HERMITE COLLOCATION ΜΕΘΟΔΟΙ ΓΙΑ ΣΤΑΘΕΡΕΣ ΚΑΤΑΣΤΑΣΕΙΣ ΠΡΟΒΛΗΜΑΤΩΝ ΜΕΤΑΦΟΡΑΣ-ΔΙΑΧΥΣΗΣ.

Μιχαήλ Π. Τσιβουράκης
Πολυτεχνείο Κρήτης
Γενικό Τμήμα

Χανιά - Οκτώβριος 2004

Ευχαριστίες

Καταρχάς θα ήθελα να ευχαριστήσω τον σύμβουλο καθηγητή μου, Καθηγητή Ιωάννη Σαριδάκη ο οποίος μου παρείχε την άρτια επιστημονική καθοδήγηση για την ολοκλήρωση της παρούσας διατριβής.

Τους καθηγητές μου στο Γενικό Τμήμα και ιδιαίτερα την Ε.Παπαδοπούλου, Αναπληρώτρια Καθηγήτρια και τον Α.Δελή Λέκτορα, για την συμμετοχή τους στην τριμελή επιτροπη.

Τον Λέκτορα Ε. Μαθιουδάκη για την παροχή πολύτιμων επιστημονικών γνώσεων-συμβουλών.

Τους συναδέλφους μου μεταπτυχιακούς φοιτητές για την ηθική συμπαράσταση που μου παρείχαν.

Τέλος ευχαριστώ τα μέλη της οικογενείας μου για την κάθε είδους βοήθεια κατά την διάρκεια των σπουδών μου.

Πρόλογος

Η παρούσα εργασία είναι δομημένη σε τέσσερα κεφάλαια.

Στο πρώτο κεφάλαιο παρουσιάζεται μια εισαγωγή στην αριθμητική μέθοδο Collocation. Στη συνέχεια γίνεται εφαρμογή Hermite Cubic Orthogonal Collocation σε προβλήματα μορφής μεταφοράς-διάχυσης. Στο τέλος του κεφαλαίου γίνεται αναφορά στα μεγάλα προβλήματα που παρουσιάζονται κατά την αριθμητική επίλυση των παραπάνω προβλημάτων με την Hermite Cubic Orthogonal Collocation.

Στο δεύτερο κεφάλαιο αναπτύσονται μέθοδοι Hermite Cubic Spline Collocation (H.C.S.C.), οι οποίες ενσωματώνουν upwinding χαρακτηριστικά. Αρχικά παρουσιάζεται μια φασματική ανάλυση των πινάκων παραγώγισης της H.C.S.C.. Στη συνέχεια βασιζόμενοι σε αυτήν την ανάλυση κατασκευάζονται σύνολα από collocation points, ώστε η μέθοδος να αποκτήσει upwinding χαρακτηριστικά. Στο τέλος του κεφαλαίου, εφαρμόζεται η Upwind H.C.S.C. για την αρθμητική επίλυση προβλήματων τύπου μεταφοράς-διάχυσης και γίνεται σύγκριση των αποτελεσμάτων με την orthogonal collocation.

Στο τρίτο κεφάλαιο αναζητούμε μια καλλίτερη διαμέριση του διαστήματος $I = [a, b]$, πάνω στο επιλύεται το πρόβλημα, ώστε η αριθμητική μέθοδος να συμπεριφέρεται καλλίτερα. Γίνεται αναφορά σε adaptive h-refinement τεχνικές που προσαρμόζουν κατάλληλα ένα πλέγμα διακριτοποίησης και γίνεται κατασκευή μιας επαναληπτικής h-refinement τεχνικής χρησιμοποιώντας τη μέθοδο της Hermite Cubic Orthogonal Collocation.

Στο τέταρτο κεφάλαιο γίνεται σύντομη περίληψη των αποτελεσμάτων της παρούσας μεταπτυχιακής εργασίας.

Συμβολισμοί

- $\overset{\circ}{\Omega}$: Το εσωτερικό ενός συνόλου Ω.
- $\partial\Omega$: Το σύνορο ενός συνόλου Ω.
- ∇ : Έστω $f(x_1, x_2, \ldots, x_n)$ μια συνάρτηση n μεταβλητών. Τότε με ∇f συμβολίζουμε την κλίση της συνάρτησης, δηλαδή $\nabla f = \Big(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \ldots, \frac{\partial f}{\partial x_n}\Big)$.
- $\triangle$: Ο Λαπλασιανός τελεστής ή απλως Λαπλασιανή. Συμβολίζεται και με ∇^2. Η Λαπλασιανή μιας συνάρτησης $f(x_1, x_2, \ldots, x_n)$ n μεταβλητών είναι $\triangle f = \dfrac{\partial^2 f}{\partial x_1^2} + \dfrac{\partial^2 f}{\partial x_2^2} + \ldots + \dfrac{\partial^2 f}{\partial x_n^2}$.
- $L_2(\Omega)$: Ο γραμμικός χώρος όλων των μετρήσιμων συναρτήσεων $f : \Omega \to K$ για τις οποίες,

$$\int_\Omega |f|^2 d\Omega < \infty.$$

- $H_2(\Omega)$: Με $H_2(\Omega)$ συμβολίζουμε τον χώρο Sobolev δευτέρας τάξεως. Γενικά ένας χώρος $H_m(\Omega)$ Sobolev $m-$οστής τάξεως, είναι ο χώρος των τετραγωνικών ολοκληρώσιμων συναρτήσεων που έχουν m μερικές παραγώγους, όπου είναι αναπαραστάσημες ως τετραγωνικά ολοκληρώσιμες συναρτήσεις,

$$H_m(\Omega) = \Big\{ u \in L_2(\Omega) \big| \frac{\partial^k u}{\partial x^k} \in L_2(\Omega), k = 1, \ldots, m \Big\}$$

εφοδιασμένος με εσωτερικό γινόμενο,

$$< u, v >_{H_m(\Omega)} = \sum_{i=0}^{m} \int_\Omega \frac{\partial^k u}{\partial x^k} \frac{\partial^k v}{\partial x^k} \, d\Omega$$

και νόρμα,

$$||u||_{H_m(\Omega)} = \sqrt{< u, v >_{H_m(\Omega)}}.$$

Περιεχόμενα

Κεφάλαιο 1

Εισαγωγή

1.1 Το πρόβλημα

Είναι γνωστό ότι παραβολικές εξισώσεις της μορφής μεταφοράς-διάχυσης (advection-diffusion) περιγράφουν μαθηματικά μοντέλα πολλών διαφορετικών διαδικασιών που εμφανίζονται σε διάφορες περιοχές της επιστήμης και της μηχανικής. Η γενικότερη τους μορφή έχει ως εξής:

$$\frac{\partial u(\mathbf{x},t)}{\partial t} + Lu(\mathbf{x},t) = f(\mathbf{x},t) \qquad , \text{ με } \mathbf{x} \in \overset{\circ}{\Omega} \text{ και } t > 0$$

$$\begin{cases} Bu(\mathbf{x},t) & = g(\mathbf{x},t) \quad , \text{με x} \in \partial\Omega \text{ και } t > 0 \\ u(\mathbf{x},t) & = u_o(\mathbf{x}) \quad , \text{όταν } t = 0 \end{cases} \tag{1.1}$$

και

$$L = -\epsilon \cdot \triangle u(\mathbf{x},t) + v \cdot \nabla u(\mathbf{x},t)$$

όπου L, B διαφορικοί τελεστές, ϵ ο συντελεστής διάχυσης (diffusion) και v το διάνυσμα ταχύτητας (συντελεστής advection[1]).

Είναι ευρέως γνωστό ότι η αριθμητική επίλυση των παραπάνω εξισώσεων με καθιερωμένες αριθμητικές μεθόδους θεωρείται ανεπαρκής όταν η ποσότητα $\frac{\epsilon}{||v||}$ είναι μικρή σε σχέση με το μέγεθος του χωρικού βήματος διακριτοποίησης της εκάστοτε αριθμητικής μεθόδου επίλυσης,[BRI02,BRI04,HUN93,LEV98].

[1] Ο συντελεστής advection καλείται επίσης και συντελεστής convection.

Για την μελέτη της συμπεριφοράς των αριθμητικών μεθόδων στην λύση των παραπάνω μοντέλων (1.1) είναι καθιερωμένη η χρήση του προβλήματος μοντέλου:

$$\begin{cases} -\epsilon \cdot \triangle u(\mathbf{x}) + v \cdot \nabla u(\mathbf{x}) &= f(\mathbf{x}) \quad , \text{με x} \in \overset{\circ}{\Omega} \\ Bu(\mathbf{x}) &= g(\mathbf{x}) \quad , \text{με x} \in \partial\Omega \end{cases} \tag{1.2}$$

Το παραπάνω πρόβλημα μοντέλο, του οποίου η επίλυση και η ανάδειξη των ιδιομορφιών που παρουσιάζει κατά την επίλυση είναι ο σκοπός της παρούσας δουλειάς, αποτελεί την steady - state κατάσταση των προβλημάτων (1.1).

Στην συνέχεια θα κάνουμε αναφορά σε θεωρητικά στοιχεία της αριθμητικής μεθόδου που θα χρησιμοποιήσουμε για την αριθμητική επίλυση των προβλημάτων (1.2).

1.2 Collocation

Η μέθοδος της collocation (καθώς και άλλες μέθοδοι όπως spectral, finite volume, finite elements, finite difference) μπορεί να παραχθεί με συγκεκριμένη εφαρμογή της μεθόδου των weighted residuals. Η μέθοδος των weighted residuals χρησιμοποιεί expansion functions[2] ως βάση συναρτήσεων για την ανάπτυξη σε μορφή πεπερασμένου αθροίσματος της λύσης μιας διαφορικής εξίσωσης. Στην συνέχεια για να εξασφαλιστεί ότι η προσεγγιστική λύση, η οποία ορίζεται από το πεπερασμένο άθροισμα των expansion functions, θα ικανοποιεί την διαφορική εξίσωση όσο το δυνατόν "καλλίτερα "(δηλαδή όσο το δυνατόν πιο κοντά στην πραγματική λύση), test functions[3] χρησιμοποιούνται για να ελαχιστοποιήσουν το υπόλοιπο, το οποίο εμφανίζεται όταν η προσεγγιστική λύση αντικαταστήσει την πραγματική στην διαφορική εξίσωση. Διαφορετικοί συνδυασμοί των expansion functions και των test functions οδηγούν σε διαφορετικές μεθόδους.

Στην συνέχεια θα παράγουμε την μέθοδο των πεπερασμένων στοιχείων της collocation όπου θα χρησιμοποιήσουμε για την διακριτοποίηση των προβλημάτων μας, ως αποτέλεσμα της μεθόδου των weighted residuals.

[2] Οι expansion functions καλούνται επίσης trial functions.

[3] Οι test functions καλούνται επίσης weighting functions.

1.2.1 Μέθοδος Weighted Residuals

Έστω L γραμμίκος διαφορικός τελεστής δευτέρας τάξεως . Θεωρούμε ένα χωρίο Ω με σύνορο $\Gamma = \partial\Omega$ και υποθέτουμε ότι $f : \Omega \to \mathbb{R}$ είναι μια δεδομένη συνάρτηση. Οπότε ορίζουμε την ακόλουθη διαφορική εξίσωση ως εξής:

$$\begin{cases} Lu - f = 0 & , \text{με } x \in \overset{\circ}{\Omega} \\ \\ u = u_\Gamma & , \text{με } x \in \Gamma = \partial\Omega \end{cases} \tag{1.3}$$

Έστω το σύνολο U των trail συναρτήσεων:

$$U = \left\{ u \;\middle|\; u \in H^2(\Omega) \text{ , } u = u_\Gamma \text{ στο } \Gamma = \partial\Omega \right\} \tag{1.4}$$

τα στοιχεία του οποίου είναι ορισμένα με τρόπο τέτοιο, ώστε να ικανοποιούν την διαφορική εξίσωση (1.3) στο σύνορο $\Gamma = \partial\Omega$.

Ορίζουμε επίσης το σύνολο των W test συναρτήσεων:

$$W = \left\{ w \;\middle|\; w \in L_2(\Omega) \text{ , } w = 0 \text{ στο } \Gamma = \partial\Omega \right\} \tag{1.5}$$

τα στοιχεία του οποίου είναι ορισμένα με τρόπο τέτοιο, ώστε να είναι μηδενικά στο σύνορο $\Gamma = \partial\Omega$.

Χρησιμοποιώντας τα σύνολα U και W, για την αναζήτηση μιας λύσης $u \in U$, ώστε να εξασφαλίζεται ότι η προβολή της συνάρτησης $Lu - f$ πάνω στο σύνολο W των test συναρτήσεων είναι μηδέν, η διαφορική εξίσωση (1.3) μπορεί να αποκτήσει ακόλουθη μορφή:

$$\begin{cases} \textit{Βρές } u \in U \textit{ τέτοια ώστε:} \\ \\ < Lu - f, w >_w = 0 & , \forall w \in W \end{cases} \tag{1.6}$$

Στον χώρο $L_2(\Omega)$ το παραπάνω εσωτερικό γινόμενο μπορεί να εκφραστεί με εξής μορφή:

$$\begin{cases} \textit{Βρές } u \in U \textit{ τέτοια ώστε :} \\ \\ \int_\Omega (Lu - f) w d\Omega = 0 & , \forall w \in W \end{cases} \tag{1.7}$$

Το επόμενο βήμα για το διακριτοποιημένο σχήμα είναι η επιλογή ενός πεπερασμένης διάστασης υποχώρου $\widehat{U} \subseteq U$ με βάση $\left\{\phi_i\right\}_{i=1}^{n}$, ο οποίος περιλαμβάνει την προσεγγιστική λύση $\hat{u}$. Οι trial συναρτήσεις $\left\{\phi_i\right\}_{i=1}^{n}$ χρησιμοποιούνται ως βάση για το ανάπτυγμα της προσεγγιστικής λύσης. Οπότε η προσεγγιστική λύση $\hat{u} \in \widehat{U}$ γράφεται:

$$\hat{u} = \sum_{i=1}^{n} c_i \phi_i \tag{1.8}$$

Από την επιλογή του χώρου $\widehat{U}$ εξαρτάται αν θα χρησιμοποιηθεί ο ακριβής διαφορικός τελεστής L ή ένας κατάλληλος διαφορικός τελεστής $\widehat{L}$. Στην περίπτωση της finite element collocation ο χώρος $\widehat{U}$ επιτρέπει την χρήση του ακριβούς διαφορικού τελεστή L, αφού αποτελείται από αναλυτικές συναρτήσεις (πολυωνυμικές συναρτήσεις). Όταν η προσεγγιστική λύση $\hat{u}$ αντικατασταθεί στην διαφορική εξίσωση (1.3), τότε αυτή δεν θα είναι ταυτοτικά μηδέν αλλά:

$$L\hat{u} - f = \hat{r} \tag{1.9}$$

όπου $\hat{r}$ καλείται υπόλοιπο ή συνάρτηση υπολοίπου (ή σφάλματος) κατά την διακριτοποιήση της διαφορικής εξίσωσης. Οι συντελεστές του αναπτύγματος της λύσης c_i από την (1.8) είναι άγνωστοι και μπορούν να αποκτηθούν απαιτώντας η προβολή του υπόλοιπου στον χώρο W να ισούται με μηδέν. Δηλαδή,

$$< \hat{r}, w >_w = 0 \quad , \quad \forall w \in W \tag{1.10}$$

ή ως πρός την L_2- νόρμα,

$$\int_{\Omega} \hat{r} w d\Omega = 0 \quad , \quad \forall w \in W. \tag{1.11}$$

Έστω υποχώρος $\widehat{W} \subseteq W$ πεπερασμένης διάστασης με βάση $\left\{\psi_i\right\}_{i=1}^{n}$. Οπότε η σχέση (1.10) γίνεται:

$$\begin{cases} \textit{Βρες } \hat{u} \in \widehat{U} \textit{ τέτοια ώστε:} \\ \\ < L\hat{u} - f, \hat{w} >_{\hat{w}} = 0 \qquad\qquad , \forall \hat{w} \in \widehat{W} \end{cases} \tag{1.12}$$

ή ισοδύναμα χρησιμοποιώντας το εσωτερικό γινόμενο του χώρου L_2 έχουμε:

$$\begin{cases} \textit{Βρες } \hat{u} \in \widehat{U} \textit{ τέτοια ώστε:} \\ \\ \int_\Omega (L\hat{u} - f)\hat{w} d\Omega = 0 \qquad , \forall \hat{w} \in \widehat{W} \end{cases} \tag{1.13}$$

όμως, όπως ήδη εχουμε αναφέρει, το παραπάνω εσωτερικό γινόμενο εξασφαλίζει ότι η προβόλη της συνάρτησης $L\hat{u} - f$ θα είναι μηδέν πάνω στόν χώρο $\widehat{W}$, άρα και σε κάθε στοίχειο της βάσης του $\{\psi_i\}_{i=1}^n$. Συνεπώς, ισοδύναμα έχουμε,

$$\int_\Omega (L\hat{u} - f)\psi_j \, d\Omega = 0 \qquad \text{με} \quad j = 1, \ldots, n \Leftrightarrow$$

$$\int_\Omega (L\sum_{i=1}^{n} c_i\phi_i - f)\psi_j \, d\Omega = 0 \quad \text{με} \quad j = 1, \ldots, n$$

και αφού L γραμμικός,

$$\sum_{i=1}^{n} c_i \int_\Omega (L\phi_i - f)\psi_j \, d\Omega = 0 \quad \text{με} \quad j = 1, \ldots, n.$$

Οπότε είμαστε έτοιμοι να ορίσουμε την διακριτή μορφή της μεθόδου των weighted residuals, ως εξής:

$$\begin{cases} \textit{Βρες } c_i \textit{ με } i = 1, \ldots, n \textit{ τέτοια ώστε:} \\ \\ \sum_{i=1}^{n} c_i \int_\Omega (L\phi_i) \cdot \psi_j d\Omega = \int_\Omega f \cdot \psi_j d\Omega \quad \text{, με } j = 1, \ldots, n \end{cases} \tag{1.14}$$

Συνεπώς χρησιμοποιώντας συμβολισμό πινάκων αποκτούμε το γραμμικό σύστημα:

$$\mathbf{L} \cdot \vec{c} = \vec{\mathbf{f}} \tag{1.15}$$

όπου τα στοιχεία L_{ij} του πίνακα $\mathbf{L}$ καθώς και τα στοιχεία του διανύσματος στήλη $\vec{\mathbf{f}}$ δίνονται από τις σχέσεις:

$$\mathbf{L}_{ij} = \int_\Omega (L\phi_i) \cdot \psi_j d\Omega \tag{1.16}$$

$$\mathbf{f}_i = \int_\Omega f \cdot \psi_j d\Omega \tag{1.17}$$

και όπου

$$\vec{c} = [c_0, c_1, \ldots, c_n]^T \tag{1.18}$$

$$\vec{\mathbf{f}} = [f_0, f_1, \ldots, f_n]^T. \tag{1.19}$$

Ακολούθως οι άγνωστοι συντελεστές c_i της προσεγγιστικής λύσης μπορούν να αποκτηθούν από την λύση του γραμμικού συστήματος (1.15). Γνωρίζοντας τους συντελεστές c_i η προσεγγιστική λύση μπορεί να υπολογιστεί από την σχέση (1.8). Διαφορετικές επιλογές των test functions οδηγούν σε διαφορετικές μεθόδους διακριτοποίησης. Στην συνέχεια θα παράγουμε την μέθοδο της Collocation.

1.2.2 Collocation Method ως Weighted Residual Method

Για την παραγωγή της μεθόδου της collocation, ένας αριθμός n από σημεία ορίζονται στο Ω, τα οποία ονομάζονται collocation points $\prod_{COL} = \left\{x_j\right\}_{j=1}^{n}$ και χρησιμοποιώντας τα ορίζονται οι test functions ψ_j, επιλέγοντάς τις να είναι Dirac δέλτα συναρτήσεις σύμφωνα με τα παρακάτω:

$$\psi_j = \delta(x - x_j) \tag{1.20}$$

με

$$\delta(x - x_j) = \begin{cases} \infty & , \text{με } x = x_j \\ 0 & , \text{με } x \neq x_j \end{cases}$$

και με την ιδιότητα

$$\begin{cases} \int_\Omega \delta(x - x_j) d\Omega = 1 & , \text{με } j = 1, \ldots, n \\ \int_\Omega \delta(x - x_j) g(x) d\Omega = g(x_j) & , \text{με } j = 1, \ldots, n \end{cases} \tag{1.21}$$

Για οποιαδήποτε συνεχή συνάρτηση g στο Ω.

Αντικαθιστώντας την (1.20) στην εξίσωση (1.14) και με βάση τις παραπάνω ιδιότητες έχουμε την ακόλουθη διατύπωση της μεθόδου της collocation:

$$\begin{cases} \textit{Βρες } \hat{u} \in \widehat{U} \textit{ τέτοια ώστε:} \\ \\ L\hat{u}\Big|_{x=x_j} = f(x_j) \qquad\qquad \text{, με } j = 1, \ldots, n \end{cases} \tag{1.22}$$

Παρατηρούμε ότι το υπόλοιπο εξαναγκάζεται σε μηδενισμό στο σύνολο το οποίο έχει ως στοιχεία του τα collocation points $\prod_{COL} = \left\{x_j\right\}_{j=1}^{n}$.

1.2.3 Finite Element Collocation Method

Τα παραπάνω γενικεύονται στην περίπτωση της μεθόδου των πεπερασμένων στοιχείων collocation, όπου το χωρίο Ω χωρίζεται σε έναν αριθμό N_{el} από υποχωρία Ω_i (στοιχεία "elements"του χώρου Ω) με την ιδιότητα,

$$\begin{cases} \bigcup_{i=1}^{N_{el}} \Omega_i &= \Omega \\ \\ \bigcap_{i=1}^{N_{el}} \overset{\circ}{\Omega}_i &= \emptyset \end{cases} \tag{1.23}$$

και θεωρώντας ως χώρο των προσεγγιστικών λύσεων $\widehat{U}$ να είναι:

$$\widehat{U} = U_{Fe} = \left\{ u \in U \;\middle|\; u|_{\Omega_i} \in P_N(\Omega_i) \right\} \tag{1.24}$$

όπου το $P_N(\Omega_i)$ δηλώνει τον χώρο των πολυωνυμικών συναρτήσεων στο υποχωρίο "element"Ω_i βαθμού $< N$. Στην περίπτωση επίλυσης μιας m-οστής τάξεως διαφορικής εξίσωσης με m - συνοριακές συνθήκες ορίζουμε ένα σύνολο απο k = N-m interior collocation points $\prod_{COL} = \left\{x_{ij}^c\right\}_{i,j=1}^{N_{el},k}$ στο εσωτερικό του κάθε υποχωρίου Ω_i και, έχουμε:

$$\begin{aligned} \int_\Omega \delta(x - x_{ij}^c) Lu(x) d\Omega = \int_{\Omega_1} \delta(x - x_{ij}^c) Lu(x) d\Omega + \int_{\Omega_2} \delta(x - x_{ij}^c) Lu(x) d\Omega + \\ + \cdots + \int_{\Omega_i} \delta(x - x_{ij}^c) Lu(x) d\Omega + \cdots + \int_{\Omega_{N_{el}}} \delta(x - x_{ij}^c) Lu(x) d\Omega. \end{aligned} \tag{1.25}$$

Οπότε,

$$Lu|_{\Omega_1}(x_{ij}^c) + Lu|_{\Omega_2}(x_{ij}^c) + \cdots + Lu|_{\Omega_i}(x_{ij}^c) + \cdots + Lu|_{\Omega_{N_{el}}}(x_{ij}^c) = Lu|_{\Omega_i}(x_{ij}^c)$$

Διότι για το x_{ij}^c collocation point έχουμε

$$x_{ij}^c \in \Omega_i \Rightarrow Lu|_{\Omega_k}(x_{ij}^c) = 0 \quad \text{για} \quad k \neq i.$$

Συνεπώς, είμαστε έτοιμοι να ορίσουμε την διατύπωση της Finite Element Collocation Method:

$$Lu|_{\Omega_i}(x_{ij}^c) = f\,(x_{ij}^c)$$

με $j = 1, \cdots, N-m$ και $i = 1, \cdots, N_{el}$. Στην περίπτωση της μιας διάστασης, με αυτόν τον τρόπο κατασκευάσαμε ένα γραμμικό σύστημα $N_{el}(N-m)$ εξισώσεων με $N_{el}(N-m)+m$ αγνώστους και τούτο διότι σε κάθε element Ω_i έχουμε N αγνώστους και m συνθήκες συνέχειας της λύσης στους $N_{el}-1$ εσωτερικούς κόμβους. Οπότε συνολικά θα έχουμε $N_{el}(N-m)+m$ αγνώστους. Οι υπόλοιπες m εξισώσεις προκύπτουν από τις m συνοριακές συνθήκες. Σε περισσότερες διαστάσεις, ο χώρος μπορεί να θεωρηθεί ότι προκύπτει ως τανυστικό γινόμενο χώρων μιας διάστασης, και αναλόγως προκύπτει το γραμμικό σύστημα (1.15).

Παράδειγμα της Finite Element Collocation είναι η Hermite Cubic Finite Element Collocation Method, της οποίας η βάση για τον χώρο των trail functions αποτελείται από τμηματικά κυβικά πολυώνυμα hermite ώστε η προσεγγιστική λύση να ορίζεται τμηματικά σε κάθε element Ω_i και θα την αναφέρουμε σε επόμενη ενότητα.

1.2.4 Collocation points - Orthogonal Finite Element Collocation

Διαφορετικές επιλογές του συνόλου των collocation points οδηγούν σε διαφορετικές ταχύτητες σύγκλισης της μεθόδου . Οι De Boor και Swartz το 1973 [BOO73] έδειξαν ότι η προσεγγιστική λύση για μια m-οστής τάξεως διαφορικής εξίσωσης με m - συνοριακές συνθήκες μπορεί να βρεθεί χρησιμοποιώντας μια τμηματική πολυωνυμική προσέγγιση βαθμού μικρότερου του m+k, έχοντας m-1 συνεχείς παραγώγους και χρησιμοποιώντας k-gauss points του κάθε υποχωρίου "element"Ω_j. Το συνολικό σφάλμα της προσεγγιστικής λύσης θα είναι

της τάξεως $O(h^{m+k})$ εάν η πραγματική λύση έχει m+2k συνεχείς παραγώγους και η διαφορική εξίσωση παράγει ικανοποιητικά ομαλές λύσεις. Στην περίπτωση της χρησιμοποίησης των gauss points του κάθε element Ω_i τότε παράγουμε την γνωστή Orthogonal Collocation. Η μέθοδος της Orthogonal Collocation βασίζεται στην ίδια ιδέα όπως και η Gaussian Quadrature, μεθόδου προσέγγισης ολοκληρώματος με ένα πεπερασμένο άθροισμα της μορφής:

$$\int_a^b f(x) = \sum_{i=1}^{m} w_i f(x_i)$$

όπου x_i τα σημεία ολοκλήρωσης (quadrature points) και w_i τα σχετιζόμενα βάρη (weights).
Έχοντας την ιδιότητα ότι πολυώνυμα βαθμού μικρότερου του 2m να ολοκληρώνονται ακριβώς. Στην περίπτωση της μεθόδου των πεπερασμένων στοιχείων Hermite Cubic Orthogonal Collocation την οποία θα αναφέρουμε στην επόμενη ενότητα, τα σχετιζόμενα βάρη $w_i = 1$ για $i = 1, 2$ και τα quadrature points x_i είναι οι ρίζες του $2^{ου}$ βαθμού πολυώνυμου του Legendre.

1.2.5 Hermite Cubic Finite Element Collocation

Στην μέθοδο των πεπερασμένων στοιχείων Hermite Cubic Collocation η βάση για το σύνολο των trial functions $\widehat{U}$ αποτελείται από τμηματικά κυβικά πολυώνυμα hermite. Αρχικά ορίζουμε τα κυβικά πολυώνυμα hermite στο διάστημα $[-1, 1]$:

$$\Phi(x) = \begin{cases} \Phi_L(x), & x \in [-1, 0] \\ \Phi_R(x), & x \in [0, 1] \\ 0, & \text{διαφορετικά} \end{cases}$$

με

$$\Phi_L(x) = (1 + x)^2(1 - 2x) \tag{1.26}$$

$$\Phi_R(x) = (1 - x)^2(1 + 2x) \tag{1.27}$$

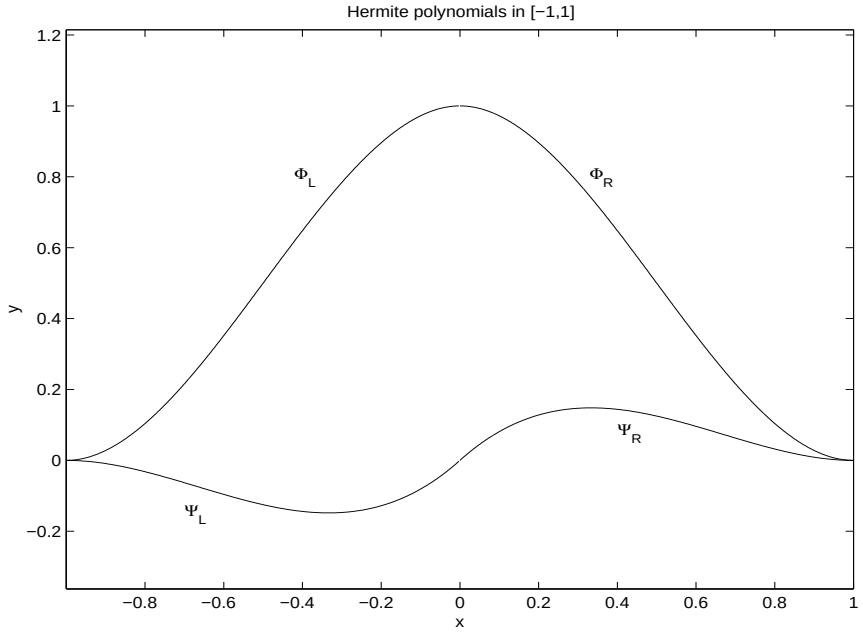
και

$$\Psi(x) = \begin{cases} \Psi_L(x), & x \in [-1, 0] \\ \Psi_R(x), & x \in [0, 1] \\ 0, & \text{διαφορετικά} \end{cases}$$

με

$$\Psi_L(x) = (1+x)^2 x \tag{1.28}$$

$$\Psi_R(x) = (1-x)^2 x \tag{1.29}$$



Σχήμα 1.1: Τα κυβικά πολυώνυμα hermite ορισμένα στο διάστημα $[-1, 1]$.

Θεωρούμε το διάστημα $I = [a, b]$ και έστω μια διαμερίση του $\left\{I_i\right\}_{i=1}^{N_{el}}$ με $I = \bigcup_{i=1}^{N_{el}} I_i$. Ορίζοντας ένα σύνολο από $N_{el}+1$ σημεία "κόμβους, " $\prod_{N_{el}} = \left\{x_i\right\}_{i=1}^{N_{el}+1}$ στο $[a, b]$, ως εξής:

$$a = x_1 < x_2 < \ldots < x_{N_{el}} < x_{N_{el}+1} = b$$

αποκτούμε μια διαμέριση του $[a, b]$ ώστε,

$$I_i = [x_i, x_{i+1}]$$

και

$$h_i = x_{i+1} - x_i$$

με $i = 1, 2, \ldots, N_{el}$.

