ΠΑΝΕΠΙΣΤΗΜΙΟ ΚΡΗΤΗΣ
ΤΜΗΜΑ ΜΑΘΗΜΑΤΙΚΩΝ

Μέθοδοι Ελαχιστοποίησης για την Επίλυση Αλγεβρικών Γραμμικών Συστημάτων και Εφαρμογή τους στην $p-$Κυκλική Περίπτωση

ΜΙΧΑΗΛ ΛΑΠΙΔΑΚΗΣ

Επιβλέπων Καθηγητής: ΑΠΟΣΤΟΛΟΣ ΧΑΤΖΗΔΗΜΟΣ

ΜΕΤΑΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ
Ηράκλειο, Οκτώβριος 2002.

Περιεχόμενα

1 Αντί Προλόγου

Πριν από την παρουσίαση της παρούσης εργασίας, θα ήθελα να ευχαριστήσω όσους βοήθησαν στην υλοποίησή της. Οφείλω, πρωτίστως, να ευχαριστήσω θερμά το δάσκαλό μου, Καθηγητή Α. Χατζηδήμο, με τη καθοδήγηση και τις πολύτιμες συμβουλές του οποίου, όχι μόνο ολοκλήρωσα τη μεταπτυχιακή μου εργασία αλλά, επιπλέον, κατανόησα βαθύτερες έννοιες του χώρου της Αριθμητικής Γραμμικής Άλγεβρας. Τον ευχαριστώ, ειλικρινά, για την ηθική συμπαράσταση και κατανόησή του καθ' όλη τη διάρκεια της συνεργασίας μας.

Θα ήθελα, επίσης, να ευχαριστήσω τα υπόλοιπα μέλη της τριμελούς επιτροπής: Τον κ. Μ. Βάβαλη (Αναπληρωτή Καθηγητή Μαθηματικών του Πανεπιστημίου Κρήτης) του οποίου η βοήθεια ήταν ιδιαίτερα σημαντική και πολύπλευρη. Τον κ. Δ. Νούτσο (Αναπληρωτή καθηγητή Μαθηματικών του Πανεπιστημίου Ιωαννίνων) οι παρατηρήσεις του οποίου συνέβαλαν στην διαμόρφωση του τελικού κειμένου της εργασίας. Επιπλέον, δε θα μπορούσα να ξεχάσω να ευχαριστήσω τους καθηγητές του Τμήματος Μαθηματικών του Πανεπιστημίου Κρήτης για τις γνώσεις που μου παρείχαν στη διάρκεια των μέχρι τώρα σπουδών μου.

Τέλος, θα ήθελα να ευχαριστήσω δύο οικεία μου πρόσωπα, την αδελφή μου Καλλιόπη Λαπιδάκη (Φιλόλογο) η οποία με υποστήριξε ηθικά και φιλολογικά, όπως επίσης, τη φίλη και συνάδελφο Ιακώβα Κοκκινάκη της οποίας η ηθική υποστήριξη ήταν ιδιαίτερα σημαντική.

2 Εισαγωγή

Σκοπός της παρούσης εργασίας είναι η επίλυση του αλγεβρικού γραμμικού συστήματος

$$Ax = b, \quad A \in \mathbb{C}^{n,n}, \ \det(A) \neq 0, \ b \in \mathbb{C}^n \backslash \{0\}. \tag{2.1}$$

Η επίλυση ενός τέτοιου συστήματος θα προέλθει μέσω μίας επαναληπτικής διαδικασίας διαδοχικών προσεγγίσεων. Εάν, π.χ., ως αρχική προσέγγιση της λύσης θεωρηθεί η $x_0 = 0$, τότε η νέα προσέγγιση $x_1 \in \text{span}\{b\}$, η επόμενη $x_2 \in \text{span}\{b, Ab\}$ κ.ο.κ. Με την ίδια διαδικασία, λαμβάνουμε νέες προσεγγίσεις x_k τέτοιες, ώστε

$$x_k \in span\{b, Ab, \ldots, A^{k-1}b\}, \ k = 1, 2, \ldots.$$

Χώροι της μορφής

$$\mathcal{K}_k(A, b) := span\{b, Ab, \ldots, A^{k-1}b\}, \ k = 1, 2, \ldots,$$

ονομάζονται "χώροι Krylov". Στην πράξη, οι παραπάνω προσεγγίσεις ίσως να μην είναι οι καλύτερες δυνατές. Για τη βελτίωσή τους, τροποποιούμε το αρχικό σύστημα λαμβάνοντας το εξής ισοδύναμο,

$$M^{-1}Ax = M^{-1}b.$$

Ο αντιστρέψιμος πίνακας $M \in \mathbb{C}^{n,n}$ καλείται προρρυθμιστής και, εάν ακολουθηθεί η προηγούμενη διαδικασία, τότε η προσεγγιστική λύση

$$x_k \in span\{M^{-1}b, M^{-1}AM^{-1}b, \ldots, (M^{-1}A)^{k-1}M^{-1}b\}.$$

Τα βασικότερα κριτήρια επιλογής του M είναι: Η επίλυση ενός συστήματος της μορφής $My = c$ να είναι σημαντικά "οικονομικότερη", σε πλήθος πράξεων, από αυτήν του $Ax = b$ και η σύγκλιση της ακολουθίας των $x_k, \ k = 0, 1, 2, \ldots$ να είναι όσο το δυνατό "ταχύτερη".

Στη συνέχεια, ακολουθεί μία σύντομη περιγραφή των περιεχομένων της παρούσης εργασίας. Αρχικά, στο δεύτερο Κεφάλαιο, γίνεται μία σύντομη ανασκόπηση βασικών αποτελεσμάτων της Γραμμικής Άλγεβρας και της Αριθμητικής Γραμμικής Άλγεβρας. Τα αποτελέσματα αυτά αφορούν στους ορισμούς εσωτερικών γινομένων, νορμών και συνθηκών σύγκλισης ακολουθίας διανυσμάτων. Επίσης, δίδονται σύντομες περιγραφές μεθόδων παραγοντοποίησης

πινάκων, όπως η "κανονική μορφή Jordan", η "παραγοντοποίηση Schur" και η "QR ανάλυση". Στο τελευταίο τμήμα του κεφαλαίου ορίζονται έννοιες, όπως εκείνη του "πεδίου τιμών" ενός πίνακα A, $\mathcal{F}(A)$ και της "αριθμητικής ακτίνας" $\nu(A)$, ενός πεδίου τιμών. Επίσης, παρουσιάζονται βασικές ιδιότητες και σχέσεις που αφορούν στις δύο παραπάνω έννοιες.

Στο τρίτο Κεφάλαιο, παρουσιάζονται οι λεγόμενες "μέθοδοι ελαχιστοποίησης" διανύσματος−υπολοίπου και διανύσματος−σφάλματος. Αρχικός σκοπός για την εισαγωγή αυτών των μεθόδων είναι η βελτίωση της "Απλής Επαναληπτικής Μεθόδου". Με τον όρο "Απλή Επαναληπτική Μέθοδος" προσδιορίζεται μία ακολουθία διαδοχικών προσεγγίσεων της λύσης, μέσω μίας επαναληπτικής διαδικασίας που εμπλέκει έναν πίνακα προρρύθμισης M, η επιλογή του οποίου οδηγεί στις βασικές επαναληπτικές μεθόδους Jacobi, Gauss–Seidel, SOR και άλλες. Έχοντας ως γνώμονα τη βελτίωση αυτής της μεθόδου, παρουσιάζουμε όλες εκείνες τις μεθόδους που αναπτύχθηκαν, κατά καιρούς, στην προσπάθεια να δοθούν καλύτερα αποτελέσματα από αυτά της "Απλής Επαναληπτικής Μεθόδου". Η βασική αρχή όλων αυτών των μεθόδων βασίζεται, όπως αναφέρθηκε προηγουμένως, στην ελαχιστοποίηση της "Ευκλείδειας" νόρμας του διανύσματος−υπολοίπου ($r_k = b - Ax_k$) ή της "A−νόρμας" του διανύσματος−σφάλματος ($e_k = A^{-1}b - x_k$), σε κάποιον χώρο Krylov της μορφής

$$span\{r_0, Ar_0, \ldots, A^{k-1}r_0\}. \tag{2.2}$$

Ξεκινώντας, λοιπόν, χρονολογικά από τις μεθόδους "Απότομης Καθόδου" και "Orthomin(1)", περιγράφουμε τη βασική μαθηματική θεωρία, δίδοντας, συγχρόνως, τις σχέσεις που αφορούν στα φράγματα για τις νόρμες σφαλμάτων και υπολοίπων αντίστοιχα. Στη συνέχεια, με βάση την εξέλιξη των μεθόδων, παρουσιάζουμε μεθόδους γενικότερες και " καλύτερες" αναφορικά με τον αριθμό των πράξεων αλλά και τη μνήμη που χρησιμοποιείται για την εύρεση της λύσης του συστήματος (2.1). Τέτοιες μέθοδοι είναι η Orthomin(j), Orthidir, MINRES, CG και η GMRES. Όπως ειπώθηκε προηγουμένως, οι παραπάνω μέθοδοι βασίζονται στην ελαχιστοποίηση της "Ευκλείδειας" νόρμας του διανύσματος−υπολοίπου ή της "A−νόρμας" του διανύσματος−σφάλματος στο χώρο Krylov (2.2). Η κατασκευή αλλά και η κανονικοποίηση της βάσης Krylov στηρίζεται στους Αλγορίθμους των Arnoldi και Lanczos. Εφαρμογή του δεύτερου αλγορίθμου γίνεται στην ειδική περίπτωση όπου ο πίνακας του αρχικού προρρυθμισμένου συστήματος είναι πραγματικός και συμμετρικός.

Στο τέταρτο Κεφάλαιο της εργασίας, παρουσιάζεται μία κατηγορία μεθόδων ελαχιστοποίησης της νόρμας των διανυσμάτων−υπολοίπων και των διανυσμά-

των−σφαλμάτων στην περίπτωση γενικού πίνακα A. Η βασική ιδέα αυτών των μεθόδων στηρίζεται σε μία υβριδική μορφή του Αλγορίθμου του Lanczos, που είναι γνωστή ως "Αμφίπλευρος Αλγόριθμος του Lanczos". Σε αυτή την κατηγορία ανήκουν μέθοδοι όπως οι BiCG, QMR, CGS και η BiCGSTAB.

Στις δύο προηγούμενες παραγράφους κατηγοριοποιήσαμε τις μεθόδους σύμφωνα με την αρχή κατασκευής τους. Η πρώτη κατηγορία μεθόδων στηρίζεται στους Αλγορίθμους των Arnoldi και Lanczos, ενώ η δεύτερη στον "Αμφίπλευρο Αλγόριθμο του Lanczos". Μπορούμε, όμως, να διακρίνουμε τις παραπάνω μεθόδους σε δύο ακόμα κατηγορίες, με κριτήριο διαχωρισμού τη θέση του διανύσματος−υπολοίπου σε σχέση με το χώρο Krylov (2.2). Σύμφωνα με αυτό, οι παραπάνω μεθόδοι διαχωρίζονται σε MR (Minimal Residual) (Ελαχίστου Υπολοίπου) και σε OR (Orthogonal Residual) (Ορθογώνιου Υπολοίπου). Στην πρώτη κατηγορία ανήκουν μέθοδοι στις οποίες το υπόλοιπο r_k ελαχιστοποιείται πάνω στο χώρο Krylov (2.2) και στη δεύτερη κατηγορία ανήκουν αυτές όπου το υπόλοιπο είναι ορθογώνιο στον ίδιο χώρο Krylov. Οι μέθοδοι Orthomin(1), Orthomin(j), Orthidir, MINRES, GMRES και QMR ανήκουν στην πρώτη κατηγορία, ενώ οι μέθοδοι "Απότομης Καθόδου", CG, BiCG, CGS και η BiCGSTAB ανήκουν στη δεύτερη κατηγορία.

Στα δύο τελευταία κεφάλαια αυτής της εργασίας εφαρμόζουμε τις μεθόδους ελαχιστοποίησης, ειδικότερα τους πυρήνες αυτών, δηλαδή, τους αλγορίθμους των Arnoldi, Lanczos και τον "Αμφίπλευρο Αλγόριθμο του Lanczos", σε μία κατηγορία πινάκων που καλούνται $p-$κυκλικοί. Σε αυτή την εργασία, θα περιορίσουμε την εφαρμογή μονάχα στην περίπτωση του Αλγορίθμου του Arnoldi, ο οποίος, βέβαια, είναι γενικότερος από αυτόν του Lanczos.

Συγκεκριμένα, στο πέμπτο Κεφάλαιο, γίνεται μία σύντομη παρουσίαση των βασικών αποτελεσμάτων που αφορούν στους $p-$κυκλικούς πίνακες, σύμφωνα πάντα με τους ορισμούς που δίδονται από το Varga [38]. Επίσης, στο τελευταίο μέρος του κεφαλαίου αυτού παρουσιάζονται κάποια αποτελέσματα που αφορούν σε ισοδύναμες μορφές του αρχικού συστήματος $Ax = b$, οι οποίες προκύπτουν από την ειδική μορφή του $p-$κυκλικού πίνακα A.

Στο έκτο Κεφάλαιο, δίδουμε ένα "μεταλλαγμένο" αλγόριθμο του Arnoldi για την περίπτωση των $p-$κυκλικών πινάκων, ο οποίος εκμεταλλεύεται πλήρως τη μορφή του πίνακα A. Επιπλέον, βλέπουμε ότι είναι σκόπιμο και εύχρηστο να υποθέσουμε ότι το αρχικό υπόλοιπο r_0 είναι χωρισμένο σε p blocks από τα οποία μόνο το πρώτο είναι μη-μηδενικό. Επίσης, παρατίθενται δύο θεωρήματα που μας επιτρέπουν να μιλήσουμε για την ισοδυναμία της εφαρμογής των μεθόδων ελαχιστοποίησης MR και OR στο αρχικό και το ανηγμένο σύστημα. Σημειώνουμε ότι "Ανηγμένο" καλείται το σύστημα που προκύπτει εκμετ-

αλλευόμενοι, κατά ένα συγκεκριμένο τρόπο, την ειδική μορφή του πίνακα A.

Στο έβδομο Κεφάλαιο δίδεται μια σύνοψη των περιεχομένων της παρούσης εργασίας και υποδεικνύονται δυνατές κατευθύνσεις για ερευνητική εκμετάλλευση των Μεθόδων Ελαχιστοποίησης σε συγκεκριμένες κατηγορίες πινάκων.

Τέλος, στο Παράρτημα, παρουσιάζονται αναλυτικά όλοι οι αλγόριθμοι των μεθόδων που περιγράφτηκαν και αναπτύχθηκαν στην παρούσα εργασία.

3 Στοιχεία Γραμμικής Άλγεβρας

3.1 Βασικοί ορισμοί

Θεωρούμε γνωστούς τους ορισμούς του "Ευκλείδειου εσωτερικού γινομένου"

$$(x,y)_2 := \sum_{i=1}^{n} \overline{x}_i y_i,\ x,y \in \mathbb{C}^n,$$

όπου $\overline{x}_i$ ο συζυγής μιγαδικός του x_i και της "διανυσματικής νόρμας" στον $\mathbb{C}^n$. Οι βασικές διανυσματικές νόρμες στον $\mathbb{C}^n$, στις οποίες κυρίως θα αναφερόμαστε, είναι οι εξής:

$$\begin{gathered}\|u\|_1 = \sum_{i=1}^{n} |u_i|\,, \\ \|u\|_2 = \left(\sum_{i=1}^{n} |u_i|^2\right)^{\frac{1}{2}} = (u,u)_2^{\frac{1}{2}}, \\ \|u\|_\infty = \max_{i=1,\ldots,n} |u_i|\,.\end{gathered}$$

Αν $G \in \mathbb{C}^{n,n}$ και G^H ο συζυγής ανάστροφος του G, τότε αν $u,v \in \mathbb{C}^n$, ο ορισμός

$$(u,v)_{G^HG} := (u, G^HGv)_2 = (Gu, Gv)_2$$

γενικεύει, προφανώς, τον ορισμό του Ευκλείδειου εσωτερικού γινομένου και καλείται "G^HG-εσωτερικό γινόμενο". Ο δε ορισμός

$$\|u\|_{G^HG} := (Gu, Gu)_2^{\frac{1}{2}} = (u,u)_{G^HG}^{\frac{1}{2}}, \tag{3.1}$$

μπορεί να αποδειχθεί ότι ορίζει μια διανυσματική νόρμα. Με βάση τον ορισμό του G^HG-εσωτερικού γινομένου, μπορούμε να ορίσουμε την G^HG-προβολή του $u \in \mathbb{C}^n$ στο $v \in \mathbb{C}^n$,

$$P^r_{G^HG}(u,v) := \frac{(u, G^HGv)_2}{(v, G^HGv)_2} v. \tag{3.2}$$

Παρατηρούμε ότι στην περίπτωση όπου $G^HG = I$, λαμβάνουμε τη γνωστή Ευκλείδεια προβολή.

Ορισμός 3.1. : Ένας πίνακας $Q \in \mathbb{R}^{n,n}$ καλείται ορθογώνιος ανν

$$Q^TQ = QQ^T = I.$$

Αν $Q \in \mathbb{C}^{n,n}$, τότε καλείται ορθοκανονικός ανν ισχύει

$$Q^HQ = QQ^H = I.$$

Θεωρώντας γνωστό τον ορισμό της νόρμας πίνακα στο $\mathbb{C}^{n,n}$, δίδουμε τον παρακάτω μερικότερο:

Ορισμός 3.2. : Έστω ότι $\|.\|$ είναι μία νόρμα διανύσματος στο $\mathbb{C}^n$. Η επαγόμενη ("φυσική") νόρμα πίνακα $A \in \mathbb{C}^{n,n}$ συμβολίζεται με $\|A\|$ και ορίζεται ως:

$$\|A\| := \sup_{x \in \mathbb{C}^n \setminus \{0\}} \frac{\|Ax\|}{\|x\|} = \max_{y \in \mathbb{C}^n,\, \|y\|=1} \|Ay\|.$$

Μπορεί να αποδειχθεί ότι οι τρεις φυσικές νόρμες πινάκων στο $\mathbb{C}^{n,n}$, που επάγονται από τις διανυσματικές νόρμες $\|.\|_1$, $\|.\|_2$, $\|.\|_\infty$, στο $\mathbb{C}^n$, είναι οι εξής:

$$\|A\|_1 = \max_j \sum_{i=1}^n |a_{ij}|,$$
$$\|A\|_2 = \rho^{\frac{1}{2}}(A^H A),$$
$$\|A\|_\infty = \max_i \sum_{j=1}^n |a_{ij}|,$$

όπου a_{ij} είναι τα στοιχεία του πίνακα A και $\rho(A)$ είναι η φασματική ακτίνα του.

Θεώρημα 3.1. *: Αν $\|.\|$ είναι η νόρμα ενός πίνακα και αν $G \in \mathbb{C}^{n,n}$ είναι αντιστρέψιμος, τότε η*

$$\|A\|_{G^H G} \equiv \left\|GAG^{-1}\right\|$$

ορίζει μία νόρμα πίνακα. Αν, επιπλέον, η $\|.\|$ είναι φυσική νόρμα, τότε η $\|.\|_{G^H G}$ επάγεται από τη διανυσματική νόρμα $\|.\|_{G^H G}$.

Θεώρημα 3.2. *: Αν $\|.\|$ είναι μια φυσική νόρμα πίνακα και $A \in \mathbb{C}^{n,n}$, τότε*

$$\rho(A) \leq \|A\|.$$

Θεώρημα 3.3. *: Έστω $A \in \mathbb{C}^{n,n}$ και $\epsilon > 0$. Υπάρχει φυσική νόρμα τ.ω.*

$$\|A\| \leq \rho(A) + \epsilon.$$

Θεώρημα 3.4. *: Έστω $A \in \mathbb{C}^{n,n}$. Τότε*

$$\lim_{k \longrightarrow \infty} A^k = O \Leftrightarrow \rho(A) < 1.$$

Πόρισμα 3.5. *: Έστω $\|.\|$ μια φυσική νόρμα. Τότε*

$$\rho(A) = \lim_{k \longrightarrow \infty} \left\|A^k\right\|^{1/k}, \forall A \in \mathbb{C}^{n,n}.$$

Ορισμός 3.3. : Δείκτης κατάστασης ενός πίνακα $A \in \mathbb{C}^{n,n}$ με $\det(A) \neq 0$, ως προς κάποια φυσική νόρμα $\|.\|$, καλείται ο πραγματικός αριθμός $\kappa(A) = \|A\| \, \|A^{-1}\|$.

Σημειώσεις: 1) Από τον ορισμό του δείκτη κατάστασης προκύπτει ότι $\kappa(A) = \|A\| \, \|A^{-1}\| \geq 1$. 2) Στην ειδική περίπτωση που η νόρμα του πίνακα A είναι η "Ευκλείδεια", ο δείκτης κατάστασης λέγεται "Ευκλείδειος δείκτης κατάστασης του A" ή "φασματικός δείκτης κατάστασης του A ως προς την αντιστροφή (inversion)" και συμβολίζεται με $\kappa_2(A)$. Επιπλέον, όταν $A = A^H$, τότε $\kappa_2(A) = \frac{|\lambda_{\max}|}{|\lambda_{\min}|}$, όπου $\lambda_{max}, \lambda_{min}$ αντιστοιχούν στην απόλυτα μέγιστη και την απόλυτα ελάχιστη ιδιοτιμή του πίνακα A[1]. 3) Ο ρόλος του δείκτη κατάστασης στη σύγκλιση των μεθόδων επίλυσης γραμμικών συστημάτων (2.1) είναι ιδιαίτερα σημαντικός για τους εξής παρακάτω λόγους: i) Όσο μικρότερος είναι ο δείκτης κατάστασης ενός πίνακα, τόσο ταχύτερη είναι η σύγκλιση των μεθόδων που χρησιμοποιούμε για την εύρεση της λύσης του (2.1). ii) Στις περιπτώσεις αριθμητικής πεπερασμένης ακρίβειας ισχύει, γενικά, ότι όσο μικρότερος είναι ο δείκτης κατάστασης, τόσο μικρότερο είναι το φράγμα για το σχετικό απόλυτο σφάλμα της λύσης. iii) Ένα σύστημα θεωρείται "καλής κατάστασης" όταν "μικρές" μεταβολές στα στοιχεία του πίνακα A ή/και του διανύσματος b επιφέρουν "ασήμαντες" μεταβολές στη λύση του συστήματος. Διαφορετικά, το σύστημα χαρακτηρίζεται ως "κακής κατάστασης". Για τη βελτίωση τέτοιων καταστάσεων, χρησιμοποιούνται αλγόριθμοι που ονομάζονται συνήθως "Αλγόριθμοι Επαναληπτικής Βελτίωσης". Για μία πλήρη ανάλυση όλων αυτών των αποτελεσμάτων, μπορεί ο αναγνώστης να καταφύγει στο βιβλίο του Wilkinson [39].

3.2 Κανονικές Μορφές και Παραγοντοποιήσεις.

Κανονική Μορφή Jordan: Έστω $A \in \mathbb{C}^{n,n}$. Τότε υπάρχει αντιστρέψιμος πίνακας $S \in \mathbb{C}^{n,n}$ τ.ω.

$$A = SJS^{-1}, \quad J = diag(J_{n_1}(\lambda_1), \ldots, J_{n_m}(\lambda_m)), \tag{3.3}$$

[1] Στη γενική περίπτωση πίνακα ο "φασματικός" δείκτης κατάστασης δίδεται από τον τύπο $\kappa_2(A) = \frac{\sigma_{\max}}{\sigma_{\min}}$, με $\sigma_{\max}, \sigma_{\min}$ να είναι η μέγιστη και η ελάχιστη "ιδιάζουσα τιμή", αντίστοιχα. Ο ορισμός της "ιδιάζουσας τιμής" θα δοθεί παρακάτω.

όπου $J_{n_i}(\lambda_i) \in \mathbb{C}^{n_i,n_i}$ είναι άνω τριγωνικός πίνακας της μορφής:

$$J_{n_i}(\lambda_i) = \begin{pmatrix} \lambda_i & 1 & & & \\ & \lambda_i & 1 & & \\ & & \ddots & \ddots & \\ & & & \lambda_i & 1 \\ & & & & \lambda_i \end{pmatrix}, \quad i = 1(1)n_i, \tag{3.4}$$

με $\sum_{i=1}^{m} n_i = n$. Η έκφραση του A στις (3.3)–(3.4) καλείται "κανονική μορφή Jordan" του πίνακα A. Ο πίνακας S έχει ως στήλες τα ιδιοδιανύσματα, καθώς και τα "γενικευμένα" ιδιοδιανύσματα του πίνακα A.

Ορισμός 3.4. : Ένας πίνακας A λέγεται διαγωνοποιήσιμος ανν στις (3.3)–(3.4) $m = n$. Δηλαδή, ο πίνακας S έχει ως στήλες του μόνο τα n (γραμμικά ανεξάρτητα) ιδιοδιανύσματα του A.

Ορισμός 3.5. : Ένας πίνακας A είναι "κανονικός" (normal,) εάν μπορεί να γραφεί στη μορφή

$$A = Q\Lambda Q^H,$$

όπου Λ είναι ένας διαγώνιος και Q ένας ορθοκανονικός πίνακας.

Πρόταση 3.6. *: Ένας πίνακας A είναι κανονικός ανν αντιμετατίθεται με τον συζυγή ανάστροφό του. Δηλαδή, αν*

$$AA^H = A^H A.$$

(Σημείωση: Προφανώς, κάθε Ερμιτιανός πίνακας ($A^H = A$) και κάθε αντι-Ερμιτιανός ($A^H = -A$) είναι κανονικός.)

Μπορούμε να αποδείξουμε επαγωγικά ότι η k δύναμη ενός Jordan block είναι:

$$J_{n_l}^k(\lambda_l) = \{d_{i,j}^{(k)}(\lambda_l)\},\ 1 \le i, j \le n_l,$$

όπου

$$d_{i,j}^{(k)}(\lambda_l) = \begin{cases} 0, & j < i \\ \binom{k}{j-1} \lambda_l^{k-j+i}, & i \le j \le \min(n_l, k+i) \\ 0, & k+i < j \le n_l \end{cases}$$

Μέσω των παραπάνω εκφράσεων και της κανονικής μορφής Jordan μπορεί να αποδειχθεί το Πόρισμα 3.5.

Θεώρημα 3.7. *: (Μορφή Schur): Έστω $A \in \mathbb{C}^{n,n}$ με ιδιοτιμές $\lambda_1, \ldots, \lambda_n$. Τότε υπάρχει ορθοκανονικός πίνακας Q τ.ω.*

$$A = QUQ^H,$$

όπου U είναι ένας άνω τριγωνικός πίνακας και $u_{ii} = \lambda_i$.

Παρατηρούμε ότι ο πίνακας S της κανονικής μορφής Jordan, (3.3) μπορεί να είναι κακής κατάστασης (δηλαδή, $\kappa(S) = \|S\| \, \|S^{-1}\| \gg 1$). Αντίθετα, όμως, ο μετασχηματισμός στην "άνω τριγωνική" μορφή (Schur) μπορεί να είναι εξαιρετικά "καλής κατάστασης", αφού

$$\kappa(U) = \|U\| \left\|U^{-1}\right\| = \left\|Q^H A^{-1} Q\right\| \left\|Q^H A Q\right\| = \|A\| \left\|A^{-1}\right\| = \kappa(A).$$

Μπορεί να αποδειχθεί ότι η μορφή Schur δεν είναι μοναδική (βλ. [20], [22]).

Θεώρημα 3.8. *(LU παραγοντοποίηση): Έστω $A \in \mathbb{C}^{n,n}$ αντιστρέψιμος. Τότε, ο A μπορεί να παραγοντοποιηθεί στη μορφή*

$$A = PLU,$$

όπου P είναι ένας πίνακας μετάθεσης, L είναι ένας κάτω τριγωνικός με μονάδες στη διαγώνιο και U είναι ένας άνω τριγωνικός πίνακας. Όταν ο A είναι Ερμιτιανός και επιπλέον θετικά ορισμένος (δηλαδή, $(x, Ax)_2 > 0, \ \forall x \in \mathbb{C}^n \backslash \{0\}$), τότε η παραγοντοποίηση του A μπορεί να πάρει και τη μορφή [2]

$$A = LL^H,$$

όπου τα διαγώνια στοιχεία του (μοναδικού) κάτω τριγωνικού πίνακα L είναι θετικά.

Θεώρημα 3.9. *(QR ανάλυση): Έστω $A \in \mathbb{C}^{n,m}$, με $m \leq n$. Υπάρχει πίνακας $Q \in \mathbb{C}^{n,m}$, με διανύσματα–στήλες ανά δύο ορθοκανονικά, και άνω τριγωνικός πίνακας $R \in \mathbb{C}^{m,m}$ τ.ω.*

$$A = QR.$$

Επιπλέον, μπορούν να προστεθούν $n - m$ στήλες στον πίνακα Q, έτσι ώστε να γίνει ορθοκανονικός πίνακας Q' τ.ω.

$$A = Q'R'$$

με τον $R' \in \mathbb{C}^{n,m}$ πίνακα να έχει τον R ως $m \times m$ κύριο υποπίνακα και οπουδήποτε αλλού μηδενικά.

