Επίλυση προβλημάτων Ελλειπτικών Μερικών Διαφορικών Εξισώσεων με την μέθοδο των Πεπερασμένων Στοιχείων καί εφαρμογές στις Εξισώσεις Μεταφοράς-Διάχυσης.

Διπλωματική εργασία

Χαβέλια Ελένη

Επιβλέπων Καθηγητής : Κ.Χρυσαφίνος
Επίκουρος Καθηγητής ΕΜΠ

Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Εφαρμοσμένων Μαθηματικών και Φυσικών Επιστημών

Αθήνα 2011

Περιεχόμενα

Ευχαριστίες

Ολοκληρώνοντας την διπλωματική μου εργασία θα ήθελα αρχικά να ευχαριστήσω θερμά τον επίκουρο καθηγητή του ΕΜΠ κ.Χρυσαφίνο Κωσταντίνο για την άριστη συνεργασία που είχαμε καί για την πολύτιμη βοήθεια που προσέφερε σε όλα τα ζητήματα που προέκυψαν κατά την διάρκεια της εργασίας. Θα ήθελα επίσης να ευχαριστήσω την οικογένεια μου για την αγάπη καί την υποστήριξη που μου έχουν προσφέρει. Τέλος δεν θα μπορούσα να ξεχάσω τους φίλους καί τις φίλες που με βοήθησαν καί με στήριξαν όλο αυτό το διάστημα καί ειδικότερα τον μεταπτυχιακό φοιτητή Ν.Παλληκαράκη για την βοήθεια του στη χρήση της γλώσσας TEX.

Κεφάλαιο 1

Εισαγωγή

1.1 Σκοπός της εργασίας

Σκοπός της διπλωματικής εργασίας είναι η μελέτη καί η επίλυση των εξισώσεων μεταφοράς-διάχυσης με χρήση προσεγγιστικών μεθόδων. Οι εξισώσεις αυτές χρησιμοποιούνται για την περιγραφή προβλημάτων μεταφοράς ενέργειας σε πολλούς επιστημονικούς τομείς ,όπως στον τομέα της μηχανικής της φυσικής κ.τ.λ. Ειδικά στην παρούσα εργασία θα ασχοληθούμε με προβλήματα όπου η μεταφορά υπερισχύει της διάχυσης. Τα προβλήματα αυτά παρουσιάζουν ιδιαίτερο ενδιαφέρον διότι εμφανίζουν προβλήματα όταν έχουμε να κάνουμε με προβλήματα που ορίζονται σε περισσότερες από μία διαστάσεις, πράγμα το οποίο αντιμετωπίζεται όπως θα δούμε καί αναλυτικά στο Κεφάλαιο 5.

Πριν παρουσιάσουμε όμως το πρόβλημα μεταφοράς-διάχυσης θα ασχοληθούμε με την ανάλυση των προσεγγιστικών μεθόδων για τα προβλήματα συνοριακών τιμών μερικών διαφορικών εξισώσεων *ελλειπτικού τύπου*. Με τον όρο ελλειπτικού τύπου εννοούμε εξισώσεις που περιγράφουν φυσικές καταστάσεις ,οι οποίες δεν εξελίσσονται στον χρόνο, δηλαδή καταστάσεις ισορροπίας. Έχουν αναπτυχθεί διάφορες κατηγορίες προσεγγιστικών μεθόδων για την επίλυση τόσο συνήθη διαφορικών εξισώσεων όσο καί μερικών διαφορικών εξισώσεων. Οι βασικότεροι μέθοδοι είναι, η μέθοδος των Πεπερασμένων Διαφορών καί η μέθοδος των Πεπερασμένων Στοιχείων με σημαντικότερο εκπρόσωπο την μέθοδο *Galerkin*, με τις οποίες καί θα ασχοληθούμε στην παρούσα εργασία. Επίσης άλλοι μέθοδοι που χρησιμοποιούνται είναι η μέθοδος Σκόπευσης καί η μέθοδος Ταξινόμησης.

1.2 Δομή της εργασίας

Για την παρουσίαση του θέματος χρησιμοποιούμε ως κύρια αναφορά το βιβλίο [1,Κεφ. 12,13,14] Στο **Κεφάλαιο 2.** γίνεται παρουσίαση του απλού συνοριακού προβλήματος. Το πρόβλημα αυτό επιλύεται με ολοκλήρωση κατά μέρη καί χρήση της συνάρτησης *Green*. Η λύση του τώρα ικανοποιεί κάποιες χαρακτηριστικές ιδιότητες, όπως *μοναδικότητα*, *μονοτονία* καί την *αρχή*

του μεγίστου.
Στην συνέχεια στο **Κεφάλαιο 3.** το πρόβλημα συνοριακών τιμών επιλύεται προσεγγιστικά με την μέθοδο των Πεπερασμένων Διαφορών. Αρχικά από το συνεχές πρόβλημα μεταβαίνουμε στο αντίστοιχο διακριτοποιημένο καί αναλύουμε την *ευστάθεια* την *συνέπεια* καί κατ'επέκταση την *σύγκλιση* της μεθόδου. Αναλυτικότερα, για να δείξουμε την ευστάθεια της παραπάνω μεθόδου αρκεί να δείξουμε ότι η προσεγγιστική λύση είναι φραγμένη. Όσο αναφορά την συνέπεια, ο ρόλος της είναι να μειώσει το σφάλμα που προκύπτει λόγω της μετάβασης από το συνεχές πρόβλημα στο αντίστοιχο διακριτοποιημένο καί τέλος η σύγκλιση της μεθόδου μας δείχνει κατά πόσο η προσεγγιστική λύση" πλησιάζει" την αντίστοιχη ακριβή λύση του προβλήματος.

Στο **Κεφάλαιο 4.** τώρα παρουσιάζουμε την μέθοδο $Galerkin$. Εδώ θεωρούμε ένα πιο γενικό πρόβλημα συνοριακών τιμών καί σε αντίθεση με το Κεφάλαιο 3., όπου περάσαμε κατευθείαν στη διακριτοποίηση του, εδώ προσδιορίζουμε πρώτα την *ασθενή διατυπώση* του. Στην συνέχεια διακριτοποιούμε το πρόβλημα χρησιμοποιώντας την μέθοδο των Πεπερασμένων Στοιχείων ,όπου η προσεγγιστική λύση δίνεται ως ένας γραμμικός συνδιασμός συνεχών τμηματικών πολυωνύμων βαθμού $K \geq 1$. Για ευκολότερη περιγραφή ορίζουμε τους διανυσματικούς χώρους $X_h^1, X_h^2, ..., X_h^k$, οι οποίοι περιέχουν τα αντίστοιχα τμηματικά πολυώνυμα καί οι οποίοι παράγονται από τις αντίστοιχες συναρτήσεις βάσης φ_i. Όσο αναφορά τις συναρτήσεις αυτές γίνεται αναλυτική παρουσίαση αυτών για την μορφή καί τις ιδιότητες τους με ιδιαίτερο ενδιαφέρον να παρουσιάζεται στην περίπτωση του χώρου X_h^1 όπου οι φ_i είναι οι γνωστές συναρτήσεις *στέγες*. Τέλος όπως καί στο Κεφάλαιο 3. γίνεται ανάλυση της ευστάθειας καί της σύγκλισης της μεθόδου.

Στο **Κεφάλαιο 5.** μελετάμε τις εξισώσεις μεταφοράς-διάχυσης καί ειδικότερα τα προβλήματα όπου η μεταφορά υπερισχύει της διάχυσης. Αφού προσδιορίσουμε πρώτα την ασθενή διατύπωση του προβλήματος στη συνέχεια περνάμε στη διακριτοποίηση αυτού χρησιμοποιώντας την μέθοδο $Galerkin$ με πεπερασμένα στοιχεία καί στην αντίστοιχη εκτίμηση του σφάλματος. Αν τώρα το πρόβλημα μας ορίζεται σε περισσότερες από μία διαστάσεις ,τότε χρησιμοποιώντας την τεχνιτή διάχυση σταθεροποιούμε την μέθοδο των πεπερασμένων στοιχείων καί υπολογίζουμε την αντίστοιχη βελτιωμένη εκτίμηση του σφάλματος.

Τέλος στο **Κεφάλαιο 6.** με χρήση του προγράμματος $Freefem$ επιλύουμε αριθμητικά το πρόβλημα $Laplace$ καί το πρόβλημα μεταφοράς-διάχυσης στις δύο διαστάσεις. Εκτός από την προσεγγιστική τους επίλυση γίνεται καί υπολογισμός των αντίστοιχων σφαλμάτων για διάφορα βήματα της μεθόδου των πεπερασμένων στοιχείων. Με αυτό τον τρόπο αποκτούμε μία πλήρη εικόνα για το τι συμβαίνει στα προβλήματα μεταφοράς-διάχυσης καθώς βλέπουμε την απόλυτη συσχέτιση που υπάρχει μεταξύ των θεωρητικών αποτελεσμάτων της εκτίμησης σφάλματος με τα αντίστοιχα αριθμητικά αποτελέσματα.

Κεφάλαιο 2

Παρουσίαση του προβλήματος.

Υποθέτουμε ότι έχουμε το ακόλουθο πρόβλημα συνοριακών τιμών,

$$u''(x) = f(x), 0 < x < 1 \tag{2.1}$$

$$u(0) = u(1) = 0 \tag{2.2}$$

όπου η συνάρτηση $u \in \mathcal{C}^2([0,1])$
Αναζητούμε λύση της διαφορικής εξίσωσης(2.1)-(2.2) της μορφής:

$$u(x) = c_1 + c_2 x - \int_0^x F(s)ds$$

όπου c_1, c_2 είναι αυθαίρετες σταθερές καί $F(s) = \int_0^s f(t)dt$. Χρησιμοποιώντας ολοκλήρωση κατά μέρη προκύπτει,

$$\int_0^x F(s)ds = \int_0^x (s)'F(s)ds = [sF(s)]_0^x - \int_0^x sF'(s)ds = \int_0^x (x-s)f(s)ds.$$

Άρα η λύση της διαφορικής εξίσωσης (2.1) μπορεί να γραφεί τώρα στην ακόλουθη μορφή,

$$u(x) = c_1 + c_2 x - \int_0^x (x-s)f(s)ds.$$

Οι σταθερές c_1, c_2 προσδιορίζονται εφαρμόζοντας τις συνοριακές συνθήκες (2.2). Από την πρώτη συνθήκη $u(0) = 0$ προκύπτει ότι $c_1 = 0$, ενώ από τη δεύτερη συνθήκη $u(1) = 0$ προκύπτει ότι $c_2 = \int_0^1 (1-s)f(s)ds$. Επομένως η λύση τώρα γράφεται,

$$u(x) = x\int_0^1 (1-s)f(s)ds - \int_0^x (x-s)f(s)ds.$$

Για κάθε x σταθερό ορίζουμε,

$$G(x,s) = \begin{cases} s(1-x), & \text{αν} \quad 0 \le s \le x \\ x(1-s), & \text{αν} \quad x \le s \le 1. \end{cases} \tag{2.3}$$

Η λύση τώρα της διαφορικής χρησιμοποιώντας την συνάρτηση G μπορεί να γραφεί σε πιο συμπτηγμένη μορφή,

$$u(x) = \int_0^1 G(x,s) f(s) ds. \tag{2.4}$$

Η συνάρτηση G είναι η συνάρτηση Green για το πρόβλημα συνοριακών τιμών (2.1)-(2.2). Είναι μια κατά τμήματα γραμμική συνάρτηση του x για σταθερό s καί αντιστρέψιμη. Επίσης $\forall x, s \in [0,1]$ είναι συνεχής, συμμετρική (δηλαδή, $G(x,s) = G(s,x)$), μη αρνητική καί μηδενίζεται αν το x ή το s είναι ίσα με το 0 ή το 1, τέλος το ολοκλήρωμα αυτής της συνάρτησης από το 0 στο 1 είναι ίσο με το εμβαδόν των τριγώνων που σχηματίζει η συνάρτηση Green για διαφορετικά x δηλαδή $\int_0^1 G(x,s)ds = \frac{1}{2}x(1-x)$.

ΠΑΡΑΤΗΡΗΣΕΙΣ ΠΑΝΩ ΣΤΗ ΣΧΕΣΗ (2.4):

Παρατήρηση 2.1 *Λόγω της σχέσης (2.4) μπορούμε να συμπεράνουμε ότι* $\forall f \in \mathcal{C}^0([0,1])$ *υπάρχει* **μοναδική λύση** $u \in \mathcal{C}^2([0,1])$ *του προβλήματος συνοριακών τιμών (2.1)-(2.2).*

Παρατήρηση 2.2 *Μία άλλη ιδιότητα της λύσης u που προκύπτει από τη σχέση (2.4) είναι ότι αν η συνάρτηση* $f \in \mathcal{C}^0([0,1])$ *είναι μη αρνητική, τότε καί η λύση u είναι επίσης μη αρνητική, διότι* $\forall x, s \in [0,1], G(x,s) \ge 0$. *Η ιδιότητα αυτή αναφέρεται ως* **ιδιότητα της μονοτονίας**.

Παρατήρηση 2.3 *Επειδή η συνάρτηση G είναι μη αρνητική έχουμε ότι,*

$$|u(x)| \le \int_0^1 G(x,s)|f(s)|ds \le \|f\|_\infty \int_0^1 G(x,s)ds = \frac{1}{2}x(1-x)\|f\|_\infty$$

στη συνέχεια χρησιμοποιώντας καί την άπειρο νόρμα που ορίζεται ως $\|u(x)\| = \max_{0\le x\le 1}|u(x)|$ *καταλήγουμε στην ακόλουθη ιδιότητα της λύσης u,*

$$\|u(x)\|_\infty \le \frac{1}{8}\|f\|_\infty. \tag{2.5}$$

Η ιδιότητα αυτή ονομάζεται **αρχή του μεγίστου** *καί ισχύει αν* $f \in \mathcal{C}^0([0,1])$.

Κεφάλαιο 3

Μέθοδος των πεπερασμένων διαφορών.

Η μέθοδος των πεπερασμένων διαφορών για την επίλυση συνοριακών προβλημάτων κατασκευάζεται, αν αντικαταστήσουμε τις παραγώγους που υπάρχουν στη διαφορική εξίσωση με κατάλληλες προσεγγίσεις.
Συγκεκριμένα στο πρόβλημα συνοριακών τιμών (2.1)-(2.2) η μέθοδος εφαρμόζεται ως εξής:
Διαιρούμε το διάστημα [0,1] σε n+1 υποδιαστήματα (όπου $n \in \mathbb{Z}$ καί $n \geq 2$) μήκους $h = \frac{1}{n}$ το καθένα. Ζητάμε προσεγγίσεις u_j της λύσης $u(x_j)$ στα σημεία x_j με $j = 1, ..., n-1$.Το σύνολο των προσεγγίσεων u_j της λύσης $u(x_j)$ είναι μια πεπερασμένη ακολουθία $\{u_j\}_{j=0}^{n}$, η οποία ορίζεται μόνο στα διακριτά σημεία $\{x_j\}_{j=0}^{n}$.
Το πρόβλημα (2.1)-(2.2) προσεγγίζεται τώρα στο σημείο x_j αντικαθιστώντας την παράγωγο $u''(x_j)$ από κατάλληλη προσέγγιση η οποία προκύπτει χρησιμοποιώντας το πολυώνυμο παρεμβολής Newton. Αναλυτικότερα,

το πολυώνυμο $P_n(x)$ βαθμού n παρεμβάλλει την συνάρτηση $u(x)$ με $x \in [0, 1]$ στα διακριτά σημεία $\{x_j\}_{j=0}^{n}$ καί την προσεγγίζει στο διάστημα [0,1]. Σε μορφή Newton καί χρησιμοποιώντας τις εμπρός διαφορές, για ισαπέχοντα σημεία απόστασης $h = \frac{1}{n}$ το πολυώνυμο γράφεται,

$$(3.1)\quad P_n(x) = P_n(x_0 + sh) = u_0 + \binom{s}{1}\Delta u_0 + \binom{s}{2}\Delta^2 u_0 + ... + \binom{s}{n}\Delta^n u_0,$$

όπου $s = \frac{x-x_0}{h}$ καί Δ είναι ο τελεστής των προς τα εμπρός διαφορών, αναλυτικότερα:

- $\Delta u_0 = u_1 - u_0$
- $\Delta^2 u_0 = \Delta(\Delta u_0) = \Delta(u_1 - u_0) = u_2 - 2u_1 - u_0$

⋮

-
- $\Delta^n u_0 = \sum_{r=0}^{n} (-1)^{n-r} \binom{n}{r} u_{0+r}$

(βλ.βιβλία [1,7,9])

Παραγωγίζοντας δύο φορές την σχέση (3.1), θέτοντας $n = 2$ (με σκοπό να προκύψει μία μη μηδενική τιμή της δεύτερης παράγωγου του πολυωνύμου $P_n(x)$) καί θεωρώντας ότι $u''(x) \approx P_n''(x)$, προκύπτει η προσέγγιση της δεύτερης παραγώγου στο κεντρικό σημείο παρεμβολής x_j.

Άρα το πρόβλημα (2.1)-(2.2) παίρνει την ακόλουθη μορφή,

$$\frac{-u_{j+1} + 2u_j - u_{j-1}}{h^2} = f(x_j) \qquad j = 1, ..., n-1, \tag{3.2}$$

με $u_0 = u(x_0) = u(0) = u(1) = u_(x_n) = u_n = 0$.

Αν θέσουμε $\mathbf{u} = (u_1, ..., u_{n-1})^T$ καί $\mathbf{f} = (f_1, ..., f_{n-1})$,με $f_i = f(x_i)$, τότε η σχέση (3.2) μπορεί να γραφεί σε ποιο συμπτηγμένη μορφή,

$$A_{fd}\mathbf{u} = \mathbf{f}, \tag{3.3}$$

όπου A_{fd} είναι ο πίνακας των πεπερασμένων διαφορών διάστασης $(n-1)\times(n-1)$ που ορίζεται ως εξής,

$$A_{fd} = h^{-2}\text{tridiag}_{n-1}(-1, 2, -1), \tag{3.4}$$

δηλαδή ο πίνακας A_{fd} έχει την ακόλουθη μορφή,

$$A_{fd} = h^{-2}\begin{bmatrix} 2 & -1 & 0 & \cdots & 0 \\ -1 & 2 & -1 & & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \cdots & -1 & 2 & -1 \\ 0 & \cdots & 0 & -1 & 2 \end{bmatrix}$$

ΠΑΡΑΤΗΡΗΣΕΙΣ ΠΑΝΩ ΣΤΗ ΣΧΕΣΗ (3.3):

Παρατήρηση 3.1 *Από την δόμη του πίνακα A_{fd} μπορούμε να συμπεράνουμε ότι, ο πίνακας είναι* συμμετρικός, *δηλαδή $(A_{fd})^T = A_{fd}$, έχει* αυστηρή διαγώνια υπεροχή κατά γραμμές *καί ειναι* θετικά ορισμένος, *διότι για κάθε διάνυσμα $\mathbf{x}\in \mathbb{R}^{n-1}$ με $\mathbf{x} \neq 0$ το παρακάτω γινόμενο είναι μεγαλύτερο του μηδενός,*

$$\mathbf{x}^T A_{fd}\mathbf{x} = h^{-2}\left[x_1^2 + x_{n-1}^2 + \sum_{i=2}^{n-1}(x_i - x_{i-1})^2\right] > 0.$$

Λόγω των παραπάνω ιδιοτήτων του πίνακα A_{fd} καταλήγουμε στο συμπέρασμα ότι πρόβλημα που περιγράφεται μέσω της σχέσης (3.3) έχει μοναδική λύση.

Παρατήρηση 3.2 *Το σύστημα (3.3) μπορεί εύκολα να επιλυθεί με μια από τις μεθόδους επίλυσης γραμμικών συστημάτων, διότι ο πίνακας A_{fd} είναι τριδιαγώνιος.*

Η σχέση (3.2) μπορεί να γραφεί καί σε μορφή τελεστών. Για το σκοπό αυτό ορίζουμε με V_h το σύνολο όλων των διακριτών συναρτήσεων που ορίζονται στα διακριτά σημεία x_j για $j = 0, .., n$, δηλαδή αν μία συνάρτηση $w_h \in V_h$, τότε η συνάρτηση $w_h(x_j)$ ορίζεται για όλα τα x_j. Για λόγους συντομογραφίας αντί για $w_h(x_j)$ θα χρησιμοποιούμε τον συμβολισμό w_j. Στη συνέχεια ορίζουμε V_h^0 να είναι υποσύνολο του V_h που περιέχει τις διακριτές συναρτήσεις οι οποίες μηδενίζονται στα σημεία x_0 καί x_n. Παρακάτω ορίζεται ο τελεστής L_h για μία συνάρτηση w_h ως εξής:

$$(L_h w_h)(x_j) = -\frac{w_{j+1} - 2w_j + w_{j-1}}{h^2}, \qquad j = 1, ..., n-1, \tag{3.5}$$

καί το πρόβλημα πεπερασμένων διαφορών (3.2) είναι τώρα ισοδύναμο με το ακόλουθο: να βρεθεί $u_h \in V_h^0$ έτσι ώστε,

$$(L_h u_h)(x_j) = f(x_j), \qquad j = 1, ..., n-1. \tag{3.6}$$

Παρατηρούμε ότι οι συνοριακές συνθήκες του προβλήματος (3.6) περιλαμβάνονται στην απαίτηση μας $u_h \in V_h^0$.

3.1 Ανάλυση της ευστάθειας της μεθόδου.

