ΠΑΝΕΠΙΣΤΗΜΙΟ ΚΡΗΤΗΣ
ΤΜΗΜΑ ΜΑΘΗΜΑΤΙΚΩΝ

ΜΙΧΑΗΛ ΛΑΠΙΔΑΚΗΣ

Επιβλέπων Καθηγητής: ΑΠΟΣΤΟΛΟΣ ΧΑΤΖΗΔΗΜΟΣ

ΔΙΔΑΚΤΟΡΙΚΗ ΔΙΑΤΡΙΒΗ
Ηράκλειο, 18 Ιουνίου 2008 .

# Περιεχόμενα

Θα ήθελα να ευχαριστήσω μερικούς από εκείνους που με βοήθησαν στην πορεία για την ολοκλήρωση της διδακτορικής μου διατριβής. Πάνω από όλους θα ήθελα να ευχαριστήσω τον "Δάσκαλο" μου κ. Α. Χατζηδήμο για την πολύτιμη βοήθεια που μου προσέφερε σε ερευνητικό επίπεδο αλλά κυρίως για τις πολύτιμες συμβουλές του που αποτέλεσαν αλλά και θα αποτελούν οδηγό στην πορεία μου . Στην συνέχεια θα ήθελα να ευχαριστήσω τα υπόλοιπα μέλη της τριμελούς συμβουλευτικής επιτροπής κ. Εμμ. Βάβαλη και κ. Χ. Μακριδάκη. Επίσης τα μέλη της επταμελούς εξεταστικής επιτροπής κ. Β. Δουγαλή κ. Θ. Παπαθεοδώρου, κ. Η. Χούστη, κ. Δ. Νούτσο, κ. Γ. Ζουράρη και βέβαια όλα τα μέλη των Τμημάτων Μαθηματικών και Εφαρμοσμένων Μαθηματικών του Πανεπιστημίου Κρήτης που υπήρξαν κατά καιρούς καθηγητές μου σε προπτυχιακό και μεταπτυχιακό επίπεδο.

Θα ήθελα να ευχαριστήσω το Ίδρυμα Κρατικών υποτροφιών που με στήριξε οικονομικά καθ'όλη την διάρκεια των σπουδών μου για την εκπόνηση της διδακτορικής μου διατριβής.

Τέλος θα ήθελα να ευχαριστήσω την σύζυγο μου Κοκκινάκη Ιακώβα καθώς και την Οικογένεια μου για την υπομονή που έδειξαν και την ηθική στήριξη που μου προσέφεραν όλα αυτά τα χρόνια.

Μ. Λαπιδάκης

# Κεφάλαιο 1

# Εισαγωγή

Σκοπός της παρούσας διατριβής είναι η επίλυση ενός αλγεβρικού γραμμικού συστήματος της μορφής

$$Ax = b,\ A \in \mathbb{C}^{n\times n},\ \det(A) \neq 0,\ b \in \mathbb{C}^n \backslash \{0\}. \tag{1.0.1}$$

Στην εξέλιξη της διατριβής, και από κάποιο σημείο και μετά, ο πίνακας $A$ θα θεωρείται πραγματικός, συμμετρικός και θετικά ορισμένος, το δε διάνυσμα $b$ πραγματικό. Η επίλυση ενός τέτοιου συστήματος θα προέλθει μέσω της Μεθόδου Συζυγών Κλίσεων και πιο συγκεκριμένα μέσω της Προρρυθμισμένης Μεθόδου Συζυγών Κλίσεων. Ιδιαίτερα, απώτερος σκοπός της παρούσας εργασίας είναι η εισαγωγή−πρόταση ενός νέου προρρυθμιστή ο οποίος βασίζεται στις Πεπλεγμένες Επαναληπτικές Μεθόδους Εναλλασσόμενων Διευθύνσεων (Alternating Direction Implicit (ADI) Iterative Methods).

Στη συνέχεια παρουσιάζουμε περιγραφικά τα βασικά στοιχεία κάθε κεφαλαίου της διατριβής.

Στο Δεύτερο Κεφάλαιο παρουσιάζονται και περιγράφονται βασικά στοιχεία Γραμμικής Άλγεβρας καθώς και στοιχεία Ανάλυσης Πινάκων, τα οποία θα χρησιμοποιούνται συνεχώς στα επόμενα κεφάλαια και η αναφορά σ΄ αυτά κρίνεται απαραίτητη.

Στο Τρίτο Κεφάλαιο παρουσιάζονται δύο σχήματα διακριτοποίησης μέσω Πεπερασμένων Διαφορών της Εξίσωσης Poisson σε ορθογώνιο χωρίο με Dirichlet συνοριακές συνθήκες. Τα σχήματα που παρουσιάζονται είναι το κλασικό σχήμα των $5-$σημείων με τάξη ακρίβειας $\mathcal{O}(h^2)$ και κυρίως ένα σχήμα $9-$σημείων με τάξη ακρίβειας $\mathcal{O}(h^4)$ σε ομοιόμορφο διαμερισμό με το ίδιο βήμα σε κάθε διεύθυνση.

Στο Τέταρτο Κεφάλαιο παρουσιάζονται και αναλύονται, ως ένα ικανοποιητικό βαθμό, οι βασικές Πεπλεγμένες Επαναληπτικές Μεθόδοι Εναλλασσόμενων Διευθύνσεων (ADI), καθώς και αρκετές παραλαγές αυτών. Μάλιστα στα ερευνητικά αποτελέσματα, που παρουσιάζονται, χρησιμοποιείται μία παραλλαγή ενός επαναληπτικού σχήματος που προτάθηκε αρχικά από τους Guittet [27] και Hadjidimos [28]. Έμφαση δίνεται στις Μεθόδους με Σταθερές Παραμέτρους Επιτάχυνσης για τις οποίες υπάρχουν σημαντικά προηγούμενα αποτελέσματα. Στο τέλος του Κεφαλαίου παρουσιάζεται ένα γενικευμένο σχήμα Παρεκβαλλόμενων (Extrapolated) (E)ADI Επαναληπτικών Μεθόδων η λεπτομερής μελέτη των οποίων συνεχίζεται και στα επόμενα κεφάλαια της διατριβής.

Στο Πέμπτο Κεφάλαιο παρουσιάζονται αφενός βασικά στοιχεία, που αφορούν στα σφάλματα της Μεθόδου Συζυγών Κλίσεων, και αφετέρου η Προρρυθμισμένη Μέθοδος Συζυγών Κλίσεων, παρουσιάζοντας εν συντομία τους βασικούς Προρρυθμιστές που χρησιμοποιούνται στη μέθοδο αυτή.

Στο Έκτο Κεφάλαιο της διατριβής παρουσιάζονται τα πρωτότυπα αποτελέσματα της διατριβής που αφορούν στους Βέλτιστούς EADI Προρρυθμιστές για τη Μέθοδο των Συζυγών Κλίσεων. Ειδικότερα, στο Πέμπτο Κεφάλαιο παρουσιάζονται αποτελέσματα που αφορούν στη μονοπαραμετρική περίπτωση της παραμέτρου επιτάχυνσης, μιας παραλλαγής του σχήματος του Guittet, βρίσκοντας αναλυτικά τις βέλτιστες παραμέτρους στις περιπτώσεις των σχημάτων διακριτοποίησης των $5-$ και των $9-$σημείων. Σημειώνεται ότι τα αποτελέσματα για το σχήμα των $9-$σημείων δίνονται για πρώτη φορά. Στο Έκτο Κεφάλαιο παρουσιάζονται αναλυτικά αποτελέσματα του διπαραμετρικού αναλόγου, σ΄ ό,τι αφορά τις παραμέτρους επιτάχυνσης, του σχήματος του Guittet. Οι αναλυτικές εκφράσεις για τις βέλτιστες τιμές των παραμέτρων στις περιπτώσεις των διακριτοποιήσεων των $5-$ και $9-$σημείων παρουσιάζονται εδώ για πρώτη φορά. Θα πρέπει ακόμη να τονιστεί ότι τα πρωτότυπα αποτελέσματα του παρόντος κεφαλαίου, που αφορούν στις βέλτιστες παραμέτρους επιτάχυνσης και παρεκβολής, είναι πρωτότυπα όχι μόνο σ΄ ό,τι αφορά τον EADI Προρρυθμιστή της Μεθόδου Συζυγών Κλίσεων αλλά και τις EADI Επαναληπτικές Μεθόδους αυτές καθεαυτές.

Στο Έβδομο Κεφάλαιο παρουσιάζεται μια σειρά Αριθμητικών Παραδειγμάτων τα αποτελέσματα των οποίων επαληθεύουν τα αντίστοιχα θεωρητικά. Κυρίως, όμως, αποδεικνύουν πειραματικά ότι ο προρρυθμιστής που έχει εισαχθεί είναι καλύτερος από πολλούς από τους μέχρι σήμερα κλασικούς προρρυθμιστές. Επιπλέον, αποδεικνύουν ότι η προτεινόμενη Προρρυθμισμένη Μέθοδος ADI-CG είναι συγκρίσιμη με τις πλέον γνωστές και δημοφιλείς μεθόδους που χρησιμοποιούνται, όπως είναι οι FFT, Cyclic Reduction και βέβαια οι Μέθοδοι Multi-

grid.

Στο Όγδοο κεφάλαιο γίνεται μια πρώτη προσπάθεια εύρεσης και εφαρμογής Βέλτιστου EADI Προρρυθμιστή για τη Μέθοδο των Συζυγών Κλίσεων, όταν η διακριτοποίηση του ελλειπτικού διαφορικού τελεστή πραγματοποιείται με μεθόδους Collocation και δίνονται βέλτιστες αναλυτικές εκφράσεις για τις διάφορες εμπλεκόμενες παραμέτρους καθώς κι ένα πλήθος αριθμητικών παραδειγμάτων προς επιβεβαίωση της αναπτυχθείσας θεωρίας.

Τέλος, παρουσιάζονται μια σειρά από χρήσιμα συμπεράσματα που προκύπτουν από την αναπτυχθείσα θεωρία όπως επίσης και από τα αριθμητικά παραδείγματα που δίνονται. Επιπλέον διατυπώνονται προτάσεις για περαιτέρω μελέτη και έρευνα τόσο σε θεωρητικό επίπεδο όσο και σε επίπεδο αριθμητικών υπολογισμών.

# Κεφάλαιο 2

# Στοιχεία Γραμμικής Άλγεβρας

## 2.1 Βασικοί ορισμοί

Στο κεφάλαιο αυτό θα παρουσιάσουμε μία σειρά από βασικά στοιχεία Γραμμικής Άλγεβρας, τα οποία θα χρησιμοποιηθούν στα επόμενα κεφάλαια της διατριβής.

Αρχικά θα ξεκινήσουμε με βασικούς ορισμούς των Ευκλείδειων νορμών διανυσμάτων και πινάκων.

**Ορισμός 2.1.1.** : Έστω $x^T = (x_1, x_2, \ldots, x_n)$ διάνυσμα του γραμμικού διανυσματικού χώρου $\mathbb{C}^n$. Τότε

$$\|x\|_2 = (x, x)^{\frac{1}{2}} = \left( \sum_{i=1}^{n} |x_i|^2 \right)^{\frac{1}{2}}. \tag{2.1.1}$$

**Ορισμός 2.1.2.** : Εάν $A \in \mathbb{C}^{n \times n}$ με ιδιοτιμές $\lambda_i$, $i = 1, 2, \ldots, n$, τότε φασματική ακτίνα του πίνακα $A$ καλείται η ποσότητα

$$\rho(A) = \max_{i=1,2,\ldots,n} |\lambda_i| . \tag{2.1.2}$$

Έχοντας τον ορισμό της φασματικής ακτίνας πίνακα, δίνουμε τον ορισμό της φυσικής νόρμας πίνακα, που επάγεται από την αντίστοιχη Ευκλείδεια διανυσματική νόρμα.

**Ορισμός 2.1.3.** : Εάν $A \in \mathbb{C}^{n \times n}$ τότε

$$\|A\|_2 = \rho^{\frac{1}{2}}(A^H A), \tag{2.1.3}$$

όπου $A^H$ είναι ο συζυγής ανάστροφος του πίνακα $A$ και $\rho(\cdot)$ συμβολίζει τη φασματική ακτίνα του πίνακα.

Θα πρέπει να σημειώσουμε εδώ ότι σε αρκετά κλασικά συγγράμματα ο παραπάνω ορισμός της "Ευκλείδειας" νόρμας ενός πίνακα $A$ δίνεται ώς ισοδύναμος ορισμός που παράγεται από τον κλασικό ορισμό της νόρμας ενός πίνακα δηλαδή,

$$\|A\|_2 := \sup_{x\in\mathbb{C}^n\setminus\{0\}} \frac{\|Ax\|_2}{\|x\|_2} \equiv \sup_{x\in\mathbb{C}^n,\, \|x\|_2=1} \|Ax\|_2 \equiv \max_{x\in\mathbb{C}^n,\, \|x\|_2=1} \|Ax\|_2.$$

Παραθέτουμε στη συνέχεια μερικούς βασικούς ορισμούς και προτάσεις, που σχετίζονται με χαρακτηριστικές ιδιοτήτες πινάκων, οι οποίες έχουν σχέση με τη μορφή ή/και τα στοιχεία ενός πίνακα $A \in \mathbb{C}^{n\times n}$.

**Ορισμός 2.1.4.** : Ένας πίνακας $A \in \mathbb{C}^{n\times n}$ είναι Ερμιτιανός εάν και μόνο εάν $A^H = A$.

**Πρόταση 2.1.1.** *: Εάν ο πίνακας $A \in \mathbb{C}^{n\times n}$ είναι Ερμιτιανός τότε*

$$\|A\|_2 = \rho(A). \tag{2.1.4}$$

*Γενικότερα, εάν $g_m(x)$ είναι ένα πραγματικό πολυώνυμο βαθμού $m$ τότε*

$$\|g_m(A)\|_2 = \rho(g_m(A)). \tag{2.1.5}$$

Στην περίπτωση όπου ο πίνακας $A$ είναι Ερμιτιανός τότε έχουμε δύο σημαντικές σχέσεις για τη μέγιστη και την ελάχιστη ιδιοτιμή του πίνακα $A$:

$$\begin{aligned} \lambda_{\max} &= \max_{v\in\mathbb{C}^n\setminus\{0\}} \frac{(v,Av)}{(v,v)} = \frac{(\widetilde{v},A\widetilde{v})}{(\widetilde{v},\widetilde{v})} \\ \lambda_{\min} &= \min_{v\in\mathbb{C}^n\setminus\{0\}} \frac{(v,Av)}{(v,v)} = \frac{(\bar{v},A\bar{v})}{(\bar{v},\bar{v})}, \end{aligned} \tag{2.1.6}$$

όπου $\widetilde{v}, \bar{v}$ είναι τα ιδιοδιανύσματα που αντιστοιχούν στη μέγιστη και στην ελάχιστη ιδιοτιμή του πίνακα $A$.

Στη συνέχεια θα δώσουμε ένα ορισμό της θετικής ορισημότητας ενός Ερμιτιανού πίνακα.

**Ορισμός 2.1.5.** : Ένας Ερμιτιανός πίνακας $A \in \mathbb{C}^{n\times n}$ είναι θετικά ορισμένος εάν ικανοποιείται η σχέση

$$(v, Av) > 0,\ \forall v \in \mathbb{C}^n\setminus\{0\}. \tag{2.1.7}$$

(Σημείωση: Στη συνέχεια όταν αναφερόμαστε σε "θετικά ορισμένο" πίνακα θα θεωρούμε ότι ο υπόψη πίνακας είναι Ερμιτιανός)

Έχοντας δώσει τον ορισμό ενός Ερμιτιανού και θετικά ορισμένου πίνακα $A$ θα ορίσουμε την τετραγωνική ρίζα του.

**Θεώρημα 2.1.2.** *: Εάν ο πίνακας $A \in \mathbb{C}^{n\times n}$ είναι θετικά ορισμένος τότε υπάρχει μοναδικός θετικά ορισμένος πίνακας $B \in \mathbb{C}^{n\times n}$ τέτοιος ώστε*

$$B^2 = A. \tag{2.1.8}$$

Ο πίνακας $B$ χαρακτηρίζεται ως η τετραγωνική ρίζα του πίνακα $A$ και συμβολίζεται με $A^{\frac{1}{2}}$.

Στη συνέχεια δίνουμε τον ορισμό της $A-$νόρμας ενός διανύσματος.

**Ορισμός 2.1.6.** : Έστω $A \in \mathbb{C}^{n\times n}$, Ερμιτιανός και θετικά ορισμένος. Τότε, για κάθε $x \in \mathcal{C}^n$ η συνάρτηση

$$\|x\|_{A^{\frac{1}{2}}} = (Ax, x)^{\frac{1}{2}}, \tag{2.1.9}$$

ορίζει μια διανυσματική νόρμα η οποία καλείται $A-$νόρμα.

Στην συνέχεια θα δώσουμε τον ορισμό της $A-$προβολής ενός διανύσματος $u$ πάνω σε ένα διάνυσμα $w$.

**Ορισμός 2.1.7.** : Έστω $u, w \in \mathbb{C}^n$ και $A \in \mathbb{C}^{n\times n}$. Ορίζουμε την "$A-$προβολή" του διανύσματος $u$ πάνω στο διάνυσμα $w$ ως το διάνυσμα $\widehat{u}$ που δίνεται από την παρακάτω σχέση

$$\widehat{u} = \frac{(u, Aw)}{(w, Aw)} w. \tag{2.1.10}$$

Έχοντας δώσει τον παραπάνω ορισμό και γνωρίζοντας τον κλασικό ορισμό της καθετότητας δύο διανυσμάτων, μπορούμε να ορίσουμε την "$A-$καθετότητα". Θα λέμε λοιπόν ότι δύο διανύσματα $u, w \in \mathbb{C}^n$ είναι $A-$κάθετα εάν ικανοποιείται η σχέση

$$(u, Aw) = 0.$$

Παρακάτω θα παρουσιάσουμε μία ικανή και αναγκαία συνθήκη για τη σύγκλιση μίας ακολουθίας πινάκων της μορφής $A, A^2, A^3, \ldots, A^n$, με $A \in \mathbb{C}^{n\times n}$, στο μηδενικό πίνακα $O$.

**Θεώρημα 2.1.3.** *: Εάν $A \in \mathbb{C}^{n\times n}$, τότε ο πίνακας αυτός συγκλίνει (στο μηδενικό πίνακα) εάν και μόνο εάν $\rho(A) < 1$.*

Σε πολλές περιπτώσεις διακριτοποίησης ο πίνακας των συντελεστών των αγνώστων που καταλήγουμε είναι $L-$πίνακας. Συγκεκριμένα:

**Ορισμός 2.1.8.** : Ένας πίνακας $A \in \mathbb{R}^{n\times n}$ είναι $L-$πίνακας εάν ικανοποιούνται οι σχέσεις

$$\alpha_{i,i} > 0,\ i = 1, 2, 3, \ldots n, \tag{2.1.11}$$

και

$$\alpha_{i,j} \leq 0,\ i \neq j,\ i, j = 1, 2, 3, \ldots n. \tag{2.1.12}$$

Μια ειδική περίπτωση $L-$πίνακα είναι ο πίνακας Stieltjes ο οποίος ορίζεται ως εξής:

**Ορισμός 2.1.9.** : Ένας πίνακας $A \in \mathbb{R}^{n\times n}$ καλείται πίνακας Stieltjes εάν είναι θετικά ορισμένος και ικανοποιεί την ιδιότητα στη (2.1.12).

Έναν ισοδύναμο ορισμό του πίνακα Stieltjes, με τη βοήθεια δυο ακόμα ιδιοτήτων του πίνακα $A$, θα δώσουμε παρακάτω. Ξεκινάμε με τον ορισμό του Ασθενώς Διαγώνια Υπέρτερου πίνακα $A$.

**Ορισμός 2.1.10.** : Ένας πίνακας $A \in \mathbb{C}^{n\times n}$ είναι Ασθενώς Διαγώνια Υπέρτερος εάν

$$|\alpha_{i,i}| \geq \sum_{j=1, j\neq i}^{n} |a_{i,j}|\,,\ i = 1, 2, 3, \ldots, n, \tag{2.1.13}$$

και για μια τουλάχιστον τιμή του $i$ ισχύει ότι

$$|\alpha_{i,i}| > \sum_{j=1, j\neq i}^{n} |a_{i,j}|\,. \tag{2.1.14}$$

Μια δεύτερη σημαντική ιδιότητα είναι αυτή του να είναι ένας πίνακας $A \in \mathbb{C}^{n\times n}$ "μη αναγώγιμος" (irreducible).

**Ορισμός 2.1.11.** : Ένας πίνακας $A \in \mathbb{C}^{n\times n}$ είναι μη αναγώγιμος εάν δεν υπάρχει πίνακας μετάθεσης $P$ τέτοιος ώστε $P^{-1}AP$ να έχει τη μορφή

$$PAP^T = \begin{pmatrix} F & O \\ G & H \end{pmatrix}, \tag{2.1.15}$$

όπου $F$ και $H$ είναι τετραγωνικοί πίνακες.

Όπως αναφέραμε και προηγουμένως θα δώσουμε μία ικανή συνθήκη ώστε να είναι ένας πίνακας $A \in \mathbb{R}^{n\times n}$ πίνακας Stieltjes.

**Θεώρημα 2.1.4.** *: Εάν έχουμε έναν $L-$πίνακα ο οποίος είναι συμμετρικός, irreducible και είναι και ασθενώς διαγώνια υπέρτερος τότε, ο πίνακας αυτός είναι ένας πίνακας Stieltjes.*

Τέλος θα δώσουμε κάποιες πολύ σημαντικές ιδιότητες μιας κατηγορίας πινάκων που ικανοποιούν τη μεταθετική ιδιότητα, δηλαδή τετραγωνικούς πίνακες για τους οποίους ισχύει ότι

$$A_1 A_2 = A_2 A_1. \tag{2.1.16}$$

Δύο πολύ σημαντικά θεωρήματα που σχετίζονται με τους "αντιμεταθέσιμους" πίνακες είναι τα εξής:

**Θεώρημα 2.1.5.** *: Έστω ότι οι δύο πίνακες $A_1, A_2 \in \mathbb{C}^{n\times n}$ είναι Ερμιτιανοί. Τότε υπάρχει ορθοκανονική βάση ιδιοδιανυσμάτων $\xi_i$, $i = 1, 2, \ldots, n$, με την ιδιότητα ότι $A_1\xi_i = \lambda_i\xi_i$ και $A_2\xi_i = \mu_i\xi_i$ για $i = 1, 2, \ldots, n$, εάν και μόνο εάν $A_1A_2 = A_2A_1$.*

**Θεώρημα 2.1.6.** *: Έστω δύο Ερμιτιανοί πίνακες $A_1, A_2 \in \mathbb{C}^{n\times n}$. Τότε υπάρχει ένας $n \times n$ πίνακας $U$ για τον οποίο ισχύει ότι οι πίνακες $UA_1U^T$ και $UA_2U^T$ είναι και οι δύο διαγώνιοι πίνακες εάν και μόνο εάν $A_1A_2 = A_2A_1$.*

Το σημαντικότατο συμπέρασμα από τα δύο παραπάνω θεωρήματα είναι ότι στην περίπτωση πινάκων που ικανοποιούν τη μεταθετική ιδιότητα μπορούμε να βρούμε ορθοκανονική (εάν είναι και Ερμιτιανοί) βάση των ίδιων ιδιδιανυσμάτων με βέβαια όχι και υποχρεωτικά τις ίδιες ιδιοτιμές. Τα δύο παραπάνω συμπεράσματα θα αποτελέσουν και τη βάση για τη μελέτη των μεθόδων που θα παρουσιάσουμε στα επόμενα κεφάλαια.

Θα αναφερθούμε τέλος σε μια άλλη κατηγορία ιδιοτήτων που αφορούν στις πράξεις μεταξύ διανυσμάτων και πινάκων, και πιο συγκεκριμένα στην πράξη του τανυστικού γινομένου πινάκων ή αλλιώς γινομένου Kronecker.

**Ορισμός 2.1.12.** : Το τανυστικό γινόμενο ενός πίνακα $A \in \mathbb{C}^{m\times n}$ και ενός πίνακα $B \in \mathbb{C}^{p\times q}$ συμβολίζεται με $A\otimes B \in \mathbb{C}^{mp\times nq}$ και ορίζεται από τον "μπλοκ" πίνακα

$$A \otimes B = \begin{bmatrix} \alpha_{11}B & \ldots & \alpha_{1n}B \\ \vdots & \ddots & \vdots \\ \alpha_{m1} & \ldots & \alpha_{mn}B \end{bmatrix} \tag{2.1.17}$$

Με βάση τον παραπάνω ορισμό δίνουμε μια σειρά από ορισμούς και ιδιότητες για το τανυστικό γινόμενο πινάκων.

**Ορισμός 2.1.13.** : Έστω $A \in \mathbb{C}^{m\times n}$ τότε το τανυστικό γινόμενο $A^{\otimes k}$ ορίζεται ως

$$A^{\otimes k} = A \otimes A^{\otimes(k-1)},\ k = 2, 3, \ldots,\ \text{με } A^{\otimes 1} = A. \tag{2.1.18}$$

Θα απαριθμήσουμε μια σειρά από βασικές ιδιότητες του τανυστικού γινομένου.

$$\begin{aligned}
(aA)\otimes B &= A\otimes(aB),\ \forall a\in\mathbb{C},\ A\in\mathbb{C}^{m\times n},\ B\in\mathbb{C}^{p\times q} && (2.1.19)\\
(A\otimes B)^H &= A^H\otimes B^H,\ A\in\mathbb{C}^{m\times n},\ B\in\mathbb{C}^{p\times q} && (2.1.20)\\
(A\otimes B)\otimes C &= A\otimes(B\otimes C),\ A\in\mathbb{C}^{m\times n},\ B\in\mathbb{C}^{p\times q},\ C\in\mathbb{C}^{r\times s} && (2.1.21)\\
(A+B)\otimes C &= A\otimes C+B\otimes C,\ A,B\in\mathbb{C}^{m\times n},\ C\in\mathbb{C}^{p\times q} && (2.1.22)\\
A\otimes(B+C) &= A\otimes B+A\otimes C,\ A\in\mathbb{C}^{m\times n},\ B,C\in\mathbb{C}^{p\times q} && (2.1.23)
\end{aligned}$$

Στη συνέχεια θα δώσουμε δυο ιδιότητες των τανυστικών γινομένων μέσω των παρακάτω λημμάτων.

**Λήμμα 2.1.7.** *: Έστω $A\in\mathbb{C}^{m\times n}$, $B\in\mathbb{C}^{p\times q}$, $C\in\mathbb{C}^{n\times r}, D\in\mathbb{C}^{q\times s}$ τότε*

$$(A\otimes B)(C\otimes D) = AC\otimes BD \tag{2.1.24}$$

**Λήμμα 2.1.8.** *: Εάν $A\in\mathbb{C}^{m\times m}$ και $B\in\mathbb{C}^{n\times n}$ είναι αντιστρέψιμοι πίνακες, τότε ο πίνακας $A\otimes B$ είναι επίσης αντιστρέψιμος και ισχύει ότι $(A\otimes B)^{-1} = A^{-1}\otimes B^{-1}$*

**Λήμμα 2.1.9.** *: Εάν $A\in\mathbb{C}^{m\times m}$ και $B\in\mathbb{C}^{n\times n}$ είναι θετικά ορισμένοι πίνακες τότε και ο $A\otimes B$ είναι θετικά ορισμένος.*

Τέλος θα δώσουμε μέσω ενός θεωρήματος τη σχέση των ιδιοτιμών και των ιδιοδιανυσμάτων δυο τετραγωνικών πινάκων $A$, $B$ με τις ιδιοτιμές και τα ιδιοδιανύσματα του πίνακα $A\otimes B$.

**Θεώρημα 2.1.10.** *: Έστω $A\in\mathbb{C}^{m\times n}$ και $B\in\mathbb{C}^{n\times n}$. Εάν $\lambda\in\sigma(A)$ και $x\in\mathbb{C}^m$ το αντίστοιχο ιδιοδιάνυσμα, και εάν $\mu\in\sigma(B)$ και $y\in\mathbb{C}^n$ το αντίστοιχο ιδιοδιάνυσμα, τότε $\lambda\mu\in\sigma(A\otimes B)$ και $x\otimes y\in\mathbb{C}^{mn}$ είναι το αντίστοιχο ιδιοδιάνυσμα του $A\otimes B$. Επιπλέον έχουμε ότι $\sigma(A\otimes B) = \sigma(B\otimes A)$.*

Σημείωση: Στο παραπάνω θεώρημα καθώς και στη συνέχεια, με το σύμβολο $\sigma(X)$, $X \in \mathbb{C}^{n\times n}$, θα εννοούμε το σύνολο των ιδιοτιμών (φάσμα των ιδιοτιμών) του πίνακα $X$, με την έννοια ότι

$$\sigma(X) := \{\lambda_1, \lambda_2, \ldots, \lambda_k\},$$

όπου $\lambda_i$, $i = 1, 2, \ldots, k$, οι διαφορετικές ιδιοτιμές του $X$ χωρίς τις πολλαπλότητες τους.

Οι αποδείξεις όλων των παραπάνω Θεωρημάτων, Λημμάτων, Πορισμάτων και Προτάσεων μπορούν να βρεθούν στα κλασικά συγγράμματα των Varga [55], Young [64] καθώς και στο δίτομο σύγγραμμα των Horn και Jonson [37], [38].

# Κεφάλαιο 3

# Διακριτά Σχήματα Πεπερασμένων Διαφορών

## 3.1 Εξίσωση Poisson, Διακριτοποίηση $5-$ και $9-$Σημείων.

Αρχίζουμε το παρόν κεφάλαιο της διατριβής με τη διακριτοποίηση της εξίσωσης Poisson, με Dirichlet συνοριακές συνθήκες, σε ορθογώνια γενικά χωρία. Για το σκοπό αυτό θεωρούμε την εξίσωση Poisson στις $p-$διαστάσεις:

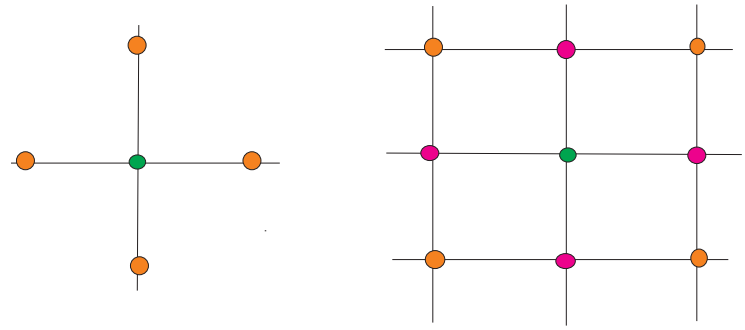
$$-\Delta u = -\sum_{i=1}^{p} \frac{\partial^2 u}{\partial x_i^2} = f(x),\ x \in \Omega, \tag{3.1.1}$$

με συνοριακές συνθήκες της μορφής

$$u = g(x),\ x \in \partial\Omega, \tag{3.1.2}$$

όπου $x = (x_1, x_2, \ldots, x_p)$.

Στο παρόν κεφάλαιο θα επικεντρωθούμε στη διακριτοποίηση της εξίσωσης Poisson με χρήση σχημάτων πεπερασμένων διαφορών δεύτερης και τέταρτης τάξης ακρίβειας. Τη διακριτοποίηση αυτή θα την εφαρμόσουμε αρχικά στην περίπτωση των δύο διαστάσεων, χρησιμοποιώντας διακριτοποίηση του διαφορικού τελεστή $\Delta u$ στα $5-$σημεία και στη συνέχεια διακριτοποίηση του ίδιου τελεστή στα $9-$σημεία, όπως αυτά παρουσιάζονται στα παρακάτω πλέγματα (βλ. Σχήματα 3.1(α), (β), αντίστοιχα).



Σχήμα 3.1: Διακριτά Πλέγματα (Stencils) 5− και 9−σημείων.

Αρχίζουμε την ανάλυσή μας περιοριζόμενοι στη δεύτερης τάξης Εξίσωση μορφής Poisson ορισμένης στην περιοχή του επιπέδου, που περιγράφεται από το ορθογώνιο χωρίο

$$\Omega := \left\{(x_1, x_2) \in \mathbb{R}^2 | 0 < x_1 < c,\, 0 < x_2 < d\right\},$$

και η οποία δίνεται από την

$$-a(x_1,x_2)u_{x_1x_1}(x_1,x_2) - b(x_1,x_2)u_{x_2x_2}(x_1,x_2) = f(x_1,x_2),\;\; f(x_1,x_2) \in C^2(\Omega), \tag{3.1.3}$$

όπου $u$ είναι η άγνωστη συνεχώς διαφορίσιμη συνάρτηση στο χωρίο $\Omega$, με γνωστή έκφραση $u(x_1, x_2) = g(x_1, x_2)$ στο σύνορο $\partial\Omega$. Οι συναρτήσεις $a := a(x_1, x_2)$ και $b := b(x_1, x_2)$ είναι συνεχείς θετικές συναρτήσεις. Στις περιπτώσεις που θα εξετάσουμε, θεωρούμε ότι οι συναρτήσεις $a, b$ είναι θετικές σταθερές και ότι $c = d = 1$, για την απλοποίηση των εκφράσεων που θα προκύψουν. Για τη διακριτοποίηση της παραπάνω διαφορικής εξίσωσης θεωρούμε ένα ομοιόμορφο διαμερισμό της κλειστότητας $\overline{\Omega}$ με βήμα διακριτοποίησης $h_1$ και $h_2$ στην $x_1-$ και $x_2-$ διεύθυνση αντίστοιχα. Οπότε, το διακριτό ανάλογο της (3.1.3) δίνεται από την εξίσωση

$$\begin{aligned} & \sqrt{\frac{a}{b}}\frac{h_2}{h_1}(-u_{i-1,j} + 2u_{ij} - u_{i+1,j}) + \sqrt{\frac{b}{a}}\frac{h_1}{h_2}(-u_{i,j-1} + 2u_{ij} - u_{i,j+1}) \\ - \;& \theta\,[4u_{ij} - 2(u_{i-1,j} + u_{i+1,j} + u_{i,j-1} + u_{i,j+1}) \\ + \;& u_{i-1,j-1} + u_{i+1,j-1} + u_{i-1,j+1} + u_{i+1,j+1}] = \frac{h_1h_2}{\sqrt{ab}}(f_{ij} + \phi_{ij}), \end{aligned} \tag{3.1.4}$$

όπου οι παράμετροι $\theta$ και $\phi$ δίνονται από τις παρακάτω εκφράσεις

$$(\theta,\phi) \;=\; \begin{cases} (0,0), \\ (\theta^*,\phi^*) = \left(\frac{1}{12}(\sqrt{\frac{a}{b}}\frac{h_2}{h_1} + \sqrt{\frac{b}{a}}\frac{h_1}{h_2}),\, \frac{1}{12}(ah_1^2 f_{x_1x_1} + bh_2^2 f_{x_2x_2})\right). \end{cases} \tag{3.1.5}$$

Το παραπάνω διακριτό σχήμα αποτελεί μία ομαδοποιημένη έκφραση του κλασικού σχήματος των $5-$σημείων και ενός σχήματος $9-$σημείων. Ειδικότερα, στην περίπτωση όπου $\theta = 0$ λαμβάνουμε το σχήμα διακριτοποίησης των $5-$σημείων τάξης ακρίβειας $\mathcal{O}(h^2)$, όταν $h_1 = h_2 = h$. Στην περίπτωση όπου $\theta = \theta^*$ έχουμε ένα σχήμα διακριτοποίησης $9-$σημείων τάξης ακρίβειας $\mathcal{O}(h^4)$, όταν $h_1 = h_2 = h$ (βλ. [48] ή [14]).

Για το σχήμα των $9-$σημείων θα ήταν σκόπιμο να δώσουμε μερικές επιπλέον επεξηγήσεις. Αρχικά ένα τέτοιο σχήμα παρουσιάστηκε και μελετήθηκε από τους Kantorovich και Krylov [41] οι οποίοι απέδειξαν ότι στην περίπτωση όπου $h_1 = h_2 = h$, η τάξη ακρίβειας του σχήματος για την εξίσωση Poisson είναι $\mathcal{O}(h^4)$, ενώ για την εξίσωση Laplace η τάξη είναι $\mathcal{O}(h^6)$. Οι αποδείξεις και στις δύο περιπτώσεις γίνονται με την χρήση αναπτυγμάτων Taylor στις δύο μεταβλητές (βλ. [41]).

Στη συνέχεια θα παρουσιάσουμε με περισσότερες λεπτομέρειες το σχήμα των $9-$σημείων στην περίπτωση όπου θεωρούμε ομοιόμορφο διαμερισμό και στις δύο διευθύνσεις $x_1$ και $x_2$, αλλά με διαφορετικό βήμα διακριτοποίησης σε κάθε διεύθυνση.

Ας δούμε, λοιπόν, το σχήμα των $9-$σημείων εφαρμοσμένο στην εξίσωση μορφής Poisson στις δύο διαστάσεις (βλ. 3.1.3). Εκφράζουμε την εξίσωση αυτή με τη βοήθεια των τελεστών $L_1, L_2$ ως εξής

$$\begin{gathered} -(a(x_1,x_2)L_1 + b(x_1,x_2)L_2)u(x_1,x_2) \;=\; f(x_1,x_2), \\ L_1 \;=\; \tfrac{\partial^2}{\partial x_1^2}, \;\; L_2 \;=\; \tfrac{\partial^2}{\partial x_2^2}. \end{gathered} \tag{3.1.6}$$

Όπως αναφέραμε προηγουμένως περιοριζόμαστε στην περίπτωση όπου $a(x_1,x_2)$, $b(x_1,x_2)$ είναι θετικές σταθερές, οι οποίες θα παραλειφθούν στη συνέχεια, για την απλοποίηση των υπολογισμών, και θα εισαχθούν στις τελικές εκφράσεις. Ο αντίστοιχος διακριτός τελεστής εκφράζεται από τη σχέση

$$\Lambda u = -(\Lambda_1 + \Lambda_2)u, \tag{3.1.7}$$

με $\Lambda_1$, $\Lambda_2$ τα διακριτά ανάλογα των $L_1$ και $L_2$. Με χρήση αναπτυγμάτων Taylor δύο μεταβλητών έχουμε για τον τελεστή του σφάλματος

$$Eu = \Lambda u - \Delta u = \frac{h_1^2}{12}L_1^2 u + \frac{h_2^2}{12}L_2^2 u + \mathcal{O}(h^4), \tag{3.1.8}$$

όπου $h^4 = \max\{h_1^4, h_2^4\}$. Εφαρμόζοντας διαδοχικά τους τελεστές $L_1, L_2$ στην εξίσωση (3.1.6) λαμβάνουμε τις παρακάτω εκφράσεις

$$-L_1^2 u = L_1 f + L_1 L_2 u, \ -L_2^2 u = L_2 f + L_2 L_1 u, \tag{3.1.9}$$

με τους τελεστές $L_1, L_2$ να αντιμετατίθενται. Επομένως μπορούμε να αντικαταστήσουμε τον τελεστή $L_2 L_1$ με τον $L_1 L_2$ και έτσι η (3.1.8) λαμβάνει τη μορφή

$$\Lambda u = -\Delta u + \frac{h_1^2}{12} L_1 f + \frac{h_2^2}{12} L_2 f + \frac{h_1^2 + h_2^2}{12} L_1 L_2 u + \mathcal{O}(h^4). \tag{3.1.10}$$

Αντικαθιστώντας από την αρχική εξίσωση Poisson την $-\Delta u = f$ και το διαφορικό τελεστή $L_1 L_2$ με ένα διακριτό ανάλογο της μορφής $\Lambda_1 \Lambda_2$ λαμβάνουμε το διακριτό ανάλογο της εξίσωσης Poisson τάξης $\mathcal{O}(h^4)$. Ο διακριτός τελεστής μπορεί να εκφραστεί με εφαρμογή ενός διακριτού πλέγματος $9-$σημείων όπως αυτό φαίνεται στο σχήμα 3.1(β). Έτσι η έκφραση για το διακριτό τελεστή είναι η εξής

$$\begin{aligned}
\Lambda_1 \Lambda_2 u &= \Lambda_1 \left[ \frac{u(x_1, x_2 - h_2) - 2u(x_1, x_2) + u(x_1, x_2 + h_2)}{h_2^2} \right] \\
&= \frac{1}{h_1^2 h_2^2} \Big\{ u(x_1 - h_1, x_2 - h_2) - 2u(x_1, x_2 - h_2) \\
&+ u(x_1 + h_1, x_2 - h_2) + 4u(x_1, x_2) \\
&- 2u(x_1 - h_1, x_2) + u(x_1 - h_1, x_2 + h_2) - 2u(x_1, x_2 + h_2) \\
&- 2u(x_1 + h_1, x_2) + u(x_1 + h_1, x_2 + h_2) \Big\}
\end{aligned} \tag{3.1.11}$$

Η έκφραση

$$\frac{u(x_1, x_2 - h_2) - 2u(x_1, x_2) + u(x_1, x_2 + h_2)}{h_2^2}$$

αντιστοιχεί στη δεύτερη τάξης προσέγγιση της μερικής παραγώγου της $u_{x_2 x_2}$. Γενικότερα, θεωρώντας τις κεντρικές διαφορές δεύτερης τάξης έχουμε ότι

$$\Lambda v = v_{xx} = \frac{-v(x+h) + 2v(x) - v(x-h)}{h^2} = v''(\xi), \ \xi = x + \theta h, \ |\theta| \leq 1. \tag{3.1.12}$$

Θεωρώντας ότι η $v(x)$ έχει συνεχή δεύτερη παράγωγο στο διάστημα $[x-h, x+h]$ προκύπτει ότι

$$\Lambda v = v_{xx} = v''(x) + \frac{h^2}{12} v^{(4)}(\xi^*), \ \xi^* = x + \theta^* h, \ |\theta^*| \leq 1. \tag{3.1.13}$$

Στην περίπτωση των δύο μεταβλητών θεωρώντας ότι η μία, π.χ. η $x_1$, παραμένει σταθερή, έχουμε την έκφραση

$$\Lambda_2 u = L_2 u(x_1, x_2) + \frac{h_2^2}{12}\frac{\partial^4 u}{\partial x_2^4}(x_1, \xi_2),\ \xi_2 = x_2 + \theta_2 h,\ |\theta_2| \leq 1. \qquad (3.1.14)$$

Στην περίπτωση του τελεστή $\Lambda_1\Lambda_2 u$ λαμβάνουμε μία ανάλογη έκφραση

$$\Lambda_1\Lambda_2 u(x_1, x_2) = \Lambda_1 L_2 u(x_1, x_2) + \frac{h_2^2}{12}\Lambda_1\frac{\partial^4 u}{\partial x_2^4}(x_1, \xi_2),\ \xi_2 = x_2 + \theta_2 h,\ |\theta_2| \leq 1. \qquad (3.1.15)$$

Εφαρμόζοντας τώρα την (3.1.13) με $v = L_2 u$ και $x = x_1$ στον πρώτο όρο έχουμε ότι

$$\Lambda_1 L_2 u(x_1, x_2) = L_1 L_2 u(x_1, x_2) + \frac{h_1^2}{12}\frac{\partial^4 u}{\partial x_1^4}(\xi_1, x_2),\ \xi_1^* = x_1 + \theta_1^* h_1,\ |\theta_1^*| \leq 1. \qquad (3.1.16)$$

Με όμοιο τρόπο εφαρμόζοντας την παραπάνω διαδικασία και για το δεύτερο όρο, σύμφωνα με τη σχέση (3.1.12), βρίσκουμε ότι

$$\frac{h_2^2}{12}\Lambda_1\frac{\partial^4 u}{\partial x_2^4}(x_1, \xi_2) = \frac{h_2^2}{12}\frac{\partial^6 u}{\partial x_1^2 \partial x_2^4}(\xi_1, \xi_2), \xi_1 = x_1 + \theta_1 h,\ |\theta_1| \leq 1, \qquad (3.1.17)$$

με το σφάλμα για τον τελεστή $\Lambda_1\Lambda_2 - L_1 L_2$ να έχει τάξη ακρίβειας

$$(\Lambda_1\Lambda_2 - L_1 L_2)u = \mathcal{O}(h_1^2) + \mathcal{O}(h_2^2) = \mathcal{O}(h^2) \qquad (3.1.18)$$

Αντικαθιστώντας στη σχέση (3.1.10) την έκφραση του $\Lambda_1\Lambda_2$ στη θέση της $L_1 L_2$ και γνωρίζοντας ότι $-\Delta u = f$ παίρνουμε τη σχέση

$$\Lambda u = \left(f + \frac{h_1^2}{12}L_1 f + \frac{h_2^2}{12}L_2 f\right) + \frac{h_1^2 + h_2^2}{12}\Lambda_1\Lambda_2 u + \mathcal{O}(h^4). \qquad (3.1.19)$$

Στη συνέχεια, ορίζοντας

$$\Lambda' u \ = \ \Lambda u + \frac{h_1^2 + h_2^2}{12}\Lambda_1\Lambda_2 u, \quad \phi \ = \ -\left(f + \frac{h_1^2}{12}L_1 f + \frac{h_2^2}{12}L_2 f\right), \qquad (3.1.20)$$

ο τελεστής $\Lambda'$ για το πλέγμα των $9-$σημείων γράφεται τελικά στην παρακάτω

μορφή, όπου εισάγονται και οι σταθερές $a$ και $b$

$$\begin{aligned}\frac{5}{3}\left(\frac{a}{h_1^2}+\frac{b}{h_2^2}\right)u &= \frac{1}{6}\left(\frac{5a}{h_1^2}-\frac{b}{h_2^2}\right)(u(x_1+h_1,x_2)+u(x_1-h_1,x_2))\\ &+ \frac{1}{6}\left(\frac{5b}{h_2^2}-\frac{a}{h_1^2}\right)(u(x_1,x_2+h_2)+u(x_1,x_2-h_2))\\ &+ \frac{1}{12}\left(\frac{a}{h_1^2}+\frac{b}{h_2^2}\right)(u(x_1+h_1,x_2+h_2)+u(x_1+h_1,x_2-h_2)\\ &+ u(x_1-h_1,x_2-h_2)+u(x_1-h_1,x_2+h_2))+\phi. \quad (3.1.21)\end{aligned}$$

Θα πρέπει επίσης να τονίσουμε ότι στην περίπτωση των $9-$σημείων ο διακριτός τελεστής (3.1.21) είναι θετικά ορισμένος εάν ικανοποιείται η συνθήκη

$$\frac{1}{5} \;\le\; \frac{bh_1^2}{ah_2^2} \;\le\; 5 \qquad (3.1.22)$$

(βλ. [48]). Η απόδειξη της παραπάνω συνθήκης είναι σχετικά απλή αφού αρκεί οι όροι εντός των παρενθέσεων του σχήματος (3.1.21) να είναι θετικοί. Αυτό συμβαίνει στην περίπτωση που ικανοποιείται η εν λόγω συνθήκη. Σημειώνουμε εδώ ότι στην περίπτωση του διακριτού σχήματος των $5-$σημείων έχουμε ότι ο διακριτός τελεστής είναι για κάθε επιλογή των $h_1, h_2$ θετικά ορισμένος.

Γράφοντας το διακριτό σχήμα (3.1.21) σε μορφή συστήματος, ο πίνακας των συντελεστών $A$ θα λαμβάνει τη μορφή

$$A=\sqrt{\frac{a}{b}}\frac{h_2}{h_1}(I_{n_2}\otimes T_{n_1})+\sqrt{\frac{b}{a}}\frac{h_1}{h_2}(T_{n_2}\otimes I_{n_1})-\theta(T_{n_2}\otimes T_{n_1}) \qquad (3.1.23)$$

ή ισοδύναμα

$$\begin{aligned}A &= \sqrt{\tfrac{a}{b}}\tfrac{h_2}{h_1}(I_{n_2}\otimes T_{n_1})+\sqrt{\tfrac{b}{a}}\tfrac{h_1}{h_2}(T_{n_2}\otimes I_{n_1})\\ &\quad -\theta\left[\sqrt{\tfrac{a}{b}}\tfrac{h_2}{h_1}(I_{n_2}\otimes T_{n_1})\cdot\sqrt{\tfrac{b}{a}}\tfrac{h_1}{h_2}(T_{n_2}\otimes I_{n_1})\right],\end{aligned} \qquad (3.1.24)$$

όπου $n_1\,(\ge 2)$ και $n_2\,(\ge 2)$ είναι οι αριθμοί που εκφράζουν το πλήθος των εσωτερικών κόμβων του διακριτού πλέγματος, ενώ οι πίνακες $T_{n_1}\in\mathbb{R}^{n_1\times n_1}$ και $T_{n_2}\in\mathbb{R}^{n_2\times n_2}$ είναι της μορφής $\text{tridiag}(-1,2,-1)$. Το σύμβολο $\otimes$ είναι αυτό του τανυστικού γινομένου δύο πινάκων (βλ. Halmos [36] ή Horn και Jonhson [38]). Η πρώτη έκφραση του πίνακα συντελεστών (3.1.23) προέρχεται άμεσα από το διακριτό σχήμα (3.1.21) ενώ η δεύτερη αποτελεί μια ισοδύναμη μορφή της

πρώτης περισσότερο πρακτική για τη μελέτη που θα ακολουθήσει. Θέτοντας στη συνέχεια

$$A_1 := \sqrt{\frac{a}{b}}\frac{h_2}{h_1}(I_{n_2} \otimes T_{n_1}) \quad \text{και} \quad A_2 := \sqrt{\frac{b}{a}}\frac{h_1}{h_2}(T_{n_2} \otimes I_{n_1}),$$

η (6.1.25) γίνεται

$$A = A_1 + A_2 - \theta A_1 A_2. \tag{3.1.25}$$

Με απλούς υπολογισμούς μπορούμε να δείξουμε ότι οι πίνακες $A_1$ και $A_2$ αντιμετατίθενται (βλ. [36]). Η συνθήκη αυτή αποτελεί και το βασικότερο εργαλείο για τη μελέτη των μεθόδων επίλυσης των συστημάτων με πίνακα συντελεστών τον πίνακα $A$ του παρόντος κεφαλαίου αλλά και τους πίνακες $A_1$ και $A_2$.

# Κεφάλαιο 4

# Πεπλεγμένες Επαναληπτικές Μέθοδοι Εναλλασσόμενων Διευθύνσεων (ADI)

## 4.1 Κλασικά Σχήματα Πεπλεγμένων Επαναληπτικών Μεθόδων Εναλλας σόμενων Διευθύνσεων (ADI)

Στο κεφάλαιο αυτό θα ασχοληθούμε με την παρουσίαση των Πεπλεγμένων Επαναληπτικών Μεθόδων Εναλλασσόμενων Διευθύνσεων (Alternating Direction Implicit Methods) γνωστές με το αρκτικόλεξο ADI.

Οι μέθοδοι ADI εμφανίστηκαν αρχικά σε μια εργασία των Peaceman και Rachford το 1955 (βλ. [45]) οι οποίες παρουσίασαν μια νέα επαναληπτική μέθοδο (παράγωγο της μεθόδου των Crank-Nickolson) για τη λύση Παραβολικών και Ελλειπτικών προβλημάτων Μερικών Διαφορικών Εξισώσεων. Στην περίπτωση των ελλειπτικών εξισώσεων και πιο συγκεκριμένα στην περίπτωση της εξίσωσης Poisson, με διακριτοποίηση $5-$σημείων, η μέθοδος την οποία οι Peaceman και Rachford πρότειναν, στηρίζεται στη διάσπαση του πίνακα $A$ του συστήματος

$$Ax = b,\ A \in \mathbb{R}^{n_1 n_2 \times n_1 n_2},\ b \in \mathbb{R}^{n_1 n_2}, \tag{4.1.1}$$

στη μορφή

$$A = A_1 + A_2 \tag{4.1.2}$$

όπου $A_1 := \sqrt{\frac{a}{b}}\frac{h_2}{h_1}(I_{n_2} \otimes T_{n_1})$ και $A_2 := \sqrt{\frac{b}{a}}\frac{h_1}{h_2}(T_{n_2} \otimes I_{n_1})$ με $n_1\,(\geq 2)$ και $n_2\,(\geq 2)$ να είναι το πλήθος των εσωτερικών κόμβων του διακριτού πλέγματος στις δύο διευθύνσεις, ενώ οι πίνακες $T_{n_1} \in \mathbb{R}^{n_1 \times n_1}$ και $T_{n_2} \in \mathbb{R}^{n_2 \times n_2}$ είναι τριδιαγώνιοι της μορφής $\text{tridiag}(-1, 2, -1)$.

Το επαναληπτικό σχήμα των Peaceman-Rachford, που αντιστοιχεί στην παραπάνω διάσπαση, είναι το εξής

$$\begin{aligned} (r_{m+1}I + A_1)u^{(m+\frac{1}{2})} &= (r_{m+1}I - A_2)u^{(m)} + b, \\ (r_{m+1}I + A_2)u^{(m+1)} &= (r_{m+1}I - A_1)u^{(m+\frac{1}{2})} + b, \end{aligned} \tag{4.1.3}$$

όπου $r_{m+1}$ είναι θετικές παράμετροι, η κατάλληλη επιλογή των οποίων μπορεί να αυξήσει τη μέση ασυμπτωτική ταχύτητα σύγκλισης της μεθόδου. Με την πάροδο του χρόνου παρουσιάστηκαν αρκετές παραλλαγές της αρχικής μεθόδου των Peaceman-Rachford. Πρώτα, το 1956, οι Douglas και Rachford (βλ. [20]) πρότειναν μία παραλλαγή της αρχικής μεθόδου όπου με αντικατάσταση του όρου $A_1u^{(m+\frac{1}{2})}$ από την πρώτη εξίσωση της σχέσης (4.1.3) στη δεύτερη παίρνουμε το σχήμα

$$\begin{aligned} (r_{m+1}I + A_1)u^{(m+\frac{1}{2})} &= (r_{m+1}I - A_2)u^{(m)} + b, \\ (r_{m+1}I + A_2)u^{(m+1)} &= A_2u^{(m)} + r_{m+1}u^{(m+\frac{1}{2})}. \end{aligned} \tag{4.1.4}$$

Χαρακτηριστικό του σχήματος (4.1.4) είναι ότι η δεύτερη εξίσωση δεν εξαρτάται από τον πίνακα $A_1$, ο οποίος ουσιαστικά εκφράζει τη διακριτή μορφή της εξίσωσης Poisson στη $x-$διεύθυνση, γεγονός που κάνει τη μέθοδό μας περισσότερο ευέλικτη σε ό,τι αφορά τον παραλληλισμό.

