i-1)}$$

για τη μέθοδο της απότομης καθόδου και αντίστοιχα

$$x^{(k)} \;=\; x^{(0)} + \sum_{i=1}^{k} \alpha_i p^{(i)}$$

για τη μέθοδο των γενικών διευθύνσεων. Επομένως και η διαφορά $x^{(k)} - x^{(0)}$ αποτελεί ένα γραμμικό συνδυασμό των θεωρηθεισών προηγούμενων διευθύνσεων, ως προς τις οποίες γίνεται κάθε φορά η ελαχιστοποίηση της τιμής του συναρτησιακού. Φυσικά, οι διευθύνσεις αυτές, για $k > n$, είναι οπωσδήποτε γραμμικά εξαρτημένες. Γεννιέται, λοιπόν, το εύλογο ερώτημα μήπως θα ήταν δυνατόν να επιλεγούν n διευθύνσεις γραμμικά ανεξάρτητες και τ.ω. η ελαχιστοποίηση της τιμής του συναρτησιακού ως προς κάθε μία χωριστά να αποτελεί συγχρόνως και ελαχιστοποίηση της τιμής αυτού ως προς όλο τον υπόχωρο των μέχρι τότε θεωρηθεισών διευθύνσεων. Μία πρώτη διερεύνηση πιθανών δυνατοτήτων, όπως γίνεται στη συνέχεια, ίσως έδινε απάντηση στο ερώτημά μας.

Εστω ότι έχουν βρεθεί n γραμμικά ανεξάρτητες διευθύνσεις $p^{(i)},\ i = 1(1)n$, τ.ω.

$$x^{(n)} - x^{(0)} \;= \sum_{i=1}^{n} \alpha_i p^{(i)},$$

με $x^{(n)} = A^{-1}b$ τη λύση του γραμμικού συστήματος $Ax = b$. Το πρόβλημά μας τότε θα ήταν η εύρεση των συντελεστών $\alpha_i,\ i = 1(1)n$. Μια πρώτη πιθανή επιλογή θα ήταν να παρθούν τα διανύσματα $p^{(i)}$ έτσι ώστε να είναι ανά δύο ορθογώνια. Τότε όμως θα είχαμε $\alpha_i = \frac{(p^{(i)}, x^{(n)} - x^{(0)})}{(p^{(i)}, p^{(i)})}$. Δυστυχώς όμως στην περίπτωση αυτή το $x^{(n)} = A^{-1}b$ είναι άγνωστο κι αυτό λόγω της παρουσίας του A^{-1}. Η παρατήρηση αυτή ίσως δίνει και τη λύση στο πρόβλημά μας διότι αν τα $p^{(i)}, i = 1(1)n$, ήταν τ.ω.

$$(Ap^{(i)}, p^{(j)}) \begin{cases} > 0, & \alpha\nu\nu\ i = j, \\ = 0, & \alpha\nu\nu\ i \neq j, \end{cases} \tag{6.16}$$

τότε θα είχαμε

$$\alpha_i = \frac{(p^{(i)}, A(x^{(n)} - x^{(0)}))}{(Ap^{(i)}, p^{(i)})} \ = \ \frac{(p^{(i)}, r^{(0)})}{(Ap^{(i)}, p^{(i)})},\ i = 1(1)n, \tag{6.17}$$

σα συναρτήσεις γνωστών στοιχείων και μόνο. Η απάντηση στο πρόβλημα της εύρεσης τέτοιων $p^{(i)}$ δίνεται στη Γραμμική Αλγεβρα, όπου υπάρχει πρόταση που υποδείχνει τον τρόπο της κατασκευής τους και που είναι γνωστή ως Θεώρημα (ή αλγόριθμος) των Gram-Schmidt.

Πριν τη διατύπωση και απόδειξη του Θεώρηματος των Gram-Schmidt θα δώσουμε ένα γενικότερο ορισμό και με βάση αυτόν ένα λήμμα που θα χρησιμοποιήσουμε στο εν λόγω θεώρημα.

Ορισμός 6.1 *Εστω $A \in \mathbb{C}^{n,n} \setminus \{0\}$ Ερμιτιανός πίνακας. Δύο διανύσματα $u, v \in \mathbb{C}^n$ θα καλούνται "A-συζυγή" ή "A-ορθογώνια" ανν $(Au, v) = 0$.*

Σημείωση: Από τον ορισμό είναι φανερό ότι η γνωστή ορθογωνιότητα δύο διανυσμάτων δεν είναι παρά μερική περίπτωση της A-ορθογωνιότητας για $A = I$.

Λήμμα 6.1 *Εστω $A \in \mathbb{C}^{n,n}$ Ερμιτιανός και θετικά (ή αρνητικά) ορισμένος πίνακας. Αν $p^{(i)} \in \mathbb{C}^n \setminus \{0\},\ i = 1(1)k \leq n$, είναι ανά δύο A-ορθογώνια, τότε είναι γραμμικά ανεξάρτητα.*

Απόδειξη: Εστω $\sum_{i=1}^{k} c_i p^{(i)} = 0$, όπου $c_i \in \mathbb{C},\ i = 1(1)k$. Σχηματίζοντας το εσωτερικό γινόμενο των μελών της ισότητας επί $Ap^{(j)},\ j = 1(1)k$, έχουμε $(Ap^{(j)}, \sum_{i=1}^{k} c_i p^{(i)}) = 0$ ή $\sum_{i=1}^{k} c_i (Ap^{(j)}, p^{(i)}) = 0$ ή, ακόμη, $c_j (Ap^{(j)}, p^{(j)}) = 0$, αφού όλοι οι άλλοι όροι του αθροίσματος θα είναι μηδέν λογω της A-ορθογωνιότητας των $p^{(i)}$ και $p^{(j)}$ για $i \neq j$. Από την τελευταία ισότητα και επειδή $(Ap^{(j)}, p^{(j)}) > 0$ (ή < 0) προκύπτει ότι $c_j = 0,\ j = 1(1)k$, που συνεπάγεται τη γραμμική ανεξαρτησία των $p^{(i)},\ i = 1(1)k$. □

Θεώρημα 6.4 *(*Gram-Schmidt*) Εστω $A \in \mathbb{R}^{n,n},\ A^T = A$ και A θετικά ορισμένος. Εστω ακόμη πίνακας $U \in \mathbb{R}^{n,k}$, $k \leq n$, τα διανύσματα-στήλες $u^j,\ j = 1(1)k$, του οποίου είναι γραμμικά*

ανεξάρτητα. Υπάρχουν μοναδικοί πίνακες $P \in \mathbb{R}^{n,k}$ *με* $P^T AP = D = \mathrm{diag}(d_1, d_2, \cdots, d_k)$, *με* $d_j > 0,\ j = 1(1)k$, *και άνω τριγωνικός* $R \in \mathbb{R}^{k,k}$, *με* $r_{jj} = 1,\ j = 1(1)k$, *τ.ω.*

$$U = PR. \tag{6.18}$$

(Σημείωση: Αν $p^{(j)},\ j = 1(1)k$, *είναι τα διανύσματα-στήλες του* P, *τότε η σχέση* $P^T AP = D = \mathrm{diag}(d_1, d_2, \cdots, d_k)$, *με* $d_j > 0,\ j = 1(1)k$, *δεν αποτελεί παρά την απαίτησή μας να ικανοποιούνται οι (6.16). Να είναι δηλαδή τα* $p^{(j)}$ *μή μηδενικά και ανά δύο A-ορθογώνια (ή A-συζυγή), οπότε με βάση το προηγούμενο λήμμα θα είναι και γραμμικά ανεξάρτητα.)*

Απόδειξη: Η αποδεικτέα σχέση (6.18) γράφεται αναλυτικά ως εξής

$$\begin{bmatrix} u^1 & u^2 & u^3 & \cdots & u^k \end{bmatrix} = \begin{bmatrix} p^{(1)} & p^{(2)} & p^{(3)} & \cdots & p^{(k)} \end{bmatrix} \begin{bmatrix} 1 & r_{12} & r_{13} & \cdots & r_{1k} \\ & 1 & r_{23} & \cdots & r_{2k} \\ & & 1 & \cdots & r_{3k} \\ & & & \ddots & \vdots \\ & & & & 1 \end{bmatrix}, \tag{6.19}$$

και είναι ισοδύναμη με τις k ισότητες

$$\begin{cases} p^{(1)} = u^1 \\ r_{12}p^{(1)} + p^{(2)} = u^2 \\ r_{13}p^{(1)} + r_{23}p^{(2)} + p^{(3)} = u^3 \\ \vdots \\ r_{1k}p^{(1)} + r_{2k}p^{(2)} + r_{3k}p^{(3)} + \cdots + r_{k-1,k}p^{(k-1)} + p^{(k)} = u^k, \end{cases} \tag{6.20}$$

όπως μπορεί πολύ εύκολα να διαπιστωθεί. Η πρώτη ισότητα από τις (6.20) ορίζει το διάνυσμα $p^{(1)} (\neq 0)$, γιατί αν $p^{(1)} = 0$, τότε και $u^1 = 0$ και άρα το σύνολο των διανυσμάτων $u^j,\ j = 1(1)k$, θα ήταν γραμμικά εξαρτημένα, πράγμα που δεν ισχύει. Από τη δεύτερη ισότητα, για να έχουμε τα $p^{(1)}$ και $p^{(2)}$ A-ορθογώνια, πρέπει και αρκεί να ορίσουμε το r_{12} ως $r_{12} = \frac{(Ap^{(1)}, u^2)}{(Ap^{(1)}, p^{(1)})}$. Τότε το $p^{(2)}$ ορίζεται ως $p^{(2)} = u^2 - r_{12}p^{(1)} = u^2 - r_{12}u^1 \neq 0$, γιατί το $p^{(2)}$ είναι ίσο με ένα γραμμικό συνδυασμό των διανυσμάτων u^2 και u^1, τα οποία είναι γραμμικά ανεξάρτητα ως υποσύνολο των γραμμικά ανεξάρτητων διανυσμάτων $u^j,\ j = 1(1)k$. Η απόδειξη μπορεί να ολοκληρωθεί με επαγωγή και είναι προφανής. Η γενική έκφραση για το τυχόν $p^{(j)},\ j = 1(1)k$, θα δίνεται αναδρομικά από τις παρακάτω σχέσεις

$$p^{(j)} = u^j - \sum_{i=1}^{j-1} r_{ij}p^{(i)}, \quad r_{ij} = \frac{(Ap^{(i)}, u^j)}{(Ap^{(i)}, p^{(i)})},\ i = 1(1)j-1, \quad j = 1(1)k. \tag{6.21}$$

Η απόδειξη για τη μοναδικότητα των P και R μπορεί να προκύψει με θεώρηση ανάλυσης του U σε ένα άλλο γινόμενο $P'R'$, όπου οι P' και R' θα είναι αντίστοιχων μορφών με αυτές των P και

R στην (6.19) και με αντίστοιχες ιδιότητες. Η ισότητα $PR = P'R'$, η εξίσωση των αντίστοιχων στηλών των γινομένων τους και εφαρμογή απλής επαγωγής αποδείχνει τη μοναδικότητα. □

Από τις σχέσεις (6.20) μπορεί να προκύψει μια προφανής πρόταση, που ήδη χρησιμοποιήθηκε στην απόδειξη των (6.21) και θα χρησιμοποιηθεί και στη συνέχεια στην κατασκευή του αντίστοιχου αλγόριθμου της Μεθόδου των Συζυγών Διευθύνσεων.

Θεώρημα 6.5

$$(Ap^{(l)}, p^{(l)}) \;=\; (Ap^{(l)}, u^l - \sum_{j=1}^{l-1} r_{jl}p^{(j)}) \;=\; (Ap^{(l)}, u^l), \quad l = 1(1)k,\ k = 1(1)n. \tag{6.22}$$

Θα πρέπει να τονιστεί ότι κάποιες από τις ιδιότητες της μεθόδου των γενικών διευθύνσεων ισχύουν κι εδώ με τη μόνη διαφορά ότι μπορούν να επεκταθούν παραπέρα λόγω της A-ορθογωνιότητας των διευθύνσεων $p^{(j)},\ j = 1(1)n$. Π.χ., ισχύει η παρακάτω πρόταση.

Θεώρημα 6.6 $(p^{(k)}, r^{(k-1)}) \;=\; (p^{(k)}, r^{(j)}),\ j = 0(1)k-1,\ k = 1(1)n.$

Απόδειξη: Επειδή $r^{(k-1)} = r^{(k-2)} - \alpha_{k-1}Ap^{(k-1)}$ έχουμε ότι

$$(p^{(k)}, r^{(k-1)}) = (p^{(k)}, r^{(k-2)}) - \alpha_{k-1}(p^{(k)}, Ap^{(k-1)}) = (p^{(k)}, r^{(k-2)}),$$

αφού $(p^{(k)}, Ap^{(k-1)}) = 0$ λόγω της A-ορθογωνιότητας των $p^{(k)}$ και $p^{(k-1)}$. Απλή επαγωγή ολοκληρώνει την απόδειξη. □

Σημείωση: Με βάση την πρόταση που μόλις αποδείχτηκε, και που αποδείχνει συγχρόνως ότι η Μέθοδος των Συζυγών Διευθύνσεων περατώνεται σε n το πολύ βήματα, θα μπορέσουμε να αντικαταστήσουμε στην έκφραση του συντελεστή α_i, της (6.17), το $(p^{(i)}, r^{(0)})$ με $(p^{(i)}, r^{(i-1)})$.

Εχοντας υπόψη τη μέχρι τώρα ανάλυση και τα Θεωρήματα 6.5 και 6.6, ο αλγόριθμος της μεθόδου των συζυγών διευθύνσεων, με ενσωματωμένο τον αλγόριθμο των Gram-Schmidt, μπορεί να δοθεί αμέσως και έχει ως εξής:

Αλγόριθμος Μεθόδου Συζυγών Διευθύνσεων:

Δεδομένα: $A \in I\!R^{n,n},\ A^T = A, A$ θετικά ορισμένος, $b \in I\!R^n,\ U \in I\!R^{n,n}$, με διανύσματα-στήλες $u^j,\ j = 1(1)n$, γραμμικά ανεξάρτητα, $\epsilon \in I\!R^+$ το επιθυμητό σφάλμα
$x^{(0)} = 0$
$r^{(0)} = b$
$k = 0$
Εφόσον $\|r^{(k)}\| > \epsilon$ και $k < n$

$k = k + 1$
$p^{(k)} = u^k$
Για $j = 1(1)k - 1$
$r_{jk} = \frac{(Ap^{(j)}, p^{(k)})}{(Ap^{(j)}, p^{(j)})}$
$p^{(k)} = p^{(k)} - r_{jk} p^{(j)}$
Τέλος 'Για'
$\alpha_k = \frac{(p^{(k)}, r^{(k-1)})}{(Ap^{(k)}, p^{(k)})}$
$x^{(k)} = x^{(k-1)} + \alpha_k p^{(k)}$
$r^{(k)} = b - Ax^{(k)} \ \ (= r^{(k-1)} - \alpha_k A p^{(k)})$
Τέλος 'Εφόσον'
Αποτέλεσμα: $x = x^{(k)}$ η προσέγγιση της λύσης

Σημείωση: Στον Αλγόριθμο που μόλις δόθηκε, το διάνυσμα $p^{(k)}$ στην έκφραση του r_{jk} και αυτό στα μέλη της σχέσης $p^{(k)} = p^{(k)} - r_{jk} p^{(j)}$ **δεν** είναι το τελικά ευρεθησόμενο διάνυσμα $p^{(k)}$ αλλά απλά ένα διάνυσμα που είναι εκάστοτε ίσο με $u^k - \sum_{i=1}^{j} r_{ik} p^{(i)}, \ j = 1(1)k - 1$.

Η Σημείωση, που ακολουθεί το Θεώρημα 6.6 σε συνδυασμό με την (6.17), αποδείχνει ότι με ακριβή αριθμητική ο Αλγόριθμος των Συζυγών Διευθύνσεων περατώνεται σε n επαναλήψεις. Ο ισχυρισμός αυτός μπορεί να καταστεί πιο φανερός και με το παρακάτω θεώρημα.

Θεώρημα 6.7 *Ο Αλγόριθμος της Μεθόδου των Συζυγών Διευθύνσεων περατώνεται σε n το πολύ επαναλήψεις. Δηλαδή*

$$r^{(n)} = 0.$$

Απόδειξη: Από την κατασκευή του αλγόριθμου και το Θεώρημα 6.6 έχουμε ότι

$$x^{(n)} \ = \ x^{(0)} + \sum_{k=1}^{n} \alpha_k p^{(k)} \iff x^{(n)} - x^{(0)} \ = \ \sum_{k=1}^{n} \frac{(p^{(k)}, r^{(0)})}{(Ap^{(k)}, p^{(k)})} p^{(k)}.$$

Αρα

$$(Ap^{(i)}, x^{(n)} - x^{(0)}) \ = \ (Ap^{(i)}, \sum_{k=1}^{n} \frac{(p^{(k)}, r^{(0)})}{(Ap^{(k)}, p^{(k)})} p^{(k)}) \ = \ (p^{(i)}, r^{(0)}), \ i = 1(1)n, \tag{6.23}$$

ή επειδή

$$(p^{(i)}, A(x^{(n)} - x^{(0)})) \ = \ (p^{(i)}, r^{(0)} - r^{(n)}) \ = \ (p^{(i)}, r^{(0)}) - (p^{(i)}, r^{(n)}), \ i = 1(1)n, \tag{6.24}$$

έχουμε από την ισότητα των αριστερών μελών των (6.23) και (6.24) ότι

$$(p^{(i)}, r^{(n)}) = 0, \ i = 1(1)n.$$

Λόγω της γραμμικής ανεξαρτησίας των $p^{(i)},\ i = 1(1)n$, η τελευταία ισότητα συνεπάγεται ότι $r^{(n)} = 0$. □

Οπως και στις δυο προηγούμενες μεθόδους οι δύο πολλαπλασιασμοί πίνακα επί διάνυσμα ανά ανακύκλωση μπορούν να συμπτυχτούν σε έναν, όπως υποδείχνεται στον αλγόριθμο, αφού μπορεί να δειχτεί αμέσως ότι $r^{(k)} = r^{(k-1)} - \alpha_k A p^{(k)}$. Σημειώσεις: α) Οπως αποδείχτηκε στο Θεώρημα 6.6 το εσωτερικό γινόμενο $(p^{(k)}, r^{(0)})$ στον αλγόριθμο της μεθόδου των συζυγών διευθύνσεων μπορεί να αντικατασταθεί με $(p^{(k)}, r^{(k-1)})$, όπως δηλαδή αυτό εμφανίζεται στον αλγόριθμο της μεθόδου των γενικών διευθύνσεων. β) Εκτός από τον ορισμό των $p^{(k)}$, τον περιορισμό του στα n βήματα και την εύρεση της λύσης $x^{(n)} = A^{-1}b$, ο αλγόριθμος της μεθόδου των συζυγών διευθύνσεων είναι σχεδόν ταυτόσημος με αυτόν των γενικών διευθύνσεων. γ) Στην προκείμενη περίπτωση, σε αντίθεση με αυτήν των γενικών διευθύνσεων, είναι δυνατόν να είναι $(p^{(k)}, r^{(k-1)}) = 0$, οπότε με βάση το θεώρημα θα είναι και $(p^{(k)}, r^{(j)}) = 0,\ j = 0(1)k-2$. Αυτό απλά σημαίνει ότι το διάνυσμα-υπόλοιπο $r^{(k-1)}$, καθώς και όλα τα προηγούμενα διανύσματα-υπόλοιπα, έχουν συνιστώσες 0 στην κατεύθυνση $p^{(k)}$ (ή είναι ορθογώνια στο $p^{(k)}$).

Τέλος δίνεται ακόμη μια Πρόταση, που είναι πολύ χρήσιμη, κυρίως δε, στη Μέθοδο των Συζυγών Κλίσεων που θα ακολουθήσει.

Θεώρημα 6.8 *Ορίζοντας $p^{(0)} = 0$ έχουμε ότι για $j = 0(1)k-1$, $k = 1(1)n$, ισχύει*

$$(r^{(k-1)}, p^{(j)}) = 0.$$

Απόδειξη: Η απόδειξη γίνεται με επαγωγή. Για $k = 1$ έχουμε $j = 0$ και άρα $(r^{(0)}, p^{(0)}) = 0$, αφού $p^{(0)} = 0$, και επομένως η πρόταση ισχύει. Υποθέτουμε ότι η πρόταση ισχύει για κάποια τιμή του k και αποδείχνουμε ότι ισχύει και για την επόμενη τιμή $k+1$. Διακρίνουμε δύο περιπτώσεις. Για $j \leq k-1$ έχουμε διαδοχικά

$$(r^{(k)}, p^{(j)}) \ = \ (r^{(k-1)} - \alpha_k A p^{(k)}, p^{(j)}) \ = \ (r^{(k-1)}, p^{(j)}) - \alpha_k (A p^{(k)}, p^{(j)}) \ = 0,$$

κι αυτό γιατί το πρώτο εσωτερικό γινόμενο είναι μηδέν από την υπόθεση της τέλειας επαγωγής και το δεύτερο είναι μηδέν από την A-ορθογωνιότητα των συζυγών διεθύνσεων. Για $j = k$ έχουμε

$$(r^{(k)}, p^{(k)}) \ = \ (r^{(k-1)} - \alpha_k A p^{(k)}, p^{(k)}) \ = \ (r^{(k-1)}, p^{(k)}) - \frac{(p^{(k)}, r^{(k-1)})}{(A p^{(k)}, p^{(k)})} (A p^{(k)}, p^{(k)}) \ = \ 0.$$

□

Πόρισμα 6.2 *Αν $e^{(k)},\ k = 1(1)n$, είναι το διάνυσμα-σφάλμα στην k επανάληψη ισχύει ότι $(A e^{(k-1)}, p^{(j)}) = 0,\ j = 1(1)k-1$.*

Είναι φανερό ότι η μέθοδος των συζυγών διευθύνσεων πλεονεκτεί σαφώς έναντι των προηγούμενων μεθόδων διότι βρίσκει την ακριβή λύση σε n επαναλήψεις. Η πολυπλοκότητα της μεθόδου, αν περιοριστούμε στους πολλαπλασιασμούς και μόνο, είναι $\mathcal{O}(n^3)$ και ο ασυμπτωτικός συντελεστής της είναι $\frac{3}{2}$ έναντι $\frac{1}{3}$ που είναι στη μέθοδο απαλοιφής του Gauss. (Σημείωση: Ο ασυμπτωτικός συντελεστής $\frac{3}{2}$ προκύπτει από το άθροισμα των ασυμπτωτικών συντελεστών 1 για την εύρεση των $Ap^{(k)},\ k = 1(1)n$, και $\frac{1}{2}$ για την εύρεση των $(Ap^{(j)}, u^{k+1}),\ j = 1(1)k, k = 1(1)n-1$.) Συνεπώς αν κάτι θα πρέπει να μας προβληματίσει ως προς τη δυνατότητα βελτίωσης της παρούσας μεθόδου είναι η κατασκευή μιας νέας μεθόδου πάνω στην ιδέα των συζυγών διευθύνσεων η οποία όμως θα "εκμεταλλεύεται" την "αυθαιρεσία" της θεώρησης του πίνακα U για την κατασκευή των $p^{(j)}$.

Αν και είναι διαισθητικά φανερό δεν έχει αποδειχτεί τυπικά με τη μέχρι τώρα ανάλυση ότι η παρούσα μέθοδος είναι καταρχάς μια μέθοδος ελαχιστοποίησης όχι μόνο ως προς τη διεύθυνση $p^{(k)}$, όπως συμβαίνει, αλλά και ως προς το μέχρις εκείνη τη στιγμή θεωρηθέντα υπόχωρο $\text{span}\{p^{(1)}, p^{(2)},\ \cdots,\ p^{(k)}\}$. Γι' αυτό, εξάλλου, βρίσκει την ακριβή λύση σε n επαναλήψεις. Για το σκοπό αυτό θα πρέπει, λοιπόν, η κάθε νέα προσέγγιση $x^{(k)}$ να βρίσκεται όχι μόνο από τη $x^{(k-1)}$ με ελαχιστοποίηση της τιμής του δοθέντος συναρτησιακού $f(x)$ κατά μήκος της $p^{(k)}$, όπως συμβαίνει, αλλά και με τη σύγχρονη ελαχιστοποίηση της τιμής $f(x^{(k)})$ στον υπόχωρο $\text{span}\{p^{(1)},\ p^{(2)},\ \cdots,\ p^{(k)}\}$. Πράγματι, τότε το $x^{(n)}$ θα ελαχιστοποιεί το $f(x)$ για $x \in\ \text{span}\{p^{(1)},\ p^{(2)},\ \cdots,\ p^{(n)}\} \equiv I\!R^n$ και άρα θα δίνει και τη λύση $A^{-1}b$ του συστήματος. Η απόδειξη μπορεί να γίνει με επαγωγή και δίνεται στη συνέχεια.

Για $k = 1$ η τιμή του $x \in\ \text{span}\{p^{(1)}\}$, που ελαχιστοποιεί το συναρτησιακό $f(x)$ στη διεύθυνση $p^{(1)}$, είναι η $x^{(1)} = x^{(0)} + \alpha p^{(1)}$ με $\alpha = \alpha_1 = \frac{(p^{(1)}, r^{(0)})}{(Ap^{(1)}, p^{(1)})}$ και προφανώς είναι η ίδια με αυτήν που ελαχιστοποιεί το συναρτησιακό στον υπόχωρο $\text{span}\{p^{(1)}\}$. Για το γενικό βήμα της μετάβασης από το $k-1$ στο k $(1 < k \leq n)$ εργαζόμαστε ως εξής. Εστω ότι το $x^{(k-1)}$ ελαχιστοποιεί το $f(x)$ κατά μήκος της διεύθυνσης $p^{(k-1)}$ και συγχρόνως στον υπόχωρο $\text{span}\{p^{(1)},\ p^{(2)},\ \cdots,\ p^{(k-1)}\}$ και έστω $x \in \text{span}\{p^{(1)},\ p^{(2)},\ \cdots,\ p^{(k)}\}$. Εστω ακόμη ότι $x = y + \alpha p^{(k)}$, όπου $y \in\ \text{span}\{p^{(1)},\ p^{(2)},\ \cdots,\ p^{(k-1)}\}$. Θα ισχύει $y = \sum_{i=1}^{k-1} \beta_i p^{(i)}$, όπου $\beta_i,\ i = 1(1)k-1$, πραγματικοί αριθμοί, και επομένως θα έχουμε διαδοχικά

$$\begin{aligned} f(x) &= f(y + \alpha p^{(k)}) = \tfrac{1}{2}(A(y + \alpha p^{(k)}), y + \alpha p^{(k)}) - (b, y + \alpha p^{(k)}) \\ &= f(y) + \tfrac{1}{2}(Ap^{(k)}, p^{(k)})\alpha^2 - (b, p^{(k)})\alpha + \alpha(Ay, p^{(k)}) \\ &= f(y) + \tfrac{1}{2}(Ap^{(k)}, p^{(k)})\alpha^2 - (b, p^{(k)})\alpha, \end{aligned}$$

αφού $(Ay, p^{(k)}) = \sum_{i=1}^{k-1} \beta_i (Ap^{(i)}, p^{(k)}) = 0$, λόγω της A-ορθογωνιότητας των $p^{(i)}$ με το $p^{(k)}$.

Παρατηρούμε τώρα ότι η $f(x)$ αποτελείται από το άθροισμα δύο ανεξάρτητων συναρτήσεων. Της $f(y)$, που είναι συνάρτηση του $y \in\ \text{span}\{p^{(1)},\ p^{(2)},\ \cdots,\ p^{(k-1)}\}$, και της $\frac{1}{2}(Ap^{(k)}, p^{(k)})\alpha^2 - (b, p^{(k)})\alpha$, που είναι συνάρτηση του α μόνο. Αρα για την ελαχιστοποίησή της πρέπει και αρκεί να ελαχιστοποιηθεί κάθε μία ξεχωριστά. Η μεν πρώτη έχει ήδη ελαχιστοποιηθεί στον υπόχωρο $\text{span}\{p^{(1)},\ p^{(2)},\ \cdots,\ p^{(k-1)}\}$, από την υπόθεση της τέλειας επαγωγής, για $y = x^{(k-1)}$. Η δε δεύτερη ελαχιστοποιείται για $\alpha = \alpha_k = \frac{(b, p^{(k)})}{(Ap^{(k)}, p^{(k)})}$, δηλαδή για την τιμή $\alpha = \alpha_k = \frac{(p^{(k)}, r^{(0)})}{(Ap^{(k)}, p^{(k)})}$. Η

τελευταία είναι ίση με $\frac{(p^{(k)},r^{(k-1)})}{(Ap^{(k)},p^{(k)})}$, από το Θεώρημα 6.6, και δίνεται αντί της προηγούμενης στον αλγόριθμο της μεθόδου των συζυγών διευθύνσεων.

Η πρόταση που μόλις αποδείχτηκε δίνεται στη συνέχεια υπό μορφή θεωρήματος.

Θεώρημα 6.9 *Στην k επανάληψη της μεθόδου των συζυγών διευθύνσεων η $x^{(k)}$, $k = 1(1)n$, επιλύει το πρόβλημα της μονοδιάστατης ελαχιστοποίησης στη διεύθυνση $p^{(k)}$*

$$\min_{x=x^{(k-1)}+\alpha p^{(k)}} f(x)$$

και συγχρόνως το πρόβλημα της ελαχιστοποίησης στον υπόχωρο $\text{span}\{p^{(1)},\ p^{(2)},\ \cdots,\ p^{(k)}\}$

$$\min_{x\in \text{span}\{p^{(1)},\ p^{(2)},\ \cdots,\ p^{(k)}\}} f(x). \tag{6.25}$$

Σημείωση: Το παραπάνω θεώρημα ισχύει για $x^{(0)} = 0$. Αν είναι $x^{(0)} \neq 0$, τότε το ελάχιστο στην (6.25) βρίσκεται για $x - x^{(0)} \in \text{span}\{p^{(1)},\ p^{(2)},\ \cdots,\ p^{(k)}\}$. Αυτό γιατί στην περίπτωση $x^{(0)} \neq 0$ μπορεί να θεωρηθεί το σύστημα $A\widehat{x}\ \left(= A(x - x^{(0)}) = b - Ax^{(0)}\right)\ =\ \widehat{b}$, οπότε η ελαχιστοποίηση του αντίστοιχου συναρτησιακού με $\widehat{x}^{(0)} = 0$ οδηγεί στο βασικό συμπέρασμα του θεωρήματος.

6.5 Μέθοδος Συζυγών Κλίσεων (Conjugate Gradients)

Η μέθοδος των συζυγών κλίσεων, που προτάθηκε από τους Hestenes και Stiefel (ο αναγνώστης παρεπέμπεται στο βιβλίο της Greenbaum [22]), είναι μία μέθοδος συζυγών διευθύνσεων απαλλαγμένη από τα μειονεκτήματα της μεθόδου, όπως αυτά τονίστηκαν στην προηγούμενη παράγραφο. Καταρχάς, ως στήλες του πίνακα U στον αλγόριθμο (θεώρημα) των Gram-Schmidt παίρνονται τα διαδοχικά διανύσματα-υπόλοιπα $u^j = r^{(j-1)},\ j = 1(1)k,\ k = 1(1)n$. Θα διαπιστωθεί αμέσως μετά, στην Πρόταση 7 του Θεωρήματος 6.10, ότι για την εύρεση της νέας διεύθυνσης $p^{(k)}$ χρειάζεται μόνο η προηγούμενη $p^{(k-1)}$ και όχι **όλες** οι προηγούμενες διευθύνσεις, όπως στη μέθοδο των συζυγών διευθύνσεων. Μπορεί να αποδειχτεί, ακόμη, ότι είναι η βέλτιστη μέθοδος μεταξύ όλων των μεθόδων συζυγών διευθύνσεων. Αυτό, γιατί η εκάστοτε νέα διεύθυνση $p^{(k)}$ δεν ανήκει απλά στον υπόχωρο που είναι ορθογώνιος στο $\text{span}\{Ap^{(1)},\ Ap^{(2)},\ \cdots,\ Ap^{(k-1)}\},\ k = 1(1)n$, αλλά αποτελεί την ορθή προβολή του $r^{(k-1)}$ στον εν λόγω υπόχωρο. Μερικοί από τους παραπάνω ισχυρισμούς που διατυπώθηκαν θα αποδειχτούν στη συνέχεια. Για το σκοπό αυτό και για να είναι συγκεντρωμένες όλες οι σχετικές προτάσεις, μερικές από τις οποίες είναι ήδη γνωστές αφού ισχύουν για οποιαδήποτε μέθοδο συζυγών διευθύνσεων, πρώτα θα διατυπωθούν και μετά, όσες από αυτές παρουσιάζονται για πρώτη φορά, θα αποδειχτούν.

Θεώρημα 6.10 *Στη μέθοδο συζυγών κλίσεων ισχύουν οι παρακάτω προτάσεις για $k = 1(1)n$, $p^{(0)} = 0$.*

Πρόταση 1: $u^j = r^{(j-1)}$, $j = 1(1)k$.
Πρόταση 2: $(r^{(k-1)}, p^{(j)}) = 0$, $j = 0(1)k-1$, $p^{(0)} = 0$.
Πρόταση 3: $(r^{(k-1)}, r^{(j-1)}) = (r^{(k-1)}, u^j) = 0$, $j = 1(1)k-1$.
Πρόταση 4: $(r^{(k-1)}, r^{(k-1)}) = (r^{(k-1)}, u^k) = (r^{(k-1)}, p^{(k)})$.
Πρόταση 5: $\alpha_k = \frac{(p^{(k)}, r^{(k-1)})}{(Ap^{(k)}, p^{(k)})} = \frac{(r^{(k-1)}, r^{(k-1)})}{(Ap^{(k)}, p^{(k)})}$.
Πρόταση 6: $r^{(k)} = r^{(k-1)} - \alpha_k A p^{(k)}$.
Πρόταση 7: Αν r_{jk}, $j = 1(1)k$, *είναι τα στοιχεία της* k *στήλης του άνω τριγωνικού πίνακα* R *της (6.18), τότε ισχύει*

$$r_{jk} = \begin{cases} 0, & j = 1(1)k-2, \\ -\frac{(r^{(k-1)}, r^{(k-1)})}{(r^{(k-2)}, r^{(k-2)})}, & j = k-1, \\ 1, & j = k. \end{cases}$$

Απόδειξη: Πρόταση 1: Ισχύει εξ ορισμού.
Πρόταση 2: Είναι το Θεώρημα 6.8.
Πρόταση 3: Εχουμε $(r^{(k-1)}, r^{(j-1)}) = (r^{(k-1)}, u^j) = (r^{(k-1)}, r_{1j}p^{(1)} + \cdots + r_{j-1,j}p^{(j-1)} + p^{(j)}) = 0$. Η πρώτη ισότητα από τα αριστερά ισχύει λόγω της Πρότασης 1, η δεύτερη από τον αλγόριθμο (θεώρημα) των Gram-Schmidt και η τρίτη λόγω της Πρότασης 2. (Σημείωση: Η παρούσα πρόταση εξασφαλίζει τη γραμμική ανεξαρτησία των διανυσμάτων $r^{(j-1)}$, $j = 1(1)k$, υπό την προϋπόθεση ότι είναι διάφορα από το μηδέν και άρα έχει εφαρμογή το Θεώρημα (και ο Αλγόριθμος) των Gram-Schmidt.)
Πρόταση 4: Αποδείχνεται ακριβώς με τον ίδιο τρόπο όπως η προηγούμενη. Η μόνη διαφορά είναι ότι το τελευταίο εσωτερικό γινόμενο $(r^{(k-1)}, p^{(k)})$ δεν είναι μηδέν.
Πρόταση 5: Είναι $\alpha_k = \frac{(p^{(k)}, r^{(k-1)})}{(Ap^{(k)}, p^{(k)})}$ από τον αλγόριθμο των συζυγών διευθύνσεων και το Θεώρημα 6.6. Αν αντικατασταθεί ο αριθμητής με βάση το συμπέρασμα της Πρότασης 4 αποδείχνεται το ζητούμενο.
Πρόταση 6: Είναι η έκφραση του υπολοίπου $r^{(k)}$ από τον αλγόριθμο των συζυγών διευθύνσεων.
Πρόταση 7: Από τον αλγόριθμο των Gram-Schmidt και τη συγκεκριμένη επιλογή των u^j έχουμε $r_{jk} = \frac{(Ap^{(j)}, u^k)}{(Ap^{(j)}, p^{(j)})} = \frac{(Ap^{(j)}, r^{(k-1)})}{(Ap^{(j)}, p^{(j)})}$. Αντικαθιστώντας το $Ap^{(j)}$ στον αριθμητή από την Πρόταση 6 και διασπώντας σε δυο κλάσματα έχουμε ότι

$$r_{jk} = \frac{1}{\alpha_j}\frac{(r^{(j-1)} - r^{(j)}, r^{(k-1)})}{(Ap^{(j)}, p^{(j)})} = \frac{1}{\alpha_j (Ap^{(j)}, p^{(j)})}\left[(r^{(j-1)}, r^{(k-1)}) - (r^{(j)}, r^{(k-1)})\right].$$

Διακρίνουμε τρεις περιπτώσεις. Αν $j < k-1$, τότε και τα δυο εσωτερικά γινόμενα μέσα στις αγκύλες είναι μηδέν από την Πρόταση 3. Αν $j = k-1$, τότε το πρώτο εσωτερικό γινόμενο μέσα στις αγκύλες είναι πάλι μηδέν και άρα $r_{k-1,k} = -\frac{1}{\alpha_{k-1}}\frac{(r^{(k-1)}, r^{(k-1)})}{(Ap^{(k-1)}, p^{(k-1)})}$, οπότε με τη χρησιμοποίηση της δεύτερης έκφρασης του α_{k-1} από την Πρόταση 5 βρίσκεται το ζητούμενο. Τέλος, για $j = k$

έχουμε $r_{kk} = 1$ από τον αλγόριθμο των Gram-Schmidt, πράγμα που μπορεί να προκύψει αμέσως και από το συνδυασμό των Προτάσεων 6, 3, 4 και 5. □

Στη συνέχεια δίνεται ο αλγόριθμος της μεθόδου των συζυγών κλίσεων. Οπως είναι δυνατόν να παρατηρήσει κανείς είναι σχεδόν ταυτόσημος με αυτόν των συζυγών διευθύνσεων εκτός από το τμήμα του που αφορά στην εύρεση της νέας διεύθυνσης $p^{(k)}$, που βρίσκεται, όπως είδαμε από την Πρόταση 7, μόνο από την προηγούμενη διεύθυνση και όχι σα συνάρτηση όλων των προηγούμενων διευθύνσεων.