Για δύο διακριτές συναρτήσεις $w_h, u_h \in V_h$ ορίζεται το διακριτό **εσωτερικό γινόμενο**

$$(w_h, u_h) = h \sum_{k=0}^{n} c_k w_k u_k, \tag{3.7}$$

με $c_0 = c_n = \frac{1}{2}$ καί $c_k = 1$ για $k = 1, .., n-1$.
Διαφορετικά η σχέση (3.7) δεν είναι τίποτα άλλο από τον σύνθετο κανόνα τραπεζίου, ο οποίος χρησιμοποιείται για τον υπολογισμό του εσωτερικού γινομένου $(w, u) = \int_0^1 w(x)u(x)dx$.
Αναλυτικότερα,
το γινόμενο $w(x)u(x)$ προσεγγίζεται από το πολυώνυμο παρεμβολής Newton με αποτέλεσμα,

$$\int_0^1 w(x)u(x)dx \approx \int_{x_0}^{x_n} p_n(x)dx = \int_{x_0}^{x_1} p_n(x)dx + \int_{x_1}^{x_2} p_n(x)dx + ... + \int_{x_{n-1}}^{x_n} p_n(x)dx. \tag{3.8}$$

Υπολογισμός των επιμέρους ολοκληρωμάτων της σχέσης (3.8):

$$\int_{x_0}^{x_1} p_n(x)dx = \int_{x_0}^{x_1} (w_0 u_0 + \binom{s}{1} \Delta w_0 u_0)dx, \qquad (n = 1),$$

όπου $s = \frac{x-x_0}{h}$ καί $h = \frac{1}{n}$.
Αλλάζοντας την μεταβλητή ολοκλήρωσης (από x σε s) το παραπάνω ολοκλήρωμα παίρνει την ακόλουθη μορφή,

$$h\int_0^1 (w_0u_0 + \binom{s}{1}\Delta w_0u_0)ds = h(w_0u_0s + \frac{s^2}{2}\Delta w_0u_0)|_0^1 = \frac{h}{2}(w_0u_0 + w_1u_1).$$

Με τον ίδιο τρόπο προκύπτει καί ο υπολογισμός των υπολοίπων ολοκληρωμάτων καί η σχέση (3.8) γράφεται τώρα ως εξής,

$$\int_0^1 w(x)u(x)dx \approx \frac{h}{2}(w_0u_0+w_1u_1)+\frac{h}{2}(w_1u_1+w_2u_2)+...+\frac{h}{2}(w_{n-1}u_{n-1}+w_nu_n) = h\sum_{k=0}^{n} c_kw_ku_k,$$

όπου $c_0 = c_n = \frac{1}{2}$ καί $c_k = 1$, για $k = 1, ..., n-1$.

Από την σχέση (3.7) καί με την υπόθεση ότι $w_h = u_h$ προκύπτει ότι,

$$(u_h, u_h)_h^{1/2} = \|u_h\|_h. \tag{3.9}$$

Θα δείξουμε τώρα ότι το παραπάνω διγραμμικό συναρτησοειδές $\|.\|_h : V_h \times V_h \to \mathbb{R}$ περιγράφει μία νόρμα του χώρου V_h. Αναλυτικότερα,

i)$\forall u_h \in V_h$ είναι προφανές ότι $\|u_h\|_h \geq 0$ καθώς,

$$\|u_h\|_h = (u_h, u_h)_h^{1/2} = \sqrt{h\sum_{k=0}^{n-1} c_ku_k^2} = \sqrt{h\left(\frac{u_0^2}{2} + u_1^2+, ..., +\frac{u_n^2}{2}\right)} \geq 0,$$

και

$$\|u_h\|_h = 0 \Leftrightarrow u_0 = u_1 = ... = u_n \Leftrightarrow u_h = 0.$$

Επίσης,
ii)$\forall u_h \in V_h$ καί $\lambda \in \mathbb{R}$ έχουμε $\|\lambda u_h\|_h = |\lambda|\|u_h\|_h$,διότι

$$\|\lambda u_h\|_h = (\lambda u_h, \lambda u_h)_h^{1/2} = \sqrt{h\sum_{k=0}^{n-1} c_k\lambda^2u_k^2} = \sqrt{\lambda^2}\sqrt{h\sum_{k=0}^{n-1} c_ku_k^2} = |\lambda|\|u_h\|_h.$$

Τέλος θα δείξουμε ότι,
iii)$\forall w_h, u_h \in V_h$ ισχύει η τριγωνική ανισότητα $\|u_h + w_h\|_h \leq \|u_h\|_h + \|w_h\|_h$,

$$\|u_h + w_h\|_h = (u_h + w_h, u_h + w_h)_h^{1/2} = \left(h\sum_{k=0}^{n-1} c_k(u_k + w_k)^2\right)^{1/2}$$

$$= \left(\left(h\sum_{k=0}^{n-1} c_ku_k^2\right) + \left(h\sum_{k=0}^{n-1} c_kw_k^2\right) + 2\left(h\sum_{k=0}^{n-1} c_ku_kw_k\right)\right)^{1/2}$$

$$= ((u_h, u_h)_h + (w_h, w_h)_h + 2(u_h, w_h)_h)^{1/2}$$
$$\leq ((u_h, u_h)_h + (w_h, w_h)_h + 2|(u_h, w_h)_h|)^{1/2}.$$

Εφαρμόζωντας την ακόλουθη **ανισότητα Cauchy-Schwarz** η οποία ισχύει στην περίπτωση του εσωτερικού γινομένου,

$$|(u_h, w_h)_h| \leq (u_h, u_h)_h^{1/2} (w_h, w_h)_h^{1/2}$$

προκύπτει ότι,

$$\|u_h + w_h\|_h \leq \left((u_h, u_h)_h + (w_h, w_h)_h + 2(u_h, u_h)_h^{1/2}(w_h, w_h)_h^{1/2}\right)^{1/2} =$$

$$= \left(\left((u_h, u_h)_h^{1/2} + (w_h, w_h)_h^{1/2}\right)^2\right)^{1/2} =$$

$$= \|u_h\|_h + \|w_h\|_h.$$

Λήμμα 3.1.1 *i)Ο τελεστής L_h είναι συμμετρικός, δηλαδή*

$$(L_h w_h, u_h)_h = (w_h, L_h u_h)_h \qquad \forall w_h, u_h \in V_h^0$$

ii)και θετικά ορισμένος, δηλαδή

$$(L_h u_h, u_h)_h \geq 0 \qquad \forall u_h \in V_h^0,$$

η ισότητα ισχύει μόνο στην περίπτωση που $u_h \equiv 0$.

Απόδειξη.

i)Από την σχέση (3.5) έχουμε ότι $(L_h w_h)(x_j) = -\frac{w_{j+1} - 2w_j + w_{j+1}}{h^2}$,
άρα το διακριτό εσωτερικό γινόμενο των διακριτών συναρτήσεων $L_h w_h$ καί u_h είναι,

(3.10)

$$(L_h w_h, u_h)_h = -h \sum_{j=0}^{n-1} \left(\frac{w_{j+1} - 2w_j + w_{j+1}}{h^2}\right) u_j = h^{-1} \sum_{j=0}^{n-1} ((w_{j+1} - w_j) - (w_j - w_{j-1}))\, u_j.$$

Χρησιμοποιώντας την ακόλουθη ταυτότητα,

$$w_{j+1}u_{j+1} - w_j u_j = (w_{j+1} - w_j)u_j + (u_{j+1} - u_j)w_{j+1}$$

καί αθροίζοντας από το 0 μέχρι το $n-1$ προκύπτει ότι, $\forall w_h, u_h \in V_h$

$$\sum_{j=0}^{n-1} (w_{j+1} - w_j)\, u_j = w_n u_n - w_0 u_0 - \sum_{j=0}^{n-1} (u_{j+1} - u_j)\, w_{j+1},$$

το οποίο αναφέρεται ως *άθροιση κατά μέρη*. Χρησιμοποιώντας δύο φορές την άθροιση κατά μέρη καί θέτοντας $\forall w_h, u_h \in V_h^0 w_{-1} = u_{-1} = 0$, παρατηρούμε ότι η σχέση (3.10) είναι ισοδύναμη με την ακόλουθη:

$$(L_h w_h, u_h)_h = h^{-1} \sum_{j=0}^{n-1} (w_{j+1} - w_j)(u_{j+1} - u_j). \tag{3.11}$$

Με παρόμοιο τρόπο υπολογίζουμε $(u_h, L_h w_h)$ καί παρατηρούμε ότι μπορούμε να εναλλάξουμε τον όρο w_j με u_j από την σχέση (3.11). Έτσι έχουμε $(L_h w_h, u_h)_h = (w_h, L_h u_h)_h$.

ii)Για να αποδείξουμε τώρα ότι ο τελεστής L_h είναι θετικά ορισμένος, θεωρούμε ότι $w_h = u_h$ καί η σχέση (3.11) παίρνει την ακόλουθη μορφή,

$$(L_h u_h, u_h)_h = h^{-1} \sum_{j=0}^{n-1} (u_{j+1} - u_j)^2 \geq 0. \tag{3.12}$$

Η ισότητα ισχύει όταν $u_{j+1} = u_j$ για $j = 0, ..., n-1$, συγκεκριμένα για $j = 0$ έχουμε ότι $u_1 = u_0$, όμως $u_0 = 0$(συνοριακή συνθήκη) άρα $u_1 = 0$. Ομοίως προκύπτει ότι $u_2 = u_3 = ... = u_n = 0$. Επομένως καταλήγουμε στο συμπέρασμα ότι η ισότητα ισχύει μόνο στην περίπτωση που $u_j = 0$ για $j = 0, ..., n$. □

Για οποιαδήποτε διακριτή συνάρτηση $u_h \in V_h$ ορίζουμε την ακόλουθη νόρμα,

$$|||u_h|||_h = \left\{ h \sum_{j=0}^{n-1} \left(\frac{u_{j+1} - u_j}{h} \right)^2 \right\}^{1/2}. \tag{3.13}$$

Επομένως η σχέση (3.12) είναι τώρα ισοδύναμη με την ακόλουθη,

$$(L_h u_h, u_h)_h = |||u_h|||_h^2 \qquad \forall u_h \in V_h^0. \tag{3.14}$$

Μέχρι στιγμής έχουμε ορίσει δύο νόρμες στον χώρο V_h^0 την $\|u_h\|_h$ καί την $|||u_h|||_h$. Παρακάτω μέσω του λήμματος (3.1.2) δίνεται η σχέση που υπάρχει μεταξύ τους.

Λήμμα 3.1.2 *Για οποιαδήποτε συνάρτηση $u_h \in V_h^0$ ισχύει ότι:*

$$\|u_h\|_h \leq \frac{1}{\sqrt{2}} |||u_h|||_h. \tag{3.15}$$

Απόδειξη.

Υποθέσαμε ότι η $u_h \in V_h^0$, άρα η u_j (δηλαδή η $u(x_j)$) ορίζεται για όλα τα j. Από τη στιγμή που $u_0 = 0$ έχουμε,

$$u_j = h\sum_{k=0}^{j-1} \frac{u_{k+1} - u_k}{h} \qquad \forall j = 1, ..., n-1,$$

διότι η παραπάνω σχέση για $j = 1$ δίνει $u_1 = u_1$ ομοίως για $j = 2$ δίνει $u_2 = u_2$ κ.ο.κ.
Επίσης,

$$u_j^2 = h^2 \left[\sum_{k=0}^{j-1} \left(\frac{u_{k+1} - u_k}{h}\right)\right]^2.$$

Χρησιμοποιώντας την ανισότητα Minkowski,

$$\left(\sum_{k=1}^{m} p_k\right)^2 \leq m \left(\sum_{k=1}^{m} p_k^2\right),$$

η οποία ισχύει $\forall m \in \mathbb{Z}$ με $m \geq 1$ καί για κάθε ακολουθία $\{p_1, ..., p_m\}$ πραγματικών αριθμών,παρατηρούμε ότι,

$$\sum_{j=1}^{n-1} u_j^2 \leq h^2 \sum_{j=1}^{n-1} j \sum_{k=0}^{j-1} \left(\frac{u_{k+1} - u_k}{h}\right)^2.$$

Τότε $\forall u_h \in V_h^0$ έχουμε,

$$\|u_h\|_h^2 = (u_h, u_h)_h = h\sum_{j=1}^{n-1} u_j^2 \leq h^2 \sum_{j=1}^{n-1} jh \sum_{k=0}^{n-1} \left(\frac{u_{k+1} - u_k}{h}\right)^2 = h^2 \frac{(n-1)n}{2} |||u_h|||_h^2.$$

Αντικαθιστώντας στην παραπάνω σχέση όπου $h = \frac{1}{n}$ προκύπτει τελικά η ζητούμενη σχέση,

$$\|u_h\|_h^2 \leq \frac{1}{n}\frac{(n-1)}{2} |||u_h|||_h^2 = (\frac{1}{2} - \frac{1}{2n}) |||u_h|||_h^2 \overset{(n\to\infty)}{=} \frac{1}{2} |||u_h|||_h^2.$$

□

ΠΑΡΑΤΗΡΗΣΕΙΣ ΠΑΝΩ ΣΤΗ ΣΧΕΣΗ (3.15):

Παρατήρηση 3.3 *Για κάθε $u_h \in V_h^0$ η διακριτή συνάρτηση $u_h^{(1)}$, με τιμές $\frac{u_{j+1}-u_j}{h}$ για $j = 0, ..., n-1$ μπορεί να θεωρηθεί ως η διακριτή παράγωγος του u_h. Η σχέση (3.15) παίρνει τότε την ακόλουθη μορφή,*

$$\|u_h\|_h \leq \frac{1}{2} \|u_h^{(1)}\|_h \qquad \forall u_h \in V_h^0,$$

διότι,

$$\|u_h^{(1)}\|_h = (u_h^{(1)}, u_h^{(1)})_h^{1/2} = \left(h\sum_{j=0}^{n-1} u_h^{(1)} u_h^{(1)}\right)^{1/2} = \left(h\sum_{j=0}^{n-1} \left(\frac{u_{j+1} - u_j}{h}\right)^2\right)^{1/2} = |||u_h|||_h.$$

Επίσης η συνάρτηση $u_h^{(1)}$ μπορεί να θεωρηθεί ως ο διακριτός μετρητής στο διάστημα [0,1] της ακόλουθης ανισότητας Poincare:∀ διάστημα [a,b] ∃ μία σταθερά $C_p > 0$ τέτοια ώστε,

$$\|u\|_{L^2(a,b)} \leq C_p \|u^1\|_{L^2(a,b)},$$

για όλες τις συναρτήσεις $u \in \mathcal{C}^1([a,b])$ τέτοιες ώστε $u(a) = u(b) = 0$ καί όπου $\|.\|_{L^2(a,b)}$ είναι η νόρμα του χώρου $L^2(a,b)$,δηλαδή $\|u\|_{L^2(a,b)} = \left(\int_a^b |u(x)|^2\right)^{1/2}$.

Παρατήρηση 3.4 *Η ανισότητα (3.15) έχει μία ενδιαφέρουσα συνέπεια. Αν πολλαπλασιάσουμε την σχέση (3.6) με u_j καί στη συνέχεια αθροίσουμε ως προς j για $j = 1, ..., n-1$ θα παρατηρήσουμε ότι,*

$$(L_h u_h, u_h)_h = (f, u_h)_h,$$

από την σχέση αυτή καί σε συνδιασμό με την (3.14) προκύπτει ότι,

$$|||u_h|||_h^2 = (f, u_h)_h,$$

με εφαρμογή τώρα της ανισότητας Cauchy-Schwarz έχουμε,

$$(f, u_h)_h = \sum_{j=0}^{n-1} f(x_j)u(x_j) \leq \sum_{j=0}^{n-1} |f(x_j)u(x_j)| \leq \left(\sum_{j=0}^{n-1} |f(x_j)|^2\right)^{1/2} \left(\sum_{j=0}^{n-1} |u(x_j)|^2\right)^{1/2} = \|f_h\|_h \|u_h\|_h,$$

δηλαδή,

$$|||u_h|||_h^2 \leq \|f_h\|_h \|u_h\|_h.$$

Στις παραπάνω σχέσεις υπενθυμίζουμε ότι το f_h ορίζεται από την f ως η διακριτοποίηση $f_h(x_j) = f(x_j)$, δηλαδή έχουμε $f_h(x) = f_h(x_j) \quad \forall x \in [x_j, x_{j+1}]$. Τέλος χρησιμοποιώντας την σχέση (3.15) καταλήγουμε στην ακόλουθη ανισοτική σχέση,

$$\|u_h\|_h \leq \frac{1}{2}\|f_h\|_h, \tag{3.16}$$

από την οποία συμπεραίνουμε ότι το πρόβλημα πεπερασμένων διαφορών (3.2) έχει μοναδική λύση. Επίσης παρατηρούμε ότι η λύση του προβλήματος χρησιμοποιώντας την μέθοδο των πεπερασμένων διαφορών είναι φραγμένη από την διακριτή συνάρτηση f_h, πράγμα που μας εξασφαλίζει την ευστάθεια της λύσης.

3.2 Ανάλυση της συνέπειας της μεθόδου.

Η μετάβαση από το συνεχές πρόβλημα συνοριακών τιμών (2.1)-(2.2) στο διακριτό (3.2)ή(3.5) μέσω της μεθόδου των πεπερασμένων διαφορών εισάγει ένα σφάλμα προσέγγισης το οποίο γενικά ονομάζεται **σφάλμα διακριτοποίησης ή αποκοπής**. Τα σφάλματα αυτά διακρίνονται σε **τοπικά** καί **ολικά**. Το τοπικό σφάλμα έχει ουσιαστικά να κάνει με το σφάλμα αποκοπής

σε ένα μόνο βήμα εφαρμογής του προσεγγιστικού μοντελου.

Συγκεκριμένα στο πρόβλημα (2.1)-(2.2), όπου η συνάρτηση $f \in \mathcal{C}^0([0,1])$ καί $u \in \mathcal{C}^2([0,1])$ είναι η λύση του προβλήματος, το τοπικό σφάλμα αποκοπής είναι μία διακριτή συνάρτηση η οποία ορίζεται ως εξής,

$$T_h(x_j) = (L_h u)(x_j) - f(x_j), \qquad j = 1, ..., n-1. \tag{3.17}$$

Ο ρόλος τώρα της συνέπειας είναι να περιορίσει το μέγεθος των τοπικών σφαλμάτων.

Η σχέση (3.17) είναι ισοδύναμη με την ακόλουθη,

$$T_h(x_j) = -h^{-2}[u(x_{j-1}) - 2u(x_j) + u(x_{j+1})] - f(x_j). \tag{3.18}$$

Χρησιμοποιώντας το ανάπτυγμα Taylor το $u(x_{j+1})$ για κατάλληλο $\eta_j \in (x_j, x_{j+1})$ αναπτύσσεται ως εξής,

$$u(x_{j+1}) = u(x_j) + hu'(x_j) + \frac{h^2}{2}u''(x_j) + \frac{h^3}{6}u'''(x_j) + \frac{h^4}{24}u^{(iv)}(\eta_j). \tag{3.19}$$

Ομοίως για κατάλληλο $\xi_j \in (x_{j-1}, x_j)$ έχουμε καί το ανάπτυγμα του $u(x_{j-1})$,

$$u(x_{j-1}) = u(x_j) - hu'(x_j) + \frac{h^2}{2}u''(x_j) - \frac{h^3}{6}u'''(x_j) + \frac{h^4}{24}u^{(iv)}(\xi_j). \tag{3.20}$$

Στην συνέχεια αντικαθιστούμε τις σχέσεις (3.19) καί (3.20) στην σχέση (3.18) με αποτέλεσμα το τοπικό σφάλμα αποκοπής να πάρει την ακόλουθη μορφή,

$$T_h(x_j) = \frac{h^2}{24}(u^{(iv)}(\xi_j) + u^{(iv)}(\eta_j)). \tag{3.21}$$

Παρατηρούμε ότι,

$$\|T_h\|_{h,\infty} \leq \frac{\|f''\|_\infty}{12}h^2. \tag{3.22}$$

Αναλυτικότερα,

$$\|T_h\|_{h,\infty} = \frac{h^2}{24}\|u^{(iv)}(\xi_j) + u^{(iv)}(\eta_j)\|_{h,\infty} \leq \frac{h^2}{24}(\|u^{(iv)}(\xi_j)\|_{h,\infty} + \|u^{(iv)}(\eta_j)\|_{h,\infty}) \leq$$

$$\leq \frac{h^2}{24}(\|f''\|_\infty + \|f''\|_\infty) = \frac{h^2}{12}\|f''\|_\infty,$$

όπου χρησιμοποιήσαμε την *διακριτή άπειρο νόρμα* η οποία ορίζεται ως εξής,

$$\|u_h\|_{h,\infty} = \max_{0 \leq j \leq n} |u_h(x_j)|.$$

Η σχέση (3.22) είναι πολύ σημαντική καθώς,

$$\lim_{h\to 0} \|T_h\|_{h,\infty} = 0.$$

Το αποτέλεσμα αυτό μας εξασφαλίζει την συνέπεια της μεθόδου.

Έστω $e = u - u_h$ η συνάρτηση του ολικού σφάλματος διακριτοποιήσης. Το ολικό σφάλμα εκτός από το τοπικό σφάλμα στο τελευταίο βήμα της μεθόδου περιλαμβάνει καί το συσσωρευμένο σφάλμα που οφείλεται στη μετάδοση όλων των προηγούμενων τοπικών σφαλμάτων αποκοπής. Παρατηρούμε ότι,

$$L_h e = L_h u - L_h u_h = L_h u - f_h = T_h. \tag{3.23}$$

Ασκηση 1 *Χρησιμοποιώντας την παραπάνω σχέση θα δείξουμε ότι,*

$$\|T_h\|_h^2 \le 3\|f\|_h^2 + 2\|f\|_{L^2(0,1)}^2 \tag{3.24}$$

Απόδειξη:

$$\|T_h\|_h^2 = (T_h, T_h)_h = (L_h u - f_h, L_h u - f_h)_h = (L_h u - f_h, L_h u)_h - (L_h u - f_h, f_h)_h =$$

$$= (L_h u - f_h, L_h u)_h - (L_h u, f_h)_h + \|f_h\|_h^2 \le |(T_h, L_h u)_h| + |(L_h u, f_h)_h| + \|f_h\|_h^2 \le$$

$$\le \|T_h\|_h \|L_h u\|_h + \|L_h u\|_h \|f_h\|_h + \|f_h\|_h^2 \le \|T_h\|_h \|f\|_{L^2(0,1)} + \|f\|_{L^2(0,1)} \|f\|_h + \|f_h\|_h^2$$

εφαρμόζωντας την ταυτότητα,

$$a \cdot b \le \frac{a^2}{2} + \frac{b^2}{2}$$

η παραπάνω σχέση είναι ισοδύναμη με την ακόλουθη,

$$\|T_h\|_h^2 \le \frac{1}{2}\|T_h\|_h^2 + \frac{1}{2}\|f\|_{L^2(0,1)}^2 + \frac{1}{2}\|f\|_{L^2(0,1)}^2 + \frac{1}{2}\|f_h\|_h^2 + \|f_h\|_h^2 \Rightarrow$$

$$\Rightarrow \frac{1}{2}\|T_h\|_h^2 \le \|f\|_{L^2(0,1)}^2 + \frac{3}{2}\|f_h\|_h^2 \Rightarrow \|T_h\|_h^2 \le 2\|f\|_{L^2(0,1)}^2 + 3\|f_h\|_h^2$$

.

