ΣΧΟΛΗ ΕΦΑΡΜΟΣΜΕΝΩΝ ΜΑΘΗΜΑΤΙΚΩΝ ΚΑΙ ΦΥΣΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
ΕΜΠ

ΣΗΜΕΙΩΣΕΙΣ ΓΙΑ ΤΟ ΜΑΘΗΜΑ

ΥΠΟΛΟΓΙΣΤΙΚΗ ΦΥΣΙΚΗ

Ε. Δρυς,
Γ. Κουτσούμπας
Ν.Δ. Τράκας

ΣΕΠΤΕΜΒΡΙΟΣ 2003
2η έκδοση

ΕΙΣΑΓΩΓΗ

Οι σημειώσεις αυτές έχουν στόχο να καλύψουν τις πρώτες ανάγκες του μαθήματος της 'Υπολογιστικής Φυσικής' που διδάσκεται στο 7ο εξάμηνο της ΣΕΜΦΕ (κατεύθυνση Φυσικού Εφαρμογών). Ο σκοπός δεν είναι να διδάξει τους φοιτητές Αριθμητική Ανάλυση, αλλά να τους εισάγει στην εφαρμογή γνωστών μεθόδων της Αριθμητικής Ανάλυσης σε φυσικά προβλήματα.

Πολλά ενδιαφέροντα θέματα δεν καλύπτονται στις σημειώσεις αυτές. Πρόθεσή μας είναι με την αλληλεπίδραση με τους φοιτητές, στην πορεία διαμόρφωσης του μαθήματος, σε μεταγενέστερες εκδόσεις να συμπληρώσουμε το παρόν κείμενο.

Θα είμαστε ευγνώμονες σε κάθε παρατήρηση/διόρθωση.

Θέλουμε να ευχαριστήσουμε τον Σταμάτη Νίκολη για την ευγενική παραχώρηση των σημειώσεών του που χρησιμοποιεί στο Πανεπιστήμιο της Tour, στη Γαλλία. Οι παρούσες σημειώσεις στηρίζονται κατά ένα μεγάλο βαθμό σ' αυτές.

Επίσης, ευχαριστούμε τους Σοφρώνη Παπαδόπουλο για την προσεκτική ανάγνωση των σημειώσεων καθώς και τους Κώστα Παρασκευαΐδη και Γιώργο Τσιπολίτη για τη συμμετοχή τους στην Επιτροπή Σύνταξης των σημειώσεων του μαθήματος. Τέλος, ευχαριστούμε τον Θεοφάνη Γραμμένο για τις επισημάνσεις και διορθώσεις του πριν την δεύτερη έκδοση.

Οι συγγραφείς

Περιεχόμενα

Κεφάλαιο 1

Σχεδίαση Δυναμικών Γραμμών

Οι ηλεκτρικές δυναμικές γραμμές και οι ισοδυναμικές επιφάνειες είναι οι πιο συνηθισμένοι τρόποι παράστασης ενός ηλεκτροστατικού πεδίου. Οι δυναμικές γραμμές ορίζονται από δύο ιδιότητες: (1) Σε κάθε σημείο, η εφαπτομένη στη δυναμική γραμμή είναι παράλληλη με το διάνυσμα του ηλεκτρικού πεδίου στο σημείο αυτό και (2) Σε κάθε σημείο ο αριθμός των δυναμικών γραμμών (που διαπερνούν κάθετα μια μοναδιαία επιφάνεια) είναι ανάλογος με την ένταση του πεδίου στο σημείο αυτό. Η δεύτερη ιδιότητα μας λέει ότι όσο οι δυναμικές γραμμές γίνονται πιο πυκνές, τόσο το ηλεκτρικό πεδίο (και επομένως και η δύναμη που ασκείται σ' ένα φορτίο) γίνεται ισχυρότερο. Η πρώτη ιδιότητα μας προσφέρει μια εύκολη υπολογιστική μέθοδο για να σχεδιάζουμε τις δυναμικές γραμμές. Η κατανόηση και η περιγραφή φυσικών καταστάσεων και φαινομένων με βάση τις δυναμικές γραμμές είναι πολύ βασική και χρήσιμη διεργασία στον ηλεκτρισμό και το μαγνητισμό. Η απόκτηση πείρας στην αναγνώριση της μορφής των δυναμικών γραμμών για ορισμένες κατανομές φορτίων, βοηθάει στην καλλίτερη κατανόηση και χρήση των ηλεκτρικών πεδίων.

Οι ισοδυναμικές επιφάνειες είναι επιφάνειες στο χώρο πάνω στις οποίες τα φορτία μπορούν να μετακινούνται χωρίς την παραγωγή (ή κατανάλωση) έργου. Οι ισοδυναμικές επιφάνεις είναι παντού κάθετες στις ηλεκτροστατικές δυνάμεις ($\mathbf{F} \cdot d\mathbf{l}=0$). Επομένως, οι ισοδυναμικές επιφάνειες είναι παντού κάθετες στις δυναμικές γραμμές του πεδίου. Η διαγραφή μιας καμπύλης που βρίσκεται πάνω σε μία ισοδυναμική επιφάνεια επιτυγχάνεται αν κινούμαστε μονίμως κάθετα στην τοπική δυναμική γραμμή. Αυτή η ιδιότητα μας βοηθά στην σχεδίαση ισοδυναμικών καμπυλών με εύκολο υπολογιστικό τρόπο.

1.1 Βασικές Εξισώσεις

Οι παρακάτω εξισώσεις σχολιάζονται πληρέστερα σ' όλα τα βιβλία φυσικής. Εδώ αναφέρουμε όσες θα μας χρειασθούν σ' αυτό το κεφάλαιο.

1. Ο νόμος του Coulomb
$$\mathbf{F} = \frac{1}{4\pi\epsilon_0}\frac{q\,q_1}{r^2}\,\hat{\mathbf{r}}$$
όπου $\mathbf{F}$ είναι η δύναμη που ασκείται στο φορτίο q από το φορτίο q_1 και $\hat{\mathbf{r}}$ είναι το μοναδιαίο διάνυσμα από το φορτίο q_1 στο φορτίο q και r η μεταξύ τους απόσταση

2. Το αξίωμα της επαλληλίας
$$\mathbf{F}_{\text{ολική}} = \mathbf{F}_1 + \mathbf{F}_2 + \mathbf{F}_3 + \cdots \mathbf{F}_N = \sum_{i=1}^{N}\mathbf{F}_i = \sum_{i=1}^{N}\frac{q\,q_i}{4\pi\epsilon_0 r_i^2}\,\hat{\mathbf{r}}_\mathbf{i}$$
όπου $\mathbf{F}_{\text{ολική}}$ είναι η δύναμη που ασκείται στο φορτίο q από όλα τα άλλα φορτία $q_1, \ldots, q_N$, r_i η απόσταση του q_i φορτίου από το q και $\hat{\mathbf{r}}_\mathbf{i}$ το αντίστοιχο μοναδιαίο διάνυσμα (από το q_i στο q).

3. Ο ορισμός του ηλεκτρικού πεδίου που οφείλεται σε σημειακό φορτίο q
$$\mathbf{E}(r) = \frac{1}{4\pi\epsilon_0}\frac{q}{r^2}\,\hat{\mathbf{r}}$$
όπου $\mathbf{E}(r)$ είναι ένα διανυσματικό πεδίο. Είναι δηλαδή μια συνάρτηση που οι τιμές της είναι διανύσματα και που αλλάζει από σημείο σε σημείο στο χώρο. Η δύναμη $\mathbf{F}$ που ασκείται σε φορτίο q_0 που βρίσκεται στο σημείο r_0 δίνεται από τη σχέση:
$$\mathbf{F} = \mathbf{E}(r_0)\,q_0$$

4. Οι δυναμικές γραμμές του ηλεκτροστατικού πεδίου ξεκινούν από θετικά φορτία και καταλήγουν σε αρνητικά φορτία ή στο άπειρο.

5. Ολα τα σημεία μιας ισοδυναμικής επιφάνειας έχουν το ίδιο ηλεκτροστατικό δυναμικό.

6. Οι ηλεκτροστατικές δυναμικές γραμμές είναι παντού κάθετες στις ισοδυναμικές επιφάνειες.

1.2 Υπολογισμός των Δυναμικών Γραμμών

Οι δυναμικές γραμμές μπορούν να υπολογιστούν αλγοριθμικά, δηλαδή με μια μέθοδο βήμα-προς-βήμα . Η διαδικασία αυτή χρησιμοποιεί την πρώτη από τις ιδιότητες που αναφέραμε παραπάνω: η εφαπτομένη στη δυναμική γραμμή είναι πάντα παράλληλη με το διάνυσμα του ηλεκτρικού πεδίου. Ας περιοριστούμε

σε προβλήματα που είναι, ή μπορεί να αναχθούν, σε δύο διαστάσεις Σχ.(1.1). Θεωρήστε το ηλεκτρικό πεδίο στο τυχαίο σημείο (x, y) του επιπέδου, που προκαλείται από ομοεπίπεδα σημειακά φορτία. Το συνολικό πεδίο $\mathbf{E}$ βρίσκεται αν αθροίσουμε τους όρους $(q/4\pi\epsilon_0)(\hat{\mathbf{r}}/r^2)$ για όλα τα φορτία. Τώρα υποθέστε ότι μετακινείστε κατα ένα μικρό διάστημα Δs πάνω στη δυναμική γραμμή, από το σημείο (x, y) έως το σημείο $(x + \Delta x, y + \Delta y)$. Για μικρό βήμα Δs, το τρίγωνο που σχηματίζεται από τα Δx, Δy και Δs είναι όμοιο με το τρίγωνο που σχηματίζουν τα E_x, E_y και $\sqrt{E_x^2 + E_y^2}$. Επομένως, τα Δx και Δy δίνονται από τις σχέσεις

$$\Delta x = \Delta s \frac{E_x}{\sqrt{E_x^2 + E_y^2}}, \qquad \Delta y = \Delta s \frac{E_y}{\sqrt{E_x^2 + E_y^2}}$$

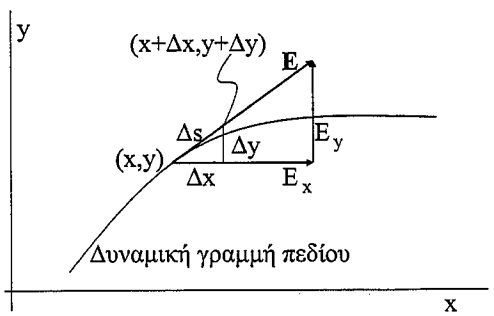
Το νέο σημείο της δυναμικής γραμμής είναι το $(x+\Delta x, y+\Delta y)$ και η διαδικασία αρχίζει ξανά. Μ' αυτό το τρόπο προχωρούμε πάνω στη δυναμική γραμμή βήμα-προς-βήμα.

Η διαδικασία για την ισοδυναμική γραμμή (δηλαδή την τομή της ισοδυναμικής επιφάνειας με το επίπεδο που βρίσκονται τα φορτία), είναι κι αυτή εύκολη. Επειδή οι ισοδυναμικές είναι κάθετες στις δυναμικές γραμμές ή στο πεδίο $\mathbf{E}$, σε κάθε βήμα η κίνηση πρέπει να είναι κάθετη στο ηλεκτρικό πεδίο. Γνωρίζουμε ότι το γινόμενο των κλίσεων δυο καθέτων ευθειών είναι -1. Επομένως αν θέλουμε να ακολουθήσουμε μια ισοδυναμική γραμμή, θα πρέπει να μετακινηθούμε κατά Δx και Δy

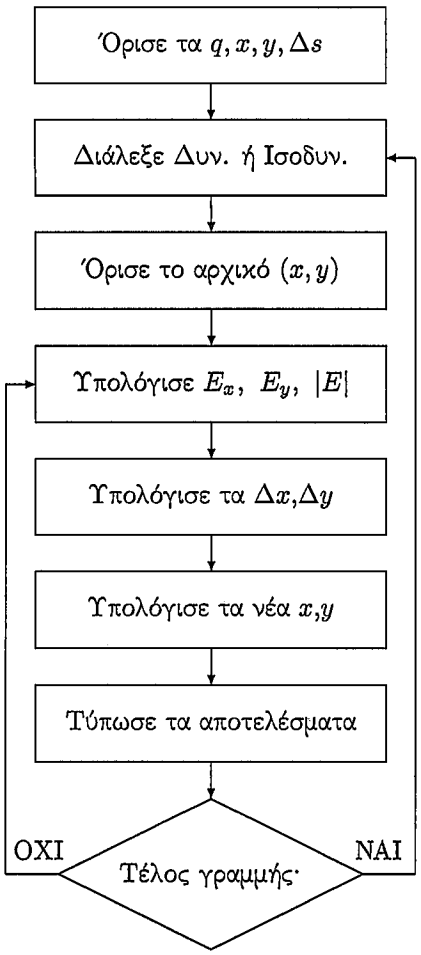
$$\Delta x = \Delta s \frac{E_y}{\sqrt{E_x^2 + E_y^2}} \quad \Delta y = \Delta s \frac{-E_x}{\sqrt{E_x^2 + E_y^2}}$$

παράλληλα με τους άξονες x και y. Η επιλογή του αρνητικού προσήμου στο Δy, και όχι στο Δx είναι τυχαία. Η διαφορά έγκειται στη φορά διαγραφής της ισοδυναμικής γραμμής.

Στο Σχ.(1.2) φαίνεται το λογικό διάγραμμα για τη διαδικασία υπολογισμού των δυναμικών και ισοδυναμικών γραμμών.



Σχήμα 1.1: Σε κάθε σημείο της δυναμικής γραμμής, τα τρίγωνα $(\Delta x, \Delta y, \Delta s)$ και $E_x, E_y, |E|)$ είναι όμοια.



Σχήμα 1.2: Διάγραμμα ροής για την σχεδίαση δυναμικών και ισοδυναμικών γραμμών. Μετά τον καθορισμό των αρχικών τιμών των παραμέτρων, ο υπολογισμός 'προχωρά' πάνω στην δυναμική ή ισοδυναμική γραμμή βήμα - προς - βήμα. Σε κάθε σημείο, η ένταση του πεδίου καθορίζει τις συντεταγμένες του επόμενου βήματος.

Παρακάτω φαίνεται ο κώδικας σε γλώσσα FORTRAN που μπορεί να χρησιμοποιηθεί για σχεδίαση δυναμικών ή ισοδυναμικών γραμμών.

```
C     ******************************
      PROGRAM ELECTRIC_FIELDS
C     ******************************
C     ******************1o TMHMA*****************************
      IMPLICIT NONE
      INTEGER I,N,KCHECK,P,ICON,L
      PARAMETER (P=20)
      REAL X(P),Y(P),Q(P),X0IN,Y0IN,X0,Y0,RCHECK,EX,EY,DL,DLIN,
     *DX,DY
      OPEN(UNIT=10,FILE="ELEC_LINE.DAT")
C     ******************2o TMHMA*************************
      PRINT *,"NUMBER OF CHARGES (MAX=20)"
      READ(5,*)N

      DO I=1,N
         PRINT *,I,"CHARGE"
         PRINT *,"GIVE THE CHARGE AND ITS POSITION (Q,X,Y)"
         READ(5,*)Q(I),X(I),Y(I)
      END DO

666   PRINT *,"LINE OF FORCE (1) OR EQUIPOTENTIAL LINE (2)"
      READ(5,*)L
      PRINT *,"CHOOSE THE STARTING POINT (X,Y)"
      READ(5,*)X0IN,Y0IN
C     ******************3o TMHMA*****************************
      X0=X0IN
      Y0=Y0IN
      DX=0.0
      DY=0.0
      DL=0.05
      EX=0.0
      EY=0.0
C     ******************4o TMHMA*************************
      DO I=1,N
         RCHECK=((X0-X(I))**2+(Y0-Y(I))**2)**(0.5)

         IF (RCHECK.LT.DL) THEN
            PRINT *,"TOO NEAR TO A CHARGE"
            GOTO 666
         END IF

      END DO
```

```
C     ******************5ο ΤΜΗΜΑ*************************
      KCHECK=1
      DO WHILE (KCHECK.EQ.1)

      WRITE(10,*)X0,Y0
C     PRINT *,X0,Y0
      CALL STRENGTH(X0,Y0,N,X,Y,Q,EX,EY,DX,DY)

      IF (L.EQ.1) THEN
         DX=DL*EX/((EX**2+EY**2)**(0.5))
         X0=X0+DX
         DY=DL*EY/((EX**2+EY**2)**(0.5))
         Y0=Y0+DY
      ELSE IF (L.EQ.2) THEN
         DX=DL*EY/((EX**2+EY**2)**(0.5))
         X0=X0+DX
         DY=-DL*EX/((EX**2+EY**2)**(0.5))
         Y0=Y0+DY
      ELSE
      END IF

      IF ((ABS(X0).GT.20.0).OR.(ABS(Y0).GT.20.0)) KCHECK=0

      IF (L.EQ.1) THEN
         DO I=1,N
            RCHECK=((X0-X(I))**2+(Y0-Y(I))**2)**(0.5)

            IF (RCHECK.LT.DL) THEN
               KCHECK=0
C              PRINT *,I,X0,Y0,RCHECK
            END IF

         END DO
      ELSE
            DLIN=ABS(X0-X0IN)+ABS(Y0-Y0IN)
            IF (DLIN.LT.0.9*DL) KCHECK=0
      ENDIF


      END DO
```

```
      ******************6ο ΤΜΗΜΑ*****************************
667   PRINT *,"NEW STARTING POINT (YES=1, NO=2)"
      READ(5,*)ICON
      IF (ICON.EQ.1) GOTO 666

      CLOSE(10)
      END
C     ****************************************
C     ******************7ο ΤΜΗΜΑ*****************************
      SUBROUTINE STRENGTH(X0,Y0,N,X,Y,Q,EX,EY,DX,DY)
      IMPLICIT NONE
      INTEGER I,N,P
      PAMETER(P=20)
      REAL X0,Y0,X(P),Y(P),Q(P),EX,EY,R,DX,DY

      EX=0.0
      EY=0.0

      DO I=1,N
         R=((X0+DX/2.0-X(I))**2+(Y0+DY/2.0-Y(I))**2)**(0.5)

         EX=EX+Q(I)*(X0+DX/2.0-X(I))/R**3
         EY=EY+Q(I)*(Y0+DY/2.0-Y(I))/R**3

      END DO

      RETURN
      END
```

Ας προσπαθήσουμε, επειδή είναι ο πρώτος κώδικας που παρουσιάζεται, να δώσουμε μια όσο το δυνατό λεπτομερή επεξήγηση.
Το *1ο τμήμα* περιέχει όλους τους ορισμούς των μεταβλητών που θα χρησιμοποιηθούν. Η εντολή:
IMPLICIT NONE
είναι πολύ βοηθητική γιατί απαγορεύει την εμφάνιση στον κώδικα, οποιαδήποτε μεταβλητής που δεν έχει οριστεί στην αρχή με τις εντολές INTEGER και REAL και έτσι αποφεύγονται λάθη που προέρχονται από κακή πληκτρολόγηση μεταβλητών. Οι πραγματικές μεταβλητές Q, X και Y, που καθορίζουν τα φορτία και τις θέσεις τους, είναι πίνακες, οπότε πρέπει να ορίσουμε την (μέγιστη) διάστασή τους. Με την εντολή:
PARAMETER (P=20)

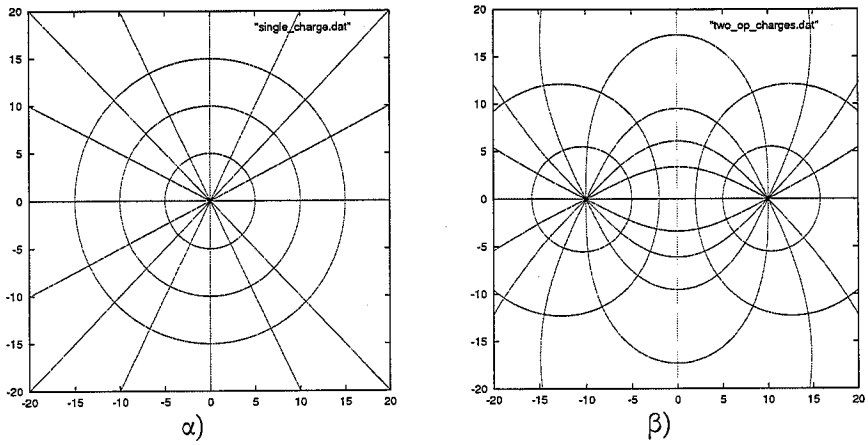
μπορούμε να ορίσουμε αυτή τη μέγιστη διάσταση. Τέλος με την εντολή:
OPEN(UNIT=10,FILE="ELEC_LINE.DAT"):
'ανοίγουμε' ένα αρχείο (με το όνομα ELEC_LINE.DAT) όπου θα 'γράφουμε' τα αποτελέσματά μας.

Στο *2ο τμήμα* δίνουμε στο πρόγραμμα τα αρχικά δεδομένα: τα φορτία και τις θέσεις τους. Στην εντολή:
READ(5,*)Q(I),X(I),Y(I)
το '5' υποδηλώνει ότι τα δεδομένα εισάγονται από το πληκτρολόγιο και ο '*' ότι ο τρόπος που δίνονται (δεκαδική μορφή, εκθετική μορφή κ.λπ.) είναι τελείος γενικός. Ακολούθως, το πρόγραμμα ρωτά αν θέλουμε να υπολογίσει μία δυναμική γραμμή ή μια ισοδυναμική γραμμή (επιλέγοντας τη τιμή της μεταβλητής L). Τέλος, δηλώνουμε τις συντεταγμένες του σημείου από το οποίο θέλουμε να ξεκινήσει η γραμμή: X0IN,Y0IN.

Στο *3ο τμήμα* γίνονται οι αρχικοποιήσεις των τιμών: τα X0,Y0 θα αποτελούν τις συντεταγμένες του εκάστοτε νέου σημείου που βρίσκουμε (τα X0IN,Y0IN τα χρειαζόμεστε για διάφορους ελέγχους κατά τη διάρκεια της διαδικασίας). Το DL είναι το μήκος του βήματος που επιλέγουμε και DX,DY είναι οι προβολές της μετακίνησής μας.

Στο *4ο τμήμα* ελέγχουμε κατά πόσο το αρχικό σημείο που επιλέξαμε είναι πολύ κοντά (μικρότερο από DL) σε κάποιο από τα φορτία του προβλήματός μας.

Το *5ο τμήμα* είναι το κυρίως πρόγραμμα. Σ' αυτό καλείται η υπορουτίνα STRENGTH όπου με εισαγωγή των συντεταγμένων του σημείου που βρισκόμαστε (X0,Y0), των φορτίων και των θέσεών τους (Q,X,Y) λαμβάνουμε ώς έξοδο τις τοπικές συντεταγμένες του διανύσματος του πεδίου (τα DX,DY χρειάζονται για την καλλίτερη προσέγγιση αυτής της αλγοριθμικής διαδικασίας, βλ. το σχολιασμό της υπορουτίνας). Με δεδομένα πια τα τοπικά EX,EY υπολογίζουμε τα νέα DX,DY καθώς και το νέο σημείο της γραμμής (διακρίνοντας την περίπτωση δυναμικής ή ισοδυναμικής γραμμής). Μετά αρχίζουν οι έλεγχοι διακοπής αυτής της διαδικασίας: (1) αν το X0 ή το Y0 'βγούν' εκτός της περιοχής που ενδιαφερόμαστε (η γραμμή απομακρύνεται), (2) αν το νέο σημείο μας X0,Y0 είναι πολύ κοντά σε κάποιο φορτίο (για την περίπτωση δυναμικής γραμμής) και (3) αν το νέο σημείο μας είναι πολύ κοντά (ουσιαστικά συμπίπτει) με το αρχικό μας σημείο (για την περίπτωση ισοδυναμικής γραμμής). Αν όλοι οι έλεγχοι ξεπεραστούν θετικά αρχίζει ο υπολογισμός του επόμενου σημείου της γραμμής (αφού αποθηκευτεί το νέο σημείο στο αρχείο που ήδη έχει ανοιχτεί), ειδάλλως, το *6ο τμήμα* του προγράμματος μας ζητά αν θέλουμε να ξεκινήσουμε ένα νέο υπολογισμό ή να τερματίσουμε όλη τη διαδικασία. Τέλος 'κλείνουμε' το αρχείο των αποτελεσμάτων μας.



Σχήμα 1.3: Δυναμικές και ισοδυναμικές γραμμές για α) σημειακό φορτίο και β) δύο αντίθετα σημειακά φορτία

Στην υπορουτίνα υπολογίζουμε την απόσταση R του σημείου που βρισκόμαστε (X0,Y0) από κάθε φορτίο, υπολογίζουμε τη συνεισφορά του φορτίου αυτού στο EX και EY και αθροίζουμε για όλα τα φορτία. Ακολουθούμε το λεγόμενο τρόπο του 'μισού βήματος' όπου η θέση του σημείου που βρισκόμαστε είναι X0+DX/2.0,Y0+DY/2.0, δηλαδή αυξάνομε τις τιμές των συντεταγμένων κατά το ήμισυ των αντιστοίχων συνιστωσών του προηγούμενου βήματος. Μία απλή εξήγηση είναι ότι είναι ακριβέστερο να βρισκόμαστε στο μέσον του διαστήματος που μετακινούμαστε παρά στην άκρη του.

Ας δούμε τώρα μερικά απλά παραδείγματα για να ελέγξουμε τη μέθοδό μας. Στην περίπτωση ενός σημειακού φορτίου οι δυναμικές γραμμές είναι ακτινικές και οι ισοδυναμικές κύκλοι. Στο Σχ.(1.3α) βλέπουμε τη σχεδίαση που παίρνουμε χρησιμοποιώντας τα αποτελέσματα του προγράμματος. Στο Σχ.(1.3β) βλέπουμε τις δυναμικές και τις ισοδυναμικές γραμμές για δύο αντίθετα φορτία.

Η μέθοδος εφαρμόζεται για απεριόριστο αριθμό σημειακών φορτίων και μπορεί εύκολα να γενικευτεί και για συνεχείς κατανομές.

Κεφάλαιο 2

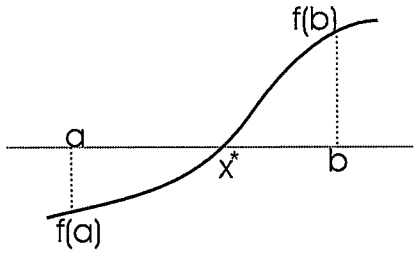
Εύρεση Ριζών μιας Εξίσωσης

Ξεκινάμε με το πιο συνηθισμένο αριθμητικό πρόβλημα, δηλαδή πώς βρίσκουμε τις ρίζες μιας εξίσωσης.

Είναι γνωστό ότι ο αναλυτικός υπολογισμός των ριζών μιας αυθαίρετης δεδομένης εξίσωσης είναι γενικά αδύνατος. Για την περίπτωση των πολυωνυμικών εξισώσεων (που και θεωρούνται οι πιο 'απλές' περιπτώσεις), πέρασαν πράγματι πολλοί αιώνες (Αναγέννηση) προτού προχωρήσουμε από τις γνωστές λύσεις της δευτεροβάθμιας σ' αυτές της τριτοβάθμιας και τεταρτοβάθμιας εξίσωσης. Για πολυωνυμικές εξισώσεις βαθμού μεγαλύτερου του τετάρτου, χρειάστηκε να φτάσουμε έως τον 18ο με 19ο αιώνα για να κατανοήσουμε (με τις έρευνες των Lagrange, Abel και Galois) ότι είναι αδύνατο να εκφράσουμε τις λύσεις μιας γενικής πολυωνυμικής εξίσωσης, βαθμού ανωτέρου του τετάρτου, με τη βοήθεια πεπερασμένου αριθμού τετραγωνικών, κυβικών ή ανώτερης τάξης ριζών ρητών εκφράσεων των συντελεστών του πολυωνύμου[1]. Φυσικά υπάρχουν κλάσεις πολυωνύμων, βαθμού μεγαλύτερου του τετάρτου, των οποίων οι λύσεις μπορούν να εκφραστούν με αυτό τον τρόπο. Για παράδειγμα: η εξίσωση $a_6x^6 + a_5x^5 + a_4x^4 + 2a_5x^3 + a_4x^2 + a_5x + a_6 = 0$ μπορεί να παραγοντοποιηθεί σε γινόμενο του παράγοντα $x^2 + 1$ επί ένα πολυώνυμο τετάρτου βαθμού. Αλλά, φυσικά είναι μια ειδική περίπτωση πολυωνυμικής εξίσωσης έκτου βαθμού και όχι η γενική μορφή της.

Για πιο περίπλοκες περιπτώσεις δεν υπάρχουν γενικές εκφράσεις που δίνουν τις ρίζες. Αλλά υπάρχει και άλλο ένα σημείο, που αποτελεί και το λόγο

[1]Το εκπληκτικό είναι ότι αυτό το αποτέλεσμα παρουσιάστηκε ως ειδική περίπτωση μιας εκτεταμένης μελέτης που αναφερόταν στις συμμετρίες μεταξύ ομάδων αντικειμένων κάτω από μετάθεση. Στην περίπτωση μας, τα αντικείμενα είναι οι συντελεστές των πολυωνύμων. Αυτή ακριβώς η μελέτη μπορεί να θεωρηθεί ως το πιστοποιητικό γέννησης της θεωρίας των ομάδων, κλάδος που έμελλε να παίξει σημαντικότατο ρόλο στα μαθηματικά, τη φυσική, τη χημεία και τη βιολογία.



Σχήμα 2.1: Το θεώρημα της ενδιάμεσης τιμής: μια συνεχής συνάρτηση σ' ένα διάστημα, της οποίας οι τιμές στα άκρα έχουν διαφορετικό πρόσημο, πρέπει να μηδενίζεται σε κάποιο σημείο του διαστήματος.

ύπαρξης αυτής της μελέτης: Δυστυχώς, όσο ωραίες και αν είναι οι γενικές λύσεις των πολυωνυμικών εξισώσεων 3ου και 4ου βαθμού, προσαρμόζονται πολύ δύσκολα σε αριθμητικούς υπολογισμούς. Αυτό που θέλουμε να πούμε είναι ότι ο αριθμητικός υπολογισμός αυτών των εκφράσεων είναι τόσο επιρρεπής σε αριθμητικά σφάλματα, ώστε προτιμώνται οι μέθοδοι εύρεσης ριζών που θα παρουσιαστούν σ' αυτό το κεφάλαιο. Επιπλέον, οι μέθοδοι αυτές είναι γενικές και εφαρμόζονται τόσο σε απλές όσο και σε δύσκολες περιπτώσεις.

Το πρόβλημα μας λοιπόν μπορεί να τεθεί με τον ακόλουθο γενικό τρόπο: Αναζητούμε τις ρίζες της εξίσωσης:

$$f(x) = 0 \tag{2.1}$$

όπου $x \in \mathbb{R}^n$ και $f : \mathbb{R}^n \to \mathbb{R}^n$. Ενδιαφερόμαστε για την περίπτωση $n = 1$, αλλά είναι φανερό ότι η περίπτωση $n = 2$ ενδέχεται να εμφανισθεί στην περίπτωση *μιγαδικών* λύσεων πραγματικών πολυωνυμικών εξισώσεων.

2.1 Οι Ρίζες είναι Πραγματικές και Απλές

Για τις περιπτώσεις αυτές, δηλαδή πραγματικών ριζών, τα πράγματα - στη θεωρία - είναι μάλλον απλά: γνωρίζουμε ότι αν η συνάρτηση είναι συνεχής στο διάστημα $[a, b]$ και επιπλέον $f(a)f(b) < 0$, υπάρχει τουλάχιστον ένα σημείο x^* για το οποίο ισχύει $f(x^*) = 0$ (βλ. το Σχ.(2.1)). Αν, αντιθέτως, η συνάρτηση δεν αλλάζει πρόσημο σε κανένα διάστημα, τότε οι λύσεις είναι μιγαδικές και τα πράγματα δυσκολεύουν.

Η αξία αυτού του θεωρήματος έγκειται στο ότι μας βεβαιώνει για την ύπαρξη λύσης με πολύ μικρό αντάλλαγμα: τη γνώση της τιμής της συνάρτησης. Επίσης είναι η αφετηρία μιας προσεγγιστικής μεθόδου εύρεσης της ρίζας καθώς και ελέγχου της ακρίβειας της μεθόδου αυτής. Η στρατηγική λοιπόν είναι η ακόλουθη: κάθε φορά που βρίσκουμε κάποιο διάστημα το οποίο περιέχει τουλάχιστον μία ρίζα, προσπαθούμε να το μειώσουμε έως ότου το μέγεθός

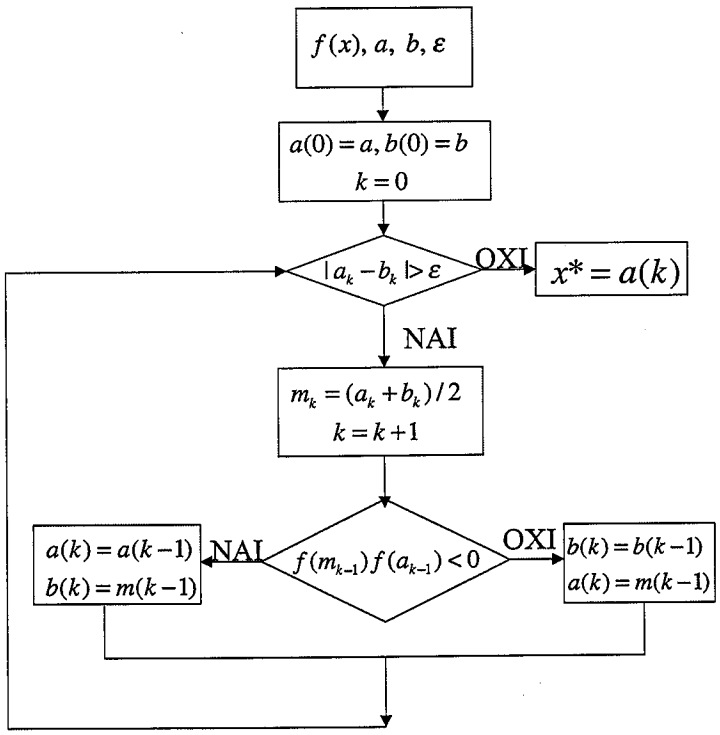
του γίνει συγκρίσιμο με την αριθμητική ακρίβεια, δηλαδή ως το σημείο που δεν μπορούμε να ξεχωρίσουμε το ένα άκρο του διαστήματος από το άλλο.

2.2 Η Μέθοδος της Διχοτόμησης

Ο αλγόριθμος μια τέτοιας διαδικασίας μπορεί να πάρει τη μορφή που φαίνεται στο Σχ.(2.2).

Η σταθερά ϵ είναι η επιθυμητή ακρίβεια του αριθμητικού υπολογισμού. Η ακρίβεια ενός τυπικού υπολογιστή (δηλαδή η μικρότερη δυνατή τιμή που μπορεί να έχει η επιθυμητή ακρίβεια), είναι της τάξης του 10^{-8} για *απλή ακρίβεια* (single precision), και 10^{-16} για *διπλή ακρίβεια* (double precision). Βλέπουμε ότι σε κάθε βήμα του αλγορίθμου, το διάστημα διαιρείται δια 2, γι' αυτό και η μέθοδος ονομάζεται μέθοδος *διχοτόμησης*.

Ας δούμε τώρα ένα σημαντικό σημείο. Με δεδομένη την ακρίβεια ϵ, πόσα βήματα θα πρέπει να κάνουμε; Η απλότητα του αλγορίθμου μας λέει αμέσως



Σχήμα 2.2: Λογικό διάγραμμα για την εύρεση των ριζών εξίσωσης με τη μέθοδο της διχοτόμησης. Η ζητούμενη λύση είναι η x^*. Θεωρούμε ότι $f(a_0) \neq 0$ και $f(b_0) \neq 0$.

ότι:

$$\epsilon = \frac{b_0 - a_0}{2^{k_{min}}} \rightarrow \ln\left(2^{k_{min}}\right) = \ln\frac{b_0 - a_0}{\epsilon} \rightarrow k_{min} = \frac{\ln\left|\frac{b_0 - a_0}{\epsilon}\right|}{\ln 2} \tag{2.2}$$

Για την καλύτερη κατανόηση, ας αναφερθούμε σε ένα συγκεκριμένο παράδειγμα.