[2] Παραγοντοποίηση Cholesky.

Για την κατασκευή ορθοκανονικών διανυσμάτων υπάρχουν διάφοροι αλγόριθμοι οι οποίοι μπορούν να χρησιμοποιηθούν. Ο γνωστότερος αλγόριθμος κατασκευής ορθοκανονικών διανυσμάτων $u_j \in \mathbb{C}^n$, $\|u_j\|_2 = 1\,,\ j = 1, \ldots, k \leq n$, από ένα σύνολο γραμμικώς ανεξάρτητων διανυσμάτων v_j, $j = 1, \ldots, k \leq n$, είναι ο αλγόριθμος των Gram–Schmidt. Σε πολλές περιπτώσεις, και κυρίως για λόγους ευστάθειας, χρησιμοποιείται ο "τροποποιημένος" αλγόριθμος των Gram–Schmidt, ο οποίος παρουσιάζεται στο Παράρτημα.

Υπάρχουν διάφοροι τρόποι να πετύχουμε την QR ανάλυση του $A \in \mathbb{C}^{n,m}$, $m \leq n$. Όπως:

1. (Τροποποιημένος) Αλγόριθμος Gram–Schmidt.
2. Ανακλάσεις (Μετασχηματισμοί) Householder.
3. Στροφές (Μετασχηματισμοί) Givens.

Ανακλάσεις Householder: Ένας ορθοκανονικός πίνακας $P \in \mathbb{C}^{n,n}$ ορίζεται ως ανάκλαση Householder αν

$$\exists u \in \mathbb{C}^n \,\mu\varepsilon\, \|u\|_2 = 1$$

τ.ω. $P = I - 2uu^H$. Οι ανακλάσεις Householder εφαρμόζονται στην QR ανάλυση με σκοπό την απαλοιφή των κάτω από τη "διαγώνιο" στοιχείων του πίνακα A. Συγκεκριμένα, αν y είναι το διάνυσμα της πρώτης στήλης του A και ξ^1 το μοναδιαίο διάνυσμα με όλες τις συνιστώσες μηδέν εκτός της πρώτης ($\xi^1 = (1, 0, \ldots, 0)^T$), τότε θεωρούμε

$$P_1 y = (I - 2uu^H)y = c\xi^1 \ \Leftrightarrow\ 2uu^H y = y - c\xi^1 \ \Leftrightarrow\ 2u^H yu = y - c\xi^1.$$

Παίρνοντας νόρμες στις παραπάνω σχέσεις έχουμε:

$$\|P_1 y\|_2 = \left\|c\xi^1\right\|_2 \Rightarrow \|y\|_2 = |c| \Rightarrow c = \pm \|y\|_2 .$$

Ορίζουμε, λοιπόν,

$$u' = y \pm \|y\|_2 \xi^1,\ u = \frac{u'}{\|u'\|_2}$$

και εφαρμόζοντας τον P_1 στον A, λαμβάνουμε τον πίνακα:

$$P_1 A \ = \ \begin{pmatrix} a_{11}^{(1)} & a_{12}^{(1)} & \ldots & a_{1m}^{(1)} \\ 0 & a_{22}^{(1)} & \ldots & a_{2m}^{(1)} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & a_{n2}^{(1)} & \ldots & a_{nm}^{(1)} \end{pmatrix}.$$

Έπειτα, μηδενίζουμε τα κάτω από τη διαγώνιο στοιχεία του κάτω δεξιά $(n-1)\times(m-1)$ υποπίνακα του A, επαναλαμβάνοντας την παραπάνω διαδικασία $m-1$ φορές. Συγκεκριμένα πολλαπλασιάζοντας διαδοχικά από τα αριστερά τον A επί $P_1,\ P_2,\ \ldots,\ P_m$, λαμβάνουμε έναν πίνακα με μηδενικά κάτω από την κύρια διαγώνιο, δηλαδή:

$$A_m = P_m P_{m-1} \ldots P_1 A.$$

Επομένως

$$A = P_1^H P_2^H \ldots P_m^H A_m.$$

Ο πίνακας Q αποτελείται από τις m πρώτες στήλες του

$$P = P_1^H P_2^H \ldots P_m^H$$

και ο R είναι ο άνω τριγωνικός $m \times m$ υποπίνακας του A_m.

Στροφές Givens: Ένας άλλος τρόπος QR παραγοντοποίησης είναι οι στροφές Givens, δηλαδή, η εφαρμογή στον πίνακα A ενός πίνακα της μορφής:

$$F_{jk}(\theta) = \begin{pmatrix} 1 & & & & & & & & \\ & 1 & & & & & & & \\ & & \ddots & & & & & & \\ & & & c & & & s & & \\ & & & & \ddots & & & & \\ & & & & & 1 & & & \\ & & & & & & \ddots & & \\ & & & -\overline{s} & & & c & & \\ & & & & & & & \ddots & \\ & & & & & & & & 1 \end{pmatrix},$$

με

$$c = \cos(\theta),\ s = \sin(\theta),$$

όπου η γωνία θ μπορεί να είναι μιγαδική. Η εφαρμογή ενός τέτοιου πίνακα σε ένα διάνυσμα στρέφει την (j,k) συνιστώσα του κατά γωνία θ. Σκοπός μας, και σε αυτή την περίπτωση, είναι ο μηδενισμός των στοιχείων κάτω από τη διαγώνιο του πίνακα A. Αν, λοιπόν, θέλουμε να μηδενίσουμε το (k,j) στοιχείο του A, εφαρμόζουμε στον A τον πίνακα–στροφή $F_{jk}(\theta)$. Παρατηρούμε, τότε, ότι για το μηδενισμό του (k,j) στοιχείου πρέπει

$$\cos(\theta) = \frac{|a_{jj}|}{\sqrt{|a_{kj}|^2 + |a_{jj}|^2}},\ \overline{\sin(\theta)} = \frac{a_{kj}}{a_{jj}}\cos(\theta),\ a_{jj} \neq 0, \qquad (3.5)$$

και

$$\overline{\sin(\theta)} = 1 \ \alpha\nu \ \cos(\theta) = 0.$$

Επιπλέον, το νέο στοιχείο a'_{jj} θα δίδεται από τη σχέση

$$a'_{jj} = a_{jj}\cos(\theta) + a_{kj}\sin(\theta) = \frac{\alpha_{jj}}{|\alpha_{jj}|}\sqrt{|a_{kj}|^2 + |a_{jj}|^2}. \tag{3.6}$$

Τα υπόλοιπα στοιχεία της j γραμμής θα δίδονται από τις σχέσεις

$$a'_{ji} = a_{ji}\cos(\theta) + a_{ki}\sin(\theta),\ i = j + 1(1)m. \tag{3.7}$$

Με παρόμοιο τρόπο, τα στοιχεία της k γραμμής θα δίδονται από τους τύπους

$$a'_{ki} = -a_{ji}\overline{\sin(\theta)} + a_{ki}\cos(\theta),\ i = j + 1(1)m. \tag{3.8}$$

Ακολουθώντας την παραπάνω διαδικασία για κάθε στοιχείο α_{ij}, $i = 1, \ldots, n$, $j = 1, \ldots, m$, $j < i$, του A λαμβάνουμε τον πίνακα

$$A_m = F_{mn} \ldots F_{12} A,$$

όπου τα κάτω από την διαγώνιο στοιχεία του είναι μηδενικά. Ο Q αποτελείται από τις πρώτες m στήλες του πίνακα

$$P = F_{12}^H \ldots F_{mn}^H$$

ενώ ο R είναι ο άνω τριγωνικός $m \times m$ υποπίνακας του A_m.

Θεώρημα 3.10. *(Ανάλυση Ιδιαζουσών Τιμών (Singular Value Decomposition (SVD))): Αν $A \in \mathbb{C}^{m,n}$ με rank$(A) = k$, τότε ο A μπορεί να γραφεί στη μορφή*

$$A = V\Sigma W^H,$$

όπου $V \in \mathbb{C}^{m,m}$, $W \in \mathbb{C}^{n,n}$ ορθοκανονικοί πίνακες και $\Sigma \in \mathbb{C}^{m,n}$ "διαγώνιος" πίνακας με

$$\sigma_{ij} = 0 \ \forall i \neq j \ \kappa\alpha\iota \ \sigma_{11} \geq \sigma_{22} \geq \ldots \geq \sigma_{kk} > \sigma_{k+1,k+1} = \ldots = \sigma_{qq} = 0,$$

όπου $q = \min(m, n)$.

Οι αριθμοί $\sigma_i = \sigma_{ii}$, $i = 1, 2, \ldots, q$ ονομάζονται "ιδιάζουσες τιμές" (singular values) του A και είναι οι μη-αρνητικές τετραγωνικές ρίζες των ιδιοτιμών του AA^H. Οι στήλες του V καλούνται "αριστερά ιδιάζοντα διανύσματα" (singular vectors) του A και είναι τα ιδιοδιανύσματα του AA^H. Οι στήλες του W καλούνται "δεξιά ιδιάζοντα διανύσματα" και είναι τα ιδιοδιανύσματα του πίνακα $A^H A$.

Θεώρημα 3.11. *(Παραγοντοποίηση Schur): Έστω ότι ο $A \in \mathbb{C}^{n,n}$ έχει ιδιοτιμές $\lambda_1, \ldots, \lambda_n$ και ιδιάζουσες τιμές $\sigma_1, \ldots, \sigma_n$ και έστω ότι*

$$A = QUQ^H$$

είναι η παραγοντοποίηση Schur του A. Έστω Λ ο πίνακας με διαγώνια στοιχεία τα διαγώνια στοιχεία του U, δηλαδή, τις ιδιοτιμές του A, και έστω T το αυστηρά άνω τριγωνικό τμήμα του U. Τότε

$$\|A\|_F^2 = \sum_{i=1}^{n} \sigma_i^2 = \|\Lambda\|_F^2 + \|T\|_F^2 .$$

Σημείωση: Το σύμβολο $\|A\|_F$ δηλώνει τη νόρμα του Frobenius του A και ορίζεται ως

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

(Η νόρμα του Frobenius **δεν** είναι φυσική.)

Για περισσότερα πάνω στις αναλύσεις και παραγοντοποιήσεις πινάκων βλ. [5], [6], [22].

3.3 Ιδιοτιμές και Πεδίο Τιμών.

Στις περιπτώσεις κανονικών πινάκων $A \in \mathbb{C}^{n,n}$ οι ιδιοτιμές τους μας δίδουν όλες τις πληροφορίες για την επίλυση του αρχικού γραμμικού συστήματος. Στην περίπτωση όπου ο πίνακας A δεν είναι κανονικός, τις αντίστοιχες πληροφορίες τις παίρνουμε από ένα άλλο σύνολο τιμών, που καλείται "πεδίο τιμών" του πίνακα A.

Θεώρημα 3.12. *(Gerschgorin): Έστω $A \in \mathbb{C}^{n,n}$ με*

$$R_i(A) = \sum_{j=1, i \neq j}^{n} |a_{ij}| , i = 1, \ldots, n.$$

Τότε όλες οι ιδιοτιμές του A βρίσκονται στην ένωση των δίσκων

$$\bigcup_{i=1}^{n} \{ z \in \mathbb{C} : |z - a_{ii}| \leq R_i(A) \} .$$

Πόρισμα 3.13. *: Έστω $A \in \mathbb{C}^{n,n}$ και έστω*

$$C_j(A) = \sum_{i=1, i\neq j}^{n} |a_{ij}|\,,\ j = 1, \ldots, n.$$

Τότε όλες οι ιδιοτιμές του A ανήκουν στην ένωση των δίσκων

$$\bigcup_{i=1}^{n} \{z \in \mathbb{C} : |z - a_{jj}| \leq C_j(A)\}\,.$$

Ορισμός 3.6. : Ένας πίνακας $A \in \mathbb{C}^{n,n}$ λέγεται "αυστηρά διαγώνια υπέρτερος κατά γραμμές", αν

$$|a_{ii}| > \sum_{j=1, i\neq j}^{n} |a_{ij}|\,,\ i = 1, \ldots, n.$$

(Σημείωση: Ανάλογα, ορίζεται και ένας "αυστηρά διαγώνια υπέρτερος κατά στήλες" πίνακας.)
Ένας "αυστηρά διαγώνια υπέρτερος κατά γραμμές (ή στήλες)" πίνακας είναι αντιστρέψιμος[3]. Στην περίπτωση που η προηγούμενη σχέση παρουσιάζεται με ισότητα, δηλαδή,

$$|a_{ii}| \geq \sum_{j=1, i\neq j}^{n} |a_{ij}|\,,\ i = 1, \ldots, n,$$

ο πίνακας καλείται "ασθενώς διαγώνια υπέρτερος κατά γραμμές".

Ορισμός 3.7. : Ένας Ερμιτιανός πίνακας $A \in \mathbb{C}^{n,n}$ καλείται θετικά ορισμένος ανν $\forall x \in \mathbb{C}^n \backslash \{0\}$ συνεπάγεται $(x, Ax)_2 > 0$. Ισοδύναμα, όταν όλες οι ιδιοτιμές του A είναι γνήσια θετικές[4].

Πρόταση 3.14. *: Ένας αυστηρά διαγώνια υπέρτερος Ερμιτιανός πίνακας A με θετικά διαγώνια στοιχεία είναι θετικά ορισμένος.*

Ορισμός 3.8. : Ένας Ερμιτιανός πίνακας $A \in \mathbb{C}^{n,n}$ καλείται θετικά ημιορισμένος ανν $\forall x \in \mathbb{C}^n \backslash \{0\}$ συνεπάγεται $(x, Ax)_2 \geq 0$. Ισοδύναμα, όταν όλες οι ιδιοτιμές του A είναι πραγματικές μη-αρνητικές.

[3]Εφαρμογή των δίσκων Gerschgorin.

[4]Είναι γνωστό ότι ένας πραγματικός συμμετρικός πίνακας έχει πραγματικές, ιδιοτιμές.

Ορισμός 3.9. : Πεδίο τιμών του $A \in \mathbb{C}^{n,n}$ είναι το σύνολο

$$\mathcal{F}(A) = \left\{ y^H A y : y \in \mathbb{C}^n, y^H y = 1 \right\}. \tag{3.9}$$

Το σύνολο $\mathcal{F}(A)$ καλείται και "αριθμητική περιοχή". Ένας ισοδύναμος ορισμός είναι και ο εξής:

$$\mathcal{F}(A) = \left\{ \frac{y^H A y}{y^H y} : y \in \mathbb{C}^n, y \neq 0 \right\}.$$

Παρατήρηση 3.1. : Το πεδίο τιμών είναι συμπαγές υποσύνολο του $\mathbb{C}$, αφού είναι η εικόνα της μοναδιαίας Ευκλείδειας μπάλας, μέσω της συνεχούς απεικόνισης που ορίζεται στη (3.9). Επίσης, μπορεί να αποδειχθεί ότι είναι και κυρτό [22],[23].

Ορισμός 3.10. : "Αριθμητική ακτίνα" πίνακα $A \in \mathbb{C}^n$, $\nu(A)$, είναι η μεγαλύτερη απόλυτη τιμή των στοιχείων του $\mathcal{F}(A)$, δηλαδή,

$$\nu(A) \equiv \max \{ |z| : z \in \mathcal{F}(A) \}.$$

Αν $A \in \mathbb{C}^{n,n}$ και $\alpha \in \mathbb{C}$ τότε

$$\begin{aligned} &1.\ \mathcal{F}(A + \alpha I) = \mathcal{F}(A) + \alpha \\ &2.\ \mathcal{F}(\alpha A) = \alpha \mathcal{F}(A) \end{aligned} \tag{3.10}$$

Παρατήρηση 3.2. : 1) Για κάθε $A \in \mathbb{C}^{n,n}$ το $\mathcal{F}(A)$ περιέχει τις ιδιοτιμές του A. 2) Αν $Q \in \mathbb{C}^{n,n}$ είναι ορθοκανονικός, τότε $\mathcal{F}(Q^H A Q) = \mathcal{F}(A)$. 3) Για κανονικούς πίνακες το πεδίο τιμών είναι η κυρτή θήκη του φάσματος των ιδιοτιμών τους.

Έστω $A \in \mathbb{C}^{n,n}$ και $H(A) = \frac{1}{2}(A + A^H)$ η Ερμιτιανή συνιστώσα του $\left(A = \frac{1}{2}(A + A^H) + \frac{1}{2}(A - A^H)\right)$. Τότε έχουμε ότι $\mathcal{F}(H(A)) = Re(\mathcal{F}(A))$.

Θεώρημα 3.15. *: Έστω $A \in \mathbb{C}^{n,n}$ και έστω $R_i(A) = \sum_{j=1, j\neq i}^{n} |a_{ij}|$, $i = 1, \ldots, n$, και $C_j(A) = \sum_{i=1, i\neq j}^{n} |a_{ij}|$, $j = 1, \ldots, n$. Τότε το πεδίο τιμών του A ανήκει στο σύνολο*

$$C_o \left(\bigcup_{i=1}^{n} \left\{ z \in \mathbb{C} : |z - a_{ii}| \leq \frac{1}{2}(R_i(A) + C_i(A)) \right\} \right),$$

όπου με C_o συμβολίζουμε την κυρτή θήκη του συνόλου.

Παρατήρηση 3.3. : Η απόδειξη αυτού του θεωρήματος δίδει και μια διαδικασία αριθμητικής προσέγγισης του πεδίου τιμών.

Επιστρέφουμε πάλι στην έννοια της αριθμητικής ακτίνας, δίδοντας κάποιες σημαντικές ιδιότητες.

1. $\nu(A+B) = \max_{\|y\|_2=1} \left|y^H(A+B)y\right| \leq \max_{\|y\|_2=1} \left|y^HAy\right| + \max_{\|y\|_2=1} \left|y^HBy\right| \leq \nu(A)+\nu(B)$.

2. Η αριθμητική ακτίνα δεν ορίζει νόρμα πίνακα, αφού δεν είναι εν γένει πολλαπλασιαστική. Δηλαδή, $\nu(AB) \not\leq \nu(A)\nu(B)$.

3. $$\frac{1}{2}\|A\|_2 \leq \nu(A) \leq \|A\|_2 . \tag{3.11}$$

4. $\nu(A^m) \leq (\nu(A))^m$, $m = 1, 2, \ldots$.

Πρόταση 3.16. *: Αν $\sigma(A)$ είναι το φάσμα των ιδιοτιμών του A και P είναι ένα πολυώνυμο με μιγαδικούς γενικά συντελεστές, τότε $\sigma(P(A)) = P(\sigma(A))$.*

Παρατήρηση 3.4. : Στην περίπτωση του πεδίου τιμών δεν υπάρχει ανάλογη πρόταση, αφού υπάρχουν περιπτώσεις όπου $\mathcal{F}(P(A)) \neq P(\mathcal{F}(A))$.

Πρόταση 3.17. *: Αν το πεδίο τιμών του A περιέχεται σε δίσκο κέντρου 0 και ακτίνας r, τότε το πεδίο τιμών του A^m περιέχεται σε δίσκο κέντρου 0 και ακτίνας r^m.*

Για τις αποδείξεις όλων των παραπάνω προτάσεων, ιδιοτήτων και παρατηρήσεων βλ. [1], [22], [23].

4 Μέθοδοι Ελαχιστοποίησης

Στο κεφάλαιο αυτό θα αναφερθούμε στις κύριες μεθόδους ελαχιστοποίησης διανύσματος−υπολοίπου ή διανύσματος−σφάλματος για Ερμιτιανούς και θετικά ορισμένους πίνακες, μίας επαναληπτικής μεθόδου. Πρωτίστως, όμως, θα προσπαθήσουμε να συνδέσουμε αυτές με τις βασικές επαναληπτικές μεθόδους, δίδοντας μία γενική μέθοδο διαδοχικών προσεγγίσεων της λύσης.

4.1 Μέθοδος της Απλής Επανάληψης

Η μέθοδος της "Απλής Επανάληψης" ή "Απλή Επαναληπτική", όπως καλείται, μπορεί να θεωρηθεί στην παρακάτω γενική μορφή των διαδοχικών προσεγγίσεων της λύσης

$$x_{k+1} = x_k + M^{-1}(b - Ax_k), \quad \forall\, x_0 \in \mathbb{C}^n,\ k = 0, 1, 2, \ldots, \tag{4.1}$$

όπου ο πίνακας M αναφέρεται στον προρρυθμιστή του αρχικού συστήματος $Ax = b$. Ανάλογα με την επιλογή του προρρυθμιστή, λαμβάνουμε τις διάφορες απλές επαναληπτικές μεθόδους, μεταξύ των οποίων συγκαταλέγονται και οι λεγόμενες "κλασικές". Έτσι, αν $M = D$ όπου $D = \mathrm{diag}(a_{11}, a_{22}, \ldots, a_{nn})$, με $\det(D) \neq 0$, τότε παίρνουμε τη σημειακή μέθοδο του Jacobi. Θεωρώντας κατάλληλο πίνακα M, μπορούμε να πάρουμε μεθόδους όπως η Gauss–Seidel, η SOR, κ.λπ. Ο αλγόριθμος της μεθόδου της Απλής Επανάληψης (4.1) δίδεται πλήρως στο Παράρτημα. Για τη μελέτη των σφαλμάτων και των ασυμπτωτικών ή μη ασυμπτωτικών, ταχυτήτων σύγκλισης των παραπάνω μεθόδων, ο αναγνώστης παραπέμπεται σε οποιοδήποτε βιβλίο που έχει σχέση με Αριθμητική Γραμμική Άλγεβρα (βλ., π.χ., [38], [40], [3]).

Το βασικό θεώρημα σύγκλισης της μεθόδου της Απλής Επανάληψης είναι το εξής:

Θεώρημα 4.1. *: Η Απλή Επαναληπτική μέθοδος (4.1) συγκλίνει στη λύση του αρχικού συστήματος (2.1) , $A^{-1}b$, για κάθε αρχική προσέγγιση $x_0 \in \mathbb{C}^n$ αν και μόνο αν $\rho(I - M^{-1}A) < 1$.*

Από την (4.1), θεωρώντας το διάνυσμα–σφάλμα

$$e_k = A^{-1}b - x_k,$$

λαμβάνουμε τη σχέση

$$e_k = (I - M^{-1}A)e_{k-1},$$

που συνδέει τα διαδοχικά σφάλματα της μεθόδου. Ομοίως, θεωρώντας το διάνυσμα–υπόλοιπο

$$r_k = b - Ax_k,$$

λαμβάνουμε την ανάλογη σχέση για τα διαδοχικά υπόλοιπα

$$r_k = (I - AM^{-1})r_{k-1}.$$

Παίρνοντας νόρμες και στις δύο παραπάνω σχέσεις έχουμε :

$$\|e_k\| \leq \left\|I - M^{-1}A\right\| \|e_{k-1}\|, \tag{4.2}$$

$$\|r_k\| \leq \left\|I - AM^{-1}\right\| \|r_{k_1}\|. \tag{4.3}$$

Παρατηρούμε, λοιπόν, ότι αν $\|I - M^{-1}A\| < 1$ ($\|I - AM^{-1}\| < 1$), για κάποια φυσική νόρμα του αντίστοιχου πίνακα, τότε το (απόλυτο) σφάλμα (υπόλοιπο) μειώνεται μονότονα, με το άνω φράγμα του ρυθμού μείωσης ανά επανάληψη να δίδεται από την παραπάνω νόρμα.

4.2 Μέθοδοι Orthomin(1) και Απότομης Καθόδου Steepest Descent

Σε αυτή την παράγραφο, θα εξετάσουμε μία νέα κατηγορία μεθόδων που αποσκοπούν στη βελτίωση της μεθόδου της Απλής Επανάληψης. Βασίζονται, κυρίως, στην ελαχιστοποίηση μίας νόρμας είτε του διανύσματος−σφάλματος είτε του διανύσματος−υπολοίπου. Εφεξής, θα υποθέτουμε ότι το αρχικό σύστημα είναι ήδη προρρυθμισμένο. Θα λέμε ότι είναι "Ερμιτιανό", αν ο αρχικός πίνακας A είναι Ερμιτιανός και ο προρρυθμιστής M είναι Ερμιτιανός και θετικά ορισμένος. Στην περίπτωση όπου ο αρχικός πίνακας A είναι Ερμιτιανός και ο προρρυθμιστής είναι Ερμιτιανός αλλά μη-ορισμένος, τότε θα θεωρούμε το σύστημα ως "μη-Ερμιτιανό". Στην περίπτωση όπου έχουμε Ερμιτιανό και θετικά ορισμένο προρρυθμιστή, τότε μπορούμε να αναλύσουμε αυτόν, σύμφωνα με την παραγοντοποίηση Cholesky, ως γινόμενο

$$M = LL^H.$$

Σε αυτή την περίπτωση, μπορούμε να εφαρμόσουμε τον καλούμενο "αριστερό–δεξιό" προρρυθμιστή, με το αρχικό σύστημα να λαμβάνει τη μορφή

$$L^{-1}AL^{-H}y = L^{-1}b,\ y = L^H x.$$

Σημείωση: Στη συνέχεια, όταν σε κάποιο εσωτερικό γινόμενο ή σε κάποια νόρμα δεν υπάρχει δείκτης, τότε θα υποθέτουμε ότι αναφερόμαστε πάντοτε στο Ευκλείδειο εσωτερικό γινόμενο ή στην Ευκλείδεια νόρμα αντίστοιχα.

Όπως αναφέρθηκε στην αρχή της παραγράφου, σκοπός μας είναι η βελτίωση της Απλής Επαναληπτικής μεθόδου. Για να το επιτύχουμε, επιλέγουμε, καταρχάς, ως προρρυθμιστή τον πίνακα

$$M_k = \frac{1}{\alpha_k} M,\ k = 0, 1, 2, \cdots,$$

όπου α_k παράμετρος μεταβαλλόμενη με τις επαναλήψεις. Η ακολουθία των διαδοχικών προσεγγίσεων της λύσης είναι η εξής:

$$x_{k+1} = x_k + \alpha_k (b - Ax_k). \tag{4.4}$$

Στη γενική περίπτωση επιδιώκουμε την ελαχιστοποίηση της Ευκλείδειας νόρμας του διανύσματος−υπολοίπου r_k $(= b - Ax_k)$ ανά επανάληψη. Στην περίπτωση όπου το σύστημα είναι Ερμιτιανό και θετικά ορισμένο, μπορούμε να επιδιώξουμε την ελαχιστοποίηση είτε της Ευκλείδειας νόρμας του διανύσματος−-υπολοίπου r_k $(= b - Ax_k)$ είτε της "$A-$νόρμας" του αντίστοιχου διανύσματος−-σφάλματος. (Σημείωση: Ο ορισμός της "$A-$νόρμας" διανύσματος θα δοθεί παρακάτω.)

Στην περίπτωση της ελαχιστοποίησης της Ευκλείδειας νόρμας του διανύσματος−υπολοίπου, $r_{k+1} = r_k - \alpha_k A r_k$, έχουμε ότι

$$(r_{k+1}, r_{k+1}) = (r_k, r_k) - 2Re(\bar{\alpha}_k (r_k, Ar_k)) + |\alpha_k|^2 (Ar_k, Ar_k).$$

Θέτοντας $\alpha_k = x + iy,\ x, y \in \mathbb{R}$, και $(r_k, Ar_k) = \gamma + i\delta,\ \gamma, \delta \in \mathbb{R}$, καταλήγουμε στην ισότητα

$$\|r_{k+1}\|^2 = \|r_k\|^2 - 2(x\gamma + y\delta) + (x^2 + y^2) \|Ar_k\|^2 .$$

Ορίζοντας ως

$$F(x, y) := ||r_{k+1}||^2 = \|r_k\|^2 - 2(x\gamma + y\delta) + (x^2 + y^2) \|Ar_k\|^2 ,$$

έχουμε ότι

$$\nabla F(x, y) = 0 \ \Rightarrow\ x = \frac{\gamma}{\|Ar_k\|^2},\ y = \frac{\delta}{\|Ar_k\|^2}.$$

Επιπλέον, λαμβάνοντας τον πίνακα των δεύτερων παραγώγων (Εσιανό)

$$H_F = \begin{bmatrix} \frac{\partial^2 F}{\partial x^2} & \frac{\partial^2 F}{\partial x \partial y} \\ \frac{\partial^2 F}{\partial y \partial x} & \frac{\partial^2 F}{\partial y^2} \end{bmatrix} = \begin{bmatrix} 2\|Ar_k\|^2 & 0 \\ 0 & 2\|Ar_k\|^2 \end{bmatrix},$$

διαπιστώνουμε αμέσως ότι είναι θετικά ορισμένος. Επομένως, η Ευκλείδεια νόρμα του διανύσματος−υπολοίπου ελαχιστοποιείται, εάν επιλέξουμε

$$x = \frac{\gamma}{\|Ar_k\|^2} \ \kappa\alpha\iota \ y = \frac{\delta}{\|Ar_k\|^2}$$

ή, ισοδύναμα,

$$\alpha_k = \frac{(r_k, Ar_k)}{(Ar_k, Ar_k)}. \tag{4.5}$$

Με την επιλογή (4.5), λαμβάνουμε μία μέθοδο που ονομάζεται "Orthomin(1)". Μπορούμε να παρατηρήσουμε ότι για να πάρουμε το υπόλοιπο $r_{k+1} = r_k - \alpha_k Ar_k$, θεωρούμε τη διαφορά του r_k από την προβολή του στο Ar_k (βλ. (3.2)). Διαπιστώνουμε, λοιπόν, ότι οι νόρμες των r_{k+1} και r_k ικανοποιούν την ανισότητα $\|r_{k+1}\| \leq \|r_k\|$. Στην περίπτωση που το r_k είναι κάθετο στο Ar_k, έχουμε ισότητα των δύο νορμών. Τότε λέμε ότι ο αλγόριθμος της μεθόδου "καταρρέει".