Παρατήρηση 3.5 *Παρατηρούμε ότι αν το δεξί μέλος της ανισότητας (3.24) είναι φραγμένο δηλαδή, οι νόρμες $\|f\|_{L^2(0,1)}^2$ καί $\|f_h\|_h^2$ είναι φραγμένες, τότε καί το $\|T_h\|_{h,\infty}$ θα είναι φραγμένο. Το συμπέρασμα αυτό προκύπτει από το γεγονός ότι στο δεξί μέλος της ανισοτικής σχέσης δεν εμφανίζεται ο παράγοντας h.*

3.3 Ανάλυση της σύγκλισης της μεθόδου.

Στις παραγράφους (3.1) καί (3.2) έγινε ανάλυση της **ευστάθειας** καί της **συνέπειας** αντίστοιχα της μεθόδου των πεπερασμένων διαφορών. Οι δύο αυτές έννοιες είναι απαραίτητες για να αποδειχθεί τώρα η **σύγκλιση** της μεθόδου, δηλαδή αν η προσεγγιστική λύση η οποία προκύπτει με εφαρμογή της μεθόδου συγκλίνει στην ακριβή λύση του προβλήματος.

Η λύση u_h που προκύπτει με εφαρμογή της μεθόδου των πεπερασμένων διαφορών μπορεί να χαρακτηριστεί από μία διακριτή συνάρτηση Green.
Αναλυτικότερα, για δοθέν διακριτό σημείο x_k ορίζεται η διακριτή συνάρτηση $G^k \in V_h^0$ ως η λύση του παρακάτω προβλήματος:

$$L_h G^k = e^k,$$

όπου η συνάρτηση $e^k \in V_h^0$ ικανοποιεί την σχέση $e^k(x_j) = \delta_{kj}$,$1 \leq j \leq n-1$,
καί

$$\delta_{kj} = \begin{cases} 1, & \text{αν} \quad k = j \\ 0, & \text{αν} \quad k \neq j. \end{cases} \tag{3.25}$$

είναι το σύμβολο του Kronecker. Επίσης παρατηρούμε ότι $G^k(x_j) = hG(x_j, x_k)$, όπου με G συμβολίζουμε την συνάρτηση Green, η οποία έχει οριστεί μέσω της σχέσης (2.3).

Για οποιαδήποτε διακριτή συνάρτηση $g \in V_h^0$ ορίζεται η συνάρτηση,

$$w_h = T_h g, \qquad w_h = \sum_{k=1}^{n-1} g(x_k) G^k.$$

Τότε,

$$L_h w_h = \sum_{k=1}^{n-1} g(x_k) L_h G^k = \sum_{k=1}^{n-1} g(x_k) e^k = g.$$

Η λύση τώρα u_h του προβλήματος (3.6) ικανοποιεί την $u_h = T_h f$, άρα σύμφωνα με τα παραπάνω,

$$u_h = \sum_{k=1}^{n-1} f(x_k) G^k, \qquad \text{καί} \qquad u_h(x_j) = h \sum_{k=1}^{n-1} G(x_j, x_k) f(x_k). \tag{3.26}$$

Θεώρημα 3.3.1 *Υποθέτουμε ότι η συνάρτηση $f \in \mathcal{C}^2([0,1])$. Τότε το ολικό σφάλμα διακριτοποίησης $e(x_j) = u(x_j) - u_h(x_j)$ ικανοποιεί την ακόλουθη σχέση:*

$$\|u - u_h\|_{h,\infty} \leq \frac{h^2}{96} \|f''\|_\infty,$$

που σημαίνει ότι η u_h συγκλίνει στην u.

Απόδειξη.

Αρχικά θα δείξουμε ότι η σχέση (2.5) $\|u\|_\infty \leq \frac{1}{8}\|f\|_\infty$, η οποία δίνει ένα άνω φράγμα της λύσης του συνεχούς προβλήματος συνοριακών τιμών (2.1)-(2.2) ισχύει καί στην περίπτωση που το πρόβλημα έχει διακριτοποιήθει μέσω της μεθόδου των πεπερασμένων διαφορών.

Αναλυτικότερα από την σχέση (3.26) προκύπτει,

$$|u_h(x_j)| \leq h\sum_{k=1}^{n-1} G(x_j,x_k)|f(x_k)| \leq \|f\|_{h,\infty}\left(h\sum_{k=1}^{n-1} G(x_j,x_k)\right) = \|f\|_{h,\infty}\frac{1}{2}x_j(1-x_j) \leq \frac{1}{8}\|f\|_{h,\infty},$$

δηλαδή,

$$\|u_h\|_{h,\infty} \leq \frac{1}{8}\|f\|_{h,\infty}. \tag{3.27}$$

Η παραπάνω σχέση μας δείχνει ότι η λύση που προκύπτει με τη μέθοδο των πεπερασμένων διαφορών είναι ευσταθής.

Στην παράγραφο (3.2) ορίσαμε με $e_h = u - u_h$ την συνάρτηση του σφάλματος αποκοπής. Παρατηρούμε ότι η συνάρτηση αυτή ικανοποιεί την σχέση $L_h e = T_h$, άρα σύμφωνα με την σχέση (3.26) η συνάρτηση e_h παίρνει την ακόλουθη μορφή,

$$e = \sum_{k=1}^{n-1} T_h(x_k)G^k \qquad \text{καί} \qquad e_h(x_j) = h\sum_{k=1}^{n-1} G(x_j,x_k)T_h(x_k),$$

από την οποία προκύπτει,

$$|e_h(x_j)| \leq h\sum_{k=1}^{n-1} G(x_j,x_k)|T_h(x_k)| \leq \left(h\sum_{k=1}^{n-1} G(x_j,x_k)\right)\|T_h(x_k)\|_{h,\infty} \leq \frac{1}{8}\|T_h(x_k)\|_{h,\infty},$$

εφαρμόζωντας τώρα τον ορισμό της διακριτής άπειρο νόρμα,

$$\|e_h(x_j)\|_{h,\infty} = \max_{1\leq j\leq n-1}|e_h(x_j)|,$$

καί αντικαθιστώντας την σχέση (3.22) προκύπτει η ζητούμενη σχέση,

$$\|e_h\|_{h,\infty} \leq \frac{1}{8}\|T_h\|_{h,\infty} \overset{(3.22)}{\leq} \frac{1}{8}\frac{\|f''\|_\infty}{12}h^2 = \frac{h^2}{96}\|f''\|_\infty.$$

□

Κεφάλαιο 4

Μέθοδος Galerkin.

Στην παράγραφο αυτή θα αναπτύξουμε την μέθοδο Galerkin με πεπερασμένα στοιχεία, για την προσέγγιση του προβλήματος συνοριακών τιμών (2.1)-(2.2). Γενικά η μέθοδος αυτή εφαρμόζεται ευρέως στις αριθμητικές προσεγγίσεις προβλημάτων συνοριακών τιμών.

Αρχικά θεωρούμε το παρακάτω πρόβλημα, το οποίο είναι μια γενικότερη μορφή του προβλήματος (2.1),

$$-(\alpha u')'(x) + (\beta u')(x) + (\gamma u)(x) = f(x) \qquad 0 < x < 1, \tag{4.1}$$

με $u(0) = u(1) = 0$, καί όπου α, β, γ είναι συνεχείς συναρτήσεις στο διάστημα [0,1] με $\alpha(x) \geq a_0 > 0 \qquad \forall x \in [0,1]$.
Στη συνέχεια πολλαπλασιάζουμε την σχέση (4.1) με μία συνάρτηση $\upsilon \in \mathcal{C}^1([0,1])$, η οποία ονομάζεται **συνάρτηση δοκιμής** καί ολοκληρώνουμε την σχέση που προκύπτει στο διάστημα [0,1],

$$-\int_0^1 (\alpha u')'\upsilon dx + \int_0^1 \beta u'\upsilon dx + \int_0^1 \gamma u\upsilon dx = \int_0^1 f\upsilon dx,$$

χρησιμοποιώντας ολοκλήρωση κατά μέρη η παραπάνω σχέση είναι ισοδύναμη με την ακόλουθη,

$$\int_0^1 \alpha u'\upsilon' dx + \int_0^1 \beta u'\upsilon dx + \int_0^1 \gamma u\upsilon dx = \int_0^1 f\upsilon dx + [\alpha u'\upsilon]_0^1.$$

Αν υποθέσουμε ότι η συνάρτηση υ μηδενίζεται στα σημεία $x = 0$ $x = 1$, τότε παρατηρούμε ότι,

$$\int_0^1 \alpha u'\upsilon' dx + \int_0^1 \beta u'\upsilon dx + \int_0^1 \gamma u\upsilon dx = \int_0^1 f\upsilon dx.$$

Ορίζουμε με V τον χώρο όλων των συναρτήσεων δοκιμής, ο οποίος αποτελείται από τις συναρτήσεις υ οι οποίες είναι συνεχείς, μηδενίζονται στα σημεία $x = 0$ καί $x = 1$, καί η πρώτη τους παράγωγος είναι κατά τμήματα συνεχής, που σημαίνει, συνεχής παντού εκτός από ένα πεπερασμένο αριθμό σημείων του διαστήματος [0,1] όπου τα αριστερά καί δεξιά όρια υ'_- καί υ'_+ υπάρχουν αλλά δεν συμπίπτουν απαραίτητα.

Ο χώρος V είναι ουσιαστικά ένας διανυσματικός χώρος ο οποίος συμβολίζεται με $H_0^1(0,1)$. Αναλυτικότερα,

$$H_0^1(0,1) = \{\upsilon \in L^2(0,1) : \upsilon' \in L^2(0,1), \upsilon(0) = \upsilon(1) = 0\}. \tag{4.2}$$

Σημείωση:Ο χώρος H_0^1 όπως γνωρίζουμε από την συναρτησιακή ανάλυση είναι ένας χώρος Sobolev. Τους χώρους αυτούς θα τους χρησιμοποιήσουμε ως εργαλεία για να διατυπώσουμε παρακάτω την ασθενή μορφή του προβλήματος (4.1).

Αν μία συνάρτηση $u \in \mathcal{C}^2([0,1])$ ικανοποιεί την σχέση (4.1) τότε η u αποτελεί επίσης λύση του ακόλουθου προβλήματος,

$$\text{να βρεθεί} \qquad u \in V : \alpha(u,\upsilon) = (f,\upsilon) \qquad \forall \upsilon \in V, \tag{4.3}$$

όπου με $(f,\upsilon) = \int_0^1 f\upsilon dx$ συμβολίζεται το βαθμωτό γινόμενο του χώρου $L^2(0,1)$ καί

$$\alpha(u,\upsilon) = \int_0^1 \alpha u'\upsilon' dx + \int_0^1 \beta u'\upsilon dx + \int_0^1 \gamma u\upsilon dx \tag{4.4}$$

είναι μία διγραμμική μορφή, δηλαδή γραμμική ως προς u καί ως προς υ. Το πρόβλημα (4.3) είναι η **ασθενής μορφή** του προβλήματος (4.1).

ΠΑΡΑΤΗΡΗΣΕΙΣ ΠΑΝΩ ΣΤΗ ΣΧΕΣΗ (4.3):

Παρατήρηση 4.1 *Παρατηρούμε ότι η σχέση (4.3) περιέχει μόνο την πρώτη παράγωγο του u, έτσι μπορεί να καλύψει περιπτώσεις όπου η λύση $u \in \mathcal{C}^2([0,1])$ του προβλήματος (4.1) δεν υπάρχει ωστόσο το φυσικό πρόβλημα είναι καλά ορισμένο. Για παράδειγμα αν θέσουμε α=1, β=γ=0, τότε το πρόβλημα (4.1) παίρνει την ακόλουθη μορφή,*

$$-u''(x) = f(x) \qquad 0 < x < 1.$$

Η λύση $u(x)$ του παραπάνω προβλήματος περιγράφει την μετατόπιση στο σημείο x μίας ελαστικής χορδής με σταθερά άκρα(x=0 καί x=1) η οποία έχει γραμμική πυκνότητα ίση με f καί η θέση της σε κατάσταση ισορροπίας είναι $u(x) = 0 \forall x \in [0,1]$. Αν υποθέσουμε ότι στο διάστημα [0,1] υπάρχουν σημεία ασυνέχειας της συνάρτησης f τότε η u'' στα σημεία αυτά δεν υπάρχει ωστόσο το φυσικό πρόβλημα είναι καλά ορισμένο.

Παρατήρηση 4.2 *Αν υποθέσουμε ότι το πρόβλημα (4.1) έχει τις ακόλουθες μη ομογενής συνοριακές συνθήκες $u(0) = u_0$ καί $u(1) = u_1$, τότε πάλι μπορούμε να παρατηρήσουμε μια διατύπωση παρόμοια με την(4.3). Θέτουμε $\bar{u}(x) = xu_1 + (1-x)u_0$ να είναι η ευθεία γραμμή που παρεμβάλλει τα δεδομένα μεταξύ των τελικών σημείων x=0 καί x=1,*
καί $\overset{0}{u} = u(x) - \bar{u}(x)$, τότε η συνάρτηση $\overset{0}{u} \in V$ ικανοποιεί το ακόλουθο πρόβλημα,

$$\text{να βρεθεί} \qquad \overset{0}{u} \in V : \alpha(\overset{0}{u},\upsilon) = (f,\upsilon) - \alpha(\bar{u},\upsilon) \qquad \forall \upsilon \in V.$$

Στην περίπτωση που έχουμε Neumann συνοριακές συνθήκες, για παράδειγμα $u'(0) = u'(1) = 0$, εφαρμόζουμε την ίδια διαδικασία όπως κάναμε για την απόκτηση της σχέσης (4.3) καί παρατηρούμε ότι η λύση u αυτού του μη ομογενούς προβλήματος Neumann ικανοποιεί το πρόβλημα (4.3) υπό την προϋπόθεση ότι ο χώρος V είναι τώρα ο $H^1(0,1)$.

4.1 Διατύπωση καί ιδιότητες της μεθόδου Galerkin.

Σε αντίθεση με την μέθοδο των πεπερασμένων διαφορών η οποία πηγάζει άμεσα από το πρόβλημα (4.1), η μέθοδος Galerkin βασίζεται στην ασθενή διατύπωση (4.3). Αν V_h είναι ένας διανυσματικός υπόχωρος του V πεπερασμένης διάστασης ο οποίος όπως θα δούμε παρακάτω κατασκευάζεται χρησιμοποιώντας την μέθοδο των πεπερασμένων στοιχείων, τότε η μέθοδος Galerkin προσεγγίζει την ακόλουθη ασθενή μορφή του προβλήματος (4.1),

$$\text{να βρεθεί} \qquad u_h \in V_h : \alpha(u_h, \upsilon_h) = (f, \upsilon_h) \qquad \forall \upsilon_h \in V_h. \tag{4.5}$$

Το παραπάνω πρόβλημα είναι ένα πρόβλημα πεπερασμένης διάστασης. Υποθέτουμε ότι το σύνολο $\{\varphi_1, ..., \varphi_N\}$ αποτελεί μία βάση του χώρου V_h, δηλαδή το σύνολο αυτό αποτελείται από N σε πλήθος γραμμικές καί ανεξάρτητες συναρτήσεις του V_h. Τότε μπορούμε να γράψουμε,

$$u_h(x) = \sum_{j=1}^{N} u_j \varphi_j(x),$$

όπου ο ακέραιος N υποδηλώνει την διάσταση του διανυσματικού χώρου V_h. Λαμβάνοντας $\upsilon_h = \varphi_i$, αποδεικνύεται ότι το πρόβλημα Galerkin (4.5) είναι ισοδύναμο με την αναζήτηση N αγνώστων συντελεστών $\{u_1, ..., u_N\}$ τέτοιοι ώστε,

$$\sum_{j=1}^{N} u_j \alpha(\varphi_j, \varphi_\iota) = (f, \varphi_\iota) \qquad \forall i = 1, ..., N, \tag{4.6}$$

όπου χρησιμοποιήσαμε την παρακάτω ιδιότητα,

$$\alpha\left(\sum_{j=1}^{N} u_j \varphi_j, \varphi_i\right) = \sum_{j=1}^{N} u_j \alpha(\varphi_j, \varphi_i).$$

Αν θέσουμε $\mathbf{u} = [u_1, ..., u_N]$ καί $\mathbf{f_G} = [f_1, ..., f_N]$, όπου $f_i = (f, \varphi_\iota)$, τότε το πρόβλημα (4.6) είναι ισοδύναμο με το ακόλουθο γραμμικό σύστημα,

$$A_G \mathbf{u} = \mathbf{f}_G, \tag{4.7}$$

όπου $A_G = (\alpha_{ij})$ με $\alpha_{ij} = \alpha(\varphi_j, \varphi_i)$ είναι ένας πίνακας, ο οποίος ονομάζεται καί πίνακας ακαμψίας (**stiffnes matrix**), με την δομή του να εξαρτάται από την μορφή που έχουν οι συναρτήσεις βάσης $\{\varphi_i\}$ καί συνεπώς από τη επιλογή του χώρου V_h.

4.2 Ανάλυση της ευστάθειας της μεθόδου.

Εφοδιάζουμε τον χώρο $H_0^1(0,1)$ με την ακόλουθη (ημι)-νόρμα,

$$|\upsilon|_{H^1(0,1)} = \left\{ \int_0^1 |\upsilon'(x)|^2 dx \right\}^{1/2}. \tag{4.8}$$

Αν οι συντελεστές α, β, γ του διαφορικού προβλήματος (4.1) ικανοποιούν την σχέση,

$$-\frac{1}{2}\beta' + \gamma \geq 0, \qquad \forall x \in [0,1], \tag{4.9}$$

τότε το πρόβλημα Galerkin (4.5) έχει **μοναδική λύση**. Εμείς θα ασχοληθούμε με την ειδική περίπτωση όπου β=0 καί $\gamma(x) \geq 0$, καί το πρόβλημα (4.1) λαμβάνει την ακόλουθη μορφή,

$$-(\alpha u')'(x) + (\gamma u)(x) = f(x), \qquad 0 < x < 1.$$

Λαμβάνοντας $\upsilon_h = u_h$ στο πρόβλημα (4.5), μπορούμε να παρατηρήσουμε ότι,

$$\alpha(u_h, u_h) = \int_0^1 \alpha u_h' u_h' dx + \int_0^1 \gamma u_h u_h dx = (f, u_h),$$

καί επειδή,

$$\alpha(u_h, u_h) = \int_0^1 \alpha(x)|u_h'|^2 dx \geq \alpha_0 \int_0^1 |u_h'|^2 dx = \alpha_0 |u_h|^2_{H^1(0,1)},$$

έχουμε ότι,

$$\alpha_0 |u_h|^2_{H^1(0,1)} \leq (f, u_h).$$

Επίσης,

$$(f, u_h) = \int_0^1 f u_h dx \leq \left(\int_0^1 |f|^2 dx\right)^{1/2} \left(\int_0^1 |u_h|^2 dx\right)^{1/2} = \|f\|_{L^2(0,1)} \|u_h\|_{L^2(0,1)},$$

όπου εφαρμόσαμε την ανισότητα Cauchy-Schwarz. Τέλος χρησιμοποιώντας την ανισότητα Poincare,

$$\|u_h\|_{L^2(0,1)} \leq C_p \|u_h^{(1)}\|_{L^2(0,1)},$$

καταλήγουμε στην ακόλουθη ανισοτική σχέση,

$$|u_h|_{H^1(0,1)} \leq \frac{C_p}{\alpha_0} \|f\|_{L^2(0,1)}. \tag{4.10}$$

Παρατήρηση 4.3 *Παρατηρούμε ότι η νόρμα της λύσης του προβλήματος Galerkin είναι φραγμένη υπό την προϋπόθεση ότι η $f \in L^2(0,1)$, συνεπώς η λύση του προβλήματος είναι ευσταθής.*

4.3 Ανάλυση της σύγκλισης της μεθόδου.

Το παρακάτω θεώρημα επιβεβαιώνει τη σύγκλιση της μεθόδου Galerkin.

Θεώρημα 4.3.1 *Έστω $C = \alpha_0^{-1}(\|\alpha\|_\infty + C_p^2 \|\gamma\|_\infty)$, αποδεικνύεται ότι,*

$$|u - u_h|_{H^1(0,1)} \leq C \min_{w_h \in V_h} |u - w_h|_{H^1(0,1)}. \tag{4.11}$$

Απόδειξη.