Αλγόριθμος Μεθόδου Συζυγών Κλίσεων:

Δεδομένα: $A \in I\!R^{n,n},\ A^T = A,\ A$ θετικά ορισμένος, $b \in I\!R^n,\ \epsilon \in I\!R^+$ το επιθυμητό σφάλμα
$x^{(0)} = 0$
$r^{(0)} = b$
$p^{(1)} = r^{(0)}$
$\alpha_1 = \frac{(r^{(0)}, r^{(0)})}{(Ap^{(1)}, p^{(1)})}$
$x^{(1)} = x^{(0)} + \alpha_1 p^{(1)}$
$r^{(1)} = b - Ax^{(1)}\ (= r^{(0)} - \alpha_1 Ap^{(1)})$
$k = 1$
Εφόσον $\|r^{(k)}\| > \epsilon$ και $k < n$
$k = k + 1$
$\beta_k = \frac{(r^{(k-1)}, r^{(k-1)})}{(r^{(k-2)}, r^{(k-2)})}$
$p^{(k)} = r^{(k-1)} + \beta_k p^{(k-1)}$
$\alpha_k = \frac{(r^{(k-1)}, r^{(k-1)})}{(Ap^{(k)}, p^{(k)})}$
$x^{(k)} = x^{(k-1)} + \alpha_k p^{(k)}$
$r^{(k)} = b - Ax^{(k)}\ \ (= r^{(k-1)} - \alpha_k Ap^{(k)})$
Τέλος 'Εφόσον'
Αποτέλεσμα: $x = x^{(k)}$ η προσέγγιση της λύσης.

Στη συνέχεια δίνουμε δυο λήμματα που χρησιμεύουν στην εύρεση ενός άνω φράγματος για το απόλυτο σφάλμα της k επανάληψης της μεθόδου των συζυγών κλίσεων σα συνάρτηση του απόλυτου αρχικού σφάλματος.

Λήμμα 6.2 *Αν* $x^{(0)} = 0$, *τότε για* $j = 1(1)k$ *και* $k = 1(1)n$ *ισχύει ότι*

$$\text{span}\{p^{(1)},\ p^{(2)},\ \cdots,\ p^{(j)}\} = \text{span}\{r^{(0)},\ r^{(1)},\ \cdots,\ r^{(j-1)}\} = \text{span}\{b,\ Ab,\ \cdots,\ A^{j-1}b\}. \quad (6.26)$$

Απόδειξη: Από τον αλγόριθμο (θεώρημα) των Gram-Schmidt και όταν τα $u^j,\ j = 1(1)k$, είναι γραμμικά ανεξάρτητα έχουμε ότι

$$\text{span}\{p^{(1)},\ p^{(2)},\ \cdots,\ p^{(j)}\} = \text{span}\{u^1,\ u^2,\ \cdots,\ u^j\},\ j = 1(1)k.$$

Επειδή στη μέθοδο των συζυγών κλίσεων είναι $u^j = r^{(j-1)},\ j = 1(1)k$, με την προϋπόθεση ότι κανένα από τα $r^{(j-1)},\ j = 1(1)k$, δεν είναι μηδέν, από την Πρόταση 3 έπεται ότι αυτά θα είναι γραμμικά ανεξάρτητα. Αρα αποδείχνεται αμέσως η ισχύς της πρώτης ισότητας από τις (6.26). Για την απόδειξη της δεύτερης ισότητας παρατηρούμε ότι για $x^{(0)} = 0$ είναι $p^{(1)} = r^{(0)} = b$ και η πρόταση ισχύει για $j = 1$. Εστω ότι η πρόταση ισχύει για κάποιο $j = 1(1)k - 1$, δηλαδή ότι

$$(\text{span}\{p^{(1)},\ p^{(2)},\ \cdots,\ p^{(j)}\} =)\ \text{span}\{r^{(0)},\ r^{(1)},\ \cdots,\ r^{(j-1)}\}\ =\ \text{span}\{b,\ Ab,\ \cdots,\ A^{j-1}b\}.$$

Από τον αλγόριθμο των συζυγών κλίσεων έχουμε ότι $r^{(j)} = r^{(j-1)} - \alpha_j A p^{(j)}$. Αλλά $r^{(j-1)} \in \text{span}\{b,\ Ab,\ \cdots, A^{j-1}b\}$ και $Ap^{(j)} \in \text{span}\{Ab,\ A^2b,\ \cdots,\ A^jb\},\ j = 1(1)k-1$, από την προηγούμενη υπόθεση. Επομένως $r^{(j)} \in \text{span}\{b,\ Ab,\ \cdots, A^jb\}$. Προφανώς, όμως, $r^{(j)} \in \text{span}\{r^{(0)},\ r^{(1)},\ \cdots,\ r^{(j)}\}$, που είναι διάστασης $j+1$. Αρα και ο υπόχωρος που παράγεται από τα $j+1$ διανύσματα $b, Ab, \cdots, A^jb$, ο $\text{span}\{b,\ Ab,\ \cdots, A^jb\}$, θα είναι της ίδιας διάστασης. Συνεπώς θα ισχύουν οι (6.26) και για $j+1$, πράγμα που ολοκληρώνει την επαγωγική απόδειξη. □

Σημειώσεις: α) Στο προηγούμενο λήμμα υποτέθηκε ότι $x^{(0)} = 0$. Αν $x^{(0)} \neq 0$, τότε για $j = 1(1)k$ το $x^{(j)}$ λύνει το πρόβλημα ελαχιστοποίησης

$$\min_{y - x^{(0)} \in \text{span}\{p^{(1)},\ p^{(2)},\ \cdots,\ p^{(j)}\}} f(y) \tag{6.27}$$

κι ακόμη ισχύει ότι

$$\text{span}\{p^{(1)},\ p^{(2)},\ \cdots,\ p^{(j)}\} = \text{span}\{r^{(0)},\ r^{(1)},\ \cdots,\ r^{(j-1)}\} = \text{span}\{r^{(0)},\ Ar^{(0)},\ \cdots,\ A^{j-1}r^{(0)}\}. \tag{6.28}$$

β) Οι χώροι των δεξιών μελών στις σχέσεις (6.26) και (6.28) είναι γνωστοί και ως υπόχωροι Krylov του πίνακα A.

Λήμμα 6.3 *Εστω ότι* $0 < \alpha < \beta$. *Τότε το* min-max *πρόβλημα*

$$\min_{p_m \in \mathcal{P}_m,\ p_m(0)=1} \left(\max_{\alpha \le z \le \beta} |p_m(z)| \right), \tag{6.29}$$

όπου $\mathcal{P}_m$ *το σύνολο των πολυωνύμων βαθμού μικρότερου ή ίσου του* m *με πραγματικούς συντελεστές, λύνεται μοναδικά από το πολυώνυμο*

$$\widetilde{p}_m(z)\ =\ \frac{T_m\left(\frac{\beta+\alpha-2z}{\beta-\alpha}\right)}{T_m\left(\frac{\beta+\alpha}{\beta-\alpha}\right)} \tag{6.30}$$

για το οποίο

$$\max_{\alpha \le z \le \beta} |\widetilde{p}_m(z)|\ =\ \frac{1}{T_m\left(\frac{\beta+\alpha}{\beta-\alpha}\right)}. \tag{6.31}$$

Απόδειξη: Η απόδειξη δεν είναι παρά αυτή του Θεωρήματος 4.1 αρκεί να παρατηρήσουμε τα εξής. Καταρχάς οι ανισότητες $0 < \alpha \leq z \leq \beta$ δίνουν ισοδύναμα $1-\beta \leq 1-z \leq 1-\alpha < 1$. Οι τελευταίες δεν είναι παρά οι αντίστοιχες των $-1 < \alpha \leq z \leq \beta < 1$ του Θεωρήματος 4.1. (Σημείωση: Το γεγονός ότι στην παρούσα περίπτωση δεν παρουσιάζεται το κάτω φράγμα -1 στο $1-\beta$ δεν επηρεάζει την κατάσταση. Στο Θεώρημα 4.1 το φράγμα -1 είχε προέρθει από κάποια προηγούμενη απαίτησή μας για τον επαναληπτικό πίνακα T, που παρουσιαζόταν εκεί, να είναι Ερμιτιανός με $\rho(T) < 1$.) Επομένως αν στα αποτελέσματα των αντίστοιχων εκφράσεων της διατύπωσης του Θεωρήματος 4.1 αντικαταστήσουμε τις σταθερές και μεταβλητή $\alpha,\ z$ και β με $1-\beta,\ 1-z$ και $1-\alpha$, αντίστοιχα, προκύπτουν τα συμπεράσματα του παρόντος λήμματος. □

Με μέθοδο αντίστοιχη αυτής του Θεωρήματος 6.3, για τη μέθοδο της απότομης καθόδου, και έχοντας υπόψη τα μέχρι στιγμής συμπεράσματα για τη μέθοδο των συζυγών κλίσεων είναι δυνατόν να διατυπωθεί και να αποδειχτεί με τη βοήθεια των δύο προηγούμενων λημμάτων ένα ανάλογο θεώρημα που δίνεται στη συνέχεια.

Θεώρημα 6.11 *Εστω $x^{(k)},\ k \geq 0$, η ακολουθία που παράγεται από τον αλγόριθμο της μεθόδου συζυγών κλίσεων για οποιοδήποτε $x^{(0)} \in I\!R^n$ και έστω x η λύση του συστήματος $Ax = b$ όπου ο $A \in I\!R^{n,n}$ είναι συμμετρικός και θετικά ορισμένος. Εστω $\kappa = \kappa_2(A) = \frac{\lambda_{\max}}{\lambda_{\min}}$, όπου $\lambda_{\max}$ και $\lambda_{\min}$ η μέγιστη και η ελάχιστη ιδιοτιμή του A, αντίστοιχα. Αν $e^{(k)} = x^{(k)} - x$ είναι το διάνυσμα-σφάλμα στην k επανάληψη τότε*

$$||e^{(k)}||_{A^{\frac{1}{2}}} \ \leq \ 2\left[\left(\frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}\right)^k + \left(\frac{\sqrt{\kappa}+1}{\sqrt{\kappa}-1}\right)^k\right]^{-1} ||e^{(0)}||_{A^{\frac{1}{2}}},\ k = 1(1)n-1. \tag{6.32}$$

Απόδειξη: Καταρχάς παρατηρούμε ότι για $y \in I\!R^n$ είναι $(A(y-x), y-x) = 2f(y) + (Ax, x)$. Επομένως το πρόβλημα ελαχιστοποίησης $\min_{y \in S} f(y)$, με $S \subset I\!R^n$, είναι ισοδύναμο με το πρόβλημα ελαχιστοποίησης $\min_{y \in S}(A(y-x), y-x)$, οπότε, σύμφωνα με τη Σημείωση αμέσως μετά το Λήμμα 6.2 και συγκεκριμένα τη σχέση (6.27), θα έχουμε

$$\begin{aligned}(Ae^{(k)}, e^{(k)}) \ &= \ (A(x^{(k)}-x), x^{(k)}-x) \\ = \ \min_{y-x^{(0)} \in \mathrm{span}\{p^{(1)},\ \cdots,\ p^{(k)}\}}(A(y-x), y-x) \ &= \ \min_{z \in \mathrm{span}\{p^{(1)},\ \cdots,\ p^{(k)}\}}(A(z+e^{(0)}), z+e^{(0)}).\end{aligned} \tag{6.33}$$

Από την (6.28) έχουμε ότι

$$z \in \mathrm{span}\{p^{(1)},\ \cdots,\ p^{(k)}\} \iff z \in \mathrm{span}\{r^{(0)},\ Ar^{(0)},\ \cdots,\ A^{k-1}r^{(0)}\} \iff z = p_{k-1}(A)r^{(0)}, \tag{6.34}$$

για κάποιο πολυώνυμο $p_{k-1} \in \mathcal{P}_{k-1}$, όπου $\mathcal{P}_{k-1}$ το σύνολο των πολυωνύμων με πραγματικούς συντελεστές βαθμού το πολύ $k-1$. Επειδή $r^{(0)} = -Ae^{(0)}$, η (6.33) λόγω της (6.34) γίνεται

$$(Ae^{(k)}, e^{(k)}) \ = \ \min_{p_{k-1} \in \mathcal{P}_{k-1}}(A(I - Ap_{k-1}(A))e^{(0)}, (I - Ap_{k-1}(A))e^{(0)}). \tag{6.35}$$

(Σημείωση: Η παραπάνω σχέση ισχύει και για $k = 0$ αν οριστεί ότι $\mathcal{P}_{-1} = \{0\}$.) Χρησιμοποιώντας τη φασματική ανάλυση του A προκύπτει με πράξεις ανάλογες αυτών του Θεωρήματος (6.3) ότι

$$(Ae^{(k)}, e^{(k)}) \leq \min_{p_{k-1}\in\mathcal{P}_{k-1}} \left((\max_{\lambda\in[\lambda_{\min},\lambda_{\max}]} |1 - \lambda p_{k-1}(\lambda)|)^2 \right) (Ae^{(0)}, e^{(0)}), \tag{6.36}$$

η οποία είναι ισοδύναμη με την

$$||e^{(k)}||_{A^{\frac{1}{2}}} \leq \min_{p_k\in\mathcal{P}_k,\ p_k(0)=1} \left(\max_{\lambda\in[\lambda_{\min},\lambda_{\max}]} |p_k(\lambda)| \right) ||e^{(0)}||_{A^{\frac{1}{2}}}. \tag{6.37}$$

Χρησιμοποιώντας το Λήμμα 6.3 για την εύρεση του συντελεστή του δεύτερου μέλους στην (6.37) προκύπτει αμέσως ότι

$$||e^{(k)}||_{A^{\frac{1}{2}}} \leq \frac{1}{T_k\left(\frac{\lambda_{\max}+\lambda_{\min}}{\lambda_{\max}-\lambda_{\min}}\right)} ||e^{(0)}||_{A^{\frac{1}{2}}} \tag{6.38}$$

ή ισοδύναμα

$$||e^{(k)}||_{A^{\frac{1}{2}}} \leq \frac{1}{T_k\left(\frac{\kappa+1}{\kappa-1}\right)} ||e^{(0)}||_{A^{\frac{1}{2}}} \tag{6.39}$$

και τέλος, επειδή $\frac{\kappa+1}{\kappa-1} > 1$, με βάση την ιδιότητα (γ) των πολυωνύμων Chebyshev καταλήγουμε στην (6.32). □

Σημειώσεις: α) Είναι δυνατόν να αποδειχτεί ότι για σταθερό k ο συντελεστής του $||e^{(0)}||_{A^{\frac{1}{2}}}$ είναι αύξουσα συνάρτηση του δείκτη κατάστασης κ. β) Είναι επίσης δυνατόν να αποδειχτεί με επαγωγή, αν γίνει η σύγκριση των εκφράσεων (6.8) και (6.32), ότι ο συντελεστής στο δεύτερο μέλος της δεύτερης είναι μικρότερος από τον αντίστοιχο συντελεστή της πρώτης (εκτός από την περίπτωση $\kappa = 1$ ή/και $k = 1$, οπότε έχουμε ισότητα), πράγμα που ήταν άλλωστε αναμενόμενο. γ) Στην πράξη και για πολλές κατηγορίες πινάκων, η μέθοδος των συζυγών κλίσεων, και ιδιαίτερα παραλλαγές αυτής, αποδείχνονται ταχύτατες ως προς τη σύγκλιση συγκρινόμενες με άλλες επαναληπτικές μεθόδους. Στην τελευταία παρατήρηση οφείλεται και το γεγονός ότι η μέθοδος των συζυγών κλίσεων είναι ιδιαίτερα δημοφιλής.

Θεωρώντας τις τρεις μεθόδους, απότομης καθόδου, συζυγών διευθύνσεων και συζυγών κλίσεων, είναι φανερό ότι παρά τα τυχόν υπάρχοντα μειονεκτήματα η τελευταία υπερτερεί σαφώς των δύο πρώτων. Καταρχάς οι δυο τελευταίες υπερτερούν της μεθόδου της απότομης καθόδου, αφού ολοκληρώνονται σε n βήματα, ενώ η πρώτη συγκλίνει στη λύση μόνο ασυμπτωτικά. Μεταξύ των μεθόδων συζυγών διευθύνσεων και συζυγών κλίσεων η δεύτερη, όπως ήδη αναφέρθηκε, είναι η βέλτιστη μέθοδος συζυγών διευθύνσεων.

6.6 Προρρυθμισμένη Μέθοδος Συζυγών Κλίσεων

Ενα πιθανό μειονέκτημα της μεθόδου συζυγών κλίσεων είναι ότι για μεγάλο n, πολλές φορές στην πράξη, φαίνεται να συγκλίνει αργά. Για τη βελτίωση της ταχύτητας σύγκλισής της χρησιμοποιείται

η μέθοδος της προρρύθμισης. Η ιδέα της προρρύθμισης είναι η ακόλουθη. Επειδή η ταχύτητα σύγκλισης είναι αύξουσα συνάρτηση του δείκτη κατάστασης $\kappa_2(A)$, προσπαθούμε να τροποποιήσουμε το αρχικό σύστημα και αντ' αυτού να επιλύσουμε ένα ισοδύναμό του με αρκετά μικρότερο δείκτη κατάστασης. Για το σκοπό αυτό χρησιμοποιείται ένας πίνακας $M \in I\!R^{n,n}$, συμμετρικός και θετικά ορισμένος, ως προρρυθμιστής, τ.ω. αφενός μεν να είναι οικονομικά αντιστρέψιμος σε σχέση με τον A αφετέρου δε να ισχύει $\frac{\lambda_{\max}(M^{-1}A)}{\lambda_{\min}(M^{-1}A)} << \kappa_2(A)$, όπου $\lambda_{\max}(\cdot)$ και $\lambda_{\min}(\cdot)$ συμβολίζουν, αντίστοιχα, τη μεγαλύτερη και τη μικρότερη ιδιοτιμή του $M^{-1}A$, που μπορεί να δειχτεί ότι είναι θετικές. Ετσι το σύστημα που θεωρούμε για επίλυση, αντί του αρχικού, είναι το

$$M^{-1}Ax \;=\; M^{-1}b. \tag{6.40}$$

Ο συντελεστής πίνακας του συστήματος (6.40) όμως δεν είναι συμμετρικός και άρα δεν είναι δυνατόν να εφαρμοστεί η μέθοδος των συζυγών κλίσεων για τη λύση του. Για τη μετατροπή του συστήματος (6.40) σε ένα άλλο ισοδύναμο με πίνακα συντελεστών αγνώστων συμμετρικό και θετικά ορισμένο με τις ίδιες ακραίες ιδιοτιμές με αυτές του $M^{-1}A$ εργαζόμαστε ως εξής. Αφού ο M είναι πραγματικός, συμμετρικός και θετικά ορισμένος θα υπάρχει (μοναδικός) πίνακας πραγματικός συμμετρικός και θετικά ορισμένος του οποίου το τετράγωνο θα είναι ο M. Εστω C ο πίνακας αυτός για τον οποίο ισχύει $C^2 = M$. Χρησιμοποιώντας τον C, η (6.40) μπορεί να γραφτεί ισοδύναμα ως

$$C^{-1}AC^{-1}Cx \;= C^{-1}b$$

ή θεωρώντας τις αντικαταστάσεις

$$\widetilde{A} = C^{-1}AC^{-1},\; \widetilde{x} = Cx,\; \widetilde{b} = C^{-1}b \tag{6.41}$$

γράφεται ισοδύναμα και ως

$$\widetilde{A}\widetilde{x} \;= \widetilde{b}. \tag{6.42}$$

Είναι προφανές ότι το νέο σύστημα (6.42) έχει πίνακα συντελεστών αγνώστων $\widetilde{A}$ που είναι πραγματικός, συμμετρικός και θετικά ορισμένος, όπως είναι εύκολο να ελεγχτεί. Ακόμη, επειδή ισχύει ότι $\sigma(\widetilde{A}) \equiv \sigma(C^{-1}AC^{-1}) = \sigma(C^{-2}A) = \sigma(M^{-1}A)$, θα ισχύει ότι $\kappa_2(\widetilde{A}) = \frac{\lambda_{\max}(M^{-1}A)}{\lambda_{\min}(M^{-1}A)}$ και άρα $\kappa_2(\widetilde{A}) << \kappa_2(A)$.

Ισως από μια πρώτη ματιά η προρρύθμιση, με τον τρόπο που εισάγει τον πίνακα C, να θεωρείται ότι δεν είναι δυνατόν να συνεισφέρει ουσιαστικά στο πρόβλημα που θέσαμε αφού για την εύρεση του C απαιτούνται όλες οι ιδιοτιμές και όλα τα ιδιοδιανύσματα του πίνακα αυτού ή, με βάση το Θεώρημα 3.13, του M. Οπως θα δούμε όμως στη συνέχεια ο πίνακας C εμφανίζεται στον τελικό αλγόριθμο ως $C^2 = M$ και άρα δε χρειάζεται να βρεθεί αναλυτικά.

Αν ο αλγόριθμος της μεθόδου συζυγών κλίσεων εφαρμοστεί για τη λύση του συστήματος (6.42) θα είναι ο παρακάτω:

$\widetilde{x}^{(0)} = 0$
$\widetilde{r}^{(0)} = \widetilde{b}$
$\widetilde{p}^{(1)} = \widetilde{r}^{(0)}$
$\widetilde{\alpha}_1 = \frac{(\widetilde{r}^{(0)}, \widetilde{r}^{(0)})}{(\widetilde{A}\widetilde{p}^{(1)}, \widetilde{p}^{(1)})}$
$\widetilde{x}^{(1)} = \widetilde{x}^{(0)} + \widetilde{\alpha}_1 \widetilde{p}^{(1)}$
$\widetilde{r}^{(1)} = \widetilde{b} - \widetilde{A}\widetilde{x}^{(1)} \ \ (= \widetilde{r}^{(0)} - \widetilde{\alpha}_1 \widetilde{A}\widetilde{p}^{(1)})$
$k = 1$
Εφόσον $\|\widetilde{r}^{(k)}\| > \epsilon$ και $k < n$
$k = k + 1$
$\widetilde{\beta}_k = \frac{(\widetilde{r}^{(k-1)}, \widetilde{r}^{(k-1)})}{(\widetilde{r}^{(k-2)}, \widetilde{r}^{(k-2)})}$
$\widetilde{p}^{(k)} = \widetilde{r}^{(k-1)} + \widetilde{\beta}_k \widetilde{p}^{(k-1)}$
$\widetilde{\alpha}_k = \frac{(\widetilde{r}^{(k-1)}, \widetilde{r}^{(k-1)})}{(\widetilde{A}\widetilde{p}^{(k)}, \widetilde{p}^{(k)})}$
$\widetilde{x}^{(k)} = \widetilde{x}^{(k-1)} + \widetilde{\alpha}_k \widetilde{p}^{(k)}$
$\widetilde{r}^{(k)} = \widetilde{b} - \widetilde{A}\widetilde{x}^{(k)} \ \ (= \widetilde{r}^{(k-1)} - \widetilde{\alpha}_k \widetilde{A}\widetilde{p}^{(k)})$
Τέλος 'Εφόσον'
Αποτέλεσμα: $\widetilde{x} = \widetilde{x}^{(k)}$ η προσέγγιση της λύσης.

Λόγω των αντικαταστάσεων (6.41) είναι λογικό να θέσουμε καταρχάς $\widetilde{x}^{(k)} = Cx^{(k)}$ και ακόμη $\widetilde{r}^{(k)} = \widetilde{b} - \widetilde{A}\widetilde{x}^{(k)} = C^{-1}b - C^{-1}AC^{-1}Cx^{(k)} = C^{-1}(b - Ax^{(k)}) = C^{-1}r^{(k)}$. Αν δε επιπλέον επιθυμούμε παραπέρα απλοποίηση του αλγόριθμου και ιδιαίτερα αν θέλουμε να έχουμε τη σχέση $\widetilde{x}^{(k)} = \widetilde{x}^{(k-1)} + \widetilde{\alpha}_k \widetilde{p}^{(k)}$ σε πλήρη αντιστοιχία με τη $x^{(k)} = x^{(k-1)} + \alpha_k p^{(k)}$ του κανονικού αλγόριθμου της μεθόδου των συζυγών κλίσεων, τότε θα πρέπει να θέσουμε $\widetilde{p}^{(k)} = Cp^{(k)}$. Είναι εύκολο να διαπιστώσει κανείς τότε ότι οι (6.41) μαζί με τις αντικαταστάσεις, που μόλις θεωρήθηκαν, μπορούν να μετασχηματίσουν τον παρόντα αλγόριθμο σε έναν ισοδύναμο όπου, όμως, μόνο οι γνωστές, ως προς το συμβολισμό, ποσότητες από τον κλασικό αλγόριθμο θα εμφανίζονται. Τέλος, με μερικούς απλούς μετασχηματισμούς, όπου παραλείπουμε το σύμβολο "˜" από τα $\widetilde{\alpha}_k$ και $\widetilde{\beta}_{k+1}$ και κυρίως αντικαθιστούμε το εμφανιζόμενο C^2 με το ίσο του M και το $M^{-1}r^{(k)} = z^{(k)}$, με το $z^{(k)}$ να βρίσκεται από τη λύση του συστήματος $Mz^{(k)} = r^{(k)}$, παίρνουμε τον αλγόριθμο της προρρυθμισμένης μεθόδου των συζυγών κλίσεων στην τελική του μορφή. Συγκεκριμένα:

Αλγόριθμος Προρρυθμισμένης Μεθόδου Συζυγών Κλίσεων:

Δεδομένα: $A \in I\!R^{n,n},\ A^T = A,\ A$ θετικά ορισμένος, $b \in I\!R^n$, $M \in I\!R^{n,n},\ M^T = M,\ M$ θετικά ορισμένος με $\frac{\lambda_{\max}(M^{-1}A)}{\lambda_{\min}(M^{-1}A)} << \kappa_2(A),\ \epsilon \in I\!R^+$ το επιθυμητό σφάλμα.
$x^{(0)} = 0$
$r^{(0)} = b$
$Mz^{(0)} = r^{(0)}$
$p^{(1)} = z^{(0)}$
$\alpha_1 = \frac{(z^{(0)}, r^{(0)})}{(Ap^{(1)}, p^{(1)})}$

$x^{(1)} = x^{(0)} + \alpha_1 p^{(1)}$
$r^{(1)} = b - Ax^{(1)} \;\; (= r^{(0)} - \alpha_1 Ap^{(1)})$
$k = 1$
Εφόσον $\|r^{(k)}\| > \epsilon$ και $k < n$
$k = k + 1$
$Mz^{(k-1)} = r^{(k-1)}$
$\beta_k = \frac{(z^{(k-1)}, r^{(k-1)})}{(z^{(k-2)}, r^{(k-2)})}$
$p^{(k)} = z^{(k-1)} + \beta_k p^{(k-1)}$
$\alpha_k = \frac{(z^{(k-1)}, r^{(k-1)})}{(Ap^{(k)}, p^{(k)})}$
$x^{(k)} = x^{(k-1)} + \alpha_k p^{(k)}$
$r^{(k)} = b - Ax^{(k)} \;\; (= r^{(k-1)} - \alpha_k Ap^{(k)})$
Τέλος 'Εφόσον'
Αποτέλεσμα: $x = x^{(k)}$ η προσέγγιση της λύσης.

Η επιλογή κατάλληλου προρρυθμιστή πίνακα M έχει τις ίδιες δυσκολίες που έχει κανείς στην επιλογή προρρυθμιστή στις γενικές επαναληπτικές μεθόδους. Λόγω της φύσης του A (πραγματικός, συμμετρικός και θετικά ορισμένος) υπάρχουν κάποιοι προρρυθμιστές που δεν είναι άλλοι παρά αυτοί που χρησιμοποιούνται στις κλασικές επαναληπτικές μεθόδους με την προϋπόθεση ότι είναι (πραγματικοί) συμμετρικοί και θετικά ορισμένοι πίνακες. Ετσι μπορεί κανείς να επιλέξει

$$M \;=\; \mathrm{diag}(a_{11},\; a_{22},\; \cdots,\; a_{nn}),$$

οπότε προκύπτει ως μέθοδος η γνωστή και ως Jacobi-Συζυγών Κλίσεων (Jacobi-CG), ή

$$M \;=\; \mathrm{diag}(A_{11},\; A_{22},\; \cdots,\; A_{pp}), \quad A_{ii} \in I\!R^{n_i,n_i},\; \sum_{i=1}^{p} n_i = n,$$

με αντίστοιχη μέθοδο τη γνωστή και ως Block Jacobi-Συζυγών Κλίσεων (Block Jacobi-CG), ή, τέλος,

$$M \;=\; \frac{1}{\omega(2-\omega)}(D - \omega L)D^{-1}(D - \omega L^T),$$

με αντίστοιχη μέθοδο τη γνωστή και ως (Block) SSOR-Συζυγών Κλίσεων ((Block) SSOR-CG). Σημειώνεται ότι η τελευταία μέθοδος χρησιμοποιείται συνήθως με $\omega = 1$.

ΑΣΚΗΣΕΙΣ

1.: Να θεωρηθεί το 2×2 γραμμικό σύστημα $Ax = b$, όπου A =diag$(1, \lambda)$, $\lambda > 0$, $b = 0$ και $x^{(0)} \in I\!R^2 \backslash \{0\}$ οποιοδήποτε.

α) Αν $x^{(k)} = [x_1^{(k)}\ x_2^{(k)}]^T$, να δειχτεί ότι η μέθοδος της Απότομης Καθόδου δίνει:

$$x^{(k+1)} = \frac{x_1^{(k)} x_2^{(k)}(\lambda-1)}{(x_1^{(k)})^2 + \lambda^3 (x_2^{(k)})^2} [\lambda^2 x_2^{(k)}\ - x_1^{(k)}]^T.$$

β) Να δειχτεί ότι αν $x^{(k)} = c[\lambda\ \pm 1]^T$, $c \in I\!R \setminus \{0\}$, τότε η σχέση

$$||e^{(k+1)}||_{A^{\frac{1}{2}}} \leq \left(\frac{\kappa-1}{\kappa+1}\right) ||e^{(k)}||_{A^{\frac{1}{2}}},$$

όπου $e^{(k)}$ το διάνυσμα-σφάλμα στην k επανάληψη και κ ο δείκτης κατάστασης, που αντιστοιχεί στη φασματική norm του πίνακα A, ισχύει ως ισότητα.

2.: α) Αν είναι $A \in \mathbb{C}^{n,n}$ Ερμιτιανός και θετικά ορισμένος πίνακας να βρεθούν οι σταθερές σύγκρισης των διανυσματικών norms $||x||_2$ και $||x||_{A^{\frac{1}{2}}}$ $\forall\ x \in \mathbb{C}^n$.
β) Να βρεθούν δυο διανύσματα $x \in \mathbb{C}^n \setminus \{0\}$, που να ικανοποιούν την κάθε μία από τις δυο ισότητες που εμφανίζονται στην παραπάνω σύγκριση.

3.: Δίνονται τα διανύσματα $x^{(1)} = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$, $x^{(2)} = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$ και $x^{(3)} = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$. Να βρεθεί μια βάση ορθογώνιων διανυσμάτων $q^{(1)}$, $q^{(2)}$, $q^{(3)}$ και να δοθούν τα $x^{(1)}$, $x^{(2)}$, $x^{(3)}$ ως γραμμικοί συνδυασμοί των $q^{(1)}$, $q^{(2)}$, $q^{(3)}$ χρησιμοποιώντας τη μέθοδο της ορθογωνιοποίησης των Gram-Schmidt. (Περιορισμός: Να γίνουν ακριβείς πράξεις χρησιμοποιώντας (ριζικά και) κλάσματα στους υπολογισμούς.)

4.: Να λυθεί το γραμμικό σύστημα $Ax = b$, με $A = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix}$ και $b = \begin{bmatrix} 2 \\ 0 \\ 2 \end{bmatrix}$, με τη μέθοδο Συζυγών Κλίσεων και αρχικό διάνυσμα $x^{(0)} = 0$.

5.: Υποτίθεται ότι στον αρχικό αλγόριθμο των Gram-Schmidt παίρνεται $U = I$.
α) Να βρεθούν αναλυτικά οι τρεις πρώτες στήλες $p^{(1)}, p^{(2)}, p^{(3)}$ του πίνακα P στην παραγοντοποίηση $U = PR$, όπου $P \in I\!R^{n,n}$ με στήλες ανά δύο $A-$ορθογώνιες ($A \in I\!R^{n,n}$, συμμετρικός και θετικά ορισμένος) και $R \in I\!R^{n,n}$ άνω τριγωνικός με διαγώνια στοιχεία μονάδες. και
β) Με βάση το παραπάνω συμπέρασμα να δοθούν, χωρίς απόδειξη, οι αναλυτικές εκφράσεις των στοιχείων της k στήλης του πίνακα P καθώς και της k γραμμής του πίνακα R.

6.: Δίνεται ο πίνακας

$$U = \begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & 1 \\ 1 & 1 & -1 \\ 1 & -1 & -1 \end{bmatrix}.$$

α) Να δειχτεί ότι τα διανύσματα-στήλες του πίνακα U είναι γραμμικά ανεξάρτητα. και
β) Βασιζόμενοι στο γεγονός ότι ο πίνακας I είναι πραγματικός, συμμετρικός και θετικά ορισμένος να αναλυθεί κατά Gram-Schmidt ο παραπάνω πίνακας U σε γινόμενο παραγόντων PR, έτσι ώστε τα διανύσματα-στήλες του πίνακα P να είναι ορθογώνια μεταξύ τους και ο πίνακας R να είναι άνω τριγωνικός με διαγώνια στοιχεία μονάδες.

7.: Δίνονται πίνακας $A \in I\!R^{n,n}$, με $A^T = A$, και διανύσματα $x_i \in I\!R^n,\ i = 1(1)n$, που είναι $A-$ορθογώνια ανά δύο και τέτοια ώστε $(Ax_i, x_i) > 0,\ i = 1(1)n$.
α) Να δειχτεί ότι τα $x_i \in I\!R^n,\ i = 1(1)n$, είναι γραμμικά ανεξάρτητα. και
β) Να δειχτεί ότι ο A είναι θετικά ορισμένος.

8.: Είναι γνωστό απο τη Γραμμική Αλγεβρα ότι αν $r \in I\!R^n$ και M υπόχωρος του $I\!R^n$, τότε $p \in M$ είναι η *ορθή προβολή* του r στον M ανν ισχύει ότι

$$(r, x) = (p, x) \ \ \forall\ x \in M$$

α) Με βάση τον παραπάνω ορισμό να αποδειχτεί ότι η ορθή προβολή p του r στον M υπάρχει, είναι μοναδική και ικανοποιεί τη σχέση

$$||p||^2 \ = \ ||r||^2 - ||r - p||^2.$$

(Υπόδειξη: Να θεωρηθούν n ορθοκανονικά διανύσματα $x^{(i)} \in I\!R^n,\ i = 1(1)n$, τα $m\ (< n)$ πρώτα από τα οποία ανήκουν στον υπόχωρο M.)
β) Να αποδειχτεί ότι το διάνυσμα $r - p$ είναι η ορθή προβολή του r στον υπόχωρο $M^\perp$, όπου $M^\perp$ συμβολίζει τον υπόχωρο που είναι κάθετος στον υπόχωρο M. και
γ) Να αποδειχτεί ότι το διάνυσμα $p^{(k)}$ της μεθόδου των Συζυγών Κλίσεων δεν είναι παρά η ορθή προβολή του διανύσματος $r^{(k-1)}$ στον υπόχωρο που είναι κάθετος στον $\text{span}\{Ap^{(1)}, Ap^{(2)}, \cdots, Ap^{(k-1)}\},\ k = 1(1)n,\ r^{(j)} \neq 0,\ j = 0(1)k-1$.

9.: Για την επίλυση του γραμμικού συστήματος (5.21) προτιθέμεθα να εφαρμόσουμε τη Μέθοδο Συζυγών Κλίσεων καθώς και τις δυο Προρρυθμισμένες Μεθόδους Συζυγών Κλίσεων με προρρυθμιστές πίνακες α) $4\,I \otimes I$ και β) $I \otimes (2I + T)$. Να βρεθούν αναλυτικά οι δείκτες κατάστασης των αντίστοιχων συντελεστών πινάκων των συστημάτων που επιλύονται και να σχολιαστεί (συγκριτικά) η σύγκλιση των τριών μεθόδων.

7 Γραμμική Μέθοδος Ελάχιστων Τετραγώνων

7.1 Εισαγωγή

Πολλά προβλήματα της Επιστήμης και της Τεχνολογίας απαιτούν τακτικές προσομοίωσης για την αναπαραγωγή φυσικών φαινομένων (καταστάσεων) σε συνθήκες Εργαστηρίου. Στα περισσότερα από αυτά προτείνεται ένα μαθηματικό πρότυπο (μοντέλο) που συνήθως είναι γραμμικό είτε για την περιγραφή του φαινομένου είτε για μια πρώτη προσέγγιση στην κατανόησή του. Για την καλύτερη δυνατή κατανόηση του φαινομένου που μελετιέται το πλήθος των διεξαγόμενων πειραμάτων και αντίστοιχων μετρήσεων (εξισώσεων) είναι συνήθως κατά πολύ μεγαλύτερο από τον αριθμό των ζητούμενων παραμέτρων (αγνώστων). Σε τέτοιες περιπτώσεις, και όχι μόνον, η μαθηματική διατύπωση του προβλήματος είναι η ακόλουθη:
Να βρεθεί διάνυσμα $x \in I\!R^m$ που ικανοποιεί κατά τον καλύτερο δυνατό τρόπο το γραμμικό σύστημα

$$Ax = b,\ \mu\epsilon\ A \in I\!R^{n,m},\ b \in I\!R^n\ \kappa\alpha\iota\ m \leq n. \tag{7.1}$$

Αν υποτεθεί ότι οι πρώτες m εξισώσεις του (7.1) έχουν μονοσήμαντα ορισμένη λύση τότε είναι μάλλον απίθανο στη γενική περίπτωση ($m < n$) η λύση αυτή να επαληθεύει και τις υπόλοιπες $n - m$ εξισώσεις του (7.1). Ετσι οδηγούμαστε στο να αναζητούμε εκείνο το x για το οποίο το διάνυσμα-σφάλμα $r = b - Ax$ είναι όσο το δυνατόν "μικρότερο". Με άλλα λόγια επιζητιέται η ελαχιστοποίηση μιας norm του r. Αν η προϋπάρχουσα εμπειρία υποδείχνει ότι η "καλύτερη" norm είναι η Ευκλείδεια, τότε το αντίστοιχο πρόβλημα της ελαχιστοποίησης της $||r||_2$, ή ισοδύναμα της $||r||_2^2$, διατυπώνεται μαθηματικά ως

$$\min_{x \in I\!R^m} ||r||_2^2 \equiv \min_{x \in I\!R^m} ||b - Ax||_2^2 = \min_{x \in I\!R^m} \sum_{i=1}^{n} \left(b_i - \sum_{j=1}^{m} a_{ij} x_j \right)^2. \tag{7.2}$$

Το πρόβλημα (7.2), για ευνόητους λόγους, είναι γνωστό ως Πρόβλημα των Ελάχιστων Τετραγώνων.

