ΣΥΝΤΟΜΗ ΕΙΣΑΓΩΓΗ ΣΤΗΝ ΑΡΙΘΜΗΤΙΚΗ ΑΝΑΛΥΣΗ ΚΑΙ ΣΤΟ MATLAB

ΠΡΟΛΟΓΟΣ

Το βιβλίο αυτό διαδέχεται δύο προηγούμενα: την «Αριθμητική Ανάλυση» που κυκλοφόρησε το 1996, και το αντίστοιχο βιβλίο «Αριθμητική Ανάλυσις» που κυκλοφόρησε το 1977. Όπως και στα παλαιότερα εκείνα συγγράμματα στόχος του παρόντος είναι να δώσει την Αριθμητική Ανάλυση από την οπτική γωνία του χρηστή και όχι του θεωρητικού μαθηματικού, χωρίς να θυσιάσει την ακρίβεια διατύπωσης των αποτελεσμάτων. Όμως δεν επιδιώκει πλέον να παρουσιάσει τα βασικά θέματά της διεξοδικά, αλλά μάλλον συνοπτικά, έτσι όπως παρουσιάζονται σε ένα εισαγωγικό μάθημα Αριθμητικής Ανάλυσης διαρκείας ενός εξαμήνου. Όποιος ενδιαφέρεται για μίαν εκτεταμένη παρουσίαση των βασικών θεωρημάτων μπορεί να την αναζητήσει στο βιβλίο του 1996 ή σε άλλα συγγράμματα που αναφέρονται εν μέρει στην βιβλιογραφία.[1]

Η Αριθμητική Ανάλυση είναι σήμερα βασικό εργαλείο για την προώθηση τόσο της ερευνάς όσο και των εφαρμογών στην Φυσική και τους διάφορους κλάδους της Τεχνολογίας. Για φανεί η σημασία των μεθόδων που παρουσιάζονται για την διερεύνηση φυσικών και τεχνικών προβλημάτων, προτάσσονται σε πολλά κεφαλαία τεχνικά και φυσικά προβλήματα που απαιτούν την χρήση μεθόδων Αριθμητικής Ανάλυσης, γιατί δεν μπορούν να επιλυθούν με τις μεθόδους της κλασσικής Ανάλυσης ή της Άλγεβρας.

Σε ορισμένα σημεία παρεμβάλλονται επίσης παρατηρήσεις που αναφέρονται σε πιο ειδικού ενδιαφέροντος θέματα (π.χ. για χαοτικά φαινόμενα και την σχέση τους με αριθμητικές μεθόδους) που μπορούν να παραλειφθούν από τον μη άμεσα ενδιαφερόμενο. Τέτοιες παρατηρήσεις σημειώνονται με αστερίσκο.

Οι προγενέστερες εκδόσεις περιείχαν προγράμματα σε γλώσσα BASIC. Στην παρούσα έκδοση τα προγράμματα αυτά παρελείφθησαν δεδομένου ότι έχουν πλέον ευρεία διάδοση προγραμματιστικά περιβάλλοντα όπως το MATLAB και το Mathematica, που επιτρέπουν την εφαρμογή πολλών από τις μεθόδους της Αριθμητικής Ανάλυσης (αλλά και της κλασσικής Ανάλυσης) με χρήση μίας μόνον εντολής και αποτελούν έτσι γλώσσες συνοπτικού προγραμματισμού για κάθε είδος μαθηματικών υπολογισμών. Έτσι προτιμήθηκε να παρατεθεί στο τέλος του βιβλίου μίαν σύντομη εισαγωγή στο MATLAB, καθώς και ορισμένα προγράμματα (scripts ή m-files) με τα οποία έχουν δημιουργηθεί οι γραφικές παραστάσεις που παρατίθενται.

Αθηνά, Μάρτιος 2004

Γιάννης Κιουστελίδης

---

[1] Όπως θα διαπιστώσει ο αναγνώστης, μεγάλα τμήματα του κειμένου δεν έχουν τόνους. Αυτό οφείλεται στο ότι τα τμήματα αυτά έχουν παρθεί αυτούσια από την προγενέστερη έκδοση του 1996, που ήταν γραμμένη με άλλον επεξεργαστή κειμένου, και κατά την μετατροπή δεν έγινε δυνατόν να διατηρηθούν οι τόνοι (καθώς και πολλοί τύποι). Η συμπλήρωσή τους εκ των υστέρων αποδείχθηκε, δυστυχώς, πολύ δύσκολη. Έτσι το βιβλίο τούτο είναι κατά κάποιον τρόπο θύμα της ραγδαίας εξέλιξης στην πληροφορική.

ΕΙΣΑΓΩΓΗ

Η Αριθμητική Ανάλυση είναι το ένα από τα δύο μέρη της Ανάλυσης: το υπολογιστικό η κατασκευαστικό. Αυτό που συνήθως λέμε "Ανάλυση", ή στοιχειωδέστερα "Απειροστικό Λογισμό", είναι στην πραγματικότητα κυρίως το ποιοτικό της μέρος, που ασχολείται με το αν υπάρχουν λύσεις των προβλημάτων της Ανάλυσης και με το τι ιδιότητες έχουν αυτές οι λύσεις. Αυτό το μέρος της Ανάλυσης περιορίζεται να δώσει λύσεις των βασικών προβλημάτων μόνο αν αυτές μπορούν να εκφρασθούν με την βοήθεια στοιχειωδών συναρτήσεων (x^p , ln x, e^x , sin x, cos x) η με την βοήθεια δυναμοσειρών. Αυτό σημαίνει ότι είναι σπανίως σε θέση να προσδιορίσει ακριβώς την λύση των προβλημάτων που μελετά.

Η λύση μιας μη γραμμικής εξίσωσης, το αποτέλεσμα μιας ολοκλήρωσης ή η λύση ενός προβλήματος αρχικών η συνοριακών τιμών για κοινές διαφορικές εξισώσεις εν γενεί δεν μπορούν να εκφρασθούν σε κλειστή μορφή με την βοήθεια στοιχειωδών συναρτήσεων, ενώ οι λύσεις με μορφή δυναμοσειρών σπάνια συγκλίνουν αρκετά γρήγορα ώστε να μπορούν να χρησιμεύσουν για πρακτικούς υπολογισμούς. Έτσι οι λύσεις αυτές παραμένουν για την κλασσική "Ανάλυση" μόνο αντικείμενα θεωρητικής μελέτης των γενικών τους χαρακτηριστικών (βαθμός ομαλότητας, μονοτονία, περιοδικότητα κ.ο.κ.) που δεν μπορεί να τα προσδιορίσει ποσοτικά.

Στόχος και βασικό αντικείμενο μελέτης της Αριθμητικής Ανάλυσης είναι ο ποσοτικός προσδιορισμός αυτών των λύσεων με όσο το δυνατό πιο αποτελεσματικό τρόπο και όσο το δυνατό μεγαλύτερη ακρίβεια. Ακόμα πριν πενήντα χρόνια η μελέτη φυσικών και τεχνικών προβλημάτων βασιζόταν σε αναλυτικές λύσεις, όπου υπάρχουν τέτοιες. Έτσι, συχνά γινόταν προσπάθεια τροποποίησης των προβλημάτων, π.χ. γραμμικοποίησης τους, Έτσι ώστε να δέχονται αναλυτικές λύσεις. Στις ημέρες μας έγινε όμως κατανοητό ότι πολλές φορές ακριβώς η πιο ενδιαφέρουσες φάσεις πολλών φαινομένων είναι έντονα μη γραμμικές και δεν μπορούν να προσεγγισθούν ικανοποιητικά με γραμμικά μοντέλα και αναλυτικές προσεγγίσεις. Έτσι, η Αριθμητική Ανάλυση έχει πάψει να είναι βοήθημα και έχει γίνει βασικό εργαλείο της μελέτης φυσικών και τεχνικών προβλημάτων. Αυτό δεν σημαίνει όμως ότι μπορεί να υποκαταστήσει την κλασσική Ανάλυση. Η επιλογή μιας γρήγορης και ακριβούς μεθόδου για τον ποσοτικό προσδιορισμό μιας λύσης συχνά απαιτεί κατ' αρχήν την ποιοτική μελέτη της λύσης αυτής. Αν π.χ. υπάρχουν περιοχές ταχείας μεταβολής της λύσης πρέπει ενδεχόμενα να χρησιμοποιηθούν διαφορετικές μέθοδοι. Ακόμα βασικότεροι όμως είναι η ποιοτική ανάλυση της λύσης για την ερμηνεία των αριθμητικών αποτελεσμάτων.

Οι μέθοδοι (αλγόριθμοι) που μας προσφέρει η Αριθμητική Ανάλυση παράγουν μόνο σύνολα αριθμών (π.χ., ακολουθίες η αριθμητικούς πίνακες). Το τι σημαίνουν τα σύνολα αυτά για το πρόβλημα που μελετούμε, το πως σχετίζονται με την λύση του, πρέπει να το προσδιορίσουμε εμείς και για τον σκοπό αυτό χρησιμοποιούμε τόσο την ποιοτική ανάλυση της λύσης του αρχικού προβλήματος αλλά και του χρησιμοποιούμενου αριθμητικού αλγορίθμου (δηλαδή, της υπολογιστικής διαδικασίας που χρησιμοποιούμε).

Βασικό εργαλείο της Αριθμητικής Ανάλυσης είναι ο Ηλεκτρονικός

Υπολογιστής (Η/Υ) γι' αυτό η μεθοδολογία της, και η νοοτροπία της, ο τρόπος που αντιμετωπίζει τα διάφορα προβλήματα και αξιολογεί τους διάφορους τρόπους υπολογισμού της λύσης, είναι βαθιά επηρεασμένος από τις δυνατότητες και τον τρόπο λειτουργίας του Η/Υ. Έτσι, λύσεις με μορφή δυναμοσειρών, σπάνια θεωρούνται ενδεδειγμένες για τον ποσοτικό υπολογισμό των λύσεων γιατί απαιτούν συνήθως πολύ μεγάλο όγκο πράξεων για να επιτευχθεί μια μέση ακρίβεια αποτελέσματος. Αντί γι' αυτές χρησιμοποιούνται από την Αριθμητική Ανάλυση εκτενέστατα οι τμηματικά πολυωνυμικές προσεγγίσεις μιας λύσης, που είναι πολύ ακριβέστερες, αλλά καταγράφονται συνήθως μόνο στην μνήμη ενός Η/Υ, γιατί περιλαμβάνουν πολλές παραμέτρους. Παρόμοια, η λύση ενός μεγάλου συστήματος γραμμικών εξισώσεων δεν γίνεται ποτέ με χρήση οριζουσών, γιατί απαιτούν υπέρμετρο όγκο πράξεων, αλλά συνήθως με την πολύ οικονομικότερη μέθοδο απαλοιφής του Gauss, ή με επαναληπτικές μεθόδους.

Για να μπορέσει μια οποιαδήποτε υπολογιστική μέθοδος να χρησιμοποιηθεί στον Η/Υ πρέπει πρώτα να πάρει μορφή αλγόριθμου, που δίνεται στον υπολογιστή σαν "πρόγραμμα" γραμμένο σε μιαν από τις γλώσσες προγραμματισμού. Βασικό χαρακτηριστικό ενός αλγόριθμου είναι ότι αποτελεί εντελώς αυτοματοποιημένη διαδικασία που δεν ζητάει ούτε επιτρέπει διαισθητικές επιλογές για την πορεία λύσης ενός προβλήματος. Όλα τα στοιχεία μιας διαδικασίας επίλυσης προβλήματος πρέπει να είναι πλήρως καθορισμένα, δηλαδή να αντιστοιχούν σε απλές διαδικασίες άμεσα εκτελέσιμες από τον Η/Υ. Έτσι, αλγόριθμος ονομάζεται στα πλαίσια της Αριθμητικής Ανάλυσης μια πεπερασμένη ακολουθία αριθμητικών πράξεων (προσθέσεων, αφαιρέσεων, πολλαπλασιασμών ή διαιρέσεων) και λογικών συνδέσεων (συζεύξεων, $A \wedge B$, διαζεύξεων, $A \vee B$, συναγωγών, $A \rightarrow B$, ή αρνήσεων, $\sim A$) που εφαρμόζονται στα αρχικά δεδομένα ενός προβλήματος και τα ενδιάμεσα αποτελέσματα για να παραγάγουν ένα τελικό αριθμητικό αποτέλεσμα. Το αποτέλεσμα αυτό ερμηνεύεται από εμάς, με βάση την μαθηματική ανάλυση της μεθόδου, σαν ορισμένης ακρίβειας προσέγγιση της λύσης του αρχικού προβλήματος. Η συγγραφή όμως ενός προγράμματος απαιτεί μεγάλη προσοχή για την αποφυγή, όχι μόνον λογικών ασυνεπειών, αλλά και γραμματικών σφαλμάτων, γιατί η αλλαγή έστω και ενός γράμματος, η σημείου της στίξεως κάνει το κείμενο ακατανόητο για τον Η/Υ, γιατί αυτός δεν υποκαθιστά μιαν εσφαλμένη λέξη ή διφορούμενη πρόταση με την πλησιέστερη που θα έδινε νόημα, όπως θα κάναμε εμείς. Έτσι προσπαθούμε να κάνουμε τα υπολογιστικά προγράμματα όσο το δυνατό συντομότερα, ώστε να μειωθεί ο κίνδυνος να γίνουν σφάλματα, αλλά και ο κόπος για τον έλεγχο της τυχόν ύπαρξης τους.

Συνέπεια αυτής της πρακτικής επιδίωξης είναι ότι η Αριθμητική Ανάλυση προτιμά κατά κανόνα τις λεγόμενες Επαναληπτικές Υπολογιστικές Μεθόδους, μικρής εκτάσεως αλγορίθμους επαναλαμβανόμενοι προσδιορίζουν με όλο και μεγαλύτερη ακρίβεια την λύση ενός προβλήματος.

Κατά την χρήση των αλγορίθμων μεγάλη σημασία έχουν τα σφάλματα, τα οποία είναι δυο ειδών:

(α) *Μεθοδικά Σφάλματα*. Αυτά οφείλονται στον τρόπο κατασκευής της μεθόδου προσέγγισης και αποτελούν βασικό κριτήριο επιλογής των μεθόδων. Συχνά, π.χ. η διακριτοποίηση ενός προβλήματος, ώστε να είναι δυνατή η χρήση του Η/Υ εισάγει "παρασιτικές λύσεις" που δεν σχετίζονται με το αρχικό

πρόβλημα. Μια μέθοδος είναι αποδεκτή μόνον αν τέτοιες παρασιτικές λύσεις όπως και οποιαδήποτε αλλά μεθοδικά σφάλματα παραμένουν φραγμένα και έχουν σχετικά μικρές τιμές.

(β) *Σφάλματα Στρογγυλέματος*. Αυτά οφείλονται στο ότι ο Η/Υ διαθέτει για την αποθήκευση κάθε ενδιάμεσου αποτελέσματος μόνο πεπερασμένο χώρο μνήμης και επομένως αναγκαστικά αποκόπτει τα τελευταία δεκαδικά η δυαδικά ψηφία. Τέτοια σφάλματα προκύπτουν μάλιστα και κατά την μετατροπή της δυαδικής μορφής των αριθμών, που χρησιμοποιείται εσωτερικά από τον Η/Υ, στην δεκαδική με την οποία εκτυπώνονται τα αποτελέσματα. Τα σφάλματα στρογγυλέματος φαίνονται κατ' αρχήν ασήμαντα, γιατί περιορίζονται στα τελευταία μόνο δεκαδικά ψηφία. Όμως σε ορισμένες μεθόδους που απαιτούν μεγάλο όγκο πράξεων μπορούν να συσσωρευτούν και να παραποιήσουν σημαντικά τα αποτελέσματα. Έτσι σε πολλά σημεία είναι απαραίτητο να μελετηθεί προσεκτικά είτε η ευστάθεια της μεθόδου, δηλαδή αν κρατάει τα μεθοδικά σφάλματα μικρά η όχι, είτε η ευστάθεια του προβλήματος, δηλαδή αν το ίδιο το πρόβλημα είναι ευπαθές σε σφάλματα, ιδίως σφάλματα στρογγυλέματος.

Αρχίζουμε την μελέτη μας με την εκτίμηση των σφαλμάτων που μπαίνουν αναγκαστικά σε κάθε διαδικασία υπολογισμού συναρτήσεων.

Συμβολισμοί

Στο βιβλίο αυτό θα χρησιμοποιήσουμε τον διεθνή συμβολισμό για δεκαδικούς αριθμούς και παριστάνουμε την υποδιαστολή με τελεία και όχι, κατά την ελληνική πρακτική, με κόμμα. Έτσι θα 0.25 σημαίνει 1/8 και 1.125 σημαίνει 1+1/8=9/8. Αυτό τι κάνουμε γιατί ο ελληνικός συμβολισμός δημιουργεί προβλήματα ερμηνείας για διάφορους καθιερωμένους συμβολισμούς. Έτσι ο μέγιστος των αριθμών 0.1, 0.2, 0.3, max{0.1,0.2,0.3}, θα γραφόταν: max{0,1,0,2,0,3}, που μπορεί να σημαίνει τον μέγιστο των έξη ακεραίων αριθμών: 0,1,0,2,0,3.

Ο συμβολισμός: α(β)γ, που χρησιμοποιείται συχνά, σημαίνει την αριθμητική πρόοδο με πρώτο όρο α, λόγο β και τελευταίο όρο γ, δηλαδή την ακολουθία α, α+β, α+2β, …,γ, όπου πρέπει να είναι γ=α+νβ με κατάλληλο ακέραιο αριθμό ν. Για παράδειγμα, 1(1)n ≡ 1, 2, 3, …,n.

Επίσης «:=» σημαίνει «εξ ορισμού ίσον». Έτσι, η παράσταση: $f(x):=x^3-1/x$ σημαίνει ότι δίνουμε στην συνάρτηση x^3-1/x το όνομα f(x).

1. ΕΚΤΙΜΗΣΗ ΑΡΙΘΜΗΤΙΚΩΝ ΣΦΑΛΜΑΤΩΝ

Παράδειγμα Φυσικής Εφαρμογής

Η σχέση μεταξύ πίεσης P και όγκου V ενός μη ιδανικού αερίου σε μιαν απόλυτη θερμοκρασία T δίνεται από την σχέση:

$$P = \frac{RT}{V-b} - \frac{a}{V^2}$$

όπου R,a,b γνωστές σταθερές. Αν η μέτρηση του όγκου γίνεται με μέγιστο σφάλμα υ, $V = V_0 \pm$ υ, και η μέτρηση της θερμοκρασίας με μέγιστο σφάλμα τ, $T = T_0 \pm$ τ, να βρεθεί ποιο μέγιστο σφάλμα προκύπτει κατά τον υπολογισμό της πίεσης με αυτόν τον τύπο.

1.1 Η Προέλευση των Αριθμητικών ή Υπολογιστικών Σφαλμάτων

κατά την μέτρηση φυσικών μεγεθών προκύπτουν σφάλματα μετρήσεως που οφείλονται στην πεπερασμένη ακρίβεια των οργάνων μέτρησης. Αν επομένως θέλουμε να προσδιορίσουμε με την βοήθεια μαθηματικών σχέσεων αλλά Φυσικά μεγέθη θα έχουμε και γι' αυτά κάποια σφάλματα οφειλόμενα στην αβεβαιότητα με την οποία είναι γνωστά τα αρχικά μεγέθη. Επίσης κατά την εισαγωγή και επεξεργασία αριθμητικών δεδομένων στον Η/Υ τα αποτελέσματα όλων των πράξεων εκφράζονται με πεπερασμένη ακρίβεια, γιατί ο υπολογιστής διαθέτει πεπερασμένο χώρο μνήμης για κάθε ενδιάμεσο αποτέλεσμα. Είναι λοιπόν σημαντικό να προσδιορισθούν σε κάθε περίπτωση άνω φράγματα για τα υπάρχοντα σφάλματα.

1.2 Φράγματα Υπολογιστικών Σφαλμάτων

σφάλματα του παραπάνω είδους μπορούν να προσδιορισθούν με την βοήθεια του ακόλουθου θεωρήματος:

Θεώρημα 1.1:

Αν για τα μεγέθη $x_1, x_2, \ldots, x_n$ είναι γνωστές οι προσεγγιστικές τιμές:

$$x_i' = x_{i0} \pm a_i \quad , i = 1(1)n \tag{1.1}$$

η δε συνάρτηση

$$y = f(x) = f(x_1, x_2, \ldots, x_n) \tag{1.2}$$

έχει συνεχείς πρώτες μερικές παραγώγους στο σύνολο

$$M := \{ x \in \mathbf{R}^n, \text{ ή } \mathbf{C}^n : |x_i - x_{i0}| \le a_i \quad , i = 1(1)n \} \tag{1.3}$$

τοτε ισχύει για το σφάλμα προσέγγισης της τιμής f(x) με την $f(x_0)$:

$$| f(x) - f(x_0) | \le \sum_{i=1}^{n} a_i \max_{x \in M} | \frac{\partial f}{\partial x_i}(x) | := a_y \tag{1.4}$$

όπου η ποσότητα a_y λέγεται *μέγιστο (απόλυτο) σφάλμα* της συνάρτησης

f(x) στο σημείο x_0.

Απόδειξη:

Σύμφωνα με το θεώρημα μέσης τιμής του διαφορικού λογισμού ισχύει:

$$f(x)-f(x_0)=\sum_{i=1}^{n}(x_i-x_{i0})\frac{\partial f}{\partial x_i}(\xi),\ \xi\in M \qquad (1.5)$$

Αν πάρουμε τις απόλυτες τιμές και στα δυο μέρη της σχέσης αυτής και κάνουμε εκτίμηση του μέγιστου σφάλματος τότε προκύπτει το αποτέλεσμα (1-4).#

Η περιοχή M υποτίθεται ότι είναι εν γένει πολύ μικρή, ώστε οι τιμές των παραγώγων να μην μεταβάλλονται κατά πολύ μέσα σ' αυτήν. Για τούτο στην πράξη παίρνουμε συνήθως κατά τον προσδιορισμό του a_y αντί για το μέγιστο κάθε παραγώγου απλώς την τιμή της στο σημείο x_0. Έτσι προκύπτει:

$$a_y \approx \sum_{i=1}^{n} a_i \,|\frac{\partial f}{\partial x_i}(x_0)| \qquad (1.6)$$

Εφαρμόζοντας αυτόν τον τύπο έχουμε για την τιμή y = f(x_0):

$$|\frac{f(x)-f(x_0)}{f(x_0)}|\le \sigma_y := \frac{a_y}{|y_0|} \approx \sum_{i=1}^{n} a_i \,|\frac{f_i(x_0)}{f(x_0)}| \qquad (1.7)$$

με

$$f_i(x) := \frac{\partial f}{\partial x_i}(x) \qquad (1.8)$$

Η ποσότητα σ_y λέγεται *μέγιστο σχετικό σφάλμα* της f(x) στο σημείο x_0 .

Βασικές εφαρμογές των παραπάνω τύπων είναι οι ακόλουθοι τύποι:

α) y = x_1x_2 ...x_n : $\sigma_y \approx \sum_{i=1}^{n}\frac{a_i}{|x_{i0}|}=\sum_{i=1}^{n}\sigma_i$ (1.9)

β) y = x_1/x_2 : $\sigma_y \approx \frac{a_1}{|x_{i0}|}+\frac{a_2}{|x_{i0}|}=\sigma_1+\sigma_2$ (1.10)

γ) y = x_1 + x_2 +...+ x_n : $a_y=\sum_{i=1}^{n}a_i$ (1.11)

δ) y = x_1 – x_2 : a_y = a_1 + a_2 (1.12)

Αν οι τιμές της συνάρτησης f(x) προσδιορίζονται με την βοήθεια μιας προσεγγίζουσας συνάρτησης τότε ισχύει το ακόλουθο θεώρημα:

Θεώρημα 1.2:

Αν ισχύουν οι προϋποθέσεις του θεωρήματος 1.1 και p(x) είναι συνάρτηση που προσεγγίζει την f(x) Έτσι ώστε να είναι:

$$|f(x) - p(x)| \leq \varepsilon \text{ για } x \in M \qquad (1.13)$$

τότε για το συνολικό σφάλμα προσέγγισης της f(x) με $p(x_0)$ ισχύει:

$$|f(x) - p(x_0)| \leq a_y + \varepsilon \text{ , για } x \in M \qquad (1.14)$$

Απόδειξη

$$|f(x) - p(x_0)| = |f(x) - f(x_0) + f(x_0) - p(x_0)| \leq |f(x) - f(x_0)| + |f(x_0) - p(x_0)| \leq a_y + \varepsilon$$

γιατι ο πρώτος προσθετέος δεν ξεπερνάει το α με βάση το προηγούμενο θεώρημα και ο δεύτερος το ε με βάση την προϋπόθεση (1.13).

Κατά την εκτίμηση σφαλμάτων συχνά χρησιμοποιείται ο ακόλουθος ορισμός:

Ορισμός 1.1:

Λέμε ότι ο αριθμός x_0 προσεγγίζει τον αριθμό x με ακρίβεια n δεκαδικών ψηφίων μετά την υποδιαστολή, αν ισχύει:

$$|x - x_0| \leq 0.5\ 10^{-n},\ n = 1, 2, \ldots \qquad (1.15)$$

Στην περίπτωση αυτή τα δεκαδικά ψηφία του x_0 συμπίπτουν με τα αντίστοιχα του x μέχρι το ψηφίο n μετά την υποδιαστολή εκτός αν Αυτό είναι "9" η "0" οπότε η σύμπτωση η μη εξαρτάται από την τιμή του n+1 ψηφίου μετά την υποδιαστολή.

Π.χ. στην περίπτωση x = 5.9996 τα x με

$$|x - x_0| < 0.5\ 10^{-3} = 0.0005$$

δεν έχουν οπωσδήποτε μέχρι και το τρίτο δεκαδικό ψηφίο μετά την υποδιαστολή ίδιο με εκείνο του x , γιατί εδώ είναι:

$$5.9991 < x < 6.0001$$

Παρατήρηση:

Οι τιμές a_i , i = 1(1)n που προσδιορίζουν με τι ακρίβεια ξέρουμε τις ανεξάρτητες μεταβλητές x_i ($|x_i - x_{i0}| \leq a_i$) μπορεί, όπως είπαμε να οφείλονται στην περιορισμένη ακρίβεια με την οποία μετρούν τα όργανα μετρήσεως φυσικών μεγεθών που διαθέτουμε, μπορεί όμως να οφείλονται απλά στο αναγκαστικό στρογγύλεμα των τιμών που κάνει ο Η/Υ για να μπορέσει να αποθηκεύσει ενδιάμεσα αποτελέσματα. Πώς, όμως μπορούμε να προσδιορίσουμε με ποιαν ακρίβεια αποθηκεύει ο Η/Υ τα ενδιάμεσα αποτελέσματα;

Ένας απλός τρόπος για να διαπιστώσουμε με τι ακρίβεια απομνημονεύει Ένας Η/Υ τα διάφορα αριθμητικά δεδομένα, είναι να διαλέξουμε μιαν φθίνουσα ακολουθία θετικών αριθμών $\{\varepsilon_\kappa\}$, να δώσουμε στον Η/Υ την εντολή να υπολογίσει τις ποσότητες $a_\kappa = 1 + \varepsilon_\kappa$ και να τυπώσει τις ποσότητες $\beta_\kappa = a_\kappa - 1$.

Ενώ αρχικά θα είναι $\beta_\kappa = a_\kappa$, όπως είναι σωστό, όταν τα ε_κ γίνουν μικρότερα από κάποιο όριο, ρ, το αποτέλεσμα θα είναι $\beta_\kappa = 0$. Αυτό σημαίνει ότι η αντίστοιχη μνήμη του Η/Υ δεν είναι αρκετά μεγάλη για να κρατήσει τον αριθμό 1 + ε και τον στρογγυλεύει σε 1.

Στην γλώσσα προγραμματισμού GWBASIC διαπιστώνουμε έτσι ότι όταν εκτελεί τις πράξεις με απλή ακρίβεια (single precision), είναι $\rho \approx 0.5\ 10^{-7}$ ως 10^{-7}, Ενώ με διπλή ακρίβεια εκτελέσεως πράξεων (double precision) έχουμε $\rho \approx 10^{-17}$ το

προγραμματιστικό περιβάλλον MATLAB εκτελεί εξ αρχής τις πράξεις με ακρίβεια 16 δεκαδικών ψηφίων, έστω και αν δεν τα τυπώνει όλα.

Παραδείγματα Εύρεσης Φραγμάτων Σφάλματος:

1. Για

$$0 \le x \le \ln 2 = 0.693$$

είναι

$$|e^{-x} - p(x)| \le 0.003$$

όπου

$$p(x) := 1 - 0{,}9664\, x + 0{,}3536\, x^2$$

(Βλ. Abramowitz - Stegun: Handbook of Mathematical Functions, Dover, σελ 71).

Αν $s = 0.5 \pm 0.001$ ($|s - 0.5| \le 0.001$) να υπολογισθεί προσεγγιστικά η τιμή e^{-s} και να εκτιμηθεί το σφάλμα προσέγγισης.

Λύση:

Σαν προσέγγιση για την e^{-s} χρησιμοποιούμε την τιμή $p(x_0)$, όπου $x_0 = 0.5$, $a_x = 0.001$. Δηλαδή

$$e^{-s} \approx p(0.5) = 0.6052$$

Το συνολικό σφάλμα είναι εδώ σύμφωνα με το θεώρημα 1.2 απόλυτα μικρότερο η ίσο με $\varepsilon + a_y$, όπου $\varepsilon = 0.003$ Ενώ

$$a_y = a_x \max_{x \in M} |f'(x)| = a_x \max_{x \in M} |e^{-x}| \cong a_x e^{-x_0}$$

όπου $M = [x_0 - a_x , x_0 + a_x] = [0.499, 0.501]$. Δηλαδή

$$a_y \approx a_x\, p(x_0) \approx 0.0006$$

$\varepsilon + a_y \cong 0.0036$ και επομένως

$$e^{-s} = 0.6052 \pm 0.0036$$

2. Ποιο είναι το άνω φράγμα της απόλυτης τιμής του σφάλματος υπολογισμού της κεντρομόλου δύναμης με βάση τον τύπο:

$$F = \frac{mv^2}{r}$$

αν $m = m_0 \pm a_m$, $v = v_0 \pm a_v$, $r = r_0 \pm a_r$;

Να θεωρηθεί ότι όλα τα παραπάνω μεγέθη έχουν αναχθεί σε καθαρούς αριθμούς με διαίρεση με τις αντίστοιχες φυσικές μονάδες.

Λύση:

σύμφωνα με τον τύπο (1.6) είναι

$$a_F = a_m \left|\frac{\partial F}{\partial m}\right| + a_v \left|\frac{\partial F}{\partial v}\right| + a_r \left|\frac{\partial F}{\partial r}\right| = a_m \frac{v_0^2}{r_0} + a_v \frac{2m_0 v_0}{r_0} + a_r \frac{m_0 v_0^2}{r_0^2}$$

(όπου οι μερικές παραγωγοί παίρνονται στο σημείο (m_0, v_0, r_0)) ή απλούστερα,

$$\frac{a_F}{F_0} \approx \frac{a_m}{m_0} + 2\frac{a_v}{v_0} + \frac{a_r}{r_0} \quad \mu\varepsilon \quad F_0 = \frac{m_0 v_0^2}{r_0}$$

Ασκήσεις:

1. Υπολογίστε το άνω φράγμα του απόλυτου σφάλματος a_y για τον προσδιορισμό της πίεσης ενός αεριού με βάση την σχέση:

$$P = \frac{RT}{V-b} - \frac{a}{V^2}$$

αν a, b, R είναι γνωστές σταθερές και $V = V_0 \pm \alpha_V$, $T = T_0 \pm \alpha_T$.

2. Να δοθεί άνω φράγμα του σφάλματος υπολογισμού της παράστασης $z = x^a y^b$ στις δυο ακόλουθες τρεις περιπτώσεις:

α) $a,b \in \mathbf{R}$ σταθερά και $x = x_0 \pm \alpha_x$, $y = y_0 \pm \alpha_y$

β) $x,y \in \mathbf{R}$ σταθερά και $a = a_0 \pm \alpha_a$,$b = b_0 \pm \alpha_b$

γ) $a,b,x,y \in \mathbf{R}$ μεταβλητά μεγέθη με τις παραπάνω διακυμάνσεις τιμών.

3. Εφαρμόσατε το αποτέλεσμα της ασκ. (2α) για $z = x^2 y^{-3}$ αν $x= 5.3 \pm 0.002$, $y = 3.1 \pm 0.003$ και της (2γ) όταν είναι $z=x^a y^b$ με τις παραπάνω περιοχές τιμών των x, y και με $a = 2 \pm 0.001$, $b = -3 \pm 0.005$.

4. Να δοθεί άνω φράγμα του σφάλματος υπολογισμού της συνάρτησης $z:= \sin x \cos y$ για $|x - \pi/6| \leq 0,1$ $|y - \pi/4| \leq 0,01$.

5. Αν $x_i = x_{i0} + \alpha_i$, i = 1,2,3 να βρεθεί το μέγιστο σχετικό σφάλμα για τις παραστάσεις x_1x_2/x_3 και $x_1/(x_2x_3)$.

6. Αν οι πράξεις γίνονται με ακρίβεια 5 δεκαδικών ψηφίων μετά την υποδιαστολή δείξτε ότι $(2/3)*6 \neq 2*(6/3)$, και γενικότερα, ότι για φυσικούς αριθμούς m, n, k , όπου m,n η k,n πρώτοι προς αλλήλους, είναι $(m/n)*k \neq m*(k/n)$.

7. Αν οι πολλαπλασιασμοί εκτελούνται με ακρίβεια 7 δεκαδικών ψηφίων μετά την υποδιαστολή, με τι σφάλμα συνδέεται ο υπολογισμός της δύναμης a^{10} , του δεκαδικού αριθμού a = 0.3333;

2. ΕΠΑΝΑΛΗΠΤΙΚΗ ΛΥΣΗ ΜΗ ΓΡΑΜΜΙΚΩΝ ΕΞΙΣΩΣΕΩΝ ΜΕ ΕΝΑΝ ΑΓΝΩΣΤΟ

Παραδείγματα Φυσικών και Τεχνικών Εφαρμογών:

1. *Ανεμόσκαλα μπροστά σε εμπόδιο:*

Ποιο είναι το ύψος, στο οποίο μπορεί να φθάσει σε έναν κάθετο τοίχο μια Ανεμόσκαλα μήκους L αν μπροστά στον τοίχο βρίσκεται ορθογώνιο εμπόδιο (π.χ. κιβώτιο) ύψους b και πλάτους a;
(Υποτίθεται ότι η σκάλα θα ακουμπήσει στον τοίχο, στην ακμή του κιβωτίου και στο δάπεδο).

Αν y είναι το ύψος στο οποίο θα φθάσει η σκάλα και x είναι η οριζόντια απόσταση της βάσης της από τον τοίχο, τότε χρησιμοποιώντας ομοιότητα τριγώνων βλέπουμε ότι

$$y = bx/(x-a) \quad \text{Ενώ} \quad x^2 + y^2 = L^2$$

από τις σχέσεις αυτές προκύπτει ότι η οριζόντια απόσταση x ικανοποιεί την πολυωνυμική εξίσωση τετάρτου βαθμού

$$x^4 + 2ax^3 + (b^2 + a^2 - L^2)x^2 + 2aL^2 x - a^2 L^2 = 0$$

Ενώ προφανώς πρέπει να είναι $a < x < L$.

2. Κάμψη ελαστικής στήλης:

(Βλέπε: M.S.El Nasie: Stress, Stability and Chaos, Mc Graw - Hill, 1990, σελ 27-28).

Λεπτή στήλη μήκους L είναι στερεωμένη αρθρωτά στο δάπεδο, έτσι ώστε να περιστρέφεται ελεύθερα, και στέκεται κατακόρυφα με την βοήθεια ελατηρίου που βρίσκεται στην βάση της. Αν στην κορυφή της στήλης ασκηθεί κατακόρυφη δύναμη μεγέθους P ,ποια είναι η γωνία α εκτροπής της στήλης από την κατακόρυφη διεύθυνση;

Η γωνία αυτή προκύπτει αν εξισώσουμε την ροπή που ασκείται από την δύναμη με την ροπή ελαστικής επαναφοράς της στήλης:

$$PL\sin\alpha = C\alpha$$

όπου C είναι συντελεστής που χαρακτηρίζει την αντίσταση του ελατηρίου, δίνει Δηλαδή την ροπή που χρειάζεται για την στρέψη του Κατά ένα ακτίνιο.

Το μοντέλο αυτό περιγράφει προσεγγιστικά και το τι συμβαίνει όταν η στήλη είναι ελαστική και κατακόρυφα μπηγμένη (πακτωμένη).
μπορούμε Δηλαδή να θεωρήσουμε, ότι η ίδια η στήλη λειτουργεί σαν ελατήριο επαναφοράς.

Αν η ασκούμενη στην κορυφή της παραπάνω δοκού δύναμη είναι οριζόντια, τότε έχουμε την σχέση:

$$PL\cos\alpha = C\alpha$$

Ενώ αν το ελατήριο στην βάση της είναι μη γραμμικό με ροπή στρέψεως $M = C\alpha^2$, τότε έχουμε την:

$$PL\cos\alpha = C\alpha^2$$

3. Πτώση αλεξιπτωτιστή:

(Βλ. Chapra - Canale: Numerical Methods for Engineers, Mc Graw - Hill, 1990) Η Πτώση ενός αλεξιπτωτιστή μπορεί να περιγραφτει προσεγγιστικά από την

διαφορική εξίσωση:

$$m\frac{dv}{dt} = mg - cv$$

όπου m η μάζα του, g η επιτάχυνση της βαρύτητας και -cv η αντίσταση του αέρα στην Πτώση, που Κατά προσέγγιση την θεωρούμε ανάλογη προς την ταχύτητα πτώσεως, Ενώ ο συντελεστής C εξαρτάται προφανώς από το μέγεθος του αλεξίπτωτου:

Η λύση αυτής της διαφ.εξ. είναι:

$$v(t) = \frac{gm}{c}(1 - e^{-(c/m)t})$$

Ενώ το διανυόμενο διάστημα σε χρόνο t είναι:

$$x(t) = \frac{gm}{c}(t + \frac{m}{c}e^{-(c/m)t} - \frac{m}{c})$$

Με την βοήθεια αυτών των σχέσεων μπορούμε να εξετάσουμε δυο βασικά προβλήματα. Πρώτον, ποιά πρέπει να είναι η σταθερά C, Δηλαδή το μέγεθος του αλεξίπτωτου, ώστε η τελική ταχύτητα να μην ξεπερνάει μιαν ορισμένη τιμή από την πρώτη σχέση βλέπουμε ότι για μεγάλους χρόνους η ταχύτητα v_0 πτώσης τείνει να γίνει σταθερή. έτσι, για να έχουμε την ίδια σταθερά (το ίδιο αλεξίπτωτο) ασχέτως ύψους πτώσης βάζουμε $c = gm/v_0$. Ο προσδιορισμός του χρόνου πτώσης από ύψος h απαιτεί όμως την λύση μιας μη γραμμικής εξίσωσης. Με $x(t)=h$, $gt/v_0=z$, $1+gh/v_0^2=\beta$ προκύπτει από την τελευταία εξίσωση:

$$\beta=z+e^{-z}$$

ενώ με $\beta-z=y$ η εξίσωση παίρνει την μορφή:

$$\ln y - y+\beta=0, \text{ με } \beta>1.$$

4. Σχεδίαση Ηλεκτρικού Κυκλώματος

(Βλ. Chapra -Canale σελ. 182-183).

Στους πόλους μιας μπαταρίας τάσεως V συνδέεται την στιγμή t=0 η παράλληλη σύνδεση ενός πυκνωτή χωρητικότητας C και της εν σειρά σύνδεσης μιας ωμικής αντίστασης R και ενός πηνίου με συντελεστή επαγωγής L. Το φορτίο του πυκνωτή μεταβάλλεται τότε με βάση την σχέση:

$$q(t) = q_0 e^{-\frac{Rt}{2L}}\cos(\omega t), \quad \omega = \sqrt{\frac{1}{LC} - \frac{R^2}{4L^2}}$$

όπου $q_0 = V_0C$. Να προσδιορισθεί η ωμική αντίσταση R, έτσι ώστε να εκπέμπει ενέργεια με ορισμένο ρυθμό, π.χ., έτσι ώστε σε χρόνο t = 0.05 sec το φορτίο να είναι 1% του αρχικού (q/q_0 = 0.01) αν L = 5 H και C = 10 F.

5. Κρίσιμο φορτίο κολώνας

(Βλ. Iyengar - Gupta: Programming Methods in Structural Design, Edward Arnold,, 1980, σελ. 13).

Μια κολώνα μήκους L είναι αναρτημένη αρθρωτά στο κάτω άκρο της ενώ το πάνω άκρο της είναι σταθερά συνδεδεμένο υπό ορθή γωνία με οριζόντια δοκό ιδίου μήκους μπηγμένη σε τοίχο. Να προσδιορισθεί το Κρίσιμο φορτίο P της κολώνας αυτής, δηλαδή η μέγιστη κάθετη δύναμη που μπορεί να ασκηθεί στο

πάνω άκρο της χωρίς αυτή να καμφθεί.

Για δεδομένο συντελεστή ελαστικότητας Ε και ροπή αδρανείας Ι της κολώνας το φορτίο αυτό δίνεται από την μη γραμμική εξίσωση:

$$\tan \lambda = 4\lambda/(4 + \lambda)$$

όπου $\lambda := P L^2/(EI)$.

6. Προσδιορισμός της Λύσης Διαφορικών Εξισώσεων

Τα προβλήματα αρχικών τιμών:

$$(\alpha)\ yy' - f'(x) = 0, y(x_0) = y_0 \quad , \quad (\beta)\ \frac{z'}{z} = \frac{f'(x)}{2f(x)}, z(x_0) = z_0$$

Έχουν αντιστοίχως τις λύσεις:

$$(\alpha)\ y = \pm\sqrt{2f(x) + y_0^2 - 2f(x_0)} \ , \ (\beta)\ z = \pm z_0\sqrt{\frac{f(x)}{f(x_0)}}$$

Οι λύσεις αυτές είναι πραγματικές μόνο όταν τα υπόριζα είναι θετικά. Για να προσδιορίσουμε τα διαστήματα του x όπου συμβαίνει αυτό πρέπει να λύσουμε τις εξισώσεις:

$$(\alpha)\ 2f(x) + y_0^2 - 2f(x_0) = 0 \ \text{ ή } (\beta)\ f(x)=0.$$

Για παράδειγμα, η λύση του προβλήματος: yy'+1-1/x=0, y(0)=1
είναι πραγματική μόνο μεταξύ των δύο ριζών της εξίσωσης: 2ln x = x-1

2.1 Γραφικός Εντοπισμός των Λύσεων

Η γενική μορφή μη γραμμικών εξισώσεων μιας μεταβλητής είναι:

$$f(x) = 0 \qquad (2.1)$$

οπου f(x) ειναι συναρτηση ωρισμενη και συνεχης σε ενα διαστημα I = [a,b] ($f \in C[a,b]$). Για την λυση τετοιων εξισωσεων χρησιμοποιουμε κατα κανονα επαναληπτικες μεθοδους, δηλ. μεθοδους της μορφης:

$$x_{k+1} = \varphi(x_k)\ ,\ k = 0,1,2,3,... \qquad (2.2)$$

όπου x_0 η αρχική τιμή και x_k το αποτέλεσμα της k επανάληψης.

Η επαναληπτική αυτή ακολουθία, εφ' όσον συγκλίνει, μας δίνει προφανώς μιαν λύση του «Προβλήματος Σταθερού Σημείου»:

$$x = \varphi(x)$$

Σε αυτήν την μορφή προσπαθούμε, λοιπόν να φέρουμε την αρχική εξίσωση ώστε να μπορέσουμε να κατασκευάσουμε μίαν επαναληπτική διαδικασία, που ενδεχόμενα θα συγκλίνει σε κάποιαν λύση της εξίσωσης. Επειδή όμως μία εξίσωση μπορεί να έχει περισσότερες από μία λύσεις, είναι φανερό ότι για να προσδιορίσουμε κάποια συγκεκριμένη από αυτές τις λύσεις πρέπει να πάρουμε μίαν αρχική τιμή x_0 που να βρίσκεται κοντά της. Για τούτο, πριν ασχοληθούμε με κατάλληλες επαναληπτικές διαδικασίες, θα εξετάσουμε το πώς προσδιορίζουμε πόσες λύσεις έχει μία εξίσωση και το πού βρίσκονται αυτές.

Ο εντοπισμος των λυσεων πριν απο τον τελικο προσδιορισμο τους ειναι σημαντικος και για εναν αλλο λογο. Οταν η (2.1) μας δινει την λυση καποιου τεχνικου προβληματος, τοτε συχνα δεν ειναι ολες οι λυσεις της αποδεκτες,

αλλα μονον οσες πληρουν προσθετους περιορισμους που προκυπτουν απο το τεχνικο προβλημα. Γι' αυτο δεν εχει νοημα η τυφλη αναζητηση μιας, οποιασδηποτε λυσης. Στο παράδειγμα 3 (πτώση αλεξιπτωτιστή), από τις λύσεις της εξίσωσης: $\ln y - y + \beta = 0$ ($\beta > 1$) ενδιαφέρουν μόνο όσες υπερβαίνουν την τιμή β, γιατί $\beta - y = z = gt/v_0 > 0$.

Ο γραφικος εντοπισμος του συνολου των πραγματικων λυσεων της (2.1) μπορει να γινει βασικα αν σχεδιασουμε την συναρτηση $y=f(x)$, εστω και μονον ποιοτικα, δηλαδη χωρις μεγαλην ακριβεια. Τα σημεια οπου η καμπυλη αυτη τεμνει τον οριζοντιο αξονα ειναι οι λυσεις της (2.1).

Ο σχεδιασμος της καμπυλης αυτης μπορει να γινει βεβαια και αυτοματα απο καποιο σχεδιαστικο προγραμμα, ομως η ποιοτικη κατανοηση της μορφης των γραφηματων που αντιστοιχουν στις διαφορες βασικες συναρτησεις ειναι παρ' ολα αυτα απαραιτητη, γιατι κανενα σχεδιαστικο προγραμμα δεν μπορει να διατρεξει ολοκληρον τον αξονα των x απο $-\infty$ ως $+\infty$. Δεν μπορουμε, δηλαδη, εν γενει να αποφυγουμε το να κανουμε οι ιδιοι ενα ποιοτικα (οχι ποσοτικα) σωστο προχειρο γραφημα της συναρτησης f.

Στην περιπτωση που η συναρτηση f είναι άθροισμα δύο ή περισσοτέρων όρων ο γραφικός εντοπισμος των λυσεων μπορει να απλοποιηθει με βαση την παρατηρηση ότι αν η f(x) είναι διαφορά δύο απλουστέρων συναρτήσεων:

$$f(x) := f_1(x) - f_2(x) \qquad (2.4)$$

τοτε η σχεση $f(s) = 0$ αντιστοιχει στην $f_1(s) = f_2(s)$. Δηλαδη τα σημεια μηδενισμου της $y = f(x)$ αντιστοιχουν στα σημεια διασταυρωσης των καμπυλων $y_1 = f_1(x)$ και $y_2 = f_2(x)$.

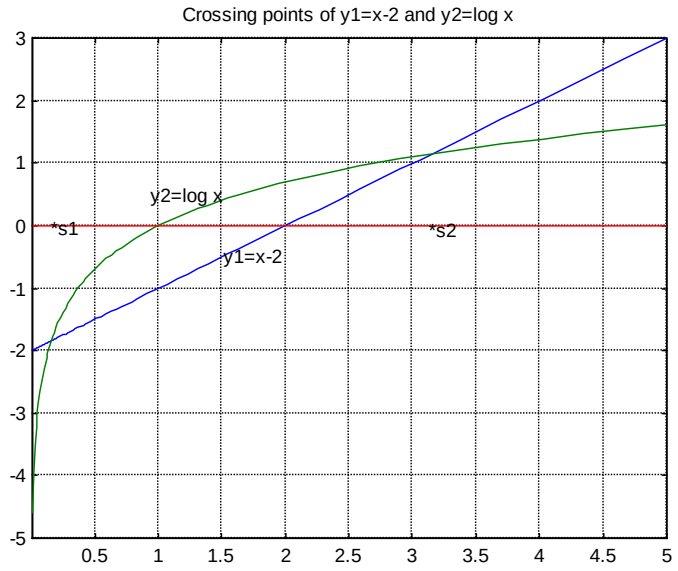
Αντι, δηλαδη, να σχεδιασουμε την $y = f(x)$ και να εντοπισουμε που τεμνει τον οριζοντιο αξονα, αρκει να σχεδιασουμε τις απλουστερες καμπυλες $y_1 = f_1(x)$ και $y_2 = f_2(x)$ και να εντοπισουμε σε ποια σημεια διασταυρωνονται.

Παραδειγματος χαριν, για την εξισωση:

$$f(x) := \ln x - x + 2 = 0 \qquad (2.5)$$

προκυπτει με $f_1(x) = \ln x$ και $f_2(x) = x - 2$ το ακολουθο σχεδιαγραμμα, από το οποίο προκύπτει ότι η εξίσωση (2.5) έχει δύο λύσεις:

$$s_1 \in (0,1) \quad \text{και} \quad s_2 > 2 \qquad (2.6)$$



Σχημα 2.1 : ln x = x - 2

Για να επιβεβαιωσουμε οτι μεσα σε καποιο διαστημα που εντοπισαμε γραφικα υπαρχει πραγματι μια λυση της (2.1) μπορει να χρησιμοποιηθει το γνωστο θεωρημα του Bolzano:

Θεώρημα 2.1 (Bolzano):

Αν η συναρτηση f ειναι συνεχης στο διαστημα [a,b] ($f \in C[a,b]$) και εχει αντιθετα προσημα στα ακρα του:

$$f(a)f(b) < 0 \qquad (2.7)$$

τοτε υπαρχει μεσα στο [a,b] τουλαχιστον ενα σημειο μηδενισμου της f.

Για την αποδειξη του θεωρηματος αυτου μπορει κανεις να ανατρεξει σε οποιοδηποτε βιβλιο Αναλυσης (Βλ.π.χ. Ι.Β.Κιουστελιδη: "Αριθμητικη Αναλυσις", Αθηνα, 1977 σελ. 11-12). Εποπτικα ομως, η ισχυς του ειναι εμφανης αν σκεφθουμε, οτι μια συνεχης καμπυλη που ξεκινα π.χ. απο καποιαν αρνητικη τιμη και καταληγει σε θετικη πρεπει οπωσδηποτε ενδιαμεσα να περασει και απο την τιμη μηδεν.

Για την εφαρμογη του θεωρηματος αυτου πρεπει ομως να προσεχθει αν το σημειο μηδενισμου s που προσπαθουμε να εντοπισουμε ειναι απλο, η πολλαπλο σημειο μηδενισμου:

Ορισμός 2.1

Το σημειο $s \in (a,b)$ λεγεται πολλαπλο σημειο μηδενισμου της συναρτησης $f \in C[a,b]$ αν ισχυει:

$$0 = f(s) = f'(s) = f''(s) = \ldots = f^{(n-1)}(s),\ f^{(n)}(s) \neq 0 \qquad (2.8)$$

Αν αναπτυξουμε τις συναρτησεις f και f' σε αναπτυγματα Taylor γυρω απο το σημειο s (με χρηση αντιστοιχα n και n-1 ορων και ορου σφαλματος), τοτε, εξ αιτιας του μηδενισμου των συντελεστων, προκυπτει οτι τοσο η f οσο και η f'

ισουνται με τον αντιστοιχο ορο σφαλματος:

$$f(x) = (x-s)^n \frac{f^{(n)}(\xi(x))}{n!}, \quad f'(x) = (x-s)^{n-1} \frac{f^{(n)}(\zeta(x))}{(n-1)!} \qquad (2.9)$$

οπου οι (κατα τα αλλα αγνωστες) θεσεις ξ(x) και ζ(x) βρισκονται μεταξυ των σημειων x και s, και τεινουν επομενως στο s οταν το x τεινει στο s.

Βλεπουμε, δηλαδη, οτι, οπως και στα πολυωνυμα, οταν μια τυχουσα συναρτηση f εχει ενα πολλαπλο σημειο μηδενισμου με πολλαπλοτητα n, τοτε περιλαμβανει τον παραγοντα $(x - s)^n$.

Επισης παρατηρουμε, οτι το πηλικο f/f' περιλαμβανει μονον τον ορο x-s (στην πρωτη δυναμη), εχει δηλαδη το s σαν απλο σημειο μηδενισμου και μαλιστα ισουται ασυμπτωτικα (για $x \to s$) με (x-s)/n.

Η παρατηρηση αυτη επιτρεπει την εφαρμογη του θεωρηματος του Bolzano σε καθε περιπτωση, δηλαδη ακομα και οταν το s ειναι αρτιας πολλαπλοτητας σημειο μηδενισμου (διπλο, τετραπλο, κοκ), οποτε το προσημο της f ειναι ιδιο και στις δυο πλευρες του σημειου s. Στην περιπτωση αυτην αρκει να εφαρμοσουμε το θεωρημα στην f/f' που εμφανιζει οπωσδηποτε εναλλαγη προσημου. Η χρηση της f' ειναι εδω επισης δυνατη, αλλα για μεγαλα n δεν συμφερει, γιατι οι τιμες της κοντα στο s θα ειναι εν γενει πολυ μικρες.

Ο γραφικός εντοπισμός των λύσεων στις ασκήσεις που ακολουθούν μπορεί να γίνει π. χ. με το υπολογιστικό περιβάλλον MATLAB. Όμως συνίσταται η σχεδίαση των καμπύλων που διασταυρώνονται να γίνει κατ' αρχήν με πρόχειρη σχεδίασή τους. Η λυση των ασκησεων γραφικου εντοπισμου λυσεων που διινονται πρεπει να γινει με την σκεψη και οχι με Η/Υ για δυο λογους:

1. Ειναι σημαντικο να μαθουμε να βλεπουμε πισω απο μιαν αναλυτικη εκφραση την γραφικη της παρασταση (ποιοτικα, στα βασικα της χαρακτηριστικα) γιατι μονον ετσι καταλαβαινουμε την φυση του φαινομενου που περιγραφει η αναλυτικη εκφραση. Ενα γραφημα μπορει να περιλαβει ενα ολοκληρο κεφαλαιο περιγραφων και επεξηγησεων, αλλα πρεπει να ξερουμε να το διαβασουμε.

2. Καμπυλες σχεδιασμενες με Η/Υ δεν μπορουν να δειξουν στο συνολο του το γραφημα μιας συναρτησης, αλλα μονον ενα μικρο τμημα του. Αν δεν ειμαστε προιδεασμενοι για τα γενικα χαρακτηριστικα του γραφηματος μπορει να μας διαφυγει το ποια περιοχη ειναι καιριο να σχεδιασθει.

Ασκήσεις:

Να γινει γραφικος εντοπισμος ολων των πραγματικων λυσεων των εξισωσεων:

1. $x = \sin x$
2. $2\cos x = |x - 1|$
3. $x/2 = \arctan x + 1$
4. $\ln|\cos x| = x - 1$
5. $x = e^{x-2}$
6. $\ln x - x + 1 = 0$
7. $\sin x - 2|x| + 1 = 0$
8. $2x + x = \cos x$

Παραδείγματα Γραφικού Εντοπισμού Λύσεων

1. Να εντοπισθουν ολες οι λυσεις της εξισωσης:

$$\ln|x + 1| = |x - 2|$$

Λύση

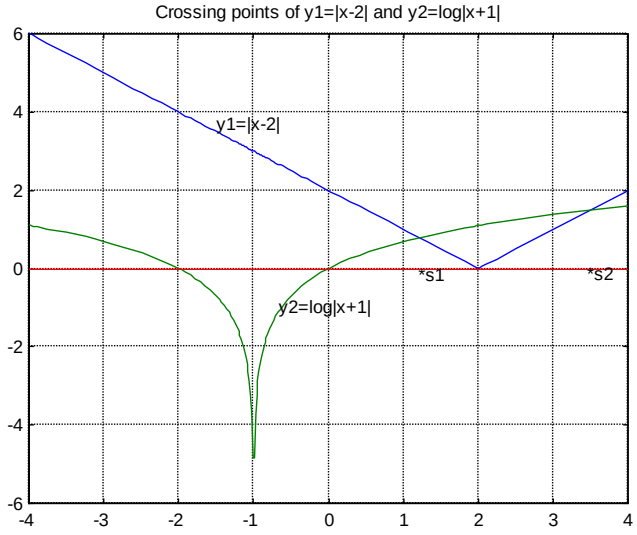
Η συναρτηση $y = |x-2|$ παριστανει μιαν τεθλασμενη ευθεια, που παραμενει παντοτε στο ανω ημιεπιπεδο και μηδενιζεται για $x=2$. Για $x < 2$ ειναι $y = 2-x$ ενω για $x > 2$ ειναι $y = x-2$.

Η συναρτηση $y = \ln|x+1|$ για $x > -1$ $(x + 1 > 0)$ εχει την κλασσικη μορφη της συναρτησης του λογαριθμου μετακινημένη στον αξονα $x = -1$ (αντι για τον $x = 0$). Για $x = -1$ "ξεκιναει" απο το $-\infty$, φθανει στο 0 για $x+1=1$, δηλ. για $x=0$ (γιατι $\ln 1=0$), και τεινει για αυξανομενα x προς το απειρο, αλλα πολυ αργα, γιατί τεινει να οριζοντιωθει αφού η παράγωγός της, $1/(x+1)$, που προσδιορίζει την κλίση της τείνει προς το μηδέν.

Σημειωστε οτι η λογαριθμικη συναρτηση τεινει προς το απειρο πιο αργα απο οποιαδηποτε θετικη δυναμη του $x+1$, γιατι για x αρκετα μεγαλο (για $a(x+1)^a>1$) ειναι

$$((x+1)^a)' = a(x+1)^{a-1} > (x+1)^{-1} = (\ln(x+1))'$$

Ο λογαριθμος δεν οριζεται, η μαλλον, δεν ειναι πραγματικη συναρτηση για αρνητικες τιμες του ορισματος του (παιρνει μιγαδικες τιμες). Στην προκειμενη περιπτωση ομως, επειδη το ορισμα του λογαριθμου ειναι μια απολυτη τιμη, η συναρτηση y ειναι ορισμενη και για $x < -1$. Μαλιστα η μορφη της ειναι συμμετρικη ως προς τον αξονα $x = -1$. Δηλαδη, εχει ιδιες τιμες σε σημεια που απεχουν ισες αποστασεις απο το $x = -1$ προς τα δεξια και τα αριστερα.



Σχημα 2.2: $\ln|x + 1| = |x - 2|$

Τα σημεια τομης των συναρτησεων y_1 και y_2 βρισκονται ολα στο ανω ηιεπιπεδο και μαλιστα στο δεξιο τεταρτημοριο. Για $x<-2$ οι συναρτησεις $y_1=2-x$ και $y_2=\ln(-x-1)$ ειναι μεν θετικες αλλα δεν διασταυρωνονται, γιατι για $x\to-\infty$ ο λογαριθμος τεινει στο απειρο πιο αργα απο την ευθεια $y_1=2-x$, όπως φαινεται αν συγκρινουμε τις παραγωγους.

Για $x > 0$ η καμπυλη του λογαριθμου συναντάει πρωτα τον κλαδο $y = 2 - x$

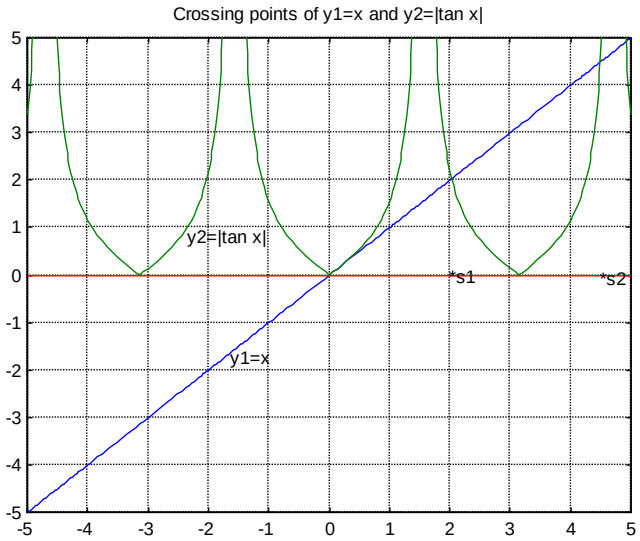
και μετα τον κλαδο $y = x - 2$ της τεθλασμενης ευθειας. Υπαρχουν, δηλαδη, τελικα δυο σημεια διασταυρωσεως, $s_1 \in (0, 2)$ και $s_2>2$. Π.χ. για $x = 4$ ειναι ηδη $y_1=2 > \ln 5 = 1,6$. Δηλαδη η ευθεια εχει ηδη ξεπερασει τον λογαριθμο ξεκινωντας απο το 0 για $x=2$. Αρα το σημειο διασταυρωσεως τους, s_2 , ειναι μεταξυ 2 και 4.

2. Να εντοπισθουν γραφικα ολες οι λυσεις της εξισωσης:

$$x = |\tan x|$$

Λύση

Η $y_1 = x$ ειναι ευθεια που περναει απο την αρχη των συντεταγμενων υπο γωνια 45^0. Η $y_2 = |\tan x|$ παραμενει παντα μη αρνητικη, και κινειται περιοδικα μεταξυ 0 και $+\infty$. Για $x=0$ ειναι 0 οπως και η ευθεια, για $x = \pi/2$ γινεται $+\infty$,



Σχημα 2.3: $x = |\tan x|$

κατεβαινει στην συνεχεια στο 0 για $x = \pi$, ανεβαινει στο $+\infty$ για $x = 3\pi/2$, κοκ. Για $x<0$ εχει τιμες συμμετρικες προς τον αξονα των y, αλλα εκει δεν μας ενδιαφερει αφου εδω οι λυσεις θα ειναι οπωσδηποτε θετικες ($x>0$ γιατι ειναι ισο με μιαν απολυτη τιμη).

Τα σημεια διασταυρωσεως αυτων των καμπυλων ειναι απειρα, ενα για καθε κλαδο της εφαπτομενης. Μαλιστα, για αυξανομενα x τεινουν ολο και περισσοτερο να συμπεσουν με τις τιμες $(2\nu-1)\pi/2$ ($\nu \in \mathbf{N}$):

$$s_\nu \cong (2\nu-1)\pi/2 \text{ , για } \nu >> 1$$

Το σημειο διασταυρωσεως $s = 0$ πρεπει να μελετηθει χωριστα, γιατι ενω η μια καμπυλη προχωρει ευθυγραμμα, η αλλη κυρτωνεται. Θα μπορουσε, λοιπον να υπαρχει και δευτερο σημειο διασταυρωσεως κοντα στο 0 (Αυτο συμβαινει, π.χ. με ολες τις ευθειες $y_1 = \alpha x$, με $\alpha >1$). Στην προκειμενη περιπτωση ομως, πρεπει να θυμηθουμε οτι $\tan x \cong x$ για $x \cong 0$. Δηλαδη οι δυο καμπυλες εχουν ιδια συμπεριφορα στη γειτονια του 0. Αυτο το βλεπουμε πιο καθαρα αν συγκρινουμε τις παραγωγους στο 0:

$$y_1'(0) = 1 = y_2'(0)$$

Δηλαδη οι δυο καμπυλες δεν συναντιωνται μονον, αλλα εχουν και κοινην εφαπτομενη, πραγμα που σημαινει οτι το σημειο 0 ειναι διπλο σημειο μηδενισμου της εξισωσης. Για $x \in (0, \pi/2)$ ειναι $y_2 = \tan x$ και επομενως

$$y_1' = 1/\cos^2 x > 1 = y_2'$$

άρα οι δυο καμπυλες δεν διασταυρωνονται ξανα σ' αυτην την περιοχη.

3. Να εντοπισθουν γραφικα ολες οι λυσεις της εξισωσης:

$$x/\ln|x| = x + 1$$

Λύση

Μετασχηματιζουμε την εξισωση στην μορφη:

$$\ln|x| = x/(x+1) = 1 - 1/(x+1)$$

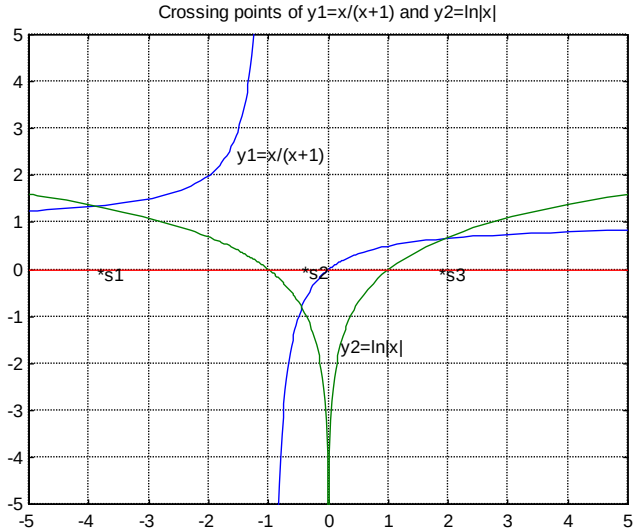
γιατι ειναι απλουστερο να σχεδιασουμε την $z = \ln|x|$ παρα την $x/\ln|x|$.

Η υπερβολη $z_2 = 1 - 1/(x+1)$, εχει για $x \to \pm\infty$ την ασυμπτωτο $y = 1$. Για $x < -1$ ανεβαινει απο το 1 στο $+\infty$ και μεταπηδα στο $-\infty$ στην θεση $x = -1$ για να ανεβει ξανά σταδιακα στο 1.

Η $z_1 = \ln|x|$ ειναι συμμετρικη ως προς τον καθετο αξονα συντεταγμενων και ξεκινα απο $-\infty$ για $x = 0$ για να καταληξει προς τα αριστερα και δεξια, δηλ. για $x \to \Upsilon\infty$ στο $+\infty$.

Ο αριστερος κλαδος της z_1 συναντεται τον αριστερο κλαδο της z_2 σε ενα σημειο $s_1 < -1$, καθως και τον δεξιο κλαδο της z_2 σε ενα σημειο $s_2 \in (-1, 0)$. Ο δεξιος κλαδος της z_1 συναντεται τον δεξιο κλαδο της z_2 σε ενα σημειο s_3 που μπορουμε να το προσδιορισουμε προσεγγιστικα αν δεχθουμε οτι εχει γινει ηδη $z_2 \cong 1$:

$\ln|s_3| \cong 1$ δινει $|s_3| \cong e \cong 2{,}7$.



Σχημα 2.4: $\ln|x| = 1 - 1/(x+1)$

4. Να εντοπισθουν γραφικα οι λυσεις της εξισωσης:

$$\ln(x + 1/x) = \sin x$$

Λύση

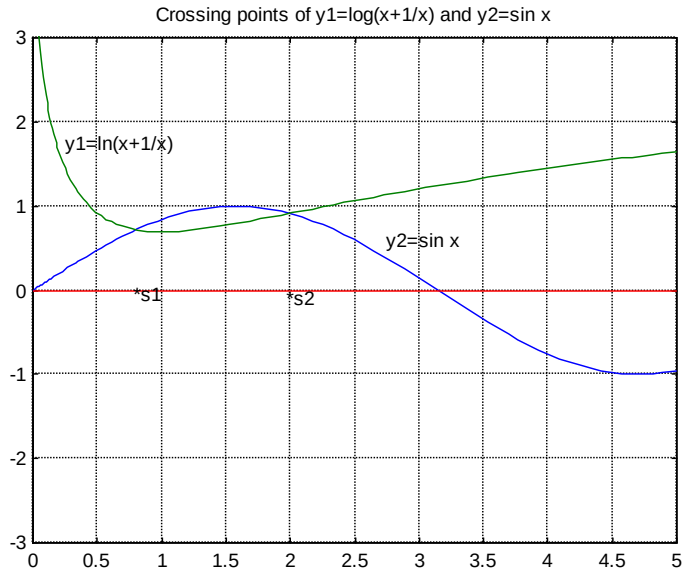
Η $y = \ln(x + 1/x)$ οριζεται μονο για $x > 0$. Για $x = 0$ υπερισχυει στο ορισμα

της ο ορος 1/x, και επομενως συμπεριφερεται οπως η ln(1/x) = -ln x. Δηλαδη στο x = 0 εχει την τιμη +∞ και κατεβαινει σταδιακα προς μικροτερες τιμες. Για μεγαλα x υπερισχυει στο ορισμα της ο ορος x, δηλ. ειναι: $y_1 \cong \ln x$ και τεινει στο +∞ για x→+∞. Το ελαχιστο της προκυπτει στην ιδια θεση με το ελαχιστο της x + 1/x, δηλ. στο σημειο x = 1 και ειναι ln2 = 0,69.

Η y_2 = sin x ειναι περιοδικη συναρτηση που κινειται συνεχως μεταξυ -1 και +1. Για $y_1 > 1$ (π.χ. για x σ e ≅ 2,7) αποκλειεται λοιπον να διασταυρωθει με την y_1 . Μεταξυ 0 και π ≅ 3,14 το ημιτονο εκιναει απο το μηδεν, φθανει στο 1 για x = π/2 ≅ 1,57 και γυριζει στο 0 για x = π. Οι δυο συναρτησεις ενδεχεται λοιπον να μην συναντιουνται καθολου, αφου η μια κατεβαινει αρχικα απο το απειρο για να ξαναανεβει σ' αυτο, ενω η αλλη δεν ξεπερνα ποτε το 1. Για να δουμε αν υπαρχουν σημεια διασταυρωσεως συγκρινουμε τις δυο συναρτησεις στο σημειο ελαχιστου της πρωτης:

$$y(1) = \ln 2 = 0{,}69 < 0{,}9 = \sin 1$$

Αυτο σημαινει οτι οι συναρτησεις πραγματι διασταυρωνονται, και μαλιστα σε δυο σημεια, ενα αριστερα και ενα δεξια απο το σημειο 1.



Σχημα 2.5: ln (x + 1/x) = sin x

5. Να προσδιορισθουν γραφικα οι λυσεις της εξισωσης:

$$\arctan x = \sin x$$

Λύση

Η συναρτηση y_1 = arctan x εχει για x→-∞ την ασυμπτωτο y = -π/2 και για για x→+∞ την ασυμπτωτο y = π/2. Στο σημειο x = 0 παιρνει ακριβως οπως και η y_2 = sin x την τιμη 0 περνωντας απο αρνητικες σε θετικες τιμες. Οι δυο συναρτησεις διασταυρωνονται επομενως σ'αυτο το σημειο και μαλιστα εχουν κοινη εφαπτομενη:

$$y_1'(0) = 1 = y_2'(0)$$

Αυτο σημαινει οτι το x = 0 ειναι τουλάχιστον διπλο σημειο μηδενισμου της εξισωσης. Αν όμως εξετάσουμε τις δεύτερες παραγώγους $y_1''=-2x/(1+x^2)^2$ και $y_2''=-\sin x$ βλέπουμε ότι είναι και $y_1''(0)=0= y_2''(0)$, δηλαδή ότι το σημείο x=0 είναι <u>τριπλό</u> σημείο μηδενισμού. Αυτο ειναι φαινομενικα και η μοναδικη της λυση, γιατι στην συνεχεια η y_1 τεινει προς π/ 2 ≅ 1,57 ενω η y_2 ποτε δεν ξεπερναει το 1. Αυτο ομως ειναι λαθος!

Για x ≅ 0, x ≠ 0, ειναι

$$y_1'(x) = 1/(1+x^2) \cong 1 - x^2 < 1 - x^2/2 \cong \cos x = y_2'(x)$$

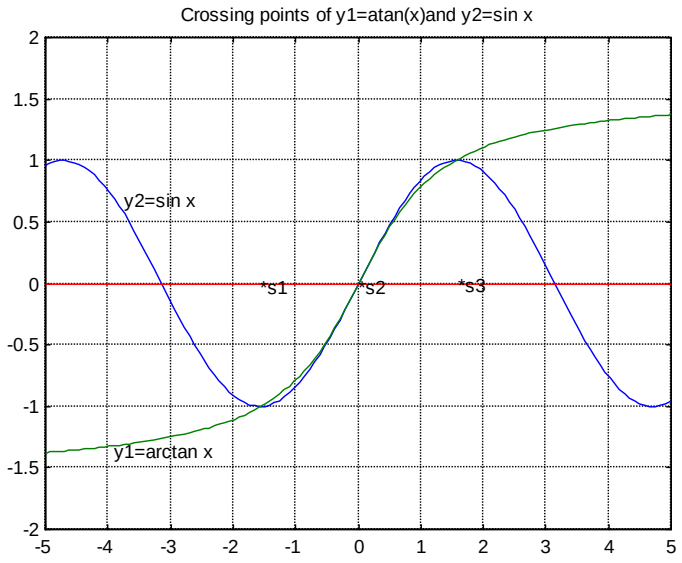
πραγμα που σημαινει οτι για θετικα x κοντα στο 0 ειναι

$$y_1(x) = \arctan x < \sin x = y_2(x)$$

ενω αργοτερα, π.χ. για x = 2 ειναι

$$y_1(2) = \arctan 2 \cong 1{,}1 > 0{,}9 \cong \sin 2 = y_2(2)$$

Δηλαδη στο διαστημα (0, 2) υπαρχει ενα ακομα σημειο μηδενισμου της εξισωσης αυτης και παρομοια υπαρχει ενα συμμετρικο του στο (-2, 0). Εδω φαινεται ποσο σημαντικη ειναι η προσεκτικη μελετη των καμπυλων και επαληθευεται το γερμανικο γνωμικο οτι ο διαβολος κρυβεται στην λεπτομερεια (Im Detail steckt der Teufel).



Σχημα 2.6: arctan x = sin x

6. Να προσδιορισθουν γραφικα οι λυσεις της εξισωσης:

$$\sin x = \frac{1}{x^2 - 1}$$

Λύση

Η μορφη της y_1 = sin x ειναι γνωστη. Γι'αυτο θα μελετησουμε την

$$y_2 = \frac{1}{x^2-1} = \frac{1}{2}\left(\frac{1}{x-1} - \frac{1}{x+1}\right)$$

Η συναρτηση αυτη ειναι αρτια (εχει για +x και -x τις ιδιες τιμες) και επομενως

ειναι συμμετρικη ως προς τον αξονα των y. Για x = -∞ εχει την τιμη 0 και αυξανεται σταδιακα μεχρι το ∞, οπου φθανει για x = -1. Για -1 < x < +1 ειναι αρνητικη και μαλιστα αμεσως μετα το -1 και αμεσως πριν απο το +1 εχει τιμη -∞. Το μεγιστο της σ' αυτην την περιοχη ειναι -1, για x = 0. Στο σημειο x = 1 μεταπηδα απο -∞ σε +∞ και τεινει παλι στο 0 για αυξανομενα x.

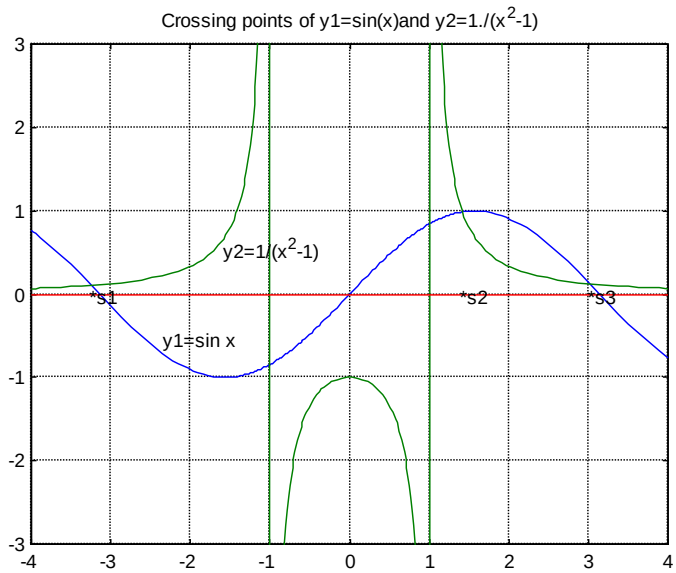
Στην περιοχη -1<x<1 η συναρτηση αυτη δεν μπορει να διασταυρωθει με το ημιτονο αφου ειναι μικροτερη απο -1 για $x \neq 0$.

Επειδη για $x \to \pm\infty$ τεινει στο 0, αναγκαστικα θα διασταυρωνεται με την $y_1 = \sin x$ κοντα στα σημεια $\pm k\pi$, $k \in \mathbf{N}$. Δηλαδη η εξισωση εχει τις λυσεις:

$$s_k \cong \pm k\pi,\ k \in \mathbf{N}$$

Εκτος απο αυτες, υπαρχει ακομα ενα σημειο διασταυρωσεως $s \in (1, \pi/2)$, γιατι $y_2(1+) = \infty > \sin 1 = y_1(1)$, ενω

$$y_1(\pi/2) = \sin(\pi/2) = 1 > y_2(\pi/2) \cong 0.68\ .$$



Σχημα 2.7: $\sin x = \dfrac{1}{x^2 - 1}$

2.2 Η Γενική Επαναληπτική Μέθοδος

Οπως ειπαμε στην αρχη του κεφαλαιου, ο βασικος τροπος λυσης της εξισωσης f(x)=0 ειναι η αναγωγη της σε μιαν μορφη x=φ(x) τετοια ωστε η επαναληπτικη διαδικασια $x_{k+1}=\varphi(x_k)$ να συγκλινει. Βασικο μας ερωτημα ειναι, δηλαδη: Ποτε συγκλινει η ακολουθία αυτή;

Ας υποθεσουμε οτι συγκλινει προς ενα οριο s:

$$s = \varphi(s) \qquad (2.10)$$

τοτε αφαιρωντας την σχέση αυτή κατα μελη απο την επαναληπτική ακολουθία

και εφαρμοζοντας το θεωρημα μεσης τιμης του Διαφορικου Λογισμου εχουμε

$$x_{k+1} - s = \varphi(x_k) - \varphi(s) = \varphi'(\xi_k)(x_k - s) \qquad (2.11)$$

οπου ξ_k ενα ενδιαμεσο σημειο μεταξυ x_k και s.

Επειδη $|x_{k+1} - s|$ ειναι η αποσταση του x_{k+1} απο το s, ενω $|x_k - s|$ ειναι η αποσταση του x_k απο το s, αν θελουμε καθε νεος ορος της ακολουθιας να πλησιαζει περισσοτερο το s, βλεπουμε οτι αρκει να ειναι:

$$|\varphi'(\xi)| < 1 \qquad (2.12)$$

μάλιστα η ταχύτητα σύγκλισης είναι τόσο μεγαλύτερη όσο μικρότερες τιμές έχει ο συντελεστής $|\varphi'(\xi_k)|$ που συνδέει τις δύο αποστάσεις.

Εφαρμογή:

Να υπολογισθουν προσεγγιστικα οι λυσεις της εξισωσης:

$$f(x) := \ln x - x + 2 = 0$$

με κατάλληλες επαναληπτικές διαδικασίες.

Λύση:

Απο το σχημα 2.1 προκυπτει, καθως ειδαμε, οτι η εξισωση εχει δυο λυσεις:

$$s_1 \in (0,1), \quad s_2 \cong 3$$

Για να εφαρμοσουμε μίαν επαναληπτικη μεθοδο μπορουμε να φερουμε την δοσμενη εξισωση στην μορφη:

$$x = \ln x + 2 := \varphi(x)$$

οποτε η επαναληπτικη ακολουθια εχει την μορφη (2.2). Η συγκλιση της ακολουθιας αυτης ειναι εξασφαλισμενη αν σε καποιο διαστημα γυρω απο το σταθερο σημειο $s = \varphi(s)$ η παραγωγος $\varphi'(x)$ ειναι απολυτως μικροτερη απο την μοναδα.

Επειδη για $x \cong 3$ ειναι

$$\varphi'(x) = 1/x \cong 1/3 < 1$$

ενω για $0 < x < 1$ ειναι $\varphi'(x) > 1$, συμπεραινουμε οτι η επαναληπτικη ακολουθια

$$x_{k+1} = \varphi(x_k) = \ln x_k + 2, \quad k = 0,1,2,.. \qquad (2.17)$$

ειναι καταλληλη μονο για τον προσδιορισμο του σημειου s_2 .

Για να προσδιορισουμε το s_1 μπορουμε να χρησιμοποιησουμε την ιδιοτητα οτι αν $x = \psi(y)$ ειναι η αντιστροφη συναρτηση της $y = \varphi(x)$ τοτε, οπως προκυπτει απο παραγωγιση των δυο μερων της ταυτοτητας $x \equiv \psi(\varphi(x))$, ειναι

$$1 = \frac{d\psi}{dy}\frac{d\varphi}{dx} \qquad (2.18)$$

και επομενως, αν η παραγωγος της φ ειναι απολυτα μεγαλυτερη απο την μοναδα, τοτε η παραγωγος της ψ ειναι απολυτα μικροτερη απο αυτην.

Αντιστρεφοντας την εξισωση:

$$x = \ln x + 2$$

εχουμε

$$x = e^{x-2} := \psi(x) \qquad (2.19)$$

οπου για $x \cong 0$ ειναι

$$\psi'(x) = e^{x-2} \cong e^{-2} < 1$$

Επομενως η επαναληπτικη ακολουθια

$$x_{k+1} = e^{x_k - 2} \qquad (2.20)$$

με $x_0 \in (0,1)$ θα συγκλινει στο s .
Με $x_0 = 3$ εχουμε απο την (2.17):

k	x_k	k	x_k	k	x_k	k	x_k
1	3.09861	5	3.1457022	9	3.1461882	13	3.146193169
2	3.13095	6	3.1460371	10	3.1461916	14	3.146193204
3	3.14134	7	3.1461436	11	3.1461927	15	3.146193215
4	3.14465	8	3.1461775	12	3.1461931	16	3.146193219

οπου οι οροι της ακολουθιας εχουν εν μερει στρογγυλευθει κατα την εκτυπωση, για να φαινεται πιο καθαρα η σταδιακη σταθεροποιηση ολο και περισσοτερων δεκαδικων ψηφιων, που ειναι η χαρακτηριστικη πρακτικη ενδειξη συγκλισης μιας ακολουθιας.

Παρατήρηση: Τι κάνει μία επαναληπτική ακολουθία όταν δεν συγκλίνει; Χαοτική Συμπεριφορά.

Ενα ενδιαφερον ερωτημα που προκυπτει εδω ειναι: Τι κανει μια επαναληπτικη ακολουθια οταν δεν συγκλινει;
Μια επιπολαιη απαντηση θα ηταν "αποκλινει". Ομως αυτο συμβαινει σε μερικες μονον περιπτωσεις. Σε πολλες αλλες περιπτωσεις η ακολουθια εχει απλως περισσοτερα απο ενα σημεια συσσωρευσεως, ενω υπαρχουν και ακολουθιες με απειρα σημεια συσσωρευσεως.
Παραδειγματος χαριν, αυτο συμβαινει με την λεγομενη "Λογιστικη Ακολουθια":

$$X_{n+1} = \lambda\, x_n\, (1 - x_n) \qquad (2.24)$$

(που περιγραφει μεταξυ αλλων και το πως εξελισσεται ο αριθμος αυτων που ασθενουν κατα την εξελιξη μιας επιδημιας) για ωρισμενες τιμες της παραμετρου λ. Αν, π.χ., ειναι λ = 4 τοτε η (2.24) ικανοποιειται απο x_n της μορφης:

$$X_n = c \sin^2(2^n a\pi) \qquad (2.25)$$

επειδη

$$\sin^2(2\beta) = 4 \sin^2(\beta)(1 - \sin^2(\beta)) \qquad (2.26)$$

οπου τα c και a προσδιοριζονται απο τους δυο πρωτους ορους

$$x_0 = c \sin^2(a\pi),\ x_1 = c \sin^2(2a\pi) \qquad (2.27)$$

Η ακολουθια (2.25) εχει ολα τα σημεια του διαστηματος [0,c] σαν σημεια συσσωρευσεως αν ο αριθμος a ειναι αρρητος.
Αυτο προκυπτει αμεσως απο την παρατηρηση οτι μονο το κλασματικο μερος της ακολουθιας $\theta_n = 2^n a$ επηρεαζει την τιμη του x_n , γιατι αφαιρεση ακεραιων πολλαπλασιων του π δεν μεταβαλλει την τιμη του ημιτονου στο τετραγωνο.
Αν τωρα θεωρησουμε οτι ο a ειναι γραμμενος σε δυαδικη μορφη:

$$a = <0.a_1 a_2 a_3 \ldots> := \sum_{k=1}^{\infty} a_k 2^{-k} \text{ με } a_k = 0 \text{ ή } 1 \qquad (2.28)$$

βλεπουμε οτι το θ_n προκυπτει απλα με μετατοπιση της υποδιαστολης κατα n θεσεις προς τα δεξια, οπου, καθως ειπαμε, το ακεραιο μερος του θ_n <u>δεν</u> επηρεαζει την τιμη του x_n . Ειναι δηλαδη:

$$x_n = c \sin^2 (<0,a_{n+1}\, a_{n+2} \ldots> \pi) \qquad (2.29)$$

Καθως βλεπουμε, η τιμη του x_n εξαρταται ουσιαστικα απο το ποια ψηφια βρισκονται μετα την n θεση στην δυαδικη παρασταση του αριθμου α. Αν ο α ειναι αρρητος, εχει απειρα μη περιοδικως επαναλαμβανομενα, οχι μονον δεκαδικα, αλλα και δυαδικα ψηφια. Ετσι για καθε n ο συντελεστης a_{n+1} και οι μετα απο αυτον μπορουν να εχουν οποιεσδηποτε τιμες. Ετσι και οι τιμες που παιρνει το ημιτονο στο τετραγωνο μεταπηδουν ακανονιστα στο διαστημα [0, 1].

Ενα αλλο χαρακτηριστικο της ακολουθιας (2.24) ειναι οτι ειναι εξαιρετικα ευπαθης σε αριθμητικα σφαλματα. Π.χ. αν αντι για α βαλουμε στην (2.25):

$$a^* = a + \pi/2^{k+1} \qquad (2.30)$$

τοτε η ακολουθια που προκυπτει εχει σαν k ορο:

$$x_k^* = c \sin^2 (2^k a^*\pi) = c \sin^2 (2^k a\pi + \pi/2) = c \cos^2 (2^k a\pi) = c - x_k$$

δηλαδη εναν αριθμο πολυ διαφορετικο απο τον x_k . Π.χ. για k = 9 προκυπτει οτι η μεταβολη του α κατα π/1024 = 0.003 εχει σαν συνεπεια οτι $x_9^*=c-x_9$.

Η συμπεριφορα αυτη ονομαζεται «χαοτική».

2.3 Μέθοδος Newton – Raphson ή Εφαπτομένης

Ανοικτο παραμενει ακομα το ερωτημα πως για τυχον σημειο μηδενισμου s μιας συναρτησης f, (f(s)=0) θα βρεθει συναρτηση φ(x) με s = φ(s) που να Εξασφαλίζει την σύγκλιση της επαναληπτικής ακολουθίας:

$$x_{k+1}=\varphi(x_k) \qquad (2.25)$$

στο s. Όπως είδαμε, αρκεί να είναι

$$|\varphi'(x)| \le L< 1 \qquad (2.26)$$

σε ενα διαστημα γυρω απο το s και μάλιστα η σύγκλιση είναι τόσο ταχύτερη όσο μικρότερες είναι οι τιμές του |φ'(x)| στην περιοχή αυτή. Για να βρουμε μιαν τετοια συναρτηση εξεταζουμε, λοιπόν, την κατηγορια των συναρτησεων:

$$\varphi(x) = x + f(x)g(x) \qquad (2.27)$$

που εχουν την ιδιοτητα οτι για f(x) = 0 ειναι φ(x) = x. Η g(x) είναι εδώ μία συνάρτηση που μπορούμε να επιλέξουμε αυθαίρετα. Για τούτο θα την προσδιορίσουμε έτσι ώστε να εξασφαλίσουμε την ισχύ της συνθήκης (2.26). Ο απλουστερος τροπος να το επιτύχουμε αυτό είναι το να απαιτήσουμε τον μηδενισμο της παραγωγου φ' στο s. Ετσι, αν η φ' ειναι συνεχης συναρτηση, τοτε προφανως γυρω απο το s θα εχει τιμες γυρω απο το μηδεν και επομενως πολύ μικρότερες του 1, πράγμα που εξασφαλίζει και την ταχεία σύγκλιση της ακολουθίας.

Απο την απαιτηση αυτη προκυπτει, εξ' αιτιας της σχεσης f(s)=0:

$$0 = \varphi'(s) = 1 + f(s)g'(s) + f'(s)g(s) = 1 + f'(s)g(s)$$

δηλαδη,

$$g(s) = -\frac{1}{f'(s)} \qquad (2.28)$$

Οποιαδηποτε συναρτηση g(x) ικανοποιει την συνθηκη (2.28) ειναι καταλληλη για την δημιουργια μιας επαναληπτικης μεθοδου μεσω των τυπων (2.27) και (2.25). Η απλουστερη τετοια συναρτηση ειναι προφανως εκείνη με:

$$g(x) = -\frac{1}{f'(x)} \tag{2.29}$$

για όλα τα x, οπότε προκύπτει η ακολουθία

$$x_{n+1} = \varphi(x_n) = x_n - \frac{f(x_n)}{f'(x_n)}, \quad n = 0,1,2,... \tag{2.30}$$

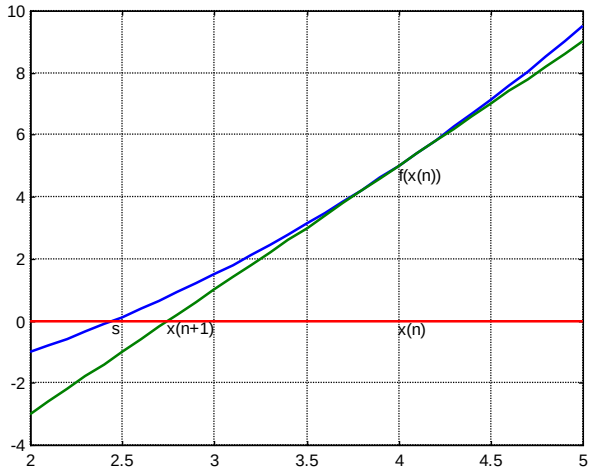
Η ακολουθία αυτη λεγεται ακολουθία Newton - Raphson και συγκλίνει εξαιρετικά γρήγορα στο σημείο μηδενισμού, s, αν ο αρχικός της όρος, x_0, επιλεγεί αρκετά κοντά σε αυτό.

Ενας αλλος τροπος για να παραχθει η (2.30) ειναι το να σχεδιασουμε την εφαπτομενη της καμπυλης y = f(x) στο τυχαιο σημειο x_n. Το σημειο οπου η εφαπτομενη αυτη τεμνει τον οριζοντιο αξονα ειναι το σημειο x_{n+1} και οπως φαινεται απο το σχημα 2.8 ειναι πολυ πλησιεστερα προς το s απ'οτι το x_n.

Συμφωνα με το σχημα, ειναι

$$f'(x_n) = \frac{f(x_n)}{x_n - x_{n+1}}$$

σχεση ισοδυναμη με την (2.30).



Σχημα 2.8: $f'(x_n) = \tan a = f(x_n) / (x_n - x_{n+1})$

Όπως είπαμε, η ακολουθια (2.30) δεν συγκλινει μονο οπωσδηποτε προς το s οταν το x ειναι αρκετα κοντα στο s, αλλα συγκλινει επι πλεον ταχυτατα. Ακριβέστερα, αυτό συμβαίνει μόνο όταν το s ειναι απλο σημειο μηδενισμου της f, δηλαδή όταν $f'(s) \neq 0$.

Για να διαπιστώσουμε την ταχύτητα σύγκλισης παρατηρούμε ότι, συμφωνα με τον τροπο κατασκευης της ακολουθιας, ειναι όχι μόνο s = φ(s), αλλά και φ'(s) = 0. Ετσι, χρησιμοποιωντας το αναπτυγμα Taylor της φ(x) περι το s εχουμε:

$$x_{n+1} - s = \varphi(x_n) - \varphi(s) = (x_n - s)\varphi'(s) + \frac{1}{2}(x_n - s)^2 \varphi''(\xi_n)$$

οπου ξ_n σημειο μεταξυ x_n και s. Επειδη ο πρωτος απο τους δυο αυτούς προσθετεους μηδενιζεται λογω μηδενισμου της παραγωγου, αν η φ″(x) είναι φραγμένη στην περιοχή του s, δηλαδή αν ισχυει:

$$|\varphi''(x)| \le 2M \text{ για } |x - s| \le \rho,$$

τότε προκυπτει οτι στο διαστημα αυτο ειναι

$$|x_{n+1} - s| \le M\,|x_n - s|^2 \qquad (2.31)$$

Λόγω του εκθέτη 2 στο δεύτερο μέρος της ανισότητας λέμε ότι η ακολουθια αυτή *συγκλίνει προς το s τετραγωνικά, ή με τάξη συγκλίσεως 2.*

Γενικα, λεμε οτι η ακολουθια συγκλινει στο s με ταξη συγκλισεως r, αν ισχυει:

$$|x_{n+1} - s| \le M\,|x_n - s|^r .$$

Η σημασία της ανισοτητας (2.31) φαίνεται καθαρά αν χρησιμοποιήσουμε την συντόμευση

$$d_k := M|x_k - s|$$

για να την γράψουμε στην μορφή

$$d_{n+1} \le d_n^2 \quad , n = 0,1,2,...$$

Τότε βλέπουμε ότι αν, για παράδειγμα, είναι $d_0 \le 10^{-1}$, θα ειναι στην συνέχεια $d_2 \le 10^{-2}$, $d_3 \le 10^{-4}$, $d_4 \le 10^{-8}$ κοκ.

Δηλαδη, $d_k \le 10^{-2^k}$. Αυτο σημαινει οτι το πληθος των μηδενικων στην δεκαδικη μορφη καθε νεου ορου της ακολουθιας $\{d_k\}$ είναι διπλασιο απο αυτο του προηγουμενου του. Αυτο σημαινει όμως και οτι σε καθε νεαν επαναληψη διπλασιαζεται περιπου ο αριθμος των σωστων δεκαδικων ψηφιων της ακολουθίας $\{x_n\}$. Αν, επομενως, αρχισουμε με ενα σημειο αρκετα κοντινο στο s, τοτε κατα την τριτη η τεταρτη επαναληψη το σφαλμα ειναι μικροτερο απο την ακριβεια που παρεχει ο υπολογιστης και δεν χρειαζεται να συνεχισουμε τις επαναληψεις.

Η προϋπόθεση f′(s) ≠ 0, είναι κατ' αρχήν απαραίτητη, γιατί αλλοιώς το κλασμα στο δευτερο μερος της (2.30) έχει απροσδιοριστη τιμή 0/0 οταν το x_n τεινει στο s. Τι κανει ομως η ακολουθια (2.30) οταν το s ειναι πολλαπλο σημειο μηδενισμου της f δηλαδή όταν f′(s)=0;

Οπως φαίνεται από την πρώτη σχέση (2.9), οταν το s ειναι σημειο μηδενισμου με πολλαπλοτητα m, τοτε η f μπορει να γραφτει στην μορφη

$$f(x) = (x - s)^m g(x), \text{ με } g(s) \ne 0 \qquad (2.32)$$

Απο την σχεση αυτην προκυπτει οτι

$$x_{n+1} - s = x_n - s - \frac{f(x_n)}{f'(x_n)} = x_n - s - \frac{(x_n - s)g(x_n)}{mg(x_n) + (x_n - s)g'(x_n)}$$

αρα

$$\lim_{x_n \to s} \frac{x_{n+1} - s}{x_n - s} = 1 - \frac{1}{m} \qquad (2.33)$$

Απο την σχεση αυτην προκυπτει οτι για αρχικη τιμη αρκετα κοντα στο s θα υπαρχει ακομα συγκλιση, αλλα πολυ βραδεια, γιατι εδω το άνω φράγμα της |φ'(x)| ειναι L = 1 - 1/m, που για μεγαλη πολλαπλοτητα m πλησιαζει πολυ την μοναδα.

Παρατηρήσεις: Επιτάχυνση σύγκλισης και Εξασφάλιση Σύγκλισης

1. Στην περιπτωση που το s ειναι πολλαπλο σημειο μηδενισμου της f, για να εξασφαλισουμε τετραγωνικη συγκλιση, αρκει να εφαρμοσουμε την μεθοδο Newton - Raphson, οχι στην συναρτηση f, αλλα στην g:=f/f', που εχει τα ιδια σημεια μηδενισμου, αλλα μονον απλα (βλέπε σελ.15).

2. Συνεπεια της τετραγωνικης συγκλισης της μεθοδου Newton - Raphson ειναι οτι κατα τον προσδιορισμο σημειων μηδενισμου δεν χρειαζεται να ανησυχουμε πολυ για το αν η αρχικη τιμη ειναι αρκετα κοντα σε καποιαν λυση για να εξασφαλισει συγκλιση. Ακομα και αν ξεκινησει κανεις την επαναληπτικη διαδικασια απο μακρυα, συνηθως η ακολουθια, εφ' οσον δεν συγκλινει αμεσα, μεταπηδα ακανονιστα σε διαφορες τιμες μεχρις οτου βρεθει τυχαια αρκετα κοντα σε μιαν λυση. Τοτε αρχιζει να συγκλινει ταχυτατα προς την λυση και μεσα σε τεσσερις ως πεντε επαναληψεις σταθεροποιειται εντελως. Αυτο κανει μερικους να αδιαφορουν για τον εντοπισμο των λυσεων, πραγμα που ειναι εσφαλμενο, γιατι μια τυχαια αναζητηση λυσης μπορει να μας οδηγησει σε λυση που δεν ανταποκρινεται στο φυσικο η τεχνικο προβλημα που αντιμετωπιζουμε.

Παράδειγμα Εφαρμογής της Μεθόδου Neuton-Raphson:

Να προσδιορισθουν με την μεθοδο Newton - Raphson οι λυσεις της εξισωσης

$$f(x):= \ln x - x + 2 = 0.$$

Λύση

Η ακολουθια Newton - Raphson εχει εδω την μορφη:

$$x_{n+1} = x_n - \frac{\ln x_n - x_n + 2}{\frac{1}{x_n} - 1}$$

Οπως ηδη ξερουμε απο τον γραφικο εντοπισμο των λυσεων, υπαρχουν δυο λυσεις, μια στο διαστημα (0,1) και μια κοντα στην τιμη 3. Το ποσο αστραπιαια συγκλινει αυτη η επαναληπτικη ακολουθια φαινεται αν ξεκινησουμε με ιδιες αρχικες τιμες, οπως και κατα την εφαρμογη των επαναληπτικων μεθοδων που χρησιμοποιήσαμε αρχικά.

Παιρνοντας μιαν φορα σαν αρχικη τιμη το 0.1 και μιαν φορα το 3 εχουμε τις εξης ακολουθιες, που σταθεροποιουνται εντελως μετα την πεμτη επαναληψη:

0.1 | 0.144731677 | 0.1578643558 | 0.1585923416 | 0.1585943395 | 0.1585943396

3 | 3.147918433 | 3.146193441 | 3.146193221

Και στις δυο περιπτωσεις ειναι δυνατον να παρατηρησουμε το φαινομενο του διπλασιασμου περιπου των σωστων δεκαδικων ψηφιων μετα απο καθε επαναληψη.

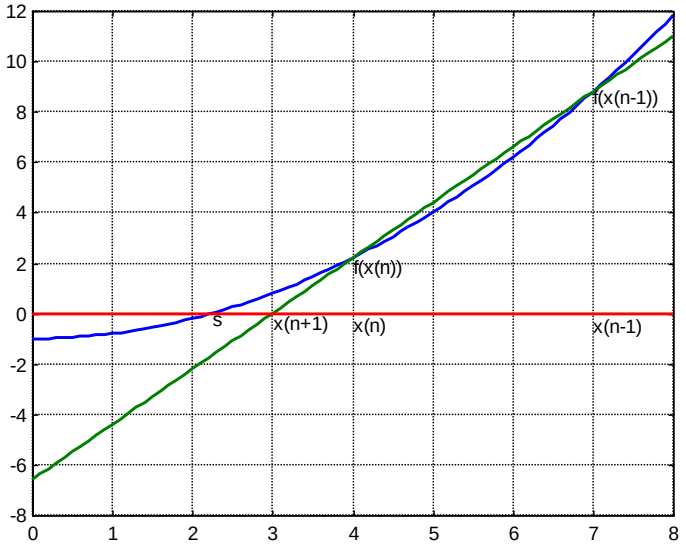
Αν δοκιμασουμε με διαφορες αρχικες τιμες, βλεπουμε οτι για $0 < x \leq 0.3$ η ακολουθια συγκλινει στο πρωτο σημειο μηδενισμου, ενω για $x > 1$ συγκλινει παντα στο δευτερο. Για $0.4 \leq x < 1$ η ακολουθια εχει μιγαδικους ορους (δεν αποκλειεται ομως να συγκλινει αν κανουμε χρηση μιγαδικης αριθμητικης στον Η/Υ).

Το ποσο ραγδαια ειναι η συγκλιση εστω και αν ξεκινησουμε απο μακρυα, φαινεται αν βαλουμε π.χ. αρχικη τιμη $x = 10000$. Τοτε προκυπτει η ακολουθια:

10000 | 10.21135161 | 3.684305505 | 3.162435176 | 3.146212575 | 3.146193221.

2.4 Μέθοδος Τέμνουσας ή Regula Falsi

Η μεθοδος τεμνουσας προκυπτει σχεδιαστικα οπως περιπου και η μεθοδος Newton - Raphson, αντι ομως να φερουμε την εφαπτομενη της $y = f(x)$ στην τελευταια θεση, x_n, παιρνουμε την τεμνουσα που περναει απο τα δυο τελευταια σημεια (x_{n-1}, f_{n-1}), (x_n, f_n) (οπου χρησιμοποιουμε την συντομογραφια $f_n := f(x_n)$).



Σχημα 2.9: Η μεθοδος Τεμνουσας

Το σημειο οπου η ευθεια αυτη τεμνει τον οριζοντιο αξονα ειναι το καινουργιο σημειο x_{n+1} και, οπως φαινεται απο το σχημα, ειναι κατα πολυ πλησιεστερο στο s απο οτι το x_n .

Αν ονομάσουμε φ την γωνία μεταξύ της τέμνουσας και του οριζόντιου άξονα βλέπουμε ότι είναι κοινή γωνία δύο ορθογωνίων τριγώνων με μήκος πλευρών αφ' ενός f_n, x_n-x_{n+1} και αφ' ετέρου f_{n-1}, x_{n-1}-x_{n+1}. Άρα ισχύει:

$$\tan\varphi = \frac{f_n}{x_n - x_{n+1}} = \frac{f_{n-1}}{x_{n-1} - x_{n+1}}$$

Επιλύοντας την τελευταία ισότητα ως προς x_{n+1} έχουμε την σχέση:

$$x_{n+1} = x_n - f_n \frac{x_n - x_{n-1}}{f_n - f_{n-1}} \qquad (2.34)$$

με n=1,2,3,….

Ενας αλλος τροπος παραγωγης της μεθοδου τεμνουσας ειναι να αντικαταστησουμε στον τυπο (2.30) την παραγωγο $f'(x_n)$ με το πηλικο διαφορων

$$\frac{f_n - f_{n-1}}{x_n - x_{n-1}}$$

που την προσεγγιζει.

Για την παραγωγη μιας ακολουθιας με τον τυπο αυτον πρεπει να χρησιμοποιησουμε δυο αρχικες τιμες, x_0, και x_1 που απλα τις παιρνουμε απο την περιοχη της λυσης s που προσπαθουμε να προσδιορισουμε.

Βασικο του πλεονεκτημα ειναι οτι δεν χρειαζεται ο υπολογισμος της παραγώγου f'.

Η ταξη συγκλισεως αυτης της μεθοδου ειναι, r = (1 + $\sqrt{5}$)/2 = 1.618, (βλέπε Κιουστελίδη:«Αριθμητική Ανάλυση», 1996, σελ.51) και επομενως, χονδρικα σε καθε επαναληψη αυξανονται κατα 61,8 % τα σωστα δεκαδικα ψηφια.

Συνολικα παρατηρουμε οτι, ενω η ταξη συγκλισεως ειναι αρκετα κοντα στην ταξη, 2, της μεθοδου Newton - Raphson, ο υπολογιστικος ογκος ειναι εδω περιπου ο μισος, γιατι σε καθε νεα επαναληψη αρκει να υπολογισουμε την τιμη f_n, αφου η f_{n-1} ειναι ηδη γνωστη. Στην μεθοδο Newton - Raphson σε καθε νεα επαναληψη υπολογιζουμε, οχι μονον την f_n , αλλα και την $f_n' := f'(x_n)$. Ετσι, ενω η μεθοδος τεμνουσας χρειαζεται περισσοτερες επαναληψεις για να επιτυχει την ιδια ακριβεια, ο συνολικος ογκος πραξεων και ο χρονος εκτελεσης ειναι μικροτεροι.

Το μονο που πρεπει να προσεξουμε κατα την εφαρμογη του τυπου (2.34) ειναι οτι η ακολουθια αυτη δεν συγκλινει οταν το σημειο s ειναι πολλαπλο σημειο μηδενισμου. Σε μιαν τετοια περιπτωση αρκει αντι της f να θεωρησουμε την συναρτηση g:= f/f' που εχει τα ιδια σημεια μηδενισμου οπως η f αλλα απλα!

Αξιο προσοχης ειναι επισης οτι η αναγκη να παρουμε δυο αρχικες τιμες, αντι να δυσκολεψει την εφαρμογη της μεθοδου, την διευκολυνει, γιατι η συγκλιση ειναι εξασφαλισμενη αν εστω και μονο μια απο αυτες τις τιμες ειναι αρκετα κοντα στην λυση. Προσεξτε τι τιμη x προκυπτει απο την (2.34) αν συμπτωματικα βαλλουμε x = s, ασχετως της τιμης του x .