Μία γενίκευση του παραπάνω σχήματος είναι η εξής

$$\begin{aligned} (r_{m+1}I + A_1)u^{(m+\frac{1}{2})} &= (r_{m+1}I - A_2)u^{(m)} + b, \\ (r_{m+1}I + A_2)u^{(m+1)} &= (A_2 - (1-\omega)r_{m+1}I)u^{(m)} + (2-\omega)r_{m+1}u^{(m+\frac{1}{2})}, \end{aligned} \tag{4.1.5}$$

όπου $\omega$ είναι μία σταθερή παράμετρος η κατάλληλη επιλογή της οποίας μπορεί να επιταχύνει τη σύγκλιση της μεθόδου. Στην περίπτωση όπου $\omega = 0$ έχουμε το σχήμα των Peaceman-Rachford και για $\omega = 1$ έχουμε το σχήμα των Douglas-Rachford. Εάν εκφράσουμε την παραπάνω μέθοδο στη μορφή μίας κλασικής επαναληπτικής μεθόδου, όπως φαίνεται παρακάτω

$$u^{(m+1)} = T_{r_{m+1}}(\omega)u^{(m)} + c(r_{m+1}, \omega), \tag{4.1.6}$$

όπου

$$\begin{aligned} T_{r_{m+1}}(\omega) &= (A_2 + r_{m+1}I)^{-1}(A_1 + r_{m+1}I)^{-1} \times \\ &\quad \left[A_1A_2 + (\omega - 1)r_{m+1}(A_1 + A_2) + r_{m+1}^2 I\right], \\ c(r_{m+1}, \omega) &= (2 - \omega)r_{m+1}(A_2 + r_{m+1}I)^{-1}(A_1 + r_{m+1}I)^{-1}b, \quad (4.1.7) \end{aligned}$$

παρατηρούμε ότι ο επαναληπτικός πίνακας της μεθόδου σχετίζεται άμεσα με τον επαναληπτικό πίνακα της μεθόδου των Peaceman-Rachford $T_{r_{m+1}}$ και δίνεται από την έκφραση

$$T_{r_{m+1}}(\omega) = \frac{1}{2}(\omega I + (2 - \omega)T_{r_{m+1}}. \quad (4.1.8)$$

Την παραπάνω μορφή θα μπορούσαμε να τη θεωρήσουμε ως μία παρεκβαλλόμενη (extrapolated) μέθοδο των Peaceman-Rachford με παράμετρο παρεκβολής $2 - \omega$. Η (4.1.8) μας δίνει μία σχέση μεταξύ του φάσματος των ιδιοτιμών του πίνακα $T_{r_{m+1}}(\omega)$ και του φάσματος ιδιοτιμών του πίνακα $T_{r_{m+1}}$, και κατ΄ επέκταση ενός φράγματος της φασματικής ακτίνας του $T_{r_{m+1}}(\omega)$

$$\rho(T_{r_{m+1}}(\omega)) \leq \frac{1}{2}(\omega I + (2 - \omega)\rho(T_{r_{m+1}})), \ 0 \leq \omega \leq 2. \quad (4.1.9)$$

Μια δεύτερη σημαντική παραλλαγή του βασικού σχήματος των Peaceman-Rachford έγινε από τους Wachspress και Habetler το 1960 (βλ. [63]). Ο μοναδιαίος πίνακας που χρησιμοποιήθηκε στα προηγούμενα επαναληπτικά σχήματα μπορεί να αντικατασταθεί με κάποιο θετικό πραγματικό διαγώνιο πίνακα $F$, όποτε το σχήμα (4.1.3) λαμβάνει τώρα τη μορφή

$$\begin{aligned} (r_{m+1}F + A_1)u^{(m+\frac{1}{2})} &= (r_{m+1}F - A_2)u^{(m)} + b, \\ (r_{m+1}F + A_2)u^{(m+1)} &= (r_{m+1}F - A_1)u^{(m+\frac{1}{2})} + b. \quad (4.1.10) \end{aligned}$$

Θέτοντας $F^{\frac{1}{2}}u = v$ το σχήμα (4.1.10) γίνεται

$$\begin{aligned} (r_{m+1}I + \widetilde{A}_1)v^{(m+\frac{1}{2})} &= (r_{m+1}I - \widetilde{A}_2)v^{(m)} + \widetilde{b}, \\ (r_{m+1}I + \widetilde{A}_2)v^{(m+1)} &= (r_{m+1}I - \widetilde{A}_1)v^{(m+\frac{1}{2})} + \widetilde{b}, \quad (4.1.11) \end{aligned}$$

όπου

$$\widetilde{A}_1 = F^{-\frac{1}{2}}A_1F^{-\frac{1}{2}}, \ \widetilde{A}_2 = F^{-\frac{1}{2}}A_2F^{-\frac{1}{2}}, \ \widetilde{b} = F^{-\frac{1}{2}}b.$$

Η μορφή (4.1.10) είναι γενικότερη της μεθόδου των Peaceman-Rachford αλλά με ένα σημαντικό μειονέκτημα το οποίο παρουσιάζεται στην επιλογή του

πίνακα $F$, ο οποίος πρέπει να είναι τέτοιος ώστε οι πίνακες $\widetilde{A}_1$ και $\widetilde{A}_2$ να αντιμετατίθενται. Οι Wachspress και Habetler απέδειξαν ότι για το πρόβλημα μοντέλο της εξίσωσης Poisson με Dirichlet συνοριακές συνθήκες στο μοναδιαίο τετράγωνο και στην περίπτωση ακόμα που δεν έχουμε ίσα βήματα διακριτοποίησης σε κάθε διεύθυνση μπορούμε να βρούμε πίνακα $F$ ο οποίος να μην προξενεί πρόβλημα αντιμετάθεσης των πινάκων $\widetilde{A}_1$ και $\widetilde{A}_2$. Γενικότερα όμως μπορεί να αποδειχτεί ότι και στην περίπτωση των γενικότερων ορθογώνιων χωρίων, ακόμη και χωρίς να είναι ορθογώνια, η συνθήκη αντιμετάθεσης των πινάκων $\widetilde{A}_1$ και $\widetilde{A}_2$ μπορεί να ικανοποιείται με κατάλληλη επιλογή του πίνακα $F$. Οι Wachspress και Habetler πρότειναν τη χρήση του πίνακα $F$ με την προοπτική ότι αυτός θα μπορούσε να ήταν ένας πίνακας ρυθμιστής (conditioning matrix) των $A_1$ και $A_2$.

## 4.2 Ανάλυση Βασικών Σχημάτων με Σταθερές Παραμέτρους Επιτάχυνσης

### 4.2.1 Εισαγωγή

Θεωρούμε το βασικό επαναληπτικό σχήμα των Peaceman-Rachford (4.1.3) με την προϋπόθεση ότι στην θέση των μεταβλητών παραμέτρων $r_{m+1}$, που εμφανίζονται ανά επανάληψη, θεωρούμε τις σταθερές παραμέτρους $r_1$ και $r_2$. Έτσι το σχήμα (4.1.3) λαμβάνει την μορφή

$$\begin{aligned} (r_1 I + A_1)u^{(m+\frac{1}{2})} &= (r_1 I - A_2)u^{(m)} + b, \\ (r_2 I + A_2)u^{(m+1)} &= (r_2 I - A_1)u^{(m+\frac{1}{2})} + b. \end{aligned} \tag{4.2.1}$$

Το αντίστοιχο επαναληπτικό σχήμα να έχει τη γενική μορφή

$$u^{(m+1)} = T_{r_1,r_2}u^{(m)} + c, \tag{4.2.2}$$

όπου επαναληπτικός πίνακας είναι ο

$$\begin{aligned} T_{r_1,r_2} &= (A_2 + r_2 I)^{-1}(A_1 - r_2 I)(A_1 + r_1 I)^{-1}(A_2 - r_2 I) \\ &= I - (r_1 + r_2)(A_2 + r_2 I)^{-1}(A_1 + r_1 I)^{-1}A \end{aligned} \tag{4.2.3}$$

και το διάνυσμα $c$ έχει την μορφή

$$c_{r_1,r_2} = (r_1 + r_2)(A_2 + r_2 I)^{-1}(A_1 + r_1 I)^{-1}b \tag{4.2.4}$$

Μπορεί να αποδειχτεί ότι στην περίπτωση όπου οι σταθερές $r_1$ και $r_2$ είναι θετικοί αριθμοί και οι πίνακες $A_1$ και $A_2$ θετικά ορισμένοι πραγματικοί πίνακες τότε η φασματική ακτίνα $\rho(T_{r_1,r_2}) < 1$. Επομένως η επαναληπτική μέθοδος συγκλίνει (βλ. Young [64]).

### 4.2.2 Επιλογή Παραμέτρων Επιτάχυνσης

Στην παρούσα παράγραφο θα μελετήσουμε διάφορες περιπτώσεις επιλογής σταθερών παραμέτρων επιτάχυνσης της μεθόδου των Peaceman-Rachford. Θα πρέπει να σημειώσουμε εδώ ότι η επιλογή των παραμέτρων θα γίνει έτσι ώστε να έχουμε "καλή" σύγκλιση και όχι πάντοτε "βέλτιστη" την οποία θα παρουσιάσουμε σε επόμενες παραγράφους.

Θα ξεκινήσουμε με την παρουσίαση της διαδικασίας εύρεσης των παραμέτρων υποθέτοντας αρχικά ότι οι ιδιοτιμές $\lambda_1$ και $\lambda_2$ των πινάκων $A_1$ και $A_2$, αντίστοιχα, ανήκουν στα διαστήματα $\alpha_1 \leq \lambda_1 \leq \beta_1$ και $\alpha_2 \leq \lambda_2 \leq \beta_2$. Τότε η φασματική ακτίνα του επαναληπτικού πίνακα θα έχει τη μορφή

$$\begin{aligned}
\rho(T_{r_1,r_2}) &= \rho[(A_2 + r_2 I)T_{r_1,r_2}(A_2 + r_2 I)^{-1}] \\
&\leq \left\|(A_1 - r_2 I)(A_1 + r_1 I)^{-1}\right\| \left\|(A_2 - r_1 I)(A_2 + r_2 I)^{-1}\right\| \\
&= \max_{\alpha_1 \leq \lambda_1 \leq \beta_1} \left|\frac{\lambda_1 - r_2}{\lambda_1 + r_1}\right| \max_{\alpha_2 \leq \lambda_2 \leq \beta_2} \left|\frac{\lambda_2 - r_1}{\lambda_2 + r_2}\right| \\
&= \max_{\substack{\alpha_1 \leq \lambda_1 \leq \beta_1 \\ \alpha_2 \leq \lambda_2 \leq \beta_2}} \left|\frac{(\lambda_1 - r_2)}{(\lambda_1 + r_1)} \frac{(\lambda_2 - r_1)}{(\lambda_2 + r_2)}\right|. \qquad (4.2.5)
\end{aligned}$$

Οι παράμετροι $r_1$ και $r_2$ επιλέγονται έτσι ώστε να ελαχιστοποιείται η συνάρτηση

$$f(\alpha_1, \alpha_2 \beta_1 \beta_2; r_1, r_2) = \max_{\substack{\alpha_1 \leq \lambda_1 \leq \beta_1 \\ \alpha_2 \leq \lambda_2 \leq \beta_2}} \left|\frac{(\lambda_1 - r_2)}{(\lambda_1 + r_1)} \cdot \frac{(\lambda_2 - r_1)}{(\lambda_2 + r_2)}\right|. \qquad (4.2.6)$$

Καταλήγουμε λοιπόν σε ένα minmax πρόβλημα η επίλυση του οποίου εν γένει παρουσιάζει αρκετές δυσκολίες. Για το παραπάνω πρόβλημα που έχουμε να λύσουμε, θα θεωρήσουμε κάποιους περιορισμούς, σε ό,τι αφορά τα διαστήματα ορισμού των ιδιοτιμών $\lambda_1$ και $\lambda_2$ καθώς επίσης και στη σύμπτωση ή όχι των παραμέτρων (μονοπαραμετρικό ή διπαραμετρικό σχήμα, αντίστοιχα).

Στη συνέχεια θα διακρίνουμε τις διάφορες περιπτώσεις που αντιστοιχούν στις παραπάνω θεωρήσεις.

1. Περίπτωση $1^{\eta}$ : $r = r_1 = r_2$ και ίδιο διάστημα ορισμού για τις ιδιοτιμές των πινάκων $A_1$ και $A_2$ ($\alpha \leq \lambda_1, \lambda_2 \leq \beta$).

   Σ′ αυτήν την περίπτωση η τιμή του $r$ που επιλύει το πρόβλημα (4.2.6) είναι

   $$r^* = \sqrt{\alpha\beta}, \tag{4.2.7}$$

   ενώ το βέλτιστο φράγμα για την φασματική ακτίνα δίνεται από την έκφραση

   $$f(\alpha, \beta; r) = \left(\frac{\sqrt{\beta} - \sqrt{\alpha}}{\sqrt{\beta} + \sqrt{\alpha}}\right)^2 \tag{4.2.8}$$

2. Περίπτωση $2^{\eta}$ : $r = r_1 = r_2$ και διαφορετικά διαστήματα ορισμού για τις ιδιοτιμές των $A_1$ και $A_2$.

   Σ′ αυτήν την περίπτωση υποθέτουμε επιπλέον ότι ικανοποιείται η συνθήκη

   $$\alpha_1\beta_1 \leq \alpha_2\beta_2, \tag{4.2.9}$$

   οπότε η τιμή του $r$ η οποία επιλύει το πρόβλημα (4.2.6) είναι

   $$r^* = \begin{cases} \sqrt{\alpha_1\beta_1} & \text{εάν } \alpha_1 \geq \alpha_2 \text{ ή } \alpha_1 \leq \alpha_2 \text{ και } \alpha_1\beta_2 \geq \alpha_2\beta_1, \\ \sqrt{\alpha_2\beta_2} & \text{εάν } \beta_1 \geq \beta_2 \text{ ή } \beta_1 \leq \beta_2 \text{ και } \alpha_1\beta_2 \leq \alpha_2\beta_1. \end{cases} \tag{4.2.10}$$

   Στην περίπτωση αυτή η φασματική ακτίνα δίνεται από τις εκφράσεις

   $$\begin{aligned} f(\alpha_1, \alpha_2, \beta_1, \beta_2; r^*) &= \left(\frac{r^* - \alpha_1}{r^* + \alpha_1}\right)\left(\frac{\beta_2 - r^*}{\beta_2 + r^*}\right) \qquad (4.2.11) \\ &= \begin{cases} \left(\frac{\sqrt{\beta_1} - \sqrt{\alpha_1}}{\sqrt{\beta_1} + \sqrt{\alpha_1}}\right)\left(\frac{\beta_2 - \sqrt{\alpha_1\beta_1}}{\beta_2 + \sqrt{\alpha_1\beta_1}}\right) & \text{εάν } r^* = \sqrt{\alpha_1\beta_1}, \\ -\left(\frac{\sqrt{\beta_2} - \sqrt{\alpha_2}}{\sqrt{\beta_2} + \sqrt{\alpha_2}}\right)\left(\frac{\alpha_1 - \sqrt{\alpha_1\beta_1}}{\alpha_1 + \sqrt{\alpha_1\beta_1}}\right) & \text{εάν } r^* = \sqrt{\alpha_2\beta_2}. \end{cases} \end{aligned}$$

   Θα πρέπει επιπλέον να τονίσουμε ότι

   $$f(\alpha_1, \alpha_2, \beta_1, \beta_2; r^*) \leq f(\alpha_1, \alpha_2, \beta_1, \beta_2; r), \tag{4.2.12}$$

   με γνήσια ανισότητα να λαμβάνεται στην περίπτωση όπου ικανοποιούνται οι συνθήκες

   $$\alpha_1 \leq \alpha_2, \ \beta_1 \leq \beta_2, \ \alpha_1\beta_2 = \alpha_2\beta_1. \tag{4.2.13}$$

   Τότε αποδεικνύεται ότι η ανισότητα (4.2.12) γίνεται γνήσια, εκτός βέβαια από την περίπτωση όπου $r = \sqrt{\alpha_1\beta_2}$ ή $r = \sqrt{\alpha_2\beta_1}$, για την οποία έχουμε ισότητα.

3. Περίπτωση $3^{η}$ : Διαφορετικά $r_1$ και $r_2$ και διαφορετικά διαστήματα για τις ιδιοτιμές των πινάκων $A_1$ και $A_2$.

Αυτή η περίπτωση είναι η γενικότερη στην κατηγορία των στάσιμων επαναληπτικών σχημάτων (δηλαδή, σχημάτων με σταθερές επαναληπτικές παραμέτρους). Η ιδέα της εύρεσης των βέλτιστων παραμέτρων $r_1$ και $r_2$ στηρίζεται σε μια τεχνική του W.B. Jordan για την επίλυση μίας περίπτωσης προβλήματος Zolotareff [65], που για πρώτη φορά εμφανίζεται στην ανάλυση Φίλτρων Επικοινωνίας και ειδικότερα στην ανάλυση του φίλτρου Gauer. Παρουσιάστηκε το 1963 σε μία εργασία του Wachspress [58]. Στην εργασία αυτή ο Jordan λύνει το πρόβλημα στη γενική περίπτωση των επαναληπτικών παραμέτρων $r_{m+1}$. Από την τεχνική αυτή μπορούμε να χρησιμοποιήσουμε τον αρχικό μετασχηματισμό έτσι ώστε και τα δυο διαστήματα ορισμού των ιδιοτιμών των δύο πινάκων $A_1$ και $A_2$ να μετασχηματίζονται σε ένα κοινό διάστημα και για τα δύο σύνολα ιδιοτιμών.

Για το σκοπό αυτό, έστω $\alpha_1 \leq \lambda_1 \leq \beta_1$ και $\alpha_2 \leq \lambda_2 \leq \beta_2$ τα διαστήματα των ιδιοτιμών των πινάκων $A_1$ και $A_2$, αντίστοιχα. Θεωρούμε τους μετασχηματισμούς

$$\lambda_1 = \frac{p + q\mu_1}{1 + s\mu_1}, \quad \lambda_2 = \frac{\widetilde{p} + \widetilde{q}\mu_2}{1 + \widetilde{s}\mu_2}, \tag{4.2.14}$$

όπου $\mu_1$ και $\mu_2$ είναι νέες μεταβλητές που αντιστοιχούν στις ιδιοτιμες $\lambda_1$ και $\lambda_2$, αντίστοιχα. Ο προσδιορισμός των παραμέτρων $p, q, s$ και $\widetilde{p}, \widetilde{q}, \widetilde{s}$ γίνεται με τέτοιον τρόπο ώστε να ικανοποιείται η συνθήκη

$$\left(\frac{\lambda_1 - r_2}{\lambda_1 + r_1}\right)\left(\frac{\lambda_2 - r_1}{\lambda_2 + r_2}\right) = \left(\frac{\mu_1 - \widetilde{r}_2}{\mu_1 + \widetilde{r}_1}\right)\left(\frac{\mu_2 - \widetilde{r}_1}{\mu_2 + \widetilde{r}_2}\right), \tag{4.2.15}$$

όπου $\widetilde{r}_1$ και $\widetilde{r}_2$ θεωρούνται οι νέες παράμετροι επιτάχυνσης μετά την εφαρμογή των μετασχηματισμών. Επιπλέον θα πρέπει να οι παράμετροι $\mu_1$ και $\mu_2$ να βρίσκονται σε ένα ορισθησόμενο κλειστό διάστημα, π.χ.

$$(0 <)\ \gamma \ \leq \ \mu_1, \mu_2 \ \leq \ \delta. \tag{4.2.16}$$

Με τη βοήθεια των μετασχηματισμών και τις συνθήκες (4.2.15) και (4.2.16) έχουμε ότι

$$\frac{\lambda_1 - r_2}{\lambda_1 + r_1} = \left(\frac{q - sr_2}{q + sr_1}\right)\frac{\mu_1 - \left(\frac{r_2 - p}{q - r_2 s}\right)}{\mu_1 + \left(\frac{r_1 + p}{q + r_1 s}\right)}, \tag{4.2.17}$$

$$\frac{\lambda_2 - r_1}{\lambda_2 + r_2} = \left(\frac{\widetilde{q} - \widetilde{s}r_1}{\widetilde{q} + \widetilde{s}r_2}\right) \frac{\mu_2 - \left(\frac{r_1 - \widetilde{p}}{\widetilde{q} - r_1\widetilde{s}}\right)}{\mu_2 + \left(\frac{r_1 + \widetilde{p}}{\widetilde{q} + r_2\widetilde{s}}\right)}. \tag{4.2.18}$$

Από τις παραπάνω συνθήκες παίρνουμε ότι

$$\widetilde{p} = -p, \quad \widetilde{q} = q, \quad \widetilde{s} = -s \tag{4.2.19}$$

και έτσι από τη σχέση (4.2.14) λαμβάνουμε τις παρακάτω εκφράσεις των ιδιοτιμών

$$\lambda_1 = \frac{p - q\mu_1}{1 + s\mu_1}, \qquad \lambda_2 = \frac{-p + q\mu_2}{1 - s\mu_2}. \tag{4.2.20}$$

Για τον υπολογισμό των παραμέτρων $p, q, s, \gamma, \delta$ απαιτούμε ότι εάν $\lambda_i = \alpha_i$ τότε $\mu_i = \gamma$, και εάν $\lambda_i = \beta_i$ τότε $\mu_i = \delta$, για $i = 1, 2$. Οι εκφράσεις για τα φράγματα των ιδιοτιμών δίνονται από τις σχέσεις

$$\begin{aligned} \alpha_1 &= \frac{p + q\gamma}{1 + s\gamma}, & \beta_1 &= \frac{p + q\delta}{1 + s\delta}, \\ \alpha_2 &= \frac{-p + q\gamma}{1 - s\gamma}, & \beta_2 &= \frac{-p + q\delta}{1 - s\delta}. \end{aligned} \tag{4.2.21}$$

Επιλύοντας τις παραπάνω εξισώσεις λαμβάνουμε τις εκφράσεις για τα $s, q, p$, οι οποίες είναι

$$s = \frac{(\beta_2 - \alpha_2) - (\beta_1 - \alpha_1)}{(\beta_1 + \beta_2)\delta - (\alpha_1 + \alpha_2)\gamma}, \tag{4.2.22}$$

$$q = \frac{(\beta_1 + \beta_2) + (\beta_1 - \beta_2)\delta s}{2\delta}, \tag{4.2.23}$$

$$p = \frac{(\beta_1 - \beta_2) + (\beta_1 + \beta_2)\delta s}{2}, \tag{4.2.24}$$

οπότε οι αντίστοιχες εκφράσεις για τις ιδιοτιμές $\lambda_1$ και $\lambda_2$ αλλά και για τις παραμέτρους $r_1$ και $r_2$ δίνονται από τις σχέσεις

$$\lambda_1 = \frac{p + (\delta q)\frac{\mu_1}{\delta}}{1 + (\delta s)\frac{\mu_1}{\delta}}, \qquad \lambda_2 = \frac{-p + (\delta q)\frac{\mu_2}{\delta}}{1 - (\delta s)\frac{\mu_2}{\delta}}, \tag{4.2.25}$$

$$r_1 = \frac{-p + (\delta q)\frac{\widetilde{r}_1}{\delta}}{1 - (\delta s)\frac{\widetilde{r}_1}{\delta}}, \qquad r_2 = \frac{p + (\delta q)\frac{\widetilde{r}_2}{\delta}}{1 + (\delta s)\frac{\widetilde{r}_2}{\delta}}. \tag{4.2.26}$$

Με χρήση απλού Απειροστικού Λογισμού μπορούμε να βρούμε ότι οι βέλτιστες παράμετροι $\widetilde{r}_1, \widetilde{r}_2$ για το minmax πρόβλημα δίνονται από τις κλασικές εκφράσεις

$$\widetilde{r}_1 = \widetilde{r}_2 = \sqrt{\gamma\delta}. \tag{4.2.27}$$

Επομένως οι βέλτιστες αρχικές παράμετροι για το πρόβλημα είναι οι παρακάτω

$$r_1 = \frac{-p + (\delta q)\sqrt{c}}{1 - (\delta s)\sqrt{c}}, \quad r_2 = \frac{p + (\delta q)\sqrt{c}}{1 + (\delta s)\sqrt{c}}, \tag{4.2.28}$$

ενώ η βέλτιστη φασματική ακτίνα δίνεται από την έκφραση

$$f(\alpha_1, \alpha_2, \beta_1, \beta_2, r_1, r_2) = \left(\frac{1 - \sqrt{c}}{1 + \sqrt{c}}\right)^2, \tag{4.2.29}$$

όπου

$$c = \frac{1}{1 + \xi + \sqrt{2\xi + \xi^2}}, \quad \xi = 2\frac{(\beta_2 - \alpha_2)(\beta_1 - \alpha_1)}{(\alpha_1 + \alpha_2)(\beta_1 + \beta_2)}. \tag{4.2.30}$$

*Παρατήρηση* 4.2.1. : Θα πρέπει και πάλι να τονίσουμε ότι οι βέλτιστες αυτές τιμές των παραμέτρων αφορούν στην επίλυση του minmax προβλήματος (4.2.6) και είναι οι βέλτιστες τιμές του φράγματος της φασματικής ακτίνας του επαναληπτικού πίνακα και όχι της φασματικής ακτίνας, τις οποίες θα προσδιορίσουμε σε επόμενο κεφάλαιο.

## 4.3 Ανάλυση Βασικών Σχημάτων με Μεταβλητές Παραμέτρους Επιτάχυνσης

### 4.3.1 Εισαγωγή

Στην παράγραφο αυτή θα ασχοληθούμε με την παρουσίαση και τη μελέτη του σχήματος των Peaceman-Rachford

$$\begin{aligned} (r_{m+1}I + A_1)u^{(m+\frac{1}{2})} &= (r_{m+1}I - A_2)u^{(m)} + b, \\ (r_{m+1}I + A_2)u^{(m+1)} &= (r_{m+1}I - A_1)u^{(m+\frac{1}{2})} + b, \end{aligned} \tag{4.3.1}$$

στη γενική περίπτωση όπου οι παράμετροι $r_{m+1}$ μεταβάλλονται από επανάληψη σε επανάληψη. Οι πρώτες προσπάθειες για την εύρεση τέτοιων παραμέτρων έγιναν από τους Peaceman και Rachford [45], το Wachspress [56] και τον Douglas

[18]. Στις τρεις αυτές προσπάθειες βρέθηκαν "καλές" παράμετροι επιτάχυνσης αλλά όχι οι βέλτιστες. Το 1962 οι Wachspress [57] και Gastinel [24], εργαζόμενοι ανεξάρτητα, έδωσαν τις βέλτιστες παραμέτρους στην ειδική περίπτωση όπου αυτές επαναλαμβάνονταν κυκλικά ανά $m-$επαναλήψεις με $m = 2^n$. Στην περίπτωση του γενικού $m$, όπως ήδη αναφέρθηκε, οι βέλτιστες παράμετροι δόθηκαν από τον W.B. Jordan και παρουσιάστηκαν σε εργασία του Wachspress το 1963 [58]. Στη συνέχεια αυτής της παραγράφου θα παρουσιάσουμε τη διαδικασία εύρεσης αυτών των παραμέτρων και για τις δύο περιπτώσεις που αναφέρθηκαν προηγουμένως.

Αρχίζουμε γράφοντας το επαναληπτικό σχήμα (4.3.1) στη μορφή μιας κλασικής επαναληπτικής μεθόδου. Συγκεκριμένα,

$$u^{(m+1)} = T_{r_{m+1}} u^{(m)} + c_{r_{m+1}}, \tag{4.3.2}$$

όπου

$$\begin{aligned} T_{r_{m+1}} &= (A_2 + r_{m+1}I)^{-1}(r_{m+1}I - A_1)(A_1 + r_{m+1}I)^{-1}(r_{m+1}I - A_2), \\ c_{r_{m+1}} &= (A_2 + r_{m+1}I)^{-1}\left((r_{m+1}I - A_1)(A_1 + r_{m+1}I)^{-1} + I\right) b. \end{aligned} \tag{4.3.3}$$

Έστω $e^{(m)} = u^{(m)} - u$ το σφάλμα στην $m-$επανάληψη της μεθόδου των Peaceman-Rachford. Τότε $e^{(m+1)} = T_{r_{m+1}} e^{(m)}$ και, γενικότερα, το σφάλμα $e^{(m)}$ μετά από $m$ επαναλήψεις θα δίνεται από την σχέση

$$e^{(m)} = \prod_{i=1}^{m} T_{r_i} e^{(0)}, \quad m > 1, \tag{4.3.4}$$

όπου $e^{(0)}$ το αρχικό σφάλμα της μεθόδου και

$$R_m := \prod_{i=1}^{m} T_{r_i} = T_{r_m} \cdot T_{r_{m-1}} \cdot \ldots \cdot T_{r_1}. \tag{4.3.5}$$

Θα πρέπει να τονίσουμε εδώ ότι για να προχωρήσουμε στη μελέτη της μεθόδου όλοι οι πίνακες που εμφανίζονται θα πρέπει να αντιμετατίθενται. Η απαίτηση αυτή ικανοποιείται σε όλες τις περιπτώσεις των προβλημάτων που εξετάζονται παρακάτω. Έχοντας υπόψη τα προαναφερθέντα και ορίζοντας με $\lambda_1$ και $\lambda_2$ τις ιδιοτιμές των πινάκων $A_1$ και $A_2$, αντίστοιχα, γνωρίζουμε ότι υπάρχει πλήρες σύστημα κοινών ιδιοδιανυσμάτων και για τους δυο πίνακες γεγονός που μας επιτρέπει να υπολογίσουμε τις ιδιοτιμές του επαναληπτικού πίνακα $T_{r_m}$ από την έκφραση

$$\lambda = \prod_{i=1}^{m} \frac{(r_i - \lambda_1)(r_i - \lambda_2)}{(r_i + \lambda_1)(r_i + \lambda_2)} \tag{4.3.6}$$

και την αντίστοιχη φασματική ακτίνα από τη

$$\rho(R_m) = \max_{\lambda_1,\lambda_2} \left| \prod_{i=1}^{m} \frac{(r_i - \lambda_1)(r_i - \lambda_2)}{(r_i + \lambda_1)(r_i + \lambda_2)} \right|. \tag{4.3.7}$$

Υποθέτουμε ότι οι ιδιοτιμές $\lambda_1$ και $\lambda_2$ βρίσκονται στα διαστήματα $\alpha_1 \leq \lambda_1 \leq \beta_1$ και $\alpha_2 \leq \lambda_2 \leq \beta_2$, αντίστοιχα. Σκοπός μας στην περίπτωση αυτή, όπως και στην περίπτωση του προβλήματος των σταθερών παραμέτρων, είναι η επίλυση του minmax προβλήματος

$$\min_{r_m} \rho(R_m). \tag{4.3.8}$$

Η περίπτωση που θα παρουσιάσουμε και, που συνήθως εξετάζεται, είναι αυτή όπου οι ιδιοτιμές $\lambda_1$ και $\lambda_2$ των πινάκων $A_1$ και $A_2$, αντίστοιχα, βρίσκονται στο ίδιο διάστημα $D$ (άλλωστε και την περίπτωση των διαφορετικών διαστημάτων ορισμού μπορούμε να την αναγάγουμε σ΄ αυτή με τη χρήση των προαναφερθέντων μετασχηματισμών), όπου

$$D \;:=\; \{\alpha = \min(\alpha_1, \alpha_2) \leq \lambda_1, \lambda_2 \leq \max(\beta_1, \beta_2) = \beta\}. \tag{4.3.9}$$

Τότε το minmax πρόβλημα που έχουμε να επιλύσουμε είναι το ακόλουθο

$$\min_{r_i,\, i=1,\ldots,m} \max_{\lambda_1,\lambda_2 \in D} \left| \prod_{i=1}^{m} \frac{(r_i - \lambda_1)(r_i - \lambda_2)}{(r_i + \lambda_1)(r_i + \lambda_2)} \right|. \tag{4.3.10}$$

Επειδή ισχύει ότι

$$\max_{\lambda_1,\lambda_2 \in D} \left| \prod_{i=1}^{m} \frac{(r_i - \lambda_1)(r_i - \lambda_2)}{(r_i + \lambda_1)(r_i + \lambda_2)} \right| \;\leq\; \left\{ \max_{\lambda \in D} \left| \prod_{i=1}^{m} \frac{(r_i - \lambda)}{(r_i + \lambda)} \right| \right\}^2,$$

το minmax πρόβλημα (4.3.10), που έχουμε να επιλύσουμε, ανάγεται σε minmax πρόβλημα μιας συνάρτησης που αποτελεί φράγμα αυτής που είχαμε. Έτσι το πρόβλημά μας μετατρέπεται στο ακόλουθο

$$\min_{r_m} \max_{\alpha \leq \lambda \leq \beta} \left| \prod_{i=1}^{m} \frac{r_i - \lambda}{r_i + \lambda} \right|. \tag{4.3.11}$$

Θέτοντας

$$Q_m(\lambda, p) = \prod_{i=1}^{m} \frac{r_i - \lambda}{r_i + \lambda}, \tag{4.3.12}$$

όπου $p \in \mathbb{R}^m$, $p = (r_1, r_2, \ldots, r_m)$, και ορίζοντας

$$H(p) = \max_{\lambda \in [\alpha,\beta]} |Q_m(\lambda, p)| \tag{4.3.13}$$

παρατηρούμε ότι $\rho(R_m) \leq H^2(p)$.

Για να επιλύσουμε το πρόβλημα (4.3.11) θα χρησιμοποιήσουμε τρία βασικά θεωρήματα της Θεωρίας Προσεγγίσεων (βλ. Achieser [2]).

**Θεώρημα 4.3.1.** *(Εναλλασσόμενη Ιδιότητα "de La Vallée Poussin"): Εάν η συνάρτηση $Q_m(\lambda, p)$, οριζόμενη στην (4.3.12), λαμβάνει τις τιμές $\gamma_1, -\gamma_2, \gamma_3, -\gamma_4, \ldots, (-1)^{m+1}\gamma_m$, με $\gamma_i > 0$, $i = 1, \ldots, m$, για γνησίως αύξουσα ακολουθία τιμών του $\lambda \in D$, και εάν η $Q$ είναι συνεχής στο $D$, τότε*

$$H \equiv \min_p H(p) = \min_p \max_{\lambda \in D} |Q_m(\lambda, p)| \tag{4.3.14}$$

*είναι κάτω φραγμένη από τη μικρότερη τιμή των $\gamma_i$, $i = 1, \ldots, m$.*

**Θεώρημα 4.3.2.** *: Έστω δυο αρχικά πολυώνυμα $P_m(\lambda)$ και $P_m(-\lambda)$ βαθμού $m$. Για κάθε θετική τιμή των $\lambda_i$ και κάθε $t \leq m$, ορίζουμε την συνάρτηση*

$$\phi(\lambda) = \lambda \prod_{i=1}^{t-1} (\lambda_i - \lambda)(\lambda_i + \lambda), \tag{4.3.15}$$

*τότε μπορεί να βρεθεί πολυώνυμο $P_s(\lambda)$ βαθμού $s$ τέτοιο ώστε*

$$\phi(\lambda) \equiv P_s(\lambda)P_m(-\lambda) - P_s(-\lambda)P_m(\lambda). \tag{4.3.16}$$

Το τρίτο θεώρημα αποδίδεται στον Chebyshev και είναι το ακόλουθο

**Θεώρημα 4.3.3.** *: Η συνάρτηση $Q_m(\lambda, p)$ η οποία έχει "ελάχιστη απόκλιση από το μηδέν" στο διάστημα $D := [\alpha, \beta]$ λαμβάνει το απόλυτο μέγιστό της $H$, $m + 1$ φορές στο $D$ με εναλλασσόμενα πρόσημα.*

### 4.3.2 Προσδιορισμός Βέλτιστων Παραμέτρων Σχήματος Peaceman-Rachford

Όπως έχουμε ήδη αναφέρει μέχρι το 1962 είχαν βρεθεί καλές παράμετροι επιτάχυνσης για τα σχήματα των ADI μεθόδων. Οι Wachspress [57] και Gastinel [24], έδωσαν βέλτιστες παραμέτρους στην ειδική περίπτωση $m = 2^n$. Ο W.B.

Jordan [57] έδωσε βέλτιστες παραμέτρους για κάθε τιμή του φυσικού $m$. Παρακάτω θα παρουσιάσουμε αυτές τις δύο εργασίες και κυρίως τη δεύτερη η οποία παρουσιάζει εξαιρετικό ενδιαφέρον εάν σκεφτεί κανείς ότι η λύση ήταν γνωστή από το 1877 από μια εργασία του Zolotareff [65] την οποία χρησιμοποίησαν αρχικά ο Gauer το 1933 για την κατασκευή ενός φίλτρου δικτύων επικοινωνίας και στην συνέχεια ο Jordan για την εύρεση των βέλτιστων παραμέτρων στο ADI σχήμα των Peaceman-Rachford το 1963. Θα πρέπει εδώ να τονίσουμε ότι δεν είναι τυχαίο ότι η λύση του Jordan δόθηκε ουσιαστικά οχτώ χρόνια μετά την πρώτη εμφάνιση του σχήματος των Peaceman-Rachford το 1955 και αυτό γιατί, όπως θα δούμε η λύση του Jordan στηρίζεται στη χρήση των Ελλειπτικών Συναρτήσεων του Jacobi οι οποίες δεν είναι ιδιαίτερα γνωστές στο ευρύ κοινό.

### 4.3.3 Βέλτιστες Παράμετροι για $m = 2^n$

Η βασική ιδέα για την εύρεση των βέλτιστων παραμέτρων στηρίζεται στο παρακάτω λήμμα

**Λήμμα 4.3.4.** *: Εάν $p_i \in p$ ($p$ είναι η ακολουθία των βέλτιστων παραμέτρων) τότε $\frac{\alpha\beta}{p_i} \in p$.*

Μία επίσης ιδιότητα του minmax προβλήματος είναι ότι η συνάρτηση $Q_m(\lambda, p)$ έχει την ιδιότητα $Q_m(\lambda, p) = Q_m(\sigma\lambda, \sigma p)$, και επομένως το διάστημα $[\alpha, \beta]$ κανονικοποιείται στο διάστημα $[\frac{\alpha}{\beta}, 1]$. Εάν λοιπόν οι παράμετροι γι΄ αυτό το διάστημα είναι $p_i$, τότε οι παράμετροι για το $[\alpha, \beta]$ είναι $\beta p_i$. Θεωρούμε τώρα το γινόμενο των όρων

$$\left[\frac{p_i - \lambda}{p_i + \lambda}\right]\left[\frac{\frac{\alpha\beta}{p_i} - \lambda}{\frac{\alpha\beta}{p_i} + \lambda}\right] = \frac{\alpha\beta + \lambda^2 - \left(p_i + \frac{\alpha\beta}{p_i}\right)\lambda}{\alpha\beta + \lambda^2 + \left(p_i + \frac{\alpha\beta}{p_i}\right)\lambda}. \tag{4.3.17}$$

Διαιρώντας τον αριθμητή και τον παρονομαστή του δεξιού μέλους με $2\lambda$ λαμβάνουμε την ισοδύναμη με τις παραπάνω έκφραση

$$\frac{\left[\frac{\frac{\alpha\beta}{\lambda}+\lambda}{2}\right] - \left[\frac{\frac{\alpha\beta}{p_i}+p_i}{2}\right]}{\left[\frac{\frac{\alpha\beta}{\lambda}+\lambda}{2}\right] + \left[\frac{\frac{\alpha\beta}{p_i}+p_i}{2}\right]}. \tag{4.3.18}$$

Ορίζοντας τώρα

$$\begin{aligned} \lambda^{(1)} &\equiv \frac{\left(\frac{\alpha\beta}{\lambda}+\lambda\right)}{2}, \\ p_i^{(1)} &\equiv \frac{\left(\frac{\alpha\beta}{p_i}+p_i\right)}{2},\ i=1,2,\ldots,\frac{m}{2}, \end{aligned} \tag{4.3.19}$$

παρατηρούμε ότι για $\lambda \in [\alpha,\beta]$ έχουμε $\lambda^{(1)} \in \left[\sqrt{\alpha\beta}, \frac{\alpha+\beta}{2}\right]$. Οπότε εάν οι βέλτιστες παράμετροι μπορούν να υπολογιστούν για το πρόβλημα $Q_{\frac{m}{2}}(\lambda^{(1)}, p^{(1)})$, τότε μπορούμε να βρούμε τις βέλτιστες παραμέτρους για το αρχικό πρόβλημα από τις εξισώσεις (4.3.19).

Στην περίπτωση λοιπόν όπου το $m$ είναι άρτιο τότε το φάσμα των ιδιοτιμών μπορεί να διχοτομηθεί έτσι ώστε η τάξη του προβλήματος να υποβιβάζεται στο $\frac{m}{2}$. Ακολουθώντας αυτήν την τεχνική καταφέρνουμε τελικά να λύσουμε το πρόβλημα στην περίπτωση του $m = 1$ για το οποίο υπάρχουν αναλυτικές εκφράσεις. Στη συνέχεια, με τη βοήθεια τύπων, όπως οι (4.3.19), καταφέρνουμε να βρούμε τις βέλτιστες παραμέτρους για το αρχικό πρόβλημα που δεν είναι τίποτα περισσότερο από την τετραγωνική ρίζα των άκρων του διαστήματος που θα προκύψει μετά $n$ διαδοχικές διχοτομήσεις. Για την αναλυτική περιγραφή της διαδιακασίας ο αναγνώστης παραπέμπεται στις εργασίες των Wachspress [57] και Gastinel [24].

### 4.3.4 Βέλτιστες Παράμετροι για Γενικό $m$

Η βασική ιδέα της εύρεσης των βέλτιστων παραμέτρων στην περίπτωση όπου το $m$ είναι ένας οποισδήποτε φυσικός αριθμός στηρίζεται στην εργασία του Zolotareff [65] του 1877 αλλά και στο W.B. Jordan [58], ο οποίος χρησιμοποιώντας τη λύση του Zolotareff υπολόγισε τις βέλτιστες παραμέτρους των *ADI* μεθόδων για το σχήμα των Peaceman-Rachford. Τα εργαλεία που χρησιμοποιήθηκαν ήταν η "εναλλακτική ιδιότητα του Chebyshev" (βλ. Achieser [2]) και οι βασικές αρχές των Ελλειπτικών Συναρτήσεων του Jacobi (βλ. Abramowitz-Stegun [1]). Στην συνέχεια θα παρουσιάσουμε την τεχνική του W.B. Jordan στην επίλυση του (4.3.10) που προκύπτει από το σχήμα των Peaceman-Rachford.

Έστω ότι οι ιδιοτιμές των πινάκων $A_1$ και $A_2$ ανήκουν στα διαστήματα $\alpha_1 \leq \lambda_1 \leq \beta_1$ και $\alpha_2 \leq \lambda_2 \leq \beta_2$, αντίστοιχα. Θεωρούμε τους μετασχηματισμούς

$$\lambda_1 = \frac{\mu_1 - h}{f - g\mu_1}, \quad \lambda_2 = \frac{\mu_2 + h}{f + g\mu_2}, \tag{4.3.20}$$

όπου οι σταθερές $f, g, h$ θα προσδιοριστούν με τέτοιο τρόπο ώστε τα διαστήματα των φασμάτων των ιδιοτιμών κάθε πίνακα να μετασχηματιστούν σε ένα κοινό διάστημα με νέες μεταβλητές $\alpha \leq \mu_1, \mu_2 \leq \beta$. Τότε το αρχικό πρόβλημα γίνεται

$$Q_m = \prod_{i=1}^{m} \left( \frac{\mu_1 - \widetilde{r}_{iA_2}}{\mu_1 + \widetilde{r}_{iA_1}} \right) \left( \frac{\mu_2 - \widetilde{r}_{iA_1}}{\mu_2 + \widetilde{r}_{iA_2}} \right), \tag{4.3.21}$$

όπου

$$\widetilde{r}_{iA_2} = \frac{f r_{iA_2} + h}{1 + g r_{iA_2}}, \quad \widetilde{r}_{iA_1} = \frac{f r_{iA_1} - h}{1 - g r_{iA_1}}. \tag{4.3.22}$$

Όπως αναφέραμε και προηγουμένως η επιλογή των παραμέτρων $f, g, h$ θα γίνει με τέτοιον τρόπο ώστε αν $\lambda_i = \alpha_i$ και $\lambda_i = \beta_i$, τότε $\mu_i = k'$ και $\mu_i = 1$, για $i = 1, 2$, αντίστοιχα. Για να επιτευχθεί αυτό θα πρέπει να επιλέξουμε το $k'$ ως τη ρίζα τις εξίσωσης

$$k'^2 - 2(1 + \xi)k' + 1 = 0, \tag{4.3.23}$$

όπου

$$\xi = \frac{2(\beta_1 - \alpha_1)(\beta_2 - \alpha_2)}{(\alpha_1 + \alpha_2)(\beta_1 + \beta_2)}. \tag{4.3.24}$$

Αφού το $\xi > 0$ τότε οι ρίζες του παραπάνω τριωνύμου δευτέρου βαθμού θα είναι πραγματικές και θετικές. Θεωρούμε ότι $k'$ είναι η μικρότερη εξ αυτών οπότε

$$k' = \frac{1}{1 + \xi + \sqrt{\xi(\xi + 2)}}, \quad 0 < k' < 1. \tag{4.3.25}$$

Έτσι οι συντελεστές $f, g, h$ των μετασχηματισμών ισούνται, αντίστοιχα, με

$$f = \frac{2 + g(\beta_1 - \beta_2)}{\beta_1 + \beta_2}, \tag{4.3.26}$$

$$g = 2\frac{k'(\beta_1 + \beta_2) - (\alpha_1 + \alpha_2)}{(\alpha_1 + \alpha_2)(\beta_1 - \beta_2) + k'(\beta_1 + \beta_2)(\alpha_2 - \alpha_1)}, \tag{4.3.27}$$

$$h = \frac{k'(\alpha_2 - a l_1 + 2\alpha_1\alpha_2 g)}{\alpha_1 + \alpha_2}. \tag{4.3.28}$$

Έχοντας τώρα τις μεταβλητές $\mu_1, \mu_2$ να ανήκουν στο ίδιο διάστημα, τα σύνολα των παραμέτρων $\widetilde{r}_{iA_1}$ και $\widetilde{r}_{iA_2}$ που επιλύουν το minmax πρόβλημα μπορούν να θεωρηθούν ότι ταυτίζονται και ότι οι παράμετροι

$$\widetilde{r}_{iA_1} = \widetilde{r}_{iA_2} = r_i, \ i = 1, \ldots, m. \tag{4.3.29}$$

Έτσι η συνάρτηση $Q_m$ γράφεται ως εξής

$$Q_m = P_m(\mu_1)P_m(\mu_2), \quad \text{όπου} \quad P_m(\mu_1) = \prod_{i=1}^{m} \frac{\mu_1 - r_i}{\mu_1 + r_i} \tag{4.3.30}$$

Για την εύρεση των παραμέτρων $\mu_i$, $i = 1, \dots m$, κάνουμε χρήση των βασικών Ελλειπτικών Συναρτήσεων του Jacobi, θέτοντας $\mu_1 = \mathrm{dn}(Kz)$ και έχοντας $\mathrm{mod}K = \sqrt{1-k'^2}$, και την παράμετρο $\tau = \frac{ik'}{K}$, η σχέση (4.3.30) μετασχηματίζεται σε

$$P(z) = \prod_{i=1}^{m} \frac{\mathrm{dn}(Kz) - r_i}{\mathrm{dn}(Kz) + r_i}. \tag{4.3.31}$$

Θεωρούμε τη συνάρτηση

$$P'(z) = \frac{\mathrm{dn}(mK_1 z) - \sqrt{k_1'}}{\mathrm{dn}(mK_1 z) + \sqrt{k_1'}}, \quad \text{με} \quad \mathrm{mod}K_1, \quad \text{και παράμετρο} \quad \tau_1 = m\tau, \tag{4.3.32}$$

που έχει τις παρακάτω ιδιότητες

1. Η μέγιστη τιμή της είναι $P'_{\max}(z) = \frac{1-\sqrt{k_1'}}{1+\sqrt{k_1'}}$, όταν $z = 0$.
2. Η ελάχιστη τιμή της είναι $P'_{\min} = -P'_{\max}$, όταν $z = \frac{1}{m}$.
3. Έχει πραγματική περίοδο $\frac{2}{m}$.
4. Η συνάρτηση $P'$ ικανοποιεί τη minmax συνθήκη του Chebyshev.
5. Έχει φανταστική περίοδο $\frac{4\tau_1}{m} = 4\tau$.
6. $P'(\tau) = 1$.

Εάν λοιπόν οι συναρτήσεις $P$ και $P'$ είναι ίσες τότε το πρόβλημα θα έχει λυθεί. Από το θεώρημα όμως του Liouville (βλ. Achieser [2]), δύο συναρτήσεις διαφέρουν κατά σταθερό παράγοντα, και άρα μπορούν να καταστούν ίσες με κανονικοποίηση, εάν οι πόλοι και οι ρίζες τους συμπίπτουν. Οι ρίζες της συνάρτησης $P'$ είναι $z = \frac{2i-1}{2m}$ ενώ οι πόλοι της είναι $z = 2\tau + \frac{2i-1}{2m}$. Για να ταυτίζονται αυτές με τις ρίζες και τους πόλους της συνάρτησης $R$ θα πρέπει να επιλεγούν οι παράμετροι $r_i$ έτσι ώστε

$$r_i = \mathrm{dn}\left(\frac{(2i-1)K}{2m}\right), \quad \mathrm{mod}k = \sqrt{1-k'^2}. \tag{4.3.33}$$

Έτσι από τους προηγούμενους μετασχηματισμούς έχουμε ότι οι παράμετροι του αρχικού προβλήματος δίνονται από τις εκφράσεις

$$r_{iA_1} = \frac{r_i - h}{f - gr_i}, \;\; r_{iA_2} = \frac{r_i + h}{f + gr_i}, \;\; i = 1, \ldots m, \tag{4.3.34}$$

και άρα η φασματική ακτίνα του επαναληπτικού πίνακα θα ισούται με

$$\rho(T_m) = \left( \frac{1 - \sqrt{k_1'}}{1 - \sqrt{k_1'}} \right)^2 . \tag{4.3.35}$$

Σημειώνεται οι παραπάνω εκφράσεις για τις βέλτιστες παραμέτρους αλλά και κατ΄ επέκταση για τη φασματική ακτίνα του επαναληπτικού πίνακα μπορούν να δοθούν με εκφράσεις που σχετίζονται με τις Συναρτήσεις Θήτα του Jacobi.