7.2 Λύση των Κανονικών Εξισώσεων

Για την ελαχιστοποίηση της έκφρασης

$$f(x) := ||r||_2^2 = \sum_{i=1}^{n} \left(b_i - \sum_{j=1}^{m} a_{ij} x_j \right)^2$$

θα πρέπει πρώτα να βρεθούν τα κρίσιμα σημεία της $f(x)$ μηδενίζοντας την αντίστοιχη κλίση. Δηλαδή

$$\nabla f(x) \;=\; \left[\frac{\partial f(x)}{\partial x_1}\;\frac{\partial f(x)}{\partial x_2}\;\cdots\;\frac{\partial f(x)}{\partial x_m}\right]^T \;=\; 0. \tag{7.3}$$

Η αναγκαία συνθήκη (7.3) είναι ισοδύναμη με το σύνολο των εξισώσεων

$$\frac{\partial f(x)}{\partial x_k} \;=\; \sum_{i=1}^{n} 2\left(b_i - \sum_{j=1}^{m} a_{ij}x_j\right)(-a_{ik}) \;=\; 0,\; k = 1(1)m.$$

Οι παραπάνω εξισώσεις γράφονται ισοδύναμα ως

$$\sum_{i=1}^{n}\left(a_{ik}\left(\sum_{j=1}^{m} a_{ij}x_j\right)\right) \;=\; \sum_{i=1}^{n} a_{ik}b_i,\; k = 1(1)m,$$

ή ως

$$\sum_{i=1}^{n}\left(a_{ik}(Ax)_i\right) \;=\; \sum_{i=1}^{n} a_{ik}b_i,\; k = 1(1)m. \tag{7.4}$$

Αν θεωρηθεί ότι $A = [a_1\; a_2\; \cdots\; a_m]$, όπου $a_k = [a_{1k}\; a_{2k}\; \cdots\; a_{nk}]^T$, $k = 1(1)m$, τότε οι εξισώσεις (7.4) γράφονται ως

$$a_k^T(Ax) \;=\; a_k^T b,\; k = 1(1)m, \tag{7.5}$$

και το σύνολο των (7.5) γράφεται υπό μορφήν πινάκων ως

$$\begin{bmatrix} a_1^T(Ax) \\ a_2^T(Ax) \\ \vdots \\ a_m^T(Ax) \end{bmatrix} \;=\; \begin{bmatrix} a_1^T b \\ a_2^T b \\ \vdots \\ a_m^T b \end{bmatrix}$$

ή ισοδύναμα ως

$$A^T Ax \;=\; A^T b. \tag{7.6}$$

Το γραμμικό σύστημα (7.6) των m εξισώσεων με τους m αγνώστους είναι γνωστό και ως σύστημα των "Κανονικών Εξισώσεων".

Στη συνέχεια θα αποδείξουμε ότι το σύστημα (7.6), που έχει πάντοτε μία ή άπειρες λύσεις, δεν αποτελεί μόνο αναγκαία αλλά και ικανή συνθήκη για να έχει το πρόβλημα (7.2) λύση.

Αποδείχνουμε πρώτα το δεύτερο ισχυρισμό.

Θεώρημα 7.1 *Αν $x \in \mathbb{R}^m$ είναι λύση του συστήματος (7.6), τότε $\forall\; y \in \mathbb{R}^m$ ισχύει*

$$||r_y||_2^2 \;\geq\; ||r_x||_2^2, \quad \textit{όπου } r_y = b - Ay \;\textit{και}\; r_x = b - Ax. \tag{7.7}$$

Απόδειξη: Από την (7.6) παίρνουμε αμέσως ότι

$$A^T r_x = 0_m \ \kappa\alpha\iota \ r_x^T A = 0_m^T. \tag{7.8}$$

Επομένως

$$\begin{array}{rcl} ||r_y||_2^2 & = & r_y^T r_y \ = \ (b-Ay)^T(b-Ay) \ = \ (r_x + A(x-y))^T(r_x + A(x-y)) \\ & = & (r_x^T + (x-y)^T A^T)(r_x + A(x-y)) \\ & = & ||r_x||_2^2 + r_x^T A(x-y) + (x-y)^T A^T r_x + ||A(x-y)||_2^2 \\ & = & ||r_x||_2^2 + ||A(x-y)||_2^2 \end{array} \tag{7.9}$$

αφού οι δυο μεσαίοι όροι στο προτελευταίο μέλος είναι μηδέν λόγω της (7.8). Επομένως $||r_y||_2^2 \geq ||r_x||_2^2$, πράγμα που αποδείχνει την (7.7). □

Για τον πίνακα των συντελεστών των αγνώστων $A^T A$ στο σύστημα των κανονικών εξισώσεων (7.6) παρατηρούμε ότι είναι πραγματικός με $A^T A \in I\!R^{m,m}$, συμμετρικός, διότι $(A^T A)^T = A^T A$, και μή αρνητικά ορισμένος, διότι $\forall\, y \in I\!R^m \setminus \{0\}$ είναι $y^T A^T A y = (Ay)^T(Ay) = ||Ay||_2^2 \geq 0$. Λόγω της τελευταίας ανισότητας θα διακρίνουμε προφανώς δύο περιπτώσεις. Την $||Ay||_2 > 0,\ \forall\, y \in I\!R^m \setminus \{0\}$ και την $||Ay||_2 = 0$ για ένα τουλάχιστον $y \in I\!R^m \setminus \{0\}$. Στην πρώτη περίπτωση ο πίνακας $A^T A$ θα είναι θετικά ορισμένος, θα έχει ορίζουσα θετική (διάφορη από το μηδέν) και άρα το σύστημα (7.6) θα έχει μία και μόνο λύση και θα λέμε ότι ο πίνακας A είναι πλήρους βαθμού, ενώ στη δεύτερη περίπτωση ο $A^T A$ θα είναι μή αρνητικά ορισμένος και άρα θα πρέπει να εξεταστεί παραπέρα για το αν και πότε το σύστημα (7.6) έχει άπειρες λύσεις ή καμία λύση, και θα λέμε ότι ο πίνακας A είναι ελλιπούς βαθμού.

Για την πρώτη περίπτωση ισχύει η παρακάτω πρόταση:

Θεώρημα 7.2 *Ο πίνακας $A^T A$ του συστήματος (7.6) είναι θετικά ορισμένος ανν τα διανύσματα-στήλες του πίνακα A είναι γραμμικά ανεξάρτητα.*

Απόδειξη: Ο πίνακας $A^T A$ είναι θετικά ορισμένος ανν $\forall\ y \in I\!R^m \setminus \{0\}$ ισχύει $y^T A^T A y > 0$ ή, ισοδύναμα, $||Ay||_2^2 > 0$ ή ακόμη $Ay \neq 0$. Αν $a_i,\ i = 1(1)m$, είναι τα διανύσματα-στήλες του A και $y_i,\ i = 1(1)m$, οι συνιστώσες του y, τότε η τελευταία σχέση είναι ισοδύναμη με την $y_1 a_1 + y_2 a_2 + \cdots + y_m a_m \neq 0,\ \forall\, y \in I\!R^m \setminus \{0\}$. Η σχέση στην οποία καταλήξαμε, όμως, είναι ισοδύναμη με το ότι τα διανύσματα $a_i,\ i = 1(1)m$, είναι γραμμικά ανεξάρτητα, που αποδείχνει την παρούσα πρόταση. □

Πόρισμα 7.1 *Αν τα διανύσματα-στήλες του πίνακα A του συστήματος των κανονικών εξισώσεων (7.6) είναι γραμμικά ανεξάρτητα, τότε το σύστημα (7.6) έχει μία και μόνο λύση.*

Αμεση συνέπεια του θεωρήματος που αποδείχτηκε είναι ότι αν τα διανύσματα-στήλες του πίνακα A είναι γραμμικά εξαρτημένα, τότε ο πίνακας $A^T A$ του συστήματος των κανονικών εξισώσεων (7.6) δεν είναι αντιστρέψιμος και άρα το σύστημα (7.6) έχει άπειρες λύσεις ή καμία λύση. Στην παρακάτω πρόταση θα αποδειχτεί ότι μόνο η πρώτη περίπτωση είναι δυνατή.

Θεώρημα 7.3 *Εστω ότι* r $(0 < r < m)$ *είναι το μέγιστο πλήθος των γραμμικά ανεξάρτητων διανυσμάτων-στηλών του πίνακα* $A \in I\!R^{n,m}$, $m \leq n$, *του συστήματος (7.1). Τότε το σύστημα των κανονικών εξισώσεων (7.6) έχει άπειρες λύσεις. (Σημείωση: Στην παρούσα περίπτωση ο* A *είναι ελλιπούς βαθμού* r $(< m)$, *πράγμα που γράφεται και ως* $\mathrm{rank}(A) = r$.)

Απόδειξη: Σύμφωνα με γνωστή πρόταση της Γραμμικής Αλγεβρας, αν r $(0 < r < m)$ είναι το μέγιστο πλήθος των γραμμικά ανεξάρτητων διανυσμάτων-στηλών του A, τότε το μέγιστο πλήθος των γραμμικά ανεξαρτήτων διανυσμάτων-γραμμών του A θα είναι κι αυτό r. Οπως επίσης γνωρίζουμε από τη Γραμμική Αλγεβρα, είναι δυνατόν, με πολλαπλασιασμό του A από τα αριστερά επί κατάλληλο μεταθετικό πίνακα $P_1 \in I\!R^{n,n}$ και με πολλαπλασιασμό του από τα δεξιά επί κατάλληλο μεταθετικό πίνακα $P_2 \in I\!R^{m,m}$, να έχουμε όχι μόνο τα διανύσματα των r πρώτων γραμμών του P_1AP_2 γραμμικά ανεξάρτητα και τα διανύσματα των r πρώτων στηλών του ομοίως γραμμικά ανεξάρτητα, αλλά συγχρόνως και όλους τους $p \times p$, $p = 1(1)r$, κύριους υποπίνακες της άνω αριστερής γωνίας του αντιστρέψιμους. Εφαρμόζοντας τότε τη μέθοδο απαλοιφής του Gauss στον πίνακα $\widetilde{A} = P_1AP_2$, για την απαλοιφή των στοιχείων του που βρίσκονται κάτω από τα διαγώνια στοιχεία του $\widetilde{a}_{ii}$, $i = 1(1)r$, όπως ακριβώς στο Θεώρημα 2.1, καταλήγουμε σε έναν $n \times m$ "άνω τριγωνικό" πίνακα $\widetilde{U}$ του οποίου όλα τα στοιχεία κάτω από τα r πρώτα διαγώνια έχουν απαλειφτεί. Επιπλέον, ο $(n-r) \times (m-r)$ υποπίνακάς του της κάτω δεξιά γωνίας του είναι ο μηδενικός. Αυτό γιατί αν κάποιο από τα στοιχεία του υποπίνακα αυτού ήταν μή μηδενικό τότε προφανώς εκτός από τα r γραμμικά ανεξάρτητα διανύσματα-στήλες των r πρώτων στηλών του $\widetilde{U}$ μαζί με τη στήλη που θα περιείχε το προαναφερόμενο μή μηδενικό στοιχείο, και άρα τα αντίστοιχα διανύσματα-στήλες των πινάκων $\widetilde{A}$ και A, θα ήταν γραμμικά ανεξάρτητα πλήθους $r+1$, που είναι άτοπο. Για την απαλοιφή που περιγράφτηκε χρησιμοποιήθηκαν προφανώς πολλαπλασιαστές, που αντιστοιχούν στις πρώτες r στήλες του $\widetilde{A}$, και οι οποίοι θα μπορούν να γραφτούν σε έναν $n \times r$ "κάτω τριγωνικό" πίνακα $\widetilde{L}$ με διαγώνια στοιχεία μονάδες. Για να μπορεί να γραφτεί ο $\widetilde{A}$ σα γινόμενο ενός "κάτω τριγωνικού" και ενός "άνω τριγωνικού" πίνακα, όπως στο Θεώρημα 2.1, επεκτείνουμε τον πίνακα των πολλαπλασιαστών έτσι ώστε να καταστεί $n \times n$. Τα στοιχεία των τελευταίων $n-r$ στηλών του νέου "κάτω τριγωνικού" πίνακα L μπορεί να είναι οποιαδήποτε κι αυτό λόγω της παρουσίας των $n-r$ μηδενικών γραμμών του πίνακα $\widetilde{U}$. Για να διατηρείται η κάτω τριγωνική μορφή του νέου πίνακα των πολλαπλασιαστών L επιλέγουμε τον άνω δεξιά $(n-r) \times r$ υποπίνακά του να είναι ο μηδενικός και τον $(n-r) \times (n-r)$ κάτω δεξιά να είναι ο μοναδιαίος. Ετσι μπορούμε να γράφουμε τον πίνακα $\widetilde{A}$ σαν το γινόμενο $\widetilde{A} = L\widetilde{U}$. Επισημαίνεται ότι ο $m \times m$ "άνω τριγωνικός" πίνακας $\widetilde{U}$ έχει στοιχεία $\widetilde{a}^{(k)}_{kj}$, $k = 1(1)r$, $j = k(1)m$, τα οποία έχουν προκύψει κατά τα r βήματα της απαλοιφής, με $\widetilde{a}^{(1)}_{1j} = \widetilde{a}_{1j}$, $j = 1(1)m$. Τέλος, για να δώσουμε στην παραγοντοποίηση του $\widetilde{A}$ μια μορφή ανάλογη αυτής της παραγοντοποίσης Crout γράφουμε τον $\widetilde{U}$ σαν ένα γινόμενο ενός "διαγώνιου" $n \times m$ πίνακα D τα διαγώνια στοιχεία του οποίου είναι τα διαγώνια στοιχεία του $\widetilde{U}$, με όλα τα άλλα στοιχεία του μηδενικά, επί έναν "άνω τριγωνικό" πίνακα U, του οποίου στοιχεία στις r πρώτες γραμμές είναι αυτά του $\widetilde{U}$ διαιρεμένα ανά γραμμή με τα αντίστοιχα οδηγά στοιχεία του.

Ακόμη παρατηρούμε ότι, λόγω της παρουσίας των μηδενικών του D, οι τελευταίες $m-r$ γραμμές του U μπορεί να είναι οποιεσδήποτε. Επιλέγουμε τον κάτω αριστερά $(m-r)\times r$ υποπίνακά του να είναι ο μηδενικός και τον κάτω δεξιά $(m-r)\times(m-r)$ να είναι ο μοναδιαίος. Ετσι τελικά έχουμε ότι

$$LDU \;=\; \widetilde{A} \;= P_1 A P_2,, \tag{7.10}$$

όπου, υπενθυμίζεται ότι, $L \in I\!R^{n,n}$ είναι κάτω τριγωνικός με $l_{ii}=1,\ i=1(1)n$, και $(n-r)\times(n-r)$ υποπίνακα της κάτω δεξιά γωνίας του τον I_{n-r}, $D \in I\!R^{n,m}$, "διαγώνιος" με $d_{ii}\neq 0,\ i=1(1)r$, και $d_{ii}=0,\ i=r+1(1)m$, και $U \in I\!R^{m,m}$ άνω τριγωνικός με διαγώνια στοιχεία $u_{ii}=1,\ i=1(1)m$, και $(m-r)\times(m-r)$ υποπίνακα της κάτω δεξιάς γωνίας του τον I_{m-r}.
Λύνοντας την (7.10) ως προς A και αντικαθιστώντας στην (7.6) έχουμε

$$P_2U^TD^TL^TP_1P_1^TLDUP_2^Tx \;=\; P_2U^TD^TL^TP_1b,$$

οπότε απλοποιώντας παίρνουμε

$$D^TL^TLDUP_2^Tx \;=\; D^TL^TP_1b.$$

Εστω $y=UP_2^Tx \in I\!R^m$ και $c=L^TP_1b \in I\!R^n$. Χρησιμοποιώντας block μορφές για τους πίνακες D και L, όπου οι block υποπίνακές τους, εκτός από τους μοναδιαίους, συμβολίζονται ενδεικτικά με τις διαστάσεις τους, καθώς και τους block διαχωρισμούς των διανυσμάτων $y=[y^{rT}\ y^{m-rT}]^T$, και $c=[c^{rT}\ c^{n-rT}]^T$, όπου $y^r, c^r \in I\!R^r,\ y^{m-r} \in I\!R^{m-r},\ c^{n-r} \in I\!R^{n-r}$, έχουμε

$$\begin{array}{l}\left[\begin{array}{c|c} D_{r,r} & 0_{r,n-r} \\ \hline 0_{m-r,r} & 0_{m-r,n-r}\end{array}\right]\left[\begin{array}{c|c} M_{r,r} & M_{r,n-r} \\ \hline M_{n-r,r} & M_{n-r,n-r}\end{array}\right]\left[\begin{array}{c|c} D_{r,r} & 0_{r,m-r} \\ \hline 0_{n-r,r} & 0_{n-r,m-r}\end{array}\right]\left[\begin{array}{c} y^r \\ \hline y^{m-r}\end{array}\right] \\ \qquad = \left[\begin{array}{c|c} D_{r,r} & 0_{r,n-r} \\ \hline 0_{m-r,r} & 0_{m-r,n-r}\end{array}\right]\left[\begin{array}{c} c^r \\ \hline c^{n-r}\end{array}\right],\end{array} \tag{7.11}$$

όπου

$$\begin{array}{rl} M \;:=\; \left[\begin{array}{c|c} M_{r,r} & M_{r,n-r} \\ \hline M_{n-r,r} & M_{n-r,n-r}\end{array}\right] &= \left[\begin{array}{c|c} L_{r,r}^T & L_{n-r,r}^T \\ \hline 0_{n-r,r} & I_{n-r}\end{array}\right]\left[\begin{array}{c|c} L_{r,r} & 0_{r,n-r} \\ \hline L_{n-r,r} & I_{n-r}\end{array}\right] \\ &= \left[\begin{array}{c|c} L_{r,r}^TL_{r,r}+L_{n-r,r}^TL_{n-r,r} & L_{n-r,r}^T \\ \hline L_{n-r,r} & I_{n-r}\end{array}\right] \;=:\; L^TL.\end{array} \tag{7.12}$$

Η (7.11) είναι ισοδύναμη με

$$\left\{\begin{array}{l} M_{r,r}D_{r,r}y^r \;=\; c^r, \\ 0\ y^{m-r} \;=\; 0_{m-r}.\end{array}\right. \tag{7.13}$$

Η δεύτερη των εξισώσεων (7.13) δίνει αμέσως

$$y^{m-r} \in I\!R^{m-r},\ \mu\epsilon\ y_i^{m-r}=y_{r+i} \in I\!R,\ i=1(1)m-r,\ \alpha\upsilon\theta\alpha\acute{\iota}\rho\epsilon\tau\alpha. \tag{7.14}$$

Για την πρώτη των εξισώσεων (7.13), με βάση την (7.12), έχουμε καταρχάς ότι $\forall\ z \in I\!R^r \setminus \{0\}$ ισχύει

$$z^T M_{r,r} z = z^T (L_{r,r}^T L_{r,r} + L_{n-r,r}^T L_{n-r,r}) z = ||L_{r,r} z||_2^2 + ||L_{n-r,r} z||_2^2 > 0,$$

αφού ο $L_{r,r}$ είναι κάτω τριγωνικός με διαγώνια στοιχεία μονάδες, άρα αντιστρέψιμος, και επομένως η πρώτη από τις δύο norms θα είναι θετική ενώ η δεύτερη, γενικά, μή αρνητική. Συνεπώς η πρώτη των εξισώσεων (7.13) λύνεται ως προς y^r και δίνει

$$y^r \ = \ D_{r,r}^{-1} M_{r,r}^{-1} c^r. \tag{7.15}$$

Από τις (7.14) και (7.15) έχουμε ότι το διάνυσμα $y \in I\!R^m$ έχει τις $m-r$ τελευταίες συνιστώσες του αυθαίρετους πραγματικούς αριθμούς. Ετσι, έχοντας βρεί το $y = [y^{rT}\ y^{m-rT}]^T$ και έχοντας υπόψη την block μορφή του πίνακα U βρίσκουμε από την $U(P_2^T x) = y$ ότι

$$\left[\begin{array}{c|c} U_{r,r} & U_{r,m-r} \\ \hline 0_{m-r,r} & I_{m-r} \end{array}\right] \left[\begin{array}{c} (P_2^T x)^r \\ \hline (P_2^T x)^{m-r} \end{array}\right] \ = \ \left[\begin{array}{c} y^r \\ \hline y^{m-r} \end{array}\right], \tag{7.16}$$

όπου $(P_2^T x)^r \in I\!R^r$ και $(P_2^T x)^{m-r} \in I\!R^{m-r}$ διανύσματα με συνιστώσες τις r πρώτες και τις $m-r$ τελευταίες του διανύσματος $P_2^T x \in I\!R^m$. Το σύστημα (7.16) δίνει ισοδύναμα τα δύο συστήματα

$$\left\{ \begin{array}{c} U_{r,r}(P_2^T x)^r + U_{r,m-r}(P_2^T x)^{m-r} \ = \ y^r \\ (P_2^T x)^{m-r} \ = \ y^{m-r} \end{array}\right.$$

Από το δεύτερο σύστημα έχουμε αμέσως τις $m-r$ (αυθαίρετες) τελευταίες συνιστώσες του διανύσματος $P_2^T x$, οπότε αντικαθιστώντας στο πρώτο σύστημα και λύνοντας ως προς $(P_2^T x)^r$, με προς τα πίσω αντικατάσταση, βρίσκουμε τις r πρώτες συνιστώσες του $P_2^T x$, ως συναρτήσεις των $m-r$ αυθαίρετων τελευταίων συνιστωσών του. Τέλος, έχοντας το διάνυσμα $P_2^T x$, με πολλαπλασιασμό από τα αριστερά επί P_2, βρίσκουμε τη λύση x. Λόγω των $m-r$ αυθαίρετων συνιστωσών του $P_2^T x$ η λύση $x \in I\!R^m$ έχει επίσης $m-r$ αυθαίρετες συνιστώσες και τις άλλες r συνιστώσες της συναρτήσεις αυτών των $m-r$ αυθαίρετων. Αρα το σύστημα των κανονικών εξισώσεων (7.6) θα έχει άπειρες λύσεις οι οποίες θα βρίσκονται με τον τρόπο που έμμεσα, αλλά εξαντλητικά, περιγράφτηκε. □

Στην περίπτωση που το σύστημα των κανονικών εξισώσεων έχει μία μόνο λύση, οπότε ο $A^T A$ είναι πραγματικός, συμμετρικός και θετικά ορισμένος, τότε η άμεση μέθοδος Cholesky ίσως συνιστάται για την επίλυσή του.

Στη γενικότερη περίπτωση, όπου norms για μή τετραγωνικούς πίνακες καθώς και γενικευμένοι αντίστροφοι μή αντιστρέψιμων ή και μή τετραγωνικών πινάκων έχουν εισαχτεί, είναι δυνατόν να αποδειχτεί ότι ο δείκτης κατάστασης, που αντιστοιχεί στη φασματική norm του πίνακα $A^T A$, είναι ίσος με το τετράγωνο του δείκτη κατάστασης του πίνακα A του αρχικού προβλήματος. Ευνόητο είναι τότε ότι το αντίστοιχο σύστημα (7.6) θα είναι πολύ χειρότερης κατάστασης σε σχέση με το αρχικό. Θα δείξουμε του λόγου το αληθές στην τετριμμένη περίπτωση όπου $A \in I\!R^{n,n}$ με $\det(A) \neq 0$. Πράγματι, τότε

$$\kappa_2(A) = ||A||_2 ||A^{-1}||_2 = \rho^{\frac{1}{2}}(A^T A)\rho^{\frac{1}{2}}(A^{-T} A^{-1}).$$

ενώ για τον πίνακα A^TA θα είναι

$$\kappa_2(A^TA) = ||A^TA||_2||(A^TA)^{-1}||_2 = \rho^{\frac{1}{2}}((A^TA)^TA^TA)\rho^{\frac{1}{2}}((A^TA)^{-T}(A^TA)^{-1}) \quad (7.17)$$
$$= \rho(A^TA)\rho\left((A^TA)^{-1}\right) = \rho(A^TA)\rho\left(A^{-1}A^{-T}\right) \quad (7.18)$$
$$= \rho(A^TA)\rho\left(A^{-T}A^{-1}\right) = (\kappa_2(A))^2. \quad (7.19)$$

Σημειώνεται ότι η τρίτη από τα δεξιά έκφραση είναι ίση με τη δεύτερη από τα δεξιά γιατί οι πίνακες $A^{-1}A^{-T}$ και $A^{-T}A^{-1}$ είναι όμοιοι και άρα έχουν το ίδιο φάσμα ιδιοτιμών.

7.3 QR Ανάλυση (Παραγοντοποίηση)

Η "χειρότερη" κατάσταση την οποία έχει το σύστημα των κανονικών εξισώσεων σε σχέση με το αρχικό μπορεί να ξεπεραστεί με την παραγοντοποίηση του αρχικού πίνακα $A \in I\!R^{n,m}$ σε γινόμενο δύο παραγόντων, ενός πίνακα $Q \in I\!R^{n,m}$, τα διανύσματα-στήλες του οποίου είναι ανά δύο ορθογώνια με Ευκλείδεια norm μονάδα, δηλαδή $Q^TQ = I_m$, και ενός άνω τριγωνικού πίνακα $R \in I\!R^{m,m}$, με θετικά διαγώνια στοιχεία. Το πλεονέκτημα μιας τέτοιας ανάλυσης, με την προϋπόθεση ότι έχουν οριστεί φυσικές norms και αντίστροφοι μή τετραγωνικών πινάκων, είναι ότι ο νέος πίνακας συντελεστής QR έχει τον ίδιο δείκτη κατάστασης με τον πίνακα A ως προς την ℓ_2-norm. Για την εύρεση των πινάκων Q και R, οι οποίοι είναι μοναδικοί ανν τα διανύσματα-στήλες του A είναι γραμμικά ανεξάρτητα, μπορεί να χρησιμοποιηθεί μία από τις μεθόδους που θα πραγματευτούμε παρακάτω. Πριν όμως προχωρήσουμε θα παραθέσουμε μια γενικότερη πρόταση ευρεία χρήση της οποίας, άμεσα ή έμμεσα και χωρίς παραπέρα επεξηγήσεις, θα γίνεται στη συνέχεια του παρόντος κεφαλαίου.

Στο σημείο αυτό υπενθυμίζεται το παρακάτω λήμμα ευρεία χρήση του οποίου θα γίνει σε ό,τι ακολουθεί.

Λήμμα 7.1 *Αν ο $Q \in \mathbb{C}^{n,m}$, $m \leq n$, είναι ένας ορθομοναδιαίος πίνακας, δηλαδή $Q^HQ = I_m$, (ορθογώνιος στην περίπτωση που ο Q είναι πραγματικός), τότε για κάθε $x \in \mathbb{C}^m$ ισχύει ότι $\|Qx\|_2 = \|x\|_2$.*

Απόδειξη: Πραγματικά, έχουμε διαδοχικά $\|Qx\|_2^2 = (Qx)^H(Qx) = x^HQ^HQx = x^HI_mx = \|x\|_2^2$. □

Αν έχει πραγματοποιηθεί η παραγοντοποίηση QR τότε αποφεύγεται η λύση του προβλήματος με τη λύση των Κανονικών Εξισώσεων, που είναι ασταθής λόγω της δυναμικής αύξησης του δείκτη κατάστασης. Για το σκοπό αυτό, θεωρούμε πρώτα τον ορθοκανονικό πίνακα $\widetilde{Q} \in I\!R^{n,n-m}$($\widetilde{Q}^T\widetilde{Q} = I_{n-m}$), τ.ω. τα διανύσματα-στήλες του να είναι ορθογώνια με όλα τα διανύσματα-στήλες του Q, ($\widetilde{Q}^TQ = 0_{n-m,m}$). Με άλλα λόγια ο $\widetilde{Q}$ συμπληρώνει τον Q για τη δημιουργία ενός τετραγωνικού ορθογώνιου πίνακα $[Q \mid \widetilde{Q}] \in I\!R^{n,n}$. Το πως υπολογίζεται ο πίνακας $\widetilde{Q}$ εξετάζεται παρακάτω.

Ξεκινώντας από την τοποθέτηση του προβλήματος (7.2) έχουμε διαδοχικά

$$\begin{aligned}
\min_{x\in\mathbb{R}^m}\|r\|_2^2 &\equiv \min_{x\in\mathbb{R}^m}\|b-Ax\|_2^2 = \min_{x\in\mathbb{R}^m}\|b-QRx\|_2^2 \\
&= \min_{x\in\mathbb{R}^m}\|[Q\mid\widetilde{Q}]^T(b-QRx)\|_2^2 = \min_{x\in\mathbb{R}^m}\left\|\begin{bmatrix}Q^T\\ \widetilde{Q}^T\end{bmatrix}(b-QRx)\right\|_2^2 \\
&= \min_{x\in\mathbb{R}^m}\left\|\begin{bmatrix}Q^T\\ \widetilde{Q}^T\end{bmatrix}b-\begin{bmatrix}Q^T\\ \widetilde{Q}^T\end{bmatrix}QRx\right\|_2^2 = \min_{x\in\mathbb{R}^m}\left\|\begin{bmatrix}Q^Tb-Rx\\ \widetilde{Q}^Tb\end{bmatrix}\right\|_2^2 \\
&= \min_{x\in\mathbb{R}^m}||Q^Tb-Rx||_2^2+||\widetilde{Q}^Tb||_2^2. \qquad (7.20)
\end{aligned}$$

Ο δεύτερος όρος του δεξιού μέλους των (7.20) είναι σταθερός και ανεξάρτητος του διανύσματος x, επομένως η λύση x του προβλήματός μας προκύπτει από τη λύση του προβλήματος $\min_{x\in\mathbb{R}^m}||Q^Tb-Rx||_2$. Παρατηρούμε όμως ότι ο R είναι τετραγωνικός πίνακας και άρα η λύση του τελευταίου προβλήματος προκύπτει από το μηδενισμό του διανύσματος Q^Tb-Rx, δηλαδή για τα διανύσματα x που αποτελούν λύση του γραμμικού συστήματος

$$Rx = Q^Tb. \qquad (7.21)$$

Αν ο πίνακας R είναι αντιστρέψιμος, που σημαίνει ότι οι στήλες του A είναι γραμμικά ανεξάρτητες ή $\mathrm{rank}(A)=m$, τότε υπάρχει μια μοναδική λύση η

$$x = R^{-1}Q^Tb.$$

Στην περίπτωση που ο A είναι ελλιπούς βαθμού, $\mathrm{rank}(A)=r<m$, χρησιμοποιούμε την τεχνική της αναδιάταξης του πίνακα A, όπως ακριβώς και στο Θεώρημα 7.3. Τότε η ανάλυση QR δίνει

$$P_1AP_2 = QR = [Q_1\mid Q_2]\left[\begin{array}{c|c}R_{r,r} & R_{r,m-r}\\ \hline 0_{m-r,r} & 0_{m-r,m-r}\end{array}\right],$$

όπου ο Q διαχωρίστηκε σε $Q_1\in\mathbb{R}^{n,r}$ και $Q_2\in\mathbb{R}^{n,m-r}$ και ο R στην παραπάνω block μορφή με $R_{r,r}\in\mathbb{R}^{r,r}$ αντιστρέψιμο. Υλοποιώντας αυτές τις αλλαγές στην (7.20) θα έχουμε

$$\begin{aligned}
\min_{x\in\mathbb{R}^m}||b-Ax||_2^2 &= \min_{x\in\mathbb{R}^m}||P_1(b-Ax)||_2^2 = \min_{x\in\mathbb{R}^m}||P_1b-P_1AP_2P_2^Tx||_2^2 \\
&= \min_{y\in\mathbb{R}^m}||\widehat{b}-P_1AP_2y||_2^2 = \min_{y\in\mathbb{R}^m}||Q^T\widehat{b}-Ry||_2^2+||\widetilde{Q}^T\widehat{b}||_2^2, \qquad (7.22)
\end{aligned}$$

όπου $\widehat{b}=P_1b$, $y=P_2^Tx$ και $\widetilde{Q}\in\mathbb{R}^{n,n-m}$ ο πίνακας που θεωρήθηκε προηγουμένως. Αν θεωρήσουμε και τον αντίστοιχο block διαχωρισμό του y ($y=[y_1^T\mid y_2^T]^T$, $y_1\in\mathbb{R}^r$, $y_2\in\mathbb{R}^{m-r}$), ο πρώτος όρος του δεξιά μέλους της (7.22) δίνει

$$\begin{aligned}
\min_{y\in\mathbb{R}^m}||Q^T\widehat{b}-Ry||_2^2 &= \min_{y\in\mathbb{R}^m}\left\|\begin{bmatrix}Q_1^T\widehat{b}-R_{r,r}y_1-R_{r,m-r}y_2\\ Q_2^T\widehat{b}\end{bmatrix}\right\|_2^2 \\
&= \min_{y\in\mathbb{R}^m}||Q_1^T\widehat{b}-R_{r,r}y_1-R_{r,m-r}y_2||_2^2+||Q_2^T\widehat{b}||_2^2. \qquad (7.23)
\end{aligned}$$

Από την τελευταία έκφραση στις (7.23) προκύπτει ότι έχουμε άπειρες λύσεις στο πρόβλημα, που ταυτίζονται με τις λύσεις των γραμμικών συστημάτων

$$R_{r,r}y_1 = Q_1^T\widehat{b} - R_{r,m-r}y_2$$

όπου το y_2 παίρνεται αυθαίρετα στον $I\!R^{m-r}$. Οι λύσεις αυτές δίνονται από τις

$$y_1 = R_{r,r}^{-1}(Q_1^T\widehat{b} - R_{r,m-r}y_2), \quad \mu\epsilon \quad y_2 \in I\!R^{m-r} \quad \alpha\upsilon\theta\alpha\acute{\iota}\rho\epsilon\tau o.$$

Είμαστε τώρα σε θέση να δώσουμε τη λύση του προβλήματος ελάχιστων τετραγώνων με ανάλυση QR ξεκινώντας από μια διαφορετική θεώρηση του προβλήματος. Το διάνυσμα υπόλοιπο $b - Ax$ μπορεί να γραφτεί και ως

$$b - Ax = (I - QQ^T + QQ^T)b - QRx = (I - QQ^T)b + Q(Q^Tb - Rx). \tag{7.24}$$

Εύκολα αποδείχνεται ότι τα διανύσματα $(I - QQ^T)b$ και $Q(Q^Tb - Rx)$ είναι ορθογώνια. Πράγματι

$$b^T(I - QQ^T)^TQ(Q^Tb - Rx) = b^T(I - QQ^T)Q(Q^Tb - Rx) = b^T(Q - Q)(Q^Tb - Rx) = 0. \tag{7.25}$$

Εφαρμόζοντας το "Πυθαγόρειο" θεώρημα στην (7.24) έχουμε

$$||b - Ax||_2^2 \;=\; ||(I - QQ^T)b||_2^2 \;+\; ||Q(Q^Tb - Rx)||_2^2 \;=\; ||(I - QQ^T)b||_2^2 \;+\; ||Q^Tb - Rx||_2^2.$$

Προέκυψε έτσι η ίδια λύση (7.21) με τη διαφορά ότι η ελάχιστη ποσότητα είναι η $||(I - QQ^T)b||_2$ αντί της $||\widetilde{Q}^Tb||_2$ στην (7.20). Αναμφισβήτητα οι δύο αυτές ποσότητες θα πρέπει να είναι ίσες μεταξύ τους. Το όφελος στην προκείμενη περίπτωση είναι ότι δε χρειάζεται να προσδιοριστεί ο βοηθητικός ορθογώνιος πίνακας $\widetilde{Q}$.

Απομένει να δειχτεί ότι η QR ανάλυση δίνει την ίδια ακριβώς λύση με εκείνη της λύσης των κανονικών εξισώσεων. Πράγματι, στην περίπτωση όπου $\mathrm{rank}(A) = m$, ξεκινώντας από τη λύση των κανονικών εξισώσεων έχουμε

$$\begin{aligned} x &= (A^TA)^{-1}A^Tb = [(QR)^T(QR)]^{-1}(QR)^Tb = (R^TQ^TQR)^{-1}R^TQ^Tb \\ &= (R^TR)^{-1}R^TQ^Tb = R^{-1}(R^T)^{-1}R^TQ^Tb = R^{-1}Q^Tb. \end{aligned} \tag{7.26}$$

Στην περίπτωση όπου $\mathrm{rank}(A) = r < m$ θα πρέπει να ακολουθηθεί πρώτα μία ανάλυση του πίνακα A, όπως αυτή του Θεωρήματος 7.3, η οποία παραλείπεται.

Στη συνέχεια δίνουμε τις μεθόδους με τις οποίες πραγματοποιείται η QR ανάλυση.

7.3.1 Gram-Schmidt **Ορθογωνιοποίηση**

Στην παρούσα παράγραφο θα εκθέσουμε τη μέθοδο και τον αλγόριθμο της Gram-Schmidt Ορθογωνιοποίησης ξεκινώντας από το βασικό της θεώρημα. Προφανώς ο υπόψη αλγόριθμος μπορεί να δοθεί ως εφαρμογή αυτού των Συζυγών Διευθύνσεων. Εντούτοις δίνεται εξαρχής για την ολοκληρωμένη παρουσίαση της ανάλυσης QR του παρόντος κεφαλαίου.