Παράδειγμα Εφαρμογής της Μεθόδου Τέμνουσας

Να υπολογισθουν με την μεθοδο τεμνουσας οι λυσεις της

$$f(x) := \ln x - x + 2 = 0$$

Λύση

Συμφωνα με τον γραφικο εντοπισμο η εξισωση εχει δυο λυσεις: $s_1 \cong 0$ και $s_2 \cong 3$. Με αρχικες τιμες $x_0 = 0.1$, $x_1 = 0.2$ εχουμε λοιπον:

0.1 | 0.2 || 0.1678727146 | 0.1571823361 | 0.1586432619 |0.1585945986 | 0.1585943395 | 0.1585943396

ενω με αρχικες τιμες $x_0 = 3$, $x_1 = 3.1$ προκυπτει:

3 | 3.1 || 3.146722257 | 3.146191387 | 3.146193221.

Παρατήρηση: Η Μέθοδος Steffensen

Μια μεθοδος προσδιορισμου των σημειων μηδενισμου της συναρτησης f(x) που συγκλινει τετραγωνικα αν η f εχει συνεχη δευτερη παραγωγο χωρις να κανει χρηση της παραγωγου f'(x) ειναι η ακολουθη μεθοδος του Steffensen:

$$x_{n+1} = x_n - \frac{f_n^2(x_n)}{f(x_n + f(x_n)) - f(x_n)}$$

που προκυπτει αν προσεγγισουμε την παραγωγο f'(x) στην γειτονια του σημειου μηδενισμου s με [f(x+f(x))-f(x)]/f(x). Η μεθοδος αυτη ειναι ισοδυναμη με την μεθοδο Newton - Raphson, γιατι εχει την ιδια ταξη συγκλισης και τον ιδιο υπολογιστικο ογκο (δυο υπολογισμους συναρτησεων ανα βημα), εχει ομως το πλεονεκτημα οτι δεν χρειαζεται να υπολογισουμε την παραγωγο, ουτε να προγραμματισουμε τον τροπο υπολογισμου της.

2.5 Έκτίμηση Σφάλματος

Μεθοδοι οπως η Μεθοδος Newton - Raphson η η Μεθοδος Τεμνουσας συνηθως δεν χρειαζονται (εφ' οσον συγκλινουν) περισσοτερες απο 4 με 5 επαναληψεις έως ότου σταθεροποιηθούν τα πρώτα 8 ή 9 δεκαδικά ψηφία μετά την υποδιαστολή. Μπαίνει όμως το ερώτημα πώς μπορούμε να βεβαιωθούμε για την ακρίβεια της τελικής προσέγγισης. Έναν απλό τρόπο μας δίνει το θεώρημα του Bolzano (θεώρημα 2.1). Εδώ θα το διατυπώσουμε σε κάπως διαφορετική μορφή που διευκολύνει την εφαρμογή του για την εκτίμηση σφάλματος:

Θεώρημα 2.4 (Bolzano):

Αν x ειναι προσεγγιση ενος σημειου μηδενισμου, s, μιας συναρτησης f που ειναι συνεχης σε ενα διαστημα γυρω απο το s και αν για καταλληλο θετικο ε ισχυει:

$$f(x - \varepsilon)f(x + \varepsilon) < 0$$

τοτε το s βρισκεται μεσα στο διαστημα (x - ε, x + ε) και επομενως ειναι:

$$|x - s| < \varepsilon.$$

Για την εφαρμογη αυτου του θεωρηματος για εντοπισμο του μεγεθους του σφαλματος αντιμετωπιζουμε κατ' αρχην την δυσκολια του πώς πρεπει να επιλεξουμε την τιμη ε. Στην πραγματικοτητα ομως, αυτο ειναι πολυ απλο. Αν κατα τις διαδοχικες επαναληψεις μιας, οποιασδηποτε, επαναληπτικης μεθοδου εχουν φαινομενικα σταθεροποιηθει n δεκαδικα ψηφια μετα την υποδιαστολη, τοτε δοκιμαζουμε τις τιμες ε_n , ή ε_{n-1}, οπου

$$\varepsilon_k = 10^{-k}$$

Μπορουμε ομως να απαλλαγουμε απο τα διλημματα οποιασδηποτε επιλογης του ε με το να βαλουμε τον Η/Υ να δοκιμασει ολες τις τιμες k (k = 1,2,3,...) μεχρι τον μεγιστο αριθμο δεκαδικων ψηφιων που διατιθενται για την

απομνημονευση ενος αριθμου στην μνημη του. Η ακριβεια υπολογισμου των τιμων της f δεν ενδιαφερει εδω, γιατι μας αρκει ο εντοπισμος της εναλλαγης προσημου.

Μοναδικη περιπτωση οπου μπορει να αποτυχει αυτη η μεθοδος εντοπισμου του σφαλματος ειναι το να ειναι το s σημειο μηδενισμου αρτιας ταξης, οποτε η f δεν εχει αντιθετα προσημα στις δυο πλευρες του s. Η περιπτωση ομως αυτη μπορει να γινει ευκολα αντιληπτη απο το γραφημα της f, γιατι τοτε η καμπυλη πλησιαζει τον οριζοντιο αξονα και απομακρυνεται παλι προς την ιδια πλευρα του, χωρις να τον κοψει. Σε μιαν τετοια περιπτωση, δεν εχουμε παρα να εξετασουμε αντι της f, την $g(x):= f(x)/f'(x)$, που θα εμφανιζει οπωσδηποτε εναλλαγες προσημου, αφου εχει τα ιδια σημεια μηδενισμου με την f, αλλα μονον απλα.

Στην περιπτωση που κανουμε τον εντοπισμο σημειων μηδενισμου της $f=f_1-f_2$ με εντοπισμο των σημειων διασταυρωσεως των f_1 και f_2 , τα αρτιας πολλαπλοτητας σημεια μηδενισμου χαρακτηριζονται απο το οτι δεν ειναι γνησια σημεια διασταυρωσεως αλλα μονο σημεια επαφης των δυο καμπυλων. Δηλαδη οι γραφικές παραστάσεις των f_1 και f_2 πλησιαζουν μεταξυ τους και στην συνεχεια απομακρυνονται η καθε μια προς την κατευθυνση απο οπου ηρθε.

Παράδειγμα Προσδιορισμού του Σφάλματος:

Με τι σφαλμα προσεγγιζει την λυση s της εξισωσης

$$f(x):= \ln x - x + 2 = 0$$

η τιμη $x^* = x_{15} = 3.146193215$

που πήραμε με την επαναληπτικη μεθοδο $x_{k+1}=\ln x_k+2$;

Απάντηση

Επειδη τα x_{15} και x_{14} συμπιπτουν στις πρωτες 7 δεκαδικες θεσεις μετα την υποδιαστολη, δοκιμαζουμε την τιμη $\varepsilon_7 = 10^{-7}$:

$$f(x^* - \varepsilon_7) = 7.2\ 10^{-8}\ ,\ \text{ενω}\ \ f(x^* + \varepsilon_7) = -6.44\ 10^{-8}$$

Αρα πραγματι ειναι

$$|x^* - s| < 10^{-7}$$

Αν τωρα δοκιμασουμε και την τιμη $\varepsilon_8 = 10^{-8}$ βρισκουμε

$$f(x^* - \varepsilon_8) = 1{,}065\ 10^{-8}\ ,\ \text{και}\ f(x^* + \varepsilon_8) = -2{,}99\ 10^{-8}$$

Δηλαδη,

$$|x^* - s| < 10^{-8}\ .$$

Η δοκιμη με ε_9 δεν δινει πια εναλλαγη προσημου. Αρα το σφαλμα ειναι της ταξης του ε_8 .

Άσκηση:

Η εξισωση

$$f(x):= \cos 2x + 4/3*\cos x + 11/9=0$$

εχει το διπλό σημείο μηδενισμού $s \cong 1.9106$. Δοκιμαστε τον εντοπισμο του σφαλματος αυτής της προσέγγισης (Η ακριβής τιμή της διπλής ρίζας είναι:

2arctan√2)[2].

Ασκήσεις και Εφαρμογές

Οι ακολουθες εξισωσεις 1 έως 4 προεκυψαν κατα την λυση φυσικων και τεχνικων προβληματων. Να γινει γραφικος εντοπισμος ολων τους των λυσεων και να υπολογισθει αριθμητικα η απολυτως μικροτερη.

1. tan(b rλ) = - rλ (Κατανομη θερμοκρασιας κατα μηκος μιας ραβδου. Βλ. N.Dorny: A Vector Space Approach to Models and Optimization σελ 94, 173).

2. 1 + cos λ cosh λ = 0 (Φυσικες συχνοτητες ταλαντωσης ενος προβολου. Βλ. Pipes - Harvill: Applied Mathematics for Engineers and Physicists, σελ. 534).

3. $\cosh^2 aL + \cos^2 aL = 0$ (Ταλαντωσεις Δοκου. Βλ. Schwarz - Friedland, σελ. 466)

4. tanh λ = - tanλ (Ευσταθεια Συστηματος Ελεγχου Θερμοκρασιων. Βλ. Schwarz - Friedland σελ. 474).

Για παραδειγματα μη γραμμικων εξισωσεων που προκυπτουν κατα την μελετη γραμμων μεταφορας ηλεκτρικου ρευματος βλεπε Pipes - Harvill σελ. 287-289, και Schwarz - Friedland σελ.448).

5. Για ποιες τιμές της παραμέτρου a έχει λύσεις η εξίσωση: lnx –x +a=0 και πόσες; Υπάρχουν μιγαδικές λύσεις εκεί που δεν υπάρχουν πραγματικές;

6. Έχει η εξίσωση $x^y=y^x$ άλλες λύσεις εκτός από την προφανή: x=y;

Υπόδειξη: Η εξίσωση αυτή ισοδυναμεί με lnx/x=lny/y. Το ερώτημα είναι, επομένως ισοδύναμο με το πόσες διαφορετικές λύσεις έχει η εξίσωση: lnt/t=a για διάφορες τιμές του a.

7. Αν $f(t)=t^a/(t^b+c)$, με a,b,c>0 υπάρχει άλλη λύση της εξίσωσης f(x)=f(y) εκτός από την προφανή, x=y; Εξετάστε ειδικότερα την περίπτωση a=c=1, b=2.

Παρατήρηση στις 4/2/2017:

Εδώ αρκεί να εξετάσουμε την μονότονη ή μη συμπεριφορά της f(t). Αν η f(t) π.χ. πρώτα αυξάνεται και μετά μειούται, τότε εκατέρωθεν της θέσησ s όπου υπάρχει μέγιστο θα υπάρχουν θέσεις x, y με x<s<y με ίδιες τιμές. Πρέπει, δηλαδή να εξετάσουμε που μηδενίζεται η f'(t), και αν αλλάζει πρόσημο περνώντας από το σημείο s μηδενισμού της.

---

[2] Γενικά οι εξισώσεις: $\cos 2t+4q\cos t+1+2q^2=0$ έχουν τις διπλές ρίζες:

$$t = \pm 2\arctan\sqrt{\frac{1+q}{1-q}}$$

3. ΠΟΛΥΩΝΥΜΙΚΕΣ ΕΞΙΣΩΣΕΙΣ

3.1 Εντοπισμός Ριζών Πολυωνύμου

Οι πολυωνυμικες εξισωσεις εχουν ιδιομορφιες που επιτρεπουν η και επιβαλλουν την χρηση εν μερει ειδικων μεθοδων για την επιλυση τους.

Κατ' αρχην πρεπει να παρατηρηθει οτι πολυ συχνα μας ενδιαφερουν και οι μιγαδικες ριζες ενος πολυωνυμου, τις οποιες ο γραφικος εντοπισμος δεν εντοπιζει, ενω και ο γραφικος εντοπισμος πραγματικων ριζων ειναι πολυπλοκος αν προκειται για πολυωνυμο με πανω απο τεσσερις ως πεντε ορους. Για τουτο θα δοθουν εδω υπολογιστικοι τροποι εντοπισμου των ριζων.

Θεώρημα 3.1

Ολες οι ριζες του πολυωνυμου:

$$P_n(z) := z^n + a_1 z^{n-1} + a_2 z^{n-2} + \ldots + a_{n-1} z + a_n \qquad (3.1)$$

βρισκονται μεσα στους δισκους του μιγαδικού επιπέδου (ή τα αντίστοιχα διαστήματα του πραγματικού άξονα):

a) $$|z| \le 1 + \max_{k=1(1)n} |a_k| \qquad (3.2)$$

b) $$|z| \le 2 \max_{k=1(1)n} |a_n|^{1/k} \qquad (3.3)$$

Απόδειξη

a) Αν ολες οι ριζες του πολυωνυμου εχουν μετρο μικροτερο απο την μοναδα, τοτε η (3.2) ισχυει. Εστω, λοιπον, s μια ριζα με μετρο μεγαλυτερο απο την μοναδα.

Απο την ισοτητα $P_n(s) = 0$ προκυπτει τοτε με διαιρεση με s^{n-1}:

$$s = -a_1 - \frac{a_2}{s} - \ldots - \frac{a_{n-1}}{s^{n-2}} - \frac{a_n}{s^{n-1}}$$

και με $A := \max_{k=1(1)n} |a_k|$

$$|s| \le A\left(1 + \frac{1}{|s|} + \ldots + \frac{1}{|s^{n-1}|}\right) < A \frac{1}{1 - \frac{1}{|s|}}$$

Η σχεση αυτη ειναι ισοδυναμη με την (3.2)

b) Η ανισότητα

$$|a_k| < |s/2|^k$$

είναι ισοδύναμη με την

$$|s|^n/2^k > |a_k||s|^{n-k}$$

Αν η ανισότητα αυτή ίσχυε για κάθε τιμή του k, (k=1,2,…,n) τότε αθροίζοντας όλες αυτές τις ανισότητες θα είχαμε:

$$\left(1 - \frac{1}{2^{n+1}}\right)|s|^n > \sum_{k=1}^{n} |a_k||s|^{n-k} \ge \left|\sum_{k=1}^{n} a_k s^{n-k}\right| = |s|^n$$

όπου η τελευταία ισότητα προκύπτει από την $P_n(s)=0$ αν την γράψουμε στην μορφή:

$$\sum_{k=1}^{n} a_k s^k = -s^n$$

και πάρουμε την απόλυτη τιμή (ή το μέτρο) και στα δύο μέρη της. Η παραπάνω ανισότητα είναι όμως αδύνατη, αφού ο συντελεστής στην αριστερή πλευρά της είναι μικρότερος από 1, ενώ αυτός στην δεξιά πλευρά είναι 1.

Επομένως, για έναν τουλάχιστον δείκτη, ν, θα είναι $|a_\nu|\sigma|s/2|^\nu$, δηλαδή:

$$|s|\chi 2(|a_\nu|)^{1/\nu}\chi 2\max_{k=1(1)n}(|a_k|)^{1/k}$$

Παράδειγμα: Φράγμα των Ριζών Πολυωνύμου

Να βρεθεί άνω φράγμα των ριζών του πολυωνύμου:

$$P_5(x) = 2x^5 - 7x^4 + 15x^3 - 30x^2 + 20x - 10$$

Λύση

Γιά να εφαρμόσουμε το θεώρημα 3.1 πρέπει ο συντελεστής της μεγαλυτέρας δυνάμεως να είναι 1. Γι' αυτό εφαρμόζουμε το θεώρημα στο πολυώνυμο:

$$P_5(x)/2 = x^5 - 3.5x^4 + 7.5x^3 - 15x + 10x - 5$$

που έχει τις ίδιες ρίζες. Με

$$a_1 = -3.5 \quad a_2 = 7.5 \quad a_3 = -15 \quad a_4 = 10 \quad a_5 = -5$$

εχουμε για το μετρο (ή την απολυτη τιμη) οποιασδηποτε ριζας s του πολυωνυμου:

$$|s| \chi\ 1 + \max |a_\kappa| = 16$$

$$|s| \chi\ 2 \max |a_\kappa|^{1/\kappa} = 2 \max \{ 3.5,\ (7.5)^{1/2}, 15^{1/3}, 10^{1/4}, 5^{1/5} \}$$

$$= 2 \max \{3.5, 2.74, 2.47, 1.78, 1.38\} = 7$$

Για να βρουμε τις ριζες πολυωνυμικων εξισωσεων μπορουμε κατ' αρχην να εφαρμοσουμε ειτε την μεθοδο Newton - Raphson, ειτε την μεθοδο τεμνουσας οπως ακριβως και για τυχουσες συναρτησεις. Ο υπολογισμος των τιμων του πολυωνυμου και των παραγωγων του διευκολυνεται στην περιπτωση αυτην απο τις ακολουθες παρατηρησεις:

Αν

$$P_n(x) := a_0 x^n + a_1 x^{n-1} + a_2 x^{n-2} + \ldots + a_{n-1}x + a_n \qquad (3.11)$$

τοτε η τιμη του πολυωνυμου στο τυχον σημειο t μπορει να υπολογισθει μονο με n πολλαπλασιασμους με τον ακολουθο τροπο:

$$P_n(t) = a_n + t\,(a_{n-1} + t\,(a_{n-2} + \ldots + t\,(a_0 t + a_1)..))$$

Αν δηλαδη δημιουργησουμε την ακολουθια:

$$b_0 = a_0,\ b_1 = tb_0 + a_1, \ldots, b_n = tb_{n-1} + a_n$$

δηλαδη

$$b_0 = a_0,\ b_k = tb_{k-1} + a_k,\ k = 1(1)n \qquad (3.12)$$

τοτε ειναι

$$P_n(t) = b_n \qquad (3.13)$$

Η ακολουθια (3.12) λεγεται ακολουθια του Horner.

Θεώρημα 3.3

Οι οροι της ακολουθιας του Horner (3.12) ειναι οι συντελεστες

του πηλικου:

$$Q(x) = b_0x^{n-1} + b_1x^{n-2} +...+ b_{n-1} \qquad (3.14)$$

της διαιρεσης του $P_n(x)$ δια $x - t$, ενω b_n ειναι το υπολοιπο:

$$P_n(x) = Q(x)\,(x - t) + R,\ R = b_n = P_n(t) \qquad (3.15)$$

Επισης ισχυει

$$Q(t) = P_n'(t) \qquad (3.16)$$

πραγμα που επιτρεπει τον υπολογισμο αυτης της παραγωγου με δημιουργια της ακολουθιας Horner για το πολυωνυμο Q(x).

Απόδειξη

Η ταυτοτητα (3.15) αποδεικνυεται με εκτελεση των πραξεων στο δευτερο μερος της και χρηση των σχεσεων (3.12). Η (3.16) προκυπτει αν διαφορισουμε την ταυτοτητα (3.15) και βαλουμε οπου x το t.

Σχηματικα ο υπολογισμος της ακολουθιας Horner γινεται με τον ακολουθο τροπο:

	a_0	a_1	a_2	...	a_{n-1}	a_n
t	0	tb_0	tb_1	...	tb_{n-2}	tb_{n-1}
+	b_0	b_1	b_2	...	b_{n-1}	b_n

Η διαταξη αυτη λεγεται σχημα του Horner.

Μειονεκτημα της απ' ευθειας εφαρμογης π.χ. της μεθοδου Newton - Raphson ειναι οτι για τον προσδιορισμο μιγαδικων ριζων οι πραξεις πρεπει να γινουν με μιγαδικες αρχικες τιμες και να γινει στον υπολογιστη χρήση μιγαδικης αριθμητικης. Ο λογος γι' αυτο ειναι οτι αν παρουμε πραγματικη αρχικη τιμη η ακολουθια που θα προκυψει θα ειναι οπωσδηποτε πραγματικη. Το μειονεκτημα αυτο μπορει να αποφευχθει αν χρησιμοποιησουμε την λεγομενη μεθοδο Bairstow, που προσδιοριζει τους συντελεστες ενος δευτεροβαθμιου παραγοντα του αρχικου πολυωνυμου, και επιτρεπει ετσι τον προσδιορισμο συζυγων μιγαδικων ριζων χωρις πραξεις με μιγαδικους, αλλα απλα με επιλυση της δευτεροβαθμιας εξισωσης που αντιστοιχει στον δευτεροβαθμιο παραγοντα που βρηκαμε. Για την μέθοδο αυτή βλέπε Κιουστελίδη: Αριθμητική Ανάλυση, 1996, σελ. 65-69.

Ασκήσεις

1. Να βρεθουν ανω φραγματα για τις ριζες των πολυωνυμων:

$P(x) = x^5 - 8x^4 + 12x^3 - 16x^2 + 20x - 24$

$Q(x) = 8x^4 + 6x^3 - 5x^2 - 3x -1$

$R(x) = 5x^4 - 4x^2 - 3x -2$

2. Να προσδιορισθει οσο το δυνατον πιο μεγαλη περιοχη τιμων $t \in \mathbf{R}$ για τις οποιες το πολυωνυμο:

$P(x) = x^4 - (t^2 -1)x^3 + |3t+2|x^2 - (2t+1)x + t$

εχει ολες τις ριζες του μεσα στον μοναδιαιο κυκλο του μιγαδικου επιπεδου

($|x|<1$).

3. Να προσδιορισθει με την μεθοδο Newton - Raphson μια πραγματικη ριζα του πολυωνυμου $p(x) = x^3 - 2x^2 + 3x - 4$

4. ΝΟΡΜ ΔΙΑΝΥΣΜΑΤΩΝ ΚΑΙ ΠΙΝΑΚΩΝ

Οι περισσοτερες απο τις επαναληπτικες μεθοδους λυσεως μη γραμμικων εξισωσεων που ειδαμε στα προηγουμενα κεφαλαια μπορουν να γενικευθουν καταλληλα και να εφαρμοσθουν και στην περιπτωση γραμμικων και μη γραμμικων συστηματων. Ετσι, για να λυσουμε την *διανυσματική* εξισωση:

$$x = \varphi(x) \qquad (4.1)$$

δηλαδη, το σύστημα

$$(x_i)=(\varphi_i(x_1, x_2, \ldots,x_n)) \text{ με } i=1, 2, \ldots, n \qquad (4.2)$$

μπορουμε να χρησιμοποιησουμε την επαναληπτικη διαδικασια:

$$x^{(k)} = \varphi(x^{(k)}) \qquad (4.3)$$

δηλαδη,

$$(x_i^{(k+1)}) = (\varphi_i (x_1^{(k)}, x_2^{(k)},\ldots, x_n^{(k)})) \qquad (4.4)$$

οπου το i διατρεχει ολες τις συνιστωσες (i = 1(1)n), και ο δεικτης επαναληψεως, (k), (k = 0, 1,2,...) μπηκε στην επανω δεξια πλευρα καθε μεταβλητης για να μην τον μπερδευουμε με τον δεικτη συνιστωσας i.

Αν η επαναληπτικη αυτη διαδικασια συγκλινει σε καποιο σημειο

$$s = \varphi(s) \qquad (4.5)$$

του n-διαστατου πραγματικου χωρου $\mathbf{R}^n$ (ή και του μιγαδικου χωρου $\mathbf{C}^n$) τοτε για να μπορεσουμε να αποδειξουμε την συγκλιση με τροπο αναλογο οπως στα προηγουμενα κεφαλαια θα επρεπε να αφαιρεσουμε τις (4.3) και (4.5) κατα μελη και χρησιμοποιωντας καταλληλες γενικευσεις σε διανυσματα της εννοιας της απολυτης τιμης και της εννοιας της παραγωγου να πουμε οτι:

$$|x^{(k+1)} - s| = |\varphi(x^{(k)}) - \varphi(s)| = |\varphi'(\xi)||x^{(k)} - s| < |x^{(k)} - s|$$

αν $|\varphi'(x)| < 1$ σε καποιαν περιοχη γυρω απο το σημειο s.

Η καταλληλη γενικευση της απολυτης τιμης ειναι η εννοια της "νορμ" διανυσματων και πινακων, για την οποια θα μιλησουμε αμεσως τωρα. Την καταλληλη γενικευση της παραγωγου, την παραγωγο Frechet (Φρεσέ), που συνηθεστερα την λεμε Ιακωβιανο πινακα, θα την γνωρισουμε στο κεφαλαιο περι μη γραμμικων συστηματων.

4.1 Νορμ Διανυσμάτων

Ορισμός 4.1 (Νορμ)

Εστω οτι **V** ειναι ο χωρος $\mathbf{R}^n$ η $\mathbf{C}^n$. Καθε πραγματικη συναρτηση $v(x):=v(x_1,x_2,\ldots,x_n)$ των στοιχειων (δηλ. διανυσματων) του χωρου τουτου λεγεται norm (norm) του **V** και συμβολιζεται με 7x7 (7x7:= v(x)) αν εχει τις ακολουθες ιδιοτητες:

1. 7x7 > 0 για $x \neq \theta$,

οπου θ παριστανει το μηδενικο διανυσμα στον V (θ = (0,0,..,0)).

2. 7x7 = 0 μονο αν x = θ.
3. 7a x7 = |a|7x7 για καθε a $\in$ **R** ή **C** και καθε x $\in$ **V**.
4. 7x + y7 $\leq$ 7x7 + 7y7 για καθε x,y $\in$ **V** (Ανισοτητα Τριγωνου).

Οπως βλεπουμε, οι ιδιοτητες αυτες ειναι αναλογες με αυτες της απολυτης τιμης ή του μετρου και γι' αυτο συμβολιζονται και οι νορμ με αναλογο τροπο. Βασικη τους διαφορα ειναι οτι για n σ 2 δεν υπαρχει μονο μια συναρτηση που να εχει αυτες τις ιδιοτητες, οπως συμβαινει στην μια διασταση με την απολυτη τιμη,

αλλα ουσιαστικα απειρες. Παραδειγματος χαριν, ολες οι συναρτησεις:

$$\|x\|_p = (\sum_{k=1}^{n} | x_k |^p)^{1/p} \quad , \text{p σ 1} \qquad (4.6)$$

ειναι νορμ, γιατι ικανοποιουν ολες τις παραπανω σχεσεις (Η τεταρτη σχεση, λεγεται στην περιπτωση αυτην ανισοτητα του Minkowski).

Απο ολες αυτες τις νορμ, στην πραξη χρησιμοποιουνται μονον τρεις:

$$\|x\|_\infty = \max_{i=1(1)n} | x_i | \quad \text{(maximum norm)} \qquad (4.7)$$

$$\|x\|_1 = \sum_{k=1}^{n} | x_k | \quad \text{(Αθροιστικη νορμ)} \qquad (4.8)$$

$$\|x\|_2 = (\sum_{k=1}^{n} | x_k |^2)^{1/2} \quad \text{(Ευκλειδια νορμ)} \qquad (4.9)$$

Απο τις τρεις η συνηθεστερα χρησιμοποιουμενη ειναι η πρωτη, η νορμ μεγιστου, γιατι ειναι πιο απλος ο υπολογισμος της. Δευτερη σε σειρα προτιμησεως ειναι για τον ιδιο λογο η αθροιστικη νορμ, ενω η Ευκλειδια νορμ χρησιμοποιειται κυριως σε σχεση με γεωμετρικες εφαρμογες.

Παραδείγματα Υπολογισμού Νορμ:

Οι παραπανω νορμ για το διανυσμα $x \in \mathbf{R}^5$:

$$x = (-1, 2, -2, -3, 1)^T$$

ειναι:

$$\|x\|_\infty = \max \{|-1|,|2|,|-2|,|-3|,|1|\} = 3$$

$$\|x\|_1 = |-1|+|2|+|-2|+|-3|+|1| = 9$$

$$\|x\|_2 = (|-1|^2+|2|^2+|-2|^2+|-3|^2+|1|^2)^{1/2} = 19^{1/2} \cong 4.3589$$

Με την βοηθεια μιας, οποιασδηποτε, νορμ μπορουμε τωρα να ορισουμε πολυ απλα την αποσταση δυο σημειων στον χωρο V:

Ορισμός 4.2

Απόσταση δυο σημειων $x \in \mathbf{V}$, $y \in \mathbf{V}$ λεγεται η νορμ της διαφορας τους:

$$\rho(x,y) := \|x - y\|$$

Εφ' οσον ο διανυσματικος χωρος V εφοδιασθει με μιαν συγκεκριμενη τετοια εννοια αποστασεως, λεγεται «χωρος νορμε» ως προς την αντιστοιχη νορμ.

Η εννοια αυτη της αποστασης ταυτιζεται με την συνηθη εννοια αποστασης μονο στην περιπτωση της ευκλειδιας νορμ. Εν τουτοις σε πολλες περιπτωσεις μπορουμε να εισαγαγουμε μιαν πιο ευστοχη εννοια αποστασης χρησιμοποιωντας μιαν αλλη νορμ. Παραδειγματος χαριν, στην περιπτωση μιας πολης με ορθογωνιο οδικο δικτυο ειναι ευκολο να ιδουμε οτι η συντομωτερη αποσταση δυο σημειων $x, y \in \mathbf{R}^2$ κατα μηκος του οδικου δικτυου (αν θεωρησουμε το πλατος των δρομων αμελητεο) ειναι:

$$\|x - y\|_1 = |x_1 - y_1 | + |x_2 - y_2|$$

Διαδρομες ελαχιστου μηκους υπαρχουν στην περιπτωση αυτη εν γενει περισσοτερες απο μια, αλλα ολες εχουν το παραπανω μηκος.

Ασκήσεις

1) Σχεδιαστε τους "νορμόκυκλους" με κεντρο θ και ακτινα R ως προς καθε μιαν απο τις παραπανω νορμ:

$$\|x\|_p \leq R\ ,\quad p = 1,2,\infty$$

2) Δειξτε οτι:

$$\lim_{p\to\infty} \|x\|_p = \|x\|_\infty \quad \text{δηλ. } \lim_{p\to\infty} (\|x\|_p / \max_{i=1(1)n}|x_i|) = 1$$

Βαζοντας y = θ στον παραπανω ορισμο βλεπουμε οτι η νορμ ενος σημειου x ταυτιζεται με την αποσταση του απο το σημειο μηδεν. Μπορουμε επισης να ειπουμε οτι μια συγκεκριμενη νορμ, εφ' οσον χρησιμοποιειται για ολα τα $x \in V$, αποτελει μετρο του μεγεθους τους.

Τωρα γινεται κατανοητο και γιατι η τεταρτη ιδιοτητα του ορισμου μιας νορμ ονομαστηκε "Ανισοτητα Τριγωνου". Αν x και y ειναι δυο πλευρες ενος τριγωνου, τοτε x + y ειναι το διανυσμα που εκφραζει την τριτη πλευρα και 7x+y7 ειναι το μηκος του (ως προς την συγκεκριμενη νορμ). Εχουμε, δηλαδη, την γενικευση της γνωστης ιδιοτητας της Ευκλειδιας Γεωμετριας, οτι το αθροισμα του μηκους των δυο πλευρων ενος τριγωνου ειναι μεγαλυτερο απο το μηκος της τριτης πλευρας.

Η ιδιοτητα αυτη μπορει καπως να επεκταθει στην λεγομενη "Διπλη Ανισοτητα τριγωνου", που ειναι αναλογη με αντιστοιχη ιδιοτητα των απολυτων τιμων:

Θεώρημα 4.1

Ολες οι διανυσματικες νορμ ικανοποιουν την λεγομενη «Διπλη Ανισοτητα Τριγωνου»:

$$|\|x\| - \|y\|| \leq \|x \pm y\| \leq \|x\|+\|y\| \qquad (4.10)$$

Απόδειξη

Η δευτερη ανισοτητα ισχυει εξ' ορισμου. Η πρωτη προκυπτει με τον ακολουθο τροπο:

$$\|x\| = \|(x + y) - y\| \leq \|x+y\| + \|y\|$$

αρα

$$\|x\|-\|y\| \leq \|x+y\|$$

Παρομοια προκυπτει:

$$\|y\|-\|x\| \leq \|x+y\|$$

δηλαδη

$$|\|x\|-\|y\|| \leq \|x+y\|$$

Οι ιδιες ανισοτητες ισχυουν και αν αντικαταστησουμε το y με -y, γιατι, με βαση την τριτη ιδιοτητα των νορμ, αλλαγη προσημου (a = -1) δεν μεταβαλλει το μεγεθος της νορμ.

Η εισαγωγη της εννοιας της νορμ μας επιτρεπει να αναγαγουμε την συγκλιση μιας ακολουθιας διανυσματων στην συγκλιση μιας ακολουθιας πραγματικων αριθμων:

Ορισμός 4.3

Λεμε οτι μια ακολουθια διανυσματων $x^{(k)} \in \mathbf{V}$, k = 0, 1, 2, ... οπου

$$x^{(k)} = (x_1^{(k)}, x_2^{(k)}, \ldots, x_n^{(k)})$$

συγκλινει (ισχυρα) προς το οριο $s \in \mathbf{V}$, οταν ισχυει:

$$\lim_{k\to\infty} \|x^{(k)} - y\| = 0$$

τοτε γραφουμε επισης:

$$\lim_{k\to\infty} x^{(k)} = s$$

Παρατήρηση: Ισοδυναμία των Νορμ ως προς την Σύγκλιση

Παρα το οτι για τον ορισμο της συγκλισης μιας ακολουθιας διανυσματων χρησιμοποιουμε καποιαν νορμ, η συγκλιση αυτη ειναι ανεξαρτητη απο το ποιαν νορμ θα παρουμε. Δηλαδη δεν μπορει μια ακολουθια διανυσματων να συγκλινει ως προς μιαν νορμ και να μην συγκλινει ως προς αλλην. Αυτο ισχυει ομως μονο για χωρους επερασμενης διαστασεως (οχι για απειροδιαστατους χωρους) και οφειλεται στην ακολουθη ιδιοτητα των νορμ:

Θεώρημα 4.2

Για καθε ζευγαρι νορμ $\|.\|_A$, $\|.\|_B$ του n-διαστατου χωρου $\mathbf{V}$ υπαρχουν θετικες σταθερες m, M τετοιες ωστε για καθε $x \in \mathbf{V}$ ισχυει:

$$m\|x\|_A \le \|x\|_B \le M\|x\|_A$$

Απόδειξη: Βλεπε Isaacson - Keller, σελ.7.

Παραδείγματα: Για $x \in \mathbf{R}^n$ ή $\mathbf{C}^n$ ειναι

$$\|x\|_\infty \le \|x\|_2 \le n^{1/2}\|x\|_\infty \qquad (4.11)$$

$$\|x\|_2 \le \|x\|_1 \le n^{1/2}\|x\|_2 \qquad (4.12)$$

Συνεπεια του θεωρηματος τουτου ειναι οτι αν καποια νορμ τεινει για μιαν ακολουθια διανυσματων στο μηδεν, τοτε και καθε αλλη νορμ θα τεινει στο μηδεν.

4.2 Νορμ Πινάκων

Το συνολο των τετραγωνικων πινακων n x n μπορει να θεωρηθει οτι αποτελει εναν γραμμικο χωρο ως προς τις γνωστες πραξεις προσθεσεως πινακων και πολλαπλασιασμου πινακα με σταθερα. Στον χωρο αυτον μπορουν επομενως να ορισθουν νορμ οπως εγινε προηγουμενως. Ομως στην περιπτωση των τετραγωνικων πινακων υπαρχει και η γνωστη πραξη του πολλαπλασιασμου A*B, που δημιουργει ομοειδη πινακα (διαστασεων nxn) οπως οι αρχικοι A, B. Εφ' οσον ο διανυσματικος χωρος των τετραγωνικων πινακων εφοδιασθει και με την πραξη αυτην λεγεται *Άλγεβρα Πινάκων*. Αντιστοιχα, οποιεσδηποτε νορμ ορισουμε για την αλγεβρα των τετραγωνικων πινακων πρεπει να εχουν καποια προσθετη ιδιοτητα που να συνδεει την νορμ του γινομενου δυο πινακων με τις νορμ καθε ενος απο τους πινακες αυτους. Ετσι, ο ορισμος 4.1 επεκτεινεται στην περιπτωση της αλγεβρας των τετραγωνικων πινακων με τον ακολουθο τροπο:

Ορισμός 4.4

Νορμ για την αλγεβρα $\mathbf{\Pi}$ των τετραγωνικων nxn πινακων λεγεται μια πραγματικη συναρτηση 7A7 των στοιχειων a_{ik} καθε πινακα $A \in \mathbf{\Pi}$ αν εχει τις

ακολουθες ιδιοτητες:

1. $\|A\| > 0$ για καθε $A \in$ **Π** διαφορετικο απο τον μηδενικο πινακα 0.
2. $\|A\| = 0$ μονο αν $A = \mathbf{0}$, δηλ. $a_{ik} = 0$ για $i,k = 0(1)n$.
3. $\|\lambda A\| = |\lambda|\|A\|$ για καθε $\lambda \in$ **R** η **C** και καθε $A \in$ **Π**.
4. $\|A+B\| \leq \|A\|+\|B\|$ για καθε $A, B \in$ **Π**.
5. $\|AB\| \leq \|A\|\|B\|$ για καθε $A, B \in$ **Π**.

Τετοιες νορμ μπορουν να ορισθουν κατα πολλους τροπους. Μια βασικη κατηγορια προκυπτει απο διανυσματικες νορμ με την βοηθεια του ακολουθου θεωρηματος:

Θεώρημα 4.3:

α) Αν ειναι $7x7_\Delta$ μια τυχουσα νορμ του διανυσματικου χωρου **V**, τοτε η συναρτηση των $A \in$ **Π**:

$$\|A\| = \max_{x \neq \theta} \frac{\|Ax\|_\Delta}{\|x\|_\Delta} = \max_{\|y\|_\Delta = 1} \|Ay\|_\Delta \qquad (4.13)$$

(με $y = x/7x7_\Delta$) αποτελει νορμ της αλγεβρας πινακων **Π**, και λεγεται "επαγομενη" απο την διανυσματικη νορμ 7.7_Δ.

β) Για καθε $z \in$ **V** η νορμ αυτη ικανοποιει την σχεση:

$$\|Az\|_\Delta \leq \|A\|\|z\|_\Delta \qquad (4.14)$$

Απόδειξη: Βλέπε Κιουστελίδη: Αριθμητική Ανάλυση, 1996, σελ.77.

Βασικωτερες επαγομενες νορμ ειναι αυτες που προκυπτουν απο τις διανυσματικες νορμ (4.7), (4.8), (4.9).

Θεώρημα 4.4

Οι απο τις διανυσματικες νορμ 7.7_∞ , 7.7_1 , 7.7_2 επαγομενες νορμ πινακων ειναι αντιστοιχα:

$$\|A\|_\infty = \max_{i=1(1)n} \sum_{k=1}^{n} |a_{ik}| \quad \text{(μεγιστο αθροισμα σειρας)} \quad (4.15)$$

$$\|A\|_1 = \max_{k=1(1)n} \sum_{i=1}^{n} |a_{ik}| \quad \text{(μεγιστο αθροισμα στηλης)} \quad (4.16)$$

$$\sigma(A) = (\max_{i=1(1)n} |\lambda_i(A^*A)|)^{1/2} \quad \text{(φασματικη νορμ)} \quad (4.17)$$

οπου $\lambda_i(A^*A)$ ειναι οι ιδιοτιμες του πινακα A^*A. A^* ειναι εδώ ο αναστροφος (transposed) του συζυγους μιγαδικου προς τον A, ή απλά ο ανάστροφος, A^T, του A, αν πρόκειται για πίνακα με πραγματικά στοιχεία.

Απόδειξη: Βλέπε Κιουστελίδη: Αριθμητική Ανάλυση, 1996, σελ.78.

Επειδή οι νορμ αυτές είναι επαγόμενες από αντίστοιχες διανυσματικές νορμ συνδέονται με αυτές με τις βασικές ανισότητες:

$$\|Ax\|_p \leq \|A\|_p\|x\|_p, \text{ για } p=1 \text{ και } \infty \qquad (4.18)$$

καθώς και

$$\|Ax\|_2 \le \sigma(A)\, \|x\|_2 \qquad (4.19)$$

Γενικά, νορμ διανυσμάτων και πινάκων που συνδέονται με ανισότητες του τύπου (4.18) λέγονται *συμβιβαστές* μεταξύ τους.

Σε συνδυασμο με την διανυσματικη νορμ $\|.\|_2$ χρησιμοποιειται συχνα η νορμ:

$$\|A\|_2 = (\sum_{i=1}^{n}\sum_{k=1}^{n} |\, a_{ik}\, |^2)^{1/2} \qquad (4.20)$$

η οποια ικανοποιει επισης την σημαντικη για εκτιμησεις σφαλματων ανισοτητα:

$$7Ax7_2 \le 7A7_2 7x7_2 \qquad (4.21)$$

Η αποδειξη οτι η (4.20) εχει ολες τις ιδιοτητες μιας νορμ πινακων μπορει να γινει με χρηση της ανισοτητας του Schwartz.

Μία χρήσιμη εφαρμογή των παραπάνω ανισοτήτων είναι ότι για ολες τις ιδιοτιμες λ_i , i= 1(1)n, ενος πινακα A ισχυει:

$$|\lambda_i| \le \|A\| \qquad (4.22)$$

οπου 7A7 οποιαδήποτε από τις παραπάνω νορμ πινακων.

Αυτό προκύπτει αμέσως απο την σχεση

$$\lambda_i\, x^{<i>} = Ax^{<i>},$$

που συνδέει τις ιδιοτιμές λ_i με τα αντίστοιχα ιδιοδιανύσματα $x^{<i>}$, αν πάρουμε κάποια νορμ και στις δύο πλευρές της:

$$|\lambda_i\, |\|x^{<i>}\| = \|\lambda_i\, x^{<i>}\| = \|Ax^{<i>}\| \le \|A\|\|x^{<i>}\|$$

Παράδειγμα υπολογισμού των βασικών νορμ ενός πίνακα

Να υπολογισθουν οι βασικες νορμ και ανω φραγματα του μετρου των ιδιοτιμων του πινακα:

$$A = \begin{pmatrix} -5 & 2 & -1 \\ 3 & -4 & 2 \\ -2 & -2 & 4 \end{pmatrix}$$

Λύση

α) $\|A\|_\infty = \max\{|-5|+|2|+|-1|, |3|+|-4|+|2|, |-2|+|-2|+|4|\} = 9$

$\|A\|_1 = \max\{|-5|+|3|+|-2|, |2|+|-4|+|2|, |-1|+|2|+|4|\} = 9$

$\|A\|_2 = (|-5|^2+|2|^2+|-1|^2+|3|^2+|-4|^2+|2|^2+|-2|^2+|-2|^2+|4|^2)^{1/2}$

$= \sqrt{80} = 8.944$

β) $|\lambda_i| \le \min\{\|A\|_1, \|A\|_2, \|A\|_\infty\} = \sqrt{80} = 8.944$

Ασκήσεις

Να υπολογισθουν οι βασικες νορμ καθως και ανω φραγματα για το μετρο των ιδιοτιμων των παρακατω πινακων:

$$\begin{pmatrix} -2 & 1 & 2 \\ 1 & -3 & -1 \\ 3 & 2 & 1 \end{pmatrix}, \begin{pmatrix} 5 & 7 & 3 \\ -7 & 11 & 2 \\ 3 & 2 & -6 \end{pmatrix}, \begin{pmatrix} 5 & -1 & 1 \\ 1 & -4 & 1 \\ 2 & 3 & 8 \end{pmatrix}$$

5. ΕΠΙΛΥΣΗ ΓΡΑΜΜΙΚΩΝ ΣΥΣΤΗΜΑΤΩΝ

Γραμμικα συστηματα εμφανιζονται σε ολους τους κλαδους της Τεχνικης των Φυσικων Επιστημων, αλλα και της Οικονομιας. Για να περιορισθουμε σε κλαδους της Τεχνικης αρκει να υπενθυμισουμε τους νομους του Kirchoff στην Ηλεκτρολογια, που εχουν μορφη γραμμικου συστηματος η τους νομους της Στατικης, που εχουν επισης μορφη γραμμικου συστηματος.

Περα ομως απο ολες αυτες τις εφαρμογες, οι ίδιες οι μέθοδοι της Αριθμητικής ανάλυσης παράγουν γραμμικά συστήματα και μάλιστα τεραστίων διαστάσεων. Η διακριτοποίηση ανάγει σε γραμμικά συστήματα, για παράδειγμα, ολοκληρωτικές εξισώσεις ή προβλήματα συνοριακών τομών για διαφορικές εξισώσεις. Το μέγεθος των συστημάτων εξαρτάται εδώ από το με ποια ακρίβεια επιθυμούμε να προσεγγίσουμε την λύση, γιατί η τιμή της λύσης σε κάθε θέση του πεδίου όπου εκτείνεται το πρόβλημα αποτελεί ουσιαστικά έναν χωριστό άγνωστο του διακριτοποιημένου προβλήματος. Έτσι το μέγεθος του συστήματος εξαρτάται από το σε πόσα σημεία θέλουμε να προσδιορίσουμε την λύση. Μία ιδιαίτερα σημαντική εφαρμογή αυτού του είδους είναι η προσεγγιστική επίλυση των διαφορικών εξισώσεων Navier-Stokes της ρευστοδυναμικής για την πρόβλεψη του καιρού των επομένων ημερών (με βάση τα τρέχοντα μετεωρολογικά δεδομένα). Μία πρόβλεψη που επιδιώκει τον προσδιορισμό των αυριανών καιρικών συνθηκών ανά δέκα με δεκαπέντε χιλιόμετρα απαιτεί την λύση γραμμικών συστημάτων με εκατοντάδες χιλιάδων εξισώσεις, που μπορεί να γίνει ικανοποιητικά μόνο με χρήση υπερϋπολογιστών.

Το πληθος των εφαρμογων γραμμικων συστηματων ειναι, δηλαδή τεραστιο, και αντίστοιχα τεραστια είναι η βιβλιογραφία μεθοδων για την επιλυση τους. Ολες αυτες οι μεθοδοι χωριζονται βασικα σε δυο κατηγοριες: τις επαναληπτικες μεθοδους, και τις μεθοδους απαλοιφης[3].

Η απο το Λυκειο γνωστη μας μεθοδος επιλυσης με την βοηθεια οριζουσων, η μεθοδος του Gramm, δεν ενδεικνυται σε καμμιαν περιπτωση σαν μεθοδος επιλυσης και χρησιμευει μονον για θεωρητικους συλλογισμους. Το μειονεκτημα της ειναι οτι απαιτει τεραστιον ογκο πραξεων σε σχεση με αλλες μεθοδους.

Θα ασχοληθουμε λοιπον εδω με τις δυο παραπανω βασικες κατηγοριες μεθοδων, και αρχιζουμε με τις επαναληπτικες.

5.1 Επαναληπτικές Μέθοδοι

Οι επαναληπτικες μεθοδοι επιλυσης χρησιμευουν μονο στην περιπτωση που το γραμμικο συστημα ειναι "σχεδον διαγωνιο", δηλαδη εχει ορους

---

[3] Σε μία τρίτη κατηγορία υπάγεται η σημαντική Μέθοδος των Συζυγών Κλίσεων, που είναι μεν επαναληπτική, αλλά ανάγεται σε μεθόδους βελτιστοποίησης. Αυτή εφαρμόζεται σε μεγάλα αραιά συστήματα με θετικά προσδιορισμένο πίνακα και είναι ιδιαίτερα ταχεία. Δεν θα την περιγράψουμε όμως εδώ, γιατί ο διαθέσιμος χρόνος για ένα εισαγωγικό μάθημα δεν επιτρέπει την παρουσίαση των μεθόδων βελτιστοποίησης και των εφαρμογών τους. Για την μέθοδο αυτήν βλέπε π. χ. D. Kincaid-W. Cheney: Numerical Analysis, Brooks/Cole Publ.Co.,1991, σελ.204-218.

διαφορετικους απο το μηδεν μονο στην κυρια διαγωνιο και μερικες παραπλευρες διαγωνιους. Στην περιπτωση πινακα με σχεδον ολα τα στοιχεια του μη μηδενικα ΔΕΝ συμφερουν γιατι ο ογκος των πραξεων ειναι πολυ μεγαλυτερος απο οτι με απαλοιφη Gauss.

Οι βασικες επαναληπτικες μεθοδοι ειναι δυο: η Μεθοδος Ολικων Βηματων και η Μεθοδος Μερικων η Μεμονωμενων Βηματων. Και οι δυο προυποθετουν οτι το συστημα εχει την μορφη:

$$x = Mx + r \tag{5.1}$$

απο την οποια μπορουμε να κατασκευασουμε μιαν επαναληπτικη διαδικασια. Το δοσμενο συστημα πρεπει, δηλαδη, να μετασχηματισθει αρχικα με καταλληλο τροπο στην μορφη αυτη. Τουτο δεν ειναι παντα απαραιτητο, γιατι σε καποιες βασικες εφαρμογες προκυπτει αμεσα στην μορφη (5.1). Αυτο συμβαινει, π.χ., κατα τον υπολογισμο των συντελεστων μιας τμηματικα πολυωνυμικης προσεγγισης τυπου Spline, η κατα την επιλυση προβληματων συνοριακων τιμων για διαφορικες εξισωσεις, ειτε με την μεθοδο Πεπερασμενων Στοιχειων, ειτε με την μεθοδο Πεπερασμενων Διαφορων.

5.1.1. Μέθοδος Ολικών Βημάτων

Η πρωτη απο τις επαναληπτικες μεθοδους επιλυσης γραμμικων συστηματων ειναι ουσιαστικα η γενικη επαναληπτικη μεθοδος. Κατ' αναλογιαν με οσα ειπαμε γι'αυτην στην περιπτωση μη γραμμικων εξισωσεων μιας μεταβλητης μπορουμε τωρα να αποδειξουμε το ακολουθο θεωρημα:

Θεώρημα 5.1 (Μεθοδος Ολικων Βηματων η Jacobi η Gauss-Jordan)

Αν για οποιαδηποτε νορμ του πινακα M ισχυει:

$$\|M\| < 1 \tag{5.2}$$

τοτε αληθευουν τα ακολουθα:

α. Το συστημα (5.1) εχει μιαν μονοσημαντα ωρισμενη λυση s.

β. Η επαναληπτικη ακολουθια διανυσματων:

$$x^{(k)} = Mx^{(k-1)} + r \tag{5.3}$$

δηλ. κατα συνιστωσες:

$$x_i^{(k)} = \sum_{j=1}^{n} m_{ij} x_j^{(k-1)} + r_i \ , i = 1,2,3,\ldots,n \tag{5.4}$$

οπου k = 0, 1, 2, 3, ... είναι ο δείκτης επανάληψης, συγκλινει για τυχον αρχικο διανυσμα $x^{(0)} \in \mathbf{R}^n$ ή $\mathbf{C}^n$ προς την λυση s του συστηματος.

γ. Για καθε διανυσματικη νορμ συμβιβαστή με την παραπανω νορμ πινακων ισχυει:

$$\left\|x^{(k)} - s\right\| \le \frac{\|M\|}{1-\|M\|}\left\|x^{(k)} - x^{(k-1)}\right\| \le \frac{\|M\|^k}{1-\|M\|}\left\|x^{(1)} - x^{(0)}\right\|, \tag{5.5}$$

όπου η πρώτη ανισότητα λέγεται «εκ των υστέρων» ή «a posteriori» εκτιμηση σφαλματος, και η δεύτερη «εκ των προτερων» η «a priori» εκτιμηση σφαλματος.

Απόδειξη:

α. Η λύση του συστήματος (5.1), (I - M)x = r, οπου I ο μοναδιαιος πινακας,

δεν είναι μονοσημαντα ορισμενη μονον οταν το αντιστοιχο ομογενες, (I -M)x = θ, εχει και μιαν μη μηδενικη λυση. Αν, δηλαδη, το (5.1) εχει δυο λυσεις u και v, τοτε αφαιρωντας τις αντιστοιχες ισοτητες κατα μελη εχουμε (I -M)t = θ, οπου t:= u - v. Αλλα αν t ειναι οποιαδηποτε λυση του ομογενους συστηματος τοτε ισχυει:

$$t = M\,t = M^2\,t = M^3\,t = \ldots = M^k\,t$$

και επομενως, επειδη 7M7 < 1,

$$7t7=7M^k\,t7\beta7M7^k7t7\to 0 \text{ για } k\to\infty$$

Δηλαδη, μοναδικη λυση του ομογενους συστηματος ειναι η μηδενικη και το μη ομογενες συστημα (5.1) εχει μιαν μονοσημαντα ωρισμενη λυση s.

β. Για την λυση s ισχυει:

$$s = Ms + r$$

Αφαιρωντας την σχεση αυτην κατα μελη απο την (5.3) εχουμε:

$$x^{(k)} - s = M(x^{(k-1)} - s)\ ,\ k = 0, 1, 2, ..$$

και επομενως,

$$x^{(k)} - s = M(x^{(k-1)} - s) = M^2(x^{(k-2)} - s)= M^3(x^{(k-3)} - s)= \ldots = M^k(x^{(0)} - s) \qquad (5.6)$$

Για οποιαδηποτε διανυσματικη νορμ συμβιβαστη με την νορμ πινακα εχουμε επομενως:

$$7x^{(k)} - s7 = 7M^k(x^{(0)} - s)\ 7\beta7M7^k\ 7x^{(0)} - s7\to 0 \text{ για } k\to\infty$$

Δηλαδη η επαναληπτικη ακολουθια διανυσματων (5.3) συγκλινει πραγματι προς την λυση s, ΑΣΧΕΤΩΣ του ποιο ειναι το αρχικο διανυσμα $x^{(0)}$.

γ. Απο την (5.6) προκυπτει:

$$7x^{(k)} - s7 = 7M(x^{(k-1)} - s)7\ \beta7M7\ 7x^{(k-1)} - s7=7M7\ 7x^{(k-1)} - x^{(k)} + x^{(k)} - s7\beta$$

$$\beta7M7\ \{7x^{(k-1)} - x^{(k)}\ 7+7x^{(k)} - s7\}$$

αρα

$$(1 - 7M7)7x^{(k)} - s7\ \beta7M7\ \{7x^{(k-1)} - x^{(k)}\ 7$$

δηλαδη, η πρώτη ανισότητα (5.5).

Για να αποδειξουμε την δεύτερη ανισότητα (5.5) αφαιρουμε κατα μελη τις σχεσεις:

$$x^{(k)}= Mx^{(k-1)}+ r, \quad x^{(k-1)}= Mx^{(k-2)}+ r$$

οποτε προκυπτει:

$$x^{(k)} - x^{(k-1)} = M(x^{(k-1)} - x^{(k-2)})$$

Απο αυτην εχουμε επαγωγικα:

$$x^{(k)} - x^{(k-1)} = M(x^{(k-1)} - x^{(k-2)})= M(x^{(k-2)} - x^{(k-3)})=\ldots =M^{k-1}(x^{(1)} - x^{(0)})$$

και επομενως,

$$7x^{(k)} - x^{(k-1)}\ 7\beta7M7^{k-1}7x^{(1)} - x^{(0)}7$$

Σε συνδυασμο με την πρώτη ανισότητα (5.5) η σχεση αυτη μας δινει την δεύτερη.

Παρατήρηση: Διαγώνια Υπεροχή

Στην περιπτωση συστηματων της μορφης Ax = b, δηλ.

$$\sum_{j=1}^{n} a_{ij}x_j = b_i\ , \quad i = 1,2,3,\ldots,n \qquad (5.7)$$

για να εφαρμοσουμε την μεθοδο ολικων βηματων μπορουμε να επιλύσουμε κάθε

μία από αυτές τις εξισώσεις ως προς διαφορετικόν άγνωστο x_j. Έτσι μπορούμε να μετατρέψουμε το σύστημα στην μορφή x=Mx+r. Για παράδειγμα, αν επιλύσουμε τις σχέσεις (5.7) ως προς τους αγνώστους που αντιστοιχουν στην κυρια διαγωνιο του πινακα A εχουμε:

$$x_i = -\frac{1}{a_{ii}} \sum_{j=1}^{n} a_{ij} x_j + \frac{b_i}{a_{ii}}, \qquad i = 1,2,3,...,n \qquad (5.9)$$

Αν χρησιμοποιησουμε την νορμ 7M7∞ (μεγιστο αθροισμα σειρας) τοτε απο το θεωρημα 5.1 προκυπτει σαν ικανη (οχι όμως αναγκαια) συνθηκη συγκλισης της ακολουθιας διανυσματων $x^{(k)}$:

$$\|M\|_\infty = \max_{i=1(1)n} \frac{1}{|a_{ii}|} \sum_{j=1}^{n} |a_{ij}| < 1 \qquad (5.10)$$

Αυτο σημαινει οτι η συγκλιση της (5.10) ειναι βεβαιη οταν:

$$|a|_{ii} > \sum_{j=1, j\neq i}^{n} |a_{ij}| \qquad (5.11)$$

δηλαδη οταν καθε διαγωνιος συντελεστης του πινακα A εχει μετρο (ή απολυτη τιμη) μεγαλυτερο απο το αθροισμα των μετρων των υπολοιπων ορων καθε σειρας. Στην περιπτωση αυτην μιλαμε για *διαγώνια υπεροχή* (υπεροχη της διαγωνιου) του πινακα A.

Αν και η συνθήκη (5.11) είναι μόνο ικανή και όχι ικανή και αναγκαία, είναι όμως η μόνη που πρακτικά επιτρέπει να αναγνωρίσουμε άμεσα, χωρίς μεγάλο όγκο πράξεων, αν ένα σύστημα είναι επαναληπτικά επιλύσιμο ή όχι.

Ικανή και αναγκαία συνθήκη σύγκλισης της μεθόδου είναι το να ισχύει:

$$\rho(M) := \max_{i=1(1)n} |\lambda_i| < 1 \qquad (5.12)$$

όπου λ_i οι ιδιοτιμές του πίνακα M[4]. Η συνθήκη όμως αυτή έχει μόνο θεωρητική αξία, γιατί ο προσδιορισμός όλων των ιδιοτιμών ενός πίνακα είναι ένα υπολογιστικά πολύ πίο απαιτητικό πρόβλημα από την λύση ενός γραμμικού συστήματος.

Η επαναληπτικη επιλυση του συστηματος Ax = b εχει επομένως νοημα μονον αν μπορουμε να το φερουμε σε μορφη οπου ο πινακας A να εμφανιζει διαγωνια υπεροχη και μαλιστα αισθητη, γιατι η ταχυτητα συγκλισης της μεθοδου είναι τόσο μεγαλύτερη όσο μικρότερη μικροτερη της μοναδας ειναι η νορμ (5.10).

Η διαγώνια υπεροχή ενδέχεται να μην είναι άμεσα εμφανής, αν οι εξισώσεις του συστήματος Ax=b δεν έχουν την κατάλληλη σειρά. Γι' αυτό συνίσταται να επιλύουμε τις εξισώσεις του, όχι ως προς κάθε έναν από τους αγνώστους της διαγωνίου, αλλά ως προς τον άγνωστο που έχει τον απόλυτα μεγαλύτερο συντελεστή. Αν αυτό συμβαίνει σε κάθε μία από τις εξισώσεις για διαφορετικόν

---

[4] Γιατί από την (5.6) προκύπτει ότι αν γράψουμε το διάνυσμα $x^{(0)} - s$ σαν γραμμικό συνδυασμό των ιδιοδιανυσμάτων $v^{<m>}$ του πίνακα M:

$x^{(0)} - s = \Sigma\, c_i v^{<i>}$,

τότε ισχύει:

$$x^{(k)} - s = M^k(x^{(0)} - s) = \Sigma c_i M^k v^{<i>} = \Sigma c_i \lambda_i^k v^{<i>} \to 0 \text{ για } k \to \infty.$$

άγνωστο, τότε μπορούμε να τους συγκεντρώσουμε σε ένα διάνυσμα x και να δημιουργήσουμε έτσι ένα σύστημα της μορφής: x=Mx+r. Αν τώρα είναι 7M7$_\infty$<1 μπορούμε να εφαρμόσουμε την επαναληπτική διαδικασία: $x^{(k)} = Mx^{(k-1)} + r$.

Παράδειγμα Εφαρμογής της Μεθόδου ολικών Βημάτων

Με την μεθοδο ολικων βηματων και με αρχικο διανυσμα το μηδενικό διάνυσμα, $x^{(0)} = \theta$, να λυθει προσεγγιστικα το συστημα:

$$\begin{aligned} (1)\quad & x_1 + 4x_2 + x_3 = 2 \\ (2)\quad & x_1 + 2x_2 + 5x_3 = 3 \\ (3)\quad & 8x_1 + 2x_2 - x_3 = 4 \end{aligned}$$

α. Μετα την πεμτη επαναληψη να γινει εκτιμηση του σφαλματος με τις νορμ 7M7$_\infty$ και 7M7$_1$.

β. Ποσες το πολυ επαναληψεις χρειαζονται για να καθορισθουν ολες οι συντεταγμενες με ακριβεια τεσσαρων δεκαδικων ψηφιων μετα την υποδιαστολη, δηλ. με σφαλμα μικροτερο απο $0{,}5\ 10^{-4}$;

Λύση

Εδώ ο πίνακας ο πινακας του συστηματος δεν εχει διαγωνια υπεροχη. Αν όμως εξετάσουμε ποιος άγνωστος έχει υπερισχύοντα συντελεστή σε κάθε εξίσωση, βλέπουμε ότι η (3) μπορεί να επιλυθεί ως προς x_1, η (1) ως προς x_2, και η (2) ως προς x_3. γράφοντας, λοιπόν τις εξισώσεις με την σειρά (3), (1), (2) έχουμε ένα σύστημα του οποίου ο πίνακας εμφανίζει διαγώνια υπεροχή και επομένως επιτρέπει την χρήση επαναληπτικής μεθόδου:

$$\begin{aligned} (3)\quad & 8x_1 + 2x_2 - x_3 = 4 \\ (1)\quad & x_1 + 4x_2 + x_3 = 2 \\ (2)\quad & x_1 + 2x_2 + 5x_3 = 3 \end{aligned}$$

Επιλύοντας, λοιπόν, κάθε εξίσωση αυτού του συστήματος ως προς τον άγνωστο που βρίσκεται στην διαγώνιο του πίνακα έχουμε άμεσα την επαναληπτικη διαδικασια:

$$\begin{aligned} x_1^{(k+1)} &= 0.5 - 0.25\, x_2^{(k)} + 0.125\, x_3^{(k)} \\ x_2^{(k+1)} &= 0.5 - 0.25\, x_1^{(k)} - 0.25\, x_3^{(k)} \\ x_3^{(k+1)} &= 0.6 - 0.2\, x_1^{(k)} - 0.4\, x_2^{(k)} \end{aligned}$$

οπου k = 0,1,2,..., δηλαδη

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}^{(k+1)} = \begin{pmatrix} 0 & -0.25 & 0.125 \\ -0.25 & 0 & -0.25 \\ -0.2 & -0.4 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}^{(k)} + \begin{pmatrix} 0.5 \\ 0.5 \\ 0.6 \end{pmatrix}$$

Οι ζητούμενες νορμ του πινακα M του δευτερου μερους ειναι:

$7M7_\infty$ = max(|-0.25| + |0.125|,|-0.25| + |-0.25|, |-0.2| + |-0.4|)
= max(0.375 , 0.5 , 0.6) = 0.6 < 1
$7M7_1$ = max(|-0.25| + |-0.2|,|-0.25| + |-0.4|, |0.125| + |-0.25|)
= max(0.45 , 0.65 , 0.375) = 0.65 < 1

Με αρχικο διανυσμα το $x^{(0)} = \theta$ προκυπτει τωρα η εξης ακολουθια διανυσματων οπου τα αποτελεσματα εχουν για λογους εποπτικοτητας γραφτει με πεντε δεκαδικα ψηφια μετα την υποδιαστολη, αλλα οι πραξεις εγιναν με ακριβεια δωδεκα δεκαδικων ψηφιων:

k	0	1	2	3	4	5	...	10	11
$x_1^{(k)}$	0	0.5	0.45	0.48125	0.47438	0.47867	...	0.47826	0.47826
$x_2^{(k)}$	0	0.5	0.225	0.31250	0.27469	0.28672	...	0.28262	0.28261
$x_3^{(k)}$	0	0.6	0.3	0.42	0.37875	0.39525	...	0.39131	0.39130

β. Με βαση τον τυπο (5.5) με k = 5 εχουμε τωρα:

$$\left\|x^{(5)} - s\right\|_p \le \frac{\|M\|_p}{1-\|M\|_p}\left\|x^{(5)} - x^{(4)}\right\|_p \quad , p = \infty, 1$$

οπου $x^{(5)} - x^{(4)} = (0.00429 , 0.01203 , 0.0165)^T$

Αρα, για i = 1,2,3:

$$|x_i^{(5)} - s_i| \beta \max_{j=1,2,3}|x_j^{(5)} - s_j| = 7x^{(5)} - s7_\infty \beta \frac{\|M\|_\infty}{1-\|M\|_\infty}\left\|x^{(5)} - x^{(4)}\right\|_\infty = \frac{0.6}{0.4}0.0165$$

= 0.02475,

ενω

$$|x_1^{(5)} - s_1| + |x_2^{(5)} - s_2| + |x_3^{(5)} - s_3| = 7x^{(5)} - s7_1 \beta \frac{\|M\|_1}{1-\|M\|_1}\left\|x^{(5)} - x^{(4)}\right\|_1 =$$

$$= \frac{0.65}{0.35}0.03282 = 0.06095$$

Οπως βλεπουμε, καθε νορμ μας δινει μιαν διαφορετικη πληροφορια. Η νορμ μεγιστου προσδιοριζει το μεγιστο δυνατο σφαλμα κατα συνιστωσα, ενω η νορμ αθροισματος μας δινει ενα ανω φραγμα του αθροισματος των σφαλματων ανα συνιστωσα.

γ. Για να εχουμε ακριβεια τεσσαρων δεκαδικων ψηφιων μετα την υποδιαστολη για καθε συνιστωσα πρεπει να ειναι

$$\max_{j=1,2,3} |x_j^{(k)} - s_j| = 7x^{(k)} - s7_\infty \beta \varepsilon = 0.5\ 10^{-4}$$

Συμφωνα με τον τυπο (5.5) αρκει επομενως να ισχυει

$$\frac{\|M\|_\infty^{\ k}}{1-\|M\|_\infty}\left\|x^{(1)} - x^{(0)}\right\|_\infty \le \varepsilon$$

δηλαδη,

$$\frac{(0.6)^k}{0.4}0.6 \le 0.5\cdot 10^{-4}$$

Η σχεση αυτη μας δινει k ≥ 20,18 δηλαδη k ≥ 21 , πραγμα που σημαινει οτι *το αργότερο* μετα την εικοστη πρωτη επαναληψη θα εχει επιτευχθει η ζητουμενη ακριβεια. Οπως φαινεται απο τον παρατιθεμενο πινακα τιμων, η ακριβεια αυτη επιτυγχανεται ηδη μετα την ενδεκατη επαναληψη, γιατι μετα απο αυτην εχει ηδη σταθεροποιηθει το τεταρτο δεκαδικο ψηφιο.

5.1.2 Μέθοδος Μερικών Βημάτων (Gauss - Seidel)

Η μεθοδος Μερικων η Μεμονωμενων Βηματων για τον προσδιορισμο της λυσης του συστηματος (5.1) βασιζεται στην ιδεα οτι ενδεχομενα η συγκλιση της επαναληπτικης διαδικασιας επιταχυνεται, αν για τον υπολογισμο καθε νεας τιμης μιας συνιστωσας $x_i^{(k)}$ χρησιμοποιουνται οχι μονον οι τιμες των συνιστωσων που προεκυψαν κατα την προηγουμενη επαναληψη, $x_i^{(k-1)}$, k = 1(1)n, αλλα και οι νεες τιμες που εχουν ηδη προκυψει για τις προηγουμενες συνιστωσες, $x_j^{(k)}$, για j=1 ως i-1. Ετσι προκυπτει η επαναληπτικη ακολουθια:

$$x_i^{(k)} = \sum_{j=1}^{i-1} m_{ij} x_j^{(k)} + \sum_{j=i}^{n} m_{ij} x_j^{(k-1)} + r_i, \quad i = 1(1)n \qquad (5.13)$$

οπου k= 1, 2, 3, ... Δηλαδη συνοπτικωτερα:

$$x^{(k)} = Lx^{(k)} + Rx^{(k-1)} + r, \; k = 1,2,.. \qquad (5.14)$$

οπου L ειναι ο αριστερος (Left) τριγωνικος υποπινακας του M, που συμπιπτουν τα στοιχεια του κατω απο την κυρια διαγωνιο με αυτα του M ενω τα υπολοιπα ειναι μηδενικα και R ειναι ο δεξιος (Right) τριγωνικος υποπινακας του M, που συμπιπτει μ' αυτον στην διαγωνιο και πανω απο αυτην, ενω κατω της εχει μηδενικα στοιχεια (M = L+R).

Για ωρισμενες κατηγοριες πινακων (οχι ομως για ολους τους πινακες) μπορει πραγματι να δειχθει οτι η μεθοδος αυτη συγκλινει ταχυτερα (Βλ. Varga: Matrix Iterative Analysis σελ. 70-71).

Κριτηρια για την συγκλιση της μεθοδου Μερικων Βηματων δινονται απο τα ακολουθα θεωρηματα:

Θεώρημα 5.3 *(Κριτηριο Sassenfeld)*

Αν για τον πινακα M ισχύει:

$$q = \max_{i=1(1)n} |q_i| < 1 \qquad (5.15)$$

οπου

$$q_1 = \sum_{j=1}^{n} |m_{ij}|, \quad q_i = \sum_{j=1}^{i-1} |m_{ij}|\, q_j + \sum_{j=i}^{n} |m_{ij}|, \quad i = 2(1)n \qquad (5.16)$$

τοτε το συστημα (5.1) εχει μονοσημαντα ωρισμενη λυση s και η ακολουθια (5.14) συγκλινει για $x^{(0)}$ τυχον προς αυτην, ενω ισχυουν οι εκτιμησεις σφαλματος:

$$\left\| x^{(k)} - s \right\|_\infty \le \frac{q}{1-q} \left\| x^{(k)} - x^{(k-1)} \right\|_\infty \le \frac{q^k}{1-q} \left\| x^{(1)} - x^{(0)} \right\|_\infty \qquad (5.17)$$

Απόδειξη: Βλεπε Stummel - Heiner σελ. 153-154.

Απλουστερο, αλλα οχι τοσο ακριβες κριτηριο συγκλισης δινεται απο το ακολουθο θεωρημα:

Θεώρημα 5.4

Αν ειναι

$$\|M\|_\infty<1 \qquad (5.18)$$

τοτε το συστημα (5.1) εχει μονοσημαντα ωρισμενη λυση s και η ακολουθια Gauss - Seidel (5.13) η (5.14) συγκλινει για $x^{(0)}$ τυχον προς την λυσην αυτη. Τοτε ισχυουν επισης οι εκτιμησεις σφαλματος:

$$\left\|x^{(k)}-s\right\|_\infty \le \frac{\|M\|_\infty}{1-\|M\|_\infty}\left\|x^{(k)}-x^{(k-1)}\right\|_\infty \le \frac{\|M\|_\infty^{\ k}}{1-\|M\|_\infty}\left\|x^{(1)}-x^{(0)}\right\|_\infty \qquad (5.19)$$

Απόδειξη

Το θεωρημα αυτο προκυπτει αμεσα απο το προηγουμενο αν προσεξουμε οτι για q < 1 ειναι q < 7M7$_\infty$.

Για να εφαρμοσουμε την μεθοδο Μερικων Βηματων για την επιλυση ενος συστηματος της μορφης Ax = b το μετασχηματιζουμε παλι στην μορφη (5.9), οποτε προκυπτει:

$$x_i^{(k)} = \frac{1}{a_{ii}}(b_i - \sum_{j=1}^{i-1} a_{ij}x_j^{(k)} - \sum_{j=i}^{n} a_{ij}x_j^{(k-1)}), \quad i = 1(1)n \qquad (5.20)$$

για k = 1,2,3,....

Με βαση την ανισοτητα (5.18) προκυπτει παλι σαν ικανη συνθηκη για την συγκλιση της μεθοδου το να ισχυει η (5.12), δηλ. το να υπαρχει διαγωνια υπεροχη.

Για να επιτευχθει ταχυτερη συγκλιση της επαναληπτικης διαδικασιας χρησιμοποιειται συχνα η ακολουθη παραλλαγη που λεγεται *Μέθοδος Χαλαρώσεως (Relaxation Method)*:

$$x_i^{(k)} = (1-\sigma)x_i^{(k-1)} + \frac{\sigma}{a_{ii}}(b_i - \sum_{j=1}^{i-1} a_{ij}x_j^{(k)} - \sum_{j=i}^{n} a_{ij}x_j^{(k-1)}), \quad i = 1(1)n \qquad (5.21)$$

οπου σ ειναι παραμετρος που την προσδιοριζουμε ετσι ωστε να εχουμε οσο το δυνατον ταχυτερη συγκλιση. Η μεθοδος αυτη συμπιπτει με την μεθοδο Μερικων Βηματων για σ = 1.

Αν ο πινακας A ειναι συμμετρικος (η γενικωτερα Ερμητιανος) και θετικα προσδιορισμενος (δηλ. αν εχει μονο θετικες ιδιοτιμες) τοτε αποδεικνυεται οτι η Μεθοδος Χαλαρωσεως δεν μπορει να συγκλινει αν το σ ειναι εξω απο το διαστημα 0 < σ < 2 (Βλ. Stummel - Hainer, σελ 159, Blum, σελ.146 η Burden - Faires σελ. 435). Ο προσδιορισμος της βελτιστης τιμης για το σ ειναι ομως δυσκολο θεμα και αποτελει αντικειμενο ευρυτατης ερευνητικης προσπαθειας (Βλ. π.χ. Varga, Gastinel, Smith).

Παράδειγμα Εφαρμογής της Μεθόδου Μερικών Βημάτων

α. Με την μεθοδο Μερικων Βηματων με αρχικο διανυσμα $x^{(0)} = \theta$ να λυθει προσεγγιστικα το συστημα του προηγουμενου παραδειγματος.

β. Μετα την πεμτη επαναληψη να γινει εκτιμηση του σφαλματος.

γ. Ποσες το πολυ επαναληψεις χρειαζονται για να καθορισθουν ολες οι συντεταγμενες με ακριβεια οκτω δεκαδικων ψηφιων μετα την υποδιαστολη, δηλ. με σφαλμα μικροτερο απο $0.5\ 10^{-8}$;

Λύση

Συμφωνα με τον τυπο (5.20) εχουμε την επαναληπτικη διαδικασια:

$$x_1^{(k+1)} = 0.5 - 0.25\, x_2^{(k)} + 0.125\, x_3^{(k)}$$
$$x_2^{(k+1)} = 0.5 - 0.25\, x_1^{(k+1)} - 0.25\, x_3^{(k)}$$
$$x_3^{(k+1)} = 0.6 - 0.2\, x_1^{(k+1)} - 0.4\, x_2^{(k+1)}$$

οπου k = 0,1,2,..., δηλαδη

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}^{(k+1)} = \begin{pmatrix} 0 & 0 & 0 \\ -0.25 & 0 & 0 \\ -0.2 & -0.4 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}^{(k+1)} + \begin{pmatrix} 0 & -0.25 & 0.125 \\ 0 & 0 & -0.25 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}^{(k)} + \begin{pmatrix} 0.5 \\ 0.5 \\ 0.6 \end{pmatrix}$$

Προς επιβεβαιωση της συγκλισης μπορουμε εδω να χρησιμοποιησουμε με βαση τα θεωρηματα 5.3 και 5.4 μονο την νορμ $7M7_\infty$ (ή τον συντελεστη q του Sassenfeld):

$\|M\|_\infty = \max(|-0.25| + |0.125|, |-0.25| + |-0.25|, |-0.2| + |-0.4|)$

$= \max(0.375, 0.5, 0.6) = 0.6 < 1$

$q_1 = 0.375$, $q_2 = 0.25\, q_1 + 0.25 = 0.34375$, $q_3 = 0.2\, q_1 + 0.4\, q_2 = 0.2125$

$q = \max_{v=1,2,3} q_v = 0{,}375$

Με αρχικο διανυσμα το $x^{(0)} = \theta$ προκυπτει τωρα η εξης ακολουθια

k	0	1	2	3	4	5	6	7
$x_1^{(k)}$	0	0.5	0.45	0.47375	0.47794	0.47827	0.47827	0.47826
$x_2^{(k)}$	0	0.375	0.3	0.28406	0.28261	0.28259	0.28261	0.28261
$x_3^{(k)}$	0	0.35	0.39	0.39163	0.39137	0.39131	0.39130	0.39130

οπου τα αποτελεσματα στρογγυλευτηκαν σε πεντε δεκαδικα ψηφια, αλλα οι πραξεις εχουν γινει με χρηση δωδεκα δεκαδικων. Συγκρινοντας αυτα τα αποτελεσματα με εκεινα της Μεθοδου Ολικων Βηματων βλεπουμε οτι εδω εχουμε φθασει ηδη κατα την εβδομη επαναληψη την ακριβεια που προηγουμενως ειχαμε πετυχει με δωδεκα επαναληψεις.

β. Με βαση τους τυπους (5.17) και (5.19) με k = 5 εχουμε τωρα:

$$\left\|x^{(5)} - s\right\|_\infty \le \frac{d}{1-d}\left\|x^{(5)} - x^{(4)}\right\|_\infty$$

οπου μπορουμε να βαλλουμε d = q η d = $\|M\|_\infty$, ενώ

$$x^{(5)} - x^{(4)} = (0.00033\ ,-0.00002\ ,-0.0006)$$

Αρα, για i = 1,2,3:

$|x_i - s_i| \le \max_{j=1,2,3} |x_j - s_j| = \|x - s\|_\infty \le \frac{d}{1-d}\ \|x^{(5)} - x^{(4)}\|_\infty =$

$= (0.6/0.4)\ 0{,}00033 = 0{,}0005$ για d = $7M7_\infty = 0{,}6$

= {

$= (0.375/0.625)\ 0{,}00033 = 0{,}0002$ για d = q = 0,375

γ. Για να εχουμε ακριβεια οκτω δεκαδικων ψηφιων μετα την υποδιαστολη για καθε συνιστωσα πρεπει να ειναι

$\max_{j=1,2,3} |x_j - s_j| = 7x - s7_\infty\ \beta\ \varepsilon = 0.5\ 10^{-8}$

Συμφωνα με τον τυπο (5.17) αρκει επομενως να ισχυει

$$\frac{q^k}{1-q}\left\|x^{(1)} - x^{(0)}\right\|_\infty\ \beta\ \varepsilon$$

δηλαδη,

$$\frac{(0.375)^k}{0.625} 0.5 \le 0.5\, 10^{-8}$$

και επομενως αρκει να ειναι k ≥ 19.26 , δηλαδη k ≥20.

Μέθοδοι Απαλοιφής του Gauss

Η σημαντικωτερη μεθοδος για την λυση συστηματων με πληρη πινακα, δηλαδη πινακα που οι περισσοτεροι συντελεστες του ειναι διαφορετικοι απο το μηδεν, ειναι η μεθοδος απαλοιφης του Gauss. Βασικη ιδεα αυτης της μεθοδου ειναι ο σταδιακος μετασχηματισμος ενος συστηματος με τετραγωνικο πινακα:

$$\begin{aligned}
&(1):\quad a_{11}x_1 + a_{12}x_2 + \dots a_{1,n-1}x_{n-1} + a_{1n}x_n = a_1\\
&(2):\quad a_{21}x_1 + a_{22}x_2 + \dots a_{2,n-1}x_{n-1} + a_{2n}x_n = a_2\\
&\qquad\dots\dots\dots\\
&(n):\quad a_{n1}x_1 + a_{n2}x_2 + \dots a_{n,n-1}x_{n-1} + a_{nn}x_n = a_n
\end{aligned} \tag{5.22}$$

σε ισοδυναμο συστημα με ανω τριγωνικο πινακα:

$$\begin{aligned}
b_{11}x_1 + b_{12}x_2 + \dots b_{1,n-1}x_{n-1} + b_{1n}x_n &= b_1\\
b_{22}x_2 + \dots b_{2,n-1}x_{n-1} + b_{2n}x_n &= b_2\\
\dots\\
b_{n-1,n-1}x_{n-1} + b_{n-1,n}x_n &= b_{n-1}\\
b_{nn}x_n &= b_n
\end{aligned} \tag{5.23}$$

με απαλοιφη των συντελεστων που βρισκονται κατω απο την κυρια διαγωνιο.

Επιλυοντας την τελευταια εξισωση ως προς x_n , την προτελευταια ως προς

x_{n-1}, κοκ, μπορουμε στην συνεχεια να βρουμε τις τιμες ολων των αγνωστων.

Η απαλοιφη των συντελεστων που βρισκονται κατω απο την κυρια διαγωνιο γινεται κατα στηλες με τον ακολουθο τροπο:

Αν ειναι $a_{11}\neq 0$ για να απαλειψουμε το x_1 απο τις εξισωσεις (2) ως (n) με την βοηθεια της (1) αρκει να τις αντικαταστησουμε με τις (2') ως (n') οπου:

$$(k') = (k) - \frac{a_{k1}}{a_{11}}(1), \quad k = 2(1)n \tag{5.24}$$

Ο συμβολισμος αυτος σημαινει απλα οτι η εξισωση (k') προκυπτει με πολλαπλασιασμο ολων των ορων της (1) επι a_{k1}/a_{11} και αφαιρεση της κατα μελη απο την (k). Το αποτελεσμα ειναι προφανως:

$$(k'):\ 0x_1 + (a_{k2} - a_{12}\frac{a_{k1}}{a_{11}})x_2 + \ldots + (a_{kn} - a_{1n}\frac{a_{k1}}{a_{11}})x_n = a_k - a_1\frac{a_{k1}}{a_{11}} \tag{5.25}$$

δηλαδη μια εξισωση που δεν περιεχει το x_1.

Τα δυο συστηματα, (1),(2'),..,(n') και (1),(2),..,(n) ειναι ισοδυναμα, γιατι οι παραπανω μετασχηματισμοι ειναι αντιστρεψιμοι, και επομενως καθε λυση του ενος λυση και του αλλου.

Για να συνεχισουμε την διαδικασια τριγωνοποιησης παρατηρουμε οτι οι εξισωσεις (2'),..,(n') αποτελουν υποσυστημα n-1 εξισωσεων με n-1 αγνωστους, που μπορει να λυθει ανεξαρτητα απο την (1) για να προσδιορισθουν οι τιμες των αγνωστων x_2 ,..,x_n. Στο υποσυστημα αυτο η πρωτη απο τις εξισωσεις, (2') μπορει να χρησιμοποιηθει, οπως ακριβως και παραπανω, για απαλοιφη του x_2 απο τις υπολοιπες, υπο την προυποθεση οτι $a'_{22}\neq 0$. Αν $a_{22}' = 0$, τοτε αρκει να αντιμεταθεσουμε τις εξισωσεις (2') και (3') και να επαναλαβουμε την διαδικασια απαλειψης.

Αν σε καποιο απο τα υποσυστηματα που παραγονται με τις διαδοχικες απαλοιφες μεταβλητων ολοι οι συντελεστες μιας μεταβλητης ειναι ισοι με μηδεν αυτο σημαινει οτι το υποσυστημα αυτο, αρα και το αρχικο, δεν δινει συγκεκριμενη τιμη στην μεταβλητη αυτην και επομενως, ειτε ειναι αοριστο, ειτε ασυνεπες (δηλ. με εξισωσεις μη συμβιβαστες μεταξυ τους).

Κατα την εφαρμογη της μεθοδου πρεπει να προσεξουμε δυο σημεια:

1. Ο Η/Υ κραταει στην μνημη του πεπερασμενον μονο αριθμο δεκαδικων ψηφιων καθε συντελεστη κοβοντας, αν ειναι αναγκη, τα τελευταια δεκαδικα ψηφια. Αν τωρα , π.χ., οι συντελεστες της (2) ειναι πολυ μικροτεροι απο αυτους της (1) μπορει κατα την μεταβαση στην (2') να εξαφανισθουν κατα τις στρογγυλοποιησεις τα περισσοτερα σημαντικα δεκαδικα ψηφια των συντελεστων της (2), πραγμα που ισοδυναμει με αισθητην παραποιηση της αρχικης εξισωσης (2) και επομενως και της λυσης του συστηματος.

Δηλαδη, για να περιορισουμε οσο το δυνατον περισσοτερο την επιρροη των σφαλματων στρογγυλεματος στο τελικο αποτελεσμα, πρεπει να φροντιζουμε να ειναι, τοσο οι συντελεστες του αρχικου συστηματος, οσο και ολοι οι συντελεστες του συστηματος που προκυπτει μετα απο την απαλοιφη μιας μεταβλητης, αν ειναι δυνατο στην ιδια ταξη μεγεθους. Για να γινει αυτο, αρκει, π.χ., να προσδιοριζουμε καθε φορα σε ποια δεκαδικη ταξη μεγεθους βρισκεται ο απολυτα μεγαλυτερος συντελεστης καθε εξισωσης και να πολλαπλασιαζουμε ολους τους συντελεστες με την δυναμη του 10 που φερνει

αυτον τον συντελεστη στην ταξη των μοναδων. Η διαδικασια αυτη λεγεται κλιμακωση (scaling) και ο αντιστοιχος συντελεστης λεγεται "συντελεστης κλιμακωσης" (scaling factor).

Αν μετα απο αυτην την κλιμακωση κατα σειρες οι συντελεστες καποιας στηλης διαφερουν κατα πολλες ταξεις μεγεθους απο την ταξη των μοναδων, μπορουμε να τους φερουμε στην ταξη αυτην αντικαθιστωντας την μεταβλητη που αντιστοιχει σε αυτην την στηλη με καταλληλο υποπολλαπλασιο της. Αν, π.χ., ολοι οι συντελεστες της μεταβλητης x_3 ειναι στην ταξη του 10^{-5}, εισαγοντας την νεα μεταβλητη $x_3' = 10^{-5}x_3$, εχουμε συντελεστες της x_3' στην ταξη των μοναδων.

2. Οι συντελεστες a_{11} που χρησιμοποιουμε σαν παρονομαστες κατα την απαλοιφη (5.24),(5.25) πρεπει να ειναι διαφορετικοι απο το μηδεν και μαλιστα με οσο το δυνατο μεγαλυτερη απολυτη τιμη, ωστε κατα την διαιρεση να μην μεγαλωνουν υπερμετρα τα πηλικα a_{k1}/a_{11}. Αυτο μπορει εξασφαλισθει με αντιμεταθεση των εξισωσεων του καθε υποσυστηματος ετσι ωστε στην πρωτη θεση να ερθει η εξισωση με τον απολυτα μεγαλυτερο συντελεστη της μεταβλητης που θελουμε να απαλειψουμε. Ετσι, ολα αυτα τα πηλικα ειναι μικροτερα απο την μοναδα. Αν τωρα οι συντελεστες των εξισωσεων (k) και (1) εχουν ηδη αναχθει στην ταξη μεγεθους των μοναδων με καταλληλη κλιμακωση, τοτε και οι συντελεστες των εξισωσεων (k') που προκυπτουν κατα την απαλοιφη θα ειναι επισης στην ταξη των μοναδων. Το ιδιο κριτηριο, δηλαδη η επιλογη του απολυτα μεγιστου, εφαρμοζεται και κατα την επιλογη των υπολοιπων συντελεστων a_{22}', a_{33}'', .. που χρησιμευουν για την απαλοιφη της αντιστοιχης μεταβλητης σε καθε νεο υποσυστημα που προκυπτει. Η διαδικασια αυτη λεγεται "απαλοιφη με μερικη εναλλαγη", αφου εναλλασσουμε τις σειρες ετσι ωστε εκεινη με τον απολυτα μεγαλυτερο συντελεστη του αγνωστου που θελουμε να απαλειψουμε να πηγαινει στην πρωτη θεση του υποσυστηματος που εξεταζουμε. Ο ορος "μερικη" μπαινει εδω για αντιδιαστολη αυτης της διαδικασιας με την διαδικασια απαλοιφης με "πληρη εναλλαγη", που δεν απαλοιφει τους αγνωστους κατα σειραν, αλλα διαλεγει καθε φορα απο τον απομενοντα υποπινακα το απολυτα μεγαλυτερο στοιχειο και απαλειφει τον αγνωστο που τυχαινει να εχει αυτον τον συντελεστη.

Παράδειγμα της Επιρροής του τρόπου Απαλοιφής στο Αριθμητικό Αποτέλεσμα

(Βλέπε: D. Kincaid-W.Cheney: Numerical Analysis, Brooks/Cole, 1990, σελ. 143-145).

Αν λύσουμε το σύστημα:

$$(1): \varepsilon x + y=1$$
$$(2): x+y=2$$

με την απαλοιφή: (2′)=(2)-(1/ε)(1) προκύπτει:

$$(1-1/\varepsilon)y=2-1/\varepsilon$$

Αν η ποσότητα ε είναι πολύ μικρή τότε η 1/ε είναι πολύ μεγάλη και για τον Η/Υ είναι: $2-1/\varepsilon \approx -1/\varepsilon$, $1-1/\varepsilon \approx -1/\varepsilon$, οπότε κατά την διαίρεσή τους προκύπτει:

$$y \approx 1,$$

Όμως, με y=1 προκύπτει από την (1):

$$x=(1-y)/\varepsilon \approx (1-1)/\varepsilon=0/\varepsilon=0,$$

ενώ η ακριβής λύση είναι:

$$x=1/(1-\varepsilon)\approx 1,\ y=(1-2\varepsilon)/(1-\varepsilon)\approx 1$$

Αντίθετα, η απαλοιφή (1′)=(1)-ε(2) μας δίνει:

$$(1-\varepsilon)y=1-2\varepsilon,\ \ y=(1-2\varepsilon)/(1-\varepsilon)\approx 1,$$

και η (2) δίνει τότε: $x=2-y\approx 2-1=1$, που είναι καλή προσέγγιση της λύσης για ε πολύ μικρό.

Παραλλαγη της μεθοδου απαλοιφης του Gauss ειναι η αναγωγη της λυσης σε αναλυση του πινακα σε γινομενο τριγωνικων πινακων, καθως και η ειδικωτερη περιπτωση αυτης της μεθοδου για συμμετρικους πινακες που λεγεται μεθοδος Cholesky, για τις οποιες θα μιλησουμε αργοτερα.

Παράδειγμα Εφαρμογής απαλοιφής Gauss

Να λυθει με την μεθοδο απαλοιφης του Gauss το συστημα:

$$\begin{aligned}(1):&\quad 8x_1+2x_2-x_3=4\\(2):&\quad x_1+4x_2+x_3=2\\(3):&\quad x_1+2x_2+5x_3=3\end{aligned}$$

Λύση

Με χρηση μερικης απαλοιφης απαλειφουμε κατ'αρχην το x_1 απο τις (2) και (3) με την βοηθεια της (1) που εχει τον απολυτα μεγαλυτερο συντελεστη για τον x_1. Έτσι προκυπτει:

$$(2') = (2) - \frac{1}{8}(1):\quad 3.75x_1+1.125x_2=1.5$$

$$(3') = (3) - \frac{1}{8}(1):\quad 1.75x_1+5.125x_2=2.5$$

Με απαλοιφη του x_2 απο την (3') προκυπτει παρομοια:

$$(3'') = (3') - \frac{1.75}{3.75}(2'):\quad 4.6x=1.8$$

Τελικα εχουμε, δηλαδη, το συστημα:

$$\begin{aligned}8x_1+2.00x_2-1.000x_3&=4\\3.75x_2+1.125x_3&=1.5\\4.6x_3&=1{,}8\end{aligned}$$

που η λυση του ειναι:

$$x_1=0.391304348\quad x_2=0.282608696\quad x_3=0.47826087$$

Αν εδω θελαμε να κανουμε ολικη εναλλαγη αντι για μερικη εναλλαγη, η μεν πρωτη απαλοιφη θα γινοταν οπως ακριβως παραπανω, αλλα στην συνεχεια θα απαλειφαμε, οχι τον αγνωστο x_2 , αλλα τον x_3 γιατι αυτος εχει στις απομενουσες δυο εξισωσεις τον απολυτα μεγαλυτερο συντελεστη. Δηλαδη θα ειχαμε:

$$\begin{aligned}(3'):&\ 5.125x_3+1.75x_2=2.5\\(2'):&\ 1.125x_3+3.75x_2=1.5\end{aligned}$$

και

$$(2'') = (2')\frac{1.125}{5.125}(3') : (3.75 - \frac{1.125}{5.125}1.75)x_2 = 1.5 - \frac{1.125}{5.125}2.5$$

δηλαδη

(2") : 3.365853659 x_2 = 0.951219512

αρα x_2 = 0.282608695 κοκ.

Παρατήρηση: Αντιστροφή ενός Πίνακος

Με την μεθοδο απαλοιφης Gauss μπορουμε επισης να αντιστρεψουμε εναν πινακα χρησιμοποιωντας το γεγονος οτι αν $b^{(i)}$ ειναι η i στηλη του αντιστροφου πινακα, A := ($b^{(1)}$, $b^{(2)}$,.., $b^{(n)}$), τοτε ειναι $Ab^{(i)}$ = $e^{(i)}$ οπου $e^{(i)}$ το μοναδιαιο διανυσμα στην διευθυνση της συντεταγμενης $x^{(i)}$. Για να βρουμε τον αντιστροφο πινακα αρκει, δηλαδη, να λυσουμε τα συστηματα $Ab^{(i)}$ = $e^{(i)}$, i = 1(1)n. Τα συστηματα αυτα μπορουν να λυθουν μαλιστα ταυτοχρονα αφου το πρωτο μερος τους ειναι ιδιο και μετασχηματιζεται αρα παντα με τον ιδιο τροπο.

5.3 Ασταθή συστήματα – Βελτίωση Προσεγγιστικής Λύσης

Τα σφαλματα που προκυπτουν κατα την εφαρμογη της μεθοδου απαλοιφης του Gauss ειναι αποκλειστικα σφαλματα στρογγυλεματος και οχι κυριως μεθοδικα, οπως ειναι κατα την εφαρμογη επαναληπτικων μεθοδων. Η επιδραση τους στα αποτελεσματα που προκυπτουν δεν ειναι ομως παντα αμελητεα. Ας υποθεσουμε, π.χ., οτι το υποσυστημα 2×2 στο οποιο καταλήγουμε με απαλοιφη ολων των αλλων μεταβλητων ενος μεγαλυτερου συστηματος 100×100 ειναι:

$$x_{99} + 2\,x_{100} = 5$$
$$2x_{99} + 4.001\,x_{100} = 3 \qquad (5.26)$$

Η λυση αυτου του συστηματος ειναι:

$$x_{99} = 14005\ ,\quad x_{100} = -7000$$

Αν τώρα λύσουμε το ίδιο σύστημα 100 εξισώσεων σε άλλον Η/Υ ή με χρήση διαφορετικού προγράμματος, ενδέχεται λόγω κάπως διαφορετικών σφαλμάτων στρογγυλεύσεως να καταλήξουμε στο κάπως παραλλαγμένο υποσύστημα 2×2:

$$x_{99} + 2\,x_{100} = 5$$
$$2x_{99} + 3.999\,x_{100} = 3 \qquad (5.27)$$

που διαφέρει από το προηγούμενο μόνο κατά 0.002 σε έναν συντελεστή. Η λυση ομως αυτου του συστηματος ειναι:

$$x_{99} = -13995\ ,\quad x_{100} = +7000$$

που έχει τεράστιες διαφορές απο την προηγουμενη όχι μονο ως προς τις αριθμητικες τιμες, αλλα ακομα και ως προς τα προσημα.

Γιατι αυτη η μεγαλη διαφορα των δύο λυσεων;

Η αιτια αυτου του φαινομενου γινεται φανερη αν χρησιμοποιησουμε γραφικη παρασταση των λυσεων: Καθε εξισωση ενος γραμμικου συστηματος 2×2 παριστανει μιαν ευθεια, ενω η λυση του συστηματος ειναι το σημειο τομης των δυο ευθειων. Στην προκειμενη περιπτωση, οι ευθειες του συστηματος (5.26) ειναι σχεδον παραλληλες, δηλαδη τεμνονται υπο μιαν πολυ μικρη γωνια. Αυτο εχει σαν

συνεπεια οτι μια μικρη μετατοπιση της μιας απο αυτες τις ευθειες μπορει να μετατοπισει σημαντικα το σημειο τομης τους.

Αναλογα ισχυουν και για ενα συστημα 3×3 η γενικωτερα, n×n εξισωσεων. Καθε εξισωση ενος τετοιου συστηματος παριστανει ενα επιπεδο, η αντιστοιχα "υπερεπιπεδο", και η λυση του συστηματος ειναι το κοινο σημειο τομης των επιπεδων η υπερεπιπεδων. Αν εδω δυο επιπεδα η υπερεπιπεδα σχηματιζουν μεταξυ τους μικρην γωνια προκυπτει προφανως παλι ασταθεια της λυσης. Συστηματα τετοιου ειδους λεγονται "ασταθη" η "κακης καταστασεως" (ill conditioned).

Η διαγνωση της ασταθειας ενος συστηματος δεν ειναι ευκολη. Παραδειγματος χαριν, το συστημα που δινεται απο τον Kahan:

$$\begin{Bmatrix} 0.2161x_1 + 0.1441x_2 = 0.144 \\ 1.2969x_1 + 0.8648x_2 = 0.8642 \end{Bmatrix} \quad (5.28)$$

εχει την λυση $x_1 = 2$, $x_2 = -2$. Η επιλυση του ομως με την μεθοδο απαλοιφης Gauss δινει σαν λυση $x_1 = 1.333179$, $x_2 = -1$, οποτε το διανυσμα σφαλματος r:= Ax - b εχει τις συνιστωσες $r_1 = -1.49\ 10^{-8}$, $r_2 = -1.192\ 10^{-7}$. Το οτι οι ευθειες που αντιστοιχουν στις εξισωσεις του συστηματος τεμνονται υπο μιαν πολυ μικρη γωνια δεν φαινεται εδω αμεσα. Μπορουμε βεβαια να το διαπιστωσουμε, π.χ., αν διαιρεσουμε ολους τους συντελεστες καθε εξισωσης με τον συντελεστη του δευτερου μερους. Τοτε το συστημα παιρνει την μορφη:

$$\begin{Bmatrix} 1.500695x_1 + 1.000695x_2 = 1 \\ 1.500694x_1 + 1.000694x_2 = 1 \end{Bmatrix} \quad (5.29)$$

απο οπου φαινεται οτι οι ευθειες ειναι σχεδον παραλληλες. Αυτο ομως δεν θα ηταν αμεσα αντιληπτο αν το συστημα ηταν μεγαλυτερων διαστασεων.

Ενα κριτηριο διαγνωσης της ασταθειας προκυπτει με τον ακολουθο τροπο:

Εστω οτι αντι για την αυστηρη λυση x ενος συστηματος Ax = b ειναι γνωστη μια προσεγγιστικη λυση x+h. Αν ειναι:

$$A(x+h) = b+r \quad (5.29)$$

τοτε προκυπτει:

$$Ah = r \quad \text{ή} \quad h = A^{-1}r$$

Αν τωρα χρησιμοποιησουμε μιαν οποιαδηποτε διανυσματικη νορμ και την απο αυτην επαγομενη νορμ πινακων, προκυπτει:

$$\frac{\|h\|}{\|r\|} = \frac{\|A^{-1}r\|}{\|r\|} \le \|A^{-1}\| \quad (5.30)$$

οπου η ισοτητα ειναι δυνατη για καταλληλο r (η h) γιατι εξ ορισμου ειναι:

$$\|A^{-1}\| = \max_{x \neq \theta} \frac{\|A^{-1}z\|}{\|z\|} \quad (5.31)$$

Παρομοια ισχυει:

$$\frac{\|b\|}{\|x\|} = \frac{\|Ax\|}{\|x\|} \le \|A\| \quad (5.32)$$

οπου η ισοτητα ειναι δυνατη για καταλληλο x, η αντιστοιχα, για καταλληλο b. Πολλαπλασιαζοντας τις (5.30) και (5.32) κατα μελη εχουμε:

$$\frac{\|h\|}{\|x\|} \le \|A\|\|A^{-1}\|\frac{\|r\|}{\|b\|} \qquad (5.33)$$

Απο αυτην την σχεση προκυπτει οτι, εστω και αν η ποσοστιαια μεταβολη του δευτερου μερους των εξισωσεων, ||r||/||b||, ειναι μικρη, η ποσοστιαια μεταβολη της λυσης, 7h7/7x7, μπορει να ειναι μεγαλη αν η ποσοτητα:

$$K(A) := \|A\|\|A^{-1}\| \qquad (5.34)$$

εχει μεγαλη τιμη. Το μεγεθος αυτο λεγεται για οποιαδηποτε νορμ πινακων «δεικτης καταστασης» (condition number) του πινακα A, και ειναι παντοτε μεγαλυτερο η ισο με την μοναδα, γιατι

$$\|I\|^2 \ge \|I\ I\| = \|I\|, \text{ αρα } \|I\| \ge 1 \qquad (5.35)$$

και επομενως

$$K(A) := \|A\|\|A^{-1}\| \ge \|A\ A^{-1}\| = \|I\| \ge 1 \qquad (5.36)$$

Η σημασια του μεγεθους αυτου ειναι περισσοτερο θεωρητικη παρα πρακτικη, γιατι ο υπολογισμος του απαιτει τον προηγουμενο υπολογισμο του αντιστροφου πινακα A . Στην περιπτωση που ο υπολογισμος αυτος εχει ηδη γινει, δηλαδη αν εχουμε εναν πινακα $C \approx A^{-1}$, το μεγεθος $\|C\|\|A\| = K(A)$ ειναι προφανως μια ενδειξη του αν ο C ειναι καλη προσεγγιση του A^{-1} η οχι[5].

Παραδειγματος χαριν, για τον πινακα A του συστηματος (5.28) βρισκουμε με την βοηθεια απαλοιφης Gauss τον ακολουθο αντιστροφο:

$$A^{-1} \approx C = \begin{pmatrix} -4.475 & 0.746 \\ 6.71 & -1.118 \end{pmatrix} \times 10^7$$

οποτε $A\|\|_\infty\|C\|_\infty = 2.1617*7.829\ 10^7 = 16.924\ 10^7 >> 1$.

Η βελτιωση των αποτελεσματων της επιλυσης ενος ασταθους συστηματος μπορει να γινει με χρήση αριθμητικης διπλης ακριβειας (double precision), πραγμα που σημαινει βασικα οτι διπλασιος αριθμος δεκαδικων ψηφιων καθε ενδιαμεσου αποτελεσματος κρατειται στην *μνημη του Η/Υ. Βλέπε Κιουστελίδη: Αριθμητική Ανάλυση, 1996, σελ.105.*

5.4 Ανάλυση σε Γινόμενο Τριγωνικών Πινάκων

Παραλλαγη της μεθοδου απαλοιφης του Gauss ειναι η μεθοδος αναγωγης του πινακα A σε γινομενο τριγωνικων πινακων που εχει το πλεονεκτημα οτι χρειαζεται περιπου τον μισο χωρο μνημης απο οτι η αμεση μεθοδος απαλοιφης. Αυτο επιτρεπει, π.χ., να εξοικονομηθει μνημη για την εκτελεση των πραξεων με διπλην ακριβεια. Η μέθοδος αυτή είναι ιδιαίτερα χρήσιμη στην περίπτωση που πρέπει να λυθεί μία σειρά από γραμμικά συστήματα με τον ίδιο πίνακα αλλά διαφορετικό δεύτερο μέρος, όπως συμβαίνει:

(α) κατά την επίλυση των συστημάτων $Ax^{(i)} = e^{(i)}$, I=1(1)n, για τον προσδιορισμό των στηλών $x^{(i)}$ του αντίστροφου πίνακα A^{-1}, και

---

[5] Ένας τρόπος προσεγγιστικού υπολογισμού του K(A) περιγράφεται στα συγγράμματα [Burden-Faires,σελ.441], και [Forsythe-Moler, σελ. 49-51].

(β) στην Μέθοδο Αντιστρόφων Δυνάμεων, με την οποία προσδιορίζουμε την απόλυτα μικρότερη ιδιοτιμή ενός πίνακα.

Βαση της μεθοδου αυτης ειναι η αναλυση του πινακα A του συστηματος Ax = b σε γινομενο ενος κατω τριγωνικου πινακα L (Lower triangular) και ενος ανω τριγωνικου πινακα U (Upper triangular).

Επειδη τοτε ειναι:

$$b = Ax = LUx = L(Ux) \qquad (5.42)$$

η λυση του αρχικου συστηματος αναγεται στην διαδοχικη επιλυση δυο τριγωνικων συστηματων:

$$Lz = b,\ Ux = z \qquad (5.43)$$

Η εξοικονομηση χωρου μνημης γινεται εδω, γιατι αντι να διατεθουν θεσεις μνημης για τον πινακα A και εναν σταδιακα τριγωνοποιουμενο πινακα B, οι συντελεστες των πινακων L και U αποθηκευονται αντιστοιχα στις θεσεις των συντελεστων του κατω τριγωνικου και πανω τριγωνικου μερους του A.