Θεώρημα 4.2. *: Η Ευκλείδεια νόρμα του υπολοίπου της επαναληπτικής μεθόδου $x_{k+1} = x_k + \alpha_k(b - Ax_k)$ με την επιλογή του $\alpha_k = \frac{(r_k, Ar_k)}{(Ar_k, Ar_k)}$ μειώνεται γνήσια μονότονα για κάθε r_0 ανν $0 \notin \mathcal{F}(A^H)$.*

Απόδειξη: Εύκολα, παρατηρεί κάποιος ότι εάν $0 \in \mathcal{F}(A^H)$, τότε θα υπήρχε διάνυσμα x_0 και άρα r_0 τέτοιο, ώστε $0 = r_0^H A^H r_0$. Έπομένως, το επόμενο διάνυσμα−υπόλοιπο θα είναι $r_1 = r_0$, το οποίο σημαίνει ότι $r_k = r_0$, $k = 1, 2, \ldots$, δηλαδή, κατάρρευση της μεθόδου, **χωρίς** εύρεση της λύσης $\square$.

Παρατήρηση 4.1. : Το παραπάνω θεώρημα ισχύει ακόμη και αν στην υπόθεση έχουμε ότι $0 \notin \mathcal{F}(A)$ αντί για $0 \notin \mathcal{F}(A^H)$, αφού το $\mathcal{F}(A)$ είναι το συζυγές μιγαδικό του $\mathcal{F}(A^H)$[5].

[5]Το συζυγές μιγαδικό ενός συνόλου $\mathcal{A} \subset \mathbb{C}$ είναι το σύνολο $\mathcal{A}^H$ με στοιχεία τα συζυγή μιγαδικά του $\mathcal{A}$.

Θεώρημα 4.3. *: Το επαναληπτικό σχήμα του προηγούμενου θεωρήματος συγκλίνει στο $A^{-1}b$ για κάθε r_0 ανν $0 \notin \mathcal{F}(A^H)$. Σε αυτήν την περίπτωση η Ευκλείδεια νόρμα του υπολοίπου ικανοποιεί τη σχέση*

$$\|r_{k+1}\| \leq \sqrt{1 - \frac{d^2}{\|A\|^2}}\, \|r_k\|\,, \tag{4.6}$$

όπου d είναι η απόσταση της αρχής των αξόνων από το πεδίο τιμών $\mathcal{F}(A^H)$.

Παρατήρηση 4.2. : Αν ο $A \in \mathbb{R}^{n,n}$ και ο $H(A) = \frac{A+A^H}{2}$ είναι θετικά ορισμένος, τότε η απόσταση d είναι η ελαχίστη ιδιοτιμή του πίνακα $H(A)$.

Πρόταση 4.4. *: Έστω ότι ο $A \in \mathbb{C}^{n,n}$ είναι Ερμιτιανός και θετικά ορισμένος. Τότε υπάρχει μοναδικός Ερμιτιανός και θετικά ορισμένος πίνακας $B \in \mathbb{C}^{n,n}$ τ.ω. $A = B^H B$. (Σημείωση: Ο πίνακας B καλείται (θετική) τετραγωνική ρίζα του πίνακα A και συμβολίζεται με $A^{\frac{1}{2}}$ (βλ. [40])).*

Ορισμός 4.1. : Έστω πίνακας $A \in \mathbb{C}^{n,n}$ Ερμιτιανός και θετικά ορισμένος. Τότε για κάθε $x \in \mathbb{C}^n$ η απεικόνιση $\|.\|_A : \mathbb{C}^{n,n} \to [0, \infty)$, που ορίζεται από τη σχέση

$$\|x\|_A = (x, Ax)^{\frac{1}{2}},$$

ορίζει μία διανυσματική νόρμα η οποία καλείται "$A-$νόρμα" ή "νόρμα ενέργειας".

Παρατήρηση 4.3. : Παρατηρούμε ότι $\|x\|_A = \left\|A^{\frac{1}{2}}x\right\|$. Επόμενως, η παραπάνω νόρμα είναι μία $G^H G-$νόρμα, με $G = A^{\frac{1}{2}}$ (βλ. (3.1)).

Υποθέτουμε ότι ο A είναι Ερμιτιανός και θετικά ορισμένος. Από την (4.4) έχουμε ότι $e_{k+1} = e_k - \alpha_k r_k$. Θεωρώντας το τετράγωνο της $A-$νόρμας του σφάλματος, παίρνουμε:

$$(e_{k+1}, Ae_{k+1}) = (e_k, Ae_k) - 2\alpha_k(r_k, r_k) + \alpha_k^2(r_k, Ar_k).$$

Η παραπάνω έκφραση, λόγω των θετικών Ευκλείδειων εσωτερικών γινομένων ελαχιστοποιείται (για α_k πραγματικό), όταν επιλέξουμε το συντελεστή α_k να ισούται με

$$\alpha_k = \frac{(e_k, Ar_k)}{(r_k, Ar_k)} = \frac{(r_k, r_k)}{(r_k, Ar_k)}. \tag{4.7}$$

Η μέθοδος που προκύπτει, σε αυτήν την περίπτωση, καλείται μέθοδος της "Απότομης Καθόδου" (Steepest Descent).

Παρατήρηση 4.4. : Μία διαφορετική διαδικασία δημιουργίας της παραπάνω μεθόδου δίδεται μέσω της ελαχιστοποίησης του συναρτησοειδούς[6]

$$f(x) = x^H Ax - 2b^H x,$$

το οποίο λαμβάνει ελάχιστο στο σημείο $x = A^{-1}b$. Τότε η διεύθυνση στην οποία έχουμε τη μεγίστη μείωση της τιμής της $f(x)$ στο $x = x_k$ είναι αυτή της $-\nabla f(x)_{|x=x_k}$, που ισούται με το διάνυσμα−υπόλοιπο $r_k = b - Ax_k$.

Σημείωση: Οι αλγόριθμοι των δύο μεθόδων παρουσιάζονται ολοκληρωμένοι στο Παράρτημα.

Στην περίπτωση όπου ο πίνακας A είναι Ερμιτιανός και θετικά ορισμένος, βελτίωση του φράγματος του σχετικού απόλυτου σφάλματος δύο οποιωνδήποτε διαδοχικών επαναλήψεων της μεθόδου Orthomin(1)

$$\frac{||r_{k+1}||}{||r_k||} \leq ||I - \alpha_k A||$$

μπορεί να γίνει αν στην θέση του α_k θεωρήσουμε σταθερό συντελεστή $\alpha \in \mathbb{R}$ και προσπαθήσουμε να ελαχιστοποιήσουμε το παραπάνω φράγμα. Έτσι η σχέση που συνδέει τις νόρμες των υπολοίπων δύο διαδοχικών επαναλήψεων παίρνει τη μορφή

$$\|r_{k+1}\| \leq \|I - \alpha A\| \, \|r_k\| \, . \tag{4.8}$$

Θεωρώντας $\lambda_i \in \sigma(A),\ i = 1, 2, \ldots, n$ τις ιδιοτιμές του A με $\lambda_{\min}$ και $\lambda_{\max}$, την ελάχιστη και μέγιστη αντίστοιχα εξ αυτών, μπορεί να δειχθεί ότι για $\alpha = \frac{2}{\lambda_{\max}+\lambda_{\min}}$ έχουμε

$$\begin{array}{c} \min_{\alpha\in\mathbb{R}} ||I - \alpha A|| = \min_{\alpha\in\mathbb{R}} \max_{\lambda_i\in\sigma(A)} |1 - \alpha\lambda_i| \leq \\ \min_{\alpha\in\mathbb{R}} \max_{\lambda\in[\lambda_{\min},\lambda_{\max}]} |1 - \alpha\lambda| = |1 - \frac{2\lambda_{\min}}{\lambda_{\max}+\lambda_{\min}}| \end{array} \tag{4.9}$$

και άρα

$$\|r_{k+1}\| \leq \left(\frac{\lambda_{\max} - \lambda_{\min}}{\lambda_{\max} + \lambda_{\min}}\right) \|r_k\| = \left(\frac{\kappa - 1}{\kappa + 1}\right) \|r_k\| \, ,$$

όπου $\kappa = \frac{\lambda_{\max}}{\lambda_{\min}}$ είναι ο δείκτης κατάστασης του πίνακα A.

[6]Η ελαχιστοποίηση αυτού του συναρτησοειδούς μπορεί να οδηγήσει στη δημιουργία και άλλων μεθόδων για Ερμιτιανούς και θετικά ορισμένους πίνακες, όπως η CG (Conjugate Gradient), την οποία θα μελετήσουμε παρακάτω.

Για τη μέθοδο της Απότομης Καθόδου και για την "$A-$νόρμα" του σφάλματος έχουμε αντίστοιχα φράγματα τα οποία δίδονται από τη σχέση

$$\|e_{k+1}\|_A \leq \left(\frac{\kappa-1}{\kappa+1}\right)\|e_k\|_A .$$

Θεώρημα 4.5. *: Έστω x_k η ακολουθία των διαδοχικών προσεγγίσεων της λύσης $A^{-1}b$ του γραμμικού συστήματος $Ax=b$, με τον πίνακα A να είναι Ερμιτιανός και θετικά ορισμένος[7], που παράγεται από τον αλγόριθμο της μεθόδου Απότομης Καθόδου για οποιαδήποτε αρχική προσσέγγιση $x_0 \in \mathbb{C}^{n,n}$. Έστω $\kappa(A)=\frac{\lambda_{\max}}{\lambda_{\min}}$, όπου $\lambda_{\max}, \lambda_{\min}$ είναι η μέγιστη και η ελάχιστη ιδιοτιμή του πίνακα A. Εάν $e_k = A^{-1}b - x_k$ είναι το διάνυσμα$-$σφάλμα στην k επανάληψη, τότε*

$$\|e_{k+1}\|_A \leq \left(\frac{\kappa-1}{\kappa+1}\right)\|e_k\|_A .$$

Για την απόδειξη του θεωρήματος ανατρέχουμε στο [6].

Στην περίπτωση μη-Ερμιτιανού συστήματος, με την επιπλέον υπόθεση ότι το πεδίο τιμών του πίνακα A^H βρίσκεται σε ένα δίσκο $D=\{z\in\mathcal{C}:|z-\bar{c}|\leq s\}$, που δεν περιέχει το μηδέν, επιλέγοντας το $\alpha=\frac{1}{c}$ στην (4.8) και χρησιμοποιώντας τις ιδιότητες του πεδίου τιμών (3.10), λαμβάνουμε τις σχέσεις:

$$\mathcal{F}(I-\frac{1}{\bar{c}}A^H) = 1-\frac{1}{\bar{c}}\mathcal{F}(A^H) \subseteq \left\{z\in\mathcal{C}:|z|\leq\frac{s}{|c|}\right\}.$$

Στη συνέχεια, χρησιμοποιώντας τη σχέση (3.11) που συνδέει τη νόρμα με την αριθμητική ακτίνα, με $\alpha=\frac{1}{c}$, έχουμε ότι

$$\|I-\alpha A\| = \left\|I-\frac{1}{\bar{c}}A^H\right\| \leq 2\nu(I-\frac{1}{\bar{c}}A^H) \leq 2\frac{s}{|c|}$$

και έτσι η σχέση (4.8) γίνεται

$$\|r_{k+1}\| \leq 2\frac{s}{|c|}\|r_k\| .$$

Παρακάτω θα διατυπώσουμε το παραπάνω συμπέρασμα σε μορφή θεωρήματος.

[7] Σύμφωνα με τον ορισμό που δόθηκε στο δεύτερο κεφάλαιο.

Θεώρημα 4.6. *: Έστω x_k η ακολουθία των διαδοχικών προσεγγίσεων της λύσης $A^{-1}b$ του γραμμικού συστήματος $Ax = b$, με $A \in \mathbb{C}^{n,n}, \det(A) \neq 0$. Έστω ότι το πεδίο τιμών του πίνακα $A^H(\mathcal{F}(A^H))$ ανήκει στο δίσκο $D = \{z \in \mathcal{C} : |z - \bar{c}| \leq s\}$, με $0 \notin D$. Τότε, εάν θεωρήσουμε $r_k = b - Ax_k$ το διάνυσμα–υπόλοιπο στην k επανάληψη του αλγορίθμου της μεθόδου $Orthomin(1)$ για το παραπάνω σύστημα, ισχύει ότι*

$$\|r_{k+1}\| \leq 2\frac{s}{|c|}\|r_k\|.$$

Μία σκιαγράφηση της απόδειξης παρουσιάσαμε λίγο πριν την εκφώνηση του θεωρήματος.

Με διαφορετικές επιλογές των s, c μπορούμε να δούμε ότι το φράγμα που βρίσκεται μπορεί να είναι ή να μην είναι καλύτερο από το (4.6) (βλ.[20]).

4.3 Μέθοδοι Orthomin(2) και Συζυγών Κλίσεων (Conjugate Gradient (CG))

Γενικεύοντας τις μεθόδους Orthomin(1) και Απότομης Καθόδου, χρησιμοποιούμε μία νέα ακολουθία διαδοχικών προσεγγίσεων της λύσης

$$x_{k+1} = x_k + \alpha_k p_k, \tag{4.10}$$

όπου τα διανύσματα p_k είναι, καταρχάς, γενικά τυχαίες διευθύνσεις. Το νέο υπόλοιπο και το σφάλμα, στην περίπτωση αυτή, δίδονται αντίστοιχα από τις σχέσεις

$$r_{k+1} = r_k - \alpha_k A p_k, \quad e_{k+1} = e_k - \alpha_k p_k.$$

Ο συντελεστής α_k επιλέγεται έτσι ώστε στην πρώτη περίπτωση το r_{k+1} να είναι ορθογώνιο στο Ap_k, ενώ στη δεύτερη το e_{k+1} να είναι $A-$ορθογώνιο στο p_k. Εάν για την εύρεση του νέου υπολοίπου στην περίπτωση της Orthomin(1), αντί να αφαιρέσουμε την προβολή του r_k στη διεύθυνση Ar_k, αφαιρέσουμε την προβολή του σε διεύθυνση Ap_k, η οποία, όπως διαπιστώνεται, είναι κάθετη στην Ap_{k-1}, όταν

$$p_k = r_k - \frac{(Ar_k, Ap_{k-1})}{(Ap_{k-1}, Ap_{k-1})}p_{k-1}, \tag{4.11}$$

τότε έχουμε ότι: $(r_{k+1}, Ap_k) = 0,$ εξ ορισμού, και

$$(r_{k+1}, Ap_{k-1}) = (r_k, Ap_{k-1}) - \alpha_k(Ap_k, Ap_{k-1}) = 0,$$

διότι $(r_k, Ap_{k-1}) = 0$, εξ ορισμού, και $(Ap_k, Ap_{k-1}) = 0$, από την (4.11). Σε αυτήν την περίπτωση το υπόλοιπο ελαχιστοποιείται στο χώρο span$\{Ap_k, Ap_{k-1}\}$ =span$\{Ar_k, Ap_{k-1}\}$ και παίρνει τη μορφή

$$r_{k+1} = r_k - \alpha_k A r_k + a_k \beta_{k-1} A p_{k-1},$$

όπου

$$\alpha_k = \frac{(r_k, Ap_k)}{(Ap_k, Ap_k)} \ \kappa\alpha\iota \ \ \beta_{k-1} = \frac{(Ar_k, Ap_{k-1})}{(Ap_{k-1}, Ap_{k-1})},$$

αντίστοιχα. Οι τιμές των α_k, β_{k-1} είναι προφανώς τ.ω. το διάνυσμα r_{k+1} να είναι ορθογώνιο στο χώρο span$\{Ap_k, Ap_{k-1}\}$ ή, ισοδύναμα, στο span$\{Ar_k, Ap_{k-1}\}$. Η μέθοδος αυτή, που παράγεται από την παραπάνω διαδικασία, καλείται Orthomin(2) και ο αλγόριθμός της δίδεται στο Παράρτημα.

Παρατήρηση 4.5. : Στην περίπτωση όπου το r_0 είναι κάθετο στο Ar_0, τότε η μέθοδος αποτυγχάνει, με την έννοια ότι δεν μπορεί να επιτευχθεί περαιτέρω μείωση της νόρμας του υπολοίπου. Μπορεί και πάλι να αποδειχθεί ότι όταν $0 \notin \mathcal{F}(A^H)$, τότε ο η μέθοδος δεν αποτυγχάνει και το φράγμα για τη νόρμα του υπολοίπου δίδεται από την (4.6).

Ειδικά, αν ο πίνακας A είναι Ερμιτιανός ισχύει το επόμενο θεώρημα.

Θεώρημα 4.7. *: Έστω ότι ο A είναι Ερμιτιανός, τα $\alpha_0, \ldots, \alpha_{k-1}$ είναι μη-αρνητικοί, και τα διανύσματα $r_1, \ldots, r_{k+1}$ και $p_1, \ldots, p_{k+1}$ στη μέθοδο της Orthomin(2) είναι καλώς ορισμένα*[8] *. Τότε*

$$(r_{k+1}, Ap_j) = (Ap_{k+1}, Ap_j) = 0 \ \ \forall j \leq k \leq n-1.$$

Επιπλέον, για κάθε διάνυσμα στο χώρο

$$r_0 + span\{Ar_0, \ldots, A^{k+1} r_0\},$$

το r_{k+1} έχει την ελαχίστη Ευκλείδεια νόρμα. Επίσης, αν $\alpha_0, \ldots, \alpha_{n-2} \neq 0$ και $r_1, \ldots, r_n$ και $p_1, \ldots, p_n$ είναι καλώς ορισμένα, τότε $r_n = 0$.

Η απόδειξη του παραπάνω θεωρήματος γίνεται με επαγωγή στο k (βλ. [20]). Η επίλυση ενός συστήματος, με την παραπάνω διαδικασία ελαχιστοποίησης

[8] Η υπόθεση αυτή δεν είναι απαραίτητη αφού μπορούμε να τη συμπεράνουμε με τη χρήση των υπολοίπων αν $\alpha_k \neq 0$. Τότε, με την παραπάνω μέθοδο, παράγονται k γραμμικώς ανεξάρτητα διανύσματα διεύθυνσης.

του υπολοίπου καλείται Μέθοδος Ελαχίστου Υπολοίπου (Minimal Residual (MINRES)).

Στην περίπτωση όπου το σύστημα είναι Ερμιτιανό και θετικά ορισμένο, τότε αυτό που κάνουμε είναι η απαλοιφή της $A-$προβολής του σφάλματος σε μία διεύθυνση $A-$ορθογώνια στην προηγούμενη. Δηλαδή, στη διεύθυνση

$$p_k = r_k - \frac{(r_k, Ap_{k-1})}{(p_{k-1}, Ap_{k-1})} p_{k-1}.$$

Σε αυτή την περίπτωση, έχουμε ότι

$$(e_{k+1}, Ap_k) = (e_{k+1}, Ap_{k-1}) = 0,$$

και άρα η "$A-$νόρμα" του σφάλματος ελαχιστοποιείται στο χώρο

$$e_k + span\{r_k, p_{k-1}\}.$$

Η μέθοδος, που εφαρμόζει τα παραπάνω, καλείται Μέθοδος "Συζυγών Κλίσεων" (Conjugate Gradient (CG)) και η υλοποίησή της παρουσιάζεται στο Παράρτημα.

Θεώρημα 4.8. *: Έστω ότι ο A είναι Ερμιτιανός και θετικά ορισμένος. Η μέθοδος της CG παράγει την ακριβή λύση του αρχικού συστήματος σε n, το πολύ, επαναλήψεις. Το σφάλμα, το υπόλοιπο και τα διανύσματα–διεύθυνσης που προκύπτουν κατά την εφαρμογή της μεθόδου είναι καλά ορισμένα*[9] *και ικανοποιούν τις σχέσεις:*

$$(e_{k+1}, Ap_j) = (p_{k+1}, Ap_j) = (r_{k+1}, r_j) = 0 \ \ \forall j \leq k \leq n-1.$$

Επίσης, από όλα τα διανύσματα που ανήκουν στο χώρο

$$e_0 + span\{Ae_0, \ldots, A^{k+1}e_0\},$$

το e_{k+1} έχει την ελαχίστη $A-$νόρμα.

Η απόδειξη είναι όμοια με αυτή του προηγούμενου θεωρήματος.

Βλέπουμε, λοιπόν, ότι και στις δύο μεθόδους, το σφάλμα e_k αλλά και το υπόλοιπο r_k μπορούν να εκφραστούν ως γραμμικοί συνδυασμοί των διανυσμάτων $A^l e_0$ και $A^l r_0$, $l = 0, 1, \ldots, k$, αντίστοιχα. Επομένως, $e_k = p_k(A)e_0$ και

[9] Όπως στο προηγούμενο θεώρημα.

$r_k = p_k(A)r_0$, όπου τα $p_k(A)$ είναι πολυώνυμα−πίνακες k βαθμού ως προς A με συντελεστές, γενικά, μιγαδικούς. Επομένως, $e_k = p_k(A)e_0$ και $r_k = p_k(A)r_0$. Έτσι, για το σφάλμα e_k της μεθόδου CG παίρνουμε τη σχέση

$$\|e_k\|_A = \min_{p_k \in \mathcal{P}_k,\ p_k(0)=1} \|p_k(A)e_0\|_A .$$

Ομοίως, για το υπόλοιπο r_k της μεθόδου MINRES έχουμε τη σχέση

$$\|r_k\| = \min_{p_k \in \mathcal{P}_k,\ p_k(0)=1} \|p_k(A)r_0\| .$$

Και στις δύο σχέσεις, το minimum λαμβάνεται πάνω στο χώρο όλων των πολυωνύμων βαθμού k, $\mathcal{P}_k$, με $p_k(0) = 1$.

Παρατηρούμε ότι, επειδή ο A είναι Ερμιτιανός θα είναι και διαγωνοποιήσιμος. Επομένως, υπάρχει ορθοκανονικός πίνακας U τέτοιος ώστε $A = U\Lambda U^H$. Ο Λ είναι διαγώνιος πίνακας με στοιχεία τις ιδιοτιμές του A.

Στην περίπτωση όπου ο A είναι επιπλέον θετικά ορισμένος, θεωρώντας γνωστή την προηγούμενη παραγοντοποίησή του, βρίσκεται ότι $A^{\frac{1}{2}} = U\Lambda^{\frac{1}{2}}U^H$, όπου τα διαγώνια στοιχεία του $\Lambda^{\frac{1}{2}}$ είναι οι θετικές τετραγωνικές ρίζες των αντίστοιχων στοιχείων του Λ. Επομένως,

$$\|e_k\|_A = \min_{p_k \in \mathcal{P}_k,\ p_k(0)=1} \left\| A^{\frac{1}{2}} p_k(A)e_0 \right\| .$$

Αντικαθιστώντας τους πίνακες A, $A^{\frac{1}{2}}$ σύμφωνα με τις παραπάνω εκφράσεις τους, έχουμε ότι

$$\|e_k\|_A \leq \min_{p_k \in \mathcal{P}_k,\ p_k(0)=1} \|p_k(\Lambda)\| \, \|e_0\|_A . \tag{4.12}$$

Με όμοιο τρόπο, λαμβάνουμε την αντίστοιχη σχέση για το υπόλοιπο

$$\|r_k\| \leq \min_{p_k \in \mathcal{P}_k,\ p_k(0)=1} \|p_k(\Lambda)\| \, \|r_0\| . \tag{4.13}$$

Και τα δύο παράπανω φράγματα μπορεί να αποδειχθεί ότι είναι sharp[10]. Οι δύο εκφράσεις στις (4.12) και (4.13) μπορούν να πάρουν τις μορφές

$$\frac{\|e_k\|_A}{\|e_0\|_A} \leq \min_{p_k \in \mathcal{P}_k,\ p_k(0)=1} \max_{i=1,\ldots,n} |p_k(\lambda_i)| , \ e_0 \in \mathbb{C}^n \backslash \{0\}, \tag{4.14}$$

[10] Δηλαδή, υπάρχουν αρχικά διανύσματα e_0 και r_0, αντίστοιχα, τ.ω. και οι δύο ανισότητες να γίνονται ισότητες.

και

$$\frac{\|r_k\|}{\|r_0\|} \leq \min_{p_k \in \mathcal{P}_k,\ p_k(0)=1} \max_{i=1,\ldots,n} |p_k(\lambda_i)|\,,\ r_0 \in \mathbb{C}^n \backslash \{0\}, \tag{4.15}$$

για τη μέθοδο CG και για τη μέθοδο MINRES, αντίστοιχα. Θεωρώντας ότι

$$0 < \lambda_{\min} \equiv \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n \equiv \lambda_{\max}$$

είναι οι ιδιοτιμές του A και γνωρίζοντας μόνο τη $\lambda_{\min}$ και τη $\lambda_{\max}$, με τη βοήθεια των πολυωνύμων του Chebyshev πρώτου είδους,

$$T_k(z),\ k = 0, 1, 2, \ldots,\ (T_0(z) = 1,\ T_1(z) = z,\ T_k(z) = 2zT_{k-1} - T_{k-2},\ k \geq 2),$$

που ορίζονται για $z \in [-1,\ 1]$ ή $z \in [1,\ \infty)$ (ή ακόμη για $z \in (-\infty,\ -1]$), μπορούμε να λάβουμε φράγματα για τα παραπάνω σχετικά απόλυτα σφάλματα. Για να επιτευχθεί αυτό, ανάγουμε την επίλυση του διακριτού minmax προβλήματος στις σχέσεις (4.14) και (4.15) σε ανάλογο συνεχές ως εξής:

$$\min_{p_k \in \mathcal{P}_k,\ p_k(0)=1} \max_{i=1,\ldots,n} |p_k(\lambda_i)| \leq \min_{p_k \in \mathcal{P}_k,\ p_k(0)=1} \max_{\lambda \in [\lambda_{\min}, \lambda_{\max}]} |p_k(\lambda)|\,. \tag{4.16}$$

Η επίλυση του συνεχούς προβλήματος αποτελεί ένα κλασικό πρόβλημα του οποίου η λύση δίδεται μέσω των πολυωνύμων του Chebyshev πρώτου είδους στο διάστημα $[\lambda_{\min}, \lambda_{\max}]$. Συγκεκριμένα, το πολυώνυμο που επιλύει το συνεχές minmax πρόβλημα στην (4.16) είναι το

$$p_k(z) = \frac{T_k\left(\frac{2z - \lambda_{\max} - \lambda_{min}}{\lambda_{\max} - \lambda_{\min}}\right)}{T_k\left(\frac{-\lambda_{\max} - \lambda_{min}}{\lambda_{\max} - \lambda_{\min}}\right)}, \tag{4.17}$$

όπου $T_k(z)$ είναι το πολυώνυμο του Chebyshev πρώτου είδους βαθμού k. Παρατηρούμε ότι, εφόσον $T_k(z) = \cos(k \arccos(z))$ για $z \in [-1, 1]$, η απόλυτη τιμή του αριθμητή του $p_k(z)$ φράσσεται από τη μονάδα, αφού για κάθε $z \in [\lambda_{\min}, \lambda_{\max}]$ το όρισμα στον αριθμητή ικανοποιεί $-1 \leq \frac{2z - \lambda_{\max} - \lambda_{min}}{\lambda_{\max} - \lambda_{\min}} \leq 1$. Για να υπολογίσουμε την απόλυτη τιμή του παρανομαστή, παρατηρούμε ότι το όρισμα στον παρονομαστή ικανοποιεί τη $\frac{-\lambda_{\max} - \lambda_{min}}{\lambda_{\max} - \lambda_{\min}} < -1$ ($\lambda_{\max} > \lambda_{min}$). Σε αυτήν την περίπτωση $T_k(z) = \cosh(k \operatorname{arccos}h(z))$ και από την επίλυση της εξίσωσης διαφορών που ορίζει τα πολυώνυμα του Chebyshev μπορεί να βρεθεί, μετά από πράξεις, η συγκεκριμένη τιμή για το $T_k\left(\frac{-\lambda_{\max} - \lambda_{min}}{\lambda_{\max} - \lambda_{\min}}\right)$, που δίδεται έμμεσα στο επόμενο Θεώρημα.

