FLORIDA INTERNATIONAL UNIVERSITY
Mechanical Engineering Department

Summer 1993 Advanced Analysis of Mechanical Systems EGM 6422

COURSE CONTENT

Review of Linear, Algebraic Sets of Equations Chapt 2 4 PER
- Direct Methods
- Iterative Methods
- Vector and Matrix Norm Definitions

Nonlinear Algebraic Sets of Equations Chapt 2 1 PER

Ordinary Differential Equations Chapt 5 5 PER
- Single step methods available
- Multistep methods available

Introduction to PDE's
- What PDE's characterize
- Classification of PDE's
- Finite Difference Notation

Parabolic Differential Equations Chapt 8 6 PER
- Explicit and Implicit Methods
- Initial and Boundary Conditions, Limiting Conditions
- Stability and Consistency
- Well Posedness and Sufficiency
- Lax–Wendroff statement of Equality (Laws of Conservation)
- Second Order Parabolic Equations and Schemes to Solve Them

Elliptic PDE's Chapt 7 6 PER
- Laplace Equation and Iterative Schemes to Solve It
- Poisson Equation
- Dealing with Limiting Conditions of Boundary Conditions
- NonCartesian Meshes and NonRegular Regions

Hyperbolic PDE's Chapt 9 4 PER
- Implicit and Explicit Schemes
- Problems of Stability in the Schemes
- Courant/Frederick Levy Condition
- Numerical Integration along Characteristics
- Second Order Equations

Text:

C.F. Gerald, <u>Applied Numerical Analysis</u>, Addison–Wesley, 3rd Ed.

References:

Elementary Numerical Analysis by Conte and deBoor, McGraw–Hill Publishers
Applied Numerical Methods by Carnahan, Krieger Publishers
Advanced Calculus for Applications by Hildebrand, Prentice–Hall Publishers
Applied Numerical Methods for Digital Computations by James, Smith and Wolford, 4th Edition, Harper Collins Publishers

Jul 1	—	
Jul 6	Ch 2	
7	"	
8	"	4
[illegible]	↓	
Jul 12	Ch 2	
13	Ch 5	
14	5	
15	5	4
Jul 19	5	
20	—	
21	Ch 8	
22	8	

Jul 26	8	
27	8	6
28	8	
29	8	
Aug 2	8	
3	Ch 7	
4	7	6
5	7	
Aug 9	7	
10	7	
11	7	
12	Ch 9	
Aug 16	9	4
17	9	
18	9	

Grade will be determined on the basis of

1 Exam 30 % each
HW/Project 30 %
Final Exam 40 %

Letter Grades will be based as follows

(A) 90 & above
(A–) 87–89
(B+) 83–86
(B) 80–82
(B–) 77–79
(F) below 60
(C+) 73–76
(C) 70–72
(C–) 67–69
(D+) 63–66
(D) 60–62

<u>This is a preliminary syllabus subject to change.</u>

Please be on time to class. Room is CP 101. Class may be cancelled 1 July because I am on Jury Duty. If so, we will meet on 6 July in class. Please read Chapter 2 of your books and be prepared to discuss.

תכנית לימודים

מס׳	הנושא	מס׳ שעות הלימוד	הפרקים בספרי הלימוד [1]	הפרקים בספרי הלימוד [2]
1	**מערכות משוואות אלגבריות לינאריות:** שיטות ישירות, שיטות איטרטיביות, נורמים של וקטורים ומטריצות, השגיאות בפתרון מערכות משוואות. Sets of Linear Algebraic ~~Matrix~~ Eqn.	4	5	2
2	מערכות משוואות אלגבריות לא לינאריות Nonlinear Algebraic ~~Matrix~~, Sets of Eqs.	1	–	2
3	**משוואות דיפרנציאליות רגילות:** שיטות צעד אחד ושיטות מרובות צעד Regular Diff Eq one step multistep	4	–	5
4	מבוא למשוואות דיפרנציאליות חלקיות Intro to PDE	1	1	–
5	**משוואות פרבוליות:** שיטות מפורשות וסתומות, תנאי גבול של נגזרות. מושגי ההתכנסות היציבות והעקביות, משפט השוויוניות של Lax. משוואות פרבוליות מסדר שני וסכמות שונות לפתרונן. Parab explicit implicit method Conditions on boundary of cuts concentration equality consistency diff Scheme to solve them.	7	2	8
6	**משוואות אליפטיות:** משוואת לפלס ושיטות איטרטיביות לפתרונה, משוואת פואסון. טיפול בתנאי גבול של נגזרות, סריגים לא קרטזיים ותחומים לא רגולריים. Ellip LAPLACE Poisson Fix based on boundary of cuts Noncartesian meshes Non regular regions	6	5	7
7	**משוואות היפרבוליות:** סכמות מפורשות וסתומות, בעיות היציבות של הסכמות. תנאי קורנט פרדריק לוי. אינטגרציה נומרית לאורך קרקטריסטיקות. משוואות מסדר שני. Hyperbolic Implicit/Exp. Schem in the Scheme Problems of Courant Cond. & Fred. Lewy Numerical Int. along characteristics 2nd order eqs.	4	4	9

MATRIX SOL

NON LIN

ODE

PDE intro

PDE paral

2nd order Parab eq

PDE ellip

PDE hyper

ספרי לימוד:

[1] G.D. Smith: Numerical Solution of Partial Differential Equations: Finite Difference Method, Oxford University Press, 3rd ed., 1985.

[2] C.F. Gerald: Applied Numerical Analysis, Addison-Wesley, 2nd ed., 1978.

(iii) Alternate from line to line by first using Saul'yev A and then B, or the reverse. This is related to *alternating direction methods* to be discussed later.

(iv) Use Saul'yev A and Saul'yev B on the same line and average the results for the final answer (A first, and then B). This is equivalent to introducing the dummy variables $P_{i,j}$ and $Q_{i,j}$ such that

$$(1+r)P_{i,j+1} = U_{i,j} + r(P_{i-1,j+1} - U_{i,j} + U_{i+1,j}), \quad (2\text{-}268a)$$

$$(1+r)Q_{i,j+1} = U_{i,j} + r(Q_{i+1,j+1} - U_{i,j} + U_{i-1,j}), \quad (2\text{-}268b)$$

and

$$U_{i,j+1} = \tfrac{1}{2}(P_{i,j+1} + Q_{i,j+1}). \quad (2\text{-}269)$$

This averaging method has some computational advantage because of the possibility of truncation error cancellation.

As an alternative to Eqns. (2-268a, b) one can retain the $P_{i,j}$ and $Q_{i,j}$ from the previous step and replace $U_{i,j}$, $U_{i+1,j}$ in (2-268a) by $P_{i,j}$ and $P_{i+1,j}$, respectively, and $U_{i,j}$, $U_{i-1,j}$ in (2-268b) by $Q_{i,j}$ and $Q_{i-1,j}$.