Θεώρημα 7.4 *Εστω $A \in I\!R^{n,m}$, $m \leq n$, τα διανύσματα-στήλες a_i, $i = 1(1)m$, του οποίου είναι γραμμικά ανεξάρτητα. Υπάρχουν μοναδικοί πίνακες $Q \in I\!R^{n,m}$, με $Q^TQ = I_m$, και άνω τριγωνικός πίνακας $R \in I\!R^{m,m}$, με $r_{ii} > 0$, $i = 1(1)m$, τ.ω.*

$$A \;=\; QR. \tag{7.27}$$

(Σημείωση: Διαπιστώνεται αμέσως ότι η απαίτηση $Q^TQ = I_m$ είναι ισοδύναμη με την απαίτηση τα διανύσματα-στήλες $q^{(i)}$, $i = 1(1)m$, του πίνακα Q, να είναι ανά δύο ορθογώνια με Ευκλείδεια norm *μονάδα.)*

Απόδειξη: Η αποδεικτέα σχέση (7.27) γράφεται αναλυτικά ως εξής

$$[a_1\ a_2\ a_3\ \cdots\ a_m] \;=\; \left[q^{(1)}\ q^{(2)}\ q^{(3)}\ \cdots\ q^{(m)}\right] \begin{bmatrix} r_{11} & r_{12} & r_{13} & \cdots & r_{1m} \\ & r_{22} & r_{23} & \cdots & r_{2m} \\ & & r_{33} & \cdots & r_{3m} \\ & & & \ddots & \vdots \\ & & & & r_{mm} \end{bmatrix}, \tag{7.28}$$

και είναι ισοδύναμη με τις m ισότητες

$$\begin{cases} r_{11}q^{(1)} \;=\; a_1 \\ r_{12}q^{(1)} + r_{22}q^{(2)} \;=\; a_2 \\ r_{13}q^{(1)} + r_{23}q^{(2)} + r_{33}q^{(3)} \;= a_3 \\ \vdots \\ r_{1m}q^{(1)} + r_{2m}q^{(2)} + r_{3m}q^{(3)} + \cdots + r_{m-1,m}q^{(m-1)} + r_{mm}q^{(m)} \;= a_m, \end{cases} \tag{7.29}$$

όπως μπορεί πολύ εύκολα να διαπιστωθεί. Από την πρώτη ισότητα, το γεγονός ότι το διάνυσμα $q^{(1)}$ είναι κανονικοποιημένο ($\|q^{(1)}\|_2 = 1$) και από την απαίτηση $r_{11} > 0$, προκύπτει ότι $r_{11} = \|a_1\|_2$. Λύνοντας στη συνέχεια ως προς $q^{(1)}$ έχουμε $q^{(1)} = \frac{a_1}{r_{11}}$. Από τη δεύτερη ισότητα, για να προσδιορίσουμε το r_{12}, θεωρούμε το εσωτερικό γινόμενο αυτής με το $q^{(1)}$. Έτσι, $r_{12}(q^{(1)}, q^{(1)})_2 + r_{22}(q^{(2)}, q^{(1)})_2 = (a_2, q^{(1)})_2$ και λόγω της απαίτησης τα $q^{(1)}$ και $q^{(2)}$ να είναι ορθογώνια, έχουμε ότι $r_{12} \;=\; (a_2, q^{(1)})_2$. Στη συνέχεια η δεύτερη ισότητα γίνεται $r_{22}q^{(2)} = a_2 - r_{12}q^{(1)}$ και τα r_{22} και $q^{(2)}$ υπολογίζονται με το ίδιο σκεπτικό που υπολογίστηκαν τα r_{11} και $q^{(1)}$ αντίστοιχα, από την

πρώτη. Έτσι, $r_{22} = \|a_2 - r_{12}q^{(1)}\|_2$ και $q^{(2)} = \frac{a_2 - r_{12}q^{(1)}}{r_{22}}$. Είναι προφανές ότι $q^{(2)} \neq 0$ διότι σε αντίθετη περίπτωση, τα a_1 και a_2 θα ήταν γραμμικά εξαρημένα. Η απόδειξη μπορεί να ολοκληρωθεί με επαγωγή και είναι προφανής. Η γενική έκφραση για το τυχόν $q^{(i)}, i = 1(1)m$, θα δίνεται αναδρομικά από τις παρακάτω σχέσεις

$$q^{(i)} = a_i - \sum_{j=1}^{i-1} r_{ji}q^{(j)}, \quad r_{ji} = (a_i, q^{(j)})_2, j = 1(1)i-1, \quad i = 1(1)m. \tag{7.30}$$

Η απόδειξη για τη μοναδικότητα των Q και R μπορεί να προκύψει με θεώρηση ανάλυσης του A σε ένα άλλο γινόμενο $Q'R'$, όπου οι Q' και R' θα είναι αντίστοιχων μορφών με αυτές των Q και R στην (7.28) και με αντίστοιχες ιδιότητες. Η ισότητα $PR = P'R'$, η εξίσωση των αντίστοιχων στηλών των γινομένων τους και εφαρμογή απλής επαγωγής αποδείχνει τη μοναδικότητα. □

Έχουμε να παρατηρήσουμε εδώ ότι το παραπάνω θεώρημα αποτελεί ειδική περίπτωση και παραλλαγή του Θεωρήματος 6.4 Gram-Schmidt του προηγούμενου Κεφαλαίου.

Στη συνέχεια δίνουμε τον αλγόριθμο της QR ανάλυσης με Gram-Schmidt ορθογωνιοποίηση, θεωρώντας ότι $\text{rank}(A) = m$, όπως διατυπώθηκε στο αντίστοιχο θεώρημα. (Σημείωση: Αν $\text{rank}(A) = r < m$, τότε θα πρέπει να προβλεφτούν μεταθέσεις στηλών του A ώστε οι γραμμικά ανεξάρτητες στήλες του να τοποθετηθούν στις πρώτες στη σειρά στήλες, οπότε και η διατύπωση του αντίστοιχου θεωρήματος θα πρέπει να δοθεί ανάλογα.)

Αλγόριθμος QR Ανάλυσης με Gram-Schmidt **Ορθογωνιοποίηση:**

Δεδομένα: $A \in I\!R^{n,m}$ με διανύσματα-στήλες $a_i,\ i = 1(1)m$, γραμμικά ανεξάρτητα.
Για $i = 1(1)m$
    $q^{(i)} = a_i$
    Για $j = 1(1)i-1$
        $r_{ji} = (a_i, q^{(j)})_2$ ή $r_{ji} = (q^{(i)}, q^{(j)})_2$
        $q^{(i)} = q^{(i)} - r_{ji}q^{(j)}$
    Τέλος 'Για'
    $r_{ii} = ||q^{(i)}||_2$
    $q^{(i)} = q^{(i)}/r_{ii}$
Τέλος 'Για'

Παρατηρούμε ότι στην εσωτερική ανακύκλωση του αλγόριθμου έχουμε δύο ισοδύναμες διαδικασίες. Η πρώτη δίνει τον κλασικό αλγόριθμο ενώ η δεύτερη τον παραλλαγμένο αλγόριθμο Gram-Schmidt ορθογωνιοποίησης. Η πρώτη διαδικασία προκύπτει άμεσα από τη θεώρηση του προβλήματος $A = QR$. Η ισοδυναμία της δεύτερης με την πρώτη αποδείχνεται ως εξής:
Είναι φανερό ότι κατά την i-οστή στη σειρά εξωτερική επανάληψη και κατά την είσοδό της στην

j-οστή εσωτερική επανάληψη, το διάνυσμα $q^{(i)}$ θα είναι

$$q^{(i)} \;=\; a_i - \sum_{k=1}^{j-1} r_{ki} q^{(k)}. \tag{7.31}$$

Το εσωτερικό γινόμενο αυτού με το $q^{(j)}$ δίνει

$$(q^{(i)}, q^{(j)})_2 \;=\; (a_i, q^{(j)})_2 - \sum_{k=1}^{j-1} r_{ki} (q^{(k)}, q^{(j)})_2 \;=\; (a_i, q^{(j)})_2, \tag{7.32}$$

λόγω της ορθογωνιότητας των διανυσμάτων $q^{(k)}, k = 1(1)j-1$, με το $q^{(j)}$.

Πρέπει να τονιστεί εδώ ότι στην πράξη προτιμιέται ο παραλλαγμένος αλγόριθμος διότι είναι περισσότερο ευσταθής από τον αντίστοιχο κλασικό, κυρίως σε προβλήματα όπου οι στήλες του A είναι "σχεδόν" γραμμικά εξαρτημένες. Αυτό οφείλεται στο γεγονός ότι με τον παραλλαγμένο αλγόριθμο συνεχίζουμε τους υπολογισμούς αφαιρώντας από τα a_i τις ορθογώνιες διευθύνσεις $q^{(k)},\ k = 1(1)j-1$, και "εργαζόμαστε" με τις υπόλοιπες, ενώ στον κλασικό υπολογίζονται κάθε φορά και τα εσωτερικά γινόμενα $(q^{(j)}, q^{(k)})_2,\ k = 1(1)j-1$, τα οποία στην πράξη δεν είναι ακριβώς μηδέν και άρα δίνουν σφάλματα η μετάδοση των οποίων από βήμα σε βήμα έχει ως συνέπεια τη συσσώρευσή τους και την αλλοίωση του αποτελέσματος.

Σημείωση: Εχουμε να παρατηρήσουμε ότι ο αλγόριθμος Gram-Schmidt όπως ακριβώς δίνεται, μπορεί να χρησιμεύσει για την εύρεση όρθογώνιων βάσεων σε χώρους Hilbert. Για παράδειγμα, αν θεωρήσουμε το χώρο των συνεχών και φραγμένων πραγματικών συναρτήσεων στο διάστημα $[\alpha, \beta]$ και το εσωτερικό γινόμενο: $(u, v) = \int_\alpha^\beta uvdx$, ορίζεται ένας απειροδιάστατος χώρος Hilbert. Επιλέγοντας ως a_i μια ακολουθία γραμμικά ανεξάρτητων συναρτήσεων, ο αλγόριθμος Gram-Schmidt παράγει την q_i, μια ακολουθία ορθογώνιων συναρτήσεων, που αποτελεί ορθογώνια βάση. Εφαρμογή του αλγόριθμου Gram-Schmidt έχουμε στη Θεωρία Προσέγγισης, για την παραγωγή βάσεων ορθογώνιων πολυωνύμων.

7.3.2 Μετασχηματισμοί ή Ανακλάσεις (Reflections) Householder

Η QR ανάλυση με μετασχηματισμούς Householder πραγματοποιείται διαμέσου μιας ακολουθίας ορθογώνιων μετασχηματισμών όπου κάθε φορά απαλείφεται το κάτω από τη διαγώνιο τμήμα μιας στήλης του πίνακα A.

Ορισμός 7.1 *Ενας ορθογώνιος πίνακας $P \in I\!R^{n,n}$ καλείται μετασχηματισμός ή ανάκλαση* Householder *αν ορίζεται διαμέσου διανύσματος $u \in I\!R^n$, με $||u||_2 = 1$, έτσι ώστε*

$$P \;=\; I - 2uu^T. \tag{7.33}$$

Ο πίνακας P στην (7.33) είναι πράγματι ορθογώνιος αφού $P^TP = (I-2uu^T)^T(I-2uu^T) = (I-2uu^T)^2 = I-4uu^T+4u(u^Tu)u^T = I-4uu^T+4uu^T = I$. Ο μετασχηματισμός ονομάστηκε και ανάκλαση από το εξής γεγονός: Εστω θ_1 η γωνία που σχηματίζει το τυχαίο διάνυσμα $x \in I\!R^n\backslash\{0\}$ με το u και θ_2 η αντίστοιχη γωνία του Px με το u. Εύκολα αποδείχνεται ότι οι γωνίες αυτές είναι παραπληρωματικές. Πράγματι, $\cos\theta_1 = \frac{u^Tx}{||u||_2||x||_2} = \frac{u^Tx}{||x||_2}$ ενώ $\cos\theta_2 = \frac{u^TPx}{||u||_2||Px||_2} = \frac{u^T(I-2uu^T)x}{||x||_2} = \frac{u^Tx-2(u^Tu)u^Tx}{||x||_2} = \frac{u^Tx-2u^Tx}{||x||_2} = -\frac{u^Tx}{||x||_2}$ = -$\cos\theta_1$. Αυτό σημαίνει ότι αν θεωρήσουμε τη θ_1 ως γωνία πρόσπτωσης, η θ_2 θεωρείται ως γωνία ανάκλασης.

Σημείωση: Για την απλοποίηση της ανάλυσης, τόσο στην παρούσα παράγραφο όπως και στην επόμενη, εκτός κι αν σαφώς τονίζεται αλλιώς, θα θεωρούμε ότι ο πίνακας $A \in I\!R^{n,m},\ m \leq n$, είναι πλήρους βαθμού m, επιπλέον δε οι m πρώτες γραμμές του είναι γραμμικά ανεξάρτητες και όλοι οι $m-1$ κύριοι υποπίνακες της άνω αριστερής γωνίας του είναι αντιστρέψιμοι.

Στη συνέχεια θεωρούμε την εφαρμογή των μετασχηματισμών Householder στην QR ανάλυση, στόχος της οποίας είναι η απαλοιφή των κάτω από τη διαγώνιο στοιχείων του πίνακα A, όπου για την απλούστευση της ανάλυσης υποθέτουμε ότι τα αντίστοιχα "οδηγά" στοιχεία είναι διαφορετικά από το μηδέν. Την πρώτη φορά θα πρέπει να απαλειφτούν όλα τα στοιχεία της πρώτης στήλης εκτός από το πρώτο. Εστω ότι $y = [a_{11}\, a_{21}\, \ldots\, a_{n1}]^T \in I\!R^n$ είναι το διάνυσμα της πρώτης στήλης του A. Θα πρέπει να προσδιορίσουμε το u έτσι ώστε

$$P_1y \;=\; (I-2uu^T)y \;=\; ce^1, \tag{7.34}$$

όπου e^1 είναι το διάνυσμα της πρώτης στήλης του $n \times n$ μοναδιαίου πίνακα I. Η (7.34) γίνεται

$$(I-2uu^T)y \;=\; ce^1 \;\Leftrightarrow\; y-2(u^Ty)u \;=\; ce^1 \;\Leftrightarrow\; 2(u^Ty)u \;=\; y-ce^1.$$

Η τελευταία έκφραση υποδείχνει ότι το διάνυσμα u είναι γραμμικός συνδυασμός των διανυσμάτων y και e^1. Η σταθερά c προσδιορίζεται από την (7.34) αν πάρουμε norms. Συγκεκριμένα

$$||P_1y||_2 \;=\; ||ce^1||_2 \;\Leftrightarrow\; |c| \;= ||y||_2 \;\Leftrightarrow\; c \;=\; \pm||y||_2.$$

Συμβολίζουμε με $\widetilde{u}$ το διάνυσμα $y \pm ||y||_2e^1$. Αυτό θα είναι παράλληλο προς το u, επομένως $u = \frac{\widetilde{u}}{||\widetilde{u}||_2}$. Παρατηρούμε ότι έχουμε δύο λύσεις στο πρόβλημά μας, που σημαίνει ότι η QR ανάλυση δεν θα είναι μονοσήμαντα ορισμένη, πράγμα που έρχεται σε αντίφαση με την Gram-Schmidt ορθογωνιοποίηση όπου αποδείχτηκε το μονοσήμαντο της QR ανάλυσης. Αυτό εξηγείται απλά από το γεγονός ότι στον αλγόριθμο των Gram-Schmidt απαιτήσαμε τα διαγώνια στοιχεία του R να είναι θετικά. Ετσι, η QR ανάλυση είναι μονοσήμαντα ορισμένη ανεξάρτητα από το θεωρούμενο πρόσημο. Πράγματι, αν αλλάξουμε το πρόσημο όλων των στοιχείων μιας στήλης του Q και της αντίστοιχης γραμμής του R, έχουμε διαφορετική ανάλυση που δεν επηρεάζει όμως το αποτέλεσμα του γραμμικού προβλήματος των ελάχιστων τετραγώνων. Στην πράξη προτιμούμε να παίρνουμε το πρόσημο του στοιχείου y_1 ώστε $|\widetilde{u}_1| > |y_1|$. Αυτό γίνεται για την αποφυγή της περίπτωσης όπου το u_1 θα γίνει μηδέν λόγω των σφαλμάτων που θα υπεισέλθουν κατά τους υπολογισμούς.

Παίρνοντας επομένως ως y το διάνυσμα της πρώτης στήλης του A, κατασκευάζοντας το μετασχηματισμό Householder $P_1 \in I\!R^{n,n}$, όπως περιγράφτηκε και θεωρώντας το γινόμενο P_1A, θα έχουμε

$$A_1 \;=\; P_1A \;=\; \left[\begin{array}{r|rrcr} a_{11}^{(1)} & a_{12}^{(1)} & a_{13}^{(1)} & \cdots & a_{1m}^{(1)} \\ \hline 0 & a_{22}^{(1)} & a_{23}^{(1)} & \cdots & a_{2m}^{(1)} \\ 0 & a_{32}^{(1)} & a_{33}^{(1)} & \cdots & a_{3m}^{(1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & a_{n2}^{(1)} & a_{n3}^{(1)} & \cdots & a_{nm}^{(1)} \end{array}\right],$$

όπου θέσαμε άνω δείκτη 1 στα μή μηδενικά στοιχεία του A_1 για να διακρίνονται από εκείνα του A. Ακολουθεί η απαλοιφή των κάτω από τη διαγώνιο στοιχείων της δεύτερης στήλης του πίνακα A_1 παίρνοντας ως $y = [a_{22}^{(1)}\ a_{32}^{(1)} \cdots a_{n2}^{(1)}]^T \in I\!R^{n-1}$. Κατασκευάζουμε έτσι το μετασχηματισμό Householder $Q_2 \in I\!R^{n-1,n-1}$. Ο αντίστοιχος μετασχηματισμός Householder διάστασης n θα είναι

$$P_2 \;=\; \left[\begin{array}{c|c} 1 & 0_{n-1}^T \\ \hline 0_{n-1} & Q_2 \end{array}\right],$$

και ο A_2 θα δίνεται ως

$$A_2 \;=\; P_2A \;=\; \left[\begin{array}{rr|ccr} a_{11}^{(1)} & a_{12}^{(1)} & a_{13}^{(1)} & \cdots & a_{1m}^{(1)} \\ 0 & a_{22}^{(2)} & a_{23}^{(2)} & \cdots & a_{2m}^{(2)} \\ \hline 0 & 0 & a_{33}^{(2)} & \cdots & a_{3m}^{(2)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & a_{n3}^{(2)} & \cdots & a_{nm}^{(2)} \end{array}\right].$$

Επαγωγικά, κατά την k επανάληψη θα έχουμε να απαλείψουμε το κάτω της διαγωνίου τμήμα της k στη σειρά στήλης. Θα κατασκευάσουμε, λοιπόν, τον Q_k παίρνοντας ως $y \in I\!R^{n-k+1}$ το $[a_{kk}^{(k-1)}\ a_{k+1,k}^{(k-1)} \cdots a_{nk}^{(k-1)}]^T$, και ο P_k θα είναι

$$P_k \;=\; \left[\begin{array}{c|c} I_{k-1} & 0_{k-1,n-k+1} \\ \hline 0_{n-k+1,k-1} & Q_k \end{array}\right].$$

Τέλος, μετά την m επανάληψη θα έχουμε

$$A_m = P_m A_{m-1} = P_m P_{m-1} A_{m-2} = \cdots = P_m P_{m-1} P_{m-2} \cdots P_2 P_1 A$$

$$= \begin{bmatrix} a_{11}^{(1)} & a_{12}^{(1)} & a_{13}^{(1)} & \cdots & a_{1,m-1}^{(1)} & a_{1m}^{(1)} \\ 0 & a_{22}^{(2)} & a_{23}^{(2)} & \cdots & a_{2,m-1}^{(2)} & a_{2m}^{(2)} \\ 0 & 0 & a_{33}^{(3)} & \cdots & a_{3,m-1}^{(3)} & a_{3m}^{(3)} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & a_{m-1,m-1}^{(m-1)} & a_{m-1,m}^{(m-1)} \\ 0 & 0 & 0 & \cdots & 0 & a_{mm}^{(m)} \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 0 \end{bmatrix}. \tag{7.35}$$

Από αυτές προκύπτει ότι

$$\begin{aligned} A &= (P_m P_{m-1} P_{m-2} \cdots P_2 P_1)^{-1} A_m = P_1^T P_2^T \cdots P_{m-1}^T P_m^T A_m \\ &= P_1 P_2 \cdots P_{m-1} P_m A_m = P A_m, \end{aligned} \tag{7.36}$$

όπου P ονομάσαμε το γινόμενο $P_1 P_2 \cdots P_{m-1} P_m$, που είναι κι αυτό ορθογώνιος πίνακας. Από την (7.36) είναι προφανές ότι η QR ανάλυση ολοκληρώθηκε. Ως άνω τριγωνικό πίνακα R έχουμε τον πίνακα που αποτελείται από τις πρώτες m γραμμές του A_m ενώ ως Q τον πίνακα που αποτελείται από τις πρώτες m στήλες του P. Ουσιαστικά ο πίνακας P είναι ο πίνακας $[Q \mid \widetilde{Q}]$ που είχαμε θεωρήσει στην (7.20).

Για την εύρεση του κόστους του αλγόριθμου της QR μεθόδου με ανακλάσεις Householder εργαζόμαστε ως εξής. Καταρχάς είναι προφανές ότι όχι μόνο δε θα πραγματοποιηθούν οι πολλαπλασιασμοί μεταξύ πινάκων, όπως φαίνεται στην (7.35), αλλά ούτε καν θα υπολογιστούν οι πίνακες $P_k,\ k = 1(1)m-1$. Θα ακολουθηθεί η επαναληπτική διαδικασία που ορίζει η (7.35) για τον υπολογισμό των A_k κάνοντας πράξεις απ΄ ευθείας μεταξύ των διανυσμάτων $\widetilde{u}_k$ και των πινάκων A_k. Ουσιαστικά υπολογίζεται το διάνυσμα $\widetilde{u}_k$ και στη συνέχεια το

$$(I - \frac{2}{||\widetilde{u}_k||_2^2} \widetilde{u}_k \widetilde{u}_k^T) A_k = A_k - \frac{2}{||\widetilde{u}_k||_2^2} \widetilde{u}_k (\widetilde{u}_k^T A_k).$$

Ακόμη, επειδή κατά την k επανάληψη αλλάζουν μόνο τα στοιχεία a_{ij} του A για τα οποία $k \le i \le m$ και $k \le j \le n$, οι πράξεις θα γίνονται μόνο στο αντίστοιχο block του πίνακα A. Μετά από τις παρατηρήσεις αυτές είμαστε σε θέση να δώσουμε το νέο αλγόριθμο της QR ανάλυσης.

Αλγόριθμος QR **Ανάλυσης με Ανακλάσεις** Householder:

Δεδομένα: $A \in I\!R^{n,m}$.
Για $k = 1(1)\min\{n-1, m\}$

$c = \sqrt{\sum_{j=k}^{n} a_{jk}^2}$ Υπολογισμός της $||y||_2$
$u_k = a_{kk} + \text{sign}(a_{kk})c$
Για $i = k+1(1)n$
$u_i = a_{ik}$ Υπολογισμός του $\widetilde{u}_k$
Τέλος 'Για'
$d = 2c(c + |a_{kk}|)$ Υπολογισμός της $||\widetilde{u}_k||_2^2$
$\gamma = 2/d$
Για $j = k(1)m$
$v_j = \sum_{l=k}^{n} u_l a_{lj}$ Υπολογισμός του $\widetilde{u}_k^T A_k$
$v_j = \gamma v_j$
Τέλος 'Για'
$a_{kk} = a_{kk} - u_k v_k$
Για $i = k(1)n$
Για $j = k+1(1)m$
$a_{ij} = a_{ij} - u_i v_j$ Υπολογισμός του $A_k - \widetilde{u}_k(\frac{2}{||\widetilde{u}_k||_2^2}\widetilde{u}_k^T A_k)$
Τέλος 'Για'
Τέλος 'Για'
Τέλος 'Για'

Για τον υπολογισμό του κόστους του αλγόριθμου, παρατηρούμε ότι σε κάθε επανάληψη έχουμε την εύρεση μιας norm ($||y||_2$), την εύρεση $m-k+1$ εσωτερικών γινομένων τάξης $n-k+1$ ($\widetilde{u}_k^T A_k$), την εκτέλεση ενός πολλαπλασιασμού σταθεράς επί διάνυσμα τάξης $m-k+1$ ($\frac{2}{||\widetilde{u}_k||_2}\widetilde{u}_k^T A_k$) και την εύρεση του κάτω δεξιά $(n-k+1)\times(m-k)$ block, που απαιτεί έναν πολλαπλασιασμό και μία αφαίρεση για κάθε στοιχείο. Το κόστος επομένως, σε πλήθος πράξεων, οφείλεται κυρίως στον τελευταίο υπολογισμό των στοιχείων του block και στα $m-k+1$ εσωτερικά γινόμενα. Κάνοντας τους υπολογισμούς βρίσκεται ότι το πλήθος πράξεων, αγνοώντας όρους μικρότερου βαθμού, είναι $m^2n - \frac{m^3}{3}$ πολλαπλασιασμοί και $\frac{m^2n}{2} - \frac{m^3}{6}$ προσθαφαιρέσεις. Επιπλέον, η λύση του τριγωνικού συστήματος με προς τα πίσω αντικατάσταση απαιτεί κόστος μικρότερης τάξης $\mathcal{O}(m^2)$. Για το κόστος του αλγόριθμου σε θέσεις μνήμης, εύκολα φαίνεται ότι ο πίνακας R μπορεί να αποθηκευτεί στο άνω τριγωνικό μέρος του πίνακα A. Οπως προαναφέρθηκε θα ήταν δαπανηρό και ως προς το πλήθος των πράξεων και ως προς τις θέσεις μνήμης, να υπολογίζονται και να αποθηκεύονται οι ορθογώνιοι πίνακες $P_k,\ k = 1(1)m$. Εκείνο που γίνεται στην πράξη είναι να αποθηκεύονται μόνο τα διανύσματα $\widetilde{u}_k$ τα οποία πολλαπλασιάζονται απ΄ ευθείας με τους πίνακες A_k. Τα διανύσματα αυτά έχουν μή μηδενικά στοιχεία από την k θέση και κάτω, επομένως αποθηκεύονται ως στήλες στο κάτω τριγωνικό μέρος του πίνακα A, εκτός από τα πρώτα στοιχεία τους που μπορούν να αποθηκευτούν σε ένα m-διάστατο διάνυσμα. Ακόμη, ένα m-διάστατο διάνυσμα χρειάζεται για την αποθήκευση των ποσοτήτων $\gamma = \frac{2}{||\widetilde{u}_k||_2^2}$ που παρουσιάζονται στον αλγόριθμο. Επιπλέον, για τη λύση του προβλήματος ελάχιστων τετραγώνων θα χρειαστεί και η αποθήκευση του n-διάστατου διανύσματος b, στις πρώτες m θέσεις του οποίου θα αποθηκευτεί τελικά η λύση x.

Στη συνέχεια παραθέτουμε ένα αριθμητικό παράδειγμα γραμμικού προβλήματος ελάχιστων τετραγώνων. Αν

$$A = \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \\ 1 & 1 & 1 \end{bmatrix}, \quad b = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \tag{7.37}$$

είναι τα δεδομένα του προβλήματος ελάχιστων τετραγώνων, τότε εφαρμόζοντας τον αλγόριθμο QR ανάλυσης με ανακλάσεις Householder έχουμε

$$Q = \begin{bmatrix} -0.8165 & 0 & 0.3106 \\ 0.4082 & -0.6386 & 0.0565 \\ 0 & 0.4264 & -0.7625 \\ -0.4082 & -0.6396 & -0.5648 \end{bmatrix}, \quad R = \begin{bmatrix} -2.4495 & 1.2247 & -0.8165 \\ 0 & -2.3452 & 0.8528 \\ 0 & 0 & -2.1462 \end{bmatrix} \tag{7.38}$$

και η λύση του προβλήματος είναι η $x = [0.4474\ 0.5263\ 0.4474]^T$. Ας σημειωθεί ότι το συγκεκριμένο παράδειγμα επιλύθηκε χρησιμοποιώντας το πακέτο προγραμμάτων MATLAB και τα αποτελέσματα δόθηκαν με τέσσερα δεκαδικά ψηφία.

7.3.3 Στροφές Givens

Ενας τρίτος τρόπος για την επίτευξη της QR ανάλυσης είναι η χρησιμοποίηση των ορθογώνιων μετασχηματισμών, των καλούμενων στροφών Givens. Είναι γνωστό ότι ο 2×2 πίνακας της μορφής $R(\theta) = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$, είναι ορθογώνιος και αν πολλαπλασιάσει από τα αριστερά ένα διάνυσμα $x \in I\!R^2$ το στρέφει κατά γωνία θ κατά τη θετική φορά. Κατ΄ επέκταση ο πίνακας $R_{kj}(\theta) \in I\!R^{n,n}$ της μορφής

$$R_{kj}(\theta) = \begin{array}{c} \\ \\ \\ \\ k \\ \\ \\ \\ j \\ \\ \\ \end{array} \begin{array}{c} \begin{array}{ccccccccccc} & & & & k & & & & j & & \end{array} \\ \begin{bmatrix} 1 & & & & & & & & & & \\ & 1 & & & & & & & & & \\ & & \ddots & & & & & & & & \\ & & & & \cos\theta & & & & -\sin\theta & & \\ & & & & & 1 & & & & & \\ & & & & & & \ddots & & & & \\ & & & & & & & 1 & & & \\ & & & & \sin\theta & & & & \cos\theta & & \\ & & & & & & & & & \ddots & \\ & & & & & & & & & & 1 \end{bmatrix} \end{array}$$

είναι ορθογώνιος και αν εφαρμοστεί με τον ίδιο τρόπο σε ένα n-διάστατο διάνυσμα το στρέφει ως προς τις συνιστώσες k και j ενώ το αφήνει αναλλοίωτο ως προς τις άλλες συνιστώσες. Ο πίνακας αυτός ονομάζεται στροφή Givens και ο αντίστοιχος μετασχηματισμός, μετασχηματισμός Givens. Οι στροφές Givens χρησιμεύουν ως ορθογώνιοι μετασχηματισμοί για την επίτευξη της QR ανάλυσης μηδενίζοντας κάθε φορά και ένα στοιχείο κάτω από τη διαγώνιο ξεκινώντας από την πρώτη στήλη και προχωρώντας προς τα δεξιά.

Για να προσδιορίσουμε τη στροφή $R_{kj}(\theta),\ j > k$, ώστε να μηδενιστεί το (j,k) στοιχείο του πίνακα $A \in I\!R^{n,n}$, $\mathrm{rank}(A) = m < n$, συμβολίζουμε με c το $\cos\theta$ και με s το $\sin\theta$ και μηδενίζουμε το (j,k) στοιχείο του γινομένου $R_{kj}(\theta)A$. Τότε,

$$sa_{kk} \ + \ ca_{jk} \ = \ 0$$

και η λύση της μας δίνει τις τιμές

$$s \ = \ -\frac{a_{jk}}{\sqrt{a_{kk}^2 + a_{jk}^2}}, \quad c \ = \ \frac{a_{kk}}{\sqrt{a_{kk}^2 + a_{jk}^2}}.$$

Με την προϋπόθεση ότι έχουν μηδενιστεί τα στοιχεία $a_{pq},\ q = 1(1)k-1,\ p = q+1(1)n$, κάτω από τη διαγώνιο, εκείνο που θα αλλάξει στον πίνακα A μετά την εφαρμογή του $R_{kj}(\theta)$ θα είναι μόνο οι γραμμές k και j, η πρώτη από το k στοιχείο και δεξιά ενώ η δεύτερη, με το μηδενισμό του (j,k) στοιχείου της, από το $k+1$ στοιχείο και δεξιά. Το (k,k) στοιχείο θα γίνει

$$a'_{kk} \ = \ ca_{kk} \ - \ sa_{jk} \ = \ \sqrt{a_{kk}^2 + a_{jk}^2}.$$

Τα υπόλοιπα στοιχεία της k γραμμής θα γίνουν

$$a'_{ki} \ = \ ca_{ki} \ - \ sa_{ji} \ = \ \frac{a_{kk}a_{ki} + a_{jk}a_{ji}}{\sqrt{a_{kk}^2 + a_{jk}^2}},\ i = k+1(1)m,$$

ενώ τα αντίστοιχα στοιχεία της j γραμμής

$$a'_{ji} \ = \ sa_{ki} \ + \ ca_{ji} \ = \ \frac{-a_{jk}a_{ki} + a_{kk}a_{ji}}{\sqrt{a_{kk}^2 + a_{jk}^2}},\ i = k+1(1)m.$$

Με την απλούστευση της υπόθεσης ότι τα "οδηγά" στοιχεία δε μηδενίζονται, είναι φανερό ότι στις πρώτες $k-1$ στήλες δεν επέρχεται καμία μεταβολή αφού τα κάτω από τη διαγώνιο στοιχεία έχουν μηδενιστεί και η εφαρμογή του μετασχηματισμού δίνει πάλι μηδέν. Ακόμη, πρέπει να τονιστεί ότι δεν παίζει ρόλο η σειρά που θα απαλειφτούν τα στοιχεία της ίδιας στήλης. Αν επιλέξουμε την ορθή σειρά, από πάνω προς τα κάτω, θα πάρουμε την εξής ακολουθία πινάκων

$$\begin{aligned} &R_{12}A,\ R_{13}R_{12}A,\ R_{14}R_{13}R_{12}A,\ \cdots,\ R_{1n}R_{1,n-1}\cdots R_{13}R_{12}A, \\ &\quad R_{23}(R_{1n}\cdots R_{12})A,\ \cdots,\ (R_{2n}\cdots R_{23})(R_{1n}\cdots R_{12})A, \\ &\cdots,\ (R_{mn}\cdots R_{m,m+1})\cdots(R_{2n}\cdots R_{23})(R_{1n}\cdots R_{12})A. \end{aligned} \tag{7.39}$$

Οι μετασχηματισμοί Givens είναι τόσοι, όσα και τα στοιχεία που πρέπει να απαλειφτούν, δηλαδή $nm - \frac{m(m+1)}{2}$. Με την ολοκλήρωση της QR ανάλυσης, το τελικό αποτέλεσμα θα είναι ο άνω τριγωνικός πίνακας R που θα αποτελείται από τις m πρώτες γραμμές του A μετά την απαλοιφή, ενώ ο πίνακας Q θα αποτελείται από τις m πρώτες στήλες του ορθογώνιου πίνακα

$$P \;=\; (R_{12}\cdots R_{1n})(R_{23}\cdots R_{2n})\cdots(R_{m,m+1}\cdots R_{mn}).$$

Ο αλγόριθμος της QR ανάλυσης με στροφές Givens δίνεται στη συνέχεια.

Αλγόριθμος QR **Ανάλυσης με Στροφές** Givens:

Δεδομένα: $A \in I\!R^{n,m}$.
Για $k = 1(1)\min\{n-1, m\}$
  Για $j = k+1(1)n$
    $r = \sqrt{a_{kk}^2 + a_{jk}^2}$
    $c = \frac{a_{kk}}{r}$
    $s = -\frac{a_{jk}}{r}$
    $a_{kk} = r$
    Για $i = k+1(1)m$
      $x = ca_{ki} - sa_{ji}$
      $y = sa_{ki} + ca_{ji}$
      $a_{ki} = x$
      $a_{ji} = y$
    Τέλος 'Για'
  Τέλος 'Για'
Τέλος 'Για'

Αν υπολογίσουμε το πλήθος των πράξεων που απαιτούνται για την πραγματοποίηση του αλγόριθμου θα παρατηρήσουμε ότι το κόστος είναι πολύ μεγαλύτερο από εκείνο των ανακλάσεων Householder. Από την άποψη αυτή, είναι ασύμφορη η χρησιμοποίηση του αλγόριθμου. Θα δούμε όμως παρακάτω, ότι η ιδέα των στροφών Givens είναι πολύ χρήσιμη και εφαρμόζεται συχνά στην QR μέθοδο για την εύρεση των ιδιοτιμών ενός πίνακα. Από πλευράς θέσεων μνήμης, οι δύο τελευταίοι αλγόριθμοι είναι ισοδύναμοι. Ο πίνακας R μπορεί να αποθηκευτεί στο άνω τριγωνικό μέρος του A. Για την αποθήκευση των πινάκων στροφών Givens R_{kj}, αρκεί η αποθήκευση των ποσοτήτων c και s κάθε φορά. Δημιουργείται έτσι η εντύπωση ότι δεν επαρκεί το κάτω τριγωνικό μέρος του A αλλά χρειάζεται διπλάσιος χώρος. Επειδή όμως οι αριθμοί c και s αποτελούν το συνημίτονο και το ημίτονο αντίστοιχα της ίδιας γωνίας, ακολουθείται το εξής τέχνασμα: Αν $|s| \leq |c|$ τότε αποθηκεύεται η ποσότητα $p = s \cdot \mathrm{sign}(c)$ αλλιώς η ποσότητα $p = \frac{\mathrm{sign}(s)}{c}$. Κατά την αντίστροφη διαδικασία της αποκρυπτογράφησης, αν $|p| \leq 1$ τότε παίρνεται ως $s = p$ και ως $c = \sqrt{1-s^2}$ αλλιώς παίρνεται ως $c = 1/p$ και ως $s = \sqrt{1-c^2}$. Είναι επίσης φανερό ότι αν (c, s) αποτελεί ένα ζεύγος στροφής Givens, τότε και το $(-c, -s)$ αποτελεί επίσης ζεύγος στροφής Givens.

7.4 Ανάλυση Ιδιαζουσών Τιμών (Singular Value Decomposition-SVD)

Η ανάλυση ιδιαζουσών τιμών ενός πίνακα $A \in I\!R^{n,m}$ είναι ένα από τα σημαντικά κεφάλαια της Γραμμικής Αλγεβρας και έχει πολλές εφαρμογές. Μια από τις εφαρμογές είναι και το γραμμικό πρόβλημα ελάχιστων τετραγώνων. Για το σκοπό αυτό δίνουμε καταρχάς το ακόλουθο βασικό θεώρημα.

Θεώρημα 7.5 *Εστω ο πίνακας $A \in I\!R^{n,m}\backslash\{0_{n,m}\}$ με $m \leq n$. Τότε αυτός μπορεί να γραφτεί υπό τη μορφή*

$$A = U\Sigma V^T,$$

όπου $U \in I\!R^{n,m}$ ορθογώνιος πίνακας ($U^TU = I_m$), $V \in I\!R^{m,m}$ ορθογώνιος πίνακας ($VV^T = I_m$) και $\Sigma \in I\!R^{m,m}$ διαγώνιος τ.ω. $\Sigma = \mathrm{diag}(\sigma_1, \sigma_2, \ldots, \sigma_m)$ με $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_m \geq 0$. (Σημείωση: Οι στήλες $u_1, u_2, \ldots, u_m$ του πίνακα U ονομάζονται αριστερά ιδιάζοντα διανύσματα, οι στήλες $v_1, v_2, \ldots, v_m$ του πίνακα V δεξιά ιδιάζοντα διανύσματα, ενώ τα σ_i ιδιάζουσες τιμές.)