Υποθέτουμε ότι έχουμε ένα υλικό σημείο μάζας m, συνδεδεμένο μ' ένα οριζόντιο ιδανικό ελατήριο σταθεράς K. Το υλικό σημείο μετακινείται κατά x από τη θέση ισορροπίας του. Η δυναμική του ενέργεια είναι $V = (1/2)Kx^2$. Αν E είναι η ολική του ενέργεια, τότε η μέγιστη απομάκρυνσή του θα δίνεται από τη ρίζα της εξίσωσης $E = V(x)$. Ένα απλό γράφημα μας βεβαιώνει ότι υπάρχουν 2 σημεία που επαληθεύουν αυτήν την εξίσωση, για οποιαδήποτε τιμή της ολικής ενέργειας. Η κίνηση είναι περιοδική και φραγμένη και οι λύσεις της εξίσωσης είναι ανεξάρτητες της μάζας του υλικού σημείου. Αν διαλέξουμε $K = 2\,\mathrm{J/m^2}$ και $E = 0.7\,\mathrm{J}$, καταλήγουμε στην εξίσωση:

$$E = \frac{1}{2}Kx^2 \Rightarrow x^2 = \frac{2E}{K} \Rightarrow x^2 = 0.7 \tag{2.3}$$

Για $x > 0$, μπορούμε εύκολα να δούμε ότι $0.64 = 0.8^2 < 0.7 < 0.81 = 0.9^2$, που μας οδηγεί στο διάστημα $[0.8, 0.9]$ ως αρχικό διάστημα με ακρίβεια πρώτου δεκαδικού ψηφίου. Δηλαδή ήδη ξέρουμε ότι $0.8 < x^* = 0.8... < 0.9$. Η Εξ.(2.2) μας λέει ότι τρεις επιπλέον επαναλήψεις είναι αρκετές για να υπολογίσουμε το επόμενο ψηφίο στην προσέγγισή μας, και ότι με 5 επαναλήψεις η ακρίβεια μας θα είναι 10^{-4}. Δηλαδή θα γνωρίζουμε 3 ψηφία μετά την υποδιαστολή. Πιο συγκεκριμένα, βρίσκουμε $a_1 = 0.8$, $b_1 = 0.85$, $a_2 = 0.825$, $b_2 = 0.85$, $a_3 = 0.825$, $b_3 = 0.8375$, $a_4 = 0.83125$, $b_4 = 0.8375$ και επομένως είμαστε σίγουροι ότι $x^* = 0.83...$

Ας προχωρήσουμε σε μια όχι τόσο τετριμμένη περίπτωση, παίρνοντας ένα πραγματικό ελατήριο που παρουσιάζει και μη γραμμικές αποκρίσεις, κάτι που εμφανίζεται όταν η απομάκρυνση του υλικού σημείου δεν μπορεί να θεωρείται 'μικρή' και η γραμμική απόκριση δεν είναι αρκετή για να περιγράψει το φαινόμενο. Μπορούμε να πάρουμε λοιπόν τη δυναμική ενέργεια $V[x] = (1/2)Kx^2 + ax^3$ με $a < 0$. Η μορφή αυτή μας βοηθά να καταλάβουμε τι γίνεται όταν το ελατήριο σπάει - και η μάζα ελευθερώνεται - οπότε και η ενέργεια Ε ξεπερνά μια κρίσιμη τιμή E_{cr}. Τότε η εξίσωση γίνεται:

$$ax^3 + (1/2)Kx^2 - E = 0 \tag{2.4}$$

Ας πάρουμε την προηγούμενη τιμή για την ενέργεια $E = 0.7\,\mathrm{J}$ και για τη σταθερά $K = 2\,\mathrm{J/m^2}$ και ας προσπαθήσουμε να καταλάβουμε τι ακριβώς γίνεται για ένα μικρό a ($a \simeq 0.1$ για παράδειγμα). Μπορούμε να θεωρήσουμε τον κυβικό όρο ως μια διαταραχή της Εξ.(2.3), αλλά είναι μια ιδιαίτερη διαταραχή εφόσον για $a = 0$ η εξίσωση έχει μόνο 2 λύσεις, ενώ για $a \neq 0$, όσο μικρό και

αν είναι, η Εξ. (2.4) έχει 3 πραγματικές λύσεις για $E < E_{cr}$, 2 πραγματικές (εκ των οποίων μια διπλή) για $E = E_{cr}$ και μια πραγματική για $E > E_{cr}$.

Η εξίσωση παίρνει λοιπόν τη μορφή:

$$0.7 = x^2 - 0.1x^3 = x^2(1 - (1/10)x) \equiv g(x) \tag{2.5}$$

και είναι εύκολο να καταλάβουμε ότι οι ρίζες βρίσκονται στα παρακάτω διαστήματα: $x_1 \in (-1, 0)$, $x_2 \in (0, 1)$, $x_3 \in (9, 10)$, δηλαδή $x_1 = -0....$, $x_2 = 0...$ και $x_3 = 9...$, εφόσον $(g(-1) - 0.7)\,(g(0) - 0.7) < 0$ κλπ.

Τέλος, όπως είδαμε στην Εξ.(2.2)

$$\epsilon = \frac{b_0 - a_0}{2^{k_{min}}}$$

και εφόσον τα διαστήματά μας έχουν εύρος μονάδα,

$$\begin{aligned} k_{min} = 4 \quad &\rightarrow \quad \epsilon = 1/16 \sim 0.6 \\ 5 \quad &\rightarrow \quad 1/32 \sim 0.3 \\ 6 \quad &\rightarrow \quad 1/64 \sim 0.02 \end{aligned}$$

δηλαδή με 6 επαναλήψεις το σφάλμα μας επιτρέπει να βρούμε δυο ψηφία μετά την υποδιαστολή.

2.3 Η Μέθοδος Newton-Raphson

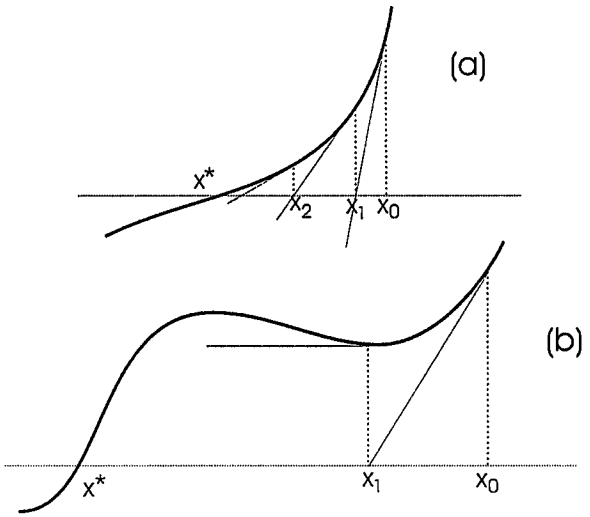
Μετά από αυτή την πρακτική άσκηση μπορούμε να συνοψίσουμε λέγοντας ότι κατέχουμε μια σίγουρη μέθοδο, και να προχωρήσουμε παρακάτω ρωτώντας αν είναι δυνατόν να αυξήσουμε τις επιδόσεις της, δηλαδή να επιταχύνουμε τη σύγκλισή της προς μια ρίζα. Τέτοια μέθοδος, η οποία χρησιμοποιείται συχνά, είναι η λεγόμενη Newton-Raphson. Το πλεονέκτημα αυτής της μεθόδου είναι ακριβώς ότι συγκλίνει γρηγορότερα από τη μέθοδο της διχοτόμησης και ότι μπορεί να γενικευτεί για περισσότερες διαστάσεις. Το μειονέκτημά της είναι ότι δεν δίνει καμιά εγγύηση για τη σύγκλιση προς τη ρίζα. Δηλαδή, αν αρχίσουμε 'κοντά' στη ρίζα η μέθοδος δουλεύει άψογα, αλλά αν είμαστε 'μακρυά' αποκλίνει θεαματικά. Φυσικά, οι έννοιες 'μακρυά' και 'κοντά' εξαρτώνται από την ίδια την εξίσωση που επιλύουμε! Επομένως, η διαίσθηση παίζει σοβαρό ρόλο και ο υπολογισμός μερικών τιμών της συνάρτησης μπορεί να δώσει πολύτιμες πληροφορίες.

Για να γίνουμε σαφείς, θα περιγράψουμε τη μέθοδο χρησιμοποιώντας το Σχ.(2.3α). Εύκολα φαίνεται ότι:

$$0 \equiv f'(x_0)x_1 + (f(x_0) - f'(x_0)x_0) \rightarrow x_1 = x_0 - \frac{f(x_0)}{f'(x_0)}, \tag{2.6}$$

σχέση που μπορεί να γενικευτεί στην:

$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)}, \quad k = 0, 1, 2, ... \tag{2.7}$$



Σχήμα 2.3: Η επαναληπτική μέθοδος Newton-Raphson. Το σημείο x_0 είναι το σημείο έναρξης, το οποίο πρέπει να είναι κοντά στη ρίζα.

Μπορούμε να δούμε ήδη μια πηγή ανησυχίας: αν η παράγωγος μηδενίζεται σε κάποιο σημείο, η μέθοδος τερματίζεται (Σχ.(2.3β)). Αλλά ακόμα πιο προβληματική είναι η περίπτωση όπου η παράγωγος στο x_k είναι περίπου μηδέν, δηλαδή $f'(x_k) \approx 0$. Τότε το σφάλμα δεν ελέγχεται πλέον, χωρίς μάλιστα να έχουμε κάποιο συγκεκριμένο 'μήνυμα σφάλματος'.

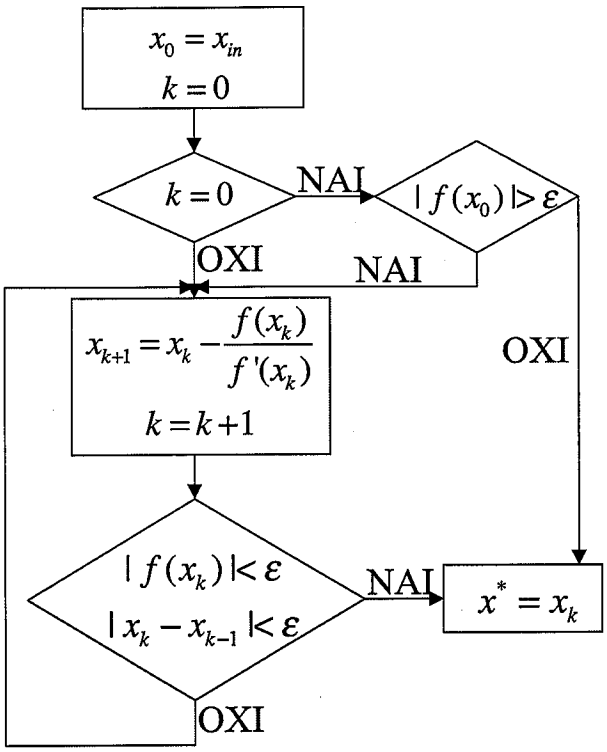
Αλλά υπάρχει και άλλο ένα πρόβλημα το οποίο σχετίζεται με τη σύγκλιση του αλγορίθμου: τίποτα δε μας βεβαιώνει, ακόμα και αν $f'(x) \neq 0$, ότι η ακολουθία $\{x_k\}$ συγκλίνει σε κάποιο όριο και ότι αυτό το όριο είναι μια λύση της εξίσωσή μας. Παρόλ' αυτά, ο αλγόριθμος των Newton-Raphson χρησιμοποιείται στις μέρες μας τόσο για την δυνατότητα του στην εύρεση ριζών όσο και για τις ιδιότητες των ακολουθιών που παράγει.

Επομένως, μετά την αναφορά όλων των μειονεκτημάτων της μεθόδου, ας περιγράψουμε και την άλλη όψη του νομίσματος: την ταχύτητα σύγκλισης όταν αυτή επιτυγχάνεται. Αν x^* είναι η ρίζα, ας συμβολίσουμε με ϵ_k το $x_k - x^*$. Χρησιμοποιούμε τώρα τη θεμελιώδη υπόθεση της μεθόδου: το ότι είμαστε αρκετά κοντά στη ρίζα και μπορούμε να αναπτύξουμε μια οριακή προσέγγιση. Από την Εξ.(2.7) παίρνουμε:

$$\epsilon_{k+1} = \epsilon_k - \frac{f(x^* + \epsilon_k)}{f'(x^* + \epsilon_k)} = \frac{\epsilon_k f'(x^* + \epsilon_k) - f(x^* + \epsilon_k)}{f'(x^* + \epsilon_k)} \qquad (2.8)$$

και από τα αναπτύγματα Taylor των $f(x)$ και $f'(x)$ περί το $x = x^*$ έχουμε:

$$\begin{aligned} f(x^* + \epsilon_k) &= 0 + \epsilon_k f'(x^*) + (1/2)\epsilon_k^2 f''(x^*) + \mathcal{O}(\epsilon_k^3) \\ f'(x^* + \epsilon_k) &= f'(x^*) + \epsilon_k f''(x^*) + (1/2)\epsilon_k^2 f'''(x^*) + \mathcal{O}(\epsilon_k^3) \end{aligned}$$



Σχήμα 2.4: Λογικό διάγραμμα για την εύρεση των ριζών εξίσωσης με τη μέθοδο Newton-Raphson. Η ζητούμενη λύση είναι η x^*.

Αντικαθιστώντας τις τελευταίες στην Εξ.(2.8) βρίσκουμε:

$$\begin{aligned} \epsilon_{k+1} &= \frac{(1/2)\epsilon_k^2 f''(x^*) + \mathcal{O}(\epsilon_k^3)}{f'(x^*) + \epsilon_k f''(x^*) + (1/2)\epsilon_k^2 f'''(x^*) + \mathcal{O}(\epsilon_k^3)} = \\ &= \epsilon_k^2 \frac{f''(x^*)}{2f'(x^*)} \frac{1 + \mathcal{O}(\epsilon_k)}{1 + \mathcal{O}(\epsilon_k)} = \epsilon_k^2 \frac{f''(x^*)}{2f'(x^*)}(1 + \mathcal{O}(\epsilon_k)) \end{aligned} \tag{2.9}$$

δηλαδή $\epsilon_{k+1} \sim \epsilon_k^2$. Μπορούμε να συγκρίνουμε αυτό το σφάλμα με το αντίστοιχο της μεθόδου της διχοτόμησης, όπου είχαμε βρει ότι (βλ. Εξ.(2.2)) $\epsilon_{k+1} = (1/2)\epsilon_k$, και να δούμε άμεσα ότι η νέα μέθοδος είναι πιο αποτελεσματική: μπορούμε να κερδίσουμε έως και 2 ψηφία ανά επανάληψη της διαδικασίας! Βέβαια η προϋπόθεση είναι να είμαστε κοντά στη ρίζα - έτσι ώστε η οριακή διαδικασία να έχει νόημα - και επίσης ο παράγοντας $f''(x^*)/(2f'(x^*))$ να είναι μικρός (της τάξης της μονάδας). Βέβαια, η δεύτερη συνθήκη μπορεί να ελεγχθεί μόνο εκ των υστέρων, ενώ η πρώτη επαφίεται στη φυσική διαίσθηση που μπορεί να αντληθεί, κατά ένα μέρος, από τον αριθμητικό υπολογισμό τιμών της συνάρτησης. Στο Σχ.(2.4 φαίνεται ο αλγόριθμος που αντιστοιχεί στη μέθοδο Newton-Raphson.

2.4 Μιγαδικές Ρίζες

Τι γίνεται τώρα όταν οι ρίζες είναι μιγαδικές; Καμιά από τις μεθόδους δεν φαίνεται να μας βοηθά. Αλλά η μέθοδος Newton-Raphson γενικεύεται με τον ακόλουθο τρόπο. Γράφοντας:

$$f(x+\epsilon) \approx f(x) + \epsilon f'(x)$$

και αν υποθέσουμε ότι $f(x+\epsilon) \approx 0$, παίρνουμε:

$$\epsilon = -\frac{f(x)}{f'(x)}$$

και μια καινούργια προσέγγιση για τη ρίζα θα είναι:

$$x_{\text{νέο}} = x_{\text{παλιό}} + \epsilon$$

(συγκρίνετε με την Εξ.(2.7)). Αυτή η αλγεβρική αναπαράσταση της μεθόδου επιτρέπει την εύκολη γενίκευσή της για την περίπτωση εύρεσης ριζών συνάρτησης πολλών πραγματικών μεταβλητών. Ένα παράδειγμα θα διαφωτίσει το θέμα.

Ας υποθέσουμε ότι θέλουμε να βρούμε τις (μιγαδικές) ρίζες της εξίσωσης $f(x) = x^2 + x + 1$, η οποία είναι και η χαρακτηριστική εξίσωση της διαφορικής εξίσωσης:

$$m\frac{d^2x}{dt^2} + \Gamma\frac{dx}{dt} + Kx = 0$$

που περιγράφει την κίνηση ενός αρμονικού ταλαντωτή με απόσβεση (για παράδειγμα, κίνηση μιας μάζας συνδεδεμένης με ελατήριο μέσα σε ρευστό για $\Gamma = 1$, $K = 1$ και $m = 1$). Γράφοντας $z = \lambda_R + i\lambda_I$, η εξίσωση $f(z) = 0$ είναι ισοδύναμη με:

$$(\lambda_R^2 - \lambda_I^2 + \lambda_R + 1) + i(2\lambda_R\lambda_I + \lambda_I) = 0$$

Από κατασκευής, $\lambda_R, \lambda_I \in \mathbb{R}$ και γνωρίζουμε πως η τελευταία εξίσωση υποδηλώνει ότι τόσο το πραγματικό όσο και το μιγαδικό τμήμα πρέπει να είναι μηδέν, δηλαδή:

$$f(z) = 0 \Leftrightarrow \begin{cases} f_R(\lambda_R, \lambda_I) = 0 \\ \text{και} \\ f_I(\lambda_R, \lambda_I) = 0 \end{cases}$$

όπου $f_R(\lambda_R, \lambda_I) \equiv \lambda_R^2 - \lambda_I^2 + \lambda_R + 1$ και $f_I(\lambda_R, \lambda_I) \equiv 2\lambda_R\lambda_I + \lambda_I$.

Εφαρμόζοντας λοιπόν τα παραπάνω, αναζητούμε τις *βελτιώσεις* $\delta\lambda_R$ και $\delta\lambda_I$ αναπτύσσοντας:

$$\begin{aligned} f_R(\lambda_R + \delta\lambda_R, \lambda_I + \delta\lambda_I) &= f_R(\lambda_R, \lambda_I) + \delta\lambda_R \left.\frac{\partial f_R}{\partial \lambda_R}\right|_{\lambda_R,\lambda_I} + \delta\lambda_I \left.\frac{\partial f_R}{\partial \lambda_I}\right|_{\lambda_R,\lambda_I} \\ f_I(\lambda_R + \delta\lambda_R, \lambda_I + \delta\lambda_I) &= f_I(\lambda_R, \lambda_I) + \delta\lambda_R \left.\frac{\partial f_I}{\partial \lambda_R}\right|_{\lambda_R,\lambda_I} + \delta\lambda_I \left.\frac{\partial f_I}{\partial \lambda_I}\right|_{\lambda_R,\lambda_I} \end{aligned}$$

Μηδενίζοντας το αριστερό μέλος των παραπάνω εξισώσεων, καταλήγουμε σε ένα γραμμικό σύστημα δυο εξισώσεων με δυο αγνώστους, $\delta\lambda_R$ και $\delta\lambda_I$. Μπορούμε να λύσουμε το σύστημα αυτό εύκολα, γράφοντάς το σε μορφή πινάκων:

$$\begin{pmatrix} -f_R(\lambda_R,\lambda_I) \\ -f_I(\lambda_R,\lambda_I) \end{pmatrix} = \begin{pmatrix} \frac{\partial f_R}{\partial \lambda_R} & \frac{\partial f_R}{\partial \lambda_I} \\ \frac{\partial f_I}{\partial \lambda_R} & \frac{\partial f_I}{\partial \lambda_I} \end{pmatrix} \begin{pmatrix} \delta\lambda_R \\ \delta\lambda_I \end{pmatrix}$$
$$= \begin{pmatrix} 2\lambda_R+1 & -2\lambda_I \\ 2\lambda_I & 2\lambda_R+1 \end{pmatrix} \begin{pmatrix} \delta\lambda_R \\ \delta\lambda_I \end{pmatrix}$$

Είναι εύκολο να υπολογίσουμε τον αντίστροφο ενός 2×2 πίνακα. Ελέγχοντας την αλήθεια της παρακάτω σχέσης:

$$\frac{1}{ad-bc}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

οδηγούμαστε στην εξής λύση για τα $\delta\lambda_R$ και $\delta\lambda_I$:

$$\begin{pmatrix} \delta\lambda_R \\ \delta\lambda_I \end{pmatrix} = \frac{1}{(2\lambda_R+1)^2+4\lambda_I^2}\begin{pmatrix} 2\lambda_R+1 & +2\lambda_I \\ -2\lambda_I & 2\lambda_R+1 \end{pmatrix}\begin{pmatrix} \lambda_I^2-\lambda_R^2-\lambda_R-1 \\ -2\lambda_R\lambda_I-\lambda_I \end{pmatrix}$$

Προσθέτοντας το παραπάνω διάνυσμα στο (λ_R,λ_I) παίρνουμε την πραγματική και φανταστική τιμή της νέας προσέγγισης της λύσης. Συνεχίζουμε με αυτόν τον τρόπο έως ότου προκύψει:

- $\sqrt{(\delta\lambda_R)^2+(\delta\lambda_I)^2} < \epsilon$
- $|f_R(\lambda_R,\lambda_I)| < \epsilon$
- $|f_I(\lambda_R,\lambda_I)| < \epsilon$

Απαιτούνται και οι τρεις συνθήκες. Όπως ακριβώς και για την περίπτωση πραγματικών ριζών, η διόρθωση καθώς και οι τιμές των συναρτήσεων (οι οποίες στο σημείο που ψάχνουμε πρέπει να μηδενίζονται), θα πρέπει να είναι μικρότερες από την ακρίβεια με την οποία εργαζόμαστε.

Πώς όμως αποφασίζουμε για το σημείο εκκίνησης; Βλέπουμε εύκολα ότι η τιμή $\lambda_R^{(0)} = 1/2$ απλοποιεί πολύ τις πράξεις. Για την $\lambda_I^{(0)}$ η τιμή $\pm1/2$ είναι επίσης καλή, εφόσον ο συνδυασμός $2\lambda_I^{(0)}$ εμφανίζεται στα στοιχεία του πίνακα. Συνιστάται πάντως να γίνει κάποιο γράφημα στο επίπεδο (λ_R,λ_I) και να ακολουθείται η κίνηση των σημείων. Τι θα παρατηρήσουμε αν πάρουμε ως αφετηρία άλλα σημεία;

Παρατήρηση 1η. Από γεωμετρική άποψη, οι ρίζες είναι τα σημεία τομής των καμπυλών:

$$f_R(\lambda_R,\lambda_I)=0, \quad f_I(\lambda_R,\lambda_I)=0$$

Παρατήρηση 2η. Παρατηρούμε ότι η εύρεση των μιγαδικών λύσεων είναι μια ειδική περίπτωση της λύσης ενός μη γραμμικού συστήματος εξισώσεων. Ως παράδειγμα, μπορούμε να αναζητήσουμε τα σημεία ισορροπίας ενός

υλικού σημείου που κινείται στο επίπεδο υπό την επίδραση ενός δυναμικού της μορφής:

$$V[x,y] = \frac{1}{2}K(x^2+y^2) - a(x^2+y^2)^2 - F_x \cdot x - F_y \cdot y \qquad (2.10)$$

Συμπέρασμα

Στο κεφάλαιο αυτό αναφερθήκαμε στις μεθόδους αριθμητικού υπολογισμού των λύσεων μιας εξίσωσης. Κάθε μια από αυτές έχει τα πλεονεκτήματά της - τα οποία πρέπει να εκμεταλλευόμαστε στο έπακρο - όπως και τα μειο- νεκτήματά της - τα οποία πρέπει να λαμβάνουμε σοβαρά υπόψη - ώστε τα αποτελέσματα να έχουν σημασία. Η μέθοδος της διχοτόμησης είναι σίγουρη, αλλά αργή. Η Newton-Raphson είναι ταχύτερη, αλλά επικίνδυνη αν δεν γνωρίζουμε τη συμπεριφορά της συνάρτησης. Επίσης δεν είναι ανταγωνιστική αν δεν μπορούμε να υπολογίσουμε την παράγωγο της συνάρτησης με την ίδια ακρίβεια που υπολογίζουμε την ίδια τη συνάρτηση. Από την άλλη, είναι η μόνη μέθοδος που μπορεί να γενικευτεί για την περίπτωση μιγαδικών λύσεων ή των λύσεων ενός μη γραμμικού συστήματος εξισώσεων.

Κεφάλαιο 3

Αριθμητική Ολοκλήρωση

Μια ενδιαφέρουσα διαφορά μεταξύ της παραγώγισης και της ολοκλήρωσης είναι ότι υπάρχουν ολοκληρώματα, τα οποία δεν μπορούν να υπολογιστούν αναλυτικά, ενώ αντιθέτως δεν υπάρχει θεωρητικό εμπόδιο στον υπολογισμό της παραγώγου μιας συνάρτησης (ή στο συμπέρασμα ότι δεν υπάρχει η παράγωγος).

Επιπλέον, όπως ήδη αναφέραμε, η αναλυτική μορφή δεν προσαρμόζεται πάντοτε αρμονικά με αριθμητικούς υπολογισμούς.

Το απλούστερο φυσικό παράδειγμα είναι πάλι η κίνηση ενός σωματιδίου μέσα σ' ένα δυναμικό, σε μια διάσταση. Όπως είδαμε στην προηγούμενη παράγραφο, όταν γνωρίζουμε την ενέργεια διαθέτουμε όλες τις αναγκαίες πληροφορίες για τον υπολογισμό των ορίων κίνησης. Εδώ θα ασχοληθούμε με το δυναμικό πρόβλημα του υπολογισμού του χρονικού διαστήματος που χρειάζεται το σωματίδιο για να μετακινηθεί από ένα σημείο σ' ένα άλλο.

Η λύση ενός τέτοιου σημαντικού προβλήματος στη φυσική, έγκειται στην ολοκλήρωση των διαφορικών εξισώσεων της κίνησης. Οι αριθμητικές μέθοδοι για το σκοπό αυτό αποτελούν το αντικείμενο του κεφαλαίου αυτού. Σε μία διάσταση, με διατήρηση της ενέργειας, καταλήγουμε εύκολα σε μια συνήθη διαφορική εξίσωση πρώτης τάξης. Η ολοκλήρωση αυτής της διαφορικής εξίσωσης μπορεί να γίνει αναλυτικά μόνο σε πολύ λίγες περιπτώσεις, οπότε η ανάγκη μιας αριθμητικής προσέγγισης είναι εμφανής.

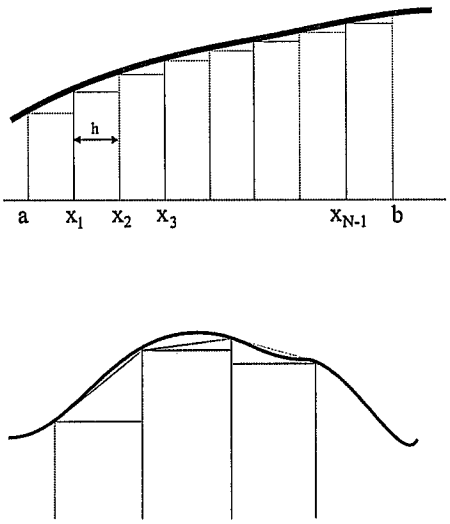
Υπενθυμίζουμε ότι η διατήρηση της ενέργειας, σε μια διάσταση, μπορεί να περιγραφεί από την εξίσωση:

$$E = \frac{1}{2}m\left(\frac{dx}{dt}\right)^2 + V(x) = \text{σταθερή} \tag{3.1}$$

Η εξίσωσης αυτή οδηγεί στην εξίσωση του Newton:

$$m\frac{d^2x}{dt^2} = -\frac{dV}{dx} \tag{3.2}$$

δηλαδή, σε μια συνήθη διαφορική εξίσωση δευτέρας τάξης, όπου βέβαια θα πρέπει να καθορίσουμε την αρχική θέση και την αρχική ταχύτητα. Αλλά πολύ



Σχήμα 3.1: Αριθμητική ολοκλήρωση

εύκολα διακρίνουμε ότι μπορούμε να λύσουμε την εξίσωση αυτή ως προς την ταχύτητα, οπότε η διαφορική εξίσωση είναι πρώτης τάξης:

$$\frac{dx}{dt} = \sqrt{\frac{2}{m}(E - V(x))} \tag{3.3}$$

με αρχική συνθήκη $x(0) = x_0$. Αυτή η εξίσωση ολοκληρώνεται άμεσα:

$$t = \int_{x_0}^{x} \frac{dx'}{\sqrt{\frac{2}{m}(E - V[x'])}} \tag{3.4}$$

η οποία, αν αναστραφεί, μας δίνει την τροχιά $x(t)$ του σωματιδίου.

3.1 Η Μέθοδοι των Ορθογωνίων και των Τραπεζίων

Για τον υπολογισμό του ολοκληρώματος, μπορούμε να ξεκινήσουμε βέβαια από τον μαθηματικό ορισμό του:

$$\int_a^b f(x)dx \approx \sum_{i=0}^{N+1} f(x_i)\delta x_i \tag{3.5}$$

(βλέπε Σχ.(3.1)). Το πρόβλημα είναι ότι η σύγκλιση αυτής της σειράς πρός το ολοκλήρωμα είναι πολύ αργή και επίσης το σφάλμα (ως προς την θεωρητική τιμή που ορίζεται βέβαια όταν $N \to \infty$), είναι σημαντικό.

Μια καλύτερη προσέγγιση έχουμε αν χρησιμοποιήσουμε, αντί των ορθογωνίων παραλληλογράμμων, τα τραπέζια:

$$\int_a^b f(x)dx \approx h\left(\frac{1}{2}f(a) + f(x_1) + ... + f(x_N) + \frac{1}{2}f(b)\right) \qquad (3.6)$$

Παράδειγμα:
Ως ένα απλό παράδειγμα ας πάρουμε την $f(x) = Ax + B$. Χρησιμοποιώντας τον ορισμό του ολοκληρώματος:

$$\int_a^b f(x)dx \equiv \lim_{N\to\infty} h\sum_{k=0}^{N-1} f(x_k)$$

όπου $x_k = a + kh$ και $h = (b-a)/N$. Για πεπερασμένο N βρίσκουμε:

$$h\sum_{k=0}^{N-1}(Ax_k + B) = h\left((B + Aa)N + Ah\sum_{k=0}^{N-1} k\right) =$$
$$= B(b-a) + (A/2)(b^2 - a^2) - (A/2N)(b-a)^2$$

Γνωρίζοντας το ακριβές αποτέλεσμα $(A(b^2 - a^2)/2 + B(b-a))$, βλέπουμε ότι το σφάλμα είναι της τάξης $\mathcal{O}(1/N)$. Αν χρησιμοποιήσουμε τα τραπέζια βρίσκουμε:

$$h\left[\frac{1}{2}(Aa + B) + Ax_1 + B + Ax_2 + B + ... + Ax_{N-1} + B + \frac{1}{2}(Ab + B)\right] =$$
$$= h\left[(A/2)(a+b) + NB + Aa(N-1) + Ah\sum_{k=0}^{N-1} k\right] =$$
$$= \frac{b-a}{N}\left[(A/2)(b+a)N + BN\right]$$

δηλαδή βλέπουμε ότι έχουμε το ακριβές αποτέλεσμα, ακόμα και για N πεπερασμένο! Στην πρώτη περίπτωση (με τα ορθογώνια) η συνάρτηση θεωρείται σταθερή σε κάθε διάστημα, ενώ στη δεύτερη περίπτωση (με τα τραπέζια) η συνάρτηση θεωρείται γραμμική σε κάθε διάστημα. Η συνάρτηση του παραδείγματός μας είναι πράγματι γραμμική. Οπότε, στην πρώτη περίπτωση το σφάλμα είναι ανάλογο του συντελεστή του γραμμικού όρου, A, ενώ στη δεύτερη το σφάλμα είναι μηδέν.

3.2 Η Μέθοδος Simpson

Μπορούμε να συνεχίσουμε με την ίδια φιλοσοφία και να φτάσουμε στον *τύπο του Simpson* που θεωρεί ότι η συνάρτηση είναι τετραγωνική σε κάθε διάστημα, δηλαδή ότι μπορεί να προσεγγιστεί από ένα πολυώνυμο δεύτερου βαθμού. Αλλά ας δούμε πιο συγκεκριμένα αυτή την περίπτωση.

Ας υποθέσουμε ότι έχουμε το διάστημα $[a, b]$ και το μέσον του $m \equiv (a+b)/2$. Υπάρχει ένα, και μόνο ένα, πολυώνυμο δεύτερου βαθμού,

το $P(x)$, το οποίο παίρνει τις τιμές $f(a)$, $f(b)$ και $f(m)$ στις θέσεις a, b και m αντίστοιχα. Για να βρούμε το πολυώνυμο αυτό, θέτουμε $P(x) = Ax^2+Bx+C$ και απαιτούμε $P(a) = f(a)$, $P(b) = f(b)$ και $P(m) = f(m)$. Με αυτό τον τρόπο, παίρνουμε ένα γραμμικό σύστημα τριων εξισώσεων με τρεις αγνώστους, τους συντελεστές A, B και C. Γι' αυτό το πολυώνυμο έχουμε:

$$\int_a^b P(x)dx = h\left(\frac{1}{3}P(a) + \frac{4}{3}P(m) + \frac{1}{3}P(b)\right) \tag{3.7}$$

όπου $h \equiv (b-a)/2$. Η απόδειξη είναι εύκολη: Αντικαταστήστε το $P(x)$ με $Ax^2 + Bx + C$ και θα καταλήξετε σε ταυτότητα. Μια πιο ενδιαφέρουσα απόδειξη είναι η ακόλουθη:
Ξέρουμε ότι

- Κάθε πολυώνυμο $P(x)$ δευτέρου βαθμού μπορεί να γραφεί ως γραμμικός συνδυασμός των μονωνύμων x^2, x και 1.
- Η ολοκλήρωση είναι μια γραμμική πράξη:

$$\int_a^b (c_1 f(x) + c_2 g(x))dx = c_1 \int_a^b f(x)dx + c_2 \int_a^b g(x)dx \tag{3.8}$$

- Αρκεί να μπορούμε να ολοκληρώσουμε στο διάστημα $[0, 1]$:

$$\begin{aligned} \int_a^b f(x)dx &= \int_0^{b-a} f(x+a)\, d(x+a) \\ &= (b-a)\int_0^1 f(\frac{x+a}{b-a}(b-a))\, d\left(\frac{x+a}{b-a}\right) \end{aligned} \tag{3.9}$$

Επομένως, μπορούμε να γράψουμε:

$$\begin{aligned} \int_0^1 x^2 dx &= \frac{1}{2}\left(\alpha\cdot 0^2 + \beta\cdot\left(\frac{1}{2}\right)^2 + \gamma\cdot 1^2\right) \\ \int_0^1 x dx &= \frac{1}{2}\left(\alpha\cdot 0^1 + \beta\cdot\left(\frac{1}{2}\right)^1 + \gamma\cdot 1^1\right) \\ \int_0^1 dx &= \frac{1}{2}(\alpha\cdot 1 + \beta\cdot 1 + \gamma\cdot 1) \end{aligned} \tag{3.10}$$

και είναι εύκολο να λυθεί το σύστημα για τους τρεις αγνώστους που δίνει $\alpha = 1/3$, $\beta = 4/3$ και $\gamma = 1/3$. Η θεμελιώδης ιδιότητα αυτών των συντελεστών είναι η ανεξαρτησία τους από το συγκεκριμένο δευτεροβάθμιο πολυώνυμο που θέλουμε να ολοκληρώσουμε, αλλά και από το διάστημα $[a, b]$ ολοκλήρωσης που μας ενδιαφέρει. Η όλη διαδικασία εξαρτάται μόνο από την εκλογή μας να προσεγγίσουμε την υπό ολοκλήρωση συνάρτηση με πολυώνυμο (το πολύ) δευτέρου βαθμού.