Αρχικά στο πρόβλημα (4.3) θέτουμε $\upsilon = \upsilon_h$ με $\upsilon_h \in V_h \subset V$, οπότε το πρόβλημα (4.3) παίρνει την ακόλουθη μορφή,

$$\text{να βρεθεί} \qquad u \in V : \alpha(u, \upsilon_h) = (f, \upsilon_h) \qquad \forall \upsilon_h \in V_h. \tag{4.12}$$

Αφαιρώντας την σχέση (4.5) από την (4.12) καί λόγω της διγραμμικής μορφής $\alpha(\cdot, \cdot)$ παρατηρούμε ότι,

$$\alpha(u, \upsilon_h) - \alpha(u_h, \upsilon_h) = 0 \Rightarrow$$
$$\Rightarrow \alpha(u - u_h, \upsilon_h) = 0. \tag{4.13}$$

Στη συνέχεια θέτοντας $e(x) = u(x) - u_h(x)$ καί προσθέτοντας καί αφαιρώντας $w_h \in V_h$ συμπεραίνουμε ότι,

$$\alpha_0 |e|^2_{H^1(0,1)} \leq \alpha(e, e) = \alpha(e, u - w_h) + \alpha(e, w_h - u_h) \qquad \forall w_h \in V_h.$$

Ο τελευταίος όρος σύμφωνα με την σχέση (4.13) είναι ίσος με μηδέν, ενώ ο όρος $\alpha(e, u - w_h)$ χρησιμοποιώντας την ανισότητα Cauchy-Schwarz γράφεται ως εξής,

$$\alpha(e, u - w_h) = \int_0^1 \alpha e'(u - w_h)' dx + \int_0^1 \gamma e(u - w_h) dx \leq$$

$$\leq \|\alpha\|_\infty |e|_{H^1(0,1)} |u - w_h|_{H^1(0,1)} + \|\gamma\|_\infty \|e\|_{L^2(0,1)} \|u - w_h\|_{L^2(0,1)}.$$

Τέλος εφαρμόζοντας την ανισότητα Poincare στους όρους $\|e\|_{L^2}(0,1)$ καί $\|u - w_h\|_{L^2(0,1)}$, δηλαδή

$$\|e\|_{L^2(0,1)} \leq C_p |e|_{H^1(0,1)}$$

καί

$$\|u - w_h\|_{L^2(0,1)} \leq C_p |u - w_h|_{H^1(0,1)},$$

καταλήγουμε στη ζητούμενη σχέση (4.11). □

Στην συνέχεια γενικεύουμε τα παραπάνω αποτελέσματα σε ασθενή προβλήματα της μορφής $\alpha(u, \upsilon) =< f, \upsilon >$. Οι βασικές υποθέσεις επιλυσιμότητας αναφέρονται στον ακόλουθο ορισμό.

Ορισμός 4.3.1 *Υποθέτουμε ότι ο χώρος V είναι ένας χώρος Hilbert εφοδιασμένος με την νόρμα $\|\cdot\|_V$. Λέμε ότι ένα διγραμμικό συναρτησοειδές $\alpha(u, \upsilon) : V \times V \to \mathbb{R}$ είναι*

(i) **πιεστικό**,*αν υπάρχει σταθερά $\alpha_0 > 0$ τέτοια ώστε*

$$\alpha(u, \upsilon) \geq \alpha_0 \|\upsilon\|^2_V \qquad \forall \upsilon \in V,$$

ii) **συνεχές**,*αν υπάρχει $M > 0$ τέτοια ώστε*

$$|\alpha(u, \upsilon)| \leq M \|u\|_V \|\upsilon\|_V \qquad \forall u, \upsilon \in V.$$

Οι ιδιότητες τώρα του διγραμμικού συναρτησιακού μας εξασφαλίζουν την επιλυσιμότητα του γραμμικού συστήματος.

Πρόταση 4.3.1 *Με χρήση του παραπάνω ορισμού θα δείξουμε ότι ο πίνακας A_G του προβλήματος (4.7) είναι θετικά ορισμένος.*

Απόδειξη:
Αρκεί να δείξουμε ότι $\forall \mathbf{v} \in \mathbb{R}^N$ με $\mathbf{v} \neq 0$ το παρακάτω γινόμενο είναι πάντα θετικό δηλαδή,

$$\mathbf{v}^T A_G \mathbf{v} > 0.$$

Αρχικά υποθέτουμε ότι οι συνιστώσες του διανύσματος $\mathbf{v} = (\upsilon_i) \in \mathbb{R}^N$ δίνονται από την ακόλουθη συνάρτηση $\upsilon_h = \sum_{j=1}^{N} \upsilon_j \varphi_j \in V_h$, καί λόγω του παραπάνω ορισμού προκύπτει ότι,

$$\mathbf{v}^T A_G \mathbf{v} = \sum_{j=1}^{N}\sum_{i=1}^{N} \upsilon_i \alpha_{ij} \upsilon_j = \sum_{j=1}^{N}\sum_{i=1}^{N} \upsilon_i \alpha(\varphi_j, \varphi_i)\upsilon_j$$

$$= \sum_{j=1}^{N}\sum_{i=1}^{N} \alpha(\upsilon_j\varphi_j, \upsilon_i\varphi_i) = \alpha\left(\sum_{j=1}^{N} \upsilon_j\varphi_j, \sum_{i=1}^{N} \upsilon_i\varphi_i\right)$$

$$= \alpha(\upsilon_h, \upsilon_h) \geq \alpha_0 \|\upsilon_h\|_V^2 \geq 0.$$

Η ισότητα ισχύει στην περίπτωση που $\upsilon_h = 0$, δηλαδή $\mathbf{v} = 0$ το οποίο είναι άτοπο καθώς υποθέσαμε ότι $\mathbf{v} \neq 0$.

Παρατήρηση 4.4 *Παρατηρούμε ότι ο πίνακας $A_G = (\alpha_{ij})$ με $\alpha_{ij} = \alpha(\varphi_j, \varphi_i)$ είναι συμμετρικός αν καί μόνο αν η διγραμμική μορφή $\alpha(\cdot, \cdot)$ είναι συμμετρική. Για παράδειγμα στο πρόβλημα (4.1) αν υποθέσουμε ότι $\beta = \gamma = 0$, τότε ο πίνακας A_G είναι συμμετρικός καί θετικά ορισμένος, ενώ αν υποθέσουμε ότι $\beta, \gamma \neq 0$, τότε ο πίνακας A_G είναι θετικά ορισμένος κάτω από την υπόθεση (4.9). Τέλος αν ο πίνακας A_G είνα συμμετρικός καί θετικά ορισμένος, τότε το γραμμικό σύστημα (4.7) επιλύεται χρησιμοποιώντας άμεσες μεθόδους όπως η παραγοντιποιήση Cholesky. (Βλ. βιβλία [1,7,9])*

Τόσο τα αποτελέσματα της ευστάθειας όσο καί της σύγλισης της μεθόδου Galerkin μπορούν να παρατηρηθούν καί κάτω από γενικότερες υποθέσεις των προβλημάτων (4.3) καί (4.5).

Θεώρημα 4.3.2 Lax Milgram *Έστω V ένας χώρος Hilbert εφοδιασμένος με τις ιδιότητες του ορισμού 4.3.1. δηλαδή, το διγραμμικό συναρτησιακό είναι πιεστικό καί συνεχές. Τότε για οποιαδήποτε $f \in V^*$ υπάρχει μοναδική λύση $u \in V$ του προβλήματος (4.3), η οποία ικανοποιεί την ακόλουθη σχέση*

$$\|u\| \leq \frac{1}{\alpha_0}\|f\|_{V^*}, \tag{4.14}$$

όπου με V^* συμβολίζουμε τον δυϊκό του χώρου V.

Επίσης το δεξί μέλος του προβλήματος (4.3) ικανοποιεί την ακόλουθη ανισοτική σχέση,

$$|(f,\upsilon)| \leq K\|\upsilon\|_V \qquad \forall \upsilon \in V. \tag{4.15}$$

Από τις σχέσεις (4.14) καί (4.15) προκύπτει τελικά ότι τα προβλήματα (4.3) καί (4.5) έχουν μοναδική λύση η οποία ικανοποιεί τις ακόλουθες σχέσεις αντίστοιχα,

$$\|u\|_V \leq \frac{K}{\alpha_0} \qquad \text{καί} \qquad \|u_h\|_V \leq \frac{K}{\alpha_0}.$$

Λήμμα 4.3.1 Ce'as *Έστω V ένας χώρος Hilbert εφοδιασμένος με τις ιδιότητες του ορισμού 4.3.1. Τότε,*

$$\|u-u_h\|_V \leq \frac{M}{\alpha_0} \min_{w_h \in V_h} \|u-w_h\|_V. \tag{4.16}$$

Η τελευταία σχέση μας εξασφαλίζει την σύγκλιση της μεθόδου Galerkin.

4.4 Μέθοδος των πεπερασμένων στοιχείων.

Η μέθοδος των πεπερασμένων στοιχείων είναι μία ειδική αριθμητική μέθοδος την οποία χρησιμοποιούμε για την κατασκευή του υποχώρου V_h πάνω στον οποίο ορίσαμε το πρόβλημα (4.5). Η μέθοδος αυτή στηρίζεται στην κατά τμήματα πολυωνυμική παρεμβολή καί για το σκοπό αυτό θεωρούμε μία διαμέριση T_h του διαστήματος $[0,1]$ σε n υποδιαστήματα $I_j = [x_j, x_{j+1}]$ με εύρος $h_j = x_{j+1} - x_j$ καί σε n+1 κόμβους

$$0 = x_0 < x_1 < ... < x_{n-1} < x_n = 1,$$

όπου $j = 1, ..., n-1$ καί $n \geq 2$. Επίσης θέτουμε $h = \max_{T_h}(h_j)$.
Στην αρχή του Κεφ.4. ορίσαμε με V τον χώρο όλων των συναρτήσεων δοκιμής καί καταλήξαμε στο συμπέρασμα από τις ιδιότητες του χώρου αυτού ότι ο V ταυτίζεται με τον χώρο Sobolev $H_0^1(0,1)$, ο οποίος περιέχει τις συναρτήσεις που είναι συνεχείς στο (0,1) καί μηδενίζονται στα σημεία x=0 καί x=1.

Ορίζουμε τώρα,

$$X_h^k = \{\upsilon \in \mathcal{C}^0([0,1]) : \upsilon|I_j \in \mathbb{P}_k(I_j) \forall I_j \in T_h\},$$

την οικογένεια των τμηματικών πολυωνύμων στο διάστημα [0,1], με $k \geq 1$. Οποιαδήποτε συνάρτηση $\upsilon_h \in X_h^k$ είναι ένα συνεχές τμηματικό πολυώνυμο στο διάστημα [0,1] του οποίου ο περιορισμός σε κάθε διάστημα $I_j \in T_h$ είναι ένα πολυώνυμο με βαθμό $\leq k$.
Στην συνέχεια θέτουμε,

$$V_h = X_h^{k,0} = \{\upsilon_h \in X_h^k : \upsilon_h(0) = \upsilon_h(1) = 0\}. \tag{4.17}$$

Η διάσταση N του χώρου V_h των πεπερασμένων στοιχείων είναι ίση με $nk-1$.

Για την μελέτη της **ακρίβειας** της μεθόδου Galerkin με πεπερασμένα στοιχεία θεωρούμε αρχικά την τμηματική πολυωνυμική **παρεμβολή Lagrange** $\Pi_h^k u$ της ακριβής λύσης $u \in V$ του προβλήματος (4.3) (για τον ορισμό βλ.βιβλία [1,7,9]) καί τότε από το λήμμα 4.3.2 έχουμε,

$$\min_{w_h \in V_h} \|u - w_h\|_{H_0^1(0,1)} \leq \|u - \Pi_h^k u\|_{H_0^1(0,1)}. \tag{4.18}$$

Από την τελευταία σχέση συμπεραίνουμε ότι το πρόβλημα εκτίμησης του *σφάλματος προσέγγισης* της μεθόδου Galerkin $\|u-u_h\|_{H_0^1(0,1)}$ είναι ισοδύναμο με το πρόβλημα εκτίμησης του *σφάλματος παρεμβολής* $\|u - \Pi_h^k u\|_{H_0^1(0,1)}$.

Λήμμα 4.4.1 *Έστω $u \in H_0^1(0,1)$ η ακριβής λύση του προβήματος (4.3) καί $u_h \in V_h$ η αντίστοιχη προσεγγιστική η οποία προκύπτει χρησιμοποιώντας συνεχή τμηματικά πολυώνυμα βαθμού $k \geq 1$. Υποθέτουμε επίσης ότι $u \in H^s$, όπου $s \geq 2$. Κάτω από τις προϋποθέσεις αυτές ισχύει η ακόλουθη ανισοτική σχέση,*

$$\|u - \Pi_h^k u\|_{H_0^1(0,1)} \leq C h^l \|u\|_{H^{l+1}(0,1)}, \tag{4.19}$$

όπου C είναι μία θετική σταθερά ανεξάρτητη του h καί $l = min(k, s-1)$.

Απόδειξη.

(βλ.βιβλία [1,7,9]) □

Τέλος αν συνδιάσουμε τις σχέσεις (4.16),(4.18) καί (4.19) προκύπτει η ακόλουθη εκτίμηση του σφάλματος της μεθόδου Galerkin με πεπερασμένα στοιχεία

$$\|u - u_h\|_{H_0^1(0,1)} \leq \frac{M}{\alpha_0} C h^l \|u\|_{H^{l+1}(0,1)}. \tag{4.20}$$

Παρατήρηση 4.5 *Υπάρχουν δύο τρόποι για την βελτίωση της ακρίβειας της μεθόδου. Ένας τρόπος είναι να αυξήσουμε τον βαθμό k των τμηματικών πολυωνύμων αλλά κάτι τέτοιο σύμφωνα με την σχέση (4.19) απαιτεί η λύση u να είναι επαρκώς ομαλή συνάρτηση, καί ο δεύτερος τρόπος τον οποίο καί εφαρμόζουμε στην μέθοδο αυτή είναι να μειώσουμε το μέγεθος του βήματος h.*

Παρατήρηση 4.6 *Από την τελευταία σχέση συμπεραίνουμε ότι η προσέγγιση του σφάλματος τείνει στο μηδέν όταν $h \to 0$, δηλαδή η μέθοδος Galerkin συγκλίνει καί η* **τάξη σύγκλισης** *είναι l.*

Παρατήρηση 4.7 *Μία από τις υποθέσεις του λήμματος 4.4.1 ήταν ότι η ακριβής λύση $u \in H^s$ όπου $s \geq 2$. Στην περίπτωση τώρα που η λύση έχει ομαλότητα s=1, τότε προφανώς για τη μελέτη της σύγκλισης της μεθόδου Galerkin δεν μπορούμε να χρησιμοποιήσουμε την σχέση (4.20). Σε αυτή την περίπτωση το λήμμα Cea's μας εξασφαλίζει την σύγκλιση της μεθόδου καθώς όταν το $h \to 0$ ο υπόχωρος V_h γίνεται πυκνός στον χώρο V.*

Παρακάτω δίνεται ο πίνακας 4.1 ο οποίος συνοψίζει την τάξη σύγκλισης της μεθόδου με πεπερασμένα στοιχεία για $k = 1, ..., 4$ καί $s = 1, ..., 5$.

k	s = 1	s = 2	s = 3	s = 4	s = 5
1	συγκλίνει	h^1	h^1	h^1	h^1
2	συγκλίνει	h^1	h^2	h^2	h^2
3	συγκλίνει	h^1	h^2	h^3	h^3
4	συγκλίνει	h^1	h^2	h^3	h^4

Πίνακας 4.1.Τάξη σύγκλισης της μεθόδου με πεπερασμένα στοιχεία συναρτήσει του k (βαθμός παρεμβολής) καί του s (βαθμός ομαλότητα της λύσης u).

Παραπάνω ορίσαμε με X_h^k τον χώρο που περιέχει τα τμηματικά πολυώνυμα στο [0,1]. Θα ασχοληθούμε τώρα με την κατασκευή μίας κατάλληλης βάσης $\{\varphi_j\}$ για τον χώρο αυτό στις περιπτώσεις όπου το k=1 καί k=2. Για τον σκοπό αυτό επιλέγουμε ένα σύνολο από κατάλληλες παραμέτρους (*ή βαθμούς ελευθερίας*) για κάθε στοιχείο I_j της διαμέρισης T_h, οι οποίες προσδιορίζουν μοναδικά μία συνάρτηση η οποία ανήκει στον χώρο X_h^k.
Επομένως η συνάρτηση υ_h στον χώρο X_h^k μπορεί να γραφεί ως γραμμικός συνδιασμός των στοιχείων της βάσης,

$$\upsilon_h(x) = \sum_{i=0}^{nk} \upsilon_i \varphi_i(x),$$

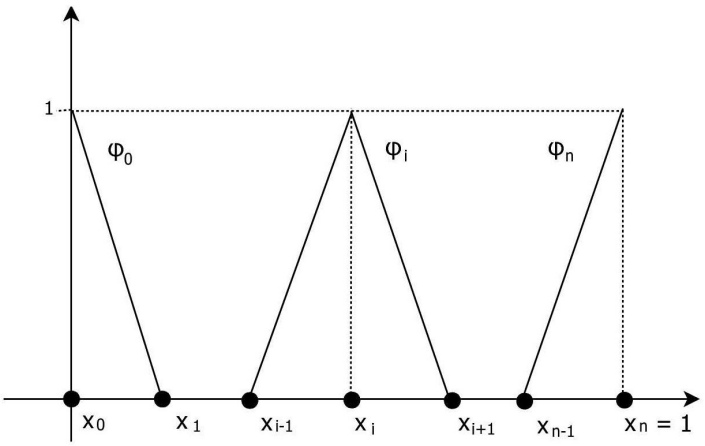
όπου με $\{\upsilon_i\}$ συμβολίζουμε το σύνολο των παραμέτρων της υ_h. Επίσης υποθέτουμε ότι οι συναρτήσεις φ_i, οι οποίες αποτελούν την βάση του χώρου X_h^k, ικανοποιούν την ιδιότητα της παρεμβολής Lagrange $\varphi_i(x_j) = \delta_{ij}, i, j = 0, ..., n$, όπου δ_{ij} είναι το σύμβολο του Kronecker(βλ.σχέση 3.25).

Ο χώρος X_h^1
Ο χώρος αυτός αποτελείται από όλες τις συναρτήσεις οι οποίες είναι συνεχείς καί γραμμικές σε κάθε υποδιάστημα I_j της διαμέρισης T_h. Από τη στιγμή που μία μοναδική γραμμή διέρχεται από δύο διακριτούς κόμβους ο αριθμός των παραμέτρων που περιγράφει μοναδικά μία συνάρτηση στον χώρο X_h^1 είναι ίσος με τον αριθμό των κόμβων, δηλαδή ίσος με n+1. Επομένως για να παράγουμε τον χώρο X_h^1 χρειάζονται n+1 συναρτήσεις $\{\varphi_i\}_{i=0}^n$. Ορίζω λοιπόν τις

συναρτήσεις φ_i με $i = 1, ..., n-1$ ως εξής,

$$\varphi_i(x) = \begin{cases} \frac{x - x_{i-1}}{x_i - x_{i-1}} & x_{i-1} \le x \le x_i, \\ \frac{x_{i+1} - x}{x_{i+1} - x_i} & x_i \le x \le x_{i+1}, \\ 0 & \text{αλλιώς.} \end{cases} \tag{4.21}$$

Κάθε συνάρτηση φ_i έχει μορφή "*στέγης*" όπως φαίνεται καί από το σχήμα που ακολουθεί, είναι συνεχής, ομαλή καί μη - μηδενική σε ένα μικρό τμήμα "περί του κόμβου i". Επίσης παίρνει την τιμή 1 στον κόμβο x_i καί την τιμή 0 στους υπόλοιπους κόμβους της διαμέρισης. Η ένωση των τμημάτων γύρω από τους κόμβους i όπου οι φ_i συναρτήσεις είναι μη - μηδενικές αποτελεί ένα υποσύνολο του διαστήματος $[0, 1]$ το οποίο ονομάζεται *υποστήριγμα (support)*. Αναλυτικότερα το υποσύνολο αυτό αποτελείται από την ένωση των διαστημάτων I_{j-1} καί I_j όταν $1 \le i \le n-1$, ενώ όταν $i = 0$ συμπίπτει με το διάστημα I_0.



Σχήμα 4.1 Συναρτήσεις στέγες

Σε οποιοδήποτε διάστημα $I_i = [x_i, x_{i+1}]$,όπου $i = 0, ..., n-1$, οι συναρτήσεις φ_i καί φ_{i+1} μπορούν να θεωρηθούν ως οι εικόνες δύο συναρτήσεων "*αναφοράς*" $\widehat{\varphi}_0$ καί $\widehat{\varphi}_1$ (οι οποίες ορίζονται στο *διάστημα αναφοράς* $[0, 1]$) μέσω της ακόλουθης ομοπαραλληλικής απεικόνισης

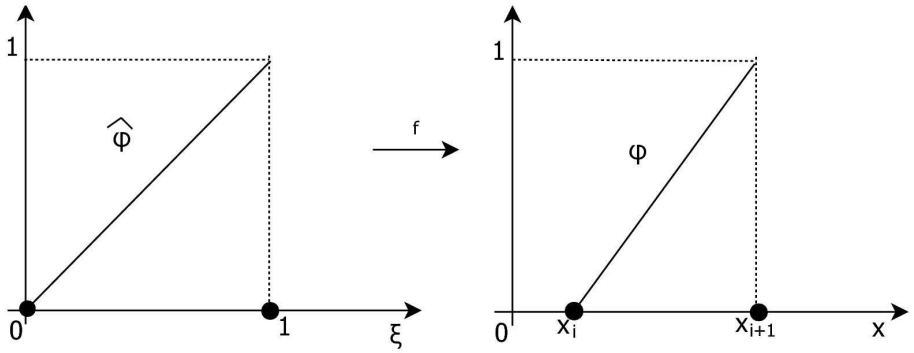
$$f : [0, 1] \to I_i$$

$$x = f(\xi) = x_i + \xi(x_{i+1} - x_i), \qquad i = 0, ..., n-1. \tag{4.22}$$

Στη συνέχεια ορίζοντας $\widehat{\varphi}_0(\xi) = 1 - \xi$ καί $\widehat{\varphi}_1(\xi) = \xi$, οι συναρτήσεις φ_i καί φ_{i+1} μπορούν να κατασκευαστούν στο διάστημα I_i ως εξής,

$$\varphi_i(x) = \widehat{\varphi}_0(\xi(x)), \qquad \varphi_{i+1} = \widehat{\varphi}_i(\xi(x)),$$

όπου $\xi(x) = (x - x_i)\backslash(x_{i+1} - x_i)$.