Αν ειναι

$$L = \begin{pmatrix} 1 & 0 & \dots & 0 \\ l_{21} & 1 & \dots & 0 \\ \dots & \dots & \dots & 0 \\ l_{n1} & l_{n2} & \dots & 1 \end{pmatrix},\ U = \begin{pmatrix} u_{11} & u_{12} & \dots & u_{1n} \\ 0 & u_{22} & \dots & u_{2n} \\ 0 & \dots & \dots & \dots \\ 0 & 0 & \dots & u_{nn} \end{pmatrix} \qquad (5.44)$$

τοτε, συμφωνα με τον κανονα πολλαπλασιασμου πινακων, για τον προσδιορισμο των L και U εχουμε τις σχεσεις:

$$a_{ik} = \sum_{j=1}^{n} l_{ij} u_{jk} = \sum_{j=1}^{\min(i,k)} l_{ij} u_{jk} \qquad (5.45)$$

οπου το ανω οριο της αθροισης ειναι min(i,k) γιατι

$$l_{ij} = 0 \text{ για } j > i,\ u_{jk} = 0 \text{ για } j > k \qquad (5.46)$$

Το συστημα αυτο εχει n^2 εξισώσεις με $n^2 + n$ αγνωστους. Επειδη οι αγνωστοι ειναι περισσοτεροι απο οτι χρειαζεται μπορουμε να μειωσουμε τον αριθμο τους κατα n βαζοντας π.χ. την προσθετη απαιτηση:

$$l_{kk} = 1,\ k = 1(1)n \qquad (5.47)$$

Ετσι εχουμε τις σχεσεις:

$$a_{ik} = l_{i1}u_{1k} + l_{i2}u_{2k} + \dots + l_{jv}u_{vk},\ v = \min(i,k) \qquad (5.48)$$

Από τις σχέσεις αυτές προκύπτει, π. χ. για n=3:

i=1: $a_{1k}=l_{11}u_{1k} = \mathbf{u_{1k}}$, k=1,2,3 k=1: $a_{i1}=\mathbf{l_{i1}}u_{11}$, i=1,2,3

i=2: $a_{2k}=l_{21}u_{1k}+l_{22}u_{2k}=l_{21}u_{1k}+\mathbf{u_{2k}}$, k=2,3 k=2: $a_{i2}=l_{i1}u_{12}+\mathbf{l_{i2}}u_{22}$, i=2,3

i=3: $a_{3k}=l_{31}u_{1k}+l_{32}u_{2k}+\mathbf{u_{3k}}$, k=3

Διαδοχική χρήση των σχέσεων αυτών μας δίνει τις τιμές των στοιχείων που έχουν τονισθεί με έντονη γραφή:

$u_{11}=a_{11}$, $u_{12}=a_{12}$, $u_{13}=a_{13}$, $l_{21}=a_{21}/u_{11}$, $l_{31}=a_{31}/u_{11}$,

$u_{22}=a_{22}-l_{21}u_{12}$, $u_{23}=a_{23}-l_{21}u_{13}$, $l_{32}=(a_{32}-l_{31}u_{12})/u_{22}$

$u_{33}=a_{33}-l_{31}u_{13}-l_{32}u_{23}$

Η επιλυση των τριγωνικων συστηματων Lz = b, Ux = z, γινεται αμεσα ξεκινωντας απο την εξισωση με εναν αγνωστο.

Παράδειγμα: Ανάλυση σε γινόμενο τριγωνικών πινάκων

Να λυθει με αναλυση σε γινομενο τριγωνικων πινακων το συστημα:

$$\begin{pmatrix} 4 & -2 & 0 \\ -1 & 4 & -1 \\ 0 & -1 & 4 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$$

Λύση

i=1: $u_{11}=a_{11}=4$, $u_{12}=a_{12}=-2$, $u_{13}=a_{13}=0$,

k=2: $l_{21}=a_{21}/u_{11}=-1/4$, $l_{31}=a_{31}/u_{11}=0$

i=2: $u_{22}= a_{22}-l_{21}u_{12}=4-(-1/4)*(-2)=7/2$

$u_{23}=a_{23}-l_{21}u_{13}=-1-(-1/4)*0=-1$

k=2: $l_{32}=(a_{32}-l_{31}u_{12})/u_{22}=(-1-0)/(7/2)=-2/7$

k=3: $u_{33}=a_{33}-l_{31}u_{13}-l_{32}u_{23} =4-0-(-2/7)*(-1)=4-2/7=26/7$

$$Lz = b \Leftrightarrow \begin{pmatrix} 1 & 0 & 0 \\ -1/4 & 1 & 0 \\ 0 & -2/7 & 1 \end{pmatrix}\begin{pmatrix} z_1 \\ z_2 \\ z_3 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$$

άρα $z_1=1$, $z_2=(1/4)z_1=1/4$, $z_3=(2/7)z_2=1/14$ και

$$Ux = z \Leftrightarrow \begin{pmatrix} 4 & -2 & 0 \\ 0 & 7/2 & -1 \\ 0 & 0 & 26/7 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 1 \\ 1/4 \\ 1/14 \end{pmatrix}$$

άρα $x_3=1/52$, $x_2=(2/7)*(1/4+x_3)= 7/26$, $x_1=1+2x_2=20/13$

Η Μέθοδος Cholesky

Η μεθοδος αυτη αποτελει εξειδικευση της παραπανω μεθοδου για συμμετρικους θετικά προσδιορισμένους πίνακες πινακες, και είναι ιδιαίτερα χρήσιμη κατά την επίλυση προβλημάτων με την Μέθοδο Ελαχίστων Τετραγώνων, γιατί τότε προκύπτουν πίνακες με αυτά τα χαρακτηριστικά.

Επειδη ενας συμμετρικος πινακας $A^T = A$ εχει $n(n+1)/2$ τυχαιους συντελεστες, ενω οι υπολοιποι $n(n-1)/2$ συμπιπτουν με τους συμμετρικους τους ($a_{ik} = a_{ki}$) ο πινακας A μπορει να γραφτει σε μιαν τετοια περιπτωση σαν γινομενο ενος κατω τριγωνικου πινακα και του συμμετρικου του:

$$A = A^T = LL^T \qquad (5.56)$$

που είναι και μη αρνητικά προσδιορισμένος, γιατί

$$x^TAx=x^TLL^Tx=(L^Tx)^TL^Tx=(L^Tx, L^Tx)=7\ L^Tx7^2\sigma0$$

Ετσι εχουμε τις σχεσεις:

$$a_{ik} = \sum_{j=1}^{n} l_{ij} u_{jk} = \sum_{j=1}^{\min(i,k)} l_{ij} u_{jk}, \quad i,k = 1(1)n \qquad (5.57)$$

δηλαδη για k = i:

$$a_{ii} = \sum_{j=1}^{i-1} l_{ij} l_{jk} + l_{ii}^{\ 2} \qquad (5.58)$$

και για k < i:

$$a_{ik} = \sum_{j=1}^{k} l_{ij} l_{jk} = l_{ik} l_{kk} + \sum_{j=1}^{k-1} l_{ij} l_{jk} \qquad (5.59)$$

Ετσι εχουμε τις σχεσεις:

$$l_{ii} = (a_{ii} - \sum_{j=1}^{i-1} l_{ij} l_{jk})^{1/2}, i = 1(1)n \qquad (5.60)$$

$$l_{ik} = \frac{1}{l_{ii}} \left(a_{ik} - \sum_{j=1}^{k-1} l_{ij} l_{jk} \right), i = k+1,...,n \qquad k = 1,...,n \qquad (5.61)$$

που χρησιμοποιουνται εναλλαξ για τον προσδιορισμο των συντελεστων του πινακα L:

k = 1: $l_{11} = \sqrt{a_{11}}, \quad l_{i1} = \frac{a_{i1}}{l_{11}}$, i = 2(1)n

k = 2: $l_{22} = \sqrt{a_{22} - l_{21}^{\ 2}}, l_{i2} = \frac{1}{l_{22}} \left(a_{i2} - l_{i1} l_{21} \right),$ i = 3(1)n

k = 3: $l_{33} = \sqrt{a_{33} - l_{31}^{\ 2} - l_{32}^{\ 2}}$

Παράδειγμα: Εφαρμογή Μεθόδου Cholesky

Να λυθει με την μεθοδο Cholesky το συστημα:

$$\begin{pmatrix} 4 & -6 & -2 \\ -6 & 25 & 15 \\ -2 & 15 & 11 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$$

Λύση

Συμφωνα με τους παραπανω τυπους, οι συντελεστες του πινακα L απο την διαγωνιο και κατω ειναι:

$$l_{11} = \sqrt{a_{11}} = \sqrt{4} = 2, \quad l_{21} = \frac{a_{21}}{l_{11}} = -3, \quad l_{31} = \frac{a_{31}}{l_{11}} = -1$$

$$l_{22} = \sqrt{a_{22} - l_{21}^{\ 2}} = \sqrt{25-9} = 4, \quad l_{32} = \frac{1}{l_{22}} \left(a_{32} - l_{31} l_{21} \right) = \frac{15-3}{4} = 3$$

$$l_{33} = \sqrt{a_{33} - l_{31}^{\ 2} - l_{32}^{\ 2}} = \sqrt{11-1-9} = 1$$

ενω οι συντελεστες πανω απο την διαγωνιο ειναι μηδεν.

Τα συστηματα που εχουμε να λυσουμε ειναι Lz = b, L^Tx = z. Ετσι προκυπτει:

$$z_1 = \frac{b_1}{l_{11}} = \frac{1}{2},\ z_2 = \frac{b_2 - l_{21}z_1}{l_{22}} = \frac{3}{8},\ z_3 = \frac{b_3 - l_{31}z_1 - l_{32}z_2}{l_{33}} = \frac{1}{2} - \frac{9}{8} = \frac{-5}{8}$$

$$x_3 = \frac{z_3}{l_{33}} = -\frac{5}{8},\ x_2 = \frac{z_2 - l_{32}x_3}{l_{22}} = \frac{9}{16},\ x_1 = \frac{z_1 - l_{21}x_2 - l_{31}x_3}{l_{11}} = \frac{25}{32}$$

5.5 Ασυνεπή Γραμμικά Συστήματα-Ψευδοαντίστροφοι Πίνακες

Ασυνεπη λεγονται τα γραμμικα συστηματα, που η λυση τους δεν ειναι μονοσημαντα ωρισμενη, ειτε γιατι ο αριθμος των εξισωσεων ειναι μεγαλυτερος απο τον αριθμο των αγνωστων, ειτε γιατι οι αριθμοι ειναι μεν ισοι, αλλα το συστημα ειτε δεν εχει λυση, ειτε εχει περισσοτερες απο μια λυσεις. Τετοια συστηματα προκυπτουν, π.χ., οταν το γραμμικο συστημα προκύπτει από μετρήσεις που γίνονται για τον προσδιορισμό παραμέτρων καποιου φυσικου φαινομενου. Επειδη οι μετρησεις μπορουν να επαναληφθουν οσες φορες θελουμε, εχουμε τοτε πολυ περισσοτερες εξισωσεις παρα αγνωστους. Οι εξισωσεις αυτες δεν ειναι ουσιαστικα ποτε συμβιβαστες μεταξυ τους αφου υπαρχουν παντα τοσο οι επιρροες εξωτερικων παραγοντων, εστω και ανεπαισθητες, οσο και τα λαθη που οφειλονται στην πεπερασμενη ακριβεια των οργανων μετρησεως.

Σε ολες αυτες τις περιπτωσεις, αντι για μιαν αυστηρη λυση του συστηματος Ay = b, αναζητουμε μιαν *γενικευμένη λύση*, δηλαδη ενα διανυσμα y που να ελαχιστοποιει καποιαν νορμ της διαφορας Ay - b, συνηθως την ευκλειδια νορμ: 7Ay - b72 . Αν η λυση αυτου του προβληματος ελαχιστοποιησης δεν ειναι μονοσημαντα ωρισμενη, τοτε απο το συνολο M των ελαχιστοποιουντων διανυσματων επιλεγουμε εκεινο που εχει ελαχιστη νορμ: xÃÂŷŵ

$$7\hat{y}7 = \min_{y \in M} 7y7 \qquad (5.68)$$

Ψευδοαντίστροφος, του πινακα A λεγεται τοτε ο πινακας Â που απεικονιζει το b στο ŷ :

$$\hat{A}\, b = \hat{y} \qquad (5.69)$$

Στην περιπτωση που προσδιοριζουμε την γενικευμενη λυση ως προς την ευκλειδια νορμ ισχυει το ακολουθο θεωρημα:

Θεώρημα 5.5

Αν ο πινακας A του συστηματος n εξισωσεων με m αγνωστους (n>m) Ay = b εχει γραμμικα ανεξαρτητες στηλες, τοτε το συνολο M των διανυσματων που ελαχιστοποιουν την ευκλειδια νορμ 7Ay - b72 αποτελειται απο ενα μονο σημειο ŷ και ισχυει:

$$\hat{y} = (A^TA)^{-1}A^T b \qquad (5.70)$$

δηλαδη,

$$\hat{A} = (A^TA)^{-1}A^T \qquad (5.71)$$

οπου A^T ο αναστροφος του πινακα Α.

Απόδειξη

Η σχεση 7Ay - $b7_2^2$ = min εχει εδω την μορφη:

$$g(x) := \sum_{i=1}^{n} \left(a_{i1}y_1 + \ldots + a_{im}y_m - b_i\right)^2 = \min \qquad (5.72)$$

(με n > m). Οι συνθηκες ελαχιστοποιησης αυτης της παραστασης ειναι:

$$0 = \frac{\partial g(y)}{\partial y_k} = 2\sum_{i=1}^{n} a_{ik}\left(a_{i1}y_1 + \ldots + a_{im}y_m - b_i\right), \quad k = 1(1)m \qquad (5.73)$$

και επομενως, για k = 1(1)m:

$$(\sum_{i=1}^{n} a_{i1}a_{ik})y_1 + (\sum_{i=1}^{n} a_{i2}a_{ik})y_2 + \ldots + (\sum_{i=1}^{n} a_{i2}a_{ik})y_m = \sum_{i=1}^{n} b_i a_{ik} \qquad (5.74)$$

Το συστημα αυτο ειναι τετραγωνικο, m εξισωσεων με m αγνωστους οι δε συντελεστες του ειναι εσωτερικα γινομενα των σειρων του A με τις στηλες του A. Δηλαδη εχει την μορφη:

$$A^T Ay = A^T b \qquad (5.75)$$

Αν οι στηλες του A ειναι γραμμικα ανεξαρτητες, τοτε απο την γραμμικη αλγεβρα ειναι γνωστο οτι ο πινακας A^TA ειναι αντιστρεψιμος. Επομενως ισχυει η (5.70).

Αν οι εξισώσεις του ασυνεπούς συστήματος Ay=b έχουν διαφορετικό βαθμό αξιοπιστίας (αντίστοιχο προς την αξιοπιστία των μετρήσεων από τις οποίες προέρχονται οι εξισώσεις) τότε για τον προσδιορισμό μίας προσεγγιστικής λύσης μπορούμε να απαιτήσουμε να γίνει:

$$p(x) := \sum_{i=1}^{n} w_i\left(a_{i1}y_1 + \ldots + a_{im}y_m - b_i\right)^2 = \min$$

όπου οι θετικοί συντελεστές βάρους w_i εκφράζουν τον βαθμό αξιοπιστίας της αντίστοιχης εξίσωσης. Με

$$w_i := d_i^2,\ i=1(1)n$$

η παράσταση αυτή παίρνει την μορφή:

$$p(x) := \sum_{i=1}^{n} \left(d_i a_{i1}y_1 + \ldots + d_i a_{im}y_m - d_i b_i\right)^2 = \min$$

Η λύση αυτού του προβλήματος είναι λύση του συστήματος:

$$Q^TQy = Q^Tb, \text{ όπου } Q = DA, \text{ και } D = \mathrm{diag}(d_i),$$

δηλαδή D είναι ο διαγώνιος πίνακας με στοιχεία της διαγωνίου $d_i = \sqrt{w_i}$, i=1(1)n.

Το τελευταίο σύστημα μπορεί να γραφτεί επομένως και στην μορφή:

$$A^TD^2Ay = A^TDb$$

Παράδειγμα Γενικευμένης Λύσης Ασυνεπούς Συστήματος

Να προσδιορισθει η γενικευμενη λυση ως προς την ευκλειδια νορμ για το συστημα:

$$y_1+y_2=5.1$$
$$y_1 - y_2=1.1$$
$$y_1 \quad =2.8$$

Λύση

Το συστημα εχει την μορφη:

$$\begin{pmatrix} 1 & 1 \\ 1 & -1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} y_1 \\ y_2 \end{pmatrix}=\begin{pmatrix} 5.1 \\ 1.1 \\ 2.8 \end{pmatrix}$$

Επομενως η γενικευμενη λυση του ως προς την ευκλειδια νορμ ειναι λυση του συστηματος:

$$\begin{pmatrix} 1 & 1 & 1 \\ 1 & -1 & 0 \end{pmatrix}\begin{pmatrix} 1 & 1 \\ 1 & -1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} y_1 \\ y_2 \end{pmatrix}=\begin{pmatrix} 1 & 1 & 1 \\ 1 & -1 & 0 \end{pmatrix}\begin{pmatrix} 5.1 \\ 1.1 \\ 2.8 \end{pmatrix}$$

δηλαδη του:

$$\begin{pmatrix} 3 & 0 \\ 0 & 2 \end{pmatrix}\begin{pmatrix} y_1 \\ y_2 \end{pmatrix}=\begin{pmatrix} 9 \\ 4 \end{pmatrix}$$

Η ζητουμενη ψευδολύση του αρχικού συστήματος, δηλαδή η λύση του τελευταίου συστήματος, ειναι:

$$\hat{y}_1 = 3 \ , \ \hat{y}_2 = 2$$

Ασκήσεις

1. Να λυθει το συστημα:

$$\begin{pmatrix} 5 & 2 & -1 \\ 2 & 10 & 4 \\ -1 & 4 & -10 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}=\begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$$

(α) με την επαναληπτικη μεθοδο ολικων βηματων, (β) με την επαναληπτικη μεθοδο μερικων βηματων, (γ) με την βασικη μεθοδο απαλοιφης του Gauss, (δ) με ανάλυση LU, (ε) την μεθοδο Cholesky (αν γίνεται).

Για τις επαναληπτικες μεθοδους καμετε δυο ως τρεις επαναληψεις με αρχικο διανυσμα το μηδενικο, και εκτιμηση σφαλματος εκ των προτερων και εκ των υστερων.

2. Ποιες μεθοδοι ειναι καταλληλες για την λυση του παρακατω συστηματος; Εφαρμοστε μιαν απο αυτες για να το λυσετε.

$$4x_1 + 2x_2 - x_3 = 4$$
$$x_1 + 4x_2 + 8x_3 = 2$$
$$x_1 + 5x_2 + 2x_3 = 3$$

3. Να λυθεί το σύστημα:

$$\begin{pmatrix} 8 & -1 & 1 \\ -1 & 5 & 1 \\ 1 & -1 & 5 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 6 \\ 7 \\ 8 \end{pmatrix}$$

α) με αναλυση του πινακα τους σε γινομενο τριγωνικων πινακων.
β) με την επαναληπτικη μεθοδο Gauss - Seidel.

4. Να λυθει με απαλοιφη Gauss το συστημα:

$$0.186\ x_1+0.123\ x_2=0.126$$
$$1.314\ x_1+0.875\ x_2=0.878$$

που εχει την ακριβη λυση x = 2, x = -2. Σε τι οφειλεται η ανακριβεια της προσεγγιστικης λυσης που βρισκουμε;

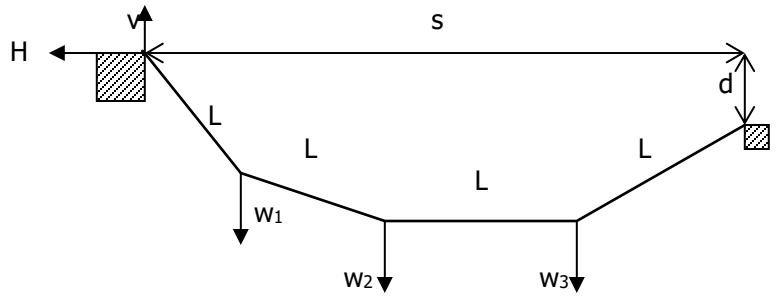
5. Να προσδιορισθει η γενικευμενη λυση (ψευδολύση) για το συστημα:

$$4x_1 + 2x_2 = 4$$
$$x_1 + 4x_2 = 2$$
$$2x_1 - 5x_2 = 3$$

6. ΕΠΑΝΑΛΗΠΤΙΚΗ ΕΠΙΛΥΣΗ ΜΗ ΓΡΑΜΜΙΚΩΝ ΣΥΣΤΗΜΑΤΩΝ

Παραδείγματα Φυσικών και Τεχνικών Εφαρμογών

1. Εξεταζουμε το καλωδιο του σχηματος οπου η οριζοντια αποσταση s και η καθετη αποσταση d των σημειων αναρτησεως του καλωδιου ειναι γνωστες. Σε ισες αποστασεις L κατα μηκος του καλωδιου ειναι αναρτημενα τρια βαρη W_i (Συνολικο μηκος του καλωδιου 4L). Ζητειται να προσδιορισθει η θεση ισοροπιας του καλωδιου, αν υποτεθει οτι τα τμηματα μεταξυ των κομβων αναρτησεως ειναι αβαρη (επομενως ευθυγραμμα) και μη επιμηκυνομενα.



Σχήμα 6.1: Θέση ισοροπίας καλωδίου με φορτία

Ως βασικες μεταβλητες θεωρουμε την οριζοντια συνιστώσα, H, και καθετη συνιστωσα, V, της τασεως του καλωδιου. Συμφωνα με τις συνθηκες ισορροπιας του καλωδιου οι οριζοντιες συνιστωσες της τασεως σε καθε τμημα του ειναι ισες με H, ενω η αντιστοιχη καθετη συνιστωσα καθως και η συνισταμενη ταση στο τμημα i ειναι αντιστοιχα:

$$V_i = V - W_i \quad , T_i = (H^2 + V_i^2)^{1/2} \quad , i = 0, 1, 2, 3$$

Επομενως, η οριζοντια και καθετη προβολη του τμηματος i του καλωδιου ειναι αντιστοιχα:

$$X_i = \frac{HL}{T_i}, Y_i = \frac{V_i L}{T_i}$$

Τα αθροισματα αυτων των μηκων πρεπει να ειναι αντιστοιχα s και d:

$$\sum_{i=0}^{3} X_i = s, \sum_{i=0}^{3} Y_i = d$$

Επομενως, για τον προσδιορισμο των H και V πρεπει να λυσουμε το συστημα μη γραμμικων εξισωσεων:

$$HL\sum_{i=0}^{3}\frac{1}{[H^2 + (V - W_i)^2]^{1/2}} - s = 0$$

$$L\sum_{i=0}^{3}\frac{V-W_i}{[H^2+(V-W_i)^2]^{1/2}}-d=0$$

(A. Jennings: Matrix Computations for Engineers and Scientists, Wiley, 1977, σελ. 65-67).

2. Να υπολογισθει προσεγγιστικα η λυση του προβληματος συνοριακων τιμων:

$$y''=\sqrt{1+y'(x)^2},\ y(0)=0,\ y(1)=1$$

(Stummel - Hainer σελ. 190-191)

Το προβλημα αυτο μπορει να αναφερεται, π.χ., στον προσδιορισμο των ενδιαμεσων θεσεων ενος σωματος, αν ειναι γνωστη η αρχικη και τελικη θεση και αν η επιταχυνση εξαρταται με τον παραπανω τροπο απο την ταχυτητα.

Για την προσεγγιστικη επιλυση ενος τετοιου προβληματος επιλεγουμε N ενδιαμεσες θεσεις x = k/N, k = 1(1)N-1 και προσεγγιζουμε τις τιμες $y'(x_k)$, $y''(x_k)$ με την βοηθεια των ακολουθων τυπων αριθμητικης διαφορισης:

$$y'(x_k)\approx\frac{y(x_k+h)-y(x_k-h)}{2h}=\frac{y(x_{k+1})-y(x_{k-1})}{2h}$$

$$y''(x_k)\approx\frac{y(x_k+h)-2y(x_k)+y(x_k-h)}{h^2}=\frac{y(x_{k+1})-2y(x_k)+y(x_{k-1})}{h^2}$$

οπου h = 1/N (για να τους αποδειξουμε αρκει να παρουμε τα αναπτυγματα Taylor περι το x_k, οποτε βλεπουμε και οτι η ακριβεια αυτων των προσεγγισεων ειναι της ταξεως του h^2).

Βαζοντας στην παραπανω διαφορικη εξισωση $x=x_k$ και προσεγγιζοντας τις παραγωγους με τους παραπανω τυπους παιρνουμε:

$$Y_{k+1}-2Y_k+Y_{k-1}=\frac{h}{2}\sqrt{4h^2+(Y_{k+1}-Y_{k-1})^2},\ k=1(1)N$$

οπου $Y_k \approx y(x_k)$, δηλαδη ενα μη γραμμικο συστημα N-1 εξισωσεων με N-1 αγνωστους.

6.1 Εντοπισμός Λύσεων

Οι βασικες μεθοδοι για την επιλυση συστηματων μη γραμμικων εξισωσεων ειναι επαναληπτικες και αποτελουν γενικευσεις αντιστοιχων μεθοδων για εξισωσεις μιας μεταβλητης. Οπως και στην περιπτωση εξισωσεων μιας μεταβλητης, η συγκλιση των μεθοδων αυτων ειναι εξασφαλισμενη μονον οταν το σημειο εκκινησεως της μεθοδου ειναι αρκετα κοντα σε καποιαν λυση. Ετσι μπαινει και εδω κατ' αρχην το βασικο προβλημα του εντοπισμου, αν ειναι δυνατον, ολων των λυσεων, δηλαδη, του προσδιορισμου περιοχων που καθε μια να περιλαμβανει μιαν μονον λυση. Το προβλημα αυτο ειναι στην προκειμενη περιπτωση πολυ δυσκολωτερο απο οτι σε εξισωσεις μιας μεταβλητης, γιατι εν γενει δεν ειναι δυνατος γραφικος εντοπισμος των λυσεων.

Ο βασικος τροπος εντοπισμου των λυσεων ενος συστηματος μη

γραμμικων εξισωσεων:

$$f(x) = \theta \qquad (6.1)$$

οπου f συνεχης απεικονιση του $\mathbf{R}^n$ (η $\mathbf{C}^n$) στον εαυτο του, δηλαδη:

$$f(x_1, x_2,...,x_n) = 0 \text{ , } i=1(1)n \qquad (6.2)$$

ειναι η αναζητηση ελαχιστων καποιας νορμ του διανυσματος f.

Συνηθεστερα χρησιμοποιειται εδω η Ευκλειδια νορμ οποτε η προς ελαχιστοποιηση συναρτηση ειναι:

$$g(x):= \|f(x)\|_2^2 = |f_1(x)|^2 + |f_2(x)|^2 + ...+ |f_n(x)|^2 \qquad (6.3)$$

Τα ελαχιστα αυτα δεν αντιστοιχουν οπωσδηποτε σε σημεια μηδενισμου, αλλα τα σημεια μηδενισμου ειναι οπωσδηποτε θεσεις (απολυτου) ελαχιστου της g(x).

Ο εντοπισμος τετοιων ελαχιστων στην απλουστερη μορφη του μπορει να γινει με αναζητηση ελαχιστων πανω στους κομβους ενος ορθογωνιου δικτυου που να καλυπτει καποιαν περιοχη του χωρου. Για ακριβεστερο προσδιορισμο των ελαχιστων μπορει στην συνεχεια καθε περιοχη του αρχικου δικτυου που εμφανιζει τοπικα ελαχιστα να καλυφθει με ενα δευτερο μικροτερης εκτασης αλλα πυκνοτερο δικτυο και να γινει παλι αναζητηση ελαχιστων στους κομβους του με απλο υπολογισμο των τιμων της g(x) σ'αυτα.

Πιο αποτελεσματικος, αλλα και πιο απαιτητικος υπολογιστικα τροπος εντοπισμου ελαχιστων ειναι η εφαρμογη στην g(x) αλγοριθμων ελαχιστοποιησης, οπως η μεθοδος ταχιστης καθοδου η οι ημινευτωνειες μεθοδοι.

6.2 Η Μέθοδος Newton – Raphson για Συστήματα

Η μέθοδος Newton-Raphson για την λύση μη γραμμικών εξισώσεων μπορεί να γενικευθεί άμεσα ώστε να είναι εφαρμόσιμη και σε μη γραμμικά συστήματα. Θα περιγράψουμε εδώ αρχικά έναν άλλο τρόπο παραγωγής της μεθόδου για την λύση απλών εξισώσεων, ο οποίος μπορεί άμεσα να εφαρμοσθεί και για την λύση συστημάτων.

Έστω ότι η συνάρτηση $f\in C^2[a,b]$ έχει ένα σημείο μηδενισμού $s\in[a,b]$ και ότι είναι ήδη γνωστή κάποια προσέγγιση x_k του s. Ζητούμε έναν διορθωτικό όρο δ_k που μεταφέρει από το x_k στο s: $x_k + \delta_k = s$. Επομένως, με χρήση του αναπτύγματος Taylor της f(x) γύρω από το σημείο x_k έχουμε:

$$0=f(s)=f(x_k + \delta_k)= f(x_k) + \delta_k f'(x_k)+(\delta_k^2/2)f''(\xi_k) \qquad (6.4)$$

Αν θεωρήσουμε ότι ο διορθωτικός όρος δ_k είναι μικρός, τότε ο συντελεστής δ_k^2 στον όρο σφάλματος είναι πολύ μικρότερος και έτσι το σφάλμα μπορεί ναθεωρηθεί αμελητέο. Έτσι έχουμε:

$$0\approx f(x_k) + \delta_k f'(x_k) \text{ ή } \delta_k\approx -f(x_k)/f'(x_k) \qquad (6.5)$$

Άρα είναι

$$s= x_k + \delta_k \approx x_k - f(x_k)/f'(x_k). \qquad (6.6)$$

και η παράσταση:

$$x_{k+1}:= x_k - f(x_k)/f'(x_k) \qquad (6.7)$$

θα αποτελεί μία πολύ καλύτερη προσέγγιση του s από ότι το x_k.

Τους ίδιους συλλογισμούς μπορούμε να επαναλάβουμε και για ένα σύστημα εξισώσεων. Θα τους επαναλάβουμε εδώ για ένα σύστημα δύο εξισώσεων με δύο αγνώστους.

Έστω ότι το σύστημα: $f(x,y)=0$, $g(x,y)=0$

έχει την λύση x=s, y=t και ότι γνωρίζουμε ήδη μίαν προσέγγιση $x^{(k)} \approx s$, $y^{(k)} \approx t$ αυτής της λύσης. Τότε, με βάση το θεώρημα Taylor για συναρτήσεις πολλών μεταβλητών, θα ισχύει:

$0 = f(s,t)= f(x^{(k)} + \delta^{(k)}, y^{(k)}+ \varepsilon^{(k)}) \approx f(x^{(k)}, y^{(k)})+ \delta^{(k)}f_x(x^{(k)}, y^{(k)})+ \varepsilon^{(k)}f_y(x^{(k)}, y^{(k)})$

$0 = g(s,t)= g(x^{(k)} + \delta^{(k)}, y^{(k)}+ \varepsilon^{(k)}) \approx g(x^{(k)}, y^{(k)})+ \delta^{(k)}g_x(x^{(k)}, y^{(k)})+ \varepsilon^{(k)}g_y(x^{(k)}, y^{(k)})$,

όπου $\delta^{(k)}$, $\varepsilon^{(k)}$ οι διορθωτικοί όροι των προσεγγιστικών τιμών $x^{(k)}$, $y^{(k)}$, και οι όροι σφάλματος στα αναπτύγματα Taylor έχουν παραλειφθεί απ'ευθείας, ενώ $f_x=\partial f/\partial x$ κ.ο.κ. Έτσι προκύπτει ένα προσεγγιστικό γραμμικό σύστημα για τον προσδιορισμό των διορθωτικών όρων. Για να απλουστεύσουμε την επεξεργασία του εισάγουμε τις συντομογραφίες:

$$f^{(k)}:= f(x^{(k)}, y^{(k)}) \text{ , } f_x^{(k)}:= f_x(x^{(k)}, y^{(k)}), \text{ κ.ο.κ.}$$

Έτσι έχουμε το σύστημα:

$$\delta^{(k)}f_x^{(k)}+ \varepsilon^{(k)}f_y^{(k)}\approx -f^{(k)} \quad \text{ή} \quad (s-x^{(k)})f_x^{(k)}+ (t-y^{(k)})f_y^{(k)}\approx -f^{(k)},$$
$$\delta^{(k)}g_x^{(k)}+ \varepsilon^{(k)}g_y^{(k)}\approx -g^{(k)} \quad \text{ή} \quad (s-x^{(k)})g_x^{(k)}+ (t-y^{(k)})g_y^{(k)}\approx -g^{(k)}$$

Λύνοντας το τελευταίο σύστημα έχουμε:

$$\begin{pmatrix} s \\ t \end{pmatrix} \approx \begin{pmatrix} x^{(k)} \\ y^{(k)} \end{pmatrix} - \begin{pmatrix} f_x^{(k)} & f_y^{(k)} \\ g_x^{(k)} & g_y^{(k)} \end{pmatrix}^{-1} \begin{pmatrix} f^{(k)} \\ g^{(k)} \end{pmatrix} := \begin{pmatrix} x^{(k+1)} \\ y^{(k+1)} \end{pmatrix} \qquad (6.8)$$

όπου η τελευταία ισότητα ορίζει τις νέες, βελτιωμένες προσεγγίσεις των s,t.

Επαναλαμβανοντας και εδω τους αντιστοιχους συλλογισμους που παραθεσαμε για εξισωσεις μιας μεταβλητης μπορουμε και εδω να δειξουμε οτι για (x,y) αρκετα κοντα στο (s,t) η μεθοδος αυτη εχει τετραγωνικη συγκλιση.

Παράδειγμα: Λύσης Συστήματος με την Μέθοδο Newton - Raphson

Να λυθει το συστημα:

$$0 = e^{-y} + e^{-x} - x \quad := f(x , y)$$
$$0 = e^{-y} - e^{-x} - y \quad := g(x , y)$$

Λύση

Με βάση την (6.8) η γενικη μορφη της μεθόδου Newton-Raphson για συστηματα δυο εξισωσεων ειναι:

$$x^{(k+1)} = x^{(k)} - \frac{f^{(k)}g_y^{(k)} - g^{(k)}f_y^{(k)}}{f_x^{(k)}g_y^{(k)} - f_y^{(k)}g_x^{(k)}}, \quad y^{(k+1)} = y^{(k)} - \frac{g^{(k)}f_x^{(k)} - f^{(k)}g_x^{(k)}}{f_x^{(k)}g_y^{(k)} - f_y^{(k)}g_x^{(k)}}$$

(6.9)

Εδω ειναι

$f_x = - e^{-x} - 1$, $f_y = - e^{-y}$, $g_x = e^{-x}$, $g_y = - e^{-y} - 1$

Αντικαθιστωντας αυτες τις παραστασεις στις σχεσεις (6.9) και με αρχικο σημειο το $(0,0)^T$ εχουμε την εξης ακολουθια:

k	0	1	2	3	4
$x^{(k)}$	0	0.8	1.04023	1.05176	1.051783072
$y^{(k)}$	0	0.4	0.35745	0.35316	0.35315439

οπου στις πρωτες επαναληψεις γραψαμε τα αποτελεσματα στρογγυλεμενα σε πεντε δεκαδικα ψηφια μετα την υποδιαστολη. Μετα την τεταρτη επαναληψη τα αποτελεσματα δεν μεταβαλλονται, γιατι εχουμε εξαντλησει την ακριβεια με την οποια εκτελει πραξεις και αποθηκευει τα αποτελεσματα ο υπολογιστης.

Ασκήσεις

1. Να λυθουν τα παρακατω συστηματα με την μεθοδο Newton - Raphson

a. $f_1 := 4x_1 - x_2 - x_1 \sin x_2 = 0$, $f_2 := x_1 + x_2 - 4\cot x_2 = 0$
(Ostrowski σελ.267).

b. $-2x^3 - 3xy + 4\sin y = -6$, $3x^2 - 2xy^2 + 3\cos x = 8$ (Antia, σελ.343)

c. $x^2 + 3y^2 = 1.1076$, $x^2 - xy + y^2 = 0.25$ (Antia, σελ.344)

2. Βρειτε τις λυσεις (πραγματικες η μιγαδικες) της εξισωσης:

$$\cos z = 1 + z/2,\ z = x + iy$$

παιρνοντας υπ'οψη οτι:

$$\cos(x+iy) = \cos x \cosh y - i \sin x \sinh y$$

3.Προβληματα συνοριακων τιμων σαν το ακολουθο:

$$u'' = c \sin u + g(t),\ u(0) = 0,\ u(1) = 0$$

χαρακτηριζουν τις εξαναγκασμενες ταλαντωσεις ενος εκκρεμους. Αν προσεγγισουμε την λυση του προβληματος αυτου αντικαθιστωντας την δευτερη παραγωγο με πηλικο διαφορων καταληγουμε στο συστημα:

$$U_{k+1} - 2U_k + U_{k-1} = c \sin U_k + h^2 g(t_k),\ h = 1/n,\ t_k = k/n,\ k = 1(1)n-1$$

Να λυθει το συστημα αυτο για $n = 4$ αν $c = 0.5$, $g(t) := 0.3\cos(2t)$.

4. Η λυση του προβληματος συνοριακων τιμων:

$$u'' + 4a^2(u')^2 + 1 = 0,\ u(0) = u(1) = 0$$

μπορει να προσεγγισθει με την λυση του συστηματος:

$$U_{k+1} - 2U_k + U_{k-1} + a^2(U_{k+1} - U_{k-1})^2 + h^2 = 0,\ k = 1(1)n-1,\ U_0 = U_n = 0$$

οπου $h = 1/n$. Δηλαδη, $U_k \approx u(k/n)$. Λυστε αυτο το συστημα για $a = 1$ και $n = 3$ και συγκρινετε την προσεγγιση με την ακριβη λυση:

$$u(t) = \frac{1}{4a^2}\ln\left(\frac{\cos(2at - a)}{\cos(a)}\right)$$

5. Προσδιοριστε με την μεθοδο Newton - Raphson τις λυσεις των ακολουθων συστηματων (Kopchenova - Maron: Computational Mathematics) :

a. $\cos(0{,}4y + x^2) + x^2 + y^2 - 1.6 = 0$
$1.5x^2 - y^2/0.36 - 1 = 0$

b. $\tan(xy + k) = x^2$, $k = 0.1m$, $m = 0(1)4$
$ax^2 + 2y^2 = 1$, $a = 0.5 + 0.1m$

c. $2x^2 - xy - 5x + 1 = 0$, $x + 3\ln x - y = 0$

e. $xy^3 - y - 4 = 0$, $ax^3 - y^2 - 1 = 0$, $a = 1 + 0.5k$, $k = 0(1)5$

f. $\sin x - y = 1.32$, $\cos y - x = -0.85$

g. $x^2 + y^2 + z^2 = 1$, $2x^2 + y^2 - 4z = 0$, $3x^2 - 4y + z^2 = 0$

h. $x = \ln(y/z) + 1$, $y = 0.4 + z^2 - 2x^2$, $z = 2 + xy/20$

i. $x + x^2 - 2yz = 0.1$, $y - y^2 + 3xz = -0.2$, $z + z^2 + 2xy = 0.3$

6. Αν για την αριθμητικη επιλυση της ολοκληρωτικης εξισωσης:

$$\int_0^1 \frac{(y(x)+y(t))^2}{1+x+t}\,dt = 1, \quad 0 \le x \le 1$$

προσεγγισουμε το ολοκληρωμα με απλη χρηση του κανονα ολοκληρωσης του τραπεζιου προκυπτει:

$$1 \approx \frac{1}{2}\left(\frac{(y(x)+y(0))^2}{1+x+0} + \frac{(y(x)+y(1))^2}{1+x+1} \right)$$

παιρνοντας την σχεση αυτην για $x=x_0=0$ και $x=x_1=1$ εχουμε το συστημα μη γραμμικων εξισωσεων:

$$4Y_0^2 + ½*(Y_0 + Y_1)^2 = 2, \quad ½*(Y_0 + Y_1)^2 + 4/3*Y_1^2 = 2,$$

οπου Y_k προσεγγισεις των τιμων $y(x_k)$, $k = 0, 1$.

Παρομοια, με απλη χρηση του κανονα του Simpson προκυπτει:

$$1 \approx \frac{1}{6}\left(\frac{(y(x)+y(0))^2}{1+x+0} + 4\frac{(y(x)+y(0.5))^2}{1+x+0.5} + \frac{(y(x)+y(1))^2}{1+x+1} \right)$$

και για $x = x_k$, $k = 0,1,2$, οπου $x_0 = 0$, $x_1 = 0.5$, $x_2 = 1$ εχουμε:

$$4Y_0^2 + 8/3*(Y_0 + Y_1)^2 + ½*(Y_0 + Y_1)^2 = 6$$

$$2/3*(Y_0 + Y_1)^2 + 8Y_1^2 + 2/5*(Y_1 + Y_2)^2 = 6$$

$$½*(Y_0 + Y_1)^2 + 8/5*(Y_1 + Y_2)^2 + 4/3*Y_2^2 = 6$$

οπου $Y_k \approx y(x_k)$, $k = 0(1)2$ (Βλ. Collatz: The Numerical Treatment of Differential Equations, Springer, 1966, σελ.477).

Να λυθουν τα παραπανω συστηματα με την μεθοδο Newton - Raphson.

7. ΠΑΡΕΜΒΟΛΗ

Η παρεμβολη αποτελει ενα τοσο πρακτικο οσο και θεωρητικο εργαλειο προσεγγισεως συναρτησεων με ευρυτατη κλιμακα εφαρμογων. Πρακτικα χρησιμοποιειται η παρεμβολη, σε διαφορες μορφες της, οπουδηποτε εχουμε να προσδιορισουμε προσεγγιστικα τις τιμες που παιρνει ενα φυσικο η τεχνικο μεγεθος σε ενδιαμεσες θεσεις οταν εχουμε προσδιορισει με μετρηση η υπολογιστικα τις τιμες του για μεμονωμενες χρονικες στιγμες η θεσεις στον χωρο.

Θεωρητικα χρησιμευει η παρεμβολη για την κατασκευη προσεγγιστικων τυπων αριθμητικης διαφορισης, αριθμητικης ολοκληρωσης, αριθμητικης επιλυσης διαφορικων και ολοκληρωτικων εξισωσεων, αλλα και για την μελετη αλλων ειδων προσεγγισεων. Π.χ. βελτιστες προσεγγισεις ως προς διαφορα κριτηρια αποδεικνυεται οτι αποτελουν παρεμβολικες προσεγγισεις της δοσμενης συναρτησης σε ειδικες κατανομες σημειων.

Η (μονοδιαστατη) παρεμβολη αποτελει εναν βασικο τροπο προσεγγισης μιας συναρτησης f(x) σε ενα ολοκληρο διαστημα [a, b] η ενα χωριο G του μιγαδικου επιπεδου (σε αντιθεση, π.χ. με το αναπτυγμα Taylor, που προσεγγιζει ικανοποιητικα την συναρτηση μονο στην γειτονια ενος σημειου). Απαιτηση της ειναι ο προσδιορισμος της προσεγγιζουσας συναρτησης φ(x) ετσι ωστε να συμπιπτει με την f(x) (να εχει τις ιδιες τιμες με αυτην) σε επιλεγμενα σημεια $x_i \in [a, b]$, ή $x_i \in G$ που λεγονται "σημεια παρεμβολης" η "κομβοι".

Η συναρτηση φ(x) ειναι κατα κανονα *γραμμικός συνδυασμός:*

$$\varphi(x) := \sum_{k=0}^{n} c_k \varphi_k(x) \tag{7.1}$$

(γενικευμενο πολυωνυμο) δοσμενων συναρτησεων $\varphi_k(x)$, k = 0(1)n, που πρεπει να ειναι γραμμικως ανεξαρτητες ωστε η παρασταση (7.1) να ειναι μονοσημαντα ωρισμενη.

Η συνθηκη αυτη σημαινει οτι μια παρασταση της μορφης (7.1) μπορει να μηδενισθει για ολα τα $x \in [a,b]$ μονον οταν ολοι οι συντελεστες c_k μηδενιζονται. Ετσι δυο παραστασεις της μορφης (7.1) συμπιπτουν για ολα τα $x \in [a, b]$ (δηλ. η διαφορα τους ειναι εκ ταυτοτητος μηδεν), μονον οταν ολοι οι συντελεστες τους συμπιπτουν.

Με βαση την απαιτηση παρεμβολης πρεπει να ειναι:

$$\sum_{k=0}^{n} c_k \varphi_k(x_i) = f(x_i) \tag{7.2}$$

Δηλαδη, εχουμε ενα γραμμικο συστημα ως προς τους αγνωστους c_k.

Για να ειναι μονοσημαντα ωρισμενη η λυση του πρεπει οι εξισωσεις, αρα και τα σημεια x_i , να ειναι οσα και οι αγνωστοι. Αρα i = 0(1)n.

Η συνθηκη αυτη δεν αρκει βεβαια για την εξασφαλιση της μονοσημαντοτητας αλλα πρεπει επι πλεον η οριζουσα $\det(\varphi_k(x_i))$ του συστηματος να ειναι διαφορετικη του μηδενος. Αυτο ειναι π.χ. εξασφαλισμενο, στην συνηθεστερη περιπτωση, που οι συναρτησεις $\varphi_k(x)$ ειναι δυναμεις του x:

$$\varphi_k(x) := x^k \text{ , } k = 0(1)n \tag{7.3}$$

(δηλ. η φ(x) πολυωνυμο του x βαθμου n) και τα σημεια x_i ειναι διαφορετικα μεταξυ τους. Τοτε η οριζουσα του συστηματος ειναι η οριζουσα του Vandermonde που ισουται με το γινομενο ολων των ορων $(x_i - x_j), i \neq j$.

Σημείωση: Η ιδια διαδικασια μπορει να ακολουθηθει και με $x \in \mathbf{R}^n$, $n \geq 2$, δηλαδη για την κατασκευη παρεμβολικων προσεγγισεων για συναρτησεις πολλων μεταβλητων. Δυστυχως ομως στην περιπτωση αυτην υπαρχουν παντα διαφορετικα μεταξυ τους σημεια τετοια ωστε η λυση του συστηματος (7.2) δεν ειναι μονοσημαντα ωρισμενη.

7.1 Λαγκρανζιανή μορφή του Πουωνύμου Παρεμβολής

Ο προσδιορισμος της παρεμβολικης προσεγγισης φ(x) με επιλυση του γραμμικου συστηματος (7.2) ειναι διαδικασια χρονοβορα. Στην περιπτωση ομως που η προσεγγιση αυτη ειναι πολυωνυμο υπαρχουν δυο τροποι προσδιορισμου του πολυωνυμου χωρις να λυσουμε το συστημα αυτο, οποτε ομως το πολυωνυμο δεν εχει την γνωστη μορφη γραμμικου συνδυασμου δυναμεων του x, αλλα αλλες καταλληλοτερες.

Η πρωτη απο αυτες τις μορφες οφειλεται στον Lagrange και παριστανει το πολυωνυμο παρεμβολης σαν αθροισμα αλλων πολυωνυμων που το καθε ενα τους παιρνει μια απο τις δοσμενες τιμες, f_k, στο αντιστοιχο σημειο x_k, και μηδενιζεται σε ολα τα αλλα σημεια x_i.

Αν θελουμε το πολυωνυμο βαθμου n $L_k(x)$ να παιρνει την τιμη 1 στο σημειο x_k και να μηδενιζεται σε ολα τα αλλα σημεια παρεμβολης, υπαρχει ενας πολυ απλος τροπος να το κατασκευασουμε. Το πολυωνυμο εχει τις n ριζες x_i, $i \in \{0, 1, 2, ..., n\}$, αλλα $i \neq k$. Αρα μπορει να γραφτει σαν γινομενο των αντιστοιχων n πρωτοβαθμιων παραγοντων:

$$L_k(x) \equiv A_k(x - x_0)..(x - x_{k-1})(x - x_{k+1})..(x - x_n) \qquad (7.4)$$

η συντομωτερα,

$$L_k(x) = A_k \prod_{i=0, i \neq k}^{n} (x - x_i) \qquad (7.5)$$

Ο αγνωστος συντελεστης A_k προκυπτει απο την συνθηκη:

$$1 = L_k(x_k) = A_k \prod_{i=0, i \neq k}^{n} (x_k - x_i) \qquad (7.6)$$

Διαιρωντας κατα μελη τις σχεσεις (8.7) και (8.8) εχουμε ετσι:

$$L_k(x) = \prod_{i=0, i \neq k}^{n} \left(\frac{x - x_i}{x_k - x_i} \right) \qquad (7.7)$$

Το πολυωνυμο βαθμου n που στην θεση x_j εχει την τιμη f_j και σε ολα τα αλλα σημεια παρεμβολης μηδενιζεται ειναι τωρα προφανως το $f_j L_j(x)$, ενω το αθροισμα ολων αυτων των πολυωνυμων:

$$P_n(x) := \sum_{k=0}^{n} f_k L_k(x) \qquad (7.8)$$

ειναι το ζητουμενο πολυωνυμο παρεμβολης. Ειναι βαθμου n γιατι ολοι οι προσθετεοι ειναι βαθμου n και στην τυχουσα θεση x_j παιρνει την τιμη f_j , γιατι

εκει μηδενιζονται ολοι οι προσθετεοι εκτος απο τον $f_jL_j(x_j) = f_j$. Ετσι εχουμε το θεωρημα:

Θεώρημα 7.1 *(Lagrange)*

Με δοσμενα τα σημεια (x_k, f_k), k = 0(1)n (με $x_k \neq x_j$ για $k \neq j$) το πολυωνυμο n (το πολυ) βαθμου που περναει απο τα σημεια αυτα μπορει να γραφτει υπο την μορφη (8.10), οπου τα πολυωνυμα $L_k(x)$ που δινονται απο την (8.9) λεγονται "στοιχειωδη πολυωνυμα του Lagrange" και μπορουν να γραφτουν επισης στην μορφη:

$$L_k(x) = \prod_{i=0, i\neq k}^{n} \left(\frac{x - x_i}{x_k - x_i} \right) \equiv \frac{\omega(x)}{(x - x_k)\omega'(x_k)} \qquad (7.9)$$

οπου

$$\omega(x) := (x - x_0)(x - x_1)...(x - x_n) = \prod_{i=0}^{n} (x - x_i) \qquad (7.10)$$

Παράδειγμα Παρεμβολής Lagrange

Να βρεθει η Λαγκρανζιανη μορφη του πολυωνυμου παρεμβολης που περναει απο τα σημεια:

k	0	1	2	3
x_k	*1*	*3*	*4*	*5*
f_k	*-3*	*0*	*1*	*2*

Λύση

Τα δοσμενα σημεια ειναι 4 = n+1. Αρα το πολυωνυμο ειναι βαθμου n = 3 και εχει την μορφη:

$$P_3(x) := \sum_{k=0}^{3} f_k L_k(x) = -3L_0(x) + 0L_1(x) + 1L_2(x) + 2L_3(x)$$

οπου

$$L_0(x) = \frac{(x - x_1)(x - x_2)(x - x_3)}{(x_0 - x_1)(x_0 - x_2)(x_0 - x_3)} = \frac{(x-3)(x-4)(x-5)}{(1-3)(1-4)(1-5)}$$

$$L_2(x) = \frac{(x - x_0)(x - x_1)(x - x_3)}{(x_2 - x_0)(x_2 - x_1)(x_2 - x_3)} = \frac{(x-1)(x-3)(x-5)}{(4-3)(4-4)(4-5)}$$

$$L_3(x) = \frac{(x - x_0)(x - x_1)(x - x_2)}{(x_3 - x_0)(x_3 - x_1)(x_3 - x_2)} = \frac{(x-1)(x-3)(x-4)}{(5-1)(5-3)(5-4)}$$

Εισαγοντας αυτες τις παραστασεις στον τυπο για το $P_3(x)$ παιρνουμε:

$$P_3(x) = \frac{1}{8}(x-3)(x-4)(x-5) - \frac{1}{3}(x-1)(x-3)(x-5) + \frac{1}{4}(x-1)(x-3)(x-4)$$

Ασκήσεις

1. Αν η διαταξη των σημειων παρεμβολης ειναι συμμετρικη ως προς το μηδεν: $x_j = -x_{n-j}$, $j = 0(1)\left[\frac{n+1}{2}\right]$ (όπου [α] το ακέραιο μέρος του α) και η συναρτηση f(x) είναι αρτια, (f(-x) = f(x)) δείξτε ότι τοτε και το πολυωνυμο παρεμβολης ειναι αρτια συναρτηση ($P_f(-x)=P_f(x)$) και ισοδυναμει με παρεμβολη της συναρτησης g(t):= f(x), t := x^2 , στα σημεια $t_j := x_j^2$, $j = 0(1)\left[\frac{n+1}{2}\right]$.

Δηλαδη, $P_f(x) = P_g(x^2)$ οπου $P_f(x)$, $P_g(t)$ τα παραπανω πολυωνυμα παρεμβολης.

2. Δειξτε οτι αν τα σημεια παρεμβολης εχουν συμμετρικη διαταξη προς το μηδεν και η f(x) ειναι περιττη συναρτηση, (f(-x) = - f(x)) τοτε και το πολυωνυμο παρεμβολης της ειναι περιττη συναρτηση ($P_f(-x) = - P_f(x)$) και ισοδυναμει με παρεμβολη της συναρτησης g(t) = f(x)/x , t:= x^2, στα σημεια $t_j = x_j^2$, $j = 0(1)\left[\frac{n+1}{2}\right]$.

Δηλαδη, $P_f(x) = x\, P_g(x^2)$, οπου $P_f(x)$, $P_g(t)$ τα παραπανω πολυωνυμα παρεμβολης.

7.2 Η νευτώνεια Μορφή του Πολυωνύμου Παρεμβολής

Η Λαγκρανζιανη μορφη του πολυωνυμου παρεμβολης εχει το πλεονεκτημα οτι εκφραζει αμεσα την εξαρτηση του πολυωνυμου αυτου απο τις δοσμενες τιμες της συναρτησης που προσεγγιζεται, αλλα και το μειονεκτημα οτι αν προστεθει εστω και ενα ακομη σημειο παρεμβολης ολες οι συναρτησεις L (x) πρεπει να υπολογισθουν ξανα.

Αυτο το μειονεκτημα αποφευγεται στην Νευτωνεια μορφη του πολυωνυμου παρεμβολης. Εδω το πολυωνυμο γραφεται υπο την μορφη:

$$P_n(x) = b_0 + b_1 (x-x_0) + b_2(x-x_0)(x-x_1) + \ldots + b_n(x-x_0)..(x-x_{n-1}) \qquad (7.11)$$

Αν χρησιμοποιησουμε εδω την απαιτηση:

$$P_n(x_j) = f_j,\ j = 0(1)n \qquad (7.12)$$

για τον προσδιορισμο των συντελεστων b_k παρατηρουμε οτι η k εξισωση περιλαμβανει μονο τους συντελεστες b_0 ως b_k γιατι οι υπολοιποι προσθετεοι της μηδενιζονται. Ετσι προκυπτει διαδοχικα:

$$f_0 = P(x_0) = b_0 ,\ f_1= P(x_1) = b_0 + b_1(x_1 - x_0) \text{ δηλ. } b_1 = \frac{f_1 - f_0}{x_1 - x_0} \text{ , κτλ.}$$

Γενικωτερα ισχυει το ακολουθο θεωρημα:

Θεώρημα 7.2 *(Newton)*

1. Αν εχουν δοθει οι τιμες

$$f(x_j) = f_j ,\ j = 0(1)n \qquad (7.13)$$

μιας συναρτησης f(x) το πολυωνυμο $P_n(x)$ παρεμβολης της στα σημεια x_j ($P_n(x_j)=f_j$, j = 0(1)n) μπορει να γραφτει με την μορφη:

$$P_n(x) = f_0 + f_{10}(x-x_0) + f_{210}(x-x_0)(x-x_1) + \ldots + f_{n,n-1,\ldots,0}(x-x_0)..(x-x_{n-1}) \qquad (7.14)$$

οπου οι συντελεστες δινονται απο τους τυπους:

$$f_{10} = \frac{f_1 - f_0}{x_1 - x_0} \qquad (7.15)$$

και γενικα

$$f_{k,k-1,\ldots,1,0} = \frac{f_{k,\ldots,1} - f_{k-1,\ldots,1,0}}{x_k - x_0} \qquad (7.16)$$

και υπολογιζονται με βαση το σχημα:

x_i	f_i	f_{ik}	f_{ikl}	
x_0	$f_0=b_0$			
		$f_{10}=b_1$		
x_1	f_1		$f_{210}=b_2$	
		f_{21}		$f_{3210}=b_3$
x_2	f_2		f_{321}	
		f_{32}		
x_3	f_3			
...	...			

(7.17): Υπολογιστικό Σχήμα Διαιρουμένων Διαφορών

οπου καθε φορα υπολογιζονται ολα τα μεγεθη μιας στηλης πριν αρχισει ο υπολογισμος των μεγεθων της επομενης.

Εξ αιτιας της μορφης τους τα μεγεθη αυτα λεγονται "πηλικα διαφορων" η "διαιρουμενες διαφορες" (divided differences). Το καθε ενα απο αυτα προκυπτει σαν πηλικο της διαφορας των δυο γειτονικων του απο την προηγουμενη στηλη με την διαφορα των ακραιων κομβων x που εμφανιζονται σε αυτα. Ετσι ειναι, π.χ.,

$$f_{i+2,i+1,i} = \frac{f_{i+2,i+1} - f_{i+1,i}}{x_{i+2} - x_i} \qquad (7.18)$$

Απόδειξη

Αν στην σχέση (7.14) βάλουμε $x=x_1$ και λάβουμε υπόψη ότι είναι $P_n(x_1)=f_1$, τότε προκύπτει αμέσως η (7.15). Παρόμοια, αν βάλουμε $x=x_k$ και λάβουμε υπόψη ότι είναι $P_n(x_k)=f_k$, τότε προκύπτουν διαδοχικά όλες οι σχέσεις (7.16).

Παρατηρήσεις: *Ιδιότητες της Παρεμβολικής Μορφής Newton*

Εδω πρεπει να προσεξουμε οτι σε περιπτωση προσθηκης νεων σημειων παρεμβολης το υπολογιστικό σχήμα (7.17) απλα επεκτεινεται προς τα κατω και προς τα δεξια. Επισης πρεπει να προσεξουμε οτι δεν ειναι απαραιτητο τα σημεια x_k να ειναι αριθμημενα κατα σειρα μεγεθους, γιατι μια τετοια προϋπόθεση δεν χρησιμοποιηθηκε πουθενα κατα την αποδειξη. Αλλωστε, μια τετοια διαταξη ειναι αδυνατη αν τα σημεια παρεμβολης εχουν μιγαδικες τιμες.

Αξιο προσοχης ειναι επισης οτι η τιμη ενος πηλικου διαφορων δεν μεταβαλλεται αν εναλλαξουμε την σειρα με την οποια παιρνουμε τα σημεια που χρησιμοποιουνται για τον σχηματισμο του, γιατι το πολυωνυμο παρεμβολης παραμενει το ιδιο ασχετως του με ποια σειρα χρησιμοποιουμε τα σημεια παρεμβολης.

Παράδειγμα Εύρεσης του Πολυωνύμου Παρεμβολής σε Μορφή Newton:

Να δοθει η μορφη Newton του πολυωνυμου που περναει απο τα ακολουθα σημεια:

k	*0*	*1*	*2*	*3*
x_k	1	2	4	5
f_k	7	6	8	7

Λύση

Το πολυωνυμο παρεμβολης ειναι εδω:

$$P_3(x) = b_0 + b_1(x-x_0) + b_2(x-x_0)(x-x_1) + b_3(x-x_0)(x-x_1)(x-x_2)$$
$$= b_0 + b_1(x-1) + b_2(x-1)(x-2) + b_3(x-1)(x-2)(x-4)$$

οπου οι συντελεστες b_κ προκυπτουν με βαση το σχημα διαιρεμενων διαφορων (7.17). Έτσι προκύπτει:

$$P_3(x) = 7 - (x-1) + \frac{2}{3}(x-1)(x-2) - \frac{1}{3}(x-1)(x-2)(x-4)$$

όπου οι συντελεστές δίνονται από το παρακάτω υπολογιστικό σχήμα:

x_i	f_i	f_{ik}	f_{ikl}	
1	7			
		(6-7)/(2-1)=-1		
2	6		(1-(-1))/(4-1)=2/3	
		(8-6)/(4-2)=1		(-2/3-2/3)/(5-1)= -1/3
4	8		(-1-1)/(5-2)=-2/3	
		(7-8)/(5-4)=-1		
5	7			
…	…			

Παρεμβολή Newton με ισαπέχοντα σημεία παρεμβολής

Ο υπολογισμος των συντελεστων b_i απλοποιειται πολυ αν τα σημεια x_j ειναι ισαπεχοντα, δηλ. αποτελουν αριθμητικη προοδο:

$$x_j = x_0 + j\,h\ ,\ j = 0(1)n \qquad (7.19)$$

Στην περιπτωση αυτην οι παρονομαστες των διαρουμενων διαφορων $f_{k,k-1,\ldots,1,0}$ εκφραζονται με την βοηθεια του βηματος h της αριθμητικης προοδου, ενω οι αριθμητες προκυπτουν με επανειλημμενο υπολογισμο διαφορων.

Για να απλοποιησουμε την παρασταση αυτων των διαφορων θα ορισουμε εδω τον τελεστη διαφορας, Δ, με την βοηθεια των ακολουθων σχεσεων:

$$\Delta f_i := f_{i+1} - f_i \qquad (7.20)$$
$$\Delta^{k+1} f_i := \Delta(\Delta^k f_i) = \Delta^k f_{i+1} - \Delta^k f_i \qquad (7.21)$$

οπου i, k = 0, 1, 2, ...

Χρησιμοποιωντας αυτον τον συμβολισμο μπορουμε να δειξουμε οτι:

$$f_{i+1,i} = \frac{\Delta f_i}{h}, \quad f_{i+2,i+1,i} = \frac{\Delta^2 f_i}{2h^2} \tag{7.22}$$

και γενικα

$$f_{i+k,\ldots,i} = \frac{\Delta^k f_i}{k!\,h^k} \tag{7.23}$$

Επομενως ισχυει το ακολουθο θεωρημα:

Θεώρημα 7.3

Αν τα σημεια παρεμβολης αποτελουν αριθμητικη προοδο με βημα h:

$$x_j = x_0 + j\,h\ ,\ j = 0(1)n \tag{7.24}$$

τοτε η νευτωνεια μορφη του πολυωνυμου παρεμβολης ειναι:

$$P_n(x) = f_0 + \frac{\Delta f_0}{h}(x - x_0) + \frac{\Delta^2 f_0}{2h^2}(x - x_0)(x - x_1) + \ldots + \frac{\Delta^n f_0}{n!\,h^n}(x - x_0)\ldots(x - x_{n-1}) \tag{7.25}$$

οπου οι διαφορες Δ f υπολογιζονται βασει των τυπων (7.20),(7.21).

Αν με την ισοτητα:

$$x = x_0 + t\,h \tag{7.26}$$

εισαγαγουμε μιαν νεα μεταβλητη t τοτε το πολυωνυμο παρεμβολης μπορει να γραφτει και με την μορφη:

$$P_n(x_0 + th) = f_0 + \frac{\Delta f_0}{1}t + \frac{\Delta^2 f_0}{2!}t(t-1) + \ldots + \frac{\Delta^n f_0}{n!}t(t-1)\ldots(t-n+1) \tag{7.27}$$

$$= f_0 + \Delta f_0\binom{t}{1} + \Delta^2 f_0\binom{t}{2} + \ldots + \Delta^n f_0\binom{t}{n} \tag{7.28}$$

όπου

$$\binom{t}{k} = \frac{t(t-1)\ldots(t-k+1)}{k!},\ k = 1,2,3,\ldots \tag{7.29}$$

Οι τυποι (7,27),(7.28) προκυπτουν αμεσα απο τον (7.25) αν προσεξουμε οτι

$$x - x_k = h(t - k) \tag{7.30}$$

Παρατήρηση: Η Σχέση Παρεμβολής Newton και Αναπτύγματος Taylor

Η (7.25) θυμιζει καπως το αναπτυγμα Taylor. Υπαρχει πραγματι συγγενεια μεταξυ τους, γιατι οταν το βημα h τεινει προς μηδεν η (7.25) τεινει προς το αθροισμα των n+1 πρωτων ορων του αναπτυγματος Taylor, γιατι τοτε οι συντελεστες $\Delta^k f_0/h^k$ τεινουν προς $f^{(k)}(x_0)$.

Οπως και στην περιπτωση των πηλικων διαφορων ο υπολογισμος των

διαφορων Δ f διευκολυνεται αν χρησιμοποιησουμε μιαν σχηματικη διαταξη των ενδιαμεσων αποτελεσματων:

x_i	f_i	Δf_i	$\Delta^2 f_i$	$\Delta^3 f_i$
x_0	f_0			
		Δf_0		
x_0+h	f_1		$\Delta^2 f_0$	
		Δf_1		$\Delta^3 f_0$
x_0+2h	f_2		$\Delta^2 f_1$	
		Δf_2		
x_0+3h	f_3			
…				

(7.31): Υπολογιστικό Σχήμα Διαφορών

Στο υπολογιστικο σχημα αυτο καθε διαφορα προκυπτει με αφαιρεση των γειτονικων της προηγουμενης στηλης.

Παράδειγμα Παρεμβολής Newton με Ισαπέχοντες Κόμβους

Να δοθει στην μορφη του Newton το πολυωνυμο παρεμβολης για τα ακολουθα σημεια:

k	0	1	2	3
x_k	10	12	14	16
f_k	-3	-1	2	5

Λύση

Επειδη εδω τα σημεια x αποτελουν αριθμητικη προοδο:

$x_j = 10 + 2j$, $j = 0, 1, 2, 3$

αντι για το σχημα υπολογισμου διαιρουμενων διαφορων χρησιμοποιουμε το σχημα υπολογισμου διαφορων:

x_i	f_i	Δf_i	$\Delta^2 f_i$	$\Delta^3 f_i$
10	-3			
		2		
12	-1		1	
		3		-1
14	2		0	
		3		
16	5			
…				

Συμφωνα με τους τυπους (7.25) και (7.27) ειναι αρα:

$$P_3(x) = -3 + \frac{2}{2}(x-10) + \frac{1}{2!2^2}(x-10)(x-12) - \frac{1}{3!2^3}(x-10)(x-12)(x-14)$$

$$P_3(10+2t) = -3+2t+\frac{1}{2!}t(t-1)-\frac{1}{3!}t(t-1)(t-2)$$

7.3 Το Σφάλμα Παρεμβολής

Στην περιπτωση που η συναρτηση f(x) που προσεγγιζουμε, καθως και τα σημεια παρεμβολης ειναι πραγματικα μεγεθη ισχυει το ακολουθο θεωρημα:

Θεώρημα 7.3

Αν η συναρτηση f(x) εχει στο διαστημα [a,b] συνεχη παραγωγο n+1 ταξεως και τα σημεια παρεμβολης x_i, i =0(1)n ανηκουν στο διαστημα αυτο, τοτε ισχυει:

$$f(x)-P_n(x)=\frac{f^{(n+1)}(\xi)}{(n+1)!}(x-x_0)(x-x_1)...(x-x_n)\,, \qquad (7.32)$$

οπου για καθε x το ξ ειναι ενα *άγνωστο* σημειο που ανηκει στο διαστημα I_x που οριζεται απο την ανισοτητα:

$$\min\{x, x_0,..., x_n\} < \xi < \max\{x, x_0, ..., x_n\} \qquad (7.33)$$

δηλαδη βρισκεται μεσα στο διαστημα που οριζεται απο τα ακραια από τα σημεια x και x_i, i = 0(1)n.

Ειδικωτερα ειναι:

$$|f(x)-P_n(x)|\le\frac{M}{(n+1)!}|(x-x_0)(x-x_1)...(x-x_n)|\,, \qquad (7.34)$$

οπου M ειναι το μεγιστο η ενα ανω φραγμα της $|f^{(n+1)}(\xi)|$ για $\xi \in I_x$.

Απόδειξη

Θα θεωρησουμε εδω οτι το x δεν συμπιπτει με κανενα απο τα σημεια παρεμβολης x_i .Επειδη το x θεωρειται εδω οτι ειναι συγκεκριμενο σημειο, εισαγουμε μιαν αλλη ελευθερη μεταβλητη t και οριζουμε την συναρτηση:

$$g(t):=f(t)-P_n(t)-\omega(t)\left(\frac{f(x)-P_n(x)}{\omega(x)}\right) \qquad (7.35)$$

οπου ω(x) ειναι το γινομενο των πρωτοβαθμιων παραγοντων $x - x_i$:

$$\omega(x) := (x - x_0)(x - x_1)...(x - x_n) \qquad (7.36)$$

Η συναρτηση g(t) εχει σαν σημεια μηδενισμου τοσο τα σημεια παρεμβολης x_i, οσο και το σημειο x, συνολικα δηλαδη n+2 σημεια.

Συμφωνα με το θεωρημα του Rolle, αναμεσα σε καθε ζευγαρι διαδοχικων τετοιων σημειων μηδενισμου της g, πρεπει να εχει τουλαχιστον ενα σημειο μηδενισμου η παραγωγος της g'. Η g'(t) εχει, δηλαδη, τουλαχιστον n+1 διαφορετικα σημεια μηδενισμου.

Η παραγωγος g"(t) της g'(t) θα εχει επομενως αντιστοιχα τουλαχιστον ενα σημειο μηδενισμου αναμεσα σε καθε ζευγαρι διαδοχικων σημειων μηδενισμου της g'(t), δηλαδη συνολικα τουλαχιστον n σημεια μηδενισμου.

Συνεχιζοντας ετσι τον συλλογισμο βλεπουμε οτι η n+1 παραγωγος $g^{(n+1)}(t)$ θα εχει τουλαχιστον ενα σημειο μηδενισμου, ξ.

Ολα αυτα τα σημεια βρισκονται αναμεσα στα σημεια μηδενισμου της αρχικης συναρτησης g(t), δηλαδη στο διαστημα που οριζεται απο τα ακραια σημεια x και

x_i.

Διαφοριζοντας n+1 φορες την g(t) βλεπουμε οτι το πολυωνυμο $P_n(t)$ εξαφανιζεται, γιατι ειναι βαθμου n, ενω η n+1 παραγωγος της ω(t) ειναι (n+1)!. Ετσι προκυπτει:

$$0 = g^{(n+1)}(\xi) = f^{(n+1)}(\xi) - (n+1)!\frac{f(x) - P_n(x)}{\omega(x)} \qquad (7.37)$$

σχεση ισοδυναμη με την (7.32).

Το Μέγεθος του Σφάλματος Παρεμβολής

Οι τυποι (7.33), (7.34) μας φανερωνουν πολλα για το μεγεθος του σφαλματος παρεμβολης.

Κατ' αρχην ειναι φανερο οτι αν το σημειο x βγει εξω απο το διαστημα που οριζεται απο τα ακραια σημεια παρεμβολης, εστω τα x_0 και x_n, και απομακρυνεται απο αυτα ειτε προς τα δεξια ειτε προς τα αριστερα, τοτε μεγαλωνει η τιμη της |ω(x)| γιατι ολοι οι παράγοντές της, $|x - x_i|$, μεγαλωνουν ταυτοχρονως. Επισης μεγαλωνει ενδεχόμενα το μεγιστο της $|f^{(n+1)}(\xi)|$ στο διαστημα I_x , γιατι το διαστημα αυτο διευρυνεται. Αρα το φραγμα του σφαλματος που δινεται απο την (7.34) συνολικα μεγαλωνει.

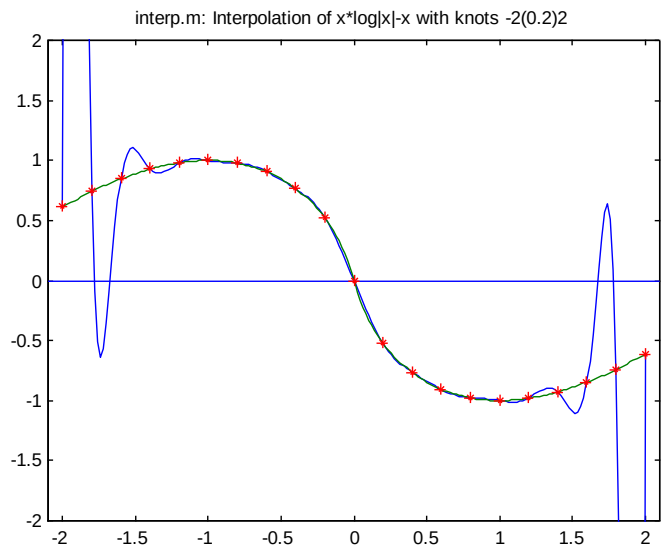
Αυτο σημαινει οτι για να προσεγγισουμε ικανοποιητικα την f(x) σε καποιο διατημα τιμων του x *δεν* μας συμφερει να χρησιμοποιησουμε σημεια παρεμβολης που να βρισκονται εξω απο αυτο το διαστημα.

Ετσι, συνηθως παιρνουμε, αν μπορουμε, τα σημεια παρεμβολης ισομοιρασμενα δεξια και αριστερα απο το σημειο x που μας ενδιαφερει.

Ενα αλλο σημειο που πρεπει να προσεξουμε ειναι οτι, αν παιρνουμε σε ενα διαστημα [a, b] ισαπεχοντα σημεια σε αποστασεις h= (b-a)/n, η ποιοτητα της προσεγγισης με αυξανομενον αριθμο σημειων n αντι να βελτιωνεται χειροτερευει!! Το σφαλμα γινεται μεν μηδεν σε εναν αυξανομενο αριθμο σημειων, αλλα το μεγιστο του σφαλματος στα ενδιαμεσα των σημειων αυτων γινεται ολο και μεγαλυτερο.

Ακριβεστερα, με αυξανομενο αριθμο σημειων παρεμβολης, n, το σφαλμα τεινει γενικα προς μηδεν σε μιαν περιοχη γυρω απο το μεσο του διαστηματος [a, b], ενω προς τα ακρα εμφανιζει αυξανομενα μεγιστα. Το κλασσικο παραδειγμα αυτης της συμπεριφορας ειναι τα-ο παραδειγμα παρεμβολης με τιμές της συναρτησης $f(x) = \frac{1}{1+x^2}$ με n+1 ισαπεχοντα σημεια κατανεμημένα π. χ., στο διαστημα [-5,5]. Ο K.Runge εδειξε οτι το πολυωνυμο παρεμβολης $P_n(x)$ συγκλινει με $n\to\infty$ προς την τιμη f(x) μονο για |x| χ 3,63.. και αποκλινει εξω απο αυτο το διαστημα (Βλ.E.K.Blum: Numerical Analysis and Computation; Theory and Practice, Addison Wesley, 1972 σελ.374-376).

Σαν παράδειγμα της συμπεριφοράς του πολυωνύμου παρεμβολής με ισαπέχοντες κόμβους (ισαπέχοντα σημεία παρεμβολής) παραθέτουμε εδώ την γραφική παράσταση της συνάρτησης x|x|-x και του πολυωνύμου παρεμβολής της στα σημεία x_k=-2+0.2k, k=0(1)20, δηλαδή ενός πολυωνύμου παρεμβολής εικοστού βαθμού.



Διάγραμμα 7-1

Ενας τροπος να αποφυγουμε αυτην την συμπεριφορα του σφαλματος ειναι να μην παρουμε τα σημεια παρεμβολης ισαπεχοντα, αλλα να τα πυκνωσουμε προς τα ακρα του διαστηματος. Αν π.χ. χρησιμοποιησουμε τα σημεια Chebyshef:

$$x_j = \frac{b+a}{2} - \frac{b-a}{2}\cos\left(\frac{2j+1}{2n+2}\pi\right),\ j = 0(1)n \qquad (7.38)$$

και η συναρτηση f ειναι στο διαστημα [a, b] δυο φορες συνεχως διαφορισιμη τοτε το σημειακο σφαλμα τεινει με αυξανομενο n για ολα τα σημεια του διαστηματος στο μηδεν. Ειδικωτερα ισχυει για τα σημεια αυτα (Βλ.Conte - de Boor: Elementary Numerical Analysis, McGraw-Hill, 1972, σελ 232):

$$\max_{x\in[a,b]} \left|\prod_{j=0}^{n}(x-x_j)\right| = 2\left(\frac{b-a}{4}\right)^{n+1} \qquad (7.39)$$

Πολυ πιο απλος τροπος να πετυχουμε καλες προσεγγισεις απο το να χρησιμοποιουμε πολυωνυμα παρεμβολης μεγαλου βαθμου με μη ισαπεχοντα σημεια παρεμβολης ειναι το να χωρισουμε το διαστημα [a, b] σε μικροτερα υποδιαστηματα και να χρησιμοποιησουμε ενα διαφορετικο πολυωνυμο παρεμβολης μικρου βαθμου για καθε τετοιο υποδιαστημα. Ετσι καταληγουμε στις τμηματικα πολυωνυμικες προσεγγισεις με τις οποιες θα ασχοληθουμε εκτενεστερα λιγο αργοτερα.

Ενα παραδειγμα του πως κατασκευαζουμε τμηματικα πολυωνυμικες προσεγγισεις με δοσμενη ακριβεια προσεγγισης δινουμε αμεσως παρακατω στο δευτερο παραδειγμα που ακολουθει.

Παρατήρηση: Ένας βασικός λόγος για τον οποίο η παρεμβολή υστερεί ως τρόπος προσέγγισης μίας λείας συνάρτησης είναι ότι το πολυώνυμο παρεμβολής $P_n(x)$ στα σημεία $x_i \in [a,b]$, i=0(1)n, μπορεί να έχει στο διάστημα αυτό μέχρι και n-1 σχετικά ακρότατα (μέγιστα και ελάχιστα εναλλάξ), τα σημεία μηδενισμού του

$P_n'(x)$. Αντίθετα, η προσεγγιζόμενη συνάρτηση, f(x), μπορεί να έχει στο ίδιο διάστημα πολύ λιγώτερα ακρότατα, οπότε η μορφή της διαφέρει αισθητά από αυτήν του $P_n(x)$. Για τον λόγο αυτό, αν θέλουμε να διατηρήσουμε την μορφή της, επιλέγουμε είτε λείες τμηματικά πολυωνυμικές προσεγγίσεις (splines), είτε την μέθοδο ελαχίστων τετραγώνων με πολυώνυμο περίπου βαθμού m, όπου m-1 ο αριθμός των σχετικών ακρότατων της f(x) στο διάστημα [a,b].

Παράδειγμα Προσδιορισμού του Σφάλματος Παρεμβολής

Να προσδιορισθει το πολυωνυμο παρεμβολης Lagrange, της συναρτησης f(x) = cos x στα σημεια:

$x_0 = -\pi/4$, $x_1 = -\pi/6$, $x_2 = 0$, $x_3 = \pi/6$, $x_4 = \pi/4$

καθως και το αντιστοιχο σφαλμα.

Λύση

Επειδη ισχύει ότι $\cos(\pm\pi/4) = \sqrt{2}/2, \cos(\pm\pi/6) = \sqrt{3}/2$ το πολυώνυμο παρεμβολής ειναι:

$$P_4(x) = \frac{24^2}{\pi^4}\left\{\frac{2\sqrt{2}}{5}x^2(x^2-\frac{\pi^2}{36}) + 1(x^2-\frac{\pi^2}{16})(x^2-\frac{\pi^2}{36}) - \frac{9\sqrt{3}}{10}x^2(x^2-\frac{\pi^2}{16})\right\}$$

ενω συμφωνα με τον τυπο (7.32) ειναι:

$$f(x) - P_4(x) = \frac{f^{(5)}(\xi)}{5!}(x-x_0)(x-x_1)...(x-x_n)$$

δηλαδη

$$\cos x - P_4(x) = \frac{-\sin\xi}{5!}x(x^2-\frac{\pi^2}{16})(x^2-\frac{\pi^2}{36})$$

με

$$\min\left\{\frac{-\pi}{4}, x\right\} < \xi < \left\{\frac{\pi}{4}, x\right\}$$

Για $x \in \left[\frac{-\pi}{4}, \frac{\pi}{4}\right]$ είναι $I_x = \left[\frac{-\pi}{4}, \frac{\pi}{4}\right]$ και $|\sin\xi| \le \sin\frac{\pi}{4} = \frac{\sqrt{2}}{2}$

αρα

$$|\cos x - P_4(x)| \le \frac{\sqrt{2}}{2\cdot 5!}|x(x^2-\frac{\pi^2}{16})(x^2-\frac{\pi^2}{36})|$$

Για $x < -\frac{\pi}{4}$ ειναι $I_x = \left[x, \frac{\pi}{4}\right]$ και για $x > \frac{\pi}{4}$ ειναι $I_x = \left[\frac{-\pi}{4}, x\right]$.

Και στις δυο περιπτωσεις το μεγιστο του |sin ξ| στο I είναι |sin x| για |x|<π/2 γιατι το ημιτονο ειναι στο διαστημα αυτο μονοτονα αυξανομενη συναρτηση.

Δηλαδη για $x \in \left[\frac{-\pi}{2}, -\frac{\pi}{4}\right]$ ή $x \in \left[\frac{\pi}{4}, \frac{\pi}{2}\right]$ ειναι:

$$|\cos x - P_4(x)| \le \frac{|\sin x|}{5!} |x(x^2 - \frac{\pi^2}{16})(x^2 - \frac{\pi^2}{36})|$$

Παράδειγμα Προσδιορισμού του Κατάλληλου Βήματος για την κατασκευή ενός Πίνακος Τιμών μίας Συνάρτησης

Να προσδιορισθει το βημα h για την κατασκευη ενος πινακα με ισαπεχουσες τιμες της f(x) = ln x στο διαστημα [1, 2] ετσι ωστε η προσεγγιση της f(x) σε τυχον σημειο x ∈ [1, 2] με δευτεροβαθμιο πολυωνυμο παρεμβολης βασιζόμενο σε τρεις διαδοχικές τιμές του πινακα να εχει μεγιστο σημειακο σφαλμα ε = 10^{-6}.

Λύση

Συμφωνα με τον τυπο του σφαλματος παρεμβολής θα ειναι

$$\ln x - P_2(x) = \frac{|f^{(3)}(\xi)|}{3!} |(x - x_{i-1})(x - x_i)(x - x_{i+1})$$

οπου για x ∈ [x_{i-1}, x_{i+1}] ⊆ [1, 2] ειναι και ξ ∈ [x_{i-1}, x_{i+1}] δηλαδή ξ ∈ [1, 2].

Στο διαστημα αυτο ειναι:

$$|f^{(3)}(\xi)| \le \frac{2}{|\xi|^3} \le 2$$

ενω με x = x_i + th προκυπτει για x ∈ [x_{i-1}, x_{i+1}] = [x - h, x +h], αρα |t|≤ 1:

$|\omega(x)| = |(x - x_{i-1})(x - x_i)(x - x_{i+1})| = h^3 |t(t^2 - 1)|$

Στο διαστημα -1 ≤ t ≤ 1 η το πολυωνυμο $t(t^2 - 1) = t^3 - t$ εχει ακροτατα για:

$0 = 3t^2 - 1$, δηλ. $t = \pm 1/\sqrt{3}$

Αρα

$$\max_{|x - x_i| \le h} |\omega(x)| = \frac{2h^3}{3\sqrt{3}}$$

και

$$\max_{|x - x_i| \le h} |f(x) - P_2(x)| = \frac{2}{3!} \frac{2h^3}{3\sqrt{3}}$$

Για να ειναι η τελευταια ποσοτητα μικροτερη απο ε = 10^{-6} αρκει να ειναι

$$h < \left(\frac{9\sqrt{3}}{2}\right)^{1/3} 10^{-2} = 1.9827 \cdot 10^{-2} < 2 \cdot 10^{-2}$$

Αρκει λοιπον να παρουμε h = 0.02 , δηλ. n = 1/h = 50.

7.4 Παρεμβολή με ομαλές τμηματικά πολυωνυμικές προσεγγίσεις (Splines)

Οπως ειπαμε και παραπανω, αν δοθουν για μιαν συναρτηση f(x) n+1 ισαπεχοντα σημεια παρεμβολης:

$$(x_i, f_i), \; i = 0(1)n, \; x_i = a + ih, \; h = (b-a)/n,$$

σε καποιο διαστημα [a, b] το πολυωνυμο παρεμβολης βαθμου n, που περναει απο τα σημεια αυτα δεν αποτελει γενικα καλην προσεγγιση της f(x) στα ενδιαμεσα των σημειων παρεμβολης οταν το n ειναι μεγαλο. Επι πλεον η χρηση αυτου του πολυωνυμου εχει το μειονεκτημα οτι ο υπολογισμος καθε τιμης του απαιτει n πολλαπλασιασμους και n+1 διαιρεσεις (αν εχουμε την λαγκρανζιανη μορφη του).

Μπορούμε βέβαια να προσεγγίσουμε την συνάρτηση με ένα πολυώνυμο μικρού βαθμού και με χρήση όλων των παραπάνω τιμών, αν καταφύγουμε στην Μέθοδο Ελαχίστων Τετραγώνων, για την οποία θα μιλήσουμε αργότερα. Όμως, η καμπύλη του προσεγγίζοντος πολυωνύμου έχει μεν τότε εν γένει την μορφή της καμπύλης της f(x), αλλά δεν διέρχεται ακριβώς από τα δοθέντα σημεία. Έτσι έχουμε διατήρηση της μορφής, αλλά χωρίς μεγάλη ακρίβεια προσέγγισης.

Αν, όμως, θέλουμε να έχουμε πιστότητα και της μορφής και της αριθμητική προσέγγισης τότε καταλήγουμε στην χρήση τμηματικά πολυωνυμικών προσεγγίσεων και ειδικότερα ομαλόποιημένων τμηματικά πολυωνυμικών προσεγγίσεων, που συντομώτερα λέγονται splines (= τερίζια), από έναν παλαιό ναυπηγικό όρο.

Οι τμηματικα πολυωνυμικες προσεγγισεις προσεγγιζουν την f(x) με ενα διαφορετικο παρεμβολικο πολυωνυμο μικρου βαθμου σε καθε ένα από τα υποδιαστήματα $[x_k, x_{k+1}]$, k = 0(1)n.

Αν δεν βαλουμε αλλες απαιτησεις, προφανως τα πολυωνυμα αυτα μπορεί να ειναι πρωτοβαθμια, και η f(x) προσεγγιζεται με μιαν πολυγωνικη γραμμη αποτελουμενη απο ευθυγραμμα τμηματα. Προσεγγισεις αυτου του ειδους χρησιμοποιουνται σε εφαρμογες, οπως η λυση προβληματων συνοριακων τιμων, οπου ο ογκος των υπολογισμων ειναι γενικα μεγαλος και δεν θελουμε να τον αυξησουμε με χρηση πολυωνυμων ανωτερου βαθμου.

Υπαρχουν ομως πολλες περιπτωσεις που επιδιωκουμε την περιγραφή μίας δοσμένης καμπύλης με *ομαλή* τμηματικά πολυωνυμική προσέγγιση (δηλαδή χωρίς τα «σπασίματα» μίας πολυγωνικής γραμμής). Αυτο συμβαινει, π.χ., κατα την αριθμητική περιγραφη του περιγράμματος της ατρακτου ενος αυτοκινητου, πλοιου ή αεροπλανου.

Σε τετοιες περιπτωσεις μπορουμε να χρησιμοποιησουμε για καθε υποδιαστημα $[x_i, x_{i+1}]$, αντί για ένα πρωτοβάθμιο πολυώνυμο, για παράδειγμα, ενα τριτοβαθμιο πολυωνυμο, $s_i(x)$ βάζοντας κάποιες πρόσθετες απαιτήσεις ομαλής μετάβασης από το ένα πολυώνυμο στο επόμενό του.

Η αύξηση του βαθμού του πολυωνύμου μας δίνει ένα περιθώριο ελευθέρων επιλογών. Οι συνθηκες παρεμβολης $s_k(x_k) = f_k$, $s_k(x_{k+1}) = f_{k+1}$ δεν αρκουν τώρα για τον μονοσήμαντο προσδιορισμο των συντελεστων των τεσσαρων δυναμεων, x^j, j = 0, 1, 2, 3 στο πολυωνυμο αυτο. Αυτο μας δινει ομως την ευχερεια να απαιτησουμε το να ειναι η συνολικη μας προσεγγιση οσο το δυνατον πιο ομαλη. Ετσι απαιτουμε να προσαρμοζονται τα πολυωνυμα αυτα με τετοιον τροπο μεταξυ τους, ωστε στους κομβους x_k να έχουν κοινή εφαπτομένη και κοινή

καμπυλότητα, δηλαδή να συμπιπτουν τοσο οι πρωτες οσο και οι δευτερες παραγωγοι τους.

Ετσι η συνολικη προσεγγιση s(x), που συντιθεται απο αυτα τα επι μερους πολυωνυμα, θα ειναι παντου τουλαχιστον δυο φορες συνεχως διαφορισιμη.

Συνολικα εχουμε λοιπον τις συνθηκες:

$$s_k(x_k) = f_k\ ,\ s_k(x_{k+1}) = f_{k+1},\ s_{k-1}'(x_k) = s_k'(x_k),\ s_{k-1}''(x_k) = s_k''(x_k) \qquad (7.39)$$

οπου $s_k(x)$ τριτοβαθμιο πολυωνυμο και

$$s(x) = s_k(x),\ \text{για}\ x \in [x_k, x_{k+1}],\ k = 0(1)n. \qquad (7.40)$$

Μια τετοια προσεγγιση ονομαζεται τριτοβαθμια συναρτηση Spline.

Εξ αιτιας της ομαλοτητας τους, οι προσεγγισεις με splines δεν διαφερουν πουθενα πολυ απο την f(x) και προσεγγιζουν ικανοποιητικα, οχι μονο την ιδια, αλλα και την πρωτη δευτερη και τριτη παραγωγο της, αν βεβαια ειναι και αυτη αρκετες φορες (τουλαχιστον τεσσερις) διαφορισιμη. Μειονεκτημα τους ειναι μονον οτι δεν μπορουμε να γραψουμε απ'ευθειας το καθε πολυωνυμο $s_k(x)$, αλλα για τον προσδιορισμο ολων των πολυωνυμων που απαρτιζουν το spline πρεπει εν γενει να λυσουμε ενα γραμμικο συστημα.

Ακριβεστερα, αυτο ισχυει αν οι κομβοι x_k δεν ειναι ισαπεχοντες. Για ισαπεχοντες κομβους υπαρχει τυπος που να περιγραφει αμεσα τα πολυωνυμα (Βλ. Hildebrand: Numerical Analysis, Mc Graw - Hill, 1974, σελ. 483) αλλα ο υπολογιστικος ογκος ειναι μικροτερος αν λυσουμε απ'ευθειας το συστημα.

Για να προσδιορισουμε αυτο το συστημα οριζουμε:

$$d_k := s_{k-1}''(x_k) = s_k''(x_k) = s''(x_k),\ k = 0(1)n \qquad (7.41)$$

Επειδη το $s_k''(x)$ ειναι πρωτοβαθμιο πολυωνυμο με:

$$s_k''(x_k) = d_k\ ,\ s_{k+1}''(x_{k+1}) = d_{k+1} \qquad (7.42)$$

συμφωνα με τον τυπο παρεμβολης Lagrange ειναι:

$$s_k''(x) = d_k \frac{x - x_{k+1}}{x_k - x_{k+1}} + d_{k+1} \frac{x - x_k}{x_{k+1} - x_k} = d_k \frac{x - x_{k+1}}{h} + d_{k+1} \frac{x - x_k}{h}\ , \qquad (7.43)$$

Ολοκληρωνοντας τα δυο μερη της (7.43) απο x_k ως x εχουμε:

$$s_k'(x) - s_k'(x_k) = -\frac{d_k}{h}\left\{\frac{(x - x_{k+1})^2}{2} - \frac{h^2}{2}\right\} + \frac{d_{k+1}}{h}\frac{(x - x_k)^2}{2} \qquad (7.44)$$

και με μιαν ακομη ολοκληρωση απο x_k ως x θα παρουμε:

$$s_k(x) = f_k + s_k'(x_k)(x - x_k) + \frac{d_k}{6h}(x_{k+1} - x)^3 + \frac{d_{k+1}}{6h}(x - x_k)^3 + \frac{d_k h}{2}(x - x_k) - \frac{d_k h^2}{6}$$

(7.45)

Ο αγνωστος συντελεστης $s_k'(x_k)$ μπορει να απαλειφθει απο αυτην την σχεση αν παρουμε υπόψη ότι:

$$f_{k+1} = s_k(x_{k+1}) = f_k + s_k'(x_k)h + \frac{d_k}{6h}h^3 + \frac{d_k h}{2}h - \frac{d_k h^2}{6}\ , \qquad (7.46)$$

Επιλυοντας αυτην την σχεση ως προς $s_k'(x_k)$ και αντικαθιστωντας το $s_k'(x_k)$ την προηγουμενη με την τιμη αυτην εχουμε:

$$s_k(x) = \frac{d_k}{6h}(x_{k+1}-x)^3 + \frac{d_{k+1}}{6h}(x-x_k)^3 + \frac{d_k h}{2}(x-x_k) + \left(\frac{f_{k+1}}{h} - \frac{d_{k+1}h}{6}\right)(x-x_k) +$$

$$+\left(\frac{f_k}{h} - \frac{d_k h}{6}\right)(x_{k+1}-x) \qquad (7.47)$$

Για τον προσδιορισμο των συντελεστων d_k μπορουμε τωρα να αξιοποιησουμε τις συνθηκες συνεχειας $s_{k-1}'(x_k) = s_k'(x_k)$, k=1(1)n-1.

Αυτες ειναι οι μονες που δεν χρησιμοποιηθηκαν ως τωρα, γιατι βαζοντας

$s_k(x_{k+1}) = f_{k+1} = s_{k+1}(x_{k+1})$ και $s_k''(x_{k+1}) = d_{k+1} = s_k''(x_k)$

εχουμε ηδη αξιοποιησει την συνεχεια των s και s".

Απο τις συνθηκες συνεχειας της s', $s_{k-1}'(x_k) = s_k'(x_k)$, παιρνουμε με την βοηθεια της (7.47):

$$d_{k-1} + 4d_k + d_{k+1} = \frac{6}{h^2}\left(f_{k+1} - 2f_k + f_{k-1}\right), k = 1(1)n-1 \qquad (7.48)$$

δηλαδη, n-1 εξισωσεις με τους n+1 αγνωστους d , i = 0(1)n.

Για να εξασφαλισουμε την μονοσημαντη επιλυσιμοτητα του συστηματος αυτου χρειαζομαστε επομενως ακομα δυο εξισωσεις. Αυτες προκυπτουν συνηθως με το να προδιαγραφουν οι τιμες ειτε των πρωτων ειτε των δευτερων παραγωγων του spline στα δυο ακρα:

(α) $s''(x_0) = f_0''$, $s''(x_n) = f_n''$, δηλ. $d_0 = f_0''$, $d_n = f_n''$. (7.49)

ή (β) $s'(x_0) = f_0'$, $s'(x_n) = f_n'$ (7.50)

Το συστημα που προκυπτει και στις δυο περιπτωσεις ειναι τριδιαγωνιο με υπερισχυουσα διαγωνιο. Αρα ειναι μονοσημαντα επιλυσιμο.

Για τις προσεγγισεις με τετοια splines αποδεικνυονται οι ακολουθες ανισοτητες, οπου j = 0, 1, 2, 3:

(α) $$\max_{x\in[a,b]} |s^{(j)}(x) - f^{(j)}(x)| \le \left(B_j + \frac{A_j}{24}\right) h^{4-j} \max_{x\in[a,b]} |f^{(4)}(x)|, \qquad (7.51)$$

(β) $$\max_{x\in[a,b]} |s^{(j)}(x) - f^{(j)}(x)| \le$$

$$\le B_j h^{4-j} \max_{x\in[a,b]} |f^{(4)}(x)| + D_j h^{5-j} \max_{x\in[a,b]} |f^{(5)}(x)| \qquad (7.52)$$

Οι ανισότητες αυτές σχύουν, προφανώς, αν η f(x) εχει τουλαχιστον τμηματικα συνεχη τεταρτη παραγωγο στην περιπτωση (α) και τμηματικα συνεχη πεμπτη παραγωγο στην περιπτωση (β). Οι σταθερές έχουν εδώ τις ακόλουθες τιμές:

$A_0 = 1/4$, $A_1 = 1$, $A_2 = 6$, $A_3 = 12$

$B_0 = 1/384$, $B_1 = r\ 3/216$, $B_2 = 1/12$, $B_3 = 1/2$

$D_0 = 1/240$, $D_1 = 1/60$, $D_2 = 1/10$, $D_3 = 1/5$.

Για τις αποδειξεις των σχεσεων αυτων βλεπε: M.H.Schultz: Spline Analysis, Prentice - Hall, 1973 σελ.57-58, και σελ.61.

Οι σχεσεις αυτες μας δειχνουν οτι η spline προσεγγιζει ομοιομορφα καλα σε ολο το διαστημα [a,b], οχι μονον την συναρτηση f(x), αλλα και τις παραγωγους της μεχρι την τριτη παραγωγο, οπου το σφαλμα προσεγγισης της f (x) με την s (x) εχει την ταξη μεγεθους του h (j = 0(1)4), οπου h η αποσταση των σημειων παρεμβολης.

Παράδειγμα Προσέγγισης με Τριτοβάθμια Συνάρτηση Spline

Να προσεγγισθει η συναρτηση f(x) = sin(πx) στο διαστημα [0,1] με τριτοβάθμια συνάρτηση spline και κομβους: $x_0 = 0$, $x_1 = 1/3$, $x_2 = 2/3$, $x_3 = 1$.

Λύση

Για $x \in [x_k, x_{k+1}]$ το αντίστοιχο πολυώνυμο $s_k(x)$ του spline δίνεται από την σχέση (7.47), οπου οι συντελεστες $d_k = s_k''(x_k)$ ειναι λυσεις του συστηματος:

k=1: $$d_0 + 4d_1 + d_2 = \frac{6}{h^2}\left(f_2 - 2f_1 + f_0\right) = -27$$

k=2: $$d_1 + 4d_2 + d_3 = \frac{6}{h^2}\left(f_3 - 2f_2 + f_1\right) = -27$$

(οπου h = 1/2, $f_0 = f_3 = 0$, $f_1 = f_2 = \sqrt{3}/2$).

Αν χρησιμοποιησουμε τις συνοριακες συνθηκες (α), δηλαδή αν προδιαγραψουμε την τιμη της δευτερης παραγωγου στα ακρα:

$$d_0 = s''(x_0) = f_0'' = 0,\ d_n = s_n''(x_n) = f_n'' = 0$$

τοτε το συστημα γινεται

$$4d_1 + d_2 = -27\sqrt{3},\ d_1 + 4d_2 = -27\sqrt{3}$$

οποτε ειναι:

$$d_1 = d_2 = -\frac{27}{5}\sqrt{3}$$

Αν χρησιμοποιησουμε τις συνοριακες συνθηκες (β), δηλαδη προδιαγραψουμε την τιμη της πρωτης παραγωγου στα ακρα, εχουμε:

$$s'(x_0) = f_0' = \pi \cos 0 = \pi,\ s'(x_3) = f_3' = \pi \cos \pi = -\pi$$

οπου

$$s_k'(x) = -\frac{d_k}{2h}(x_{k+1} - x)^2 + \frac{d_{k+1}}{2h}(x - x_k)^2 + \frac{f_{k+1} - f_k}{h} + \frac{(d_{k+1} - d_k)h}{6}$$

δηλαδη

$$d_1 + 2d_0 = 27 - 18\pi,\ 2d_3 + d_2 = 27 - 18\pi$$

Ετσι το συνολικο προς επιλυση συστημα ειναι:

$$2d_0 + d_1 = 27 - 18\pi$$
$$d_0 + 4d_1 + d_2 = -27$$
$$d_1 + 4d_2 + d_3 = -27$$
$$d_2 + 2d_3 = 27 - 18\pi$$

Η λυση αυτου του συστηματος ειναι:

$$d_0 = d_3 = -72 + 10\pi,\ d_1 = d_2 = -9 + 2\pi$$

Ασκήσεις

1. Προσεγγιστε την τιμη της συναρτησης f(x) = ln x στο σημειο x = 2,5 με χρηση παρεμβολης σε μορφη Newton με τις τιμες f(1) = 0, f(1.5) = 0.4054651, f(2) = 0.6931472, f(3) = 1.0986123. Ποσο μεγαλο ειναι το πολυ το απολυτο σφαλμα προσεγγισης;

2. Συγκρινετε την προσεγγιση της συναρτησης f(x) = cos(πx) στο διαστημα

[-1, 1] με τα πολυωνυμα παρεμβολης:

(α) στα ισαπεχοντα σημεια $x_j = -1 + 2j/n$, $j = 0(1)n$, με n=3.

(β) στα σημεια Chebysheff $x_j = -\cos\left(\frac{2j+1}{2n+2}\pi\right), j = 0(1)n$ με n=3.

Δωστε στις δυο περιπτωσεις ενα ανω φραγμα της απολυτης τιμης του σφαλματος προσεγγισης για $x = -1 + 1/n$ (όπου n=3).

3. Να προσεγγισθει η συναρτηση $f(x) = \cos \pi x$ στο διαστημα [0,1] με κυβικη spline και κομβους: $x = 0$, $x = \pm 1/3$, $x = \pm 2/3$, $x = \pm 1$.

Ποιο ειναι το μεγιστο απολυτο σφαλμα προσεγγισης της f(x) και των παραγωγων της με αυτον τον τροπο;

4. Με τι ακριβεια μπορουμε να υπολογισουμε την ποσοτητα $\sqrt{117}$ χρησιμοποιωντας το πολυωνυμο παρεμβολης της συναρτησης $f(x) = \sqrt{x}$ στα σημεια $x_0 = 100$, $x_1 = 121$, $x_2 = 144$;

5. Προσδιοριστε την προσεγγιση με κυβικη συναρτηση spline με s"(0)=0, s"(3) = 0, για μιαν συναρτηση f(x) με f(0)=2, f(1)=6, f(2) = 12, f(3) = 9. Χρησιμοποιωντας αυτην την προσεγγιση βρειτε προσεγγιστικες τιμες των f'(1.5) και f'(2.5).

6. Αν ειναι γνωστες οι τιμες $y_k = f(x_k)$, $k = 0(1)n$ μιας συναρτησης $y = f(x)$ και η ακολουθια τιμων $\{y_k\}$ ειναι μονοτονα αυξανομενη ή μειουμενη τοτε μπορουμε να χρησιμοποιησουμε τα ζευγη (x_k, y_k) για να προσεγγισουμε την αντιστροφη συναρτηση, $x = \varphi(y)$, στο διαστημα $\mathrm{int}(y_0, y_n)$ με ενα πολυωνυμο παρεμβολης στα σημεια $x_k = \varphi(y_k)$.

α) Γραψτε την γενικη μορφη του πολυωνυμου αυτου.

β) Αν για μιαν συναρτηση ειναι γνωστο οτι f(5) = 3, f(10) = 7, f(12) = 12, f(15) = 17 βρειτε την τιμη του x που αντιστοιχει σε y = 10.

γ) Αν f(4) = -9, f(6) = -3, f(8) = 5, f(10) = 8 προσδιοριστε το σημειο μηδενισμου της f(x) χρησιμοποιωντας το πολυωνυμο παρεμβολης της αντιστροφης συναρτησης.

δ) Προσεγγιστε στο διαστημα [1, 2] την αντιστροφη της συναρτησης $y = f(x) := x + \ln x$ με καταλληλο πολυωνυμο παρεμβολης παιρνοντας τις τιμες της στα σημεια $x = 1 + 0{,}25k$, $k = 0(1)4$.

ε) Προσεγγιστε στο διαστημα [-1, 1] τις αντιστροφες των συναρτησεων $y = f(x) := x + \sin x$, $y = f(x) := x - \cos x$ με πολυωνυμο παρεμβολης με χρηση των σημειων Chebysheff $x_j = -\cos\left(\frac{2j+1}{2n+2}\pi\right), j = 0(1)n$ με n = 3.

8. ΠΟΛΥΔΙΑΣΤΑΤΗ ΠΑΡΕΜΒΟΛΗ

8.1 Πολυωνυμική Παρεμβολή και Ρητή Παρεμβολή για Συναρτήσεις με Δύο Μεταβλητές

Ενας τροπος να προσεγγισουμε μιαν συναρτηση π.χ. των δυο μεταβλητων, x και y, με χρηση παρεμβολης ειναι να παρουμε μιαν ορθογωνια κατανομη κομβων (x_i, y_j), i = 0(1)m, j = 0(1)n και να εφαρμοσουμε μονοδιαστατη παρεμβολη μια φορα ως προς x και μια ως προς y. Ετσι προκυπτει αμεσα οτι το πολυωνυμο m (το πολυ) βαθμου ως προς x και n (το πολυ) βαθμου ως προς y που ικανοποιει τις συνθηκες:

$$P(x_i, y_j) = f_{ij}\ ,\ \ i = 0(1)m,\ j = 0(1)n \qquad (8.1)$$

ειναι

$$P(x,y) = \sum_{i=0}^{m}\sum_{j=0}^{n} f_{ij} L_i(x) L_j(y) \qquad (8.2)$$

οπου $L_i(x)$, $L_j(y)$ τα γνωστα μας στοιχειωδη πολυωνυμα του Lagrange που εχουν την τιμη 1 στον κομβο με τον αντιστοιχο δεικτη i, j και μηδενιζονται σε ολους τους αλλους κομβους $\{x_r\}$, $\{y_s\}$.

$$L_i(x) = \prod_{r=0, r\neq i}^{m}\left(\frac{x - x_r}{x_i - x_r}\right), \quad L_j(y) = \prod_{s=0, s\neq i}^{n}\left(\frac{x - x_s}{x_i - x_s}\right) \qquad (8.3)$$

Αν ομως η κατανομη των κομβων παρεμβολης στο επιπεδο *δεν* ειναι ορθογωνια τοτε η υπαρξη και μονοσημαντοτητα του πολυωνυμου παρεμβολης *δεν* ειναι εξασφαλισμενες, οπως ειπαμε ηδη στην αρχη του κεφαλαιου 7. Αν θελησουμε να προσδιορισουμε τους συντελεστες της συναρτησης:

$$\varphi(x, y) := axy + bx + cy + d \qquad (8.4)$$

ετσι ωστε να ισχυει:

$$\varphi(x_i\ , y_i\ ) = f_i\ ,\ i = 1(1)4 \qquad (8.5)$$

διαπιστωνουμε οτι η οριζουσα του συστηματος αυτου μηδενιζεται π.χ. οταν τα σημεια (x_i ,y_i) βρισκονται πανω σε μιαν ευθεια, η πανω σε μιαν υπερβολη.