Συμπέρασμα

Για τον αριθμητικό υπολογισμό ενός ολοκληρώματος μιας συνάρτησης με τη

μέθοδο του Simpson, αντικαθιστούμε τη συνάρτηση με ένα πολυώνυμο δευτέρου βαθμού το οποίο στά άκρα και στο μέσον του διαστήματος ολοκλήρωσης παίρνει τις τιμές που έχει η συνάρτηση για τα ίδια σημεία, και ολοκληρώνουμε αυτό το συγκεκριμένο πολυώνυμο ακριβώς.

Παράδειγμα:

$$I = \int_0^{\pi/2} \sin x \, dx$$

Θα υπολογίσουμε το ολοκλήρωμα αυτό με τρεις διαφορετικές μεθόδους:
α) αναλυτικά, β) υπολογίζοντας ρητά το πολυώνυμο $P(x)$ και ολοκληρώνοντας το πολυώνυμο και, γ) εφαρμόζοντας τον κανόνα του Simpson.
α) Το ολοκλήρωμα υπολογίζεται εύκολα: $I = -\cos(\pi/2) + \cos(0) = 1$
β) Γράφουμε το πολυώνυμο ως $P(x) = Ax^2 + Bx + C$. Τα A, B και C θα είναι οι λύσεις του συστήματος:

$$\begin{aligned} C = f(0) &= 0 \\ A\pi^2/16 + B\pi/4 + C &= 1/\sqrt{2} \\ A\pi^2/4 + B\pi/2 + C &= 1 \end{aligned}$$

Επιλύοντας το σύστημα βρίσκουμε $A = -(8/\pi^2)(\sqrt{2}-1)$ και $B = (8/\pi)(1/\sqrt{2}-1/4)$. Αν αντικαταστήσουμε στο ολοκλήρωμα I το $\sin x$ με το πολυώνυμο $P(x)$ και ολοκληρώσουμε, βρίσκουμε:

$$I_{\text{πολ}} = \int_0^{\pi/2} P(x)\, dx = A\frac{\pi^3}{24} + B\frac{\pi^2}{8} = \frac{\pi}{4}\frac{2\sqrt{2}+1}{3} = 1.00228$$

γ) Υπολογίζουμε το I με τον κανόνα του Simpson:

$$\begin{aligned} I_{Simpson} &= \frac{1}{2}\frac{\pi}{2}\left[\frac{1}{3}\sin(0) + \frac{4}{3}\sin(\pi/4) + \frac{1}{3}\sin(\pi/2)\right] \\ &= \frac{\pi}{4}\left[\frac{4}{3\sqrt{2}} + \frac{1}{3}\right] = \frac{\pi}{4}\frac{2\sqrt{2}+1}{3} = 1.00228 \end{aligned}$$

βρίσκουμε δηλαδή το ίδιο αποτέλεσμα. Τι θα συνέβαινε αν, αντί να παίρναμε το πολυώνυμο που περνά από τα σημεία $(0,0)$, $(\pi/4, 1/\sqrt{2})$ και $(\pi/2, 1)$, θεωρούσαμε ως πολυώνυμο το ανάπτυγμα του $\sin x$ γύρω από το σημείο 0 ως δεύτερη τάξη, δηλαδή $\sin x = x + \mathcal{O}(x^3)$; Τότε θα βρίσκαμε:

$$I_{Taylor} = \int_0^{\pi/2} x\, dx = \frac{\pi^2}{8} = 1.2337$$

Αυτό το αποτέλεσμα είναι βέβαια μικρότερης ακρίβειας από το προηγούμενο. Εδώ όμως, γνωρίζουμε το ακριβές αποτέλεσμα. Μπορούμε να συμπεράνουμε ότι το πολυώνυμο $P(x)$ θα δίνει πάντοτε καλύτερο αποτέλεσμα από το ανάπτυγμα Taylor τάξης 2; Διαισθητικά θα απαντούσαμε θετικά, διότι το $P(x)$ 'περνά' όχι μόνο από το σημείο $x = 0$ (από το οποίο περνά και το ανάπτυγμα κατά Taylor) αλλά και από τα σημεία $x = m$ και $x = b$. Όμως, θα μπορούσαμε

να προβλέψουμε ότι δεν αποκλείεται το σφάλμα που υπεισέρχεται από τη προσέγγιση του πολυωνύμου να αντισταθμίζει την ικανότητά του να περνά από τα σημεία αυτά και το τελικό αποτέλεσμα να είναι χειρότερο. Το μόνο που μπορούμε να πούμε με σιγουριά είναι ότι, αν έχουμε να ολοκληρώσουμε μια συνάρτηση 2ου βαθμού, τόσο το $P(x)$ όσο και και η ανάπτυξη κατά Taylor θα δώσουν τα ίδια αποτελέσματα.

Παράδειγμα:

Ας δοκιμάσουμε τώρα τη συνάρτηση $f(x) = \cos x$ για την οποία ξέρουμε βέβαια ότι η ανάπτυξή της γύρω από το $x = 0$ είναι $\cos x = 1 - x^2/2 + \mathcal{O}(x^4)$.

α) Το ακριβές αναλυτικό αποτέλεσμα είναι:

$$I = \int_0^{\pi/2} \cos x \, dx = \sin(\pi/2) = 1$$

β) Το πολυώνυμο $P(x) = Ax^2 + Bx + C$ δίνει για συντελεστές $A = (8/\pi^2)(\sqrt{2} - 1)$ $B = (2/\pi)(4\sqrt{2} - 3)$ και $C = 1$. Ολοκληρώνοντας παίρνουμε:

$$I_{\pi o\lambda} = \int_0^{\pi/2} P(x)dx = A(\pi^3/24) + B(\pi^2/8) + C(\pi/2) = (\pi/12)(1 + 2\sqrt{2}) = 1.00228 \tag{3.11}$$

γ) Η μέθοδος Simpson δίνει άμεσα:

$$I_{simpson} = \frac{\pi}{4}\left[\frac{1}{3} + \frac{1}{4}\frac{1}{\sqrt{2}}\right] = 1.00228$$

δ) Τέλος, αν αντικαταστήσουμε τη συνάρτηση με το ανάπτυγμά της γύρω από το $x = 0$ βρίσκουμε:

$$I_{Taylor} = \int_0^{\pi/2} \left(1 - \frac{x^2}{2}\right) dx = (\pi/2)\left(1 - \frac{\pi^2}{24}\right) = 0.9248$$

Το παράδειγμα αυτό δείχνει ότι, ακόμα και για μια συνάρτηση όπως το $\cos x$, το πολυώνυμο $P(x)$ δίνει καλύτερο αποτέλεσμα από ό,τι το ανάπτυγμα κατά Taylor. Βέβαια αυτό δεν αποτελεί απόδειξη, αλλά δείχνει ότι τα αντι-παραδείγματα δεν θα πρέπει να είναι τετριμμένα.

Μπορούμε να γενικεύσουμε τον κανόνα για N ενδιάμεσα σημεία (αντί ενός που είχαμε έως τώρα, αλλά προσοχή, γιατί το N πρέπει να είναι περιττός ώστε να ισχύει ο τύπος), εφαρμόζοντας τη μέθοδο σε κάθε ενδιάμεσο διάστημα:

$$\int_a^b f(x)dx = h\left(\frac{1}{3}f(a) + \frac{4}{3}(f(x_1) + f(x_3) + ... + f(x_N)) + \frac{2}{3}(f(x_2) + f(x_4) + ... + f(x_{N-1})) + \frac{1}{3}f(b)\right) \tag{3.12}$$

όπου τώρα $h = (b - a)/(N + 1)$

Είναι ευνόητο ότι μπορούμε να συνεχίσουμε με αυτήν την πορεία και να παράγουμε ανάλογους τύπους που είναι ακριβείς για πολυώνυμα τάξης ανώτερης του 2.

Τέλος, θα πρέπει να πούμε ότι η χρήση της μεθόδου Simpson μπορεί να γίνει όχι μόνο στο αρχικό ολοκλήρωμα αλλά και σ' αυτό που προκύπτει από έναν κατάλληλο μετασχηματισμό μεταβλητής. Το τελευταίο μπορεί να υπολογίζεται ακριβώς με τη μέθοδο ή να παρουσιάζει μικρότερο σφάλμα από την εφαρμογή της μεθόδου στο αρχικό ολοκλήρωμα. Μερικά παραδείγματα θα διευκρινίσουν το σημείο αυτό.

Παράδειγμα.

$$I = \int_0^{\pi/2} \sin 3x \, dx$$

Η υπό ολοκλήρωση συνάρτηση δεν είναι πολυώνυμο του x οπότε ούτε η μέθοδος του τραπεζίου ούτε αυτή του Simpson εφαρμόζονται. Ωστόσο, μπορούμε να βρούμε μια αλλαγή μεταβλητών η οποία θα μας δώσει ένα ολοκλήρωμα τέτοιας μορφής ώστε τουλάχιστον μια από τις μεθόδους αυτές να δίνει ακριβές αποτέλεσμα.

α) $\sin 3x \, dx = -\frac{1}{3} d(\cos 3x)$. Οπότε, θέτοντας $u = \cos 3x$ παίρνουμε:

$$I = \frac{1}{3} \int_1^0 (-du) = \frac{1}{3} \int_0^1 du = \frac{1}{3}$$

Η σταθερή συνάρτηση μπορεί να ολοκληρωθεί δίνοντας ακριβές αποτέλεσμα και με τις δύο μεθόδους (τραπεζίου ή Simpson):

$$I = 1 \times \left[\frac{1}{2} + \frac{1}{2}\right] = \frac{1}{2} \times \left[\frac{1}{3} + \frac{4}{3} + \frac{1}{3}\right]$$

β) $\sin 3x = \sin(2x + x) = \sin 2x \cos x + \cos 2x \sin x = \sin x(4\cos^2 x - 1)$

Το ολοκλήρωμα γράφεται, επομένως:

$$\begin{aligned} I = \int_0^{\pi/2} \sin 3x \, dx &= \int_0^{\pi/2} -(4\cos^2 x - 1) \, d\cos x \\ &= \int_1^0 -(4u^2 - 1) \, du = \int_0^1 (4u^2 - 1) \, du \end{aligned}$$

όπου κάναμε αλλαγή την μεταβλητής $u = \cos x$. Τώρα το ολοκλήρωμα είναι ένα πολυώνυμο 2ου βαθμού και η μέθοδος του Simpson θα δώσει ακριβές αποτέλεσμα:

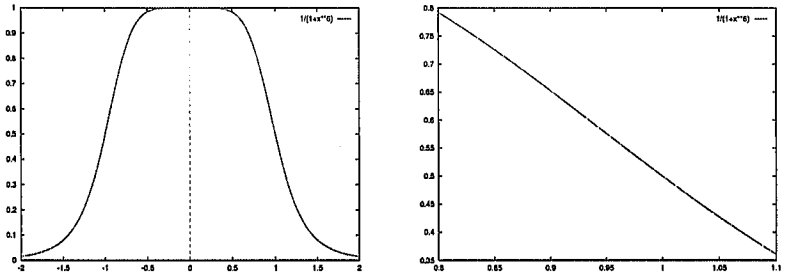
$$I = \frac{1}{2}\left[\frac{1}{3}(-1) + \frac{4}{3}(0) + \frac{1}{3}(3)\right] = \frac{1}{3}$$

Παράδειγμα.

$$I = \int_{100}^{100+a} \frac{dx}{1 + x^6}$$

με $0 < a << 100$. Εδώ μπορούμε να εφαρμόσουμε τη μέθοδο του τραπεζίου

$$\frac{1}{1 + 100^6} - \frac{1}{1 + (100 + a)^6} = \frac{6}{1 + 100^6} \frac{a}{100} + \mathcal{O}((\frac{a}{100})^2)$$



Σχήμα 3.2: Η καμπύλη $f(x) = 1/(x^6 + 1)$. Στο σχήμα δεξιά βλέπουμε μια λεπτομέρεια της καμπύλης.

όπου χρησιμοποιήσαμε:

$$(1+y)^n - 1 = ny + \mathcal{O}(y^2), \quad \frac{1}{1+y} = 1 - y + \mathcal{O}(y^2)$$

Τα παραπάνω μας δείχνουν ότι η συνάρτηση $f(x) = 1/(1 + x^6)$ μπορεί να προσεγγιστεί από μια γραμμική συνάρτηση αν $x >> 1$ και το εύρος του διαστήματος ολοκλήρωσης είναι $<< x$ (δηλαδή το εύρος του διαστήματος μπορεί να είναι μεγαλύτερο του 1).

Αντίθετα, πολύ κοντά στο 0, τα πράγματα αλλάζουν. Εκεί η συνάρτηση έχει μέγιστο, $f(x) = 1 - x^6 + \mathcal{O}(x^{12})$, το οποίο δείχνει ότι κοντά στο 0 η συνάρτηση είναι περίπου σταθερή οπότε η μέθοδος του τραπεζίου ή του Simpson μπορούν να εφαρμοστούν με επιτυχία. Για μεγαλύτερες τιμές του x αρχίζουμε να 'αισθανόμαστε' την καμπύλωση. Περνάμε από ένα σημείο καμπής και στην περιοχή περίπου $[0.8, 1]$ η συνάρτηση μπορεί να προσεγγιστεί από μια γραμμική συνάρτηση. Για μεγαλύτερες τιμές του x αισθανόμαστε εκ νέου μια καμπύλωση και τέλος η καμπύλη πλατύνεται, όπως δείχνει το Σχ.(3.2).

Όσον αφορά την προσαρμογή αντίστοιχων αλγορίθμων τα πράγματα είναι απλά. Η μέθοδος του τραπεζίου απαιτεί τον υπολογισμό ενός αθροίσματος σ' όλα τα ενδιάμεσα σημεία με ίδιο συντελεστή (δηλαδή 1), ενώ η μέθοδος Simpson χρησιμοποιεί δύο συντελεστές (έναν για τα άρτια σημεία και άλλον για τα περιττά). Το πρόβλημα της ποιότητας της προσέγγισης δεν είναι καθόλου τετριμμένο, αλλά για μια διάσταση μπορούμε να το ελέγξουμε ικανοποιητικά. Επίσης, μπορούμε να πάρουμε ίδιο βήμα ολοκλήρωσης h για όλη την περιοχή ολοκλήρωσης. Για παραπάνω από μία διάσταση, χάνουμε αυτήν την πολυτέλεια και θα πρέπει να έχουμε πληροφορίες για τα διαστήματα όπου η συνάρτηση μεταβάλλεται γρήγορα ή να ακολουθήσουμε άλλες τεχνικές, όπως η *πολυπλεγματική προσέγγιση ή μέθοδος Monte Carlo.*

Τέλος, υπάρχει το θέμα των *ιδιαζόντων σημείων.* Υπάρχουν δύο τύποι: τα ολοκληρώσιμα και τα μη ολοκληρώσιμα. Για τη δεύτερη περίπτωση δεν υπάρχει κανένα αριθμητικό τέχνασμα. Όσον αφορά τον πρώτο τύπο, πάντοτε υπάρχει μια κατάλληλη αλλαγή μεταβλητής για τον υπολογισμό του ολοκληρώματος.

Ένα κλασικό παράδειγμα είναι ο υπολογισμός της περιόδου κίνησης ενός υλικού σημείου σ' ένα δυναμικό:

$$T \equiv 2t(x_1 \to x_2) = 2\sqrt{\frac{m}{2}} \int_{x_1}^{x_2} \frac{dx}{\sqrt{E - V(x)}}$$

όπου $E = V(x_1) = V(x_2)$. Βλέπουμε αμέσως ότι δεν μπορούμε να χρησιμοποιήσουμε τη μέθοδο Simpson (ούτε αυτή του τραπεζίου) επειδή το ολοκλήρωμα απειρίζεται στα σημεία x_1 και x_2. Ευτυχώς, γνωρίζουμε ότι αν τα σημεία αυτά δεν είναι ακρότατα του δυναμικού $V(x)$, το ολοκλήρωμα πρέπει *να συγκλίνει*, η περίοδος είναι πεπερασμένη. Αντίθετα, αρκεί ένα από τα σημεία να αποτελεί ακρότατο του $V(x)$ ώστε να κάνει το ολοκλήρωμα να αποκλίνει. Πιο συγκεκριμένα, αν το δυναμικό έχει πολυωνυμική μορφή μπορούμε να ακολουθήσουμε την παρακάτω διαδικασία:

$$E - V(x) \equiv \mathcal{P}(x) = (x - x_1)P(x) = (x_2 - x)Q(x)$$

όπου υποθέτουμε ότι $P(x_1) \neq 0$ και $Q(x_2) \neq 0$. Αν y είναι ένα σημείο του διαστήματος (x_1, x_2) με την ιδιότητα $P(y) \neq 0$ και $Q(y) \neq 0$, μπορούμε να γράψουμε:

$$T = 2\sqrt{\frac{m}{2}} \left(\int_{x_1}^{y} \frac{dx}{\sqrt{\mathcal{P}(x)}} + \int_{y}^{x_2} \frac{dx}{\sqrt{\mathcal{P}(x)}} \right) \tag{3.13}$$

Χρησιμοποιούμε τώρα το ανάπτυγμα $\mathcal{P}(x) = (x - x_1)P(x)$ στο πρώτο διάστημα $[x_1, y]$ και το $\mathcal{P}(x) = (x_2 - x)Q(x)$ για το δεύτερο $[y, x_2]$:

$$\int_{x_1}^{y} \frac{dx}{\sqrt{(x - x_1)P(x)}} = \int_{x_1}^{y} \frac{d[2\sqrt{x - x_1}]}{\sqrt{P(x)}} = \int_{0}^{2\sqrt{y - x_1}} \frac{du}{\sqrt{P(x_1 + \frac{u^2}{4})}} \tag{3.14}$$

και το τελευταίο ολοκλήρωμα δεν παρουσιάζει κανένα πρόβλημα. Με τον ίδιο τρόπο βρίσκουμε:

$$\begin{aligned} \int_{y}^{x_2} \frac{dx}{\sqrt{(x_2 - x)Q(x)}} &= \int_{y}^{x_2} \frac{d[-2\sqrt{x_2 - x}]}{\sqrt{Q(x)}} = \\ &= \int_{-2\sqrt{x_2 - y}}^{0} \frac{dv}{\sqrt{Q(x_2 - \frac{v^2}{4})}} \end{aligned} \tag{3.15}$$

Ένα ενδιαφέρον σημείο είναι ότι, αν μπορούμε να υπολογίσουμε αναλυτικά τα δύο παραπάνω ολοκληρώματα, το τελικό αποτέλεσμα είναι ανεξάρτητο από την εκλογή του σημείου y. Αντίθετα, αν χρησιμοποιήσουμε μια προσεγγιστική μέθοδο, όπως για παράδειγμα του τραπεζίου ή τη Simpson, το αποτέλεσμα, γενικά, εξαρτάται από το σημείο y. Οπότε μπαίνει το ερώτημα αν μπορούμε να διαλέξουμε το σημείο y με τέτοιο τρόπο ώστε να ανακτήσουμε ορισμένες ιδιότητες που εκ των προτέρων γνωρίζουμε ότι υπάρχουν. Το επόμενο παράδειγμα θα διαλευκάνει το ζήτημα αυτό.

Παράδειγμα:

Ας υποθέσουμε ότι ένα σωμάτιο με μάζα m κινείται σε δυναμικό $V(x) = kx^2$ με ολική ενέργεια E. Θέλουμε να υπολογίσουμε την περίοδο της κίνησής του με τη μέθοδο του τραπέζιου.

Δείξτε ότι αν υπολογίσουμε την περίοδο αναλυτικά, το αποτέλεσμα δεν εξαρτάται από τη συνολική ενέργεια ούτε από το σημείο που θα διαλέξουμε να κόψουμε το ολοκλήρωμα.

Δείξτε επίσης ότι αν χρησιμοποιήσουμε τη μέθοδο του τραπεζίου, το αποτέλεσμα εξαρτάται από την ενέργεια καθώς και από το σημείο y, αλλά μπορούμε πάντοτε να διαλέξουμε το σημείο αυτό έτσι ώστε το αποτέλεσμα να μην εξαρτάται από την ενέργεια.

Συμπέρασμα

Η αριθμητική ολοκλήρωση περιέχει δύο ελεύθερες παραμέτρους: τον αριθμό των σημείων στα οποία υπολογίζουμε την υπό ολοκλήρωση συνάρτηση καθώς και το βάρος που βάζουμε σε κάθε σημείο. Στην απλή περίπτωση που εμπνέεται από τον κλασικό ορισμό της ολοκλήρωσης, η κατανομή των ενδιαμέσων σημείων είναι ομοιόμορφη και τα αντίστοιχα βάρη είναι 1. Στις μεθόδους του τραπεζίου και του Simpson η κατανομή παραμένει ομοιόμορφη αλλά το βάρος κάθε σημείου διαφέρει. Μπορεί κάποιος να αλλάξει επίσης την κατανομή των ενδιαμέσων σημείων, οπότε έχουμε και τις πιο εξελιγμένες αριθμητικές μεθόδους ολοκλήρωσης.

Στα συγκεκριμένα προβλήματα, η εκλογή των ενδιαμέσων σημείων δεν είναι πάντοτε αθώα. Θα πρέπει να γίνεται με τέτοιο τρόπο, ώστε το τελικό αποτέλεσμα να διατηρεί τις ποιοτικές ιδιότητες του 'ακριβούς αποτελέσματος' που ψάχνουμε.

Κεφάλαιο 4

Αριθμητική Λύση Συνήθων Διαφορικών Εξισώσεων

Οι διαφορικές εξισώσεις παίζουν, από την εποχή του Newton, σημαντικό ρόλο στην περιγραφή της φύσης. Αλλά η διατύπωση μιας διαφορικής εξίσωσης από τη μια και η λύση της από την άλλη είναι δύο τελείως διαφορετικές ασκήσεις. Για πολλά χρόνια δεν είχαμε στη διάθεσή μας παρά μόνο αναλυτικά εργαλεία, γεγονός που περιόριζε σημαντικά το ερευνητικό πεδίο. Η ανακάλυψη των αριθμητικών μεθόδων απελευθέρωσε τη φυσική, τη χημεία, τη βιολογία κλπ., από τον περιορισμό της ανάπτυξης απλοποιημένων μοντέλων (προτύπων) στην προσπάθειά τους να περιγράψουν την πολυπλοκότητα των φυσικών φαινομένων. Σκοπός αυτού του κεφαλαίου είναι η περιγραφή των ευρέως χρησιμοποιουμένων σήμερα μεθόδων για τη λύση συνήθων διαφορικών εξισώσεων που συναντάμε σε εφαρμογές της φυσικής (ή της χημείας ή της βιολογίας, όπου βρίσκουμε τις ίδιες εξισώσεις σ' ένα τελείως διαφορετικό πλαίσιο).

Θα παρουσιάσουμε τις μεθόδους αυτές σ' ένα πλαίσιο γενικό. Ας υποθέσουμε ότι έχουμε την εξαρτημένη μεταβλητή x και t είναι η ανεξάρτητη μεταβλητή. Ψάχνουμε να βρούμε λύση της εξίσωσης:

$$\frac{dx}{dt} = F(x(t), t) \tag{4.1}$$

με 'αρχικές' συνθήκες $x(0) = x_0$. Αν οι εξισώσεις είναι βαθμού ανώτερου του πρώτου (όπως είναι οι εξισώσεις της κίνησης στη Μηχανική), οδηγούμαστε σε *σύστημα* εξισώσεων, οπότε χρειαζόμαστε τη γραμμική άλγεβρα, κυρίως αν είμαστε σε χώρο με διαστάσεις μεγαλύτερες του 1. Για μία διάσταση δεν έχουμε ανάγκη τέτοιων εργαλείων. Για παράδειγμα, για την εξίσωση του

Νεύτωνα $md^2x/dt^2 = f(x(t),t)$, απλά γράφουμε

$$\begin{aligned} \frac{dx}{dt} &= v \\ \frac{dv}{dt} &= \frac{1}{m} f\,(x(t),t) \end{aligned} \tag{4.2}$$

με αρχικές συνθήκες $x(0) = x_0$ και $v(0) = v_0$.

Η κεντρική ιδέα πίσω από τις αριθμητικές μεθόδους είναι η διακριτοποίηση του χρόνου (της ανεξάρτητης μεταβλητής) και η αντικατάσταση των παραγώγων από πεπερασμένες διαφορές:

$$\frac{dx}{dt} \to \frac{\Delta x}{\Delta t} \equiv \frac{x(t+h) - x(t)}{h} \tag{4.3}$$

4.1 Οι Μέθοδοι του Euler και του Ενδιάμεσου Σημείου

Η παραπάνω διαδικασία οδηγεί αβίαστα στη *μέθοδο του Euler* που για μεν την Εξ.(4.1) γράφεται:

$$x(t+h) = x(t) + hF(x(t),t) \tag{4.4}$$

ενώ για τις Εξ.(4.2) έχει τη μορφή:

$$\begin{aligned} x(t+h) &= x(t) + hv(t) \\ v(t+h) &= v(t) + hf(x(t),t) \end{aligned} \tag{4.5}$$

Παρατηρούμε αμέσως ότι αυτές οι εξισώσεις δεν αντιστοιχούν σε τίποτα άλλο παρά στον πρώτο όρο αναπτύγματος, ως προς το χρόνο, γύρω από το $t = 0$. Δηλαδή η ακρίβεια της μεθόδου είναι $\mathcal{O}(h^2)$. Μπορούμε λοιπόν να ρωτήσουμε πώς είναι δυνατόν να αυξήσουμε την ακρίβεια, ώστε το σφάλμα να είναι $\mathcal{O}(h^3)$ τουλάχιστον.

Αυτό επιτυγχάνεται ως ακολούθως:

$$\begin{aligned} x(t+h) &= x(t) + hv(t) + (h^2/2)f(x(t),t) + (h^3/6)f'(x(t),t) + ... \\ x(t-h) &= x(t) - hv(t) + (h^2/2)f(x(t),t) - (h^3/6)f'(x(t),t) + ... \end{aligned} \tag{4.6}$$

Πολύ εύκολα βλέπουμε ότι

$$x(t+h) = x(t-h) + 2hv(t) + \mathcal{O}(h^3) \tag{4.7}$$

δηλαδή στο $t = 0$ κάνουμε ένα βήμα πίσω για να υπολογίσουμε το $x(-h)$, χρησιμοποιώντας μια από τις άλλες μεθόδους, για παράδειγμα την καθιερωμένη μέθοδο του Euler, και έπειτα υπολογίζουμε τα $x(h)$, $x(3h)$, ... με τη νέα μέθοδο που ονομάζεται *μέθοδος του ενδιάμεσου σημείου*.

Παράδειγμα: Για την εξίσωση $x' = -x$ με αρχική συνθήκη $x(0) = 1$ και βήμα $h = 0.01$, στον Πίνακα 4.1 μπορούμε να συγκρίνουμε τη μέθοδο του

t	Euler	Ενδ.σημ.	Ακριβές αποτ.
1.00000×10^{-2}	0.990000	0.990000	0.990050
2.00000×10^{-2}	0.980100	0.980200	0.980199
3.00000×10^{-2}	0.970299	0.970396	0.970446
4.00000×10^{-2}	0.960596	0.960792	0.960789
5.00000×10^{-2}	0.950990	0.951180	0.951229
6.00000×10^{-2}	0.941480	0.941768	0.941765
7.00000×10^{-2}	0.932065	0.932345	0.932394
8.00000×10^{-2}	0.922745	0.923122	0.923116
9.00000×10^{-2}	0.913517	0.913882	0.913931
1.00000×10^{-1}	0.904382	0.904844	0.904837

Πίνακας 4.1: Σύγκριση της μεθόδου του Euler με αυτή του ενδιαμέσου σημείου

Euler και τη μέθοδο του ενδιάμεσου σημείου. Στην περίπτωσή μας $dx/dt = f(x(t),t) = -x(t)$, οπότε η μέθοδος του Euler δίνει:

$$x(t+h) = x(t) + hf(x(t),t) = x(t) + h(-x(t)) = x(t) - hx(t) = x(t)(1-h)$$

Στην περίπτωση της μεθόδου του ενδιάμεσου σημείου, βάζοντας βήμα h αντί $2h$, η Εξ.(4.7) γράφεται για την περίπτωσή μας:

$$\begin{aligned} x(t+h) &= x(t) + hf\left(x(t+h/2), t+h/2\right) \\ &= x(t) + hf\left(x(t) + (h/2)f(x(t),t), t+h/2\right) \\ &= x(t) + hf\left(x(t) + (h/2)(-x(t)), t+h/2\right) \\ &= x(t) + h\left(-x(t) + (h/2)x(t)\right) = x(t)\left(1 - h + h^2/2\right) \end{aligned}$$

όπου στη δεύτερη ισότητα αντικαταστήσαμε το όρισμα $x(t+h/2)$ της f με το ανάπτυγμά του, δηλαδή χρησιμοποιώντας τη μέθοδο του Euler: $x(t+h/2) = x(t) + (h/2)f(x(t),t)$, ενώ στην τρίτη και τέταρτη ισότητα χρησιμοποιήσαμε ότι $f(x(t),t) = -x$ στο συγκεκριμένο παράδειγμά μας.

Υπάρχει όμως και ένας άλλος, περισσότερο διαισθητικός, τρόπος εφαρμογής της μεθόδου του ενδιάμεσου σημείου. Θεωρώντας πάλι το διάστημα $[t, t+h]$, όπως κάναμε στο παραπάνω παράδειγμα, αντί του $[t-h, t+h]$, οι παρακάτω σχέσεις είναι ουσιαστικά ισοδύναμες με την Εξ.(4.7):

$$\begin{aligned} k_1 &= hf(x(t),t) \\ k_2 &= hf(x(t) + k_1/2,\, t + h/2) \\ x(t+h) &= x(t) + k_2 \end{aligned} \tag{4.8}$$

Η ιδέα είναι πως η μέθοδος του Euler θεωρεί ότι το βασικό σημείο στο διάστημα $[t, t+h]$ είναι το t: $f(x(t),t)$. Η μέθοδος του ενδιαμέσου σημείου θεωρεί ότι το βασικό σημείο του διαστήματος είναι το ενδιάμεσο: $f(x(t+h/2), t+h/2)$.

4.2 Η Μέθοδος Runge-Kutta

Είναι δυνατό να προχωρήσουμε ακόμα περισσότερο και να φτάσουμε σε σφάλμα της τάξης του $\mathcal{O}(h^5)$. Αυτή είναι η *μέθοδος Runge-Kutta τετάρτης τάξης*, η οποία παίρνει την ακόλουθη μορφή:

$$\begin{aligned} k_1 &= hf(x(t),t) \\ k_2 &= hf(x(t)+k_1/2, t+h/2) \\ k_3 &= hf(x(t)+k_2/2, t+h/2) \\ k_4 &= hf(x(t)+k_3, t+h) \\ x(t+h) &= x(t) + \frac{1}{6}(k_1+2k_2+2k_3+k_4)) + \mathcal{O}(h^5) \end{aligned} \tag{4.9}$$

Σε προβλήματα μηχανικής, όπου η διαφορική εξίσωση είναι δεύτερου βαθμού, ορίζουμε την ταχύτητα $dx/dt = v$, και η διαφορική εξίσωση γίνεται $dv/dt = f(x(t),t)/m$. Σ' αυτήν την περίπτωση εφαρμόζουμε τις Εξ.(4.9) δύο φορές, μια για την ταχύτητα και μια για την επιτάχυνση.

Παράδειγμα: Αρμονικός ταλαντωτής. Θα ασχοληθούμε με συντομία με ένα κλασικό παράδειγμα: ένα οριζόντιο ελατήριο με μιά μάζα m στο ένα άκρο του. Η εξίσωση της κίνησης είναι:

$$m\frac{d^2x}{dt^2} = -kx \tag{4.10}$$

όπου x η απομάκρυνση από το σημείο ισορροπίας. Θέτουμε τις αρχικές συνθήκες $x(0)$ και $v(0) \equiv \frac{dx}{dt}|_{t=0}$ και μετατρέπουμε τη δεύτερου βαθμού εξίσωση σε σύστημα δύο εξισώσεων πρώτου βαθμού:

$$\begin{aligned} v &= \frac{dx}{dt} \\ -\frac{k}{m}x &= \frac{dv}{dt} \end{aligned} \tag{4.11}$$

Η πρακτική ερώτηση που παρουσιάζεται τώρα είναι πώς μπορούμε να ελέγξουμε τις μεθόδους του Euler, του ενδιάμεσου σημείου ή Runge-Kutta σε συγκεκριμένα παραδείγματα. Μια πολύ ισχυρή μέθοδος είναι αυτή των νόμων διατήρησης. Με άλλα λόγια, στη μηχανική, όταν δεν έχουμε τριβές, η ολική ενέργεια διατηρείται. Μπορούμε λοιπόν να θέσουμε την ερώτηση κατά πόσον αυτές οι προσεγγιστικές μέθοδοι ικανοποιούν αυτήν τη διατήρηση. Ας το δούμε σε ένα παράδειγμα.

Παράδειγμα: Έστω η μάζα m, η οποία βρίσκεται στην άκρη ενός ελατηρίου σταθεράς k. Η εξισώσεις της κίνησης, όπως είπαμε και παραπάνω, είναι:

$$\begin{aligned} \frac{dx}{dt} &= v \\ \frac{dv}{dt} &= -\frac{k}{m}x \end{aligned} \tag{4.12}$$

οι οποίες και διατηρούν την ολική ενέργεια $E = mv^2/2 + kx^2/2$. Χρησιμοποιώντας τη μέθοδο του Euler, οι εξισώσεις γράφονται:

$$\begin{aligned} x_{n+1} &= x_n + v_n h \\ v_{n+1} &= v_n + (-(k/m)x_n)\, h \end{aligned} \tag{4.13}$$

και η ολική ενέργεια γράφεται αμέσως:

$$E_{n+1} = mv_{n+1}^2/2 + kx_{n+1}^2/2 = (1+\delta)E_n$$

όπου $\delta = h^2 k/m$. Μπορούμε εύκολα να γράψουμε την E_{n+1} συναρτήσει της E_0:

$$E_{n+1} = (1+\delta)^n E_0 = \left(1 + n\delta + \mathcal{O}(n^2\delta^2)\right) E_0$$

Βλέπουμε λοιπόν ότι η μέθοδος του Euler δεν διατηρεί την ενέργεια, αλλά η διαφορά για κάθε βήμα μπορεί να θεωρηθεί «μικρή» αφού είναι της ίδιας τάξης με το σφάλμα της μεθόδου, $\mathcal{O}(h^2)$. Όμως, το σφάλμα είναι προσθετικό, και μετά από n βήματα γίνεται $n\delta$, δηλαδή πρώτης τάξης στο δ. Αν το $n\delta$ γίνει συγκρίσιμο με τη μονάδα, η προσεγγιστική μέθοδός μας δεν έχει πια έννοια. Συμπέρασμα: το σφάλμα, όντας προσθετικό, θέτει όριο στον αριθμό των βημάτων που μπορούμε να χρησιμοποιήσουμε στη μέθοδο του Euler:

$$n_{max} = 1/\delta = \frac{m}{h^2 k}$$

Βλέπουμε ότι ο μέγιστος αριθμός των βημάτων είναι ανάλογος της μάζας και αντιστρόφως ανάλογος της σταθεράς του ελατηρίου και του βήματος. Παρατηρούμε επίσης ότι η ποσότητα $n_{max}h$ έχει διαστάσεις χρόνου και επειδή η περίοδος είναι ανάλογη με το $\sqrt{m/k}$, μας ενδιαφέρει το $n_{max}h$ να είναι τουλάχιστον όσο η περίοδος. Οι συνθήκες αυτές θέτουν κάθε άλλο παρά τετριμμένους περιορισμούς στην εκλογή του h, με δεδομένες τις σταθερές του προβλήματος m και k.