Απόδειξη: Το θεώρημα θα αποδειχτεί με τη μέθοδο της τέλειας επαγωγής ως προς m και n. Για $m = 1$ και για n οποιοδήποτε αριθμό ο A θα είναι ένα n-διάστατο διάνυσμα. Τότε η ανάλυση ιδιαζουσών τιμών πραγματοποιείται αν λάβουμε $U \in I\!R^n$ με $U = \frac{1}{||A||_2}A$, $\Sigma \in I\!R^{1,1}$ με $\Sigma = [||A||_2]$ και $V \in I\!R^{1,1}$ με $V = [1]$.
Υποθέτοντας ότι η ανάλυση ιδιαζουσών τιμών μπορεί να πραγματοποιηθεί για όλους τους πίνακες $A \in I\!R^{n-1,m-1}$ θα αποδείξουμε ότι πραγματοποιείται και για τους $A \in I\!R^{n,m}$. (Σημείωση: Στη συνέχεια δεχόμαστε ότι τουλάχιστον οι φυσικές norms πινάκων μπορούν να επεκταθούν και για μή τετραγωνικούς πίνακες, πράγμα που είναι αληθές.) Εστω, λοιπόν, $A \in I\!R^{n,m}$. Επιλέγουμε ως v_1 το διάνυσμα εκείνο για το οποίο ισχύει $||A||_2 = ||Av_1||_2$. Τέτοιο διάνυσμα υπάρχει, από τον ορισμό της $||A||_2$ ($||A||_2 = \max_{||v||_2=1} ||Av||_2$). Ως u_1 επιλέγουμε το διάνυσμα $u_1 = \frac{Av_1}{||Av_1||_2}$ το οποίο είναι τ.ω. $||u_1||_2 = 1$. Στη συνέχεια συμπληρώνουμε πίνακες $\widetilde{U}$ και $\widetilde{V}$ με τους υποπίνακες $U' \in I\!R^{n,n-1}$ και $V' \in I\!R^{m,m-1}$ ώστε οι $\widetilde{U} = [u_1 \mid U']$ και $\widetilde{V} = [v_1 \mid V']$ να είναι ορθογώνιοι. Εφαρμόζουμε τους πίνακες αυτούς στον A ως εξής:

$$\widetilde{U}^T A \widetilde{V} \;=\; \left[\frac{u_1^T}{U'^T}\right] A[v_1 \mid V'] \;=\; \left[\begin{array}{c|c} u_1^T Av_1 & u_1^T AV' \\ \hline U'^T Av_1 & U'^T AV' \end{array}\right]. \tag{7.40}$$

Απο τη θεώρηση των u_1 και v_1 έχουμε

$$u_1^T Av_1 \;=\; \frac{(Av_1)^T}{||Av_1||_2}(Av_1) \;=\; \frac{||Av_1||_2^2}{||Av_1||_2} \;=\; ||Av_1||_2 \;=\; ||A||_2 \;=\; \sigma_1$$

και

$$U'^T Av_1 \;=\; U'^T u_1 ||Av_1||_2 \;=\; 0$$

λόγω της ορθογωνιότητας του $\widetilde{U}$. Ισχυριζόμαστε ότι και $u_1^T A V' = 0$ γιατί αλλιώς θα είχαμε

$$\begin{aligned} \sigma_1 &= ||A||_2 = ||A\widetilde{V}||_2 = ||u_1^T||_2 ||A\widetilde{V}||_2 \geq ||u_1^T A\widetilde{V}||_2 \\ &= |||[u_1^T A\upsilon_1 | u_1^T A V']||_2 = |||[\sigma_1 | u_1^T A V']||_2 > \sigma_1 \end{aligned} \tag{7.41}$$

που είναι άτοπο. Η σχέση (7.40) μετά την παρατήρηση αυτή γίνεται

$$\widetilde{U}^T A\widetilde{V} = \left[\begin{array}{c|c} \sigma_1 & 0_{m-1}^T \\ \hline 0_{n-1} & U'^T A V' \end{array}\right]. \tag{7.42}$$

Ο πίνακας $A' = U'^T A V'$ είναι τάξης $(n-1)\times(m-1)$ και σύμφωνα με την υπόθεση της τέλειας επαγωγής, πραγματοποιείται γι' αυτόν η ανάλυση ιδιαζουσών τιμών. Εστω

$$U'^T A V' = U_1 \Sigma_1 V_1^T \tag{7.43}$$

όπου $U_1 \in I\!R^{n-1,m-1}$, $V_1 \in I\!R^{m-1,m-1}$ ορθογώνιοι πίνακες και $\Sigma_1 = \mathrm{diag}(\sigma_2, \sigma_3, \ldots, \sigma_m)$. Τότε η (7.42) γίνεται

$$\widetilde{U}^T A\widetilde{V} = \left[\begin{array}{cc} \sigma_1 & 0_{m-1}^T \\ 0_{n-1} & U_1\Sigma_1 V_1^T \end{array}\right] = \left[\begin{array}{cc} 1 & 0_{m-1}^T \\ 0_{n-1} & U_1 \end{array}\right] \left[\begin{array}{cc} \sigma_1 & 0_{m-1}^T \\ 0_{m-1} & \Sigma_1 \end{array}\right] \left[\begin{array}{cc} 1 & 0_{m-1}^T \\ 0_{m-1} & V_1^T \end{array}\right] \tag{7.44}$$

ή

$$A = \widetilde{U} \left[\begin{array}{cc} 1 & 0_{m-1}^T \\ 0_{n-1} & U_1 \end{array}\right] \Sigma \left[\begin{array}{cc} 1 & 0_{m-1}^T \\ 0_{m-1} & V_1^T \end{array}\right] \widetilde{V}^T = U\Sigma V^T,$$

όπου $U \in I\!R^{n,m}$, $V \in I\!R^{m,m}$ ορθογώνιοι πίνακες, αφού είναι γινόμενα ορθογώνιων πινάκων.
Το γεγονός ότι οι ιδιάζουσες τιμές εμφανίζονται κατά φθίνουσα τάξη μεγέθους αποδείχνεται αν δείξουμε ότι $\sigma_1 \geq \sigma_2$. Τότε η απόδειξη για τις υπόλοιπες θα είναι προφανής από την αρχή της τέλειας επαγωγής. Από την (7.43) προκύπτει ότι η σ_2 θα είναι η πρώτη ιδιάζουσα τιμή του πίνακα $U'^T A V'$. Για τον ίδιο ακριβώς λόγο για τον οποίο $\sigma_1 = ||A||_2$, αν επιλέξουμε με τον ίδιο τρόπο τα διανύσματα $u_2 \in I\!R^{n-1}$ και $\upsilon_2 \in I\!R^{m-1}$, θα έχουμε και

$$\sigma_2 = ||A'||_2 = ||U'^T A V'||_2 \leq ||U'^T||_2 ||A||_2 ||V'||_2 = ||A||_2 = \sigma_1.$$

Ετσι ολοκληρώθηκε η απόδειξη του βασικού θεωρήματος της ανάλυσης των ιδιαζουσών τιμών. □

Σημείωση: Η ανάλυση ιδιαζουσών τιμών που δόθηκε στο θεώρημα ισχύει και στην περίπτωση όπου $A \in \mathbb{C}^{n,m}$. Η μόνη διαφορά τότε είναι ότι οι πίνακες U και V είναι ορθομοναδιαίοι ($U^H U = I_m$, $VV^H = I_m$), οπότε κι οι διαφορές στην απόδειξη είναι προφανείς.

Είναι φανερό ότι μπορεί να αλλαχτεί η διάταξη των ιδιαζουσών τιμών σ_i στον πίνακα Σ, αρκεί να γίνουν οι αντίστοιχες αλλαγές στις στήλες των U και V. Στην ουσία ο πίνακας A αναλύεται στο άθροισμα των πινάκων

$$A = \sum_{i=1}^{m} \sigma_i u_i \upsilon_i^T$$

με $\text{rank}([u_i v_i^T]) = 1,\ i = 1(1)m$.

Η ανάλυση ιδιαζουσών τιμών έχει ιδιότητες που την καθιστούν χρησιμότατη σε πολλές εφαρμογές. Συγκεκριμένα, από τον ορισμό της έχουμε ότι

$$AV \ = \ U\Sigma.$$

Από αυτήν προκύπτει εύκολα η εξής γεωμετρική ερμηνεία: Αν θεωρήσουμε ένα ορθογώνιο σύστημα συντεταγμένων του $I\!R^m$ με άξονες τα διανύσματα v_i του V, τότε ο πίνακας A εφαρμοζόμενος στη μοναδιαία σφαίρα του $I\!R^m$ τη μετασχηματίζει σε ένα ελλειψοειδές του $I\!R^n$ με άξονες τα διανύσματα $\sigma_i u_i$.
Παραθέτουμε μια σειρά από χρήσιμες ιδιότητες χωρίς απόδειξη.

i) Η ανάλυση ιδιαζουσών τιμών ενός πίνακα $A \in I\!R^{n,m}$, εκτός από τυχόν μεταθέσεις, είναι μοναδική.
ii) Αν ο πίνακας A είναι τετραγωνικός ($A \in I\!R^{n,n}$) και αντιστρέψιμος, τότε $\sigma_n = ||A^{-1}||_2^{-1}$, ενώ ο δείκτης κατάστασης του A ως προς τη $||\cdot||_2$, θα είναι $\kappa_2(A) = \frac{\sigma_1}{\sigma_n}$.
iii) Αν ο πίνακας A είναι συμμετρικός ($A^T = A$), τότε η ανάλυση ιδιαζουσών τιμών είναι $A = U\Sigma V^T$, ενώ η κανονική μορφή Jordan θα είναι $A = U\Lambda U^T$, όπου $U, V \in I\!R^{n,n}$ ορθογώνιοι πίνακες, $\Sigma = \text{diag}(\sigma_1, \sigma_2, \ldots, \sigma_n)$ και $\Lambda = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$, λ_i οι ιδιοτιμές του A. Επιπλέον ισχύουν οι σχέσεις

$$\sigma_i \ = \ |\lambda_i|,\ v_i \ = \ \text{sign}(\lambda_i) u_i,\ i = 1(1)n,\ \text{με sign}(0) = 1.$$

iv) Αν ο πίνακας A είναι συμμετρικός και θετικά ορισμένος, τότε η ανάλυση ιδιαζουσών τιμών ταυτίζεται με την κανονική μορφή Jordan του A. Η πρόταση αυτή είναι άμεση συνέπεια της προηγούμενης.
v) Αν $A \in I\!R^{n,m},\ m < n$, τότε οι ιδιοτιμές του συμμετρικού πίνακα $A^T A \in I\!R^{m,m}$ θα είναι σ_i^2 με αντίστοιχα ορθοκανονικά ιδιοδιανύσματα τα δεξιά ιδιάζοντα διανύσματα v_i του A.
vi) Αν $A \in I\!R^{n,m},\ m < n$, τότε οι ιδιοτιμές του συμμετρικού πίνακα $AA^T \in I\!R^{n,n}$ θα είναι οι σ_i^2 και $n - m$ το πλήθος μηδενικές, με αντίστοιχα ορθοκανονικά ιδιοδιανύσματα τα αριστερά ιδιάζοντα διανύσματα u_i του A για τις σ_i^2, ενώ ως ιδιοδιανύσματα των μηδενικών ιδιοτιμών μπορούν να ληφτούν οποιαδήποτε διανύσματα του $I\!R^n$ ορθογώνια προς όλα τα $u_i,\ i = 1(1)m$.
vii) Εστω $H = \begin{bmatrix} 0_{n,n} & A^T \\ A & 0_{n,n} \end{bmatrix} \in I\!R^{2n,2n}$, με $A \in I\!R^{n,n}$ τετραγωνικό πίνακα και $A = U\Sigma V^T$ η ανάλυση ιδιαζουσών τιμών του A. Τότε, οι $2n$ ιδιοτιμές του πίνακα H είναι οι $\pm\sigma_i$ με αντίστοιχα ορθοκανονικά ιδιοδιανύσματα τα $\frac{1}{\sqrt{2}} \begin{bmatrix} v_i \\ \pm u_i \end{bmatrix}$.
viii) Εστω $A \in I\!R^{n,m},\ m < n$, και $A = U\Sigma V^T$, η ανάλυση ιδιαζουσών τιμών του A ή ως άθροισμα πινάκων με βαθμό 1, $A = \sum_{i=1}^{m} \sigma_i u_i v_i^T$. Τότε, ο πλησιέστερος προς τον A πίνακας με βαθμό $k < m$, μετρούμενος με τη $||\cdot||_2$, θα είναι ο $A_k = \sum_{i=1}^{k} \sigma_i u_i v_i^T$ ή $A_k = U\Sigma_k V^T$, με $\Sigma_k = \text{diag}(\sigma_1, \sigma_2, \ldots, \sigma_k, 0, \ldots, 0)$, η δε απόστασή του από τον A θα είναι σ_{k+1}. Δηλαδή

$$\min_{B \in I\!R_k^{n,m}} ||A - B||_2 \ = \ ||A - A_k||_2 \ = \ \sigma_{k+1},$$

όπου $I\!R_k^{n,m}$ είναι ο χώρος όλων των $n \times m$ πινάκων βαθμού k.

Η τελευταία ιδιότητα έχει μεγάλη εφαρμογή στην επεξεργασία εικόνας (συμπίεση εικόνας) [11]. Για το σκοπό αυτό διακριτοποιούμε την εικόνα χρησιμοποιώντας ορθογώνιο πλέγμα με πολλά και πολύ μικρά στοιχεία (pics). Σε κάθε στοιχείο, ανάλογα με τη σκιερότητα ή το χρωματισμό του, αντιστοιχίζεται ένας πραγματικός αριθμός. Η διάταξη των αριθμών σύμφωνα με τη διάταξη των στοιχείων στην εικόνα, δημιουργεί έναν πολύ μεγάλο πίνακα A, με συνέπεια την πολύ δαπανηρή απεικόνισή του στον Υπολογιστή. Η ανάλυση των ιδιαζουσών τιμών του πίνακα A, δίνει τη δυνατότητα αναπαραγωγής της εικόνας χρησιμοποιώντας τον πίνακα A_k για αρκετά μικρό k (σε σχέση με τους n και m) χωρίς να υπάρχει μεγάλο σφάλμα στην εικόνα. Για την αποθήκευση απαιτούνται θέσεις μνήμης μόνο για τα ιδιάζοντα διανύσματα $u_1, u_2, \ldots, u_k, \upsilon_1, \upsilon_2, \ldots, \upsilon_k$ και για τις ιδιάζουσες τιμές $\sigma_1, \sigma_2, \ldots, \sigma_k$.
Μια άλλη σημαντική εφαρμογή βρίσκεται στο χώρο της Στατιστικής και συγκεκριμένα στην ανάλυση αντιστοιχιών (Correspondence Analysis-CA) των πινάκων συνάφειας. Ο πίνακας συνάφειας $A \in I\!R^{n,m}$, $n > m$, είναι ο πίνακας συχνοτήτων ενός διδιάστατου δείγματος. Η CA, διαμέσου της ανάλυσης ιδιαζουσών τιμών, αναλύει τον πίνακα A στο άθροισμα $A = \sigma_1 u_1 \upsilon_1^T + \sum_{i=2}^{m} \sigma_i u_i \upsilon_i^T$. Ο πρώτος όρος αντιστοιχεί στην ανεξαρτησία των μεταβλητών, ενώ ο δεύτερος τυποποιεί (μοντελοποιεί) την απόκλιση από αυτή. Στόχος της CA είναι ο προσδιορισμός της μικρότερης δυνατής διάστασης που ερμηνεύει κατά τον καλύτερο τρόπο τη συσχέτιση των μεταβλητών. Δηλαδή την προσέγγιση του $\sum_{i=2}^{m} \sigma_i u_i \upsilon_i^T$ από το $\sum_{i=2}^{k} \sigma_i u_i \upsilon_i^T$, όπου ο k είναι πρακτικά πολύ μικρότερος από τον m.

Οπως φάνηκε από τις ιδιότητες που προαναφέρθηκαν, οι ιδιάζουσες τιμές και τα ιδιάζοντα διανύσματα συνδέονται άμεσα με τις ιδιοτιμές και τα ιδιοδιανύσματα, αντίστοιχα, πινάκων που σχετίζονται με τον A. Για το λόγο αυτό και οι αλγόριθμοι για την ανάλυση ιδιαζουσών τιμών είναι άμεσα συνδεδεμένοι με αλγόριθμους για την εύρεση ιδιοτιμών και ιδιοδιανυσμάτων. Δε θα ασχοληθούμε επομένως στην παράγραφο αυτή με αλγόριθμους της αντίστοιχης ανάλυσης, αλλά στο επόμενο κεφάλαιο όπου θα ασχοληθούμε κυρίως με μεθόδους για την αριθμητική εύρεση ιδιοτιμών και ιδιοδιανυσμάτων.

Απομένει να δοθεί η λύση στο γραμμικό πρόβλημα ελάχιστων τετραγώνων χρησιμοποιώντας την ανάλυση ιδιαζουσών τιμών. Η τεχνική που ακολουθείται για τη λύση του προβλήματος αυτού είναι ακριβώς η ίδια με εκείνη της QR μεθόδου. Το ρόλο του πίνακα $Q \in I\!R^{n,m}$ παίζει ο πίνακας $U \in I\!R^{n,m}$, ενώ το ρόλο του R ο πίνακας ΣV^T. Στην περίπτωση που ο A είναι πλήρους βαθμού, δηλαδή $\mathrm{rank}(A) = m$, θεωρούμε $\widetilde{U} \in I\!R^{n,n-m}$ τον πίνακα που συμπληρώνει τον πίνακα U σε τετραγωνικό ορθογώνιο πίνακα. Ακολουθώντας την ίδια πορεία όπως στις ισότητες (7.20) έχουμε

$$\begin{aligned}
\min_{x \in I\!R^m} ||r||_2^2 \;&\equiv\; \min_{x \in I\!R^m} ||b - Ax||_2^2 = \min_{x \in I\!R^m} ||b - U\Sigma V^T x||_2^2 \\
&= \min_{x \in I\!R^m} ||[U \mid \widetilde{U}]^T (b - U\Sigma V^T x)||_2^2 = \min_{x \in I\!R^m} \left\| \begin{bmatrix} U^T \\ \widetilde{U}^T \end{bmatrix} b - \begin{bmatrix} U^T \\ \widetilde{U}^T \end{bmatrix} U\Sigma V^T x) \right\|_2^2
\end{aligned}$$

$$= \min_{x\in I\!R^m} \left\| \begin{bmatrix} U^Tb - \Sigma V^Tx \\ \widetilde{U}^Tb \end{bmatrix} \right\|_2^2 = \min_{x\in I\!R^m} \left\|U^Tb - \Sigma V^Tx\right\|_2^2 + ||\widetilde{U}^Tb||_2^2 \qquad (7.45)$$

και η λύση x δίνεται από τη

$$x \;=\; V\Sigma^{-1}U^Tb.$$

Στην περίπτωση που ο A είναι ελλιπούς βαθμού, έστω ότι $\mathrm{rank}(A) = r < m$, τότε σύμφωνα με την (7.45), η λύση θα δοθεί ξανά από το μηδενισμό της $U^Tb - \Sigma V^Tx$. Η μόνη διαφορά θα έγκειται στο γεγονός ότι ο Σ θα είναι μή αντιστρέψιμος πίνακας της μορφής

$$\Sigma \;=\; \begin{bmatrix} \Sigma_1 & 0_{r,m-r} \\ 0_{m-r,r} & 0_{m-r,m-r} \end{bmatrix}$$

με $\Sigma_1 \;=\; \mathrm{diag}(\sigma_1, \sigma_2, \ldots, \sigma_r)$. Τότε, διαχωρίζουμε τους πίνακες U και V σε blocks ώστε οι διαστάσεις των blocks να συμφωνούν με τη νέα μορφή του Σ, δηλαδή $U = [U_1 \mid U_2]$, $U_1 \in I\!R^{n,r}$, $U_2 \in I\!R^{n,m-r}$ και $V = [V_1 \mid V_2]$, $V_1 \in I\!R^{m,r}$, $V_2 \in I\!R^{m,m-r}$. Η λύση τότε δίνεται από τη

$$\min_{x\in I\!R^m} ||U^Tb - \Sigma V^Tx||_2^2 = \min_{x\in I\!R^m} \left\| \begin{bmatrix} U_1^T \\ U_2^T \end{bmatrix} b - \begin{bmatrix} \Sigma_1 & 0_{r,m-r} \\ 0_{m-r,r} & 0_{m-r,m-r} \end{bmatrix} \begin{bmatrix} V_1^T \\ V_2^T \end{bmatrix} x \right\|_2^2$$

$$= \min_{x\in I\!R^m} \left\| \begin{bmatrix} U_1^Tb - \Sigma_1 V_1^Tx \\ U_2^Tb \end{bmatrix} \right\|_2^2 = \min_{x\in I\!R^m} ||U_1^Tb - \Sigma_1 V_1^Tx||_2^2 + ||U_2^Tb||_2^2 \;=\; ||U_2^Tb||_2^2. \qquad (7.46)$$

Η ελαχιστοποίηση πετυχαίνεται για

$$x \;=\; V_1\Sigma_1^{-1}U_1^Tb \;+\; V_2y \qquad (7.47)$$

για όλα τα $y \in I\!R^{m-r}$, αφού $V_1^TV_1 = I$ και $V_1^TV_2y = 0$ για όλα τα y. Προφανώς η συσχέτιση των (7.45) και (7.46) δίνει

$$\min_{x\in I\!R^m} ||b - Ax||_2^2 \;=\; ||U_2^Tb||_2^2 \;+\; ||\widetilde{U}^Tb||_2^2.$$

7.5 Ευστάθεια και Κόστος των Μεθόδων για το Γραμμικό Πρόβλημα Ελάχιστων Τετραγώνων

Στην παράγραφο αυτή θα γίνει μια περιγραφή και σύγκριση των μεθόδων που αναπτύχτηκαν σε σχέση με το κόστος και την ευστάθειά τους. Για το κόστος έγινε ήδη αναφορά κατά την παρουσίαση του αλγόριθμου της κάθε μεθόδου. Εδώ θα γίνει προσπάθεια να δοθεί ένα είδος σύγκρισης του κόστους σε σχέση με την ευστάθεια.
Πριν προχωρήσουμε θα δώσουμε τους ορισμούς του ψευδοαντίστροφου πίνακα κατά Moore-Penrose και του δείκτη κατάστασης για $n \times m$ πίνακες.

Ορισμός 7.2 *Εστω ότι ο πίνακας* $A \in I\!R^{n,m}$, $m \leq n$, *είναι πλήρους βαθμού, με* $A = QR = U\Sigma V^T$ *την* QR *ανάλυση και την ανάλυση ιδιαζουσών τιμών του* A, *αντίστοιχα. Τότε ο πίνακας*

$$A^\dagger = (A^T A)^{-1} A^T = R^{-1} Q^T = V \Sigma^{-1} U^T$$

καλείται ψευδοαντίστροφος κατά Moore-Penrose *του πίνακα* A.

Είναι φανερό ότι ο $A^\dagger$ είναι αριστερός αντίστροφος του A αφού $A^\dagger A = (A^T A)^{-1} A^T A = I$. Ακόμη, η λύση του γραμμικού προβλήματος ελάχιστων τετραγώνων θα δίνεται υπό τη μορφή

$$x = A^\dagger b. \tag{7.48}$$

Θα πρέπει να παρατηρήσουμε ότι αν $m = n$ τότε ο $A^\dagger$ συμπίπτει με τον A^{-1}. Στην περίπτωση που $n < m$, ο ψευδοαντίστροφος ορίζεται ως $A^\dagger = A^T (A A^T)^{-1}$ και είναι δεξιός αντίστροφος.
Με τον ορισμό αυτόν η (7.48) δίνει τη λύση μόνο στην περίπτωση που ο A είναι πλήρους βαθμού. Θα πρέπει να επεκτείνουμε τον ορισμό και για τους πίνακες βαθμού $r < m$ ($\leq n$), ώστε η (7.48) να δίνει μία από τις λύσεις του προβλήματος. Οι λύσεις του προβλήματος με ανάλυση ιδιαζουσών τιμών δίνονται από την (7.47). Αν σ΄ αυτή θέσουμε $y = 0$ τότε παίρνουμε τη λύση

$$x \;=\; V_1 \Sigma_1^{-1} U_1^T b. \tag{7.49}$$

Η λύση αυτή είναι η λύση με τη μικρότερη Ευκλείδεια norm. Πράγματι, κάθε άλλη λύση θα έχει $\|x\|_2^2 \;=\; \|V_1 \Sigma_1^{-1} U_1^T b\|_2^2 \;+\; \|V_2 y\|_2^2$ λόγω του Πυθαγορείου θεωρήματος αφού τα διανύσματα $V_1 \Sigma_1^{-1} U_1^T b$ και $V_2 y$ είναι ορθογώνια. Δίνουμε, λοιπόν, τον ορισμό του ψευδοαντίστροφου ώστε η λύση (7.49) να συμπίπτει με την (7.48).

Ορισμός 7.3 *Εστω ότι ο πίνακας* $A \in I\!R^{n,m}$, $m \leq n$, *είναι ελλιπούς βαθμού, με* $\mathrm{rank}(A) = r < m$ *και* $A = U\Sigma V^T = U_1 \Sigma_1 V_1^T$ *η ανάλυση ιδιαζουσών τιμών του* A, *όπως ορίστηκε στην προηγούμενη παράγραφο. Τότε ο πίνακας*

$$A^\dagger = V_1 \Sigma_1^{-1} U_1^T \tag{7.50}$$

καλείται ψευδοαντίστροφος κατά Moore-Penrose *του πίνακα* A.

Η σχέση (7.50) γράφεται και ως

$$A^\dagger = V \Sigma^\dagger U^T, \quad \Sigma^\dagger = \begin{bmatrix} \Sigma_1 & 0_{r,m-r} \\ 0_{m-r,r} & 0_{m-r,m-r} \end{bmatrix}^\dagger \;=\; \begin{bmatrix} \Sigma_1^{-1} & 0_{r,m-r} \\ 0_{m-r,r} & 0_{m-r,m-r} \end{bmatrix}.$$

Εύκολα μπορεί να παρατηρήσει κανείς ότι ο ορισμός αυτός καλύπτει πλήρως και τον προηγούμενο ορισμό.

Στη συνέχεια δίνουμε την επέκταση του ορισμού του δείκτη κατάστασης για $n \times m$ πίνακες.

Ορισμός 7.4 *Ο δείκτης κατάστασης ενός πίνακα $A \in I\!R^{n,m}$, $m \leq n$, ορίζεται ως*

$$\kappa_2(A) = \frac{\sigma_{\max}}{\sigma_{\min}},$$

όπου $\sigma_{\max}$ και $\sigma_{\min}$, η μεγαλύτερη και η μικρότερη (θετική) ιδιάζουσα τιμή του A αντίστοιχα.

Από τη ιδιότητα (ii) των ιδιαζουσών τιμών της προηγούμενης παραγράφου φαίνεται καθαρά ότι ο ορισμός αυτός καλύπτει πλήρως και τον ορισμό του δείκτη κατάστασης για τετραγωνικούς πίνακες. Επίσης, χρησιμοποιώντας την ιδιότητα (v) των ιδιαζουσών τιμών, αποδείχνεται η ισχύς της σχέσης (7.17) για το δείκτη κατάστασης του συστήματος των κανονικών εξισώσεων. Πράγματι

$$\begin{array}{rcl} \kappa_2(A^TA) & = & ||A^TA||_2||(A^TA)^{-1}||_2 \;=\; \rho(A^TA)\rho\left((A^TA)^{-1}\right) \\ & = & \frac{\sigma_{\max}^2}{\sigma_{\min}^2} \;=\; (\kappa_2(A))^2\,. \end{array} \tag{7.51}$$

Πρέπει να παρατηρήσουμε εδώ ότι αν ο A είναι ελλιπούς βαθμού τότε ο A^TA είναι μή αντιστρέψιμος πίνακας και $\sigma_{\min} = 0$, οπότε θεωρείται ότι ο δείκτης κατάστασης είναι ίσος με ∞.

Μετά από τους τρεις αυτούς απαραίτητους ορισμούς μπορούμε να μελετήσουμε την κατάσταση του γραμμικού προβλήματος ελάχιστων τετραγώνων. Μπορεί να αναπτυχτεί μια θεωρία διατάραξης αντίστοιχη εκείνης των γραμμικών συστημάτων του Κεφαλαίου 2. Δεν θα αναπτύξουμε εδώ τη θεωρία αυτή διότι αφενός δε διαφέρει στη φιλοσοφία της από εκείνη του Κεφαλαίου 2 και αφετέρου γίνεται τεχνικά πιο δύσκολη. Θα παραθέσουμε απλώς το αποτέλεσμά της στο παρακάτω θεώρημα.

Θεώρημα 7.6 *Υποθέτουμε ότι $A \in I\!R^{n,m}$, $m \leq n$, είναι ελλιπούς βαθμού και ότι το διάνυσμα x ελαχιστοποιεί τη $||Ax - b||_2$. Αν δA είναι η διατάραξη του A, δb η διατάραξη του b και $\epsilon \equiv \max(\frac{||\delta A||_2}{||A||_2}, \frac{||\delta b||_2}{||b||_2}) < \frac{1}{\kappa_2(A)} = \frac{\sigma_{\min}}{\sigma_{\max}}$, τότε η σχετική απόλυτη διατάραξη της λύσης δίνεται από τη σχέση*

$$\frac{||\delta x||_2}{||x||_2} = \epsilon\left(\frac{2\kappa_2(A)}{\cos\theta} + \tan\theta\kappa_2^2(A)\right) + \mathcal{O}(\epsilon^2), \tag{7.52}$$

όπου $\sin\theta \;=\; \frac{||Ax-b||_2}{||b||_2}$.

Από την (7.52) φαίνεται ότι η κατάσταση του προβλήματος εξαρτιέται από την ποσότητα $\frac{2\kappa_2(A)}{\cos\theta} + \tan\theta\kappa_2^2(A)$, που ονομάζεται δείκτης κατάστασης του προβλήματος ελάχιστων τετραγώνων και συμβολίζεται με κ_{LS}. Παρατηρούμε ότι ο κ_{LS} εξαρτιέται όχι μόνο από τον $\kappa_2(A)$ αλλά και από τη γωνία που σχηματίζουν τα διανύσματα Ax και b. Αν τα διανύσματα αυτά είναι συνευθειακά, τότε εξαρτιέται μόνο από τον $\kappa_2(A)$, ο οποίος αν είναι σχετικά μικρός προσδίδει καλή κατάσταση στο πρόβλημα (well-conditioned). Αν δεν είναι συνευθειακά, τότε εξαρτιέται από τον $\kappa_2^2(A)$ και το πρόβλημα χάνει την καλή κατάστασή του (ill-conditioned) όσο μεγαλώνει η γωνία, με ακραία την

περίπτωση όπου $\kappa_{LS} = \infty$ όταν $\theta = \frac{\pi}{2}$. Στην τελευταία περίπτωση έχουμε τη λύση $x = 0$.

Προχωρούμε στη μελέτη της ευστάθειας και τη σύγκρισή της σε σχέση και με το κόστος των μεθόδων που αναπτύχθηκαν. Αρχικά μελετάμε το πρόβλημα όταν ο πίνακας είναι πλήρους βαθμού.

Ως προς το κόστος, οικονομικότερη μέθοδος είναι η επίλυση του συστήματος των κανονικών εξισώσεων, ακολουθεί η QR ανάλυση, ενώ η ανάλυση ιδιαζουσών τιμών είναι η πιο δαπανηρή. Τα πράγματα όμως αλλάζουν όσον αφορά την ευστάθεια των μεθόδων. Η επίλυση των κανονικών εξισώσεων είναι η πιο ασταθής μέθοδος αφού, όπως φαίνεται από τις (7.17) και (7.51), έχουμε τετραγωνική αύξηση του δείκτη κατάστασης. Αυτό σημαίνει ότι χάνεται σε ακρίβεια διπλάσιος αριθμός σημαντικών ψηφίων, από εκείνα που χάνονται από τις άλλες δύο μεθόδους. Συνιστάται να εφαρμόζεται μόνο σε προβλήματα με καλή κατάσταση επειδή το όφελος σε πλήθος πράξεων για την επίλυση είναι μεγάλο. Τα περισσότερα όμως προβλήματα που προέρχονται από διάφορες εφαρμογές δεν έχουν καλή κατάσταση. Για τη λύση αυτών συνιστάται η QR ανάλυση. Πρέπει όμως να γίνει σύγκριση των επί μέρους αλγόριθμων της QR ανάλυσης.
Εχει αναφερθεί ότι και έχει εξηγηθεί γιατί ο κλασικός αλγόριθμος της Gram-Schmidt ορθογωνιοποίησης είναι ασταθής. Ο παραλλαγμένος αλγόριθμος είναι αρκετά πιο ευσταθής, αλλά κι αυτός δε δίνει καλά αποτελέσματα ειδικά όταν ο πίνακας A είναι "σχεδόν" ελλιπούς βαθμού. Αυτό οφείλεται στο γεγονός ότι ο αλγόριθμος βρίσκει ακολουθιακά μία προς μία τις στήλες του ορθογώνιου πίνακα Q, με συνέπεια τα σφάλματα που εισχώρησαν σε προηγούμενες στήλες, να μεταβιβάζονται και να συσσωρεύονται στις επόμενες με αποτέλεσμα ο πίνακας Q να απέχει αρκετά από ορθογώνιο πίνακα. Δε συνιστάται επομένως ούτε η παραλλαγμένη μέθοδος Gram-Schmidt, η οποία είναι πολύ χρήσιμη σε άλλα ειδικά προβλήματα ορθογωνιοποίησης, όπως η εύρεση των ιδιοδιανυσμάτων τριδιαγώνιων και συμμετρικών πινάκων. Σε αντίθεση προς αυτήν, οι αλγόριθμοι Householder και Givens είναι πιο ενδεδειγμένοι γιατί σ' αυτούς εκτελούνται μόνο διαδοχικοί πολλαπλασιασμοί από αριστερά με ορθογώνιους πίνακες των οποίων το γινόμενο είναι ένας ορθογώνιος πίνακας που έχει την ίδια τάξη σφάλματος με εκείνη των επί μέρους παραγόντων. Αυτό φαίνεται καθαρά από την ακόλουθη ανάλυση σφαλμάτων.
Εστω ότι επιθυμούμε να βρούμε το γινόμενο XA και ότι έχουμε στη διάθεσή μας την προσεγγιστική τιμή $fl(X)$ του X με την ακρίβεια του Υπολογιστή ϵ. Στην πράξη δε θα βρεθεί το γινόμενο $fl(X)A$ αλλά η στρογγύλευσή του στο $fl(fl(X)A)$. Υπενθυμίζεται ότι η συνάρτηση $fl(\cdot)$ που προσεγγίζει έναν αριθμό x στον πλησιέστερο αριθμό του Υπολογιστή, δίνεται ως $fl(x) = x(1+\delta)$ με $|\delta| \leq \epsilon$ ή $\delta = \mathcal{O}(\epsilon)$. Για την παράσταση ενός πίνακα στον Υπολογιστή ισχύει κάτι αντίστοιχο, συγκεκριμένα

$$fl(X) = X(I + \Delta),\ ||\Delta||_2 = \mathcal{O}(\epsilon).$$

Εφαρμόζοντας τη σχέση αυτή στο παραπάνω γινόμενο παίρνουμε

$$\begin{aligned} fl\left(fl(X)A\right) &= fl\left(X(I+\Delta_1)A\right) = X(I+\Delta_1)A(I+\Delta_2) \\ &= X(A + \Delta_1 A + A\Delta_2 + \Delta_1 A\Delta_2) = X(A + \delta A). \end{aligned} \tag{7.53}$$

Από την (7.53) διαπιστώνεται ότι ο πολλαπλασιασμός αυτός θα δώσει το ίδιο αποτέλεσμα με εκείνον του πολλαπλασιασμού του ακριβούς πίνακα X με το διαταραγμένο πίνακα $A + \delta A$. Υπολογίζοντας τη $||\delta A||_2$, έχουμε

$$\begin{aligned} ||\delta A||_2 &= ||\Delta_1 A + A\Delta_2 + \Delta_1 A \Delta_2||_2 \\ &\leq ||\Delta_1||_2||A||_2 + ||\Delta_2||_2||A||_2 + ||\Delta_1||_2||\Delta_2||_2||A||_2 = \mathcal{O}(\epsilon)||A||_2. \end{aligned} \tag{7.54}$$

Η συνολική διατάραξη θα είναι $X\delta A$ και επομένως το συνολικό σφάλμα είναι

$$||E||_2 = ||X\delta A||_2 \leq ||X||_2||\delta A||_2 = \mathcal{O}(\epsilon)||X||_2||A||_2.$$

Αν πολλαπλασιάσουμε διαδοχικά από αριστερά με τους πίνακες $X_1, X_2, \ldots, X_k$, τότε θα πάρουμε φράγμα για το ολικό σφάλμα

$$||E_{\text{ολ}}||_2 = \mathcal{O}(\epsilon)||X_1||_2||X_2||_2 \cdots ||X_k||_2||A||_2.$$

Στις μεθόδους Householder και Givens οι πίνακες X_i, $i = 1(1)k$, είναι ορθογώνιοι και κατά συνέπεια οι φασματικές norms τους μονάδα. Επομένως τα σφάλματα στις δύο αυτές μεθόδους θα είναι

$$||E_H||_2 = \mathcal{O}(\epsilon)||A||_2, \quad ||E_G||_2 = \mathcal{O}(\epsilon)||A||_2,$$

αντίστοιχα. Η ίδια θεωρία μας δίνει και το ολικό σφάλμα του ορθογώνιου πίνακα Q, αρκεί στη θέση του A να θέσουμε το μοναδιαίο πίνακα I, έτσι

$$||E_Q||_2 = \mathcal{O}(\epsilon).$$

Επιπλέον με την QR μέθοδο εξασφαλίζεται το αμετάβλητο του δείκτη κατάστασης του πίνακα A, αφού αυτός πολλαπλασιάζεται με μια πολύ καλή προσέγγιση του ορθογώνιου πίνακα Q^T. Ισχύει δε και για την επέκταση του ορισμού του δείκτη κατάστασης ότι

$$\kappa_2(R) = \kappa_2(Q^T A) = \kappa_2(A)$$

επειδή οι $Q^T A$ και A έχουν τις ίδιες ακριβώς ιδιάζουσες τιμές.
Ακριβώς για τις ιδιότητές τους αυτές, όπως φαίνεται κι από την παραπάνω ανάλυση της θεωρίας σφαλμάτων, επινοήθηκαν και χρησιμοποιήθηκαν οι ορθογώνιοι μετασχηματισμοί για τη λύση του γραμμικού προβλήματος ελάχιστων τετραγώνων και όχι κάποιοι άλλοι, όπως π.χ. οι τριγωνικοί.
Ως προς την ευστάθεια οι αλγόριθμοι Householder και Givens αποδείχτηκαν ισοδύναμοι. Επειδή δε το κόστος του αλγόριθμου Householder είναι μικρότερο, αυτός είναι εκείνος που συνιστάται για την επίλυση του προβλήματος με QR ανάλυση.