Χρησιμοποιώντας απλούς γραμμικούς μετασχηματισμούς μπορούμε να ορίσουμε τα κυβικά πολυώνυμα hemite σε κάθε υποδιάστημα $I_i = [x_i, x_{i+1}]\ , \quad i = 1, \ldots, N_{el}$. Οπότε έχουμε:

$$\begin{cases} \xi_{1,i}(s) &= (1-s)^2(1+2s) \\ \xi_{2,i}(s) &= s(1-s)^2 \\ \xi_{3,i}(s) &= \xi_{1,i}(1-s) \\ \xi_{4,i}(s) &= -\xi_{2,i}(1-s) \end{cases} \tag{1.30}$$

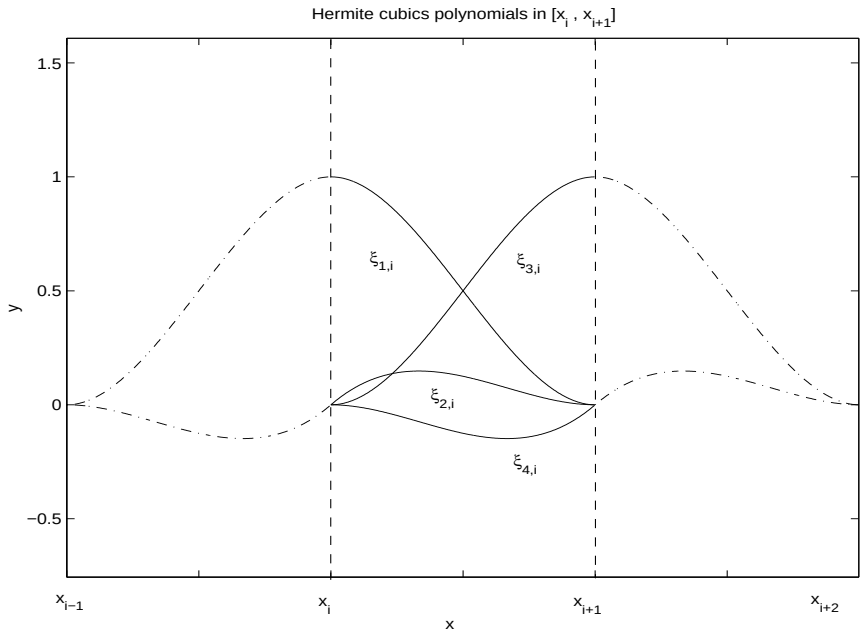
όπου

$$s = \frac{x - x_i}{h_i}$$

Παρατηρούμε ότι ισχύουν οι εξής ιδιότητες:

$$\begin{cases} \xi_{1,i}(x_j) &= \delta_{ij} \\ \dfrac{\partial \xi_{1,i}(x_j)}{\partial x} &= 0 \\ \xi_{2,i}(x_j) &= 0 \\ \dfrac{\partial \xi_{2,i}(x_j)}{\partial x} &= \dfrac{1}{h_i}\delta_{ij} \end{cases} \tag{1.31}$$

όταν $j = 1, 2, \ldots, N_{el}, N_{el} + 1$ και



Σχήμα 1.2: Τα τμηματικά κυβικά πολυώνυμα hermite ορισμένα στο element $I_i = [x_i, x_{i+1}]$.

$$\begin{cases} \xi_{3,i}(x_{j+1}) & = & \delta_{ij} \\ \dfrac{\partial \xi_{3,i}(x_{j+1})}{\partial x} & = & 0 \\ \xi_{4,i}(x_{j+1}) & = & 0 \\ \dfrac{\partial \xi_{4,i}(x_{j+1})}{\partial x} & = & \dfrac{1}{h_i}\delta_{ij} \end{cases} \tag{1.32}$$

όταν $j = 0, 1, \ldots, N_{el} - 1, N_{el}$.
Όπου δ_{ij} το δέλτα του Kronecker,

$$\delta_{ij} = \begin{cases} 1 & , i = j \\ 0 & , i \neq j. \end{cases}$$

Έχοντας κατασκευάσει την βάση για το σύνολο των trial συναρτήσεων, ορίζουμε την τμηματική πολυωνύμικη προσέγγιση για την λύση στο υποδιάστημα $I_i = [x_i, x_{i+1}]$ η οποία θα δίνεται από την σχέση:

$$\hat{u}(x) = c_{1,i}\xi_{1,i}(s) + c_{2,i}\xi_{2,i}(s) + c_{3,i}\xi_{3,i}(s) + c_{4,i}\xi_{4,i}(s) \tag{1.33}$$

Από τις σχέσεις (1.31), (1.32) παρατηρούμε ότι ισχύουν τα εξής:

$$\begin{cases} c_{1,i} &= \hat{u}(x_i) \\ c_{2,i} &= h_i\hat{u'}(x_i) \\ c_{3,i} &= \hat{u}(x_{i+1}) \\ c_{4,i} &= h_i\hat{u'}(x_{i+1}) \end{cases} \tag{1.34}$$

Οπότε θεωρώντας $u_i = \hat{u}(x_i)$ και $u'_i = \hat{u'}(x_i)$ η τμηματική πολυωνύμικη προσέγγιση παίρνει την ακόλουθη μορφή:

$$\hat{u}(x) = u_i\xi_{1,i}(s) + h_iu'_i\xi_{2,i}(s) + u_{i+1}\xi_{3,i}(s) + h_iu'_{i+1}\xi_{4,i}(s) \tag{1.35}$$

στο υποδιάστημα $I_i = [x_i, x_{i+1}]$.
Θεωρούμε το σύνολο

$$\Omega_c = \left\{ (x,y) \;\middle|\; x < y \text{ τέτοια ώστε } x, y \in (0,1) \right\} \in \mathbb{R}^2 \tag{1.36}$$

Για ένα δεδομένο στοιχείο (σ_1, σ_2) του συνόλου Ω_c ορίζουμε το σύνολο των collocation points για την διαμέριση $\prod_{N_{el}}$ ως $\prod_{COL} = \left\{x^c_{ij}\right\}_{i,j=1}^{N_{el},2}$, όπου:

$$x^c_{i1} = x_i + \sigma_1(x_{i+1} - x_i) \tag{1.37}$$
$$x^c_{i2} = x_i + \sigma_2(x_{i+1} - x_i) \tag{1.38}$$

με $j = 1, 2, \ldots, N_{el}$ και

$$x^c_{11} < x^c_{12} < x^c_{21} < x^c_{22} < \ldots < x^c_{N_{el}1} < x^c_{N_{el}2}$$

Για παράδειγμα θεωρούμε το μονοδιάστατο γραμμικό πρόβλημα συνοριακών τιμών, των μοντέλων (1.2):

$$\begin{cases} L\upsilon(x) &= f(x) \quad , \text{με } x \in (a,b) \\ B\upsilon(x) &= g(x) \quad , \text{με } x = a \text{ ή } x = b \end{cases} \tag{1.39}$$

όπου

$$L = -\epsilon \frac{\partial^2 \cdot}{\partial x^2} + p(x) \frac{\partial \cdot}{\partial x}$$

Οι συνοριακές συνθήκες του προβλήματος που θα χρησιμοποιήσουμε θα είναι διαχωρίσιμες τύπου Dirichlet ή Neumann.
Καλούμε την προσεγγιστική λύση $\hat{u}$ και την αντικαθιστούμε στην παραπάνω διαφορική εξίσωση οπότε έχουμε:

$$L\left[\hat{u}\right](x) - f\left(x\right) = E(x) \tag{1.40}$$

όπου $E(x)$ η συνάρτηση υπολοίπου (σφάλματος). Σκοπός μας είναι να επιλέξουμε τα collocation points για να υπολογίσουμε τους αγνώστους u_i, u_i' για $i = 1, \ldots, N_{el}+1$, ελέγχοντας ταυτόχρονα και την συνάρτηση σφάλματος $E(x)$. Απαιτώντας να ικανοποιείται η διαφορική εξίσωση στο σύνολο των $2N_{el}$ collocation points $\prod_{COL} = \left\{x_{ij}^c\right\}_{i,j=1}^{N_{el},2}$ η συνάρτηση υπολοίπου μηδενίζεται $E(x_{ij}^c) = 0$ και αποκτούμε το διακριτό μας μοντέλο:

$$L\left[\hat{u}\right](x_{ij}^c) = f\left(x_{ij}^c\right) \tag{1.41}$$

όπου $i = 1, \ldots, N_{el}$ και $j = 1, 2$.
Η εξίσωση (1.41) παριστάνει ένα σύστημα $2N_{el}$ εξισώσεων με $2N_{el}+2$ αγνώστους u_i, u_i' με $i = 1, \ldots, N_{el}+1$. Σε μορφή πινάκων το γραμμικό σύστημα εκφράζεται ως εξής:

$$\widetilde{C}\widetilde{x} = \widetilde{b} \tag{1.42}$$

με

$$\widetilde{C} \in \mathbb{R}^{2N_{el}, 2N_{el}+2}, \widetilde{x} \in \mathbb{R}^{2N_{el}+2}, \widetilde{b} \in \mathbb{R}^{2N_{el}}.$$

Στην περίπτωση μας χρησιμοποιούμε για τους αγνώστους κανονική αρίθμηση δηλαδή:

$$\widetilde{x} = \left[u_1, u_1', u_2, u_2', \ldots, u_{N_{el}}, u_{N_{el}}', u_{N_{el}+1}, u_{N_{el}+1}'\right]^T.$$

Οπότε ο πίνακας $\widetilde{C}$ που αντιστοιχεί σε αυτήν την αρίθμηση έχει την παρακάτω

μορφή:
$\widetilde{C} =$

$$\begin{pmatrix}
C_{1,1} & C_{2,1} & C_{3,1} & C_{4,1} & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 & 0 \\
C_{5,1} & C_{6,1} & C_{7,1} & C_{8,1} & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & C_{1,2} & C_{2,2} & C_{3,2} & C_{4,2} & \cdots & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & C_{5,2} & C_{6,2} & C_{7,2} & C_{8,2} & \cdots & 0 & 0 & 0 & 0 & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
0 & 0 & 0 & 0 & 0 & 0 & \cdots & C_{1,N_{el}-1} & C_{2,N_{el}-1} & C_{3,N_{el}-1} & C_{4,N_{el}-1} & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & \cdots & C_{5,N_{el}-1} & C_{6,N_{el}-1} & C_{7,N_{el}-1} & C_{8,N_{el}-1} & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & C_{1,N_{el}} & C_{2,N_{el}} & C_{3,N_{el}} & C_{4,N_{el}} \\
0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & C_{5,N_{el}} & C_{6,N_{el}} & C_{7,N_{el}} & C_{8,N_{el}}
\end{pmatrix}$$

όπου

$$\begin{aligned}
C_{1,i} &= L[\xi_{1,i}](x^c_{i1}) & \qquad C_{5,i} &= L[\xi_{1,i}](x^c_{i2}) \\
C_{2,i} &= h_i L[\xi_{2,i}](x^c_{i1}) & \qquad C_{6,i} &= h_i L[\xi_{2,i}](x^c_{i2}) \\
C_{3,i} &= L[\xi_{3,i}](x^c_{i1}) & \qquad C_{7,i} &= L[\xi_{3,i}](x^c_{i2}) \\
C_{4,i} &= h_i L[\xi_{4,i}](x^c_{i1}) & \qquad C_{8,i} &= h_i L[\xi_{4,i}](x^c_{i2})
\end{aligned}$$

και

$$\widetilde{b} = [f\,(x^c_{11}), f\,(x^c_{12}), \ldots, f\,(x^c_{N_{el}1}), f\,(x^c_{N_{el}2})]^T$$

Εισάγοντας τις συνοριακές συνθήκες του προβλήματος στο αριστερό άκρο $x = a$ και στο δεξιό $x = b$ το παραπάνω γραμμικό σύστημα θα μετασχηματιστεί σε ένα σύστημα $2N_{el}$ εξισώσεων με $2N_{el}$ αγνώστους, αφού καθορίζονται 2 από τους προηγούμενους $2N_{el} + 2$. Για παράδειγμα θεωρούμε συνοριακές συνθήκες τύπου Neumann:

$$u(a) = u_a$$
$$u'(b) = u'_b$$

Οπότε το γραμμικό σύστημα (1.42) γίνεται:

$$Cx = b$$

όπου

$$C \in \mathbb{R}^{2N_{el},2N_{el}}, x \in \mathbb{R}^{2N_{el}}, b \in \mathbb{R}^{2N_{el}}$$

Οπότε το διάνυσμα των αγνώστων έχει ως εξής:

$$x = [u'_1, u_2, u'_2, u_3, \ldots, u_{N_{el}-1}, u_{N_{el}}, u'_{N_{el}}, u_{N_{el}+1}]^T$$

Και ο πίνακας των αγνώστων C παίρνει την παρακάτω μορφή:

$$C = \begin{pmatrix} C_{2,1} & C_{3,1} & C_{4,1} & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 \\ C_{6,1} & C_{7,1} & C_{8,1} & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 \\ 0 & C_{1,2} & C_{2,2} & C_{3,2} & C_{4,2} & \cdots & 0 & 0 & 0 & 0 & 0 \\ 0 & C_{5,2} & C_{6,2} & C_{7,2} & C_{8,2} & \cdots & 0 & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & \cdots & C_{1,N_{el}-1} & C_{2,N_{el}-1} & C_{3,N_{el}-1} & C_{4,N_{el}-1} & 0 \\ 0 & 0 & 0 & 0 & 0 & \cdots & C_{5,N_{el}-1} & C_{6,N_{el}-1} & C_{7,N_{el}-1} & C_{8,N_{el}-1} & 0 \\ 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & C_{1,N_{el}} & C_{2,N_{el}} & C_{3,N_{el}} \\ 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & C_{5,N_{el}} & C_{6,N_{el}} & C_{7,N_{el}} \end{pmatrix}.$$

Το δεξιό μέλος του γραμμικού συστήματος γίνεται:

$$b = \begin{pmatrix} f(x^c_{11}) - L[\xi_{1,1}](x^c_{11})u_a \\ f(x^c_{12}) - L[\xi_{1,1}](x^c_{12})u_a \\ f(x^c_{21}) \\ f(x^c_{22}) \\ \vdots \\ f(x^c_{N_{el}-1\ 1}) \\ f(x^c_{N_{el}-1\ 2}) \\ f(x^c_{N_{el}\ 1}) - L[\xi_{4,N_{el}}](x^c_{N_{el}\ 1})u'_b \\ f(x^c_{N_{el}\ 2}) - L[\xi_{4,N_{el}}](x^c_{N_{el}\ 2})u'_b \end{pmatrix}.$$

Παρατηρήσεις

- Ο Collocation πίνακας C έχει μπλοκ τριδιαγώνια μορφή με μπλοκ διάστασης 2×2.
- Δεν είναι συμμετρικός ούτε θετικά ορισμένος.
- Είναι αραιός καθώς και πίνακας ζώνης.

1.3 Εφαρμογή της μεθόδου στο πρόβλημα

Θεωρούμε την 1-dimensional steady-state προβλήματος advection- diffusion που δίνεται από την παρακάτω σχέση:

$$-\epsilon \cdot u''(x) + \beta \cdot u'(x) = 1 \text{ , οταν } 0 < x < 1 \tag{1.43}$$

με συνοριακές συνθήκες τύπου Dirichlet

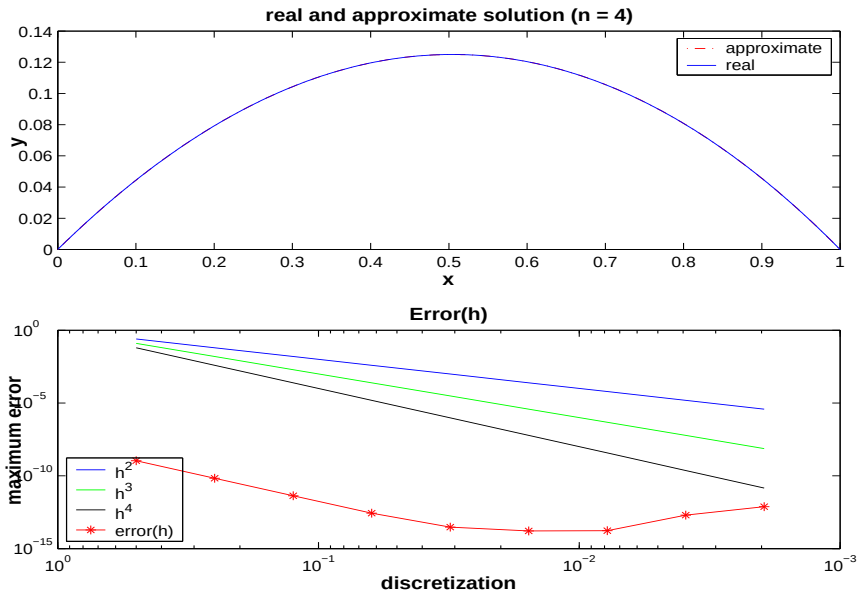
$$u(0) = 0$$

$$u(1) = 0$$

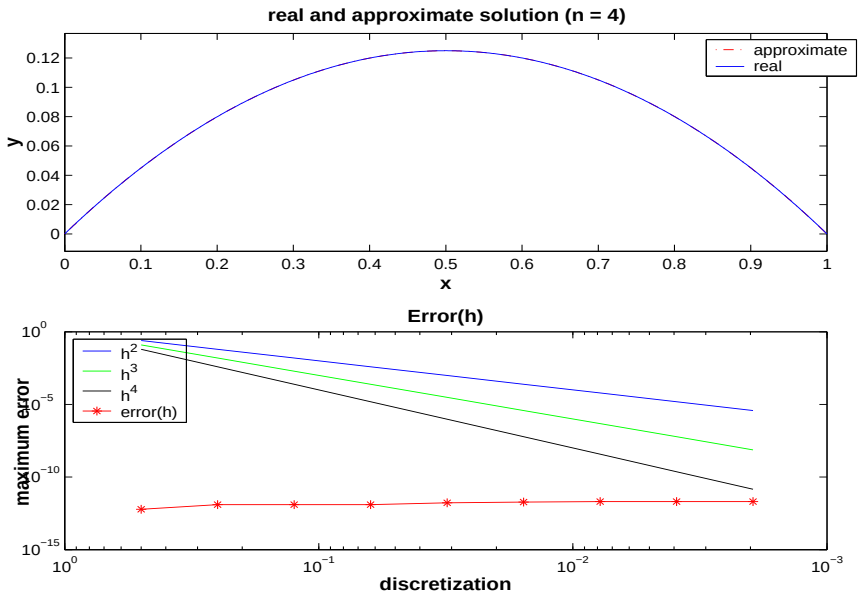
Η αναλυτική λύση του προβλήματος είναι:

$$u(x) = \frac{1 - e^{\frac{\beta x}{\epsilon}}}{e^{\frac{\beta}{\epsilon}} - 1} + \frac{x}{\beta}$$

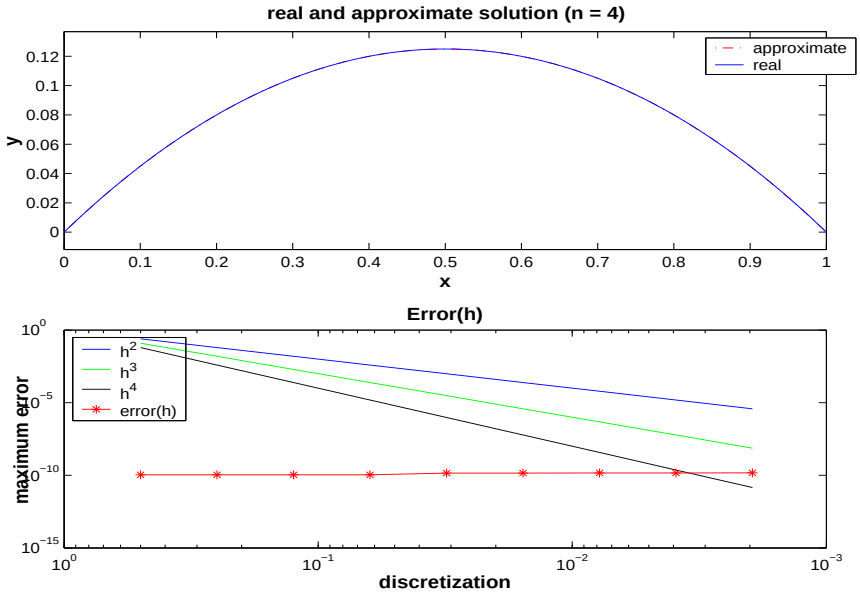
Θα επιχειρήσουμε να λύσουμε αριθμητικά το παραπάνω πρόβλημα χρησιμοποιώντας την Orthogonal FEM Hermite Collocation. Κρατώντας τον συντελεστή διάχυσης $\epsilon > 0$ σταθερό και θεωρώντας ότι συντελεστής *advection* $\beta \to 0$ με $\beta > 0$, παρατηρούμε ότι η αριθμητική επίλυση του παραπάνω προβλήματος δεν παρουσιάζει καμία δυσκολία. Για παράδειγμα θέτοντας $\epsilon = 1$ και $\beta \to 0$ έχουμε τα παρακάτω αποτελέσματα:



Σχήμα 1.3: Η πραγματική, προσεγγιστική λύση καθώς η ταχύτητα σύγκλισης για $\epsilon = 1$ και $\beta = 0.1$ ($\frac{\epsilon}{||\beta||} = 10 \quad h = 0.25$).



Σχήμα 1.4: Η πραγματική, προσεγγιστική λύση καθώς η ταχύτητα σύγκλισης για $\epsilon = 1$ και $\beta = 0.01$ ($\frac{\epsilon}{||\beta||} = 100 \quad h = 0.25$).

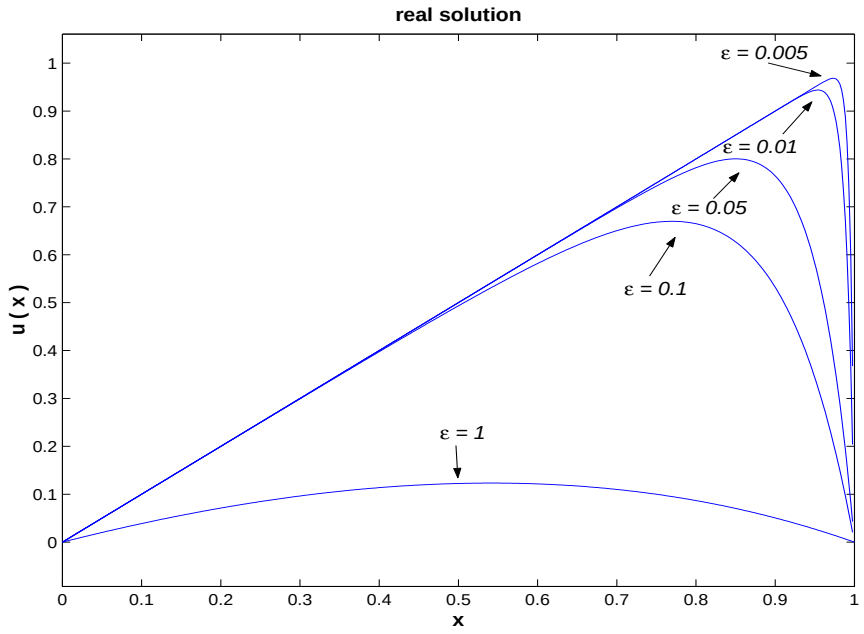


Σχήμα 1.5: Η πραγματική, προσεγγιστική λύση καθώς η ταχύτητα σύγκλισης για $\epsilon = 1$ και $\beta = 0.001$ ($\frac{\epsilon}{||\beta||} = 1000 \quad h = 0.25$).

Παρατηρήσεις

- Ο συντελεστής διάχυσης υπερέχει του συντελεστή advection.
- Η ποσότητα $\frac{\epsilon}{||\beta||}$ γίνεται αρκετά μεγάλη καθώς $\beta \to 0$ σε σχέση με το βήμα διακριτοποίησης της μεθόδου, όπως φαίνεται απο τα παραπάνω σχήματα.
- Η πραγματική λύση παραμένει ομαλή καθώς $\beta \to 0$.
- Ικανοποιούνται τα κριτήρια τάξης σύγκλισης της αριθμητικής μεθόδου.

Στην αντίθετη περίπτωση, που ο συντελεστής ταχύτητας β είναι πολύ μεγαλύτερος από τον συντελεστή διάχυσης ϵ η πραγματική λύση της διαφορικής εξίσωσης τείνει σε μια μη-συνεχή συνάρτηση η οποία την τελευταία στιγμή παίρνει την τιμή 0. Αυτή η περιοχή της απότομης μετάβασης της λύσης καλείται συνοριακό στρώμα "boundary layer". Για παράδειγμα διατηρώντας το β σταθερό και ίσο με 1 και υποθέτοντας ότι $0 < \epsilon \lll 1$ έχουμε:



Σχήμα 1.6: Η πραγματική λύση καθώς $\epsilon \to 0$.

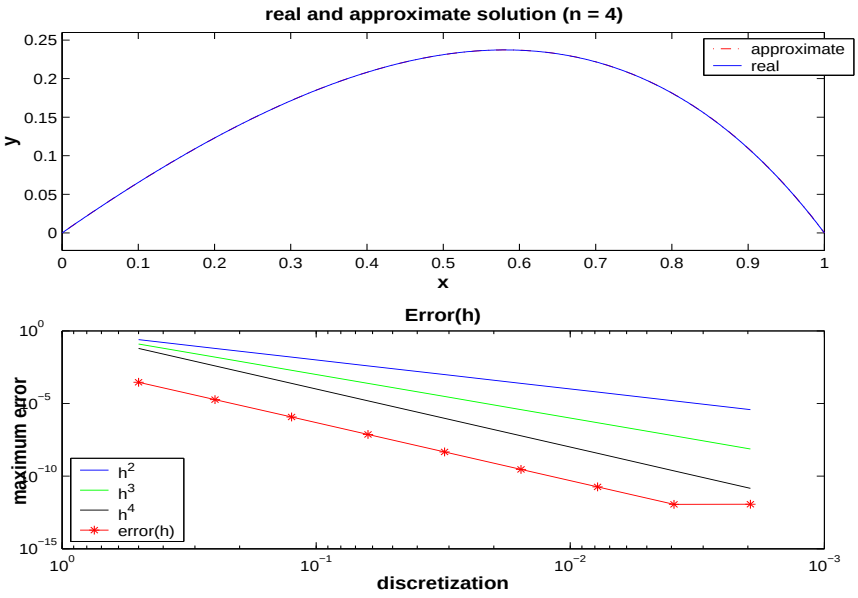
Παρατήρηση

- Παρατηρούμε ότι στο συγκεκριμένο παράδειγμα το πλάτος του boundary layer είναι της τάξεως $O(\epsilon)$.

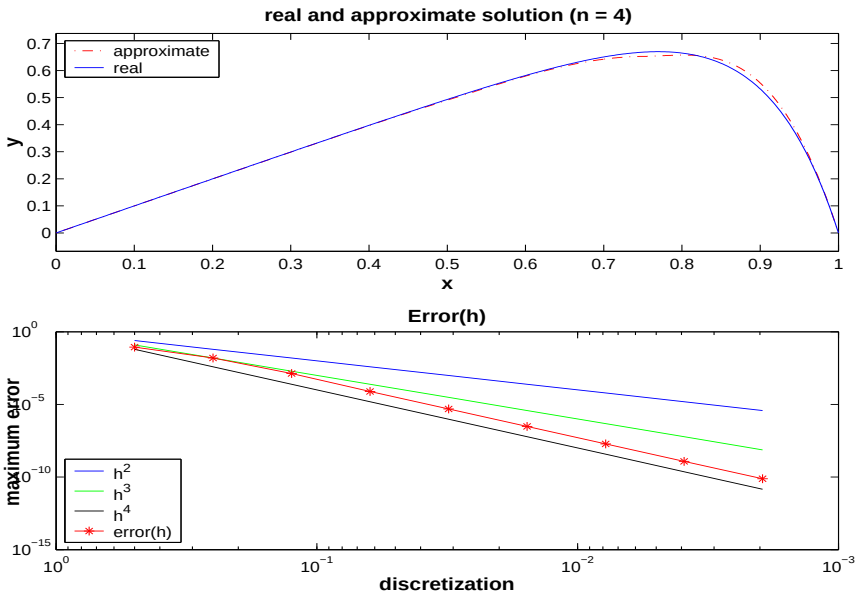
Επιχειρώντας την αριθμητική επίλυση καθώς $\epsilon \to 0$ και $\beta = 1$ παρατηρούμε ότι η συμπεριφορά της μεθόδου κρίνεται ανεπαρκής διότι:

- Καθώς $\epsilon \to 0$ η τάξη σύγκλισης παύει πια να είναι της τάξεως $O(h^4)$ και καθώς το ϵ γίνεται μικρό η τάξη σύγκλισης συνεχώς ελαττώνεται όπως φαίνεται στα παρακάτω σχήματα.

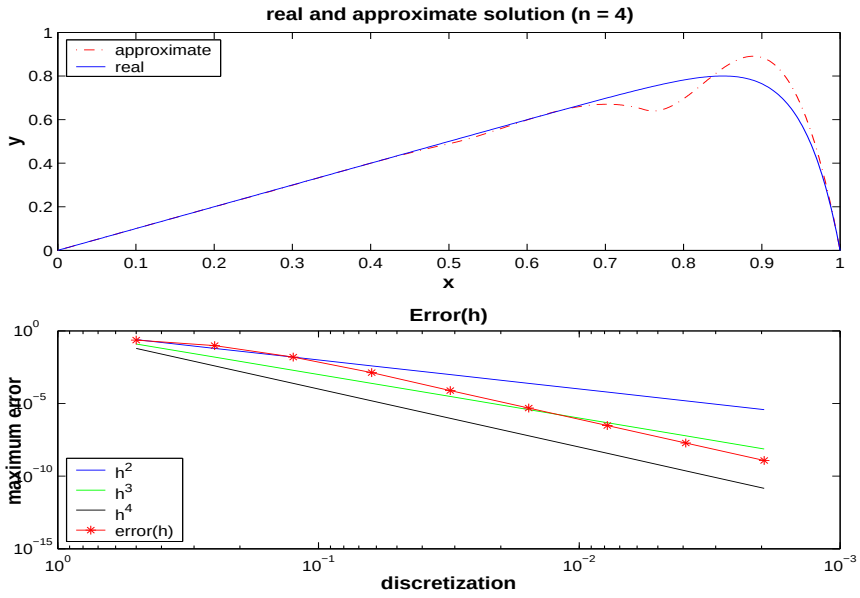
- Επίσης προσεγγιστική λύση παρουσιάζει ταλαντώσεις η κοντά στην περιοχή του boundary layer για μικρές διαμερίσεις.



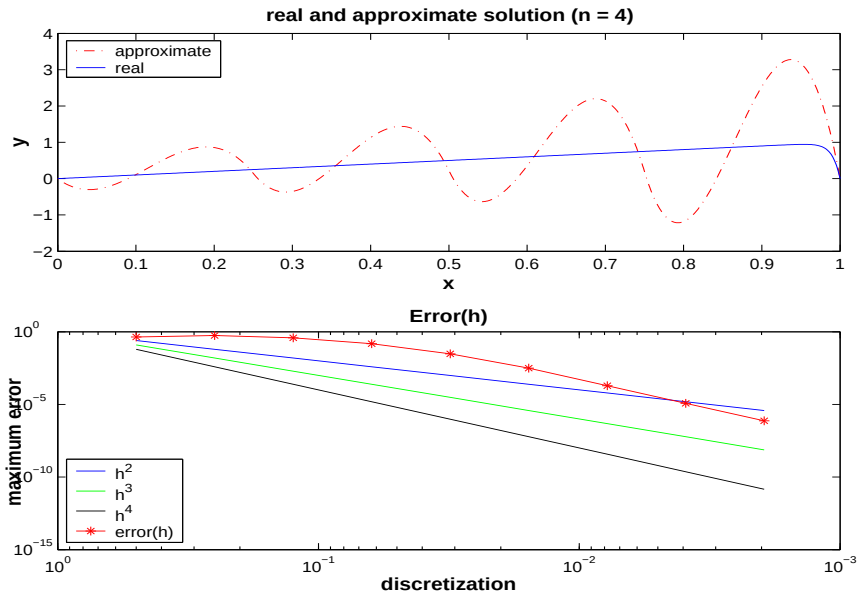
Σχήμα 1.7: Η πραγματική, προσεγγιστική λύση καθώς η ταχύτητα σύγκλισης για $\beta = 1$ και $\epsilon = 0.5$ ($\frac{\epsilon}{||\beta||} = 0.5 \quad h = 0.25$).



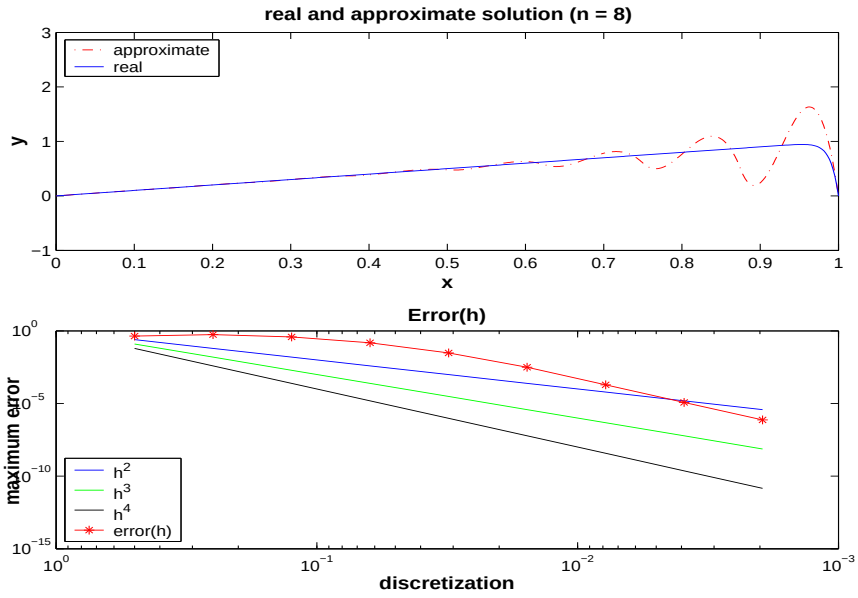
Σχήμα 1.8: Η πραγματική, προσεγγιστική λύση καθώς η ταχύτητα σύγκλισης για $\beta = 1$ και $\epsilon = 0.1$ ($\frac{\epsilon}{||\beta||} = 0.1 \quad h = 0.25$).



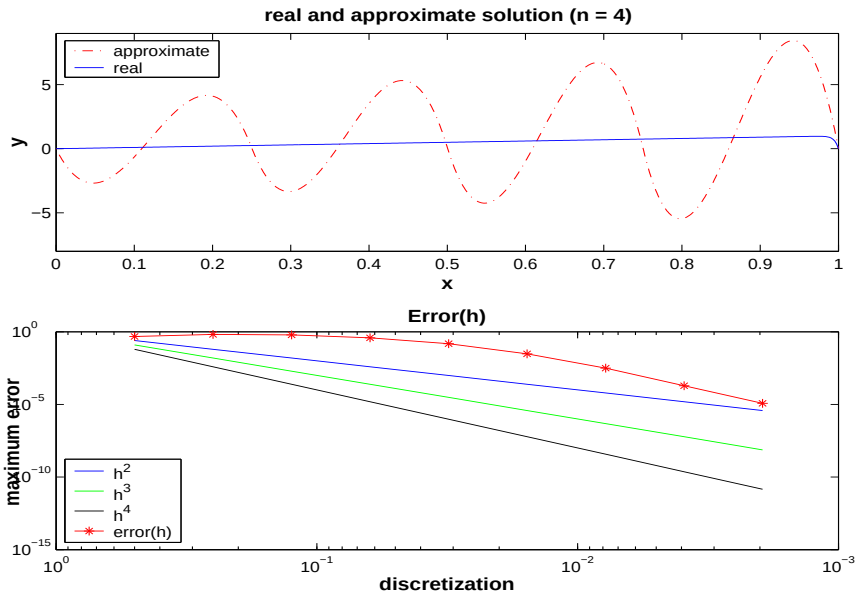
Σχήμα 1.9: Η πραγματική, προσεγγιστική λύση καθώς η ταχύτητα σύγκλισης για $\beta = 1$ και $\epsilon = 0.05$ ($\frac{\epsilon}{||\beta||} = 0.05 \quad h = 0.25$).



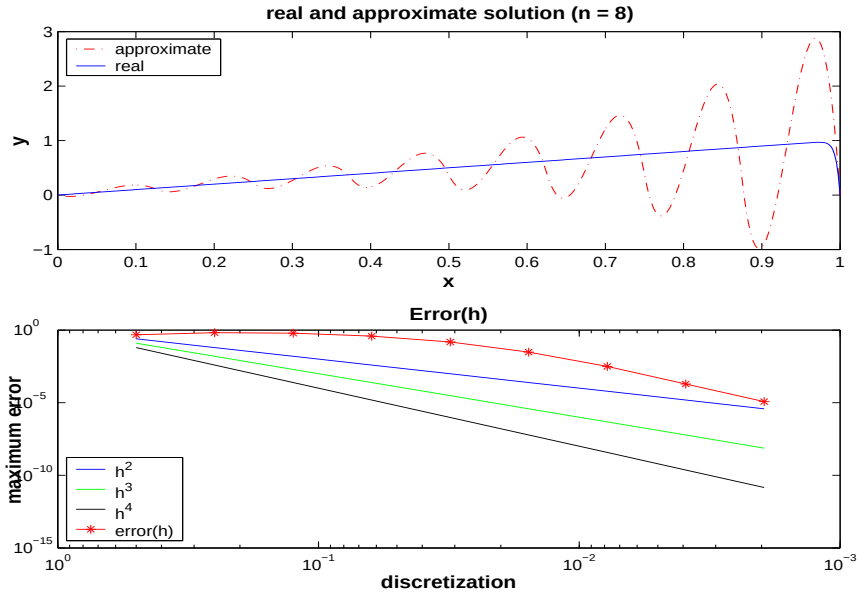
Σχήμα 1.10: Η πραγματική, προσεγγιστική λύση καθώς η ταχύτητα σύγκλισης για $\beta = 1$ και $\epsilon = 0.01$ ($\frac{\epsilon}{||\beta||} = 0.01 \quad h = 0.25$).



Σχήμα 1.11: Η πραγματική, προσεγγιστική λύση καθώς η ταχύτητα σύγκλισης για $\beta = 1$ και $\epsilon = 0.01$ ($\frac{\epsilon}{||\beta||} = 0.01 \quad h = 0.125$).



Σχήμα 1.12: Η πραγματική, προσεγγιστική λύση καθώς η ταχύτητα σύγκλισης για $\beta = 1$ και $\epsilon = 0.005$ ($\frac{\epsilon}{||\beta||} = 0.005 \quad h = 0.25$).



Σχήμα 1.13: Η πραγματική, προσεγγιστική λύση καθώς η ταχύτητα σύγκλισης για $\beta = 1$ και $\epsilon = 0.005$ ($\frac{\epsilon}{||\beta||} = 0.005 \quad h = 0.125$).

Παρατηρήσεις

- Παρατηρούμε ότι όσο μικρότερη είναι η ποσότητα $\frac{\epsilon}{||\beta||}$ σε σχέση με το βήμα διακριτοποίησης τόσο περισσότερες ταλαντώσεις παρουσιάζει η προσεγγιστική λύση, σχήματα 1.7 εώς 1.13. Το χαρακτηριστικό αυτό, όπως ήδη έχουμε αναφέρει εμφανίζεται στα περισσότερα αριθμητικά σχήματα [BRI02,BRI04,HUN93,LEV98].

- Επίσης παρατηρούμε ότι όσο πλησιάζουμε κοντά στην περιοχή του boundary layer τόσο μεγαλύτερα σφάλματα εισέρχονται στη προσεγγιστική λύση.

Θα λέγαμε ότι αυτή η μικρή διαταραχή στην λύση καθώς $\epsilon \to 0$, τείνει να αλλάξει εντελώς τον χαρακτήρα της διαφορικής εξίσωσης από $2^{ας}$ τάξεως, όπου δυο συνοριακές συνθήκες απαιτούνται για την μοναδικότητα της λύσης, σε μια $1^{ης}$ τάξεως, όπου μια αρχική συνθήκη είναι αρκετή για τον μοναδικό ορισμό της λύσης.

Γενικότερα κάθε διαφορική εξίσωση στην οποία μια μικρή παράμετρος πολλαπλασιάζεται με την μέγιστης τάξης παράγωγο που περιέχει, οδηγεί σε ένα

ιδιόμορφο πρόβλημα διαταραχών "singularly pertubed problem". Τέτοιου είδους προβλήματα προκαλούν πολλές αριθμητικές δυσκολίες διότι αλλάζει η λύση ταχύτατα σε μικρά διαστήματα του πεδίου ορισμού της. Σε αυτές τις ιδιόμορφες περιοχές (όπως boundary, interior layers) η $\|u'(x)\|$ γίνεται πολύ μεγάλη προκαλώντας μεγάλα σφάλματα κατά την αριθμητική τους επίλυση κοντά σε αυτές τις περιοχές.

Κεφάλαιο 2

Upwind Hermite Collocation

Η finite element spline collocation έχει χρησιμοποιηθεί ευρέως στην αριθμητική επίλυση Μ.Δ.Ε. και Σ.Δ.Ε. [ASC95], λόγω της υψηλής της τάξης ακρίβειας και της εύκολης εφαρμογής της. Η μέθοδος της collocation στα gauss points, έχει αποδειχθεί ότι δίνει την μέγιστη τάξη ακρίβειας σε σχέση με οποιαδήποτε αλλά collocation points για ικανοποιητικά ομαλές λύσεις. Στην περίπτωση της hermite collocation και για ικανοποιητικά ομαλές λύσεις έχουμε σφάλματα τάξης $O(h^4)$ [BOO73, ASC95]. Όμως σε προβλήματα μεταφοράς-διάχυσης αλλά και γενικότερα σε ιδιόμορφα προβλήματα διαταραχών όπου η πραγματική λύση μεταβάλλεται απότομα σε μικρές περιοχές του πεδίου ορισμού της, η μέθοδος αδυνατεί να προσεγγίσει ικανοποιητικά την λύση, προσδίδοντας ταλαντώσεις στην προσεγγιστική λύση.