Συμπέρασμα

Παραπάνω παρουσιάσαμε τα κυριότερα και άμεσης χρήσης σημεία της αριθμητικής επίλυσης συνήθων διαφορικών εξισώσεων. Ένα πολύ σοβαρό πρακτικό πρόβλημα είναι η εκλογή του βέλτιστου βήματος h και της μεταβολής του: πιο συγκεκριμένα, στις περισσότερες εφαρμογές πρέπει να μεταβάλλουμε το βήμα το οποίο θα πρέπει να είναι μικρό όταν, για παράδειγμα, οι δυνάμεις μεταβάλλονται γρήγορα με το χρόνο και μεγαλύτερο όταν η μεταβολή τους είναι αργή. Μπορούμε επίσης να «συντονίσουμε» την εκλογή του βήματος με την ταχύτητα. Παρόλα αυτά, με τις τεχνικές που παρουσιάστηκαν στο κεφάλαιο αυτό μπορούμε να λύσουμε όλα τα προβλήματα κλασικής μηχανικής με ένα βαθμό ελευθερίας και αυτό δεν είναι λίγο.

Κεφάλαιο 5

Πίνακες και εφαρμογές τους

Οι ηλεκτρονικοί υπολογιστές συναντούν πολλές δυσκολίες στην επεξεργασία πινάκων και μόνο πρόσφατα, με την εισαγωγή της FORTRAN90, οι διαδικασίες με πίνακες και διανύσματα μπορούν να εκτελεστούν με σοβαρές επιδόσεις. Θα ξεκινήσουμε με φυσικά προβλήματα που περιγράφονται από περισσότερους από ένα βαθμούς ελευθερίας οι οποίοι και αλληλεπιδρούν. Θα εξετάσουμε δύο αντιπροσωπευτικά παραδείγματα: τη λύση ενός γραμμικού συστήματος και τον υπολογισμό των ιδιοτιμών ενός πίνακα.

5.1 Λύση Γραμμικού Συστήματος Εξισώσεων

Ένα συχνό μαθηματικό πρόβλημα που συναντάμε είναι η λύση ενός γραμμικού συστήματος εξισώσεων, δηλαδή

$$A \cdot \mathbf{x} = \mathbf{b} \tag{5.1}$$

όπου $A \equiv A_{ij}$ είναι ένας $n \times n$ πίνακας, $\mathbf{x}$ είναι ένα άγνωστο διάνυσμα και $\mathbf{b}$ γνωστό διάνυσμα. Ο σκοπός μας είναι να υπολογίσουμε τις συνιστώσες του $\mathbf{x}$. Όπως είναι γνωστό, το πρόβλημα έχει λύση αν, και μόνον αν, $\det A \neq 0$, και η λύση γράφεται συμβολικά

$$\mathbf{x} = A^{-1} \cdot \mathbf{b} \tag{5.2}$$

Η ανησυχία έγκειται στο γεγονός ότι ο υπολογισμός του αντίστροφου ενός πίνακα είναι μια διαδικασία με κόστος και πολύ επιρρεπής σε αριθμητικές αστάθειες. Θα αναφέρουμε εδώ δύο μεθόδους οι οποίες οδηγούν στη λύση της Εξ.(5.1) μ' ένα τρόπο σταθερό: είναι η μέθοδος Gauss-Jordan και η μέθοδος LU.

5.2 Η Μέθοδος Gauss-Jordan

Η βασική ιδέα αυτής της μεθόδου είναι η αναγωγή του πίνακα A σε τριγωνικό και ακολούθως ο υπολογισμός των συνιστωσών x_i. Ο λόγος είναι γιατί αν ο A είναι τριγωνικός

$$A = \begin{pmatrix} a_{11} & a_{12} & ... & a_{1n} \\ 0 & a_{22} & ... & a_{2n} \\ ... & .... & .... & ... \\ 0 & 0 & 0 & a_{nn} \end{pmatrix} \tag{5.3}$$

βρίσκουμε αμέσως ότι

$$x_n = b_n/a_{nn} \tag{5.4}$$

$$x_{n-1} = (b_{n-1} - a_{n-1,n}x_n)/a_{n-1,n-1} \tag{5.5}$$

ή γενικά

$$x_{n-k} = \left(b_{n-k} - \sum_{l=n}^{n-k+1} a_{n-k,l}x_l \right) / a_{n-k,n-k}, \qquad k = 1, ..., n-1 \tag{5.6}$$

Το όλο πρόβλημα ανάγεται λοιπόν στο μετασχηματισμό του πίνακα A σε τριγωνική μορφή. Ακολουθούμε τον παρακάτω τρόπο:

1. Αντικαθιστούμε την τελευταία εξίσωση που περιγράφει η (5.1) με ένα γραμμικό συνδυασμό της ίδιας και της πρώτης εξίσωσης, κατάλληλο ώστε το στοιχείο a_{n1} να μηδενιστεί. Δηλαδή, πολλαπλασιάζουμε τη γραμμή $(a_{11}, ..., a_{1n})$ με $-a_{n1}/a_{11}$ και αντικαθιστούμε τη γραμμή $(a_{n1}, a_{n2}, ..., a_{nj}, ..., a_{nn})$ με την

   $$\left(0, a_{n2} - a_{12}\frac{a_{n1}}{a_{11}}, ..., a_{nj} - a_{1j}\frac{a_{n1}}{a_{11}}, ..., a_{nn} - a_{1n}\frac{a_{n1}}{a_{11}} \right)$$

   Βέβαια αντικαθιστούμε το b_n με το $b_n - b_1\left(\frac{a_{n1}}{a_{11}}\right)$.

2. Συνεχίζουμε με αυτόν τον τρόπο και μηδενίζουμε τα στοιχεία της πρώτης στήλης του πίνακα, εκτός βέβαια του a_{11}, καταλήγοντας στον πίνακα

   $$A = \begin{pmatrix} a_{11} & a_{12} & ... & a_{1n} \\ 0 & a_{22} & ... & a_{2n} \\ 0 & a_{32} & ... & a_{3n} \\ ... & .... & .... & ... \\ 0 & a_{k2} & ... & a_{kn} \\ ... & .... & .... & ... \\ 0 & a_{n2} & .... & a_{nn} \end{pmatrix} \tag{5.7}$$

   όπου έχουμε χρησιμοποιήσει τα ίδια σύμβολα για τα στοιχεία των γραμμών $2, ..., n$.

3. Συνεχίζουμε με ανάλογο τρόπο στο μηδενισμό των στοιχείων a_{n2}, $a_{n-1,2}$, ..., a_{32}. Θεωρούμε τον πίνακα

$$\begin{pmatrix} a_{22} & ... & a_{2n} \\ a_{32} & ... & a_{3n} \\ .... & .... & ... \\ a_{k2} & ... & a_{kn} \\ .... & .... & ... \\ a_{n2} & .... & a_{nn} \end{pmatrix} \tag{5.8}$$

πολλαπλασιάζουμε τα στοιχεία $(a_{22}, ..., a_{2n})$ με $-a_{n2}/a_{22}$ και αντικαθιστούμε την τελευταία γραμμή με το αποτέλεσμα το οποίο έχει $a_{32} = ... = a_{n2} = 0$.

4. Είναι εύκολο πλέον να πεισθούμε ότι στο τέλος του m βήματος, θα έχουμε μηδενίσει τα στοιχεία $a_{n1}, a_{n-1,1}, ..., a_{21}; a_{n2}, a_{n-1,2}, ..., a_{32}$; $..., a_{n,m-1}, ..., a_{n-1,m-1}, ..., a_{n,m-1}$. Στο τέλος του $n-1$ βήματος ο πίνακας θα είναι τριγωνικός.

Παράδειγμα:

$$\begin{pmatrix} 1 & 2 & 1 \\ 1 & 1 & 1 \\ 1 & -1 & -1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} \tag{5.9}$$

Πολλαπλασιάζουμε την πρώτη γραμμή επί 1 και αντικαθιστούμε την τρίτη με $(0, 3, 2)$. Βρίσκουμε

$$\begin{pmatrix} 1 & 2 & 1 \\ 1 & 1 & 1 \\ 0 & 3 & 2 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ -2 \end{pmatrix}$$

Τώρα, πολλαπλασιάζουμε την πρώτη γραμμή επί 1 και την αφαιρούμε από τη δεύτερη (για να απαλλείψουμε το a_{21}). Το σύστημα γίνεται

$$\begin{pmatrix} 1 & 2 & 1 \\ 0 & 1 & 0 \\ 0 & 3 & 2 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \\ -2 \end{pmatrix}$$

Πολλαπλασιάζουμε τη δεύτερη γραμμή επί 3 και την αφαιρούμε από την τρίτη, για να απαλλείψουμε το a_{32}, και τελειώσαμε. Παίρνουμε

$$\begin{pmatrix} 1 & 2 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \\ -1 \end{pmatrix}$$

και βρίσκουμε άμεσα ότι $z = 1/2$, $y = -1$ και $x = 5/2$. Αντικαθιστώντας τις τιμές αυτές στο αρχικό σύστημα της Εξ.(5.9), επαληθεύουμε την ορθότητα της λύσης.

Έχουμε ήδη παρατηρήσει τη σπουδαιότητα που παίζουν οι διαγώνιοι όροι a_{nn}, καθώς διαιρούμε με αυτούς. Οπότε, η ερώτηση είναι άμεση: τι γίνεται αν ένας οι περισσότεροι από αυτούς είναι μηδενικοί; Η ιδέα είναι ότι αν το σύστημα δεν είναι ιδιάζον, είναι δυνατό, με εναλλαγές των εξισώσεων ή/και των αγνώστων, να καταλήξουμε σε τριγωνικό πίνακα με διαγώνια στοιχεία μη μηδενικά. Πράγματι, το σύστημα δεν αλλάζει αν εναλλάξουμε δύο εξισώσεις. Αντίθετα, αν εναλλάξουμε δύο στήλες, οι μεταβλητές μπερδεύονται. Όμως αυτό το μπέρδεμα αντιστοιχεί σε γραμμικό μετασχηματισμό. Ο τελικός σκοπός της διαδικασίας αυτής είναι κάθε διαγώνιο στοιχείο να έχει τη μεγαλύτερη απόλυτη τιμή από όλα τα στοιχεία της αντίστοιχης γραμμής και στήλης. Αυτή η διαδικασία ονομάζεται *περιστροφή* και το διαγώνιο στοιχείο *περιστροφέας*. Θα χρησιμοποιήσουμε τη *μερική περιστροφή* που είναι απλούστερη σε εφαρμογή και ικανή να κρατήσει την όλη διαδικασία σταθερή. Μ' αυτή τη διαδικασία, εναλλάσουμε τις εξισώσεις ώστε κάθε διαγώνιο στοιχείο να έχει τη μεγαλύτερη απόλυτη τιμή στην αντίστοιχη στήλη (ενώ η πλήρης περιστροφή απαιτεί τη μεγαλύτερη απόλυτη τιμή και στην αντίστοιχη γραμμή). Ένα παράδειγμα θα μας διαφωτίσει.

Παράδειγμα:

$$\begin{pmatrix} 2 & 2 & 1 \\ 1 & 1 & 1 \\ 1 & -1 & -1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} \tag{5.10}$$

Βρίσκουμε

$$\begin{pmatrix} 2 & 2 & 1 \\ 1 & 1 & 1 \\ 0 & 4 & 3 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ -5 \end{pmatrix}$$

$$\begin{pmatrix} 2 & 2 & 1 \\ 0 & 0 & -1 \\ 0 & 4 & 3 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 1 \\ -3 \\ -5 \end{pmatrix}$$

Βλέπουμε ότι $a_{22} = 0$. Αλλά μπορούμε να εναλλάξουμε τις γραμμές 2 και 3 χωρίς να αλλάξουμε τους αγνώστους. Οπότε καταλήγουμε

$$\begin{pmatrix} 2 & 2 & 1 \\ 0 & 4 & 3 \\ 0 & 0 & -1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 1 \\ -5 \\ -3 \end{pmatrix}$$

Γενικά, θα πρέπει να ελέγχουμε την απόλυτη τιμή $|a_{mm}|$. Αν $|a_{mm}| < \epsilon$, θα πρέπει να εναλλάξουμε τη γραμμή m με τη γραμμή j, όπου $|a_{jm}| = \max_{l=1,\ldots,n}\{|a_{lm}|\}$. Βέβαια θα πρέπει $b_m \to b_j$.

5.3 Η Μέθοδος LU

Η μέθοδος αυτή στηρίζεται επίσης στην ιδέα της επίλυσης ενός συστήματος μετατρέποντας τον αντίστοιχο πίνακα σε τριγωνικό. Η ιδέα έγκειται στην παραγοντοποίηση του αρχικού πίνακα σε γινόμενο δύο πινάκων: ενός κάτω-τριγωνικού (Lower-triangular), δηλαδή $a_{ij} = 0$ για $j > i$ και ενός άνω-τριγωνικού (Upper-triangular), δηλαδή $a_{ij} = 0$ για $j < i$.

$$A \cdot x = (L \cdot U) \cdot x = L \cdot (U \cdot x) = b \tag{5.11}$$

και αν θέσουμε $y = U \cdot x$ μπορούμε να υπολογίσουμε τις συνιστώσες του y άμεσα, με αντικατάσταση, και ύστερα τις συνιστώσες του x. Πώς βρίσκουμε λοιπόν τους πίνακες L και U; Ας πάρουμε το παράδειγμα ενός πίνακα 4×4.

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{pmatrix} = \begin{pmatrix} l_{11} & 0 & 0 & 0 \\ l_{21} & l_{22} & 0 & 0 \\ l_{31} & l_{32} & l_{33} & 0 \\ l_{41} & l_{42} & l_{43} & l_{44} \end{pmatrix} \cdot \begin{pmatrix} u_{11} & u_{12} & u_{13} & u_{14} \\ 0 & u_{22} & u_{23} & u_{24} \\ 0 & 0 & u_{33} & u_{34} \\ 0 & 0 & 0 & u_{44} \end{pmatrix} \tag{5.12}$$

Καταλαβαίνουμε ότι η Εξ.(5.12) αποτελεί ένα σύστημα n^2 εξισώσεων (όσα και τα στοιχεία του πίνακα A) με $n^2 + n$ αγνώστους (τις μεταβλητές l_{ij} και u_{ij}). Δηλαδή, μπορούμε να θέσουμε n συνθήκες. Για να απλοποιήσουμε τα πράγματα διαλέγουμε

$$l_{ii} = 1, \quad i = 1, ..., n$$

και μπορούμε να υπολογίσουμε τα υπόλοιπα στοιχεία με τον *αλγόριθμο του Crout*:

Για κάθε $j = 1, 2, 3, ..., n$

- Για $i = 1, 2, ..., j$

$$u_{ij} = a_{ij} - \sum_{k=1}^{i-1} l_{ik} u_{kj}$$

  (το άθροισμα είναι 0 για $i = 1$).

- Για $i = j + 1, ..., n$

$$l_{ij} = \frac{1}{u_{jj}} \left(a_{ij} - \sum_{k=1}^{j-1} l_{ik} u_{kj} \right)$$

Παράδειγμα:

Θα λύσουμε το προηγούμενο παράδειγμα με τη μέθοδο LU

$$\begin{pmatrix} 1 & 2 & 1 \\ 1 & 1 & 1 \\ 1 & -1 & -1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$$

Θέτουμε $l_{11} = l_{22} = l_{33} = 1$.

$j = 1$

$i = 1, \quad u_{11} = a_{11} = 1$

$i = 2, \quad l_{21} = \frac{1}{u_{11}} a_{21} = 1$

$i = 3, \quad l_{31} = \frac{1}{u_{11}} a_{31} = 1$

$j = 2$

$i = 1, \quad u_{12} = a_{12} = 2$

$i = 2, \quad u_{22} = a_{22} - l_{21}u_{12} = 1 - 1 \times 2 = -1$

$i = 3, \quad l_{32} = \frac{1}{-1}(a_{32} - l_{31}u_{12}) = -(-3) = +3$

$j = 3$

$i = 1, \quad u_{13} = a_{13} = 1$

$i = 2, \quad u_{23} = a_{23} - l_{21}u_{13} = 1 - 1 \times 1 = 0$

$i = 3, \quad u_{33} = a_{33} - (l_{31}u_{13} + l_{32}l_{23}) = -1 - (1 \times 1 + 3 \times 0) = -2$

Επομένως, οδηγούμαστε στα παρακάτω συστήματα:

$$\begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 3 & 1 \end{pmatrix} \begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 2 & 1 \\ 0 & -1 & 0 \\ 0 & 0 & -2 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} x' \\ y' \\ z' \end{pmatrix}$$

και βρίσκουμε $x' = 1, y' = 1, z' = 1$ και $x = 5/2, y = -1, z = 1/2$.

Ένα μαθηματικό ερώτημα παραμένει αναπάντητο όμως: ποιες είναι οι συνθήκες, αναγκαίες και ικανές, που επιτρέπουν την παραγοντοποίηση ενός πίνακα 4×4 σε γινόμενο $L \cdot U$; Είναι δυνατό να αποδειχτεί ότι αρκεί όλοι οι *ελάσσονες* του A να είναι μη μηδενικοί. Αλλά είναι δυνατό να έχουμε ένα μη ιδιάζον σύστημα με κάποιον ελάσσονα μηδενικό. Η απάντηση είναι ότι μπορούμε να παραγοντοποιήσουμε στη μορφή $L \cdot U$ όχι τον A αλλά τον πίνακα όπου έχουμε εναλλάξει τις γραμμές του. Είναι και πάλι μια περιστροφή στον αλγόριθμο του Crout. Η ιδέα προέρχεται από το δεύτερο βήμα του αλγορίθμου αυτού. Αντιλαμβανόμαστε ότι μπορούμε να υπολογίσουμε τον αριθμητή και τον παρονομαστή ανεξάρτητα. Αφού υπολογίσουμε τους αριθμητές, τότε μπορούμε να αποφασίσουμε. Επομένως διαλέγουμε, το μεγαλύτερο από τα στοιχεία u της στήλης j, εναλλάσσουμε τη γραμμή j με τη γραμμή i_{max} και διαιρούμε *μόνο* τα στοιχεία του δεύτερου βήματος με τα καινούργια u_{jj}. Ο λόγος που ένα τέτοιο τέχνασμα επιτυγχάνει είναι ότι δεν χρειάζεται να αποθηκεύουμε στη μνήμη *τρεις* πίνακες $n \times n$ (τους πίνακες A, L και U) αλλά τους αντικαθιστούμε «επί τόπου», επειδή κάθε στοιχείο a_{ij} δεν χρησιμοποιείται παρά μόνο μια φορά.

5.4 Υπολογισμός Ορίζουσας

Ένα πολύ ενδιαφέρον μέγεθος, τόσο από μαθηματική σκοπιά όσο και από τη σκοπιά της φυσικής, είναι η *ορίζουσα* ενός πίνακα. Η έννοιά της στη φυσική είναι αυτή του «όγκου» –σε δύο διαστάσεις αντιστοιχεί σε (προσανατολισμένη) «επιφάνεια». Στη γραμμική άλγεβρα μαθαίνουμε ότι η ορίζουσα αποτελεί τμήμα των «αναλλοίωτων» ποσοτήτων –ποσότητες οι οποίες δεν αλλάζουν τιμή σε μετασχηματισμούς ομοιότητας, $A \to SAS^{-1}$. Μαθαίνουμε επίσης πώς να υπολογίζουμε την ορίζουσα (αναλύοντάς την σε ελάσσονες ορίζουσες). Στη γενική περίπτωση, αυτή η ανάλυση χρειάζεται $\mathcal{O}(n^3)$ επιμέρους υπολογισμούς, αλλά απαιτούνται μόνο $\mathcal{O}(n)$ υπολογισμοί για την περίπτωση που ο πίνακας είναι τριγωνικός –γιατί γνωρίζουμε ότι η ορίζουσα τριγωνικού πίνακα είναι ίση με το γινόμενο των διαγωνίων όρων. Παίρνοντας υπόψη ότι τόσο η μέθοδος Gauss-Jordan όσο και η μέθοδος LU μετασχηματίζουν τον αρχικό πίνακα σε τριγωνικό, καταλαβαίνουμε πόσο εύκολος είναι ο υπολογισμός μιας ορίζουσας. Αλλά απαιτείται προσοχή. Γνωρίζουμε ότι το πρόσημο μιας ορίζουσας αλλάζει με την εναλλαγή δύο γραμμών ή δύο στηλών. Απαιτείται λοιπόν η ύπαρξη μιας βοηθητικής μεταβλητής η οποία θα αλλάζει πρόσημο σε κάθε εναλλαγή γραμμών κατά τη διάρκεια της περιστροφής. Αυτή η μεταβλητή θα δίνει και το τελειωτικό πρόσημο στην ορίζουσα

$$\det A = \text{ομοτιμία} \times \prod_{k=1}^{n} a_{kk}$$

Ποια είναι τα σημαντικά σημεία; Το γινόμενο μπορεί να γίνει πολύ μεγάλο ή πολύ μικρό (σε απόλυτη τιμή) –επομένως θα πρέπει να αθροίσουμε λογαρίθμους. Αλλά θα πρέπει επίσης να βλέπουμε και από τη σκοπιά της φυσικής. Δηλαδή, αν μία (ή περισσότερες) ιδιοτιμές τείνουν στο μηδέν (οπότε η ορίζουσα τείνει επίσης στο μηδέν), είναι σημάδι ότι ίσως υπάρχει μια «ψευδοσυμμετρία», την οποία αξίζει να μελετήσουμε.

Κεφάλαιο 6

Ιδιοτιμές και Ιδιοδιανύσματα Συμμετρικού Πίνακα

Ο πίνακας αποτελεί την αναπαράσταση ενός τελεστή. Είναι γνωστό ότι αντικατοπτρίζει την εκλογή μιας *βάσης* στο χώρο αυτό και μπορεί να τεθεί το ερώτημα αν υπάρχουν ποσότητες, χαρακτηριστικές μιας αναπαράστασης, οι οποίες είναι ανεξάρτητες από την εκλογή της βάσης. Με άλλα λόγια, αν μπορούμε να ορίσουμε κατά κάποιο τρόπο «κλάσεις ισοδυναμίας». Όλοι οι πίνακες που έχουν τις ίδιες ανεξάρτητες ποσότητες θα ανήκουν στην ίδια κλάση. Είναι γνωστό ότι όλοι οι πίνακες που συνδέονται μεταξύ τους με μετασχηματισμό *ομοιότητας*

$$A \to S^{-1}AS \tag{6.1}$$

έχουν τις ίδιες ιδιοτιμές, δηλαδή οι λύσεις της εξίσωσης

$$\det(A - \lambda 1) = 0 \tag{6.2}$$

είναι οι ίδιες. Επομένως, είναι προφανές ότι αυτές οι ποσότητες, οι ιδιοτιμές, αποτελούν πραγματικά ενδιαφέροντα στοιχεία, κυρίως από πλευράς φυσικής[1]. Είναι επίσης διασκεδαστικό να παρατηρήσει κανείς ότι η προσπάθεια *κατάταξης των αναλλοίωτων* είχε αρχίσει από τους μαθηματικούς πολύ πιο νωρίς: τυπικά από τους Sylvester και Cayley στα μέσα του 19ου αιώνα, με εργασίες όπου εμπλέκονται και τα ονόματα των Klein, Jordan, Hilbert αλλά ο κατάλογος αυτός είναι σίγουρα ημιτελής. Δεν θα πρέπει τέλος να λησμονήσουμε τους θεμελιωτές της προβολικής γεωμετρίας στις αρχές του 19ου αιώνα όπως τους Poncelet και Monge. Άμεσα μπαίνει η πρακτική ερώτηση: πώς τις υπολογίζουμε; Στη γενική περίπτωση είναι ένα δύσκολο εγχείρημα. Αλλά αν περιοριστούμε

[1] Μία πολύ ωραία συζήτηση μπορείτε να βρείτε στο βιβλίο *The Character of Physical Law του R.P. Feynman, MIT Press (1965).*

στις περιπτώσεις που σχετίζονται με τις φυσικές επιστήμες, καταλαβαίνουμε γρήγορα ότι αναφερόμαστε σε συμμετρικούς πίνακες[2]στις περισσότερες των περιπτώσεων. Θα περιοριστούμε λοιπόν σ' αυτούς τους πίνακες.

Επίσης, είναι γνωστό ότι υπάρχει ένα θεώρημα το οποίο μας βεβαιώνει ότι οι ιδιοτιμές ενός συμμετρικού πίνακα είναι *πραγματικές* και τα αντίστοιχα *ιδιοδιανύσματα* μπορούν πάντοτε να εκλεγούν με τέτοιο τρόπο ώστε να είναι ορθογώνια ανά δύο. Αποτέλεσμα: μπορούμε να περιορίσουμε τη μελέτη μας σε συμμετρικούς πίνακες A καθώς και σε ορθογώνιους πίνακες S, δηλαδή γι' αυτούς που ισχύει $S^{-1} = S^T$. Η ιδέα είναι να καταλήξουμε σε ένα διαγώνιο πίνακα A (επομένως με τα διαγώνια στοιχεία να είναι οι ιδιοτιμές του πίνακα), μετά από μια σειρά ορθογωνίων μετασχηματισμών που σκοπό έχουν να μηδενίσουν τα μη διαγώνια στοιχεία

$$A \to P_{12}^T A P_{12} \to P_{13}^T P_{12}^T A P_{12} P_{13} \to \ldots \left(\prod P\right)^T A \prod P = \Lambda \qquad (6.3)$$

Το τέχνασμα λοιπόν έγκειται στην εξεύρεση μιας οικογένειας πινάκων P των οποίων ο πολλαπλασιασμός να γίνεται εύκολα. Ο Jacobi (1846) πρότεινε ο μηδενισμός των μη διαγωνίων στοιχείων να γίνεται με συνεχείς *επίπεδες* στροφές. Μ' αυτό εννοούμε ότι για το μηδενισμό του στοιχείου a_{kl} του πίνακα A, διαλέγουμε έναν πίνακα P_{kl} ο οποίος είναι ο ταυτοτικός πίνακας εκτός των στοιχείων $p_{kk} = p_{ll} = \cos\theta$ και $p_{kl} = -p_{lk} = \sin\theta$, και διαλέγουμε τη γωνία θ τέτοια ώστε $a'_{kl} = a'_{lk} = (P^T A P)_{kl} = 0$. Συνεχίζουμε με τον τρόπο αυτό για όλα τα μη διαγώνια στοιχεία. Είναι φανερό λοιπόν ότι για πίνακα $N \times N$, το τίμημα για τόσους πολλαπλασιασμούς πινάκων γίνεται απαγορευτικό. Από την άλλη, πολλές πράξεις είναι άχρηστες, εφόσον μόνο 2 στήλες και 2 γραμμές του πίνακα A αλλάζουν σε κάθε βήμα της μεθόδου Jacobi: οι γραμμές k και l και οι στήλες k και l. Επομένως είναι προτιμότερο να γράψουμε αναλυτικά τα στοιχεία του πίνακα που αλλάζουν

$$a'_{kl} = \sin\theta\cos\theta(a_{kk} - a_{ll}) + a_{kl}(\cos^2\theta - \sin^2\theta) \qquad (6.4)$$

η οποία σχέση δίνει για τη γωνία που μηδενίζει το αντίστοιχο στοιχείο

$$\theta^*_{kl} = \frac{1}{2}\tan^{-1}\frac{-2a_{kl}}{a_{kk} - a_{ll}} \qquad (6.5)$$

Η έκφραση αυτή παραμένει πολύ θεωρητική ακόμα. Στην πράξη, θα πρέπει να ελέγξουμε αν $|a_{kk} - a_{ll}| < \epsilon$. Σ' αυτήν την περίπτωση θέτουμε $\theta^*_{kl} = \pi/4$. Επίσης, αν$|a_{kl}| < \epsilon$, δεν είναι ανάγκη να εκτελέσουμε την περιστροφή και θέτουμε απ' ευθείας το στοιχείο αυτό ίσο με μηδέν. Τελικά, μετά την εκτέλεση $n(n-1)/2$ περιστροφών δεν θα καταλήξουμε σ' έναν τριγωνικό πίνακα, ακόμα και με την ακρίβεια του συγκεκριμένου υπολογιστικού συστήματος. Η αιτία είναι ότι ο αλγόριθμος αυτός, γενικά, δεν συγκλίνει σε πεπερασμένο αριθμό

[2]Στην περίπτωση πραγματικών πινάκων. Για μιγαδικούς πίνακες, οι οποίοι επίσης παίζουν ρόλο στη φυσική, αναφερόμαστε σε *ερμιτιανούς* (hermitian) πίνακες.

βημάτων. Αλλά, στην πράξη, 4-10 επαναλήψεις, από τις $n(n-1)/2$ που θα πρέπει να γίνουν συνολικά, είναι αρκετές για την αναγωγή της απόλυτης τιμής των μη διαγωνίων όρων κάτω από την ακρίβεια του υπολογιστή, οπότε και σταματάμε τη διαδικασία. Αν ο πίνακας έχει μια ιδιαίτερη συμμετρία, η όλη διαδικασία συγκλίνει πολύ γρήγορα.

Παράδειγμα: $n = 2$. Υποθέστε ότι έχετε τον 2×2 πίνακα

$$A = \begin{pmatrix} a & b \\ b & d \end{pmatrix} \tag{6.6}$$

Πρέπει να μηδενίσουμε το στοιχείο $a_{12} = b$ οπότε μία στροφή είναι αρκετή

$$A' = \begin{pmatrix} \cos\theta & -sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \cdot \begin{pmatrix} a & b \\ b & d \end{pmatrix} \cdot \begin{pmatrix} \cos\theta & sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \tag{6.7}$$

και τελικά βρίσκουμε

$$\theta^*_{12} = \frac{1}{2}\tan^{-1}\left(\frac{-2b}{a-d}\right) \tag{6.8}$$

Τα στοιχεία a'_{11} και a'_{22} εξαρτώνται, γενικά, από τη γωνία θ. Για $\theta = \theta^*_{12}$, είναι ακριβώς οι ιδιοτιμές του πίνακα A και οι στήλες του πίνακα στροφής αποτελούν τα *ιδιοδιανύσματα*, δηλαδή τα διανύσματα για τα οποία η δράση του πίνακα αποτελεί απλά μια *διαστολή* (ή συστολή).

Παράδειγμα: $n = 3$. Υποθέστε ότι έχετε τον 3×3 πίνακα

$$A = \begin{pmatrix} a & b & b \\ b & a & b \\ b & b & a \end{pmatrix} \tag{6.9}$$

Καταρχήν χρειαζόμαστε 3 επαναλήψεις για το μηδενισμό των στοιχείων $a_{12} = b$, $a_{13} = b$, $a_{23} = b$. Οι τρεις πίνακες στροφής είναι

$$P_{12} = \begin{pmatrix} \cos\theta & \sin\theta & 0 \\ -\sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{6.10}$$

$$P_{13} = \begin{pmatrix} \cos\theta & 0 & \sin\theta \\ 0 & 1 & 0 \\ -\sin\theta & 0 & \cos\theta \end{pmatrix} \tag{6.11}$$

$$P_{23} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & \sin\theta \\ 0 & -\sin\theta & \cos\theta \end{pmatrix} \tag{6.12}$$

Αρχίζουμε από τον υπολογισμό της ποσότητας $P_{12}^T A P_{12}$. Ο άμεσος υπολογισμός, ή η εφαρμογή του γενικού τύπου, οδηγούν στη γωνία $\theta^*_{12} = \pi/4$, η οποία μηδενίζει όχι μόνο το στοιχείο a_{12}, αλλά και το a_{13} και επομένως δεν χρειάζεται να κάνουμε τη στροφή που μηδενίζει το a_{13}. Έτσι έχουμε ήδη μία ιδιοτιμή, $\lambda_1 = a - b$. Απομένει τώρα να μηδενίσουμε το στοιχείο $a^*_{23}(\theta^*12) = b\sqrt{2}$,

χρησιμοποιώντας τη στροφή P_{23}. Σημειώνουμε ότι, στη συγκεκριμένη περίπτωση, η εφαρμογή της τελευταίας στροφής δεν επηρεάζει την τιμή των μη διαγωνίων στοιχείων, τα οποία είναι ήδη μηδέν, οπότε μπορούμε να απομονώσουμε τον υποπίνακα 2×2 και να γράψουμε μια στροφή σε 2 διαστάσεις. Βρίσκουμε έτσι ότι $\tan 2\theta_{23}^* = -2\sqrt{2}$ και $a''_{22}(\theta_{23}^*) = a - b$, $a''_{33}(\theta_{23}^*) = a + 2b$. Τα ιδιοδιανύσματα είναι οι στήλες του πίνακα $P_{12}(\theta_{12}^*) \cdot P_{23}(\theta_{23}^*)$ και αποτελούν ορθογώνιο σύστημα.

Βιβλιογραφία

1. *Numerical Recipies: The Art of Scientific Computing in Fortran, W. Press et al. Cambridge University Press (1992).* Το βιβλίο αυτό μπορείτε να το βρείτε ακόμα και στο διαδίκτυο: http://cfatab.harvard.edu/nr/nronline.html.

2. Για μια ωραία συζήτηση σχετική με την αριθμητική επίλυση διαφορικών εξισώσεων καθώς και για αριθμητικές προσεγγίσεις εν γένει, βλέπε R.P. Feynman, *The Feynman Lectures in Physics*, vol I. Addisson-Wesley (1963).

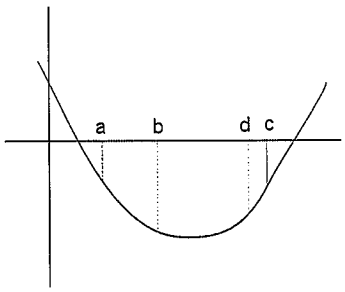
Κεφάλαιο 7

Ελαχιστοποίηση συναρτήσεων

Το πρόβλημα της εύρεσης του ελαχίστου μιας συνάρτησης (μιας ή περισσότερων μεταβλητών) είναι αρκετά δύσκολο αλλά με μια μεγάλη σειρά εφαρμογών που δικαιολογούν απόλυτα την προσπάθεια που απαιτείται. Για παράδειγμα, τα ελάχιστα ενός δυναμικού είναι τα σημεία ισορροπίας, με προφανή φυσική σημασία. Σημειώνουμε ότι η εύρεση του μέγιστου μιας συνάρτησης $f(x_1, x_2, \ldots, x_N)$ ανάγεται στην εύρεση του ελάχιστου της συνάρτησης $-f(x_1, x_2, \ldots, x_N)$, οπότε δεν είναι καινούργιο πρόβλημα. Το πρόβλημα για συναρτήσεις μιας μεταβλητής είναι απλούστερο, ενώ για περισσότερες διαστάσεις η δυσκολία αυξάνει εντυπωσιακά. Αξίζει να σημειώσουμε ότι η δοσμένη συνάρτηση ενδέχεται να έχει περισσότερα από ένα τοπικά ελάχιστα. Το ποιό από τα ελάχιστα θα εντοπιστεί με τις μεθόδους που θα αναπτύξουμε εξαρτάται από το σημείο εκκίνησης, άρα είναι γενικά απρόβλεπτο σε ποιο ελάχιστο θα καταλήξουμε. Απαιτείται, λοιπόν, κάποια στοιχειώδης γνώση της συνάρτησης και κάποια έστω και όχι πολύ σαφής γνώση της θέσης των ελαχίστων, αλλιώς βαδίζει κανείς χωρίς καθοδήγηση και τα αποτελέσματά του ενδεχομένως να μην είναι διαφωτιστικά.

7.1 Μονοδιάστατη Ελαχιστοποίηση

Θα περιγράψουμε μια μέθοδο που δίνει αποτέλεσμα για μια αρκετά ευρεία κατηγορία συναρτήσεων $f(x)$ μιας πραγματικής μεταβλητής. Απαραίτητη προϋπόθεση για να αρχίσει κανείς τη μέθοδο είναι να έχει βρεί προηγουμένως τρία σημεία, έστω τα $a < b < c$, τέτοια ώστε η τιμή της συνάρτησης $f(b)$ στο μεσαίο σημείο να είναι μικρότερη και από τις δύο τιμές $f(a)$ και $f(c)$. Αυτό εγγυάται ότι το ελάχιστο βρίσκεται κάπου μέσα στο διάστημα $[a, c]$. Σημειώνουμε ότι το σχετικό μέγεθος των $f(a)$ και $f(c)$ δεν παίζει ρόλο. Το επόμενο



Σχήμα 7.1: Η περίπτωση $f(b) < f(d)$.