ΑΣΚΗΣΕΙΣ

1.: Να λυθεί το γραμμικό πρόβλημα των ελάχιστων τετραγώνων $\min_{x\in\mathbb{R}^3} ||b - Ax||_2$, όπου $A = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 0 & -1 \\ 0 & 1 & 1 \\ 0 & 1 & -1 \end{bmatrix}$ και $b = [1\ 1\ 1\ 1]^T$.
α) Με τη λύση του συστήματος των κανονικών εξισώσεων.
β) Με την QR ανάλυση χρησιμοποιώντας τον αλγόριθμο ορθογωνιοποίησης των Gram-Schmidt. και
γ) Τέλος, να βρεθεί η τιμή $\min_{x\in\mathbb{R}^3} ||b - Ax||_2$.
(**Περιορισμός**: Να γίνουν ακριβείς πράξεις διατηρώντας ριζικά και κλάσματα στους υπολογισμούς.)

2.: Να λυθεί το γραμμικό σύστημα $Ax = b$, με $A = \begin{bmatrix} 2 & -1 & 0 & -1 \\ 1 & 2 & -1 & 0 \\ 0 & 1 & 2 & -1 \\ 1 & 0 & 1 & 2 \end{bmatrix}$ και $b = [4\ -\ 2\ 2\ 0]^T$, με τη μέθοδο της QR ανάλυσης χρησιμοποιώντας τον αλγόριθμο ορθογωνιοποίησης των Gram-Schmidt (**Περιορισμός**: Να γίνουν ακριβείς πράξεις διατηρώντας ριζικά και κλάσματα στους υπολογισμούς.)

3.: Να λυθεί το γραμμικό σύστημα $Ax = b$, με $A = \begin{bmatrix} 3 & 2 & 1 \\ 2 & 2 & 1 \\ 1 & 1 & 1 \end{bmatrix}$ και $b = [2\ 1\ 1]^T$, με την QR ανάλυση χρησιμοποιώντας τον αλγόριθμο ορθογωνιοποίησης των Gram-Schmidt. (**Περιορισμός**: Να γίνουν ακριβείς πράξεις διατηρώντας ριζικά και κλάσματα στους υπολογισμούς.)

4.:
α) Να αποδειχτεί ότι η λύση των κανονικών εξισώσεων για το πρόβλημα ελάχιστων τετραγώνων συμπίπτει με τη λύση που δίνει η μέθοδος QR παραγοντοποίησης, για την περίπτωση που ο συντελεστής πίνακας A είναι πλήρους βαθμού. και
β) Να λυθεί το γραμμικό σύστημα ελάχιστων τετραγώνων, με $A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}$ και $b = [1\ 1\ 1\ 1]^T$, με την QR ανάλυση χρησιμοποιώντας τον αλγόριθμο ορθογωνιοποίησης των Gram-Schmidt. (**Περιορισμός**: Να γίνουν ακριβείς πράξεις διατηρώντας ριζικά και κλάσματα στους υπολογισμούς.)

5.: Να λυθεί το γραμμικό πρόβλημα ελάχιστων τετραγώνων $\min_{x\in\mathbb{R}^3} ||b - Ax||_2$, όπου $A =$

$\begin{bmatrix} 3 & 0 & 4 \\ 4 & 0 & 3 \\ 0 & 4 & 3 \\ 0 & 3 & 4 \end{bmatrix}$ και $b = [1\ 1\ 1\ 1]^T$, χρησιμοποιώντας τον αλγόριθμο της QR με Gram-Schmidt ορθογωνιοποίηση και να βρεθεί η τιμή $\min_{x\in\mathbb{R}^3} ||b - Ax||_2$. (**Περιορισμός**: Να γίνουν ακριβείς πράξεις διατηρώντας ριζικά και κλάσματα στους υπολογισμούς.)

6.: Να λυθεί το γραμμικό πρόβλημα ελάχιστων τετραγώνων $\min_{x\in\mathbb{R}^3} ||b - Ax||_2$, όπου $A = \begin{bmatrix} 1 & 3 & 4 \\ 1 & 3 & 2 \\ 1 & 1 & 0 \\ -1 & -1 & 2 \end{bmatrix}$ και $b = [1\ 1\ 1\ 1]^T$, χρησιμοποιώντας τον αλγόριθμο της QR με Gram-Schmidt ορθογωνιοποίηση και να βρεθεί η τιμή $\min_{x\in\mathbb{R}^3} ||b - Ax||_2$. (**Περιορισμός**: Να γίνουν ακριβείς πράξεις διατηρώντας ριζικά και κλάσματα στους υπολογισμούς.)

7.: Δίνεται το γραμμικό σύστημα $Ax = b$, όπου

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \quad \kappa\alpha\iota \quad b = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}.$$

α) Να σχηματιστεί το σύστημα των κανονικών εξισώσεων και να λυθεί με τη μέθοδο του Cholesky. και
β) Να επαληθευτεί ότι το διάνυσμα $r = Ax^* - b$, όπου x^* η λύση που παίρνεται παραπάνω, είναι ορθογώνιο προς κάθε ένα από τα διανύσματα-στήλες του A.

8.: Δίνεται το γραμμικό σύστημα $Ax = b$, όπου

$$A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & 1 \\ 3 & -1 & 3 \\ 2 & 0 & 2 \end{bmatrix} \quad \kappa\alpha\iota \quad b = \begin{bmatrix} 3 \\ 1 \\ 5 \\ 6 \end{bmatrix}.$$

α) Να βρεθεί $x \in \mathbb{R}^3$ τέτοιο ώστε η ποσότητα $||b - Ax||_2$ να γίνεται ελάχιστη. και
β) Να βρεθεί η ελάχιστη τιμή της παραπάνω ποσότητας.

8 Αριθμητικές Μέθοδοι για τον Υπολογισμό Ιδιοτιμών και Ιδιοδιανυσμάτων

8.1 Εισαγωγή

Στο παρόν κεφάλαιο θα ασχοληθούμε με αριθμητικές μεθόδους για τον υπολογισμό των ιδιοτιμών $\lambda_i, i = 1(1)n$, και των ιδιοδιανυσμάτων $x_i \in \mathbb{C}^n \setminus \{0\},\ i = 1(1)n$, του πίνακα $A \in \mathbb{C}^{n,n} (A \in \mathbb{R}^{n,n})$. Οπως είδαμε σε προηγούμενα κεφάλαια, η γνώση του φάσματος των ιδιοτιμών του πίνακα των συντελεστών των αγνώστων ενός γραμμικού συστήματος έχει μεγάλη σημασία στην επιλογή κατάλληλης μεθόδου επίλυσής του. Εκτός όμως από την παραπάνω χρησιμότητά τους, οι ιδιοτιμές αποτελούν σημαντικές οντότητες σε μια πληθώρα προβλημάτων στις Θετικές Επιστήμες και την Τεχνολογία. Αναφέρουμε εδώ ενδεικτικά μερικά παραδείγματα: Στη Θεωρία Πληροφοριών οι ιδιοτιμές αποτελούν μέτρα πληροφορίας. Στη Στατιστική εμφανίζονται στην πολυδιάστατη ανάλυση και στις Μαρκοβιανές στοχαστικές διαδικασίες. Στη Μηχανική, κατά τη μελέτη των ελεύθερων ταλαντώσεων παραμορφώσιμων σωμάτων (πλάκες, κελύφη), προκύπτουν πίνακες που οι ιδιοτιμές τους είναι τετράγωνα των ιδιοσυχνοτήτων των ταλαντώσεων. Στη Θεωρία Στερεών Σωμάτων, οι ιδιοτιμές είναι ιδιοσυχνότητες των ταλαντώσεων. Στην Πυρηνική Φυσική είναι ιδιοενέργειες των σωματιδίων του πυρήνα (νουκλεόνια). Στην Ατομική Φυσική είναι ιδιοενέργειες των ηλεκτρονίων. Ακόμη, οι ιδιοτιμές εμφανίζονται στην Κβαντική Χημεία και σε πολλές άλλες επιστήμες. Πριν προχωρήσουμε σε συγκεκριμένες μεθόδους για την εύρεση των ιδιοτιμών και των ιδιοδιανυσμάτων θα δώσουμε μερικούς ορισμούς, κάποιοι από τους οποίους είναι ήδη γνωστοί, απαραίτητους για την ανάπτυξη της βασικής θεωρίας.

8.2 Βασική Θεωρία

Ορισμός 8.1 *Αν $A \in \mathbb{C}^{n,n}$, $x \in \mathbb{C}^n \setminus \{0\}$ και $\lambda \in \mathbb{C}$ πληρούν τη σχέση $Ax = \lambda x$ τότε η λ είναι ιδιοτιμή του πίνακα A και το x αντίστοιχο (δεξιό) ιδιοδιάνυσμα. Το διάνυσμα $y \in \mathbb{C}^n \setminus \{0\}$ για το οποίο ισχύει η σχέση $y^H A = \lambda y^H$, θα λέγεται αριστερό ιδιοδιάνυσμα.*

Είναι φανερό ότι για να υπάρξει μή μηδενική λύση x στο σύστημα $Ax = \lambda x$, θα πρέπει ο πίνακας $(A - \lambda I)$ να είναι μή αντιστρέψιμος. Οι ιδιοτιμές λ θα δίνονται ως ρίζες του πολυωνύμου $P(\lambda) = \det(A - \lambda I)$. Το πολυώνυμο αυτό λέγεται χαρακτηριστικό πολυώνυμο και είναι βαθμού ακριβώς n. Ακριβώς n το πλήθος θα είναι και οι ιδιοτιμές του A, κάποιες από τις οποίες μπορεί να είναι και πολλαπλές. Είναι προφανές ότι αν $A \in \mathbb{C}^{n,n}$ τότε, γενικά, $\lambda \in \mathbb{C}$ ενώ αν $A \in \mathbb{R}^{n,n}$ τότε οι ιδιοτιμές θα είναι πραγματικές ή ζεύγη συζυγών μιγαδικών αριθμών. Ενώ οι ιδιοτιμές είναι ακριβώς n το πλήθος, τούτο δεν ισχύει και για τα ιδιοδιανύσματα. Υπάρχουν πίνακες που δεν έχουν βάση n γραμμικά ανεξάρτητων ιδιοδιανυσμάτων. Αυτό φαίνεται από την κανονική μορφή Jordan

του πίνακα A που έχει δοθεί στο Θεώρημα 3.2. Σε κάθε block J_i της μορφής Jordan αντιστοιχεί μόνο ένα ιδιοδιάνυσμα, επομένως αν p είναι τα blocks, p θα είναι και τα γραμμικά ανεξάρτητα ιδιοδιανύσματα του A. Πρέπει να παρατηρήσουμε εδώ ότι p θα είναι και τα γραμμικά ανεξάρτητα αριστερά ιδιοδιανύσματα. (Σημείωση: Τα ιδιοδιανύσματα αντιστοιχούν στην πρώτη θέση του κάθε block ενώ τα αριστερά ιδιοδιανύσματα στην τελευταία.) Η εύρεση της κανονικής μορφής Jordan θα έλυνε συγχρόνως και το πρόβλημα της εύρεσης των ιδιοτιμών και των ιδιοδιανυσμάτων. Τούτο όμως είναι δαπανηρό από πλευράς τόσο του κόστους όσο και της ευστάθειας. Μια άλλη κανονική μορφή που οφείλεται σε ορθογώνιο μετασχηματισμό ομοιότητας (θα τον λέμε ορθομοναδιαίο, από τη λεξη unitary, στην περίπτωση μιγαδικών πινάκων), είναι η κανονική μορφή Schur. Λόγω της χρήσης ορθογώνιων μετασχηματισμών, εξασφαλίζεται η ευστάθεια σε ικανοποιητικό βαθμό, ενώ όπως θα δούμε παρακάτω ότι έχουν επινοηθεί αποτελεσματικοί αλγόριθμοι για την υλοποίησή της. Δίνουμε την κανονική μορφή Schur με το θεώρημα που ακολουθεί.

Θεώρημα 8.1 *(Κανονική μορφή* Schur*) Δοθέντος ενός πίνακα $A \in \mathbb{C}^{n,n}$, υπάρχει ένας ορθομοναδιαίος πίνακας $Q \in \mathbb{C}^{n,n}$ και ένας άνω τριγωνικός πίνακας $T \in \mathbb{C}^{n,n}$ τ.ω.*

$$Q^H A Q \;=\; T. \tag{8.1}$$

Προφανώς οι ιδιοτιμές του A θα είναι τα διαγώνια στοιχεία του T.

Απόδειξη: Η απόδειξη θα γίνει επαγωγικά. Είναι προφανές ότι η (8.1) ισχύει για 1×1 πίνακες. Υποθέτουμε ότι ο ισχυρισμός του θεωρήματος αληθεύει για όλους τους $(n-1) \times (n-1)$ πίνακες και θα αποδείξουμε την ισχύ του για τους $n \times n$ πίνακες. Εστω $A \in \mathbb{C}^{n,n}$, $\lambda \in \mathbb{C}$ μια ιδιοτιμή του A και $u \in \mathbb{C}^n \setminus \{0\}$ το αντίστοιχο ιδιοδιάνυσμα, κανονικοποιημένο ($\|u\|_2 = 1$). Συμπληρώνουμε το χώρο θεωρώντας τον πίνακα $\widetilde{U} \in \mathbb{C}^{n,n-1}$ με στήλες ορθομοναδιαία διανύσματα με το διάνυσμα u και μεταξύ τους. Ετσι ο πίνακας $U = [u|\widetilde{U}]$, είναι ένας τετραγωνικός ορθομοναδιαίος πίνακας. Κατασκευάζουμε το γινόμενο

$$U^H A U \;=\; \left[\frac{u^H}{\widetilde{U}^H}\right] A[u|\widetilde{U}] \;=\; \left[\begin{array}{c|c} u^H A u & u^H A\widetilde{U} \\ \hline \widetilde{U}^H A u & \widetilde{U}^H A\widetilde{U} \end{array}\right] \;=\; \left[\begin{array}{cc} \lambda & u^H A\widetilde{U} \\ \lambda\widetilde{U}^H u & \widetilde{U}^H A\widetilde{U} \end{array}\right] \tag{8.2}$$

αλλά $\widetilde{U}^H u = 0$ επειδή ο πίνακας U κατασκευάστηκε έτσι ώστε να είναι ορθομοναδιαίος. Η (8.2) επομένως γίνεται

$$U^H A U \;=\; \left[\begin{array}{cc} \lambda & u^H A\widetilde{U} \\ 0 & \widetilde{U}^H A\widetilde{U} \end{array}\right]. \tag{8.3}$$

Συμβολίζουμε με $\widetilde{A}$ τον $(n-1) \times (n-1)$ πίνακα $\widetilde{U}^H A\widetilde{U}$. Από την υπόθεση της τέλειας επαγωγής, υπάρχει ορθομοναδιαίος πίνακας $P \in \mathbb{C}^{n-1,n-1}$ που δίνει την κανονική μορφή Schur. Συγκεκριμένα,

$$P^H \widetilde{A} P \;=\; \widetilde{T} \;\Leftrightarrow\; \widetilde{A} \;=\; P\widetilde{T}P^H.$$

Αντικαθιστώντας στην (8.3) έχουμε

$$\begin{aligned} U^H A U &= \begin{bmatrix} \lambda & u^H A\widetilde{U} \\ 0 & \widetilde{A} \end{bmatrix} = \begin{bmatrix} \lambda & u^H A\widetilde{U} \\ 0 & P\widetilde{T}P^H \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 \\ 0 & P \end{bmatrix} \begin{bmatrix} \lambda & u^H A\widetilde{U}P \\ 0 & \widetilde{T} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & P^H \end{bmatrix}. \end{aligned} \tag{8.4}$$

Είναι προφανές ότι ο $\begin{bmatrix} 1 & 0 \\ 0 & P \end{bmatrix}$ είναι ορθομοναδιαίος $n \times n$ πίνακας, καθώς και το γινόμενο $Q = U \begin{bmatrix} 1 & 0 \\ 0 & P \end{bmatrix}$ ως γινόμενο ορθομοναδιαίων πινάκων. Ετσι η (8.4) γίνεται

$$Q^H A Q = \begin{bmatrix} 1 & 0 \\ 0 & P^H \end{bmatrix} U^H A U \begin{bmatrix} 1 & 0 \\ 0 & P \end{bmatrix} = \begin{bmatrix} \lambda & u^H A\widetilde{U}P \\ 0 & \widetilde{T} \end{bmatrix} = T \tag{8.5}$$

και η απόδειξη ολοκληρώθηκε. □

Σημειώσεις: α) Πρέπει να παρατηρήσουμε ότι αν ο A είναι Ερμιτιανός πίνακας, τότε

$$u^H A\widetilde{U}P = (Au)^H \widetilde{U}P = \bar{\lambda} u^H \widetilde{U}P = 0,$$

αφού ο U είναι ορθομοναδιαίος. β) Στην περίπτωση (α), αν ακολουθήσουμε την πορεία της απόδειξης του θεωρήματος προκύπτει ότι ο πίνακας T θα είναι διαγώνιος, οπότε η κανονική μορφή Schur συμπίπτει με την κανονική μορφή Jordan. Ετσι αποδείχνεται και η γνωστή πρόταση: Κάθε Ερμιτιανός πίνακας έχει n γραμμικά ανεξάρτητα ιδιοδιανύσματα που αποτελούν ορθομοναδιαία βάση. Στην περίπτωση των πραγματικών συμμετρικών πινάκων, κάθε πραγματικός συμμετρικός πίνακας έχει n γραμμικά ανεξάρτητα ιδιοδιανύσματα που αποτελούν ορθογώνια βάση. γ) Από την απόδειξη του θεωρήματος φαίνεται ακόμη ότι η κανονική μορφή Schur δεν είναι μονοσήμαντα ορισμένη, αλλά εξαρτιέται από τη σειρά που λαμβάνονται οι ιδιοτιμές στο διαγώνιο μέρος του άνω τριγωνικού πίνακα.

Το θεώρημα της κανονικής μορφής Schur ισχύει για κάθε μιγαδικό. πίνακα. Η υλοποίησή του επομένως από κάποιον αλγόριθμο, προϋποθέτει αριθμητική μιγαδικών αριθμών. Ακόμη και στην περίπτωση πραγματικού πίνακα, εκ πρώτης όψεως, φαίνεται ότι δεν μπορούμε να αποφύγουμε την αριθμητική μιγαδικών αριθμών, επειδή κάποιες ιδιοτιμές μπορεί να είναι ζεύγη συζυγών μιγαδικών αριθμών και τα αντίστοιχα ιδιοδιανύσματα συζυγή μιγαδικά. Μπορούμε όμως να εξειδικεύσουμε την κανονική μορφή Schur για πραγματικούς πίνακες, παραλλάσσοντάς την ώστε να απαιτείται αριθμητική πραγματικών αριθμών μόνο. Την πραγματική αυτή κανονική μορφή Schur δίνουμε με το ακόλουθο θεώρημα.

Θεώρημα 8.2 *(Πραγματική κανονική μορφή* Schur*) Δοθέντος ενός πίνακα $A \in \mathbb{R}^{n,n}$, υπάρχει πραγματικός ορθογώνιος πίνακας $Q \in \mathbb{R}^{n,n}$ και πραγματικός "σχεδόν" άνω τριγωνικός πίνακας*

$T \in \mathbb{R}^{n,n}$ τ.ω. $Q^T A Q = T$. Το "σχεδόν" άνω τριγωνικός σημαίνει ότι ο T θα είναι block *άνω τριγωνικός με διαγώνια* blocks 1×1, *που θα αντιστοιχούν στις πραγματικές ιδιοτιμές ή 2×2, που θα αντιστοιχούν σε ζεύγη μιγαδικών ιδιοτιμών.*

Απόδειξη: Αν θεωρήσουμε ότι η ιδιοτιμή λ είναι πραγματική, τότε ακολουθούμε την ίδια ακριβώς πορεία απόδειξης όπως στο προηγούμενο θεώρημα με τέλεια επαγωγή. Εστω όμως ότι η λ είναι μια μιγαδική ιδιοτιμή του A με αντίστοιχο (μιγαδικό) ιδιοδιάνυσμα u. Τότε και η $\bar{\lambda}$ θα είναι ιδιοτιμή του A με αντίστοιχο (μιγαδικό) ιδιοδιάνυσμα $\bar{u}$. Τα συζυγή αυτά μιγαδικά ιδιοδιανύσματα παράγουν ένα διδιάστατο υπόχωρο αποτελούμενο από όλα τα διανύσματα, γραμμικούς συνδυασμούς των u και $\bar{u}$, το $\mathrm{span}\{u, \bar{u}\}$. Θεωρούμε τώρα τα διανύσματα u_R, το πραγματικό μέρος του u, και u_I, το φανταστικό μέρος του u. Είναι προφανές ότι τα διανύσματα αυτά παράγουν τον ίδιο υπόχωρο ($\mathrm{span}\{u_R, u_I\} = \mathrm{span}\{u, \bar{u}\}$). Εστω ότι $P = [u_R \mid u_I]$ και $P = QR$, η QR παραγοντοποίηση του P ($P, Q \in \mathbb{R}^{n,2},\ R \in \mathbb{R}^{2,2}$). Τα διανύσματα στήλες του P θα είναι γραμμικοί συνδυασμοί των στηλών του Q. Ο 2×2 πίνακας R αποτελείται από τους συντελεστές των γραμμικών συνδυασμών των στηλών του P ως προς τις στήλες του Q. Επομένως

$$\mathrm{span}\{Q\} = \mathrm{span}\{P\} = \mathrm{span}\{u, \bar{u}\}. \tag{8.6}$$

Από την (8.6) προκύπτει ότι οι στήλες του Q αποτελούν γραμμικούς συνδυασμούς των u και $\bar{u}$. Οι στήλες τότε του AQ θα είναι γραμμικοί συνδυασμοί των Au και $A\bar{u}$ και άρα των u και $\bar{u}$, αφού αυτά είναι ιδιοδιανύσματα του A. Αυτό σημαίνει ότι

$$\mathrm{span}\{AQ\} = \mathrm{span}\{u, \bar{u}\} = \mathrm{span}\{Q\},$$

που οδηγεί στο συμπέρασμα ότι θα υπάρχει ένας πίνακας $B \in \mathbb{R}^{2,2}$ τ.ω.

$$AQ = QB.$$

Ας σημειωθεί ότι ο πίνακας B αποτελείται από τους συντελεστές των γραμμικών συνδυασμών που δίνουν τις στήλες του AQ σα συναρτήσεις των στηλών του Q. Θεωρούμε τον πίνακα $\widetilde{Q} \in \mathbb{R}^{n,n-2}$, που συμπληρώνει τον ορθογώνιο πίνακα Q, στον τετραγωνικό ορθογώνιο πίνακα U. Ακολουθούμε τώρα αντίστοιχη πορεία με την απόδειξη του προηγουμένου θεωρήματος με τη μέθοδο της τέλειας επαγωγής. Εστω ότι ο ισχυρισμός του θεωρήματος αληθεύει για 2×2 πίνακες. Υποθέτουμε ότι αληθεύει για $(n-2) \times (n-2)$ και θα αποδείξουμε την ισχύ του για $n \times n$ πραγματικούς πίνακες. Θεωρούμε το γινόμενο

$$U^T A U = \left[\begin{array}{c} Q^T \\ \hline \widetilde{Q}^H \end{array}\right] A [Q | \widetilde{Q}] = \left[\begin{array}{cc} Q^T A Q & Q^T A \widetilde{Q} \\ \widetilde{Q}^T A Q & \widetilde{Q}^T A \widetilde{Q} \end{array}\right] = \left[\begin{array}{cc} B & Q^T A \widetilde{Q} \\ 0 & \widetilde{Q}^T A \widetilde{Q} \end{array}\right], \tag{8.7}$$

αφού $Q^T A Q = Q^T Q B = B$ και $\widetilde{Q}^T A Q = \widetilde{Q}^T Q B = 0$. Ακολουθεί στη συνέχεια η ίδια ακριβώς επαγωγική διαδικασία του προηγούμενου θεωρήματος, και η απόδειξη ολοκληρώνεται. □

Στην περίπτωση που έχει επιτευχτεί η εύρεση της κανονικής μορφής Schur, απομένει η εύρεση των ιδιοδιανυσμάτων που αντιστοιχούν στην κάθε μια ιδιοτιμή. Αν $y^{(i)}$ είναι το ιδιοδιάνυσμα του πίνακα T που αντιστοιχεί στην ιδιοτιμή λ_i, τότε το ιδιοδιάνυσμα του A θα είναι

$$x^{(i)} \;=\; Qy^{(i)}.$$

Το $y^{(i)}$ βρίσκεται από τη λύση του συστήματος $(\lambda_i I_n - T)y^{(i)} = 0$ ως εξής. Αν η ιδιοτιμή λ_i βρίσκεται στην i-οστή στη σειρά θέση της διαγωνίου του T, τότε το προς λύση σύστημα, σε block μορφή, θα είναι

$$\left[\begin{array}{c|c|c} \lambda_i I_{i-1} - T_{11} & t_2 & T_{13} \\ \hline & 0 & t_3^T \\ \hline & & \lambda_i I_{n-i} - T_{33} \end{array}\right] \left[\begin{array}{c} y_1^{(i)} \\ \hline y_2^{(i)} \\ \hline y_3^{(i)} \end{array}\right] \;=\; 0, \tag{8.8}$$

όπου έγινε ένας block διαχωρισμός του T της i-οστής γραμμής και i-οστής στήλης. Είναι φανερό, αν υποθέσουμε ότι η λ_i είναι απλή ιδιοτιμή (εύκολα γενικεύεται όταν είναι πολλαπλή), ότι οι πίνακες $\lambda_i I_{i-1} - T_{11}$ και $\lambda_i I_{n-i} - T_{33}$ είναι αντιστρέψιμοι. Ακολουθώντας τη διαδικασία της επίλυσης του συστήματος με προς τα πίσω αντικατάσταση έχουμε $y_3^{(i)} = 0$, το σύστημα αληθεύει για οποιαδήποτε τιμή του $y_2^{(i)}$ οπότε θέτοντας, π.χ. $y_2^{(i)} = 1$, το $y_1^{(i)}$ βρίσκεται με προς τα πίσω αντικατάσταση από το σύστημα $(\lambda_i I_{i-1} - T_{11})y_1^{(i)} = -t_2$.

Αρχικά οι ερευνητές, για τον υπολογισμό των ιδιοτιμών και των αντίστοιχων ιδιοδιανυσμάτων, έθεσαν ως στόχο την εύρεση του χαρακτηριστικού πολυωνύμου P_n του πίνακα A, η εύρεση των ριζών του οποίου δίνει συγχρόνως και τις ιδιοτιμές του A. Η μέθοδος Krylov είναι μια τέτοια μέθοδος. Βασίζεται στην ακολουθία διανυσμάτων $x^{(0)}, x^{(1)} = Ax^{(0)}, x^{(2)} = Ax^{(1)} = A^2x^{(0)}, \cdots, x^{(n-1)} = Ax^{(n-2)} = \cdots = A^{n-1}x^{(0)}$, που ορίζει έναν υπόχωρο Krylov με $x^{(0)} \in I\!R^n \setminus \{0\}$ αυθαίρετο αρχικό διάνυσμα. Θεωρούμε τον πίνακα $K \in I\!R^{n,n}$ ως τον πίνακα αποτελούμενο από τα πρώτα n διανύσματα που ορίζουν τον υπόχωρο Krylov

$$K \;=\; \left[\; x^{(0)} \quad x^{(1)} \quad x^{(2)} \quad \cdots \quad x^{(n-1)} \;\right].$$

Θεωρούμε στη συνέχεια το γινόμενο

$$AK = \left[\; Ax^{(0)} \quad Ax^{(1)} \quad Ax^{(2)} \quad \cdots \quad Ax^{(n-1)} \;\right] = \left[\; x^{(1)} \quad x^{(2)} \quad \cdots \quad x^{(n-1)} \quad A^n x^{(0)} \;\right].$$

Παρατηρούμε ότι ο πίνακας AK προέκυψε από τον K με μετατόπιση όλων των στηλών του κατά μία θέση αριστερά, με εξαφάνιση του $x^{(0)}$ από την πρώτη θέση και εμφάνιση του $A^n x^{(0)}$ στην τελευταία, που είναι ο επόμενος όρος της ακολουθίας Krylov. Αυτό, σε μορφή πινάκων, σημαίνει ότι

$$AK \;=\; K\left[\; e^2 \quad e^3 \quad \cdots \quad e^n \quad -a \;\right] \;=\; KF,$$

όπου $e^i, i = 2(1)n$, είναι η i-οστή στήλη του μοναδιαίου πίνακα και $a = [a_0\ a_1\ a_2\ \cdots\ a_{n-1}]^T = -K^{-1}A^n x^{(0)}$. Ετσι κατασκευάστηκε ο πίνακας

$$F \ = \ K^{-1}AK \ = \ \begin{bmatrix} 0 & 0 & \cdots & 0 & -a_0 \\ 1 & 0 & \cdots & 0 & -a_1 \\ 0 & 1 & \ddots & \vdots & \vdots \\ \vdots & \vdots & \ddots & 0 & -a_{n-2} \\ 0 & 0 & \cdots & 1 & -a_{n-1} \end{bmatrix}.$$

Ο πίνακας αυτός είναι γνωστός ως ο "συνοδεύων" πίνακας Frobenius και είναι όμοιος προς τον A. Επομένως έχει το ίδιο χαρακτηριστικό πολυώνυμο και εύκολα αποδείχνεται ότι αυτό είναι το

$$P_n(x) \ = \ x^n + a_{n-1}x^{n-1} + a_{n-2}x^{n-2} + \cdots + a_1 x + a_0.$$

Ο αλγόριθμος επομένως για την εύρεση του P_n συνίσταται στην εύρεση του πίνακα K, στην εύρεση του επόμενου όρου $x^{(n)} = Ax^{(n-1)} = A^n x^{(0)}$ και στη συνέχεια στη λύση του συστήματος

$$Ka \ = \ -x^{(n)}. \tag{8.9}$$

Ο αλγόριθμος αυτός δεν είναι αποτελεσματικός ούτε από την άποψη του κόστους ούτε από αυτήν της ευστάθειας. Το κόστος αφορά σε n πολλαπλασιασμούς πίνακα επί διάνυσμα και στη λύση ενός συστήματος. Επιπλέον επιβαρύνεται και με το κόστος της εύρεσης των ριζών του P_n, που αποτελεί ένα εντελώς διαφορετικό πρόβλημα. Αν το P_n δεν είναι κάποιας συγκεκριμένης μορφής, τότε το πρόβλημα αυτό είναι αρκετά πελύπλοκο με μεγάλο κόστος.
Οσον αφορά στην ευστάθεια, το πρόβλημα της επίλυσης του συστήματος (8.9) δεν έχει καλή κατάσταση (ill-conditioned). Αυτό οφείλεται στο γεγονός ότι η ακολουθία των διανυσμάτων που ορίζουν τον υπόχωρο Krylov, όπως θα δούμε παρακάτω στη μέθοδο δυνάμεων, τείνει σε ένα ιδιοδιάνυσμα. Ετσι, όσο προχωρούμε προς τα δεξιά, οι στήλες του πίνακα K τείνουν να γίνουν παράλληλες.

Δε συνιστάται επομένως η μέθοδος αυτή, εκτός από προβλήματα με πολύ μικρή διάσταση. Στην περίπτωση των μεγάλων αραιών πινάκων, δε συνιστάται για έναν ακόμη λόγο. Διότι ο πίνακας K γίνεται πυκνός.

Για τους παραπάνω λόγους η έρευνα στράφηκε προς την κατεύθυνση του υπολογισμού των ιδιοτιμών και ιδιοδιανυσμάτων, απευθείας με τεχνικές της Γραμμικής Αλγεβρας, που οδήγησε σε μια σειρά από επαναληπτικές κυρίως μεθόδους. Η μέθοδος Krylov θεωρείται ως άμεση μέθοδος, αφού υποτίθεται ότι υπολογίζει τις ακριβείς τιμές των ιδιοτιμών αν εφαρμοστεί με ακριβή αριθμητική.

Σε κάποιες μεθόδους απαιτείται ο εντοπισμός των ιδιοτιμών οι οποίες βρίσκονται σε κάποιο χωρίο του μιγαδικού επιπέδου. Ενας τέτοιος εντοπισμός δίνεται από τη γνωστή σχέση $\rho(A) \leq \|A\|$ για κάθε φυσική norm. Αν επομένως γνωρίζουμε μία norm του A έχουμε ως συμπέρασμα ότι όλες

οι ιδιοτιμές ανήκουν στο δίσκο του μιγαδικού επιπέδου που ορίζεται ως $\{z \in \mathbb{C} : |z| \leq \|A\|\}$. Ο εντοπισμός αυτός είναι αρκετά χοντρικός ένας πιο εκλεπτισμένος δίνεται από το ακόλουθο Θεώρημα του Gerschgorin.

Θεώρημα 8.3 *(*Gerschgorin*) Αν $A \in \mathbb{C}^{n,n}$, τότε οι ιδιοτιμές του $\lambda_i, i = 1(1)n$, ανήκουν στην ένωση των δίσκων του μιγαδικού επιπέδου*

$$\bigcup_{i=1}^{n}\{z \in \mathbb{C} : |z - a_{ii}| \leq \sum_{j=1, j\neq i}^{n} |a_{ij}|\}. \tag{8.10}$$

Απόδειξη: Εστω $\lambda \in \mathbb{C}$ ιδιοτιμή του A και x το αντίστοιχο ιδιοδιάνυσμα, κανονικοποιημένο ως προς τη $\|\cdot\|_\infty$ ($\|x\|_\infty = 1$). Εστω επίσης ότι $|x_i| = \|x\|_\infty = 1$, τότε η σχέση $Ax = \lambda x$ δίνει

$$\begin{aligned}
\sum_{j=1}^{n} a_{ij}x_j = \lambda x_i \quad &\Leftrightarrow \quad (\lambda - a_{ii})x_i \;=\; \sum_{j=1, j\neq i}^{n} a_{ij}x_j \\
&\Rightarrow \quad |\lambda - a_{ii}||x_i| \;=\; |\sum_{j=1, j\neq i}^{n} a_{ij}x_j| \;\leq\; \sum_{j=1, j\neq i}^{n} |a_{ij}||x_j| \qquad (8.11)\\
&\Rightarrow \quad |\lambda - a_{ii}| \;\leq\; \sum_{j=1, j\neq i}^{n} |a_{ij}|.
\end{aligned}$$

Η τελευταία σχέση θα ισχύει για όλες τις ιδιοτιμές του A, επομένως αυτές θα ανήκουν στην ένωση των δίσκων (8.10). □

Σημειώνεται ότι, επειδή οι ιδιοτιμές του A είναι και ιδιοτιμές του A^T, το Θεώρημα του Gerschgorin ισχύει και για τον A^T. Επομένως οι ιδιοτιμές του θα ανήκουν και στην ένωση των δίσκων

$$\bigcup_{i=1}^{n}\{z \in \mathbb{C} : |z - a_{ii}| \leq \sum_{j=1, j\neq i}^{n} |a_{ji}|\}.$$

Μετά τη βασική θεωρία και τις εισαγωγικές παρατηρήσεις, που προηγήθηκαν, είμαστε σε θέση να προχωρήσουμε στη μελέτη επαναληπτικών μεθόδων για την αριθμητική εύρεση ιδιοτιμών και ιδιοδιανυσμάτων, ξεκινώντας από την πιο απλή και βασική, τη μέθοδο δυνάμεων.

8.3 Μέθοδος Δυνάμεων

Η μέθοδος δυνάμεων είναι μία από τις πιο απλές επαναληπτικές μεθόδους και χρησιμεύει για τον υπολογισμό μιας μόνο ιδιοτιμής, της απόλυτα μεγαλύτερης.