# 4.4 Παρεκβαλλόμενες Πεπλεγμένες Μέθοδοι Εναλλασσόμενων Διευθύνσεων (Extrapolated(E)ADI)

## 4.4.1 Εισαγωγή

Στην παράγραφο αυτή θα παρουσιάσουμε τις Παρεκβαλλόμενες (Extrapolated) Μεθόδους Πεπλεγμένων Εναλλασσόμενων Διευθύνσεων ή αλλιώς, συντομογραφικά, EADI. Οι μέθοδοι αυτές εισήχθησαν σχεδόν ταυτόχρονα και ανεξάρτητα από τους Guittet [27] και Hadjidimos [28] το 1967. Οι EADI μέθοδοι αποτελούν ένα συνδυασμό των ADI μεθόδων που παρουσιάσαμε στην προηγούμενη παράγραφο και της τεχνικής της παρεκβολής (extrapolation). Η βασική ιδέα ξεκίνησε από το σχήμα των Douglas-Rachford για την επίλυση της εξίσωσης Poisson με Dirichlet συνοριακές συνθήκες γενικευμένο στις $p-$διαστάσεις ορισμένης στον $p-$διάστατο υπερκύβο με το ίδιο βήμα διακριτοποίησης σε κάθε διεύθυνση και με μια μικρή επιπλέον τροποποίηση

$$\begin{aligned} (I + rA_1)u^{(m+\frac{1}{q})} &= (I + r(A_1 - A))u^{(m)} + rb \\ (I + rA_i)u^{(m+\frac{i+1}{q})} &= rA_i u^{(m)} + u^{(m+\frac{i}{q})}, \; i = 1, 2, \ldots, q-1. \end{aligned} \tag{4.4.1}$$

Θα πρέπει να τονίσουμε εδώ ότι στο σχήμα (4.4.1) η παράμετρος επιτάχυνσης $r$ αντιστοιχεί στο $\frac{1}{r}$ του αρχικού σχήματος των Douglas-Rachford. Ο επαναληπ-

τικός πίνακας του σχήματος αυτού έχει τη μορφή

$$T_{DR} = I - (\mathbf{1})rA\prod_{i=1}^{q}(I + rA_i)^{-1}. \tag{4.4.2}$$

Ένα επίσης ενδιαφέρον σχήμα το οποίο εισήγαγε ο Douglas και το οποίο αποτελεί μία παραλλαγή του σχήματος (4.4.1) παρουσιάζεται παρακάτω

$$\begin{aligned}(I + rA_1)u^{m+\frac{1}{q}} &= (I + r(A_1 - 2A))u^m + 2rb \\ (I + rA_i)u^{m+\frac{1}{q}} &= rA_iu^m + u^{m+\frac{1}{q}},\ i = 2, 3, \ldots, q,\end{aligned} \tag{4.4.3}$$

με επαναληπτικό πίνακα

$$T_{DR} = I - (\mathbf{2})rA\prod_{i=1}^{q}(I + rA_i)^{-1}. \tag{4.4.4}$$

Παρατηρούμε ότι τα δύο αυτά σχήματα ουσιαστικά διαφέρουν στη τιμή μίας σταθεράς όπου στο πρώτο αυτή έχει την τιμή $\mathbf{1}$ ενώ στο δεύτερο την τιμή $\mathbf{2}$. Γενικεύοντας την παραπάνω έκφραση για τον επαναληπτικό πίνακα λαμβάνουμε την έκφραση

$$T = I - \omega rA\prod_{i=1}^{q}(I + rA_i)^{-1}. \tag{4.4.5}$$

Με βάση την παραπάνω έκφραση ο Guittet έδωσε ένα επαναληπτικό σχήμα που αποτελεί μία παραλλαγή του σχήματος των Douglas-Rachford (4.1.4) και του σχήματος του Douglas (4.4.1), στο οποίο παρουσιάζεται η δεύτερη εξίσωση απαλλαγμένη από την προηγούμενη προσέγγιση. Το γενικό σχήμα του Guittet έχει την παρακάτω μορφή

$$\begin{aligned}(I + rA_1)u^{(m+\frac{1}{q})} &= \left(\prod_{i=1}^{q}(I + rA_i) - \omega rA\right)u^{(m)} + \omega rb \\ (I + rA_i)u^{(m+\frac{i+1}{q})} &= u^{(m+\frac{i}{q})}, \quad i = 1, 2, \ldots, q-1,\end{aligned} \tag{4.4.6}$$

με επαναληπτικό πίνακα

$$T_\omega = I - \omega rA\prod_{i=1}^{q}(I + rA_i)^{-1}, \tag{4.4.7}$$

ή σε μία ισοδύναμη μορφή

$$T_\omega = I - \omega A \prod_{i=1}^{q} (I + rA_i)^{-1}, \tag{4.4.8}$$

όπου η παράμετρος $\omega$ ισούται με το $\omega r$ της προηγούμενης έκφρασης (4.4.9). Η παραπάνω μορφή του πίνακα δεν εκφράζει τίποτα άλλο από τον επαναληπτικό πίνακα της Extrapolation ADI μεθόδου με το βασικό επαναληπτικό πίνακα να δίνεται από την έκφραση

$$T = A \prod_{i=1}^{q} (I + rA_i)^{-1}. \tag{4.4.9}$$

Ορίζουμε ως $\sigma(A_i)$ το φάσμα ιδιοτιμών $\lambda_i$ των πινάκων $A_i$, $i = 1, \ldots, q$, και $\alpha \leq \lambda_i \leq \beta$, $i = 1, \ldots, q$, να αντιστοιχούν στις ιδιοτιμές των πινάκων $A_i$. Έχοντας υπόψη ότι $A = \sum_{i=1}^{q} A_i$, ενώ οι πίνακες $A_i$ αντιμετατίθενται, μπορούμε να θεωρήσουμε το ίδιο σύστημα ιδιοδιανυσμάτων για όλους τους εμπλεκόμενους πίνακες με τις ιδιοτιμές του πίνακα $T_\omega$ να δίνονται από την έκφραση

$$\lambda(T_\omega) = 1 - \omega \frac{\sum_{i=1}^{q} \lambda_i}{\prod_{i=1}^{q} (1 + r\lambda_i)}. \tag{4.4.10}$$

### 4.4.2 Εύρεση Βέλτιστων Παραμέτρων

Η βασική ιδέα στην οποία στηρίζεται η διαδικασία που θα ακολουθήσουμε για την εύρεση των βέλτιστων παραμέτρων θα είναι παρόμοια με αυτήν που ακολουθήσαμε και προηγουμένως. Για το σκοπό αυτό θεωρούμε τη διακριτή συνάρτηση

$$f_i(\lambda_1, \lambda_2, \ldots, \lambda_q) = \frac{\sum_{i=1}^{q} \lambda_i}{\prod_{i=1}^{q} (1 + r\lambda_i)}, \tag{4.4.11}$$

όπου $\lambda_i$, $i = 1, 2, \ldots, q$, είναι οι ιδιοτιμές των πινάκων $A_i$, $i = 1, 2, \ldots, q$. Οι πίνακες αυτοί είναι συμμετρικοί και θετικά ορισμένοι κι έτσι οι ιδιοτιμές $\lambda_i$, $i = 1, 2, \ldots, q$, είναι πραγματικοί θετικοί αριθμοί. Ορίζουμε το συνεχές ανάλογο της συνάρτησης (4.4.11) με την

$$f := f(x_1, x_2, \ldots, x_q) = \frac{\sum_{i=1}^{q} x_i}{\prod_{i=1}^{q} (1 + rx_i)}, \tag{4.4.12}$$

όπου $x_i \in [\alpha, \beta],\ i = 1, 2, \ldots, q$. Έχουμε ότι

$$\inf_{\substack{x_i \in [\alpha,\beta] \\ i=1,2,\ldots,q}} (1 - \omega f) \leq (1 - \omega f_i) \leq \sup_{\substack{x_i \in [\alpha,\beta] \\ i=1,2,\ldots,q}} (1 - \omega f) \tag{4.4.13}$$

κι επομένως λαμβάνουμε τη σχέση

$$\rho(T) \leq \sup_{\substack{x_i \in [\alpha,\beta] \\ i=1,2,\ldots,q}} |1 - \omega f| \tag{4.4.14}$$

Υπολογίζοντας τη μερική παράγωγο της συνάρτησης $f$ ως προς τη μεταβλητή $x_i$ έχουμε ότι ο αριθμητής του λόγου

$$\frac{1}{f}\frac{\partial f}{\partial x_i} = \frac{1 - \sum_{j=1, j\neq i}^{q} x_j}{(1 + r x_i)\sum_{i=1}^{q} x_i} \tag{4.4.15}$$

είναι ανεξάρτητος από τη μεταβλητή $x_i$ ως προς την οποία παραγωγίσαμε την $f$. Το αποτέλεσμα αυτό έχει ως συνέπεια ότι τα ακρότατα της $f$ λαμβάνονται στα άκρα του διαστήματος ορισμού $[a\, \beta]$, αφού η $f$ είναι μονότονη ως προς $x_i$, $i = 1, 2, \ldots, q$. Επομένως τα ακρότατα της συνάρτησης $f$ θα δίνονται από τη διακριτή συνάρτηση

$$g(p) = \frac{p\alpha r + (q - p)\beta r}{(1 + \alpha r)^p (1 + \beta r)^{q-p}},\ p \in \{0, 1, 2, \ldots, q\}. \tag{4.4.16}$$

Συνεπώς η μέγιστη και η ελάχιστη τιμή της συνάρτησης $f$ ταυτίζονται με τη μέγιστη και την ελάχιστη τιμή της συνάρτησης $g$. Έχουμε λοιπόν τις σχέσεις

$$\sup_{\substack{x_i \in [\alpha,\beta] \\ i=1,2,\ldots,q}} f = \sup_{\substack{x_i \in \{\alpha,\beta\} \\ i=1,2,\ldots,q}} f = \sup_{p \in \{0,1,\ldots,q\}} g(p) = G \tag{4.4.17}$$

$$\inf_{\substack{x_i \in [\alpha,\beta] \\ i=1,2,\ldots,q}} f = \inf_{\substack{x_i \in \{\alpha,\beta\} \\ i=1,2,\ldots,q}} f = \inf_{p \in \{0,1,\ldots,q\}} g(p) = g, \tag{4.4.18}$$

όπου $G$ και $g$ αντιστοιχούν στη μέγιστη και στην ελάχιστη τιμή της $g(p)$ και κατ΄ επέκταση της $f$. Με χρήση βασικών στοιχείων Απειροστικού Λογισμού μπορούμε να αποδείξουμε ότι η συνάρτηση $g(p)$ λαμβάνει την ελάχιστη τιμή όταν $p = q$ ή όταν $p = 0$. Επίσης από την θεωρία της extrapolation τεχνικής η τιμή για το $\omega$ που ελαχιστοποιεί την έκφραση της φασματικής ακτίνας λαμβάνεται από την έκφραση

$$\omega^* = \frac{2}{g + G}, \tag{4.4.19}$$

και για τη φασματική ακτίνα έχουμε ότι ισχύει η ανισότητα

$$\rho(T) \leq \{|1-\omega g|, |1-\omega G|\}. \tag{4.4.20}$$

Βασική μας επιδίωξη είναι να βρεθούν οι εμπλεκόμενες παράμετροι ώστε να ελαχιστοποιηθεί η έκφραση στο δεξιό μέλος της ανισότητας (4.4.20). Αντικαθιστώντας λοιπόν το $\omega = \omega^*$ στη (4.4.20) έχουμε τη σχέση

$$\rho(T) \leq \frac{1-\frac{g}{G}}{1+\frac{g}{G}}. \tag{4.4.21}$$

Για την ελαχιστοποίηση αυτής της ποσότητας θα πρέπει να μεγιστοποιήσουμε το λόγο $\frac{g}{G}$ ως προς $r$. Μετά από κάποιες, όχι και τόσο απλές, πράξεις βρίσκουμε ότι το βέλτιστο $r = r^*$ έχει την τιμή

$$r^* = \frac{\beta^{\frac{1}{q}} - \alpha^{\frac{1}{q}}}{\alpha^{\frac{1}{q}}\beta - \alpha\beta^{\frac{1}{q}}}. \tag{4.4.22}$$

Η βέλτιστη τιμή για το $\omega$ είναι τότε

$$\omega^* = \frac{2(1-\nu)^q}{\nu^{\frac{q-p-1}{q}}(q\nu^{\frac{p}{q}} + p\nu + q - p)(1-\nu^{\frac{q-1}{q}})^{q-1}(1-\nu^{\frac{1}{q}})}, \tag{4.4.23}$$

ενώ η βέλτιστη φασματική ακτίνα έχει την τιμή

$$\rho^* = \frac{1-\left(\frac{q\nu^{\frac{p}{q}}}{p\nu} + q - p\right)}{1-\left(\frac{q\nu^{\frac{p}{q}}}{p\nu} + q - p\right)}. \tag{4.4.24}$$

Σημειώνεται ότι σε κάθε μία από τις παραπάνω σχέσεις έχει τεθεί $\nu = \frac{\alpha}{\beta}$.

Μετά από την εργασία του Guittet, η οποία αναφέρεται σε στατικό μονοπαραμετρικό (μιας παραμέτρου επιτάχυνσης) επαναληπτικό σχήμα, δημοσιεύτηκαν μια σειρά από εργασίες στις οποίες βρέθηκαν βέλτιστες τιμές για τις παραμέτρους επιτάχυνσης στις περιπτώσεις των διπαραμετρικών σχημάτων, βασισμένων στο EADI σχήμα του Guittet, στις περιπτώσεις προβλημάτων μοντέλων με διακριτοποίηση πεπερασμένων διαφορών σε πλέγμα των 5−σημείων. Σε επόμενες παραγράφους θα παρουσιάσουμε την εύρεση των βέλτιστων παραμέτρων στη γενικότερη περίπτωση των 9− σημείων για την περίπτωση ενός μονοπαραμετρικού σχήματος βασισμένου στο σχήμα του Guittet αλλά και ενός διπαραμετρικού ανάλογου σχήματος.

# Κεφάλαιο 5

# Προρρυθμισμένη Μέθοδος Συζυγών Κλίσεων(PCG)

## 5.1 Μέθοδος Συζυγών Κλίσεων

Στην περίπτωση κατά την οποία ο πίνακας που προκύπτει από τη διακριτοποίηση του αρχικού προβλήματος είναι Ερμιτιανός και θετικά ορισμένος τότε η επαναληπτική μέθοδος που χρησιμοποιείται συνήθως για την επίλυση του συστήματος $Ax = b$ είναι η μέθοδος των Συζυγών Κλίσεων γνωστή ως Conjugate Gradient (CG). Η μέθοδος αυτή ανήκει στην κατηγορία των μεθόδων ελαχιστοποίησης και αποτελεί βελτίωση της Απλής Επαναληπτικής Μεθόδου, δηλαδή των γνωστών μας μεθόδων, όπως η μέθοδος Jacobi, η μέθοδος Gauss-Seidel (GS) και η SOR, καθώς και γενίκευση της μεθόδου της Απότομης Καθόδου (βλ. [26]).

Γενικεύοντας λοιπόν τη μέθοδο της Απότομης Καθόδου, χρησιμοποιούμε μία νέα ακολουθία διαδοχικών προσεγγίσεων της λύσης

$$x_{k+1} = x_k + \alpha_k p_k, \tag{5.1.1}$$

όπου τα διανύσματα $p_k$ είναι, καταρχάς, γενικά τυχαίες διευθύνσεις. Το σφάλμα, στην περίπτωση αυτή, δίνεται από τη σχέση

$$e_{k+1} = e_k - \alpha_k p_k.$$

Ο συντελεστής $\alpha_k$ επιλέγεται έτσι ώστε το $e_{k+1}$ να είναι $A-$ορθογώνιο στο $p_k$.

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

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

Σ' αυτήν 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

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

Η μέθοδος, που εφαρμόζει τα εκτεθέντα, καλείται Μέθοδος "Συζυγών Κλίσεων" (Conjugate Gradient (CG) ) και η υλοποίησή της παρουσιάζεται στην Greenbaum (βλ. [26]).

**Θεώρημα 5.1.1.** *: 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 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 ελάχιστη $A-$νόρμα.*

Για το σφάλμα της μεθόδου των Συζυγών Κλίσεων παραθέτουμε το παρακάτω θεώρημα

**Θεώρημα 5.1.2.** *: Έστω ότι $e_k$ είναι το σφάλμα στην $k$ επανάληψη της μεθόδου CG εφαρμοσμένης σε Ερμιτιανό και θετικά ορισμένο σύστημα. Τότε*

$$\frac{\|e_k\|_A}{\|e_0\|_A} \leq 2\left[\left(\frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}\right)^k + \left(\frac{\sqrt{\kappa}+1}{\sqrt{\kappa}-1}\right)^k\right]^{-1} \leq 2\left(\frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}\right)^k, \qquad (5.1.2)$$

*όπου $\kappa = \frac{\lambda_{\max}}{\lambda_{\min}}$ είναι ο δείκτης κατάστασης 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, στην περίπτωση αυτή ισχύει το παρακάτω θεώρημα για το σφάλμα

**Θεώρημα 5.1.3.** *: 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*

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

*Παρατήρηση* 5.1.1. : Από τη σχέση 5.1.2 παρατηρούμε ότι το η $A-$νόρμα του σφάλματος της μεθόδου των Συζυγών Κλίσεων φράσσετε από έναν όρο που είναι αύξουσα συνάρτηση του δείκτη κατάστασης. Επομένως, για να ελαχιστοποιήσουμε το φράγμα αυτό θα πρέπει κατά κάποιον τρόπο, να ελαχιστοποιήσουμε το δείκτη κατάστασης του πίνακα-συντελεστή του συστήματός μας. Γνωρίζουμε, βέβαια, ότι στην περίπτωση όπου ο πίνακας $A$ είναι Ερμιτιανός και θετικά ορισμένος ο δείκτης κατάστασης δίνεται από την έκφραση $\kappa = \frac{\lambda_{\max}}{\lambda_{\min}}$. Έτσι θα πρέπει να ελαχιστοποιήσουμε το λόγο της μέγιστης προς την ελάχιστη ιδιοτιμή του πίνακα–συντελεστή του συστήματος. Όπως διαπιστώνουμε κι από τα φράγματα που βρέθηκαν θα πρέπει να προσπαθούμε να περιορίζουμε το εύρος της διασποράς των ιδιοτιμών έτσι ώστε οι περισσότερες εξ αυτών να συγκεντρώνονται σε ένα διάστημα και μόνο. Οι υπόλοιπες, ελάχιστες σε πλήθος, ιδιοτιμές θα πρέπει να βρίσκονται εκτός του προαναφερθέντος διαστήματος και τότε δε θα επηρεάζουν σοβαρά την ταχύτητα σύγκλισης της μεθόδου.

Η καλύτερη τεχνική για την πραγματοποίηση των όσων επισημάνθηκαν στην παραπάνω παρατήρηση είναι η χρήση ενός προρρυθμιστή (preconditioner) για την μέθοδο των Συζυγών Κλίσεων.

## 5.2 Προρρυθμισμένη Μέθοδος Συζυγών Κλίσεων (PCG)

Η χρήση προρρυθμιστή για τις διάφορες επαναληπτικές μεθόδους μπορεί κάποιος να πει ότι είναι το άλφα και το ωμέγα για τη βελτίωση της ταχύτητας σύγκλισης

των μεθόδων αυτών. Η ανάγκη για τη χρήση του προρρυθμιστή προέρχεται, όπως είδαμε και στην προηγούμενη παράγραφο, από το γεγονός ότι η ταχύτητα σύγκλισης της επαναληπτικής μεθόδου, και ειδικότερα της μεθόδου Συζυγών Κλίσεων (CG), εξαρτάται από τις ιδιότητες του φάσματος των ιδιοτιμών και κατ΄ επέκταση από τη φασματική ακτίνα εφόσον αναφερόμαστε στην ασυμπτωτική συμπεριφορά των μεθόδων. Θα ήταν, λοιπόν, επιθυμητό να θεωρούσαμε ένα μετασχηματισμό του αρχικού συστήματος,

$$Ax = b, \quad A \in \mathbb{C}^{n\times n},\, b \in \mathbb{C}^n \tag{5.2.1}$$

σε ένα ισοδύναμό του σε ό,τι αφορά τη λύση αλλά με καλύτερες ιδιότητες του φάσματος του νέου πίνακα συντελεστών. Το μετασχηματισμένο σύστημα θα έχει τη μορφή

$$M^{-1}Ax = M^{-1}b,\ A, M \in \mathbb{C}^{n\times n},\, b \in \mathbb{C}^n. \tag{5.2.2}$$

Η επιλογή ενός προρρυθμιστή βασίζεται στην ιδέα ότι ο πίνακας $M$ θα αποτελεί μία προσέγγιση του αρχικού πίνακα $A$. Η επιλογή του $M$ ως τον πίνακα $A$ δεν είναι σίγουρα η καλύτερη επιλογή αφού η εύρεση του $M^{-1}$ θα στοίχιζε από την άποψη του κόστους σε πράξεις ανά επανάληψη, περισσότερο από ό,τι η επίλυση του αρχικού συστήματος με, π.χ., Απαλοιφή Gauss. Λόγω του αυξημένου κόστους της εύρεσης του $M^{-1}$ οδηγούμαστε λοιπόν στην επιλογή του πίνακα $M$ έτσι ώστε αυτός να είναι αντιστρέψιμος με τον εξής περιορισμό. Ένα σύστημα της μορφής

$$Mx = c \tag{5.2.3}$$

να λύνεται με πολύ μικρότερο κόστος σε σχέση με αυτό που έχει ένα σύστημα με πίνακα συντελεστών $A$.

Στην περίπτωση όπου ο αρχικός πίνακας είναι Ερμιτιανός και θετικά ορισμένος, ο προρρυθμιστής που συνήθως χρησιμοποιούμε παρουσιάζει τις ίδιες με αυτές που προαναφέρθηκαν ιδιότητες. Επιπλέον στις περιπτώσεις των μεθόδων, όπως η CG, χρησιμοποιούμε, ως επί το πλείστον και "δεξιό" και "αριστερό" προρρυθμιστή, οπότε το προρρυθμισμένο σύστημα έχει την μορφή

$$M_1 A M_2^{-1} y = M_1^{-1} b, \quad y = M_2 x,\ A, M_1, M_2 \in \mathbb{C}^{n\times n},\, b \in \mathbb{C}^n. \tag{5.2.4}$$

Στην περίπτωση όπου ο πινακάς μας είναι Ερμιτιανός, τότε ο προρρυθμιστής πίνακας $M$ μπορεί να ορθοκανονικοποιηθεί με αποτέλεσμα να ικανοποιείται η σχέση $M^{-1} = M^H$. Επομένως και ο προρρυθμισμένος πίνακας $M_1^H A M_2^H$ είναι

με τη σειρά του Ερμιτιανός και θετικά ορισμένος. Στην επαναληπτική διαδικασία με τη μέθοδο των Συζυγών Κλίσεων (CG) το επιπλέον ουσιαστικά κόστος είναι η επίλυση ενός συστήματος με πίνακα συντελεστών αγνώστων τον προρρυθμιστή πίνακα $M$. Εδώ πρέπει να τονιστεί ότι, από πρώτης όψεως, το γεγονός ότι ένα σύστημα $Mx = c$ θα λύνεται σε κάθε επανάληψη δίνει επιπλέον κόστος για την επαναληπτική μέθοδο. Στην πραγματικότητα αυτό δεν ισχύει γιατί δε χρειάζεται σε κάθε επανάληψη να παραγοντοποιούμε τον πίνακα $M$, αφού παραμένει ο ίδιος σε κάθε επαναληπτικό βήμα. Έτσι η παραγοντοποίηση γίνεται μόνο στην αρχή κι αυτό που επαναλαμβάνεται ανά βήμα είναι οι προς τα πίσω αντικαταστάσεις με διαφορετικά δεξιά μέλη κάθε φορά. Οι πράξεις αυτές δεν έχουν πολύ κόστος και τα οφέλη που έχουμε ουσιαστικά ελαχιστοποιούν το επιπλέον κόστος. Στη συνέχεια θα δώσουμε μία σειρά από βασικούς προρρυθμιστές για τη μέθοδο Συζυγών Κλίσεων.

### 5.2.1 Προρρυθμιστής Jacobi

Είναι ο απλούστερος όλων τον προρρυθμιστών αφού επιλέγουμε το $M$ να είναι ο διαγώνιος πίνακας με στοιχεία τα διαγώνια στοιχεία του πίνακα $A$. Στην περίπτωση όπου ο πίνακας $A$ είναι Ερμιτιανός και θετικά ορισμένος μπορούμε να παραγοντοποιήσουμε των διαγώνιο πίνακα του προρρυθμιστή Jacobi σε $M = M^{\frac{1}{2}} M^{\frac{1}{2}}$, με τον πίνακα $M^{\frac{1}{2}}$, να είναι η καλούμενη "τετραγωνική ρίζα" του πίνακα $M$. Μ΄ αυτόν τον τρόπο μπορούμε στη μέθοδο των Συζυγών Κλίσεων (CG) να χρησιμοποιήσουμε δεξιό και αριστερό προρρυθμιστή. Το κόστος με τη χρήση του προρρυθμιστή αυξάνεται ελάχιστα αλλά και τα οφέλη από την χρησιμοποίηση αυτού του προρρυθμιστή, αν υπάρχουν, είναι επίσης ελάχιστες. Βελτιώνοντας τον παραπάνω προρρυθμιστή εισάγουμε τη μορφή του "μπλοκ" (block) προρρυθμιστή Jacobi. Η βασική ιδέα είναι η διάσπαση του πίνακα $A$ σε "μπλοκ" υποπίνακες. Είναι προφανές, λοιπόν, ότι η διάσπαση αυτή δεν είναι μοναδική. Τα βασικά κριτήρια που καθορίζουν τη διάσπαση αυτή είναι η φύση του προβλήματος, όπως σε προβλήματα Μερικών Διαφορικών Εξισώσεων, όπου η φύση των προβλημάτων προβλέπει ένα συγκεκριμένο διαχωρισμό κατά γραμμές, στην περίπτωση των διδιάστατων ($2D$) και σε επίπεδα στην περίπτωση των τριδιάστατων ($3D$) προβλημάτων. Επίσης, η μέθοδος αλλά και η τεχνική (παράλληλη επεξεργασία), που επιλέγουμε για την επίλυση του προβλήματος, πολλές φορές μας επιβάλλουν ένα συγκεκριμένο διαχωρισμό. Τέλος, στην περίπτωση των μεθόδων, όπως η CG, θα πρέπει η επιλογή του προρρυθμιστή να είναι τέτοια ώστε το ισοδύναμο προρρυθμισμένο σύστημα να εξακολουθεί να έχει πίνακα συντελεστών Ερμιτιανό και θετικά ορισμένο, ώστε

να μπορεί να εφαρμοστεί η μέθοδος CG και στο προρρυθμισμένο σύστημα. Ο "μπλοκ" Jacobi προρρυθμιστής έχει κι αυτός μικρό κόστος και εν γένει δίνει καλύτερα αποτελέσματα σε σχέση με τον απλό προρρυθμιστή Jacobi.

### 5.2.2 Προρρυθμιστής SSOR

Ένας άλλος προρρυθμιστής που λόγω της δομής του χρησιμοποιείται ευρύτατα είναι ο SSOR Προρρυθμιστής. Ο προρρυθμιστής αυτός προέρχεται από τον αρχικό πίνακα $A$ μέσω του διαχωρισμού $A = D - L - L^T$. Ο προρρυθμιστής πίνακας που προκύπτει από τη διάσπαση αυτή είναι

$$M = (D - L)D^{-1}(D - L)^T, \tag{5.2.5}$$

ή στη μορφή

$$M(\omega) = \frac{1}{2-\omega}\left(\frac{1}{\omega}D - L\right)\left(\frac{1}{\omega}D\right)^{-1}\left(\frac{1}{\omega}D - L\right)^T. \tag{5.2.6}$$

Βέβαια, έδω θα πρέπει να τονίσουμε ότι παρά το γεγονός ότι ο προρρυθμιστής για τη χρησιμοποιηθησόμενη βέλτιστη τιμή του $\omega$ ($\omega_{opt}$) μπορεί θεωρητικά να δώσει πολύ καλά αποτελέσματα $\kappa(M_{\omega_{opt}}^{-1}A) = \mathcal{O}(\sqrt{\kappa(A)})$. Όμως το κόστος για την εύρεση του $\omega_{opt}$ είναι απαγορευτικό στη χρήση ενός τέτοιου προρρυθμιστή, γι' αυτό και χρησιμοποιείται η τιμή $\omega = 1$.

### 5.2.3 Προρρυθμιστές Ατελούς Παραγοντοποίησης

Μία από τις πιο γνωστές αλλά και ευρείας χρήσης κατηγορίες προρρυθμιστών είναι αυτή που βασίζεται στην ιδέα της τεχνικής της Ατελούς Παραγοντοποίησης (Incomplete Factorization) του πίνακα $A$. Η τεχνική αυτή έχει σαν βασική ιδέα την προσπάθεια να βρεθεί μια καλή προσέγγιση για τους παράγοντες $L$ και $U$ στην $LU$ παραγοντοποίηση του πίνακα $A\,(= LU)$. Αυτό που προσπαθούμε να επιτύχουμε είναι κατά την απαλοιφή του Gauss στον πίνακα $A$, να χρησιμοποιούνται μη μηδενικά στοιχεία στους προσεγγιστικούς παράγοντες $L$ και $U$ μόνο στις θέσεις όπου ο πίνακας $A$ έχει μη μηδενικά στοιχεία. Αυτή η κατηγορία της Incomplete Factorization τεχνικής χαρακτηρίζεται ως ILU(0) και είναι η πιο διαδεδομένη. Σε άλλες περιπτώσεις είναι δυνατόν να επιτρέπονται και άλλα μη μηδενικά στοιχεία στους προσεγγιστικούς παραγοντες. Για παράδειγμα σε ένα ή δύο στοιχεία συμμετρικά στις υπερ- και υπο-διαγωνίους σε σχέση με τα μη μηδενικά στοιχεία του αρχικού πίνακα, οπότε τις μεθόδους

αυτές τις χαρακτηρίζουμε ως ILU(1) και ILU(2), αντίστοιχα. Με την τεχνική αυτή παράγουμε μία προσέγγιση των πινάκων $L$ και $U$ της κλασικής απαλοιφής Gauss. Στην περίπτωση όπου ο πίνακας $A$ είναι Ερμιτιανός και θετικά ορισμένος τότε αντί για την ILU χρησιμοποιούμε Incomplete Cholesky (IC), μία τεχνική που βρίσκει μια προσέγγιση για τον πίνακα $L$ και κατά συνέπεια και του γινομένου $LL^T$ της παραγοντοποίησης Cholesky. Η διαδικασία με την οποία μπορούμε να επιλύσουμε το σύστημα $Mx = c$, που εμφανίζεται στον αλγόριθμο της CG, μπορεί να είναι η κλασική με προς τα μπρος και προς τα πίσω αντικαταστάσεις με πίνακες συντελεστών τους $L$ και $L^T$, αντίστοιχα. Επίσης μία άλλη θεώρηση του πίνακα $M$ σε $M = (D - L)D^{-1}(D - L^T)$ θα έδινε το εξής σύστημα ισοδύναμων συστημάτων

$$(D - L)z = c, \qquad (D - L^T)x = Dz. \tag{5.2.7}$$

Και τα δύο αυτά συστήματα είναι εύκολο να λυθούν με τη διαδικασία της προς τα μπρος και προς τα πίσω αντικαταστάσης, αντίστοιχα. Περισσότερα σε ό,τι αφορά την ύπαρξη αλλά και την τεχνική της ILU-IC παραγοντοποίησης υπάρχουν σε διάφορα βιβλία αναφοράς, όπως, π.χ., στο βιβλίο της Greenbaum [26] αλλά και στις εργασίες των Meijerink και van der Vorst [43].

# Κεφάλαιο 6

# Βέλτιστοι EADI Προρρυθμιστές Μεθόδου Συζυγών Κλίσεων

## 6.1 Βέλτιστοι Μονοπαραμετρικοί EADI Προρρυθμιστές

### 6.1.1 Εισαγωγή

Στο προηγούμενο κεφάλαιο μελετήσαμε τη μέθοδο Συζυγών Κλίσεων και μια σειρά από τους κλασικούς προρρυθμιστές της μεθόδου αυτής. Στο παρόν κεφάλαιο θα παρουσιάσουμε μία νέα κατηγορία προρρυθμιστών της μεθόδου Συζυγών Κλίσεων (CG).

Στη διάρκεια των τελευταίων χρόνων μια σειρά από εργασίες έρχονται να επαναφέρουν τις ADI μεθόδους αυτή την φορά ως προρρυθμιστές των μεθόδων τύπου CG αλλά και ως "ομαλοποιητές" των multigrid μεθόδων (βλ., [50], [23], [51], [31], [52], [61], [62], [54], [32], [9], [10], [11], κ.τ.λ.). Οι παραπάνω αναφορές στάθηκαν η αφετηρία της προσπάθειάς μας εφαρμογής των EADI μεθόδων ως προρρυθμιστών για τις μεθόδους τύπου Συζυγών Κλίσεων. Αρχικά, θα δείξουμε ότι, σε ό,τι αφορά το πρόβλημα μοντέλο της εξίσωσης Poisson με Dirichlet συνοριακές συνθήκες στο μοναδιαίο τετράγωνο με ομοιόμορφη διακριτοποίηση του τελεστή με πεπερασμένες διαφορές 5−σημείων και ίδιο βήμα διακριτοποίησης, έχουμε ότι το προρρυθμισμένο σύστημα, με προρρυθμιστή αυτόν του βέλτιστου EADI σχήματος. Το τελευταίο σύστημα έχει

αισθητά μικρότερο δείκτη κατάστασης σε σχέση με αυτόν αντίστοιχων συστημάτων με κλασικούς προρρυθμιστές, όπως τον απλό Jacobi και τον "μπλοκ" Jacobi. Επίσης θα πρέπει να τονίσουμε ότι ο δείκτης κατάστασης με βέλτιστο EADI προρρυθμιστή είναι καλύτερος και από τον SSOR προρρυθμιστή. Στη συνέχεια, για την εύρεση του EADI προρρυθμιστή στην περίπτωση της διακριτοποιήσης $9-$σημείων, χρειάστηκε να βρεθούν οι βέλτιστες παραμέτροι και επομένως το αντίστοιχο βέλτιστο EADI σχήμα.

### 6.1.2 Σύγκριση Δεικτών Κατάστασης CG και PCG Μεθόδων

Αρχίζουμε με τη σύγκριση των δεικτών κατάστασης της μεθόδου CG με τους κλασικούς προρρυθμιστές Jacobi, "μπλοκ" Jacobi και SSOR αλλά και με την CG μέθοδο, όπου χρησιμοποιούμε το "βέλτιστο" EADI προρρυθμιστή.

Για το σκοπό αυτό θεωρούμε την εξίσωση Poisson στις δύο διαστάσεις με Dirichlet συνοριακές συνθήκες στο μοναδιαίο τετράγωνο με τετράγωνο πλέγμα βήματος $h = \frac{1}{n+1}$. Διακριτοποιούμε το συνεχή τελεστή με το διακριτό σχήμα των $5-$σημείων σε κάθε εσωτερικό κόμβο του πλέγματος και καταλήγουμε σε ένα $n^2 \times n^2$ πραγματικό συμμετρικό και θετικά ορισμένο σύστημα

$$Ax = c. \tag{6.1.1}$$

Στην (6.1.1) ο πίνακας $A$ έχει τη μορφή

$$A \;=\; I_n \otimes T + T \otimes I_n, \tag{6.1.2}$$

όπου $\otimes$ συμβολίζει το τανυστικό γινόμενο δύο πινάκων (βλ. Halmos [36]), $T = \text{tridiag}(-1, 2, -1) \in \mathbb{R}^{n\times n}$, και $I_n$ είναι ο μοναδιαίος πίνακας τάξης $n$. Στην περίπτωση αυτή και λόγω της ιδιότητας του πίνακα $A$ να είναι συμμετρικός και θετικά ορισμένος η κατάλληλη μέθοδος για την επίλυση του συστήματος (6.1.1) είναι η μέθοδος Συζυγών Κλίσεων (CG).

Είναι γνωστό από προηγούμενο κεφάλαιο ότι η $A-$νόρμα του σφάλματος στην $k-$επανάληψη σε σχέση με την $A-$νόρμα του αρχικού σφάλματος ικανοποιούν τη σχέση

$$\left\|e^{(k)}\right\|_A \;\leq\; \frac{1}{T_k\left(\frac{\kappa(A)+1}{\kappa(A)-1}\right)} \left\|e^{(0)}\right\|_A, \tag{6.1.3}$$

όπου $T_k(\cdot)$ είναι το πολυώνυμο του Chebyshev βαθμού $k$ και $\kappa(A)$ συμβολίζει το δείκτη κατάστασης του πίνακα $A$, που αντιστοιχεί στη φασματική νόρμα.

Αντικαθιστώντας την έκφραση για το πολυώνυμο Chebyshev λαμβάνουμε τη σχέση

$$\frac{\|e_k\|_A}{\|e_0\|_A} \leq 2\left[\left(\frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}\right)^k + \left(\frac{\sqrt{\kappa}+1}{\sqrt{\kappa}-1}\right)^k\right]^{-1} \leq 2\left(\frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}\right)^k. \qquad (6.1.4)$$

Στο προηγούμενο κεφάλαιο είδαμε τα θεωρήματα στα οποία αναφέρονται οι παραπάνω σχέσεις ενώ οι αντίστοιχες αποδείξεις βρίσκονται στο βιβλίο αναφοράς της Greenbaum [26]. Αφού ο $A$ είναι πραγματικός συμμετρικός και θετικά ορισμένος τότε ο δείκτης κατάστασης δίνεται από τη σχέση

$$\kappa(A) \;=\; \frac{\lambda_{\max}(A)}{\lambda_{\min}(A)}, \qquad (6.1.5)$$

όπου $\lambda_{\max}$ και $\lambda_{\min}$ είναι η μέγιστη και η ελάχιστη ιδιοτιμή του πίνακα $A$. Στην περίπτωση του προβλήματος μοντέλου που εμείς εξετάζουμε οι ιδιοτιμές του πίνακα $A$ δίνονται από της εκφράσεις

$$\lambda_i = 4\sin^2\left(\frac{i\pi}{2(n+1)}\right) + 4\sin^2\left(\frac{j\pi}{2(n+1)}\right),\ i,j = 1,2,\ldots,n. \qquad (6.1.6)$$

Έτσι, η μέγιστη και η ελάχιστη ιδιοτιμές δίνονται από τις εκφράσεις $\lambda_{\max} = 8\cos^2\left(\frac{\pi}{2(n+1)}\right)$ και $\lambda_{\min} = 8\sin^2\left(\frac{\pi}{2(n+1)}\right)$. Επομένως, από την (6.1.5), έχουμε ότι

$$\kappa(A) \;=\; \frac{8\cos^2\left(\frac{\pi}{2(n+1)}\right)}{8\sin^2\left(\frac{\pi}{2(n+1)}\right)} \;=\; \cot^2\left(\frac{\pi}{2(n+1)}\right). \qquad (6.1.7)$$

Για να βελτιώσουμε την ταχύτητα σύγκλισης της μεθόδου των Συζυγών Κλίσεων, για το διακριτό πρόβλημα της εξίσωσης Poisson με Dirichlet συνοριακές συνθήκες, κάνουμε χρήση προρρυθμιστή πίνακα στο αρχικό σύστημα. Ο προρρυθμιστής πίνακας που θα χρησιμοποιήσουμε θα πρέπει να είναι κι αυτός πραγματικός συμμετρικός και θετικά ορισμένος. Έτσι θεωρούμε πίνακα $M$, πραγματικό συμμετρικό και θετικά ορισμένο, γεγονός που μας επιτρέπει να ορίσουμε τη (θετική) τετραγωνική του ρίζα $M^{\frac{1}{2}}$, που θα έχει ακριβώς τις ίδιες ιδιότητες με τον $M$. Ο πίνακας $M^{\frac{1}{2}}$ χρησιμοποιείται ως δεξιός και αριστερός προρρυθμιστής για το αρχικό σύστημα το οποίο θα πάρει τελικά την ισοδύναμη μορφή

$$\widetilde{A}\widetilde{x} \;=\; \widetilde{b}, \qquad (6.1.8)$$

όπου $\widetilde{A} = M^{-\frac{1}{2}} A M^{-\frac{1}{2}}$, $\widetilde{x} = M^{\frac{1}{2}} x$ και $\widetilde{b} = M^{-\frac{1}{2}} b$. Ο πίνακας $\widetilde{A}$ είναι προφανώς πραγματικός συμμετρικός και θετικά ορισμένος, επομένως μπορούμε στο νέο σύστημα να εφαρμόσουμε τη μέθοδο Συζυγών Κλίσεων. Η μόνη διαφορά στον αλγόριθμό της, σε σχέση με αυτόν της απλής CG του αρχικού συστήματος, θα έγκειται στην επίλυση ενός επιπλέον γραμμικού συστήματος με πίνακα συντελεστών τον πίνακα $M$. Η πολυπλοκότητα της επίλυσης ενός τέτοιου συστήματος είναι αισθητά μικρότερη από αυτήν του αρχικού.

Παρατηρούμε τώρα ότι οι πίνακες $M^{-\frac{1}{2}} A M^{-\frac{1}{2}}$ και $M^{-1}A$ είναι όμοιοι και επομένως έχουν το ίδιο φάσμα ιδιοτιμών. Έτσι ο νέος δείκτης κατάστασης του προρρυθμισμένου συστήματος έχει τη μορφή

$$\kappa(\widetilde{A}) = \frac{\lambda_{\max}(M^{-1}A)}{\lambda_{\min}(M^{-1}A)}. \tag{6.1.9}$$

Από τη σχέση (6.1.9) παρατηρούμε ότι για να μπορέσουμε να εξετάσουμε την ταχύτητα σύγκλισης του προρρυθμισμένου συστήματος είναι ανάγκη να υπολογίσουμε τη μέγιστη και την ελάχιστη ιδιοτιμή του πίνακα $M^{-1}A$. Γενικά, αυτό το πρόβλημα δεν είναι εύκολο στις περισσότερες περιπτώσεις, κι αυτό έχει ως αποτέλεσμα να καταφεύγουμε στην εύρεση φραγμάτων των ακραίων ιδιοτιμών. Στην περίπτωση όμως του προβλήματος που μελετάμε είμαστε σε θέση να κάνουμε αυτόν τον υπολογισμό.

Στην περίπτωση του σημειακού προρρυθμιστή Jacobi μπορούμε με προφανή τρόπο να αποδείξουμε ότι

$$\kappa(\widetilde{A}) = \kappa(A) \tag{6.1.10}$$

μιας και ο πίνακας $M = cI$ και επομένως ο δείκτης κατάστασης παραμένει αναλλοίωτος. Αντίθετα, στην περίπτωση του "μπλοκ" προρρυθμιστή Jacobi αποδεικνύουμε ό,τι ακολουθεί. Θεωρούμε ως πίνακα $M$, τον πίνακα $M = D$, όπου $D$ ο "μπλοκ" διαγώνιος πίνακας με στοιχεία τα κεντρικά "μπλοκς" του πίνακα $A$. Έτσι για τον πίνακα $M$ έχουμε ότι

$$M \;=\; M_1 \;=\; I_n \otimes \text{tridiag}(-1, 4, -1) \;=\; I_n \otimes (2I_n + T). \tag{6.1.11}$$

Για την προρρυθμισμένη μέθοδο Συζυγών Κλίσεων (CG) με προρρυθμιστή τον "μπλοκ" Jacobi (Block Jacobi-CG) πίνακα εργαζόματε ως εξής. Οι ιδιοτιμές του $M_1^{-1}A$ δίνονται από τις εκφράσεις

$$\lambda_{k,l}(M_1^{-1}A) \;=\; \frac{2\sin^2\left(\frac{k\pi}{2(n+1)}\right) + 2\sin^2\left(\frac{l\pi}{2(n+1)}\right)}{1 + 2\sin^2\left(\frac{l\pi}{2(n+1)}\right)}, \quad k, l = 1, \ldots, n. \tag{6.1.12}$$

Επομένως οι ακραίες ιδιοτιμές είναι οι

$$\lambda_{\max}(M_1^{-1}A) \;=\; \frac{2}{1+2\sin^2\left(\frac{\pi}{2(n+1)}\right)} \;\text{και}\; \lambda_{\min}(M_1^{-1}A) \;=\; \frac{4\sin^2\left(\frac{\pi}{2(n+1)}\right)}{1+2\sin^2\left(\frac{\pi}{2(n+1)}\right)}. \tag{6.1.13}$$

Συνεπώς, ο δείκτης κατάστασης του προρρυθμισμένου συστήματος έχει την παρακάτω μορφή

$$\kappa(\widetilde{A}) \;=\; \frac{\lambda_{\max}(M_1^{-1}A)}{\lambda_{\min}(M_1^{-1}A)} \;=\; \frac{1}{2\sin^2\left(\frac{\pi}{2(n+1)}\right)}. \tag{6.1.14}$$

*Παρατήρηση* 6.1.1. : Παρατηρούμε ότι ασυμπτωτικά, ο δείκτης κατάστασης της Block Jacobi-CG είναι ο μισός σε σχέση με αυτόν της απλής μεθόδου CG.

Στην περίπτωση όπου ο προρρυθμιστής που χρησιμοποιούμε είναι ο SSOR προρρυθμιστής μπορεί να αποδειχτεί (βλ. Axelsson-Barker [4]) ότι

$$\kappa(\widetilde{A}) = \sqrt{\kappa(A)} = \cot\left(\frac{\pi}{2(n+1)}\right). \tag{6.1.15}$$

Εξετάζοντας τώρα και την περίπτωση των προρρυθμιστών, και συγκεκριμένα αυτόν των Peaceman-Rachford των βέλτιστων ADI μεθόδων, έχουμε ότι ο πίνακας προρρυθμιστής $M$ έχει τη μορφή

$$M \;=\; M_2 \;=\; (rI_n\otimes I_n + I_n\otimes T)(rI_n\otimes I_n + T\otimes I_n) \;=\; (rI_n+T)\otimes(rI_n+T), \tag{6.1.16}$$

με $r = 2\sin\left(\frac{\pi}{n+1}\right)$ (βλ. [55], [64]). Οι ακραίες ιδιοτιμές του $M_2^{-1}A$ δίνονται από τις εκφράσεις

$$\lambda_{\max}(M_2^{-1}A) \;=\; \frac{1}{4\sin\left(\frac{\pi}{2(n+1)}\right)\cos\left(\frac{\pi}{2(n+1)}\right)\left(\sin\left(\frac{\pi}{2(n+1)}\right)+\cos\left(\frac{\pi}{2(n+1)}\right)\right)^2}, \tag{6.1.17}$$

$$\lambda_{\min}(M_2^{-1}A) \;=\; \frac{1}{2\left(\sin\left(\frac{\pi}{2(n+1)}\right)+\cos\left(\frac{\pi}{2(n+1)}\right)\right)^2}. \tag{6.1.18}$$

Έτσι

$$\kappa(\widetilde{A}) \;=\; \frac{\lambda_{\max}(M_2^{-1}A)}{\lambda_{\min}(M_2^{-1}A)} \;=\; \frac{1}{2\sin\left(\frac{\pi}{2(n+1)}\right)\cos\left(\frac{\pi}{2(n+1)}\right)}. \tag{6.1.19}$$

*Παρατήρηση* 6.1.2. : Παρατηρούμε λοιπόν ότι ασυμπτωτικά, από όλους τους κλασικούς προρρυθμιστές που χρησιμοποιούνται για τη μέθοδο Συζυγών Κλίσεων ο ADI προρρυθμιστής δίνει δείκτη κατάστασης του γραμμικού συστήματος κατά μια τάξη μικρότερο από αυτόν της προρρυθμισμένης σημειακής Jacobi, της προρρυθμισμένης "μπλοκ" Jacobi και βέβαια από αυτόν της μεθόδου Συζυγών Κλίσεων χωρίς προρρύθμιση. Ακόμη έχει τον μισό δείκτη κατάστασης σε σχέση με αυτόν της προρρυθμισμένης SSOR. Τέλος, μπορούμε να παρατηρήσουμε ότι στην περίπτωση που χρησιμοποιούμε την τεχνική της παρεκβολής (extrapolation) για τους προρρυθμιστές ο δείκτης κατάστασης παραμένει αμετάβλητος σε σχέση με αυτόν χωρίς παρεκβολή.

### 6.1.3 Βέλτιστοι Μονοπαραμετρικοί EADI Προρρυθμιστές

Αρχίζουμε με την εξίσωση Poisson σε ορθογώνιο χωρίο

$$\Omega := \left\{(x,y) \in \mathbb{R}^2 | 0 < x < c,\, 0 < y < d\right\}$$

$$-a(x,y)u_{xx}(x,y) - b(x,y)u_{yy}(x,y) = f(x,y),\ f \in C^2 \tag{6.1.20}$$

με Dirichlet συνοριακές συνθήκες στο σύνορο $\partial\Omega$ του $\Omega$, $u(x,y) = \gamma(x,y)$. Θεωρούμε ότι οι συναρτήσεις $a := a(x,y)$ και $b := b(x,y)$ είναι συνεχείς, θετικές και κάτω φραγμένες. Στην περίπτωση που θα μελετήσουμε θα τις θεωρήσουμε απλά θετικές σταθερές. Στη συνέχεια επιθέτουμε ένα ομοιόμορφο διακριτό πλέγμα στο $\overline{\Omega} := \Omega \cup \partial\Omega$, με βήμα διακριτοποίησης $h_1$ και $h_2$ στη $x-$ και στην $y-$ διεύθυνση, αντίστοιχα. Για την απλοποίηση των υπολογισμών θα θεωρούμε εφεξής ότι $c = d = 1$. Σε κάθε εσωτερικό κόμβο του πλέγματος η εξίσωση (6.1.20) προσεγγίζεται, όπως ήδη έχουμε δεί σε προηγούμενο κεφάλαιο, από το σχήμα διαφορών

$$\begin{aligned} & \sqrt{\frac{a}{b}}\frac{h_2}{h_1}(-u_{i-1,j} + 2u_{ij} - u_{i+1,j}) + \sqrt{\frac{b}{a}}\frac{h_1}{h_2}(-u_{i,j-1} + 2u_{ij} - u_{i,j+1}) \\ - \quad & \theta\,[4u_{ij} - 2(u_{i-1,j} + u_{i+1,j} + u_{i,j-1} + u_{i,j+1}) \\ + \quad & u_{i-1,j-1} + u_{i+1,j-1} + u_{i-1,j+1} + u_{i+1,j+1}] = \frac{h_1 h_2}{\sqrt{ab}}(f_{ij} + \phi_{ij}), \end{aligned} \tag{6.1.21}$$

όπου οι παράμετροι $\theta$ και $\phi$ λαμβάνουν τις τιμές

$$(\theta,\phi) \;=\; \begin{cases} (0,0), \\ (\theta^*,\phi^*) = \left(\frac{1}{12}(\sqrt{\frac{a}{b}}\frac{h_2}{h_1} + \sqrt{\frac{b}{a}}\frac{h_1}{h_2}), \frac{1}{12}(ah_1^2 f_{xx} + bh_2^2 f_{yy})\right). \end{cases} \tag{6.1.22}$$

Θα σημειώσουμε για άλλη μία φορά ότι εάν $\theta = 0$, (6.1.21), τότε έχουμε το σχήμα των $5-$σημείων, ενώ εάν $\theta = \theta^*$, τότε έχουμε ένα σχήμα $9-$σημείων. Επίσης θα πρέπει να τονίσουμε ότι όπως αναφέραμε σε προηγούμενο κεφάλαιο ο διακριτός τελεστής των $9-$σημείων είναι θετικά ορισμένος αν και μόνον αν ικανοποιούνται οι παρακάτω συνθήκες

$$\frac{1}{5} \;\leq\; \frac{bh_1^2}{ah_2^2} \;\leq\; 5 \tag{6.1.23}$$

(βλ. [48]). Ο πίνακας συντελεστών $A$ που αντιστοιχεί στο γραμμικό σύστημα που προέρχεται από τη σχέση (6.1.21) μπορεί να γραφτεί στην μορφή

$$A = \sqrt{\frac{a}{b}}\frac{h_2}{h_1}(I_{n_2} \otimes T_{n_1}) + \sqrt{\frac{b}{a}}\frac{h_1}{h_2}(T_{n_2} \otimes I_{n_1}) - \theta(T_{n_2} \otimes T_{n_1}) \tag{6.1.24}$$

ή ισοδύναμα στη μορφή

$$\begin{aligned} A \;=\; & \sqrt{\tfrac{a}{b}}\tfrac{h_2}{h_1}(I_{n_2} \otimes T_{n_1}) + \sqrt{\tfrac{b}{a}}\tfrac{h_1}{h_2}(T_{n_2} \otimes I_{n_1}) \\ & -\theta\left[\sqrt{\tfrac{a}{b}}\tfrac{h_2}{h_1}(I_{n_2} \otimes T_{n_1}) \cdot \sqrt{\tfrac{b}{a}}\tfrac{h_1}{h_2}(T_{n_2} \otimes I_{n_1})\right], \end{aligned} \tag{6.1.25}$$

όπου $n_1\,(\geq 2)$ και $n_2\,(\geq 2)$ είναι ο αριθμός των εσωτερικών κόμβων του πλέγματος σε κάθε διεύθυνση. Οι πίνακες $T_{n_1} \in \mathbb{R}^{n_1 \times n_1}$ και $T_{n_2} \in \mathbb{R}^{n_2 \times n_2}$ είναι συμμετρικοί και θετικά ορισμένοι της μορφής $\text{tridiag}(-1, 2, -1)$. Θέτοντας $A_1 := \sqrt{\frac{a}{b}}\frac{h_2}{h_1}(I_{n_2} \otimes T_{n_1})$ και $A_2 := \sqrt{\frac{b}{a}}\frac{h_1}{h_2}(T_{n_2} \otimes I_{n_1})$, η σχέση (6.1.25) λαμβάνει τη μορφή

$$A = A_1 + A_2 - \theta A_1 A_2. \tag{6.1.26}$$

Εκτελώντας απλές πράξεις με βάση τις ιδιότητες των τανυστικών γινόμενων διαπιστώνουμε ότι οι πίνακες $A_1$ και $A_2$ αντιμετατίθενται (βλ. [36]). Το γεγονός αυτό θα μας βοηθήσει πολύ στην μετέπειτα μελέτη των μεθόδων, που θα χρησιμοποιηθούν, για την λύση του συστήματος $Ax = c$.

Θα πρέπει και πάλι στο σημείο αυτό να τονίσουμε ότι βασική μας επιδίωξη είναι να λύσουμε ένα γραμμικό σύστημα με πίνακα συντελεστών αγνώστων

τον πίνακα (6.1.26), χρησιμοποιώντας την Προρρυθμισμένη μέθοδο Συζυγών Κλίσεων με προρρυθμιστή τον πίνακα της EADI μεθόδου όπως αυτή ορίστηκε από τον Guittet [27]. Στην πορεία όμως αυτής της μελέτης, όπως είδαμε σε προηγούμενο κεφάλαιο, καταλήγουμε στη μελέτη και την εύρεση των βέλτιστων τιμών των παραμέτρων επιτάχυνσης του EADI σχήματος, καθώς επίσης και στην εύρεση της βέλτιστης παραμέτρου παρεκβολής. Το γενικό σχήμα για τις $p-$διαστάσεις που πρότεινε ο Guittet είναι το παρακάτω

$$\begin{aligned}(I+rA_1)u^{(m+\frac{1}{p})} &= [\prod_{i=p}^{1}(I+rA_i)-\omega rA]u^{(m)}+\omega rb,\\ (I+rA_j)u^{(m+\frac{j}{p})} &= u^{(m+\frac{j-1}{p})},\ j=2,\ldots,p,\end{aligned} \tag{6.1.27}$$

όπου $A=\sum_{i=1}^{p}A_i$. Στη συνέχεια θα χρησιμοποιήσουμε μια παραλλαγή του παραπάνω σχήματος στην περίπτωση των δύο διαστάσεων με σταθερή παράμετρο επιτάχυνσης το οποίο θα έχει την παρακάτω μορφή

$$\begin{aligned}(I+rA_1)u^{(m+\frac{1}{2})} &= [(I+rA_2)(I+rA_1)-\omega A]u^{(m)}+\omega b,\\ (I+rA_2)u^{(m+1)} &= u^{(m+\frac{1}{2})}.\end{aligned} \tag{6.1.28}$$

Σ′ αυτή τη μορφή $A=A_1+A_2-\theta A_1A_2$ κι ακόμη το γινόμενο $\omega r$ του (6.1.27) έχει αντικατασταθεί από $\omega$ στην (6.1.28). Καθίσταται φανερό ότι ο EADI προρρυθμιστής θα δίνεται από την έκφραση

$$M = \frac{1}{\omega}(I+rA_2)(I+rA_1). \tag{6.1.29}$$

Το επαναληπτικό σχήμα, που προέρχεται από την (6.1.28), με απαλοιφή του ενδιάμεσου διανύσματος $u^{(m+\frac{1}{2})}$ είναι το παρακάτω

$$u^{(m+1)}=T_{EADI}u^{(m)}+c_{EADI}, \tag{6.1.30}$$

όπου

$$T_{EADI}=I-\omega(I+rA_1)^{-1}(I+rA_2)^{-1}A,\quad c_{EADI}=(I+rA_1)^{-1}(I+rA_2)^{-1}\omega b. \tag{6.1.31}$$

Οι παράμετροι $r,\omega\in\mathbb{R}_+$ υπολογίζονται έτσι ώστε να επιταχύνουν τη σύγκλιση. Για τον υπολογισμό τους με τον καλύτερο δυνατό τρόπο (βέλτιστο) θα αξιοποιήσουμε τις ιδιοτιμές των πινάκων $A_i$, $i=1,2$. Έστω ότι οι ιδιοτιμές αυτές ανήκουν στο σύνολο

$$S := \{\lambda_1,\lambda_2\in\mathbb{R}_+|\alpha_1\le\lambda_1\le\beta_1,\ \alpha_2\le\lambda_2\le\beta_2\},$$

όπου $\alpha_i$, $\beta_i \in \mathbb{R}_+$, $i = 1, 2$. Το γεγονός ότι οι πίνακες $A_1$ και $A_2$ είναι συμμετρικοί, θετικά ορισμένοι και αντιμετατίθενται, μας επιτρέπει να έχουμε ένα πλήρες σύστημα ορθοκανονικών ιδιοδιανυσμάτων το οποίο μπορεί να είναι κοινό και για τους δύο πίνακες. Έχοντας αυτό υπόψη μπορούμε να βρούμε ότι οι ιδιοτιμές του επαναληπτικού πίνακα $T_{EADI}$ δίνονται από τις εκφράσεις

$$\lambda_{T_{EADI}} = 1 - \omega \frac{\lambda_1 + \lambda_2 - \theta\lambda_1\lambda_2}{(1 + r\lambda_1)(1 + r\lambda_2)}. \tag{6.1.32}$$

Ορίζοντας στη συνέχεια το δεξιό μέλος ως συνάρτηση των $\lambda_i$, $i = 1, 2$, έχουμε ότι

$$f \equiv f(\lambda_1, \lambda_2) := \frac{\lambda_1 + \lambda_2 - \theta\lambda_1\lambda_2}{(1 + r\lambda_1)(1 + r\lambda_2)}. \tag{6.1.33}$$

Λόγω του γεγονότος ότι ο πίνακας $A$ είναι θετικά ορισμένος έχουμε ότι ο αριθμητής της συνάρτησης είναι θετικός. Το ίδιο βέβαια ισχύει και για τον παρονομαστή αφού τόσο οι ιδιοτιμές $\lambda_1$ και $\lambda_2$ όσο και η παράμετρος $r$ είναι θετικές. Έτσι οι ιδιοτιμές του EADI ικανοποιούν τις παρακάτω ανισότητες

$$\inf_{\lambda_1,\lambda_2 \in S}(1 - \omega f) \leq \lambda_{T_{EADI}} \leq \sup_{\lambda_1,\lambda_2 \in S}(1 - \omega f).$$

Η φασματική ακτίνα του πίνακα $T_{EADI}$ ικανοποιεί με τη σειρά της την παρακάτω ανισότητα

$$\rho(T_{EADI}) \leq \sup_{\lambda_1,\lambda_2 \in S} |1 - \omega f| . \tag{6.1.34}$$

Για τον υπολογισμό του $\sup_{\lambda_1,\lambda_2 \in S} |1 - \omega f|$ χρειάζεται να υπολογίσουμε τη μέγιστη και την ελάχιστη τιμή της συνάρτησης $f$. Θεωρούμε, λοιπόν, τους συμβολισμούς

$$G := \max_{\lambda_1,\lambda_2 \in S} f \quad \text{και} \quad g := \min_{\lambda_1,\lambda_2 \in S} f. \tag{6.1.35}$$

Για τον υπόλογισμό αυτών των τιμών θα χρησιμοποιήσουμε ένα θεώρημα του Απειροστικού Λογισμού σύμφωνα με το οποίο εάν το πρόσημο της ποσότητας $\frac{\partial f}{\partial \lambda_i}$, $i = 1, 2$, είναι ανεξάρτητο από τη μεταβλητή $\lambda_i$, τότε οι ακραίες τιμές της $f$ θα λαμβάνονται στα άκρα του πεδίου ορισμού τους. Θα πρέπει να θυμηθούμε ότι παρόμοια το ίδιο θεώρημα χρησιμοποιήθηκε και στην περίπτωση της μελέτης του σχήματος του Guittet. Στη δική μας περίπτωση έχουμε το πρόσημο της έκφρασης

$$\frac{\partial f}{\partial \lambda_i} = \frac{\lambda_j \left( (\frac{1}{\lambda_j} - \theta) - r \right)}{(1 + r\lambda_i)^2(1 + r\lambda_j)}, \; i \neq j = 1, 2, \tag{6.1.36}$$

είναι ανεξάρτητο από την τιμή της μεταβλητής $\lambda_i$. Έτσι η μέγιστη τιμή $G$ και η ελάχιστη τιμή $g$ λαμβάνονται στις κορυφές του ορθογωνίου $S$. Επομένως οι ακραίες τιμές της είναι κάποιες έκ των $f(\alpha_1, \alpha_2)$, $f(\alpha_1, \beta_2)$, $f(\beta_1, \alpha_2)$ και $f(\beta_1, \beta_2)$.