Σχήμα 4.2 Ομοπαραλληλική απεικόνιση f από το διάστημα αναφοράς στο γενικό διάστημα της διαμέρισης

Ο χώρος X_h^2

Ο χώρος αυτός αποτελείται από όλες τις συναρτήσεις υ_h οι οποίες είναι τμηματικά πολυώνυμα δευτέρου βαθμού σε κάθε διάστημα I_i. Για να προσδιορίσουμε τα πολυώνυμα αυτά χρειαζόμαστε τρείς τιμές αυτών σε τρεία διακριτά σημεία των διαστημάτων I_i καί για να εξασφαλίσουμε την συνέχεια της συνάρτησης υ_h στο διάστημα [0,1] οι παράμετροι που προσδιορίζουν μοναδικά την υ_h, επιλέγονται να είναι οι τιμές της συνάρτησης στους κόμβους x_i της διαμέρισης T_h με $i = 0, ..., n$, καί στα μέσα των διαστημάτων I_i, με $i = 0, ..., n-1$. Έτσι ο αριθμός των παραμέτρων είναι ίσος με 2n+1, επομένως χρειζόμαστε 2n+1 συναρτήσεις φ_i με $i = 0, ..., 2n$ για να παράγουμε τον χώρο X_h^2.

Πριν ορίσουμε τις συναρτήσεις αυτές είναι βολικό, αφού πρώτα έχουμε διαμερίσει το διάστημα [0,1] σε 2n+1 κόμβους ξεκινώντας από $x_0 = 0$ μέχρι $x_{2n} = 1$, να αντιστοιχίσουμε τα μέσα των διαστημάτων στους κόμβους με περιττό δείκτη καί τα άκρα των διαστημάτων στους κόμβους με άρτιο δείκτη.

Επιλέγουμε τώρα τις συναρτήσεις φ_i ανάλογα με το αν ο δείκτης i είναι άρτιος ή περιττός, έτσι για i άρτιο, δηλαδή στα άκρα των διαστημάτων οι συναρτήσεις φ_i έχουν την ακόλουθη μορφή,

$$\varphi_i(x) = \begin{cases} \frac{(x-x_{i-1})(x-x_{i-2})}{(x_i-x_{i-1})(x_i-x_{i-2})} & x_{i-2} \le x \le x_i, \\ \frac{(x_{i+1}-x)(x_{i+2}-x)}{(x_{i+1}-x_i)(x_{i+2}-x_i)} & x_i \le x \le x_{i+2}, \\ 0 & \text{αλλιώς.} \end{cases} \tag{4.23}$$

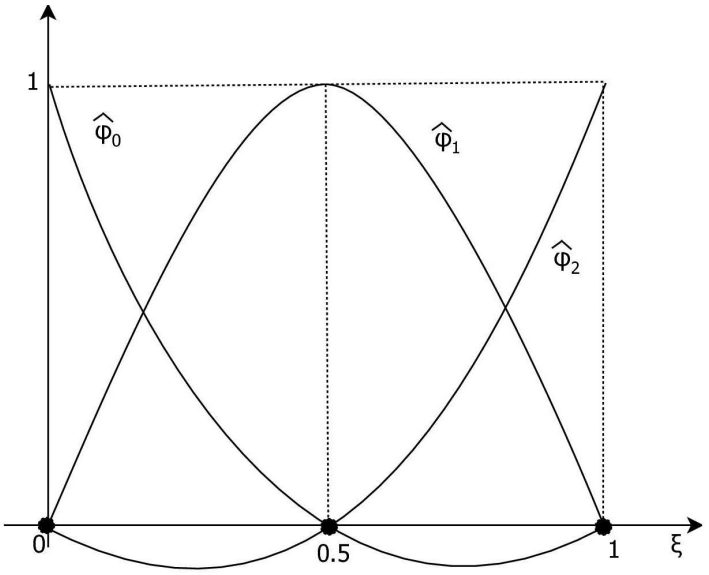
ενώ για i περιττό, δηλαδή στα μέσα των διαστημάτων,

$$\varphi_i(x) = \begin{cases} \frac{(x_{i+1}-x)(x-x_{i-1})}{(x_{i+1}-x_i)(x_i-x_{i-1})} & x_{i-1} \le x \le x_{i+1}, \\ \\ 0 & \text{αλλιώς.} \end{cases} \tag{4.24}$$

Κάθε συνάρτηση φ_i ικανοποιεί την ιδιότητα της παρεμβολής Lagrange $\varphi_i(x_j) = \delta_{ij}$, όπου $i, j = 0, ..., 2n$. Οι αντίστοιχες τώρα συναρτήσεις στο *διάστημα αναφοράς* [0,1] οι οποίες απεικονίζονται καί στο σχήμα που ακολουθεί είναι,

$$\widehat{\varphi}_0(\xi) = (1-\xi)(1-2\xi), \widehat{\varphi}_1(\xi) = 4(1-\xi)\xi, \widehat{\varphi}_2(\xi) = \xi(2\xi-1). \tag{4.25}$$

Όπως ακριβώς καί στην περίπτωση του χώρου X_h^1 οι συναρτήσεις (4.23) καί (4.24) αποτελούν τις εικόνες των συναρτήσεων (4.25) μέσω της ομοπαραλληλικής απεικόνισης (4.22).



Σχήμα 4.3 Συναρτήσεις βάσης του χώρου X_h^2 στο διάστημα αναφοράς.

Μέχρι στιγμής έχουμε ασχοληθεί μόνο με τις συναρτήσεις βάσης οι οποίες ικανοποιούν τις ιδιότητες της παρεμβολής Lagrange. Αν αυτός ο περιορισμός πάψει να υφίσταται, τότε θα προκύψουν διαφορετικά είδη συναρτήσεων βάσης. Ένα παράδειγμα στο *διάστημα αναφοράς* αποτελούν οι *ιεραρχικές* συναρτήσεις βάσης,

$$\widehat{\psi}_0(\xi) = 1-\xi, \widehat{\psi}_1(\xi) = (1-\xi)\xi, \widehat{\psi}_2(\xi) = \xi, \tag{4.26}$$

οι οποίες ονομάζονται έτσι, διότι παράγονται χρησιμοποιώντας τις συναρτήσεις βάσης του υποχώρου με την αμέσως μικρότερη διάσταση, δηλαδή του X_h^1. Για την ακρίβεια η συνάρτηση

$\widehat{\psi_1} \in X_h^2$ προκύπτει ως συνδιασμός των συναρτήσεων $\widehat{\psi_0}$ καί $\widehat{\psi_2}$ οι οποίες ανήκουν στον χώρο X_h^1.
Για να ελέγξουμε τώρα ότι οι συναρτήσεις (4.26) αποτελούν όντως μία βάση του χώρου X_h^2, αρκεί να επαληθεύσουμε ότι οι συναρτήσεις αυτές είναι γραμμικά ανεξάρτητες,δηλαδή

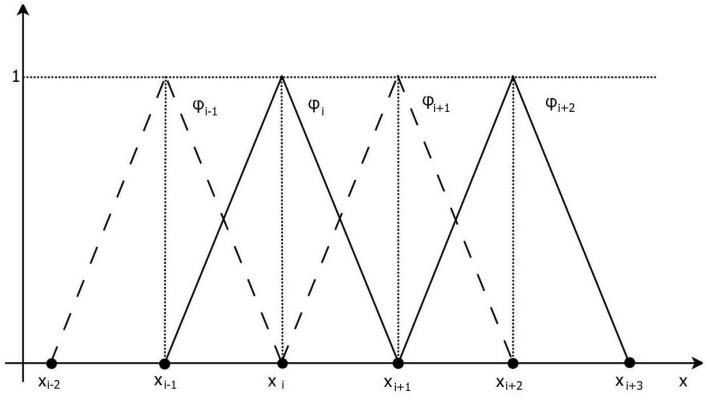
$$\alpha_0\widehat{\psi_0}(\xi) + \alpha_1\widehat{\psi_1}(\xi) + \alpha_2\widehat{\psi_2}(\xi) = 0, \forall \xi \in [0,1] \Leftrightarrow \alpha_0 = \alpha_1 = \alpha_2 = 0,$$

πράγμα που στην περίπτωση μας ισχύει καθώς αν

$$\sum_{i=0}^{2} \alpha_i\widehat{\psi_i}(\xi) = \alpha_0 + \xi(\alpha_1 - \alpha_0 + \alpha_2) - \alpha_1\xi^2 = 0 \qquad \forall \xi \in [0,1],$$

τότε υποχρεωτικά έχουμε ότι $\alpha_0 = \alpha_1 = 0$ καί άρα $\alpha_2 = 0$.

Ανάλογες διαδικασίες με παραπάνω μπορούν να εφαρμοστούν καί για την κατασκευή της βάσης οποιουδήποτε υποχώρου X_h^k, όπου το k είναι αυθαίρετο. Επισημαίνουμε ότι η αύξηση του βαθμού k της πολυωνυμικής προσέγγισης καί κατ'επέκταση του πλήθους των βαθμών ελευθερίας έχει ως άμεση συνέπεια την αύξηση του κόστους υπολογισμού της λύσης του γραμμικού συστήματος (4.7).
Θα ασχοληθούμε τώρα με την δομή καί τις ιδιότητες του πίνακα ακαμψίας A_G του γραμμικού συστήματος (4.7) στην περίπτωση της μεθόδου με πεπερασμένα στοιχεία ($A_G = A_{fe}$). Η βασική ιδέα της μεθόδου είναι να ανάγουμε το υπολογιστικό πρόβλημα στη λύση ενός γραμμικού συστήματος του οποίου ο πίνακας περιέχει "λίγα " μη μηδενικά στοιχεία.
Πράγματι, από την στιγμή που οι συναρτήσεις βάσης του υποχώρου X_h^k έχουν τοπικό υποστήριγμα τα περισσότερα στοιχεία του πίνακα A_{fe} είναι μηδενικά. Συγκεκριμένα στην περίπτωση όπου $k = 1$ το υποστήριγμα της συνάρτησης στέγης φ_i όπως έχουμε αναφέρει είναι η ένωση των διαστημάτων I_{i-1} καί I_i όταν $1 \leq i \leq n-1$, ενώ όταν το $i = 0$ συμπίπτει με το διάστημα I_0. Άμεση συνέπεια των παραπάνω είναι ότι για ένα σταθερό $i = 1, ..., n-1$ οι συναρτήσεις στέγες φ_{i-1} καί φ_{i+1} είναι οι μοναδικές που έχουν μη - μηδενικό υποστήριγμα, το οποίο διασταυρώνεται με το υποστήριγμα της φ_i (βλ.σχήμα 4.4). Έτσι καταλήγουμε στο συμπέρασμα ότι ο πίνακας A_{fe} είναι τριδιαγώνιος καθώς $\alpha_{ij} = 0$ όταν $j \notin \{i-1, i, i+1\}$.



Σχήμα 4.4 Οι συναρτήσεις στέγες φ_{i-1}, φ_i, φ_{i+1} καί φ_{i+2}.

Θα υπολογίσουμε τώρα αναλυτικά τα στοιχεία α_{ij} του πίνακα A_{fe}. Αρχικά θεωρούμε μία ομοιόμορφη διαμέριση του διαστήματος [0,1] με $x_i = x_{i-1} + h$, $i = 1, ..., n$ καί $h = \frac{1}{n}$. Επίσης όπου φ_i είναι οι συναρτήσεις που δίνονται από την σχέση (4.21), ενώ όπου φ_i' είναι οι αντίστοιχες παράγωγοι τους, δηλαδή

$$\varphi_i'(x) = \begin{cases} \frac{1}{h} & x_{i-1} \le x \le x_i, \\ -\frac{1}{h} & x_i \le x \le x_{i+1}. \end{cases} \tag{4.27}$$

Στην συνέχεια χρησιμοποιώντας την διγραμμική μορφή (4.4) καί υποθέτοντας ότι α, β, γ είναι σταθερές έχουμε ότι,

$$\alpha_{i,i-1} = \alpha(\varphi_i, \varphi_{i-1}) = \int_0^1 \alpha\varphi_i'(x)\varphi_{i-1}'(x)dx + \int_0^1 \beta\varphi_i'(x)\varphi_{i-1}(x)dx + \int_0^1 \gamma\varphi_i(x)\varphi_{i-1}(x)dx =$$

$$= \sum_{i=1}^{n} \left(\int_{x_{i-1}}^{x_i} \alpha\varphi_i'(x)\varphi_{i-1}'(x)dx + \int_{x_{i-1}}^{x_i} \beta\varphi_i'(x)\varphi_{i-1}(x)dx + \int_{x_{i-1}}^{x_i} \gamma\varphi_i(x)\varphi_{i-1}(x)dx \right) = -\frac{\alpha}{h} + \frac{\beta}{2} + \frac{\gamma h}{6},$$

Ομοίως προκύπτουν καί τα υπόλοιπα στοιχεία των διαγωνίων του πίνακα A_{fe},

$$\alpha_{i,i+1} = -\frac{\alpha}{h} - \frac{\beta}{2} + \frac{\gamma h}{6} \qquad \alpha_{i,i} = \frac{2\alpha}{h} + \frac{2}{3}\gamma h,$$

καί ο πίνακας A_{fe} έχει τελικά την ακόλουθη μορφή,

$$A_{fe} = \begin{bmatrix} \frac{2\alpha}{h} + \frac{2\gamma h}{3} & -\frac{\alpha}{h} - \frac{\beta}{2} + \frac{\gamma h}{6} & 0 & \cdots & 0 \\ -\frac{\alpha}{h} + \frac{\beta}{2} + \frac{\gamma h}{6} & \frac{2\alpha}{h} + \frac{2\gamma h}{3} & -\frac{\alpha}{h} - \frac{\beta}{2} + \frac{\gamma h}{6} & & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \cdots & -\frac{\alpha}{h} + \frac{\beta}{2} + \frac{\gamma h}{6} & \frac{2\alpha}{h} + \frac{2\gamma h}{3} & -\frac{\alpha}{h} - \frac{\beta}{2} + \frac{\gamma h}{6} \\ 0 & \cdots & 0 & -\frac{\alpha}{h} + \frac{\beta}{2} + \frac{\gamma h}{6} & \frac{2\alpha}{h} + \frac{2\gamma h}{3} \end{bmatrix}$$

Παρακάτω δίνεται ο δείκτης κατάστασης του πίνακα A_{fe} χρησιμοποιώντας την $\|\cdot\|_2$,

$$K_2(A_{fe}) = \|A_{fe}\|_2 \|A_{fe}^{-1}\|_2 = \mathcal{O}(h^{-2}).$$

Ο δείκτης αυτός μας δίνει ένα φράγμα για το πόσο ανακριβής μπορεί να είναι η λύση του συστήματος (4.7) μέτα την εφαρμογή της μεθόδου των πεπεραμένων στοιχείων. Έχουμε

αναφέρει ότι για να βελτιώσουμε την ακρίβεια της μεθόδου αρκεί να μειώσουμε το μέγεθος του βήματος h. Παρατηρούμε όμως ότι όταν το $h \to 0$ το $K_2(A_{fe}) \to \infty$ το οποίο έρχεται σε αντίφαση με τα προπάνω.

Παρατήρηση 4.8 *Επειδή ο δείκτης κατάστασης του γραμμικού συστήματος (4.7) είναι πολύ μεγαλύτερος του 1* ($>> 1$) *το σύστημα καλείται "κακής κατάστασης" καί για την επίλυση του απαιτούνται ειδικές τεχνικές. Επισημαίνεται πάντως πως η επιλυσιμότητα των γραμμικών συστημάτων που προκύπτουν με τη διαδικασία διακριτοποίησης των πεπερασμένων στοιχείων, εξασφαλίζεται με σχετικά γρήγορο καί ευσταθή τρόπο. (βλ. κεφάλαια 3,4,5 του [1])*

4.5 Φασματικές μέθοδοι

Η βασική διαφορά των φασματικών μεθόδων σε σχέση με τις μεθόδους πεπερασμένων στοιχείων είναι ότι οι πρώτες προσεγγίζουν την λύση ως γραμμικό συνδιασμό συνεχών συναρτήσεων οι οποίες είναι μη-μηδενικές σε ολόκληρη την περιοχή της λύσης,ενώ στην μέθοδο των πεπερασμένων στοιχείων οι συναρτήσεις είναι μη-μηδενικές σε μικρές υποπεριοχές. Έτσι στην πρώτη περίπτωση έχουμε μία συνολική προσέγγιση της λύσης σε αντίθεση με την δεύτερη περίπτωση όπου η προσέγγιση γίνεται τοπικά.
Η μέθοδος αυτή εφαρμόζεται είτε χρησιμοποιώντας την μέθοδο ταξινόμησης είτε την μέθοδο Galerkin. Στην παράγραφο αυτή θα ασχοληθούμε με το πως διαμορφώνεται το πρόβλημα προσέγγισης της μεθόδου Galerkin (4.5) χρησιμοποιώντας τις φασματικές μεθόδους.

Αρχικά θεωρούμε ότι το πρόβλημα (4.1) ορίζεται στο διάστημα $(-1, 1)$. Στη συνέχεια ορίζουμε μία κατανομή κόμβων $\{x_0, ..., x_n\}$, η οποία συμπίπτει με τα $n+1$ Legendre-Gauss-Lobatto σημεία. Τα σημεία αυτά είναι οι $n+1$ ρίζες των πολυωνύμων P_{n+1} βαθμού $n+1$ στο διάστημα $[-1, 1]$, όπου $\{P_k\}$ είναι μία ακολουθία ορθογωνίων πολυωνύμων στο [-1,1] με συνάρτηση βάρους $w = 1$. Παρακάτω δίνεται ο ορισμός των ορθογωνίων πολυωνύμων.

Ορισμός 4.5.1 *Τα πολυώνυμα P_k με $k = 0, 1, ...,$ καλούνται ορθογώνια με συνάρτηση βάρους $w = 1$, στο γενικό διάστημα $[a, b]$ αν,*

$$(i) \textit{ο βαθμός του} P_k \textit{είναι} k, k = 0, 1, 2, ...,$$

καί

$$(ii) \int_a^b P_i P_j dx = \begin{cases} 0 & \alpha\nu \quad i \neq j, \\ \neq 0 & \alpha\nu \quad i = j. \end{cases}$$

Η ακολουθία τώρα $\{P_k\}$ των ορθογωνίων πολυωνύμων μπορεί να κατασκευαστεί στο γενικό διάστημα $[a, b]$ από τα αντίστοιχα πολυώνυμα Legendre p_k στο $[-1, 1]$ με τον τύπο

$$P_k(x) = \frac{2}{b-a} p_k(\frac{a+b}{2} + \frac{b-a}{2}x),$$

καί τα πολυώνυμα Legendre με την σειρά τους, με τον επαναληπτικό τύπο

$$(k+1)p_{k+1} - (2k+1)xp_k + kp_{k-1} = 0 \qquad \text{με} \qquad p_0 = 1 \qquad \text{καί} \qquad p_1 = x.$$

Σύμφωνα με τον κανόνα τεραγωνισμού του Gauss το βαθμωτό γινόμενο $(u, \upsilon) = \int_{-1}^{1} u\upsilon dx$ στον χώρο $L^2(-1,1)$ προσεγγίζεται από το ακόλουθο άθροισμα,

$$(u, \upsilon)_n \approx \sum_{j=0}^{n} u(x_j)\upsilon(x_j)w_j, \tag{4.28}$$

όπου w είναι μία αυστηρά θετική ολοκληρώσιμη συνάρτηση η οποία καλείται συνάρτηση βάρους του τύπου τετραγωνισμού Legendre-Gauss-Lobatto.
Επίσης υποθέτουμε ότι η προσεγγιστική λύση u_h είναι ένα πολυώνυμο βαθμού n καί ορίζουμε με $\mathbb{P}_n^0$ το σύνολο των πολυωνύμων $p \in \mathbb{P}_n([-1,1])$ έτσι ώστε $p(-1) = p(1) = 0$.
Το πρόβλημα Galerkin (4.5) παίρνει τώρα την ακόλουθη μορφή,

$$\text{να βρεθεί} \qquad u_n \in \mathbb{P}_n^0 : a_n(u_n, \upsilon_n) = (f, \upsilon_n)_n \qquad \forall \upsilon_n \in \mathbb{P}_n^0, \tag{4.29}$$

όπου α_n είναι μία διγραμμική μορφή η οποία προκύπτει από την σχέση (4.4) αντικαθιστώντας τα ολοκληρώματα από την σχέση (4.28), δηλαδή

$$\alpha_n(u_n, \upsilon_n) = (au_n', \upsilon_n')_n + (\beta u_n', \upsilon_n)_n + (\gamma u_n, \upsilon_n)_n.$$

Η παραπάνω προσέγγιση (4.29) ονομάζεται τώρα *γενικευμένη προσέγγιση Galerkin*. Η ανάλυση της απαιτεί περισσότερη προσοχή σε σχέση με την ανάλυση της μεθόδου Galerkin καί βρίσκεται εκτός πλαισίου αυτής της διπλωματικής εργασίας.

Κεφάλαιο 5

Εξισώσεις μεταφοράς-διάχυσης.

Τα συνοριακά προβλήματα της μορφής (4.1) χρησιμοποιούνται για να περιγράψουν διαδικασίες διάχυσης, μεταφοράς καί απορρόφησης μίας συγκεκριμένης ποσότητας η οποία συμβολίζεται με u(x). Με τον όρο $-(au')'$ συμβολίζουμε την διάχυση, με $\beta u'$ την μεταφορά καί με γu την την απορρόφηση.
Για παράδειγμα αν θεωρήσουμε ότι $\alpha = \varepsilon$, $\gamma = 0$ καί $\beta =$ σταθερά $>> 1$ τότε έχουμε το ακόλουθο πρόβλημα συνορικών τιμών,

$$-\varepsilon u'' + \beta u' = 0, \qquad 0 < x < 1, \tag{5.1}$$

$$u(0) = 0, \qquad u(1) = 1, \tag{5.2}$$

όπου ε, β είναι δύο θετικές σταθερές τέτοιες ώστε $\frac{\varepsilon}{\beta} << 1$. Το παραπάνω πρόβλημα μπορεί να είναι απλό, αλλά περιγράφει ένα πολύ ενδιαφέρον πρόβλημα μεταφοράς-διάχυσης, όπου η μεταφορά υπερισχύει της διάχυσης.

Ορισμός 5.0.2 *Ορίζουμε τον παγκόσμιο αριθμό Péclet ως,*

$$\mathbb{P}e_{gl} = \frac{|\beta|L}{2\varepsilon}, \tag{5.3}$$

όπου L είναι το εύρος του πεδίου ορισμού (στην περίπτωση μας είναι ίσο με 1). Ο αριθμός Peclet μετράει την υπεροχή του όρου που συμβολίζει την μεταφορά έναντι του όρου που συμβολίζει την διάχυση.