Ενας απλος τροπος να κατασκευασουμε μιαν μονοσημαντα ωρισμενη παρεμβολικη προσεγγιση εχει δοθει απο τον Shepard. Για να κατασκευασει μιαν συναρτηση SL(x,y) που συμπιπτει με την f(x, y) στα σημεια $(x_k, y_k)^T$ δηλ. με:

$$SL(x_k, y_k) = f(x_k, y_k) = f_k\ ,\ k = 0(1)n \qquad (8.6)$$

χρησιμοποιει την ακολουθη παραλλαγη του τυπου παρεμβολης Lagrange:

$$SL(x,y) = \sum_{k=0}^{n} f_k \prod_{i=0, i\neq k}^{n} \frac{\|(x,y) - (x_k, y_k)\|^p}{\|(x_i, y_i) - (x_k, y_k)\|^p} \qquad (8.7)$$

οπου 7.7 τυχουσα διανυσματικη νορμ, και p > 0 τυχων εκθετης.

Παράδειγμα Παρεμβολής Shepard - Lagrange

Αν βαλουμε p = 2 και 7.7= 7.7_2 δηλ. την ευκλειδια νορμ στον παραπανω τυπο εχουμε τον τυπο πολυωνυμικης παρεμβολης:

$$SL(x,y)=\sum_{k=0}^{n} f_k \prod_{i=0,i\neq k}^{n} \frac{(x-x_k)^2+(y-y_k)^2}{(x_i-x_k)^2+(y_i-y_k)^2} \qquad (8.8)$$

Ενας αλλος ενδιαφερων τυπος παρεμβολης, οπου η προσεγγιζουσα συναρτηση *δεν* ειναι ομως πολυωνυμο αλλα σε ωρισμενες περιπτωσεις πηλικο πολυωνυμων, οφειλεται επισης στον Shepard και ειναι:

$$S(x,y)=\sum_{k=0}^{n} f_k r_k(x,y) \qquad (8.9)$$

με

$$r_k(x,y)=\frac{\|(x,y)-(x_k,y_k)\|^{-p}}{\sum_{j=0}^{n}\|(x,y)-(x_j,y_j)\|^{-p}} \qquad (8.10)$$

οπου p > 0, και 7(x,y)7 τυχουσα διανυσματικη νορμ.

Οι συναρτησεις $r_k(x,y)$ εχουν την ιδιοτητα ότι

$$r_k(x_i,y_i) = \delta_{ik}, \text{ με } \delta_{ik} = 0 \text{ για } i\neq k, \text{ και } \delta_{kk}=1 \qquad (8.11)$$

και επομενως εξασφαλιζουν αμεσα την ισχυ των συνθηκων παρεμβολης.

Παράδειγμα Ρητής Παρεμβολής Shepard

Αν 7.7= 7.7_p , $p \geq 1$, τοτε ειναι:

$$S(x,y)=\frac{1}{\sum_{j=0}^{n}\left(|x-x_j|^p+|y-y_j|^p\right)^{-1}}\sum_{k=0}^{n}\frac{f_k}{|x-x_j|^p+|y-y_j|^p} \qquad (8.12)$$

Η προσεγγιση S(x,y) εχει σε συγκριση με την λαγκρανζιανη παρεμβολη το πλεονεκτημα οτι, ασχετως του ποσα ειναι τα σημεια παρεμβολης και που βρισκονται, ποτε δεν φευγει πολυ εξω απο την περιοχη τιμων της συναρτησης f(x,y), γιατι για ολα τα $(x,y)^T \in \mathbf{R}^2$ ειναι:

$$\min_{i=0(1)n} f_i \leq S(x,y) \leq \max_{i=0(1)n} f_i \qquad (8.13)$$

επειδη

$$r_k(x,y) \in [0, 1] \text{ και } \sum_{k=0}^{n} r_k(x,y)=1. \qquad (8.14)$$

Παρ' ολα αυτα η ποιοτητα της προσεγγισης με την S(x,y) δεν ειναι ιδαιτερα καλη (το σφαλμα ειναι εδω αναλογο του $n^{-1/2}$ και γενικα για παρεμβολη στον $\mathbf{R}^m$, δηλ. με συναρτησεις m μεταβλητων, $n^{-1/m}$).

Η προσεγγιση αυτη εχει επισης το μειονεκτημα οτι ο υπολογισμος των τιμων της απαιτει γενικα μεγαλον ογκο πραξεων.

Ολα αυτα τα μειονεκτηματα αποφευγονται αν χρησιμοποιησει κανεις, οπως και στην μονοδιαστατη περιπτωση, μιαν *τμηματικά πολυωνυμική*

παρεμβολικη προσεγγιση με πολυωνυμα μικρου βαθμου, συνηθως πρωτοβαθμια. Οι προσεγγισεις αυτου του ειδους λεγονται Προσεγγισεις με Πεπερασμενα Στοιχεια. Αν προσεγγίζουμε με αυτά μίαν συναρτηση δύο μεταβλητών, z = f(x,y), τότε αντιστοιχούν στην προσέγγιση μίας τυχούσας καμπύλης επιφάνειας με μίαν πολυεδρική επιφάνεια, συνήθως με τριγωνικές επίπεδες έδρες, που συμπιπτει με την προσεγγιζόμενη επιφάνεια στις κορυφές των τριγωνικών εδρών.

9. ΠΡΟΣΕΓΓΙΣΕΙΣ ΜΕ ΤΗΝ ΜΕΘΟΔΟ ΕΛΑΧΙΣΤΩΝ ΤΕΤΡΑΓΩΝΩΝ

Εφαρμογές

1. Η σχεση τασης σ προς επιμηκυνση Δl/l κατα την ασκηση δυναμεων σε καποιο υλικο μπορει να μετρηθει με οπτικες μεθοδους σαν συνεχης καμπυλη σ = f(Δl/l) = f(x). Η καμπυλη αυτη εχει στην αρχη την μορφη ευθειας που ξεκιναει απο το μηδεν με καποιαν γωνια α ως προς τον οριζοντιο αξονα, και στην συνεχεια κυρτωνεται, φθανει ενα μεγιστο, και πεφτει αρκετα αποτομα στο μηδεν, που το φθανει για μιαν τιμη x:= Δl/l = δ. Μεχρις αυτο το σημειο η καμπυλη εχει περιπου την μορφη του γραφηματος της συναρτησης:

$$\varphi(x) = \frac{Ax+B}{x^2+1} + \Gamma = \frac{Ax+B}{x^2+1} - \mathrm{B}$$

οπου Γ = -Β, για να ειναι φ(0) = 0.

Για την προσαρμογη της μορφης φ(x) στην καμπυλη f(x) που προκυπτει απο τις μετρησεις μπορει να χρησιμοποιηθει παρεμβολη σε δυο σημεια, π.χ., φ(δ)= 0, φ(x) = σ , οπου x η θεση οπου η ταση, σ, φθανει την μεγιστη τιμη της. Για την καλυτερη προσαρμογη της μορφης φ(x) στις μετρησεις f(x) σε olo το διαστημα [0, δ] ενδεικνυται για τον προσδιορισμο των Α, Β να χρησιμοποιησουμε το αιτημα της μεθοδου ελαχιστων τετραγωνων:

$$\int_0^{\delta} \left(f(x) - \varphi(x)\right)^2 dx = \min$$

2. Ο πληθυσμός της Γης κατά τις τελευταίες δύο χιλιετίες εξελίχθηκε σύμφωνα με τον ακόλουθο πίνακα (βλ. ΙΣΤΟΡΙΑ ΕΙΚΟΝΟΓΡΑΦΗΜΕΝΗ, Νοεμβριος 1997, σελ. 14-15):

0	1000	1500	1750	1830	1930	1960	1975	1987	2000
200	250	460	830	1000	2000	3000	4000	5000	6000

όπου η πρώτη σειρά αναφέρεται στην χρονολογία σε έτη και η δεύτερη στον πληθυσμό σε εκατομμύρια (προφανώς έως τον 19ο αιώνα οι τιμές αυτές έχουν υπολογισθεί κατ' εκτίμηση). Να περιγραφεί η εξέλιξη αυτή με μίαν συνάρτηση της μορφής $p=e^{at+b}$, όπου p θα είναι ο πληθυσμός σε εκατομμύρια, και t η χρονολογία σε χιλιετίες (π. χ. p(1)=0.25, p(1,83)=1, p(2)=6).

9.1 Πως προκύπτει η μέθοδος αυτή

Η προσεγγιση μιας συναρτησης f(x) σε καποιο διαστημα [a, b] με ενα πολυωνυμο παρεμβολης, p(x), σε δοσμένα σημεία x_i, i=0(1)n, εχει, οπως ειδαμε, το μειονεκτημα οτι αν τα σημεια παρεμβολης ειναι ισαπεχοντα τοτε προς τα ακρα του διαστηματος, το σφαλμα στα ενδιαμεσα των σημειων παρεμβολης γινεται υπερμετρα μεγαλο. Αυτο οδηγει στο αιτημα μετακινησης των σημειων παρεμβολης ετσι ωστε να αποφευγονται οι υπερμετρες αποκλισεις του σφαλματος.

Ετσι καταληγουμε στο αιτημα βελτιστης προσεγγισης κατα Τσεμπυσεφφ:

$$\max_{i=0(1)n} |f(x_i) - p(x_i)| = \min \qquad (9.1)$$

Ένα άλλο είδος βέλτιστης προσεγγισης προκυπτει απο την απαιτηση να εχει η προσεγγιζουσα συναρτηση τόσες μεταπτωσεις (μεγιστα και ελαχιστα) όσα και η προσεγγιζόμενη συνάρτηση. Για μεγάλο πλήθος, n+1, των σημείων παρεμβολής αυτό σημαίνει ότι από το προσεγγίζον πολυώνυμο θα πρέπει να έχει εν γένει βαθμό, m, πολύ μικρότερο από τον βαθμό, n, του πολυωνύμου παρεμβολής.

Αν δεν θελουμε να απορριψουμε κανενα απο τα ζεύγη τιμών (x_i,f_i), i = 0(1)n, που μας εχουν δοθει (όπου $f_i=f(x_i)$) μπορούμε, για παράδειγμα, να απαιτήσουμε να είναι

$$\sum_{i=0}^{n} | f_i - p(x_i) | = \min \qquad (9.2)$$

όπου p(x) πολυώνυμο βαθμού $m \le n$ (μας ενδιαφερει να ειναι μικρα τα σφαλματα ασχετως του προσημου τους). Για m=n η απαίτηση αυτή μπορεί να εκπληρωθεί από το πολυώνυμο παρεμβολής, που μηδενίζει μάλιστα το παραπάνω άθροισμα, γιατί μηδενίζει κάθε προσθετέο του. Αν όμως θέλουμε να είναι m<n τότε δεν μπορούμε να πετύχουμε να είναι ακριβώς $p(x_i) = f_i$, i=0(1)n. Θα είναι μόνο $p(x_i) \approx f_i$, i=0(1)n, δηλαδή το προσεγγίζον πολυωνυμο θα περναει κοντα απο τα σημεια παρεμβολής, χωρις να περναει εν γένει ακριβώς από αυτά. Επειδη ομως η χρηση απολυτων τιμων δυσκολευει τον προσδιορισμό του $p_m(x)$ (για δοσμένο m) συχνά βαζουμε το υπολογιστικα απλουστερο αιτημα:

$$\sum_{i=0}^{n} (f_i - p(x_i))^2 = \min \qquad (9.3)$$

που μας απαλλασσει απο τις απολυτες τιμες. Ετσι καταληγουμε στο αιτημα "ελαχιστων τετραγωνων".

Αν η συναρτηση f(x) ειναι γνωστη σε ολα τα σημεια ενος διαστηματος [a, b], και θέλουμε να την προσεγγίσουμε με ένα πολυώνυμο p(x), αντι για το αιτημα αυτο μπορει να απαιτησουμε να ειναι ελαχιστο το ολοκληρωμα του τετραγωνου του σφαλματος πανω σε ολο το [a, b]:

$$\int_a^b \left(f(x) - p(x)\right)^2 dx = \min \qquad (9.4)$$

Αν τα σημεια (x_i ,f_i) προερχονται απο μετρησεις διαφορετικης αξιοπιστιας μπορουμε στον τύπο (9.3) να εισαγαγουμε θετικους συντελεστες, w_i, τα λεγομενα "βαρη", που προσδιοριζουν τι σχετικη βαρυτητα δινουμε σε καθε μετρηση. Ετσι, η (9.3) γινεται:

$$\sum_{i=0}^{n} w_i (f_i - p(x_i))^2 = \min \qquad (9.5)$$

Οσο πιο μεγαλος ειναι ενας τετοιος συντελεστης σε σχεση με τους αλλους τοσο πιο κοντα θα βρισκεται η $p(x_i)$ προς την τιμη f_i .

Παρομοια, στην (9.4) μπορουμε να εισαγαγουμε μιαν μη αρνητικη συναρτηση βαρους w(x) που να προσδιοριζει ποιες τιμες f(x) θελουμε να προσεγγιζει η p(x) ιδιαιτερα καλα. Ετσι εχουμε την απαιτηση:

$$\int_a^b w(x)\big(f(x)-p(x)\big)^2 dx = \min \qquad (9.6)$$

Εδω θα υποθεσουμε οτι η προσεγγιζουσα συναρτηση p(x) ειναι γραμμικος συνδυασμος m + 1 γραμμικα ανεξαρτητων συναρτησεων $p_k(x)$, k = 0(1)m :

$$p(x) = \sum_{k=0}^{n} c_k p_k(x) \qquad (9.7)$$

Στην περιπτωση που εξεταζουμε προσεγγισεις πανω σε ενα διακριτο συνολο σημειων $\{x_i\}$, i = 0(1)n με n > m, αυτο σημαινει οτι τυχων γραμμικος συνδυασμος των συναρτησεων $p_k(x)$, k = 0(1)m, μπορει τοτε μονον να μηδενισθει πανω σε *όλα* τα σημεια του συνολου $\{x_i\}$ οταν ολοι του οι συντελεστες μηδενιζονται. Αν εξεταζουμε προσεγγισεις πανω σε ενα διαστημα [a, b], τοτε οι συναρτησεις $p_k(x)$, k = 0(1)m, θεωρουνται γραμμικα ανεξαρτητες οταν τυχων γραμμικος συνδυασμος τους μηδενιζεται τοτε μονον για ολα τα $x \in [a, b]$ οταν ολοι του οι συντελεστες μηδενιζονται.

Υπο αυτες τις προυποθεσεις η η απαιτηση (9.5) παιρνει την μορφη:

$$P(c) := \sum_{i=0}^{n} w_i (f_i - c_0 p_0(x_i) - c_1 p_1(x_i) - \ldots - c_m p_m(x_i))^2 = \min \qquad (9.8)$$

οπου $P(c) = P(c_0, c_1, \ldots, c_m)$ πολυωνυμο των αγνωστων συντελεστων c_k.

Η ελαχιστοποιηση μιας τετοιας παραστασης γινεται πολυ απλα με μηδενισμο των μερικων παραγωγων ως προς τους αγνωστους c_k:

$$0 = \frac{\partial P(c)}{\partial c_k} = -2\sum_{i=0}^{n} w_i (f_i - c_0 p_0(x_i) - c_1 p_1(x_i) - \ldots - c_m p_m(x_i)) p_k(x_i)$$
$$= \{[f, p_k] - c_0[p_0, p_k] - \ldots - c_m[p_m, p_k]\} \qquad (9.9)$$

οπου για συντομια χρησιμοποιησαμε τον συμβολισμο:

$$[f, g] := \sum_{i=0}^{n} w_i f(x_i) g(x_i) \qquad (9.10)$$

που λεγεται "Εσωτερικο Γινομενο" των συναρτησεων f και g (ως προς τα βαρη w_i και τα σημεια x_i).

Για την ακριβεια, οι συνθηκες (9.9) προσδιοριζουν τα σημεια στασιμοτητας της P(c), αλλα στην προκειμενη περιπτωση ειναι φανερο οτι το μοναδικο σημειο στασιμοτητας ειναι θεση ελαχιστου, γιατι με $c_k \to \infty$ η P(c) τεινει στο απειρο.

Ετσι εχουμε για τον προσδιορισμο των συντελεστων που προσδιοριζουν την βελτιστη προσεγγιση της f(x) υπο την εννοια (9.8) το λεγομενο "Συστημα Κανονικων Εξισωσεων":

$$c_0[p_0, p_k] + \ldots + c_m[p_m, p_k] = [f, p_k] \quad k = 0(1)m \qquad (9.11)$$

Η λυση αυτου του συστηματος ειναι μονοσημαντα ωρισμενη. Η οριζουσα του ειναι διαφορετικη απο το μηδεν, γιατι αν το αντιστοιχο ομογενες συστημα:

$$0 = d_0[p_0, p_k] + \ldots + d_m[p_m, p_k], \; k = 0(1)m \qquad (9.12)$$

ειχε μη μηδενικη λυση, τοτε πολλαπλασιαζοντας καθε μία από τις εξισώσεις (9.12)

με τον αντιστοιχο συντελεστη d_k και αθροιζοντας εχουμε:

$$0=[\mathrm{c}_0 p_0+\ldots+\mathrm{c}_\mathrm{m} p_m, \mathrm{c}_0 p_0+\ldots+\mathrm{c}_m p_m]=\sum_{i=0}^{n} w_i\left(\mathrm{c}_0 p_0(x_i)+\ldots+\mathrm{c}_\mathrm{m} p_m(x_i)\right)^2 \tag{9.13}$$

Αυτο το αποτελεσμα ειναι αδυνατο. Οι συναρτησεις $p_k(x)$, k=0(1)m, υποτιθεται οτι διαλεχτηκαν ετσι ωστε να ειναι γραμμικα ανεξαρτητες πανω στο συνολο σημειων $\{x_i\}$, i = 0(1)n, n> m, πραγμα που σημαινει οτι ενας γραμμικος συνδυασμος τους μπορει τοτε μονον να μηδενιζεται σε ολα αυτα τα σημεια αν ολοι οι συντελεστες του ειναι μηδεν.

Με παρομοιο τροπο μπορουμε να επεξεργασθουμε και την απαιτηση (9.6):

$$p(c):=\int_a^b w(x)\left(f(x)-\sum_{k=0}^{m} c_k p_k(x)\right)^2 dx=\min \tag{9.14}$$

Μηδενιζοντας τις μερικες παραγωγους ως προς τους αγνωστους c_k εχουμε:

$$0=\frac{\partial p(c)}{\partial c_j}=-2\int_a^b w(x)\left(f(x)-\sum_{k=0}^{m} c_k p_k(x)\right)p_j(x)dx=$$

$$=-2\left\{(f,p_j)-\sum_{k=0}^{m} c_k(p_k,p_j)\right\} \tag{9.15}$$

οπου χρησιμοποιουμε την συντομογραφια:

$$(f,g):=\int_a^b w(x)f(x)g(x)dx \tag{9.16}$$

που λεγεται "Εσωτερικο Γινομενο" των συναρτησεως f(x), g(x) στο διαστημα [a, b] με συναρτηση βαρους $w(x) \geq 0$.

Ετσι προκυπτει το "Συστημα Κανονικων Εξισωσεων":

$$c_0(p_0, p_k) + \ldots + c_m(p_m, p_k) = (f, p_k) \quad k = 0(1)m \tag{9.17}$$

που ειναι και σε αυτην την περιπτωση μονοσημαντα επιλυσιμο.

Αν χρησιμοποιουμε πολυωνυμα για την κατασκευη μιας τετοιας προσεγγισης ανακυπτει παλι το ερωτημα κατα ποσον το σημειακο σφαλμα μειωνεται η οχι οταν αυξανουμε τον βαθμο του πολυωνυμου.

Στην περιπτωση $w(x)\equiv 1$, [a, b] = [-1, 1] μπορει πραγματι να δειχθει, οτι αν η συναρτηση f(x) εχει στο [-1,1] συνεχη δευτερη παραγωγο, τοτε για m αρκετα μεγαλο και τυχον $\varepsilon > 0$ ισχυει:

$$|f(x)-Q_m(x)|\leq \frac{\varepsilon}{\sqrt{m}} \to 0, \quad \gamma\iota\alpha \quad m\to\infty \tag{9.18}$$

οπου $Q_m(x)$ το πολυωνυμο βαθμου m που προσεγγιζει την f(x) με την μεθοδο ελαχιστων τετραγωνων (με $w(x)\equiv 1$). Το αποτελεσμα αυτο αποδεικνυεται με την βοηθεια αναπτυγματων σε πολυωνυμα Legendre στο βιβλιο: E.Isaacson - H.B.Keller: Analysis of Numerical Methods, Wiley, 1966, σελ. 206. Αναλογα αποτελεσματα μπορουν να δειχθουν και για αλλες συναρτησεις βαρους $w(x)\geq 0$, οχι ομως για ολες (Βλ. Isaacson - Keller σελ. 209-211). Τα αποτελεσματα αυτα

ισχυουν και για τυχον διαστημα [a, b] αφου αρκει μια απλη αλλαγη μεταβλητης:

$$x = \frac{2t-(a+b)}{b-a} \tag{9.19}$$

για να μετατραπουν οι συναρτησεις του t∈[a,b] σε συναρτήσεις του x∈[-1,1].

Παράδειγμα 1

Να προσδιορισθει η λυση του τεχνικου προβληματος που υπαρχει στην αρχη του κεφαλαιου.

Λύση

Μηδενιζοντας τις μερικες παραγωγους ως προς Α και Β της

$$P(A,B) = \int_0^{\delta}\left(f(x) - \frac{Ax+B}{x^2+1} + B\right)^2 dx = \int_0^{\delta}\left(f(x) - \frac{Ax-Bx^2}{x^2+1}\right)^2 dx$$

εχουμε το συστημα:

$$\int_0^{\delta} f(x)\frac{x}{x^2+1}dx = A\int_0^{\delta}\frac{x^2}{(x^2+1)^2}dx - B\int_0^{\delta}\frac{x^3}{(x^2+1)^2}dx$$

$$\int_0^{\delta} f(x)\frac{x^2}{x^2+1}dx = A\int_0^{\delta}\frac{x^3}{(x^2+1)^2}dx - B\int_0^{\delta}\frac{x^4}{(x^2+1)^2}dx$$

Τα ολοκληρωματα του δευτερου μερους μπορουν να υπολογισθουν σεκλειστη μορφη. Αυτα ομως του πρωτου μερους υπολογιζονται μονον αριθμητικα, αφου η f(x) ειναι δοσμενη μονο σαν καμπυλη.

Μια απλουστερη παραλλαγη αυτης της διαδικασιας προκυπτει αν θεωρησουμε οτι το σημειο δ οπου μηδενιζεται η f(x) ειναι γνωστο με αρκετην ακριβεια. Βαζοντας τοτε φ(δ) = 0 εχουμε Α = Βδ και επομενως αρκει να ελαχιστοποιησουμε την παρασταση:

$$\int_0^{\delta}\left(f(x) - B\frac{\delta x - x^2}{x^2+1}\right)^2 dx$$

Ετσι παιρνουμε για το Β την μοναδικη εξισωση:

$$\int_0^{\delta} f(x)\frac{\delta x - x^2}{x^2+1}dx = B\int_0^{\delta}\left(\frac{\delta x - x^2}{x^2+1}\right)^2 dx$$

Παράδειγμα 2

Να προσδιορισθεί η συνάρτηση $p(t)=e^{at+b}$ που προσδιορίζει την εξέλιξη του πληθυσμού της Γης με βάση τα δεδομένα από το 1830 και μετά του πίνακα στην εφαρμογή 2.

Λύση

Αν εκφράσουμε τα δεδομένα από το 1830 και μετά σε χιλιετίες και εκατομμύρια έχουμε τον πίνακα τιμών:

k	*1*	*2*	*3*	*4*	*5*	*6*
t_k	1.83	1.93	1.96	1.975	1.987	2.0
p_k	1.0	2.0	3.0	4.0	5.0	6.0
ln p_k	0	0.69315	1.09861	1.38629	1.60944	1.79176

Αν εργασθούμε άτεχνα θα επιδιώξουμε την ελαχιστοποίηση της παράστασης:

$$S(a,b)=\sum_{k=1}^{6}\left(p_k-e^{at_k+b}\right)^2$$

μηδενίζοντας τις μερικές παραγώγους της ως προς a και b και θα πρέπει να λύσουμε το μη γραμμικό σύστημα:

$$0=\frac{\partial S}{\partial a}=-2\sum_{k=1}^{6}t_k\left(p_k-e^{at_k+b}\right),\quad 0=\frac{\partial S}{\partial b}=-2\sum_{k=1}^{6}(p_k-e^{at_k+b})$$

Μπορούμε όμως να χειρισθούμε το πρόβλημα πιο επιδέξια, αν συλλογισθούμε ότι $p(t) = e^{at+b}$ ισοδυναμεί με: $\ln p(t) = at+b$. Δηλαδή, αντί να επιδιώξουμε να είναι $p(t_k)\approx p_k$, όπου p_k οι τιμές του πληθυσμού που δίνονται στον παραπάνω πίνακα, μπορούμε να επιδιώξουμε να είναι $\ln p(t_k)\approx \ln p_k$. Δηλαδή ελαχιστοποιούμε την παράσταση:

$$T(a,b)=\sum_{k=1}^{6}\left(\ln\, p_k-at_k-b\right)^2$$

και έχουμε έτσι το σύστημα

$$0=\frac{\partial T}{\partial a}=-2\sum_{k=1}^{6}\left(\ln\, p_k-at_k-b\right)t_k,\quad 0=\frac{\partial T}{\partial b}=-2\sum_{k=1}^{6}\left(\ln\, p_k-at_k-b\right)$$

δηλαδή

$$a\sum_{k=1}^{6}t_k^{\,2}+b\sum_{k=1}^{6}t_k=\sum_{k=1}^{6}t_k\ln\, p_k,\quad a\sum_{k=1}^{6}t_k+6b=\sum_{k=1}^{6}\ln\, p_k$$

που είναι γραμμικό.

Παράδειγμα 3

Να βρεθει πολυωνυμο δευτερου βαθμου που να προσεγγιζει στο διαστημα [-1, 1] την συναρτηση:

$$f(x)=\frac{1}{1+x^2}$$

με την μεθοδο ελαχιστων τετραγωνων: (α) με χρηση ολων των τιμων της συναρτησεως, (β) με χρηση μονο των τιμων της στα σημεια:

$x_0 = -1$, $x_1 = -0{,}5$, $x_2 = 0$, $x_3 = 0{,}5$, $x_4 = 1$.

Λύση

α) Επειδη $p(x) = c_0 + c_1x + c_2x^2$ οι κανονικες εξισωσεις ειναι εδω:

$$c_0\int_{-1}^{1}dx + c_1\int_{-1}^{1}xdx + c_2\int_{-1}^{1}x^2dx = \int_{-1}^{1}\frac{1}{1+x^2}dx$$

$$c_0\int_{-1}^{1}xdx + c_1\int_{-1}^{1}x^2dx + c_2\int_{-1}^{1}x^3dx = \int_{-1}^{1}\frac{x}{1+x^2}dx$$

$$c_0\int_{-1}^{1}x^2dx + c_1\int_{-1}^{1}x^3dx + c_2\int_{-1}^{1}x^4dx = \int_{-1}^{1}\frac{x^2}{1+x^2}dx$$

Απο αυτες προκυπτει:

$$2c_0 + 0c_1 + \frac{2}{3}c_2 = 2\arctan 1$$

$$0c_0 + \frac{2}{3}c_1 + 0c_2 = 0$$

$$\frac{2}{3}c_0 + 0c_1 + \frac{2}{5}c_2 = 2(1-\arctan 1)$$

και επομενως με ακριβεια πεντε δεκαδικων ψηφιων μετα το κομμα:

$c_0 = 0.96239$ $c_1 = 0$ $c_2 = -0.53097$

Το ζητουμενο πολυωνυμο ειναι δηλαδη:

$P(x) = 0.96239 - 0.53097\, x^2$

Εδω μπορουσαμε βεβαια να παραλειψουμε εξ αρχης ενα μερος των εξισωσεων αν παρατηρουσαμε οτι η συναρτηση f(x) ειναι αρτια, δηλαδη συμμετρικη ως προς τον καθετο αξονα του συστηματος συντεταγμενων (f(-x) = f(x)), και οτι και το διαστημα [-1, 1], στο οποιο προσεγγιζουμε ειναι επισης συμμετρικο ως προς αυτον τον αξονα. Αυτο σημαινει οτι η προσεγγιζουσα συναρτηση θα ειναι και αυτη αρτια (p(-x) = p(x)), δηλαδη θα περιεχει μονο αρτιες δυναμεις ($p(x) = c_0 + c_2 x^2$) οποτε το συστημα κανονικων εξισωσεων ειναι:

$$c_0\int_{-1}^{1}dx + c_2\int_{-1}^{1}x^2dx = \int_{-1}^{1}\frac{1}{1+x^2}dx$$

$$c_0\int_{-1}^{1}x^2dx + c_2\int_{-1}^{1}x^4dx = \int_{-1}^{1}\frac{x^2}{1+x^2}dx$$

β) Για να προσδιορισουμε την προσεγγιση ελαχιστων τετραγωνων με χρηση μονο των μεμονωμενων σημειων που εχουν δοθει, προσεχουμε οτι εκτος απο τις συμμετριες που διαπιστωσαμε παραπανω, εδω ειναι και τα σημεια που χρησιμοποιουμε συμμετρικα τοποθετημενα ως προς τον καθετο αξονα: $x_0=-x_4$, $x_1 = x_3$, $x_2 = 0 = -x_2$. Για τον λογον αυτο θα ειναι και η προσεγγιζουσα συναρτηση συμμετρικη, δηλαδη αρτια. Για να δειξουμε ομως εδω την πληρη μορφη των κανονικων εξισωσεων δεν θα αξιοποιησουμε αυτην την συμμετρια.

Ετσι, χρησιμοποιουμε το συστημα κανονικων εξισωσεων:

$$c_0\sum_{i=0}^{4}x_i^{\,0}+c_1\sum_{i=0}^{4}x_i^{\,1}+c_2\sum_{i=0}^{4}x_i^{\,2}=\sum_{i=0}^{4}f_i$$
$$c_0\sum_{i=0}^{4}x_i^{\,1}+c_1\sum_{i=0}^{4}x_i^{\,2}+c_2\sum_{i=0}^{4}x_i^{\,3}=\sum_{i=0}^{4}x_i f_i$$
$$c_0\sum_{i=0}^{4}x_i^{\,2}+c_1\sum_{i=0}^{4}x_i^{\,3}+c_2\sum_{i=0}^{4}x_i^{\,4}=\sum_{i=0}^{4}x_i^{\,2} f_i$$

απο τις οποιες προκυπτει:

$$5\,c_0 + 0\,c_1 + 2.5\,c_2 = 3.6$$
$$0\,c_0 + 2.5\,c_1 + 0\,c_2 = 0$$
$$2.5\,c_0 + 0\,c_1 + 2.125\,c_2 = 1.4$$

και επομενως:

$$c_0 = 0.94857 \quad c_1 = 0 \quad c_2 = -0.45714$$

Το ζητουμενο πολυωνυμο ειναι δηλαδη:

$$Q(x) = 0.94857 - 0.45714\,x^2$$

Ασκήσεις

1. Με την μεθοδο Ελαχιστων Τετραγωνων να προσδιορισετε τα πολυωνυμα τριτου (το πολυ) βαθμου που προσεγγιζουν τις συναρτησεις f(x) = cos x , και g(x) = sin x στο διαστημα [-π/4,π/4] με τους ακολουθους δυο τροπους:

α. Κανοντας χρηση των τιμων των συναρτησεων αυτων μονο στα σημεια

$x_0 = -\pi/4$, $x_1 = -\pi/6$, $x_2 = 0$, $x_3 = \pi/6$, $x_4 = \pi/4$

β. Κανοντας χρηση ολου του διαστηματος.

2. Εστω οτι προσεγγιζουμε μιαν συναρτηση f(x) με την μεθοδο ελαχιστων τετραγωνων σε ενα διαστημα [a, b] που ειναι συμμετρικο ως προς το μηδεν (a = -b), με συναρτηση βαρους που ειναι αρτια (w(-x) = w(x)). Δειξτε οτι αν η f(x) ειναι αρτια (f(-x) = f(x)) τοτε ειναι και η προσεγγιση της αρτια (p(-x) = p(x)) και αν η f(x) ειναι περιττη (f(-x) = - f(x)) ειναι και η προσεγγιση της περιττη (p(-x) = - p(x)).

3.α. Δειξτε οτι αν το διαστημα [a, b] ειναι συμμετρικο ως προς το μηδεν, δηλαδή a = -b, και η συναρτηση βαρους αρτια (w(-x) = w(x)), τοτε οι αρτιες και οι περιττες συναρτησεις ειναι καθετες μεταξυ τους με την έννοια του εσωτερικού γινομένου (9.16).

β. Εστω οτι το συστημα γραμμικα ανεξαρτητων συναρτησεων που χρησιμοποιουμε για την κατασκευη της βελτιστης προσεγγισης με την εννοια της μεθοδου ελαχιστων τετραγωνων αποτελειται εναλλαξ απο αρτιες και περιττες συναρτησεις: $p(-x) = (-1)^k p(x)$, k = 0(1)m.

Χρησιμοποιωντας την ιδιοτητα (α) δειξτε οτι αν η f(x) δεν ειναι ουτε αρτια, ουτε περιττη, μπορουμε να αντικαταστησουμε το συστημα m+1 κανονικων εξισωσεων με m+1 αγνώστους για την f(x), με δυο συστηματα περίπου (m+1)/2 εξισώσεων με (m+1)/2 αγνώστους το κάθε ένα, που προσεγγιζουν χωριστα το αρτιο μερος της f(x): $f_A(x) := ½\{f(x) + f(-x)\}$, και χωριστα το

περιττο, $f_{\Pi}(x) := ½\{f(x) - f(-x)\}$.

Υπόδειξη: Με $\varepsilon(x) := f(x) - P(x)$ και $<\varepsilon_A, \varepsilon_{\Pi}> = 0$ προκυπτει:

$$<\varepsilon_A + \varepsilon_{\Pi}, \varepsilon_A + \varepsilon_{\Pi}> = <\varepsilon_A, \varepsilon_A> + <\varepsilon_{\Pi}, \varepsilon_{\Pi}>$$

4. Δειξτε οτι αν εχουμε ενα διακριτο εσωτερικο γινομενο οπου οι χρησιμοποιουμενοι κομβοι x_j ειναι συμμετρικοι ως προς το μηδεν:

$x_{n-j} = -x_j$, $j = 0(1)m$ (οπου m το ακεραιο μερος του αριθμου $(n+1)/2$), ενω τα βαρη συμμετρικα ευρισκομενων κομβων ειναι ισα:

$w_{n-j} = w_j$, $j = 0(1)\ m$ τοτε η προσεγγιση της f(x) με την μεθοδο ελαχιστων τετραγωνων ειναι αρτια αν η f(x) ειναι αρτια και περιττη αν η f(x) ειναι περιττη. Αξιοποιωντας την ιδιοτητα αυτην δειξτε, οπως ακριβως και στην ασκηση 3, πως αν $p_k(-x) = (-1)^k\ p(x)$ για $k = 0(1)m$, τοτε ακομα και στην περιπτωση που η f(x) δεν ειναι αρτια ουτε περιττη, μπορουμε να αντικαταστησουμε το συστημα κανονικων εξισωσεων με δυο μικροτερων διαστασεων που προσεγγιζουν χωριστα το αρτιο και το περιττο μερος της f(x).

5. Να προσεγγισθει με την μεθοδο ελαχιστων τετραγωνων η συναρτηση:

$$f(x) := \int_0^x \frac{\sin t}{t}\, dt$$

στο διαστημα $x \in [-\pi/2, \pi/2]$ με πολυωνυμο τριτου βαθμου και με χρηση των τιμων: $f(0)=0$, $f(\pm 0.1\pi)=\pm 0.31244$, $f(\pm 0.2\pi)=\pm 0.6147$, $f(\pm 0.3\pi)=\pm 0.89718$, $f(\pm 0.4\pi) = \pm 1.15147$, $f(\pm 0.5\pi) = \pm 1.37076$.

10. ΑΡΙΘΜΗΤΙΚΗ ΔΙΑΦΟΡΙΣΗ

Το προβλημα αριθμητικου προσδιορισμου των παραγωγων μιας συναρτησης προκυπτει κατ'αρχην οταν η συναρτηση ειναι γνωστη μονο σε μεμονωμενα σημεια, π.χ. βασει μετρησεων. Οι τυποι αριθμητικης διαφορισης χρησιμοποιουνται ομως συνηθεστερα για την μετατροπη διαφορικων εξισωσεων σε εξισωσεις διαφορων, δηλ. αναδρομικες σχεσεις που μπορουν να επιλυθουν αλγεβρικα.

Ο βασικος τροπος παραγωγης τετοιων τυπων ειναι με διαφοριση καποιας προσεγγισης της αρχικης συναρτησης, συνηθως μιας παρεμβολικης προσεγγισης της. Ετσι απο τον τυπο (7.32) για το σφαλμα παρεμβολης της f(x) στα σημεια $\{x_k\}$, k=0(1)n:

$$f(x)-P_n(x)=\frac{f^{(n+1)}(\xi)}{(n+1)!}(x-x_0)(x-x_1)\ldots(x-x_n) \qquad (10.1)$$

οπου

$$\min\{x, x_0,\ldots, x_n\} < \xi < \max\{x, x_0, \ldots, x_n\} \qquad (10.2)$$

και

$$\omega(x) := (x - x_0)(x - x_1)\ldots(x - x_n) \qquad (10.3)$$

προκυπτει με διαφοριση:

$$f'(x)-P_n'(x)=\frac{d}{dx}\left\{\frac{f^{(n+1)}(\xi(x))}{(n+1)!}\right\}\omega(x)+\frac{f^{(n+1)}(\xi(x))}{(n+1)!}\frac{d}{dx}\omega(x)$$

οπου ο η παράγωγος του $f^{(n+1)}(\xi(x))$ εμφανίζεται γιατί βάσει της (10.2) το ξ είναι συνάρτηση του x, αλλά δεν ειναι ευκολο να υπολογισθει. Αν ομως μας ενδιαφερουν οι τιμες της πρωτης παραγωγου μονο στα σημεια παρεμβολης x , τοτε βαζοντας εδω $x = x_k$ εχουμε:

$$f'(x_k)-P_n'(x_k)=\frac{f^{(n+1)}(\xi(x))}{(n+1)!}\omega'(x_k)=\frac{f^{(n+1)}(\xi(x))}{(n+1)!}\prod_{i=0,i\neq k}^{n}(x_k-x_i) \qquad (10.4)$$

επειδή είναι

$$\omega'(x_k)=\prod_{i=0,i\neq k}^{n}(x_k-x_i) \qquad (10.5)$$

Εδω ειναι $x = x_k$, αρα με βαση την (10.2):

$$\min\{x_i\} < \xi < \max\{x_i\} \quad \text{η} \quad \xi \in (x_0, x_n) \qquad (10.6)$$

αν τα σημεια x_i ειναι διατεταγμενα κατα σειρα μεγεθους.

Επομενως, καταληγουμε στο θεωρημα:

Θεώρημα 10.1

Αν ειναι γνωστες οι τιμες $f_i := f(x_i)$, i = 0(1)n, της συναρτησης f(x) και

στο διαστημα I των σημειων παρεμβολης ειναι συνεχης η $f^{(n+1)}$, τοτε ισχυει:

$$f'(x_k) - P_n'(x_k) = \frac{f^{(n+1)}(\xi(x))}{(n+1)!} \prod_{i=0, i\neq k}^{n} (x_k - x_i) \qquad (10.7)$$

και ειδικωτερα,

$$| f'(x_k) - P_n'(x_k) | = \frac{M_{n+1}}{(n+1)!} | \prod_{i=0, i\neq k}^{n} (x_k - x_i) | \qquad (10.8)$$

οπου

$$M_{n+1} := \max_{\xi\in I} |f^{(n+1)}(\xi)| \ , \ I = [x_0, x_n] \qquad (10.9)$$

Παρατήρηση: Ελαχιστοποίηση του Φράγματος του Σφάλματος

Αν τα σημεια x ειναι ισαπεχοντα:

$$x_i = x_0 + i\,h,\ i = 0(1)n \qquad (10.10)$$

τοτε $x_k - x_i = (k-i)h$ και επομενως,

$$| f'(x_k) - P_n'(x_k) | = \frac{M_{n+1}}{(n+1)!} h^n k!(n-k)! \qquad (10.11)$$

Αν στην θεση του σημειου x_k βαλουμε το σημειο x_{k+1} τοτε ο τελευταιος ορος πολλαπλασιαζεται επι (k+1)/(n-k). Ο ορος αυτος ειναι μικροτερος απο την μοναδα η ισος με αυτην για $k\leq(n+1)/2$. Επομενως, το ανω φραγμα του σφαλματος, που δινει ο τυπος (10.11), εχει την μικροτερη δυνατη τιμη για k = [(n+1)/2] (= ακέραιο μέρος του (n+1)/2), δηλαδη οταν το x βρισκεται περιπου στην μεση του διαστηματος $[x_0, x_n]$. Αυτο μας δειχνει οτι οταν τα σημεια παρεμβολης μπορουν να επιλεγουν ελευθερα, π.χ. απο εναν πινακα τιμων της f, τοτε μας συμφερει να τα παρουμε συμμετρικα γυρω απο το σημειο οπου θελουμε να υπολογισουμε την παραγωγο.

Ενας πιο αμεσος τροπος να παραγαγουμε τους συνηθεστερα χρησιμοποιουμενους τυπους αριθμητικης διαφορισης ειναι η χρηση αναπτυγματων Taylor. Χρησιμοποιωντας π.χ. τον τυπο:

$$f(x+h) = f(x) + hf'(x) + \frac{h^2}{2} f''(\xi), \quad \xi \in \mathrm{int}(x, x+h) \qquad (10.12)$$

εχουμε:

$$\frac{f(x+h) - f(x)}{h} = f'(x) + \frac{h}{2} f''(\xi) \qquad (10.13)$$

δηλαδη μιαν προσεγγιση της f'(x) με σφαλμα της ταξεως μεγεθους του h.

Αν ομως χρησιμοποιησουμε συμμετρικα προς το x ευρισκομενες θεσεις για την δημιουργια του πηλικου διαφορων εχουμε:

$$\frac{f(x+h) - f(x-h)}{2h} = f'(x) + \frac{h^2}{6} f^{(3)}(\xi), \quad \xi \in \mathrm{int}(x-h, x+h) \qquad (10.14)$$

οπου h > 0, δηλαδη μιαν προσεγγιση της f'(x) με σφαλμα της ταξεως μεγεθους του h^2.

Εδω καναμε χρηση των τυπων:

$$f(x+h)=f(x)+hf'(x)+\frac{h^2}{2}f''(x)+\frac{h^3}{6}f^{(3)}(\xi_1) \qquad (10.15)$$

$$f(x-h)=f(x)-hf'(x)+\frac{h^2}{2}f''(x)-\frac{h^3}{6}f^{(3)}(\xi_2) \qquad (10.16)$$

και της ιδιοτητας οτι αν η $f^{(3)}(x)$ ειναι συνεχης συναρτηση, τοτε το ημιαθροισμα δυο τιμων της, αφου βρισκεται αναμεσα σε αυτες τις δυο τιμες, ειναι ισο με την τιμη της συναρτησης αυτης σε καποιο ενδιαμεσο σημειο:

$$f^{(3)}(\xi_1) + f^{(3)}(\xi_2) = 2\, f^{(3)}(\xi) \qquad (10.17)$$

Παρομοια, προσθετοντας κατα μελη τους τυπους:

$$f(x+h)=f(x)+hf'(x)+\frac{h^2}{2}f''(x)+\frac{h^3}{6}f^{(3)}(x)+\frac{h^4}{24}f^{(4)}(\xi_1) \qquad (10.18)$$

$$f(x-h)=f(x)-hf'(x)+\frac{h^2}{2}f''(x)-\frac{h^3}{6}f^{(3)}(x)+\frac{h^4}{24}f^{(4)}(\xi_2) \qquad (10.19)$$

και παιρνοντας υπ'οψη οτι αν η f (x) ειναι συνεχης τοτε:

$$f^{(4)}(\xi_1) + f^{(4)}(\xi_2) = 2\, f^{(4)}(\xi) \qquad (10.17)$$

οπου ξ μια ενδιαμεση τιμη μεταξυ ξ_1 και ξ_2 εχουμε:

$$\frac{f(x+h)-2f(x)+f(x-h)}{h^2}=f''(x)+\frac{h^2}{12}f^{(4)}(\xi) \qquad (10.18)$$

με $\xi \in (x-h,x+h)$.

Παρατήρηση: Η επίδραση των σφαλμάτων στρογγύλευσης

Οι τύποι (10.14) και (10.18) μας δίνουν την εντύπωση ότι προσεγγίζουν την τιμή αντίστοιχα της πρώτης και δευτέρας παραγώγου με τόσο μεγαλύτερη ακρίβεια όσο μικρότερο επιλέξουμε το βήμα h. Η εντύπωση όμως αυτή είναι παραπλανητική, γιατί εξ αιτίας των σφαλμάτων στρογγύλευσης, η ακρίβεια της προσέγγισης μειώνεται αν πάρουμε εξαιρςτικά μικρές τιμές για το h.

Για να εξετάσουμε ακριβέστερα αυτό το φαινόμενο θα υιοθετήσουμε την σύμβαση ότι μικρά γράμματα παριστάνουν τις ακριβείς τιμές των υπολογιζόμενων μεγεθών, ενώ τα αντίστοιχά τους κεφαλαία παριστάνουν τις προσεγγιστικές τιμές που μας δίνει ο υπολογιστής. Έτσι, αν f(a) είναι η ακριβής τιμή της συνάρτησης f στην θέση a, τότε F(a) είναι η προσέγγιση αυτής της τιμής που μας δίνει ο υπολογιστής. Οι τιμές αυτές διαφέρουν, γιατί ο υπολογιστής διαθέτει για την αποθήκευση κάθε αριθμού έναν πεπερασμένο χώρο μνήμης, δηλαδή καταγράφει μόνο ένα πεπερασμένο πλήθος δεκαδικών ψηφίων κάθε αποτελέσματος. Αν, για παράδειγμα, μία απλή θέση μνήμης ενός υπολογιστή μπορεί να χωρέσει ένδεκα δεκαδικά ψηφία μετά την υποδιαστολή, τότε όλοι οι υπολογισμοί συνδέονται με ένα σφάλμα της τάξης του 10^{-11}, αν και στην οθόνη του Η/Υ συνήθως εμφανίζονται μόνο τα πρώτα εννέα δεκαδικά μετά την υποδιαστολή. Έτσι θα είναι $F(a)=f(a)\pm\rho$, όπου $\rho\cong 10^{-11}$ (ή 10^{-9}, αν δεν λάβουμε υπόψη τα δύο μη άμεσα εμφανή δεκαδικά ψηφία). Τα δύο πρόσθετα, μη παρουσιαζόμενα στην οθόνη, δεκαδικά μπορούμε να τα εμφανίσουμε αν αφαιρέσουμε από το αποτέλεσμα που έχει στην μνήμη ο Η/Υ

την προσέγγιση που εκτυπώνει. για παράδειγμα, η εντολή √2 θα εμφανίσει στην οθόνη τον αριθμό 1.414213562. Αν τώρα δώσουμε την εντολή √2 - 1.414213562 τότε στην οθόνη θα εμφανισθεί: $3.7*10^{-10}$.

Κάποιες γλώσσες προγραμματισμού δίνουν και την δυνατότητα χρήσης διπλάσιας ακρίβειας (double precision), πράγμα που σημαίνει ότι για την αποθήκευση κάθε αποτελέσματος χρησιμοποιούνται *δύο* θέσεις μνήμης αντί για μία. Αυτό διπλασιάζει τα δεκαδικά ψηφία που κρατάει στην μνήμη ο Η/Υ. Η γλώσσα MATLAB δίνει την δυνατότητα εκτέλεσης πράξεων με οσαδήποτε θέλουμε δεκαδικά ψηφία, αλλά δεν πρέπει να ξεχνούμε ότι ο διπλασιασμός της ακρίβειας, δηλαδή των δεκαδικών ψηφίων με τα οποία γίνονται οι πράξεις, σημαίνει τετραπλασιασμό του χρόνου εκτέλεσης.

Η ανάλυση αυτή των υπολογιστικών σφαλμάτων ίσως φαίνεται επουσιώδης, αλλά δεν είναι έτσι, γιατί τα σφάλματα στρογγύλευσης επηρεάζουν δραματικά την ακρίβεια της προσέγγισης για πολύ μικρές τιμές του βήματος h. Ακριβώς αντίστοιχα αποτελέσματα ισχύουν μάλιστα και κατά την αριθμητική λύση διαφορικών εξισώσεων.

Για να κατανοήσουμε την επίδραση των σφαλμάτων στρογγυλεύσεως αρκεί να μεταγράψουμε τους τύπους (10.14) και (10.18) χρησιμοποιώντας την παραπάνω σύμβαση. έτσι θα έχουμε αντί για τον (10.14):

$$f'(x) = \frac{(F(x+h)\pm\rho)-(F(x-h)\pm\rho)}{2h} - \frac{h^2}{6} f^{(3)}(\xi) =$$

$$= \frac{F(x+h)-F(x-h)}{2h} - \frac{h^2}{6} f^{(3)}(\xi) \pm \frac{2\rho}{2h}$$

Δηλαδή,

$$\left| f'(x) - \frac{F(x+h)-F(x-h)}{2h} \right| \le \frac{h^2}{6} | f^{(3)}(\xi) | + \frac{\rho}{h} \le \frac{h^2}{6} M_3 + \frac{\rho}{h} = E_1(h), \quad (10.19)$$

όπου

$$M_3 = \max_{\xi\in[x-h,x+h]} | f^{(3)}(\xi) | \quad (10.20)$$

Αν δεχθούμε ότι $\rho \cong 10^{-11}$, τότε έχουμε:

$$E_1(10^{-3}) = 10^{-6}(M_3/6) + 10^{-8} \cong 10^{-6}(M_3/6)$$
$$E_1(10^{-4}) = 10^{-8}(M_3/6) + 10^{-7} \cong 10^{-7}$$
$$E_1(10^{-5}) = 10^{-10}(M_3/6) + 10^{-6} \cong 10^{-6}$$

Βλέπουμε, δηλαδή, ότι το φράγμα $E_1(h)$ έχει μίαν ελάχιστη τιμή περίπου 10^{-7} την οποία φθάνει για $h=10^{-4}$, ενώ για μικρότερες τιμές του h η τιμή του αυξάνεται πάλι.

Αντίστοιχα, για την αριθμητική προσέγγιση της δευτέρας παραγώγου ισχύει με βάση τον τύπο (10.18):

$$f''(x) = \frac{(F(x+h)\pm\rho) - 2(F(x)\pm\rho) + (F(x-h)\pm\rho)}{h^2} - \frac{h^2}{12} f^{(4)}(\xi)$$

δηλαδή,

$$f''(x) - \frac{F(x+h) - 2F(x) + F(x-h)}{h^2} = -\frac{h^2}{12} f^{(4)}(\xi) \pm \frac{4\rho}{h^2}$$

και επομένως,

$$\left| f''(x) - \frac{F(x+h) - 2F(x) + F(x-h)}{h^2} \right| \le \frac{h^2}{12} | f^{(4)}(\xi) | + \frac{4\rho}{h^2} \le$$

$$\le \frac{h^2}{12} M_4 + \frac{4\rho}{h^2} = E_2(h) \qquad (10.21)$$

όπου

$$M_4 = \max_{\xi \in [x-h, x+h]} | f^{(4)}(\xi) | \qquad (10.22)$$

Αν δεχθούμε ότι $\rho \cong 10^{-11}$, τότε έχουμε:

$E_2(10^{-2}) = 10^{-4}(M_4/12) + 4*10^{-7} \cong 10^{-4}(M_4/12)$

$E_2(10^{-3}) = 10^{-6}(M_4/12) + 4*10^{-5} \cong 4*10^{-5}$

$E_2(10^{-4}) = 10^{-8}(M_4/12) + 4*10^{-3} \cong 4*10^{-3}$

Βλέπουμε, δηλαδή, ότι το φράγμα $E_2(h)$ είναι πολύ πιο ευαίσθητο στην μείωση του βήματος h. Έχει μίαν ελάχιστη τιμή περίπου $4*10^{-5}$, την οποία φθάνει για $h=10^{-3}$, ενώ για μικρότερες τιμές του h η τιμή του αυξάνεται πάλι.

Για παράδειγμα, αν $f(x)=x^4$, τότε $f''(x)=12x^2$, και για x=2 έχουμε f''(2)=48. Η προσέγγιση αυτής της τιμής που παίρνουμε με $h=10^{-4}$ είναι:

$$\frac{\Delta^2 F(x)}{h^2} = \frac{F(x+h) - 2F(x) + F(x-h)}{h^2} = \frac{(2.0001)^4 - 2^5 + (1.9999)^4}{10^{-8}} = 47.99$$

και επομένως έχουμε ένα σφάλμα $f''(2)-F''(2)=48-47.99=10^{-2}$

Αν θέλουμε να έχουμε καλύτερες προσεγγίσεις των παραγώγων, τότε πρέπει να χρησιμοποιήσουμε υπολογισμούς με διπλάσια ακρίβεια. Αν τότε είναι, π. χ., $\rho \cong 10^{-25}$, οι ελάχιστες τιμές των φραγμάτων σφάλματος που προκύπτουν με βάση τους παραπάνω τύπους θα είναι: $E_1(10^{-9}) \cong 10^{-16}$ και $E_2(10^{-7}) \cong 4*10^{-11}$.

Ασκήσεις:

1. Τι βήμα h πρέπει να χρησιμοποιήσουμε και με τι ακρίβεια εκτέλεσης των πράξεων για να μπορεσουμε να προσδιορισουμε την παραγωγο f'(x) της συναρτησης $f(x) = \ln(1 + x^2)$ στο διαστημα [1, 2] με την βοηθεια του τυπου (10.14) με ακριβεια 5 δεκαδικων ψηφιων μετα την υποδιαστολη ($\varepsilon \le 0.5\ 10^{-5}$);

2. Απαντήστε τις ίδιες ερωτήσεις για την δεύτερη παράγωγο.

11. ΑΡΙΘΜΗΤΙΚΗ ΟΛΟΚΛΗΡΩΣΗ

Παραδείγματα Φυσικών και Τεχνικών Εφαρμογών

1. Για την περιοδο, Τ, ενος μαθηματικου εκκρεμους ισχυει ο τυπος:

$$\frac{T}{\tau}=\int_0^{\pi/2}\frac{d\varphi}{\sqrt{1-k^2\sin^2\varphi}}=\int_0^1\frac{du}{\sqrt{(1-u^2)(1-k^2u^2)}},\ k:=\sin\left(\frac{a}{2}\right)$$

οπου τ:= $4(l/g)^{1/2}$, l το μηκος του εκκρεμους, g η επιταχυνση της βαρυτητας, και α η μεγιστη γωνια αποκλισης του εκκρεμους απο την θεση ηρεμιας (Βλ. Pipes - Harvill, Applied Mathematics for Engineers and Physicists, McGraw - Hill, 1970, σελ. 598-601).

Να προσδιορισθει για δοσμενο τ η περιοδος Τ οταν α = 60^0.

2. Για την ενεργεια, Ε, που εκπεμπεται απο ενα απολυτα μαυρο σωμα οταν εχει απολυτη θερμοκρασια Τ ισχυει:

$$\frac{E}{\varepsilon}=\int_0^\infty\frac{x^3dx}{e^x-1},\quad \varepsilon:=\frac{2\pi}{c^3h^3}(kT)^4$$

Να υπολογισθει το ολοκληρωμα αυτο με ακριβεια τριων δεκαδικων θεσεων (BIT 19(1979), σελ.552).

11.1 Ολοκλήρωση με Χρήση Παρεμβολής σε Ισαπέχοντα Σημεία

Οπως εχει αποδειχθει ηδη απο τα μεσα του $19^{ου}$ αιωνα απο τον Liouville, το αοριστο ολοκληρωμα μιας πεπερασμενης εκφρασης που προκυπτει με συνδυασμο των στοιχειωδων συναρτησεων (x^p, e^x, ln x, sin x, cos x) δεν μπορει παντα να εκφρασθει επισης με παρομοια μορφη. Έτσι, δεν μπορουμε να υπολογισουμε τα ωρισμενα ολοκληρωματα απλά με χρηση τιμων των στοιχειωδων συναρτησεων. Ακομα και για την ολοκληρωση απλων παραστασεων που συνδυαζουν στοιχειωδεις συναρτησεις, ειναι συχνά απαραιτητο να καταφυγουμε σε μεθοδους προσεγγιστικης ολοκληρωσης. Π.χ., τα αοριστα ολοκληρωματα των παραστασεων:

$$\frac{e^x}{x},\frac{x}{\ln x},\frac{\sin x}{x},\frac{\cos x}{x},e^{-x^2},\frac{x}{\cos x},\frac{x}{\sin x}$$

δεν μπορουν να εκφρασθουν σαν πεπερασμενοι συνδυασμοι στοιχειωδων συναρτησεων.

Ενας τροπος να υπολογισει κανεις τετοια ολοκληρωματα ειναι να αναπτυξει τις ολοκληρωτεες συναρτησεις σε δυναμοσειρα γυρω απο καποιο σημειο και να ολοκληρωσει κατα μελη. Η δυναμοσειρα που προκυπτει ετσι δεν συγκλινει ομως παντοτε αρκετα γρηγορα. Για τουτο καταφευγουμε συνηθως στις πολυ πιο αμεσες και πιο απλες μεθοδους της αριθμητικης ολοκληρωσης. Αυτες προκυπτουν βασικα με ολοκληρωση μίας προσέγγισης της ολοκληρωτέας συνάρτησης με πολυώνυμο παρεμβολής σε ωρισμενα σημεια.

Αν, π.χ., προσεγγισουμε την f(x) στο διαστημα [a, b] με πολυωνυμο παρεμβολης Lagrange εχουμε:

$$\int_a^b f(x)dx = \sum_{k=0}^{n} f(x_k)\int_a^b L_k(x)dx + \int_a^b R_f(x)dx \qquad (11.1)$$

οπου $x_k \in [a,b]$ είναι τα σημεία παρεβολής, $L_k(x)$ ειναι τα στοιχειωδη πολυωνυμα Lagrange, και $R_f(x)$ το σφαλμα παρεμβολης. Ισχυει, δηλαδη, με $f_k := f(x_k)$:

$$\int_a^b f(x)dx = \sum_{k=o}^{n} \alpha_k f_k + E_n(f) \qquad (11.2)$$

οπου

$$\alpha_k := \int_a^b L_k(x)dx = \int_a^b \prod_{i=0, i\neq k}^{n} \left(\frac{x - x_i}{x_k - x_i} \right) dx \qquad (11.3)$$

και, αν η συναρτηση f εχει συνεχη παραγωγο n+1 ταξεως στο διάστημα [a,b], το σφαλμα ολοκληρωσης εχει την μορφη:

$$E_n(f) = \frac{1}{(n+1)!}\int_a^b f^{(n+1)}(\xi(x))(x - x_0)...(x - x_n)dx \qquad (11.4)$$

με

$$\min\{x_0, x\} < \xi(x) < \max\{x_n, x\} \qquad (11.5)$$

αν τα x_k ειναι αριθμημενα κατα αυξανομενη σειρα μεγεθους: $x_0 < x_1 < ... < x_n$.

Για να ειναι το σφαλμα παρεμβολης οσο το δυνατον μικροτερο, πρεπει επισης να επιλέξουμε τα x_k μεσα στο διαστημα ολοκληρώσεως I = [a,b], οποτε ειναι και $\xi \in I = [a, b]$. Για το σφαλμα ολοκληρωσης εχουμε τότε:

$$|E_n(f)| \leq \frac{1}{(n+1)!} \cdot \max_{\xi \in I} |f^{(n+1)}(\xi)| \cdot \int_a^b |(x - x_0)...(x - x_n)|\, dx \qquad (11.6)$$

Απο τον τυπο αυτον φαινεται οτι το σφαλμα μηδενιζεται αν η f(x) ειναι πολυωνυμο βαθμου μικροτερου ή ισου προς n.

Το ερωτημα που απομενει τωρα ειναι πως επιλεγουμε το πληθος και την θεση των "κομβων ολοκληρωσης" (σημειων παρεμβολης) x_k.

Οπως ειπαμε κατα την μελετη της παρεμβολης, αν εχουμε ισαπεχοντες κομβους:

$$x_k = x_0 + k\,h,\ k = 0(1)n,\ h := \frac{b-a}{n} \qquad (11.7)$$

η ποιοτητα της προσεγγισης με παρεμβολη δεν βελτιωνεται με αυξανομενο πλήθος σημείων, n, αλλ' αντιθετα προς τα ακρα του [a, b] χειροτερευει σημαντικα στα ενδιαμεσα των κομβων. Αυτο εχει σαν συνεπεια οτι στην περιπτωση ισαπεχοντων κομβων παρεμβολης το σφαλμα $E_n(f)$ δεν μειωνεται με αυξανομενο n για ολες τις συναρτησεις f.

Παραδειγματος χαριν για το ολοκληρωμα:

$$\int_{-1}^{1} \frac{dt}{1 + 25t^2} = \frac{2}{5}\arctan 5 = 0.54936$$

ο τυπος ολοκληρωσης με χρηση n+1 ισαπεχοντων σημειων δινει:

n	2	4	6	8	10	12	14	16	18
S_n	1.358	0.474	0.774	0.300	0.934	-0.062	1.579	-1.248	3.775

(Βλ. Colin W. Cryer: Numerical Functional Analysis, Oxford, 1982, σελ.125).

Για να πετυχουμε ικανοποιητικη ακριβεια, μπορουμε να πυκνωσουμε τους κομβους προς τα ακρα, ώστε να μην γίνεται το σφάλμα παρεμβολής υπ΄ρμετρα μεγάλο στα ενδιάμεσα των κόμβων παρεμβολής. Έτσι καταλήγουμε σε τύπους όπως οι τύποι ολοκλήρωσης του Gauss, στους οποίους το σφαλμα ολοκληρωσης μηδενιζεται για πολυωνυμα μεχρι βαθμου 2n+1, και συγκλινουν παντα αν η συναρτηση ειναι αρκετα ομαλη. Για αυτους θα μιλησουμε αργοτερα.

Ικανοποιητική ακρίβεια υπολογισμού ενός ολοκληρώματος μπορούμε όμως να πετύχουμε, όπως θα ιδούμε, και με κατάλληλη χρήση των απλούστερων τύπων αριθμητικής ολοκλήρωσης. Έτσι περιοριζόμαστε σε τύπους που προκύπτουν με παρεμβολή σε ένα, δύο, ή τρία ισαπέχοντα σημεία.

Οι δυο απλουστεροι απο τους τυπους που παράγονται έτσι ειναι ο τυπος μεσου σημειου, που χρησιμοποιει την τιμη της συναρτησης μονο στο μεσον του διαστηματος και ο τυπος τραπεζιου που χρησιμοποιει τις τιμες της συναρτησης στα ακραια σημεια a, b.

Ετσι εχουμε:

1. Τύπος Μέσου Σημείου

Αυτός προκύπτει άμεσα από το ανάπτυγμα Taylor της ολοκληρωτέας συνάρτησης $f \in C^2[a,b]$ γύρω από το μέσον, (a+b)/2, του διαστήματος ολοκλήρωσης:

$$f(x) = f(c) + f'(c)(x-c) + f''(\xi)\frac{(x-c)^2}{2}, \quad c = \frac{a+b}{2}$$

Ολοκληρώνοντας τα δύο μέρη αυτής της ισότητας στο διάστημα [a,b] διαπιστώνουμε ότι το ολοκλήρωμα του x-c είναι μηδέν, οπότε έχουμε:

$$\int_a^b f(x)dx = (b-a)f\left(\frac{a+b}{2}\right) + E_1(f), \tag{11.8}$$

όπου

$$E_1(f) = \int_a^b f''(\xi(x))\frac{(x-c)^2}{2}dx = f''(\eta)\int_a^b \frac{(x-c)^2}{2}dx = \frac{(b-a)^3}{24}f''(\eta), \eta \in (a,b) \tag{11.9}$$

είναι το σφάλμα ολοκληρωσης, ενώ (b-a)f(c) είναι η προσέγγιση του ολοκληρώματος. Ο μετασχηματισμός του ολοκληρώματος, που εκφράζει το σφάλμα, βασίζεται στο λεγόμενο 2ο θεώρημα μέσης τιμής του ολοκληρωτικού λογισμού, που λεει ότι, αν φ(x) και θ(x) είναι ολοκληρώσιμες συναρτήσεις, και η θ(x) *δεν* αλλάζει πρόσημο στο [a,b] (είναι π.χ. μόνιμα μη αρνητική), τότε:

$$\int_a^b \varphi(x)\theta(x)dx = \varphi(\eta)\int_a^b \theta(x)dx, \eta \in [a,b] \qquad (11.10)$$

όπου φ(η) μία κατάλληλη ενδιάμεση τιμή της φ(x). Εδώ είναι φ(x)=f″(x) και θ(x)=(x-c)^2/2.

2. Τύπος του Τραπεζίου

Για f $\in$ C^2[a, b] ειναι

$$\int_a^b f(x)dx = \frac{b-a}{2}\big(f(a)+f(b)\big) - \frac{(b-a)^3}{12} f''(\eta), \eta \in (a,b) \qquad (11.11)$$

Ο τυπος αυτος προκυπτει απο τον (11.2), (11.3) για n=2, x_0 = a, και x_1 = b. Το σφαλμα εχει εδω κατ'αρχην την μορφη (11.4). Επειδη ομως ο ορος v(x)= (x-x_0)(x-x_1) = (x-a)(x-b) δεν αλλαζει προσημο μεσα στο διαστημα [a, b], συμφωνα με το 2ο θεωρημα μεσης τιμης του ολοκληρωτικου λογισμου προκυπτει:

$$E_2(f) = \frac{1}{2}\int_a^b f''(\xi(x))(x-a)(x-b)dx = f''(\eta)\int_a^b (x-a)(x-b)dx = -\frac{(b-a)^3}{12} f''(\eta)$$

με η $\in$(a,b).

Ο συνηθεστερα χρησιμοποιουμενος τυπος ειναι όμως ο ακολουθος:

3. Τύπος του Simpson

Για f $\in$ C^4 [a, b] ειναι

$$\int_a^b f(x)dx = \frac{b-a}{6}\left\{f(a)+4f\left(\frac{a+b}{2}\right)+f(b)\right\} - \frac{(b-a)^5}{2880} f^{(4)}(\eta) \qquad (11.12)$$

με η $\in$ (a, b).

Ο τυπος αυτος προκυπτει αν χρησιμοποιησουμε παρεμβολη με δοσμενες τις τιμες της συναρτησης στα σημεια a, c = (a+b)/2, και b, αλλά και την τιμη της παραγωγου, f′(c), στο μεσαιο σημειο c.[6]

Κατα την ολοκληρωση μηδενιζεται ομως ο συντελεστης του f'(c) και ετσι απομενουν μονον οι τιμες της συναρτησης. Το σφαλμα εχει εδω κατ'αρχην την μορφη:

$$E_4(f) = \frac{1}{4!}\int_a^b f^{(4)}(\xi(x))(x-a)(x-c)^2(x-b)dx$$

με εφαρμογη ομως του 2ου θεωρηματος μεσης τιμης του ολοκληρωτικού λογισμού παιρνει την τιμη που δινεται στην σχεση (11.10).

---

[6] Η παρεμβολή, που χρησιμοποιεί και τιμές των παραγώγων της f, λέγεται παρεμβολή Hermite. Εδώ παραλείψαμε την περιγραφή της γιατί υπερβαίνει τα πλαίσια του βασικού μαθήματος Αριθμητικής Ανάλυσης. Βλεπε Κιουστελίδη, Αριθμητική Ανάλυση, Αθήνα 1996, σελ. 186-190.

Σύνθετοι Τύποι Ολοκλήρωσης

Για να πετυχουμε αποτελεσματα μεγαλης ακριβειας κατα τον υπολογισμο ολοκληρωματων, προφανως δεν μπορουμε να εφαρμοσουμε τους παραπανω τυπους στο συνολον του διαστηματος [a, b], γιατί το ευρος b-a του διαστηματος ειναι πολυ μεγαλυτερο απο την μοναδα τοτε οι οροι σφαλματος εχουν τεραστιες τιμες.

Για να εφαρμόσουμε, λοιπόν, επιτυχώς αυτούς τους τύπους, χωρίζουμε το διάστημα [a,b] σε υποδιαστήματα με εύρος μικρότερο της μονάδας, συνήθως ισομήκη. Έτσι έχουμε:

$$\int_a^b f(x)dx = \sum_{k=0}^{N} \int_{x_k}^{x_{k+1}} f(x)dx \qquad (11.13)$$

με

$$x_0 = a,\ x_k = x_0 + k\,h,\ k = 0(1)N-1,\ x_N = b,\ h := \frac{b-a}{N} \qquad (11.14)$$

και χρησιμοποιούμε τον εκάστοτε τυπο προσεγγιστικής ολοκλήρωσης χωριστά γιά καθε υποδιαστημα. Ετσι προκυπτουν οι ακόλουθοι συνθετοι τυποι προσεγγιστικής ολοκληρωσης.

1. Σύνθετος Τύπος του Τραπεζίου

Για $f \in C^2$ [a, b] ειναι

$$\int_a^b f(x)dx = h\left\{\frac{1}{2}(f_0 + f_N) + \sum_{k=1}^{N-1} f_k\right\} - (b-a)\frac{h^2}{12} f''(\eta),\ \eta \in (a,b)$$

(11.15)

οπου $f_k := f(x_k)$ και N ο αριθμος των υποδιαστηματων στα οποια χωριζουμε το [a, b]. Το σφαλμα εχει εδω κατ' αρχην την μορφη

$$E_N(f) = -\frac{h^3}{12}\sum_{k=1}^{N} f''(\eta_k),\ \eta_k \in (x_{k-1}, x_k) \qquad (11.16)$$

Ομως η μεση τιμη $\frac{1}{N}\sum_{k=1}^{N} f''(\eta_k)$ βρισκεται μεταξυ της μεγαλυτερης και της μικροτερης απο τις τιμες $f''(\eta_k)$. Επειδη η f'' υποτιθεται συνεχης κινουμενη μεταξυ αυτων των δυο τιμων θα περασει και απο καθε ενδιαμεση τιμη. Αυτο σημαινει οτι για καποιο $\eta \in$[a, b] θα ειναι και

$$f''(\eta) = \frac{1}{N}\sum_{k=1}^{N} f''(\eta_k)$$

οποτε το αθροισμα στον τυπο (11.16) μπορει να αντικατασταθει με $Nf''(\eta)$.

Οπως βλεπουμε, για το σφαλμα ισχυει εδω:

$$|E_N(f)| \le (b-a)\frac{h^2}{12}\max_{\eta \in I} |f''(\eta)|$$

Έτσι, μικραινοντας αρκετα το ευρος h των υποδιαστηματων, δηλ.

αυξανοντας το πληθος τους, μπορουμε να κανουμε το σφαλμα οσοδηποτε μικρο θελουμε (μεσα στα οριια της ακριβειας με την οποια κανει τις πραξεις και απομνημονευει τα αποτελεσματα ο Η/Υ).

2. Σύνθετος Τύπος του Simpson

Για f $\in$ C^4[a, b] και

$$x_k = a + kh,\ k = 0(1)2M,\ h = \frac{b-a}{2M} \qquad (11.17)$$

προκύπτει:

$$\int_a^b f(x)dx = \sum_{k=0}^{M-1} \int_{x_{2k}}^{x_{2k+2}} f(x)dx = \frac{2h}{6}\left\{\sum_{k=0}^{M-1}\left(f_{2k} + 4f_{2k+1} + f_{2k+2}\right)\right\} - \frac{(2h)^5}{2880}\sum_{k=0}^{M-1} f^{(4)}(\eta_k)$$

όπου και εδώ ισχύει για κατάλληλο η$\in$(a,b):

$$\frac{1}{M}\sum_{k=0}^{M} f^{(4)}(\eta_k) = f^{(4)}(\eta)$$

Έτσι, ο τύπος παίρνει τελικά την μορφή

$$\int_a^b f(x)dx = \frac{h}{3}\left\{f_0 + f_{2M} + 2\sum_{k=1}^{M-1} f_{2k} + 4\sum_{k=1}^{M} f_{2k+1}\right\} - \frac{h^5}{180} f^{(4)}(\eta),\ \eta \in (a,b),$$

(11.18)

Παρατήρηση: Η Αριθμητική Ολοκλήρωση μόνο Συνεχών Συναρτήσεων

Συμφωνα με τους παραπανω τυπους αν η συναρτηση f εχει συνεχη δευτερη παραγωγο (αν προκειται για τον συνθετο κανονα τραπεζιου) η τεταρτη παραγωγο (αν προκειται για τον συνθετο κανονα του Simpson) τοτε με h που τεινει στο μηδεν, δηλ. N που τεινει στο απειρο, το αθροισμα του δευτερου μερους συγκλινει πραγματι προς την ακριβη τιμη του ολοκληρωματος. Πως θα μπορουσαμε ομως να προσεγγισουμε τα ολοκληρωματα συναρτησεων που ειναι μονο συνεχεις;

Η απαντηση ειναι απλη: Με τους ιδιους τυπους!

Μπορει πραγματι να αποδειχθει οτι αν ενας απλος κανονας ολοκληρωσης εχει μονο θετικους συντελεστες για τις τιμες της συνεχους συναρτησης f που χρησιμοποιει και ισχυει χωρις σφαλμα για πολυωνυμα μεχρι καποιον βαθμο τοτε ο αντιστοιχος συνθετος τυπος συγκλινει προς το ολοκληρωμα της f οταν το h τεινει στο μηδεν. Η συγκλιση βεβαια μπορει να ειναι πολυ αργη αν η f ειναι μονο συνεχης . Η αποδειξη αυτου του αποτελεσματος δινεται π.χ. στις σελιδες 341-343 του βιβλιου: Isaacson - Keller: Analysis of Numerical Methods, Wiley 1966. Ειδικα για την συνθετη μεθοδο τραπεζιου μπορουμε ομως αμεσα να διαπιστωσουμε την συγκλιση αφου συμφωνα με τον ορισμο του ολοκληρωματος:

$$\int_a^b f(x)dx = \lim_{h\to 0} h\sum_{k=0}^{N} f_k = \lim_{h\to 0}\left\{T_N - \frac{h}{2}(f_0 + f_N)\right\}$$

οπου T_N ειναι η προσεγγιση του ολοκληρωματος για N ισομηκη

υποδιαστηματα συμφωνα με τον συνθετο κανονα τραπεζιου, δηλ. το δευτερο μερος της ισοτητας (11.15) αν παραλειψουμε τον ορο σφαλματος.

Παραδείγματα Εφαρμογής των Συνθέτων Τύπων

1. α) Με τον συνθετο τυπο του Simpson για δυο υποδιαστηματα (N=2) να υπολογισθει προσεγγιστικα το ολοκληρωμα:

$$J = \int_1^2 \frac{e^x}{x} dx$$

β) Να γινει εκτιμηση του σφαλματος.

γ) Σε ποσα το πολυ ισομηκη υποδιαστηματα πρεπει να χωρισθει το διαστημα [1, 2] ωστε να προκυψει ακριβεια εξη δεκαδικων ψηφιων ($E_N(f) \leq 0.5*10^{-6}$);

Λύση

α) Μπορούμε να εφαρμόσουμε τους τυπους (11.17),(11.18) με h = (b-a)/2M, δηλαδή h= 1/4. Μπορουμε επισης απλα να θεωρησουμε τα ολοκληρωματα στα διαστηματα [1,3/2] και [3/2, 2] χωριστα και να εφαρμοσουμε τον απλο κανονα του Simpson σε καθένα. Σε καθε περιπτωση προκυπτει με $f(x) = e^x/x$:

$J \approx S_2 = (1/12)\{f(1) + 4f(1.25) + 2f(1.5) + 4f(1.75) + f(2)\} = 3.060663$

β) Ισχυει

$$f^{(4)}(x) = e^x\left(\frac{24}{x^5} - \frac{24}{x^4} + \frac{12}{x^3} - \frac{4}{x^2} + \frac{1}{x}\right)$$

αρα για $x \in [1, 2]$

$$|f^{(4)}(x)| \leq e^x\left(\frac{24}{x^4}\left|1 - \frac{1}{x}\right| + \frac{4}{x^2}\left|\frac{3}{x} - 1\right| + \frac{1}{x}\right) \leq e^x\left(\frac{12}{x^4} + \frac{8}{x^3} + \frac{1}{x}\right)$$

γιατι $|1/x - 1| \leq 1/2$, $|3/x - 1| \leq 2$ και επομενως

$$|f^{(4)}(x)| \leq 21e^2$$

Για το σφαλμα $E_2(f) = J - S_2$ ισχυει επομενως:

$$|E_2(f)| \leq (2-1)\frac{(0.25)^4}{180} 21e^2 = 3.367 \cdot 10^{-3}$$

γ) Αν S_M ειναι η προσεγγιση του J για M υποδιαστηματα εύρους 2h το καθένα, και $E_M(f)$ είναι το αντιστοιχο σφαλμα, τότε αρκει να ισχυει:

$$|E_M(f)| = |J - S_M| \leq \frac{21e^2}{180}\left(\frac{1}{2M}\right)^4 \leq 0.5 \cdot 10^{-6}$$

Απο την ανισοτητα αυτην προκυπτει $M \leq 18.11$ αρα $M \leq 19$. Δηλαδη η προσεγγιση S_M πρεπει να εχει ηδη την ζητουμενη ακριβεια. Επειδη το ολοκληρωμα της e /x δεν μπορει να εκφρασθει σαν πεπερασμενη παρασταση στοιχειωδων συναρτησεων δεν εχουμε αμεσο τροπο να διαπιστωσουμε ποτε οι προσεγγισεις S εχουν για πρωτην φορα την ζητουμενη ακριβεια. Στην προκειμενη περιπτωση

μπορουμε ομως να χρησιμοποιησουμε τον πινακα τιμων του ολοκληρωματος Ei(x) που υπαρχει στην γνωστη συλλογη τυπων και πινακων: Abramowitz - Stegun: Handbook of Mathematical Functions, Dover Publications, στις σελ.240-241.

Απο τον πινακα τιμων προκυπτει:

J = Ei(2) - Ei(1) = 3.059116540 = 3.059117

Συγκρινοντας τις προσεγγισεις S με αυτην την τιμη βλεπουμε οτι ηδη για N = 16 ειναι:

S_{16} = 3.05911707 ≅ 3.059117

2. Σε ποσα το πολυ υποδιαστηματα πρεπει να χωρισθει το διαστημα [0, 1] ωστε το σφαλμα προσεγγισης του ολοκληρωματος:

$$J = \int_0^1 \ln(1+x^3)dx$$

με τον συνθετο κανονα τραπεζιου να μην υπερβαινει απολυτα την τιμη $0.5*10^{-6}$;

Λύση

$$J = \int_0^1 \ln(1+x^3)dx \approx T_N := \frac{1}{N}\left\{\frac{1}{2}(f_0+f_N)+\sum_{k=1}^{N-1} f_k\right\}$$

οπου

$f_k = \ln(1 + x_k^3) = \ln\{1 + (k/N)^3\}$, k = 0(1)N

Εδώ, με I:= [0, 1], ειναι:

$$|J - T_N| \le \frac{h^2}{12}\max_{\eta\in I}|f''(\eta)| = \frac{1}{12N^2}\max_{\eta\in[0,1]}\frac{3\eta^3(2-\eta^3)}{(1+\eta^3)^2} \le \frac{6}{12N^2} = \frac{1}{2N^2}$$

Αρκει επομενως να ειναι η τελευταια αυτη παρασταση μικροτερη απο $0.5*10^{-6}$, αρα $N^2 \ge 10^6$ ή $N \ge 10^3$. Δηλαδη υποδιαιρεση σε 1000 υποδιαστηματα αρκει για να εχουμε την ζητουμενη ακριβεια.

Ασκήσεις

1.ως 3. Σε ποσα το πολυ υποδιαστηματα πρεπει να χωρισθει σε καθε περιπτωση το διαστημα ολοκληρωσης ωστε το σφαλμα προσεγγισης των ολοκληρωματων:

$$J_1 = \int_0^1 x\sqrt{1+x^2}\,dx, \quad J_2 = \int_0^1 x\arctan x\,dx, \quad J_3 = \int_1^2 \frac{\ln x}{1+x}dx$$

με τον συνθετο κανονα τραπεζιου να μην υπερβαινει απολυτα την τιμη $0.5*10^{-6}$;

Υπόδειξη: Για την τριτη ολοκληρωτεα συναρτηση ισχυει:

$x(1+x^2)|f''(x)| = (2x+1)/x + r(x)/(x+1)$, με $r(x) := x+1-2x\ln x \le r(1) = 2$ γιατι η r(x) πεφτει μονοτονα. Αρα $|f''(x)| \le 11/6 < 2$ στο [1, 2].

11.2 Οι τύποι ολοκλήρωσης του Gauss

Στους τύπους αυτούς προσεγγίζουμε την ολοκληρωτέα συνάρτηση με παρεμβολή όχι σε ισαπέχοντα σημεία, αλλά έτσι ώστε ο τύπος ολοκλήρωσης να ισχύει ακριβώς για πολυώνυμα όσο το δυνατόν μεγαλύτερου βαθμού. Η κατασκευή τέτοιων τύπων ανάγεται στον Karl Friedrich Gauss και γίνεται με αξιοποίηση των ιδιοτήτων των λεγομένων ορθογωνίων πολυωνύμων (βλέπε Κιουστελίδη [1996, σελ. 244-250]). Οι τυποι ολοκληρωσης Gauss που αντιστοιχουν στις βασικες κατηγοριες ορθογωνιων πολυωνυμων δινονται π.χ. στις σελ. 887-890 και 916-924 της κλασσικης συλλογης τυπων και πινακων τιμων: M.Abramowitz - I.Stegun: Handbook of Mathematical Functions, Dover Publications. Εδώ δεν θα αναπτύξουμε τον τρόπο κατασκευής τους και περιοριζόμαστε να αναφέρουμε τους δύο απλούστερους τύπους Gauss- Legendre, που μπορουν ενδεχομενα να χρησιμοποιηθουν αντι για τον κανονα του Simpson:

$$\int_a^b f(x)dx = \frac{b-a}{2}\left\{f\left(c-\frac{h}{\sqrt{3}}\right)+f\left(c+\frac{h}{\sqrt{3}}\right)\right\}+\frac{(b-a)^5}{4320}f^{(4)}(\xi) \qquad (11.19)$$

$$\int_a^b f(x)dx = \frac{b-a}{18}\left\{5f\left(c-h\frac{\sqrt{15}}{5}\right)+f(c)+5f\left(c+h\frac{\sqrt{15}}{5}\right)\right\}+\frac{(b-a)^7}{201600}f^{(6)}(\xi), \qquad (11.20)$$

με $\xi \in (a, b)$, και

$$c=\frac{b+a}{2}, \quad h=\frac{b-a}{2} \qquad (11.21)$$

11.3 Γενικευμένα Ολοκληρώματα

Τα γενικευμενα ολοκληρωματα λεγονται ετσι γιατι ειναι οριαι κλασσικων ολοκληρωματων και ανηκουν σε δυο βασικες κατηγοριες: εκεινα που εκτεινονται σε διαστημα απειρου μηκους ($(-\infty,+\infty)$, ή $[a, +\infty)$, ή $(-\infty, b]$) και εκεινα που εχουν σε καποιο σημειο του διαστηματος ολοκληρωσης ειτε αλγεβρικη ανωμαλια (πολο) ειτε λογαριθμικη ανωμαλια.

Υπαρχει, βεβαια, και η περιπτωση να ειναι το διαστημα απειρο και να υπαρχουν σ' αυτο και σημεια ανωμαλιας. Στην περιπτωση αυτην μπορουμε να χωρισουμε το διαστημα ολοκληρωσης σε υποδιαστηματα, ετσι ωστε το ολοκληρωμα σε καθε υποδιαστημα να ανηκει σε μιαν μονο απο τις παραπανω κατηγοριες.

Στην περιπτωση απειρου διαστηματος το ολοκληρωμα εχει νοημα μονο αν η συναρτηση $|f(x)|$ φθινει αρκετα γρηγορα για $|x|\to\infty$, και ακριβεστερα αν

$$|f(x)| \le \frac{c}{|x|^{1+\gamma}}, \text{ γιά } |x| > M \qquad (11.22)$$

με M, c, γ θετικές σταθερές, γιατι το ολοκληρωμα απο $a>0$ ως ∞ της συναρτησης $\frac{c}{x}$ *δεν* εχει νοημα (ειναι απειρο), ενω για οποιοδηποτε $\gamma > 0$ είναι

$$\int_a^{\infty} \frac{c}{x^{1+\gamma}} dx = ca^{-\gamma}$$

Στην περιπτωση αλγεβρικης ανωμαλιας σε καποιο σημειο c, το ολοκληρωμα της f(x) απο η μεχρι το σημειο c εχει νοημα, μονον αν η |f(x)| δεν τεινει για $x \to c$ γρηγορωτερα στο απειρο απο μιαν συναρτηση της μορφης $|x-c|^{-\gamma}$ με $\gamma \in (0, 1)$, δηλ. αν

$$\lim_{x \to c} |f(x)||x-c|^{\gamma} = M \text{ , } \gamma \in (0, 1) \qquad (11.23)$$

γιατι, το ολοκληρωμα στο [a, c] της $|x-c|^{-\gamma}$ υπαρχει για $\gamma < 1$ αλλα οχι για $\gamma = 1$.

Οι βασικες μεθοδοι υπολογισμου τετοιων ολοκληρωματων ειναι οι ακολουθες:

11.3.1 Ολοκληρώματα Απείρου Διαστήματος

Αν τα ολοκληρωματα αυτα εχουν τις μορφες:

$$\int_0^{\infty} e^{-x} f(x)dx, \quad \int_{-\infty}^{\infty} e^{-x^2} f(x)dx$$

τοτε μπορουμε να χρησιμοποιησουμε για τον υπολογισμο τους τους αντιστοιχους τυπους Gauss με συναρτησεις βαρους e^{-x} και e^{-x^2} που βασιζονται αντιστοιχα στα ορθογωνια πολυωνυμα Laguerre και Hermite (βλ. Abramowitz - Stegun σελ. 890 και 923-924. Οι απλουστεροι απο αυτους ειναι:

$$\int_0^{\infty} e^{-ax} f(x)dx = \frac{1}{4a}\left\{(2+\sqrt{2})f\left(\frac{2-\sqrt{2}}{a}\right) + (2-\sqrt{2})f\left(\frac{2+\sqrt{2}}{a}\right)\right\} + \frac{f^{(4)}(\xi)}{6a^5}, \xi > 0$$

(11.24)

$$\int_{-\infty}^{\infty} e^{-a^2x^2} f(x)dx = \frac{\sqrt{\pi}}{6a}\left\{f\left(-\frac{\sqrt{6}}{a}\right) + 4f(0) + f\left(\frac{\sqrt{6}}{a}\right)\right\} + \frac{f^{(6)}(\xi)}{960a^7}, \xi \in R$$

(11.25)

Τυποι αυτου του ειδους ειναι ομως μονο τοτε χρησιμοι, αν οι παραγωγοι της f(x) ειναι φραγμενες για $\xi > 0$, η αντιστοιχα, $\xi \in \mathbf{R}$. Ετσι, για παράδειγμα, με $f(x) = x^p$ στην πρωτη περιπτωση δεν μπορουμε να φραξουμε το σφαλμα αν είναι $p \geq 5$ και στην δευτερη αν είναι $p \geq 7$. Εκτος απο τετοιες δυσκολιες, εχουμε και το προβλημα οτι πρεπει ενδεχομενα να χρησιμοποιησουμε τυπους Gauss - Laguerre και Gauss - Hermite με μεγαλο αριθμο κομβων τους οποιους πρεπει να βρουμε σε καποιον πινακα η με την βοηθεια καταλληλου προγραμματος.

Ο βασικος τροπος να αποφυγουμε τετοιες δυσκολιες ειναι να μετασχηματισουμε τα ολοκληρωματα αυτα σε ολοκληρωματα πανω σε πεπερασμενο διαστημα. Αν γενικα η f(x) εχει την μορφη:

$$f(x) = \frac{g(x)}{x^{1+\gamma}}, \gamma > 0, |g(x)| < M \text{ γιά } x > a > 0 \qquad (11.26)$$

τοτε

$$\int_a^\infty f(x)dx = \int_a^\infty \frac{g(x)}{x^{1+\gamma}}dx = \int_a^\infty g(x)d\left\{\frac{x^{-\gamma}}{\gamma}\right\} = \frac{1}{\gamma}\int_0^{a^{-\gamma}} g(t^{-1/\gamma})dt \qquad (11.27)$$

Δηλαδη το ολοκληρωμα απειρου διαστηματος μετατρεπεται σε ενα πεπερασμενου διαστηματος με την βοηθεια του μετασχηματισμου:

$$x = t^{-1/\gamma} \quad , t = x^{-\gamma} \qquad (11.28)$$

Ειδικα για $f(x) = g(x)/x^2$ ($\gamma = 1$) μπορουμε να βαλουμε $x = 1/t$.

Παραδείγματα υπολογισμού ολοκληρωμάτων απείρου διαστήματος

1. Να υπολογισθει με καταλληλο μετασχηματισμο μεταβλητης το ολοκληρωμα

$$J = \int_1^\infty \frac{e^{-x}}{x}dx$$

Λύση

Η συναρτηση f(x) εχει μεν την μορφη (11.26) με g(x) φραγμενο, αλλα εδω ειναι $\gamma = 0$. Για να εχουμε μιαν συναρτηση f(x) ακριβως της μορφης (11.26) παρατηρουμε οτι, οχι μονον η e^{-x} , αλλα και καθε συναρτηση $x^k e^{-x}$ ειναι φραγμενη για $x \in [1, \infty)$. Ετσι, π.χ., πολλαπλασιαζοντας αριθμητη και παρονομαστη της f με x εχουμε:

$$J = \int_1^\infty \frac{xe^{-x}}{x^2}dx = \int_1^\infty \frac{g(x)}{x^2}dx = \int_1^\infty g(x)d\left\{\frac{1}{x}\right\} = \int_0^1 g(1/t)dt = \int_0^1 \frac{1}{t}e^{-1/t}dt = \int_0^1 \varphi(t)dt$$

οπου $g(x) := xe^{-x}$ και $\varphi(t) := e^{-1/t}/t$. Συμφωνα με τον συνθετο τυπο τραπεζιου (11.15) ειναι:

$$\int_0^1 \varphi(x)dx = h\left\{\frac{1}{2}(\varphi_0 + \varphi_N) + \sum_{k=1}^{N-1}\varphi_k\right\} - (b-a)\frac{h^2}{12}\varphi''(\eta), \eta \in (0,1)$$

με $h = 1/N$, $x_k = k/N$ και

$$\varphi_0 = \varphi(0) = \lim_{t\to 0} g(1/t) = \lim_{x\to\infty} g(x) = \lim_{x\to\infty} x/e^x = \lim_{x\to\infty} 1/e^x = 0$$

οπου η προτελευταια ισοτητα προκυπτει με εφαρμογη του κανονα του de l'Hospital.

Για την εκτιμηση σφαλματος παρατηρουμε οτι:

$$\varphi''(t) = \frac{1}{t^5}e^{-1/t}(2t^2 - 4t + 1) = \frac{1}{t^5}e^{-1/t}(2(t-1)^2 - 1)$$

Για $t \in [0, 1]$ η $(1/t^5)e^{-1/t}$ έχει μέγιστη τιμή 21.056, που προκυπτει (με μηδενισμο της παραγωγου) για $t = 1/5 = 0.2$. Η εντος παρενθεσεως παρασταση ειναι στο [0, 1] απολυτα μικροτερη απο 3.

Δηλαδη για $t \in [0, 1]$ ειναι:

$$|\varphi''(t)| \le 3*21.056 = 63.168 < 64$$

και επομενως

$$|J - S_N| \le \frac{64}{12N^2} < \varepsilon \ \gamma\iota\alpha \ N > \left(\frac{16}{3\varepsilon}\right)^{1/2}$$

οπου S_N η προσεγγιζουσα παρασταση. Π.χ. για ε $=10^{-6}/3 = 0{,}33..10^{-6}$ προκυπτει ετσι $N \geq 4000$.

2. Να υπολογισθουν τα ολοκληρωματα:

$$J_a = \int_1^\infty \frac{\ln x}{x^{3/2}}\,dx,\ J_b = \int_1^\infty \frac{\sin x}{1+x^2}\,dx,\ J_c = \int_1^\infty \frac{x^3}{e^x - 1}\,dx$$

Λύση

α. Στο πρωτο ολοκληρωμα ο αριθμητης τεινει στο απειρο με $x\to\infty$. Για τουτο διαιρουμε τοσο τον αριθμητη οσο και τον παρονομαστη π.χ. με $x^{1/4}$ ετσι ωστε ο μεν αριθμητης να τεινει τωρα στο μηδεν, ο δε παρονομαστης να εχει ακομα εκθετη μεγαλυτερο απο την μοναδα:

$$J_a = \int_1^\infty \frac{\ln x}{x^{3/2}}\,dx = \int_1^\infty \frac{\ln x / x^{1/4}}{x^{5/4}}\,dx = \int_1^\infty \frac{g(x)}{x^{1+1/4}}\,dx = \int_1^\infty g(x)d\left\{\frac{x^{-1/4}}{-1/4}\right\}$$

με $g(x) := x^{-1/4}\ln x$ οποτε $g(\infty\infty) = \lim_{x\to\infty} g(x) = 0$, οπως προκυπτει με εφαρμογη του κανονα του del'Hospital.

Εισαγοντας την νεα μεταβλητη $t = x^{-1/4}$, $x = t^{-4}$ εχουμε τωρα:

$$J_a = 4\int_0^1 g(t^{-4})dt = 4\int_0^1 t\ln(t^{-4})dt = -16\int_0^1 t\ln t\,dt$$

οπου η $\varphi(t) := t\ln t$ μηδενιζεται και στα δυο ακρα του διαστηματος.

Το ολοκληρωμα αυτο μπορει επομενως να υπολογισθει με οποιονδηποτε συνθετο κανονα ολοκληρωσης, π.χ.τον συνθετο κανονα τραπεζιου.

Αξιο προσοχης ειναι εδω ομως οτι $\varphi''(t) = 1/t \to\infty$ για $t \to 0$ και επομενως *δεν* μπορουμε να βρουμε ενα πεπερασμενο ανω φραγμα του σφαλματος. Παρ' ολα αυτα ο συνθετος τυπος συγκλινει προς το ολοκληρωμα αφου η $\varphi(t)$ ειναι συνεχης συναρτηση. Αν χωρισουμε το [0, 1] σε N ισομηκη υποδιαστηματα τοτε για την προσεγγιση S_N του J_a προκυπτουν οι ακολουθες τιμες:

$S_{10} = 3.929496$, $S_{100} = 3.998989$, $S_{1000} = 3.999989$, $S_{10000} = 3{,}999991$

Προφανως ειναι $J_a = 4$, πραγμα που επιβεβαιωνεται αν παρουμε υπ'οψη οτι το αοριστο ολοκληρωμα της $t\ln t$ ειναι $(\ln t - 1/2)t^2/2$.

b) Εδω ο αριθμητης ειναι μεν φραγμενος για ολα τα x, αλλα δεν εχει συγκεκριμενο οριο για $x\to\infty$ για τουτο διαιρουμε και εδω τον αριθμητη και παρονομαστη π.χ. με $x^{1/2}$. Ετσι εχουμε:

$$J_b = \int_1^\infty \frac{\sin x}{1+x^2}\,dx = \int_1^\infty \frac{\sin x / x^{1/2}}{x^{-1/2} + x^{3/2}}\,dx = \int_1^\infty \frac{\sin x / x^{1/2}}{1/x^2 + 1}\,\frac{dx}{x^{3/2}} = \int_1^\infty f(x)d\left\{\frac{x^{-1/2}}{-1/2}\right\} = 2\int_0^1 f(1/t^2)dt$$

με

$$f(x) := \frac{\sin x}{x^{1/2}}\,\frac{1}{1+1/x^2}$$

οπου η συναρτηση f(x) ειναι συνεχης και φραγμενη στο διαστημα $[1,\infty)$ με $f(\infty) = 0$. Στο τελευταιο ολοκληρωμα μπορουμε οπως και στο προηγουμενο παραδειγμα να εφαρμοσουμε οποιονδηποτε συνθετο τυπο ολοκληρωσης.

c) $J_c = \int_1^{\infty} \frac{x^3}{e^x - 1} dx = \int_0^1 g(t)dt, \mu\varepsilon\ x = 1/t, g(t) = \frac{1}{t^5(e^{1/t} - 1)}$

Η συναρτηση g(t) εχει για t $\to$ 0 το οριο μηδεν (g(0) = 0) και ειναι σε ολο το [0, 1] συνεχης. Αρα μπορει και εδω να εφαρμοσθει οποιοσδηποτε τυπος συνθετης ολοκληρωσης. Η επιλογη του μετασχηματισμου x = 1/t οφειλεται στο οτι η f(x) τεινει για x $\to\infty$ προς το μηδεν πιο γρηγορα απο οποιανδηποτε αρνητικη δυναμη του x, αφου ο παρονομαστης e^x-1 περιλαμβανει οσοδηποτε μεγαλες δυναμεις του x. Θεωρωντας λοιπον οτι f(x) = g(x)/x με g(x) φραγμενο εχουμε γ = 1 στις (11.27), (11.28).

Ασκήσεις

Μετασχηματιστε σε ολοκληρωμα πεπερασμενου διαστηματος τα επομενα ολοκληρωματα και υπολογιστε τα προσεγγιστικα με τον συνθετο κανονα Simpson με δυο υποδιαστηματα:

$$J_a = \int_1^{\infty} \frac{\ln(x+2)}{(x+1)\sqrt{x}} dx, J_b = \int_1^{\infty} \frac{\cos x}{x^{3/2}} dx, J_c = \int_1^{\infty} \frac{\ln x}{x^2 + 3x + 2} dx$$

11.3.2 Ολοκληρώματα Πεπερασμένου Διαστήματος με Αλγεβρική Ανωμαλία

Η βασικη μορφη τετοιων ολοκληρωματων ειναι η ακολουθη:

$$J = \int_a^b \frac{g(x)}{(x-a)^p} dx, p < 1, g \in C^1[a,b] \qquad (11.29)$$

Το σημειο ανωμαλιας η ιδιομορφιας ειναι εδω το x = a οπου η ολοκληρωτεα συναρτηση απειριζεται αν g(a) $\neq$ 0. Το ολοκληρωμα εχει εν τουτοις νοημα. Π.χ. με g(x) = 1 η τιμη του ειναι $\frac{(b-a)^{1-p}}{1-p}$.

Απο το αποτελεσμα αυτο φαινεται και οτι η προυποθεση p < 1 ειναι απαραιτητη γιατι αλλοιως το ολοκληρωμα δεν εχει νοημα. Η προυποθεση οτι το σημειο απειρισμου της συναρτησης ειναι στο ακρο του διαστηματος ολοκληρωσεως δεν εχει ιδιαιτερη σημασια, γιατι αν το σημειο απειρισμου ειναι καποιο σημειο c $\in$ (a, b) τοτε μπορουμε να σπασουμε το ολοκληρωμα σε δυο αλλα πανω στα διαστηματα [a, c) και (c, b] που εχουν το σημειο ανωμαλιας στο πανω η το κατω ακρο τους.

Ο βασικος τροπος μετασχηματισμου του ολοκληρωματος αυτου ειναι η λεγομενη "Αποσπαση του Ανωμαλου Ορου", δηλαδη η αποσπαση του ορου $g(a)/(x-a)^p$ που περιγραφει την συμπεριφορα της ολοκληρωτεας συναρτησης στην γειτονια του σημειου a. Τοτε εχουμε:

$$J = \int_a^b \frac{g(x)}{(x-a)^p} dx = \int_a^b \frac{g(x) - g(a)}{(x-a)^p} dx + g(a)\int_a^b \frac{dx}{(x-a)^p} = \int_a^b \varphi(x)dx - g(a)\frac{(b-a)^{1-p}}{1-p}$$

(11.30)

οπου

$$\varphi(x) := \big(g(x) - g(a)\big)(x-a)^{-p}, \text{ γιά } x > a, \varphi(a) := 0 \qquad (11.31)$$

Η τιμη φ(a) = 0 προκυπτει αν στην πρωτη παρασταση (11.31) βαλουμε g(x) - g(a) = g'(ξ)(x-a) οπου ξ $\in$ (a, x), με βαση το θεωρημα μεσης τιμης του διαφορικου λογισμου οποτε εχουμε:

$$\varphi(x) = g'(\xi)(x-a)^{1-p} \to 0 \text{ για } x \to a+0,$$

δηλ. για x που τεινει στο a απο τιμες μεγαλυτερες του a.

Το ολοκληρωμα που απομενει στην τελευταια σχεση (11.30) ειναι ολοκληρωμα μιας, οχι μονον φραγμενης αλλα και συνεχους συναρτησης που μπορει να υπολογισθει π.χ. με οποιονδηποτε συνθετο τυπο ολοκληρωσης.

Στην μορφη ομως αυτη του τελευταιου ολοκληρωματος δεν μπορει να εφαρμοσθει π.χ. ο τυπος εκτιμησης σφαλματος για τον συνθετο τυπο τραπεζιου γιατι η φ(x) *δεν* εχει στο σημειο a φραγμενη παραγωγο. Αν θελουμε να εφαρμοσουμε αυτον τον τυπο εκτιμησης σφαλματος, οπότε η φ(x) θα έπρεπε να έχει συνεχή δεύτερη παράγωγο, πρεπει στο αρχικο ολοκληρωμα να αποσπασουμε απο την g(x) τουλαχιστον τρεις ορους του αναπτυγματος της κατα Taylor στο σημειο a, οποτε εχουμε:

$$\int_a^b \frac{g(x)}{(x-a)^p}dx = \int_a^b \psi(x)dx - g(a)\frac{(b-a)^{1-p}}{1-p} - g'(a)\frac{(b-a)^{2-p}}{2-p} - \frac{g''(a)}{2}\frac{(b-a)^{3-p}}{3-p}$$

,

(11.32)

με

$$\psi(x) := d(x)(x-a)^{-p}, \text{ γιά } x > a, \psi(a) := \lim_{x\to a}\psi(x) = 0 \qquad (11.33)$$

$$d(x) := g(x) - g(a) - g'(a)(x-a) - \frac{g''(a)}{2}(x-a)^2 \qquad (11.34)$$

οπου το οριο σχηματιζεται για x που τεινουν στο a απο μεγαλυτερες τιμες.

Αν τωρα στο ολοκληρωμα του δευτερου μερους της (11.32) εφαρμοσουμε τον συνθετο κανονα τραπεζιου θα χρειαστουμε για την εκτιμηση σφαλματος ενα ανω φραγμα της απολυτης τιμης της δευτερης παραγωγου της ψ(x) στο διαστημα [a, b]. Γι' αυτην ισχυει:

$$\psi''(x) = p(p+1)(x-a)^{-(p+2)}d(x) - 2p(x-a)^{-(p+1)}d'(x) + (x-a)^{-p}d''(x) =$$

$$= \left\{ p(p+1)\frac{g^{(3)}(\xi)}{3!} - 2p\frac{g^{(3)}(\zeta)}{2!} + g^{(3)}(\eta) \right\}(x-a)^{1-p}$$

(11.35)

οπου η τελευταια παρασταση προκυπτει αν χρησιμοποιησουμε τα υπολοιπα των αναπτυγματων Taylor για να εκφρασουμε τις d(x), d'(x), d''(x). Επομενως ειναι:

$$\max_{x\in[a,b]} |\psi(x)| \le (b-a)^{1-p}\left(\frac{p(p+1)}{6} + p + 1\right)\max_{\xi\in[a,b]} |g^{(3)}(\xi)|$$

(11.36)

Βασικο πλεονεκτημα της αποσπασης του ανωμαλου ορου ειναι οτι το

μεγαλυτερο μερος της τιμης του ολοκληρωματος υπολογιζεται αμεσα και το ολοκληρωμα που απομενει αποτελει διορθωτικο μονο ορο που ο έστω και χονδροειδης προσδιορισμος του δεν επηρεαζει σημαντικα το ολο αποτελεσμα. Αυτο το πλεονεκτημα ειναι τοσο σημαντικο ωστε συμφερει η αποσπαση ανωμαλου ορου ακομα και αν το ολοκληρωμα δεν αρχιζει απο το σημειο ιδιομορφιας, a, αλλα απο καποιο γειτονικο σημειο a+ε, ετσι ωστε το ολοκληρωμα να μην ειναι γενικευμενο αλλα συνηθισμενο.

Ενας αλλος τροπος για να απαλλαχθει το ολοκληρωμα (11.29) απο την αλγεβρικη ανωμαλια ειναι με μετασχηματισμο της μεταβλητης:

$$J=\int_a^b \frac{g(x)}{(x-a)^p}dx=\int_a^b g(x)d\left\{\frac{(x-a)^{1-p}}{1-p}\right\}=\frac{1}{1-p}\int_0^{(b-a)^{1-p}} g(a+t^{1/(1-p)})dt \tag{11.37}$$

με t:= $(x-a)^{1-p}$ (1 > p). Ο τροπος αυτος εχει ομως το μειονεκτημα οτι δεν αποσπαται αμεσως το μεγαλυτερο μερος του ολοκληρωματος, οπως παραπανω. Ετσι χρειαζεται περισσοτερες διαμερισεις για να φθασει την ιδια ακριβεια προσεγγισης οπως η αποσπαση του ανωμαλου ορου (Βλ. το επομενο παραδειγμα).