*Παρατήρηση* 6.1.3. : Στο σημείο αυτό θα προσπαθήσουμε να συνδέσουμε τις βέλτιστες EADI μεθόδους (6.1.28) με αυτές των αντίστοιχων βέλτιστων Προρρυθμιστών της μεθόδου των Συζυγών Κλίσεων με τον προρρυθμιστή πίνακα να δίνεται από την έκφραση (6.1.29). Για την EADI μέθοδο οι βέλτιστες τιμές των παραμέτρων $r$, $\omega$, τις οποίες συμβολίζουμε με $r^*$, $\omega^*$ (βλ. [27]), μπορούν να βρεθούν με την ελαχιστοποίηση του λόγου $\frac{G}{g}$, όπου $G^*$ και $g^*$ να είναι οι αντίστοιχες βέλτιστες τιμές των $G$ και $g$, αντίστοιχα, έτσι ώστε ο βέλτιστος (μικρότερος) λόγος να δίνεται από την έκφραση

$$\frac{G^*}{g^*} = \min_{r\in(0,\infty)} \frac{G}{g}. \tag{6.1.37}$$

Σ' αυτήν την περίπτωση η βέλτιστη τιμή της παραμέτρου $\omega$ θα δίνεται από την

$$\omega^* = \frac{2}{G^* + g^*}. \tag{6.1.38}$$

Για την αντίστοιχη Προρρυθμισμένη μέθοδο Συζυγών Κλίσεων, όπως έχουμε ήδη αναφέρει, η παράμετρος παρεκβολής (extrapolation) δεν επηρεάζει τη σύγκλιση στην προρρύθμιση και η βέλτιστη μέθοδος θα είναι αυτή για την οποία ο λόγος $\kappa(M^{-1}A) = \frac{\lambda_{\max}(M^{-1}A)}{\lambda_{\min}(M^{-1}A)}$ θα έχει την ελάχιστη τιμή. Βέβαια, αυτό συμβαίνει όταν ο λόγος $\frac{G}{g}$ ελαχιστοποιείται.

Όλα τα παραπάνω συνοψίζονται στο παρακάτω θεώρημα.

**Θεώρημα 6.1.1.** *Σύμφωνα με τους μέχρι τώρα συμβολισμούς που έχουμε εισαγάγει και τις υποθέσεις που έχουμε θεωρήσει, ο βέλτιστος (μικρότερος) δείκτης κατάστασης, με βάση τον EADI Προρρυθμιστή $(I + rA_2)(I + rA_1)$, για τη μέθοδο των Συζυγών Κλίσεων λαμβάνεται για την τιμή της παραμέτρου επιτάχυνσης $r = r^*$, η οποία βελτιστοποιεί το αντίστοιχο EADI πρόβλημα στην (6.1.37). Έτσι ο βέλτιστος δείκτης κατάστασης δίνεται από την έκφραση*

$$\kappa^* = \frac{G^*}{g^*}. \tag{6.1.39}$$

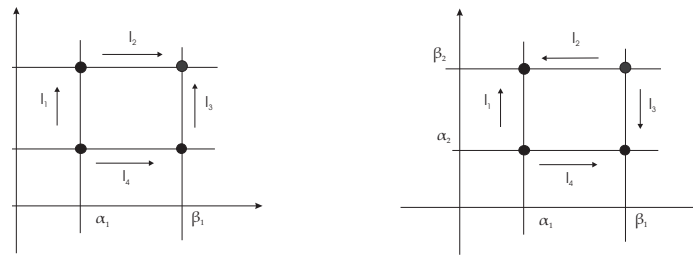
Όπως αναφέρθηκε και προηγουμένως είναι απαραίτητη η εύρεση των κορυφών του $S$ στις οποίες λαμβάνονται οι τιμές $G$ και $g$ της συνάρτησης $f$. Από τη σχέση (6.1.36) έχουμε ότι το πρόσημο της παραγώγου της συνάρτησης $f$ εξαρτάται από την παράμετρο $r$, η οποία ορίζεται σε διαστήματα της μορφής $[\frac{1}{\beta_i} - \theta, \frac{1}{\alpha_i} - \theta]$. $i = 1, 2$, Αυτή, με τη σειρά της, εξαρτάται από τη διάταξη των άκρων των διαστημάτων $[\alpha_1, \beta_1]$ και $[\alpha_2, \beta_2]$. Θεωρούμε λοιπόν τις παρακάτω τρεις διατάξεις

$$\begin{array}{lll} A: & \alpha_2 < \alpha_1 < \beta_2 < \beta_1, \\ B: & \alpha_2 < \alpha_1 < \beta_1 < \beta_2, \\ C: & \alpha_2 < \beta_2 < \alpha_1 < \beta_1. \end{array} \tag{6.1.40}$$

*Παρατήρηση* 6.1.4. : Θα πρέπει εδώ να τονιστεί ότι υπάρχουν και άλλες τρεις ακόμη διατάξεις οι οποίες λαμβάνονται από τις τρεις προηγούμενες με αντιμετάθεση των δεικτών 1 και 2. Συνεπώς ό,τι προκύψει από την ανάλυση που θα ακολουθήσει με βάση τις παραπάνω διατάξεις θα μπορεί αμέσως να αναφερθεί και στις τρεις παραλειφθείσες. Επίσης, σε περίπτωση όπου κάποια γνήσια ανισότητα καταστεί ισότητα τότε την τελευταία μπορούμε να τη θεωρήσουμε ως οριακή περίπτωση αυτής της γνήσιας ανισότητας.

Για να απλοποιήσουμε την ανάλυσή μας παραμετρικοποιούμε τις πλευρές του $S$, όπως φαίνεται στις σχέσεις (6.1.41) και όπως παρουσιάζεται στο Σχήμα 6.1

$$\begin{cases} l_1(t) = (\alpha_1, \alpha_2 + t(\beta_2 - \alpha_2)), & t \in [0,1], \\ l_2(t) = (\alpha_1 + t(\beta_1 - \alpha_1), \beta_2), & t \in [0,1], \\ l_3(t) = (\beta_1, \alpha_2 + t(\beta_2 - \alpha_2)), & t \in [0,1], \\ l_4(t) = (\alpha_1 + t(\beta_1 - \alpha_1), \alpha_2), & t \in [0,1]. \end{cases} \tag{6.1.41}$$




Σχήμα 6.1: Μονοτονία της $f$ κατά μήκος των πλευρών του ορθογωνίου $S$ σε δύο συγκεκριμένες περιπτώσεις.

Για να απλοποιήσουμε ακόμη περισσότερο την ανάλυσή μας θα θεωρήσουμε ότι $\frac{1}{\beta_1}, \frac{1}{\beta_2} > \theta$ έτσι ώστε $0 < \frac{1}{\beta_i} - \theta < \frac{1}{\alpha_i} - \theta$, $i = 1, 2$. Σημειώνεται ότι

για $\theta = 0$ οι ανισότητες αυτές ικανοποιούνται. Στην περίπτωση όμως, κατά την οποία $\theta = \theta^*$, το $\theta^*$ μπορεί να μην είναι το κάτω φράγμα του $\frac{1}{\beta_i}$ και του $\frac{1}{\alpha_i}$. Οι περιπτώσεις αυτές που είναι δυνατόν να προκύψουν θα εξεταστούν αργότερα.

Στη συνέχεια, λαμβάνοντας μερικές παραγώγους της $f$ κατά μήκος των πλευρών $l_i(t),\ i = 1, \ldots, 4$, του $S$ ως προς $t$, για παράδειγμα η παράγωγος για την πλευρά $l_1$ δίνεται από την έκφραση

$$\frac{df(l_1(t))}{dt} = \nabla f \cdot \frac{dl_1(t)}{dt} = \frac{\partial f(l_1(t))}{\partial \lambda_2}(\beta_2 - \alpha_2) = \frac{\lambda_1 \left( (\frac{1}{\lambda_1} - \theta) - r \right)}{(1 + r\lambda_2)^2(1 + r\lambda_1)}(\beta_2 - \alpha_2). \tag{6.1.42}$$

Από την παραπάνω έκφραση βλέπουμε ότι η παραγωγός είναι θετική εάν $r < \frac{1}{\lambda_1 - \theta}$. Με όμοιο τρόπο βρίσκουμε τις παραγώγους της $f$ κατά μήκος των τριών άλλων πλευρών. Βρίσκουμε λοιπόν τη μονοτονία της $f(l_i)(t) : [0, 1] \to \mathbb{R},\ i = 1, \ldots, 4$, της οποίας το πρόσημο των παραγώγων παρουσιάζουμε στον Πίνακα 6.1.

| $r$ | $\frac{1}{\alpha_1} - \theta$ | |
|---|---|---|
| $\frac{df(l_1)}{dt}$ | $+$ | $-$ |

| $r$ | $\frac{1}{\beta_2} - \theta$ | |
|---|---|---|
| $\frac{df(l_2)}{dt}$ | $+$ | $-$ |

| $r$ | $\frac{1}{\beta_1} - \theta$ | |
|---|---|---|
| $\frac{df(l_3)}{dt}$ | $+$ | $-$ |

| $r$ | $\frac{1}{\alpha_2} - \theta$ | |
|---|---|---|
| $\frac{df(l_4)}{dt}$ | $+$ | $-$ |

Πίνακας 6.1: Πρόσημα παραγώγων κατά μήκος των πλευρών του Ορθογωνίου $S$.

Στη συνέχεια με τη βοήθεια του παραπάνω πίνακα θα δώσουμε σε κάθε διάστημα ορισμού του $r$, όπως αυτά προκύπτουν από την διάταξη των ακροτάτων της $f$ κατά μήκος των πλευρών του ορθογωνίου, τη μονοτονία της συνάρτησης $f$. Για παράδειγμα: Στην πρώτη από τις περιπτώσεις διάταξης των άκρων των διαστημάτων έχουμε: 1) Στο αριστερό σχήμα του 6.1 τα βέλη κατά μήκος κάθε πλευράς του ορθογωνίου $S$ δείχνουν ότι η συνάρτηση $f$ αυξάνει καθώς το $t$ αυξάνει από το 0 στο 1. Η πρώτη αυτή περίπτωση λαμβάνεται όταν $r \in \left(0, \frac{1}{\beta_1} - \theta\right]$. Έτσι η μέγιστη τιμή της $f$ δίνεται από την $G = f(\beta_1, \beta_2)$ και η ελάχιστη από την $g = f(\alpha_1, \alpha_2)$. 2) Στο δεξιό σχήμα του 6.1 τα βέλη δείχνουν πότε η $f$ αυξάνει και πότε ελαττούται κατά μήκος των πλευρών του $S$. Η εικόνα αυτή λαμβάνεται όταν $r \in \left[\frac{1}{\beta_2} - \theta, \frac{1}{\alpha_1} - \theta\right]$. Στην περίπτωση αυτή η

μέγιστη τιμή της $f$ δίνεται από την έκφραση $G = \max\{f(\alpha_1, \beta_2), f(\beta_1, \alpha_2)\}$ και η ελάχιστη από την $g = \min\{f(\alpha_1, \alpha_2), f(\beta_1, \beta_2)\}$. Η αναφερθείσα διαδικασία επαναλαμβάνεται σε κάθε διάστημα ορισμού του $r$ με τη θεώρηση του προσήμου των αντίστοιχων μερικών παραγώγων του Πίνακα 6.1 και του Σχήματος 6.1. Εργαζόμενοι με παρόμοιο τρόπο και στις περιπτώσεις των δύο άλλων διατάξεων (6.1.40) καταλήγουμε σε αποτελέσματα τα οποία παρουσιάζουμε στους πίνακες 6.2, 6.3, 6.4, όπου η $G$ και η $g$ υπολογίζονται και για τις τρεις περιπτώσεις διάταξης, $A$, $B$, $C$, αντίστοιχα, καθώς το $r$ αυξάνει στο $(0, +\infty)$.

| $r$ | $0$ | $\frac{1}{\beta_1} - \theta$ | $\frac{1}{\beta_2} - \theta$ | $\frac{1}{\alpha_1} - \theta$ | $\frac{1}{\alpha_2} - \theta$ | $+\infty$ |
|---|---|---|---|---|---|---|
| $G$ | $f(\beta_1, \beta_2)$ | $f(\beta_1, \alpha_2)$ | $\max\{f(\alpha_1, \beta_2), f(\beta_1, \alpha_2)\}$ | $f(\beta_1, \alpha_2)$ | $f(\alpha_1, \alpha_2)$ | |
| $g$ | $f(\alpha_1, \alpha_2)$ | $f(\alpha_1, \alpha_2)$ | $\min\{f(\alpha_1, \alpha_2), f(\beta_1, \beta_2)\}$ | $f(\beta_1, \beta_2)$ | $f(\beta_1, \beta_2)$ | |

Πίνακας 6.2: Περίπτωση $A$ $(\alpha_2 < \alpha_1 < \beta_2 < \beta_1)$: Μέγιστη $G$ και ελάχιστη $g$ τιμή της $f$

| $r$ | $0$ | $\frac{1}{\beta_2} - \theta$ | $\frac{1}{\beta_1} - \theta$ | $\frac{1}{\alpha_1} - \theta$ | $\frac{1}{\alpha_2} - \theta$ | $+\infty$ |
|---|---|---|---|---|---|---|
| $G$ | $f(\beta_1, \beta_2)$ | $f(\alpha_1, \beta_2)$ | $\max\{f(\alpha_1, \beta_2), f(\beta_1, \alpha_2)\}$ | $f(\beta_1, \alpha_2)$ | $f(\alpha_1, \alpha_2)$ | |
| $g$ | $f(\alpha_1, \alpha_2)$ | $f(\alpha_1, \alpha_2)$ | $\min\{f(\alpha_1, \alpha_2), f(\beta_1, \beta_2)\}$ | $f(\beta_1, \beta_2)$ | $f(\beta_1, \beta_2)$ | |

Πίνακας 6.3: Περίπτωση $B$ $(\alpha_2 < \alpha_1 < \beta_1 < \beta_2)$: Μέγιστη $G$ και ελάχιστη $g$ τιμή της $f$

| $r$ | $0$ | $\frac{1}{\beta_1} - \theta$ | $\frac{1}{\alpha_1} - \theta$ | $\frac{1}{\beta_2} - \theta$ | $\frac{1}{\alpha_2} - \theta$ | $+\infty$ |
|---|---|---|---|---|---|---|
| $G$ | $f(\beta_1, \beta_2)$ | $f(\beta_1, \alpha_2)$ | $f(\beta_1, \alpha_2)$ | $f(\beta_1, \alpha_2)$ | $f(\alpha_1, \alpha_2)$ | |
| $g$ | $f(\alpha_1, \alpha_2)$ | $f(\alpha_1, \alpha_2)$ | $f(\alpha_1, \beta_2)$ | $f(\beta_1, \beta_2)$ | $f(\beta_1, \beta_2)$ | |

Πίνακας 6.4: Περίπτωση $C$ $(\alpha_2 < \beta_2 < \alpha_1 < \beta_1)$: Μέγιστη $G$ και ελάχιστη $g$ τιμή της $f$

Με τη βοήθεια των παραπάνω αποτελεσμάτων θα παρουσιάσουμε στην επόμενη παράγραφο τη διαδικασία για την εύρεση των βέλτιστων παραμέτρων επιτάχυνσης $r$ και παρεκβολής $\omega$.

### 6.1.4 Βέλτιστη Παράμετρος Επιτάχυνσης

Στην παράγραφο αυτή θα παρουσιάσουμε τη διαδικασία εύρεσης των βέλτιστων παραμέτρων επιτάχυνσης $r$ στις τρεις περιπτώσεις διάταξης (6.1.40). Σε ό,τι αφορά την παράμετρο παρεκβολής $\omega$ αυτή υπολογίζεται αμέσως από τη σχέση (6.1.38). Κατά την ανάλυσή μας θα χρησιμοποιήσουμε το σύμβολο " $\sim$ " και θα γράφουμε

$$E_1 \ \sim \ E_2 \tag{6.1.43}$$

για να υποδηλώνουμε ότι οι εκφράσεις ή ποσότητες $E_1$ και $E_2$ έχουν το ίδιο πρόσημο.

**Περίπτωση** $A$ $(\alpha_2 < \alpha_1 < \beta_2 < \beta_1)$

Θεωρούμε το λόγο $\frac{G}{g}$, για τις τιμές του $r$ σε κάθε ένα από τα πέντε διαστήματα του Πίνακα 6.2 και βρίσκουμε τη συμπεριφορά του, θεωρώντας τις αντίστοιχες παραγώγους ως προς $r$. Για παράδειγμα, έστω ότι $r \in \left(0, \frac{1}{\beta_1} - \theta\right]$. Τότε θα έχουμε

$$\frac{G}{g} \ = \ \frac{f(\beta_1,\beta_2)}{f(\alpha_1,\alpha_2)} \ = \ \frac{(\beta_1+\beta_2-\theta\beta_1\beta_2)\,(1+\alpha_1 r)(1+\alpha_2 r)}{(\alpha_1+\alpha_2-\theta\alpha_1\alpha_2)\,(1+\beta_1 r)(1+\beta_2 r)}.$$

Λαμβάνοντας παράγωγο στην παραπάνω έκφραση και γνωρίζοτας το πρόσημο του παρονομαστή, της έκφρασης που θα προκύψει, και το γεγονός ότι ο λόγος $\frac{\beta_1+\beta_2-\theta\beta_1\beta_2}{\alpha_1+\alpha_2-\theta\alpha_1\alpha_2}$ είναι θετικός επικεντρωνόμαστε στο πρόσημο του αριθμητή της έκφρασης της παραγώγου. Έχουμε λοιπόν ότι

$$\begin{aligned}
\frac{d\left(\frac{G}{g}\right)}{dr} \ &\sim \ (2r_1r_2+(\alpha_1+\alpha_2))((1+\beta_1 r)(1+\beta_2 r)) \\
&\quad -(2r\beta_1\beta_2+(\beta_1+\beta_2))((1+\alpha_1 r)(1+\alpha_2 r)) \\
&\sim \ 2r\alpha_1\alpha_2(1+\beta_1 r)(1+\beta_2 r)-2r\beta_1\beta_2(1+\alpha_1 r)(1+\alpha_2 r) \\
&\quad +(\alpha_1+\alpha_2)(1+\beta_1 r)(1+\beta_2 r) \\
&\quad -(\beta_1+\beta_2)(1+\alpha_1 r)(1+\alpha_2 r) \\
&\sim \ 2r^2(\beta_1+\beta_2)\alpha_1\alpha_2-r^2(\alpha_1+\alpha_2)\beta_1\beta_2+2r\alpha_1\alpha_2 \\
&\quad -2r\beta_1\beta_2+(\alpha_1+\alpha_2)-(\beta_1+\beta_2) \\
&\sim \ -[\alpha_1\beta_1(\beta_2-\alpha_2)+\alpha_2\beta_2(\beta_1-\alpha_1)]r^2-(\beta_1\beta_2-\alpha_1\alpha_2)r \\
&\quad -[(\beta_1-\alpha_1)+(\beta_2-\alpha_2)]<0,
\end{aligned} \tag{6.1.44}$$

αφού όλοι οι συντελεστές του παραπάνω τριωνύμου είναι αρνητικοί και το $r$ θετικό. Έτσι συμπεραίνουμε ότι ο λόγος $\frac{G}{g}$ είναι φθίνουσα συνάρτηση του

$r$ στο διάστημα $\left(0, \frac{1}{\beta_1} - \theta\right]$. Θεωρώντας τώρα ότι το $r$ ανήκει στο διάστημα $\left[\frac{1}{\beta_1} - \theta, \frac{1}{\beta_2} - \theta\right]$, έχουμε ότι

$$\begin{aligned} \frac{G}{g} &= \frac{f(\beta_1,\alpha_2)}{f(\alpha_1,\alpha_2)} = \frac{(\beta_1+\alpha_2-\theta\beta_1\alpha_2)(1+\alpha_1 r)(1+\alpha_2 r)}{(\alpha_1+\alpha_2-\theta\alpha_1\alpha_2)(1+\beta_1 r)(1+\alpha_2 r)} \\ &= \frac{(\beta_1+\alpha_2-\theta\beta_1\alpha_2)(1+\alpha_1 r)}{(\alpha_1+\alpha_2-\theta\alpha_1\alpha_2)(1+\beta_1 r)}, \end{aligned}$$

Όμοια με πριν παραγωγίζοντας ως προς $r$ το λόγο $\frac{G}{g}$ έχουμε την ισοδύναμη έκφραση για την παράγωγο, οπότε προκύπτει ότι

$$\frac{d\left(\frac{G}{g}\right)}{dr} \sim \alpha_1 - \beta_1 < 0.$$

Αυτό ισχύει γιατί ο πίνακας $A_1 + A_2 - \theta A_1 A_2$ είναι θετικά ορισμένος και τα $\beta_1, \alpha_1$ και $\beta_2, \alpha_2$ είναι ιδιοτιμές των $A_1$ και $A_2$, αντίστοιχα. Συμπεραίνουμε, λοιπόν, ότι ο λόγος για τον οποίο ενδιαφερόμαστε είναι φθίνουσα συνάρτηση του $r$ στο εν λόγω διάστημα. Στο τρίτο, στη σειρά, διάστημα του $r$, δηλαδή στο $\left[\frac{1}{\beta_2} - \theta, \frac{1}{\alpha_1} - \theta\right]$, παρατηρούμε ότι η μονότονη συμπεριφορά του λόγου $\frac{G}{g}$ είναι ανεξάρτητη από το ποια από τις δυο ποσότητες, του μεγίστου είναι η μεγαλύτερη και αυτή εξαρτάται μόνον από ποια από τις δύο ποσότητες του ελαχίστου είναι η μικρότερη. Από αυτήν την παρατήρηση, έχουμε ότι όποια εκ των $f(\alpha_1, \beta_2)$ και $f(\beta_1, \alpha_2)$ θεωρηθεί ως η μέγιστη και εάν η ελάχιστη είναι η $f(\alpha_1, \alpha_2)$, τότε αποδεικνύεται ότι για τους λόγους

$$\frac{G}{g} = \frac{f(\alpha_1, \beta_2)}{f(\alpha_1, \alpha_2)} = \frac{(\alpha_1 + \beta_2 - \theta\alpha_1\beta_2)}{(\alpha_1 + \alpha_2 - \theta\alpha_1\alpha_2)} \frac{(1 + r\alpha_2)}{(1 + r\beta_2)},$$

$$\frac{G}{g} = \frac{f(\beta_1, \alpha_2)}{f(\alpha_1, \alpha_2)} = \frac{(\beta_1 + \alpha_2 - \theta\beta_1\alpha_2)}{(\alpha_1 + \alpha_2 - \theta\alpha_1\alpha_2)} \frac{(1 + r\alpha_1)}{(1 + r\beta_1)}$$

έχουμε τις εκφράσεις για το πρόσημο της παραγώγου σε κάθε μία από τις περιπτώσεις που αναφέρθηκαν πριν

$$\frac{d\left(\frac{G}{g}\right)}{dr} \sim \alpha_1 - \beta_1 < 0 \quad \text{ή} \quad \frac{d\left(\frac{G}{g}\right)}{dr} \sim \alpha_2 - \beta_2 < 0$$

για την πρώτη και την δεύτερη περίπτωση, αντίστοιχα. Το γεγονός αυτό μας υποδεικνύει ότι σε κάθε μία περίπτωση ο λόγος αποτελεί φθίνουσα συνάρτηση του $r$. Εξάλλου, εάν το ελάχιστο δίνεται από την $f(\beta_1, \beta_2)$, μπορεί να αποδειχτεί, με όμοιο τρόπο, ότι ο λόγος $\frac{G}{g}$ είναι αύξουσα συνάρτηση του $r$. Είναι απαραίτητο, λοιπόν, να βρεθεί ποια από τις δύο ποσότητες $f(\beta_1, \beta_2)$ και $f(\alpha_1, \alpha_2)$ αντιστοιχεί στην ελάχιστη τιμή για την $f$. Για το σκοπό αυτό θεωρούμε τη διαφορά

$$\begin{array}{rcl} q(r) & = & f(\beta_1,\beta_2) - f(\alpha_1,\alpha_2) = \frac{\beta_1+\beta_2-\theta\beta_1\beta_2}{(1+r\beta_1)(1+r\beta_2)} - \frac{\alpha_1+\alpha_2-\theta\alpha_1\alpha_2}{(1+r\alpha_1)(1+r\alpha_2)} \ \sim \\ & & (\beta_1+\beta_2-\theta\beta_1\beta_2)(1+r\alpha_1)(1+r\alpha_2) \\ & - & (\alpha_1+\alpha_2-\theta\alpha_1\alpha_2)(1+r\beta_1)(1+r\beta_2) \ \sim \\ & & (\beta_1+\beta_2-\theta\beta_1\beta_2)(1+(\alpha_1+\alpha_2)r+\alpha_1\alpha_2 r^2) \\ & - & (\alpha_1+\alpha_2-\theta\alpha_1\alpha_2)(1+(\beta_1+\beta_2)r+\beta_1\beta_2 r^2) \sim \\ & - & \beta_2\alpha_2(\beta_1-\alpha_1)+\beta_1\alpha_1(\beta_2-\alpha_2)r^2 \\ & - & \theta[\beta_2\alpha_2(\beta_1-\alpha_1)+\beta_1\alpha_1(\beta_2-\alpha_2)]r \\ & + & [(1-\theta\alpha_1)(\beta_2-\alpha_2)+(1-\theta\beta_2)(\beta_1-\alpha_1)]. \end{array} \tag{6.1.45}$$

Παρατηρούμε ότι στο τριώνυμο του δεξιού μέλους ο πρώτος συντελεστής είναι αρνητικός ενώ ο τελευταίος είναι θετικός. Συνεπώς μόνο μία από τις ρίζες του τριωνύμου είναι θετική, την οποία και συμβολίζουμε με $r_{AB}$. Για να επιβεβαιώσουμε ότι η $r_{AB}$ βρίσκεται στο συγκεκριμένο διάστημα σημειώνουμε ότι λόγω της συνέχειας η ελάχιστη τιμή $g$ θα δίνεται στα αριστερά του διαστήματος από την ποσότητα $f(\alpha_1, \alpha_2)$ και ο λόγος $\frac{G}{g}$ θα φθίνει ενώ στα δεξιά αυτού από την ποσότητα $f(\beta_1, \beta_2)$ και ο αντίστοιχος λόγος θα αυξάνει. Σαν αποτέλεσμα της παραπάνω παρατήρησης έχουμε ότι το σημείο στο οποίο ο λόγος $\frac{G}{g}$ έχει ολικό ελάχιστο αντιστοιχεί σ΄ αυτό όπου παρουσιάζεται και το βέλτιστο. Για να το επιβεβαιωθεί το συμπέρασμά μας και σ΄ αυτό το σημείο μπορούμε να ελέγξουμε και να διαπιστώσουμε ότι όσο το $r$ αυξάνει στα δύο άλλα διαστήματα, $\left[\frac{1}{\alpha_1}-\theta, \frac{1}{\alpha_2}-\theta\right]$ και $\left[\frac{1}{\alpha_2}-\theta, +\infty\right)$, τότε και ο λόγος $\frac{G}{g}$ αυξάνει κι αυτός γνήσια μονότονα.

Έστω ότι οι συντελεστές του τριωνύμου στην (6.2.26) συμβολίζονται με

$$\begin{array}{rcl} \gamma_{AB} & = & -[\beta_2\alpha_2(\beta_1-\alpha_1)+\beta_1\alpha_1(\beta_2-\alpha_2)], \\ \delta_{AB} & = & -\theta[\beta_2\alpha_2(\beta_1-\alpha_1)+\beta_1\alpha_1(\beta_2-\alpha_2)], \\ \varepsilon_{AB} & = & (1-\theta\alpha_1)(\beta_2-\alpha_2)+(1-\theta\beta_2)(\beta_1-\alpha_1), \end{array}$$

αντίστοιχα. Τότε η βέλτιστη τιμή του $r$, δηλαδή η $r^* = r_{AB}$, θα δίνεται από την έκφραση

$$r^* = r_{AB} \ = \ \frac{-\delta_{AB} - (\delta_{AB}^2 - 4\gamma_{AB}\varepsilon_{AB})^{\frac{1}{2}}}{2\gamma_{AB}}. \tag{6.1.46}$$

**Περίπτωση** $B$ $(\alpha_2 < \alpha_1 < \beta_1 < \beta_2)$

Σ′ αυτήν την περίπτωση η διαδικασία που θα ακολουθήσουμε είναι ακριβώς η ίδια και θα οδηγήσει στα ίδια ακριβώς αποτελέσματα. Έτσι το βέλτιστο $r$ και στην παρούσα περίπτωση είναι αυτό που δίνεται από την έκφραση $r^* = r_{AB}$ στην (6.1.46).

**Περίπτωση** $C$ $(\alpha_2 < \beta_2 < \alpha_1 < \beta_1)$

Είναι εύκολο να ελέγξουμε ότι ο λόγος $\frac{G}{g}$ ως συνάρτηση του $r$ είναι γνησίως φθίνουσα στα πρώτα δύο διαστήματα και γνησίως αύξουσα στα δύο τελευταία. Θα δούμε στην συνέχεια αναλυτικότερα πως προέκυψε το παραπάνω συμπέρασμα. Θεωρούμε τους λόγους $\frac{G}{g}$ σε κάθε ένα από τα διαστήματα του Πίνακα 6.4.

Έχουμε λοιπόν τους λόγους:

$$\frac{G}{g} = \frac{f(\beta_1, \beta_2)}{f(\alpha_1, \alpha_2)}, \quad \frac{G}{g} = \frac{f(\beta_1, \alpha_2)}{f(\alpha_1, \alpha_2)},$$

$$\frac{G}{g} = \frac{f(\beta_1, \alpha_2)}{f(\beta_1, \beta_2)}, \quad \frac{G}{g} = \frac{f(\alpha_1, \alpha_2)}{f(\beta_1, \beta_2)}.$$

Για κάθε ένα από τους παραπάνω λόγους λαμβάνουμε τις παραγώγους ως προς $r$ το πρόσημο των οποίων δίνεται παρακάτω. Για τον πρώτο και τον τέταρτο λόγο γνωρίζουμε ήδη το πρόσημο της παραγώγου από προηγούμενη ανάλυση. Για τους δύο μεσαίους έχουμε τις εκφράσεις:

$$\frac{d\left(\frac{G}{g}\right)}{dr} = \frac{\beta_1 + \alpha_2 - \theta\beta_1\alpha_2}{\alpha_1 + \alpha_2 - \theta\alpha_1\alpha_2} \frac{d}{dt}\left(\frac{1 + r\alpha_1}{1 + r\beta_1}\right) \sim \alpha_1 - \beta_1 < 0,$$

$$\frac{d\left(\frac{G}{g}\right)}{dr} = \frac{\beta_1 + \alpha_2 - \theta\beta_1\alpha_2}{\beta_1 + \beta_2 - \theta\beta_1\beta_2} \frac{d}{dt}\left(\frac{1 + r\beta_2}{1 + r\alpha_2}\right) \sim \beta_2 - \alpha_2 > 0.$$

Έτσι, η μελέτη μας περιορίζεται στο μεσαίο διάστημα $\left[\frac{1}{\alpha_1} - \theta, \frac{1}{\beta_2} - \theta\right]$. Δηλαδή στη μελέτη του λόγου

$$\frac{G}{g} = \frac{f(\beta_1, \alpha_2)}{f(\alpha_1, \beta_2)}.$$

Θεωρώντας στη συνέχεια την παράγωγο του παραπάνω λόγου ως προς $r$, σ΄ αυτό το διάστημα, έχουμε τη σχέση

$$\begin{aligned} p(r) &= \frac{d\left(\frac{G}{g}\right)}{dr} = \frac{d}{dt}\left(\frac{f(\beta_1,\alpha_2)}{f(\alpha_1,\beta_2)}\right) \\ &\sim [(\alpha_1+\beta_2)+2r\alpha_1\beta_2](1+r(\beta_1+\alpha_2)+r^2\beta_1\alpha_2)- \\ &\quad [(\beta_1+\alpha_2)+2r\beta_1\alpha_2](1+r(\alpha_1+\beta_2)+r^2\alpha_1\beta_2) \\ &\sim (\alpha_1+\beta_2)+2r\alpha_1\beta_2+r^2\alpha_1\beta_2(\beta_1+\alpha_2)-(\beta_1+\alpha_2)- \\ &\quad 2r\beta_1\alpha_2-r^2\beta_1\alpha_2(\alpha_1+\beta_2) \\ &\sim [\beta_1\alpha_1(\beta_2-\alpha_2)-\beta_2\alpha_2(\beta_1-\alpha_1)]r^2+ \\ &\quad 2(\alpha_1\beta_2-\beta_1\alpha_2)r+(\beta_2-\alpha_2)-(\beta_1-\alpha_1). \end{aligned} \tag{6.1.47}$$

Λαμβάνοντας τη διακρίνουσα $D$ του τριωνύμου του δεξιού μέλους της (6.1.47) μπορούμε να βρούμε ότι

$$D = 4(\beta_1-\beta_2)(\alpha_1-\alpha_2)(\beta_1-\alpha_1)(\beta_2-\alpha_2) > 0. \tag{6.1.48}$$

Το γεγονός ότι η διακρίνουσα είναι θετική μας οδηγεί στο συμπέρασμα ότι η (6.1.47) έχει δύο πραγματικές άνισες ρίζες. Έστω ότι συμβολίζουμε τους συντελεστές του τριωνύμου με $\gamma_C$, $\delta_C$, $\varepsilon_C$, αντίστοιχα. Δηλαδή,

$$\gamma_C = \beta_1\alpha_1(\beta_2-\alpha_2)-\beta_2\alpha_2(\beta_1-\alpha_1),\ 2\delta_C = 2(\alpha_1\beta_2-\alpha_2\beta_1),\ \varepsilon_C = (\beta_2-\alpha_2)-(\beta_1-\alpha_1). \tag{6.1.49}$$

Εκτός από τη θέση των δύο αυτών ριζών του τριωνύμου στην (6.1.47) θα χρειαστούμε επίσης τις τιμές του λόγου $\frac{G}{g}$ στα άκρα του $\left[\frac{1}{\alpha_1}-\theta, \frac{1}{\beta_2}-\theta\right]$. Μπορούμε να βρούμε, μετά από κάποιους υπολογισμούς, ότι

$$\begin{aligned} &\frac{G}{g}\left(\frac{1}{\alpha_1}-\theta\right)-\frac{G}{g}\left(\frac{1}{\beta_2}-\theta\right) = \frac{(1+(\frac{1}{a_1}-\theta)\alpha_1)(1+(\frac{1}{a_1}-\theta)\beta_2)}{(1+(\frac{1}{a_1}-\theta)\beta_1)(1+(\frac{1}{a_1}-\theta)\alpha_2)}-\frac{(1+(\frac{1}{\beta_2}-\theta)\alpha_1)(1+(\frac{1}{\beta_2}-\theta)\beta_2)}{(1+(\frac{1}{\beta_2}-\theta)\beta_1)(1+(\frac{1}{\beta_2}-\theta)\alpha_2)} \\ &\sim (2-\theta\alpha_1)(1+\tfrac{\beta_2}{\alpha_1}-\theta\beta_2)(1+\tfrac{\beta_1}{\beta_2}-\theta\beta_1)(1+\tfrac{\alpha_2}{\beta_2}-\theta\alpha_2)- \\ &(2-\theta\beta_2)(1+\tfrac{\beta_1}{\alpha_1}-\theta\beta_1)(1+\tfrac{\alpha_1}{\beta_2}-\theta\alpha_1)(1+\tfrac{\alpha_2}{\alpha_1}-\theta\alpha_2) \\ &\sim \alpha_1(2-\theta\alpha_1)(\alpha_1+\beta_2-\theta\beta_2\alpha_1)(\beta_1+\beta_2-\theta\beta_1\beta_2)(\beta_2+\alpha_2-\theta\alpha_2\beta_2) \\ &-\beta_2(2-\theta\beta_2)(\alpha_1+\beta_1-\theta\alpha_1\beta_1)(\beta_2+\alpha_1-\theta\alpha_1\beta_2)(\alpha_1+\alpha_2-\theta\alpha_1\alpha_2) \\ &\sim -2\delta_C+\theta\gamma_C = (\beta_1-\alpha_1)\alpha_2(2-\theta\beta_2)-(\beta_2-\alpha_2)\alpha_1(2-\theta\beta_1). \end{aligned} \tag{6.1.50}$$

Θεωρώντας ότι $\delta_C < 0$ έχουμε ισοδύναμα ότι $\frac{\beta_2}{\beta_1} < \frac{\alpha_2}{\alpha_1}$ $(< 1)$. Από το τελευταίο μπορούμε εύκολα να πάρουμε ότι $\frac{\beta_2-\alpha_2}{\beta_1-\alpha_1} < \frac{\beta_2}{\beta_1} < 1$, από το οποίο έχουμε ότι $\beta_2-\alpha_2 < \beta_1-\alpha_1$ ή ότι $\varepsilon_C < 0$. Στη συνέχεια εξετάζουμε δύο υποπεριπτώσεις που εξαρτώνται από το πρόσημο του $\gamma_C$.

Υποπερίπτωση $C_1$ : $\gamma_C > 0$. Εφόσον $\frac{\varepsilon_C}{\gamma_C} < 0$, οι δύο ρίζες του τριωνύμου στην

(6.1.47) θα πρέπει να έχουν διαφορετικό πρόσημο. Η θετική από αυτές θα είναι η

$$r_C \;=\; \frac{-\delta_C + (\delta_C^2 - \gamma_C \varepsilon_C)^{\frac{1}{2}}}{\gamma_C}. \tag{6.1.51}$$

Παρατηρούμε ότι εξαιτίας του γεγονότος ότι $-2\delta_C + \theta\gamma_C > 0$, από την (6.1.50), έχουμε ότι $\frac{G}{g}\left(\frac{1}{\alpha_1} - \theta\right) > \frac{G}{g}\left(\frac{1}{\beta_2} - \theta\right)$. Το τελευταίο συνεπάγεται ότι δεν είναι δυνατόν να έχουμε $r_C < \frac{1}{\alpha_1} - \theta$, εφόσον τότε ο λόγος $\frac{G}{g}$ θα αυξάνει στο διάστημα $\left[\frac{1}{\alpha_1} - \theta, \frac{1}{\beta_2} - \theta\right]$ πράγμα το οποίο είναι άτοπο. Εάν $r_C \in \left(\frac{1}{\alpha_1} - \theta, \frac{1}{\beta_2} - \theta\right)$ τότε ό λόγος $\frac{G}{g}$ είναι γνήσια φθίνουσα συνάρτηση στο διάστημα $\left[\frac{1}{\alpha_1} - \theta, r_C\right]$ και γνήσια αύξουσα στο διάστημα $\left[r_C, \frac{1}{\beta_2} - \theta\right]$. Συνεπώς $r^* = r_C$. Εάν $r_C > \frac{1}{\beta_2} - \theta$, τότε καθώς το $p(r_C)$ φθίνει στο διάστημα $\left(\frac{1}{\alpha_1} - \theta, \frac{1}{\beta_2} - \theta\right)$, έχουμε ότι $r* = \frac{1}{\beta_2} - \theta$.
Υποπερίπτωση $C_2$ : $\gamma_C < 0$. Σ′ αυτήν την περίπτωση και οι τρεις συντελεστές στη σχέση (6.1.47) είναι αρνητικοί με αποτέλεσμα και το $p(r)$ να είναι επίσης αρνητικό, γεγονός το οποίο συνεπάγεται ότι $r^* = \frac{1}{\beta_2} - \theta$.

*Παρατήρηση* 6.1.5. Θα πρέπει να τονίσουμε εδώ ότι η περίπτωση $\delta_C < 0$ **δεν είναι δυνατόν** να υπάρχει στη διακριτοποίηση των 9−σημείων. Αυτό δικαιολογείται ως εξής. Εάν κάτι τέτοιο συνέβαινε, τότε από το γεγονός ότι $\delta_C < 0$ θα είχαμε ισοδύναμα $\alpha_1\beta_2 < \alpha_2\beta_1$. Θέτοντας $4\sqrt{\frac{a}{b}}\frac{h_2}{h_1}\sin^2\frac{\pi}{2(n_1+1)}$, $4\sqrt{\frac{a}{b}}\frac{h_2}{h_1}\cos^2\frac{\pi}{2(n_1+1)}$, $4\sqrt{\frac{b}{a}}\frac{h_1}{h_2}\sin^2\frac{\pi}{2(n_2+1)}$, $4\sqrt{\frac{b}{a}}\frac{h_1}{h_2}\cos^2\frac{\pi}{2(n_2+1)}$ για $\alpha_1$, $\beta_1$, $\alpha_2$, $\beta_2$, αντίστοιχα, από την παραπάνω ανισότητα θα έχουμε ότι $\tan\frac{\pi}{2(n_1+1)} < \tan\frac{\pi}{2(n_2+1)}$, από την οποία συνεπάγεται ότι $n_2 < n_1$. Από τη συνθήκη διάταξης όμως έχουμε ότι $\beta_2 < \alpha_1$. Κάνοντας τις ίδιες αντικαταστάσεις λαμβάνουμε

$$\cos^2\frac{\pi}{2(n_2+1)} \;<\; \frac{a}{b}\frac{h_2^2}{h_1^2}\sin^2\frac{\pi}{2(n_1+1)}. \tag{6.1.52}$$

Η μικρότερη τιμή του αριστερού μέλους, λαμβάνεται για $n_2 = 2$ και η μεγαλύτερη τιμή του δεξιού μέλους για $\frac{a}{b}\frac{h_2^2}{h_1^2} = 5$. Αυτό προέρχεται από τον περιορισμό της ύπαρξης θετικά ορισμένου πίνακα $A$ (6.1.23) και από το $n_1 = 3$. Αλλά τότε το αριστερό μέλος γίνεται $\frac{3}{4}$ και το δεξιό μέλος $5 \cdot \frac{2-\sqrt{2}}{4}$ το οποίο μας δίνει, ισοδύναμα, ότι $50 < 49$, το οποίο είναι άτοπο.

Στη συνέχεια εξετάζουμε την περίπτωση όπου $\delta_C > 0$. Από τη συνθήκη αυτή έχουμε ότι

$$\frac{\beta_2 - \alpha_2}{\beta_1 - \alpha_1} > \frac{\beta_2}{\beta_1} > \frac{\alpha_2}{\alpha_1} > \frac{\beta_2\alpha_2}{\beta_1\alpha_1}, \tag{6.1.53}$$

όπου η τελευταία δεξιά ανισότητα προέρχεται από το γεγονός ότι οι δύο λόγοι στο μέσον είναι γνήσια μικρότεροι του 1. Η γνήσια ανισότητα μεταξύ των δύο ακραίων λόγων στις σχέσεις (6.1.53) δίνει ότι $\gamma_C > 0$. Έτσι, θα πρέπει να μελετηθούν πάλι δύο περιπτώσεις .
Υποπερίπτωση $C_3$ : $\varepsilon_C > 0$. Εφόσον και οι τρεις συντελεστες $\gamma_C$, $\delta_C$, $\varepsilon_C$ είναι θετικοί τότε και το $p(r)$ είναι επίσης θετικό, επομένως $r^* = \frac{1}{\alpha_1} - \theta$.
Υποπερίπτωση $C_4$ : $\varepsilon_C < 0$. Στην περίπτωση αυτή έχουμε να διακρίνουμε υποπεριπτώσεις ανάλογα με τη θέση του $r_C$ σε σχέση με τα άκρα των διαστημάτων που θεωρήσαμε. Σημειώνουμε ότι η θετική ρίζα του τριωνύμου $p(r) = 0$, $r_C$, αυτή τη φορά θα πρέπει να ικανοποιεί την $r_C < \frac{1}{\alpha_1} - \theta$ και άρα το $p(r)$ θα είναι επίσης θετικό. Συνεπώς, $r^* = \frac{1}{\alpha_1} - \theta$. Εάν το $r_C \in \left(\frac{1}{\alpha_1} - \theta, \frac{1}{\beta_2} - \theta\right)$ τότε θα έχουμε ότι $r^* = r_C$, ενώ στην περίπτωση όπου $r_C > \frac{1}{\beta_2} - \theta$ έχουμε ότι $r^* = \frac{1}{\beta_2} - \theta$.

*Παρατήρηση* 6.1.6. Σημειώνουμε ότι για το σχήμα των 9−σημείων, η περίπτωση για την οποία $\delta_C > 0$ **δεν** μπορεί να συμβεί **εκτός** από την περίπτωση όπου $n_1 = 2$ και $n_2 \geq 3$. Για να αποδείξουμε τον ισχυρισμό αυτόν εργαζόμαστε με τον ίδιο τρόπο που εργαστήκαμε στην Παρατήρηση 6.1.5 για την περίπτωση όπου ήταν $\delta_C < 0$. Σ′ αυτήν την περίπτωση έχουμε $n_2 > n_1$. Θεωρώντας ξανά τη συνθήκη διάταξης, που αναφέρεται στην περίπτωση όπου $\beta_2 < \alpha_1$, ολοκληρώνουμε την απόδειξη με τη σχέση (6.1.52). Η μεγαλύτερη τιμή του δεξιού μέλους λαμβάνεται για $\frac{a}{b}\frac{h_2^2}{h_1^2} = 5$ και για $n_1 = 2$ ενώ η ελάχιστη του αριστερού μέλους για $n_2 = 3$. Γι′ αυτές τις τιμές έχουμε ότι $\frac{2+\sqrt{2}}{4} < 5 \cdot \frac{1}{4}$ ή $\sqrt{2} < 3$, το οποίο αληθεύει. Έτσι η $\beta_2 < \alpha_1$ είναι επίσης αληθής για κάθε $n_2 \geq 3$ δίνοντας $n_1 = 2$. Για $n_1 = 3$ και $n_2 = 4$ η συνθήκη, (6.1.23) για το θετικά ορισμένο του πίνακα $A$, δίνει ότι $\cos^2\frac{\pi}{10} < 5 \cdot \sin^2\frac{\pi}{8}$ ή $\left(\frac{\sqrt{10+2\sqrt{5}}}{4}\right)^2 < 5 \cdot \left(\frac{\sqrt{5}-1}{4}\right)^2$ ή, ισοδύναμα, $720 < 400$, το οποίο δεν ισχύει.

Όλα τα παραπάνω αποτελέσματα συνοψίζονται στο επόμενο θεώρημα.

**Θεώρημα 6.1.2.** *Με τους μέχρι τώρα συμβολισμούς και τις υποθέσεις, η βέλτιστη παράμετρος επιτάχυνσης $r = r^*$ και στις τρεις Περιπτώσεις $A$, $B$, $C$, της σχέσης (6.1.40), σύμφωνα με το Θεώρημα 6.1.1, βρίσκονται με την ελαχιστοποίηση του λόγου $\frac{G}{g}$, όπου οι εκφράσεις για τα $G$ και $g$ δίνονται πάντα από τις μεσαίες εκφράσεις των Πινάκων 6.2, 6.3, 6.4, αντίστοιχα. Οι ακριβείς τιμές του $r^*$ σε κάθε περίπτωση δίνονται αναλυτικά στον Πίνακα 6.5.*

| Περίπτωση | | | | $r^*$ |
|---|---|---|---|---|
| $A$ | | | | ${r_{AB}}^1$ |
| $B$ | | | | $r_{AB}$ |
| $C$ | ${\delta_C}^2 < 0$ | ${\gamma_C}^2 > 0$ | ${r_C}^3 \in (\frac{1}{\alpha_1} - \theta, \frac{1}{\beta_2} - \theta)$ | $r_C$ |
| | | | $r_C \in [\frac{1}{\beta_2} - \theta, +\infty)$ | $\frac{1}{\beta_2} - \theta$ |
| | | $\gamma_C < 0$ | | $\frac{1}{\beta_2} - \theta$ |
| | $\delta_C > 0$ | ${\varepsilon_C}^2 > 0$ | | $\frac{1}{\alpha_1} - \theta$ |
| | | $\varepsilon_C < 0$ | $r_C \in (0, \frac{1}{\alpha_1} - \theta]$ | $\frac{1}{\alpha_1} - \theta$ |
| | | | $r_C \in (\frac{1}{\alpha_1} - \theta, \frac{1}{\beta_2} - \theta)$ | $r_C$ |
| | | | $r_C \in [\frac{1}{\beta_2} - \theta, +\infty)$ | $\frac{1}{\beta_2} - \theta$ |

Πίνακας 6.5: Βέλτιστη Παράμετρος Επιτάχυνσης $r^*$
(${}^1r_{AB}$ δίνεται από την (6.1.46). ${}^2$Οι συντελεστές $\gamma_C$, $\delta_C$, $\varepsilon_C$ δίνονται στην (6.1.49). ${}^3r_C$ δίνεται από την (6.1.51).)

### 6.1.5 Άλλες Δυνατές Περιπτώσεις

Στην προηγούμενη παράγραφο εξετάσαμε την περίπτωση όπου $\theta = 0$ και εκείνη για την οποία $\theta = \theta^* < \frac{1}{\beta_i}$, $i = 1, 2$. Όπως θα δούμε άλλες δυνατές περιπτώσεις που μπορούν να υπάρξουν είναι **μόνον** οι δύο επόμενες: $\frac{1}{\beta_i} < \theta^* < \frac{1}{\beta_j}$, $i \neq j = 1, 2$. Αρχικά, μπορούμε να ελέγξουμε ότι δεν είναι δυνατόν να έχουμε $\frac{1}{\alpha_i} < \theta^*$, $i = 1$ ή 2. Εάν συνέβαινε κάτι τέτοιο τότε, για $i = 1$, αντικαθιστούμε την τιμή του $\theta^*$ από τη σχέση (6.2.3) και τη μικρότερη ιδιοτιμή του $A_1$, που είναι η $\alpha_1 = \sqrt{\frac{a}{b}}\frac{h_2}{h_1}\sin^2\left(\frac{\pi}{2(n_1+1)}\right)$, στη σχέση (6.1.25). Θα έχουμε ισοδύναμα ότι $\frac{1}{3}\left(\frac{ah_2^2}{bh_1^2} + 1\right)\sin^2\left(\frac{\pi}{2(n_1+1)}\right) > 1$. Όμως, οι μεγαλύτερες τιμές για $\frac{ah_2^2}{bh_1^2}$ και $\sin^2\left(\frac{\pi}{2(n_1+1)}\right)$ είναι η 5 και η $\frac{1}{4}$, αντίστοιχα. Η πρώτη προκύπτει από το γεγονός ότι ο πίνακας $A$, του σχήματος των 9−σημείων, είναι θετικά ορισμένος, δηλαδή

από τη συνθήκη (6.1.23), και η δεύτερη λαμβάνεται για $n_1 = 2$. Έτσι έχουμε ότι $\frac{1}{3}(5+1)\frac{1}{4} > 1$ το οποίο είναι άτοπο. Η δεύτερη περίπτωση προκύπτει, εάν $\frac{1}{\beta_i} < \theta^*,\ i = 1, 2$. Τότε έχουμε ότι για $i = 1$, $1 < \frac{1}{3}\left(\frac{ah_2^2}{bh_1^2} + 1\right)\cos^2\left(\frac{\pi}{2(n_1+1)}\right)$ και για $i = 2$, $1 < \frac{1}{3}\left(\frac{bh_1^2}{ah_2^2} + 1\right)\cos^2\left(\frac{\pi}{2(n_2+1)}\right)$. Λύνοντας την πρώτη ως προς $\frac{ah_2^2}{bh_1^2}$ και τη δεύτερη ως προς $\frac{bh_1^2}{ah_2^2}$ θα έχουμε ότι

$$\frac{ah_2^2}{bh_1^2} > \frac{3}{\cos^2\left(\frac{\pi}{2(n_1+1)}\right)} - 1 \quad \text{και} \quad \frac{bh_1^2}{ah_2^2} > \frac{3}{\cos^2\left(\frac{\pi}{2(n_2+1)}\right)} - 1,$$

αντίστοιχα. Εφόσον τα δεξιά μέλη των δύο παραπάνω ανισοτήτων βρίσκονται στο διάστημα $(2, 3]$, τα αριστερά τους μέλη θα είναι τουλάχιστον μεγαλύτερα από 2. Όμως εάν κάποιο από αυτά ανήκει στο διάστημα $(2, 5]$, το άλλο θα ανήκει στο διάστημα $[\frac{1}{5}, \frac{1}{2})$. Έτσι **δε** θα ικανοποιούν την αντίστοιχη ανισότητα.