βήμα είναι να θεωρήσει κανείς ένα τέταρτο σημείο, το d, και να συγκρίνει την τιμή $f(d)$ της συνάρτησης στο νέο σημείο με την $f(b)$. Για να είμαστε σαφείς, θεωρούμε ότι $d > b$. Αν $f(d) > f(b)$, όπως συμβαίνει για τη συνάρτηση του Σχ.(7.1), το συμπέρασμα είναι ότι το ελάχιστο θα βρίσκεται στο διάστημα $[a, d]$, ενώ οι μέχρι τώρα πληροφορίες μας το εντόπιζαν στο μεγαλύτερο διάστημα $[a, c]$. Αν, αντίθετα, $f(d) < f(b)$, όπως συμβαίνει για τη συνάρτηση του σχήματος Σχ.(7.2), το συμπέρασμα είναι ότι το ελάχιστο θα βρίσκεται στο διάστημα $[b, c]$, το οποίο επίσης είναι μικρότερο από το αρχικό $[a, c]$. Σε οποιαδήποτε περίπτωση, λοιπόν, έχουμε κάνει ένα βήμα προς τον ακριβέστερο προσδιορισμό του ελάχιστου: το ελάχιστο θα βρίσκεται όχι απλά στο διάστημα $[a, c]$, αλλά σε κάποιο από τα μικρότερα: $[a, d]$ ή $[b, c]$, ανάλογα με το πρόσημο της παράστασης $f(d) - f(b)$. Δηλαδή τη θέση της αρχικής τριάδας σημείων $a < b < c$ την παίρνει μια καινούρια τριάδα: είτε η $a < b < d$ είτε η $b < d < c$, όπως μόλις εξηγήσαμε. Τη συνέχεια είναι εύκολο να τη φανταστεί κανείς: ξεκινώντας από την καινούργια τριάδα που προσδιορίστηκε, θεωρεί ένα τέταρτο σημείο και επαναλαμβάνει τη διαδικασία. Το αποτέλεσμα θα είναι να εντοπιστεί το ελάχιστο σ' ένα ακόμη μικρότερο διάστημα. Αυτή η διαδικασία θα συνεχιστεί μέχρι να επιτευχθεί ένας ικανοποιητικός προσδιορισμός του ελάχιστου. Μέχρι στιγμής έχουμε υποθέσει ότι $d > b$. Αν $d < b$ συμβαίνουν ανάλογα πράγματα. Είμαστε σε θέση τώρα να διατυπώσουμε τον κανόνα:

Σχήμα 7.2: Η περίπτωση $f(b) > f(d)$.

Περίπτωση $d > b$	Αν $f(d) > f(b)$	η επόμενη τριάδα είναι η	(a, b, d)
	Αν $f(d) < f(b)$		(b, d, c)
Περίπτωση $d < b$	Αν $f(d) > f(b)$		(d, b, c)
	Αν $f(d) < f(b)$		(a, d, b)

Πόσο όμως μας επιτρέπει ο υπολογιστής να πλησιάσουμε το ελάχιστο; Απλοϊκά σκεπτόμενοι θα μπορούσαμε να περιορίσουμε το $x_{\epsilon\lambda}$ στη περιοχή $[(1-\epsilon)x_{\epsilon\lambda}, (1+\epsilon)x_{\epsilon\lambda}]$, όπου ϵ είναι η (σχετική) ακρίβεια του υπολογιστή ($3\cdot 10^{-8}$ για απλή ακρίβεια και 10^{-15} για διπλή). Δυστυχώς όμως τα πράγματα είναι χειρότερα! Κοντά σε ελάχιστο, κάθε συνάρτηση μπορεί να αναπτυχθεί κατά Taylor

$$f(x) \approx f(x_{\epsilon\lambda}) + \frac{1}{2} f''(x_{\epsilon\lambda})(x - x_{\epsilon\lambda})^2$$

Η απαίτηση ο δεύτερος όρος να είναι ίσος με ϵ επί τον πρώτο, δηλαδή νά έχουμε αγγίξει την ακρίβεια του υπολογιστή, $(1/2)f''(x_{\epsilon\lambda})(x - x_{\epsilon\lambda})^2 \approx \epsilon f(x_{\epsilon\lambda})$ γράφεται

$$|x - x_{\epsilon\lambda}| = \sqrt{\epsilon}\sqrt{\frac{2|f(x_{\epsilon\lambda})|}{f''(x_{\epsilon\lambda})}} = \sqrt{\epsilon}\,|x_{\epsilon\lambda}|\sqrt{\frac{2|f(x_{\epsilon\lambda})|}{x_{\epsilon\lambda}^2 f''(x_{\epsilon\lambda})}}$$

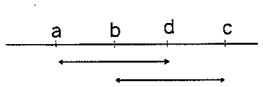
Ο λόγος που γράψαμε την τελευταία ισότητα είναι ότι η τετραγωνική ρίζα είναι της τάξης της μονάδας, για κάθε συνάρτηση. Αυτό μπορούμε να το δούμε εύκολα αν σκεφτούμε ότι γύρω από ελάχιστο κάθε συνάρτηση συμπεριφέρεται σαν τετραγωνική. Επομένως, βλέπουμε ότι το σχετικό μας σφάλμα ($|(x - x_{\epsilon\lambda})/x_{\epsilon\lambda}|$) είναι της τάξης της τετραγωνικής ρίζας του ϵ (10^{-4} και 10^{-8} για μονή και διπλή ακρίβεια αντίστοιχα).

Μέχρι στιγμής είμαστε τελείως αόριστοι σχετικά με τις ακριβείς θέσεις των κάθε φορά ενδιάμεσων σημείων b και d. Θα εξηγήσουμε τώρα ότι υπάρχει ένας ενδεδειγμένος τρόπος για αυτές τις επιλογές. Θεωρούμε τα σημεία $a < b < d < c$, όπως στο Σχ.(7.3). Το ελάχιστο βρίσκεται κάπου μέσα στο διάστημα $[a, c]$. Ανάλογα με το πρόσημο της παράστασης $f(d) - f(b)$, το νέο διάστημα, μέσα στο οποίο θα εντοπίζεται το ελάχιστο θα είναι είτε το $[b, c]$ είτε το $[a, d]$. Επιλέγουμε αυτά τα δύο διαστήματα να είναι ίσα:

$$d - a = c - b$$

έτσι ώστε η ταχύτητα σύγκλισης της μεθόδου να μην εξαρτάται από τη συμπεριφορά της συνάρτησης. Εισάγουμε επί πλέον τους συμβολισμούς:

$$w \equiv \frac{b-a}{c-a}, \quad z \equiv \frac{d-b}{c-a}.$$



Σχήμα 7.3: Τα διαστήματα (a, d) και (b, c).

Η σχέση $d-a=c-b$ συνεπάγεται διαδοχικά:

$$\begin{aligned}(d-b)+(b-a) &= (c-a)-(b-a)\\ z\cdot(c-a)+w\cdot(c-a) &= (c-a)-w(c-a)\\ z+w &= 1-w,\end{aligned}$$

δηλαδή

$$z=1-2w.$$

Παρατηρούμε ότι, ενώ το w είναι μια θετική ποσότητα, δεν ισχύει το ίδιο και για το z. Η περίπτωση το z να είναι αρνητικό αντιστοιχεί στη δυνατότητα το νέο σημείο d να βρίσκεται αριστερά του b αντί δεξιά του. Η σχέση $d-a=c-b$ συνεπάγεται επίσης ότι: $d-b=a+c-2b$. Αρα το $d-b$ είναι θετικό όταν $b<\frac{a+c}{2}$. Συνεπώς, όταν το $[b,c]$ είναι το μεγαλύτερο από τα δύο διαστήματα $[a,b]$ και $[b,c]$ στα οποία χωρίζεται το διάστημα $[a,c]$, το νέο σημείο d θα βρίσκεται δεξιά του b, δηλαδή θα περιλαμβάνεται σ' αυτό το μεγαλύτερο διάστημα $[b,c]$. Εξ άλλου, το $d-b$ είναι αρνητικό όταν $b>\frac{a+c}{2}$, δηλαδή όταν το μεγαλύτερο διάστημα είναι το $[a,b]$. Αρα σ' αυτήν την περίπτωση το d βρίσκεται αριστερά του b, δηλαδή και πάλι βρίσκεται μέσα στο μεγαλύτερο διάστημα. Μπορούμε επομένως να διατυπώσουμε τον κανόνα ότι, **με δεδομένη την αρχική τριάδα των σημείων** $a<b<c$, **το τέταρτο σημείο** d **θα βρίσκεται πάντοτε στο μεγαλύτερο μεταξύ των δύο διαστημάτων** $[a,b]$ **και** $[b,c]$. Ας υποθέσουμε τώρα ότι το μεγαλύτερο διάστημα είναι το $[b,c]$. Το d θα βρίσκεται μέσα σ' αυτό. Έστω ότι η επόμενη τριάδα, μετά την $a<b<c$, είναι η $b<d<c$. Απαιτούμε το d να χωρίζει το διάστημα $[b,c]$ κατά το ίδιο πηλίκο όπως κάνει το b με το διάστημα $[a,c]$, δηλαδή να ισχύει η ισότητα

$$\frac{d-b}{c-b}=\frac{b-a}{c-a}\to\frac{\frac{d-b}{c-a}}{\frac{c-b}{c-a}}=\frac{b-a}{c-a}\to\frac{z}{1-w}=w.$$

Ο συνδυασμός της τελευταίας σχέσης με την $z=1-2w$ που παρουσιάσαμε προηγουμένως δίνει την εξίσωση

$$w^2-3w+1=0,$$

που δίνει τη λύση $w=\frac{3-\sqrt{5}}{2}\approx 0.38$ Αυτός ο αριθμός δεν είναι άλλος από την περίφημη χρυσή τομή, που είναι γνωστή από τους Αρχαίους Ελληνες. Θα διατυπώσουμε τον τελικό κανόνα για το πώς προσδιορίζεται το τέταρτο σημείο: στην περίπτωση που το μεγαλύτερο από τα δύο διαστήματα είναι το $[b,c]$, το τέταρτο σημείο d θα βρίσκεται μέσα στο διάστημα αυτό και θα απέχει από το μεσαίο σημείο b της τριάδας $a<b<c$ απόσταση ίση με το 0.38 επί το μήκος $c-b$ του διαστήματος αυτού. Στην περίπτωση που το μεγαλύτερο διάστημα είναι το $[a,b]$, το d θα βρίσκεται αριστερά του b, δηλαδή μέσα στο διάστημα

$[a, b]$ και θα απέχει από το b απόσταση ίση με $0.38(b-a)$.

Παράδειγμα: Θεωρούμε τη συνάρτηση

$$f(x) = \frac{1}{20}x^2 - \frac{2}{5}x + 1$$

η οποία έχει ένα (μοναδικό) ελάχιστο στο $x = 4$. Θα κάνουμε μερικά βήματα της μεθόδου, ώστε να δούμε πως θα προσεγγίσουμε αυτό το ελάχιστο με αριθμητικές μεθόδους, όπως αναγκαστικά θα κάνουμε σε άλλες περιπτώσεις, όπου η αναλυτική αντιμετώπιση δεν είναι δυνατή. Όσα θα πούμε περιέχονται συνοπτικά στον παρακάτω πίνακα. Έστω ότι ξεκινάμε από την αρχική τριάδα $a = 0, b = 2, c = 8$. Μπορεί κανείς να ελέγξει ότι το $f(2)$ είναι πράγματι μικρότερο από τα $f(0)$ και $f(8)$, επομένως το ελάχιστο πρέπει να βρίσκεται μεταξύ του 0 και του 8. από τα δύο διαστήματα στα οποία χωρίζεται το $[0, 8]$ μεγαλύτερο είναι το $[2, 8]$ με μήκος 6. Το σημείο d θα βρίσκεται δεξιά του 2 σε απόσταση $0.38 \cdot 6$, θα είναι δηλαδή το σημείο $2 + 0.38 \cdot 6 = 4.28$. Μπορεί να ελεγχθεί ότι $f(4.28) < f(2)$, οπότε η νέα τριάδα θα είναι η $a = 2, b = 4.28, c = 8$. Το μεγαλύτερο διάστημα είναι τώρα το $[4.28, 8]$, άρα το καινούργιο σημείο d θα βρίσκεται στη θέση $4.28 + 0.38 \cdot (8 - 4.28) = 5.6936$. Η σχέση $f(5.6936) > f(4.28)$ υποδεικνύει ως νέα τριάδα την $a = 2, b = 4.28, c = 5.6936$. Το μεγαλύτερο διάστημα είναι το $[2, 4.28]$, οπότε $d = 4.28 - 0.38 \cdot (4.28 - 2) = 3.4136$. Το νέο στοιχείο εδώ είναι ότι το σημείο d βρίσκεται **αριστερά** του μεσαίου σημείου, οπότε έχουμε την εμφάνιση ενός αρνητικού προσήμου πριν από τον παράγοντα 0.38. Η σχέση $f(3.4136) > f(4.28)$ δείχνει ότι η νέα τριάδα είναι η $a = 3.4136, b = 4.28, c = 5.6936$. Μέχρι στιγμής έχουμε περιορίσει το διάστημα που περιέχει το ελάχιστο από το $[0, 8]$, που ήταν στην αρχή, στο μικρότερο $[3.4136, 5.6936]$. Η διαδικασία μπορεί να συνεχιστεί με τον ίδιο τρόπο. Μια καλή άσκηση είναι να επαληθεύσει ο αναγνώστης ότι οι τριάδες είναι οι:

a	b	c	d	$f(a)$	$f(b)$	$f(c)$	$f(d)$	νέα τριάδα
0	2	8	4.28	1	0.4	1	0.204	(b, d, c)
2	4.28	8	5.6936	0.4	0.204	1	0.343	(a, b, d)
2	4.28	5.6936	3.4136	0.4	0.204	0.343	0.217	(d, b, c)
3.4136	4.28	5.6936	4.8172	0.217	0.204	0.343	0.233	(a, b, d)
3.4136	4.28	4.8172	3.9508	0.217	0.204	0.233	0.2001	(a, d, b)
3.4136	3.9508	4.28	3.7467	0.217	0.2001	0.204	0.203	(d, b, c)
3.7467	3.9508	4.28	4.0759	0.203	0.2001	0.204	0.2003	(a, b, d)
3.7467	3.9508	4.0759						

Η τελευταία τριάδα εντοπίζει το ελάχιστο στο διάστημα $[3.74665, 4.07588]$, που αντιπροσωπεύει μια σαφή πρόοδο σε σχέση με το αρχικό διάστημα $[0, 8]$. Το διάστημα αυτό ξέρουμε ότι πράγματι περικλείει το ελάχιστο που εμφανίζεται στο $x = 4$.

7.2 Πολυδιάστατη Ελαχιστοποίηση

Η εύρεση του ελαχίστου μιας συνάρτησης N μεταβλητών είναι ένα πολύ δύσκολο πρόβλημα. Αν και έχουν αναπτυχθεί διάφορες μέθοδοι για τη επίλυσή του, καμμιά δεν είναι πλήρως ικανοποιητική. Θα περιγράψουμε μια από τις δημοφιλέστερες, που ονομάζεται **μέθοδος των συζυγών διευθύνσεων (conjugate gradient method).** Μια αρχική ιδέα που θα μπορούσε να έχει κανείς για την αντιμετώπιση του προβλήματος (για τη σχετικά απλή περίπτωση μιας συνάρτησης δύο μεταβλητών $f(x,y)$, αλλά τα επιχειρήματα γενικεύονται για οποιονδήποτε αριθμό διαστάσεων) θα ήταν η εξής: να ξεκινήσει κανείς από κάποιο αρχικό σημείο (x_0,y_0) και να ελαχιστοποιήσει τη συνάρτηση κατά μήκος του ενός άξονα, έστω του άξονα των x, δηλαδή να βρεί το σημείο x_1 που ελαχιστοποιεί τη συνάρτηση μιας μεταβλητής $f(x,y_0)$. Στη συνέχεια να σταθεροποιηθεί το x στην τιμή x_1 και ξεκινώντας από το αρχικό σημείο (x_1,y_0) να βρει το ελάχιστο κατά μήκος του άλλου άξονα, δηλαδή να βρεί το σημείο y_1 που ελαχιστοποιεί τη συνάρτηση μιας μεταβλητής $f(x_1,y)$. Αφού ελαχιστοποιήθηκε η συνάρτηση ως προς και τους δύο άξονες θα περίμενε κανείς ότι βρέθηκε το ελάχιστο ως προς και τις δύο μεταβλητές.

Τα πράγματα όμως δεν είναι τόσο απλά. Το πρόβλημα είναι ότι όταν επιχειρείται η δεύτερη ελαχιστοποίηση (ως προς y), επηρεάζεται πολύ η ελαχιστοποίηση ως προς x που υποτίθεται ότι έχει ήδη ολοκληρωθεί στο προηγούμενο στάδιο. Το αποτέλεσμα είναι ότι, αν και η θέση (x_1,y_0) είναι θέση ελαχίστου ως προς x, η θέση (x_1,y_1) δεν αντιπροσωπεύει ελάχιστο ως προς x. Δηλαδή οι δύο κατευθύνσεις γενικά αλληλοεπηρεάζονται κι αυτό είναι το πολύ σοβαρό πρόβλημα που πρέπει να αντιμετωπιστεί. Αναρωτιέται κανείς αν υπάρχει ένα σύνολο διευθύνσεων στο N-διάστατο χώρο των ανεξάρτητων μεταβλητών, ώστε να αποφεύγεται το πρόβλημα αυτό. Πιο συγκεκριμένα, πρέπει η προβολή της βαθμίδας (gradient) της συνάρτησης $f(x_1,x_2,\ldots,x_N)$ κατά μήκος μιας κατεύθυνσης που την εκφράζει ένα διάνυσμα $\mathbf{u}$ να παραμένει μηδέν, καθώς το σημείο $(x_1,x_2,\ldots,x_N)$ κινείται κατά μήκος της άλλης κατεύθυνσης που εκφράζεται από το διάνυσμα $\mathbf{v}$. Διευθύνσεις που ικανοποιούν αυτή τη συνθήκη λέγονται συζυγείς διευθύνσεις. Η σημασία τους έγκειται στο ότι η ελαχιστοποίηση κατά μήκος της μιας δεν επηρεάζει την αντίστοιχη διαδικασία κατά μήκος της άλλης, οπότε μπορεί να ακολουθηθεί η διαδικασία που περιγράψαμε προηγουμένως. Σ' αυτήν την περίπτωση το ελάχιστο θα μπορούσε να βρεθεί σε N βήματα για N-διάστατο χώρο (μέχρι να εξαντληθούν όλες οι ανεξάρτητες διευθύνσεις).

Πρέπει λοιπόν να βρεθούν N ανεξάρτητες συζυγείς διευθύνσεις. Το πρόβλημα μπορεί να λυθεί ακριβώς όταν η συνάρτηση είναι το πολύ δεύτερου βαθμού ως προς τις ανεξάρτητες μεταβλητές:

$$f(x_1,x_2,\ldots,x_N) = f_0 + \sum_{i=1}^{N} c_i x_i + \frac{1}{2}\sum_{k,l=1}^{N} x_k H_{kl} x_l, \qquad (7.1)$$

όπου τα f_0, c_i και H_{kl} είναι σταθερές. Ο πίνακας H_{kl} είναι συμμετρικός και θετικός (δηλαδή έχει θετικές ιδιοτιμές). Αν η συνάρτηση δεν είναι αυτής της μορφής, μπορεί υπό ορισμένες συνθήκες να προσεγγιστεί κοντά σ' ένα ελάχιστο από μια τέτοια συνάρτηση και οι διευθύνσεις που θα προκύψουν θα είναι κατά προσέγγιση συζυγείς. Αν τις χρησιμοποιήσει κανείς, η σύγκλιση θα βελτιωθεί θεαματικά και γι' αυτές τις συναρτήσεις. Όπως είπαμε προηγουμένως, ένα βασικό στοιχείο στον ορισμό των συζυγών διευθύνσεων είναι η βαθμίδα. Για την τετραγωνική συνάρτηση (7.1) η βαθμίδα υπολογίζεται εύκολα:

$$\nabla f = H\,\mathbf{r} + \mathbf{c},$$

που σημαίνει σε συνιστώσες ότι: $(\nabla f)_k = \sum_l H_{kl} x_l + c_k$, όπου χρησιμοποιήσαμε το συμβολισμό: $\mathbf{r} \equiv (x_1, x_2, \ldots, x_N)$. Η μεταβολή της βαθμίδας για μια αλλαγή $\delta\mathbf{r}$ του διανύσματος θέσης $\mathbf{r}$ είναι:

$$\delta(\nabla f) = H\,\delta\mathbf{r}.$$

Έστω ότι επιχειρείται ελαχιστοποίηση κατά μήκος της κατεύθυνσης που χαρακτηρίζεται από το διάνυσμα $\mathbf{v}$, δηλαδή ότι το $\delta\mathbf{r}$ είναι παράλληλο προς το $\mathbf{v}$. Η μεταβολή της βαθμίδας κατά τη διαδικασία αυτή είναι πολλαπλάσιο του διανύσματος $H \cdot \mathbf{v}$ και η απαίτηση η $\mathbf{v}$ να είναι συζυγής προς κάποια κατεύθυνση $\mathbf{u}$, ως προς την οποία είχε γίνει ενδεχομένως κάποια προηγούμενη ελαχιστοποίηση, είναι το $H \cdot \mathbf{v}$ να έχει μηδενική προβολή πάνω στο $\mathbf{u}$:

$$0 = \mathbf{u} \cdot \delta(\nabla f) = \mathbf{u} \cdot H\,\mathbf{v}.$$

Έχουμε λοιπόν παραγάγει μια ακριβή συνθήκη που χαρακτηρίζει δύο κατευθύνσεις ως συζυγείς:

$$\mathbf{u} \cdot H\,\mathbf{v} = 0.$$

Όπως είπαμε και προηγουμένως, μια μέθοδος που θα φαινόταν στην αρχή καλή είναι να γίνει ελαχιστοποίηση κατά μήκος της βαθμίδας σε κάποιο αρχικό σημείο $\mathbf{r}_0$: Η κατεύθυνση της πιο γρήγορης μείωσης της συνάρτησης είναι η $-\nabla f|_{\mathbf{r}_0} \equiv -\nabla f_0$. Άρα βρίσκει κανείς το ελάχιστο της συνάρτησης $f(\mathbf{r}_0 - \lambda \nabla f_{\mathbf{r}_0})$[1]ως προς λ και όταν βρεί το λ_0 πηγαίνει στο αντίστοιχο διάνυσμα θέσης $\mathbf{r}_1 \equiv \mathbf{r}_0 - \lambda \nabla f_0$, υπολογίζει την αντίστοιχη βαθμίδα $-\nabla f_{\mathbf{r}_1} \equiv -\nabla f_1$ σ' αυτή τη θέση και συνεχίζει τη διαδικασία. Ένα σημαντικό στοιχείο αυτής της διαδικασίας είναι το απλό γεγονός ότι:

$$\frac{d}{d\lambda} f(\mathbf{r}_0 - \lambda \nabla f_0) = 0 \rightarrow \nabla f|_{\lambda_0} \cdot (-\nabla f_0) = 0 \rightarrow \nabla f_1 \cdot \nabla f_0 = 0,$$

δηλαδή η βαθμίδα στο σημείο $\mathbf{r}_1$ είναι κάθετη προς τη βαθμίδα στο σημείο $\mathbf{r}_0$. Αυτό σημαίνει ότι οι βαθμίδες είναι ορθογώνιες και όχι συζυγείς, όπως θα έπρεπε προκειμένου η μία ελαχιστοποίηση να μην επηρεάζει την άλλη. Το

[1] Υπενθυμίζεται ότι η ευθεία που περνά από το σημείο $\mathbf{r}_0$ και είναι παράλληλη με την $\nabla f|_{\mathbf{r}_0}$ περιγράφεται από τη σχέση $\mathbf{r}_0 - \lambda \nabla f_{\mathbf{r}_0}$.

αποτέλεσμα αυτής της μεθόδου θα ήταν γενικά μια πληθώρα πολλών μικρών βημάτων ορθογώνιων μεταξύ τους, για τα οποία είναι απρόβλεπτο πού θα καταλήξουν. Αυτός είναι ο λόγος για τον οποίο δεν ακολουθείται αυτή η μέθοδος, αλλά η εξής:

1) Ξεκινάμε από το αρχικό σημείο $\mathbf{r}_0$ και ελαχιστοποιούμε κατά μήκος της γραμμής $\mathbf{r}_0 - \lambda \nabla f_0$, όπως περιγράφτηκε προηγουμένως. Φτάνουμε στο σημείο $\mathbf{r}_1$.

2) Αντί να ελαχιστοποιήσουμε κατά μήκος της βαθμίδας ∇f_1 σ' αυτό το σημείο αναζητούμε μια νέα διεύθυνση $\mathbf{h}$, η οποία θα είναι συζυγής προς την προηγούμενη κατεύθυνση που δεν ήταν άλλη από την $\mathbf{r}_1 - \mathbf{r}_0$. (Το διάνυσμα αυτό είναι παράλληλο προς το ∇f_0, όπως μπορεί κανείς να δει απλά.) Για το $\mathbf{h}$ απαιτούμε να ικανοποιεί την

$$\mathbf{h} \cdot H(\mathbf{r}_1 - \mathbf{r}_0) = 0.$$

3) Αφού $\nabla f = H\mathbf{r} - \mathbf{c}$, βλέπουμε ότι $\mathbf{h} \cdot H(\mathbf{r}_1 - \mathbf{r}_0) = \mathbf{h} \cdot (\nabla f_1 - \nabla f_0) = 0$, οπότε η μέθοδος μπορεί να γενικευτεί και για συναρτήσεις όχι υποχρεωτικά τετραγωνικές.

4) Για να προσδιορίσουμε το $\mathbf{h}$ είναι αρκετό να περιοριστούμε σ' ένα γραμμικό συνδυασμό των ∇f_0 και ∇f_1

$$\mathbf{h} = \nabla f_1 + \mu \nabla f_0.$$

Η συνθήκη $\mathbf{h} \cdot (\nabla f_1 - \nabla f_0) = 0$ δίνει:

$$(\nabla f_1 + \mu \nabla f_0) \cdot (\nabla f_1 - \nabla f_0) = 0 \ \rightarrow \mu = \frac{(\nabla f_1)^2}{(\nabla f_0)^2},$$

όπου χρησιμοποιήσαμε τη σχέση $\nabla f_1 \cdot \nabla f_0 = 0$, που αποδείξαμε προηγουμένως.

5) Αφού προσδιορίστηκε η συζυγής κατεύθυνση $\mathbf{h}$, ελαχιστοποιούμε τη συνάρτηση ως προς λ κατά μήκος της γραμμής $\mathbf{r}_1 - \lambda \mathbf{h}$, βρίσκουμε το $\mathbf{r}_2$ και συνεχίζουμε τη διαδικασία.

Παράδειγμα: Έστω ότι πρέπει να ελαχιστοποιηθεί η συνάρτηση

$$f(x) = x^2 + \frac{1}{2}y^2.$$

Έχουμε δηλαδή ένα διδιάστατο χώρο με συντεταγμένες (x, y) για ένα τυπικό σημείο του. Ένα πρώτο, αναγκαίο, βήμα είναι να υπολογίσουμε τη βαθμίδα (gradient) της συνάρτησης:

$$\nabla f(x, y) = 2x\hat{\mathbf{x}} + y\hat{\mathbf{y}}$$

Επιλέγουμε ως αρχικό σημείο το $\mathbf{r}_0 = (1, 1) \rightarrow \nabla f|_0 = (2, 1)$. Ελαχιστοποιούμε ως προς λ κατά μήκος της γραμμής $\mathbf{r} = \mathbf{r}_0 - \lambda \nabla f|_0 = (1 - 2\lambda, 1 - \lambda)$ τη συνάρτηση $f_\lambda = (1 - 2\lambda)^2 + \frac{1}{2}(1 - \lambda)^2$. Είναι εύκολο να δει κανείς ότι

$$\frac{df_\lambda}{d\lambda} = -\nabla f|_0 \cdot \nabla f(\mathbf{r})$$

Άρα $-\nabla f|_0 \cdot (2(1-2\lambda)\hat{x}+(1-\lambda)\hat{y} = 0 \rightarrow 2(2-4\lambda)+(1-\lambda) = 0 \rightarrow \lambda_0 = \frac{5}{9}$, που δίνει με τη σειρά του τα αποτελέσματα

$$\mathbf{r}_1 = (-\frac{1}{9}, \frac{4}{9}), \quad \nabla f|_1 = (-\frac{2}{9}, \frac{4}{9})$$

Έχουμε τα εφόδια για να υπολογίσουμε την ποσότητα μ_1 : $\mu_1 = \frac{(\nabla f|_1)^2}{(\nabla f|_0)^2} = \frac{4}{81}$, οπότε η συζυγής κατεύθυνση, κατά μήκος της οποίας θα γίνει η επόμενη ελαχιστοποίηση είναι η $\mathbf{h}_1 = \nabla f|_1 + \mu_1 \nabla f|_0 = \left(-\frac{10}{81}, \frac{40}{81}\right)$. Κινούμαστε, λοιπόν, πάνω στην ευθεία $\mathbf{r} = \mathbf{r}_1 - \lambda \mathbf{h}_1 = \left(\frac{10\lambda-9}{81}, \frac{36-40\lambda}{81}\right)$. Η $\frac{df}{d\lambda} = 0$ δίνει

$$\left(\frac{2}{9}, -\frac{4}{9}\right)\left(2\frac{10\lambda-9}{81}, \frac{36-40\lambda}{81}\right) = 0 \rightarrow \lambda_1 = \frac{9}{10},$$

οπότε $\mathbf{r}_2 = (0,0)$ και $\nabla f|_2 = (0,0)$. Παρατηρούμε ότι η τελική βαθμίδα είναι μηδέν, οπότε η μέθοδος έχει βρει το ελάχιστο με ελαχιστοποίηση κατά μήκος δύο κατευθύνσεων. Αυτό είναι γενικό χαρακτηριστικό: Αν είχαμε εξάρτηση από n μεταβλητές θα χρειάζονταν n ελαχιστοποιήσεις.

Κεφάλαιο 8

Εισαγωγή στις Στατιστικές Μεθόδους Φυσικής

8.1 Πειραματικές Μετρήσεις και Αβεβαιότητα

Το αποτέλεσμα ενός πειράματος που μετρά κάποια παράμετρο x δίνεται σε απλές περιπτώσεις, συνήθως ως εξής:

$$x = a \pm \sigma$$

Το a είναι η πιο πιθανή τιμή και το σ καθορίζει την αβεβαιότητα στη μέτρηση (λέγεται σφάλμα της μέτρησης). Αν η κατανομή πιθανότητας της μέτρησης υποτεθεί ότι είναι Γκαουσιανή κατανομή, με τυπική απόκλιση σ, τότε η ολική πιθανότητα να βρίσκεται η πραγματική τιμή μέσα στο διάστημα $(a-\sigma, a+\sigma)$ είναι 68 %. Αν το σφάλμα δεν είναι 'συμμετρικό' γράφουμε

$$x = a^{+\sigma_1}_{-\sigma_1}$$

8.2 Τυχαίες και Συστηματικές Αβεβαιότητες

Η τυχαία (στατιστική) αβεβαιότητα οφείλεται στην ενδογενή τυχαιότητα της διαδικασίας που ακολουθείται στη μέτρηση. Η συστηματική αβεβαιότητα οφείλεται στην αβεβαιότητα της συμπεριφοράς της πειραματικής διάταξης.
Παράδειγμα: Μέτρηση της ενεργότητας ραδιενεργού πηγής.
Μετριέται ο αριθμός n των σωματιδίων μέσω ενός ανιχνευτή που καλύπτει

στερεά γωνία Ω. Ο ανιχνευτής έχει απόδοση ε. Η μέτρηση γίνεται σε χρονικό διάστημα (χρόνο) t_a. Η στατιστική αβεβαιότητα είναι η αβεβαιότητα στο n (για μεγάλα n, η κατανομή πιθανότητας του στατιστικού σφάλματος, που γενικά ακολουθεί την κατανομή του Poisson, γίνεται γκαουσιανή οπότε $\sigma = \sqrt{n}$). Η συστηματική αβεβαιότητα οφείλεται στο ότι τα t_a, Ω και ε δεν είναι γνωστά με απόλυτη ακρίβεια.

Όταν ταυτόχρονα υπάρχουν τυχαία και συστηματικά σφάλματα, το αποτέλεσμα πολλές φορές δίνεται ως

$$x \pm \sigma^{stat} \pm \sigma^{sys}$$

όπου σ^{stat} είναι το τυχαίο (στατιστικό) σφάλμα και σ^{sys} το συστηματικό. Το ολικό σφάλμα είναι

$$\sigma = \sqrt{(\sigma^{stat})^2 + (\sigma^{sys})^2}$$

Σε όσα ακολουθούν θα δούμε στοιχεία από πιθανότητες και στατιστική, που σχετίζονται με την εξαγωγή συμπερασμάτων από πειραματικές μετρήσεις.

8.3 Τυχαίες μεταβλητές. Δειγματικός Χώρος

Οι τυχαίες ή στοχαστικές ή στατιστικές μεταβλητές σχετίζονται με πειράματα τύχης στα οποία δεν είναι δυνατόν να προβλεφθεί το αποτέλεσμα (η τιμή της τυχαίας μεταβλητής). Υποτίθεται ότι το πείραμα μπορεί να επαναληφθεί πολλές φορές υπό τις ίδιες συνθήκες.

Τα δυνατά (πιθανά) αποτελέσματα ενός πειράματος τύχης λέγονται απλά ή βασικά ενδεχόμενα. Το σύνολο των βασικών ενδεχομένων λέγεται δειγματικός χώρος. Κάθε υποσύνολο του δειγματικού χώρου λέγεται ενδεχόμενο ή συμβάν ή γεγονός. Αν ρίξουμε ένα νόμισμα και ενδιαφερόμαστε για το αποτέλεσμα κορώνα ή γράμματα, τότε τα απλά ενδεχόμενα είναι δύο: K (κορώνα) και Γ (γράμματα). Ο δειγματικός χώρος είναι $\Delta = \{K, \Gamma\}$, δηλαδή περιέχει δύο στοιχεία. Τυχαία ή στατιστική μεταβλητή είναι κάθε πραγματική συνάρτηση με πεδίο ορισμού ένα δειγματικό χώρο. Μια τυχαία μεταβλητή μπορεί να λαμβάνει είτε διακριτές τιμές, οπότε είναι ασυνεχής, είτε συνεχείς τιμές, οπότε είναι συνεχής. Αν υποθέσουμε ότι ρίχνουμε ένα ζάρι και ενδιαφερόμαστε για την ένδειξή του, τότε έχουμε το δειγματικό χώρο $\Delta = \{1, 2, 3, 4, 5, 6\}$, ενώ το πεδίο ορισμού είναι οι τιμές 1,...,6. Μπορούμε να ορίσουμε μια τυχαία συνάρτηση με την αντιστοιχία

$$1 \to 0, \quad 2 \to 0, \quad 3 \to 0, \quad 4 \to 0, \quad 5 \to 0, \quad 6 \to 1,$$

έχουμε δηλαδή την τυχαία μεταβλητή Y=πλήθος εμφανίσεων της ένδειξης 6, η οποία παίρνει τις τιμές 0 και 1.

Αν η τυχαία μεταβλητή X μπορεί να πάρει τις διακριτές τιμές x_i, $i = 1, ..., N$, τότε σε κάθε τιμή x_i αντιστοιχεί μια πιθανότητα P_i,

$$P(X = x_i) = P_i,$$

Πρέπει το άθροισμα των πιθανοτήτων όλων των δυνατών αποτελεσμάτων (τιμών x_i) να ισούται με 1:

$$\sum_{i=1}^{N} P_i = 1$$

Τα γεγονότα x_i λέγονται αποκλειστικά ή ξένα ή ασυμβίβαστα αν η πραγματοποίηση του ενός αποκλείει τό άλλο. Αν τα γεγονότα είναι ξένα και ισχύει η ανωτέρω σχέση κανονικοποίησης, τότε λέμε ότι τα γεγονότα έχουν πληρότητα (είναι πλήρη).