Θεώρημα 4.9. *: Έστω ότι e_k είναι το σφάλμα στην k επανάληψη της μεθόδου CG εφαρμοσμένης σε Ερμιτιανό και θετικά ορισμένο σύστημα. Τότε*

$$\frac{\|e_k\|_A}{\|e_0\|_A} \leq 2\left[\left(\frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}\right)^k + \left(\frac{\sqrt{\kappa}+1}{\sqrt{\kappa}-1}\right)^k\right]^{-1} \leq 2\left(\frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}\right)^k, \qquad (4.18)$$

όπου $\kappa = \frac{\lambda_{max}}{\lambda_{min}}$ είναι ο δείκτης κατάστασης του πίνακα A.

Στην προηγούμενη περίπτωση, μας αρκούσαν η μέγιστη και η ελάχιστη ιδιοτιμή του πίνακα A. Εάν, επιπλέον, γνωρίζουμε ότι, π.χ., η μέγιστη ιδιοτιμή είναι "μακριά" από τις υπόλοιπες και πιο συγκεκριμένα

$$\lambda_1 \leq \ldots \leq \lambda_{n-1} \ll \lambda_n,$$

τότε η επίλυση του minmax προβλήματος (4.16) ανάγεται στην επίλυση ενός άλλου minmax προβλήματος στο χώρο των πολυωνύμων $p_{k-1} \in \mathcal{P}_{k-1}$,$p_{k-1}(0) = 1$. Συγκεκριμένα, αν θέσουμε $p_k(z) = ap_{k-1}(z)(z - \lambda_{\max})$, ο περιορισμός $p_k(0) = 1$ και ο $p_{k-1}(0) = 1$ δίδουν $a = -\frac{1}{\lambda_{\max}}$, οπότε έχουμε:

$$p_k(z) = p_{k-1}(z)\left(\frac{\lambda_{\max} - z}{\lambda_{\max}}\right).$$

Παρατηρούμε ότι για το πολυώνυμο $p_k(z)$ ισχύει $p_k(\lambda_{\max}) = 0$. Έτσι για την επίλυση του minmax προβλήματος (4.16), αρκεί να βρεθεί η μέγιστη τιμή του πολυωνύμου $p_k \in \mathcal{P}_k$, $p_k(0)= 1$ πάνω από τις $n-1$ πρώτες ιδιοτιμές λ_i, $i = 1, 2, \ldots, n-1$, όπως φαίνεται στη συνέχεια. Συγκεκριμένα

$$\min_{p_k\in\mathcal{P}_k,\ p_k(0)=1} \max_{\lambda_i\in\sigma(A),\ i=1,2,\ldots,n} |p_k(\lambda_i)| \leq$$

$$\min_{p_{k-1}\in\mathcal{P}_{k-1},\ p_{k-1}(0)=1} \left[\max_{\lambda_i\in\sigma(A),\ i=1,2,\ldots,n-1} |p_{k-1}(\lambda)| \max_{\lambda_i\in\sigma(A)} \left|\frac{\lambda_{\max} - \lambda_i}{\lambda_{\max}}\right|\right] \leq$$

$$\min_{p_{k-1}\in\mathcal{P}_{k-1},\ p_{k-1}(0)=1} \left[\max_{\lambda\in[\lambda_{\min},\lambda_{n-1}]} |p_{k-1}(\lambda)| \max_{\lambda\in[\lambda_{\min},\lambda_{n-1}]} \left|\frac{\lambda_n - \lambda}{\lambda_n}\right|\right] \leq$$

$$\left(\frac{\lambda_n - \lambda_1}{\lambda_n}\right) \min_{p_{k-1}\in\mathcal{P}_{k-1},\ p_{k-1}(0)=1} \max_{\lambda\in[\lambda_{\min},\lambda_{n-1}]} |p_{k-1}(\lambda)| <$$

$$\min_{p_{k-1}\in\mathcal{P}_{k-1},\ p_{k-1}(0)=1} \max_{\lambda\in[\lambda_{\min},\lambda_{n-1}]} |p_{k-1}(\lambda)|.$$

Το τελευταίο minmax πρόβλημα μπορεί να επιλυθεί, όπως προηγουμένως, με τη χρήση των πολυωνύμων του Chebyshev πρώτου είδους βαθμού $k-1$ στο διάστημα $[\lambda_{\min}, \lambda_{n-1}]$. Έτσι, λοιπόν, το μετασχηματισμένο αρχικό minmax πρόβλημα επιλύεται θεωρώντας ουσιαστικά το πολυώνυμο :

$$p_k(z) = \frac{T_{k-1}\left(\frac{2z-\lambda_{n-1}-\lambda_1}{\lambda_{n-1}-\lambda_1}\right)}{T_{k-1}\left(\frac{-\lambda_{n-1}-\lambda_1}{\lambda_{n-1}-\lambda_1}\right)} \left(\frac{\lambda_n - z}{\lambda_n}\right).$$

Με βάση το αποτέλεσμα (4.18) βρίσκεται ότι:

$$\frac{\|e_k\|_A}{\|e_0\|_A} \leq 2\left(\frac{\sqrt{\kappa_{n-1}}-1}{\sqrt{\kappa_{n-1}}+1}\right)^{k-1}, \; \kappa_{n-1} = \frac{\lambda_{n-1}}{\lambda_1}.$$

Ομοίως, αν συμβαίνει να γνωρίζουμε ότι κάποιες από τις ιδιοτιμές του A, όπως π.χ., οι $n-l$ μεγαλύτερες, απέχουν αρκετά από τις υπόλοιπες, δηλαδή,

$$\lambda_1 \leq \ldots \leq \lambda_{n-l} \ll \lambda_{n-l+1} \leq \ldots \leq \lambda_n,$$

τότε με σκεπτικό ανάλογο προς το προηγούμενο καταλήγουμε σε ένα πολυώνυμο της μορφής:

$$p_k(z) = \frac{T_{k-1}\left(\frac{2z-\lambda_{n-l}-\lambda_1}{\lambda_{n-l}-\lambda_1}\right)}{T_{k-1}\left(\frac{-\lambda_{n-l}-\lambda_1}{\lambda_{n-l}-\lambda_1}\right)} \prod_{i=n-l+1}^{n} \left(\frac{\lambda_i - z}{\lambda_i}\right),$$

οπότε μπορούμε να λάβουμε φράγματα της μορφής:

$$\frac{\|e_k\|_A}{\|e_0\|_A} \leq 2\left(\frac{\sqrt{\kappa_{n-l}}-1}{\sqrt{\kappa_{n-l}}+1}\right)^{k-l}, \; \kappa_{n-l} = \frac{\lambda_{n-l}}{\lambda_1}.$$

Όλα τα παραπάνω συνοψίζονται στο επόμενο γενικό θεώρημα, η απόδειξη του οποίου δόθηκε ουσιαστικά στις δύο προηγούμενες σελίδες.

Θεώρημα 4.10. *: Έστω ότι e_k είναι το σφάλμα στην k επανάληψη της μεθόδου CG εφαρμοσμένης σε Ερμιτιανό και θετικά ορισμένο σύστημα. Εάν επιπλέον, οι ιδιοτιμές του πίνακα A είναι διατεταγμένες*

$$\lambda_1 \leq \ldots \leq \lambda_{n-l} \ll \lambda_{n-l+1} \leq \ldots \leq \lambda_n,$$

δηλαδή, οι $n-l$ πρώτες απέχουν αρκετά από τις υπόλοιπες, τότε:

$$\frac{\|e_k\|_A}{\|e_0\|_A} \leq 2\left(\frac{\sqrt{\kappa_{n-l}}-1}{\sqrt{\kappa_{n-l}}+1}\right)^{k-l}, \; \kappa_{n-l} = \frac{\lambda_{n-l}}{\lambda_1}.$$

Ανάλογα αποτελέσματα λαμβάνουμε και στην περίπτωση της μεθόδου MINRES, εφαρμοσμένης σε Ερμιτιανά και θετικά ορισμένα συστήματα, όπου προκύπτουν ακριβώς οι ίδιοι τύποι, χρησιμοποιώντας όμως, τώρα, την Ευκλείδεια νόρμα του υπολοίπου, αντί της "$A-$νόρμας" του σφάλματος.

Στην περίπτωση Ερμιτιανών αλλά μη-ορισμένων συστημάτων, με ιδιοτιμές να βρίσκονται σε δύο διαστήματα της μορφής

$$\sigma(A) \in [a,b] \bigcup [c,d],\ \mu\varepsilon\ a < b < 0 < c < d\ \kappa\alpha\iota\ b-a = d-c,$$

το πολυώνυμο Chebyshev, το οποίο επιλύει το αντίστοιχο minmax πρόβλημα, είναι το

$$p_k(z) = \frac{T_l(q(z))}{T_l(q(0))}$$

με $q(z) = 1 + \frac{2(z-b)(z-c)}{ad-bc}$, όπου $l = [k/2]$ [11]. Επιλέγοντας το παραπάνω πολυώνυμο, το φράγμα για το σφάλμα που παίρνουμε είναι

$$\frac{\|r_k\|}{\|r_0\|} \leq 2\left(\frac{\sqrt{|ad|}-\sqrt{|bc|}}{\sqrt{|ad|}+\sqrt{|bc|}}\right)^l$$

(βλ. [13]).

4.4 Μέθοδοι Orthodir, Γενικευμένη Ελαχίστου Υπολοίπου (Generalized Minimal Residual (GMRES)) και Ελαχίστου Υπολοίπου (Minimal Residual (MINRES))

Σε αυτή την παράγραφο θα ασχοληθούμε με την περίπτωση του γενικού πίνακα $A \in \mathbb{C}^{n,n}$, $\det(A) \neq 0$ και θα προσπαθήσουμε να ελαχιστοποιήσουμε την Ευκλείδεια νόρμα του υπολοίπου $\|r_{k+1}\|$ πάνω σε έναν j–διάστατο χώρο της μορφής

$$r_k + span\{Ap_k, Ap_{k-1}, \ldots, Ap_{k-j+1}\},$$

όπου

$$p_k = r_k - \sum_{l=1}^{j-1} \beta_{k-l}^{(k)} p_{k-l}, \quad \beta_{k-l}^{(k)} = \frac{(Ar_k, Ap_{k-l})}{(Ap_{k-l}, Ap_{k-l})}.$$

[11] Με [.] συμβολίζουμε το ακέραιο μέρος ενός πραγματικού αριθμού.

Η μέθοδος που προκύπτει με αυτή τη διαδικασία καλείται Orthomin(j) και είναι, όπως μπορεί να παρατηρηθεί, γενίκευση της Orthomin(1). Σε αυτή την περίπτωση, όπως και με τις προηγούμενες των Orthomin(1) και Orthomin(2), ο αλγόριθμος καταρρέει όταν $0 \in \mathcal{F}(A^H)$. Μία λύση για την αποφυγή αυτής της κατάστασης είναι η αντικατάσταση του r_k των παραπάνω σχέσεων με Ap_{k-1}. Τότε η μέθοδος που προκύπτει καλείται Orthodir. Όπως μπορεί κάποιος να παρατηρήσει, η μέθοδος αυτή είναι αρκετά δαπανηρή από άποψη πράξεων και μνήμης, αφού σε κάθε επανάληψη πρέπει να υπολογίζουμε διαφορετικά γινόμενα πίνακα−διανύσματος και εσωτερικά γινόμενα. Επιπλέον, θα πρέπει να αποθηκεύουμε τουλάχιστον j διαφορετικές διευθύνσεις για την εύρεση της νέας διεύθυνσης.

Στην ανάλυση που ακολουθεί χρειάζεται ο παρακάτω ορισμός:

Ορισμός 4.2. Ένας πίνακας $A \in \mathbb{C}^{m,n}$ καλείται άνω Hessenberg αν

$$a_{ij} = 0,\ i = 3, 4, \ldots, m,\ j = 1, 2, \ldots, \min\{i-2, n\}.$$

Μία άλλη μέθοδος για την εύρεση της προσεγγιστικής λύσης στην περίπτωση γενικού πίνακα είναι η Γενικευμένη Ελαχίστου Υπολοίπου (Generalized Minimal Residual) (GMRES)), η οποία απαιτεί τη μισή μνήμη από ότι η Orthodir(n) και έχει καλύτερη ακρίβεια. Η μέθοδος αυτή χρησιμοποιεί την παραλλαγή της Gram–Schmidt για την ορθοκανονικοποίηση του χώρου Krylov $\text{span}\{r_0, Ar_0, \ldots, A^k r_0\}$. Ο αλγόριθμος αυτής της ορθοκανονικοποίησης καλείται αλγόριθμος του Arnoldi και δίδεται στο Παράρτημα. Οι βασικές σχέσεις στο k βήμα του αλγορίθμου σε μορφή πινάκων δίδονται από τις

$$AQ_k = Q_k H_{kk} + h_{k+1}(\xi^k)^T = Q_{k+1} H_{k+1,k}. \tag{4.19}$$

Στις (4.19), ο πίνακας Q_k είναι ένας $n \times k$ ορθοκανονικός πίνακας με στήλες τα ορθοκανονικά διανύσματα q_k που παράγονται από τον ίδιο αλγόριθμο, ο πίνακας H_{kk} είναι ένας $k \times k$ άνω Hessenberg πίνακας με στοιχεία τα εσωτερικά γινόμενα $h_{ij} = (\widetilde{q}_{j+1}, q_i)$, όπου q_i είναι τα ορθοκανονικά διανύσματα που κατασκευάζονται από τον αλγόριθμο του Arnoldi, ενώ τα $\widetilde{q}_{j+1}$ είναι τα ορθογώνια ανά δύο διανύσματα τα οποία, επίσης, κατασκευάζονται από τον ίδιο αλγόριθμο. Ο πίνακας $H_{k+1,k}$ είναι ο προηγούμενος πίνακας H_{kk}, αυξημένος κατά μία επιπλέον γραμμή, της οποίας το μόνο μη-μηδενικό στοιχείο είναι το $h_{k+1,k}$. Το στοιχείο αυτό βρίσκεται από τον αλγόριθμο του Arnoldi. Το διάνυσμα h_{k+1} είναι το ανάστροφο του διανύσματος−γραμμής της τελευταίας γραμμής του πίνακα $H_{k+1,k}$. Το διάνυσμα ξ^k, που εμφανίζεται στις παραπάνω σχέσεις, είναι το γνωστό βασικό μοναδιαίο διάνυσμα διάστασης k με 1 στην k συνιστώσα.

Στη μέθοδο GMRES η προσέγγιση της λύσης δίδεται από τη σχέση

$$x_k = x_0 + Q_k y_k,$$

η οποία θα μπορούσαμε να πούμε ότι αποτελεί μία διαδικασία διαδοχικών βελτιώσεων της αρχικής προσέγγισης x_0, μέσω γραμμικών συνδυασμών των διανυσμάτων του χώρου Krylov $K_k(A, r_0) = \text{span}\{r_0, Ar_0, \ldots, A^{k-1}r_0\}$. Η αντίστοιχη σχέση για τα διαδοχικά υπόλοιπα δίδεται από την έκφραση

$$r_k = r_0 - AQ_k y_k.$$

Για να βρούμε πότε το r_k έχει ελαχίστη Ευκλείδεια νόρμα, θα πρέπει να λύσουμε το πρόβλημα των ελαχίστων τετραγώνων:

$$\min_{y\in\mathbb{C}^k} \|r_0 - AQ_k y\| \overset{(4.19)}{=} \min_{y\in\mathbb{C}^k} \|r_0 - Q_{k+1}H_{k+1,k}y\| =$$

$$\min_{y\in\mathbb{C}^k} \left\|Q_{k+1}(\tau\xi^1 - H_{k+1,k}y)\right\| = \min_{y\in\mathbb{C}^k} \left\|\tau\xi^1 - H_{k+1,k}y\right\|, \qquad (4.20)$$

όπου $\tau = \|r_0\|$ και ξ^1 είναι το βασικό μοναδιαίο διάνυσμα[12] διάστασης $k+1$. Η τελευταία ισότητα, στην παραπάνω σχέση, ισχύει γιατί ο πίνακας Q_{k+1} είναι ορθοκανονικός και επομένως, διατηρεί αναλλοίωτα τα Ευκλείδεια εσωτερικά γινόμενα.

Η περισσότερο διαδεδομένη μέθοδος για την επίλυση τέτοιων προβλημάτων ελαχίστων τετραγώνων είναι αυτή που εφαρμόζει την QR ανάλυση στον άνω Hessenberg πίνακα $H_{k+1,k}$. Η μέθοδος που συνήθως χρησιμοποιείται για την υλοποίηση της QR ανάλυσης είναι η εφαρμογή διαδοχικών στροφών Givens στον άνω Hessenberg πίνακα $H_{k+1,k}$. Η λύση y_k του προβλήματος ελαχίστων τετραγώνων λαμβάνεται από την επίλυση του άνω τριγώνικου συστήματος

$$R_{k\times k}y = \tau(F\xi^1)_{k\times 1},$$

όπου $R_{k\times k}$ είναι ο $k\times k$ κύριος υποπίνακας του άνω τριγωνικού πίνακα R που προέρχεται από την QR ανάλυση, ο δε πίνακας F^H είναι ο πίνακας Q της QR ανάλυσης.

Αυτό που τώρα μας απασχολεί είναι το πώς μπορεί να παραχθεί η QR ανάλυση του $H_{k+2,k+1}$, αν γνωρίζουμε την QR ανάλυση του $H_{k+1,k}$, με όσο το δυνατόν λιγότερες πράξεις. Η διαδικασία που χρησιμοποιείται είναι η εξής:

[12] Με μονάδα στη πρώτη συνιστώσα.

Έστω $F_{i,i+1}$ ο πίνακας−στροφή των μοναδιαίων διανυσμάτων ξ^i, ξ^{i+1} κατά μία γωνία θ_i:

$$F_{i,i+1} = \begin{pmatrix} I_{i-1} & & & \\ & c_i & s_i & \\ & -\bar{s}_i & c_i & \\ & & & I_{k-i} \end{pmatrix}, \quad i = 1, 2, \ldots, k. \tag{4.21}$$

Εφαρμόζοντας διαδοχικά τις στροφές $F_{i,i+1}$ στον πίνακα $H_{k+1,k}$, αυτός παίρνει τη μορφή:

$$(F_{k,k+1}F_{k-1,k}\cdots F_{12})H_{k+1,k} = R^{(k)} = \begin{pmatrix} x & x & \cdots & x \\ & x & \cdots & x \\ & & \ddots & \vdots \\ & & & x \\ 0 & 0 & \cdots & 0 \end{pmatrix}.$$

Για να πάρουμε τον παράγοντα $R^{(k+1)}$ από τον $H_{k+2,k+1}$, εφαρμόζουμε τις προηγούμενες στροφές $F_{12}, \ldots, F_{k,k+1}$ **μόνο** στην τελευταία στήλη του $H_{k+2,k+1}$ και καταλήγουμε σε πίνακα της μορφής:

$$\begin{pmatrix} x & x & \cdots & x & x \\ & x & \cdots & x & x \\ & & \ddots & \vdots & \vdots \\ & & & x & x \\ 0 & 0 & \cdots & 0 & d \\ 0 & 0 & \cdots & 0 & h \end{pmatrix}.$$

Στη συνέχεια, εφαρμόζουμε στον παραπάνω πίνακα τη στροφή $F_{k+1,k+2}$, επιλέγοντας, από τη (3.5), τη γωνία στροφής κατάλληλα ώστε να μηδενιστεί το στοιχείο $h_{k+2,k+1} = h$. Τότε το νέο $(k+1, k+1)$ στοιχείο του πίνακα είναι ίσο με $\frac{d}{|d|}\sqrt{|d|^2 + |h|^2}$ (βλ. (3.6)). Ο πίνακας $F = (F_{k+1,k+2}F_{k,k+1}\cdots F_{12})^H$ είναι αυτός που εφαρμόζεται στο δεξιό μέλος του συστήματος

$$R_{k\times k}y = \tau(F\xi^1)_{k\times 1}. \tag{4.22}$$

Παρατηρούμε ότι η απόλυτη τιμή του τελευταίου στοιχείου του διανύσματος $\tau F\xi^1$ του δεξιού μέλους είναι η Ευκλείδεια νόρμα του r_k. Αυτό συμβαίνει διότι από τις σχέσεις

$$\|b - Ax_k\| = \left\|\tau\xi^1 - F^H R y_k\right\| = \left\|\tau F\xi^1 - Ry_k\right\|,$$

με βάση την $\tau F\xi^1 = Ry_k$, έχουμε την ισότητα των k πρώτων συνιστωσών, από τη σχέση (4.22), εκτός από την τελευταία συνιστώσα του διανύσματος $\tau F\xi^1$. Βλέπουμε, λοιπόν, ότι με τη διαδικασία που περιγράφτηκε, μπορούμε, χωρίς επιπλέον κόστος, να ελέγχουμε σε κάθε βήμα τη νόρμα του υπολοίπου. Ο πλήρης αλγόριθμος της μεθόδου GMRES παρουσιάζεται στο Παράρτημα.

Παρατήρηση 4.6. : Ο αλγόριθμος της πλήρους GMRES είναι σχετικά δαπανηρός. Αφού εάν εφαρμοστεί $n-$φορές τότε η τάξη μεγέθους των πράξεων μόνο κατά τον αλγόριθμο του Arnoldi είναι $O(n^3)$. Επιπλέον το κόστος αποθήκευσης των $\{q_i\}_{i=1}^n$ είναι μεγάλο. Τα δύο παραπάνω γεγονότα φτάνουν να χαρακτηρίσουν τον αλγόριθμο της GMRES(n) ασύμφορο. Αυτό, λοιπόν, που συνήθως γίνεται στην πράξη είναι η "επανεκκίνηση" της μεθόδου μετά από λίγες σχετικά επαναλήψεις, με εκάστοτε αρχική προσέγγιση την τελευταία του προηγούμενου κύκλου της. Η νέα αυτή μέθοδος καλείται GMRES(j), με το j να αναφέρεται στην επανάληψη αμέσως μετά την οποία γίνεται η επανεκκίνηση της μεθόδου.

Σκοπός μας και εδώ είναι η ελαχιστοποίηση της Ευκλείδειας νόρμας του υπολοίπου της μεθόδου GMRES. Όπως και στην περίπτωση της μεθόδου MINRES, η Ευκλείδεια νόρμα του υπολοίπου εκφράζεται ως

$$\|r_k\| = \min_{p_k\in\mathcal{P}_k,\ p_k(0)=1} \|p_k(A)r_0\|\,,$$

με πίνακα $A \in \mathbb{C}^{n,n}$ και μόνες υποθέσεις ότι $\det(A) \neq 0$ και ότι ο πίνακας A είναι διαγωνιοποιήσιμος. Θέλοντας να δώσουμε φράγματα ανεξάρτητα από το r_0, εφαρμόζουμε και πάλι τη διαδικασία που ακολουθήθηκε για την εύρεση φραγμάτων στη MINRES στην περίπτωση του Ερμιτιανού συστήματος. Θεωρούμε, λοιπόν, χρησιμοποιώντας τη μορφή Jordan, την παραγοντοποίηση του πίνακα A σε $A = V\Lambda V^{-1}$, όπου Λ είναι διαγώνιος πίνακας που περιέχει τις ιδιοτιμές του πίνακα A και V ο πίνακας των αντίστοιχων ιδιοδιανυσμάτων. Τότε από τη σχέση $\|r_k\| = \min_{p_k\in\mathcal{P}_k,\ p_k(0)=1} \|p_k(A)r_0\|$ έχουμε:

$$\|r_k\| = \min_{p_k\in\mathcal{P}_k,\ p_k(0)=1} \left\|Vp_k(\Lambda)V^{-1}r_0\right\| \leq \kappa(V) \min_{p_k\in\mathcal{P}_k,\ p_k(0)=1} \|p_k(\Lambda)\|\,\|r_0\|\,, \tag{4.23}$$

που ισοδύναμα γίνεται

$$\frac{\|r_k\|}{\|r_0\|} = \min_{p_k\in\mathcal{P}_k,\ p_k(0)=1} \left\|Vp_k(\Lambda)V^{-1}\right\| \leq \kappa(V) \min_{p_k\in\mathcal{P}_k,\ p_k(0)=1} \|p_k(\Lambda)\|\,,$$

και όπου με $\kappa(V)$ συμβολίζουμε το δείκτη κατάστασης του πίνακα V. Παρατηρούμε ότι αυτό το φράγμα δεν είναι πάντοτε καλό, αφού ο δείκτης κατάστασης

του πίνακα V μπορεί να είναι οσοδήποτε μεγάλος και, επιπλέον, το πολυώνυμο p_k που ελαχιστοποιεί την έκφραση $\|Vp_k(\Lambda)V^{-1}r_0\|$ δεν είναι απαραίτητα αυτό που ελαχιστοποιεί τη $\|p_k(\Lambda)\|$, όπου $p_k \in \mathcal{P}_k$ με $p_k(0) = 1$. Στην ειδική περίπτωση κανονικών (normal) πινάκων, όπου $\kappa = 1$, το πρόβλημα ανάγεται σε ένα πρόβλημα βελτιστοποίησης πάνω στα μιγαδικά πολυώνυμα $\mathcal{P}_k$ με $p_k(0) = 1$. Και σε αυτή, όμως, την περίπτωση, δεν έχουμε "απλά" φράγματα, όπως αυτά των μεθόδων MINRES και CG.

Ένας άλλος τρόπος για την εύρεση φραγμάτων για το υπόλοιπο της μεθόδου είναι μέσω του πεδίου τιμών του πίνακα A. Έστω $0 \notin \mathcal{F}(A)$ και ότι το $\mathcal{F}(A)$ ανήκει σε ένα δίσκο $D = \{z \in \mathbb{C} : |z - c| \leq s\}$. Τότε από τις ιδιότητες του πεδίου τιμών (3.10) έχουμε:

$$\mathcal{F}(I - \frac{1}{c}A) = 1 - \frac{1}{c}\mathcal{F}(A) \subseteq \{z \in \mathbb{C} : |z| \leq \frac{s}{|c|}\}.$$

Έχοντας ότι η αριθμητική ακτίνα είναι $\nu(I - \frac{1}{c}A) \leq \frac{s}{|c|}$, παίρνουμε ότι

$$\nu((I - \frac{1}{c}A)^k) \leq (\frac{s}{|c|})^k.$$

Επομένως,

$$\left\|p_{k p_k \in \mathcal{P}_k,\ p_k(0)=1}(A)\right\| \leq 2\left(\frac{s}{|c|}\right)^k.$$

Τότε το φράγμα για το σχετικό απόλυτο σφάλμα του υπολοίπου στην k επανάληψη της GMRES παίρνει τη μορφή:

$$\frac{\|r_k\|}{\|r_0\|} \leq 2\left(\frac{s}{|c|}\right)^k.$$

Τα παραπάνω συμπεράσματα συνοψίζονται στο ακόλουθο θεώρημα.

Θεώρημα 4.11. *: Έστω r_k το υπόλοιπο στην k επανάληψη του αλγορίθμου της μεθόδου GMRES. Εάν, επιπλέον, $0 \notin \mathcal{F}(A)$ και $\mathcal{F}(A) \subseteq D = \{z \in \mathbb{C} : |z - c| \leq s\}$. Τότε*

$$\frac{\|r_k\|}{\|r_0\|} \leq 2\left(\frac{s}{|c|}\right)^k.$$

Διαφορετικά φράγματα μπορούν να προκύψουν, αν χρησιμοποιηθεί θεωρία πολυωνύμων Faber (βλ. [20], [7], [8]), όπως επίσης και με την εισαγωγή της έννοιας των "ψευδοϊδιοτιμών" (pseudo-eigenvalues) (βλ. [35]).

Στο τελευταίο μέρος αυτής της παραγράφου θα περιγράψουμε πως είναι δυνατόν, με τη βοήθεια του αλγορίθμου του Lanczos[13] για Ερμιτιανούς πίνακες,[14] να κατασκευάσουμε τη μέθοδο MINRES.