Παραδείγματα Υπολογισμού Ολοκληρωμάτων με Ιδιομορφίες

1. Να υπολογισθει προσεγγιστικα το γενικευμενο ολοκληρωμα:

$$K=\int_0^1 \frac{e^x}{\sqrt{x}}dx$$

(α) με αποσπαση του ανωμαλου ορου, (β) με μετασχηματισμο της μεταβλητης και στην συνεχεια με εφαρμογη του συνθετου τυπου Simpson με τον ιδιο αριθμο υποδιαστηματων στα μετασχηματισμενα ολοκληρωματα. Ποια προσεγγιση ειναι ακριβεστερη;

Λύση

α) $$K=\int_0^1 \frac{e^x}{\sqrt{x}}dx=\int_0^1 \frac{e^x-1}{\sqrt{x}}dx+\int_0^1 \frac{dx}{\sqrt{x}}=2+\int_0^1 \varphi(x)dx$$

οπου φ(x):= $(e^x-1)/\sqrt{x}$ για x > 0, φ(0):= 0.

Αν παραστησουμε με K(N) την προσεγγιση του K με τον συνθετο κανονα τραπεζιου με N υποδιαστηματα τοτε εχουμε:

K(1) = 2.859 , K(10) = 2.9202 , K(100) = 2.9251,
K(1000) = 2.925297, K(1400) = 2.9253

οπου για μεγαλυτερο αριθμο διαμερισεων η προσεγγιση φαινεται να σταθεροποιειται σε αυτην την τιμη (με μικρες διακυμανσεις στο εκτο δεκαδικο ψηφιο εξ αιτιας σφαλματων στρογγυλευσεως).

β) $$K=\int_0^1 \frac{e^x}{\sqrt{x}}dx=2\int_0^1 e^x d\left\{\sqrt{x}\right\}=2\int_0^1 e^{t^2}dt$$

Αν παραστησουμε με K*(N) την προσεγγιση του K με τον συνθετο κανονα

τραπεζιου με N υποδιαστηματα τοτε εχουμε:

K*(1) = 3.7182 , K*(10) = 2.9343 , K*(100) = 2.9254 ,
K*(1000) = 2.925305, K*(2000) = 2.925304

Στο παραδειγμα αυτο η προσεγγιση με μετασχηματισμο της μεταβλητης φαινεται να ειναι λιγωτερο ακριβης για μικρο αριθμο διαμερισεων ενω για μεγαλον αριθμο διαφερει ελαχιστα απο αυτην που αντιστοιχει σε αποσπαση ανωμαλου ορου.

2. Να υπολογισθουν προσεγγιστικα τα επομενα ολοκληρωματα με απλη χρηση του κανονα του Simpson:

$$J_a = \int_0^1 \frac{\ln(x+2)}{(x+1)\sqrt{x}}\,dx, \quad J_b = \int_0^2 \frac{\ln(x+2)}{\sqrt{x(2-x)}}\,dx$$

Λύση

α) $J_a = \int_0^1 \frac{\ln(x+2)}{(x+1)\sqrt{x}}\,dx = \int_0^1 \frac{\ln(x+2)/(x+1)}{\sqrt{x}}\,dx = \ln 2 \int_0^1 \frac{dx}{\sqrt{x}} + \int_0^1 \varphi(x)dx$

με

$$\varphi(x) := \left(\frac{\ln(x+2)}{x+1} - \ln 2\right)\frac{1}{\sqrt{x}}, \text{ γιά } x > 0, \varphi(0) := 0$$

$$J_a \approx 2\ln 2 + \frac{1}{6}\left(\varphi(0) + 4\varphi(0.5) + \varphi(1)\right)$$

β) $J_b = \int_0^2 \frac{\ln(x+2)}{\sqrt{x(2-x)}}\,dx = \int_0^1 \frac{\ln(x+2)/\sqrt{2-x}}{\sqrt{x}}\,dx + \int_1^2 \frac{\ln(x+2)/\sqrt{x}}{\sqrt{2-x}}\,dx$

$$J_1 = \int_0^1 \frac{\ln(x+2)/\sqrt{2-x}}{\sqrt{x}}\,dx = \frac{\ln 2}{\sqrt{2}}\int_0^1 \frac{dx}{\sqrt{x}} + \int_0^1 \frac{1}{\sqrt{x}}\left\{\frac{\ln(x+2)}{\sqrt{2-x}} - \frac{\ln 2}{\sqrt{2}}\right\}dx \approx$$

$$\approx 2\frac{\ln 2}{\sqrt{2}} + \frac{1}{6}\left\{4\varphi_1(0.5) + \varphi_1(1)\right\}, \quad \varphi_1(x) := \frac{\ln(x+2)}{\sqrt{2-x}} - \frac{\ln 2}{\sqrt{2}}$$

$$J_2 = \int_1^2 \frac{\ln(x+2)/\sqrt{x}}{\sqrt{2-x}}\,dx = \frac{\ln 4}{\sqrt{2}}\int_1^2 \frac{dx}{\sqrt{2-x}} + \int_1^2 \frac{1}{\sqrt{2-x}}\left\{\frac{\ln(x+2)}{\sqrt{x}} - \frac{\ln 4}{\sqrt{2}}\right\}dx \approx$$

$$\approx 2\frac{\ln 4}{\sqrt{2}} + \frac{1}{6}\left\{\varphi_2(1) + 4\varphi_2(1.5)\right\}, \quad \varphi_2(x) := \frac{\ln(x+2)}{\sqrt{x}} - \frac{\ln 4}{\sqrt{2}}$$

Κατα την εφαρμογη του κανονα του Simpson παραλειψαμε τις τιμες $\varphi_1(0) := \lim_{x\to 0} \varphi_1(x)$, $\varphi_2(2) := \lim_{x\to 2} \varphi_2(x)$ γιατι ισουνται με μηδεν.

3. Με καταλληλους μετασχηματισμους και απλη χρηση του κανονα του Simpson υπολογιστε προσεγγιστικα το ολοκληρωμα:

$$J := \int_0^1 \frac{e^x}{(1-\cos x)^{1/4}}\,dx$$

Υποδειξη: Προσεξτε οτι $\cos x \approx 1 - x^2/2$ για $x \approx 0$.

Λύση

Χρησιμοποιωντας την υποδειξη βλεπουμε οτι:

$$f(x) := \frac{e^x}{(1-\cos x)^{1/4}} \approx \frac{2^{1/4} e^x}{x^{1/2}}, \text{ γιά } x \approx 0$$

Για να φερουμε το παραπανω ολοκληρωμα στην μορφη (11.29) γραφουμε επομενως την ολοκληρωτεα συναρτηση στην μορφη:

$$f(x) := \frac{x^2/2}{(1-\cos x)^{1/4}} \frac{e^x}{x^2/2} = g(x) \frac{2^{1/4} e^x}{x^{1/2}}$$

οπου g(0) = 1. Αρα:

$$J = \int_0^1 \frac{2^{1/4} e^x g(x)}{x^{1/2}} dx = \int_0^1 \frac{2^{1/4}(e^x g(x) - e^0 g(0)}{x^{1/2}} dx + 2^{1/4} e^0 g(0) \int_0^1 \frac{dx}{x^{1/2}} =$$

$$= 2^{5/4} + 2^{1/4} \int_0^1 \varphi(x) dx, \quad \varphi(x) := \frac{e^x g(x) - 1}{x^{1/2}}, \text{ γιά } x > 0, \varphi(0) := 0$$

$$J = 2^{5/4} + 2^{1/4} \frac{1}{6} \{4\varphi(0.5) + \varphi(1)\}$$

Ασκήσεις

1-5. Με καταλληλο μετασχηματισμο να μετατραπουν σε κοινα τα επομενα γενικευμενα ολοκληρωματα και να υπολογισθουν προσεγγιστικα με καποιον κανονα αριθμητικης ολοκληρωσης:

$$K := \int_0^\pi \frac{\cos x}{\sqrt{x(\pi - x)}} dx, \; J := \int_1^2 \frac{\cos x}{(x^2-1)^{1/3}} dx, \; L := \int_0^1 \frac{dx}{(1-x^4)^{1/2}}, \; N := \int_{-1}^1 \frac{dx}{(1-x^2)^{1/3}}$$

6. Υπολογιστε το παραπανω ολοκληρωμα L με την βοηθεια του τυπου Γκαους - Τσεμπυσεφφ:

$$\int \frac{f(x)}{(1-x^2)^{1/2}} dx = \frac{\pi}{n} \sum_{k=1}^n f(x_k) + E_n(f), \quad E_n(f) = \frac{2\pi}{2^{2n}(2n)!} f^{(2n)}(\xi), \; \xi \in (0,1)$$

(11.38)

όπου $$x_k = \cos\left(\frac{2k-1}{2n}\pi\right), \; k = 1(1)n$$

(11.39)

Παρατήρηση: Η Περίπτωση Λογαριθμικής Ανωμαλίας

Η μέθοδος απόσπασης ανωμάλου όρου μπορεί να χρησιμοποιηθεί και όταν το

ολοκλήρωμα εμφανίζει λογαριθμική ανωμαλία, δηλαδή όταν είναι της μορφής:

$$L := \int_a^b g(x)\,|\ln(x-a)|^q\,dx \qquad (11.40)$$

με q > 0, g $\in$ C^1 [a, b].

Η απόσπαση ανωμάλου όρου γίνεται ως εξής:

$$L := \int_a^b \varphi(x)dx + g(a)\int_a^b |\ln(x-a)|^q\,dx \qquad (11.41)$$

με

$$\varphi(x) := (g(x) - g(a))\,|\ln(x-a)|^q,\ \varphi(a) := 0 \qquad (11.42)$$

όπου $\varphi(a) := \lim_{x\to a} \varphi(x) = 0$, γιατί g(x) - g(a) = (x - a) g'(ξ) με ξ $\in$(a, x) και με t = x-a, r = 1/q > 0 ειναι:

$\lim_{x\to a} (x-a)|\ln(x-a)|^q = \lim_{t\to 0} t|\ln t|^q = (\lim_{t\to 0} t^{1/q}(-\ln t))^q =$

$(\lim_{t\to 0} (-\ln t)/t^{-r})^q = (\lim_{t\to 0} (-1/t)/(-rt^{-r-1}))^q = (\lim_{t\to 0} (t^r/r))^q = 0$

Τα ορια παιρνονται εδω απο τα δεξια, δηλαδη για x > a που τεινει στο a και t > 0 που τεινει στο 0.

Το δευτερο ολοκληρωμα στην (11.41) χωριζεται σε ενα γενικευμενο ολοκληρωμα και ενα κοινο

$$J := \int_a^b |\ln(x-a)|^q\,dx = \int_0^{b-a} |\ln t|^q\,dt = \int_0^{\varepsilon} |\ln t|^q\,dt + \int_{\varepsilon}^{b-a} |\ln t|^q\,dt,$$

με ε $\in$ (0, 1], οπου μονο το ολοκληρωμα στο διαστημα [0, ε] ειναι γενικευμενο.

Για να το υπολογισουμε το μετασχηματιζουμε με τον ακολουθο τροπο

$$\int_0^{\varepsilon} |\ln t|^q\,dt = \int_0^{\varepsilon} (-\ln t)^q\,dt = \int_{-\ln\varepsilon}^{\infty} e^{-u}u^q du = \int_{-\ln\varepsilon}^{\infty} e^{-u}u^{q+2} d(1/u) = \int_0^{-1/\ln\varepsilon} e^{-1/s}s^{-(q+2)}ds \qquad (11.43)$$

με $t = e^{-u}$, s = 1/u, οπου η ολοκληρωτεα παρασταση στο τελευταιο ολοκληρωμα μηδενιζεται στο σημειο s = 0 και επομενως το ολοκληρωμα δεν ειναι γενικευμενο.

Αν ειδικωτερα ειναι q = 1 τοτε ο υπολογισμος του ολοκληρωματος (11.40) ειναι απλος:

$$L := \int_a^b g(x)\ln(x-a)dx = \int_a^b \varphi(x)dx + g(a)\int_a^b \ln(x-a)dx =$$

$$= \int_a^b \varphi(x)dx + g(a)(b-a)\big(\ln(b-a) - 1\big) \qquad (11.44)$$

με $\varphi(x) := (g(x) - g(a))\ln(x-a)$, $\varphi(a) := 0$ (11.45)

12. ΤΟ ΠΡΟΒΛΗΜΑ ΑΡΧΙΚΩΝ ΤΙΜΩΝ ΓΙΑ ΚΟΙΝΕΣ ΔΙΑΦΟΡΙΚΕΣ ΕΞΙΣΩΣΕΙΣ

Φυσικά και Τεχνικά Παραδείγματα

1.Ενα απο τα βασικα προβληματα της θεωριας μη γραμμικων κυκλωματων ειναι ο προσδιορισμος των περιοδικων ταλαντωσεων σε στασιμη κατασταση ενος κυκλωματος που περιεχει ενα μη γραμμικο επαγωγικο στοιχειο. Τυπικο παραδειγμα ειναι το ακολουθο:

Στα ακρα πηνιου που βρισκεται γυρω απο σιδερενιο πυρηνα εφαρμοζεται ξαφνικα περιοδικη ταση

$$u = u_0 \sin \omega t$$

Αν η αποκριση αυτης της διαταξης σε μαγνητικα πεδια ειναι μη γραμμικη και δινεται απο την σχεση

$$i = a\psi + b\psi^3$$

οπου i η ενταση του ρευματος, και ψ η μαγνητικη ροη μεσα απο το πηνιο, τοτε εχουμε την διαφορικη εξισωση

$$u = Ri + \frac{d\psi}{dt}$$

οπου R η ωμικη αντισταση του πηνιου. Με

$$y = \frac{aR\psi}{u_0}, \quad x = aRt$$

και $bu_0^2 = 2a^3R^2$, $2aR = \omega$, εχουμε τοτε το προβλημα αρχικων τιμων:

$$y' = -y - 2y^3 + \sin 2x \ , \ y(0) = 0$$

(Βλ. Collatz: The Numerical Treatment of Differential Equations, Springer, 1966, σελ.90-91).

Ιδιου τυπου προβλημα προκυπτει (αυτην την φορα για την μαγνητικη ροη ψ) και οταν η ωμικη αντισταση ειναι αμελητεα αλλα υπαρχει στο κυκλωμα πυκνωτης συνδεδεμενος σε σειρα με το πηνιο. Η διαφορικη εξισωση αυτη ειναι σημαντικη και λεγεται διαφ.εξ. του Duffing (Βλ.Condon - Odishaw: Handbook of Physics, σελ. 4-42).

2. Κυκλωμα με μη γραμμικη ωμικη αντισταση:

Σε εναν θυριστορα η σχεση τασεως και εντασεως του ρευματος ειναι μη γραμμικη και δινεται απο την σχεση:

$$u = R_0 i^k$$

οπου R_0, k θετικες σταθερες. Αν στην συνδεση σε σειρα ενος θυριστορα και μιας επαγωγικης αντιστασης, L, εφαρμοσουμε σταθερη ταση E προκυπτει η διαφορικη εξισωση (προβλημα αρχικης τιμης):

$$L\frac{di}{dt} + R_0 i^k = E, \quad i(0) = 0$$

Η λύση αυτού του προβλήματος μπορεί να γραφτεί στην μορφή:

$$t = L\int_0^1 \frac{dx}{E - R_0 x^k}$$

που μας δινει τον χρονο σαν συναρτηση της εντασης του ρευματος και οχι την ενταση σαν συναρτηση του χρονου. Η τελευταια αυτη σχεση μπορει εν γενει να προσδιορισθει μονο με αριθμητικες μεθοδους.

Ειδικα για k = 2 ειναι ομως:

$$i(t) = \sqrt{\frac{E}{R_0}} \tanh\left(\sqrt{\frac{ER_0}{L}} \cdot t\right)$$

3. Μη γραμμικες ταλαντωσεις

Η διαφορικη εξισωση που περιγραφει τις ταλαντωσεις σε στασιμη κατασταση (μετα την αποσβεση της διεγερσης) ενος σωματος μαζας m αναρτημενου απο γραμμικο ελατηριο με σταθερα k αλλα υπο την επιδραση *αρνητικής τριβής* ειναι

$$mx'' + kx - f(x') = 0$$

οπου η δυναμη τριβης f εχει τα χαρακτηριστικα:

$$f(-v) = -f(v), \quad f(v) > 0 \text{ για } v \in [0,v_0], \quad f(v) < 0 \text{ για } v > v_0$$

οπου v = x' η ταχυτητα. Μια τετοια δυναμη ισοδυναμει με αρνητικη τριβη στο διαστημα $[-v_0,v_0]$ και πραγματοποιειται στην φυση σαν *αεροδυναμική άνωση*. Σε ηλεκτρικα κυκλωματα ενας τετοιος ορος αντιστοιχει σε *αρνητική αντίσταση* και προκυπτει σε κυκλωματα αναδρασης.

Π.χ. με $f(v) = h(1 - v^2)v$ εχουμε την Δ.Ε. του van der Pol:

$$mx'' - h(1 - (x')^2)x' + kx = 0$$

Η εξισωση αυτη περιγραφει το ρευμα που διαρεει μιαν διοδο. Να λυθει για m = k = 1, h = 0.1 με αρχικες τιμες (a) x(0) = 0.5 v(0) = 0,και (b) x(0) = 4, v(0) = 0 (Βλ. Lax - Burstein - Lax: Calculus with Applications and Computing, vol.1, Springer, 1976, σελ.430).

4. Πρόβλημα Διώκτη - Θηράματος

Προβληματα αυτου του ειδους ανακυπτουν κατα την μελετη της επιδρασης διαφορων παραγοντων στην εξελιξη ενος πληθυσμου (π.χ. σπανιοτητα τροφης, προσβολη απο επιδημιες κοκ). Ενα απο τα απλουστερα προβληματα αυτου του ειδους ειναι το ακολουθο (Βλ. Lax - Burstein - Lax, σελ.455):

Εξεταζουμε δυο πληθυσμους που ζουν στην ιδια περιοχη: τους διωκτες και τα θηραματα. Τα θηραματα υποτιθεται οτι εχουν παντα επαρκεια τροφης, ενω οι διωκτες ζουν αποκλειστικα τρωγοντας αυτα. Πως αλληλοεπηρεαζονται τα μεγεθη των δυο πληθυσμων;

Αν δ(t) ειναι το πληθος των διωκτων κατα την στιγμη t και θ(t) το πληθος των θηραματων, τοτε οι μεταβολες των δυο πληθυσμων εξαρτωνται καθε μια απο καποιαν σταθερη διαφορα των ρυθμων γεννησεων και θανατων απο φυσικες αιτιες καθως και απο την συχνοτητα των συναντησεων ατομων των δυο πληθυσμων. Αυτην μπορουμε να την θεωρησουμε αναλογη του γινομενου

του μεγεθους των δυο πληθυσμων. Ετσι εχουμε:

$$\frac{d\theta(t)}{dt} = \alpha\theta(t) - \beta\theta(t)\delta(t), \quad \frac{d\delta(t)}{dt} = -\gamma\delta(t) + \varepsilon\theta(t)\delta(t)$$

οπου α,β,γ,ε θετικες σταθερες (Εξισωσεις Voltera - Lotka). Οι εξισώσεις αυτές μπορούν εν γένει να λυθούν μόνο αριθμητικά.

5. Κινητική Χημικών Αντιδράσεων

Εξεταζουμε στοιχειωδεις αντιδρασεις του τυπου:

$$A_2 + B_2 \rightarrow 2AB$$

Ο ρυθμος μεταβολης της πυκνοτητας x(t) του προιοντος της αντιδρασης ειναι προφανως αναλογος της πυκνοτητας a(t), b(t) των δυο επι μερους ουσιων:

$$\frac{dx}{dt} = kab$$

Αν a_0 , b_0 ειναι αντιστοιχα οι αρχικες πυκνοτητες αυτων των ουσιων, τοτε ειναι

$$a(t) = a_0 - x/2 \text{ , } b(t) = b_0 - x/2$$

γιατι καθε μοριο A και B δινει δυο μορια AB. Ετσι εχουμε:

$$\frac{dx}{dt} = k\left(a_0 - \frac{x(t)}{2}\right)\left(b_0 - \frac{x(t)}{2}\right)$$

με x(0) = 0, αν αρχικα δεν υπαρχει καθολου προιον AB (Βλ.Lax - Burstein -Lax σελ.470).

Με αναλογο τροπο προκυπτει οτι οι πυκνοτητες των προιοντων διασπασεως μιας ενωσης σε δυο απλουστερες περιγραφονται απο το συστημα διαφορικων εξισώσεων (Βλ.Lax - Burstein - Lax, σελ.474-475):

$$a' = -ka^2 + rab, \; b' = ka^2 - rab - db$$

6. *Η Σχεδίαση του Γραφήματος Πεπλεγμένης Συνάρτησης με Αναγωγή σε Πρόβλημα αρχικών Τιμών*

Πώς θα μπορούσαμε να σχεδιάσουμε το γράφημα μίας πεπλεγμένης συνάρτησης f(x,y)=0;

Μπορούμε, βέβαια, να επιλέξουμε μίαν ακολουθία τιμών του x, $x_k=x_0+kh$, k=0,1,2,… και να προσδιορίσουμε εκάστοτε μίαν λύση y_k της εξίσωσης $f(x_k,y)=0$ με κάποιαν επαναληπτική μέθοδο. Μία εναλλακτική δυνατότητα είναι, όμως η ακόλουθη:

Θεωρούμε ότι υπάρχει y=y(x) που να επαληθεύει την f(x,y)=0. Αυτό σημαίνει ότι: $f(x,y(x))\equiv 0$.

Παραγωγίζοντας αυτήν την ταυτότητα ως προς x έχουμε: $f_x+f_y y'=0$, δηλαδή την διαφορική εξίσωση: $y'=- f_x/f_y$, την οποία μπορούμε να λύσουμε αριθμητικά με αρχική συνθήκη $y(x_0)=y_0$, όπου x_0,y_0 κάποια λύση της αρχικής εξίσωσης: $f(x_0,y_0)=0$.

Για παράδειγμα, η πεπλεγμένη συνάρτηση: $y^x x^y = a$

(αν λογαριθμήσουμε και παραγωγίσουμε) μας οδηγεί στο πρόβλημα αρχικών τιμών:

$$y' = -\frac{\ln y + y/x}{\ln x + x/y}, \quad y(1) = a$$

12.1 Η Γενική Μορφή του Προβλήματος Αρχικών Τιμών

Η αναγκη αριθμητικης επιλυσης διαφορικων εξισωσεων προκυπτει απο το γεγονος οτι, οπως και στην περιπτωση αοριστων ολοκληρωματων, η λυση μιας κοινης διαφορικης εξισωσης δεν μπορει παντα να γραφτει σαν πεπερασμενη παρασταση αποτελουμενη απο συνδυασμο στοιχειωδων συναρτησεων (Για την αποδειξη του γεγονοτος αυτου βλεπε:G.N.Watson: A Treatise on the Theory of Bessel Functions, Cambridge Univ.Press, 1958, σελ.111-123). Οι κοινες διαφορικες εξισωσεις (ΚΔΕ) που μπορουν να λυθουν ετσι σε κλειστη μορφη αποτελουν μαλιστα ενα ελαχιστο ποσοστο των ΚΔΕ που συναντάμε στην Φυσικη η την Τεχνικη, ενω συχνα προτιμαται η αριθμητικη λυση, ακομα και αν υπαρχει κλειστη λυση, αν αυτη τυχαινει να ειναι πολυ περιπλοκη.

Το βασικο προβλημα που θα αντιμετωπισουμε αριθμητικα, λεγεται «προβλημα αρχικων τιμων για ΚΔΕ» και εχει την ακολουθη γενικη μορφη:

Να βρεθει συναρτηση y(x) διαφορισιμη στο διαστημα I, που να ικανοποιει εκει την διαφορικη εξισωση

$$y'(x) = f(x,y(x)) \tag{12.1}$$

και να εχει την αρχικη τιμη

$$y(x_0) = y_0 \tag{12.2}$$

οπου $x \in I$.

Η συναρτηση f(x,y) θεωρειται συνεχης σε μιαν ορθογωνια περιοχη:

$$|x - x_0| \le a, \ |y - y_0| \le b \tag{12.3}$$

οπου a, b θετικες σταθερες.

Η υπαρξη μονοσημαντα ωρισμενης λυσης του προβληματος (12.1), (12.2) ειναι εξασφαλισμενη, π.χ., αν στην περιοχη αυτην η f(x,y) ικανοποιει επι πλεον την ανισοτητα Lipschitz:

$$|f(x, y) - f(x, \hat{y})| \le L|y - \hat{y}| \tag{12.4}$$

οπου L θετικη σταθερα, η απλουστερα, αν υπαρχει η μερικη παραγωγος f_y της f ως προς y:

$$|f_y(x,y)| \le L \tag{12.5}$$

δεδομενου οτι, συμφωνα με το θεωρημα μεσης τιμης ειναι τοτε:

$$|f(x, y) - f(x, \hat{y})| = |f_y(x,\xi)||y - \hat{y}| \le L|y - \hat{y}|$$

οπου ξ μια ενδιαμεση θεση.

Υπο αυτες μαλιστα τις προυποθεσεις αποδεικνυεται οτι η λυση y(x) της (12.1) (12.2) δεν ειναι μονο διαφορισιμη, αλλα συνεχως διαφορισιμη.

12.2 Συνέπειες της Παράβασης της Συνθήκης Lipschitz

Τι συμβαίνει αν παραβούμε την συνθήκη Lipschitz;

Αυτό φαίνεται καθαρά αν προσπαθήσουμε να λύσουμε αριθμητικά το πρόβλημα:

$$y' = -x/y, \ y(-1) = 0.2$$

Το παράδειγμα αυτό δίνεται από τον Amelkin στο βιβλίο του: Differential Equations in Applications, MIR, Moscow, 1990.

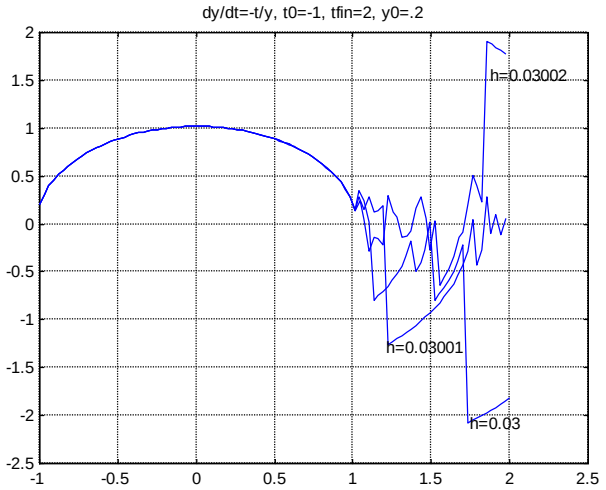
Η μέθοδος που εφαρμόζουμε είναι η μέθοδος Heun (έγινε χρήση του MATLAB script *deq.m*), για την οποία θα μιλήσουμε λίγο αργότερα. Η λύση δίνεται γραφικά, για να φανεί το πώς επηρεάζεται από διαφορετικές τιμές του βήματος h, που όμως εδώ διαφέρουν ελάχιστα μεταξύ τους.

Στο γράφημα που ακολουθεί παρατηρούμε ότι το πρώτο κομμάτι της λύσης είναι ίδιο και για τις τρεις τιμές του βήματος h, και φαινεται να ειναι ενα ημικυκλιο, που αντιστοιχεί στην ακριβή λύση της Δ.Ε.:

$$x^2+y^2=x_0^2+y_0^2=1.04$$

Στην συνεχεια ομως, λιγο μετα το σημειο x=1, παρουσιαζονται αποτομες ταλαντωσεις, που διαφερουν ριζικα μεταξυ τους αναλογα με το τι βημα h χρησιμοποιουμε.

Τι μπορει να σημαινουν αυτες οι ταλαντωσεις;



Σχήμα 12.1
Οι συνέπειες της παράβασης της συνθήκης Lipschitz

Αν εξετασουμε την συνθηκη μονοσημαντοτητας της λυσης του προβληματος αρχικων τιμων, βλεπουμε οτι με βαση την (12.5) η υπαρξη μονοσημαντης λυσης ειναι εδω εξασφαλισμενη αν ειναι φραγμενη η συνάρτηση $|f_y|= |x/y^2(x)|$. Αλλα εδω βλεπουμε στο γράφημα ότι, οσο το x πλησιαζει στο 1, τοσο η λυση y πλησιαζει στο 0, οποτε δεν αποκλειεται το f να τεινει στο απειρο.

Στην προκειμενη περιπτωση μπορουμε να διαπιστωσουμε ακριβως τι συμβαινει, γιατι η ακριβης λυση της Δ.Ε. μας ειναι γνωστη:

$$y=\sqrt{1.04-x^2}$$

και επομένως

$$f_y\left(\sqrt{1.04}\right)=f_y(1.0198..)=\infty$$

Στο σημειο x=1.0198.. παραβιαζεται η συνθηκη μονοσημαντοτητας, γιατί μηδενίζεται ο παρονομαστής της f_y. Απο το σημειο αυτο και μετα η υποτιθεμενη αριθμητικη λυση δεν εχει καμμιαν σχεση με την διαφορικη εξισωση, αφου μαλιστα η παραπανω παρασταση για το y παιρνει φανταστικες τιμες για x > 1.0198….

Εδώ βλέπουμε οτι η συνθήκη (12.5) δεν έχει μονο θεωρητικη αξια. Όταν λύνουμε αριθμητικά ένα πρόβλημα αρχικών τιμών πρέπει να ελέγχουμε κατά τον προοδευτικό ποσδιορισμό των σημείων της λύσης αν η συνθήκη αυτή εξακολουθεί να ικανοποιείται, γιατί από το σημείο που ενδεχόμενα θα παραβιασθεί και μετά δεν έχει νόημα να συνεχίσουμε την αριθμητική λύση.

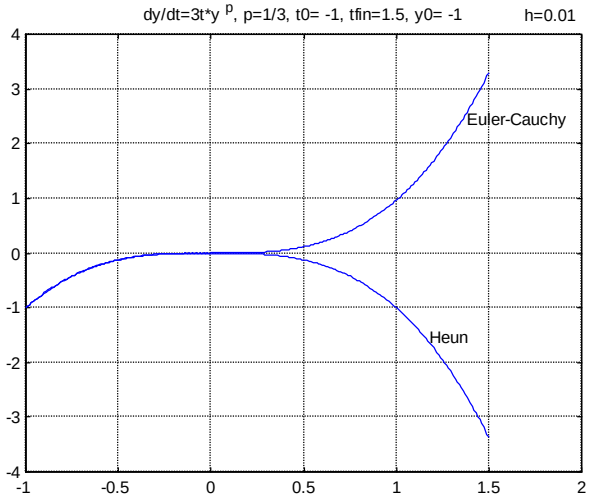
Δεύτερο Παράδειγμα

Να δοθεί γραφικά και να μελετηθεί η μορφή της λύσης του προβλήματος:

$$y' = 3x*y^{1/3},\ y(-1) = -1$$

τόσο με την μέθοδο Euler-Cauchy όσο και με την μέθοδο Heun (βλέπε το script deqf3.m).

Σχόλιο: Για $x\in[-1,0)$ η λυση αυτού του προβλήματος φαινεται να είναι μια μονοτονα αυξανομενη συνάρτηση που εχει ενα σημειο καμπης στο σημειο x=0. Απο το σημειο αυτο και μετα συνεχιζεται ομαλώτατα, σε έναν θετικό κλάδο με την μέθοδο Euler-Cauchy (βλέπε ec.m) ή έναν αρνητικό κλάδο με την μέθοδο Heun (βλέπε hn.m), δηλαδή ως μονοτόνως αύξουσα ή ως μονοτόνως πίπτουσα συνάρτηση. Έχουμε, λοιπόν για x>0 δύο εναλλακτικές λύσεις. Πώς είναι αυτό δυνατόν;



Μία ακριβής λύση του προβλήματος είναι:

$$y(x) = x^3$$

όπως μπορούμε να διαπιστώσουμε με εισαγωγή της στις παραπάνω σχέσεις. Η λύση αυτή είναι μονοτόνως αύξουσα και για x>0 και έχει θετικές τιμές. Γιατί όμως η ακριβέστερη μέθοδος Heun μας δίνει σε αυτήν την περιοχή μίαν αρνητική λύση;

Αν εξετάσουμε την $|f_y| = |x||y|^{-2/3}$ στην γειτονια του x = 0 βλέπουμε από την προσεγγιστική λύση ότι έχει απροσδιοριστη τιμη: 0/0. Δεν αποκλειεται επομενως το ενδεχομενο απειρισμου της. Τι ακριβως συμβαινει φαινεται εδω αν χρησιμοποιησουμε την λυση:

$$y(x) = x^3$$

που ικανοποιεί ακριβώς την Δ.Ε. Αυτη μας δινει $|f_y| = 1/|x| \to \infty$ για $x \to 0$. Στην προκειμενη περιπτωση, λοιπόν, στο σημείο x=0 παυει να ισχυει η συνθηκη μονοσημαντοτητας. Αυτό σημαίνει ότι μετά από αυτό το σημείο ενδέχεται να υπάρχουν πολλές διαφορετικές λύσεις. Για x>0 υπάρχουν, πράγματι, δυο διαφορετικες λυσεις, που επαληθεύουν εξ ίσου την διαφορική εξίσωση: $y(x) = x^3$ ή $y(x) = - x^3$.

12.3 Η Αναγωγή του Προβλήματος Αρχικών Τιμών για Συστήματα ΚΔΕ στην Μορφή $y'=f(x,y)$, $y(x_0)=y_0$

Στην μορφη (12.1) (12.2):

$$y'(x) = f(x,y(x)),\ \ y(x_0) = y_0,\ x \in I$$

μπορουμε να συμπεριλαβουμε και την περιπτωση ενος συστηματος διαφορικων εξισωσεων πρωτης ταξεως, αν θεωρησουμε οτι το y ειναι ενα διανυσμα: $y(x) = (y_i(x))$, i = 1(1)n, οτι $y_0 \in \mathbf{R}^n$ και οτι η f(x,y) ειναι μια διανυσματικη συναρτηση n+1 μεταβλητων[7]:

$$f(x,y) = (f_i(x,y_1,y_2,...,y_n)),\ i = 1(1)n.$$

Η υπαρξη λυσης εξασφαλιζεται στην περιπτωση αυτην υπο αναλογες με τις παραπανω προυποθεσεις. Στην θεση των απολυτων τιμων πρεπει απλα να βαλουμε τωρα μιαν οποιαδηποτε διανυσματικη νορμ και στην θεση της παραγωγου f_y τον ιακωβιανο πινακα της f ως προς τις μεταβλητες y :

$$f_y = (\partial f_i / \partial y_k),\ i,k = 1(1)n.$$

Χρησιμοποιωντας αυτες τις αντιστοιχιες μπορουμε να μεταφερουμε πολλα απο τα αποτελεσματα που θα αποδειχθουν παρακατω και σε συστηματα διαφορικων εξισωσεων.

Παράδειγμα

Οι εξισώσεις του προβλήματος «διώκτη-θηράματος» μπορούν να γραφτούν στην μορφή (12.1) αν θεωρήσουμε ότι y, $f \in \mathbf{R}^2$, και ότι:

$$y_1(t):=\theta(t),\ y_2(t):=\delta(t),\ f_1(y_1,y_2):=\alpha y_1 - \beta y_1 y_2,\ f_2(y_1,y_2):=-\gamma y_1 + \varepsilon y_1 y_2$$

12.4 Η Αναγωγή του Προβλήματος Αρχικών Τιμών για ΚΔΕ Ανωτέρας Τάξεως στην Μορφή $y'=f(x,y)$, $y(x_0)=y_0$

Διαφορικες εξισωσεις ανωτερας ταξεως μπορουν, οπως ειναι γνωστο να αναχθουν σε συστηματα Δ.Ε. πρωτης ταξεως. Π.χ. το δευτερας ταξεως προβλημα αρχικων τιμων:

---

[7] Διανυσματικά μεγέθη δεν θα τα σηματοδοτήσουμε εδώ με κάποιον ειδικό συμβολισμό, παρά μόνον αν είναι αυτό απαραίτητο για λόγους σαφήνειας. Τότε θα τα σηματοδιτήσουμε είτε με υπογράμμιση, π. χ. $\underline{y}$, είτε με έντονη γραφή, π.χ. **y**.

$$y'' = f(x,y,y'),\ y(x_0) = y_0,\ y'(x_0) = y_0' \qquad (12.6)$$

μπορει να γραφτει στην μορφη:

$$\begin{pmatrix} y_1' \\ y_2' \end{pmatrix} = \begin{pmatrix} f_1(x, y_1, y_2) \\ f_2(x, y_1, y_2) \end{pmatrix}, \quad \begin{pmatrix} y_1(x_0) \\ y_2(x_0) \end{pmatrix} = \begin{pmatrix} y_{10} \\ y_{20} \end{pmatrix} \qquad (12.7)$$

αν βαλουμε

$$y_1 := y\ ,\ y_2 := y',\ y_{10} := y_0\ ,\ y_{20} := y_0' \qquad (12.8)$$

$$f_1(x, y_1, y_2) := y_2\ ,\ f_2(x, y_1, y_2) := f(x, y_1, y_2) \qquad (12.9)$$

Μπορει κανεις επομενως να θεωρησει οτι το προβλημα αρχικων τιμων και για κοινες διαφορικες εξισωσεις ανωτερας ταξεως μπορει να αναχθει στην μορφη (12.1) και να αντιμετωπισθει με τις μεθοδους που θα περιγραψουμε παρακατω. Για ωρισμενες ομως ειδικες περιπτωσεις, με μεγαλη πρακτικη σημασια, π.χ. για την Δ.Ε. y" = f(x,y) εχουν κατασκευασθει ειδικες μεθοδοι που μπορουν να βρεθουν στην ειδικωτερη βιβλιογραφια.

Παράδειγμα μετατροπής Δ.Ε. 2ας τάξεως σε σύστημα Δ.Ε. 1ης τάξεως

Το πρόβλημα αρχικών τιμών για μία Δ.Ε. δευτέρας τάξεως:

$$y''=x+y^2-2(y')^3,\ y(0)=1, y'(0)=-1$$

μπορεί έτσι να μετατραπει στο σύστημα Δ.Ε. πρώτης τάξεως:

$$\begin{pmatrix} y_1' \\ y_2' \end{pmatrix} = \begin{pmatrix} y_2 \\ x + y_1 - 2y_2^{\ 3} \end{pmatrix}, \quad \begin{pmatrix} y_1(0) \\ y_2(0) \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

12.5 Ο Τρόπος Παραγωγής Αριθμητικών Μεθόδων

Αριθμητικη επιλυση του προβληματος (12.1) (12.2) δεν ειναι τιποτα αλλο παρα η επιλογη μιας ακολουθιας καταλληλων σημειων $x_k \in I$ (π.χ. ισαπεχοντων σημειων $x_k = a + kh$, $k = 0, 1, 2, \ldots$) και ο προσδιορισμος τιμων Y_k που να προσεγγιζουν τις τιμες $y_k := y(x_k)$ που εχει η λυση του προβληματος (12.1), (12.2) στα σημεια αυτα.

Ο προσδιορισμος των προσεγγισεων Y_k γινεται αναδρομικα. Δηλαδη για την ευρεση του Y_{k+1} προστιθεται καταλληλος διορθωτικος ορος στην τιμη Y_k (ή αλλες προγενέστερες τιμες που ηδη τις γνωριζουμε).

Ένας τρόπος προσδιορισμού του y_{n+1} με βάση το y_n είναι η χρήση του αναπτύγματος Taylor: $y_{n+1}=y(x_{n+1})=y(x_n+h)=y_n+hy'(x_n)+\ldots$
που θα τον περιγράψουμε αμέσως μετά. Επειδή όμως η μέθοδος αυτή απαιτεί την χρήση παραγώγων ως προς x της συνάρτησης f(x,y(x)) συνήθως καταφεύγουμε σε μεθόδους που δεν απαιτούν τον υπολογισμό παραγώγων αυτές βασιζονται συνήθως στον μετασχηματισμο της (12.1) στην μορφη:

$$y_{n+1} - y_n = y(x_{n+1}) - y(x_n) = \int_{x_n}^{x_{n+1}} f(t, y(t))dt \qquad (12.10)$$

που προκυπτει αν ολοκληρωσουμε την (12.1) κατα μελη απο x_n ως x_{n+1} και διαφερουν κατα τον τροπο με τον οποιο προσεγγιζεται το ολοκληρωμα.

Βασικα υπαρχουν δυο τροποι προσέγγισής του και επομενως δυο

κατηγοριες μεθοδων. Κατα τον ενα τροπο χρησιμοποιουνται οι τιμες της f(t,y(t)) σε θεσεις ενδιαμεσες μεταξυ x_n και x_{n+1}, ενώ κατά τον αλλο, το ολοκληρωμα προσεγγιζεται με χρηση σημειων που βρισκονται εξω απο το διαστημα ολοκληρωσεως, δηλαδη μερικων απο τα τελευταια σημεια x_{n-k}, για τα οποια εχουν ηδη βρεθει προσεγγισεις.

Οι μεθοδοι που προκυπτουν με τον πρωτο τροπο προσεγγισης λεγονται Μεθοδοι Runge - Kutta, ή και "Μεθοδοι ενος Βηματος", γιατι χρησιμοποιουν ουσιαστικα μονο την τελευταια προσεγγιση που εχει βρεθει. Αυτες που προκυπτουν με τον δευτερο τροπο προσεγγισης μπορουν να ονομασθουν Μεθοδοι Adams, γιατι αναπτυχθηκαν απο αυτον και τους μαθητες του, ή "Πολυβηματικες Μεθοδοι" επειδη χρησιμοποιουν περισσοτερες απο μια προγενεστερες θεσεις.

12.6 Η Μέθοδος Taylor

Όπως είπαμε, η μέθοδος αυτή θεωρεί ότι έχει ήδη προσδιορισθεί μία ικανοποιητική προσέγγιση Y_n της τιμής $y_n=y(x_n)$ και χρησιμοποιεί ένα πεπερασμένο πλήθος όρων του αναπτύγματος Taylor της y(x) γύρω από το σημείο x_n για να προσεγγίσει την τιμή $y_{n+1}=y(x_{n+1})$. Αν χρησιμοποιήσουμε για την προσέγγιση τους τρεις πρώτους όρους του αναπτύγματος έχουμε:

$$y_{n+1} = y(x_{n+1}) = y(x_n + h) = y(x_n) + hy'(x_n) + \frac{h^2}{2} y''(x_n) + \frac{h^3}{6} y^{(3)}(\xi_n),$$

με $\xi_n \in (x_n, x_{n+1})$, δηλαδή

$$y_{n+1} = y_n + hf(x_n, y_n) + \frac{h^2}{2}\left(f_x + f_y f\right)_n + O(h^3) \qquad (12.11)$$

όπου $(\ldots)_n$ υποδηλώνει ότι η εντός παρενθέσεως παράσταση λαμβάνεται στο σημείο (x_n, y_n), και ο συμβολισμός $O(h^3)$ υποδηλώνει ότι ο όρος σφάλματος έχει ουσιαστικά την τάξη μεγέθους της ποσότητας h^3. Ακριβέστερα, ισχύει ο ορισμός του Landau:

Ορισμός 12.1:

Ο συμβολισμός: $g(h)=O(h^k)$ (12.12)

σημαίνει ότι

$$\lim_{h\to 0} \frac{g(h)}{h^k} = A \qquad (12.13)$$

όπου A είναι κάποια σταθερά, δηλαδή ότι

$$g(h) \approx Ah^k \text{ για } h \approx 0 \qquad (12.14)$$

Αν αγνοήσουμε τον όρο σφάλματος, $O(h^3)$, και αντικαταστήσουμε παντού την ακριβή τιμή y_n με την προσέγγισή της, Y_n, έχουμε, λοπόν την προσέγγιση:

$$Y_{n+1} = Y_n + hf(x_n, Y_n) + \frac{h^2}{2}\left(f_x + f_y f\right)(x_n, Y_n) \qquad (12.15)$$

όπου χρησιμοποιήσαμε την συντομογραφία

$$(f_x + f_y f)(x_n,Y_n) := f_x(x_n,Y_n) + f_y(x_n,Y_n)\, f(x_n,Y_n)$$

Η σχέση (12.15) ορίζει μίαν αναδρομική ακολουθία που επιτρέπει τον υπολογισμό προσεγγίσεων της λύσης του προβλήματος (12.1),(12.2) διαδοχικά, στα σημεία x_1, x_2, x_3,....

Παράδειγμα εφαρμογής της μεθόδου Taylor

Για να προσεγγίσουμε την λύση του προβλήματος αρχικών τιμών

$$y' = x^2 + y^2,\ y(0) = 0$$

στα σημεία $x_n = x_0 + nh = nh$ μπορούμε, με βάση την (12.15), να κατασκευάσουμε την αναδρομική ακολουθία:

$$Y_{n+1} = Y_n + h(x_n^{\ 2} + Y_n^{\ 2}) + h^2\left(x_n + Y_n \cdot \left(x_n^{\ 2} + Y_n^{\ 2}\right)\right), \quad n = 1,2,3,4,....$$

όπου ελάβαμε υπόψη ότι $f(x,y) = x^2 + y^2$, $f_x = 2x$, $f_y = 2y$. Έτσι, για $h=0.01$ έχουμε: $Y_1 = Y_0 = y_0 = 0$, $Y_2 = Y_1 + hx_1^2 + h^2 x_1$ με $x_1 = h$, δηλαδή $Y_2 = 2h^3 = 0.002$, κ.ο.κ.

Το σφάλμα προσέγγισης της τιμής y_{n+1} με την την Y_{n+1} θα ήταν εδώ $O(h^3)$ αν για τον υπολογισμό της Y_{n+1} χρησιμοποιούσαμε την τιμή y_n. Επειδή όμως ξεκινούμε από την τιμή Y_n, που είναι και αυτή συνδεδεμένη με ανάλογο σφάλμα, έχουμε μίαν συνεχή συσσώρευση των σφαλμάτων. Δηλαδή σε κάθε βήμα, σε κάθε νέα θέση, προστίθεται ένα επί πλέον σφάλμα της τάξεως του h^3 στο βήμα n έχουμε, επομένως, ένα συνολικό σφάλμα τάξεως: $nh^3 = (x_n - x_0)h^2$. Είναι, δηλαδή,

$$|y_n - Y_n| = O(h^2)$$

Για τον λόγο αυτό η μέθοδος (12.15) λέγεται *μέθοδος δευτέρας τάξεως* (ως προς h). Ακριβέστερα όμως, όπως θα ιδούμε και κατά την μελέτη των επομένων μεθόδων, το ολικό σφάλμα εξαρτάται και από την ακρίβεια με την οποία ο Η/Υ εκτελεί τους υπολογισμούς και δίνεται από έναν τύπο της μορφής:

$$f_y(x_n,Y_n)|y_n - Y_n| = O(h^2) + \rho/h$$

όπου ρ η οριακή τιμή που κρατάει ο Η/Υ σε μία θέση μνήμης. Δηλαδή, ρ είναι η τιμή που μικρότερές της στρογγυλεύονται αυτομάτως σε 0, γιατί ο Η/Υ κρατάει μόνο πεπερασμένο πλήθος ψηφίων στις μνήμες του. Για παράδειγμα, αν οι πράξεις εκτελούνται με απλήν ακρίβεια, μπορεί να είναι $\rho \approx 10^{-11}$. Ο προσδιορισμός του ρ είναι απλός. Για παράδειγμα, αν ζητήσουμε την τιμή του $\sqrt{2}$, ο Η/Υ μπορεί να εμφανίσει στην οθόνη του τον αριθμό $a = 1.414213562$. Αν τώρα δώσουμε την εντολή $\sqrt{2} - a$, στην οθόνη θα εμφανισθεί ένα υπόλοιπο $3.7\ 10^{-10}$, δηλαδή τα πρόσθετα δεκαδικά ψηφία του $\sqrt{2}$, που κράτησε ο Η/Υ στην μνήμη του χωρίς να τα εμφανίσει αρχικά. Στην περίπτωση αυτή το όριο ακρίβειας εκτέλεσης των πράξεων είναι προφανώς της τάξεως του 10^{-10}.

Συνέπει της ύπαρξης του όρου ρ/h είναι ότι δεν μπορούμε να αυξήσουμε την ακρίβεια προσδιορισμού του y_n με μείωση του βήματος h. Με $h = 10^{-3}$ θα είναι:
$f_y(x_n,Y_n)|y_n - Y_n| = O(h^2) + \rho/h = O(10^{-6}) + 10^{-10}/10^{-3} = O(10^{-6})$, με $h = 10^{-4}$ θα είναι $f_y(x_n,Y_n)|y_n - Y_n| = O(10^{-8}) + 10^{-10}/10^{-4} \approx 10^{-10}/10^{-4} = 10^{-6}$, αλλά με $h = 10^{-5}$ είναι μόνο $f_y(x_n,Y_n)|y_n - Y_n| = O(10^{-10}) + 10^{-10}/10^{-5} \approx 10^{-10}/10^{-5} = 10^{-5}$.
Δηλαδή με μία μέθοδο δευτέρας τάξεως δεν έχει νόημα να πάρουμε h μικρότερο

από 10^{-4} ή 10^{-3}.

Αν θέλουμε μεγαλύτερη ακρίβεια αποτελεσμάτων τότε μία δυνατότητα είναι να ζητήσουμε από τον Η/Υ διπλάσια ακρίβεια υπολογισμών με μίαν εντολή double precision, οπότε χρησιμοποιεί για την καταγραφή ενός αριθμού δύο θέσεις μνήμης. Έτσι είναι, για παράδειγμα, $\rho \approx 10^{-22}$ και μπορούμε να χρησιμοποιήσουμε $h=10^{-8}$ για να πετύχουμε σφάλμα της τάξεως του 10^{-16}. Πολύ καλυτερο είναι όμως το να χρησιμοποιήσουμε μίαν μέθοδο τρίτης τάξεως, δηλαδή να φθάσουμε στο ανάπτυγμα Taylor μέχρι τον όρο $O(h^4)$:

$$y_{n+1} = y(x_n + h) = y(x_n) + hy'(x_n) + \frac{h^2}{2} y''(x_n) + \frac{h^3}{6} y^{(3)}(x_n) + \frac{h^4}{24} y^{(4)}(\xi_n),$$

με $\xi_n \in (x_n, x_{n+1})$, δηλαδή

$$y_{n+1} = y_n + hf(x_n, y_n) + \frac{h^2}{2}\left(f_x + f_y f\right)_n + \frac{h^3}{6}\left(f_{xx} + 2f_{xy} f + f_{yy} f^2 + f_y\left(f_x + f_y f\right)\right)_n + O(h^4)$$

και επομένως,

$$Y_{n+1} = Y_n + hf(x_n, Y_n) + \frac{h^2}{2}\left(F_x + F_y F\right)_n + \frac{h^3}{6}\left(F_{xx} + 2F_{xy} F + F_{yy} F^2 + F_y\left(F_x + F_y F\right)\right)_n$$

όπου τα κεφαλαία F, F_x, F_y,... υποδηλώνουν ότι οι συναρτήσεις αυτές υπολογίζονται με χρήση της τιμής Y_n και όχι της y_n. Δηλαδή

$$(F_x + F_y F)_n = (f_x + f_y f)(x_n, Y_n) := f_x(x_n, Y_n) + f_y(x_n, Y_n)\, f(x_n, Y_n)$$

κ.ο.κ.

Εδώ θα είναι

$$f_y(x_n, Y_n)|y_n - Y_n| = O(h^3) + \rho/h$$

και για $\rho \approx 10^{-10}$ με $h=10^{-3}$ έχουμε

$$f_y(x_n, Y_n)|y_n - Y_n| = O(h^3) + \rho/h = O(10^{-9}) + 10^{-7} \approx 10^{-7}$$

δηλαδή μεγαλύτερη ακρίβεια αποτελεσμάτων από εκείνην της μεθόδου δευτέρας τάξεως, παρά το ότι χρησιμοποιούμε μεγαλύτερο βήμα.

Στην τελευταία αναδρομική σχέση βλέπουμε όμως το μεγάλο μειονέκτημα της μεθόδου Taylor: την ανάγκη υπολογισμού πλήθους παραγώγων της f(x,y). Αυτός είναι ο λόγος για τον οποίο χρησιμοποιείται σπάνια και προτιμώνται μέθοδοι όπως οι επόμενες.

12.7 Μέθοδοι Runge-Kutta

Ένας απλός τρόπος προσέγγισης του ολοκληρώματος στην (12.10) είναι η εφαρμογή του κανόνα του μέσου σημείου. Με την βοηθεια του εχουμε:

$$y_{n+1} = y_n + \int_{x_n}^{x_{n+1}} f(t, y(t))dt = y_n + hf\left(x_n + \frac{h}{2}, y\left(x_n + \frac{h}{2}\right)\right) + \frac{h^3}{24} y^{(3)}(\xi_n)$$

οπου ο τελευταιος ορος ειναι ο ορος σφαλματος. Εδω βλεπουμε ποιαν δυσκολια αντιμετωπιζουν οι μεθοδοι που χρησιμοποιουν εσωτερικα σημεια. Η

τιμη $y(x_n + h/2)$ δεν μας ειναι γνωστη, ουτε κατα προσεγγιση. Για να την προσεγγισουμε μπορουμε όμως να χρησιμοποιησουμε το αναπτυγμα Taylor:

$$y\left(x_n + \frac{h}{2}\right) = y(x_n) + \frac{h}{2}y'(x_n) + \frac{h^2}{8}y''(x_n)$$

Ετσι προκυπτει

$$\begin{aligned} y_{n+1} &= y_n + hf\left(x_n + \frac{h}{2}, y_n + \frac{h}{2}y'(x_n) + O(h^2)\right) + O(h^3) = \\ &= y_n + hf\left(x_n + \frac{h}{2}, y_n + \frac{h}{2}f(x_n, y_n)\right) + hO(h^2)f_y(\ldots) + O(h^3) \end{aligned} \qquad (12.11)$$

οπου η τελευταια ισοτητα προκυπτει παλι με αναπτυγμα Taylor της f ως προς την δευτερη της μεταβλητη γυρω απο την θεση $y_n+(h/2)f(x_n, y_n)$.

Η τελικη μορφη της μεθοδου αυτης ειναι επομενως:

$$Y_{n+1} = Y_n + hf\left(x_n + \frac{h}{2}, Y_n + \frac{h}{2}f(x_n, Y_n)\right), n = 1,2,3,\ldots \qquad (12.12)$$

οπου $Y_0=y_0$. Χαρακτηριστικο της ειναι, οπως βλεπουμε, η αντικατασταση της f μεσα στον εαυτο της. Αυτο ειναι και το γενικο χαρακτηριστικο ολων των μεθοδων Runge - Kutta.

Το σφαλμα που γινεται κατα την χρηση της (12.12) ειναι φαινομενικα, οπως φαινεται απο την (12.11) της μορφης $O(h^3)$. Δεν πρεπει ομως να ξεχναμε οτι αυτο το σφαλμα προκυπτει κατα την μεταβαση απο την θεση x_n στην x_{n+1}. Επειδη το Y_{n+1} προκυπτει απο το Y_n με προσθηκη ενος διορθωτικου ορου, ουσιαστικα τα σφαλματα αυτα αθροιζονται, οποτε η ταξη μεγεθους του σφαλματος $Y_n - y_n$ ειναι

$$n\,O(h^3) = (x_n - x_0)O(h^2)$$

δηλαδη, τελικα $O(h^2)$. Για τον λογο αυτο η μεθοδος (12.12) λεγεται μεθοδος Runge - Kutta δευτερας ταξεως (ως προς h).

Ο τροπος παραγωγης της μεθοδου (12.12) βασιζεται, οπως ειδαμε στην χρηση ενδιαμεσων θεσεων μεταξυ x_n και x_{n+1} και στην επανειλημμενη χρηση αναπτυγματων Taylor. Για τουτο, για την παραγωγη ολων των μεθοδων αυτου του ειδους, αντι να χρησιμοποιουμε καθε φορα διαφορετικο τυπο αριθμητικης ολοκληρωσης και στην συνεχεια αναπτυγματα Taylor, γραφουμε απ'ευθειας

$$y(x_n + h) = y_n + c_1 k_{1n} + c_2 k_{2n} + \ldots + c_m k_{mn} + O(h^{p+1}) \qquad (12.13)$$

με

$$\begin{aligned} k_{1n} &= hf(x_n, y_n), \\ k_{2n} &= hf(x_n + a_2 h, y_n + b_2 k_{1n}), \\ k_{3n} &= hf(x_n + a_3 h, y_n + b_{31}k_{1n} + b_{32}k_{2n}), \ldots \\ k_{mn} &= hf(x_n + a_m h, y_n + b_{m1}k_{1n} + b_{m2}k_{2n} + \ldots + b_{m,n-1}k_{m-1,n}) \end{aligned} \qquad (12.14)$$

Για τον προσδιορισμο των συντελεστων a_i, b_{ik} κανουμε τα αναπτυγματα Taylor, τοσο της $y(x_n + h)$ οσο και των k_{in} και εξισωνουμε τους συντελεστες ιδιων δυναμεων του h, ωστε η διαφορα πρωτου και δευτερου μερους της (15.13)

να εμφανιζεται σε οσο το δυνατον μεγαλυτερη δυναμη του h. Έτσι υποκαθιστούμε τους πολλαπλούς υπολογισμούς παραγώγων, που θα είχαμε σε μίαν μέθοδο Taylor ανωτέρας τάξεως με πολλαπλές χρήσεις της συνάρτησης f(x,y).

Η αντιστοιχη προς την (12.13) μεθοδος προκυπτει τώρα με παραλειψη του ορου σφαλματος και αντικατασταση του y_n, οπου εμφανιζεται, με την προσεγγιση του, Y_n. Το ολικο σφαλμα $y_n - Y_n$ ειναι, οπως και στη περιπτωση της (12.12), $O(h^p)$, δηλαδη ταξεως p και οχι p+1.

Εν γενει προκυπτουν κατ' αυτον τον τροπο περισσοτερες απο μια μεθοδοι ιδιας ταξεως p. Η συνηθεστερη μεθοδος ταξεως 4, η λεγομενη "Κλασσικη Μεθοδος Runge - Kutta", ειναι η ακολουθη:

$$Y_{n+1} = Y_n + \frac{1}{6}\left(K_{1n} + 2K_{2n} + 2K_{3n} + K_{4n}\right) \tag{12.15}$$

$$\begin{aligned} K_{1n} &= hf(x_n, Y_n) \\ K_{2n} &= hf\left(x_n + \frac{h}{2}, Y_n + \frac{K_{1n}}{2}\right) \\ K_{3n} &= hf\left(x_n + \frac{h}{2}, Y_n + \frac{K_{2n}}{2}\right) \\ K_{1n} &= hf(x_n + h, Y_n + K_{3n}) \end{aligned} \tag{12.16}$$

Μια αλλη ενδιαφερουσα μεθοδος Runge - Kutta τεταρτης ταξεως ειναι η ακολουθη, η οποία φαίνεται να χρησιμοποιείται και από το MATLAB με την εντολή ode45(…):

$$\begin{aligned} K_{1n} &= hf(x_n, Y_n) \\ K_{2n} &= hf\left(x_n + \frac{h}{2}, Y_n + \frac{K_{1n}}{2}\right) \\ K_{3n} &= hf\left(x_n + \frac{h}{2}, Y_n + \frac{K_{1n}}{4} + \frac{K_{2n}}{4}\right) \\ K_{4n} &= hf(x_n + h, Y_n - K_{2n} + 2K_{3n}) \end{aligned} \tag{12.17}$$

$$Y_{n+1} = Y_n + \frac{1}{6}\left(K_{1n} + 4K_{3n} + K_{4n}\right) \tag{12.18}$$

Αν συμπληρωσουμε τους παραπανω διορθωτικους ορους (12.17) με δυο ακομα ορους:

$$\begin{aligned} K_{5n} &= hf\left(x_n + \frac{2}{3}h, Y_n + \frac{7}{27}K_{1n} + \frac{10}{27}K_{2n} + \frac{1}{27}K_{4n}\right) \\ K_{6n} &= hf\left(x_n + \frac{1}{5}h, Y_n + \frac{28}{625}K_{1n} - \frac{1}{5}K_{2n} + \frac{546}{625}K_{3n} + \frac{54}{625}K_{4n} - \frac{378}{625}K_{5n}\right) \end{aligned}$$

(12.19)

τοτε με

$$\widetilde{Y}_{n+1} = Y_n + \frac{1}{24}K_{1n} + \frac{5}{48}K_{4n} + \frac{27}{56}K_{5n} + \frac{125}{336}K_{6n} \qquad (12.20)$$

εχουμε μιαν μεθοδο πεμτης ταξεως.

Σημαντικο ειναι στην προκειμενη περιπτωση, οτι το σφαλμα $y_{n+1} - Y_{n+1}$ της μεθοδου τεταρτης ταξεως, δηλαδη της (12.17), (12.18) μπορει να εκτιμηθει με την βοηθεια του υπολογισμου μονο δυο προσθετων ορων (12.19). Αν, δηλαδη ονομασουμε E_{n+1} και $\widetilde{E}_{n+1}$ τα ολικα σφαλματα των δυο μεθοδων στην θεση x_{n+1}:

$$E_{n+1} := Y_{n+1} - y_{n+1}, \quad \widetilde{E}_{n+1} := \widetilde{Y}_{n+1} - y_{n+1} \qquad (12.21)$$

Τοτε αφαιρωντας κατα μελη αυτες τις δυο σχεσεις εχουμε:

$$Y_{n+1} - \widetilde{Y}_{n+1} = E_{n+1} - \widetilde{E}_{n+1} \approx E_{n+1} \qquad (12.22)$$

οπου η τελευταια σχεση οφειλεται στο οτι $E_{n+1}=O(h^4)$ ενω $\widetilde{E}_{n+1} = O(h^5)$, οποτε ειναι αμελητεο σε σχεση με το E_{n+1}.

Αφαιρωντας κατα μελη την (12.20) απο την (12.18) εχουμε επομενως για το σφαλμα E_{n+1} (Βλ.Ralston -Rabinowitz σελ. 218-220):

$$y_{n+1} - Y_{n+1} = E_{n+1} \approx \frac{1}{8}K_{1n} + \frac{2}{3}K_{2n} + \frac{1}{16}K_{4n} - \frac{27}{56}K_{5n} - \frac{125}{336}K_{6n} \qquad (12.23)$$

Κατα την χρηση μιας τετοιας μεθοδου μπορει να γινει σημαντικη οικονομια χρονου με το να μεταβαλλουμε το βημα h αναλογα με το ποσο γρηγορα μεταβαλλονται οι τιμες Y_n που προκυπτουν. Αν οι τιμες αυτες μεταβαλλονται πολυ αργα, διπλασιαζουμε π.χ. το βημα, ενω αν μεταβαλλονται γρηγορα, το μειωνουμε στο μισο (Για λεπτομερειες, βλ. π.χ. Ralston -Rabinowitz σελ. 220-221).

Παράδειγμα Εφαρμογής της Μεθόδου Runge-Kutta

Να λυθει με την κλασσικη μεθοδο Runge - Kutta το προβλημα αρχικων τιμων:

$$y' = -2xy^2 \ , \ y(0) = 1$$

στο διαστημα [0, 1] με βημα h = 0.2.

Λύση

$x_0 = 0$, $Y_0 = 1$

$x_1 = 0.2$:

$K_{10} = hf(x_0 , Y_0) = -2x_0Y_0 = 0$

$K_{20} = h\, f(x_0 + h/2, Y_0 + K_{10}/2) = -2h(x_0 + h/2)(Y_0 + K_{10}/2)^2 = -0.04$

$K_{30} = h\, f(x_0 + h/2, Y_0 + K_{20}/2) = -2h(x_0 + h/2)(Y_0 + K_{20}/2)^2 = -0.038416$

$K_{40} = h\, f(x_0 + h, Y_0 + K_{30}) = -2h(x_0 + h)(Y_0 + K_{30})^2 = -0.0739715$

$y(0.2) \approx Y_1 = Y_0 + (1/6)(K_{10} + 2K_{20} + 2K_{30} + K_{40}) = 0.9615328$

$x_1 = 0.4$:

$K_{11} = hf(x_1, Y_1) = -2h\, x_1 Y_1^2 = -0.0739636$

$K_{21} = h\, f(x_1 + h/2, Y_1 + K_{11}/2) = -2h(x_1 + h/2)(Y_1 + K_{11}/2)^2 = -0.1025754$

$K_{31} = h\, f(x_1 + h/2, Y_1 + K_{21}/2) = -2h(x_0 + h/2)(Y_1 + K_{21}/2)^2 = -0.0994255$

$K_{41} = h\, f(x_0 + h, Y_0 + K_{31}) \quad = -2h(x_0 + h)(Y_0 + K_{31})^2 \quad = -0.1189166$

$y(0.4) \approx Y_2 = Y_1 + (1/6)(K_{11} + 2K_{21} + 2K_{31} + K_{41}) \quad = 0.8620525$

Παρομοια προκυπτει:

$y(0.6) \approx Y_3 = 0.7352784$

$y(0.8) \approx Y_4 = 0.6097519$

$y(1.0) \approx Y_5 = 0.5000073$

Για διαπιστωση της ακριβειας αυτης της μεθοδου μπορει κανεις να συγκρινει αυτες τις τιμες με εκεινες της ακριβους λυσης που είναι: $y(x) = \dfrac{1}{x^2+1}$.

Παράδειγμα Λύσης Συστήματος ΚΔΕ με την Μέθοδο Runge-Kutta-O(h^2)

Να λυθει με την μεθοδο Runge - Kutta - $O(h^2)$ το συστημα των Δ.Ε. Voltera - Lotka:

$$d\theta/dt = \theta(\alpha - \beta\delta),\ d\delta/dt = \delta(-\gamma + \varepsilon\theta)$$

με $\alpha = 1$, $\beta = 1$, $\gamma = 3$, $\delta = 2$, $\theta_0 = \theta(0) = 0.8$, $\delta_0 = \delta(0) = 0.2$

Λύση

Με $\theta := y_1$, $\delta := y_2$ και $t := x$ το συστημα παιρνει την μορφη:

$$y_1' = y_1(1 - y_2) := f_1(y_1, y_2)$$

$$y_2' = y_2(-3 + 2y_1) := f_2(y_1, y_2)$$

Θεωρωντας ολα τα μεγεθη διανυσματικα:

$$y := (y_1, y_2)^T \approx (Y_1, Y_2)^T := Y$$

$$f(y) := (f_1(y_1, y_2), f_2(y_1, y_2))^T$$

εχουμε με βαση την (12.12):

$$Y_{n+1} = Y_n + hf\left(x_n + \frac{h}{2}, Y_n + \frac{h}{2} f(x_n, Y_n)\right) = Y_n + hf\left(Y_n + \frac{h}{2} f(Y_n)\right)$$

Για να απλοποιησουμε τις παραστασεις που ακολουθουν θα παραλειψουμε τους δεικτες βαζοντας $Y_{n+1} := \hat{Y}$, $Y_{n+1} := Y$ και θα βαλουμε $h = 2\hat{h}$. Επισης χρησιμοποιουμε βοηθητικα τις παραστασεις:

$$Z_1 := Y_1 + \hat{h}\, F_1(Y) = Y_1 + \hat{h}\, Y_1(1 - Y_2)$$

$$Z_2 := Y_2 + \hat{h}\, F_2(Y) = Y_2 + \hat{h} Y_2(-3 + 2Y_1)$$

Ετσι εχουμε:

$$\hat{Y}_1 = Y_1 + 2\hat{h} Z_1(1 - Z_2)$$

$$\hat{Y}_2 = Y_2 + 2\hat{h} Z_2(-3 + 2Z_1)$$

Χρησιμοποιωντας αυτην την σχεση επαναληπτικα, π.χ. με $\hat{h} = 0.05$,($h = 0.1$), μπορουμε να κατασκευασουμε μιαν ακολουθια προσεγγισεων της λυσης στα σημεια $t_n = hn = 0.1n$.

12.8 Η Γενική Μορφή του Ολικού Σφάλματος για Αριθμητικές Μεθόδους Λύσης του Προβλήματος Αρχικών Τιμών

Όπως φάνηκε ήδη κατά την μελέτη της μεθόδου Taylor, το ολικό σφάλμα προσέγγισης για μία μέθοδο τάξης k έχει γενικά την μορφή:

$$f_y(x_n, Y_n)|y_n - Y_n| = O(h^k) + \rho/h$$

αν ικανοποιείται η λεγόμενη «συνθήκη ευστάθειας της μεθόδου», για την οποία θα μιλήσουμε σε λίγο. Εδώ είναι ρ η ακρίβεια με την οποία διατηρεί ο Η/Υ τα αποτελέσματα στις μνήμες του. Αυτό ισχύει για όλες τις μεθόδους , ασχέτως του πως παράγονται.

Ένας τέτοιος τύπος είναι σημαντικός, γιατί μας επιτρέπει να προσδιορίσουμε το κατάλληλο βήμα h, για να έχουμε την μεγαλύτερη δυνατή ακρίβεια προσέγγισης. Όπως είδαμε, δεν μπορούμε να επιλέξουμε το h οσοδήποτε μικρό θέλουμε για να επιτύχουμε μεγαλύτερη ακρίβεια προσέγγισης, αλλά υπάρχει μία τάξη μεγέθους του h κάτω από την οποία το ολικό σφάλμα αρχίζει να διογκώνεται πάλι λόγω των σφαλμάτων στρογγυλεύσεως.

Είναι, λοιπόν σημαντικό να δούμε πώς ακριβώς προκύπτει ένας τέτοιος τύπος, δηλαδή πώς εξαρτάται το ολικό σφάλμα προσέγγισης από τις παραμέτρους του προβληματος. Εδώ θα κάνουμε μίαν τέτοια ανάλυση σφαλμάτων ενδεικτικά για την μέθοδο Runge-Kutta δευτέρας τάξεως.

Με βάση τις (12.11),(12.12) ισχύει:

$$y_{n+1} = y_n + hf\left(x_n + \frac{h}{2}, y_n + \frac{h}{2} f(x_n, y_n)\right) + A_n h^3 \qquad (12.24)$$

$$Y_{n+1} = Y_n + hf\left(x_n + \frac{h}{2}, Y_n + \frac{h}{2} f(x_n, Y_n)\right) - \rho \qquad (12.25)$$

όπου A_n ο συντελεστής του σφάλματος προσέγγισης στην θέση x_n. Εδώ θα τον θεωρήσουμε ανεξάρτητο του n, ίσο με μία σταθερή τιμή A, για να απλοποιήσουμε την μελέτη της συσσώρευσης των σφαλμάτων. Επίσης στον τύπο που παράγει τις προσεγγίσεις προσθέσαμε έναν όρο σφάλματος στρογγύλευσης, ρ (έστω $\rho=\pm 10^{-10}$), δεδομένου ότι οι πράξεις εκτελούνται με πεπερασμένη ακρίβεια.

Αφαιρώντας κατά μέλη αυτές τις ισότητες και κάνοντας διπλή χρήση του θεωρήματος μέσης τιμής του διαφορικού λογισμού:

$$f(x,y)-f(x,Y)=(y-Y)f_y(x,\xi)$$

έχουμε:

$$E_{n+1}=E_n+hf_y(x_n+h/2,\ldots)\{E_n+(h/2)f_y(x_n,Y_n)E_n\}+Ah^3+\rho$$

όπου

$$E_k:=y_k - Y_k \qquad (12.26)$$

Εδώ θα κάνουμε μίαν ακόμη απλοποιητική παραδοχή, ότι η f_y έχει σταθερή ή σχεδόν σταθερή τιμή, έστω $f_y=\lambda$, πράγμα που σημαίνει ότι η ανάλυσή μας ισχύει αυστηρά μόνο για προβλήματα με f(x,y) της μορφής λy+μ(x), ή και $\lambda(y-y_0)+\mu(x)$, δηλαδή προβλήματα της μορφής:

$$y'=\lambda(y-y_0)+\mu(x),\ y(x_0)=y_0. \qquad (12.27)$$

Τότε έχουμε

$$E_{n+1}=E_n\{1+hf_y+(hf_y)^2/2\}+Ah^3+\rho= E_n\{1+h\lambda+(h\lambda)^2/2\}+Ah^3+\rho$$

δηλαδή

$$E_{n+1}=\alpha E_n+\beta,\ n=1,2,3,\ldots \qquad (12.28)$$

με

$$\alpha:=1+h\lambda+(h\lambda)^2/2,\ \ \beta:=Ah^3+\rho \qquad (12.29)$$

Από αυτήν την αναδρομική σχέση προκύπτει:
$E_2=\alpha E_1+\beta$, $E_3=\alpha E_2+\beta=\alpha^2E_1+\alpha\beta+\beta$, $E_4=\alpha E_3+\beta=\alpha^3E_1+\alpha^2\beta+\alpha\beta+\beta$,
και γενικά,

$$E_{n+1}=\alpha^nE_1+\beta(1+\alpha+\alpha^2+\ldots+\alpha^{n-1})$$

δηλαδή

$$E_{n+1}=\alpha^nE_1+\beta\frac{1-\alpha^n}{1-\alpha},\quad n=1,2,3,\ldots \tag{12.30}$$

Η σχέση αυτή μας δείχνει ότι το σφάλμα E_n τείνει για $n\to\infty$ είτε προς το άπειρο, αν είναι $|\alpha|>1$, είτε να σταθεροποιηθεί στην τιμή

$$E_\infty=\frac{\beta}{1-\alpha} \tag{12.31}$$

αν είναι $|\alpha|<1$.

Το ολικό σφάλμα είναι λοιπόν φραγμένο αν ισχύει:

$$|\alpha|=\left|1+h\lambda+\frac{(h\lambda)^2}{2}\right|<1 \tag{12.32}$$

δηλαδή αν ισχύει η διπλή ανισότητα

$$-1<1+h\lambda+\frac{(h\lambda)^2}{2}<1.$$

Η πρώτη από αυτές τις ανισότητες ισχύει πάντα, ενώ η δεύτερη ισοδυναμεί με $h\lambda(h\lambda+2)<0$, δηλαδή

$$-2<h\lambda<0 \tag{12.33}$$

Αν ισχύει αυτή η συνθήκη τότε είναι

$$E_\infty=\frac{\beta}{1-\alpha}=-\frac{Ah^3+\rho}{h\lambda+\frac{(h\lambda)^2}{2}}\approx-\frac{1}{\lambda}\left(Ah^2+\frac{\rho}{h}\right) \tag{12.34}$$

όπου ο τετραγωνικός όρος στον παρονομαστή θεωρήθηκε αμελητέος.

Τι σημαίνει η συνθήκη (12.33);

Δεδομένου ότι $h>0$, κατ' αρχήν μας λέει ότι για να έχουμε φραγμένο σφάλμα πρέπει να είναι $\lambda=f_y<0$. Κατ' αρχήν πρέπει να τονισθεί ότι για κάθε μέθοδο επίλυσης Δ.Ε. μπορεί να αποδειχθεί μία ανάλογη συνθήκη, δηλαδή μία συνθήκη του τύπου:

$$-\gamma<h\lambda<0 \tag{12.35}$$

όπου η αρνητική σταθερά γ εξαρτάται από την μέθοδο που εξετάζουμε[8]. Για παράδειγμα, για την μέθοδο (12.16), δηλαδή την κλασσική μέθοδο Runge-Kutta-$O(h^4)$, είναι $\gamma=2.785$.

[8] Αν η μέθοδος εφαρμόζεται σε ένα σύστημα διαφορικών εξισώσεων της μορφής $y'=f(x,y)$ όπου οι συναρτήσεις y και f είναι διανυσματικές, τότε μπορούμε να κάνουμε μίαν παρόμοια ανάλυση σφαλμάτων, μόνο που $f_y(x,y)$ είναι τώρα ο πίνακας των πρώτων παραγώγων του $f(x,y)$ ως προς τις συνιστώσες του y, δηλαδή ο ιακωβιανός πίνακας. Στην θέση του λ

Μπαίνει, λοιπόν, άμεσα το ερώτημα: όλες αυτές οι μέθοδοι μπορούν να εφαρμοσθούν μόνο σε Δ.Ε. με $f_y<0$; Τι θα κάνουμε αν $f_y>0$;

Εδώ πρέπει να προσέξουμε ότι το πρόβλημα (12.27), που μελετούμε αντιπροσωπευτικά αντί για το γενικό πρόβλημα (12.1)-(12.2), έχει την λύση:

$$y(x) = y_0 + \int_{x_0}^{x} e^{\lambda(x-t)}\mu(t)dt \qquad (12.36)$$

από την οποία προκύπτει με την βοήθεια του δεύτερου θεωρήματος μέσης τιμής του ολοκληρωτικού λογισμού:

$$y(x) - y_0 = \mu(\xi)\int_{x_0}^{x} e^{\lambda(x-t)}dt = \mu(\xi)\frac{e^{\lambda(x-x_0)}-1}{\lambda}, \quad \mu\varepsilon \quad \xi \in (x_0, x) \qquad (12.37)$$

Προφανώς, αν είναι λ>0, λύση αυτή τείνει με αυξανόμενα x προς το άπειρο. Αυτό σημαίνει ότι στην περίπτωση αυτή δεν μας απασχολεί και το αν το σφάλμα μίας αριθμητικής λύσης αυξάνεται συνεχώς, υπό τον όρο ότι αποτελεί μικρό ποσοστό της ακριβούς λύσης. Το ίδιο ισχύει και στην περίπτωση (12.1)-(12.2) με $f_y>0$. Δηλαδή, σε μίαν τέτοια περίπτωση ενδιαφέρει να είναι φραγμένο το σχετικό σφάλμα: $|y_n-Y_n|/|Y_n|$. Αυτό είναι εν γένει πιο εύκολο να εξασφαλισθεί. Γι' αυτό επικεντρώνουμε εδώ το ενδιαφέρον μας στην περίπτωση λ<0, που αντιπροσωπεύει φθίνουσες λύσεις. Τέτοιες λύσεις έχουν και τα περισσότερα τεχνικά ή φυσικά προβλήματα, δεδομένου ότι τα περισσότερα φαινόμενα στην φύση έχουν την τάση για απόσβεση με την πάροδο του χρόνου, παρά για αύξηση.

Η (12.33) μας λέει ότι για την μέθοδο Runge-Kutta-$O(h^2)$ θα έχουμε ένα σφάλμα φραγμένο του τύπου (12.34) αν είναι:

$$-\frac{2}{\lambda} = \left|\frac{2}{\lambda}\right| > h > 0 \qquad (12.38)$$

δηλαδή,

$$-\frac{2}{f_y(x_n, Y_n)} = \left|\frac{2}{f_y(x_n, Y_n)}\right| > h > 0 \qquad (12.39)$$

Αν, λοιπόν, η $f_y(x_n,Y_n)$ δεν έχει πολύ μεγάλες τιμές μπορούμε εν γένει να επιλέξουμε το βήμα h έτσι ώστε να ικανοποιεί την (12.39). Συνήθως η συνθήκη αυτή ικανοποιείται άνετα για h τάξεως 10^{-3} ή 10^{-4}, που χρησιμοποιούμε για να είναι το ολικό σφάλμα (12.34) μικρό.

Εδώ προκύπτει, βέβαια, το ερώτημα κατά πόσο μία τέτοια ανάλυση έχει αξία για προβλήματα με $f_y(x_n,Y_n)$ που δεν είναι περίπου σταθερό. Αν η παράσταση αυτή μεταβάλλεται σχετικά αργά, μπορούμε να θεωρήσουμε ότι το πρόβλημα (12.27) περιγράφει την συμπεριφορά του (12.1)-(12.2).

Με την βοήθεια του θεωρήματος μέσης τιμής $g(y)-g(y_0)=(y-y_0)g_y(\xi)$ προκύπτει, ειδικότερα:

μπαίνουν τότε οι ιδιοτιμές αυτού του πίνακα, που ενδέχεται να είναι και μιγαδικές. η συνθήκη (12.35) γίνεται τότε:

$$-\gamma<\mathrm{Re}\{h\lambda\}<0$$

$$y'(x)=f(x,y)=f(x,y_0)+f(x,y)-f(x,y_0)=f(x,y_0)+(y-y_0)f_y(x,\xi(x)) \qquad (12.40)$$

δηλαδή μία μορφή ανάλογη προς την (12.27) με

$$\mu(x)=f(x,y_0), \text{ και } \lambda= f_y(x,\xi(x)) \qquad (12.41)$$

Αν, λοιπόν, η $f_y(x_n,Y_n)$ μεταβάλλεται αργά, μπορούμε να θεωρήσουμε ότι η (12.27) με λ σταθερό περιγράφει την συμπεριφορά της Δ.Ε. (12.1) αρκετά καλά και ότι η παραπάνω ανάλυση σφαλμάτων μπορεί να εφαρμοσθεί στο πρόβλημα (12.1)-(12.2). Αν πάλι η $f_y(x_n,Y_n)$ είναι έντονα μεταβαλλόμενη, αντιστοιχεί και σε έντονες μεταβολές της λύσης. Τότε πρέπει σε κάθε περίπτωση να εξετάσουμε και το κατά πόσον το h πρέπει να επιλεχθεί υπερβολικά μικρό για να ικανοποιεί την (12.39).

Σε κάθε περίπτωση μπορούμε να περικλείσουμε την λύση του προβλήματος (12.1)-(12.2) με βάση το ακόλουθο θεώρημα:

Θεώρημα 12.1

Αν είναι

$$\lambda_m < f_y(x,\xi(x))< \lambda_M \qquad (12.42)$$

τότε ισχύει ότι

$$y_m(x) \le y(x) \le y_M(x) \qquad (12.43)$$

ή

$$y_m(x) \ge y(x) \ge y_M(x) \qquad (12.43)$$

όπου $y_m(x)$, $y_M(x)$ είναι οι λύσεις της (12.27) με $\mu(x)=f(x,y_0)$ και αντίστοιχα $\lambda=\lambda_m$ και $\lambda=\lambda_M$, που δίνονται από την (12.36). Δηλαδή, η y(x) βρίσκεται πάντα μεταξύ $y_m(x)$ και $y_M(x)$.

Απόδειξη: Με βάση την (12.40) ισχύει είτε

$$\lambda_m(y-y_0)+\mu(x)<y'(x)<\lambda_M(y-y_0)+\mu(x) \qquad (12.44)$$

είτε

$$\lambda_m(y-y_0)+\mu(x)>y'(x)>\lambda_M(y-y_0)+\mu(x) \qquad (12.45)$$

αναλόγως του τι πρόσημο έχει η $y-y_0$. Έστω ότι ισχύει η (12.44). Αν μετονομάσουμε την μεταβλητή x σε t το δεξιό μέρος της (12.44):

$$y'(t)<\lambda_M(y(t)-y_0)+\mu(t)$$

ισοδυναμεί προς

$$y'(t)-\lambda_M(y(t)-y_0)<\mu(t)$$

ή

$$\left(e^{-\lambda_M t}(y-y_0)\right)' = e^{-\lambda_M t}\left(y'-\lambda_M(y-y_0)\right)< e^{-\lambda_M t}\mu(t)$$

ολοκληρώνοντας την σχέση αυτή από x_0 ως x έχουμε αμέσως:

$$e^{-\lambda_M x}(y-y_0) < \int_{x_0}^{x} e^{-\lambda_M t}\mu(t)dt$$

ή

$$y(x) < y_0 + \int_{x_0}^{x} e^{\lambda_M (x-t)} \mu(t)dt = y_M(x)$$

Αναλόγως αποδεικνύονται και οι υπόλοιπες ανισότητες.

Το αποτέλεσμα αυτό μας οδηγεί στο συμπέρασμα ότι, για να διαπιστώσουμε την επίδραση της συσσώρευσης των σφαλμάτων στην ακρίβεια της προσέγγισης, αρκεί να εξετάσουμε αντιπροσωπευτικά την περίπτωση εφαρμογής της εκάστοτε μεθόδου στην περίπτωση $y'=\lambda y$, δηλαδή $f(x,y)=\lambda y$. Η απάιτηση να είναι φραγμένο το ολικό σφάλμα αυτό οδηγεί για κάθε μέθοδο σε μίαν ανισότητα παρόμοια με την (12.32) που είναι ισοδύναμη με μίαν ανισότητα του τύπου (12.35). Στην περίπτωση της Κλασσικής Μεθόδου Runge-Kutta αυτό οδηγεί σε μίαν σχέση του τύπου (12.28) με

$$\alpha(h\lambda):=1+h\lambda+(h\lambda)^2/2 +(h\lambda)^3/6+(h\lambda)^4/24, \text{ και } \beta:=Ah^5+\rho$$

και η σχέση $|\alpha(h\lambda)|<1$ ισοδυναμεί με

$$h\lambda\{1+(h\lambda)/2 +(h\lambda)^2/6+(h\lambda)^3/24\}<0$$

δηλαδή, $\lambda<0$ και

$$P(h\lambda):= 1+(h\lambda)/2 +(h\lambda)^2/6+(h\lambda)^3/24>0$$

Το πολυώνυμο $P(h\lambda)$ έχει μόνο μίαν πραγματική ρίζα, την $h\lambda=-2.78529356\ldots$. Άρα παραμένει θετικό για $h\lambda\in(-2.785, 0)$.

12.9 Η Μέθοδος Heun (Χόιν)

Η απλουστερη μεθοδος πολλων βηματων, δηλαδή μέθοδος που για τον προσδιορισμό της Y_{n+1} χρησιμοποιεί όχι μόνο την Y_n αλλά και άλλες τιμές, προκυπτει με εφαρμογη του κανονα του τραπεζιου για να προσεγγισθει το ολοκληρωμα στο δευτερο μερος της (12.10). Ετσι προκυπτει:

$$y_{n+1} = y_n + \int_{x_n}^{x_{n+1}} f(t, y(t))dt = y_n + \frac{h}{2}\left(f(x_n, y_n) + f(x_{n+1}, y_{n+1})\right) - \frac{h^3}{12} y^{(3)}(\xi) \tag{12.46}$$

Στον τυπο αυτον βασιζεται η λεγομενη μεθοδος Heun:

$$Y_{n+1} = Y_n + \frac{h}{2}\left(f(x_n, Y_n) + f(x_{n+1}, Y_{n+1})\right) \tag{12.47}$$

Η ακριβεια που προσφερει η μεθοδος ειναι $y_n - Y_n = O(h^2)$, γιατι και εδω, τα τοπικα σφαλματα $O(h^3)$ μεταφερονται αθροιστικα στο επομενο βημα, ωστε στο βημα n να ειναι $n0(h^3) = (x_n - x_0)O(h^2)$.

Οπως και ολες οι αλλες μεθοδοι απο την κατηγορια μεθοδων Adams - Moulton στην οποια ανηκει, η μεθοδος αυτη εχει μορφη πεπλεγμενης εξισωσης ως προς Y . Για την ευρεση επομενως του Y πρεπει να χρησιμοποιηθει καποια επαναληπτικη μεθοδος π.χ. η γενικη επαναληπτικη μεθοδο:

$$Y_{n+1}^{(k+1)} = Y_n + \frac{h}{2}\left(f(x_n, Y_n) + f(x_{n+1}, Y_{n+1}^{(k)})\right) \tag{12.48}$$

Η συνθηκη συγκλισης αυτης της μεθοδου ειναι, οπως ξερουμε, το να ειναι απολυτα μικροτερη απο την μοναδα η παραγωγος του δευτερου μερους ως προς

τον αγνωστο Y_{n+1}:

$$L = \frac{h}{2} | f_y(x_{n+1}, Y_{n+1}) | < 1 \qquad (12.49)$$

Η συνθηκη αυτη ειναι ευκολο να εκπληρωθει γιατι μπορουμε να επιλεξουμε το h τοσο μικρο οσο χρειαζεται. Υπο την προυποθεση (12.49) ισχυει τωρα, με βαση την εκ των προτερων εκτιμηση σφαλματος της επαναληπτικης μεθοδου:

$$| Y_{n+1} - Y_{n+1}^{(k+1)} | = \frac{L^{k+1}}{1-L} | Y_{n+1}^{(1)} - Y_{n+1}^{(0)} | = O(h^{k+1}) | Y_{n+1}^{(1)} - Y_{n+1}^{(0)} |$$

Δηλαδη, σε καθε επαναληψη βελτιωνεται η ακριβεια της προσεγγισης του Y_{n+1} κατα μιαν ταξη μεγεθους ως προς h. Αν $|Y_{n+1}^{(1)} - Y_{n+1}^{(0)}| = O(h)$ αυτο σημαινει οτι για την ευρεση ικανοποιητικης προσεγγισης του Y_{n+1} αρκουν μια, ή δυο επαναληψεις, δεδομενου οτι τοτε ειναι ηδη $|Y_{n+1}^{(1)} - Y_{n+1}| = O(h^2)$ και $|Y_{n+1}^{(2)} - Y_{n+1}| = O(h^3)$. Μεγαλυτερη ακριβεια δεν χρειαζεται αφου θα είναι σε κάθε περίπτωση $y_{n+1} - Y_{n+1} = O(h^2)$.

Η ευρεση αρχικης τιμης για την επαναληπτικη διαδικασια (12.48) βασιζεται στην προσεγγιση που προκυπτει αν στην (12.10) χρησιμοποιησουμε τον απλουστερο δυνατο κανονα ολοκληρωσης που προσεγγιζει την f(t,y(t)) με την τιμη της στο σημειο x_n:

$$y_{n+1} = y_n + \int_{x_n}^{x_{n+1}} f(t, y(t))dt = y_n + hf(x_n, y_n) + \frac{h^2}{2} y''(\xi) \qquad (12.50)$$

Η μεθοδος που προκυπτει λεγεται μεθοδος Euler - Cauchy και μας δινει τον εξης τυπο για την ευρεση αρχικων τιμων:

$$Y_{n+1}^{(0)} = Y_n + hf(x_n, Y_n) \qquad (12.51)$$

Ο τυπος αυτος λεγεται predictor (προβλεπτης), γιατι "προβλεπει" την τιμη του Y_{n+1}. Αντιστοιχα, ο τυπος (12.48) λεγεται corrector (διορθωτης), γιατι διορθωνει την αρχικη προβλεψη.

Παράδειγμα Εφαρμογής της Μεθόδου Heun

Να προσεγγισθει με την μεθοδο Heun η λυση του προβληματος αρχικων τιμων

$$y' = -2xy^2 \ , \ y(0) = 1$$

στο διαστημα [0, 1] με βημα h = 0.2 και μιαν χρηση του corrector ανα βημα.

Λύση

$x_0 = 0$, $Y_0 = 1$

$x_1 = 0.2$: $Y_1^{(0)} = Y_0 + hf(x_0, Y_0) = Y_0 - 2h\, x_0 Y_0^2 = 1$

$Y_1 = Y_1^{(1)} = Y_0 + h/2\{f(x_0,Y_0) + f(x_1,Y_1^{(0)})\}$

$= Y_0 - h\{x_0Y_0^2 + x_1(Y_1^{(0)})^2\} = 0.96$

$x = 0.4$: $Y_2^{(0)} = Y_1 + hf(x_1, Y_1) = Y_1 - 2h\, x_1Y_1^2 = 0.886272$

$Y_2 = Y_2^{(1)} = Y_1 + h/2\{f(x_1,Y_1) + f(x_2,Y_2^{(0)})\}$

$= Y_1 - h\{x_1Y_1^2 + x_2(Y_2^{(0)})^2\} = 0.8602978$

κοκ.

12.10 Μέθοδοι Πολλών Βημάτων Ανωτέρας Τάξεως

Αρχη των πολυβηματικων μεθοδων ειναι η προσεγγιση του ολοκληρωματος στην σχεση (12.10) με την βοηθεια προσεγγισης της f(t,y(t)) με πολυωνυμο παρεμβολης στα ηδη γνωστα σημεια (x_{n-k},y_{n-k}), k = 0, 1, 2, .. και ενδεχομενα στο (x_{n+1}, y_{n+1}).

Αν, π.χ., χρησιμοποιησουμε σαν σημεια παρεμβολης τα τεσσερα σημεια (x_{n-k},y_{n-k}), k = 0, 1, 2, 3, τοτε εχουμε:

$$F(t, y(t) = \sum_{k=0}^{3} f_{n-k} L_{n-k}(t) + \frac{\omega(t)}{4!} y^{(5)}(\xi_n) \qquad (12.52)$$

οπου

$$f_{n-k} = f(x_{n-k}, y_{n-k}) \qquad (12.53)$$

$$L_{n-k}(t) = \prod_{j=0, j\neq k}^{3} \left(\frac{x - x_{n-j}}{x_{n-k} - x_{n-j}} \right), \quad \omega(t) = \prod_{j=0}^{3} (x - x_{n-j}) \qquad (12.54)$$

ενω για x $\in$[x_n, x_{n+1}] ειναι $\xi_n \in$ (x_n, x_{n+1}).

Λογω της (12.1) στον ορο σφαλματος αντικαταστησαμε την τεταρτη παραγωγο της f ως προς t με $y^{(5)}(t)$.

Ολοκληρωνοντας την (12.52) κατα μελη απο x_n ως x_{n+1} και εισαγοντας το αποτελεσμα στην (12.10) εχουμε:

$$y_{n+1} = y_n + \frac{h}{24}\left(55 f_n - 59 f_{n-1} + 37 f_{n-2} - 9 f_{n-3}\right) + \frac{251}{720} h^5 y^{(5)}(\xi_n) \qquad (12.55)$$

Παραλειποντας τον ορο σφαλματος εχουμε ετσι τον αναδρομικο τυπο Adams - Bashford 4^{ης} ταξεως για τον διαδοχικο προσδιορισμο των y_k.

Επειδη για την κατασκευη αυτου του τυπου χρησιμοποιηθηκαν σημεια παρεμβολης μη ευρισκομενα μεσα στο διαστημα ολοκληρωσης [x_n, x_{n+1}] για τουτο, τοσο το σφαλμα παρεμβολης οσο και το σφαλμα ολοκληρωσης ειναι αρκετα μεγαλο. Ετσι ο τυπος αυτος χρησιμοποιειται κατα βασιν σαν predictor σε μιαν διαδικασια predictor - corrector (προβλεψης - διορθωσης):

$$Y_{n+1}^{(0)} = Y_n + \frac{h}{24}\left(55 F_n - 59 F_{n-1} + 37 F_{n-2} - 9 F_{n-3}\right) \qquad (12.56)$$

οπου

$$F_k := f(x_k, Y_k) \qquad (12.57)$$

Ο αντιστοιχος corrector προκυπτει αν χρησιμοποιησουμε σαν σημεια παρεμβολης τα x_{n+1}, x_n, x_{n-1}, x_{n-2}, δηλ. το x_{n+1} αντι του x_{n-3}.

Με τον ιδιο ακριβως τροπο, εχουμε τοτε τον τυπο:

$$y_{n+1} = y_n + \frac{h}{24}\left(9 f_{n+1} + 19 f_n - 5 f_{n-1} + f_{n-2}\right) - \frac{19}{720} h^5 y^{(5)}(\vartheta_n) \qquad (12.58)$$

οπου ο συντελεστης του ορου σφαλματος ειναι δεκατρεις φορες μικροτερος απο οτι προηγουμενως (251/19 ≈ 13).

Παραλειποντας εδω τον ορο σφαλματος εχουμε τον τυπο Adams - Moulton 4ης ταξεως:

$$Y_{n+1} = Y_n + \frac{h}{24}\left(9F_{n+1} + 19F_n - 5F_{n-1} + F_{n-2}\right) \qquad (12.59)$$

Ο τυπος αυτος ειναι παλι εξισωση πεπλεγμενης μορφης γιατι $F_{n+1}= f(x_{n+1},Y_{n+1})$ και επομενως πρεπει να λυθει επαναληπτικα, π.χ. με την μεθοδο:

$$Y_{n+1}^{(k+1)} = Y_n + \frac{h}{24}\left(9F_{n+1}^{(k)} + 19F_n - 5F_{n-1} + F_{n-2}\right) \quad k = 0,1,... \qquad (12.60)$$

οπου

$$F_{n+1}^{(k)} := f(x_k, Y_{n+1}^{(k)}) \qquad (12.61)$$

Η συνθηκη συγκλισης αυτης της επαναληπτικης διαδικασιας ειναι παλι το να ειναι η παραγωγος του δευτερου μερους ως προς τον αγνωστο απολυτα μικροτερη απο την μοναδα:

$$L = \frac{9h}{24} | f_y(x_{n+1}, Y_{n+1}) | < 1 \qquad (12.62)$$

Τοσο ο τυπος (12.59), οσο και ο (12.56) μας παρεχουν ακριβεια $y_{n+1} - Y_{n+1} = O(h^4)$, παρα το οτι σε καθε χρηση τους προκυπτει σφαλμα $O(h^5)$ (βλ. (12.58), (12.55)) γιατι, λογω της αναδρομης στην αμεσως προηγουμενη τιμη για καθε νεα τιμη, τα σφαλματα $O(h^5)$ αθροιζονται, οποτε στην θεση x_{n+1} αντιστοιχει συνολικο σφαλμα $(n+1)O(h^5) = (x_{n+1} - x_0)O(h^4)$.

Σφαλμα αυτης της ταξεως προκυπτει βεβαια κατ' αρχην με πληρη συγκλιση της επαναληπτικης διαδικασιας (12.60), αλλα στη πραξη αρκουν μονο λιγες επαναληψεις, γιατι οπως και στην μεθοδο Heun, ετσι και εδω σε καθε επαναληψη βελτιωνεται η ακριβεια της προσεγγισης του Y_{n+1} κατα μιαν ταξη μεγεθους ως προς h. Για τον λογο αυτο, συνηθως αρκουν μια ως δυο χρησεις του corrector (k = 0, 1).

Σημείωση

Αφαιρωντας την (12.56) κατα μελη απο την (12.60) για k = 0 εχουμε:

$$Y_{n+1}^{(1)} - Y_{n+1}^{(0)} = \frac{3}{8}h\left(F_{n+1}^{(0)} - F_n\right) + \frac{9}{8}h\left(F_{n-1} - F_n\right) + \frac{9}{8}h\left(F_{n-1} - F_{n-2}\right) + \frac{3}{8}h\left(F_{n-3} - F_{n-2}\right) = O(h^2)$$

γιατι

$$F_{k+1} - F_k = F(x_{k+1}, Y_{k+1}) - F(x_k, Y_k) = hf_x + (Y_{k+1} - Y_k)f_y = O(h) + O(h) = O(h)$$

οποτε

$$| Y_{n+1}^{(k)} - Y_{n+1} | = \frac{L^k}{1-L} | Y_{n+1}^{(1)} - Y_{n+1}^{(0)} | = O(h^{k+2})$$

Επειδη

$$y_{n+1} - Y_{n+1} = O(h^4) \text{ και } Y_{n+1} - Y_{n+1}^{(2)} = O(h^4)$$

ειναι αρα και

$$y_{n+1} - Y_{n+1}^{(2)} = O(h^4)$$

Για να περιγραψουμε μιαν πληρη διαδικασια predictor - corrector συνηθως χρησιμοποιουμε το συμβολο P για καθε χρηση του Predictor, το συμβολο E για καθε υπολογισμο (Evaluation) της τιμης $F_{n+1}^{(k)}= f(x_{n+1},Y_{n+1}^{(k)})$ και το συμβολο C για καθε χρηση του Corrector, οποτε π.χ. η διαδικασια με δυο χρησεις του corrector, δηλ. ο υπολογισμος του $Y_{n+1}^{(2)}$, περιγραφεται σαν διαδικασια PECEC = $P(EC)^2$. Στο σημειο αυτο πρεπει να αποφασισθει επισης αν για το επομενο βημα θα χρησιμοποιηθει η τιμη $F_{n+1}^{(1)}$ ή η τιμη $F_{n+1}^{(2)}$. Στην δευτερη περιπτωση, πρεπει να υπολογισθει και αυτη η νεα τιμη, οποτε η διαδικασια περιγραφεται ακριβεστερα σαν $P(EC)^2E$.

Για να δημιουργησουμε μιαν πληρη μεθοδο με την βοηθεια των τυπων (12.56) και (12.60) πρεπει να προσεξουμε οτι για να εφαρμοσουμε για πρωτη φορα τον τυπο (12.56) πρεπει ο μικροτερος δεικτης του να εχει την τιμη 0, γιατι μικροτερες τιμες του δεικτη δεν υπαρχουν. Με n-3 = 0 εχουμε ομως n+1 = 4. Η πρωτη τιμη Y_k που μπορουμε, επομενως, να υπολογισουμε ειναι δηλαδη η Y_4, οποτε μας λειπουν οι Y_1, Y_2, Y_3, γιατι μονο η $Y_0 = y_0$ ειναι γνωστη. Οι τιμες αυτες πρεπει να προσδιορισθουν με αλλη μεθοδο.

Εδω πρεπει να προσεξουμε οτι, επειδη ολες αυτες οι μεθοδοι χρησιμοποιουν την τελευταια γνωστη τιμη, Y_n, για να βρουν την επομενη, Y_{n+1}, οποιοδηποτε λαθος γινει κατα τον προσδιορισμο μιας απο αυτες τις τιμες αυτοματα θα κληρονομηθει και απο ολες τις επομενες. Αν, π.χ. ο predictor ειχε ακριβεια $O(h^2)$ τοτε ακριβεια $O(h^4)$ στον corrector θα ηταν αχρηστη γιατι η τελικη προσεγγιση θα ειχε ακριβεια $O(h^2) + O(h^4) = O(h^2)$.

Παρομοια, αν προσδιορισουμε τις αρχικες τιμες Y_1, Y_2, Y_3 της διαδικασιας με ακριβεια $O(h^2)$, π.χ. με την μεθοδο Heun, τοτε ολες οι τιμες Y_n θα εχουν ακριβεια $O(h^2)$. Για τουτο συνηθως αυτες οι πρωτες τρεις προσδιοριζονται με την μεθοδο Runge - Kutta 4ης ταξεως. Αν, παρ' ολα αυτα θελαμε να χρησιμοποιησουμε την μεθοδο Heun για τις αρχικες τιμες, τοτε αντι για βημα h θα επρεπε να παρουμε βημα $\hat{h}= h^2$ ωστε να εχουμε ακριβεια $O(\hat{h}^2) = O(h^4)$, π.χ. βημα $\hat{h} = 0.01$ αντι $h = 0.1$, οποτε για να παρουμε τις τιμες Y_1, Y_2, Y_3 θα χρειαζομασταν 30 βηματα Heun.

Ενα ευλογο ερωτημα ειναι βεβαια, γιατι αφου ξεκιναμε με την μεθοδο Runge - Kutta να μην συνεχισουμε με αυτην. Η απαντηση βρισκεται στον ογκο υπολογισμων που απαιτει καθε διαδικασια. Η κλασσικη μεθοδος Runge - Kutta απαιτει τεσσερις υπολογισμους της συναρτησης f(x,y) σε καθε βημα. Η συνθετη μεθοδος Adams 4ης ταξεως ως προς h απαιτει σε μορφη PECE μονο δυο υπολογισμους της f, και σε μορφη $P(EC)^2E$ τρεις υπολογισμους της f, δηλαδη εχει περιπου ειτε το 1/2 ειτε τα 3/4 του υπολογιστικου ογκου της μεθοδου Runge - Kutta.

Μειονεκτημα της συνθετης μεθοδου Adams που περιγραψαμε ειναι οτι απαιτει κατ' αρχην την χρηση σταθερου βηματος, h, γιατι αν αλλαξει το βημα στην θεση x_n, πρεπει να προσδιορισθουν με το καινουργιο βημα $\hat{h}$, προσεγγισεις για τα σημεια $x_n - \hat{h}$, $x_n - 2\hat{h}$, $x_n - 3\hat{h}$, που χρειαζονται για να

προχωρησουμε στο επομενο βημα. Αυτο μπορει να γινει βεβαια, ειτε παλι με την μεθοδο Runge - Kutta, ειτε με την βοηθεια ενος πολυωνυμου παρεμβολης που χρησιμοποιει τις ηδη γνωστες προγενεστερες τιμες, π.χ. του (12.52).

12.11 Ο Έλεγχος του Σφάλματος της Μεθόδου Predictor-Corrector

Η διαπιστωση οτι σε μια μεθοδο 4ης ταξεως ειναι $y_n - Y_n = O(h^4)$ μας λεει βεβαια κατι για το αναμενομενο μεγεθος σφαλματος αλλα ειναι πολυ ασαφης. $O(h^4)$ σημαινει $\approx Ah^4$, αλλα ο συντελεστης A μπορει να ειναι πολυ μεγαλος. Ενας αμεσος τροπος ελεγχου του ολικου σφαλματος στις μεθοδους predictor - corrector βασιζεται στην συγκριση των τοπικων σφαλματων του predictor και του corrector.

Η εκτιμηση σφαλματος αυτη βασιζεται στις ακολουθες προυποθεσεις που δεν ισχυουν μεν ακριβως, αλλα αληθευουν κατα προσεγγιση:

1. Υποθετουμε οτι για k = 0(1)n ειναι $Y_k \approx y_k$ και επομενως, επισης $F_k \approx f_k$.

Απο τους τυπους (12.55) και (12.56) προκυπτει τοτε με αφαιρεση κατα μελη:

$$y_{n+1} - Y_{n+1}^{(0)} \approx \frac{251}{720} h^5 y^{(5)}(\xi) \qquad (12.63)$$

Παρομοια, με την προυποθεση οτι $y_{n+1} \approx Y_{n+1}^{(0)}$ προκυπτει απο τους (12.58) και (12.60) για k = 0:

$$y_{n+1} - Y_{n+1}^{(1)} \approx \frac{-19}{720} h^5 y^{(5)}(\vartheta) \qquad (12.64)$$

2. Υποθετουμε οτι η $y^{(5)}(x)$ δεν μεταβαλλεται πολυ στην γειτονια του σημειου x_n οπου βρισκονται τα ξ και ζ και βαζουμε:

$$y^{(5)}(\xi) \approx y^{(5)}(\theta) \qquad (12.65)$$

Απο τις (12.63) και (12.64) προκυπτει τοτε:

$$y_{n+1} - Y_{n+1}^{(1)} \approx \frac{-19}{720}\left\{y_{n+1} - Y_{n+1}^{(0)}\right\} \qquad (12.66)$$

και επομενως

$$y_{n+1} - Y_{n+1}^{(1)} \approx \frac{-19}{720}\left\{y_{n+1} - Y_{n+1}^{(0)}\right\} \approx \frac{-1}{14}\left\{Y_{n+1}^{(1)} - Y_{n+1}^{(0)}\right\} \qquad (12.67)$$

Για να ειναι

$$|y_{n+1} - Y_{n+1}^{(1)}| \in [\varepsilon_1 , \varepsilon_2] \qquad (12.68)$$

αρκει να ειναι

$$|Y_{n+1}^{(1)} - Y_{n+1}^{(0)}| \in [14\varepsilon_1 , 14\varepsilon_2] \qquad (12.69)$$

Αν η ποσοτητα $|Y_{n+1}^{(1)} - Y_{n+1}^{(0)}|$ υπερβαινει το $14\varepsilon_2$ μπορουμε, π.χ. να μειωσουμε το βημα στο μισο, $\hat{h} = h/2$, ενω αν ειναι μικροτερη απο $14\varepsilon_1$ τοτε μπορουμε να το διπλασιασουμε.