Αν η τυχαία μεταβλητή X παίρνει συνεχείς τιμές x, τότε ορίζεται η συνάρτηση πυκνότητας πιθανότητας $f(x)$

$$f(x)dx = dP(x \leq X \leq x + dx),$$

όπου dP είναι η (στοιχειώδης) πιθανότητα η μεταβλητή X να έχει τιμές μεταξύ x και $x+dx$. Είναι ευνόητο ότι αν 'αθροίσουμε' όλες τις πιθανότητες θα έχουμε

$$\int_{\Omega} f(x)\, dx = 1$$

όπου Ω είναι ο δειγματικός χώρος που καθορίζεται από όλες τις δυνατές τιμές του X (αποτελέσματα πειράματος τύχης, μέτρησης της X). Πολλές φορές δε θα κάνουμε διάκριση μεταξύ X και x και θα χρησιμοποιούμε μόνο το σύμβολο x.

Συνάρτηση ολικής κατανομής πιθανότητας. Ως συνάρτηση ολικής κατανομής πιθανότητας ορίζεται η συνάρτηση

$$F(x) = \int_{x_{min}}^{x} f(x')\, dx', \quad x_{min} \leq x \leq x_{max}$$

με ευνόητες σχέσεις τις $F(x_{min}) = 0$ και $F(x_{max}) = 1$. Οι στατιστικολόγοι προτιμούν πολλές φορές τον αντίστροφο ορισμό, ξεκινώντας από την $F(x)$ που ονομάζουν συνάρτηση κατανομής. Στη συνέχεια η $f(x)$ ορίζεται από την $F(x)$ με τη σχέση

$$f(x) = \frac{dF(x)}{dx}$$

Είναι ευνόητο ότι απλή περίπτωση τυχαίας μεταβλητής είναι η συνάρτηση (στο δειγματικό χώρο) $x_i \to x_i$ ή, για συνεχείς τιμές, $x \to x$. Κάθε συνάρτηση τυχαίας μεταβλητής είναι επίσης τυχαία μεταβλητή.

Χαρακτηριστικά του νόμου πιθανοτήτων. Η $f(x)$ ορίζει μια θεωρητική κατανομή της μεταβλητής X που χαρακτηρίζεται από θέση και διασπορά. Ορίζεται

ως μαθηματική αναμενόμενη τιμή (ή αναμενόμενη τιμή ή ελπίδα) μιας συνάρτησης $g(x)$ η ποσότητα

$$E[g(x)] = \int_\Omega g(x) f(x)\, dx$$

όπου $f(x)$ η πυκνότητα πιθανότητας της x. Ω είναι (όλη) η περιοχή ορισμού της τυχαίας μεταβλητής x. Η αναμενόμενη τιμή $E[g(x)]$ είναι η μέση τιμή ή κεντρική τιμή της $g(x)$ με βάρος $f(x)$, είναι δηλαδή ο βαρυκεντρικός μέσος ή βαρυκεντρική μέση τιμή. Εύκολα, από τον ορισμό, προκύπτουν οι σχέσεις που ακολουθούν

$$E[a] = a, \quad E[ag(x)] = aE[g(x)],$$
$$E[a_1 g_1(x) + a_2 g_2(x)] = a_1 E[g_1(x)] + a_2 E[g_2(x)]$$

όπου a_1 και a_2 είναι σταθερές. Από τα ανωτέρω προκύπτει ότι η 'πράξη' E είναι γραμμικός τελεστής. Το μέγεθος αυτό χαρακτηρίζει τη 'θέση' της $g(x)$ ως προς τη μεταβλητή x.

Ορίζεται ως διασπορά ή διακύμανση της συνάρτησης $g(x)$ της μεταβλητής x, και παριστάνεται με $V[g(x)]$ ή σ_g^2 ή $\Delta[g(x)]$, η ποσότητα

$$V[g(x)] = \sigma_g^2 = \int_\Omega \left(g(x) - E[g(x)]\right)^2 f(x)\, dx$$

Το μέγεθος αυτό χαρακτηρίζει τη διασπορά της $g(x)$ περί την αναμενόμενη τιμή της, $E[g(x)]$. Η $V[g(x)]$ είναι η αναμενόμενη τιμή της $(g(x) - E[g(x)])^2$. Είναι ευνόητο ότι αν οι μεταβλητές δεν είναι συνεχείς, τα ολοκληρώματα αντικαθίστανται με αντίστοιχα αθροίσματα, όπως:

$$E[g(x)] = \sum_{i=1}^{N} g(x_i) P(x_i)$$
$$V[g(x)] = \sum_{i=1}^{N} \left(g(x_i) - E[g(x_i)]\right)^2 P(x_i)$$

Σημειώνουμε ότι για τη διακύμανση ισχύει

$$V[ag(x)] = a^2 V[g(x)]$$

8.4 Μέση Τιμή και Διασπορά Τυχαίας Μεταβλητής

Αν θέσουμε $g(x) = x$ στις ανωτέρω σχέσεις έχουμε την αναμενόμενη τιμή (ή ελπίδα) και την απόκλιση ή διασπορά της μεταβλητής x με πυκνότητα πιθανότητας $f(x)$. Η ποσότητα

$$\mu = E[x] = \int_\Omega x f(x)\, dx$$

ονομάζεται μέση τιμή ή μέσος της x με πυκνότητα πιθανότητας $f(x)$. Η διασπορά ή απόκλιση της x είναι η

$$\sigma^2 = V[x] = E[(x-\mu^2)] = \int_\Omega (x-\mu)^2 f(x)\, dx$$

Το σ λέγεται τυπική απόκλιση της x ($\sigma = \sqrt{\sigma^2} = \sqrt{V[x]}$).

Από τη γραμμικότητα του τελεστή E προκύπτει ότι:

$$\begin{aligned}\sigma^2 = E[(x-\mu)^2] = E[x^2 - 2x\mu + \mu^2] = E[x^2] - 2\mu E[x] + \mu^2 = \\ = E[x^2] - 2\mu\mu + \mu^2 = E[x^2] - \mu^2 \Rightarrow \\ \Rightarrow E[x^2] = \sigma^2 + \mu^2 = E[x^2] - (E[x])^2\end{aligned}$$

8.5 Κατανομές Πολλών Τυχαίων Μεταβλητών

Οι n τυχαίες μεταβλητές $x_1, x_2, ..., x_n$ χαρακτηρίζονται από τη συνδυασμένη συνάρτηση πυκνότητας πιθανότητας $f(x_1, x_2, ..., x_n)$. Εδώ έχουμε σημεία $(x_1, x_2, ..., x_n)$ σε χώρο n διαστάσεων. Η $f(x_1, x_2, ..., x_n)$ είναι κανονικοποιημένη:

$$\int_\Omega f(x_1, x_2, ..., x_n)\, dx_1 dx_2 ... dx_n = 1$$

Η αναμενόμενη τιμή και η διακύμανση συνάρτησης των n τυχαίων μεταβλητών ορίζονται ως εξής:

$$\begin{aligned}E\left[g(x_1, ..., x_n)\right] &= \int_\Omega g(x_1, ..., x_n) f(x_1, ..., x_n)\, dx_1 ... dx_n \\ V\left[g(x_1, ..., x_n)\right] &= E\left[\left(g(x_1, ..., x_n) - E[g(x_1, ..., x_n)]\right)^2\right] = \\ &= \int_\Omega \left(g(x_1, ..., x_n) - E[g(x_1, ..., x_n)]\right)^2 f(x_1, ..., x_n)\, dx_1 ... dx_n\end{aligned} \tag{8.1}$$

8.6 Συναλλοίωτη Μήτρα. Συντελεστής Συσχέτισης

Για κάθε μια από τις τυχαίες μεταβλητές x_i, έχουμε τη μέση τιμή

$$\mu_i = E[x_i] = \int_\Omega x_i f(x_1, ..., x_n)\, dx_1 ... dx_n$$

Η γενίκευση, για πολλές μεταβλητές, της διασποράς που ορίστηκε για μια μεταβλητή οδηγεί στη συναλλοίωτη μήτρα $V(x_1, ..., x_n)$ της οποίας τα στοιχεία είναι:

$$V_{ij} = E[(x_i - \mu_i)(x_j - \mu_j)] = \int_\Omega (x_i - \mu_i)(x_j - \mu_j) f(x_1, ..., x_n)\, dx_1 ... dx_n$$

όπου μ_i, μ_j είναι οι αναμενόμενες τιμές των x_i, x_j. Ισχύουν τα εξής,

1. Η $V(x_1,...,x_n)$ είναι συμμετρική, $V_{ij} = V_{ji}$.

2. Το διαγώνιο στοιχείο V_{ii} λέγεται διακύμανση σ_i^2 της x_i. Το σ_i^2 είναι μη αρνητικό,

$$\sigma_i^2 = V_{ii} = E[(x_i - \mu_i)^2] = \int_\Omega (x_i - \mu_i)^2 f(x_1...x_n)\, dx_1...dx_n$$

και ισχύει $\sigma_i^2 = V_{ii} = E[x_i^2] - (E[x_i])^2$.

3. Το μη διαγώνιο στοιχείο V_{ij} όπου $i \neq j$, λέγεται συναλλοίωτο των x_i και x_j και συμβολίζεται με $\mathrm{cov}(x_i, x_j)$,

$$\mathrm{cov}(x_i, x_j) = V_{ij} = E[x_i x_j] - E[x_i]E[x_j]$$

Ως μέτρο της συσχέτισης μεταξύ δύο μεταβλητών x_i και x_j χρησιμοποιείται συχνά ο *συντελεστής συσχέτισης* $\varrho(x_i, x_j)$, ο οποίος ορίζεται ως

$$\varrho(x_i, x_j) = \frac{V_{ij}}{(V_{ii}V_{jj})^{1/2}} = \frac{\mathrm{cov}(x_i, x_j)}{\sigma_i \sigma_j}$$

Αποδεικνύεται ότι ισχύει $-1 \leq \varrho(x_i, x_j) \leq +1$. Αν $\varrho(x_i, x_j) = +1$ ή $\varrho(x_i, x_j) = -1$ τότε λέμε ότι οι δύο τυχαίες μεταβλητές x_i και x_j είναι πλήρως θετικά ή πλήρως αρνητικά συσχετισμένες. Αν $\varrho(x_i, x_j) = 0$, οι μεταβλητές είναι μη συσχετισμένες.

8.7 Ανεξάρτητες Μεταβλητές

Οι μεταβλητές $x_1, ..., x_n$ είναι μεταξύ τους ανεξάρτητες αν η συνδυασμένη συνάρτηση πυκνότητας πιθανότητάς τους παραγοντοποιείται πλήρως, δηλαδή,

$$f(x_1, ..., x_n) = f_1(x_1)...f_n(x_n)$$

(οι $f_i(x_i)$ θεωρούνται κανονικοποιημένες). Το συναλλοίωτο στοιχείο V_{ij} $(i \neq j)$, και άρα ο συντελεστής συσχέτισης ανεξαρτήτων μεταβλητών, μηδενίζεται. Η απόδειξη είναι εύκολη από τους σχετικούς ορισμούς. Παρόλο που οι ανεξάρτητες μεταβλητές είναι κατ' ανάγκη μη συσχετισμένες, το αντίστροφο δεν είναι πάντα αληθές, δηλαδή μη συσχετισμένες μεταβλητές δεν είναι πάντα ανεξάρτητες. Μπορεί δηλαδή να ισχύει ότι ο συντελεστής συσχέτισης $\varrho(x_i, x_j) = 0$, αλλά η $f(x_1, x_2)$ να μην παραγοντοποιείται, άρα οι x_1 και x_2 να είναι εξαρτημένες. Δηλαδή η ανεξαρτησία είναι πιο δυνατός ισχυρισμός από το μη συσχετισμό $(\rho(x_i, x_j) = 0)$.

Αν έχουμε μια συνάρτηση δύο μεταβλητών, η οποία μπορεί να παραγοντοποιηθεί ως εξής

$$g(x_i, x_j) = u(x_i)v(x_j)$$

ισχύει ότι, *αν οι x_i και x_j είναι μεταξύ τους ανεξάρτητες, τότε η αναμενόμενη τιμή του γινομένου $u \cdot v$ είναι ίση με το γινόμενο των αναμενόμενων τιμών των u και v.* Πράγματι,

$$E[g(x_i, x_j)] = \int u(x_i) f_i(x_i)\, dx_i \cdot \int v(x_j) f_j(x_j)\, dx_j = E[u(x_i)]E[v(x_j)]$$

Ισχύει ότι η αναμενόμενη τιμή γινομένου ανεξάρτητων μεταβλητών ισούται με το γινόμενο των αναμενόμενων τιμών των επί μέρους μεταβλητών.

8.8 Περιεκτικές Κατανομές και Περιορισμένες Κατανομές

Ας θεωρήσουμε τη $n-$διάστατη συνάρτηση πυκνότητας κατανομής $f(x_1, ..., x_n)$. Αν την ολοκληρώσουμε ως προς όλες τις μεταβλητές εκτός από μία, έστω την x_1, έχουμε την περιεκτική κατανομή αυτής της μεταβλητής, την $h_1(x_1)$,

$$h_1(x_1) = \int_{x_{2min}}^{x_{2max}} \int_{x_{nmin}}^{x_{nmax}} dx_2 \ldots dx_n\, f(x_1, x_2, ..., x_n)$$

Ανάλογα ισχύουν για τις υπόλοιπες μεταβλητές. Αν οι μεταβλητές είναι μεταξύ τους ανεξάρτητες, οπότε η συνάρτηση κατανομής τους παραγοντοποιείται, έχουμε

$$h_1(x_1) = f_1(x_1) \int_{x_{2min}}^{x_{2max}} dx_2 f_2(x_2) \ldots \int_{x_{nmin}}^{x_{nmax}} dx_n\, f_n(x_n)$$

Αφού οι $f_k(x_k)$ είναι, κάθε μια, κανονικοποιημένη, έχουμε

$$h_1(x_1) = f_1(x_1), \quad ..., \quad h_n(x_n) = f_n(x_n)$$

Δηλαδή,οι ανεξάρτητες μεταξύ τους τυχαίες μεταβλητές έχουν συνολική συνάρτηση πυκνότητας πιθανότητας η οποία παραγοντοποιείται στις περιεκτικές συναρτήσεις πυκνότητας.

Η περιορισμένη συνάρτηση πυκνότητας κατανομής για όλες τις μεταβλητές εκτός μιας, έστω της x_1, ορίζεται ως

$$f(x_2, ..., x_n|x_1) = \frac{f(x_1, x_2, ..., x_n)}{h_1(x_1)}$$

Θεωρούμε ότι η x_1 είναι σταθερή και η περιορισμένη (υπό συνθήκη) συνάρτηση κατανομής είναι συνάρτηση των υπολοίπων μεταβλητών $x_2, ..., x_n$.

Ορίζεται η *περιορισμένη αναμενόμενη τιμή* μιας συνάρτησης $u(x_2, ..., x_n|x_1)$, όπου το x_1 θεωρείται σταθερό (δεδομένο), ως εξής:

$$E[u(x_2, ..., x_n|x_1)] = \int_{x_{2min}}^{x_{2max}} \int_{x_{nmin}}^{x_{nmax}} dx_2 \ldots dx_n\, u(x_2, ..., x_n|x_1) f(x_2, ..., x_n|x_1)$$

Προφανώς, ανάλογες σχέσεις μπορούν να γραφτούν για κάθε μια από τις υπόλοιπες μεταβλητές αντί της x_1.

8.9 Γραμμικές Συναρτήσεις Τυχαίων Μεταβλητών

Έστω ότι

$$g(x_1, ..., x_n) = \sum_{i=1}^{n} a_i x_i \tag{8.2}$$

όπου a_i σταθερές. Ισχύει προφανώς

$$\begin{aligned}
E\left[\sum_{i=1}^{n} a_i x_i\right] &= \sum_{i=1}^{n} E[a_i x_i] = \sum_{i=1}^{n} a_i E[x_i] = \sum_{i=1}^{n} a_i \mu_i \qquad (8.3)\\
V\left[\sum_{i=1}^{n} a_i x_i\right] &= E\left[\left(\sum_{i=1}^{n} a_i x_i - E\left[\sum_{i=1}^{n} a_i x_i\right]\right)^2\right]\\
&= E\left[\left(\sum_{i=1}^{n} a_i x_i - \sum_{i=1}^{n} a_i \mu_i\right)^2\right] = E\left[\left(\sum_{i=1}^{n} a_i (x_i - \mu_i)\right)^2\right]\\
&= E\left[\sum_{i=1}^{n} a_i^2 (x_i - \mu_i)^2 + \sum_{i=1}^{n} \sum_{j=1, i\neq j}^{n} a_i a_j (x_i - \mu_i)(x_j - \mu_j)\right]\\
&= \sum_{i=1}^{n} a_i^2 E\left[(x_i - \mu_i)^2\right] + \sum_{i=1}^{n} \sum_{j=1, i\neq j}^{n} a_i a_j E\left[(x_i - \mu_i)(x_j - \mu_j)\right]\\
&= \sum_{i=1}^{n} a_i^2 V_{ii} + \sum_{i=1}^{n} \sum_{j=1, i\neq j}^{n} a_i a_j \, cov(x_i, x_j)
\end{aligned}$$

και τελικά

$$V\left[\sum_{i=1}^{n} a_i x_i\right] = \sum_{i=1}^{n} a_i^2 V_{ii} + 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} a_i a_j V_{ij}$$

Αν οι μεταβλητές είναι μεταξύ τους ασυσχέτιστες, τότε ισχύει η απλή σχέση

$$V\left[\sum_{i=1}^{n} a_i x_i\right] = \sum_{i=1}^{n} a_i^2 V_{ii} \tag{8.4}$$

Παράδειγμα: Αριθμητικός μέσος ανεξαρτήτων μεταβλητών με ίδιο μέσο και απόκλιση

Έστω οι ανεξάρτητες μεταβλητές $x_1, ..., x_n$, όπου $\mu_i = \mu$ και $\sigma_i^2 = \sigma^2$. Ας

σχηματίσουμε το γραμμικό συνδυασμό τους $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ που είναι ο αριθμητικός μέσος τους. Προφανώς είναι ειδική περίπτωση της σχέσης (8.2), όπου $a_i = 1/n$, άρα από τις σχέσεις (8.3) και (8.4) βρίσκουμε για την αναμενόμενη τιμή και την απόκλιση (διασπορά) του $\bar{x}$

$$E[\bar{x}] = \sum_{i=1}^{n} a_i \mu_i = \sum_{i=1}^{n} \frac{1}{n}\mu = \mu$$

$$V[\bar{x}] = \sum_{i=1}^{n} a_i^2 \sigma_i^2 = \sum_{i=1}^{n} \left(\frac{1}{n}\right)^2 \sigma^2 = \frac{\sigma^2}{n}$$

8.10 Αλλαγή Μεταβλητών

Έστω ότι η x είναι μια συνεχής τυχαία μεταβλητή με συνάρτηση πυκνότητας κατανομής $f(x)$. Ας υποθέσουμε ότι δίνεται η σχέση

$$y = y(x) \tag{8.5}$$

Μπορούμε να βρούμε τη συνάρτηση πυκνότητας πιθανότητας $g(y)$, της νέας μεταβλητής y.

Ας υποθέσουμε ότι υπάρχει αντιστοίχιση ένα-προς-ένα του διαστήματος $(x, x+dx)$ στο διάστημα $(y, y+dy)$. Απαιτούμε να ισχύει

$$f(x)dx = g(y)dy, \quad \text{πρέπει } dx, dy > 0$$

Επειδή η εξάρτηση δεν πρέπει να οδηγεί σε αρνητική $g(y)$, οδηγούμαστε στη σχέση

$$g(y) = f(x)\left|\frac{dx}{dy}\right|$$

Αν ο μετασχηματισμός (8.5) δεν είναι ένα προς ένα αλλά πολλά τμήματα $(x, x+dx)$ μετασχηματίζονται (απεικονίζονται) στο διάστημα $(y, y+dy)$, τότε έχουμε,

$$g(y) = \sum f(x)\left|\frac{dx}{dy}\right|$$

δηλαδή αθροίζουμε πάνω σε όλα τα τμήματα.

Αν έχουμε πολλές μεταβλητές, ισχύει

$$g(y_1, ..., y_n) = f(x_1, ..., x_n)\left|\frac{\partial(x_1, ..., x_n)}{\partial(y_1, ..., y_n)}\right|$$

Χρησιμοποιούμε τη συντομογραφία

$$g(\underline{y}) = f(\underline{x})|J|$$

όπου $|J|$ είναι η απόλυτη τιμή της Ιακωβιανής ορίζουσας του μετασχηματισμού, δηλαδή

$$|J| = \left|\frac{\partial(x_1, ..., x_n)}{\partial(y_1, ..., y_n)}\right| = \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} & \cdots & \frac{\partial x_1}{\partial y_n} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} & \cdots & \frac{\partial x_2}{\partial y_n} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial x_n}{\partial y_1} & \frac{\partial x_n}{\partial y_2} & \cdots & \frac{\partial x_n}{\partial y_n} \end{vmatrix}$$

8.11 Διάδοση Σφαλμάτων

Έστω ότι η τυχαία μεταβλητή y εξαρτάται από τις τυχαίες μεταβλητές $x_1, ..., x_n$ σύμφωνα με τη σχέση

$$y = y(x_1, ..., x_n) = y(\underline{x})$$

Ας υποθέσουμε ότι ξέρουμε τη μήτρα συναλλοιώτου των $x_1, ..., x_n$, δηλαδή την $V(x_1, ..., x_n) = V(\underline{x})$. Αναπτύσσουμε κατά Taylor την y στο σημείο $\underline{\mu} = (\mu_1, ..., \mu_n)$ που αντιπροσωπεύει τις μέσες τιμές των $(x_1, ..., x_n)$ (δηλαδή του $\underline{x}$). Έχουμε τότε

$$y(\underline{x}) = y(\underline{\mu}) + \sum_{i=1}^{n} (x_i - \mu_i) \left.\frac{\partial y}{\partial x_i}\right|_{\underline{x}=\underline{\mu}} + \text{όροι ανώτερης τάξης} \qquad (8.6)$$

Αν πάρουμε τις αναμενόμενες τιμές των διαφόρων όρων, οι αναμενόμενες τιμές των όρων πρώτης τάξης μηδενίζονται, άρα έχουμε

$$E\left[y(\underline{x})\right] = y(\underline{\mu}) + \text{όροι ανώτερης τάξης}$$

Αν οι διαφορές $(x_i - \mu_i)$ είναι μικρές, τότε οι όροι ανώτερης τάξης μπορεί να αγνοηθούν οπότε, κατά προσέγγιση,

$$E\left[y(\underline{x})\right] = y(\underline{\mu})$$

Αν εισαγάγουμε αυτήν τη σχέση στην (8.1) για την απόκλιση της $y(\underline{x})$, βρίσκουμε κατά προσέγγιση

$$V\left[y(\underline{x})\right] = E\left[\left(y(\underline{x}) - E[y(\underline{x})]\right)^2\right] = E\left[\left(y(\underline{x}) - y(\underline{\mu})\right)^2\right] \qquad (8.7)$$

Από την (8.6), αγνοώντας τους όρους τάξης ανώτερης της πρώτης, βρίσκουμε κατά προσέγγιση

$$y(\underline{x}) - y(\underline{\mu}) = \sum_{i=1}^{n} (x_i - \mu_i) \left.\frac{\partial y}{\partial x_i}\right|_{\underline{x}=\underline{\mu}}$$

Με τη βοήθεια της (8.7) βρίσκουμε την προσεγγιστική σχέση

$$V\left[y(\underline{x})\right] = \sum_{i=1}^{n}\sum_{j=1}^{n} \left.\frac{\partial y}{\partial x_i}\right|_{\underline{x}=\underline{\mu}} \left.\frac{\partial y}{\partial x_j}\right|_{\underline{x}=\underline{\mu}} E\left[(x_i - \mu_i)(x_j - \mu_j)\right]$$

Όμως, $V_{ij}(\underline{x}) = E\left[(x_i - \mu_i)(x_j - \mu_j)\right]$. Άρα

$$V\left[y(\underline{x})\right] = \sum_{i=1}^{n}\sum_{j=1}^{n} \left.\frac{\partial y}{\partial x_i}\right|_{\underline{x}=\underline{\mu}} \left.\frac{\partial y}{\partial x_j}\right|_{\underline{x}=\underline{\mu}} V_{ij}(\underline{x}) \qquad (8.8)$$

Αυτή η σχέση, (8.8), είναι η σχέση μετάδοσης σφαλμάτων. Αν η y συνδέεται γραμμικά με τα $x_1, ..., x_n$, τότε η τελευταία σχέση είναι ακριβής (όχι προσέγγιση).

Για n ανεξάρτητες μεταξύ τους μεταβλητές έχουμε (αφού η $V(\underline{x})$ είναι διαγώνια)

$$V[y(\underline{x})] = \sum_{i=1}^{n} \left(\frac{\partial y}{\partial x_i}\right)^2 V_{ii}(\underline{x})$$

Παράδειγμα: Διασπορά του αριθμητικού μέσου ανεξάρτητων μεταβλητών. Έστω οι ανεξάρτητες μεταβλητές $x_1, ..., x_n$, με ίδια διασπορά σ^2. Έστω η συνάρτηση

$$y = y(x_1, ..., x_n) = y(\underline{x}) = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Τότε

$$\frac{\partial y}{\partial x_i} = \frac{1}{n}, \quad \text{για όλα τα } i$$

(λόγω της γραμμικότητας οι ανώτερες παράγωγοι μηδενίζονται άρα η σχέση που ακολουθεί είναι ακριβής)

$$\sigma_y^2 = V[y(\underline{x})] = V[\bar{x}] = \sum_{i=1}^{n} \left(\frac{\partial y}{\partial x_i}\right)^2 \sigma^2 = \frac{\sigma^2}{n}$$

8.12 Πολλές Συναρτήσεις. Συμβολισμός Μητρών

Έστω ότι ισχύουν οι σχέσεις $y_k = y_k(x_1, ..., x_n) = y_k(\underline{x})$ με $k = 1, 2, ..., m$. Το ανάπτυγμα Taylor περί το $\underline{x} = \underline{\mu}$ δίνει

$$y_k(\underline{x}) = y_k(\underline{\mu}) + \sum_i (x_i - \mu_i) \left.\frac{\partial y_k}{\partial x_i}\right|_{\underline{x}=\underline{\mu}} + \cdots, \quad k = 1, 2, ..., n$$

Γενικά, κατά προσέγγιση ισχύουν

$$E[y_k(\underline{x})] = y_k(\underline{\mu}), \quad k = 1, \cdots, m$$

Βρίσκουμε, επομένως, για τη συναλλοίωτη των y_k, y_l:

$$V_{kl}(\underline{y}) = E[(y_k(\underline{x}) - E[y_k(\underline{x})]) \cdot (y_l(\underline{x}) - E[y_l(\underline{x})])]$$

ή

$$V_{kl}(\underline{y}) = \sum_{i=1}^{n}\sum_{j=1}^{n} \left.\frac{\partial y_k}{\partial x_i}\right|_{\underline{x}=\underline{\mu}} \left.\frac{\partial y_l}{\partial x_j}\right|_{\underline{x}=\underline{\mu}} E[(x_i - \mu_i)(x_j - \mu_j)]$$

και τελικά

$$V_{kl}(\underline{y}) = \sum_{i=1}^{n}\sum_{j=1}^{n} \left.\frac{\partial y_k}{\partial x_i}\right|_{\underline{x}=\underline{\mu}} \left.\frac{\partial y_l}{\partial x_j}\right|_{\underline{x}=\underline{\mu}} V_{ij}(\underline{x})$$

Αυτή είναι η γενικευμένη μορφή του νόμου διάδοσης σφαλμάτων. Οι όροι $V_{kl}(\underline{y})$ ορίζουν τη μήτρα συναλλοιώτου $V(\underline{y})$ των εξαρτημένων μεταβλητών $\underline{y}$ από τα $\underline{x}$. Τα σφάλματα των $y_1, ..., y_n$ είναι τα διαγώνια στοιχεία $V_{kk}(\underline{y})$ αν τα διαγώνια στοιχεία της $V(\underline{x})$ είναι τα σφάλματα των $x_1, ..., x_n$.

Τα διαγώνια στοιχεία $V_{kk}(\underline{y})$, γενικά, εξαρτώνται από τα μη διαγώνια στοιχεία της $V(\underline{x})$, αφού

$$V_{kk}(\underline{y}) = \sum_{i=1}^{n}\sum_{j=1}^{n} \left.\frac{\partial y_k}{\partial x_i}\right|_{\underline{x}=\underline{\mu}} \left.\frac{\partial y_k}{\partial x_j}\right|_{\underline{x}=\underline{\mu}} V_{ij}(\underline{x})$$

Αν οι $x_1, ..., x_n$ είναι μη συσχετισμένες τότε προφανώς

$$V_{kk}(\underline{y}) = \sum_{i=1}^{n} \left(\frac{\partial y_k}{\partial x_i}\right)^2 \Bigg|_{\underline{x}=\underline{\mu}} V_{ii}(\underline{x})$$

ή

$$\sigma_k^2(\underline{y}) = \sum_{i=1}^{n} \left(\frac{\partial y_k}{\partial x_i}\right)^2 \Bigg|_{\underline{x}=\underline{\mu}} \sigma_i^2(\underline{x})$$

Ακόμη και αν οι αρχικές μεταβλητές $\underline{x}$ δεν είναι συσχετισμένες, οι μεταβλητές $\underline{y}$ μπορεί να είναι συσχετισμένες, γιατί το $V_{kl} = \sum_{i=1}^{n}(\partial y_k/\partial x_i)(\partial y_l/\partial x_i)V_{ii}(\underline{x})$ δεν είναι υποχρεωτικά μηδέν.

Μπορούμε να χρησιμοποιήσουμε για ευκολία συμβολισμό μητρών. Αν τα $\underline{x}$ και τα $\underline{y}$ είναι διανύσματα που έχουν n και m στοιχεία αντίστοιχα, μπορούμε να γράψουμε $\underline{y} = \underline{C} + S\underline{x} +$ όροι ανώτερης τάξης. Το $\underline{C}$ είναι διάνυσμα (στήλη) με m συνιστώσες σταθερές. Το S είναι μήτρα $m \times n$ που αντιπροσωπεύει το γραμμικό μέρος του μετασχηματισμού $\underline{x} \to \underline{y}$

$$S_{ki} = \left.\frac{\partial y_k}{\partial x_i}\right|_{\underline{x}=\underline{\mu}}$$

Άρα γράφουμε

$$V_{kl}(\underline{y}) = \sum_{i=1}^{n}\sum_{j=1}^{n} S_{ki}V_{ij}(\underline{x})S_{li}$$

Ο νόμος διάδοσης των σφαλμάτων γίνεται

$$V(\underline{y}) = SV(\underline{x})S^T$$

όπου S^T είναι η ανάστροφη μήτρα της S.

8.13 Διακριτές Κατανομές Πιθανότητας

Οι σχέσεις που αναφέραμε τροποποιούνται και έχουμε

$$\sum_r P_r = 1$$

Η αναμενόμενη τιμή του r και η διασπορά είναι

$$E[r] = \sum_r rP_r, \quad V[r] = \sum_r (r - E[r])^2 P_r = E[r^2] - (E[r])^2$$

8.14 Δειγματοληψία

Όταν κάνουμε ένα στατιστικό πείραμα με n παρατηρήσεις, αναφερόμαστε συνήθως σε μέρος μόνο του υπό μελέτη πληθυσμού, δηλαδή, σε ένα δείγμα του. Στόχος μας είναι, από τις n περιορισμένες παρατηρήσεις (μετρήσεις) να συμπεράνουμε την κατανομή του όλου πληθυσμού ως προς κάποια ιδιότητα των στοιχείων του. Η συνάρτηση πυκνότητας πιθανότητας $f(x)$, για συνεχή τυχαία μεταβλητή, ή ισοδύναμα, το σύνολο (όλων) των πιθανοτήτων για μη συνεχή (διακριτή) τυχαία μεταβλητή, περιγράφουν τις ιδιότητες του (όλου) πληθυσμού. Στη Φυσική, αντιστοιχίζουμε τυχαίες μεταβλητές με παρατηρήσεις σε κάποιο φυσικό σύστημα και η $f(x)$ συνοψίζει το αποτέλεσμα όλων των δυνατών μετρήσεων αυτού του συστήματος αν οι μετρήσεις επαναλαμβάνονταν άπειρες φορές υπό τις ίδιες πειραματικές συνθήκες. Οι μετρήσεις στην πράξη δεν είναι όπως οι ανωτέρω και η έννοια του (όλου) πληθυσμού είναι μια εξιδανίκευση. Η $f(x)$ λέγεται και θεωρητική κατανομή.

Ένα πείραμα αποτελείται από περιορισμένο αριθμό μετρήσεων και η σειρά των μετρήσεων $x_1, ..., x_n$ κάποιας ποσότητας X αποτελεί δείγμα μεγέθους n. Το δείγμα είναι ένα υποσύνολο του πληθυσμού, είναι 'ένα δείγμα μεγέθους n που λήφθηκε από τον πληθυσμό'. Θεωρούμε ότι οι συνθήκες είναι τυπικές με την έννοια ότι επαναλαμβανόμενα πειράματα με τον ίδιο αριθμό μετρήσεων θα δώσουν λίγο-πολύ ίδιο αποτέλεσμα. Αυτή είναι η έννοια των τυχαίων δειγμάτων.

Αν το πείραμά μας αναλύεται σε ανεξάρτητες δοκιμές, θεωρούμε ότι για κάθε μια δοκιμή i, μια αντίστοιχη τυχαία μεταβλητή X_i ακολουθεί την (θεωρητική) κατανομή της X. Η ποσότητα x_i που παρατηρήθηκε (μετρήθηκε) είναι μια τιμή της X_i και οι μετρήσεις $x_1, ..., x_n$ θεωρούνται τιμές των n τυχαίων μεταβλητών $X_1, ..., X_n$, οι οποίες είναι ανεξάρτητες και ακολουθούν την κατανομή της X.

Ας θεωρήσουμε μια συνάρτηση $\Psi = \Psi(X_1, ..., X_n)$ των τυχαίων μεταβλητών $X_1, ..., X_n$. Η Ψ είναι τυχαία μεταβλητή της οποίας η κατανομή εξαρτάται από την κατανομή των $X_1, ..., X_n$. Μια τιμή της Ψ που προκύπτει από το πείραμά μας είναι η $\psi = \Psi(x_1, ..., x_n)$. Η κατανομή της συνάρτησης Ψ λέγεται κατανομή δειγματοληψίας. Για κάθε συνάρτηση $\Psi = \Psi(X_1, ..., X_n)$ υποθέτουμε ότι τα $X_1, ..., X_n$ είναι ανεξάρτητες τυχαίες μεταβλητές και έχουν την ίδια (θεωρητική) κατανομή με $E[x_i] = \mu$ και $V[x_i] = \sigma^2$. Τα αντίστοιχα μεγέθη που προκύπτουν από τις συναρτήσεις αυτές λέγονται πειραματικά ή εμπειρικά ή δειγματικά μεγέθη, π.χ. δειγματικός ή εμπειρικός μέσος.

8.15 Εκτιμητική

Σε ένα στατιστικό πείραμα υποθέτουμε ότι η άγνωστη κατανομή πληθυσμού ακολουθεί μια θεωρητική κατανομή f. Στο δείγμα των μετρήσεων $x_1, ..., x_n$

οι ποσότητες αυτές θεωρούνται τιμές των ανεξάρτητων τυχαίων μεταβλητών $X_1, ..., X_n$, που όλες ακολουθούν τη θεωρητική κατανομή f που υποθέσαμε. Η εκλεγείσα f έχει παραμέτρους a_i που πρέπει να προσδιοριστούν. Έχουμε δηλαδή τη συνάρτηση πυκνότητας πιθανότητας $f(x_1, ..., x_n; a_1, ..., a_k)$. Οι παράμετροι a_i είναι γενικά αδύνατο να προσδιοριστούν πλήρως. Αντικείμενο της εκτιμητικής είναι η εύρεση, με χρήση πειραμάτων, κατάλληλων τιμών $\hat{a}_i$ για τις a_i, ώστε με βάση αυτές τις τιμές να μπορούμε να εκτιμήσουμε (μέσω της f) τα μεγέθη που μας ενδιαφέρουν.