Ο αλγόριθμος του Lanczos γράφεται σε μορφή πινακων ως εξής:

$$AQ_k = Q_k T_{kk} + \beta_k q_{k+1}(\xi^k)^T = Q_{k+1}T_{k+1,k},$$

όπου Q_k είναι ένας $n \times k$ ορθοκανονικός πίνακας με στήλες τα ορθοκανονικά διανύσματα q_j, $j = 1, 2, \ldots, k+1,\ k = 1, 2, \ldots$ που παράγονται από τον αλγόριθμο. Ο T_{kk} είναι ένας τριδιαγώνιος συμμετρικός πίνακας με $T_{k+1,k}$ να έχει ως κύριο υποπίνακά του τον T_{kk} και την τελευταία γραμμή του μηδενική, εκτός του τελευταίου στοιχείου της, το οποίο ισούται με $T(k+1,k) = \beta_k(\xi^k)^T$, όπου $\beta_k = \|\widetilde{q}_{k+1}\|$ με $\widetilde{q}_{k+1}$ το διάνυσμα που είναι ορθογώνιο στα q_j, $j = 1, \ldots, k$. Θεωρώντας το $q_1 = \frac{r_0}{\tau}$ με $\tau = \|r_0\|$, η μέθοδος παράγει διαδοχικές προσεγγίσεις της λύσης της μορφής $x_k = x_0 + Q_k y_k$, όπου $Q_k y_k \in \mathcal{K}_k(A, r_0)$ και y_k είναι η λύση του προβλήματος ελαχίστων τετραγώνων:

$$\begin{array}{c} \min_{y\in\mathbb{C}^k} \|r_0 - AQ_k y\| = \min_{y\in\mathbb{C}^k} \|r_0 - Q_{k+1}T_{k+1,k}y\| = \\ \min_{y\in\mathbb{C}^k} \|Q_{k+1}(\tau\xi^1 - T_{k+1,k}y)\| = \min_{y\in\mathbb{C}^k} \|\tau\xi^1 - T_{k+1,k}y\|. \end{array} \tag{4.24}$$

Οι παραπάνω ισότητες ισχύουν από τους ορισμούς των τ, Q_k, Q_{k+1} και από το γεγονός ότι οι πίνακες Q_k και Q_{k+1} είναι ορθοκανονικοί. Το διάνυσμα $Q_k y_k$ αποτελεί έναν γραμμικό συνδυασμό των διανυσμάτων βάσης του χώρου Krylov, $K_k(A, r_0)$ =span$\{r_0, Ar_0, \ldots, A^{k-1}r_0\}$. Σε αυτή την περίπτωση δε χρειάζεται να αποθηκεύουμε σε κάθε επανάληψη περισσότερα του ενός ορθοκανονικά διανύσματα που κατασκευάζει ο αλγόριθμος του Lanczos. Μπορούμε να λύσουμε το πρόβλημα με απλή αντικατάσταση ως εξής: Ορίζουμε $P_k \equiv Q_k R_{k\times k}^{-1}$ και έχοντας $p_0 = \tau q_1$, μπορούμε να υπολογίσουμε τις στήλες του P_k από τη σχέση

$$p_{k-1} = \frac{\left(q_k - \beta_{k-2}^{(k-1)} p_{k-2} - \beta_{k-3}^{(k-1)} p_{k-3}\right)}{\beta_{k-1}^{(k-1)}},$$

όπου $\beta_{k-l}^{(k-1)},\ l = 1, 2, 3$ είναι το $(k-l+1, k)$ στοιχείο του $R_{k\times k}$. Βλέπουμε, λοιπόν, ότι η προσέγγιση της λύσης στην k επανάληψη δίδεται από τη σχέση:

$$x_k = x_0 + P_k \tau(F\xi^1)_{k\times 1} = x_{k-1} + \alpha_{k-1}p_{k-1},$$

[13] Χωρίς "look-ahead".

[14] Στην περίπτωση των Ερμιτιανών πινάκων μπορεί να αποδειχθεί εύκολα, με επαγωγή, ότι ο αλγόριθμος του Lanczos ταυτίζεται με αυτόν του Arnoldi.

όπου α_{k-1} είναι η k συνιστώσα του διανύσματος $\tau(F\xi^1)_{k\times 1}$. Ολοκληρωμένος ο αλγόριθμος παρουσιάζεται στο Παράρτημα.

Παρατήρηση 4.7. : Με παρόμοια διαδικασία, στην περίπτωση όπου ο A είναι Ερμιτιανός και θετικά ορισμένος, μπορούμε να κατασκευάσουμε τον αλγόριθμο της μεθόδου CG, χρησιμοποιώντας και πάλι τον αλγόριθμο του Lanczos(βλ. [20]).

5 Μέθοδος Δισυζυγών Κλίσεων (Biconjugate Gradient (BiCG)) και Σχετικές Μέθοδοι

Όπως ήδη παρατηρήσαμε και στην προηγούμενη παράγραφο, η πλήρης GMRES (GMRES(n)) απαιτεί αυξημένο πλήθος πράξεων και χώρο μνήμης ανά επανάληψη. Για τη βελτίωση των αυξημένων απαιτήσεων της πλήρους GMRES, παρουσιάσαμε μεθόδους, όπως η Orhomin(j) και η GMRES(j), οι αλγόριθμοι των οποίων βελτιώνουν σημαντικά το κόστος επίλυσης. Είδαμε, όμως, ότι υπάρχουν περιπτώσεις στις οποίες οι παραπάνω μέθοδοι αποτυγχάνουν, όπως για παράδειγμα στην περίπτωση όπου η νόρμα του υπολοίπου σταματάει να μειώνεται. Μία τέτοια περίπτωση έχουμε όταν ο συντελεστής $\alpha_{k-1} = \frac{(r_{k-1}, Ap_{k-1})}{(Ap_{k-1}, Ap_{k-1})}$ που συνδέει δύο διαδοχικά υπόλοιπα είναι μηδέν.

Στην παράγραφο αυτή, θα μελετήσουμε μεθόδους οι οποίες συμπεριφέρονται καλύτερα από τις παραπάνω[15] και, επιπλέον, η πιθανότητα αποτυχίας εξαλείφεται με τη χρήση της τεχνικής look-ahead [29] βημάτων στους αλγορίθμους τους. Η χρήση, όμως, τέτοιων τεχνικών μπορεί να έχει ως συνέπεια την αύξηση των επαναλήψεων και του χώρου μνήμης που απαιτείται για τη σύγκλιση των μεθόδων αυτών. Επιπλέον, σε αυτές τις περιπτώσεις, δεν έχουμε εκ των προτέρων θεωρητικές εκτιμήσεις σφαλμάτων.

5.1 Αμφίπλευρος Αλγόριθμος του Lanczos

Στο προηγούμενο κεφάλαιο είδαμε ότι στην περίπτωση όπου ο πίνακας A είναι Ερμιτιανός, ο αλγόριθμος του Lanczos έδιδε την ορθοκανονικοποίηση της βάσης Krylov, με μια ακολουθία τριών βημάτων. Στην περίπτωση, όμως, που ο πίνακας δεν είναι Ερμιτιανός, δε γίνεται κάτι τέτοιο. Μπορούμε, όμως, να κατασκευάσουμε μία ακολουθία τριών βημάτων για τους πίνακες A και A^H χωριστά, δημιουργώντας διορθογώνιες βάσεις για τους χώρους Krylov που αντιστοιχούν στους παραπάνω πίνακες.

Έστω ότι $\mathcal{K}_k(B, v)$ είναι ο χώρος Krylov $\text{span}\{v, Bv, \ldots, B^{k-1}v\}$. Κατασκευάζουμε δύο σύνολα διανυσμάτων

$$v_1, v_2, \ldots, v_k \in \mathcal{K}_k(A, r_0)$$

και

$$w_1, w_2, \ldots, w_k \in \mathcal{K}_k(A^H, \widehat{r}_0),$$

[15] Με την έννοια ότι έχουν σταθερό χώρο εργασίας.

με την ιδιότητα ότι

$$(v_i, w_j) = 0 \quad \forall i, j = 1, \ldots, k, \ i \neq j.$$

Τα διανύσματα $r_0, \widehat{r}_0$ αναφέρονται στα διανύσματα–υπόλοιπα της αρχικής επανάληψης x_0 για τα συστήματα $Ax = b$ και $A^H x = b$, αντίστοιχα. Ο τρόπος κατασκευής των δύο παραπάνω συνόλων διανυσμάτων καλείται "αμφίπλευρος αλγόριθμος του Lanczos" και παρουσιάζεται αναλυτικά στο Παράρτημα. Στις επόμενες παραγράφους, όταν θα αναφερόμαστε στον αλγόριθμο του Lanczos, θα εννοούμε τον προαναφερθέντα "αμφίπλευρο αλγόριθμο του Lanczos".

Θεωρώντας V_k και W_k τους πίνακες με στήλες τα $v_1, v_2, \ldots, v_k$ και $w_1, w_2, \ldots, w_k$ αντίστοιχα, η k επανάληψη του αλγορίθμου του Lanczos σε μορφή πινάκων είναι:

$$AV_k = V_k T_{kk} + \gamma_k v_{k+1} (\xi^k)^T = V_{k+1} T_{k+1,k}, \tag{5.1}$$

$$A^H W_k = W_k T_{kk}^H + \bar{\beta}_k w_{k+1} (\xi^k)^T = W_{k+1} \widetilde{T}_{k+1,k}, \tag{5.2}$$

με

$$T_{kk} = \begin{pmatrix} \alpha_1 & \beta_1 & & \\ \gamma_2 & \alpha_2 & & \\ & & \ddots & \beta_{k-1} \\ & & \gamma_k & \alpha_k \end{pmatrix}.$$

Οι πίνακες $T_{k+1,k}$ και $\widetilde{T}_{k+1,k}$ είναι αυτοί που έχουν ως κύριους υποπίνακες τους T_{kk} και T_{kk}^H αντίστοιχα, ενώ η τελευταία τους γραμμή είναι μηδενική, εκτός από τα στοιχεία τους $T(k+1, k) = \gamma_{k+1}$ και $\widetilde{T}(k+1, k) = \bar{\beta}_k$. Από την κατασκευή των v_i, w_i, μέσω του αλγορίθμου του Lanczos, προκύπτει η συνθήκη διορθογωνιότητας

$$V_k^H W_k = I_k.$$

Έχουμε, λοιπόν, το επόμενο θεώρημα:

Θεώρημα 5.1. *: Εάν τα διανύσματα που παράγονται από τον αμφίπλευρο αλγόριθμο του Lanczos είναι καλά ορισμένα στις $k+1$ πρώτες επαναλήψεις, δηλαδή, εάν $(v_j, w_j) \neq 0$, $j = 1, 2, \ldots, k+1$, τότε*

$$(v_i, w_j) = 0 \quad \forall i, j = 1, \ldots, k+1, \ i \neq j. \tag{5.3}$$

Η απόδειξη του παραπάνω θεωρήματος γίνεται με επαγωγή στο k (βλ. [20]).

Παρατήρηση 5.1. : Τα διανύσματα που παράγονται από τον αλγόριθμο του Lanczos δεν ορίζονται σε δύο περιπτώσεις: Πρώτον, στην περίπτωση όπου κάποιο από τα $\widetilde{v}_{k+1}$ και $\widetilde{w}_{k+1}$ είναι μηδέν. Τότε είτε τα διανύσματα $v_1, v_2, \ldots, v_k$ παράγουν έναν $A-$αναλλοίωτο υπόχωρο είτε τα $w_1, w_2, \ldots, w_k$ παράγουν έναν A^H-αναλλοίωτο υπόχωρο. Αυτός ο τρόπος τερματισμού του αλγορίθμου καλείται "κανονικός" τερματισμός.

Η δεύτερη περίπτωση τερματισμού καλείται "κατάρρευση" (break-down) και συμβαίνει όταν $(\widetilde{v}_{k+1}, \widetilde{w}_{k+1}) = 0$ αλλά κανένα εκ των $\widetilde{v}_{k+1}$ και $\widetilde{w}_{k+1}$ δεν είναι μηδέν. Έτσι, δεν υπάρχουν διανύσματα $\widetilde{v}_{j+1} \in \mathcal{K}_{j+1}(A, r_0)$ και $\widetilde{w}_{j+1} \in \mathcal{K}_{j+1}(A^H, \widehat{r_0})$ που να ικανοποιούν τις συνθήκες $(v_{j+1}, w_i) = (w_{j+1}, v_i)$ $= 0, \forall i \leq j$. Είναι δυνατόν, βέβαια, στην $j+1$ επανάληψη να μην υπάρχουν τέτοια διανύσματα αλλά σε κάποια επανάληψη στη συνέχεια να υπάρχουν $\widetilde{v}_{j+l} \in \mathcal{K}_{j+l}(A, r_0)$ και $\widetilde{w}_{j+l} \in \mathcal{K}_{j+l}(A^H, \widehat{r_0})$ τ.ω. να είναι ορθογώνια στους χώρους $\mathcal{K}_{j+l-1}(A^H, \widehat{r_0})$ και $\mathcal{K}_{j+l-1}(A, r_0)$ αντίστοιχα. Η διαδικασία που ουσιαστικά προκαλεί "μικρές διαταραχές" στα ενδιάμεσα βήματα του αλγορίθμου του Lanczos, στα οποία υποθέτουμε ότι θα μπορούσαμε να έχουμε "κατάρρευση", καλείται "αμφίπλευρος αλγόριθμος του Lanczos με look-ahead" (βλ. [26], [29], [4]).

5.2 Μέθοδος Δισυζυγών Κλίσεων (Biconjugate Gradient (BiCG))

Στην προηγούμενη παράγραφο, είδαμε αναλυτικά τον αμφίπλευρο αλγόριθμο του Lanczos για μη-Ερμιτιανούς πίνακες. Σε αυτήν την παράγραφο, θα μελετήσουμε μία νέα μέθοδο η οποία στηρίζεται στον αλγόριθμο του Lanczos και καλείται Μέθοδος Δισυζυγών Κλίσεων (Biconjugate Gradient (BiCG)).

Έστω $A \in \mathbb{C}^{n,n}$ μη-Ερμιτιανός, γενικά, πίνακας και έστω $Ax = b$ το σύστημα που θέλουμε να επιλύσουμε με $b \in \mathbb{C}^n$. Θεωρούμε ότι η ακολουθία διανυσμάτων που παράγει ο αλγόριθμος του Lanczos είναι πλήρης, δηλαδή δεν έχουμε περίπτωση "κατάρρευσης". Τα διανύσματα βάσης του χώρου Krylov, που αναφέραμε προηγουμένως και τα οποία παράγονται από τον αλγόριθμο του Lanczos, μπορούν να χρησιμοποιηθούν στην προσέγγιση της λύσης του γραμμικού συστήματος. Εδώ, όπως και στην περίπτωση των μεθόδων GMRES, MINRES και CG, η προσέγγιση x_k της λύσης του γραμμικού συστήματος δίδεται από μία σχέση της μορφής:

$$x_k = x_0 + V_k y_k. \tag{5.4}$$

Η παραπάνω σχέση αποτελεί μία διαδικασία διαδοχικών βελτιώσεων της αρχικής προσέγγισης x_0, μέσω ενός γραμμικού συνδυασμού των διανυσμάτων βάσης του χώρου Krylov $\mathcal{K}_k(A, r_0)$. Μία δυνατή επιλογή του y_k είναι να απαιτήσουμε το υπόλοιπο στην k επανάληψη,

$$r_k = r_0 - AV_k y_k,$$

να είναι ορθογώνιο στα $w_1, w_2, \ldots, w_k$, που κατασκεύαζονται από τον αλγόριθμο του Lanczos. Αυτό που απαιτούμε είναι το υπόλοιπο r_k να είναι ορθογώνιο στον χώρο $\mathcal{K}_k(A^H, \widehat{r_0})$. Λαμβάνουμε έτσι την εξίσωση

$$W_k^H r_k = W_k^H r_0 - W_k^H A V_k y_k = 0. \tag{5.5}$$

Από την (5.1) και την (5.3) έχουμε ότι

$$W_k^H A V_k = W_k^H V_{k+1} T_{k+1,k} = T_{kk}$$

και

$$W_k^H r_0 = \beta \xi^1,\ \beta = \|r_0\|.$$

Από τα παραπάνω, παρατηρούμε ότι το y_k προκύπτει ως λύση του τριδιαγώνιου συστήματος

$$T_{kk} y_k = \beta \xi^1.$$

Παρατήρηση 5.2. : Αν ο πίνακας T_{kk} είναι μη-αντιστρέψιμος τότε το παραπάνω σύστημα δεν έχει λύση και αυτό σημαίνει ότι δεν υπάρχει διάνυσμα x_k τ.ω. $W_k^H r_k = 0$.

Ο πλήρης αλγόριθμος της Biconjugate Gradient μεθόδου δίδεται στο Παράρτημα.

5.3 Μέθοδος Ημιελαχίστου Υπολοίπου (Quasi-minimal Residual (QMR))

Στη μέθοδο Ημιελαχίστου Υπολοίπου (Quasi-minimal Residual (QMR)) οι προσεγγίσεις της λύσης δίδονται και σε αυτή την περίπτωση από τη σχέση $x_k = x_0 + V_k y_k$, όπου το y_k είναι τέτοιο ώστε να ελαχιστοποιεί μία ποσότητα που σχετίζεται με την Ευκλείδεια νόρμα του υπολοίπου. Έχουμε, λοιπόν, ότι η έκφραση

$$r_k = r_0 - AV_k y_k$$

μπορεί να γραφεί στη μορφή

$$r_k = V_{k+1}(\beta\xi^1 - T_{k+1,k}y_k),$$

όπου V_{k+1} και $T_{k+1,k}$ είναι οι πίνακες των οποίων τα στοιχεία προέρχονται από τον αλγόριθμο του Lanczos. Οι στήλες του πίνακα V_{k+1} **δεν** είναι ορθογώνιες, πράγμα που δημιουργεί πρόβλημα στην επιλογή y_k τέτοιου ώστε να ελαχιστοποιεί τη $\|r_k\|$. Λαμβάνοντας νόρμες, παίρνουμε τη σχέση

$$\|r_k\| \leq \|V_{k+1}\| \left\|\beta\xi^1 - T_{k+1,k}y_k\right\|. \tag{5.6}$$

Σε αυτήν την περίπτωση, το y_k επιλέγεται έτσι ώστε να ελαχιστοποιεί τον παράγοντα $\|\beta\xi^1 - T_{k+1,k}y_k\|$. Έχοντας, επιπλέον, γνωστό ότι κάθε στήλη του πίνακα V_{k+1} έχει νόρμα ίση με ένα, λαμβάνουμε τη σχέση $\|V_k\| \leq \sqrt{k+1}$, άρα,

$$\|r_k\| \leq \sqrt{k+1}\left\|\beta\xi^1 - T_{k+1,k}y_k\right\|.$$

Στη μέθοδο QMR λύνουμε το πρόβλημα των ελαχίστων τετραγώνων

$$\min_{y\in\mathbb{C}^k}\left\|\beta\xi^1 - T_{k+1,k}y\right\|,$$

το οποίο έχει πάντα λύση ακόμα και αν ο πίνακας T_{kk} είναι μη-αντιστρέψιμος, εκτός βέβαια από την περίπτωση όπου ο αλγόριθμος του Lanczos καταρρέει.

Στη συνέχεια, παρουσιάζουμε ένα θεώρημα το οποίο συνδέει τις νόρμες των υπολοίπων των μεθόδων GMRES και QMR.

Θεώρημα 5.2. *: Αν με r_k^G συμβολίσουμε το υπόλοιπο της μεθόδου GMRES στην k επανάληψη και με r_k^Q το αντίστοιχο της μεθόδου QMR, τότε:*

$$\left\|r_k^Q\right\| \leq \kappa(V_{k+1})\left\|r_k^G\right\|, \tag{5.7}$$

όπου V_{k+1} είναι ο πίνακας με στήλες τα διανύσματα βάσης που κατασκευάζονται από τον αλγόριθμο του Lanczos και κ συμβολίζει το δείκτη κατάστασης.

Δυστυχώς, δεν υπάρχουν εκ των προτέρων φράγματα για τον δείκτη κατάστασης του V_{k+1}, ο οποίος ενδέχεται να είναι αρκετά μεγάλος, με αποτέλεσμα το φράγμα να είναι αρκετά απαισιόδοξο.

Η μέθοδος που ακολουθείται για την επίλυση του προβλήματος ελαχίστων τετραγώνων είναι αυτή που είδαμε στο προηγούμενο κεφάλαιο, στην περίπτωση

της MINRES μεθόδου. Δηλαδή, η εφαρμογή των στροφών Givens στον τριδιαγώνιο πίνακα $T_{k+1,k}$, με σκοπό το μηδενισμό των στοιχείων κάτω από τη διαγώνιο. Καταλήγουμε, τελικά, ότι το πρόβλημα ελαχίστων τετραγώνων

$$\min_{y\in\mathbb{C}^k} \left\|\beta\xi^1 - T_{k+1,k}y\right\|$$

$$= \min_{y\in\mathbb{C}^k} \left\|\beta\xi^1 - F^H Ry\right\| = \min_{y\in\mathbb{C}^k} \left\|\beta F\xi^1 - Ry\right\|$$

ισοδυναμεί με την επίλυση του συστήματος

$$R_{k\times k}y_k = g_{k\times 1}, \tag{5.8}$$

όπου $g_{k\times 1}$ είναι ένα διάνυσμα με στοιχεία τα πρώτα k στοιχεία του διανύσματος $g = (F_{k,k+1}\cdots F_{12})^H\beta\xi$[16] και $R_{k\times k}$ είναι ένας άνω τριγωνικός πίνακας με δύο υπερδιαγωνίους. Ακολουθώντας τη διαδικασία της μεθόδου MINRES, θέτοντας, δηλαδή, $P_k \equiv V_k R_{k\times k}^{-1}$, και $y = y_k$ τη λύση του συστήματος (5.8), έχουμε ότι

$$x_k = x_0 + V_k y_k = x_0 + P_k g_{k\times 1}.$$

Επομένως, και η προηγούμενη επανάληψη έχει τη μορφή: $x_k = x_0 + P_{k-1}g_{(k-1)\times 1}$. Συνδυάζοντας τις δύο παραπάνω σχέσεις, λαμβάνουμε έναν αναδρομικό τύπο για τις προσεγγίσεις της λύσης:

$$x_k = x_{k-1} + \alpha_{k-1}p_{k-1},$$

όπου α_{k-1} είναι το k στοιχείο του διανύσματος g. Η παραπάνω έκφραση μας επιτρέπει να παρατηρήσουμε ότι η προσέγγιση x_k ισούται με τη x_{k-1} διορθωμένη κατά ένα διάνυσμα ανάλογο της διεύθυνσης p_{k-1}. Από τον ορισμό του $P_k = (p_0\ p_1\ \ldots\ p_{k-1})$, έχουμε ότι

$$p_{k-1} = \frac{1}{\beta_k}(v_k - \beta_{k-1}p_{k-2} - \beta_{k-2}p_{k-3}),$$

όπου $\beta_l = R(k+1-l,k),\ l = 1,2,3$. Ο πλήρης αλγόριθμος της QMR μεθόδου παρουσιάζεται στο Παράρτημα.

[16] Οι $F_{12}, F_{23}, \ldots, F_{k,k+1}$ είναι οι πίνακες−στροφή με εκφράσεις αυτές που δόθηκαν στις (3.5)–(3.8).

5.4 Τετραγωνική Μέθοδος Συζυγών Κλίσεων (Conjugate Gradient Squared) (CGS))

Στις μεθόδους BiCG και QMR, σε κάθε επανάληψη, απαιτούνται πολλαπλασιασμοί των πινάκων A και A^H επί διανύσματα. Το γεγονός αυτό απαιτεί επιπλέον κόστος σε πράξεις. Επίσης, σε πολλές περιπτώσεις, ο πολλαπλασιασμός με τον A^H δεν είναι "βολικός" σε σύγκριση με αυτόν του A (βλ. [20])[17]. Για αυτούς τους λόγους, προσπαθούμε να κατασκευάσουμε επαναληπτικές μεθόδους, τέτοιες ώστε από το ένα μέρος να απαιτούν πολλαπλασιασμό **μόνο** με τον πίνακα A και από το άλλο μέρος να παράγουν καλές προσεγγίσεις της λύσης από τους χώρους Krylov, που κατασκευάζονται από τον πολλαπλασιασμό του πίνακα A επί διάνυσμα.

Μία τέτοια μέθοδος, η οποία βελτιώνει τις μεθόδους BiCG και QMR, είναι η Τετραγωνική Μέθοδος Συζυγών Κλίσεων (Conjugate Gradient Squared) (CGS)).

Από τη μέθοδο BiCG έχουμε ότι τα $r_k, \widehat{r}_k$ και τα $p_k, \widehat{p}_k$ μπορούν να εκφραστούν στις μορφές:

$$r_k = \phi_k(A)r_0, \quad \widehat{r}_k = \bar{\phi}_k(A^H)\widehat{r}_0 \quad p_k = \psi_k(A)r_0, \quad \widehat{p}_k = \bar{\psi}_k(A^H)\widehat{r}_0 \qquad (5.9)$$

για συγκεκριμένα πολυώνυμα ϕ_k και ψ_k βαθμού k. Αν η μέθοδος της BiCG συγκλίνει καλώς, δηλαδή, αν η $\|\phi_k(A)r_0\|$ είναι αρκετά μικρή, τότε περιμένουμε η $\|\phi_k^2(A)r_0\|$ να είναι ακόμα μικρότερη. Αυτό ισχύει γιατί, εάν υποθέσουμε ότι το πολυώνυμο $\phi_k(A)$ είναι ένας τελεστής μείωσης του αρχικού υπολοίπου r_0, τότε αυτό που θα συμβαίνει είναι ότι καθώς θα αυξάνεται η διάσταση του χώρου Krylov, $K_k(A, r_0)$, ο τελεστής θα μειώνει το εκάστοτε υπόλοιπο, με αποτέλεσμα η μέθοδος να συγκλίνει. Κάνοντας λοιπόν αυτή την παραδοχή για το $\phi_k(A)$, είναι λογικό να περιμένουμε ότι η εφαρμογή του τελεστή αυτού στο διάνυσμα $\phi_k(A)r_0$ θα προκαλεί μεγαλύτερη μείωση του υπολοίπου. Στην αντίθετη περίπτωση, δηλαδή, της μη-σύγκλισης του υπολοίπου της BiCG, $\phi_k(A)r_0$, η περαιτέρω εφαρμογή του τελεστή δημιουργεί επιπλέον προβλήματα

[17]Με την έκφραση ότι ο πολλαπλασιασμός του πίνακα A με ένα διάνυσμα είναι "βολικός" σε σύγκριση με αυτόν του A^H, εννοούμε το εξής: Σε αρκετές περιπτώσεις, λόγω του τρόπου αποθήκευσης των πινάκων στον Υπολογιστή, υπάρχουν υπορουτίνες που λαμβάνουν υπόψη και εκμεταλλεύονται το συγκεκριμένο γεγονός κατά τον πολλαπλασιασμό πίνακα επί διάνυσμα. Αντιθέτως δεν υπάρχει αντίστοιχη πρόβλεψη για τον πολλαπλασιασμό του A^H επί διάνυσμα, πράγμα που καθιστά τον αντίστοιχο πολλαπλασιασμό στην πράξη πιο χρονοβόρο. Επίσης, στην περίπτωση που χρησιμοποιείται παράλληλη επεξεργασία, ο πολλαπλασιασμός του πίνακα A^H επί διάνυσμα απαιτεί επιπλέον χρόνο επικοινωνίας μεταξύ των επεξεργαστών.

στη σύγκλιση της μεθόδου. Περισσότερα πάνω σ' αυτό το θέμα θα αναπτυχθούν παρακάτω.

Θα πρέπει να τονιστεί ότι, εάν ο υπολογισμός του $\phi_k^2(A)r_0$ γίνεται το ίδιο εύκολα με τον $\phi_k(A)r_0$, τότε η αντίστοιχη μέθοδος θα μπορέσει να χαρακτηριστεί καλύτερη από τις προηγούμενες, αφού θα ικανοποιεί τις προϋποθέσεις που έχουμε θέσει. Σε αυτή την παρατήρηση στηρίζεται και η μέθοδος της CGS.

Από τη μέθοδο της BiCG και τις (5.9) έχουμε ότι:

$$\phi_k(A)r_0 = \phi_{k-1}(A)r_0 - \alpha_{k-1}A\psi_{k-1}(A)r_0 \tag{5.10}$$

και

$$\psi_k(A)r_0 = \phi_k(A)r_0 + \beta_k\psi_{k-1}(A)r_0, \tag{5.11}$$

όπου

$$\alpha_{k-1} = \frac{(\phi_{k-1}(A)r_0, \bar{\phi}_{k-1}(A^H)\widehat{r}_0)}{(A\psi_{k-1}(A)r_0, \bar{\psi}_{k-1}(A^H)\widehat{r}_0)} = \frac{(\phi_{k-1}^2(A)r_0, \widehat{r}_0)}{(A\psi_{k-1}^2(A)r_0, \widehat{r}_0)}.$$

Από τις σχέσεις (5.10), (5.11) θεωρούμε τα πολυώνυμα

$$\phi_k(z) = \phi_{k-1}(z) - \alpha_{k-1}z\psi_{k-1}(z) \tag{5.12}$$

και

$$\psi_k(z) = \phi_k(z) + \beta_k\psi_{k-1}(z). \tag{5.13}$$

Τετραγωνίζοντας τις σχέσεις (5.12) και (5.13) λαμβάνουμε:

$$\phi_k^2(z) = \phi_{k-1}^2(z) - 2\alpha_{k-1}z\phi_{k-1}(z)\psi_{k-1}(z) + \alpha_{k-1}^2z^2\psi_{k-1}^2(z)$$

και

$$\psi_k^2(z) = \phi_k^2(z) + 2\beta_k\phi_k(z)\psi_{k-1}(z) + \beta_k^2\psi_{k-1}^2(z).$$

Επίσης, με πολλαπλασιασμό των ίδιων σχέσεων λαμβάνουμε:

$$\phi_k(z)\psi_k(z) = \phi_k^2(z) + \beta_k\phi_k(z)\psi_{k-1}(z)$$

και

$$\begin{aligned}\phi_k(z)\psi_{k-1}(z) &= \phi_{k-1}(z)\psi_{k-1}(z) - \alpha_{k-1}z\psi_{k-1}^2(z)\\ &= \phi_{k-1}^2(z) + \beta_{k-1}\phi_{k-1}(z)\psi_{k-2}(z) - \alpha_{k-1}z\psi_{k-1}^2(z)\end{aligned}$$

Ορίζοντας $\Phi_k \equiv \phi_k^2$, $\Theta_k \equiv \phi_k\psi_{k-1}$, $\Psi_{k-1} \equiv \psi_{k-1}^2$ έχουμε τις σχέσεις:

$$\Phi_k(z) = \Phi_{k-1}(z) - 2\alpha_{k-1}z(\Phi_{k-1}(z) + \beta_{k-1}\Theta_{k-1}(z)) + \alpha_{k-1}^2z^2\Psi_{k-1}(z),$$

$$\Theta_k(z) = \Phi_{k-1}(z) + \beta_{k-1}\Theta_{k-1}(z) - \alpha_{k-1}z\Psi_{k-1}(z)$$

και

$$\Psi_k(z) = \Phi_k(z) + 2\beta_k\Theta_k(z) + \beta_k^2\Psi_{k-1}.$$

Στη συνέχεια, χρησιμοποιώντας τις παραπάνω εκφράσεις, "μεταλλάσσουμε" τον αλγόριθμο της BiCG σε μία νέα μορφή, παράγοντας αυτόν της CGS. Ο πλήρης αλγόριθμος της CGS παρουσιάζεται αναλυτικά στο Παράρτημα.