12.12 Απόλυτη Ευστάθεια – Ανάλυση Σφαλμάτων

Η αναλυση του ολικου σφαλματος αριθμητικων μεθοδων για ΚΔΕ που ακολουθει, εχει *ποιοτικό* και οχι ποσοτικό χαρακτηρα. Εν τουτοις ειναι πολυ σημαντικη, γιατι αποκαλυπτει ποιοι παραγοντες προσδιοριζουν την ακριβεια

μιας μεθοδου και πως πρεπει να την χειρισθουμε για να εχουμε αξιοπιστα αποτελεσματα. Βασικα, θα φανει το πως επηρεαζει το μεγεθος του βηματος την ποιοτητα της προσεγγισης και ετσι θα οδηγηθουμε αυτοματα και στην ταξη μεγεθους ως προς h καθε μεθοδου.

Τα αποτελεσματα που προκυπτουν απο μιαν τετοια αναλυση ειναι παρομοια για ολες τις μεθοδους, είτε ενός βήματος είτε πολλών βημάτων. Εδώ θα μελετησουμε σε βαθος ενδεικτικά για όλες τις μεθόδους πολλών βημάτων, δηλαδή μεθόδους predictor-corrector, την περιπτωση της σχετικα απλης μεθοδου Heun και θα αναφερθουμε συντομωτερα στα αναλογα αποτελεσματα για τις μεθόδους 4ης ταξεως.

Συνθήκες Απόλυτης Ευστάθειας της Μεθόδου Heun

Συμφωνα με οσα ειπαμε στη παραγραφο 12.9 η εφαρμογη του ολοκληρωτικου τυπου του τραπεζιου στο δευτερο μερος της ισοτητας (12.10), με βαση την οποια παραγουμε ολες αυτες τις μεθοδους, μας δινει:

$$y_{n+1} = y_n + \int_{x_n}^{x_{n+1}} f(t, y(t))dt = y_n + \frac{h}{2}\left(f(x_n, y_n) + f(x_{n+1}, y_{n+1})\right) - \frac{h^3}{12} y^{(3)}(\xi) \tag{12.70}$$

Με απορριψη του ορου σφαλματος εχουμε επομενως την προσεγγιστικη μεθοδο Heun:

$$Y_{n+1} = Y_n + \frac{h}{2}\left(f(x_n, Y_n) + f(x_{n+1}, Y_{n+1})\right) + \rho_n \tag{12.71}$$

Ο καινουργιος ορος ρ_n, που εισαγουμε εδω, ειναι το σφαλμα στρογγυλεματος, που οφειλεται, όπως έχουμε πει, στην εκτελεση των πραξεων και την διατηρηση των ενδιαμεσων αποτελεσματων με πεπερασμενη ακριβεια στην μνημη του Η/Υ. Δεδομένου ότι είναι πάντα της ιδίας τάξεως μεγέθους, εδώ θα τον θεωρήσουμε σταθερό ($\rho_n \approx \rho$) και θα πάρουμε ενδεικτικά $\rho = 10^{-10}$ για την περίπτωση που οι πραξεις εκτελούνται από τον Η/Υ με απλην ακριβεια (single precision). Διαφορες αποτελεσματων μικροτερες απο ρ δεν λαμβανονται υπ'οψη κατα την εκτελεση των πραξεων, οπως φαινεται π.χ. αν ζητησουμε να εκτυπωθει ο αριθμος (1 + ρ/2) -1. Αντι για τον ρ/2 θα εκτυπωθει ο αριθμός: 0.

Αφαιρωντας την (12.71) απο την (12.70) εχουμε τωρα για το σφαλμα:

$$E_k := y_k - Y_k \tag{12.72}$$

την αναδρομικη σχεση:

$$E_{n+1} = E_n + \frac{h}{2}\left(f_y(x_n, \zeta_n)E_n + f_y(x_{n+1}, \zeta_{n+1})E_{n+1}\right) - \rho - \frac{h^3}{12} y^{(3)}(\xi_n) \tag{12.73}$$

οπου εφαρμοσαμε δυο φορες το θεωρημα μεσης τιμης:

$$f(x_k, y_k) - f(x_k, Y_k) = (y_k - Y_k) f_y(x_k, \zeta_k) = E_k f_y(x_k, \zeta_k) \tag{12.74}$$

Για να μελετησουμε την εξελιξη των σφαλματων με βαση την σχεση (12.73) κανουμε πάλι ωρισμενες απλουστευτικες υποθεσεις, που περιοριζουν μεν την αυστηρη ισχυ των αποτελεσματων, αλλα δεν επηρεαζουν την ποιοτικη ισχυ των συμπερασματων.

Θα υποθεσουμε, λοιπον, οτι:

1. $\rho_n \approx \rho$, όπου το ρ είναι σταθερο
2. $y^{(3)}(\xi_n) \approx 12A$, με A σταθερό.
3. $f_y(x_k, \zeta_k) \approx \lambda$, με λ σταθερό.

Η πρωτη υποθεση είναι, όπως είπαμε, ευλογη, αφου ολα τα σφαλματα στρογγυλεματος ειναι της ιδίας ταξης μεγέθους. Η δευτερη και η τριτη ειναι ευλογες, αν το βημα h ειναι αρκετά μικρο και οι συναρτήσεις $y^{(3)}(x)$ και $f_y(x,y(x))$ ομαλές. Για παράδειγμα, αν $h=10^{-4}$ τότε ακόμα και 1000 διαδοχικά βήματα ισοδυναμούν με μία μεταβολή του x μόνο κατά 1000h=0.1 και επομένως με μίαν σχετικά μικρή μεταβολή των παραπάνω συναρτήσεων αν είναι συνεχείς στην περιοχή αυτή. Αν έχουν ασυνέχειες ή ιδιόμορφα σημεία, τότε η ανάλυση που κάνουμε ενδέχεται πράγματι να μην ισχύει.

Όπως έχουμε πει και προηγουμένως, η υποθεση (3) ισχυει σε καθε περιπτωση αυστηρα αν ειναι f(x, y) = λy + μ(x), δηλαδη αν η διαφορικη εξισωση ειναι γραμμικη με σταθερο συντελεστη του y.

Σημειωνουμε επισης, οτι, λογω της (12.71), υποτιθεται πως ο corrector της μεθοδου Heun χρησιμοποιειται, οχι μια η δυο φορες, αλλα μεχρι πληρους συγκλισεως, δηλαδη εως οτου δεν μεταβαλλεται κανενα απο τα δεκαδικα ψηφια μετα την υποδιαστολη που μας δινει ο Η/Υ.

Με βαση τις παραπανω υποθεσεις η (12.73) γινεται:

$$E_{n+1} = E_n + \frac{h\lambda}{2}\left(E_n + E_{n+1}\right) - \rho - Ah^3 \qquad (12.75)$$

Δηλαδή

$$E_{n+1} = \alpha E_n + \beta, \quad n = 1,2,3,...$$

με

$$\alpha = \frac{1 + h\lambda/2}{1 - h\lambda/2} \quad \beta = -\frac{\rho + Ah^3}{1 - h\lambda/2}$$

Επομένως έχουμε πάλι μίαν σχέση ανάλογη προς την (12.30):

$$E_{n+1} = \alpha^n E_1 + \beta\frac{1-\alpha^n}{1-\alpha}, \quad n = 1,2,3,... \qquad (12.76)$$

Η σχέση αυτή μας δείχνει ότι το σφάλμα E_n τείνει για $n\to\infty$ είτε προς το άπειρο, αν είναι $|\alpha|>1$, είτε να σταθεροποιηθεί στην τιμή

$$E_\infty = \frac{\beta}{1-\alpha} = \frac{\rho + Ah^3}{h\lambda} \qquad (12.77)$$

αν είναι $|\alpha|<1$, δηλαδή:

$$|\alpha(h\lambda)| = \left|\frac{1 + h\lambda/2}{1 - h\lambda/2}\right| < 1 \qquad (12.78)$$

Η ανισότητα αυτή λεγεται "Συνθηκη Απολυτης Ευσταθειας" της μεθοδου και ισοδυναμεί, με

$$h\lambda<0 \qquad (12.79)$$

δηλαδή ισχύει για όλες τις τιμές h>0 αν είναι το $\lambda=f_y$ αρνητικό, δηλαδή η λύση φραγμένη.

Προϋπόθεση για την ισχύ της (12.79) είναι, όπως είπαμε το να γίνονται οι επαναλήψεις (12.48) μέχρι τελικής συγκλίσεως, ή εναλλακτικά το να προσδιορίζεται η λύση της εξίσωσης (12.47) με μίαν μέθοδο ταχείας σύγκλισης, όπως η μέθοδος Newton-Raphson ή η Μέθοδος Τέμνουσας. Αν κάνουμε μόνο μία ή δύο χρήσεις του corrector, τότε καταλήγουμε πάλι σε μίαν συνθήκη απόλυτης ευστάθειας του τύπου (12.35), όπως θα ιδούμε.

Η συνθηκη (12.79) σημαινει οτι, για να παραμενει φραγμενο το ολικο σφαλμα (15.58) πρεπει οπωσδηποτε να ειναι $\lambda = f_y < 0$.

Βασικα σημαινει οτι φραγμενο σφαλμα μπορουμε να εχουμε μονο αν η λυση του προβληματος αρχικων τιμων τεινει σε μιαν στασιμη κατασταση. Για παράδειγμα, η λυση του προβληματος

$$y' = \lambda y + \mu \ , \ y(x\ ) = y$$

ειναι

$$y(x) = -\ \mu/\lambda + (y_0 + \mu/\lambda)e^{\lambda x} \rightarrow -\mu/\lambda$$

για $x \rightarrow \infty$.

Υπο την προυποθεση, λοιπον, οτι $f_y<0$, το συνολικο σφαλμα (12.76) παραμενει στην προκειμενη περιπτωση φραγμενο για $n\rightarrow\infty$ ασχετως του τι τιμη εχει το βημα h (αρκει να ειναι θετικη). Λογω της (12.78) ειναι μαλιστα τοτε ο πρωτος προσθετεος στη (12.76) για μεγαλα n αμελητεος και ισχυει:

$$f_y E_{n+1} \approx Ah^2 + \frac{\rho}{h} \qquad (12.80)$$

Η σχεση αυτη μας λεει οτι το μεγεθος του $f_y(y_{n+1} - Y_{n+1})$ ειναι ουσιαστικα αναλογο του h , αν το βημα h δεν ειναι παρα πολυ μικρο. Με απλη ακριβεια εκτελεσης των πραξεων ($\rho\approx10^{-10}$), αυτο σημαινει οτι η μικρότερη τιμή του σφάλματος, $f_yE_n\approx10^{-6}$ προκύπτει για $h\approx10^{-4}$. Αν επιλέξουμε $h\approx10^{-5}$, τοτε ο πρωτος μεν ορος στην (12.80) ειναι της ταξεως του 10^{-10}, αλλα ο δευτερος ειναι 10^{-5} και υπερισχυει. Αν με διπλην ακριβεια εκτελεσης πραξεων, ειναι $\rho\approx10^{-22}$ μπορουμε να παμε μεχρι $h\approx10^{-7}$ για να εχουμε σφαλμα $f_yE_n\approx O(10^{-15})$.

Απόλυτη Ευστάθεια της Μεθόδου Heun σε Μορφή PECE

Ο τυπος (12.77) για το ολικο σφαλμα, και η συνθήκη απολυτης ευσταθειας (12.78) ισχυουν μονον αν ο corrector της μεθοδου Heun χρησιμοποιειται μεχρι την πληρη συγκλιση του. Τι συμβαινει ομως αν χρησιμοποιουμε μονο μια φορα τον corrector;

Υπο τις ιδιες τρεις απλοποιητικες προυποθεσεις που χρησιμοποιησαμε και προηγουμενα, ισχυει για τον predictor και τον corrector:

α. predictor:

$$y_{n+1} = y_n + hf(x_n, y_n) + \frac{h^2}{2} y''(\xi_n) \qquad (12.81)$$

$$Y_{n+1}^{(0)} = Y_n + hf(x_n, Y_n) + \rho \qquad (12.82)$$

και με αφαιρεση κατα μελη,

$$E_{n+1}^{(0)} = y_{n+1} - Y_{n+1}^{(0)} = E_n + h\lambda E_n + Bh^2 - \rho \qquad (12.83)$$

οπου χρησιμοποιησαμε την προυποθεση οτι $f_y \approx \lambda$ = σταθερό και θεωρησαμε οτι $y''(\xi_n) \approx 2B$ = σταθερό.

β. corrector:

$$y_{n+1} = y_n + \frac{h}{2}\left(f(x_n, y_n) + f(x_{n+1}, y_{n+1})\right) - \frac{h^3}{12} y^{(3)}(\xi) \qquad (12.84)$$

$$Y_{n+1} = Y_n + \frac{h}{2}\left(f(x_n, Y_n) + f(x_{n+1}, Y_{n+1})\right) + \rho_n \qquad (12.85)$$

Άρα,

$$E_{n+1}{}^{(1)} = y_{n+1} - Y_{n+1}{}^{(1)} = E_n + \frac{h\lambda}{2}\left(E_n + E_{n+1}{}^{(0)}\right) - Ah^3 - \rho \qquad (12.86)$$

Συνολικά είναι, δηλαδή (με $E_k := E_k^{(1)}$):

$$E_{n+1} = \left(1 + h\lambda + \frac{1}{2}(h\lambda)^2\right)E_n - \left(A - \frac{B\lambda}{2}\right) - \left(1 + \frac{h\lambda}{2}\right)\rho \qquad (12.87)$$

Δηλαδή

$$E_{n+1} = \alpha E_n + \beta, \quad n = 1,2,3,... \qquad (12.88)$$

με

$$\alpha = 1 + h\lambda + \frac{1}{2}(h\lambda)^2, \quad \beta = -\left(A - \frac{B\lambda}{2}\right) - \left(1 + \frac{h\lambda}{2}\right)\rho \qquad (12.89)$$

Επομένως έχουμε πάλι μίαν σχέση ανάλογη προς την (12.30) και την (12.76):

$$E_{n+1} = \alpha^n E_1 + \beta \frac{1-\alpha^n}{1-\alpha}, \quad n = 1,2,3,... \qquad (12.90)$$

και καταλήγουμε στο συμπέρασμα, ότι αν

$$|\alpha(h\lambda)| = \left|1 + h\lambda + \frac{1}{2}(h\lambda)^2\right| < 1 \qquad (12.91)$$

τότε για n>>1 θα είναι

$$\lambda E_{n+1} \approx \lambda E_\infty = \frac{\beta}{1-\alpha} = \frac{1}{1 + h\lambda/2}\left\{\left(A + \frac{B\lambda}{2}\right)h^2 + \frac{\rho}{h}\left(1 + \frac{h\lambda}{2}\right)\right\} \qquad (12.92)$$

συνθηκη που ισοδυναμει με:

$$-2 < h\lambda < 0 \qquad (12.93)$$

οταν το λ ειναι πραγματικο.

Βλεπουμε, δηλαδη, οτι η συνθηκη απολυτης ευσταθειας περιοριζεται στην μορφη PECE της μεθοδου σημαντικα και παιρνει την μορφη:

$$-2 < hf_y < 0 \qquad (12.94)$$

Με $f_y < 0$, αυτο σημαινει οτι πρεπει να ειναι

$$0 < h < \frac{2}{|f_y|} \qquad (12.95)$$

Το ότι η ανάλυση σφαλμάτων αυτού του είδους δεν είναι πάντοτε επιτυχής φαίνεται από το ακόλουθο παράδειγμα. Η λύση του προβλήματος δεν είναι

φραγμένη ($y(x)\approx x^2-2$ για $x>>1$), παρά το ότι ισχύει η σύνθήκη: $f_y=-x<0$, και επομένως καμμία επιλογή βήματος δεν μπορεί να εξασφαλίσει το να παραμένουν φραγμένα τα σφάλματα.

Παράδειγμα

Για το πρόβλημα:

$$y'=-x(y-x^2),\ y(1)=2 \qquad (12.96)$$

που η ακριβής λύση του είναι

$$y(x)=3\exp\left(\frac{1-x^2}{2}\right)+x^2-2 \qquad (12.97)$$

η συνθήκη απόλυτης ευστάθειας για την μέθοδο Heun σε μορφή PECE μας δίνει:

$$-2<hf_y=-xh<0$$

δηλαδή

$$0<h<2/x$$

Αυτό σημαίνει ότι για μεγάλες τιμές του x το h πρέπει να γίνεται όλο και μικρότερο για να έχουμε φραγμένο το σφάλμα της μεθόδου (τους όρους που εξαρτώνται από το a^n), πράγμα που θα κάνει υπέρμετρα μεγάλο τον όρο σφάλματος που οφείλεται στις στρογγυλεύσεις.

Απόλυτη Ευστάθεια της μεθόδου Adams-O(h^4) σε Μορφή PECE

Μία ανάλογη ανάλυση μας δίνει για τον corrector της μεθόδου Adams 4ης τάξεως, δηλαδή τον τύπο Adams-Moulton την συνθήκη απόλυτης ευστάθειας:

$$-3<hf_y<0 \qquad (12.98)$$

Αν όμως χρησιμοποιούμε τον corrector μόνο μία φορά σε κάθε βήμα και όχι μέχρι την πλήρη του σύγκλιση μέχρι το όριο του αριθμητικού σφάλματος, τότε η συνθήκη είναι

$$-1.25<hf_y<0 \qquad (12.99)$$

και περιορίζει πολύ περισσότερο την επιλογή του βήματος h (Βλέπε Κιουστελίδη, 1996, σελ. 314-317).

Άσκηση

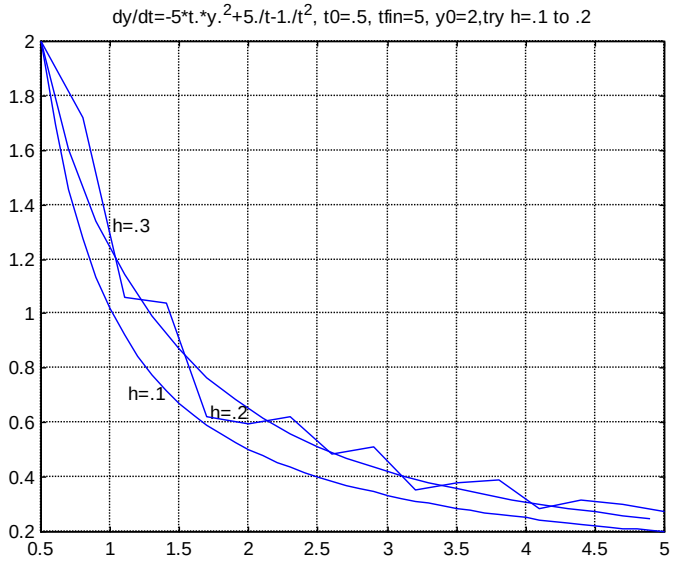
Με εφαρμογή των μεθόδων Heun ή Runge-Kutta 2ας τάξεως στο MATLAB σχεδιάστε την λύση του προβλήματος:

$$y' = -5xy^2 + 5/x - 1/x^2,\ y(0.5) = 2 \text{ ή } y(1) = 1 \qquad (12.100)$$

που η ακριβης του λυση ειναι $y(x) = 1/x$, τόσο με $h=0.1$ όσο και με $h=0.2$ και $h=0.3$. Προσέξτε ότι η συνθηκη απολυτης ευσταθειας είναι εδώ: $-2 < hf_y < 0$ και επομενως οι τιμες του h που εξασφαλιζουν απόλυτη ευσταθεια ειναι:

$$0 < h < 0.2/(x_nY_n)$$

Το βημα h πρεπει δηλαδη να ειναι μικροτερο απο 0.2 αφου εδώ ισχύει $x_ky_k=1$. Πραγματι, με $h = 0.1$ εχουμε ως αριθμητικη λυση μιαν καμπυλη που προσεγγίζει ικανοποιητικά την $y = 1/x$, ενω με $h = 0.2$ και $h=0.3$ οι παραγόμενες προσεγγίσεις να διαφέρουν αισθητά, όπως φαίνεται από το επόμενο γράφημα των λύσεων αυτών.



12.13 Σχετική Ευστάθεια

Ειδαμε παραπανω οτι η εξασφαλιση απολυτης ευσταθειας μιας μεθοδου δεν εξαρταται μονο απο το ποσο μεγαλο επιλεγουμε το βημα, αλλα πριν απο όλα απο το αν ισχυει $f_y < 0$ ή οχι! Τι κανουμε ομως αν για το προβλημα που αντιμετωπιζουμε ειναι $f_y > 0$;

Για παράδειγμα, πώς θα προσεγγισουμε αριθμητικα τη λυση του προβληματος:

$$y' = x^2 + y^2 \text{ , } y(0) = y_0 > 0; \qquad (12.101)$$

Εδω ειναι $y'(x) > 0$, $y_0 > 0$, αρα και $y(x) > 0$, καθως και $f_y(x,y) = 2y > 0$.

Ενα τεχνασμα που μπορουμε να μεταχειρισθουμε τοτε, αν τυχον ισχυει $f_y>0$, ειναι το να γραψουμε το προβλημα αρχικων τιμων στην μορφη:

$$\frac{dx}{dy} = \frac{1}{f(x,y)} := \varphi(x,y),\ x(y_0) = x_0 \qquad (12.102)$$

οποτε εχουμε

$$\varphi_x = -\frac{f_x}{f^2} < 0 \qquad (12.103)$$

Το τεχνασμα αυτο ομως δεν ειναι εφαρμοσιμο στην περίπτωση συστημάτων διαφορικών εξισώσεων. Για τις περισσοτερες μεθοδους προκύπτει όμως για $\lambda>0$, δηλαδη $f_y>0$, ότι το απολυτο σφαλμα τεινει μεν στο απειρο, αλλα επειδη και η λυση αυξανεται συνεχως, το σχετικο σφαλμα E_n/y_n τεινει στο μηδεν (Για την κλασσικη μεθοδο Runge - Kutta 4ης ταξεως βλεπε, π.χ. Jain σελ .281-282). Αυτα συμβαινουν, αν οχι για ολες τις τιμες του h (οπως, π.χ. στην κλασσικη μεθοδο Runge - Kutta) , ουσιαστικα για τοσο ευρειες περιοχες τιμων του h, ωστε σπανια

γινεται σαφης αναφορα σ'αυτες.

Απο τις παραπανω αναλυσεις του ολικου σφαλματος φαινεται οτι καθε διαφορικη εξισωση, ειτε με $f_y<0$, ειτε με $f_y> 0$, μπορει να επιλυθει ικανοποιητικα με την βοηθεια μιας καταλληλης αριθμητικης μεθοδου. Η εντυπωση αυτη ειναι ομως εν μερει παραπλανητικη. Τι γινεται αν η συναρτηση $f_y(x,y)$ αλλαζει προσημο μεσα στο διαστημα που μας ενδιαφερει η λυση;

Στα διαστηματα οπου εχουμε απολυτη ευσταθεια το απολυτο σφαλμα ειναι μικρο. Στα διαστηματα ομως οπου εχουμε σχετικη ευσταθεια μονο το σχετικο σφαλμα ειναι μικρο ενω το απολυτο τεινει να αυξηθει. Αν τετοια διαστηματα εναλλασσονται τοτε η ολη αξιοπιστια της προσεγγισης με την αριθμητικη λυση ειναι συζητησιμη.

Αλλου ειδους προβλημα με την επιλογη του βηματος h εχουμε στις λεγομενες "Δυσκαμπτες Δ.Ε."(stiff differential equations). Οι λυσεις τετοιων Δ.Ε. περιλαμβανουν ορους που αλλοι μεν μεταβαλλονται (π.χ. σβυνουν) εξαιρετικα γρηγορα, και αλλοι μεταβαλλονται εξαιρετικα αργα. Η λυση εχει επομενως, αλλου περιοχες γρηγορης μεταβολης, οποτε το βημα πρεπει να ειναι πολυ μικρο για να συλλαβει αυτες τις γρηγορες μεταβολες, και αλλου περιοχες πολυ αργης μεταβολης, οποτε το βημα πρεπει να ειναι μεγαλο, αφου οι μεταβολες ειναι αλλοιως ανεπαισθητες. Μεγαλες τιμες του βηματος αποκλειονται ομως γενικα απο τις μεθοδους ανωτερας ταξεως (π.χ. για την μεθοδο Runge - Kutta ειναι $h \leq 2.785/|f_y|$) ενω πολυ μικρες τιμες του βηματος αυξανουν σημαντικα τον υπολογιστικο ογκο, αλλα και το σφαλμα στρογγυλευσεως.

Για τετοιου τυπου Δ.Ε. εχουν κατασκευασθει ειδικες μεθοδοι, με πολυ ευρυτερα διαστηματα απολυτης ευσταθειας απο αυτα των κλασσικων μεθοδων, ωστε να ειναι δυνατη η χρηση μεγαλου βηματος οπου η λυση μεταβαλλεται πολυ αργα. Γι' αυτες θα μιλησουμε λιγο αργοτερα.

12.14 Χαοτική Συμπεριφορά Μη Γραμμικών Συστημάτων

Πολυ χειροτερη ειναι ωρισμενες φορες η θεση μας αν ζητουμε να προσεγγισουμε αριθμητικα την λυση Δ.Ε. που περιγραφουν την συμπεριφορα μη γραμμικων συστηματων που μπορουν να εμφανισουν ιδιοταλαντωσεις.

Αν σε ενα τετοιο συστημα εφαρμοσθει μια εξωτερικη περιοδικη δυναμη, τοτε η κινηση που προκυπτει μπορει να ειναι περιοδικη, ημιπεριοδοκη (δηλ. οριακως περιοδικη) η χαοτικη, δηλαδη φαινομενικα εντελως ακανονιστη και τυχαια.

Στην τελευταια αυτην περιπτωση η ευαισθησια της λυσης σε απειροελαχιστες μεταβολες των αρχικων τιμων ειναι τοσο μεγαλη, ωστε καθε αριθμητικη λυση ειναι φαινομενικα αχρηστη.

Αυτο μπορει να προκυψει π.χ. σε τοσο απλα συστηματα οπως ενα μαθηματικο εκκρεμες υπο την επιδραση εξωτερικης συνημιτονοειδους περιοδικης ωθησης, οποτε εχουμε την Δ.Ε.:

$$\theta'' + \sin\theta = A\cos\omega t \qquad (12.104)$$

οπου θ η γωνια αποκλισης απο την καθετο και ω η συχνοτητα της περιοδικης ωθησης.

Σε ενα τετοιο συστημα, υπαρχουν τιμες της παραμετρου, η των παραμετρων, (εδω της A) που αν μεταβληθουν ανεπαισθητα η κινηση γινεται απο ημιπεριοδικη χαοτικη.

Αν σχεδιασει κανεις την ταχυτητα σε εξαρτηση απο την θεση σε ενα τετοιο συστημα (εδω $\theta'=f(\theta)$), τοτε η περιοδικη κινηση αντιστοιχει σε μιαν κλειστη καμπυλη (αφου το συστημα επανερχεται περιοδικα στην ιδια θεση με την ιδια ταχυτητα). Η ημιπεριοδικη κινηση αντιστοιχει σε μιαν ελικοειδη τροχια που τεινει να συμπεσει με μιαν κλειστη καμπυληενω η χαοτικη κινηση αντιστοιχει σε μιαν τροχια που διερχεται απειρες φορες απο την ιδια περιοχη φαινομενικα εντελως ακανονιστα, ωστε να δινει την εντυπωση οτι ειναι τυχαια.

Αν το συστημα εχει απωλειες ενεργειας (dissipative system) οποτε η Δ.Ε. γινεται:

$$\theta'' + \theta/q + \sin\theta = A\cos\omega t \qquad (12.105)$$

οπου q η παραμετρος αποσβεσης, τοτε το συστημα μπορει να εμφανισει αναλογα με το ποιες τιμες παιρνει το q ολες τις μορφες χαους που ειναι γνωστες (Βλεπε Baker - Golub: Chaotic Dynamics, Cambridge U.P., 1990, οπου μελεταται διεξοδικα αυτη η Δ.Ε). Για ωρισμενες τιμες του q η τροχια τεινει να περιστρεφεται περι ωρισμενα σημεια που αποτελουν στασιμες λυσεις, μεταπηδωντας ομως με τροπο ακανονιστο και μη προβλεψιμο απο το ενα στο αλλο. Για τουτο τετοια σημεια λεγονται "παραξενοι ελκυστες" (strange attractors). Σε αλλες περιοχες τιμων του q η τροχια εμφανιζει με ανεπαισθητες μεταβολες του q συνεχεις διπλασιασμους της περιοδου οποτε συντομα η τροχια εμφανιζει ενα χαος περιοδων (και συχνοτητων) ενω για αλλα q η τροχια εμφανιζει τις μορφες χαους που λεγονται ομοκλινικο και ετεροκλινικο μπερδεμα (homoclinic - or heteroclinic tangle), ενα εντελως ακανονιστο κουβαριασμα.

Αν το συστημα ειναι συντηρητικο, δηλ. δεν εχει απωλειες ενεργειας, τοτε δεν μπορει να εχει παραξενους ελκυστες. Εν τουτοις μπορει παλι σε ωρισμενες περιοχες τιμων των παραμετρων να εμφανισει μεγαλην ευαισθησια σε μικρες μεταβολες των παραμετρων. Τοτε η τροχια ποτε γινεται φαινομενικα περιοδικη η ημιπεριοδικη και ποτε εξελισσεται σε ενα εντελως ακανονιστο κουβαριασμα.

Ακριβεστερη σχεδιαση των φαινομενικα ημιπεριοδικων τμηματων της τροχιας δειχνει οτι και αυτες αποτελουνται απο περιοχες ημιπεριοδικης τροχιας περιβαλλομενες απο τμηματα εντελως ακανονιστης κινησης. Το φαινομενο αυτο επαναλαμβανεται απεριοριστες φορες, οσο ακριβεστερα σχεδιαζουμε την τροχια, και λεγεται "αυτοομοιοτητα" (self - similarity).

Οπως ειπαμε, τετοιες Δ.Ε. εχουν για ωρισμενες περιοχες τιμων των παραμετρων τους μεγαλην ευαισθησια σε μεταβολες των αρχικων τιμων ωστε εστω και πολυ μικρες αριθμητικες ανακριβειες να παραποιουν συντομα σημαντικα την λυση. Αυτο σημαινει οτι οι αριθμητικες λυσεις δεν συμπιπτουν εν γενει αριθμητικα με την ακριβη λυση εστω και κατα τα πρωτα δεκαδικα ψηφια.

Παρ' ολα αυτα, η αριθμητικη λυση εχει σε πολλες περιπτωσεις ποιοτικην αξια, γιατι μπορει να αποδειχθει οτι μια "ψευδοτροχια" προερχομενη απο αριθμητικη λυση, ακολουθει στενα μιαν ακριβη τροχια, που δεν εχει ομως το ιδιο σημειο εκκινησεως, δηλ. τις ιδιες αρχικες τιμες. Το φαινομενο αυτο λεγεται "shadowing" ('στενη παρακολουθηση'). Αριθμητικες λυσεις εχουν, δηλαδη, ποιοτικη και οχι ποσοτικη σημασια στην περιπτωση περιοχων χαοτικης συμπεριφορας ενος συστηματος (Βλεπε, π.χ., επισης: Parker - Chua : Practical

Numerical Algorithms for Chaotic Systems, Springer, 1989, σελ.110).

Εδω πρεπει ισως να επισημανθει οτι χαοτικη συμπεριφορα μπορουν να εμφανισουν μονο μη γραμμικες Δ.Ε., ειτε τουλαχιστον 2ας ταξεως με μη ομογενη ορο, ειτε 3ης ταξεως και ανω. Μια μονο μη γραμμικη Δ.Ε. πρωτης ταξεως αποκλειεται να συμπεριφερθει χαοτικα. Χαοτικη συμπεριφορα μπορουν να εμφανισουν και αυτονομα συστηματα μη γραμμικων Δ.Ε. πρωτης ταξεως (δηλ. συστηματα της μορφης $y' = f(y)$, οπου το δευτερο μερος δεν εξαρταται αμεσα απο την ανεξαρτητη μεταβλητη x) αν εχουν τρεις τουλαχιστον αγνωστες συναρτησεις ($y \in \mathbf{R}^3$, $n \geq 3$).

Παραδείγματα Δ.Ε. με Χαοτική Συμπεριφορά

1. Οι εξαναγκασμενες (με περιοδικη διεγερση) πλευρικες ταλαντωσεις μιας καμπτομενης δοκου (Buckled Beam under forced lateral vibrations) περιγραφονται απο την ακολουθη Δ.Ε.:

$$y'' - y + 2y^3 = \varepsilon \cos t \qquad (12.106)$$

Η λυση της Δ.Ε. αυτης μπορει να συμπεριφερθει χαοτικα για ωρισμενες τιμες της παραμετρου ε (Βλ. Kirchgraber - Stoffer: Chaotic Behaviour in Simple Dynamical Systems, SIAM Review, vol. 32, pp. 424 - 452 σελ.449).

2. Αν ενα μαθηματικο εκκρεμες με αποσβεση εχει περιοδικα μετακινουμενο σημειο αναρτησεως τοτε οι ταλαντωσεις του περιγραφονται απο την Δ.Ε.:

$$\varphi'' + \varepsilon\lambda\varphi' + (1 - \varepsilon a \sin t)\sin \varphi = 0 \qquad (12.107)$$

που μπορει επισης να συμπεριφερθει χαοτικα για ωρισμενες τιμες της παραμετρου ε (Kirchgraber - Stoffer).

3. Ενα απλο αυτονομο συστημα, με γνωση μαλιστα λυση, που να συμπεριφερεται χαοτικα μπορουμε να κατασκευασουμε αν χρησιμοποιησουμε μια μικρη γενικευση της συναρτησης (2.25) που μελετησαμε στο κεφαλαιο περι μη γραμμικων εξισωσεων:

$$y(t) = \sin (a\pi e^{bt}) \qquad (12.108)$$

Διαφοριζοντας δυο φορες την συναρτηση αυτην καταληγουμε στο συστημα:

$$\begin{aligned} x' &= 2bx & x(0) &= a^2b^2\pi^2 \\ y' &:= z & y(0) &= \sin(a\pi) \\ z' &= -xy + bz & z(0) &= ab\pi\cos(a\pi) \end{aligned} \qquad (12.109)$$

που έχει την λυση:

$$y(t) = \sin(a\pi e^{bt}),\ x(t) = a^2b^2\pi^2e^{2bt},\ z(t) = y'(t) \qquad (12.110)$$

Η ευαισθησια της λυσης σε μικρες μεταβολες των αρχικων τιμων φαινεται αν αντι της παραμετρου a μπει στις αρχικες τιμες η παραμετρος:

$$a^* = a + de^{-b\tau}$$

Τοτε για την λυση

$$y^*(t) = \sin (a^*\pi e^{bt})$$

ισχυει

$$y^*(\tau) = \sin(a\pi e^{b\tau} + \pi d) \quad \text{ενω} \quad y(\tau) = \sin (a\pi e^{b\tau})$$

Για d=1 προκυπτει ετσι $y(\tau) = -y(\tau)$, ενω για d = 1/2 είναι $y^*(\tau) = \sqrt{1 - y^2(\tau)}$ και για d = 3/2 ειναι $y^*(\tau) = -\sqrt{1 - y^2(\tau)}$.

Προσεξτε οτι Δ.Ε. με $f_y < 0$, οπως η:

$$y' = -\alpha y + \beta \sin x$$

με α > 0 που ειναι ευσταθεις, δηλ. η λυση τους μεταβαλλεται ελαχιστα για μικρες μεταβολες της αρχικης τιμης, εχουν και ασυμπτωτικα ευσταθη λυση με οποιαδηποτε αριθμητικη μεθοδο ανωτερας ταξεως για καταλληλο h. Η αριθμητικη λυση προσεγγιζει επομενως εν γενει ικανοποιητικα την ακριβη λυση *για όλες τις τιμές του* x, πραγμα που στο παραπανω συστημα ειναι προφανως αδυνατο, αφου ελαχιστες μεταβολες των αρχικων τιμων μεταβαλλουν λιγο αργοτερα σημαντικα την λυση.

12.15 Δύσκαμπτες Δ.Ε. (Stiff Differential Eq.)

Δυσκαμπτες ονομαζονται, καθως ειπαμε, διαφορικες εξισωσεις που οι λυσεις τους περιλαμβανουν ορους που αλλοι μεν μεταβαλλονται (π.χ. σβυνουν) εξαιρετικα γρηγορα, και αλλοι μεταβαλλονται εξαιρετικα αργα. Η λυση εχει επομενως, αλλου περιοχες γρηγορης μεταβολης, οποτε το βημα πρεπει να ειναι πολυ μικρο για να συλλαβει αυτες τις γρηγορες μεταβολες, και αλλου περιοχες πολυ αργης μεταβολης, οποτε το βημα πρεπει να ειναι μεγαλο, αφου οι μεταβολες ειναι αλλοιως ανεπαισθητες.

Για να ιδουμε πιο καθαρα την παθολογια τετοιων προβληματων ας κατασκευασουμε κατ' αρχην ενα τυπικο παραδειγμα αυτου του ειδους χρησιμοποιωντας το προβλημα αρχικων τιμων:

$$y' = -\alpha y + \beta \sin x, \quad y(0) = y_0 \qquad (12.111)$$

που εχει την λυση:

$$y(x) = \left(y_0 + \frac{\beta}{\alpha^2 + 1} \right) e^{-\alpha x} + \frac{\beta}{\alpha^2 + 1} \left(\alpha \sin x - \cos x \right) \qquad (12.112)$$

Για α = β = 100 και y_0 = 0 το προβλημα αυτο παιρνει την μορφη:

$$y' = 100(\sin x - y) \quad y(0) = 0 \qquad (12.113)$$

με την λυση:

$$y(x) = \frac{1}{1.0001} \left(\sin x - 0.01 \cos x + 0.01 e^{-100x} \right) \qquad (12.114)$$

(βλεπε Ralston - Rabinowitz, σελ. 228).

Ο εκθετικος ορος στην παρασταση αυτην σβυνει τοσο γρηγορα ωστε επηρεαζει την λυση π.χ. μονο για $0 \le x \le 0.1$, γιατι για x = 0.1 ειναι ηδη μικροτερος απο 10^{-6}. Παρ' ολα αυτα, η συνθηκη απολυτης ευσταθειας π.χ. για την μεθοδο Runge - Kutta 4ης ταξεως:

$$-2.785 < h f_y < 0$$

μας δινει εδω (με $f_y = -100$):

$$h < 0.027..$$

δηλαδη ενα βημα στην ταξη των δυο εκατοστων. Το βημα αυτο ειναι τοσο μικρο, ωστε π.χ. για x > 0.1 οι μεταβολες της λυσης απο βημα σε βημα ειναι ανεπαισθητες. Εδω ενα βημα της ταξεως του 0.1, δηλαδη 5 φορες μεγαλυτερο, θα ηταν υπεραρκετο αν η συνθηκη ευσταθειας δεν το απεκλειε. Ετσι, ηδη με h = 0.03 (λιγο εξω απο το διαστημα απολυτης ευσταθειας) προκυπτει $y(3) \cong 6.7*10^{11}$ ενω με h = 0.025 (μεσα στο διαστημα ευσταθειας) ειναι $y(3) \cong 0.150943$, τιμη που αποτελει καλην προσεγγιση της ακριβους λυσης.

Εδω βλεπουμε οτι για να μπορουμε να προσαρμοσουμε το βημα στην ταχυτητα μεταβολης της λυσης πρεπει να εχουμε ενα διαστημα ευσταθειας οσο το δυνατον μεγαλυτερο. Τετοιο διαστημα ευσταθειας εχει π.χ. η μεθοδος τραπεζίου:

$$Y_{n+1} = Y_n + \frac{h}{2}\left(f(x_n, Y_n) + f(x_{n+1}, Y_{n+1})\right) \qquad (12.115)$$

($hf_y < 0$ ή $Re\{hf_y\} < 0$) αλλα monon an to Y_{n+1} *προσδιορίζεται με πλήρη ακρίβεια (μέχρι το όριο των σφαλμάτων στρογγύλευσης),* π.χ. με χρηση της μεθοδου Newton - Raphson εως οτου το σφαλμα της επαναληπτικης διαδικασιας γινει μικροτερο απο το σφαλμα στρογγύλευσης.

Η συνθηκη ευσταθειας της μεθοδου αυτης είναι:

$$E_{n+1} = E_n + \frac{h\lambda}{2}\left(E_n + E_{n+1}\right) - \rho - Ah^3 \qquad (12.116)$$

οπότε

$$E_{n+1} = \alpha^n\left(E_1 - \frac{\beta}{1-\alpha}\right) + \frac{\beta}{1-\alpha}, \quad n = 1,2,3,...$$

με

$$\alpha = \frac{1 + h\lambda/2}{1 - h\lambda/2} \quad \beta = -\frac{\rho + Ah^3}{1 - h\lambda/2}$$

δηλαδή

$$E_{n+1} = \left(\alpha = \frac{1 + h\lambda/2}{1 - h\lambda/2}\right)^n \left(E_1 - \frac{\rho + Ah^3}{h\lambda}\right) + \frac{\rho + Ah^3}{h\lambda} \qquad (12.117)$$

Απο την σχεση αυτην προκυπτει ομως οτι ο πρωτος ορος του σφαλματος σβυνει για μεγαλες τιμες του $|h\lambda| = |hf_y|$ πολυ αργα με αυξανομενο n. Π.χ. για $h\lambda = -20$ ο πρωτος ορος αποτελειται απο τις δυναμεις του αριθμου $-19/21 \approx -0.9$, πραγμα που σημαινει οτι σε καθε επαναληψη ο ορος αυτος γινεται 90% της προηγουμενης του τιμης.

Μιαν πολυ ταχυτερη αποσβεση του μεταβατικου ορου εχει η λεγομενη "Αντιστροφη Μεθοδος Euler" (Backward Euler Method):

$$Y_{n+1} = Y_n + hf(x_{n+1}, Y_{n+1}) \qquad (12.118)$$

που βασίζεται στην προσεγγιση:

$$y_{n+1} = y_n + \int_{x_n}^{x_{n+1}} f(t, y(t))dt = y_n + hf(x_{n+1}, y_{n+1}) + h^2 y''(\xi_n) \qquad (12.119)$$

Αυτή προκυπτει αν προσεγγισουμε την f(t,y(t)) με τους δυο πρωτους ορους του αναπτυγματος της κατα Taylor στο σημειο x_{n+1}.

Αφαιρωντας απο την (12.119) την:

$$Y_{n+1} = Y_n + hf(x_{n+1}, Y_{n+1}) \pm \rho \qquad (12.120)$$

που προκυπτει αν στην (12.118) λαβουμε υπ' οψη και το σφαλμα στρογγυλευσεως, εχουμε για $f_y \approx \lambda$=σταθ.:

$$E_{n+1} = E_n + h\lambda E_{n+1} + Ah^2 \pm \rho \qquad (12.121)$$

Η σχεση αυτη εχει την στασιμη λυση $E_\infty = E$, οπου

$$E = E + h\lambda E + Ah^2 \pm \rho \qquad (12.122)$$

δηλαδη

$$E = \frac{A}{\lambda} h \pm \rho \qquad (12.123)$$

Αφαιρεση της (12.122) απο την (12.121) μας δινει επομενως:

$$E_{n+1}-E = E_n-E + h\lambda (E_{n+1}-E) \qquad (12.124)$$

δηλαδη

$$E_{n+1} - E = \left(\frac{1}{1+h\lambda}\right)(E_n - E) = \left(\frac{1}{1+h\lambda}\right)^n (E_1 - E) \qquad (12.125)$$

Το σφαλμα E_{n+1} τεινει επομενως προς την στασιμη τιμη E αν είναι

$$|1 - h\lambda| > 1 \qquad (12.126)$$

δηλαδη για ολα τα $h\lambda = hf_y$ που ειναι μικροτερα του μηδενος ή εχουν πραγματικο μερος μικροτερο του μηδενος ($Re\{hf\} < 0$).

Αντιθετα με το σφαλμα της μεθοδου τραπεζιου (12.115), το σφαλμα της αντιστροφης μεθοδου Euler τεινει για μεγαλες (απολυτες) τιμες του hλ ταχυτατα στην στασιμη τιμη του. Αυτο σημαινει οτι η αντιστροφη μεθοδος Euler μπορει να χρησιμοποιηθει για δυσκαμπτες Δ.Ε. αφου μπορουμε να μεταβαλλουμε το βημα h χωρις δυσκολιες. Μειονεκτημα της μεθοδου ειναι βεβαια οτι το ολικο σφαλμα ειναι στην ταξη μεγεθους του h.

Για παραδείγματα δύσκαμπτων Δ.Ε. και για μεθόδους ανωτέρας τάξεως για τον προσδιορισμό της λύσης τους βλέπε [Κιουστελίδη, 1996, σελ. 337-339].

Ασκήσεις

Να προσδιορισθει προσεγγιστικα η λυση των επομενων προβληματων αρχικων τιμων στο διαστημα I που δινεται με μιαν απο τις παραπανω αριθμητικες μεθοδους και να γινει συγκριση των αριθμητικων προσεγγισεων με την ακριβη λυση, αν αυτη δινεται:

1. $y' = y^2 + 1$, $y(0) = 0$, $I = [0, 1.5]$ ($y(x) = \tan x$)
2. $y' = -y + ty^2$, $y(0) = 1$, $I = [0, 1]$
3. $y' = \cos^2 y$, $y(0) = 0$, $I = [0, 2]$ ($y(x) = \arctan x$)

Για τις παρακατω ασκησεις βλεπε Kopchenova - Maron: Computational Mathematics, MIR, σελ.226 και μετα. Μια ενδεικτικη τιμη βηματος (που πρεπει ενδεχομενα να τροποποιηθει με βαση τις συνθηκες ευσταθειας) ειναι h = 0.1. Οι παραμετροι a,b , οπου εμφανιζονται εχουν τις τιμες:

$a = 1 + 0.4n$, $b = 1 + 0.8k$, $n,k = 0(1)5$.

4. $y' = - \sinh(y/2 + x) + y/2$, $y(0) = 0$, $I = [0, 0.5]$
5. $y' = y/x- y^2$, $y(1) = 1$, $I = [1, 2]$
6. $y' = \cos(ax + y) + x - y$, $y(0) = 0$.
7. $y' = \cos x/(a+y) + y^2$, $y(0) = 0$
8. $y' = 1 - \sin(ax + y) + by/(2+x)$, $y(0) = 0$

Στα επομενα συστήματα ειναι $a = 2 + 0.5n$, $b = 2 + 0.5k$, $n,k=0(1)5$.

9. $$x' = \sqrt{t^2 + ax^2} + y,\ x(0) = 0.5$$
$$y' = \cos(by) + x, \qquad y(0) = 1$$

10. $$x' = \cos(x + ay) + b,\ x(0) = 1$$
$$y' = \frac{a}{t + bx^2} + t + 1,\ y(0) = 0.05$$

Εφαρμογή

Χρησιμοποιήσετε την εντολή ODE45 της MATLAB για να βρείτε την μορφή της πεπλεγμένης συνάρτησης $y^x x^y = a$. Δείξετε ότι αν πάρουμε τον λογάριθμο στα δύο μέλη αυτής της ισότητας και παραγωγίσουμε ως προς x καταλήγουμε στο πρόβλημα αρχικών τιμών:

$$y' = -\frac{\ln y + y/x}{\ln x + x/y}, \quad y(1) = a$$

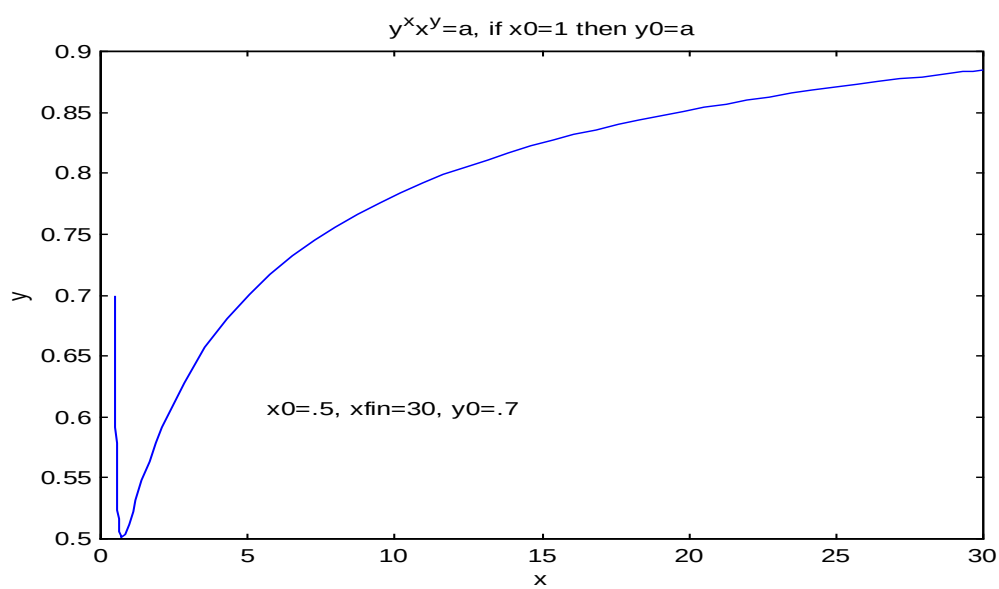
Μελετήστε την μορφή της καμπύλης λύσης αυτού του προβλήματος για διάφορες τιμές του a κάνοντας χρήση των προγραμμάτων difeq και diffunc που ακολουθούν:

difeq.m

```
xfin=input('final x=');
x0=input('x0=');
y0=input('y0=');
[x,y]=ode45('diffunc',[x0,xfin],y0);
plot(x,y);xlabel('x');ylabel('y');
title('y^xx^y=a, if x0=1 then y0=a');
```

diffunc.m

```
function yton=diffunc(x,y)
yton=-(log(y)+y./x)./(log(x)+x./y);
```



13. ΤΟ ΠΡΟΒΛΗΜΑ ΣΥΝΟΡΙΑΚΩΝ ΤΙΜΩΝ ΣΕ ΚΟΙΝΕΣ ΔΙΑΦΟΡΙΚΕΣ ΕΞΙΣΩΣΕΙΣ

Παραδείγματα Φυσικών και Τεχνικών Εφαρμογών

1. Οι διαμηκεις ταλαντωσεις ενος οριζοντιου προβολου με μηκος L, πυκνοτητα ρ, συντελεστη ελαστικοτητας E και επιφανεια διατομης F(x), περιγραφονται απο την εξισωση:

$$-E(Fy')' = \omega^2 \rho F y$$

με συνοριακες συνθηκες

$$y'(0) = 0,\ y(L) = 0$$

οπου x ειναι η αποσταση απο το ελευθερο ακρο του προβολου, και y ειναι η καθετη μετατοπιση.

Ζητειται να προσδιορισθουν οι φυσικες συχνοτητες της, ω, δηλ. οι ιδιοσυχνοτητες του συστηματος.

Με

$$F(x) = F_0(1 + x/L),\ \lambda := \omega^2 \rho / E,\ L = 1$$

προκυπτει ετσι το προβλημα:

$$-(1 + x)y'' - y' = -[(1 + x)y']' = \lambda(1 + x)y,\ y'(0) = 0,\ y(1) = 0$$

(Βλεπε L.Collatz: The Numerical Treatment of Differential Equations, Springer, 1966, σελ. 147-148).

2. Να ευρεθει η στασιμη κατανομη θερμοκρασιων σε μιαν ομογενη ραβδο μηκους L στην οποια παραγεται θερμοτητα με ρυθμο f(y) ανα μοναδα χρονου και ανα μοναδα μηκους, π.χ. σαν συνεπεια καποιας χημικης αντιδρασης, οπου τα ακρα της ραβδου x=0, x=L, διατηρουνται σε σταθερες θερμοκρασιες. Αν η f(y) εκφραζει την διαφορα θερμοκρασιας της ραβδου απο το περιβαλλον της και το x μετριεται κατα μηκος της ραβδου, εχουμε τοτε το μη γραμμικο προβλημα συνοριακων τιμων:

$$y'' = -c\, f(y),\ y(0) = y_0,\ y(L) = y_1$$

οπου c δοσμενη σταθερα. Π.χ.,

$$c\, f(y) = 1 + e^y,\ y(0) = 0,\ y(1) = 1$$

13.1 Οι Βασικές Μέθοδοι

Το προβλημα τουτο προκυπτει στην περιπτωση Δ.Ε. δευτερης η ανωτερης ταξης αν ειναι δοσμενη η τιμη της λυσης η και των παραγωγων της σε περισσοτερα απο ενα σημεια. Ζητειται δηλαδη, π.χ., να προσεγγισθει η λυση y(x) της Δ.Ε.:

$$y'' = f(x, y, y'),\ x \in [a, b] \qquad (13.1)$$

υπο τις συνοριακες συνθηκες:

$$y(a) = y_a,\ y(b) = y_b \qquad (13.2)$$

Για την λυση ενος τετοιου προβληματος χρησιμοποιουνται τρεις κυριως κατηγοριες μεθοδων:

α) Η Μεθοδος Βολης

β) Η Μεθοδος Πεπερασμενων Διαφορων

γ) Η Μεθοδος πεπερασμενων Στοιχειων.

Εδω θα ασχοληθουμε κυριως με τις δυο τελευταιες γιατι μπορουν να εφαρμοσθουν και σε πολυδιαστατα προβληματα, δηλαδη σε προβληματα συνοριακων συνθηκων για Μερικες Διαφορικες Εξισωσεις.

Την πρωτη μεθοδο θα την περιγραψουμε μονο συντομα.

13.2 Η Μέθοδος Βολής

Η μεθοδος βολης αναγει το προβλημα συνοριακων τιμων σε προβλημα αρχικων τιμων ωστε να μπορουν να εφαρμοσθουν οι αντιστοιχες μεθοδοι αριθμητικης επιλυσεως.

Στην περιπτωση του προβληματος (13.1), (13.2) τουτο θα σημαινε την αντικατασταση της συνθηκης y(b) = y με μιαν συνθηκη της μορφης

$$y'(a) = c \qquad (13.3)$$

Επειδη η λυση του προβληματος αρχικων τιμων ειναι μονοσημαντα ωρισμενη για f και f φραγμενα, μπορουμε να θεωρησουμε την λυση του προβληματος αρχικων τιμων (13.1), με $y(a) = y_a$, $y'(a) = c$, ως συναρτηση του c, y(x,c), και να προσδιορισουμε το καταλληλο c λυνοντας την εξισωση:

$$f(c) := y(b,c) - y = 0 \qquad (13.4)$$

Αυτο γινεται επαναληπτικα, π.χ. ,με την μεθοδο τεμνουσας:

Για δυο αρχικες κλισεις c_0, c_1 προσδιοριζουμε τις τιμες $f(c_0) := y(b,c_0) - y_0$, $f(c_1) := y(b,c_1) - y_1$ λυνοντας τα αντιστοιχα προβληματα αρχικων τιμων:

$$y'' = f(x,y,y'),\ y(a) = y_a,\ y'(a) = c_i,\ i = 0, 1 \qquad (13.5)$$

και προσδιοριζουμε το σημειο

$$c_2 = c_0 - f(c_0)\frac{c_1 - c_0}{f(c_1) - f(c_0)} \qquad (13.6)$$

οπου η ευθεια που περναει απο τα σημεια $(c_i, f(c_i))$, i = 0, 1 τεμνει τον οριζοντιο αξονα. Η διαδικασια αυτη μπορει να επαναληφθει μεχρι την τελικη συγκλιση της ακολουθιας των c_i.

Στην περιπτωση που η Δ.Ε. (13.1) ειναι γραμμικη ως προς y, δηλ. της μορφης:

$$y'' - p(x)y' - q(x)y = r(x) \qquad (13.7)$$

η τιμη (13.6) ειναι ηδη η ζητουμενη λυση της εξισωσης (13.4), γιατι η y(x,c) εξαρταται γραμμικα απο το c, ισχυει δηλαδη η αρχη της επαλληλιας.

13.2 Η Μέθοδος Πεπερασμένων Διαφορών

Η μεθοδος αυτη επιλεγει ισαπεχοντα σημεια στο διαστημα [a,b]:

$$x = a + k\,h,\ h = \frac{b-a}{n},\ k = 0(1)n \qquad (13.8)$$

οποτε $x_0 = a$, $x_n = b$, και αναζητει τιμες Y_k που να ικανοποιουν προσεγγιστικα την Δ.Ε. (13.1) στα σημεια x_k :

$$y_k'' = f(x_k, y_k, y_k'),\ k = 1(1)n-1 \qquad (13.9)$$

καθως και τις συνοριακες συνθηκες.

Αυτο γινεται με χρηση στην (13.9) τυπων προσεγγιστικης διαφορισης οπως οι:

$$y_k' = y'(x_k) = \frac{y_{k+1} - y_{k-1}}{2h} - \frac{h^2}{6} y^{(3)}(\xi_k) \qquad (13.10)$$

οπου $\xi_k \in (x_k - h, x_k + h)$ και

$$y_k'' = y''(x_k) = \frac{y_{k+1} - 2y_k + y_{k-1}}{h^2} - \frac{h^2}{12} y^{(4)}(\zeta_k) \qquad (13.11)$$

με $\zeta_k \in (x_k - h, x_k + h)$, που ισχυουν αντιστοιχα αν μεχρι και η $y^{(3)}$ ή η $y^{(4)}$ ειναι συνεχεις.

Ετσι προκυπτει απο την (13.9):

$$\frac{y_{k+1} - 2y_k + y_{k-1}}{h^2} = f(x_k, y_k, \frac{y_{k+1} - y_{k-1}}{2h}) + O(h^2) \qquad (13.12)$$

και με παραλειψη του ορου σφαλματος:

$$\frac{Y_{k+1} - 2Y_k + Y_{k-1}}{h^2} = f(x_k, Y_k, \frac{Y_{k+1} - Y_{k-1}}{2h}),\ k = 1(1)n-1 \qquad (13.13)$$

οπου με βαση π.χ.τις συνοριακες συνθηκες (13.2) ειναι

$$Y_0 = y_a\ ,\ Y_n = y_b \qquad (13.14)$$

Ετσι εχουμε ενα γενικα *μη γραμμικό* συστημα n-1 εξισωσεων με τους n-1 αγνωστους Y_k, k = 1(1)n-1.

Για την λυση αυτου του συστηματος αποδεικνυεται οτι, αν η συναρτηση f(x, y, y') εχει συνεχεις μερικες παραγωγους ως προς y και y', και η y εχει μεχρι και την τεταρτη παραγωγο της συνεχη τοτε ισχυει

$$y_k = Y_k + O(h^2) \qquad (13.15)$$

Η μεθοδος αυτη αφαρμοζεται συνηθως στην περιπτωση γραμμικων Δ.Ε. οπως η (13.7), οποτε το το προς επιλυση συστημα (13.13) ειναι γραμμικο και παιρνει την μορφη:

$$\left(1 + \frac{h}{2} p_k\right) Y_{k-1} - \left(2 + h^2 q_k\right) Y_k + \left(1 - \frac{h}{2} p_k\right) Y_{k+1} = h^2 r_k \qquad (13.16)$$

οπου $p_k := p(x_k)$, $q_k := q(x_k)$, $r_k := (x_k)$, k = 1(1)n-1.

Το συστημα αυτο εχει τριδιαγωνιο πινακα, δηλαδη πινακα οπου μονο η κυρια διαγωνιος και οι δυο γειτονικες της διαγωνιοι εχουν μη μηδενικα στοιχεια και μπορει να λυθει π.χ. με την μεθοδο απαλοιφης Gauss.

Η μονοσημαντοτητα της λυσης του συστηματος αυτου ειναι εγγυημενη αν η συναρτηση q(x) ειναι *θετική* στο διαστημα [a, b]

$$q(x) \geq Q > 0 \qquad (13.17)$$

και το βημα h ειναι τοσο μικρο ωστε να ισχυει:

$$h|p_k|/ 2 < 1,\ k = 1(1)n-1 \qquad (13.18)$$

οποτε προκυπτει

$$2 + h^2 q_k > |1 + hp_k/2| + |1 - hp_k/2| = 2\ ,\ k = 1(1)n-1 \qquad (13.19)$$

Το συστημα εχει τοτε, δηλαδη, υπερισχυουσα διαγωνιο, και μπορει να λυθει με επαναληπτικες μεθοδους, οπως η μεθοδος Gauss - Jordan ή η μεθοδος Gauss - Seidel.

Η αποδειξη της (13.15) ειναι εδω απλη. Αν παρουμε την (13.7) για $x = x_k$ και χρησιμοποιησουμε τις (13.10),(13.11) εχουμε:

$$\left(1+\frac{h}{2}p_k\right)y_{k-1}-\left(2+h^2q_k\right)y_k+\left(1-\frac{h}{2}p_k\right)y_{k+1}=h^2(r_k+\tau_k) \qquad (13.20)$$

οπου

$$\tau_k=\frac{h^2}{12}y^{(4)}(\xi_k)+p_k\frac{h^2}{6}y^{(3)}(\zeta_k)=O(h^2)$$

και αφαιρωντας την (13.16) απο αυτην:

$$\left(1+\frac{h}{2}p_k\right)\varepsilon_{k-1}-\left(2+h^2q_k\right)\varepsilon_k+\left(1-\frac{h}{2}p_k\right)\varepsilon_{k+1}=h^2\tau_k \qquad (13.21)$$

οπου

$$\varepsilon_k := y_k - Y_k \ , \ k = 0(1)n \qquad (13.22)$$

Με

$$\varepsilon := m\ a\ x\ |\varepsilon_k| \qquad (13.23)$$

εχουμε τοτε:

$$\left(2+h^2q_k\right)\varepsilon_k=\left(1+\frac{h}{2}p_k\right)\varepsilon_{k-1}+\left(1+\frac{h}{2}p_k\right)\varepsilon_{k+1}-h^2\tau_k\le 2\varepsilon-h^2\tau_k$$

και επομενως:

$$(2+h^2Q)\varepsilon\le 2\varepsilon-h^2\tau_k \qquad (13.24)$$

ή

$$|\varepsilon_k|\le\varepsilon\le\frac{h^2}{2Q}\left(\frac{M_4}{12}+P\frac{M_3}{6}\right) \qquad (13.25)$$

οπου M_4 ανω φραγμα της $|y^{(4)}(x)|$, M_3 ανω φραγμα της $|y^{(3)}(x)|$, και P ανω φραγμα της $|p(x)|$ στο διαστημα [a, b].

Εδω βλεπουμε και γιατι η προυποθεση (13.17) ειναι απαραιτητη. Οσο πιο μικρο ειναι το κατω φραγμα Q της q(x) τοσο χειροτερη ειναι η εκτιμηση σφαλματος (13.25).

<u>Προσοχή</u>: Εξ αιτίας των σφαλμάτων στρογγυλεύσεως ο τύπος (13.10) περιέχει ένα πρόσθετο σφάλμα $\pm\rho/h$ (βλ. (10.19)-(10.22)). Επομένως η σχέση $|y_k - Y_k| = O(h^2)$ ισχύει μόνο όταν $4\rho/h^2 << h^2$, δηλ. $4\rho << h^4$.

Μικτές Συνοριακές Συνθήκες

Οπως ειπαμε και στην αρχη, συχνα οι συνοριακες συνθηκες δεν προδιαγραφουν απλα τις τιμες της λυσης στα δυο συνοριακα σημεια, αλλα τις τιμες, π.χ., γραμμικων συνδυασμων της λυσης και της πρωτης παραγωγου της. Ενα τετοιο προβλημα ειναι, π.χ., το να προσδιορισουμε την λυση της Δ.Ε. (16.7) υπο τις συνοριακες συνθηκες:

$$y'(a) + cy(a) = d,\ y(b) = y \qquad (13.26)$$

Στην περιπτωση αυτη μπορουμε να εφαρμοσουμε την ιδια τεχνικη οπως παραπανω, μονο που θα πρεπει να προσεγγισουμε και την τιμη y'(a) με πηλικο διαφορων. Εδω ομως χρειαζεται προσοχη, γιατι αν χρησιμοποιησουμε την προσεγγιση:

$$y'(a) = y'(x_0) = \frac{y(x_0+h) - y(x_0)}{h} + O(h) = \frac{y_1 - y_0}{h} + O(h)$$

(13.27)

προκυπτει μια προσεγγιση με σφαλμα O(h) και οχι O(h^2), εστω και αν σε ολα τα αλλα σημεια χρησιμοποιησουμε τους τυπους (13.10), (13.11).

Σε τετοιες περιπτωσεις προεκτεινουμε την λυση τεχνητα *έξω* απο το διαστημα [a, b], εισαγοντας εναν ορο Y_{-1} = y(a-h), ωστε να μπορεσουμε να εφαρμοσουμε τον τυπο (13.10) και για k = 0.

Οι συνοριακες συνθηκες παιρνουν λοιπον την μορφη:

$$\frac{Y_1 - Y_{-1}}{2h} + cY_0 = d, Y_n = y_b \qquad (13.28)$$

Το συστημα (13.16), (13.28) εχει ομως τωρα εναν αγνωστο παραπανω απο τον αριθμο των εξισωσεων του. Για τουτο το συμπληρωνουμε με την αντιστοιχη εξισωση (13.16) για k = 0:

$$\left(1 + \frac{h}{2}p_0\right)Y_{-1} - \left(2 + h^2 q_0\right)Y_0 + \left(1 - \frac{h}{2}p_0\right)Y_1 = h^2 r_0 \qquad (13.29)$$

Το παραπανω τεχνασμα μπορει να εφαρμοσθει τοσο στο ενα, οσο και στο αλλο συνοριακο σημειο οταν εχουμε μονοδιαστατα προβληματα, δηλαδη κοινες Δ.Ε.

Παραδείγματα Εφαρμογής της Μεθόδου Πεπερασμένων Διαφορών

1. Να λυθει το προβλημα συνοριακων τιμων:

$$y'' - (1 + x)y' + x^2y = 1 + x^2 , y'(0) - y(0) = 1, y'(1) = 2$$

με την μεθοδο πεπερασμενων διαφορων.

Λύση

Αν στον τυχοντα εσωτερικο κομβο x_k αντικαταστησουμε τις παραγωγους με τα πηλικα διαφορων (13.10), (13.11) εχουμε:

$$\frac{Y_{k+1} - 2Y_k + Y_{k-1}}{h^2} - (1 + x_k)\frac{Y_{k+1} - Y_{k-1}}{2h} + x_k{}^2 Y_k = 1 + x_k{}^2, k = 1(1)n-1$$

οπου $x_0 = 0$, $x_k = k/n$, $x_n = 1$, δηλαδή h = 1/n.

Επειδη και οι δυο συνοριακες συνθηκες περιλαμβανουν παραγωγους πρεπει εδω να εισαγαγουμε δυο ακομα εξωτερικους κομβους:

$$x_{-1} = -h, x_{n+1} = 1 + h$$

οποτε με χρηση πεπερασμενων διαφορων οι συνοριακες συνθηκες γινονται:

$$\frac{Y_1 - Y_{-1}}{2h} - Y_0 = 1, \quad \frac{Y_{n+1} - Y_{n-1}}{2h} = 2$$

Συνολικα εχουμε ετσι n-1 + 2 = n+1 εξισωσεις με n+3 αγνωστους, τους Y_{-1}, Y_0, Y_1, ..., Y_n, Y_{n+1}. Χρειαζομαστε δηλαδη δυο ακομα εξισωσεις. Για τουτο χρησιμοποιουμε την πρωτη διαφοροεξισωση και για k = 0, k = n.

Ειδικα για n = 2, h = 0.5 εχουμε ετσι το συστημα:

$$Y_1 - Y_{-1} - Y_0 = 1$$

k=0: $\quad 4(Y_1 - 2Y_0 + Y_{-1}) - (Y_1 - Y_{-1}) = 1$

k=1: $4(Y_2 - 2Y_1 + Y_0) - 1.5\,(Y_2 - Y_0) + 0.25\,Y_1 = 1.25$
k=2: $4(Y_3 - 2Y_2 + Y_1) - 2(Y_3 - Y_1) + Y_2 = 2$
$Y_3 - Y_1 = 2$

2. Να λυθει το *μη γραμμικό* προβλημα συνοριακων τιμων:

$$y'' - x(y')^2 + 4x^2y = 2x^2+1\ ,\ \ y(0) = 1,\ y(1) = 4$$

με την μεθοδο πεπερασμενων διαφορων.

Λύση

Αντικατάσταση των παραγώγων με πηλίκα διαφορών μας δίνει εδώ το σύστημα μη γραμμικών εξισώσεων:

$$\frac{Y_{k+1} - 2Y_k + Y_{k-1}}{h^2} - x_k\left(\frac{Y_{k+1} - Y_{k-1}}{2h}\right)^2 + 4x_k{}^2Y_k = 2x_k{}^2 + 1,\ k = 1(1)n-1$$

όπου $Y_0=1$, $Y_n=4$. Το σύστημα αυτό μπορεί να λυθεί με κάποιαν επαναληπτική μέθοδο για μη γραμμικά συστήματα.

13.3 Η Μέθοδος Πεπερασμένων Στοιχείων

Η όλη διαδικασία της αντικατάστασης των παραγώγων με πηλίκα διαφορών γίνεται εξαιρετικά δυσχερής σε πολυδιάστατα προβλήματα, δηλ. σε Δ.Ε. με μερικές παραγώγους που πρέπει να ισχύουν, π.χ., στο εσωτερικό ενός διδιάστατου χωρίου ενώ είναι δοσμένες, οι τιμές της ζητούμενης συνάρτησης πάνω στα σημεία της συνοριακής καμπύλης του χωρίου αυτού.

Άμεσα μπορούν να γενικευθουν οσα ειπαμε παραπανω μονον οταν το χωριο εχει σχημα ορθογωνιο, γιατί σε αυτό μπορούν να επιλεγούν με ομοιόμορφο τρόπο ισαπέχοντα σημεία. Αν το σύνορο του χωρίου είναι ακανόνιστο, τοτε για την προσεγγιση των παραγωγων κοντα στο συνορο ενδεχεται να χρειασθουμε σημεια που βρισκονται εξω απο αυτό, όπως έγινε ήδη παραπάνω στην μονοδιάστατη περίπτωση, και μαλιστα μεγάλο πλήθος τέτοιων σημείων. Αυτο δυσχεραινει πολυ την αυτοματοποιηση της ολης διαδικασιας.

Αυτος ειναι ο λογος για τον οποιο κατα τις τελευταιες δεκαετιες εχει επικρατησει η χρηση της μεθοδου Πεπερασμενων Στοιχειων για ολα τα πολυδιαστατα προβληματα. Η μεθοδος αυτη αξιοποιει αμεσα και χωρις δυσκολιες τις συνοριακες συνθηκες, ειτε ειναι απλες, ειτε συνθετες, αλλα οδηγει σε περιπλοκώτερα γραμμικα συστηματα (αν η Δ.Ε. ειναι γραμμικη) που οι συντελεστες τους πρεπει να προσδιορισθουν με αριθμητικες ολοκληρωσεις.

Εδω θα περιγραφει ο τροπος εφαρμογης της μεθοδου αυτης μονο σε μονοδιαστατα προβληματα, οπου στην πραξη δεν εχει κανενα πλεονεκτημα. Στοχος μας ειναι ομως, οπως και σε πολλα αλλα σημεια αυτου του βιβλιου, να διευκολυνθει η κατανοηση των βασικων χαρακτηριστικων της μεθοδου αυτης χωρις την επιβαρυνση απο προσθετες τεχνικες λεπτομερειες που επιβαλλει η αντιμετωπιση πολυδιαστατων προβληματων. Ετσι οποιος ενδιαφερεται, θα μπορει να την εφαρμοσει και σε πολυδιαστατα προβληματα διαβαζοντας με μεγαλυτερη ανεση την εκτενεστατη βιβλιογραφια που υπαρχει γι' αυτα.

Πιο συγκεκριμενα, θα μελετησουμε εδω την λυση του προβληματος προσεγγισης της συναρτησης $y(x)$ ($y \in C\,[a, b]$) που ικανοποιει τις σχεσεις:

$$y''(x) + p(x)y'(x) + q(x)y(x) = r(x) \text{ για } x \in (a, b) \qquad (13.30)$$

και

$$y(a) = y_a \,,\; y(b) = y_b \qquad (13.31)$$

ή εναλλακτικα,

$$y(a) = y_a,\; y'(b) + c\, y(b) = d \qquad (13.32)$$

οπου p, q, r $\in$ C[a, b].

Η προσεγγιση που ζητουμε βασιζεται σε μιαν *διαμέριση* $\{a = x_0 < x_1 < \ldots < x_n = b\}$ του διαστηματος [a, b] και εχει την μορφη:

$$Y(x) = \sum_{k=0}^{n} Y_k \varphi_k(x) \qquad (13.33)$$

οπου Y_k προσεγγιση της τιμης $y_k = y(x_k)$ και $\varphi_k(x)$ μια *μη αρνητική* συναρτηση με μεγιστη τιμη 1, που βρισκεται στην θεση x_k και μηδενιζεται σε ολο το διαστημα [a, b] εκτος απο τα γειτονικά υποδιαστήματα $[x_{k-1}, x_k)$ και $[x_k, x_{k+1})$, που εδώ ονομάζονται «στοιχεία» της διαμέρισης. Έτσι θα είναι $Y(x_k) = Y_k$, για k = 0(1)n.

Παραδειγματος χαριν, μπορουμε να βαλουμε:

$$\varphi_k(x) = \left[\left(\frac{x - x_{k-1}}{x_k - x_{k-1}}\right)\left(\frac{x_{k+1} - x}{x_{k+1} - x_k}\right)\right]^{\alpha}, x \in [x_{k-1}, x_{k+1}] \qquad (13.34)$$

και $\varphi_k(x) = 0$, για $x \notin [x_{k-1}, x_{k+1}]$, με α = 1, αν θελουμε οι συναρτήσεις $\varphi_k(x)$ (άρα και η Y(x)) να εχουν τμηματικα συνεχη πρωτη παραγωγο, ή με α = 2, αν θελουμε οι $\varphi_k(x)$, Y(x) να εχουν συνεχη ως και την δευτερη παραγωγο.

Στην πραξη ομως προτιμουμε τις "τριγωνικες", η "πυραμιδοειδεις" συναρτησεις $\varphi_k(x) := N_k(x)$ οπου:

$$N_k(x) = \begin{cases} \dfrac{x - x_{k-1}}{x_k - x_{k-1}}, x \in [x_{k-1}, x_k) \\ \dfrac{x_{k+1} - x}{x_{k+1} - x_k}, x \in [x_k, x_{k+1}) \end{cases} \qquad (13.35)$$

και $N_k(x) = 0$, για $x \notin [x_{k-1}, x_{k+1})$, που δεν εχουν παρα μονον τμηματικα συνεχη πρωτη παραγωγο.

Ο λογος της προτιμησεως αυτης ειναι οτι μειωνεται σημαντικα ο ογκος των πραξεων, αφου οι συναρτησεις $N_k(x)$ ειναι τμηματικώς πρωτοβάθμια πολυώνυμα (αντι δευτεροβάθμια ή τριτοβάθμια).

Για να προσδιορισουμε τους αγνωστους Y_k (που προσεγγιζουν τις τιμες y_k) ξεκιναμε απο την σχεση:

$$Dy(x) - r(x) = 0,\; x \in (a, b) \qquad (13.36)$$

οπου

$$Dy := y'' + py' + qy \qquad (13.37)$$

Αν στην θεση της y(x) βαλουμε την προσεγγιση Y(x), δεν μπορουμε βεβαια να περιμενουμε οτι η συναρτηση DY(x) - r(x) θα ειναι εν γενει μηδεν για $x \in (a, b)$, δηλ. για απειρες τιμες του x, γιατι οι αγνωστοι ειναι πεπερασμενοι (τα Y_k). Μπορουμε ομως να αναγκασουμε την συναρτηση DY - r να εχει εν γενει τιμες γυρω απο το μηδεν.

Προς τουτο αρκει να απαιτησουμε να ειναι:

$$\int_a^b \left(DY(x) - r(x)\right)\varphi_j(x)dx = 0,\ j = 1(1)n-1 \qquad (13.38)$$

Η απαιτηση αυτη λεγεται απαιτηση Galerkin.

Επειδη τα $\varphi_j(x)$ μηδενιζονται εξω απο το διαστημα $[x_{j-1}, x_{j+1}]$ οι σχεσεις αυτες σημαινουν οτι εξαναγκαζουμε το εμβαδον της συναρτησης $(DY - r)\varphi_j$ στο διαστημα $[x_{j-1}, x_{j+1}]$ να μηδενισθει.

Επειδη, ομως, η $\varphi_j(x)$ εχει στο (x_{j-1}, x_{j+1}) μονο θετικες τιμες, τουτο σημαινει οτι η $DY(x) - r(x)$ πρεπει να εχει εκει τοσο θετικες, οσο και αρνητικες τιμες και μαλιστα ετσι ωστε να μηδενιζεται το εμβαδον της συνάρτησης $(DY - r)\varphi_j$.

Αν εχουν δοθει οι συνοριακες συνθηκες $y(a)=y_a$, $y(b)=y_b$, τοτε μπορουμε να ορισουμε απ' ευθειας:

$$Y_0 := y_a,\ Y_n := y_b \qquad (13.39)$$

Ετσι απομενουν οι n-1 αγνωστοι $Y_1, \ldots, Y_{n-1}$ και για τον προσδιορισμο τους αρκουν n-1 εξισωσεις. Αυτος ειναι ο λογος για τον οποιο χρησιμοποιουμε τις εξισωσεις (13.38) μονο για j = 1 ως n-1.

Αναλυτικωτερα, οι εξισωσεις αυτες εχουν την μορφη:

$$\int_a^b r(x)\varphi_j(x)dx = \int_a^b DY(x)\varphi_j(x)dx = \int_a^b \left(Y''(x) + p(x)Y'(x) + q(x)Y(x)\right)\varphi_j(x)dx$$

και επομένως,

$$\int_a^b r(x)\varphi_j(x)dx = \left[Y'(x)\varphi_j(x)\right]_a^b +$$

$$+\int_a^b \left(-Y'(x)\varphi_j'(x) + p(x)Y'(x)\varphi_j(x) + q(x)Y(x)\varphi_j(x)\right)dx \qquad (13.40)$$

οπου η τελευταια μορφη προκυπτει αν μετασχηματισουμε το ολοκληρωμα της $Y''\varphi_j = (Y')'\varphi_j$ χρησιμοποιωντας παραγοντικη ολοκληρωση.

Στην τελευταια αυτην μορφη εχουν το σημαντικο πλεονεκτημα οτι ειναι απαλλαγμενες απο δευτερην παραγωγο. Αυτο μας επιτρεπει να επιλεξουμε για τη κατασκευη της Y(x) συναρτήσεις $\varphi_k(x)$ που να μην έχουν δεύτερη παραγωγο. Ετσι αντι των σχεσεων (13.38) μπορουμε να απαιτησουμε να ισχυουν απ' ευθειας οι σχεσεις (13.40) και αντι για $\varphi_k(x)$ της μορφης (13.34) μπορουμε να χρησιμοποιησουμε φ_k της μορφης (13.35) που απλοποιουν τις ολοκληρωσεις (τα $\varphi_k'(x) = N_k'(x)$ ειναι, τότε τμηματικα σταθερες συναρτησεις).

Στην περιπτωση των συνοριακων συνθηκων (13.31) χρησιμοποιουμε, οπως ειπαμε, τις σχεσεις (13.40) μονο για j = 1(1)n-1. Αυτο μας επιτρεπει να παραλειψουμε απο ολες τον ορο

$$[Y'(x)\varphi_j(x)]^b_a = Y'(b)\varphi_j(b) - Y'(a)\varphi_j(a) \qquad (13.41)$$

γιατι για j = 1(1)n-1 ειναι $\varphi_j(a) = 0 = \varphi_j(b)$.

Αντικαθιστωντας τωρα το Y(x) με βαση την (13.33) στις σχεσεις (13.40) εχουμε για j = 1(1)n-1:

$$\int_a^b r(x)\varphi_j(x)dx = \sum_{k=0}^{n} Y_k \int_a^b \left(-\varphi_k'(x)\varphi_j'(x) + p(x)\varphi_k'(x)\varphi_j(x) + q(x)\varphi_k(x)\varphi_j(x)\right)dx$$

(13.42)

Το συστημα αυτο ειναι προφανως γραμμικο ως προς τους αγνωστους Y_k. Για να το λυσουμε πρεπει πρωτα να υπολογισουμε, ειτε ακριβως, ειτε προσεγγιστικα ολα τα παραπανω ολοκληρωματα και να λαβουμε υπ'οψη οτι $Y_0 := y_a$, $Y_n := y_b$. Αυτο σημαινει οτι, για να γινει διαχωρισμος αγνωστων απο γνωστα μεγεθη, οι προσθετεοι με συντελεστη Y_0 και Y_n στο παραπανω αθροισμα πρεπει να μεταφερθουν στο πρωτο μερος των εξισωσεων.

Στην περιπτωση (13.32), οπου μας εχει δοθει μια "μικτη" συνοριακη συνθηκη, y'(b)+cy(b)=d, το Y_n δεν μας ειναι αμεσα γνωστο. Ετσι οι εξισωσεις (13.40) ή (13.42) με j = 1(1)n-1 *δεν* αρκουν.

Εδω ενδεικνυται να χρησιμοποιησουμε την (13.40) και για j = n.

Τοτε εχουμε την προσθετη σχεση:

$$\int_a^b r(x)\varphi_n(x)dx = Y(b) + \int_a^b \left(-Y'(x)\varphi_n'(x) + p(x)Y'(x)\varphi_n(x) + q(x)Y(x)\varphi_n(x)\right)dx$$

γιατι $\varphi_n(b) = 1$. Στην θεση του Y'(b) βαζουμε τωρα $d-cY(b) = d-cY_n$ με βαση την σχεση y'(b) = d - c y(b) που προκυπτει απο την μικτη συνοριακη συνθηκη. Απαιτουμε δηλαδη να ισχυει και Y'(b) = d-cY(b).

Ετσι εχουμε μιαν προσθετη εξισωση για τον προσδιορισμο του προσθετου αγνωστου Y_n.

Παράδειγμα Εφαρμογής της Μεθόδου Πεπερασμένων Στοιχείων

Να λυθει το ιδιο οπως προηγουμενα προβλημα συνοριακων τιμων:

$$y'' - (1 + x)y' + x^2y = 1 + x^2, \quad y'(0) - y(0) = 1, \quad y'(1) = 2$$

με την μεθοδο πεπερασμενων στοιχειων.

Λύση

Αν διαλεξουμε και εδω h = 0.5, δηλαδη τους κομβους $x_0=0$, $x_1=0.5$, $x_2=1$, εχουμε την προσεγγιση:

$$y(x) \approx Y(x) = \sum_{k=0}^{2} Y_k\varphi_k(x) = Y_0\varphi_0(x) + Y_1\varphi_1(x) + Y_2\varphi_2(x)$$

Η απαιτηση Galerkin παιρνει εδω την μορφη:

$$\int_0^1 (1+x^2)\varphi_j(x)dx = \int_0^1 \left(Y''(x) - (1+x)Y'(x) + x^2Y(x)\right)\varphi_j(x)dx$$

και επομένως,

$$\int_0^1 (1+x^2)\varphi_j(x)dx = \left[Y'(x)\varphi_j(x)\right]_0^1 + \int_0^1 \left(-Y'(x)\varphi_j'(x) - (1+x)Y'(x)\varphi_j(x) + x^2Y(x)\varphi_j(x)\right)dx$$

όπου

$$\varphi_k(x) = N_k(x) = \begin{cases} \dfrac{x - x_{k-1}}{h}, & x \in [x_{k-1}, x_k) \\ \dfrac{x_{k+1} - x}{h}, & x \in [x_k, x_{k+1}) \end{cases}$$

και $\varphi_k(x) = 0$, για $x \notin (x_{k-1}, x_{k+1})$.

Έτσι έχουμε τις εξισώσεις:

$$\int_0^1 (1+x^2) N_0(x) dx = -(1+Y_0) + \int_0^1 \left(-Y'(x)N_0'(x) - (1+x)Y'(x)N_0(x) + x^2 Y(x) N_0(x)\right) dx$$

$$\int_0^1 (1+x^2) N_1(x) dx = \int_0^1 \left(-Y'(x)N_1'(x) - (1+x)Y'(x)N_1(x) + x^2 Y(x) N_1(x)\right) dx$$

$$\int_0^1 (1+x^2) N_2(x) dx = \int_0^1 \left(-Y'(x)N_2'(x) - (1+x)Y'(x)N_2(x) + x^2 Y(x) N_2(x)\right) dx$$

όπου στην πρώτη εξίσωση είναι

$[Y'(x)N_0(x)]^1{}_0 = - Y'(0)N_0(0) = - Y'(0) = -(1 + Y(0)) = -(1 + Y_0)$

εξ αιτιας της συνοριακης συνθηκης: $y'(0) = 1 + y(0)$, και στην τριτη εξισωση ειναι

$[Y'(x)N_2(x)]^1{}_0 = Y'(1)N_2(1) = Y'(1) = 2$

εξ αιτιας της συνοριακης συνθηκης: $y'(1) = 2$.

Εδω ειναι:

$$N_0(x) = \begin{cases} \dfrac{x_1 - x}{h} = 1 - 2x & x \in [x_0, x_1) = [0, 0.5) \\ 0 & x \in [x_1, x_2] = [0.5, 1] \end{cases}$$

$$N_1(x) = \begin{cases} \dfrac{x - x_0}{h} = 2x & x \in [x_0, x_1) = [0, 0.5) \\ \dfrac{x_2 - x}{h} = 2(1-x) & x \in [x_1, x_2) = [0.5, 1] \end{cases}$$

$$N_2(x) = \begin{cases} 0 & x \in [x_0, x_1) = [0, 0.5) \\ \dfrac{x_1 - x}{h} & x \in [x_1, x_2) = [0.5, 1] \end{cases}$$

Ασκήσεις

1. Να λυθει το προβλημα συνοριακων τιμων:

$y'' - (2 + x)y' + 3xy = 1 + 5x$, $y'(0) - 2y(0) = 1$, $y'(1) = 2$

α) με την μεθοδο πεπερασμενων στοιχειων

β) με την μεθοδο πεπερασμενων διαφορων
με χρηση των κομβων: $x_k = k/n$, $n = 0(1)n$, με $n = 3$.

2. Να λυθει το προβλημα συνοριακων τιμων:
$y'' - (2 + x^2)y' + 3x^2y = 1 + 5x^4$, $y'(1) - 2y(1) = 3$, $y(2) = 1$
με την μεθοδο πεπερασμενων διαφορων με χρηση των κομβων:
$x_k= 1 + k/n$, $n = 0(1)n$, με $n = 3$.

3. Να λυθουν με την μεθοδο πεπερασμενων διαφορων με βημα $h = 0.1$ και συνοριακες συνθηκες $y(0) = y(1) = 0$ τα ακολουθα προβληματα συνοριακων τιμων (Kopchenova-Maron σελ.264):
a. $y'' + (1+x^3)y' + (1-x^2)y = \exp(1-2x^2)$
b. $y'' + x^2y' + (1-x)y = x/(x^2+2)$
c. $y'' + y'\sin x + y = 1/(2 + \sin x)$
d. $y'' + (x^2+ 2)^{-1/2}y' + y = x$

4. Να λυθει με την μεθοδο πεπερασμενων διαφορων το προβλημα:
$$y''-(1.7953 + 0.0001 f(x))y'+ \cos(0.7x)y = 2x^2 + 2x - 4,\ y(0)=y(1)=0$$
με $f(0)=2.3$, $f(0.2)=9$, $f(0.4)=12.1$, $f(0.6)=11.5$, $f(0.8)=7.5$, $f(1) = 0$.

5. Να λυθει με την μεθοδο πεπερασμενων στοιχειων το προβλημα συνοριακων τιμων:
$$y'' -2xy' +2y = 3x^2 + x - 1,\ y(0) = 0,\ y'(1) = 1$$

6. Να λυθουν με την μεθοδο πεπερασμενων στοιχειων τα προβληματα συνοριακων τιμων:
$$y'' +x^2y' -xy = f(x),\ y(0) = 0,\ y(1) = 1$$
με (a) $f(x) = e^x$, (b) $f(x) = \sin x$, (c) $f(x) = \cos x$.

7. Να λυθεί τοσο με την μεθοδο πεπερασμενων στοιχειων οσο και με την μεθοδο πεπερασμενων διαφορων το πρόβλημα συνοριακων τιμων:
$$y'' + y' -y/x = 5x^3-3x^2+x ,\ y(0) = 0,\ y'(1) = 1$$

8. Να λυθούν με την μεθοδο πεπερασμενων διαφορων τα προβληματα:
$$y'' - y'\cos x + y \sin x = f(x) ,\ y(-\pi) = y(\pi) = 2$$
με (a) $f(x) = \sin x$, (b) $f(x) = \cos x$, (c) $f(x) = \cos 2x$

9. Σχεδιάστε με την βοήθεια του MATLAB τα γραφήματα των ακόλουθων πεπλεγμένων συναρτήσεων, δηλαδή συναρτήσεων της μορφής $f(x,y)=c$, ανάγοντάς τες σε λύσεις του προβλήματος αρχικών τιμών: $y' = -f_x/f_y$, $y(a)=y_a$ όπου $f(a, y_a) = c$:
$y\sin(y\sin x)=.5$, $y\cos x + x\sin y = 1$, $\sin(xy) + \sin(x+y) = 2/3$, $(x+y)^{xy} = 2$
$(xy)^{x+y}= 2$, $(x/y)^{x+y}= 2$, $(xy)^{x/y}= 2$, $x^{2y}+y^{3x}= 4$, $x^yy^x = 2$, $x^y+y^x=3$, $x^{1/y}y^{1/x}=2$
$x^{1/y}+y^{1/x}=2$, $ye^x+xe^y=3$, $x^{xy}+y^{x/y}=2$, $x^{\sin y}+y^{\sin x}=2$, $x^{\ln y}+y^{\ln x}=2$, $x^3y+x^2y^2+xy^3=4$

14. ΠΡΟΣΔΙΟΡΙΣΜΟΣ ΙΔΙΟΤΙΜΩΝ ΚΑΙ ΙΔΙΟΔΙΑΝΥΣΜΑΤΩΝ

Παραδείγματα Φυσικών και Τεχνικών Εφαρμογών

1. Εξεταζουμε τις διαμηκεις ταλαντωσεις ενος πρόβολου μηκους L, πυκνοτητας ρ, με συντελεστη ελαστικοτητας E και επιφανεια διατομης F(x), οπου η συντεταγμενη x μετριεται κατα μηκος του αξονα του προβολου με το σημειο μηδεν στο ελευθερο ακρο του. Η μετατοπιση y ικανοποιει τοτε την διαφορικη εξισωση:

$$-E\,F'y' = \omega\,\rho\,F\,y$$

και τις συνοριακες συνθηκες:

$$y'(0) = 0,\ y(L) = 0$$

Ζητειται ο προσδιορισμος των φυσικων συχνοτητων ω, αν η επιφανεια διατομης μεταβαλλεται γραμμικα:

$$F(x) = F_0(1 + x/L)$$

Με $\lambda := \omega^2\rho/E$ και L = 1 εχουμε τοτε το προβλημα:

$$-(1+x)y'' - y' = -[(1+x)y']' = \lambda(1+x)y\ ,\ y'(0) = 0,\ y(1) = 0$$

Αν τωρα διαλεξουμε N ισαπεχοντες κομβους x_k μεταξυ 0 και L=1 και προσεγγισουμε τις παραγωγους με πηλικα διαφορων εχουμε:

$$-(1+x_k)\frac{Y_{k-1}-2Y_k+Y_{k+1}}{h^2}-\frac{Y_{k+1}-Y_{k-1}}{2h}=\Lambda(1+x_k)Y_k$$

για k = 0(1)N-1 με h = 1/N , και $x_k = k/N$, οπου $Y_k = y(x_k)$ και $\Lambda \approx \lambda$ (Δηλ. Λ ειναι προσεγγιση μιας αντιστοιχης ιδιοτιμης ,λ, του αρχικου προβληματος).

Απο τις συνοριακες συνθηκες εχουμε αντιστοιχα: $Y_N = 0$ (επειδη y(1) = 0) και $Y_1 - Y_{-1} = 0$ (επειδη $0 = y'(0) \approx (Y_1 - Y_{-1})/2h$).

Επειδη με την τελευταια αυτη συνοριακη τιμη εισαγεται στο προβλημα ο προσθετος αγνωστος Y ειμαστε υποχρεωμενοι για να τον προσδιορισουμε να χρησιμοποιησουμε την παραπανω διαφοροεξισωση που αντιστοιχει στην διαφορικη εξισωση και για k = 0 (αλλοιως θα την παιρναμε μονο για k = 1(1)N-1). Το παραδειγμα αυτο δινεται στην σελ.147 του βιβλιου:

L.Collatz: The Numerical Treatment of Differential Equations, Springer, 1966.

2. Παρόμοια, μπορούμε να εξετάσουμε τις διαμήκεις ταλαντώσεις σε ένα σύστημα δύο ελατηρίων, που συνδέουν μεταξύ τους σώματα με μάζες m_i. Το πρώτο είναι αναρτημένο από έναν τοίχο και έχει στο δεύτερο άκρο του σώμα μάζας m_1. Το δεύτερο ελατήριο είναι συνδεδεμένο με το πρώτο σώμα και έχει στο άλλο άκρο του σώμα μάζας m_2.

Αν στο σύστημα αυτό δοθεί μία στιγμιαία διέγερση με δύναμη που εφαρμόζεται στην μάζα m_2 στην κατεύθυνση που έχουν τα ελατήρια, ζητείται να προσδιορισθεί η συχνότητα των στάσιμων ταλαντώσεων που εκτελεί το σύστημα αυτό, δηλαδή των ταλαντώσεων που διατηρούνται μακροχρόνια, όταν πάψει η επίδραση της στιγμιαία εξασκηθείσας δύναμης.

Το σύστημα αυτό μπορεί να γενικευθεί σε οποιοδήποτε σύστημα ελαστικά συνδεδεμένων μεταξύ τους σωμάτων, π.χ. τα βαγόνια ενός τραίνου ή ένα μεταλλικό ικρίωμα. Στην προκειμένη περίπτωση, αν ονομάσουμε y_1, y_2 τις μετατοπίσεις αντίστοιχα των μαζών m_1, m_2 από την θέση ισορροπίας τους και s_1, s_2 τις σταθερές των ελατηρίων, τότε έχουμε τις εξισώσεις:

$$m_2y_2''=-s_2(y_2-y_1)$$

γιατί το δεύτερο ελατήριο επιμηκύνεται μόνο κατά y_2-y_1, και

$$m_1y_1''=-s_1y_1+s_2(y_2-y_1)$$

Σε στάσιμη κατάσταση το σύστημα αυτό, που δεν περιλαμβάνει αποσβέσεις της κίνησης λόγω τριβών, εκτελεί μίαν καθαρά περιοδική κίνηση. Αν θεωρήσουμε ότι είναι:

$$y_k(t)=Y_k\cos(\omega t+\varphi),\ k=1,2$$

τότε, με εισαγωγή αυτών των σχέσεων στο παραπάνω σύστημα, έχουμε για τον προσδιορισμό των σταθερών Y_1, Y_2 και ω το σύστημα:

$$\begin{pmatrix} \frac{s_2+s_1}{m_1} & -\frac{s_2}{m_1} \\ -\frac{s_2}{m_2} & \frac{s_2}{m_2} \end{pmatrix}\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = \omega^2\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix}$$

Ειδικώτερα, αν $s_1=s_2=s$, $m_1=m_2=m$, τότε προκύπτει από τις εξισώσεις αυτές ότι το σύστημα έχει δύο ιδιοσυχνότητες:

$$\omega_{1,2}=\sqrt{\frac{3\pm\sqrt{5}}{2}}\cdot\sqrt{\frac{s}{m}}$$

3. Ιδιοταλαντώσεις εμφανίζονται σε όλα τα γραμμικά μηχανικά ηλεκτρομαγνητικά συστήματα, αν υποστούν στιγμιαία εξωτερική διέγερση και αφεθούν στην συνέχεια να φθάσουν σε μίαν στάσιμη κατάσταση.

Η γενική μορφή που περιγράφει τέτοια συστήματα είναι:

α. Αν προκειται για μηχανικο συστημα που υποκειται σε ευθυγραμμες μετατοπισεις των στοιχειων του[9]:

$$\mathbf{Mx}'' + \mathbf{Rx}' + \mathbf{Kx} = \mathbf{0}$$

οπου **M** ο πινακας των μαζων, **K** ο πινακας των συντελεστων δυσκαμψιας και **R** ο πινακας των συντελεστων αποσβεσης των επι μερους στοιχειων απο τα οποια αποτελειται το συστημα. Αντιστοιχα ειναι **x** = **x**(t) το διανυσμα που καθε συνιστωσα του οριζει πως μετατοπιζεται ενα απο τα στοιχεια του ολου συστηματος δηλ. το διανυσμα των στιγμιαιων θεσεων.

β. Αν προκειται για μηχανικο συστημα που κανει περιστροφικες κινησεις:

$$\mathbf{Ju}'' + \mathbf{Ru}' + \mathbf{Ku} = \mathbf{0}$$

οπου **J** ο πινακας ροπων αδρανειας, **R** ο πινακας συντελεστων περιστροφικης αποσβεσης, **K** ο πινακας συντελεστων περιστροφικης δυσκαμψιας, u το διανυσμα γωνιωδων μετατοπισεων των συνιστωσων του συστηματος και u' το διανυσμα γωνιωδων ταχυτητων.

γ. Αν προκειται για ηλεκτρομαγνητικο συστημα:

$$\mathbf{Lq}'' + \mathbf{Rq}' + \mathbf{Sq} = \mathbf{0} \quad \text{ή} \quad \mathbf{Li}'' + \mathbf{Ri}' + \mathbf{Si} = \mathbf{0}$$

---

[9] Στο κεφάλαιο αυτό διανύσματα και πίνακες συμβολίζονται με έντονη γραφή για να είναι ευδιάκριτα. Έτσι, **A** = (a_{ik}) είναι ένας πίνακας, ενώ **x** = $(x_i)=(x_1, x_2, \ldots, x_n)^T$ είναι ένα διάνυσμα. Αντίστοιχα, ο συμβολισμός $\mathbf{x}_k$ παριστάνει το k διάνυσμα από ένα σύνολο διανυσμάτων: $\mathbf{x}_k = (x_{1k}, x_{2k}, \ldots, x_{nk})^T$, ενώ ο συμβολισμός $\mathbf{x}^{(k)}= (x_1^{(k)}, x_2^{(k)}, \ldots, x_n^{(k)})^T$ παριστάνει το k διάνυσμα που προκύπτει από μίαν επαναληπτική διαδικασία. Κατ' εξαίρεση, τα σύμβολα **R** και **C** παριστάνουν σε ωρισμένες περιπτώσεις αντίστοιχα τα σύνολα των πραγματικών και μιγαδικών αριθμών.

οπου **L** ο των πινακας συντελεστων επαγωγης των συνιστωσών του συστηματος, **R** ο πινακας των αντιστασεων, και **S** ο πινακας με στοιχεια τα αντιστροφα $1/C_{ik}$ των πυκνοτητων C_{ik} των συνιστωσών του συστηματος. **q** ειναι εδω το διανυσμα των φορτιων που βρισκονται στιγμιαια στα στοιχεια του συστηματος και **i** = **q**' το διανυσμα των ρευματων που διαρρεουν τα στοιχεια αυτα.

Σε καθε περιπτωση εχουμε δηλαδη ενα γραμμικο και ομογενες συστημα διαφορικων εξισωσεων της μορφης:

$$\mathbf{A}\mathbf{y}'' + \mathbf{B}\mathbf{y}' + \mathbf{C}\mathbf{y} = \mathbf{0}$$

Η λυση ενος τετοιου συστηματος ειναι γραμμικος συνδυασμος παραστασεων της μορφης $\mathbf{Y}e^{\lambda t}$, οπου λ βαθμωτη πραγματικη η μιγαδικη σταθερα, και **Y** σταθερο (ανεξαρτητο απο τον χρονο t) διανυσμα.

Βαζοντας $\mathbf{Y}e^{\lambda t}$ στην θεση του y στην παραπανω εξισωση καταληγουμε στην σχεση:

$$(\lambda^2\mathbf{A} + \lambda\mathbf{B} + \mathbf{C})\mathbf{Y} = \mathbf{0}$$

Αν εδω ειναι **B** = **0** τοτε η σχεση αυτη παιρνει αμεσως την κλασσικη μορφη ενος προβληματος ευρεσης ιδιοτιμων, μ, και ιδιοδιανυσματων, **Y**, ενος τετραγωνικου πινακα **M**:

$$\mathbf{M}\mathbf{Y} = \mu\mathbf{Y}\ ,\ \text{με}\ \mathbf{M} = -\mathbf{A}^{-1}\mathbf{C},\ \mu = \lambda^2$$

Αν ο πινακας **B** δεν ειναι μηδενικος τοτε μπορουμε να φερουμε την παραπανω σχεση στην κλασσικη μορφη ενος προβληματος ιδιοτιμων με χρηση της προσθετης μεταβλητης:

$$\mathbf{Z} := \lambda\mathbf{Y}$$

με

$$\mathbf{M} = -\mathbf{A}^{-1}\mathbf{C}\ ,\ \mathbf{N} = -\mathbf{A}^{-1}\mathbf{B}$$

εχουμε τοτε την σχεση:

$$\begin{pmatrix} N & M \\ I & O \end{pmatrix}\begin{pmatrix} Z \\ Y \end{pmatrix} = \lambda\begin{pmatrix} Z \\ Y \end{pmatrix}$$

ή την ισοδύναμη,

$$\begin{pmatrix} -B & -C \\ I & O \end{pmatrix}\begin{pmatrix} Z \\ Y \end{pmatrix} = \lambda\begin{pmatrix} A & O \\ O & I \end{pmatrix}\begin{pmatrix} Z \\ Y \end{pmatrix}$$

όπου όμως ο νέος πίνακας ειναι διπλασίων διαστάσεων απο τους αρχικούς. Το **I** συμβολίζει εδω τον μοναδιαιο πινακα και το **O** τον μηδενικο πινακα.

Οι ιδιοσυχνότητες ενός συστήματος έχουν ιδιαίτερο ενδιαφέρον σε πολλές περιπτώσεις. Ορισμένες φορές μας ενδιαφέρουν για να συντηρήσουμε τις ταλαντώσεις, π. χ. όταν συντονίζουμε τον δέκτη του ραδιοφώνου ή της τηλεόρασης στην συχνότητα εκπομπής ενός σταθμού. Σε άλλες περιπτώσεις, όμως, μας ενδιαφέρουν για να αποφύγουμε την εμφάνιση ιδιοταλαντώσεων που μπορεί να καταστρέψουν μίαν κατασκευή. Τέτοιες μπορεί να προκύψουν σε μίαν οικοδομή από σεισμό αλλά και σε μίαν κρεμαστή γέφυρα υπό την επίδραση του ανέμου (θυμηθείτε την γέφυρα της Τακόμα).

Το ακόλουθο απλό παραδειγμα τετοιου συστηματος δινεται απο τον Jennings στις σελ. 225 - 229 του βιβλιου: Matrix Computation for Engineers and Scientists, Wiley, 1977:

Ένας οριζόντιος πρόβολος φέρει κατά διαστήματα συγκεντρωμένες μάζες, συνολικά τρείς, που μπορούν να εμφανίσουν κάθετες προς τον πρόβολο ταλαντώσεις υπο την επίδραση εξωτερικών δυνάμεων που μπορεί να δράσουν στιγμιαία πάνω στις μάζες αυτές. Αυτό θα μπορούσε, π.χ., να θεωρηθεί σαν ενα πολύ απλοποιημένο μοντέλο των πτερυγίων ενός αεροπλάνου.

Ζητείται ο προσδιορισμός των ιδιοσυχνοτήτων του συστήματος αν ειναι δοσμένες οι μάζες m_1, m_2, m_3 καθως και ο πίνακας ευκαμψίας **F** του πρόβολου που προσδιορίζει τις κάθετες μετατοπίσεις x_i των μαζών υπό την επίδραση καθέτων εξωτερικών δυνάμεων p_i , i = 1,2,3, (στην περιοχή γραμμικής εξαρτήσεως των μεγεθών αυτών): $\mathbf{x} = \mathbf{F}\,\mathbf{p}$

Αν M ειναι ο πινακας 3*3 με m_1, m_2, m_3 σαν στοιχεια της κυριας διαγωνιου και μηδενικα ολα τα αλλα του στοιχεια τοτε εχουμε:

$$\mathbf{p} = -\mathbf{M}\,\mathbf{x}''$$

και επομενως:

$$\mathbf{FMx}'' + \mathbf{x} = \mathbf{0}$$

Βαζοντας εδω:

$$\mathbf{x}(t) = \mathbf{X}\sin(\omega t + \varepsilon)$$

οπου ειναι το **X** σταθερο διανυσμα των μεγιστων αποκλισεων των μαζων, εχουμε το προβλημα ιδιοτιμων:

$$\mathbf{FMX} = \lambda\mathbf{X},\ \lambda := 1/\omega^2$$

Η υπαρξη εδω τριων ιδιοτιμων και ιδιοδιανυσματων σημαινει οτι υπαρχουν τρια ειδη ταλαντωσεων που μπορουν να εμφανισθουν στον προβολο αυτον υπο την επιδραση καταλληλων εξωτερικων διεγερσεων.

Θεωρητικα οι ταλαντωσεις αυτες συνεχιζονται επ'απειρον, αλλα στην πραγματικοτητα βεβαια σβυνουν σταδιακα υπο την επιδραση της αντιστασης του αερα και των εσωτερικων τριβων του υλικου.

Σημειωνουμε οτι μονο το σχετικο μεγεθος των μεγιστων αποκλισεων των μαζων μπορει να προσδιορισθει με αυτον τον τροπο (τα ιδιοδιανυσματα δεν εχουν ωρισμενο μετρο) γιατι το απολυτο μεγεθος τους εξαρταται απο το μεγεθος των εξωτερικων διεγερσεων.

14.1 Εισαγωγή

Οπως ειναι γνωστο, ενας τετραγωνικος πινακας A διαστασεων n*n οριζει μιαν γραμμικη απεικονιση ενος διανυσματικου χωρου, X, π.χ. του $\mathbf{R}^n$, ή $\mathbf{C}^n$ στον εαυτο του. Δηλαδη το διανυσμα $\mathbf{x} \in X$

απεικονιζεται απο τον πινακα **A** στο διανυσμα

$$\mathbf{y} = \mathbf{Ax} \in X. \qquad (14.1)$$

Ο βασικος τροπος να χαρακτηρισουμε μιαν τετοια απεικονιση ειναι να προσδιορισουμε σε ποια διανυσματα απεικονιζει τα μοναδιαια διανυσματα στις διευθυνσεις των αξονων συντεταγμενων.Αν $\mathbf{a}_1, \mathbf{a}_2,..,\mathbf{a}_n$ ειναι οι εικονες αντιστοιχα των μοναδιαιων διανυσματων συντεταγμενων $\mathbf{e}_1, \mathbf{e}_2,..,\mathbf{e}_n$ ($\mathbf{Ae}_k = \mathbf{a}_k$, k = 1(1)n) τοτε ο πινακας **A** εχει τις στηλες $\mathbf{a}_k$, δηλαδη $\mathbf{A} = (\mathbf{a}_1, \mathbf{a}_2,.., \mathbf{a}_n)$.

Ενας αλλος τροπος να χαρακτηρισουμε μιαν τετοια απεικονιση ειναι να προσδιορισουμε τα διανυσματα που η διευθυνση τους δεν μεταβαλλεται κατα την απεικονιση αυτην, δηλαδη τα διανυσματα x με

$$\mathbf{Ax} = \lambda\,\mathbf{x} \qquad (14.2)$$

οπου λ βαθμωτος συντελεστης ($\lambda \in \mathbf{R}$ ή $\mathbf{C}$). Επειδη μαζυ με το $\mathbf{x}$ και καθε διανυσμα c $\mathbf{x}$, οπου c τυχουσα σταθερα, ικανοποιει επισης την σχεση (14.2), για να ορισουμε τα $\mathbf{x}$ μονοσημαντα τα κανονικοποιουμε απαιτωντας ,π.χ., επι πλεον να ειναι ιση με 1 η απολυτα (η κατα μετρο) μεγαλυτερη συνιστωσα τους. Αντι γι' αυτο μπορουμε επισης να απαιτησουμε να ισουται με 1 το ευκλειδιο μηκος τους και να εχουν μιαν συνιστωσα με θετικο πραγματικο μερος. Τα διανυσματα αυτα λεγονται τοτε «ιδιοδιανυσματα» του πινακα $\mathbf{A}$, οι δε αντιστοιχοι συντελεστες συμπτυξης ή επιμηκυνσης, λ , "ιδιοτιμες" του $\mathbf{A}$. Τα ιδιοδιανυσματα αντιστοιχουν δηλαδη στις σταθερες διευθυνσεις κατα την απεικονιση (14.1).

Απο φυσικη πλευρα μπορουμε να θεωρησουμε τα ιδιοδιανυσματα σαν αυτοσυντηρουμενες διαταραχες ενος συστηματος αναδρασεως που η εισοδος και η εξοδος του συνδεονται με την σχεση (14.1) ενω η εξοδος y επανερχεται σαν εισοδος με εναν συντελεστη συζευξης 1/λ:

$$\begin{array}{ccc} \mathbf{x}\rightarrow & [\mathbf{A}] & \rightarrow \mathbf{y} \\ \uparrow & & \downarrow \\ \mathbf{y}/\lambda\leftarrow & [1/\lambda] & \hookleftarrow \end{array}$$

Στο παραπανω συστημα ισχυει:

$$\mathbf{y}^{(j+1)}= \mathbf{A}\, \mathbf{x}^{(j)}$$

οπου οι δεικτες (j), (j+1) αναφερονται στις χρονικες στιγμες, και

$$\mathbf{x}^{(j+1)}= \mathbf{y}^{(j+1)}/\lambda$$

άρα,

$$\mathbf{x}^{(j+1)}= \mathbf{A}\, \mathbf{x}^{(j)}/ \lambda$$

Αν ειναι $\mathbf{x}^{(0)}= \mathbf{x}_k$ και $\lambda = \lambda_k$ τοτε το σημα αυτο διατηρειται συνεχως. Αν αντιθετως ειναι $\mathbf{x}^{(0)} \neq \mathbf{x}_k$ ή $\lambda \neq \lambda_k$ τοτε η αποκριση του συστηματος τεινει να αποσβεσθει η να αυξηθει απεριοριστα μεχρι την καταστροφη του συστηματος. Π.χ. για $\mathbf{x}^{(0)}= \mathbf{x}_k$ αλλα $\lambda \neq \lambda_k$ προκυπτει:

$$\vec{x}^{(j)} = \left(\frac{\lambda_k}{\lambda}\right)^j \vec{x}_k, \quad j = 0,1,2,...$$

οποτε το $\mathbf{x}^{(j)}$ τεινει με αυξανομενο j στο μηδεν αν $|\lambda| > |\lambda_k|$ και στο απειρο αν $|\lambda| < |\lambda_k|$.

14.2 Γενικές Ιδιότητες των Ιδιοδιανυσμάτων και Ιδιοτιμών

Το συστημα (14.2) μπορει να γραφτει και στην μορφη:

$$(\mathbf{A} - \lambda\mathbf{I})\mathbf{x}= \mathbf{0} \qquad (14.3)$$

οπου $\mathbf{I}$ ειναι ο μοναδιαιος πινακας ιδιων διαστασεων n*n με τον $\mathbf{A}$ και $\mathbf{0}$ το μηδενικο διανυσμα στον διανυσματικο χωρο X.

Το γραμμικο αυτο συστημα ειναι ομογενες. Οι λυσεις του διαφερουν επομενως τοτε μονον απο το μηδεν οταν η οριζουσα του πινακα του μηδενιζεται:

$$\det(\mathbf{A} - \lambda\mathbf{I}) = 0 \qquad (14.4)$$

Οι ιδιοτιμες του πινακα A ειναι αρα οι ριζες του πολυωνυμου n βαθμου:

$$P(\lambda) = \det(\mathbf{A} - \lambda\mathbf{I}) \qquad (14.5)$$

που λεγεται «*χαρακτηριστικό πολυώνυμο*» του πινακα $\mathbf{A}$ και είναι επομένως ακριβώς n. Αν οι ιδιοτιμες ειναι διαφορετικες μεταξυ τους, δηλ. το χαρακτηριστικο πολυωνυμο εχει μονο απλες ριζες, τοτε τα ιδιοδιανυσματα ειναι επισης n,

ενα για καθε ιδιοτιμη.

Επειδη ιδιοδιανυσματα που αντιστοιχουν σε διαφορετικες ιδιοτιμες ενος πινακα ειναι γραμμικως ανεξαρτητα μεταξυ τους, αν ενας πινακας εχει μονο απλες ιδιοτιμες τοτε τα ιδιοδιανυσματα του αποτελουν ενα συστημα n γραμμικα ανεξαρτητων διανυσματων, δηλαδη μιαν βαση του διανυσματικου χωρου που εχουμε.

Αν ο πινακας **A** εχει μιαν πολλαπλη ιδιοτιμη, λ_i ,τοτε το συστημα:

$$(\mathbf{A} - \lambda_i \mathbf{I})\mathbf{x} = \mathbf{0} \qquad (14.6)$$

εχει το πολυ τοσες γραμμικα ανεξαρτητες λυσεις οση ειναι η πολλαπλοτητα της ιδιοτιμης. Δηλαδη η διασταση του *ιδιοχωρου* που αντιστοιχει στην ιδιοτιμη λ_i (του υποχωρου των διανυσματων **x** με $\mathbf{Ax} = \lambda_i\mathbf{x}$) δεν ξεπερνα την πολλαπλοτητα της. Δεν ειναι ομως απαραιτητο να εχει διασταση ιση με την πολλαπλοτητα της ιδιοτιμης.

Π.χ. αν:

$$A_\varepsilon = \begin{pmatrix} a & 1 \\ 0 & a+\varepsilon \end{pmatrix}, \quad B_\varepsilon = \begin{pmatrix} a & 0 \\ 0 & a+\varepsilon \end{pmatrix}$$

τοτε και οι δυο πινακες εχουν τις ιδιοτιμες $\lambda_1=a$, $\lambda_2=a+\varepsilon$. Τα αντιστοιχα τους ιδιοδιανυσματα ειναι ομως:

$$A_\varepsilon : \vec{x}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \vec{x}_2 = \begin{pmatrix} 1 \\ \varepsilon \end{pmatrix}, \quad B_\varepsilon : \vec{y}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \vec{y}_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

Για $\varepsilon \to 0$ οι δυο ιδιοτιμες συμπιπτουν σε μιαν διπλη, $\lambda_{1,2}=a$. Ομως στην περιπτωση του $\mathbf{A}_0$ ο ιδιοχωρος της διπλης αυτης ιδιοτιμης ειναι μονοδιαστατος και αποτελειται απο τα πολλαπλασια του $\mathbf{x}_1$, γιατι τοτε το $\mathbf{x}_2$ συμπιπτει με το $\mathbf{x}_1$, ενω ο ιδιοχωρος του **B** ειναι ολος ο $\mathbf{R}^2$, δηλαδη ισχυει $\mathbf{B}_0\mathbf{y} = a\mathbf{y}$ για τυχον $\mathbf{y}\in\mathbf{R}^2$.

Για να απλοποιησουμε τις αποδειξεις των σχεσεων που ακολουθουν εδω θα θεωρησουμε οτι τυχον υπαρχουσες πολλαπλες ιδιοτιμες του πινακα **A** που εξεταζουμε εχουν ιδιοχωρο με διασταση ιση με την πολλαπλοτητα της ιδιοτιμης ωστε να ειναι δυνατον να τον περιγραψουμε σαν γραμμικο συνδυασμο αντιστοιχου πληθους ιδιοδιανυσματων. Ετσι θεωρουμε οτι ενα τυχον διανυσμα $\mathbf{x}\in\mathbf{R}^n$ μπορει να γραφτει σαν γραμμικος συνδυασμος n ιδιοδιανυσματων:

$$\mathbf{x} = c_1\,\mathbf{x}_1 + c_2\,\mathbf{x}_2 + \ldots + c_n\,\mathbf{x}_n \qquad (14.7)$$

Ειδικα για συμμετρικους πραγματικους πινακες ($a_{ik}=a_{ki}\in\mathbf{R}$, για i,k = 1(1)n, ή συντομώτερα, $\mathbf{A}^T=\mathbf{A}$) που εμφανιζονται πολυ συχνα σε εφαρμογες, η γενικώτερα Ερμιτιανούς πινακες ($a_{ki} = \underline{a}_{ik}$, i,k=1(1)n, η συντομώτερα, $\underline{\mathbf{A}}^T = \mathbf{A}$, οπου $\underline{\mathbf{A}}^T$ ο αναστροφος του συζυγους μιγαδικου του **A**)[10], ισχυει οτι ολες οι ιδιοτιμες τους ειναι πραγματικοι αριθμοι ενω τα ιδιοδιανυσματα που ανηκουν σε διαφορετικες ιδιοτιμες ειναι καθετα μεταξυ τους.

Αυτο προκυπτει αμεσως με χρηση των σχεσεων $\mathbf{Ax}_k = \lambda_k\mathbf{x}_k$, $\mathbf{Ax}_j = \lambda_j\mathbf{x}_j$ αν προσεξουμε οτι συμφωνα με τον ορισμο του εσωτερικου γινομενου δυο διανυσματων $\mathbf{x},\mathbf{y} \in\mathbf{C}^n$:

---

[10] Με υπογράμμιση θα παραστήσουμε εδώ τον συζυγή μιγαδικό ενός αριθμού. Αν z=x+iy τότε $\underline{z}$=x-iy.

$$(\vec{x},\vec{y}) := \sum_{i=1}^{n} \underline{x_i} y_i \qquad (14.8)$$

είναι

$\underline{\lambda}_k(\mathbf{x}_j,\mathbf{y}_k) = (\mathbf{x}_j,\lambda_k\mathbf{x}_k) = (\mathbf{x}_j,\mathbf{A}\mathbf{x}_k) \equiv (\underline{\mathbf{A}}^T\mathbf{x}_k,\mathbf{x}_j) = (\mathbf{A}\mathbf{x}_j,\mathbf{x}_k) = (\lambda_j\mathbf{x}_j,\mathbf{x}_k)=\lambda_j(\mathbf{x}_j\ ,\ \mathbf{x}_k)$

δηλαδη,

$$(\underline{\lambda}_k - \lambda_j)(\mathbf{x}_j\ ,\mathbf{x}_k) = 0 \qquad (14.9)$$

Για j = k η σχεση αυτη μας λεει οτι:

$$\underline{\lambda}_k = \lambda_k \qquad (14.10)$$

αρα $\lambda_k \in \mathbf{R}$. Για $j \neq k$, αρα και $\lambda_j \neq \lambda_k$, η σχεση αυτη ισοδυναμει με

$$(\mathbf{x}_j\ ,\mathbf{x}_k) = 0 \qquad (14.11)$$

που σημαινει οτι τα ιδιοδιανυσματα που ανηκουν σε διαφορετικες ιδιοτιμες ειναι καθετα μεταξυ τους. Οταν ο πινακας **A** ειναι πραγματικος και συμμετρικος τοτε τα ιδιοδιανυσματα αυτα ειναι πραγματικα, γιατι ο πινακας $\mathbf{A}-\lambda\mathbf{I}$ ειναι πραγματικος και επομενως το συστημα $(\mathbf{A} - \lambda\ \mathbf{I})\mathbf{x} = \mathbf{0}$ εχει πραγματικες λυσεις.

Αν οι ιδιοτιμες ενος πινακα ειναι γνωστες, ο προσδιορισμος των ιδιοδιανυσματων γινεται με επιλυση των γραμμικων συστηματων της μορφης (14.6) π.χ. με την μεθοδο απαλοιφης.

Ο προσδιορισμος των ιδιοτιμων θα μπορουσε να γινει με προσδιορισμο των συντελεστων του πολυωνυμου (14.5) και ευρεση των ριζων του με μιαν απο τις γνωστες μεθοδους επιλυσης μη γραμμικων εξισωσεων. Η διαδικασια αυτη αποφευγεται ομως για μεγαλους πινακες, γιατι κατα τον τον προσδιορισμο του αναπτυγματος της οριζουσας (14.5) ειναι αναποφευκτα τα σφαλματα στρογγυλεματος που μπορουν να μεταβαλλουν αισθητα τις ιδιοτιμες. Οπως εδειξε ο J.H. Wilkinson, πολλες απο τις ριζες ωρισμενων πολυωνυμων μεταβαλλονται σημαντικα αν αλλαξουμε ελαχιστα εστω και εναν μονο απο τους συντελεστες τους. Ετσι, αν αντι για το πολυωνυμο:

$$P(z):= (z-1)(z-2)(z-3)...(z-19)(z-20) \qquad (14.12)$$

με ριζες $z_k = k$, k = 1(1)20 εχουμε το

$$Q(z):= P(z) - 2^{-23}z^{19} \qquad (14.13)$$

οπου $2^{-23} \approx 1.19\ 10^{-7}$ τοτε για τις ριζες του, w_k, ισχυει $w_k \approx z_k$ μονο για k = 1(1)8. Ετσι ειναι $w_9 \approx 8.91725 \neq 9 = z_9$, $w_{20} \approx 20.8469 \neq 20 = z_{20}$, ενω οι υπολοιπες ριζες ειναι μιγαδικες με $|w_k|\approx|z_k|$. Π.χ., $w_{10}\approx 10.09526 \pm i\ 0.6435$, και $w_{19} \approx 19.5024 \pm i\ 1.94033$ (Βλεπε J.H.Wilkinson: Rounding Errors in Algebraic Processes, Dover, 1994, σελ. 43).

Αντι να κανουμε αμεσο υπολογισμο της οριζουσας (14.5) συνηθως μετασχηματιζουμε τον πινακα **A** σε εναν τριδιαγωνιο η διαγωνιο πινακα $\mathbf{B} = \mathbf{P}^{-1}\mathbf{A}\ \mathbf{P}$ με χρηση καταλληλου αντιστρεψιμου πινακα **P**. Τοτε ειναι:

$$\det(\mathbf{B} - \lambda\mathbf{I}) = \det(\mathbf{P}^{-1}\mathbf{A}\mathbf{P} - \lambda\mathbf{I}) = \det(\mathbf{P}^{-1}(\mathbf{A} - \lambda\mathbf{I})\mathbf{P}) = \det(\mathbf{P}^{-1})\det(\mathbf{A} - \lambda\mathbf{I})\det(\mathbf{P}) \qquad (14.14)$$

οπου η οριζουσα του **P** δεν ειναι μηδεν γιατι ειναι αντιστρεψιμος. Επομενως, οι

οριζουσες det(**B** - λ**I**) και det(**A** - λ**I**) μηδενιζονται ταυτοχρονως.

Αν ο πινακας **B** εχει διαγωνια μορφη, τοτε τα στοιχεια της διαγωνιου του ειναι οι ιδιοτιμες του. Αν ειναι τριδιαγωνιος, τοτε, οπως θα ιδουμε, ο προσδιορισμος των ιδιοτιμων του μπορει να γινει χωρις υπολογισμο των συντελεστων του πολυωνυμου det(**B** - λ**I**).

14.3 Φράγματα Ιδιοτιμών

Για τον προσδιορισμο των ιδιοτιμων ενος πινακα ειναι χρησιμα τα φραγματα που δινονται απο τα ακολουθα θεωρηματα:

Θεώρημα 14.1 (Gerschgorin)

Αν γ_i, σ_k ειναι αντιστοιχα τα αθροισματα των απολυτων τιμων (ή μετρων) των μη διαγωνιων στοιχειων των γραμμων και των στηλων του πινακα **A**:

$$\gamma_i := \sum_{k=1,k\neq i}^{n} |a_{ik}|, \quad \sigma_k := \sum_{i=1,i\neq k}^{n} |a_{ik}| \tag{14.15}$$

τοτε ισχυει οτι:

α) Καθε ιδιοτιμη του **A** βρισκεται μεσα στην ενωση των κυκλων:

$$\Gamma_i := \{z \in \mathbf{C}: |z - a_{ii}| \leq \gamma_i\}, \; i = 1(1)n \tag{14.16}$$

β) Καθε ιδιοτιμη του A βρισκεται μεσα στην ενωση των κυκλων:

$$\Sigma_k := \{z \in \mathbf{C}: |z - a_{kk}| \leq \gamma_i\}, \; i = 1(1)n \tag{14.17}$$

Δηλαδη καθε ιδιοτιμη ικανοποιει μιαν τουλαχιστον απο τις ανισότητες (14.16) και μιαν τουλαχιστον απο τις (14.17).

Απόδειξη

α) Η σχεση **Ax** = λ**x** ισοδυναμει με:

$$(\lambda - a_{ii})x_i = \sum_{k=1,k\neq i}^{n} a_{ik}x_k, \; i = 1(1)n$$

Αρα για ολους τους δεικτες i ισχυει:

$$|\lambda - a_{ii}| = \sum_{k=1,k\neq i}^{n} |a_{ik}| \left|\frac{x_k}{x_i}\right|$$

Αν θεωρησουμε ειδικωτερα εναν δεικτη j με $|x_j| = \max_{k=1(1)n} |x_k|$ τοτε οι λογοι $|x_k / x_j|$ ειναι ολοι μικροτεροι απο 1 ή ισοι με 1 και επομενως προκυπτει:

$$|\lambda - a_{jj}| = \sum_{k=1,k\neq i}^{n} |a_{jk}| = \gamma_j$$

β) Συμφωνα με τις ιδιοτητες των οριζουσων ειναι det(**B**) = det($\mathbf{B}^T$) οπου $\mathbf{B}^T$ ο αναστροφος του πινακα **B**.

Στην προκειμενη περιπτωση ειναι αρα

$$\det(\mathbf{A} - \lambda\mathbf{I}) = \det(\mathbf{A}^T - \lambda\mathbf{I})$$

Επομενως ο πινακας **A** εχει τις ιδιες ιδιοτιμες με τον ανάστροφό του, $\mathbf{A}^T$. Αν εφαρμοσουμε στον $\mathbf{A}^T$ τα συμπερασματα της περιπτωσης (α) εχουμε αμεσως αυτα της περιπτωσης (β) γιατι οι γραμμες του πινακα $\mathbf{A}^T$ ειναι στηλες του **A**.

Παρατήρηση: Απλά Φράγματα Ιδιοτιμών

Απο το θεωρημα αυτο προκυπτουν αμεσως οι απλουστερες ανισοτητες

$$|\lambda| \le \max_{i=1(1)n} \sum_{k=1}^{n} |a_{ik}| \qquad (14.18)$$

και

$$|\lambda| \le \max_{k=1(1)n} \sum_{i=1}^{n} |a_{ik}| \qquad (14.19)$$

Αν ειναι γνωστες προσεγγισεις λ και **x** καποιας ιδιοτιμης και του αντιστοιχου ιδιοδιανυσματος ενος ερμητιανου πινακα $\mathbf{A} = \mathbf{A}^T$ (δηλαδη απλα ενος συμμετρικου πινακα στην περιπτωση που τα στοιχεια του ειναι πραγματικοι αριθμοι) τοτε μπορουμε να προσδιορισουμε το σφαλμα αυτων των προσεγγισεων με το ακολουθο θεωρημα:

Θεώρημα 14.2

Αν ο **A** ειναι ερμητιανος πινακας με ιδιοτιμες λ_i , i = 1(1)n και αν

$$\mathbf{y} := \mathbf{A}\mathbf{x} - \lambda\mathbf{x},\ \mathbf{x} \neq \mathbf{0} \qquad (14.20)$$

τοτε ισχυει:

α) $$\min_{i=1(1)n} |\lambda_i - \lambda| \le \frac{\|y\|_2}{\|x\|_2} = \frac{\|Ax - \lambda x\|_2}{\|x\|_2}$$

(14.21)

β) Το μηκος του διανυσματος σφαλματος 7**y**72 ειναι ελαχιστο για δοσμενο **x** αν βάλουμε:

$$\lambda = R(\vec{x}) := \frac{(\vec{x}, A\vec{x})}{(\vec{x}, \vec{x})} \qquad (14.22)$$

οπου η τελευταια παρασταση λεγεται «πηλίκο του Ρέιλη (Rayleigh)».

Απόδειξη

α) Αν $\{\mathbf{x}_i\}$ ειναι το ορθογωνιο συστημα των ιδιοδιανυσματων του πινακα $\mathbf{A} = \mathbf{A}^T$ που προκυπτει αν περιγραψουμε τον ιδιοχωρο πολλαπλων ιδιοτιμων σαν γραμμικο συνδυασμο ορθογωνιων μεταξυ τους ιδιοδιανυσματων τοτε ισχυει:

$$\vec{x} = \sum_{i=1}^{n} c_i \vec{x}_i, \quad \vec{y} = \sum_{i=1}^{n} c_i (\lambda_i - \lambda) \vec{x}_i$$

Εξ αιτιας της ορθογωνιοτητας των x ειναι τωρα:

$$\|\vec{x}\|_2^{\,2} = (\vec{x},\vec{x}) = \left(\sum_{i=1}^{n} c_i\vec{x}_i, \sum_{k=1}^{n} c_k\vec{x}_k\right) = \sum_{i,k=1}^{n} \underline{c}_i c_k (\vec{x}_i,\vec{x}_k) = \sum_{i,k=1}^{n} \underline{c}_i c_k \delta_{ik} = \sum_{i=1}^{n} |\, c_i \,|^2$$

και παρομοια

$$\|\vec{y}\|_2^{\,2} = (\vec{y},\vec{y}) = \sum_{i=1}^{n} |\, c_i \,|^2 |\, \lambda_i - \lambda \,|^2$$

Αρα

$$\|\vec{y}\|_2^{\,2} \geq \min_{\,i=1(1)n} |\, \lambda_i - \lambda \,|^2 \sum_{i=1}^{n} |\, c_i \,|^2 = \min_{\,i=1(1)n} |\, \lambda_i - \lambda \,|^2 \, \|\vec{x}\|_2^{\,2}$$

β) Επειδη ο **A** ειναι ερμιτιανος ισχυει:

$7\mathbf{y}7_2^2 = (\underline{\mathbf{x}}^T\underline{\mathbf{A}}^T - \lambda\underline{\mathbf{x}}^T)(\mathbf{Ax}-\lambda\mathbf{x}) = (\underline{\mathbf{x}}^T\mathbf{A}^T - \lambda\underline{\mathbf{x}}^T)(\mathbf{Ax}-\lambda\mathbf{x}) = \underline{\mathbf{x}}^T\mathbf{A}^2\mathbf{x} - 2\lambda\underline{\mathbf{x}}^T\mathbf{Ax} + \lambda^2\underline{\mathbf{x}}^T\mathbf{x}$

$= (\mathbf{x}, \mathbf{A}^2\mathbf{x}) - 2\lambda(\mathbf{x},\mathbf{Ax}) + \lambda^2(\mathbf{x},\mathbf{x}) = (\mathbf{x},\mathbf{x})(\lambda - R(\mathbf{x}))^2 + (\mathbf{x},\mathbf{A}^2\,\mathbf{x}) - R(\mathbf{x})^2$

Η τελευταια παρασταση ελαχιστοποιειται προφανως για $\lambda = R(x)$.

Θεώρημα 14.3

Αν ο **A** ειναι ερμιτιανος πινακας με ιδιοτιμες και ιδιοδιανυσματα αντιστοιχα λ_i, $\mathbf{x}_i$ και αν **x** ειναι τυχον διανυσμα τοτε ισχυει:

α) $|\lambda_1| = \max_{i=1(1)n} |\lambda_i| \geq |R(\mathbf{x})|^2$ (14.23)

β) $R(\mathbf{x}_i) = \lambda_i$, $i = 0(1)n$ (14.24)

αρα $\max_{\mathbf{x}\in V} |R(\mathbf{x})| = |\lambda_1|$ (14.25)

οπου V ο χωρος $\mathbf{R}^n$ ή $\mathbf{C}^n$.

Απόδειξη

Αν γραψουμε παλι οπως προηγουμενως:

$$\vec{x} = \sum_{i=0}^{n} c_i\vec{x}_i$$

προκυπτει:

$$R(\vec{x}) = \frac{\sum_{i=0}^{n} \lambda_i \,|\, c_i \,|^2}{\sum_{i=0}^{n} |\, c_i \,|^2} \leq \frac{\sum_{i=0}^{n} |\, \lambda_i \,|\, |\, c_i \,|^2}{\sum_{i=0}^{n} |\, c_i \,|^2} \leq \max_{\,i=0(1)n} |\, \lambda_i \,| \qquad (14.26)$$

β) Για $\mathbf{x} = \mathbf{x}_k$ ειναι c_i για $i \neq k$, $c_k = 1$ και απο την πρωτη σχεση (14.26) προκυπτει αμεσως $R(\mathbf{x}_k) = \lambda_k$.

14.4 Προσδιορισμός της Μεγίστης Ιδιοτιμής – Μέθοδος των Δυνάμεων

Σε πολλες περιπτωσεις μας αρκει ο προσδιορισμος της μεγιστης κατα απολυτη τιμη ή μετρο ιδιοτιμης ενος πινακα **A**, γιατι, οπως στο παραπανω παραδειγμα του οριζοντιου προβολου, αντιστοιχει στην μικροτερη ιδιοσυχνοτητα ταλαντωσεων του συστηματος (**A** = **FM**, ενω $\lambda = 1/\omega^2$).

Στις περιπτωσεις αυτες μπορει να χρησιμοποιηθει η μεθοδος των δυναμεων που προκυπτει με τον ακολουθο τροπο:

Εστω οτι ισχυει:

$$|\lambda_1|>|\lambda_2|\geq|\lambda_3|\geq\ldots\geq|\lambda_n| \qquad (14.27)$$

Αν με τυχον αρχικο διανυσμα $\mathbf{x}^{(0)}$ δημιουργησουμε την ακολουθια:

$$\mathbf{x}^{(k)}=\mathbf{A}\mathbf{x}^{(k-1)}=\ldots=\mathbf{A}^n\,\mathbf{x}^{(0)}\quad,\ k=0,1,2,\ldots \qquad (14.28)$$

τοτε με βαση την (14.7) προκυπτει:

$$\vec{x}^{(0)}=\sum_{i=1}^{n}c_i\vec{x}_i \qquad (14.29)$$

και

$$\vec{x}^{(k)}=\sum_{i=1}^{n}\lambda_i^{\,k}c_i\vec{x}_i=\lambda_1^{\,k}\left\{c_1\vec{x}_1+\sum_{i=2}^{n}\left(\frac{\lambda_i}{\lambda_1}\right)^k c_i\vec{x}_i\right\} \qquad (14.30)$$

Οι λογοι λ_i/λ_1, i = 2,3,..,n ειναι εδω ολοι απολυτα ή κατα μετρο μικροτεροι απο την μοναδα. Επομενως οι αυξανομενες δυναμεις τους τεινουν στο μηδεν. Αν το τυχον αρχικο διανυσμα $\mathbf{x}^{(0)}$ εχει μιαν συνιστωσα στην διευθυνση του ιδιοδιανυσματος $\mathbf{x}_1$, δηλ. αν $c_1\neq 0$, τοτε ισχυει για τα πηλικα της ιδιας συνιστωσας, j, διαδοχικων διανυσματων της ακολουθιας:

$$\frac{x_j^{\,(k+1)}}{x_j^{\,(k)}}=\lambda_1\frac{\left\{c_1\vec{x}_1+\sum_{i=2}^{n}\left(\frac{\lambda_i}{\lambda_1}\right)^{k+1}c_i\vec{x}_i\right\}_j}{\left\{c_1\vec{x}_1+\sum_{i=2}^{n}\left(\frac{\lambda_i}{\lambda_1}\right)^{k}c_i\vec{x}_i\right\}_j}\to\lambda_1\quad \gamma\iota\alpha\ k\to\infty \qquad (14.31)$$

Αξιο προσοχης ειναι επισης οτι η ακολουθια $\{\mathbf{x}^{(k)}\}$ επιτρεπει επισης τον προσδιορισμο του ιδιοδιανυσματος $\mathbf{x}_i$, γιατι

$$\frac{1}{\lambda_1^{\,k}}\vec{x}^{(k)}\to c_1\vec{x}_1\quad \gamma\iota\acute{\alpha}\ k\to\infty \qquad (14.32)$$

Απο την σχεση αυτην προκυπτει ομως επισης οτι οι συνιστωσες των διανυσματων $\mathbf{x}^{(k)}$ τεινουν με αυξανομενο k προς το μηδεν ή το απειρο αναλογα με το αν $|\lambda_1|<1$ ή $|\lambda_1|>1$. Επομενως, για να χρησιμοποιησουμε πρακτικα αυτην την επαναληπτικη διαδικασια ειναι απαραιτητο σε καθε επαναληψη να κανονικοποιουμε τα διανυσματα αυτα, π.χ. ετσι ωστε η μεγιστη κατα μετρο συνιστωσα του να ισουται με την μοναδα. Στην πραξη, δηλαδη, υπολογιζουμε την ακολουθια:

$$\vec{z}^{(0)}=\vec{x}^{(0)},\ \vec{y}^{(k)}=A\vec{z}^{(k-1)},\ \vec{z}^{(k)}=\frac{1}{\max(\vec{y}^{(k)})}\vec{y}^{(k)} \qquad (14.33)$$

με k = 1,2,3,... , οπου $\max(\mathbf{y}^{(k)})$ μια συνιστωσα του $\mathbf{y}^{(k)}$ με μεγιστο μετρο (ή μεγιστη απολυτη τιμη), δηλαδη:

$$\max(\mathbf{y})=y_M\ ,\ |y_M|=\max_{j=1(1)n}|y_j| \qquad (14.34)$$

οποτε προκυπτει:

$$\vec{z}^{(k)} \to \frac{1}{\max(\vec{x}_1)}\vec{x}_1, \quad \max(\vec{y}^{(k)}) \to \lambda_1 \quad \gamma\iota\acute{\alpha}\ k \to \infty \tag{14.35}$$

γιατι απο τις σχεσεις (14.33) προκυπτει:

$$\mathbf{z}^{(k)} = d_k\mathbf{A}\mathbf{z}^{(k-1)} \quad , d_k = 1/\max(\mathbf{y}^{(k-1)}) \tag{14.36}$$

αρα

$$\mathbf{z}^{(k)} = d_k\mathbf{A}\mathbf{z}^{(k-1)} = d_k d_{k-1}\mathbf{A}^2\mathbf{z}^{(k-2)} = \ldots = d_k d_{k-1} \ldots d_1\mathbf{A}^k\mathbf{z}^{(0)} \tag{14.37}$$

και

$$\max(\mathbf{z}^{(k)}) = d_k d_{k-1} \ldots d_1 \max(\mathbf{A}^k\mathbf{z}^{(0)}) \tag{14.38}$$

δηλαδη,

$$\vec{z}^{(k)} = \frac{1}{\max(A^k\vec{x}^{(0)})}A^k\vec{x}^{(0)} = \frac{\vec{x}^{(k)}}{\max(\vec{x}^{(k)})} \tag{14.39}$$

$$\vec{y}^{(k)} = \frac{1}{\max(A^{k-1}\vec{x}^{(0)})}A^k\vec{x}^{(0)} = \frac{\vec{x}^{(k)}}{\max(\vec{x}^{(k-1)})} \tag{14.40}$$

Εξ αιτιας της πρωτης απο τις σχεσεις (14.35) και της ταυτοτητας $R(d\mathbf{x}_i) \equiv \lambda_i$ με τυχον $d \in \mathbf{R}$ ή $\mathbf{C}$ προκυπτει επισης οτι

$$R(\vec{z}^{(k)}) = \frac{(\vec{z}^{(k)}, A\vec{z}^{(k)})}{(\vec{z}^{(k)}, \vec{z}^{(k)})} \to \lambda_1 \quad \gamma\iota\acute{\alpha}\ k \to \infty \tag{14.41}$$

Αν ο πινακας A ειναι ερμιτιανος, τοτε η τελευταια ακολουθια συγκλινει μαλιστα πολυ ταχυτερα προς λ_1 απο οτι η $\max(\mathbf{y}^{(k)})$.

Αυτο οφειλεται στο οτι για μεγαλες τιμες του k ειναι

$$\vec{y}^{(k)} = \frac{\lambda_1^{\,k}c_1\vec{x}_1 + \lambda_2^{\,k}c_2\vec{x}_2 + \ldots}{\max\left(\lambda_1^{\,k-1}c_1\vec{x}_1 + \lambda_2^{\,k-1}c_2\vec{x}_2 + \ldots\right)} \cong \frac{\lambda_1^{\,k}c_1\vec{x}_1 + \lambda_2^{\,k}c_2\vec{x}_2 + \ldots}{\max\left(\lambda_1^{\,k-1}c_1\vec{x}_1\right)}$$

$$\vec{y}^{(k)} - \lambda_1\frac{\vec{x}_1}{\max(\vec{x}_1)} \cong \frac{\lambda_1^{\,k}c_1\vec{x}_1 + \lambda_2^{\,k}c_2\vec{x}_2 + \ldots}{\max\left(\lambda_1^{\,k-1}c_1\vec{x}_1\right)} - \lambda_1\frac{\vec{x}_1}{\max(\vec{x}_1)} \cong \frac{\lambda_1\left(\lambda_2/\lambda_1\right)^{k-1}c_2\vec{x}_2 + \ldots}{\max\left(c_1\vec{x}_1\right)} \tag{14.42}$$

και

$$\max(\vec{y}^{(k)}) - \lambda_1 \cong \lambda_1(\lambda_2/\lambda_1)^{k-1}\frac{c_2\max(\vec{x}_2)}{c_1\max(\vec{x}_1)} \tag{14.43}$$

οπου $A \cong B$ σημαινει εδώ οτι οι ποσοτητες A και B οριακα (για $k \to \infty$) συμπιπτουν, δηλαδη οτι $A_k/B_k \to 1$ για $k \to \infty$, οποτε λεγονται και "ασυμπτωτικα ισες".

Αντιστοιχα ειναι λογω της (14.30) και της ορθογωνιοτητας των $\mathbf{x}_i$:

$$R(\vec{z}^{(k)}) = R(\vec{x}^{(k)}) = \frac{(\vec{x}^{(k)}, \vec{x}^{(k+1)})}{(\vec{x}^{(k)}, \vec{x}^{(k)})} = \frac{\sum \lambda_i^{\,2k+1}|c_i|^2}{\sum \lambda_i^{\,2k}|c_i|^2} \tag{14.44}$$

και

$$R(\vec{z}^{(k)}) - \lambda_1 \cong \frac{\lambda_1^{\,2k+1}|c_1|^2 + \lambda_2^{\,2k+1}|c_2|^2}{\lambda_1^{\,2k}|c_1|^2} - \lambda_1 \cong (\lambda_2 - \lambda_1)\left(\frac{\lambda_2}{\lambda_1}\right)^{2k}\left|\frac{c_2}{c_1}\right|^2 \tag{14.45}$$

Η τελευταια παρασταση τεινει πολυ πιο γρηγορα στο μηδεν απο οτι η (14.43) γιατι η (14.45) τεινει προς μηδεν τοσο γρηγορα οσο ο λογος $(\lambda_2/\lambda_1)^k$ ενω η (14.43) οσο ο λογος $(\lambda_2/\lambda_1)^k$.

Το Σφάλμα της Προσέγγισης αν $A^T=A$

Αν ο **A** είναι ερμητιανός πίνακας, $\mathbf{A}^T=\mathbf{A}$, και μ_1 κάποια προσέγγιση της λ_1, π.χ. $\mu_1=R(\mathbf{z}^{(k)})$ για κάποια τιμή του k, τότε με βάση το θεώρημα 14.2 ισχύει:

$$|\lambda_1-\mu_1|\le\frac{\left\|A\vec{z}^{(k)}-\mu_1\vec{z}^{(k)}\right\|_2}{\left\|\vec{z}^{(k)}\right\|_2}=\frac{\left\|\vec{y}^{(k+1)}-\mu_1\vec{z}^{(k)}\right\|_2}{\left\|\vec{z}^{(k)}\right\|_2}$$

Παρατήρηση

Η μέθοδος των δυνάμεων μπορεί να εφαρμοσθεί και στην περίπτωση που υπαρχουν δύο συζυγείς μιγαδικές ιδιοτιμές ή δύο αντίθετες πραγματικές ιδιοτιμές με μέγιστο μέτρο. για την περίπτωση αυτήν βλέπε Ι.Β.Κιουστελίδη: Αριθμητική Ανάλυση, 1996, σελ. 374.

Παράδειγμα Εύρεσης της Μέγιστης Ιδιοτιμής με την Μέθοδο των Δυνάμεων

Να προσδιορισθει με την μεθοδο των δυναμεων η μεγιστη ιδιοτιμή των πινακων:

$$A=\begin{pmatrix}1&1&0\\1&2&0.5\\0&2&3\end{pmatrix},\quad B=\begin{pmatrix}1&1&0\\1&2&1\\0&1&3\end{pmatrix}$$

και να συγκριθουν στις δυο περιπτωσεις οι προσεγγισεις $\max(\mathbf{y}^{(k)})$ και $R(\mathbf{z}^{(k)})$.

Λύση

Οι ιδιοτιμες και των δυο πινακων ειναι ιδιες:

$$\lambda_1=2=\sqrt{3},\ \lambda_2=2,\ \lambda_3=2-\sqrt{3}$$

αλλα ενω ο **B** ειναι συμμετρικος, ($\mathbf{B}^T=\mathbf{B}$), ο **A** δεν ειναι.

Εφαρμοζοντας την μεθοδο των δυναμεων στον πινακα **A** με αρχικο διανυσμα $\mathbf{x}^{(0)}=(1,1,1)^T$ και εκτελώντας τις πράξεις με ακρίβεια 15 δεκαδικων ψηφιων μετα την υποδιαστολη, έχουμε μετα απο 60 επαναληψεις τοσο με χρηση της

προσεγγισης $\max(\mathbf{y}^{(k)})$ όσο και της $R(\mathbf{z}^{(k)})$:

$$\max(\mathbf{y}^{(k)})=R(\mathbf{z}^{(k)})=3.732050807568877$$

$$\mathbf{z}^{(60)}=(0.1339745962155614,\ 0.3660254037844387,\ 1)\ .$$

Αντιθετα στην περιπτωση του πινακα **B**, επειδη ειναι συμμετρικος, ο συντελεστης Ρέιλη φθανει αυτην την ακριβεια μετα απο 27 επαναληψεις ενω

η τιμη max($\mathbf{y}^{(27)}$) συμπιπτει με την ιδιοτιμη λ_1 μονο κατα τα πρωτα 7 δεκαδικα ψηφια μετα την υποδιαστολη.

14.5 Προσδιορισμός της Ελαχίστης Ιδιοτιμής με την Μέθοδο των αντιστρόφων Δυνάμεων

Στην περίπτωση που μας ενδιαφέρει η ελαχίστη κατά μέτρο ιδοτιμή λ_n του πίνακα **A**, αντί για την ακολουθία:

$$\mathbf{x}^{(k+1)}=\mathbf{A}\mathbf{x}^{(k)},\ k=1,2,3,\ldots$$

χρησιμοποιούμε την ακολουθία των αντιστρόφων δυνάμεων:

$$\mathbf{x}^{(k+1)}=\mathbf{A}^{-1}\mathbf{x}^{(k)},\ k=1,2,3,\ldots$$

ή ισοδύναμα,

$$\mathbf{A}\mathbf{x}^{(k+1)}=\mathbf{x}^{(k)},\ k=1,2,3,\ldots$$

γιατί αν ο **A** έχει ως ιδιοτιμή ελαχίστου μέτρου την λ_n τότε ο $\mathbf{A}^{-1}$ έχει την $1/\lambda_n$ ως ιδιοτιμή μεγίστου μέτρου, για την οποία ισχύουν όλα όσα είπαμε πιο πάνω.

Στην προκειμένη περίπτωση δημιουργούμε, λοιπόν, την ακολουθία:

$$\vec{z}^{(0)}=\vec{x}^{(0)},\ A\vec{y}^{(k)}=\vec{z}^{(k-1)},\ \vec{z}^{(k)}=\frac{\vec{y}^{(k)}}{\max(\vec{y}^{(k)})}=d_k\vec{y}^{(k)}=d_kA^{-1}\vec{z}^{(k-1)}$$

με k=1,2,3,…. Το διάνυσμα $\mathbf{y}^{(k)}$ προσδιορίζεται κάθε φορά με λύση του συστήματος $\mathbf{Ay}=\mathbf{z}^{(k-1)}$, κατά προτίμηση με ανάλυση του πίνακα **A** σε γινόμενο ένος κάτω τριγωνικού πίνακα **L** και ένος άνω τριγωνικού πίνακα **U**: **A**=**LU**. Αυτό επιτρέπει τον προσδιορισμό των $\mathbf{y}^{(1)}$, $\mathbf{y}^{(2)}$, $\mathbf{y}^{(3)}$, … με άμεση επίλυση κάθε φορά δύο τριγωνικών συστημάτων.

Με παρόμοιο τρόπο όπως παραπάνω μπορούμε να αποδείξουμε εδώ τις σχέσεις:

$$\vec{z}^{(k)}=\frac{A^{-k}\vec{z}^{(0)}}{\max(A^{-k}\vec{z}^{(0)})}\cong\frac{\vec{x}_n}{\max(\vec{x}_n)}=\vec{x}_n\quad,\quad \vec{y}^{(k+1)}=A^{-1}\vec{z}^{(k)}\cong\frac{1}{\lambda_n}\vec{x}_n$$

και

$$\widetilde{R}(\vec{z}^{(k)}):=\frac{(\vec{z}^{(k)},A^{-1}\vec{z}^{(k)})}{(\vec{z}^{(k)},\vec{z}^{(k)})}=\frac{(\vec{z}^{(k)},\vec{y}^{(k+1)})}{(\vec{z}^{(k)},\vec{z}^{(k)})}\cong\frac{1}{\lambda_n}$$

όπου κανονικοποιούμε τα ιδιοδιανύσματα έτσι ώστε να είναι max($\mathbf{x}_k$)=1.

Αν είναι $\mathbf{A}^T=\mathbf{A}$ και $\mu_n=1/\widetilde{R}(\vec{z}^{(k)})$,τότε ισχύει και εδώ με βάση το θεώρημα 14.2 η εκτίμηση σφάλματος :

$$|\lambda_n-\mu_n|\le\frac{\left\|A\vec{z}^{(k)}-\mu_n\vec{z}^{(k)}\right\|_2}{\left\|\vec{z}^{(k)}\right\|_2}$$

όπως επίσης η

$$\left|\frac{1}{\lambda_n}-\frac{1}{\mu_n}\right| \le \frac{\left\|A^{-1}\vec{z}^{(k)}-\vec{z}^{(k)}/\mu_n\right\|_2}{\left\|\vec{z}^{(k)}\right\|_2}=\frac{\left\|\vec{y}^{(k+1)}-\vec{z}^{(k)}/\mu_n\right\|_2}{\left\|\vec{z}^{(k)}\right\|_2}$$

Ασκήσεις

1. Με την βοηθεια του θεωρηματος του Gerschgorin εντοπιστε περιοχες μεσα στις οποιες βρισκονται οι ιδιοτιμες των παρακατω πινακων.

2. Προσδιοριστε τοσο με την μεθοδο των δυναμεων οσο και με χρηση των συντελεστων Ρειλη την μεγαλυτερη και την μικρότερη ιδιοτιμη των πινακων:

$$A=\begin{pmatrix} 4 & -2 & 2 \\ -2 & 5 & 1 \\ 2 & 1 & 6 \end{pmatrix}, B=\begin{pmatrix} 1 & 2 & -3 \\ 2 & 1 & 2 \\ -3 & 2 & 1 \end{pmatrix}, C=\begin{pmatrix} 4 & 1 & 0 \\ 1 & -2 & 1 \\ 0 & 1 & 1 \end{pmatrix}$$

ΒΙΒΛΙΟΓΡΑΦΙΑ

1. Ελληνική

Ακρίβης, Γ.Δ. – Δουγαλής, Β. Α.: Εισαγωγή στην Αριθμητική Ανάλυση, Πανεπιστημιακές Εκδόσεις Κρήτης, Ηράκλειο, 1997.

Βραχάτης, Μ.Ν.: Αριθμητική Ανάλυση, Ελληνικά Γράμματα, Αθήνα 2002

Βραχάτης, Μ.Ν.: Αριθμητική Επίλυση Συνήθων Διαφορικών Εξισώσεων, Ελληνικά Γράμματα, Αθήνα 2003.

Κιουστελίδης, Ι.Β.: Αριθμητική Ανάλυσις, Αθήναι, 1977.

Κιουστελίδης, Ι.Β.: Αριθμητική Ανάλυση, Μία Εισαγωγή με Φυσικές και Τεχνικές Εφαρμογές και Προγράμματα σε BASIC, Αθήναι, 1996.

Μπακόπουλος Α. – Χρυσοβέργης, Ι.: Εισαγωγή στην Αριθμητική Ανάλυση, με Βιβλιοθήκη Προγραμμάτων και Δισκέττα, Εκδ. Συμεών, Αθήνα.

Παπαγεωργίου Γ.Σ. – Τσίτουρας Χ.Γ.: Αριθμητική Ανάλυση με εφαρμογές σε MATLAB και Mathematica, Εκδ. Συμεών, Αθήνα 2000.

2. Ξενόγλωσση

Abramowitz,M. - Stegun,I: Handbook of Mathematical Functions, Dover, New York, 1965.

Amelkin : Differential Equations in Applications, MIR, Moscow, 1990.

Antia, H.M.: Numerical Methods for Scientists and Engineers, Tata McGraw-Hill, New Delhi, 1991.

Baker,G.L. - Golub,J.P.: Chaotic Dynamics, Cambridge U.P., 1990.

Beresin,I.S.-Shidkow, N.P.: Numerische Methoden, Deutscher Verlag der Wissenschaften, Berlin, Bd.1:1970, Bd.2:1971.

Blum, E. K.: Numerical Analysis and Computation; Theoryand Practice, Addison Wesley, 1972.

Burden, R.L. - Faires, J.D.: Numerical Analysis, 3.Ed., Prindle, Weber & Smith, Boston, 1985.

Conte,S.D. - de Boor,C.: Elementary Numerical Analysis, 2.Ed. McGraw-Hill, Kogacusha, 1972.

Chapra - Canale: Numerical Methods for Engineers, Mc Graw-Hill, 1990.

Cheney, E.W.: Introduction to Approximation Theory, McGraw - Hill, New York, 1966

Collatz,L.: The Numerical Treatment of Differential Equations, 3.Ed., Springer, Berlin 1966.

Condon - Odishaw: Handbook of Physics Cryer, Colin W.: Numerical Functional Analysis, Oxford, 1982

Danilina-Dubrovskaya-Kvasha- Smirnov: Computational Mathematics, MIR, 1988.

Dorny,N.: A Vector Space Approach to Models and Optimization

El Nasie, M.S.: Stress, Stability and Chaos, Mc Graw - Hill, 1990,

Fletcher,R.: Practical Methods of Optimization, Wiley, 1987.

Freud, G.: Orthogonal Polynomials, Pergamon Press, Oxford, 1971.

Gastinel, N.: Linear Numerical Analysis, Herman, Paris, 1970.

Isaacson,E. - Keller,H.B.: Analysis of Numerical Methods, Wiley, 1966.

Jain, M.K.- Iyengar,S.R.K.- Jain,R.K.: Numerical Methods for Scientific and Engineering Computation, Wiley Eastern Ltd., 1989.

Iyengar,N.G.R.- Gupta,S.K.: Programming Methods in Structural Design, Edward Arnold, London, 1980.

Jennings,A. :Matrix Computation for Engineers and Scientists, Wiley, 1977.

Kincaid, D. – Cheney, W.: Numerical analysis, Brooks/Cole, Pacific Grove, 1991.

Kopchenova, N.V.-Maron, I.A.: Computational Mathematics, MIR, Moscow 1975.

Lax,P. - Burstein,S. -Lax,A.: Calculus with Applications and Computing,vol.1, Springer, 1976.

Lambert, J.D.: Computatinal Methods for Ordinary Differential Equations, Wiley, 1973.

Luenberger,D.G.: Introduction to Linear and Nonlinear Programming, Addison - Wesley, 1973.

Ostrowski, A.M.: Solution of Equations in Euclidean and Banach Spaces, Academic Press, New York, 1970.

Parker,T.S. - Chua,L.O. : Practical Numerical Algorithms for Chaotic Systems, Springer, 1989.

Pipes,L.A.- Harvill,L.R., Applied Mathematics for Engineers and Physicists, McGraw - Hill, 3.Ed., 1970.

Phillips,G.M. - Taylor,P.J.: Theory and Applications of Numerical Analysis, Academic Press, 1973.

Ralston,A. - Rabinowitz,P.: A First Course in Numerical Analysis, Mcgraw - Hill Kogakusha, 2.Ed.,1978.

Rivlin, T.J.: An Introduction to the Approximation of Functions, Dover, 1981.

Schultz, M.H.: Spline Analysis, Prentice - Hall, 1973

Schwarz, R.J. - Friedland,B.: Linear Systems, McGraw -Hill, 1965.

Shoup,T.E.: A Practical Guide to Computer Methods for Engineers, Prentice - Hall, 1979.

Smith,G.D.: Numerical Solution of Partial Differential Equations, Oxford, 1969.

Stummel,F. - Hainer,K.: Praktische Mathematik, Teubner, 1971.

Szego, G.: Orthogonal Polynomials, American Mathematical Society, New York, 1967.

Varga,R.: Matrix Iterative Analysis, Prentice Hall, 1962.

G.N.Watson: A Treatise on the Theory of Bessel Functions, Cambridge Univ.Press, 1958.

Werner,H. - Schaback,R.: Praktische Mathematik II, Springer 1972.

Wilkinson,J.H.: Rounding Errors in Algebraic Processes, Dover, 1994.

ΣΥΝΤΟΜΗ ΕΙΣΑΓΩΓΗ ΣΤΟ MATLAB[11]

Γενική παρατήρηση:

Ο καλύτερος τρόπος για να μάθουμε να χρησιμοποιούμε το MATLAB είναι ο ίδιος που χρησιμοποιούμε για να μάθουμε την χρήση ποδηλάτου: άμεση χρήση. Στην προκειμένη περίπτωση συμφέρει η άμεση χρήση των εντολών με κατασκευή μικρών προγραμμάτων, καθώς και απομίμηση διαφόρων παραδειγμάτων. Για τον λόγο αυτό δεν παραθέτουμε εδώ πλήρεις ορισμούς των εντολών ή «συναρτήσεων» του πακέτου αλλά τις παρουσιάζουμε μέσα από παραδείγματα. Λεπτομέρειες για κάθε τέτοια εντολή δίνονται από το ίδιο το προγραμματιστικό περιβάλλον του MATLAB με την εντολή:

» **help** <όνομα εντολής>.

Γενικά χαρακτηριστικά:

1. Το MATLAB είναι ένα ενιαίο προγραμματιστικό περιβάλλον που επιτρέπει την εκτέλεση διαφόρων αλγορίθμων της Αριθμητικής Ανάλυσης και την ανταλλαγή δεδομένων μεταξύ τους. Οι βασικοί αλγόριθμοι της Αριθμητικής Ανάλυσης υπάρχουν έτοιμοι σ' αυτό το περιβάλλον υπό μορφήν «συναρτήσεων» (functions).

2. Γιά ειδικές κατηγορίες εφαρμογών, π.χ. βελτιστοποίηση, επεξεργασία σημάτων κ.α., υπάρχουν εκτενείς τέτοιες συλλογές συναρτήσεων που ονομάζονται toolboxes (εργαλειοθήκες).

3. Όλες οι μεταβλητές στο MATLAB θεωρούνται ως πίνακες που οι διαστάσεις τους καθορίζονται κατά την εισαγωγή των δεδομένων. Έτσι η ισότητα A = 3 ορίζει τον πίνακα A διαστάσεων 1x1 με μοναδικό στοιχείο το 3, A=[1,2] ορίζει ένα διάνυσμα γραμμής, A=[1;2] ορίζει ένα διάνυσμα στήλης και A=[1,2;3,4] ορίζει έναν πίνακα 2x2 με διάνυσματα γραμμής A(1,:)=[1,2] και A(2,:)=[3,4].

4. Εξ αιτίας της ευχέρειας με την οποία εκτελεί πράξεις μεταξύ διάνυσματων και πινάκων το πακέτο MATLAB μπορεί να θεωρηθεί σαν γλώσσα προγραμματισμού ειδικά σχεδιασμένη γιά την ταχεία και ακριβή εφαρμογή αλγορίθμων της Αριθμητικής Ανάλυσης.

5. Το MATLAB διαθέτει σήμερα και ένα εκτενές πακέτο συναρτήσεων συμβολικής επεξεργασίας παραστάσεων, το Symbolic Manipulation Toolbox, που βασίζεται στην βιβλιοθήκη προγραμμάτων του πακέτου MAPLE.

Το MAPLE όπως και το MATHEMATICA είναι κατά βάση περιβάλλοντα συμβολικής επεξεργασίας και δημιουργίας γραφικών παραστάσεων με άμεση εκτέλεση των εντολών. Σε νεώτερες εκδόσεις έχουν την δυνατότητα συγγραφής προγραμμάτων και αριθμητικής επίλυσης προβλημάτων, αλλά μάλλον γιά περιστασιακή παρά γιά συνεχή χρήση.

---

[11] Η εισαγωγή αυτή έγινε για ιδία χρήση του συγγραφέα και δεν επιδιώκει την συστηματική παρουσίαση της MATLAB αλλά μόνο την καθοδήγηση στην άμεση χρήση ωρισμένων από τα μαθηματικά εργαλεία που προσφέρει, καθώς και στην γραφική παρουσίαση των αποτελεσμάτων.

ΤΟ ΠΕΡΙΒΑΛΛΟΝ ΕΡΓΑΣΙΑΣ ΤΟΥ MATLAB

Προσοχή:

1. Το σύμβολο ' ↵ ' παριστάνει εδώ την εντολή ENTER.

2. Οι εντολές του MATLAB παρουσιάζονται εδώ με έντονη γραφή για να είναι ευδιάκριτες αν αναζητηθούν. Στο MATLAB εισάγονται με κανονική γραφή.

3. Το σύμβολο '»' είναι το «prompt» του MATLAB, προσδιορίζει δηλαδή την θέση εισαγωγής μίας νέας εντολής προς άμεση εκτέλεση μέσα στο περιβάλλον εργασίας του MATLAB.

4. Πληροφορίες γιά μιαν εντολή παίρνουμε με:

» **help** <όνομα εντολής> ↵

5. Οι υπολογισμοί γίνονται με διπλήν ακρίβεια, δηλ. με 15 σημαντικά δεκαδικά ψηφία αλλά τα αποτελέσματα δίνονται με 4 η 15 δεκαδικά ψηφία αναλόγως του αν έχει δοθεί εντολή:

» **format short** ↵ ή » **format long** ↵

ΤΡΟΠΟΣ ΧΡΗΣΗΣ ΤΟΥ ΠΕΡΙΒΑΛΛΟΝΤΟΣ MATLAB

Τον παρουσιάζουμε συνοπτικά με ενδεικτικά παραδείγματα.

1. » 2+3*4/5 ↵

```
ans=
    4.4000
```

2. Οι μεταβλητές i, j έχουν αρχικά και οι δύο την τιμή $\sqrt{-1}$, αλλά μπορούμε να την αλλάξουμε, π.χ. με την εντολή: i=-2 ↵. Δηλ. 1+5i, 1-3j είναι κατ' αρχήν μιγαδικοί αριθμοί.

3. Εισαγωγή διάνυσματος γραμμής:

```
» x=[1,2,sqrt(3),2:3:8] ↵
  x=
   1.0000  2.0000  1.7321  2.0000  5.0000  8.0000
```

Εδώ ο συμβολισμος a:b:c σημαίνει την ακολουθία με αρχικό όρο a, βήμα b, και τελικό όρο που ΔΕΝ υπερβαίνει την τιμή c.

4. Εισαγωγή διάνυσματος στήλης:

```
» v=[2;3;4;5] ↵ ή
» v=[2:5]' ↵
  v=
   2
   3
   4
   5
```

ΠΑΡΑΤΗΡΗΣΗ: Γενικά και στην εισαγωγή πινάκων τα στοιχεία μίας σειράς χωρίζονται με κόμμα ή κενό, ενώ διαδοχικές σειρές χωρίζονται με ελληνικό ερωτηματικό « ; » (semicolon):

```
» a=[1,2,3,4;5,6,7,8] ↵
  a=
   1  2  3  4
   5  6  7  8
```

Εδώ είναι a(:,2) η δεύτερη στηλη , a(1,:) η πρώτη σειρα, ενώ a(:,2:4) είναι ο υποπίνακας που αποτελείται από την δεύτερη ως την τέταρτη στήλη του a: a(:,2:4)=[2,3,4;6,7,8].

5. Σύνθεση πινάκων

Με χρήση των ιδίων συμβόλων «,» και «;» μπορούμε να συνθέσουμε πίνακες σε πίνακες μεγαλυτέρων διαστάσεων. Για παράδειγμα, αν a = [1,2] και b = [3,4] τότε [a,b] είναι το διάνυσμα [1,2,3,4] ενώ [a;b] είναι ο πίνακας [1,2;3,4] και [a,5;b,6;b,7] είναι ο πίνακας με σειρές [a,5] [b,6] και [b,7]:

```
» c=[a;b] ↵
  c=
   1  2
   3  4
```

```
» d=[a,5;b,6;b,7] ↵
  d=
     1   2   5
     3   4   6
     3   4   7
```

Προσοχή:

Η εντολή » **clear**↵ καταργεί τις διαστάσεις και τις τιμές όλων των (διάνυσματικών) μεταβλητών που έχουν ήδη χρησιμοποιηθεί ώστε να μπορούν να ξαναορισθούν, π.χ. όταν τρέξουμε κάποιο πρόγραμμα. »clear a,b↵ καταργεί μόνο τις διαστάσεις και τιμές των μεταβλητων a,b.

6. Εκτέλεση πράξεων με διανύσματα και πίνακες

<u>α. Γινόμενα:</u>

```
» u=[1,2,3] ; v=[4;5;6] ; x=u*v ↵
x=
   32   (Εσωτερικό Γινόμενο)
```

ΠΡΟΣΟΧΗ. Αν δώσουμε μίαν σειρά από εντολές που χωρίζονται με ελληνικό ερωτηματικό τότε στην οθόνη εμφανίζεται μόνο το τελικό αποτέλεσμα και όχι τα ενδιάμεσα.

```
» x=v*u ↵
```

```
x=
   4    5   6
   8   10  12
  12  15  18    (Δυαδικό Γινόμενο)

» x=u.*v' ↵
x=
   4  10  18   (Γινόμενο κατα συνιστώσα)
```

β. Πηλίκο κατά συνιστώσα

Ανάλογα προκύπτει:

```
x=u./v' ↵
x=
   0.2500  0.4000  0.5000    ( Πηλίκο κατά συνιστώσα)
```

γ. Ύψωση συνιστωσών σε δύναμη:

```
» x=u.^3 ↵
x=
   1  8  27
```

ΠΡΟΣΟΧΗ: Το γινόμενο, το πηλίκο καθώς και η ύψωση σε δύναμη κατα συνιστώσα προκύπτουν εάν πριν ακριβώς από τον τελεστή της πράξεως υπάρχει τελεία.

δ. Συναρτήσεις διανυσμάτων και πινάκων

Οι συναρτήσεις διανυσμάτων και πινάκων λαμβάνονται κατά συνιστώσα. Π.χ.

```
» x=sin(v)  ↵
x=
   0.8415  0.9093  0.1411
```

ε. Ύψωση συνιστωσών σε δυνάμεις:

```
» x=u.^v ↵
x=
   1  32  729
```

Ύψωση κάθε συνιστώσας του v στην αντίστοιχη συνιστώσα του u.

7. Νορμ και άλλες συναρτήσεις πινάκων

```
» a=[-2,3,-4;5,-6,7;-1:1]  ↵
a=    -2  3  -4
       5  -6  7
      -1   0  1
```

» ni=**norm(a,inf)** ↵
ni=
18.0000 (μέγιστο άθροισμα απολύτων τιμών σειράς)

» n1=**norm(a,1)** ↵
n1=
12.0000 (μέγιστο άθροισμα απολύτων τιμών στήλης)
» nf=**norm(a,fro)** ↵
Δίνει την φασματική νορμ, την νορμ του Frobenius.
» n2=**norm(a)** ↵
Δίνει την ευκλείδια νορμ.

» b=**inv(a)** ↵
Δίνει τον αντίστροφο πίνακα του a
» c=a' ↵
Δίνει τον ανάστροφο του a.
» d=**det(a)** ↵
Δίνει την ορίζουσα του a.

» t=**min(a)** ↵
t=
-4.0000
-6.0000
-1.0000 (Διάνυσμα ελάχιστου στοιχείου κάθε σειράς)
» t=**max(min(a))** ↵
t=
-1.0000 (Μέγιστο των ελάχιστων κάθε σειράς)
» [r c]=size(a) ↵
r =
3 (three rows)
c=
3 (three columns)

s=**sum(a)**, p=**prod(a)** ορίζουν αντίστοιχα τα διανύσματα του αθροίσματος και του γινομένου των συνιστωσών κάθε σειράς του πίνακα a.

8. Τμήματα πίνακα και ειδικοί πίνακες

a(:,3), a(2,:) είναι αντίστοιχα η τρίτη στήλη και η δεύτερη σειρά του πίνακα a.

b(:,4:9) είναι ο πίνακας που αποτελείται από τις στήλες 4 ως 9 του πίνακα b.

triu(a), tril(a) είναι αντίστοιχα το άνω και κάτω τριγωνικό μέρος του πίνακα a.

diag(a) ορίζει το διαγώνιο μέρος του a
eye(4) ορίζει τον μοναδιαίο πίνακα διαστάσεων 4x4.
ones(3,4) είναι πίνακας 3x4 με μοναδιαία στοιχεία.

zeros(3,4) είναι πίνακας 3x4 με μηδενικά στοιχεία.

9. Λύση Γραμμικού Συστήματος

x=A\b ορίζει την λύση του συστήματος Ax = b.

10. Εύρεση ιδιοδιανυσμάτων και ιδιοτιμών

d=**eig(a)** είναι το διάνυσμα των ιδιοτιμών του πίνακα a.

[u, d]=eig(a) δίνει τον πίνακα u των ιδιοδιανυσμάτων και τον διαγώνιο πίνακα d των ιδιτιμών του πίνακα a.

ΟΙ ΒΑΣΙΚΕΣ ΣΥΝΑΡΤΗΣΕΙΣ ΤΟΥ MATLAB

abs, round, rem (remainder), sin, cos, tan, asin, acos, atan (αντίστροφες των τριών προηγουμένων), exp, log (φυσικός λογάριθμος), log10 (δεκαδικός λογάριθμος), sign, gcd (μέγιστος κοινός διαιρέτης), lcm (ελάχιστο κοινό πολλαπλάσιο), bessel, beta, gamma, rat (ρητή προσέγγιση), erf (error function),...

ΠΡΟΓΡΑΜΜΑΤΙΣΜΟΣ ΣΤΟ MATLAB

Τα προγράμματα του Matlab λέγονται **m-files** (αρχεία-m) και αποτελούν σειρές εντολών γραμμένες με οποιονδήποτε επεξεργαστή κειμένου, π.χ. τον Microsoft Write. Αυτός ενεργοποιείται και από το περιβάλλον του Matlab με Alt-F, O-pen. Στην παραθυρική έκδοση του MATLAB ο επεξεργαστης κειμένου ενεργοποιείται με την επιλογή File, Create, M-file από την ράβδο επιλογών στο πάνω μέρος της οθόνης.

Το όνομα αυτών των αρχείων έχει την επέκταση .m: **<όνομα>.m**

Οι μορφές των προγραμμάτων αυτών είναι δύο: **scripts** που αντιστοιχούν στις *subroutines* άλλων γλωσσών και **functions**.

Καλούμε ένα αρχείο script απλά με το όνομά του οπότε εκτελούνται κατά σειράν όλες οι εντολές του προγραμμάτος αυτού.

Αν δεν θέλουμε να εμφανισθούν ενδιάμεσα αποτελέσματα πρέπει αυτές οι εντολές να τελειώνουν με semicolon «;» .

Ένα αρχείο-m είναι function όταν η πρώτη γραμμή του περιλαμβάνει την λέξη 'function'. Π.χ.

```
function y=sinmean(x)
m=size(x);
y=sum(sin(x))/m;
end
```

Αν z=[1,2,3] και δώσουμε την εντολή:

» r=sinmean(z) ↵

θα πάρουμε σαν απάντηση τον μέσο όρο των ημιτόνων των συνιστωσών του z:

r=
0.6306

Παρατήρηση: Γιά να «τρέξουμε» ένα m-file αρκεί να δώσουμε την εντολή: » <όνομα αρχείου> ↵

Το αρχείο αυτό αναζητείται τότε στον κατάλογο MATLAB\BIN. Γιά να τρέξουμε όμως το m-file: Fig1.m που βρίσκεται π.χ. στον υποκατάλογο EVASPFIG του BIN πρέπει να δώσουμε με την σειρά τις εντολές/επιλογές: File, Run Script, EVASPFIG\Fig1 ↵

Εισαγωγή Δεδομένων

Η εισαγωγή δεδομένων μπορεί να γίνει και μέσα από το περιβάλλον εργασίας του Matlab πριν καλέσουμε εκεί γιά εκτέλεση το πρόγραμμα. Μπορεί όμως να γίνει εισαγωγή δεδομένων στο περιβάλλον εργασίας και από ένα αρχείο ASCII γραμμένο με τυχόντα επεξεργαστή. Ένα τέτοιο αρχείο γιά την εισαγωγή ενός πίνακα πρέπει να αποτελείται μόνο από μίαν ορθογώνια διάταξη αριθμών (των στοιχείων του πίνακα). Αν το όνομα του αρχείου είναι π.χ. matrix1.dat τότε η εντολή εισαγωγής του είναι:

» **load** matrix1

ΠΡΟΣΟΧΗ: Αρχείο matlab.mat

Τα αριθμητικά δεδομένα (όχι οι εντολές) του ίδιου του περιβάλλοντος εργασίας μπορούν να σωθούν στον σκληρό δίσκο ή σε δισκέττα ώστε να μπορούμε να συνεχίσουμε από εκεί που σταματήσαμε και να έχουμε έτσι διαθέσιμα τα αριθμητικά δεδομένα χωρίς να χρειάζεται επανεισαγωγή τους. Το σώσιμο γίνεται απλά με την εντολή save, και η επανάκληση με την εντολή load.

Καταγραφή Προγράμματος

Ένα αρχείο-m μπορεί να γραφτεί με οποιονδήποτε επεξεργαστή κειμένου αλλά και μεσα από το περιβάλλον εργασίας του Matlab. Προς τούτο αρκεί να κάνουμε τις επιλογές: File, Open (με ενδεχόμενη αλλαγή καταλόγου από c:\Wmatlab σε άλλον), New, M-file. Τότε εμφανίζεται το πλαίσιο του Notepad με ένδειξη Untitled. Αφού γράψουμε τις εντολές του προγράμματος μπορούμε να το σώσουμε με: Save As, ...

Μπορούμε επίσης να αλλάξουμε και τον κατάλογο που σώζεται με την εντολή **cd**. Π.χ.: »cd c:\wmatlab\progr (ενώ αρχικό μόνοπάτι (path) είναι:
c:\wmatlab\bin.

Στοιχεία *της δομής των προγραμμάτων*

Σχόλια:
Σχόλια που παρεμβάλλονται σε ένα πρόγραμμα αρχίζουν με '%'. Π.χ.
%Επίλυση Εξισώσεων

Εισαγωγή Δεδομένων κατά την εκτέλεση προγράμματος:

Var=**input**('κείμενο')

Παρουσιάζει στην οθόνη το 'κείμενο' και ζητεί την εισαγωγή ΑΡΙΘΜΗΤΙΚΩΝ δεδομένων

Var=input('κείμενο', 's')

Εδώ το s ζητεί την εισαγωγή αλυσσίδας ψηφίων (αυτό επιτρέπει την επεξεργασία κειμένων).

ΠΡΟΣΟΧΗ: Μεταβλητές που εισάγονται σε ένα υποπρόγραμμα (script ή function) έχουν κατά βάση μόνο τοπική ισχύ, δηλ. οι τιμές τους δεν διατηρούνται όταν βγούμε από το υποπρόγραμμα. Αυτό γίνεται γιά να μην είναι ανάγκη να αναζητούμε κατά τον προγραμματισμό συνέχεια νέα ονόματα μεταβλητών. Αν κατ' εξαίρεση θέλουμε μία μεταβλητή να έχει τόσο στο κύριο πρόγραμμα όσο και στα υποπρογράμματα την ίδια τιμή τότε πρέπει να την δηλώσουμε ως καθολική με την εντολή:

global <όνομα μεταβλητής>

που μπαίνει τόσο στο κύριο πρόγραμμα όσο και στα υποπρογράμματα.

Βρόγχοι For-end:

```
» for i=1:4 ↵
» x(i) = i^2; ↵
 » end ↵
» x ↵
   x=
      1  4  9  16
```

Βρόγχος while-end:

Μορφή

```
        while {σχέση}
              {εντολές}
         end.
```

Π.χ. αν καλέσουμε το ακόλουθο script sum.m

```
  s=1;
   while s<99
    s=s+1;
   end
```

με

```
» sum ↵
```

και

```
» s ↵
```

παίρνουμε την απάντηση

```
s=
   4851
```

Σχεσιακοί Τελεστές

Προσοχή: Σε περίπτωση συσχέτισης συμβολικών παραστάσεων οι σχεσιακοί τελεστές ανισότητας είναι οι '>', '<','<=' (μικρότερο ή ίσο),'>=' (μεγαλύτερο ή ίσο), '~=' (διάφορο), αλλά ο τελεστής ισότητας είναι '==''. Π.χ. while s==sin(x). Επίσης υπάρχουν οι λογικοί τελεστές: &, |, ~

A&B σημαίνει ταυτόχρονη ισχύ των σχέσεων A και B.

A|B σημαίνει ισχύ είτε της A είτε της B.

~A είναι η άρνηση της A. Π.χ.:

```
if (x~=0)&(x<y)
b=sqrt(y-x)/x;
disp(b);
end
```

Βρόγχοι if-end, if-elseif-end

Μορφές

```
if {σχεση}
  {εντολές};
end.

if {σχέση}
  {εντολές};
elseif {σχέση}
  {εντολές};
else
  {εντολές};
end.
```

Παράδειγμα

```
for k=1:n
  for p=1:m
    if k==p
      z(k,p)=1
      sum=sum+z(k,p);
    elseif k<p
      z(k,p)=-1
      sum=sum+z(k,p);
    else
      z(k,p)=0;
    end
  end
end
```

Εντολές Εκτύπωσης

disp(['κειμενο1', num2str(x),'κειμενο2'])

Τυπώνει το 'κείμενο1', την αριθμητική τιμή του x που η συνάρτηση num2str(x) την μετατρέπει σε αλφαριθμητικό κείμενο, και τέλος το 'κείμενο2'.

fprintf('όνομα αρχείου', 'μορφή εκτύπωσης', κατάλογος μεταβλητών)

Εδώ το όνομα αρχείου μπαίνει μόνο αν επιθυμούμε την καταγραφή των αποτελεσμάτων σε αρχείο και όχι στην οθόνη. Ο κατάλογος μεταβλητών περιλαμβάνει τις μεταβλητές χωρισμένες με κόμματα ενώ η μορφή εκτύπωσης έχει τις ακόλουθες εναλλακτικές επιλογές

%P.Qe γιά παρουσίαση σε εκθετική μορφή

%P.Qf γιά παρουσίαση σε δεκαδική μορφή με σταθερή υποδιαστολή με P δεκαδικά ψηφία (μαζύ με την υποδιαστολή) όπου Q είναι ο αριθμός των ψηφίων μετά την υποδιαστολή.

%P.Qg γιά επιλογή της συντομώτερης από τις παραπάνω παραστάσεις.

/n γιά αλλαγή σειράς. Π.χ.:

fprintf('/n x=%8.2f y=%8.6f, x,y)

Χρονομέτρηση Προγράμματος:

tic-toc: Παρεμβολή της εντολής tic μέσα στο πρόγραμμα αρχίζει την μέτρηση χρόνου, ενώ η toc έχει την τιμή του χρόνου που πέρασε από το τελευταίο tic.

flops: Η εντολή αυτή μετράει τον αριθμό των πράξεων κινητής υποδιαστολής που έχουν γίνει. Με την εντολή flops(0) αρχίζει η μέτρηση από το μηδέν.

Εντολές Γραφικών Παραστάσεων

1. Γραφήματα συναρτήσεων μίας μεταβλητής.

Εντολή plot.

Π.χ. η σειρά εντολών:

```
» x=1:0.5:20; ↵
» y=sin(x)./x; ↵
» plot(x,y) ↵
» xlabel('x');ylabel('y') ↵
```

σχεδιάζει την συνάρτηση sin(x)/x με συνεχή γραμμή παίρνοντας τιμές ανά 0.5. Ενώ με » plot(x,y,'*r') εμφανίζεται μία σειρά από κόκκινα αστεράκια στις αντίστοιχες θέσεις.

Εδώ έχουμε τις επιλογές: *, +, x,., ° γιά τα σύμβολα και:

r=red, y=yellow, m=magenta, c=cyan, g=green, b=blue, w=white,b=black γιά τα χρώματα.

» plot(x,y,x,y,'*') ↵ σχεδιάζει μίαν συνεχή γραμμή και βάζει και αστεράκια στις αντίστοιχες θέσεις.

» plot(x,y,'*',x,z,'+') σχεδιάζει τόσο το γράφημα της συνάρτησης y όσο και αυτό της z, αν αυτή έχει ορισθεί (π.χ. με >>z=x*sin(x);).

Το είδος της γραμμής που σχεδιάζεται μπορεί να επιλεγεί με τα σύμβολα: -solid, -- dash, : dotted, -. dashdot.

Προσδιορισμός των Διαστημάτων επί των αξόνων συντεταγμένων

Τα διαστήματα των αξόνων x- και y- μεταξύ των οποίων παρουσιάζεται το γράφημα προσδιορίζονται κατ' αρχήν αυτόματα. Μπορούμε όμως να προσδιορίσουμε και διαστήματα δικής μας επιλογής με την εντολή: » **axis([xmin,xmax,ymin,ymax])** ↵

Επιστρέφουμε σε αυτόματο προσδιορισμό των διαστημάτων με: »**axis auto** ↵, ενώ αντίθετα: » **axis manual**↵ διατηρεί τα διατήματα που επιλέξαμε και σε επόμενα γραφήματα.

Η εντολή: »axis off↵ αποτρέπει την αυτόματη παρουσίαση υποδιαιρέσεων των αξόνων ενώ: » **axis on**↵ ενεργοποιεί πάλι αυτήν την δυνατότητα.

Εισαγωγή Τίτλων, Επιγραφών, Κειμένου στο Γράφημα

title('κείμενο')	εισάγει το κείμενο στο πάνω μέρος του γραφήματος.
xlabel ('συμβολα')	ορίζει την επιγραφή του οριζόντιου άξονα.
ylabel ('συμβολα')	ορίζει την επιγραφή του κάθετου άξονα.
gtext('κειμενο')	επιτρέπει να τοποθετήσουμε το κείμενο οπουδήποτε στο γράφημα με την βοήθεια του ποντικιού.

ΠΡΟΣΟΧΗ :
Αν έχουμε ελαχιστοποιήσει σε εικονίδιο ένα γράφημα, π.χ., την figure1, και τρέξουμε ένα πρόγραμμα γιά την σχεδίαση κάποιου άλλου γραφήματος τότε φαινομενικά δεν συμβαίνει τίποτα. Στην πραγματικότητα το νέο γράφημα αντικαθιστά το προηγούμενο, αλλά παραμένει ελαχιστοποιημένο και δεν φαίνεται. Πρέπει είτε να κλείσουμε το αρχικό γράφημα εξ αρχής, είτε να ξαναμεγεθύνουμε την figure1 γιά να ιδούμε το νέο γράφημα.

Γράφημα Παραμετρικά Οριζόμενης Καμπύλης

t=0:.001:2*pi; x=cos(3*t); y=sin(2*t); plot(x,y)

Πολλαπλά Γραφήματα με την plot

Υπάρχουν τρείς τρόποι γιά την σχεδίαση πολλών καμπύλων στο ίδιο γράφημα, που τους παρουσίαζουμε με τα ακόλουθα παραδείγματα:

(1): Καταγραφή με την σειρά των συντεταγμένων τους μέσα στην plot.

```
x=0:.01:2*pi;
y1=cos(x); y2=cos(2*x); y3=cos(4*x);
plot(x,y1,x,y2,x,y3)
```

(2): Δημιουργιά πίνακα Y που να περιέχει τις τιμές

```
x=0:.01:2*pi;
Y=[cos(x)', cos(2*x)', cos(4*x)'];
plot(x,Y)
```

(3): Χρήση της εντολής: » **hold**

Η εντολή αυτή κρατάει ανοιχτό το τελευταίο γράφημα και επιτρέπει εγγραφή του νέου μέσα σ' αυτό. Η εντολή: »**hold off** αναστέλλει την » hold.

Εντολή fplot

Η εντολή αυτή έχει την μορφή:

fplot('<όνομα συνάρτησης>',[xmin,xmax])

όπου [xmin, xmax] το διάστημα που διατρέχει η μεταβλητή x, και επιτρέπει να σχεδιασθεί μία προγενέστερα δοσμένη συνάρτηση μεταξύ κάποιων ορίων. Ο αριθμός και η θέση των σημείων που θα χρησιμοποιηθούν προσδιορίζονται εδώ αυτόματα αναλόγως του πόσο απότομα μεταβάλλεται η συνάρτηση. Η fplot κάνει, δηλαδή, αυτόματη επιλογή πυκνώνοντας τα σημεία όπου οι τιμές μεταβάλλονται απότομα. Η μέγιστη και ελάχιστη τιμή της συνάρτησης προσδιορίζονται επίσης αυτόματα και χρησιμοποιούνται έτσι ώστε η καμπύλη να είναι πάντα μέσα στα πλαισία του γραφήματος.

Π.χ. αν το αρχείο f11.m περιέχει την συνάρτηση:

```
function v=f11(x)
v=sin(x.^3);
```

τότε μπορούμε να την σχεδιάσουμε στο διάστημα [2,4] με την εντολή:

```
fplot('f11',[2,4])
```

Μπορεί κανείς να εισαγάγει και άμεσα την συνάρτηση, π.χ. με:

» fplot('sin(x)',[2,4]) ή και » fplot('sin',[2,4])

όπου [2,4] είναι το διάστημα των τιμών του x ενώ το διάστημα των τιμών του y προσδιορίζεται αυτόματα.

Μπορούμε όμως να επιλέξουμε εμείς και το διάστημα των τιμών του y (π.χ. αν η συνάτηση έχει κάπου σημείο ιδιομορφίας) ώστε να ιδούμε μεγεθυμένες κάποιες λεπτομέρειες. Αυτό γίνεται με την παραλλαγή:

» fplot('όνομα συνάρτησης',[xmin,xmax,ymin,ymax])

π.χ.: » fplot('tan', [-6,6,-3,3])

Πολλαπλά Γραφήματα

Με την fplot μπορούμε να σχεδιάσουμε και περισσότερες από μίαν συναρτήσεις μέσα στο ίδιο διάγραμμα. Αυτό γίνεται με:

» fplot('[f1,f2,...,fn]',[xmin,xmax])

όπου f1, f2, f3, ...,fn ονόματα συναρτήσεων.
Π.χ. αν σώσω ως graph1.m το αρχείο:

```
fplot('[sin(abs(x-1)),f12(x),0]',[-5,5])
```

και ως f12,m το αρχείο:

```
function v=f12(x)
y=log(abs(x));
if y>-2 v=y;
else v=-2;
end
```

τότε με: » graph1 ↵
έχω αμέσως το κοινό γράφημα των συναρτήσεων: y1=sin|x-1| και y2=ln|x| ώστε να βρώ τα σημεία διασταυρώσεώς τους, που είναι λύσεις της εξίσωσης:

0 = f(x):= sin|x-1| - ln|x|

Έναλλακτικά μπορώ να δώσω την εντολή
» fplot('[sin(abs(x-1)),log(abs(x)),0]', [-5,5,-2,2]) ↵

Γιά σύγκριση της εντολής plot με την fplot δοκιμάστε, π.χ., το script:

```
x=2:.04:4;
y=f11(x);
subplot(1,2,1);plot(x,y);xlabel('x');ylabel('y');
subplot(1,2,2);fplot('f11',[2,4]);xlabel('x');ylabel('y');
```

Παράλληλη παρουσίαση περισσότερων γραφημάτων: subplot

Ο βασικός τρόπος χρήσης της έχει την μορφή:

subplot(p,q,r);plot(x,fr);title('f(x)');xlabel('x'); ylabel('y');

όπου p ο αριθμός των γραφημάτων ανά σειρά και q ο αριθμός των σειρών με παράλληλα γραφήματα, r ο αύξων αριθμός γραφήματος, ενώ fr είναι η συνάρτηση που απεικονίζεται στο r υπογράφημα.

2. Γραφικές Απεικονίσεις Επιφανειών ή Καμπύλων στον χώρο

Τριδιάστατη καμπύλη: plot3

Παράδειγμα: Έλικα

```
t=.01:.01:20*pi;
x=cos(t);y=sin(t);z=t.^3;
plot3(x,y,z)
```

Εντολές xlabel, ylabel, zlabel μπορούν να προστεθούν και εδώ γιά να ορίσουν τις επιγραφές των τριών αξόνων.

Πολυεδρική προσέγγιση επιφανείας: mesh, surf και meshgrid

Γιά να σχεδιάσουμε την επιφάνεια z=f(x,y) πάνω από μίαν ορθογώνια περιοχή του επιπέδου x-y δημιουργούμε αρχικά μίαν διαμέριση των πλευρών του ορθογωνίου με την εντολή meshgrid.

Η εντολή **mesh(z)** δημιουργεί στην συνέχεια μίαν προοπτική απεικόνιση ραβδογράμματος, όπου τα στοιχεία του πίνακα z είναι τα ύψη των κομβικών σημείων πάνω από το επίπεδο x-y. Π.χ.

```
xx=-2:.2:2;
yy=xx;
[x,y]=meshgrid(xx,yy);
z=exp(-x.^2-y.^2);
mesh(z)
```

Αντίστοιχα η εντολή **surf(z)** δημιουργεί μίαν τριδιάστατη πολυεδρική επιφάνεια με ύψη των κορυφών τα στοιχεία του πίνακα z. Π.χ.:

```
[x,y]=meshgrid(-2:.2:2,-2:.2:2);
z=exp(-x.^2-y.^2);
surf(z)
```

3. Προσδιορισμός Προοπτικής Άποψης από ωρισμένη θέση ή κατεύθυνση: view

Η εντολή αυτή μπαίνει μετά από οποιανδήποτε εντολή παρουσίασης τριδιαστάτου γραφήματος γιά να αλλάξει την κατεύθυνση παρατήρησης. Η βασική μορφή της είναι:

view(AL,EL)

όπου AL η γωνία περιστροφής του γραφήματος μέσα στο επίπεδο x-y, και EL η γωνία παρατήρησης ως προς τον άξονα των z. Οι προκαθορισμένες τιμές αυτών των γωνιών σε όλες τις εντολές τριδιάστατης απεικόνισης είναι: AL=-37.5, EL= 30 (μοιρες). Δοκιμάστε π.χ. την εντολή view(-20,60).

Παράδειγμα:

```
» xx=-2:0.2:2;
» yy=xx;
» [x,y]=meshgrid(xx,yy);
» z=cos(x.^2+y.^2);
» mesh(z) ή >>surf(z)
» view(-20,60)
```

4. Χρωματικό Προφίλ: colormap (χρωματικός συνδυασμός)

Η παρουσίαση επιφανειών στον χώρο γίνεται όχι μόνο με σχεδίαση δικτυωτού που περιγράφει μίαν πολυεδρική προσέγγιση, αλλά και με εναλλαγή χρωμάτων που κάνουν πιο εμφανείς τις διάφορες υψομετρικές ζώνες. Το αρχικό φάσμα χρωμάτων είναι: κόκκινο-κίτρινο-πράσινο-μπλέ- μωβ- κόκκινο. Μπορεί όμως να αλλαχθεί με την εντολή **colormap**.

Π.χ. προσθήκη της εντολής colormap(cool) μετά από τις παραπάνω εντολές αλλάζει το φάσμα σε γάλαζιο-μωβ.

Χρωματικοί Συνδυασμοί είναι: hsv(αρχικός), hot, cool, jet, pink, copper, flag, gray, bone, prism, white.

5. Χρωματική Σκίαση: shading

shading flat: Προσθήκη της εντολής αυτής μετά τις παραπάνω εντολές διώχνει τις γραμμές του δικτυωτού.

shading interp: (=interpolated) κάνει την μετάβαση από το ένα χρώμα στο άλλο σταδιακή, δηλ. πιο ομαλή.

shading faceted: επαναφέρει τις γραμμές δικτυώσεως.

Εκτύπωση Γραφικών

Η εντολή »**print** μόνη της στέλνει ένα αντίγραφο υψήλης ανάλυσης του τρέχοντος γραφήματος στον εκτυπωτή. Ο προσδιορισμός των παγίων ρυθμίσεων γιά την εκτέλεση της εντολής αυτής γίνεται με το αρχείο printopt.m. Αν χρειάζεται μπορούμε να αλλάξουμε αυτό το αρχείο με βάση τις οδηγίες που παίρνουμε με την εντολή: >>help printopt.

Η εντολή »print συνοδευόμενη από ένα όνομα αρχείου σώζει το γράφημα σε ένα αρχείο με αυτό το όνομα. Αν δεν δώσουμε προέκταση στο όνομα τότε προστίθεται αυτόματα μία κατάλληλη. Π.χ. η εντολή »print lorenz δημιουργεί ένα αρχείο postscript: lorenz.ps. Αν υπάρχει ήδη αρχείο με το όνομα που διαλέξαμε

τότε το νέο γράφημα θα αντικαταστήσει το παλιό. Γιά να σώσουμε το νέο γράφημα παράλληλα με το παλιό σε ένα αρχείο με δεδομένο όνομα δίνουμε την παραλλαγμένη εντολή: » **print -append** <όνομα αρχείου>.

ΑΡΙΘΜΗΤΙΚΟΑΝΑΛΥΤΙΚΕΣ ΣΥΝΑΡΤΗΣΕΙΣ ΤΗΣ MATLAB

Σημειωση: Γιά να ιδείτε περισσότερες λεπτομέρειες γιά τον τρόπο χρήσης των συναρτήσεων αυτών χρησιμοποιήστε την εντολή:

»help <όνομα συνάρτησης> ↵.

Παρατήρηση: Οι παρακάτω «συναρτήσεις» αντιπροσωπεύουν αλγορίθμους της Αριθμητικής Ανάλυσης. Η συνάρτηση που θέλουμε να υποστεί επεξεργασία προσδιορίζεται στην αντίστοιχη θέση με τρείς τρόπους:

(α) με το όνομά ενός αρχείου-m που την περιγράφει βάλμένο μέσα σε απλά εισαγωγικά π. χ. 'lorenz'.

(β) με μίαν άμεση συναρτησιακή έκφραση, όπως sin (x) βαλμένη μέσα σε απλά εισαγωγικά.

(γ) με μίαν συνάρτηση εισαγόμενη άμεσα με την βοήθεια της εντολής inline.

inline (παράσταση) είναι μία εντολή που ερμηνεύει την εντός παρενθέσεων παράσταση σαν συνάρτηση του MATLAB. Οι παράμετροι της συνάρτησης αυτής προσδιορίζονται αυτόματα με αναζήτηση των συμβόλων ή συμβολοσειρών που παριστάνουν μεταβλητές.

Προσοχή: Μερικές από τις παρακάτω «συναρτήσεις» του MATLAB, όπως οι quad, quad8, και feval δέχονται μίαν συνάρτηση προς επεξεργασία μόνο αν αυτή περιγράφεται σε ένα αρχείο-m ή εισάγεται κατ' ευθείαν με την εντολή inline. Έτσι προκύπτει π. χ. ο ακόλουθος «διάλογος»:

```
»y= feval('x.^2-5*x+2',2.5) ↵
??? Cannot find function 'x.^2-5*x+2'
»y= feval(inline('x.^2-5*x+2'),2.5) ↵
y= -4.25
```

x=feval('F', a) Προδιορίζει την τιμή της συνάρτησης F στην θέση a.

x=fzero('FUN',x0,tol), Προσδιορίζει ένα σημείο μηδενισμού της συνάρτησης μίας μεταβλητής FUN (βλέπε την προηγούμενη παρατήρηση). x0 είναι το σημείο εκκίνησης της αναζήτησης (με την μέθοδο του Brent) και tol η ακρίβεια της λύσης. Π. χ. »fzero('sin(x)',7) ↵ δίνει την απάντηση: ans = 6.2832.

x=fsolve('FUN',x0) Προσδιορίζει ένα σημείο μηδενισμού του συστήματος μη γραμμικών εξισώσεων FUN ξεκινώντας από το σημείο x0.

x=fmin('F',x1,x2) Προδιορίζει στο διάστημα [X0,X1] μίαν θέση σχετικού ελαχίστου της συνάρτησης F, που είναι συνάρτηση μίας μεταβλητης.

x=fmins('F',x0) Προσδιορίζει ένα τοπικό ελάχιστο της συνάρτησης πολλών μεταβλητών F που να βρίσκεται κοντά στο σημείο x0 (Το x0 είναι εδώ ένα διάνυσμα που ορίζει την θέση εκκίνησης της αναζήτησης.

I = quad('F',a,b,tol) Υπολογίζει το ολοκλήρωμα της συνάρτησης F στο διάστημα [a,b] με τον σύνθετο κανόνα Simpson. Η παράμετρος tol προσδιορίζει τα ανεκτά όρια ακριβείας: tol = [RelTol, AbsTol] όπου RelTol η σχετική ακρίβεια και AbsTol η απόλυτη. Εναλλακτικές τιμές τους είναι RelTol = 1e-4, δηλαδή 10^{-4} και AbsTol = 1e-6, δηλαδή 10^{-6}.

I = quad8('F',a,b,tol) Υπολογίζει το ολοκλήρωμα με τον κανόνα Newton-Cotes 8 σημείων.

I = trapz('F',a,b,tol) Υπολογίζει το ολοκλήρωμα με τον σύνθετο κανόνα τραπεζίου.

[t,y]=ode45('F',[t0,tfin],y0,options) Εφαρμόζει την μέθοδο Runge-Kutta-Fehlberg 3ης/4ης τάξεως για να προσδιορίσει την λύση του προβλήματος αρχικών τιμών: y'=F(t,y), y(t0)=y0 στο διάστημα [t0, tfin]. Η παράμετρος options προσδιορίζει τα ανεκτά όρια ανοχής σχετικού και απολύτου σφάλματος, π. χ.: options=odeset('RelTol',1e-4, 'AbsTol',1e-5) αν παραλειφθεί τότε οι εναλλακτικές τιμές είναι 1e-4 για την σχετική ανοχή RelTol και 1e-6 για την απόλυτη ανοχή AbsTol. Για να προσδιορίσουμε την λύση στις θέσεις t0, t1, t2, …, tfin αντικαθιστούμε στην παραπάνω εντολή το χρονικό διάστημα με **[t0, t1, t2, ..., tfin]**.

ode23 Εφαρμόζει την μέθοδο Runge-Kutta-Fehlberg 2ας/3ης τάξεως γιά την επίλυση του προβλήματος αρχικών τιμών.

ode23p Κάνει χρήση της ode23 και παρουσιάζει την λύση γραφικά.

Π.χ. γιά να λύσουμε το σύστημα Δ.Ε. του Lorenz:

$$x' = 10(y-x),\ y' = rx-y-xz,\ z' = xy-8z/3$$

χρησιμοποιούμε το script:

```
%Λύση των Δ.Ε. Lorenz
global r
r=input('δώσε τιμή γιά το r');
maxtime=input('δώσε τον τελικό χρόνο');
acc=input('δώσε την απαιτούμενη ακρίβεια');
x0=[-7.69,-15.61,90.39]';
[t,x]=ode45('lor',0,maxtime,x0,acc);
figure(1);plot(t,x);xlabel('time');ylabel('x');
figure(2);plot(x(:,1)x(:,3));xlabel('x');ylabel('z');
```

που καλεί την συνάρτηση lor:

```
function fv=lor(t,x)
% x=x(1), y=x(2), z=x(3)
global r
fv=zeros(3,1);
fv(1)=10*(x(2)-x(1)); fv(2)=r*x(1)-x(2)-x(1)*x(3);fv(3)=x(1)*x(2)-
8*x(3)/3;
```

Εδώ « global r » σημαίνει οτι η τιμή που δίνουμε στο r δεν ισχύει μόνο γιά το αρχικό script αλλά και γιά την function. figure(1), και figure(2) είναι δύο γραφήματα που μπορούμε να τα 'ανοίξουμε' έναλλάξ γιά να τα παρατηρήσουμε. Γιά r > 24.7 συνήθως έχουμε χαοτική συμπεριφορά.

ΒΙΒΛΙΟΓΡΑΦΙΑ

G.Lindfield - J.Penny: Numerical Methods Using MATLAB, Ellis Horwood, N.Y., 1995

Pärt-Enander, Eva – Sjöberg, Anders: The MATLAB 5 Handbook, Addison-Wesley, 1999.

K. Sigmon: MATLAB Primer, fourth edition, CRC Press, Boca Raton, 1994.

ΕΠΕΞΕΡΓΑΣΙΑ ΣΥΜΒΟΛΙΚΩΝ ΠΑΡΑΣΤΑΣΕΩΝ ΜΕ ΤΟ MATLAB

Το MATLAB μπορεί να κάνει και συμβολική (αντί γιά αριθμητική) επεξεργασία μαθηματικών παραστάσεων. Αυτό γίνεται με τις εντολές που περιλαμβάνει το Symbolic Math Toolbox που χρησιμοποιεί τμήματα του προγράμματος συμβολικής επεξεργασίας MAPLE.

Ορισμός συμβολικών αντικειμένων

Προϋπόθεση γιά να γίνει συμβολική επεξεργασία είναι το να δηλωθεί η αντίστοιχη παράσταση σαν συμβολικό αντικειμενο. Αυτό γίνεται με την συνάρτηση/εντολή **sym**.

Π.χ. οι εντολές

» x = sym('x')

» **syms('a','b','c')** ή ισοδύναμα » **syms a b c**

ορίζουν αντίστοιχα συμβολικές μεταβλητές ή παραμέτρους x, a, b, c. Κάθε παράσταση που περιέχει αυτές τις μεταβλητές (π.χ. η » f=a + b*sin(x)) θεωρείται στην συνέχεια επίσης σαν συμβολική παράσταση, χωρίς να έχει δηλωθεί έτσι. Αυτό επιτρέπει την αλγεβρική ή αναλυτική της επεξεργασία. Π.χ. η παράγωγος και το ολοκλήρωμα της f προκύπτουν ως εξής:

» syms a b c x ↵

» f = a + b*sin(c*x); ↵

» diff(f) ↵

```
ans =
b*cos(c*x)*c
» g=int(f) ↵
g =
a*x – cos(c*x)/c *b
» pretty(g) ↵          cos(c x) b
                  a x - ----------
                              c
```

Σαν συμβολικές παραστάσεις μπορούν να δηλωθούν και αριθμητικές παραστάσεις. Π.χ. y = sym(1/3). Αυτό επιτρέπει την εκτέλεση πράξεων εν γένει χωρίς σφάλματα στρογγυλεύσεως, γιατί το MATLAB παριστάνει εσωτερικά συμβολικές παραστάσεις σαν παραστάσεις που περιέχουν συμβολικά αντικειμένα.

Κατά τον ίδιο τρόπο μπορούν να μετατραπούν σε συμβολικές παραστάσεις και πίνακες αριθμητικών τιμών. Π.χ.:

```
» h=hilb(3) ↵
h=
    1.0000   0.5000   0.3333
    0.5000   0.3333   0.2500
    0.3333   0.2500   0.2000
» H=sym(h) ↵
H=
    [  1,   1/2,  1/3]
    [1/2,  1/3,  1/4]
    [1/3,  1/4, 1/ 5]
```

Αριθμητικά δεδομένα είτε συμβολικού είτε αριθμητικού τύπου έχουν διάφορες μορφές παράστασης. Η παρουσίαση τέτοιων εναλλακτικών μορφών γίνεται με τις εντολές >> help sym και >>help format.

Ο τύπος αριθμητικών δεδομένων (συμβολικός ή αριθμητικός) μπορεί να ανιχνευθεί με την εντολή class. Π.χ.

```
» class(h) ↵
ans=
      double
» class(H) ↵
ans=
      sym
```

'double' σημαίνει εδώ "double precision", γιάτι με αυτήν την ακρίβεια γίνονται εσωτερικά οι πράξεις έστω και αν τα αποτελέσματα παρουσιάζονται (σε μορφή format short) μόνο με τέσσερα σημαντικά ψηφία μετά την υποδιαστολή.

Συμβολικές Μεταβλητές

Όταν μια συμβολική παράσταση περιέχει περισσότερες από μία συμβολικές μεταβλητές μόνο μία από αυτές λαμβάνεται σαν ανεξάρτητη μεταβλητή. Αν δεν δηλωθεί ποιά από αυτές είναι η ανεξάρτητη τότε το MATLAB θεωρεί σαν ανεξάρτητη μεταβλητή την πλησιέστερη προς το 'x'.

Απλοποίηση γραφής

Το MATLAB θεωρεί κάθε εντολή της μορφής: {εντολή} {όρισμα} σαν ισοδύναμη με: {εντολή}(‘{όρισμα}’). Π.χ. «syms a b» είναι ισοδύναμο με «syms(‘a’,’b’)».

Αλγεβρικές Πράξεις

1. Προσδιορισμός Αριθμητή-Παρονομαστή: Εντολή **numden**

Αν η εισαγόμενη παράσταση είναι άθροισμα ρητών παραστάσεων τότε με την εντολή numden δημιουργείται μια ενιαία παράσταση και μας δίνεται ο αριθμητής (numerator) και ο παρονομαστής (denominator) της:

```
» h= (x^2+3)/(2*x-1)+3*x/(x-1); ↵
» [n,d]=numden(h) ↵
n=
   x^3+5*x^2-3
d=
  (2*x-1)*(x-1)
```

Ο διαχωρισμός πάλι σε επί μέρους κλάσματα γίνεται αν δώσουμε εντολή διαίρεσης των n και d:

```
» h2=n/d ↵
h2=
    (x^2+3)/(2*x-1)+3*x/(x-1);
```

Η εντολή [n,d]=numden(M) μπορεί να εφαρμοσθεί και σε έναν πίνακα, M, με ρητες παραστάσεις. n είναι τότε ο πίνακας των αριθμητών και d ο πίνακας των παρονομαστών. Αντίστοιχα γιά να ανακτήσουμε τον πίνακα M πρέπει να δώσουμε την εντολή: n./d .

Οι βασικές αλγεβρικές πράξεις με συμβολικές παραστάσεις εκτελούνται κατά τον συνήθη τρόπο. Άλλες εντολές είναι:

2. Προσδιορισμός της συνάρτησης f(g(x)): Εντολή **compose(f,g)**

```
» syms x u v; ↵
» f=1/(1+x^2); ↵
» g=sin(x); ↵
» compose(f,g) ↵
 ans =
       1/(1+sin(x)^2)
```

3. Αντιστροφή της Συναρτήσεως g: **finverse(g)**

```
» finverse(x^2) ↵
```

```
ans =
      x^(1/2)
```

Η συνάρτηση αυτή προσδιορίζει μόνο μιαν αντίστροφη έστω και αν η αντίστροφη δεν είναι μόνοσημαντα ωρισμένη.

4. Άθροιση Συμβολικής Σειράς: **symsum**

symsum(f) δίνει το άθροισμα $\sum_{k=0}^{x-1} f(k)$.

symsum(f,s) δίνει το άθροισμα $\sum_{k=0}^{s-1} f(k)$.

symsum(f,a,b) δίνει το άθροισμα $\sum_{k=a}^{k=b} f(k)$. Π.χ.:

```
» symsum(n^2) ↵
ans =
      1/3*n^3-1/2*n^2+1/6*n
» factor(ans) ↵
ans =
      1/6*n*(2*n-1)*(n-1)
» symsum(1/(2*n-1)^2),1,inf) ↵
ans =
       1/8*pi^2
```

5. Αντικαταστάση μεταβλητής με παράσταση: **subs(f,old,new)**

Με την εντολή αυτήν μπορούμε να αντικαταστήσουμε στην συμβολική παράσταση f την συμβολική μεταβλητή old με άλλη συμβολική μεταβλητή, παράσταση, πίνακα ή αριθμητική τιμή, την new.

```
» syms a b c s x r↵
» f=a*x^2+b*x+c↵
f=
   a*x^2+b*x+c
» subs(f,x,s) ↵
ans=
      a*s^2+b*s+c
» subs(f,a,[r;s]) ↵
ans =
       [r*x^2+b*x+c]
       [s*x^2+b*x+c]
» g=3*x^2+5*x-4↵
g =
    3*x^2+5*x-4
» h=subs(g,x,2) ↵
h =
    18
» class(h) ↵
```

```
ans =
        sym
```

6. Μετατροπή συμβολικής σε αριθμητική σταθερά: **double(.)**

Όπως φαίνεται παραπάνω η σταθερά h = 18 είναι συμβολική σταθερά και όχι αριθμητική τιμή. Η μεταβολή της σε αριθμητική τιμή γίνεται με την εντολή double(h):

```
» double(h) ↵
ans =
       18
» class(ans) ↵
ans =
        double
```

Η μετατροπή αυτή είναι π.χ. απαραίτητη όταν λύνουμε μίαν εξίσωση. Π.χ. γιά την εξίσωση x^2-x+1=0 προκύπτει:

```
» syms x ↵
» t=solve(x^2-x+1) ↵
t =
   [1/2*5^(1/2)+1/2]
   [1/2-1/2*5^(1/2)]
» double(t) ↵
ans =
        1.618
       -0.618
» format long ↵
» ans ↵
ans =
        1.61803398874990
       -0.61803398874989
```

6. Άλλες αλγεβρικές πράξεις φαίνονται συνοπτικά στα παρακάτω παραδείγματα:

```
» syms x ↵
» f= (x^2-1)*(x-2)*(x-3) ↵
» collect(f) ↵
ans =
        x^4-5*x^3+5*x^2+5*x-6
» horner(ans) ↵
ans =
       -6+(5+(5+(x-5)*x)*x)*x
» factor(ans) ↵
ans =
        (x-1)*(x-2)*(x-3)*(x+1)
» expand(f) ↵
ans =
        x^4-5*x^3+5*x^2+5*x-6
```

simplify(.):

Η εντολή simplify(.) απλοποιεί παραστάσεις όλων των ειδών χρησιμοποιώντας ταυτότητες. Π.χ.:

```
» syms x y a ↵
» simplify(sin(x)^2+3*x+cos(x)^2-5) ↵
ans =
        -4+3*x
» simplify((-a^2+1)/(1-a)) ↵
ans =
        a+1
```

simple(.):

Η εντολή αυτή προσπαθεί να βρει την ισοδύναμη παράσταση ελαχίστου μήκους.

pretty(.):

Η εντολή αυτή, π.χ. pretty(ans), γράφει τα αποτελέσματα σε πιό ευανάγνωστη μορφή. Π.χ. βάζει στα κλάσματα τον παρονομαστή κάτω (αντί γιά δίπλα) από τον αριθμητή.

Μεταβολή Αριθμητικής Ακρίβειας

Το Symbolic Math Toolbox χρησιμοποιεί κατ' αρχήν ακρίβεια 32 ψηφίων μετά την υποδιαστολή. Η ακρίβεια αυτή μπορεί όμως να μεταβληθεί σε n ψηφία με την εντολή **digits(n)**. Ο τρέχων αριθμός ψηφίων παρουσιάζεται αν δώσουμε απλά την εντολή **digits**.

Η εντολή **vpa(.)**, π.χ. **vpa(pi)** (από τα αρχικά των λέξεων Variable Precision Arithmetic) υπολογίζει στην συνέχεια την οποιαδήποτε συμβολική παράσταση με την προσδιωρισμένη ακρίβεια. Μπορεί όμως κάνεις να πετύχει απ' ευθείας το ίδιο αποτέλεσμα με την εντολή **vpa(s,n)**, όπου s η συμβολική παράσταση και n ο ζητούμενος αριθμός ψηφίων. Π.χ.

```
» vpa(sin(.5),20) ↵
ans =
        .47942553860420300538
```

Αναλυτικές Πράξεις

1. Διαφόριση: **diff(.)**

Η διαφόριση έχει τέσσερις μορφές:

diff(f) : Διαφορίζει την f ως προς x.
diff(f,a): Διαφορίζει την f ως προς a.
diff(f,2): Διαφορίζει την f δυο φορες ως προς x.
diff(f,a,2): Διαφορίζει την f δυο φορες ως προς a.

```
» syms a b c x ↵
» f=a*x^2+b*x+c ↵
f =
    a*x^2+b*x+c
```

```
» diff(f) ↵
ans =
      2*a*x+b
» diff(f,a) ↵
ans =
      x^2
» diff(f,a,2) ↵
ans =
       0
```

Η εντολή diff μπορεί να εφαρμοσθεί και σε πίνακες:

```
» F=[a*x,b*x^2; b*x,c] ↵
F =
    [ a*x, b*x^2]
    [ b*x,       c]
» diff(F) ↵
ans =
      [ a, 2*b*x]
      [ b,       0]
```

ΠΡΟΣΟΧΗ: Η εντολή diff χρησιμοποιείται και στο βασικό MATLAB για τον υπολογισμό διαφορών των διαδοχικών συντεταγμένων ενός αριθμητικού διανύσματος ή πίνακος.

2. Ολοκλήρωση: **int(.)**

Η ολοκλήρωση είναι αντίστροφη της προηγουμένης διαδικασίας. Δηλ. int(f) είναι η λύση του προβλήματος diff(F) = f. Έχει και αυτή διάφορες μορφές:

int(f) : Προσπαθεί να βρεί το ολοκλήρωμα της f ως προς την αρχικά υποτιθέμενη μεταβλητή x.

int(f,s): Προσπαθεί ολοκλήρωση ως προς την συμβολική μεταβλητή s.

int(f,a,b): Προσπαθεί να βρει συμβολική έκφραση γιά το ωρισμένο ολοκλήρωμα ως προς x από a ως b.

int(f,s,a,b): Προσπαθεί να βρει συμβολική έκφραση γιά το ωρισμένο ολοκλήρωμα ως προς s από a ως b.

Τα a, b μπορούν να είναι εδώ είτε συμβολικοί αριθμοί είτε συμβολικές μεταβλητές.

```
» syms x s a b ↵
» f=sin(s+2*x) ↵
f =
    sin(s+2*x)
» int(f) ↵
ans =
      -1/2*cos(s+2*x)
» int(f,s) ↵
ans =
      -cos(s+2*x)
» int(f,pi/2,pi) ↵
```

```
ans =
        -cos(s)
» int(f,s,pi/2,pi) ↵
ans =
        2*cos(x)^2-1-2*sin(x)*cos(x)
» g=simple(int(f,a,b)) ↵
g =
    -1/2*cos(s+2*b)+1/2*cos(s+2*a)
```

Η εντολή simple χρησιμοποιήθηκε εδώ γιά να απλοποιήσει το αποτέλεσμα της ολοκλήρωσης.

Όπως και η διαφόριση μπορεί και η συμβολική ολοκλήρωση να εφαρμοσθεί σε διανύσματα και πίνακες.

Ανάλυση ρητής παραστάσεως

Οι εντολές int και diff μπορούν να χρησιμοποιηθούν σε συνδυασμό γιά την ανάλυση μίας ρητής παραστάσεως r(x) σε επί μέρους κλάσματα. Ο υπολογισμός της R = int(r) κάνει αναγκαία την ανάλυση σε τέτοια κλάσματα. Στην συνέχεια η diff(R) = r μας δίνει την r αναλυμένη σε τέτοια κλάσματα. Μπορούμε βέβαια και να συνδυάσουμε τις εντολές σε μία: » **diff(int(r))**. Π.χ.

```
» syms x↵
» r=(x^2-x+2)/(x^2-3*x+2) ↵
» int(r) ↵
ans =
       x-2*log(-1+x)+4*log(x-2)
» diff(ans) ↵
ans =
       1-2/(-1+x)+4/(x-2)
```

3. Ανάπτυγμα Taylor: **taylor(.)**

Η εντολή αυτή έχει τις μορφές:

taylor(f): Ανάπτυγμα taylor της f(x) γύρω από το x=0 μέχρι τον όρο x^5.

taylor(f,n): Ανάπτυγμα γύρω από το x=0 μέχρι τον όρο x^(n-1).

taylor(f,a,n): Ανάπτυγμα taylor γύρω από το σημείο x=a μέχρι τον όρο (x-a)^(n-1). Π.χ.:

```
» taylor(sin(x),pi/2,6) ↵
ans =
        1-1/2*(x-1/2*pi)^2-1/24*(x-1/2*pi)^4
```

Γραφική Παράσταση Συμβολικών Παραστάσεων μίας Μεταβλητής

Αυτή γίνεται με την εντολή **ezplot** (= easy plot). Οι μορφές της είναι οι ακόλουθες:

ezplot(f) : Σχεδιάζει την f στο διάστημα x∈[-2*pi, 2*pi].

ezplot(f,xmin,xmax): Σχεδιάζει την f στο διάστημα x∈[xmin,xmax].
ezplot(f,[xmin,xmax]): Σχεδιάζει την f στο διάστημα x∈[xmin,xmax].
Π.χ. » ezplot('erf(x)') ή » ezplot erf(x).

Στην γραφική παράσταση μπορούν να προστεθούν και άλλα στοιχεία, όνομασία αξόνων, τίτλος γραφήματος κ.ο.κ. ακριβώς όπως και σε γραφήματα αριθμητικών παραστάσεων.

Υπάρχει επίσης η εντολή **zoom on** που επιτρέπει την χρήση του ποντικιού γιά μεγέθυνση ή σμύκρινση οποιουδήποτε τμήματος του διαγράμματος.

Επίλυση Αλγεβρικών Εξισώσεων

Η επίλυση μίας εξίσωσης f(x)=c γίνεται με την εντολή **solve(f(x)-c) .** Το ίδιο μπορεί να επιτευχθεί χωρίς ορισμό συμβολικών αντικειμένων με εισαγωγή της εξίσωσης εντός εισαγωγικών: **solve('f(x)=c')**. Π.χ.

```
» syms a b c x ↵
» solve(a*x^2+b*x+c) ↵
x =
    [1/2/a*(-b+(b^2-4*a*c)^(1/2)]
    [1/2/a*(-b-(b^2-4*a*c)^(1/2)]
```

Το ίδιο αποτέλεσμα προκύπτει και με την εντολή:

```
» solve('a*x^2+b*x=(-c)') ↵
```

Ανάλογα λύνονται και μη πολυωνυμικές εξισώσεις. Π.χ.:

```
» t=solve(tan(x)-sin(2*x)) ↵
t =
    [        0]
    [       pi]
    [  1/4*pi]
    [ -3/4*pi]
» solve('z*sin(x)=3*y','y') ↵
ans =
      1/3*z*sin(x)
```

Επίλυση Συστήματος Εξισώσεων

Αυτή γίνεται με την εντολή: **[a1, a2,...,aN]=solve(f1,f2,...,fN)** ή και **[a1, a2,...,aN]=solve(f1,f2,...,fN,v1,v2,...,vN)** όπου v1, v2, …, vN είναι οι άγνωστοι. Π.χ.:

```
» syms x y ↵
» [a1 a2]=solve(x^2+x*y+y-3, x^2-4*x+3) ↵
X =
    [  1]
    [  3]
Y =
    [   1]
    [-3/2]
```

Η εντολή **S=solve(f1,f2,...,fN)** καταχωρεί την λύση στην δομή S που κάθε πεδίο της, S.vi, αντιστοιχεί σε έναν άγνωστο, vi.

```
» S=solve(x^2+x*y+y-3, x^2-4*x+3) ↵
S =
    X: [  2X1 sym]
    Y: [  2x1 sym]
» S.x ↵
ans =
     [  1]
     [  3]
» S.y ↵
ans =
     [   1]
     [-3/2]
```

Επίλυση Διαφορικών Εξισώσεων

Αυτή γίνεται με την εντολή **dsolve(.)** όπου η παράσταση εντός παρενθέσεων πρέπει οπωσδήποτε να περιέχει το σύμβολο ισότητας '='. Οι παράγωγοι ορίζονται με πρόταξη κεφαλαίου γράμματος D γιά την πρώτη παράγωγο, D2 γιά την δεύτερη, D3 γιά την τρίτη, κοκ. Δηλ. η παράσταση 'D2y=0' παριστάνει την εξίσωση: $d^2y/dt^2=0$. Η ανεξάρτητη μεταβλητή μπορεί να ορισθεί άλλως θεωρείται ότι είναι η t. Π.χ.:

```
» dsolve('Dy=1+y^2') ↵

ans =
        tan(t-c1)
» dsolve('Dy=1+y^2, y(0)=1') ↵
ans =
       tan(t+1/4*pi)

» dsolve('Dy=1+y^2, y(0)=1', 'x') ↵
ans =
       tan(x+1/4*pi)

» y=dsolve('D2y=cos(2*t)-y, Dy(0)=0, y(0)=1') ↵
y =
    -2/3*cos(t)^2+1/3+4/3*cos(t)
» y=simple(y) ↵
y =
    -1/3*cos(2*t)+4/3*cos(t)
» ezplot(y,[-6 2]) ↵
```

Η λύση συστημάτων διαφορικών εξισώσεων γίνεται με ανάλογο τρόπο. Π.χ.:

```
» [f,g]=dsolve('Df=3*f+4*g, Dg=-4*f+3*g') ↵
f =
```

```
    exp(3*t)*cos(4*t)*c1+exp(3*t)*sin(4*t)*c2
g =
    -exp(3*t)*sin(4*t)*c1+exp(3*t)*cos(4*t)*c2

» S=dsolve('Df=3*f+4*g, Dg=-4*f+3*g, f(0)=0, g(0)=1') ↵
S =
    f: [1x1 sym]
    g:[1x1 sym]
» S.f ↵
f =
    exp(3*t)*sin(4*t)
» S.g ↵
g =
    exp(3*t)*cos(4*t)
```

ΣΥΜΒΟΛΙΚΗ ΓΡΑΜΜΙΚΗ ΑΛΓΕΒΡΑ

Οι δυνατότητες συμβολικής επεξεργασίας διάνυσματων, πινάκων και γραμμικών εξισώσεων θα παρουσιάσθουν πιό απλά με παραδείγματα:

```
» syms a b c s t ↵
» A=[a,b;b,c] ↵
A =
    [a, b, c]
    [b, c, a]

» [m,n]=size(A) ↵
m =
    2
n =
    3
» A(2,1) ↵ (Στοιχείο της 2ης σειράς και 1ης στήλης).
A =
    b

» A(2,2)=sym('d') ↵    (Αλλαγή του στοιχείου A(2,2)).
A =
    [a, b, c]
    [b, d, a]
» A(3,3)=sym('e') ↵ (Επέκταση του πίνακα A με προσθήκη του A(3,3)).
A =
    [a, b, c]
    [b, d, a]
    [0, 0, e]
```

Οι εντολές **diag(A,k), triu(A,k), tril(A,k)** δίνουν αντίστοιχα την k διαγώνιο, τον άνω τριγωνικό πίνακα από την k διαγώνιο και μετά, και τον κάτω

τριγωνικό πίνακα από την k διαγώνιο και μετά. Η κυρία διαγώνιος έχει εδώ τον αριθμό k=0. Η παράμετρος k μπορεί να παραλειφθεί αν αναφερόμαστε στην κυρία διαγώνιο.

```
» triu(A,1) ↵
ans =
      [0, b, c]
       [0, 0, a]
       [0, 0, 0]
```

Οι περισσότερες αλγεβρικές πράξεις γίνονται εδώ όπως συνήθως. Π.χ.:

```
» syms a b c d t↵
» B=[a,b;c,d] ↵
B =
    [a, b]
    [c, d]
» B+t↵    (Προσθέτει t σε κάθε στοιχείο του B).
 ans =
        [a+t, b+t]
        [c+t, d+t]
» B*B↵  (ή και >>B^2↵)
ans =
       [a^2+b*c, a*b+b*d]
       [c*a+c*d, c*b+d^2]
» B.*B↵
ans =
       [a^2, b^2]
       [c^2, d^2]
» t=det(B) ↵
t =
       a*d-b*c
» t*inv(B) ↵
ans =
      [d –b]
      [-c  a]
» inv(B) ↵
ans =
      [d/( a*d-b*c), -b/( a*d-b*c)]
      [-c/( a*d-b*c), a/( a*d-b*c)]
```

Ιδιοδιανύσματα και Ιδιοτιμές

Ο προσδιορισμός του διάνυσματος ιδιοτιμών ενός πίνακα M γίνεται με την εντολή: » **eig(M)** ↵

Ο προσδιορισμός του διαγώνιου πίνακα ιδιοτιμών D καθώς και του πίνακα V που έχει στήλες τα ιδιδιανύσματα του M (M*V=V*D) γίνεται με την εντολή:

» **[V,D]=eig(M)** ↵

```
» H=sym(hilb(2)) ↵
H =
     [  1,  1/2]
     [1/2, 1/3]
» inv(H) ↵
ans =
      [ 4,   -6]
      [-6,  12]

» [v,d]=eig(H) ↵
v =
[                    1,                    1]
[-2/3+1/3*13^(1/2), -2/3-1/3*13^(1/2)]

d =
[2/3+1/6*13^(1/2),                    0]
[                 0, 2/3-1/6*13^(1/2)]
```

» **poly(H)** ↵ (Χαρακτηριστικό πολυώνυμο του πίνακα H).

```
ans =
      x^2-4/3*x+1/12
```

Η εντολή jordan(M) ή **[V,D]=jordan(M)** προσπαθεί να βρει έναν μη ιδιόμορφο πίνακα V τέτοιον ώστε inv(V)*M*V να είναι η κανονική μορφή κατά Jordan του πίνακα M, καθώς και τον πίνακα D ιδιοτιμών του M.

Η εντολή **[U,S,V]=svd(M)** (svd = Singular Value Decomposition) προσδιορίζει έναν διαγωνιο πίνακα D καθώς και μοναδιαίους (unitary) πίνακες U, V τέτοιους ώστε U*S*V′ = M.

Προσδιορισμός Ιακωβιανού Πίνακα

Η εντολή **jacobian(f,v)** υπολογίζει τον ιακωβιανό πίνακα της απεικόνισης f, δηλαδή τον πίνακα (df(i)/dv(j)). Όταν η f είναι βαθμωτή τότε η εντολή αυτή υπολογίζει την κλίση, grad(f), της f.

```
» syms u v ↵
» jacobian(u*exp(v),[u,v]) ↵
ans =
      [   exp(v), u*exp(v)]
```

ΜΕΤΑΣΧΗΜΑΤΙΣΜΟΙ Laplace, Fourier, Z

Ο μετασχηματισμός Laplace:

$$L(s)=\int_0^{\infty} f(t)e^{-st}dt$$

μιας συνάρτησης f(t) γίνεται με την εντολή: **laplace(f,t,s)**

```
» syms a s t w ↵
» f=exp(-a*t)*cos(w*t) ↵
f =
   exp(-a*t)*cos(w*t)
» L=laplace(f,t,s) ↵
 L =
     (s+a)/((s+a)^2+w^2)

» laplace(sym('Dirac(t)')) ↵
ans =
        1
» laplace(sym('Heaviside(t)')) ↵
ans =
       1/s
```

όπου Heaviside(t-a)= 1 για t>a, 0 για t<a και Dirac(t-a) η γνωστή συνάρτηση (το συναρτησίακο) δ(t-a) που είναι μηδέν για $t \neq a$ και απειρίζεται για t=a έτσι ώστε το 'ολοκλήρωμά' της να ισούται με 1 .

Ο αντίστροφος μετασχηματισμός προκύπτει με την εντολή: **ilaplace(L,s,t)**

```
» ilaplace(L,s,t) ↵
ans =
      exp(-a*t)*cos(w*t)
```

Αντίστοιχα ο μετασχηματισμός Fourier F(ω) μίας συνάρτησης f(t):

$$F(\omega)=\int_{-\infty}^{\infty} f(t)e^{-j\omega t}dt \qquad f(t)=\int_{-\infty}^{\infty} F(\omega)e^{j\omega t}dt$$

καθώς και ο αντίστροφός του προκύπτουν με τις εντολές: **fourier(f,t,w) , ifourier(F,w,t)**

```
» syms t w ↵
» f=t*exp(-t^2) ↵
f =
    t*exp(-t^2)/t
» F=fourier(f,t,w) ↵
F =
    -1/2*i*pi^(1/2)*w*exp(-1/4*w^2)
» ifourier(F,w,t) ↵
ans =
      ½*4^(1/2)*t*exp(-t^2)
» simplify(ans) ↵
ans =
      t*exp(-t^2)
```

Ο μετασχηματισμός Z, που χρησιμοποιείται γιά την ανάλυση συστημάτων διακριτού χρόνου, έχει την μορφή:

$$F(z) = \sum_{n=0}^{\infty} f(n)\, z^{-n}$$

όπου z μιγαδική μεταβλητή. Ο μετασχηματισμός αυτός και ο αντίστροφός του προκύπτουν αντίστοιχα με τις εντολές **ztrans(f,n,z)** και **iztrans(F,z,n)**:

```
» syms n z ↵
» f=2^n/7-(-5)^n/7
f =
    1/7*2^n-1/7*(-5)^n
» G=ztrans(f,n,z) ↵
G =
     Z/(z-2)/(z+5)
» pretty(G) ↵
        ____z_____
        (z-2)  (z+5)

» iztrans(G,z,n) ↵
ans =
      1/7*2^n-1/7*(-5)^n
```

ΠΑΡΑΔΕΙΓΜΑΤΑ ΠΡΟΓΡΑΜΜΑΤΩΝ ΤΟΥ MATLAB

Σημείωση: Τα παρακάτω προγράμματα (m-files) του MATLAB δεν αποτελούν μόνο παραδείγματα προγραμματισμού αλλά διερευνούν και διάφορα ενδιαφέροντα φαινόμενα σχετικά με την φύση των μεθόδων της Αριθμητικής Ανάλυσης αλλά και με την φύση των ιδίων των μαθηματικών προβλημάτων που εξετάζονται. Ιδιαίτερα γίνεται εκμετάλλευση των δυνατοτήτων γραφικής παρουσίασης των αποτελεσμάτων από το MATLAB για να γίνουν εποπτικά εμφανή τα φαινόμενα αυτά. Τροποποιώντας, δηλαδή τις παραμέτρους κάθε προβλήματος μπορούμε να κάνουμε ένα είδος «Πειραματικών Μαθηματικών»: μελέτη των χαρακτηριστικών ενός προβλήματος ή μίας μεθόδου όχι με την βοήθεια συμβολικής επεξεργασίας και ανάλυσης παραστάσεων αλλά με άμεση εποπτική παρατήρηση αυτών των χαρακτηριστικών.

Παρατηρήσεις για τα προγράμματα που ακολουθούν

Τα CHAOS, CHAOS1, CHAOS2 δείχνουν την χαοτική συμπεριφορά επαναληπτικών μεθόδων εύρεσης σημείων μηδενισμού.

Το INTERP είναι πρόγραμμα παρεμβολής Lagrange που χρησιμοποιεί την μέθοδο του AITKEN για να υπολογίσει τις τιμές του πολυωνύμου παρεμβολής.
Τα INTERPOL και LAGRANGE κατασκευάζουν το πολυώνυμο παρεμβολής με βάση τον τύπο του Lagrange.

Το PEOPLE είναι ένα πρόγραμμα πολυωνυμικής προσέγγισης με την μέθοδο ελαχίστων τετραγώνων.

Τα AMELKIN, DRIVA YPLOT DEQF1 και DEQ είναι προγράμματα λύσης διαφορικών εξισώσεων (προβλημάτων αρχικών τιμών).
Το DRIVA χρησιμοποιεί τις συναρτήσεις EULERT και FUNC.
Το YPLOT χρησιμοποιεί τις συναρτήσεις EULERT και FUNC.
Το DEQF1 χρησιμοποιεί τις συναρτήσεις HEUN και F1.
Το DEQ χρησιμοποιεί τις συναρτήσεις HN και F1.
Το HN εφαρμόζει την μέθοδο Heun με μίαν χρήση του corrector.
Τα FUNC και F1 ορίζουν συναρτήσεις f(t,y) για την διαφ. εξ. y'= f(t,y)
Το DIFEQ λύνει προβλήματα αρχικών τιμών για κοινές διαφορικές εξισώσεις με την βοήθεια της συνάρτησης ODE45 που περιέχει το MATLAB. Η συνάρτηση του δευτ΄ςρου μέρους της Δ.Ε. f(t,y) ορίζεται από το πρόγραμμα DIFFUNC.

Χαοτική συμπεριφορά μεθόδων εύρεσης σημείων μηδενισμού

Τα CHAOS, CHAOS1, CHAOS2 δείχνουν την χαοτική συμπεριφορά επαναληπτικών μεθόδων εύρεσης σημείων μηδενισμού με τάξη σύγκλισης μεγαλύτερη της μονάδας, όπως η μέθοδος Newton-Raphson. Με τυχαίες αρχικές τιμές συγκλίνουν μεν πάντα ή σχεδόν πάντα προς κάποιο σημείο μηδενισμού της εξεταζόμενης συνάρτησης (αν υπάρχει) αλλά δεν είναι πάντα προβλέψιμο σε ποια ρίζα θα μας οδηγήσει κάποια αρχική τιμή. Σε ορισμένες περιοχές εκκίνησης, αρχικές τιμές που διαφέρουν απειροελάχιστα οδηγούν σε διαφορετικές ρίζες. Εδώ εξετάζεται ένα πολυώνυμο τρίτου βαθμού που έχει μόνο πραγματικές ρίζες και χρωματίζονται διάφορα διαστήματα του πραγματικού άξονα αναλόγως του σε ποιάν από τις τρεις ρίζες καταλήγει κάθε σημείο του διαστήματος. Κάνοντας πυκνότερη την σάρωση των σημείων παρατηρούμε ότι περιοχές που «μακροσκοπικά» φαίνονται μονόχρωμες διακόπτονται από μικρά διαστήματα διαφορετικού χρώματος. Το φαινόμενο αυτό συνεχίζεται όσο πυκνότερη σάρωση κάνουμε (όσο μικραίνουμε τις αποστάσεις των δοκιμαζομένων αρχικών τιμών).

Chaos.m

```
%chaos.m, chaos1.m and chaos2.m show the sensitive dependence of the limit
point of
%a Newton-Raphson sequence from its initial values
axis([-2.5,2.5,-1,1])
for i=1:200;
  x(i)=-1+0.01*i;
  t=x(i);
  d=(t^3-t)/(3*t^2-1);
  %d=0.5*(t^3-t), which is used in Short92.m, does not provide convergence for
all x-values
  while abs(d)>0.05
    y=t-d;
```

```
    t=y;
    d=(t^3-t)/(3*t^2-1);
  end
  r(i)=t;
  if abs(r(i)+1)<0.1
       plot(x(i),0,'.b')
      %plot(x(i),-.1,'.b')
      hold on
  elseif abs(r(i)-1)<0.1
       plot(x(i),0,'.r')
      %plot(x(i),.1,'.r')
    else
       plot(x(i),0,'.g')
  end
end
hold off
```

Chaos1.m

```
axis([-2.5,2.5,-1,1])
for i=1:200;
  x(i)=-1+0.01*i;
  t=x(i);
  d=(t^3-t)/(3*t^2-1);
  %d=0.5*(t^3-t) does not provide convergence for all x-values
  while abs(d)>0.005
    y=t-d;
    t=y;
    d=(t^3-t)/(3*t^2-1);
  end
  r(i)=t;
  if abs(r(i)+1)<0.1
      %plot(x(i),0,'.b')
       plot(x(i),-.1,'.b')
      hold on
  elseif abs(r(i)-1)<0.1
      %plot(x(i),0,'.r')
       plot(x(i),.1,'.r')
    else
       plot(x(i),0,'.g')
  end
end
hold off
```

Chaos2.m

```
axis([-1,1,-1,1])
```

```
for i=1:200;
   x(i)=-.6+0.001*i;
   t=x(i);
   d=(t^3-t)/(3*t^2-1);
   %d=0.5*(t^3-t) does not provide convergence for all x-values
   while abs(d)>0.005
      y=t-d;
      t=y;
      d=(t^3-t)/(3*t^2-1);
   end
   r(i)=t;
   if abs(r(i)+1)<0.1
         plot(x(i),0,'.b')
        %plot(x(i),-.1,'.b')
        hold on
   elseif abs(r(i)-1)<0.1
         plot(x(i),0,'.r')
        %plot(x(i),.1,'.r')
      else
         plot(x(i),0,'.g')
   end
end
hold off
```

Πολυωνυμική παρεμβολή Lagrange

Τα προγράμματα παρεμβολής σε ισαπέχοντα σημεία, που ακολουθούν, δείχνουν ότι για ορισμένες συναρτήσεις το σφάλμα παρεμβολής γίνεται σε κάποιες θέσεις προς τα άκρα του διαστήματος παρεμβολής όλο και μεγαλύτερο όσο αυξάνουμε τον αριθμό των σημείων παρεμβολής.

interp.m

```
%use 'clear' to redefine the vector
%lengths after changing the interval
clear
x=-2:.2:2;
y=x.*log(abs(x)+.001)-x;
t=-2:.02:2;
n=length(t);
for i=1:n
   z(i)=aitken(x,y,t(i));
   w(i)=t(i)*log(abs(t(i))+.001)-t(i);
end
%plot of y=0 in order to have x-span
%[-2.1,2.1] and y-span [-2,2]
fplot('x.^0-1',[-2.1,2.1,-2,2])
```

```
hold
plot(t,z,t,w,x,y,'*')
hold off
```

interp2.m

```
%use 'clear' to redefine the vector
%lengths after changing the interval
clear
x=-2:.25:2;
y=1./(1+x.^2);
t=-2:.01:2;
n=length(t);
for i=1:n
   z(i)=aitken(x,y,t(i));
   w(i)=1/(1+t(i)^2);
end
plot(t,z,t,w,x,y,'*')
```

Aitken.m

(Υπολογίζει σημειακά τις τιμές του πολυωνύμου παρεμβολής Lagrange. Καλείται από το interp.m και το interp2.m)

```
function q=aitken(x,y,xval)
n=length(x);
p=zeros(n);
p(1,:)=y;
for j=1:n-1
   for i=j+1:n
      p(j+1,i)=(p(j,i)*(xval-x(j))-p(j,j)*(xval-x(i)))/(x(i)-x(j));
   end
end
q=p(n,n);
```

interpol.m

(Εφαρμόζει τον τύπο του Lagrange για τον υπολογισμό του πολυωνύμου παρεμβολής)

```
function f=interpol(x,y,xval)
%Lagrange interpolation at the point xval
%for given interpolation points (x(i),y(i))
%where x, y are given vectors.
%EQUIVALENT TO lagrange.m
n=length(x);
%coefficients
for i=1:n
      denom=1;
```

```
  for j=1:n
    if j~=i
      denom =denom*(x(i)-x(j));
    end
    c(i)=y(i)/denom;
  end
end
%This last'end' was missing
%and caused index-error
%%%%%%
 %calculation of f-value
    prod=1;
  for i=1:n
       prod=prod*(xval-x(i));
  end
    sum=0;
    for i=1:n
      if xval==x(i)
        f=y(i);
        return
      else
    term(i)=c(i)/(xval-x(i));
    sum=sum+term(i);
      end
   end
   r=sum*prod;
       prod;
       f=r;
```

lagrange.m

(Εφαρμόζει τον τύπο του Lagrange για τον υπολογισμό του πολυωνύμου παρεμβολής)

```
function p=lagrange(x,y,xv)
%For given interpolation pairs (x,y)
%it determines the value of the interpolation
%polynomial at xv
%EQUIVALENT TO interpol.m
%corrected
%The line 'for j=1:i-1,i+1:n' was wrong!!!
n=length(x);
%computing w(xv)= prod(xv-x(i))
w=1;
for i=1:n
  w=w*(xv-x(i));
end
%computing c(i)=y(i)/prod(x(i)-x(j)),j><i
```

```
for i=1:n
      c(i)=y(i);
   for j=1:i-1
      c(i)=c(i)/(x(i)-x(j));
   end
   for j=i+1:n
      c(i)=c(i)/(x(i)-x(j));
   end
end
%computing p(xv)=w(xv)*sum{c(i)/(xv-x(i))}
s=0;
for i=1:n
   if xv==x(i)
      p=y(i);
      return
   else
       s=s+c(i)/(xv-x(i));
   end
end
   p=w*s;
```

Αντίστροφη Παρεμβολή

invint.m

```
x=[.4,.6,.8,1,1.2];
y=[-1.1106,-0.102,.4674,.8415,1.0987];
n=length(x);
q=polyfit(y,x,n-1);
%q is a least squares polynomial data-fit
%but for degree=n-1 it coincides with
%the interpolation polynomial
yv=-1.5:.05:1.5;
xv=polyval(q,yv);
plot(y,x,'-*',yv,xv);
title('Inverse Interpolation of sin(x)+1-1./x')
xlabel('y'),ylabel('x')
grid
r=polyval(q,0)
%In order to print r in the graph window
%use: gtext('r=<r-value>') EXTERNALLY
```

lsq2.m

```
%lsq2=least squares approximation
n=input('n=')
a=-2:.5:2;
%since these are 9 points for n=8 we have interpolation
b=a.*log(abs(a)+.001)-a;
p=polyfit(a,b,n);
x=-2:.05:2;
y=x.*log(abs(x)+.001)-x;
d=polyval(p,x);
axis([-2.1,2.1,-3,3])
plot(x,d,'r')
title('x*log(|x|+.001)-x (green) versus n.degree polynomial fit at the points a=-
2:.5:2 (red)')
hold on
plot(x,y,'g')
plot([-2.1,2.1],[0,0])
hold off
grid
```

lsq3.m

Το ακόλουθο παράδειγμα δείχνει πόσο ασταθές είναι το σύστημα κανονικών εξισώσεων που παράγει η μέθοδος ελαχίστων τετραγώνων. Αφού χρησιμοποιούμε 9 τιμές για να προσεγγίσουμε ένα πολυώνυμο μόνο τρίτου βαθμού, $y=x^3$, θα έπρεπε το πολυώνυμο που προκύπτει με την μέθοδο ελαχίστων τετραγώνων να συμπίπτει με το προσεγγιζόμενο πολυώνυμο, x^3. Όμως εδώ έχουμε εμφανή απόκλιση των καμπύλων προς τα άκρα του διαστήματος [-2,2]. Οι αποκλίσεις γίνονται ακόμα μεγαλύτερες αν κατασκευάσουμε προσεγγίζον πολυώνυμο ακόμα μεγαλυτέρου βαθμού.

```
%lsq3=least squares approximation
n=input('n=')
a=-2:.5:2;
%since these are 9 points for n=8 we have interpolation
b=a.^3;
p=polyfit(a,b,n);
x=-2:.05:2;
y=x.^3;
d=polyval(p,x);
%axis([-2.1,2.1,-3,3])
plot(x,d,'r')
title('green function versus red n.degree polynomial fit at the points a=-2:.5:2')
hold on
plot(x,y,'g')
```

```
plot([-2.1,2.1],[0,0])
hold off
grid
```

people.m

```
%people.m
%The population explosion
%Input data are estimates p(k) of the world population in billions
%in the corresponding year t(k) expressed as a fraction of 1000 years.
clear;
%t=input('Enter the years","s");
%p=input('Enter the population","s");
t=[1 1.5 1.75 1.83 1.93 1.96 1.975 1.987 1.995];
p=[.25 .46 .83 1 2 3 4 5 5.75];
q=log(p);
e=ones(size(t));
m=[e;t];
%The system of canonical equations for a least squares approximation is:
%m*m'*a=m*q'
C=m*m';b=m*q';
a=b\C
end;
```

populat.m

```
t=[1.5,1.75,1.83,1.93,1.96,1.975,1.987,2,2,2];
s=[.45,.78,1,2,3,4,5,6,6,6];
r=log(s);
%The data pair (t,s)=(2,6) is repeated
%three times in order to increase it's
%weight, so that tv=2 approaches zv=6
n=input('n=');
p=polyfit(t,r,n);
%q=polyfit(r,t,n);
%tv=linspace(1,2,100);
tv=1.5:.01:2.05;
zv=polyval(p,tv);
pv=exp(zv);
%plot(t,r,'-o',tv,zv);
plot(t,s,'-*',tv,pv);
xlabel('milleniums')
%ylabel('billions')
title('World Population Least squares curve fitting')
grid
gtext('billions')
```

Spline.m

```
x=0:10;
y=sin(x);
xi=0:.25:10;
yi=interp1(x,y,xi,'spline');
plot(x,y,'o',xi,yi)
```

Μελέτη προβλημάτων αρχικών τιμών για ΚΔΕ

Amelkin.m

Η διαφορική εξίσωση $y'=-x/y$ που λύνεται εδώ προσεγγιστικά με την μέθοδο τραπεζίου έχει για $y(x_0)=y_0$ την γενική λύση $y=(y_0^2+ x_0^2 - x^2)^{1/2}$ που δεν έχει νόημα για $x>(y_0^2+ x_0^2)^{1/2}$, επειδή παραβιάζεται η συνθήκη μονοσημαντότητας του Lipschitz. Όμως, η αριθμητική ακολουθία συνεχίζει να παράγει τιμές που δεν έχουν πια καμία σχέση με την πραγματική λύση!

```
%Solves dy/dt=f(t,y). start=starting t-val., finish=final t-value
% startval = initial y, step = t-increment
t=-1;y=.2;
start=-1;finish=3;step=.05;startval=.2;

steps=(finish-start)/step+1
y=startval;t=start;
yvals=startval;tvals=start;
for i=2:steps
  y1=y-step*t/y;
  t1=t+step;
  y2=y-step*(t/y+t1/y1)/2;
    diff=y1-y2;
    y1=y2;
tvals=[tvals,t1];yvals=[yvals,y1];
t=t1;
%a=input('stp')
y=y1;
end;
plot(tvals,yvals)
```

deq.m

Η διαφορική εξίσωση $y'=-x/y$ που λύνεται εδώ προσεγγιστικά με την μέθοδο τραπεζίου έχει για $y(x_0)=y_0$ την γενική λύση $y=(y_0^2+ x_0^2 - x^2)^{1/2}$ που δεν έχει νόημα για $x>(y_0^2+ x_0^2)^{1/2}$, επειδή παραβιάζεται η συνθήκη μονοσημαντότητας του Lipschitz. Όμως, η αριθμητική ακολουθία συνεχίζει να παράγει τιμές που δεν έχουν πια καμία σχέση με την πραγματική λύση!

Δοκιμάστε τι προκύπτει για διαφορετικές τιμές του βήματος h έστω και πολύ κοντινές.

```
%deq.m : Shows what happens if the Lipschitz condition
% is not fulfilled
h=input('h=');
[tv,yv]=hn('f1',-1,2,.2,h);
plot(tv,yv);
grid;
%gtext('dy/dt=-t/y');
title('dy/dt=-t/y, t0=-1, tfin=2, y0=.2, h=.03');
hold;
```

deqerr.m

```
%deqerr.m : Shows what happens if h is
% too small. Try h=10^-6 and h=10^-7
k=10^10;
h=input('h=');
[tv,yv]=hn('f5',1,2,1,h);
zv=k*(yv-1./tv);
plot(tv,zv);
grid;
title('dy/dt=-5*t*y^2+5/t-1/t^2, t0=1, tfin=2, y0=1');
hold;
```

deqf.m

```
%deqf.m : Shows what happens if the
%Lipschitz condition is not fulfilled
%The irregular curve pieces are circles
%of different radii. To see this try
%y0=.5 and y0=5 or y0=3 and y0=4
%y0=input('y0=');
y0=.2;
%h=.03003;
h=input('h=');
[tv,yv]=ec('f1',1,2.8,y0,h);
tv,yv
plot(tv,yv);
grid on;
title('dy/dt=-t/y, t0=.9, tfin=4, y0=.5+d, h=.01');
hold on;
```

ec.m (Euler-Cauchy)

```
function [tv,yv]=ec(f,t0,tf,y0,h)
%Solves dy/dt=f(t,y) with y(t0)=y0
%tf = final t-value, h = t-increment
```

```
%by means of the Euler-Caucy method
y=y0;t=t0;
yv=y0;tv=t0;
n=(tf-t0)/h+1;
for k=2:n
    yn=y+h*feval(f,t,y);
    tn=t+h;
    tv=[tv,tn];
  yv=[yv,yn];
  t=tn;y=yn;
end;
```

deqf1.m

```
clear
[tv,yv]=heun('f1',-1,1,.2,.03);
plot(tv,yv);
grid;
```

deqf3.m

```
%deqf3.m : Shows what happens if the Lipschitz condition
% is not fulfilled. It uses alternatively
%Heun's method or the Euler-Cauchy method
h=input('h=');
[tv,yv]=hn('f3',-1,1.5,-1,h);
%[tv,yv]=ec('f3',-1,1.5,-1,h);
plot(tv,yv)
grid;
title('dy/dt=3t*y ^p, p=1/3, t0= -1, tfin=1.5, y0= -1');
```

f1.m

```
%f1.m
function v=fct(t,y);
   v=-t./y;
```

f2.m

```
function v=fct(t,y);
   v=-2.*t.*y.^2;
```

f3.m

```
%f3.m
function v=fct(t,y);
```

```
   v=3*t.*sign(y).*abs(y).^(1/3);
```

f4.m

```
%f4.m
function v=fct(t,y);
   v=1+y.^2;
```

f5.m

```
%f5.m
function v=fct(t,y);
   v=-5*t.*y.^2+5./t-1./t.^2;
```

func.m

```
function v=func(t,y);
v=sin(y-t^2);
```

g1.m

```
%g1.m
function v=fct(t,y);
%v=sin(t.*y);
v=cos(t.*y);
%v=sin(t+y);
```

deqg1.m

```
%deqg1.m : Shows what happens if the
%absolute stability condition
% is not always fulfilled
h=input('h=');
[tv,yv]=ode45('g1',0,10,1,h);
%[tv,yv]=hn('g1',0,10,.1,h);
plot(tv,yv);
grid;
title('dy/dt=sin(t*y), t0=0, tfin=10, y0=1, h=.1');
hold on;
```

g1.m

```
%g1.m
function v=fct(t,y);
%v=sin(t.*y);
v=cos(t.*y);
%v=sin(t+y);
```

difeq.m

Μία εφαρμογή της αριθμητικής λύσης του προβλήματος αρχικών τιμών είναι για τον προσδιορισμό της γραφικής παράστασης μίας πεπλεγμένης συνάρτησης f(x,y)=a. Αρκεί να σχεδιάσουμε την λύση της Δ.Ε. $y'=-f_x/f_y$ με αρχικές τιμές x_0, y_0 τέτοιες ώστε $f(x_0,y_0)=a$. Εδώ κατασκευάζεται προσεγγιστικά το γράφημα της εξίσωσης: $x^y y^x=a$.

```
xfin=input('final x=');
x0=input('x0=');
y0=input('y0=');
[x,y]=ode45('diffunc',[x0,xfin],y0);
plot(x,y);xlabel('x');ylabel('y');
title('y^xx^y=a, if x0=1 then y0=a');
```

diffunc.m

```
function yton=diffunc(x,y)
yton=-(log(y)+y./x)./(log(x)+x./y);
```

heun.m

```
function [tv,yv]=heun(f,t0,tf,y0,h)
%Solves dy/dt=f(t,y) with y(t0)=y0. tf=final t-value, h=t-increment
y=y0;t=t0;
yv=y0;tv=t0;
n=(tf-t0)/h+1;
%yn=y-new,tn=t-new,yi=y-improved
for k=2:n
    yn=y+h*feval(f,t,y);
    tn=t+h;
    iter=0;dy=1;
  while abs(dy)>h^2
    iter=iter+1;
    yi=y+h*(feval(f,t,y)+feval(f,tn,yn))/2;
    dy=yi-yn;yn=yi;
  end;
tv=[tv,tn];yv=[yv,yn];
t=tn;y=yn;
end;
```

hn.m

```
function [tv,yv]=hn(f,t0,tf,y0,h)
%Solves dy/dt=f(t,y) with y(t0)=y0. tf=final t-value, h=t-increment
y=y0;t=t0;
yv=y0;tv=t0;
```

```
n=(tf-t0)/h+1;
%yn=y-new,tn=t-new,yi=y-improved
for k=2:n
   yn=y+h*feval(f,t,y);
   tn=t+h;
   yi=y+h*(feval(f,t,y)+feval(f,tn,yn))/2;
   yn=yi;
   tv=[tv,tn];
  yv=[yv,yn];
  t=tn;y=yn;
end;
```

Yplot.m

```
clear
[tv,yv]=heun('func',0,50,.5,.1);
plot(tv,yv);
grid;
%gtext('dy/dt=sin(y-t^2)');
hold off
```

yplot20.m

```
clear
[tv,yv]=heun('func',0,20,.5,.1);
plot(tv,yv);
grid;
%gtext('dy/dt=sin(y-t^2)');
hold off
```

<u>Σημείωση:</u> Τα παραπάνω προγράμματα αντιγράφηκαν από τον MATLAB editor/debugger με την ακόλουθη διαδικασία:

1. Άνοιγμα ενός νέου Word document.
2. Ελαχιστοποίηση του παραθύρου Word και άνοιγμα του MATLAB Command Window.
3. Άνοιγμα εκεί του φάκελλου περιεχομένων. Αυτός μας παρουσιάζει τα m-files που περιέχει.
4. Άνοιγμα ενός από αυτά με διπλό κλικ στο όνομά του.
5. Μαρκάρισμα όλου του περιεχομένου με shift ↓ και αντιγραφή του στην βοηθητική μνήμη με Ctrl C.
6. Ελαχιστοποίηση όλων των παραθύρων του MATLAB και μεγιστοποίηση του παραθύρου του Word.
7. Αντιγραφή του περιεχομένου του αρχείου στο παράθυρο του Word με Ctrl V. Αλλαγή της γραμματοσειράς σε Tahoma.

Υπόδειγμα Λύσης των Θεμάτων μίας Τυπικής Εξέτασης

Εξέταση Χημικών: 3.3.2003

1. (α) Να γίνει γραφικός εντοπισμός όλων των λύσεων της επόμενης εξίσωσης και να προσεγγισθεί η θετική λύση με την μέθοδο Newton-Raphson με μίαν επανάληψη:

$$x^2-2 = \ln|x+1|$$

(β) Περιγράψτε πώς προκύπτει αυτή η μέθοδος γραφικά ή αναλυτικά. **(2)**

2.(α) Να γραφτεί σε μορφή Lagrange το πολυώνυμο παρεμβολής στα σημεία $x_0=0$, $x_1=0.1$, $x_2=0.2$ που προσεγγίζει την συνάρτηση f(x):= $\sqrt{1-2x}$.

(β) Να δοθεί άνω φράγμα του σφάλματος προσέγγισης για το διάστημα: $-0.2 \le x \le 0.2$ σε μορφή $|f(x) - p(x)| \le E(x)$. $f(x)-p(x)=\dfrac{f^{(n+1)}(\xi)}{(n+1)!}\omega(x)$ **(2)**

3. Προσεγγίσετε τη συνάρτηση f(x):= $\sqrt{2-x}$ με την Μέθοδο Ελαχίστων Τετραγώνων με ένα πολυώνυμο δευτέρου βαθμού και με χρήση των τιμών της συνάρτησης στα σημεία: $x_0=-1$, $x_1=-0.5$, $x_2= 0$, $x_3=0.5$, $x_4=1$. Αρκεί να γράψετε το σύστημα που προσδιορίζει τους συντελεστές και να εξηγήσετε πώς προκύπτει. **(1)**

4.(α) Πώς προκύπτει υπολογιστικά ο απλός τύπος του τραπεζίου και πώς κατασκευάζεται με αυτόν ο σύνθετος τύπος για N υποδιαστήματα;

(β) Αν S_N είναι η προσέγγιση με αυτόν τον τύπο του ολοκληρώματος:

$$I=\int_0^1 \sin(1+x^2)dx$$

ποια τιμή του N αρκεί για να έχουμε $|I- S_N| \le 10^{-5}$; **(2)**

$$I-S_N=-(b-a)\frac{h^2}{12}f''(\xi)$$

5. Με την μέθοδο Μερικών Βημάτων (Gauss-Seidel) να προσεγγισθεί η λύση του συστήματος:

$$\begin{aligned} 10x_1+x_2 \quad &=20 \\ x_1+10x_2+2x_3&=30 \\ 2x_2+10x_3&=40 \end{aligned}$$

(α) Με αρχική τιμή (0,0,0) να γίνουν δύο επαναλήψεις και με κατάλληλη νόρμ να γίνει εκ των υστέρων εκτίμηση του σφάλματος κατά συνιστώσα.

(β) Πόσες επαναλήψεις αρκούν για να είναι το σφάλμα κατά συνιστώσα μικρότερο από 10^{-6}; **(2)**

6.(α) Με την Μέθοδο Ελαχίστων Τετραγώνων για συστήματα να προσδιοριστεί η «ψευδολύση» του ασυνεπούς γραμμικού συστήματος:

$$2x-y=3,\ x+y=3,\ x-2y=-1$$

Πώς προκύπτει το σύστημα που δίνει την ψευδολύση;

(β) Να προσεγγισθεί με την μέθοδο Newton-Raphson και με μία μόνο επανάληψη η λύση του μη γραμμικού συστήματος:

$$x^y y^x=2,\ xy+2x/y=6$$

Για διευκόλυνση των πράξεων μπορείτε να μετασχηματίσετε σε άλλη, ισοδύναμη μορφή την πρώτη εξίσωση. $J_f(x^{(k)})(x^{(k+1)}-x^{(k)})=-f(x^{(k)})$ **(1)**

Υποδείξεις:

1. Η διάρκεια της εξέτασης είναι: 2,5 ώρες
2. Δεν έχει τόση σημασία η ακρίβεια των πράξεων όσο η σαφής τεκμηρίωση της μεθόδου λύσεως (ποιοι τύποι εφαρμόζονται και πώς ακριβώς).
3. Κατά την διαφόριση ή ολοκλήρωση τα τόξα των τριγωνομετρικών αριθμών πρέπει να μετρώνται σε ΑΚΤΙΝΙΑ (RADIANS).
4. Η συνολική βαθμολογία κάθε θέματος δίνεται στο τέλος του με έντονη γραφή.

Απαντήσεις των Θεμάτων[12]

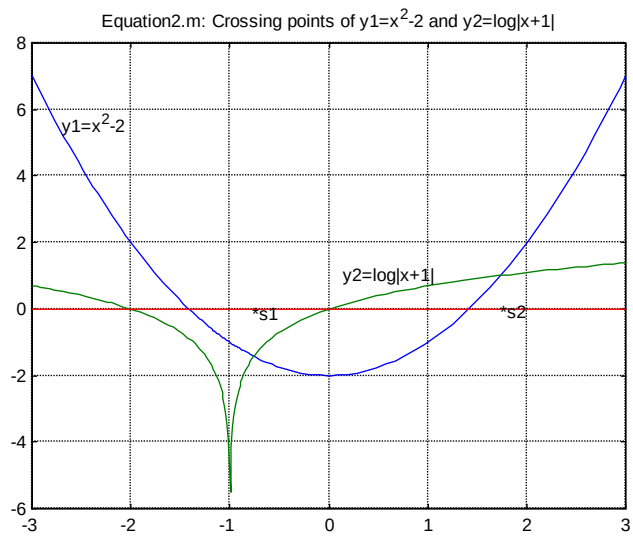
1*. (α) Να γίνει γραφικός εντοπισμός όλων των λύσεων της επόμενης εξίσωσης και να προσεγγισθεί η θετική λύση με την μέθοδο Newton-Raphson με μίαν επανάληψη:*

$$x^2-2 = \ln|x+1|$$

(β) Περιγράψτε πώς προκύπτει αυτή η μέθοδος γραφικά ή αναλυτικά.

Απάντηση

1.α. Αν $f_1(x)=x^2-2$ και $f_2(x)=\ln|x+1|$ τότε τα σημεία διασταυρώσεως των καμπύλων $y_1=f_1(x)$ και $y_1=f_1(x)$ είναι οι λύσεις της εξίσωσης $f_1(x)=f_2(x)$.



[12] Για διευκόλυνση του αναγνώστη επαναλαμβάνουμε την εκφώνηση κάθε θέματος πριν από την απάντησή του.

Από την γραφική παράσταση αυτών των καμπύλων φαίνεται ότι οι λύσεις είναι δύο: $s_0 \in (-1, 0)$ και $s_1 > \sqrt{2}$, γιατί $f_1(\sqrt{2})=0$ ενώ $f_2(\sqrt{2})>0$.
Για x>-1 είναι

$$f(x):=f_1(x)-f_2(x)=x^2-2-\ln(x+1)\text{ , και } f'(x)=2x-1/(x+1)$$

άρα η μέθοδος Newton-Raphson έχει την μορφή της επαναληπτικής ακολουθίας:

$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)} = x_k - \frac{x_k^{\,2} - 2 - \ln(x_k+1)}{2x_k - \dfrac{1}{x_k+1}} \qquad k = 0,1,2,\ldots$$

Για τον προσδιορισμό της θετικής λύσης s_2 αρκεί να επιλέξουμε μίαν αρχική τιμή που να βρίσκεται κοντά στο s_2, π.χ. $x_0=1.5$. Τότε προκύπτει:

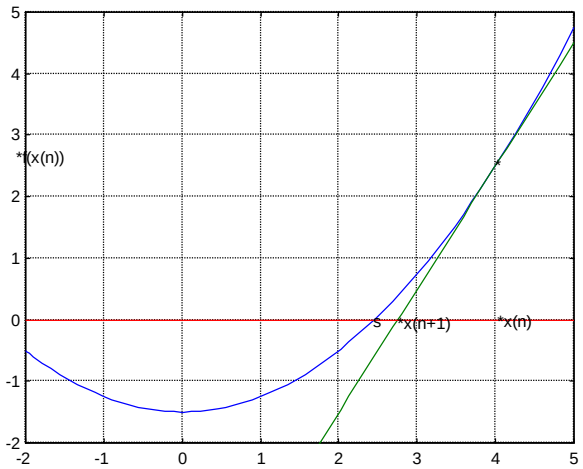
$$x_1 = 1.5 - \frac{2.25 - 2 - \ln(2.5)}{3 - \dfrac{1}{2.5}} = 1.5 - \frac{0.25 - 0.91629}{2.6} \approx 1.756266$$

β) Σχεδιάζουμε την καμπύλη y = f(x) και φέρουμε την εφαπτομένη της στο τυχαίο σημείο x_n, που θεωρούμε ότι βρίσκεται κοντά στην ζητούμενη λύση s. Το σημείο όπου η εφαπτομένη αυτή τέμνει τον οριζόντιο άξονα το ονομάζουμε x_{n+1} και όπως φαίνεται από το σχήμα είναι πολύ πλησιέστερα προς το s από το x_n. Γι' αυτήν την καλύτερη προσέγγιση ισχύει με βάση το σχήμα:

$$f'(x_n) = \tan a_n = \frac{f(x_n)}{x_n - x_{n+1}}$$

όπου a_n είναι η γωνία που σχηματίζει η εφαπτομένη με τον οριζόντιο άξονα. Από την σχέση αυτή προκύπτει με επίλυση ως προς x_{n+1}:

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$$



***2**.(α) Να γραφτεί σε μορφή Lagrange το πολυώνυμο παρεμβολής στα σημεία $x_0=0$, $x_1=0.1$, $x_2=0.2$ που προσεγγίζει την συνάρτηση $f(x):=\sqrt{1-2x}$.*

(β) Να δοθεί άνω φράγμα του σφάλματος προσέγγισης για το διάστημα:

-0.2 ≤ x ≤ 0.2 σε μορφή |f(x) - p(x)| ≤ E(x). $f(x)-p(x)=\dfrac{f^{(n+1)}(\xi)}{(n+1)!}\omega(x)$

Απάντηση

α) $f_0=f(x_0)=f(0)=1$, $f_1=f(x_1)=f(0.1)=0.894427$, $f_2=f(x_2)=f(0.2)=0.774597$

Επειδή

$$P_2(x)=f_0\frac{(x-x_1)(x-x_2)}{(x_0-x_1)(x_0-x_2)}+f_1\frac{(x-x_0)(x-x_2)}{(x_1-x_0)(x_1-x_2)}+f_2\frac{(x-x_0)(x-x_1)}{(x_2-x_0)(x_2-x_1)}$$

θα είναι επομένως

$$P_2(x)=\frac{(x-0.1)(x-0.2)}{(0-0.1)(0-0.2)}+0.89443\frac{(x-0)(x-0.2)}{(0.1-0)(0.1-0.2)}+0.7746\frac{(x-0)(x-0.1)}{(0.2-0)(0.2-0.1)}$$

ή

$$P_2(x)=44.72135(x-0.1)(x-0.2)-89.4427x(x-0.2)+38.72985x(x-0.1)$$

β) Ο τύπος που μας δίνει το σφάλμα προσέγγισης παίρνει εδώ την μορφή:

$$f(x)-P_2(x)=\frac{f^{(3)}(\xi)}{3!}x(x-0.1)(x-0.2)$$

όπου το σημείο ξ μας είναι άγνωστο αλλά βρίσκεται μέσα στην ένωση του διαστήματος όπου θα χρησιμοποιηθεί η προσέγγιση και του διαστήματος των σημείων παρεμβολής:

$$\xi\in[x_{min}, x_{max}]\cap[x_0, x_2]=[-0.2,0.2]\cap[0, 0.2]=[-0.2, 0.2]$$

Επειδή

$$f(x)=(1-2x)^{1/2},\ f'(x)=-(1-2x)^{-1/2},\ f''(x)=-(1-2x)^{-3/2},\ f^{(3)}(x)=-3(1-2x)^{-5/2}$$

ισχύει, επομένως,

$$|f(x)-P_2(x)|\le\frac{M}{6}|x(x-0.1)(x-0.2)|$$

όπου

$$M=\max_{\xi\in[-0.2,0.2]}|f^{(3)}(\xi)|=3\max_{\xi\in[-0.2,0.2]}(1-2\xi)^{-5/2}=3(1-2\times0.2)^{-5/2}=$$
$$=3/(0.6)^{5/2}=10.75829$$

Επειδή είναι M/6=1.793048, θα ισχύει άρα για κάθε $x\in[-0.2,0.2]$:

$$|(1-2x)^{1/2}-P_2(x)|\le1.793048|x(x-0.1)(x-0.2)|$$

3*. Προσεγγίσετε τη συνάρτηση f(x):=* $\sqrt{2-x}$ *με την Μέθοδο Ελαχίστων Τετραγώνων με ένα πολυώνυμο δευτέρου βαθμού και με χρήση των τιμών της συνάρτησης στα σημεία:* $x_0=-1$, $x_1=-0.5$, $x_2=0$, $x_3=0.5$, $x_4=1$. *Αρκεί να γράψετε το σύστημα που προσδιορίζει τους συντελεστές και να εξηγήσετε πώς προκύπτει.*

Απάντηση

Οι τιμές της f στα σημεία x_i, i=0(1)4 είναι αντίστοιχα (με ακρίβεια 5 δεκαδικών ψηφίων μετά την υποδιαστολή):
$f_0=\sqrt{3}=1.73205$, $f_1=\sqrt{2.5}=1.58114$, $f_2=\sqrt{2}=1.41421$, $f_3=\sqrt{1.5}=1.22474$, $f_4=1$.
Η μέθοδος ελαχίστων τετραγώνων προσδιορίζει το πολυώνυμο:

$$P_2(x)=a_0+a_1x+a_2x^2$$

που θα προσεγγίζει την f(x) (βασικά, στο διάστημα $[x_0,x_4]$) έτσι ώστε να είναι:

$$E(a_0,a_1,a_2):=\sum_{k=0}^{4}(P_2(x_k)-f_k)^2=\sum_{k=0}^{4}(a_0+a_1x_k+a_2{x_k}^2-f_k)^2=\min$$

Δηλαδή προσδιορίζει τους συντελεστές a_j, j=0,1,2 λύνοντας το σύστημα:

$$0=\frac{\partial E}{\partial a_j}=2\sum_{k=0}^{4}(a_0+a_1x_k+a_2{x_k}^2-f_k){x_k}^j,\quad j=0,1,2$$

που μπορεί να γραφτεί και στην μορφή:

$$a_0[x^0,x^j]+a_1[x^1,x^j]+a_2[x^2,x^j]=[f,x^j],\quad j=0,1,2$$

όπου για δοσμένες συναρτήσεις g(x), h(x) ο συμβολισμός [g,h] σημαίνει:

$$[g,h]=\sum_{k=0}^{4}g(x_k)h(x_k)=\sum_{k=0}^{4}g_kh_k$$

Εδώ, δηλαδή, είναι:

$$[x^q,x^j]=\sum_{k=0}^{4}{x_k}^q{x_k}^j=\sum_{k=0}^{4}{x_k}^{q+j}=(-1)^{q+j}+(-0.5)^{q+j}+0^{q+j}+(0.5)^{q+j}+1$$

και

$$[f,x^j]=\sum_{k=0}^{4}f_k{x_k}^j=(-1)^jf_0+(-0.5)^jf_1+0^jf_2+(0.5)^jf_3+f_4$$

άρα

$[x^0,x^0]=5$ (επειδή $0^0=1$), $[x^0,x^1]=0$, $[x^0,x^2]=2.5$
$[x^1,x^0]=0$, $[x^1,x^1]=2.5$, $[x^1,x^2]=0$
$[x^0,x^2]=2.5$, $[x^1,x^2]=0$, $[x^2,x^2]=2.125$

και

$[f,x^0]=6.95214$, $[f,x^1]=0.55385$, $[f,x^2]=3.43352$

Το ζητούμενο σύστημα είναι, επομένως:

$$5a_0+0+2.5a_2=6.95214$$
$$0+2.5a_1+0=0.55385$$
$$2.5a_0+0+2.125a_2=3.43352$$

***4**.(α) Πώς προκύπτει υπολογιστικά ο απλός τύπος του τραπεζίου και πώς κατασκευάζεται με αυτόν ο σύνθετος τύπος για N υποδιαστήματα;*
(β) Αν S_N είναι η προσέγγιση με αυτόν τον τύπο του ολοκληρώματος:

$$I=\int_0^1\sin(1+x^2)dx$$

ποια τιμή του N αρκεί για να έχουμε $|I-S_N|\le 10^{-5}$; $I-S_N=-(b-a)\frac{h^2}{12}f''(\xi)$

Απάντηση

α) Ο απλός τύπος του τραπεζίου για ένα ολοκλήρωμα πάνω στο διάστημα [a,b] προκύπτει αν προσεγγίσουμε την ολοκληρωτέα συνάρτηση f(x) με ένα πολυώνυμο παρεμβολής στα σημεία x_0=a, x_1=b και χρησιμοποιήσουμε το ολοκλήρωμα αυτού του πολυωνύμου ως προσέγγιση του ολοκληρώματος της δοσμένης συνάρτησης:

$$f(x) = f(a)\frac{x-b}{a-b} + f(b)\frac{x-a}{b-a} + \frac{f''(\xi)}{2!}(x-a)(x-b), \quad \xi \in [a,b]$$

Ολοκληρώνοντας τα δύο μέρη αυτής της ισότητας έχουμε:

$$\int_a^b f(x)dx = \frac{b-a}{2}(f(a)+f(b)) - \frac{(b-a)^3}{12} f''(\eta), \quad \eta \in [a,b]$$

όπου ο πρώτος όρος του 2ου μέρους είναι η προσέγγιση και προκύπτει από την ολοκλήρωση του πολυωνύμου παρεμβολής,

$$\int_a^b f(x)dx \approx f(a)\int_a^b \frac{x-b}{a-b}dx + f(b)\int_a^b \frac{x-a}{b-a}dx = f(a)\frac{b-a}{2} + f(b)\frac{b-a}{2}$$

ενώ ο δεύτερος όρος προσδιορίζει το σφάλμα αυτής της προσέγγισης και προκύπτει από το ολοκλήρωμα του σφάλματος παρεμβολής με εφαρμογή του 2ου θεωρήματος μέσης τιμής του ολοκληρωτικού λογισμού:

$$\frac{1}{2!}\int_a^b f''(\xi(x))(x-a)(x-b)dx = \frac{1}{2}f''(\eta)\int_a^b (x-a)(x-b)dx = -\frac{(b-a)^3}{12}f''(\eta)$$

Ο σύνθετος τύπος του τραπεζίου προκύπτει αν χωρίσουμε το διάστημα [a,b] σε N ισομήκη υποδιαστήματα: [a,b]=[x_0, x_N]= [x_0, x_1]∩ [x_1, x_2]∩… ∩[x_{N-1}, x_N] όπου, αν η είναι το εύρος κάθε υποδιαστήματος, ισχύει:

$$x_k = x_0 + kh = a + kh, \quad k = 0,1,2,\ldots,N$$

και

$$h = \frac{x_N - x_0}{N} = \frac{b-a}{N}$$

Έτσι έχουμε

$$\int_a^b f(x)dx = \int_{x_0}^{x_N} f(x)dx = \sum_{k=0}^{N-1}\int_{x_k}^{x_{k+1}} f(x)dx$$

όπου, επειδή είναι

$$x_{k+1}-x_k=h$$

ισχύει

$$\int_{x_k}^{x_{k+1}} f(x)dx = \frac{h}{2}(f_k + f_{k+1}) - \frac{h^3}{12}f''(\eta_k), \quad \eta_k \in [x_k, x_{k+1}]$$

Με άθροιση αυτών των σχέσεων κατά μέλη έχουμε

$$\int_a^b f(x)dx = \sum_{k=0}^{N-1}\int_{x_k}^{x_{k+1}} f(x)dx = \sum_{k=0}^{N-1}\left\{\frac{h}{2}(f_k + f_{k+1}) - \frac{h^3}{12}f''(\eta_k)\right\} = S_N + E_N$$

όπου

$$S_N = h\left\{\frac{f_0 + f_N}{2} + \sum_{k=1}^{N-1} f_k\right\}$$

και

$$E_N = -\frac{h^3}{12}\sum_{k=0}^{N-1} f''(\eta_k) = -\frac{h^3}{12} N f''(\eta) = -(b-a)\frac{h^2}{12} f''(\eta), \quad \eta \in [a,b]$$

β) Με βάση τον δοσμένο τύπο του σφάλματος ισχύει:

$$|I - S_N| \le (b-a)\frac{h^2}{12}\max_{\xi\in[a,b]} |f''(\xi)|$$

όπου εδώ είναι [a,b]=[0,1] και $f(x)=\sin(1+x^2)$. Επομένως είναι:

$f'(x)=2x\cos(1+x^2)$, και $f''(x)=2\cos(1+x^2)+4x^2\sin(1+x^2)$,

οπότε ισχύει:

$$|f''(x)| \le 2|\cos(1+x^2)| + 4x^2|\sin(1+x^2)| \le 2+4x^2$$

και

$$\max_{\xi\in[0,1]}|f''(\xi)| \le \max_{\xi\in[0,1]}(2+4\xi^2)=6$$

Επομένως είναι

$$|I - S_N| \le \frac{h^2}{12}\cdot 6 = \frac{h^2}{2} = \frac{1}{2N^2}$$

γιατί h=(b-a)/N=1/N.

Για να ισχύει

$$|I - S_N| \le 10^{-5}$$

αρκεί, άρα, να είναι

$$\frac{1}{2N^2} \le 10^{-5}$$

Η ανισότητα αυτή ισχύει για N>223.6. Άρα αρκεί να είναι: N≥224.

***5**. Με την μέθοδο Μερικών Βημάτων (Gauss-Seidel) να προσεγγισθεί η λύση του συστήματος:*

$$\begin{aligned} 10x_1 + x_2 \qquad &= 20 \\ x_1 + 10x_2 + 2x_3 &= 30 \\ 2x_2 + 10x_3 &= 40 \end{aligned}$$

(α) Με αρχική τιμή (0,0,0) να γίνουν δύο επαναλήψεις και με κατάλληλη νόρμ να γίνει εκ των υστέρων εκτίμηση του σφάλματος κατά συνιστώσα.

(β) Πόσες επαναλήψεις αρκούν για να είναι το σφάλμα κατά συνιστώσα μικρότερο από 10^{-6};

Απάντηση

Επιλύουμε κάθε μία από αυτές τις εξισώσεις ως προς τον άγνωστο με τον απόλυτα μεγαλύτερο συντελεστή και φέρνουμε έτσι το σύστημα από την μορφή Ax=b στην μορφή x=c+Mx:

$$\begin{aligned} x_1 &= 2 - 0.1x_2 \\ x_2 &= 3 - 0.1x_1 - 0.2x_3 \\ x_3 &= 4 - 0.2x_2 \end{aligned}$$

δηλαδή,

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 2 \\ 3 \\ 4 \end{pmatrix} + \begin{pmatrix} 0 & -0.1 & 0 \\ -0.1 & 0 & -0.2 \\ 0 & -0.2 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$$

όπου M είναι ο πίνακας του 2ου μέρους και c το σταθερό διάνυσμα.
Η επαναληπτική Μέθοδος Μερικών Βημάτων συνίσταται στην άμεση χρήση κάθε νέας τιμής που προκύπτει για μίαν συνιστώσα για τον υπολογισμό βελτιωμένων τιμών για τις επόμενές της συνιστώσες. Έχει, δηλαδή, την μορφή:

$$x_1^{(k+1)} = 2 - 0.1x_2^{(k)}$$
$$x_2^{(k+1)} = 3 - 0.1x_1^{(k+1)} - 0.2x_3^{(k)}$$
$$x_3^{(k+1)} = 4 - 0.2x_2^{(k+1)}$$

με k=0,1,2,…. Ικανή συνθήκη για την σύγκλιση αυτής της διαδικασίας είναι το να ισχύει: $7M7_\infty<1$. Εδώ είναι:

$$7M7_\infty = \max_{i=1,2,3}\{|m_{i1}|+|m_{i2}|+|m_{i3}|\} = \max\{0.1,\ 0.1+0.2, 0.2\} = 0.3 < 1$$

Άρα θα υπάρχει σύγκλιση, πράγμα που φαίνεται και από το αρχικό σύστημα, γιατί έχει διαγώνια υπεροχή.
Με βάση την παραπάνω επαναληπτική διαδικασία έχουμε λοιπόν:

(k)	(0)	(1)	(2)	(2)-(1)
$x_1^{(k)}$	0	2	1.8	-0.2
$x_2^{(k)}$	0	2.8	2.132	-0.668
$x_3^{(k)}$	0	3.44	3.5736	0.1336

Η εκτίμηση σφάλματος κατά συνιστώσα μπορεί να γίνει μόνο με την διανυσματική νορμ

$$7x7_\infty := \max_{i=1,2,3}|x_i|$$

η οποία είναι και η μόνη για την οποία υπάρχει τύπος εκτίμησης του σφάλματος στην περίπτωση της μεθόδου Μερικών Βημάτων. Αν ονομάσουμε $s=(s_1, s_2, s_3)^T$ την ακριβή λύση του συστήματος, ενώ $x^{(k)} =(x_1^{(k)}, x_2^{(k)}, x_3^{(k)})^T$ είναι η k προσέγγισή της,τότε ισχύει για την μέθοδο Μερικών Βημάτων:

$$\left\|s - x^{(k)}\right\|_\infty \le \frac{\|M\|_\infty}{1-\|M\|_\infty}\left\|x^{(k)} - x^{(k-1)}\right\|_\infty \le \frac{\|M\|_\infty^{\ k}}{1-\|M\|_\infty}\left\|x^{(1)} - x^{(0)}\right\|_\infty$$

Άρα με

$$7x^{(2)}-x^{(1)}7_\infty := \max_{i=1,2,3}|x_i^{(2)}-x_i^{(1)}| = \max\{0.2, 0.668, 0.1336\} = 0.668$$

προκύπτει

$$\left\|s - x^{(2)}\right\|_\infty \le \frac{\|M\|_\infty}{1-\|M\|_\infty}\left\|x^{(2)} - x^{(1)}\right\|_\infty = \frac{0.3}{0.7}0.668 = 0.28629$$

Αυτό σημαίνει ότι

$$\max_{i=1,2,3}|s_i - x_i^{(2)}| := 7s - x^{(2)}7_\infty \le 0.28629$$

άρα, για i=1,2,3:

$$|s_i - x_i^{(2)}| \le 0.28629$$

ή

$$s_i = x_i^{(2)} \pm 0.28629\ ,\ i=1,2,3$$

β) Η δεύτερη από τις ανισότητες εκτίμησης σφάλματος μας δίνει ότι:

$$\left\|s-x^{(k)}\right\|_\infty \le \frac{\|M\|_\infty^{\,k}}{1-\|M\|_\infty}\left\|x^{(1)}-x^{(0)}\right\|_\infty = \frac{(0.3)^k}{0.7}\times 3.44$$

Για να είναι 7s - $x^{(k)}$7$_\infty$ ≤10^{-6}, αρκεί, επομένως να ισχύει:

$$\frac{(0.3)^k}{0.7}\times 3.44 \le 10^{-6}$$

ή, με λήψη των δεκαδικών λογαρίθμων στα δύο μέρη:

$$k\log(0.3)+\log(3.44)-\log(0.7)\le -6$$

Άρα

$$k \ge \frac{-6-\log(3.44)+\log(0.7)}{\log(0.3)} \approx 12.797$$

αυτό σημαίνει ότι για k≥13 θα έχουμε την απαιτούμενη ακρίβεια.

5*.(α) Με την Μέθοδο Ελαχίστων Τετραγώνων για συστήματα να προσδιοριστεί η «ψευδολύση» του ασυνεπούς γραμμικού συστήματος:*

2x-y=3, x+y=3, x-2y=-1

Πώς προκύπτει το σύστημα που δίνει την ψευδολύση;

(β) Να προσεγγισθεί με την μέθοδο Newton-Raphson και με μία μόνο επανάληψη η λύση του μη γραμμικού συστήματος:

$x^y y^x$=2, xy+2x/y=6

Για διευκόλυνση των πράξεων μπορείτε να μετασχηματίσετε σε άλλη, ισοδύναμη μορφή την πρώτη εξίσωση. $J_f(x^{(k)})(x^{(k+1)}-x^{(k)}) = -f(x^{(k)})$

Απάντηση

α) Εδώ έχουμε ένα υπερπροσδιορισμένο σύστημα Av=b (δηλαδή ένα σύστημα που έχει περισσότερες εξισώσεις από ότι αγνώστους). Το σύστημα αυτό μπορεί να γραφτεί και στην μορφή Av-b=θ, όπου θ το μηδενικό διάνυσμα. Ψευδολύση ενός τέτοιου συστήματος λέγεται το διάνυσμα v που ελαχιστοποιεί το τετράγωνο της ευκλείδιας νορμ: 7Av-b7$_2^2$.

Αν μηδενίσουμε τις μερικές παραγώγους αυτής της παράστασης ως προς τις συνιστώσες του v καταλήγουμε στο σύστημα:

$$A^TAv=A^Tb$$

που έχει τόσες εξισώσεις όσους και αγνώστους και επομένως είναι εν γένει μονοσήμαντα επιλύσιμο. A^T είναι εδώ ο ανάστροφος του πίνακα A.

Στην προκειμένη περίπτωση ζητούμε το v=$(x,y)^T$, για το οποίο είναι:

$$7Av-b7_2^2=(2x-y-3)^2+(x+y-3)^2+(x-2y+1)^2=\min$$

Επειδή το αρχικό σύστημα μπορεί να γραφτεί στην μορφή:

$$\begin{pmatrix} 2 & -1 \\ 1 & 1 \\ 1 & -2 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 3 \\ 3 \\ -1 \end{pmatrix}$$

το σύστημα που δίνει την ψευδολύση είναι:

$$\begin{pmatrix} 2 & 1 & 1 \\ -1 & 1 & -2 \end{pmatrix}\begin{pmatrix} 2 & -1 \\ 1 & 1 \\ 1 & -2 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 2 & 1 & 1 \\ -1 & 1 & -2 \end{pmatrix}\begin{pmatrix} 3 \\ 3 \\ -1 \end{pmatrix}$$

δηλαδή

$$\begin{pmatrix} 6 & -3 \\ -3 & 6 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 8 \\ 2 \end{pmatrix}$$

Η λύση αυτού του συστήματος, δηλαδή η ψευδολύση του αρχικού, είναι:

x=2, y=4/3

β) Αν λογαριθμίσουμε την πρώτη από τις εξισώσεις που δόθηκαν το σύστημα παίρνει την μορφή:

$$0 = f(x,y) := y\ln x + x\ln y - \ln 2, \quad 0 = g(x,y) := xy + \frac{2x}{y} - 6$$

Η επαναληπτική μέθοδος Newton-Raphson είναι:

$$\begin{pmatrix} f_x^{(k)} & f_y^{(k)} \\ g_x^{(k)} & g_y^{(k)} \end{pmatrix}\begin{pmatrix} x^{(k+1)} - x^{(k)} \\ y^{(k+1)} - y^{(k)} \end{pmatrix} = -\begin{pmatrix} f^{(k)} \\ g^{(k)} \end{pmatrix}$$

όπου $f_x := \partial f/\partial x$, $f_y := \partial f/\partial y$, $f^{(k)} := f(x^{(k)}, y^{(k)})$, κ.ο.κ.

Άρα έχουμε εδώ:

$$\begin{pmatrix} y/x + \ln y & \ln x + x/y \\ y + 2/y & x - 2x/y^2 \end{pmatrix}^{(k)}\begin{pmatrix} x^{(k+1)} - x^{(k)} \\ y^{(k+1)} - y^{(k)} \end{pmatrix} = -\begin{pmatrix} y\ln x + x\ln y - \ln 2 \\ xy + 2x/y - 6 \end{pmatrix}^{(k)}$$

όπου $(\ldots)^{(k)}$ σημαίνει ότι όλες οι συναρτήσεις μέσα στην παρένθεση λαμβάνονται στο σημείο $(x^{(k)}, y^{(k)})^T$. Το σημείο $(x^{(0)}, y^{(0)})^T$ πρέπει να είναι εδώ τέτοιο ώστε να μην γίνεται ο πίνακας του συστήματος ιδιόμορφος σε αυτό το σημείο ($\det(J_f) \neq 0$). Αν, λοιπόν, βάλουμε: $x^{(0)} = y^{(0)} = 1$ έχουμε:

$$\begin{pmatrix} 1 & 1 \\ 3 & -1 \end{pmatrix}\begin{pmatrix} x^{(1)} - 1 \\ y^{(1)} - 1 \end{pmatrix} = -\begin{pmatrix} -\ln 2 \\ -3 \end{pmatrix}$$

και λύνοντας αυτό το σύστημα βρίσκουμε:

$$x^{(1)} = \frac{\ln 2 + 7}{4} = 1.92329, \; y^{(1)} = \frac{3\ln 2 + 1}{4} = 0.76986$$

@@@

Θέμα από άλλη εξέταση

7.(α) Να λυθεί με την Μέθοδο Πεπερασμένων Διαφορών με h = 0.25 το επόμενο πρόβλημα συνοριακών τιμών (Αρκεί η αναγραφή του συστήματος):

$3y''-(1+x)y'+2xy=1+x$, $y(2)=0$, $y(3)=1$

(β) Να κατασκευασθεί η μέθοδος Taylor δευτέρας τάξεως για την αριθμητική λύση του προβλήματος αρχικών τιμών:

$y' = xe^y+ye^x$, $y(0)=0.5$

και να προσεγγισθεί με αυτόν η τιμή $y(0.2)$ με βήμα $h = 0.1$.

Απάντηση

7α) Στο διάστημα μεταξύ των συνοριακών σημείων, [a,b]=[2,3], επιλέγουμε τα ισαπέχοντα σημεία:

$$x_k=a+kh=2+0.25k,\ k=0,1,2,3,4$$

και θα προσδιορίσουμε σε αυτές τις θέσεις προσεγγίσεις Y_k των τιμών $y_k:=y(x_k)$ της ζητούμενης λύσης $y(x)$. Δηλαδή, θα είναι $Y_k\approx y_k=y(x_k)$, $k=0(1)4$.
Οι δύο συνοριακές τιμές είναι ήδη γνωστές:

$$Y_0=y_0=y(2)=0,\ Y_4=y_4=y(3)=1$$

Για τον προσδιορισμό των ενδιαμέσων τιμών Y_1, Y_2, Y_3, εφαρμόζουμε την διαφορική εξίσωση στα ενδιάμεσα σημεία x_1, x_2,x_3,

$$3y''(x_k)-(1+x_k)y'(x_k)+2x_ky_k =1+x_k,\ \ k=1,2,3$$

και αντικαθιστούμε τις τιμές $y'(x_k)$, $y''(x_k)$ με τις προσεγγίσεις τους:

$$y'(x_k)=\frac{y_{k+1}-y_{k-1}}{2h}+O(h^2),\quad y''(x_k)=\frac{y_{k+1}-2y_k+y_{k-1}}{h^2}+O(h^2)$$

Παραλείποντας τους όρους σφάλματος, $O(h^2)$, έχουμε έτσι το σύστημα εξισώσεων:

$$3\frac{Y_{k+1}-2Y_k+Y_{k-1}}{h^2}-(1+x_k)\frac{Y_{k+1}-Y_{k-1}}{2h}+2x_kY_k=1+x_k,\quad k=1,2,3$$

ή

$$Y_{k-1}\left(3+\frac{h}{2}(1+x_k)\right)+Y_k\left(-6+2x_kh^2\right)+Y_{k+1}\left(3-\frac{h}{2}(1+x_k)\right)=(1+x_k)h^2,\quad k=1,2,3$$

Έτσι έχουμε τελικά το σύστημα:

$$k=1:\quad Y_0\left(3+\frac{h}{2}(1+x_1)\right)+Y_1\left(-6+2x_1h^2\right)+Y_2\left(3-\frac{h}{2}(1+x_1)\right)=(1+x_1)h^2$$

$$k=2:\quad Y_1\left(3+\frac{h}{2}(1+x_2)\right)+Y_2\left(-6+2x_2h^2\right)+Y_3\left(3-\frac{h}{2}(1+x_2)\right)=(1+x_2)h^2$$

$$k=3:\quad Y_2\left(3+\frac{h}{2}(1+x_3)\right)+Y_3\left(-6+2x_3h^2\right)+Y_4\left(3-\frac{h}{2}(1+x_3)\right)=(1+x_3)h^2$$

όπου είναι $x_0=2$, $x_1=2.25$, $x_2=2.5$, $x_3=2.75$, $x_4=3$.

7β) Η μέθοδος Taylor για τον προσδιορισμό της λύσης ενός προβλήματος αρχικών τιμών:

$$y'=f(x,y),\ y(x_0)=y_0$$

υπολογίζει προσεγγίσεις Y_k των τιμών y_k:=$y(x_k)$ της ζητούμενης λύσης y(x) στα τα ισαπέχοντα σημεία:

$$x_k=x_0+kh,\ k=1,2,3,\ldots$$

χρησιμοποιώντας άμεσα το ανάπτυγμα Taylor της y(x) γύρω από την εκάστοτε θέση x_k:

$$y_{k+1}=y(x_{k+1})=y(x_k)+hy'(x_k)+\frac{h^2}{2}y''(x_k)+O(h^3)$$

όπου είναι

$$y'(x_k)=f(x_k,y_k)$$

ενώ

$$y''(x)=\frac{dy'}{dx}=\frac{d}{dx}f(x,y(x))=\frac{\partial f}{\partial x}(x,y)+\frac{\partial f}{\partial y}(x,y)\cdot y'=f_x(x,y)+f_y(x,y)\cdot f(x,y)$$

άρα

$$y''(x_k)=(f_x+f_y\cdot f)(x_k,y_k)$$

Αντικαθιστώντας τα $y'(x_k)$, $y''(x_k)$ στο ανάπτυγμα Taylor με τις τιμές που δίνουν οι σχέσεις αυτές και παραλείποντας τον όρο σφάλματος, $O(h^3)$, έχουμε έτσι την προσέγγιση:

$$Y_{k+1}=Y_k+hf(x_k,Y_k)+\frac{h^2}{2}(f_x+f_y\cdot f)(x_k,Y_k)$$

που προσδιορίζει αναδρομικά κάθε νέα τιμή Y_{k+1} με βάση την προηγούμενή της, Y_k, k=0,1,2,3,….

Εδώ είναι

$$x_k=0+kh=0.1k,\quad k=0,1,2,\ldots$$

και

$$f(x,y)=xe^y+ye^x,\ f_x=e^y+ye^x,\ f_y=xe^y+e^x$$

Άρα η αναδρομική αυτή σχέση παίρνει την μορφή:

$$Y_{k+1}=Y_k+h\left(x_ke^{Y_k}+Y_ke^{x_k}\right)+\frac{h^2}{2}\left(e^{Y_k}+Y_ke^{x_k}+\left(x_ke^{Y_k}+e^{x_k}\right)\cdot\left(x_ke^{Y_k}+Y_ke^{x_k}\right)\right)$$

Επειδή είναι h=0.1, αν η σχέση αυτή εφαρμοσθεί για k=0 με $Y_0=y(x_0)=y(0)=0.5$ μας δίνει την προσέγγιση Y_1 της τιμής $y_1=y(x_1)=y(0.1)$. Αν εφαρμόσουμε ξανά την αναδρομική αυτή σχέση για k=1, τότε παίρνουμε την ζητούμενη προσέγγιση Y_2 για την τιμή $y_2=y(x_2)=y(0.2)$.

Παρατηρήσεις:

1. Σε άλλες εξετάσεις έχει ζητηθεί ο προσδιορισμός λύσεων των εξής εξισώσεων:
sin x =1/x –1, sin 2x =1/x, 4 – x^4 = |x - 1|, cos2x=1/x, x^2-1=ln(x+1), cosx=1/x, 1/x=ln|x+1|, x^2-1=ln|x+1|, x^2-2=1/x, sin(1+x)=1/x, 3-x^2=sin(1+x) sinx=1/(x-1) 1/x= $\sqrt[3]{1+x}$, cos x=ln|1+x|.

2. Κατασκευή του πολυωνύμου παρεμβολής Lagrange έχει ζητηθεί για τις συναρτήσεις:
f(x):= ln(x^2+3x+2) με σημεία παρεμβολής: x_0=0, x_1=0.25, x_2=0.5, καθώς και άνω φράγμα του σφάλματος για x∈[-0.5,1]
f(x):= ln(1 - x^2) = ln{(1-x)(1+x)} στα σημεία x_0=0, x_1=0.25, x_2=0.5, και φράγμα για x∈[-0.5, 0.5],
f(x)= cos(0.5x+1), x_0=0, x_1=0.25, x_2=0.5, για –1≤x≤1.
f(x)= cos(x+0.5), x_0=0, x_1=0.25, x_2=0.5, για –0.5 ≤ x ≤ 1.
f(x):= $\sqrt[3]{1+x}$, x_0=0, x_1=0.1, x_2=0.2, για –0.2 ≤ x ≤ 0.2
f(x):= $\sqrt{1+2x}$ Lagrange x_0=0, x_1=0.1, x_2=0.2, γιά –0.2 ≤ x ≤ 0.2.
f(x)= $(1+x)^{1/3}$ στα σημεία x_0=0, x_1=0.2, x_2=0.4, για το [-0.2,0.2]
f(x):= $\sqrt{1-2x}$, x_0=0, x_1=0.1, x_2=0.2, για -0.2 ≤ x ≤ 0.2
f(x)= $(2+x)^{1/3}$ x_0=0, x_1=0.2, x_2=0.4 για το [-0.2,0.2] .
f(x):= ln(1-x^2)=ln(1-x)+ln(1+x), x_0=0, x_1=0.1, x_2=0.2, –0.2 ≤ x ≤ 0.2
f(x):= $\sqrt[3]{1-2x}$, x_0=0, x_1=0.1, x_2=0.2, για –0.2 ≤ x ≤ 0.2
3. Άλλο θέμα παρεμβολής ήταν:
(β) Δίνεται ο πίνακας τιμών της y = f(x):

j	1	2	3	4
x_j	-2	1	0	1
y_j	-3	-1	1	2

Με κατασκευή του πολυωνύμου παρεμβολής x= p(y) προσεγγίστε το σημείο x όπου είναι y=0.

4. Έχει ζητηθεί η κατασκευή του σύνθετου τύπου ολοκλήρωσης των μεθόδων τραπεζίου ή Simpson για τα ολοκληρώματα:

$$I=\int_1^3 \ln\{(x+1)(x+2)\}dx,\ I=\int_1^3 \ln\{(2x+1)(x+1)\}dx,\ I=\int_0^1 \ln(2+x^3)dx,$$

$$I=\int_0^1 \sin(1+x^2)dx,\ I=\int_0^1 \cos(1+x^2)dx,\ I=\int_0^1 \ln(1+x^2)dx,$$

$$I=\int_1^2 \ln(9-x^2)dx,\ I=\int_0^2 \sin(6-x^2)dx,\ I=\int_0^2 \cos(1+x^3)dx.$$

5. Έχει ζητηθεί η κατασκευή με την μέθοδο ελαχίστων τετραγώνων πολυωνύμου $2^{ου}$ βαθμού, που να προσεγγίζει τις ακόλουθες συναρτήσεις με βάση τις τιμές τους σε ωρισμένα σημεία:

$f(x):= \sqrt{2-x}$: $x_0=-1$, $x_1=-0.5$, $x_2= 0$, $x_3=0.5$, $x_4=1$.

$f(x):= 1/(2-x^2)$: $x_0=-1$, $x_1=-0.5$, $x_2= 0$, $x_3=0.5$, $x_4=1$.

$f(x):= 1/(1+x^2)$: $x_0=-1$, $x_1=-0.5$, $x_2= 0$, $x_3=0.5$, $x_4=1$.

$f(x):= \ln(2+x^2)$: $x_0=-1$, $x_1=-0.5$, $x_2= 0$, $x_3=0.5$, $x_4=1$.

$f(x):= \ln(1+x^2)$: $x_0=-1$, $x_1=-0.5$, $x_2= 0$, $x_3=0.5$, $x_4=1$.

6. Η Μέθοδος Πεπερασμένων Διαφορών με h = 0.25 ή h=0.2 ζητήθηκε για προβλήματα όπως το ακόλουθο:

$$y''-(1+x)y'-x^2y=1/x,\ y(2)=-1,\ y(3)=0$$

7. Η Μέθοδος Taylor ζητήθηκε για $y' = \sin(x^2+y)$, $y(1)=0.5$, με εύρεση του $y(1.2)$ με $h = 0.1$.

$y' = \ln(2x+y)$, $y(1)=0.5$, $y(1.2)$ με βήμα $h = 0.1$.

$y' = \ln(x+3y)$, $y(1)=0.5$

$y' = \ln(x+y^2)$, $y(1)=0.5$, $y(1.2)$ με $h = 0.1$.

$y' = \cos(x+y^2)$, $y(1)=0.5$, $y(1.2)$ με βήμα $h = 0.1$.

8. Η λύση Μη Γραμμικών Συστημάτων με την μέθοδο Newton – Raphson έχει ζητηθεί για τα συστήματα:

α) $\sin(x_1+ x_2) - x_1x_2 = 2$, $x_1^3+ x_2^2 -2 x_1x_2 = 3$

β) $x^yy^x=2$, $xy+x/y=4$

γ) $x^yy^x=2$, $xy+2x/y=6$

δ) $\sin(x+y) =1/2$, $xy+x/y=4$

9. Έχει ζητηθεί η «ψευδολύση» συστημάτων όπως το ακόλουθο:

$2x-y=3$, $x+y=3$, $x-2y=1$, $3x-2y=2$

ΠΙΝΑΚΑΣ ΠΕΡΙΕΧΟΜΕΝΩΝ

ΠΡΟΛΟΓΟΣ.....1
ΕΙΣΑΓΩΓΗ 3
1. ΕΚΤΙΜΗΣΗ ΑΡΙΘΜΗΤΙΚΩΝ ΣΦΑΛΜΑΤΩΝ.....6
Παράδειγμα Φυσικής Εφαρμογής 6
1.1 Η Προέλευση των Αριθμητικών ή Υπολογιστικών Σφαλμάτων 6
1.2 Φράγματα Υπολογιστικών Σφαλμάτων 6
Απόδειξη 8
2. ΕΠΑΝΑΛΗΠΤΙΚΗ ΛΥΣΗ ΜΗ ΓΡΑΜΜΙΚΩΝ ΕΞΙΣΩΣΕΩΝ ΜΕ ΕΝΑΝ ΑΓΝΩΣΤΟ.....11
Παραδείγματα Φυσικών και Τεχνικών Εφαρμογών:.....11
2.1 Γραφικός Εντοπισμός των Λύσεων.....13
Παραδείγματα Γραφικού Εντοπισμού Λύσεων 16
Παρατηρήσεις: Επιτάχυνση σύγκλισης και Εξασφάλιση Σύγκλισης.....28
Παράδειγμα Εφαρμογής της Μεθόδου Τέμνουσας 30
Παρατήρηση: Η Μέθοδος Steffensen 31
2.5 Έκτίμηση Σφάλματος.....31
Ασκήσεις και Εφαρμογές 33
3. ΠΟΛΥΩΝΥΜΙΚΕΣ ΕΞΙΣΩΣΕΙΣ.....34
3.1 Εντοπισμός Ριζών Πολυωνύμου 34
Παράδειγμα: Φράγμα των Ριζών Πολυωνύμου 35
Ασκήσεις.....36
4. ΝΟΡΜ ΔΙΑΝΥΣΜΑΤΩΝ ΚΑΙ ΠΙΝΑΚΩΝ 38
4.1 Νορμ Διανυσμάτων.....38
Ασκήσεις.....40
4.2 Νορμ Πινάκων.....41
Ασκήσεις.....43
5. ΕΠΙΛΥΣΗ ΓΡΑΜΜΙΚΩΝ ΣΥΣΤΗΜΑΤΩΝ.....44
5.1 Επαναληπτικές Μέθοδοι.....44
5.1.1. Μέθοδος Ολικών Βημάτων.....45
Παρατήρηση: Διαγώνια Υπεροχή 46
Παράδειγμα Εφαρμογής της Μεθόδου ολικών Βημάτων.....48
5.1.2 Μέθοδος Μερικών Βημάτων (Gauss - Seidel).....50
Παράδειγμα της Επιρροής του τρόπου Απαλοιφής στο Αριθμητικό Αποτέλεσμα 55
Παράδειγμα Εφαρμογής απαλοιφής Gauss.....56
Παρατήρηση: Αντιστροφή ενός Πίνακος.....57
5.3 Ασταθή συστήματα – Βελτίωση Προσεγγιστικής Λύσης.....57
5.4 Ανάλυση σε Γινόμενο Τριγωνικών Πινάκων 59

Η Μέθοδος Cholesky 61
5.5 Ασυνεπή Γραμμικά Συστήματα-Ψευδοαντίστροφοι Πίνακες 63
Παράδειγμα Γενικευμένης Λύσης Ασυνεπούς Συστήματος 64
Ασκήσεις 65
6. ΕΠΑΝΑΛΗΠΤΙΚΗ ΕΠΙΛΥΣΗ ΜΗ ΓΡΑΜΜΙΚΩΝ ΣΥΣΤΗΜΑΤΩΝ 67
Παραδείγματα Φυσικών και Τεχνικών Εφαρμογών 67
6.1 Εντοπισμός Λύσεων 68
6.2 Η Μέθοδος Newton – Raphson για Συστήματα 69
Ασκήσεις 71
7. ΠΑΡΕΜΒΟΛΗ 73
7.1 Λαγκρανζιανή μορφή του Πουωνύμου Παρεμβολής 74
Ασκήσεις 76
7.2 Η νευτώνεια Μορφή του Πολυωνύμου Παρεμβολής 76
Παρεμβολή Newton με ισαπέχοντα σημεία παρεμβολής 78
Παρατήρηση: Η Σχέση Παρεμβολής Newton και Αναπτύγματος Taylor 79
Παράδειγμα Παρεμβολής Newton με Ισαπέχοντες Κόμβους 80
7.3 Το Σφάλμα Παρεμβολής 81
Το Μέγεθος του Σφάλματος Παρεμβολής 82
Παράδειγμα Προσδιορισμού του Σφάλματος Παρεμβολής 84
Παράδειγμα Προσδιορισμού του Κατάλληλου Βήματος για την κατασκευή ενός Πίνακος Τιμών μίας Συνάρτησης 85
7.4 Παρεμβολή με ομαλές τμηματικά πολυωνυμικές προσεγγίσεις (Splines) 86
Παράδειγμα Προσέγγισης με Τριτοβάθμια Συνάρτηση Spline 89
Ασκήσεις 89
8. ΠΟΛΥΔΙΑΣΤΑΤΗ ΠΑΡΕΜΒΟΛΗ 91
8.1 Πολυωνυμική Παρεμβολή και Ρητή Παρεμβολή για Συναρτήσεις με Δύο Μεταβλητές 91
Παράδειγμα Παρεμβολής Shepard - Lagrange 91
Παράδειγμα Ρητής Παρεμβολής Shepard 92
9. ΠΡΟΣΕΓΓΙΣΕΙΣ ΜΕ ΤΗΝ ΜΕΘΟΔΟ ΕΛΑΧΙΣΤΩΝ ΤΕΤΡΑΓΩΝΩΝ 94
Εφαρμογές 94
9.1 Πως προκύπτει η μέθοδος αυτή 94
Παράδειγμα 1 98
Παράδειγμα 2 98
Παράδειγμα 3 99
Ασκήσεις 101
10. ΑΡΙΘΜΗΤΙΚΗ ΔΙΑΦΟΡΙΣΗ 103
Παρατήρηση: Ελαχιστοποίηση του Φράγματος του Σφάλματος 104
Παρατήρηση: Η επίδραση των σφαλμάτων στρογγύλευσης 105

Ασκήσεις: 107
11. ΑΡΙΘΜΗΤΙΚΗ ΟΛΟΚΛΗΡΩΣΗ 108
Παραδείγματα Φυσικών και Τεχνικών Εφαρμογών 108
11.1 Ολοκλήρωση με Χρήση Παρεμβολής σε Ισαπέχοντα Σημεία 108
1. Τύπος Μέσου Σημείου 110
2. Τύπος του Τραπεζίου 111
3. Τύπος του Simpson 111
Σύνθετοι Τύποι Ολοκλήρωσης 112
1. Σύνθετος Τύπος του Τραπεζίου 112
2. Σύνθετος Τύπος του Simpson 113
Παρατήρηση: Η Αριθμητική Ολοκλήρωση μόνο Συνεχών Συναρτήσεων 113
Παραδείγματα Εφαρμογής των Συνθέτων Τύπων 114
Ασκήσεις 115
11.2 Οι τύποι ολοκλήρωσης του Gauss 116
11.3 Γενικευμένα Ολοκληρώματα 116
11.3.1 Ολοκληρώματα Απείρου Διαστήματος 117
Παραδείγματα υπολογισμού ολοκληρωμάτων απείρου διαστήματος 118
Ασκήσεις 120
11.3.2 Ολοκληρώματα Πεπερασμένου Διαστήματος με Αλγεβρική Ανωμαλία 120
Παραδείγματα Υπολογισμού Ολοκληρωμάτων με Ιδιομορφίες 122
Ασκήσεις 124
Παρατήρηση: Η Περίπτωση Λογαριθμικής Ανωμαλίας 124
12. ΤΟ ΠΡΟΒΛΗΜΑ ΑΡΧΙΚΩΝ ΤΙΜΩΝ ΓΙΑ ΚΟΙΝΕΣ ΔΙΑΦΟΡΙΚΕΣ ΕΞΙΣΩΣΕΙΣ 126
Φυσικά και Τεχνικά Παραδείγματα 126
12.1 Η Γενική Μορφή του Προβλήματος Αρχικών Τιμών 129
12.2 Συνέπειες της Παράβασης της Συνθήκης Lipschitz 129
Δεύτερο Παράδειγμα 131
12.3 Η Αναγωγή του Προβλήματος Αρχικών Τιμών για Συστήματα ΚΔΕ στην Μορφή $y'=f(x,y)$, $y(x_0)=y_0$ 132
Παράδειγμα 132
12.4 Η Αναγωγή του Προβλήματος Αρχικών Τιμών για ΚΔΕ Ανωτέρας Τάξεως στην Μορφή $y'=f(x,y)$, $y(x_0)=y_0$ 132
Παράδειγμα μετατροπής Δ.Ε. $2^{ας}$ τάξεως σε σύστημα Δ.Ε. $1^{ης}$ τάξεως 133
12.5 Ο Τρόπος Παραγωγής Αριθμητικών Μεθόδων 133
12.6 Η Μέθοδος Taylor 134
Παράδειγμα εφαρμογής της μεθόδου Taylor 135
12.7 Μέθοδοι Runge-Kutta 136
Παράδειγμα Εφαρμογής της Μεθόδου Runge-Kutta 139

Παράδειγμα Λύσης Συστήματος ΚΔΕ με την Μέθοδο Runge-Kutta-O(h^2)... 140
12.8 Η Γενική Μορφή του Ολικού Σφάλματος για Αριθμητικές Μεθόδους Λύσης του Προβλήματος Αρχικών Τιμών ... 140
12.9 Η Μέθοδος Heun (Χόιν) ... 145
Παράδειγμα Εφαρμογής της Μεθόδου Heun ... 146
12.10 Μέθοδοι Πολλών Βημάτων Ανωτέρας Τάξεως ... 147
*Σημείωση ... 148
12.11 Ο Έλεγχος του Σφάλματος της Μεθόδου Predictor-Corrector ... 150
12.12 Απόλυτη Ευστάθεια – Ανάλυση Σφαλμάτων ... 150
Συνθήκες Απόλυτης Ευστάθειας της Μεθόδου Heun ... 151
Απόλυτη Ευστάθεια της Μεθόδου Heun σε Μορφή PECE ... 153
Παράδειγμα ... 155
Απόλυτη Ευστάθεια της μεθόδου Adams-O(h^4) σε Μορφή PECE ... 155
Άσκηση ... 155
12.13 Σχετική Ευστάθεια ... 156
12.14 Χαοτική Συμπεριφορά Μη Γραμμικών Συστημάτων ... 157
*Παραδείγματα Δ.Ε. με Χαοτική Συμπεριφορά ... 159
12.15 Δύσκαμπτες Δ.Ε. (Stiff Differential Eq.) ... 160
Ασκήσεις ... 162
Εφαρμογή ... 163
Παραδείγματα Φυσικών και Τεχνικών Εφαρμογών ... 164
13.1 Οι Βασικές Μέθοδοι ... 164
13.2 Η Μέθοδος Βολής ... 165
13.2 Η Μέθοδος Πεπερασμένων Διαφορών ... 165
Μικτές Συνοριακές Συνθήκες ... 167
Παραδείγματα Εφαρμογής της Μεθόδου Πεπερασμένων Διαφορών ... 168
13.3 Η Μέθοδος Πεπερασμένων Στοιχείων ... 169
Παράδειγμα Εφαρμογής της Μεθόδου Πεπερασμένων Στοιχείων ... 172
Ασκήσεις ... 173
14. ΠΡΟΣΔΙΟΡΙΣΜΟΣ ΙΔΙΟΤΙΜΩΝ ΚΑΙ ΙΔΙΟΔΙΑΝΥΣΜΑΤΩΝ ... 175
Παραδείγματα Φυσικών και Τεχνικών Εφαρμογών ... 175
14.1 Εισαγωγή ... 178
14.2 Γενικές Ιδιότητες των Ιδιοδιανυσμάτων και Ιδιοτιμών ... 179
14.3 Φράγματα Ιδιοτιμών ... 182
Παρατήρηση: Απλά Φράγματα Ιδιοτιμών ... 183
14.4 Προσδιορισμός της Μεγίστης Ιδιοτιμής – Μέθοδος των Δυνάμεων ... 184
Το Σφάλμα της Προσέγγισης αν $A^T=A$ ... 187
Παράδειγμα Εύρεσης της Μέγιστης Ιδιοτιμής με την Μέθοδο των Δυνάμεων ... 187

14.5 Προσδιορισμός της Ελαχίστης Ιδιοτιμής με την Μέθοδο των αντιστρόφων Δυνάμεων 188
Ασκήσεις 189
ΒΙΒΛΙΟΓΡΑΦΙΑ 190
ΣΥΝΤΟΜΗ ΕΙΣΑΓΩΓΗ ΣΤΟ MATLAB 192
ΤΟ ΠΕΡΙΒΑΛΛΟΝ ΕΡΓΑΣΙΑΣ ΤΟΥ MATLAB 193
ΤΡΟΠΟΣ ΧΡΗΣΗΣ ΤΟΥ ΠΕΡΙΒΑΛΛΟΝΤΟΣ MATLAB 193
ΟΙ ΒΑΣΙΚΕΣ ΣΥΝΑΡΤΗΣΕΙΣ ΤΟΥ MATLAB 197
ΠΡΟΓΡΑΜΜΑΤΙΣΜΟΣ ΣΤΟ MATLAB 197
Εισαγωγή Δεδομένων 198
Καταγραφή Προγράμματος 198
Στοιχεία της δομής των προγραμμάτων 198
Εισαγωγή Δεδομένων κατά την εκτέλεση προγράμματος: 199
Εντολές Εκτύπωσης 201
Χρονομέτρηση Προγράμματος: 201
Εντολές Γραφικών Παραστάσεων 201
1. Γραφήματα συναρτήσεων μίας μεταβλητής. 201
Εισαγωγή Τίτλων, Επιγραφών, Κειμένου στο Γράφημα 202
Γράφημα Παραμετρικά Οριζόμενης Καμπύλης 202
Πολλαπλά Γραφήματα με την plot 203
Πολλαπλά Γραφήματα 204
Παράλληλη παρουσίαση περισσότερων γραφημάτων: subplot 204
2. Γραφικές Απεικονίσεις Επιφανειών ή Καμπύλων στον χώρο 205
Πολυεδρική προσέγγιση επιφανείας: mesh, surf και meshgrid 205
3. Προσδιορισμός Προοπτικής Άποψης από ωρισμένη θέση ή κατεύθυνση: view 205
4. Χρωματικό Προφίλ: colormap (χρωματικός συνδυασμός) 206
5. Χρωματική Σκίαση: shading 206
Εκτύπωση Γραφικών 206
ΑΡΙΘΜΗΤΙΚΟΑΝΑΛΥΤΙΚΕΣ ΣΥΝΑΡΤΗΣΕΙΣ ΤΗΣ MATLAB 207
ΒΙΒΛΙΟΓΡΑΦΙΑ 209
ΕΠΕΞΕΡΓΑΣΙΑ ΣΥΜΒΟΛΙΚΩΝ ΠΑΡΑΣΤΑΣΕΩΝ ΜΕ ΤΟ MATLAB 209
Ορισμός συμβολικών αντικειμένων 209
Συμβολικές Μεταβλητές 210
Απλοποίηση γραφής 211
Μεταβολή Αριθμητικής Ακρίβειας 214
Αναλυτικές Πράξεις 214
Ανάλυση ρητής παραστάσεως 216
Γραφική Παράσταση Συμβολικών Παραστάσεων μίας Μεταβλητής 216

Επίλυση Αλγεβρικών Εξισώσεων 217
Επίλυση Διαφορικών Εξισώσεων 218
ΣΥΜΒΟΛΙΚΗ ΓΡΑΜΜΙΚΗ ΑΛΓΕΒΡΑ 219
Ιδιοδιανύσματα και Ιδιοτιμές 220
Προσδιορισμός Ιακωβιανού Πίνακα 221
ΜΕΤΑΣΧΗΜΑΤΙΣΜΟΙ Laplace, Fourier, Z 221
ΠΑΡΑΔΕΙΓΜΑΤΑ ΠΡΟΓΡΑΜΜΑΤΩΝ ΤΟΥ MATLAB 223
Παρατηρήσεις για τα προγράμματα που ακολουθούν 223
Χαοτική συμπεριφορά μεθόδων εύρεσης σημείων μηδενισμού 224
Πολυωνυμική παρεμβολή Lagrange 226
Αντίστροφη Παρεμβολή 229
Μέθοδος ελαχίστων τετραγώνων για πολυωνυμικές προσεγγίσεις 230
Μελέτη προβλημάτων αρχικών τιμών για ΚΔΕ 232
Σημείωση: Τα παραπάνω προγράμματα αντιγράφηκαν από τον MATLAB editor/debugger με την ακόλουθη διαδικασία: 237
Υπόδειγμα Λύσης των Θεμάτων μίας Τυπικής Εξέτασης 238
Υποδείξεις: 239
Απαντήσεις των Θεμάτων 239
Θέμα από άλλη εξέταση 247