Η αντίστοιχη τώρα ασθενής διατύπωση του προβλήματος (5.1) είναι,

$$\text{να βρεθεί} \qquad u \in H_0^1(0,1) : a(u,\upsilon) = 0 \qquad \forall \upsilon \in H_0^1(0,1), \tag{5.4}$$

όπου $\alpha(\cdot,\cdot)$ είναι η διγραμμική μορφή η οποία δίνεται από την ακόλουθη σχέση,

$$\alpha(u,\upsilon) = \varepsilon \int_0^1 u'\upsilon' dx + \beta \int_0^1 u'\upsilon dx.$$

Λόγω της συνέχειας του διγραμμικού συναρτησοειδούς έχουμε ότι,

$$|\alpha(u,\upsilon)| \leq \varepsilon \|u'\|_{L^2(0,1)} \|\upsilon'\|_{L^2(0,1)} + \beta \|u'\|_{L^2(0,1)} \|\upsilon\|_{L^2(0,1)}.$$

Εφαρμόζωντας την ανισότητα Poincare,

$$\|\upsilon\|_{L^2(0,1)} \leq C_p \|\upsilon'\|_{L^2(0,1)},$$

προκύπτει τελικά ότι,

$$|a(u,\upsilon)| \leq \varepsilon \|u'\|_{L^2(0,1)} \|\upsilon'\|_{L^2(0,1)} + \beta C_p \|u'\|_{L^2(0,1)} \|\upsilon'\|_{L^2(0,1)} \leq$$

$$\leq (\varepsilon + \beta C_p) \|u'\|_{L^2(0,1)} \|\upsilon'\|_{L^2(0,1)}.$$

Στην συνέχεια θέτουμε $M = \varepsilon + \beta C_p$ καί $\alpha_0 = \varepsilon$ καί η εκτίμηση σφάλματος της μεθόδου Galerkin η οποία δίνεται από την σχέση (4.20) στην ειδική περίπτωση όπου το $k = 1$ παίρνει τώρα την ακόλουθη μορφή,

$$\|u - u_h\|_{H_0^1(0,1)} \leq \frac{(\varepsilon + \beta C_p)}{\varepsilon} ch \|u\|_{H^2(0,1)} \leq ch \|u\|_{H^2(0,1)} + \frac{\beta C_p}{\varepsilon} ch \|u\|_{H^2(0,1)}$$

Από την τελευταία σχέση παρατηρούμε ότι η εκτίμηση του σφάλματος για να έχει νόημα καί να είναι τουλάχιστον φραγμένη για μικρές τιμές του ε πρέπει

$$h \leq \frac{\varepsilon}{\beta C_p c} \qquad \text{ή ισοδύναμα όταν} \qquad h \approx \left(\frac{M}{a_0}\right)^{-1}.$$

Έτσι καταλήγουμε στο συμπέρασμα ότι στην περίπτωση που ο όρος α είναι πολύ μικρότερος από τον όρο β ή τον όρο γ του προβλήματος (4.1) η μέθοδος Galerkin μπορεί να είναι ακατάλληλη στο να επιφέρει ακριβή αριθμητικά αποτελέσματα.

Υπολογίζουμε τώρα την ακριβή λύση του προβλήματος (5.1)-(5.2). Παρατηρούμε ότι πρόκειται για μία διαφορική εξίσωση με σταθερούς συντελεστές καί με χαρακτηριστική εξίσωση την $-\varepsilon\lambda^2 + \beta\lambda = 0$, οι ρίζες της οποίας είναι $\lambda_1 = 0$ καί $\lambda_2 = \beta/\varepsilon$. Έτσι έχουμε την ακόλουθη λύση,

$$u(x) = C_1 e^{\lambda_1 x} + C_2 e^{\lambda_2 x} = C_1 + C_2 e^{\frac{\beta}{\varepsilon} x},$$

όπου C_1 καί C_2 είναι αυθαίρετες σταθερές. Τέλος εφαρμόζωντας τις συνοριακές συνθήκες έχουμε ότι $C_1 = -\frac{1}{e^{\beta/\varepsilon}-1} = -C_2$ με αποτέλεσμα η λύση τώρα να γράφεται ως εξής,

$$u(x) = \frac{exp(\beta x/\varepsilon) - 1}{exp(\beta/\varepsilon) - 1}.$$

Παρατηρούμε ότι αν το $\beta/\varepsilon << 1$ μπορούμε να αναπτύξουμε τα εκθετικά καί τότε έχουμε ότι,

$$u(x) = (1 + \frac{\beta}{\varepsilon} x + ... - 1)/(1 + \frac{\beta}{\varepsilon} + ... - 1) \approxeq (\beta x/\varepsilon)(\beta/\varepsilon) = x,$$

δηλαδή η λύση είναι κοντά με την λύση του οριακού προβλήματος $-\varepsilon u'' = 0$, η οποία είναι μία ευθεία γραμμή η οποία παρεμβάλει τα συνοριακά στοιχεία.
Από την άλλη αν το $\beta/\varepsilon >> 1$ τα εκθετικά παίρνουν μεγάλες τιμές με αποτέλεσμα,

$$u(x) \approxeq \frac{exp(\beta/\varepsilon x)}{exp(\beta/\varepsilon)} = exp\big[-\frac{\beta}{\varepsilon}(1-x)\big].$$

Από την στιγμή που ο εκθέτης είναι μεγάλος καί αρνητικός η λύση είναι σχεδόν παντού μηδενική εκτός από μία μικρή περιοχή γύρω από το σημείο $x = 1$ όπου ο όρος $1 - x$ γίνεται πολύ μικρός καί η λύση τότε παίρνει την τιμή 1 με εκθετική συμπεριφορά. Το εύρος αυτής της περιοχής είναι της τάξεως ε/β καί είναι αρκετά μικρό.
Οι παραπάνω παρατηρήσεις επισημαίνουν την πολυπλοκότητα του προβλήματος καί την εξάρτηση της λύσης από το ε.

5.1 Διακριτοποίηση καί επίλυση του προβλήματος.

Το συνεχές πρόβλημα (5.1)-(5.2) διακριτοποιείται εφαρμόζοντας την μέθοδο Galerkin με πεπερασμένα στοιχεία στην ειδική περίπτωση που ο χώρος των πεπερασμένων στοιχείων είναι ο X_h^1 ως εξής,

$$\text{να βρεθεί} \qquad u_h \in X_h^1 : a(u_h, \upsilon_h) = 0 \qquad \forall \upsilon_h \in X_h^{1,0}, \tag{5.5}$$

όπου $u_h(0) = 0$ καί $u_h(1) = 1$ είναι οι αντίστοιχες συνοριακές συνθήκες καί $\alpha(\cdot,\cdot)$ είναι η ακόλουθη διγραμμική μορφή,

$$\alpha(u_h, \upsilon_h) = \int_0^1 (\varepsilon u_h' \upsilon_h' + \beta u_h' \upsilon_h) dx. \tag{5.6}$$

Λαμβάνοντας στο πρόβλημα (5.1) ως συνάρτηση δοκιμής υ_h τη γενική συνάρτηση βάσης φ_i έχουμε την εξής μορφή του προβλήματος,

$$\int_0^1 \varepsilon u_h' \varphi_i' dx + \int_0^1 \beta u_h' \varphi_i dx = 0, \qquad i = 1, ..., n-1.$$

Στη συνέχεια θέτοντας $u_h(x) = \sum_{j=0}^n = u_j \varphi_j(x)$ καί υπενθυμίζοντας ότι οι συναρτήσεις βάσης φ_i είναι μη-μηδενικές στο διάστημα $[x_{i-1}, x_{i+1}]$ για $i = 1, ..., n-1$ το παραπάνω πρόβλημα ανάγεται στο ακόλουθο,

$$\varepsilon \left[u_{i-1} \int_{x_{i-1}}^{x_i} \varphi_{i-1}' \varphi_i' dx + u_i \int_{x_{i-1}}^{x_{i+1}} (\varphi_i')^2 dx + u_{i+1} \int_{x_i}^{x_{i+1}} \varphi_i' \varphi_{i+1}' dx \right]$$

$$+\beta \left[u_{i-1} \int_{x_{i-1}}^{x_i} \varphi_{i-1}' \varphi_i dx + u_i \int_{x_{i-1}}^{x_{i+1}} \varphi_i' \varphi_i dx + u_{i+1} \int_{x_i}^{x_{i+1}} \varphi_{i+1}' \varphi_i dx \right] = 0.$$

Θεωρούμε τώρα μία ομοιόμορφη διαμέριση του διαστήματος $[0,1]$ με $x_i = x_{i-1} + h$ για $i = 1, ..., n$ καί $h = \frac{1}{n}$. Στην συνέχεια αντικαθιστούμε τις συναρτήσεις φ_i καί φ_i' από τις σχέσεις (4.21) καί (4.27) αντίστοιχα με αποτέλεσμα να προκύψει ότι,

$$\frac{\varepsilon}{h}\left(-u_{i-1} + 2u_i - u_{i+1}\right) + \frac{1}{2}\beta(u_{i+1} - u_{i-1}) = 0, \qquad i = 1, ..., n-1. \tag{5.7}$$

Αν πολλαπλασιάσουμε την τελευταία σχέση με h/ε καί ορίσουμε τον *τοπικό αριθμό Péclet* να είναι

$$\mathbb{P}e = \frac{|\beta|h}{2\varepsilon},$$

το πρόβλημα θα πάρει τελικά την ακόλουθη μορφή,

$$(\mathbb{P}e - 1)u_{i+1} + 2u_i - (\mathbb{P}e + 1)u_{i-1} = 0, \qquad i = 0, ..., n-1. \tag{5.8}$$

Αυτή είναι μία εξίσωση διαφορών της οποίας η λύση έχει την μορφή $u_i = A_1\rho_1^i + A_2\rho_2^i$, όπου A_1, A_2 είναι δύο κατάλληλες σταθερές καί ρ_1, ρ_2 είναι οι ρίζες της ακόλουθης χαρακτηριστικής εξίσωσης,

$$(\mathbb{P}e - 1)\rho^2 + 2\rho - (\mathbb{P}e + 1) = 0.$$

Αναλυτικότερα,

$$\rho_{1,2} = \frac{-1 \overset{+}{-} \sqrt{1 + \mathbb{P}e^2 - 1}}{\mathbb{P}e - 1} = \begin{cases} \frac{1+\mathbb{P}e}{1-\mathbb{P}e}, \\ \\ 1. \end{cases}$$

Εφαρμόζοντας τις συνοριακές συνθηκές στα σημεία $x = 0$ καί $x = 1$ βρίσκουμε ότι,

$$A_1 = 1/\left(1 - \left(\frac{1+\mathbb{P}e}{1-\mathbb{P}e}\right)^n\right) \qquad \text{καί} \qquad A_2 = -A_1,$$

καί η λύση τελικά του προβλήματος (5.8) είναι,

$$u_i = \left(1 - \left(\frac{1+\mathbb{P}e}{1-\mathbb{P}e}\right)^i\right) / \left(1 - \left(\frac{1+\mathbb{P}e}{1-\mathbb{P}e}\right)^n\right), \qquad i = 0, ..., n.$$

Παρατήρηση 5.1 *Παρατηρούμε ότι αν* $\mathbb{P}e > 1$ *εμφανίζεται στον αριθμητή της παραπάνω λύσης μία δύναμη με αρνητική βάση η οποία έχει ως αποτέλεσμα να έχουμε μία "ταλαντούμενη" λύση. Για να αποφύγουμε τέτοιου είδους λύσεις επιλέγουμε το μέγεθος του βήματος* h *να είναι μικρό με έναν τέτοιο τρόπο ώστε* $\mathbb{P}e < 1$. *Η προσέγγιση αυτή πολλές φορές δεν είναι πρακτική, για παράδειγμα αν το* β*=1 καί* $\varepsilon = 5 \cdot 10^{-5}$ *θα πρέπει το βήμα* h *να είναι μικρότερο του* 10^{-4} $(h < 10^{-4})$, *πράγμα που σημαίνει ότι θα πρέπει να υποδιαιρέσουμε το διάστημα* $[0,1]$ *σε* 10000 *υποδιαστήματα, μία στρατηγική η οποία παρουσιάζει προβλήματα όταν έχουμε να αντιμετωπίσουμε πολυδιάστατα προβλήματα.*

5.2 Σύγκριση των μεθόδων των Π.Σ καί των Π.Δ.

Θα εξετάσουμε τώρα τη συμπεριφορά της μεθόδου των πεπερασμένων διαφορών όταν εφαρμόζεται στην επίλυση προβλημάτων μεταφοράς-διάχυσης σε σχέση με την μέθοδο των πεπερασμένων στοιχείων. Υποθέτουμε πάλι ότι έχουμε το πρόβλημα (5.1) − (5.2) καί για να εξασφαλίσουμε ότι το τοπικό σφάλμα διακριτοποίησης είναι δεύτερης τάξης προσεγγίζουμε τις παραγώγους $u'(x_i)$ καί $u''(x_i)$,$i = 1, ..., n-1$ στο κεντρικό σημείο παρεμβολής x_i (βλ.κεφ 3) καί έτσι προκύπτει το ακόλουθο πρόβλημα πεπερασμένων διαφορών,

$$-\varepsilon\frac{u_{i+1}-2u_i+u_{i-1}}{h^2}+\beta\frac{u_{i+1}-u_{i-1}}{2h}=0, \qquad i=1,...,n-1, \tag{5.9}$$

$$u_0=0, \qquad u_n=1. \tag{5.10}$$

Παρατηρούμε τώρα ότι αν πολλαπλασιάσουμε την παραπάνω σχέση με h θα μας δώσει ακριβώς την ίδια σχέση (5.7) η οποία είχε προκύψει από την εφαρμογή της μέθοδο Galerkin με πεπερασμένα στοιχεία.
Η σχέση αυτή μεταξύ των δύο μεθόδων μπορεί να επιφέρει μία λύση στο πρόβλημα της "ταλαντούμενης " λύσης η οποία εμφανίζεται κατά την προσεγγιστική επίλυση του προβλήματος (5.7) όταν ο τοπικός αριθμός Peclet είναι μεγαλύτερος από 1. Η σημαντικότερη παρατήρηση στο σημείο αυτό είναι ότι η αστάθεια που παρουσιάζεται στη λύση εφαρμόζοντας τη μέθοδο των πεπερασμένων διαφορών προέρχεται από το γεγονός ότι η διακριτοποιήση του προβλήματος γίνεται στο κεντρικό σημείο παρεμβολής. Ένας πιθανός τρόπος αντιμετώπισης αυτού του προβλήματος είναι να προσεγγίσουμε την πρώτη παράγωγο είτε χρησιμοποιώντας τις εμπρός διαφορές είτε τις πίσω διαφορές ανάλογα με το πρόσημο τις σταθεράς β. Αναλυτικότερα,

$$u' \approxeq \begin{cases} \frac{u_i-u_{i-1}}{h} & \text{όταν} \quad \beta>0 \quad \textit{(πίσω διαφορές)}, \\ \\ \frac{u_{i+1}-u_i}{h} & \text{όταν} \quad \beta<0 \quad \textit{(εμπρός διαφορές)}. \end{cases}$$

Στο συγκεκριμένο πρόβλημα μεταφοράς-διάχυσης όπου έχουμε θεωρήσει ότι το β είναι μία θετική σταθερά πολύ μεγαλύτερη της μονάδας χρησιμοποιώντας τις πίσω διαφορές για την προσέγγιση της πρώτης παραγώγου το πρόβλημα (5.9) παίρνει τώρα την ακόλουθη μορφή,

$$-\varepsilon\frac{u_{i+1}-2u_i+u_{i-1}}{h^2}+\beta\frac{u_i-u_{i-1}}{h}=0, \qquad i=1,...,n-1, \tag{5.11}$$

όπου για $\varepsilon = 0$ παίρνει την μορφή $u_i = u_{i-1}$ από την οποία προκύπτει καί η σταθερή λύση του οριακού προβλήματος $\beta u' = 0$. Αυτός ο τρόπος διακριτοποιήσης της πρώτης παραγώγου όπου χρησιμοποιούμε είτε τις εμπρός είτε τις πίσω διαφορές ονομάζεται καί *upwind differencing*. Το τίμημα τώρα που πληρώνουμε προκυμένου να ενισχύσουμε την ευστάθεια της λύσης είναι ότι χάνουμε μία τάξη ακρίβειας καθώς όταν χρησιμοποιούμε *upwind differencing* έχουμε ένα τοπικό σφάλμα διακριτοποίησης της τάξεως $\mathcal{O}(h)$ καί όχι της τάξεως $\mathcal{O}(h^2)$ όπως συνέβαινε όταν η διακριτοποίηση της πρώτης παραγώγου γινόταν στο κεντρικό σημείο παρεμβολής.
Παρατηρούμε επίσης

$$\frac{u_i-u_{i-1}}{h}=\frac{u_{i+1}-u_{i-1}}{2h}-\frac{h}{2}\frac{u_{i+1}-2u_i+u_{i-1}}{h^2},$$

δηλαδή η προσέγγιση της πρώτης παραγώγου χρησιμοποιώντας τις πίσω διαφορές μπορεί να γραφεί ως το άθροισμα της προσέγγισης της πρώτης παραγώγου στο κεντρικό σημείο παρεμβολής καί ενός όρου που είναι ανάλογος της διακριτοποίησης της δεύτερης τάξεως παραγώγου. Συνεπώς το πρόβλημα (5.11) μπορεί τώρα να πάρει την ισοδύναμη μορφή

$$-\varepsilon_h \frac{u_{i+1} - 2u_i + u_{i-1}}{h^2} + \beta \frac{u_{i+1} - u_{i-1}}{2h} = 0 \qquad i = 1, ..., n-1, \tag{5.12}$$

όπου $\varepsilon_h = (1 + \mathbb{P}e)$. Βέβαια αυτό προϋποθέτει την αντικατάσταση της διαφορικής εξίσωσης (5.1) με την αντίστοιχη *διαταραγμένη*,

$$-\varepsilon_h u'' + \beta u' = 0, \tag{5.13}$$

όπου στην συνέχεια χρησιμοποιώντας την μέθοδο των πεπερασμένων διαφορών προσεγγίζουμε στο κεντρικό σημείο παρεμβολής τις παραγώγους u' καί u''. Η ακόλουθη *διαταραχή*

$$-\varepsilon \mathbb{P}e u'' = -\frac{\beta h}{2} u'' \tag{5.14}$$

ονομάζεται *τεχνιτή διάχυση (artificial diffusion ή numerical viscosity).*
Επίσης μπορούμε να παράγουμε μία γενικότερη μορφή του προβλήματος (5.11) χρησιμοποιώνατς την ακόλουθη διάχυση

$$\varepsilon_h = \varepsilon(1 + \varphi(\mathbb{P}e)), \tag{5.15}$$

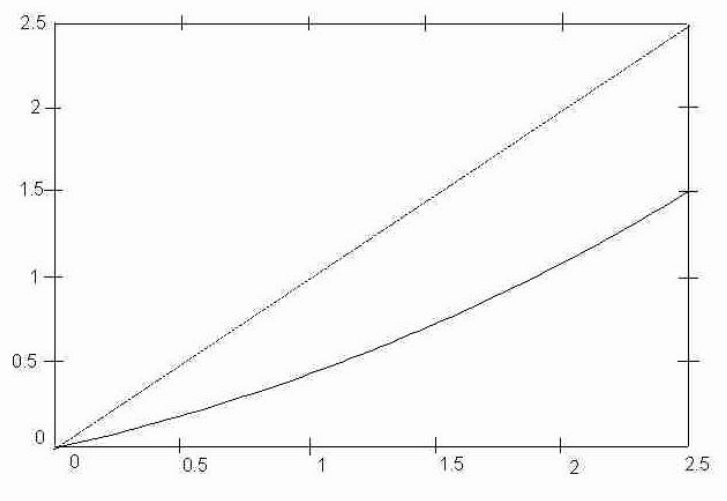
όπου φ είναι μία κατάλληλη συνάρτηση του τοπικού αριθμού Peclet η οποία ικανοποιεί την σχέση,

$$\lim_{t \to 0+} \varphi(t) = 0.$$

Παρατηρούμε ότι όταν $\varphi(t) = 0$ για όλα τα t, τότε παίρνουμε το πρόβημα (5.9) του οποίου η διακριτοποίηση έγινε χρησιμοποιώντας την μέθοδο των πεπερασμένων διαφορών στο κεντρικό σημείο παρεμβολής, ενώ όταν $\varphi(t) = t$ προκύπτει το πρόβλημα (5.11) ή ισοδύναμα το (5.12) του οποίου η διαριτοποίηση έγινε προσεγγίζοντας την πρώτη παράγωγο χρησιμοποιώντας τις πίσω διαφορές (*upwind differencing*). Επίσης μπορούμε να θεωρήσουμε καί άλλες μορφές της συνάρτησης φ. Για παράδειγμα μπορούμε να θεωρήσουμε την ακόλουθη συνάρτηση,

$$\varphi(t) = t - 1 + B(2t), \tag{5.16}$$

όπου B(t) είναι η αντίστροφη συνάρτηση Bernoulli η οποία ορίζεται ως $B(t) = t/(e^t - 1)$ για $t \neq 0$ καί όταν $t = 0$ τότε $B(0) = 1$.
Η μέθοδος αυτή είναι γνωστή και ως *Scharfetter-Gummel method (SG).*



Σχήμα 5.1 Οι συναρτήσεις φ^{UP}(διακεκομμένη γραμμή) καί φ^{SG} (ευθεία γραμμή) .