Παρατήρηση 5.3. : Ο αλγόριθμος της CGS απαιτεί δύο πολλαπλασιασμούς πίνακα επί διάνυσμα σε κάθε επανάληψη αλλά δεν απαιτεί, πλέον, πολλαπλασιασμούς με τον A^H. Στις περιπτώσεις όπου η BiCG συγκλίνει καλώς, τότε η CGS απαιτεί τις μισές επαναλήψεις για τη σύγκλισή της και, επομένως, και το μισό κόστος εργασίας. Εάν όμως η νόρμα του υπολοίπου της BiCG, $\|r_k\|$, παρουσιάζει αυξομειωτικές διακυμάνσεις από επανάληψη σε επανάληψη, τότε η συμπεριφορά της νόρμας του υπολοίπου της CGS, από επανάληψη σε επανάληψη, θα παρουσιάζει μεγαλύτερες αυξομειώσεις (βλ. [24], [36], [37]).

Συγκρίνοντας τώρα το υπόλοιπο της CGS, r_k^S, με αυτό της BiCG, r_k^B, έχουμε τη σχέση

$$r_k^S = P_k^2(A)r_0 = P_k(A)P_k(A)r_0 = P_k(A)r_k^B. \tag{5.14}$$

Υποθέτοντας ότι ο A είναι διαγωνοποιήσιμος, έχουμε ότι υπάρχει $S \in \mathbb{C}^{n,n}$ αντιστρέψιμος πίνακας, τέτοιος ώστε $A = S\Lambda S^{-1}$. Θέτοντας αυτή τη σχέση στην (5.14) έχουμε ότι

$$r_k^S = SP_k(\Lambda)S^{-1}r_k^B. \tag{5.15}$$

Λαμβάνοντας νόρμες, βρίσκουμε ότι

$$\left\|r_k^S\right\| \leq \left\|SP_k(\Lambda)S^{-1}\right\| \left\|r_k^B\right\| \leq \kappa(S)\left\|P_k(\Lambda)\right\| \left\|r_k^B\right\|, \tag{5.16}$$

όπου με $\kappa(S)$ συμβολίζουμε το δείκτη κατάστασης του πίνακα S. Αυτό λοιπόν που παρατηρεί κάποιος είναι: Πρώτον, η εξάρτηση του συντελεστή $\kappa(S)\left\|P_k(\Lambda)\right\|$ από τον δείκτη κατάστασης του πίνακα S, ο οποίος μπορεί να είναι αρκετά μεγάλος, και δεύτερον, από την τιμή $\left\|P_k(\Lambda)\right\|$, η οποία για κάποια ιδιοτιμή μπορεί να είναι μεγάλη. Η εξάρτηση της έκφρασης του φράγματος του δεξιού μέλους της σχέσεως (5.16) από τους δύο αυτούς παράγοντες ίσως να είναι η αιτία του φαινομένου αυτών των αυξομειωτικών τάσεων στη σύγκλιση της μεθόδου CGS.

Για την κατά το δυνατόν εξάλειψη ή ελάττωση αυτών των αυξομειώσεων στην περίπτωση σύγκλισης, χρησιμοποιούμε μεθόδους ευστάθειας της σύγκλισης της CGS, οι οποίες παράγουν νέες μεθόδους που έχουν πάντα ως βάση την BiCG.

5.5 Ευσταθειοποιημένη Μέθοδος Δισυζυγών Κλίσεων (Biconjugate Gradient Stabilized (BiCGSTAB))

Για την αποφυγή των αυξομειώσεων της νόρμας του υπολοίπου κατά τη σύγκλιση της CGS μεθόδου, θεωρούμε την παρακάτω τροποποιημένη μορφή υπολοίπου

$$r_k = \chi_k(A)\phi_k(A)r_0, \tag{5.17}$$

όπου ϕ_k είναι το πολυώνυμο της BiCG ενώ το χ_k επιλέγεται έτσι ώστε να διατηρεί τη νόρμα του υπολοίπου σε κάθε επανάληψη μικρή, ακολουθώντας την ταχεία σύγλιση της CGS. Έτσι, εάν το πολυώνυμο $\chi_k(z)$ είναι της μορφής

$$\chi_k(z) = (1-\omega_k z)(1-\omega_{k-1}z)\dots(1-\omega_1 z), \tag{5.18}$$

τότε οι συντελεστές $\omega_j,\ j = 1, 2, \dots, k$, μπορεί να επιλεγούν σε κάθε επανάληψη έτσι, ώστε να ελαχιστοποιούν τη

$$\|r_j\| = \|(I-\omega_j A)\chi_{j-1}(A)\phi_j(A)r_0\|\,,\ j = 1, \dots, k. \tag{5.19}$$

Αυτή η διαδικασία οδηγεί στην καλούμενη "Ευσταθειοποιημένη Μέθοδο Δισυζυγών Κλίσεων" (Biconjugate Gradient Stabilized (BiCGSTAB)). Πιο συγκεκριμένα, ορίζουμε τα $\phi_k(A)r_0$, $\psi_k(A)r_0$ από τις σχέσεις (5.10), (5.11) και θεωρούμε τα τροποποιημένα διανύσματα

$$r_k = \chi_k(A)\phi_k(A)r_0, \quad p_k = \chi_k(A)\psi_k(A)r_0$$

για το υπόλοιπο και το τυχαίο διάνυσμα–διεύθυνσης στην k επανάληψη, αντίστοιχα. Από τις σχέσεις (5.18), (5.10), (5.11) λαμβάνουμε τις παρακάτω μορφές για τα r_k και p_k :

$$\begin{aligned} r_k &= (I-\omega_k A)\chi_{k-1}(A)[\phi_{k-1}(A) - \alpha_{k-1}A\psi_{k-1}(A)]r_0 \\ &= (I-\omega_k A)[r_{k-1} - \alpha_{k-1}Ap_{k-1}] \end{aligned}$$

και

$$\begin{aligned} p_k &= \chi_k(A)[\phi_k(A) + \beta_k\psi_{k-1}(A)]r_0 \\ &= r_k + \beta_k(I-\omega_k A)p_{k-1}. \end{aligned}$$

Τέλος, χρειάζεται να εκφράσουμε τους συντελεστές α_{k-1} και β_k της μεθόδου CGS της προηγούμενης παραγράφου, συναρτήσει των νέων διανυσμάτων r_k, p_k.

Χρησιμοποιώντας τις διορθογώνιες ιδιότητες των πολυωνύμων της BiCG, δηλαδή,

$$(\phi_k(A)r_0, A^{H^j}\widehat{r}_0) = (A\psi_k(A)r_0, A^{H^j}\widehat{r}_0) = 0,\ j = 0, 1, \ldots, k-1,$$

και τις σχέσεις (5.10), (5.11), παράγονται οι επόμενες εκφράσεις για τα εσωτερικά γινόμενα, που παρουσιάζονται συναρτήσει των συντελεστών $\alpha_{j-1}\ \beta_j,\ j = 1, 2, \ldots, k$,

$$(\phi_{k-1}(A)r_0, \bar{\phi}_{k-1}(A^H)\widehat{r}_0) = (-1)^{k-1}\alpha_{k-2}\ldots\alpha_0(\phi_{k-1}(A)r_0, A^{H^{k-1}}\widehat{r}_0)$$

και

$$(A\psi_{k-1}(A)r_0, \bar{\psi}_{k-1}(A)r_0, \bar{\psi}_{k-1}(A^H)\widehat{r}_0) = (-1)^{k-1}\alpha_{k-2}\ldots\alpha_0(\psi_{k-1}(A)r_0, A^{H^{k-1}}\widehat{r}_0).$$

Από τις ίδιες συνθήκες, έχουμε ότι τα διανύσματα–υπόλοιπα και τυχαίων διευθύνσεων ικανοποιούν τις σχέσεις:

$$(r_{k-1}, \widehat{r}_0) = (\phi_{k-1}(A)r_0, \bar{\chi}_{k-1}(A^H)\widehat{r}_0) = (-1)^{k-1}\omega_{k-1}\ldots\omega_1(\phi_{k-1}(A)r_0, A^{H^{k-1}}\widehat{r}_0)$$

και

$$(Ap_{k-1}, \widehat{r}_0) = (A\psi_{k-1}(A)r_0, \bar{\chi}_{k-1}(A^H)\widehat{r}_0) = (-1)^{k-1}\omega_{k-1}\ldots\omega_1(A\psi_{k-1}(A)r_0, A^{H^{k-1}}\widehat{r}_0),$$

αντίστοιχα. Έτσι οι εκφράσεις των συντελεστών α_{k-1}, β_k της BiCG μεθόδου μπορούν να αντικατασταθούν από τις παρακάτω:

$$\alpha_{k-1} = \frac{(r_{k-1}, \widehat{r}_0)}{(Ap_{k-1}, \widehat{r}_0)}, \quad \beta_k = \frac{\alpha_{k-1}}{\omega_k}\frac{(r_k, \widehat{r}_0)}{(r_{k-1}, \widehat{r}_0)}, \quad k = 1, 2, \ldots.$$

Ο πλήρης αλγόριθμος της BiCGSTAB μεθόδου παρουσιάζεται στο Παράρτημα.

6 $p-$Κυκλικοί Πίνακες

6.1 Εισαγωγή

Έστω ότι θέλουμε να επιλύσουμε ένα γραμμικό σύστημα της μορφής

$$Ax = b, \tag{6.1}$$

με $A \in \mathbb{C}^{n,n},\ \det(A) \neq 0$, και $b \in \mathbb{C}^n$. Έστω ότι ο πίνακας A διαχωρίζεται σε blocks ως ακολούθως:

$$A = \begin{pmatrix} A_{11} & A_{12} & \dots & A_{1p} \\ A_{21} & A_{22} & \dots & A_{2p} \\ \vdots & & & \vdots \\ A_{p1} & A_{p2} & \dots & A_{pp} \end{pmatrix}, \tag{6.2}$$

με τους διαγώνιους υποπίνακες

$$A_{ii} \in \mathbb{C}^{n_i,n_i},\ i = 1, \dots, p,\ \sum_{i=1}^{p} n_i = n,$$

να είναι αντιστρέψιμοι $(\det(A_{ii}) \neq 0)$ και τα διανύσματα x, b, να διαχωρίζονται ανάλογα με το διαχωρισμό του πίνακα A σε

$$x = (\bar{x}_1^T\, \bar{x}_2^T \dots \bar{x}_p^T)^T,\ b = (\bar{b}_1^T\, \bar{b}_2^T \dots \bar{b}_p^T)^T \quad \bar{x}_i, \bar{b}_i \in \mathbb{C}^{n_i}.$$

Θεωρούμε τη διάσπαση του πίνακα A

$$A = D - L - U, \tag{6.3}$$

όπου D ο block διαγώνιος πίνακας $D = \mathrm{diag}(A_{11}, \dots, A_{pp})$, L ο block αυστηρά κάτω τριγωνικός πίνακας και U ο block αυστηρά άνω τριγωνικός πίνακας, που αντιστοιχούν στο διαχωρισμό (6.2) του πίνακα A. Τότε ο block επαναληπτικός πίνακας του Jacobi που προκύπτει από την παραπάνω διάσπαση δίδεται από τον τύπο

$$B_J = I - D^{-1}A. \tag{6.4}$$

Η μορφή του πίνακα A με την οποία κυρίως θα ασχοληθούμε είναι η εξής:

$$A = \begin{pmatrix} A_{11} & O & O & \dots & O & A_{1p} \\ A_{21} & A_{22} & O & \dots & O & O \\ O & A_{32} & A_{33} & \dots & O & O \\ \vdots & \ddots & \ddots & \ddots & \vdots & \vdots \\ O & O & O & \dots & A_{p,p-1} & A_{pp} \end{pmatrix}, \tag{6.5}$$

με αντίστοιχο επαναληπτικό πίνακα του Jacobi τον παρακάτω:

$$B_J = \begin{pmatrix} O & O & O & \dots & O & B_{1p} \\ B_{21} & O & O & \dots & O & O \\ O & B_{32} & O & \dots & O & O \\ \vdots & \ddots & \ddots & \ddots & \vdots & \vdots \\ O & O & O & \dots & B_{p,p-1} & O \end{pmatrix}, \tag{6.6}$$

όπου

$$B_{i,i-1} = -A_{ii}^{-1} A_{i,i-1},\ i = 1, \dots, p,$$

και

$$B_{10} = -A_{11}^{-1} A_{10} = -A_{11}^{-1} A_{1p} = B_{1p}.$$

Σημείωση: Τα διαγώνια blocks στον πίνακα (6.6) είναι προφανώς τετραγωνικοί $n_i \times n_i$ πίνακες.

Ορισμός 6.1. : Έστω $B \in \mathbb{C}^{n,n}$. Ο B καλείται "ασθενώς κυκλικός δείκτη $p\,(>1)$" (weakly cyclic of index p), αν υπάρχει μεταθετικός πίνακας P τ.ω. ο πίνακας PBP^T να είναι της μορφής (6.6).

Ορισμός 6.2. : Έστω ότι ο "ασθενώς κυκλικός δείκτη $p\,(>1)$" πίνακας $B \in \mathbb{C}^{n,n}$ είναι ήδη στη μορφή (6.6). Τότε λέμε ότι ο B είναι στην "κανονική" μορφή του.

Θεώρημα 6.1. *: Έστω $B \in \mathbb{C}^{n,n}$ "ασθενώς κυκλικός δείκτη $p\,(>1)$". Τότε το χαρακτηριστικό πολυώνυμό του δίδεται από τον τύπο:*

$$\phi(t) = \det(tI - B) = t^m \prod_{j=1}^{r} (t^p - \lambda_j^p), \tag{6.7}$$

με $m + rp = n$. Οι λ_j αντιστοιχούν σε μη-μηδενικές ιδιοτιμές του πίνακα B και m είναι ένας μη-αρνητικός ακέραιος.

Βασικό στοιχείο για την απόδειξη του παραπάνω θεωρήματος είναι η εξής παρατήρηση:

Παρατήρηση 6.1. : Εάν ο πίνακας B είναι "ασθενώς κυκλικός δείκτη $p\,(>1)$", τότε υπάρχει μεταθετικός πίνακας P, που μεταθέτει τα blocks του B, τ.ω. $PBP^T = B_J$, όπου ο B_J είναι της μορφής (6.6). Οι πίνακες B και B_J είναι "όμοιοι", όπως είναι και οι B^p και B_J^p. Άρα, έχουν τις ίδιες ιδιοτιμές, όπως

έχουν και οι B^p και B_J^p. Ο πίνακας B_J^p μπορεί να επαληθευθεί ότι είναι ένας block διαγώνιος πίνακας με block διαγώνια στοιχεία τις κυκλικές μεταθέσεις του γινομένου των blocks $B_{i,i-1}$, $i = 1, \ldots, p$. Συγκεκριμένα,

$$B^p = diag(B_1^{(p)}, B_2^{(p)}, \ldots, B_p^{(p)}), \tag{6.8}$$

όπου

$$\begin{aligned} B_1^{(p)} &= B_{1p}B_{p,p-1}\cdots B_{21}, \\ B_2^{(p)} &= B_{21}B_{1p}\cdots B_{32}, \\ &\vdots \\ B_p^{(p)} &= B_{p,p-1}B_{p-1,p-2}\cdots B_{1p}. \end{aligned}$$

Παρατηρούμε, λοιπόν, ότι, εφόσον το καθένα από τα διαγώνια blocks $B_i^{(p)}$, $i = 1, \ldots, p$, αποτελείται από ένα κυκλικό γινόμενο των ίδιων πινάκων $B_{i,i-1}$, $i = 1, \ldots, p$, $B_{10} = B_{1p}$, θα έχουν το ίδιο φάσμα μη-μηδενικών ιδιοτιμών, αφού είναι γνωστό ότι για δυο πίνακες $E \in \mathbb{C}^{m,n}$ και $F \in \mathbb{C}^{n,m}$ ισχύει ότι $\sigma(EF)\backslash\{0\} \equiv \sigma(FE)\backslash\{0\}$. Έτσι κάθε μη-μηδενική ιδιοτιμή του B^p είναι και ιδιοτιμή καθενός από τους $B_i^{(p)}$, $i = 1, \ldots, p$, συμπέρασμα που μας δίδει την παραπάνω (6.7) μορφή για το χαρακτηριστικό πολυώνυμο του πίνακα B.

Παρατήρηση 6.2. : Από τη σχέση (6.7) παρατηρούμε ότι οι μη-μηδενικές λύσεις της εξίσωσης $\det(tI - B) = 0$ δίδονται από τις

$$t_{j,k} = |\lambda_j| \exp\left(i\left(\frac{2k\pi}{p} + \theta_j\right)\right), \; k = 0, 1, \ldots, p-1, \; j = 1, 2, \ldots, r, \tag{6.9}$$

όπου θ_j το πρωτεύον όρισμα (Arg) της ιδιοτιμής λ_j. Επιπλέον, στροφή του μιγαδικού επιπέδου περί την αρχή κατά ακέραιο πολλαπλάσιο της γωνίας $\frac{2\pi}{p}$ απεικονίζει το φάσμα των ιδιοτιμών του B, $\sigma(B)$, στον εαυτό του.

Ορισμός 6.3. : Αν ο block πίνακας του Jacobi ($B_J = I - D^{-1}A$), που αντιστοιχεί στον block διαχωρισμό (6.2) είναι "ασθενώς κυκλικός δείκτη $p\,(> 1)$", τότε ο αντίστοιχος πίνακας A, από τον οποίο προέρχεται ο B_J, καλείται "$p-$κυκλικός" σε σχέση με τον block διαχωρισμό του A.

Ορισμός 6.4. : Έστω ότι ο $A \in \mathbb{C}^{n,n}$ είναι $p-$κυκλικός. Τότε ο A καλείται "συνεπώς διατεταγμένος", αν όλες οι ιδιοτιμές του πίνακα

$$B_J(\alpha) = D^{-1}\left(\alpha L + \alpha^{-(p-1)}U\right), \forall \alpha \in \mathbb{C}\backslash\{0\}, \tag{6.10}$$

με L, U, όπως ορίστηκαν στην (6.3), είναι ανεξάρτητες του α.

Με βάση τη μορφή (6.6) για τον πίνακα B_J, η αντίστοιχη μορφή του πίνακα $B_J(\alpha)$ της (6.10) δίδεται από την εκφράση:

$$B_J(\alpha) = \begin{pmatrix} O & O & O & \dots & O & \alpha^{-(p-1)}B_{1p} \\ \alpha B_{21} & O & O & \dots & O & O \\ O & \alpha B_{32} & O & \dots & O & O \\ \vdots & \ddots & \ddots & \ddots & \vdots & \vdots \\ O & O & O & \dots & \alpha B_{p,p-1} & O \end{pmatrix}. \tag{6.11}$$

Αποδεικνύεται με απλούς διαδοχικούς πολλαπλασιασμούς ότι

$$[B_J(\alpha)]^p = (B_J)^p.$$

Έχοντας ότι οι ιδιοτιμές του $B_J(\alpha)$ είναι ανεξάρτητες του α, ο πίνακας A, από τον οποίο προέρχεται ο B_J, είναι "$p-$κυκλικός και συνεπώς διατεταγμένος".

6.2 Ανηγμένο Σύστημα

Θεωρούμε ότι ο πίνακας A του αρχικού συστήματος (6.1), κατάλληλα διαχωρισμένος σε blocks, έχει την εξής ιδιότητα: Ο επαναληπτικός πίνακας Jacobi B, που αντιστοιχεί στον προαναφερθέντα block διαχωρισμό του A, είναι "ασθενώς κυκλικός δείκτη p" στην "κανονική" του μορφή. Τότε, το αρχικό σύστημα θα γράφεται ισοδύναμα ως:

$$(I - B)x = c, \tag{6.12}$$

όπου ο B έχει τη μορφή (6.6) και

$$c = (\bar{c}_1^T, \bar{c}_2^T, \dots, \bar{c}_p^T)^T,\ \bar{c}_j = (diag(A_{jj}))^{-1}\bar{b}_j \in \mathbb{C}^{n_j},\ j = 1, 2, \dots, p.$$

Είναι προφανές ότι, λόγω της αντιστρεψιμότητας του A, $1 \notin \sigma(B)$. Λόγω αυτού του συμπεράσματος και της ασθενώς κυκλικής δείκτη p ιδιότητας του B, από την (6.9) θα έχουμε και

$$\exp\left(i\frac{2k\pi}{p}\right) \notin \sigma(B),\ k = 1, \dots, p-1.$$

Άρα, πολλαπλασιάζοντας τη σχέση (6.12) με τον $I + B + B^2 + \ldots + B^{p-1}$ και χρησιμοποιώντας την ταυτότητα

$$(I + B + B^2 + \ldots + B^{p-1})(I - B) = I - B^p, \tag{6.13}$$

καταλήγουμε στο ισοδύναμο σύστημα:

$$(I - B^p)x = d, \quad d = (I + B + B^2 + \ldots + B^{p-1})c. \tag{6.14}$$

Γράφοντας αναλυτικά το σύστημα (6.14), δηλαδή,

$$\begin{pmatrix} I_{n_1} - B_1^{(p)} & & & & & & \\ & I_{n_2} - B_2^{(p)} & & & & & \\ & & \ddots & & & & \\ & & & I_{n_j} - B_j^{(p)} & & & \\ & & & & \ddots & & \\ & & & & & I_{n_p} - B_p^{(p)} \end{pmatrix} \begin{pmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \vdots \\ \bar{x}_j \\ \vdots \\ \bar{x}_p \end{pmatrix} = \begin{pmatrix} \bar{d}_1 \\ \bar{d}_2 \\ \vdots \\ \bar{d}_j \\ \vdots \\ \bar{d}_p \end{pmatrix},$$

με $\bar{d}_j \in \mathbb{C}^{n_j},\ j = 1, 2, \ldots, p$ να είναι το j block−διάνυσμα, που διαχωρίζεται αντίστοιχα το διάνυσμα d, καταλήγουμε σε p μικρότερης τάξεως ($n_j \times n_j,\ j = 1, 2, \ldots, p$) συστήματα. Επομένως, κάθε σύστημα που παράγεται με τον τρόπο αυτό έχει τη μορφή

$$(I_{n_j} - B_j^{(p)})\bar{x}_j = \bar{d}_j, \quad j = 1, \ldots, p, \tag{6.15}$$

όπου $\bar{d}_j$ δίδεται από τον τύπο

$$\bar{d}_j = \bar{c}_j + B_j \bar{c}_{\pi(j)} + B_j^{(2)} \bar{c}_{\pi^2(j)} + \ldots + B_j^{(p-1)} \bar{c}_{\pi^{p-1}(j)},$$

με $\bar{c}_j \in \mathbb{C}_j^n$ να είναι το j block του c. Το σύμβολο $\pi(j)$ αντιστοιχεί στην κυκλική μετάθεση κατά μία block θέση (από την j στην $j+1$, όπου το $p+1$ ερμηνεύεται ως 1) του j block στοιχείου του διανύσματος c. Το σύστημα (6.15) καλείται "κυκλικώς ανηγμένο σύστημα" (cyclically reduced system).

Παρατήρηση 6.3. : Παρατηρούμε ότι, εν γένει, η επίλυση ενός κατάλληλα επιλεγμένου συστήματος από τα (6.15) είναι "λιτότερη" σε πλήθος πράξεων από αυτήν του συστήματος (6.12). Αυτό συμβαίνει, αρχικά, γιατί επιλύοντας ένα από τα συστήματα

$$(I_{n_j} - B_j^{(p)})\bar{x}_j = \bar{d}_j,\ j = 1, \ldots, p-1,$$

βρίσκουμε τη λύση $\bar{x}_j$, η οποία αντικαθιστώμενη στο σύστημα (6.12), μας επιτρέπει να υπολογίσουμε τα άλλα block−διανύσματα $\bar{x}_i,\ i \neq j$, μέσω σχέσεων της μορφής

$$-B_{j,j-1}\bar{x}_j + \bar{x}_{j-1} = \bar{c}_j,\ j = 1, 2, \ldots, p,$$

όπου δείκτες $j-1$ ή $j+1$ εκτός του διαστήματος $[1,p]$ ερμηνεύονται, πάντοτε, ως $j-1+p$ ή $j+1-p$, αντίστοιχα. Κυρίως, όμως, υπάρχει κάποιος άλλος λόγος που φαίνεται καλύτερα στο παρακάτω παράδειγμα.

Για $p=2$ έχουμε ότι ο πίνακας των συντελεστων στο σύστημα (6.12) γράφεται ως

$$B=\begin{pmatrix} O & B_{12} \\ B_{21} & O \end{pmatrix},$$

οπότε το αντίστοιχο σύστημα (6.14) που προκύπτει είναι το εξής:

$$\begin{pmatrix} I_{n_1}-B_{12}B_{21} & O \\ O & I_{n_2}-B_{21}B_{12} \end{pmatrix}\begin{pmatrix} \bar{x}_1 \\ \bar{x}_2 \end{pmatrix}=\begin{pmatrix} \bar{c}_1+B_{12}\bar{c}_2 \\ \bar{c}_2+B_{21}\bar{c}_1 \end{pmatrix}. \tag{6.16}$$

Επιλύοντας κάποιο από τα δύο ανηγμένα συστήματα και βρίσκοντας την τιμή του διανύσματος $\bar{x}_1$ ($\bar{x}_2$), μπορούμε να βρούμε την τιμή του $\bar{x}_2$ ($\bar{x}_1$), αντικαθιστώντας στην (6.12), μέσω της αλγεβικής εξίσωσης

$$\bar{x}_2=B_{21}\bar{x}_1+\bar{c}_2(\bar{x}_1=B_{12}\bar{x}_2+\bar{c}_1),$$

το διάνυσμα $\bar{x}_2$ ($\bar{x}_1$) (βλ. [10]). Στο σημείο αυτό, γεννιέται το ερώτημα εάν αυτή η διαδικασία είναι πραγματικά συμφέρουσα σε σχέση με την απευθείας απαλοιφή του Gauss και προς τα πίσω αντικατάσταση. Βλέπουμε, λοιπόν, ότι ακολουθώντας τη διαδικασία της απαλοιφής του Gauss σε ένα σύστημα με πίνακα συντελεστών της μορφής

$$\begin{pmatrix} I_{n_1} & -B_{12} \\ -B_{21} & I_{n_2} \end{pmatrix},$$

χρειάζεται να γίνει απαλοιφή μόνο στον $n_2\times n_1$ υποπίνακα B_{21}. Παρατηρούμε ότι αυτή ουσιαστικά γίνεται πολλαπλασιάζοντας το (6.12) με τον πίνακα

$$\begin{pmatrix} I_{11} & O \\ B_{21} & I_{22} \end{pmatrix}$$

με κόστος σε πολλαπλασιασμούς $n_2^2n_1$. Το σύστημα που προκύπτει είναι το εξής:

$$\begin{pmatrix} I_{n_1} & -B_{12} \\ O & I_{n_2}-B_{21}B_{12} \end{pmatrix}\begin{pmatrix} \bar{x}_1 \\ \bar{x}_2 \end{pmatrix}=\begin{pmatrix} \bar{c}_1+B_{12}\bar{c}_2 \\ \bar{c}_2+B_{21}\bar{c}_1 \end{pmatrix}. \tag{6.17}$$

Για την επίλυση ενός τέτοιου συστήματος, επιλύουμε το σύστημα

$$(I_{n_2}-B_{21}B_{12})\bar{x}_2=\bar{c}_2+B_{21}\bar{c}_1,$$

με κόστος σε πολλαπλασιασμούς $\frac{1}{3}n_2^3$, περίπου, βρίσκοντας το $\bar{x}_2$ και έπειτα αντικαθιστώντας στη σχέση $\bar{x}_1 - B_{12}\bar{x}_2 = \bar{c}_1$, βρίσκουμε το $\bar{x}_1$. Το συνολικό κόστος σε πλήθος πολλαπλασιασμών είναι περίπου

$$n_2^2 n_1 + \frac{1}{3}n_2^3.$$

Αν είχε ακολουθηθεί η διαδικασία της απαλοιφής του $-B_{12}$, τότε το πλήθος των αντίστοιχων πολλαπλασιασμών για την επίλυση του συστήματος θα ήταν $n_1^2 n_2 + \frac{1}{3}n_1^3$ περίπου. Η διαφορά είναι ίση με

$$n_2^2 n_1 + \frac{1}{3}n_2^3 - n_1^2 n_2 - \frac{1}{3}n_1^3 \;=\; \frac{1}{3}(n_2 - n_1)\left((n_2 + n_1)^2 + 2n_1 n_2\right).$$

Η τελευταία έκφραση της διαφοράς υποδεικνύει ότι έχοντας τη μορφή (6.16) μπορούμε να βρίσκουμε και να επιλύουμε πρώτα, ως συμφερότερο, εκείνο από τα δύο συστήματα που αντιστοιχεί στο μικρότερο εκ των n_1 και n_2. Κάτι αντίστοιχο ισχύει και στη γενικότερη περίπτωση του (6.15).