Υποθέτουμε ότι $A \in I\!R^{n,n}$, και ότι οι ιδιοτιμές του $\lambda_i, i = 1(1)n$, ικανοποιούν τις σχέσεις

$$|\lambda_1| \ > \ |\lambda_2| \ \geq \ |\lambda_3| \ \geq \ \cdots \ \geq \ |\lambda_n|,$$

δηλαδή ότι η απόλυτα μεγαλύτερη ιδιοτιμή είναι απλή, πραγματική και ότι υπάρχει βάση των n γραμμικά ανεξάρτητων αντίστοιχων ιδιοδιανυσμάτων $x_i, i = 1(1)n$. Θεωρούμε επίσης μια norm και επιλέγουμε ένα αρχικό διάνυσμα $x^{(0)} \in I\!R^n \setminus \{0\}$ κανονικοποιημένο ώστε $\|x^{(0)}\| = 1$. Στη συνέχεια κατασκευάζουμε την κανονικοποιημένη ακολουθία $x^{(k)} \in I\!R^n \setminus \{0\}, k = 1, 2, 3, \cdots$, των δυνάμεων του A επί το διάνυσμα $x^{(0)}$

$$x^{(k)} \ = \ \frac{A^k x^{(0)}}{\|A^k x^{(0)}\|}. \tag{8.12}$$

Αποδείχνεται ότι η ακολουθία αυτή έχει όριο το ιδιοδιάνυσμα x_1, που αντιστοιχεί στην απόλυτα μεγαλύτερη ιδιοτιμή λ_1. Πραγματικά, θεωρούμε ότι το $x^{(0)}$ δίνεται ως γραμμικός συνδυασμός των ιδιοδιανυσμάτων ως εξής:

$$x^{(0)} \ = \sum_{i=1}^{n} c_i x_i, \ c_i \in I\!R, i = 1(1)n, \ \mu\epsilon \ c_1 \neq 0. \tag{8.13}$$

Τότε η (8.12) γίνεται

$$\begin{aligned} x^{(k)} \ &= \ \frac{A^k x^{(0)}}{\|A^k x^{(0)}\|} \ = \ \frac{\sum_{i=1}^{n} c_i A^k x_i}{\|\sum_{i=1}^{n} c_i A^k x_i\|} \\ &= \ \frac{\sum_{i=1}^{n} c_i \lambda_i^k x_i}{\|\sum_{i=1}^{n} c_i \lambda_i^k x_i\|} \ = \ \frac{\lambda_1^k \sum_{i=1}^{n} c_i \left(\frac{\lambda_i}{\lambda_1}\right)^k x_i}{|\lambda_1|^k \|\sum_{i=1}^{n} c_i \left(\frac{\lambda_i}{\lambda_1}\right)^k x_i\|} \end{aligned} \tag{8.14}$$

και

$$\begin{aligned} \lim_{k\to\infty} x^{(k)} \ &= \ \lim_{k\to\infty} \frac{\lambda_1^k}{|\lambda_1|^k} \frac{c_1 x_1 + \sum_{i=2}^{n} c_i \left(\frac{\lambda_i}{\lambda_1}\right)^k x_i}{\|c_1 x_1 + \sum_{i=2}^{n} c_i \left(\frac{\lambda_i}{\lambda_1}\right)^k x_i\|} \\ &= \ \lim_{k\to\infty} \frac{\lambda_1^k}{|\lambda_1|^k} \frac{c_1 x_1 + \sum_{i=2}^{n} c_i \lim_{k\to\infty} \left(\frac{\lambda_i}{\lambda_1}\right)^k x_i}{\|c_1 x_1 + \sum_{i=2}^{n} c_i \lim_{k\to\infty} \left(\frac{\lambda_i}{\lambda_1}\right)^k x_i\|} \\ &= \ \lim_{k\to\infty} \frac{\lambda_1^k}{|\lambda_1|^k} \frac{c_1 x_1}{\|c_1 x_1\|} \ = \ \lim_{k\to\infty} \frac{c_1 \lambda_1^k}{|c_1 \lambda_1^k|} \cdot \frac{x_1}{\|x_1\|}, \end{aligned} \tag{8.15}$$

αλλά το πηλίκο $\frac{c_1 \lambda_1^k}{|c_1 \lambda_1^k|}$ παίρνει τιμές 1 ή -1 και επειδή το ιδιοδιάνυσμα x_1 είναι ανεξάρτητο προσήμου και σταθερού παράγοντα, έχουμε ότι η ακολουθία $x^{(k)}$ συγκλίνει στο ιδιοδιάνυσμα x_1. Αφού η $x^{(k)}$

συγκλίνει στο x_1, η ιδιοτιμή λ_1 θα είναι το όριο του πηλίκου Rayleigh

$$\lambda_1 \;=\; \lim_{k\to\infty} \frac{x^{(k)^T} A x^{(k)}}{x^{(k)^T} x^{(k)}}.$$

Πράγματι,

$$\lim_{k\to\infty} \frac{x^{(k)^T} A x^{(k)}}{x^{(k)^T} x^{(k)}} \;=\; \frac{\lim_{k\to\infty} x^{(k)^T} A \lim_{k\to\infty} x^{(k)}}{\lim_{k\to\infty} x^{(k)^T} \lim_{k\to\infty} x^{(k)}} \;=\; \frac{x_1^T A x_1}{x_1^T x_1} \;=\; \lambda_1.$$

Ο αλγόριθμος επομένως της μεθόδου δυνάμεων, παράγει επαναληπτικά την ακολουθία $x^{(k)}$, $k = 1, 2, \cdots$, και θεωρεί, στην k επανάληψη, το διάνυσμα $x^{(k)}$ ως την προσέγγιση του x_1 και το πηλίκο Rayleigh $\lambda_1^{(k)} = \frac{x^{(k)^T} A x^{(k)}}{x^{(k)^T} x^{(k)}}$ ως την προσέγγιση της ιδιοτιμής λ_1.
Στην πράξη, δεν υπολογίζεται το γινόμενο $A^k x^{(0)}$ για μεγάλο k και μετά να γίνει η κανονικοποίηση. Αν γινόταν αυτό θα ήταν επικίνδυνο, αν $\lambda_1 > 1$, να αυξηθούν τα μέτρα των συνιστωσών των διανυσμάτων, ώστε να βρεθούν έξω από τη χωρητικότητα της μνήμης του Υπολογιστή (overflow) ή αν $\lambda_1 < 1$, να μηδενιστεί η ακολουθία διανυσμάτων (underflow). Εκείνο που γίνεται είναι να κανονικοποιείται σε κάθε βήμα το διάνυσμα

$$y^{(k)} \;=\; A x^{(k-1)},$$

που θα παράγει την ίδια ακολουθία $x^{(k)}$. Πραγματικά, η απόδειξη γίνεται εύκολα με τη μέθοδο της τέλειας επαγωγής. Ο ισχυρισμός αληθεύει για $k = 1$ αφού

$$\frac{y^{(1)}}{\|y^{(1)}\|} \;=\; \frac{A x^{(0)}}{\|A x^{(0)}\|} \;=\; x^{(1)}.$$

Υποθέτουμε ότι αληθεύει για $k = m$, δηλαδή

$$\frac{y^{(m)}}{\|y^{(m)}\|} \;=\; x^{(m)} \;=\; \frac{A^m x^{(0)}}{\|A^m x^{(0)}\|}$$

και αποδείχνουμε στη συνέχεια την αλήθεια του για $k = m + 1$

$$\frac{y^{(m+1)}}{\|y^{(m+1)}\|} \;=\; \frac{A x^{(m)}}{\|A x^{(m)}\|} \;=\; \frac{A \frac{A^m x^{(0)}}{\|A^m x^{(0)}\|}}{\|A \frac{A^m x^{(0)}}{\|A^m x^{(0)}\|}\|} \;=\; \frac{A^{m+1} x^{(0)}}{\|A^{m+1} x^{(0)}\|} \;=\; x^{(m+1)}.$$

Στη συνέχεια δίνουμε τον αλγόριθμο της μεθόδου δυνάμεων. Ως norm επιλέγουμε την Ευκλείδεια norm, όπου το πηλίκο Rayleigh γίνεται $x^{(k)^T} A x^{(k)}$, αφού $x^{(k)^T} x^{(k)} = 1$.

Αλγόριθμος Μεθόδου Δυνάμεων:

Δεδομένα: $A \in I\!R^{n,n}$.
Επιλογή $x^{(0)} \in I\!R^n$, ώστε $\|x^{(0)}\|_2 = 1$
Για $k = 1$ έως ότου υπάρξει σύγκλιση
$y^{(k)} = Ax^{(k-1)}$
$\lambda^{(k-1)} = x^{(k-1)^T} y^{(k)}$
$x^{(k)} = \frac{y^{(k)}}{\|y^{(k)}\|_2}$
Τέλος 'Για'
Αποτέλεσμα: $\lambda^{(k-1)}$ είναι η προσέγγιση της λ_1 και $x^{(k)}$ η προσέγγιση του x_1.

Παρατηρούμε ότι το κόστος του αλγορίθμου, ανά επανάληψη, είναι ένας πολλαπλασιασμός πίνακα επί διάνυσμα, ο υπολογισμός ενός εσωτερικού γινομένου, ο υπολογισμός της norm διανύσματος και η διαίρεση διανύσματος δια αριθμού. Ο πολλαπλασιασμός πίνακα επί διάνυσμα χαρακτηρίζει το κόστος που είναι $\mathcal{O}(n^2)$. Αν m είναι το πλήθος των επαναλήψεων, τότε το συνολικό κόστος ανέρχεται σε $\mathcal{O}(mn^2)$. Ως κριτήριο τερματισμού των επαναλήψεων λαμβάνεται η προσέγγιση του σχετικού απόλυτου σφάλματος της ιδιοτιμής: $\frac{|\lambda^{(k-1)} - \lambda^{(k)}|}{|\lambda^{(k)}|} \leq \epsilon$, ή η προσέγγιση του απόλυτου σφάλματος του ιδιοδιανύσματος $\|x^{(k)} - x^{(k+1)}\| \leq \epsilon$, όπου ϵ είναι "μικρός" αριθμός επιλεγόμενος από το χρήστη. Το πρώτο κριτήριο προτιμάται ως το λιγότερο δαπανηρό. Το δεύτερο κριτήριο είναι συγχρόνως απόλυτο και σχετικό απόλυτο σφάλμα αφού τα ιδιοδιανύσματα λαμβάνονται κανονικοποιημένα.
Σ΄ ό,τι αφορά την ταχύτητα σύγκλισης της μεθόδου έχουμε, από τις διαδοχικές ισότητες στις σχέσεις (8.16), ότι το σφάλμα αποκοπής οφείλεται στο άθροισμα $\sum_{j=2}^{n} c_i \left(\frac{\lambda_i}{\lambda_1}\right)^k x_i$. Ο μεγαλύτερος απόλυτα όρος του αθροίσματος, από κάποιο k και πέρα, θα είναι ο $c_2 \left(\frac{\lambda_2}{\lambda_1}\right)^k x_2$. Αυτός είναι και ο όρος που χαρακτηρίζει την τάξη σύγκλισης. Μπορούμε επομένως να πούμε ότι το σφάλμα είναι της τάξης $\mathcal{O}\left(|\frac{\lambda_2}{\lambda_1}|^k\right)$. Η σύγκλιση θα είναι ταχύτερη όταν το κλάσμα $|\frac{\lambda_2}{\lambda_1}|$ είναι όσο το δυνατόν μικρότερο.

Στη συνέχεια δίνουμε μια σειρά από παρατηρήσεις για τη συμπεριφορά της μεθόδου δυνάμεων σε ορισμένες ειδικές περιπτώσεις.

i) Στο γραμμικό συνδυασμό (8.13) υποθέτουμε ότι $c_1 \neq 0$. Το διάνυσμα όμως $x^{(0)}$ το επιλέγουμε αυθαίρετα χωρίς να γνωρίζουμε το ιδιοδιάνυσμα x_1, επομένως δε γνωρίζουμε αν το x_1 υπεισέρχεται στο γραμμικό συνδυασμό. Στην περίπτωση οπου το $x^{(0)}$ δίνεται ως γραμμικός συνδυασμός άλλων ιδιοδιανυσμάτων εκτός του x_1, δηλαδή $c_1 = 0$, τότε η θεωρία που αναπτύχθηκε για τη μέθοδο δυνάμεων, ισχύει για το υποσύνολο των ιδιοτιμών των οποίων τα αντίστοιχα ιδιοδιανύσματα υπεισέρχονται στο γραμμικό συνδυασμό. Δηλαδή η μέθοδος προσεγγίζει την απόλυτα μεγαλύτερη ιδιοτιμή του υποσυνόλου, εφόσον αυτή είναι απλή και δεν υπάρχει άλλη με το ίδιο μέτρο. Στην πράξη όμως συμβαίνει κάτι άλλο. Ενώ στο ανάπτυγμα του $x^{(0)}$ δεν υπάρχει το x_1, κατά τη δι-

αδικασία εφαρμογής της μεθόδου δημιουργούνται σφάλματα στρογγύλευσης από τον Υπολογιστή, με συνέπεια να θεωρείται μια προσέγγιση $x^{(0)^*}$ ως αρχικό διάνυσμα, στο ανάπτυγμα του οποίου υπεισέρχεται το x_1 με πολύ μικρό απόλυτα συντελεστή $c_1^* \neq 0$. Με την έναρξη των επαναλήψεων η μέθοδος κατευθύνει τη σύγκλιση προς την απόλυτα μεγαλύτερη ιδιοτιμή του υποσυνόλου. Από κάποια επανάληψη όμως και μετά ο όρος $c_1^* x_1$ θα υπερισχύσει του $c_2 \left(\frac{\lambda_2}{\lambda_1}\right)^k x_2$, αφού $|\lambda_1| > |\lambda_2|$. Τότε δημιουργείται μία "αναταραχή" στη μέθοδο η οποία αλλάζει την κατεύθυνση της σύγκλισης και συγκλίνει στη λ_1 και στο ιδιοδιάνυσμα x_1.
ii) Στην περίπτωση όπου η λ_1 είναι πολλαπλότητας $n_1 > 1$ και υπάρχουν n_1 γραμμικά ανεξάρτητα ιδιοδιανύσματα που αντιστοιχούν σ΄ αυτήν, τότε ο γραμμικός συνδυασμός (8.13) γράφεται ως

$$x^{(0)} = \sum_{i=1}^{n_1} \alpha_i x_i + \sum_{i=n_1+1}^{n} c_i x_i, \ \alpha_i, c_i \in I\!R, \ \sum_{i=1}^{n_1} \alpha_i^2 \neq 0,$$

όπου $x_i, i = 1(1)n_1$, είναι τα ιδιοδιανύσματα της λ_1. Το διάνυσμα όμως $x = \sum_{i=1}^{n_1} \alpha_i x_i$ θεωρείται ως ιδιοδιάνυσμα του A αντίστοιχο της λ_1. Τότε, η όλη αναπτυχθείσα θεωρία ισχύει και η μέθοδος των δυνάμεων θα συγκλίνει στην πολλαπλή ιδιοτιμή λ_1 και στο ιδιοδιάνυσμα x.
iii) Στην περίπτωση όπου η λ_1 είναι πολλαπλότητας $n_1 > 1$ αλλά δεν υπάρχουν n_1 γραμμικά ανεξάρτητα ιδιοδιανύσματα που να αντιστοιχούν σ΄ αυτήν, τότε η παραπάνω θεωρία δεν ισχύει. Μια παραπλήσια θεωρία αναπτύσσεται κάνοντας ευρεία χρήση της κανονικής μορφής Jordan και των δυνάμεων αυτής. Η θεωρία αυτή αποδείχνει ότι και τότε η μέθοδος συγκλίνει στην ιδιοτιμή λ_1 και στο ιδιοδιάνυσμα που αντιστοιχεί στο μεγαλύτερης τάξης block της λ_1 στην κανονική μορφή Jordan. Η σύγκλιση όμως είναι βραδεία και το σφάλμα αποκοπής είναι $\mathcal{O}\left(\frac{1}{k}\right)$ αντί $\mathcal{O}\left(|\frac{\lambda_2}{\lambda_1}|^k\right)$, που είναι στην ομαλή περίπτωση.
iv) Οταν η λ_1 αντιστοιχεί σε ζεύγος συζυγών μιγαδικών ιδιοτιμών ή σε ζεύγος αντίθετων ιδιοτιμών, η μέθοδος δυνάμεων δεν συγκλίνει.
v) Από τον αλγόριθμο της μεθόδου παρατηρούμε ότι αφού το διάνυσμα $x^{(k-1)}$ τείνει προς το x_1, τότε και τό $y^{(k)} = Ax^{(k-1)}$ θα τείνει προς το $\lambda_1 x^{(k-1)}$. Επομένως η ιδιοτιμή λ_1 μπορεί να προσεγγιστεί και από το πηλίκο των συνιστωσών $\frac{y_i^{(k)}}{x_i^{(k-1)}}$, αρκεί $x_i^{(k-1)} \neq 0$. Αυτό εφαρμόζεται κυρίως όταν χρησιμοποιούμε ως norm τη $\|\cdot\|_\infty$. Η προσέγγιση της λ_1 τότε θα είναι η $\lambda^{(k-1)} = \frac{y_i^{(k)}}{x_i^{(k-1)}}$, όπου i θα είναι η συνιστώσα που αντιστοιχεί στη μέγιστη απόλυτα τιμή, δηλαδή το $x_i^{(k-1)}$ θα είναι 1 ή -1. Μπορούμε επομένως να διατυπώσουμε παραλλαγμένα τον αλγόριθμο χρησιμοποιώντας τη $\|\cdot\|_\infty$ ως εξής:

Αλγόριθμος Μεθόδου Δυνάμεων με $\|\cdot\|_\infty$:

Δεδομένα: $A \in I\!R^{n,n}$.
Επιλογή $x^{(0)} \in I\!R^n$, ώστε $\|x^{(0)}\|_\infty = 1$

Αποθήκευση της συνιστώσας i για την οποία $|x_i^{(0)}| = 1$
Για $k = 1$ έως ότου υπάρξει σύγκλιση
$y^{(k)} = Ax^{(k-1)}$
$\lambda^{(k-1)} = \frac{y_i^{(k)}}{x_i^{(k-1)}}$
Υπολογισμός της $\|y^{(k)}\|_\infty$
Αποθήκευση της συνιστώσας i για την οποία $|y_i^{(k)}| = \|y^{(k)}\|_\infty$
$x^{(k)} = \frac{y^{(k)}}{\|y^{(k)}\|_\infty}$
Τέλος 'Για'
Αποτέλεσμα: $\lambda^{(k-1)}$ είναι η προσέγγιση της λ_1 και $x^{(k)}$ η προσέγγιση του x_1.

vi) Οταν θέλουμε να υπολογίσουμε τη μικρότερη ή τη μεγαλύτερη ιδιοτιμή του πίνακα A εφαρμόζουμε τη μέθοδο δυνάμεων στον πίνακα $A + \|A\|I$ ή στον πίνακα $A - \|A\|I$, αντίστοιχα, για οποιαδήποτε norm. Οι ιδιοτιμές των πινάκων αυτών θα είναι $\lambda_i + \|A\|$ ή $\lambda_i - \|A\|$ αντίστοιχα. Επειδή $|\lambda_i| \leq \|A\|, i = 1(1)n$, θα έχουμε $\lambda_i + \|A\| \geq 0$ ενώ $\lambda_i - \|A\| \leq 0$. Στην πρώτη περίπτωση η μέθοδος δυνάμεων υπολογίζει τη μεγαλύτερη των $\lambda_i + \|A\|$ ενώ στη δεύτερη, την μικρότερη των $\lambda_i - \|A\|$. Τελικά αν αφαιρέσουμε ή προσθέσουμε αντίστοιχα τη $\|A\|$ θα έχουμε υπολογίσει τη μεγαλύτερη ή τη μικρότερη ιδιοτιμή.

Παράδειγμα: Δίνουμε εδώ ένα παράδειγμα εφαρμογής του αλγόριθμου για την προσέγγιση της απόλυτα μεγαλύτερης ιδιοτιμής και του αντίστοιχου ιδιοδιανύσματος, του πίνακα $\begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}$, ξεκινώντας με $x^{(0)} = [0\ 1\ 0]^T$, κάνοντας τρείς επαναλήψεις και διατηρώντας τρία δεκαδικά ψηφία στους υπολογισμούς.

$$x^{(0)} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix},\ y^{(1)} = Ax^{(0)} = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix},\ \lambda^{(0)} = {x^{(0)}}^T y^{(1)} = 0,\ \|y^{(1)}\|_2 = \sqrt{2} = 1.414$$

$$x^{(1)} = \begin{bmatrix} 0.707 \\ 0 \\ 0.707 \end{bmatrix},\ y^{(2)} = Ax^{(1)} = \begin{bmatrix} 0.707 \\ 1.414 \\ 0.707 \end{bmatrix},\ \lambda^{(1)} = {x^{(1)}}^T y^{(2)} = 1,\ \|y^{(2)}\|_2 = \sqrt{3} = 1.732$$

$$x^{(2)} = \begin{bmatrix} 0.408 \\ 0.816 \\ 0.408 \end{bmatrix}, y^{(3)} = Ax^{(2)} = \begin{bmatrix} 1.224 \\ 0.816 \\ 1.224 \end{bmatrix}, \lambda^{(2)} = {x^{(2)}}^T y^{(3)} = 1.665, \|y^{(3)}\|_2 = 1.914$$

$x^{(3)} = [0.639\ 0.426\ 0.639]^T$.

Ας σημειωθεί ότι η απόλυτα μεγαλύτερη ιδιοτιμή είναι η $\lambda_1 = 2$ με αντίστοιχο ιδιοδιάνυσμα το

$[1\ 1\ 1]^T$, όπου και θα συγκλίνει ο αλγόριθμος αν συνεχιστεί η εφαρμογή του.

Στη συνέχεια παρουσιάζουμε μια παραλλαγή της μεθόδου δυνάμεων, με την οποία έχουμε τη δυνατότητα να υπολογίζουμε και άλλες ιδιοτιμές εκτός από την απόλυτα μεγαλύτερη, τη μέθοδο των αντίστροφων δυνάμεων.

8.3.1 Μέθοδος Αντίστροφων Δυνάμεων ή Αντίστροφης Επανάληψης

Με την προϋπόθεση ότι ο πίνακας A είναι αντιστρέψιμος, η μέθοδος των αντίστροφων δυνάμεων βασίζεται στο γεγονός ότι οι ιδιοτιμές του A^{-1} είναι οι αντίστροφες των ιδιοτιμών του A. Ετσι, η εφαρμογή της μεθόδου δυνάμεων στον πίνακα A^{-1} οδηγεί στον υπολογισμό της απόλυτα μεγαλύτερης ιδιοτιμής του A^{-1}, το αντίστροφο της οποίας είναι η απόλυτα μικρότερη ιδιοτιμή του A. Η μέθοδος αυτή υπολογίζει ουσιαστικά την πλησιέστερη προς το μηδέν ιδιοτιμή και το αντίστοιχο ιδιοδιάνυσμα. Εκμεταλλευόμενοι και το γεγονός ότι μία μετατόπιση ενός πίνακα κατά σI δίνει μετατόπιση σ σε όλες τις ιδιοτιμές, δηλαδή αν $\lambda_i, i = 1(1)n$, είναι ιδιοτιμές του A τότε $\lambda_i - \sigma$ θα είναι ιδιοτιμές του $A-\sigma I$, μπορούμε να εφαρμόσουμε τη μέθοδο δυνάμεων στον πίνακα $(A-\sigma I)^{-1}$ για να υπολογίσουμε την πλησιέστερη ιδιοτιμή προς τον πραγματικό αριθμό σ. Η μέθοδος αυτή μπορεί να οδηγήσει και στον υπολογισμό όλων των ιδιοτιμών του πίνακα A, αρκεί να επιλεγούν μεθοδικά διαφορετικά σ_i μέσα στο φάσμα των ιδιοτιμών του A και να εφαρμόζεται κάθε φορά ο αλγόριθμος των αντίστροφων δυνάμεων. Θα έλεγε κανείς ότι δε χρειάζεται να δοθεί ξεχωριστός αλγόριθμος αλλά να χρησιμοποιηθεί ο αλγόριθμος της μεθόδου δυνάμεων, όπου στη θέση του A θα τεθεί ο $(A-\sigma I)^{-1}$. Ομως, επειδή στην πράξη αποφεύγεται ο υπολογισμός του αντίστροφου πίνακα ως ασύμφορος και λύνεται ένα γραμμικό σύστημα με πίνακα τον $A-\sigma I$, ο αντίστοιχος αλγόριθμος θα είναι:

Αλγόριθμος Μεθόδου Αντίστροφων Δυνάμεων:

Δεδομένα: $A \in I\!R^{n,n}$, αντιστρέψιμος, $\sigma \in I\!R$.
Επιλογή $x^{(0)} \in I\!R^n$, ώστε $\|x^{(0)}\|_2 = 1$
Για $k = 1$ έως ότου υπάρξει σύγκλιση
$\quad$ Λύση του συστήματος $(A - \sigma I)y^{(k)} = x^{(k-1)}$
$\quad \lambda^{(k-1)} = x^{(k-1)^T} y^{(k)}$
$\quad \mu^{(k-1)} = \frac{1}{\lambda^{(k-1)}} + \sigma$
$\quad x^{(k)} = \frac{y^{(k)}}{\|y^{(k)}\|_2}$
Τέλος 'Για'
Αποτέλεσμα: $\mu^{(k-1)}$ είναι η προσέγγιση της πλησιέστερης προς το σ ιδιοτιμής και
$x^{(k)}$ η προσέγγιση του αντίστοιχου ιδιοδιανύσματος.

Το κόστος του αλγόριθμου χαρακτηρίζεται από το κόστος λύσης συστήματος. Πρέπει να παρατηρήσουμε εδώ ότι οποιαδήποτε από τις μεθόδους που αναπτύχθηκαν σε προηγούμενα κεφάλαια, για την επίλύση συστημάτων, μπορεί να εφαρμοστεί εφόσον το επιτρέπει η κατάσταση του προβλήματος και το συνιστούν τυχόν ιδιότητες του πίνακα A. Το κόστος επομένως του αλγόριθμου ποικίλει. Αν ο πίνακας είναι καλής κατάστασης, τότε προτιμάται η LU παραγοντοποίηση η οποία γίνεται μια φορά στην αρχή με κόστος $\mathcal{O}(n^3)$. Στη συνέχεια, και σε κάθε επανάληψη λύνονται ένα κάτω τριγωνικό και ένα άνω τριγωνικό σύστημα, με προς τα μπρος και προς τα πίσω αντικαταστάσεις, αντίστοιχα, που κοστίζουν $\mathcal{O}(n^2)$. Ετσι, το κόστος του αλγόριθμου είναι $\mathcal{O}(n^3)$, κατά μία τάξη μεγέθους μεγαλύτερο από αυτό της μεθόδου δυνάμεων.

Η ταχύτητα σύγκλισης της μεθόδου θα είναι η ίδια με εκείνη της μεθόδου δυνάμεων για τον πίνακα $(A-\sigma I)^{-1}$. Αν μ_i είναι η πλησιέστερη προς το σ ιδιοτιμή του A που αναζητούμε και μ_j η αμέσως πιο απομακρυσμένη, ώστε η $\lambda_j = \frac{1}{\mu_j-\sigma}$ να είναι η δεύτερη σε απόλυτη τιμή ιδιοτιμή, τότε η ταχύτητα σύγκλισης θα είναι τάξης $\mathcal{O}(|\frac{\lambda_j}{\lambda_i}|^k)$ σύμφωνα με τη θεωρία της μεθόδου δυνάμεων. Αντικαθιστώντας θα έχουμε τάξη σύγκλισης $\mathcal{O}(|\frac{\mu_i-\sigma}{\mu_j-\sigma}|^k)$. Μπορούμε να παρατηρήσουμε ότι αν σ είναι μια καλή προσέγγιση της λ_i τότε η ταχύτητα σύγκλισης θα είναι πολύ μεγάλη. Επομένως η μέθοδος είναι αποτελεσματική όταν υπάρχουν καλές προσεγγίσεις των ιδιοτιμών, ίσως από άλλη μέθοδο, και επιθυμούμε να τις προσεγγίσουμε καλύτερα και να υπολογίσουμε και τα αντίστοιχα ιδιοδιανύσματα.

Παράδειγμα: Δίνουμε στη συνέχεια ένα παράδειγμα όπου φαίνεται η μεγάλη ταχύτητα σύγκλισης. Ο πίνακας $A = \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix}$ έχει ιδιοτιμές τις 2, $2+\sqrt{2}$ και $2-\sqrt{2}$. Η ιδιοτιμή $2-\sqrt{2}$ $(=0.58578644)$ βρίσκεται κοντά στο 0.5, ενώ το αντίστοιχο ιδιοδιάνυσμα κανονικοποιημένο, είναι $[\frac{1}{2}\ \frac{\sqrt{2}}{2}\ \frac{1}{2}]^T$. Θα εφαρμόσουμε τη μέθοδο των αντίστροφων δυνάμεων με $\sigma = 0.5$ χρησιμοποιώντας την LU παραγοντοποίηση.

$$A-\sigma I = \begin{bmatrix} 1.5 & -1 & 0 \\ -1 & 1.5 & -1 \\ 0 & -1 & 1.5 \end{bmatrix} = \begin{bmatrix} 1 & & \\ -\frac{2}{3} & 1 & \\ 0 & -1.2 & 1 \end{bmatrix} \begin{bmatrix} 1.5 & -1 & 0 \\ & \frac{5}{6} & -1 \\ & & 0.3 \end{bmatrix} = LU.$$

Επιλέγουμε $x^{(0)} = [0\ 1\ 0]^T$ και αρχίζουμε με την πρώτη επανάληψη.

$Lz^{(1)} = x^{(0)} \Rightarrow z^{(1)} = [0\ 1\ 2]^T,\ Uy^{(1)} = z^{(1)} \Rightarrow y^{(1)} = [4\ 6\ 4]^T,$

$\lambda^{(0)} = x^{(0)^T} y^{(1)} = 6,\ \mu^{(0)} = \frac{1}{\lambda^{(0)}} + \sigma = 0.66666667,\ \|y^{(1)}\|_2 = 8.24621125,$
$x^{(1)} = \frac{y^{(1)}}{\|y^{(1)}\|_2} = [0.48507125\ 0.72760688\ 0.48507125]^T.$
Ακολουθώντας την ίδια ακριβώς διαδικασία, η δεύτερη και τρίτη επανάληψη δίνουν.

$\Delta\epsilon\acute{\upsilon}\tau\epsilon\rho\eta$	$T\rho\acute{\iota}\tau\eta$	$A\kappa\rho\iota\beta\epsilon\acute{\iota}\varsigma\ \tau\iota\mu\acute{\epsilon}\varsigma$
$\mu^{(1)} = 0.58585859$	$\mu^{(2)} = 0.58578650$	$\mu = 0.58578644$

$$x^{(2)} = \begin{bmatrix} 0.49956654 \\ 0.70771926 \\ 0.49956654 \end{bmatrix} \quad x^{(3)} = \begin{bmatrix} 0.49998725 \\ 0.70712482 \\ 0.49998725 \end{bmatrix} \quad x = \begin{bmatrix} 0.5 \\ 0.70710678 \\ 0.5 \end{bmatrix}.$$

Παρατηρούμε ότι στη δεύτερη επανάληψη έχουμε σύγκλιση με ακρίβεια τριών δεκαδικών ψηφίων ενώ στην τρίτη με ακρίβεια έξι για την ιδιοτιμή και τεσσάρων δεκαδικών ψηφίων για το ιδιοδιάνυσμα. Οι πράξεις έγιναν με διπλή ακρίβεια με το πρόγραμμα MATLAB.

8.3.2 Τεχνική της Υποτίμησης (Deflation)

Η τεχνική της υποτίμησης οδηγεί ακολουθιακά στον υπολογισμό και των άλλων ιδιοτιμών και ιδιοδιανυσμάτων, με βάση τη μέθοδο δυνάμεων. Βασίζεται στην ιδέα της εξάλειψης, με κάποιον τρόπο, της απόλυτα μεγαλύτερης ιδιοτιμής που μόλις βρέθηκε, μετασχηματίζοντας τον πίνακα A, ώστε η εφαρμογή της μεθόδου δυνάμεων στο νέο πίνακα, να υπολογίζει τη δεύτερη απόλυτα μεγαλύτερη ιδιοτιμή. Εχουν επινοηθεί δύο τέτοιες τεχνικές, μία για συμμετρικούς πίνακες και μια δεύτερη, γενικά, για κάθε πίνακα $A \in I\!R^{n,n}$.

Για την πρώτη τεχνική, έστω ότι ο πίνακας $A \in I\!R^{n,n}$ είναι συμμετρικός με ιδιοτιμές τις $\lambda_i,\ i = 1(1)n$, ώστε $|\lambda_1| > |\lambda_2| > |\lambda_3| \geq \cdots \geq |\lambda_n|$ και αντίστοιχα ιδιοδιανύσματα $x_i,\ i = 1(1)n$, κανονικοποιημένα. Με εφαρμογή της μεθόδου δυνάμεων βρίσκεται αρχικά μια προσέγγιση της λ_1 και του x_1. Εχοντας αυτά ως γνωστά κατασκευάζουμε τον πίνακα

$$A_1 \ = \ A - \lambda_1 x_1 x_1^T.$$

Ο A_1 έχει ιδιοτιμές τις $\lambda_i,\ i = 2(1)n$, με αντίστοιχα ιδιοδιανύσματα $x_i,\ i = 2(1)n$, και τη 0 με αντίστοιχο ιδιοδιάνυσμα το x_1. Πραγματικά

$$A_1 x_i \ = \ (A - \lambda_1 x_1 x_1^T) x_i \ = \ A x_i - \lambda_1 x_1 x_1^T x_i \ = \ \lambda_i x_i,\ i = 2(1)n. \tag{8.16}$$

Στην (8.16) $x_1^T x_i \ = \ 0,\ i = 2(1)n$, λόγω της ορθογωνιότητας των ιδιοδιανυσμάτων ενός συμμετρικού πίνακα. Επίσης

$$A_1 x_1 \ = \ (A - \lambda_1 x_1 x_1^T) x_1 \ = \ A x_1 - \lambda_1 x_1 x_1^T x_1 \ = \ \lambda_1 x_1 - \lambda_1 x_1 \ = \ 0.$$

Εφαρμόζοντας τη μέθοδο δυνάμεων στον A_1, προσεγγίζουμε τη λ_2 και το x_2. Μπορούμε να συνεχίσουμε τη διαδικασία παίρνοντας $A_2 \ = \ A_1 - \lambda_2 x_2 x_2^T$ κ.ο.κ.

Πρέπει να σημειώσουμε εδώ ότι αν λάβουμε υπόψη την παρατήρηση (ii) της παραγράφου 14.1, συμπεραίνουμε ότι η τεχνική αυτή μπορεί να εφαρμοστεί και όταν η λ_1 είναι πολλαπλή ιδιοτιμή. Τότε, με την τεχνική υποτίμησης, η μία από τις πολλές λ_1 γίνεται μηδέν και οι άλλες παραμένουν ως έχουν. Με την εφαρμογή της μεθόδου δυνάμεων θα βρεθεί ξανά η λ_1 με αντίστοιχο ιδιοδιάνυσμα κάποιο άλλο, ορθογώνιο προς το πρώτο. Με άλλα λόγια, η τεχνική υποτίμησης υποβιβάζει την πολλαπλότητα της λ_1 κατά ένα κάθε φορά. Μπορούμε λοιπόν να πούμε ότι η τεχνική υποτίμησης οδηγεί πάντα στον υπολογισμό όλων των ιδιοτιμών και ιδιοδιανυσμάτων του A. Εκείνο που πρέπει να τονιστεί είναι ότι χρειάζεται μεγάλη ακρίβεια στις προσεγγίσεις με τη μέθοδο δυνάμεων, διότι με την τεχνική υποτίμησης μεταφέρονται τα σφάλματα από βήμα σε βήμα και συσσωρεύονται, με συνέπεια να αλλοιώνονται τα αποτελέσματα από ένα σημείο και πέρα.

Στη συνέχεια παρουσιάζουμε τη δεύτερη τεχνική. Εστω πίνακας $A \in I\!R^{n,n}$, $\lambda_i,\ i = 1(1)n$, οι ιδιοτιμές του, τ.ω. $|\lambda_1| > |\lambda_2| > |\lambda_3| \geq \cdots \geq |\lambda_n|$, με αντίστοιχα, κανονικοποιημένα ως προς τη $\|\cdot\|_\infty$, ιδιοδιανύσματα x_i. Εστω επίσης ότι οι λ_1 και λ_2 είναι πραγματικές απλές ιδιοτιμές και επομένως τα ιδιοδιανύσματα x_1 και x_2 υπάρχουν και είναι γραμμικά ανεξάρτητα. Εφαρμόζοντας τη μέθοδο δυνάμεων προσεγγίζουμε την ιδιοτιμή λ_1 και το ιδιοδιάνυσμα x_1. Εχοντας αυτά ως γνωστά μπορούμε να προχωρήσουμε για την προσέγγιση των στοιχείων λ_2 και x_2 ως εξής:

Εστω ότι η k-οστή συνιστώσα του x_1 είναι μονάδα ($(x_1)_k = \|x_1\|_\infty = 1$). Συμβολίζουμε με a_k^T την k-οστή γραμμή του A και κατασκευάζουμε τον πίνακα

$$A_1 \ = \ A - x_1 a_k^T.$$

Ο πίνακας A_1 έχει ιδιοτιμές τις $\lambda_i,\ i = 2(1)n$, με αντίστοιχα ιδιοδιανύσματα $x_i - x_1$, και την ιδιοτιμή 0 με αντίστοιχο ιδιοδιάνυσμα το x_1. Πραγματικά, θεωρώντας ότι και το x_i έχει μονάδα στην k-οστή συνιστώσα του ($(x_i)_k = 1$), αφού είναι ιδιοδιάνυσμα και άρα ανεξάρτητο από σταθερό παράγοντα, έχουμε

$$\begin{aligned} A_1(x_i - x_1) &= (A - x_1 a_k^T)(x_i - x_1) \ = \ Ax_i - Ax_1 - x_1 a_k^T x_i + x_1 a_k^T x_1 \\ &= \lambda_i x_i - \lambda_1 x_1 - \lambda_i x_1 + \lambda_1 x_1 \ = \ \lambda_i (x_i - x_1),\ i = 2(1)n, \end{aligned} \tag{8.17}$$

αφού $a_k^T x_1 = \lambda_1$ επειδή a_k^T είναι η k-οστή γραμμή του A και x_1 ιδιοδιάνυσμα με $(x_1)_k = 1$ και $a_k^T x_i = \lambda_i$ για τον ίδιο λόγο. Επίσης

$$A_1 x_1 \ = \ (A - x_1 a_k^T) x_1 \ = \ Ax_1 - x_1 a_k^T x_1 \ = \ \lambda_1 x_1 - \lambda_1 x_1 \ = \ 0.$$

Ακόμη, ο πίνακας A_1 έχει μηδενική ολόκληρη την k-οστή γραμμή, αφού ο πίνακας $x_1 a_k^T$ έχει την a_k^T ως k-οστή γραμμή του. Αν θεωρήσουμε την ορίζουσα $\det(A_1 - \lambda I_n)$ και αναπτύξουμε ως προς την k-οστή γραμμή της, θα έχουμε

$$\det(A_1 - \lambda I_n) \ = \ -\lambda \det(B_1 - \lambda I_{n-1}), \tag{8.18}$$

όπου ο $B_1 \in I\!R^{n-1,n-1}$ προκύπτει από τον A_1 αν διαγράψουμε την k-οστή γραμμή και στήλη. Η (8.18) δίνει ότι ο πίνακας B_1 έχει όλες τις ιδιοτιμές του A_1 εκτός από τη μηδενική. Εφαρμόζοντας λοιπόν τη μέθοδο δυνάμεων στον πίνακα B_1, προσεγγίζουμε την απόλυτα μεγαλύτερη ιδιοτιμή του που είναι η λ_2 και το αντίστοιχο ιδιοδιάνυσμα $y_2 \in I\!R^{n-1}$. Απομένει να ακολουθήσουμε την αντίστροφη πορεία για την προσέγγιση του $x_2 \in I\!R^n$ έχοντας γνωστό το $y_2 \in I\!R^{n-1}$.
Για το σκοπό αυτό επεκτείνουμε το διάνυσμα y_2 στο $z_2 \in I\!R^n$ θέτοντας τη συνιστώσα 0 μεταξύ των $k-1$ και k στη σειρά συνιστωσών του y_2, δηλαδή

$$z_2 \;=\; [(y_2)_1 \; (y_2)_2 \; \cdots \; (y_2)_{k-1} \; 0 \; (y_2)_k \; \cdots \; (y_2)_{n-1}]^T.$$

Εύκολα προκύπτει, από το γεγονός ότι υπάρχει 0 στην k-οστή συνιστώσα του z_2 και μηδενική γραμμή στην k-οστή γραμμή του A_1, ότι το διάνυσμα z_2 είναι ιδιοδιάνυσμα του A_1 με ιδιοτιμή τη λ_2. Από την (8.17) έχουμε ότι $x_2 - x_1$ είναι το ιδιοδιάνυσμα του A_1 με ιδιοτιμή τη λ_2. Επομένως το $x_2 - x_1$ θα είναι πολλαπλάσιο του z_2, δηλαδή

$$x_2 - x_1 \;=\; cz_2 \;\Leftrightarrow\; x_2 \;=\; x_1 + cz_2,\; c \in I\!R \setminus \{0\}. \tag{8.19}$$

Για τον προσδιορισμό της σταθεράς c πολλαπλασιάζουμε με a_k^T από αριστερά τα δύο μέλη της (8.19), οπότε

$$a_k^T x_2 \;=\; a_k^T x_1 + c a_k^T z_2 \;\Leftrightarrow\; \lambda_2 \;=\; \lambda_1 + c a_k^T z_2 \;\Leftrightarrow\; c \;=\; \frac{\lambda_2 - \lambda_1}{a_k^T z_2}.$$

Εδώ ολοκληρώθηκε η διαδικασία υπολογισμού και της δεύτερης ιδιοτιμής με το αντίστοιχο ιδιοδιάνυσμα. Απομένει να διευκρινίσουμε μια λεπτομέρεια, την περίπτωση όπου $a_k^T z_2 = 0$. Στην περίπτωση αυτή το c θεωρείται άπειρο ενώ το $\frac{1}{c}$ μηδέν. Τότε, επειδή το x_2 είναι ιδιοδιάνυσμα, μπορούμε να διαιρέσουμε με τη σταθερά c το δεύτερο μέλος της (8.19), οπότε θα έχουμε $x_2 = \frac{1}{c}x_1 + z_2$ ή $x_2 = z_2$. Στην περίπτωση αυτή το z_2 θα είναι το ιδιοδιάνυσμα. Εδώ αίρουμε κάποια ασάφεια στη θεώρηση $(x_i)_k = 1$ στη σχέση (8.17). Εκεί δεν αναφέρθηκε κάτι για την περίπτωση $(x_i)_k = 0$. Διευκρινίζουμε εδώ ότι αν $(x_2)_k = 0$, τότε $x_2 = z_2$. Δίνουμε στη συνέχεια την παραπάνω περιγραφή σε μορφή αλγόριθμου:

Αλγόριθμος Τεχνικής Υποτίμησης :

Δεδομένα: $A \in I\!R^{n,n}$.
Υλοποίηση αλγόριθμου μεθόδου δυνάμεων - Υπολογισμός των λ_1 και x_1
Κανονικοποίηση του x_1 ώστε $\|x_1\|_\infty = 1$ - Εύρεση του k ώστε $(x_1)_k = 1$
$A_1 \;=\; A - x_1 a_k^T$ (a_k^T η k γραμμή του A)
Υποτίμηση του πίνακα A_1 στον πίνακα $B_1 \in I\!R^{n-1,n-1}$
Υλοποίηση αλγόριθμου μεθόδου δυνάμεων στον B_1 - Υπολογισμός των λ_2 και $y_2 \in I\!R^{n-1}$
Επέκταση του y_2 στο $z_2 \in I\!R^n$
$s \;=\; a_k^T z_2$
Αν $s \;=\; 0$ τότε

$x_2 = z_2$

αλλιώς

$c = \frac{\lambda_2 - \lambda_1}{s}$

$x_2 = x_1 + cz_2$

Τέλος 'Αν'

Αποτέλεσμα: λ_1 και λ_2, ιδιοτιμές του A με αντίστοιχα ιδιοδιανύσματα x_1 και x_2.