8.16 Εκτιμητές (εκτιμητήρες) και Εκτιμήσεις

Έστω ότι δίνεται μια παράμετρος (δηλαδή μια χαρακτηριστική ποσότητα) t μιας θεωρητικής κατανομής. Ονομάζουμε εκτιμητήρα της t κάθε τυχαία μεταβλητή $\Psi_t(X_1, ..., X_n)$ που είναι συνάρτηση των τυχαίων μεταβλητών και στο πεδίο τιμών της περιέχεται η t. Υπάρχουν πολλοί εκτιμητήρες της t.

Η τιμή $\Psi_t(x_1, x_2, ..., x_n)$ λέγεται εκτίμηση της t και παριστάνεται με $\hat{t}$. Ένας εκτιμητήρας της t, $\Psi_t(X_1, ..., X_n)$, θα λέγεται:

1. Αμερόληπτος, αν ο μαθηματικός μέσος του ισούται με t, δηλαδή $E[\Psi_t] = t$.

2. Συγκλίνων ή συνεπής, αν ο μαθηματικός μέσος του έχει όριο την t και η διακύμανσή του έχει όριο το μηδέν, δηλαδή

$$\lim_{n\to\infty} E[\Psi_t] = t, \quad \lim_{n\to\infty} V[\Psi] = 0$$

3. Αποτελεσματικός, αν έχει την ελάχιστη διακύμανση μεταξύ όλων των εκτιμητήρων της t.

Θα δείξουμε ότι αν οι μεταβλητές $X_1, ..., X_n$ ακολουθούν ανεξάρτητα η μια από την άλλη οποιαδήποτε κατανομή, ο δειγματικός μέσος

$$\bar{x} = \frac{1}{n}(x_1 + x_2 + ... + x_n)$$

είναι πάντοτε αμερόληπτος εκτιμητής της μαθηματικής μέσης τιμής μ της κατανομής.
Πράγματι έχουμε,

$$E[\bar{x}] = \frac{1}{n}(E[x_1] + E[x_2] + ... + E[x_n])$$

Αλλά ισχύει

$$E[x_1] = E[x_2] = ... = E[x_n] = \mu$$

και επομένως $E[\bar{x}] = n\mu/n = \mu$, πράγμα που αποδεικνύει τον αρχικό ισχυρισμό.

Επίσης θα δείξουμε ότι αν οι μεταβλητές $X_1, ..., X_n$ ακολουθούν ανεξάρτητα η μια από την άλλη οποιαδήποτε κατανομή, η τυχαία μεταβλητή

$$S^2 = \frac{1}{n-1}\left((x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + ... + (x_3 - \bar{x})^2\right)$$

είναι πάντοτε αμερόληπτος εκτιμητήρας της διακύμανσης σ^2 της κατανομής. Ισχύουν $\bar{x}_i = \mu$ και $\bar{x} = (1/n)\sum_{i=1}^{n} x_i$. Πράγματι έχουμε

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left((x_i - \mu) - (\bar{x} - \mu)^2\right)$$

ή

$$\begin{aligned}
(n-1)S^2 &= \sum_{i=1}^{n}\left((x_i - \mu) - \frac{1}{n}\sum_{j=1}^{n}(x_j - \mu)\right)^2 = \\
&= \sum_{i=1}^{n}\left((x_i - \mu)^2 + \frac{1}{n^2}\left(\sum_{j=1}^{n}(x_j - \mu)\right)^2 - \right. \\
&\quad \left. -\frac{2}{n}(x_i - \mu)\sum_{j=1}^{n}(x_j - \mu)\right) = \\
&= \sum_{i=1}^{n}(x_i - \mu)^2 + \frac{1}{n^2}\sum_{i=1}^{n}\left(\sum_{j=1}^{n}(x_j - \mu)\right)^2 - \\
&\quad - \frac{2}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}(x_i - \mu)(x_j - \mu) = \\
&= \sum_{i=1}^{n}(x_i - \mu)^2 + \frac{1}{n^2}n\left(\sum_{j=1}^{n}(x_j - \mu)\right)^2 - \\
&\quad - \frac{2}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}(x_i - \mu)(x_j - \mu) = \\
&= \sum_{i=1}^{n}(x_i - \mu)^2 + \frac{1}{n}\sum_{j=1}^{n}(x_j - \mu)^2 + \\
&\quad + \frac{1}{n}2\sum_{k=1}^{n}\sum_{\substack{l\neq k\\ l=1}}^{n}(x_k - \mu)(x_l - \mu)) - \\
&\quad - \frac{2}{n}\sum_{i=1}^{n}(x_i - \mu)^2 - \frac{2}{n}\sum_{i=1}^{n}\sum_{\substack{i\neq j\\ j=1}}^{n}(x_i - \mu)(x_j - \mu)
\end{aligned}$$

Επομένως, αφού τα ζεύγη x_k, x_l για $k \neq l$ είναι ανεξάρτητα μεταξύ τους, θα

ισχύει

$$\begin{aligned}(n-1)E[S^2] &= nE\left[(x_i-\mu)^2\right] + \frac{1}{n}nE\left[(x_i-\mu)^2\right] + \\ &+ \frac{1}{n}\cdot 0 - \frac{2}{n}nE\left[(x_j-\mu)^2\right] - \frac{2}{n}\cdot 0 \Rightarrow \\ (n-1)E[S^2] &= nE\left[(x_i-\mu)^2\right] + E\left[(x_i-\mu)^2\right] - 2E\left[(x_j-\mu)^2\right] \\ &= (n-1)E\left[(x_k-\mu)^2\right]\end{aligned}$$

Όμως η διακύμανση του κάθε x_k είναι $\sigma_x^2 = E\left[(x_k-\mu)^2\right]$. Άρα $(n-1)E[S^2] = (n-1)\sigma_x^2$, άρα

$$E[S^2] = \sigma_x^2$$

Για κάποιον εκτιμητήρα Ψ_t του t ορίζεται ως μεροληψία η ποσότητα $b(\Psi_t) = E[\Psi_t] - t$. Αν $b(\Psi_t) \to 0$ καθώς $n \to \infty$, ο εκτιμητήρας είναι συγκλίνων ή συνεπής. Αν $b(\Psi_t) = 0$, για κάθε n, ο εκτιμητήρας είναι αμερόληπτος.

8.17 Κεντρικό Οριακό Θεώρημα

Δίνουμε, χωρίς απόδειξη, το πιο σημαντικό θεώρημα της θεωρίας πιθανοτήτων το οποίο έχει μεγάλες θεωρητικές και πρακτικές συνέπειες.

Έστω $x_1, ..., x_n$ ένα σύνολο από ανεξάρτητες τυχαίες μεταβλητές που η κάθε μια έχει μέσο μ_i και (πεπερασμένη) απόκλιση σ_i. Τότε, η μεταβλητή

$$\frac{\sum_{i=1}^n (x_i - \mu_i)}{\sqrt{\sum_{i=1}^n \sigma_i^2}}$$

στο όριο που το $n \to \infty$ τείνει στην ανηγμένη κανονική κατανομή

$$N(0,1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$$

με μέσον ίσο με μηδέν και διασπορά 1.

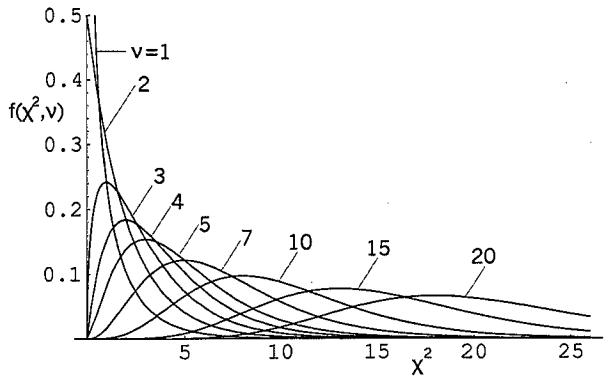
8.18 Ο Νόμος του χ^2

Ορισμός. Θεωρούμε τις $x_1, ..., x_\nu$ (συνηθίζεται το ελληνικό ν αντί του n) οι ανεξάρτητες μεταξύ τους τυχαίες μεταβλητές που η κάθε μια ακολουθεί κανονική κατανομή

$$N(\mu_i, \sigma_i^2) = \frac{1}{\sqrt{2\pi}\sigma_i} e^{-\frac{1}{2}\left(\frac{x_i-\mu_i}{\sigma_i}\right)^2}$$

Ορίζεται το μέγεθος χ^2 ως το άθροισμα

$$\chi^2 = \sum_{i=1}^{\nu} \left(\frac{x_i - \mu_i}{\sigma_i}\right)^2$$



Σχήμα 8.1: Κατανομή χ^2 για διάφορους βαθμούς ελευθερίας ν.

Μπορεί να δειχτεί ότι αυτή η τυχαία μεταβλητή έχει συνάρτηση πυκνότητας πιθανότητας που δίνεται από τη σχέση

$$f(\chi^2,\nu) = \frac{1}{2^{\nu/2}\Gamma(\nu/2)} \left(\chi^2\right)^{\frac{\nu}{2}-1} e^{-\frac{\chi^2}{2}}$$

Η συνάρτηση αυτή ονομάζεται κατανομή χ^2 με ν βαθμούς ελευθερίας. Υπενθυμίζεται ότι ισχύουν για τη συνάρτηση Γ: $\Gamma(x+1) = x\Gamma(x)$, $\Gamma(1/2) = \sqrt{\pi}$ και $\Gamma(1) = 1$.

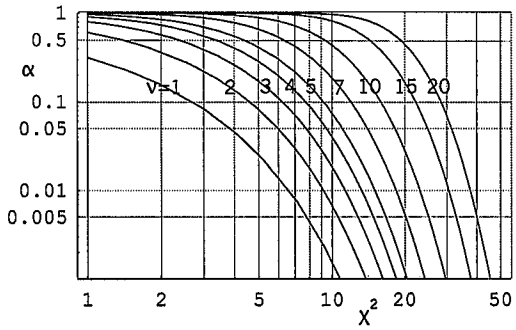
Η μέση τιμή και η απόκλιση του χ^2 είναι ν και 2ν αντίστοιχα. Για $\nu \to \infty$ η κατανομή χ^2 τείνει στην κανονική κατανομή, συγκεκριμένα στη $N(\nu, 2\nu)$. Για $\nu \leq 2$ η κατανομή χ^2 είναι μονότονα φθίνουσα καθώς αυξάνεται το χ^2. Για $\nu = 1$ η $f(\chi^2, 1) \to \infty$ καθώς $\chi^2 \to 0$. Για $\nu = 2$, $\lim_{\chi^2 \to 0} f(\chi^2, 2) \to 0.5$. Για $\nu \geq 2$ η f έχει μέγιστο στη θέση $\chi^2 = \nu - 2$. Λέμε ότι η τυχαία μεταβλητή χ^2 ακολουθεί το νόμο $\chi^2(\nu)$, με ν βαθμούς ελευθερίας, ενώ η $f(\chi^2, \nu)$ είναι η πυκνότητα πιθανότητας.

Στην πράξη ενδιαφερόμαστε συνήθως για την ολική κατανομή χ^2 (ή απλώς συνάρτηση κατανομής), $F(\chi^2_\alpha, \nu)$ που ορίζεται ως

$$F(\chi^2_\alpha, \nu) = \int_0^{\chi^2_\alpha} f(\chi^2, \nu)\, d\chi^2 = 1 - \alpha \quad .$$

Στα Σχ.(8.1) και (8.2) φαίνονται οι κατανομές $f(u,\nu)$, $F(\chi^2_\alpha, \nu)$ και $\alpha(\chi^2_\alpha)$, όπου $u = \chi^2$. Οι τιμές της $F(\chi^2_\alpha, \nu)$ μπορούν να βρεθούν και σε σχετικούς πίνακες (για παράδειγμα, βλέπε Πίνακα 7.1 στο τέλος του κεφαλαίου).

Ισχύει το εξής θεώρημα άθροισης για μεταβλητές που ακολουθούν κατανομές χ^2. Έστω ότι οι $u_1, ..., u_r$ είναι ένα σύνολο από ανεξάρτητες μεταξύ τους μεταβλητές που έχουν κατανομές χ^2 με $\nu_1, ..., \nu_r$ βαθμούς ελευθερίας αντίστοιχα. Τότε το άθροισμα $u_1 + ... + u_r$ ακολουθεί κατανομή χ^2 με $\nu_1 + ... + \nu_r$ βαθμούς ελευθερίας.



Σχήμα 8.2: Κατανομή $\alpha(\chi^2_\alpha)$ συναρτήσει του χ^2_α για διάφορους βαθμούς ελευθερίας ν.

8.19 Κατανομή-t του Student

Έστω x μια μεταβλητή που ακολουθεί την ανηγμένη γκαουσιανή (κανονική) κατανομή $N(0,1)$ και $u = \chi^2$ είναι μια μεταβλητή που ακολουθεί κατανομή χ^2 με ν βαθμούς ελευθερίας, $\chi^2(\nu)$. Έστω ότι οι x και u είναι ανεξάρτητες. Ορίζουμε την μεταβλητή

$$t = \frac{x}{\sqrt{u/\nu}}, \quad \nu > 0, -\infty \leq t \leq \infty$$

Αυτή η μεταβλητή έχει συνάρτηση κατανομής πυκνότητας την

$$f(t,\nu) = \frac{\Gamma\left((\nu+1)/2\right)}{\sqrt{\pi\nu}\,\Gamma(\nu/2)} \frac{1}{\left(1+\dfrac{t^2}{\nu}\right)^{(\nu+1)/2}}$$

και λέγεται κατανομή-t του student με ν βαθμούς ελευθερίας. Η $f(t,\nu)$ είναι συμμετρική περί το $t = 0$ και έχει ένα μέγιστο στο $t = 0$. Για $\nu \to \infty$, η $f(t,\nu) \to (1/\sqrt{2\pi})\exp(-t^2/2)$, δηλαδή γκαουσιανή $N(0,1)$, με $E[t] = 0$ και $V[t] = \nu/(\nu-2)$, $(\nu > 2)$.

Η ολική (αθροιστική) κατανομή-t του Student είναι η

$$F(t_\alpha,\nu) = \int_{-\infty}^{t_\alpha} f(t,\nu)\,dt = 1-\alpha$$

Οι τιμές της βρίσκονται σε σχετικούς πίνακες (για παράδειγμα, βλέπε Πίνακα 7.2 στο τέλος του κεφαλαίου). Για $\nu = 1$ παίρνομε την κατανομή Cauchy

$$f(t,1) = \frac{1}{\pi}\frac{1}{1+t^2}, \quad \int_{-\infty}^{t_\alpha} f(t,1)\,dt = 1$$

Από αυτήν τη συνάρτηση δεν μπορεί να υπολογιστεί μέση τιμή για το t γιατί το σχετικό ολοκλήρωμα δεν υπάρχει. Το ίδιο ισχύει για όλες τις ροπές, δηλαδή δεν υπάρχουν τα ολοκληρώματα των ροπών

$$\int_{-\infty}^{\infty} t^k \frac{1}{\pi}\frac{1}{1+t^2}\,dt$$

8.20 Σύγκριση Πειραματικών Δεδομένων με τη Θεωρία

Οσα αναφέραμε μέχρι τώρα σχετίζονται με ιδιότητες κατανομών πιθανοτήτων που χρησιμοποιούνται συχνά στη Φυσική. Τα πειραματικά δεδομένα δεν μπορούν κάθε φορά να συγκριθούν άμεσα με τις ιδανικές, θεωρητικές, μαθηματικές κατανομές. Σχεδόν πάντοτε το ιδανικό μοντέλο πρέπει να τροποποιηθεί προτού γίνουν συγκρίσεις (που να έχουν νόημα), μεταξύ πρόβλεψης και παρατήρησης. Αυτό οφείλεται στο γεγονός ότι η θεωρητική κατανομή περιγράφει ιδανικό πείραμα που εκτελείται υπό εξιδανικευμένες συνθήκες που στην πράξη δεν ικανοποιούνται. Οταν, για παράδειγμα, γράφουμε τον νόμο κατανομής του χρόνου ζωής ραδιενεργού υλικού, $f(t,\lambda) = \lambda \exp(-\lambda t)$, για να μπορεί να γίνει άμεσα σύγκριση με τα πειραματικά αποτελέσματα μέτρησης του χρόνου ζωής, πρέπει ο ανιχνευτής να έχει άπειρη έκταση, πρέπει ακόμα να γίνουν διορθώσεις για πειραματικές αβεβαιότητες και διάφορες συστηματικές επιδράσεις. Εδώ θα αναφερθούμε πολύ συνοπτικά σε μερικά από τα θέματα που σχετίζονται με αυτό το δύσκολο αντικείμενο.

8.21 Πειραματικά Σφάλματα Μετρήσεων. Συνάρτηση Διακριτικής Ικανότητας.

Ένεκα της αβεβαιότητας των μετρήσεων, μια μεταβλητή με πραγματική τιμή x μπορεί να παρατηρηθεί με μια διαφορετική τιμή x'. Γι' αυτό πρέπει, προτού γίνει η σύγκριση με τα δεδομένα του πειράματος, να τροποποιηθεί η θεωρητική ιδανική κατανομή, ώστε να αντιστοιχεί στην "παρατηρούμενη" ποσότητα x'. Η συνάρτηση $r(x',x)$ που περιγράφει την παρατηρούμενη ποσότητα x' για δεδομένη αληθή ποσότητα x, λέγεται συνάρτηση διακριτικής ικανότητας της x (resolution function). Αν η αληθής συνάρτηση κατανομής πιθανότητας είναι $f(x,\underline{\theta})$, όπου $\underline{\theta}$ είναι κάποιες αρχικές παράμετροι, τότε ισχύει για την τροποποιημένη κατανομή ότι:

$$f'(x') = \int f(x,\underline{\theta}) r(x',x) dx.$$

Η παρατηρούμενη $f'(x')$ εξαρτάται από τις αρχικές παραμέτρους $\underline{\theta}$ και από άλλες σταθερές που περιέχει η $r(x',x)$. Αν η $r(x',x)$ ισούται με τη συνάρτηση του Dirac, $r(x',x) = \delta(x'-x)$, δηλαδή έχουμε ιδανικά καλή διακριτική ικανότητα, τότε $f'(x') = f(x',\underline{\theta})$. Αν η $r(x',x)$ είναι πολύ κακή και έχει μεγάλες τιμές ακόμη και εκεί που η $f(x,\underline{\theta})$ είναι μικρή, τότε μπορούμε να ισχυριστούμε ότι $f(x',\underline{\theta}) = \delta(x-x_0)$, οπότε $f'(x') = r(x',x_0)$, δηλαδή η παρατηρούμενη κατανομή πιθανότητας είναι η συνάρτηση διακριτικής ικανότητας. Ενα απλό παράδειγμα είναι η περίπτωση γκαουσιανής διακριτικής ικανότητας και γκαουσιανής επίσης συνάρτησης πυκνότητας πιθανότητας. Εστω η συνάρτηση

πυκνότητας πιθανότητας:

$$f(x, x_0, \Gamma) = \frac{1}{\sqrt{2\pi}\Gamma} \exp[-\frac{(x' - x_0)^2}{2\Gamma^2}].$$

Εστω επίσης η

$$r(x', x) = \frac{1}{\sqrt{2\pi}R} \exp[-\frac{(x' - x)^2}{2R^2}].$$

Κάνοντας την ολοκλήρωση προκύπτει:

$$f'(x') = \frac{1}{\sqrt{2\pi}\sqrt{\Gamma^2 + R^2}} \exp[-\frac{(x' - x_0)^2}{2(\Gamma^2 + R^2)}].$$

Για να ληφθεί υπόψη η απόδοση του ανιχνευτή, μπορεί να τροποποιηθεί κατάλληλα η $f(x, \underline{\theta})$ ή να δίνεται κατάλληλο βάρος στα παρατηρούμενα γεγονότα (μετρήσεις). Παράδειγμα είναι το "κόψιμο" εκθετικής κατανομής. Εστω η κατανομή $f(t, \lambda) = \lambda \exp(-\lambda t)$, που αναφέρεται σε διασπάσεις ασταθών σωματιδίων. Αν η γεωμετρία του ανιχνευτή είναι τέτοια που να επιτρέπεται η ανίχνευση μόνο χρόνων ζωής t έτσι ώστε $t_{min} \leq t \leq t_{max}$, τότε έχουμε την

$$f'(t, \lambda) = \frac{\exp[-\lambda t]}{\int_{t_{min}}^{t_{max}} \exp[-\lambda t] dt} = \frac{\lambda \exp[-\lambda t]}{\exp[-\lambda t_{min}] - \exp[-\lambda t_{max}]}.$$

Δηλαδή η τροποποίηση γίνεται έτσι ώστε:

$$\int_{t_{min}}^{t_{max}} f'(t, \lambda) dt = 1.$$

Γενικά:

$$f'(x, \underline{\theta}) = \frac{f(x, \underline{\theta})}{\int_A^B f(x, \underline{\theta}) dx},$$

όπου τα A και B καθορίζουν την περιοχή των μετρούμενων x.

8.22 Στατιστική Εκτίμηση για Δείγματα από Γκαουσιανές Κατανομές

Ορισμοί: Θεωρούμε ένα τυχαίο δείγμα $x_1, x_2, \ldots, x_n$ από έναν πληθυσμό που χαρακτηρίζεται από μια πυκνότητα πιθανότητας η οποία εξαρτάται από άγνωστη παράμετρο θ. Θέλουμε να προσδιορίσουμε το μέγεθος της θ από το δείγμα. Εστω ότι $t(x_1, x_2, \ldots, x_n)$ είναι συνάρτηση των μεταβλητών του δείγματος που δεν εξαρτάται από άγνωστες παραμέτρους. Πρόκειται για συνάρτηση εκτίμησης, αφού υποθέτουμε ότι η $t(x_1, x_2, \ldots, x_n)$ έχει κάποια αντιστοιχία με την θ. Θα δούμε ότι η $t(x_1, x_2, \ldots, x_n)$ είναι μια εκτιμήτρια της θ. Εστω $f(t)$ η πυκνότητα πιθανότητας της μεταβλητής t. Μπορούμε να καθορίσουμε δύο τιμές t_a και t_b έτσι ώστε

$$\int_{t_a}^{t_b} f(t) dt = \gamma, \quad 0 \leq \gamma \leq 1.$$

Αν για την παράμετρο θ η πιθανότητα είναι τέτοια ώστε

$$P(t_a \leq t \leq t_b) = \gamma,$$

τότε το κλειστό διάστημα $[t_a, t_b]$ λέγεται ένα διάστημα εμπιστοσύνης $100 \cdot \gamma\,\%$ για την θ. Το γ λέγεται συντελεστής εμπιστοσύνης και τα t_a, t_b όρια εμπιστοσύνης.

Αν μας δίνεται ένα συγκεκριμένο δείγμα μεγέθους n, για παράδειγμα n μετρήσεις, τα όρια t_a και t_b είναι συγκεκριμένα και μπορούν να υπολογιστούν από τις μετρήσεις. Αρα από το συγκεκριμένο δείγμα συνάγεται ένα συγκεκριμένο διάστημα $[t_a, t_b]$, το οποίο είτε θα περιλαμβάνει την πραγματική τιμή της θ είτε όχι. Ενα δεύτερο δείγμα, επίσης μεγέθους n, θα οδηγήσει γενικά σε διαφορετικό διάστημα, το οποίο μπορεί να περιλαμβάνει την πραγματική τιμή της παραμέτρου θ αλλά μπορεί και να μην την περιλαμβάνει. Η σχετική πρόταση για ένα συγκεκριμένο δείγμα είναι επομένως:

$$\begin{aligned} P(t_a \leq \theta \leq t_b) &= 1, \quad \text{αν} \quad \theta \in [t_a, t_b], \\ P(t_a \leq \theta \leq t_b) &= 0, \quad \text{αν} \quad \theta \in [t_a, t_b]. \end{aligned}$$

Το νόημα της σχέσης

$$P(t_a \leq \theta \leq t_b) = \gamma$$

είναι το εξής: Υπάρχει πιθανότητα γ ότι το τυχαίο διάστημα $[t_a, t_b]$ θα περιλαμβάνει την πραγματική τιμή της θ. Αν εξεταστεί μεγάλο πλήθος δειγμάτων μεγέθους n, δηλαδή άν το πείραμα γίνει πολλές φορές υπό τις ίδιες συνθήκες, τότε η θ θα βρίσκεται στα υπολογιζόμενα διαστήματα $[t_a, t_b]$ σε ποσοστό $100 \cdot \gamma\,\%$ των περιπτώσεων. Με άλλα λόγια αναμένεται ότι για πολλές επαναλήψεις του ίδιου πειράματος, τα υπολογιζόμενα όρια t_a και t_b είναι τέτοια ώστε η σχέση $t_a \leq \theta \leq t_b$ είναι ορθή στο $100 \cdot \gamma\,\%$ των περιπτώσεων. Ο συντελεστής αξιοπιστίας γ επομένως, ανταναχλά την αξιοπιστία που αποδίδουμε στη σχέση $t_a \leq \theta \leq t_b$. Όποτε χρειάζεται να διατυπωθεί μια πρόταση για την άγνωστη παράμετρο σε σχέση με ένα διάστημα εμπιστοσύνης, παρουσιάζεται το εξής δίλημμα: Η εκλογή μεγάλου διαστήματος αντιστοιχεί σε μεγάλη πιθανότητα η άγνωστη παράμετρος να βρίσκεται μέσα στο διάστημα, αλλά τότε η παράμετρος δεν είναι καλά καθορισμένη. Από την άλλη μεριά, αν δοθεί ένα μικρό διάστημα, αυτό σημαίνει καλύτερο προσδιορισμό της παραμέτρου, αλλά τότε η πιθανότητα να βρίσκεται η πραγματική τιμή στο μικρό διάστημα είναι μικρή. Στην πράξη, οι περισσότεροι προτιμούν την πρώτη δυνατότητα και επιλέγουν τον συντελεστή αξιοπιστίας γ κοντά στο 1. Κοινές επιλογές είναι 0,90, 0,95, 0,99.

8.23 Διαστήματα Εμπιστοσύνης για τον Μέσο

Πολλές φορές μπορούμε να υποθέσουμε ότι το αποτέλεσμα μιας μέτρησης κάποιας ποσότητας είναι μια τυχαία μεταβλητή που έχει γκαουσιανή κατανομή περί την μέση τιμή μ. Θα υποθέσουμε ότι η μ είναι άγνωστη, για την οποία θέλουμε να πούμε κάτι με βάση n ανεξάρτητες μετρήσεις που έγιναν υπό τις ίδιες συνθήκες. Δηλαδή θέλουμε να εικάσουμε κάτι για το μ με βάση το τυχαίο δείγμα μεγέθους n από τον γκαουσιανό πληθυσμό (κανονική κατανομή) $N(\mu, \sigma^2)$. Το σ^2 δίνει το μέτρο της ακρίβειας των μετρήσεων και θα θεωρήσουμε ότι είναι γνωστό. Εχουμε, λοιπόν, n ανεξάρτητες μετρήσεις $x_1, x_2, \ldots, x_n$ με γνωστό σφάλμα σ (τυπική απόκλιση ίδια για κάθε μέτρηση). Θεωρούμε τον δειγματικό μέσο

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i.$$

Αυτό το στατιστικό μέγεθος αποδεικνύεται ότι έχει κατανομή $N\left(\mu, \frac{\sigma^2}{n}\right)$. Επίσης αποδεικνύεται ότι η μεταβλητή $y = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}}$ έχει κατανομή ανηγμένη γκαουσιανή (κανονική): $N(0,1)$. Εύκολα, επομένως, προκύπτει ότι για το διάστημα των ± 2 τυπικών αποκλίσεων, $[-2\sigma, +2\sigma] = [-2, +2]$ (αφού $\sigma = 1$) ισχύει:

$$P\left(-2 \leq \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} \leq +2\right) = \int_{-2}^{+2} g(y)dy = 0,954,$$

όπου

$$g(y) = \frac{1}{\sqrt{2\pi}} \exp[-\frac{y^2}{2}] = N(0,1).$$

Αυτό σημαίνει ότι η τυχαία μεταβλητή $y = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}}$ έχει πιθανότητα $0,954$ να βρίσκεται στο διάστημα $[-2, +2]$. Αυτό με τη σειρά του σημαίνει ότι

$$P\left(\bar{x} - 2\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + 2\frac{\sigma}{\sqrt{n}}\right) = 0,954.$$

Μπορούμε να πάρουμε άλλα διαστήματα εμπιστοσύνης με άλλες πιθανότητες. Γενικά:

$$P\left(a \leq \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} \leq b\right) = \int_a^b g(y)dy = \gamma.$$

Οι τιμές του γ, για διάφορα $[a, b]$, βρίσκονται σε σχετικούς πίνακες (για παράδειγμα, βλέπε Πίνακα 7.3 στο τέλος του κεφαλαίου).

Ας δούμε τώρα την περίπτωση που το σ^2 δεν είναι γνωστό. Αν $x_1, x_2, \ldots, x_n$ είναι ένα τυχαίο δείγμα κατανομής $N(\mu, \sigma^2)$, τότε μπορούν να φτιαχτούν δύο μεταβλητές με γνωστές ιδιότητες: Η μεταβλητή $\frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}}$ με $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ που

ακολουθεί κατανομή $N(0,1)$ και η $\frac{(n-1)s^2}{\sigma^2}$ όπου $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$ που ακολουθεί κατανομή $\chi^2(n-1)$. Οι μεταβλητές αυτές είναι ανεξάρτητες και έτσι η μεταβλητή

$$t = \frac{\frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}}}{\sqrt{\frac{(n-1)s^2}{\sigma^2}/(n-1)}} = \frac{\bar{x}-\mu}{\frac{s}{\sqrt{n}}}$$

ακολουθεί κατανομή του student (μεταβλητή t) με $(n-1)$ βαθμούς ελευθερίας.

Εχουμε δηλαδή την κατανομή του Student με $\nu = n-1$ βαθμούς ελευθερίας,

$$f(t,\nu) = \frac{\Gamma(\frac{1}{2}(\nu+1))}{\sqrt{\pi\nu}\Gamma(\frac{1}{2}\nu)} \cdot \frac{1}{(1+\frac{t^2}{\nu})^{\frac{1}{2}(\nu+1)}}$$

Βλέπουμε ότι η άγνωστη παράμετρος σ απαλείφεται και μένει μόνος άγνωστος το μ.

Μπορούμε να γράψουμε, ανάλογα με πριν, σχέση της μορφής:

$$P\left(a \le \frac{\bar{x}-\mu}{\frac{s}{\sqrt{n}}} \le b\right) = \int_a^b f(t,n-1)dt = \gamma.$$

Επειδή υπάρχει συμμετρία της $f(t,n-1)$ περί το $t=0$, συνηθίζεται να θεωρούμε διαστήματα συμμετρικά περί το 0 :

$$P\left(-b \le \frac{\bar{x}-\mu}{\frac{s}{\sqrt{n}}} \le b\right) = \int_{-b}^b f(t,n-1)dt = \gamma.$$

Ο πίνακας 2 βοηθά στη χρήση αυτής της σχέσης. Μπορούμε να γράψουμε τη σχέση με την πιθανότητα και ως εξής:

$$P\left(\bar{x} - b\frac{s}{\sqrt{n}} \le \mu \le \bar{x} + b\frac{s}{\sqrt{n}}\right) = \gamma.$$

8.24 Διαστήματα Εμπιστοσύνης για την Απόκλιση

Έστω το τυχαίο δείγμα $x_1, x_2, \ldots, x_n$ μιας κατανομής $N(\mu, \sigma^2)$. Θέλουμε να υπολογίσουμε διαστήματα εμπιστοσύνης του σ^2. Έστω ότι το μ είναι γνωστό. Δηλαδή είναι η περίπτωση που το μ είναι γνωστό και μετριέται με διαδικασία άγνωστου σφάλματος (τυπικής απόκλισης) σ. Μια συνάρτηση εκτίμησης που αντιστοιχεί στο σ^2 είναι η $\frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2$. Ξέρουμε ότι η μεταβλητή $\sum_{i=1}^{n}\left(\frac{x_i-\mu}{\sigma}\right)^2$ κατανέμεται ως $\chi^2(n)$. Για την $\chi^2(n)$ μπορούμε να βρούμε δύο παραμέτρους a και b έτσι ώστε για $0 \le \gamma \le 1$ να έχουμε:

$$P\left(a \le \sum_{i=1}^{n}\left(\frac{x_i-\mu}{\sigma}\right)^2 \le b\right) = \int_a^b f(u,n)du = \gamma,$$

όπου $f(u,n)$ η κατανομή χ^2 με n βαθμούς ελευθερίας $\left(\chi^2 = \sum_{i=1}^n \left(\frac{x_i-\mu}{\sigma}\right)^2 = u\right)$.

Μπορούμε να γράψουμε τη σχέση αυτή και με τη μορφή:

$$P\left(\frac{\sum_{i=1}^n (x_i-\mu)^2}{b} \leq \sigma^2 \leq \frac{\sum_{i=1}^n (x_i-\mu)^2}{a}\right) = \gamma.$$

Συνηθίζεται τα όρια a και b να επιλέγονται έτσι ώστε οι "ουρές" πέραν του b και πριν από το a να αντιστοιχούν σε πιθανότητες $\frac{1}{2}(1-\gamma)$. Ο πίνακας 1 μας βοηθά στους σχετικούς υπολογισμούς.

Όταν το μ δεν είναι γνωστό, σκεπτόμαστε ως εξής: Το δείγμα είναι το $x_1, x_2, \ldots, x_n$ και έχει κατανομή $N(\mu, \sigma^2)$. Μπορούμε να γράψουμε τον αριθμητικό μέσον $\bar{x} = \frac{1}{n}\sum_{i=1}^n x_i$ και ξέρουμε ότι η μεταβλητή $\sum_{i=1}^n \left(\frac{x_i-\mu}{\sigma}\right)^2$ ακολουθεί την κατανομή $\chi^2(n-1)$. Επομένως έχουμε:

$$P\left(a \leq \sum_{i=1}^n \left(\frac{x_i-\mu}{\sigma}\right)^2 \leq b\right) = \int_a^b f(u, n-1)du = \gamma.$$

Κάνουμε και πάλι χρήση του σχετικού πίνακα. Μπορούμε να γράψουμε τη σχέση και ως εξής:

$$P\left(\frac{\sum_{i=1}^n (x_i-\bar{x})^2}{b} \leq \sigma^2 \leq \frac{\sum_{i=1}^n (x_i-\bar{x})^2}{a}\right) = \gamma.$$

Τα a και b υπολογίζονται όπως στην προηγούμενη περίπτωση.

8.25 Περιοχές Εμπιστοσύνης για το Μέσο και τη Διασπορά μαζί

Ας υποθέσουμε ότι θέλουμε να δώσουμε συνδυασμένη περιοχή εμπιστοσύνης για τον μέσο και την διασπορά (απόκλιση) μαζί. Έχουμε το δείγμα $x_1, x_2, \ldots, x_n$ με κατανομή $N(\mu, \sigma^2)$. Ξέρουμε ότι οι μεταβλητές $\bar{x}$ και s^2 είναι ανεξάρτητες, αφού τα x_i έχουν κανονική (γκαουσιανή) κατανομή. Η μεταβλητή $\frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}}$ ακολουθεί την κατανομή $N(0,1)$, ενώ η $\sum_{i=1}^n \left(\frac{x_i-\bar{x}}{\sigma}\right)^2$ ακολουθεί την κατανομή $\chi^2(n-1)$, άρα μπορούμε να γράψουμε:

$$P\left(-a \leq \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} \leq a\right) = \sqrt{\gamma},$$

$$P\left(b \leq \sum_{i=1}^n \left(\frac{x_i-\bar{x}}{\sigma}\right)^2 \leq b'\right) = \sqrt{\gamma},$$

όπου το a προσδιορίζεται από την κατανομή $N(0,1)$ και τα b, b' από την $\chi^2(n-1)$.