7 Εφαρμογή των Μεθόδων Ελαχιστοποίησης σε $p-$Κυκλικούς Πίνακες

Υπό τις υποθέσεις της προηγούμενης παραγράφου, το σύστημα (6.1) μετασχηματίζεται ισοδύναμα στο (6.12):

$$(I - B)x = c,$$

όπου ο πίνακας $B \in \mathbb{C}^{n,n}$ είναι "ασθενώς κυκλικός δείκτη $p(> 1)$" (weakly cyclic of index p) και το διάνυσμα $c \in \mathbb{C}^n$. Παρατηρούμε ότι για το χώρο Krylov έχουμε

$$\mathcal{K}_k(I - B, r_0) \equiv \mathcal{K}_k(B, r_0).$$

Επιπλέον, υποθέτοντας ότι το αρχικό υπόλοιπο $r_0 \in \mathbb{C}^n$ έχει τη μορφή

$$r_0 = \begin{pmatrix} \bar{r}_0^1 \\ 0_{n_2} \\ \vdots \\ 0_{n_p} \end{pmatrix}, \tag{7.1}$$

με block διαχωρισμό ανάλογο με αυτόν σε blocks του πίνακα B (βλ.6.6). Δηλαδή, $\bar{r}_0^1 \in \mathbb{C}^{n_1}$ και τα μηδενικά blocks $0_{n_i} \in \mathbb{C}^{n_i}$, $i = 2, \ldots, p$, έτσι ώστε $\sum_{i=1}^{p} n_i = n$. Η παραπάνω επιλογή του υπολοίπου r_0 παράγεται επιλέγοντας τυχαίες τιμές για το πρώτο block της αρχικής προσέγγισης x_0:

$$x_0 = \begin{pmatrix} \bar{x}_0^1 \\ \bar{x}_0^2 \\ \vdots \\ \bar{x}_0^p \end{pmatrix}$$

και υπολογίζοντας αλγεβρικά τα υπόλοιπα blocks, $\bar{x}_0^i$, $i = 2, \ldots, p$, από τη σχέση:

$$r_0 = \begin{pmatrix} \bar{r}_0^1 \\ 0_{n_2} \\ \vdots \\ 0_{n_p} \end{pmatrix} = \begin{pmatrix} \bar{c}^1 \\ \bar{c}^2 \\ \vdots \\ \bar{c}^p \end{pmatrix} - (I - B) \begin{pmatrix} \bar{x}_0^1 \\ \bar{x}_0^2 \\ \vdots \\ \bar{x}_0^p \end{pmatrix},$$

με τον πίνακα B να είναι της μορφής (6.6). Καταλήγουμε, λοιπόν, στο εξής αλγεβρικό σύστημα:

$$\bar{c}_1 - \bar{x}_0^1 + B_{1p}\bar{x}_0^p = \bar{r}_0^1$$

$$\bar{c}_j - \bar{x}_0^j + B_{j,j-1}\bar{x}_0^{j-1} = 0,\ j = 2, \ldots, p,$$

η επίλυση του οποίου υποδεικνύει την τιμή της αρχικής προσέγγισης, ώστε το υπόλοιπο να έχει τη μορφή (7.1).

Όπως είπαμε και προηγουμένως, λόγω της ειδικής μορφής του πίνακα B και του αρχικού υπολοίπου r_0, ο χώρος Krylov $\mathcal{K}_k(B, r_0)$ που παράγεται είναι της μορφής:

$$\mathcal{K}_k(B, r_0) = span\left\{ \begin{pmatrix} \bar{q}_1^1 \\ 0_{n_2} \\ \vdots \\ 0_{n_p} \end{pmatrix}, \begin{pmatrix} 0_{n_1} \\ \bar{q}_2^2 \\ \vdots \\ 0_{n_p} \end{pmatrix}, \ldots, \begin{pmatrix} 0_{n_1} \\ 0_{n_p} \\ \vdots \\ \bar{q}_p^p \end{pmatrix}, \begin{pmatrix} \bar{q}_1^{p+1} \\ 0_{n_2} \\ \vdots \\ 0_{n_p} \end{pmatrix}, \ldots \right\}, \quad (7.2)$$

όπου τα διανύσματα

$$\bar{q}_j^i \in \mathbb{C}^{n_j}, i = mp + j,\ j = 1, \ldots, p,\ m = 0, 1, \ldots$$

είναι της μορφής

$$\bar{q}_j^i = B^{(i-1)}\bar{q}_1^1,\ i = mp + j,\ m = 0, 1, \ldots\ j = 1, \ldots, p.$$

Παρατηρούμε, λοιπόν, ότι ο παραπάνω χώρος Krylov μπορεί να γραφεί σαν ευθύ άθροισμα χώρων Krylov της μορφής,

$$\mathcal{K}_m(B^p, B^j r_0),\ j = 0, \ldots, p-1,\ k = mp + j,\ m = 0, 1, \ldots,$$

ως εξής:

$$\mathcal{K}_k(B, r_0) = \bigoplus_{j=0}^{i-1} \mathcal{K}_{m+1}(B^p, B^j r_0) \bigoplus_{j=i}^{p-1} \mathcal{K}_m(B^p, B^j r_0),\ k > p. \quad (7.3)$$

Η κατασκευή του παραπάνω χώρου Krylov μέσω του αλγορίθμου του Arnoldi, που παρουσιάστηκε στο Κεφάλαιο 3, παράγει μία ειδική μορφή για τον άνω Hessenberg πίνακα, $H_{k+1,k}$, της σχέσεως (4.19). Η μορφή του πίνακα $H_{k+1,k}$ είναι η εξής:

$$\begin{pmatrix} 0 & 0 & 0 & \ldots & x & \ldots & x & \ldots & \ldots & \ldots \\ x & 0 & 0 & \ldots & 0 & x & \ldots & x & \ldots & \ldots \\ 0 & x & 0 & \ldots & 0 & 0 & x & \ldots & x & \ldots \\ \vdots & \ddots & \ddots & \ddots & \vdots & \vdots & \ddots & \ddots & \ddots & \ddots \end{pmatrix}. \quad (7.4)$$

Αυτό που μπορεί να δεί εύκολα κάποιος είναι ότι τα μη-μηδενικά στοιχεία του άνω Hessenberg πίνακα $H_{k+1,k}$ βρίσκονται στην κύρια υποδιαγώνιο και σε κάθε mp, $m = 0, 1, 2, \ldots$, υπερδιαγώνιο του πίνακα, μετρώντας από την κύρια διαγώνιο στην οποία αντιστοιχίζουμε την τιμή $m = 0$.

Εάν τώρα αναδιατάξουμε τα ορθοκανονικά διανύσματα που παράγονται από τον αλγόριθμο του Arnoldi, έτσι, ώστε τα διανύσματα με την ίδια διάταξη blocks να είναι συνεχόμενα, όπως ακριβώς φαίνεται από τη διάταξη των υπόχωρων Krylov της σχέσεως (7.3), ο αντίστοιχος πίνακας $H_{k+1,k}$ που παράγεται είναι $p-$κυκλικός[18] και μάλιστα τα μη-μηδενικά του blocks έχουν μία συγκεκριμένη και αρκετά απλή δομή.

Στη συνέχεια δίδουμε μερικές επιπλέον λεπτομέρειες σχετικά με τα παραπάνω συμπεράσματα.

Από τη δομή του πίνακα $H_{k+1,k}$ και από τον τρόπο σύνδεσης των στοιχείων μεταξύ τους, παρατηρούμε μία $p-$κυκλικότητα της σύνδεσης αυτής. Αυτό το συμπέρασμα μπορεί να γίνει εύκολα κατανοητό, εάν κατασκευάσουμε το πλήρες point κατευθυνόμενο γράφημα του πίνακα $H_{k+1,k}$ (βλ. [38]). Στο γράφημα αυτό παρατηρούμε ότι όλοι οι κύκλοι (κυκλώματα) αποτελούνται από mp, $m \in \mathbb{N}$, κορυφές. Το χαρακτηριστικό αυτό μας επιτρέπει να συμπεράνουμε ότι ο πίνακας $H_{k+1,k}$ μπορεί να πάρει μια "κανονική" $p-$κυκλική μορφή (6.6). Υπάρχουν, λοιπόν, μεταθετικοί πίνακες $P_1 \in \mathbb{R}^{k,k}$, $P_2 \in \mathbb{R}^{k+1,k+1}$ τέτοιοι, ώστε η σχέση

$$(I - B)Q_k = Q_{k+1}H_{k+1,k}$$

να λαμβάνει την εξής ισοδύναμη μορφή:

$$(I - B)Q_kP_1 = Q_{k+1}P_2P_2^TH_{k+1,k}P_1. \tag{7.5}$$

[18] Καλούμε τον πίνακα $H_{k+1,k}$ $p-$κυκλικό παρόλο που κάποιο από τα διαγώνια blocks δεν θα είναι τετραγωνικό, όπως προϋποθέτει ο κλασικός ορισμός ([38]).

Έτσι, ο πίνακας $\widetilde{H}_{k+1,k} = P_2^T H_{k+1,k} P_1$ έχει τη μορφή που φαίνεται παρακάτω:

$$\widetilde{H}_{k+1,k} = \begin{pmatrix} I_{m+1} & & & & & & & -\widetilde{T}_1^{(m)} \\ -T_2^{(m+1)} & I_{m+1} & & & & & & \\ & \ddots & \ddots & & & & & \\ & & -T_i^{(m+1)} & \widetilde{I}_m & & & & \\ & & & -T_{i+1}^{(m)} & I_m & & & \\ & & & & \ddots & \ddots & & \\ & & & & & -T_p^{(m)} & I_m & \end{pmatrix}, \tag{7.6}$$

με τους πίνακες $T_j^{(m+1)} \in \mathbb{C}^{m+1,m+1}$, $j = 2, \ldots, i$, $T_j^{(m)} \in \mathbb{C}^{m,m}$, $j = i+1, \ldots, p$ να είναι άνω τριγωνικοί και τον πίνακα $\widetilde{T}_1^m \in \mathbb{C}^{m+1,m}$ να είναι άνω Hessenberg. Επιπλέον, ο πίνακας $\widetilde{I}_m \in \mathbb{R}^{m+1,m}$ έχει ως κύριο υποπίνακα το μοναδιαίο $m \times m$ και το τελευταίο $1 \times m$ διάνυσμα−γραμμή είναι το μηδενικό. Το στοιχείο h προκύπτει από τον αλγόριθμο του Arnoldi, που δόθηκε στο Κεφάλαιο 3.

Παρατήρηση 7.1. : Ο παραπάνω άνω Hessenberg πίνακας προκύπτει από την εφαρμογή του αλγορίθμου του Arnoldi στην κατασκευή του χώρου Krylov $\mathcal{K}_k((I-B), r_0)$, όπου ο πίνακας B είναι στην κανονική $p-$κυκλική μορφή και το αρχικό διάνυσμα−υπόλοιπο έχει τη μορφή (7.1). Εάν θελήσουμε να βρούμε τη μορφή που θα έχει ο άνω Hessenberg πίνακας, στην περίπτωση της εφαρμογής του αλγορίθμου του Arnoldi στην κατασκευή του χώρου Krylov $\mathcal{K}_k(B, r_0)$, η μόνη διαφορά είναι ότι ο πίνακας (7.6) έχει διαγώνια blocks μηδενικά.

Θα συμβολίζουμε τον άνω Hessenberg πίνακα της παραπάνω μορφής $\widetilde{T}_{k+1,k}$, όπου k η διάσταση του χώρου Krylov $\mathcal{K}_k(B, r_0)$.

Κάνοντας την παραπάνω παρατήρηση για τη μορφή του άνω Hessenberg πίνακα στην περίπτωση του χώρου Krylov $\mathcal{K}_k(B, r_0)$, ο αλγόριθμος του Arnoldi δίδεται από τη σχέση:

$$B[Q_1^{(m+1)} Q_2^{(m+1)} \ldots Q_j^{(m+1)} \ldots Q_p^{(m)}] = [Q_1^{(m+1)} Q_2^{(m+1)} \ldots Q_j^{(m+1)} Q_{j+1}^{(m+1)} \ldots Q_p^{(m)}] \widetilde{T}_{k+1,k}, \tag{7.7}$$

όπου

$$Q_{j+1}^{(m)} = \mathcal{K}_m(B^p, B^j r_0),\ j = 0, \ldots, p-1,\ k = mp + j,\ m = 0, 1, \ldots.$$

Από τη σχέση (7.7), λαμβάνουμε την εξής εξίσωση:

$$BQ_i^{(m)} = Q_{i+1}^{(m)} T_{i+1}^{(m)}.$$

Πολλαπλασιάζοντας αυτή τη σχέση με τον πίνακα B, καταλήγουμε στην εξίσωση:

$$B^2 Q_i^{(m)} = B Q_{i+1}^{(m)} T_{i+1}^m = Q_{i+2}^{(m)} T_{i+2}^{(m)} T_{i+1}^{(m)}.$$

Επαγωγικά, λοιπόν, και εκμεταλλευόμενοι την κανονική μορφή του πίνακα B, όπως και τη μορφή των υπόχωρων Krylov $\mathcal{K}_m(B^p, B^j r_0)$, καταλήγουμε στην εξίσωση:

$$B^p Q_i^{(m)} = Q_i^{(m+1)} \prod_{j=i}^{2} T_j^{(m+1)} \widetilde{T}_1^{(m)} \prod_{j=p}^{i+1} T_j^{(m)} = Q_i^{(m+1)} T_1 \widetilde{T}_1^{(m)} T_2,\ i = 1, \ldots, p,$$

όπου οι πίνακες T_1, T_2 είναι οι πίνακες γινόμενα

$$T_1 = \prod_{j=i}^{2} T_j^{(m+1)}, T_2 = \prod_{j=p}^{i+1} T_j^{(m)}.$$

Χρησιμοποιώντας την παραπάνω σχέση, λαμβάνουμε ότι

$$(I - B^p) Q_i^{(m)} = Q_i^{(m+1)} \left(\widetilde{I}_m - T_1 \widetilde{T}_1^{(m)} T_2 \right). \tag{7.8}$$

Η παραπάνω εξίσωση είναι ο αλγόριθμος του Arnoldi για τη δημιουργία του χώρου Krylov $\mathcal{K}_m((I - B^p), r_0)$. Εκμεταλλευόμενοι, όμως, και την ειδική μορφή του αρχικού υπολοίπου (7.1), παρατηρούμε ότι ο προηγούμενος χώρος Krylov ταυτίζεται με το χώρο Krylov $\mathcal{K}_m((I - B_i^{(p)}), r_0)$, που προκύπτει εφαρμόζοντας τον αλγόριθμο του Arnoldi στο "κυκλικώς ανηγμένο σύστημα" (6.15).

Γεννιέται, λοιπόν, το ερώτημα εάν η εφαρμογή των μεθόδων ελαχιστοποίησης υπολοίπου στο σύστημα (6.12), στην $k = mp + j$ επανάληψη, ταυτίζεται με το υπόλοιπο στην m επανάληψη των μεθόδων ελαχιστοποίησης, εφαρμοσμένων στο "κυκλικώς ανηγμένο σύστημα" (6.15). Η απάντηση σε αυτό το ερώτημα δεν είναι καθολική για όλες τις μεθόδους. Έτσι, θα διαχωρίσουμε τις μεθόδους σε δύο κύριες κατηγορίες: Στις μεθόδους όπου κατασκευάζουμε το υπόλοιπο r_{k+1} να είναι κάθετο σε ολόκληρο το χώρο Krylov $\mathcal{K}_k((I - B), r_0)$, τις οποίες καλούμε $OR-$μεθόδους (Orthogonal Residual) και σε αυτές όπου ελαχιστοποιούμε την Ευκλείδεια νόρμα του υπολοίπου, τις οποίες καλούμε $MR-$μεθόδους (Minimal Residual). Στην πρώτη κατηγορία μεθόδων ανήκουν, από αυτές που μελετήσαμε, οι μέθοδοι Απότομης Καθόδου, CG, BiCG, CGS και BiCGSTAB. Στη δεύτερη κατηγορία ανήκουν οι μέθοδοι Orthomin(j), Orthodir, MINRES, GMRES και QMR. Για την πρώτη κατηγορία μεθόδων έχουμε το ακόλουθο αποτέλεσμα:

Θεώρημα 7.1. *: Εάν οι $OR-$μέθοδοι εφαρμοστούν στο σύστημα (6.12) με τον πίνακα B στην κανονική $p-$κυκλική μορφή και το αρχικό υπόλοιπο r_0 να έχει την μορφή (7.1), τότε το υπόλοιπο στην επανάληψη $k = mp+i$ είναι μηδέν εκτός από το $i+1$ block, το οποίο συμπίπτει με το υπόλοιπο στην επανάληψη m των $OR-$μεθόδων, εφαρμοσμένων στο $i+1$ block του "κυκλικώς ανηγμένου συστήματος" (6.15).*

Απόδειξη: Από την έκφραση (4.20), έχουμε ότι το υπόλοιπο

$$r_{mp+i} \in \mathcal{K}_{mp+i+1}((I-B), r_0) \equiv \mathcal{K}_{mp+i+1}(B, r_0).$$

Επιπλέον για τις $OR-$μεθόδους το r_{mp+i} είναι κάθετο στον χώρο $\mathcal{K}_{mp+i}((I-B), r_0)$. Παρατηρώντας, τώρα, τη μορφή του υπολοίπου r_{mp+i}, το οποίο έχει μη-μηδενικά στοιχεία μόνο στο $i+1$ block, βλέπουμε ότι τα μη-μηδενικά εσωτερικά γινόμενα

$$(r_{mp+i}, q_j),\ j = 1, \ldots, mp+i$$

είναι αυτά ανάμεσα στο r_{mp+i} και στο χώρο Krylov $\mathcal{K}_m(B^p, B^i r_0)$, τα διανύσματα βάσεις του οποίου είναι μη-μηδενικά μόνο στο $i+1$ block. Άρα αρκεί το υπόλοιπο r_{mp+i} να είναι κάθετο στον χώρο Krylov $\mathcal{K}_m(B^p, B^i r_0)$. Το τελευταίο ισοδυναμεί, λόγω της μορφής του πίνακα B αλλά και του αρχικού υπολοίπου r_0, με το ότι το $r_{mp+i}^{(i+1)}$ block του r_{mp+i} είναι κάθετο στο χώρο $\mathcal{K}_m(B_{i+1}^{(p)}, B^i r_0^{(i+1)})$. Η προηγούμενη έκφραση μπορεί να θεωρηθεί ως το χαρακτηριστικό των $OR-$μεθόδων, εφαρμοσμένων στο $i+1$ block του "κυκλικώς ανηγμένου συστήματος" (6.15). □

Στη συνέχεια, θα δώσουμε ένα ισοδύναμο θεώρημα στην περίπτωση των $MR-$μεθόδων.

Θεώρημα 7.2. *: Εάν οι $MR-$μέθοδοι εφαρμοστούν στο σύστημα (6.12) με τον πίνακα B στην κανονική $p-$κυκλική μορφή και το αρχικό υπόλοιπο r_0 να έχει τη μορφή (7.1), τότε το υπόλοιπο στην επανάληψη $k = mp+i$ είναι μηδέν, εκτός από το $i+1$ block, το οποίο συμπίπτει με το υπόλοιπο στην επανάληψη m των $MR-$μεθόδων, εφαρμοσμένων στο $i+1$ block του "κυκλικώς ανηγμένου συστήματος" (6.15), μόνο εάν το προαναφερθέν υπόλοιπο είναι μηδέν.*

Απόδειξη: Στην κατηγορία των $MR-$μεθόδων επιλέγουμε διανύσματα βελτίωσης του αρχικού υπολοίπου, με σκοπό την ελαχιστοποίηση της Ευκλείδειας νόρμας του υπολοίπου

$$r_k = r_0 - (I-B)Q_{kk}y_k$$

με το διάνυσμα $Q_{kk}y_k \in \mathcal{K}_k((I-B), r_0)$. Όπως ήδη έχουμε δεί, ο χώρος Krylov $\mathcal{K}_k((I-B), r_0)$ διαχωρίζεται σε ευθύ άθροισμα υπόχωρων Krylov, σύμφωνα με τη σχέση (7.3). Από τη μορφή του πίνακα B αλλά και του αρχικού υπολοίπου r_0 στη σχέση (7.1), λαμβάνουμε τις σχέσεις:

$$BQ_j^{(m+1)} = Q_{j+1}^{(m+1)},\ j = 1, \ldots, i, \tag{7.9}$$

$$BQ_j^{(m)} = Q_{j+1}^{(m)},\ j = i+1, \ldots, p-1, \tag{7.10}$$

$$BQ_p^{(m)} = B^p Q_1^{(m)}. \tag{7.11}$$

Από την επιλογή του r_0, έχουμε ότι αυτό είναι κάθετο σε κάθε υπόχωρο $(I-B)Q_j^{(m)}$ για $j = 2, \ldots, i$, και $(I-B)Q_j^{(m+1)}$ για $j = i+1, \ldots, p-1$. Έτσι, οι χώροι οι οποίοι συμμετέχουν στην κατασκευή της διόρθωσης του αρχικού υπολοίπου r_0 είναι οι $(I-B)Q_1^{(m+1)}$ και $(I-B)Q_p^{(m)}$. Λόγω αυτού του χαρακτηριστικού, επιλέγουμε τη διάσταση του χώρου να είναι πολλαπλάσια του p, δηλαδή, $k = mp$.

Ορίζουμε με $\widetilde{r}_m \in Q_1^{(m+1)}$ το υπόλοιπο της m επανάληψης των $MR-$ μεθόδων, εφαρμοσμένων στο πρώτο block του "κυκλικώς ανηγμένου συστήματος" (6.15). Επιπλέον, θα πρέπει να ισχύει η ισότητα $\widetilde{r}_m = r_{mp}$, προκειμένου να λάβουμε το συμπέρασμα του θεωρήματος. Θα πρέπει, λοιπόν, το $\widetilde{r}_m$ να είναι κάθετο στο χώρο Krylov $\mathcal{K}_k((I-B), r_0)$. Η καθετότητα αυτού του υπολοίπου με τους χώρους $(I-B)Q_j^{(m)}$ για $j = 2, \ldots, i$ και $(I-B)Q_j^{(m+1)}$ για $j = i+1, \ldots, p-1$, εξασφαλίζεται λόγω της ειδικής μορφής του πίνακα B καθώς και του αρχικού υπολοίπου r_0. Επιπλέον, έχουμε ότι:

$$(\widetilde{r}_m, (I-B)q_j) = (\widetilde{r}_m, q_j) = 0,\ \forall\, q_j \in Q_1^{(m)}$$

και

$$(\widetilde{r}_m, (I-B)q_j) = (\widetilde{r}_m, q_j) = 0,\ \forall\, q_j \in Q_p^{(m)}.$$

Από τις δύο παραπάνω σχέσεις έχουμε ότι το $\widetilde{r}_m$ είναι κάθετο στον υπόχωρο $Q_1^{(m)} + B^p Q_1^{(m)} \equiv Q_1^{(m+1)}$ και εφόσον το $\widetilde{r}_m \in Q_1^{(m+1)}$ είναι επιπλέον κάθετο στον υπόχωρο $Q_1^{(m+1)}$ θα πρέπει $\widetilde{r}_m = 0$. □

Παρατήρηση 7.2. : Παρατηρούμε, λοιπόν, ότι στην περίπτωση των $OR-$μεθόδων, όπου με την εφαρμογή τους στο σύστημα (6.12) βρίσκουμε τη λύση σε $k = mp + i$ επαναλήψεις, ισοδυναμεί με την εφαρμογή των $OR-$μεθόδων στο "κυκλικώς ανηγμένο σύστημα" (6.15) και την εύρεση της λύσης σε m επαναλήψεις, ενώ, αντιθέτως, στην περίπτωση των $MR-$μεθόδων δεν ισχύει η παραπάνω ισοδύναμία, εκτός βέβαια από την περίπτωση όπου $\widetilde{r}_m = 0$.

8 Επίλογος

Στην παρούσα εργασία παρουσιάσαμε συνοπτικά τη θεωρία των βασικών μεθόδων ελαχιστοποίησης. Σε αυτό που κυρίως επικεντρώσαμε την προσοχή μας ήταν στο να δούμε την εξέλιξη των μεθόδων αυτών, επισημαίνοντας τα σημεία στα οποία κάποια μέθοδος παρουσίαζε ένα μειονέκτημα, όπως και το πώς μία νέα μέθοδος μπορούσε να αναπτυχθεί, ώστε να βελτιώνει ή και να ξεπερνάει το εν λόγω μειονέκτημα. Χαρακτηριστικό παράδειγμα είναι αυτό των μεθόδων BiCG, CGS και BiCGSTAB. Κύριο γνώρισμα των τριών αυτών μεθόδων είναι ότι η επόμενη αποτελεί μια εξέλιξη της προηγούμενης και η κάθε μία εξέλιξη ξεπερνάει κάποιο πρόβλημα που παρουσιάζει η προηγούμενη μέθοδος. Οι τρεις συγκεκριμένες μέθοδοι αναφέρονται σε προβλήματα με γενικό πίνακα συντελεστών και ο βασικός τους πυρήνας είναι ο "Αμφίπλευρος Αλγόριθμος του Lanczos".

Η εξέλιξη αυτών των μεθόδων δε σταματά εδώ, αφού με τη χρήση κατάλληλων προρρυθμιστών, λαμβάνουμε καλύτερα αποτελέσματα. Σε αυτό το σημείο, βεβαίως, θα πρέπει να τονιστεί ότι η χρήση ενός προρρυθμιστή, προφανώς, δεν αλλάζει τη μέθοδο αλλά οι αλλαγές γίνονται κυρίως στον αλγόριθμο της μεθόδου. Αντίστοιχες εξελίξεις παρουσιάζονται και στην κατηγορία μεθόδων που στηρ΄ιζονται στους αλγορίθμους των Arnoldi και Lanczos.

Στην περίπτωση των Ερμιτιανών πινάκων η εξέλιξη, ουσιαστικά, σταματά στην CG, με τις διάφορες, βέβαια, μετεξελίξεις που αναφέρονται κυρίως στο επίπεδο των προρρυθμιστών και, κατά συνέπεια, του αλγορίθμου της μεθόδου και όχι, όπως αναφέραμε, στην ίδια τη μέθοδο. Στη γενική περίπτωση πίνακα, η εξέλιξη των μεθόδων Orthomin(1), Orthomin(j), Orthodir(1) και MINRES σταματά στην GMRES. Στο εξής, κάθε εξέλιξη αφορά σε ειδικές μορφές της GMRES, όπως GMRES(j), δηλαδή, μία μέθοδος GMRES με επανεκκίνηση μετά από κάθε j βήματα. Άλλες εξελίξεις αφόρουν στους προρρυθμιστές της μεθόδου.

Σε ό,τι αφορά τον τομέα των σφαλμάτων είδαμε ότι σε πολλές μεθόδους υπάρχουν αρκετά αποτελέσματα, που δίδουν πολύ ικανοποιητικά φράγματα για τα σφάλματα αυτά. Υπάρχουν, όμως, και μέθοδοι όπως η GMRES και οι μέθοδοι που στηρίζονται στον "Αμφίπλευρο Αλγόριθμο του Lanczos", για τις οποίες αποτελέσματα στον τομέα των σφαλμάτων είτε δεν υπάρχουν ή όταν υπάρχουν δεν είναι τόσο απλά στην κατανόηση και ιδιαίτερα στην υλοποίησή τους. Αυτό συμβαίνει γιατί χρησιμοποιούν διάφορα πολύπλοκα θεωρητικά εργαλεία, όπως στην περίπτωση της μεθόδου GMRES, τη θεωρία των πολυωνύμων Faber. Έτσι χρησιμοποιώντας μόνο εργαλεία της Μιγαδικής Ανάλυσης, της Θεωρίας

Προσεγγίσεων και της Αριθμητικής Ανάλυσης μπορούν να δοθούν σε κάθε συγκεκριμένη περίπτωση κι αυτά μόνο προσεγγιστικά.