Θα πρέπει να παρατηρήσουμε εδώ ότι η διαδικασία της υποτίμησης μπορεί να εφαρμοστεί και για τον υπολογισμό των στοιχείων λ_3 και x_3, αρκεί στη θέση του A να θεωρήσουμε τον πίνακα $B_1 \in I\!R^{n-1,n-1}$ και να εφαρμόσουμε τον παραπάνω αλγόριθμο. Αυτός όμως υπολογίζει το ιδιοδιάνυσμα y_3 του B_1. Θα πρέπει επομένως να το επεκτείνουμε ξανά με τη διαδικασία που αναφέραμε για να μεταβούμε στο ιδιοδιάνυσμα x_3 του A. Μ' αυτή τη διαδικασία μπορούμε να πάμε σε περισσότερο βάθος και να υπολογίσουμε όλες τις ιδιοτιμές, αν είναι δυνατόν, και στη συνέχεια να επιστρέψουμε με συνεχείς επεκτάσεις, μέχρι την κορυφή, για τον υπολογισμό των αντίστοιχων ιδιοδιανυσμάτων. Ακόμη, η τεχνική της υποτίμησης μπορεί να εφαρμοστεί και χωρίς τη μέθοδο δυνάμεων στην αρχή, αρκεί να γνωρίζουμε τα στοιχεία λ_1 και x_1, από οποιαδήποτε άλλη πηγή, και ας μην είναι αναγκαστικά η λ_1 η απόλυτα μεγαλύτερη ιδιοτιμή.

Παράδειγμα: Δίνουμε στη συνέχεια ένα παράδειγμα τεχνικής υποτίμησης χωρίς να απαιτείται καθόλου εφαρμογή μεθόδου δυνάμεων. Ο πίνακας $A = \begin{bmatrix} 1 & 2 & 3 \\ -1 & 3 & 4 \\ -1 & 2 & 5 \end{bmatrix}$ έχει ιδιοτιμή τη $\lambda_1 = 6$ με αντίστοιχο ιδιοδιάνυσμα το $x_1 = [1\ 1\ 1]^T$. Θα προσπαθήσουμε να υπολογίσουμε τις άλλες δύο ιδιοτιμές και τα αντίστοιχα ιδιοδιανύσματα, με την τεχνική υποτίμησης.
Παρατηρούμε ότι το x_1 είναι κανονικοποιημένο, ως προς τη $\|\cdot\|_\infty$, με μονάδες σε όλα τα στοιχεία, επομένως μπορούμε να θεωρήσουμε οποιαδήποτε από τις τρεις γραμμές του A ως a_k^T. Επιλέγουμε την πρώτη $a_1^T = [1\ 2\ 3]$. Τότε

$$A_1 = A - x_1 a_1^T = \begin{bmatrix} 1 & 2 & 3 \\ -1 & 3 & 4 \\ -1 & 2 & 5 \end{bmatrix} - \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ -2 & 1 & 1 \\ -2 & 0 & 2 \end{bmatrix},$$

οπότε B_1 θα είναι ο πίνακας $\begin{bmatrix} 1 & 1 \\ 0 & 2 \end{bmatrix}$. Αυτός είναι άνω τριγωνικός επομένως οι ιδιοτιμές του θα είναι $\lambda_2 = 2$ και $\lambda_3 = 1$. Τα αντίστοιχα ιδιοδιανύσματα προκύπτουν εύκολα από τη λύση των συστημάτων $(B_1 - \lambda_i I)y = 0,\ i = 1, 2$, και είναι $y_2 = [1\ 1]^T$ και $y_3 = [1\ 0]^T$.
Επεκτείνουμε το y_2 στο $z_2 = [0\ 1\ 1]^T$. Στη συνέχεια βρίσκουμε το $c_2 = \frac{\lambda_2 - \lambda_1}{a_1^T z_2} = \frac{2-6}{5} = -0.8$ και το $x_2 = x_1 + c_2 z_2 = [1\ 1\ 1]^T - 0.8[0\ 1\ 1]^T = [1\ 0.2\ 0.2]^T$.
Η ίδια ακριβώς διαδικασία για το τρίτο ιδιοδιάνυσμα δίνει
$y_3 = [1\ 0]^T$, $z_3 = [0\ 1\ 0]^T$, $c_3 = \frac{\lambda_3 - \lambda_1}{a_1^T z_3} = \frac{1-6}{2} = -2.5$ και $x_3 = x_1 + c_3 z_3 =$

$[1\ 1\ 1]^T\ -\ 2.5[0\ 1\ 0]^T\ =\ [1\ -1.5\ 1]^T.$

Ισχύουν κι εδώ οι παρατηρήσεις που κάναμε για τη συσσώρευση σφαλμάτων στην τεχνική υποτίμησης για συμμετρικούς πίνακες. Κι εδώ όσο προχωρούμε σε βάθος, τόσο αυξάνουν τα σφάλματα, με συνέπεια να μην είναι αποτελεσματική η μέθοδος από ένα σημείο και μετά. Το κόστος της χαρακτηρίζεται κυρίως από το κόστος της μεθόδου δυνάμεων που είναι $\mathcal{O}(mn^2)$. Αν θελήσουμε όμως να υπολογίσουμε όλες τις ιδιοτιμές, θα εφαρμόσουμε τη μέθοδο δυνάμεων $n-1$ φορές, κάθε φορά όμως και με πίνακα διάστασης ελαττωμένης κατά ένα. Αν θεωρήσουμε ότι m είναι ο μέσος όρος των επαναλήψεων σε κάθε εφαρμογή και αθροίσουμε, παίρνουμε συνολικό κόστος $\mathcal{O}(mn^3)$.

Λόγω του ότι αφενός οι τεχνικές υποτίμησης έχουν προβλήματα συσσώρευσης σφαλμάτων και αφετέρου η μέθοδος δυνάμεων δε λύνει το πρόβλημα σ΄ όλες τις περιπτώσεις, οι μέθοδοι αυτές γίνονται αναποτελεσματικές. Ετσι, η έρευνα στράφηκε σε επινόηση επαναληπτικών μεθόδων που θα προσεγγίζουν συγχρόνως όλες τις ιδιοτιμές και κυρίως όχι με διαδικασία αλληλουχίας. Τέτοιες μέθοδοι θα μελετηθούν στη συνέχεια.

8.4 Μέθοδος QR

Η μέθοδος QR προήρθε ως επέκταση της μεθόδου δυνάμεων για την ταυτόχρονη εύρεση όλων των ιδιοτιμών. Βασίζεται στην QR παραγοντοποίηση που αναπτύχτηκε στο προηγούμενο κεφάλαιο. Ξεκινώντας με την ιδέα της επέκτασης της μεθόδου δυνάμεων θεωρούμε αρχικά ότι ο πίνακας $A \in I\!R^{n,n}$ έχει διακεκριμένες ιδιοτιμές $\lambda_i,\ i = 1(1)n$, με $|\lambda_1| > |\lambda_2| > \cdots > |\lambda_n|$. Στη θέση του αρχικού διανύσματος $x^{(0)} \in I\!R^n$, θεωρούμε τον ορθογώνιο πίνακα $X_0 \in I\!R^{n,p},\ p \leq n$. Στη συνέχεια παράγουμε τις λεγόμενες ορθογώνιες επαναλήψεις ως εξής

$$Y_{i+1}\ =\ AX_i, \tag{8.20}$$

ενώ ως X_{i+1} παίρνουμε τον ορθογώνιο παράγοντα του πίνακα Y_{i+1} κατά την QR παραγοντοποίηση

$$Y_{i+1}\ =\ X_{i+1}R_{i+1}, \tag{8.21}$$

όπου $R_{i+1} \in I\!R^{p,p}$ είναι άνω τριγωνικός πίνακας. Είναι φανερό ότι για $p = 1$ έχουμε ακριβώς τη μέθοδο δυνάμεων. Για $p > 1$, από την (8.21) παίρνουμε ότι ο Y_{i+1} παράγεται από γραμμικούς συνδυασμούς των στηλών του X_{i+1}, επομένως οι δύο αυτοί πίνακες παράγουν τον ίδιο χώρο

$$\text{span}\{X_{i+1}\}\ =\ \text{span}\{Y_{i+1}\}\ =\ \text{span}\{AX_i\}. \tag{8.22}$$

Θεωρούμε τη μορφή Jordan του πίνακα A

$$A\ =\ S\Lambda S^{-1}\ =\ S\text{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n)S^{-1}.$$

Η σχέση (8.22) δίνει επαγωγικά ότι

$$\text{span}\{X_i\} \;=\; \text{span}\{A^i X_0\} \;=\; \text{span}\{S\Lambda^i S^{-1} X_0\}.$$

Κατασκευάζουμε στη συνέχεια τον πίνακα $S\Lambda^i S^{-1} X_0$

$$S\Lambda^i S^{-1} X_0 \;=\; \lambda_p^i S \begin{bmatrix} \left(\frac{\lambda_1}{\lambda_p}\right)^i & & & & & \\ & \left(\frac{\lambda_2}{\lambda_p}\right)^i & & & & \\ & & \ddots & & & \\ & & & 1 & & \\ & & & & \ddots & \\ & & & & & \left(\frac{\lambda_n}{\lambda_p}\right)^i \end{bmatrix} S^{-1} X_0 \;= \lambda_p^i S \begin{bmatrix} U_1^{(i)} \\ U_2^{(i)} \end{bmatrix},$$

όπου θεωρήσαμε τα blocks $U_1^{(i)} \in I\!R^{p,p}$ και $U_2^{(i)} \in I\!R^{n-p,p}$. Επειδή $|\frac{\lambda_j}{\lambda_p}| < 1, j = p+1(1)n$, η ακολουθία πινάκων $U_2^{(i)}$ συγκλίνει στο μηδενικό πίνακα με ταχύτητα $\mathcal{O}(|\frac{\lambda_{p+1}}{\lambda_p}|^i)$, αφού το $\frac{\lambda_{p+1}}{\lambda_p}$ είναι το απόλυτα μεγαλύτερο στοιχείο στο αντίστοιχο block του διαγώνιου πίνακα. Θεωρούμε τον αντίστοιχο διαχωρισμό του πίνακα S

$$S \;=\; [S_1|S_2],\; S_1 \in I\!R^{n,p},\; S_2 \in I\!R^{n,n-p}.$$

Τότε,

$$S\Lambda^i S^{-1} X_0 \;=\; \lambda_p^i (S_1 U_1^{(i)} + S_2 U_2^{(i)})$$

και επειδή $\lim_{i\to\infty} U_2^{(i)} = 0$, έχουμε

$$\text{span}\{X_i\} \;=\; \text{span}\{S\Lambda^i S^{-1} X_0\} \;=\; \text{span}\{S_1 U_1^{(i)} + S_2 U_2^{(i)}\} \;\to\; \text{span}\{S_1 U_1^{(i)}\}. \tag{8.23}$$

Θεωρούμε ότι ο πίνακας $U_1^{(i)}$ είναι πλήρους βαθμού ($\text{rank}(U_1^{(i)}) = p$). Η προϋπόθεση αυτή πετυχαίνεται αν ξεκινήσουμε με πίνακα $U_1^{(0)}$ πλήρους βαθμού. Το τελευταίο είναι η γενίκευση της υπόθεσης $c_1 \neq 0$ στο γραμμικό συνδυασμό (8.13) της μεθόδου δυνάμεων. Με την προϋπόθεση αυτή έχουμε ότι

$$\text{span}\{S_1 U_1^{(i)}\} \;=\; \text{span}\{S_1\}$$

και σύμφωνα με την (8.23), ο χώρος που παράγεται από τον πίνακα X_i τείνει προς τον χώρο που παράγεται από τον πίνακα S_1, που αποτελείται από τα πρώτα p ιδιοδιανύσματα του A. Εδώ πρέπει να σημειώσουμε ότι αν θεωρήσουμε έναν ακέραιο $r < p$ και λάβουμε τις r πρώτες στήλες της ακολουθίας X_i, από την παραπάνω διδικασία, θα έχουμε την ίδια ακριβώς ακολουθία $X_i' \in I\!R^{n,r}$ με εκείνη που θα προερχόταν αν εξαρχής θεωρούσαμε $p = r$. Θεωρώντας τώρα όλα τα r από 1 έως p, θα ισχύουν όλα τα παραπάνω συμπεράσματα, επομένως ο $U_1^{(i)}$ θα τείνει σε άνω τριγωνικό πίνακα. Στη συνέχεια θεωρούμε $p = n$ και $X_0 = I$. Η παραπάνω θεωρία των ορθογώνιων επαναλήψεων δίνει την επαναληπτική μέθοδο QR, βάση της οποίας είναι το επόμενο θεώρημα.

Θεώρημα 8.4 *Εστω $A \in I\!R^{n,n}$, με διακεκριμένες απόλυτες τιμές ιδιοτιμών. Θεωρούμε τις ορθογώνιες επαναλήψεις $X_i \in I\!R^{n,n}$, $(p = n)$ $i = 1, 2, \cdots$ με $X_0 = I$ και τον πίνακα S αποτελούμενο από τα ιδιοδιανύσματα του A, που αντιστοιχούν στη φθίνουσα τάξη των μέτρων των αντίστοιχων ιδιοτιμών. Αν όλοι οι κύριοι υποπίνακες $S_{jj}, j = 1(1)n$, που αποτελούνται από τις πρώτες j γραμμές και στήλες του S, είναι πλήρους βαθμού (*rank*$(S_{jj}) = j$), τότε η ακολουθία πινάκων $A_i = X_i^T A X_i$ συγκλίνει στην κανονική μορφή* Schur, *δηλαδή σε άνω τριγωνικό πίνακα όπου στη διαγώνιο εμφανίζονται οι ιδιοτιμές του A κατά φθίνουσα τάξη απόλυτου μεγέθους.*

Απόδειξη: Η απαίτηση $S_{jj}, j = 1(1)n$, να είναι πλήρους βαθμού, προέρχεται από την απαίτηση ο $U_1^{(0)}$ να είναι πλήρους βαθμού, όπως αναφέρθηκε προηγουμένως. Διαχωρίζουμε τον πίνακα X_i στην block μορφή $X_i = [X_{1i}|X_{2i}], X_{1i} \in I\!R^{n,p}, X_{2i} \in I\!R^{n,n-p}$. Τότε η ακολουθία A_i δίνεται ως

$$A_i \quad = \quad X_i^T A X_i \; = \; \begin{bmatrix} X_{1i}^T A X_{1i} & X_{1i}^T A X_{2i} \\ X_{2i}^T A X_{1i} & X_{2i}^T A X_{2i} \end{bmatrix}.$$

Σύμφωνα με την παραπάνω θεωρία των ορθογώνιων επαναλήψεων, ο υπόχωρος span$\{X_{1i}\}$ συγκλίνει στον span$\{S_1\}$ στον οποίο συγκλίνει και ο span$\{AX_{1i}\}$, επομένως

$$\text{span}\{X_{1i}\} \; \rightarrow \; \text{span}\{AX_{1i}\}.$$

Τότε το γινόμενο $X_{2i}^T A X_{1i}$ συγκλίνει στο $X_{2i}^T X_{1i} = 0$. Αρα ο πίνακας A_i συγκλίνει σε block άνω τριγωνικό, και επειδή αυτό ισχύει για όλα τα p από 1 έως $n-1$, προκύπτει ότι η ακολουθία A_i συγκλίνει σε άνω τριγωνικό πίνακα. □

Θα πρέπει να σημειωθεί ότι η τάξη σύγκλισης για τα στοιχεία του $X_{2i}^T A X_{1i}$, θα είναι $\mathcal{O}(|\frac{\lambda_{p+1}}{\lambda_p}|^i)$ ενώ στην p διαγώνια θέση θα έχουμε σύγκλιση στην ιδιοτιμή λ_p με τάξη σύγκλισης $\mathcal{O}(\max\{|\frac{\lambda_{p+1}}{\lambda_p}|^i, |\frac{\lambda_p}{\lambda_{p-1}}|^i\})$.
Ακόμη έχουμε να παρατηρήσουμε ότι αν δεν ισχύει η απαίτηση να είναι πλήρους βαθμού οι πίνακες S_{jj}, τότε οι ορθογώνιες επαναλήψεις συγκλίνουν αν θεωρήσουμε ως S μια μικρή αναδιάταξη των στηλών του έτσι ώστε να ισχύει η απαίτηση. Ετσι όμως θα παρουσιαστούν το ίδιο αναδιατεταγμένες και οι ιδιοτιμές στη διαγώνιο. Στην πράξη, λόγω σφαλμάτων στρογγύλευσης, θα παρουσιαστεί το φαινόμενο που περιγράφτηκε στην παρατήρηση (i) της μεθόδου δυνάμεων.

Στη συνέχεια δίνεται ο αλγόριθμος της μεθόδου QR.

Αλγόριθμος Μεθόδου QR:

Δεδομένα: $A \in I\!R^{n,n}$.
$A_0 = A$

Για $i = 0$ έως ότου υπάρξει σύγκλιση
$A_i = Q_i R_i$ (QR Παραγοντοποίηση του A_i)
$A_{i+1} = R_i Q_i$
Τέλος 'Για'
Αποτέλεσμα: A_{i+1} είναι η προσέγγιση της κανονικής μορφής Schur του πίνακα A.

Απομένει να αποδείξουμε ότι ο αλγόριθμος, που δόθηκε, παράγει την ίδια ακολουθία A_i, με εκείνη των ορθογώνιων επαναλήψεων. Αυτό θα γίνει με τη μέθοδο της τέλειας επαγωγής.
Για $i = 1$, από τις ορθογώνιες επαναλήψεις έχουμε $A_1 = X_1^T A X_1$ αλλά $A = X_1 R_1$ είναι η QR παραγοντοποίηση, επομένως $A_1 = X_1^T X_1 R_1 X_1 = R_1 X_1$, που δίνει τον A_1 του αλγόριθμου QR. Εστω ότι ο $A_i = X_i^T A X_i$ προέρχεται και από τον αλγόριθμο QR. Θα αποδείξουμε ότι ο ισχυρισμός αληθεύει και για την $i+1$ επανάληψη. Εχουμε

$$A_{i+1} \;=\; X_{i+1}^T A X_{i+1} \;=\; (X_{i+1}^T A X_i) X_i^T X_{i+1}, \tag{8.24}$$

αλλά από τις ορθογώνιες επαναλήψεις (8.20) και (8.21) παίρνουμε ότι

$$A X_i \;=\; X_{i+1} R_{i+1}. \tag{8.25}$$

Αντικαθιστώντας στην (8.24) έχουμε

$$A_{i+1} \;=\; (X_{i+1}^T X_{i+1} R_{i+1}) X_i^T X_{i+1} \;=\; R_{i+1}(X_i^T X_{i+1}). \tag{8.26}$$

Από την υπόθεση της τέλειας επαγωγής και την (8.25) παίρνουμε

$$A_i \;= X_i^T A X_i \;=\; X_i^T X_{i+1} R_{i+1}. \tag{8.27}$$

Παρατηρούμε ότι η (8.27) δίνει την QR παραγοντοποίηση του A_i, με $Q = X_i^T X_{i+1}$ και $R = R_{i+1}$, ενώ η (8.26) δίνει τον A_{i+1} ως το γινόμενο RQ. Αυτό όμως είναι μια επανάληψη της μεθόδου QR, που δίνει τον A_{i+1} από τον A_i και η απόδειξη ολοκληρώθηκε.

Για την υλοποίηση του αλγόριθμου QR χρειάζεται να γίνει η QR παραγοντοποίηση, αλγόριθμοι της οποίας δόθηκαν στο προηγούμενο κεφάλαιο. Οι πιο αποτελεσματικοί είναι οι αλγόριθμοι Householder και Givens. Αποδείχνεται ότι ένας συνδυασμός των δύο αυτών αλγόριθμων καθιστά την μέθοδο QR ως την πιο αποτελεσματική μέθοδο. Το κόστος της ανέρχεται συνολικά σε $\mathcal{O}(n^3)$, ενώ είναι ευσταθής λόγω της χρήσης ορθογώνιων μετασχηματισμών.

ΑΣΚΗΣΕΙΣ

1.: Χρησιμοποιώντας κατάλληλη επαναληπτική μέθοδο και αρχικό διάνυσμα $x^{(0)} = [1\ 1\ 1]^T$, να βρεθούν κατά προσέγγιση, μετά από δυο επαναλήψεις, η απόλυτα μικρότερη ιδιοτιμή και το αντίστοιχο ιδιοδιάνυσμα του πίνακα $A = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & -1 \\ 1 & -1 & 0 \end{bmatrix}$.

2.: Δίνεται ο πίνακας $A = \begin{bmatrix} 1 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 1 \end{bmatrix}$. Χρησιμοποιώντας τη μέθοδο των δυνάμεων, με αρχικό διάνυσμα $x^{(0)} = [1\ 0\ 1]^T$, και την επέκτασή της (τεχνική υποτίμησης), να βρεθούν όλες οι ιδιοτιμές του με προσέγγιση τεσσάρων σηματικών ψηφίων.

3.: Δίνεται ο πίνακας $A = \begin{bmatrix} 6 & 1 & 1 \\ 1 & 3 & 2 \\ 1 & 2 & 3 \end{bmatrix}$. Αφού αποδειχτεί ότι ο A είναι θετικά ορισμένος να προσεγγιστεί με κάποια παραλλαγή της μεθόδου των δυνάμεων η φασματική ακτίνα του αντιστρόφου του εκτελώντας τρεις επαναλήψεις.

4.: Δίνεται ο πίνακας $A = \begin{bmatrix} 3 & 2 & 1 \\ 2 & 3 & 2 \\ 1 & 2 & 3 \end{bmatrix}$. Να προσεγγιστεί η ιδιοτιμή του A, που βρίσκεται κοντά στον αριθμό 3, με κάποια παραλλαγή της μεθόδου των δυνάμεων και με αρχικό διάνυσμα $x^{(0)} = [1\ 0\ 0]^T$, εκτελώντας τρεις επαναλήψεις.

5.: Δίνεται ότι ο πίνακας $A = \begin{bmatrix} 6 & -2 & 2 \\ -2 & 5 & 0 \\ 2 & 0 & 7 \end{bmatrix}$ έχει ως απόλυτα μεγαλύτερη ιδιοτιμή τη $\lambda_1 = 9$ με αντίστοιχο ιδιοδιάνυσμα το $x^{(1)} = [1\ -0.5\ 1]^T$. Χρησιμοποιώντας την αρχή της υποτίμησης να προσεγγιστεί η αμέσως μικρότερη σε απόλυτη τιμή ιδιοτιμή λ_2 και το αντίστοιχο ιδιοδιάνυσμα $x^{(2)}$, εκτελώντας τρεις επαναλήψεις με αρχικό διάνυσμα το $[1\ 1]^T$.

6.: Δίνεται ο πίνακας $A = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}$. Να προσεγγιστεί η ιδιοτιμή του A, που βρίσκεται κοντά στον αριθμό 0.5, με κάποια παραλλαγή της μεθόδου των δυνάμεων και με αρχικό διάνυσμα $[1\ 1\ -1]^T$, εκτελώντας τρεις επαναλήψεις.

7.: Δίνεται ο πίνακας $A = \begin{bmatrix} 2 & 1 & 0 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix}$. Να προσεγγιστεί η μεγαλύτερη ιδιοτιμή του A καθώς και το αντίστοιχο ιδιοδιάνυσμα, με τη μέθοδο των δυνάμεων και με αρχικό διάνυσμα

το $x^{(0)} = [0\ 1\ 0]^T$, χρησιμοποιώντας τη $||\cdot||_\infty$ και εκτελώντας τρεις επαναλήψεις.

8.: Αφού αποδειχτεί ότι ο πίνακας $A = \begin{bmatrix} 1 & -1 & 1 \\ -1 & 2 & -2 \\ 1 & -2 & 3 \end{bmatrix}$ είναι θετικά ορισμένος , να προσεγγιστεί η μικρότερη ιδιοτιμή του και το αντίστοιχο ιδιοδιάνυσμα, εκτελώντας τρεις επαναλήψεις με τη μέθοδο των αντίστροφων δυνάμεων και με αρχικό διάνυσμα $[0\ 1\ 0]^T$. Για τη λύση γραμμικών συστημάτων να χρησιμοποιηθεί η ανάλυση Cholesky διατηρώντας ακρίβεια ενός δεκαδικού ψηφίου στους ενδιάμεσους υπολογισμούς.

9.: Δίνεται ο πίνακας $A = \begin{bmatrix} 3 & 2 & 1 \\ 2 & 2 & 1 \\ 1 & 1 & 1 \end{bmatrix}$. Χρησιμοποιώντας τη μέθοδο των δυνάμεων με αρχικό διάνυσμα $x^{(0)} = [0\ 1\ 0]^T$ να προσεγγιστεί η απόλυτα μεγαλύτερη ιδιοτιμή του A και το αντίστοιχο ιδιοδιάνυσμα, εκτελώντας δύο επαναλήψεις.

10.: Δίνεται ο πίνακας $A = \begin{bmatrix} -3 & 4 & 1 \\ -2 & 2 & 2 \\ -1 & 1 & 1 \end{bmatrix}$. Χρησιμοποιώντας τη μέθοδο των δυνάμεων με αρχικό διάνυσμα $x^{(0)} = [0\ 1\ 0]^T$, να γίνουν τρεις επαναλήψεις για την εύρεση της απόλυτα μεγαλύτερης ιδιοτιμής και το αντίστοιχου ιδιοδιανύσματος. Στη συνέχεια να βρεθούν ολες οι ιδιοτιμές με τη μέθοδο του Krylov, με το ίδιο αρχικό διάνυσμα. (**Περιορισμός**: Να γίνουν ακριβείς πράξεις διατηρώντας ριζικά και κλάσματα στους υπολογισμούς.)

11.: Δίνεται ο πίνακας $A = \begin{bmatrix} 3 & -1 & 0 \\ -1 & 4 & -1 \\ 0 & -1 & 3 \end{bmatrix}$. Να προσεγγιστεί η πλησιέστερη προς τη μονάδα ιδιοτιμή του και το αντίστοιχο ιδιοδιάνυσμα, εκτελώντας τρεις επαναλήψεις με τη μέθοδο των δυνάμεων και με αρχικό διάνυσμα το $[0\ 1\ 0]^T$. Για τη λύση γραμμικών συστημάτων να χρησιμοποιηθεί η LU παραγοντοποίηση διατηρώντας ακρίβεια δύο δεκαδικών ψηφίων στους ενδιάμεσους υπολογισμούς.

12.: Δίνεται ο πίνακας $A = \begin{bmatrix} 3 & -1 & 0 \\ -2 & 3 & -2 \\ 0 & -1 & 3 \end{bmatrix}$. Χρησιμοποιώντας τους αλγόριθμους της μεθόδου των δυνάμεων με $||\cdot||_2$ και $||\cdot||_\infty$ και αρχικό διάνυσμα $x^{(0)} = [0\ 1\ 0]^T$, να προσεγγιστεί η απόλυτα μεγαλύτερη ιδιοτιμή του A και το αντίστοιχο ιδιοδιάνυσμα, εκτελώντας τρεις επαναλήψεις με τον καθένα. (**Περιορισμός**: Να γίνουν ακριβείς πράξεις διατηρώντας ριζικά και κλάσματα στους υπολογισμούς.)

13.: Αφού αποδειχτεί ότι ο πίνακας $A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix}$ είναι θετικά ορισμένος, να προσεγγιστεί η μικρότερη ιδιοτιμή του και το αντίστοιχο ιδιοδιάνυσμα, εκτελώντας τρεις επαναλήψεις με τη μέθοδο των αντίστροφων δυνάμεων και με αρχικό διάνυσμα το $[0\ 1\ 0]^T$. Για τη λύση γραμμικών συστημάτων να χρησιμοποιηθεί η ανάλυση Cholesky. (**Περιορισμός**: Να γίνουν ακριβείς πράξεις διατηρώντας ριζικά και κλάσματα στους υπολογισμούς.)

14.: Να γίνουν τρεις επαναλήψεις της QR μεθόδου για την προσέγγιση των ιδιοτιμών του πίνακα

$$A = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}.$$

Αναφορές

[1] O. Axelsson. *Iterative Solution Methods*. Cambridge University Press, Cambridge, England, 1994.

[2] R. Bellman. *Introduction to Matrix Analysis*, 2nd edition. Classics in Applied Mathematics 19, SIAM, Philadelphia, 1997.

[3] A. Berman, M. Neumann and R.J. Stern. *Nonnegative Matrices in Dynamic Systems*. John Wiley and Sons, New York, 1989.

[4] A. Berman and R.J. Plemmons. *Nonnegative Matrices in the Mathematical Sciences*. Classics in Applied Mathematics 9, SIAM, Philadelphia, 1994.

[5] G. Birkhoff and S. MacLane. *A Survey in Modern Algebra,* revised edition. The MacMillan Company, New York, 1953.

[6] E. Bodewig. *Matrix Calculus,* revised edition. Interscience Publishers, Inc., New York, 1959.

[7] W. Cheney and D. Kincaid. *Numerical Mathematics and Computing*, fourth edition. Brooks/Cole Publishing Company, Pacific Grove, CA, 1999.

[8] P. Ciarlet. *Introduction to Linear Algebra and Optimization*. Cambridge University Press, Cambridge, England, 1989.

[9] S.D. Conte and C. de Boor. *Elementary Numerical Analysis*, 2nd edition. McGraw-Hill Book Co., Inc., New York, 1972.

[10] G. Dahlquist and Å. Björk. *Numerical Methods*, translated by N. Anderson. Prentice-Hall, Englewood Cliffs, N.J., 1974.

[11] J.W. Demmel. *Applied Numerical Linear Algebra*. SIAM, Philadelphia, 1996.

[12] Β. Δουγαλής. Αριθμητική Λύση Γραμμικών Συστημάτων στον Υπολογιστή. Μαθηματικό Τμήμα Πανεπιστημίου Κρήτης και Ινστιτούτο Υπολογιστικών Μαθηματικών Ερευνητικού Κέντρου Κρήτης, Ηράκλειο, 1986.

[13] D.K. Fadeev and V.N. Fadeeva. *Computational Methods in Linear Algebra*. W.H. Freeman and Company, San Francisco, CA, 1963.

[14] L.V. Fausett. *Applied Numerical Analysis Using MATLAB*. Prentice-Hall, Upper Saddle River, N.J., 1999.

[15] G.E. Forsythe, M.A. Malcolm and C.B. Moler. *Computer Methods for Mathematical Computations*. Prentice-Hall, Englewood Cliffs, N.J., 1977.

[16] G.E. Forsythe and W.R. Wasow. *Finite Difference Methods for Partial Differential Equations*. John Wiley and Sons, Inc., New York, 1960.

[17] L. Fox. *Introduction to Numerical Linear Algebra*. Clarendon Press, Oxford, 1964.

[18] F.R. Gantmakher. *Matrix Theory*, 2 vols. Chelsea, New York, 1959.

[19] F.R. Gantmakher. *Applications of the Theory of Matrices*, translated and revised by J.L. Brenner. Interscience Publishers, Inc., New York, 1959.

[20] N. Gastinel. *Linear Numerical Analysis*. Academic Press, London, 1970.

[21] G.H. Golub and C.F. Van Loan. *Matrix Computations*. The John Hopkins University Press, Baltimore and London, 1989.

[22] A. Greenbaum. *Iterative Methods for Solving Linear Systems*. SIAM, Philadelphia, 1997.

[23] A. Hadjidimos. *Computational Methods in Analysis*. Lecture Notes, Department of Computer Sciences, Purdue University, West Lafayette, IN, 1988.

[24] L.A. Hageman and D.M. Young. *Applied Iterative Methods*. Academic Press, New York, 1981.

[25] P.R. Halmos. *Finite-Dimensional Vector Spaces*. Van Nostrand, Princeton, N.J., 1958.

[26] P. Henrici. *Applied and Computational Complex Analysis*, 3 vols. John Wiley and Sons, New York, 1974.

[27] R.A. Horn and C.R. Johnson. *Matrix Analysis*. Cambridge University Press, Cambridge, England, 1985.

[28] R.A. Horn and C.R. Johnson. *Topics in Matrix Analysis*. Cambridge University Press, Cambridge, England, 1991.

[29] A.S. Householder. *Principles in Numerical Analysis*. McGraw-Hill Book Co., Inc., New York, 1953.

[30] A.S. Householder. *The Theory of Matrices in Numerical Analysis*. Blaisdell, New York, 1965.

[31] C.T. Kelly. *Iterative Methods for Linear and Nonlinear Equations*. SIAM, Philadelphia, 1995.

[32] P. Lancaster. *Theory of Matrices*. Academic Press, New York, 1969. (2nd edition by P. Lancaster and M. Tismenetsky: 1985)

[33] C. D. Meyer Jr. *Matrix Analysis and Applied Linear Algebra*. SIAM, Philadelphia, 2000.

[34] B. Noble. *Applied Linear Algebra*. Prentice-Hall, Englewood Cliffs, N.J., 1969. (2nd edition by P. Noble and J.W. Daniel: 1997; 3rd edition 1988.)

[35] Δ. Νούτσος. Σημειώσεις από τις παραδόσεις του μαθήματος Αριθμητική Γραμμική Αλγεβρα. Τμήμα Μαθηματικών, Πανεπιστήμιο Ιωαννίνων, Ιωάννινα, 1998.

[36] J.M. Ortega. *Matrix Theory: A Second Course*. Plenum Press, New York, 1987.

[37] J.R. Rice. *Numerical Methods, Software and Analysis*. McGraw-Hill Book Co., Inc., New York, 1983.

[38] Y. Saad. *Iterative Methods for Sparse Linear Systems*. PWS Publishing Company, Boston, MA, 1996.

[39] R.V. Southwell. *Relaxation Methods in Theoretical Physics*. Clarendon Press, Oxford, 1946.

[40] G.W. Stewart. *Introduction to Matrix Computations*. Academic Press, New York, 1973.

[41] G. Strang. *Linear Algebra and its Applications*, 3rd edition. Harcourt Brace Jovanovich, San Diego, 1988.

[42] L.N. Trefethen and D. Bau III. *Numerical Linear Algebra*. SIAM, Philadelphia, 1997.

[43] R.S. Varga. *Matrix Iterative Analysis*, 2nd, revised and expanded, edition. Springer, Berlin, 2000.

[44] E.L. Wachspress. *Iterative Solution of Elliptic Systems*. Prentice-Hall, Inc., Englwood Cliffs, N.J., 1966.

[45] D.S. Watkins. *Fundamentals of Matrix Computations*. John Wiley and Sons, New York, 1991.

[46] J.H. Wilkinson. *The Algebraic Eigenvalue Problem*. Clarendon Press, Oxford, 1965.

[47] G. Williams. *Linear Algebra With Applications*, 4th edition. Jones and Bartlett Publishers, Sudbury, MA, 2001.

[48] D.M. Young. *Iterative Solution of Large Linear Systems*. Academic Press, New York, 1971.

[49] J.Y. Yuan. *Applied Iterative Analysis*. Lecture Notes, Departamento de Matemática, Universidade Federal do Paraná, Curitiba, Paraná, Brazil, 2002.