Παρόμοια προβλήματα εμφανίζονται και σε συμμετρικά σχήματα πεπερασμένων διαφορών και πεπερασμένων στοιχείων στην επίλυση τέτοιων προβλημάτων. Για την αντιμετώπιση αυτών των προβλημάτων σε αυτές τις μεθόδους, χρησιμοποιούνται upwind σχήματα, τα οποία είναι μη-συμμετρικά. Κάποιες μέθοδοι collocation με upwinding χαρακτηριστικά προτάθηκαν από τους [ASC86, BRI02, BRI04, MAH93, RIN84, SUN00]. Στα σχήματα αυτά χρησιμοποιήθηκαν μη συμμετρικά collocation points ώστε η προσεγγιστική λύση να μην περιέχει ταλαντώσεις.

Θεωρούμε το απλό πρόβλημα συνοριακών τιμών,

$$-\epsilon u_{xx} + p u_x = d(x) \qquad \epsilon > 0 \tag{2.1}$$

με κατάλληλες συνοριακές συνθήκες. Όταν ένα αριθμητικό σχήμα εφαρμοστεί

στο (2.1), τότε έχουμε την ακόλουθη διακριτή μορφή,

$$-\epsilon T_2 \vec{u} + pT_1 \vec{u} = \vec{d}$$

όπου T_1, T_2 οι πίνακες πρώτης και δεύτερης τάξης παραγώγισης αντίστοιχα του σχήματος που χρησιμοποιείται. Η ευστάθεια του αριθμητικού σχήματος εξαρτάται από τις ιδιοτιμές του πίνακα.

$$M = -\epsilon T_2 + pT_1$$

Για αρκετά μικρό ϵ το σχήμα θεωρείται ευσταθές (δηλαδή το αριθμητικό σχήμα δεν προσδίδει ταλαντώσεις στην προσεγγιστική λύση) όταν όλες οι ιδιοτιμές του πίνακα πρώτης παραγώγισης T_1 του σχήματος βρίσκονται σε ένα από τα δυο πραγματικά ημι-επίπεδα. Πιο συγκεκριμένα εάν $p > 0$ τότε οι ιδιοτιμές του T_1 πρέπει να βρίσκονται στο δεξιό πραγματικό ημι-επίπεδο, ενώ εάν $p < 0$ στο αριστερό πραγματικό ημι-επίπεδο, όπως ακριβώς και στην περίπτωση των κλασσικών σχημάτων upwind πεπερασμένων διαφορών [LEV98]. Σχήματα των οποίων οι ιδιοτιμές του T_1 έχουν και θετικά και αρνητικά πραγματικά μέρη δεν είναι ευσταθή.

Μερικά upwinding χαρακτηριστικά για pseudospectral collocation μελετήθηκαν από τους [HUN93]. Αριθμητικά αποτελέσματα και θεωρητική ανάλυση στο πρόβλημα

$$4\epsilon u_{xx} + 2u_x = \frac{1+x}{2} \qquad x \in (-1,1) \quad \epsilon > 0$$
$$u(-1) = u(1) = 0$$

υποθέτουν ότι οι ιδιοτιμές του πρώτης τάξεως πίνακα παραγώγισης, πρέπει να εντοπίζονται στο αριστερό πραγματικό ημι-επίπεδο ώστε το σχήμα να είναι ευσταθές προσδίδοντας λύσεις χωρίς ταλαντώσεις. Αυτό είναι το κύριο χαρακτηριστικό του σχήματος. Το χαρακτηριστικό αυτό εμφανίζεται όπως ήδη αναφέραμε προηγουμένως και στο κλασσικό upwind σχήμα των πεπερασμένων διαφορών.

Σε αυτό το κεφαλαίο θα αναπτύξουμε Hermite Cubic Spline Collocation (H.C.S.C.) μεθόδους όπου θα ενσωματώνουν τα παραπάνω upwinding χαρακτηριστικά. Στην αρχή του κεφαλαίου θα παρουσιάσουμε μια φασματική ανάλυση των πινάκων παραγώγισης της H.C.S.C. σύμφωνα με τους [RUS97, SUN99, WEI88]. Στην συνέχεια βασιζόμενοι σε αυτήν την ανάλυση θα προσπαθήσουμε να κατασκευάσουμε σύνολα από collocation points ώστε η μέθοδος μας

να καλύπτει τις παραπάνω προϋποθέσεις για ένα ευσταθές σχήμα. Επίσης θα ορίσουμε το σύνολο των collocation points για το οποίο η μέθοδος δεν είναι ευσταθής. Στο τέλος του κεφαλαίου θα εφαρμόσουμε την Upwind H.C.S.C. για να επιλύσουμε αριθμητικά κάποια προβλήματα μορφής μεταφοράς-διάχυσης και θα συγκρίνουμε τα αποτελέσματα μας με την orthogonal collocation.

2.1 Collocation Πίνακες Παραγώγισης

Έχουμε ήδη αναφέρει στο πρώτο κεφάλαιο πως κατασκευάζεται το διακριτό σχήμα της Hermite Collocation. Έστω $\Pi_N = \left\{x_i\right\}_{i=1}^{n+1}$ μια ομοιόμορφη διαμέριση του $[0,1]$, ώστε $h = x_{i+1} - x_i$, $i = 1, \ldots, n$. Τότε σύμφωνα με την (1.35) η κυβική Hermite προσέγγιση της λύσης ορίζεται,

$$v(x) = u_i\xi_1(s) + hu_i'\xi_2(s) + u_{i+1}\xi_3(s) + hu_{i+1}'\xi_4(s) \tag{2.2}$$

σε κάθε element $[x_i, x_{i+1}]$ με $i = 1, \ldots, n$. Όπου $\xi_i(s)$ με $i = 1, \ldots, 4$ και $s = \dfrac{x - x_i}{h}$, τα κυβικά πολυώνυμα Hermite στο $[x_i, x_{i+1}]$, όπως ορίζονται από την σχέση (1.30). Προφανώς $v(x_i) = u_i$ και $v(x_{i+1}) = u_{i+1}$. Έστω Ω_c το σύνολο όπως ορίζεται από την σχέση (1.36) και $(\sigma_1, \sigma_2) \in \Omega_c$. Σύμφωνα με τις σχέσεις (1.37),(1.38) ορίζεται το σύνολο των collocation points για την δεδομένη διαμέριση $\Pi_{COL} = \left\{x_{ij}^c\right\}_{i,j=1}^{n,2}$.

Η spline collocation διακριτή μορφή του προβλήματος (2.1) δίνεται,

$$-\epsilon v_{xx}(x_{ij}^c) + pv_x(x_{ij}^c) = d(x_{ij}^c) \qquad i = 1, \ldots, n \quad j = 1, 2$$

η οποία σε μορφή πινάκων γίνεται,

$$-\epsilon A_2\vec{u} + pA_1\vec{u} = \vec{d}$$

όπου A_1, A_2 οι collocation πίνακες πρώτης και δεύτερης παραγώγου αντίστοιχα. Για την διάταξη του διανύσματος $\vec{u}$ χρησιμοποιούμε κανονική αρίθμηση των αγνώστων,

$$\vec{u} = [u_1, hu_1', u_2, hu_2', \ldots, u_n, hu_n', u_{n+1}, hu_{n+1}']^T$$

Έστω ότι με $\vec{v}^k$, $k = 0, 1, 2$ συμβολίσουμε την $k-$οστή παράγωγο, δηλαδή

$$\vec{v}^{(k)} = [v^{(k)}(x^c_{11}), v^{(k)}(x^c_{12}), \ldots, v^{(k)}(x^c_{n1}), v^{(k)}(x^c_{n2})]^T \qquad k = 0, 1, 2$$

τότε παρατηρούμε

$$A_k \vec{u} = \vec{v}^{(k)}. \tag{2.3}$$

Οι collocation πίνακες παραγώγισης A_k με $k = 0, 1, 2$ έχουν την εξής αραιή μορφή όπως προκύπτει απο τις σχέσεις (2.2),(2.3).

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

Χρησιμοποιώντας συνοριακές συνθήκες τύπου Neumann η Dirichlet τότε οι πίνακες A_k με $k = 0, 1, 2$ έχουν την παρακάτω αραιή δομή, αφού λόγω των συνοριακών συνθηκών κάποιοι άγνωστοι θα έχουν πλέον προσδιοριστεί.

$$\begin{pmatrix}
x & x & x & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 & 0 \\
x & x & x & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & x & x & x & x & \cdots & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & x & x & x & x & \cdots & 0 & 0 & 0 & 0 & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
0 & 0 & 0 & 0 & 0 & \cdots & x & x & x & x & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & \cdots & x & x & x & x & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & x & x & x \\
0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & x & x & x
\end{pmatrix}. \tag{2.4}$$

Από την (2.4) παρατηρούμε ότι οι πίνακες A_k έχουν σχεδόν μπλοκ διαγώνια μορφή και περιέχουν $n-2$ μη-μηδενικά 2×4 μπλοκ καθώς και δυο 2×3 μπλοκ. Από τις σχέσεις (2.2),(2.3) έχουμε ότι όλα τα 2×4 μπλοκ είναι ίδια και εκφράζονται ως εξής

$$h^{-k} \begin{pmatrix}
\xi_1^{(k)}(\sigma_1) & \xi_2^{(k)}(\sigma_1) & \xi_3^{(k)}(\sigma_1) & \xi_4^{(k)}(\sigma_1) \\
\xi_1^{(k)}(\sigma_2) & \xi_2^{(k)}(\sigma_2) & \xi_3^{(k)}(\sigma_2) & \xi_4^{(k)}(\sigma_2)
\end{pmatrix} \tag{2.5}$$

όπου $\xi_i^{(k)}(s) = \dfrac{d^k \xi_i}{ds^k}$. Για συνοριακές συνθήκες τύπου Dirichlet το πρώτο και το τελευταίο 2×3 μπλοκ είναι

$$h^{-k} \begin{pmatrix} \xi_2^{(k)}(\sigma_1) & \xi_3^{(k)}(\sigma_1) & \xi_4^{(k)}(\sigma_1) \\ \xi_2^{(k)}(\sigma_2) & \xi_3^{(k)}(\sigma_2) & \xi_4^{(k)}(\sigma_2) \end{pmatrix}$$

$$h^{-k} \begin{pmatrix} \xi_1^{(k)}(\sigma_1) & \xi_2^{(k)}(\sigma_1) & \xi_4^{(k)}(\sigma_1) \\ \xi_1^{(k)}(\sigma_2) & \xi_2^{(k)}(\sigma_2) & \xi_4^{(k)}(\sigma_2) \end{pmatrix}$$

αντίστοιχα, ενώ για συνοριακές τύπου Neumann έχουμε:

$$h^{-k} \begin{pmatrix} \xi_1^{(k)}(\sigma_1) & \xi_3^{(k)}(\sigma_1) & \xi_4^{(k)}(\sigma_1) \\ \xi_1^{(k)}(\sigma_2) & \xi_3^{(k)}(\sigma_2) & \xi_4^{(k)}(\sigma_2) \end{pmatrix}$$

$$h^{-k} \begin{pmatrix} \xi_1^{(k)}(\sigma_1) & \xi_2^{(k)}(\sigma_1) & \xi_3^{(k)}(\sigma_1) \\ \xi_1^{(k)}(\sigma_2) & \xi_2^{(k)}(\sigma_2) & \xi_3^{(k)}(\sigma_2) \end{pmatrix}.$$

Από την σχέση (2.3) έχουμε ότι $\vec{v} = A_0 \vec{u}$. Οπότε $\vec{u} = A_0^{-1} \vec{v}$. Ο πίνακας A_0 υποθέσαμε ότι αντιστρέφεται. Στην παράγραφο 2.2.1 αποδεικνύουμε ότι για οποιαδήποτε επιλογή των collocation points με συνοριακές συνθήκες τύπου Dirichlet ή Neumann ο A_0^{-1} υπάρχει. Συνεπώς το διακριτό μας μοντέλο,

$$-\epsilon A_2 \vec{u} + p A_1 \vec{u} = \vec{d}$$

ισοδύναμα γράφεται,

$$-\epsilon A_2 A_0^{-1} \vec{v} + p A_1 A_0^{-1} \vec{v} = \vec{d} \Leftrightarrow$$

$$-\epsilon T_2 \vec{v} + p T_1 \vec{v} = \vec{d}$$

με $T_1 = A_1 A_0^{-1}$ και $T_2 = A_2 A_0^{-1}$.

Στην συνέχεια αυτού του κεφαλαίου χρησιμοποιώντας το τελευταίο ισοδύναμο διακριτό μοντέλο θα προσπαθήσουμε να κατασκευάσουμε σύνολα από collocation points, ώστε η μέθοδος μας να καλύπτει τις προϋποθέσεις που αναφέραμε στην αρχή του κεφαλαίου αυτού για ένα ευσταθές upwind σχήμα.

2.2 Φασματική Ανάλυση

Υποθέτουμε το ακόλουθο πρόβλημα ιδιοτιμών, με $k = 0, 1, 2$

$$T_k v = \lambda v$$

όμως $T_k = A_k A_0^{-1}$ άρα,

$$A_k A_0^{-1} v = \lambda v \Leftrightarrow$$
$$A_k v = \lambda A_0 v$$

και θεωρούμε τον πίνακα $W(\lambda) = A_k - \lambda A_0$. Για συνθήκες τύπου Dirichlet ορίζουμε τον πίνακα,

$$C_k^D(\lambda) = \begin{pmatrix} h^{-k}\xi_2^{(k)}(\sigma_1) - \lambda\xi_2(\sigma_1) & h^{-k}\xi_4^{(k)}(\sigma_1) - \lambda\xi_4(\sigma_1) \\ h^{-k}\xi_2^{(k)}(\sigma_2) - \lambda\xi_2(\sigma_2) & h^{-k}\xi_4^{(k)}(\sigma_2) - \lambda\xi_4(\sigma_2) \end{pmatrix} \tag{2.6}$$

και με $\widetilde{C_k^D}$ συμβολίζουμε τον συζυγή πίνακα του C_k^D. Όποτε:

$$\widetilde{C_k^D}(\lambda) = \begin{pmatrix} h^{-k}\xi_4^{(k)}(\sigma_2) - \lambda\xi_4(\sigma_2) & -h^{-k}\xi_4^{(k)}(\sigma_1) + \lambda\xi_4(\sigma_1) \\ -h^{-k}\xi_2^{(k)}(\sigma_2) + \lambda\xi_2(\sigma_2) & h^{-k}\xi_2^{(k)}(\sigma_1) - \lambda\xi_2(\sigma_1) \end{pmatrix}. \tag{2.7}$$

Για παράδειγμα στην περίπτωση όπου $n = 3$ οι πίνακες A_k περιέχουν ένα μη-μηδενικό μπλοκ 2×4 και προφανώς δύο μη-μηδενικά μπλοκ 2×3. Τότε πολλαπλασιάζοντας απο αριστερά με τον μπλοκ διαγώνιο πίνακα $diag(\widetilde{C_k^D}, \widetilde{C_k^D}, \widetilde{C_k^D})$ με τον πίνακα $W(\lambda) = A_k - \lambda A_0$ έχουμε ότι,

$$diag(\widetilde{C_k^D}, \ldots, \widetilde{C_k^D})(A_k - \lambda A_0) = \begin{pmatrix} z & x_2 & 0 & & & \\ 0 & x_2' & z & & & \\ & x_1 & z & x_2 & 0 & \\ & x_1' & 0 & x_2' & z & \\ & & & x_1 & z & 0 \\ & & & x_1' & 0 & z \end{pmatrix}$$

όπου

$$z = \det(C_k^D(\lambda)) = \det\begin{pmatrix} h^{-k}\xi_2^{(k)}(\sigma_1) - \lambda\xi_2(\sigma_1) & h^{-k}\xi_4^{(k)}(\sigma_1) - \lambda\xi_4(\sigma_1) \\ h^{-k}\xi_2^{(k)}(\sigma_2) - \lambda\xi_2(\sigma_2) & h^{-k}\xi_4^{(k)}(\sigma_2) - \lambda\xi_4(\sigma_2) \end{pmatrix}$$

$$x_1 = \det\begin{pmatrix} h^{-k}\xi_1^{(k)}(\sigma_1) - \lambda\xi_1(\sigma_1) & h^{-k}\xi_4^{(k)}(\sigma_1) - \lambda\xi_4(\sigma_1) \\ h^{-k}\xi_1^{(k)}(\sigma_2) - \lambda\xi_1(\sigma_2) & h^{-k}\xi_4^{(k)}(\sigma_2) - \lambda\xi_4(\sigma_2) \end{pmatrix}$$

$$x_1' = -\det\begin{pmatrix} h^{-k}\xi_1^{(k)}(\sigma_1) - \lambda\xi_1(\sigma_1) & h^{-k}\xi_2^{(k)}(\sigma_1) - \lambda\xi_2(\sigma_1) \\ h^{-k}\xi_1^{(k)}(\sigma_2) - \lambda\xi_1(\sigma_2) & h^{-k}\xi_2^{(k)}(\sigma_2) - \lambda\xi_2(\sigma_2) \end{pmatrix}$$

$$x_2 = \det\begin{pmatrix} h^{-k}\xi_3^{(k)}(\sigma_1) - \lambda\xi_3(\sigma_1) & h^{-k}\xi_4^{(k)}(\sigma_1) - \lambda\xi_4(\sigma_1) \\ h^{-k}\xi_3^{(k)}(\sigma_2) - \lambda\xi_3(\sigma_2) & h^{-k}\xi_4^{(k)}(\sigma_2) - \lambda\xi_4(\sigma_2) \end{pmatrix}$$

$$x_2' = \det\begin{pmatrix} h^{-k}\xi_2^{(k)}(\sigma_1) - \lambda\xi_2(\sigma_1) & h^{-k}\xi_3^{(k)}(\sigma_1) - \lambda\xi_3(\sigma_1) \\ h^{-k}\xi_2^{(k)}(\sigma_2) - \lambda\xi_2(\sigma_2) & h^{-k}\xi_3^{(k)}(\sigma_2) - \lambda\xi_3(\sigma_2) \end{pmatrix}.$$

Γενικεύοντας για οποιοδήποτε n έχουμε,

$$diag(\widetilde{C_k^D}, \ldots, \widetilde{C_k^D})(A_k - \lambda A_0) = \begin{pmatrix} z & x_2 & 0 & & & & & & \\ 0 & x_2' & z & & & & & & \\ & x_1 & z & x_2 & 0 & & & & \\ & x_1' & 0 & x_2' & z & & & & \\ & & & x_1 & z & x_2 & 0 & & \\ & & & x_1' & 0 & x_2' & z & & \\ & & & & & & \ddots & & \\ & & & & & & & x_1 & z & 0 \\ & & & & & & & x_1' & 0 & z \end{pmatrix}. \tag{2.8}$$

Στον πίνακα που προκύπτει από την (2.8), αφαιρώντας την $2j-1$ γραμμή από την $2j$, τότε αποκτούμε μια γραμμή με τρεις μη μηδενικές εισόδους για $j = 2, \ldots, n-2$ και μια με δύο μη μηδενικές εισόδους για $j = 1$ και $j = n-1$. Πίο συγκεκριμένα γράφουμε,

$$G \cdot diag(\widetilde{C_k^D}, \ldots, \widetilde{C_k^D}) \cdot (A_k - \lambda A_0) =$$

$$\begin{pmatrix}
z & x_2 & 0 & 0 & & & & & & & \\
0 & x_2' - x_1 & 0 & -x_2 & & & & & & & \\
 & x_1 & z & x_2 & 0 & 0 & & & & & \\
 & x_1' & 0 & x_2' - x_1 & 0 & -x_2 & & & & & \\
 & & & x_1 & z & x_2 & 0 & 0 & & & \\
 & & & x_1' & 0 & x_2' - x_1 & 0 & -x_2 & & & \\
 & & & & & & \ddots & \ddots & & & \\
 & & & & & & & & x_1 & z & 0 \\
 & & & & & & & & x_1' & 0 & z
\end{pmatrix} \tag{2.9}$$

όπου,

$$G = diag(1, G_1, \ldots, G_1, 1)$$

με

$$G_1 = \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix}.$$

Όμως αναδιατάσσοντας τις γραμμές και τις στήλες του πίνακα που προκύπτει από την σχέση (2.9) έχουμε,

$$G \cdot diag(\widetilde{C_k^D}, \ldots, \widetilde{C_k^D}) \cdot (A_k - \lambda A_0) = P \cdot \begin{pmatrix} B_1^k(\lambda) & O \\ X & B_2^k(\lambda) \end{pmatrix} \cdot P \tag{2.10}$$

με πίνακα $P = P_{2n-3,2n-2} \cdot P_{2n-5,2n-3} \cdot P_{2n-7,2n-4} \cdots P_{1,n} \in \mathbb{R}^{2n,2n}$. Συμβολίζουμε με $P_{i,j}$ τον πίνακα μετάθεσης, ο οποίος προκύπτει από τον ταυτοτικό πίνακα εάν με εναλλάξουμε την i με την j γραμμή του. Το μπλοκ κομμάτι $B_k^1(\lambda)$ του πίνακα $G \cdot diag(\widetilde{C_k^D}, \ldots, \widetilde{C_k^D}) \cdot (A_k - \lambda A_0)$ είναι ένας τριδιαγώνιος πίνακας με διάσταση $n-1 \times n-1$ και έχει την εξής δομή,

$$B_k^1 = h^{-2k} tridiag(a_k(\lambda)), b_k(\lambda)), c_k(\lambda))) \tag{2.11}$$

με

$$
\begin{aligned}
a_k(\lambda) &= -\det\begin{pmatrix} \xi_1^{(k)}(\sigma_1) - \lambda h^k \xi_1(\sigma_1) & \xi_2^{(k)}(\sigma_1) - \lambda h^k \xi_2(\sigma_1) \\ \xi_1^{(k)}(\sigma_2) - \lambda h^k \xi_1(\sigma_2) & \xi_2^{(k)}(\sigma_2) - \lambda h^k \xi_2(\sigma_2) \end{pmatrix} \\
b_k(\lambda) &= \det\begin{pmatrix} \xi_2^{(k)}(\sigma_1) - \lambda h^k \xi_2(\sigma_1) & \xi_3^{(k)}(\sigma_1) - \lambda h^k \xi_3(\sigma_1) \\ \xi_2^{(k)}(\sigma_2) - \lambda h^k \xi_2(\sigma_2) & \xi_3^{(k)}(\sigma_2) - \lambda h^k \xi_3(\sigma_2) \end{pmatrix} \\
&\quad - \det\begin{pmatrix} \xi_1^{(k)}(\sigma_1) - \lambda h^k \xi_1(\sigma_1) & \xi_4^{(k)}(\sigma_1) - \lambda h^k \xi_4(\sigma_1) \\ \xi_1^{(k)}(\sigma_2) - \lambda h^k \xi_1(\sigma_2) & \xi_4^{(k)}(\sigma_2) - \lambda h^k \xi_4(\sigma_2) \end{pmatrix} \\
c_k(\lambda) &= -\det\begin{pmatrix} \xi_3^{(k)}(\sigma_1) - \lambda h^k \xi_3(\sigma_1) & \xi_4^{(k)}(\sigma_1) - \lambda h^k \xi_4(\sigma_1) \\ \xi_3^{(k)}(\sigma_2) - \lambda h^k \xi_3(\sigma_2) & \xi_4^{(k)}(\sigma_2) - \lambda h^k \xi_4(\sigma_2) \end{pmatrix}
\end{aligned}
\tag{2.12}
$$

ενώ το μπλοκ κομμάτι $B_k^2(\lambda)$ έχει διάσταση $n+1 \times n+1$ και η δομή του είναι η παρακάτω.

$$B_k^2 = \det(C_k^D(\lambda)) I_{n+1}.$$

Από τις παραπάνω σχέσεις (2.12) προκύπτει άμεσα ότι τα $a_k(\lambda), b_k(\lambda), c_k(\lambda)$ είναι τετραγωνικά πολυώνυμα ως προς λ.

Λήμμα 2.1 *Από την σχέση (2.10) προκύπτουν τα παρακάτω.*

$$\det(A_k - \lambda A_0) = h^{-2k(n-1)} \det(C_k^D(\lambda)) \cdot \det(tridiag(a_k(\lambda), b_k(\lambda), c_k(\lambda))). \tag{2.13}$$

Η $\det(C_k^D(\lambda))$ *είναι τετραγωνικό πολυώνυμο ως προς* λ *ενώ η* $\det(tridiag(a_k(\lambda), b_k(\lambda), c_k(\lambda))$ *είναι πολυώνυμο ως προς* λ *βαθμού* $2n-2$.

Για συνοριακές συνθήκες τύπου Neumann υπάρχει μια αλλαγή στο πρώτο και στο τελευταίο μπλοκ του $A_k - \lambda A_0$. *Όμως θεωρώντας,*

$$C_k^N(\lambda) = \begin{pmatrix} h^{-k}\xi_1^{(k)}(\sigma_1) - \lambda \xi_1(\sigma_1) & h^{-k}\xi_3^{(k)}(\sigma_1) - \lambda \xi_3(\sigma_1) \\ h^{-k}\xi_1^{(k)}(\sigma_2) - \lambda \xi_1(\sigma_2) & h^{-k}\xi_3^{(k)}(\sigma_2) - \lambda \xi_3(\sigma_2) \end{pmatrix} \tag{2.14}$$

και χρησιμοποιώντας τον $C_k^N(\lambda)$ *αντί του* $C_k^D(\lambda)$ *για την παραπάνω διαδικασία, αποκτούμε ομοίως*

$$\det(A_k - \lambda A_0) = h^{-2k(n-1)} \det(C_k^N(\lambda)) \cdot \det(tridiag(a_k(\lambda), b_k(\lambda), c_k(\lambda))). \tag{2.15}$$

□

Είναι φανερό από τα παραπάνω ότι το χαρακτηριστικό πολυώνυμο, για συνοριακές συνθήκες Dirichlet αλλά και Neumann έχει έναν κοινό παράγοντα βαθμού $2n-2$.

2.2.1 Μερικές Ιδιότητες των Πινάκων Παραγώγισης

Απλοποιώντας την σχέση (2.12) για $k = 0, 1, 2$ γράφουμε,

$k = 0$

$$\begin{aligned}
a_0(\lambda) =& -(\lambda-1)^2(\sigma_1-1)^2(\sigma_2-1)^2(\sigma_2-\sigma_1) \\
b_0(\lambda) =& \quad (\lambda-1)^2(\sigma_2-\sigma_1)\big(2\sigma_1\sigma_2(1-\sigma_1)(1-\sigma_2)+\sigma_1(1-\sigma_1)+\sigma_2(1-\sigma_2)\big) \\
c_0(\lambda) =& -(\lambda-1)^2\sigma_1^2\sigma_2^2(\sigma_2-\sigma_1)
\end{aligned} \tag{2.16}$$

$k = 1$

$$\begin{aligned}
a_1(\lambda) =& -(1-\sigma_1)(1-\sigma_2)(\sigma_2-\sigma_1)\big((1-\sigma_1)(1-\sigma_2)h^2\lambda^2+2(2-\sigma_1-\sigma_2)h\lambda+6\big) \\
b_1(\lambda) =& \quad (\sigma_2-\sigma_1)\Big([2\sigma_1\sigma_2(1-\sigma_1)(1-\sigma_2)+\sigma_1(1-\sigma_1)+\sigma_2(1-\sigma_2)]h^2\lambda^2 \\
& \quad -2(1-\sigma_1-\sigma_2)(1+\sigma_1+\sigma_2-2\sigma_1\sigma_2)h\lambda+6(1+2\sigma_1\sigma_2-\sigma_1-\sigma_2)\Big) \\
c_1(\lambda) =& -\sigma_1\sigma_2(\sigma_2-\sigma_1)\big(\sigma_1\sigma_2h^2\lambda^2-2(\sigma_1+\sigma_2)h\lambda+6\big)
\end{aligned} \tag{2.17}$$

$k = 2$

$$\begin{aligned}
a_2(\lambda) =& -(\sigma_2-\sigma_1)\big((1-\sigma_1)^2(1-\sigma_2)^2h^4\lambda^2 - 2(\sigma_2-\sigma_1)^2h^2\lambda + 12\big)\\
b_2(\lambda) =& \quad (\sigma_2-\sigma_1)\Big([2\sigma_1\sigma_2(1-\sigma_1)(1-\sigma_2)+\sigma_1(1-\sigma_1)+\sigma_2(1-\sigma_2))]h^4\lambda^2\\
&+(12-4(\sigma_2-\sigma_1)^2)h^2\lambda+24\Big)\\
c_2(\lambda) =& -(\sigma_2-\sigma_1)(\sigma_1^2\sigma_2^2h^4\lambda^2 - 2(\sigma_2-\sigma_1)^2h^2\lambda+12).
\end{aligned} \tag{2.18}$$

Θέτοντας $\lambda = 0$ στις παραπάνω σχέσεις (2.16),(2.17),(2.18) από το λήμμα 2.1 έχουμε,

$$\det(A_k) = h^{-2k(n-1)}\det(C_k^D(0))\cdot\det(tridiag(a_k(0),b_k(0),c_k(0))) \tag{2.19}$$

με

$$\det(C_k^D(0)) = h^{-2k}\det\begin{pmatrix}\xi_2^{(k)}(\sigma_1) & \xi_4^{(k)}(\sigma_1)\\ \xi_2^{(k)}(\sigma_2) & \xi_4^{(k)}(\sigma_2)\end{pmatrix} \tag{2.20}$$

$$\begin{aligned}
a_0(0) =& -(1-\sigma_1)^2(1-\sigma_2)^2(\sigma_2-\sigma_1)\\
b_0(0) =& \quad (\sigma_2-\sigma_1)\big(2\sigma_1\sigma_2(1-\sigma_1)(1-\sigma_2)+\sigma_1(1-\sigma_1)+\sigma_2(1-\sigma_2)\big)\\
c_0(0) =& -\sigma_1^2\sigma_2^2(\sigma_2-\sigma_1)
\end{aligned} \tag{2.21}$$

και

$$\begin{aligned}
a_1(0) =& -6(1-\sigma_1)(1-\sigma_2)(\sigma_2-\sigma_1)\\
b_1(0) =& \quad 6(1+2\sigma_1\sigma_2-\sigma_1-\sigma_2)(\sigma_2-\sigma_1)\\
c_1(0) =& -6\sigma_1\sigma_2(\sigma_2-\sigma_1)
\end{aligned} \tag{2.22}$$

$$\begin{aligned}
a_2(0) =& -12(\sigma_2-\sigma_1)\\
b_2(0) =& \quad 24(\sigma_2-\sigma_1)\\
c_2(0) =& -(\sigma_2-\sigma_1)(\sigma_1^2\sigma_2^2h^4\lambda^2 - 2(\sigma_2-\sigma_1)^2h^2\lambda+12).
\end{aligned}$$

(2.23)

Από την εξίσωση (2.13) του λήμματος 2.1, προκύπτουν εύκολα ορισμένες ιδιότητες των πινάκων παραγώγισης, όπως για παράδειγμα η τάξη τους.

Θεώρημα 2.1 *Για συνοριακές συνθήκες τύπου Dirichlet, ο πίνακας* A_0 *είναι αντιστρέψιμος για οποιαδήποτε επιλογή των* $(\sigma_1, \sigma_2) \in \Omega_c$.

ΑΠΟΔΕΙΞΗ

Με βάση αλγεβρικές ιδιότητες για τριδιαγώνιους και Toeplitz πίνακες, [MEY00 p.514], έχουμε ότι

$$\det(tridiag(a_0(0), b_0(0), c_0(0))) = \prod_{j=1}^{n-1} \Big(b_0(0) + 2\sqrt{a_0(0)c_0(0)} \cos\frac{j\pi}{n} \Big). \quad (2.24)$$

Όμως,

$$b_0(0) - 2\sqrt{a_0(0)c_0(0)} = (\sigma_1 - \sigma_2)^2 \big((\sigma_1(1-\sigma_1) + \sigma_2(1-\sigma_2)) \big) > 0$$

συνεπώς,

$$b_0(0) + 2\sqrt{a_0(0)c_0(0)} \cos\frac{j\pi}{n} > 0$$

για κάθε j, όποτε

$$\det(tridiag(a_0(0), b_0(0), c_0(0))) \neq 0. \quad (2.25)$$

Από την σχέση (2.20),

$$\det(C_0^D(0)) = \det \begin{pmatrix} \xi_2(\sigma_1) & \xi_4(\sigma_1) \\ \xi_2(\sigma_2) & \xi_4(\sigma_2) \end{pmatrix} = \sigma_1\sigma_2(1-\sigma_1)(1-\sigma_2)(\sigma_2-\sigma_1) \neq 0.$$

Οπότε ο A_0 είναι αντιστρέψιμος για κάθε επιλογή των $(\sigma_1, \sigma_2) \in \Omega_c$. □

Θεώρημα 2.2 *Για συνοριακές συνθήκες τύπου Neumann, ο πίνακας A_0 είναι αντιστρέψιμος για οποιαδήποτε επιλογή των $(\sigma_1, \sigma_2) \in \Omega_c$.*

ΑΠΟΔΕΙΞΗ

Από την (1.25) έχουμε,

$$\det(tridiag(a_0(0), b_0(0), c_0(0))) \neq 0$$

και από την σχέση (2.20),

$$\det(C_2^N(0)) = \det \begin{pmatrix} \xi_1(\sigma_1) & \xi_1(\sigma_1) \\ \xi_1(\sigma_2) & \xi_3(\sigma_2) \end{pmatrix} = (\sigma_2 - \sigma_1)\big(3(\sigma_1 + \sigma_2) - 2(\sigma_1^2 + \sigma_1\sigma_2 + \sigma_2^2)\big) =$$
$$(\sigma_2 - \sigma_1)\big(2(\sigma_1 + \sigma_2) - 2(\sigma_1^2 + \sigma_2^2) + (\sigma_1 + \sigma_2) - 2\sigma_1\sigma_2\big) > 0.$$

Διότι,

$$0 < \sigma_1 < 1 \Rightarrow 2\sigma_1 > 2\sigma_1^2$$
$$0 < \sigma_2 < 1 \Rightarrow 2\sigma_2 > 2\sigma_2^2$$
$$\sigma_1 + \sigma_2 > \sigma_1^2 + \sigma_2^2 > 2\sigma_1\sigma_2 > 0.$$

Οπότε ο A_0 είναι αντιστρέψιμος για οποιαδήποτε επιλογή των $(\sigma_1, \sigma_2) \in \Omega_c$. □

Θεώρημα 2.3 *Για συνοριακές συνθήκες τύπου Dirichlet, ο πίνακας A_2 είναι αντιστρέψιμος για οποιαδήποτε επιλογή των $(\sigma_1, \sigma_2) \in \Omega_c$.*

ΑΠΟΔΕΙΞΗ

Όπως με προηγουμένως έχουμε,

$$\det(tridiag(a_2(0), b_2(0), c_2(0))) = \prod_{j=1}^{n-1} \Big(b_2(0) + 2\sqrt{a_2(0)c_2(0)}\cos\frac{j\pi}{n}\Big). \quad (2.26)$$

Όμως,

$$b_2(0) - 2\sqrt{a_2(0)c_2(0)} = 48(\sigma_2 - \sigma_1) > 0$$

συνεπώς,

$$b_2(0) + 2\sqrt{a_2(0)c_2(0)}\cos\frac{j\pi}{n} > 0$$

για κάθε j, οπότε

$$\det(tridiag(a_2(0), b_2(0), c_2(0))) \neq 0. \quad (2.27)$$

Από την σχέση (2.20),

$$\det(C_2^D(0)) = \det \begin{pmatrix} \xi_2^{(2)}(\sigma_1) & \xi_4^{(2)}(\sigma_1) \\ \xi_2^{(2)}(\sigma_2) & \xi_4^{(2)}(\sigma_2) \end{pmatrix} = 12(\sigma_1 - \sigma_2) \neq 0.$$

Οπότε ο A_2 είναι αντιστρέψιμος για κάθε επιλογή των $(\sigma_1, \sigma_2) \in \Omega_c$. □

Θεώρημα 2.4 *Για συνοριακές συνθήκες τύπου Neumann, ο πίνακας A_2 δεν είναι αντιστρέψιμος για καμία επιλογή των $(\sigma_1, \sigma_2) \in \Omega_c$.*

ΑΠΟΔΕΙΞΗ

Από την σχέση (2.14),

$$h^{-2k} \det(C_2^N(0)) = \det \begin{pmatrix} \xi_1^{(2)}(\sigma_1) & \xi_3^{(2)}(\sigma_1) \\ \xi_1^{(2)}(\sigma_2) & \xi_3^{(2)}(\sigma_2) \end{pmatrix} = 0.$$

Οπότε ο A_2 δεν είναι αντιστρέψιμος για καμία επιλογή των $(\sigma_1, \sigma_2) \in \Omega_c$. □

2.2.2 Ευαστηθή Collocation points

Όπως ήδη αναφέραμε ένα διακριτό σχήμα της μορφής (2.1) θεωρείται ευσταθές όταν όλες οι ιδιοτιμές του πίνακα πρώτης τάξεως παραγώγισης βρίσκονται σε ένα από τα δυο πραγματικά ημι-επίπεδα, ανάλογα με το πρόσημο του συντελεστή advection. Αυτό είναι και το κύριο χαρακτηριστικό των upwinding σχημάτων των πεπερασμένων διάφορων. Σε αυτήν την παράγραφο θα καθορίσουμε το ασταθές σύνολο των collocation points για το οποίο οι ιδιοτιμές του $A_1 A_0^{-1}$ έχουν θετικά και αρνητικά πραγματικά μέρη, καθώς και τα δυο σταθερά σύνολα των collocation points στα οποία το πραγματικό μέρος των ιδιοτιμών είναι θετικό ή αρνητικό.

Θεωρούμε δυο ιδιοτιμές του πίνακα $A_1 A_0^{-1}$ για συνοριακές συνθήκες Dirichlet, που ικανοποιούν την ακόλουθη τετραγωνική εξίσωση, όπως αυτή προκύπτει από την σχέση (2.6),

$$\det(C_1^D(\lambda)) = h^{-2}(\sigma_1 - \sigma_2)(\gamma_0 + \gamma_1 h\lambda + \gamma_2 h^2 \lambda^2) \tag{2.28}$$

όπου

$$\begin{aligned}\gamma_0 =& 6\sigma_1\sigma_2 + 2 - 3(\sigma_1 + \sigma_2)\\ \gamma_1 =& (1 - \sigma_1 - \sigma_2)(2\sigma_1\sigma_2 - \sigma_1 - \sigma_2)\\ \gamma_2 =& \sigma_1\sigma_2(1 - \sigma_1)(1 - \sigma_2) > 0.\end{aligned} \tag{2.29}$$

Θεωρούμε την διακρίνουσα της σχέσης ως προς $h\lambda$,

$$\Delta^D(\sigma_1, \sigma_2) = \gamma_1^2 - 4\gamma_0\gamma_2 \tag{2.30}$$

και χρησιμοποιώντας τις σχέσεις (2.29) έχουμε,

$$\begin{aligned}\Delta^D(\sigma_1, \sigma_2) =& (1 + 2\sigma_1\sigma_2 - \sigma_1 - \sigma_2)(2\sigma_1^3\sigma_2 + 2\sigma_1\sigma_2^3 - 8\sigma_1^2\sigma_2^2\\ &- \sigma_1^3 - \sigma_2^3 + 5\sigma_1^2\sigma_2 + 5\sigma_1\sigma_2^2 - 6\sigma_1\sigma_2 + \sigma_1^2 + \sigma_2^2).\end{aligned} \tag{2.31}$$

Εάν το $\gamma_0 < 0$ τότε η διακρίνουσα $\Delta^D(\sigma_1, \sigma_2) > 0$, αφου $\gamma_2 > 0$, και η εξίσωση (2.28) θα εχει δυο ριζες.

$$\begin{aligned}h\lambda_1 =& \frac{-\gamma_1 + \sqrt{\gamma_1^2 - 4\gamma_0\gamma_2}}{2\gamma_2}\\ h\lambda_2 =& \frac{-\gamma_1 - \sqrt{\gamma_1^2 - 4\gamma_0\gamma_2}}{2\gamma_2}\end{aligned}$$

Αφού,

$$\gamma_2 > 0, \gamma_0 < 0 \Rightarrow \left|\gamma_1\right| < \sqrt{\gamma_1^2 - 4\gamma_0\gamma_2}$$

άρα εάν,

$$\gamma_1 > 0 \Rightarrow \frac{-\gamma_1 + \sqrt{\gamma_1^2 - 4\gamma_0\gamma_2}}{2\gamma_2} > 0$$

$$\gamma_1 > 0 \Rightarrow \frac{-\gamma_1 - \sqrt{\gamma_1^2 - 4\gamma_0\gamma_2}}{2\gamma_2} < 0$$

$$\gamma_1 < 0 \Rightarrow \frac{-\gamma_1 - \sqrt{\gamma_1^2 - 4\gamma_0\gamma_2}}{2\gamma_2} < 0$$

$$\gamma_1 < 0 \Rightarrow \frac{-\gamma_1 + \sqrt{\gamma_1^2 - 4\gamma_0\gamma_2}}{2\gamma_2} > 0$$

οπότε η εξίσωση (2.28), θα έχει σε κάθε περίπτωση δυο ρίζες με διαφορετικά πρόσημα.

Εάν $\gamma_0 > 0$ και $\Delta^D(\sigma_1, \sigma_2) < 0$ τότε θα έχουμε δυο μιγαδικές ρίζες,

$$\begin{aligned} h\lambda_1 =& \frac{-\gamma_1 + i\sqrt{4\gamma_0\gamma_2 - \gamma_1^2}}{2\gamma_2} \\ h\lambda_2 =& \frac{-\gamma_1 - i\sqrt{4\gamma_0\gamma_2 - \gamma_1^2}}{2\gamma_2}. \end{aligned}$$

Σε αυτήν την περίπτωση, εάν $\gamma_1 > 0$ τότε έχουμε δυο μιγαδικές ρίζες με αρνητικά πραγματικά μέρη, ενώ εάν $\gamma_1 < 0$ τότε έχουν θετικά πραγματικά μέρη. Σε κάθε περίπτωση τα πραγματικά τους μέρη έχουν το ίδιο πρόσημο.

Εάν $\gamma_0 > 0$ και $\Delta^D(\sigma_1, \sigma_2) \geq 0$ τότε θα έχουμε δυο πραγματικές ρίζες.

$$\begin{aligned} h\lambda_1 =& \frac{-\gamma_1 + \sqrt{\gamma_1^2 - 4\gamma_0\gamma_2}}{2\gamma_2} \\ h\lambda_2 =& \frac{-\gamma_1 - \sqrt{\gamma_1^2 - 4\gamma_0\gamma_2}}{2\gamma_2} \end{aligned}$$

Αφού,

$$\gamma_2 > 0, \gamma_0 > 0 \Rightarrow \left|\gamma_1\right| > \sqrt{\gamma_1^2 - 4\gamma_0\gamma_2}$$

άρα εάν,

$$\gamma_1 > 0 \Rightarrow \frac{-\gamma_1 + \sqrt{\gamma_1^2 - 4\gamma_0\gamma_2}}{2\gamma_2} < 0$$

$$\gamma_1 > 0 \Rightarrow \frac{-\gamma_1 - \sqrt{\gamma_1^2 - 4\gamma_0\gamma_2}}{2\gamma_2} < 0$$

$$\gamma_1 < 0 \Rightarrow \frac{-\gamma_1 - \sqrt{\gamma_1^2 - 4\gamma_0\gamma_2}}{2\gamma_2} > 0$$
$$\gamma_1 < 0 \Rightarrow \frac{-\gamma_1 + \sqrt{\gamma_1^2 - 4\gamma_0\gamma_2}}{2\gamma_2} > 0$$

όποτε η εξίσωση (2.28), θα έχει σε κάθε περίπτωση δυο πραγματικές ρίζες με το ίδιο πρόσημο. Οι υπόλοιπες ιδιοτιμές του $A_1A_0^{-1}$ καθορίζονται από την σχέση,

$$\det(tridiag(a_1(\lambda), b_1(\lambda), c_1(\lambda))) = \prod_{j=1}^{n-1} \Big(b_1(\lambda) + 2\sqrt{a_1(\lambda)c_1(\lambda)}\cos\frac{j\pi}{n}\Big) = 0 \tag{2.32}$$

Οπότε είμαστε έτοιμοι να ορίσουμε την περιοχή των collocation points για την οποία προβλήματα της μορφής (2.1) δεν επιλύονται ικανοποιητικά,

$$U^D = \Big\{(\sigma_1, \sigma_2) \in \Omega_c \;\Big|\; \gamma_0 < 0\Big\}. \tag{2.33}$$

Τα σταθερά σύνολα των collocation points ορίζονται ως εξής,

$$\begin{aligned} S_1^{D^+} =& \Big\{(\sigma_1, \sigma_2) \in \Omega_c \;\Big|\; Re\big(\lambda(A_1A_0^{-1}) > 0\big)\Big\} \\ S_2^{D^-} =& \Big\{(\sigma_1, \sigma_2) \in \Omega_c \;\Big|\; Re\big(\lambda(A_1A_0^{-1}) < 0\big)\Big\} \end{aligned} \tag{2.34}$$

Το πραγματικό μέρος των ιδιοτιμών όπως προκύπτει από την προηγούμενη ανάλυση είναι θετικό όταν $\gamma_0 > 0$ και $\gamma_1 < 0$ και αρνητικό όταν $\gamma_0 > 0$ και $\gamma_1 > 0$. Επίσης το γ_1 από την σχέση (2.29) μπορεί να γράφει,

$$\gamma_1 = (1 - \sigma_1 - \sigma_2)(\sigma_1(\sigma_2 - 1) + \sigma_2(\sigma_1 - 1))$$

όμως

$$(\sigma_1(\sigma_2 - 1) + \sigma_2(\sigma_1 - 1)) < 0$$

οπότε

$$\gamma_1 > 0 \Leftrightarrow 1 - \sigma_1 - \sigma_2 < 0$$

$$\gamma_1 < 0 \Leftrightarrow 1 - \sigma_1 - \sigma_2 > 0.$$

Συνεπώς

$$\begin{aligned} S_1^{D^+} =& \Big\{ (\sigma_1, \sigma_2) \in \Omega_c \;\Big|\; \gamma_0 > 0 \;,\; \sigma_1 + \sigma_2 < 1 \Big\} \\ S_2^{D^-} =& \Big\{ (\sigma_1, \sigma_2) \in \Omega_c \;\Big|\; \gamma_0 > 0 \;,\; \sigma_1 + \sigma_2 > 1. \Big\}. \end{aligned} \tag{2.35}$$

Τα σύνολα παριστάνονται γραφικά, σχήμα 2.1. Αριθμητικά έλεγχοι για την επίλυση των αλγεβρικών εξισώσεων,

$$\det(tridiag(a_1(\lambda), b_1(\lambda), c_1(\lambda))) = b_1(\lambda) + 2\sqrt{a_1(\lambda)c_1(\lambda)}p = 0$$

με $-1 < p < 1$ επιβεβαιώνουν τα συμπεράσματα του σχήματος 2.1 και είναι αληθές για όλες τις ιδιοτιμές του $A_1 A_0^{-1}$ αλλά δεν μπορούμε να δώσουμε μια αναλυτική απόδειξη.

Για συνοριακές συνθήκες τύπου Neumann από την σχέση (2.14) θα έχουμε,

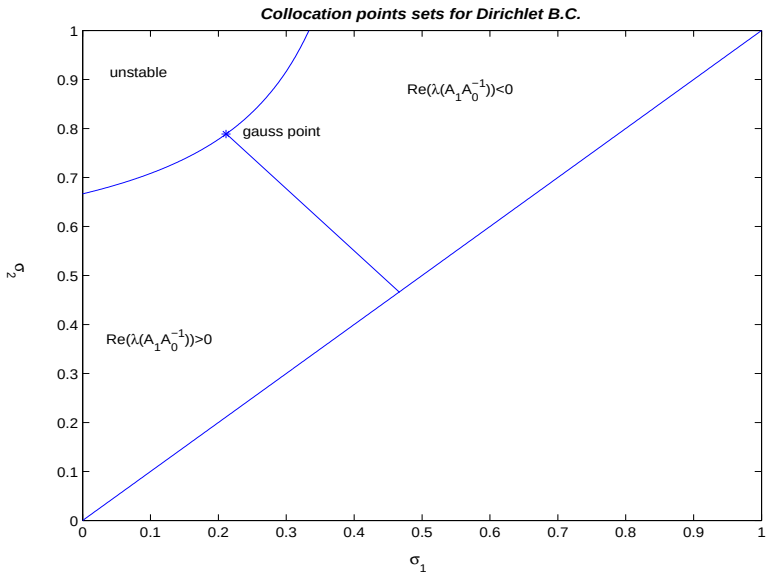
$$\det(C_1^N(\lambda)) = h^{-2}(\sigma_1 - \sigma_2)(\gamma_2' h^2 \lambda^2 + \gamma_1' h \lambda) \tag{2.36}$$

με

$$\begin{aligned} \gamma_2' =& -\big((1-\sigma_2)(2\sigma_2 + \sigma_1) + (1-\sigma_1)(2\sigma_1 + \sigma_2)\big) < 0 \\ \gamma_1' =& \quad 6(1 - \sigma_1 - \sigma_2). \end{aligned} \tag{2.37}$$

Στην περίπτωση αυτή η διακρίνουσα θα είναι,

$$\Delta^N(\sigma_1, \sigma_2) = \gamma_1^2 \geq 0. \tag{2.38}$$



Σχήμα 2.1: Τα σύνολα των collocation points για συνοριακές συνθήκες τύπου Dirichlet.

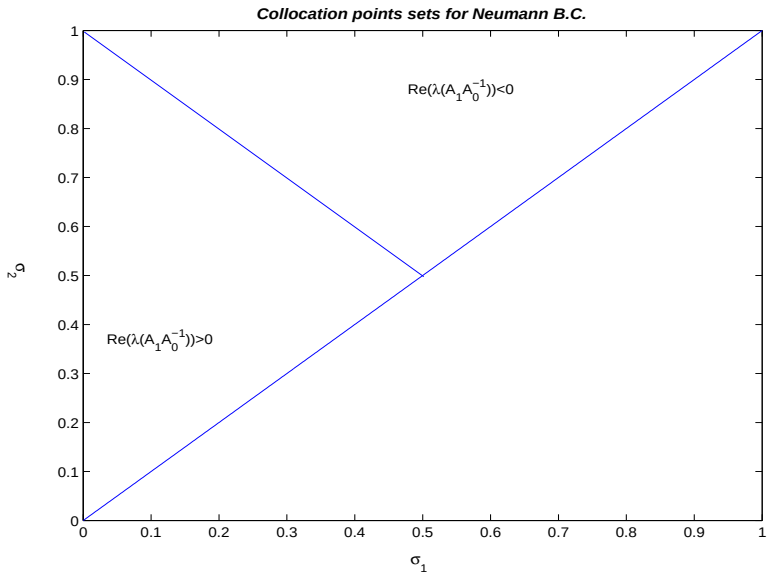
Οπότε θα έχουμε δυο πραγματικές ρίζες,

$$h\lambda_1 = \frac{-\gamma_1' + \sqrt{\gamma_1'^2}}{2\gamma_2'}$$
$$h\lambda_2 = \frac{-\gamma_1' - \sqrt{\gamma_1'^2}}{2\gamma_2'}.$$

Εάν $\gamma_1' > 0$ έχουμε μια μηδενική ιδιοτιμή και μια με θετικό πρόσημο διότι $\gamma_2' < 0$. Αντίθετα, εάν $\gamma_1 < 0$ τότε έχουμε μια μηδενική και μια με αρνητικό πρόσημο. Οπότε τα σταθερά σύνολα για τα collocation points στην περίπτωση συνοριακών συνθηκών τύπου Neumann αναλόγως με προηγουμένως θα είναι,

$$\begin{aligned} S_1^{N+} &= \Big\{ (\sigma_1, \sigma_2) \in \Omega \;\Big|\; \sigma_1 + \sigma_2 < 1 \Big\} \\ S_2^{N-} &= \Big\{ (\sigma_1, \sigma_2) \in \Omega \;\Big|\; \sigma_1 + \sigma_2 > 1 \Big\}. \end{aligned} \tag{2.39}$$

και παριστάνονται γραφικά στο σχήμα (2.2).



Σχήμα 2.2: Τα σύνολα των collocation points για συνοριακές συνθήκες τύπου Neumann.

2.3 Αριθμητικά Αποτελέσματα

Στην ενότητα αυτή θα παρουσιάσουμε τα αριθμητικά αποτελέσματα για δυο προβλήματα συνοριακών τιμών, επιβεβαιώνοντας την παράπανω θεωρητική ανάλυση.

Παράδειγμα 1 Θεωρούμε το πρόβλημα συνοριακών τιμών,

$$-\epsilon u_{xx} + u_x = 1 \qquad 0 < x < 1 \tag{2.40}$$

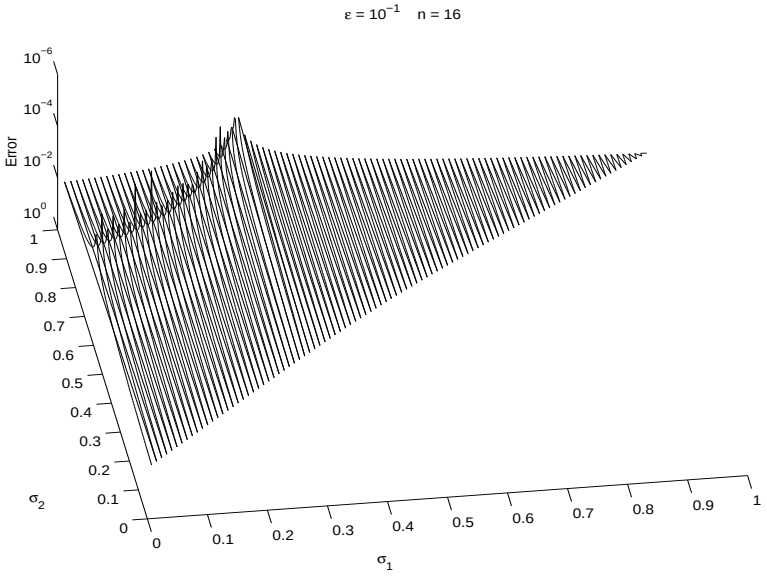
$$u(0) = u(1) = 0$$

το οποίο έχει αναλυτική λύση,

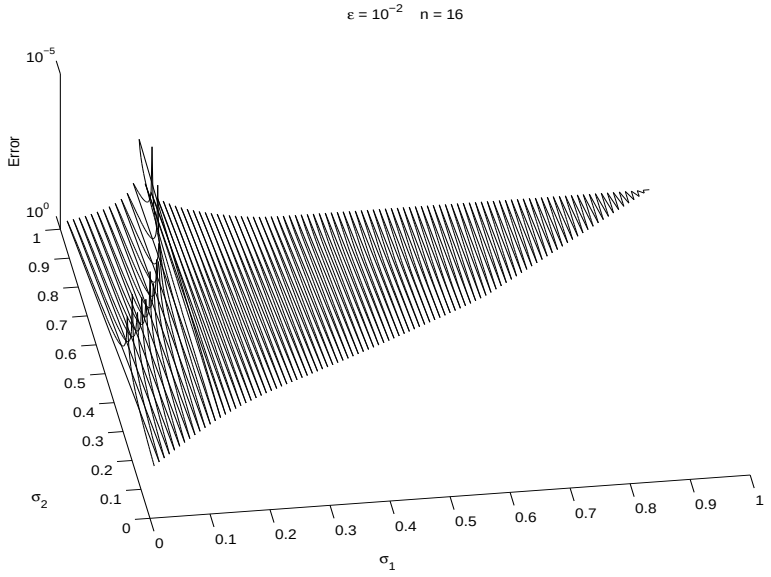
$$u(x) = \frac{1 - e^{\frac{x}{\epsilon}}}{e^{\frac{1}{\epsilon}} - 1} + x.$$

Για $0 < \epsilon \ll 1$ παρουσιάζεται ένα συνοριακό στρώμα στο $x = 1$. Στην συνέχεια θα παρουσιάσουμε τα επίπεδα σφάλματος $\left(\sigma_1, \sigma_2, e = \max_{1<i<n} |u(x_{i+1}) -\right.$

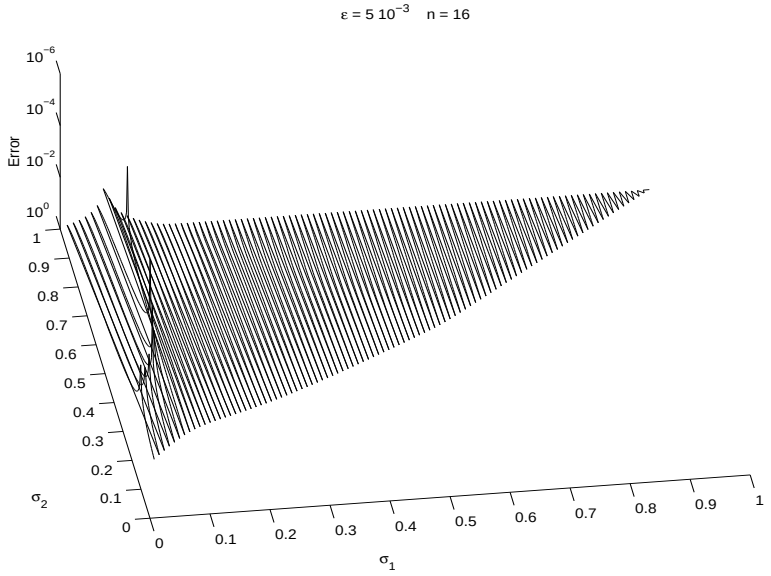
$u(x_i)|\big)$ για διάφορες τιμές του ϵ, καθώς για διάφορες διαμερίσεις n, επαληθεύοντας έτσι την παραπάνω θεωρητική ανάλυση. Επίσης κάποια νέα upwinding χαρακτηριστικά θα αναγνωριστούν, γεννώντας νέα, πολύ σημαντικά ερωτήματα.



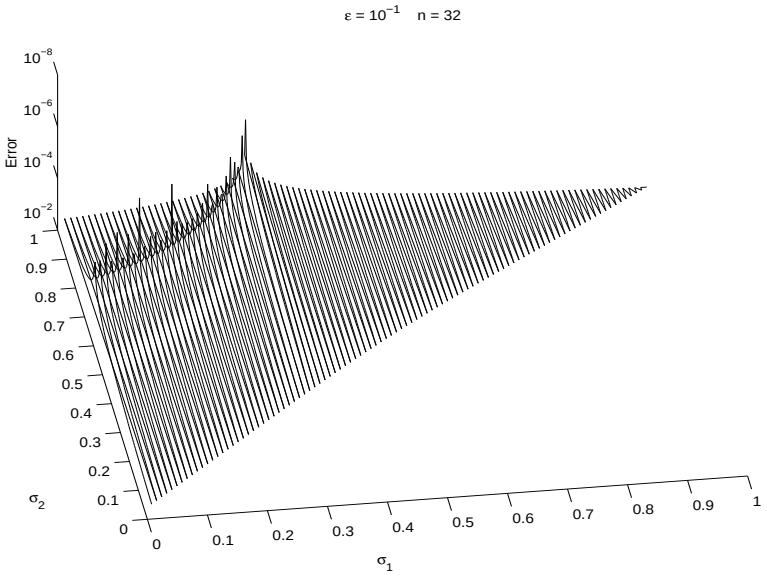
Σχήμα 2.3: Το επίπεδο (σ_1, σ_2, e) για το πρόβλημα (2.40) με $\epsilon = 10^{-1}$ και $n = 16$.



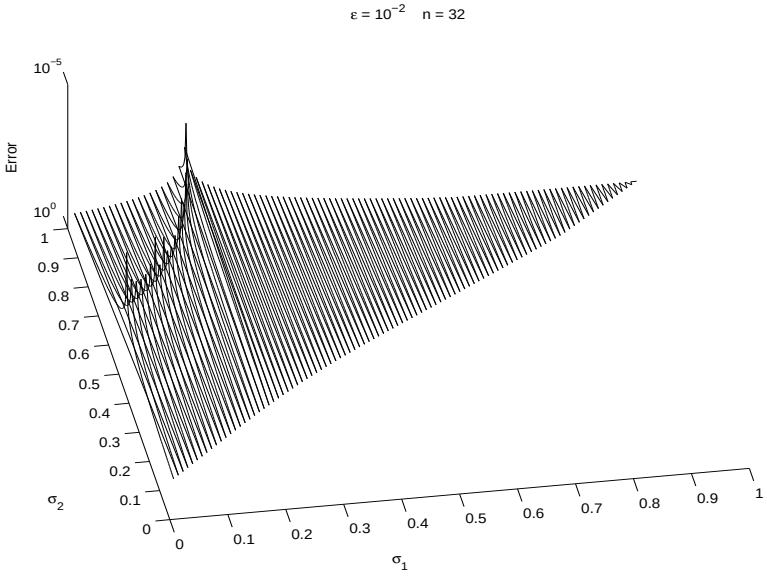
Σχήμα 2.4: Το επίπεδο (σ_1, σ_2, e) για το πρόβλημα (2.40) με $\epsilon = 10^{-2}$ και $n = 16$.



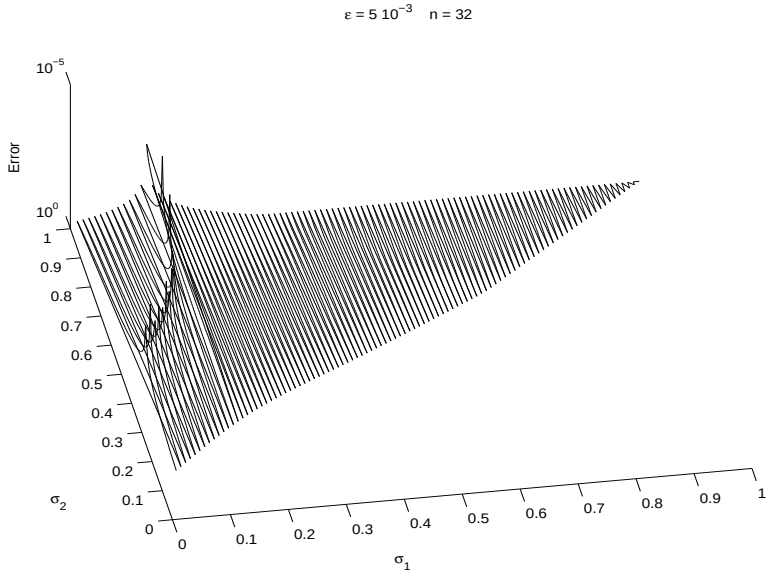
Σχήμα 2.5: Το επίπεδο (σ_1, σ_2, e) για το πρόβλημα (2.40) με $\epsilon = 5 \cdot 10^{-3}$ και $n = 16$.



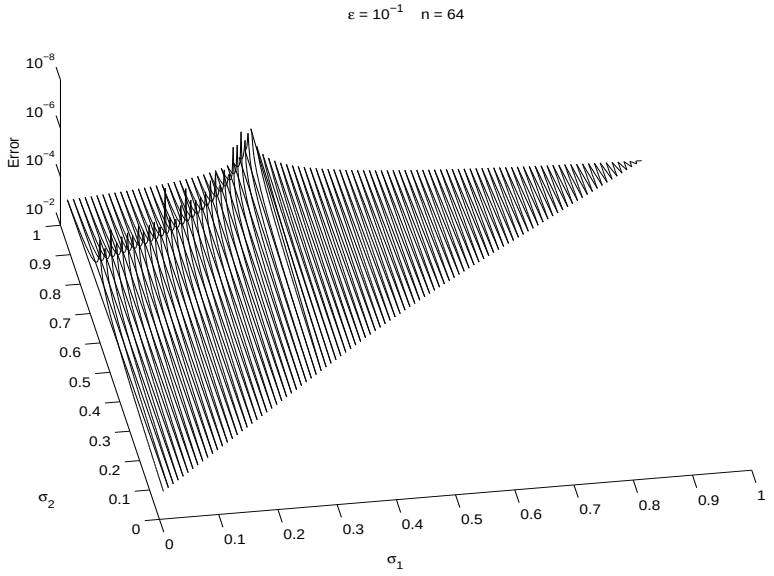
Σχήμα 2.6: Το επίπεδο (σ_1, σ_2, e) για το πρόβλημα (2.40) με $\epsilon = 10^{-1}$ και $n = 32$.



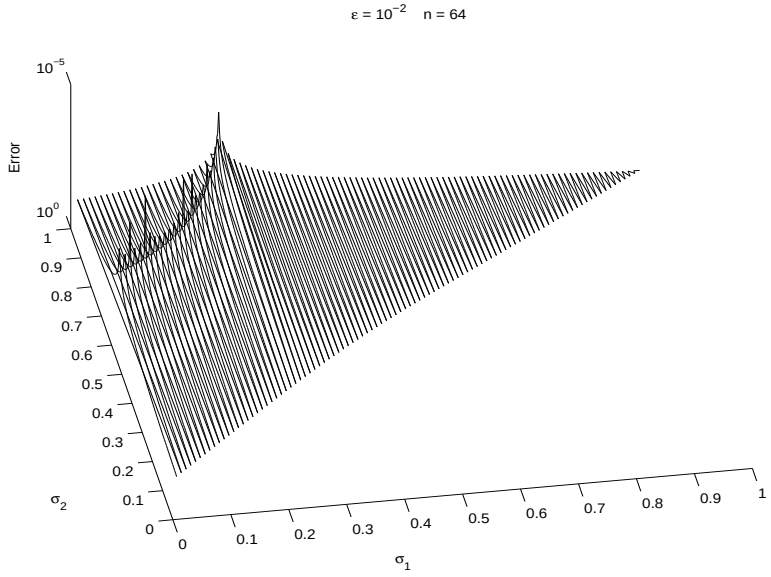
Σχήμα 2.7: Το επίπεδο (σ_1, σ_2, e) για το πρόβλημα (2.40) με $\epsilon = 10^{-2}$ και $n = 32$.



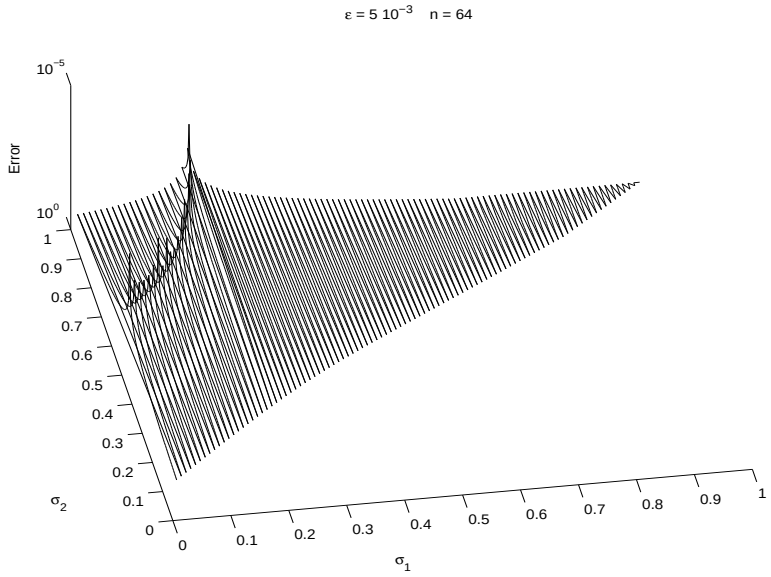
Σχήμα 2.8: Το επίπεδο (σ_1, σ_2, e) για το πρόβλημα (2.40) με $\epsilon = 5 \cdot 10^{-3}$ και $n = 32$.



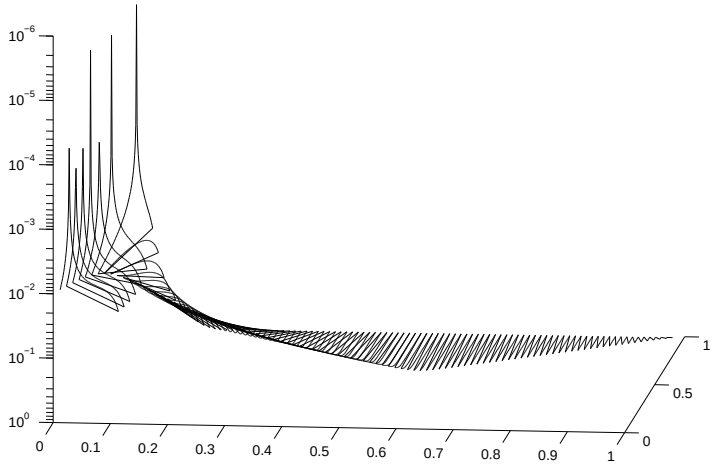
Σχήμα 2.9: Το επίπεδο (σ_1, σ_2, e) για το πρόβλημα (2.40) με $\epsilon = 10^{-1}$ και $n = 64$.



Σχήμα 2.10: Το επίπεδο (σ_1, σ_2, e) για το πρόβλημα (2.40) με $\epsilon = 10^{-2}$ και $n = 64$.



Σχήμα 2.11: Το επίπεδο (σ_1, σ_2, e) για το πρόβλημα (2.40) με $\epsilon = 5 \cdot 10^{-3}$ και $n = 64$.



Σχήμα 2.12: Το επίπεδο (σ_1, σ_2, e) για το πρόβλημα (2.40) με $\epsilon = 5 \cdot 10^{-3}$ και $n = 16$ από διαφορετική οπτική γωνία.

Παρατηρήσεις

- Από το σχήμα (2.12) επιβεβαιώνεται η θεωρητική ανάλυση της παραγράφου αφού το σφάλμα αρχίζει να ελαχιστοποιείται όταν τα collocation points βρίσκονται στο σύνολο ${S_1}^{D^+}$. Επίσης μια καμπύλη εμφανίζεται στο ${S_1}^{D^+}$ στην οποία το σφάλμα πάνω στους κόμβους ελαχιστοποιείται. Η καμπύλη αυτή διέρχεται και μέσα από την μη-σταθερή περιοχή U^D. Όπως θα δείξουμε παρακάτω μπορεί η καμπύλη με το ελάχιστο σφάλμα στους κόμβους να διέρχεται από την U^D αλλά η collocation λύσεις παρουσιάζουν ταλαντώσεις σε αυτήν την περιοχή σχήματα (2.14),(2.17). Ταλαντώσεις βεβαίως δεν εμφανίζονται στην καμπύλη εντός της περιοχής ${S_1}^{D^+}$.

- Η καμπύλη εντός της περιοχής ${S_1}^{D^+}$, καθορίζει το βέλτιστο upwind σχήμα, καθώς $\epsilon \to 0$, το οποίο δεν προσδίδει ταλαντώσεις στην προσεγγιστική λύση, σχήματα(2.15),(2.18).

- Παρατηρούμε ότι καθώς το ϵ μικραίνει για μια συγκεκριμένη διαμέριση π.χ. $n = 16$, σχήματα (2.3),(2.4),(2.5), η καμπύλη των collocation points μετατοπίζεται πλησιάζοντας όλο και περισσότερο των άξονα σ_2. Δηλαδή όσο το ϵ γίνεται μικρότερο τα collocation points (σ_1, σ_2) που δίνουν το

ελάχιστο σφάλμα, παίρνουν συνεχώς μικρότερες τιμές, προσδίδοντας στο σχήμα συνεχώς μεγαλύτερη ασυμμετρία.

- Επίσης παρατηρούμε ότι η πληροφορία που δίνουμε στο αριθμητικό μας σχήμα από τα collocation points, χρησιμοποιώντας βέλτιστο upwinding, είναι πλησιέστερα στο κόμβο x_i όσο ο συντελεστής διάχυσης μικραίνει για το element $[x_i, x_{i+1}]$, που συμφωνεί απόλυτα με το κλασσικό upwind σχήμα των πεπερασμένων διάφορων. Όταν ο συντελεστής advection είναι αρνητικός, όπως θα δούμε στο επόμενο παράδειγμα η καμπύλη των collocation points με το ελάχιστο σφάλμα κινείται αντίστροφα αναζητώντας πληροφορία κοντά στον κόμβο x_{i+1}.

- Όταν το μέγεθος της διαμέρισης αυξάνεται για συγκεκριμένη τιμή του ϵ παρατηρούμε το σχήμα παύει να αναζητά συνεχώς μια μη-συμμετρική πληροφορία. Πιο συγκεκριμένα αναφέρουμε, ότι οι καμπύλες των collocation points με το ελάχιστο σφάλμα τείνουν να ταυτιστούν με την καμπύλη $\gamma_0 = 0$, η οποία περιέχει και το gauss point, σχήματα (2.3 εώς 2.11). Στην καμπύλη αυτή για ικανοποιητικά ομαλές λύσεις έχουμε ελάχιστο σφάλμα.

- Η τάξη σύγκλισης του βέλτιστου upwind hermite collocation σχήματος εξαρταται από το μέγεθος της διαμέρισης, από τον λόγο του συντελεστή διάχυσης προς τον συντελεστή της ταχύτητας καθώς και από το μήκος του διαστήματος $I = [a, b]$ όπου επιλύεται η διαφορική εξίσωση.

  Αριθμητικές μας μετρήσεις δείχνουν ότι καθώς το $\epsilon \to 0$, το βέλτιστο upwind σχήμα μπορεί να δώσει σφάλματα τάξεως ως και $O(h^4)$. Απαραίτητη προϋπόθεση για να έχουμε σφάλματα τάξης $O(h^4)$ το γινόμενο $\frac{\epsilon}{p}n(b-a)$ να είναι αρκετά μικρό, όπου σε αυτήν την περίπτωση η orthogonal collocation δεν δίνει ικανοποιητικά αποτελέσματα προσδίδοντας ταλαντώσεις στην προσεγγιστική λύση. Αριθμητικές μετρήσεις για το πρόβλημα (2.1), δείχνουν ότι αν ο αριθμός $\frac{\epsilon}{p}n(b-a) < 0.2$, τότε το βέλτιστο upwind σχήμα είναι ικανό να δώσει σφάλματα τάξεως ως $O(h^4)$. Για $0.2 < \frac{\epsilon}{p}n(b-a) < 0.5$ το βέλτιστο upwind σχήμα είναι ικανό να δώσει σφάλματα τάξεως ως $O(h^3)$. Εν γένει όλα τα σημεία πάνω στην καμπύλη ελάχιστου σφάλματος του παρουσιάζουν σφάλματα άνω της τάξης $O(h^2)$, για $\frac{\epsilon}{p}n(b-a) < 0.5$. Παρατηρούμε ότι όταν το $\frac{\epsilon}{p}$, ή το n αυξάνεται, τότε ο αριθμός $\frac{\epsilon}{p}n(b-a)$ μεγαλώνει με αποτέλεσμα το upwind σχήμα να δίνει σφάλματα μικρότερης τάξης. Αυτό θεωρείται απολύτως λογικό,

διότι όταν $\frac{\epsilon}{p}$ (π.χ. για το πρόβλημα (2.1) με $\frac{\epsilon}{p} > 0.5$) είναι μεγάλο τότε η λύση είναι ομαλότερη ως προς την διαμέριση, οπότε ένα συμμετρικό σχήμα όπως η orthogonal collocation θα προτιμάτε.

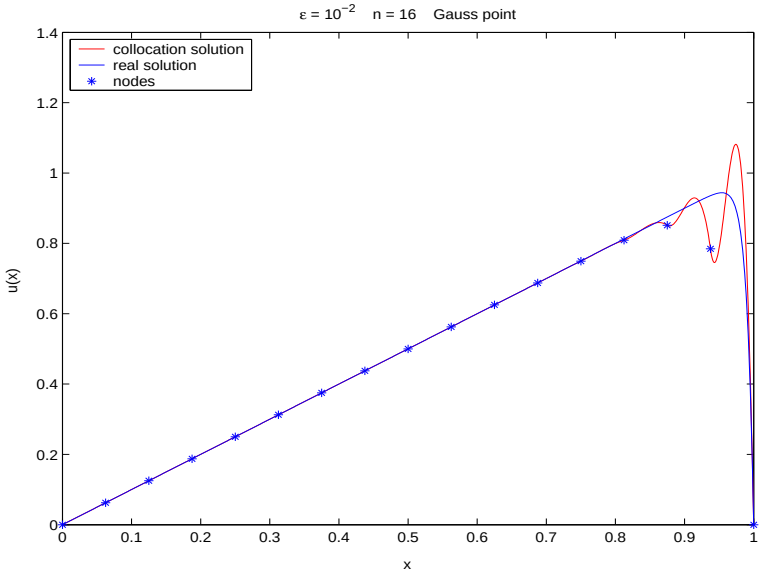
$\epsilon = 10^{-2}, p = 1$	$n = 16, \frac{\epsilon}{p}n = 0.16$	$n = 32, \frac{\epsilon}{p}n = 0.32$	$n = 64, \frac{\epsilon}{p}n = 0.64$
$(\sigma_1, \sigma_2) \in S_1^{D^+}$	$(0.06, 0.40)$	$(0.07, 0.54)$	$(0.15, 0.72)$
Error	$7\ 10^{-5}\ \ O(h^4)$	$5\ 10^{-5}\ \ O(h^3)$	$8\ 10^{-5}\ \ O(h^2)$
gauss point	(σ_1^G, σ_2^G)	(σ_1^G, σ_2^G)	(σ_1^G, σ_2^G)
Error	$1.5\ 10^{-1}\ \ O(h)$	$3\ 10^{-2}\ \ O(h)$	$7\ 10^{-4}\ \ O(h^2)$

- Επιλέγοντας collocation point (σ_1, σ_2) πάνω στην καμπύλη ελάχιστου σφάλματος, με $\sigma_1 < \sigma_1^G$ και $\sigma_2 < \sigma_2^G$ [1] παρατηρούμε ότι έχουμε βέλτιστα upwind αποτελέσματα. Θεωρητικά δεν μπορούμε να το στοιχειοθετήσουμε το παραπάνω συμπέρασμα.

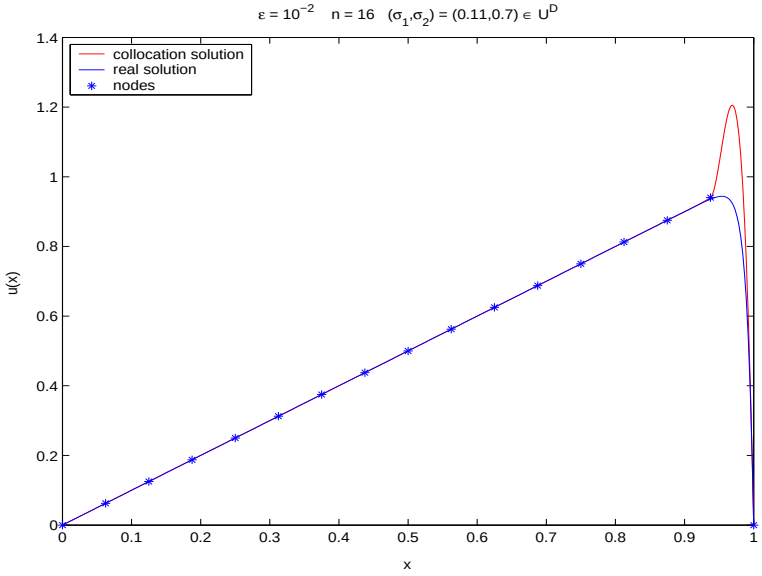
- Όπως μπορούμε να εκτιμήσουμε από τα παραπάνω, οι καμπύλες ελάχιστου σφάλματος η οποία προσδίδει την βέλτιστη upwind collocation εξαρτώνται από τον λόγο του συντελεστή διάχυσης προς τον συντελεστή ταχύτητας $\frac{\epsilon}{p}$, καθώς και από το μέγεθος της διαμέρισης n. Επίσης υπάρχει εξάρτηση του πρόσημου του συντελεστή advection $sign(p)$, που καθορίζει σε πιο από τα δυο σύνολα ${S_1}^{D^+}$, ${S_2}^{D^-}$ εντοπίζεται η καμπύλη κάθε φορά. Το κύριο ερώτημα που γεννιέται από αυτό το κεφάλαιο είναι ο καθορισμός αυτής της "βέλτιστης "καμπύλης. **Θεωρούμε ότι είναι ένα πάρα πολύ σημαντικό αντικείμενο μελλοντικής μελέτης**.

Παρακάτω παριστάνουμε γραφικά την προσεγγιστική λύση με collocation points να είναι αντίστοιχα gauss point, $(\sigma_1, \sigma_2) \in U^D$, $(\sigma_1, \sigma_2) \in {S_1}^{D^+}$, καθώς και τις ιδιοτιμές των πινάκων $A_1 A_0^{-1}$ για gauss point και $(\sigma_1, \sigma_2) \in {S_1}^{D^+}$, για διάφορες τιμές του ϵ.

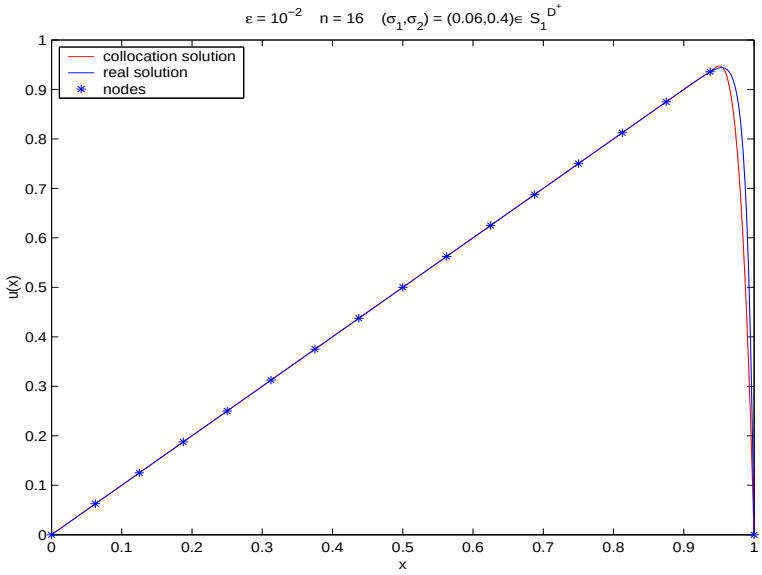
[1] Οπου $(\sigma_1^G, \sigma_2^G) = (\frac{1}{2} - \frac{1}{2\sqrt{3}}, \frac{1}{2} - \frac{1}{2\sqrt{3}}) = gauss\ \ point.$



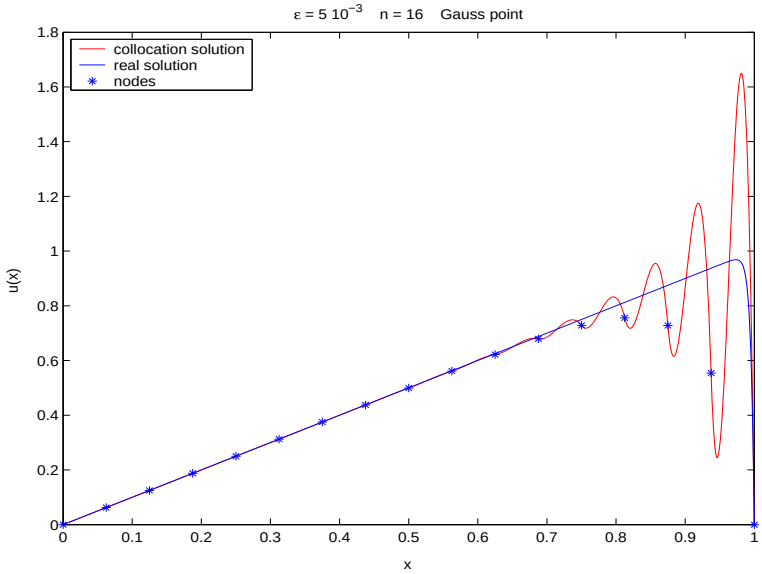
Σχήμα 2.13: Προσεγγιστική και πραγματική λύση για $\epsilon = 10^{-2}$, $n = 16$ με $(\sigma_1, \sigma_2) = \text{gauss point}$.



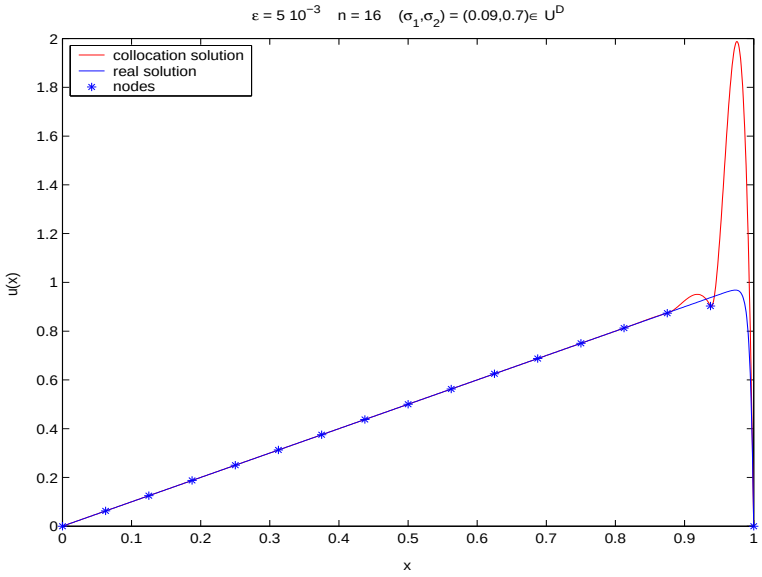
Σχήμα 2.14: Προσεγγιστική και πραγματική λύση για $\epsilon = 10^{-2}$, $n = 16$ με $(\sigma_1, \sigma_2) = (0.11, 0.7) \in U^D$.



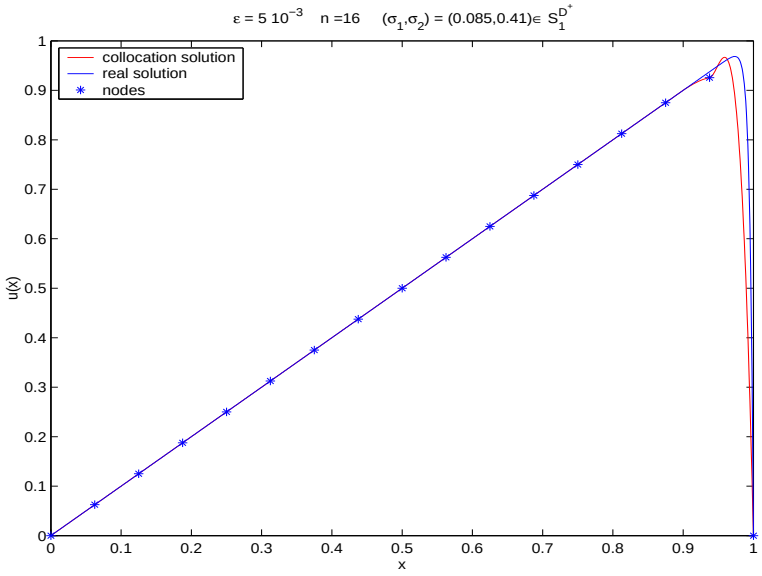
Σχήμα 2.15: Προσεγγιστική και πραγματική λύση για $\epsilon = 10^{-2}$, $n = 16$ με $(\sigma_1, \sigma_2) = (0.06, 0.4) \in S_1^{D+}$.



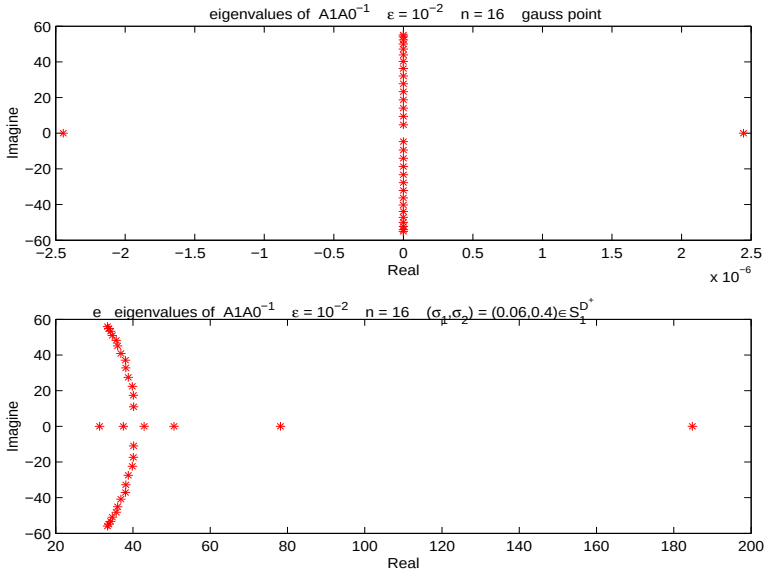
Σχήμα 2.16: Προσεγγιστική και πραγματική λύση για $\epsilon = 5 \cdot 10^{-3}$, $n = 16$ με $(\sigma_1, \sigma_2) =$ gauss point.



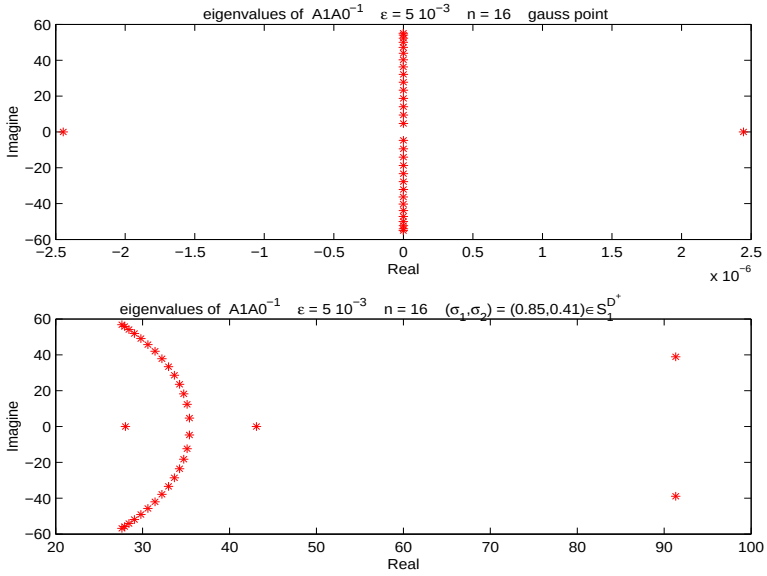
Σχήμα 2.17: Προσεγγιστική και πραγματική λύση για $\epsilon = 5 \cdot 10^{-3}$, $n = 16$ με $(\sigma_1, \sigma_2) = (0.09, 0.7) \in U^D$.



Σχήμα 2.18: Προσεγγιστική και πραγματική λύση για $\epsilon = 5 \cdot 10^{-3}$, $n = 16$ με $(\sigma_1, \sigma_2) = (0.085, 0.41) \in S_1^{D+}$.



Σχήμα 2.19: Οι ιδιοτιμές του πίνακα $A_1 A_0^{-1}$, για $\epsilon = 10^{-2}$, $n = 16$ με $(\sigma_1, \sigma_2) =$ gauss point και $(\sigma_1, \sigma_2) = (0.06, 0.4)$ αντίστοιχα.



Σχήμα 2.20: Οι ιδιοτιμές του πίνακα $A_1 A_0^{-1}$, για $\epsilon = 5 \cdot 10^{-3}$, $n = 16$ με $(\sigma_1, \sigma_2) =$ gauss point και $(\sigma_1, \sigma_2) = (0.085, 0.41)$ αντίστοιχα.

Παράδειγμα 2 Θεωρούμε το πρόβλημα συνοριακών τιμών,

$$-\epsilon u_{xx} - u_x = -1 \qquad 0 < x < 1 \tag{2.41}$$

$$u(0) = u(1) = 0$$

το οποίο έχει αναλυτική λύση,

$$u(x) = \frac{(e^{\frac{x}{\epsilon}} - 1)e^{\frac{1}{\epsilon}}}{e^{\frac{1}{\epsilon}} - 1} + x.$$

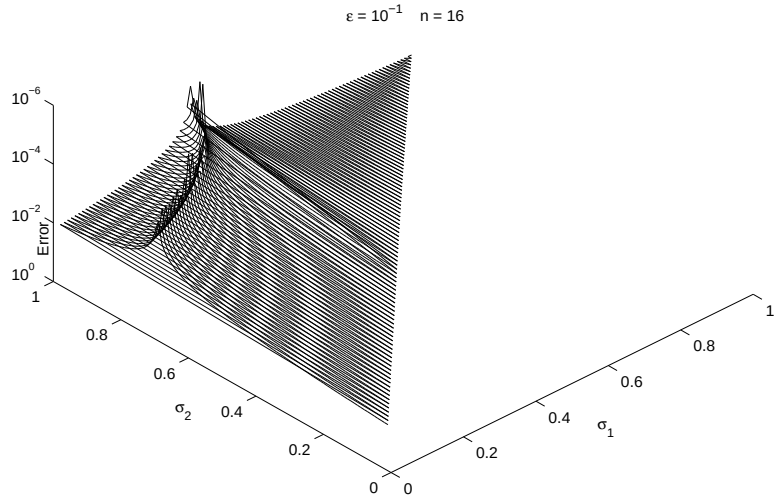
Σε αυτήν την περίπτωση ο συντελεστής advection έχει αρνητικό πρόσημο. Οπότε τα upwinding χαρακτηριστικά όπως ήδη προαναφέραμε, είναι αντίθετα σε σχέση με την προηγούμενη περίπτωση όπου ο συντελεστής της ταχύτητας ήταν θετικός. Συνοπτικά αναφέρουμε.

- Για $0 < \epsilon \ll 1$ παρουσιάζεται ένα συνοριακό στρώμα στο $x = 0$.
- Η καμπύλη που ελαχιστοποιεί το σφάλμα στους κόμβους διέρχεται από την ${S_2}^{D^-}$ και U^D. Όμως στην περιοχή U^D, παρόλο που έχουμε ελάχιστο σφάλμα πάνω στους κόμβους, η προσεγγιστική λύση παρουσιάζει ταλαντώσεις. Βεβαίως μέσα στο σύνολο ${S_2}^{D^-}$, η collocation λύση δεν παρουσιάζει ταλαντώσεις.
- Η τάξη σύγκλισης του βέλτιστου upwind hermite collocation σχήματος εξαρτάται από το μέγεθος της διαμέρισης, από τον λόγο του συντελεστή διάχυσης προς τον συντελεστή της ταχύτητας καθώς και από το μήκος του διαστήματος $I = [a, b]$ όπου επιλύεται η διαφορική εξίσωση. Για αρκετά μικρές τιμές του $\frac{\epsilon}{|p|}n(b-a)$ το βέλτιστο upwind σχήμα είναι ικανό να δώσει σφάλματα τάξεως $O(h^4)$. Εν γένει όλα τα σημεία πάνω στην καμπύλη ελάχιστου σφάλματος, δίνουν σφάλματα τάξεως άνω του $O(h^2)$. Όσο μεγαλώνει ο αριθμός $\frac{\epsilon}{|p|}n(b-a)$ τα σφάλματα του upwind σχήματος είναι μικρότερης τάξης, οπότε και η orthogonal collocation προτιμάτε.
- Επιλέγοντας collocation point (σ_1, σ_2) πάνω στην καμπύλη ελάχιστου σφάλματος, με $\sigma_1 > \sigma_1^G$ και $\sigma_2 > \sigma_2^G$ [2] παρατηρούμε ότι έχουμε βέλτιστα upwind αποτελέσματα. Θεωρητικά δεν μπορούμε να το στοιχειοθετήσουμε το παραπάνω συμπέρασμα.

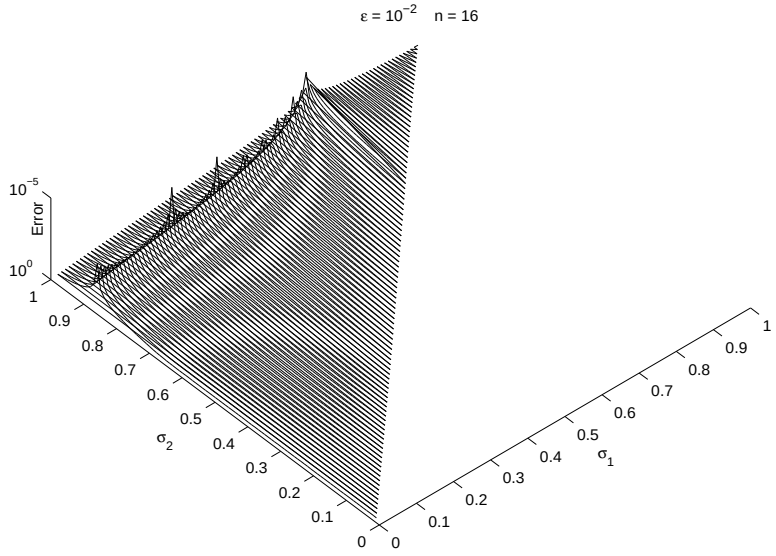
[2] Οπου $(\sigma_1^G, \sigma_2^G) = (\frac{1}{2} - \frac{1}{2\sqrt{3}}, \frac{1}{2} - \frac{1}{2\sqrt{3}}) = gauss\ \ point.$

- Καθώς $\epsilon \to 0$ το upwind σχήμα δεν προσδίδει ταλαντώσεις στην προσεγγιστική λύση.
- Καθώς το ϵ γίνεται μικρό για μια συγκεκριμένη διαμέριση παρατηρούμε ότι τα collocation points παίρνουν συνεχώς μεγαλύτερες τιμές.
- Όσο ο συντελεστής διάχυσης ϵ μικραίνει το upwind σχήμα αναζητά πληροφορία πιο κοντά στον κόμβο x_{i+1} για κάθε element $[x_i, x_{i+1}]$.
- Παρατηρούμε ότι καθώς η διαμέριση μεγαλώνει, η καμπύλη ελάχιστου σφάλματος τείνει να ταυτιστεί με την καμπύλη $\gamma_0 = 0$ πάνω στην οποία βρίσκεται και το gauss point.

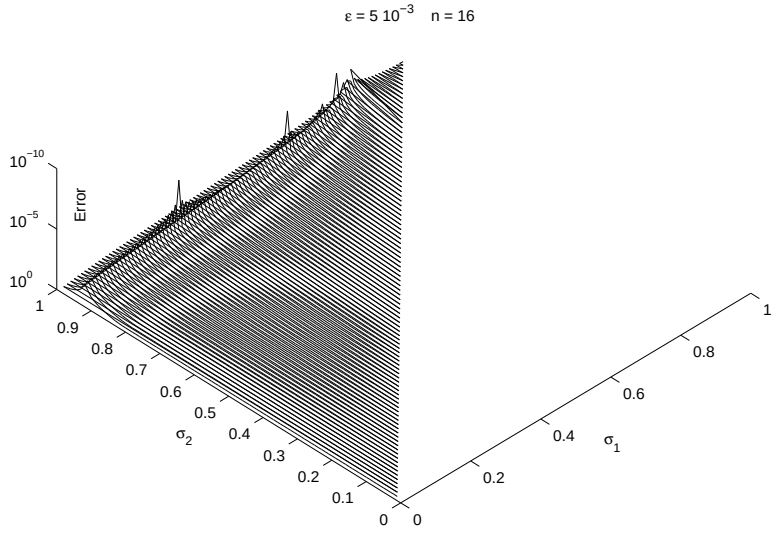
Στην συνέχεια παριστάνουμε γραφικά τα αποτελέσματα μας, όπου στοιχειοθετούν τις παραπάνω παρατηρήσεις μας για αυτήν την περίπτωση.



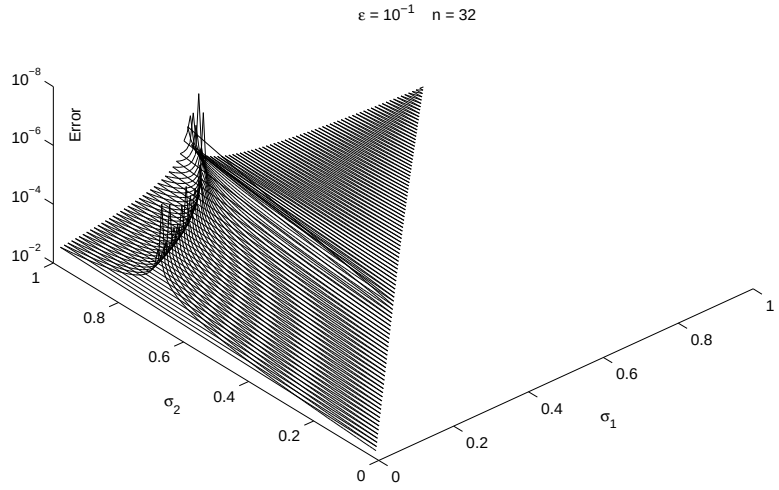
Σχήμα 2.21: Το επίπεδο (σ_1, σ_2, e) για το πρόβλημα (2.41) με $\epsilon = 10^{-1}$ και $n = 16$.



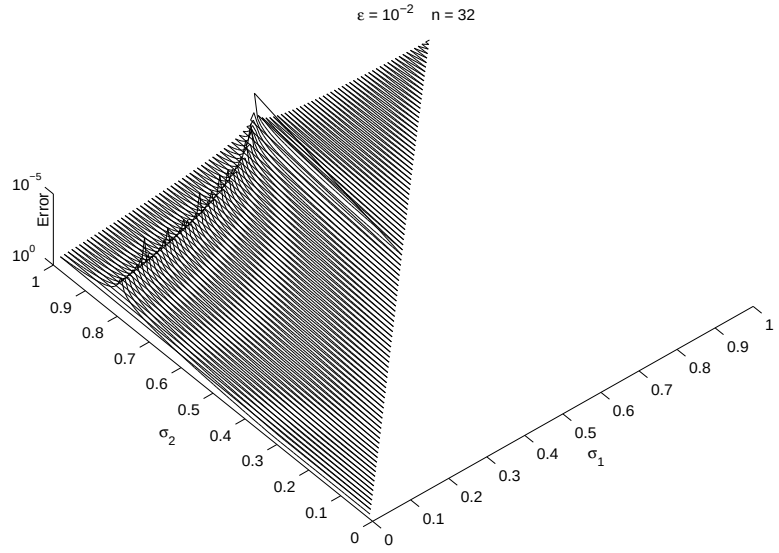
Σχήμα 2.22: Το επίπεδο (σ_1, σ_2, e) για το πρόβλημα (2.41) με $\epsilon = 10^{-2}$ και $n = 16$.



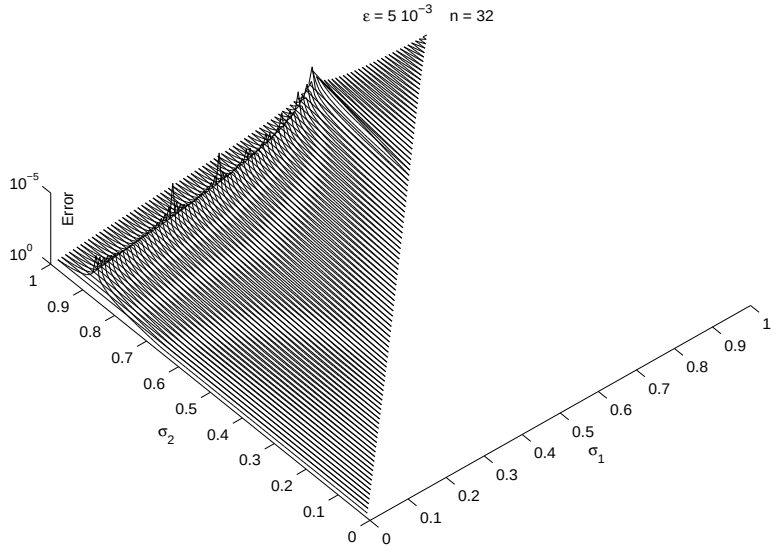
Σχήμα 2.23: Το επίπεδο (σ_1, σ_2, e) για το πρόβλημα (2.41) με $\epsilon = 5 \cdot 10^{-3}$ και $n = 16$.



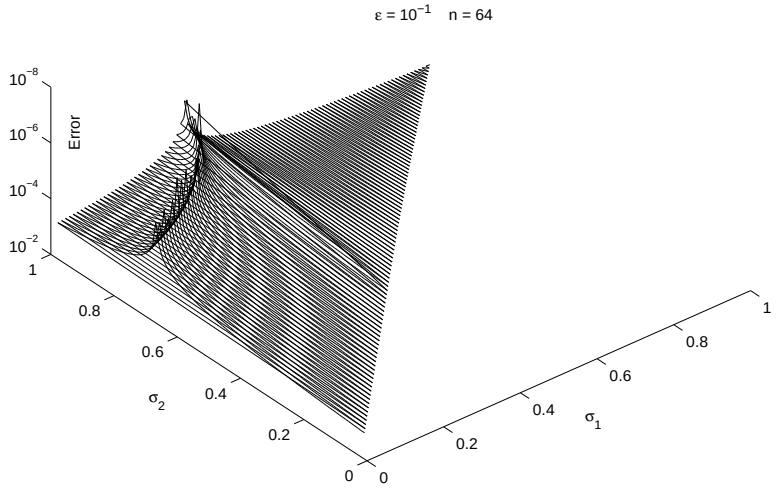
Σχήμα 2.24: Το επίπεδο (σ_1, σ_2, e) για το πρόβλημα (2.41) με $\epsilon = 10^{-1}$ και $n = 32$.



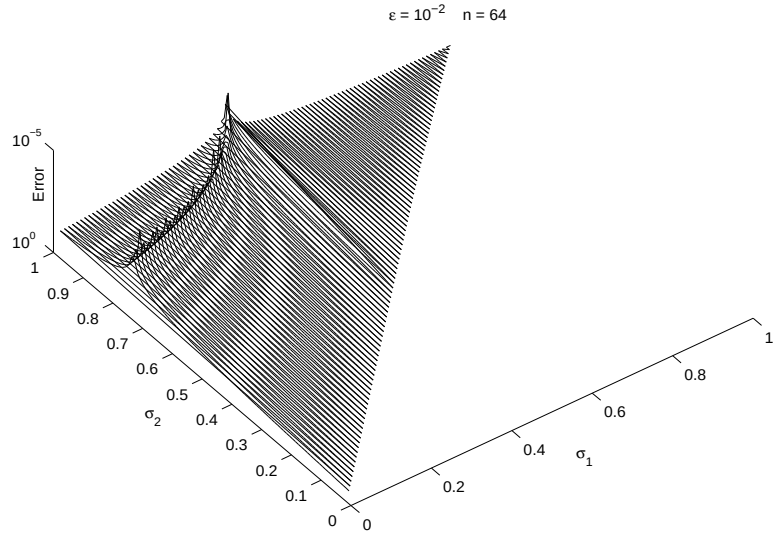
Σχήμα 2.25: Το επίπεδο (σ_1, σ_2, e) για το πρόβλημα (2.41) με $\epsilon = 10^{-2}$ και $n = 32$.



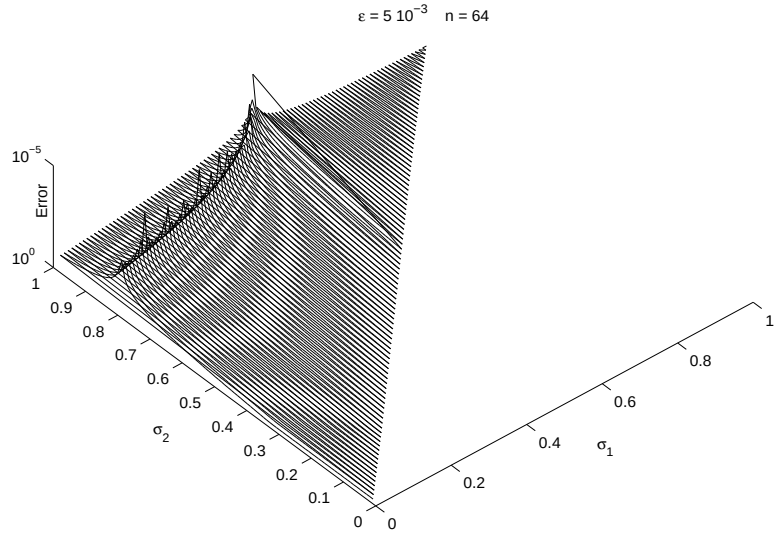
Σχήμα 2.26: Το επίπεδο (σ_1, σ_2, e) για το πρόβλημα (2.41) με $\epsilon = 5 \cdot 10^{-3}$ και $n = 32$.



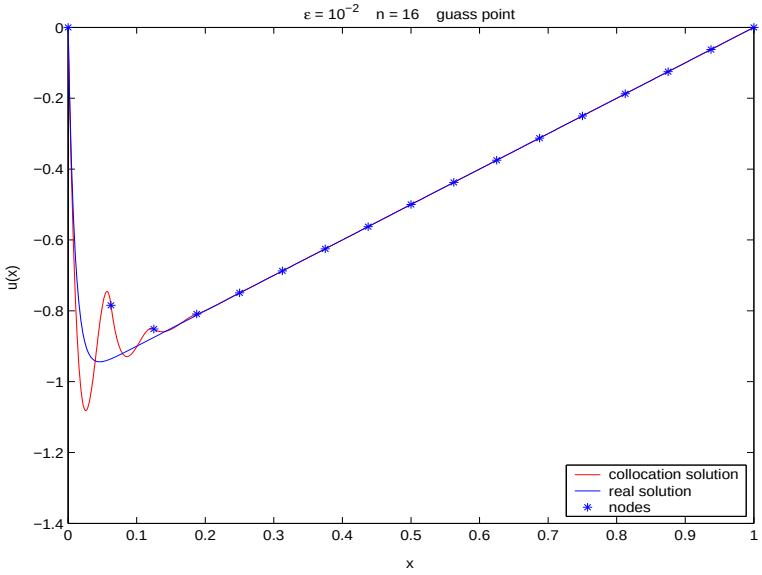
Σχήμα 2.27: Το επίπεδο (σ_1, σ_2, e) για το πρόβλημα (2.41) με $\epsilon = 10^{-1}$ και $n = 64$.



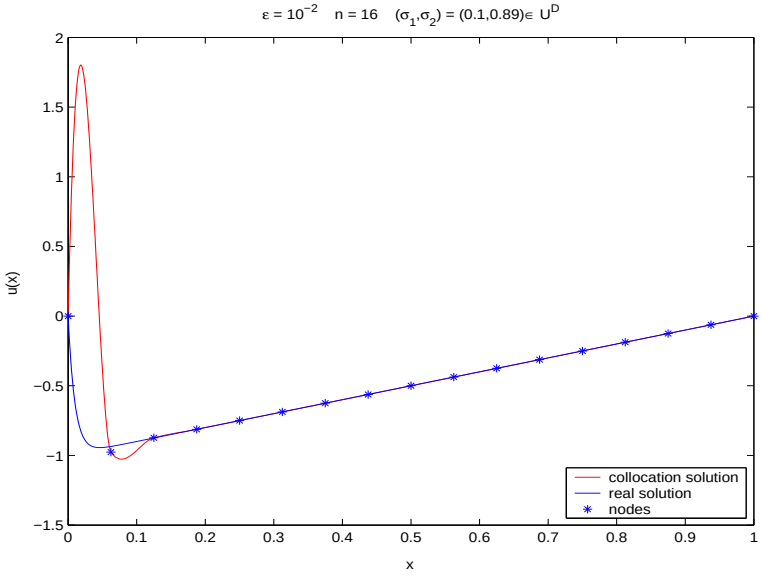
Σχήμα 2.28: Το επίπεδο (σ_1, σ_2, e) για το πρόβλημα (2.41) με $\epsilon = 10^{-2}$ και $n = 64$.



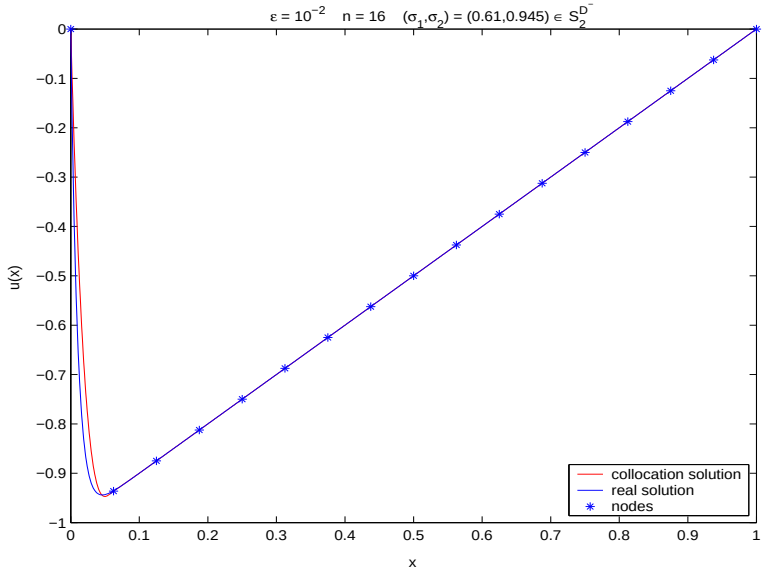
Σχήμα 2.29: Το επίπεδο (σ_1, σ_2, e) για το πρόβλημα (2.41) με $\epsilon = 5 \cdot 10^{-3}$ και $n = 64$.



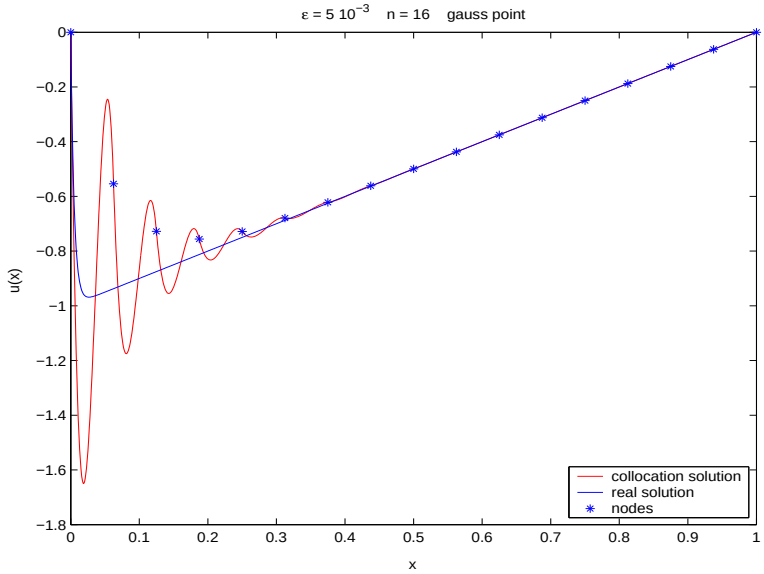
Σχήμα 2.30: Προσεγγιστική και πραγματική λύση για $\epsilon = 10^{-2}$, $n = 16$ με $(\sigma_1, \sigma_2) =$ gauss point.



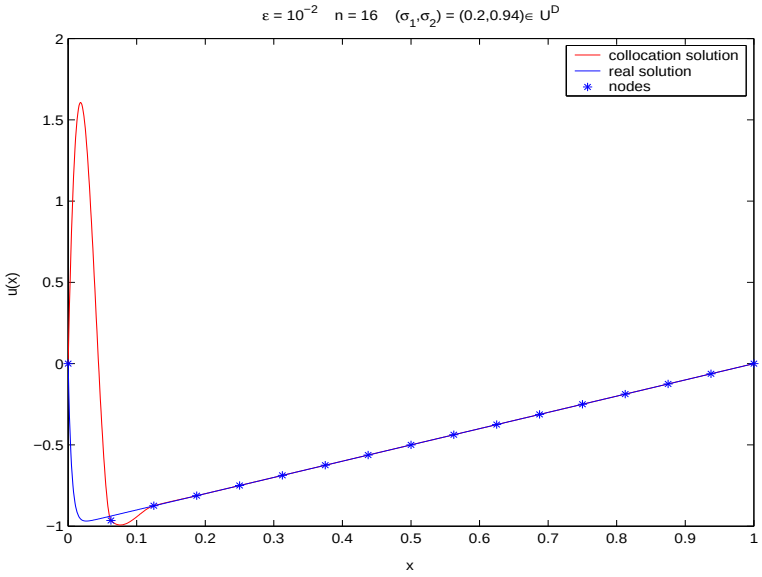
Σχήμα 2.31: Προσεγγιστική και πραγματική λύση για $\epsilon = 10^{-2}$, $n = 16$ με $(\sigma_1, \sigma_2) = (0.1, 0.89) \in U^D$.



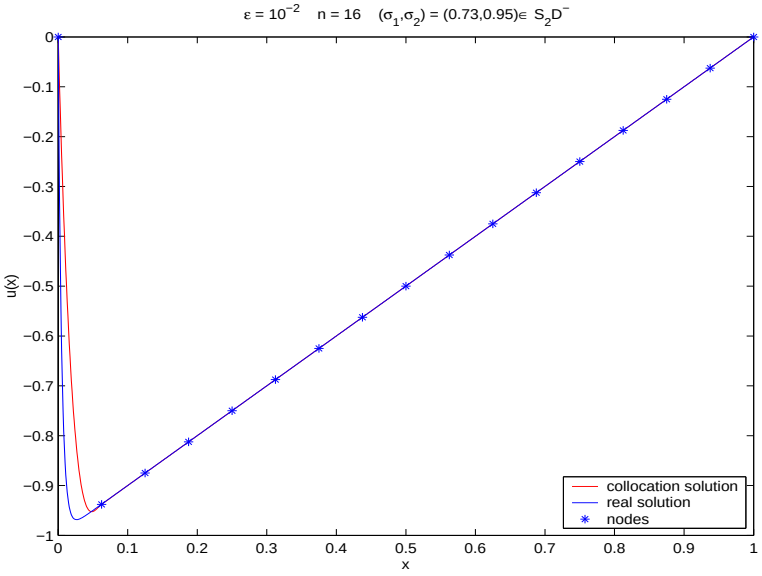
Σχήμα 2.32: Προσεγγιστική και πραγματική λύση για $\epsilon = 10^{-2}$, $n = 16$ με $(\sigma_1, \sigma_2) = (0.61, 0.945) \in S_2^{D^-}$.



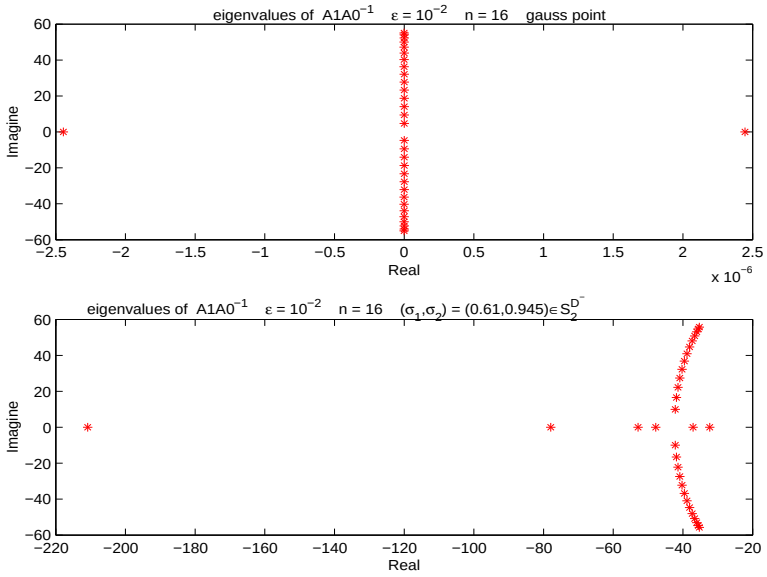
Σχήμα 2.33: Προσεγγιστική και πραγματική λύση για $\epsilon = 5 \cdot 10^{-3}$, $n = 16$ με $(\sigma_1, \sigma_2) =$ gauss point.



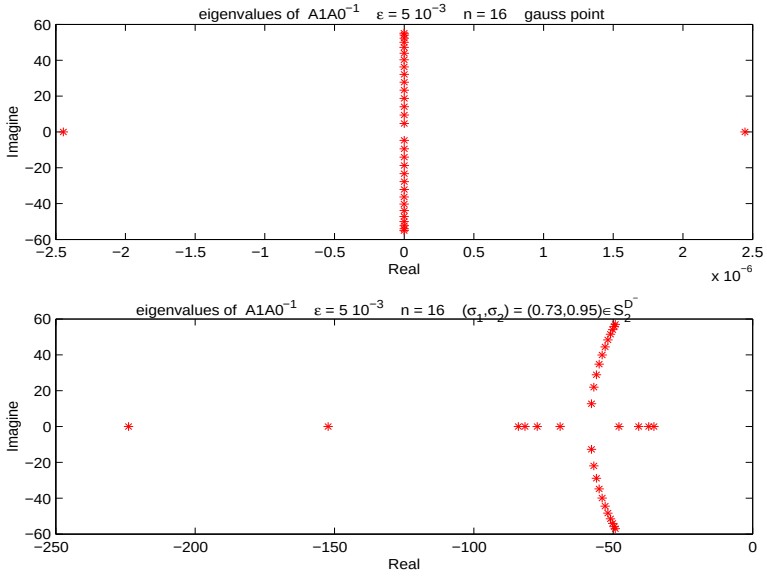
Σχήμα 2.34: Προσεγγιστική και πραγματική λύση για $\epsilon = 5 \cdot 10^{-3}$, $n = 16$ με $(\sigma_1, \sigma_2) = (0.2, 0.94) \in U^D$.



Σχήμα 2.35: Προσεγγιστική και πραγματική λύση για $\epsilon = 5 \cdot 10^{-3}$, $n = 16$ με $(\sigma_1, \sigma_2) = (0.73, 0.95) \in S_2^{D^-}$.



Σχήμα 2.36: Οι ιδιοτιμές του πίνακα $A_1 A_0^{-1}$, για $\epsilon = 10^{-2}$, $n = 16$ με $(\sigma_1, \sigma_2) =$ gauss point και $(\sigma_1, \sigma_2) = (0.61, 0.945)$ αντίστοιχα.



Σχήμα 2.37: Οι ιδιοτιμές του πίνακα $A_1 A_0^{-1}$, για $\epsilon = 5 \cdot 10^{-3}$, $n = 16$ με $(\sigma_1, \sigma_2) =$ gauss point και $(\sigma_1, \sigma_2) = (0.73, 0.95)$ αντίστοιχα.

Κεφάλαιο 3

Adaptive h-Refinement Μεθόδοι

Στα δυο πρώτα κεφάλαια χρησιμοποιήσαμε ομοιόμορφη διαμέριση του διαστήματος $I = [a, b]$ για την αριθμητική επίλυση των προβλημάτων μας. Όταν η ποσότητα $\frac{\epsilon}{|p|}n(b-a)$ είναι αρκετά μικρή, τότε ένα upwind σχήμα είναι επιθυμητό, αφού η γνωστή μας orthogonal collocation προσδίδει ταλαντώσεις στην προσεγγιστική λύση κοντά στην περιοχή όπου η $\left\|u'(x)\right\|$ της πραγματικής λύσης είναι αρκετά μεγάλη. Παρόλα αυτά το βέλτιστο upwind collocation σχήμα δύσκολα μπορεί να προσδιοριστεί. Επίσης παρατηρήσαμε ότι όταν η ποσότητα $\frac{\epsilon}{|p|}n(b-a)$ μεγαλώνει, τότε η orthogonal collocation προτιμάται δίνοντας σφάλματα τάξεως $O(h^4)$. Όπως έχουμε ήδη αναφέρει, όταν ο αριθμός $\frac{\epsilon}{|p|}n(b-a)$ είναι μεγάλος τότε υπάρχει ικανοποιητικός αριθμός κόμβων ώστε η λύση να θεωρείται ομαλή ως προς την ομοιόμορφη αυτή διαμέριση. Στην πραγματικότητα ικανοποιητικός αριθμός κόμβων βρίσκεται στην περιοχή όπου η $\left\|u'(x)\right\|$ μεταβάλλεται απότομα. Εάν καταφέρουμε, κάθε φορά που λύνουμε ενα πρόβλημα αυτής της μορφής, να συσσωρεύουμε ικανοποιητικό αριθμό κόμβων κοντά στην ιδιάζουσα περιοχή, χωρίς κατά ανάγκη να χρησιμοποιούμε ομοιόμορφη διαμέριση και χωρίς να αυξάνουμε σημαντικά το μέγεθος της διαμέρισης, τότε η ποσότητα $\frac{\epsilon}{|p|}n$, τοπικά σε αυτήν την περιοχή θεωρείται αρκετά μεγάλη ώστε η orthogonal collocation να συμπεριφέρεται καλλίτερα. Οπότε θέτουμε το ερώτημα που θα μας απασχολήσει στο παρόν κεφάλαιο.

' Μπορούμε να βρούμε μια καλλίτερη διαμέριση του διαστήματος

$I = [a, b]$ ώστε η αριθμητική μας μέθοδος να συμπεριφέρεται καλλίτερα σε τέτοιου είδους προβλήματα; '

Όπως γνωρίζουμε το παραπάνω ερώτημα απασχολεί τις περισσότερες αριθμητικές μεθόδους, αφού σε κάποιο στάδιο τους εμπλέκουν την διακριτοποίηση της διαφορικής εξίσωσης πάνω σε ένα πλέγμα. Ο σκοπός του κεφαλαίου αυτού είναι η πρακτική επιλογή ενός *καλού* πλέγματος με αντικειμενικό στόχο την επιτυχή και ακριβή αριθμητική επίλυση μιας διαφορικής εξίσωσης της μορφής μεταφοράς - διάχυσης, αλλά και οποιαδήποτε άλλης singular διαφορικής εξίσωσης πληρώνοντας όσο το δυνατό λιγότερο. Η επιλογή ενός *καλού* πλέγματος είναι ουσιώδης για την αποτελεσματική επίλυση προβλημάτων των οποίων η λύση μεταβάλλεται απότομα σε διάφορες μικρές περιοχές του πεδίου ορισμού της. Βεβαίως μια ομοιόμορφη διαμεριση με ένα πολύ μεγάλο αριθμό σημείων θα μπορούσε να λύσει με ακρίβεια ένα ιδιόμορφο πρόβλημα. Όμως ένα πλέγμα αυτής της μορφής δεν θα μπορούσε να θεωρηθεί *καλό*, διότι ο μεγάλος αριθμός των σημείων του αυξάνει πολύ το υπολογιστικό κόστος. Άρα αρχή μας είναι η εύρεση ενός πιθανού αραιού πλέγματος, πάνω στο οποίο η προσεγγιστική λύση θα διαφέρει από την πραγματική λύση επαρκώς σε σχέση με το μέγεθος του πλέγματος διακριτοποίησης. Συνεπώς ένα πλέγμα θα λέμε ότι θεωρείται ικανοποιητικά καλό, όταν στην πραγματικότητα μας δίνει καλή προσεγγιστική λύση η ακόμα καλλίτερα όταν το διακριτό πρόβλημα συνοριακών τιμών (Π.Σ.Τ.) αναπαριστά επαρκώς το συνεχές Π.Σ.Τ .

3.1 Επιλογή Πλέγματος Διακριτοποίησης

Το πρόβλημα επιλογής πλέγματος τίθεται ως εξής:

' Δεδομένου ενός Π.Σ.Τ. μορφής μεταφοράς - διάχυσης (ή οποιασδήποτε άλλης μορφής)

$$\begin{aligned} -\epsilon u''(x) + p(x)u'(x) &= f(x) \quad \text{με} \quad a < x < b \\ g(u(a), u(b)) &= 0 \end{aligned}$$

το οποίο(όπως και κάθε άλλο Π.Σ.Τ.) μπορεί να γράφει

$$\begin{aligned} y'(x) &= f(x, y(x)) \\ b(y(a), y(b)) &= 0 \end{aligned}$$

με $y(x) = (u(x), u'(x))^T$ και

$$f(x, y(x)) = \begin{pmatrix} 0 & 1 \\ 0 & -\dfrac{p(x)}{\epsilon} \end{pmatrix} y(x) + \begin{pmatrix} 0 \\ f(x) \end{pmatrix}$$

και ενός σφάλματος ανοχής TOL, βρες ένα πλέγμα

$$\pi : a = x_1 < x_2 < \ldots < x_n < x_{n+1} = b$$

με

$$h = \max_{1 \le i \le n} h_i \qquad h_i = x_{i+1} - x_i \qquad i = 1, 2, \cdots n$$

τέτοιο ώστε το μέγεθος του πλέγματος n να είναι μικρό και το σφάλμα της προσεγγιστικής λύσης $y_\pi(x)$ από την πραγματική λύση $y(x)$ να είναι μικρότερη από το TOL(χρησιμοποιώντας κατάλληλη νόρμα και μετρώντας με απόλυτα ή σχετικά σφάλματα).'

Η ιδέα για ένα *μικρό* n θα λέγαμε ότι είναι περισσότερο ποιοτική παρά ποσοτική. Η ιδανική απάντηση για το προηγούμενο πρόβλημα, θα ήταν η εύρεση ενός βέλτιστου πλέγματος διακριτοποίησης, δηλαδή ενός πλέγματος με το μικρότερο δυνατό μέγεθος n για το δεδομένο σφάλμα ανοχής TOL. Στην πράξη όμως, η εύρεση ενός τέτοιου πλέγματος διακριτοποίησης οδηγεί σε ένα πρόβλημα βελτιστοποίησης του οποίου η επίλυση, είναι πάρα πολύ δύσκολη, λόγω υπερβολικού κόστους, αν όχι αδύνατη πρακτικά. Γι αυτό λοιπόν, για να απαντήσουμε στο πρόβλημα επιλογής πλέγματος και να επιλέξουμε ένα ποιοτικό πλέγμα μεγέθους n, θα ασχοληθούμε με adaptive h-refinement techniques [WAN03, WRI03, ASC95]. Αρχικά θα μιλήσουμε για την βασική μέθοδο των h-refinement techniques, όπως προκύπτει από την ομοιόμορφη κατανομή ενός μέτρου συμπεριφοράς της λύσης, και προτάθηκε στο [WHI79]. Η διατύπωση για την μέθοδο αυτή έχει ως εξής, δεδομένου ενός μεγέθους n επέλεξε ένα πλέγμα το οποίο θα ελαχιστοποιεί το σφάλμα ως προς ένα μέτρο που ελέγχει την συμπεριφορά της λύσης. Η μέθοδος αυτή προκύπτει από έναν έμμεσο μετασχηματισμό και θα αναφερθούμε εκτενώς σε επόμενες παραγράφους. Στην συνέχεια θα προσπαθήσουμε να απαντήσουμε στο πρόβλημα επιλογής πλέγματος, κατασκευάζοντας μια επαναληπτική μέθοδο, iterative h-refinement technique, η οποία θα χρησιμοποιεί την hermite cubic finite element orthogonal collocation. Στην μέθοδο που θα παρουσιάσουμε, θα αναζητήσουμε πλέγματα διακριτοποίησης με μέγεθος όχι πολύ μεγαλύτερο από εκείνο του βέλτιστου. Πιο συγκεκριμένα μπορούμε να υποθέσουμε μια ακολουθία από σφάλματα ανοχής TOL και πλέγματα μεγέθους $n = n(TOL)$ τέτοια ώστε,

$$n(TOL) \leqq Cn^*(TOL)$$

όπως $TOL \to 0$, με $n^*(TOL)$ το βέλτιστο μέγεθος πλέγματος για το αντίστοιχο σφάλμα ανοχής TOL και C σταθερά με $C \approx 1$ ανεξάρτητη του TOL.

Με δεδομένα ένα Π.Σ.Τ. και αριθμητικής μεθόδου, η καταλληλότητα του διακριτού μας μοντέλου καθορίζεται και από το πλέγμα διακριτοποίησης. Γενικότερα θα λέγαμε ότι ένα *καλό* πλέγμα θα πρέπει να είναι πυκνό σε περιοχές όπου η επιθυμητή λύση αλλάζει απότομα αλλά και σχετικά αραιό στις υπόλοιπες περιοχές ώστε το διακριτό σχήμα να προσφέρει καλή λύση. Για να γίνει κάτι τέτοιο απαραίτητη και βασική προϋπόθεση είναι ο

' Έλεγχος του σφάλματος διακριτοποίησης '

όποτε και μια καλή προσεγγιστική λύση είναι δυνατή. Αυτή είναι η βασική προϋπόθεση των adaptive h-refinement μεθόδων. Στην επόμενη παράγραφο θα αναφέρουμε την h-refinement τεχνική όπως αυτή προκύπτει από μεθόδους μετασχηματισμών.

3.2 Μέθοδοι Μετασχηματισμών

Στις μεθόδους μετασχηματισμών μια αλλαγή μεταβλητών εφαρμόζεται στο Π.Σ.Τ.,

$$\begin{aligned} &y'(x) = f(x, y(x)) \qquad\qquad a < x < b \\ &g(y(a), y(b)) = 0 \end{aligned} \tag{3.1}$$

ελπίζοντας ότι το παραγόμενο νέο μετασχηματισμένο πρόβλημα θα είναι περισσότερο αποτελεσματικό σε αριθμητικούς υπολογισμούς. Γενικότερα θα λέγαμε ότι ψάχνουμε για μετασχηματισμούς συντεταγμένων $x(t)$ της ανεξάρτητης μεταβλητής x, όπου θα έχουν ως στόχο η λύση $y(x(t))$ να είναι να είναι ομαλότερη(θα μεταβάλλεται αργά στο πεδίο ορισμού της) ως συνάρτηση της μεταβλητής t.

3.2.1 Άμεσοι Μετασχηματισμοί

Η ιδέα ενός μετασχηματισμού συντεταγμένων για την επίτευξη μιας εκλέπτυνσης του πλέγματος διακριτοποίησης είναι αρκετά χρήσιμη. Για έναν τέτοιο μετασχηματισμό κρίνεται αναγκαία η εύρεση μιας ανεξάρτητης μεταβλητής $x(t)$,

η οποία θα αποκλείει την ανάγκη για ένα μη-ομοιόμορφο πλέγμα διακριτοποίησης για την προσέγγιση της λύσης $y(x(t))$. Οπότε εάν η νέα μεταβλητή επιλέγει σωστά, το νέο μετασχηματισμένο Π.Σ.Τ. θα μπορεί να επιλυθεί αριθμητικά και με ακρίβεια, από έναν μικρό αριθμό σημείων που προκύπτουν από μια ομοιόμορφη διαμέριση για την νέα μετασχηματισμένη μεταβλητή.

Εάν είναι διαθέσιμη εκ των προτέρων κάποια γνώση για την λύση του Π.Σ.Τ. τότε ένας άμεσος μετασχηματισμός συντεταγμένων $x(t)$ μπορεί να επιλεγεί κατάλληλα, ώστε να είμαστε σε θέση να μορφοποιήσουμε κατάλληλα την συμπεριφορά της λύσης. Προφανώς για έναν τέτοιο μετασχηματισμό θα πρέπει να ισχύουν τα εξής:

$$\frac{dx}{dt} > 0 \qquad\qquad 0 < t < 1$$

$$x(0) = a \qquad\qquad x(1) = b$$

οπότε το Π.Σ.Τ. γράφεται:

$$\begin{aligned} \hat{y}'(t) = \hat{f}(t, \hat{y}(t))x'(t) \quad 0 < t < 1 \\ g(\hat{y}(0), \hat{y}(1)) = 0 \end{aligned}$$

όπου $\hat{y}(t) = y(x(t))$ και $\hat{f}(t, \hat{y}(t)) = f(x(t), y(x(t)))$. Συνεπώς, κάθε φορά θέλουμε να διαλέγουμε τον μετασχηματισμό $x(t)$ με τέτοιο τρόπο ώστε η λύση του μετασχηματισμένου Π.Σ.Τ. να είναι ομαλή(καλά συμπεριφερόμενη-χωρίς απότομες αλλαγές) και το νέο μετασχηματισμένο Π.Σ.Τ. να λύνεται αποτελεσματικά.

Στην πραγματικότητα ψάχνουμε μετασχηματισμούς $x(t)$ με αντίστροφο $t(x) = x^{-1}(t)$ που καταφέρνουν μια επέκταση των συντεταγμένων στις ευαίσθητες περιοχές της λύσης, επιδιώκοντας την επιτυχή αριθμητική επίλυση του μετασχηματισμένου Π.Σ.Τ. σε ένα ομοιόμορφο πλέγμα.

Παράδειγμα 3 Θεωρούμε το πρόβλημα συνοριακών τιμών

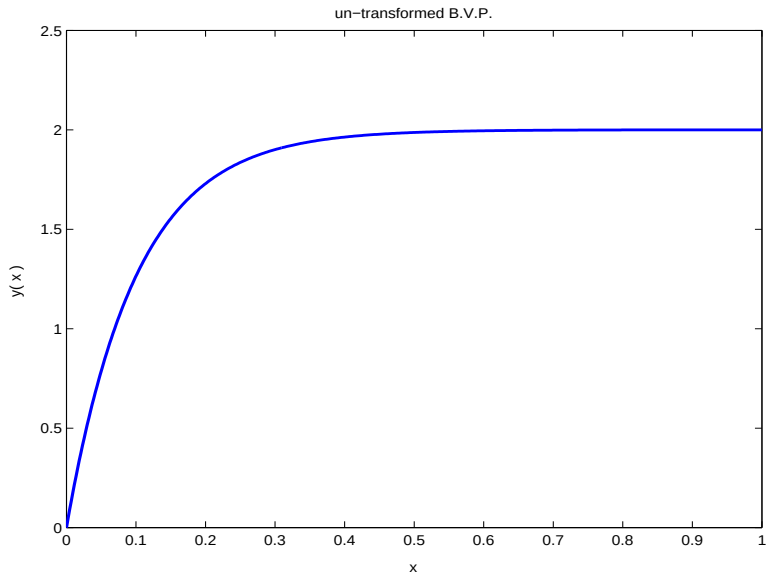
$$\epsilon u''(x) + u'(x) = 0 \qquad\qquad 0 < x < 1$$

$$u(0) = u(1) = 0$$

η αναλυτική λύση του προβλήματος είναι

$$u(x) = \frac{-2 + 2e^{\frac{x}{\epsilon}}}{-1 + e^{\frac{1}{\epsilon}}}$$

η λύση του Π.Σ.Τ. για $\epsilon = 0.01$ παρουσιάζει boundary layer στο $x = 0$.



Σχήμα 3.1: Η πραγματική λύση για $\epsilon = 0.01$ του αρχικου Π.Σ.Τ.

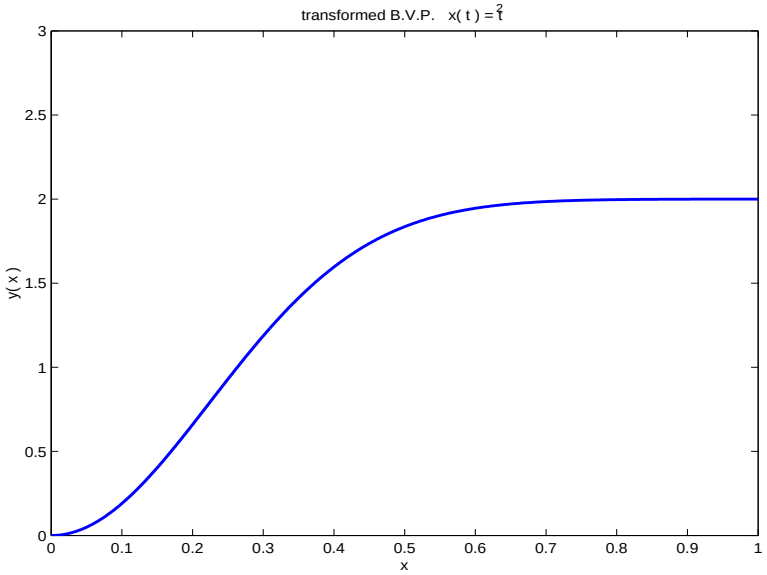
Όμως θεωρώντας τον μετασχηματισμό συντεταγμένων

$$x(t) = t^2 \qquad 0 < t < 1$$
$$x(0) = 0 \qquad x(1) = 1$$

τότε το πρόβλημα συνοριακών τιμών γίνεται

$$\frac{\epsilon}{2t}\hat{u}''t + (1 - \frac{\epsilon}{2t^2})\hat{u}'(t) = 0 \qquad 0 < t < 1$$

$$\hat{u}(0) = \hat{u}(1) = 0$$



Σχήμα 3.2: Η πραγματική λύση για $\epsilon = 0.01$ του μετασχηματισμένου Π.Σ.Τ.

Παρατηρούμε ότι η μέθοδος αυτή έχει δυο σημαντικά μειονεκτήματα:

- Η εκ των προτέρων πληροφορία της λύσης για τον εντοπισμό των ιδιάζουσων περιοχών(boundary-interior layers) της, όπου η λύση δεν είναι αρκετά ομαλή, πρέπει να γίνει πριν ξεκινήσουμε την αριθμητική επίλυση του προβλήματος και τις περισσότερες φορές είναι δύσκολο αν όχι αδύνατη στην πράξη.

- Δεν υπάρχουν πάντοτε ουσιαστικές ενδείξεις ότι το μετασχηματισμένο Π.Σ.Τ. είναι λιγότερο δύσκολο από το γνήσιο.

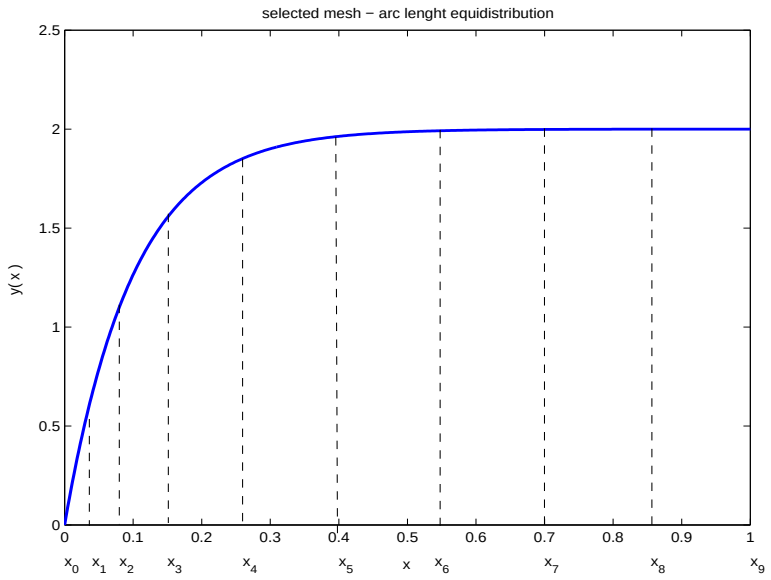
3.2.2 Έμμεσοι Μετασχηματισμοί

Η εκ των προτέρων γνώση της συμπεριφοράς της λύσης $y(x)$ σε συνδυασμό με την τυχόν κακή της συμπεριφορά σε διάφορες περιοχές της καθιστούν την επιλογή ενός explicit μετασχηματισμού συντεταγμένων, όπως αναφέραμε προηγουμένως παρά πολύ δύσκολη αν όχι αδύνατη στην πράξη. Για τον λόγο αυτό, είναι αναγκαία μια διαφορετική προσέγγιση για την επιλογή ενός επιθυμητού μετασχηματισμού συντεταγμένων $x(t)$ η $t(x)$, ώστε η λύση $y(t)$ να συμπεριφέρεται καλλίτερα. Κάτι τέτοιο μπορεί να γίνει χρησιμοποιώντας μια έμμεση

διαδικασία μετασχηματισμού. Η διαδικασία αυτή αποτελεί την πρώτη προσέγγιση των h-refinement μεθόδων όπως αυτή προτάθηκε από [WHI79] και έχει ως εξής:

Ομοιόμορφη κατανομή του μήκους τόξου

Θεωρούμε το πρόβλημα εύρεσης ενός πλέγματος $\{x_i\}$ ώστε το μήκος τόξου της πραγματικής λύσης $y(x)$ να κατανέμεται ομοιόμορφα σε κάθε ένα από τα υποδιαστήματα $[x_i, x_{i+1}]$ με $i = 1, \cdots, n$. Για παράδειγμα έστω ότι η $y(x)$ είναι μια βαθμωτή συνάρτηση, θεωρούμε το συνολικό μήκος του γραφήματος να είναι ίσο με θ. Τότε λοιπόν θέλουμε να βρούμε ένα πλέγμα $\{x_i\}$ με την ιδιότητα ότι το μήκος της καμπύλης από το (x_i, y_i) μέχρι το (x_{i+1}, y_{i+1}) θα είναι $\dfrac{\theta}{n}$.



Σχήμα 3.3: Ομοιόμορφη κατανομή του μήκους του γραφήματος του σχήματος 1.2

Οπότε είμαστε έτοιμοι να ορίσουμε τον μετασχηματισμό συντεταγμένων, ως προς το μήκος τόξου της πραγματικής λύσης,

$$\bar{t}(x) = \int_a^x (1 + \|y'(s)\|_2^2)^{\frac{1}{2}} ds$$

όπου $y(x)$ η ακριβής λύση του Π.Σ.Τ. Όποτε γράφοντας τον παραπάνω μετασχηματισμό σε κανονικοποιημένη μορφή, θα έχει ως εξής:

$$t(x) = \frac{1}{\theta}\int_a^x (1 + \left\|y'(s)\right\|_2^2)^{\frac{1}{2}} ds$$

ή

$$t(x) = \frac{1}{\theta}\int_a^x (1 + \left\|f(s, y(s))\right\|_2^2)^{\frac{1}{2}} ds$$

με

$$\theta = \int_a^b (1 + \left\|f(s, y(s))\right\|_2^2)^{\frac{1}{2}} ds$$

Προφανώς για τον μετασχηματισμό $t(x)$ και για το θ θα ισχύουν:

$$\frac{dt}{dx} = \frac{(1 + \left\|f(x, y(x))\right\|_2^2)^{\frac{1}{2}}}{\theta} \qquad a < x < b \tag{3.2}$$

$$\frac{d\theta}{dx} = 0 \qquad a < x < b \tag{3.3}$$

$$t(a) = 0 \qquad t(b) = 1 \tag{3.4}$$

Η προϋπόθεση ότι τα σημεία του πλέγματος $\{x_i\}$ διαλέγονται με τέτοιο τρόπο ώστε το μήκος της καμπύλης της πραγματικής λύσης $y(x)$ να κατανέμεται ομοιόμορφα μπορεί να γράφει ως εξής:

$$t(x_{i+1}) - t(x_i) = \frac{1}{n} = \sigma\tau\alpha\theta\epsilon\rho\acute{\alpha} \tag{3.5}$$

με $i = 1, \cdots, n$.