Στη συνέχεια, είδαμε πώς οι μέθοδοι ελαχιστοποίησης μπορούν να εφαρμοστούν σε ειδικές κατηγορίες πινάκων και συγκεκριμένα στους $p-$κυκλικούς πίνακες. Παρατηρήσαμε, λοιπόν, ότι εκμεταλλευόμενοι την ειδική μορφή του πίνακα, μπορούμε να απλοποιήσουμε αισθητά τους αλγορίθμους, βλέποντας, μάλιστα, ότι σε κάποιες κατηγορίες αλγορίθμων η επίλυση του αρχικού συστήματος $Ax = b$ ισοδυναμεί με την επίλυση ενός άλλου συστήματος, που προκύπτει από την εκμετάλλευση της ειδικής μορφής του πίνακα A. Η επίλυση αυτού του νέου συστήματος είναι συμφερότερη έναντι αυτής του αρχικού συστήματος. Αυτή η σύγκριση έγινε μέσω των δύο θεωρημάτων που δόθηκαν στο έκτο Κεφάλαιο της παρούσης εργασίας. Βέβαια, υπάρχουν ακόμα πολλά αναπάντητα ερωτήματα, όπως π.χ το πότε συμφέρει να επιλυθεί, μέσω μίας μεθόδου ελαχιστοποίησης το αρχικό σύστημα και πότε το "Ανηγμένο". Μία μερική απάντηση σε αυτό το ερώτημα, που αφορά κυρίως στις πράξεις που απαιτούνται για την επίλυση των δύο συστημάτων, δόθηκε στο πέμπτο Κεφάλαιο της εργασίας.

Ιδιαίτερα αναφέρεται το γεγονός ότι φαίνεται να είναι δυνατή η περαιτέρω εκμετάλλευση των Μεθόδων Ελαχιστοποίσης στην $p-$κυκλική περίπτωση με βάση τις εργασίες [15], [17], [10] και μάλιστα όταν είναι γνωστή(ές) ιδιότητα(ες) των ιδιοτιμών του επαναληπτικου πίνακα του Jacobi, όπως περιγράφονται στα κλασικά βιβλία [38], [40], [3]. Προς αυτήν την κατεύθυνση εργαζόμαστε ήδη.

Ένα γενικό συμπέρασμα, το οποίο συνάγεται από όσα μέχρι τώρα ειπώθηκαν στα πλαίσια της συγκεκριμένης εργασίας, είναι ότι μπορούμε να εκμεταλλευτούμε την ειδική μορφή των πινάκων, ακόμα και για μεθόδους, όπως αυτές που παρουσιάσαμε στην εργασία μας. Θα πρέπει, επίσης, να σημειώσουμε ότι η εφαρμογή των απλών επαναληπτικών μεθόδων σε τέτοιες κατηγορίες πινάκων έχει δώσει πολύ σημαντικά αποτελέσματα. Ίσως, λοιπόν, να κρίνεται σκόπιμη και η αντίστοιχη μελέτη στην περίπτωση των μεθόδων ελαχιστοποίησης που χρησιμοποιούνται ευρύτατα τις δυο τελευταίες δεκαετίες στις πρακτικές εφαρμογές και η μέχρι τώρα εξέλιξή τους είναι ραγδαία.

9 Παράρτημα

1) "Τροποποιημένος" Αλγόριθμος Gram–Schmidt:

Δεδομένα: $v_1, v_2, \ldots, v_n \in \mathbb{C}^n$ γραμμικά ανεξάρτητα
$u_1 = v_1 / \|v_1\|_2$
Για $k = 1, \ldots, n$
$u'_k = v_k$
Για $i = 1, \ldots, k-1$
$u'_k \leftarrow u'_k - (u'_k, u_i)_2 u_i$
$u_k = u'_k / \|u_k\|_2$.

2) Αλγόριθμος Arnoldi:

Δεδομένα: $A \in \mathbb{C}^{n,n},\ \det(A) \neq 0$
Επιλογή: $q_1 \in \mathbb{C}^n \backslash \{0\},\ \|q_1\| = 1$
Για $j = 1, 2, \ldots$
$q'_{j+1} = Aq_j$
Για $i = 1, \ldots, j$
$h_{ij} = (q'_{j+1}, q_i)$
$q'_{j+1} \leftarrow q'_{j+1} - h_{ij} q_i$
$h_{j+1,j} = \|q'_{j+1}\|$
$q_{j+1} = q'_{j+1} / h_{j+1,j}$.

3)Αλγόριθμος Lanczos:

Δεδομένα: $A \in \mathbb{C}^{n,n},\ \det(A) \neq 0,\ A^H = A$
Επιλογή: $q_1 \in \mathbb{C}^n \backslash \{0\},\ \|q_1\| = 1,\ \beta_0 = 0$
Για $j = 1, 2, \ldots$
$q'_{j+1} = Aq_j - \beta_{j-1} q_{j-1}$
$\alpha_j = (q'_{j+1}, q_j)$
$q'_{j+1} \leftarrow q'_{j+1} - \alpha_j q_j$
$\beta_j = \|q'_{j+1}\|$
$q_{j+1} = q'_{j+1} / \beta_j$.

4) Αλγόριθμος Απλής Επανάληψης[19]**:**

Δεδομένα: $A, M \in \mathbb{C}^{n,n},\ \det(A) \neq 0,\ \det(M) \neq 0,\ b \in \mathbb{C}^n$
$x_0 \in \mathbb{C}^n$ (αρχική εκτίμηση)
$r_0 = b - Ax_0,\ Mz_0 = r_0$
Για $k = 1, 2, \ldots$
$x_k = x_{k-1} + z_{k-1}$
$r_k = b - Ax_k$
$Mz_k = r_k.$

5) Αλγόριθμος Απότομης Καθόδου:

Δεδομένα: $A \in \mathbb{C}^{n,n},\ \det(A) \neq 0,\ A^H = A$, θετικά ορισμένος, $b \in \mathbb{C}^n$
$x_0 \in \mathbb{C}^n$ (αρχική εκτίμηση)
$r_0 = b - Ax_0$
Για $k = 1, 2, \ldots$
Εύρεση: Ar_{k-1}
$\alpha_{k-1} = \frac{(r_{k-1}, r_{k-1})}{(Ar_{k-1}, r_{k-1})}$
$x_k = x_{k-1} + \alpha_{k-1} r_{k-1}$
$r_k = b - Ax_k.$

[19] Στον παρόντα καθώς και σε όλους τους άλλους αλγορίθμους που ακολουθούν οι επαναλήψεις τερματίζονται μόλις ικανοποιηθεί κάποιο κριτήριο που έχει τεθεί εκ των προτέρων.

5) Αλγόριθμος Orthomin(2):

Δεδομένα: $A \in \mathbb{C}^{n,n},\ \det(A) \neq 0,\ b \in \mathbb{C}^n$
$x_0 \in \mathbb{C}^n$ (αρχική εκτίμηση)
$r_0 = b - Ax_0,\ p_0 = r_0$
Για $k = 1, 2, \ldots$
Εύρεση: $Ap_{k-1},\ (Ap_{k-1}, Ap_{k-1})$
$\alpha_{k-1} = \frac{(r_{k-1}, Ap_{k-1})}{(Ap_{k-1}, Ap_{k-1})}$
$x_k = x_{k-1} + \alpha_{k-1} p_{k-1}$
$r_k = r_{k-1} - \alpha_{k-1} Ap_{k-1}$
$\beta_{k-1} = \frac{(Ar_k, Ap_{k-1})}{(Ap_{k-1}, Ap_{k-1})}$
$p_k = r_k - \beta_{k-1} p_{k-1}$.

6) Αλγόριθμος Μεθόδου Συζυγών Κλίσεων(CG) [20]:

Δεδομένα: $A \in \mathbb{C}^{n,n},\ \det(A) \neq 0,\ A^H = A$, θετικά ορισμένος, $b \in \mathbb{C}^n$
$x_0 \in \mathbb{C}^n$ (αρχική εκτίμηση)
$r_0 = b - Ax_0,\ p_0 = r_0$
Για $k = 1, 2, \ldots$
Εύρεση: Ap_{k-1}
$\alpha_{k-1} = \frac{(r_{k-1}, r_{k-1})}{(p_{k-1}, Ap_{k-1})}$
$x_k = x_{k-1} + \alpha_{k-1} p_{k-1}$
$r_k = r_{k-1} - \alpha_{k-1} Ap_{k-1}$
$\beta_{k-1} = \frac{(r_k, r_k)}{(r_{k-1}, r_{k-1})}$ [21]
$p_k = r_k + \beta_{k-1} p_{k-1}$.

[20] Για Ερμιτιανούς και θετικά ορισμένους πίνακες.

[21] Σε αρκετές περιπτώσεις ο αλγόριθμος μπορεί να δοθεί και στη μορφή όπου τα $\alpha_{k-1},\ \beta_{k-1}$ λαμβάνουν τις τιμές $\alpha_{k-1} = \frac{(r_{k-1}, p_{k-1})}{(p_{k-1}, Ap_{k-1})},\ \beta_{k-1} = -\frac{(r_k, Ap_k)}{(p_{k-1}, Ap_{k-1})}$, αντίστοιχα.

7) Αλγόριθμος Μεθόδου GMRES:

Δεδομένα: $A \in \mathbb{C}^{n,n},\ \det(A) \neq 0,\ b \in \mathbb{C}^n$
$x_0 \in \mathbb{C}^n$ (αρχική εκτίμηση), $\xi^1 = (1, 0, \ldots, 0)^T$
$r_0 = b - Ax_0$, $q_1 = \frac{r_0}{\|r_0\|}$, $\beta = \|r_0\|$
Για $k = 1, 2, \ldots$
q_{k+1}, $h_{ik} \equiv H(i,k),\ i = 1, \ldots, k+1$ (Υπολογίζονται από τον αλγόριθμο του Arnoldi)
$F_1, \ldots, F_{k-1}$ (Εφαρμογή στροφών Givens στην τελευταία στήλη του H)

$$\begin{pmatrix} H(i,k) \\ H(i+1,k) \end{pmatrix} \leftarrow \begin{pmatrix} c_i & s_i \\ \bar{s}_i & c_i \end{pmatrix} \begin{pmatrix} H(i,k) \\ H(i+1,k) \end{pmatrix}$$

$$\begin{pmatrix} \xi_k^1 \\ \xi_{k+1}^1 \end{pmatrix} \leftarrow \begin{pmatrix} c_i & s_i \\ \bar{s}_i & c_i \end{pmatrix} \begin{pmatrix} \xi_k^1 \\ 0 \end{pmatrix}$$

$$H(k,k) \leftarrow c_k H(k,k) + s_k H(k+1,k),\ H(k+1,k) \leftarrow 0$$

Εάν $\beta \left|\xi_{k+1}^1\right| \ll 1$
$H_{k\times k} y_k = \beta \xi_{k\times 1}^1$
$x_k = x_0 + Q_k y_k$.

8) Αλγόριθμος Μεθόδου MINRES:

Δεδομένα: $A \in \mathbb{C}^{n,n},\ \det(A) \neq 0,\ A^H = A$, θετικά ή μη-ορισμένος, $b \in \mathbb{C}^n$
$x_0 \in \mathbb{C}^n$ (αρχική εκτίμηση), $\xi^1 = (1, 0, \ldots, 0)^T$
$r_0 = b - Ax_0$, $q_1 = \frac{r_0}{\|r_0\|}$, $\beta = \|r_0\|$
Για $k = 1, 2, \ldots$

$q_{k+1},\ \alpha_k \equiv T(k,k),\ \beta_k \equiv T(k+1,k) \equiv T(k,k+1)$ (Υπολογίζονται από τον αλγόριθμο του Lanczos)
$F_{k-2},\ F_{k-1}$ (Εφαρμογή των στροφών Givens)

$$\begin{pmatrix} T(k-2,k) \\ H(k-1,k) \end{pmatrix} \leftarrow \begin{pmatrix} c_{k-2} & s_{k-2} \\ \bar{s}_{k-2} & c_{k-2} \end{pmatrix} \begin{pmatrix} 0 \\ T(k-1,k) \end{pmatrix},\ k > 2$$

$$\begin{pmatrix} T(k-1,k) \\ H(k,k) \end{pmatrix} \leftarrow \begin{pmatrix} c_{k-1} & s_{k-1} \\ \bar{s}_{k-1} & c_{k-1} \end{pmatrix} \begin{pmatrix} T(k-1,k) \\ T(k,k) \end{pmatrix},\ k > 1$$

(Εφαρμόζεται η k στροφή στο ξ^1 και στην τελευταία στήλη του T)

$$\begin{pmatrix} \xi_k^1 \\ \xi_{k+1}^1 \end{pmatrix} \leftarrow \begin{pmatrix} c_k & s_k \\ \bar{s}_k & c_k \end{pmatrix} \begin{pmatrix} \xi_k^1 \\ 0 \end{pmatrix}$$

$$T(k,k) \leftarrow c_k T(k,k) + s_k T(k+1,k),\ T(k+1,k) \leftarrow 0$$

$p_{k-1} = [q_k - T(k-1,k)p_{k-2} - T(k-2,k)p_{k-3}]/T(k,k)$
$\alpha_{k-1} = \beta\xi_k^1$
$x_k = x_{k-1} + \alpha_{k-1}p_{k-1}$.

9) Αμφίπλευρος Αλγόριθμος Lanczos (χωρίς look-ahead):

Δεδομένα: $A \in \mathbb{C}^{n,n},\ \det(A) \neq 0$
Επιλογή: $r_0 \in \mathbb{C}^n \backslash \{0\}$
Επιλογή: $r_0' \in \mathbb{C}^n \backslash \{0\}$ τ.ω. $(r_0, r_0')_2 \neq 0$
$v_0 = w_0 = 0,\ v_1 = \frac{r_0}{\|r_0\|},\ w_1 = \frac{r_0'}{(r_0', v_1)},\ \beta_0 = \gamma_0 = 0$
Για $j = 1, 2, \ldots$
$\quad$ Εύρεση: $Av_j,\ A^H w_j$
$\quad \alpha_j = (Av_j, w_j)$
$\quad v_{j+1}' = Av_j - \alpha_j v_j - \beta_{j-1} v_{j-1}$
$\quad w_{j+1}' = A^H w_j - \bar{\alpha}_j w_j - \gamma_{j-1} w_{j-1}$
$\quad \gamma_j = \left\| v_{j+1}' \right\|$
$\quad v_{j+1} = v_{j+1}'/\gamma_j$
$\quad \beta_j = (v_{j+1}, w_{j+1}')$
$\quad w_{j+1} = w_{j+1}'/\bar{\beta}_j.$

10) Αλγόριθμος Μεθόδου Biconjugate Gradient:

Δεδομένα: $A \in \mathbb{C}^{n,n},\ \det(A) \neq 0,\ b \in \mathbb{C}^n$
$x_0 \in \mathbb{C}^n$ (αρχική εκτίμηση)
$r_0 = b - Ax_0,\ p_0 = r_0$
Επιλογή: $\hat{r}_0 \in \mathbb{C}^n$ τ.ω. $(r_0, \hat{r}_0) \neq 0$
$\hat{p}_0 = \hat{r}_0$
Για $k = 1, 2, \ldots$
$\quad \alpha_k = \frac{(r_{k-1}, \hat{r}_{k-1})}{(Ap_{k-1}, \hat{p}_{k-1})}$
$\quad x_k = x_{k-1} + \alpha_k p_{k-1}$
$\quad r_k = r_{k-1} - \alpha_{k-1} A p_{k-1},\ \hat{r}_k = \hat{r}_{k-1} - \bar{\alpha}_{k-1} A^H \hat{p}_{k-1}$
$\quad \beta_{k-1} = \frac{(r_k, \hat{r}_k)}{(r_{k-1}, \hat{r}_{k-1})}$
$\quad p_k = r_k + \beta_{k-1} p_{k-1},\ \hat{p}_k = \hat{r}_k + \bar{\beta}_{k-1} \hat{p}_{k-1}.$

11) Αλγόριθμος Μεθόδου QMR (χωρίς look ahead):

Δεδομένα: $A \in \mathbb{C}^{n,n},\ \det(A) \neq 0,\ b \in \mathbb{C}^n$
$x_0 \in \mathbb{C}^n$ (αρχική εκτίμηση), $\xi^1 = (1, 0, \ldots, 0)^T$
$r_0 = b - Ax_0$, $v_1 = \frac{r_0}{\|r_0\|}$, $\beta = \|r_0\|$
Επιλογή: $\hat{r}_0 \in \mathbb{C}^n$,
$w_1 = \frac{\hat{r}_0}{\|\hat{r}_0\|}$
Για $k = 1, 2, \ldots$

v_{k+1}, w_{k+1}, $\alpha_k \equiv T(k,k)$, $\beta_k \equiv T(k,k+1)$, $\gamma_k \equiv T(k+1,k)$
(Εφαρμόζονται οι στροφές F_{k-2} και F_{k-1} στην τελευταία στήλη του T)

$$\begin{pmatrix} T(k-2,k) \\ T(k-1,k) \end{pmatrix} \leftarrow \begin{pmatrix} c_{k-2} & s_{k-2} \\ -\bar{s}_{k-2} & c_{k-2} \end{pmatrix} \begin{pmatrix} 0 \\ T(k-1,k) \end{pmatrix}, \quad k > 2$$

$$\begin{pmatrix} T(k-1,k) \\ T(k,k) \end{pmatrix} \leftarrow \begin{pmatrix} c_{k-1} & s_{k-1} \\ -\bar{s}_{k-1} & c_{k-1} \end{pmatrix} \begin{pmatrix} T(k-1,k) \\ T(k,k) \end{pmatrix}, \quad k > 1$$

(Εφαρμόζεται η k στροφή στο διάνυσμα ξ^1 και στην τελευταία στήλη του T)

$$\begin{pmatrix} \xi_k^1 \\ \xi_{k+1}^1 \end{pmatrix} \leftarrow \begin{pmatrix} c_k & s_k \\ -\bar{s}_k & c_k \end{pmatrix} \begin{pmatrix} \xi_k^1 \\ 0 \end{pmatrix}$$

$$T(k,k) \leftarrow c_k T(k,k) + s_k T(k+1,k),\ T(k+1,k) \leftarrow 0$$

$p_{k-1} = [v_k - T(k-1,k)p_{k-2} - T(k-2,k)p_{k-3}]/T(k,k)$ (ορισμένοι όροι για $k \leq 2$ είναι μηδέν)
$x_k = x_{k-1} + \beta \xi_k^1 p_{k-1}$.

12) Αλγόριθμος Μεθόδου CGS:

Δεδομένα: $A \in \mathbb{C}^{n,n},\ \det(A) \neq 0,\ b \in \mathbb{C}^n$
$x_0 \equiv x_0^S \in \mathbb{C}^n$ (αρχική εκτίμηση)
$r_0 \equiv r_0^S = b - Ax_0^S$
$u_0^S = r_0^S$, $p_0^S = r_0^S$, $q_0^S = 0$, $v_0^S = Ap_0^S$
Επιλογή: $\widehat{r}_0 \in \mathbb{C}^n$ τ.ω. $(r_0^S, \widehat{r}_0) \neq 0$
Για $k = 1, 2, \ldots$

$$\alpha_{k-1} = \frac{(r_{k-1}^S, \widehat{r}_0)}{(v_{k-1}^S, \widehat{r}_0)}$$
$$q_k^S = u_{k-1}^S - \alpha_{k-1} v_{k-1}^S$$
$$x_k^S = x_{k-1}^S + \alpha_{k-1}(u_{k-1}^S + q_k^S)$$
$$r_k^S = r_{k-1}^S - \alpha_{k-1} A(u_{k-1}^S + q_k^S)$$
$$\beta_k = \frac{(r_k^S, \widehat{r}_0)}{(r_{k-1}^S, \widehat{r}_0)}$$
$$u_k^S = r_k^S + \beta_k q_k^S$$
$$p_k^S = u_k^S + \beta_k(q_k^S + \beta_k p_{k-1}^S)$$
$$v_k^S = Ap_k^S.$$

13) Αλγόριθμος Μεθόδου BiCGSTAB:

Δεδομένα: $A \in \mathbb{C}^{n,n},\ \det(A) \neq 0,\ b \in \mathbb{C}^n$
$x_0 \in \mathbb{C}^n$(αρχική εκτίμηση)
$r_0 = b - Ax_0,\ p_0 = r_0$
Επιλογή: $\widehat{r}_0 \in \mathbb{C}^n$ τ.ω. $(r_0, \widehat{r}_0) \neq 0$
Για $k = 1, 2, \ldots$
Εύρεση: Ap_{k-1}
$\alpha_{k-1} = \frac{(r_{k-1}, \widehat{r}_0)}{(Ap_{k-1}, \widehat{r}_0)}$
$x_{k-\frac{1}{2}} = x_{k-1} + \alpha_{k-1} p_{k-1}$
$r_{k-\frac{1}{2}} = r_{k-1} - \alpha_{k-1} Ap_{k-1}$
Εύρεση: $Ar_{k-\frac{1}{2}}$
$\omega_k = \frac{(r_{k-\frac{1}{2}}, Ar_{k-\frac{1}{2}})}{(Ar_{k-\frac{1}{2}}, Ar_{k-\frac{1}{2}})}$
$x_k = x_{k-\frac{1}{2}} + \omega_k r_{k-\frac{1}{2}}$
$r_k = r_{k-\frac{1}{2}} - \omega_k Ar_{k-\frac{1}{2}}$
$\beta_k = \frac{\alpha_{k-1}}{\omega_k} \frac{(r_k, \widehat{r}_0)}{(r_{k-1}, \widehat{r}_0)}$
$p_k = r_k + \beta_k (p_{k-1} - \omega_k Ap_{k-1})$.

Αναφορές

[1] O. Axelsson, *Iterative Solution Methods*, Cambridge University Press, London, 1994.

[2] R. Barrett, M. Berry, T.F. Chan, J. Demmel, J. Donato, J. Dongatta, V. Eijkhout, R. Pozo, C. Romine and H. van der Vorst, *Templates for the Solution of Linear Systems: Building Blocks for Iterative Methods*, SIAM, Philadelphia, PA, 1995.

[3] A. Berman and R.J. Plemmons, *Nonnegative Matrices in the Mathematical Sciences*, SIAM, Philadelphia, 1994.

[4] C. Brezinski, M. Redivo Zaglia and H. Sadok, *Avoiding breakdown and near-breakdown in Lanczos type algorithms*, Numer. Algorithms, **1** (1991), pp. 199–206.

[5] J.W. Demmel, *Applied Numerical Linear Algebra*, SIAM, Philadelphia, 1997.

[6] Β. Δουγαλής, Δ. Νούτσος και Α. Χατζηδήμος, *Σημειώσεις Αριθμητικής Γραμμικής Αλγεβρας*, Ηράκλειο, 2001.

[7] M. Eiermann, *Fields of values and iterative methods*, Linear Algebra Appl., **180** (1993), pp. 167–197.

[8] M. Eiermann, *Fields of values and iterative methods*, Talk presented at Oberwolfach Meeting on Iterative Methods and Scientific Computing, Oberwolfash, Germany, April, 1997, to appear.

[9] M. Eiermann and O. Ernst, *Geometric aspects in the theory of Krylov subspace methods*, 1998, to appear in Acta Numerica.

[10] O. Ernst, *Equivalent iterative methods for p–cyclic matrices.*

[11] V. Faber and T. Manteuffel, *Necessary and sufficient conditions for the existence of a conjugate gradient method*, SIAM J. Numer. Anal., **21** (1984), pp. 352–362.

[12] V. Faber and T. Manteuffel, *Orthogonal error methods*, SIAM J. Numer. Anal., **24** (1987), pp. 170–187.

[13] B. Fischer, *Polynomial Based Iteration Methods for Symmetric Linear Systems*, Wiley–Teubner, Leipzig, 1996.

[14] R. Fletcher, *Conjugate gradient methods for indefinite systems*, in Proc. Dundee Biennial Conference on Numerical Analysis, G.A. Watson, ed., Springer–Verlag, Berlin, 1975.

[15] R.W. Freund, *A note on two block-SOR methods for sparse least square problems*, Linear Algebra Appl., **88/89** pp. 211–221, 1987.

[16] R.W. Freund, *A transpose-free quasi-minimal residual algorithm for non-Hermitian linear systems*, SIAM J. Sci. Comput., **14** (1993), pp. 470–482.

[17] R.W. Freund, G.H. Golub and M. Hochbruck, *Krylov subspace methods for non-Hermitian p−cyclic matrices*, unbublished manustript.

[18] R.W. Freund and N.M. Nachtigal, QMR: *A quasi-minimal residual methods for non-Hermitial linear systems*, Numer. Math., **60** (1991), pp. 315–339.

[19] G.H. Golub and R.S. Varga, *Chebychev semi-iterative methods, successive overrelaxation iterative methods, and second-order Richardson iterative methods, part I and II*, Numer. Math., **3** (1961), pp. 147–168.

[20] A. Greenbaum, *Iterative Methods for Solving Linear Systems*, SIAM, PA, 1997.

[21] W. Hackbusch, *Iterative Solution of Large Sparse Systems of Equations*, Springer–Verlag, Berlin, 1994.

[22] R.A. Horn and C. Johnson, *Matrix Analysis*, Cambridge University Press, London, U.K., 1985.

[23] R.A. Horn and C. Johnson, *Topics in Matrix Analysis*, Cambridge University Press, London, U.K., 1991.

[24] E.F. Kaasschieter, *The solution of non-symmetric linear systems by bi-conjugate gradients or conjugate gradients squared*, Delft University of Technology, Faculty of Mathematics and Informatics, Report 86–121, 1986.

[25] C. Lanczos, *An iteration method for the solution of the eigenvalue problem of linear differential and integral operators*, J. Res. Nat. Bur. Standards, **45** (1950), pp. 255–282.

[26] N.M. Nachtigal, *A look-ahead variant of the Lanczos algorithm and its applications to the quasi-minimal residual method for non-Hermitian linear systems*, Ph.D. dissertation, Massachusetts Institute of Technology, Cambridge, MA, 1991.

[27] N.M. Nachtigal, S. Reddy and L.N. Trefethen, *How fast are nonsymmetric matrix iterations?*, SIAM J. Matrix Anal. Appl., **13** (1992), pp. 778–795.

[28] C.C. Paige and M.A. Saunders, *Solution of sparse indefinite systems of linear equations*, SIAM J. Numer. Anal., **11** (1974), pp. 197–209.

[29] B.N. Parlett, D.R. Taylor and Z.A. Liu, *A look-ahead Lanczos algorithm for unsymmetric matrices*, Math. Comp., **44** (1985), pp. 105–124.

[30] V. Romanovsky, *Recherches sur les chaines de Markoff*, Acta Math., **66** (1936), 147–251.

[31] Y. Saad, *Krylov Subspace Methods for Solving Large Unsymmetric Linear Systems*, Math. Comp., **37** (1981), pp. 105–126.

[32] Y. Saad, *Iterative Methods for Sparse Linear Systems*, PWS Pub. Co., Boston, MA, 1996.

[33] Y. Saad and M.H. Schultz, GMRES: *A generalized minimal residual algorithm for solving nonsymmetric linear systems*, SIAM J. Sci. Statist. Comput., **7** (1986), pp. 856–869.

[34] P. Sonneveld, CGS: *a fast Lanczos-type solver for nonsymmetric linear systems*, SIAM J. Sci. Statist. Comput., **10** (1989), pp. 36–52.

[35] L.N. Trefethen, *Approximation theory and numerical linear algebra*, in Algorithms for Approximation II, J. Mason and M. Cox, eds., Chapman and Hall, London, U.K., 1990.

[36] H. van der Vorst, *The convergence behavior of preconditioned CG and CG-S in the presence of rounding errors*, in Preconditioned Conjugate Gradient Methods, O. Axelsson and L. Kolotilina, eds., Lecture Notes in Mathematics 1457, Springer–Verlag, Berlin, 1990.

[37] H. van der Vorst, Bi-CGSTAB: *A fast and smoothly converging variant of Bi-CG for the solution of nonsymmetric linear systems*, SIAM J. Sci. Comput., **13** (1992), pp. 631–644.

[38] R.S. Varga, *Matrix Iterative Analysis*, Springer–Verlag, Berlin/Heidelberg, second edition, 2000.

[39] J.H. Wilkinson, *The Algebraic Eigenvalue Problem*, Oxford University Press, 1965.

[40] D.M. Young, *Iterative Solution of Large Linear Systems*, Academic Press, NY, 1971.

[41] D.M. Young and K.C. Jea, *Generalized conjugate gradient acceleration of nonsymmetrizable iterative methods*, Linear Algebra Appl. **34**, (1980), pp. 159–194.