Πολλαπλασιάζοντας αυτές τις ανεξάρτητες πιθανότητες βρίσκουμε τη συνδυασμένη πιθανότητα:

$$P\left(-a \leq \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} \leq a, b \leq \sum_{i=1}^n \left(\frac{x_i-\bar{x}}{\sigma}\right)^2 \leq b'\right) = \gamma.$$

Οι ανισότητες αυτές καθορίζουν την περιοχή εμπιστοσύνης στον χώρο των παραμέτρων που φαίνεται στη γραμμοσκιασμένη περιοχή στο σχήμα. Η περιοχή περιορίζεται μεταξύ των δύο οριζόντιων ευθειών με $\sigma^2 = \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{b'}$ και $\sigma^2 = \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{b}$ και της παραβολής $\sigma^2 = \frac{n(\mu-\bar{x})^2}{a^2}$, που δίνει τη σχέση μεταξύ σ^2 και μ.

8.26 Εκτίμηση Παραμέτρων

Ο όρος εκτιμητήρας αντιστοιχεί σε συνάρτηση των παρατηρήσεων ή σε μέθοδο που χρησιμοποιείται για να βρεθεί η τιμή άγνωστης παραμέτρου. Εκτίμηση είναι η τιμή της παραμέτρου που βρίσκεται από τη χρήση του εκτιμητήρα για συγκεκριμένο σύνολο παρατηρήσεων. Αν η παράμετρος είναι θ, η εκτίμησή της είναι $\hat{\theta}$. Έγινε εισαγωγή της συνάρτησης εκτίμησης ως συνάρτηση μιας ή περισσότερων τυχαίων μεταβλητών που δεν εξαρτάται από άγνωστες παραμέτρους. Γενικώς, αυτή η συνάρτηση $t(x_1, x_2, \ldots, x_n)$ θα είναι εκτιμητήρας του άγνωστου θ ή κάποιας συνάρτησης του θ.

Αν ένας πληθυσμός έχει κατανομή πιθανότητας $f(x|\theta)$, για δεδομένο θ η αληθοφάνεια των παρατηρήσεων $x_1, x_2, \ldots, x_n$ δίνεται από τη σχέση:

$$L(x_1, x_2, \ldots, x_n|\theta) = \prod_{i=1}^{n} f(x_i|\theta).$$

Το γινόμενο εκφράζει την (υπό όρους) πιθανότητα να επιτευχθούν οι μετρήσεις $x_1, x_2, \ldots, x_n$ για δεδομένο θ. Επειδή η $L(x_1, x_2, \ldots, x_n|\theta)$ μπορεί να θεωρείται συνάρτηση του θ, λέγεται και συνάρτηση αληθοφάνειας. Χρησιμοποιείται και ο συμβολισμός $L(\underline{x}|\theta)$. Μια άλλη άποψη θεωρεί την $L(\underline{x}|\theta)$ ως την συνδυασμένη συνάρτηση πυκνότητας πιθανότητας για τις παρατηρούμενες μεταβλητές x_i για δεδομένο θ. Έχουμε ήδη αναφέρει διάφορες ιδιότητες που πρέπει να έχει ένας καλός εκτιμητήρας.

8.27 Μέθοδος Μέγιστης Αληθοφάνειας

Αυτή η μέθοδος της μέγιστης αληθοφάνειας (ML, Maximum Likelihood) είναι ισχυρή και γενική μέθοδος υπολογισμού παραμέτρων όταν είναι γνωστή η μορφή της συνάρτησης της θεωρητικής κατανομής. Για μεγάλα δείγματα οι εκτιμήσεις έχουν γκαουσιανή κατανομή. Αυτό βοηθά στον εύκολο προσδιορισμό των διακυμάνσεων.

Έχουμε

$$L(\underline{x}|\theta) = \prod_{i=1}^{n} f(x_i|\theta),$$

όπου $\int f(x_i|\theta)dx_i = 1$ και άρα $\int_\Omega L(\underline{x}|\theta)dx_1...dx_n = 1$, όπου Ω είναι ο χώρος των $x_1, x_2, \ldots, x_n$. Σύμφωνα με την Αρχή Μέγιστης Αληθοφάνειας πρέπει να επιλεγεί ως εκτίμηση της παραμέτρου θ η τιμή $\hat{\theta}$, μέσα στον επιτρεπτό χώρο των θ, η οποία κάνει το $L(\underline{x}|\theta)$ όσο μεγαλύτερο γίνεται. Δηλαδή ισχύει:

$$L(\underline{x}|\hat{\theta}) \geq L(\underline{x}|\theta)$$

για όλες τις τιμές του θ. Αν η $L(\underline{x}|\theta)$ έχει πρώτη και δεύτερη παράγωγο ως προς θ, τότε μπορούμε να βρούμε την $\hat{\theta}$ λύνοντας την εξίσωση

$$\frac{\partial L(\underline{x}|\theta)}{\partial \theta} = \frac{\partial}{\partial \theta}\prod_{i=1}^{n} f(x_i|\theta) = 0,$$

με τη συνθήκη ότι η δεύτερη παράγωγος στη θέση $\theta = \hat{\theta}$ είναι αρνητική:

$$\left.\frac{\partial^2 L(\underline{x}|\theta)}{\partial \theta^2}\right|_{\theta=\hat{\theta}} = \left.\frac{\partial^2}{\partial \theta^2}\prod_{i=1}^{n} f(x_i|\theta)\right|_{\theta=\hat{\theta}} < 0.$$

Αν υπάρχουν περισσότερα του ενός μέγιστα πρέπει να αναζητηθεί πρόσθετη πληροφορία για να επιλεγεί το ένα.

Είναι ευκολότερο να εργαζόμαστε με τον λογάριθμο της $L(\underline{x}|\theta)$ και να ζητούμε τη θέση μεγίστου του λογαρίθμου που συμπίπτει με τη θέση μεγίστου της $L(\underline{x}|\theta)$.

$$\left.\frac{\partial \ln L(\underline{x}|\theta)}{\partial \theta}\right|_{\theta=\hat{\theta}} = \left.\frac{\partial}{\partial \theta}\sum_{i=1}^{n} \ln f(x_i|\theta)\right|_{\theta=\hat{\theta}} = 0.$$

Πρέπει

$$\left.\frac{\partial^2 \ln L(\underline{x}|\theta)}{\partial \theta^2}\right|_{\theta=\hat{\theta}} = \left.\frac{\partial^2}{\partial \theta^2}\sum_{i=1}^{n} \ln f(x_i|\theta)\right|_{\theta=\hat{\theta}} < 0.$$

Αν έχουμε πολλές άγνωστες παραμέτρους, $\underline{\theta} = (\theta_1, \theta_2, \ldots, \theta_k)$, τότε έχουμε $L(\underline{x}|\underline{\theta})$ και

$$\frac{\partial}{\partial \theta_j}\ln L(\underline{x}|\underline{\theta}) = 0, j = 1, 2, \ldots, k,$$

απ'οπου βρίσκουμε τις εκτιμήσεις $\underline{\hat{\theta}} = (\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_k)$. Μια ικανή συνθήκη ώστε να έχουμε μέγιστο είναι η μήτρα $U(\underline{\hat{\theta}})$ με

$$U_{ij}(\underline{\hat{\theta}}) = \left.\frac{\partial^2 \ln L}{\partial \theta_i \partial \theta_j}\right|_{\underline{\theta}=\underline{\hat{\theta}}}$$

να είναι αρνητικά ορισμένη.

Απλό παράδειγμα μετρήσεων γκαουσιανής κατανομής με διαφορετικά σφάλματα σ_i. Εύρεση μέσου με βάρη. Έχουμε τις μετρήσεις $x_1, x_2, \ldots, x_n$ με σφάλματα $\sigma_1, \sigma_2, \ldots, \sigma_n$. Προφανώς

$$L(x_1, x_2, \ldots, x_n; \sigma_1, \sigma_2, \ldots, \sigma_n|\mu) = L(\underline{x}; \underline{\sigma}|\mu) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma_i} e^{-\frac{1}{2}\left(\frac{x_i-\mu}{\sigma_i}\right)^2}$$

άρα

$$\frac{\partial \ln L}{\partial \mu} = -\sum_{i=1}^{n} \frac{x_i - \mu}{\sigma_i} \frac{1}{\sigma_i} = 0,$$

απ' όπου

$$\hat{\mu} = \frac{\sum_{i=1}^{n} \frac{x_i}{\sigma_i^2}}{\sum_{i=1}^{n} \frac{1}{\sigma_i^2}}.$$

Αν $\sigma_i = \sigma$, τότε $\hat{\mu} = (1/n)\sum_{i=1}^{n} x_i$, ο γνωστός μέσος χωρίς βάρη. Εύκολα προκύπτει ότι $\frac{\partial^2 \ln L}{\partial \mu^2} < 0$, άρα το $\hat{\mu}$ αντιστοιχεί σε μέγιστα των L και $\ln L$.

8.28 Απόκλιση Εκτιμητήρων με τη Μέθοδο Μέγιστης Αληθοφάνειας

Έστω η συνάρτηση αληθοφάνειας

$$L(\underline{x}|\underline{\theta}) = \prod_{i=1}^{n} f(x_i|\underline{\theta})$$

με $\underline{\theta} = (\theta_1, \theta_2, \ldots, \theta_k)$. Αν οι εκτιμήσεις μπορούν να γραφτούν ως συναρτήσεις των x_i, δηλαδή $\hat{\theta}_i = \hat{\theta}_i(x_1, x_2, \ldots, x_n), i = 1, 2, \ldots, k$. Τότε ο όρος συναλλοιώτου $V_{ij}(\underline{\hat{\theta}})$ δίνεται από τη σχέση:

$$V_{ij}(\underline{\hat{\theta}}) = \int (\hat{\theta}_i - \theta_i)(\hat{\theta}_j - \theta_j) L(\underline{x}|\underline{\theta}) d\underline{x}.$$

Ο τύπος αυτός μπορεί να χρησιμοποιηθεί για τον υπολογισμό της μήτρας συναλλοιώτου. Αν η $f(\underline{x}|\theta)$ οδηγεί σε αποτελεσματικό εκτιμητήρα για την παράμετρο θ τότε

$$V(\hat{\theta}) = \left. \frac{\left(1 + \frac{\partial b}{\partial \theta}\right)^2}{\left(-\frac{\partial^2 \ln L}{\partial \theta^2}\right)} \right|_{\theta=\hat{\theta}},$$

όπου $b(\theta)$ η μεροληψία. Αν έχουμε αμερόληπτο εκτιμητήρα, τότε $b = 0$, άρα

$$V(\hat{\theta}) = \left. \frac{1}{\left(-\frac{\partial^2 \ln L}{\partial \theta^2}\right)} \right|_{\theta=\hat{\theta}}.$$

Σε περιπτώσεις πολλών παραμέτρων τα πράγματα είναι πιο πολύπλοκα. Αν όμως υπάρχει ένα σύνολο εκτιμητήρων $t_1, t_2, \ldots, t_k$ για τις παραμέτρους $\theta_1, \theta_2, \ldots, \theta_k$ που είναι αποτελεσματικοί εκτιμητήρες, τότε για πολύ μεγάλα δείγματα ισχύει:

$$V_{ij}^{-1}(\underline{\hat{\theta}}) = \left. \left(-\frac{\partial^2 \ln L}{\partial \theta_i \partial \theta_j}\right) \right|_{\underline{\theta}=\underline{\hat{\theta}}}.$$

Μπορούμε να βρούμε διαστήματα εμπιστοσύνης στην περίπτωση που ο αριθμός των παρατηρήσεων (μέγεθος δείγματος) είναι πολύ μεγάλος. Τότε, για

ένα θ, η $L(\underline{x}|\theta)$ παίρνει τη μορφή κανονικής κατανομής ως προς θ με μέσο $\hat{\theta}$ και απόκλιση σ^2. Γράφομε τότε $L(\underline{x}|\theta) = L(\theta) = L(max)\exp[-\frac{1}{2}Q]$ όπου $Q = \left(\frac{\theta-\hat{\theta}}{\sigma}\right)^2$ ή $L(\theta) = \ln L(max) - \frac{1}{2}Q$. Επομένως $P(\theta_a \leq \theta \leq \theta_b) = F\left(\frac{\theta_b-\hat{\theta}}{\sigma}\right) - F\left(\frac{\theta_a-\hat{\theta}}{\sigma}\right)$, όπου F είναι η συνολική ανηγμένη γκαουσιανή κατανομή (βλέπε σχετικό πίνακα). Αν κάνουμε τα διαστήματα συμμετρικά ως προς $\hat{\theta}$ για πιθανότητα γ (με ίδιες ουρές στα άκρα της κατανομής, $\frac{1}{2}(1-\gamma)$), έχουμε $P(\hat{\theta} - m\sigma \leq \theta \leq \hat{\theta} + m\sigma) = 2F(m) - 1 = \gamma$.

8.29 Μέθοδος Ελαχίστων Τετραγώνων (LS)

Η αρχή των ελαχίστων τετραγώνων:
Στα σημεία παρατήρησης $x_1, x_2, \ldots, x_N$ έχουμε ένα σύνολο από αντίστοιχες ανεξάρτητες πειραματικές τιμές $y_1, y_2, \ldots, y_N$. Οι αληθείς τιμές $u_1, u_2, \ldots, u_N$ των παρατηρήσιμων μεγεθών δεν είναι γνωστές, αλλά υποθέτουμε ότι υπάρχει κάποιο θεωρητικό μοντέλο που προβλέπει τη σωστή τιμή που σχετίζεται με κάθε x_i με βάση κάποια συνάρτηση $f_i = f_i(\theta_1, \theta_2, \ldots, \theta_L; x_i)$, $i = 1, ..., N$. Τα θ_j είναι παράμετροι και $L \leq N$. Σύμφωνα με την αρχή των ελαχίστων τετραγώνων οι άριστες τιμές των αγνώστων παραμέτρων είναι αυτές που κάνουν την έκφραση (άθροισμα τετραγώνων)

$$X^2 = \sum_{i=1}^{N} w_i(y_i - f_i)^2$$

να έχει ελάχιστη τιμή. Ο συντελεστής w_i είναι το βάρος που δίνεται στην $i-$οστή παρατήρηση. Το σύνολο των παραμέτρων $\underline{\hat{\theta}} = (\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_L)$ που δίνει την ελάχιστη τιμή του X^2 λέγεται εκτίμηση παραμέτρων ελαχίστων τετραγώνων. Το βάρος w_i εκφράζει την ακρίβεια της μέτρησης y_i. Μπορεί τα w_i να είναι όλα ίδια, οπότε η έκφραση για το άθροισμα των τετραγώνων απλοποιείται. Τότε έχουμε

$$X^2 = \sum_{i=1}^{N}(y_i - f_i)^2$$

και λέμε ότι έχουμε εκτίμηση με ελάχιστα τετράγωνα χωρίς βάρη. Συνήθως γράφουμε $w_i = \frac{1}{\sigma_i^2}$, όπου σ_i είναι τα σφάλματα μέτρησης και ελαχιστοποιούμε το

$$X^2 = \sum_{i=1}^{N}\left(\frac{y_i - f_i}{\sigma_i}\right)^2.$$

Πολλές φορές τα σ_i δεν είναι γνωστά. Πρέπει να κάνουμε κάποια εκτίμηση των επί μέρους σφαλμάτων. Ας υποθέσουμε, για παράδειγμα, ότι y_i είναι ο αριθμός των γεγονότων στην κλάση (διάστημα) i. Μπορούμε να χρησιμοποιήσουμε την προσέγγιση $\sigma_i^2 \approx f_i$, που σημαίνει ότι θεωρούμε τα αληθή u_i ως μεταβλητές

με κατανομή poisson με μέση τιμή και απόκλιση f_i, έχουμε επομένως:

$$X^2 = \sum_{i=1}^{N} \frac{(y_i - f_i)^2}{f_i}.$$

Αν η f_i είναι πολύπλοκη συνάρτηση μπορούμε για απλοποίηση των υπολογισμών να θέσουμε $\sigma_i^2 \approx y_i$, οπότε

$$X^2 = \sum_{i=1}^{N} \frac{(y_i - f_i)^2}{y_i}.$$

Έχουμε τότε απλοποιημένη εκτίμηση ελαχίστων τετραγώνων. Αν υπάρχουν συσχετίσεις μεταξύ των παρατηρήσεων, τότε ελαχιστοποιούμε την έκφραση:

$$X^2 = \sum_{i=1}^{N}\sum_{j=1}^{N}(y_i - f_i)V_{ij}^{-1}(y_j - f_j).$$

8.30 Σχέση Μεταξύ των Μεθόδων Ελαχίστων Τετραγώνων και Μέγιστης Αληθοφάνειας

Αν οι μετρήσεις y_i έχουν κανονική (γκαουσιανή) κατανομή, $N(u_i, \sigma_i^2)$, γύρω από τις αληθείς άγνωστες τιμές τους, u_i, με αποκλίσεις σ_i^2, τότε η αληθοφάνεια για την παρατήρηση της σειράς $y_1, y_2, \ldots, y_N$ είναι:

$$L = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}\sigma_i} e^{-\frac{1}{2}\left(\frac{y_i - u_i}{\sigma_i}\right)^2} = \left(\prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}\sigma_i}\right) e^{-\frac{1}{2}\sum_{i=1}^{N}\left(\frac{y_i - u_i}{\sigma_i}\right)^2}.$$

Το L έχει μέγιστο προφανώς όταν το $\sum_{i=1}^{N}\left(\frac{y_i - u_i}{\sigma_i}\right)^2$ γίνεται ελάχιστο. Αυτό είναι ισοδύναμο με την αρχή των ελαχίστων τετραγώνων με την προϋπόθεση ότι τα βάρη σχετίζονται με τα σφάλματα των επί μέρους ανεξάρτητων μετρήσεων με τις σχέσεις: $w_i = \frac{1}{\sigma_i^2}$.

8.31 Το γραμμικό Μοντέλο Ελαχίστων Τετραγώνων

Σε αυτήν την περίπτωση όπου η εξάρτηση από τις παραμέτρους είναι γραμμική και τα βάρη ανεξάρτητα από τις παραμέτρους, βρίσκουμε αναλυτικές λύσεις στο πρόβλημα. Επίσης, οι εκτιμητήρες έχουν τις θεωρητικές άριστες ιδιότητες της μοναδικότητας, της αμεροληψίας και της ελάχιστης απόκλισης. Έστω ότι οι παρατηρήσεις είναι $(y_i \pm \sigma_i), i = 1, 2, \ldots, N$ και αναφέρονται στα σημεία $x_1, x_2, \ldots, x_N$. Έστω ότι το μοντέλο είναι

$$f_i = f_i(\theta_1, \theta_2, \ldots, \theta_L; x_i) = \sum_{l=1}^{L} a_{il}\theta_l, \quad i = 1, 2, \ldots, N, L \leq N.$$

Οι συντελεστές a_{il} είναι, γενικά, συναρτήσεις των x_i. Έχουμε να ελαχιστοποιήσουμε την έκφραση:

$$X^2 = \sum_{i=1}^{N} \left(\frac{y_i - f_i}{\sigma_i} \right)^2 = \sum_{i=1}^{N} \frac{1}{\sigma_i^2} \left(y_i - \sum_{l=1}^{L} a_{il}\theta_l \right)^2 .$$

Πρέπει $\frac{\partial X^2}{\partial \theta_k} = 0, k = 1, 2, \ldots, L$ άρα

$$\sum_{l=1}^{L} \left(\sum_{i=1}^{N} \frac{a_{ik}a_{il}}{\sigma_i^2} \right) \theta_l = \sum_{i=1}^{N} \frac{a_{ik}y_i}{\sigma_i^2}, \quad k = 1, ..., L.$$

Η λύση του μη ομογενούς συστήματος των L εξισώσεων με L αγνώστους δίνει τις τιμές $\hat{\theta}_l$ των παραμέτρων θ_l.

Με συμβολισμό μητρών έχουμε:

$$\underline{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{pmatrix}, \underline{f} = \begin{pmatrix} f_1 \\ f_2 \\ \vdots \\ f_N \end{pmatrix}, \underline{\theta} = \begin{pmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_N \end{pmatrix}, V = V(\underline{y}) = \begin{pmatrix} \sigma_1^2 & 0 & 0 & 0 \\ 0 & \sigma_2^2 & 0 & 0 \\ & & \vdots & \\ 0 & 0 & 0 & \sigma_N \end{pmatrix}$$

Η μήτρα των σφαλμάτων είναι διαγώνια διότι υποθέσαμε ότι οι μετρήσεις είναι ανεξάρτητες. Επίσης

$$A = \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1L} \\ a_{21} & a_{22} & \ldots & a_{2L} \\ \vdots & \vdots & \vdots & \vdots \\ a_{N1} & a_{N2} & \ldots & a_{NL} \end{pmatrix}$$

Επομένως: $\underline{f} = A\underline{\theta}$ και η προς ελαχιστοποίηση ποσότητα είναι η $X^2 = (\underline{y} - A\underline{\theta})^T V^{-1} (\underline{y} - A\underline{\theta})$. Τελικώς παίρνοντας τις παραγώγους ως προς τα $\theta_1, \theta_2, \ldots, \theta_L$ και θέτοντάς τις ίσες με μηδέν καταλήγουμε:

$$\hat{\underline{\theta}} = (A^T V^{-1} A)^{-1} A^T V^{-1} \underline{y}.$$

Η εφαρμογή των μεθόδων διάδοσης σφαλμάτων οδηγεί στον υπολογισμό της μήτρας σφαλμάτων (συναλλοιώτων)

$$V(\hat{\underline{\theta}}) = (A^T V^{-1} A)^{-1}.$$

Οι τύποι είναι πιο γενικοί απ' ό,τι φαίνεται, διότι ισχύουν και όταν η μήτρα $V = V(\underline{y})$ δεν είναι διαγώνια, οπότε οι μετρήσεις δεν είναι ανεξάρτητες.

8.32 Ορθογώνια Πολυώνυμα

Όταν τα a_{il} είναι πολυώνυμα των x_i μεγάλου βαθμού, υπάρχουν υπολογιστικά προβλήματα με αριθμητικές ανακρίβειες στους υπολογισμούς. Μπορεί να αποφευχθούν τέτοια σφάλματα που σχετίζονται με το περιορισμένο του πλήθους

των σημαντικών ψηφίων ακόμη και με υπολογιστές, όταν χρησιμοποιούνται ορθογώνια πολυώνυμα στο γραμμικό μοντέλο. Θα θεωρήσουμε μιαν απλή περίπτωση όπου οι μετρήσεις δεν είναι συσχετισμένες και έχουν ίδιο σφάλμα. Τότε $V(\underline{y}) = \sigma^2 I_N, V^{-1}(\underline{y}) = \frac{1}{\sigma^2} I_N$, όπου η μοναδιαία μήτρα:

$$I_N = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \dots & \dots & \dots & \dots \end{pmatrix}$$

Τα πολυώνυμα είναι τα $\xi_l(x), l = 1, 2, \dots, L$ και η ορθογωνιότητά τους πάνω στις μετρήσεις x_i σημαίνει: $\sum_{i=1}^{N} \xi_k(x_i)\xi_l(x_i) = \delta_{kl}, k, l = 1, 2, \dots, L$. Το μοντέλο είναι:

$$f_i = \sum_{l=1}^{L} \xi_l(x_i)\theta_l, i = 1, 2, \dots, N.$$

Έχουμε προφανώς:

$$\underline{\hat{\theta}} = (A^T A)^{-1} A^T \underline{y}, \quad A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1L} \\ a_{21} & a_{22} & \dots & a_{2L} \\ \vdots & \vdots & \vdots & \vdots \\ a_{N1} & a_{N2} & \dots & a_{NL} \end{pmatrix},$$

$$(A)_{il} = (A^T)_{li} = a_{il} = \xi_l(x_i),$$

$$(A^T A)_{kl} = \sum_{i=1}^{N} (A^T)_{ki}(A)_{il} = \sum_{i=1}^{N} \xi_k(x_i)\xi_l(x_i) = \delta_{kl}.$$

Άρα $A^T A = I_L$, η μοναδιαία μήτρα διαστάσεων $L \times L$. Επομένως η λύση απλοποιείται στη μορφή $\underline{\hat{\theta}} = A^T \underline{y}$ ή $\hat{\theta}_l = (A^T \underline{y})_l = \sum_{i=1}^{N} \xi_l(x_i) y_i, l = 1, 2, \dots, L$ και $V(\underline{\hat{\theta}}) = A^T V(\underline{y}) A = \sigma^2 I_L$

Στην περίπτωση που δεν υπάρχει γραμμική εξάρτηση των παραμέτρων στο μοντέλο, για να βρεθεί η λύση, πρέπει να χρησιμοποιηθούν αριθμητικές τεχνικές, όπως επαναληπτικές μέθοδοι κλπ.

8.33 Ποιότητα Προσαρμογής

Ουσιαστικά οι μέθοδοι που εξετάσαμε κάνουν προσαρμογή δεδομένων σε θεωρητικά μοντέλα. Γενικά, αν οι μετρήσεις δεν είναι κατ' ανάγκην ανεξάρτητες, ακολουθούν όμως γκαουσιανές κατανομές σε πολλές διαστάσεις περί τις αληθείς τιμές τους, τότε η έκφραση

$$X^2_{min} = \sum_{i=1}^{N} \sum_{j=1}^{N} (y_i - \hat{z}_i) V^{-1}_{ij} (y_j - \hat{z}_j)$$

ακολουθεί κατανομή χ^2 με $N - L + K$ βαθμούς ελευθερίας, όπου K είναι το πλήθος των πιθανών δεσμευτικών σχέσεων που περιορίζουν τις L παραμέτρους και δεν εξετάσαμε εδώ. Αυτό έχει μια σημαντική πρακτική συνέπεια.

Μπορεί να χρησιμοποιηθεί για να ελεγχθεί η ποιότητα της προσαρμογής. Ας υποθέσουμε ότι το πρόβλημα περιλαμβάνει ν βαθμούς ελευθερίας. Λέμε τότε ότι έχουμε να κάνουμε με πρόβλημα προσαρμογής $\nu-$δεσμεύσεων ή νC (ν-constraints). Από τον σχετικό πίνακα μπορούμε να βρούμε την πιθανότητα $P_{\chi^2} = \int_{X^2_{min}}^{\infty} f(u,\nu)du = 1 - F(X^2_{min},\nu)$, όπου $u = \chi^2$. Αυτή η ποσότητα είναι η πιθανότητα, σε μια νέα ελαχιστοποίηση με παρόμοιες μετρήσεις και το ίδιο μοντέλο, να βρεθεί μεγαλύτερη τιμή για το X^2_{min}. Μικρή τιμή του X^2_{min} αντιστοιχεί σε μεγάλο P_{χ^2}, ή καλή προσαρμογή, ενώ πολύ μεγάλη τιμή του X^2_{min} αντιστοιχεί σε μικρό P_{χ^2}, δηλαδή κακή προσαρμογή.

Βιβλιογραφία

1. Probability and Statistics in Particle Physics, A.G.Frodesen, O. Skjeggestad, H.Toeffe, Columbia University Press, 1978.

2. Statistical Methods in Experimental Physics, W.T. Earlie, D.Drijard, F.G.James,M.Roos, B.Sadonlat, North Holland, 1971.

3. Statistics for Nuclear and Particle Physicists, L.Lyons, Cambridge University Press, 1986.

4. Probability, Statistics and Monte Carlo, in Reviews of Particle Physics, The European Physical Journal, C15 (2000) 191-204.

n	$\chi^2_{.995}$	$\chi^2_{.99}$	$\chi^2_{.975}$	$\chi^2_{.95}$	$\chi^2_{.90}$	$\chi^2_{.75}$	$\chi^2_{.50}$	$\chi^2_{.25}$	$\chi^2_{.10}$	$\chi^2_{.05}$	$\chi^2_{.025}$	$\chi^2_{.01}$	$\chi^2_{.005}$
1	7.38	6.63	5.02	3.84	2.71	1.32	.455	.102	.0158	.0039	.0010	.0002	.0000
2	10.6	9.21	7.38	5.99	4.61	2.77	1.39	.575	.211	.103	.0506	.0201	.0100
3	12.8	11.3	9.35	7.81	6.25	4.11	2.37	1.21	.584	.352	.216	.115	.072
4	14.9	13.3	11.1	9.49	7.78	5.39	3.36	1.92	1.06	.711	.484	.297	.207
5	16.7	15.1	12.8	11.1	9.24	6.63	4.35	2.67	1.61	1.15	.831	.554	.412
6	18.5	16.8	14.4	12.6	10.6	7.84	5.35	3.45	2.20	1.64	1.24	.872	.676
7	20.3	18.5	16.0	14.1	12.0	9.04	6.35	4.25	2.83	2.17	1.69	1.24	.989
8	22.0	20.1	17.5	15.5	13.4	10.2	7.34	5.07	3.49	2.73	2.18	1.65	1.34
9	23.6	21.7	19.0	16.9	14.7	11.4	8.34	5.90	4.17	3.33	2.70	2.09	1.73
10	25.2	23.2	20.5	18.3	16.0	12.5	9.34	6.74	4.87	3.94	3.25	2.56	2.16
11	26.8	24.7	21.9	19.7	17.3	13.7	10.3	7.58	5.58	4.57	3.82	3.05	2.60
12	28.3	26.2	23.3	21.0	18.5	14.8	11.3	8.44	6.30	5.23	4.40	3.57	3.07
13	29.8	27.7	24.7	22.4	19.8	16.0	12.3	9.30	7.04	5.89	5.01	4.11	3.57
14	31.3	29.1	26.1	23.7	21.1	17.1	13.3	10.2	7.79	6.57	5.63	4.66	4.07
15	32.8	30.6	27.5	25.0	22.3	18.2	14.3	11.0	8.55	7.26	6.26	5.23	4.60
16	34.3	32.0	28.8	26.3	23.5	19.4	15.3	11.9	9.31	7.96	6.91	5.81	5.14
17	35.7	33.4	30.2	27.6	24.8	20.5	16.3	12.8	10.1	8.67	7.56	6.41	5.70
18	37.2	34.8	31.5	28.9	26.0	21.6	17.3	13.7	10.9	9.39	8.23	7.01	6.26
19	38.6	36.2	32.9	30.1	27.2	22.7	18.3	14.6	11.7	10.1	8.91	7.63	6.84
20	40.0	37.6	34.2	31.4	28.4	23.8	19.3	15.5	12.4	10.9	9.59	8.26	7.43
21	41.4	38.9	35.5	32.7	29.6	24.9	20.3	16.3	13.2	11.6	10.3	8.90	8.03
22	42.8	40.3	36.8	33.9	30.8	26.0	21.3	17.2	14.0	12.3	11.0	9.54	8.64
23	44.2	41.6	38.1	35.2	32.0	27.1	22.3	18.1	14.8	13.1	11.7	10.2	9.26
24	45.6	43.0	39.4	36.4	33.2	28.2	23.3	19.0	15.7	13.8	12.4	10.9	9.89
25	46.9	44.3	40.6	37.7	34.4	29.3	24.3	19.9	16.5	14.6	13.1	11.5	10.5
26	48.3	45.6	41.9	38.9	35.6	30.4	25.3	20.8	17.3	15.4	13.8	12.2	11.2
27	49.6	47.0	43.2	40.1	36.7	31.5	26.3	21.7	18.1	16.2	14.6	12.9	11.8
28	51.0	48.3	44.5	41.3	37.9	32.6	27.3	22.7	18.9	16.9	15.3	13.6	12.5
29	52.3	49.6	45.7	42.6	39.1	33.7	28.3	23.6	19.8	17.7	16.0	14.3	13.1
30	53.7	50.9	47.0	43.8	40.3	34.8	29.3	24.5	20.6	18.5	16.8	15.0	13.8
40	66.8	63.7	59.3	55.8	51.8	45.6	39.3	33.7	29.1	26.5	24.4	22.2	20.7
50	79.5	76.2	71.4	67.5	63.2	56.3	49.3	42.9	37.7	34.8	32.4	29.7	28.0
60	92.0	88.4	83.3	79.1	74.4	67.0	59.3	52.3	46.5	43.2	40.5	37.5	35.5
70	104.2	100.4	95.0	90.5	85.5	77.6	69.3	61.7	55.3	51.7	48.8	45.4	43.3
80	116.3	112.3	106.6	101.9	96.6	88.1	79.3	71.1	64.3	60.4	57.2	53.5	51.2
90	128.3	124.1	118.1	113.1	107.6	98.6	89.3	80.6	73.3	69.1	65.6	61.8	59.2
100	140.2	135.8	129.6	124.3	118.5	109.1	99.3	90.1	82.4	77.9	74.2	70.1	67.3

Πίνακας 8.1:

n	$t_{.005}$	$t_{.01}$	$t_{.025}$	$t_{.05}$	$t_{.10}$	$t_{.20}$	$t_{.25}$	$t_{.30}$	$t_{.40}$	$t_{.45}$
1	63.66	31.82	12.71	6.31	3.08	1.376	1.000	.727	.325	.158
2	9.92	6.96	4.30	2.92	1.89	1.061	.816	.617	.289	.142
3	5.84	4.54	3.18	2.35	1.64	.978	.765	.584	.277	.137
4	4.60	3.75	2.78	2.13	1.53	.941	.741	.569	.271	.134
5	4.03	3.36	2.57	2.02	1.48	.920	.727	.559	.267	.132
6	3.71	3.14	2.45	1.94	1.44	.906	.718	.553	.265	.131
7	3.50	3.00	2.36	1.90	1.42	.896	.711	.549	.263	.130
8	3.36	2.90	2.31	1.86	1.40	.889	.706	.546	.262	.130
9	3.25	2.82	2.26	1.83	1.38	.883	.703	.543	.261	.129
10	3.17	2.76	2.23	1.81	1.37	.879	.700	.542	.260	.129
11	3.11	2.72	2.20	1.80	1.36	.876	.697	.540	.260	.129
12	3.06	2.68	2.18	1.78	1.36	.873	.695	.539	.259	.128
13	3.01	2.65	2.16	1.77	1.35	.870	.694	.538	.259	.128
14	2.98	2.62	2.14	1.76	1.34	.868	.692	.537	.258	.128
15	2.95	2.60	2.13	1.75	1.34	.866	.691	.536	.258	.128
16	2.92	2.58	2.12	1.75	1.34	.865	.690	.535	.258	.128
17	2.90	2.57	2.11	1.74	1.33	.863	.689	.534	.257	.128
18	2.88	2.55	2.10	1.73	1.33	.862	.688	.534	.257	.127
19	2.86	2.54	2.09	1.73	1.33	.861	.688	.533	.257	.127
20	2.84	2.53	2.09	1.72	1.32	.860	.687	.533	.257	.127
21	2.83	2.52	2.08	1.72	1.32	.859	.686	.532	.257	.127
22	2.82	2.51	2.07	1.72	1.32	.858	.686	.532	.256	.127
23	2.81	2.50	2.07	1.71	1.32	.858	.685	.532	.256	.127
24	2.80	2.49	2.06	1.71	1.32	.857	.685	.531	.256	.127
25	2.79	2.48	2.06	1.71	1.32	.856	.684	.531	.256	.127
26	2.78	2.48	2.06	1.71	1.32	.856	.684	.531	.256	.127
27	2.77	2.47	2.05	1.70	1.31	.855	.684	.531	.256	.127
28	2.76	2.47	2.05	1.70	1.31	.855	.683	.530	.256	.127
29	2.76	2.46	2.04	1.70	1.31	.854	.683	.530	.256	.127
30	2.75	2.46	2.04	1.70	1.31	.854	.683	.530	.256	.127
40	2.70	2.42	2.02	1.68	1.30	.851	.681	.529	.255	.126
60	2.66	2.39	2.00	1.67	1.30	.848	.679	.527	.254	.126
120	2.62	2.36	1.98	1.66	1.29	.845	.677	.526	.254	.126
∞	2.58	2.33	1.96	1.645	1.28	.842	.674	.524	.253	.126

Πίνακας 8.2:

x	0	1	2	3	4	5	6	7	8	9
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5754
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7258	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7518	.7549
0.7	.7580	.7612	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7996	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998
3.5	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998
3.6	.9998	.9998	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999
3.7	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999
3.8	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999
3.9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Πίναχας 8.3: