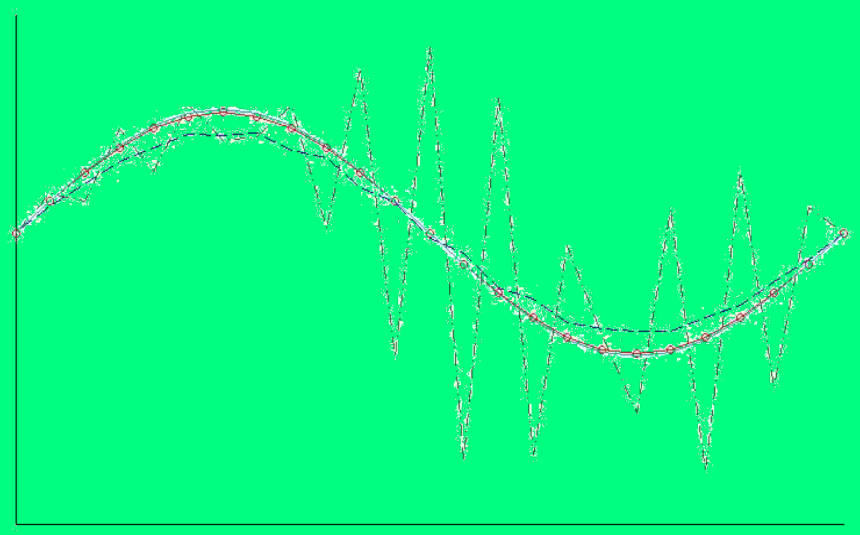
Μιχάλης Πλεξουσάκης
Παναγιώτης Χατζηπαντελίδης

Αριθμητική επίλυση μερικών διαφορικών εξισώσεων



Ελληνικά Ακαδημαϊκά Ηλεκτρονικά Συγγράμματα και Βοηθήματα
www.kallipos.gr

HEALLINK
Σύνδεσμος Ελληνικών Ακαδημαϊκών Βιβλιοθηκών

Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο

ΕΠΙΧΕΙΡΗΣΙΑΚΟ ΠΡΟΓΡΑΜΜΑ
ΕΚΠΑΙΔΕΥΣΗ ΚΑΙ ΔΙΑ ΒΙΟΥ ΜΑΘΗΣΗ
επένδυση στην κοινωνία της γνώσης
ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ ΚΑΙ ΘΡΗΣΚΕΥΜΑΤΩΝ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ
Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης

ΕΣΠΑ 2007-2013
ΕΥΡΩΠΑΪΚΟ ΚΟΙΝΩΝΙΚΟ ΤΑΜΕΙΟ

Αριθμητική Επίλυση Μερικών Διαφορικών Εξισώσεων

Μιχάλης Πλεξουσάκης και Παναγιώτης Χατζηπαντελίδης
Τμήμα Μαθηματικών και Εφαρμοσμένων Μαθηματικών
Πανεπιστήμιο Κρήτης

Αριθμητική Επίλυση Μερικών Διαφορικών Εξισώσεων

Συγγραφή
Μιχάλης Πλεξουσάκης
Παναγιώτης Χατζηπαντελίδης

Κριτικός Αναγνώστης
Γεώργιος Ακρίβης

Συντελεστές Έκδοσης
Γλωσσική Επιμέλεια : Αναστασία Τσιαδήμου
Τεχνική Επιμέλεια : Παναγιώτης Χατζηπαντελίδης
Γραφιστική Επιμέλεια : Παναγιώτης Χατζηπαντελίδης

ISBN: 978-960-603-481-7

Copyright ©ΣΕΑΒ, 2015



Το παρόν έργο αδειοδοτείται υπό τους όρους της άδειας Creative Commons Αναφορά Δημιουργού - Μη Εμπορική Χρήση - Παρόμοια Διανομή 3.0. Για να δείτε ένα αντίγραφο της άδειας αυτής επισκεφτείτε τον ιστότοπο

https://creativecommons.org/licenses/by-nc-sa/3.0/gr/

Σύνδεσμος Ελληνικών Ακαδημαϊκών Βιβλιοθηκών
Εθνικό Μετσόβιο Πολυτεχνείο
Ηρώων Πολυτεχνείου 9, 15780 Ζωγράφου

http://www.kallipos.gr

Περιεχόμενα

Κατάλογος Σχημάτων **vii**

Κατάλογος Πινάκων **x**

Ακρωνύμια **xiii**

1 Εισαγωγή **1**
1.1 Διαφορικές εξισώσεις 1
1.2 Επίλυση προβλημάτων συνοριακών ή και αρχικών τιμών 6
1.2.1 Πρόβλημα δύο σημείων 6
1.2.2 Εξίσωση της θερμότητας 7
1.2.3 Εξίσωση μεταφοράς 10
1.2.4 Εξίσωση κύματος 12
1.2.5 Ελλειπτική εξίσωση 13
1.3 Σειρές Fourier 15

2 Πεπερασμένες διαφορές **21**
2.1 Προσέγγιση πρώτης παραγώγου 21
2.2 Προσέγγιση δεύτερης παραγώγου 26
2.3 Ασκήσεις 29

3 Πρόβλημα δύο σημείων **33**
3.1 Διακριτοποίηση 33
3.2 Επίλυση τριδιαγώνιου γραμμικού συστήματος 36
3.3 Σύγκλιση της μεθόδου πεπερασμένων διαφορών 38
3.4 Μέθοδος ενέργειας 41
3.5 Συνοριακές συνθήκες Neumann 45
3.6 Ένα γενικότερο πρόβλημα 49
3.7 Μη ομογενείς συνοριακές συνθήκες 53
3.8 Άλλες συνοριακές συνθήκες 54
3.9 Ταινίες γραφικών παραστάσεων 55

3.10 Ασκήσεις 55

4 Πρόβλημα δύο σημείων: Πεπερασμένα Στοιχεία 59
4.1 Μεταβολικό πρόβλημα 59
4.2 Συνθήκες Dirichlet 62
4.3 Συνθήκες Neumann 68
4.4 Υλοποίηση της μεθόδου Galerkin 69
4.5 Κυβικές splines 72
4.6 Άλλες μέθοδοι και προβλήματα 73
4.7 Ασκήσεις 73

5 Εξίσωση θερμότητας: Πεπερασμένες διαφορές 75
5.1 Άμεση μέθοδος του Euler 75
5.2 Πεπλεγμένη μέθοδος του Euler 84
5.3 Μέθοδος των Crank-Nicolson 90
5.4 Άλλες μέθοδοι και προβλήματα 97
5.5 Ταινίες γραφικών παραστάσεων 97
5.6 Ασκήσεις 98

6 Εξίσωση θερμότητας: Πεπερασμένα στοιχεία 103
6.1 Ημιδιακριτό πρόβλημα 103
6.2 Πεπλεγμένη μέθοδος του Euler 108
6.3 Μέθοδος των Crank–Nicolson 111
6.4 Άμεση μέθοδος του Euler 114
6.5 Άλλες μέθοδοι και προβλήματα 118
6.6 Ασκήσεις 118

7 Εξίσωση μεταφοράς 121
7.1 Μέθοδοι upwind και downwind 121
7.1.1 Μέθοδος upwind 122
7.1.2 Μέθοδος downwind 124
7.1.3 Χωρίο υπολογιστικής εξάρτησης 128
7.1.4 Ευστάθεια και σύγκλιση 130
7.2 Μέθοδος των Lax–Wendroff 133
7.2.1 Μια μη ευσταθής μέθοδος 133
7.2.2 Μέθοδος Lax–Wendroff 135
7.2.3 Ευστάθεια και σύγκλιση 138
7.3 Άλλες μέθοδοι και προβλήματα 141
7.4 Ταινίες γραφικών παραστάσεων 142
7.5 Ασκήσεις 147

8 Εξίσωση κύματος **149**
8.1 Μέθοδος των Courant, Friedrichs και Lewy 149
8.2 Άλλες μέθοδοι και προβλήματα 157
8.3 Ταινίες γραφικών παραστάσεων 157
8.4 Ασκήσεις 160

9 Ελλειπτική εξίσωση: Πεπερασμένες διαφορές **161**
9.1 Ελλειπτική εξίσωση 161
9.2 Σύγκλιση 165
9.3 Άλλες μέθοδοι και προβλήματα 166
9.4 Ασκήσεις 167

10 Πεπερασμένα στοιχεία στις πολλές διαστάσεις **169**
10.1 Μεταβολικό πρόβλημα 169
10.2 Η μέθοδος Galerkin για ελλειπτικά προβλήματα 176
10.3 Η εξίσωση της θερμότητας 180
10.3.1 Ημιδιακριτή προσέγγιση Galerkin 181
10.3.2 Πλήρως διακριτές προσεγγίσεις Galerkin 182
10.4 Άλλες μέθοδοι και προβλήματα 185
10.5 Ασκήσεις 185

Λεξικό Αγγλικών Όρων **189**

Κατάλογος σχημάτων

1.1 Η ακριβής λύση u του Παραδείγματος 1.1 9
1.2 Οι χαρακτηριστικές ευθείες $x - \alpha t = c$ της εξίσωσης μεταφοράς . 10
1.3 Η αρχική συνθήκη g και η ακριβής λύση u του Παραδείγματος 1.2 11
1.4 Η ακριβής λύση u του Παραδείγματος 1.3 14
1.5 Η ακριβής λύση u του Παραδείγματος 1.4 15
1.6 Η συνάρτηση $f(x) = |x|$ και τα πεπερασμένα αθροίσματα $S_n(x)$ της σειράς Fourier με $n = 1, 3,$ και 5 18

2.1 Γεωμετρική ερμηνεία της $\delta_h^+ f(x_0)$ 22
2.2 Γεωμετρική ερμηνεία της $\delta_h^- f(x_0)$ 23
2.3 Γεωμετρική ερμηνεία της $\delta_h^c f(x_0)$ 24

3.1 Η ακριβής και οι προσεγγιστικές λύσεις του Παραδείγματος 3.1 . 39
3.2 Οι προσεγγιστικές λύσεις του Παράδειγματος 3.4 53
3.3 Ταινίες των γραφικών παραστάσεων των Παραδειγμάτων 3.1 και 3.4 55

5.1 Τα σημεία (x_i, t^j) του χωρίου $[0, L] \times [0, T]$ στα οποία αναζητούμε τις τιμές U_i^j 76
5.2 Σημεία του πλέγματος που χρησιμοποιούμε την άμεση μέθοδο του Euler 79
5.3 Η ακριβής λύση u του Παραδείγματος 5.1 και οι προσεγγίσεις της με την άμεση μέθοδο του Euler 84
5.4 Σημεία του πλέγματος που χρησιμοποιούμε την πεπλεγμένη μέθοδο του Euler 86
5.5 Η ακριβής λύση u του Παραδείγματος 5.1 και οι προσεγγίσεις της με την πεπλεγμένη μέθοδο του Euler 89
5.6 Σημεία του πλέγματος που χρησιμοποιούμε τη μέθοδο των Crank-Nicolson 93
5.7 Η ακριβής λύση u του Παραδείγματος 5.1 και οι προσεγγίσεις της με τη μέθοδο των Crank-Nicolson 95

5.8 Ταινίες των γραφικών παραστάσεων των Παραδείγματων 5.1–5.3 για την άμεση και πεπλεγμένη μέθοδο του Euler και τη μέθοδο των Crank-Nicolson 98

7.1 Σημεία ενός πλέγματος για τον υπολογισμό της λύσης με τη μέθοδο upwind 123
7.2 Η ακριβής λύση u του Παραδείγματος 7.1 και οι προσεγγίσεις της με τη μέθοδο upwind 124
7.3 Σημεία ενός πλέγματος για τον υπολογισμό της λύσης με τη μέθοδο downwind 125
7.4 Η ακριβής λύση u του Παραδείγματος 7.2 και οι προσεγγίσεις της με τη μέθοδο downwind 127
7.5 Η ακριβής λύση u του Παραδείγματος 7.3 και οι προσεγγίσεις της με τη μέθοδο downwind 128
7.6 Χωρίο εξάρτησης της προσεγγιστικής λύσης με τη μέθοδο upwind 129
7.7 Χωρίο εξάρτησης της προσεγγιστικής λύσης με τη μέθοδο downwind 130
7.8 Σημεία ενός πλέγματος για τον υπολογισμό της λύσης με τη μέθοδο (7.20) 134
7.9 Χωρίο εξάρτησης της προσεγγιστικής λύσης με τις μεθόδους (7.21) και Lax–Wendroff 135
7.10 Η ακριβής λύση u του Παραδείγματος 7.4 και οι προσεγγίσεις της με τη μέθοδο Lax-Wendroff 139
7.11 Η ακριβής λύση u του Παραδείγματος 7.4 και οι προσεγγίσεις της με τη μέθοδο Lax-Wendroff 140
7.12 Ταινίες των γραφικών παραστάσεων του Παραδείγματος 7.1 με $\alpha > 0$ για τη μέθοδο upwind 143
7.13 Ταινίες των γραφικών παραστάσεων του Παραδείγματος 7.2 με $\alpha > 0$ για τη μέθοδο downwind 144
7.14 Ταινίες των γραφικών παραστάσεων του Παραδείγματος 7.3 με $\alpha < 0$ για τη μέθοδο downwind 145
7.15 Ταινίες των γραφικών παραστάσεων του Παραδείγματος 7.4 με $\alpha > 0$ για τη μέθοδο των Lax–Wendroff 146

8.1 Σημεία ενός πλέγματος για τον υπολογισμό της λύσης με τη μέθοδο των Courant, Friedrichs και Lewy 152
8.2 Χωρίο εξάρτησης της προσεγγιστικής λύσης με τη μέθοδο των Courant, Friedrichs και Lewy 153
8.3 Η ακριβής λύση u του Παραδείγματος 8.1 και οι προσεγγίσεις της με τη μέθοδο των Courant, Friedrichs και Lewy 156

8.4 Ταινίες των γραφικών παραστάσεων του Παραδείγματος 8.1 με τη μέθοδο των Courant, Friedrichs και Lewy 158
8.5 Ταινίες των γραφικών παραστάσεων του Παραδείγματος 8.2 με τη μέθοδο των Courant, Friedrichs και Lewy 159

9.1 Τα σημεία (x_i, y^j) του χωρίου Ω όπου αναζητούμε τις τιμές U_i^j . . 162
9.2 Σημεία ενός πλέγματος για τον υπολογισμό της λύσης με τη μέθοδο (9.5) 163

10.1 Μια υποδιαίρεση ενός τετραγώνου (αριστερά) και ένας τριγωνισμός του (δεξιά). 177
10.2 Αρίθμηση των κόμβων ενός τριγωνισμού του $[0, 1]^2$. 179

Κατάλογος πινάκων

2.1 Τιμές των προσεγγίσεων της $f'(1.1) = 1/1.1 \approx 0.90909$ 24
2.2 Τα σφάλματα $E_{\delta_h^+}$, $E_{\delta_h^-}$ και $E_{\delta_h^c}$ των προσεγγίσεων $\delta_h^+ f$, $\delta_h^- f$ και $\delta_h^c f$, αντίστοιχα, της $f'(1.1) = 1/1.1$ και η προσεγγιστική τάξη ακρίβειας p 26
2.3 Οι τιμές της $\delta_{h,2}^c$ και το σφάλμα $E_{\delta_{h,2}^c}$, της προσέγγισης της $f''(1.1) = -1/(1.1)^2 \approx -0.82645$, καθώς και η προσεγγιστική τάξη ακρίβειας p 29

3.1 Τα σφάλματα του Παραδείγματος 3.1 στα σημεία x_1 και x_3 και το μέγιστο σφάλμα σε κάθε σημείο του διαμερισμού 39
3.2 Το σφάλμα της λύσεως του (3.18) στο Παράδειγμα 3.1 και η κατά προσέγγιση τάξη ακρίβειας p 41
3.3 Το σφάλμα της λύσεως του (3.52) στο Παράδειγμα 3.3 και η κατά προσέγγιση τάξη ακρίβειας p 49

5.1 Το μέγιστο σφάλμα $\bar{E}^j$ του Παραδείγματος 5.1 83
5.2 Το σφάλμα $\bar{E}^j$ του Παραδείγματος 5.2 90
5.3 Το σφάλμα $\bar{E}^j$ του Παραδείγματος 5.3 96
5.4 Τα σφάλματα των μεθόδων Euler και Crank–Nicolson $\bar{E}_{BE}$ και $\bar{E}_{CN}$, αντίστοιχα, στο Παράδειγμα 5.4 και η κατά προσέγγιση τάξη ακρίβειάς τους p 97

7.1 Τα σφάλματα των μεθόδων upwind και Lax–Wendroff του Παραδείγματος 7.5, καθώς και η κατά προσέγγιση τάξη ακρίβειάς τους p 141

8.1 Το σφάλμα που προκύπτει για το Παράδειγμα 8.2 με τη μεθόδο των Courant, Friedrichs και Lewy και η κατά προσέγγιση τάξη ακρίβειας p 155

Ακρωνύμια

Δ.Ε. Διαφορική Εξίσωση.

Μ.Δ.Ε. Μερική Διαφορική Εξίσωση.

Σ.Δ.Ε. Συνήθης Διαφορική Εξίσωση.

Κεφάλαιο 1

Εισαγωγή

Θα παρουσιάσουμε τις διαφορικές εξισώσεις και τα αντίστοιχα προβλήματα αρχικών και συνοριακών τιμών που θα συναντήσουμε στα επόμενα κεφάλαια. Επίσης, θα δούμε ορισμένες ιδιότητες και συμπεράσματα για τη λύση αυτών των προβλημάτων που θα χρειαστούν για την κατανόηση των αριθμητικών μεθόδων που θα χρησιμοποιήσουμε.

1.1 Διαφορικές εξισώσεις

Διαφορική Εξίσωση (Δ.Ε.) καλούμε μια εξίσωση η οποία συσχετίζει μια άγνωστη παραγωγίσιμη συνάρτηση u με κάποιες από τις παραγώγους της. Αν η u είναι συνάρτηση μιας μεταβλητής, τότε η διαφορική εξίσωση καλείται *Συνήθης Διαφορική Εξίσωση* (Σ.Δ.Ε.). Ένα απλό παράδειγμα είναι η εξίσωση

$$u'(t) = u(t), \quad t \in \mathbb{R}. \tag{1.1}$$

Αν η άγνωστη συνάρτηση u είναι δύο ή περισσοτέρων μεταβλητών, τότε η διαφορική εξίσωση συσχετίζει τη u με κάποιες από τις μερικές παραγώγους της και καλείται *Διαφορική Εξίσωση με Μερικές Παραγώγους* ή *Μερική Διαφορική Εξίσωση* (Μ.Δ.Ε). Παραδείγματος χάριν, αν συμβολίσουμε με u_{xx} και u_{yy} τις δεύτερες μερικές παραγώγους ως προς τις μεταβλητές x και y, αντίστοιχα, τότε ένα παράδειγμα Μ.Δ.Ε. είναι η

$$u_{xx}(x, y) + u_{yy}(x, y) = 0.$$

Κάθε ομαλή συνάρτηση ϕ που ικανοποιεί μια διαφορική εξίσωση καλείται *λύση* αυτής της εξίσωσης. Παραδείγματος χάριν, όλες οι συναρτήσεις $\phi(t) = ce^t$, όπου c είναι μια αυθαίρετη σταθερά, αποτελούν λύσεις της (1.1). Επομένως, είναι

δυνατόν να υπάρχουν άπειρες συναρτήσεις οι οποίες μπορούν να ικανοποιούν μια διαφορική εξίσωση.

Αν η υψηλότερη σε τάξη παράγωγος, που εμφανίζεται σε μία Σ.Δ.Ε. είναι κ, τότε λέμε ότι η Σ.Δ.Ε. είναι *τάξης* κ. Αντίστοιχα, μια Μ.Δ.Ε. λέμε ότι είναι *τάξης* κ, αν η μεγαλύτερη σε τάξη μερική παράγωγος που εμφανίζεται είναι κ.

Τη μερική παράγωγο μιας συνάρτησης μπορούμε να τη συμβολίσουμε με τον ακόλουθο τρόπο. Έστω $d \geq 1$ ένας ακέραιος αριθμός. Τότε, ορίζουμε ως πολυδείκτη $\alpha = (\alpha_1, \ldots, \alpha_d)$ ένα διάνυσμα του $\mathbb{R}^d$, όπου οι συνιστώσες α_i είναι μη αρνητικοί ακέραιοι. Το μήκος $|\alpha|$ του πολυδείκτη α ορίζεται ως $|\alpha| = \sum_{i=1}^{d} \alpha_i$. Για μια συνάρτηση $u : \mathbb{R}^d \to \mathbb{R}$, συμβολίζουμε τις μερικές παραγώγους τάξεως $|\alpha|$ ως προς τις μεταβλητές $x_1, \ldots, x_d$ με

$$D^\alpha u = \frac{\partial^{|\alpha|} u}{\partial x_1^{\alpha_1} \ldots \partial x_d^{\alpha_d}}.$$

Οι διαφορικές εξισώσεις κατηγοροποιούνται σε γραμμικές και μη γραμμικές εξισώσεις. Μια Μ.Δ.Ε. k τάξεως καλείται *γραμμική*, αν μπορεί να γραφεί στη μορφή

$$\sum_{|\alpha| \leq k} a_\alpha D^\alpha u = f, \tag{1.2}$$

όπου a_α, f είναι κατάλληλα ομαλές συναρτήσεις ανεξάρτητες της u, ορισμένες σε ένα υποσύνολο $\Omega \subset \mathbb{R}^d$, διαφορετικά καλείται *μη γραμμική*. Επίσης, αν η f είναι η μηδενική συνάρτηση, τότε η διαφορική εξίσωση (1.2) καλείται *ομογενής*, διαφορετικά καλείται *μη ομογενής*.

Εκτός από τον παραπάνω συμβολισμό της μερικής παραγώγου μιας συνάρτησης, θα χρησιμοποιήσουμε και δείκτες, π.χ.

$$u_t = \frac{\partial u}{\partial t}, \quad u_{xx} = \frac{\partial^2 u}{\partial x^2}, \quad u_{xy} = \frac{\partial^2 u}{\partial x \partial y}.$$

Στα επόμενα κεφάλαια θα θεωρήσουμε Μ.Δ.Ε., το πολύ, δεύτερης τάξεως, όπου η άγνωστη συνάρτηση u είναι δύο μεταβλητών, π.χ. x και y, και περιγράφεται από μια σχέση της μορφής

$$F(x, y, u, u_x, u_y, u_{xx}, u_{xy}, u_{yy}) = 0. \tag{1.3}$$

Μερικά παραδείγματα Μ.Δ.Ε. είναι τα ακόλουθα, ορισμένα από τα οποία θα δούμε στα επόμενα κεφάλαια:

1. Η εξίσωση της θερμότητας

$$u_t(t, x) - \alpha u_{xx}(t, x) = f(t, x), \quad \alpha > 0. \tag{1.4}$$

2. Η εξίσωση μεταφοράς

$$u_t(t,x) + \alpha u_x(t,x) = 0, \quad \alpha \in \mathbb{R},\ \alpha \neq 0. \tag{1.5}$$

3. Η εξίσωση του κύματος

$$u_{tt}(t,x) - \alpha^2 u_{xx}(t,x) = f(t,x), \quad \alpha > 0. \tag{1.6}$$

4. Η εξίσωση του Poisson

$$u_{xx}(x,y) + u_{yy}(x,y) = f(x,y). \tag{1.7}$$

5. Η εξίσωση του Burgers

$$u_t(t,x) + u(t,x)u_x(t,x) = 0. \tag{1.8}$$

6. Η εξίσωση του Korteweg-de Vries (KdV)

$$u_t(t,x) + u(t,x)u_x(t,x) + u_{xxx}(t,x) = 0. \tag{1.9}$$

Στα παραπάνω παραδείγματα, οι εξισώσεις του Laplace, κύματος, μεταφοράς και θερμότητας είναι *γραμμικές*, ενώ οι εξισώσεις των Burgers και Korteweg-de Vries είναι *μη γραμμικές*. Στα επόμενα κεφάλαια θα θεωρήσουμε και θα μελετήσουμε αριθμητικές μεθόδους για την επίλυση γραμμικών Μ.Δ.Ε.

Τις γραμμικές Μ.Δ.Ε. δεύτερης τάξεως μπορούμε να τις ταξινομήσουμε με τον ακόλουθο τρόπο, όπου για ευκολία θα θεωρήσουμε ότι η άγνωστη συνάρτηση u είναι συνάρτηση δύο μεταβλητών π.χ., x και y. Έστω Ω ένα υποσύνολο του $\mathbb{R}^2$. Τότε, μια γραμμική Μ.Δ.Ε. δεύτερης τάξεως μπορεί να γραφεί στη μορφή

$$a_{11}u_{xx} + a_{12}u_{xy} + a_{22}u_{yy} + a_1 u_x + a_2 u_y + a_0 u = f, \text{ στο } \Omega \subset \mathbb{R}^2, \tag{1.10}$$

όπου οι $a_{11}, a_{12}, a_{22}, a_1, a_{12}, a_0$ και f είναι ομαλές συναρτήσεις στο Ω, ανεξάρτητες της u. Με βάση το πρόσημο της ποσότητας

$$\mathcal{D} = a_{12}^2 - a_{11}a_{22},$$

στο σημείο $(x_0, y_0) \in \Omega$, η (1.10) καλείται υπερβολική, παραβολική ή ελλειπτική στο (x_0, y_0), αν η $\mathcal{D}$ είναι γνήσια θετική, μηδέν ή γνήσια αρνητική στο (x_0, y_0). Η (1.10) καλείται υπερβολική, παραβολική ή ελλειπτική στο $\Omega \subset \mathbb{R}^2$, αν είναι υπερβολική, παραβολική ή ελλειπτική σε κάθε σημείο $(x_0, y_0) \in \Omega$. Επομένως, η εξίσωση του κύματος (1.6) είναι υπερβολική, γιατί η αντίστοιχη ποσότητα $\mathcal{D} =$

$1 > 0$, η εξίσωση της θερμότητας (1.4) είναι παραβολική, γιατί $\mathcal{D} = 0$ και η εξίσωση του Laplace (1.7) είναι ελλειπτική, διότι $\mathcal{D} = -1 < 0$.

Εύκολα μπορούμε να δούμε ότι οι διαφορικές εξισώσεις μπορεί να έχουν πολλές λύσεις. Παραδείγματος χάριν, αν θεωρήσουμε την εξίσωση του Laplace (1.7) με $f = 0$, όλες οι συναρτήσεις της μορφής

$$u(x, y) = ax + by + c, \text{ με } a, b, c \in \mathbb{R},$$

αποτελούν λύσεις της (1.7). Επίσης, όλες οι συναρτήσεις της μορφής

$$u(x, t) = e^{-a^2 t} \sin(ax), \text{ με } a \in \mathbb{R},$$

αποτελούν λύσεις της εξίσωσης της θερμότητας (1.4) με $f = 0$.

Μια βασική ιδιότητα των γραμμικών διαφορικών εξισώσεων είναι η αρχή της *υπέρθεσης* ή *επαλληλίας*, σύμφωνα με την οποία αν έχουμε n λύσεις $u_1, u_2, \ldots, u_n$, μιας ομογενούς γραμμικής διαφορικής εξίσωσης, τότε και οποιοσδήποτε γραμμικός συνδυασμός τους

$$c_1 u_1 + \cdots + c_n u_n, \text{ με } c_1, \ldots, c_n \in \mathbb{R},$$

είναι και αυτός λύση της ίδιας διαφορικής εξίσωσης.

Ένας βασικός λόγος του ενδιαφέροντος που υπάρχει για τη μελέτη και την επίλυση μιας διαφορικής εξίσωσης είναι ότι χρησιμοποιούνται για την περιγραφή διάφορων φυσικών φαινόμενων. Έτσι, μπορούμε να βρούμε την τιμή της θερμοκρασίας μιας μεταλλικής ράβδου, το ύψος ενός κύματος, την ταλάντωση μιας χορδής ή την παραμόρφωση ενός σχοινιού λόγω ενός βάρους που κρεμάμε. Αν θεωρήσουμε ότι κάτω από τις ίδιες συνθήκες ένα φυσικό φαινόμενο πρέπει να δίνει το ίδιο αποτέλεσμα, τότε και η διαφορική εξίσωση που θέλουμε να περιγράφει αυτό το φαινόμενο πρέπει να έχει μοναδική λύση. Για αυτό τον λόγο, επιβάλουμε σε αυτή τη διαφορική εξίσωση βοηθητικές συνθήκες οι οποίες πρέπει να είναι τέτοιες ώστε να έχει μοναδική λύση. Αναλόγως με το φυσικό φαινόμενο που θέλουμε να περιγράψουμε, αυτές οι βοηθητικές συνθήκες καλούνται είτε *συνοριακές συνθήκες* είτε *αρχικές και συνοριακές συνθήκες*.

Συγκεκριμένα, μια *αρχική συνθήκη* καθορίζει την τιμή της λύσης σε μια καθορισμένη χρονική στιγμή t_0. Παραδείγματος χάριν, για την εξίσωση της θερμότητας, (1.4), μια αρχική συνθήκη είναι

$$u(x, t_0) = \phi(x), \text{ για } x \in [a, b]. \tag{1.11}$$

Αν υποθέσουμε ότι έχουμε ένα μοναδιάστατο σύρμα από το σημείο $x = a$ έως το σημείο $x = b$, τότε η λύση $u(x, t)$ της εξίσωσης (1.4) περιγράφει την κατανομή της θερμότητας στο σημείο x και στον χρόνο t, και η αρχική συνθήκη (1.11) περιγράφει

την κατανομή της θερμότητας στο $[a, b]$ στο χρόνο t_0. Αν τώρα υποθέσουμε ότι μας δίνονται δύο συναρτήσεις g και h, και διατηρήσουμε τη θερμοκρασία $u(x, t)$ ίση με $g(t)$ στο $x = a$ και ίση με $h(t)$ στο $x = b$, για $t \geq t_0$, τότε μπορούμε να θεωρήσουμε την ακόλουθη συνοριακή συνθήκη για την (1.4)

$$u(a, t) = g(t), \quad u(b, t) = h(t), \quad \text{για } t \geq t_0. \tag{1.12}$$

Οι συνθήκες (1.12) όπου προσδιορίζεται η τιμή της λύσης u στα άκρα του $[a, b]$ καλούνται *συνοριακές συνθήκες Dirichlet*. Αν οι τιμές στα άκρα είναι ίσες με μηδέν, τότε οι συνοριακές συνθήκες (1.12) καλούνται *ομογενείς συνοριακές συνθήκες Dirchlet*. Άλλες συνοριακές συνθήκες είναι οι *συνοριακές συνθήκες Neumann*, οι οποίες περιγράφονται από τις

$$u_x(a, t) = g(t), \quad u_x(b, t) = h(t), \quad \text{για } t \geq t_0, \tag{1.13}$$

όπου g, h είναι γνωστές συναρτήσεις. Παραδείγματος χάριν, αν μονώσουμε τα άκρα $x = a$ και $x = b$ του μονοδιάστατου σύρματος, τότε, επειδή δεν υπάρχει ροή θερμότητας από τα άκρα προς τα έξω, οι συναρτήσεις g και h στην (1.13) μηδενίζονται, δηλαδή οι συνοριακές συνθήκες Neumann γίνονται *ομογενείς*.

Υπάρχουν και άλλες συνοριακές συνθήκες, με τις οποίες όμως δεν θα ασχοληθούμε στα επόμενα κεφάλαια, όπως είναι οι *συνοριακές συνθήκες Robin*,

$$\begin{cases} -u_x(a, t) + \sigma_1 u(a, t) = 0, \\ \quad u_x(b, t) + \sigma_2 u(b, t) = 0, \end{cases} \quad \text{για } t \geq t_0, \ \text{ με } \sigma_1, \sigma_2 \geq 0 \tag{1.14}$$

και οι *περιοδικές συνοριακές συνθήκες*

$$u(a, t) = u(b, t), \quad u_x(a, t) = u_x(b, t), \ \text{για } t \geq t_0. \tag{1.15}$$

Για μεθόδους που υλοποιούν αυτές τις συνοριακές συνθήκες (1.14), (1.15), βλ. π.χ. στο (Ακρίβης & Δουγαλής, 2005).

Οι διαφορικές εξισώσεις της θερμότητας, της μεταφοράς, του κύματος και του Poisson, (1.4)–(1.7), θεωρούνται θεμελιώδεις εξισώσεις, γιατί η μελέτη τους οδηγεί στην κατανόηση και ανάλυση πολλών άλλων διαφορικών εξισώσεων. Στις επόμενες παραγράφους αυτού του κεφαλαίου θα παρουσιάσουμε ορισμένες ιδιότητες αυτών των διαφορικών εξισώσων και στα επόμενα κεφάλαια θα μελετήσουμε αριθμητικές μεθόδους για την προσέγγιση της λύσης τους. Μια πιο αναλυτική παρουσίαση των διαφορικών εξισώσεων που θα θεωρήσουμε υπάρχει π.χ. στα (Ακρίβης & Αλικάκος, 2012· Logan, 2014).

1.2 Επίλυση προβλημάτων συνοριακών τιμών και αρχικών και συνοριακών τιμών

Σε αυτή την παράγραφο θα δούμε ορισμένες ιδιότητες των διαφορικών εξισώσεων που θα συναντήσουμε, καθώς και των λύσεων τους.

1.2.1 Πρόβλημα δύο σημείων

Θα ασχοληθούμε πρώτα με μια απλή διαφορική εξίσωση σε μία διάσταση, το πρόβλημα δύο σημείων. Έστω u τέτοια, ώστε

$$-u''(x) + q(x)u(x) = f(x), \quad \text{για } x \in [a, b]. \tag{1.16}$$

Είναι γνωστό, βλ. π.χ. (Logan, 2014), ότι η γενική λύση u της (1.16) προκύπτει ως γραμμικός συνδυασμός δύο γραμμικώς ανεξαρτήτων λύσων της αντίστοιχης ομογενούς, u_1 και u_2, και μιας λύσης u_p της μη ομογενούς (1.16),

$$u = c_1 u_1 + c_2 u_2 + u_p,$$

όπου c_1, c_2 πραγματικές σταθερές. Αν θεωρήσουμε τώρα ότι η λύση u του (1.16) ικανοποιεί, επιπλέον, τις συνθήκες

$$u(a) = a_0, \quad u'(a) = a_1, \tag{1.17}$$

όπου a_0, a_1 δοσμένες σταθερές, τότε έχουμε ένα πρόβλημα αρχικών τιμών, το οποίο έχει μία ακριβώς λύση, σύμφωνα με το ακόλουθο θέωρημα, [βλ. π.χ. (Logan, 2014, Chapter 3.4)].

Θεώρημα 1.1. *Αν οι συναρτήσεις f, q είναι συνεχείς στο διάστημα $[a, b]$, τότε το πρόβλημα αρχικών τιμών* (1.16)–(1.17) *έχει μοναδική λύση στο $[a, b]$.*

Αν τώρα, αντί για τις αρχικές συνθήκες (1.17), θεωρήσουμε ότι η u ικανοποιεί τις *ομογενείς συνοριακές συνθήκες Dirichlet*,

$$u(a) = u(b) = 0, \tag{1.18}$$

τότε δεν είναι σίγουρο ότι υπάρχει μοναδική λύση u του προβλήματος (1.16) και (1.18), για κάθε συνεχή συνάρτηση f και q. Όμως, μπορούμε να δείξουμε το ακόλουθο θεώρημα, βλ. π.χ. (Brezis, 1997, Κεφάλαιο VIII).

Θεώρημα 1.2. *Αν οι συναρτήσεις f, q είναι συνεχείς στο διάστημα $[a, b]$ και $q \geq 0$, τότε έχει μοναδική λύση ακόλουθο πρόβλημα: Ζητείται $u : [a, b] \to \mathbb{R}$, τέτοια ώστε*

$$-u'' + qu = f, \quad \text{στο } [a, b], \ \text{με } u(a) = u(b) = 0, \tag{1.19}$$

Η υπόθεση ότι $q \geq 0$ είναι σημαντική για την ύπαρξη και μοναδικότητα της λύσης του (1.19). Αν θεωρήσουμε, για παράδειγμα, την

$$-u'' - \pi u = 0, \quad \text{στο } [0,1], \ \text{με } u(0) = u(1) = 0, \tag{1.20}$$

μπορούμε να δούμε ότι οι συναρτήσεις $c\sin(\pi x)$, $c \in \mathbb{R}$ αποτελούν λύσεις του (1.20) και επομένως, το πρόβλημα (1.20) έχει άπειρες λύσεις. Στη συνέχεια, αν τροποποιήσουμε το (1.20) και θεωρήσουμε το πρόβλημα

$$-u'' - \pi u = \pi, \quad \text{στο } [0,1], \ \text{με } u(0) = u(1) = 0 \tag{1.21}$$

από τη θεωρία των Σ.Δ.Ε., βλ. (Logan, 2014), έχουμε ότι η γενική λύση αυτής της Δ.Ε. είναι η $c_1 \sin(\pi x) + c_2 \cos(\pi x) - 1$, με $c_1, c_2 \in \mathbb{R}$. Όμως για καμία επιλογή των c_1, c_2 δεν προκύπτει λύση του (1.21). Συνεπώς, το (1.21) δεν έχει καμμία λύση.

Αν τροποποιήσουμε τώρα τις συνοριακές συνθήκες Dirichlet στο πρόβλημα (1.19) και θεωρήσουμε το αντίστοιχο πρόβλημα με *ομογενείς συνοριακές συνθήκες Neumann*, δηλαδή: Ζητείται, u τέτοια ώστε

$$-u'' + qu = f, \quad \text{στο } [a,b], \ \text{με } u'(a) = u'(b) = 0, \tag{1.22}$$

τότε, όπως και στο Θεώρημα 1.2 μπορούμε να δείξουμε μοναδικότητα της λύσης αν q, f είναι συνεχείς συναρτήσεις στο $[a,b]$ και $q > 0$. Διαφορετικά, αν π.χ. $q = f = 0$, τότε το (1.22) γίνεται

$$-u'' = 0, \quad \text{στο } [a,b], \ \text{με } u'(a) = u'(b) = 0. \tag{1.23}$$

Είναι προφανές ότι όλες οι συναρτήσεις της μορφής $u(x) = c$, με $c \in \mathbb{R}$ αποτελούν λύση και, άρα, έχουμε άπειρες λύσεις του προβλήματος.

1.2.2 Εξίσωση της θερμότητας

Θα θεωρήσουμε τώρα μια απλή παραβολική διαφορική εξίσωση και συγκεκριμένα την εξίσωση της θερμότητας (1.4), όπου η χωρική μεταβλητή x ανήκει στο διάστημα $[0, L]$, με $L > 0$ και θα θεωρήσουμε, επιπλέον, *αρχικές και ομογενείς συνοριακές συνθήκες Dirichlet*. Έτσι, έχουμε το ακόλουθο πρόβλημα: Ζητείται $u : [0, L] \times [0, +\infty) \to \mathbb{R}$, τέτοια ώστε

$$u_t = u_{xx}, \qquad \text{στο } [0, L] \times [0, +\infty) \tag{1.24}$$

$$u(0,t) = u(L,t) = 0, \qquad \text{για } t \geq 0, \tag{1.25}$$

$$u(x,0) = g(x), \qquad \text{στο } [0, L], \tag{1.26}$$

με g μια συνεχή συνάρτηση.

Ένας τρόπος για να λύσουμε την (1.24) είναι να υποθέσουμε ότι $u(x,t) = X(x)T(t)$, δηλαδή να θεωρήσουμε τη μέθοδο των χωριζομένων μεταβλητών. Σύμφωνα με αυτή τη μέθοδο, βλ. π.χ. (Logan, 2014), οι συναρτήσεις u_n, με $u_n = X_nT_n$, $n = 1, 2, \ldots$, όπου

$$\begin{aligned} u_n &= X_nT_n, \quad \text{με} \\ X_n(x) &= c\sin(\lambda_n x), \ \text{ και } T_n(t) = de^{-\lambda_n^2 t} \text{ με } \lambda_n = \frac{n\pi}{L},\ c, d \in \mathbb{R}, \end{aligned} \tag{1.27}$$

ικανοποιούν την εξίσωση (1.24) και τις συνοριακές συνθήκες (1.25). Λόγω της αρχής της υπέρθεσης, κάθε γραμμικός συνδυασμός λύσεων της μορφής (1.27) ικανοποιεί τις (1.24)-(1.25). Επομένως, αν θεωρήσουμε μια άπειρη σειρά συναρτήσεων της μορφής

$$u(x,t) = \sum_{n=1}^{\infty} c_n e^{-\lambda_n^2 t}\sin(\lambda_n x), \quad \text{με } c_n \in \mathbb{R}, \tag{1.28}$$

η οποία συγκλίνει για κάθε σημείο $x \in [0, L]$ και $t \geq 0$, τότε αποτελεί και λύση του προβλήματος (1.24)-(1.25). Μια λύση u της μορφής (1.28) θα ικανοποιεί και την αρχική συνθήκη (1.26), μόνο αν

$$u(x,0) = \sum_{n=1}^{\infty} c_n \sin(\lambda_n x) = g(x), \quad \text{με } c_n \in \mathbb{R}. \tag{1.29}$$

Ένα ανάπτυγμα της μορφής (1.29) καλείται σειρά Fourier ημιτόνων της g, βλ. π.χ. την Παράγραφο 1.3.

Είναι γνωστό ότι το προβλήμα (1.24)-(1.26) έχει μοναδική λύση, βλ. π.χ. (Logan, 2014).

Θεώρημα 1.3. *Αν η συνάρτηση g είναι συνεχής στο $[0, L]$, τότε υπάρχει μοναδική λύση u του προβλήματος* (1.24)–(1.26)*.*

Έστω τώρα ότι γνωρίζουμε το ανάπτυγμα στη μορφή (1.29) της συνάρτησης $g(x)$, της αρχικής συνθήκης (1.26), δηλαδή είναι γνωστές οι σταθερές c_n, $n = 1, 2, \ldots$, στην (1.29). Τότε, η λύση u του προβλήματος (1.24)–(1.26) δίνεται από την (1.28), με σταθερές c_n αυτές της (1.29).

Παρόμοια αποτελέσματα ισχύουν αν αντί για τις συνοριακές συνθήκες Dirichlet (1.25), θεωρήσουμε *ομογενείς συνοριακές συνθήκες Neumann*, δηλαδή αν αναζητήσουμε u, τέτοια ώστε

$$u_t = u_{xx}, \qquad \text{στο } [0, L] \times [0, +\infty) \tag{1.30}$$

$$u_x(0,t) = u_x(L,t) = 0, \qquad \text{για } t \geq 0, \tag{1.31}$$

$$u(x,0) = g(x), \qquad \text{στο } [0, L], \tag{1.32}$$

τότε οι συναρτήσεις

$$u_n(x,t) = e^{-\lambda_n^2 t}\cos(\lambda_n x), \quad \lambda_n = \frac{n\pi}{L},\ n = 0,1,2,\ldots,$$

ικανοποιούν τις (1.30)–(1.31). Έστω τώρα ότι υπάρχουν σταθερές $c_n \in \mathbb{R}$, τέτοιες ώστε

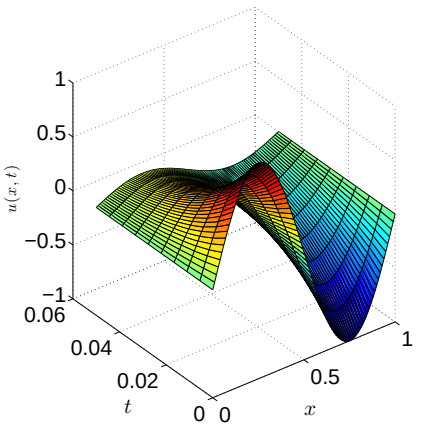
$$u(x,0) = g(x) = \sum_{n=0}^{\infty} c_n \cos(\lambda_n x), \tag{1.33}$$

όπου η σειρά συγκλίνει για κάθε $x \in [0,L]$. Αν επιπλέον για κάθε $x \in [0,L]$ και $t \geq 0$ η σειρά $\sum_{n=1}^{\infty} c_n e^{-\lambda_n^2 t}\cos(\lambda_n x)$ συγκλίνει, τότε η

$$u(x,t) = \sum_{n=1}^{\infty} c_n e^{-\lambda_n^2 t}\cos(\lambda_n x), \quad \text{με } c_n \in \mathbb{R}, \tag{1.34}$$

είναι η μοναδική λύση του προβλήματος αρχικών και συνοριακών τιμών (1.30)–(1.32). Από τη μορφή της λύσης (1.28) ή (1.34), μπορούμε να παρατηρήσουμε ότι η u φθίνει κατά απόλυτη τιμή, καθώς το t τείνει στο άπειρο.

Παράδειγμα 1.1. Αν $u(x,t) = e^{-4\pi^2 t}\sin(2\pi x)$, τότε εύκολα βλέπουμε ότι η u είναι λύση του προβλήματος (1.24)–(1.26), με $L = 1$ και $g(x) = \sin(2\pi x)$. Στο Σχήμα 1.1 απεικονίζεται η u στο $[0,1]\times[0,T]$, με $T = 0.05$.



Σχήμα 1.1: Η ακριβής λύση u του Παραδείγματος 1.1.

1.2.3 Εξίσωση μεταφοράς

Θα θεωρήσουμε τώρα μια απλή μερική διαφορική εξίσωση πρώτης τάξης, την εξίσωση μεταφοράς και το αντίστοιχο πρόβλημα αρχικών τιμών: Ζητείται μια συνάρτηση $u \in C^1(\mathbb{R} \times [0,T])$, $T > 0$, τέτοια ώστε

$$u_t(x,t) + \alpha u_x(x,t) = 0, \qquad x \in \mathbb{R},\ t \in [0,T], \tag{1.35}$$

$$u(x,0) = g(x), \qquad x \in \mathbb{R}, \tag{1.36}$$

όπου $\alpha \in \mathbb{R}$ και g μια δοσμένη συνάρτηση. Αν θεωρήσουμε την ακόλουθη αλλαγή μεταβλητών

$$x = \alpha r + s, \quad t = r,$$

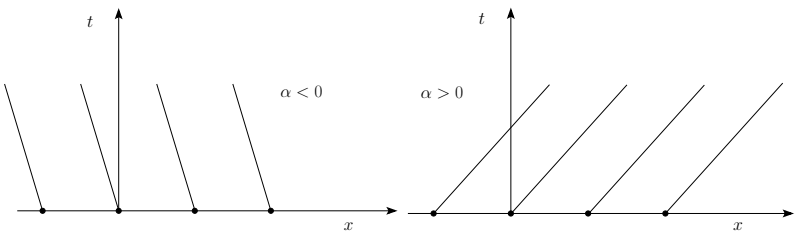
τότε εύκολα μπορούμε να δούμε ότι αν η u είναι λύση του (1.35), τότε

$$\frac{\partial u}{\partial r} = \frac{\partial x}{\partial r}\frac{\partial u}{\partial x} + \frac{\partial t}{\partial r}\frac{\partial u}{\partial t} = \alpha u_x + u_t = 0.$$

Επομένως, αν θεωρήσουμε ότι

$$z(s,r) := u(x - \alpha t, t),$$

έχουμε $\dfrac{\partial z}{\partial r} = 0$, δηλαδή η z δεν εξαρτάται από το r και είναι σταθερή ως προς r. Συνεπώς, η u πάνω σε όλες τις ευθείες $x - \alpha t = c$, με c σταθερά, δεν μεταβάλλει την τιμή της και, έτσι, $u(x,t) = z(s,r) = z(s,0) = u(x - \alpha t, 0) = g(x - \alpha t)$. Λόγω αυτού του γεγονότος, οι ευθείες $x - \alpha t = c$ καλούνται *χαρακτηριστικές* για τη διαφορική εξίσωση (1.35), βλ. Σχήμα 1.2.



Σχήμα 1.2: Οι χαρακτηριστικές ευθείες $x - \alpha t = c$ πάνω στις οποίες είναι σταθερή η λύση u της διαφορικής εξίσωσης (1.35).

Παράδειγμα 1.2. Ας θεωρήσουμε το πρόβλημα (1.35) και ότι η λύση u για $t = 0$ δίνεται από τη συνάρτηση

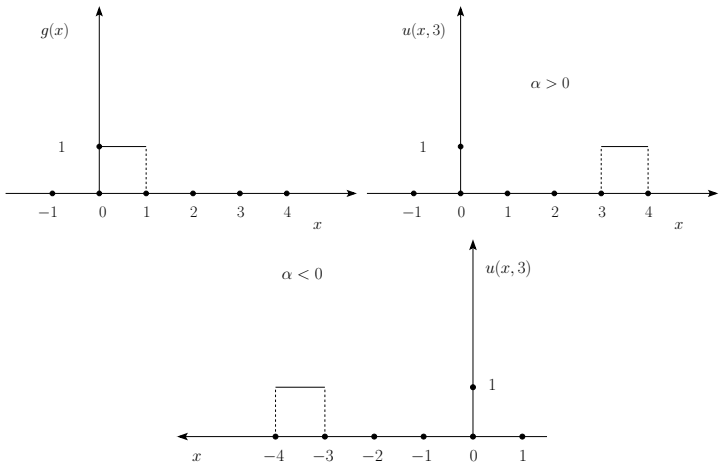
$$g(x) = \begin{cases} 1 & \text{αν } 0 \le x \le 1 \\ 0 & \text{διαφορετικά.} \end{cases}$$

Τότε η λύση είναι

$$u(x,t) = \begin{cases} 1 & \text{αν } 0 \leq x - \alpha t \leq 1 \\ 0 & \text{διαφορετικά,} \end{cases}$$

ή ισοδύναμμα

$$u(x,t) = \begin{cases} 1 & \text{αν } \alpha t \leq x \leq 1 + \alpha t \\ 0 & \text{διαφορετικά,} \end{cases}$$



Σχήμα 1.3: Η αρχική συνθήκη g και η ακριβής λύση u του Παραδείγματος 1.2 στην περίπτωση που $\alpha > 0$ ή $\alpha < 0$.

Παρατηρούμε ότι για να προσδιορίσουμε με μοναδικό τρόπο τη λύση της διαφορικής εξίσωσης (1.35), αρκεί να γνωρίζουμε την αρχική συνθήκη $u(x,0) = g(x)$. Επίσης, αν $\alpha > 0$, μπορούμε να διατυπώσουμε το πρόβλημα (1.35) ως πρόβλημα αρχικών και συνοριακών τιμών ως εξής: Ζητείται μια συνάρτηση $u \in C^1([a,b] \times [0,T])$, $a < b, T > 0$, τέτοια ώστε

$$\begin{aligned} u_t(x,t) + \alpha u_x(x,t) = 0, \quad & x \in [a,b],\ t \in [0,T], \\ u(x,0) = g(x), \quad & x \in [a,b], \\ u(a,t) = \phi_0(t), \quad & t \in [a,b], \end{aligned} \tag{1.37}$$

όπου $\phi_0(t) = g(a - \alpha t)$. Συνεπώς, σε αντίθεση με τα προβλήματα που είδαμε στις Παραγράφους 1.2.1 και 1.2.2, αρκεί μια συνοριακή συνθήκη για το σημείο $x = a$ για να προσδιορίσουμε μοναδικά τη λύση του προβλήματος (1.37). Όμως, στην περίπτωση τώρα που $\alpha < 0$, οι χαρακτηριστικές θα έχουν διαφορετική κατεύθυνση

και διατυπώνουμε το πρόβλημα (1.35) ως πρόβλημα συνοριακών τιμών ως εξής: Ζητείται μια συνάρτηση $u \in C^1([a,b] \times [0,T])$, $a < b$, $T > 0$, τέτοια ώστε

$$\begin{aligned} u_t(x,t) + \alpha u_x(x,t) = 0, \quad x \in [a,b],\ t \in [0,T], \\ u(x,0) = g(x), \quad x \in [a,b], \\ u(b,t) = \phi_1(t), \quad t \in [a,b], \end{aligned} \tag{1.38}$$

όπου $\phi_1(t) = g(a-\alpha t)$. Και σε αυτή την περίπτωση αρκεί μια συνοριακή συνθήκη για $x = b$ για να προσδιορίσουμε μοναδικά τη λύση του προβλήματος (1.38).

Είναι προφανές ότι η λύση u του προβλήματος (1.35) διατηρεί τη μορφή της αρχικής συνάρτησης, καθώς ο χρόνος t αυξάνει και ιδιαίτερα η μέγιστη κατά απόλυτο τιμή παραμένει σταθερή. Στο Σχήμα 1.3 βλέπουμε τη λύση του Παραδείγματος 1.2, για $t = 3$, αν η σταθερά α είναι θετική ή αρνητική.

1.2.4 Εξίσωση κύματος

Θα θεωρήσουμε τώρα μια απλή υπερβολική διαφορική εξίσωση και συγκεκριμένα την εξίσωση του κύματος (1.6), όπου η χωρική μεταβλητή x ανήκει στο διάστημα $[0,L]$, με $L > 0$ και θα θεωρήσουμε, επιπλέον, *αρχικές και ομογενείς συνοριακές συνθήκες Dirichlet*. Έτσι, έχουμε το ακόλουθο πρόβλημα: Ζητείται $u : [0,L] \times [0,+\infty) \to \mathbb{R}$, τέτοια ώστε

$$u_{tt} = \alpha^2 u_{xx}, \qquad \text{στο } [0,L] \times [0,+\infty) \tag{1.39}$$

$$u(0,t) = u(L,t) = 0, \qquad \text{για } t \geq 0, \tag{1.40}$$

$$u(x,0) = g(x), \ \text{ και } \ u_t(x,0) = g_1(x), \qquad \text{στο } [0,L], \tag{1.41}$$

με $\alpha > 0$ και g, g_1 συνεχείς συναρτήσεις.

Χρησιμοποιώντας τη μέθοδο των χωριζομένων μεταβλητών, βλ. π.χ. (Logan, 2014), οι συναρτήσεις u_n, με $u_n = X_n T_n$, $n = 1, 2, \ldots$, όπου

$$u_n = X_n T_n, \tag{1.42}$$

$$X_n(x) = c \sin(\lambda_n x), \ \text{ και} \tag{1.43}$$

$$T_n(t) = d_1 \sin(\lambda_n \alpha t) + d_2 \cos(\lambda_n \alpha t) \text{ με } \lambda_n = \frac{n\pi}{L}, \ c, d_1, d_2 \in \mathbb{R}, \tag{1.44}$$

ικανοποιούν την εξίσωση (1.39) και τις συνοριακές συνθήκες (1.40). Λόγω της αρχής της υπέρθεσης, κάθε γραμμικός συνδυασμός λύσεων της μορφής (1.42) ικανοποιεί τις (1.39)-(1.40). Επομένως, αν θεωρήσουμε μια άπειρη σειρά συναρτήσεων της μορφής

$$u(x,t) = \sum_{n=1}^{\infty} \sin(\lambda_n x)(c_n \sin(\lambda_n \alpha t) + d_n \cos(\lambda_n \alpha t)), \quad \text{με } c_n, d_n \in \mathbb{R}, \tag{1.45}$$

η οποία συγκλίνει για κάθε $x \in [0, L]$ και $t \geq 0$, τότε αυτή αποτελεί και λύση του προβλήματος (1.39)-(1.41). Μια λύση u της μορφής (1.45) θα ικανοποιεί και την αρχική συνθήκη (1.41), αν

$$u(x,0) = \sum_{n=1}^{\infty} d_n \sin(\lambda_n x) = g(x), \quad \text{με } d_n \in \mathbb{R}, \tag{1.46}$$

$$u_t(x,0) = \sum_{n=1}^{\infty} c_n \lambda_n \alpha \sin(\lambda_n x) = g_1(x), \quad \text{με } c_n \in \mathbb{R}. \tag{1.47}$$

Ένας διαφορετικός τρόπος για την περιγραφή της λύσης του προβλήματος (1.39)–(1.41) είναι η *αναπαράσταση d'Alembert* της λύσης, η οποία οδηγεί στην

$$u(x,t) = \Phi(x - \alpha t) + \Psi(x + \alpha t), \tag{1.48}$$

όπου Φ, Ψ είναι συνεχείς συναρτήσεις που εξαρτώνται από τις αρχικές συνθήκες (1.41). Μάλιστα, εύκολα μπορούμε να δούμε ότι στην περίπτωση που στην αρχική συνθήκη (1.41) η συνάρτηση g_1 μηδενίζεται στο $[0, L]$, τότε η ακριβής λύση u του προβλήματος (1.39)–(1.41) δίνεται από

$$u(x,t) = \frac{1}{2}(g(x - \alpha t) + g(x - \alpha t)). \tag{1.49}$$

Παράδειγμα 1.3. Αν $u(x,t) = \sin(2\pi x)(\sin(2\pi t) + \cos(2\pi t))$, τότε εύκολα βλέπουμε ότι η u είναι λύση του προβλήματος (1.24)–(1.26), με $L = \alpha = 1$, $g(x) = \sin(2\pi x)$ και $g_1(x) = 2\pi \sin(2\pi x)$. Στο Σχήμα 1.4, απεικονίζεται η u στο $[0,1] \times [0,T]$, με $T = 1$.

1.2.5 Ελλειπτική εξίσωση

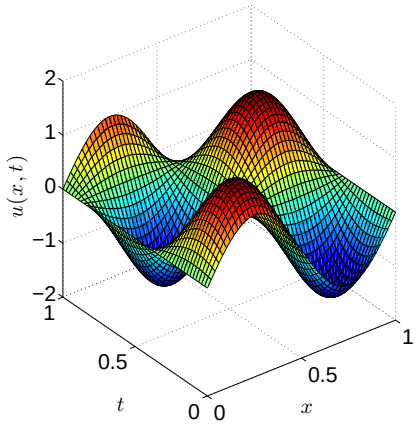
Θεωρούμε το ακόλουθο πρόβλημα: Ζητείται μια συνάρτηση $u : \Omega = (0,1) \times (0,1) \to \mathbb{R}$, τέτοια ώστε

$$\begin{aligned} -(u_{xx}(x,y) + u_{yy}(x,y)) + q(x,y)u(x,y) &= f(x,y), \quad (x,y) \in \Omega, \\ u(x,y) &= g(x,y), \quad (x,y) \in \partial\Omega, \end{aligned} \tag{1.50}$$

όπου $\partial\Omega$ είναι το σύνορο του Ω, $q, f \in C(\Omega)$, $g \in C(\partial\Omega)$ και $q(x,y) \geq 0$, για κάθε $(x,y) \in \Omega$. Στη συνέχεια, θα συμβολίζουμε με $q_{\min} = \min_{(x,y)\in\Omega} q(x,y)$ και $\bar{\Omega} = [0,1] \times [0,1]$.

Η υπόθεση $q \geq 0$ είναι σημαντική για την ύπαρξη και μοναδικότητα της λύσης του προβλήματος (1.50). Παραδείγματος χάριν, αν θεωρήσουμε την

$$\begin{aligned} -(u_{xx}(x,y) + u_{yy}(x,y)) - 2\pi^2 u(x,y) &= 0, \quad (x,y) \in \Omega, \\ u(x,y) &= g(x,y), \quad (x,y) \in \partial\Omega, \end{aligned} \tag{1.51}$$



Σχήμα 1.4: Η ακριβής λύση u του Παραδείγματος 1.3.

τότε οι συναρτήσεις $u(x,y) = c\sin(\pi x)\sin(\pi y)$, $c \in \mathbb{R}$, αποτελούν λύσεις και, επομένως, το πρόβλημα (1.51) έχει άπειρες λύσεις.

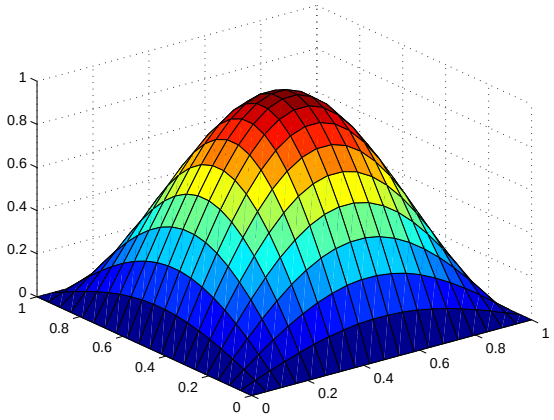
Για τη διαφορική εξίσωση (1.50) ισχύει η αρχή του μεγίστου, η οποία δίνεται από το ακόλουθο θεώρημα, βλ. π.χ. (Larsson & Thomée, 2009).

Θεώρημα 1.4. *Έστω ότι η λύση u του* (1.50) *είναι τέτοια, ώστε $u \in C^2(\bar{\Omega})$. Τότε*

1. *Αν*
$$f(x,y) \le 0, \quad \textit{για κάθε} (x,y) \in \Omega,$$
έχουμε ότι
$$\max_{(x,y)\in\bar{\Omega}} u(x,y) \le \begin{cases} \max\limits_{(x,y)\in\partial\Omega} g(x,y), & \textit{αν } q \equiv 0, \\ \max\{\max\limits_{(x,y)\in\partial\Omega} g(x,y), 0\}, & \textit{αν } q \ge 0. \end{cases} \tag{1.52}$$

2. *Αν*
$$f(x,y) \ge 0, \quad \textit{για κάθε} (x,y) \in \Omega,$$
έχουμε ότι
$$\min_{(x,y)\in\bar{\Omega}} u(x,y) \ge \begin{cases} \min\limits_{(x,y)\in\partial\Omega} g(x,y), & \textit{αν } q \equiv 0, \\ \min\{\min\limits_{(x,y)\in\partial\Omega} g(x,y), 0\}, & \textit{αν } q \ge 0. \end{cases} \tag{1.53}$$

Παράδειγμα 1.4. Αν $u(x,t) = \sin(\pi x)\sin(\pi y)$, τότε εύκολα βλέπουμε ότι η u είναι λύση του προβλήματος (1.50), με $g = 0$, $q = 1$ και $f(x,y) = (2\pi^2 + 1)\sin(\pi x)\sin(\pi y)$. Στο Σχήμα 1.5, απεικονίζεται η u στο $[0,1] \times [0,1]$.



Σχήμα 1.5: Η ακριβής λύση u του Παραδείγματος 1.4.

1.3 Σειρές Fourier

Όπως είδαμε στις προηγούμενες παραγράφους, μπορούμε να αναπαραστήσουμε τη λύση ορισμένων διαφορικών εξισώσεων, όπως π.χ. την εξίσωση της θερμότητας, χρησιμοποιώντας συγκλίνουσες σειρές τριγωνομετρικών συναρτήσεων. Μια κατηγορία τριγωνομετρικών σειρών συναρτήσεων είναι οι σειρές Fourier, τις οποίες θα παρουσιάσουμε εν συντομία σε αυτή την παράγραφο.

Γνωρίζουμε ότι μπορούμε να χρησιμοποιήσουμε πολυώνυμα για να προσεγγίσουμε μία αρκετά ομαλή συνάρτηση f κοντά σε ένα δοσμένο σημείο του πεδιού ορισμού της, π.χ. αν f είναι μια $n+1$ φορές συνεχώς παραγωγίσιμη συνάρτηση κοντά στο x_0, τότε

$$f(x) \approx \sum_{k=0}^{n} (x - x_0)^k \frac{f^{(k)}(x_0)}{k!}, \qquad \text{για } x \approx x_0.$$

Ένας άλλος τρόπος για την προσέγγιση μιας συνάρτησης f είναι η χρήση τριγωνομετρικών συναρτήσεων με το ανάπτυγμα σε σειρές Fourier της f. Σε αυτή την περίπτωση αντί να χρησιμοποιήσουμε πολυώνυμα, δηλαδή δυνάμεις του x ως

βασικές συναρτήσεις, χρησιμοποιούμε τριγωνομετρικές συναρτήσεις π.χ., $\sin(kx)$ και $\cos(kx)$, για $k = 0, 1, 2, \ldots$

Ορισμός 1.1. Το ανάπτυγμα σε *σειρά Fourier* μιας συνάρτησης f με περίοδο $2L$, $L > 0$, δίνεται ως

$$\frac{a_0}{2} + \sum_{k=1}^{\infty} \left[a_k \cos(\frac{k\pi x}{L}) + b_k \sin(\frac{k\pi x}{L}) \right], \tag{1.54}$$

όπου

$$\begin{aligned} a_0 &= \frac{1}{L} \int_{-L}^{L} f(x)\, dx, \\ a_k &= \frac{1}{L} \int_{-L}^{L} f(x) \cos(\frac{k\pi x}{L})\, dx, \quad k = 1, 2, \ldots, \\ b_k &= \frac{1}{L} \int_{-L}^{L} f(x) \sin(\frac{k\pi x}{L})\, dx, \quad k = 1, 2, \ldots. \end{aligned}$$

Παράδειγμα 1.5. Αν θέλουμε να βρούμε το ανάπτυγμα σε σειρά Fourier της περιοδικής συνάρτησης f, με περίοδο 2, η οποία στο $[-1, 1]$ ορίζεται ως $f(x) = |x|$, τότε για $k = 1, 2, \ldots$, έχουμε

$$\begin{aligned} a_k &= \int_{-1}^{1} |x| \cos(k\pi x)\, dx = \int_{-1}^{0} (-x) \cos(k\pi x)\, dx + \int_{0}^{1} x \cos(k\pi x)\, dx \\ &= \left[-x \frac{\sin(k\pi x)}{k\pi} \right]_{-1}^{0} + \int_{-1}^{0} \frac{\sin(k\pi x)}{k\pi}\, dx \\ &\quad + \left[x \frac{\sin(k\pi x)}{k\pi} \right]_{0}^{1} - \int_{0}^{1} \frac{\sin(k\pi x)}{k\pi}\, dx \\ &= \left[-\frac{\cos(k\pi x)}{(k\pi)^2} \right]_{-1}^{0} + \left[-\frac{\cos(k\pi x)}{(k\pi)^2} \right]_{0}^{1} \\ &= 2\frac{(-1)^k - 1}{(k\pi)^2} \end{aligned}$$

και

$$\begin{aligned} b_k &= \int_{-1}^{1} |x| \sin(k\pi x)\, dx = \int_{-1}^{0} (-x) \sin(k\pi x)\, dx + \int_{0}^{1} x \sin(k\pi x)\, dx \\ &= \left[x \frac{\cos(k\pi x)}{k\pi} \right]_{-1}^{0} - \int_{-1}^{0} \frac{\cos(k\pi x)}{k\pi}\, dx \end{aligned}$$

$$+\left[-x\frac{\cos(k\pi x)}{k\pi}\right]_0^1+\int_0^1\frac{\cos(k\pi x)}{k\pi}\,dx$$
$$=\frac{(-1)^k}{k\pi}-\left[\frac{\sin(k\pi x)}{(k\pi)^2}\right]_{-1}^0+\frac{(-1)^k}{k\pi}+\left[\frac{\sin(k\pi x)}{(k\pi)^2}\right]_0^1=0.$$

Ακόμα

$$a_0=\int_{-1}^1|x|\,dx=1.$$

Επομένως, το ανάπτυγμα σε σειρά Fourier της f είναι

$$\frac{1}{2}+\sum_{k=1}^{\infty}2\frac{(-1)^k-1}{(k\pi)^2}\cos(k\pi x).$$

Αν θεωρήσουμε τα πεπερασμένα αθροίσματα, $S_n(x)$, αυτής της σειράς,

$$S_n(x)=\frac{1}{2}+\sum_{k=1}^{n}2\frac{(-1)^k-1}{(k\pi)^2}\cos(k\pi x),$$

μπορούμε να δούμε ότι, καθώς αυξάνει το n, το γράφημα τους πλησιάζει αυτό της f, όπως φαίνεται και από το Σχήμα 1.6.

Ένα προφανές ερώτημα που προκύπτει είναι αν η $f(x)$ είναι ίση με την αντίστοιχη τιμή του ανάπτυγματος Fourier της $f(x)$. Μια απάντηση σε αυτό δίνεται στο παρακάτω θεώρημα.

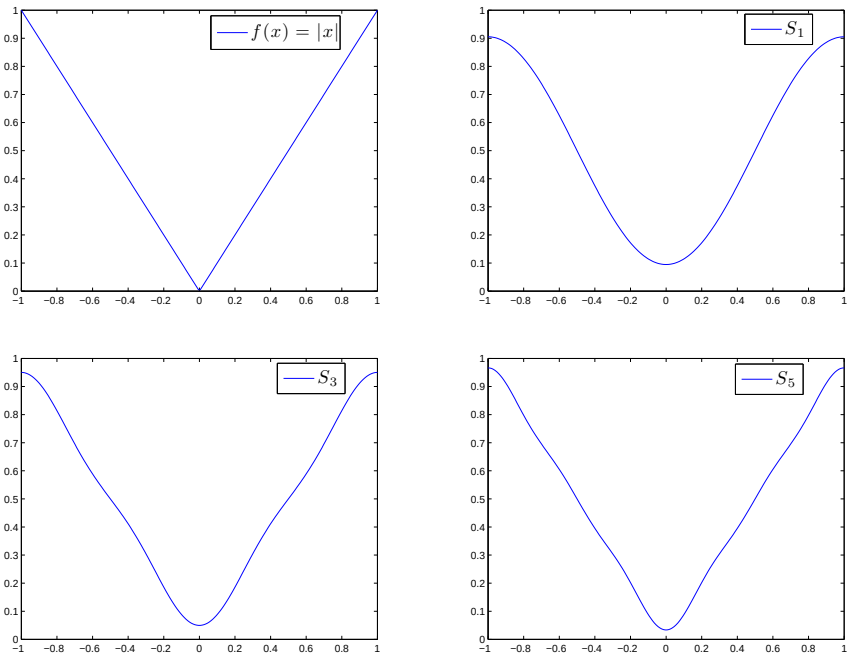
Θεώρημα 1.5. *Έστω ότι μια συνάρτηση $f:[-L,L]\to\mathbb{R}$ και η παράγωγος της, f', είναι φραγμένες και συνεχείς παντού στο $[-L,L]$, εκτός, ίσως, από ένα πεπερασμένο πλήθος σημείων. Τότε*

$$f(x)=\frac{a_0}{2}+\sum_{n=1}^{\infty}a_n\cos(\frac{k\pi x}{L})+b_n\sin(\frac{k\pi x}{L}),$$

σε όλα τα σημεία συνέχειας της f.

Μια βασική ιδιότητα που έχουν οι τριγωνομετρικές συναρτήσεις του αναπτύγματος (1.54) είναι ότι είναι ορθογώνιες μεταξύ τους, αν τις ολοκληρώσουμε στο διάστημα $[-L,L]$, δηλαδή

$$\int_{-L}^{L}\cos(\frac{k\pi x}{L})\cos(\frac{n\pi x}{L})\,dx=\begin{cases}L, & \text{αν } k=n,\\ 0, & \text{διαφορετικά,}\end{cases}\tag{1.55}$$



Σχήμα 1.6: Η συνάρτηση $f(x) = |x|$ και τα πεπερασμένα αθροίσματα $S_n(x)$ της σειράς Fourier με $n = 1, 3,$ και 5.

$$\int_{-L}^{L} \sin(\frac{k\pi x}{L}) \sin(\frac{n\pi x}{L})\, dx = \begin{cases} L, & \text{αν } k = n, \\ 0, & \text{διαφορετικά,} \end{cases} \tag{1.56}$$

$$\int_{-L}^{L} \cos(\frac{k\pi x}{L}) \sin(\frac{n\pi x}{L})\, dx = 0 \quad \text{για κάθε } k, n. \tag{1.57}$$

Επίσης, μπορούμε να δούμε και ότι αν η συνάρτηση f είναι άρτια στο $[-L, L]$, δηλ. $f(-x) = f(x)$, τότε επειδή η $f(x)\sin(\frac{k\pi x}{L})$ είναι περιττή στο $[-L, L]$, οι συντελεστές b_k στην (1.54) θα μηδενίζονται. Επομένως, το ανάπτυγμα σε σειρά Fourier μιας άρτιας συνάρτησης έχει μόνο όρους συνημιτόνων, δηλαδή είναι της μορφής

$$f(x) = \frac{a_0}{2} + \sum_{k=1}^{\infty} a_k \cos(k\pi x),$$

και μάλιστα

$$a_k = \frac{2}{L} \int_0^L f(x) \cos(\frac{k\pi x}{L})\, dx, \quad k = 0, 1, \ldots \tag{1.58}$$

΄Ομοια για μια περιοδική περιττή συνάρτηση f μπορούμε να δούμε ότι $a_k = 0, k = 0, 1, \ldots$ Επομένως, το ανάπτυγμά της σε σειρά Fourier έχει μόνο όρους ημιτόνων, δηλαδή είναι της μορφής

$$f(x) = \sum_{k=1}^{\infty} b_k \sin(\frac{k\pi x}{L}), \tag{1.59}$$

με

$$b_k = \frac{2}{L} \int_0^L f(x) \sin(\frac{k\pi x}{L})\, dx, \quad k = 1, 2, \ldots \tag{1.60}$$

΄Εστω τώρα ότι έχουμε μια συνεχή συνάρτηση f στο $[0, L]$. Η f μπορεί να επεκταθεί στο $[-L, L]$ ως άρτια ή ως περιττή συνάρτηση. Επομένως, το ανάπτυγμα της f σε σειρά Fourier μπορεί να γραφεί είτε ως σειρά συνημιτόνων είτε ως σειρά ημιτόνων, αντίστοιχα.

Παράδειγμα 1.6. ΄Εστω $f(x) = \sin(x)$, $x \in [0, \pi]$. Μπορούμε να επεκτείνουμε την f στο $[-\pi, 0]$ ως μια άρτια συνάρτηση. Έτσι

$$f(x) = \begin{cases} -\sin(x), & x \in [-\pi, 0], \\ \sin(x), & x \in [0, \pi]. \end{cases}$$

Συνεπώς, η $f(x)$ μπορεί να γραφεί ως σειρά συνημιτόνων και, υπολογίζοντας τους αντίστοιχους συντελεστές a_k της (1.58), έχουμε

$$f(x) = \frac{2}{\pi} - \frac{4}{\pi} \sum_{k=1}^{\infty} \frac{1}{2k^2 - 1} \cos(2kx).$$

Βιβλιογραφία

Ακρίβης, Γ., & Αλικάκος, Ν. (2012). *Μερικές Διαφορικές Εξισώσεις*. Σύγχρονη Εκδοτική, Αθήνα.

Ακρίβης, Γ., & Δουγαλής, Β. (2005). *Αριθμητικές Μέθοδοι για Μερικές Διαφορικές Εξισώσεις*. Ιωάννινα. (Πανεπιστημιακές Σημειώσεις).

Brezis, H. (1997). *Συναρτησιακή Ανάλυση: Θεωρία και Εφαρμογές*. Πανεπιστημιακές Εκδόσεις Ε.Μ.Π., Αθήνα. (Ελληνική μετάφραση από τους Δ. Κραββαρίτη και Ι. Χρυσοβέργη).

Larsson, S., & Thomée, V. (2009). *Partial differential equations with numerical methods* (Vol. 45). Springer-Verlag, Berlin. (Paperback reprint of the 2003 edition).

Logan, D. J. (2014). *Εισαγωγή στις Διαφορικές Εξισώσεις*. Liberal Books, Αθήνα. (Ελληνική μετάφραση από τον Ι. Πλατή).

Κεφάλαιο 2

Πεπερασμένες διαφορές

Αυτό το κεφάλαιο αποτελεί μια εισαγωγή στο αντικείμενο των πεπερασμένων διαφορών για την επίλυση διαφορικών εξισώσεων. Θα εισαγάγουμε ποσότητες που προκύπτουν από διαφορές τιμών μιας πραγματικής διαφορίσιμης συνάρτησης f μιας μεταβλητής για την προσέγγιση παραγώγων της f. Επίσης, θα εκτιμήσουμε το σφάλμα που προκύπτει από αυτή την προσέγγιση.

2.1 Προσέγγιση πρώτης παραγώγου

Από τον ορισμό της παραγώγου για συναρτήσεις μιας μεταβλητής γνωρίζουμε ότι η παράγωγος μιας συνάρτησης $f : [a, b] \to \mathbb{R}$ σε ένα σημείο $x_0 \in (a, b)$ ορίζεται ως

$$f'(x_0) = \lim_{h \to 0} \frac{f(x_0 + h) - f(x_0)}{h}.$$

Χρησιμοποιώντας αυτόν τον ορισμό μπορούμε να προσεγγίσουμε την τιμή της $f'(x_0)$ με τον ακόλουθο λόγο, για μικρές τιμές h,

$$f'(x_0) \approx \frac{f(x_0 + h) - f(x_0)}{h}, \quad h > 0.$$

Με τον ίδιο τρόπο, μπορούμε να προσεγγίσουμε την $f'(x_0)$ με τον ακόλουθο λόγο,

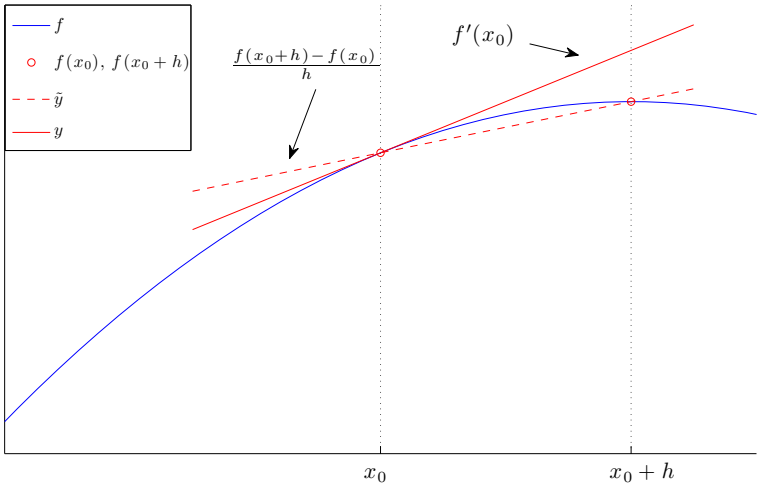
$$f'(x_0) \approx \frac{f(x_0 - h) - f(x_0)}{-h} = \frac{f(x_0) - f(x_0 - h)}{h}, \quad h > 0,$$

για μικρές τιμές h. Θα καλούμε τον πρώτο λόγο *διαφορά προς τα εμπρός* και τον δεύτερο λόγο *διαφορά προς τα πίσω* και θα χρησιμοποιήσουμε τον ακόλουθο συμβολισμό

$$\begin{aligned} \delta_h^+ f(x_0) &\equiv \frac{f(x_0+h)-f(x_0)}{h}, \quad h>0, \\ \delta_h^- f(x_0) &\equiv \frac{f(x_0)-f(x_0-h)}{h}, \quad h>0. \end{aligned} \tag{2.1}$$

Γεωμετρική ερμηνεία

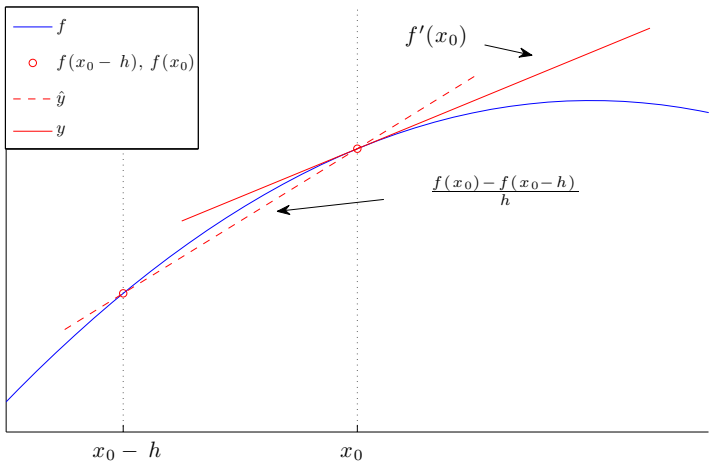
Μια γεωμετρική ερμηνεία γιατί οι διαφορές $\delta_h^+ f(x_0)$ και $\delta_h^- f(x_0)$ προσεγγίζουν την $f'(x_0)$ είναι η ακόλουθη. Επειδή η παράγωγος μιας συνάρτησης f στο σημείο x_0 είναι η κλίση της ευθείας y που εφάπτεται του γραφήματος της f στο σημείο $(x_0, f(x_0))$, μπορούμε να την προσεγγίσουμε με την κλίση της ευθείας $\tilde{y}$ η οποία διέρχεται από τα σημεία $(x_0, f(x_0))$ και $(x_0+h, f(x_0+h))$, βλέπε το Σχήμα 2.1. Παρόμοια ισχύουν και για την κλίση της ευθείας $\hat{y}$ που διέρχεται από τα σημεία $(x_0-h, f(x_0-h))$, και $(x_0, f(x_0))$, βλέπε το Σχήμα 2.2.



Σχήμα 2.1: Γεωμετρική ερμηνεία της $\delta_h^+ f(x_0)$. Η ευθεία y που έχει κλίση $f'(x_0)$ και διέρχεται από το $(x_0, f(x_0))$ και η ευθεία $\tilde{y}$ που έχει κλίση $\dfrac{f(x_0+h)-f(x_0)}{h}$.

Ένας άλλος τρόπος προσέγγισης της $f'(x_0)$, είναι η *κεντρική διαφορά* , η οποία ορίζεται από τον ακόλουθο λόγο, για μικρές τιμές h,

$$f'(x_0) \approx \frac{f(x_0+h)-f(x_0-h)}{2h}, \quad h>0.$$



Σχήμα 2.2: Γεωμετρική ερμηνεία της $\delta_h^- f(x_0)$. Η ευθεία y που έχει κλίση $f'(x_0)$ και διέρχεται από το $(x_0, f(x_0))$ και η ευθεία $\hat{y}$ που έχει κλίση $\dfrac{f(x_0)-f(x_0-h)}{h}$.

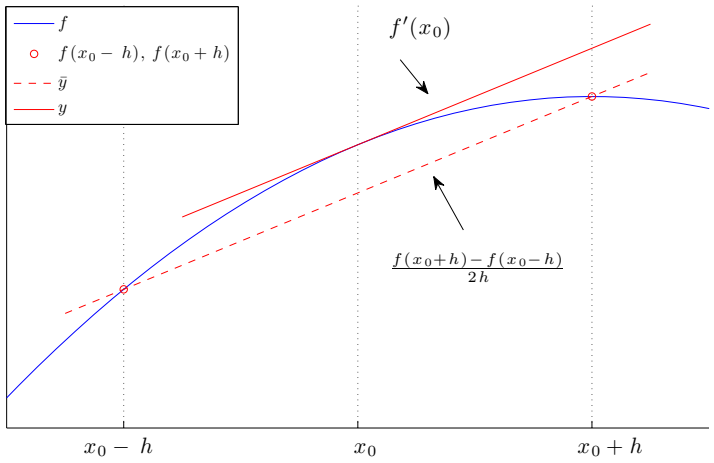
Θα τη συμβολίζουμε με

$$\delta_h^c f(x_0) \equiv \frac{f(x_0+h)-f(x_0-h)}{2h}, \quad h>0. \tag{2.2}$$

Ανάλογη γεωμετρική ερμηνεία με αυτή για τις $\delta_h^+ f(x_0)$ και $\delta_h^- f(x_0)$ υπάρχει και για την $\delta_h^c f(x_0)$. Στο Σχήμα 2.3 απεικονίζουμε την ευθεία $\bar{y}$ που διέρχεται από τα σημεία $(x_0-h, f(x_0-h))$ και $(x_0+h, f(x_0+h))$, και της οποίας η κλίση προσεγγίζει την κλίση της εφαπτομένης του γραφήματος της f στο $(x_0, f(x_0))$.

Οι διαφορές $\delta_h^+ f(x)$, $\delta_h^- f(x)$ και $\delta_h^c f(x)$ για την προσέγγιση παραγώγων μιας συνάρτησης f καλούνται και *πεπερασμένες διαφορές*. Μια φυσική ερώτηση που τίθεται είναι "πόσο καλές" είναι αυτές οι προσεγγίσεις για την εκτίμηση της παραγώγου. Ας θεωρήσουμε τη συνάρτηση $f(x)=\ln(x)$ και το σημείο $x_0=1.1$. Στον Πίνακα 2.1 δίνουμε τις τιμές των παραπάνω προσεγγίσεων για την $f'(1.1)=1/1.1\approx 0.90909$. Παρατηρούμε λοιπόν ότι, καθώς η απόσταση h των σημείων που χρησιμοποιούμε για να υπολογίσουμε τις $\delta_h^+ f(1.1)$, $\delta_h^- f(1.1)$ και $\delta_h^c f(1.1)$ μικραίνει, τόσο πλησιάζουμε την τιμή της παραγώγου. Μάλιστα, στο ακόλουθο λήμμα δείχνουμε ότι το σφάλμα της προσέγγισης θα τείνει στο μηδέν, καθώς η απόσταση h τείνει στο μηδέν, αν η συνάρτηση f είναι κατάλληλα ομαλή.

Λήμμα 2.1. *Έστω $f:[a,b]\to\mathbb{R}$, $f\in C^2[a,b]$, $x_0\in(a,b)$ και $h>0$, τέτοιο ώστε*



Σχήμα 2.3: Γεωμετρική ερμηνεία της $\delta_h^c f(x_0)$. Η ευθεία y που έχει κλίση $f'(x_0)$ και διέρχεται από το $(x_0, f(x_0))$ και η ευθεία $\bar{y}$ που έχει κλίση $\dfrac{f(x_0+h)-f(x_0-h)}{2h}$.

h	$\delta_h^+ f(1.1)$	$\delta_h^- f(1.1)$	$\delta_h^c f(1.1)$
0.50	0.74939	1.21227	0.98083
0.10	0.87011	0.95310	0.91161
0.05	0.88904	0.93040	0.90972
0.01	0.90498	0.91325	0.90912

Πίνακας 2.1: Τιμές των προσεγγίσεων της $f'(1.1) = 1/1.1 \approx 0.90909$.

$x_0 \pm h \in [a,b]$. *Τότε ισχύουν οι ακόλουθες εκτιμήσεις:*

$$\begin{aligned} |\delta_h^+ f(x_0) - f'(x_0)| &\le \frac{h}{2} \max_{x\in[a,b]} |f''(x)|, \\ |\delta_h^- f(x_0) - f'(x_0)| &\le \frac{h}{2} \max_{x\in[a,b]} |f''(x)|. \end{aligned} \tag{2.3}$$

Αν επιπλέον $f \in C^3[a,b]$, *τότε*

$$|\delta_h^c f(x_0) - f'(x_0)| \le \frac{h^2}{6} \max_{x\in[a,b]} |f'''(x)|. \tag{2.4}$$

Απόδειξη. Αναπτύσσοντας κατά Taylor ως προς το σημείο x_0, έχουμε

$$f(x_0+h) = f(x_0) + hf'(x_0) + \frac{h^2}{2}f''(\xi_1), \quad \text{με } \xi_1 \in (x_0, x_0+h). \tag{2.5}$$

Επίσης,

$$f(x_0-h) = f(x_0) - hf'(x_0) + \frac{h^2}{2}f''(\xi_2), \quad \text{με } \xi_2 \in (x_0-h, x_0). \tag{2.6}$$

Από τις σχέσεις (2.5) και (2.6) εύκολα προκύπτουν οι ζητούμενες εκτιμήσεις (2.3). Αν τώρα η $f \in C^3[a,b]$, μπορούμε να αναπτύξουμε και πάλι κατά Taylor και να οδηγηθούμε στις παρακάτω σχέσεις

$$\begin{aligned} f(x_0+h) &= f(x_0) + hf'(x_0) + \frac{h^2}{2}f''(x_0) + \frac{h^3}{6}f'''(\zeta_1), \\ f(x_0-h) &= f(x_0) - hf'(x_0) + \frac{h^2}{2}f''(x_0) - \frac{h^3}{6}f'''(\zeta_2), \end{aligned} \tag{2.7}$$

με $\zeta_1 \in (x_0, x_0+h)$ και $\zeta_2 \in (x_0-h, x_0)$. Αφαιρώντας τώρα κατά μέλη τις δύο σχέσεις της (2.7), έχουμε

$$f(x_0+h) - f(x_0-h) = 2hf'(x_0) + \frac{h^3}{6}(f'''(\zeta_1) + f'''(\zeta_2)),$$

από όπου εύκολα προκύπτει η ζητούμενη εκτίμηση (2.4). □

Παρατήρηση 2.1. Από το Λήμμα 2.1 φαίνεται ότι το σφάλμα της προσέγγισης $\delta_h^c f(x_0)$ είναι μικρότερο, για αρκετά μικρό h, από τα αντίστοιχα των προσεγγίσεων $\delta_h^+ f(x_0)$ και $\delta_h^- f(x_0)$, και εξηγεί γιατί στον Πίνακα 2.1 η $\delta_h^c f(1.1)$ προσεγγίζει καλύτερα την $f'(1.1)$ από τις $\delta_h^+ f(1.1)$ και $\delta_h^- f(1.1)$. Η συμμετρία που υπάρχει στον ορισμό της προσέγγισης $\delta_h^c f(x_0)$ είναι ο λόγος που το φράγμα (2.4) είναι μικρότερο των αντίστοιχων για τις $\delta_h^+ f(1.1)$ και $\delta_h^- f(1.1)$. Αυτό φαίνεται στην (2.7), όπου οι όροι $\frac{h^2}{2}f''(x_0)$ αλληλοαναιρούνται αφαιρώντας τις δύο σχέσεις.

Ορισμός 2.1. Έστω μια προσέγγιση της παραγώγου μιας συνάρτησης f, η οποία ορίζεται μέσω τιμών της f σε ισαπέχοντα σημεία που απέχουν κατά βήμα h. Λέμε ότι αυτή η προσέγγιση έχει *τάξη ακρίβειας* κ, αν το σφάλμα που προκύπτει φράσσεται κατ' απόλυτη τιμή από το γινόμενο μιας σταθεράς, η οποία δεν εξαρτάται από το h, επί h^κ.

Παρατήρηση 2.2. Οι $\delta_h^+ f(x_0)$ και $\delta_h^- f(x_0)$ είναι προσεγγίσεις της $f'(x_0)$ τάξεως ακρίβειας ένα, επειδή τα φράγματα των σφαλμάτων (2.3) είναι το γινόμενο μιας σταθεράς, που δεν εξαρτάται από το h, επί h στην πρώτη δύναμη. Ανάλογα, η $\delta_h^c f(x_0)$ είναι προσέγγιση της $f'(x_0)$ με τάξη ακρίβειας δύο, επειδή το φράγμα του σφάλματος (2.4) είναι το γινόμενο μιας σταθεράς, που δεν εξαρτάται από το h, επί h στη δεύτερη δύναμη.

h	$E_{\delta_h^+}$	p	$E_{\delta_h^-}$	p	$E_{\delta_h^c}$	p
0.50	0.15970		0.30318		0.07174	
0.10	0.03897	0.876	0.04401	1.199	0.00251	2.081
0.05	0.02005	0.959	0.02131	1.046	0.0.0006	2.005
0.01	0.00411	0.985	0.00416	1.015	0.00003	2.000

Πίνακας 2.2: Τα σφάλματα $E_{\delta_h^+}$, $E_{\delta_h^-}$ και $E_{\delta_h^c}$ των προσεγγίσεων $\delta_h^+ f$, $\delta_h^- f$ και $\delta_h^c f$, αντίστοιχα, της $f'(1.1) = 1/1.1$ και η προσεγγιστική τάξη ακρίβειας p.

Παρατήρηση 2.3. Μπορούμε να δούμε ότι οι προσεγγίσεις $\delta_h^+ f$ και $\delta_h^- f$ έχουν τάξη ακρίβειας ακριβώς ένα, διότι αν, παραδείγματος χάριν, θεωρήσουμε την $f(x) = x^2$, τότε $\delta_h^+ f(x) = 2x + h$ και $\delta_h^- f(x) = 2x - h$. Αντίστοιχα, η προσέγγιση $\delta_h^c f$ έχει τάξη ακριβώς δύο, διότι για την $f(x) = x^3$, $\delta_h^c f(x) = 3x^2 + h^2$.

Ένας τρόπος για να επαληθεύσουμε "πειραματικά", δηλαδή με τη χρήση H/Y, την τάξη ακρίβειας των παραπάνω προσεγγίσεων είναι ο ακόλουθος. Έστω ότι το σφάλμα ε μιας προσέγγισης ικανοποιεί $\varepsilon \approx Ch^p$, για μικρό βήμα h, όπου C είναι μια θετική σταθερά ανεξάρτητη του h. Τότε, αν θεωρήσουμε δύο προσεγγίσεις ε_1 και ε_2 με βήματα h_1 και h_2, αντίστοιχα, έχουμε ότι ο λόγος των αντίστοιχων σφαλμάτων θα ικανοποιεί

$$\frac{\varepsilon_1}{\varepsilon_1} \approx (\frac{h_1}{h_2})^p, \quad \text{οπότε} \quad p \approx \frac{\log(\frac{\varepsilon_1}{\varepsilon_1})}{\log(\frac{h_1}{h_2})}.$$

Στον Πίνακα 2.2 βλέπουμε τα σφάλματα των προσεγγίσεων $\delta_h^+ f$, $\delta_h^- f$ και $\delta_h^c f$ για την $f'(1.1) = 1/1.1$ όπου $E_{\delta_h^+} = |\delta_h^+ f(1.1) - f'(1.1)|$, $E_{\delta_h^-} = |\delta_h^- f(1.1) - f'(1.1)|$ και $E_{\delta_h^c} = |\delta_h^c f(1.1) - f'(1.1)|$. Παρατηρούμε ακόμα ότι η πειραματική τάξη ακρίβειας για τις $\delta_h^+ f$ και $\delta_h^- f$ τείνει να γίνει ένα, καθώς το h ελαττώνεται και αυτή της $\delta_h^c f$ τείνει στο δύο, τα οποία είναι σύμφωνα με τα αποτελέσματα του Λήμματος 2.1.

2.2 Προσέγγιση δεύτερης παραγώγου

Στη συνέχεια, θα θεωρήσουμε και μια πεπερασμένη διαφορά για την προσέγγιση της δεύτερης παραγώγου μιας διαφορίσιμης συνάρτησης f.

Από τον ορισμό της δεύτερης παράγωγου της f σε ένα σημείο x_0 έχουμε

$$f''(x_0) = \lim_{h \to 0} \frac{f'(x_0 + h) - f'(x_0)}{h}.$$

Επομένως, μπορούμε να προσεγγίσουμε την $f''(x_0)$ χρησιμοποιώντας μία από τις προσεγγίσεις $\delta_h^+ f'(x_0)$, $\delta_h^- f'(x_0)$ ή $\delta_h^c f'(x_0)$. Αν, όμως, θέλουμε να χρησιμοποιήσουμε μόνο τιμές της f, θα πρέπει να αντικαταστήσουμε την $f'(x_0)$ με κάποια προσέγγισή της. Έτσι, ένας τρόπος είναι

$$\begin{aligned} f''(x_0) \approx & \delta_h^+ f'(x_0) = \frac{f'(x_0+h) - f'(x_0)}{h} \approx \frac{\delta_h^- f(x_0+h) - \delta_h^- f(x_0)}{h} \\ = & \delta_h^+ \delta_h^- f(x_0). \end{aligned}$$

Από τον ορισμό των δ_h^+ και δ_h^- προκύπτει ότι

$$\begin{aligned} \delta_h^+ \delta_h^- f(x_0) = & \frac{1}{h}\left(\frac{f(x_0+h) - f(x_0)}{h} - \frac{f(x_0) - f(x_0-h)}{h}\right) \\ = & \frac{f(x_0+h) - 2f(x_0) + f(x_0-h)}{h^2}. \end{aligned}$$

Επομένως, χρησιμοποιώντας την παραπάνω πεπερασμένη διαφορά μπορούμε να προσεγγίσουμε την $f''(x_0)$. Παρόμοια, μπορούμε να προσεγγίσουμε την $f''(x_0)$ ως

$$\begin{aligned} f''(x_0) \approx & \delta_h^- f'(x_0) = \frac{f'(x_0) - f'(x_0-h)}{h} \approx \frac{\delta_h^+ f(x_0+h) - \delta_h^+ f(x_0)}{h} \\ = & \delta_h^- \delta_h^+ f(x_0). \end{aligned}$$

Από εδώ προκύπτει

$$\begin{aligned} \delta_h^- \delta_h^+ f(x_0) = & \frac{1}{h}\left(\frac{f(x_0+h) - f(x_0)}{h} - \frac{f(x_0) - f(x_0-h)}{h}\right) \\ = & \frac{f(x_0+h) - 2f(x_0) + f(x_0-h)}{h^2}, \end{aligned}$$

δηλαδή ότι $\delta_h^- \delta_h^+ f = \delta_h^+ \delta_h^- f$. Άρα και οι δύο παραπάνω μεθοδολογίες οδηγούν στην ίδια πεπερασμένη διαφορά για την προσέγγιση της f''.

Επίσης, ισχύει ότι η ίδια πεπερασμένη διαφορά προκύπτει και χρησιμοποιώντας κατάλληλα την $\delta_h^c f$. Πράγματι, έχουμε

$$\begin{aligned} \delta_{h/2}^c \delta_{h/2}^c f(x_0) = & \frac{\delta_{h/2}^c f(x_0 + h/2) - \delta_{h/2}^c f(x_0 - h/2)}{h} \\ = & \frac{1}{h}\left(\frac{f(x_0 + \frac{h}{2} + \frac{h}{2}) - f(x_0 + \frac{h}{2} - \frac{h}{2})}{h} - \frac{f(x_0 - \frac{h}{2} + \frac{h}{2}) - f(x_0 - \frac{h}{2} - \frac{h}{2})}{h}\right) \\ = & \frac{f(x_0+h) - 2f(x_0) + f(x_0-h)}{h^2}. \end{aligned}$$

Συμβολίζουμε λοιπόν $\delta^c_{h,2}f$ την ακόλουθη πεπερασμένη διαφορά,

$$\delta^c_{h,2}f(x_0) \equiv \frac{f(x_0+h)-2f(x_0)+f(x_0-h)}{h^2}. \tag{2.8}$$

Αυτός λοιπόν ο λόγος ορίζει μια προσέγγιση της $f''(x_0)$, ο οποίος καλείται και *κεντρική διαφορά για την προσέγγιση της δεύτερης παραγώγου*. Άρα σύμφωνα με τα παραπάνω, θα έχουμε ότι

$$\delta^c_{h,2}f(x_0) = \delta^+_h\delta^-_h f(x_0) = \delta^-_h\delta^+_h f(x_0) = \delta^c_{h/2}\delta^c_{h/2}f(x_0).$$

Επίσης, με ανάλογο τρόπο όπως στο Λήμμα 2.1 μπορούμε να δείξουμε την ακόλουθη εκτίμηση του σφάλματος της προσέγγισης της $f''(x_0)$.

Λήμμα 2.2. *Έστω $f : [a,b] \to \mathbb{R}$, $f \in C^4[a,b]$, $x_0 \in (a,b)$ και $h > 0$, τέτοιο ώστε $x_0 \pm h \in [a,b]$. Τότε ισχύει η ακόλουθη εκτίμηση*

$$|\delta^c_{h,2}f(x_0) - f''(x_0)| \le \frac{h^2}{12}\max_{x\in[a,b]}|f^{(4)}(x)|. \tag{2.9}$$

Απόδειξη. Για $f \in C^4[a,b]$, μπορούμε να αναπτύξουμε και πάλι κατά Taylor και να πάρουμε τις παρακάτω δύο σχέσεις

$$\begin{aligned} f(x_0+h) &= f(x_0) + hf'(x_0) + \frac{h^2}{2}f''(x_0) + \frac{h^3}{6}f'''(x_0) + \frac{h^4}{24}f^{(4)}(\zeta_1), \\ f(x_0-h) &= f(x_0) - hf'(x_0) + \frac{h^2}{2}f''(x_0) - \frac{h^3}{6}f'''(x_0) + \frac{h^4}{24}f^{(4)}(\zeta_2), \end{aligned} \tag{2.10}$$

με $\zeta_1 \in (x_0, x_0+h)$, $\zeta_2 \in (x_0-h, x_0)$. Προσθέτοντας κατά μέλη τις δύο αυτές σχέσεις, έχουμε

$$f(x_0+h)+f(x_0-h) = 2f(x_0) + h^2f''(x_0) + \frac{h^4}{24}[f^{(4)}(\zeta_1)+f^{(4)}(\zeta_2)].$$

Συνεπώς, σύμφωνα με το θεώρημα ενδιάμεσης τιμής, έχουμε

$$\frac{f(x_0+h)-2f(x_0)+f(x_0-h)}{h^2} = f''(x_0) + \frac{h^2}{12}f^{(4)}(\xi),\ \xi \in (\zeta_2,\zeta_1), \tag{2.11}$$

από όπου εύκολα προκύπτει η ζητούμενη εκτίμηση (2.9). □

Παρατήρηση 2.4. Για τον ίδιο λόγο όπως και στην Παρατήρηση 2.2, η τάξη ακρίβειας της προσέγγισης $\delta^c_{h,2}f$ είναι δύο. Μάλιστα, αν θεωρήσουμε την $f(x) = x^4$, βλέπουμε ότι $\delta^c_{h,2}f(x) = 12x^2 + 2h^2$, δηλαδή η τάξη ακρίβειας της $\delta^c_{h,2}f$ είναι ακριβώς δύο.

h	$\delta^c_{h,2} f(1.1)$	$E_{\delta^c_{h,2}}$	p
0.50	-0.92577	0.09932	
0.10	-0.82988	0.00343	2.090
0.05	-0.82730	0.00085	2.005
0.01	-0.82648	0.00003	2.000

Πίνακας 2.3: Οι τιμές της $\delta^c_{h,2}$ και το σφάλμα $E_{\delta^c_{h,2}}$, της προσέγγισης της $f''(1.1) = -1/(1.1)^2 \approx -0.82645$, καθώς και η προσεγγιστική τάξη ακρίβειας p.

Στον Πίνακα 2.3 δίνουμε τιμές της προσέγγισης $\delta^c_{h,2} f(x_0)$ για τη συνάρτηση $f(x) = \ln(x)$ στο $x_0 = 1.1$, όπου $f''(1.1) = -1/(1.1)^2 \approx -0.82645$, καθώς και το σφάλμα $E_{\delta^c_{h,2}} = |\delta^c_{h,2} f(1.1) - f''(1.1)|$ και την αντίστοιχη προσεγγιστική τάξη ακρίβειας p.

Παρατήρηση 2.5. Ένα ερώτημα που τίθεται είναι αν θα μπορούσαμε να χρησιμοποιήσουμε την $\delta^-_h \delta^-_h f(x_0)$ ή την $\delta^+_h \delta^+_h f(x_0)$ για την προσέγγιση της $f''(x_0)$. Μπορούμε να δούμε ότι αν $f : [a,b] \to \mathbb{R}$, $f \in C^3[a,b]$, $x_0 \in (a,b)$ και $h > 0$, τέτοιο ώστε $x_0 \pm h \in [a,b]$, τότε ισχύει η ακόλουθη εκτίμηση, βλ. Άσκηση 2.2,

$$|\delta^+_h \delta^+_h f(x_0) - f''(x_0)| \leq \frac{5}{3} h \max_{x \in [a,b]} |f^{(3)}(x)|. \tag{2.12}$$

Επομένως, η $\delta^+_h \delta^+_h f(x)$ είναι προσέγγιση της $f''(x)$, όμως η τάξη ακρίβειας είναι ένα. Παρόμοια εκτίμηση μπορούμε να δείξουμε για την $\delta^-_h \delta^-_h f$.

Χρησιμοποιώντας τις προσεγγίσεις για την πρώτη και τη δεύτερη παράγωγο μιας συνάρτησης f που θεωρήσαμε, στα επόμενα κεφάλαια θα κατασκευάσουμε μεθόδους πεπερασμένων διαφορών για τη προσέγγιση της λύσης διαφορικών εξισώσεων πρώτης και δεύτερης τάξεως, καθώς και, παραδείγματος χάριν, στα (Ακρίβης & Δουγαλής, 2013· Ακρίβης & Δουγαλής, 2005· Holmes, 2007· Richtmyer & Morton, 1967· Jovanović & Süli, 2014· Strikwerda, 2004).

2.3 Ασκήσεις

2.1. Έστω $f : [a,b] \to \mathbb{R}$, $f \in C^4[a,b]$, $x_0 \in (a,b)$ και $h_1, h_2 > 0$, τέτοιο ώστε $x_0 - h_1, x_0 + h_2 \in [a,b]$. Γράψτε την πεπερασμένη διαφορά για την προσέγγιση της $f''(x_0)$ χρησιμοποιώντας τις τιμές της f στα $x_0 - h_1$, x_0 και $x_0 + h_2$. Εκτιμήστε το σφάλμα της προσέγγισης της $f''(x_0)$. Είναι η τάξη ακρίβειας δύο;

2.2. Έστω $f : [a,b] \to \mathbb{R}$, $f \in C^3[a,b]$, $x_0 \in (a,b)$ και $h > 0$, τέτοιο ώστε $x_0 \pm h \in [a,b]$. Τότε ισχύει η ακόλουθη εκτίμηση

$$|\delta_h^+ \delta_h^+ f(x_0) - f''(x_0)| \leq \frac{5}{3} h \max_{x\in[a,b]} |f^{(3)}(x)|.$$

2.3. Έστω $f : [a,b] \to \mathbb{R}$, $f \in C^3[a,b]$, $x_0 \in (a,b)$ και $h > 0$, τέτοιο ώστε $x_0 \pm h \in [a,b]$. Τότε ισχύει η ακόλουθη εκτίμηση

$$|\delta_h^- \delta_h^- f(x_0) - f''(x_0)| \leq \frac{5}{3} h \max_{x\in[a,b]} |f^{(3)}(x)|.$$

2.4. Έστω $f : [a,b] \to \mathbb{R}$, $f \in C^5[a,b]$, $x_0 \in (a,b)$ και $h > 0$, τέτοιο ώστε $x_0 \pm 2h \in [a,b]$. Χρησιμοποιήστε τα σημεία $x_0 - 2h$, $x_0 - h$, $x_0 + h$, $x_0 + 2h$, για να δείξτε ότι η πεπερασμένη διαφορά

$$\hat{\delta}_h f(x_0) = \frac{1}{12h}(-f(x_0+2h) + 8f(x_0+h) - 8f(x_0-h) + f(x_0-2h)),$$

προσεγγίζει την $f'(x_0)$. Βρείτε το σφάλμα της προσέγγισης και την τάξη ακρίβειας.

2.5. Έστω $f : [a,b] \to \mathbb{R}$, $f \in C^6[a,b]$, $x_0 \in (a,b)$ και $h > 0$, τέτοιο ώστε $x_0 \pm 2h \in [a,b]$. Χρησιμοποιήστε τα σημεία $x_0 - 2h$, $x_0 - h$, x_0, $x_0 + h$, $x_0 + 2h$, για να δείξετε ότι η πεπερασμένη διαφορά

$$\begin{aligned}\hat{\delta}_{h,2} f(x_0) = \frac{1}{12h^2}(&-f(x_0+2h) + 16f(x_0+h) - 30f(x_0) \\ &+ 16f(x_0-h) - f(x_0-2h)),\end{aligned}$$

προσεγγίζει την $f''(x_0)$. Βρείτε το σφάλμα της προσέγγισης και την τάξη ακρίβειας.

Βιβλιογραφία

Ακρίβης, Γ., & Δουγαλής, Β. (2005). *Αριθμητικές Μέθοδοι για Μερικές Διαφορικές Εξισώσεις*. Ιωάννινα. (Πανεπιστημιακές Σημειώσεις).

Ακρίβης, Γ., & Δουγαλής, Β. (2013). *Αριθμητικές Μέθοδοι για Συνήθεις Διαφορικές Εξισώσεις*. Πανεπιστημιακές Εκδόσεις Κρήτης, Ηράκλειο.

Holmes, M. H. (2007). *Introduction to numerical methods in differential equations* (Vol. 52). Springer, New York.

Jovanović, B. S., & Süli, E. (2014). *Analysis of finite difference schemes* (Vol. 46). Springer, London.

Richtmyer, R. D., & Morton, K. W. (1967). *Difference methods for initial-value problems*. Interscience Publishers John Wiley & Sons, Inc., New York-London-Sydney.

Strikwerda, J. C. (2004). *Finite difference schemes and partial differential equations* (Second ed.). Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA.

Κεφάλαιο 3

Πρόβλημα δύο σημείων

Σε αυτό το κεφάλαιο θα μελετήσουμε τη μεθόδο πεπερασμένων διαφορών για προβλήματα Σ.Δ.Ε. δεύτερης τάξεως, τα οποία καλούνται και προβλήματα δύο σημείων. Ο λόγος που θα ασχοληθούμε με την προσέγγιση της λύσης αυτών των προβλημάτων είναι για την καλύτερη κατανόηση των αριθμητικών μεθόδων για προβλήματα Μ.Δ.Ε. δεύτερης τάξεως που θα συναντήσουμε στα επόμενα κεφάλαια. Θα θεωρήσουμε μια Σ.Δ.Ε. δεύτερης τάξεως και έναν ομοιόμορφο διαμερισμό του πεδίου ορισμού της. Θα κατασκευάσουμε προσεγγίσεις των παραγώγων της άγνωστης λύσης u της διαφορικής εξίσωσης, χρησιμοποιώντας τις προσεγγίσεις που θεωρήσαμε στο Κεφάλαιο 2. Με αυτό τον τρόπο, θα οδηγηθούμε σε ένα νέο πρόβλημα του οποίου η λύση αποτελεί προσέγγιση των τιμών της ακριβούς λύσης στα σημεία του διαμερισμού. Θα δούμε, επίσης, ότι αυξάνοντας τον αριθμό των σημείων της διαμέρισης, οδηγούμαστε σε καλύτερη προσέγγιση της λύσης του προβλήματος.

3.1 Διακριτοποίηση

Θεωρούμε το πρόβλημα δύο σημείων για μια συνήθη διαφορική εξίσωση δεύτερης τάξης, με ομογενείς συνοριακές συνθήκες Dirichlet: Ζητείται μια συνάρτηση $u \in C^2[a,b]$, τέτοια ώστε

$$-u''(x) + q(x)u(x) = f(x), \quad \text{για } x \in [a,b], \quad \text{με } u(a) = u(b) = 0, \qquad (3.1)$$

όπου $a, b \in \mathbb{R}$, $a < b$, $q, f \in C[a,b]$ και $q(x) \geq 0$, για κάθε $x \in [a,b]$. Στη συνέχεια, θα συμβολίζουμε με $q_{\min} = \min_{x\in[a,b]} q(x)$.

Θα θεωρήσουμε έναν φυσικό αριθμό N και μια διαμέριση του διαστήματος $[a,b]$ από $N+2$ ισαπέχοντα σημεία $a = x_0 < x_1 < \cdots < x_N < x_{N+1} = b$,

όπου $h = x_{i+1} - x_i$, $i = 0, \ldots, N$. Τότε, σε κάθε σημείο του διαμερισμού x_i, $i = 1, \ldots, N$, θα ισχύει:

$$-u''(x_i) + q(x_i)u(x_i) = f(x_i), \quad i = 1, \ldots, N. \tag{3.2}$$

Σκοπός μας είναι να κατασκευάσουμε προσεγγίσεις των τιμών $u(x_i)$ της ακριβούς λύσης του (3.1), τις οποίες θα συμβολίζουμε με U_i, $i = 0, \ldots, N+1$. Λόγω των συνοριακών συνθηκών $u(x_0) = u(x_{N+1}) = 0$, θέτουμε λοιπόν $U_0 = U_{N+1} = 0$. Οι τιμές των U_i, $i = 1, \ldots, N$, προκύπτουν με τον ακόλουθο τρόπο.

Για να προσεγγίσουμε την $u''(x)$ στα σημεία $x_i, i = 1, \ldots, N$, χρησιμοποιούμε την προσέγγιση $\delta^c_{h,2}$ που θεωρήσαμε στην (2.8). Έτσι, αν υποθέσουμε ότι $u \in C^4[a, b]$, λόγω της (2.9), η (3.2) γίνεται,

$$-\frac{u(x_{i+1}) - 2u(x_i) + u(x_{i-1})}{h^2} + q(x_i)u(x_i) = f(x_i) + \eta_i, \quad i = 1, \ldots, N, \tag{3.3}$$

όπου

$$|\eta_i| \leq \frac{h^2}{12} \max_{x\in[a,b]} |u^{(4)}(x)|. \tag{3.4}$$

Συνεπώς, έχουμε ότι ισχύει το ακόλουθο λήμμα.

Λήμμα 3.1. *Έστω u η λύση του* (3.1) *με $u \in C^4[a, b]$. Τότε για την η_i που δίνεται στην* (3.3) *έχουμε ότι υπάρχει σταθερά C ανεξάρτητη του h, τέτοια ώστε*

$$\max_{1\leq i\leq N} |\eta_i| \leq Ch^2. \tag{3.5}$$

Για να κατασκευάσουμε, λοιπόν, προσεγγίσεις U_i των $u(x_i), i = 0, \ldots, N+1$, θεωρούμε τις ακόλουθες εξισώσεις

$$-\frac{U_{i+1} - 2U_i + U_{i-1}}{h^2} + q(x_i)U_i = f(x_i), \quad i = 1, \ldots, N, \tag{3.6}$$

$$U_0 = U_{N+1} = 0. \tag{3.7}$$

Επομένως, οι εξισώσεις (3.6)–(3.7) δίνουν ένα αριθμητικό σχήμα για την προσέγγιση της ακριβούς λύσης του προβλήματος (3.1). Τότε, για αυτό το αριθμητικό σχήμα, το σφάλμα η_i που δίνεται στην (3.3) καλείται *τοπικό σφάλμα διακριτοποίησης*.

Αν το τοπικό σφάλμα διακριτοποίησης ενός αριθμητικού σχήματος σε ένα σημείο ενός διαμερισμού με βήμα h, φράσσεται κατά απόλυτη τιμή από το γινόμενο μιας θετικής σταθεράς, ανεξάρτητης του h, επί h υψωμένο σε μια δύναμη $\kappa \geq 1$, τότε λέμε ότι το τοπικό σφάλμα διακριτοποίησης έχει *τάξη ακρίβειας* κ.

Στην (3.4) η ποσότητα h είναι υψωμένη στη δεύτερη δύναμη, οπότε η *τάξη ακρίβειας* του τοπικού σφάλματος διακριτοποίησης της μεθόδου (3.6)–(3.7) είναι δύο.

Μια επιθημητή ιδιότητα του τοπικού σφάλματος διακριτοποίησης είναι αυτό να τείνει στο μηδέν κατά απόλυτη τιμή, καθώς το βήμα h του διαμερισμού τείνει στο μηδέν. Αυτή η ιδιότητα καλείται *συνέπεια* ενός αριθμητικού σχήματος.

Ορισμός 3.1. Ένα αριθμητικό σχήμα ή μια μέθοδος λέγεται *συνεπές ή συνεπής*, αντίστοιχα, αν υπό κατάλληλες συνθήκες ομαλότητας της συνάρτησης που θέλουμε να προσεγγίσουμε, το τοπικό σφάλμα διακριτοποίησης τείνει στο μηδέν, καθώς το βήμα h του διαμερισμού, τείνει στο μηδέν.

Επομένως, λόγω του Λήμματος 3.1, η μέθοδος πεπερασμένων διαφορών (3.6)–(3.7) είναι *συνεπής*.

Αν συμβολίσουμε τώρα με $U \in \mathbb{R}^N$ το διάνυσμα με συνιστώσες $U_1, \ldots, U_N$, $U = (U_1, \ldots, U_N)^T$, μπορούμε να γράψουμε το σύστημα των εξισώσεων (3.6)-(3.7) ισοδύναμα ως γραμμικό σύστημα

$$(A + h^2 Q)U = h^2 F, \tag{3.8}$$

όπου A είναι ο $N \times N$ πίνακας

$$A = \begin{pmatrix} 2 & -1 & 0 & & 0 \\ -1 & 2 & -1 & 0 & \\ 0 & \ddots & \ddots & \ddots & 0 \\ & & -1 & 2 & -1 \\ & & & -1 & 2 \end{pmatrix}, \tag{3.9}$$

Q είναι ένας διαγώνιος $N \times N$ πίνακας με στοιχεία $q(x_i)$, $i = 1, \ldots, N$, στη διαγώνιο και $F = (f(x_1), \ldots, f(x_N))^T$.

Στη συνέχεια, δίνουμε τον ακόλουθο ορισμό:

Ορισμός 3.2. Ένας $N \times N$ πίνακας A με στοιχεία (a_{ij}) λέμε ότι έχει *αυστηρά κυριαρχική διαγώνιο*, αν

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

Οι πίνακες με αυστηρά κυριαρχική διαγώνιο είναι αντιστρέψιμοι, βλ. π.χ. (Ακρίβης & Δουγαλής, 2015, Πρόταση 3.3). Εύκολα μπορούμε να δούμε ότι ο πίνακας $A + h^2 Q$ της (3.8) έχει αυστηρά κυριαρχική διαγώνιου, αν $q_{\min} > 0$, και, επομένως, το

γραμμικό σύστημα (3.8) έχει μοναδική λύση. Όμως, όπως θα δούμε στη συνέχεια, η υπόθεση ότι $q_{\min} > 0$ δεν είναι απαραίτητη για την ύπαρξη μοναδικής λύσης του (3.8).

Επίσης, παρατηρούμε ότι ο πίνακας A στην (3.9) είναι *τριδιαγώνιος*, δηλαδή αν με a_{ij}, $i, j = 1, \ldots, N$, συμβολίζουμε τα στοιχεία του A, τότε $a_{ij} = 0$, για $|i - j| > 1$. Προφανώς, τότε και ο $A + h^2 Q$ είναι τριδιαγώνιος. Στην επόμενη παράγραφο θα δούμε έναν αλγόριθμο για την επίλυση ενός γραμμικού συστήματος με τριδιαγώνιο πίνακα, χωρίς να είναι απαραίτητο ο πίνακας αυτός να έχει αυστηρά κυριαχική διαγώνιο, βλέπε π.χ. (Ακρίβης & Δουγαλής, 2015, Κεφάλαιο 3).

3.2 Επίλυση τριδιαγώνιου γραμμικού συστήματος

Έστω ότι θέλουμε να λύσουμε το γραμμικό σύστημα $Ay = z$, δηλαδή να βρούμε το $y \in \mathbb{R}^N$, όπου A είναι ένας $N \times N$ τριδιαγώνιος πίνακας με στοιχεία

$$A = \begin{pmatrix} a_1 & b_1 & & 0 & \\ c_2 & a_2 & b_2 & & \\ & \ddots & \ddots & \ddots & \\ & 0 & c_{N-1} & a_{N-1} & b_{N-1} \\ & & & c_N & a_N \end{pmatrix}, \tag{3.10}$$

και $z \in \mathbb{R}^N$ ένα δοσμένο διάνυσμα. Για τα στοιχεία του πίνακα A θα κάνουμε τις ακόλουθες υποθέσεις

$$|a_1| > |b_1|, \quad |a_k| \geq |b_k| + |c_k|,\ k = 2, \ldots, N-1, \quad |a_N| > |c_N|. \tag{3.11}$$

Για να λύσουμε το γραμμικό σύστημα $Ay = z$ μπορούμε να εφαρμόσουμε διάφορους αλγόριθμους όπως είναι η απαλοιφή Gauss. Στην περίπτωση, όμως, του πίνακα A, είναι προτιμότερο να χρησιμοποιήσουμε έναν αλγόριθμο που έχει σχεδιαστεί ειδικά για τριδιαγώνιους πίνακες, όπως ο ακόλουθος:

Ο πίνακας A μπορεί να γραφεί ως *γινόμενο* LU δύο πινάκων L και U, με L κάτω τριγωνικό και U άνω τριγωνικό, που έχουν τη μορφή

$$L = \begin{pmatrix} d_1 & & & \\ c_2 & d_2 & 0 & \\ & \ddots & \ddots & \\ & 0 & c_N & d_N \end{pmatrix}, \; U = \begin{pmatrix} 1 & e_1 & & & \\ & 1 & e_2 & 0 & \\ & & \ddots & \ddots & \\ & & & 1 & e_{N-1} \\ & & & 0 & 1 \end{pmatrix}, \tag{3.12}$$

δηλαδή έχουν μη μηδενικά στοιχεία στη διαγώνιο και ο L στην πρώτη υποδιαγώνιο και ο U στην πρώτη υπερδιαγώνιο. Είναι απλό να δούμε ότι οι αριθμοί $d_1, \ldots, d_N$ και $e_1, \ldots, e_{N-1}$ προκύπτουν με τον ακόλουθο αλγόριθμο

$$
\begin{aligned}
&d_1 = a_1, e_1 = b_1/d_1 \\
&\text{για } k = 2, 3, \ldots, N-1 \\
&\qquad d_k = a_k - c_k e_{k-1} \\
&\qquad e_k = b_k/d_k \\
&\text{τέλος για} \\
&d_N = a_k - c_N e_{N-1}.
\end{aligned}
\tag{3.13}
$$

Η υπάρξη των πινάκων L και U και η ολοκλήρωση του αλγορίθμου (3.13) αποδεικνύεται στο ακόλουθο λήμμα.

Λήμμα 3.2. *Έστω A ένας τριδιαγώνιος πίνακας της μορφής* (3.10)*, τέτοιος ώστε να ισχύουν οι υποθέσεις* (3.11)*. Τότε υπάρχουν πίνακες L και U της μορφής* (3.12)*, τέτοιοι ώστε $A = LU$ και ο αλγόριθμος* (3.13) *είναι καλά ορισμένος και ολοκληρώνεται.*

Απόδειξη. Για να είναι ο αλγόριθμος (3.13) καλά ορισμένος και, συνεπώς, να υπάρχει η ανάλυση του $A = LU$ με πίνακες L και U της μορφής (3.12), αρκεί να ισχύει $d_k \neq 0$, $k = 1, \ldots, N$. Από τις υποθέσεις (3.11) έχουμε ότι $|a_1| > |b_1|$, οπότε $|e_1| < 1$. Επαγωγικά μπορούμε να αποδείξουμε ότι

$$d_k \neq 0, k = 1, \ldots, N, \quad |e_k| < 1.$$

Πράγματι, έστω ότι ισχύει $d_{k-1} \neq 0$, $|e_{k-1}| < 1$ για κάποιο k. Τότε $|d_k| = |a_k - c_k e_{k-1}| \geq |a_k| - |c_k|\,|e_{k-1}| > |a_k| - |c_k| \geq |b_k| > 0$. Επιπλέον $|e_k| = |b_k|/|d_k| < 1$. □

Εφόσον έχουμε αποδείξει ότι $A = LU$, για να λύσουμε τώρα το γραμμικό σύστημα $LUy = z$, λύνουμε πρώτα το $Lw = z$ εφαρμόζοντας τον ακόλουθο αλγόριθμο

$$
\begin{aligned}
&w_1 = z_1/d_1 \\
&\text{για } k = 2, 3, \ldots, N-1 \\
&\qquad w_k = (z_k - c_k w_{k-1})/d_k \\
&\text{τέλος για}
\end{aligned}
\tag{3.14}
$$

και, στη συνέχεια, το διάνυσμα y προκύπτει ως λύση του γραμμικού συστήματος $Uy = w$

$$
\begin{aligned}
&y_N = w_N \\
&\text{για } k = N-1, N-2, \ldots, 1 \\
&\qquad y_k = w_k - e_k y_{k+1} \\
&\text{τέλος για}
\end{aligned}
\tag{3.15}
$$

Παρατήρηση 3.1. Παρατηρούμε, λοιπόν, ότι αν ένας πίνακας της μορφής (3.10) δεν έχει αυστηρά κυριαρχική διαγώνιο, αλλά ικανοποιεί τις υποθέσεις (3.11), τότε αντιστρέφεται. Στο πρόβλημα (3.1), υποθέσαμε ότι $q_{\min} \geq 0$. Αυτό σημαίνει ότι ο πίνακας $A + h^2 Q$ που θεωρήσαμε στην (3.8), ικανοποιεί τις υποθέσεις (3.11). Συνεπώς, εφαρμόζοντας τους αλγορίθμους (3.14) και (3.15), διαπιστώνουμε ότι το γραμμικό σύστημα (3.8) έχει μοναδική λύση.

Παράδειγμα 3.1. Θεωρούμε το ακόλουθο πρόβλημα συνοριακών τιμών

$$-u''(x) + u(x) = \sin(2\pi x), \quad 0 < x < 1, \quad \text{με } u(0) = u(1) = 0. \tag{3.16}$$

Η ακριβής λύση u αυτού του προβλήματος είναι

$$u(x) = \frac{\sin(2\pi x)}{1 + 4\pi^2}, \tag{3.17}$$

οπότε η (3.6) γίνεται τώρα

$$-\frac{U_{i+1} - 2U_i + U_{i-1}}{h^2} + U_i = \sin(2\pi x_i), \quad i = 1, \ldots, N. \tag{3.18}$$

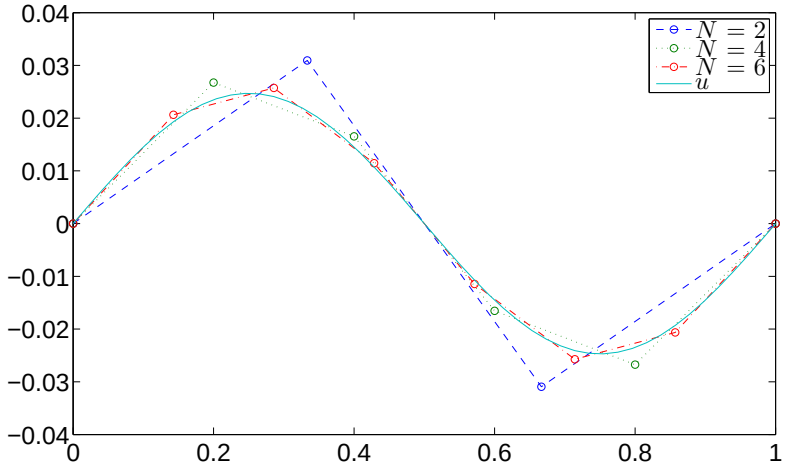
Διαμερίζουμε το $[0, 1]$ σε $N + 2$ ισαπέχοντα σημεία και χρησιμοποιώντας τους παραπάνω αλγορίθμους (3.13), (3.14) και (3.15), μπορούμε να βρούμε U_i που προσεγγίζουν την ακριβή λύση στο αντίστοιχο σημείο x_i του διαμερισμού. Στο Σχήμα 3.1 εμφανίζονται οι προσεγγίσεις για $N = 2, 4$ και 6 και το γράφημα της ακριβούς λύσης u. Σε αυτό το σχήμα, μπορούμε να παρατηρήσουμε ότι καθώς το πλήθος των σημείων της διαμέρισης αυξάνει, και άρα το βήμα h ελαττώνεται, φαίνεται ότι η αντίστοιχη προσέγγιση πλησιάζει καλύτερα την ακριβή λύση. Αυτό εμφανίζεται και στον Πίνακα 3.1, όπου βλέπουμε τα σφάλματα της προσέγγισης της $u(x_i)$ από την U_i στα σημεία x_i, με $i = 1, 3$ και $N = 5, 7, 9, 11$. Το σημείο x_1 είναι το πρώτο σημείο στο οποίο υπολογίζουμε την προσέγγιση και, καθώς το N αυξάνει, αυτό θα πλησιάζει το σημείο $x_0 = 0$. Αν και το σημείο x_1 μεταβάλλεται για κάθε διαμερισμό, το σφάλμα $|U_1 - u(x_1)|$ ελαττώνεται, καθώς το N αυξάνει. Αυτό ισχύει για όλα τα σημεία του διαμερισμού και πραγματικά η ποσότητα $\max_{1 \leq i \leq N} |U_i - u(x_i)|$ φθίνει, καθώς το πλήθος των σημείων αυξάνει. Αυτήν την προσεγγιστική ιδιότητα που έχει η μέθοδος πεπερασμένων διαφορών (3.6)–(3.7) αποδεικνύουμε παρακάτω στο Θεώρημα 3.2.

3.3 Σύγκλιση της μεθόδου πεπερασμένων διαφορών

Σε αυτή την παράγραφο θα δείξουμε ότι η λύση U του γραμμικού συστήματος (3.8) προσεγγίζει το αντίστοιχο διάνυσμα των τιμών της ακριβούς λύσης u, με συνιστώσες τις τιμές $u(x_1), \ldots, u(x_N)$.

N	x_1	$\lvert U_1 - u(x_1)\rvert$	x_3	$\lvert U_3 - u(x_3)\rvert$	$\max_{0\le i\le N+1} \lvert U_i - u(x_i)\rvert$
5	0.2	0.0051	0.6	0.046	0.0080
7	0.1429	0.0021	0.4286	0.0072	0.0072
9	0.1111	0.0010	0.3333	0.0056	0.0070
11	0.0909	0.0005	0.2727	0.0041	0.0069

Πίνακας 3.1: Τα σφάλματα του Παραδείγματος 3.1 στα σημεία x_1 και x_3 και το μέγιστο σφάλμα σε κάθε σημείο του διαμερισμού, για $N = 5, 7, 9, 11$.



Σχήμα 3.1: Η ακριβής λύση και οι προσεγγιστικές λύσεις του Παραδείγματος 3.1 για $N = 2, 4, 6$.

Ορισμός 3.3. Έστω H ένας πραγματικός γραμμικός χώρος και ϕ μια απεικόνιση $\phi : H \to \mathbb{R}$. Η ϕ καλείται *νόρμα* στον H, αν ικανοποιεί τις ακόλουθες ιδιότητες

$$\begin{aligned} &\text{Για } v \in H, \phi(v) = 0, \text{ αν και μόνο αν } v = 0, \\ &\phi(\lambda v) = |\lambda| v, \quad \forall \lambda \in \mathbb{R},\ v \in H, \\ &\phi(v + w) \le \phi(v) + \phi(w). \end{aligned} \tag{3.19}$$

Είναι απλό να δούμε ότι η απεικόνιση $\phi(v) = \max_{1\le i\le N} |v_i|$, όπου $v \in \mathbb{R}^N$, αποτελεί μια νόρμα στον $\mathbb{R}^N$.

Ορισμός 3.4. Μια αριθμητική μέθοδος για το πρόβλημα 3.1 λέγεται *ευσταθής*, αν μια νόρμα της αριθμητικής λύσης φράσσεται από μια σταθερά επί μια ποσότητα που εξαρτάται μόνο από τα δεδομένα του προβλήματος.

Στο ακόλουθο θεώρημα δείχνουμε την ευστάθεια της λύσης του (3.6)–(3.7).

Θεώρημα 3.1. *Έστω U_i, $i = 0, \ldots, N+1$, η λύση του προβλήματος* (3.6)–(3.7)*, και $q_{\min} > 0$. Τότε, υπάρχει σταθερά C ανεξάρτητη του h, τέτοια ώστε*

$$\max_{0\le i\le N+1} |U_i| \le C \max_{x\in[a,b]} |f(x)|. \tag{3.20}$$

Απόδειξη. Γράφουμε τη σχέση (3.6) στη μορφή

$$(2 + h^2 q(x_i))U_i = U_{i+1} + U_{i-1} + h^2 f(x_i), \quad 1 \le i \le N.$$

Στη συνέχεια, επειδή $q_{\min} > 0$, η παραπάνω ισότητα δίνει, για κάθε $i = 1, \ldots, N$,

$$\begin{aligned}(2 + h^2 q_{\min})|U_i| &\le |U_{i+1}| + |U_{i-1}| + h^2 |f(x_i)| \\ &\le 2 \max_{0\le i\le N+1} |U_i| + h^2 \max_{x\in[a,b]} |f(x)|.\end{aligned}$$

Επομένως,

$$(2 + h^2 q_{\min}) \max_{1\le i\le N} |U_i| \le 2 \max_{0\le i\le N+1} |U_i| + h^2 \max_{x\in[a,b]} |f(x)|,$$

η οποία εύκολα δίνει τη ζητούμενη εκτίμηση (3.20). □

Παρατήρηση 3.2. Η ευστάθεια του αριθμητικού σχήματος είναι μια εσωτερική ιδιότητα του σχήματος, δηλαδή δεν έχει σχέση με το συγκεκριμμένο πρόβλημα που θέλουμε να λύσουμε. Χρησιμοποιώντας το Θεώρημα 3.1 μπορούμε να αποδείξουμε ότι το γραμμικό σύστημα (3.8) έχει μοναδική λύση. Αν θεωρήσουμε το αντίστοιχο ομογενές γραμμικό σύστημα, τότε σύμφωνα με το Θεώρημα 3.1, οδηγούμαστε στο συμπέρασμα ότι η μοναδική λύση είναι η μηδενική λύση $U_i = 0, i = 0, \ldots, N+1$.

Στη συνέχεια, χρησιμοποιώντας την ευστάθεια και τη συνέπεια του αριθμητικού σχήματος (3.6)–(3.7), δείχνουμε τη σύγκλισή του.

Θεώρημα 3.2. *Έστω U_i, $i = 0, \ldots, N+1$, η λύση του προβλήματος* (3.6)–(3.7)*, και u η λύση του προβλήματος* (3.1)*, με $u \in C^4[a, b]$. Τότε, αν $q_{\min} > 0$, υπάρχει μια σταθερά C, ανεξάρτητη του h, τέτοια ώστε*

$$\max_{0\le i\le N+1} |U_i - u(x_i)| \le Ch^2. \tag{3.21}$$

Απόδειξη. Θέτουμε $E_i = U_i - u(x_i)$, $i = 0, \ldots, N+1$, όπου λόγω των σχέσεων $U_0 = u(a) = 0$ και $U_{N+1} = u(b) = 0$, έχουμε $E_0 = E_{N+1} = 0$. Αφαιρούμε τώρα κατά μέλη τις (3.6) και (3.3), οπότε παίρνουμε

$$E_{i+1} - (2 + q(x_i)h^2)E_i + E_{i-1} = h^2\eta_i, \quad i = 1, \ldots, N. \tag{3.22}$$

Θέτουμε τώρα $\bar{E} = \max_{1\le i\le N} |E_i|$, $\bar{\eta} = \max_{1\le i\le N} |\eta_i|$ και επειδή $q_{\min} > 0$, από την (3.22) προκύπτει

$$(2 + q_{\min}h^2)|E_i| \le 2\bar{E} + h^2\bar{\eta}.$$

Συνεπώς

$$q_{\min}h^2 \max_{1\le i\le N} |E_i| \le h^2\bar{\eta},$$

η οποία λόγω του Λήμματος 3.1 δίνει τη ζητούμενη ανισότητα. □

Παράδειγμα 3.2. Θεωρούμε και πάλι το πρόβλημα του Παραδείγματος 3.1 και υπολογίζουμε το σφάλμα $\bar{E} = \max_{0\le i\le N+1} |U_i - u(x_i)|$, με $h = 1/(N+1)$ και $h = 0.1, 0.05, 0.0250, 0.0125$. Στον Πίνακα 3.2 βλέπουμε το σφάλμα $\bar{E}$ και την αντίστοιχη προσεγγιστική τάξη ακρίβειας p, όπου παρατηρούμε ότι φαίνεται να τείνει στο δύο, καθώς το h ελαττώνεται.

h	$\bar{E}$	p
0.1000	0.00077	
0.0500	0.00020	1.948
0.0250	0.00005	2.005
0.0125	0.00001	2.001

Πίνακας 3.2: Το σφάλμα $\bar{E} = \max_{0\le i\le N+1} |U_i - u(x_i)|$ της λύσεως του (3.18) στο Παράδειγμα 3.1, όπου $h = 1/(N+1)$, και η κατά προσέγγιση τάξη ακρίβειας p.

3.4 Μέθοδος ενέργειας

Στο Θεώρημα 3.2 είδαμε ότι αν $q_{\min} > 0$, τότε μπορούμε να δείξουμε ότι τα U_i που ικανοποιούν την (3.6)–(3.7) συγκλίνουν στις $u(x_i)$, $i = 0, \ldots, N+1$, όπου u η ακριβής του (3.1). Σε αυτή την παράγραφο θα δούμε μια διαφορετική απόδειξη της παραπάνω σύγκλισης, όπου $q_{\min}$ μπορεί να μηδενίζεται. Επίσης, αυτή η ανάλυση μας επιτρέπει να δείξουμε εκτιμήσεις σφάλματος παρόμοιες με την (3.21) για προβλήματα συνοριακών τιμών με διαφορετικές συνοριακές συνθήκες, όπως π.χ. με συνθήκες Neumann.

Συμβολίζουμε τώρα με $\mathbb{R}_0^{N+2}$ τα διανύσματα του $\mathbb{R}^{N+2}$ όπου η πρώτη και τελευταία συνιστώσα να είναι ίση με μηδέν, θεωρούμε το εσωτερικό γινόμενο $(\cdot,\cdot)_h$ στον $\mathbb{R}_0^{N+2}$, το οποίο ορίζεται ως

$$(V,W)_h = h\sum_{i=1}^{N} V_i W_i, \quad \text{με } V,W \in \mathbb{R}_0^{N+2}, \tag{3.23}$$

και την αντίστοιχη νόρμα $\|\cdot\|_h$,

$$\|V\|_h = \big(h\sum_{i=1}^{N} |V_i|^2\big)^{1/2}, \quad V \in \mathbb{R}_0^{N+2}. \tag{3.24}$$

Ακόμα, θεωρούμε και τη νόρμα $|\cdot|_{1,h}$, στον $\mathbb{R}_0^{N+2}$

$$|V|_{1,h} = \big(h\sum_{i=0}^{N} |\frac{V_{i+1}-V_i}{h}|^2\big)^{1/2}, \quad V \in \mathbb{R}_0^{N+2}. \tag{3.25}$$

Το ότι η $|\cdot|_{1,h}$ ορίζει μια νόρμα στον $\mathbb{R}_0^{N+2}$ μπορούμε εύκολα να το δούμε, επαληθεύοντας τις ιδιότητες (3.19), βλ. Άσκηση 3.1.

Παρατήρηση 3.3. Από το γεγονός ότι η $(\cdot,\cdot)_h$ είναι ένα εσωτερικό γινόμενο και $\|\cdot\|_h$ η αντίστοιχη νόρμα, έχουμε ότι ισχύει η *ανισότητα Cauchy–Schwarz*

$$|(V,W)_h| \le \|V\|_h \, \|W\|_h, \quad \forall V,W \in \mathbb{R}_0^{N+2}. \tag{3.26}$$

Στη συνέχεια, θα δείξουμε ανισότητες οι οποίες συνδέουν τις νόρμες (3.24) και (3.25) με τη νόρμα μεγίστου που χρησιμοποιήσαμε στην προηγούμενη παράγραφο.

Λήμμα 3.3. *Αν $V \in \mathbb{R}_0^{N+2}$, τότε ισχύουν οι ακόλουθες ανισότητες*

$$\|V\|_h \le \sqrt{b-a}\max_{0\le i\le N+1}|V_i|, \tag{3.27}$$

$$\max_{0\le i\le N+1}|V_i| \le \sqrt{b-a}\,|V|_{1,h}, \tag{3.28}$$

$$\|V\|_h \le (b-a)|V|_{1,h}. \tag{3.29}$$

Απόδειξη. Επειδή $(N+1)h = (b-a)$, εύκολα μπορούμε να δούμε ότι ισχύει η πρώτη ανισότητα (3.27). Για να δείξουμε τη δεύτερη ανισότητα (3.28), παρατη-

ρούμε ότι για $V \in \mathbb{R}_0^{N+2}$ και $j = 1, \ldots, N$, έχουμε

$$\begin{aligned} |V_j|^2 = |\sum_{i=0}^{j-1} V_{i+1} - V_i|^2 &\leq j \sum_{i=0}^{j-1} |V_{i+1} - V_i|^2 \\ &\leq (N+1) \sum_{i=0}^{N} |V_{i+1} - V_i|^2 = (N+1)h^2 \sum_{i=0}^{N} \left| \frac{V_{i+1} - V_i}{h} \right|^2 \\ &= (b-a)h \sum_{i=0}^{N} \left| \frac{V_{i+1} - V_i}{h} \right|^2 \leq (b-a) \, |V|_{1,h}^2. \end{aligned} \tag{3.30}$$

Τελικά, συνδυάζοντας τις (3.27) και (3.28), παίρνουμε την τελευταία ανισότητα, (3.29). □

Παρατήρηση 3.4. Λόγω του Λήμματος 3.3 και του Θεωρήματος 3.1 μπορούμε εύκολα να δούμε ότι αν θέσουμε $U \in \mathbb{R}_0^{N+2}$ με $U = (U_0, U_1, \ldots, U_{N+1})^T$ και U_i, $i = 0, \ldots, N+1$, τη λύση των (3.6)–(3.7) τότε

$$\|U\|_h \leq C\sqrt{b-a} \max_{x \in [a,b]} |f(x)|, \tag{3.31}$$

όπου C είναι η σταθερά της (3.20). Επίσης, αν $E \in \mathbb{R}_0^{N+2}$ με $E = (E_0, E_1, \ldots, E_{N+1})^T$ και $E_i = U_i - u(x_i)$, $i = 0, \ldots, N+1$, τότε από το Θεώρημα 3.2 έχουμε

$$\|E\|_h \leq Ch^2\sqrt{b-a}, \tag{3.32}$$

όπου C είναι τώρα η σταθερά της (3.21).

Στη συνέχεια, θα δείξουμε ότι η εκτίμηση σφάλματος (3.21) ισχύει και στην περίπτωση που $q_{\min} = 0$. Θα θεωρήσουμε τώρα την απεικόνιση $\Delta_h : \mathbb{R}_0^{N+2} \to \mathbb{R}_0^{N+2}$, η οποία ορίζεται ως

$$\Delta_h V_i \equiv (\Delta_h V)_i = \begin{cases} \frac{V_{i+1} - 2V_i + V_{i-1}}{h^2}, & i = 1, \ldots, N, \\ 0, & i = 0, N+1, \end{cases} \tag{3.33}$$

για την οποία ισχύει το ακόλουθο λήμμα.

Λήμμα 3.4. *Για $V, W \in \mathbb{R}_0^{N+2}$ έχουμε*

$$-(\Delta_h V, W)_h = h \sum_{i=0}^{N} \frac{V_{i+1} - W_i}{h} \, \frac{W_{i+1} - W_i}{h}. \tag{3.34}$$

Απόδειξη. Από τον ορισμό της $(\cdot,\cdot)_h$, αλλάζοντας τη μεταβλητή στη σειρά άθροισης και χρησιμοποιώντας το γεγονός ότι $W_0 = W_{N+1} = 0$, παίρνουμε

$$\begin{aligned}
(\Delta_h V, W)_h &= h\sum_{i=1}^{N} \frac{V_{i+1} - 2V_i + V_{i-1}}{h^2} W_i \\
&= h\left\{ \sum_{i=1}^{N} \frac{V_{i-1} - V_i}{h^2} W_i - \sum_{i=1}^{N} \frac{V_i - V_{i+1}}{h^2} W_i \right\} \\
&= h\left\{ \sum_{i=0}^{N-1} \frac{V_i - V_{i+1}}{h^2} W_{i+1} - \sum_{i=1}^{N} \frac{V_i - V_{i+1}}{h^2} W_i \right\} \\
&= h\left\{ \sum_{i=0}^{N} \frac{V_i - V_{i+1}}{h^2} W_{i+1} - \sum_{i=0}^{N} \frac{V_i - V_{i+1}}{h^2} W_i \right\} \\
&= -h\sum_{i=0}^{N} \frac{V_{i+1} - V_i}{h} \frac{W_{i+1} - W_i}{h}.
\end{aligned}$$

□

Στη συνέχεια, αποδεικνύουμε ένα παρόμοιο αποτέλεσμα με το Θεώρημα 3.2, όπου τώρα η συνάρτηση q επιτρέπεται να λαμβάνει και μηδενικές τιμές.

Θεώρημα 3.3. *Έστω $U \in \mathbb{R}_0^{N+2}$ με $U = (U_0, U_1, \ldots, U_{N+1})^T$, U_i, $i = 0, \ldots, N+1$, η λύση του προβλήματος* (3.6)–(3.7)*, όπου $q_{\min} \geq 0$, και u η λύση του προβλήματος* (3.1)*, με $u \in C^4[a,b]$. Τότε υπάρχει σταθερά C, ανεξάρτητη του h, τέτοια ώστε*

$$\max_{0 \leq i \leq N+1} |U_i - u(x_i)| \leq Ch^2. \tag{3.35}$$

Απόδειξη. Χρησιμοποιώντας τον ορισμό της απεικόνισης Δ_h, εύκολα βλέπουμε ότι η εξίσωση σφάλματος (3.22) μπορεί να γραφεί ως

$$-\Delta_h E_i + q(x_i)E_i = \eta_i, \quad i = 1, \ldots, N.$$

Στη συνέχεια, πολλαπλασιάζοντας με E_i κάθε εξίσωση και αθροίζοντας, παίρνουμε

$$-(\Delta_h E, E)_h + h\sum_{i=1}^{N} q(x_i)|E_i|^2 = (\eta, E)_h, \tag{3.36}$$

όπου $E, \eta \in \mathbb{R}_0^{N+2}$ τα διάνυσματα με συνιστώσες E_0, E_1, …, E_{N+1} και η_0, η_1, …, η_{N+1}, αντίστοιχα.

Μπορούμε να παρατηρήσουμε ότι λόγω του Λήμματος 3.4, έχουμε

$$-(\Delta_h V, V)_h = |V|_{1,h}^2, \quad \forall V \in \mathbb{R}_0^{N+2}. \tag{3.37}$$

Χρησιμοποιώντας τώρα την (3.37) και το γεγονός ότι $q_{\min} \geq 0$, η (3.36) δίνει

$$|E|_{1,h}^2 \leq |E|_{1,h}^2 + h\sum_{i=1}^{N} q(x_i)|E_i|^2 = (\eta, E)_h \leq \|\eta\|_h \|E\|_h. \tag{3.38}$$

Επομένως, εφαρμόζοντας την (3.29) στην (3.38), παίρνουμε

$$|E|_{1,h} \leq (b-a)\|\eta\|_h,$$

από όπου, λόγω της (3.28), προκύπτει

$$\max_{0\leq i\leq N+1} |E_i| \leq (b-a)^{3/2}\|\eta\|_h.$$

Στη συνέχεια, λόγω του Λήμματος 3.1, είναι απλό να δούμε ότι

$$\|\eta\|_h \leq Ch^2,$$

όπου η σταθερά C είναι ανεξάρτητη του h. Συνεπώς, συνδυάζοντας τις δύο τελευταίες ανισότητες παίρνουμε τη ζητούμενη (3.35). □

Παρατήρηση 3.5. Εύκολα βλέπουμε ότι, αν η συνάρτηση q επιτρέπεται να λαμβάνει και μηδενικές τιμές, τότε λόγω του Θεωρήματος 3.3 και του Λήμματος 3.3, ισχύει η εκτίμηση σφάλματος (3.32).

3.5 Συνοριακές συνθήκες Neumann

Θεωρούμε τώρα ένα παρόμοιο προβλήμα συνοριακών τιμών με το (3.1) όπου τώρα έχουμε τροποποιήσει τις συνοριακές συνθήκες και θεωρούμε ομογενείς συνοριακές συνθήκες Neumann. Ζητούμε μια συνάρτηση $u \in C^2[a,b]$, τέτοια ώστε

$$-u''(x) + q(x)u(x) = f(x), \quad x \in [a,b], \quad \text{με } u'(a) = u'(b) = 0, \tag{3.39}$$

όπου $a, b \in \mathbb{R}$, $a < b$, $q, f \in C[a,b]$ και $q_{\min} > 0$. Για αυτό το πρόβλημα, σε αντίθεση με εκείνο με συνοριακές συνθήκες Dirichlet που είδαμε στις προηγούμενες παραγράφους, είναι απαραίτητο να ισχύει ότι $q_{\min} > 0$, γιατί διαφορετικά δεν έχουμε μοναδική λύση του (3.39), όπως είδαμε στην Παράγραφο 1.2.1.

Θεωρούμε και πάλι έναν φυσικό αριθμό N και μια διαμέριση του διαστήματος $[a,b]$ από $N+2$ ισαπέχοντα σημεία $a = x_0 < x_1 < \cdots < x_N < x_{N+1} = b$,

όπου $h = x_{i+1} - x_i$, $i = 0, \ldots, N$. Στα σημεία x_i, $i = 1, \ldots, N$, θα ισχύει η (3.2) και σκοπός μας είναι και πάλι να κατασκευάσουμε προσεγγίσεις U_i των τιμών $u(x_i)$, χρησιμοποιώντας την (3.6). Όμως, σε αντίθεση, με τη μέθοδο πεπερασμένων διαφορών που είδαμε στην Παραγράφο 3.1, για το πρόβλημα (3.1), δεν γνωρίζουμε τις τιμές $u(x_0)$ και $u(x_{N+1})$. Έτσι, τώρα θα χρειαστούμε δύο επιπλέον εξισώσεις εκτός από τις (3.6), για να υπολογίσουμε τα U_i, $i = 0, \ldots, N+1$.

Ένας τρόπος για να το κάνουμε αυτό είναι να θεωρήσουμε ότι η u επεκτείνεται άρτια αριστερά του a και δεξιά του b, δηλαδή $u(a+h) = u(a-h)$ και $u(b-h) = u(b+h)$, $h > 0$. Ο λόγος που θεωρούμε άρτια επέκταση είναι διότι αν π.χ. η u είναι άρτια γύρω από το a, τότε $u'(a) = \lim_{h\to 0}(u(a+h) - u(a-h))/(2h) = 0$. Επομένως, η προσέγγιση της $u''(a)$, $\delta^c_{h,2}u(a)$ γίνεται

$$\delta^c_{h,2}u(a) = \frac{u(a+h) - 2u(a) + u(a-h)}{h^2} = 2\frac{u(a+h) - u(a)}{h^2}.$$

Ανάλογα, παίρνουμε

$$\delta^c_{h,2}u(b) = 2\frac{u(b-h) - u(b)}{h^2}.$$

Άρα, χρησιμοποιώντας τις παραπάνω προσεγγίσεις $\delta^c_{h,2}u(a)$ και $\delta^c_{h,2}u(b)$ στην (3.39), έχουμε

$$\begin{aligned} -2\frac{u(x_1) - u(x_0)}{h^2} + q(x_0)u(x_0) &= f(x_0) + \eta_0, \\ -2\frac{u(x_N) - u(x_{N+1})}{h^2} + q(x_{N+1})u(x_{N+1}) &= f(x_{N+1}) + \eta_{N+1}, \end{aligned} \tag{3.40}$$

όπου

$$|\eta_i| \le \frac{h}{3}\max_{a\le x\le b}|u^{(3)}(x)|, \quad i = 0, N+1. \tag{3.41}$$

Συνεπώς, οι δύο επιπλέον σχέσεις που συμπληρώνουν τις (3.6) εδώ είναι

$$\begin{aligned} -2\frac{U_1 - U_0}{h^2} + q(x_0)U_0 &= f(x_0), \\ -2\frac{U_N - U_{N+1}}{h^2} + q(x_{N+1})U_{N+1} &= f(x_{N+1}). \end{aligned}$$

Για να κατασκευάσουμε λοιπόν προσεγγίσεις U_i της λύσης u του προβλήματος (3.39) στα σημεία x_i, $i = 0, \ldots, N+1$, θεωρούμε τις ακόλουθες εξισώσεις

$$-\frac{U_{i+1} - 2U_i + U_{i-1}}{h^2} + q(x_i)U_i = f(x_i), \quad i = 1, \ldots, N, \tag{3.42}$$

$$-2\frac{U_1 - U_0}{h^2} + q(x_0)U_0 = f(x_0), \tag{3.43}$$

$$-2\frac{U_N - U_{N+1}}{h^2} + q(x_{N+1})U_{N+1} = f(x_{N+1}). \tag{3.44}$$

Επομένως, αν συμβολίσουμε με $U \in \mathbb{R}^{N+2}$ το διάνυσμα με συνιστώσες $U_0, \ldots, U_{N+1}, U = (U_0, \ldots, U_{N+1})^T$, μπορούμε να γράψουμε το νέο σύστημα εξισώσεων ισοδύναμα ως

$$(A + h^2 Q)U = h^2 F, \tag{3.45}$$

όπου A είναι ο $(N+2) \times (N+2)$ πίνακας

$$A = \begin{pmatrix} 2 & -2 & 0 & & 0 \\ -1 & 2 & -1 & 0 & \\ 0 & \ddots & \ddots & \ddots & 0 \\ & & -1 & 2 & -1 \\ & & & -2 & 2 \end{pmatrix},$$

Q είναι ένας διαγώνιος $(N+2)\times(N+2)$ πίνακας με στοιχεία $q(x_i), i = 0, \ldots, N+1$, στη διαγώνιο και $F = (f(x_0), \ldots, f(x_{N+1}))^T$. Εύκολα βλέπουμε ότι ο πίνακας $A + h^2 Q$ είναι τριδιαγώνιος με αυστηρά κυριαρχική διαγώνιο, διότι $q_{\min} > 0$, και άρα αντιστρέφεται. Συνεπώς, το γραμμικό σύστημα (3.45) έχει μοναδική λύση.

Λόγω των (3.3) και (3.41) για το τοπικό σφάλμα διακριτοποίησης της μεθόδου (3.42)–(3.44) έχουμε ότι ισχύει το ακόλουθο λήμμα,

Λήμμα 3.5. *Έστω u η λύση του* (3.1) *με $u \in C^4[a,b]$. Τότε για την η_i που δίνεται στην* (3.3) *και* (3.41) *έχουμε ότι υπάρχει σταθερά ανεξάρτητη του h, τέτοια ώστε*

$$\max_{0\le i\le N+1} |\eta_i| \le Ch. \tag{3.46}$$

Επομένως λόγω του Λήμματος 3.5, αν η ακριβή λύση u είναι αρκετά ομαλή, τότε το σφάλμα διακριτοποίησης η_i τείνει στο μηδέν καθώς το h τείνει στο μηδέν. Επομένως ισχύει ότι η μέθοδος πεπερασμένων διαφορών (3.42)–(3.44) είναι *συνεπής*.

Με όμοια επιχειρήματα όπως και στην περίπτωση του Θεωρήματος 3.1, προκύπτει και η ευστάθεια της μεθόδου για το πρόβλημα (3.39).

Θεώρημα 3.4. *Έστω U_i, $i = 0, \ldots, N+1$, η λύση του προβλήματος* (3.42)–(3.44). *Τότε υπάρχει μια σταθερά C, ανεξάρτητη του h, τέτοια ώστε*

$$\max_{0\le i\le N+1} |U_i| \le C \max_{x\in[a,b]} |f(x)|. \tag{3.47}$$

Απόδειξη. Η απόδειξη γίνεται με ανάλογο τρόπο όπως και αυτή του Θεωρήματος 3.1. □

Στη συνέχεια, δείχνουμε τη σύγκλιση της μεθόδου (3.42)–(3.44) στην ακριβή λύση του (3.39).

Θεώρημα 3.5. *Έστω U_i, $i = 0, \ldots, N+1$, η λύση του προβλήματος* (3.42)–(3.44) *και u η λύση του προβλήματος* (3.39)*, με $u \in C^4[a, b]$. Τότε υπάρχει μια σταθερά C, ανεξάρτητη του h, τέτοια ώστε*

$$\max_{0 \le i \le N+1} |U_i - u(x_i)| \le Ch. \tag{3.48}$$

Απόδειξη. Η απόδειξη γίνεται με ανάλογο τρόπο όπως και αυτή του Θεωρήματος 3.2. Έτσι μπορούμε να δείξουμε ότι

$$\max_{0 \le i \le N+1} |U_i - u(x_i)| \le C \max_{0 \le i \le N+1} |\eta_i|,$$

για μια σταθερά C, ανεξάρτητη του h. Λόγω τώρα του Λήμματος 3.5, έχουμε τη ζητούμενη ανισότητα. □

Παρατήρηση: Η ανισότητα (3.48) που προκύπτει στο Θεώρημα 3.5 είναι διαφορετική από την αντίστοιχη του Θεωρήματος 3.2, ως προς την τάξη του h. Αυτό οφείλεται στον τρόπο απόδειξης του θεωρήματος. Στην πραγματικότητα μπορούμε να αποδείξουμε ότι και για το πρόβλημα (3.39) ισχύει, βλ. π.χ. (Ακρίβης & Δουγαλής, 2005· Ακρίβης & Δουγαλής, 2013),

$$\max_{0 \le i \le N+1} |U_i - u(x_i)| \le Ch^2. \tag{3.49}$$

Η απόδειξη αυτής της νέας ανισότητας είναι πιο πολύπλοκη, γίνεται χρησιμοποιώτας ανάλογα επιχειρήματα όπως στη Παράγραφο 3.4 και δεν είναι στους σκοπούς αυτών των σημειώσεων να δείξουμε την (3.49).

Παράδειγμα 3.3. Θεωρούμε το ακόλουθο πρόβλημα συνοριακών τιμών

$$-u''(x) + u(x) = \cos(2\pi x), \quad 0 < x < 1, \quad \text{με } u'(0) = u'(1) = 0. \tag{3.50}$$

Η ακριβής λύση u αυτού του προβλήματος είναι

$$u(x) = \frac{\cos(2\pi x)}{1 + 4\pi^2}. \tag{3.51}$$

Η εξίσωση πεπερασμένων διαφορών (3.42)–(3.44) γίνεται τώρα

$$\begin{aligned} -\frac{U_{i+1} - 2U_i + U_{i-1}}{h^2} + U_i &= \cos(2\pi x_i), \quad i = 1, \ldots, N, \\ -2\frac{U_1 - U_0}{h^2} + U_0 &= 0, \\ -2\frac{U_N - U_{N+1}}{h^2} + U_{N+1} &= 0. \end{aligned} \tag{3.52}$$

h	$\bar{E}$	p
0.1000	0.00081	
0.0500	0.00020	2.020
0.0250	0.00005	2.005
0.0125	0.00001	2.001

Πίνακας 3.3: Το σφάλμα $\bar{E} = \max_{0\leq i\leq N+1} |U_i - u(x_i)|$ της λύσεως του (3.52) στο Παράδειγμα 3.3, όπου $h = 1/(N+1)$, και η κατά προσέγγιση τάξη ακρίβειας p.

Αν θεωρήσουμε τώρα μια διαμέριση του $[0, 1]$ με βήμα $h = 0.1, 0.05, 0.025, 0.0125$, στον Πίνακα 3.3 βλέπουμε το σφάλμα $\bar{E} = \max_{0\leq i\leq N+1} |U_i - u(x_i)|$ της λύσεως του (3.52) όπου $h = 1/(N+1)$, καθώς και την αντίστοιχη προσεγγιστική τάξη ακρίβειας p, η οποία τείνει στο δύο.

3.6 Ένα γενικότερο πρόβλημα

Θεωρούμε τώρα μια πιο γενική διαφορική εξίσωση, από αυτές που συναντήσαμε στις προηγούμενες παραγράφους αυτού του κεφαλαίου, και το παρόμοιο, με το (3.1), πρόβλημα δύο σημείων με ομογενείς συνοριακές συνθήκες Dirichlet

$$\begin{aligned} -u''(x) + p(x)u'(x) + q(x)u(x) = f(x), \quad x \in [a, b], \\ u(a) = u(b) = 0, \end{aligned} \tag{3.53}$$

όπου $a, b \in \mathbb{R}$, $a < b$, $p, q, f \in C[a, b]$ και $q(x) > 0$, για κάθε $x \in [a, b]$.

Όπως και στα προηγούμενα κεφάλαια θεωρούμε έναν ομοιόμορφο διαμερισμό $N + 2$ σημείων του $[a, b]$, x_i, $i = 0, \ldots, N + 1$, και θέλουμε και εδώ να κατασκευάσουμε προσεγγίσεις U_i των τιμών $u(x_i)$ της ακριβούς λύσης του (3.53). Σε κάθε σημείο του διαμερισμού x_i, $i = 1, \ldots, N$, θα ισχύει:

$$-u''(x_i) + p(x_i)u'(x_i) + q(x_i)u(x_i) = f(x_i), \quad i = 1, \ldots, N. \tag{3.54}$$

Λόγω των συνοριακών συνθηκών $u(x_0) = u(x_{N+1}) = 0$, θέτουμε λοιπόν $U_0 = U_{N+1} = 0$. Στη συνέχεια, όπως και πριν στις Παραγράφους 3.1 και 3.5, οι τιμές U_i προκύπτουν προσεγγίζοντας κατάλληλα τις παραγώγους στην (3.54). Αντικαθιστούμε, λοιπόν, την $u''(x_i)$ στην (3.54) και πάλι με την $\delta^c_{h,2}u(x_i)$, οπότε αν $u \in C^4[a, b]$, λόγω της (2.9), η (3.54) γίνεται

$$\begin{aligned} -\frac{u(x_{i+1}) - 2u(x_i) + u(x_{i-1})}{h^2} + p(x_i)u'(x_i) + q(x_i)u(x_i) \\ = f(x_i) + \eta^1_i, \quad i = 1, \ldots, N, \end{aligned} \tag{3.55}$$

όπου

$$|\eta_i^1| \leq \frac{h^2}{12} \max_{x\in[a,b]} |u^{(4)}(x)|, \quad i = 1, \ldots, N. \tag{3.56}$$

Στη συνέχεια, για την προσέγγιση της $u'(x_i)$, χρησιμοποιούμε την $\delta_h^c u(x_i)$ που θεωρήσαμε στη (2.2). Συνεπώς, αν $u \in C^3[a,b]$, λόγω τώρα της (2.4), η (3.54) γίνεται

$$\begin{aligned} -\frac{u(x_{i+1}) - 2u(x_i) + u(x_{i-1})}{h^2} + p(x_i)\frac{u(x_{i+1}) - u(x_{i-1})}{2h} \\ + q(x_i)u(x_i) = f(x_i) + \eta_i^1 + \eta_i^2, \quad i = 1, \ldots, N, \end{aligned} \tag{3.57}$$

όπου

$$|\eta_i^2| \leq \frac{h^2}{6} \max_{x\in[a,b]} |u^{(3)}(x)|, \quad i = 1, \ldots, N, \tag{3.58}$$

Επομένως, λόγω των (3.56) και (3.58) έχουμε ότι ισχύει το ακόλουθο λήμμα.

Λήμμα 3.6. *Έστω u η λύση του* (3.1) *με $u \in C^4[a,b]$. Αν θέσουμε $\eta_i = \eta_i^1 + \eta_i^2$, με η_i^1, η_i^2 που δίνονται στις* (3.55) *και* (3.57)*, τότε έχουμε ότι υπάρχει σταθερά ανεξάρτητη του h, τέτοια ώστε*

$$\max_{1\leq i\leq N} |\eta_i| \leq Ch^2. \tag{3.59}$$

Παρατήρηση 3.6. Αντί της $\delta_h^c u(x_i)$ για την προσέγγιση της $u'(x_i)$ θα μπορούσαμε να χρησιμοποιήσουμε την $\delta_h^+ u(x_i)$ είτε την $\delta_h^- u(x_i)$. Σε αυτή την περίπτωση αντί της (3.58) θα ισχύει μια διαφορετική εκτίμηση για το η_i, όπου η τάξη του h θα είναι ένα και κατά συνέπεια το συμπέρασμα του Λήμματος 3.6 θα είναι διαφορετικό, βλ. Άσκηση 3.8.

Επομένως, οι προσεγγίσεις U_i, $i = 0, \ldots, N+1$, των $u(x_i)$ προκύπτουν σύμφωνα με τις ακόλουθες εξισώσεις

$$\begin{aligned} -\frac{U_{i+1} - 2U_i + U_{i-1}}{h^2} + p(x_i)\frac{U_{i+1} - U_{i-1}}{2h} + q(x_i)U_i \\ = f(x_i), \quad i = 1, \ldots, N, \end{aligned} \tag{3.60}$$

$$U_0 = U_{N+1} = 0. \tag{3.61}$$

Αν συμβολίσουμε με $U \in \mathbb{R}^N$, το διάνυσμα με συνιστώσες $U_1, \ldots, U_N$, $U = (U_1, \ldots, U_N)^T$, μπορούμε να γράψουμε το σύστημα εξισώσεων (3.60)–(3.61) ισοδύναμα

$$(A + h^2 Q)U = h^2 F, \tag{3.62}$$

όπου $A \in \mathbb{R}^{N\times N}$ είναι ο πίνακας

$$A = \begin{pmatrix} 2 & -1+p(x_1)\frac{h}{2} & 0 & & \\ -1-p(x_2)\frac{h}{2} & 2 & -1+p(x_2)\frac{h}{2} & 0 & \\ & \ddots & \ddots & & \\ & -1-p(x_{N-1})\frac{h}{2} & 2 & -1+p(x_{N-1})\frac{h}{2} & \\ & 0 & -1-p(x_N)\frac{h}{2} & 2 & \end{pmatrix},$$

και όπως στην (3.8) ο Q είναι ένας διαγώνιος $N \times N$ πίνακας και $F \in \mathbb{R}^N$. Προκύπτει λοιπόν ότι ο $A + h^2Q$ είναι ένας τριδιαγώνιος πίνακας, ο οποίος για να έχει αυστηρά κυριαρχική διαγώνιο πρέπει

$$\begin{aligned} &|2+h^2q(x_i)| > |1+p(x_i)\frac{h}{2}| + |1-p(x_i)\frac{h}{2}|, \quad i = 2,\dots,N-1, \\ &|2+h^2q(x_1)| > |1-p(x_1)\frac{h}{2}|, \quad \text{και } |2+h^2q(x_N)| > |1+p(x_N)\frac{h}{2}|. \end{aligned} \tag{3.63}$$

Αν $\max_{x\in[a,b]} |p(x)|\frac{h}{2} < 1$, τότε έχουμε ότι $1 + p(x_i)\frac{h}{2} > 0$ και $1 - p(x_i)\frac{h}{2} > 0$, $i = 1,\dots,N$. Οπότε σε αυτήν την περίπτωση ισχύουν οι συνθήκες (3.63).

Λόγω του Λήμματος 3.6, αν η ακριβής λύση u είναι αρκετά ομαλή, τότε το σφάλμα διακριτοποίησης η_i τείνει στο μηδέν, καθώς το h τείνει στο μηδέν. Επομένως, ισχύει ότι η μέθοδος πεπερασμένων διαφορών (3.60)–(3.61) είναι *συνεπής*.

Με όμοια επιχειρήματα όπως και στην περίπτωση του Θεωρήματος 3.1, προκύπτει το ακόλουθο θεώρημα για την ευστάθεια της μεθόδου.

Θεώρημα 3.6. *Έστω U_i, $i = 0,\dots,N+1$, η λύση του προβλήματος* (3.60)–(3.61)*, $q_{\min} > 0$ και $\max_{x\in[a,b]} |p(x)|\frac{h}{2} < 1$. Τότε υπάρχει σταθερά C ανεξάρτητη του h, τέτοια ώστε*

$$\max_{0\le i\le N+1} |U_i| \le C \max_{x\in[a,b]} |f(x)|. \tag{3.64}$$

Απόδειξη. Η απόδειξη γίνεται με ανάλογο τρόπο όπως και αυτή του Θεωρήματος 3.1. Λόγω της (3.60), έχουμε

$$(2+h^2q(x_i))U_i = (1-p(x_i)\frac{h}{2})U_{i+1} + (1+p(x_i)\frac{h}{2})U_{i-1} + h^2f(x_i), \quad 1 \le i \le N.$$

Στη συνέχεια, επειδή $q_{\min} > 0$ και $\max_{x\in[a,b]} |p(x)|\frac{h}{2} < 1$, έχουμε

$$\begin{aligned} (2+h^2q_{\min})|U_i| &\le |1-p(x_i)\frac{h}{2}|\,|U_{i+1}| + |1+p(x_i)\frac{h}{2}|\,|U_{i-1}| + h^2|f(x_i)| \\ &\le 2\max_{0\le i\le N+1} |U_i| + h^2 \max_{x\in[a,b]} |f(x)|, \end{aligned}$$

από όπου εύκολα παίρνουμε τη ζητούμενη ανισότητα. □

Θεώρημα 3.7. *Έστω U_i, $i = 0, \ldots, N+1$, η λύση του προβλήματος* (3.60)–(3.61)*, και u η λύση του προβλήματος* (3.53)*, με $u \in C^4[a,b]$. Τότε, αν $q_{\min} > 0$, και $\max_{x\in[a,b]} |p(x_i)|\frac{h}{2} < 1$ υπάρχει μια σταθερά C, ανεξάρτητη του h, τέτοια ώστε*

$$\max_{0\le i\le N+1} |U_i - u(x_i)| \le Ch^2. \tag{3.65}$$

Απόδειξη. Η απόδειξη γίνεται με ανάλογο τρόπο όπως και αυτή του Θεωρήματος 3.6. Έτσι μπορούμε να δείξουμε ότι

$$\max_{1\le i\le N} |U_i - u(x_i)| \le C \max_{1\le i\le N} |\eta_i|,$$

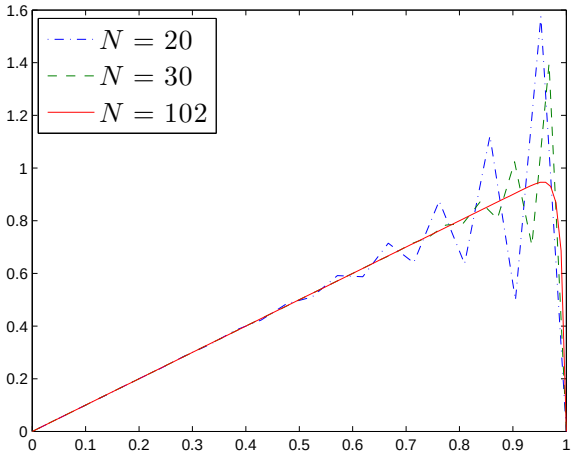
για μια σταθερά C, ανεξάρτητη του h. Λόγω τώρα της (3.62) έχουμε τη ζητούμενη ανισότητα. □

Παρατήρηση 3.7. Η υπόθεση $q_{\min} > 0$ δεν είναι απαραίτητη για να δείξουμε την ευστάθεια ή τη σύγκλιση της μεθόδου, αν ακολουθήσουμε παρόμοια ανάλυση όπως στην Παράγραφο 3.4. Όμως, η υπόθεση $\max_{x\in[a,b]} |p(x)|\frac{h}{2} < 1$ είναι αναγκαία, όπως φαίνεται και στο επόμενο παράδειγμα. Λόγω αυτής της ιδιότητας, της ύπαρξης δηλαδή συνθηκών που πρέπει να ικανοποιεί η διαμέριση, η μέθοδος (3.60)–(3.61) καλείται *ευσταθής υπό συνθήκες*.

Παράδειγμα 3.4. Για να δούμε ότι είναι αναγκαία η υπόθεση $\max_{x\in[a,b]} |p(x)|\frac{h}{2} < 1$ για την προσέγγιση της ακριβούς λύσης θα θεωρήσουμε το ακόλουθο πρόβλημα συνοριακών τιμών:

$$-u''(x) + pu' + u(x) = f(x), \quad 0 < x < 1, \quad \text{με } u(0) = u(1) = 0, \tag{3.66}$$

με $p = 200$ και $f(x) = 9900x^{98} + p(1 - 100x^{99}) + x - x^{100}$. Η ακριβής λύση u αυτού του προβλήματος είναι $u(x) = x - x^{100}$. Αν θεωρήσουμε έναν διαμερισμό του $[0, 1]$ με $N = 20, 30$, τότε, προφανώς, δεν ικανοποιείται η συνθήκη $p\frac{h}{2} < 1$, ενώ ισχύει αν $N = 102$. Στο Σχήμα 3.2 βλέπουμε ότι η προσέγγιση αποτυγχάνει κοντά στο σημείο $x = 1$ αν $N = 20, 30$.



Σχήμα 3.2: Οι προσεγγιστικές λύσεις του Παράδειγματος 3.4.

3.7 Μη ομογενείς συνοριακές συνθήκες

Στη συνέχεια θα τροποποιήσουμε τις ομογενείς συνοριακές συνθήκες Dirichlet του προβλήματος (3.1), ζητώντας τώρα μια συνάρτηση $u \in C^2[a,b]$, τέτοια ώστε

$$-u''(x) + q(x)u(x) = f(x), \quad x \in [a,b], \quad \text{με } u(a) = a_0,\ u(b) = b_0, \tag{3.67}$$

όπου $a, b \in \mathbb{R}$, $a < b$, $q, f \in C[a,b]$ και $q(x) \geq 0$, για κάθε $x \in [a,b]$ και $a_0, b_0 \in \mathbb{R}$.

Όπως και στις προηγούμενες παραγράφους θεωρούμε έναν φυσικό αριθμό N και μια διαμέριση του διαστήματος $[a,b]$ από $N+2$ ισαπέχοντα σημεία $a = x_0 < x_1 < \cdots < x_N < x_{N+1} = b$, όπου $h = x_{i+1} - x_i$, $i = 0, \ldots, N$. Επειδή το πρόβλημα (3.67) διαφέρει από το (3.1) μόνο ως προς τις συνοριακές συνθήκες, θα κατασκευάσουμε προσεγγίσεις U_i των τιμών $u(x_i)$, $i = 0, \ldots, N+1$, όπου θα ικανοποιούν την (3.6) και λόγω των συνοριακών συνθηκών $u(x_0) = a_0$, $u(x_{N+1}) = b_0$, θα θέσουμε $U_0 = a_0$, $U_{N+1} = b_0$.

Επομένως, κατασκευάζουμε προσεγγίσεις U_i των $u(x_i)$, $i = 0, \ldots, N+1$, σύμφωνα με τις ακόλουθες εξισώσεις

$$-\frac{U_{i+1} - 2U_i + U_{i-1}}{h^2} + q(x_i)U_i = f(x_i), \quad i = 1, \ldots, N, \tag{3.68}$$

$$U_0 = a_0, \quad U_{N+1} = b_0. \tag{3.69}$$

Συνεπώς, το τοπικό σφάλμα διακριτοποίησης θα ικανοποιεί την (3.3) και θα ισχύει για αυτό το ανάλογο του Λήμματος 3.1.

Έτσι, αν συμβολίσουμε με $U \in \mathbb{R}^N$ το διάνυσμα με συνιστώσες $U_1, \ldots, U_N$, $U = (U_1, \ldots, U_N)^T$, μπορούμε να γράψουμε το σύστημα των εξισώσεων (3.68)-(3.69) ισοδύναμα ως γραμμικό σύστημα $(A+h^2Q)U = h^2F$, όπου ο πίνακας A δίνεται όπως στην (3.9), Q είναι ένας διαγώνιος $N \times N$ πίνακας με στοιχεία $q(x_i), i = 1, \ldots, N$, στη διαγώνιο και $F = (f(x_1) + U_0/h^2, f(x_2), \ldots, f(x_{N-1}), f(x_N) + U_{N+1}/h^2)^T$.

Στη συνέχεια, χρησιμοποιώντας παρόμοια επιχειρήματα όπως στο Θεώρημα 3.2, δείχνουμε τη σύγκλιση της μεθόδου (3.68)–(3.69).

Θεώρημα 3.8. *Έστω U_i, $i = 0, \ldots, N+1$, η λύση του προβλήματος* (3.68)–(3.69)*, και u η λύση του προβλήματος* (3.67)*, με $u \in C^4[a, b]$. Τότε, αν $q_{\min} > 0$, υπάρχει μια σταθερά C, ανεξάρτητη του h, τέτοια ώστε*

$$\max_{0 \le i \le N+1} |U_i - u(x_i)| \le Ch^2. \tag{3.70}$$

Απόδειξη. Θέτουμε $E_i = U_i - u(x_i)$, $i = 0, \ldots, N+1$, όπου λόγω των σχέσεων $U_0 = u(a)$ και $U_{N+1} = u(b)$, έχουμε $E_0 = E_{N+1} = 0$. Αφαιρούμε τώρα κατά μέλη τις (3.68) και (3.3), οπότε παίρνουμε

$$E_{i+1} - (2 + q(x_i)h^2)E_i + E_{i-1} = h^2\eta_i, \quad i = 1, \ldots, N. \tag{3.71}$$

Συνεχίζουμε όπως στην απόδειξη του Θεωρήματος (3.2) και καταλήγουμε στη ζητούμενη ανισότητα. □

3.8 Άλλες συνοριακές συνθήκες

Ακολουθώντας την ανάλυση προβλημάτων με συνοριακές συνθήκες τύπου Dirichlet ή Neumann που είδαμε σε αυτό το κεφάλαιο, μπορούμε να τροποποιήσουμε ανάλογα τις μεθόδους που παρουσιάσαμε, ώστε να ισχύουν αντίστοιχα αποτελέσματα και για άλλου τύπου συνοριακές συνθήκες, όπως π.χ. Robin. Επίσης, μπορούμε να θεωρήσουμε και άλλες μεθόδους, οι οποίες βασίζονται στις πεπερασμένες διαφορές, για την προσέγγιση της λύσης του προβλήματος (3.1), όπως είναι η μέθοδος των πεπερασμένων όγκων (finite volume method). Παραπέμπουμε τον αναγνώστη στα (Ακρίβης & Δουγαλής, 2013· Holmes, 2007· Morton & Mayers, 2005· Roos, Stynes, & Tobiska, 2008· Thomas, 1995) για πιο λεπτομερή παρουσίαση.

3.9 Ταινίες γραφικών παραστάσεων

Σε αυτή την παράγραφο εμφανίζονται δυο ταινίες με τις γραφικές παραστάσεις που παρουσιάστηκαν στα Παραδείγματα 3.1 και 3.4. Η προβολή των ταινιών στην οθόνη του Η/Υ μπορεί να γίνει αν μετακινήσουμε τον κέρσορα (δείκτη) της οθόνης, και επιλέξουμε, π.χ. με τη χρήση του "ποντίκιου", το αντίστοιχο παράδειγμα και, στη συνέχεια, την επιλογή "Play/Pause". Στην περίπτωση που η προβολή αυτού του βιβλίου στην οθόνη γίνεται μέσω του αντίστοιχου αρχείου μορφής pdf, συνίσταται η χρήση του πρόγραμματος Adobe Reader.

Play/Pause | Παράδειγμα 3.1 | Παράδειγμα 3.4

Σχήμα 3.3: Ταινίες των γραφικών παραστάσεων των Παραδειγμάτων 3.1 και 3.4.

3.10 Ασκήσεις

3.1. Δείξτε ότι η $|\cdot|_{1,h}$ που ορίζεται στην (3.25) ορίζει μια νόρμα στον $\mathbb{R}_0^{N+2}$.

3.2. Έστω u η λύση του προβλήματος συνοριακών τιμών

$$\begin{aligned} -x^2u''(x) - xu'(x) + 4u(x) = 20x^3, \quad \text{με } x \in [1,2], \\ u(1) = 0,\ u(2) = 0. \end{aligned}$$

Γράψτε το αριθμητικό σχήμα πεπερασμένων διαφορών χρησιμοποιώντας κεντρικές διαφορές δ_h^c και $\delta_{h,2}^c$. Ποιος είναι ο περιορισμός για το βήμα h, ώστε ο αντίστοιχος πίνακας που χρησιμοποιούμε για την προσέγγιση της λύσης να είναι αντιστρέψιμος;

3.3. Έστω u η λύση του προβλήματος συνοριακών τιμών

$$\begin{aligned} -u''(x) + u'(x) = 1, \quad \text{με } x \in [a,b], \\ u(a) = 0,\ u(b) = 0. \end{aligned}$$

(α΄) Έστω ότι προσεγγίζουμε τη δεύτερη παράγωγο, $u''(x_i)$, με την κεντρική διαφορά $(u(x_{i+1}) - 2u(x_i) + u(x_{i-1}))/h^2$ και την πρώτη παράγωγο, $u'(x_i)$, με τη διαφορά $(u(x_i) - u(x_{i-1}))/h$. Ποιο θα είναι το διακριτό σχήμα και ποιο το τοπικό σφάλμα διακριτοποίησης;

(β΄) Γράψτε τη μέθοδο σε μορφή γραμμικού συστήματος, όπως η (3.8). Για να είναι αντιστρέψιμος ο πίνακας, χρειάζεται περιορισμός στο βήμα h ;

3.4. Έστω u η λύση του προβλήματος συνοριακών τιμών, με $a, b > 0$

$$\begin{aligned} -u''(x) + u(x) = f(x), \quad \text{με } x \in [0,1], \\ au(0) + bu'(0) = c,\ u(1) = 0. \end{aligned}$$

(α΄) Διατυπώστε ένα διακριτό σχήμα με τοπικό σφάλμα διακριτοποίησης τάξης ακρίβειας δύο.

(β΄) Γράψτε τη μέθοδο σε μορφή γραμμικού συστήματος.

3.5. Έστω u η λύση του προβλήματος συνοριακών τιμών

$$\begin{aligned} -u''(x) + u(x) = f(x), \quad \text{με } x \in [0,1], \\ u(0) = u(1),\ u'(0) = u'(1). \end{aligned}$$

(α΄) Διατυπώστε ένα διακριτό σχήμα με τοπικό σφάλμα διακριτοποίησης τάξης ακρίβειας δύο.

(β΄) Γράψτε τη μέθοδο σε μορφή γραμμικού συστήματος.

3.6. (α΄) Χρησιμοποιώντας το θεώρημα του Taylor, αποδείξτε ότι

$$u(x_{i+1}) - 2u(x_i) + u(x_{i-1}) = h^2 u''(x_i) + \frac{1}{12}h^4 u''''(x_i) + \eta_i$$

και ότι

$$\begin{aligned} &u(x_{i+1}) - 2u(x_i) + u(x_{i-1}) \\ &\quad = \frac{1}{12}h^2(u''(x_{i+1}) + 10u''(x_i) + u''(x_{i-1})) + \tilde{\eta}_i, \end{aligned}$$

όπου $\max\{|\tilde{\eta}_i|, |\eta_i|\} \leq Ch^6$, με C ανεξάρτητη του h.

(β΄) Αν υποθέσουμε ότι η u ικανοποιεί τη Δ.Ε. $u''(x) = F(x,u)$, χρησιμοποιήστε το παραπάνω αποτέλεσμα για να καταλήξετε στη μέθοδο πεπερασμένων διαφορών

$$-(U_{i+1} - 2U_i + U_{i+1}) = \frac{h^2}{12}(F_{i+1} + 10F_i + F_{i+1}),$$

με $F_i = F(x_i, u(x_i))$.

(γ΄) Διατυπώστε τη μέθοδο, όταν $F(x,u) = f(x) - q(x)u$. Γράψτε τη μέθοδο σε μορφή γραμμικού συστήματος.

3.7. Θεωρούμε το πρόβλημα

$$\begin{aligned} -\frac{d}{dx}(D(x)\frac{d}{dx}u(x)) + u(x) = f(x), \quad &\text{με } x \in [0,1], \\ u(0) = u(1) = 0. & \end{aligned}$$

όπου D είναι ομαλή θετική συνάρτηση.

(α΄) Γράψτε ένα αριθμητικό σχήμα με τοπικό σφάλμα διακριτοποίησης τάξης ακρίβειας δύο. Εκφράστε τη μέθοδο και σε μορφή γραμμικού συστήματος.

(β΄) Μπορείτε να δείξετε ένα ανάλογο αποτέλεσμα με το Θεώρημα 3.1;

3.8. Θεωρούμε το πρόβλημα

$$\begin{aligned} -u''(x) + p(x)u' + q(x)u(x) = f(x), \quad &\text{με } x \in [0,1], \\ u(0) = u(1) = 0. & \end{aligned}$$

Θεωρούμε έναν ομοιόμορφο διαμερισμό του διαστήματος $[0,1]$, $a = x_0 < x_1 < \cdots < x_N < x_{N+1} = b$ και συμβολίζουμε με $h = x_i - x_{i-1}$, $i = 1, \ldots, N+1$. Αν προσεγγίσουμε τη $u'(x)$ με τη $\delta_h^- u(x)$ είτε με την $\delta_h^+ u(x)$ γράψτε τα αντίστοιχα αριθμητικά σχήματα και βρείτε το τοπικό σφάλμα διακριτοποίησης, καθώς και την τάξη του. Είναι οι μέθοδοι ευσταθείς ή υπό συνθήκη ευσταθείς;

3.9. Θεωρούμε το πρόβλημα

$$-u''(x) + p(x)u' + q(x)u(x) = f(x), \quad \text{με } x \in [0,1],$$
$$u(0) = u(1) = 0.$$

Θεωρούμε έναν μη ομοιόμορφο διαμερισμό του διαστήματος $[0,1]$, $a = x_0 < x_1 < \cdots < x_N < x_{N+1} = b$ και συμβολίζουμε με $h_i = x_i - x_{i-1}$, $i = 1, \ldots, N+1$.

(α΄) Εκφράστε με πεπερασμένες διαφορές την προσέγγιση της πρώτης και της δεύτερης παραγώγου στο x_i και βρείτε το τοπικό σφάλμα διακριτοποίησης. Οι προσεγγίσεις πρέπει να είναι συνεπείς, δηλαδή αν h_i και h_{i+1} πάει στο μηδέν, τότε το σφάλμα τείνει και αυτό στο μηδέν.

(β΄) Χρησιμοποιήστε τα αποτελέσματα του προηγούμενου ερωτήματος για να διατυπώστε ένα σχήμα πεπερασμένων διαφορών για την παραπάνω διαφορική εξίσωση.

Βιβλιογραφία

Ακρίβης, Γ., & Δουγαλής, Β. (2005). *Αριθμητικές Μέθοδοι για Μερικές Διαφορικές Εξισώσεις*. Ιωάννινα. (Πανεπιστημιακές Σημειώσεις).

Ακρίβης, Γ., & Δουγαλής, Β. (2013). *Αριθμητικές Μέθοδοι για Συνήθεις Διαφορικές Εξισώσεις*. Πανεπιστημιακές Εκδόσεις Κρήτης, Ηράκλειο.

Ακρίβης, Γ., & Δουγαλής, Β. (2015). *Εισαγωγή στην Αριθμητική Ανάλυση*. Πανεπιστημιακές Εκδόσεις Κρήτης, Ηράκλειο.

Holmes, M. H. (2007). *Introduction to numerical methods in differential equations* (Vol. 52). Springer, New York.

Morton, K. W., & Mayers, D. F. (2005). *Numerical solution of partial differential equations* (Second ed.). Cambridge University Press, Cambridge.

Roos, H.-G., Stynes, M., & Tobiska, L. (2008). *Robust numerical methods for singularly perturbed differential equations* (Second ed., Vol. 24). Springer-Verlag, Berlin.

Thomas, J. W. (1995). *Numerical partial differential equations: finite difference methods* (Vol. 22). Springer-Verlag, New York.

Κεφάλαιο 4

Η μέθοδος των πεπερασμένων στοιχείων για το πρόβλημα δύο σημείων

Η μέθοδος των πεπερασμένων στοιχείων είναι μια τεχνική για την κατασκευή προσεγγιστικών λύσεων μερικών και ολοκληρωτικών διαφορικών εξισώσεων. Σε αυτό το κεφάλαιο θα μελετήσουμε αυτή τη μέθοδο για το πρόβλημα δύο σημείων που θεωρήσαμε στο Κεφάλαιο 3.

4.1 Μεταβολικό πρόβλημα

Θεωρούμε το πρόβλημα δύο σημείων με συνοριακές συνθήκες Dirichlet (3.1), δηλαδή ζητούμε μια συνάρτηση $u \in C^2[a,b]$ η οποία να ικανοποιεί

$$-u''(x) + q(x)u(x) = f(x), \text{ για } x \in [a,b], \text{ με } u(a) = u(b) = 0, \tag{4.1}$$

όπου a, b πραγματικοί αριθμοί με $a < b$ και $q, f \in C[a,b]$ με την q να λαμβάνει μη αρνητικές τιμές για κάθε $x \in [a,b]$. Όπως είδαμε στο Θεώρημα 1.2 το πρόβλημα (4.1) έχει μοναδική λύση.

Για την ανάλυση της μεθόδου των πεπερασμένων στοιχείων για το πρόβλημα δύο σημείων θα θεωρήσουμε το ακόλουθο εσωτερικό γινόμενο

$$(v,w) = \int_a^b v(x)w(x)\,dx, \quad v, w \in C[a,b],$$

και την αντίστοιχη νόρμα $\|\cdot\|$ που παράγεται από αυτό, δηλαδή

$$\|v\| = (v,v)^{1/2} = \left(\int_a^b v^2(x)\,dx\right)^{1/2}.$$

Μπορούμε να δούμε, βλέπε Άσκηση 4.1, ότι ισχύει η *ανισότητα Cauchy–Schwarz*

$$|(v,w)| \le \|v\|\,\|w\| \quad v,w \in C[a,b].$$

Στη συνέχεια, θα συμβολίζουμε με $C_0^\kappa[a,b]$ τις κ φορές συνεχώς παραγωγίσιμες συναρτήσεις, $\kappa \ge 0$, οι οποίες μηδενίζονται στα άκρα του $[a,b]$, $C_0^\kappa[a,b] = \{v \in C^\kappa[a,b] : v(a) = v(b) = 0\}$. Αν τώρα θεωρήσουμε το εσωτερικό γινόμενο και των δύο μελών της εξίσωσης (4.1) με μια συνάρτηση $v \in C_0^1[a,b]$, παίρνουμε

$$-(u'',v) + (qu,v) = (f,v) \quad \forall\, v \in C_0^1[a,b]. \tag{4.2}$$

Στη συνέχεια, ολοκληρώνοντας κατά μέρη τον πρώτο όρο στο αριστερό μέλος της (4.2) και χρησιμοποιώντας το γεγονός ότι $v(a) = v(b) = 0$, έχουμε

$$(u',v') + (qu,v) = (f,v) \quad \forall\, v \in C_0^1[a,b]. \tag{4.3}$$

Είναι χρήσιμο να γενικεύσουμε τον παραπάνω χαρακτηρισμό της λύσης του προβλήματος (4.1) ως προς την απαιτούμενη συνθήκη ομαλότητας της συνάρτησης v, δηλαδή για συναρτήσεις οι οποίες δεν είναι $C^2[a,b]$. Θα θεωρήσουμε, στη συνέχεια, τον ακόλουθο υπόχωρο V του $C_0[a,b]$,

$$V = \{v \in C_0[a,b] : v \text{ είναι κατά τμήματα συνεχώς παραγωγίσιμη}\}. \tag{4.4}$$

Λήμμα 4.1 (Ανισότητα Poincaré–Friedrichs)**.** *Έστω $v \in C_0^1[a,b]$. Τότε ισχύει η ακόλουθη ανισότητα*

$$\|v\| \le (b-a)\|v'\|. \tag{4.5}$$

Απόδειξη. Επειδή $v(a) = 0$, έχουμε

$$v(x) = \int_a^x v'(s)\,ds, \quad x \in (a,b). \tag{4.6}$$

Χρησιμοποιώντας τώρα την ανισότητα Cauchy–Schwarz, παίρνουμε ότι

$$|v(x)|^2 = \left(\int_a^x v'(s)\,ds\right)^2 \le \int_a^x 1\,ds \int_a^x [v'(s)]^2\,ds. \tag{4.7}$$

Συνεπώς, λαμβάνουμε ως αποτέλεσμα την ακόλουθη σχέση

$$|v(x)|^2 \le (b-a)\,\|v'\|^2, \quad x \in [a,b]. \tag{4.8}$$

Τελικά, η ανισότητα Poincaré–Friedrichs (4.5) προκύπτει ολοκληρώνοντας τα δύο μέλη της (4.8) στο $[a,b]$. □

Παρατήρηση 4.1. Το Λήμμα 4.1 ισχύει και για συναρτήσεις $v \in V$. Πράγματι, έστω μια διαμέριση του $[a,b]$, $a = x_0 < x_1 < \cdots < x_{N+1} = b$, τέτοια ώστε $v \in C^1(x_i, x_{i+1}), i = 0, \ldots, N+1$. Επίσης, έστω $x \in (a,b)$ και j τέτοιο, ώστε $x \in (x_j, x_{j+1})$ για $0 \le j \le N$. Τότε, αν $j = 0$, προφανώς ισχύει η (4.6), διαφορετικά για $j \ge 1$, επειδή $v(a) = 0$, έχουμε

$$v(x) = v(x) - v(x_j) + \sum_{i=0}^{j-1} [v(x_i) - v(x_{i+1})] = \int_a^x v'(s)\, ds, \quad x \in (a,b).$$

Χρησιμοποιώντας τώρα την ανισότητα Cauchy–Schwarz παίρνουμε την (4.7). Στη συνέχεια, δείχνουμε την (4.5), ακολουθώντας τα ίδια βήματα όπως στο Λήμμα 4.1.

Είναι εύκολο να δούμε ότι έχουμε τώρα

$$(u', v') + (qu, v) = (f, v) \quad \forall\, v \in V, \tag{4.9}$$

όπου, φυσικά, v' είναι κατά τμήματα παράγωγος της v. Θα αναφερόμαστε στον παραπάνω *χαρακτηρισμό* της λύσης του προβλήματος δύο σημείων ως *ασθενής* ή *μεταβολική μορφή* του (4.1), μια και η συνάρτηση v μπορεί να επιλεγεί αυθαίρετα στον χώρο V. Οι συναρτήσεις v του (4.1) καλούνται *συναρτήσεις δοκιμής*. Η συγκεκριμένη σχέση μας επιτρέπει όχι μόνο να γενικεύσουμε την έννοια της λύσης του προβλήματος (4.1), αλλά και να παράγουμε μια σημαντική κατηγορία μεθόδων για τη λύση του, τις λεγόμενες μεθόδους πεπερασμένων στοιχείων ή μεθόδους Galerkin που θα θεωρήσουμε στη συνέχεια. Παρατηρήστε ότι η σχέση (4.9) ισχύει, ακόμα και αν $u \in V$, δηλαδή όταν η u έχει λιγότερη από την απαιτούμενη *κλασική* ομαλότητα $u \in C^2[a,b]$. Η λύση u του (4.9) καλείται *ασθενής λύση* του (4.1).

Σημειώνουμε ακόμα ότι αν υποθέσουμε ότι η λύση του προβλήματος (4.1) έχει την κλασική ομαλότητα $u \in C^2[a,b]$, τότε μπορούμε να δείξουμε το ακόλουθο θεώρημα.

Θεώρημα 4.1. *Αν $u \in C_0^2[a,b]$ η λύση του προβλήματος* (4.1)*, τότε υπάρχει μια σταθερά C η οποία εξαρτάται από τα δεδομένα a, b και q, τέτοια ώστε*

$$\|u\| + \|u'\| + \|u''\| \le C\, \|f\|. \tag{4.10}$$

Απόδειξη. Από το πρόβλημα (4.3) έχουμε

$$\|u'\|^2 \le \|u'\|^2 + (qu, u) = (f, u),$$

οπότε χρησιμοποιώντας την ανισότητα Cauchy–Schwarz,

$$\|u'\|^2 \le \|f\|\, \|u\|. \tag{4.11}$$

Στη συνέχεια, λόγω της ανισότητας Poincaré–Friedrichs (4.5), παίρνουμε

$$\|u'\| \leq (b-a)\|f\|. \tag{4.12}$$

Εφαρμόζοντας άλλη μια φορά την (4.5), λαμβάνουμε

$$\|u\| \leq (b-a)^2\|f\|. \tag{4.13}$$

Επίσης, από την (4.1) έχουμε

$$u'' = qu - f,$$

επομένως, συνδυάζοντας τις (4.12)–(4.13) παίρνουμε

$$\|u''\| \leq \max_{a\leq x\leq b} |q(x)|\,\|f\|. \tag{4.14}$$

Άρα, από τις (4.12)–(4.14) λαμβάνουμε τη ζητούμενη ανισότητα. □

Η σχέση (4.10) αναφέρεται ως ανισότητα της *ελλειπτικής ομαλότητας* και θα δούμε τη χρησιμότητά της στη μελέτη της μεθόδου Galerkin που θα συναντήσουμε στη συνέχεια.

4.2 Μέθοδος πεπερασμένων στοιχείων για ομογενείς συνοριακές συνθήκες Dirichlet

Η *μέθοδος πεπερασμένων στοιχείων* ή *μέθοδος Galerkin* για το συγκεκριμένο πρόβλημα δύο σημείων (4.1), που θα ορίσουμε στη συνέχεια, μπορεί να τεθεί περιορίζοντας τόσο την αναζήτηση της λύσης όσο και τις *συναρτήσεις δοκιμής*, δηλαδή, τη συνάρτηση v στη σχέση (4.9), σε υπόχωρους του V πεπερασμένης διάστασης. Η απλούστερη μέθοδος Galerkin μπορεί να οριστεί ως εξής: θεωρούμε έναν διαμερισμό του $[a,b]$ σε $N+2$ σημεία, $N \geq 0$, $a = x_0 < x_1 < \cdots < x_{N+1} = b$ και ορίζουμε το ακόλουθο σύνολο συναρτήσεων

$$V_h = \{\chi \in C[a,b] : \chi(a) = \chi(b) = 0,\ \chi|_{[x_j,x_{j+1}]} \in \mathbb{P}_1\},$$

όπου $\mathbb{P}_1$ είναι ο χώρος των πολυωνύμων βαθμού το πολύ ένα και $h = \max_j(x_{j+1} - x_j)$. Μπορούμε εύκολα να δούμε ότι $\dim V_h = N$ και ότι οι συναρτήσεις

$$\phi_j(x) = \begin{cases} \frac{x-x_j}{x_j-x_{j-1}}, & x_{j-1} \leq x \leq x_j, \\ \frac{x_{j+1}-x}{x_{j+1}-x_j}, & x_j \leq x \leq x_{j+1}, \\ 0, & \text{διαφορετικά}, \end{cases} \qquad j = 1,\ldots,N, \tag{4.15}$$

αποτελούν βάση του χώρου V_h, το οποίο φαίνεται από το ακόλουθο λήμμα.

Λήμμα 4.2. *Οι συναρτήσεις* $\{\phi_j\}_{j=1}^N$ *της* (4.15) *αποτελούν βάση του χώρου* V_h.

Απόδειξη. Αρκεί να δείξουμε ότι οι $\{\phi_j\}_{j=1}^N$ είναι γραμμικώς ανεξάρτητες και παράγουν τον χώρο V_h. Είναι φανερό ότι οι συναρτήσεις $\{\phi_j\}_{j=1}^N$ είναι στοιχεία του V_h. Θεωρούμε τώρα έναν γραμμικό συνδυασμό των συναρτήσεων ϕ_j ο οποίος μηδενίζεται στο $[a, b]$, δηλαδή

$$\sum_{j=1}^{N} \lambda_j \phi_j(x) = 0 \quad \forall\, x \in [a, b].$$

Τότε, επειδή $\phi_i(x_j) = 0$ για $i \neq j$ και $\phi_i(x_i) = 1$, έχουμε ότι $\lambda_i = 0, i = 1, \ldots, N$. Επομένως, $\{\phi_j\}_{j=1}^N$ είναι γραμμικώς ανεξάρτητες. Επιπλέον, αν $v \in V_h$, εύκολα βλέπουμε ότι

$$v(x) = \sum_{j=1}^{N} v(x_j)\phi_j(x) \quad \forall\, x \in [a, b],$$

διότι οι v και $\sum_{j=1}^N v(x_j)\phi_j$, είναι γραμμικά πολυώνυμα σε κάθε διαστήμα $[x_i, x_{i+1}]$ και ταυτίζονται στα άκρα του. Συνεπώς, οι συναρτήσεις $\{\phi_j\}_{j=1}^N$ αποτελούν βάση του χώρου V_h. □

Θεωρούμε τώρα το ακόλουθο πρόβλημα το οποίο καλείται *μέθοδος Galerkin* ή *πεπερασμένων στοιχείων*: Ζητείται $u_h \in V_h$, τέτοια ώστε

$$(u_h', \chi') + (qu_h, \chi) = (f, \chi), \quad \forall\, \chi \in V_h. \tag{4.16}$$

Παρατηρούμε ότι η σχέση (4.16) είναι η ασθενής ή μεταβολική μορφή του (4.1) περιορισμένη στον χώρο V_h. Θα δείξουμε καταρχήν ότι η λύση u_h του (4.16) υπάρχει και ορίζεται μονοσήμαντα.

Αφού η (4.16) ισχύει, για κάθε $\chi \in V_h$ θα ικανοποιείται και για τις ϕ_j, $j = 1, \ldots, N$, επομένως

$$(u_h', \phi_i') + (qu_h, \phi_i) = (f, \phi), \quad \text{για } i = 1, \ldots, N. \tag{4.17}$$

Επειδή οι $\{\phi_j\}_{j=1}^N$ αποτελούν μια βάση του V_h, τότε, αν υπάρχει η λύση u_h του (4.16), θα γράφεται ως γραμμικός συνδυασμός των $\{\phi_j\}_{j=1}^N$, δηλαδή $u_h = \alpha_1\phi_1 + \cdots + \alpha_N\phi_N$. Επομένως, η (4.17) είναι ισοδύναμη με την

$$\sum_{j=1}^{N} \alpha_j(\phi_j', \phi_i') + \sum_{j=1}^{N} \alpha_j(q\phi_j, \phi_i) = (f, \phi_i), \quad i = 1, \ldots, N. \tag{4.18}$$

Η οποία με τη σειρά της είναι ισοδύναμη με το ακόλουθο γραμμικό σύστημα

$$A\alpha = F, \tag{4.19}$$

όπου A είναι ένας $N \times N$ πίνακας με στοιχεία $A_{ij} = (\phi_j', \phi_i') + (q\phi_j, \phi_i),\ i, j = 1, \ldots, N$, $\alpha = (\alpha_1, \ldots, \alpha_N)^T$ και $F = (F_1, \ldots, F_N)^T$ με $F_i = (f, \phi_i),\ i = 1, \ldots, N$. Είναι εύκολο να δούμε ότι ο πίνακας A είναι συμμετρικός και θετικά ορισμένος. Πράγματι, έχουμε ότι

$$A_{ij} = (\phi_j', \phi_i') + (q\phi_j, \phi_i) = (\phi_i', \phi_j') + (q\phi_i, \phi_j) = A_{ji}.$$

Επίσης, για $w \in \mathbb{R}^N$, $w = (w_1, \ldots, w_N)^T$,

$$w^T A w = \sum_{i,j=1}^{N} w_i A_{ij} w_j = (v', v') + (qv, v) = \|v'\|^2 + \|\sqrt{q}v\|^2 \geq 0,$$

όπου $v = \sum_{i=1}^{N} w_i \phi_i$. Επειδή $v \in V_h$ εύκολα βλέπουμε ότι αν $w^T A w = 0$, τότε η v ειναι η σταθερή συνάρτηση και επειδή $v(a) = 0$, θα έχουμε ότι $v = 0$ στο $[a, b]$, δηλαδή $w = 0$. Επομένως, αφού ο A είναι συμμετρικός και θετικά ορισμένος πίνακας, αντιστρέφεται και άρα υπάρχει μοναδικό $\alpha \in \mathbb{R}^N$ λύση του (4.19). Συνεπώς, η λύση u_h του (4.16) υπάρχει και είναι μοναδική.

Παρατήρηση 4.2. Η επιλογή των συναρτήσεων $\{\phi_j\}_{j=1}^{N}$ της (4.15) ως βάση του χώρου V_h δεν είναι απαραίτητη για την απόδειξη της ύπαρξης και μοναδικότητας της λύσης u_h του (4.16). Η παραπάνω απόδειξη της υπάρξης και μοναδικότητας της u_h γενικεύεται για κάθε βάση του V_h. Μάλιστα η μέθοδος των πεπερασμένων στοιχείων μπορεί να ορισθεί με ανάλογο τρόπο για κάθε υπόχωρο S_h του V, στον οποίο έχουμε θεωρήσει ένα σύνολο συναρτήσεων που αποτελούν βάση.

Παρατήρηση 4.3. Αν και η u_h ορίζεται μονοσήμαντα για οποιαδήποτε επιλογή της βάσης του χώρου V_h, είναι προφανές ότι η αριθμητική επίλυση του γραμμικού συστήματος $A\alpha = F$ εξαρτάται σε μεγάλο βαθμό από τη συγκεκριμένη επιλογή. Για παράδειγμα, αν επιλέξουμε ως βάση του V_h τις συναρτήσεις (4.15), τότε ο πίνακας είναι τριαδιαγώνιος. Παρατηρήστε ότι, εκ κατασκευής, $(q\phi_j, \phi_i) = 0$ και $(\phi_j', \phi_i') = 0$ για $|i - j| \geq 2$. Συνεπώς, το γραμμικό σύστημα $A\alpha = F$ μπορεί να λυθεί με τον αλγόριθμο της Παραγράφου 3.2, δείτε για παράδειγμα (Ακρίβης & Δουγαλής, 2015).

Παρατήρηση 4.4. Ορίζουμε τη διγραμμική μορφή $a(\cdot, \cdot) : V \times V \to \mathbb{R}$ ως

$$a(v, w) = (v', w') + (qv, w).$$

Τότε, το συνεχές πρόβλημα (4.9) και το διακριτό πρόβλημα (4.16) γράφονται, αντίστοιχα, ως

$$a(u, v) = (f, v), \quad \forall\, v \in V, \tag{4.20}$$
$$a(u_h, \chi) = (f, \chi), \quad \forall\, \chi \in V_h. \tag{4.21}$$

Αφαιρώντας κατά μέλη τις δύο αυτές σχέσεις, έχουμε

$$a(u - u_h, \chi) = 0, \quad \forall\, \chi \in V_h, \tag{4.22}$$

η οποία μπορεί να ερμηνευθεί ως μια σχέση *ορθογωνιότητας* του σφάλματος $u - u_h$ προς τα στοιχεία του χώρου V_h.

Έχοντας εξασφαλίσει την ύπαρξη της προσέγγισης u_h, αντιμετωπίζουμε τώρα το πρόβλημα της εκτίμησης του σφάλματος $u - u_h$. Ειδικώτερα, στην περίπτωση μιας *ομοιόμορφης* διαμέρισης του $[a, b]$ με $h = x_{j+1} - x_j = 1/(N + 1)$, όπου N θετικός ακέραιος, θα δείξουμε ότι $\|u - u_h\| \to 0$ όταν $h \to 0$, για κάποια κατάλληλη νόρμα $\|\cdot\|$. Για τον σκοπό αυτό, ορίζουμε τον τελεστή παρεμβολής $I_h : C_0[a, b] \to V_h$, όπου η παρεμβάλουσα $I_h v$ μιας συνάρτησης $v \in C_0[a, b]$ ορίζεται ως το στοιχείο του V_h

$$(I_h v)(x) = \sum_{j=1}^{N} v(x_j)\phi_j(x).$$

Εύκολα μπορούμε να δούμε ότι

$$(I_h v)(x_j) = v(x_j), \quad j = 1, \ldots, N, \tag{4.23}$$

και μάλιστα είναι το μοναδικό στοιχείο του V_h με αυτή την ιδιότητα, βλ. π.χ. (Ακρίβης & Δουγαλής, 2015). Επίσης, μπορούμε να δείξουμε χρησιμοποιώντας τη θεωρία παρεμβολής Lagrange, βλ. π.χ. (Ακρίβης & Δουγαλής, 2015) ότι

$$\max_{a \le x \le b} |v(x) - (I_h v)(x)| \le \frac{h^2}{8} \max_{a \le x \le b} |v''(x)|.$$

Επιπλέον, η I_h έχει την ακόλουθη προσεγγιστική ιδιότητα,

Λήμμα 4.3. *Αν $v \in C_0^2[a, b]$, τότε υπάρχει σταθερά C, ανεξάρτητη των h και v, τέτοια ώστε*

$$\|v - I_h v\| + h\|(v - I_v h)'\| \le C\, h^2 \|v''\|. \tag{4.24}$$

Απόδειξη. Γράφουμε $e = v - I_h v$. Λόγω της (4.23), έχουμε $e(x_j) = 0$, $j = 0, \ldots, N+1$. Επομένως, σε κάθε διάστημα $[x_i, x_{i+1}]$ η $e \in C_0^1[x_i, x_{i+1}]$, οπότε από την ανισότητα Poincaré–Friedrichs (4.5) λαμβάνουμε

$$\int_{x_i}^{x_{i+1}} e^2(x)\,dx \le h^2 \int_{x_i}^{x_{i+1}} [e'(x)]^2\,dx, \quad i = 0, 1, \ldots, N, \tag{4.25}$$

και αθροίζοντας όλες τις παραπάνω σχέσεις, παίρνουμε

$$\|e\| \le h\|e'\|. \tag{4.26}$$

Το θεώρημα του Rolle, στο διάστημα $[x_i, x_{i+1}]$ μας εξασφαλίζει την ύπαρξη ενός πραγματικού αριθμού $\xi \in (x_i, x_{i+1})$, τέτοιου ώστε $e'(\xi) = 0$. Τότε,

$$e'(x) = \int_\xi^x e''(s)\,ds = \int_\xi^x v''(s)\,ds, \quad x \in [x_i, x_{x+1}],$$

από την οποία προκύπτει, με την ανισότητα Cauchy–Schwarz,

$$|e'(x)|^2 \le h \int_{x_i}^{x_{i+1}} |v''(s)|^2\,ds.$$

Ολοκληρώνοντας και τα δύο μέλη στο διάστημα $[x_i, x_{i+1}]$, παίρνουμε

$$\int_{x_i}^{x_{i+1}} |e'(x)|^2\,dx \le h^2 \int_{x_i}^{x_{i+1}} |v''(s)|^2\,ds.$$

Τέλος, αθροίζοντας από $i = 0$ έως $i = N$, έχουμε

$$\|e'\| \le h\|v''\|. \tag{4.27}$$

Ο ισχυρισμός του λήμματος προκύπτει τώρα συνδυάζοντας τις (4.26) και (4.27). □

Οι ιδιότητες προσέγγισης της παρεμβάλουσας που αποδείχθηκαν στο Λήμμα 4.3 μας επιτρέπουν τώρα να αποδείξουμε εκτιμήσεις για τα σφάλματα $u - u_h$ και $u' - u_h'$:

Θεώρημα 4.2. *Εστω $u \in C^2[a, b]$ η λύση του προβλήματος* (4.1) *και $u_h \in V_h$ η λύση του μεταβολικού προβλήματος* (4.16)*. Τότε υπάρχει σταθερά C, ανεξάρτητη των u και h, τέτοια ώστε*

$$\|u - u_h\| + h\|u' - u_h'\| \le Ch^2\|u''\|. \tag{4.28}$$

Απόδειξη. Από τον ορισμό της διγραμμικής μορφής $a(\cdot,\cdot)$ και τις ανισότητες Cauchy–Schwarz και Poincaré–Friedrichs, έχουμε, για $v, w \in V$,

$$|a(v,w)| \leq \|v'\|\|w'\| + \max_x q(x)\|v\|\|w\| \leq C\|v'\|\|w'\|. \tag{4.29}$$

Επειδή $q(x) \geq 0$ στο $[a,b]$, λαμβάνουμε τη θεμελιώδη, για τη συνέχεια της απόδειξης, σχέση

$$a(v,v) \geq \|v'\|^2. \tag{4.30}$$

Από τη σχέση ορθογωνιότητας (4.22) έχουμε για $\chi \in V_h$ ότι

$$\begin{aligned} a(u-u_h, u-u_h) &= a(u-u_h, u) - a(u-u_h, u_h) = a(u-u_h, u) \\ &= a(u-u_h, u) - a(u-u_h, \chi) \\ &= a(u-u_h, u-\chi). \end{aligned}$$

Επομένως, από τις σχέσεις (4.29) και (4.30) λαμβάνουμε

$$\|u'-u_h'\|^2 \leq C\,\|u'-u_h'\|\|u'-\chi'\|,$$

ή ισοδύναμα,

$$\|u'-u_h'\| \leq C\,\|u'-\chi'\|, \quad \forall\, \chi \in V_h. \tag{4.31}$$

Επιλέγοντας $\chi = I_h u$ στην (4.31) και χρησιμοποιώντας τη σχέση (4.24), λαμβάνουμε την εκτίμηση του σφάλματος

$$\|u'-u_h'\| \leq Ch\|u''\|. \tag{4.32}$$

Για την εκτίμηση του σφάλματος $\|u-u_h\|$ θα χρησιμοποιήσουμε το λεγόμενο *δυικό επιχείρημα* ή *τέχνασμα του Nitsche*. Έστω $\psi \in V$ η λύση του προβλήματος

$$a(\psi, v) = (u-u_h, v), \quad \forall\, v \in V.$$

Επίσης, λόγω της ανισότητας της ελλειπτικής ομαλότητας (4.10) έχουμε ότι

$$\|\psi''\| \leq C\|u-u_h\|. \tag{4.33}$$

Θέτοντας $v = u-u_h$ λαμβάνουμε, χρησιμοποιώντας ξανά τη σχέση ορθογωνιότητας (4.22),

$$\|u-u_h\|^2 = (u-u_h, u-u_h) = a(\psi, u-u_h) = a(u-u_h, \psi-\chi), \tag{4.34}$$

για οποιδήποτε $\chi \in V_h$. Επιλέγουμε τώρα $\chi = I_h\psi$ στην (4.34) και χρησιμοποιώντας την ιδιότητα προσέγγισης (4.24) της παρεμβάλλουσας I_h και τη σχέση (4.29), έχουμε

$$\|u-u_h\|^2 \leq a(u-u_h, \psi - I_h\psi) \leq Ch\,\|u'-u_h'\|\,\|\psi''\|. \tag{4.35}$$

Στη συνέχεια, από τις σχέσεις (4.35) και (4.33) λαμβάνουμε

$$\|u - u_h\| \leq Ch\,\|u' - u_h'\|. \tag{4.36}$$

Συνεπώς, συνδυάζοντας τις (4.32) και (4.36) προκύπτει η ζητούμενη (4.28). □

Παρατήρηση 4.5. Από την απόδειξη του Θεωρήματος 4.2 μπορούμε να δούμε ότι οι εκτιμήσεις σφάλματος (4.28) βασίζονται κατά κύριο λόγο στην προσεγγιστική ιδιότητα που δείξαμε για την παρεμβάλλουσα I_h στην (4.24). Έστω ότι υπάρχει ένας υπόχωρος S_h^r του V, για τον οποίο μπορούμε να δείξουμε ότι ικανοποιεί την ακόλουθη προσεγγιστική ιδιότητα: Υπάρχει σταθερά C ανεξάρτητη του h, τέτοια ώστε για κάθε $v \in C^r[a,b]$, υπάρχει $\chi \in S_h^r$, τέτοιο ώστε

$$\|v - \chi\| + h\|v' - \chi'\| \leq Ch^r\|v\|_r,$$

με $\|v\|_r = (\sum_{s=0}^r \|v^{(s)}\|^2)^{1/2}$. Τότε μπορούμε να δείξουμε το ανάλογο της (4.28), βλ. Άσκηση 4.4. Δηλαδή ότι, αν $u \in C^r[a,b]$, τότε η λύση $u_h \in S_h^r$ του αντίστοιχου προβλήματος πεπερασμένων στοιχείων ικανοποιεί τη σχέση

$$\|u - u_h\| + h\|u' - u_h'\| \leq Ch^r\|u\|_r.$$

4.3 Μέθοδος πεπερασμένων στοιχείων για ομογενείς συνοριακές συνθήκες Neumann

Σε αυτήν την παράγραφο θα θεωρήσουμε και πάλι το πρόβλημα δύο σημειών (4.1) αλλά με *ομογενείς συνοριακές συνθήκες Neumann*, δηλαδή

$$-u''(x) + q(x)u(x) = f(x), \quad \text{για } x \in (a,b), \text{ με } u'(a) = u'(b) = 0. \tag{4.37}$$

Για να διασφαλίσουμε τη μοναδικότητα της λύσης του συγκεκριμένου προβλήματος θα υποθέσουμε ότι $q_{\min} = \min_{x\in[a,b]} q(x) > 0$. Συμβολίζουμε τώρα με $\widetilde{V}$ τον υπόχωρο του $C[a,b]$,

$$\widetilde{V} = \{v \in C[a,b] : v \text{ είναι κατά τμήματα συνεχώς παραγωγίσιμη}\}, \tag{4.38}$$

οι οποίες σε αντίθεση με τον V στην (4.4), δεν μηδενίζονται αναγκαστικά στα άκρα του $[a,b]$. Η ανάλογη με την (4.20) μεταβολική μορφή του προβλήματος (4.37), είναι

$$a(u,v) = (f,v), \quad \forall v \in \widetilde{V}. \tag{4.39}$$

Στη συνέχεια θεωρούμε και πάλι έναν διαμερισμό του $[a,b]$, $a = x_0 < x_1 < \cdots < x_{N+1} = b$ και συμβολίζουμε με $\widetilde{V}_h$ τον χώρο των συνεχών κατά τμήματα γραμμικών συναρτήσεων,

$$\widetilde{V}_h = \{\chi \in C[a,b] : \chi|_{[x_j,x_{j+1}]} \in \mathbb{P}_1\},$$

όπου $h = \max_j(x_{j+1} - x_j)$. Μια βάση του $\widetilde{V}_h$ αποτελείται από τις συναρτήσεις $\{\phi_i\}_{i=0}^{N+1}$, όπου οι συναρτήσεις ϕ_i, $1 \leq i \leq N$, είναι αυτές που ορίστηκαν στη σχέση (4.15) και οι ϕ_0, ϕ_{N+1} ορίζονται ως

$$\phi_0(x_0) = 1, \quad \phi_0(x_j) = 0,\ j \neq 0,$$
$$\phi_{N+1}(x_{N+1}) = 1, \quad \phi_{N+1}(x_j) = 0,\ j \neq N+1.$$

Επομένως, $\dim\widetilde{V}_h = N + 2$. Θεωρούμε τώρα το ακόλουθο πρόβλημα: Ζητείται $u_h \in \widetilde{V}_h$, τέτοια ώστε

$$a(u_h, \chi) = (f, \chi), \quad \forall\, \chi \in \widetilde{V}_h. \tag{4.40}$$

Παρόμοια, όπως και στην προηγούμενη παράγραφο, μπορούμε να γράψουμε το (4.40) ισοδύναμα ως ένα γραμμικό σύστημα,

$$Ac = F,$$

όπου ο A είναι ένας $(N + 2) \times (N + 2)$ συμμετρικός, θετικά ορισμένος, τριδιαγώνιος πίνακας, με στοιχεία $A_{ij} = a(\phi_j, \phi_i)$, $i, j = 0, \ldots, N + 1$, $F = ((f, \phi_0), \ldots, (f, \phi_{N+1}))^T$ και $c = (c_0, \ldots, c_{N+1})^T$, όπου $u_h = \sum_{j=0}^{N+1} c_j\phi_j$.

Για τη λύση $u_h \in V_h$ του (4.40) μπορούμε να δείξουμε το ακόλουθο θεώρημα, βλ. π.χ. (Ακρίβης & Δουγαλής, 2013· Ακρίβης & Δουγαλής, 2005).

Θεώρημα 4.3. *Έστω $u \in C^2[a, b]$ η λύση του προβλήματος* (4.37) *και $u_h \in \widetilde{V}_h$ η λύση του μεταβολικού προβλήματος* (4.40)*. Τότε υπάρχει σταθερά C, ανεξάρτητη των u και h, τέτοια ώστε*

$$\|u - u_h\| + h\|u' - u_h'\| \leq Ch^2\|u''\|. \tag{4.41}$$

4.4 Υλοποίηση της μεθόδου Galerkin

Το βασικό πρόβλημα στην υλοποίηση της μεθόδου των πεπερασμένων στοιχείων είναι η κατασκευή του πίνακα (και, κατά δεύτερο λόγο του δεξιού μέλους) του συστήματος $Ac = F$ που προκύπτει από τη μέθοδο Galerkin. Δεδομένης της διαμέρισης $a = x_0 < x_1 < \cdots < x_N < x_{N+1} = b$ του $[a, b]$, θα αναφερόμαστε σε κάθε διάστημα $I_e = [x_{e-1}, x_e]$, $e = 1, \ldots, N + 1$ ως *πεπερασμένο στοιχείο*. Το κάθε στοιχείο I_e έχει δύο *κόμβους*, τους x_{e-1} και x_e, που αντιστοιχούν στους *τοπικούς κομβικούς δείκτες* $j = 0$ και $j = 1$, αντίστοιχα. Ο δείκτης κάθε στοιχείου και οι τοπικοί κομβικοί δείκτες καθορίζουν τον λεγόμενο *καθολικό κομβικό δείκτη* i κάθε κόμβου x_i διαμέσου της απλής σχέσης

$$i = i(e, j) = e + j - 1, \quad e = 1, 2, \ldots, N + 1,\ j = 0, 1.$$

Οι συναρτήσεις $\{\phi_i\}_{i=0}^{N+1}$, η βάση του χώρου πεπερασμένων στοιχείων της προηγούμενης παραγράφου, μπορούν να περιγραφούν με έναν εξαιρετικά απλό και αποτελεσματικό τρόπο, για την κατασκευή του πίνακα A: συμβολίζουμε με ϕ_j^e, $j = 0, 1$, τον περιορισμό στο διάστημα $[x_{e-1}, x_e]$ των συναρτήσεων ϕ_{e-1}, ϕ_e, αντίστοιχα, και αναφερόμαστε σε αυτές ως τις *τοπικές συναρτήσεις βάσης*. Αν τώρα,

$$\Phi_0(y) = \begin{cases} 1-y & \text{αν } 0 \le y \le 1 \\ 0 & \text{διαφορετικά} \end{cases}, \qquad \Phi_1(y) = \begin{cases} y & \text{αν } 0 \le y \le 1 \\ 0 & \text{διαφορετικά} \end{cases},$$

τότε

$$\phi_j^e(x) = \Phi_j\left(\frac{x - x_{e-1}}{h_e}\right), \quad j = 0, 1,$$

όπου $h_e = x_e - x_{e-1}$. Ισοδύναμα, μπορούμε να αντιστοιχίσουμε σε κάθε διάστημα $I_e = [x_{e-1}, x_e]$ τον ομοπαραλληλικό μετασχηματισμό $x = h_e y + x_{e-1}$ και να ορίσουμε τις *τοπικές συναρτήσεις* βάσης ϕ_0^e, ϕ_1^e ως

$$\phi_0^e(x) = \Phi_0(y),\ \phi_1^e(x) = \Phi_1(x),\ \text{όταν } x = h_e y + x_{e-1}.$$

Συνεπώς, αν v είναι μια συνάρτηση στον χώρο που παράγεται από τις συναρτήσεις $\{\phi_i\}_{i=0}^{N+1}$, λαμβάνουμε την αναπαράσταση

$$v(x) = \sum_{i=0}^{N+1} \sum_{j=0}^{1} v_{i(e,j)} \phi_j^e(x),$$

όπου έχουμε θέσει $v_i = v(x_i)$. Ειδικώτερα, στο διάστημα (x_{e-1}, x_e) έχουμε

$$v(x) = v_{e-1}\phi_0^e(x) + v_e \phi_1^e(x). \tag{4.42}$$

Ο συμβολισμός και οι παρατηρήσεις της προηγούμενης παραγράφου μας επιτρέπουν τώρα την εύκολη κατασκευή του πίνακα και του δεξιού μέλους των εξισώσεων της μεθόδου Galerkin. Στη γλώσσα των πεπερασμένων στοιχείων, η κατασκευή του πίνακα του συστήματος αναφέρεται ως *συναρμολόγηση* (assembly), ορολογία που δικαιολογείται από το γεγονός ότι

$$a(v, w) = \sum_e a_e(v, w),$$

όπου

$$a_e(v, w) = \int_{I_e} v'w' + qvw = \int_{x_{e-1}}^{x_e} [v'(x)w'(x) + q(x)v(x)w(x)]\, dx,$$

για οποιεσδήποτε συναρτήσεις v, w του χώρου πεπερασμένων στοιχείων. Χρησιμοποιώντας την αναπαράσταση (4.42) και τον ομοπαραλληλικό μετασχηματισμό $x \mapsto (x - x_{e-1})/h_e$, έχουμε διαδοχικά

$$\begin{aligned} a_e(v,w) &= \int_{x_{e-1}}^{x_e} \left(\sum_j v_{i(e,j)} \phi_j^e(x) \right)' \left(\sum_j w_{i(e,j)} \phi_j^e(x) \right)' dx \\ &\quad + \int_{x_{e-1}}^{x_e} q(x) \left(\sum_j v_{i(e,j)} \phi_j^e(x) \right) \left(\sum_j w_{i(e,j)} \phi_j^e(x) \right) dx \\ &= \frac{1}{h_e} \int_0^1 \left(\sum_j v_{i(e,j)} \Phi_j(y) \right)' \left(\sum_j w_{i(e,j)} \Phi_j(y) \right)' dy \\ &\quad + h_e \int_0^1 q(x_{e-1} + h_e y) \left(\sum_j v_{i(e,j)} \Phi_j(y) \right) \left(\sum_j w_{i(e,j)} \Phi_j(y) \right) dy. \end{aligned}$$

Μπορούμε, τέλος, να γράψουμε την παραπάνω σχέση σε μορφή πίνακα επί διάνυσμα ως

$$a_e(v,w) = \frac{1}{h_e} \begin{bmatrix} v_{i(e,0)} & v_{i(e,1)} \end{bmatrix} S_e \begin{bmatrix} w_{i(e,0)} \\ w_{i(e,1)} \end{bmatrix} + h_e \begin{bmatrix} v_{i(e,0)} & v_{i(e,1)} \end{bmatrix} M_e \begin{bmatrix} w_{i(e,0)} \\ w_{i(e,1)} \end{bmatrix},$$

όπου S_e είναι ο 2×2 *τοπικός πίνακας ακαμψίας*

$$S_e = \begin{bmatrix} \int_0^1 (\Phi_0')^2 & \int_0^1 \Phi_0' \Phi_1' \\ \int_0^1 \Phi_1' \Phi_0' & \int_0^1 (\Phi_1')^2 \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$$

και M_e είναι ο 2×2 *τοπικός πίνακας μάζας*

$$M_e = \begin{bmatrix} \int_0^1 \tilde{q}_e (\Phi_0)^2 & \int_0^1 \tilde{q}_e \Phi_0 \Phi_1 \\ \int_0^1 \tilde{q}_e \Phi_1 \Phi_0 & \int_0^1 \tilde{q}_e (\Phi_1)^2 \end{bmatrix}.$$

Εδώ, έχουμε θέσει $\tilde{q}_e(y) = q(x_{e-1} + h_e y)$. Η προσέγγιση των στοιχείων του πίνακα M_e μπορεί να γίνει στην πράξη με τη χρήση ενός κανόνα αριθμητικής ολοκλήρωσης. Στην περίπτωση που εξετάζουμε στη συγκεκριμένη παράγραφο, μπορεί να αποδειχθεί ότι η χρήση του κανόνα του τραπεζίου είναι ικανή για να εξασφαλίσει το αποτέλεσμα του Θεωρήματος 4.2.

4.5 Κυβικές splines

Κατασκευάζουμε έναν χώρο πεπερασμένων στοιχείων για τον οποίο η προσέγγιση u_h της λύσης u του προβλήματος (4.1) ή του ιδίου προβλήματος αλλά με συνοριακές συνθήκες Neumann (4.37), έχει υψηλότερη τάξη ακρίβειας από την τάξη ακρίβειας δύο που αποδείχθηκε στο Θεώρημα 4.2.

Θεωρούμε έναν ομοιόμορφο διαμερισμό του $I = [a, b]$ με βήμα $h = (b - a)/(N+1)$ με κόμβους $x_i = a+ih$, $i = 0, \ldots, N+1$, και τον χώρο συναρτήσεων

$$S_h = \{\phi \in C^2[a, b] : \phi\big|_{[x_i,x_{i+1}]} \in \mathbb{P}_3, 0 \le i \le N\},$$

τον λεγόμενο χώρο των *κυβικών splines*. Μπορούμε να κατασκευάσουμε μια βάση του χώρου S_h ως εξής: η συνάρτηση

$$S(x) = \begin{cases} \frac{1}{4}(x+2)^3 & -2 \le x \le -1, \\ \frac{1}{4}[1 + 3(x+1) + 3(x+1)^2 - 3(x+1)^3] & -1 \le x \le 0, \\ \frac{1}{4}[1 + 3(1-x) + 3(1-x)^2 - 3(1-x)^3] & 0 \le x \le 1, \\ \frac{1}{4}(2-x)^3 & 1 \le x \le 2, \\ 0 & \text{διαφορετικά}, \end{cases}$$

είναι η μοναδική συνάρτηση στον χώρο $C^2[-2, 2]$ για την οποία $\text{supp}(S) = [-2, 2]$, $S|_{[k,k+1]} \in \mathbb{P}_3$, για $k = -2, -1, 0, 1$, και, τέλος, $S(\pm 2) = S'(\pm 2) = S''(\pm 2) = 0, S(0) = 1$. Χρησιμοποιώντας την $S(x)$ ορίζουμε τις συναρτήσεις $\{\phi_j\}$, $-1 \le j \le N + 2$, στο $[a, b]$ από τις σχέσεις

$$\phi_j(x) = S\left(\frac{x - x_j}{h}\right)\Bigg|_{[a,b]}, \quad -1 \le j \le N + 2,$$

όπου $x_{-1} = a-h$ και $x_{N+2} = b+h$. Είναι εύκολο να δει κανείς ότι $\phi_j \in S_h$, $-1 \le j \le N + 2$, ότι αυτές οι συναρτήσεις αποτελούν μια βάση του χώρου S_h [δείτε, για παράδειγμα, (Δουγαλής, 2013)] και η προσέγγιση Galerkin u_h από τον χώρο S_h του προβλήματος (4.37) ικανοποιεί την εκτίμηση

$$\|u - u_h\| \le Ch^4\|u^{(4)}\|.$$

Παρατήρηση 4.6. Στην περίπτωση του προβλήματος δύο σημείων με ομογενείς συνοριακές συνθήκες τύπου Dirichlet (4.1) αναζητούμε την προσέγγιση Galerkin u_h από τον χώρο $S_h^0 = \{\phi \in S_h : \phi(a) = \phi(b) = 0\}$, η διάσταση του οποίου είναι $N + 2$. Μια βάση του χώρου αυτού αποτελείται από τις κυβικές splines ϕ_j, $2 \le j \le N - 1$ και τις συναρτήσεις $\tilde{\phi}_0, \tilde{\phi}_1, \tilde{\phi}_N, \tilde{\phi}_{N+1}$ οι οποίες ορίζονται

ως γραμμικοί συνδιασμοί των $\phi_{-1}, \phi_0, \phi_1$ και $\phi_N, \phi_{N+1}, \phi_{N+2}$ και είναι τέτοιες, ώστε $\tilde{\phi}_0(a) = \tilde{\phi}_1(a) = 0$ και $\tilde{\phi}_N(b) = \tilde{\phi}_{N+1}(b) = 0$. Για παράδειγμα, θα μπορούσαμε να πάρουμε

$$\tilde{\phi}_0 = \phi_0 - 4\phi_{-1},\ \tilde{\phi}_1 = \phi_1 - \phi_{-1},\ \tilde{\phi}_N = \phi_N - \phi_{N+2},\ \tilde{\phi}_{N+1} = \phi_{N+1} - 4\phi_{N+2}.$$

4.6 Άλλες μέθοδοι και προβλήματα

Η μέθοδος πεπερασμένων στοιχείων, άλλα και άλλες μέθοδοι που βασίζονται στα πεπερασμένα στοιχεία, μπορούν να χρησιμοποιηθούν για την προσέγγιση του προβλήματος δύο σημείων με διάφορες συνοριακές συνθήκες, όπως και πιο γενικών προβλημάτων δεύτερης τάξεως. Παραπέμπουμε τον αναγνώστη στα (Ακρίβης & Δουγαλής, 2013· Ακρίβης, 2005· Johnson, 1987· Morton & Mayers, 2005· Roos, Stynes, & Tobiska, 2008· Larsson & Thomée, 2009) για πιο λεπτομερή παρουσίαση.

4.7 Ασκήσεις

4.1. Δείξτε την ανισότητα Cauchy–Schwarz

$$|(v, w)| \leq \|v\|\,\|w\|.$$

(Υπόδειξη: Χρησιμοποιήστε τη μη αρνητικότητα του $\|v + tw\|^2$, $t \in \mathbb{R}$.)

4.2. Βρείτε έναν χώρο πεπερασμένων στοιχείων για το πρόβλημα

$$\begin{cases} -u'' + qu = f & \text{στο } [a, b], \\ u(a) = u'(b) = 0. \end{cases}$$

4.3. Έστω ότι η συνάρτηση q του (4.1) λαμβάνει μη αρνητικές τιμές και J, $J(v) := (v', v') - 2(f, v)$. Δείξτε ότι η λύση u του (4.1) είναι το μόνο στοιχείο του V που η J λαμβάνει την ελάχιστη τιμή της στον V.

4.4. Έστω ότι S_h^r είναι ο υπόχωρος του χώρου V που θεωρήσαμε στην (4.4), για τον οποίο μπορούμε να δείξουμε ότι ικανοποιεί την ακόλουθη προσεγγιστική ιδιότητα: Υπάρχει σταθερά C ανεξάρτητη του h, τέτοια ώστε για κάθε $v \in C^r[a, b]$, υπάρχει $\chi \in S_h^r$ όπου

$$\|v - \chi\| + h\|v' - \chi'\| \leq Ch^r\|v\|_r,$$

με $\|v\|_r = (\sum_{s=0}^{r} \|v^{(s)}\|^2)^{1/2}$. Τότε, αν $u \in C_0^r[a,b]$, η λύση $u_h \in S_h^r$ του αντίστοιχου προβλήματος πεπερασμένων στοιχείων,

$$a(u_h, \chi) = (f, \chi), \quad \forall \chi \in S_h^r,$$

ικανοποιεί τη σχέση

$$\|u - u_h\| + h\|u' - u_h'\| \leq Ch^r \|u\|_r.$$

Βιβλιογραφία

Ακρίβης, Γ. (2005). *Μέθοδοι Πεπερασμένων Στοιχείων*. Λευκωσία, Κύπρος. (Πανεπιστημιακές Σημειώσεις).

Ακρίβης, Γ., & Δουγαλής, Β. (2005). *Αριθμητικές Μέθοδοι για Μερικές Διαφορικές Εξισώσεις*. Ιωάννινα. (Πανεπιστημιακές Σημειώσεις).

Ακρίβης, Γ., & Δουγαλής, Β. (2013). *Αριθμητικές Μέθοδοι για Συνήθεις Διαφορικές Εξισώσεις*. Πανεπιστημιακές Εκδόσεις Κρήτης, Ηράκλειο.

Ακρίβης, Γ., & Δουγαλής, Β. (2015). *Εισαγωγή στην Αριθμητική Ανάλυση*. Πανεπιστημιακές Εκδόσεις Κρήτης, Ηράκλειο.

Δουγαλής, Β. (2013). *Finite element methods for the numerical solution of partial differential equations*. Αθήνα. (Πανεπιστημιακές Σημειώσεις).

Johnson, C. (1987). *Numerical solution of partial differential equations by the finite element method*. Cambridge University Press, Cambridge.

Larsson, S., & Thomée, V. (2009). *Partial differential equations with numerical methods* (Vol. 45). Springer-Verlag, Berlin.

Morton, K. W., & Mayers, D. F. (2005). *Numerical solution of partial differential equations* (Second ed.). Cambridge University Press, Cambridge.

Roos, H.-G., Stynes, M., & Tobiska, L. (2008). *Robust numerical methods for singularly perturbed differential equations* (Second ed., Vol. 24). Springer-Verlag, Berlin.

Κεφάλαιο 5

Η μέθοδος των πεπερασμένων διαφορών για την εξίσωση θερμότητας

Σε αυτό το κεφάλαιο θεωρούμε μια απλή παραβολική εξίσωση, την εξίσωση της θερμότητας, στη μια διάσταση ως προς τον χώρο. Θα κατασκευάσουμε μεθόδους πεπερασμένων διαφορών για το πρόβλημα αρχικών και συνοριακών τιμών με ομογενείς συνθήκες Dirichlet. Στη συνέχεια, θα δείξουμε ιδιότητες ευστάθειας, καθώς και κατάλληλες συνθήκες που πρέπει να ικανοποιεί η μέθοδος για να είναι ευσταθής. Τέλος, θα δείξουμε τη σύγκλιση της προσεγγιστικής λύσης στη ακριβή λύση.

5.1 Άμεση μέθοδος του Euler

Θεωρούμε το ακόλουθο πρόβλημα: Ζητείται μια συνάρτηση $u : [0, L] \times [0, T] \to \mathbb{R}$, τέτοια ώστε

$$\begin{aligned} u_t(x,t) &= u_{xx}(x,t), \qquad x \in [0,L],\ t \in [0,T], \\ u(0,t) &= u(L,t) = 0, \quad t \in [0,T], \\ u(x,0) &= g(x), \qquad\quad x \in [0,L], \end{aligned} \tag{5.1}$$

όπου $L > 0$, $g \in C[0, L]$. Όπως είπαμε στην Παραγράφο 1.2.2 η λύση του (5.1) δίνεται από τη

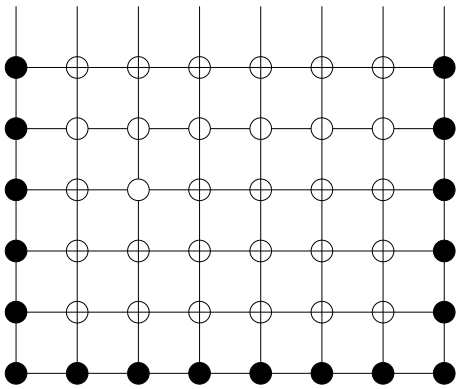
$$u(x,t) = \sum_{n=1}^{\infty} c_n e^{-\lambda_n^2 t} \sin(\lambda_n x), \tag{5.2}$$

με $\lambda_n = \dfrac{n\pi}{L}$ και $c_n = \dfrac{2}{L} \displaystyle\int_0^L g(x) \sin(\lambda_n x)$.

Όπως και στην προηγούμενη παράγραφο θα θεωρήσουμε ορισμένα σημεία στο χωρίο $[0, L] \times [0, T]$ και θα κατασκευάσουμε προσεγγίσεις της ακριβούς λύσης του (5.1) σε αυτά τα σημεία. Θεωρούμε λοιπόν έναν φυσικό αριθμό N και τη διαμέριση του διαστήματος $[0, L]$ από $N + 2$ ισαπέχοντα σημεία $0 = x_0 < x_1 < \cdots < x_N < x_{N+1} = L$, όπου $h = x_{i+1} - x_i$, $i = 0, \ldots, N$. Επίσης, θεωρούμε και έναν φυσικό αριθμό M και τη διαμέριση του $[0, T]$ από $M + 1$ ισαπέχοντα σημεία $0 = t^0 < t^1 < \cdots < t^M = T$, όπου $k = t^{j+1} - t^j$, $j = 0, \ldots, M - 1$, βλ. Σχήμα 5.1. Τότε σε κάθε σημείο (x_i, t^j) του διαμερισμού του $[0, L] \times [0, T]$, θα ισχύει:

$$u_t(x_i, t^j) = u_{xx}(x_i, t^j), \quad i = 1, \ldots, N, \ j = 0, \ldots, M. \tag{5.3}$$

Θα θεωρήσουμε λοιπόν προσεγγίσεις των τιμών της u στα παραπάνω σημεία (x_i, t^j) τις οποίες θα συμβολίσουμε με U_i^j, $i = 0, \ldots, N + 1$, $j = 0, \ldots, M$. Λόγω των συνοριακών συνθηκών $u(0, t^j) = u(L, t^j) = 0$, $j = 0, \ldots, M$, θέτουμε $U_0^j = U_{N+1}^j = 0$, $j = 0, \ldots, M$, και επειδή η λύση u της (5.1) είναι γνωστή για $t = 0$, θέτουμε $U_i^0 = g(x_i)$, $i = 1, \ldots, N$. Οι υπόλοιπες τιμες $u(x_i, t^j)$, $i = 1, \ldots, N$, $j = 1, \ldots, M$, είναι άγνωστες και θα τις προσεγγίσουμε. Στο Σχήμα



Σχήμα 5.1: Τα σημεία (x_i, t^j) του χωρίου $[0, L] \times [0, T]$, όπου σημειώνουμε με άσπρο κύκλο τα σημεία στα οποία αναζητούμε τις τιμές U_i^j και με έντονο μαύρο τα σημεία που αντιστοιχούν στις γνωστές συνοριακές τιμές.

5.1, για ένα παράδειγμα πλέγματος σημείων, σημειώνουμε με άσπρο κύκλο τα σημεία που αναζητούμε στις τιμές $u(x_i, t^j)$ και με έντονο μαύρο τα σημεία που αντιστοιχούν στις γνωστές συνοριακές τιμές της u. Τις άγνωστες τιμές $u(x_i, t^j)$ τις προσεγγίζουμε με τις τιμές U_i^j, $i = 1, \ldots, N$, $j = 1, \ldots, M$, οι οποίες προκύπτουν με τον ακόλουθο τρόπο.

Για να προσεγγίσουμε τις $u_t(x_i, t^j)$, $i = 1, \ldots, N$, $j = 0, \ldots, M - 1$, στην

(5.3) χρησιμοποιούμε τη $\delta_k^+ u(x_i, t^j)$ που θεωρήσαμε στην (2.1). Έτσι, αν υποθέσουμε ότι $u_{tt} \in C([0, L] \times [0, T])$, λόγω της (2.3), η (5.3) γίνεται για $i = 1, \ldots, N$,

$$\frac{u(x_i, t^{j+1}) - u(x_i, t^j)}{k} = u_{xx}(x_i, t^j) + \hat{\eta}_i^j, \quad j = 0, \ldots, M-1, \tag{5.4}$$

όπου

$$|\hat{\eta}_i^j| \leq \frac{k}{2} \max_{t \in [0,T]} |u_{tt}(x_i, t)|. \tag{5.5}$$

Στη συνέχεια, για να προσεγγίσουμε τις $u_{xx}(x_i, t^j)$, $i = 1, \ldots, N$, $j = 0, \ldots, M-1$, στην (5.4) χρησιμοποιούμε την προσέγγιση $\delta_{h,2}^c u(x_i, t^j)$ που θεωρήσαμε στην (2.8). Έτσι, αν υποθέσουμε ότι $\frac{\partial^4}{\partial x^4} u \in C([0, L] \times [0, T])$, λόγω της (2.9), η (5.4) γίνεται για $i = 1, \ldots, N$, και $j = 0, \ldots, M-1$,

$$\begin{aligned} \frac{u(x_i, t^{j+1}) - u(x_i, t^j)}{k} = & \frac{u(x_{i+1}, t^j) - 2u(x_i, t^j) + u(x_{i-1}, t^j)}{h^2} \\ & + \hat{\eta}_i^j + \tilde{\eta}_i^j, \end{aligned} \tag{5.6}$$

και

$$|\tilde{\eta}_i^j| \leq \frac{h^2}{12} \max_{x \in [0,L]} |\frac{\partial^4}{\partial x^4} u(x, t^j)|. \tag{5.7}$$

Αν θέσουμε τώρα $\eta_i^j = \hat{\eta}_i^j + \tilde{\eta}_i^j$, η (5.6) δίνει για $i = 1, \ldots, N$, και $j = 0, \ldots, M-1$,

$$\delta_h^+ u(x_i, t^j) - \delta_{h,2}^c u(x_i, t^j) = \eta_i^j, \tag{5.8}$$

και λόγω των (5.5) και (5.7) εύκολα βλέπουμε ότι ισχύει το ακόλουθο λήμμα.

Λήμμα 5.1. *Έστω u η λύση του* (5.1) *με $\frac{\partial^4}{\partial x^4} u \in C([0, L] \times [0, T])$ και $u_{tt} \in C([0, L] \times [0, T])$. Τότε, για την η_i^j που δίνεται στην* (5.8) *έχουμε ότι υπάρχει σταθερά C ανεξάρτητη των k και h, τέτοια ώστε*

$$\max_{0 \leq j \leq M-1} \max_{1 \leq i \leq N} |\eta_i^j| \leq C(k + h^2). \tag{5.9}$$

Θα κατασκευάσουμε λοιπόν προσεγγίσεις U_i^j της λύσης u του προβλήματος (5.1) στα σημεία (x_i, t^j), $i = 0, \ldots, N+1$, $j = 0, \ldots, M$, θεωρώντας τις ακόλουθες εξισώσεις,

$$\frac{U_i^{j+1} - U_i^j}{k} = \frac{U_{i+1}^j - 2U_i^j + U_{i-1}^j}{h^2}, \quad i = 1, \ldots, N, j = 0, \ldots, M-1, \tag{5.10}$$

$$\begin{cases} U_0^j = U_{N+1}^j = 0, \quad j = 0, \ldots, M \\ U_i^0 = g(x_i), \quad i = 1, \ldots, N. \end{cases} \tag{5.11}$$

Κατ' αναλογία με το Κεφάλαιο 3, το σφάλμα η_i^j που δίνεται στις (5.8) καλείται *τοπικό σφάλμα διακριτοποίησης* για τη μέθοδο (5.10)–(5.11).

Έστω τώρα ότι έχουμε θεωρήσει έναν διαμερισμό σημείων $\{x_i\}$ του $[0, L]$, με βήμα h και $\{t^j\}$ του $[0, T]$, με βήμα k, τότε προκύπτει ένας διαμερισμός σημείων $\{(x_i, t^j)\}$ του $[0, L] \times [0, T]$. Αν το τοπικό σφάλμα διακριτοποίησης η_i^j, ενός αριθμητικού σχήματος σε ένα σημείο ενός διαμερισμού $\{(x_i, t^j)\}$, φράσσεται κατά απόλυτη τιμή από το γινόμενο μιας θετικής σταθεράς, που δεν εξαρτάται από τα h και k, επί h^κ και k^μ, τότε λέμε ότι έχει *τάξη ακρίβειας κ ως προς h και μ ως προς k*.

Επομένως, λόγω των (5.5) και (5.7), το τοπικό σφάλμα διακριτοποίησης της μεθόδου (5.10)-(5.11) έχει τάξη ακρίβειας δύο ως προς h και ένα ως προς k.

Επίσης, κατ'αναλογία του Ορισμού 3.1, αν το τοπικό σφάλμα διακριτοποίησης μιας μεθόδου, η_i^j, τείνει στο μηδέν, καθώς τα βήματα h και k του διαμερισμού τείνουν στο μηδέν, τότε η μέθοδος λέγεται *συνεπής*.

Συνεπώς σύμφωνα με το Λήμμα 5.1, το τοπικό σφάλμα διακριτοποίησης τείνει στο μηδέν, καθώς τα h και k τείνουν στο μηδέν, οπότε η (5.10)-(5.11) είναι μια *συνεπής* μέθοδος.

Αν συμβολίσουμε τώρα με λ τον λόγο $\dfrac{k}{h^2}$, τότε η (5.10) γράφεται ως

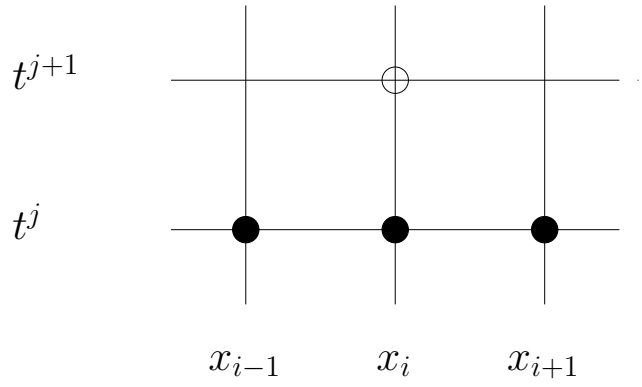
$$U_i^{j+1} = \lambda U_{i+1}^j + (1 - 2\lambda)U_i^j + \lambda U_{i-1}^j, \tag{5.12}$$

για $i = 1, \ldots, N$, $j = 0, \ldots, M - 1$. Ξεκινώντας από το χρονικό επίπεδο $t^0 = 0$, χρησιμοποιώντας δηλαδή τις αρχικές συνθήκες (5.11), από την (5.12) μπορούμε να υπολογίσουμε άμεσα την προσέγγιση στο χρονικό επίπεδο t^1. Συνεχίζοντας με αυτό τον τρόπο, λαμβάνουμε, με άμεσο τρόπο μέσω της (5.12), την προσέγγιση στο χρονικό επίπεδο t^{j+1}, από τις ήδη γνωστές τιμές στο χρονικό επίπεδο t^j.

Η μέθοδος (5.10)-(5.11) καλείται *άμεση μέθοδος του Euler*. Στο Σχήμα 5.2 φαίνεται ένα παράδειγμα πλέγματος, κοντά στο (x_i, t^j), όπου σημειώνουμε με άσπρο κύκλο το σημείο του χρονικού επιπέδου t^{j+1}, όπου σύμφωνα με την (5.12), αντιστοιχεί σε άγνωστη τιμή της προσέγγισης και με μαύρο αυτά του χρονικού επιπέδου t^j που αντιστοιχούν σε γνωστές τιμές της προσέγγισης.

Αν συμβολίσουμε με $U^j \in \mathbb{R}^N$ το διάνυσμα με συνιστώσες, $U_1^j, \ldots, U_N^j$, $U^j = (U_1^j, \ldots, U_N^j)^T$, μπορούμε να γράψουμε το σύστημα των εξισώσεων (5.12) ισοδύναμα ως

$$U^{j+1} = AU^j, \quad j = 0, \ldots, M - 1, \text{ με } U^0 = G, \tag{5.13}$$



Σχήμα 5.2: Σημεία ενός πλέγματος για τον υπολογισμό της λύσης με την άμεση μέθοδο του Euler. Με άσπρο σημειώνουμε τα σημεία που αντιστοιχούν σε άγνωστες τιμές της προσέγγισης και με μαύρο αυτά που αντιστοιχούν σε γνωστές τιμές.

όπου A είναι ο $N \times N$ πίνακας

$$A = \begin{pmatrix} 1-2\lambda & \lambda & 0 & & 0 \\ \lambda & 1-2\lambda & \lambda & 0 & \\ 0 & \ddots & \ddots & \ddots & 0 \\ & & \lambda & 1-2\lambda & \lambda \\ & & & \lambda & 1-2\lambda \end{pmatrix},$$

και $G = (g(x_1), \ldots, g(x_N))^T$.

Επομένως, για να υπολογίσουμε την προσέγγιση στο χρονικό επίπεδο t^{j+1}, απλώς πολλαπλασιάζουμε τον πίνακα A με τη γνωστή, από το προηγούμενο βήμα, U^j. Επειδή για να υπολογίσουμε την προσέγγιση σε κάθε χρονικό βήμα δεν χρειάζεται να αντιστρέψουμε κάποιον πίνακα, δηλαδή να λύσουμε κάποιο γραμμικό σύστημα με αριθμητικές μεθόδους, η μέθοδος καλείται *άμεση*.

Η ύπαρξη και η μοναδικότητα της προσεγγιστικής λύσης σε κάθε χρονικό βήμα είναι δεδομένη, διότι για να την υπολογίσουμε αρκεί να πολλαπλασιάζουμε έναν γνωστό πίνακα με τη γνωστή από το προηγούμενο χρονικό βήμα προσεγγιστική λύση.

Στη συνέχεια, θα δείξουμε το ακόλουθο θεώρημα για την *ευστάθεια* της άμεσης μεθόδου του Euler.

Θεώρημα 5.1. *Έστω $U^j \in \mathbb{R}^N$, $j = 0, \ldots, M$, τα διανύσματα που προκύπτουν από την* (5.13)*, με $U_0^j = U_{N+1}^j = 0$, $j = 0, \ldots, M$. Τότε, αν $\lambda = \frac{k}{h^2} \leq \frac{1}{2}$, ισχύει η*

ακόλουθη ανισότητα,

$$\max_{0\le i\le N+1} |U_i^j| \le \max_{0\le i\le N+1} |U_i^0|, \quad \textit{για } j = 0, 1, 2, \ldots, M.$$

Απόδειξη. Από τη σχέση (5.12) και το γεγονός $\lambda \le \frac{1}{2}$, εύκολα παίρνουμε

$$|U_i^{j+1}| \le \lambda(|U_{i+1}^j| + |U_{i-1}^j|) + (1-2\lambda)|U_i^j|, \quad 1 \le i \le N.$$

Αν θέσουμε λοιπόν $\bar{U}^j = \max_{0\le i\le N+1} |U_i^j|$, $j = 0, \ldots, M$, έχουμε

$$\bar{U}^{j+1} \le \bar{U}^j \le \cdots \le \bar{U}^0,$$

από όπου προκύπτει η ζητούμενη εκτίμηση. □

Για να αποδείξουμε την ευστάθεια της μεθόδου στο Θεώρημα 5.1 υποθέσαμε ότι $\lambda \le 1/2$. Λόγω της ύπαρξης μιας συνθήκης ανάμεσα στις παραμέτρους διακριτοποίησης k και h, λέμε ότι η μέθοδος είναι *ευσταθής υπό συνθήκες*. Το γεγονός ότι αυτή η συνθήκη είναι αναγκαία και δεν οφείλεται σε τυχόν αδυναμία της απόδειξης του Θεώρηματος 5.1 που δώσαμε φαίνεται από το ακόλουθο παράδειγμα.

Είναι απλό να δούμε ότι αν $L = 1$ και $g(x) = \sin(n\pi x)$, $n = 1, 2, \ldots$, τότε το πρόβλημα (5.1) γίνεται

$$u_t(x,t) = u_{xx}(x,t), \qquad x \in [0,1],\ t \in [0,T], \tag{5.14}$$
$$u(0,t) = u(1,t) = 0, \qquad t \in [0,T], \tag{5.15}$$
$$u(x,0) = \sin(n\pi x), \qquad x \in [0,1], \tag{5.16}$$

και η ακριβής λύση u είναι η $u(x,t) = e^{-n^2\pi^2 t}\sin(n\pi x)$, μια δηλαδή από τις θεμελιώδεις συναρτήσεις που εμφανίζονται στην έκφραση (5.2). Παρατηρούμε λοιπόν ότι σε κάθε χρονικό επίπεδο $t \ge 0$ καθεμιά από αυτές τις θεμελιώδεις συναρτήσεις είναι μια ημιτονοειδής συνάρτηση, της οποίας το πλάτος μειώνεται συνεχώς με εκθετικό τρόπο, καθώς το χρονικό επίπεδο αυξάνει. Είναι προφανές ότι, αν θεωρήσουμε για αρχική συνθήκη g μια από τις $\sin(n\pi x)$, $n = 1, 2, \ldots$, και η προσεγγιστική λύση που λαμβάνουμε με τη μέθοδο (5.10), δεν αποτυπώνει αυτήν την ιδιότητα, ή τουλάχιστον το πλάτος δεν παραμένει φραγμένο, δεν θα προσεγγίζει "καλά" την αντίστοιχη θεμελιώδη συνάρτηση.

Επειδή η ακριβής λύση του (5.14)–(5.16) αποτελείται από το γινόμενο της $e^{-n^2\pi^2 t}$, που εξαρτάται μόνο από τον χρόνο t, και της $\sin(n\pi x)$, υποθέτουμε ότι η λύση U_i^j της (5.10) έχει τη μορφή $U_i^j = w_j \sin(n\pi x_i)$. Δηλαδή αποτελείται από δύο παράγοντες, ο ένας από τους οποίους εξαρτάται μόνο από τον χρόνο t^j και ο άλλος είναι ακριβώς η ημιτονοειδής συνάρτηση που βρίσκεται στην έκφραση της

ακριβούς λύσης του (5.14). Αντικαθιστούμε λοιπόν την παραπάνω μορφή της U_i^j στην εξίσωση (5.12), οπότε παίρνουμε για $i = 1, \ldots, N, j = 0, \ldots, M-1$,

$$\begin{aligned} w_{j+1}\sin(n\pi x_i) = \lambda w_j \sin(n\pi x_{i+1}) + (1-2\lambda) w_j \sin(n\pi x_i) \\ + \lambda w_j \sin(n\pi x_{i-1}). \end{aligned} \tag{5.17}$$

Στη συνέχεια, εφαρμόζοντας την τριγωνομετρική ταυτότητα

$$\sin(\alpha \pm \beta) = \sin(\alpha)\cos(\beta) \pm \cos(\alpha)\sin(\beta), \tag{5.18}$$

η (5.17) γίνεται

$$\begin{aligned} w_{j+1}\sin(n\pi x_i) &= 2\lambda w_j \sin(n\pi x_i)\cos(\pi h) + (1-2\lambda) w_j \sin(n\pi x_i) \\ &= (2\lambda(\cos(\pi h) - 1) + 1) w_j \sin(n\pi x_i). \end{aligned} \tag{5.19}$$

Επίσης, λόγω της τριγωνομετρικής ιδιότητας

$$2\sin^2(\frac{\theta}{2}) = 1 - \cos(\theta), \quad \theta \in [0, 2\pi], \tag{5.20}$$

η (5.19) δίνει

$$w_{j+1} = (1 - 4\lambda \sin^2(\frac{n\pi h}{2})) w_j. \tag{5.21}$$

Συνεπώς, $w_j = \kappa^j w_0$, με $\kappa = 1 - 4\lambda \sin^2(\frac{n\pi h}{2})$. Άρα, η μέγιστη τιμή της U_i^j είναι φραγμένη, αν και μόνο αν $|\kappa| \leq 1$. Θα πρέπει, λοιπόν,

$$-1 \leq 1 - 4\lambda \sin^2(\frac{n\pi h}{2})) \leq 1 \text{ ή } 2\lambda \sin^2(\frac{n\pi h}{2})) \leq 1.$$

Επομένως, επειδή $|\sin(\frac{n\pi h}{2})| \leq 1$, για $n = 1, 2, \ldots$, αρκεί $\lambda \leq \dfrac{1}{2}$. Βλέπουμε ότι η συνθήκη ευστάθειας στο Θεώρημα 5.1 είναι αναγκαία για να έχουμε "καλές" προσεγγίσεις με την άμεση μέθοδο του Euler.

Μπορούμε να γενικεύσουμε την παραπάνω απόδειξη θεωρώντας τώρα ότι

$$U_i^j = w_j e^{r x_i I}, \quad r \in \mathbb{R},$$

με $I = \sqrt{-1}$ τη φανταστική μονάδα, ικανοποιούν την εξίσωση (5.12). Αν οι τιμές $|U_i^j|$ παραμένουν φραγμένες για κάθε $r \in \mathbb{R}$, θα καλούμε αυτή την ιδιότητα *ευστάθεια von Neumann*. Είναι φανερό ότι για να είναι φραγμένες οι $|U_i^j|$, αρκεί να φράσσονται οι $|w_j|$ για κάθε j. Μια πιο διεξοδική μελέτη της ευστάθειας von Neumann ο ενδιαφερόμενος αναγνώστης μπορεί να βρεί στο (Strikwerda, 2004).

Μπορεί να αποδειχθεί ότι για την εξίσωση της θερμότητας η ευστάθεια von Neumann συνεπάγεται την ευστάθεια μιας μεθόδου, βλ. π.χ. (Morton & Mayers,

2005). Στη συνέχεια, θα δείξουμε ότι για να είναι η άμεση μέθοδος του Euler von Neumann ευσταθής πρέπει και πάλι να ισχύει $\lambda \leq \frac{1}{2}$.

Υποθέτουμε λοιπόν ότι οι

$$U_i^j = w_j e^{rx_i I}, \quad r \in \mathbb{R},$$

με $I = \sqrt{-1}$ τη φανταστική μονάδα, ικανοποιούν την εξίσωση (5.12) και, άρα,

$$\begin{aligned} w_{j+1} e^{rx_i I} &= \lambda w_j e^{rx_{i+1} I} + (1-2\lambda) w_j e^{rx_i I} + \lambda w_j e^{rx_{i-1} I} \\ &= (\lambda e^{rhI} + 1 - 2\lambda + \lambda e^{-rhI}) w_j e^{rx_i I}. \end{aligned} \tag{5.22}$$

Λόγω τώρα της ταυτότητας

$$e^{\theta I} = \cos(\theta) + I \sin(\theta),$$

έχουμε

$$\cos(\theta) = \frac{1}{2}(e^{\theta I} + e^{-\theta I}), \quad \theta \in [0, 2\pi]. \tag{5.23}$$

Οπότε η (5.22) λόγω της (5.20) γίνεται

$$\begin{aligned} w_{j+1} &= (1 - 2\lambda + 2\cos(rh)) w_j e^{rx_i I} \\ &= (1 - 4\lambda \sin^2(\frac{rh}{2})) w_j. \end{aligned}$$

Συνεπώς $w_j = \kappa^j w_0$, με $\kappa = 1 - 4\lambda \sin^2(\frac{rh}{2})$. Για να παραμένει η λύση U_i^j φραγμένη κατ'απόλυτη τιμή θα πρέπει $|\kappa| \leq 1$, για κάθε $r \in \mathbb{R}$. Καταλήγουμε δηλαδή στο ίδιο αποτέλεσμα όπως και παραπάνω, δηλαδή ότι $\lambda \leq \frac{1}{2}$.

Στη συνέχεια, δείχνουμε το ακόλουθο για τη σύγκλιση της μεθόδου.

Θεώρημα 5.2. *Έστω ότι η λύση u του προβλήματος* (5.1) *είναι αρκετά ομαλή, και U_i^j, $i = 1, \ldots, N$, $j = 0, \ldots, M$, η λύση της μεθόδου* (5.4)*. Τότε, αν $\lambda = \frac{k}{h^2} \leq \frac{1}{2}$, υπάρχει μια σταθερά C, ανεξάρτητη των k, h, τέτοια ώστε*

$$\max_{1 \leq j \leq M} \max_{0 \leq i \leq N+1} |U_i^j - u(x_i, t^j)| \leq C(k + h^2). \tag{5.24}$$

Απόδειξη. Θέτουμε $E_i^j = U_i^j - u(x_i, t^j)$, $i = 0, \ldots, N+1$, $j = 0, \ldots, M$, όπου λόγω των σχέσεων $U_0^j = u(0, t^j) = 0$ και $U_{N+1}^j = u(L, t^j) = 0$, έχουμε $E_0^j = E_{N+1}^j = 0$. Αφαιρούμε τώρα κατά μέλη τις (5.10) και (5.6), οπότε παίρνουμε

$$E_i^{j+1} = \lambda E_{i+1}^j + (1-2\lambda) E_i^j + \lambda E_{i-1}^j + k\eta_i^j, \ i = 1, \ldots, N, \ j = 1, \ldots, M-1.$$

	$t = 0.025$		$t = 0.05$		$t = 0.0625$		$t = 0.075$	
M	j	$\bar{E}^j$	j	$\bar{E}^j$	j	$\bar{E}^j$	j	$\bar{E}^j$
24	6	0.0302	12	0.0216	15	0.0181	18	2.2244
32	8	0.0217	16	0.0157	20	0.1358	24	174.6062
128	32	0.0036	64	0.0027	80	0.0020	96	0.0015

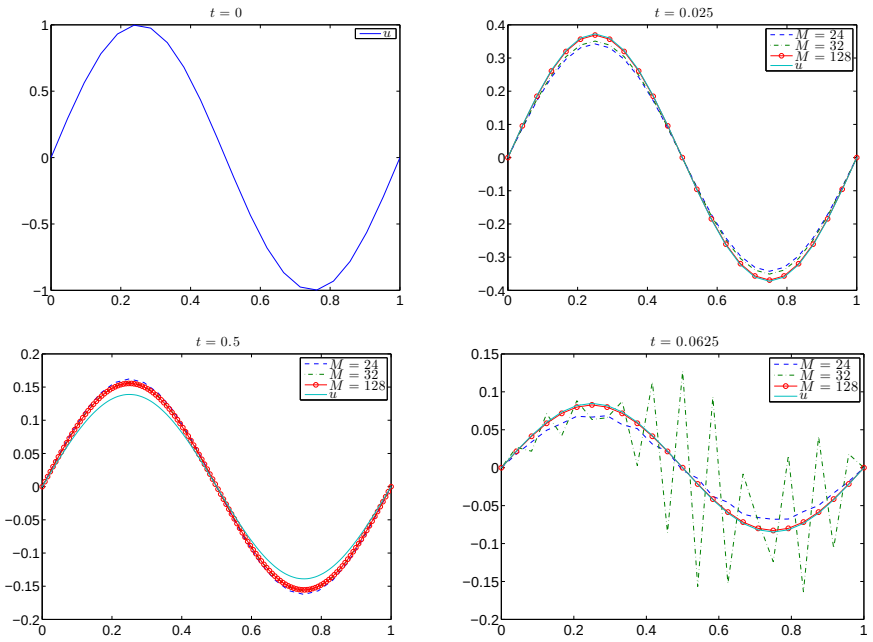
Πίνακας 5.1: Το μέγιστο σφάλμα $\bar{E}^j = \max_{1\le i\le N} |U_i^j - u(x_i, t^j)|$ στα χρονικά επίπεδα $t = 0.025, 0.05$ και 0.0625 του Παραδείγματος 5.1 για $N = 23$ και $M = 24, 32, 128$. Στη στήλη αριστερά του κάθε σφάλματος δίνεται το χρονικό βήμα j τέτοιο ώστε $t = jk$, με $k = 1/M$.

Θέτουμε, στη συνέχεια, $\bar{E}^j = \max_{1\le i\le N} |E_i^j|$ και $\bar{\eta} = \max_{1\le j\le M} \max_{0\le i\le N+1} |\eta_i^j|$. Οπότε έχουμε

$$\bar{E}^{j+1} \le 2\lambda\bar{E}^j + (1-2\lambda)\bar{E}^j + k\bar{\eta} \le \bar{E}^j + k\bar{\eta} \le \cdots \le \bar{E}^0 + jk\bar{\eta}.$$

Από εδώ, επειδή $jk \le T$ και λόγω του Λήμματος 5.1, προκύπτει η ζητούμενη εκτίμηση. □

Παράδειγμα 5.1. Ας θεωρήσουμε το πρόβλημα (5.14) με $n = 2$ και $T = 0.1$, όπου η ακριβής λύση είναι $u(x,t) = e^{-4\pi^2 t}\sin(2\pi x)$. Θεωρούμε έναν διαμερισμό του $[0,1]$ με $N = 23$, και M ισαπέχοντα σημεία του $[0,T]$, με $M = 24, 32, 128$, και υπολογίζουμε τη λύση για καθέναν από τους τρεις αυτούς διαμερισμούς χρησιμοποιώντας το αριθμητικό σχήμα (5.10). Για $t = 0.025, 0.05, 0.0625$ μπορούμε να βρούμε τις αντίστοιχες προσέγγισεις U_i^j της $u(x_i, t^j)$, με j τέτοιο ώστε $jk = 0.025, 0.05, 0.0625$ και $k = T/M$. Στο Σχήμα 5.3 απεικονίζονται οι γραφικές παραστάσεις της u για $t = 0$, καθώς και οι προσεγγίσεις για $t = 0.025, 0.05, 0.0625$, με $M = 24, 32, 128$. Επίσης, στον Πίνακα 5.1 βλέπουμε το μέγιστο σφάλμα στα χρονικά επίπεδα $t = 0.025, 0.05, 0.0625$. Στο Σχήμα 5.3 παρατηρούμε ότι η προσεγγιστική λύση για $M = 32$ δεν φαίνεται να προσεγγίζει την ακριβή λύση, καθώς το χρονικό επίπεδο t μεγαλώνει, το οποίο φαίνεται και στα αποτελέσματα του Πίνακα 5.1. Αυτό όμως δεν συμβαίνει στην περίπτωση $M = 128$. Αυτή η συμπεριφορά των προσεγγιστικών λύσεων οφείλεται στην αστάθεια της λύσης για $M = 32$ και της ευστάθειας για $M = 128$, διότι η παράμετρος λ είναι $2.4, 1.8, 0.45$ για $M = 24, 32, 128$, αντίστοιχα.



Σχήμα 5.3: Η ακριβής λύση $u(x,t) = e^{-4\pi^2 t}\sin(2\pi x)$ του Παραδείγματος 5.1 και οι προσεγγίσεις της για $N = 23$ και $M = 24, 32, 128$, με την άμεση μέθοδο του Euler για $t = 0, 0.025, 0.05$ και 0.0625.

5.2 Πεπλεγμένη μέθοδος του Euler

Σε αυτή την παράγραφο θα μελετήσουμε τη μέθοδο που προκύπτει αν αντί για την προσέγγιση, δ_k^+ της u_t που χρησιμοποιήσαμε στην άμεση μέθοδο του Euler, θεωρήσουμε την δ_k^-.

Έτσι, θα θεωρήσουμε και πάλι μια διαμέριση του $[0, L] \times [0, T]$ όπως και πριν και θα κατασκευάσουμε προσεγγίσεις U_j^i των τιμών $u(x_i, t^j)$, $i = 0, \ldots, N+1$, $j = 0, \ldots, M$, της ακριβούς λύσης του (5.1), βλέπε Σχήμα 5.1. Λόγω των αρχικών και συνοριακών συνθηκών του προβλήματος (5.1) θέτουμε και πάλι $U_0^j = U_{N+1}^j = 0$, $j = 0, \ldots, M$, και $U_i^0 = g(x_i)$, $i = 1, \ldots, N$. Στη συνέχεια, προσεγγίζουμε την $u_t(x_i, t^j)$ στις εξισώσεις (5.3) χρησιμοποιώντας τη διαφορά $\delta_k^- u(x_i, t^j)$ που θεωρήσαμε στην (2.1).

Έτσι, αν υποθέσουμε ότι $u_{tt}(x,t) \in C([0, L] \times [0, T])$, λόγω της (2.3), η (5.3) γίνεται για $i = 1, \ldots, N$,

$$\frac{u(x_i, t^j) - u(x_i, t^{j-1})}{k} = u_{xx}(x_i, t^j) + \hat{\eta}_i^j, \quad j = 1, \ldots, M, \tag{5.25}$$

όπου

$$|\hat{\eta}_i^j| \leq \frac{k}{2} \max_{t\in[0,T]} |u_{tt}(x_i,t)|. \tag{5.26}$$

Στη συνέχεια, για να προσεγγίσουμε την $u_{xx}(x_i,t^j)$, στην (5.25) χρησιμοποιούμε, όπως και στην άμεση μέθοδο του Euler, την $\delta_{h,2}^c u(x_i,t^j)$. Έτσι, αν υποθέσουμε ότι $\frac{\partial^4}{\partial x^4}u(x,t) \in C([0,L]\times[0,T])$, λόγω της (2.9), η (5.25) γίνεται για $i=1,\ldots,N$, και $j=1,\ldots,M$,

$$\begin{aligned}\frac{u(x_i,t^j)-u(x_i,t^{j-1})}{k} = &\frac{u(x_{i+1},t^j)-2u(x_i,t^j)+u(x_{i-1},t^j)}{h^2}\\ &+\hat{\eta}_i^j+\tilde{\eta}_i^j,\end{aligned} \tag{5.27}$$

όπου η $\tilde{\eta}_i^j$ φράσσεται και πάλι όπως στην (5.7).

Συνεπώς, για $\eta_i^j=\hat{\eta}_i^j+\tilde{\eta}_i^j$, η (5.27) δίνει για $i=1,\ldots,N$, και $j=1,\ldots,M$,

$$\delta_k^- u(x_i,t^j)-\delta_{h,2}^c u(x_i,t^j)=\eta_i^j, \tag{5.28}$$

και λόγω των (5.26) και (5.7) εύκολα βλέπουμε ότι ισχύει το ακόλουθο λήμμα.

Λήμμα 5.2. *Έστω u η λύση του* (5.1) *με $\frac{\partial^4}{\partial x^4}u \in C([0,L]\times[0,T])$ και $u_{tt}\in C([0,L]\times[0,T])$. Τότε, για την η_i^j που δίνεται στην* (5.28)*, έχουμε ότι υπάρχει σταθερά C ανεξάρτητη των k και h, τέτοια ώστε*

$$\max_{1\leq j\leq M}\max_{1\leq i\leq N}|\eta_i^j| \leq C(k+h^2). \tag{5.29}$$

Για να κατασκευάσουμε, λοιπόν, προσεγγίσεις U_i^j των $u(x_i,t^j)$, $i=1,\ldots,N$, $j=1,\ldots,M$, θεωρούμε την ακόλουθη μέθοδο

$$\frac{U_i^j-U_i^{j-1}}{k}=\frac{U_{i+1}^j-2U_i^j+U_{i-1}^j}{h^2},\quad i=1,\ldots,N,\ j=1,\ldots,M, \tag{5.30}$$

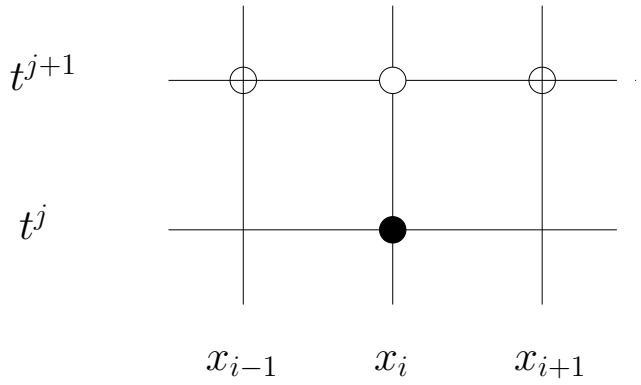
$$\begin{cases} U_0^j=U_{N+1}^j=0, \quad j=0,\ldots,M,\\ U_i^0=g(x_i), \quad i=1,\ldots,N.\end{cases} \tag{5.31}$$

Είναι φανερό ότι, λόγω της (5.28), το τοπικό σφάλμα διακριτοποίησης της μεθόδου (5.30)–(5.31), δίνεται από την η_i^j. Συνεπώς, σύμφωνα με το Λήμμα 5.2, το τοπικό σφάλμα διακριτοποίησης τείνει στο μηδέν, καθώς τα h και k τείνουν στο μηδέν, οπότε η (5.30) είναι μια *συνεπής* μέθοδος.

Συμβολίζουμε και πάλι με λ τον λόγο $\dfrac{k}{h^2}$, οπότε η (5.30) γράφεται ως

$$-\lambda U_{i+1}^j+(1+2\lambda)U_i^j-\lambda U_{i-1}^j=U_i^{j-1}, \tag{5.32}$$

με $i = 1, \ldots, N$ και $j = 1, \ldots, M$. Ξεκινώντας από το χρονικό επίπεδο $t^0 = 0$ και χρησιμοποιώντας τις αρχικές συνθήκες (5.31), παρατηρούμε ότι, σε αντίθεση με την άμεση μέθοδο του Euler (5.12), δεν μπορούμε άμεσα να υπολογίσουμε την προσέγγιση στο χρονικό επίπεδο t^1 από την (5.32). Στο Σχήμα 5.4 φαίνεται ένα παράδειγμα πλέγματος όπου με άσπρο κύκλο σημειώνουμε το σημείο του χρονικού επιπέδου t^{j+1}, όπου σύμφωνα με (5.32), αντιστοιχεί σε άγνωστες τιμές της προσέγγισης και με μαύρο αυτά του χρονικού επιπέδου t^j που αντιστοιχούν στη γνωστή τιμή της προσέγγισης.



Σχήμα 5.4: Σημεία ενός πλέγματος για τον υπολογισμό της λύσης με την πεπλεγμένη μέθοδο του Euler. Με άσπρο σημειώνουμε τα σημεία που αντιστοιχούν σε άγνωστες τιμές της προσέγγισης και με μαύρο αυτά που αντιστοιχούν σε γνωστές τιμές.

Αν θέσουμε $U^j \in \mathbb{R}^N$ το διάνυσμα με συνιστώσες $U_1^j, \ldots, U_N^j$, $U = (U_1^j, \ldots, U_N^j)^T$, μπορούμε να γράψουμε το σύστημα των εξισώσεων (5.32) ισοδύναμα ως το γραμμικό σύστημα

$$BU^j = U^{j-1}, \quad j = 1, \ldots, M, \text{ με } U^0 = G, \tag{5.33}$$

όπου B είναι ο $N \times N$ πίνακας

$$B = \begin{pmatrix} 1+2\lambda & -\lambda & 0 & & 0 \\ -\lambda & 1+2\lambda & -\lambda & 0 & \\ 0 & \ddots & \ddots & \ddots & 0 \\ & & -\lambda & 1+2\lambda & -\lambda \\ & & & -\lambda & 1+2\lambda \end{pmatrix},$$

και $G = (g(x_1), \ldots, g(x_N))^T$.

Αφού γνωρίζουμε τη U^0 για να υπολογίσουμε τώρα την προσέγγιση στο χρονικό επίπεδο t^1 χρειάζεται να λύσουμε το γραμμικό σύστημα $BU^1 = U^0$. Είναι φανερό ότι ο B έχει αυστηρά κυριαρχική διαγώνιο και, άρα, είναι αντιστρέψιμος. Επομένως, το γραμμικό σύστημα $BU^1 = U^0$ έχει μοναδική λύση. Επειδή ο πίνακας B είναι τριδιαγώνιος για τον υπολογισμό της U^1, μπορούμε να χρησιμοποιήσουμε τη μέθοδο της Παραγράφου 3.2. Συνεχίζουμε με αυτό τον τρόπο, οπότε σε κάθε χρονικό βήμα γνωρίζουμε την προσέγγιση στο χρονικό επίπεδο t^{j-1} και για να υπολογίσουμε την προσέγγιση στο χρονικό επίπεδο t^j λύνουμε το γραμμικό σύστημα (5.33). Επειδή, κάθε φορά χρειάζεται να λύσουμε ένα γραμμικό σύστημα, η μέθοδος καλείται *πεπλεγμένη ή έμμεση*. Συγκεκριμένα, τη μέθοδο (5.30) ή (5.32) την ονομάζουμε *πεπλεγμένη μέθοδο του Euler*.

Ο πίνακας B στο γραμμικό σύστημα (5.33) είναι τριδιαγώνιος και έχει αυστηρά κυριαρχική διαγώνιο, οπότε είναι αντιστρέψιμος και άρα η ύπαρξη και η μοναδικότητα της προσεγγιστικής λύσης σε κάθε χρονικό βήμα είναι δεδομένη.

Για την ευστάθεια της μεθόδου μπορούμε να δείξουμε το ακόλουθο θεώρημα.

Θεώρημα 5.3. *Έστω $U^j \in \mathbb{R}^N$, $j = 0, \ldots, M$, τα διανύσματα που ικανοποιούν την* (5.33)*, και $U_0^j = U_{N+1}^j = 0$, $j = 0, \ldots, M$. Τότε, ισχύει η ακόλουθη ανισότητα*

$$\max_{0\le i\le N+1} |U_i^j| \le \max_{0\le i\le N+1} |U_i^0|, \quad \textit{για } j = 0, 1, 2, \ldots M.$$

Απόδειξη. Από τη σχέση (5.32) εύκολα παίρνουμε

$$(1+2\lambda)|U_i^j| \le \lambda(|U_{i+1}^j| + |U_{i-1}^j|) + |U_i^{j-1}|, \quad 1 \le i \le N, j = 1, \ldots, M.$$

Αν θέσουμε λοιπόν $\bar{U}^j = \max_{0\le i\le N+1} |U_i^j|$, $j = 0, \ldots, M$, παίρνουμε

$$(1+2\lambda)|U_i^j| \le 2\lambda\bar{U}^j + \bar{U}^{j-1}, \quad j = 1, \ldots, M,$$

και άρα

$$\bar{U}^j \le \bar{U}^{j-1} \le \cdots \le \bar{U}^0,$$

από όπου προκύπτει η ζητούμενη εκτίμηση. □

Στη συνέχεια, θα μελετήσουμε την *ευστάθεια von Neumann* για την πεπλεγμένη μέθοδο του Euler. Όπως και στην άμεση μέθοδο του Euler, υποθέτουμε ότι οι

$$U_i^j = w_j e^{rx_i I}, \quad r \in \mathbb{R},$$

ικανοποιούν την (5.32). Οπότε

$$-w_j(\lambda e^{rx_{i+1}I} - (1+2\lambda)e^{rx_iI} + \lambda e^{rx_{i-1}I}) = w_{j-1}e^{rx_iI}. \tag{5.34}$$

Χρησιμοποιούμε τώρα και πάλι τις ιδιότητες (5.23) και (5.20) και έχουμε

$$\begin{aligned} -w_j(2\lambda\cos(rh)-(1+2\lambda)) &= w_j(1+4\lambda\sin^2(\frac{rh}{2})) \\ &= w_{j-1}. \end{aligned} \tag{5.35}$$

Συνεπώς, $w_j = \kappa^j w_0$, με $\kappa = 1/(1+4\lambda\sin^2(\frac{rh}{2}))$. Άρα, επειδή για κάθε λ και r, ισχύει $|\kappa| \leq 1$, η απόλυτη τιμή της U_i^j είναι φραγμένη, για κάθε σημείο του διαμερισμού, και άρα δεν είναι αναγκαία καμία συνθήκη για την ευστάθεια της πεπλεγμένης Euler.

Στη συνέχεια δείχνουμε το ακόλουθο θεώρημα για τη σύγκλιση της μεθόδου.

Θεώρημα 5.4. *Έστω ότι η λύση u του προβλήματος* (5.1) *είναι αρκετά ομαλή και U_i^j, $i = 0,\ldots,N+1$, $j = 0,\ldots,M$, η λύση της μεθόδου* (5.32)*. Τότε, υπάρχει μια σταθερά C, ανεξάρτητη των k, h, τέτοια ώστε*

$$\max_{0\leq j\leq M}\max_{0\leq i\leq N+1}|U_i^j - u(x_i,t^j)| \leq C(k+h^2). \tag{5.36}$$

Απόδειξη. Θέτουμε $E_i^j = U_i^j - u(x_i,t^j)$, $i = 0,\ldots,N+1$, $j = 0,\ldots,M$, όπου λόγω των σχέσεων $U_0^j = u(0,t^j) = 0$ και $U_{N+1}^j = u(L,t^j) = 0$, έχουμε $E_0^j = E_{N+1}^j = 0$. Αφαιρούμε τώρα κατά μέλη τις (5.32) και (5.27), οπότε παίρνουμε

$$-\lambda E_i^j + (1+2\lambda)E_i^j - \lambda E_{i-1}^j = E_i^{j-1} + k\eta_i^j, \quad i = 1,\ldots,N,\ j = 1,\ldots,M.$$

Θέτουμε, στη συνέχεια, $\bar{E}^j = \max_{1\leq i\leq N}|E_i^j|$, $\bar{\eta} = \max_{1\leq i\leq N,1\leq j\leq M}|\eta_i^j|$ και έχουμε

$$(1+2\lambda)E_i^j \leq 2\lambda\bar{E}^j + \bar{E}^{j-1} + k\bar{\eta},$$

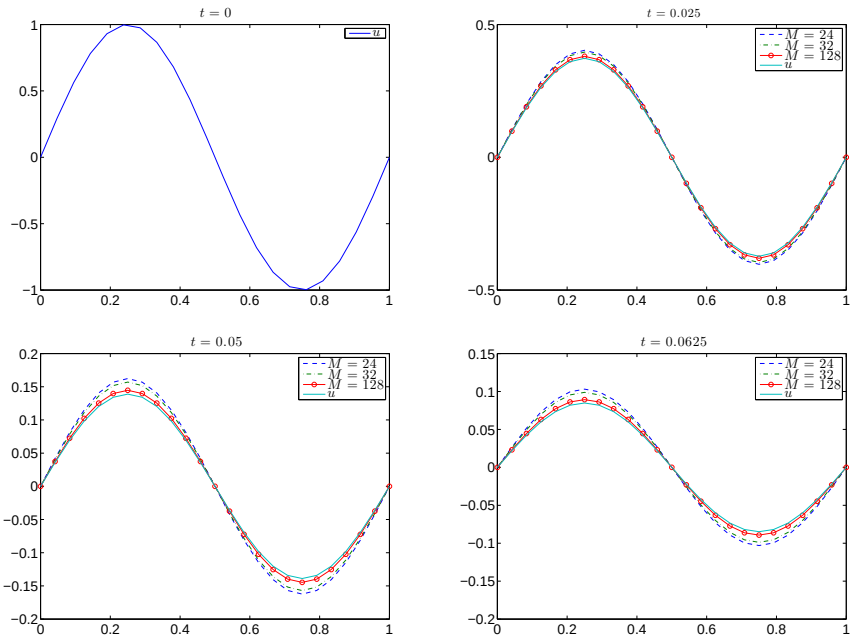
οπότε

$$\bar{E}^j \leq \bar{E}^{j-1} + k\bar{\eta} \leq \cdots \leq \bar{E}^0 + jk\bar{\eta}.$$

Από εδώ, επειδή $jk \leq T$ και λόγω του Λήμματος 5.2, προκύπτει η ζητούμενη εκτίμηση. □

Παράδειγμα 5.2. Αν στο Παράδειγμα 5.1 χρησιμοποιήσουμε την πεπλεγμένη μέθοδο του Euler, παρατηρούμε ότι οι αντίστοιχες προσεγγίσεις U_i^j της $u(x_i,t^j)$ είναι καλύτερες. Στο Σχήμα 5.5 απεικονίζονται οι γραφικές παραστάσεις της u για $t = 0$, καθώς και οι προσεγγίσεις για $t = 0.025, 0.05, 0.0625$, με $M = 24, 32, 128$. Επίσης, στον Πίνακα 5.2 βλέπουμε το μέγιστο σφάλμα στα χρονικά επίπεδα $t = 0.025, 0.05, 0.0625, 0.075$. Στο Σχήμα 5.5 παρατηρούμε ότι όλες οι προσεγγιστικές λύσεις φαίνεται να προσεγγίζουν την ακριβή λύση, καθώς το χρονικό επίπεδο

t μεγαλώνει, το οποίο φαίνεται και στα αποτελέσματα του Πίνακα 5.1. Αυτή η συμπεριφορά των προσεγγιστικών λύσεων οφείλεται στην ευστάθεια της μεθόδου για όλες τις τιμές της παράμετρου λ, σε αντίθεση με τη συμπεριφορά της άμεσης μεθόδου του Euler, βλ. Παράδειγμα 5.1.



Σχήμα 5.5: Η ακριβής λύση $u(x,t) = e^{-4\pi^2 t}\sin(2\pi x)$ του Παραδείγματος 5.1 και οι προσεγγίσεις της για $N = 23$ και $M = 24, 32, 128$, με την πεπλεγμένη μέθοδο του Euler για $t = 0, 0.025, 0.05$ και 0.0625.

	$t = 0.025$		$t = 0.05$		$t = 0.0625$		$t = 0.075$	
M	j	$\bar{E}^j$	j	$\bar{E}^j$	j	$\bar{E}^j$	j	$\bar{E}^j$
24	6	0.0303	12	0.0235	15	0.0183	18	0.0137
32	8	0.0236	16	0.0181	20	0.0140	24	0.0105
128	32	0.0077	64	0.0058	80	0.0044	96	0.0033

Πίνακας 5.2: Το σφάλμα $\bar{E}^j = \max_{1\le i\le N} |U_i^j - u(x_i, t^j)|$ στα χρονικά επίπεδα $t = 0.025, 0.05$ και 0.0625 του Παραδείγματος 5.2 για $N = 23$ και $M = 24, 32, 128$. Στη στήλη αριστερά του κάθε σφάλματος δίνεται το χρονικό βήμα j, τέτοιο ώστε $t = jk$, με $k = 1/M$.

5.3 Μέθοδος των Crank-Nicolson

Μια άλλη μέθοδος προκύπτει αν αντί για τις προσεγγίσεις δ_k^+ ή δ_k^- της u_t που θεωρήσαμε προηγουμένως, θεωρήσουμε τη δ_k^c. Όπως και πριν θα θέσουμε $U_0^j = U_{N+1}^j = 0$ για $j = 0, \ldots, M$, και $U_i^0 = g(x_i)$ και θα κατασκευάσουμε τώρα προσεγγίσεις U_i^j, $i = 1, \ldots, N$, $j = 1, \ldots, M$, των τιμών $u(x_i, t^j)$ της ακριβούς λύσης του (5.1), μεγαλύτερης τάξης ακρίβειας.

Λόγω του Λήμματος 2.1 και της εξίσωσης (5.1) έχουμε για $x = 1, \ldots, N$,

$$\begin{aligned} \frac{u(x_i, t^j) - u(x_i, t^{j-1})}{k} &= u_t(x_i, t^j - \frac{k}{2}) + \hat{\eta}_i^j \\ &= u_{xx}(x_i, t^j - \frac{k}{2}) + \hat{\eta}_i^j, \quad j = 1, \ldots, M, \end{aligned} \tag{5.37}$$

όπου

$$|\hat{\eta}_i^j| \le \frac{k^2}{6} \max_{t\in[0,T]} |u_{ttt}(x_i, t)| \tag{5.38}$$

Επειδή τα σημεία $(x_i, t^j - \frac{k}{2})$ δεν αποτελούν μέρος του διαμερισμού που έχουμε θεωρήσει, προσεγγίζουμε την $u_{xx}(x_i, t^j - \frac{k}{2})$, $i = 1, \ldots, N$, $j = 1, \ldots, M$, στην (5.37) με τον μέσο όρο της u_{xx} στα (x_i, t^j) και (x_i, t^{j-1}). Έτσι, είναι απλό να δούμε με τη βοήθεια αναπτυγμάτων Taylor ότι

$$u_{xx}(x_i, t^j - \frac{k}{2}) = \frac{1}{2}(u_{xx}(x_i, t^j) + u_{xx}(x_i, t^{j-1})) - \frac{k^2}{2} u_{xxtt}(x_i, \xi^j), \tag{5.39}$$

με $\xi^j \in (t^{j-1}, t^j)$. Επομένως, συνδυάζοντας αυτή τη σχέση με την (5.37), έχουμε για $j = 1, \ldots, M$,

$$\frac{u(x_i, t^j) - u(x_i, t^{j-1})}{k} = \frac{1}{2}(u_{xx}(x_i, t^j) + u_{xx}(x_i, t^{j-1})) + \bar{\eta}_i^j + \hat{\eta}_i^j, \tag{5.40}$$

με

$$|\bar{\eta}_i^j| \leq \frac{k^2}{6} \max_{t\in[0,T]} |u_{xxtt}(x_i,t)|. \tag{5.41}$$

Στη συνέχεια, για την προσέγγιση των $u_{xx}(x_i,t^j)$, $j=1,\ldots,M$, στην (5.40) εφαρμόζουμε τη $\delta^c_{h,2}u(x_i,t^j)$ που θεωρήσαμε στην (2.8). Έτσι, αν υποθέσουμε ότι $\frac{\partial^4}{\partial x^4}u(x,t)\in C([a,b]\times[0,T])$, λόγω της (2.9), η (5.40) γίνεται για $i=1,\ldots,N$, και $j=1,\ldots,M$,

$$\begin{aligned}\frac{u(x_i,t^j)-u(x_i,t^{j-1})}{k} &= \frac{1}{2}\frac{u(x_{i+1},t^j)-2u(x_i,t^j)+u(x_{i-1},t^j)}{h^2}\\ &+\frac{1}{2}\frac{u(x_{i+1},t^{j-1})-2u(x_i,t^{j-1})+u(x_{i-1},t^{j-1})}{h^2}\\ &+\hat{\eta}_i^j+\bar{\eta}_i^j+\tilde{\eta}_i^j,\end{aligned} \tag{5.42}$$

όπου η $\tilde{\eta}_i^j$ φράσσεται όπως στην (5.7). Συνεπώς, για $\eta_i^j=\hat{\eta}_i^j+\bar{\eta}_i^j+\tilde{\eta}_i^j$, η (5.42) δίνει για $i=1,\ldots,N$, και $j=1,\ldots,M$,

$$\delta^c_h u(x_i,t^j-\frac{k}{2})-\frac{1}{2}(\delta^c_{h,2}u(x_i,t^j)+\delta^c_{h,2}u(x_i,t^{j-1}))=\eta_i^j, \tag{5.43}$$

και λόγω των (5.38), (5.41) και (5.7) εύκολα βλέπουμε ότι ισχύει το ακόλουθο λήμμα.

Λήμμα 5.3. *Έστω u η λύση του* (5.1) *με $\frac{\partial^4}{\partial x^4}u, u_{tt}, u_{ttxx}\in C([0,L]\times[0,T])$. Τότε για την η_i^j που δίνεται στην* (5.43) *έχουμε ότι υπάρχει σταθερά C ανεξάρτητη των k και h, τέτοια ώστε*

$$\max_{1\leq j\leq M}\max_{1\leq i\leq N}|\eta_i^j|\leq C(k^2+h^2). \tag{5.44}$$

Για να κατασκευάσουμε, λοιπόν, προσεγγίσεις U_i^j των $u(x_i,t^j)$, $i=1,\ldots,N$, $j=0,\ldots,M$, θεωρούμε την ακόλουθη μέθοδο

$$\frac{U_i^j-U_i^{j-1}}{k} = \frac{1}{2}\frac{U_{i+1}^j-2U_i^j+U_{i-1}^j}{h^2}+\frac{1}{2}\frac{U_{i+1}^{j-1}-2U_i^{j-1}+U_{i-1}^{j-1}}{h^2}, \tag{5.45}$$
$$\text{για}\quad i=1,\ldots,N,\ j=1,\ldots,M,$$

$$\begin{cases} U_0^j=U_{N+1}^j=0, \quad j=0,\ldots,M,\\ U_i^0=g(x_i), \quad i=1,\ldots,N.\end{cases} \tag{5.46}$$

Το τοπικό σφάλμα διακριτοποίησης η_i^j της (5.45)–(5.46) δίνεται από την (5.43). Επίσης από το Λήμμα 5.3–(5.46) έχουμε ότι η η_i^j τείνει στο μηδέν, καθώς k και h τείνουν στο μηδέν. Επομένως, η μέθοδος (5.45)–(5.46) είναι *συνεπής*.

Αν συμβολίσουμε και πάλι με λ τον λόγο $\dfrac{k}{h^2}$, η (5.45) γράφεται ως

$$-\frac{1}{2}\lambda U_{i+1}^j+(1+\lambda)U_i^j-\frac{1}{2}\lambda U_{i-1}^j=\frac{1}{2}\lambda U_{i+1}^{j-1}+(1-\lambda)U_i^{j-1}+\frac{1}{2}\lambda U_{i-1}^{j-1}, \quad (5.47)$$

με $i=1,\ldots,N$, $j=1,\ldots,M$. Όμοια όπως και στην πεπλεγμένη μέθοδο του Euler, (5.32), αν ξεκινήσουμε από το χρονικό επίπεδο $t^0=0$, δεν μπορούμε άμεσα να υπολογίσουμε την προσέγγιση στο χρονικό επίπεδο t^1 από την (5.47). Στο Σχήμα 5.6 φαίνεται ένα παράδειγμα πλέγματος όπου σημειώνουμε τις γνωστές ή άγνωστες τιμές της προσέγγισης σε κάθε βήμα με τη μέθοδο (5.47).

Επομένως, αν $U^j\in\mathbb{R}^N$ το διάνυσμα με συνιστώσες $U_1^j,\ldots,U_N^j$, $U=(U_1^j,\ldots,U_N^j)^T$, το σύστημα των εξισώσεων (5.47) μπορούμε να το γράψουμε ως

$$BU^j=AU^{j-1},\quad j=1,\ldots,M,\qquad \text{με } U^0=G, \quad (5.48)$$

όπου B και A είναι οι $N\times N$ πίνακες

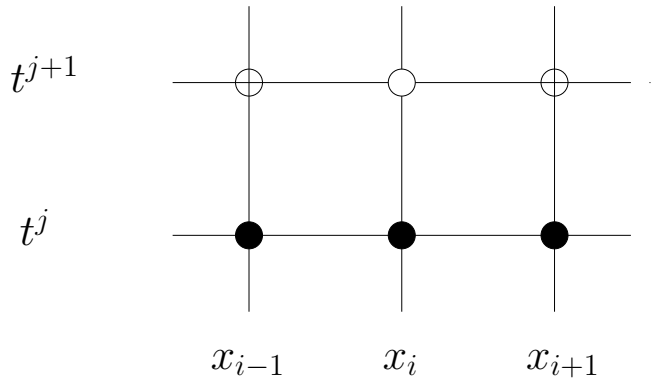
$$B=\begin{pmatrix} 1+\lambda & -\lambda/2 & 0 & & 0\\ -\lambda/2 & 1+\lambda & -\lambda/2 & 0 & \\ 0 & \ddots & \ddots & \ddots & 0\\ & & -\lambda/2 & 1+\lambda & -\lambda/2\\ & & & -\lambda/2 & 1+\lambda \end{pmatrix},$$

$$A=\begin{pmatrix} 1-\lambda & \lambda/2 & 0 & & 0\\ \lambda/2 & 1-\lambda & \lambda/2 & 0 & \\ 0 & \ddots & \ddots & \ddots & 0\\ & & \lambda/2 & 1-\lambda & \lambda/2\\ & & & \lambda/2 & 1-\lambda \end{pmatrix},$$

και $G=(g(x_1),\ldots,g(x_N))^T$.

Από την (5.45) έχουμε ότι η προσέγγιση U^0 στο χρονικό επίπεδο t^0 είναι γνωστή, οπότε για να υπολογίσουμε τη λύση στο χρονικό επίπεδο t^1 αρκεί να λύσουμε το γραμμικό σύστημα $BU^1=AU^0$. Εύκολα βλέπουμε ότι ο B έχει αυστηρά κυριαρχική διαγώνιο, επομένως είναι αντιστρέψιμος και επειδή είναι τριδιαγώνιος μπορούμε να εφαρμόσουμε τον αλγόριθμο της Παραγράφου 3.2 για τον υπολογισμό της U^1. Συνεχίζουμε με αυτό τον τρόπο και αν στο χρονικό επίπεδο t^{j-1} γνωρίζουμε την προσέγγιση U^{j-1}, λύνοντας το γραμμικό σύστημα (5.48), βρίσκουμε την προσέγγιση U^j στο χρονικό επίπεδο t^j. Για τον ίδιο λόγο, όπως και για την πεπλεγμένη μέθοδο του Euler, η μέθοδος (5.45) είναι *πεπλεγμένη ή έμμεση* και καλείται μέθοδος των Crank–Nicolson.

Για την ευστάθεια της μεθόδου μπορούμε να δείξουμε το ακόλουθο θεώρημα.



Σχήμα 5.6: Σημεία ενός πλέγματος για τον υπολογισμό της λύσης με τη μέθοδο των Crank-Nicolson. Με άσπρο σημειώνουμε τα σημεία που αντιστοιχούν σε άγνωστες τιμές της προσέγγισης και με μαύρο αυτά που αντιστοιχούν σε γνωστές τιμές.

Θεώρημα 5.5. *Έστω $U^j \in \mathbb{R}^N$, $j = 0, \ldots, M$, η λύση του προβλήματος* (5.47)*, με $U_0^j = U_{N+1}^j = 0$, $j = 0, \ldots, M$. Τότε, αν $\lambda \leq 1$, ισχύει η ακόλουθη ανισότητα*

$$\max_{0\leq i\leq N+1} |U_i^j| \leq \max_{0\leq i\leq N+1} |U_i^0|, \quad \text{για } j = 0, 1, 2, \ldots, M. \tag{5.49}$$

Απόδειξη. Από τη σχέση (5.47) εύκολα παίρνουμε για $i = 1, \ldots, N$, $j = 1, \ldots, M$,

$$(1+\lambda)|U_i^j| \leq \frac{\lambda}{2}(|U_{i+1}^j| + |U_{i-1}^j|) + (1-\lambda)|U_i^{j-1}| + \frac{\lambda}{2}(|U_{i+1}^{j-1}| + |U_{i-1}^{j-1}|).$$

Αν θέσουμε λοιπόν $\bar{U}^j = \max_{0\leq i\leq N+1} |U_i^j|$, $j = 0, \ldots, M$, έχουμε

$$(1+\lambda)|U_i^j| \leq \lambda\bar{U}^j + \bar{U}^{j-1},$$

και άρα

$$\bar{U}^j \leq \bar{U}^{j-1} \leq \cdots \leq \bar{U}^0,$$

από όπου προκύπτει η ζητούμενη ανισότητα. □

Παρατηρούμε ότι για να δείξουμε το Θεώρημα 5.5, υποθέσαμε ότι $\lambda \leq 1$. Θα δούμε παρακάτω ότι αυτή η υπόθεση δεν είναι απαραίτητη για να δείξουμε την *ευστάθεια von Neumann* για τη μέθοδο των Crank-Nicolson, επομένως η υπόθεση $\lambda \leq 1$ στο Θεώρημα 5.5 δεν είναι ουσιαστική και γίνεται μόνο για να μπορέσουμε να αποδείξουμε το θεώρημα, ακολουθώντας τα βήματα που χρησιμοποιήσαμε στην αντίστοιχη απόδειξη.

Θεωρούμε και πάλι ότι οι $U_i^j = w_j e^{rx_i I}$ ικανοποιούν την (5.47), οπότε

$$\begin{aligned} &-\frac{\lambda}{2}w_j e^{rx_{i+1}I} + (1+\lambda)w_j e^{rx_i I} - \frac{\lambda}{2}w_j e^{rx_{i-1}I} \\ &= \frac{\lambda}{2}w_{j-1}e^{rx_{i+1}I} + (1-\lambda)w_{j-1}e^{rx_i I} + \frac{\lambda}{2}w_{j-1}e^{rx_{i-1}I}, \end{aligned} \tag{5.50}$$

για $i = 1, \ldots, N, j = 1, \ldots, M$. Χρησιμοποιούμε τώρα και πάλι τις ιδιότητες (5.23) και (5.20) και έχουμε

$$-2\lambda w_j e^{rx_i I}\cos(rh) + (1+\lambda)w_j e^{rx_i I} = 2\lambda w_{j-1}e^{rx_i I}\cos(rh) + (1-\lambda)w_{j-1}e^{rx_i I},$$

από την οποία προκύπτει

$$(1 + 4\lambda \sin^2(\frac{rh}{2}))w_j = (1 - 4\lambda\sin^2(\frac{rh}{2}))w_{j-1}.$$

Συνεπώς

$$w_j = \frac{1 - 4\lambda\sin^2(\frac{rh}{2})}{1 + 4\lambda\sin^2(\frac{rh}{2})}w_{j-1} = \kappa w_{j-1} = \cdots = \kappa^j w_0. \tag{5.51}$$

Είναι απλό να δούμε ότι σε αυτήν την περίπτωση ισχύει $|\kappa| \leq 1$ για κάθε λ και r. Συνεπώς, η απόλυτη τιμή της U_i^j είναι φραγμένη ανεξάρτητα της τιμής του λ, οπότε δεν είναι αναγκαία καμιά συνθήκη για στις παραμέτρους της διαμέρισης h και k, ώστε να είναι ευσταθής η μέθοδος Crank–Nicolson.

Θεώρημα 5.6. *Έστω ότι η λύση u του προβλήματος* (5.1) *είναι αρκετά ομαλή και U_i^j, $i = 0, \ldots, N+1$, $j = 0, \ldots, M$, η λύση της μεθόδου* (5.47)*. Τότε, υπάρχει μια σταθερά C, ανεξάρτητη των k, h, τέτοια ώστε*

$$\max_{0\leq j\leq M}\max_{0\leq i\leq N+1}|U_i^j - u(x_i, t^j)| \leq C(k^2 + h^2). \tag{5.52}$$

Απόδειξη. Θέτουμε $E_i^j = U_i^j - u(x_i, t^j)$, $i = 0, \ldots, N+1$, $j = 0, \ldots, M$, όπου λόγω των σχέσεων $U_0^j = u(0, t^j) = 0$ και $U_{N+1}^j = u(L, t^j) = 0$, έχουμε $E_0^j = E_{N+1}^j = 0$. Αφαιρούμε τώρα κατά μέλη τις (5.45) και (5.42), οπότε παίρνουμε για $i = 1, \ldots, N$, $j = 1, \ldots, M$,

$$(1+\lambda)E_i^j = \frac{\lambda}{2}(E_{i+1}^j + E_{i-1}^j) + (1-\lambda)E_i^{j-1} + \frac{\lambda}{2}(E_{i+1}^{j-1} + E_{i-1}^{j-1}) + k\eta_i^j.$$

Θέτουμε, στη συνέχεια, $\bar{E}^j = \max_{1\leq i\leq N}|E_i^j|$, $\bar{\eta} = \max_{1\leq i\leq N, 1\leq j\leq M}|\eta_i^j|$ και έχουμε

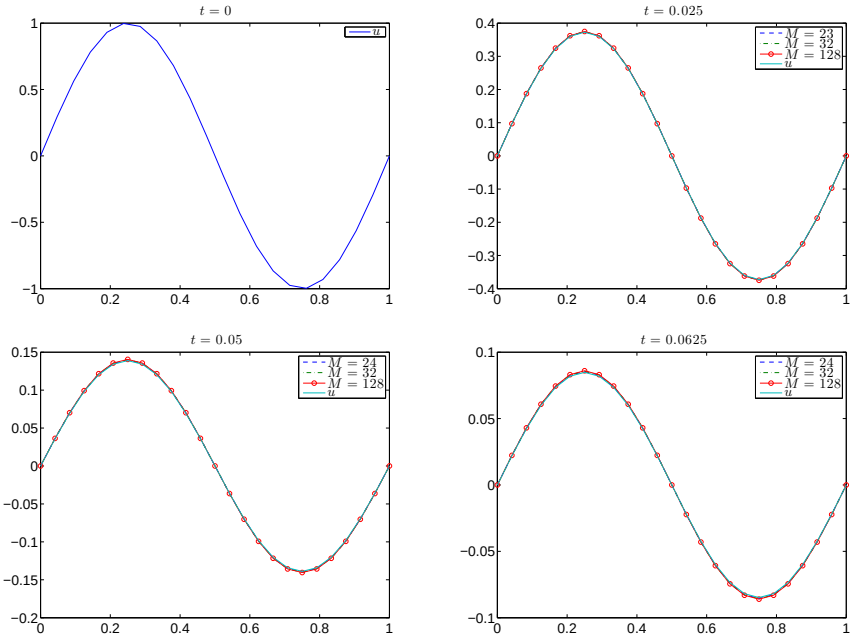
$$(1+\lambda)E_i^j \leq \lambda\bar{E}^j + (1-\lambda)\bar{E}^{j-1} + \lambda\bar{E}^{j-1} + k\bar{\eta},$$

οπότε

$$\bar{E}^j \leq \bar{E}^{j-1} + k\bar{\eta} \leq \cdots \leq \bar{E}^0 + jk\bar{\eta}.$$

Από εδώ, επειδή $jk \leq T$ και λόγω του Λήμματος 5.3, προκύπτει η ζητούμενη εκτίμηση. □

Παράδειγμα 5.3. Επαναλαμβάνουμε και εδώ το Παράδειγμα 5.1 όπου χρησιμοποιούμε τώρα τη μέθοδο των Crank-Nicolson. Στο Σχήμα 5.7 απεικονίζονται οι γραφικές παραστάσεις της u για $t = 0$, καθώς και οι προσεγγίσεις για $t = 0.025$, 0.05, 0.0625, με $M = 24, 32, 128$. Επίσης, στον Πίνακα 5.3 βλέπουμε το μέγιστο σφάλμα στα χρονικά επίπεδα $t = 0.025, 0.05, 0.0625, 0.075$.



Σχήμα 5.7: Η ακριβής λύση $u(x,t) = e^{-4\pi^2 t}\sin(2\pi x)$ του Παραδείγματος 5.1 και οι προσεγγίσεις της για $N = 23$ και $M = 24, 32, 128$, με τη μέθοδο των Crank-Nicolson για $t = 0, 0.025, 0.05$ και 0.0625.

Παρατηρούμε από τα Παραδείγματα 5.2 και 5.3, ότι το σφάλμα σε κάθε χρονικό επίπεδο είναι μικρότερο στη μέθοδο Crank–Nicolson από ότι στη πεπλεγμένη

	$t = 0.025$		$t = 0.05$		$t = 0.0625$		$t = 0.075$	
M	j	$\bar{E}^j$	j	$\bar{E}^j$	j	$\bar{E}^j$	j	$\bar{E}^j$
24	6	0.0013	12	0.0010	15	0.0007	18	0.0005
32	8	0.0016	16	0.0012	20	0.0008	24	0.0007
128	32	0.0021	64	0.0015	80	0.0012	96	0.0009

Πίνακας 5.3: Το σφάλμα $\bar{E}^j = \max_{1\le i\le N} |U_i^j - u(x_i, t^j)|$ στα χρονικά επίπεδα $t = 0.025, 0.05$ και 0.0625 0.075 του Παραδείγματος 5.3 για $N = 23$ και $M = 24, 32, 128$. Στη στήλη αριστερά του κάθε σφάλματος δίνεται το χρονικό βήμα j, τέτοιο ώστε $t = jk$, με $k = 1/M$.

μέθοδο του Euler. Αυτό οφείλεται στη μεγαλύτερη τάξη ως προς k που έχει το σφάλμα για τη μέθοδο Crank–Nicolson από το αντίστοιχο σφάλμα για τη πεπλεγμένη μέθοδο του Euler, βλ. Θεωρήματα 5.4 και 5.6. Στο επόμενο παράδειγμα παραθέτουμε ορισμένα αριθμητικά αποτελέσματα, όπου φαίνεται ότι τα σφάλματα που προκύπτουν με τη μέθοδο Crank–Nicolson έχουν τάξη ακρίβειας k^2.

Παράδειγμα 5.4. Αν οι παράμετροι k, h της διαμέρισης των $[0, L] \times [0, T]$ είναι ίσες, τότε το σφάλμα για την πεπλεγμένη μέθοδο του Euler σύμφωνα με το Θεώρημα 5.4, θα ικανοποιεί

$$\max_{0\le j\le M} \max_{0\le i\le N+1} |U_i^j - u(x_i, t^j)| \le Ck, \quad \text{για } k = h, \text{ και } k \text{ μικρό},$$

και το σφάλμα για τη μέθοδο των Crank–Nicolson σύμφωνα με το Θεώρημα 5.6, θα ικανοποιεί

$$\max_{0\le j\le M} \max_{0\le i\le N+1} |U_i^j - u(x_i, t^j)| \le Ck^2, \quad \text{για } k = h, \text{ και } k \text{ μικρό}.$$

Αν υποθέσουμε λοιπόν ότι το μέγιστο σφάλμα $\bar{E} = \max_{0\le i\le N+1} |U_i^M - u(x_i, t^M)| \approx Ck^p$, στο χρονικό επίπεδο $t^M = T$, μπορούμε να προσδιορίσουμε πειραματικά, με τη χρήση H/Y, τη δύναμη p, από τον λόγο $p \approx (\log(E_1/E_2)/\log(k_1/k_2))$, όπου E_1 είναι το σφάλμα που αντιστοιχεί στη διαμέριση με βήμα k_1 και E_2 το σφάλμα για τη διαμέριση με βήμα k_2.

Αν θεωρήσουμε και πάλι το πρόβλημα των Παραδείγματων 5.1 και 5.2, μπορούμε να υπολογίσουμε κατά προσέγγιση την τάξη σύγκλισης για την πεπλεγμένη μέθοδο του Euler και τη μέθοδο των Crank–Nicolson για $t = T = 0.1$. Στον Πίνακα 5.4, εμφανίζεται η κατά προσέγγιση δύναμη p και παρατηρούμε ότι για τη μέθοδο Euler είναι κατά προσέγγιση ένα, ενώ για τη μέθοδο Crank–Nicolson είναι δύο.

	BE		CN	
$k = h$	$\bar{E}_{BE}$	p	$\bar{E}_{CN}$	p
1/20	0.095012		0.01918	
1/40	0.045123	1.074	0.00592	1.69539
1/80	0.021160	1.092	0.00150	1.97976
1/160	0.010064	1.072	0.00037	1.99623
1/250	0.004876	1.045	0.00009	1.99912

Πίνακας 5.4: Τα σφάλματα των μεθόδων Euler και Crank–Nicolson στο χρονικό επίπεδο $t = T = 0.1$, $\bar{E}_{BE}$ και $\bar{E}_{CN}$, αντίστοιχα, στο Παράδειγμα 5.4 και η κατά προσέγγιση τάξη ακρίβειάς τους p, αν θεωρήσουμε $k = h$.

5.4 Άλλες μέθοδοι και προβλήματα

Ακολουθώντας την ανάλυση προβλημάτων με συνοριακές συνθήκες τύπου Neumann που είδαμε στο Κεφάλαιο 3, μπορούμε να τροποποιήσουμε ανάλογα τις μεθόδους Euler και Crank–Nicolson, ώστε να ισχύουν αντίστοιχα αποτελέσματα με αυτά που παρουσιάσαμε σε αυτό το κεφάλαιο. Επίσης, για την απόδειξη των διαφόρων θεωρημάτων μπορούμε να χρησιμοποιήσουμε τη μέθοδο της ενέργειας που ακολουθήσαμε στη Παράγραφο 3.4. Δεν θα παρουσιάσουμε όμως αυτά τα αποτελέσματα σε αυτές τις σημειώσεις και παραπέμπουμε τον αναγνώστη στα (Ακρίβης & Δουγαλής, 2005· Holmes, 2007· Iserles, 2009· Larsson & Thomée, 2009· Morton & Mayers, 2005· Strikwerda, 2004· Thomas, 1995) για πιο λεπτομερή παρουσίαση των αποτελεσμάτων αυτού του κεφαλαίου καθώς και άλλων.

5.5 Ταινίες γραφικών παραστάσεων

Σε αυτή την παράγραφο εμφανίζονται τρεις ταινίες με τις γραφικές παραστάσεις που παρουσιάστηκαν στο Παραδείγμα 5.1 για τις μεθόδους Άμεση Euler (Forward Euler), Πεπλεγμένη Euler (Backward Euler) και Crank-Nicolson. Η προβολή των ταινιών στην οθόνη του H/Y μπορεί να γίνει, αν μετακινήσουμε τον κέρσορα (δείκτη) της οθόνης και επιλέξουμε, π.χ. με τη χρήση του "ποντίκιου", το αντίστοιχο παράδειγμα και, στη συνέχεια, την επιλογή "Play/Pause". Στην περίπτωση που η προβολή αυτού του βιβλίου στην οθόνη γίνεται μέσω του αντίστοιχου αρχείου μορφής pdf, συνίσταται η χρήση του πρόγραμματος Adobe Reader.

Play/Pause | Forward Euler | Backward Euler | Crank-Nicolson

Σχήμα 5.8: Ταινίες των γραφικών παραστάσεων των Παραδείγματων 5.1–5.3 για την άμεση και πεπλεγμένη μέθοδο του Euler και τη μέθοδο των Crank-Nicolson.

5.6 Ασκήσεις

5.1. Θεωρούμε τη μέθοδο $U_i^{j+1} = \lambda U_{i+1}^j + (1 - 2\lambda - \beta k)U_i^j + \lambda U_{i-1}^j$ για την επίλυση της εξίσωσης της θερμότητας (5.1).

(α΄) Είναι η μέθοδος άμεση ή πεπλεγμένη;

(β΄) Είναι η μέθοδος ευσταθής;

(γ΄) Για ποιες τιμές του β, αν υπάρχουν, είναι η μέθοδος συνεπής;

5.2. Θεωρούμε την εξίσωση της θερμότητας (5.1) όπου έχουμε τροποποιήσει τις συνοριακές συνθήκες ως εξής, $u_x(0,t) = \alpha$, και $u(1,t) = 0$.

(α΄) Γράψτε ένα πεπλεγμένο αριθμητικό σχήμα όπου το αντίστοιχο με την (5.28) σφάλμα διακριτοποίησης η_i^j ικανοποιεί μια σχέση όπως αυτή της

(5.29), δηλαδή

$$|\eta_i^j| \leq C(k + h^2),$$

με C ανεξάρτητη των k και h. Στη συνέχεια, εκφράστε τη μέθοδο σε μορφή γραμμικού συστήματος, όπως η (5.33).

(β΄) Γράψτε ένα πεπλεγμένο αριθμητικό σχήμα όπου το σφάλμα διακριτοποίησης η_i^j ικανοποιεί μια αντίστοιχη σχέση όπως αυτή της (5.44), δηλαδή

$$|\eta_i^j| \leq C(k^2 + h^2),$$

με C ανεξάρτητη των k και h. Επίσης, εκφράστε τη μέθοδο σε μορφή γραμμικού συστήματος.

5.3. Έστω N, M φυσικοί αριθμοί, $h = L/(N+1)$, $x_i = ih$, $i = 0, \ldots, N$, μια διαμέριση του $[0, L]$, καθώς και $k = T/M$, $t^j = jk$, $j = 0, \ldots, M$, μια διαμέριση του $[0, T]$, με $L, T > 0$. Θεωρούμε ένα αριθμητικό σχήμα που χρησιμοποιεί τις τιμές στα (x_{i-1}, t^{j+1}), (x_i, t^{j+1}), (x_{i+1}, t^{j+1}), (x_{i-1}, t^j) και (x_{i+1}, t^j) για την προσέγγιση της λύσης της εξίσωσης της θερμότητας (5.1) στο σημείο (x_i, t^j).

(α΄) Δείξτε την ύπαρξη ένος τέτοιου αριθμητικού σχήματος, καθώς και την αντίστοιχη εκτίμηση του σφάλματος διακριτοποίησης η_i^j.

(β΄) Είναι η μέθοδος συνεπής;

(γ΄) Είναι ευσταθής;

5.4. Έστω N, M φυσικοί αριθμοί, $h = L/(N+1)$, $x_i = ih$, $i = 0, \ldots, N$, μια διαμέριση του $[0, L]$, καθώς και $k = T/M$, $t^j = jk$, $j = 0, \ldots, M$, μια διαμέριση του $[0, T]$, με $L, T > 0$. Θεωρούμε ένα αριθμητικό σχήμα που χρησιμοποιεί τις τιμές στα (x_{i-1}, t^{j+1}), (x_{i+1}, t^{j+1}), (x_{i-1}, t^j), (x_{i-1}, t^{j-1}) και (x_{i+1}, t^{j-1}) για την προσέγγιση της λύσης της εξίσωσης της θερμότητας (5.1) στο σημείο (x_i, t^j).

(α΄) Δείξτε την ύπαρξη ένος τέτοιου αριθμητικού σχήματος καθώς και την αντίστοιχη εκτίμηση του σφάλματος διακριτοποίησης η_i^j.

(β΄) Είναι η μέθοδος συνεπής;

(γ΄) Είναι ευσταθής;

5.5. Θεωρούμε το πρόβλημα

$$\begin{aligned} u_t(x,t) &= Du_{xx}(x,t) - \beta u(x,t), \quad (x,t) \in [0,L] \times [0,T], \\ u(x,0) &= g(x), \quad x \in [0,L], \end{aligned}$$

$$u(0,t) = u(L,t) = 0, \quad t \in [0,T],$$

όπου D, β είναι θετικές σταθερές.

(α΄) Γράψτε ένα πεπλεγμένο αριθμητικό σχήμα με σφάλμα διακριτοποίησης η_i^j το οποίο να ικανοποιεί μια σχέση της μορφής

$$|\eta_i^j| \leq C(k + h^2),$$

με C ανεξάρτητη των k και h. Στη συνέχεια, εκφράστε τη μέθοδο σε μορφή γραμμικού συστήματος.

(β΄) Είναι αυτή η μέθοδος ευσταθής; Είναι ευσταθής υπό συνθήκες;

5.6. Θεωρούμε το πρόβλημα

$$\begin{aligned} u_t(x,t) &= Du_{xx}(x,t) - \alpha u_x(x,t), \quad (x,t) \in [0,L] \times [0,T], \\ u(x,0) &= g(x), \quad x \in [0,L], \\ u(0,t) &= u(L,t) = 0, \quad t \in [0,T] \end{aligned}$$

όπου D, α είναι σταθερές και $D > 0$.

(α΄) Γράψτε ένα πεπλεγμένο αριθμητικό σχήμα με σφάλμα διακριτοποίησης η_i^j το οποίο να ικανοποιεί μια σχέση της μορφής

$$|\eta_i^j| \leq C(k + h^2),$$

με C ανεξάρτητη των k και h. Στη συνέχεια, εκφράστε τη μέθοδο σε μορφή γραμμικού συστήματος.

(β΄) Είναι αυτή η μέθοδος ευσταθής; Είναι ευσταθής υπό συνθήκες;

5.7. Θεωρούμε το πρόβλημα

$$\begin{aligned} u_t(x,t) &= Du_{xx}(x,t) + \mu(x)u_x(x,t), \quad (x,t) \in [0,L] \times [0,T], \\ u(x,0) &= g(x), \quad x \in [0,L], \\ u(0,t) &= u(L,t) = 0, \quad t \in [0,T], \end{aligned}$$

όπου D, μ είναι θετικές σταθερές.

(α΄) Γράψτε ένα πεπλεγμένο αριθμητικό σχήμα με σφάλμα διακριτοποίησης η_i^j το οποίο να ικανοποιεί μια σχέση της μορφής

$$|\eta_i^j| \leq C(k + h^2),$$

με C ανεξάρτητη των k και h. Στη συνέχεια, εκφράστε τη μέθοδο σε μορφή γραμμικού συστήματος.

(β΄) Είναι αυτή η μέθοδος ευσταθής; Είναι ευσταθής υπό συνθήκες;

5.8. Θεωρούμε το πρόβλημα

$$\begin{aligned} u_t(x,t) &= Du_{xx}(x,t) - \beta u(x,t), \quad (x,t) \in [0,1] \times [0,T], \\ u(x,0) &= g(x), \quad x \in [0,1], \\ u(0,t) &= h(t),\ u(1,t) = 0, \quad t \in [0,T], \end{aligned}$$

όπου D, β είναι θετικές σταθερές.

(α΄) Γράψτε ένα πεπλεγμένο αριθμητικό σχήμα με σφάλμα διακριτοποίησης η_i^j το οποίο να ικανοποιεί μια σχέση της μορφής

$$|\eta_i^j| \leq C(k^2 + h^2),$$

με C ανεξάρτητη των k και h. Στη συνέχεια, εκφράστε τη μέθοδο σε μορφή γραμμικού συστήματος.

(β΄) Είναι αυτή η μέθοδος ευσταθής; Είναι ευσταθής υπό συνθήκες;

Βιβλιογραφία

Ακρίβης, Γ., & Δουγαλής, Β. (2005). *Αριθμητικές Μέθοδοι για Μερικές Διαφορικές Εξισώσεις*. Ιωάννινα. (Πανεπιστημιακές Σημειώσεις).

Holmes, M. H. (2007). *Introduction to numerical methods in differential equations* (Vol. 52). Springer, New York.

Iserles, A. (2009). *A first course in the numerical analysis of differential equations* (Second ed.). Cambridge University Press, Cambridge.

Larsson, S., & Thomée, V. (2009). *Partial differential equations with numerical methods* (Vol. 45). Springer-Verlag, Berlin.

Morton, K. W., & Mayers, D. F. (2005). *Numerical solution of partial differential equations* (Second ed.). Cambridge University Press, Cambridge.

Strikwerda, J. C. (2004). *Finite difference schemes and partial differential equations* (Second ed.). Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA.

Thomas, J. W. (1995). *Numerical partial differential equations: finite difference methods* (Vol. 22). Springer-Verlag, New York.

Κεφάλαιο 6

Η μέθοδος των πεπερασμένων στοιχείων για την εξίσωση της θερμότητας

Σε αυτό το κεφάλαιο θεωρούμε την εξίσωση της θερμότητας στη μια διάσταση ως προς τον χώρο και θα κατασκευάσουμε μεθόδους πεπερασμένων στοιχείων για το πρόβλημα αρχικών και συνοριακών τιμών με ομογενείς συνοριακές συνθήκες Dirichlet. Η κατασκευή της μεθόδου θα γίνει σε δύο βήματα. Πρώτα, θα θεωρήσουμε ένα βοηθητικό πρόβλημα αρχικών τιμών που καλείται ημιδιακριτό πρόβλημα, όπου χρησιμοποιούμε μια μέθοδο πεπερασμένων στοιχείων για τη διακριτοποίηση μόνο ως προς τον χώρο. Στη συνέχεια, η πλήρως διακριτή μέθοδος προκύπτει, αν διακριτοποιήσουμε το ημιδιακριτό πρόβλημα χρησιμοποιώντας μια μέθοδο για προβήματα αρχικών τιμών όπως είναι η άμεση μέθοδος του Euler ή η πεπλεγμένη μέθοδο του Euler ή η μέθοδος των Crank–Nicolson.

6.1 Ημιδιακριτό πρόβλημα

Θεωρούμε το πρόβλημα αρχικών και συνοριακών τιμών για την εξίσωση της θερμότητας με ομογενείς συνοριακές συνθήκες Dirichlet (5.1), δηλαδή ζητούμε μια συνάρτηση $u : [0, L] \times [0, T] \to \mathbb{R}$ η οποία ικανοποιεί

$$\begin{aligned} u_t(x,t) &= u_{xx}(x,t), \qquad x \in [0,L],\ t \in [0,T], \\ u(0,t) &= u(L,t) = 0, \quad t \in [0,T], \\ u(x,0) &= g(x), \qquad\quad x \in [0,L], \end{aligned} \tag{6.1}$$

όπου $L > 0$, $g \in C[0, L]$.

Έστω μια διαμέριση του $[0, L]$, με $0 = x_0 < x_1 < \cdots < x_{N+1} = L$, με βήμα $h = L/(N + 1)$, $N \geq 1$. Θεωρούμε και πάλι τον χώρο συναρτήσεων V, βλ. π.χ. (4.4), που αποτελείται από τις συνεχείς συναρτήσεις οι οποίες είναι κατά τμήματα C^1 συνεχείς και μηδενίζονται στα άκρα του $[0, L]$. Επίσης, συμβολίζουμε με $(\cdot, \cdot)$ το εσωτερικό γινόμενο

$$(f, g) = \int_0^L f(x)g(x)\, dx,$$

και με $\| \cdot \|$ την αντίστοιχη παραγόμενη νόρμα.

Για μια συνάρτηση $v \in V$ θεωρούμε το εσωτερικό γινόμενο και ως προς τα δύο μέλη της εξίσωσης (6.1), οπότε έχουμε την ακόλουθη μεταβολική μορφή του προβλήματος (6.1): Ζητείται $u(\cdot, t)$, $t \in [0, T]$, τέτοια ώστε

$$(u_t(\cdot, t), v) + (u_x(\cdot, t), v') = 0, \quad \forall v \in V, \quad u(\cdot, 0) = g. \tag{6.2}$$

Αν θεωρήσουμε τώρα έναν υπόχωρο V_h του V και ακολουθήσουμε ανάλογα βήματα όπως στο Κεφάλαιο 4, προκύπτει μια προσέγγιση $u_h(\cdot, t) \in V_h, t \in [0, T]$, της $u(\cdot, t)$ η οποία αποτελεί λύση ενός νέου πρόβληματος, του λεγομένου ημιδιακριτού προβλήματος του (6.2).

Ως V_h μπορούμε, παραδείγματος χάριν, να θεωρήσουμε τον χώρο των συνεχών συναρτήσεων που είναι κατά τμήματα γραμμικά πολυώνυμα και μηδενίζονται στα άκρα του $[0, L]$, δηλαδή

$$V_h = \{\chi \in C_0[0, L] : \chi|_{[x_j, x_{j+1}]} \in \mathbb{P}_1,\ j = 0, \ldots, N\}. \tag{6.3}$$

Ορίζουμε τότε ως ημιδιακριτή λύση τη $u_h(\cdot, t) \in V_h$, $t \in [0, T]$, τέτοια ώστε

$$(u_{h,t}(\cdot, t), \chi) + (u_{h,x}(\cdot, t), \chi') = 0, \quad \forall \chi \in V_h, \quad u_h(\cdot, 0) = g_h, \tag{6.4}$$

όπου $g_h \in V_h$ μια προσέγγιση της g.

Στη συνέχεια, ως V_h θα θεωρήσουμε τον χώρο των κατά τμήματα γραμμικών συναρτήσεων (6.3) όμως τα συμπεράσματα που έπονται επεκτείνονται ανάλογα και για άλλους υπόχωρους του V. Είδαμε στο Λήμμα 4.2 ότι οι συναρτήσεις $\{\phi_j\}_{j=1}^N$ της (4.15) αποτελούν μια βάση του χώρου V_h. Οπότε, επειδή $u_h(\cdot, t) \in V_h$, υπάρχουν $\alpha_j(t)$, $t \in [0, T]$, τέτοια ώστε

$$u_h(x, t) = \sum_{j=1}^{N} \alpha_j(t)\phi_j(x), \quad x \in [0, L],\ t \in [0, T].$$

Τότε, το πρόβλημα (6.4) γράφεται ισοδύναμα

$$\begin{cases} \displaystyle\sum_{j=1}^{N} \alpha_j'(t)(\phi_j, \chi) + \sum_{j=1}^{N} \alpha_j(t)(\phi_j', \chi') = 0, \quad \forall \chi \in V_h, \\ \\ \alpha_j(0) = \gamma_j, \quad j = 1, \ldots, N, \end{cases}$$

όπου $g_h = \sum_{j=1}^{N} \gamma_j \phi_j$. Στη συνέχεια, επιλέγοντας ως $\chi = \phi_i$, $i = 1, \ldots, N$, λαμβάνουμε

$$\begin{cases} \displaystyle\sum_{j=1}^{N} \alpha_j'(t)(\phi_j, \phi_i) + \sum_{j=1}^{N} \alpha_j(t)(\phi_j', \phi_i') = 0, \quad i = 1, \ldots, N, \\ \\ \alpha_j(0) = \gamma_j, \quad j = 1, \ldots, N. \end{cases} \tag{6.5}$$

Μπορούμε να δούμε τώρα ότι η (6.5) γράφεται και ως πρόβλημα αρχικών τιμών για ένα γραμμικό σύστημα συνήθων διαφορικών εξισώσεων,

$$\mathcal{M}\alpha'(t) + \mathcal{S}\alpha(t) = 0, \quad t \in [0, T], \quad \text{με } \alpha(0) = \Gamma, \tag{6.6}$$

όπου $\mathcal{M}$ και $\mathcal{S}$ είναι $N \times N$ πίνακες με στοιχεία $\mathcal{M}_{ij} = (\phi_j, \phi_i)$ και $\mathcal{S}_{ij} = (\phi_j', \phi_i')$, $i, j = 1, \ldots, N$, αντίστοιχα, $\alpha(t) = (\alpha_1(t), \ldots, \alpha_N(t))^T$ και $\Gamma = (\gamma_1, \ldots, \gamma_N)^T$. Ο πίνακας $\mathcal{M}$ καλείται *πίνακας μάζας* και ο πίνακας $\mathcal{S}$ *πίνακας ακαμψίας*. Μπορούμε να παρατηρήσουμε ότι και οι δύο πίνακες $\mathcal{M}$ και $\mathcal{S}$ είναι συμμετρικοί. Επίσης, είναι και θετικά ορισμένοι. Πράγματι, αν $w = (w_1, \ldots, w_N)^T \in \mathbb{R}^N$, τότε

$$w^T \mathcal{M} w = \|v\|^2 \geq 0 \text{ και } w^T \mathcal{S} w = \|v'\|^2 \geq 0, \quad \text{όπου } v = \sum_{i=1}^{N} w_i \phi_i.$$

Ακόμα εύκολα βλέπουμε από την παραπάνω σχέση ότι αν $w^T \mathcal{M} w = 0$ ή $w^T \mathcal{S} w = 0$, τότε $w = 0$. Συνεπώς, ο $\mathcal{M}$ είναι αντιστρέψιμος, βλ. π.χ. (Ακρίβης & Δουγαλής, 2015), οπότε το σύστημα διαφορικών εξισώσεων (6.6) γράφεται

$$\alpha'(t) + \mathcal{M}^{-1}\mathcal{S}\alpha(t) = 0, \quad t \in [0, T], \quad \text{με } \alpha(0) = \Gamma,$$

το οποίο λύνεται μονοσήμαντα.

Στη συνέχεια, θα δείξουμε την ευστάθεια του ημιδιακριτού προβλήματος (6.4).

Θεώρημα 6.1. *Εστω $u_h(\cdot, t)$, $t \in [0, T]$, η λύση του* (6.4)*. Τότε*

$$\|u_h(t)\| \leq \|g_h\|, \quad \forall t \in [0, T]. \tag{6.7}$$

Απόδειξη. Επιλέγοντας $\chi = u_h(\cdot, t)$ στην (6.4), έχουμε

$$(u_{h,t}(\cdot, t), u_h(\cdot, t)) + (u_{h,x}(\cdot, t), u_{h,x}(\cdot, t)) = 0. \tag{6.8}$$

Παρατηρούμε τώρα ότι $\frac{1}{2}\frac{d}{dt}(\|u_h(\cdot, t)\|^2) = (u_{h,t}(\cdot, t), u_h(\cdot, t))$, οπότε η (6.8) γίνεται

$$\frac{1}{2}\frac{d}{dt}(\|u_h(\cdot, t)\|^2) + \|u_{h,x}(\cdot, t)\|^2 = 0.$$

Συνεπώς,

$$\frac{d}{dt}(\|u_h(\cdot, t)\|^2) \leq 0,$$

από όπου εύκολα προκύπτει η ζητούμενη (6.7). □

Θα θεωρήσουμε τώρα μια προβολή $R_h : V \to V_h$ η οποία ορίζεται με τον ακόλουθο τρόπο

$$((R_h v)', \chi') = (v', \chi'), \quad \forall \chi \in V_h. \tag{6.9}$$

Η προβολή R_h καλείται *ελλειπτική προβολή* ή *προβολή Ritz*.

Λήμμα 6.1. *Έστω $v \in C_0^2[0, L]$. Η προβολή R_h ορίζεται μονοσήμαντα και ικανοποιεί τη σχέση*

$$\|R_h v - v\| + h\|R_h v' - v'\| \leq Ch^2\|v''\|. \tag{6.10}$$

Απόδειξη. Θέτουμε $w = -v''$ και θεωρούμε το ακόλουθο βοηθητικό πρόβλημα

$$-v''(x) = w(x), \quad x \in [0, L], \quad v(0) = v(L) = 0.$$

Εύκολα βλέπουμε τώρα ότι η $R_h v$ όπως ορίστηκε στην (6.9) είναι η λύση με τη μέθοδο των πεπερασμένων στοιχείων από τον χώρο V_h. Τότε, σύμφωνα με το Θεώρημα 4.2, έχουμε ότι ισχύει η ζητούμενη (6.10). □

Λήμμα 6.2. *Έστω $u \in C^2([0, L] \times [0, T])$ η λύση του* (6.1) *και $R_h u(\cdot, t)$, $t \in [0, T]$, η ελλειπτική προβολή της u που ορίζεται από την* (6.9)*. Τότε υπάρχει σταθερά C, τέτοια ώστε*

$$\|R_h u(\cdot, t) - u(\cdot, t)\| + \|(R_h u)_t(\cdot, t) - u_t(\cdot, t)\| \leq Ch^2, \quad t \in [0, T]. \tag{6.11}$$

Απόδειξη. Από τον ορισμό της R_h (6.9) και το γεγονός ότι $u \in C^2([0, L] \times [0, T])$ μπορούμε να δούμε ότι $R_h u_t = (R_h u)_t$. Πράγματι,

$$\begin{aligned}((R_h u_t)_x, \chi') &= ((u_t)_x, \chi') = (\frac{\partial}{\partial t}(u_x), \chi') = \frac{\partial}{\partial t}(u_x, \chi') \\ &= \frac{\partial}{\partial t}((R_h u)_x, \chi') = (\frac{\partial}{\partial t}((R_h u)_x), \chi') = ((R_h u)_t)_x, \chi').\end{aligned}$$

Επομένως, λόγω του Λήμματος 6.1, παίρνουμε την επιθυμητή (6.11). □

Θεώρημα 6.2. *Έστω $u \in C^2([0,L]\times[0,T])$ η λύση της* (6.1) *με $g \in C^2[0,L]$ και u_h η λύση της* (6.4) *με $g_h = R_h g$. Τότε, υπάρχει σταθερά C, ανεξάρτητη του h, τέτοια ώστε*

$$\max_{0\le t\le T} \|u_h(\cdot,t) - u(\cdot,t)\| \le Ch^2. \tag{6.12}$$

Απόδειξη. Έστω $W(x,t) = R_h u(x,t)$. Συμβολίζουμε με $\rho(t) = W(\cdot,t) - u(\cdot,t)$ και $\vartheta(t) = u_h(\cdot,t) - W(\cdot,t)$, οπότε $u_h(\cdot,t) - u(\cdot,t) = \vartheta(t) + \rho(t)$. Λόγω του Λήμματος 6.1, έχουμε ότι

$$\|\rho(t)\| \le Ch^2. \tag{6.13}$$

Στη συνέχεια, μπορούμε να δούμε ότι η ϑ ικανοποιεί την ακόλουθη σχέση

$$\begin{aligned}
&(\vartheta_t(t),\chi) + (\vartheta_x(t),\chi') \\
&= (u_{h,t}(\cdot,t),\chi) + (u_{h,x}(\cdot,t),\chi') - (W_t(\cdot,t),\chi) - (W_x(\cdot,t),\chi') \\
&= (u_t(\cdot,t),\chi) + (u_x(\cdot,t),\chi') - (W_t(\cdot,t),\chi) - (W_x(\cdot,t),\chi') \\
&= -(\rho_t(t),\chi) + (u_x(\cdot,t) - W_x(\cdot,t),\chi') = -(\rho_t(t),\chi).
\end{aligned}$$

Επομένως

$$(\vartheta_t(t),\chi) + (\vartheta_x(t),\chi') = -(\rho_t(t),\chi). \tag{6.14}$$

Επιλέγοντας τώρα $\chi = \vartheta(t)$, έχουμε

$$(\vartheta_t(t),\vartheta(t)) + \|\vartheta_x(t)\|^2 = -(\rho_t(t),\vartheta(t)). \tag{6.15}$$

Επειδή $\dfrac{1}{2}\dfrac{d}{dt}\|\vartheta(t)\|^2 = (\vartheta_t(t),\vartheta(t))$, η (6.15) γίνεται

$$\frac{d}{dt}\|\vartheta(t)\|\,\|\vartheta(t)\| \le \|\rho_t(t)\|\,\|\vartheta(t)\|, \tag{6.16}$$

οπότε

$$\frac{d}{dt}\|\vartheta(t)\| \le \|\rho_t(t)\|. \tag{6.17}$$

Ολοκληρώνοντας στη συνέχεια στο $[0,t]$, παίρνουμε

$$\|\vartheta(t)\| \le \|\vartheta(0)\| + \int_0^t \|\rho_t(s)\|\,ds. \tag{6.18}$$

Η ζητούμενη σχέση τώρα έπεται άμεσα από το γεγονός ότι $\theta(0) = 0$, την (6.13) και το Λήμμα 6.2. □

6.2 Πεπλεγμένη μέθοδος του Euler

Θα θεωρήσουμε τώρα ένα πλήρως διακριτό σχήμα, διακριτοποιώντας και ως προς τον χρόνο το ημιδιακριτό σχήμα (6.4) με την πεπλεγμένη μέθοδο του Euler. Για έναν φυσικό αριθμό $M \geq 1$ θέτουμε $k = T/M$ και $t^j = jk$, $j = 0, \ldots, M$, μια διαμέριση του $[0, T]$.

Θεωρούμε λοιπόν τις προσεγγίσεις $U^n \in V_h$, $n = 1, \ldots, M$, τέτοιες ώστε

$$(\frac{U^n - U^{n-1}}{k}, \chi) + ((U^n)', \chi') = 0, \quad \forall \chi \in V_h, \quad U^0 = g_h, \tag{6.19}$$

όπου $g_h \in V_h$ μια προσέγγιση της g. Η (6.19) γράφεται ισοδύναμα ως

$$(U^n, \chi) + k((U^n)', \chi') = (U^{n-1}, \chi), \quad \forall \chi \in V_h, \quad U^0 = g_h. \tag{6.20}$$

Αν θεωρήσουμε ως βάση του V_h τις συναρτήσεις $\{\phi_j\}_{j=1}^N$, τότε οι προσεγγίσεις U^n του (6.20) γράφονται ως γραμμικός συνδυασμός των ϕ_j ως

$$U^n(x) = \sum_{j=0}^{N} \alpha_j^n \phi_j(x), \quad x \in [0, L].$$

Επιλέγοντας τώρα $\chi = \phi_i$, $i = 1, \ldots, N$, στην (6.20) παρατηρούμε ότι οδηγούμαστε σε ένα γραμμικό σύστημα της μορφής

$$(\mathcal{M} + k\mathcal{S})\alpha^n = \mathcal{M}\alpha^{n-1},$$

όπου $\mathcal{M}$ και $\mathcal{S}$ είναι οι πίνακες της (6.6) και α^j διανύσματα με συνιστώσες $\alpha^j = (\alpha_1^j, \ldots, \alpha_N^j)^T$, $j = 0, \ldots, N$. Είναι απλό να δούμε ότι ο πίνακας $\mathcal{M} + k\mathcal{S}$ είναι συμμετρικός και θετικά ορισμένος και, άρα, αντιστρέφεται. Επομένως,

$$\alpha^n = (\mathcal{M} + k\mathcal{S})^{-1}\mathcal{M}\alpha^{n-1}. \tag{6.21}$$

Άρα, ξεκινώντας από τη γνωστή προσέγγιση $U^0 = g_h$ της g, όπου γνωρίζουμε το διάνυσμα α^0, μπορούμε χρησιμοποιώντας την (6.21) να προσδιορίσουμε αναδρομικά τα α^n και κατ'επέκταση την προσέγγιση U^n της $u(\cdot, t^n)$.

Στη συνέχεια, θα δείξουμε την ευστάθεια της μεθόδου (6.19).

Θεώρημα 6.3. *Έστω ότι οι $U^n \in V_h$, $n = 0, \ldots, M$, ικανοποιούν την* (6.19). *Τότε*

$$\max_{0 \leq n \leq M} \|U^n\| \leq \|U^0\|. \tag{6.22}$$

Απόδειξη. Επιλέγουμε $\chi = U^n$ στην (6.19), οπότε λαμβάνουμε

$$(\frac{U^n - U^{n-1}}{k}, U^n) + \|(U^n)'\|^2 = 0. \tag{6.23}$$

Επομένως, έχουμε

$$\|U^n\|^2 \leq (U^{n-1}, U^n) \leq \|U^{n-1}\| \, \|U^n\|.$$

Άρα

$$\|U^n\| \leq \|U^{n-1}\|,$$

από όπου προκύπτει η ζητούμενη σχέση (6.22). □

Στη συνέχεια, αποδεικνύουμε τη σύγκλιση της πεπλεγμένης μεθόδου του Euler.

Θεώρημα 6.4. *Έστω $u \in C^2([0, L] \times [0, T])$ η λύση της* (6.1) *με $g \in C^2[0, L]$ και $U^n \in V_h$, $n = 0, \ldots, M$, ικανοποιούν την* (6.19) *με $U^0 = R_h g$. Τότε υπάρχει σταθερά C, ανεξάρτητη των k και h, τέτοια ώστε*

$$\max_{0 \leq n \leq M} \|U^n - u(\cdot, t^n)\| \leq C(k + h^2). \tag{6.24}$$

Απόδειξη. Θέτουμε $\rho^n = W(\cdot, t^n) - u(\cdot, t^n)$ και $\vartheta^n = U^n - W(\cdot, t^n)$, οπότε $U^n - u(\cdot, t^n) = \vartheta^n + \rho^n$. Λόγω του Λήμματος 6.2, έχουμε ότι

$$\max_{0 \leq n \leq M} \|\rho^n\| \leq Ch^2. \tag{6.25}$$

Στη συνέχεια, μπορούμε να δούμε ότι η ϑ^n ικανοποιεί την ακόλουθη σχέση

$$(\frac{\vartheta^n - \vartheta^{n-1}}{k}, \chi) + ((\vartheta^n)', \chi') = -(\omega^n, \chi), \quad \forall \chi \in V_h, \tag{6.26}$$

όπου $\omega^n = \omega_1^n + \omega_2^n$ και

$$\omega_1^n = \frac{\rho^n - \rho^{n-1}}{k}, \quad \omega_2^n = \frac{u^n - u^{n-1}}{k} - u_t^n,$$

με $u^n = u(\cdot, t^n)$ και $u_t^n = u(\cdot, t^n)$. Πράγματι έχουμε

$$\begin{aligned}
&(\frac{\vartheta^n - \vartheta^{n-1}}{k}, \chi) + ((\vartheta^n)', \chi') \\
&= -(\frac{R_h u^n - R_h u^{n-1}}{k}, \chi) - ((R_h u^n)', \chi') \\
&= -(\frac{\rho^n - \rho^{n-1}}{k}, \chi) - (\frac{u^n - u^{n-1}}{k}, \chi) - ((u^n)_x, \chi')
\end{aligned}$$

$$= -(\frac{\rho^n - \rho^{n-1}}{k}, \chi) - (\frac{u^n - u^{n-1}}{k}, \chi) + (u_t^n, \chi)$$
$$= -(\omega_1^n + \omega_2^n, \chi).$$

Στη συνέχεια, θα δείξουμε ότι

$$\|\omega_1^n\| \leq Ch^2 \quad \text{και} \quad \|\omega_2^n\| \leq Ck.$$

Εύκολα βλέπουμε ότι

$$\omega_1^n = \frac{1}{k} \int_{t^{n-1}}^{t^n} \rho(t)\, dt,$$

οπότε λόγω της (6.13)

$$\|\omega_1^n\| \leq \frac{1}{k} \int_{t^{n-1}}^{t^n} \|\rho(t)\|\, dt \leq Ch^2. \tag{6.27}$$

Επίσης, λόγω του αναπτύγματος Taylor έχουμε

$$\|\omega_2^n\| = \|\frac{u^n - u^{n-1}}{k} - u_t^n\| = k^{-1}\| \int_{t^{n-1}}^{t^n} (t - t^{n-1}) u_{tt}(\cdot, t)\, dt\|.$$

Συνεπώς

$$\|\omega_2^n\| \leq Ck. \tag{6.28}$$

Άρα, συνδυάζοντας τις (6.27), (6.28), παίρνουμε

$$\|\omega_n\| \leq C(k + h^2), \quad n = 1, \ldots, M. \tag{6.29}$$

Επιλέγοντας τώρα $\chi = \vartheta^n$ στην (6.26), λαμβάνουμε

$$\|\vartheta^n\|^2 + k\|(\vartheta^n)'\|^2 = (\vartheta^{n-1}, \vartheta^n) - k(\omega^n, \vartheta^n) \leq (\|\vartheta^{n-1}\| + k\|\omega^n\|)\|\vartheta^n\|.$$

Οπότε

$$\|\vartheta^n\| \leq \|\vartheta^{n-1}\| + k\|\omega^n\|,$$

από την οποία παίρνουμε την

$$\|\vartheta^n\| \leq \|\vartheta^0\| + k \sum_{j=1}^{n} \|\omega^j\| \leq \|\vartheta^0\| + Mk \max_{0 \leq j \leq M} \|\omega^j\|.$$

Συνεπώς, λόγω των (6.25), (6.29) και του γεγονότος ότι $\vartheta(0) = 0$, έχουμε ότι ισχύει η ζητούμενη σχέση. □

6.3 Μέθοδος των Crank–Nicolson

Στη συνέχεια, θα θεωρήσουμε ένα πλήρως διακριτό σχήμα, διακριτοποιώντας και ως προς τον χρόνο το ημιδιακριτό σχήμα (6.4) με την μέθοδο των Crank–Nicolson. Θέτουμε και πάλι $k = T/M$ και $t^j = jk$, $j = 0, \ldots, M$, μια διαμέριση του $[0, T]$.

Θεωρούμε λοιπόν τις προσεγγίσεις $U^n \in V_h$, $n = 1, \ldots, M$, τέτοιες ώστε

$$(\frac{U^n - U^{n-1}}{k}, \chi) + ((U^{n-1/2})', \chi') = 0, \quad \forall \chi \in V_h, \quad U^0 = g_h, \tag{6.30}$$

όπου $g_h \in V_h$ μια προσέγγιση της g και $U^{n-1/2} = \frac{1}{2}(U^n + U^{n-1})$. Η (6.30) γράφεται και ως εξής

$$\begin{aligned}(U^n, \chi) + \frac{k}{2}((U^n)', \chi') = (U^{n-1}, \chi) - \frac{k}{2}((U^{n-1})', \chi'), \quad \forall \chi \in V_h,\\ U^0 = g_h.\end{aligned} \tag{6.31}$$

Με ανάλογο τρόπο όπως και για την πεπλεγμένη μέθοδο του Euler στην προηγούμενη παράγραφο, η (6.31) είναι ισοδύναμη με ένα γραμμικό σύστημα της μορφής

$$(\mathcal{M} + \frac{k}{2}\mathcal{S})\alpha^n = (\mathcal{M} - \frac{k}{2}\mathcal{S})\alpha^{n-1},$$

όπου $\mathcal{M}$ και $\mathcal{S}$ είναι οι πίνακες της (6.6) και α^j διανύσματα με συνιστώσες $\alpha^j = (\alpha_1^j, \ldots, \alpha_N^j)^T$, $j = 0, \ldots, N$. Εύκολα βλέπουμε ότι ο πίνακας $\mathcal{M} + \frac{k}{2}\mathcal{S}$ είναι συμμετρικός και θετικά ορισμένος και, άρα, αντιστρέφεται. Επομένως,

$$\alpha^n = (\mathcal{M} + \frac{k}{2}\mathcal{S})^{-1}(\mathcal{M} - \frac{k}{2}\mathcal{S})\alpha^{n-1}. \tag{6.32}$$

Συνεπώς, επειδή γνωρίζουμε το διάνυσμα α^0, χρησιμοποιώντας την (6.32) μπορούμε να προσδιορίσουμε αναδρομικά τα α^n και κατ'επέκταση την προσέγγιση U^n της $u(\cdot, t^n)$.

Στη συνέχεια, θα δείξουμε την ευστάθεια της μεθόδου (6.30).

Θεώρημα 6.5. *Έστω ότι οι $U^n \in V_h$, $n = 0, \ldots, M$, ικανοποιούν την* (6.30)*. Τότε*

$$\max_{0 \le n \le M} \|U^n\| \le \|U^0\|. \tag{6.33}$$

Απόδειξη. Επιλέγουμε $\chi = U^{n-1/2}$ στην (6.30). Οπότε

$$(\frac{U^n - U^{n-1}}{k}, U^{n-1/2}) + \|(U^{n-1/2})'\|^2 = 0. \tag{6.34}$$

Επομένως, έχουμε ότι

$$(U^n - U^{n-1}, U^n + U^{n-1}) \leq 0,$$

από όπου εύκολα βλέπουμε ότι

$$\|U^n\|^2 \leq \|U^{n-1}\|^2 \leq \cdots \leq \|U^0\|^2,$$

δηλαδή τη ζητούμενη ανισότητα. □

Θεώρημα 6.6. *Έστω $u \in C^4([0,L] \times [0,T])$ η λύση της* (6.1) *με $g \in C^2[0,L]$ και $U^n \in V_h$, $n = 0, \ldots, M$, ικανοποιούν την* (6.30) *με $U^0 = R_h g$. Τότε, υπάρχει σταθερά C, ανεξάρτητη των k και h, τέτοια ώστε*

$$\max_{0 \leq n \leq M} \|U^n - u(\cdot, t^n)\| \leq C(k^2 + h^2). \tag{6.35}$$

Απόδειξη. Θέτουμε και πάλι $\rho^n = W(\cdot, t^n) - u(\cdot, t^n)$ και $\vartheta^n = U^n - W(\cdot, t^n)$, οπότε $U^n - u(\cdot, t^n) = \vartheta^n + \rho^n$. Λόγω της (6.25) αρκεί να εκτιμήσουμε την $\|\vartheta^n\|$. Μπορούμε να δούμε ότι το ϑ^n ικανοποιεί την ακόλουθη σχέση

$$(\frac{\vartheta^n - \vartheta^{n-1}}{k}, \chi) + ((\vartheta^{n-1/2})', \chi') = -(\omega^n, \chi), \quad \forall \chi \in V_h, \tag{6.36}$$

όπου $\omega^n = \omega_1^n + \omega_2^n + \omega_3^n$ και

$$\omega_1^n = \frac{\rho^n - \rho^{n-1}}{k}, \quad \omega_2^n = \frac{u^n - u^{n-1}}{k} - u_t(\cdot, t^{n-1/2}),$$
$$\omega_3^n = u_{xx}(\cdot, t^{n-1/2}) - u_{xx}^{n-1/2}.$$

Πράγματι έχουμε

$$\begin{aligned}
&(\frac{\vartheta^n - \vartheta^{n-1}}{k}, \chi) + ((\vartheta^{n-1/2})', \chi') \\
&= -(\frac{R_h u^n - R_h u^{n-1}}{k}, \chi) - ((R_h u^{n-1/2})', \chi') \\
&= -(\frac{\rho^n - \rho^{n-1}}{k}, \chi) - (\frac{u^n - u^{n-1}}{k}, \chi) - (u_x^{n-1/2}, \chi') \\
&= -(\frac{\rho^n - \rho^{n-1}}{k}, \chi) - (\frac{u^n - u^{n-1}}{k}, \chi) + (u_{xx}^{n-1/2}, \chi) \\
&= -(\frac{\rho^n - \rho^{n-1}}{k}, \chi) - (\frac{u^n - u^{n-1}}{k}, \chi) + (u_t(\cdot, t^{n-1/2}), \chi) \\
&\quad - (u_t(\cdot, t^{n-1/2}), \chi) + (u_{xx}^{n-1/2}, \chi)
\end{aligned}$$

$$
\begin{aligned}
&= -(\frac{\rho^n - \rho^{n-1}}{k}, \chi) - (\frac{u^n - u^{n-1}}{k}, \chi) + (u_t(\cdot, t^{n-1/2}), \chi) \\
&\quad - (u_{xx}(\cdot, t^{n-1/2}), \chi) + (u_{xx}^{n-1/2}, \chi) \\
&= -(\omega_1^n + \omega_2^n + \omega_3^n, \chi).
\end{aligned}
$$

Από την (6.27) έχουμε ότι $\|\omega_1^n\| \leq Ch^2$. Επίσης, λόγω του αναπτύγματος Taylor έχουμε

$$
\begin{aligned}
\omega_2^n &= \frac{u^n - u^{n-1}}{k} - u_t(\cdot, t^{n-1/2}) \\
&= (2k)^{-1} \int_{t^{n-1}}^{t^{n-1/2}} (t - t^{n-1})^2 u_{ttt}(t)\, dt + (2k)^{-1} \int_{t^{n-1/2}}^{t^n} (t - t^n)^2 u_{ttt}(t)\, dt,
\end{aligned}
$$

από όπου προκύπτει ότι $\|\omega_2^n\| \leq Ck^2$. Τέλος, χρησιμοποιώντας και πάλι το ανάπτυγμα Taylor, έχουμε

$$
\begin{aligned}
2\omega_3^n &= 2(u_{xx}(\cdot, t^{n-1/2}) - (u_{xx})^{n-1/2}) \\
&= -\int_{t^{n-1}}^{t^{n-1/2}} (t - t^{n-1}) u_{xxtt}(t)\, dt + \int_{t^{n-1/2}}^{t^n} (t - t^n) u_{xxtt}(t)\, dt.
\end{aligned}
$$

Συνεπώς $\|\omega_3^n\| \leq Ck^2$. Συνδυάζοντας τις παραπάνω εκτιμήσεις, λαμβάνουμε

$$
\|\omega^n\| \leq C(k^2 + h^2). \tag{6.37}
$$

Επιλέγοντας τώρα $\chi = \vartheta^{n-1/2}$ στην (6.36), παίρνουμε

$$
\begin{aligned}
\|\vartheta^n\|^2 - \|\vartheta^{n-1}\|^2 + 2k\|(\vartheta^{n-1/2})'\|^2 &= -k(\omega^n, \vartheta^n + \vartheta^{n-1}) \\
&\leq k\|\omega^n\|(\|\vartheta^n\| + \|\vartheta^{n-1}\|).
\end{aligned}
$$

Οπότε

$$
(\|\vartheta^n\| - \|\vartheta^{n-1}\|)(\|\vartheta^n\| + \|\vartheta^{n-1}\|) \leq k\|\omega^n\|(\|\vartheta^n\| + \|\vartheta^{n-1}\|),
$$

από την οποία παίρνουμε

$$
(\|\vartheta^n\| \leq \|\vartheta^{n-1}\| + k\|\omega^n\|.
$$

Συνεπώς, έχουμε

$$
\|\vartheta^n\| \leq \|\vartheta^0\| + k\sum_{j=1}^{n} \|\omega^j\| \leq \|\vartheta^0\| + Mk \max_{0 \leq j \leq M} \|\omega^j\|.
$$

Επομένως, λόγω των (6.25), (6.37) και του γεγονότος ότι $\vartheta(0) = 0$, έχουμε ότι ισχύει η επιθυμητή σχέση (6.35). □

6.4 Άμεση μέθοδος του Euler

Στη συνέχεια, θα χρησιμοποιήσουμε την άμεση μέθοδο του Euler για τη διακριτοποίηση ως προς τον χρόνο της (6.4). Θέτουμε και πάλι $k = T/M$ και $t^j = jk$, $j = 0, \ldots, M$, μια διαμέριση του $[0, T]$.

Θεωρούμε λοιπόν τις προσεγγίσεις $U^n \in V_h$, $n = 1, \ldots, M$, τέτοιες ώστε

$$(\frac{U^n - U^{n-1}}{k}, \chi) + ((U^{n-1})', \chi') = 0, \quad \forall \chi \in V_h, \quad U^0 = g_h, \tag{6.38}$$

όπου $g_h \in V_h$ μια προσέγγιση της g. Η (6.38) γράφεται τώρα ως

$$(U^n, \chi) = (U^{n-1}, \chi) - k((U^{n-1})', \chi'), \quad \forall \chi \in V_h, \quad U^0 = g_h. \tag{6.39}$$

Με όμοιο τρόπο όπως και προηγουμένως η (6.39) είναι ισοδύναμη με ένα γραμμικό σύστημα της μορφής

$$\mathcal{M}\alpha^n = (\mathcal{M} - k\mathcal{S})\alpha^{n-1},$$

όπου $\mathcal{M}$ και $\mathcal{S}$ είναι οι πίνακες της (6.6) και α^j διανύσματα με συνιστώσες $\alpha^j = (\alpha_1^j, \ldots, \alpha_N^j)^T$, $j = 0, \ldots, N$. Επειδή ο πίνακας $\mathcal{M}$ είναι αντιστέψιμος, έχουμε

$$\alpha^n = \mathcal{M}^{-1}(\mathcal{M} - k\mathcal{S})\alpha^{n-1}. \tag{6.40}$$

Συνεπώς, επειδή γνωρίζουμε το διάνυσμα α^0, χρησιμοποιώντας την (6.40), μπορούμε να προσδιορίσουμε αναδρομικά τα α^n και κατ'επέκταση την προσέγγιση U^n της $u(\cdot, t^n)$.

Σε αυτό το σημείο και για να δείξουμε την ευστάθεια της μεθόδου (6.38), θα χρειαστεί να υποθέσουμε την ακόλουθη σχέση την οποία πρέπει να πληρούν οι συναρτήσεις του χώρου V_h. Υποθέτουμε λοιπόν ότι υπάρχει σταθερά C_1 ανεξάρτητη του h, τέτοια ώστε

$$\|\chi'\| \leq C_1 h^{-1} \|\chi\|, \quad \forall \chi \in V_h. \tag{6.41}$$

Η ανισότητα (6.41) καλείται *αντίστροφη ανισότητα*. Στην περίπτωση που V_h είναι οι συνεχείς κατά τμήματα γραμμικές συναρτήσεις σε έναν ομοιόμορφο διαμερισμό, τότε ικανοποιείται η (6.41), βλ. Άσκηση 6.3.

Στη συνέχεια, θα δείξουμε την ευστάθεια της μεθόδου (6.38).

Θεώρημα 6.7. *Έστω ότι οι $U^n \in V_h$, $n = 0, \ldots, M$, ικανοποιούν τη* (6.38) *και ότι ο V_h ικανοποιεί τη* (6.41)*. Τότε, αν $\dfrac{k}{h^2} \leq \dfrac{2}{C_1^2}$, όπου C_1 είναι η σταθερά στην* (6.41)*, έχουμε*

$$\max_{0 \leq n \leq M} \|U^n\| \leq \|U^0\|. \tag{6.42}$$

Απόδειξη. Επιλέγουμε $\chi = U^{n-1}$ στην (6.38), οπότε

$$(U^n, U^{n-1}) - \|U^{n-1}\|^2 + k\|(U^{n-1})'\| = 0. \tag{6.43}$$

Επίσης, εύκολα μπορούμε να δούμε ότι

$$2(U^n, U^{n-1}) = \|U^n\|^2 + \|U^{n-1}\|^2 - \|U^n - U^{n-1}\|^2.$$

Επομένως, η (6.43) γίνεται

$$\frac{1}{2}(\|U^n\|^2 - \|U^{n-1}\|^2 - \|U^n - U^{n-1}\|^2) + k\|(U^{n-1})'\| = 0. \tag{6.44}$$

Επίσης, αν επιλέξουμε $\chi = U^n - U^{n-1}$ στην (6.38), έχουμε

$$\|U^n - U^{n-1}\|^2 - k((U^{n-1})', (U^n)' - (U^{n-1})') = 0. \tag{6.45}$$

Στη συνέχεια, χρησιμοποιώντας την ανισότητα Cauchy–Schwarz και την (6.41), η (6.45) δίνει

$$\begin{aligned}\|U^n - U^{n-1}\|^2 &\le k\|(U^{n-1})'\|\,\|(U^n)' - (U^{n-1})'\| \\ &\le C_1\frac{k}{h}\|(U^{n-1})'\|\,\|U^n - U^{n-1}\|.\end{aligned}$$

Επομένως, λαμβάνουμε

$$\|U^n - U^{n-1}\| \le C_1\frac{k}{h}\|(U^{n-1})'\|. \tag{6.46}$$

Αν χρησιμοποιήσουμε τώρα την (6.46) στην (6.44), παίρνουμε

$$\|U^n\|^2 - \|U^{n-1}\|^2 \le (C_1^2\frac{k}{h^2} - 2)k\|(U^{n-1})'\|. \tag{6.47}$$

Συνεπώς, λόγω της υπόθεσης $C_1^2\dfrac{k}{h^2} - 2 \le 0$, η (6.47) δίνει τη ζητούμενη σχέση (6.42). □

Θεώρημα 6.8. *Έστω $u \in C^2([0,L]\times[0,T])$ η λύση της* (6.1) *με $g \in C^2[0,L]$ και $U^n \in V_h$, $n = 0,\dots,M$, ικανοποιούν την* (6.38) *με $U^0 = R_h g$ και ο V_h την* (6.41). *Τότε, αν $\dfrac{k}{h^2} \le \dfrac{2}{C_1^2}$, όπου C_1 είναι η σταθερά στην* (6.41)*, έχουμε ότι υπάρχει σταθερά C, ανεξάρτητη των k και h, τέτοια ώστε*

$$\max_{0\le n\le M}\|U^n - u(\cdot,t^n)\| \le C(k + h^2). \tag{6.48}$$

Απόδειξη. Γράφουμε και πάλι $U^n - u(\cdot, t^n) = \vartheta^n + \rho^n$ με $\rho^n = W(\cdot, t^n) - u(\cdot, t^n)$ και $\vartheta^n = U^n - W(\cdot, t^n)$. Όπως και προηγουμένως, λόγω της (6.25), αρκεί να εκτιμήσουμε την $\|\vartheta^n\|$. Μπορούμε να δούμε ότι το ϑ^n ικανοποιεί την ακόλουθη σχέση

$$(\frac{\vartheta^n - \vartheta^{n-1}}{k}, \chi) + ((\vartheta^{n-1})', \chi') = -(\omega^n, \chi), \quad \forall \chi \in V_h, \tag{6.49}$$

όπου $\omega^n = \omega_1^n + \omega_2^n$ και

$$\omega_1^n = \frac{\rho^n - \rho^{n-1}}{k}, \quad \omega_2^n = \frac{u^n - u^{n-1}}{k} - u_t^{n-1}.$$

Παρόμοια, όπως και στο Θεώρημα 6.4, έχουμε

$$\begin{aligned}
&(\frac{\vartheta^n - \vartheta^{n-1}}{k}, \chi) + ((\vartheta^{n-1})', \chi') \\
&= -(\frac{R_h u^n - R_h u^{n-1}}{k}, \chi) - ((R_h u^{n-1})', \chi') \\
&= -(\frac{\rho^n - \rho^{n-1}}{k}, \chi) - (\frac{u^n - u^{n-1}}{k}, \chi) - ((u^{n-1})_x, \chi') \\
&= -(\frac{\rho^n - \rho^{n-1}}{k}, \chi) - (\frac{u^n - u^{n-1}}{k}, \chi) + (u_t^{n-1}, \chi) \\
&= -(\omega_1^n + \omega_2^n, \chi),
\end{aligned}$$

Από την (6.27) έχουμε ότι $\|\omega_1^n\| \leq Ch^2$. Επίσης, λόγω του αναπτύγματος Taylor, παίρνουμε

$$\|\omega_2^n\| = \|\frac{u^n - u^{n-1}}{k} - u_t^{n-1}\| = k^{-1}\| \int_{t^{n-1}}^{t^n} (t - t^n) u_{tt}(\cdot, t)\, dt\|.$$

Συνεπώς, λαμβάνουμε

$$\|\omega_2^n\| \leq Ck. \tag{6.50}$$

Άρα, συνδυάζοντας τις (6.27) και (6.50), έχουμε

$$\|\omega_n\| \leq C(k + h^2), \quad n = 1, \ldots, M. \tag{6.51}$$

Στη συνέχεια, θα ακολουθήσουμε τα βήματα της απόδειξης του Θεωρήματος 6.7. Επομένως, επιλέγουμε $\chi = \vartheta^{n-1}$ στην (6.49) και λαμβάνουμε

$$\begin{aligned}
(\vartheta^n, \vartheta^{n-1}) - \|\vartheta^{n-1}\|^2 + k\|(\vartheta^{n-1})'\|^2 &= -k(\omega^n, \vartheta^{n-1}) \\
&\leq k\|\omega^n\|\|\vartheta^{n-1}\|.
\end{aligned} \tag{6.52}$$

Επειδή

$$2(\vartheta^n, \vartheta^{n-1}) = \|\vartheta^n\|^2 + \|\vartheta^{n-1}\|^2 - \|\vartheta^n - \vartheta^{n-1}\|^2.$$

η (6.52) γίνεται

$$\frac{1}{2}(\|\vartheta^n\|^2 - \|\vartheta^{n-1}\|^2 - \|\vartheta^n - \vartheta^{n-1}\|^2) + k\|(\vartheta^{n-1})'\|^2 \leq k\|\omega^n\|\|\vartheta^{n-1}\|. \quad (6.53)$$

Στη συνέχεια, επιλέγοντας $\chi = \vartheta^n - \vartheta^{n-1}$ στην (6.49), έχουμε

$$\|\vartheta^n - \vartheta^{n-1}\|^2 - k((\vartheta^{n-1})', (\vartheta^n)' - (\vartheta^{n-1})') = -k(\omega^n, \vartheta^n - \vartheta^{n-1}). \quad (6.54)$$

Στη συνέχεια, χρησιμοποιώντας την ανισότητα Cauchy–Schwarz και την (6.41), η (6.54) δίνει

$$\begin{aligned}\|\vartheta^n - \vartheta^{n-1}\|^2 &\leq k\|(\vartheta^{n-1})'\|\,\|(\vartheta^n)' - (\vartheta^{n-1})'\| + k\|\omega^n\|\,\|\vartheta^n - \vartheta^{n-1}\| \\ &\leq C_1\frac{k}{h}\|(\vartheta^{n-1})'\|\,\|\vartheta^n - \vartheta^{n-1}\| + k\|\omega^n\|\,\|\vartheta^n - \vartheta^{n-1}\|.\end{aligned}$$

Επομένως, λαμβάνουμε

$$\|\vartheta^n - \vartheta^{n-1}\| \leq C_1\frac{k}{h}\|(\vartheta^{n-1})'\| + k\|\omega^n\|.$$

Στη συνέχεια, υψώνοντας στο τετράγωνο και τα δύο μέλη της προηγούμενης ανισότητας παίρνουμε

$$\|\vartheta^n - \vartheta^{n-1}\|^2 \leq C_1^2\frac{k^2}{h^2}\|(\vartheta^{n-1})'\|^2 + k^2\|\omega^n\|^2 + 2C_1\frac{k^2}{h}\|(\vartheta^{n-1})'\|\,\|\omega^n\|. \quad (6.55)$$

Χρησιμοποιώντας τώρα την (6.55) στην (6.53) παίρνουμε

$$\begin{aligned}\|\vartheta^n\|^2 - \|\vartheta^{n-1}\|^2 \leq (C_1^2\frac{k}{h^2} - 2)k\|(\vartheta^{n-1})'\|^2 + k^2\|\omega^n\|^2 \\ + 2C_1\frac{k^2}{h}\|(\vartheta^{n-1})'\|\,\|\omega^n\| + 2k\|\omega^n\|\|\vartheta^{n-1}\|.\end{aligned}$$

Στη συνέχεια, λόγω της υπόθεσης $C_1^2\dfrac{k}{h^2} - 2 \leq 0$, παίρνουμε

$$\|\vartheta^n\|^2 - \|\vartheta^{n-1}\|^2 \leq k^2\|\omega^n\|^2 + 2C_1\frac{k^2}{h}\|(\vartheta^{n-1})'\|\,\|\omega^n\| + 2k\|\omega^n\|\|\vartheta^{n-1}\|. \quad (6.56)$$

Εύκολα μπορούμε να δούμε ότι για κάθε $\epsilon > 0$ ισχύει η ανισότητα $2ab \leq (\epsilon a)^2 + (1/\epsilon)b^2$ με $a, b \in \mathbb{R}$. Έτσι, επειδή $k \leq 1$, υπάρχουν σταθερές c και C, τέτοιες ώστε από την (6.56) έχουμε

$$\|\vartheta^n\|^2 \leq (1 + ck)\|\vartheta^{n-1}\|^2 + Ck\|\omega^n\|^2. \quad (6.57)$$

Αν θέσουμε $\bar{\omega} = \max_{0\le n\le M} \|\omega^n\|$, από την (6.57) προκύπτει τώρα

$$\begin{aligned}\|\vartheta^n\|^2 &\le (1+ck)^n\|\vartheta^0\|^2 + Ck\sum_{j=0}^{n-1}(1+ck)^j\|\omega^n\|^2\\ &\le e^{ckn}\|\vartheta^0\|^2 + Ck\bar{\omega}^2\sum_{j=0}^{M-1}e^{ckj}\\ &\le e^{ckn}\|\vartheta^0\|^2 + Ck\bar{\omega}^2\frac{e^{ckM}-1}{e^{ck}-1}\\ &\le e^{cT}\|\vartheta^0\|^2 + \frac{C}{c}\bar{\omega}^2.\end{aligned}$$

Συνεπώς, λόγω της (6.51) έχουμε

$$\|\vartheta^n\|^2 \le C\|\vartheta^0\|^2 + C(k+h^2)^2.$$

Επομένως, λόγω της (6.25) και του γεγονότος ότι $\vartheta(0) = 0$, έχουμε ότι ισχύει η επιθυμητή σχέση (6.48). □

6.5 Άλλες μέθοδοι και προβλήματα

Μέθοδοι πεπερασμένων στοιχείων για προβλήματα παραβολικών εξισώσεων με συνορικακές συνθήκες τύπου Dirichlet ή και με άλλες συνοριακές συνθήκες όπως π.χ. τύπου Neumann, αναλύονται με παρόμοιο τρόπο όπως με αυτά που παρουσιάσαμε σε αυτό το κεφάλαιο. Για τη μελέτη αυτών των αριθμητικών μεθόδων καθώς και για μία πιο λεπτομερή παρουσίαση των αποτελεσμάτων αυτού του κεφαλαίου, παραπέμπουμε τον αναγνώστη στα (Ακρίβης & Δουγαλής, 2005· Ακρίβης, 2005· Johnson, 1987· Larsson & Thomée, 2009· Morton & Mayers, 2005· Strikwerda, 2004).

6.6 Ασκήσεις

6.1. Θεωρούμε το πρόβλημα

$$\begin{aligned} u_t(x,t) &= u_{xx}(x,t) + f(x,t), && x\in[0,L],\ t\in[0,T],\\ u(0,t) &= u(L,t) = 0, && t\in[0,T],\\ u(x,0) &= g(x), && x\in[0,L],\end{aligned} \tag{6.58}$$

όπου $L>0$, $f, g \in C[0,L]$. Διατυπώστε και αποδείξτε τα θεωρήματα ευστάθειας και σύγκλισης για τις μεθόδους πεπερασμένων στοιχείων χρησιμοποιώντας για τη διακριτοποίηση ως προς τον χρόνο την πεπλεγμένη μέθοδο του Euler και τη μέθοδο Crank–Nicolson.

6.2. Θεωρούμε το πρόβλημα

$$\begin{aligned} u_t(x,t) &= u_{xx}(x,t), && x \in [0,L],\ t \in [0,T], \\ u_x(0,t) &= u_x(L,t) = 0, && t \in [0,T], \\ u(x,0) &= g(x), && x \in [0,L], \end{aligned} \tag{6.59}$$

όπου $L > 0$, $g \in C[0,L]$. Διατυπώστε τις μεθόδους πεπερασμένων στοιχείων χρησιμοποιώντας για τη διακριτοποίηση ως προς τον χρόνο τη πεπλεγμένη μέθοδο του Euler, τη μέθοδο Crank–Nicolson και την άμεση μέθοδο του Euler.

6.3. Έστω $N \in \mathbb{N}$, $h = L/(N+1)$, και $x_i = ih$, $i = 0, \ldots, N+1$, ένας διαμερισμός του $[0,L]$. Τότε για τη συνάρτηση

$$v(x) = \begin{cases} \dfrac{x - x_0}{h}, & x_0 \le x \le x_1, \\ \dfrac{x_2 - x}{h}, & x_1 \le x \le x_2, \\ 0, & \text{διαφορετικά}, \end{cases}$$

έχουμε $\|v'\| \le \sqrt{3}\, h^{-1} \|v\|$.

Βιβλιογραφία

Ακρίβης, Γ. (2005). *Μέθοδοι Πεπερασμένων Στοιχείων.* Λευκωσία, Κύπρος. (Πανεπιστημιακές Σημειώσεις).

Ακρίβης, Γ., & Δουγαλής, Β. (2005). *Αριθμητικές Μέθοδοι για Μερικές Διαφορικές Εξισώσεις.* Ιωάννινα. (Πανεπιστημιακές Σημειώσεις).

Ακρίβης, Γ., & Δουγαλής, Β. (2015). *Εισαγωγή στην Αριθμητική Ανάλυση*. Πανεπιστημιακές Εκδόσεις Κρήτης, Ηράκλειο.

Johnson, C. (1987). *Numerical solution of partial differential equations by the finite element method*. Cambridge University Press, Cambridge.

Larsson, S., & Thomée, V. (2009). *Partial differential equations with numerical methods* (Vol. 45). Springer-Verlag, Berlin.

Morton, K. W., & Mayers, D. F. (2005). *Numerical solution of partial differential equations* (Second ed.). Cambridge University Press, Cambridge.

Strikwerda, J. C. (2004). *Finite difference schemes and partial differential equations* (Second ed.). Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA.

Κεφάλαιο 7

Εξίσωση μεταφοράς

Σε αυτό το κεφάλαιο θεωρούμε μια απλή διαφορική εξίσωση πρώτης τάξης, την εξίσωση μεταφοράς, στη μια διάσταση ως προς τον χώρο και ως προς τον χρόνο. Θα κατασκευάσουμε μεθόδους πεπερασμένων διαφορών για το αντίστοιχο πρόβλημα αρχικών τιμών και συνοριακών τιμών, καθώς και θα θεωρήσουμε κατάλληλες συνθήκες για να είναι μέθοδοι ευσταθείς και να συγκλίνουν στην ακριβή λύση.

7.1 Μέθοδοι upwind και downwind

Θεωρούμε το ακόλουθο πρόβλημα: Ζητείται μια συνάρτηση $u \in C^1([a,b]\times[0,T])$, τέτοια ώστε

$$\begin{aligned} u_t(x,t) + \alpha u_x(x,t) = 0, \quad & x \in [a,b],\ t \in [0,T], \\ u(x,0) = g(x), \quad & x \in [a,b], \\ u(a,t) = \phi_0(t), \quad & t \in [0,T], \end{aligned} \tag{7.1}$$

όπου $\alpha, L, T > 0$ και g μια δοσμένη συνάρτηση. Σύμφωνα με την Παράγραφο 1.2.3 η ακριβής λύση του (7.1), προσδιορίζεται από τις αντίστοιχες αρχικές και συνοριακές συνθήκες. Επομένως, φαίνεται ότι μια αριθμητική μέθοδος δεν είναι απαραίτητη για την προσέγγιση της ακριβούς λύσης. Όμως, η μελέτη αριθμητικών μεθόδων για αυτά τα απλά προβλήματα φανερώνει ιδιότητες και συνθήκες που πρέπει να ικανοποιούνται για τη σύγκλιση των προσεγγίσεων στην ακριβή λύση και οδηγεί στην καλύτερη κατανόηση των μεθόδων αυτών αλλά και σε αντίστοιχες ιδιότητες για μη γραμμικά προβλήματα πρώτης τάξεως, όπως π.χ. το (1.8).

7.1.1 Μέθοδος upwind

Θεωρούμε έναν διαμερισμό του $[a,b]\times[0,T]$ ανάλογα όπως στο Κεφάλαιο 5. Συμβολίζουμε, λοιπόν, x_i, $i=0,\ldots,N+1$, τα $N+2$ ισαπέχοντα σημεία, όπου $x_i=a+ih$, με $h=(b-a)/(N+1)$ και t^j, $j=0,\ldots,M$, τα $M+1$ ισαπέχοντα σημεία, όπου $t^j=jk$, με $k=T/M$. Σε κάθε σημείο (x_i,t^j) του διαμερισμού του $[a,b]\times[0,T]$ θα ισχύει

$$u_t(x_i,t^j)+\alpha u_x(x_i,t^j)=0,\quad i=0,\ldots,N+1,\ j=0,\ldots,M, \tag{7.2}$$

και, στη συνέχεια, θα κατασκευάσουμε και πάλι προσεγγίσεις U_i^j της $u(x_i,t^j)$ προσεγγίζοντας τις παραγώγους της u στην (7.2) χρησιμοποιώντας πεπερασμένες διαφορές που θεωρήσαμε στο Κεφάλαιο 2.

Λόγω των αρχικών και συνοριακών συνθηκών στην (7.1) θέτουμε $U_i^0=g(x_i)$, $i=0,\ldots,N+1$, και $U_0^j=\phi_0(t^j)$, $j=0,\ldots,M$. Στη συνέχεια, χρησιμοποιώντας τις $\delta_k^+u(x_i,t^j)$ και $\delta_h^-u(x_i,t^j)$ που θεωρήσαμε στην (2.1) για να προσεγγίσουμε τις $u_t(x_i,t^j)$ και $u_x(x_i,t^j)$, αντίστοιχα, έχουμε για $i=1,\ldots,N+1$, και $j=0,\ldots,M-1$,

$$\frac{u(x_i,t^{j+1})-u(x_i,t^j)}{k}+\alpha\frac{u(x_i,t^j)-u(x_{i-1},t^j)}{h}=\eta_i^j, \tag{7.3}$$

όπου, αν $u_{tt},u_{xx}\in C([a,b]\times[0,T])$, λόγω της (2.3),

$$|\eta_i^j|\le\frac{k}{2}\max_{t\in[0,T]}|u_{tt}(x,t)|+\frac{h}{2}\max_{x\in[a,b]}|u_{xx}(x,t)|. \tag{7.4}$$

Συνεπώς, ισχύει το ακόλουθο λήμμα.

Λήμμα 7.1. *Έστω u η λύση του* (7.1) *με $u_{tt},u_{xx}\in C([a,b]\times[0,T])$. Τότε, για την η_i^j που δίνεται στην* (7.3) *έχουμε ότι υπάρχει σταθερά C ανεξάρτητη των k και h, τέτοια ώστε*

$$\max_{0\le j\le M-1}\max_{1\le i\le N+1}|\eta_i^j|\le C(k+h). \tag{7.5}$$

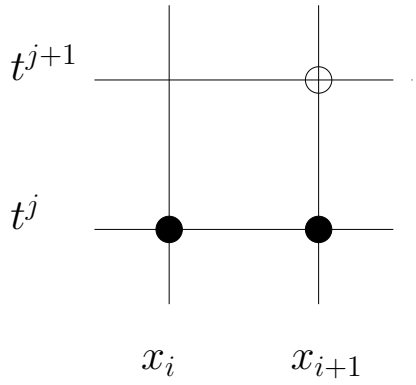
Κατασκευάζουμε λοιπόν προσεγγίσεις U_i^j των τιμών $u(x_i,t^j)$, σύμφωνα με την ακόλουθη μέθοδο, την οποία καλούμε *upwind*,

$$\begin{aligned}\frac{U_i^{j+1}-U_i^j}{k}+\alpha\frac{U_i^j-U_{i-1}^j}{h}&=0,\quad i=1,\ldots,N+1,\ j=0,\ldots,M-1,\\ U_i^0&=g(x_i),\quad i=0,\ldots,N+1,\\ U_0^j&=\phi_0(t^j),\quad j=0,\ldots,M.\end{aligned} \tag{7.6}$$

Αν συμβολίσουμε τώρα με λ τον λόγο $\dfrac{\alpha k}{h}$, η (7.6) μπορεί να γραφεί

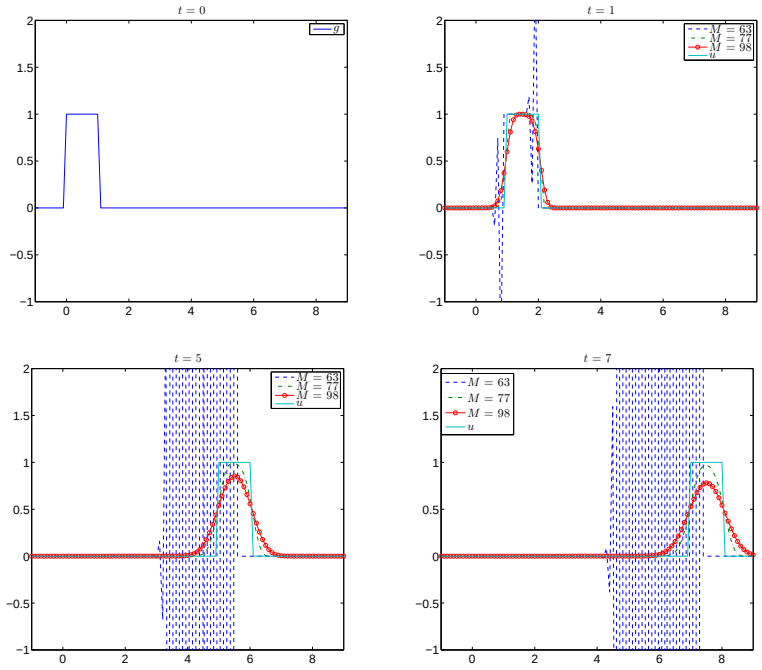
$$\begin{aligned} U_i^{j+1} &= (1-\lambda)U_i^j + \lambda U_{i-1}^j, \quad && i = 1, \ldots, N+1,\ j = 0, \ldots, M-1, \\ U_i^0 &= g(x_i), && i = 0, \ldots, N+1, \\ U_0^j &= \phi_0(t^j), && j = 0, \ldots, M. \end{aligned} \tag{7.7}$$

Εύκολα μπορούμε να δούμε ότι ξεκινώντας από το χρονικό επίπεδο $t = 0$, χρησιμοποιώντας την (7.7), μπορούμε να βρούμε με άμεσο τρόπο την προσέγγιση U_i^1, $i = 0, \ldots, N+1$, στο επόμενο χρονικό επίπεδο $t^1 = k$ και να συνεχίσουμε με αυτό τον τρόπο, ώστε να βρούμε τις προσεγγίσεις και σε όλα τα επόμενα χρονικά επίπεδα t^j, $j = 1, \ldots, M$.



Σχήμα 7.1: Σημεία ενός πλέγματος για τον υπολογισμό της λύσης με τη μέθοδο upwind. Με άσπρο σημειώνουμε τα σημεία που αντιστοιχούν σε άγνωστες τιμές της προσέγγισης και με μαύρο αυτά που αντιστοιχούν σε γνωστές τιμές.

Παράδειγμα 7.1. Θεωρούμε το πρόβλημα (7.1) στο διάστημα $[-1, 9]$, με g όπως στο Παράδειγμα 1.2, $\phi_0(t) = g(x - \alpha t)$ και $\alpha = 1$. Επειδή γνωρίζουμε ότι $u(a, t) = 0$, για $t \in [0, 7]$, μπορούμε να θέσουμε $U_0^j = 0$, $j = 0, \ldots, M$. Επίσης, διαμερίζουμε το $[-1, 9]$ σε $N + 2$ σημεία, με $N = 99$ και το $[0, 7]$ σε $M + 1$, με $M = 63, 77, 98$. Στο Σχήμα 7.1 βλέπουμε την ακριβή λύση u και τις προσεγγίσεις με τη μέθοδο upwind για $t = 0, 1, 5, 7$. Παρατηρούμε ότι αν $\lambda = 70/63 > 1$, για $M = 63$, τότε το σφάλμα της προσεγγιστικής λύσης είναι μεγάλο, ενώ για $\lambda = 70/77 < 1$, για $M = 77$, προκύπτει το μικρότερο σφάλμα. Επίσης, βλέπουμε ότι για το μικρότερο λ, το οποίο προκύπτει για $M = 98$, η προσέγγιση χειροτερεύει.



Σχήμα 7.2: Η ακριβής λύση $u(x,t)$ του Παραδείγματος 7.1 και οι προσεγγίσεις της για $N = 99$ και $M = 63, 77, 98$, με τη μέθοδο upwind για $t = 0, 1, 5$ και 7.

7.1.2 Μέθοδος downwind

Αν για την προσέγγιση της $u_x(x_i, t^j)$ στην (7.2) χρησιμοποιήσουμε την $\delta_h^+ u(x_i, t^j)$ αντί για την $\delta_h^- u(x_i, t^j)$, προκύπτει για $i = 0, \ldots, N$, και $j = 0, \ldots, M-1$, η εξίσωση

$$\frac{u(x_i, t^{j+1}) - u(x_i, t^j)}{k} + \alpha \frac{u(x_{i+1}, t^j) - u(x_i, t^j)}{h} = \eta_i^j, \tag{7.8}$$

όπου, αν $u_{tt}, u_{xx} \in C([a,b] \times [0,T])$, λόγω της (2.3),

$$|\eta_i^j| \leq \frac{k}{2} \max_{t \in [0,T]} |u_{tt}(x,t)| + \frac{h}{2} \max_{x \in [a,b]} |u_{xx}(x,t)|. \tag{7.9}$$

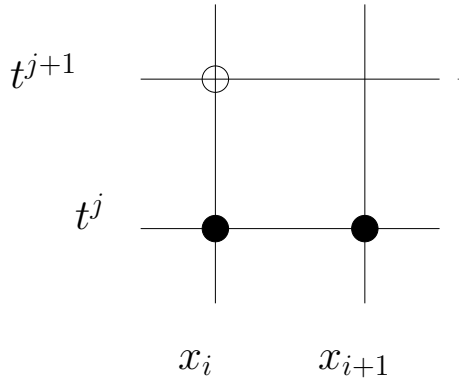
Όπως και στη μέθοδο upwind, εύκολα βλέπουμε ότι το σφάλμα η_i^j στην (7.8) θα ικανοποιεί το Λήμμα 7.1. Στη συνέχεια, θεωρούμε την ακόλουθη μέθοδο για

τον υπολογισμό των U_i^j, την οποία καλούμε *downwind*,

$$\begin{aligned} \frac{U_i^{j+1}-U_i^j}{k}+\alpha\frac{U_{i+1}^j-U_i^j}{h}=0, \quad i=0,\ldots,N,\ j=0,\ldots,M-1,\\ U_i^0=g(x_i), \quad i=0,\ldots,N+1,\\ U_0^j=\phi_0(t^j), \quad j=0,\ldots,M. \end{aligned} \tag{7.10}$$

Αν συμβολίσουμε και πάλι λ τον λόγο $\dfrac{\alpha k}{h}$, τότε η (7.10) μπορεί να γραφεί

$$\begin{aligned} U_i^{j+1}=(1+\lambda)U_i^j-\lambda U_{i+1}^j, \quad i=0,\ldots,N,\ j=0,\ldots,M-1,\\ U_i^0=g(x_i), \quad i=0,\ldots,N+1,\\ U_0^j=\phi_0(t^j), \quad j=0,\ldots,M. \end{aligned} \tag{7.11}$$



Σχήμα 7.3: Σημεία ενός πλέγματος για τον υπολογισμό της λύσης με τη μέθοδο downwind. Με άσπρο σημειώνουμε τα σημεία που αντιστοιχούν σε άγνωστες τιμές της προσέγγισης και με μαύρο αυτά που αντιστοιχούν σε γνωστές τιμές.

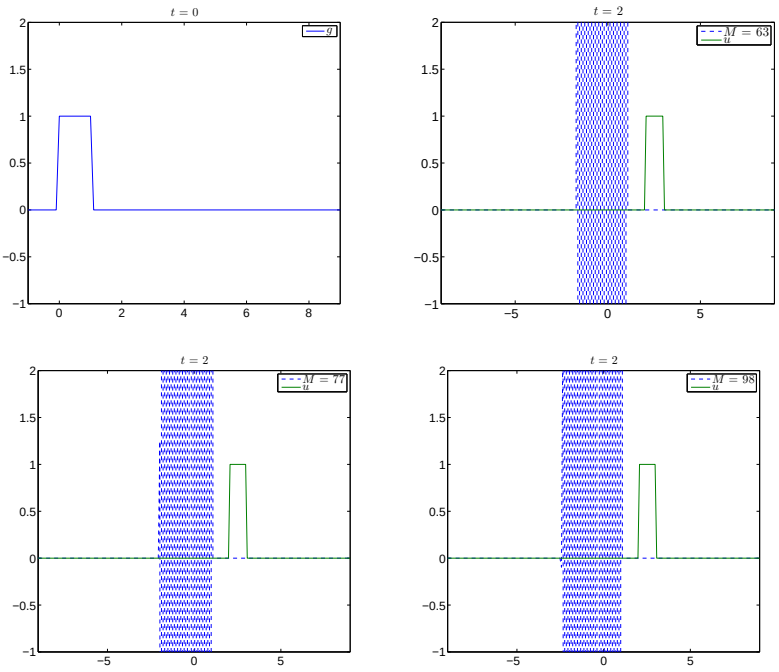
Αν και οι δύο παραπάνω μέθοδοι upwind και downwind μοιάζουν, έχουν πολύ διαφορετική συμπεριφορά και δεν δίνουν παρόμοια αποτελέσματα. Πράγματι, ξεκινώντας από το χρονικό επίπεδο $t=0$ όπου γνωρίζουμε ότι $U_i^0=g(x_i)$, $i=0,\ldots,N+1$, και χρησιμοποιώντας την (7.11), μπορούμε να βρούμε με άμεσο τρόπο την προσέγγιση U_i^1, $i=0,\ldots,N$, στο επόμενο χρονικό επίπεδο $t^1=k$, όμως παρατηρούμε ότι δεν μπορούμε να υπολογίσουμε την U_{N+1}^1. Επομένως, δεν μπορούμε να χρησιμοποιήσουμε τη μέθοδο (7.11), όπως την έχουμε θεωρήσει, για να βρούμε όλες τις τιμές U_i^j στα επόμενα χρονικά επίπεδα t^j, $j=1,\ldots,M$. Μια λύση αυτού του προβλήματος είναι να θεωρήσουμε γνωστές τις τιμές U_{N+1}^j,

$j = 1, \ldots, M$. Έτσι, αν γνωρίζουμε ότι $u(b,t) = \phi_1(t), t \in [0,T]$, για μια δοσμένη συνάρτηση ϕ_1, θέτουμε τότε $U_{N+1}^1 = \phi(t^j), j = 0, \ldots, M$. Έτσι, η μέθοδος (7.11) γίνεται

$$\begin{aligned} U_i^{j+1} &= (1+\lambda)U_i^j - \lambda U_{i+1}^j, \quad i = 0, \ldots, N,\ j = 0, \ldots, M-1, \\ U_i^0 &= g(x_i), \quad i = 0, \ldots, N+1, \\ U_{N+1}^j &= \phi_1(t^j), \quad j = 0, \ldots, M. \end{aligned} \tag{7.12}$$

Τώρα χρησιμοποιώντας την (7.12), μπορούμε να βρούμε με άμεσο τρόπο τις προσέγγισεις U_i^j, σε όλα τα χρονικά επίπεδα t^j, $j = 1, \ldots, M$.

Παράδειγμα 7.2. Ας θεωρήσουμε και πάλι το Παράδειγμα 7.1. Είναι απλό να δούμε ότι $u(b,t) = 0$, για $t \in [0,7]$, επομένως μπορούμε να θέσουμε $U_{N+1}^j = 0$, $j = 0, \ldots, M$. Για λόγο που θα φανεί από τα αριθμητικά αποτελέσματα, ας θεωρήσουμε ένα μεγαλύτερο διάστημα από αυτό του Παραδείγματος 7.1. Έτσι διαμερίζουμε το διάστημα $[-11, 9]$ σε $N+2$ σημεία, με $N = 199$, έτσι ώστε το βήμα h να είναι το ίδιο όπως στο Παράδειγμα 7.1. Επίσης, διαμερίζουμε το $[0,7]$ σε $M+1$ σημεία, με $M = 63, 77, 98$. Στο Σχήμα 7.2 βλέπουμε την ακριβή λύση u για $t = 0$ και τις προσεγγίσεις με τη μέθοδο downwind για $t = 2$. Παρατηρούμε ότι η προσέγγιση με τη μέθοδο downwind (7.12) και στις τρεις περιπτώσεις δίνει εσφαλμένες προσεγγίσεις για την ακριβή λύση.



Σχήμα 7.4: Η ακριβής λύση $u(x,t)$ του Παραδείγματος 7.2 και οι προσεγγίσεις της για $N = 199$ και $M = 63, 77, 98$, με τη μέθοδο downwind για $t = 0$ και 2.

Θεωρούμε τώρα το ανάλογο πρόβλημα με το (7.1) όπου τώρα $\alpha < 0$, δηλαδή: Ζητείται μια συνάρτηση $u \in C^1([a,b] \times [0,T])$, τέτοια ώστε

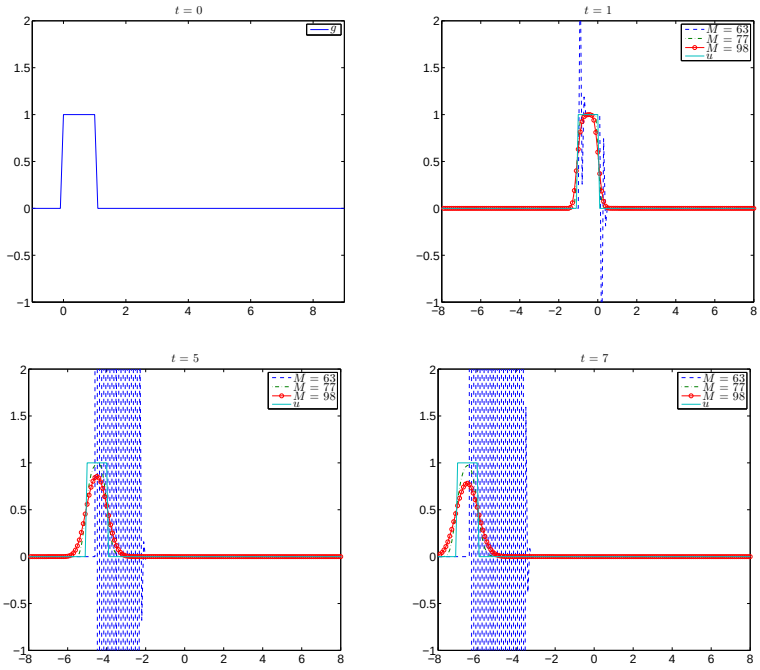
$$\begin{aligned} u_t(x,t) + \alpha u_x(x,t) &= 0, \qquad && x \in [a,b],\ t \in [0,T], \\ u(x,0) &= g(x), && x \in [a,b], \\ u(b,t) &= \phi_1(t), && t \in [0,T], \end{aligned} \tag{7.13}$$

όπου $T > 0$ και g, ϕ_1 δοσμένες συναρτήσεις.

Για αυτό το πρόβλημα θεωρούμε τη μέθοδο downwind (7.12) και, σύμφωνα με το επόμενο παράδειγμα, παρατηρούμε μια διαφορετική συμπεριφορά της μεθόδου από το Παράδειγμα 7.2.

Παράδειγμα 7.3. Θεωρούμε το πρόβλημα (7.13) στο διάστημα $[-11, 9]$, με g όπως στο Παράδειγμα 1.2, $\phi_1(t) = g(x - \alpha t)$ και $\alpha = -1$. Επίσης, διαμερίζουμε το διάστημα $[-11, 9]$ σε $N+2$ σημεία, με $N = 199$ και το $[0, 7]$ σε $M+1$ σημεία, με $M = 63, 77, 98$, όπως και στο Παράδειγμα 7.2. Επειδή γνωρίζουμε ότι $u(b,t) = 0$, για $t \in [0, 7]$, μπορούμε να θέσουμε $U^j_{N+1} = 0$, $j = 0, \ldots, M$. Στο Σχήμα 7.3

βλέπουμε την ακριβή λύση u και τις προσεγγίσεις με τη μέθοδο downwind για $t = 0, 1, 5, 7$. Παρατηρούμε ότι η μεθόδος downwind σε αυτό το παράδειγμα έχει παρόμοια συμπεριφορά με τη μέθοδο upwind στο Παράδειγμα 7.1.



Σχήμα 7.5: Η ακριβής λύση $u(x, t)$ του Παραδείγματος 7.3 και οι προσεγγίσεις της για $N = 199$ και $M = 63, 77, 98$, με τη μέθοδο downwind για $t = 0, 1, 5$ και 7.

Παρατηρούμε λοιπόν ότι αν η σταθερά α είναι θετική, τότε η μέθοδος upwind δίνει καλύτερα αποτελέσματα από τη μέθοδο downwind. Ενώ αν η σταθερά α είναι αρνητική, τότε η downwind δίνει καλύτερα αποτελέσματα από την upwind. Στην επόμενη παράγραφο θα εξηγήσουμε αυτή τη συμπεριφορά.

7.1.3 Χωρίο υπολογιστικής εξάρτησης

Ως πρώτο βήμα για την κατανόηση των αποτελεσμάτων των Παραδειγμάτων 7.1–7.3, δίνουμε τον ακόλουθο ορισμό.

Ορισμός 7.1. Για κάθε σημείο (x_i, t^j) του πλέγματος στο οποίο προσεγγίζουμε τη λύση u, μπορούμε να ορίσουμε τα σημεία $(x_\ell, 0)$, τέτοια ώστε οι τιμές U_ℓ^0 καθορίζουν την τιμή της U_i^j, σύμφωνα με το αριθμητικό σχήμα που χρησιμοποιούμε. Το

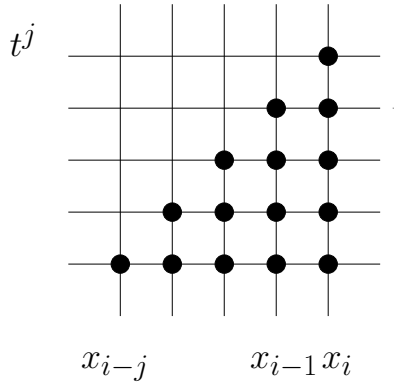
σύνολο αυτών των σημείων $(x_\ell, 0)$, το ονομάζουμε *χωρίο υπολογιστικής εξάρτησης* του (x_i, t^j), για το σχήμα που χρησιμοποιούμε.

Εύκολα παρατηρούμε ότι, χρησιμοποιώντας την μέθοδο upwind, η τιμή της προσέγγισης στο σημείο (x_i, t^j), εξαρτάται τελικά από τις τιμές στα σημεία $(x_{i-j}, 0)$, $\ldots$, $(x_i, 0)$. Ανάλογα, το χωρίο εξάρτησης της προσέγγισης στο (x_i, t^j) με τη μέθοδο downwind αποτελείται από τα σημεία $(x_i, 0), \ldots, (x_{i+j}, 0)$, βλ. Σχήματα 7.6 και 7.7, αντίστοιχα. Γνωρίζουμε ότι η ακριβής λύση στο σημείο (x_i, t^j) εξαρτάται μόνο από την τιμή της $u(x_i - \alpha t^j, 0) = g(x_i - \alpha t^j)$. Άρα, η τιμή της ακριβούς λύσης στο σημείο (x_i, t^j) καθορίζεται μοναδικά από την τιμή στο σημείο $\bar{x}_0 = x_i - \alpha t^j$ του άξονα των x. Για αυτόν ακριβώς τον λόγο, επιθυμούμε το $\bar{x}_0$ να περιέχεται στο χωρίο υπολογιστικής εξάρτησης της μεθόδου. Αυτή η ιδιότητα δίνεται από τον ακόλουθο ορισμό.

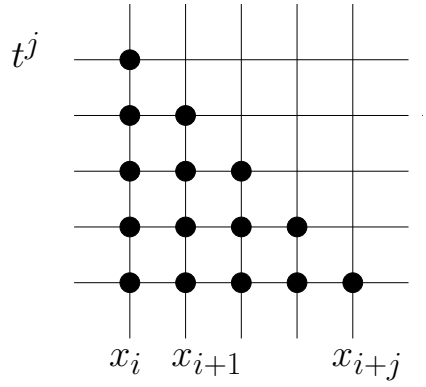
Ορισμός 7.2. *Συνθήκη CFL* καλούμε την ιδιότητα που ικανοποιεί μια μέθοδος αν το διάστημα ή τα σημεία εξάρτησης της ακριβούς λύσης σε ένα σημείο του πεδίου ορισμού της, περιέχονται στο *χωρίο υπολογιστικής εξάρτησης* αυτής της μεθόδου.

Για να ικανοποείται η συνθήκη CFL, για την μέθοδο upwind αρκεί το διάστημα $[x_{i-j}, x_i]$, το οποίο περιέχει τα σημεία του χωρίου υπολογιστικής εξάρτησης $\{x_{i-j}, \ldots, x_i\}$, να περιέχει το $\bar{x}_0 = x_i - \alpha t^j$. Επομένως, αρκεί $\alpha > 0$ και $x_i - jh \leq \bar{x}_0$, το οποίο ισχύει αν $\lambda = \dfrac{\alpha k}{h} \leq 1$.

Στην περίπτωση τώρα της μεθόδου downwind, μπορούμε να δούμε ότι το διάστημα $[x_i, x_{i+j}]$, το οποίο περιέχει το αντίστοιχο χωρίο υπολογιστικής εξάρτησης, δεν περιέχει το σημείο $\bar{x}_0$, αν $\alpha > 0$, και, άρα, η μέθοδος δεν ικανοποιεί τη συνθήκη CFL. Αντίθετα, αν $\alpha < 0$, τότε $\bar{x}_0 \leq x_i + jh$ αν ισχύει $\lambda = \dfrac{\alpha k}{h} \geq -1$.



Σχήμα 7.6: Χωρίο εξάρτησης της προσεγγιστικής λύσης με τη μέθοδο upwind.



Σχήμα 7.7: Χωρίο εξάρτησης της προσεγγιστικής λύσης με τη μέθοδο downwind.

7.1.4 Ευστάθεια και σύγκλιση

Θα μελετήσουμε τώρα την *ευστάθεια von Neumann* των μεθόδων upwind και downwind που θεωρήσαμε στις προηγούμενες παραγράφους. Αν η αρχική συνάρτηση g είναι μια ημιτονοειδής συνάρτηση, π.χ. $g(x) = \sin(x)$, τότε η ακριβής λύση του προβλήματος (7.1) θα έχει την ίδια μορφή και το ίδιο πλάτος ανεξάρτητα από το πρόσημο της σταθεράς α. Θεωρούμε, λοιπόν, ότι η προσεγγιστική λύση είναι της μορφής $U_i^j = w_j \sin(x_i)$ και θα βρούμε συνθήκες, ώστε το πλάτος w_j του ημιτονοειδούς "κύματος" να παραμένει φραγμένο.

Στην περίπτωση της μεθόδου upwind έχουμε, σύμφωνα με την (7.7),

$$w_{j+1} \sin(x_i) = (1 - \lambda) w_j \sin(x_i) + \lambda w_j \sin(x_{i-1}).$$

Χρησιμοποιώντας την τριγωνομετρική ταυτότητα (5.18), έχουμε

$$w_{j+1} \sin(x_i) = (1 - \lambda) w_j \sin(x_i) + \lambda w_j (\sin(x_i) \cos(h) - \sin(h) \cos(x_i)).$$

Συνεπώς

$$w_{j+1} \sin(x_i) = w_j ((1 - \lambda(1 - \cos(h)) \sin(x_i) - \lambda \sin(h) \cos(x_i))).$$

Αν θέσουμε τώρα $C_1 = 1 - \lambda(1 - \cos(h))$ και $C_2 = -\lambda \sin(h)$, τότε υπάρχουν σταθερές A και $\phi \in [0, 2\pi)$, τέτοιες ώστε

$$C_1 \sin(x_i) + C_2 \cos(x_i) = A \sin(x_i + \phi)$$

όπου $A^2 = C_1^2 + C_2^2$ και $\tan(\phi) = C_1/C_2$. Επομένως

$$w_{j+1} \sin(x_i) = A w_j \sin(x_i + \phi).$$

Οπότε για να είναι φραγμένη η προσεγγιστική λύση U_i^j πρέπει $|A| \leq 1$. Έχουμε, λοιπόν,

$$\begin{aligned} A^2 = C_1^2 + C_2^2 &= (1 - \lambda(1 - \cos(h)))^2 + \lambda^2 \sin(h)^2 \\ &= (1-\lambda)^2 + \lambda^2 + 2\lambda(1-\lambda)\cos(h) \\ &= 1 - 2\lambda(1-\lambda)(1-\cos(h)) = 1 - 4\lambda(1-\lambda)\sin^2(h/2). \end{aligned} \tag{7.14}$$

Είναι προφανές ότι, αν $0 \leq \lambda(1-\lambda)$, τότε η μέθοδος upwind δίνει φραγμένες προσεγγίσεις αν $\lambda > 0$ και $\lambda \leq 1$. Οπότε αν η σταθερά $\alpha > 0$ και $\lambda \in (0, 1]$, τότε η μέθοδος είναι von Neumann ευσταθής.

Στο ίδιο συμπέρασμα μπορούμε να καταλήξουμε, αν θεωρήσουμε ότι οι

$$U_i^j = w_j e^{r x_i I}, \quad r \in \mathbb{R},$$

με $I = \sqrt{-1}$ τη φανταστική μονάδα, ικανοποιούν την (7.7). Για να έχουμε ευστάθεια θα εξετάσουμε αν οι w_j παραμένουν φραγμένες. Τότε έχουμε ότι

$$\begin{aligned} w_{j+1} e^{r x_i I} &= (1-\lambda) w_j e^{r x_i I} + \lambda w_j e^{r x_{i-1} I} \\ &= (1-\lambda) w_j e^{r x_i I} + \lambda w_j e^{r x_i I} e^{-rhI}. \end{aligned}$$

Οπότε

$$\begin{aligned} w_{j+1} &= (1-\lambda) w_j + \lambda w_j e^{-rhI} \\ &= w_j (1 - \lambda + \cos(rh) - I \sin(rh)). \end{aligned}$$

Επομένως, εύκολα βλέπουμε ότι $w_j = \kappa^j w_0$, με $\kappa = 1 - \lambda + \cos(rh) - I \sin(rh)$. Στη συνέχεια, θεωρώντας την απόλυτη τιμή $|\kappa|$, έχουμε

$$\begin{aligned} |\kappa|^2 &= (1 - \lambda + \lambda \cos(rh))^2 + \lambda^2 \sin^2(rh) \\ &= 1 - 2\lambda(1-\lambda) + 2(1-\lambda)\cos(rh) \\ &= 1 - 4\lambda(1-\lambda)\sin^2(\frac{rh}{2}). \end{aligned}$$

Για να παραμένουν φραγμένες οι w_j, αρκεί $|\kappa|^2 \leq 1$, το οποίο είναι ισοδύναμο με

$$0 \leq \lambda(1-\lambda)\sin^2(\frac{rh}{2}).$$

Συνεπώς, για $0 \leq \lambda \leq 1$, και για κάθε $r \in \mathbb{R}$, έχουμε ότι w_j παραμένουν φραγμένες και, άρα, η μέθοδος upwind είναι von Neumann ευσταθής.

Στη συνέχεια, θα εξετάσουμε την ευστάθεια von Neumann της μεθόδου downwind. Θεωρούμε λοιπόν ότι οι $U_i^j = w_j e^{r x_i I}$ ικανοποιούν την (7.11), οπότε

$$w_{j+1} e^{r x_i I} = (1+\lambda) w_j e^{r x_i I} - \lambda w_j e^{r x_{i+1} I},$$

Συνεπώς, έχουμε

$$w_{j+1}e^{rx_iI} = (1+\lambda)w_je^{rx_iI} - \lambda w_je^{rx_iI}e^{rhI}.$$

Επομένως, $w_j = \kappa w_0$, με $\kappa = 1+\lambda-\lambda\cos(rh) - I\lambda\sin(rh)$. Οπότε για την απόλυτη τιμή του $|\kappa|$ έχουμε

$$\begin{aligned}|\kappa|^2 &= (1+\lambda-\lambda\cos(rh))^2 + \lambda^2\sin^2(rh)\\ &= 1+2\lambda(1+\lambda) - 2(1+\lambda)\cos(rh)\\ &= 1+4\lambda(1+\lambda)\sin^2(\frac{rh}{2}).\end{aligned}$$

Επομένως, αν $0 \geq \lambda(1+\lambda)$, τότε η μέθοδος downwind δίνει φραγμένες προσεγγίσεις αν $\lambda < 0$ και $\lambda \geq -1$. Οπότε αν $\alpha < 0$ και $\lambda \in [-1, 0)$, τότε η μέθοδος downwind είναι von Neumann ευσταθής.

Στη συνέχεια δείχνουμε τα ακόλουθα θεωρήματα για τη σύγκλιση των μεθόδων upwind και downwind.

Θεώρημα 7.1. *Έστω ότι η λύση u του προβλήματος* (7.1) *είναι αρκετά ομαλή με $\alpha > 0$, και U_i^j, $i = 1, \ldots, N$, $j = 0, \ldots, M$, η λύση της μεθόδου upwind* (7.6)*. Τότε, αν $0 \leq \lambda \leq 1$, υπάρχει μια σταθερά C, ανεξάρτητη των k, h, τέτοια ώστε*

$$\max_{1\leq j\leq M}\max_{0\leq i\leq N+1}|U_i^j - u(x_i,t^j)| \leq C(k+h). \tag{7.15}$$

Απόδειξη. Θέτουμε $E_i^j = U_i^j - u(x_i,t^j)$, $i = 0, \ldots, N+1$, $j = 0, \ldots, M$, όπου λόγω της σχέσης $U_0^j = u(0,t^j) = 0$ έχουμε $E_0^j = 0$. Αφαιρούμε τώρα κατά μέλη τις (7.3) και (7.6), οπότε παίρνουμε

$$E_i^{j+1} = (1-\lambda)E_i^j + \lambda E_{i-1}^j + k\eta_i^j,\ i = 1, \ldots, N,\ j = 0, \ldots, M-1.$$

Θέτουμε, στη συνέχεια, $\bar{E}^j = \max_{1\leq i\leq N}|E_i^j|$ και $\bar{\eta} = \max_{1\leq j\leq M}\max_{0\leq i\leq N+1}|\eta_i^j|$. Οπότε έχουμε

$$\bar{E}^{j+1} \leq (1-\lambda)\bar{E}^j + \lambda\bar{E}^j + k\bar{\eta} \leq \bar{E}^j + k\bar{\eta} \leq \cdots \leq \bar{E}^0 + jk\bar{\eta}.$$

Από εδώ, επειδή $jk \leq T$ και λόγω του Λήμματος 7.1, προκύπτει η ζητούμενη εκτίμηση. □

Θεώρημα 7.2. *Έστω ότι η λύση u του προβλήματος* (7.13) *είναι αρκετά ομαλή με $\alpha < 0$, και U_i^j, $i = 1, \ldots, N$, $j = 0, \ldots, M$, η λύση της μεθόδου downwind* (7.12)*. Τότε, αν $0 \geq \lambda \geq -1$, υπάρχει μια σταθερά C, ανεξάρτητη των k, h, τέτοια ώστε*

$$\max_{1\leq j\leq M}\max_{0\leq i\leq N+1}|U_i^j - u(x_i,t^j)| \leq C(k+h). \tag{7.16}$$

Απόδειξη. Θέτουμε και πάλι $E_i^j = U_i^j - u(x_i, t^j), i = 0, \ldots, N+1, j = 0, \ldots, M$. Οπότε με ανάλογο τρόπο όπως στο Θεώρημα 7.1, παίρνουμε

$$E_i^{j+1} = (1+\lambda)E_i^j - \lambda E_{i-1}^j + k\eta_i^j,\ i = 1, \ldots, N,\ j = 0, \ldots, M-1.$$

Θέτουμε, στη συνέχεια, $\bar{E}^j = \max_{1\le i\le N} |E_i^j|$ και $\bar{\eta} = \max_{1\le j\le M} \max_{0\le i\le N+1} |\eta_i^j|$. Οπότε έχουμε

$$\bar{E}^{j+1} \le (1-|\lambda|)\bar{E}^j + |\lambda|\bar{E}^j + k\bar{\eta} \le \bar{E}^j + k\bar{\eta} \le \cdots \le \bar{E}^0 + jk\bar{\eta}.$$

Από εδώ, επειδή $jk \le T$ και λόγω του γεγονός ότι το σφάλμα η_i^j για τη μέθοδο downwind ικανοποιεί το Λήμμα 7.1, προκύπτει η ζητούμενη εκτίμηση. □

7.2 Μέθοδος των Lax–Wendroff

7.2.1 Μια μη ευσταθής μέθοδος

Στη συνέχεια, θα θεωρήσουμε ένα αριθμητικό σχήμα που να προσεγγίζει με μεγαλύτερη ακρίβεια τη u_x, από τις δύο προηγούμενες μεθόδους upwind και downwind. Έτσι, αν στην (7.2) θεωρήσουμε για την προσέγγιση της $u_x(x_i, t^j)$ την $\delta_h^c u(x_i, t^j)$ και για την $u_t(x_i, t^j)$ την $\delta_k^+ u(x_i, t^j)$, τότε έχουμε για $i = 1, \ldots, N$, και $j = 0, \ldots, M-1$,

$$\frac{u(x_i, t^{j+1}) - u(x_i, t^j)}{k} + \alpha \frac{u(x_{i+1}, t^j) - u(x_{i-1}, t^j)}{2h} = \eta_i^j, \tag{7.17}$$

όπου, αν $u_{tt}, u_{xxx} \in C([a,b]\times[0,T])$, λόγω των (2.3) και (2.4),

$$|\eta_i^j| \le \frac{k}{2} \max_{t\in[0,T]} |u_{tt}(x,t)| + \frac{h^2}{6} \max_{x\in[a,b]} |u_{xxx}(x,t)|. \tag{7.18}$$

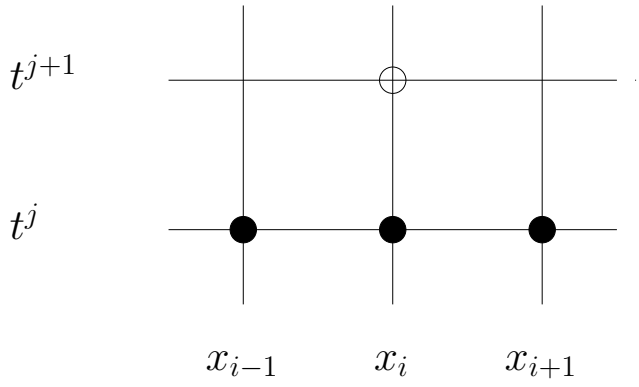
Συνεπώς, ισχύει το ακόλουθο λήμμα.

Λήμμα 7.2. *Έστω u η λύση του* (7.1) *με $u_{tt}, u_{xxx} \in C([a,b]\times[0,T])$. Τότε για την η_i^j που δίνεται στην* (7.17)*, έχουμε ότι υπάρχει σταθερά C ανεξάρτητη των k και h, τέτοια ώστε*

$$\max_{0\le j\le M-1} \max_{1\le i\le N} |\eta_i^j| \le C(k + h^2). \tag{7.19}$$

Θεωρούμε, λοιπόν, την ακόλουθη μέθοδο για να κατασκευάσουμε προσεγγίσεις U_i^j των τιμών $u(x_i, t^j)$,

$$\begin{aligned} \frac{U_i^{j+1} - U_i^j}{k} + \alpha \frac{U_{i+1}^j - U_{i-1}^j}{2h} &= 0, \quad i = 1, \ldots, N,\ j = 0, \ldots, M-1, \\ U_i^0 &= g(x_i), \quad i = 0, \ldots, N+1, \\ U_0^j &= \phi_0(t^j), \quad j = 0, \ldots, M. \end{aligned} \tag{7.20}$$



Σχήμα 7.8: Σημεία ενός πλέγματος για τον υπολογισμό της λύσης με τη μέθοδο (7.20). Με άσπρο σημειώνουμε τα σημεία που αντιστοιχούν σε άγνωστες τιμές της προσέγγισης και με μαύρο αυτά που αντιστοιχούν σε γνωστές τιμές.

Ξεκινώντας από το χρονικό επίπεδο $t = 0$ όπου γνωρίζουμε ότι $U_i^0 = g(x_i)$, $i = 0, \ldots, N+1$, και χρησιμοποιώντας την (7.20), μπορούμε να βρούμε με άμεσο τρόπο την προσέγγιση U_i^1, $i = 0, \ldots, N$, στο επόμενο χρονικό επίπεδο $t^1 = k$, όμως παρατηρούμε ότι δεν μπορούμε να υπολογίσουμε την U_{N+1}^1. Επομένως, όπως και στη μέθοδο (7.11), για να βρούμε όλες τις τιμές U_i^j στα επόμενα χρονικά επίπεδα t^j, $j = 1, \ldots, M$, χρειάζεται να γνωρίζουμε την $u(b,t)$, για $t \in [0,T]$. Έτσι, υποθέτουμε ότι στο πρόβλημα (7.1), η ακριβής λύση u ικανοποιεί επιπλέον και τη συνοριακή συνθήκη $u(b,t) = \phi_1(t)$, $t \in [0,T]$, οπότε θέτουμε $U_{N+1}^j = \phi_1(t^j)$, $j = 0, \ldots, M$.

Συνεπώς, αν συμβολίσουμε τώρα $\lambda = \dfrac{\alpha k}{h}$, τότε η (7.20) μπορεί να γραφεί

$$\begin{aligned} U_i^{j+1} &= -\frac{\lambda}{2}U_{i+1}^j + U_i^j + \frac{\lambda}{2}U_{i-1}^j, && i = 1, \ldots, N,\ j = 0, \ldots, M-1, \\ U_i^0 &= g(x_i), && i = 0, \ldots, N+1, \\ U_0^j &= \phi_0(t^j) \text{ και } U_{N+1}^j = \phi_1(t^j), && j = 0, \ldots, M. \end{aligned} \tag{7.21}$$

Οπότε τώρα χρησιμοποιώντας την (7.21), βρίσκουμε με άμεσο τρόπο όλες τις τιμές U_i^j στα επόμενα χρονικά επίπεδα t^j, $j = 1, \ldots, M$.

Εύκολα παρατηρούμε ότι χρησιμοποιώντας τη μέθοδο (7.21), η τιμή της προσέγγισης στο σημείο (x_i, t^j), εξαρτάται τελικά από τις τιμές στα σημεία $(x_{i-j}, 0)$, $\ldots$, $(x_i, 0)$, $\ldots$, $(x_{i+j}, 0)$, βλ. Σχήμα 7.9. Για να ικανοποείται τώρα η συνθήκη CFL αρκεί το διάστημα $[x_{i-j}, x_{i+j}]$, το οποίο περιέχει τα σημεία του χωρίου υπολογιστικής εξάρτησης της μεθόδου, να περιέχει το $\bar{x}_0 = x_i - \alpha t^j$, το οποίο ισχύει

ανεξάρτητα από το πρόσημο της σταθεράς α, αν $|\lambda| = |\frac{\alpha k}{h}| \leq 1$. Επίσης, ακολουθώντας τα βήματα για τη μέλετη της *ευστάθιας von Neumann* για τις μεθόδους upwind και downwind, έχουμε ότι, αν και πάλι θέσουμε $U_i^j = w_j e^{rx_i I}$, με $r \in \mathbb{R}$,

$$w_{j+1}e^{rx_iI} = -\frac{\lambda}{2}w_j e^{rx_{i+1}I} + w_j e^{rx_iI} + \frac{\lambda}{2}w_j e^{rx_{i-1}I},$$

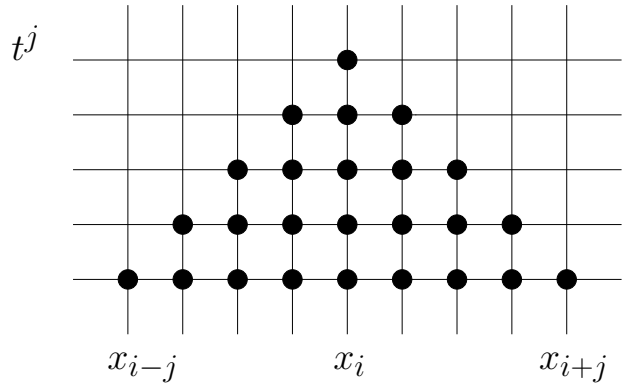
Οπότε, εφαρμόζοντας την τριγωνομετρική ταυτότητα (5.18), παίρνουμε

$$\begin{aligned} w_{j+1} &= w_j(-\frac{\lambda}{2}w_j e^{rhI} + 1 + \frac{\lambda}{2}w_j e^{-rhI}) \\ &= w_j(1 - I\lambda \sin(rh)). \end{aligned}$$

Συνεπώς, $w_j = \kappa^j w_0$, με $\kappa = 1 - I\lambda \sin(rh)$, και

$$|\kappa|^2 = 1 + \lambda^2 \sin^2(rh).$$

Επομένως, η μέθοδος (7.21) δεν είναι von Neumann ευσταθής, γιατί $|\kappa| > 1$.



Σχήμα 7.9: Χωρίο εξάρτησης της προσεγγιστικής λύσης με τις μεθόδους (7.21) και Lax–Wendroff.