Από τις δύο δυνατές περιπτώσεις θα εξετάσουμε συνοπτικά αυτή για την οποία $\frac{1}{\beta_1} < \theta^* < \frac{1}{\beta_2}$, που αντιστοιχεί σε μία εκ των δύο περιπτώσεων διάταξης $A$ ή $C$ των σχέσεων (6.1.40). Έστω ότι θεωρούμε την περίπτωση $A$. Προφανώς, ο Πίνακας 6.1 τώρα αλλάζει εξαιτίας του περιορισμού του $\beta_1$, αφού $\frac{df(l_3)}{dt} < 0$ $\forall\, r \in (0, \infty)$. Εξαιτίας αυτής της αλλαγής, ο Πίνακας 6.2 περιορίζεται έτσι ώστε το αριστερό του τμήμα να μην υπάρχει πια. Όλα τα άλλα αποτελέσματα στα υπόλοιπα τέσσερα διαστήματα του Πίνακα 6.2 παραμένουν τα ίδια όπως αυτά παρουσιάζονται στον Πίνακα 6.6.

| $r$ | $0$ | $\frac{1}{\beta_2} - \theta$ | $\frac{1}{\alpha_1} - \theta$ | $\frac{1}{\alpha_2} - \theta$ | $+\infty$ |
|---|---|---|---|---|---|
| $G$ | $f(\beta_1, \alpha_2)$ | $\max\{f(\alpha_1, \beta_2), f(\beta_1, \alpha_2)\}$ | $f(\beta_1, \alpha_2)$ | $f(\alpha_1, \alpha_2)$ | |
| $g$ | $f(\alpha_1, \alpha_2)$ | $\min\{f(\alpha_1, \alpha_2), f(\beta_1, \beta_2)\}$ | $f(\beta_1, \beta_2)$ | $f(\beta_1, \beta_2)$ | |

Πίνακας 6.6: Περίπτωση $A$ ($\alpha_2 < \alpha_1 < \beta_2 < \beta_1$): Μέγιστη $G$ και Ελάχιστη $g$ τιμή της $f$

Συνοψίζουμε τα όσα εξετάσαμε παραπάνω στο επόμενο θεώρημα:

**Θεώρημα 6.1.3.** *Με τους μέχρι τώρα συμβολισμούς και υποθέσεις, καθώς και τις επιπλέον υποθέσεις ότι $\frac{1}{\beta_1} < \theta^*$ για την Περίπτωση A (αντίστοιχα για την Περίπτωση C) να ισχύουν, τότε, οι τιμές $G$ και $g$ δίνονται από τον Πίνακα 6.2 χωρίς το αριστερό του τμήμα για το $r$, από τον Πίνακα 6.6 (αντίστοιχα από τον Πίνακα 6.4 χωρίς το αριστερό του διάστημα). Οι ακριβείς τιμές του $r^*$ σε κάθε περίπτωση δίνονται πάλι από αυτές του Πίνακα 6.5.*

## 6.2 Βέλτιστοι Διπαραμετρικοί EADI Προρρυθμιστές

### 6.2.1 Εισαγωγή

Στην προηγούμενη παράγραφο παρουσιάσαμε και μελετήσαμε την οικογένεια των μονοπαραμετρικών ADI Προρρυθμιστών για την μέθοδο των Συζυγών Κλίσεων(CG). Στην παράγραφο αυτή θα μελετήσουμε την περίπτωση των σταθερών διπαραμετρικών ADI Προρρυθμιστών. Με αυτόν τον τρόπο περιμένουμε ότι οι ADI προρρυθμιστές να καταστούν περισσότερο αποτελεσματικοί.

Θα υπενθυμήσουμε ότι στην περίπτωση που έχουμε ένα γραμμικό σύστημα $Au = c$, με $A \in \mathbb{C}^{n,n}$ Ερμιτιανό και θετικά ορισμένο, και $c \in \mathbb{C}^n$, η μέθοδος των Συζυγών Κλίσεων είναι η πλέον κατάλληλη για την επίλυσή του. Παρόλα αυτά εάν ο δείκτης κατάστασης του πίνακα $A$, ο οποίος εκφράζεται από τη σχέση $\kappa(A) = \frac{\lambda_{\max}(A)}{\lambda_{\min}(A)}$, με $\lambda_{\max}(A)$ και $\lambda_{\min}(A)$ να είναι η μέγιστη και η ελάχιστη ιδιοτιμή του πίνακα $A$, είναι μεγάλος τότε χρησιμοποιούμε έναν κατάλληλο προρρυθμιστή $M$, με $M \in \mathbb{C}^{n,n}$ Ερμιτιανό και θετικά ορισμένο, τέτοιον ώστε $\kappa(M^{-1}A) = \frac{\lambda_{\max}(M^{-1}A)}{\lambda_{\min}(M^{-1}A)} \ll \kappa(A)$ (βλ. [26]).

Στην ανάλυσή που θα ακολουθήσει, θεωρούμε και πάλι την εξίσωση Poisson

$$-au_{xx}(x,y) - bu_{yy}(x,y) = f(x,y),\ f \in C^2 \tag{6.2.1}$$

ορισμένη στο ορθογώνιο $\Omega := \{(x,y) \in \mathbb{R}^2 | 0 < x < l_1,\ 0 < y < l_2\}$, όπου η $u(x,y)$ είναι μία συνεχώς διαφορίσιμη συνάρτηση με Dirichlet συνοριακές συνθήκες $u(x,y) = g(x,y)$ στο $\partial\Omega$, και $a$ και $b$ είναι θετικές σταθερές. Θεωρώντας ένα ομοιόμορφο διαμερισμό βήματος $h_1$ και $h_2$ στη $x-$ και $y-$διεύθυνση, αντίστοιχα, στο $\overline{\Omega} := \Omega \cup \partial\Omega$ προσεγγίζουμε την εξίσωση (6.2.1) σε κάθε κόμβο του πλέγματος από το σχήμα των διαφορών

$$\begin{aligned} & \sqrt{\frac{a}{b}}\frac{h_2}{h_1}(-u_{i-1,j} + 2u_{ij} - u_{i+1,j}) + \sqrt{\frac{b}{a}}\frac{h_1}{h_2}(-u_{i,j-1} + 2u_{ij} - u_{i,j+1}) \\ - \quad & \theta\,[4u_{ij} - 2(u_{i-1,j} + u_{i+1,j} + u_{i,j-1} + u_{i,j+1}) \\ + \quad & u_{i-1,j-1} + u_{i+1,j-1} + u_{i-1,j-1} + u_{i+1,j+1}] = \frac{h_1 h_2}{\sqrt{ab}}(f_{ij} + \phi_{ij}). \end{aligned} \tag{6.2.2}$$

(*Σημ.:* Είναι φυσικό να θεωρήσουμε ότι $\sin(\pi h_i) < \cos(\pi h_i)$, $i = 1, 2$, αφού πρέπει να έχουμε υπόψη μας ότι $h_i \to 0$, $i = 1, 2$.) Οι παράμετροι $\theta$ και $\phi$ στη

σχέση (6.2.2) λαμβάνουν τις τιμές

$$(\theta,\phi) \;=\; \begin{cases} (0,0), \\ (\theta^*,\phi^*) = \left(\frac{1}{12}(\sqrt{\frac{a}{b}}\frac{h_2}{h_1} + \sqrt{\frac{b}{a}}\frac{h_1}{h_2}), \frac{1}{12}(ah_1^2 f_{xx} + bh_2^2 f_{yy})\right), \end{cases} \tag{6.2.3}$$

όπου η (6.2.2), για $\theta = 0$, δίνει το σχήμα διαφορών των $5-$σημείων και για $\theta = \theta^*$ το σχήμα διαφορών των $9-$σημείων. Για άλλη μία φορά θα πρέπει να τονίσουμε ότι ο διακριτός τελεστής στο σχήμα των $9-$σημείων είναι θετικά ορισμένος εάν

$$\frac{1}{5} \;\le\; \frac{bh_1^2}{ah_2^2} \;\le\; 5 \tag{6.2.4}$$

(βλ. [48]), και έτσι $\theta^* \in [\frac{1}{6}, \frac{1}{2\sqrt{5}}]$. Υιοθετώντας μία φυσική διάταξη των κόμβων γραμμή προς γραμμή από κάτω προς τα πάνω κι από αριστερά προς τα δεξιά, αρχίζοντας από την κάτω αριστερή γωνία, το γραμμικό σύστημα που λαμβάνουμε είναι της μορφής

$$Au \;=\; c, \tag{6.2.5}$$

όπου, από τη γενική έκφραση (6.2.2), ο πίνακας $A$ μπορεί να γραφτεί ως εξής:

$$A = \sqrt{\frac{a}{b}}\frac{h_2}{h_1}(I_{n_2} \otimes T_{n_1}) + \sqrt{\frac{b}{a}}\frac{h_1}{h_2}(T_{n_2} \otimes I_{n_1}) - \theta\left[\sqrt{\frac{a}{b}}\frac{h_2}{h_1}(I_{n_2} \otimes T_{n_1}) \cdot \sqrt{\frac{b}{a}}\frac{h_1}{h_2}(T_{n_2} \otimes I_{n_1})\right]. \tag{6.2.6}$$

Στην (6.2.6) $n_1$ και $n_2$ είναι το πλήθος των εσωτερικών κόμβων στην $x-$ και $y-$διεύθυνση, αντίστοιχα, $T_{n_1} \in \mathbb{R}^{n_1 \times n_1}$ και $T_{n_2} \in \mathbb{R}^{n_2 \times n_2}$ είναι της μορφής $\mathrm{tridiag}(-1, 2, -1)$ και έτσι προκύπτει ότι είναι συμμετρικοί και θετικά ορισμένοι. Θέτοντας

$$A_1 := \sqrt{\frac{a}{b}}\frac{h_2}{h_1}(I_{n_2} \otimes T_{n_1}) \text{ και } A_2 := \sqrt{\frac{b}{a}}\frac{h_1}{h_2}(T_{n_2} \otimes I_{n_1}), \tag{6.2.7}$$

στη σχέση (6.2.6) έχουμε ότι

$$A = A_1 + A_2 - \theta A_1 A_2. \tag{6.2.8}$$

Τονίζεται ότι λόγω του πραγματικού, συμμετρικού και θετικά ορισμένου των πινάκων $T_{n_1}$ και $T_{n_2}$ καθώς και των ιδιοτήτων του τανυστικού γινομένου, θα είναι και οι $A_1$ και $A_2$, πραγματικοί συμμετρικοί και θετικά ορισμένοι. Επιπλέον, με απλούς υπολογισμούς αποδεικνύεται ότι οι $A_1$ και $A_2$ αντιμετατίθενται (βλ. [36]).

### 6.2.2 Διπαραμετρικό EADI Σχήμα

Παραλλάσσοντας κατά τι το σχήμα του Guittet λαμβάνουμε τη Διπαραμετρική EADI μέθοδο την οποία και θα μελετήσουμε σ΄ αυτήν την παράγραφο. Ειδικότερα, το προαναφερθέν διπαραμετρικό σχήμα είναι της μορφής:

$$\begin{aligned}(I + r_1A_1)u^{(m+\frac{1}{2})} &= [(I + r_2A_2)(I + r_1A_1) - \omega A]u^{(m)} + \omega c, \quad (6.2.9)\\ (I + r_2A_2)u^{(m+1)} &= u^{(m+\frac{1}{2})},\end{aligned}$$

όπου οι πίνακες $A, A_1, A_2$ δίνονται από τις σχέσεις (6.2.8) και (6.2.7), $r_1, r_2 > 0$ είναι οι παράμετροι επιτάχυνσης και $\omega$ είναι η παράμετρος παρεκβολής. Απαλείφοντας το διάνυσμα $u^{(m+\frac{1}{2})}$ από τις σχέσεις (6.2.9) λαμβάνουμε το επόμενο επαναληπτικό σχήμα.

$$u^{(m+1)} = T_{EADI}u^{(m)} + c_{EADI}, \tag{6.2.10}$$

όπου

$$T_{EADI} = I-\omega(I+r_1A_1)^{-1}(I+r_2A_2)^{-1}A, \quad c_{EADI} = (I+r_1A_1)^{-1}(I+r_2A_2)^{-1}\omega c. \tag{6.2.11}$$

Θεωρώντας στη συνέχεια ότι οι ιδιοτιμές $\lambda_i$ των πινάκων $A_i$, $i = 1, 2$, ανήκουν στο ορθογώνιο

$$S := \{\lambda_1, \lambda_2 \in \mathbb{R}_+ | \alpha_1 \leq \lambda_1 \leq \beta_1,\, \alpha_2 \leq \lambda_2 \leq \beta_2\},$$

όπου $\alpha_i,\, \beta_i \in \mathbb{R}_+$, $i = 1, 2$, τότε εξαιτίας του γεγονότος ότι οι πίνακες $A_1$ και $A_2$ αντιμετατίθενται και άρα μπορούν να έχουν κοινό σύστημα ιδιοδιανυσμάτων, οι ιδιοτιμές του $T_{EADI}$ δίνονται από τις εκφράσεις

$$\lambda_{T_{EADI}} = 1 - \omega\frac{\lambda_1 + \lambda_2 - \theta\lambda_1\lambda_2}{(1 + r_1\lambda_1)(1 + r_2\lambda_2)}. \tag{6.2.12}$$

Ορίζοντας το κλάσμα της σχέσης (6.2.12) με $f$ ως συνάρτηση των $\lambda_1$ και $\lambda_2$, έχουμε

$$f \equiv f(\lambda_1, \lambda_2) := \frac{\lambda_1 + \lambda_2 - \theta\lambda_1\lambda_2}{(1 + r_1\lambda_1)(1 + r_2\lambda_2)}. \tag{6.2.13}$$

Σημειώνουμε εδώ ότι λόγω του θετικά ορισμένου πίνακα $A$ στη σχέση (6.2.4) ο αριθμητής της $f$ είναι θετικός. Από τις σχέσεις (6.2.12) και (6.2.13) λαμβάνουμε

$$\rho(T_{EADI}) \leq \sup_{\lambda_1,\lambda_2\in S} |1 - \omega f|. \tag{6.2.14}$$

Για να υπολογίσουμε τη μέγιστη και την ελάχιστη τιμή της $f$, τις οποίες συμβολίζουμε με

$$G := \max_{\lambda_1,\lambda_2 \in S} f \quad \text{και} \quad g := \min_{\lambda_1,\lambda_2 \in S} f, \tag{6.2.15}$$

βρίσκουμε την $\frac{\partial f}{\partial \lambda_i},\ i = 1, 2$, από την οποία λαμβάνουμε την

$$\frac{\partial f}{\partial \lambda_i} = \frac{\lambda_j \left( (\frac{1}{\lambda_j} - \theta) - r_i \right)}{(1 + r_i \lambda_i)^2 (1 + r_j \lambda_j)},\ i \neq j = 1, 2. \tag{6.2.16}$$

Παρατηρούμε, λοιπόν, ότι ο αριθμητής των παραπάνω εκφράσεων είναι ανεξάρτητος από τη μεταβλητή $\lambda_i$ ως προς την οποία παραγωγίζουμε. Έτσι, η μέγιστη $G$ και η ελάχιστη $g$ τιμή της συνάρτησης $f$ λαμβάνονται στις κορυφές του ορθογωνίου $S$. Άρα οι τιμές αυτές πρέπει να βρίσκονται μεταξύ των $f(\alpha_1, \alpha_2)$, $f(\alpha_1, \beta_2)$, $f(\beta_1, \alpha_2)$ και $f(\beta_1, \beta_2)$, ακριβώς ανάλογα με την μονοπαραμετρική περίπτωση που μελετήσαμε στην προηγούμενη παράγραφο.

Είναι προφανές ότι για τη επίλυση του συστήματος (6.2.5) με χρήση των EADI μεθόδων (6.2.9), χρησιμοποιούμε τον προρρυθμιστή

$$M = \frac{1}{\omega}(I + r_2 A_2)(I + r_1 A_1). \tag{6.2.17}$$

Γι΄ αυτόν το λόγο οι βέλτιστες τιμές για τις παραμέτρους $r_1$, $r_2$, $\omega$, τις οποίες συμβολίζουμε με $r_1^*$, $r_2^*$, $\omega^*$ (βλ. [27]) μπορούν να βρεθούν ελαχιστοποιώντας το λόγο $\frac{G}{g}$. Έστω ότι $G^*$ και $g^*$ είναι οι αντίστοιχες βέλτιστες τιμές για τις $G$ και $g$. Τότε,

$$\frac{G^*}{g^*} = \min_{r_1, r_2 \in (0,\infty)} \frac{G}{g}. \tag{6.2.18}$$

Σ΄ αυτήν την περίπτωση η βέλτιστη τιμή για το $\omega$ θα δίνεται από τη

$$\omega^* = \frac{2}{G^* + g^*}. \tag{6.2.19}$$

Για τον αντίστοιχο Προρρυθμιστή της μεθόδου των Συζυγών Κλίσεων, ο βέλτιστος θα είναι αυτός για τον οποίο $\kappa(M^{-1}A) = \frac{\lambda_{\max}(M^{-1}A)}{\lambda_{\min}(M^{-1}A)}$ ελαχιστοποιείται. Παρόλα αυτά, η παράμετρος χαλάρωσης $\omega$ απλοποιείται και δεν επηρεάζει τον Προρρυθμιστή της μεθόδου των Συζυγών Κλίσεων, το οποίο συμβαίνει όταν ο λόγος $\frac{G}{g}$ ελαχιστοποιείται. Έτσι η λύση του βέλτιστου EADI προβλήματος θα δίνει επίσης και τη λύση για το βέλτιστου Προρρυθμιστή της μεθόδου των Συζυγών Κλίσεων. Ειδικότερα έχουμε αποδείξει ότι:

**Θεώρημα 6.2.1.** *Με βάση τους μέχρι τώρα συμβολισμούς και τις υποθέσεις, ο βέλτιστος (μικρότερος) δείκτης κατάστασης για το διακριτό πρόβλημα Poisson (6.2.5), χρησιμοποιώντας τον ADI Προρρυθμιστή (6.2.17) για τη μέθοδο των Συζυγών Κλίσεων, βρίσκεται από τις βέλτιστες τιμές των παραμέτρων επιτάχυνσης $r_1 = r_1^*$ και $r_2 = r_2^*$, οι οποίες βελτιστοποιούν το αντίστοιχο EADI πρόβλημα (6.2.18). Συνεπώς ο βέλτιστος δείκτης κατάστασης δίνεται από την έκφραση*

$$\kappa^*(M^{-1}A) \;=\; \frac{G^*}{g^*}. \tag{6.2.20}$$

### 6.2.3 Προσδιορισμός των Εκφράσεων $G$ και $g$

Για την απλοποίηση της ανάλυσης, θεωρούμε ότι οι λόγοι $\frac{1}{\beta_1},\ \frac{1}{\beta_2} > \theta$ έτσι ώστε

$$0 \;<\; \frac{1}{\beta_i} - \theta \;\;<\; \frac{1}{\alpha_i} - \theta, \;\; i = 1, 2. \tag{6.2.21}$$

Προφανώς, η αριστερή ανισότητα (6.2.21) ικανοποιείται για $\theta = 0$ ενώ δεν ισχύει για κάθε τιμή του $\theta = \theta^*$. Οι περιπτώσεις που μπορεί να προκύψουν όταν η προαναφερθείσα σχέση (6.2.21) δεν ικανοποιείται θα εξεταστούν στην Παράγραφο 6.2.5. Έστω ότι $V_{\alpha_1\alpha_2}, V_{\beta_1\alpha_2}, V_{\beta_1\beta_2}, V_{\alpha_1\beta_2}$ είναι οι τέσσερις κορυφές του ορθογωνίου $S$. Αφού οι ακραίες τιμές της συνάρτησης $f$ λαμβάνονται στις κορυφές του $S$, για να βρούμε τις τιμές αυτές, παραγωγίζουμε την $f$ κατά μήκος κάθε πλευράς. Σε αυτή την διπαραμετρική περίπτωση όπως και στην μονοπαραμετρική περίπτωση η παραμετρικοποίηση των πλευρών του ορθογωνίου δίνεται από τις εκφράσεις (6.1.41) και οι εκφράσεις των σύνθετων παραγώγων δίνονται παρακάτω

$$\frac{df(l_i(t))}{dt} = \nabla f \cdot \frac{dl_i(t)}{dt} = \frac{\lambda_i \left( (\frac{1}{\lambda_i} - \theta) - r_j \right)}{(1 + r_j\lambda_j)^2(1 + r_i\lambda_i)} (\beta_j - \alpha_j),\ i, j = 1, 2,\ i \neq j \tag{6.2.22}$$

Από τις παραπάνω εκφράσεις παρατηρούμε ότι το πρόσημο της παραγώγου εξαρτάται από την έκφραση του αριθμητή $(\frac{1}{\lambda_i} - \theta) - r_j$, $i, j = 1, 2$ $i \neq j$. Έχουμε λοιπόν ότι σε αυτή την περίπτωση οι παράγωγοι σε κάθε περίπτωση είναι θετικές όταν το $r_j < \frac{1}{\lambda_i} - \theta$. Σε κάθε περίπτωση αναλυτικά τα πρόσημα των αντίστοιχων μερικών παραγώγων εμφανίζονται στον Πίνακα 6.7.

Έτσι για παράδειγμα, στην περίπτωση $r_i \in \left(0, \frac{1}{\beta_j} - \theta\right]$, $i \neq j = 1, 2$, βρίσκεται εύκολα από τον Πίνακα 6.7 ότι η $f$ αυξάνει κατά μήκος των πλευρών

| $r_1$ | $\frac{1}{\beta_2}-\theta$ | |
|---|---|---|
| $\frac{\partial f(V_{\alpha_1\beta_2}\overrightarrow{V_{\beta_1\beta_2}})}{\partial r_1}$ | $+$ | $-$ |

| $r_2$ | $\frac{1}{\alpha_1}-\theta$ | |
|---|---|---|
| $\frac{\partial f(V_{\alpha_1\alpha_2}\overrightarrow{V_{\alpha_1\beta_2}})}{\partial r_2}$ | $+$ | $-$ |

| $r_1$ | $\frac{1}{\alpha_2}-\theta$ | |
|---|---|---|
| $\frac{\partial f(V_{\alpha_1\alpha_2}\overrightarrow{V_{\beta_1\alpha_2}})}{\partial r_1}$ | $+$ | $-$ |

| $r_2$ | $\frac{1}{\beta_1}-\theta$ | |
|---|---|---|
| $\frac{\partial f(V_{\beta_1\alpha_2}\overrightarrow{V_{\beta_1\beta_2}})}{\partial r_2}$ | $+$ | $-$ |

Πίνακας 6.7: Τα πρόσημα των $\frac{\partial f}{\partial r_i}$, $i=1,2$, κατά μήκος των πλευρών του ορθογωνίου $S$.

$V_{\alpha_1\alpha_2}\overrightarrow{V_{\beta_1\alpha_2}}$ και $V_{\beta_1\alpha_2}\overrightarrow{V_{\beta_1\beta_2}}$ όπως και των $V_{\alpha_1\alpha_2}\overrightarrow{V_{\alpha_1\beta_2}}$ και $V_{\alpha_1\beta_2}\overrightarrow{V_{\beta_1\beta_2}}$. Σαν αποτέλεσμα έχουμε ότι $G=f(\beta_1,\beta_2)$ και $g=f(\alpha_1,\alpha_2)$. Αυτό φαίνεται στο κάτω αριστερά χωρίο του πρώτου τεταρτημορίου του επιπέδου των $r_1$, $r_2$ του Πίνακα 6.8. Με τον ίδιο ακριβώς τρόπο εξετάζονται και όλες οι άλλες οκτώ περιπτώσεις χρησιμοποιώντας το πρόσημο των μερικών παραγώγων από τον Πίνακα 6.7. Το τελικό αποτέλεσμα παρουσιάζεται στον Πίνακα 6.8.

| $r_2$ | | | |
|---|---|---|---|
| $+\infty$ | $G=f(\beta_1,\alpha_2)$<br>$g=f(\alpha_1,\beta_2)$<br>$A$ | $G=f(\beta_1,\beta_2)$<br>$g=f(\beta_1,\beta_2)$<br>$B$ | $G=f(\alpha_1,\alpha_2)$<br>$g=f(\beta_1,\beta_2)$ |
| $\frac{1}{\alpha_1}-\theta$ | $G=f(\beta_1,\alpha_2)$<br>$g=f(\alpha_1,\alpha_2)$<br>$D$ | $G=\max\{f(\alpha_1,\beta_2),f(\beta_1,\alpha_2)\}$<br>$g=\min\{f(\alpha_1,\alpha_2),f(\beta_1,\beta_2)\}$<br>$C$ | $G=f(\alpha_1,\beta_2)$<br>$g=f(\beta_1,\beta_2)$ |
| $\frac{1}{\beta_1}-\theta$ | $G=f(\beta_1,\beta_2)$<br>$g=f(\alpha_1,\alpha_2)$ | $G=f(\alpha_1,\beta_2)$<br>$g=f(\alpha_1,\alpha_2)$ | $G=f(\alpha_1,\beta_2)$<br>$g=f(\beta_1,\alpha_2)$ |
| $0$ | $\frac{1}{\beta_2}-\theta$ | $\frac{1}{\alpha_2}-\theta$ | $+\infty$ $r_1$ |

Πίνακας 6.8: Το Μέγιστο και το Ελάχιστο της $f$ σε κάθε χωρίο του πρώτου τεταρτημορίου του επιπέδου των $r_1$, $r_2$.

Η μόνη περίπτωση που χρειάζεται περαιτέρω διερεύνηση είναι αυτή όπου τα $(r_1, r_2)$ ανήκουν στο χωρίο

$$(r_1, r_2) \ \in \ ABCD \ := \ \left[\frac{1}{\beta_2} - \theta, \frac{1}{\alpha_2} - \theta\right] \times \left[\frac{1}{\beta_1} - \theta, \frac{1}{\alpha_1} - \theta\right], \quad (6.2.23)$$

την οποία θα μελετήσουμε αργότερα.

Ας εξετάσουμε τώρα πως συμπεριφέρεται ο λόγος $\frac{G}{g}$ όταν βρισκόμαστε σε κάποιο από τα οκτώ κελιά που ορίζονται στον Πίνακα 6.8, κινούμενοι σε κάθε μία από τις $r_1-$ ή $r_2-$διευθύνσεις διατηρώντας την άλλη σταθερή. Για παράδειγμα, έστω ότι θεωρούμε το κελί

$$r_1 \in \left(0, \frac{1}{\beta_2} - \theta\right], \quad r_2 \in \left[\frac{1}{\alpha_1} - \theta, \infty\right).$$

Τότε θα έχουμε

$$\frac{G}{g} \ = \ \frac{f(\beta_1, \alpha_2)}{f(\alpha_1, \beta_2)} \ = \ \frac{(\beta_1 + \alpha_2 - \theta\beta_1\alpha_2)\,(1 + \alpha_1 r_1)(1 + \beta_2 r_2)}{(\alpha_1 + \beta_2 - \theta\alpha_1\beta_2)\,(1 + \beta_1 r_1)(1 + \alpha_2 r_2)}.$$

Λόγω του θετικού προσήμου των σταθερών πρώτων παραγόντων στους όρους του κλάσματος, λαμβάνοντας μερικές παραγώγους πρώτα σε σχέση με το $r_1$ και έπειτα σε σχέση με το $r_2$ και εκμεταλλευόμενοι το συμβολισμό που εισαγάγαμε στη (6.1.43) έχουμε ότι

$$\frac{\partial\left(\frac{G}{g}\right)}{\partial r_1} \ \sim \ \frac{d}{dr_1}\left(\frac{1 + \alpha_1 r_1}{1 + \beta_1 r_1}\right) \ \sim \ \alpha_1 - \beta_1 < 0,$$

$$\frac{\partial\left(\frac{G}{g}\right)}{\partial r_2} \ \sim \ \frac{d}{dr_2}\left(\frac{1 + \beta_2 r_2}{1 + \alpha_2 r_2}\right) \ \sim \ \beta_2 - \alpha_2 > 0.$$

Το πρώτο αποτέλεσμα μας δείχνει ότι θεωρώντας σταθερό το $r_2$, ο λόγος $\frac{G}{g}$ αποτελεί φθίνουσα συνάρτηση του $r_1$. Άρα ελαχιστοποιείται όταν το $r_1$ λάβει τη μέγιστη δυνατή τιμή του, το οποίο συμβαίνει όταν $r_1 = \frac{1}{\beta_2} - \theta$. Από την άλλη μεριά το δεύτερο αποτέλεσμα μας δείχνει ότι ο θεωρούμενος λόγος είναι αύξουσα συνάρτηση του $r_2$, με αποτέλεσμα να ελαχιστοποιείται για $r_2 = \frac{1}{\alpha_1} - \theta$. Όπως καθίσταται φανερό ο συγκεκριμένος λόγος, για όλα τα ζεύγη $(r_1, r_2)$ του θεωρηθέντος χωρίου, ελαχιστοποιείται για το ζεύγος των συνιστωσών της κάτω δεξιά γωνίας του χωρίου, η οποία είναι αποτελεί την κορυφή $A$ του χωρίου $ABCD$ της σχέσης (6.2.23).

Παρόμοια μελέτη σε κάθε ένα από τα άλλα τρία ακραία χωρία (άνω δεξιά, κάτω δεξιά και κάτω αριστερά), που ορίζονται στον Πίνακα 6.8, οδηγεί στο συμπέρασμα ότι το ελάχιστο του λόγου $\frac{G}{g}$ λαμβάνει χώρα στα σημεία $B$, $C$, $D$, αντίστοιχα. Για τα τέσσερα χωρία που μοιράζονται κοινό σύνορο με το $ABCD$ τα τελικά συμπεράσματα είναι κάπως διαφορετικά. Για παράδειγμα, έστω ότι θεωρούμε το χωρίο

$$r_1 \in \left[\frac{1}{\beta_2} - \theta, \frac{1}{\alpha_2} - \theta\right], \quad r_2 \in \left[\frac{1}{\alpha_1} - \theta, \infty\right).$$

Λαμβάνοντας μερικές παραγώγους του λόγου

$$\frac{G}{g} = \frac{f(\beta_1, \alpha_2)}{f(\beta_1, \beta_2)} = \frac{(\beta_1 + \alpha_2 - \theta\beta_1\alpha_2)\,(1 + \beta_2 r_2)}{(\beta_1 + \beta_2 - \theta\beta_1\beta_2)\,(1 + \alpha_2 r_2)},$$

ως προς το $r_2$ μόνο, αφού ο λόγος που μας ενδιαφέρει, είναι ανεξάρτητος από το $r_1$, έχουμε ότι

$$\frac{\partial\left(\frac{G}{g}\right)}{\partial r_2} \sim \frac{d}{dr_2}\left(\frac{1 + \beta_2 r_2}{1 + \alpha_2 r_2}\right) \sim \beta_2 - \alpha_2 > 0.$$

Άρα ο θεωρούμενος λόγος $\frac{G}{g}$ αποτελεί αύξουσα συνάρτηση του $r_2$ και επομένως το ελάχιστό του λαμβάνεται για $r_2 = \frac{1}{\alpha_1} - \theta$, το οποίο παρουσιάζεται στο ευθύγραμμο τμήμα $AB$. Με τον ίδιο τρόπο εργαζόμενοι μπορούμε να βρούμε ότι η ελάχιστη τιμή του λόγου, σε κάθε ένα από τα τρία άλλα χωρία που συνορεύουν με το ορθογώνιο $ABCD$, λαμβάνεται στις πλευρές $BC$, $CD$, $DA$ του κεντρικού αυτού χωρίου (ορθογωνίου), αντίστοιχα.

Συνεπώς το τελικό συμπέρασμα που εξάγεται είναι ότι το ολικό ελάχιστο του λόγου $\frac{G}{g}$ λαμβάνεται σε κάποιο σημείο του ορθογωνίου $ABCD$. Θα πρέπει, λοιπόν, να βρεθεί καταρχάς ποια από τις δύο εκφράσεις σε κάθε ένα από τα άγκιστρα, που εμφανίζονται στο ορθογώνιο $ABCD$ του Πίνακα 6.8, δίνουν τις τιμές για τις $G$ και $g$. Για το σκοπό αυτό θεωρούμε τις διαφορές

$$Q(r_1, r_2) = f(\alpha_1, \beta_2) - f(\beta_1, \alpha_2) \quad \text{και} \quad q(r_1, r_2) = f(\beta_1, \beta_2) - f(\alpha_1, \alpha_2), \tag{6.2.24}$$

οπότε και θα προσπαθήσουμε να βρούμε το πρόσημο καθεμιάς.

Αρχίζουμε με την έκφραση

$$\begin{aligned}
Q(r_1,r_2) &= \frac{\alpha_1+\beta_2-\theta\alpha_1\beta_2}{(1+\alpha_1 r_1)(1+\beta_2 r_2)} - \frac{\beta_1+\alpha_2-\theta\beta_1\alpha_2}{(1+\beta_1 r_1)(1+\alpha_2 r_2)} \\
&\sim (\alpha_1+\beta_2-\theta\alpha_1\beta_2)(1+\beta_1 r_1)(1+\alpha_2 r_2) \\
&- (\beta_1+\alpha_2-\theta\beta_1\alpha_2)(1+\alpha_1 r_1)(1+\beta_2 r_2) \\
&\sim [(\alpha_1+\beta_2)-\theta\alpha_1\beta_2+(\alpha_1+\beta_2)\alpha_2 r_2-\theta\alpha_1\alpha_2\beta_2 r_2 \\
&+ (\alpha_1+\beta_2)\beta_1 r_1-\theta r_1\beta_1\alpha_1\beta_2+(\alpha_1+\beta_2)\beta_1\alpha_2 r_1 r_2] \\
&- [(\beta_1+\alpha_2)-\theta\beta_1\alpha_2+(\beta_1+\alpha_2)\beta_2 r_2-\theta\beta_2\beta_1\alpha_2 r_2 \\
&+ (\beta_1+\alpha_2)\alpha_1 r_1-\theta\alpha_1\beta_1\alpha_2 r_1+(\beta_1+\alpha_2)\alpha_1\beta_2 r_1 r_2] \\
&\sim r_1 r_2[\beta_2\alpha_2(\beta_1-\alpha_1)-\beta_1\alpha_1(\beta_2-\alpha_2)] \\
&+ (r_1-r_2)[\beta_1(\beta_2-\alpha_2)+\alpha_2(\beta_1-\alpha_1)] \\
&- \theta r_1\beta_1\alpha_1(\beta_2-\alpha_2)+\theta r_2\beta_2\alpha_2(\beta_1-\alpha_1) \\
&- \theta\beta_1(\beta_2-\alpha_2)+\theta\beta_2(\beta_1-\alpha_1)+(\beta_2-\alpha_2)-(\beta_1-\alpha_1) \qquad (6.2.25)
\end{aligned}$$

Θεωρώντας ότι ο συντελεστής του γινομένου $r_1 r_2$ είναι διαφορετικός του μηδενός, η συνάρτηση της σχέσης (6.2.25) παριστάνει γεωμετρικά ένα μονόχωνο υπερβολοειδές με καμπύλες στάθμης υπερβολές.

*Σημ.:* Εάν ο συντελεστής του γινομένου $r_1 r_2$ είναι μηδέν, που συμβαίνει εάν $\beta_2\alpha_2(\beta_1-\alpha_1)=\beta_1\alpha_1(\beta_2-\alpha_2)$, ή όταν $\frac{1}{\alpha_1}-\frac{1}{\beta_1}=\frac{1}{\alpha_2}-\frac{1}{\beta_2}$, έχουμε ισοδύναμα ότι

$$\frac{\cot(\pi h_1)}{\cot(\pi h_2)}\frac{\sin(\pi h_2)}{\sin(\pi h_1)} = \frac{a^2 h_2^4}{b^2 h_1^4},$$

όπως, για παράδειγμα, στην περίπτωση όπου $\alpha_2=\alpha_1$ και $\beta_2=\beta_1$. Τότε το υπερβολοειδές γίνεται επίπεδο και οι καμπύλες στάθμης είναι ευθείες γραμμές. Διαφορετικά, τίποτα δεν αλλάζει στην ανάλυση που θα ακολουθήσει. Έτσι, από εδώ και πέρα όταν χρησιμοποιούμε την έκφραση "υπερβολή" θα συμπεριλαμβάνουμε και την περίπτωση της ευθείας γραμμής.

Στη συνέχεια εξετάζουμε το πρόσημο της συνάρτησης $Q(r_1,r_2)$ σε κάθε κορυφή του ορθογωνίου $ABCD$. Στην κορυφή $A$ έχουμε ότι

$$Q\left(\frac{1}{\beta_2}-\theta,\frac{1}{\alpha_1}-\theta\right)$$

$$= \frac{\alpha_1+\beta_2-\theta\alpha_1\beta_2}{\left(1+\left(\frac{1}{\beta_2}-\theta\right)\alpha_1\right)\left(1+\left(\frac{1}{\alpha_1}-\theta\right)\beta_2\right)} - \frac{\beta_1+\alpha_2-\theta\beta_1\alpha_2}{\left(1+\left(\frac{1}{\beta_2}-\theta\right)\beta_1\right)\left(1+\left(\frac{1}{\alpha_1}-\theta\right)\alpha_2\right)}$$

$$= \alpha_1\beta_2\,\frac{\alpha_1+\beta_2-\theta\alpha_1\beta_2}{(\alpha_1+\beta_2-\theta\alpha_1\beta_2)^2} - \alpha_1\beta_2\frac{\beta_1+\alpha_2-\theta\beta_1\alpha_2}{(\beta_1+\beta_2-\theta\beta_1\beta_2)(\alpha_1+\alpha_2-\theta\alpha_2\alpha_1)}$$

$$\sim (\beta_1+\beta_2-\theta\beta_1\beta_2)(\alpha_1+\alpha_2-\theta\alpha_2\alpha_1)-(\alpha_1+\beta_2-\theta\alpha_1\beta_2)(\beta_1+\alpha_2-\theta\beta_1\alpha_2)$$
$$\sim\ -(\beta_1-\alpha_1)(\beta_2-\alpha_2)<0,$$

αυτό και διότι οι παρονομαστές είναι θετικοί από το θετικά ορισμένο των πινάκων $A_1$ και $A_2$. Στην κορυφή $B$ έχουμε ότι

$$Q\left(\frac{1}{\alpha_2}-\theta,\frac{1}{\alpha_1}-\theta\right)$$
$$= \frac{\alpha_1+\beta_2-\theta\alpha_1\beta_2}{\left(1+\left(\frac{1}{\alpha_2}-\theta\right)\alpha_1\right)\left(1+\left(\frac{1}{\alpha_1}-\theta\right)\beta_2\right)}-\frac{\beta_1+\alpha_2-\theta\beta_1\alpha_2}{\left(1+\left(\frac{1}{\alpha_2}-\theta\right)\beta_1\right)\left(1+\left(\frac{1}{\alpha_1}-\theta\right)\alpha_2\right)}$$
$$\sim \alpha_1\alpha_2\frac{\alpha_1+\beta_2-\theta\alpha_1\beta_2}{(\alpha_1+\beta_2-\theta\alpha_1\beta_2)(\alpha_1+\alpha_2-\theta\alpha_1\alpha_2)}-\alpha_1\alpha_2\frac{\beta_1+\alpha_2-\theta\beta_1\alpha_2}{(\beta_1+\alpha_2-\theta\beta_1\alpha_2)(\alpha_1+\alpha_2-\theta\alpha_1\alpha_2)}$$
$$=0.$$

Στην κορυφή $C$ έχουμε ότι

$$Q\left(\frac{1}{\alpha_2}-\theta,\frac{1}{\beta_1}-\theta\right)$$
$$= \frac{\alpha_1+\beta_2-\theta\alpha_1\beta_2}{\left(1+\left(\frac{1}{\alpha_2}-\theta\right)\alpha_1\right)\left(1+\left(\frac{1}{\beta_1}-\theta\right)\beta_2\right)}-\frac{\beta_1+\alpha_2-\theta\beta_1\alpha_2}{\left(1+\left(\frac{1}{\alpha_2}-\theta\right)\beta_1\right)\left(1+\left(\frac{1}{\beta_1}-\theta\right)\alpha_2\right)}$$
$$\sim \frac{\alpha_1+\beta_2-\theta\alpha_1\beta_2}{(\alpha_1+\alpha_2-\theta\alpha_1\alpha_2)(\beta_1+\beta_2-\theta\beta_1\beta_2)}-\frac{1}{\beta_1+\alpha_2-\theta\beta_1\alpha_2}$$

$$\sim (\alpha_1+\beta_2-\theta\alpha_1\beta_2)(\beta_1+\alpha_2-\theta\beta_1\alpha_2)-(\alpha_1+\alpha_2-\theta\alpha_1\alpha_2)(\beta_1+\beta_2-\theta\beta_1\beta_2)$$

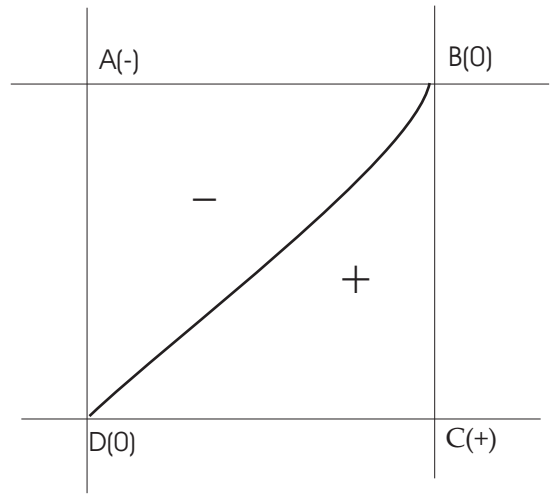
$$\sim\ (\beta_1-\alpha_1)(\beta_2-\alpha_2)>0.$$

Τέλος, στην κορυφή $D$ η τιμή της συνάρτησης είναι

$$Q\left(\frac{1}{\beta_2}-\theta,\frac{1}{\beta_1}-\theta\right)=$$
$$\frac{\alpha_1+\beta_2-\theta\alpha_1\beta_2}{\left(1+\left(\frac{1}{\beta_2}-\theta\right)\alpha_1\right)\left(1+\left(\frac{1}{\beta_1}-\theta\right)\beta_2\right)}-\frac{\beta_1+\alpha_2-\theta\beta_1\alpha_2}{\left(1+\left(\frac{1}{\beta_2}-\theta\right)\beta_1\right)\left(1+\left(\frac{1}{\beta_1}-\theta\right)\alpha_2\right)}$$
$$=\beta_1\beta_2\left(\frac{1}{\beta_1+\beta_2-\theta\beta_1\beta_2}-\frac{1}{\beta_1+\beta_2-\theta\beta_2\beta_1}\right)=0.$$

Βασιζόμενοι στα προηγούμενα αποτελέσματα, έχοντας δηλαδή ότι η συνάρτηση $Q(r_1,r_2)$ έχει αρνητική τιμή στην κορυφή $A$ μηδενική τιμή στις κορυφές

$B$ και $D$ και τέλος θετική στην κορυφή $C$, συνάγεται ότι η συνάρτηση $Q(r_1, r_2)$ είναι αρνητική αριστερά του τμήματος της υπερβολής $DB$, μηδενίζεται πάνω σ′ αυτό και είναι θετική δεξιά του, όπως φαίνεται στο Σχήμα 6.2. Αυτό απλά σημαίνει ότι από τις δύο εκφράσεις για την $G$ των οποίων η διαφορά είναι η $Q(r_1, r_2)$ στη σχέση (6.2.24), η πρώτη έκφραση, $f(\alpha_1, \beta_2)$, δίνει τη μέγιστη τιμή για την $f$ αριστερά του $DB$, η $f(\alpha_2, \beta_1)$ δίνει το αντίστοιχο μέγιστο δεξιά του $DB$, ενώ πάνω στο $DB$ οι δυο εκφράσεις ταυτίζονται.




Σχήμα 6.2: Πρόσημα της $Q(r_1, r_2)$ στο $ABCD$.

Για να βρούμε το ελάχιστο της $f$ στο $ABCD$ εργαζόμαστε με τον ίδιο ακριβώς τρόπο χρησιμοποιώντας τώρα τη διαφορά $q(r_1, r_2)$, που ορίζεται στη σχέση (6.2.24). Σ′ αυτή την περίπτωση έχουμε ότι

$$
\begin{aligned}
q(r_1,r_2) &= \frac{\beta_1+\beta_2-\theta\beta_1\beta_2}{(1+r_1\beta_1)(1+r_2\beta_2)} - \frac{\alpha_1+\alpha_2-\theta\alpha_1\alpha_2}{(1+r_1\alpha_1)(1+r_2\alpha_2)} \\
&\sim (\beta_1+\beta_2-\theta\beta_1\beta_2)(1+r_1\alpha_1)(1+r_2\alpha_2) \\
&- (\alpha_1+\alpha_2-\theta\alpha_1\alpha_2)(1+r_1\alpha_1)(1+r_2\alpha_2) \\
&= [(\beta_1+\beta_2)-\theta\beta_1\beta_2+r_2\alpha_2(\beta_1+\beta_2) \\
&- \theta r_2\alpha_2\beta_1\beta_2+(\beta_1+\beta_2)r_1\alpha_1 \\
&- \theta r_1\alpha_1\beta_1\beta_2+(\beta_1+\beta_2)r_1r_2\alpha_1\alpha_2] \\
&- [(\alpha_1+\alpha_2)-\theta\alpha_1\alpha_2 \\
&+ r_1\beta_1(\beta_1+\beta_2)+r_2\beta_2(\beta_1+\beta_2) \\
&- \theta r_1\beta_1\alpha_1\alpha_2-\theta r_2\beta_2\alpha_1\alpha_2+(\beta_1+\beta_2)r_1r_2\beta_1\beta_2] \\
&\sim -r_1r_2[\beta_2\alpha_2(\beta_1-\alpha_1)+\beta_1\alpha_1(\beta_2-\alpha_2)] \\
&+ (r_1-r_2)[\alpha_1(\beta_2-\alpha_2)-\alpha_2(\beta_1-\alpha_1)] \\
&- \theta r_1\beta_1\alpha_1(\beta_2-\alpha_2)-\theta r_2\beta_2\alpha_2(\beta_1-\alpha_1) \\
&- \theta\alpha_1(\beta_2-\alpha_2)-\theta\beta_2(\beta_1-\alpha_1)+(\beta_2-\alpha_2)+(\beta_1-\alpha_1) \quad (6.2.26)
\end{aligned}
$$

Είναι εύκολο να διαπιστώσουμε ότι το δεξιό μέλος παριστάνει γεωμετρικά πάντα ένα μονόχωνο υπερβολοειδές με τις καμπύλες στάθμης του να είναι πάντα υπερβολές αφού ο συντελεστής του γινομένου $r_1r_2$ είναι αρνητικός. Εξετάζουμε το πρόσημο της συνάρτησης $q(r_1,r_2)$, όπως αυτό έγινε πριν με την $Q(r_1,r_2)$, στις κορυφές του ορθογωνίου $ABCD$. Έχουμε λοιπόν τις ακόλουθες εκφράσεις. Για την κορυφή $A$ η τιμή της συνάρτησης $g$ είναι:

σμαλλ

$$
\begin{aligned}
q\left(\frac{1}{\beta_2}-\theta,\ \frac{1}{\alpha_1}-\theta\right) &= \frac{\beta_1+\beta_2-\theta\beta_1\beta_2}{(1+\left(\frac{1}{\beta_2}-\theta\right)\beta_1)(1+\left(\frac{1}{\alpha_1}-\theta\right)\beta_2)} \\
&- \frac{\alpha_1+\alpha_2-\theta\alpha_1\alpha_2}{(1+\left(\frac{1}{\beta_2}-\theta\right)\alpha_1)(1+\left(\frac{1}{\alpha_1}-\theta\right)\alpha_2)} \\
&= \alpha_1\beta_2\left(\frac{1}{\alpha_1+\beta_2-\theta\alpha_1\beta_2}-\frac{1}{\alpha_1+\beta_2-\theta\alpha_1\beta_2}\right)=0
\end{aligned}
$$

Έπειτα θα βρούμε διαδοχικά και τις τιμές της συνάρτησης και στις υπόλοιπες κορυφές $B, C, D$.

$$q\left(\frac{1}{\alpha_2}-\theta,\,\frac{1}{\alpha_1}-\theta\right)=\frac{\beta_1+\beta_2-\theta\beta_1\beta_2}{(1+\left(\frac{1}{\alpha_2}-\theta\right)\beta_1)(1+\left(\frac{1}{\alpha_1}-\theta\right)\beta_2)}$$

$$-\frac{\alpha_1+\alpha_2-\theta\alpha_1\alpha_2}{(1+\left(\frac{1}{\alpha_2}-\theta\right)\alpha_1)(1+\left(\frac{1}{\alpha_1}-\theta\right)\alpha_2)}$$

$$\sim(\beta_1+\beta_2-\theta\beta_1\beta_2)(\alpha_1+\alpha_2-\theta\alpha_1\alpha_2)-(\alpha_2+\beta_1-\theta\alpha_2\beta_1)(\alpha_1+\beta_2-\theta\alpha_1\beta_2)$$

$$=[(\beta_1+\beta_2)(\alpha_1+\alpha_2)-\theta\beta_1\beta_2(\alpha_1+\alpha_2)-\theta\alpha_1\alpha_2(\beta_1+\beta_2)]$$

$$-[(\alpha_2+\beta_1)(\alpha_1+\beta_2)-\theta\alpha_1\beta_2(\alpha_2+\beta_1)-\theta\alpha_2\beta_1(\alpha_1+\beta_2)$$

$$=-(\beta_2-\alpha_2)(\beta_1-\beta_2)<0$$

$$q\left(\frac{1}{\alpha_2}-\theta,\,\frac{1}{\beta_1}-\theta\right)$$

$$=\frac{\beta_1+\beta_2-\theta\beta_1\beta_2}{(1+\left(\frac{1}{\alpha_2}-\theta\right)\beta_1)(1+\left(\frac{1}{\beta_1}-\theta\right)\beta_2)}-\frac{\alpha_1+\alpha_2-\theta\alpha_1\alpha_2}{(1+\left(\frac{1}{\alpha_2}-\theta\right)\alpha_1)(1+\left(\frac{1}{\beta_1}-\theta\right)\alpha_2)}$$

$$\sim\left(\frac{1}{\alpha_2+\beta_1-\theta\alpha_2\beta_1}-\frac{1}{\alpha_2+\beta_1-\theta\alpha_2\beta_1}\right)=0$$

$$q\left(\frac{1}{\beta_2}-\theta,\,\frac{1}{\beta_1}-\theta\right)$$

$$=\frac{\beta_1+\beta_2-\theta\beta_1\beta_2}{(1+\left(\frac{1}{\beta_2}-\theta\right)\beta_1)(1+\left(\frac{1}{\beta_1}-\theta\right)\beta_2)}-\frac{\alpha_1+\alpha_2-\theta\alpha_1\alpha_2}{(1+\left(\frac{1}{\beta_2}-\theta\right)\alpha_1)(1+\left(\frac{1}{\beta_1}-\theta\right)\alpha_2)}$$

$$\sim\frac{1}{\beta_1+\beta_2-\theta\beta_1\beta_2}-\frac{\alpha_1+\alpha_2-\theta\alpha_1\alpha_2}{(\alpha_1+\beta_2-\theta\alpha_1\beta_2)(\beta_1+\alpha_2-\theta\alpha_2\beta_1)}$$

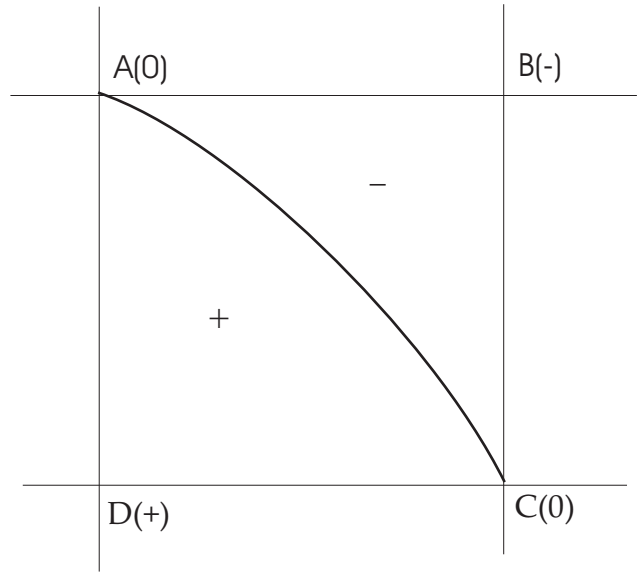
$$\sim(\alpha_1+\beta_2-\theta\alpha_1\beta_2)(\beta_1+\alpha_2-\theta\alpha_2\beta_1)-(\alpha_1+\alpha_2-\theta\alpha_1\alpha_2)(\beta_1+\beta_2-\theta\beta_1\beta_2)>0.$$

Η τελευταία έκφραση είναι η αντίθετη της έκφρασης που αφορά στην κορυφή $B$ επομένως έχει και αντίθετο πρόσημο.

Τα παραπάνω αποτελέσματα συνοψίζονται στον ακόλουθο πίνακα

$$\begin{cases} q(r_1, r_2) = 0 & \text{στην κορυφή } A, \\ q(r_1, r_2) < 0 & \text{στην κορυφή } B, \\ q(r_1, r_2) = 0 & \text{στην κορυφή } C, \\ q(r_1, r_2) > 0 & \text{στην κορυφή } D. \end{cases}$$

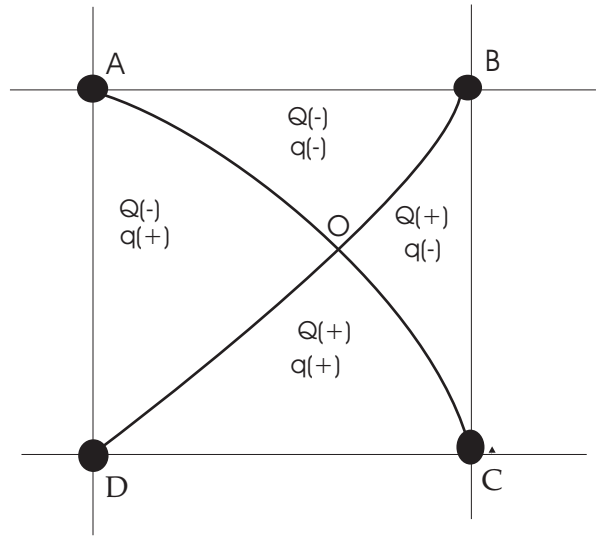
Έτσι, η συνάρτηση $q(r_1, r_2)$ είναι αρνητική στα δεξιά του τμήματος της υπερβολής $AC$, μηδενίζεται πάνω σ΄ αυτήν και είναι θετική στα αριστερά αυτής (βλ. Σχήμα 6.3).




Σχήμα 6.3: Πρόσημα της $q(r_1, r_2)$ στο $ABCD$.

Εάν συγκρίνουμε τα αποτελέσματα που παίρνουμε από τα πρόσημα των δύο συναρτήσεων $Q(r_1, r_2)$ και $q(r_1, r_2)$, μπορούμε να διαπιστώσουμε ότι το ορθογώνιο $ABCD$ χωρίζεται από τα τμήματα των δύο υπερβολών σε τέσσερα μέρη, στα οποία παρουσιάζονται διαφορετικά ζεύγη από πρόσημα, και ως εκ τούτου διαφορετικές εκφράσεις για τη μέγιστη και την ελάχιστη τιμή της συνάρτησης $f$. Τα πρόσημα των δυο συναρτήσεων σε καθένα από τους τέσσερις τομείς στους οποίους χωρίζεται το ορθογώνιο $ABCD$, παρουσιάζονται στο Σχήμα 6.4.

Βασιζόμενοι στα πρόσημα των συναρτήσεων $Q$ και $q$, που ορίζονται στις σχέσεις (6.2.24), σε κάθε έναν από τους τέσσερις τομείς του $ABCD$ έχουμε τα επόμενα αποτελέσματα, θεωρώντας τη μέγιστη και την ελάχιστη τιμή των $G$




Σχήμα 6.4: Τα πρόσημα των $Q(r_1, r_2)$ και $q(r_1, r_2)$ στο ορθογώνιο $ABCD$.

και $g$, αντίστοιχα:

$$\begin{array}{lll} \text{Τομέας}(OAB) : & G = f(\beta_1, \alpha_2), & g = f(\beta_1, \beta_2), \\ \text{Τομέας}(OBC) : & G = f(\alpha_1, \beta_2), & g = f(\beta_1, \beta_2), \\ \text{Τομέας}(OCD) : & G = f(\alpha_1, \beta_2), & g = f(\alpha_1, \alpha_2), \\ \text{Τομέας}(ODA) : & G = f(\beta_1, \alpha_2), & g = f(\alpha_1, \alpha_2). \end{array} \tag{6.2.27}$$

### 6.2.4 Βέλτιστες Παράμετροι της EADI Μεθόδου

Έχοντας βρεί τις εκφράσεις για τις συναρτήσεις $G$ και $g$ στους τέσσερις τομείς του $ABCD$ στις σχέσεις (6.2.27) επιστρέφουμε στο βασικό μας, σκοπό όπως περιγράφηκε στην προηγούμενη παράγραφο, δηλαδή η ελαχιστοποίηση του λόγου $\frac{G}{g}$ σε κάθε τομέα χωριστά.

Έστω ότι θεωρούμε τον τομέα $OAB$, όπως αυτός φαίνεται στο Σχήμα 6.4. Θεωρούμε το λόγο $\frac{G}{g}$ που έχει την έκφραση

$$\frac{G}{g} = \frac{f(\beta_1, \alpha_2)}{f(\beta_1, \beta_2)} = \frac{(\beta_1 + \alpha_2 - \theta\beta_1\alpha_2)}{(\beta_1 + \beta_2 - \theta\alpha_1\alpha_2)} \cdot \frac{(1 + \beta_1 r_1)}{(1 + \beta_1 r_1)} \cdot \frac{(1 + \beta_2 r_2)}{(1 + \alpha_2 r_2)}.$$

Εφόσον το δεύτερο κλάσμα στην παραπάνω έκφραση είναι ίσο με 1, ο ζητούμενος λόγος εξαρτάται μόνο από το $r_2$. Έτσι, παίρνοντας τη μερική παράγωγο

ως προς αυτό βρίσκουμε ότι $\frac{\partial(\frac{G}{g})}{\partial r_2} \sim \beta_2 - \alpha_2 > 0$. Ο λόγος, λοιπόν, $\frac{G}{g}$ είναι αύξουσα συνάρτηση του $r_2$ και παίρνει την ελάχιστη τιμή, εφόσον είναι ανεξάρτητος από το $r_1$, στο "χαμηλότερο" σημείο του τομέα $OAB$, που είναι το σημείο $O$.

Θεωρώντας τώρα τον τομέα $OCD$ παίρνουμε το λόγο

$$\frac{G}{g} \;=\; \frac{f(\alpha_1,\beta_2)}{f(\alpha_1,\alpha_2)} \;=\; \frac{(\alpha_1+\beta_2-\theta\alpha_1\beta_2)}{(\alpha_1+\alpha_2-\theta\alpha_1\alpha_2)}\cdot\frac{(1+\alpha_1 r_1)}{(1+\alpha_1 r_1)}\cdot\frac{(1+\alpha_2 r_2)}{(1+\beta_2 r_2)}.$$

Η περίπτωση αυτή είναι αντίστοιχη με την προηγούμενη. Κι εδώ έχουμε ότι $\frac{\partial(\frac{G}{g})}{\partial r_2} \sim \alpha_2 - \beta_2 < 0$. Το συμπέρασμα είναι ότι ο λόγος $\frac{G}{g}$ είναι φθίνουσα συνάρτηση ως προς $r_2$ κι επόμενως η ελάχιστη τιμή του λαμβάνεται στο "ψηλότερο" σημείο του τομέα $OCD$, που είναι και πάλι το $O$.

Συγκρίνοντας το τελευταίο αποτέλεσμα με το προηγούμενο έχουμε ότι το ολικό ελάχιστο (βέλτιστο) του λόγου λαμβάνεται στο σημείο $O$. Έτσι η βέλτιστη λύση του προβλήματός μας δίνεται από τις συντεταγμένες $(r_1^*, r_2^*)$ του $O$. Από το τελευταίο μπορούμε εύκολα να βρούμε το βέλτιστο $\omega$, το $(\omega^*)$, και το βέλτιστο δείκτη κατάστασης $\kappa^*(M^{-1}A)$, που δίνονται από τις σχέσεις (6.1.38) και (6.1.39), αντίστοιχα.

*Σημ.:* Σημειώνουμε ότι εάν χρησιμοποιήσουμε τους άλλους δύο τομείς στους οποίους χωρίζεται το ορθογώνιο λαμβάνουμε ακριβώς το ίδιο αποτέλεσμα.

Φυσικά για να βρούμε τις συντεταγμένες του σημείου $O$ πρέπει να λύσουμε το σύστημα των δύο εξισώσεων $Q(r_1,r_2)=0$ και $q(r_1,r_2)=0$. Λαμβάνοντας τις δύο συναρτήσεις από τις σχέσεις (6.2.25) και (6.2.26) καταλήγουμε να έχουμε προς επίλυση το παρακάτω σύστημα

$$\begin{aligned}
[\beta_2\alpha_2(\beta_1-\alpha_1)-\beta_1\alpha_1(\beta_2-\alpha_2)]r_1r_2+[\beta_1(\beta_2-\alpha_2)+\alpha_2(\beta_1-\alpha_1)](r_1-r_2)&\\
-\theta\beta_1\alpha_1(\beta_2-\alpha_2)r_1+\theta\beta_2\alpha_2(\beta_1-\alpha_1)r_2&\\
-\theta\beta_1(\beta_2-\alpha_2)+\theta\beta_2(\beta_1-\alpha_1)+(\beta_2-\alpha_2)-(\beta_1-\alpha_1) \;&=\; 0,\\
-[\beta_2\alpha_2(\beta_1-\alpha_1)+\beta_1\alpha_1(\beta_2-\alpha_2)]r_1r_2+[\alpha_1(\beta_2-\alpha_2)-\alpha_2(\beta_1-\alpha_1)](r_1-r_2)&\\
-\theta\beta_1\alpha_1(\beta_2-\alpha_2)r_1-\theta\beta_2\alpha_2(\beta_1-\alpha_1)r_2&\\
-\theta\alpha_1(\beta_2-\alpha_2)-\theta\beta_2(\beta_1-\alpha_1)+(\beta_2-\alpha_2)+(\beta_1-\alpha_1) \;&=\; 0. \qquad (6.2.28)
\end{aligned}$$

Αθροίζοντας και αφαιρώντας κατά μέλη τις δύο εξισώσεις λαμβάνουμε το ισοδύναμο σύστημα

$$\begin{aligned}
-\beta_1\alpha_1 r_1r_2+\frac{1}{2}(\alpha_1+\beta_1)(r_1-r_2)-\theta\beta_1\alpha_1 r_1-\frac{\theta}{2}(\alpha_1+\beta_1)+1=0,&\\
\beta_2\alpha_2 r_1r_2+\frac{1}{2}(\alpha_2+\beta_2)(r_1-r_2)+\theta\beta_2\alpha_2 r_2+\frac{\theta}{2}(\alpha_2+\beta_2)-1=0 \qquad (6.2.29)&
\end{aligned}$$

Πολλαπλασιάζοντας την πρώτη με $\alpha_2\beta_2$ και τη δεύτερη με $\alpha_1\beta_1$, προσθέτοντας τις εξισώσεις που προκύπτουν και λύνοντας ως προς $r_1 - r_2$ λαμβάνουμε τη σχέση

$$r_1 - r_2 \;=\; \frac{2(\beta_1\alpha_1 - \beta_2\alpha_2) + \theta\left[\beta_2\alpha_2(\alpha_1+\beta_1) - \beta_1\alpha_1(\alpha_2+\beta_2)\right]}{\beta_2\alpha_2(\alpha_1+\beta_1) + \beta_1\alpha_1(\alpha_2+\beta_2) - 2\theta\beta_1\alpha_1\beta_2\alpha_2} \;=:\; H(\theta) \equiv H. \tag{6.2.30}$$

Αντικαθιστώντας την τιμή του $r_1 = r_2 + H(\theta)$ στη δεύτερη σχέση (6.2.29) και λύνοντας την εξίσωση δευτέρου βαθμού που προκύπτει ως προς $r_2$ έχουμε

$$r_2 \;=\; \frac{-\beta_2\alpha_2(H+\theta) \pm \left[\beta_2^2\alpha_2^2(H+\theta)^2 - 2\beta_2\alpha_2[(\alpha_2+\beta_2)(H+\theta) - 2]\right]^{\frac{1}{2}}}{2\beta_2\alpha_2}. \tag{6.2.31}$$

Στη συνέχεια αντικαθιστώντας στη σχέση $r_1 = r_2 + H$ έχουμε ότι

$$r_1 \;=\; \frac{\beta_2\alpha_2(H-\theta) \pm \left[\beta_2^2\alpha_2^2(H+\theta)^2 - 2\beta_2\alpha_2[(\alpha_2+\beta_2)(H+\theta) - 2]\right]^{\frac{1}{2}}}{2\beta_2\alpha_2}. \tag{6.2.32}$$

Σημειώνουμε ότι για την ανάλυση που έγινε τα δυο τμήματα υπερβολών που θεωρήσαμε πρέπει να έχουν το μόνο σημείο τομής τους $O$ αυστηρά μέσα στο ορθογώνιο $ABCD$. Σημειώνουμε επίσης ότι οι εκφράσεις στις σχέσεις (6.2.32) και (6.2.31), που δίνουν τα ζεύγη $(r_1, r_2)$ που λύνουν το πρόβλημα, πρέπει να έχουν το ίδιο πρόσημο μέσα στις τετραγωνικές ρίζες. Θεωρούμε, ωστόσο, ότι εάν πάρουμε τα ελάχιστα των τετραγωνικών ριζών έχουμε

$$r_1 + r_2 \;=\; \frac{-\beta_2\alpha_2\theta - \left[\beta_2^2\alpha_2^2(H+\theta)^2 - 2\beta_2\alpha_2[(\alpha_2+\beta_2)(H+\theta) - 2]\right]^{\frac{1}{2}}}{\beta_2\alpha_2} \;<\; 0,$$

το οποίο είναι αδύνατο. Έτσι οι βέλτιστες τιμές για τα $r_1$ και $r_2$ δίνονται τελικά από τις επόμενες εκφράσεις

$$\begin{aligned} r_1^* \;&=\; \tfrac{1}{2}\left\{(H-\theta) + \left[(H+\theta)^2 - \tfrac{2}{\beta_2\alpha_2}[(\alpha_2+\beta_2)(H+\theta) - 2]\right]^{\frac{1}{2}}\right\}, \\ r_2^* \;&=\; \tfrac{1}{2}\left\{-(H+\theta) + \left[(H+\theta)^2 - \tfrac{2}{\beta_2\alpha_2}[(\alpha_2+\beta_2)(H+\theta) - 2]\right]^{\frac{1}{2}}\right\}. \end{aligned} \tag{6.2.33}$$

Συνεπώς, έχουμε αποδείξει ότι:

**Θεώρημα 6.2.2.** *Με βάση τους μέχρι τώρα συμβολισμούς και τις υποθέσεις, οι βέλτιστες τιμές των παραμέτρων επιτάχυνσης $r_1 = r_1^*$ και $r_2 = r_2^*$ της διακριτής εξίσωσης Poisson (6.2.5), χρησιμοποιώντας τους ADI Προρρυθμιστές (6.2.17), δίνονται από τις εκφράσεις (6.2.33), όπου $H$ δίνεται από τη σχέση (6.2.30). Οι βέλτιστες τιμές για την παράμετρο χαλάρωσης $\omega^*$ και του δείκτη κατάστασης $\kappa^*(M^{-1}A)$, δίνονται από τις εκφράσεις (6.2.19) και (6.2.20), αντίστοιχα. Οι τιμές αυτές βρίσκονται αφού έχουμε ήδη βρει τις τιμές των $G^*$ και $g^*$, χρησιμοποιώντας κάποια από τις δύο εκφράσεις του Πίνακα 6.8.*

### 6.2.5 Άλλες Δυνατές Περιπτώσεις

Στην προηγούμενη παράγραφο η περίπτωση $\theta = \theta^* < \frac{1}{\beta_i},\ i = 1, 2$, εξετάστηκε και αντιμετωπίστηκε επιτυχώς. Η περίπτωση που εξετάστηκε καλύπτει επίσης και την περίπτωση των $5-$σημείων, όταν $\theta = 0$ καθώς επίσης και τις περιπτώσεις $\theta = \theta^* \in [\frac{1}{6}, \frac{1}{2\sqrt{5}}]$. Παρόλα αυτά όπως μπορούμε να αναφερθούμε στην εργασία [;], που εύκολα μπορούμε να ελέγξουμε στην περίπτωση μας, ότι οι επιπλέον δυνατές περιπτώσεις είναι αυτές στις οποίες μία από τις δύο ανισότητες $\theta = \theta^* > \frac{1}{\beta_i},\ i = 1, 2$, ισχύει.

Θα εξετάσουμε πολύ συνοπτικά μία από τις δύο δυνατές περιπτώσεις. Έστω αυτή για την οποία ισχύει $\frac{1}{\beta_2} < \theta^* < \frac{1}{\beta_1}$. Τα συμπεράσματα για την άλλη δυνατή περίπτωση προκύπτουν με ανάλογο τρόπο και γι΄ αυτό παραλείπονται. Προφανώς, σε αυτήν την περίπτωση ο Πίνακας 6.7 θα αλλάξει εξαιτίας του περιορισμού του $\beta_2$, οπότε θα ισχύει ότι $\frac{\partial f(V_{\alpha_1\beta_2}\overrightarrow{V}_{\beta_1\beta_2})}{\partial r_1} < 0$ για κάθε $r_1 \in (0, \infty)$. Λόγω αυτής της αλλαγής, ο Πίνακας 6.8 περιορίζεται στο τμήμα του αρχικού εκτός των τριών αριστερών χωρίων τα οποία τώρα δεν υπάρχουν. Τα αναφερθέντα παραπάνω παρουσιάζονται στον επόμενο Πίνακα 6.9, όπου όλα τα συμπεράσματα στα υπόλοιπα κελιά του Πίνακα 6.8 παραμένουν αμετάβλητα.

Για να βρούμε το μέγιστο και το ελάχιστο της συνάρτησης $f$ στο αντίστοιχο, με την προηγούμενη γενική περίπτωση, κεντρικό ορθογώνιο $A'BCD'$ εργαζόμαστε με τον ίδιο ακριβώς τρόπο που εργαστήκαμε στις προηγούμενες παραγράφους. Έτσι, υπολογίζουμε τις εκφράσεις $Q(r_1, r_2) = f(\alpha_1, \beta_2) - f(\beta_1, \alpha_2)$ και $q(r_1, r_2) = f(\beta_1, \beta_2) - f(\alpha_1, \alpha_2)$ στη σχέση (6.2.24) και βρίσκουμε τα πρόσημα και τις τιμές στις κορυφές $A'$, $B$, $C$, $D'$. Οι συντεταγμένες των κορυφών του ορθογωνίου $A'BCD'$ είναι $A'\left(0, \frac{1}{\alpha_1} - \theta^*\right),\ D'\left(0, \frac{1}{\beta_1} - \theta^*\right)$, ενώ αυτές των δύο άλλων κορυφών παραμένουν αμετάβλητες. Έτσι, συνοψίζουμε

| $r_2$ | | |
|---|---|---|
| $+\infty$ | $G = f(\beta_1, \alpha_2)$<br>$g = f(\beta_1, \beta_2)$<br><br>$A'$ | $G = f(\alpha_1, \alpha_2)$<br>$g = f(\beta_1, \beta_2)$<br><br>$B$ |
| $\frac{1}{\alpha_1} - \theta^*$ | $G = \max\{f(\alpha_1, \beta_2), f(\alpha_2, \beta_1)\}$<br>$g = \min\{f(\alpha_1, \alpha_2), f(\beta_1, \beta_2)\}$<br><br>$D'$ | $G = f(\alpha_1, \beta_2)$<br>$g = f(\beta_1, \beta_2)$<br><br>$C$ |
| $\frac{1}{\beta_1} - \theta^*$ | $G = f(\alpha_1, \beta_2)$<br>$g = f(\alpha_1, \alpha_2)$ | $G = f(\alpha_1, \beta_2)$<br>$g = f(\beta_1, \alpha_2)$ |
| $0$ | $\frac{1}{\alpha_2} - \theta^*$ | $+\infty \quad r_1$ |

Πίνακας 6.9: Μέγιστο και ελάχιστο της $f$ σε κάθε χωρίο του πρώτου τεταρτημορίου του επιπέδου των $r_1$, $r_2$.

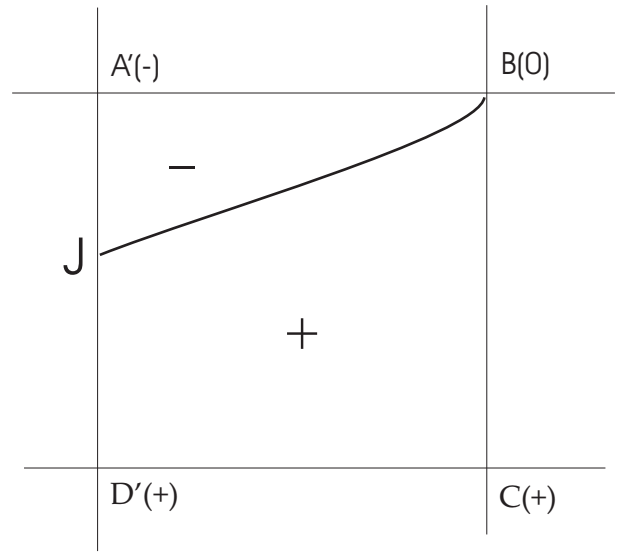
με τα επόμενα συμπεράσματα:

$$\begin{cases} Q(r_1, r_2) < 0 & \text{στην κορυφή } A', \\ Q(r_1, r_2) = 0 & \text{στην κορυφή } B, \\ Q(r_1, r_2) > 0 & \text{στην κορυφή } C, \\ Q(r_1, r_2) > 0 & \text{στην κορυφή } D', \end{cases} \tag{6.2.34}$$

και

$$\begin{cases} q(r_1, r_2) < 0 & \text{στην κορυφή } A', \\ q(r_1, r_2) < 0 & \text{στην κορυφή } B, \\ q(r_1, r_2) = 0 & \text{στην κορυφή } C, \\ q(r_1, r_2) > 0 & \text{στην κορυφή } D'. \end{cases} \tag{6.2.35}$$

Θα πρέπει να σημειώσουμε ότι οι μόνες διαφορές από τις βασικές περιπτώσεις που εξετάσαμε στις προηγούμενες παραγράφους έγκεινται στα πρόσημα της συνάρτησης $Q(r_1, r_2)$ στις κορυφές $D'$ το οποίο γίνεται αρνητικό αντί για 0, όπως ήταν στην $D$, και σ΄ αυτό της συνάρτησης $q(r_1, r_2)$ στην κορυφή $A'$, το οποίο γίνεται αρνητικό αντί να είναι 0, όπως ήταν στην $A$ (βλ. Σχήματα 6.5 και 6.6, αντίστοιχα).
Αυτό απλά σημαίνει ότι τα δύο τμήματα των υπερβολών $Q(r_1, r_2) = 0$ και $q(r_1, r_2) = 0$ θα έχουν ένα σημείο τομής η καθεμιά με την πλευρά $A'D'$



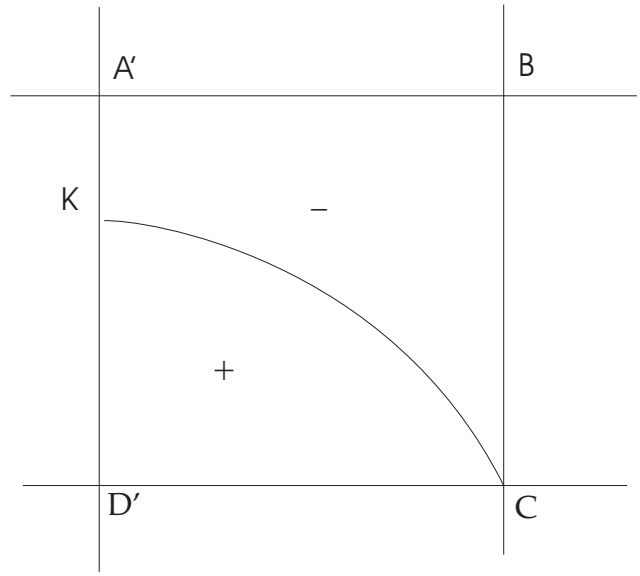

Σχήμα 6.5: Πρόσημο της $Q(r_1, r_2)$ στο $ABCD$.

του ορθογωνίου $A'BCD'$. Έστω ότι το σημείο αυτό είναι το $J$, με συντεταγμένες $(r_1, r_2) = (0, r_{2J})$, για την $Q(r_1, r_2)$, και $K$, με συντεταγμένες $(r_1, r_2) = (0, r_{2J})$, για την $q(r_1, r_2)$. Τα σημεία αυτά εμφαίνονται στα Σχήματα 6.5 και 6.6.

Απομένει να δείξουμε ότι οι τα δύο τμήματα των υπερβολών τέμνονται σε σημείο $O$, το οποίο, όπως και πριν, εξακολουθεί να βρίσκεται στο εσωτερικό του ορθογωνίου στο οποία αναφερόμαστε. Ειδικότερα, έχουμε ισχύουσα την πρακάτω πρόταση.

**Θεώρημα 6.2.3.** *Με βάση τους μέχρι τώρα συμβολισμούς και τις υποθέσεις και την επιπλέον υπόθεση ότι* $\frac{1}{\beta_2} - \theta^* < 0$, *οι δύο υπερβολές που ορίζονται από από τις συναρτήσεις* $Q(r_1, r_2) = 0$ *και* $q(r_1, r_2) = 0$, *που ορίζονται στις σχέσεις (6.2.25) και (6.2.26), τέμνονται στο σημείο* $O$, *το οποίο βρίσκεται αυστηρά εντός του ορθογωνίου* $A'BCD'$, *όπως αυτό εμφαίνεται στο Σχήμα 6.7.*

Απόδειξη: Αρχίζουμε από τις βασικές περιπτώσεις των προηγούμενων παραγράφων 6.2.3, οι οποίες ήταν $\frac{1}{\beta_i} - \theta^* > 0$, $i = 1, 2$, και το σημείο $O$ βρισκόταν αυστηρά στο εσωτερικό του ορθογωνίου $ABCD$, όπως φαίνεται στο Σχήμα 6.4. Έστω τώρα ότι η ποσότητα $\frac{1}{\beta_2} - \theta^*$ ελαττώνεται συνεχώς πηγαίνοντας από τις θετικές τιμές στο μηδέν και στη συνέχεια στις αρνητικές. Στην αρχή εξετάζουμε την περίπτωση της μηδενικής τιμής, η οποία λαμβάνεται στην

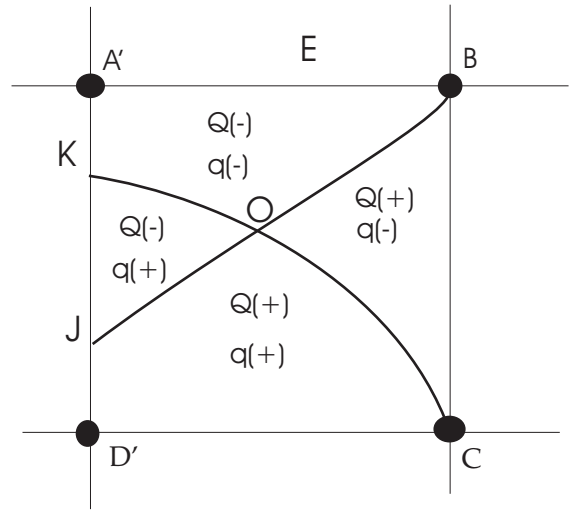



Σχήμα 6.6: Πρόσημο του $q(r_1, r_2)$ στο $ABCD$.

περίπτωση όπου $\frac{1}{\beta_2} = \theta^*$ και $ABCD \equiv A'BCD'$. Έστω $r_{2J}$ και $r_{2K}$ ότι είναι οι τιμές του $r_2$ για $r_1 = 0 = \frac{1}{\beta_2} - \theta^*$ των σημείων $J$ και $K$, αντίστοιχα. Θέτοντας και στις δύο εξισώσεις της σχέσης (6.2.28), $r_1 = 0$ και $\theta = \theta^* = \frac{1}{\beta_2}$ και λύνοντας την πρώτη εξίσωση ως προς $r_2 = r_{2J}$ και τη δεύτερη ως προς $r_2 = r_{2K}$, λαμβάνουμε

$$r_{2J} = \frac{\beta_2 - \beta_1}{\beta_1\beta_2}, \quad r_{2K} = \frac{\beta_2 - \beta_1}{\alpha_1\beta_2} \quad \text{και} \quad r_{2J} - r_{2K} = \frac{(\beta_2 - \beta_1)(\alpha_1 - \beta_1)}{\alpha_1\beta_1\beta_2} < 0, \tag{6.2.36}$$

εφόσον $\alpha_1 < \beta_1 < \beta_2$. Το αποτέλεσμα αυτό απλά σημαίνει ότι το $O$ βρίσκεται και σ΄ αυτή την περίπτωση στο εσωτερικό του $A'BCD'$.
Ας θεωρήσουμε τώρα ότι η ποσότητα $\frac{1}{\beta_2} - \theta^*$ ελαττώνεται συνέχως από τη μηδενική τιμή. Απαιτούμε κατά τη συνεχή ελάττωσή της προς το σημείο τομής $O$ να βρίσκεται πάντα προς τα δεξιά της πλευράς $A'D'$ του ορθογωνίου $A'BCD'$. Εάν για τις αρνητικές τιμές της παραπάνω ποσότητας το $O$ βρισκόταν στα αριστερά της $A'D'$ τότε θα υπήρχε μία τιμή της ποσότητας αυτής τέτοια ώστε το $O$ να έχει μηδενική τετμημένη. Σ' αυτήν την περίπτωση, προφανώς, $r_1^* = 0$. Τότε, από τη σχέση (6.2.30) θα έχουμε ότι $-r_2^* = H =: H_0$, έτσι $H_0 < 0$. Εξάλλου, από την πρώτη έκφραση των σχέσεων (6.2.33), για $r_1^* = 0$, μετά από κάποιους




Σχήμα 6.7: Πρόσημο των $Q(r_1, r_2)$ και $q(r_1, r_2)$ στο $ABCD$.

υπολογισμούς, και λύνοντας ως προς $H = H_0$, έχουμε ότι

$$H_0 = \frac{2 - (\alpha_2 + \beta_2)\theta^*}{\alpha_2 + \beta_2 - 2\theta^* \alpha_2 \beta_2}. \tag{6.2.37}$$

Παρατηρώντας ότι η ελάχιστη τιμή του αριθμητή της σχέσης (6.2.37) λαμβάνεται για τη μέγιστη τιμή της έκφρασης $(\alpha_2 + \beta_2)\theta^*$, η οποία είναι ίση με

$$\max\left\{4\sqrt{\frac{b}{a}}\frac{h_2}{h_1}\frac{1}{12}\left(\sqrt{\frac{b}{a}}\frac{h_2}{h_1} + \sqrt{\frac{a}{b}}\frac{h_1}{h_2}\right)\right\} = 4\sqrt{5}\frac{1}{12}\left(\sqrt{5} + \frac{1}{\sqrt{5}}\right) = 2,$$

έχουμε ότι ο αριθμητής είναι πάντα μη αρνητικός. Αν θεωρήσουμε τον παρανομαστή της σχέσης (6.2.37), αυτός γράφεται στη μορφή

$$4\sqrt{\frac{b}{a}}\frac{h_2}{h_1}\left\{1 - 2\sqrt{\frac{b}{a}}\frac{h_2}{h_1}\,\sin^2(\pi h_2)\,\theta^*\right\}.$$

Ο δεύτερο όρος στις αγκύλες της παραπάνω έκφρασης έχει πάνω φράγμα ίσο με

$$2\sqrt{5}\left(\frac{\sqrt{2}}{2}\right)^2\frac{1}{12}\left(\sqrt{5} + \frac{1}{\sqrt{5}}\right) = \frac{1}{2},$$

κάνοντας τον παρονομαστή να είναι πάντα θετικός. Τα τελευταία δύο αποτελέσματα ορίζουν τα πρόσημα των όρων του κλάσματος στη σχέση (6.2.37), από τα οποία προκύπτει ότι το $H_0$ **δεν** μπορεί να είναι αρνητικό. Αυτό αποδεικνύει την απαίτηση μας και συνεπάγεται την απόδειξη του θεωρήματος. □

Το Θεώρημα 6.2.3, που μόλις αποδείχτηκε, έχει ως επιπλέον συνέπεια και το ακόλουθο θεώρημα.

**Θεώρημα 6.2.4.** *Με βάση τους μέχρι τώρα συμβολισμούς και τις υποθέσεις και την επιπλέον υπόθεση ότι η ανισότητα* $\frac{1}{\beta_1} < \theta^*$ *(αντίστοιχα,* $\frac{1}{\beta_2} < \theta^*$*) ικανοποιείται, τότε, οι τιμές των* $G$ *και* $g$ *δίνονται από τον Πίνακα 6.8 χωρίς τα τρία αριστερότερα κελιά του για* $r_1$*, όπως παρουσιάζονται στον Πίνακα 6.9 (αντίστοιχα, στον Πίνακα 6.8 χωρίς τα τρία τελευταία κελιά για* $r_2$*). Παρόλα αυτά, οι βέλτιστες τιμές των* $r_1^*$ *και* $r_2^*$*, όπως και για όλες οι άλλες παράμετροι, που εμπλέκονται, δίνονται από τους ίδιους ακριβώς τύπους και εκφράσεις όπως στο Θεώρημα 6.2.2.*

# Κεφάλαιο 7

# Αριθμητικά Παραδείγματα

## 7.1 Εισαγωγή

Στην παράγραφο αυτή θα παρουσιάσουμε μία σειρά από αριθμητικά παραδείγματα που επαληθεύουν τα θεωρητικά αποτελέσματα που ήδη παρουσιάσαμε σε προηγούμενα κεφάλαια. Στα αριθμητικά πειράματα που εκτελέσαμε συγκρίνουμε την Προρρυθμισμένη Μέθοδο Συζυγών Κλίσεων, με προρρυθμιστή τον (E)ADI Προρρυθμιστή, $M = (I + rA_2)(I + rA_1)$ για τη μονοπαραμετρική και τον $M = (I + r_2A_2)(I + r_1A_1)$ για τη διπαραμετρική περίπτωση, με την απλή μέθοδο Συζυγών Κλίσεων αλλά κυρίως με την προρρυθμισμένη CG μέθοδο χρησιμοποιώντας προρρυθμιστές τους Line-Jacobi, Block-Jacobi, ILU(0) καθώς και με μεθόδους που θεωρούνται οι καταλληλότερες (δημοφιλέστερες) για την επίλυση τέτοιων προβλημάτων όπως είναι οι Cyclic Reduction, FTT-Cyclic Reduction και φυσικά οι Multigrid μέθοδοι. Σε κάθε περίπτωση από τις παραπάνω, τα αποτελέσματα ήταν υπέρ της ADI-CG ή τουλάχιστον ήταν συγκρίσιμα με αυτά κάποιων από τις μεθόδους αυτές. Θα πρέπει εδώ να σημειώσουμε ότι οι συγκρίσεις, που αναφέρουμε και θα παρουσιάσουμε παρακάτω, έγιναν μεταξύ ενός δικού μας προγράμματος και προγραμμάτων τα οποία προήλθαν από έτοιμα πακέτα λογισμικών προγραμμάτων.

Αρχικά θα παρουσιάσουμε αριθμητικά παραδείγματα που αφορούν στην περίπτωση των μονοπαραμετρικών ADI για διακριτά πλέγματα που προέρχονται από την διακριτοποίηση των $5-$ και $9-$σημείων. Στη συνέχεια θα επεκταθούμε σε παραδείγματα και θα παρουσιάσουμε αριθμητικά αποτελέσματα στην περίπτωση των διπαραμετρικών ADI και για τις δύο περιπτώσεις διακριτοποίησης που αναφέραμε παραπάνω με έμφαση σ΄ αυτή των $9-$σημείων η οποία παρουσιάζει

και το μεγαλύτερο ενδιαφέρον.

## 7.2 Αριθμητικά Παραδείγματα - Μονοπαραμετρική Περίπτωση

Παρουσιάζουμε παρακάτω μία σειρά από αριθμητικά παραδείγματα που αφορούν στο πρόβλημα της διδιάστατης εξίσωσης Poisson στο ανοιχτό τετράγωνο $(0,1)\times(0,1)$ με Dirichlet συνοριακές συνθήκες. Το δεξιό μέλος $f(x,y)$ της εξίσωσης καθώς και η έκφραση για την συνοριακή συνθήκη $\gamma(x,y)$, σε κάθε διαφορική εξίσωση προκύπτουν όταν απαιτήσουμε να έχουμε μία προκαθορισμένη έκφραση για την πραγματική λύση. Σε κάθε παράδειγμα που επεξεργαστήκαμε τα αποτελέσματα ήταν λίγο-πολύ τα ίδια γι΄ αυτό και παρακάτω θα παρουσιάσουμε μόνο την περίπτωση όπου η πραγματική λύση δίνεται από την έκφραση:

$$u(x,y) \;=\; \exp(x+y)\sin\left(\frac{\pi x}{2}\right)\sin\left(\frac{\pi y}{2}\right). \tag{7.2.1}$$

Θεωρούμε ένα ομοιόμορφο διαμερισμό με βήμα διακριτοποίησης $h=\frac{1}{n+1}$, $n=5\times 2^l$, $l=0,\ldots,5$. Τα γραμμικά συστήματα που παράγονται και στις δύο περιπτώσεις διακριτοποίησης του τελεστή με πλέγμα των $5-$ και των $9-$σημείων λύνονται χρησιμοποιώντας την Μέθοδο των Συζυγών Κλίσεων (CG), με προρρυθμιστή τον "μπλοκ" (Block) Jacobi και με το Βέλτιστο (E)ADI Προρρυθμιστή. Για την προρρυθμισμένη μέθοδο Συζυγών Κλίσεων με (E)ADI Προρρυθμιστή (ADI-CG) και για τις δύο περιπτώσεις διακριτοποίησης χρησιμοποιούμε τα παρακάτω δεδομένα:

$$\begin{gathered} a=b=1,\quad n_1=n_2=n,\quad h_1=h_2=h=\tfrac{1}{n+1},\\ \alpha_1=\alpha_2=\alpha=4\sin^2(\tfrac{\pi}{2(n+1)}),\quad \beta_1=\beta_2=\beta=4\cos^2(\tfrac{\pi}{2(n+1)}). \end{gathered} \tag{7.2.2}$$

Για τις δύο περιπτώσεις των $5-$ και των $9-$σημείων, αναφερόμαστε στις περιπτώσεις $A$ ή $B$ των διατάξεων που αναφερθήκαμε στην θεωρία (βλ. (6.1.40) ). Έτσι,

$$\theta=0,\; r^* \;=\; \frac{1}{\sqrt{\alpha\beta}}^{1},\quad \omega^* \;=\; \frac{2}{\sqrt{\alpha\beta}} \tag{7.2.3}$$

[1]Οι εκφράσεις για το βέλτιστο $r^*$ των εκφράσεων (4.1.3) και (6.1.28) είναι η μία αντίστροφη της άλλης. Το γεγονός αυτό εξηγεί την διαφορά μεταξύ των εκφράσεων του $r$ που χρησιμοποιείται στο Προρρυθμιστή $M_2$ της (6.1.16) και σ΄ αυτόν που δίνεται εδώ.

και

$$\theta = \theta^* = \tfrac{1}{6}, \quad r^* = \frac{\sqrt{\left(1-\frac{1}{12}\alpha\right)\left(1-\frac{1}{12}\beta\right)}-\frac{1}{12}\sqrt{\alpha\beta}}{\sqrt{\alpha\beta}}, \quad \omega^* = \frac{2\sqrt{\left(1-\frac{1}{12}\alpha\right)\left(1-\frac{1}{12}\beta\right)}}{\sqrt{\alpha\beta}}, \tag{7.2.4}$$

αντίστοιχα. Να σημειώσουμε εδώ ότι οι σχέσεις (7.2.3) και (7.2.4) είναι ειδικές περιπτώσεις των γενικότερων εκφράσεων

$$r^* = \frac{\sqrt{\left(1-\frac{1}{2}\theta\alpha\right)\left(1-\frac{1}{2}\theta\beta\right)}-\frac{1}{2}\theta\sqrt{\alpha\beta}}{\sqrt{\alpha\beta}}, \quad \omega^* = \frac{2\sqrt{\left(1-\frac{1}{2}\theta\alpha\right)\left(1-\frac{1}{2}\theta\beta\right)}}{\sqrt{\alpha\beta}}, \tag{7.2.5}$$

για $\theta = 0$ και $\theta = \frac{1}{6}$, αντίστοιχα. Στην υλοποίηση των αλγορίθμων χρησιμοποιήσαμε FORTRAN 77 με αριθμητική διπλής ακρίβειας, ενώ το κριτήριο σταματήματος και για τις δύο περιπτώσεις των 5− και των 9−σημείων είναι:

$$\frac{||e^{(k)}||_A}{||e^{(0)}||_A} \equiv \frac{(r^{(k)}, e^{(k)})^{\frac{1}{2}}_2}{(r^{(0)}, e^{(0)})^{\frac{1}{2}}_2} < 10^{-10} \quad \text{και} \quad \frac{||e^{(k)}||_{\widetilde{A}}}{||e^{(0)}||_{\widetilde{A}}} \equiv \frac{(z^{(k)}, e^{(k)})^{\frac{1}{2}}_2}{(z^{(0)}, e^{(0)})^{\frac{1}{2}}_2} < 10^{-10}.$$

Στις εκφράσεις αυτών των κριτηρίων η συνάρτηση $u$ είναι η ακριβής λύση του γραμμικού συστήματος, $u^{(k)}$, $e^{(k)} = u - u^{(k)}$, και $r^{(k)} = b - Au^{(k)}$ είναι η προσεγγιστική λύση, το διάνυσμα του σφάλματος και το διάνυσμα του υπολοίπου στην $k$ επανάληψη της αντίστοιχης μεθόδου, αντίστοιχα, η $z^{(k)}$ είναι η λύση του συστήματος του $Mz^{(k)} = r^{(k)}$, όπου $M$ είναι ο προρρυθμιστής που χρησιμοποιούμε. Επίσης ορίζουμε $\widetilde{A} = M^{-\frac{1}{2}}AM^{-\frac{1}{2}}$ να είναι ο προρρυθμισμένος πίνακας. Το πρώτο κριτήριο αναφέρεται στη μέθοδο των Συζυγών Κλίσεων CG και το δεύτερο στην Προρρυθμισμένη Μέθοδο Συζυγών Κλίσεων PCG. Για αρχικό διάνυσμα της προσέγγισης της λύσης θεωρούμε το $u^{(0)} = [1\,1\,\ldots\,1]^T \in \mathbb{R}^{n^2}$.

Τα αριθμητικά αποτελέσματα που παράγονται σε κάθε μία από τις έξι περιπτώσεις παρουσιάζονται στους Πίνακες 7.1 και 7.2 για τα 5− και τα 9−σημεία, αντίστοιχα. Σε όλες τις περιπτώσεις τα αντίστοιχα κριτήρια σταματήματος ικανοποιήθηκαν. Σε κάθε Πίνακα όπου παρουσιάζεται μία επαναληπτική μέθοδος υπάρχουν τρεις στήλες από τις οποίες η πρώτη παρουσιάζει τις επαναλήψεις που χρειάζεται η επαναληπτική μέθοδος για να ικανοποιήσει το κριτήριο σταματήματος (iter), η δεύτερη στήλη παρουσιάζει το χρόνο σε δευτερόλεπτα που χρειάζεται μέχρι την σύγκλιση CPU time (cpu_time) και η τελευταία παρουσιάζει το σχετικό απόλυτο σφάλμα $\frac{||u-u^{(\text{iter})}||_\infty}{||u||_\infty}$. $u$ είναι το διάνυσμα της

πραγματικής λύσης με $n^2$ συνιστώσες $u(ih, jh)$, $i, j = 1, \ldots, n$, της συνάρτησης $u(x, y)$ στους εσωτερικούς κόμβους του πλέγματος, η οποία ικανοποιεί την εξίσωση Poisson.

Αρχικά, από τον Πίνακα 7.1 μπορούμε να διαπιστώσουμε ότι η τάξη σύγκλισης $\mathcal{O}(h^2)$, που επιθυμούμε, λαμβάνεται σε κάθε μία από τις περιπτώσεις που παρουσιάζονται. Θα πρέπει να σημειώσουμε ότι για $n \geq 20$ το σχετικό απόλυτο σφάλμα γίνεται καλύτερο καθώς το $n$ αυξάνει, και είναι το ίδιο σε κάθε μία από τις τρεις περιπτώσεις των μεθόδων που εξετάζονται. Παρατηρούμε από τα αποτελέσματα του πίνακα ότι η μέθοδος ADI-CG χρειάζεται πολύ λιγότερο CPU time από ό,τι οι άλλες δύο παρά το ότι λύνουμε δύο τριδιαγώνια συστήματα σε κάθε επανάληψη της μεθόδου των Συζυγών Κλίσεων. Θα πρέπει εδώ βέβαια να τονίσουμε ότι ουσιαστικά χρειάζεται μόνο μία παραγοντοποίηση των πινάκων που χρησιμοποιούνται στα συστήματα ενώ στη συνέχεια σε κάθε επανάληψη εκτελούμε μόνο προς τα πίσω αντικαταστάσεις. Η μέθοδος "μπλοκ" Jacobi CG φαίνεται ότι για μεγάλα $n$ είναι τόσο γρήγορη όσο και η απλή CG γεγονός που πρέπει να οφείλεται στο κόστος της επίλυσης του επιπλέον τριδιαγώνιου συστήματος σε κάθε επανάληψη της CG. Σημειώνουμε ότι η "μπλοκ" Jacobi CG έχει δείκτη κατάστασης το μισό από αυτόν της απλής CG, γεγονός που εξηγεί τα αποτελέσματα που λαμβάνουμε.

Για τον Πίνακα 7.2 έχουμε να κάνουμε παρόμοια σχόλια. Θα πρέπει όμως να τονίσουμε ότι παρά το γεγονός ότι στην περίπτωση των $9-$σημείων το κόστος θεωρητικά θα έπρεπε να ήταν μεγαλύτερο, στην πραγματικότητα δε συμβαίνει κάτι τέτοιο αφού και ο χρόνος που απαιτείται είναι ουσιαστικά ο ίδιος με αυτόν των $5-$σημείων. Μία απλή εξήγηση που θα μπορούσε να δώσει κάποιος είναι ότι και στην περίπτωση των $9-$σημείων το κόστος σε χρόνο παρουσιάζεται στο σημείο της επίλυσης των δύο συστημάτων σε κάθε επανάληψη της μεθόδου των Συζυγών Κλίσεων. Όπως αναφέραμε και προηγουμένως το κόστος αυτό δεν έχει ουσιαστική επιβάρυνση στην πραγματικότητα αφού μία φορά παραγοντοποιούμε τους πίνακες και σε κάθε επανάληψη λύνουμε απλά τριγωνικά συστήματα με ελάχιστο κόστος πράξεων.

Οι συναρτήσεις και οι υπορουτίνες που χρησιμοποιήθηκαν, εκτός αυτών της ADI-CG, είναι από το πακέτο έτοιμων προγραμμάτων NSPCG (βλ. [44] και επίσης [12]). Τα πειράματα που έγιναν για κάθε διακριτό πλέγμα δείχνουν ότι η ADI-CG είναι ταχύτερη σε σχέση με κάθε μία από τις έτοιμες συνάρτήσεις του πακέτου NSPCG, όπως η (Incomplete Cholesky (IC) και οι παραλλάγες αυτής). Κατά συνέπεια, για το σχήμα των $9-$σημείων συγκρίναμε την ADI-CG με την Block Modified IC -CG (MBIC-CG), η οποία ήταν και η γρηγορότερη από τις υπόλοιπες μεθόδους του πακέτου. Επίσης τη συγκρίναμε με τις, κατά

| | | CG (5-points) | | | Block Jacobi CG (5-points) | | | ADI CG (5-points) | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | iter | cpu_time | error | iter | cpu_time | error | iter | cpu_time | error |
| $u(x,y) = e^{x+y}\sin\frac{\pi x}{2}\sin\frac{\pi y}{2}$ | n=5 | 13 | 0 | 1,64071951E-03 | 20 | 0 | 1,64071951E-03 | 9 | 0 | 1,64071951E-03 |
| | n=10 | 49 | 0,01 | 4,15572589E-04 | 40 | 0 | 4,15572589E-04 | 27 | 0 | 4,15572589E-04 |
| | n=20 | 88 | 0,0100144 | 1,04130152E-04 | 75 | 0,010014 | 1,04130152E-04 | 36 | 0,010014 | 1,04130152E-04 |
| | n=40 | 162 | 0,06 | 2,60300717E-05 | 145 | 0,08 | 2,60300717E-05 | 52 | 0,04 | 2,60300717E-05 |
| | n=80 | 316 | 0,5 | 6,50587117E-06 | 256 | 0,58 | 6,50587159E-06 | 71 | 0,21 | 6,50587126E-06 |
| | n=160 | 593 | 4,306 | 1,62678010E-06 | 477 | 4,917 | 1,62678432E-06 | 93 | 1,2 | 1,62678269E-06 |

Πίνακας 7.1: Παρουσίαση των Τριών σχημάτων $5-$σημείων

| | | CG (9-points) | | | Block Jacobi CG (9-points) | | | ADI CG (9-points) | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | iter | cpu_time | error | iter | cpu_time | error | iter | cpu_time | error |
| $u(x,y) = e^{x+y}\sin\frac{\pi x}{2}\sin\frac{\pi y}{2}$ | n=5 | 14 | 0 | 8,28416617E-06 | 19 | 0 | 8,28416617E-06 | 13 | 0 | 8,28416617E-06 |
| | n=10 | 32 | 0 | 6,08503099E-07 | 31 | 0 | 6,08518509E-07 | 25 | 0 | 6,08503202E-07 |
| | n=20 | 60 | 0,01 | 4,13791887E-08 | 64 | 0,01 | 4,13848789E-08 | 29 | 0,01 | 4,13974005E-08 |
| | n=40 | 116 | 0,06 | 3,02718541E-09 | 120 | 0,09 | 3,01911486E-09 | 41 | 0,04 | 3,00112607E-09 |
| | n=80 | 231 | 0,45 | 1,19675917E-09 | 233 | 0,59 | 1,19903849E-09 | 55 | 0,19 | 1,18727336E-09 |
| | n=160 | 455 | 4,1 | 1,03637452E-09 | 446 | 5,48789 | 1,04157338E-09 | 77 | 1,192 | 1,04104154E-09 |

Πίνακας 7.2: Παρουσίαση των Τριών σχημάτων $9-$σημείων

γενική ομολογία, καλύτερες μεθόδους για το πρόβλημά μας και ειδικότερα με: 1) Τη Fast Fourier Transform (FFT9) με Block Cyclic Reduction (BCR) των Houstis και Papatheodorou [39], οι οποίοι χρησιμοποίησαν διακριτοποίηση των $9-$σημείων για το διαφορικό τελεστή και διακριτοποίηση των $5-$σημείων για το δεξιό μέλος της εξίσωσης Poisson (βλ. (2.1) στο [39]). 2) Την BCR από το έτοιμο πακέτο προγραμμάτων FISHPACK (www.cisl.ucar.edu/css/software/fishpack) των Swarztrauber και Sweet [53]. και 3) Τη μέθοδο Multigrid (MG)(MUDPACK) του Adams [3] (www.cisl.ucar.edu/css/software/mudpack).

Εκτελέσαμε μία σειρά από αριθμητικά παραδείγματα κάτω από τις προϋποθέσεις που αναφέραμε στην αρχή αυτής της παραγράφου. Επιπλέον θεωρήσαμε

ότι το πλήθος των εσωτερικών κόμβων στη μία διεύθυνση είναι διπλάσιο από αυτό στην άλλη. Επομένως έχουμε ότι $n_2 = 2n_1$ ($h_1 = \frac{1}{n_1+1}$ και $h_2 = \frac{1}{n_2+1}$). Με την επιλογή αυτή του πλήθους των εσωτερικών κόμβων χρησιμοποιήσαμε τις βέλτιστες παραμέτρους που ορίζονται για την Περίπτωση $B$ της διάταξης των ακραίων ιδιοτιμών. Στη συνέχεια Θα παρουσιαστούν τα αριθμητικά αποτελέσματα στην περίπτωση όπου η πραγματική λύση είναι η

$$u(x, y) \;=\; (x + y) \sin\left(\frac{\pi x}{2}\right) \sin\left(\frac{\pi y}{2}\right). \tag{7.2.6}$$

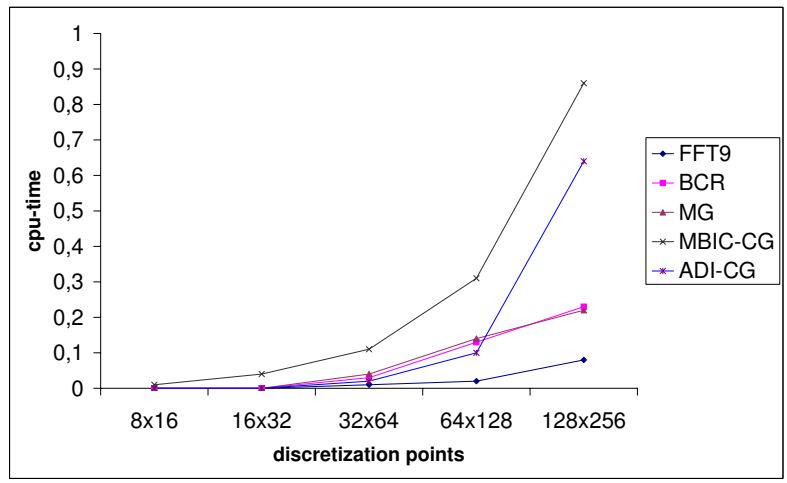
Σ′ αυτή την περίπτωση όλα τα προγράμματα είναι γραμμένα σε απλή ακρίβεια και σε FORTRAN 77 έτσι ώστε η σύγκριση με τις υπόλοιπες μεθόδους FFT9 [39], BCR [53] και MG [3], οι οποίες είναι γραμμένες επίσης σε απλή ακρίβεια, να είναι δίκαιη. Επίσης χρησιμοποιήσαμε διαμέριση με κόμβους της μορφής $n_1 \times n_2 = 2^k \times 2^l$, έτσι ώστε να είμαστε συμβατοί με αυτούς των έτοιμων προγραμμάτων FFT9 και MG. Όλα τα σχήματα διακριτοποίησης που χρησιμοποιήσαμε είναι τάξης ακρίβειας $\mathcal{O}(h^4)$. Για τις επαναληπτικές μεθόδους θεωρήσαμε ως αρχική προσέγγιση το μηδενικό διάνυσμα. Για λόγους σύγκρισης, τρέξαμε τα προγράμματα FFT9 και BCR και υπολογίσαμε την ποσότητα $\max_{i=1,\ldots,n_1,\, j=1,\ldots,n_2} \left| \frac{u_{i,j}^{(\text{approx})} - u_{i,j}}{u_{i,j}} \right|$ ώστε να μπορέσουμε να βρούμε την ακρίβεια για τη λύση. Η οντότητα $u^{(\text{approx})}$ είναι το διάνυσμα της προσεγγιστικής λύσης την οποία πήραμε. Στη συνέχεια, για τις MBIC-CG και ADI-CG μεθόδους χρησιμοποιήσαμε ένα κριτήριο σταματήματος της μορφής $\frac{\|r^{(k)}\|_2}{\|c\|_2} < tol$, όπου η ποσότητα $tol$ προσαρμόστηκε έτσι ώστε να έχουμε το ίδιο σχετικό απόλυτο σφάλμα για κάθε μία από τις μεθόδους FFT9 και BCR. Το κριτήριο σταματήματος για την MG είναι το πρότυπο που συνήθως χρησιμοποιείται. Συγκεκριμένα, $\frac{\|u^{(k)} - u^{(k-1)}\|_2}{\|u^{(k)}\|_2} < \epsilon$.