Παρατήρηση 5.2 *Συμβολίζοντας με φ^C, φ^{UP} καί φ^{SG} τις τρεις προηγούμενες συναρτήσεις, δηλαδή $\varphi^C = 0, \varphi^{UP}(t) = t$ καί $\varphi^{SG}(t) = t - 1 + B(2t)$, παρατηρούμε από το διάγραμμα ότι καθώς $\mathbb{P}e \to +\infty$ $\varphi^{SG} \approx \varphi^{UP}$ ενώ $\varphi^{SG} = \mathcal{O}(h^2)$ καί $\varphi^{UP} = \mathcal{O}(h)$ όταν $\mathbb{P}e \to 0^+$. Έτσι η μέθοδος SG έχει ακρίβεια δεύτερης τάξης καί γι'αυτό τον λόγο θεωρείται καί ως η βέλτιστη μέθοδος. Έπισης μπορεί να αποδειχθεί ότι αν η συνάρτηση f είναι κατά τμήματα σταθερή σε όλο το πλέγμα της διαμέρισης, τότε η SG μέθοδος δίνει μία αριθμητική λύση u_h^{SG} η οποία είναι ακριβής, δηλαδή $u_h^{SG}(x_i) = u(x_i)$ σε κάθε κόμβο x_i.*

Ο καινούριος τώρα τοπικός αριθμός Peclet ο οποίος σχετίζεται με τις σχέσεις (5.12) καί (5.15) ορίζεται ως εξής,

$$\mathbb{P}e^* = \frac{|\beta|h}{2\varepsilon_h} = \frac{\mathbb{P}e}{(1 + \varphi(\mathbb{P}e))}.$$

Καί στα δύο προβλήματα έχουμε $\mathbb{P}e^* < 1$ για οποιοδήποτε h. Αυτό μας εξασφαλίζει ότι η αριθμητική λύση u_h ικανοποιεί το Θεώρημα 3.3 (βλ.Κεφ.3. Παρ.3.3. Ανάλυση της σύγκλισης της μεθόδου των πεπερασμένων διαφορών.)

5.3 Σταθεροποίηση της μεθόδου των πεπερασμένων στοιχείων.

Στην ενότητα αυτή θα επεκτείνουμε την χρήση της *τεχνιτής διάχυσης*, την οποία παρουσιάσαμε στην προηγούμενη ενότητα για την μέθοδο των πεπερασμένων διαφορών, στην μέθοδο Galerkin με πεπερασμένα στοιχεία αυθαίρετου βαθμού $k \geq 1$. Για τον σκοπό αυτό θεωρούμε το πρόβλημα μεταφοράς-διάχυσης (5.1) − (5.2) όπου στη συνέχεια αντικαθηστούμε την σταθερά διάχυσης ε με την σχέση (5.15).

Αυτό έχει ως αποτέλεσμα το διαριτοποιημένο πρόβλημα (5.5), το οποίο προέκυψε από το αντίστοιχο συνεχές εφαρμόζωντας την μέθοδο Galerkin με πεπερασμένα στοιχεία, να πάρει τώρα την ακόλουθη μορφή,

$$\text{να βρεθεί} \qquad \overset{0}{u}_h \in X_h^{k,0} = \{\upsilon_h \in X_h^k : \upsilon_h(0) = \upsilon_h(1) = 0\} \qquad \text{έτσι ώστε}$$

$$\alpha_h(\overset{0}{u}_h, \upsilon_h) = -\int_0^1 \beta \upsilon_h dx \qquad \forall \upsilon_h \in X_h^{k,0}, \tag{5.17}$$

όπου

$$\alpha_h(u, \upsilon) = \alpha(u, \upsilon) + b(u, \upsilon),$$

και ο όρος $b(u, \upsilon)$ ο οποίος δίνεται από την ακόλουθη σχέση

$$b(u, \upsilon) = \varepsilon\varphi(\mathbb{P}e) \int_0^1 u'\upsilon' dx,$$

λέγεται και όρος *σταθεροποιήσης (stabilization term).*
Από την στιγμή που $\alpha_h(\upsilon, \upsilon) = \varepsilon_h|\upsilon|_1^2$ για κάθε $\upsilon \in H_0^1(0,1)$ και $\varepsilon_h/\varepsilon = (1 + \varphi(\mathbb{P}e)) \geq 1$, το τροποποιημένο πρόβλημα (5.17) ικανοποιεί περισσότερες ιδιότητες μονοτονίας σε σχέση με το αντίστοιχο μη σταθεροποιήσιμο πρόβλημα Galerkin (5.7).

Για να αποδείξουμε τώρα την σύγκλιση αρκεί να δείξουμε ότι η λύση του προβλήματος (5.17) $\overset{0}{u}_h$ συγλίνει στο $\overset{0}{u}$ καθώς το $h \to 0$, όπου $\overset{0}{u}(x) = u(x) - x$. Αυτό γίνεται στο ακόλουθο θεώρημα. Πρίν διατυπώσουμε καί αποδείξουμε το βασικό θεώρημα θα διατυπώσουμε ένα θεώρημα καί έναν ορισμό τα αποία καί θα χρησιμοποιήσουμε στην απόδειξη του βασικού θεωρήματος.

Θεώρημα 5.3.1 *Υποθέτουμε ότι $0 \leq m \leq k+1$ με $k \geq 1$ καί $u^{(m)} \in L^2(\alpha, b)$ όταν $0 \leq m \leq k+1$. Τότε υπάρχει μία θετική σταθερά C, η οποία εξαρτάται από το h, τέτοια ώστε*

$$||(u - \Pi_h^k u)^m||_{L^2(\alpha,b)} \leq C h^{k+1-m} ||u^{(k+1)}||_{L^2(\alpha,b)}.$$

Συγκεκριμένα όταν το $k = 1$,καί το $m = 0$ ή το $m = 1$ τότε παρατηρούμε αντίστοιχα,

$$||u - \Pi_h^1||_{L^2(\alpha,b)} \leq C_1 h^2 ||u''||_{L^2(\alpha,b)},$$

$$||(u - \Pi_h^1)'||_{L^2(\alpha,b)} \leq C_2 h ||u''||_{L^2(\alpha,b)}, \tag{5.18}$$

για κατάλληλες θετικές σταθερές C_1 καί C_2.

Ορισμός 5.3.1 *Ανισότητα Young:*

$$ab \leq \varepsilon\alpha^2 + \frac{1}{4\varepsilon} b^2, \qquad \forall \alpha, b \in \mathbb{R}, \qquad \forall \varepsilon > 0. \tag{5.19}$$

Θεώρημα 5.3.2 *Αν $k = 1$ τότε*

$$|\overset{0}{u} - \overset{0}{u}_h|_{H^1(0,1)} \leq ChG(\overset{0}{u}), \tag{5.20}$$

όπου $C \geq 0$ είναι μία κατάλληλη σταθερά η οποία εξαρτάται από το h καί $\overset{0}{u}$ καί

$$G(\overset{0}{u}) = \begin{cases} |\overset{0}{u}|_{H^1(0,1)} + |\overset{0}{u}|_{H^2(0,1)} & \text{για τις πίσω διαφορές,} \\ \\ |\overset{0}{u}|_{H^2(0,1)} & \text{για την μέθοδο SG.} \end{cases}$$

Επιπλέον αν $k = 2$ η μέθοδος SG μας δίνει την ακόλουθη βελτιωμένη εκτίμηση σφάλματος

$$|\overset{0}{u} - \overset{0}{u}_h|_{H^1(0,1)} \leq Ch^2(|\overset{0}{u}|_{H^1(0,1)} + |\overset{0}{u}|_{H^3(0,1)}). \tag{5.21}$$

Απόδειξη.

Από το πρόβλήμα (5.1) − (5.2) παρατηρούμε ότι,

$$\alpha(\overset{0}{u}, \upsilon_h) = -\int_0^1 \beta \upsilon_h dx, \qquad \forall \upsilon_h \in X_h^{k,0}.$$

Συγρίνοντας τώρα την παραπάνω σχέση με την (5.17) καθώς τα δεύτερα μέλη τους είναι ίσα παίρνουμε την ακόλουθη σχέση,

$$\alpha_h(\overset{0}{u} - \overset{0}{u}_h, \upsilon_h) = b(\overset{0}{u}, \upsilon_h), \qquad \forall \upsilon_h \in X_h^{k,0}. \tag{5.22}$$

Αναλτικότερα η σχέση (5.22) προκύπτει ως εξής,

$$\alpha_h(\overset{0}{u}_h, \upsilon_h) = \alpha(\overset{0}{u}, \upsilon_h) \Rightarrow \alpha(\overset{0}{u}_h, \upsilon_h) + b(\overset{0}{u}_h, \upsilon_h) = \alpha(\overset{0}{u}, \upsilon_h)$$

$$\Rightarrow \alpha(\overset{0}{u} - \overset{0}{u}_h, \upsilon_h) = b(\overset{0}{u}_h, \upsilon_h).$$

Συμβολίζουμε με $E_h = \overset{0}{u} - \overset{0}{u}_h$ το σφάλμα διακριτοποίησης καί υπενθυμίζουμε ότι έχουμε εφοδιάσει τον χώρο $H_0^1(0,1)$ με την νόρμα (4.8). Έτσι λοιπόν έχουμε,

$$\varepsilon_h |E_h|^2_{H_0^1(0,1)} = \alpha_h(E_h, E_h) = \alpha_h(E_h, \overset{0}{u} - \Pi_h^k \overset{0}{u}) + \alpha_h(E_h, \Pi_h^k \overset{0}{u} - \overset{0}{u}_h)$$

$$= \alpha_h(E_h, \overset{0}{u} - \Pi_h^k \overset{0}{u}) + b(\overset{0}{u}, \Pi_h^k \overset{0}{u} - \overset{0}{u}_h),$$

όπου εφαρμόσαμε την σχέση (5.22) με $\upsilon_h = \Pi_h^k \overset{0}{u} - \overset{0}{u}_h$. Στην συνέχεια αναπτύσουμε το δεύτερο μέλος της παραπάνω σχέσης καί εφαρμόζουμε την ανισότητα Cauchy-Schwarz καθώς καί την ανισότητα Poincare με αποτέλεσμα να έχουμε,

$$\varepsilon_h |E_h|^2_{H^1(0,1)} = \int_0^1 (\varepsilon_h E_h'(\overset{0}{u} - \Pi_h^k \overset{0}{u})' + \beta(E_h'(\overset{0}{u} - \Pi_h^k \overset{0}{u}))dx + \varepsilon\varphi(\mathbb{P}e) \int_0^1 (\overset{0}{u})'(\Pi_h^k \overset{0}{u} - \overset{0}{u}_h)' dx$$

$$\leq \varepsilon_h|E_h|_{H^1(0,1)}|\overset{0}{u}-\Pi_h^k\overset{0}{u}|_{H^1(0,1)}+|\beta|C_p|E_h|_{H^1(0,1)}|\overset{0}{u}-\Pi_h^k\overset{0}{u}|_{H^1(0,1)}+\varepsilon\varphi(\mathbb{P}e)\int_0^1(\overset{0}{u})'(\Pi_h^k\overset{0}{u}-\overset{0}{u}_h)'dx$$

$$=(\varepsilon_h+|\beta|C_p)|E_h|_{H^1(0,1)}|\overset{0}{u}-\Pi_h^k\overset{0}{u}|_{H^1(0,1)}+\varepsilon\varphi(\mathbb{P}e)\int_0^1(\overset{0}{u})'(\Pi_h^k\overset{0}{u}-\overset{0}{u}_h)'dx$$

$$=M_h|E_h|_{H^1(0,1)}|\overset{0}{u}-\Pi_h^k\overset{0}{u}|_{H^1(0,1)}+\varepsilon\varphi(\mathbb{P}e)\int_0^1(\overset{0}{u})'(\Pi_h^k\overset{0}{u}-\overset{0}{u}_h)'dx, \tag{5.23}$$

όπου $M_h=\varepsilon_h+|\beta|C_p$ είναι η σταθερά συνέχειας της διγραμμικής μορφής $\alpha(\cdot,\cdot)$ καί C_p είναι η σταθερά της ανισότης Poincare.

Παρατηρούμε ότι στην περίπτωση που το $k=1$ καί $\varphi=\varphi^{SG}$ η δεύτερη ποσότητα μέσα στο ολοκλήρωμα είναι ίση με μηδέν καθώς σύμφωνα με την παρατήρηση (5.2) έχουμε $\overset{0}{u}_h=\Pi_h^1\overset{0}{u}$. Έτσι η σχέση (5.23) παίρνει τώρα την ακόλουθη μορφή,

$$|E_h|_{H^1(0,1)}\leq\left(1+\frac{|\beta|C_p}{\varepsilon_h}\right)|\overset{0}{u}-\Pi_h^1\overset{0}{u}|_{H^1(0,1)}.$$

Παρατηρώντας τώρα ότι $\varepsilon_h>\varepsilon$ καί χρησιμοποιώντας τις σχέσεις (5.3) καί (5.18) έχουμε τελικά το ακόλουθο φράγμα του σφάλματος,

$$|E_h|_{H^1(0,1)}\leq C(1+2\mathbb{P}e_{gl}C_ph|\overset{0}{u}|_{H^2(0,1)}).$$

Στην γενική περίπτωση η ανισότητα σφάλματος (5.21) μπορεί να υπολογιστεί χρησιμοποιώνατς την ανισότητα Cauchy-Schwarz καθώς καί την τριγωνική ανισότητα. Αναλυτικότερα,

$$\int_0^1(\overset{0}{u})'(\Pi_h^k\overset{0}{u}_h)'dx\leq|\overset{0}{u}|_{H^1(0,1)}||\Pi_h^k\overset{0}{u}-\overset{0}{u}_h|_{H^1(0,1)}\leq|\overset{0}{u}|_{H^1(0,1)}|\left(|\Pi_h^k\overset{0}{u}-\overset{0}{u}|_{H^1(0,1)}+|\overset{0}{u}-\overset{0}{u}_h|_{H^1(0,1)}\right)$$

$$=|\overset{0}{u}|_{H^1(0,1)}|\left(|\Pi_h^k\overset{0}{u}-\overset{0}{u}|_{H^1(0,1)}+|E_h|_{H^1(0,1)}\right).$$

Αντικαθιστούμε στην σχέση (5.23) καί έχουμε,

$$\varepsilon_h|E_h|^2_{H^1(0,1)}\leq|E_h|_{H^1(0,1)}\left(M_h|\overset{0}{u}-\Pi_h^k\overset{0}{u}|_{H^1(0,1)}+\varepsilon\varphi(\mathbb{P}e)|\overset{0}{u}|_{H^1(0,1)}\right)+\varepsilon\varphi(\mathbb{P}e)|\overset{0}{u}|_{H^1(0,1)}|\overset{0}{u}-\Pi_h^k\overset{0}{u}|_{H^1(0,1)}.$$

Στη συνέχεια εφαρμόζοντας πάλι την σχέση (5.18) προκύπτει η ακόλουθη σχέση,

$$\varepsilon_h|E_h|^2_{H^1(0,1)}\leq|E_h|_{H^1(0,1)}\left(M_hCh^k|\overset{0}{u}|_{H^{k+1}(0,1)}+\varepsilon\varphi(\mathbb{P}e)|\overset{0}{u}|_{H^1(0,1)}\right)+C\varepsilon\varphi(\mathbb{P}e)|\overset{0}{u}|_{H^1(0,1)}h^k|\overset{0}{u}|_{H^{k+1}(0,1)}.$$

Εφαρμόζοντας τώρα την ανισότητα Young έχουμε,

$$\varepsilon_h|E_h|^2_{H^1(0,1)}\leq\frac{\varepsilon_h|E_h|^2_{H^1(0,1)}}{2}+\frac{3}{4\varepsilon_h}\left[(M_hCh^k|\overset{0}{u}|_{H^{k+1}(0,1)})^2+(\varepsilon\varphi(\mathbb{P}e)|\overset{0}{u}|_{H^1(0,1)})^2\right],$$

από την οποία προκύπτει ότι,

$$|E_h|^2_{H^1(0,1)} \leq \frac{3}{2}\left(\frac{M_h}{\varepsilon_h}\right)^2 C^2 h^{2k} |^0_u|^2_{H^{k+1}(0,1)} + \frac{3}{2}\left(\frac{\varepsilon}{\varepsilon_h}\right)^2 + \varphi(\mathbb{P}e)^2 |^0_u|^2_{H^1(0,1)} +$$

$$+\frac{2\varepsilon}{\varepsilon_h}\varphi(\mathbb{P}e)|^0_u|_{H^1(0,1)} C h^k |^0_u|_{H^{k+1}(0,1)}.$$

Χρησιμοποιώντας πάλι το γεγονός ότι $\varepsilon_h > \varepsilon$ καί τον ορισμό (5.3) έχουμε ότι $(M_h/\varepsilon_h) \leq (1 + 2C_p \mathbb{P}e_{gl})$ καί τότε η παραπάνω σχέση παίρνει την ακόλουθη μορφή,

$$|E_h|^2_{H^1(0,1)} \leq \frac{3}{2} C^2 (1 + 2C_p \mathbb{P}e_{gl})^2 h^{2k} |^0_u|^2_{H^{k+1}(0,1)}$$

$$+2\varphi(\mathbb{P}e) C h^k |^0_u|_{H^1(0,1)} |^0_u|_{H^{k+1}(0,1)} + \frac{3}{2}\varphi(\mathbb{P}e)^2 |^0_u|^2_{H^1(0,1)},$$

το οποίο μπορεί να γραφεί καί ως εξής,

$$|E_h|^2_{H^1(0,1)} \leq \mathcal{M}\left[h^{2k} |^0_u|_{H^{k+1}(0,1)} + \varphi(\mathbb{P}e) h^k |^0_u|_{H^1(0,1)} + \varphi(\mathbb{P}e)^2 |^0_u|^2_{H^1(0,1)}\right] \tag{5.24}$$

όπου $\mathcal{M}$ είναι μία κατάλληλη θετική σταθερά.

Αν έχουμε τώρα $\varphi^{UP} = \mathcal{C}_\varepsilon h$, όπου $\mathcal{C}_\varepsilon = \beta/\varepsilon$, τότε παρατηρούμε ότι

$$|E_h|^2_{H^1(0,1)} \leq C h^2 \left[h^{2k-2} |^0_u|^2_{H^{k+1}(0,1)} + h^{k-1} |^0_u|_{H^1(0,1)} |^0_u|_{H^{k+1}(0,1)} + |^0_u|^2_{H^1(0,1)}\right],$$

η οποία μας δείχνει ότι χρησιμοποιώντας την μέθοδο των πεπερασμένων στοιχείων στην ειδική περίπτωση που το $k = 1$ όπως επίσης καί την τεχνιτή διάχυση στην περίπτωση των πίσω διαφορών προκύπτει η γραμμική εκτίμηση σύγκλισης (5.20).

Στην περίπτωση τώρα που έχουμε $\varphi = \varphi^{SG}$ καί υποθέτοντας ότι το h είναι αρκετά μικρό έτσι ώστε $\varphi^{SG} \leq K h^2$, για μία κατάλληλη θετική σταθερά K, τότε παρατηρούμε ότι,

$$|E_h|^2_{H^1(0,1)} \leq C h^4 \left[h^{2(k-2)} |^0_u|^2_{H^{k+1}(0,1)} + h^{k-2} |^0_u|_{H^1(0,1)} |^0_u|_{H^{k+1}(0,1)} + |^0_u|^2_{H^1(0,1)}\right],$$

η οποία μας δείχνει ότι στην περίπτωση που το $k = 2$ σε συνδιασμό με το ότι χρησιμοποιήσαμε την τεχνιτή διάχυση προκύπτει η δεύτερης τάξεως εκτίμηση της σύγκλισης (5.21). □

5.4 Περιγραφή του προβλήματος στις δύο διαστάσεις.

Μέχρι στιγμής έχουμε παρουσιάσει τις βασικές ιδέες αριθμητικής επίλυσης προβλήματων ελλειπτικού τύπου τα οποία ορίζονται πάνω σε μονοδιάστατες περιοχές χρησιμοποιώνατς τις μεθόδους των Πεπερασμένων Διαφορών καί την μέθοδο Galerkin με Πεπερασμένα Στοιχεία. Τι γίνεται όμως όταν το πρόβλημα μας ορίζεται στις δύο διαστάσεις;

Το αρχικό πρόβλημα $(2.1)-(2.2)$ το οποίο παρουσιάσαμε στο Κεφάλαιο 2 μετατρέπεται τώρα στο γνωστό *πρόβλημα Poisson* με ομογενής συνοριακές συνθήκες Dirichlet,

$$-\Delta u = f \qquad \text{στον χώρο } \Omega \text{ με} \tag{5.25}$$

$$u = 0 \qquad \text{πάνω στο} \qquad \partial\Omega, \tag{5.26}$$

όπου $\Delta u = \partial^2 u/\partial x^2 + \partial^2 u/\partial y^2$ είναι ο τελεστής Laplace καί Ω είναι η δισδιάστατη φραγμένη περιοχή όπου το σύνορο της είναι το $\partial\Omega$. Αν υποθέσουμε ότι ο χώρος Ω είναι το μοναδιαίο τεράγωνο $\Omega = (0,1)^2$, η προσέγγιση του προβλήματος $(5.25)-(5.26)$ χρησιμοποιώντας την μέθοδο των πεπερασμένων διαφορών σε αντιστοιχεία με την προσέγγιση του μονοδιάστατου προβλήματος (3.6) είναι,

$$L_h u_h(x_{i,j}) = f(x_{i,j}) \qquad \text{όπου} \qquad i,j = 1,...,N-1, \tag{5.27}$$

$$u_h(x_{i,j}) = 0 \qquad \text{αν} \qquad i = 0 \qquad \text{ή} \qquad N, \qquad \text{καί} \qquad j = 0 \qquad \text{ή} \qquad N, \tag{5.28}$$

όπου $x_{i,j} = (ih, jh)(h = 1/N)$ είναι τα διακριτά σημεία καί u_h η αντίστοιχη διακριτή συνάρτηση.
Τέλος με τον τελεστή L_h συμβολίζουμε την προσέγγιση του τελεστη Laplace στον οποίο αν αντικαταστήσουμε τις παραγώγους δεύτερης τάξης με τις αντίστοιχες διακριτοποιημένες, οι οποίες προκύπτουν χρησιμοποιώντας την μέθοδο των πεπερασμένων διαφορών στο κεντρικό σημείο παρεμβολής παίρνει την ακόλουθη μορφή,

$$L_h u_h(x_{i,j}) = \frac{1}{h^2}(4u_{i,j} - u_{i+1,j} - u_{i-1,j} - u_{i,j+1} - u_{i,j-1}), \tag{5.29}$$

όπου $u_{i,j} = u_h(x_{i,j})$. Ο πίνακας A_{fd} ο οποίος σχετίζεται με το παραπάνω πρόβλημα έχει $(N-1)^2$ γραμμές ,είναι πενταδιαγώνιος με την i-οστή γραμμή του να δίνεται από την ακόλουθη σχέση,

$$(\alpha_{fd})_{i,j} = \frac{1}{h^2}\begin{cases} 4 & \text{αν} \quad j = 1, \\ -1 & \text{αν} \quad j = i-N-1, i-1, i+1, i+N+1, \\ 0 & \text{διαφορετικά.} \end{cases} \tag{5.30}$$

Επίσης ο πίνακας A_{fd} είναι συμμετρικός καί θετικά ορισμένος. Όσο αναφορά τώρα το σφάλμα της συνεπείας όσο καί της διακριτοποίησης $||u-u_h||_{h,\infty}$ που σχετίζονται με το πρόβλημα (5.29) είναι δεύτερης τάξης όπως ακριβώς καί στην περίπτωση που το πρόβλημα μας ήταν μονοδιάστατο.