7.2.2 Μέθοδος Lax–Wendroff

Στη συνέχεια, θα κατασκευάσουμε μια νέα μέθοδο για την προσέγγιση της (7.1) η οποία θα χρησιμοποιεί τα ίδια σημεία του πλέγματος όπως η (7.21), θα είναι von Neumann ευσταθής και θα έχει και αυτή τοπικό σφάλμα η_i^j το οποίο θα τείνει στο μηδέν, καθώς k και h τείνουν στο μηδέν. Θέτουμε, λοιπόν, για $i = 1, \ldots, N$, $j = 0, \ldots, M-1$,

$$u(x_i, t^{j+1}) = Au(x_{i+1}, t^j) + Bu(x_i, t^j) + \Gamma u(x_{i-1}, t^j) + k\eta_i^j, \qquad (7.22)$$

και θέλουμε να προσδιορίσουμε τα A, B και Γ, έτσι ώστε το τοπικό σφάλμα η_i^j να είναι όσο το δυνατόν μικρότερο. Χρησιμοποιώντας το γεγονός ότι $u_t = -\alpha u_x$, $u_{tt} = -\alpha u_{xt} = -\alpha(u_t)_x = \alpha^2 u_{xx}$ και $u_{ttt} = -\alpha^3 u_{xxx}$ και αναπτύσσοντας σύμφωνα με το Θεώρημα Taylor, έχουμε

$$\begin{aligned} u(x_i, t^j + k) &= u(x_i, t^j) + k u_t(x_i, t^j) + \frac{k^2}{2} u_{tt}(x_i, t^j) + \frac{k^3}{6} u_{ttt}(x_i, t^j) \\ &\quad + \hat{\eta}_i^j \\ &= u(x_i, t^j) - \alpha k u_x(x_i, t^j) + \frac{\alpha^2 k^2}{2} u_{xx}(x_i, t^j) \\ &\quad - \frac{\alpha^3 k^3}{6} u_{xxx}(x_i, t^j) + \hat{\eta}_i^j, \end{aligned} \tag{7.23}$$

με

$$|\hat{\eta}_i^j| \leq \frac{k^4}{24} \max_{t \in [0,T]} |\frac{\partial}{\partial t^4} u(x_i, t)|, \tag{7.24}$$

και όμοια

$$\begin{aligned} u(x_i \pm h, t^j) = u(x_i, t^j) &\pm h u_x(x_i, t^j) + \frac{h^2}{2} u_{xx}(x_i, t^j) \\ &\pm \frac{h^3}{6} u_{xxx}(x_i, t^j) + \tilde{\eta}_{i\pm}^j, \end{aligned} \tag{7.25}$$

με

$$|\tilde{\eta}_{i\pm}^j| \leq \frac{h^4}{24} \max_{x \in [a,b]} |\frac{\partial}{\partial x^4} u(x, t^j)|. \tag{7.26}$$

Αν αντικαταστήσουμε τις (7.23) και (7.25) στην (7.22), έχουμε

$$\begin{aligned} k\eta_i^j &= u(x_i, t^{j+1}) - A u(x_{i+1}, t^j) - B u(x_i, t^j) - \Gamma u(x_{i-1}, t^j) \\ &= u(x_i, t^j) - \alpha k u_x(x_i, t^j) + \frac{\alpha^2 k^2}{2} u_{xx}(x_i, t^j) + \frac{\alpha^3 k^3}{6} u_{xxx}(x_i, t^j) + \hat{\eta}_i^j \\ &\quad - A(u(x_i, t^j) + h u_x(x_i, t^j) + \frac{h^2}{2} u_{xx}(x_i, t^j) + \frac{h^3}{6} u_{xxx}(x_i, t^j) + \tilde{\eta}_{i^+}^j) \\ &\quad - B u(x_i, t^j) \\ &\quad - \Gamma(u(x_i, t^j) - h u_x(x_i, t^j) + \frac{h^2}{2} u_{xx}(x_i, t^j) - \frac{h^3}{6} u_{xxx}(x_i, t^j) + \tilde{\eta}_{i^-}^j). \end{aligned}$$

Επομένως

$$\begin{aligned}
k\eta_i^j &= (1-A-B-\Gamma)u(x_i,t^j) - (\alpha k + Ah - \Gamma h)u_x(x_i,t^j) \\
&\quad + (\frac{\alpha^2k^2}{2} - (A+\Gamma)\frac{h^2}{2})u_{xx}(x_i,t^j) \\
&\quad + (\frac{\alpha^3k^3}{6} - (A-\Gamma)\frac{h^3}{6})u_{xxx}(x_i,t^j) \\
&\quad + \hat{\eta}_i^j - A\tilde{\eta}_{i^+}^j - \Gamma\tilde{\eta}_{i^-}^j \\
&= (1-A-B-\Gamma)u(x_i,t^j) - h(\lambda + A - \Gamma)u_x(x_i,t^j) \\
&\quad + \frac{h^2}{2}(\lambda^2 - (A+\Gamma))u_{xx}(x_i,t^j) \\
&\quad + \frac{h^3}{6}(\lambda^3 - A + \Gamma)u_{xxx}(x_i,t^j) \\
&\quad + \hat{\eta}_i^j - A\tilde{\eta}_{i^+}^j - \Gamma\tilde{\eta}_{i^-}^j.
\end{aligned}$$

Στη συνέχεια, εξισώνοντας τους συντελεστές των u, u_x, u_{xx} στην παραπάνω εξίσωση με το μηδέν, προκύπτουν οι ακόλουθες σχέσεις

$$\begin{aligned}
1 - A - B - \Gamma &= 0 \\
\lambda + A - \Gamma &= 0 \\
\lambda^2 - (A+\Gamma) &= 0.
\end{aligned}$$

Επομένως, έχουμε $A = -\frac{1}{2}\lambda(1-\lambda)$, $B = 1-\lambda^2$ και $\Gamma = \frac{1}{2}\lambda(1+\lambda)$. Συνεπώς, για αυτήν την επιλογή των A, B και Γ έχουμε

$$\begin{aligned}
k|\eta_i^j| &\le \frac{h^3}{6}(|\lambda^3| + |\lambda|)\,|u_{xxx}(x_i,t^j)| + |\hat{\eta}_i^j| + \lambda^2 \max(|\tilde{\eta}_{i^+}^j|, |\tilde{\eta}_{i^-}^j|) \\
&\le \frac{\alpha k}{6}(\alpha^2k^2 + h^2)|u_{xxx}(x_i,t^j)| + |\hat{\eta}_i^j| + \frac{\alpha^2k^2}{h^2}\max(|\tilde{\eta}_{i^+}^j|, |\tilde{\eta}_{i^-}^j|).
\end{aligned} \tag{7.27}$$

Οπότε, λόγω των (7.24) και (7.26), έχουμε ότι υπάρχει σταθερά C ανεξάρτητη των k και h, τέτοια ώστε

$$\max_{1\le i\le N, 0\le j\le M-1} |\eta_i^j| \le C(k^2 + h^2).$$

Η μέθοδος που προκύπτει για την παραπάνω επιλογή των συντελεστών A, B και Γ ονομάζεται μέθοδος των Lax–Wendroff και για να μπορέσουμε να βρούμε την προσέγγιση σε κάθε χρονικό επίπεδο t^j, $j = 1, \ldots, M$, χρειάζεται, όπως και στην

(7.21) να θεωρήσουμε γνωστή την ακριβή λύση για $x = a$ και $x = b$. Έτσι, η μέθοδος Lax–Wendroff ορίζεται ως εξης

$$\begin{aligned} U_i^{j+1} &= -\frac{1}{2}\lambda(1-\lambda)U_{i+1}^j + (1-\lambda^2)U_i^j + \frac{1}{2}\lambda(1+\lambda)U_{i-1}^j, \\ &\qquad i = 1,\ldots,N,\ j = 0,\ldots,M-1, \\ U_i^0 &= g(x_i), \quad i = 0,\ldots,N+1, \\ U_0^j &= \phi_0(t^j),\ \text{και}\ U_{N+1}^j = \phi_1(t^j), \quad j = 0,\ldots,M. \end{aligned} \tag{7.28}$$

Εύκολα μπορούμε να δούμε ότι το χωρίο υπολογιστικής εξάρτησης για το (x_i, t^j) είναι το ίδιο με τη μέθοδο (7.21) και εξαρτάται από τα σημεία $\{x_i - jh, \ldots, x_i, \ldots, x_i + jh\}$, βλέπε Σχήμα 7.9. Επίσης για να ικανοποείται η συνθήκη CFL πρέπει το σημείο $\bar{x}_0 = x_i - \alpha t^j$ να περιέχεται στο διάστημα $[x_{i-j}, x_{i+j}]$, το οποίο ισχύει ανεξάρτητα από το πρόσημο της σταθεράς α, αν $|\lambda| = |\frac{\alpha k}{h}| \leq 1$.

Παράδειγμα 7.4. Θεωρούμε και πάλι το πρόβλημα των Παραδειγμάτων 7.1 και 7.2, στο διάστημα $[-1, 9]$, με g όπως στο Παράδειγμα 1.2, $\phi_0(t) = \phi_1(t) = g(x - \alpha t)$ και $\alpha = 1$. Επειδή γνωρίζουμε ότι $u(a,t) = u(b,t) = 0$, για $t \in [0,7]$, μπορούμε να θέσουμε $U_0^j = U_{N+1}^j = 0$, $j = 0,\ldots,M$. Επίσης, διαμερίζουμε το $[-1,9]$ σε $N+2$ σημεία, με $N = 99$ και το $[0,7]$ σε $M+1$, με $M = 63, 77, 98$. Στα Σχήματα 7.10 και 7.11 βλέπουμε την ακριβή λύση u και τις προσεγγίσεις με τη μέθοδο των Lax–Wendroff (7.28) για $t = 0, 1, 5$. Παρατηρούμε ότι η προσεγγιστική λύση έχει παρόμοια συμπεριφορά όπως και αυτή της μεθόδου upwind στο Παράδειγμα 7.1. Έτσι αν $\lambda = 70/63 > 1$, για $M = 63$, τότε το σφάλμα της προσεγγιστικής λύσης είναι μεγάλο, ενώ το μικρότερο σφάλμα προκύπτει αν $\lambda < 1$ και είναι πιο κοντά στη μονάδα.

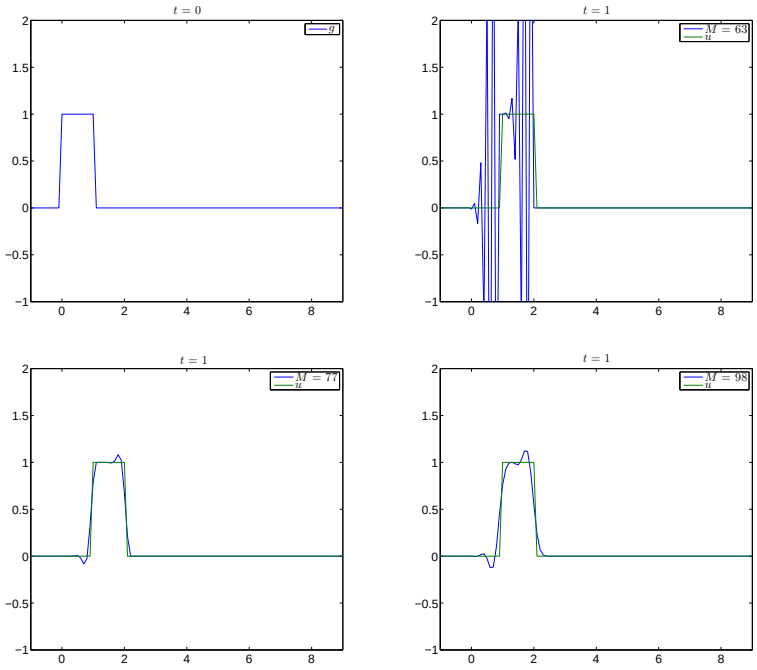
7.2.3 Ευστάθεια και σύγκλιση

Για τη μελέτη της *ευστάθειας von Neumann* για τη μέθοδο των Lax–Wendroff, θα θεωρήσουμε και πάλι ότι η προσεγγιστική λύση είναι της μορφής $U_i^j = w_j e^{rx_i I}$, $r \in \mathbb{R}$, οπότε σύμφωνα με την (7.28), έχουμε

$$\begin{aligned} w_{j+1}e^{rx_iI} = &-\frac{1}{2}\lambda(1-\lambda)w_j e^{rx_{i+1}I} + (1-\lambda^2)w_j e^{rx_iI} \\ &+ \frac{1}{2}\lambda(1+\lambda)w_j e^{rx_{i-1}I}. \end{aligned}$$

Χρησιμοποιώντας τώρα τις (5.23) και (5.18), παίρνουμε

$$w_{j+1} = (-\frac{1}{2}\lambda(1-\lambda) + \frac{1}{2}\lambda(1+\lambda))w_j \cos(h)$$



Σχήμα 7.10: Η ακριβής λύση $u(x,t)$ του Παραδείγματος 7.4 και οι προσεγγίσεις της για $N = 99$ και $M = 63, 77, 98$, με τη μέθοδο Lax-Wendroff για $t = 0$ και 1.

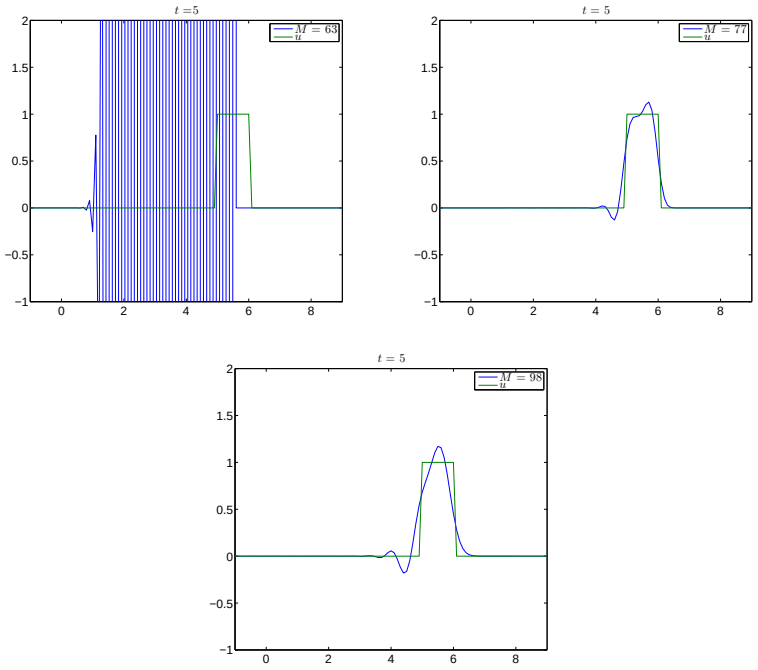
$$
\begin{aligned}
&+ (1-\lambda^2)w_j - I(\frac{1}{2}\lambda(1-\lambda) + \frac{1}{2}\lambda(1+\lambda))w_j \sin(h) \\
&= w_j(1 - \lambda^2 + \lambda^2\cos(h) - I\lambda\sin(rh)).
\end{aligned}
$$

Συνεπώς, $w_j = \kappa w_0$, με $\kappa = 1 - \lambda^2 + \lambda^2\cos(h) - I\lambda\sin(rh)$. Οπότε

$$
\begin{aligned}
|\kappa|^2 &= (1-\lambda^2+\lambda^2\cos(h))^2 + \lambda^2\sin^2(h) \\
&= (1-\lambda^2)^2 + \lambda^4\cos^2(h) + 2\lambda^2(1-\lambda^2)\cos(h) + \lambda^2\sin^2(h) \\
&= (1-\lambda^2)^2 + \lambda^4(1-\sin^2(h)) + 2\lambda^2(1-\lambda^2)\cos(h) + \lambda^2\sin^2(h) \\
&= 1 - 2\lambda^2(1-\lambda^2) + \lambda^2(1-\lambda^2)\sin^2(h) + 2\lambda^2(1-\lambda^2)\cos(h) \\
&= 1 - 2\lambda^2(1-\lambda^2)[1-\cos(h)] + \lambda^2(1-\lambda^2)\sin^2(h) \\
&= 1 - 4\lambda^2(1-\lambda^2)\sin^2(h/2) + \lambda^2(1-\lambda^2)\sin^2(h).
\end{aligned}
$$

Χρησιμοποιώντας τώρα την τριγωνομετρική ιδιότητα $\sin^2(h) = 4\sin^2(h/2) - 4\sin^4(h/4)$, η παραπάνω σχέση γίνεται

$$|\kappa|^2 = 1 - 4\lambda^2(1-\lambda^2)\sin^4(h/4).$$



Σχήμα 7.11: Η ακριβής λύση $u(x,t)$ του Παραδείγματος 7.4 και οι προσεγγίσεις της για $N = 99$ και $M = 63, 77, 98$, με τη μέθοδο Lax-Wendroff για $t = 5$.

Είναι προφανές ότι, αν $\lambda^2 \leq 1$, τότε $|\kappa| \leq 1$, και, άρα, η μέθοδος Lax–Wendroff είναι von Neumann ευσταθής, αν $\lambda \in [-1, 1]$.

Για να δείξουμε τη σύγκλιση της μεθόδου Lax–Wendroff χρειάζεται να εφαρμόσουμε παρόμοια επιχειρήματα όπως αυτά της μεθόδου ενέργειας στην Παράγραφο 3.4, βλ. παραδείγματος χάριν (Larsson & Thomée, 2009, Παράγραφος 12.3), η απόδειξη του οποίου είναι πολύπλοκη και ξεφεύγει από τους σκοπούς αυτών των σημειώσεων. Διατυπώνουμε, επομένως, το επόμενο θεώρημα χωρίς απόδειξη.

Θεώρημα 7.3. *Έστω ότι η λύση u του προβλήματος* (7.1) *είναι αρκετά ομαλή, και U_i^j, $i = 1, \ldots, N$, $j = 0, \ldots, M$, η λύση της μεθόδου Lax–Wendroff* (7.28)*. Τότε, αν $0 \leq \lambda \leq 1$, υπάρχει μια σταθερά C, ανεξάρτητη των k, h, τέτοια ώστε*

$$\max_{1\leq j\leq M} \max_{0\leq i\leq N+1} |U_i^j - u(x_i, t^j)| \leq C(k^2 + h^2). \tag{7.29}$$

Στο ακόλουθο παράδειγμα βλέπουμε και πειραματικά ότι το σφάλμα με τη μέθοδο Lax–Wendroff έχει μεγαλύτερη τάξη προσέγγισης ως προς k και h από το σφάλμα της μεθόδου upwind ή downwind.

	Upwind		Lax–Wendroff	
k	$\bar{E}_{upwind}$	p	$\bar{E}_{lax}$	p
0.05	0.0483		0.01965	
0.025	0.0325	0.574	0.00960	1.034
0.0125	0.0208	0.642	0.00413	1.216
0.0063	0.0127	0.713	0.00151	1.456
0.0031	0.0073	0.785	0.00045	1.743
0.0016	0.0040	0.852	0.00012	1.848
0.0008	0.0021	0.906	0.00003	1.992

Πίνακας 7.1: Τα σφάλματα των μεθόδων upwind και Lax–Wendroff στο χρονικό επίπεδο $t = T = 7$, $\bar{E}_{upwind}$ και $\bar{E}_{lax}$, αντίστοιχα, του Παραδείγματος 7.5, καθώς και η κατά προσέγγιση τάξη ακρίβειάς τους p, αν θεωρήσουμε διαμερίσεις όπου $k/h = 0.9$.

Παράδειγμα 7.5. Βασική υπόθεση στα θεωρήματα σύγκλισης για τις μεθόδους που μελετήσαμε σε αυτό το κεφάλαιο είναι ότι η ακριβής λύση u, και κατά επέκταση η συνάρτηση της αρχικής συνθήκης g, είναι αρκετά ομάλη. Είναι προφανές ότι η g που θεωρήσαμε στα προηγούμενα παραδείγματα δεν είναι ομαλή συνάρτηση, για αυτό τον λόγο τώρα θεωρούμε το πρόβλημα (7.13) στο διάστημα $[-11, 9] \times [0, 7]$, με

$$g(x) = \begin{cases} e^{-\frac{1}{1-x^2}}, & \text{αν } -1 \leq x \leq 1, \\ 0 & \text{διαφορετικά.} \end{cases}$$

Επειδή η σταθερά α του (7.13) είναι θετική για να ικανοποιείται η συνθήκη CFL θα χρησιμοποιήσουμε τις μεθόδους upwind και Lax–Wendroff. Θεωρούμε τώρα διαμερίσεις του $[0, 7]$ με $k = 0.05, 0.025, 0.0125, 0.0063, 0.0031, 0.0016, 0.0008$ και του $[-11, 9]$, με αντίστοιχο βήμα h, τέτοιο ώστε $k/h = 0.9$. Θέτουμε $\bar{E}_{upwind} = \max_{0 \leq N+1} |U_i^M - u(x_i, t^M)|$, με $t^M = T = 7$ το σφάλμα της προσεγγιστικής λύσης που προκύπτει με τη μεθόδο upwind και με ανάλογο τρόπο ορίζουμε το σφάλμα $\bar{E}_{lax}$ με τη μέθοδο των Lax–Wendroff. Στον πίνακα 7.1 βλέπουμε τα σφάλματα αυτών των δύο μεθόδων και την προσεγγιστική τάξη ακρίβειας p.

7.3 Άλλες μέθοδοι και προβλήματα

Στις παραγράφους 7.1 και 7.2 είδαμε τρεις άμεσες μεθόδους για την αριθμητική προσέγγιση της λύσης της εξίσωσης της μεταφοράς. Στη βιβλιογραφία, υπάρχουν και άλλες άμεσες ή πεπλεγμένες μέθοδοι, τις οποίες δεν θα μελετήσουμε εδώ. Για

την ανάλυση αυτών των αριθμητικών μεθόδων, καθώς και για μια πιο λεπτομερή παρουσίαση των αποτελεσμάτων αυτού του κεφαλαίου παραπέμπουμε τον αναγνώστη στα (Holmes, 2007· Iserles, 2009· Morton & Mayers, 2005· Richtmyer & Morton, 1967· Strikwerda, 2004· Thomas, 1995).

7.4 Ταινίες γραφικών παραστάσεων

Σε αυτή τη παράγραφο εμφανίζονται ταινίες με τις γραφικές παραστάσεις που παρουσιάστηκαν στα Παραδείγματα 7.1–7.4 για τις μεθόδους upwind, downwind και Lax–Wendroff. Η προβολή των ταινιών στην οθόνη του Η/Υ μπορεί να γίνει αν μετακινήσουμε τον κέρσορα (δείκτη) της οθόνης και επιλέξουμε, π.χ. με τη χρήση του "ποντίκιου", το αντίστοιχο παράδειγμα και, στη συνέχεια, την επιλογή "Play/Pause". Στην περίπτωση που η προβολή αυτού του βιβλίου στην οθόνη γίνεται μέσω του αντίστοιχου αρχείου μορφής pdf, συνίσταται η χρήση του προγράμματος Adobe Reader.

Play/Pause | Upwind με $\lambda = 1.1$ | Upwind με $\lambda = 0.91$ | Upwind με $\lambda = 0.71$

Σχήμα 7.12: Ταινίες των γραφικών παραστάσεων του Παραδείγματος 7.1 με $\alpha > 0$ για τη μέθοδο upwind.

Play/Pause Downwind με $\lambda = 1.1$ Downwind με $\lambda = 0.91$

Downwind με $\lambda = 0.71$

Σχήμα 7.13: Ταινίες των γραφικών παραστάσεων του Παραδείγματος 7.2 με $\alpha > 0$ για τη μέθοδο downwind.

Play/Pause Downwind με $\lambda = 1.1$ Downwind με $\lambda = 0.91$

Downwind με $\lambda = 0.71$

Σχήμα 7.14: Ταινίες των γραφικών παραστάσεων του Παραδείγματος 7.3 με $\alpha < 0$ για τη μέθοδο downwind.

Play/Pause Lax–Wendroff με $\lambda = 1.1$ Lax–Wendroff με $\lambda = 0.91$

Lax–Wendroff με $\lambda = 0.71$

Σχήμα 7.15: Ταινίες των γραφικών παραστάσεων του Παραδείγματος 7.4 με $\alpha > 0$ για τη μέθοδο των Lax–Wendroff.

7.5 Ασκήσεις

7.1 Δείξτε ότι οι μέθοδοι upwind και Lax–Wendroff δίνουν την ακριβή λύση για $\lambda = 1$.

7.2 Θεωρούμε τη μέθοδο

$$(1+\lambda)U_i^{j+1} + (1-\lambda)U_{i-1}^{j+1} = (1-\lambda)U_i^j + (1+\lambda)U_{i-1}^j,$$

για την επίλυση του προβλήματος (7.1), όπου με λ συμβολίζουμε τον λόγο $\frac{\alpha k}{h}$.

(α΄) Βρείτε το σφάλμα διακριτοποίησης αυτής της μεθόδου.

(β΄) Ποια είναι η συνθήκη CFL για αυτή τη μέθοδο;

(γ΄) Αν για κάθε j η τιμή U_0^{j+1} είναι γνωστή, δείξτε ότι το σχήμα είναι άμεσο.

(δ΄) Είναι von Neumann ευσταθής;

7.3 Θεωρούμε τη μέθοδο

$$U_i^{j+1} = U_i^{j-1} - \lambda(U_{i+1}^j - U_{i-1}^j),$$

για την επίλυση του προβλήματος (7.1), όπου με λ συμβολίζουμε τον λόγο $\frac{\alpha k}{h}$.

(α΄) Βρείτε το σφάλμα διακριτοποίησης αυτής της μεθόδου.

(β΄) Ποια είναι η συνθήκη CFL για αυτή τη μέθοδο;

(γ΄) Αν για κάθε j η τιμή U_0^{j+1} είναι γνωστή, δείξτε ότι το σχήμα είναι άμεσο.

(δ΄) Δείξτε ότι είναι von Neumman ευσταθής για $\lambda < 1$ και ασταθής για $\lambda > 1$;

7.4 Θεωρούμε τη μέθοδο

$$U_i^{j+1} = \frac{1}{4}\lambda(\lambda-1)U_{i-2}^j + \frac{1}{4}\lambda(5-\lambda)U_{i-1}^j - \frac{1}{4}(\lambda-1)(\lambda+4)U_i^j + \frac{1}{4}\lambda(\lambda-1)U_{i+1}^j.$$

(α΄) Βρείτε το σφάλμα διακριτοποίησης της μεθόδου.

(β΄) Ποια είναι η συνθήκη CFL για αυτή τη μέθοδο;

(γ΄) Είναι η μέθοδος von Neumman ευσταθής;

7.5 Θεωρούμε ένα αριθμητικό σχήμα που χρησιμοποιεί τις τιμές U_i^{j+1}, U_{i-1}^j, U_i^j και U_{i-1}^{j-1}, για την προσέγγιση της λύσης της εξίσωσης (1.35).

(α΄) Βρείτε το σχήμα, ώστε η μέθοδος να είναι συνεπής και το τοπικό σφάλμα διακριτοποίησης να είναι δεύτερης τάξης ως προς h και πρώτης ως προς k.

(β΄) Είναι η μέθοδος άμεση ή πεπλεγμένη;

(γ΄) Ποια είναι η συνθήκη CFL για αυτή τη μέθοδο;

(δ΄) Είναι von Neumann ευσταθής;

7.6 Θεωρούμε ένα αριθμητικό σχήμα που χρησιμοποιεί τις τιμές $U_{i+1}^{j+1}, U_i^{j+1}, U_{i-1}^{j+1}$, U_{i+1}^j, U_i^j και U_{i-1}^j, για την προσέγγιση της λύσης της εξίσωσης (1.35).

(α΄) Βρείτε το σχήμα, ώστε η μέθοδος να είναι συνεπής και το σφάλμα διακριτοποίησης να είναι δεύτερης τάξης ως προς h και k.

(β΄) Είναι η μέθοδος άμεση ή πεπλεγμένη;

(γ΄) Ποια είναι η συνθήκη CFL για αυτή τη μέθοδο;

(δ΄) Είναι von Neumann ευσταθής;

Βιβλιογραφία

Holmes, M. H. (2007). *Introduction to numerical methods in differential equations* (Vol. 52). Springer, New York.

Iserles, A. (2009). *A first course in the numerical analysis of differential equations* (Second ed.). Cambridge University Press, Cambridge.

Larsson, S., & Thomée, V. (2009). *Partial differential equations with numerical methods* (Vol. 45). Springer-Verlag, Berlin.

Morton, K. W., & Mayers, D. F. (2005). *Numerical solution of partial differential equations* (Second ed.). Cambridge University Press, Cambridge.

Richtmyer, R. D., & Morton, K. W. (1967). *Difference methods for initial-value problems*. Interscience Publishers John Wiley & Sons, Inc., New York-London-Sydney.

Strikwerda, J. C. (2004). *Finite difference schemes and partial differential equations* (Second ed.). Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA.

Thomas, J. W. (1995). *Numerical partial differential equations: finite difference methods* (Vol. 22). Springer-Verlag, New York.

Κεφάλαιο 8

Εξίσωση κύματος

Σε αυτό το κεφάλαιο θεωρούμε μια υπερβολική εξίσωση δεύτερης τάξεως, την εξίσωση του κύματος, στη μια διάσταση ως προς τον χώρο. Θα κατασκευάσουμε μεθόδους πεπερασμένων διαφορών για το πρόβλημα αρχικών και συνοριακών τιμών με ομογενείς συνθήκες Dirichlet. Στη συνέχεια, θα δείξουμε ιδιότητες ευστάθειας καθώς και κατάλληλες συνθήκες που πρέπει να ικανοποιεί η μέθοδος για να είναι ευσταθής. Τέλος, θα δείξουμε τη σύγκλιση της προσεγγιστικής λύσης στην ακριβή λύση.

8.1 Μέθοδος των Courant, Friedrichs και Lewy

Θεωρούμε το ακόλουθο πρόβλημα: Ζητείται μια συνάρτηση $u : [0, L] \times [0, T] \to \mathbb{R}$, τέτοια ώστε

$$\begin{aligned} u_{tt}(x,t) &= \alpha^2 u_{xx}(x,t), \qquad x \in [0,L],\ t \in [0,T], \\ u(0,t) &= u(L,t) = 0, \quad t \in [0,T], \\ u(x,0) &= g_0(x), \qquad x \in [0,L], \\ u_t(x,0) &= g_1(x), \qquad x \in [0,L], \end{aligned} \tag{8.1}$$

όπου $\alpha, L > 0$, $g_0, g_1 \in C[0, L]$. Είναι γνωστό, βλ. π.χ. (Ακρίβης & Αλικάκος, 2012), ότι η λύση του (8.1) δίνεται από τη σειρά

$$u(x,t) = \sum_{n=1}^{\infty} [A_n \cos(\alpha\lambda_n t) + B_n \sin(\alpha\lambda_n t)] \sin(\lambda_n x),\ \text{με } \lambda_n = \frac{n\pi}{L},$$

και $A_n = \dfrac{2}{L}\displaystyle\int_0^L g_0(x)\sin(\lambda_n x)$, $B_n = \dfrac{2}{\alpha n\pi}\displaystyle\int_0^L g_1(x)\sin(\lambda_n x)$.

Θα θεωρήσουμε έναν φυσικό αριθμό N και τη διαμέριση του διαστήματος $[0, L]$ από $N+2$ ισαπέχοντα σημεία $0 = x_0 < x_1 < \cdots < x_N < x_{N+1} = L$, όπου $h = x_{i+1} - x_i$, $i = 0, \ldots, N$. Επίσης, θεωρούμε και έναν φυσικό αριθμό M και τη διαμέριση του $[0, T]$ από $M+1$ ισαπέχοντα σημεία $0 = t^0 < t^1 < \cdots < t^M = T$, όπου $k = t^{j+1} - t^j$, $j = 0, \ldots, M-1$, βλ. Σχήμα 5.1. Τότε, σε κάθε σημείο (x_i, t^j) του διαμερισμού του $[0, L] \times [0, T]$, θα ισχύει:

$$u_{tt}(x_i, t^j) = \alpha^2 u_{xx}(x_i, t^j), \quad i = 1, \ldots, N,\ j = 0, \ldots, M. \tag{8.2}$$

Θα προσεγγίσουμε καταρχήν τις u_{xx} και u_{tt} στα σημεία του διαμερισμού και θα χρησιμοποιούμε τις προσέγγισεις $\delta^c_{h,2} u(x_i, t^j)$ και $\delta^c_{,2} u(x_i, t^j)$, αντίστοιχα, που θεωρήσαμε στην (2.8). Έτσι, αν υποθέσουμε ότι $\frac{\partial^4 u}{\partial x^4}, \frac{\partial^4 u}{\partial t^4} \in C([0, L] \times [0, T])$, λόγω της (2.11), η (8.2) γίνεται

$$\begin{aligned} &\frac{u(x_i, t^{j+1}) - 2u(x_i, t^j) + u(x_i, t^{j-1})}{k^2} \\ &= \alpha^2 \frac{u(x_{i+1}, t^j) - 2u(x_i, t^j) + u(x_{i-1}, t^j)}{h^2} + \eta_i^j, \end{aligned} \tag{8.3}$$

για $i = 1, \ldots, N$, $j = 1, \ldots, M-1$, όπου

$$|\eta_i^j| \leq \frac{k^2}{12} \max_{t \in [0,T]} |\frac{\partial^4 u}{\partial t^4}(x_i, t)| + \frac{\alpha^2 h^2}{12} \max_{x \in [0,L]} |\frac{\partial^4 u}{\partial x^4}(x, t^j)|. \tag{8.4}$$

Θα κατασκευάσουμε, λοιπόν, προσεγγίσεις U_i^j της λύσης u του προβλήματος (8.1) στα σημεία (x_i, t^j), $i = 1, \ldots, N$, $j = 1, \ldots, M$, θεωρώντας τις ακόλουθες εξισώσεις,

$$\frac{U_i^{j+1} - 2U_i^j + U_i^{j-1}}{k^2} = \alpha^2 \frac{U_{i+1}^j - 2U_i^j + U_{i-1}^j}{h^2}, \tag{8.5}$$
$$\text{για } i = 1, \ldots, N, j = 1, \ldots, M-1,$$

$$\begin{cases} U_0^j = U_{N+1}^j = 0, \quad j = 0, \ldots, M, \\ U_i^0 = g_0(x_i), \quad i = 1, \ldots, N. \end{cases} \tag{8.6}$$

Συμβολίζουμε τώρα με λ τον λόγο $\dfrac{\alpha k}{h}$, οπότε η (8.5) γράφεται ως

$$U_i^{j+1} = \lambda^2 U_{i+1}^j + 2(1 - \lambda^2) U_i^j + \lambda^2 U_{i-1}^j - U_i^{j-1}, \tag{8.7}$$

για $i = 1, \ldots, N$, $j = 1, \ldots, M-1$. Παρατηρούμε ότι χρησιμοποιώντας την (8.7) για να βρούμε την προσέγγιση στο χρονικό επίπεδο t^{j+1}, χρειάζεται να γνωρίζουμε

τις προσεγγίσεις στα χρονικά επίπεδα t^j και t^{j-1}. Έτσι ενώ, λόγω των (8.6), γνωρίζουμε τις $U_i^0, i = 0, \ldots, N+1$, για να βρούμε την προσέγγιση στο χρονικό επίπεδο t^2, χρειάζεται να γνωρίζουμε την προσέγγιση και στο χρονικό επίπεδο t^1. Η οποία όμως δεν είναι γνωστή και πρέπει να την προσδιορίσουμε με μια άλλη μέθοδο.

Αν υποθέσουμε ότι $\frac{\partial^3 u}{\partial t^3} \in C([0, L] \times [0, T])$, χρησιμοποιώντας το ανάπτυγμα Taylor της u στα σημεία (x_i, t^1), $i = 1, \ldots, N$, έχουμε

$$\begin{aligned} u(x_i, t^1) &= u(x_i, t^0) + k u_t(x_i, 0) + \frac{k^2}{2} u_{tt}(x_i, 0) + \hat{\eta}_i^0 \\ &= g_0(x_i) + \alpha^2 k g_1(x_i) + \frac{\alpha^4 k^2}{2} u_{xx}(x_i, 0) + \hat{\eta}_i^0, \end{aligned} \tag{8.8}$$

όπου

$$|\hat{\eta}_i^0| \leq \frac{k^3}{6} \max_{t \in [0,T]} |\frac{\partial^3 u}{\partial t^3}(x_i, t)| \tag{8.9}$$

Στη συνέχεια, αν $\frac{\partial^4}{\partial x^4} u \in C([0, L] \times [0, T])$ και θεωρήσουμε την $\delta_{h,2}^c u(x_i, 0)$, βλ. (2.8), για να προσεγγίσουμε την $u_{xx}(x_i, 0)$, για $i = 1, \ldots, N$ στην (8.8), λόγω της (2.9), έχουμε

$$\begin{aligned} u(x_i, t^1) &= g_0(x_i) + \alpha^2 k g_1(x_i) + \frac{\lambda^2}{2}(u(x_{i+1}, 0) - 2u(x_i, 0) + u(x_{i-1}, 0)) \\ &\quad + \hat{\eta}_i^0 + \tilde{\eta}_i^0 \\ &= g_0(x_i) + \alpha^2 k g_1(x_i) + \frac{\lambda^2}{2}(g(x_{i+1}) - 2g(x_i) + g(x_{i-1})) + \eta_i^0, \end{aligned} \tag{8.10}$$

όπου $\eta_i^0 = \hat{\eta}_i^0 + \tilde{\eta}_i^0$ και

$$|\tilde{\eta}_i^0| \leq \frac{\alpha^4 k^2 h^2}{24} \max_{x \in [0,L]} |\frac{\partial^4 u}{\partial x^4}(x, 0)|. \tag{8.11}$$

Λόγω τώρα των (8.4), (8.9) και (8.11), έχουμε ότι ισχύει το ακόλουθο λήμμα.

Λήμμα 8.1. *Έστω u η λύση του* (8.1) *με $\frac{\partial^4}{\partial x^4} u$, $\frac{\partial^4}{\partial t^4} u \in C([0, L] \times [0, T])$. Τότε για την η_i^j που δίνεται στην* (8.3) *και* (8.10)*, έχουμε ότι υπάρχει σταθερά C ανεξάρτητη των k και h, τέτοια ώστε*

$$\max_{0 \leq j \leq M-1} \max_{1 \leq i \leq N} |\eta_i^j| \leq C(k^2 + h^2). \tag{8.12}$$

Επομένως, συμπληρώνουμε τις εξισώσεις (8.3) με τις ακόλουθες προσεγγίσεις U_i^1 της λύσης u του προβλήματος (8.1) στα σημεία (x_i, t^1), $i = 0, \ldots, N+1$, οι

οποίες ικανοποιούν για $i = 1, \ldots, N$,

$$\begin{aligned} U_i^1 &= g_0(x_i) + \alpha^2 k g_1(x_i) + \frac{\lambda^2}{2}(g_0(x_{i+1}) - 2g_0(x_i) + g_0(x_{i-1})) \\ &= \frac{\lambda^2}{2} g_0(x_{i+1}) + (1-\lambda^2) g_0(x_i) + \frac{\lambda^2}{2} g_0(x_{i-1}) + \alpha^2 k g_1(x_i). \end{aligned} \tag{8.13}$$

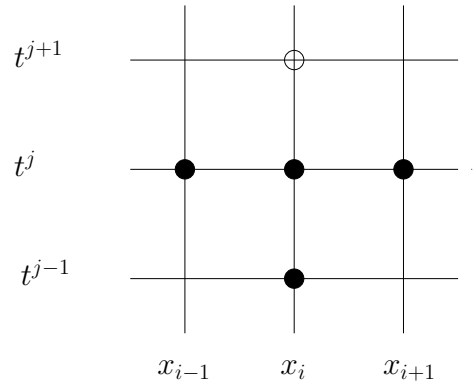
Στη συνέχεια, αν συμβολίσουμε με $U^j \in \mathbb{R}^N$ το διάνυσμα με συνιστώσες $U_1^j, \ldots, U_N^j$, $U^j = (U_1^j, \ldots, U_N^j)^T$, μπορούμε να γράψουμε το σύστημα των εξισώσεων (8.3) και (8.13) ισοδύναμα ως

$$\begin{aligned} U^{j+1} &= AU^j - U^{j-1}, \quad j = 1, \ldots, M-1, \\ U^1 &= \frac{1}{2} A G_0 + \alpha^2 k G_1, \quad \text{με } U^0 = G_0, \end{aligned} \tag{8.14}$$

όπου A είναι ο $N \times N$ πίνακας

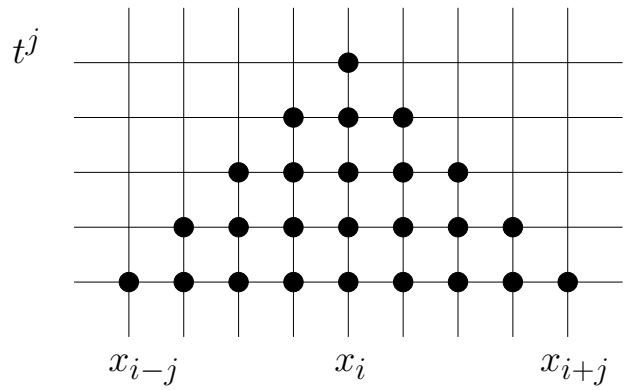
$$A = \begin{pmatrix} 2(1-\lambda^2) & \lambda^2 & 0 & & 0 \\ \lambda^2 & 2(1-\lambda^2) & \lambda^2 & 0 & \\ 0 & \ddots & \ddots & \ddots & 0 \\ & & \lambda^2 & 2(1-\lambda^2) & \lambda^2 \\ & & & \lambda^2 & 2(1-\lambda^2) \end{pmatrix},$$

και $G_0 = (g_0(x_1), \ldots, g_0(x_N))^T$ και $G_1 = (g_1(x_1), \ldots, g_1(x_N))^T$.



Σχήμα 8.1: Σημεία ενός πλέγματος για τον υπολογισμό της λύσης με τη μέθοδο των Courant, Friedrichs και Lewy. Με άσπρο σημειώνουμε τα σημεία που αντιστοιχούν σε άγνωστες τιμές της προσέγγισης και με μαύρο αυτά που αντιστοιχούν σε γνωστές τιμές.

Εύκολα παρατηρούμε ότι, χρησιμοποιώντας τη μέθοδο (8.3), η τιμή της προσέγγισης στο σημείο (x_i, t^j), εξαρτάται τελικά από τις τιμές στα σημεία $(x_{i-j}, 0)$, $\ldots$, $(x_i, 0)$, $\ldots$, $(x_{i+j}, 0)$, βλ. Σχήμα 8.2. Άρα, το χωρίο υπολογιστικής εξάρτησης της μεθόδου (8.14) για το (x_i, t^j) είναι το διάστημα $[x_{i-j}, x_{i+j}]$ του $x-$άξονα. Επίσης, λόγω της αναπαράστασης d'Alembert της ακριβούς λύσης u, βλ. (1.48), η τιμή της u στο σημείο (x_i, t^j) εξαρτάται από τις τιμές στα σημεία $(x_i - \alpha t^j)$ και $(x_i + \alpha t^j)$. Επομένως για να ικανοποιείται η συνθήκη CFL για τη μέθοδο (8.14) πρέπει τα σημεία $x_i \pm \alpha t^j \in [x_i - jh, x_i + jh]$, δηλαδή $\lambda \leq 1$.



Σχήμα 8.2: Χωρίο εξάρτησης της προσεγγιστικής λύσης με τη μέθοδο των Courant, Friedrichs και Lewy.

Όπως και για άλλες μεθόδους που είδαμε στις προηγούμενες παραγράφους για να δείξουμε τη σύγκλιση της μεθόδου (8.14) χρειάζεται να εφαρμόσουμε παρόμοια επιχειρήματα όπως αυτά της μεθόδου ενέργειας, βλ. Παράγραφο 3.4. Επομένως, διατυπώνουμε το ακόλουθο θεωρήμα χωρίς απόδειξη, λόγω της πολυπλόκοτητας αυτών των επιχειρημάτων, τα οποία δεν είναι στους σκοπούς αυτού του βιβλίου.

Θεώρημα 8.1. *Έστω ότι η λύση u του προβλήματος* (8.1) *είναι αρκετά ομαλή και U_i^j, $i = 1, \ldots, N$, $j = 0, \ldots, M$ η λύση της μεθόδου των Courant, Friedrichs και Lewy,* (8.14)*. Τότε, αν $0 < \lambda \leq 1$, υπάρχει μια σταθερά C, ανεξάρτητη των k, h, τέτοια ώστε*

$$\max_{1\leq j\leq M} \max_{0\leq i\leq N+1} |U_i^j - u(x_i, t^j)| \leq C(k^2 + h^2). \tag{8.15}$$

Παράδειγμα 8.1. Θεωρούμε το πρόβλημα (8.1), με $\alpha = 1$ και g_0 που δίνεται από την

$$g_0(x) = \begin{cases} \dfrac{1}{2}(1 - \cos(\dfrac{2\pi x}{c})), & 0 \leq x \leq c, \\ 0, & \text{διαφορετικά}, \end{cases} \tag{8.16}$$

όπου $0 < c < 1$ και $g_1(x) = -g_0'(x)$. Μπορούμε να δούμε ότι η ακριβής λύση στο $[0,1] \times [0,T]$, με $T < 2 - c$ δίνεται από τη συνάρτηση

$$u(x,t) = \begin{cases} g_0(x-t), & 0 \le x \le 1-c \\ g_0(x-t) - g_0(2-x-t), & 1-c \le t \le 2-c. \end{cases} \tag{8.17}$$

Υποθέτουμε ότι $c = 0.09$ και $T = 1.8$ οπότε για $N = 199$, θέτουμε $M = 330, 360$ και 400. Συνεπώς, για $M = 330$ ο λόγος $\lambda > 1$, ενώ για $M = 360$ είναι $\lambda = 1$ και για $M = 400$, $\lambda < 1$. Στο Σχήμα 8.3 βλέπουμε την ακριβή λύση u και τις προσεγγίσεις με τη μέθοδο Courant, Friedrichs και Lewy, για $t = 0, 0.1, 0.5, 0.9, 1.2$ και 1.8. Παρατηρούμε, λοιπόν, ότι για $M = 330$ η προσέγγιση δεν δίνει καλά αποτελέσματα, το οποίο οφείλεται στο γεγονός ότι ο λόγος $\lambda > 1$, ενώ στις άλλες δύο περιπτώσεις το σφάλμα είναι μικρό.

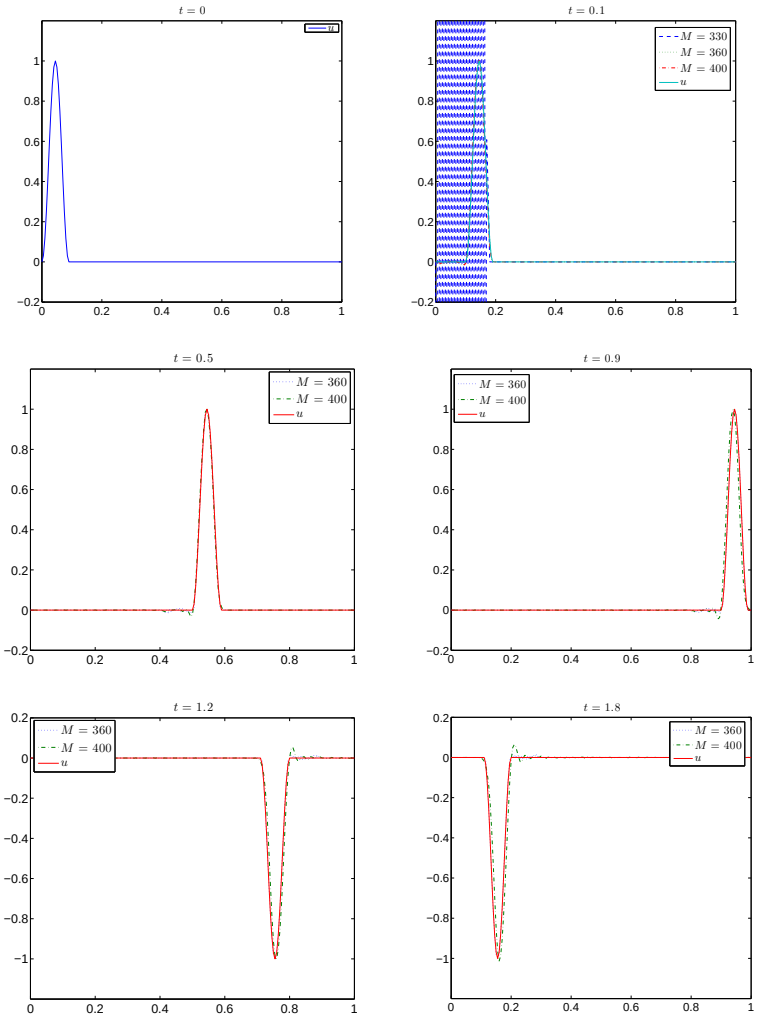
Παράδειγμα 8.2. Θεωρούμε το πρόβλημα (8.1) στο $[-2,2] \times [0,1]$, με $\alpha = 1$ και

$$g(x) = \begin{cases} e^{-\dfrac{1}{1-x^2}}, & \text{αν } -1 \le x \le 1, \\ 0, & \text{διαφορετικά,} \end{cases}$$

και $g_1 = 0$. Μπορούμε να δούμε ότι η ακριβής λύση στο $[-2,2] \times [0,1]$ είναι η $u(x,t) = \frac{1}{2}(g_0(x+t) + g_0(x-t))$. Χρησιμοποιούμε τη μέθοδο των Courant, Friedrichs και Lewy, όπου θεωρούμε διαμερίσεις του $[0,1]$ με $k = 0.05$, 0.025, 0.0125, 0.0063, 0.0031 και 0.0016 και του $[-2,2]$, με αντίστοιχο βήμα h, τέτοιο ώστε $k/h = 0.75$. Στον Πίνακα 8.1 βλέπουμε το σφάλμα $\bar{E} = \max_{0 \le i \le N+1} |U_i^M - u(x_i, t^M)|$ στο χρονικό επίπεδο $t^M = T = 1$ και την προσεγγιστική τάξη ακρίβειας p του σφάλματος.

k	$\bar{E}$	p
0.05	0.00640	
0.025	0.00117	2.450
0.0125	0.00043	1.433
0.0063	0.00011	1.915
0.0031	0.00003	1.933
0.0016	0.00001	2.023

Πίνακας 8.1: Το σφάλμα $\bar{E}$, στο χρονικό επίπεδο $t = T = 1$ που προκύπτει για το Παράδειγμα 8.2 με τη μεθόδο των Courant, Friedrichs και Lewy με $k/h = 0.75$ και η κατά προσέγγιση τάξη ακρίβειας p.



Σχήμα 8.3: Η ακριβής λύση $u(x,t)$ του Παραδείγματος 8.1 και οι προσεγγίσεις της για $N = 199$ και $M = 330, 360, 400$, με τη μέθοδο των Courant, Friedrichs και Lewy για $t = 0, 0.1, 0.5, 0.9, 1.2, 1.8$.

Θα μελετήσουμε τώρα την *ευστάθεια von Neumann* της μεθόδου των Courant, Friedrichs και Lewy. Αν υποθέσουμε, λοιπόν, ότι οι $U_i^j = \kappa^j e^{rx_i I}$ ικανοποιούν την (8.7), έχουμε

$$\kappa^{j+1} e^{rx_i I} = \kappa^j (\lambda^2 e^{rx_{i+1} I} + 2(1-\lambda^2) e^{rx_i I}) + \lambda^2 e^{rx_{i-1} I}) - \kappa^{j-1} e^{rx_i I}.$$

Χρησιμοποιώντας τώρα τις (5.23) και (5.18) έχουμε,

$$\begin{aligned} \kappa^2 =& \kappa(\lambda^2 e^{rhI} + 2(1-\lambda^2) + \lambda^2 e^{-rhI}) - 1 \\ =& \kappa(2\lambda^2 \cos(rh) + 2(1-\lambda^2)) - 1 \\ =& 2\kappa(1 - 2\lambda^2 \sin^2(\frac{rh}{2})) - 1. \end{aligned}$$

Από το οποίο προκύπτει

$$\kappa^2 - 2\kappa\sigma + 1 = 0, \quad \text{με } \sigma = (1 - 2\lambda^2 \sin^2(\frac{rh}{2})). \tag{8.18}$$

Συνεπώς

$$\kappa = \sigma \pm \sqrt{\sigma^2 - 1}.$$

Εύκολα βλέπουμε ότι αν $\sigma^2 > 1$, τότε υπάρχει μια ρίζα της (8.18) με $|\kappa| > 1$, και αν $\sigma^2 \leq 1$ τότε ότι $|\kappa|^2 = 1$. Συνεπώς, για να ισχύει ότι $\sigma^2 \leq 1$, αρκεί $\lambda \leq 1$. Άρα, η (8.7) είναι von Neumann ευσταθής, αν $\lambda \leq 1$.

8.2 Άλλες μέθοδοι και προβλήματα

Ακολουθώντας παρόμοια βήματα όπως στα Κεφάλαια 3 και 5 μπορούμε να θεωρήσουμε και άλλες μεθόδους πεπλεγμένες ή άμεσες για την επίλυση του (8.1) καθώς και προβλημάτων με συνοριακές συνθήκες τύπου Neumann ή άλλες. Για πιο λεπτομερή παρουσιάση παραπέμπουμε τον αναγνώστη στα (Iserles, 2009· Richtmyer & Morton, 1967· Morton & Mayers, 2005· Strikwerda, 2004).

8.3 Ταινίες γραφικών παραστάσεων

Σε αυτή την παράγραφο εμφανίζονται ταινίες με τις γραφικές παραστάσεις που παρουσιάστηκαν στα Παραδείγματα 8.1 και 8.2 για τη μεθόδο των Courant, Friedrichs και Lewy. Η προβολή των ταινιών στην οθόνη του H/Y μπορεί να γίνει, αν μετακινήσουμε τον κέρσορα (δείκτη) της οθόνης και επιλέξουμε, π.χ. με τη χρήση του "ποντίκιου", το αντίστοιχο παράδειγμα και, στη συνέχεια, την επιλογή "Play/Pause". Στην περίπτωση που η προβολή αυτού του βιβλίου στην οθόνη γίνεται μέσω του αντίστοιχου αρχείου μορφής pdf, συνίσταται η χρήση του πρόγραμματος Adobe Reader.

Play/Pause CFL μέθοδος με $\lambda = 1.1$ CFL μέθοδος με $\lambda = 1$

CFL μέθοδος με $\lambda = 0.9$

Σχήμα 8.4: Ταινίες των γραφικών παραστάσεων του Παραδείγματος 8.1 με τη μέθοδο των Courant, Friedrichs και Lewy.

Play/Pause CFL μέθοδος με $\lambda = 1.1$ CFL μέθοδος με $\lambda = 1$
CFL μέθοδος με $\lambda = 0.75$

Σχήμα 8.5: Ταινίες των γραφικών παραστάσεων του Παραδείγματος 8.2 με τη μέθοδο των Courant, Friedrichs και Lewy.

8.4 Ασκήσεις

8.1. Δείξτε ότι το αριθμητικό σχήμα

$$4\frac{U_i^{j+1}-2U_i^j+U_i^{j-1}}{k^2}=\alpha^2\big(\frac{U_{i+1}^{j+1}-2U_i^{j+1}+U_{i-1}^{j+1}}{h^2}$$
$$+2\frac{U_{i+1}^{j}-2U_i^{j}+U_{i-1}^{j}}{h^2}+\frac{U_{i+1}^{j-1}-2U_i^{j-1}+U_{i-1}^{j-1}}{h^2}\big),$$
$$\text{για } i=1,\dots,N, j=1,\dots,M-1,$$

προσεγγίζει την εξίσωση του κύματος (8.1) και το τοπικό σφάλμα διακριτοποιήσης η_i^j, ικανοποιεί

$$|\eta_i^j|\le C(k^2+h^2),$$

με C ανεξάρτητη των k και h.

8.2. Κατασκευάστε ένα αριθμητικό σχήμα για να προσεγγίσετε τη λύση της διαφορικής εξίσωσης

$$\begin{aligned} u_{tt}(x,t)+u_{tx}&=\alpha^2u_{xx}(x,t), \qquad x\in[0,L],\ t\in[0,T],\\ u(0,t)&=u(L,t)=0, \quad t\in[0,T],\\ u(x,0)&=g_0(x), \qquad x\in[0,L],\\ u_t(x,0)&=g_1(x), \qquad x\in[0,L]. \end{aligned}$$

Βρείτε το τοπικό σφάλμα διακριτοποίησης. Είναι von Neumann ευσταθές;

Βιβλιογραφία

Ακρίβης, Γ., & Αλικάκος, Ν. (2012). *Μερικές Διαφορικές Εξισώσεις*. Σύγχρονη Εκδοτική, Αθήνα.

Iserles, A. (2009). *A first course in the numerical analysis of differential equations* (Second ed.). Cambridge University Press, Cambridge.

Morton, K. W., & Mayers, D. F. (2005). *Numerical solution of partial differential equations* (Second ed.). Cambridge University Press, Cambridge.

Richtmyer, R. D., & Morton, K. W. (1967). *Difference methods for initial-value problems*. Interscience Publishers John Wiley & Sons, Inc., New York-London-Sydney.

Strikwerda, J. C. (2004). *Finite difference schemes and partial differential equations* (Second ed.). Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA.

Κεφάλαιο 9

Πεπερασμένες διαφορές για την ελλειπτική εξίσωση στις δύο διαστάσεις

Σε αυτό το κεφάλαιο θεωρούμε μια απλή ελλειπτική εξίσωση, στις δύο διαστάσεις. Θα κατασκευάσουμε μεθόδους πεπερασμένων διαφορών για το πρόβλημα συνοριακών τιμών με ομογενείς συνοριακές συνθήκες Dirichlet. Στη συνέχεια, θα δείξουμε ιδιότητες ευστάθειας, καθώς και κατάλληλες συνθήκες που πρέπει να ικανοποιεί η μέθοδος για να είναι ευσταθής. Τέλος, θα δείξουμε τη σύγκλιση της προσεγγιστικής λύσης στην ακριβή λύση.

9.1 Ελλειπτική εξίσωση

Θεωρούμε το ακόλουθο πρόβλημα: Ζητείται μια συνάρτηση $u : \Omega = (0,1) \times (0,1) \to \mathbb{R}$, τέτοια ώστε

$$\begin{aligned} -(u_{xx}(x,y) + u_{yy}(x,y)) + q(x,y)u(x,y) = f(x,y), \quad (x,y) \in \Omega, \\ u(x,y) = 0, \quad (x,y) \in \partial\Omega, \end{aligned} \tag{9.1}$$

όπου $q, f \in C(\Omega)$ και $q(x,y) \geq 0$, για κάθε $(x,y) \in \Omega$. Στη συνέχεια, θα συμβολίζουμε με $q_{\min} = \min_{(x,y)\in\Omega} q(x,y)$.

Θεωρούμε και πάλι έναν φυσικό αριθμό N και διαμερίζουμε το διάστημα $[0,1]$ σε $N+2$ ισαπέχοντα σημεία $0 = x_0 < x_1 < \cdots < x_N < x_{N+1} = 1$, όπου $h = x_{i+1} - x_i$, $i = 0, \ldots, N$. Επίσης θέτουμε $y^i = x_i$, $i = 0, \ldots, N+1$ και τότε τα σημεία (x_i, y^j), $i, j = 0, \ldots, N+1$ δημιουργούν έναν διαμερισμό του Ω. Τότε,

σε κάθε σημείο του διαμερισμού (x_i, y^j), $i, j = 1, \ldots, N$, θα ισχύει:

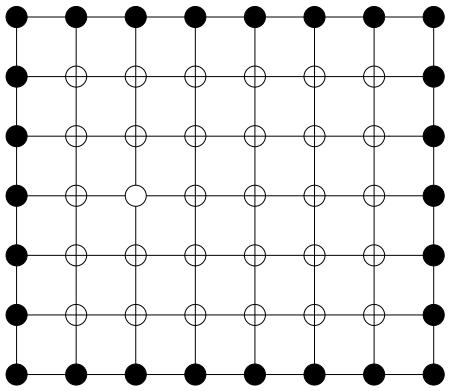
$$-\left(u_{xx}(x_i, y^j) + u_{yy}(x_i, y^j)\right) + q(x_i, y^j)u(x_i, y^j) = f(x_i, y^j), \tag{9.2}$$

για $i, j = 1, \ldots, N$. Θα κατασκευάσουμε όπως και στα προηγούμενα κεφάλαια προσεγγίσεις U_i^j των τιμών $u(x_i, y^j)$, $i, j = 0, \ldots, N+1$. Λόγω των ομογενών συνοριακών συνθηκών Dirichlet, θέτουμε $U_0^j = U_{N+1}^j = U_i^0 = U_i^{N+1} = 0$, $i, j = 0, \ldots, N+1$. Για την κατασκευή των υπολοίπων U_i^j θα προσεγγίσουμε τις $u_{xx}(x_i, y^j)$ και $u_{yy}(x_i, y^j)$ στην (9.2) με την $\delta_{h,2}^c u(x_i, y^j)$, θεωρώντας την πεπερασμένη διαφορά τη μία φορά ως προς τη μεταβλητή x και την αλλή ως προς τη μεταβλητή y, βλ. (2.8). Στο Σχήμα 9.1, βλέπουμε ένα παράδειγμα πλέγματος σημείων, όπου σημειώνουμε τα σημεία των άγνωστων ή γνωστών τιμών της λύσης u του (9.1). Έτσι, αν υποθέσουμε ότι $\frac{\partial^4 u}{\partial x^4}, \frac{\partial^4 u}{\partial y^4} \in C(\Omega)$, λόγω της (2.11), η (9.2) γίνεται,

$$\begin{aligned} -\Big(&\frac{u(x_{i+1}, y^j) - 2u(x_i, y^j) + u(x_{i-1}, y^j)}{h^2} \\ &+ \frac{u(x_i, y^{j+1}) - 2u(x_i, y^j) + u(x_i, y^{j-1})}{h^2}\Big) + q(x_i, y^j)u(x_i, y^j) \\ &= f(x_i, y^j) + \eta_i^j, \quad i, j = 1, \ldots, N, \end{aligned} \tag{9.3}$$

όπου

$$|\eta_i^j| \leq \frac{h^2}{12}\Big(\max_{(x,y)\in\Omega} |\frac{\partial^4 u}{\partial x^4} u(x, y)| + \max_{(x,y)\in\Omega} |\frac{\partial^4 u}{\partial y^4} u(x, y)|\Big). \tag{9.4}$$

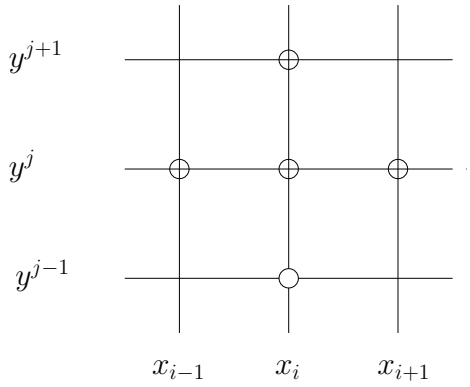


Σχήμα 9.1: Τα σημεία (x_i, y^j) του χωρίου Ω. Σημειώνουμε με άσπρο κύκλο τα σημεία όπου αναζητούμε τις τιμές U_i^j και με έντονο μαύρο τα σημεία που αντιστοιχούν στις γνωστές συνοριακές τιμές.

Θεωρούμε λοιπόν τις προσεγγίσεις U_i^j της $u(x_i, y^j)$, $i, j = 1, \ldots, N$, οι οποίες θα ικανοποιούν τις ακόλουθες εξισώσεις

$$-\Big(\frac{U_{i+1}^j - 2U_i^j + U_{i-1}^j}{h^2} + \frac{U_i^{j+1} - 2U_i^j + U_i^{j-1}}{h^2}\Big) + q(x_i, y^j)U_i^j \\ = f(x_i, y^j), \quad i, j = 1, \ldots, N. \tag{9.5}$$

Στο Σχήμα 9.2 φαίνεται ένα παράδειγμα πλέγματος όπου σημειώνουμε τις άγνωστες τιμές της προσέγγισης σύμφωνα με τη μέθοδο (9.5).



Σχήμα 9.2: Σημεία ενός πλέγματος για τον υπολογισμό της λύσης με τη μέθοδο (9.5). Με άσπρο σημειώνουμε τα σημεία που αντιστοιχούν σε άγνωστες τιμές της προσέγγισης.

Επομένως, αν συμβολίσουμε με $U \in \mathbb{R}^{N^2}$ το διάνυσμα με συνιστώσες $U_1^1, \ldots, U_N^1, U_1^2, \ldots, U_N^2, \ldots, U_1^N, \ldots, U_N^N$, $U = (U_1^1, \ldots, U_N^1, U_1^2, \ldots, U_N^2, \ldots, U_1^N, \ldots, U_N^N)^T$, μπορούμε να γράψουμε τις εξισώσεις (9.5) ως γραμμικό σύστημα

$$AU = h^2 F,$$

όπου $F = (f(x_1, y^1), \ldots, f(x_N, y^1), f(x_1, y^2), \ldots, f(x_N, y^2), \ldots, f(x_1, y^N), \ldots, f(x_N, y^N))^T$.

Για να κατανοήσουμε τη δομή του πίνακα A, θεωρούμε την περίπτωση $q = 0$ και παρατηρούμε ότι για $i = 1, \ldots, N$ και $j = 1$ οι εξισώσεις (9.5) γίνονται

$$-(U_{i+1}^1 - 4U_i^1 + U_{i-1}^1 + U_i^{j+1}) = h^2 f(x_i, y^j), \quad i = 1, \ldots, N. \tag{9.6}$$

Επομένως, οι N πρώτες γραμμές του πίνακα Α είναι

$$\underbrace{\left(\begin{array}{ccccc} 4 & -1 & 0 & & 0 \\ -1 & 4 & -1 & 0 & 0 \\ 0 & \ddots & \ddots & \ddots & 0 \\ & & -1 & 4 & -1 \\ & & & -1 & 4 \end{array}\right.}_{N \text{ στήλες}} \Bigg| \underbrace{\begin{array}{ccccc} 1 & 0 & 0 & & 0 \\ 0 & 1 & 0 & & 0 \\ 0 & & \ddots & & 0 \\ 0 & & & 1 & 0 \\ 0 & & & 0 & 1 \end{array}}_{N \text{ στήλες}} \Bigg| \underbrace{\left.\begin{array}{ccc} 0 & & 0 \\ 0 & \dots & 0 \\ 0 & \dots & 0 \\ 0 & & 0 \\ 0 & & 0 \end{array}\right)}_{N^2-2N \text{ στήλες}} \tag{9.7}$$

Συμβολίζουμε με B τον $N \times N$ πίνακα

$$B = \begin{pmatrix} 4 & -1 & 0 & & 0 \\ -1 & 4 & -1 & 0 & 0 \\ 0 & \ddots & \ddots & \ddots & 0 \\ & & -1 & 4 & -1 \\ & & & -1 & 4 \end{pmatrix}$$

και με $\mathcal{I}$ τον μοναδιαίο $N \times N$ πίνακα, και λόγω των εξισώσεων (9.5) για $i = 1, \dots, N$ και $j = 2$, οι $N+1$ έως $2N$ γραμμές του πίνακα Α είναι

$$\underbrace{\left(\begin{array}{ccccc} & & & & \\ \mathcal{I} & & B & & \mathcal{I} \\ & & & & \end{array}\right.}_{N \text{ στήλες}} \Bigg| \underbrace{\left.\begin{array}{ccc} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{array}\right)}_{N^2-3N \text{ στήλες}} \tag{9.8}$$

Οπότε, ο πίνακας A έχει τη μορφή

$$A = \begin{pmatrix} B & \mathcal{I} & 0 & 0 \\ \mathcal{I} & B & \mathcal{I} & 0 \\ 0 & \ddots & \ddots & \ddots \\ & \mathcal{I} & B & \mathcal{I} \\ & & \mathcal{I} & B \end{pmatrix}, \tag{9.9}$$

ο οποίος δεν έχει αυστηρά κυριαρχική διαγώνιο, είναι όμως αντιστρέψιμος. Αν υποθέσουμε τώρα ότι $q \neq 0$ και ότι $q_{\min} \geq 0$, τότε το γραμμικό σύστημα που προκύπτει από τις (9.5) είναι

$$(A + h^2 Q)U = h^2 F, \tag{9.10}$$

A, U και F όπως παραπάνω και Q ένας διαγώνιος $N^2 \times N^2$ πίνακας με στοιχεία στη διαγώνιο $q(x_1, y^1), \ldots, q(x_N, y^1), q(x_1, y^2), \ldots, q(x_N, y^2), \ldots, q(x_1, y^N), \ldots, q(x_N, y^N)$. Είναι απλό να δούμε ότι ο πίνακας $A + h^2 Q$ έχει όμοια μορφή με αυτή της (9.9), διότι αλλάζουν μόνο τα διαγώνια στοιχεία και έχει αυστηρά κυριαρχική διαγώνιο, αν $q_{\min} > 0$.

Επομένως, μπορούμε να επιλέξουμε μια άμεση μέθοδο, όπως είναι αυτή της απαλοιφής Gauss, για να λύσουμε το γραμμικό σύστημα (9.10), βλ. π.χ. (Ακρίβης & Δουγαλής, 2015). Ο πίνακας του γραμμικού συστήματος (9.10) είναι μεγάλος σε διάσταση και είναι και αραιός, έχει δηλαδή πολλά μηδενικά στοιχεία. Λόγω της (9.6), μπορούμε να δούμε ότι κάθε γραμμή του πίνακα A έχει το πολύ 5 μη-μηδενικά στοιχεία. Στην πράξη, δεν επιλέγουμε τη μέθοδο απαλοιφής Gauss για την αριθμητική επίλυση μεγάλων αραιών γραμμικών συστημάτων, αλλά άλλες μεθόδους όπως είναι οι επαναληπτικές μέθοδοι, Jacobi, Gauss-Seidel, SOR και άλλες, βλ. π.χ. (Ακρίβης & Δουγαλής, 2015· Strikwerda, 2004).

9.2 Σύγκλιση

Σε αυτήν την παράγραφο θα δείξουμε ότι οι συνιστώσες του διανύσματος U που αποτελεί τη λύση του γραμμικού συστήματος (9.10) προσεγγίζουν τις αντίστοιχες συνιστώσες του διανύσματος της ακριβούς λύσης u, $(u(x_1, y^1), \ldots, u(x_N, y^1), u(x_1, y^2), \ldots, u(x_N, y^2), \ldots, u(x_1, y^N), \ldots, u(x_N, y^N))^T$.

Με ανάλογο τρόπο όπως για το αντίστοιχο πρόβλημα στη μια διάσταση που είδαμε στο Κεφάλαιο 3 δείχνουμε την ευστάθεια και σύγκλιση της μεθόδου (9.5).

Θεώρημα 9.1. *Έστω $U \in \mathbb{R}^N$ η λύση του προβλήματος* (9.5)*, με $U_0^j = U_{N+1}^j = U_i^0 = U_i^{N+1} = 0$ και $q_{\min} > 0$. Τότε, ισχύει η ακόλουθη ανισότητα*

$$\max_{0 \le i,j \le N+1} |U_i^j| \le C \max_{(x,y)\in\Omega} |f(x,y)|. \tag{9.11}$$

Απόδειξη. Για να δείξουμε αυτό το θεώρημα βασιζόμαστε στην αντίστοιχη απόδειξη του Θεωρήματος 3.1. Επομένως, γράφουμε τη σχέση (9.5) στη μορφή

$$(4 + h^2 q(x_i)) U_i^j = U_{i+1}^j + U_{i-1}^j + U_i^{j+1} + U_i^{j-1} + h^2 f(x_i, y^j), \quad 1 \le i, j \le N.$$

Στη συνέχεια, επειδή $q_{\min} > 0$, έχουμε

$$(4 + h^2 q_{\min}) |U_i^j| \le 4 \max_{0 \le i,j \le N+1} |U_i| + h^2 \max_{(x,y)\in\Omega} |f(x,y)|,$$

από την οποία εύκολα προκύπτει η ζητούμενη εκτίμηση (9.11). □

Στη συνέχεια, δείχνουμε τη σύγκλιση της μεθόδου (9.5).

Θεώρημα 9.2. *Έστω ότι η λύση u του προβλήματος* (9.1) *είναι αρκετά ομαλή, $u \in C^4(\Omega)$. Τότε, αν $q_{\min} = \min_{(x,y)\in\Omega} q(x,y) > 0$, υπάρχει μια σταθερά C, ανεξάρτητη του h, τέτοια ώστε*

$$\max_{0\le i,j\le N+1} |U_i^j - u(x_i, y^j)| \le Ch^2. \tag{9.12}$$

Απόδειξη. Η απόδειξη προκύπτει με ανάλογο τρόπο όπως και η αντίστοιχη του Θεωρήματος 3.2. Θέτουμε $E_i^j = U_i^j - u(x_i, y^j)$, $i, j = 0, \ldots, N+1$, όπου λόγω των σχέσεων $U_0^j = U_{N+1}^j = U_i^0 = U_i^{N+1} = 0$, έχουμε $E_0^j = E_{N+1}^j = E_i^0 = E_i^{N+1} = 0$. Αφαιρούμε κατά μέλη τις (9.5) και (9.3), οπότε παίρνουμε για $i, j = 1, \ldots, N$,

$$(4 + q(x_i, y^j)h^2)E_i^j = E_i^{j+1} + E_i^{j-1} + E_{i+1}^j + E_{i-1}^j + h^2\eta_i^j, \tag{9.13}$$

όπου, λόγω του Λήμματος 2.2, έχουμε ότι

$$\max_{1\le i\le N} |\eta_i| \le \frac{h^2}{12} \max_{a\le x\le b} |u^{(4)}(x)|. \tag{9.14}$$

Θέτουμε τώρα $\bar{E} = \max_{1\le i\le N} |E_i|$, $\bar{\eta} = \max_{1\le i\le N} |\eta_i|$ και επειδή $q_{\min} > 0$ από την (3.22) προκύπτει

$$(2 + q_{\min}h^2)|E_i| \le 2\bar{E} + h^2\bar{\eta}.$$

Συνεπώς

$$q_{\min}h^2 \max_{1\le i\le N} |E_i| \le h^2\bar{\eta},$$

η οποία λόγω της (9.14) δίνει τη ζητούμενη ανισότητα. □

Παρατήρηση 9.1. Η υπόθεση ότι $q_{\min} > 0$ στο Θεώρημα 9.2 γίνεται για καθαρά τεχνικούς λόγους που αφορούν την απόδειξη του θεωρήματος. Χρησιμοποιώντας παρόμοια επιχειρήματα όπως αυτά της μεθόδου ενέργειας στη Παράγραφο 3.4 μπορούμε να δείξουμε ανάλογα αποτελέσματα όπως αυτό του Θεωρήματος 9.2, βλ. π.χ. (Larsson & Thomée, 2009).

9.3 Άλλες μέθοδοι και προβλήματα

Σε αυτό το κεφάλαιο θεωρήσαμε ότι το χωρίο Ω είναι τετραγωνικό και ότι η διαμέριση είναι ομοιόμορφη. Παρόμοια αποτελέσματα, όμως, ισχύουν και σε άλλα τετράπλευρα χωρία και για μη ομοιόμορφους διαμερισμούς καθώς και για άλλες συνοριακές συνθήκες όπως π.χ. Neumann ή Robin, βλ. (Holmes, 2007· Iserles, 2009· Knabner & Angermann, 2003· Larsson & Thomée, 2009· Morton & Mayers, 2005· Jovanović & Süli, 2014· Strikwerda, 2004).

9.4 Ασκήσεις

9.1. Θεωρήστε έναν διαμερισμό του $[0,1]$ σε $N+2$ σημεία με βήμα $h = 1/(N+1)$, $x_i = ih$, $i = 0, \ldots, N+1$, και άλλον έναν σε $M+2$ σημεία με βήμα $k = 1/(M+2)$, $y^j = jk$, $j = 0, \ldots, M+1$. Για τον διαμερισμό του $[0,1] \times [0,1]$ στα σημεία (x_i, y^j), $i = 1, \ldots, N$, $j = 1, \ldots, M$, διατυπώστε την αριθμητική μέθοδο για την προσέγγιση της λύσης u της (9.1) με τοπικό σφάλμα διακριτοποίησης η_i^j

$$|\eta_i^j| \leq C(k^2 + h^2),$$

όπου C σταθερά ανεξάρτητη των k, h.

9.2. Θεωρήστε έναν διαμερισμό του $[0,1]$ σε $N+2$ σημεία με βήμα $h = 1/(N+1)$, $x_i = y^i = ih$, $i = 0, \ldots, N+1$. Για τον διαμερισμό του $[0,1] \times [0,1]$ στα σημεία (x_i, y^j), $i, j = 1, \ldots, N$, διατυπώστε μια αριθμητική μέθοδο για τη προσέγγιση της λύσης u του προβλήματος

$$\begin{aligned} -(u_{xx}(x,y) + u_{yy}(x,y)) + au_x(x,y) + bu_y(x,y) + cu(x,y) \\ = f(x,y), \quad (x,y) \in \Omega, \\ u(x,y) = 0, \quad (x,y) \in \partial\Omega, \end{aligned}$$

με $a, b, c \geq 0$ σταθερές, με τοπικό σφάλμα διακριτοποίησης η_i^j

$$|\eta_i^j| \leq Ch^2,$$

όπου C σταθερά ανεξάρτητη του h.

9.3. Θεωρήστε έναν διαμερισμό του $[0,1]$ σε $N+2$ σημεία με βήμα $h = 1/(N+1)$, $x_i = y^i = ih$, $i = 0, \ldots, N+1$. Για τον διαμερισμό του $[0,1] \times [0,1]$ στα σημεία (x_i, y^j), $i, j = 0, \ldots, N+1$, διατυπώστε μια αριθμητική μέθοδο για τη προσέγγιση της λύσης u του προβλήματος

$$\begin{aligned} -(u_{xx}(x,y) + u_{yy}(x,y)) + u(x,y) = f(x,y), \quad (x,y) \in \Omega, \\ u_x(0,y) = u_x(1,y) = 0, \quad \text{για } y \in (0,1), \\ u_y(x,0) = u_y(0,1) = 0, \quad \text{για } x \in (0,1), \end{aligned}$$

Ποιο είναι το τοπικό σφάλμα διακριτοποίησης η_i^j;

Βιβλιογραφία

Ακρίβης, Γ., & Δουγαλής, Β. (2015). *Εισαγωγή στην Αριθμητική Ανάλυση*. Πανεπιστημιακές Εκδόσεις Κρήτης, Ηράκλειο.

Holmes, M. H. (2007). *Introduction to numerical methods in differential equations* (Vol. 52). Springer, New York.

Iserles, A. (2009). *A first course in the numerical analysis of differential equations* (Second ed.). Cambridge University Press, Cambridge.

Jovanović, B. S., & Süli, E. (2014). *Analysis of finite difference schemes* (Vol. 46). Springer, London.

Knabner, P., & Angermann, L. (2003). *Numerical methods for elliptic and parabolic partial differential equations* (Vol. 44). Springer-Verlag, New York.

Larsson, S., & Thomée, V. (2009). *Partial differential equations with numerical methods* (Vol. 45). Springer-Verlag, Berlin.

Morton, K. W., & Mayers, D. F. (2005). *Numerical solution of partial differential equations* (Second ed.). Cambridge University Press, Cambridge.

Strikwerda, J. C. (2004). *Finite difference schemes and partial differential equations* (Second ed.). Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA.

Κεφάλαιο 10

Πεπερασμένα στοιχεία στις πολλές διαστάσεις

Σε αυτή την παράγραφο θα μελετήσουμε τη μέθοδο πεπερασμένων στοιχείων για τη διακριτοποίηση μιας διαφορικής εξίσωσης στις πολλές διαστάσεις. Πιο συγκεκριμένα, θα θεωρήσουμε την ελλειπτική εξίσωση σε ένα χωρίο $\Omega \subset \mathbb{R}^d$, $d = 2$, 3 και θα τη διακριτοποιήσουμε χρησιμοποιώντας τη μέθοδο πεπερασμένων στοιχείων. Στη συνέχεια, θα θεωρήσουμε την εξίσωση της θερμότητας στο $[0, T] \times \Omega$, όπου $T > 0$, $\Omega \subset \mathbb{R}^d$, και θα χρησιμοποιήσουμε τη μέθοδο πεπερασμένων στοιχείων για τη διακριτοποίηση της διαφορικής εξίσωσης ως προς τη χωρική μεταβλητή στο Ω. Στη συνέχεια, θα διακριτοποιήσουμε και ως προς τη χρονική μεταβλήτη στο $[0, T]$, χρησιμοποιώντας μια από τις μεθόδους που θεωρήσαμε στα προηγούμενα κεφάλαια για την εξισώση της θερμότητας.

10.1 Μεταβολικό πρόβλημα

Στην παράγραφο αυτή θα υποθέσουμε ότι το Ω είναι ένα φραγμένο χωρίο στον $\mathbb{R}^d$, $d \geq 2$, με σύνορο $\partial\Omega$ και θα θεωρήσουμε προβλήματα συνοριακών τιμών για *ελλειπτικές* μερικές διαφορικές εξισώσεις. Τυπικά παραδείγματα ελλειπτικών μερικών διαφορικών εξισώσεων αποτελούν, για παράδειγμα, η εξίσωση του Laplace $\Delta u = 0$ και η εξίσωση του Poisson $-\Delta u = f$, όπου για $u : \mathbb{R}^d \to \mathbb{R}$, $x = (x_1, \ldots, x_d)^T$, συμβολίζουμε $\Delta u = \sum_{i=1}^{d} \frac{\partial^2 u}{\partial x_i^2}$. Γενικότερα, θεωρούμε τη μερική διαφορική εξίσωση δεύτερης τάξης

$$-\sum_{i,j=1}^{d} \frac{\partial}{\partial x_i}\left(a_{ij}(x)\frac{\partial u(x)}{\partial x_j}\right) + a_0(x)u(x) = f(x), \quad x \in \Omega, \tag{10.1}$$

όπου $a_{ij} \in C^1(\bar{\Omega})$, $i,j = 1,\ldots,d$ και με $\bar{\Omega}$ συμβολίζουμε τη κλειστότητα του Ω, $\bar{\Omega} = \Omega \cup \partial\Omega$. Επίσης, υποθέτουμε ότι $a_{ij} = a_{ji}$, $i,j = 1,\ldots,d$ και ότι ο πίνακας $A = [a_{ij}]_{i,j=1}^d$ είναι ομοιόμορφα θετικά ορισμένος, δηλαδή υπάρχει σταθερά $\underline{c} > 0$, τέτοια ώστε

$$\sum_{i,j=1}^{d} a_{ij}(x)\xi_i\xi_j \geq \underline{c}\sum_{i=1}^{d}\xi_i^2, \quad \forall\, x \in \bar{\Omega},\ \xi \in \mathbb{R}^d. \tag{10.2}$$

Υποθέτουμε, επίσης, ότι $f, a_0 \in C(\bar{\Omega})$ και ότι $a_0 \geq 0$ στο $\bar{\Omega}$. Η συνθήκη (10.2) καλείται συνθήκη της *ομοιόμορφης ελλειπτικότητας* και η εξίσωση (10.1) *ελλειπτική* εξίσωση. Σε προβλήματα συνοριακών τιμών, η εξίσωση (10.1) συνοδεύεται, συνήθως, από μια από τις παρακάτω συνοριακές συνθήκες, στο σύνορο του Ω, $\partial\Omega$, όπου g είναι μια δοσμένη συνάρτηση:

α΄. αν $u = g$ στο $\partial\Omega$, τότε η συνοριακή συνθήκη είναι τύπου Dirichlet,

β΄. αν $\sum_{i,j=1}^d a_{ij}\frac{\partial u}{\partial x_i}\nu_j = g$ στο $\partial\Omega$, όπου $\nu = (\nu_1,\ldots,\nu_d)^T$ το μοναδιαίο εξωτερικό κάθετο διάνυσμα στο $\partial\Omega$, τότε η συνοριακή συνθήκη είναι τύπου Neumann,

γ΄. και αν $\sum_{i,j=1}^d a_{ij}\frac{\partial u}{\partial x_i}\nu_j + \sigma u = g$ στο $\partial\Omega$, όπου $\sigma \geq 0$ στο $\partial\Omega$, τότε η συνοριακή συνθήκη είναι τύπου Robin.

Ξεκινούμε με τη μελέτη του ελλειπτικού προβλήματος με ομογενείς συνοριακές συνθήκες τύπου Dirichlet, δηλαδή,

$$-\sum_{i,j=1}^{d}\frac{\partial}{\partial x_i}\left(a_{ij}(x)\frac{\partial u}{\partial x_j}\right) + a_0(x)u = f(x), \quad x \in \Omega, \tag{10.3}$$

$$u = 0, \quad x \in \partial\Omega, \tag{10.4}$$

όπου οι a_{ij}, a_0 και f είναι όπως και στην (10.1). Όπως και στο Κεφάλαιο 4 για το πρόβλημα των δύο σημείων, θα γενικεύσουμε την έννοια της λύσης του προβλήματος (10.3)–(10.4) εξασθενώντας τις προφανείς συνθήκες ομαλότητας που πρέπει να ικανοποιεί. Συγκεκριμένα, θα λέμε ότι μια συνάρτηση $u \in C^2(\Omega) \cap C(\bar{\Omega})$ η οποία ικανοποιεί τις (10.3) και (10.4) ονομάζεται *κλασική λύση*. Από τη θεωρία των μερικών διαφορικών εξισώσεων, δείτε για παράδειγμα (Gilbarg & Trudinger, 1983), προκύπτει ότι αν το $\partial\Omega$ είναι αρκετά ομαλό και οι συναρτήσεις a_{ij}, a_0 και f ικανοποιούν τις παραπάνω συνθήκες ομαλότητας και τη συνθήκη ομοιόμορφης ελλειπτικότητας (10.2), τότε το πρόβλημα (10.3)–(10.4) έχει μοναδική κλασική λύση.

Στη συνέχεια, θα συμβολίζουμε με $C_0^\kappa(\Omega)$, $C_0^\kappa(\Omega) = \{v \in C^\kappa(\Omega) : v = 0 \text{ στο } \partial\Omega\}$. Αν τώρα u είναι η κλασική λύση του (10.3)–(10.4), τότε για κάθε $v \in C_0^1(\Omega)$ έχουμε

$$-\sum_{i,j=1}^{d} \int_\Omega \frac{\partial}{\partial x_i}\left(a_{ij}(x)\frac{\partial u(x)}{\partial x_j}\right) v(x)\,dx + \int_\Omega a_0(x)u(x)v(x)\,dx = \int_\Omega f(x)v(x)\,dx.$$

Ολοκληρώνοντας κατά μέρη το αριστερό μέλος της προηγούμενης σχέσης και χρησιμοποιώντας το γεγονός ότι $v = 0$ στο $\partial\Omega$, λαμβάνουμε για κάθε $v \in C_0^1(\Omega)$,

$$\sum_{i,j=1}^{d} \int_\Omega a_{ij}(x)\frac{\partial u(x)}{\partial x_j}\frac{\partial v(x)}{\partial x_i}\,dx + \int_\Omega a_0(x)u(x)v(x)\,dx = \int_\Omega f(x)v(x)\,dx. \tag{10.5}$$

Για την ισχύ της (10.5) δεν είναι απαραίτητο να υποθέσουμε ότι $u \in C^2(\Omega)$. Αρκεί, για παράδειγμα, η u όσο και οι μερικές παράγωγοί $\frac{\partial u}{\partial x_i}$, $i = 1, \ldots, d$, να είναι τετραγωνικά ολοκληρώσιμες συναρτήσεις και, επιπλέον η u να μηδενίζεται στο σύνορο του Ω. Η περαιτέρω γενίκευση του χαρακτηρισμού της λύσης του προβλήματος (10.3)–(10.4) απαιτεί στοιχεία συναρτησιακής ανάλυσης τα οποία θα παρουσιάσουμε εν συντομία και παραπέμπουμε τον ενδιαφερόμενο αναγνώστη στα βιβλία (Adams, 1975· Brezis, 1997· Royden, 1988).

Οι χώροι Sobolev

Θα περιοριστούμε σε πραγματικές συναρτήσεις ορισμένες στο χωρίο Ω οι οποίες είναι ολοκληρώσιμες με την έννοια του Lebesgue, βλ. π.χ. (Brezis, 1997· Royden, 1988). Θα συμβολίζουμε με

$$\int_\Omega f(x)\,dx,$$

το ολοκλήρωμα Lebesgue της συνάρτησης f στο χωρίο Ω. Αν μια συνάρτηση f είναι Riemann ολοκληρώσιμη, τότε το ολοκλήρωμα Lebesgue της f ταυτίζεται με το ολοκλήρωμα Riemann της f. Επίσης, ορίζουμε την ακόλουθη νόρμα

$$\|f\|_{L_p(\Omega)} = \begin{cases} \left(\int_\Omega |f(x)|^p\,dx\right)^{1/p}, & \text{για } 1 \le p < \infty, \\ \inf\{C : |f(x)| \le C : \text{ σχεδόν παντού στο } \Omega\}, & \text{για } p = \infty. \end{cases}$$

Οι χώροι Lebesgue ορίζονται τώρα ως εξής:

$$L_p(\Omega) = \{f : \|f\|_{L_p(\Omega)} < \infty\}.$$

Στη συνέχεια, για χάριν συντομίας, θα παραλείπουμε τον δείκτη $L_2(\Omega)$ και θα γράφουμε $\|\cdot\| = \|\cdot\|_{L_2(\Omega)}$. Στον χώρο $L_2(\Omega)$ μπορούμε να ορίσουμε το ακόλουθο εσωτερικό γινόμενο ως εξής:

$$(u, v) = \int_\Omega u(x)v(x)\,dx.$$

Για μια παραγωγίσιμη συνάρτηση v στο Ω, έχουμε

$$\int_\Omega v(x)\frac{\partial\phi(x)}{\partial x_i}\,dx = -\int_\Omega \frac{\partial v(x)}{\partial x_i}\phi(x)\,dx, \quad 1 \le i \le d,\ \forall\,\phi \in C_0^1(\Omega).$$

Θα γενικέυσουμε την έννοια της μερικής παραγώγου μιας συνάρτησης με τον ακόλουθο τρόπο. Θα λέμε ότι v έχει *ασθενείς μερικές παραγώγους*, αν υπάρχουν συναρτήσεις g_i, $i = 1, \ldots, d$, τέτοιες ώστε

$$\int_\Omega v\frac{\partial\phi}{\partial x_i}\,dx = -\int_\Omega g_i\phi\,dx, \quad 1 \le i \le d,\ \forall\,\phi \in C_0^1(\Omega).$$

Αν η συνάρτηση $v \in C^1(\bar{\Omega})$, τότε οι ασθενείς μερικές παράγωγοι ταυτίζονται με την κλασική μερική παράγωγο $\frac{\partial v}{\partial x_i}$. Όταν υπάρχει η ασθενής μερική παράγωγος μιας συνάρτησης v, θα την συμβολίζουμε και πάλι με $\frac{\partial v}{\partial x_i}$.

Με όμοιο τρόπο τώρα ορίζουμε και την ασθενή μερική παράγωγο $D^\alpha v$, όπου

$$D^\alpha v = \frac{\partial^{|\alpha|} v}{\partial x_1^{\alpha_1} \ldots \partial x_d^{\alpha_d}}.$$

με α έναν πολυδείκτη $\alpha = (\alpha_1, \ldots, \alpha_d)$. Έτσι $D^\alpha v$ είναι η ασθενής παράγωγος της v, αν

$$\int_\Omega vD^\alpha\phi\,dx = (-1)^{|\alpha|}\int_\Omega D^\alpha v\phi\,dx, \quad 1 \le i \le d,\ \forall\,\phi \in C_0^{|\alpha|}(\Omega).$$

Ορίζουμε τώρα τον χώρο Sobolev $H^k(\Omega)$ ως τον χώρο των συναρτήσεων v όπου όλες οι ασθενείς παράγωγοι μέχρι τάξεως k ανήκουν στον $L_2(\Omega)$,

$$H^k(\Omega) = \{v \in L_2(\Omega) : D^\alpha v \in L_2(\Omega) \text{ για } |\alpha| \le k\}.$$

Επίσης, θεωρούμε το εσωτερικό γινόμενο $(v, w)_k$ και την αντίστοιχη νόρμα $\|v\|_k$

$$(v, w)_k = (v, w)_{H^k} = \sum_{|\alpha|\le k} \left(D^\alpha v, D^\alpha w\right),$$

$$\|v\|_k = \|v\|_{H^k} = (v,v)_k^{1/2} = \left(\sum_{|\alpha| \le k} \|D^\alpha v\|^2 \right)^{1/2}.$$

Σημειώνουμε ακόμα ότι αν $v \in L_2(\Omega)$, τότε ο περιορισμός της v σύνορο $\partial\Omega$ του χωρίου Ω δεν είναι μια καλά ορισμένη συνάρτηση. Φυσικά, αν $v \in C(\bar{\Omega})$, τότε η v μπορεί να ορισθεί στο $\partial\Omega$. Μπορούμε να δείξουμε, βλ. π.χ. (Adams, 1975· Brezis, 1997), ότι αν $v \in H^1(\Omega)$ και το $\partial\Omega$ είναι ομαλό ή πολυγωνικό, τότε ο περιορισμός της v στο $\partial\Omega$ είναι μια καλά ορισμένη συνάρτηση.

Θα συμβολίζουμε στη συνέχεια $H_0^1(\Omega)$ τον χώρο των συναρτήσεων του $H^1(\Omega)$ που μηδενίζονται στο σύνορο $\partial\Omega$,

$$H_0^1(\Omega) = \{v \in H^1(\Omega) : v\big|_{\partial\Omega} = 0\}.$$

Επιστρέφοντας τώρα στη σχέση (10.5) παρατηρούμε ότι μπορούμε να γενικεύσουμε την έννοια της λύσης του προβλήματος συνοριακών τιμών (10.3) υποθέτοντας ότι $u \in L_2(\Omega)$ και $\frac{\partial u}{\partial x_i} \in L_2(\Omega), i = 1, \ldots, d$. Μια και η u πρέπει να ικανοποιεί την ομογενή συνοριακή συνθήκη Dirichlet, είναι φυσικό να θεωρήσουμε το παρακάτω πρόβλημα: Να βρεθεί $u \in H_0^1(\Omega)$, τέτοια ώστε

$$\sum_{i,j=1}^{d} (a_{ij} \frac{\partial u}{\partial x_j}, \frac{\partial v}{\partial x_i}) + (a_0 u, v) = (f, v), \quad \forall\, v \in H_0^1(\Omega). \tag{10.6}$$

Στη συνέχεια, εισάγουμε τη διγραμμική μορφή $a : H^1(\Omega) \times H^1(\Omega) \to \mathbb{R}$,

$$a(v,w) = \sum_{i,j=1}^{d} (a_{ij} \frac{\partial v}{\partial x_j}, \frac{\partial w}{\partial x_i}), + (a_0 v, w), \quad \forall\, v, w \in H^1(\Omega) \tag{10.7}$$

και το γραμμικό συναρτησιακό $l : L_2(\Omega) \to \mathbb{R}$,

$$l(v) = (f, v) \quad \forall\, v \in L_2(\Omega). \tag{10.8}$$

Με το συμβολισμό αυτό, το πρόβλημα (10.6) γράφεται: Να βρεθεί $u \in H_0^1(\Omega)$, τέτοια ώστε

$$a(u,v) = l(v) \quad \forall\, v \in H_0^1(\Omega). \tag{10.9}$$

Η εξίσωση (10.6) (ή (10.9)) ονομάζεται *ασθενής* ή *μεταβολική μορφή* του προβλήματος συνοριακών τιμών (10.3)–(10.4) και η λύση του *ασθενής λύση*.

Αν σε έναν γραμμικό χώρο V με εσωτερικό γινόμενο $(\cdot,\cdot)_V$ κάθε ακολουθία Cauchy $\{v_n\}_{n=1}^{\infty} \subset V$ είναι συγκλίσουσα σε ένα στοιχείο $v \in V$, ως προς τη νόρμα $\|\cdot\|_V$ που παράγεται από το $(\cdot,\cdot)_V$, τότε ο χώρος καλείται χώρος Hilbert. Η

ύπαρξη μοναδικής ασθενούς λύσης μπορεί να αποδειχθεί με τη βοήθεια του θεωρήματος Lax–Milgram, η απόδειξη του οποίου μπορεί να βρεθεί, για παράδειγμα, στα (Brezis, 1997· Δουγαλής, 2013). Στη συνέχεια, διατυπώνουμε το Θεώρημα Lax–Milgram.

Θεώρημα 10.1 (Lax–Milgram)**.** *Έστω $(V, \|\cdot\|_V)$ ένας (πραγματικός) χώρος Hilbert και $a(\cdot,\cdot)$ μια διγραμμική μορφή στο $V \times V$, τέτοια ώστε*

$$\exists\, \alpha > 0 \;\; \forall\, v \in V \;\; a(v,v) \geq \alpha \|v\|_V^2, \tag{10.10}$$

$$\exists\, \beta > 0 \;\; \forall\, v, w \in V \;\; |a(v,w)| \leq \beta \|v\|_V \|w\|_V. \tag{10.11}$$

Έστω, επίσης, $l : V \to \mathbb{R}$ ένα γραμμικό συναρτησιακό για το οποίο

$$\exists\, \gamma > 0 \;\; \forall\, v \in V \;\; |l(v)| \leq \gamma \|v\|_V. \tag{10.12}$$

Τότε, υπάρχει μοναδικό $u \in V$, τέτοιο ώστε $a(u,v) = l(v)$ για κάθε $v \in V$.

Είναι εύκολο να επαληθεύσουμε τις υποθέσεις (10.10)–(10.12) του Θεωρήματος Lax–Milgram για το πρόβλημα (10.3)–(10.4). Πράγματι, με $V = H_0^1(\Omega)$ και $\|\cdot\|_V = \|\cdot\|_1$, έχουμε από την (10.7) με χρήση της ανισότητας Cauchy–Schwarz

$$\begin{aligned} |a(v,w)| &\leq \sum_{i,j=1}^{d} \max_{x\in\bar{\Omega}} |a_{ij}(x)| \| \frac{\partial v}{\partial x_j} \| \, \| \frac{\partial w}{\partial x_i} \| + \max_{x\in\bar{\Omega}} |a_0(x)| \|v\| \, \|w\| \\ &\leq C \left\{ \sum_{i,j=1}^{d} (\| \frac{\partial v}{\partial x_j} \| \, \| \frac{\partial w}{\partial x_i} \|) + \|v\| \, \|w\| \right\} \leq \beta \|v\|_1 \|w\|_1, \end{aligned}$$

όπου

$$C = \max \left\{ \max_{1\leq i,j\leq d} \max_{x\in\bar{\Omega}} |a_{ij}(x)|, \; \max_{x\in\bar{\Omega}} |a_0(x)| \right\}.$$

Οπότε ισχύει η (10.11). Για να δείξουμε τη σχέση (10.10), θα χρειαστούμε το ακόλουθο λήμμα, η απόδειξη του οποίου μπορεί να βρεθεί π.χ. στα (Brezis, 1997· Ciarlet, 2002).

Λήμμα 10.1 (Ανισότητα Poincaré–Friedrichs)**.** *Για $v \in H_0^1(\Omega)$, έχουμε*

$$\|v(x)\| \leq c_* \left(\sum_{i=1}^{d} \| \frac{\partial v}{\partial x_i} \|^2 \right)^{1/2} \qquad \forall\, v \in H_0^1(\Omega). \tag{10.13}$$

Στη συνέχεια, χρησιμοποιώντας την (10.2) και το γεγονός ότι $a_0 \geq 0$ στο $\bar{\Omega}$, έχουμε

$$a(v,v) \geq \underline{c} \sum_{i=1}^{d} \| \frac{\partial v}{\partial x_i} \|^2 + (a_0 v, v) \geq \underline{c} \sum_{i=1}^{d} \| \frac{\partial v}{\partial x_i} \|^2. \tag{10.14}$$

Επομένως, συνδυάζοντας τις (10.13) και (10.14), λαμβάνουμε

$$a(v,v) \geq \frac{\underline{c}}{c_*}\|v\|^2, \tag{10.15}$$

και αθροίζοντας τις (10.14) και (10.15), παίρνουμε την (10.10) με $\alpha = \underline{c}/(1+c_*)$. Τέλος, με χρήση της ανισότητας Cauchy–Schwarz, έχουμε

$$|l(v)| = |(f,v)| \leq \|f\|\|v\| \leq \|f\|\|v\|_1, \tag{10.16}$$

επομένως, ισχύει η (10.12) με $\gamma = \|f\|$. Έχοντας επαληθεύσει όλες τις υποθέσεις του θεωρήματος Lax–Milgram, μπορούμε να συμπεράνουμε ότι το πρόβλημα (10.9) έχει μοναδική λύση. Επιπλέον, από τις σχέσεις (10.16) και (10.10) έχουμε

$$\alpha\|u\|_1^2 \leq a(u,u) = l(u) \leq \|f\|\,\|u\| \leq \|f\|\,\|u\|_1,$$

από την οποία λαμβάνουμε το εκ των προτέρων φράγμα για τη λύση

$$\|u\|_1 \leq \frac{1}{\alpha}\|f\|.$$

Παρατήρηση 10.1. Θεωρούμε το πρόβλημα συνοριακών τιμών στο $\Omega = (-1,1)$,

$$-\Delta u = \operatorname{sgn}\left(\frac{1}{2} - x\right), \quad x \in \Omega, \tag{10.17}$$

$$u = 0, \quad x \in \partial\Omega. \tag{10.18}$$

Είναι προφανές ότι το πρόβλημα (10.17)–(10.18) δεν μπορεί να έχει κλασική λύση $u \in C^2(\Omega)\cap C(\bar{\Omega})$, γιατί διαφορετικά η συνάρτηση sgn θα έπρεπε να είναι συνεχής στο Ω, το οποίο δεν συμβαίνει. Παρόλα αυτά είναι εύκολο να βεβαιωθεί ότι ισχύουν όλες οι υποθέσεις του θεωρήματος Lax–Milgram και, επομένως, το πρόβλημα (10.17)–(10.18) έχει μοναδική λύση $u \in H_0^1(\Omega)$.

Μπορούμε, με παρόμοιο τρόπο, να αποδείξουμε την ύπαρξη και μοναδικότητα της λύσης ελλειπτικών προβλημάτων συνοριακών τιμών με άλλες συνοριακές συνθήκες, όπως τις ομογενείς συνθήκες Neumann ή τις συνθήκες Robin, βλ. π.χ. (Brezis, 1997). Μπορούμε, επίσης, να θεωρήσουμε προβλήματα μεικτών συνοριακών συνθηκών, όπως το

$$\begin{aligned} -\Delta u + u &= f \quad \text{στο } \Omega \\ u &= 0 \quad \text{στο } \Gamma_1 \text{ και } \frac{\partial u}{\partial n} = g \quad \text{στο } \Gamma_2, \end{aligned}$$

όπου Γ_1 είναι μη κενό και $\Gamma_1 \cup \Gamma_2 = \partial\Omega$ και $f \in L_2(\Omega), g \in L_2(\Gamma_2)$. Ακολουθώντας την περίπτωση των ομογενών συνθηκών Dirichlet, ορίζουμε

$$\tilde{H}_0^1(\Omega) = \{v \in H^1(\Omega) : v = 0 \ \text{ στο } \ \Gamma_1\},$$

και θεωρούμε το μεταβολικό πρόβλημα της εύρεσης $u \in \tilde{H}_0^1(\Omega)$, τέτοιο ώστε

$$a(u, v) = l(v), \quad \forall\, v \in \tilde{H}_0^1(\Omega), \tag{10.19}$$

όπου τώρα

$$a(v, w) = \int_\Omega \sum_{i=1}^{d} \frac{\partial v}{\partial x_i}(x) \frac{\partial w}{\partial w_i}(x)\, dx + \int_\Omega v(x) w(x)\, dx$$

και

$$l(v) = \int_\Omega f(x) v(x)\, dx + \int_{\Gamma_2} g(s) v(s)\, ds.$$

Εφαρμόζοντας το θεώρημα Lax–Milgram με $V = \tilde{H}_0^1(\Omega)$, μπορούμε με ανάλογα επιχειρήματα όπως και προηγουμένως να αποδείξουμε την ύπαρξη και τη μοναδικότητα της ασθενούς λύσης του (10.19).

10.2 Η μέθοδος Galerkin για ελλειπτικά προβλήματα

Όπως είδαμε και στο Κεφάλαιο 4, η μέθοδος Galerkin για την προσέγγιση της λύσης u του προβλήματος (10.3)–(10.4) έγκειται στην κατασκευή μιας οικογένειας $\{S_h\}_{0<h<1}$, υποχώρων του χώρου V, και την εύρεση $u_h \in S_h$, τέτοιων ώστε $a(u_h, \chi) = l(\chi)$ για κάθε $\chi \in S_h$. Εδώ, η διαγραμμική μορφή $a(\cdot, \cdot)$ και το γραμμικό συναρτησιακό $l(\cdot)$ υποθέτουμε ότι ικανοποιούν τις υποθέσεις του θεωρήματος Lax–Milgram.

Ως ένα παράδειγμα, θα θεωρήσουμε το ακόλουθο πρόβλημα με ομογενείς συνθήκες Dirichlet

$$-\Delta u(x) = f(x), \quad x \in \Omega, \quad \text{με } u(x) = 0, \quad x \in \partial\Omega, \tag{10.20}$$

όπου Ω είναι ένα φραγμένο, κυρτό, πολυγωνικό χωρίο στον $\mathbb{R}^2$. Θα υποθέσουμε ότι για το πολυγωνικό χωρίο Ω υπάρχει μια πεπερασμένη συλλογή ανοιχτών συνόλων $\{K_i\}$, τέτοια ώστε

$$K_i \cap K_j = \emptyset, \quad \text{αν} \quad i \neq j, \quad \bigcup_i \bar{K}_i = \bar{\Omega}. \tag{10.21}$$

Η συλλογή αυτή θα ονομάζεται *τριγωνισμός*, αν τα σύνολα K_i είναι τρίγωνα και είναι τέτοια ώστε, καμιά κορυφή ενός τριγώνου να μην βρίσκεται στο εσωτερικό της πλευράς ενός άλλου τριγώνου, δείτε το Σχήμα 10.1. Θα συμβολίζουμε με h_K τη διάμετρο του τριγώνου K και $h = \max_K h_K$. Επίσης, θα συμβολίζουμε με $\mathcal{T}_h$ έναν τριγωνισμό με μέγιστη πλευρά τριγώνων h.

Θα αναζητήσουμε τη συνάρτηση u_h στον χώρο S_h, τον χώρο των συνεχών συναρτήσεων στο $\bar{\Omega}$ οι οποίες μηδενίζονται στο $\partial\Omega$ και είναι πολυώνυμα βαθμού το πολύ ένα σε κάθε τρίγωνο K του τριγωνισμού $\mathcal{T}_h$. Η μέθοδος πεπερασμένων στοιχείων για το πρόβλημα (10.20) είναι

$$a(u_h, v_h) = l(v_h) \quad \text{για κάθε } \ v_h \in S_h, \tag{10.22}$$

όπου

$$a(u,v) = \int_\Omega \left(\frac{\partial u}{\partial x}\frac{\partial v}{\partial x} + \frac{\partial u}{\partial y}\frac{\partial v}{\partial y} \right) dxdy, \qquad l(v) = \int_\Omega fv\, dxdy.$$

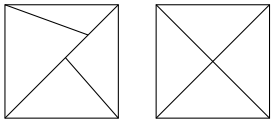
Αν N είναι το πλήθος των εσωτερικών *κόμβων*, P_i, $i = 1, \ldots, N$, δηλαδή των κορυφών των τριγώνων του $\mathcal{T}_h$ που δεν βρίσκονται στο $\partial\Omega$, τότε κάθε στοιχείο $v_h \in S_h$ ορίζεται μονοσήμαντα από τις τιμές του σε αυτά τα σημεία. Μια κατάλληλη βάση για τον χώρο S_h αποτελείται από τις συναρτήσεις $\phi_i \in S_h,\ i = 1, \ldots N$, οι οποίες ικανοποιούν

$$\phi_i(P_j) = \delta_{ij}, \quad 1 \le i, j \le N.$$

Πράγματι, ο φορέας της συνάρτησης ϕ_i αποτελείται από τα τρίγωνα του $\mathcal{T}_h$ τα οποία έχουν το P_i ως κορυφή. Οι ϕ_i είναι γραμμικά ανεξάρτητες, γιατί αν $\sum_{i=1}^N c_i\phi_i = 0$, τότε για $x = x_j$ έχουμε $c_j = 0, j = 1, \ldots, N$. Επιπλέον, αν $\psi \in S_h$, μπορούμε να γράψουμε

$$\psi(x) = \sum_{i=1}^{N} \psi(P_i)\phi_i(x), \quad x \in \bar{\Omega}, \tag{10.23}$$

γιατί οι συναρτήσεις στο αριστερό και δεξιό μέλος της (10.23) είναι στον S_h και οι τιμές τους συμφωνούν στους εσωτερικούς κόμβους P_i. Επομένως οι $\{\phi_i\}_{i=1}^N$



Σχήμα 10.1: Μια υποδιαίρεση ενός τετραγώνου (αριστερά) και ένας τριγωνισμός του (δεξιά).

αποτελούν βάση του S_h. Μπορεί να δειχθεί ότι ο χώρος S_h είναι υπόχωρος του $H_0^1(\Omega)$, δείτε για παράδειγμα (Δουγαλής, 2013). Σημειώνουμε ότι οι συναρτήσεις του S_h είναι, προφανώς, στον $L^2(\Omega)$ και μηδενίζονται κατά σημείο στο σύνορο του Ω.

Μπορεί να αποδειχθεί, δείτε για παράδειγμα (Brenner & Scott, 2008· Δουγαλής, 2013), ότι ο $S_h = \{\phi \in C(\bar{\Omega}) : \phi\big|_K \in \mathbb{P}_1,\ \forall\, K \in \mathcal{T}_h,\ \phi\big|_{\partial\Omega} = 0\}$, ικανοποιεί τη

$$\inf_{\chi \in S_h} \{\|v - \chi\| + h\|\nabla(v - \chi)\|\} \leq Ch^2\|v\|_2 \quad \text{για}\ \ v \in H^2(\Omega) \cap H_0^1(\Omega).$$

Υποθέτουμε ότι έχουμε μια διαμέριση του Ω με παράμετρο διαμέρισης h και μια οικογένεια υποχώρων πεπερασμένης διάστασης $\{S_h^r\}$ του $H_0^1(\Omega)$, με r ακέραιο, $r \geq 2$. Επίσης, υποθέτουμε ότι ο S_h^r έχει την ακόλουθη ιδιότητα προσέγγισης, για $2 \leq s \leq r$

$$\inf_{\chi \in S_h^r} \{\|v - \chi\| + h\|\nabla(v - \chi)\|\} \leq Ch^s\|v\|_s \quad \text{για}\ \ v \in H^s(\Omega) \cap H_0^1(\Omega), \tag{10.24}$$

όπου C είναι μια θετική σταθερά ανεξάρτητη των h και v.

Η ιδιότητα προσέγγισης (10.24) μας επιτρέπει τώρα να αποδείξουμε εκτιμήσεις για τα σφάλματα $u - u_h$ και $\nabla(u - u_h)$ με ανάλογο τρόπο όπως στο Θεώρημα 4.2, βλ. π.χ. (Larsson & Thomée, 2009· Johnson, 1987).

Θεώρημα 10.2. *Έστω $u \in H^r(\Omega) \cap H_0^1(\Omega)$ η λύση του προβλήματος* (10.20) *και $u_h \in S_h^r$ η λύση του μεταβολικού προβλήματος* (10.22)*. Τότε υπάρχει σταθερά C, ανεξάρτητη των u και h, τέτοια ώστε*

$$\|u - u_h\| + h\|\nabla(u - u_h)\| \leq Ch^r\|u\|_r. \tag{10.25}$$

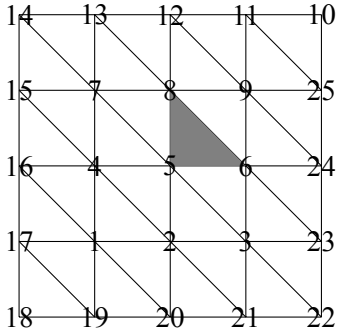
Όπως και στην περίπτωση του προβλήματος δύο σημείων, ένα σημαντικό πρόβλημα στην υλοποίηση της μεθόδου των πεπερασμένων στοιχείων είναι η διαχείριση των βαθμών ελευθερίας. Έστω τώρα $\Omega \subset \mathbb{R}^2$ ένα πολυγωνικό χωρίο και $\mathcal{T}_h$ ένας τριγωνισμός του Ω, της μορφής (10.21), ο οποίος αποτελείται από τρίγωνα K με $h = \max_K \operatorname{diam}(K)$. Συγκεκριμένα, έστω K ένα τρίγωνο του τριγωνισμού $\mathcal{T}_h$ με κορυφές $P_i^{(K)}$ για $i = 1, 2, 3$. Οι *τοπικές συναρτήσεις βάσης* $\phi_j^{(K)}$, $j = 1, 2, 3$ ανήκουν στον χώρο $\mathbb{P}_1(K)$ και ικανοποιούν $\phi_j^{(K)}(P_i^{(K)}) = \delta_{ij}$ για $1 \leq i, j \leq 3$. Επομένως, για τη λύση πεπερασμένων στοιχείων u_h έχουμε

$$u_h(x) = \sum_{j=1}^{3} \phi_j^{(K)}(x) u_h(P_j^{(K)}), \quad x \in K. \tag{10.26}$$

Οι πραγματικοί αριθμοί $\{u_h(P_j^{(K)})\}$, $j = 1, 2, 3$, είναι οι τοπικοί βαθμοί ελευθερίας οι οποίοι ορίζουν μονοσήμαντα την $u_h(x)$ στο τρίγωνο K. Αν τώρα ο τριγωνισμός $\mathcal{T}_h$ αποτελείται από N κόμβους P_i, $i = 1, \ldots, N$, (απαριθμούμε και αυτούς που βρίσκονται στο σύνορο), τότε οι κόμβοι $P_1^{(K)}, P_2^{(K)}, P_3^{(K)}$ του τριγώνου K αντιστοιχούν σε κάποιους κόμβους $P_{j_1}, P_{j_2}, P_{j_3}$ στην καθολική αρίθμηση των κόμβων του τριγωνισμού. Μας ενδιαφέρει να εκφράσουμε τους τοπικούς βαθμούς ελευθερίας $\{u_h(P_j^{(K)})\}, j = 1, 2, 3$ συναρτήσει των *καθολικών βαθμών ελευθερίας* $u_h(P_j)$, $j = 1, \ldots, N$. Ως παράδειγμα, ας πάρουμε το χωρίο Ω να είναι το τετράγωνο $[0, 1]^2$, το οποίο διαμερίζουμε σε 32 τρίγωνα και $N = 25$ κόμβους, όπως φαίνεται στο Σχήμα 10.2. Για χάρη ευκολίας, οι εσωτερικοί κόμβοι αριθμούνται *πριν* τους συνοριακούς κόμβους. Αν τ είναι ο αριθμός του σκιασμένου πεπερασμένου στοιχείου στο Σχήμα 10.2, τότε οι κόμβοι P_5, P_6, P_8, στην καθολική αρίθμηση των κόμβων, αντιστοιχούν στους κόμβους $P_1^{(\tau)}, P_2^{(\tau)}, P_3^{(\tau)}$ στην τοπική αρίθμηση των κόμβων του τριγώνου τ. Αν οι βαθμοί ελευθερίας $U_j = u_h(P_j)$, $j = 1, \ldots, N$ περιέχονται στο διάνυσμα $U = (U_1, U_2, \cdots, U_N)^T$ και $U^\tau = (U_1^\tau, U_2^\tau, U_3^\tau)^T$, όπου $U_j^\tau = u_h(P_j^{(\tau)}), j = 1, 2, 3$ τότε μπορούμε να γράψουμε $U^\tau = \mathcal{G}U$, όπου $\mathcal{G}$ είναι ένας $3 \times N$ πίνακας με στοιχεία 0 ή 1. Στο παράδειγμά μας, θα είναι της μορφής

$$\begin{pmatrix} U_1^\tau \\ U_2^\tau \\ U_3^\tau \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & \cdots \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & \cdots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & \cdots \end{pmatrix} \begin{pmatrix} U_1 \\ U_2 \\ U_3 \\ \vdots \\ U_N \end{pmatrix}.$$

Γενικότερα, αν τα τρίγωνα του τριγωνισμού $\mathcal{T}_h$ είναι $\{\tau\}$ με $\tau = 1, 2, \ldots, J$ και



Σχήμα 10.2: Αρίθμηση των κόμβων ενός τριγωνισμού του $[0, 1]^2$.

έστω $i = g(\tau, j)$, η συνάρτηση η οποία αντιστοιχεί τον τοπικό δείκτη j, $1 \le j \le 3$ των κόμβων P_j^τ με τον καθολικό δείκτη i, $1 \le i \le N$, των κόμβων P_i, τότε η g είναι μια συνάρτηση ορισμένη στο σύνολο $\{1, 2, \ldots, J\} \times \{1, 2, 3\}$, με τιμές στο σύνολο $\{1, 2, \ldots, N\}$, η οποία υπολογίζεται και αποθηκεύεται εύκολα.

Σημειώνουμε ακόμα, ότι αν $\Phi^\tau = [\phi_1^\tau, \phi_2^\tau, \phi_3^\tau]$ είναι το διάνυσμα των τοπικών συναρτήσεων βάσης, τότε μπορούμε να γράψουμε την αναπαράσταση (10.26) της u_h στο τρίγωνο τ ως $u_h(x) = \Phi(x)^\tau U^\tau$ και, φυσικά, $\nabla u_h(x) = D\Phi^\tau(x) \cdot U^\tau$, όπου $D\Phi^\tau$ συμβολίζει τον πίνακα

$$D\Phi^\tau = \begin{pmatrix} \frac{\partial \phi_1^\tau}{\partial x_1} & \frac{\partial \phi_2^\tau}{\partial x_1} & \frac{\partial \phi_3^\tau}{\partial x_1} \\ \frac{\partial \phi_1^\tau}{\partial x_2} & \frac{\partial \phi_2^\tau}{\partial x_2} & \frac{\partial \phi_3^\tau}{\partial x_2} \end{pmatrix}.$$

Μπορούμε τώρα να γράψουμε τη μέθοδο πεπερασμένων στοιχείων για το πρόβλημα (10.20) στη μορφή

$$\sum_{\tau \in \mathcal{T}_h} \int_\tau \nabla u_h \cdot \nabla v_h \, dx = \sum_{\tau \in \mathcal{T}_h} \int_\tau f(x) v_h \, dx, \quad \forall v_h \in S_h.$$

Αν τώρα θέσουμε $v_h = \Phi^\tau V^\tau$, τότε η προηγούμενη σχέση γίνεται

$$\sum_{\tau \in \mathcal{T}_h} \int_\tau (V^\tau (D\Phi^\tau)^T D\Phi^\tau U^\tau) \, dx = \sum_{\tau \in \mathcal{T}_h} \int_\tau (V^\tau)^T (\Phi^\tau)^T f(x) \, dx, \quad \forall v_h \in S_h.$$

Η τελευταία αυτή σχέση μας επιτρέπει να κατασκευάσουμε τόσο τον πίνακα ακαμψίας όσο και το δεξί μέλος αθροίζοντας τις συνεισφόρες από κάθε πεπερασμένο στοιχείο τ του τριγωνισμού $\mathcal{T}_h$.

10.3 Η εξίσωση της θερμότητας

Έστω Ω ένα φραγμένο χωρίο στον $\mathbb{R}^d$, $d = 2$ ή 3. Αναζητούμε μια πραγματική συνάρτηση $u = u(x, t)$, $x \in \bar{\Omega}$, $t \in [0, T]$, $T > 0$, τέτοια ώστε

$$\begin{cases} u_t(x,t) - \Delta u(x,t) = f(x,t), & x \in \Omega,\ t \in [0,T], \\ u(x,t) = 0, & x \in \partial\Omega,\ t \in [0,T], \\ u(x,0) = u^0(x), & x \in \bar{\Omega}, \end{cases} \tag{10.27}$$

όπου $f = f(x, t)$ είναι μια πραγματική συνάρτηση στο $\Omega \times [0, T]$ και η αρχική συνθήκη u^0 είναι, επίσης, μια πραγματική συνάρτηση στο $\bar{\Omega}$. Θα υποθέσουμε ότι το πρόβλημα αρχικών και συνοριακών τιμών (10.27) έχει μοναδική και αρκετά ομαλή λύση, έτσι ώστε να ισχύουν οι εκτιμήσεις που παρουσιάζονται στη συνέχεια.

Για τις ανάγκες της εφαρμογής της μεθόδου Galerkin ή πεπερασμένων στοιχείων θα υποθέσουμε ότι έχουμε μια διαμέριση του Ω με παράμετρο διαμέρισης h και μια οικογένεια υποχώρων πεπερασμένης διάστασης $\{S_h^r\}$ του $H_0^1(\Omega)$ με r ακέραιο $r \geq 2$, οι οποίοι ικανοποιούν την ιδιότητα (10.24).

Στο αντίστοιχο πρόβλημα της θερμότητας που μελετήσαμε στο Κεφάλαιο 6 θεωρήσαμε τον τελεστή της ελλειπτικής προβολής, βλ. (6.9), ο οποίος είχε σημαντικό ρόλο στη μελέτη του σφάλματος των μεθόδων που παρουσιάσαμε. Τώρα ο τελεστής της *ελλειπτικής προβολής* ορίζεται ως εξής: Για $v \in H^1(\Omega)$ η ελλειπτική προβολή του v στον χώρο S_h^r ορίζεται ως η γραμμική απεικόνιση $R_h : H^1(\Omega) \to S_h^r$, έτσι ώστε

$$(\nabla R_h v, \nabla \chi) = (\nabla v, \nabla \chi), \quad \forall \chi \in S_h^r. \tag{10.28}$$

Είναι προφανές ότι η ελλειπτική προβολή είναι καλά ορισμένη και ικανοποιεί τη σχέση $\|\nabla R_h v\| \leq \|\nabla v\|$, για $v \in H^1(\Omega)$. Οι ιδιότητες προσέγγισης της ελλειπτικής προβολής αποδεικνύονται στην παρακάτω πρόταση.

Λήμμα 10.2. *Έστω ότι $v \in H^s(\Omega) \cap H_0^1(\Omega)$ με $2 \leq s \leq r$. Τότε υπάρχει μια σταθερά C ανεξάρτητη του v και του h, τέτοια ώστε*

$$\|R_h v - v\| + h\|\nabla(R_h v - v)\| \leq C h^s \|v\|_s, \quad 2 \leq s \leq r. \tag{10.29}$$

Απόδειξη. Η απόδειξη γίνεται με ανάλογο τρόπο όπως στο Λήμμα 6.1, βλ. π.χ. (Larsson & Thomée, 2009· Johnson, 1987) □

10.3.1 Ημιδιακριτή προσέγγιση Galerkin

Για μια συνάρτηση $v \in H_0^1(\Omega)$ θεωρώντας το εσωτερικό γινόμενο και ως προς τα δύο μέλη της (10.27) και ολοκληρώνοντας κατά μέρη παίρνουμε τη μεταβολική μορφή του προβλήματος

$$(u_t(\cdot,t), v) + (\nabla u(\cdot,t), \nabla v) = (f(\cdot,t), v), \quad \forall v \in H_0^1(\Omega),\ t \in [0,T]. \tag{10.30}$$

Στη συνέχεια, οδηγούμαστε στο αντίστοιχο ημιδιακριτό πρόβλημα στον S_h^r: Ζητείται $u_h(t) = u_h(\cdot,t) \in S_h^r$, $t \in [0,T]$, τέτοια ώστε

$$\begin{cases} (u_{ht}(\cdot,t), \chi) + (\nabla u_h(\cdot,t), \nabla \chi) = (f(\cdot,t), \chi),\ \forall \chi \in S_h^r,\ t \in [0,T], \\ u_h(0) = u_h^0, \end{cases} \tag{10.31}$$

όπου u_h^0 είναι μια προσέγγιση του u^0 από τον S_h^r την οποία θα επιλέξουμε αργότερα. Αν $\{\phi_i\}_{i=1}^N$ με $N = \dim(S_h^r)$, είναι μια βάση του S_h^r και

$$u_h(x,t) = \sum_{j=1}^N \alpha_j(t)\phi_j(x), \quad u_h^0(x) = \sum_{j=1}^N \gamma_j \phi_j(x),$$

τότε, όπως και στην Παράγραφο 6.1, το (10.31) είναι ισοδύναμο με ένα γραμμικό σύστημα συνήθων διαφορικών εξισώσεων. Έτσι, παίρνουμε το ανάλογο πρόβλημα της (6.6)

$$\begin{cases} \mathcal{M}\alpha'(t) + \mathcal{S}\alpha(t) = F(t), \quad t \in [0,T], \\ \alpha(0) = \Gamma, \end{cases} \tag{10.32}$$

όπου οι πίνακες μάζας $\mathcal{M}$ και ακαμψίας $\mathcal{S}$ είναι οι $N \times N$ πίνακες με στοιχεία $\mathcal{M}_{ij} = (\phi_j, \phi_i)$ και $\mathcal{S}_{ij} = (\nabla\phi_j, \nabla\phi_i)$, $i,j = 1,\ldots,N$, αντίστοιχα, $\alpha(t) = (\alpha_1(t),\ldots,\alpha_N(t))^T, \Gamma = (\gamma_1,\ldots,\gamma_N)^T, F(t) = (F_1(t),\ldots,F_N(t))^T$, με $F_i(t) = (f(\cdot,t),\phi_i)$, $i = 1,\ldots,N$. Τόσο ο πίνακας μάζας $\mathcal{M}$ όσο και ο πίνακας ακαμψίας $\mathcal{S}$ είναι συμμετρικοί και θετικά ορισμένοι πίνακες και ειδικότερα ο $\mathcal{M}$ αντιστρέφεται, το οποίο οδηγεί στη μονοσήμαντη λύση του (10.32).

Με ανάλογο τρόπο όπως και στην Παράγραφο 6.1 μπορούμε να αποδείξουμε το ακόλουθο θεώρημα, για την απόδειξη του οποίου παραπέμπουμε στα (Larsson & Thomée, 2009· Johnson, 1987).

Θεώρημα 10.3. *Έστω u και u_h οι λύσεις των προβλημάτων* (10.27) *και* (10.31)*, αντίστοιχα. Τότε,*

$$\|u_h(t) - u(t)\| \le \|u_h^0 - u^0\| + Ch^r \left(\|u^0\|_r + \int_0^t \|u_t\|_r \, ds \right), \quad t \ge 0, \tag{10.33}$$

για κάποια σταθερά C ανεξάρτητη του h.

Παρατήρηση 10.2. Είναι δυνατόν να επιλέξουμε την αρχική τιμή u_h^0, έτσι ώστε $\|u_h^0 - u^0\| \le Ch^r$, με σταθερά C ανεξάρτητη του h, διατηρώντας με αυτό τον τρόπο την τάξη r του σφάλματος στη σχέση (10.33). Μπορούμε, για παράδειγμα, να πάρουμε $u_h^0 = R_h u^0$ ή να επιλέξουμε ως u_h^0 την L_2 προβολή του u^0 στον S_h^r. Και στις δύο περιπτώσεις $\|u_h^0 - u^0\| \le Ch^r \|u^0\|_r$.

10.3.2 Πλήρως διακριτές προσεγγίσεις Galerkin

Για τη λύση του γραμμικού συστήματος συνήθων διαφορικών εξισώσεων (10.31) είναι απαραίτητο να διακριτοποιήσουμε τη χρονική μεταβλητή t, όπου μπορούμε να χρησιμοποιήσουμε π.χ. την πεπλεγμένη μέθοδο του Euler ή τη μέθοδο Crank–Nicolson.

Για έναν φυσικό αριθμό $M \ge 1$ θέτουμε $k = T/M$ και $t^n = nk, n = 0, 1, \ldots, M$. Η πεπλεγμένη μέθοδος του Euler για τη διακριτοποίηση στον χρόνο του ημιδιακριτού σχήματος (10.31) ορίζεται ως εξής: για $n = 1, 2, \ldots, M$, αναζητούμε προσεγγίσεις $U^n \in S_h^r$ των $u_h(t^n)$ που ορίζονται από τις σχέσεις

$$\left(\frac{U^n - U^{n-1}}{k}, \chi \right) + (\nabla U^n, \nabla\chi) = (f^n, \chi), \quad \forall \chi \in S_h^r, \quad U^0 = u_h^0, \tag{10.34}$$

με $f^n = f(\cdot, t^n)$. Αν $\{\phi_j\}_{j=1}^N$ είναι μια βάση του S_h^r και $U^n = \sum_{j=1}^N \alpha_j^n \phi_j$, με $\alpha^n = (\alpha_1^n, \ldots, \alpha_N^n)^T \in \mathbb{R}^N$, για $n \geq 0$, τότε οι σχέσεις (10.34) είναι ισοδύναμες με το γραμμικό σύστημα εξισωσεων

$$(\mathcal{M} + k\mathcal{S})\alpha^n = \mathcal{M}\alpha^{n-1} + kF^{n-1}, \quad n \geq 1, \tag{10.35}$$

όπου $F^n = (F_1^n, \ldots, F_N^n)^T$, με $F_i^n = (f^n, \phi_i)$, $1 \leq i \leq N$. Επειδή ο $\mathcal{M} + k\mathcal{S}$ είναι συμμετρικός και θετικά ορισμένος, είναι αντιστρέψιμος και, άρα, τα α^n, $n \geq 1$ προσδιορίζονται μοναδικά και κατ' επέκταση οι U^n. Για την επίλυση του (10.35) μπορούν να εφαρμοσθούν διάφορες μέθοδοι, συνήθως χρησιμοποιείται είτε η ανάλυση Cholesky είτε η (προρυθμισμένη) μέθοδος των συζυγών κλίσεων. Με ανάλογο τρόπο όπως και στην Παράγραφο 6.2 μπορούμε να αποδείξουμε το ακόλουθο θεώρημα, για την απόδειξη του οποίου παραπέμπουμε στα (Larsson & Thomée, 2009· Johnson, 1987).

Θεώρημα 10.4. *Έστω U^n, $u(t) = u(\cdot, t)$ οι λύσεις των* (10.34)*,* (10.27)*, αντίστοιχα. Τότε υπάρχει θετική σταθερά C, ανεξάρτητη των h και k, τέτοια ώστε*

$$\begin{aligned}\max_{0\leq n\leq M} \|U^n - u(t^n)\| \leq \|u_h^0 - u^0\| + Ch^r \left[\|u^0\|_r + \int_0^T \|u_t\|_r \, ds\right] \\ + Ck \int_0^T \|u_{tt}\| \, ds.\end{aligned} \tag{10.36}$$

Στη συνέχεια, μπορούμε να θεωρήσουμε και τη μέθοδο των Crank–Nicolson για τη διακριτοποίηση του ημιδιακριτού σχήματος (10.31). Έτσι, αναζητούμε προσεγγίσεις $U^n \in S_h^r$ των $u_h(t^n)$, $n = 0, 1, \ldots, M$, τέτοιες ώστε

$$\begin{aligned}\left(\frac{U^n - U^{n-1}}{k}, \chi\right) + (\nabla U^{n-1/2}, \nabla\chi) = (f^{n-1/2}, \chi), \quad \forall \chi \in S_h^r, \\ U^0 = u_h^0,\end{aligned} \tag{10.37}$$

όπου $U^{n-1/2} = (U^n + U^{n-1})/2$ και $f^{n-1/2} = f(\cdot, t^n - \frac{k}{2})$. Με ανάλογο τρόπο, όπως προηγουμένως για την πεπλεγμένη μέθοδο του Euler μπορούμε να θεωρήσουμε το ανάλογο γραμμικό σύστημα με το (10.35). Έτσι, οι (10.37) είναι ισοδύναμες με το γραμμικό σύστημα

$$\left(\mathcal{M} + \frac{k}{2}\mathcal{S}\right)\alpha^n = \left(\mathcal{M} - \frac{k}{2}\mathcal{S}\right)\alpha^{n-1} + kF^{n-1/2}, \quad n \geq 1, \tag{10.38}$$

όπου $F^{n-1/2} = (F_1^{n-1/2}, \ldots, F_N^{n-1/2})^T$, με $F_i^{n-1/2} = (f^{n-1/2}, \phi_i)$, $1 \leq i \leq N$. Εύκολα μπορούμε να δούμε ότι ο πίνακας $\mathcal{M} + \frac{k}{2}\mathcal{S}$ είναι συμμετρικός και θετικά

ορισμένος, οπότε το σύστημα εξισώσεων (10.38) έχει μοναδική λύση. Με ανάλογο τρόπο, όπως και στην Παράγραφο 6.3, μπορούμε να αποδείξουμε το ακόλουθο θεώρημα, για την απόδειξη του οποίου παραπέμπουμε στα (Larsson & Thomée, 2009· Johnson, 1987).

Θεώρημα 10.5. *Έστω $U^n, u(t) = u(\cdot, t^n)$ οι λύσεις των προβλημάτων* (10.27)*,* (10.37)*, αντίστοιχα. Τότε υπάρχει θετική σταθερά C, ανεξάρτητη των h και k, τέτοια ώστε*

$$\max_{0\le n\le M} \|U^n - u(t^n)\| \le \|u_h^0 - u^0\| + Ch^r \left[\|u^0\|_r + \int_0^T \|u_t(s)\|_r \, ds\right] + Ck^2 \int_0^T (\|u_{ttt}(s)\|_r + \|\Delta u_{tt}(s)\|_r) \, ds. \tag{10.39}$$

Επίσης, μπορούμε να θεωρήσουμε και την άμεση μέθοδο του Euler για τη διακριτοποίηση του ημιδιακριτού σχήματος (10.31). Έτσι, αναζητούμε προσεγγίσεις $U^n \in S_h^r$ των τιμών $u(t^n) = u(\cdot, t^n)$ για $n \ge 0$ οι οποίες ικανοποιούν

$$\left(\frac{U^n - U^{n-1}}{k}, \chi\right) + (\nabla U^{n-1}, \nabla\chi) = (f^{n-1}, \chi), \ \forall \chi \in S_h^r, \qquad U^0 = u_h^0. \tag{10.40}$$

Χρησιμοποιώντας τον συμβολισμό προηγούμενων παραγράφων, βλέπουμε ότι ο προσδιορισμός του U^n για $n \ge 1$, δεδομένου του U^{n-1}, απαιτεί τη λύση του γραμμικού συστήματος

$$\mathcal{M}\alpha^n = (\mathcal{M} - k\mathcal{S})\alpha^{n-1} = kF^{n-1}, \quad n \ge 1. \tag{10.41}$$

Το σύστημα αυτό έχει μοναδική λύση, μια και ο πίνακας μάζας $\mathcal{M}$ είναι συμμετρικός και θετικά ορισμένος.

Για να δείξουμε την ευστάθεια και τη σύγκλιση της μεθόδου είναι απαραίτητη η υπόθεση ότι οι συναρτήσεις του S_h^r ικανοποιούν μια *αντίστροφη ανισότητα*, ανάλογη της (6.41), έτσι υποθέτουμε ότι

$$\|\nabla\chi\| \le C_* h^{-1} \|\chi\|, \quad \chi \in S_h^r, \tag{10.42}$$

για κάποια σταθερά C_* ανεξάρτητη των χ και h.

Αν $\Omega \subset \mathbb{R}^2$ είναι ένα πολυγωνικό χωρίο και υποθέσουμε ότι ο τριγωνισμός $\mathcal{T}_h$ του Ω είναι ημιομοιόμορφος, δηλαδή ότι υπάρχει σταθερά ν ανεξάρτητη του τριγωνισμού, τέτοια ώστε

$$\frac{h}{h_K} \le \nu, \quad \forall\, K \in \mathcal{T}_h,$$

όπου $h = \max_K h_K$, και θεωρήσουμε τον χώρο S_h^2 των κατά τμήματα γραμμικών πολυωνύμων στον $\mathcal{T}_h$, τότε ο S_h^2 ικανοποιεί την (10.42). Ο ενδιαφερόμενος αναγνώστης μπορεί να βρει την απόδειξη της αντίστροφης ανισότητας στο βιβλίου του Ciarlet (Ciarlet, 2002).

Έστω τώρα ότι

$$\frac{k}{h^2} \leq \frac{2}{C_*^2}, \tag{10.43}$$

τότε για το σφάλμα της άμεσης μεθόδου του Euler έχουμε το ακόλουθο αποτέλεσμα:

Θεώρημα 10.6. *Έστω u και U^n οι λύσεις των προβλημάτων* (10.27) *και* (10.40)*, αντίστοιχα. Υποθέτουμε ότι ισχύουν οι σχέσεις* (10.42) *και* (10.43)*. Τότε υπάρχει θετική σταθερά C, ανεξάρτηση των h και k τέτοια ώστε,*

$$\max_{0\leq n\leq M} \|U^n - u(t^n)\| \leq \|u_h^0 - u^0\| + Ch^r \left[\|u^0\|_r + \int_0^T \|u_t(s)\|_r \, ds\right] + Ck \int_0^T \|u_{tt}(s)\|_r \, ds.$$

10.4 Άλλες μέθοδοι και προβλήματα

Μέθοδοι πεπερασμένων στοιχείων για γενικότερα ελλειπτικά προβλήματα με συνορικακές συνθήκες τύπου Dirichlet ή και με άλλες συνοριακές συνθήκες όπως π.χ. τύπου Neumann, αναλύονται με παρόμοιο τρόπο όπως με αυτά που παρουσιάσαμε σε αυτό το κεφάλαιο. Για τη μελέτη αυτών των αριθμητικών μεθόδων, καθώς και για πιο λεπτομερή παρουσίαση των αποτελεσμάτων αυτού του κεφαλαίου, παραπέμπουμε τον αναγνώστη στα (Ciarlet, 2002· Johnson, 1987· Knabner & Angermann, 2003· Larsson & Thomée, 2009· Morton & Mayers, 2005).

10.5 Ασκήσεις

10.1. Δείξτε ότι η συνάρτηση $f(x) = x^{-\beta}$, $x \in (0,1)$ ανήκει στον $L_2(0,1)$, αν $\beta < 1/2$.

10.2. Δείξτε ότι η συνάρτηση $f(x) = |x|^{-\beta}$, $x \in \Omega$, με $\Omega = (0,1)\times(0,1)$ ανήκει στον $L_2(\Omega)$, αν $\beta < 1$.

10.3. Δείξτε ότι η συνάρτηση $f(x) = |x|^{\beta}$, $x \in \Omega$, με $\Omega = (0,1)\times(0,1)$ ανήκει στον $H^1(\Omega)$, αν $\beta > 0$.

10.4. Δείξτε ότι το πρόβλημα (10.19) έχει μοναδική λύση στον $\tilde{H}_0^1(\Omega)$ και διατυπώστε μια μέθοδο πεπερασμένων στοιχείων για αυτό το πρόβλημα.

10.5. Αποδείξτε το Θεώρημα 10.2.

10.6. Αποδείξτε το Λήμμα 10.2.

10.7. Αποδείξτε το Θεώρημα 10.3.

Βιβλιογραφία

Δουγαλής, Β. (2013). *Finite element methods for the numerical solution of partial differential equations*. Αθήνα. (Πανεπιστημιακές Σημειώσεις).

Adams, R. A. (1975). *Sobolev spaces*. Academic Press, New York-London.

Brenner, S. C., & Scott, L. R. (2008). *The mathematical theory of finite element methods* (Third ed., Vol. 15). Springer, New York.

Brezis, H. (1997). *Συναρτησιακή Ανάλυση: Θεωρία και Εφαρμογές*. Πανεπιστημιακές Εκδόσεις Ε.Μ.Π., Αθήνα. (Ελληνική μετάφραση από τους Δ. Κραββαρίτη και Ι. Χρυσοβέργη).

Ciarlet, P. G. (2002). *The finite element method for elliptic problems* (Vol. 40). Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA.

Gilbarg, D., & Trudinger, N. S. (1983). *Elliptic partial differential equations of second order* (Second ed., Vol. 224). Springer-Verlag, Berlin.

Johnson, C. (1987). *Numerical solution of partial differential equations by the finite element method*. Cambridge University Press, Cambridge.

Knabner, P., & Angermann, L. (2003). *Numerical methods for elliptic and parabolic partial differential equations* (Vol. 44). Springer-Verlag, New York.

Larsson, S., & Thomée, V. (2009). *Partial differential equations with numerical methods* (Vol. 45). Springer-Verlag, Berlin.

Morton, K. W., & Mayers, D. F. (2005). *Numerical solution of partial differential equations* (Second ed.). Cambridge University Press, Cambridge.

Royden, H. L. (1988). *Real analysis* (Third ed.). Macmillan Publishing Company, New York.

Λεξικό Αγγλικών Όρων

άμεση μέθοδος	*explicit method.*
αρχικές συνθήκες	*initial value conditions.*
διαφορά προς τα εμπρός	*forward difference.*
διαφορά προς τα πίσω	*backward difference.*
ελλειπτική εξίσωση	*elliptic equation.*
ελλειπτική ομαλότητα	*elliptic regularity.*
εξίσωση θερμότητας	*heat equation.*
εξίσωση κύματος	*wave equation.*
εξίσωση μεταφοράς	*transport equation.*
ευστάθεια	*stability.*
ημιδιακριτό πρόβλημα	*semidiscrete problem.*
κεντρική διαφορά	*central difference.*
μέθοδος πεπερασμένων όγκων	*finite volume method.*
μερική διαφορική εξίσωση	*partial differential equation.*
μεταβολική μορφή	*variational form.*
ομογενείς συνοριακές συνθήκες	*homogenous boundary conditions.*
πίνακας ακαμψίας	*stiffness matrix.*
πίνακας μάζας	*mass matrix.*
πίνακας με αυστηρά κυριαρχική διαγώνιο	*strictly diagonally dominant matrix.*
μέθοδος πεπερασμένων στοιχείων	*finite element method.*
πεπερασμένη διαφορά	*finite difference.*
πεπλεγμένη μέθοδος	*implicit method.*
συνέπεια	*consistency.*
συνήθης διαφορική εξίσωση	*ordinary differential equation.*
συναρμολόγηση	*assembly.*

συνοριακό πρόβλημα δύο σημείων	*two point boundary value problem.*
τάξη ακρίβειας	*order of accuracy.*
τάξη διαφορικής εξίσωσης	*order of differential equation.*
τοπικό σφάλμα διακριτοποίησης	*local discretization error.*
τριγωνισμός	*triangulation.*
τριδιαγώνιος πίνακας	*tridiagonal matrix.*

Ευρετήριο

Α

ανάπτυγμα σε σειρά Fourier, 16
ανισότητα
  Cauchy–Schwarz, 42, 60
  Poincaré–Friedrichs, 60
  αντίστροφη, 114
αρχή της υπέρθεσης, 4
ασθενής
  μερική παράγωγος, 174
  μορφή προβλήματος, 175

Δ

Μ.Δ.Ε., 1
  γραμμική, 2, 3
  μη γραμμική, 2, 3
  μη ομογενής, 2
  ομογενής, 2
  τάξη, 2
Σ.Δ.Ε., 1
  τάξη, 2
διαφορική εξίσωση, 1
  λύση, 1

Ε

ελλειπτική ομαλότητα, 62
εξίσωση
  ελλειπτική, 13, 163, 171
  θερμότητας, 7, 75, 103, 182
  κύματος, 149
  κύμματος, 12
  μεταφοράς, 10, 121
ευστάθεια, 39
  von Neumann, 81, 87, 93, 130, 135, 138, 157
  υπό συνθήκες, 52, 80

Θ

Θεώρημα Lax–Milgram, 176

Κ

κομβικοί δείκτες
  καθολικοί, 69
  τοπικοί, 69

Λ

λύση
  αναπαράσταση d'Alembert, 13
  ασθενής, 61, 175
  κλασική, 172
  χαρακτηριστικές, 10

Μ

μέθοδος
  Courant, Friedrichs και Lewy, 149
  Crank–Nicolson, 92
  downwind, 124, 125
  Galerkin, 62, 63
  Lax–Wendroff, 133, 137
  upwind, 122
  άμεση Euler, 78
  πεπερασμένων στοιχείων, 62
  πεπλεγμένη Euler, 87
μεταβολική
  μορφή προβλήματος, 175

Ν

νόρμα, 39

Π

πίνακας

ακαμψίας, 105
ακαμψίας (τοπικός), 71
ανάλυση LU, 36
αυστηρά κυριαρχική διαγώνιος, 35
μάζας, 105
μάζας (τοπικός), 71
συναρμολόγηση, 70
τριδιαγώνιος, 36
πεπερασμένες διαφορές, 23
διαφορά προς τα εμπρός, 21
διαφορά προς τα πίσω, 21
κεντρική διαφορά για τη 2η παράγωγο, 28
κεντρική διαφορά για την 1η παράγωγο, 22
πεπερασμένο στοιχείο, 69
πρόβλημα δύο σημείων
μεταβολική μορφή, 61

Σ

κυβικές splines, 72
σειρές Fourier, 15
συνέπεια, 78, 85
συναρτήσεις βάσης, 62, 63, 72
τοπικές, 70, 180
συναρτήσεις δοκιμής, 61, 62
συνεπής μέθοδος, 35, 78
συνθήκες
αρχικές, 4
αρχικές και συνοριακές, 4, 7, 12
συνοριακές, 4
συνθήκη
ομοιόμορφης ελλειπτικότητας, 172
συνθήκη CFL, 129
συνοριακές συνθήκες
Dirichlet, 5, 53
Neumann, 5
Robin, 5
ομογενείς Dirchlet, 5–7, 12, 33, 49
ομογενείς Neumann, 5, 7, 8, 45
περιοδικές, 5

Τ

τάξη ακρίβειας
πειραματική, 26
προσέγγισης, 25
τοπικού σφάλματος διακριτοποίησης, 34, 78
τοπικό σφάλμα διακριτοποίησης, 34, 78
τριγωνισμός, 178, 180

Χ

χωρίο υπολογιστικής εξάρτησης, 128
χώροι Lebesgue, 174
χώροι Sobolev, 173