Από τον Πίνακα 7.4 και το Σχήμα 7.6, όπου παρουσιάζεται ο χρόνος σε σχέση με το πλήθος των εσωτερικών κόμβων του πλέγματος, έχουμε ότι για τα μεγέθη $8 \times 16$, $16 \times 32$, $32 \times 64$, $64 \times 128$, η μέθοδος ADI-CG είναι καλύτερη ή συγκρίσιμη με όλες τις υπόλοιπες σε σχέση πάντα με το χρόνο και τα σφάλματα. Για την περίπτωση του πλέγματος $64 \times 128$ δεν είναι καλύτερη από την FFT9, η οποία, όμως, έχει λιγότερο καλά σφάλματα με ένα λόγο της τάξης του 10 σε σχέση με αυτά της ADI-CG. Τα ίδια ακριβώς ισχύουν και στην περίπτωση των μεθόδων BCR και MBIC-CG. Για τη MG έχουμε να πούμε ότι είναι καλύτερη και από τις τρεις προηγούμενες. Για το πλέγμα $128 \times 256$ η FFT9 φαίνεται να είναι καλύτερη σε σχέση με τις υπόλοιπες σε ό,τι αφορά το χρόνο. Παρό-

λα αυτά η BCR είναι η "χειρότερη" από όλες σε σχέση με τα σφάλματα ενώ η MG είναι η καλύτερη από όλες τις άλλες FFT9, MBIC-CG και ADI-CG. Το ότι η μέθοδος αυτή είναι καλύτερη οφείλεται σε μία διαδικασία, αυτή των "Διαφορικών Διορθώσεων (difference corrections)", που χρησιμοποιείται στο αντίστοιχο πακέτο προγραμμάτων και παρουσιάζεται στο [3].

| $n_1 \times n_2$ | $8 \times 16$ | | $16 \times 32$ | | $32 \times 64$ | | $64 \times 128$ | | $128 \times 256$ | |
|---|---|---|---|---|---|---|---|---|---|---|
| | time | error | time | error | time | error | time | error | time | error |
| FFT9 | 0 | 4.3E-5 | 0 | 1.96E-4 | 0.01 | 9.6E-4 | 0.02 | 1.5E-2 | 0.08 | 6.7E-2 |
| BCR | 0 | 2.09E-4 | 0 | 1.87E-4 | 0.03 | 6.2E-3 | 0.13 | 9.95E-3 | 0.23 | 5.2E-1 |
| MG | 0 | 2.2E-3 | 0 | 2.98E-4 | 0.04 | 4.8E-5 | 0.14 | 7.2E-6 | 0.22 | 1.9E-6 |
| MBIC-CG | 0.01 | 1.1E-4 | 0.04 | 1.2E-4 | 0.11 | 4.4E-4 | 0.31 | 2.3E-3 | 0.86 | 1.8E-2 |
| ADI-CG | 0 | 1.4E-4 | 0 | 1.5E-4 | 0.02 | 2.01E-4 | 0.10 | 2.5E-3 | 0.64 | 2.3E-2 |

Πίνακας 7.3: Σχήμα Διακριτοποίησης 9−Σημείων




Σχήμα 7.1: Σχήμα Διακριτοποίησης 9−Σημείων

## 7.3 Αριθμητικά Παραδείγματα - Διπαραμετρική Περίπτωση

Θα θεωρήσουμε τρεις διδιάστατες εξισώσεις Poisson με ακριβείς λύσεις

$$\begin{aligned} u(x,y) &= \sin\left(\frac{\pi x}{2}\right)\sin\left(\frac{\pi y}{2}\right),\ u(x,y)=\exp(x+y), \\ u(x,y) &= \exp(x+y)\sin\left(\frac{\pi x}{2}\right)\sin\left(\frac{\pi y}{2}\right) \end{aligned} \tag{7.3.1}$$

στο ανοιχτό τετράγωνο $\Omega = (0,1)\times(0,1)$ με Dirichlet συνοριακές συνθήκες που ορίζονται από τις ακριβείς λύσεις (7.3.1). Θεωρούμε τον ομοιόμορφο διαμερισμό με βήματα διακριτοποίησης στις δύο κατευθύνσεις να είναι $h_i = \frac{1}{n_i+1},\ i=1,2$, όπου $n_1 = 40, 80, 160$ και $n_2 = \frac{n_1}{2}$, όπως επίσης και αντιστρόφως. Δηλαδή, $n_1 = 20, 40, 80$ και $n_2 = 2n_1$, στο χωρίο $\overline{\Omega}$. Η παραπάνω διακριτοποίηση χρησιμοποιήθηκε και στις δύο περιπτώσεις διακριτού πλέγματος που εξετάζουμε δηλαδή αυτό των $5-$σημείων ($\theta = 0$) και αυτό των $9-$σημείων ($\theta = \theta^*$), που χρησιμοποιούνται για την προσέγγιση της εξίσωσης Poisson στους εσωτερικούς κόμβους. Τα συστήματα που προκύπτουν από τις παραπάνω διακριτοποιήσεις λύνονται με τις επόμενες μεθόδους: Conjugate Gradient (CG), Cholesky (C), Optimal (E)ADI-CG, Incomplete Cholesky (IC)-CG, Block (B) IC-CG, Modified (M) IC-CG and Modified Block (MB) IC-CG. Η γλώσσα προγραμματισμού που χρησιμοποιήθηκε είναι πάλι η FORTRAN 77. Σε όλα τα υποπρογράμματα και τις συναρτήσεις χρησιμοποιήθηκε διπλή ακρίβεια. Το κριτήριο σταματήματος για τις έξι επαναληπτικές μεθόδους είναι το παρακάτω:

$$\frac{||e^{(k)}||_A}{||e^{(0)}||_A} \equiv \frac{(r^{(k)},e^{(k)})_2^{\frac{1}{2}}}{(r^{(0)},e^{(0)})_2^{\frac{1}{2}}} < 10^{-10} \text{ και } \frac{||e^{(k)}||_{M^{-1}A}}{||e^{(0)}||_{M^{-1}A}} \equiv \frac{(z^{(k)},e^{(k)})_2^{\frac{1}{2}}}{(z^{(0)},e^{(0)})_2^{\frac{1}{2}}} < 10^{-10}. \tag{7.3.2}$$

Στην (7.3.2), το πρώτο κριτήριο αναφέρεται στην CG και το δεύτερο στις υπόλοιπες πέντε προρρυθμισμένες μεθόδους. Επίσης ορίζουμε ώς, $u^{(k)}$, $e^{(k)} = u - u^{(k)}$, και $r^{(k)} = c - Au^{(k)}$ την ακριβή λύση το σφάλμα στην $k$ επανάληψη της λύσης και το υπόλοιπο στην $k$ επανάληψη αντίστοιχα. Η $z^{(k)}$ είναι η λύση του συστήματος $Mz^{(k)} = r^{(k)}$, όπου $M$ είναι ο προρρυθμιστής που χρησιμοποιήθηκε, σε κάθε μία περίπτωση. Η αρχική προσέγγιση της ακριβής λύσης $u^{(0)}$ πάρθηκε σε κάθε περίπτωση να είναι $[1\,1\,\ldots\,1]^T \in \mathbb{R}^{n_1,\ldots n_2}$.

Τα αριθμητικά αποτελέσματα που πήραμε σε κάθε μία από τις επτά περιπτώσεις για το ίδιο παράδειγμα και με την ίδια διαμέριση είχαν την ίδια περίπου

συμπεριφορά. Έτσι, επιλέξαμε μόνο το τρίτο παράδειγμα της (7.3.1) το οποίο και παρουσιάζεται στα Σχήματα 7.2, 7.3, 7.4 και 7.5. Από όλες τις μεθόδους, η μέθοδος της Παραγοντοποίησης Cholesky ήταν η χειρότερη, από την άποψη του χρόνου αλλά και των σφαλμάτων στην περίπτωση πυκνής διαμέρισης. Τα σφάλματα γι΄ αυτήν την περίπτωση είναι ικανοποιητικά μόνο για μικρές τιμές των $n_1$ και $n_2$ αλλά και σ΄ αυτές τις περιπτώσεις ο χρόνος που χρειάζεται για να λυθεί το πρόβλημα (7.3.1) είναι τουλάχιστον 5 φορές μεγαλύτερος από αυτόν της λιγότερο καλής επαναληπτικής μεθόδου. Βέβαια κάτι τέτοιο είναι αναμενόμενο αφού η μέθοδος είναι άμεση και δεν είναι η καταλληλότερη για αραιά συστήματα, όπως αυτά που προκύπτουν από την διακριτοποίηση του συγκεκριμένου διαφορικού τελεστή. Λόγω του γεγονότος αυτού δεν έχουμε συμπεριλάβει τη μέθοδο αυτή σε κανένα από τα παρακάτω σχήματα. Τα αποτελέσματα για όλες τις επαναληπτικές μέθοδους για το τρίτο παράδειγμα και για τη διακριτοποίηση των 5−σημείων παρουσιάζεται στα Σχήματα 7.2 και 7.3 ενώ για την διακριτοποίηση των 9−σημείων στα Σχήματα 7.4 και 7.5.

Αρχικά, από τα Σχήματα 7.2 και 7.3, στην περίπτωση των 5−σημείων φαίνεται ότι η μέθοδος ADI-CG είναι καλύτερη από την CG και από την IC-CG και χειρότερη από τις άλλες τρεις μεθόδους. Παρόλα αυτά, τα σχετικά σφάλματα που παίρνουμε από τις εκφράσεις $\frac{||u^{(\text{iter})}-u||_\infty}{||u||_\infty}$, όπου $u$ είναι το διάνυσμα της ακριβούς λύσης της συνάρτησης $u(x,y)$, με $n_1 \times n_2$ τιμές $u(i_1h_1, i_2h_2)$, $i_l = 1, \ldots n_l$, $l = 1, 2$, στους εσωτερικούς κόμβους και $u^{(\text{iter})}$ η προσεγγιστική λύση μετά την ικανοποίηση του κριτηρίου σταματήματος. Τα απόλυτα σφάλματα που λαμβάνουμε ($||u^{(\text{iter})} - u||_\infty$) είναι λίγο πολύ τα ίδια και είναι της τάξης του $10^{-4}$ για "μικρά" $n_i$, $i = 1, 2$, και της τάξης $10^{-6}$ για τις μεγαλύτερες τιμές αυτών.

Από τα Σχήματα 7.4 και 7.5, δηλαδή της περίπτωσης των 9−σημείων, φαίνεται ότι η μέθοδος ADI-CG είναι συγκρίσιμη και μάλλον καλύτερη από όλες τις μεθόδους με τις οποίες κάναμε σύγκριση. Τα απόλυτα σφάλματα είναι περίπου τα ίδια σε κάθε περίπτωση και είναι της τάξης του $10^{-8}$ για "μικρές" τιμές των $n_i$, $i = 1, 2$, και της τάξης του $10^{-9}$ για τις μεγαλύτερες.

Οι υπορουτίνες και οι συναρτήσεις που χρησιμοποιήθηκαν πάρθηκαν όλες από το το πακέτο έτοιμων συναρτήσεων NSPCG (βλ. [44] και [12]), ενώ το πρόγραμμα της μεθόδου ADI-CG είναι γραμμένο από εμάς. Επίσης θα πρέπει να σημειώσουμε ότι για κάθε διαμέριση η μέθοδος ADI-CG είναι ταχύτερη από κάθε άλλη μέθοδο του πακέτου NSPCG, όπως διαπιστώθηκε από τα πειράματα που έγιναν. Η περισσότερο συγκρίσιμη μέθοδος είναι η MBIC-CG και είναι αυτή που παρουσιάζουμε στα επόμενα σχήματα.

Έχοντας συγκρίνει τον ADI προρρυθμιστή με τους προρρυθμιστές IC, BIC και τις παραλλαγές τους συνεχίζουμε τη σύγκριση με μια σειρά από άμεσες και επαναληπτικές μεθόδους για τη διακριτοποίηση των 9−σημείων. Οι Άμεσες Μέθοδοι που χρησιμοποιήσαμε είναι: 1) Η Fast Fourier Transform (FFT9) με Block Cyclic Reduction (BCR) των Houstis και Papatheodorou [39]. Αυτή η μέθοδος χρησιμοποιεί διακριτοποίηση 9−σημείων για το διαφορικό τελεστή και διακριτοποίηση 5−σημείων για το δεξιό μέλος της εξίσωσης Poisson (βλ. (2.1) της [39]). και 2) Η BCR από το πακέτο FISHPACK (www.cisl.ucar.edu/css/software/fishpack) των Swarztrauber και Sweet [53]. Ως Επαναληπτική Μέθοδο χρησιμοποιούμε την CG ως τον βασικό επιλυτή (solver) με προρρυθμιστή την MBIC. Τη μέθοδο αυτή την πήραμε από το πακέτο NSPCG [44] όπως και πριν. Τέλος συγκρίναμε τη μέθοδο ADI-CG με μία μέθοδο Multigrid (MG) (MUDPACK program) του Adams [3] (www.cisl.ucar.edu/css/software/mudpack).

Όπως και στην περίπτωση της μίας παραμέτρου όλες οι συναρτήσεις που χρησιμοποιήθηκαν είναι γραμμένες σε FORTRAN 77 απλής ακρίβειας για καθαρά λόγους δίκαιης σύγκρισης. Επίσης Θεωρούμε πλέγματα της μορφής $n_1 \times n_2 = 2^k \times 2^l$, και πάλι γιατί τα έτοιμα πακέτα που χρησιμοποιήσαμε εργάζονται με πλέγματα τέτοιας μορφής. Κάθε σχήμα διακριτοποίησης στα παραπάνω προγράμματα είναι τάξης, θεωρητικά τουλάχιστο, $\mathcal{O}(h^4)$. Για τις επαναληπτικές μεθόδους πήραμε ως αρχική προσέγγιση το μηδενικό διάνυσμα. Αρχικά, και σ΄ αυτήν την περίπτωση τρέξαμε τις άμεσες μεθόδους FFT9 και BCR έτσι ώστε οι εκφράσεις

$$\max_{i=1,\ldots n_1,\, j=1,\ldots,n_2} \left| \frac{u_{i,j}^{(\text{approx})} - u_{i,j}}{u_{i,j}} \right| \tag{7.3.3}$$

χρησιμοποιήθηκαν για να βρεθεί η ακρίβεια της μεθόδου. Στις παραπάνω εκφράσεις το $u^{(\text{approx})}$ είναι ως γνωστό το διάνυσμα της προσέγγισης της λύσης, ενώ το $u$ εκφράζει την ακριβή λύση. Έπειτα, για τις μεθόδους MBIC-CG και ADI-CG χρησιμοποιήσαμε ένα κριτήριο της μορφής

$$\frac{\left\|r^{(k)}\right\|_2}{\|c\|_2} \;<\; tol, \tag{7.3.4}$$

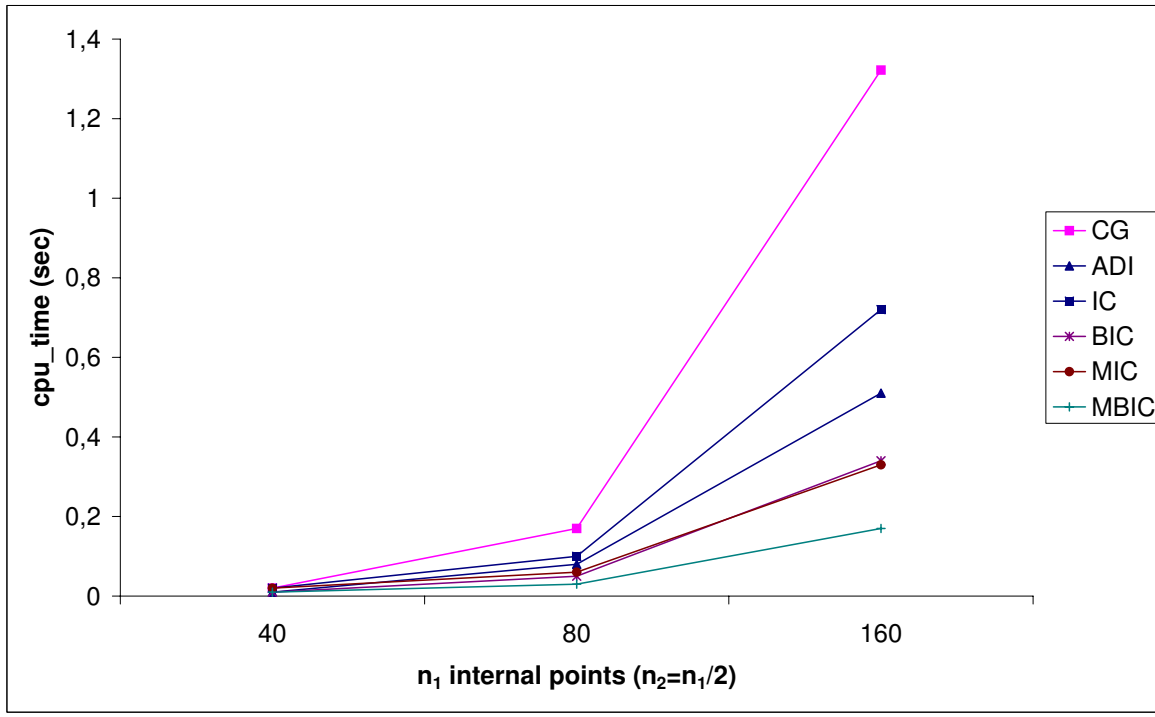
όπου το $tol$ στην (7.3.4) επιλέχτηκε έτσι ώστε να μπορέσουμε να λάβουμε το ίδιο σχετικό απόλυτο σφάλμα με αυτό της έκφρασης (7.3.3) για τις FFT9 και BCR. Να σημειώσουμε εδώ ότι στο πακέτο NSPCG το κριτήριο σταματήματος

στην (7.3.4) είναι ανεξάρτητο από τον προρρυθμιστή που χρησιμοποιήσαμε. Ώς κριτήριο σταματήματος για τη μέθοδο MG χρησιμοποιήσαμε το κλασικό κριτήριο, $\frac{\left\|u^{(k)}-u^{(k-1)}\right\|_2}{\left\|u^{(k)}\right\|_2} < \epsilon$, όπου $u^{(k)}$ είναι η προσέγγιση στην $k$ επανάληψη της λύσης.

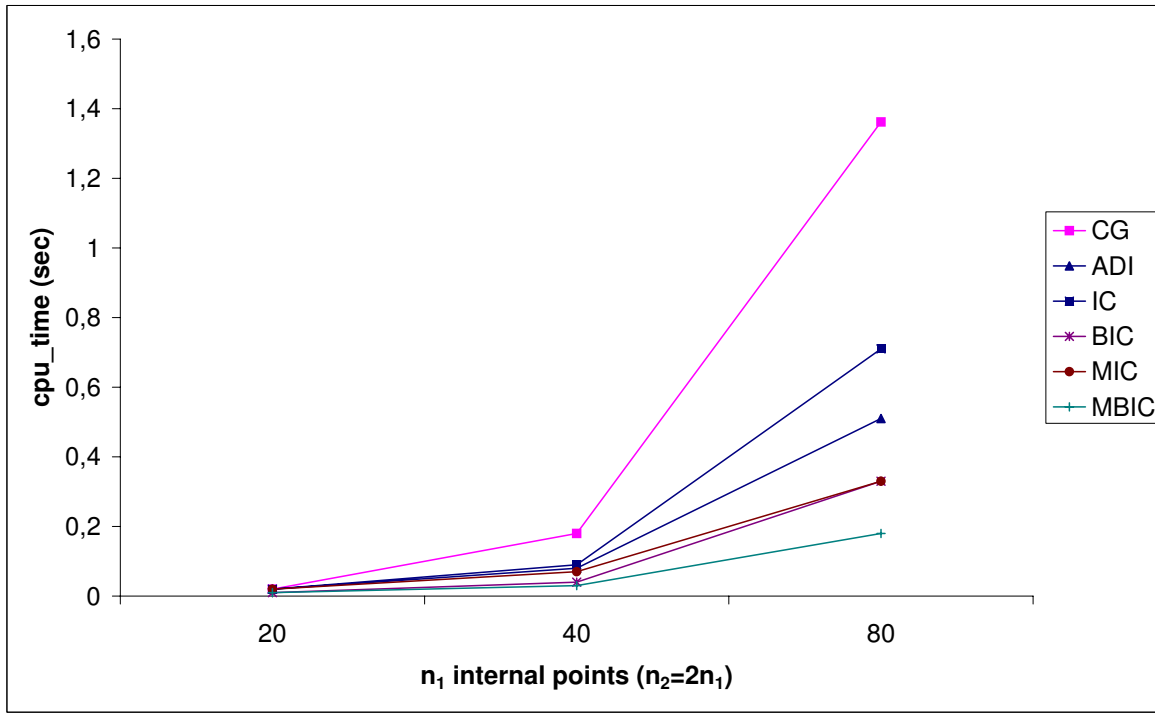
Από τον Πίνακα 7.4 ή από το Σχήμα 7.6 μπορούμε να διαπιστώσουμε ότι για μικρές διαμερίσεις ($8\times16$, $16\times32$, $32\times64$) η μέθοδος ADI-CG είναι συγκρίσιμη με την FFT9, και καλύτερη από την BCR του FISHPACK και την MG του MUDPACK. Για τη διαμέριση $64\times128$ γίνεται συγκρίσιμη με τις BCR και MG αλλά όχι καλύτερη από την FFT9. Παρόλα αυτά, θα πρέπει να τονίσουμε ότι, εάν θεωρήσουμε την ακρίβεια των σφαλμάτων η μέθοδος, FFT9 είναι η λιγότερο καλή με λόγο της τάξης του 10 σε σχέση με αυτά των υπολοίπων, ADI-CG, BCR και MBIC-CG ενώ η MG είναι καλύτερη με λόγο της τάξης του $10^3$. Μία εξήγηση γι΄ αυτήν την συμπεριφορά είναι η χρήση κάποιων τεχνικών βελτίωσης της λύσης. Για τη διαμέριση $128 \times 256$, η FFT9 είναι καλύτερη σε σχέση με το χρόνο. Παρόλα αυτά, από τον Πίνακα 7.4, η BCR είναι η λιγότερο καλή, σε ό,τι αφορά τα σφάλματα, από όλες τις άλλες, ενώ η μέθοδος MG είναι και πάλι η καλύτερη με λόγο της τάξης του $10^3$ σε σχέση με τις άλλες μεθόδους.

| $n_1 \times n_2$ | $8 \times 16$ | | $16 \times 32$ | | $32 \times 64$ | | $64 \times 128$ | | $128 \times 256$ | |
|---|---|---|---|---|---|---|---|---|---|---|
| | time | error | time | error | time | error | time | error | time | error |
| FFT9 | 0 | 1.9E-5 | 0 | 4.9E-5 | 0.01 | 1.1E-4 | 0.02 | 1.01E-3 | 0.09 | 2.0E-3 |
| BCR | 0 | 3.9E-5 | 0.01 | 3.8E-5 | 0.04 | 7.7E-4 | 0.12 | 7.5E-4 | 0.26 | 1.7E-2 |
| MG | 0 | 7.4E-5 | 0 | 9.8E-6 | 0.02 | 6.7E-7 | 0.06 | 3.5E-7 | 0.26 | 5.6E-7 |
| MBIC-CG | 0.01 | 5.41E-5 | 0.02 | 3.72E-5 | 0.12 | 6.34E-5 | 0.31 | 2.2E-4 | 0.94 | 6.3E-4 |
| ADI-CG | 0 | 7.55E-5 | 0 | 3.48E-5 | 0.01 | 9.39E-5 | 0.09 | 4.7E-4 | 0.54 | 7.9E-4 |

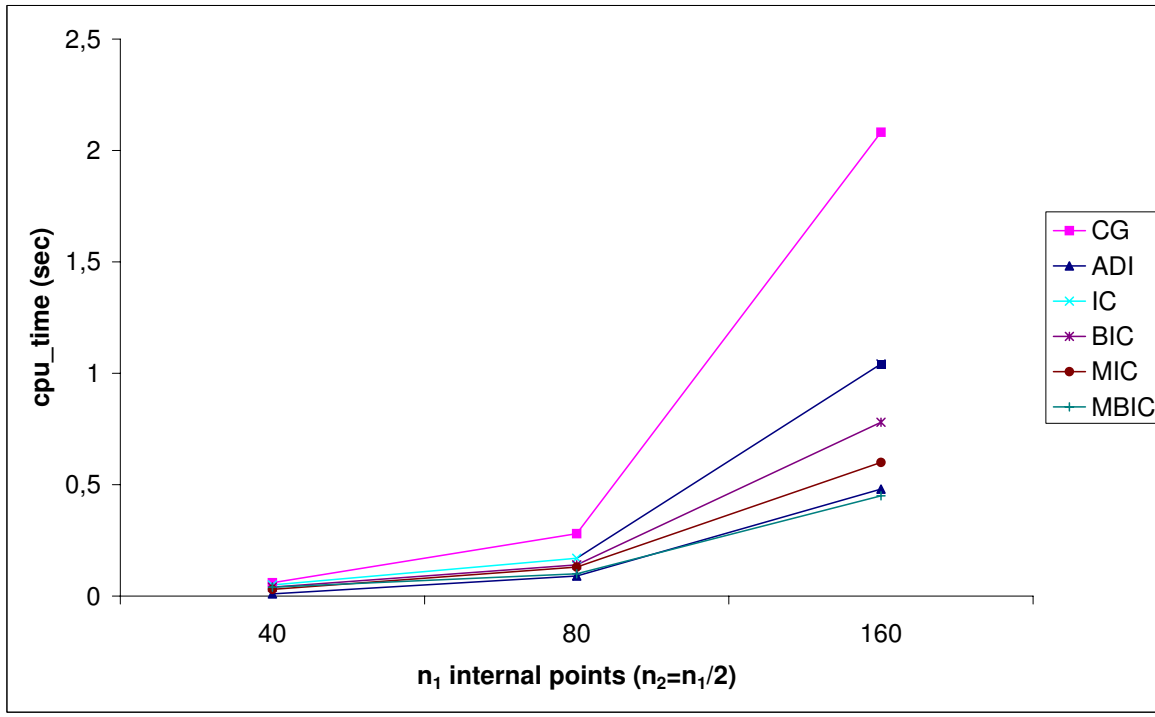
Πίνακας 7.4: Σχήμα Διακριτοποίησης 9−σημείων



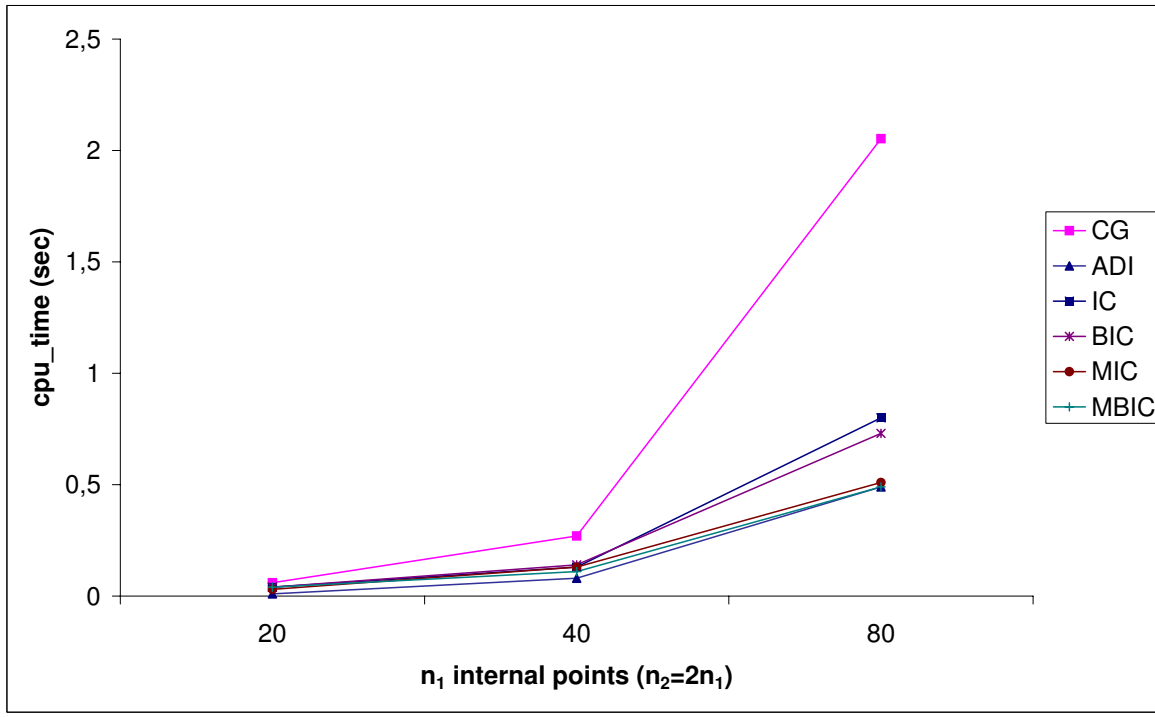

Σχήμα 7.2: Σχήμα Διακριτοποίησης 5−σημείων



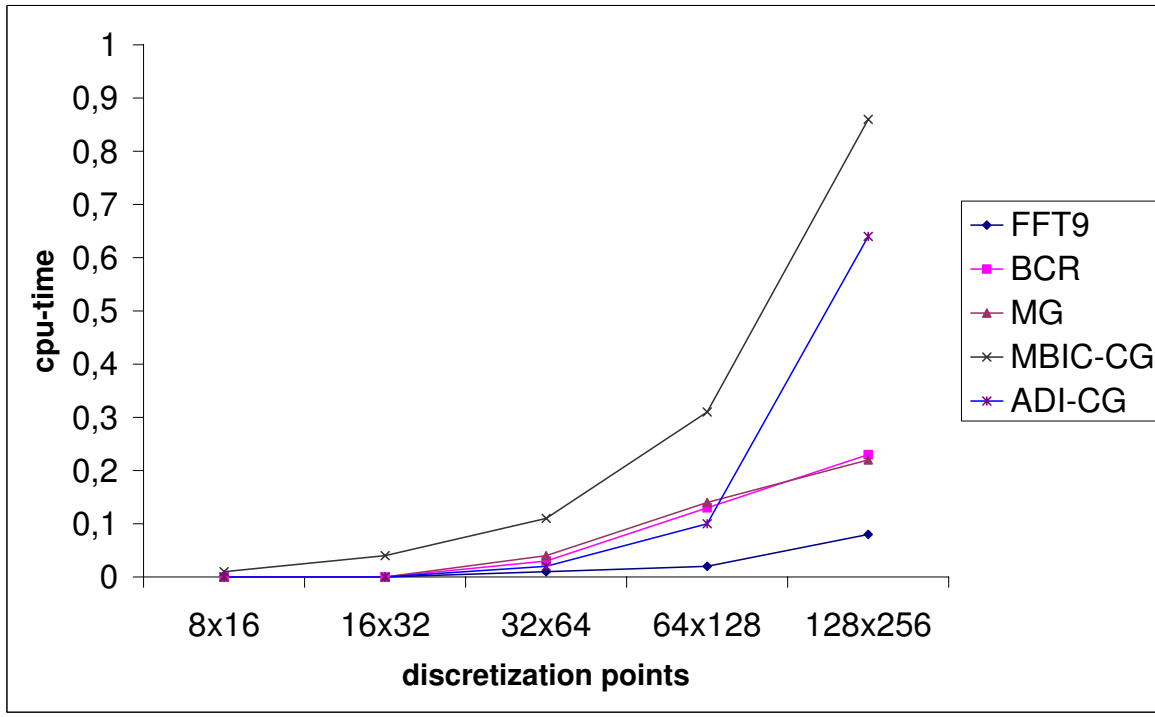

Σχήμα 7.3: Σχήμα Διακριτοποίησης 5−σημείων




Σχήμα 7.4: Σχήμα Διακριτοποίησης 9−σημείων




Σχήμα 7.5: Σχήμα Διακριτοποίησης $9-$σημείων




Σχήμα 7.6: Σχήμα Διακριτοποίησης $9-$σημείων

# Κεφάλαιο 8

# Βέλτιστοι EADI Προρρυθμιστές Μεθόδου Συζυγών Κλίσεων Κυβικής Spline Collocation

## 8.1 Εισαγωγή

### 8.1.1 Κυβικές Splines

Στα προηγούμενα κεφάλαια η Ελλειπτική Μερική Διαφορική Εξίσωση προσεγγίστηκε με τη μέθοδο των Πεπερασμένων Διαφορών. Στο παρόν η προσέγγισή της θα πραγματοποιηθεί με τη μέθοδο της "Collocation". Στη συνέχεια περιγράφεται μία από τις "Collocation" μεθόδους που θα χρησιμοποιηθεί.

Πιο συγκεκριμένα, βασικός στόχος αυτής της παραγράφου είναι να βρεθεί μία ακολουθία από συναρτήσεις $\phi_i$, $i = 0, \ldots, n$, οι οποίες θα προσεγγίζουν με τον πιο κατάλληλο τρόπο δοσμένη συνάρτηση $f$ στη μορφή

$$f(t) = c_0\phi_0(t) + c_1\phi_1(t) + \ldots \phi_n(t), \tag{8.1.1}$$

όπου $c_i$, $i = 0, \ldots, n$, σταθερές, που ικανοποιούν ένα συνδυασμό συνθηκών παρεμβολής και ομαλότητας. Οι συναρτήσεις $\phi_i$, $i = 0, \ldots n$, θεωρείται ότι είναι γραμμικώς ανεξάρτητες σε ένα διάστημα $[\alpha, \beta]$ και είναι τέτοιες ώστε να παράγουν έναν $(n+1)-$διάστατο υπόχωρο του συνόλου των συνεχών συναρτήσεων στο κλειστό διάστημα $[\alpha, \beta]$ ($\mathcal{C}[\alpha, \beta]$). Βασικό παράδειγμα τέτοιων συναρ-

τήσεων βάσης είναι οι καλούμενες $B-$Splines (βλ. de Boor [16]), που θα παρουσιάσουμε και θα χρησιμοποιήσουμε στη συνέχεια.

Θεωρούμε το χώρο $S_3(\Delta)$ ως το χώρο όλων των συναρτήσεων $s(t) \in \mathcal{C}^2[\alpha,\beta]$ που ορίζονται να είναι κυβικά πολυώνυμα σε κάθε υποδιάστημα της μορφής $[t_i, t_{i+1}]$, $i = 0,\ldots,n-1$, του $[\alpha,\beta]$ και έστω ότι $\Delta$ είναι η αντίστοιχη διαμέριση του $[\alpha,\beta]$ με κόμβους τα σημεία $t_i$, $i = 0,\ldots,n$. Εκ του ορισμού του χώρου $S_3(\Delta)$, μπορούμε να παρατηρήσουμε ότι είναι γραμμικός και εφόσον αποτελείται από το σύνολο όλων των κυβικών πολυωνύμων έχει άπειρη διάσταση. Μπορεί όμως να αποδειχτεί ότι κάτω από ορισμένες συνθήκες η συνάρτηση $s(t) \in S_3(\Delta)$ μπορεί να είναι μοναδική. Μία τέτοια περίπτωση μοναδικής συνάρτησης $s(t)$ μπορεί να βρεθεί εάν θέσουμε τους παρακάτω περιορισμούς

$$\begin{aligned} s'(t_0) &= f'(t_0), \\ s(t_i) &= f(t_i),\ i = 0,\ldots,n, \\ s'(t_n) &= f'(t_n). \end{aligned} \tag{8.1.2}$$

Μπορούμε επίσης να αποδείξουμε ότι το πρόβλημα της παρεμβολής με κυβικές Splines μπορεί να λυθεί μοναδικά θεωρώντας επίσης, πέρα από το ότι η εν λόγω spline ανήκει στο χώρο $S_3(\Delta)$ και ικανοποιεί τις συνθήκες της σχέσης (8.1.2), ικανοποιεί και τις συνθήκες

$$s^{(j)}(t_i) = f^{(j)}(t_i),\ i = 1,\ldots,n-1,\ j = 0,1,2 \tag{8.1.3}$$

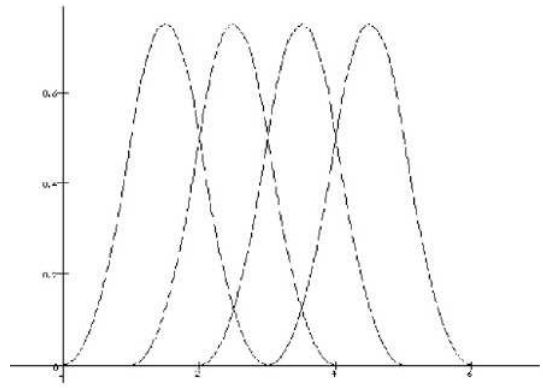
(βλ. Ahlberg, Nilson και Walsh [5]). Στην περίπτωση όπου $j = 2$ τότε αναφερόμαστε στις λεγόμενες "φυσικές" κυβικές Splines. Για ένα γενικότερο ορισμό της συνάρτησης $s(t)$ θεωρούμε επιπλέον δύο κόμβους πέρα από κάθε άκρο της διαμέρισης $\Delta$. Οι κόμβοι οι οποίοι θα προστεθούν είναι οι $t_{-2} < t_{-1} (< t_0)$ από την αριστερή πλευρά της διαμέρισης και οι κόμβοι $(t_n <)\, t_{n+1} < t_{n+2}$ από τη δεξιά πλευρά. Οι κόμβοι αυτοί είναι καθαρά βοηθητικοί και οι τιμές της συνάρτησης $s(t)$ σ' αυτούς καθορίζονται από την αρχή. Με βάση όλα τα παραπάνω ορίζουμε τις συναρτήσεις $B_i(t)$ να έχουν τη μορφή

$$B_i(t) = \frac{1}{h^3}\begin{cases} (t-t_{i-2})^3 & \text{αν } t \in [t_{i-2}, t_{i-1}], \\ h^3 + 3h^2(t-t_{i-1}) + 3h(t-t_{i-1})^2 - 3(t-t_{i-1})^3 & \text{αν } t \in [t_{i-1}, t_i], \\ h^3 + 3h^2(t_{i+1}-t) + 3h(t_{i+1}-t)^2 - 3(t_{i+1}-t)^3 & \text{αν } t \in [t_i, t_{i+1}], \\ (t_{i+2}-t)^3 & \text{αν } t \in [t_{i+1}, t_{i+2}], \\ 0 & \text{αλλού.} \end{cases} \tag{8.1.4}$$

Είναι πολύ εύκολο λοιπόν να βρει κανείς ότι οι τιμές των συναρτήσεων $B_i(t_j)$ δίνονται από τις παρακάτω εκφράσεις

$$B_i(t_j) = \begin{cases} 4 & \text{αν } j = i, \\ 1 & \text{αν } j = i-1 \text{ ή } j = i+1, \\ 0 & \text{αν } j = i-2 \text{ ή } j = i+2, \end{cases} \tag{8.1.5}$$

και $B_i(t) \equiv 0$ για κάθε $t < t_{j-2}$ ή $t > t_{j+2}$. Στο παρακάτω σχήμα παρουσιάζουμε μία τέτοια συνάρτηση.



Σχήμα 8.1: Γραφική Παράσταση B-Spline Συναρτήσεων

Παρατηρούμε, λοιπόν, ότι οι συναρτήσεις $B_i(t)$, $i = -2, \ldots, n+2$, είναι συναρτήσεις με φορέα το διάστημα $[t_{j-2}, t_{j+2}]$. Επίσης, από τον τύπο των παραπάνω συναρτήσεων παρατηρούμε ότι ανήκουν στο χώρο $S_3(\Delta)$ και είναι γραμμικώς ανεξάρτητες. Μπορεί λοιπόν να αποδειχτεί ότι οι παραπάνω συναρτήσεις αποτελούν βάση του χώρου των κυβικών Splines.

## 8.2 Κυβική Spline Collocation - Διακριτοποίηση

Στην παρούσα παράγραφο θα παρουσιάσουμε μία πολύ δυναμική μέθοδο διακριτοποίησης για την επίλυση Ελλειπτικών Μερικών Διαφορικών Εξισώσεων. Η μέθοδος αυτή καλείται Collocation μέθοδος. Θα ξεκινήσουμε παρουσιάζοντας τη γενική ιδέα της μεθόδου αυτής. Στη συνέχεια θα επικεντρωθούμε στην περίπτωση της Collocation μεθόδου με συναρτήσεις βάσης κυβικές Splines,

παρουσιάζοντας αρχικά τη μέθοδο στην περίπτωση των Συνήθων Διαφορικών Εξισώσεων και στη συνέχεια στην περίπτωση των Μερικών Διαφορικών Εξισώσεων, η οποία παρουσιάζει και το μεγαλύτερο ενδιαφέρον.

### 8.2.1 Βασική Ιδέα των Collocation Μεθόδων

Η βασική ιδέα των Collocation μεθόδων στηρίζεται στην παρεμβολή της άγνωστης συνάρτησης του αριστερού μέλους της διαφορικής εξίσωσης με τη βοήθεια ενός συνδυασμού των εικόνων των συναρτήσεων βάσης μέσω του διαφορικού τελεστή της γραμμικής διαφορικής εξίσωσης. Η παραπάνω παραδοχή είναι διαφορετική από αυτήν της περίπτωσης άλλων τεχνικών (π.χ. Πεπερασμένων Στοιχείων), όπου συνδυασμοί των συναρτήσεων της βάσης παρεμβάλλουν τη συνάρτηση λύσης της διαφορικής εξίσωσης που έχουμε να επιλύσουμε. Η παραπάνω ιδέα όταν μπορεί να εφαρμοστεί αποτελεί ένα πολύ ισχυρό μέσο διακριτοποίησης το οποίο εκτός των άλλων είναι και εύκολα υλοποιήσιμο σε σχέση με άλλες τεχνικές διακριτοποίησης.

Έστω $Q$ ένας γραμμικός υπόχωρος του χώρου $L_2(D)$, δηλαδή του χώρου των τετραγωνικά ολοκληρώσιμων συναρτήσεων με πεδίο ορισμού το χωρίο $D$ το οποίο μπορεί να είναι ένας υπόχωρος της πραγματικής ευθείας ή ένας υπόχωρος του επιπέδου. Θεωρούμε επίσης το γραμμικό τελεστή $L$ με πεδίο ορισμού το $Q$ και εικόνα στο ίδιο σύνολο $Q$. Θεωρούμε τον $(n+1)-$διάστατο υπόχωρο του $Q$, τον $X_{n+1} = \text{span}\{\phi_0, \phi_2, \ldots, \phi_n\}$, που παράγεται από τις γραμμικώς ανεξάρτητες συναρτήσεις $\phi_i$, $i = 0, \ldots, n$. Θεωρώντας ότι μας δίνεται η γραμμική εξίσωση

$$Lx = y, \tag{8.2.1}$$

όπου $y$ είναι μία συνάρτηση του $Q$, προσεγγίζουμε τη συνάρτηση $x$ με τη μέθοδο Collocation με τη συνάρτηση

$$x_{n+1}(t) = \sum_{i=0}^{n} a_i \phi_i(t) \tag{8.2.2}$$

στον υπόχωρο $X_{n+1}$, λύνοντας το γραμμικό σύστημα που προκύπτει από την εφαρμογή του τελεστή $L$ στην προσεγγιστική λύση $x_{n+1}$. Παίρνουμε, λοιπόν, το παρακάτω γραμμικό σύστημα

$$Lx_{n+1}(t_j) = \sum_{i=0}^{n} a_i L\phi_i(t_j), \;\; j = 0, \ldots, n, \tag{8.2.3}$$

όπου $t_j$, $j = 0, \ldots, n$, αποτελούν διακριτούς κόμβους στο χωρίο $D$. Λέμε, λοιπόν, ότι η συνάρτηση $x_{n+1}$ collocates τη συνάρτηση $y$ στα σημεία $t_0, t_1, \ldots, t_n$. Από την παραπάνω διαδικασία προκύπτουν μερικά πολύ σημαντικά ερωτήματα,τα οποία και θα μας απασχολήσουν στη συνέχεια. Θα πρέπει να σταθούμε ιδιαίτερα στην κατάλληλη επιλογή της βάσης των συναρτήσεων που θα χρησιμοποιηθούν. Τα κριτήρια μιας επιλογής θα πρέπει να είναι τέτοια ώστε: Οι συνάρτησεις της βάσης που θα επιλεγούν να έχουν αρκετή ομαλότητα ώστε να μπορούμε να εφαρμόσουμε σ΄ αυτές το διαφορικό τελεστή. Επιπλέον, η κατασκευή τέτοιων βάσεων να είναι εύκολη και, γενικότερα, εύκολα επεκτάσιμη στις περισσότερες διαστάσεις. Τελευταίο, αλλά συγχρόνως και πολύ σημαντικό, είναι ότι οι συναρτήσεις βάσης απαιτείται να έχουν όσο το δυνατόν μικρότερο φορέα. Το να έχουν οι συναρτήσεις βάσεις μικρό φορέα έχει άμεση συνέπεια στην πυκνότητα του πίνακα του γραμμικού συστήματος που παράγεται κατά τη μέθοδο της Collocation. Μικρός φορέας έχει ως συνέπεια ο πίνακας του γραμμικού συστήματος να είναι αραιός και επομένως θα είναι δυνατόν να προσφέρονται περισσότερες επιλογές για τις μεθόδους επίλυσης. Κάτι πολύ σημαντικό που πρέπει να προσέξει κάποιος στην επιλογή των βάσεων είναι αυτές να είναι συμμετρικές και θετικά ορισμένες, σε σχέση πάντοτε με το σύστημα που προκύπτει από την εφαρμογή σ΄ αυτές του Διαφορικού Τελεστή.

Έχοντας υπόψη τους παραπάνω περιορισμούς μια επιλογή για τη βάση της Collocation μεθόδου είναι οι $B-$Splines τον ορισμό των οποίων είδαμε στην προηγούμενη παράγραφο. Ο φορέας των συναρτήσεων $B_i(t)$ περιέχει 5 κόμβους της διακριτοποίησης από τους οποίους στους δύο ακραίους η τιμή των συναρτήσεων $B_i(t)$ είναι ίση με 0. Εξαιτίας αυτού το γραμμικό σύστημα που προκύπτει κατά την Collocation μέθοδο είναι ένα συμμετρικό και θετικά ορισμένο τριδιαγώνιο σύστημα. Αυτό το γεγονός μαζί με το ότι οι συναρτήσεις αυτές είναι εύκολα κατασκευάσιμες μας επιτρέπει να έχουμε μια κατάλληλη βάση για τη μέθοδο της Collocation.

Ένα επίσης σημαντικό πλεονέκτημα που θα πρέπει να έχει η βάση, που θα επιλέξουμε, όπως αναφέραμε και προηγουμένως, είναι η εύκολη επεκτασιμότητά της στις περισσότερες διαστάσεις. Η επιλογή των $B-$Splines ως βάσης για την Collocation μέθοδο παρουσιάζει κι αυτό το πλεονέκτημα. Θα προσπαθήσουμε στη συνέχεια να παρουσιάσουμε μία επέκταση της μεθόδου της Collocation στην περίπτωση των Μερικών Διαφορικών Εξισώσεων και πιο συγκεκριμένα στο πρόβλημα της εξίσωσης Poisson με Dirichlet συνοριακές συνθήκες.

## 8.3 Κυβική Spline Collocation - Διακριτοποίηση για ΜΔΕ

Αρχίζοντας, διακριτοποιούμε ομοιόμορφα κάθε διάστημα του $\Omega = \prod \otimes [a_i, b_i]$, $i = 1, 2$, και ορίζουμε την παρακάτω διακριτοποίηση

$$\Delta_i = \left\{ t_l^i = a + lh_i,\, l = -1, \ldots, n_i + 1,\, h_i = \frac{b_i - a_i}{n_i} \right\}.$$

Σ' αυτήν την περίπτωση ο ομοιόμορφος διαμερισμός του χωρίου $\Omega$ είναι $\Delta = \Delta_1 \otimes \Delta_2$. Συμβολίζουμε με $S_{3,\Delta_i} = \mathbf{P}_{3,\Delta_i} \cap \mathbf{C}^2([a_i, b_i])$ το χώρο των μονοδιάστατων splines που ορίζονται από τη διαμέριση $\Delta_i$ των διαστημάτων $[a_i, b_i]$. Τα στοιχεία της βάσης του διδιάστατου χώρου των splines $S_{3,\Delta}$ παράγονται παίρνοντας τα τανυστικά γινόμενα των στοιχείων της βάσης $B_l^i$ του μονοδιάστατου χώρου των splines $S_{3,\Delta_i}$. Η τεχνική της διακριτοποίησης με κυβικές splines δίνει την προσέγγιση στο χωρίο $\Delta$ ως $u_\Delta := u \in S_{3,\Delta_i}$

$$u(\mathbf{x}) = \sum_{i=-1}^{n_1} \sum_{j=-1}^{n_2} U_{ij} B_i^1(x_1) B_j^2(x_2),$$

όπου $\mathbf{x} = (x_1, x_2)$ είναι ένα σημείο του $\Delta$ και $U_{ij}$ είναι οι άγνωστοι συντελεστές της κυβικής spline collocation διακριτοποίησης. Για να προσδιορίσουμε αυτούς τους συντελεστές απαιτούμε από την συνάρτηση $u$ να ικανοποιεί τη Μερική Διαφορική Εξίσωση σε όλα τα σημεία του $\Delta$ και τις συνοριακές συνθήκες σε όλα τα σημεία του συνόρου $\Delta \cap \partial\Omega$ :

$$\begin{aligned} -Lu(x_i, y_j) &= f_{x_i,y_j}, \quad (x_i, y_j) \in \Omega \\ -Lu(x_i, y_j) &= f_{x_i,y_j}, \quad (x_i, y_j) \in \partial\Omega \\ D_x^2 D_y^2 u &= D_y^2 f, \quad (x_i, y_j) \in \Omega_c, \end{aligned} \tag{8.3.1}$$

όπου $\partial\Omega$ και $\Omega_c$ συμβολίζουν τα συνοριακά σημεία και τα σημεία στις τέσσερις κορυφές του ορθογώνιου χωρίου $\Omega$, αντίστοιχα. Με τις εκφράσεις $D_x^2 u$, $D_y^2 u$ συμβολίζουμε τη δεύτερη μερική παράγωγο της $u$ ως προς $x$ και $y$, αντίστοιχα.

Είναι γνωστό από την θεωρία της παρεμβολής με splines ότι η λύση $u$ της Διαφορικής Εξίσωσης ικανοποιεί τις collocation εξισώσεις με τάξη ακρίβειας $\mathcal{O}(h^2)$. Για να μπορέσουμε να έχουμε τη βέλτιστη προσέγγιση με splines τάξης

$\mathcal{O}(h^4)$ μπορούμε να χρησιμοποιήσουμε μία διατάραξη του βασικού τελεστή $L$, την οποία συμβολίζουμε με $L'$, η οποία έχει το επόμενο πλέγμα διακριτοποίησης:

$$\begin{pmatrix} & \frac{\beta}{12}D_y^2S(x_i,y_{j+1}) & \\ \frac{\alpha}{12}D_x^2S(x_{i-1},y_j) & 10\left[\frac{\alpha}{12}D_x^2S(x_i,y_j)+\frac{\beta}{12}D_y^2S(x_i,y_j)\right] & \frac{\alpha}{12}D_y^2S(x_{i+1},y_j) \\ & \frac{\beta}{12}D_y^2S(x_i,y_{j-1}) & \end{pmatrix} \tag{8.3.2}$$

(Για περισσότερες λεπτομέριες βλ. [40].)

Χρησιμοποιώντας τον ορισμό των $B-$splines και το πλέγμα (8.3.2) οι εσωτερικές εξισώσεις, δηλαδή οι αναφερόμενες στους εσωτερικούς κόμβους, της collocation μεθόδου μπορούν να παρασταθούν στη μορφή συστήματος, όπως φαίνεται παρακάτω

$$-AU = F. \tag{8.3.3}$$

Χρησιμοποιώντας μία διάσπαση για τον πίνακα $A$ ως άθροισμα δύο πινάκων $A_1$ και $A_2$, όπως η παρακάτω

$$-(A_1+A_2)U = F, \quad A_1,\, A_2 \in \mathbb{R}^{(n_1-1)(n_2-1)\times(n_1-1)(n_2-1)}, \tag{8.3.4}$$

όπου για την εξίσωση Poisson με Dirichlet συνοριακές συνθήκες οι εκφράσεις για τους πίνακες $A_1$ και $A_2$ δίνονται με μορφή τανυστικών γινομένων και έχουν τις παρακάτω μορφές. Για το σχήμα τάξης $\mathcal{O}(h^2)$ οι πίνακες $A_1$ και $A_2$ δίνονται από τις εκφράσεις

$$A_1 = \frac{1}{6h_1^2}T_4^{n_2-1}\otimes T_{-2}^{n_1-1}, \quad A_2 = \frac{1}{6h_2^2}T_{-2}^{n_2-1}\otimes T_4^{n_1-1},$$

όπου με $T_\alpha$ συμβολίζουμε τον τριδιαγώνιο πίνακα $\text{tridiag}(1,\alpha,1)$. Στην περίπτωση του σχήματος $\mathcal{O}(h^4)$ οι πίνακες $A_1$ και $A_2$ δίνονται από παρόμοιες εκφράσεις

$$A_1 = \frac{1}{72h_1^2}T_4^{n_2-1}\otimes T_{10}^{n_1-1}T_{-2}^{n_1-1}, \quad A_2 = \frac{1}{72h_2^2}T_{10}^{n_2-1}T_{-2}^{n_2-1}\otimes T_4^{n_1-1},$$

με $T_4 = 6I+T_{-2}$ και $T_{10} = 12I+T_{-2}$, όπου $I$ ο μοναδιαίος πίνακας κατάλληλης διάστασης. Έτσι για τους πίνακες $A_1, A_2$ έχουμε τις εκφράσεις:

$$A_1 = -\frac{1}{6h_1^2}(6I^{n_2-1}-T_{-2}^{n_2-1})\otimes T_{-2}^{n_1-1}$$

$$A_2 = -\frac{1}{6h_2^2}T_{-2}^{n_2-1}\otimes(6I^{n_1-1}-T_{-2}^{n_1-1})$$

για το σχήμα τάξης $\mathcal{O}(h^2)$ και

$$A_1 = \frac{1}{72h_1^2}(6I^{n_2-1} + T_{-2}^{n_2-1}) \otimes \left[(12I^{n_1-1} + T_{-2}^{n_1-1})T_{-2}^{n_1-1}\right]$$

$$A_2 = \frac{1}{72h_2^2}\left[(12I^{n_2-1} + T_{-2}^{n_2-1})T_{-2}^{n_2-1}\right] \otimes (6I^{n_1-1} + T_{-2}^{n_1-1})$$

για το σχήμα τάξης $\mathcal{O}(h^4)$. Στη συνέχεια και προκειμένου να απλουστεύσουμε τις πράξεις μας ορίζουμε τους πίνακες

$$\begin{array}{rclcrcl} L_1 &=& 6I^{n_1-1} - T_2^{n_1-1}, & & L_2 &=& 6I^{n_2-1} - T_2^{n_2-1}, \\ M_1 &=& \frac{1}{6h_2^2}T_2^{n_1-1}, & & M_2 &=& \frac{1}{6h_1^2}T_2^{n_2-1}. \end{array}$$

για το σχήμα τάξης $\mathcal{O}(h^2)$ και

$$\begin{array}{rclcrcl} L_1 &=& 6I^{n_1-1} - T_2^{n_1-1}, & & L_2 &=& 6I^{n_2-1} - T_2^{n_2-1}, \\ M_1 &=& \frac{1}{72h_2^2}(12I^{n_1-1} - T_2^{n_1-1})T_2^{n_1-1}, & & M_2 &=& \frac{1}{72h_1^2}(12I^{n_2-1} - T_2^{n_2-1})T_2^{n_2-1}. \end{array}$$

για το σχήμα τάξης $\mathcal{O}(h^4)$, όπου "ειδικά" εδώ $T_2$ =tridiag$(-1, 2, -1)$. Με βάση τους παραπάνω συμβολισμούς έπεται ότι $A_1 = -L_2 \otimes M_2$ και $A_2 = -L_1 \otimes M_1$ και άρα ο πίνακας $A$ γράφεται ως

$$A = -(L_2 \otimes M_1 + M_2 \otimes L_1). \tag{8.3.5}$$

Εφεξής θα περιοριστούμε στη μελέτη του σχήματος τάξης $\mathcal{O}(h^4)$ όπως επίσης και στο γεγονός ότι ο πίνακας $A$ που θα διαχειριστούμε στη συνέχεια θα είναι ο αντίθετος του πίνακα $A$ που έχουμε παραπάνω. Ο βασικός λόγος για κάτι τέτοιο είναι η θετική ορισημότητα που απαιτείται για τη χρήση της μεθόδου των Συζυγών Κλίσεων. Αρχικά θα εξετάσουμε το τι συμβαίνει με το σύνορο του χωρίου $\Omega$ αλλά και με τα σημεία των κορυφών του, τα οποία παρουσιάζουν ιδιαίτερο ενδιαφέρον. Θα ξεκινήσουμε με τα γωνιακά σημεία της διαμέρισης. Από την τελευταία έκφραση της σχέσης (8.3.1), εκφράζοντας τη συνάρτηση $u$ μέσω των συναρτήσεων βάσης $B_i(t)$, $i = 0, 1, \ldots, n$, η έκφραση $D_x^2 D_y^2 u = D_x^2 f$ λαμβάνει τη μορφή των παρακάτω τεσσάρων αλγεβικών εξισώσεων

$$\begin{array}{c} U_{0,0} = \frac{h_1^2 h_2^2}{36} D_x^2 f_{0,0},\ U_{0,n_2} = \frac{h_1^2 h_2^2}{36} D_x^2 f_{0,n_2}, \\ U_{n_1,0} = \frac{h_1^2 h_2^2}{36} D_x^2 f_{n_1,0},\ U_{n_1,n_2} = \frac{h_1^2 h_2^2}{36} D_x^2 f_{n_1,n_2}. \end{array} \tag{8.3.6}$$

Γνωρίζοντας τις τιμές των $U_{0,0}, U_{0,n_2}, U_{n_1,0}$ και $U_{n_1,n_2}$, από τις παραπάνω εξισώσεις, και ορίζοντας τον $(k-1)\times(k-1)$ πίνακα $Q_k$ =tridiag$(1,4,1)$, οι εκφράσεις των Collocation εξισώσεων στο σύνορο του χωρίου $\Omega$ λαμβάνουν τη μορφή συστημάτων, όπως αυτά παρουσιάζονται παρακάτω

$$\begin{aligned} Q_{n_2} v_0^{(1)} &= w_0^{(1)}, \qquad & Q_{n_2} v_{n_1}^{(1)} &= w_{n_1}^{(1)} \\ Q_{n_1} v_0^{(2)} &= w_0^{(2)}, \qquad & Q_{n_1} v_{n_2}^{(2)} &= w_{n_2}^{(2)}, \end{aligned} \tag{8.3.7}$$

όπου

$$\begin{aligned} v_0^{(1)} &= [U_{0,1}, U_{0,2}, \ldots, U_{0,n_2-2}, U_{0,n_2-1}]^T \\ v_0^{(2)} &= [U_{1,0}, U_{2,0}, \ldots, U_{n_1-2,0}, U_{n_1-1,0}]^T \\ v_{n_1}^{(1)} &= [U_{n_1,1}, U_{n_1,2}, \ldots, U_{n_1,n_2-2}, U_{n_1,n_2-1}]^T \\ v_{n_2}^{(2)} &= [U_{1,n_2}, U_{2,n_2}, \ldots, U_{n_1-2,n_2}, U_{n_1-1,n_2}]^T \end{aligned} \tag{8.3.8}$$

ενώ τα αντίστοιχα δεξιά μέλη έχουν τις μορφές

$$\begin{aligned} w_0^{(1)} &= [h_1^2 f_{0,1} - U_{0,0}, h_1^2 f_{0,2}, \ldots, h_1^2 f_{0,n_2-2}, h_1^2 f_{0,n_2-1} - U_{0,n_2}] \\ w_0^{(2)} &= [h_2^2 f_{1,0} - U_{0,0}, h_2^2 f_{2,0}, \ldots, h_2^2 f_{n_1-2,0}, h_2^2 f_{n_1-1,0} - U_{n_1,0}] \\ w_{n_1}^{(1)} &= [h_1^2 f_{n_1,1} - U_{n_1,0}, h_1^2 f_{n_1,2}, \ldots, h_1^2 f_{n_1,n_2-2}, h_1^2 f_{n_1,n_2-1} - U_{n_1,n_2}] \\ w_{n_2}^{(2)} &= [h_2^2 f_{1,n_2} - U_{0,n_2}, h_2^2 f_{2,n_2}, \ldots, h_2^2 f_{n_1-2,n_2}, h_2^2 f_{n_1-1,n_2} - U_{n_1,n_2}] \end{aligned} \tag{8.3.9}$$

Υπολογίζοντας, από τα παραπάνω συστήματα, τις τιμές της συνάρτησης στο σύνορο του $\Omega$ χρησιμοποιούμε αυτές για τη λύση του συστήματος των εσωτερικών εξισώσεων. Το σύστημα των εσωτερικών εξισώσεων αποτελείται από ένα block 5-διαγώνιο πίνακα όπου κάθε ένα από τα τρία εσωτερικά blocks είναι με τη σειρά του 5-διαγώνιος πίνακας ενώ τα δύο ακραία blocks είναι τριδιαγώνια. Ο πίνακας που προκύπτει για τους εσωτερικούς κόμβους της Collocation μεθόδου είναι συμμετρικός και θετικά ορισμένος κάτι που είναι προφανές από τις εκφράσεις των $L_1, L_2, M_1, M_2$ αλλά και τις εκφράσεις των αντίστοιχων ιδιοτιμών από τις οποίες αποδεικνύεται ότι ο πίνακας $A$ είναι θετικά ορισμένος.

Στην επόμενη παράγραφο θα προτείνουμε μία παραλλαγή της ADI-CG μεθόδου, που έχουμε μελετήσει σε προηγούμενα κεφάλαια, για την επίλυση του συστήματος των εσωτερικών εξισώσεων με την προϋπόθεση ότι έχουμε ήδη βρει τις τιμές της συνάρτησης λύσης στο σύνορο του χωρίου από τη λύση των παραπάνω συστημάτων.

## 8.4 Διπαραμετρικό EADI Σχήμα για τις Κυβικές Spline Collocation Εξισώσεις

Στην παράγραφο αυτή θα προσπαθήσουμε να συνδέσουμε τη βέλτιστη ADI-CG μέθοδο ως βασικό επιλυτή για το σύστημα των εσωτερικών κόμβων που προκύπτει από τη διακριτοποίηση του διαφορικού τελεστή με μεθόδους κυβικών spline collocation, όπως αυτές παρουσιάστηκαν αναλυτικά στην προηγούμενη παράγραφο.

Υπενθυμίζουμε ότι εάν έχουμε ένα γραμμικό σύστημα της μορφής $Au = c$, με $A \in \mathbb{C}^{n,n}$ Ερμιτιανό και θετικά ορισμένο και $c \in \mathbb{C}^n$, η μέθοδος των Συζυγών Κλίσεων (CG) είναι η καταλληλότερη για την επίλυση του συστήματος. Βέβαια, όπως έχουμε τονίσει αρκετές φορές σε προηγούμενα κεφάλαια, η CG είναι μία καλή μέθοδος αλλά σχεδόν πάντα χρειάζεται προρρύθμιση ώστε να καταστεί καλύτερη. Στα προηγούμενα κεφάλαια της διατριβής προτείναμε και παρουσιάσαμε αναλυτικά έναν πολύ καλό προρρυθμιστή για τη μέθοδο αυτή.

Για την επίλυση του συστήματος (8.3.3) προτείνουμε το σχήμα των Guittet [27] και Hadjidimos [28], με μια μικρή παραλλαγή, η ιδέα της οποίας στηρίζεται σε μια παραλλαγή του σχήματος των Peaceman-Rachford [20], που προτάθηκε από τον Dyksen [22] και το οποίο παρουσιάζεται παρακάτω

$$\begin{aligned} [(L_1 + r_1 M_1) \otimes M_2]\, U^{(m+\frac{1}{2})} &= [M_1 \otimes (L_2 + r_2 M_2)(L_1 + r_1 M_1) \otimes M_2 - \omega A]\, U^m + \omega b \\ [M_1 \otimes (L_2 + r_2 M_2)]\, U^{m+1} &= U^{(m+\frac{1}{2})}. \end{aligned} \tag{8.4.1}$$

Πολλαπλασιάζοντας το αρχικό σύστημα (8.3.3) επί $M_1 \otimes M_2$ από τα αριστερά, λαμβάνουμε τον επόμενο επαναληπτικό πίνακα για το σχήμα (8.4.1), ο οποίος έχει την ακόλουθη μορφή.

$$I - \omega \widetilde{T} = I - \omega (M_1 \otimes M_2)[M_1^{-1}(M_1^{-1}L_1 + r_1 I_1)^{-1} M_1^{-1} L_1 \otimes (M_2^{-1}L_2 + r_2 I_2)^{-1} M_2^{-1}$$

$$+ \left((M_1^{-1}L_1 + r_1 I_1)^{-1} M_1^{-1}\right) \otimes \left((M_2^{-1}(M_2^{-1}L_2 + r_2 I_2)^{-1} M_2^{-1} L_2)\right)]$$

Γενικός μας σκοπός είναι να βελτιστοποιήσουμε το δείκτη κατάστασης του αρχικού προβλήματος που έχουμε να επιλύσουμε. Όπως γνωρίζουμε κάτι τέτοιο μπορεί να συμβεί χρησιμοποιώντας έναν καλό προρρυθμιστή, στην περίπτωσή μας συμμετρικό και θετικά ορισμένο, έτσι ώστε να μπορέσουμε να ελαχιστοποιήσουμε το δείκτη κατάστασης που είναι και η μεταβλητή εξάρτησης του σφάλματος της μεθόδου των Συζυγών Κλίσεων που χρησιμοποιούμε σε κάθε τέτοια περίπτωση. Τη ζητούμενη ελαχιστοποίηση την επιτυγχάνουμε ελαχιστοποιώντας το δείκτη κατάστασης του παραπάνω πίνακα $\widetilde{T}$ που ουσιαστικά είναι ο πί-

νακας που προκύπτει μετά την προρρύθμιση του αρχικού συστήματος. Στην περίπτωση των συμμετρικών και θετικά ορισμένων πινάκων έχουμε ώς αντικειμενικό σκοπό την ελαχιστοποίηση του λόγου $\frac{\lambda_{\max}}{\lambda_{\min}}$, ο οποίος εκφράζει την ελαχιστοποίηση του δείκτη κατάστασης του προρρυθμισμένου συστήματος, βρίσκοντας την κατάλληλη επιλογή των παραμέτρων επιτάχυνσης $r_1, r_2$. Κάνοντας χρήση της αντιμεταθετικής ιδιότητας των πινάκων $L_1, L_2, M_1, M_2$ σ′ αυτήν την περίπτωση ο πίνακας $\widetilde{T}$ έχει τη μορφή

$$\begin{aligned}\widetilde{T} = (M_1^{-1}L_1 + r_1 I_1)M_1^{-1}L_1 \quad &\otimes \quad (M_2^{-1}L_2 + r_2 I_2)^{-1} \\ +(M_1^{-1}L_1 + r_1 I_1)^{-1} \quad &\otimes \quad (M_2^{-1}L_2 + r_2 I_2)^{-1} M_2^{-1}L_2). \quad (8.4.2)\end{aligned}$$

Θα πρέπει εδώ να τονίσουμε για άλλη μία φορά ότι ο πίνακας $\widetilde{T}$ είναι ο πίνακας του αρχικού συστήματος $A$ προρρυθμισμένος από τα αριστερά με έναν πίνακα $M$ που έχει την έκφραση:

$$M = \frac{1}{\omega}(M_2 \otimes M_1)^{-1}\left[(L_2 + r_2 M_2) \otimes M_1\right]\left[M_2 \otimes (L_1 + r_1 M_1)\right] \qquad (8.4.3)$$

Σημειώνουμε ότι εφεξής σε κάθε μας έκφραση, π.χ. των ιδιοτιμών, δε θα λαμβάνουμε υπόψη την παράμετρο παρεκβολής $\omega$ διότι, όπως είδαμε στη μελέτη για την εύρεση των βέλτιστων $r_1, r_2$, ελαχιστοποιώντας το λόγο $\frac{\lambda_{\max}}{\lambda_{\min}}$ η παράμετρος $\omega$ απαλείφεται.

Αν $\lambda_1 \in \sigma(M_1^{-1}L_1)$ και $\lambda_2 \in \sigma(M_2^{-1}L_2)$, τότε οι ιδιοτιμές του πίνακα $\widetilde{T}$ είναι

$$\lambda(\widetilde{T}) = \frac{\lambda_1 + \lambda_2}{(\lambda_1 + r_1)(\lambda_2 + r_2)} \qquad (8.4.4)$$

ή ισοδύναμα

$$\lambda(\widetilde{T}) = \frac{\frac{1}{\lambda_1} + \frac{1}{\lambda_2}}{(1 + \frac{1}{\lambda_1}r_1)(1 + \frac{1}{\lambda_2}r_2)}. \qquad (8.4.5)$$

Οι εκφράσεις $\mu_1 = \frac{1}{\lambda_1}$ και $\mu_2 = \frac{1}{\lambda_2}$ είναι οι ιδιοτιμές των πινάκων $L_1^{-1}M_1$ και $L_2^{-1}M_2$, αντίστοιχα. Χρησιμοποιώντας τις εκφράσεις $\mu_1, \mu_2$ οι ιδιοτιμές του πίνακα $\widetilde{T}$ δίνονται από τις σχέσεις

$$\lambda(\widetilde{T}) = \frac{\mu_1 + \mu_2}{(1 + \mu_1 r_1)(1 + \mu_2 r_2)}, \qquad (8.4.6)$$

όπου

$$\mu_1^i = \frac{8\left(3 - \sin^2\left(\frac{i\pi}{2(n_1+1)}\right)\right)}{3 - 2\sin^2\left(\frac{i\pi}{2(n_1+1)}\right)}\sin^2\left(\frac{i\pi}{2(n_1+1)}\right),\ i = 1, \ldots, n_1$$

$$\mu_2^i = \frac{8\left(3-\sin^2\left(\frac{i\pi}{2(n_2+1)}\right)\right)}{3-2\sin^2\left(\frac{i\pi}{2(n_2+1)}\right)}\sin^2\left(\frac{i\pi}{2(n_2+1)}\right),\ i=1,\ldots,n_2$$

Οι μέγιστες και οι ελάχιστες τιμές των ιδιοτιμών λαμβάνονται όταν $i = n_1$ ή $i = n_2$ και $i = 1$ αντίστοιχα. Παρακάτω παρουσιάζουμε τη μέγιστη και την ελάχιστη τιμή των ιδιοτιμών αυτών για τις οποίες χρησιμοποιούμε συμβολισμούς ανάλογους με αυτούς που χρησιμοποιήσαμε σε προηγούμενα κεφάλαια.

$$\min(\mu_1)=\alpha_1,\ \max(\mu_1)=\beta_1 \quad \min(\mu_2)=\alpha_2,\ \max(\mu_2)=\beta_2.$$

Οι εκφράσεις λοιπόν των ακραίων ιδιοτιμών είναι οι παρακάτω

$$\begin{aligned}
\alpha_1 &= \frac{2(5+\cos(\frac{\pi}{n_1+1}))}{2+\cos(\frac{\pi}{n_1+1})}\left(1-\cos(\frac{\pi}{n_1+1})\right),\ \alpha_2 = \frac{2(5+\cos(\frac{\pi}{n_2+1}))}{2+\cos(\frac{\pi}{n_2+1})}\left(1-\cos(\frac{\pi}{n_2+1})\right)\\
\beta_1 &= \frac{2(5+\cos(\frac{n_1\pi}{n_1+1}))}{2+\cos(\frac{n_1\pi}{n_1+1})}\left(1-\cos(\frac{n_1\pi}{n_1+1})\right),\ \beta_2 = \frac{2(5+\cos(\frac{n_2\pi}{n_2+1}))}{2+\cos(\frac{n_2\pi}{n_2+1})}\left(1-\cos(\frac{n_2\pi}{n_2+1})\right)
\end{aligned}$$

Γνωρίζοντας τις ακραίες ιδιοτιμές $\alpha_1, \alpha_2, \beta_1, \beta_2$, και την έκφραση (8.4.6) των ιδιοτιμών του επαναληπτικού πίνακα μπορούμε, εκμεταλλευόμενοι τη θεωρία που αναπτύξαμε στο Κεφάλαιο 6 (βλ. [6], [35]) για την εύρεση του βέλτιστου EADI προρρυθμιστή στην περίπτωση των διπαραμετρικών σχημάτων και θέτοντας τη μεταβλητή $\theta$ της γενικής έκφρασης ίση με μηδέν, να έχουμε ότι οι εκφράσεις για τις βέλτιστες παραμέτρους επιτάχυνσης δίνονται από τις παρακάτω σχέσεις

$$\begin{aligned}
r_1^* &= \frac{1}{2}\left\{H+\left[H^2-\frac{2}{\beta_2\alpha_2}[(\alpha_2+\beta_2)H-2]\right]^{\frac{1}{2}}\right\},\\
r_2^* &= \frac{1}{2}\left\{-H+\left[H^2-\frac{2}{\beta_2\alpha_2}[(\alpha_2+\beta_2)H-2]\right]^{\frac{1}{2}}\right\},
\end{aligned} \tag{8.4.7}$$

όπου

$$H := r_1 - r_2 = \frac{2(\beta_1\alpha_1-\beta_2\alpha_2)}{\beta_2\alpha_2(\alpha_1+\beta_1)+\beta_1\alpha_1(\alpha_2+\beta_2)}\ . \tag{8.4.8}$$

Χρησιμοποιώντας τις παραπάνω βέλτιστες εκφράσεις για τις παραμέτρους επιτάχυνσης του σχήματος (8.4.1) είμαστε σε θέση, τουλάχιστον θεωρητικά, να πούμε ότι ο προρρυθμιστής $M$ που εδώ χρησιμοποιήσαμε είναι ο βέλτιστος στην περίπτωση όπου οι παράμετροι $r_1, r_2$ λαμβάνουν τις βέλτιστες τιμές τους.

Στη συνέχεια θα δώσουμε μερικά αριθμητικά παραδείγματα που μας επιτρέπουν να είμαστε αρκετά αισιόδοξοι ως προς τη χρήση του προρρυθμιστή αυτού στην περίπτωση του συστήματος των collocation εξισώσεων για τους εσωτερικούς κόμβους του πλέγματος διακριτοποίησης.

### 8.4.1 Αριθμητικά Παραδείγματα

Στην παράγραφο αυτή του παρόντος κεφαλαίου θα παρουσιάσουμε μερικά αριθμητικά παραδείγματα που προέκυψαν από τη διακριτοποίηση της εξίσωσης Poisson σε ορθογώνιο χωρίο $[a, b] \times [c, d]$ με χρήση της μεθόδου των κυβικών spline collocation τέταρτης τάξης και Dirichlet συνοριακές συνθήκες. Τα αριθμητικά πειράματα υλοποιήθηκαν σε κώδικα Matlab και κάθε σύγκριση έγινε με έτοιμες συναρτήσεις που υπάρχουν στη v 7.1.0.246 R(14) της Matlab χρησιμοποιώντας έναν υπολογιστή Intel Centrino 1.6 σε περιβάλλον Windows xp. Σ′ αυτά τα αριθμητικά παραδείγματα όπως και στα προηγούμενα θεωρήσαμε δεδομένη λύση για την εξίσωσης Poisson και με βάση αυτή κατασκευάσαμε την συνάρτηση του δεξιού μέλους της εξίσωσης. Η συνάρτηση που χρησιμοποιήσαμε είναι

$$u(x, y) = 3e^{x+y}xy(x - 1)(y - 1),$$

και το χωρίο ορισμού είναι το μοναδιαίο τετράγωνο. Από την μορφή της συνάρτησης αυτής έχουμε ότι στο σύνορο του χωρίου η συνάρτηση έχει μηδενική τιμή και έτσι μπορούμε να εφαρμόσουμε την Cubic Spline Collocation μέθοδο με B-Splines συναρτήσεις βάσης στις κλασικές τους εκφράσεις χωρίς μετασχηματισμούς ώστε να ικανοποιούνται μηδενικές συνοριακές συνθήκες. Θα πρέπει επίσης να πούμε ότι χρησιμοποιήσαμε ομοιόμορφη διαμέριση με το ίδιο βήμα και για στις δύο διευθύνσεις. Παρακάτω παρουσιάζουμε τους πίνακες 8.1, 8.2 και τον 8.3 στους οποίους εμφαίνονται τα αριθμητικά αποτελέσματα τα οποία πήραμε από την εκτέλεση των αντίστοιχων πειραμάτων.

| $n_1 = n_2 = 4$ | time | iter | error | rel-res |
|---|---|---|---|---|
| ADI-CG | 0.0012 | 3 | 1.1E-1 | 1.4E-2 |
| J-CG | 0.0016 | 7 | 1.1E-1 | 1.31E-7 |
| IC-CG | 0.11 | 390 | 1.1E-1 | 1,34E-7 |

Πίνακας 8.1: Αριθμητικά αποτελέσματα στην περίπτωση χρήσης του $\mathcal{O}(h^4)$ προρρυθμιστή $M$.

Από μια πρώτη ματιά στα παραπάνω αποτελέσματα διαπιστώνεται ότι οι θεωρητικά καλύτεροι προρρυθμιστές όπως οι ADI, IC σε σύγκριση με τον Jacobi προρρυθμιστή δίνουν ο μεν πρώτος συγκρίσιμα αποτελέσματα ο δε δεύτερος όχι καλά. Βέβαια εάν συνεχίσουμε τα πειράματα για μεγαλύτερα $n_1, n_2$ τα αποτελέσματα που λαμβάνονται είναι λιγότερο καλά και για τον ADI προρρυθμιστή. Φαίνεται λοιπόν ότι αυτό οφείλεται καθαρά και μόνο στις πράξεις που

| $n_1 = n_2 = 8$ | time | iter | error | rel-res |
|---|---|---|---|---|
| ADI-CG | 0.002 | 3 | 3.7E-2 | 1.36E-2 |
| J-CG | 0.003 | 13 | 3.7E-2 | 1.66E-7 |
| IC-CG | 0.17 | 25 | 3.7E-2 | 1,61E-7 |

Πίνακας 8.2: Αριθμητικά αποτελέσματα στην περίπτωση χρήσης του $\mathcal{O}(h^4)$ προρρυθμιστή $M$.

| $n_1 = n_2 = 16$ | time | iter | error | rel-res |
|---|---|---|---|---|
| ADI-CG | 0.04 | 3 | 1.1E-2 | 5.2E-3 |
| J-CG | 0.01 | 25 | 1.1E-2 | 1.6E-7 |
| IC-CG | 0.39 | 38 | 1.1E-2 | 1,6E-7 |

Πίνακας 8.3: Αριθμητικά αποτελέσματα στην περίπτωση χρήσης του $\mathcal{O}(h^4)$ προρρυθμιστή $M$.

εκτελούνται και βέβαια στον τρόπο με τον οποίο η Matlab χειρίζεται αυτές. Οι δύο προρρυθμιστές ADI και IC αποτελούνται από σχετικά αραιούς πίνακες, συγκεκριμένα αναφερόμαστε σε block 5-διαγώνιους με τα τρία εσωτερικά blocks να είναι 5-διαγώνιοι πίνακες ενώ τα δύο ακραία blocks να είναι 3-διαγώνιοι πίνακες. Θα πρέπει εδώ να αναφέρουμε ότι κάνοντας πειράματα με προρρυθμιστή, στο Collocation σύστημα τάξης $\mathcal{O}(h^4)$ με τον πίνακα $M$ που προκύπτει από τάξης $\mathcal{O}(h^2)$ της Collocation διακριτοποίησης, τα αποτελέσματα ήταν περίπου τα ίδια, κάτι που γίνεται φανερό από τη εξέταση των πινάκων 8.4, 8.5 και 8.6 οι οποίοι παρουσιάζονται στη συνέχεια. Όμως εάν χρησιμοποιήσουμε ως προρρυθμιστή το διαγώνιο πίνακα που προκύπτει από τα διαγώνια στοιχεία του πίνακα $M$ τα αποτελέσματα είναι πολύ καλά ακόμα και για μεγάλα $n_1, n_2$, όπως φαίνεται από τους πίνακες 8.7, 8.8, 8.9 και 8.10, που παρουσιάζονται παρακάτω. Είναι λοιπόν προφανές ότι τα αρχικά, όχι και τόσο καλά, αποτελέσματα οφείλονται στον τρόπο με τον οποίο η Matlab διαχειρίζεται αραιούς πίνακες και όχι σ΄ αυτούς καθαυτούς τους προρρυθμιστές.

| $n_1 = n_2 = 4$ | time | iter | error | rel-res |
|---|---|---|---|---|
| ADI-CG | 0.0012 | 3 | 1.1E-1 | 1.4E-2 |
| J-CG | 0.0017 | 7 | 1.1E-1 | 4.7E-7 |
| IC-CG | 0.11 | 390 | 1.1E-1 | 4.99E-3 |

Πίνακας 8.4: Αριθμητικά αποτελέσματα στην περίπτωση χρήσης του $\mathcal{O}(h^2)$ προρρυθμιστή $M$.

| $n_1 = n_2 = 8$ | time | iter | error | rel-res |
|---|---|---|---|---|
| ADI-CG | 0.002 | 3 | 3.7E-2 | 1.36E-2 |
| J-CG | 0.003 | 13 | 3.7E-2 | 1.66E-7 |
| IC-CG | 0.17 | 25 | 3.16E-2 | 2.38E-2 |

Πίνακας 8.5: Αριθμητικά αποτελέσματα στην περίπτωση χρήσης του $\mathcal{O}(h^2)$ προρρυθμιστή $M$.

| $n_1 = n_2 = 16$ | time | iter | error | rel-res |
|---|---|---|---|---|
| ADI-CG | 0.05 | 3 | 1.1E-2 | 5.2E-3 |
| J-CG | 0.01 | 25 | 1.1E-2 | 1.6E-7 |
| IC-CG | 0.39 | 38 | 1.4E-2 | 5.1E-7 |

Πίνακας 8.6: Αριθμητικά αποτελέσματα στην περίπτωση χρήσης του $\mathcal{O}(h^2)$ προρρυθμιστή $M$.

Αξίζει πραγματικά να σημειωθεί εδώ το γεγονός ότι στα προηγούμενα αποτελέσματα, με φράγμα τάξης $10^{-1}$ ή $10^{-2}$ στο σχετικό σφάλμα των υπολοίπων στη μέθοδο των Συζυγών Κλίσεων, παίρνουμε ακριβώς το ίδιο σχετικό απόλυτο σφάλμα με αυτό στην περίπτωση φράγματος τάξης $10^{-6}$ ή $10^{-7}$ για τις άλλες μεθόδους. Το γεγονός αυτό μας κάνει να ελπίζουμε ότι εάν μπορέσουμε να εκμεταλλευτούμε τις δυνατότητες των γρήγορων υπολογισμών της Matlab, τότε πραγματικά θα μπορούμε να έχουμε πολύ καλά αποτελέσματα και από την άποψη του χρόνου υπολογισμών.

| $n_1 = n_2 = 4$ | time | iter | error | rel-res |
|---|---|---|---|---|
| ADI-CG | 0.0012 | 3 | 1.1E-1 | 2.76E-2 |
| J-CG | 0.0017 | 7 | 1.1E-1 | 4.7E-7 |
| IC-CG | 0.11 | 390 | 1.1E-1 | 4.99E-3 |

Πίνακας 8.7: Αριθμητικά αποτελέσματα στην περίπτωση χρήσης ως προρρυθμιστή τη διαγώνιο του πίνακα $M$.

| $n_1 = n_2 = 8$ | time | iter | error | rel-res |
|---|---|---|---|---|
| ADI-CG | 0.002 | 6 | 3.7E-2 | 1.8E-2 |
| J-CG | 0.003 | 13 | 3.7E-2 | 1.66E-7 |
| IC-CG | 0.17 | 25 | 3.16E-2 | 2.38E-2 |

Πίνακας 8.8: Αριθμητικά αποτελέσματα στην περίπτωση χρήσης ως προρρυθμιστή τη διαγώνιο του πίνακα $M$.

| $n_1 = n_2 = 16$ | time | iter | error | rel-res |
|---|---|---|---|---|
| ADI-CG | 0.005 | 11 | 1.1E-2 | 3.3E-2 |
| J-CG | 0.01 | 25 | 1.1E-2 | 1.6E-7 |
| IC-CG | 0.39 | 38 | 1.4E-2 | 5.1E-7 |

Πίνακας 8.9: Αριθμητικά αποτελέσματα στην περίπτωση χρήσης ως προρρυθμιστή τη διαγώνιο του πίνακα $M$.

| $n_1 = n_2 = 32$ | time | iter | error | rel-res |
|---|---|---|---|---|
| ADI-CG | 0.03 | 22 | 3.9E-3 | 3.48E-2 |
| J-CG | 0.07 | 48 | 3.1E-3 | 2.92E-7 |

Πίνακας 8.10: Αριθμητικά αποτελέσματα στην περίπτωση χρήσης ως προρρυθμιστή τη διαγώνιο του πίνακα $M$.

# Κεφάλαιο 9

# Συμπεράσματα – Επίλογος

Περατώνοντας την παρούσα διατριβή κρίνουμε ότι θα ήταν σκόπιμο να δώσουμε κάποια συμπεράσματα και να προτείνουμε γενικεύσεις που μπορούν να πραγματοποιηθούν με βάση τη μέχρι τώρα έρευνά μας και τα προκύψαντα αποτελέσματά της. Αρχικά όμως θα γίνει εν συντομία μια μικρή αναδρομή στα μέχρι τώρα αναφερθέντα στη διατριβή.

Βασικός σποπός μας ήταν η επίλυση του συστήματος $Ax = b$, $A \in \mathcal{C}^{n\times n}$, $b \in \mathcal{C}^n$, όπου $A$ είναι Ερμιτιανός και θετικά ορισμένος πίνακας. Για να λύσουμε το εν λόγω σύστημα προτείναμε ως βασικό επιλυτή την CG και εισαγάγαμε ένα νέο προρρυθμιστή γι΄ αυτήν τον ADI-CG. Ο συγκεκριμένος προρρυθμιστής προέρχεται από ένα επαναληπτικό σχήμα EADI, που αρχικά προτάθηκε από τους Guittet [27] και Hadjidimos [28].

Κατά τη μελέτη του προρρυθμισμένου συστήματος, στις δυο διαστάσεις που εργαστήκαμε, χρειάστηκε να υπολογιστούν οι βέλτιστες παράμετροι επιτάχυνσης $r_1$ και $r_2$ καθώς και η παράμετρος παρεκβολής $\omega$. Βρέθηκαν λοιπόν αναλυτικές εκφράσεις για τις παραμέτρους αυτές σε ό,τι αφορά τα μονοπαραμετρικά αλλά και τα διπαραμετρικά επαναληπτικά σχήματα. Πολλά από αυτά που προέκυψαν αποτελούν πρωτότυπα αποτελέσματα όχι μόνο σε επίπεδο προρρυθμιστή αλλά και σ΄ αυτό καθαυτό των EADI επαναληπτικών σχημάτων.

Στη συνέχεια, έχοντας υπολογίσει τις βέλτιστες παραμέτρους επιτάχυνσης και παρεκβολής, χρησιμοποίησαμε αυτές σε ένα θεωρητικό πρότυπο (μοντέλο) της εξίσωσης Poisson. Παρατηρήσαμε ότι, παρά το γεγονός ότι οι κώδικες προγραμματισμού δεν είχαν γραφεί από "ειδικούς" (προγραμματιστές), τα αποτελέσματα που προέκυψαν ήταν πολύ καλύτερα από αυτά με τους κλασικούς προρρυθμιστές των έτοιμων πακέτων προγραμμάτων αλλά και συγκρίσιμα με αυτά που πήραμε με τις πλέον γνωστές και ευρέως χρησιμοποιούμενες μεθόδους, όπως εί-

ναι οι FFT, Cyclic Reduction και Multigrid μέθοδοι. Εδώ πρέπει να τονίσουμε ότι στην περίπτωση της διακριτοποίησης των 9−σημείων όλα τα αποτελέσματα που δόθηκαν ήταν πρωτότυπα και ακόμη, χωρίς επιπλέον υπολογιστικό κόστος σε σχέση με αυτό της διακριτοποίησης των 5−σημείων. Το γεγονός αυτό, κατά τη γνώμη μας, καθιστά τη μέθοδο ιδιαίτερα ελκυστική και λόγω του μικρού κόστους της αλλά και της ευκολίας και απλότητας κατασκευής και υλοποίησης του κώδικα προγραμματισμού.

Επίσης, θα πρέπει να τονιστεί ότι δε χρησιμοποιήθηκε καμία τεχνική βελτίωσης του σφάλματος και του κόστους των πράξεων κάτι που ίσως να έκανε τη μέθοδο αρκετά πιο γρήγορη. Η επιλογή αυτή θα μπορούσαμε να πούμε ότι ήταν ηθελημένη ώστε να φανεί η πραγματική αξία της μεθόδου που εισαγάγαμε.

Στη συνέχεια, έχοντας υπολογίσει στα Κεφ. 6 και Κεφ. 7 τις βέλτιστες παραμέτρους επιτάχυνσης και παρεκβολής, χρησιμοποιήσαμε αυτές σε ένα πρόβλημα διακριτοποίησης της εξίσωσης Poisson με Κυβικές Spline Collocation μεθόδους και με σχήμα τάξης ακρίβειας $\mathcal{O}(h^4)$ [40]. Γράφοντας ένα απλό κώδικα προγραμματισμού παρατηρήσαμε ότι (βλ. Κεφ. 8) τα αριθμητικά παραδείγματα έδωσαν αποτελέσματα που μπορεί να είναι ικανοποιητικά σε ένα βαθμό αλλά σίγουρα με τη χρήση διαφόρων υπολογιστικών τεχνικών μπορούν να επιτευχθούν πολύ καλύτερα. Τελειώνοντας την περιγραφή των μέχρι τώρα αποτελεσμάτων και συμπερασμάτων της διατριβής θα διατυπώσουμε μερικές προτάσεις για περαιτέρω μελέτη και έρευνα που προκύπτουν από τα μέχρι τώρα επιτευχθέντα αποτελέσματα και συμπεράσματα.

Θα ξεκινήσουμε αρχικά με προτάσεις για περαιτέρω έρευνα σε θεωρητικό επίπεδο.

Εκμεταλλευόμενοι την τεχνική που παρουσιάσαμε στο Κεφ. 6 και στο Κεφ. 7 της διατριβής για την εύρεση των βέλτιστων παραμέτρων επιτάχυνσης και παρεκβολής, μπορούμε να εργαστούμε πάνω στο θέμα αυτό και να δώσουμε βέλτιστες σταθερές παραμέτρους για μονοπαραμετρικό ή τριπαραμετρικό EADI επαναληπτικό σχήμα (δεύτερης και τέταρτης τάξης ακρίβειας) παρόμοιο με αυτό των Guittet [27] και Hadjidimos [28] στην περίπτωση βέβαια της τριδιάστατης εξίσωσης Poisson. Χρησιμοποιώντας ένα σχήμα διακριτοποίησης ανώτερης τάξης, που προτείνεται από το Samarskii (βλ. [48]) για ελλειπτικούς τελεστές, μπορούμε να κατασκευάσουμε έναν πίνακα συντελεστών αγνώστων ο οποίος να έχει την παρακάτω γενική μορφή

$$A = A_1 + A_2 + A_3 + k_1 A_1 A_3 + k_2 A_1 A_2 + k_3 A_2 A_3 + k_4 A_1 A_2 A_3,$$

όπου οι πίνακες $A_i$, $i = 1, 2, 3$, αντιμετατίθενται, ενώ οι συντελεστές $k_1$, $k_2$, $k_3$, $k_4$, έχουν γνωστές εκφράσεις οι οποίες εξαρτώνται από τα βήματα της δι-

ακριτοποίησης $h_1, h_2, h_3$, τις εκφράσεις των οποίων είδαμε στα προηγούμενα κεφάλαια. Στη συνέχεια, επιλέγοντας το επαναληπτικό σχήμα

$$\begin{aligned}
(I + r_1 A_1) u^{(m+\frac{1}{3})} &= ((I + r_3 A_3)(I + r_2 A_2)(I + r_1 A_1) - \omega A)\, u^{(m)} + \omega b \\
(I + r_2 A_2) u^{(m+\frac{2}{3})} &= u^{(m+\frac{1}{3})} \\
(I + r_3 A_3) u^{(m+1)} &= u^{(m+\frac{2}{3})}, \qquad (9.0.1)
\end{aligned}$$

μπορούμε να καταλήξουμε στη μελέτη ένος προβλήματος αντίστοιχου με αυτό των Κεφ. 6 και 7. Η δυσκολία της μελέτης ενός τέτοιου προβλήματος οφείλεται, μεταξύ των άλλων, στις πάρα πολλές διαφορετικές περιπτώσεις που πρέπει να λάβει κάποιος υπόψη, να αναλύσει και να εξετάσει. Όμως, παρά τη δεδομένη δυσκολία που παρουσιάζει το πρόβλημα αυτό, τα μέχρι τώρα ληφθέντα αποτελέσματα είναι ιδιαίτερα ενθαρρυντικά, και υπάρχουν βάσιμες θεωρητικές ενδείξεις ως προς την ορθότητα της πορείας, που πρέπει να ακολουθηθεί, για την εύρεση της λύσης.

Στη συνέχεια θα αναφερθούμε σε κάποιες προτάσεις, που αφορούν στη χρήση του προρρυθμιστή ADI-CG σε διάφορα προβλήματα υπολογιστικής επίλυσης συστημάτων. Αρχικά προτείναμε τον EADI προρρυθμιστή για συμμετρικά και θετικά ορισμένα συστήματα. Όμως η ιδέα ένος τέτοιου προρρυθμιστή μπορεί να εφαρμοστεί και στην περίπτωση συστημάτων που δεν έχουν αυτές τις ιδιότητες, και όπου θα πρέπει να είμαστε σε θέση να γνωρίζουμε τουλάχιστον τις ακραίες ιδιοτιμές του προρρυθμισμένου συστήματος ή φράγματα αυτών (πεδίο τιμών). Κάτι τέτοιο θα μπορούσε να δώσει βέλτιστες παραμέτρους με αποτέλεσμα να έχουμε έναν ισχυρό προρρυθμιστή για μεθόδους όπως οι MINRES, GMRES, BICG και άλλες με σημαντικές αναφορές στην ερευνητική εργασία του Starke (βλ. [50], [51] και [52]). Επίσης, λόγω της απλής μορφής του προρρυθμιστή μπορεί να χρησιμοποιηθεί αυτός ως ομαλοποιητής ή προρρυθμιστής σε ομαλοποιητή μεθόδων Multigrid (βλ. [17]). Πάνω σ΄ αυτό μπορούμε να πούμε ότι η μορφή των ADI μεθόδων, δηλαδή η ιδιαιτερότητά τους στην επίλυση συστημάτων χρησιμοποιώντας εναλλασσόμενες διευθύνσεις, καθιστά τη μέθοδο κατάλληλη για χρήση παράλληλης επεξεργασίας κάτι που ήδη έχει εφαρμοστεί από μηχανικούς για την λύση πολλών προβλημάτων (βλ. [46], [47] κ.τ.λ.). Βέβαια, σε καμιά από αυτές τις περιπτώσεις εφαρμογής δεν υπάρχει θεωρητικός τρόπος εύρεσης των βέλτιστων παραμέτρων επιτάχυνσης και παρεκβολής αλλά γίνεται καθαρά χρήση υπολογιστικών βέλτιστων παραμέτρων και σχεδόν σε κάθε περίπτωση εφαρμόζεται το σχήμα των Peaceman-Rachford [45].

Σε ό,τι αφορά τώρα την εφαρμογή σε Κυβικές Spline Collocation μεθόδους,

παρόλο που τα μέχρι στιγμής αποτελέσματα, όπως τα περιγράψαμε και στο Κεφ. 8, δεν είναι και τα πλέον ικανοποιητικά, η ανάλυσή μας μπορεί να γενικευτεί και στην περίπτωση των Ορθογώνιων Spline Collocation μεθόδων (βλ. [13]) και να δώσει ακόμα καλύτερα αποτελέσματα λόγω του συνδυασμού ενός ισχυρότατου εργαλείου διακριτοποίησης, όπως αυτό των μεθόδων των Ορθογώνιων Spline Collocation, και ενός πολύ καλού προρρυθμιστή για τη μέθοδο επίλυσης του συστημάτος στους εσωτερικούς κόμβους.

Ό,τι περιγράψαμε μέχρι τώρα αφορά στην περίπτωση των σταθερών παραμέτρων επιτάχυνσης και παρεκβολής ("στατικά" EADI επαναληπτικά σχήματα). Άλλωστε και όλα τα θεωρητικά μας αποτελέσματα αφορούσαν σ΄ αυτήν την περίπτωση. Υπάρχει όμως, όπως γνωρίζουμε, και η περίπτωση των μεταβλητών παραμέτρων ("μη στατικά" EADI επαναληπτικά σχήματα). Σ΄ αυτή την περίπτωση και παρόλο που έγινε μία πρώτη προσπάθεια δεν μπορέσαμε να έχουμε αποτελέσματα αφού η τεχνική των Jordan-Wachspress (βλ. [58], [62]) με τη συγκεκριμένη μορφή των Ελλειπτικών Συναρτήσεων του Jacobi δεν είναι εύκολο να εφαρμοστεί για το σχήμα των Guittet [27] και Hadjidimos [28], που χρησιμοποιήσαμε. Ο βασικός λόγος που δεν μπορούμε να εφαρμόσουμε αυτήν την τεχνική είναι η μέχρι τώρα αδυναμία μας να "προτείνουμε" μια συνάρτηση αντίστοιχη αυτής του Jordan (βλ. [58], [62]). Όμως, μπορούμε να υπερνικήσουμε το πρόβλημα αυτό χρησιμοποιώντας μεταβλητές παραμέτρους της μορφής αυτών των Douglas-Rachford [20] ή του Douglas [19] ή των Samarskii-Andreev [49], όπως αυτές τροποποιήθηκαν από το Hadjidimos [29]. Θεωρούμε ότι η χρήση αυτών των παραμέτρων θα μπορέσει να δώσει πολύ καλύτερα αποτελέσματα σε σχέση με αυτά που έχουμε ήδη λάβει.

Σίγουρα, λοιπόν, τα αποτελέσματα μας είναι αρκετά ενθαρρυντικά και μπορούν να αποτελέσουν την απαρχή μιας νέας προσπάθειας για την εύρεση περισσότερων και καλύτερων μεθόδων. Άλλωστε, πολλές από τις λεγόμενες "παλιές", και για κάποιους "ξεπερασμένες", μεθόδους και τεχνικές επανέρχονται (βλ. Golub [25]) και μπορούν να δώσουν πραγματικά πολύ σημαντικά θεωρητικά αλλά και πρακτικά αποτελέσματα σε πολλά σύγχρονα προβλήματα που μας απασχολούν, όπως επανειλημμένα έχει παρατηρηθεί.

# Βιβλιογραφία


[1] M. Abramowitz and I.A. Stegun. *Handbook of Mathematical Functions With Formulas, Graphs, and Mathematical Tables.* Applied Mathematics Series - 55, National Bureau of Standards, Issued June 1964. Tenth Printing, December 1972.

[2] N.I. Achieser. *Theory of Approximation.* Frederick Ungar Publishing Co., New York, 1956.

[3] J.C. Adams. *Mudpack: Multigrid Portable FORTRAN Software for the Efficient Solution of Linear Elliptic Partial Differential Equations.* Appl. Math. Comput., 34 (1989), 113–146.

[4] O. Axelsson and A. Barker. *Finite Element Solution of Boundary Value Problems. Theory and Computation.* Academic Press, Orlando, FI., 1984.

[5] J.H. Ahlberg, E.N. Nilson and J.L. Walsh. *The Theory of Splines and Their Applications.* Academic Press, New York, 1967.

[6] G. Avdelas and A. Hadjidimos. *Optimum Biparametric E.A.D.I. and A.D.P. Schemes for the Numerical Solution of* $2-$*Dimensional Elliptic Problems.* Rev. Roum. Math. Pures et Appl., XXIV (1979), 999–1012.

[7] Z.-Z. Bai, G.H. Golub and M. Ng. *"Hermitian/Skew-Hermitian Splittings"* (see [25]).

[8] Z.-Z. Bai, G.H. Golub and M. Ng. *"Normal/Skew-Hermitian Splittings"* (see [25]).

[9] Z.-Z. Bai, G.H. Golub and M. Ng. *Hermitian and Skew-Hermitian Splitting Methods for Non-Hermitian Positive Definite Linear Systems.* SIAM J. Matrix Anal. Appl., 24 (2003), 603–626.

[10] Z.-Z. Bai, G.H. Golub and J.-Y. Pan. *Preconditioned Hermitian and Skew-Hermitian Splitting Methods for Non-Hermitian Positive Semidefinite Linear Systems.* Numer. Math., 98 (2004), 1–32.

[11] Z.-Z. Bai, G.H. Golub, L.-C. Lu and J.-F. Yin. *Block Triangular and Skew-Hermitian Splitting Methods for Positive-Definite Linear Systems.* SIAM J. Sci Comput., 26 (2005), 844–863.

[12] R. Barrett, M. Berry, T.F. Chan, J. Demmel, J. Donato, J. Dongarra, 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.

[13] B. Bialecki and R.I. Fernandes. *Orthogonal Spline Collocation Laplace-Modified and Alternating-Direction Methods for Parabolic on Rectangles.* Mathematics of Computation, 60 (1993), 545–573.

[14] G. Birkhoff and R.E. Lynch. *Numerical Solution of Elliptic Problems.* SIAM, Philadelphia, PA, 1984.

[15] S.D. Conte and R.T. Dames. *An Alternating Direction Methods for Solving the Biharmonic Equation.* M.T.A.C., 12 (1958), 198–205.

[16] C. De Boor. *A Practical Quide to Splines.* Springer, Berlin, 1978.

[17] C.C. Douglas, S. Malhotra and M.H. Schultz *Parallel Multigrid with ADI-like Smoothers in Two Dimensions.* 5th European Multigrid Conference Special Topics and Applications, 1998.

[18] J. Douglas, Jr. *Alternating Direction Iteration for Mildy Nonlinear Elliptic Defference Equations.* Numer. Math., 3 (1961), 92–98.

[19] J. Douglas, Jr. *Alternating Direction Methods for Three Space Variables.* Numer. Math., 4 (1962), 41–63.

[20] J. Douglas, Jr. and H.M. Rachford. *On the Numerical Solution of Heat Conduction Problems in Two and Three Space Variables.* Trans. Amer. Math. Soc., 82 (1956), 421–439.

[21] E.G. D'Yakonov. *The Method of Variable Directions in Solving Systems of Finite Difference Equations.* Soviet Math. Dokl., 2 (1961), 577–580. TOM 138, 271–274.

[22] W.R. Dyksen. *Tensor Product Generalized Adi Methods for Separable Elliptic Problems.* SIAM J. Numer. Anal., 24 (1987), 59–76.

[23] N.S. Ellner and E.L. Wachspress. *Alternating Direction Implicit Iteration for Systems With Complex Spectra.* SIAM J. Numer. Anal., 28 (1991), 859–870.

[24] N. Gastinel. *Sur le Meilleur Chois des Paramètres de Surrelaxation.* Chiffres, 5 (1962), 109–126.

[25] G.H. Golub. *Solution of Non-Symmetric, Real Positive Linear Systems.* Paper presented at the Milovy 2002 Conference on "Computational Linear Algebra with Applications", August 5–9, 2002, Milovy, Czech Republic.

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

[27] J. Guittet. *Une Nouvelle Méthode de Directions Alternées à q Variable.* J. Math. Anal. Appl., 17 (1967), 199–213.

[28] A. Hadjidimos. *Extrapolated Alternating Direction Implicit Methods for the Numerical Solution of Elliptic Partial Differential Equations.* Ph.D. Dissertation, University of Liverpool, Liverpool, England, U.K., 1968.

[29] A. Hadjidimos. *Extrapolated Alternating Direction Implicit Iterative Methods.* BIT, 10 (1969), 465–475.

[30] A. Hadjidimos. *The Numerical Solution of a Model Problem Biharmonic Equation by Using Extrapolated Alternating Direction Implicit Iterative Methods.* Numer. Math., 17 (1971), 301–317.

[31] A. Hadjidimos, E.N. Houstis, J.R. Rice and E.A. Vavalis. *Iterative Line Cubic Spline Collocation Methods for Elliptic Partial Differential Equations in Several Dimensions.* SIAM J. Sci. Comput., 14 (1993), 715–734.

[32] A. Hadjidimos, E.N. Houstis, J.R. Rice and E.A. Vavalis. *Analysis of Iterative Line Spline Collocation Methods for Elliptic Partial Differential Equations.* SIAM J. Matrix Anal. Appl., 21 (1999), 508–521.

[33] A. Hadjidimos and K. Iordanidis. *Solving Laplace's Equation in a Rectangle by Alternating Direction Implicit Methods.* J. Math. Anal. Appl., 48 (1974), 353–367.

[34] A. Hadjidimos and M. Lapidakis. *Optimal Alternating Direction Implicit Preconditioners for Conjugate Gradient Methods.* J. Appl. Math. Comput., 183 (2006), 559–574.

[35] A. Hadjidimos and M. Lapidakis. *Stationary Biparametric ADI Preconditioners for Conjugate Gradient Methods.* J. Comput. Appl. Maths, 205 (2007), 364–381.

[36] P.R. Halmos. *Finite-Dimensional Vector Spaces.* 2nd ed., Van Nostrand, Princeton, NJ, 1958.

[37] R.A. Horn and C.R. Johnson. *Matrix Analysis.* Cambridge University Press, Cambridge, UK, 1985.

[38] R.A. Horn and C.R. Johnson. *Topics in Matrix Analysis.* Cambridge University Press, Cambridge, UK, 1991.

[39] E.N. Houstis and T.S. Papatheodorou. *High Order Fast Elliptic Equation Solver.* ACM Trans. Math. Software, 5 (1979), 431–441.

[40] E.N. Houstis, E.A. Vavalis and J.R.Rice *Convergence of $O(h^4)$ Cubic Spline Collocation Methods for Elliptic Partial Differential Equations.* SIAM J. Numer. Anal., 25 (1988), 54–74.

[41] L.V. Kantorovich and V.I. Krylov. *Approximate Methods of Higher Analysis.* (Translated from Russian by C.D. Benster.) Interscience Publishers, Inc., New York, 1958.

[42] R.E. Lynch, J.R. Rice and D.H. Thomas. *Tensor Product Analysis of Alternating Direction Implicit Methods.* SIAM J. Appl. Math., 13 (1965), 995–1006.

[43] J.A. Meijerink and H.A. Van der Vorst. *An Iterative Solution Method for Linear Systems of which the Coefficient Matrix is Symmetric M-Matrix*. Math. Comput., 31 (1977), 148–162.

[44] T.C. Oppe, W.D. Joubert and D.R. Kincaid. *NSPCG User's Guide, Version 1.0: Package for Solving Large Sparse Linear Systems by Various Iterative Methods*. T.R. CNA-216, Center for Numerical Analysis, University of Texas at Austin, Austin, TX, April 1988.

[45] D.W. Peaceman and H.H. Rachford, Jr. *The Numerical Solution of Parabolic and Elliptic Differencial Equations*. SIAM J. Appl. Math., 3 (1955), 28–41.

[46] A. Povitsky. *Parallel ADI solver based on processor scheduling*. J. Appl. Math. Comput., 133 (2002), 43–81.

[47] I. V. Schevtschenko. *A Parallel ADI Method for a Nonlinear Equation Describing Gravitational Flow of Ground Water*. Lecture Notes in Computer Science, 2073 (2001), 904–910.

[48] A.A. Samarskii. *Theory of Difference Schemes*. Nauka, Moscow, 1977 (in Russian). (English translation: Marcel Dekker, Inc., New York, 2001.)

[49] A.A. Samarskii and V.B. Andreev. *Alternating Direction Iterational Schemes Schemes for the Numerical Solution of the Dirichlet Problem*. Z. Vycisl. Mat. i. Mat. Fi., 4 (1964), 1025–1036.

[50] G. Starke. *Optimal Alternating Direction Implicit Parameters for Nonsymmetric Systems of Linear Equations*. SIAM J. Numer. Anal., 28 (1991), 1431–1445.

[51] G. Starke. *Fields of Values and the ADI Method for Non-Normal Matrices*. Linear Algebra Appl., 180 (1993), 199–218.

[52] G. Starke. *Alternating Direction Preconditioning for Nonsymmetric Systems of Linear Equations*. SIAM J. Sci. Comput., 15 (1994), 369–384.

[53] P. Swarztrauber and R. Sweet. *Efficient Fortran Subprograms for the Solution of Elliptic Partial Differential Equations*. NCAR Tech. Note, NCAR-TN/IA-109, (1975), 135–137.

[54] P. Tsompanopoulou and E. Vavalis. *ADI Methods for Cubic Spline Collocation Discretization of Elliptic PDEs.* SIAM J. Sci. Comput., 19 (1998), 341–363.

[55] R.S. Varga. *Matrix Iterative Analysis.* 2nd revised and expanded edition, Springer, Berlin, 2000.

[56] E.L. Wachspress. *CURE: A Generalized Two-Space-Dimension Multigroup Coding for the IBM-704.* Report KAPL-1724, Knolls Atomic Energy Laboratory, Schenectady, New York, 1957.

[57] E.L. Wachspress. *Optimum Alternating-Direction-Implicit Iteration Parameters for a Model Problem.* J. Soc. Indust. Appl. Math., 10 (1962), 339–350.

[58] E.L. Wachspress. *Extended Application of Alternating Direction Implicit Model Problem Theory.* J. Soc. Indust. Appl. Math., 11 (1963), 994–1016.

[59] E.L. Wachspress. *Iterative Solution of Elliptic Systems.* Prentice Hall, Englewood Cliffs, NJ, 1966.

[60] E.L. Wachspress. *Generalized ADI Preconditioning.* Computers Math. Applic., 10 (1984), 457–461.

[61] E.L. Wachspress. *Three-Variable Alternating-Direction-Implicit Iteration.* Computers Math. Applic., 27 (1994), 1–7.

[62] E.L. Wachspress. *The ADI Model Problem.* Monograph, Windsor, CA, 1995.

[63] E.L. Wachspress and G.J. Habetler. *An Alternating-Direction-Implicit Iteration Technique.* J. Soc. Indust. Appl. Math., 8 (1960), 403–424.

[64] D.M. Young. *Iterative Solution of Large Linear Systems.* Academic Press, New York, 1971.

[65] E. Zolotareff. *Anwendung der Elliptischen Funkctionen auf Probleme über Funktionen, die von Null am Wenigstem Oder Meisten Abweichen.* Abh. St. Petersburg 30, 1877.