Η επέκταση τώρα της μεθόδου Galerkin στις δυο διαστάσεις προκύπτει άμεσα από το μονοδίαστατο πρόβλημα (4.5) αφού πρώτα φυσικά προσαρμόσουμε κατάλληλα τον χώρο των

συναρτήσεων δοκιμής V_h καί τη διγραμμική μορφή $\alpha(\cdot,\cdot)$. Έτσι χρησιμοποιώντας την μέθοδο των πεπερασμένων στοιχείων ορίζουμε στο δισδιάστατο πρόβλημα τον χώρο των συναρτήσεων δοκιμής V_h ως εξής,

$$V_h = \{\upsilon_h \in C^0(\bar{\Omega}) : \upsilon_h|_T \in \mathbb{P}_k(T) \qquad \forall T \in \mathcal{T}_h, \qquad \upsilon_h|_{\partial\Omega} = 0\}, \tag{5.31}$$

Παρατηρούμε ότι ο παραπάνω ορισμός του χώρου των συναρτήσεων δοκιμής δεν είναι τίποτα άλλο από την επέκταση του ορισμού (4.17) ο οποίος ισχύει στα προβλήματα μίας διάστασης. Όσο αναφορά τώρα την διγραμμική μορφή $\alpha(\cdot,\cdot)$ αυτή προκύπτει εφαρμόζωντας τους ίδιους μαθηματικούς υπολογισμούς που χρησιμοποιήσαμε καί στο Κεφάλαιο 4 για την αντίστοιχη εύρεση της διγραμμικής μορφής στην περίπτωση του μονοδιάστατου προβλήματος. Έτσι έχουμε,

$$\alpha(u_h, \upsilon_h) = \int_\Omega \nabla u_h \cdot \nabla \upsilon_h dxdy,$$

όπου χρησιμοποιήσαμε τον ακόλουθο τύπο του Green ο οποίος αποτελεί καί μία γενίκευση της ολοκλήρωσης κατά μέρη

$$\int_\Omega -\Delta u \upsilon dxdy = \int_\Omega \nabla u \cdot \nabla \upsilon dxdy - \int_{\partial\Omega} \nabla u \cdot \mathbf{n} \upsilon d\gamma, \tag{5.32}$$

με υ,u να είναι επαρκώς ομαλές συναρτήσεις καί όπου $\mathbf{n}$ να είναι το κανονικό μοναδιαίο διάνυσμα πάνω στο σύνορο $\partial\Omega$.
Η εκτίμηση τώρα του σφάλματος, το οποίο προκύπτει λόγω της προσέγγισης του προβλήματος $(5.25)-(5.26)$ με τη μεθόδο των πεπερασμένων στοιχείων, προκύπτει όπως καί στην μονοδιάστατη περίπτωση μέσω του συνδιασμού του λήμματος Cea's καί της εκτίμησης του σφάλματος παρεμβολής. Έτσι στην περίπτωση των δύο διαστάσεων ισχύει το ακόλουθο λήμμα.

Λήμμα 5.4.1 *Έστω ότι η $u \in H_0^1(\Omega)$ είναι η ακριβής ασθενής λύση του προβλήματος (5.25)-(5.26) καί η $u_h \in V_h$ η αντίστοιχη προσεγγιστική η οποία προκύπτει χρησιμοποιώντας κατά τμήματα συνεχή πολυώνυμα βαθμού $k \geq 1$. Υποθέτουμε επίσης ότι $u \in H^s(\Omega)$ για $s \geq 2$. Τότε ισχύει η ακόλουθη εκτίμηση σφάλματος*

$$\|u - u_h\|_{H_0^1(\Omega)} \leq \frac{M}{\alpha_0} C h^l \|u\|_{H^{l+1}(\Omega)}, \tag{5.33}$$

όπου $l = min(k, s-1)$. Κάτω υπό τις ίδιες υποθέσεις μπορεί επίσης να αποδειχθεί ότι,

$$\|u - u_h\|_{L^2(\Omega)} \leq C h^{l+1} \|u\|_{H^{l+1}(\Omega)}. \tag{5.34}$$

Σημειώνουμε ότι για οποιοδήποτε ακέραιο $s \geq 0$ ο χώρος Sobolev $H^s(\Omega)$ τον οποίο χρησιμοποιήσαμε παραπάνω ορίζεται ως ο χώρος των συναρτήσεων με τις πρώτες s μερικές παραγώγους να ανήκουν στον $L^2(\Omega)$. Επιπλέον ο $H_0^1(\Omega)$ είναι ο χώρος των συναρτήσεων που ανήκουν στον $H^1(\Omega)$ καί παράλληλα η $u = 0$ πάνω στο σύνορο $\partial\Omega$. Έτσι μία συνάρτηση η οποία ανήκει στον χώρο $H_0^1(\Omega)$ δεν σημαίνει απαραίτητα ότι είναι συνεχής παντού.

Συνεχίζοντας να ακολουθούμε τις ίδιες διαδικασίες με το Κεφάλαιο 4, μπορούμε να γράψουμε την λύση που προκύπτει εφαρμόζωντας την μέθοδο των πεπερασμένων στοιχείων ως εξής,

$$u_h(x,y) = \sum_{j=1}^{N} u_j \varphi_j(x,y),$$

όπου $\{\varphi_j\}_{j=1}^{N}$ είναι οι συναρτήσεις βάσης του χώρου V_h. Στην περίπτωση που το $k=1$ έχουμε τις περίφημες *συναρτήσεις στέγες*(βλ.Κεφ.4). Η μέθοδος λοιπόν Galerkin με πεπερασμένα στοιχεία οδηγεί στην λύση του γραμμικού συστήματος $A_{fe}\mathbf{u} = \mathbf{f}$, όπου $(\alpha_{fe})_{i,j} = \alpha(\varphi_j, \varphi_i)$.

Ακριβώς όπως συμβαίνει καί στην περίπτωση της μίας διάστασης, ο πίνακας A_{fe} είναι συμμετρικός, θετικά ορισμέμος καί γενικά αραιός. Τέλος ο δείκτης κατάστασης συνεχίζει να είναι της ίδιας τάξης $\mathcal{O}(h^{-2})$ καί το παραπάνω σύστημα καλείται "*κακής κατάστασης*".

Κεφάλαιο 6

Αριθμητικά παραδείγματα καί άναλυση των αποτελεσμάτων.

Πρόβλημα 1. Επίλυση του προβλήματος Laplace.

Ορίζουμε αρχικά τον χώρο C πάνω στον οποίο θα επιλύσουμε το πρόβλημα μας,

$$C = \{f(x, y) : x = cost, y = sint, 0 \leq t \leq 2\pi\}.$$

Υποθέτουμε τώρα ότι έχουμε τον ακόλουθο πρόβλημα Laplace,

$$-\Delta u = f \qquad \text{στο } \Omega, \tag{6.1}$$

όπου $f = -1$ καί Ω=$\{(x, y) : x^2 + y^2 < 1\}$, με $\partial\Omega = C$, καί με συνοριακή συνθήκη,

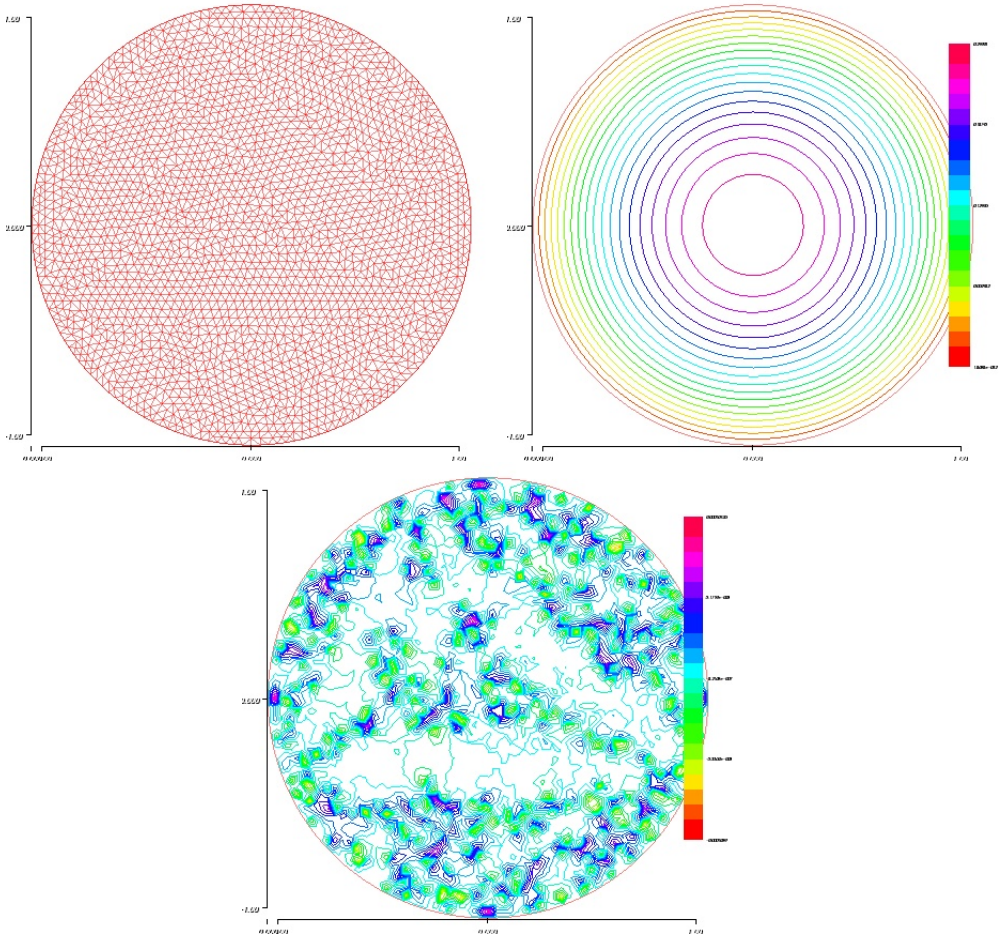
$$u = 0 \qquad \text{στο} \qquad \partial\Omega.$$

Η ακριβής λύση του παραπάνω προβλήματος είναι $u_e = \frac{1-x^2-y^2}{4}$

Με χρήση της γλώσσας προγραμματισμού Freefem (βλ.[14])επιλύσαμε αριθμητικά το παραπάνω πρόβλημα καί υπολογίσαμε τα σφάλματα L_2 καί H_1, μειώνοντας συνεχώς το βήμα της μεθόδου των Πεπερασμένων Στοιχείων, έτσι ώστε κάποια στιγμή το$h \to 0$. Παρακάτω δίνεται ο αντίστοιχος κώδικας.

```
real pi=4*atan(1);

border a(t=0,2*pi){ x = cos(t); y = sin(t);label=1;};


mesh disk = buildmesh(a(160));

plot(disk);

fespace femp1(disk,P1);

femp1 u,v;


problem laplace(u,v) =

   int2d(disk)( dx(u)*dx(v) + dy(u)*dy(v) )     //  bilinear form

 + int2d(disk)( -1*v )                    //  linear form

 + on(1,u=0) ;                          // boundary condition


laplace;

femp1 err=u-(1-x^2-y^2)/4;


plot (u,value=true,wait=true);

plot(err,value=true,wait=true);


cout << "error L2=" << sqrt(int2d(disk)( (u-(1-x^2-y^2)/4) ^2) )<< endl;

cout << "error H10=" << sqrt(   int2d(disk)((dx(u)+x/2)^2)

                    + int2d(disk)((dy(u)+y/2)^2))<< endl;
```

Πρόγραμμα 1. Αριθμητική επίλυση του προβλήματος *Laplace* καί υπολογισμός των αντίστοιχων σφαλμάτων.



Διαγράμματα_ 1.Διαγράμματα που προκύπτουν από το Πρόγραμμα 1.

n	L^2-error	H^1-error
10	0.0483756	0.176984
20	0.0130963	0.0948276
40	0.00325794	0.0463907
80	0.000818174	0.0234118
160	0.000201246	0.0115037
320	5.08133e- 005	0.00581992
640	1.27084e- 005	0.00291345
1280	3.16416e- 006	0.00145069

Πίνακας 1. Αποτελέσματα του Προγράμματος 1.

Πρόβλημα 2.Επίλυση του προβλήματος Μεταφοράς-Διάχυσης.
Υποθέτουμε ότι έχουμε το ακόλουθο πρόβλημα μεταφοράς-διάχυσης,

$$-\varepsilon\Delta u + v\nabla u = f, \qquad \text{στο } \Omega \tag{6.2}$$

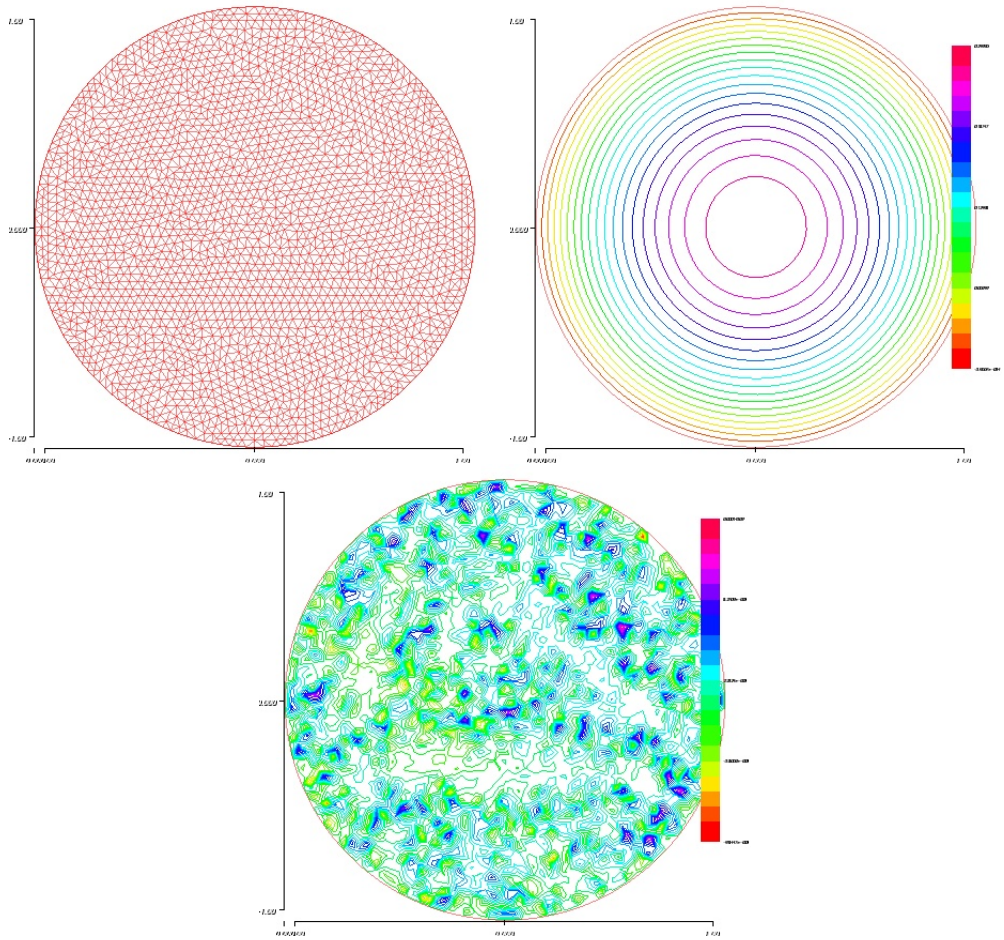
το οποίο ακριβώς όπως καί στο πρόβλημα 1 το επιλύουμε πάνω στον δίσκο C , με συνορικακή συνθήκη $u = 0$ στο $\partial\Omega$ όπου $\partial\Omega = C$ καί την ίδια ακριβή λύση.

Αν θεωρήσουμε τώρα ότι ο συντελεστής διάχυσης είναι ίσος με $\varepsilon = 0.01$ και $v = [1, -1]$,όπου με v συμβολίζουμε το πεδίο ταχυτήτων, τότε έχουμε :

```
real e=0.01;

real pi=4*atan(1);

border a(t=0,2*pi){ x = cos(t); y = sin(t);label=1;};


mesh disk = buildmesh(a(160));

plot(disk);

fespace femp1(disk,P1);


femp1 u,v;

func f=e-(x/2)+(y/2);

problem laplace(u,v) =

   int2d(disk)( e*dx(u)*dx(v) +e*dy(u)*dy(v)+dx(u)*v-dy(u)*v)    //  bilinear form

 - int2d(disk)( f*v )                    //  linear form

 + on(1,u=0) ;                          // boundary condition


laplace;

femp1 err=u-(1-x^2-y^2)/4;


plot (u,value=true,wait=true);

plot(err,value=true,wait=true);


cout << "error L2=" << sqrt(int2d(disk)( (u-(1-x^2-y^2)/4) ^2) )<< endl;

cout << "error H10=" << sqrt(   int2d(disk)((dx(u)+x/2)^2)

                    + int2d(disk)((dy(u)+y/2)^2))<< endl;
```

Πρόγραμμα 2. Αριθμητική επίλυση του προβλήματος μεταφοράς-διάχυσης καί υπολογισμός

των αντίστοιχων σφαλμάτων.



Διαγράμματα_ 2. Διαγράμματα που προκύπτουν από το Πρόγραμμα 2.

n	L^2 -error	H^1 -error
10	0.0508908	0.208642
20	0.0153761	0.107401
40	0.00393641	0.0517348
80	0.000898231	0.0245366
160	0.000187602	0.0117098
320	5.31537e- 005	0.00585641
640	1.26742e- 005	0.00291918
1280	3.25198e- 006	0.00145153

Πίνακας 2. Αποτελέσματα του Προγράμματος 2.

Παρατήρηση 6.1 *Από τα αποτελέσματα των δύο προγραμμάτων βλέπουμε ότι κάθώς το* $h \to 0$ *τα σφάλματα* H^1 *καί* L^2 *τείνουν καί αυτά με τη σειρά τους στο μηδέν. Τα παραπάνω αποτελέσματα έρχονται σε απόλυτη συμφωνία με τις αντίστοιχες εκτιμήσεις σφαλμάτων* (5.33) *καί* (5.34).

Παρατήρηση 6.2 *Στο πρόγραμμα 2. αν θεωρήσουμε ότι ο συντελεστής διάχυσης* ε *είναι ακόμη πιο μικρός για παράδειγμα* $\varepsilon = 10^{-7}$, *τότε για* $n = 10$ *έχουμε* $L^2 = 1097,31$ *καί* $H^1 = 5823,83$ *καί για* $n = 160$ *έχουμε αντίστοιχα* $L^2 = 0,09119922$ *καί* $H^1 = 0,78434$. *Παρατηρούμε ότι σε σχέση με τα αποτελέσματα του προγράμματος 2. ότι αν μειώσουμε τον συντελεστή διάχυσης τα σφάλματα αυξάνονται καί φτάνουμε σε ένα σημείο όπου για πολύ μεγάλο* n *το πρόγραμμα μας σταματάει να τρέχει.*

Βιβλιογραφία

[1] Alfio Qourteran,Riccardo Succo,Fausto Saleti.,Numerical Mathematics,Springer, 2007

[2] Brezis H., Συναρτησιακή Ανάλυση Θεωρία και Εφαρμογές, Πανεπιστημιακές Εκδόσεις Ε.Μ.Π., 1997

[3] W.E.Boyse-R.C.Diprima,Στοιχειώδεις Διαφορικές Εξισώσεις κ Προβλήματα Συνοριακών Τιμών.,Πανεπιστημιακές Εκδόσεις Ε.Μ.Π.,1999

[4] Γ.Δάσιος-Κ.Κυριάκη,Μερικές Διαφορικές Εξισώσεις,Αθήνα 1994

[5] Γ.Ν.Παντελίδη-Δ.Χ.Κραββαρίτη-Ν.Σ.Χατζησάββα,Συνήθεις Διαφορικές Εξισώσεις.,Εκδόσεις Ζήττη.,Αθήνα 1990

[6] Ν.Σταυρακάκης,Συνήθεις Διαφορικές Εξισώσεις., Εκδόσεις Παπασωτηρίου.,1997

[7] Γ.Παπαγεωργίου.,Αριθμητική Ανάλυση των διαφορικών εξισώσεων.,Εκδόσεις Συμεών.,Αθήνα 2005

[8] Γ.Σ.Παπαγεωργίου-Χ.Γ.Τσίτουρος.,Αριθμητική Ανάλυση με Εφαρμογές σε Matlab καί Mathematica.,Εκδόσεις Συμεών.,Αθήνα 2006

[9] Α.Μπακόπουλος-Ι.Χρυσοβέργης.,Εισαγωγή στην Αριθμητική Ανάλυση.,Εκδόσεις Συμεών.,Αθήνα 1994

[10] Μ.Μισυρλής.,Αριθμητική Ανάλυση.,Αθήνα 2009

[11] Logan J.D., Εφαρμοσμένα Μαθηματικά, Πανεπιστημιακές Εκδόσεις Κρήτης, 1999

[12] Maz'ya V., Sobolev Spaces in Mathematics II, Applications in Analysis and Partial Differential Equations, 2009

[13] Showalter R.E., Hilbert space methods for Partial Differential Equations, 1994

[14] Hecht, A partial guide to Freefemlab

Ευρετήριο