Συνδυάζοντας το Π.Σ.Τ. και τις σχέσεις (3.2),(3.3),(3.4),(3.5) παρατηρούμε ότι το πρόβλημα της ομοιόμορφης κατανομής του arc-length της λύσης, απαιτεί την επίλυση των παρακάτω διαφορικών εξισώσεων.

$$\frac{dy}{dx} = f(x, y(x)) \qquad a < x < b \tag{3.6}$$

$$\frac{dt}{dx} = \frac{(1 + \left\|f(x, y(x))\right\|_2^2)^{\frac{1}{2}}}{\theta} \qquad a < x < b \tag{3.7}$$

$$\frac{d\theta}{dx} = 0 \qquad a < x < b \tag{3.8}$$

για $x \in (a, b)$ μαζί με τις συνοριακές συνθήκες,

$$b(y(a), y(b)) = 0 \tag{3.9}$$

$$t(a) = 0 \qquad t(b) = 1 \tag{3.10}$$

και βεβαίως χρησιμοποιώντας την συνθήκη της ομοιόμορφης κατανομής του μήκος της καμπύλης της συνάρτησης της λύσης,

$$t(x_{i+1}) - t(x_i) = \frac{1}{n} \tag{3.11}$$

με $i = 1, \cdots, n$.

Το παραπάνω σύστημα διαφορικών εξισώσεων στην πιο χρηστική μετασχηματισμένη του μορφή σε arc-length συντεταγμένες γράφεται,

$$\hat{y}'(t) = \hat{f}(t, \hat{y}(t))x'(t) \qquad 0 < t < 1 \tag{3.12}$$

$$x'(t) = \frac{\theta}{(1 + \left\|\hat{f}(t, \hat{y}(t))\right\|_2^2)^{\frac{1}{2}}} \qquad 0 < t < 1 \tag{3.13}$$

$$\theta'(t) = 0 \qquad 0 < t < 1 \tag{3.14}$$

με $t \in (0, 1)$ και με κατάλληλες μετασχηματισμένες τις συνοριακές συνθήκες,

$$b(\hat{y}(0), \hat{y}(1)) = 0 \tag{3.15}$$

$$x(0) = a \qquad x(1) = b \tag{3.16}$$

Παρατηρούμε ότι το τελευταίο Π.Σ.Τ. αποκτά την τυπική μετασχηματισμένη του μορφή όταν η διαφορική εξίσωση (3.13) αντικατασταθεί στην (3.12). Προφανώς το μετασχηματισμένο Π.Σ.Τ. είναι μη γραμμικό ακόμη και αν το αρχικό Π.Σ.Τ. είναι γραμμικό. Επίσης παρατηρούμε ότι ο εξαναγκασμός της ομοιόμορφης κατανομής του arc-length (3.11) μπορεί να ικανοποιηθεί, κατά την επίλυση του μετασχηματισμένου Π.Σ.Τ. από την σχέση,

$$x_i = x(\frac{i-1}{n}) \tag{3.17}$$

με $i = 1, \cdots, n$.

Οπότε έχοντας με αυτόν τον τρόπο μια περισσότερο ομαλή λύση για το μετασχηματισμένο Π.Σ.Τ. (3.12),(3.13),(3.14), μπορούμε να την υπολογίσουμε με μια ομοιόμορφη διαμέριση για την μεταβλητή t ώστε να ικανοποιείται η συνθήκη (3.11) δηλαδή,

$$\left\{0, \frac{1}{n}, \frac{2}{n}, \cdots, 1\right\}$$

και όπου $y(x_i) = \hat{y}(\frac{i-1}{n})$ με $i = 1, \cdots, n$, οι προσεγγιστικές τιμές της λύσης στα σημεία x_i που αποκτούνται από την σχέση (3.17).

Όπως έχουμε ήδη αναφέρει, βασική προϋπόθεση στην επιλογή ενός *καλού* πλέγματος είναι ο έλεγχος του σφάλματος διακριτοποίησης. Κατανέμοντας ομοιόμορφα το arc-length της πραγματικής λύσης καταφέραμε να κατασκευάσουμε ένα πλέγμα, πάνω στο οποίο η λύση μπορεί να θεωρηθεί ομαλή, δηλαδή με άλλα λόγια ελέγξαμε με αυτόν τον τρόπο το σφάλμα της διακριτοποίησης. Όμως εκτός από το arc-length της πραγματικής λύσης υπάρχουν και άλλα μεγέθη που μπορούμε να επιλέξουμε για να μετράμε την συμπεριφορά της λύσης $y(x)$ σε σχέση με το διακριτό μας μοντέλο, όπως για παράδειγμα το τοπικό σφάλμα αποκοπής της εκάστοτε αριθμητικής μεθόδου που χρησιμοποιείται. Αυτά λοιπόν τα μεγέθη που μετρούν την συμπεριφορά της λύσης, τα οποία θα κατανέμουμε ομοιόμορφα κάθε φορά ώστε να παράγουμε ένα καλό πλέγμα και μια καλή προσεγγιστική λύση, θα καλούνται monitor functions. Βεβαίως όμως κάθε monitor συνάρτηση πρέπει να ικανοποιεί κάποιες συνθήκες ώστε να θεωρείται αποδεκτή. Στην επόμενη ενότητα θα γενικεύσουμε όλα όσα αναφέραμε και για την περίπτωση των αποδεκτών monitor συναρτήσεων.

Ομοιόμορφη κατανομή monitor συναρτήσεων

Θα εισάγουμε την γενικευμένη ιδέα μιας monitor function για την διαφορική εξίσωση (3.1). Ο ορισμός που θα δώσουμε παρακάτω θα περιλαμβάνει και την προηγούμενη περίπτωση της arc-lenght monitor συνάρτησης, που είδη συζητήσαμε.

Ορισμός 3.1 *Θα λέμε ότι ο μετασχηματισμός $t(x)$ είναι αποδεκτός monitor μετασχηματισμός για την διαφορική εξίσωση (3.1), εάν υπάρχει μια συνάρτηση $m(x,v)$, που θα καλείται monitor function, η οποία*

1. *έχει συνεχείς μερικές παραγώγους στο σύνολο $\Big\{(x,v) : v \in S_p(y(x)), a < x < b\Big\}$, για κάποια μπάλα S_p ακτίνας p και κέντρου $y(x)$, και είναι τέτοια ώστε:*
2. $\dfrac{dt}{dx} = \dfrac{m(x,y(x))}{\theta} \geq \delta > 0$ *για κάθε* $x \in [a.b]$ *με* $t(a) = 0$ *και* $\theta = \int_a^b m(x,y(x))dx$. $\square$

Η ιδιότητα 2 μας επιτρέπει να χρησιμοποιήσουμε τον μετασχηματισμό $t(x)$ ως ανεξάρτητη μεταβλητή για το Π.Σ.Τ., ενώ η ιδιότητα 1 μας εξασφαλίζει ότι οι λύσεις $y(x(t)) \equiv \hat{y}(t)$ δεν χάνουν την συνέχεια κάτω από τον συγκεκριμένο μετασχηματισμό.

Ορισμός 3.2 *Θα λέμε ότι η* $t(x)$ *κατανέμεται ομοιόμορφα πάνω στο πλέγμα* $\pi = \left\{x_i\right\}$ *αν,*

$$t(x_{i+1}) - t(x_i) = \frac{1}{n}$$

με $i = 1, \cdots, n$.

ή διαφορετικά ένα πλέγμα $\pi = \left\{x_i\right\}$ *θα καλείται ομοιόμορφα κατανεμημένο σε σχέση με την* $monitor$ *συνάρτηση* $m(x, y(x))$ *στο διάστημα* $[a, b]$ *εάν για κάποια σταθερά* λ *ισχύει,*

$$\int_{x_i}^{x_{i+1}} m(x, y(x))dx = \lambda$$

με $i = 1, \cdots, n$ *και,*

$$\lambda = \frac{\theta}{n}$$

όπου

$$\theta = \int_a^b m(x, y(x))dx$$

□

Οπότε το μετασχηματισμένο Π.Σ.Τ. σε monitor συντεταγμένες, θα ακολουθεί το παρακάτω σύστημα διαφορικών εξισώσεων.

$$\frac{d\hat{y}}{dt} = \frac{\theta \hat{f}(t, \hat{y}(t))}{m(t, \hat{y}(t))} \qquad 0 < t < 1 \tag{3.18}$$

$$\frac{dx}{dt} = \theta \frac{1}{m(t, \hat{y}(t))} \qquad 0 < t < 1 \tag{3.19}$$

$$\frac{d\theta}{dt} = 0 \qquad 0 < t < 1 \tag{3.20}$$

με $t \in (0, 1)$ και με συνοριακές συνθήκες,

$$b(\hat{y}(0), \hat{y}(1)) = 0 \tag{3.21}$$

$$x(0) = a \qquad x(1) = b \tag{3.22}$$

Παρατηρούμε ότι η αρχική ιδέα της ομοιόμορφης κατανομής του μήκος τόξου της λύσης ακολουθεί όλους τους παραπάνω ορισμούς, καθώς και το σύστημα των διαφορικών εξισώσεων είναι το ίδιο αν θεωρηθεί ως $m(x, y(x)) = 1 + \|y'(x)\|_2^2$. Επίσης, θα πρέπει να σημειώσουμε ότι οι monitor συναρτήσεις εξαρτώνται από την πραγματική λύση, όποτε πάντοτε ένας τέτοιος έμμεσος μετασχηματισμός θα οδηγεί σε ένα μη-γραμμικό σύστημα διαφορικών εξισώσεων ακόμη και αν το αρχικό Π.Σ.Τ. ήταν γραμμικό.

Στην ενότητα αυτή παρουσιάσαμε την h-refinement τεχνική όπως αυτή προκύπτει από τις μεθόδους μετασχηματισμών. Στην συνέχεια θα ασχοληθούμε με επαναληπτικές h-refinement τεχνικές, και προς το τέλος αυτού του κεφαλαίου θα παρουσιάσουμε την επαναληπτική h-refinement τεχνική επιλογής πλέγματος χρησιμοποιώντας την μέθοδο της finite element hermite cubic orthogonal collocation.

3.3 Άμεσες Μεθόδοι Επιλογής Πλέγματος Διακριτοποίησης

Μια άμεση μέθοδος επιλογής πλέγματος (και επίλυσης) περιγράφεται ως εξής:

Διαδικασία : Άμεσης Μεθόδου

Δεδομένου ενός αριθμητικού σχήματος διακριτοποίησης και μιας αρχικής προσεγγιστικής λύσης

Repeat:

- Καθόρισε ένα πλέγμα από την τρέχουσα λύση.
- Λύσε το Π.Σ.Τ. για την νέα y_π πάνω στο νέο πλέγμα.

Until:Μέχρι ένα σφάλμα ανοχής TOL να ικανοποιείται

Συνήθως η αρχική προσεγγιστική λύση δίνεται λύνοντας το αρχικό Π.Σ.Τ.

πάνω σε ένα αραιό αρχικό πλέγμα. Το κύριο ερώτημα που τίθεται στην επαναληπτική μέθοδο πύκνωσης είναι το εξής:

> ' Δεδομένης της προσεγγιστικής λύσης καθορισμένης από το τρέχον πλέγμα π, πως καθορίζεται το νέο εκλεπτυσμένο πλέγμα π^*; '

Για να απαντήσουμε σε αυτό το ερώτημα, θα πρέπει πρώτα να μιλήσουμε για τον τρόπο με τον οποίο θα κατασκευάζουμε και θα επιλέγουμε την monitor συνάρτηση, την οποία και θα χρησιμοποιούμε για να ελέγχουμε το σφάλμα διακριτοποίησης κατά την αριθμητική επίλυση του Π.Σ.Τ. και στην συνέχεια για να επιλέγουμε το νέο πλέγμα.

Αυτό που πραγματικά θέλουμε από την λύση του προβλήματος επιλογής πλέγματος είναι η απόκτηση μιας καλής προσεγγιστικής λύσης για το Π.Σ.Τ. με το λιγότερο δυνατό κόστος. Άρα μια στρατηγική που γενικά θα επιδίωκε την χρησιμοποίηση των συγκεκριμένων ιδιοτήτων της εκάστοτε αριθμητικής μεθόδου που χρησιμοποιείται, όπως για παράδειγμα το σφάλμα της θα μπορούσε να θεωρηθεί επιτυχής.

Ομοιόμορφη κατανομή σφάλματος

Δεδομένου ενός πλέγματος π και μιας προσεγγιστικής λύσης $y_\pi(x)$ πάνω σε αυτό το πλέγμα, συμβολίζουμε με e_i ένα μέτρο σφάλματος της προσεγγιστικής λύσης από την πραγματική στο ι-οστό υποδιάστημα του πλέγματος $[x_i, x_{i+1})$. Αυτό το μέτρο σφάλματος μπορεί να είναι απόλυτο, σχετικό ή συνδυασμός και τον δυο. Για παράδειγμα θεωρούμε ως e_i,

$$e_i = |y(x_i) - y_\pi(x_i)|$$

με $i = 1, \cdots, n+1$, όταν για παράδειγμα ένα απλό σχήμα πεπερασμένων στοιχείων χρησιμοποιηθεί στο π ή

$$e_i = \max_{x_i \le x \le x_{i+1}} |y(x) - y_\pi(x)| = \left\| y - y_\pi \right\|_i$$

με $i = 1, \cdots, n$, όταν η προσεγγιστική λύση ορίζεται συνεχώς.
Το σφάλμα e_i εξαρτάται από τα υποδιαστήματα $[x_i, x_{i+1}]$ και γενικώς αυξάνει όσο το μήκος του υποδιαστήματος αυξάνει. Όποτε μπορούμε να ορίσουμε κατάλληλα ένα αντίστοιχο μέτρο σφάλματος d_i το οποίο θα μεταβάλλεται γραμμικά σε σχέση με το μήκος h_i του κάθε υποδιαστήματος, δηλαδή

$$d_i = h_i \cdot m_i \tag{3.23}$$

όπου m_i ανεξάρτητο από το h_i. Για παράδειγμα εάν η μέθοδος διακριτοποίησης ήταν ένα σχήμα τάξεως s τότε μπορούμε να γράψουμε κάτω από κανονικές συνθήκες:

$$|y(x_i) - y_\pi(x_i)| = Ch_i^s |y^{(s)}(x_i)| + O(h_i^{s+1}) + O(h^s)$$

οπότε μπορούμε να γράψουμε την έκφραση (3.23) ως εξής:

$$m_i = C^{\frac{1}{s}} |y^{(s)}(x_i)|^{\frac{1}{s}} \tag{3.24}$$

Παρατηρούμε ότι $e_i \approx d_i^s$. Η ποσότητα m_i όπως φαίνεται από την σχέση (3.24) εξαρτάται από την διαμέριση αλλά όχι με ουσιαστικό τρόπο.

Δεδομένου ενός μεγέθους n, θα προσπαθήσουμε να διαλέξουμε ένα πλέγμα π, ώστε η ποσότητα $\max |e_i|$ να ελαχιστοποιείται. Μια απλούστερη επιλογή για την ελαχιστοποίηση της παραπάνω ποσότητας είναι η ελαχιστοποίηση της ποσότητας d_i. Οπότε προκύπτει το ακόλουθο minmax πρόβλημα με μια μόνο συνθήκη περιορισμού που τίθεται ως εξής:

$$\min_{1 \le i \le n} \left\{ |d_i| : \sum_{i=1}^{n} h_i = b - a \right\} \tag{3.25}$$

Η λύση του παραπάνω προβλήματος βελτιστοποίησης, προϋποθέτει ότι όλες οι ποσότητες d_i γίνονται ίσες με μια σταθερά λ, οπότε έχουμε τις σχέσεις:

$$h_i = \frac{\lambda}{m_i} \qquad \lambda = \frac{b-a}{\sum_{j=1}^{n} m_j^{-1}}$$

με $i = 1, \cdots, n$.

Γενικεύοντας τα παραπάνω υποθέτουμε μια ομαλή monitor συνάρτηση m, αντί ενός συνόλου διακριτών τιμών της πάνω στο πλέγμα π. Στην συνεχεία θα προσπαθήσουμε να διανέμουμε ομοιόμορφα το τοπικό σφάλμα αποκοπής ενός σχήματος διακριτοποίησης, χρησιμοποιώντας ως monitor την συνάρτηση m που εμπεριέχει υψηλές παραγωγούς της λύσης $y(x)$.

Ομοιόμορφη κατανομή τοπικού σφάλματος αποκοπής

Πριν ξεκινήσουμε θα δώσουμε έναν ορισμό:

Ορισμός 3.3 *Μια ακολουθία από πλέγματα* $\left\{\pi_n\right\}_{n=n_o}^{\infty}$ *θα καλείται ασυμπτωτικά ομοιόμορφα κατανεμημένη (asymptotically equidistributing (as.eq.)) σε σχέση με μια monitor συνάρτηση* $m(x, y(x))$ *εάν ισχύει,*

$$\int_{x_i}^{x_{i+1}} m(x, y(x))dx = \lambda(1 + O(h)) \tag{3.26}$$

με $i = 1, \ldots, n.$ □

Ο παραπάνω ορισμός υποθέτει ότι για ικανοποιητικά μεγάλο n, το πλέγμα διακριτοποίησης θα είναι ομοιόμορφα κατανεμημένο σύμφωνα με τον ορισμό 3.2. Οπότε μια καλή προσεγγιστική λύση υπάρχει στην πράξη σε σχέση με την monitor συνάρτηση, για μια κατάλληλη ακαθόριστη ακολουθία από πλέγματα διακριτοποίησης. Είμαστε τώρα έτοιμοι να ερευνήσουμε τον τρόπο με τον οποίο θα επιλέξουμε μια monitor συνάρτηση για να διαλέξουμε ασυμπτωτικά ομοιόμορφα κατανεμημένα πλέγματα.

Υποθέτουμε ότι η μέθοδος διακριτοποίησης που χρησιμοποιείται για να λύσει το (3.1), δίνει προσεγγιστική λύση που ικανοποιεί το εξής φράγμα σφάλματος.

$$\left\| y_\pi - y \right\| \leq K \max_{1 \leq i \leq n} |\tau_i[y]| \tag{3.27}$$

Κάτω από κανονικές συνθήκες το τοπικό σφάλμα αποκοπής ενός σχήματος διακριτοποίησης τάξης s ικανοποιεί την σχέση,

$$\tau_i[y] = h_i^s T(x_i) + O(h_i^p) \qquad (p > s) \tag{3.28}$$

όπου $T(x)$ είναι συνεχής συνάρτηση που εμπλέκει υψηλές παραγώγους της $y(x)$. Άρα μια φυσική επιλογή για την monitor συνάρτηση θα είναι:

$$m(x, y(x)) = |T(x)|^{\frac{1}{s}}$$

Θεώρημα 3.4 *Υποθέτουμε ένα ασυμπτωτικά ομοιόμορφο πλέγμα* π *σε σχέση με την monitor συνάρτηση* $|T(x)|^{\frac{1}{s}}$ *δηλαδή,*

$$\int_{x_i}^{x_{i+1}} |T(x)|^{\frac{1}{s}} dx = \lambda(1 + O(h)) \qquad i = 1, \ldots, n \tag{3.29}$$

οπου

$$\lambda = \frac{\theta}{n} \qquad \theta = \int_a^b |T(x)|^{\frac{1}{s}} dx \tag{3.30}$$

Τότε αντίστοιχα για την μέθοδο διακριτοποίησης που ικανοποιεί τις σχέσεις (3.26),(3.27) έχουμε την εξής σχέση σφάλματος.

$$\left\| y_\pi - y \right\| \leq K\lambda^s(1 + O(h)) + O(h) \tag{3.31}$$

ΑΠΟΔΕΙΞΗ

Από την σχέση (3.28) έχουμε

$$h_i |T(x_i)|^{\frac{1}{s}} \equiv \frac{\theta}{n}(1 + O(h))$$

οπότε

$$h_i^s |T(x_i)| = \left(\frac{\theta}{n}\right)^s (1 + O(h))$$

χρησιμοποιώντας τις σχέσεις (3.27) και (3.28) έχουμε το ζητούμενο

$$\left\| y_\pi - y \right\| \leq K\lambda^s(1 + O(h)) + O(h^p)$$

□

Παρατηρήσεις

- Εάν κάποιος θελήσει να περιορίσει το βήμα της διακριτοποίησης $h = \max_{1\leq i\leq n}(x_{i+1}-x_i)$ για τα επιλεγόμενα πλέγματα, τότε μπορεί να επιλέξει ως monitor συνάρτηση,

$$m(x, y(x)) = \max\left\{|T(x)|^{\frac{1}{s}}, v\right\}$$

για κάποιο v. Το προηγούμενο θεώρημα συνεχίζει να ισχύει και για αυτήν την επιλογή της monitor συνάρτησης αλλά οι σχέσεις του προηγούμενου θεωρήματος αλλάζουν ανάλογα.

- Δεν είναι πάντα προφανές πως η $T(x)$ μπορεί να υπολογιστεί στην πράξη. Εάν μια προσέγγιση της λύσης $y(x)$ είναι διαθέσιμη, τότε οι παράγωγοί της μπορούν να χρησιμοποιηθούν για να προσεγγίσουν της παραγώγους της λύσης που εμφανίζονται στην $T(x)$. Π.χ. παίρνοντας παραγώγους από μια κατάλληλη παρεμβολή, όπως θα συζητήσουμε στην επόμενη ενότητα.

- Η παράμετρος θ μπορεί να προσεγγιστεί χρησιμοποιώντας την monitor function. Από τις σχέσεις (3.27),(3.28),(3.31) και από την σχέση (3.30) μπορούμε να εκτιμήσουμε σε γενικές γραμμές την ποσότητα του σφάλματος $K\lambda^s$, οπότε δεδομένου ενός σφάλματος ανοχής TOL θα πρέπει $N \geq \theta\Big(\frac{K}{TOL}\Big)^{\frac{1}{s}}$. Συνεπώς θα μπορέσουμε να χρησιμοποιήσουμε την ποσότητα $\theta\Big(\frac{K}{TOL}\Big)^{\frac{1}{s}}$ ως προβλεπόμενη τιμή για το μέγεθος του νέου πλέγματος. Γνωρίζοντας την προβλεπόμενη τιμή για το μέγεθος του πλέγματος μπορούμε να υπολογίσουμε το λ, οπότε και το νέο πλέγμα π^* μπορεί να κατασκευαστεί από την σχέση (3.29).

- Είναι πολύ σημαντικό να παρατηρήσουμε ότι η μοναδική προϋπόθεση του θεωρήματος, είναι το πλέγμα π να ικανοποιεί την σχέση (3.26), δηλαδή να είναι το πλέγμα ασυμπτωτικά ομοιόμορφα κατανεμημένο. Αυτό στην πραγματικότητα σημαίνει ότι μπορούν να επιλέγουν και πολύπλοκες monitor συναρτήσεις και ακόμη το θεώρημα θα ισχύει, αλλά επίσης δεν είναι επιθυμητή μια ακριβής επίλυση για την κατασκευή του πλέγματος $\pi = \{x_i\}_{i=1}^{n+1}$, όπως επιβεβαιώνεται από την σχέση (3.29) του προηγούμενου θεωρήματος.

3.4 Adaptive Hermite Collocation

Στην προηγούμενη ενότητα δώσαμε μια γενική προσέγγιση επιλογής πλέγματος, η οποία επιλύει ένα Π.Σ.Τ., χρησιμοποιώντας ένα αριθμητικό σχήμα και προσαρμόζοντας το πλέγμα διακριτοποίησης σε κάθε βήμα. Η προσαρμογή του πλέγματος σε κάθε βήμα γίνεται με βάση την monitor συνάρτηση η οποία εμπλέκει την τρέχον προσεγγιστική λύση. Στην συνέχεια θα αναπτύξουμε μια στρατηγική για την πρακτική επιλογή πλέγματος χρησιμοποιώντας την μέθοδο της collocation. Ένας αποτελεσματικός αλγόριθμος προκύπτει εάν επιλέξουμε ως monitor συνάρτηση τον κύριο όρο του τοπικού σφάλματος. Προφανώς η στρατηγική που θα ακολουθήσουμε δεν είναι η μοναδική λογική, αλλά όμως θεωρούμε ότι είναι πολύ πρακτική και αποτελεσματική και οδηγεί σε πάρα πολύ καλά αριθμητικά αποτελέσματα, όπως θα δούμε στην τελευταία ενότητα αυτού του κεφαλαίου.

Αλγόριθμος πρακτικής επιλογής πλέγματος για την Collocation

Θεωρούμε για παράδειγμα ένα γραμμικό $2^{\alpha\varsigma}$ τάξεως Π.Σ.Τ.

$$Lu(x) = f(x) \qquad a < x < b$$
$$B_1 y(a) + B_2 y(b) = \beta$$

όπου $y(x) = (u(x), u'(x))^T$. Η μέθοδος που θα χρησιμοποιήσουμε για την κατασκευή της στρατηγικής μας θα είναι hermite cubic finite element orthogonal collocation. Έστω $u_\pi(x)$ η προσεγγιστική λύση, για ένα δεδομένο πλέγμα διακριτοποίησης π. Για την απόκτηση της $u_\pi(x)$ θα χρησιμοποιούνται ως collocation points σε κάθε element τα 2 gauss points του υποδιαστήματος. Η λύση $u_\pi(x)$ είναι τμηματική πολυωνυμική συνάρτηση με $u_\pi(x) \in P_{3,\pi} \cap C^1[a,b]$ και ικανοποιεί την διαφορική εξίσωση στα collocation points. Για $x_i \le x \le x_{i+1}$ το σφάλμα της προσεγγιστικής λύσης $u_\pi(x)$ από την πραγματική $u(x)$ δίνεται από την σχέση [ASC95]:

$$u^{(j)} - u_\pi^{(j)} = h_i^{4-j} u^{(4)}(x_i) P^{(j)}\Big(\frac{x - x_i}{h_i}\Big) + O(h_i^{5-j}) + O(h^4) \tag{3.32}$$

με $j = 0, 1, 2, 3$ και για $j = 0$ έχουμε,

$$u(x) - u_\pi(x) = h_i^4 u^{(4)}(x_i) P\Big(\frac{x - x_i}{h_i}\Big)(1 + O(h_i)) + O(h^4) \tag{3.33}$$

όπου

$$P(\xi) = \frac{1}{2}\int_0^\xi (t - \xi)\prod_{l=1}^{2}(t - p_l)dt$$

και p_l τα gauss points στο διάστημα $[0, 1]$. Από την σχέση αποκτούμε μια τοπική αναπαράσταση του σφάλματος. Οπότε

$$\big\|u(x) - u_\pi(x)\big\|_i = \max_{x_i \le x \le x_{i+1}} |u(x) - u_\pi(x)| \le C_i h_i^4 \big\|u^{(4)}(x)\big\|_i + O(h^4) \tag{3.34}$$

όπου

$$C_i = \widehat{C}(1 + O(h_i))$$

και

$$\widehat{C} = \max_{0 \le \xi \le 1} P(\xi) \tag{3.35}$$

Όπως παρατηρούμε ο κύριος όρος του σφάλματος εξαρτάται από τοπικές ποσότητες. Οπότε υπολογίζοντας με κάποιο τρόπο την $\big\|u^{(4)}(x)\big\|_i$ θα είχαμε μια εκτίμηση για το τοπικό σφάλμα. Συνεπώς, μια φυσική επιλογή για monitor function, σύμφωνα με όσα έχουμε ήδη αναφέρει σε προηγούμενες παραγράφους θα είναι:

$$\Big|u^{(4)}(x)\Big|^{1/4} \qquad\qquad x_i \le x \le x_{i+1} \tag{3.36}$$

Η ακριβής λύση $u(x)$ και συνεπώς η $u^{(4)}(x)$ είναι άγνωστες και δεν μπορούμε απευθείας να αντικαταστήσουμε την $u(x)$ από την $u_\pi(x)$ στην σχέση (3.36), διότι $u_\pi^{(4)}(x) \equiv 0$. Όμως μπορούμε με την εξής διαδικασία να εκτιμήσουμε την $u^{(4)}(x)$. Υποθέτουμε ότι η $v(x) \in P_{1,\pi} \cap C[a, b]$ είναι μια τμηματική γραμμική συνάρτηση η οποία παρεμβάλει την $u_\pi^{(3)}(x)$. Οπότε για τις εκτιμήσεις μας μπορούμε να ορίζουμε ως monitor συνάρτηση την:

$$m(x, v(x)) = \Big|v'(x)\Big|^{1/4} \qquad\qquad x_i \le x \le x_{i+1} \tag{3.37}$$

Τα σημεία που θα χρησιμοποιήσουμε για να παρεμβάλουμε γραμμικά την $v(x)$ στην $u_\pi^{(3)}$ θα είναι τα μέσα του κάθε υποδιαστήματος $[x_i, x_{i+1}]$,

$$v(x_{i+1/2}) = u_\pi^{(3)}(x_{i+1/2}) \qquad 1 \le i \le n$$

διότι τότε από την σχέση (3.32) έχουμε $P^{(3)}\Big(\frac{1}{2}\Big) = 0$ άρα,

$$v(x_{i+1/2}) = u_\pi^{(3)}(x_{i+1/2}) = u^{(3)}(x_{i+1/2}) + O(h_i^2) \qquad 1 \le i \le n$$

οπότε η προσέγγιση της $u^{(4)}(x)$ από την $v'(x)$ θα είναι τάξης $O(h)$. Δηλαδή,

$$v'(x) = u^{(4)}(x)\big(1 + O(h)\big).$$

Συνεπώς και η $m(x, v(x))$ θα είναι $O(h)$ προσέγγιση της monitor συνάρτησης (3.36). Αφού η $v'(x)$ είναι τμηματικά σταθερή συνάρτηση, ο υπολογισμός της ποσότητας

$$\theta = \int_a^b \Big|v'(x)\Big|^{1/4} dx \tag{3.38}$$

είναι μια εύκολη υπόθεση. Άρα ο υπολογισμός του νέου *as.eq.* πλέγματος διακριτοποίησης μπορεί να γίνει από την σχέση,

$$\int_{x_i}^{x_{i+1}} \Big|v'(x)\Big|^{1/4} dx \equiv \frac{\theta}{n} \tag{3.39}$$

και η συνάρτηση $t(x)$,

$$t(x) = \frac{1}{\theta}\int_a^x \Big|v'(\xi)\Big|^{1/4} d\xi \tag{3.40}$$

θα είναι τμηματικά γραμμική. Κατά συνέπεια εάν γνωρίζουμε το x_i τότε μπορούμε να υπολογίσουμε το x_{i+1} ώστε,

$$t(x_{i+1}) = \frac{i}{n} \tag{3.41}$$

Σύμφωνα με το θεώρημα (3.4), εάν π^* είναι το νέο εκλεπτυσμένο πλέγμα διακριτοποίησης που προκύπτει από την σχέση (3.40), το σφάλμα της προσεγγιστικής collocation λύσης από την πραγματική θα ικανοποιεί την σχέση:

$$\left\|u(x) - u_\pi(x)\right\|_i \leq \widehat{C}\Big(\frac{\theta}{n}\Big)^4\big(1 + O(h)\big) + O(h^4) \tag{3.42}$$

Οπότε ένα νέο μέγεθος πλέγματος n μπορεί να επιλέγει, ώστε η collocation λύση στο αντίστοιχο *as.eq.* πλέγμα να έχει ομοιόμορφο τοπικό σφάλμα μικρότερο από TOL. Από την σχέση (3.42) μπορούμε να προβλέψουμε το μέγεθος του πλέγματος διακριτοποίησης n, ώστε να μοιράσουμε ομοιόμορφα το τοπικό σφάλμα διαλέγοντας:

$$n = \theta\Big(\frac{\widehat{C}}{TOL}\Big)^{1/4} \tag{3.43}$$

όπου $\widehat{C}, \theta$ δίνονται από τις σχέσεις (3.35),(3.38) αντίστοιχα. Συνοψίζοντας όλα τα παραπάνω έχουμε τον ακόλουθο αλγόριθμο.

$**********Algorithm1**********$

`Input:` Ενα Π.Σ.Τ., ένα σφάλμα ανοχής TOL, ένα αρχικό πλέγμα π

`Output:` Μια κατάλληλη προσεγγιστική λύση u_π για το δεδομένο TOL

$**********$

1. `Flag = false`

2. `Repeat Until [ Flag=true ]`

   2.1 Λύσε το Π.Σ.Τ. και απόκτησε την collocation λύση u_π για το δεδομένο πλέγμα π.

   2.2 Κατασκεύασε την τμηματική γραμμική παρεμβολή $v(x)$ για την $u_\pi^{(3)}$, στα μέσα του κάθε υποδιαστήματος του τρέχον πλέγματος.

   2.3 Υπολόγισε την παράμετρο θ από την σχέση (3.38).

   2.4 `if` $\widehat{C}\Big(\frac{\theta}{n}\Big) \leq TOL$ `then`

```
            Flag=true
        else
            Βρες ένα νέο n από την σχέση (3.43).
            Κατασκεύασε ένα νέο πλέγμα από την σχέση (3.39),(3.40),(3.41).
        endif

    end Repeat
```

Στην συνέχεια θα τροποποιήσουμε τον παραπάνω αλγόριθμο, χρησιμοποιώντας κάποια επιπλέον κριτήρια, ώστε να εξασφαλιστεί μια αποτελεσματική πρακτική επιλογή πλέγματος. Για την αποτελεσματικότητα του νέου αλγορίθμου, τα σφάλματα θα εκτιμώνται και τα κριτήρια τερματισμού θα υπολογίζονται με δυο διαφορετικούς τρόπους. Για κριτήριο τερματισμού αλλά και ως κριτήριο εκτίμησης σφάλματος για ενδιάμεσα βήματα, θα χρησιμοποιούμε την σχέση (3.44), [ASC95]. Πιο συγκεκριμένα, όταν δυο προσεγγιστικές collocation λύσεις u_{π_1}, u_{π_2} υπολογιστούν στα πλέγματα π_1, π_2, όπου το πλέγμα π_2 προκύπτει από το π_1 όταν κάθε του υποδιάστημα διχοτομηθεί.

$$u(x) - u_{\pi_2}(x) = \frac{1}{15}(u_{\pi_1}(x) - u_{\pi_2}(x)) \tag{3.44}$$

Δεδομένης της προσεγγιστικής collocation λύσης υπολογίζουμε τις παρακάτω ποσότητες.

$$r_1 = \max_{1\le i\le n} h_i \Big(\frac{\widehat{C}\big\|v'(x)\big\|_i}{TOL}\Big)^{1/4} \qquad r_2 = \sum_{i=1}^{n} h_i \Big(\frac{\widehat{C}\big\|v'(x)\big\|_i}{TOL}\Big)^{1/4} \qquad r_3 = \frac{r_2}{n}$$

Δηλώνουμε ότι η ποσότητα r_1 παριστάνει ένα μέτρο εκτίμησης για το μέγιστο σφάλμα σε κάθε υποδιάστημα, ενώ η ποσότητα r_3 είναι ένα μέτρο εκτίμησης για το μέσο σφάλμα στα υποδιαστήματα. Ο λόγος $\dfrac{r_1}{r_3}$ είναι ένας δείκτης κατανομής του σφάλματος πάνω σε κάθε υποδιάστημα. Ειδικότερα εάν ο λόγος αυτός είναι μεγάλος, τότε η εκτίμηση για το μέγιστο σφάλμα των υποδιαστημάτων θα είναι σημαντικά μεγαλύτερο από εκείνο του μέσου σφάλματος και συνεπώς το πλέγμα μας δεν μπορεί να θεωρηθεί καλά κατανεμημένο. Οπότε ελέγχοντας τον λόγο αυτό έχουμε μια εκτίμηση, για την ποιότητα του επιλεγμένου πλέγματος διακριτοποίησης. Ένα πλέγμα θα θεωρείται ομοιόμορφα κατανεμημένο εάν ο λόγος,

$$\frac{r_1}{r_3} \le 2 \tag{3.45}$$

Εάν η σχέση (3.45) ικανοποιείται, τότε το πλέγμα θεωρείται ικανοποιητικά κατανεμημένο και δεν απαιτούμε μια νέα κατασκευή ενός νέου συνόλου σημείων $\{x_i\}$. Τέλος το τρέχον πλέγμα διπλασιάζεται

$$\pi^* = \left\{x_1, x_{1+1/2}, x_2, x_{2+1/2}, \ldots, x_n, x_{n+1/2}, x_{n+1}\right\}$$

και χρησιμοποιώντας την σχέση (3.44) αποκτούμε μια εκτίμηση για το σφάλμα της προσεγγιστικής λύσης. Οι διπλασιασμοί συνεχίζονται μέχρι να ικανοποιείται το δεδομένο TOL.

Εάν η σχέση (3.45) δεν ικανοποιείται, τότε η παράμετρος καθορίζει των αριθμό των σημείων που απαιτούνται, ώστε να είναι δυνατή η απόκτηση μιας προσεγγιστικής λύσης που να ικανοποιεί το δεδομένο TOL, όπως ακριβώς και στην σχέση (3.43). Στον νέο τροποποιημένο αλγόριθμο θα χρησιμοποιούμε ως προβλεπόμενη τιμή για το μέγεθος του νέου πλέγματος διακριτοποίησης.

$$n^* = \min\left\{n, \frac{1}{2}\max(n, r_2)\right\} \tag{3.46}$$

Με αυτόν τον τρόπο επιλογής του νέου μεγέθους n^* προστατεύεται ο αλγόριθμος από λανθασμένα συμπεράσματα νωρίς κατά την διαδικασία επιλογής πλέγματος. Στην συνέχεια το νέο πλέγμα π^* θα καθορίζεται από τις σχέσεις (3.40),(3.41). Επίσης, πρέπει να προστατεύσουμε την διαδικασία μας από τυχόν ανακυκλώσεις καθώς και από πρόωρα συμπεράσματα για τυχόν επιθυμητά πλέγματα. Για να το επιτύχουμε το παραπάνω πρέπει να εξασφαλίσουμε ότι το μέγεθος n θα αυξάνεται βαθμιαία. Η στρατηγική που θα ακολουθήσουμε είναι η εξής:

Πρόβλεψε το νέο μέγεθος του πλέγματος n^* από την σχέση (3.46), εάν όμως ισχύει μια από τις παρακάτω συνθήκες, τότε απέκτησε το νέο πλέγμα π^* διπλασιάζοντας το τρέχον πλέγμα π.

1. Εάν το νέο μέγεθος πλέγματος n^* είναι μικρότερο από το μέγεθος του πλέγματος προηγούμενου του π.

2. Εάν το τρέχον n έχει χρησιμοποιηθεί τρεις συνεχόμενες φορές.

3. Εάν έχουμε τρεις συνεχόμενες ομοιόμορφες κατανομές και διπλασιασμούς π.χ. $\dfrac{n}{2}$ και n σημεία πλέγματος χρησιμοποιήθηκαν τρεις συνεχόμενες φορές.

Τέλος, όταν η επιλογή $n^* = \frac{r_2}{2} < n$ τεθεί, τότε ένας διπλασιασμός ελπίζουμε να γίνει που θα μας προσφέρει μια προσεγγιστική λύση (καθώς μια εκτίμηση σφάλματος) που θα ικανοποιεί το δεδομένο TOL. Παρακάτω δίνουμε σχηματικά τον αλγόριθμο.

$**********Algorithm2**********$

`Input:` Ένα Π.Σ.Τ., ένα σφάλμα ανοχής TOL, ένα αρχικό πλέγμα π

`Output:` Μια κατάλληλη προσεγγιστική λύση u_π για το δεδομένο TOL

****

1. `Flag = false`

2. `Repeat Until [ Flag=true ]`

   2.1 Λύσε το Π.Σ.Τ. και απόκτησε την collocation λύση u_π για το δεδομένο πλέγμα π.

   2.2 `if` (π προκύπτει από διπλασιασμό) `then`
   -Έλεγχος κριτηρίου τερματισμού (3.44)
   αν ικανοποιείται τότε `Flag=true`.
   `else`
   -Έλεγχος κριτηρίου τερματισμού $\widehat{C}\left(\frac{\theta}{n}\right) \leq TOL$
   αν ικανοποιείται τότε `Flag=true`.
   `endif`

   2.3 Κατασκεύασε την τμηματική γραμμική παρεμβολή $v(x)$ για την $u_\pi^{(3)}$, στα μέσα του κάθε υποδιαστήματος του τρέχον πλέγματος.

   2.4 Υπολόγισε τις παραμέτρους r_1, r_2, r_3.

   2.5 `if` $r_1 \leq 2r_3$ `then`
   -Διπλασίασε το τρέχον πλέγμα μέχρι το κριτήριο

τερματισμού (3.44) να ικανοποιείται
και τότε θέσε `Flag=true`.
`else`
-Επίλεξε $n^* = \min\left\{n, \frac{1}{2}\max(n, r_2)\right\}$.
`if` (συνθήκες $1, 2, 3$ σελ. 94-5) `then`
-Διπλασίασε το τρέχον πλέγμα $n^* = 2n$.
`else`
-Κατασκεύασε ένα νέο πλέγμα για n^*
από την σχέση (3.39),(3.40),(3.41).
`endif`
`endif`

`end Repeat`

3.4.1 Αριθμητικά αποτελέσματα

Παράδειγμα 4 Θεωρούμε το πρόβλημα μοντέλο steady state advection-diffusion,

$$-\epsilon u''(x) + u'(x) = 1 \qquad 0 < x < 1$$

$$u(0) = u(1) = 0$$

η πραγματική λύση του προβλήματος είναι,

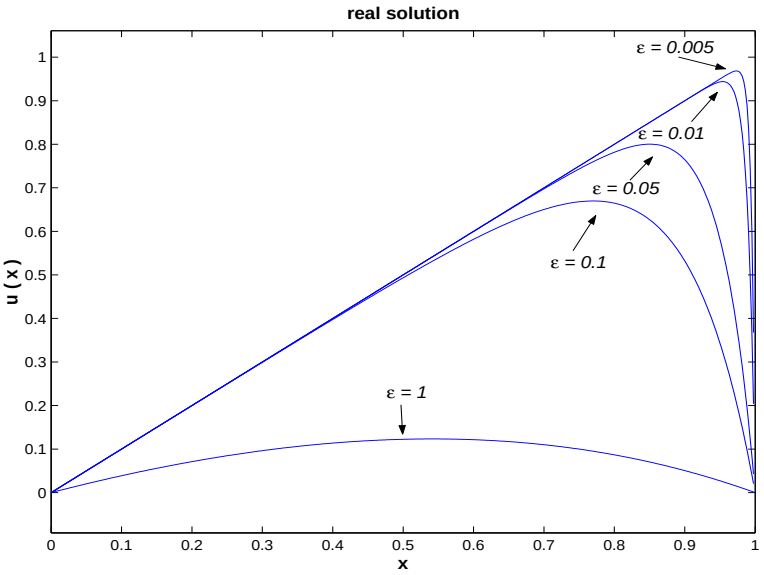
$$u(x) = -\frac{e^{\frac{x}{\epsilon}} - 1}{e^{\frac{1}{\epsilon}} - 1} + x$$

και για $0 < \epsilon \ll 1$ παρουσιάζει ένα boundary layer πλάτους $O(\epsilon)$ στο $x = 1$. Θα χρησιμοποιήσουμε τον Αλγόριθμο 2 για να επιλύσουμε το παραπάνω πρόβλημα για $\epsilon = 10^0, 10^{-1}, 5 \cdot 10^{-2}, 10^{-2}, 5 \cdot 10^{-3}$. Ως δεδομένο σφάλμα ανοχής θα έχουμε TOL=1e-07, και ως αρχικό πλέγμα διακριτοποίησης θα θεωρούμε μια ομοιόμορφη διαμέριση του $[0, 1]$ με μέγεθος $n = 5$.

Για $\epsilon = 10^0, 10^{-1}, 5 \cdot 10^{-2}, 10^{-2}, 5 \cdot 10^{-3}$ οι ακολουθίες πλεγμάτων που προκύπτουν μέχρι το επιθυμητό TOL να ικανοποιείται είναι:

$$\epsilon = 10^0 : 5 \quad 5 \quad 5 \quad 10$$

$$\epsilon = 10^{-1} : 5[1],\ 5[2],\ 5[2],\ 10[4],\ 10[5],\ 10[5],\ 20[9],\ 12[6],\ 12[5],\ 12[5],\ 24[10],\ 12[6],\ 24[11],\ 12[6],\ 24[11],\ 48[21]$$



Σχήμα 3.4: Πραγματική λύση για διάφορες τιμές του ϵ

$$\epsilon = 5 \cdot 10^{-2} : 5[2],\ 5[2],\ 5[2],\ 10[6],\ 10[6],\ 10[6],\ 20[12],\ 13[9],\ 13[8],\ 13[8],\ 26[16],\ 13[9],\ 26[17], 13[9],\ 26[17],\ 52[33]$$

$$\epsilon = 10^{-2} : 5[1],\ 5[1],\ 5[1],\ 10[1],\ 10[2],\ 10[7],\ 20[13],\ 13[9],\ 13[9],\ 13[9],\ 26[17],\ 13[9],\ 26[17],\ 13[9],\ 26[17],\ 52[34],\ 26[19],\ 52[37]$$

$$\epsilon = 5 \cdot 10^{-3} : 5[1],\ 5[1],\ 5[1],\ 10[2],\ 10[3],\ 10[5],\ 20[10],\ 11[5],\ 11[5],\ 11[6],\ 22[11],\ 11[6],\ 22[11],\ 18[13],\ 36[26],\ 18[14],\ 36[27],\ 18[14],\ 36[27],\ 72[54]$$

Όπου [.] αριθμούμε το πλήθος των σημείων του πλέγματος που βρίσκονται κοντά στην ευαίσθητη περιοχή του boundary layer. Στην περίπτωση όπου το $\epsilon = 10^{-1}, 5 \cdot 10^{-2}$ αριθμούμε το πλήθος των κόμβων που βρίσκονται στο διάστημα $(0.8, 1)$, ενώ για $\epsilon = 10^{-2}, 5 \cdot 10^{-3}$ αριθμούμε στο διάστημα $(0.9, 1)$.

Οι εκτιμήσεις σφάλματος μετά τους διπλασιασμούς στα ενδιάμεσα βήματα και στο τελικό βήμα, καθώς και το πραγματικό σφάλμα φαίνονται στους παρακάτω πίνακες.

$\epsilon = 1$	error estimate	real error
10	2.6892e-009	3.4303e-009

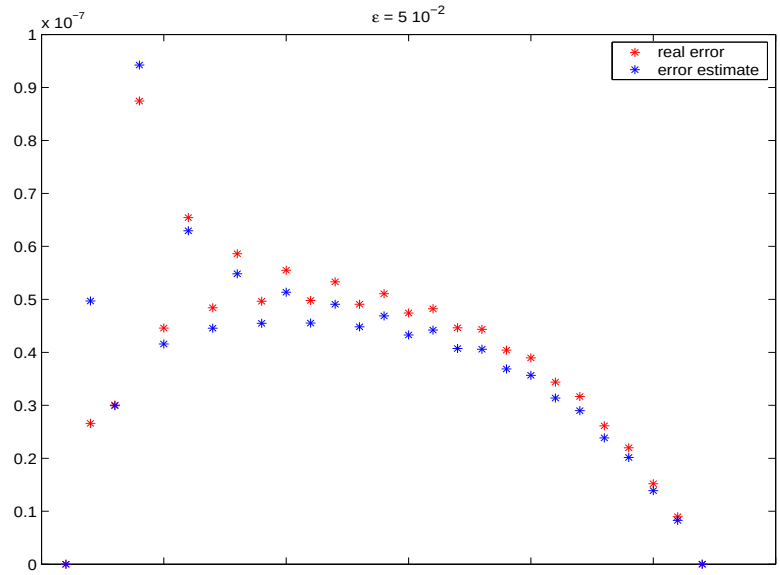
$\epsilon = 0.1$	error estimate	real error
10	3.5170e-005	4.6347e-005
20	1.2321e-006	1.7837e-006
24	4.7246e-007	6.6838e-007
24	1.0856e-007	4.8971e-007
24	1.9643e-007	5.0862e-007
48	3.1765e-008	3.3555e-008

$\epsilon = 0.05$	error estimate	real error
10	0.00012313	0.00013906
20	2.9148e-006	4.2377e-006
26	7.6896e-007	1.1074e-006
26	7.2975e-007	1.555e-006
26	4.8012e-007	1.5011e-006
52	9.4243e-008	1.146e-007

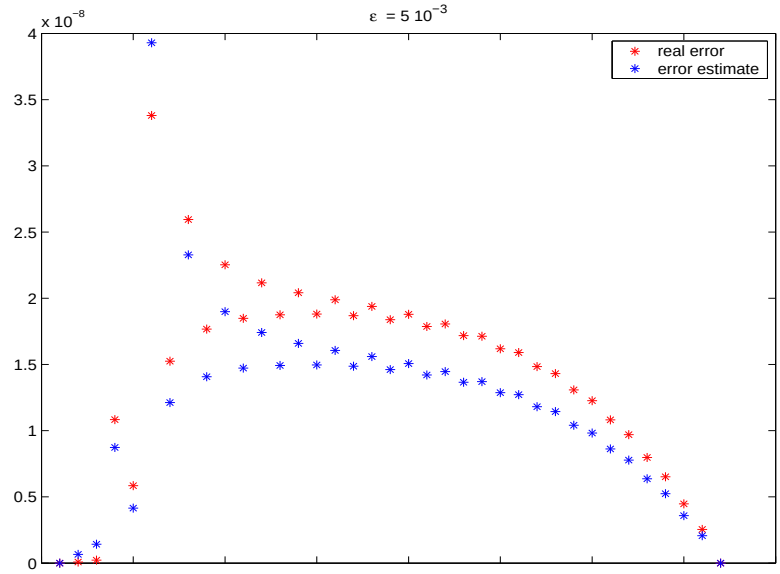
$\epsilon = 0.01$	error estimate	real error
10	0.2441	0.1817
20	8.5594e-006	8.808e-006
26	2.2802e-006	2.4323e-006
26	1.3828e-006	2.8999e-006
26	9.8085e-007	2.0115e-006
52	1.8857e-007	6.7602e-007
52	2.8252e-008	4.7521e-008

$\epsilon = 0.005$	error estimate	real error
10	0.025235	0.4906
20	0.0063072	0.015584
22	2.402e-005	3.1724e-005
22	8.7527e-007	2.3749e-006
36	7.1122e-007	2.7729e-006
36	1.5605e-007	4.2418e-007
36	1.1928e-007	6.2322e-007
72	3.9294e-008	6.3481e-008

Στην συνέχεια παριστάνουμε γραφικά τις εκτιμήσεις σφάλματος καθώς και το πραγματικό σφάλμα στο τελευταίο βήμα για $\epsilon = 0.05, 0.005$.

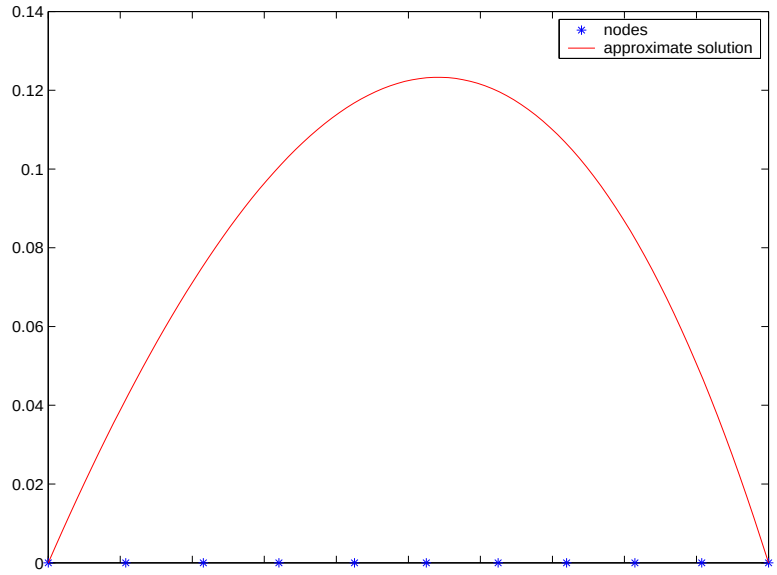


Σχήμα 3.5: Το πραγματικό και το σφάλμα εκτίμησης στο τελευταίο βήμα για $\epsilon = 0.05$

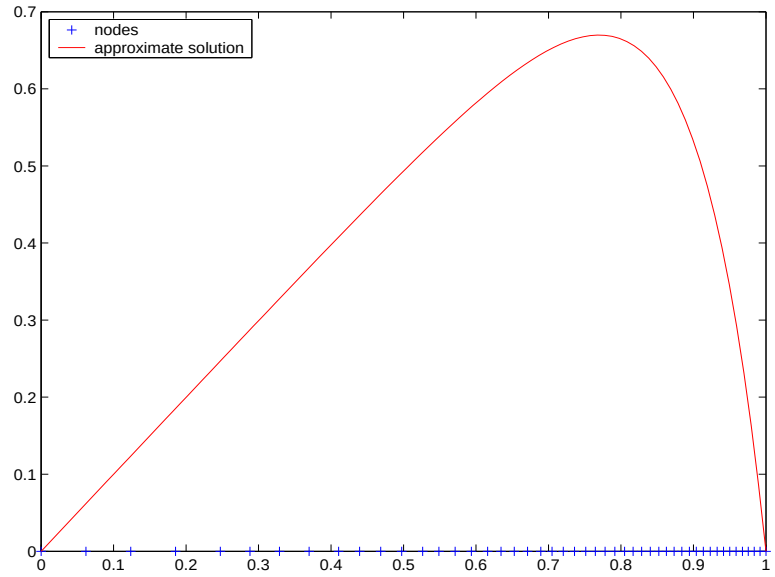


Σχήμα 3.6: Το πραγματικό και το σφάλμα εκτίμησης το τελευταίο βήμα για $\epsilon = 0.005$

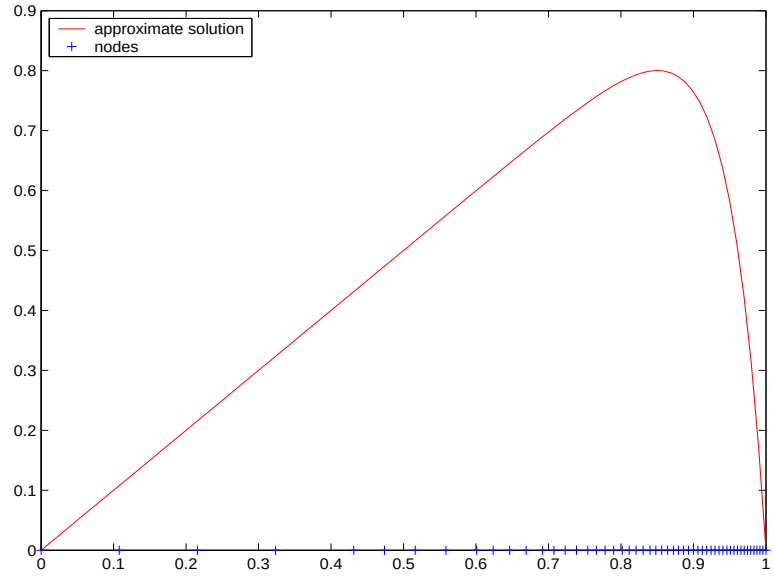
Τα πλέγματα τα επιλέχθηκαν ώστε το επιθυμητό σφάλμα TOL να ικανοποιείται είναι τα παρακάτω.



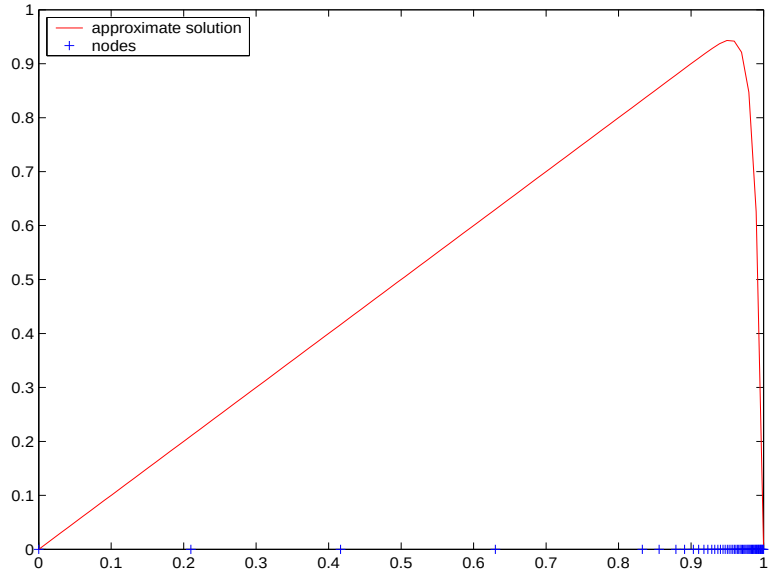
Σχήμα 3.7: Το τελικό πλέγμα διακριτοποίησης και η προσεγγιστική λύση πάνω σε αυτό για $\epsilon = 1$.



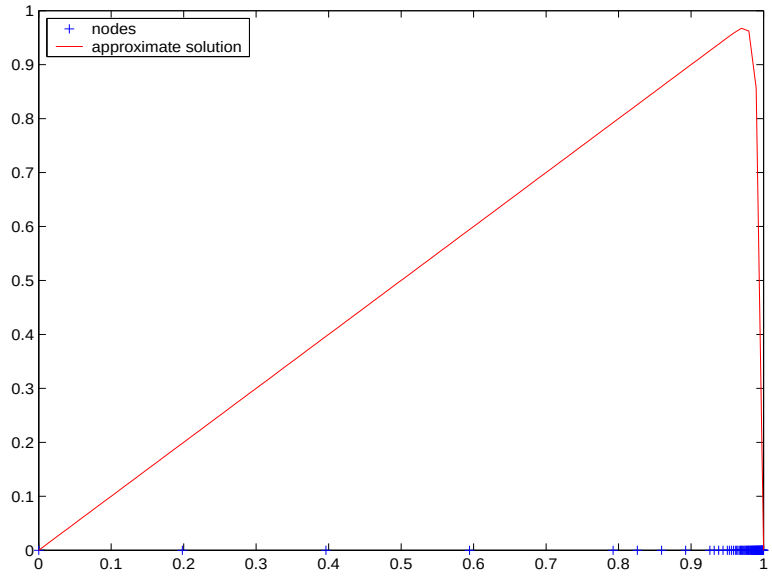
Σχήμα 3.8: Το τελικό πλέγμα διακριτοποίησης και η προσεγγιστική λύση πάνω σε αυτό για $\epsilon = 0.1$.



Σχήμα 3.9: Το τελικό πλέγμα διακριτοποίησης και η προσεγγιστική λύση πάνω σε αυτό για $\epsilon = 0.05$.



Σχήμα 3.10: Το τελικό πλέγμα διακριτοποίησης και η προσεγγιστική λύση πάνω σε αυτό για $\epsilon = 0.01$.



Σχήμα 3.11: Το τελικό πλέγμα διακριτοποίησης και η προσεγγιστική λύση πάνω σε αυτό για $\epsilon = 0.005$.

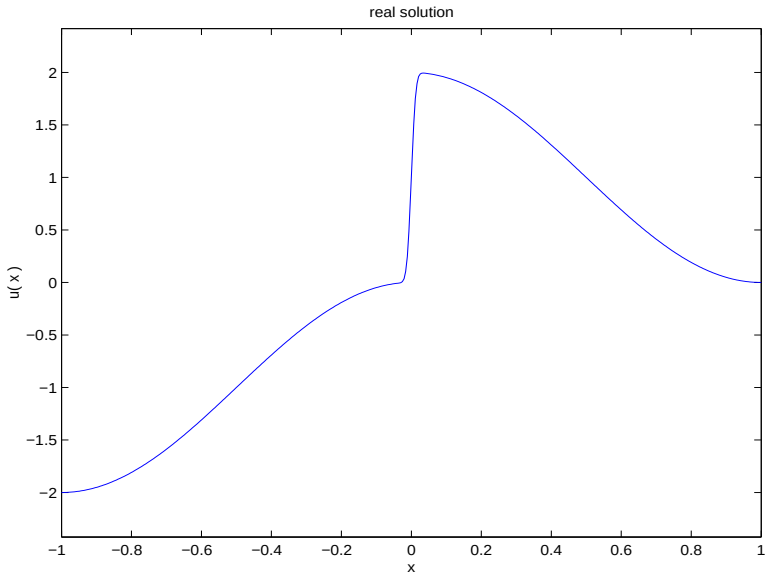
Παράδειγμα 5 Θεωρούμε το ιδιόμορφο πρόβλημα συνοριακών τιμών, όπου δεν ανήκει στην παραπάνω κατηγόρια των προβλημάτων μεταφοράς-διάχυσης.

$$\epsilon u''(x) + xu'(x) = -\epsilon \cos \pi x - (\pi x) \sin \pi x \qquad -1 < x < 1$$
$$u(-1) = -2 \qquad u(1) = 0$$

με πραγματική λύση

$$u(x) = \cos \pi x + erf(\frac{x}{\sqrt{2\epsilon}})/erf(\frac{1}{\sqrt{2\epsilon}})$$

παρατηρούμε ότι παρουσιάζεται ένα interior layer για $0 < \epsilon << 1$ στο $x = 0$.



Σχήμα 3.12: Γράφημα πραγματικής λύσης για $\epsilon = 10^{-4}$.

Η ακολουθία πλεγμάτων που προκύπτει κατά την επίλυση του Π.Σ.Τ. με $\epsilon = 10^{-4}$ για δεδομένο σφάλμα ανοχής TOL = 1e-006 και αρχική ομοιόμορφη διαμέριση του $[-1, 1]$ μεγέθους $n = 5$ είναι:

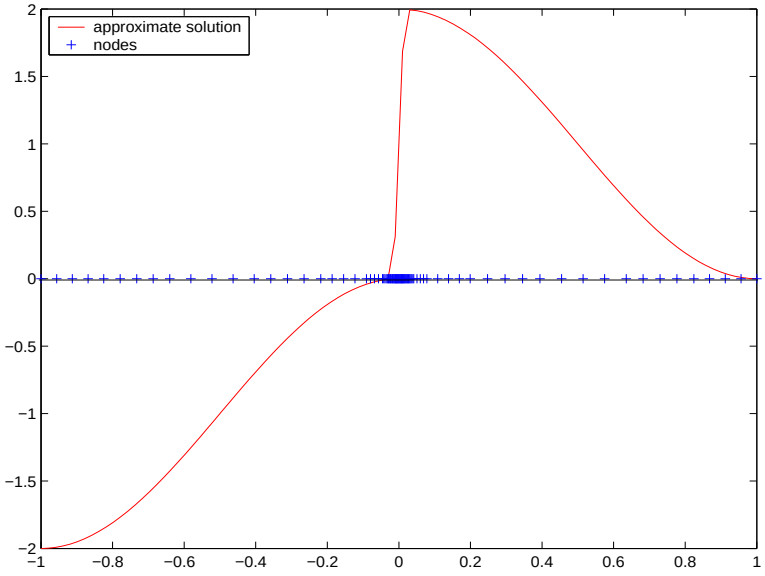
5[0], 5[0], 5[0], 10[1], 10[1], 10[1], 20[2], 20[3], 20[4], 40[8],
24[10], 48[21], 24[13], 48[25], 24[13], 48[25], 96[49]

όπου το πλήθος των κόμβων που βρίσκονται στο διάστημα $(-0.05, 0.05)$. Οι

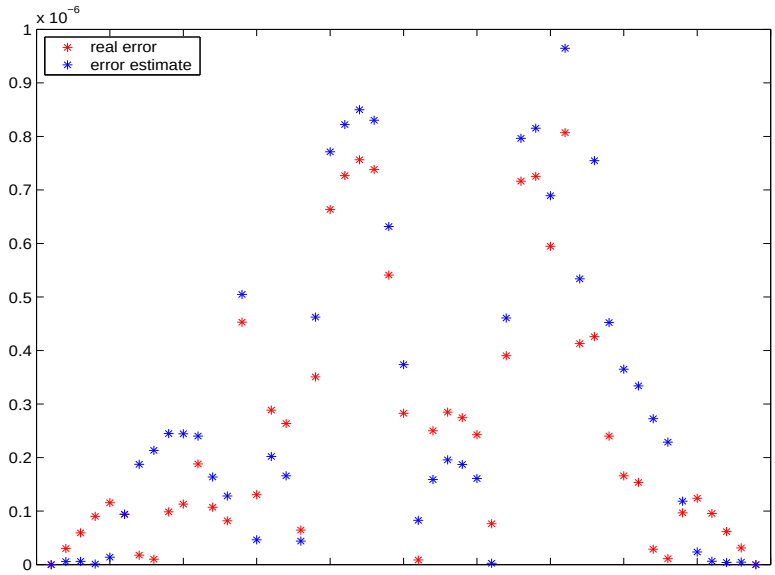
εκτιμήσεις σφάλματος μετά τους διπλασιασμούς στα ενδιάμεσα βήματα και στο τελικό βήμα, καθώς και το πραγματικό σφάλμα φαίνονται στον παρακάτω πίνακα.

$\epsilon = 10^{-4}$	error estimate	real error
10	0.026841	0.3918
20	0.016949	0.19388
40	0.0029048	0.0035868
48	4.9832e-005	5.0599e-005
48	1.1422e-005	1.3427e-005
48	1.3713e-005	1.5276e-005
96	9.6463e-007	1.9688e-006

Η τελική προσεγγιστική λύση καθώς και η τελευταία εκτίμηση σφάλματος παριστάνονται γραφικά.



Σχήμα 3.13: Γράφημα προσεγγιστικής λύσης για το τελικό επιλεγόμενο πλέγμα.



Σχήμα 3.14: Το πραγματικό και το σφάλμα εκτίμησης για το τελικό επιλεγόμενο πλέγμα.

Κεφάλαιο 4

Συμπεράσματα

Στο πρώτο κεφάλαιο αναδείξαμε τα προβλήματα που παρουσιάζει η μέθοδος της Orthogonal Collocation κατά την αριθμητική επίλυση προβλημάτων τύπου μεταφοράς-διάχυσης. Στη συνέχεια της εργασίας τροποποιήσαμε κατάλληλα την παραπάνω μέθοδο της Collocation, με δύο τρόπους, ώστε να επιλύει με ακρίβεια και με το ελάχιστο δυνατό κόστος προβλήματα της προηγούμενης μορφής.

Στο πρώτο μέρος της εργασίας, το οποίο παρουσιάζεται στο δεύτερο κεφάλαιο, καταφέρνουμε να ενσωματώσουμε στη μέθοδο της Hermite Cubic Spline Collocation upwinding χαρακτηριστικά. Αυτό γίνεται επιλέγοντας κατάλληλα τα collocation points μέσα από σύνολα, που εφοδιάζουν το αριθμητικό μας σχήμα με το ακόλουθο φυσικό χαρακτηριστικό,

$$\text{συντελεστής μεταφοράς} > 0 \Rightarrow Re(\lambda(T_1)) > 0$$

$$\text{συντελεστής μεταφοράς} < 0 \Rightarrow Re(\lambda(T_1)) < 0$$

όπου $Re(\lambda(T_1))$ το πραγματικό μέρος των ιδιοτιμών του πίνακα πρώτης τάξεως παραγώγισης T_1 της προσεγγιστικής λύσης. Τα παραπάνω σύνολα τα προσδιορίζονται θεωρητικά. Στη συνέχεια επιβεβαιώνοντας την προηγούμενη θεωρητική ανάλυση και αριθμητικά, αναδεικνύονται κάποιες πολύ σημαντικές πληροφορίες για τον τρόπο με τον οποίο πρέπει να επιλέγονται τα collocation points μέσα από σύνολα αυτά, ώστε με το μικρότερο δυνατό κόστος (δηλαδή μικρές διαμερίσεις) να έχουμε σφάλματα μέχρι και τάξεως $O(h^4)$.

Στο δεύτερο μέρος της εργασίας, το οποίο παρουσιάζεται στο τρίτο κεφάλαιο, αναζητούμε μια καλλίτερη διαμέριση του διαστήματος $I = [a, b]$, πάνω στο επιλύεται το πρόβλημα, ώστε η αριθμητική μέθοδος της Hermite Cubic Orthogonal Collocation να συμπεριφέρεται καλλίτερα.Γίνεται αναφορά σε adaptive h-refinement τεχνικές που προσαρμόζουν κατάλληλα ένα πλέγμα διακριτοποίησης, κατανέμοντας ομοιόμορφα ένα μέτρο εκτίμησης της συμπεριφοράς της λύσης. Κατόπιν βασιζόμενοι στις μεθόδους αυτές, κατανέμουμε ομοιόμορφα το τοπικό σφάλμα της αριθμητικής μεθόδου Hermite Cubic Orthogonal Collocation μέχρι το ολικό σφάλμα να είναι μικρότερο από ένα δεδομένο σφάλμα ανοχής TOL και γίνεται κατασκευή της επαναληπτικής adaptive h-refinement Hermite Cubic Orthogonal Collocation.

Βιβλιογραφία

[ASC86] U. Ascher, G. Bader, **Stability of Collocation at Gaussian Points**, *SIAM J. Numer. Anal. (23)*, 412-422, 1986.

[ASC88] U. M. Ascher, J. Christiansen, R. D. Russell, **Collocation Software for Boundary Value ODEs**, *ACM Transactions on Mathematical Software 7*, 209-222, 1988.

[ASC95] U. M. Ascher, R. M. M. Mattheij, R. D. Russell, **Numerical Solution of Boundary Value Promblems for Ordinary Differential Equations**, *SIAM*, 1995.

[BOO73] C. De. Boor, B. Swartz, **Collocation at Gaussian Points**, *SIAM J. Numer. Anal. (10)*, 582-606, 1973.

[BRI02] S. H. Brill, **Analytical Solution of Hermite Collocation Discretization of the One-Dimensional Steady-State Convection-Diffusion Equation**, *Int. J. of Diff. Eqns. and Applics. (4)*, 141-155, 2002.

[BRI04] S. H. Brill, **Optimal Upstream Collocation Solution of the One-Dimensional Steady-State Convection-Diffusion Equation**, *preprint*, 2004.

[HUN93] W. Huang, D. M. Sloan, **A New Psedospectral Method with Upwind features**, *IMA J. Numer. Anal. (13)*, 413-430, 1993.

[LEV98] R. J. LeVeque, **Finite Difference Methods for Differential Equations**, Washington 1998, lecture notes.

[MAH93] WD. Mahmood, M. R. Osborne, **Collocation for BVPs: Compact and Non-Compact Schemes**, *CTAC93*, World Scientific, Singapore 1994, 354-361, 1993.

[MEY00] C. D. Meyer, **Matrix Analysis and Applied Linear Algebra**, SIAM, 2000.

[RIN84] C. Ringhofer, **On Collocation Schemes for Quasilinear Singularly Perturbed Boundary Value Problems**, *SIAM J. Numer. Anal. (21)*, 864-882, 1984.

[RUS97] R. D. Russell, W. Sun, **Spline Collocation Differentiation Matrices**, *SIAM J. Numer. Anal. (34)*, 2274-2287, 1997.

[SUN99] W. Sun, **Spectral Analysis of Hermite Cubic Spline Collocation Systems**, *SIAM J. Numer. Anal. (36)*, 1962-1975, 1999.

[SUN00] W. Sun, **Hermite Cubic Spline Collocation Methods with Upwind features**, *ANZIAM*, 1962-1975, 2000.

[TRE97] L. N. Trefethen and D. Bau III , **Numerical Linear Algebra**, SIAM, 1997.

[WAN03] R. Wang, **High Order Adaptive Collocation Software for 1-D Parabolic PDEs**, *Phd. Thesis* , Dalhousie Univesity 2002.

[WEI88] J. A. C. Weideman, L. N. Trefethen, **The Eigenvalues of Second-Order Spectral Differentiation Martices**, *SIAM J. Numer. Anal. (25)*, 1279-1298, 1988.

[WHI79] A. B. White, **On Selection of Equidistributing Meshes for Two-Point Boundary-Value Problems**, *SIAM J. Numer. Anal. (16)*, 472- 502, 1979.

[WRI03] K. Wright, **Adaptive Methods for Piecewise Polynomial Collocation for Ordinary Differential Equations**, *Comp. and Math. with Applics.*, *(12)*, 1053-1059, 2003.