Liu [88] used higher order approximations and obtained the algorithms (Liu A)

$$(3r+2)U_{i,j+1} = 2(1-2r)U_{i,j}$$
$$+ r(U_{i-1,j} + 3U_{i+1,j} - U_{i-2,j+1} + 4U_{i-1,j+1}),$$

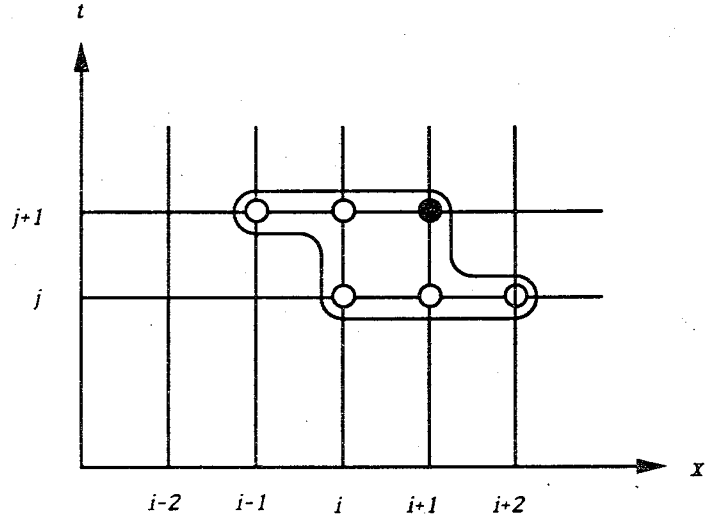


Figure 2-18(a) Molecule of Liu A: L to R

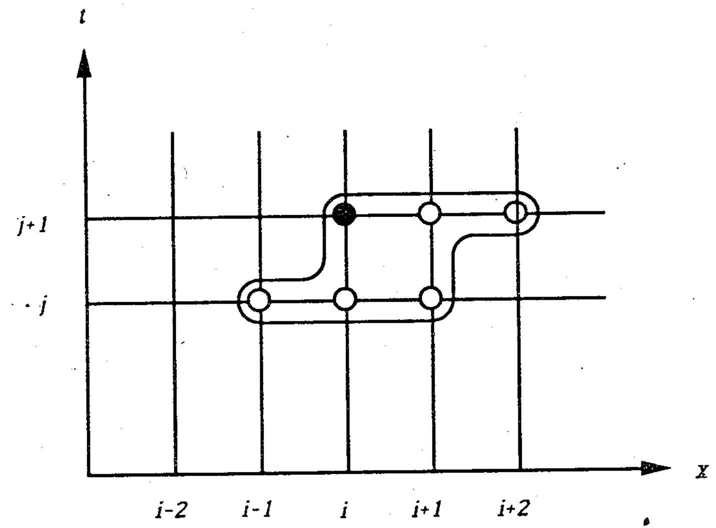


Figure 2-18(b) Molecule of Liu B: R to L

with the computational molecule shown in (Fig. 2-18a). Liu B is given by

$$(3r+2)U_{i,j+1} = 2(1-2r)U_{i,j}$$
$$+ r(U_{i+1,j} + 3U_{i-1,j} - U_{i+2,j+1} + 4U_{i+1,j+1}).$$

These are analogous to the Saul'yev schemes except that the first point on any line (either from the left or the right) must be obtained by some other means. As with the Saul'yev forms, combinations can be used.

PROBLEMS

2-58 Develop the algorithm, Eqn. (2-267), for Saul'yev B.

2-59 Show that the Saul'yev A approximation is unconditionally stable by the Fourier method. *Hint:* Use $U_{i,j} = \xi^j \exp(\sqrt{-1}\,\beta ih)$.

2-60 Show that the Liu A method is always stable.

2-61 Using Taylor series expansions for $u_t = u_{xx}$ show that the truncation error for Saul'yev A is

$$\text{TE} = \frac{k}{h}u_{xxx} - \frac{h^2}{12}u_{xxxx} - \frac{k^2}{12}u_{xxxxxx},$$

and for Saul'yev B is

$$\text{TE} = -\frac{k}{h}u_{xxx} - \frac{h^2}{12}u_{xxxx} - \frac{k^2}{12}u_{xxxxxx},$$

where $k = \Delta t$ and $h = \Delta x$.

1) $X T' - X'' T = 0$ $\quad \frac{T'}{T} = \frac{X''}{X} = -k^2$ $\quad T' + k^2 T = 0$ $\quad T = Ce^{-k^2 t}$

$X(0) = X(2) = 0$ $\quad X'' + k^2 X = 0$ $\quad X = A\sin kx + B\cos kx$

$0 = B$

$X(2) = A \sin k \cdot 2 = 0$

$\sin 2k = 0$

$2k = n\pi$ or

$k = \frac{n\pi}{2}$

$$T = Ce^{-\frac{n^2\pi^2}{4}t}$$

$$X = A \sin \frac{n\pi}{2} x$$

$$TX = \mathscr{C} e^{-\frac{n^2\pi^2}{4}t} \sin \frac{n\pi}{2} x$$

$$x(t) = e^{-\pi^2/4\, t} \sin \pi/2\, x$$

1) Using a backward difference implicit scheme solve

$$\frac{\partial u}{\partial t} - \frac{\partial^2 u}{\partial x^2} = 0 \qquad t \le 0.1$$

$$u(0,t) = 0 \quad u(2,t) = 0$$

$$u(x,0) = \sin \frac{\pi x}{2} \quad \text{using} \quad \Delta x = .5 \quad \Delta t = .05$$

(25)

2) Find the actual solution using separation of variables and compare the results of this solution to the finite difference solution

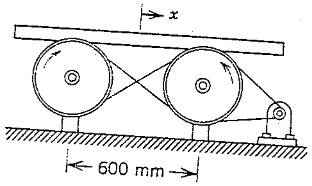
3. Solve using Runge Kutta method

$$d^2y/dx^2 + y = 0 \qquad \text{for} \quad 0 \le x \le \pi \qquad y(0) = 1,\ y(\pi) = -1$$

use $\Delta x = \pi/10$

obtain the closed form solution & compare.

4. Solve problem 26 p. 403

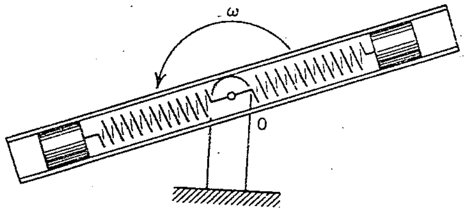


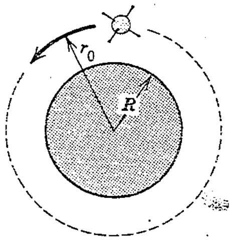
PROBLEM 2.24 A device designed to measure the kinetic coefficient of friction consists of two 90° Vee-grooved pulleys, rotating in opposite directions, across which a cylindrical bar of some known material is placed. When displaced, the bar will perform simple harmonic motion. Derive an expression for the kinetic coefficient of friction μ in terms of the frequency of vibration in cycles per second.

Answer: $\mu = 0.854 f^2$

PROBLEM 2.25 Two sliders are constrained to move within a smooth tube that is rotating in the horizontal plane about the fixed axis *0*. Each of the sliders is elastically suspended from identical springs with a modulus k. The ends of the spring are fixed at *0* and the unstretched length of the spring is r_0. Determine the frequency of vibration for a constant angular velocity ω.

Answer: $f_n = \dfrac{1}{2\pi}\sqrt{\dfrac{k}{m} - \omega^2}$





PROBLEM 2.26 A small orbiting body is displaced from its circular orbit a small distance δ. Determine the period and the equation of motion for the disturbance.

Hint: Kepler's law is that the radial line to an orbiting body sweeps equal areas of space in equal times ($r^2\dot{\theta} = \text{constant}$).

Answer: $\tau = 2\pi\sqrt{\dfrac{r_0^3}{gR^2}}$

2.4. TORSIONAL VIBRATION

Torsional vibration refers to vibration of a rigid body about a specific reference axis. In this case, displacement is measured in terms of an angular coordinate. The restoring moment may be either due to the torsion of an elastic member or to the unbalanced moment of a force or a couple.

In Figure 2.3, one end of a long rod supports a disk that has mass, which is large relative to that of the rod, and the other end of the rod is fixed to a rigid foundation. If the rod is elastic, any angular displacement of the disk away from the equilibrium position will create a restoring moment of

$$M = \frac{JG\theta}{l} \tag{2.12}$$

where J is the second moment of area (polar moment of inertia of area) about the axis of the shaft, G is the modulus of rigidity, l is the length of shaft, and θ is the angular coordinate measure of the displacement of the disk about the axis of the rod. This restoring moment is linearly proportional to the angle θ and the constant of proportionality is defined as the torsional spring constant.

$$K = \frac{M}{\theta} = \frac{JG}{l} \tag{2.13}$$

The symbol for the torsional spring constant is K and the units for the torsional spring constant are torque per unit of angular displacement, Newton metres per radian (N·m/rad). Taking the moment sum about the axis of the rod, Newton's second law of motion can be stated as

$$\sum \mathbf{M}_0 = I_0\ddot{\boldsymbol{\theta}}$$
$$-K\theta = I_0\ddot{\theta}$$

The restoring moment is $K\theta$. The equation of motion is

$$\ddot{\theta} + \frac{K}{I_0}\theta = 0 \tag{2.14}$$

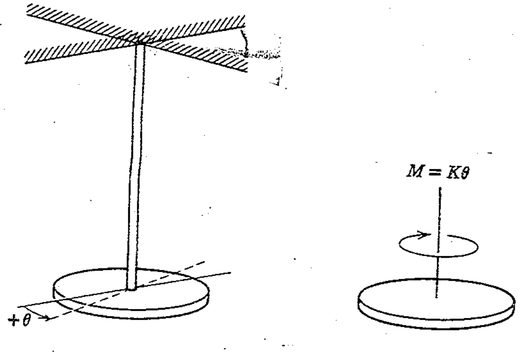


Fig. 2.3

(iii) Alternate from line to line by first using Saul'yev A and then B, or the reverse. This is related to *alternating direction methods* to be discussed later.

(iv) Use Saul'yev A and Saul'yev B on the same line and average the results for the final answer (A first, and then B). This is equivalent to introducing the dummy variables $P_{i,j}$ and $Q_{i,j}$ such that

$$(1 + r)P_{i,j+1} = U_{i,j} + r(P_{i-1,j+1} - U_{i,j} + U_{i-1,j}), \quad (2\text{-}268a)$$

$$(1 + r)Q_{i,j+1} = U_{i,j} + r(Q_{i+1,j+1} - U_{i,j} + U_{i+1,j}), \quad (2\text{-}268b)$$

and

$$U_{i,j+1} = \tfrac{1}{2}(P_{i,j+1} + Q_{i,j+1}). \quad (2\text{-}269)$$

This averaging method has some computational advantage because of the possibility of truncation error cancellation.

As an alternative to Eqns. (2-268a, b) one can retain the $P_{i,j}$ and $Q_{i,j}$ from the previous step and replace $U_{i,j}$, $U_{i+1,j}$ in (2-268a) by $P_{i,j}$ and $P_{i+1,j}$, respectively, and $U_{i,j}$, $U_{i-1,j}$ in (2-268b) by $Q_{i,j}$ and $Q_{i-1,j}$.

Liu [88] used higher order approximations and obtained the algorithms (Liu A)

$$(3r + 2)U_{i,j-1} = 2(1 - 2r)U_{i,j}$$
$$+ r(U_{i-1,j} + 3U_{i+1,j} - U_{i-2,j+1} + 4U_{i-1,j+1}).$$

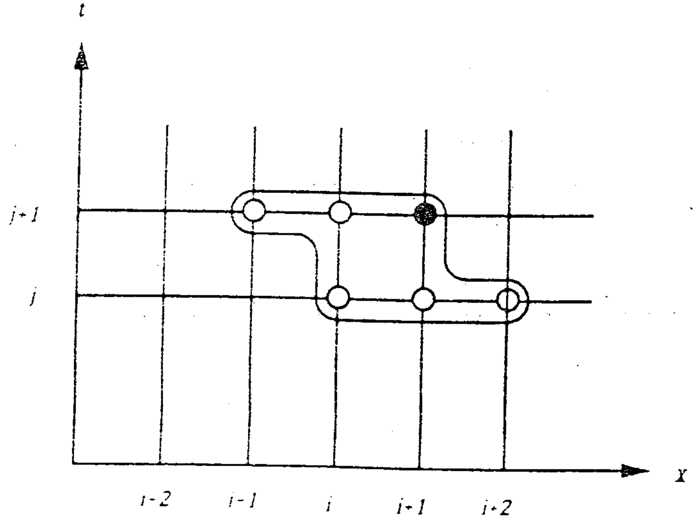


Figure 2-18(a) Molecule of Liu A: L to R

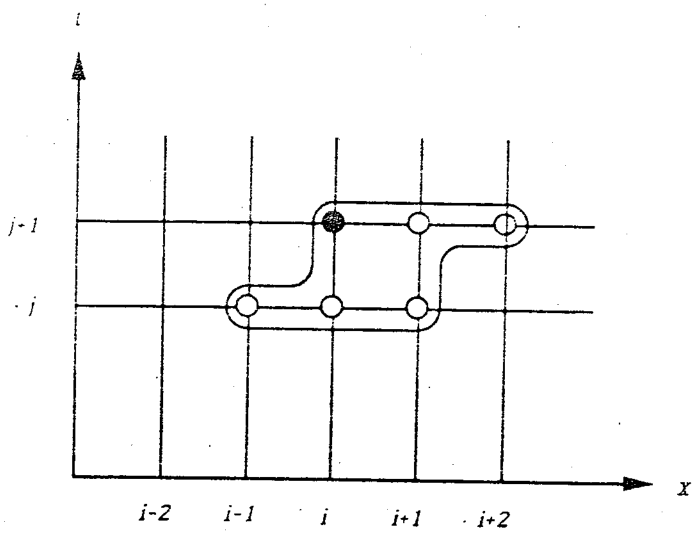


Figure 2-18(b) Molecule of Liu B: R to L

with the computational molecule shown in (Fig. 2-18a). Liu B is given by

$$(3r + 2)U_{i,j+1} = 2(1 - 2r)U_{i,j}$$
$$+ r(U_{i+1,j} + 3U_{i-1,j} - U_{i+2,j+1} + 4U_{i+1,j+1}).$$

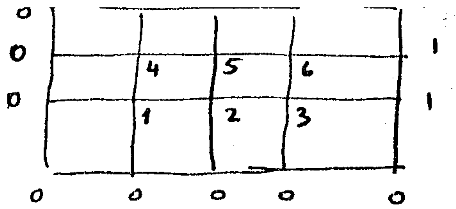
These are analogous to the Saul'yev schemes except that the first point on any line (either from the left or the right) must be obtained by some other means. As with the Saul'yev forms, combinations can be used.

YH 749

H 750 Applied Numerical Analysis by Gerald & Wheatley

751 An Intro to Numerical Computations by Yakowitz & Szidarovszky

H 752 Introductory Engineering Modeling by Rieder & Busby



$$T_2 = \frac{1}{4}\left(T_1 + T_3 + 0 + T_5\right)$$
$$T_3 = \frac{1}{4}\left(T_2 + 1 + T_6 + 0\right)$$
$$T_4 = \frac{1}{4}\left(T_1 + T_5 + 0 + 1\right)$$
$$T_5 = \frac{1}{4}\left(T_4 + T_2 + T_6 + 1\right)$$
$$T_6 = \frac{1}{4}\left(T_5 + 1 + T_3 + 1\right).$$

$$\begin{bmatrix} 4 & -1 & 0 & -1 & 0 & 0 \\ -1 & 4 & -1 & 0 & -1 & 0 \\ 0 & -1 & 4 & 0 & 0 & -1 \\ -1 & 0 & 0 & 4 & -1 & 0 \\ 0 & -1 & 0 & -1 & 4 & -1 \\ 0 & 0 & -1 & 0 & -1 & 4 \end{bmatrix} \begin{bmatrix} T_1 \\ T_2 \\ T_3 \\ T_4 \\ T_5 \\ T_6 \end{bmatrix} = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 1 \\ 1 \\ 2 \end{pmatrix}$$

$\nabla^2 T = 0$ (steady state heat conduction)

a_{ij}

when $i=j$ diagonal $\quad i>j$ subdiag

$i<j$ superdiagonal

- if for $i \neq j$ $a_{ij}=0$ then the matrix is diagonal if $\underset{\sim}{A}$ is square

  if for $i>j$ $a_{ij}=0$ upper triangular

  $i<j$ $a_{ij}=0$ lower triangular

- The identity matrix is a square matrix for which $I_{ij} = \begin{matrix} 0 & i \neq j \\ 1 & i=j \end{matrix}$

- $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = I$

- $I \cdot A = A \cdot I$ if A is square

- Like multiplication if $ax = b$ $\quad x = b/a =$ ~~b/a~~ $a^{-1}b$

- $\underset{\sim}{A}\underset{\sim}{x} = \underset{\sim}{b}$ has a solution if $\underset{\sim}{A}$ is square & determinant $A \neq$

  $\Rightarrow \underset{\sim}{x} = \underset{\sim}{A}^{-1}\underset{\sim}{b}$

Then $\underset{\sim}{A}^{-1}$ is defined so that $\underset{\sim}{A} \cdot \underset{\sim}{A}^{-1} = \underset{\sim}{I}$ or $\underset{\sim}{A}^{-1}\underset{\sim}{A} = \underset{\sim}{I}$

- How to find the inverse of a square matrix

  is it necessary to do so to find the solution to $\underset{\sim}{A}\underset{\sim}{x} = \underset{\sim}{b}$

- Eliminating Unknowns. for 2 unknowns
- Use of determinants (Cramer's Rule) for 3 or 2 unknowns
- Matrix methods

A matrix having m rows & n columns is defined by

$$\underset{\sim}{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & & a_{mn} \end{bmatrix}$$

entries are a_{ij}

we say $\underset{\sim}{A}$ is an $m \times n$ matrix or of order $m \times n$

- if $m = n$ then $\underset{\wedge}{A}$ is a square matrix of order n

A vector is a matrix with 1 column

$$\underset{\sim}{b} = \begin{bmatrix} b_{11} \\ \vdots \\ b_{p1} \end{bmatrix} = \begin{bmatrix} b_1 \\ \vdots \\ b_p \end{bmatrix} = b_i$$

if $\underset{\sim}{A} = \underset{\sim}{B} \Rightarrow a_{ij} = b_{ij}$

Two matricies may be mult. if the

$$\underset{\sim}{A} \cdot \underset{\sim}{B} = [m \times n]\,[p \times q] \quad \text{\#columns of } \underset{\sim}{A} = \text{\#rows of } \underline{B} \quad (n = p)$$

$$= \underset{\sim}{C}\ [m \times q]$$

$$C_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj}$$

$$\begin{bmatrix} 3 & 2 & 1 \\ -5 & -4 & -6 \end{bmatrix} \begin{bmatrix} -1 & 5 & 8 & 0 \\ 2 & 7 & -13 & 2 \\ 0 & -2 & +5 & 1 \end{bmatrix}$$

2×3 $\quad 3 \times 4 \quad \Rightarrow 2 \times 4$

- in general $\underset{\sim}{A} \cdot \underset{\sim}{B} \neq \underset{\sim}{B} \cdot \underset{\sim}{A}$

exaple. $C_{23} = \sum_{K=1}^{3} a_{2K} b_{K3} = 5 \cdot 8 - 4 \cdot 13 + 6 \cdot 5 = 18$

$$c_{ij} = \sum_{l=1}^{n} a_{il} b_{lj}$$

$i = 1 \ldots, m$
$j = 1, \ldots, q$

- in general if $\underset{\sim}{A} \cdot \underset{\sim}{B}$ is defined $\underline{B} \cdot \underset{\sim}{A}$ may not be
- LINEAR ALGEBRAIC EQNS CAN BE WRITTEN AS

$$A\underset{\sim}{x} = \underset{\sim}{b}$$

$\underset{\sim}{x}$ is a vector of unknowns

- also $\underline{A}(\underline{B} \cdot \underline{C}) = (\underset{\sim}{A} \cdot \underline{B}) \cdot \underset{\sim}{C}$

§ 2.3

Gauss Elimination
$$a_{11}x_1 + a_{12}x_2 + a_{13}x_3 = b_1 = a_{14}$$
$$a_{21}x_1 + a_{22}x_2 + a_{23}x_3 = b_2 = a_{24}$$
$$a_{31}x_1 + a_{32}x_2 + a_{33}x_3 = b_3 = a_{34}$$

$K=1$

$$\rightarrow \quad x_1 + \frac{a_{12}}{a_{11}}x_2 + \frac{a_{13}}{a_{11}}x_3 = \frac{a_{14}}{a_{11}}$$

$(1) \times \left(\frac{-a_{21}}{a_{11}}\right) + \text{row}(2)$ $\qquad$ $(1) \times \left(\frac{-a_{31}}{a_{11}}\right) + \text{row}(3)$

$$0 + \left(a_{22} - a_{21}\frac{a_{12}}{a_{11}}\right)x_2 + \left(a_{23} - a_{21}\frac{a_{13}}{a_{11}}\right)x_3 = a_{24} - a_{21}\frac{a_{14}}{a_{11}}$$
$$0 + \left(a_{32} - a_{31}\frac{a_{12}}{a_{11}}\right)x_2 + \left(a_{33} - a_{31}\frac{a_{13}}{a_{11}}\right)x_3 = a_{34} - a_{31}\cdot\frac{a_{14}}{a_{11}}$$

$$\boxed{a_{ij} - a_{iK}\frac{a_{Kj}}{a_{KK}} = a_{ij}^{new}}$$

$i \geq K+1$
$j \geq K+1$

~~$a_{ij}^{new} = a_{ij}/a_{KK}$~~

$i = K$
$j \geq K$

New matrix
$$x_1 + a''_{12}x_2 + a'_{13}x_3 = a'_{14}$$

$K=2$
$$a'_{22}x_2 + a'_{23}x_3 = a'_{24} \rightarrow \quad a''_{Kj} = a'_{Kj}/a'_{KK}$$
$$x_2 + a'_{23}/a'_{22}\, x_3 = a'_{24}/a'_{22}$$
$$a'_{32}x_2 + a'_{33}x_3 = a'_{34}$$

$(2) \times \left(\frac{-a'_{32}}{a'_{22}}\right) + (3)$

$$= a'_{33} - a'_{32}\left(\frac{a'_{23}}{a'_{22}}\right)x_3 = a'_{34} - a'_{32}\left(\frac{a'_{24}}{a'_{22}}\right)$$

$$a'_{11}x_1 + a'_{12}x_2 + a'_{13}x_3 = a'_{14}$$
$$a''_{22}x_2 + a''_{23}x_3 = a''_{24}$$
$$+ a''_{33}x_3 = a''_{34} \longrightarrow \quad x_3 = \frac{a_{34}}{a_{33}}.$$

$$\left[\begin{array}{ccc:c} \bullet & \bullet & \bullet & \bullet \\ 0 & \bullet & \bullet & \bullet \\ 0 & 0 & \bullet & \bullet \end{array}\right]$$

$$x_2 = \left(a_{24} - \sum_{j}^{n} a_{2j}x_j\right)/a_{22}$$
$$x_1 = \left(a_{14} - a_{12}x_2 + a_{13}x_3\right)/a_{11}$$
$$x_n = a_{nm}/a_{nn}$$
$$\boxed{x_i = \left(a_{im} - \sum_{j=i+1}^{n} a_{ij}x_j\right)/a_{ii}}$$

$i = n-1, n-2, \ldots, 1$

$DET = a_{11}a'_{22}$

- pivot elements at each step is the elemet on diagonal
- Eliminated to an upper triangular matrix
- what happens if one of the pivot elements is zero

PIVOTAL STRATEGY & round off error.

- IF any element on the diagonal is <u>zero</u> then the matrix is not invertible

the determinant of the upper triang square matrix $= a_{11}a_{22}a_{33} \dots a_{nn}$

" " " a lower triangular square matrix $= \prod_{i=1}^{n} a_{ii}$

- the solution of $\underset{\sim}{A}\,\underset{\sim}{x} = \underset{\sim}{b}$ & $\underset{\sim}{A}'\,\underset{\sim}{x} = \underset{\sim}{b}'$ are the same if $\underset{\sim}{A}^{-1}$ exists and

$\underset{\sim}{A}'$ is obtained by (1) multiplication of one row by a non zero constant

(2) addition of two rows

(3) interchange of two equations.

$\text{DET} = \prod_{i=1}^{n} a_{ii}\,(-1)^q$

q = # of interchanges

- define the augmented matrix $[a_{ij} \,\vdots\, b_i]$

initialize the $\underset{\sim}{p}$ vector so that $p_i = i$

```
FOR  k = 1, ...., n-1  do:                     k is the pivotal row
    find the smallest row number $j \geq k$ so that $a_{jk} \neq 0$
    IF no such j exists  matrix is not invertible & stop
    otherwise exchange $p_k$ & $p_j$
    ~~exchange rows [illegible] j and k~~
    for  $i \geq k+1$  do
        $\text{QUOT}_{p_i k} = a_{p_i k} / a_{p_k k}$
        for  $j \geq k+1 \dots n+1$  do:
            set  $a_{p_i j} = a_{p_i j} - \text{QUOT}_{p_i k}\, a_{p_k j}$
IF  $a(p_n, n) = 0$  matrix is not invertible
    RESULTS GIVE an augmented upper triangular matrix  $[U_{ij} \,\vdots\, \tilde{b}_i]$
```

$\underset{\sim}{p}$ just keeps track of which rows have been used as pivotal rows doesn't effect answers

solve by back substitution to find x_j

$x_n = a(p_n, n+1) / a(p_n, n)$

$x_i = \dfrac{\left[a(\overset{p_i}{\cancel{n+1}}, \overset{n+1}{\cancel{i}}) - \sum_{j=i+1}^{n} a(p_i, j)\, x_j\right]}{a(i,i)} / a(\overset{p}{i}, i)$

~~then reshuffle the x's to get them in the proper order~~

They are in order

No pivoting

$$\left[\begin{array}{ccc|c} 2 & 3 & -1 & 5 \\ 4 & 4 & -3 & 3 \\ -2 & 3 & -1 & 1 \end{array}\right] \rightarrow \left[\begin{array}{ccc|c} 2 & 3 & -1 & 5 \\ 0 & -2 & -1 & -7 \\ 0 & 6 & -2 & 6 \end{array}\right] \begin{array}{c} \\ +\frac{4}{2} \\ -1 \end{array}$$

$$\rightarrow \left[\begin{array}{ccc|c} 2 & 3 & -1 & 5 \\ 0 & -2 & -1 & -7 \\ 0 & 0 & -5 & -15 \end{array}\right] \quad \frac{6}{-2}$$

$$x_3 = 3$$

$$x_2 = \frac{-7 + 1(3)}{-2} = 2$$

$$x_1 = \frac{5 + 1(3) - 3(2)}{2} = 1$$

$$\begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ -1 & -3 & 1 \end{bmatrix} \begin{bmatrix} 2 & 3 & -1 \\ 0 & -2 & -1 \\ 0 & 0 & -5 \end{bmatrix}$$

$$\left[\begin{array}{ccc|c} 2 & 3 & -1 & 5 \\ 4 & 4 & -3 & 3 \\ -2 & 3 & -1 & 1 \end{array}\right] \qquad \begin{array}{l} 2x_1 + 3x_2 - x_3 = 5 \\ 4x_1 + 4x_2 - 3x_3 = 3 \\ -2x_1 + 3x_2 - x_3 = 1 \end{array}$$

$$\underset{\sim}{P}^T = [1, 2, 3]$$

1) a_{21} is largest in first colum

$$\left[\begin{array}{ccc|c} 4 & 4 & -3 & 3 \\ 2 & 3 & -1 & 5 \\ -2 & 3 & -1 & 1 \end{array}\right] \Rightarrow \begin{bmatrix} 4 & 4 & -3/4 & 3/4 \\ 0 & 1 & 1/2 & 7/2 \\ 0 & 5 & -5/2 & 5/2 \end{bmatrix} \quad \begin{array}{l} -\boxed{2\frac{(1)}{4}} + (2) \to (2') \\ -\boxed{-2\frac{(1)}{4}} + (3) \to (3') \end{array}$$

$$\underset{\sim}{P}^T = [2, 1, 3]$$

$$\begin{bmatrix} 4 & 4 & -3/4 & 3/4 \\ 0 & 5 & -5/2 & 5/2 \\ 0 & 1 & 1/2 & 7/2 \end{bmatrix} \Longrightarrow \begin{bmatrix} 4 & 4 & -3/4 & 3/4 \\ 0 & 5\,(1) & -5/2\,(-1/2) & 5/2\,(1/2) \\ 0 & 0 & 1 & 3 \end{bmatrix} \quad \begin{array}{l} \div \cancel{5/5} \\ \curvearrowright (2') \,\&\, (3') \\ -\boxed{1\frac{(2')}{5}} + (3') \to (3'') \end{array}$$

$$\underset{\sim}{P}^T = [2, 3, 1]$$

$\Longrightarrow$ **P just keeps track of which equations have been used for pivotal eqns and doesn't effect the answers.**

$\bar{x}_3 = 3$

$\bar{x}_2 = +2$

$\bar{x}_1 = \underline{1}$

$$U \begin{bmatrix} 4 & 4 & -3 \\ 0 & 5 & -5/2 \\ 0 & 0 & 1 \end{bmatrix} = u_{ij} \qquad P^{-1} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} \qquad \tilde{P} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ +1 & 0 & 0 \end{bmatrix}$$

$$P = \tilde{P}^T = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

$$L = \begin{bmatrix} 1 & 0 & 0 \\ -1/2 & 1 & 0 \\ 1/2 & 1/5 & 1 \end{bmatrix} = l_{ij}$$

these include row interchanges

these are a_{ij} $i>j$ a_{jj} pivots

$$LU = P^{-1}A = \begin{bmatrix} 4 & 4 & -3 \\ -2 & 3 & -1 \\ 2 & 3 & -1 \end{bmatrix} = \tilde{A}$$

<u>HW</u> p. 187 (4-9) out of wolford/smith

scaled partial pivoting

to define P

$$\begin{bmatrix} 2 & 3 & -1 \\ 4 & 4 & -3 \\ -2 & 3 & -1 \end{bmatrix}$$

search along row for largest element

define di $\begin{bmatrix} 3 \\ 4 \\ 3 \end{bmatrix}$

$p = [1, 2, 3]$

form for 1st column $\frac{a_{i1}}{d_i}$ $i>1$ $\frac{2}{3}$ $\boxed{\frac{4}{4}}$ $\frac{-2}{3}$ **find largest**

2nd row largest. interchange 2 & 1 → $p = [2, 1, 3]$

new d_i after interch $\begin{bmatrix} 4 \\ 3 \\ 3 \end{bmatrix}$

do elim

$$\begin{bmatrix} 4 & 4 & -3 \\ 0 & 1 & 1/2 \\ 0 & 5 & -5/2 \end{bmatrix}$$

form $\frac{a_{i2}}{d_i}$ $i>2$ ~~$\frac{4}{3}$ [illegible] $\frac{1}{4}$ $\frac{3}{5}$~~

$\cancel{\frac{4}{4}}$ $\frac{1}{3}$ $\boxed{\frac{5}{3}}$

interchange 3 & 2 → $p\ [2, 3, 1]$

only do this for large scale mismatch

Given $PLU\underset{\sim}{x} = \underset{\sim}{b}$

$$LU\underset{\sim}{x} = P^{-1}\underset{\sim}{b} = \underset{\sim}{\hat{b}}$$

$$L\underset{\sim}{y} = \underset{\sim}{\hat{b}}$$

$$U\underset{\sim}{x} = \underset{\sim}{y}$$

Doolittle's

Use forward subst

for $k=1,\dots,n$ do:

$$y_k = \hat{b}_k - \sum_{j=1}^{k-1} l_{kj} y_j$$

Use back subst now

$k = n, n-1, \dots, 1$

$$x_k = \frac{y_k - \sum_{j=k+1}^{n} u_{kj} x_j}{u_{kk}}$$

$$\begin{bmatrix} \ddots & & U \\ & \ddots & \\ L & & \ddots \end{bmatrix} = W$$

$\therefore \; l_{kj} = w_{kj}$ for $j < i$

$u_{kj} = w_{kj}$ for $j \geq i$

This method is Doolittle's Method

Crout or Cholesky LU Decomposition is $\begin{bmatrix} x & & & \\ x & x & & 0 \\ x & x & x & \\ x & x & x & x \end{bmatrix} \begin{bmatrix} 1 & x & x & x \\ & 1 & x & x \\ & & 1 & x \\ 0 & & & 1 \end{bmatrix}$

int 2 & 3
row
pre
~~pre~~ mult
by A

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}\begin{bmatrix} 1 & 2 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

interch 2 & 3 rows

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Take identity matrix
interchange rows
& ~~post~~ multiply it ~~to~~ by B to interch rows
post

$$\begin{bmatrix} 1 & 2 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 2 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

B ~~post~~ post mult by A

interch col 2 & 3

Take identity matrix
interchange columns
& ~~post~~multiply it by B
pre-
to interch rows.

do · 4.1-3, 4

Convert

$$\begin{bmatrix} 4 & 1 & 2 & 1 \\ 3 & 2 & 1 & 1 \\ 1 & 2 & 0 & 1 \\ 1 & 1 & 0 & 1 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 2 \\ 1 & 1 & 3 & 2 \\ 1 & 2 & 4 & 1 \end{bmatrix}$$

$$\begin{bmatrix} 4 & 2 & 4 & -1 \\ 9 & 1 & 8 & 8 \\ 1 & 3 & 9 & 5 \\ 7 & 6 & 2 & 7 \end{bmatrix} \longrightarrow \begin{bmatrix} 8 & 3 & 6 & -1 \\ 8 & 9 & 2 & 4 \\ 1 & 5 & 7 & 2 \\ 9 & 7 & 1 & 4 \end{bmatrix}$$

Crout

$$l_{ij} = \tilde{a}_{ij} - \sum_{k=1}^{j-1} l_{ik} u_{kj} \qquad j<i$$

$$u_{ij} = \frac{\tilde{a}_{ij} - \sum_{k=1}^{i-1} l_{ik} u_{kj}}{l_{ii}} \qquad i \le j \qquad u_{ii} = 1$$

get cols of L first then rows of U

the pivoting scheme & the triangularization is equivalent to

$$PLU = A$$

$$\tilde{A}\underline{x} = LU\underset{\sim}{x} = P^{-1}\underset{\sim}{b} = \underset{\sim}{\check{b}}$$

$$U\underset{\sim}{x} = L^{-1}\underset{\sim}{\check{b}} = \underset{\sim}{\tilde{b}}$$

for Doolittle

$$\left\{\begin{array}{ll} u_{ij} = \tilde{a}_{ij} - \sum_{r=1}^{i-1} l_{ir} u_{rj} & i \le j \\ l_{ij} = \dfrac{\tilde{a}_{ij} - \sum_{r=1}^{j-1} l_{ir} u_{rj}}{u_{jj}} & i > j \quad \& \ l_{ii} = 1 \end{array}\right.$$

we calculate row of U first then colum of L

$\tilde{A} = P^{-1}A$ how do we find P^{-1} and $P = (P^{-1})^T$

if we look at $\underset{\sim}{p}^T$ $[3 \ 1 \ 2]$ $P^{-1} = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$

move the rows of I $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ $P = (P^{-1})^T = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$

what is $A^{-1} = (PLU)^{-1} = U^{-1}L^{-1}P^{-1}$

$U \,\vdots\, e_i \Rightarrow$ ~~first co~~ i^{th} column of U^{-1} use back substitution :

$L \,\vdots\, e_i \Rightarrow$ i^{th} colum of L^{-1} use forward substitution

IF we want the inverse of a matrix without pivoting then $\tilde{a}_{ij}$ replace by a_{ij}

Gauss-Jordan Elimination Technique eliminates a variable from all equations even those above the pivot row - 50% more operations than gauss

Gauss-Seidel best for diag dominant lets $x_i = b_i/a_{ii}$ then $x_{i\,new} = b_i - \left[\sum_{j \ne i} a_{ij}x_j\right] / a_{ii}$

Pathology - singular matrices

1. If more unknown than equations (infinite no. of solutions) example Eigenvalues

2. If more equations than unknowns ~~(infinite no. of solutions)~~

   a. If solution satisfies all eqns we have redundancy (solve n eqns n unknowns)

   b. If solution does not satisfy all eqs we have no solution

- we can determine whether a solution exists by the rank of the matrix

  triangularize matrix; then look for zeros on diagonal if no zeros matrix will give unique solutions

- look at pg 116

$$\left[\begin{array}{ccc|c} 1 & -2 & 3 & 5 \\ 0 & 8 & -7 & -3 \\ 0 & 0 & 0 & 1 \end{array}\right]$$ inconsistent no solutions

$$\left[\begin{array}{ccc|c} 1 & -2 & 3 & 5 \\ 0 & 8 & -7 & -3 \\ 0 & 0 & 0 & 0 \end{array}\right]$$ consistent; x_3 can be any no.

$x_1 = 5 - 3a + 2(-3+7a)/8$
$x_2 = (-3+7a)/8$
$x_3 = a$

$x_1 = 34/8 - \frac{10a}{8}$
$x_2 = -3/8 + 7a/8$
$x_3 = 0 + a$

- determinants of Δ matrices is $\prod_{i=1}^{n} a_{ii}$

  determinants of row reduced matrices is $\prod_{i=1}^{n} a_{ii}(-1)^q$

  q is no. of row interchange operations

$\underset{\sim}{x} = \text{solution} + \underset{\sim}{x}_{ns}$

where $\underset{\sim}{x}_{ns}$ is such that $A\underline{x}_{ns} = \underline{0}$

ns = null space

Implementation of this project within the Department of Mechanical Engineering would make a strong program even more attractive to undergraduate students and would encourage them to continue beyond the bachelor's level in engineering.

I strongly urge you to fund this program to build strength upon special strength.

Sincerely,

Modesto A. (Mitch) Maidique

in general do not evaluate the inverse but we use error equations to increase accuracy of solution

let $\underset{\sim}{A}\,\underset{\sim}{x} = \underset{\sim}{b}$.

Choose a vector $\underset{\sim}{x}_0$ approx solution $\underset{\sim}{A}\,\underset{\sim}{x}_0 = \underset{\sim}{b}_0$ appox rhs

then $\underset{\sim}{A}\,(\overset{e}{\underset{\sim}{x}-\underset{\sim}{x}_0}) = \overset{r_0}{\cancel{e}} = \underset{\sim}{b} - b_0$ $\quad \underset{\sim}{r}$ is residual $\quad$ residual eqns.

Solve $\underset{\sim}{A}\,\underset{\sim}{e}^{(1)} = \underset{\sim}{r}_0$ for $\underset{\sim}{e}'$ is the correction to the problem

using gaussian Elim

$\& \quad \underset{\sim}{x}_1 = \underset{\sim}{x}_0 + \underset{\sim}{e}^{(1)}$

now

$\underset{\sim}{r}_1 = \underset{\sim}{b} - A\,\underset{\sim}{x}_1$ and solve $\underset{\sim}{A}\,\underset{\sim}{e}^{(2)} = \underset{\sim}{r}_1$ $\qquad A\underline{x}_1 = b_1$

$\& \quad \underset{\sim}{x}_2 = \underset{\sim}{x}_1 + \underset{\sim}{e}^{(2)}$ $\qquad \underline{r}_1 = \underline{b} - \underline{b}_1 = \underline{b} - Ax_1$

etc. when do we converge?: when $\|e\| / \|\cancel{x}\|$ is small.

what does $\|e\|$ mean $\quad \|e\|_n = \sqrt[n]{e_1^{n} + e_2^{n} + \dots}$

as $n \to \infty$ $\quad \|e\|_n \to \max e_i$

- we can define a norm for a matrix $\|A\|_\infty = \max_{1 \le i \le n} \sum_{j=1}^{n} |a_{ij}|$ max row sum

$\|A\|_1 = \max_{1 \le j \le n} \sum_{i=1}^{n} |a_{ij}|$ max col sum

- Gauss-Seidel method is also iterative and works well when

$$|a_{ii}| > \sum_{j=1}^{n} a_{ij} \quad j \ne i$$

$\|A\|_e = \left(\sum_i \sum_j a_{ij}^2\right)^{1/2}$ $\qquad \|A\|_2 = \sqrt{\text{max Evalue of } A^T A}$

cond $(A) = \|A\|\,\|A^{-1}\|$ large cond A, matrix ill conditioned - method ~~will~~ may not converge - must use double prec for iterative improvement (calc. residual)

- ~~Non homogeneous Algebraic Equations~~

~~$\underset{\sim}{A}\,\underset{\sim}{x} = \lambda \underset{\sim}{x}$~~

~~when is $(\underset{\sim}{A} - \lambda \underset{\sim}{I})\underset{\sim}{x} = \underset{\sim}{0}$ either $\underset{\sim}{x} = 0$ or $\det(\underset{\sim}{A} - \lambda \underset{\sim}{I}) = 0$~~

~~Hw Use matrix algebra & find A, B, C so that $y = Ax^2 + Bx + C$~~

~~where parabola pass through (1,8) (2,13) (3,20)~~

A method of solution to $A\underset{\sim}{x} = \underset{\sim}{b}$ using fixed pt iteration

$$\underset{\sim}{x} = g(\underset{\sim}{x}) = C\underset{\sim}{b} + (I - CA)\underset{\sim}{x} \qquad C \text{ is approximate inverse of } A$$

is $$\underset{\sim}{x}^{(m+1)} = \underset{\sim}{x}^{(m)} + C\left(b - A\underset{\sim}{x}^{(m)}\right) = Cb + (I - CA)\underset{\sim}{x}^{(m)}$$

$$\|x^{m+1}\| \le \|Cb\| + \|I - CA\|\,\|\underset{\sim}{x}^{m}\|$$

where $\|I - CA\| < 1$

- for a strictly row diagonally dominant method
$$|a_{ii}| > \sum_{j \ne i} |a_{ij}| \qquad i = 1, \dots, n$$
IF $A = L + D + U$ $\quad \left[\begin{smallmatrix} \diagdown & 0 \\ L & \diagdown \end{smallmatrix}\right] + \left[\begin{smallmatrix} \diagdown & 0 \\ 0 & D\diagdown \end{smallmatrix}\right] + \left[\begin{smallmatrix} \diagdown & U \\ 0 & \diagdown \end{smallmatrix}\right]$

let $$C = D^{-1} = \begin{bmatrix} a_{11} & & & 0 \\ & a_{22} & & \\ & & \ddots & \\ 0 & & & a_{nn} \end{bmatrix}^{-1} \qquad \begin{bmatrix} 4 & -1 & & \\ -1 & 4 & -1 & \\ & -1 & 4 & -1 \\ & & & \ddots \end{bmatrix}$$

C can be a $(\text{diagonal matrix})^{-1}$ Jacobi $\quad x_i^{(m+1)} = \left(b_i - \sum_{j \ne i} a_{ij} x_j^{(m)}\right) / a_{ii}$

C can be $(PLU)^{-1}$ $\quad = x_i^{(m)} + \left(b_i - \sum_{i=1}^{n} a_{ij} x_j^{(m)}\right) / a_{ii}$

$$A = [L\diagdown] + [\diagdown D \diagdown] + [\diagdown U \diagdown]$$

- C is the approximate inverse of A if the max eigenvalue of
$$(I - CA) < 1$$
- this maximum eigenvalue is called the spectral radius of $A^T \cdot A$
- as spectral radius decreases fixed point iteration converges quickly

Jacobi assumes $C^{-1} = [\diagdown D \diagdown]$ if A is diagonally dominant

Gauss Seidel assumes $C^{-1} = [L \diagdown D \diagdown] = L + D$

if A is row-diagonally dominant, tridiagonal

- iterative techniques are used for large systems of linear equations
sparse coeff matrix

Jacobi Method

$$Ax = b$$

$$[L + D + U]x = b$$

$$[L + U]x + DX = b \quad \text{thus}$$

$$x = D^{-1}b - D^{-1}[L+U]x$$

$$X = D^{-1}\{b - [L+U]X\} \qquad \text{here } D = \begin{bmatrix} a_{11} & & 0 \\ & a_{22} & \\ 0 & & \ddots a_{nn} \end{bmatrix}$$

$$x_i^{(n+1)} = \frac{b_i}{a_{ii}} - \sum_{\substack{j=1 \\ i\neq j}}^{n} \frac{a_{ij}}{a_{ii}} x_i^{(n)} = x_i^{(n)} + \frac{b_i}{a_{ii}} - \sum_{j=1}^{n} \frac{a_{ij}}{a_{ii}} x_i^{(n)}$$

condition for convergence is $|a_{ii}| > \sum_{\substack{j=1 \\ i\neq j}}^{n} |a_{ij}|$

normally written as $x^{(n+1)} = x^{(n)} + D^{-1}(b - Ax^{(n)})$

$$D^{-1}b - D^{-1}(L+D+U)x^{(n)}$$

$$\cancel{x^{(n)}} + D^{-1}b - \{D^{-1}(L+U)x^{(n)} + \cancel{x^{(n)}}\}$$

Gauss - Seidel

$$x^{(n+1)} = x^{(n)} + (L+D)^{-1}\{b - Ax^{(n)}\}$$

$$= x^{(n)} + (L+D)^{-1}b - [L+D]^{-1}[L+D]x^{(n)} - (L+D)^{-1}Ux^{(n)}$$

$$x^{(n+1)} = x^{(n)} + (L+D)^{-1}b - x^{(n)} - (L+D)^{-1}Ux^{(n)} \qquad \checkmark\checkmark$$

$$= (L+D)^{-1}[b - Ux^{(n)}] \qquad \text{or} \qquad (L+D)x^{(n+1)} = b - Ux^{(n)}$$

$$Dx^{(n+1)} = b - Lx^{(n+1)} - Ux^{(n)}$$

$$x^{(n+1)} = D^{-1}[b - Lx^{(n+1)} - Ux^{(n)}]$$

$$Ax = b$$

$$\overset{A}{(L+D+U)} - (L+D)$$

$$(L+D)\underset{\sim}{x} + U\underset{\sim}{x} = b$$

$$\underset{\sim}{x} = (L+D)^{-1}\underset{\sim}{b} - (L+D)^{-1}U\underset{\sim}{x}$$

$$= (L+D)^{-1}[\underset{\sim}{b} - A\underset{\sim}{x}] + \underset{\sim}{x}$$

$$\underset{\sim}{x}^{(n+1)} = \underset{\sim}{x}^{(n)} + (L+D)^{-1}[\underset{\sim}{b} - A\underset{\sim}{x}^{(n)}]$$

$$(L+D)\underset{\sim}{x}^{(n+1)} = (L+D)\underset{\sim}{x}^{(n)} + b - A\underset{\sim}{x}^{(n)}$$

$$= -U\underset{\sim}{x}^{(n)} + b \qquad \text{or } x^{(n+1)} = [-Lx^{(n+1)} - Ux^{(n)} + b]/a_{ii}$$

$-1x_1 - .125x_2 + .125x_3 + 1 = R_1$

$.143x_1 - 1x_2 + .286x_3 + .571 = R_2$

$-.222x_1 - .111x_2 - 1x_3 + 1.333 = R_3$

(0,0,0)

$R_1 = 1$
$R_2 = .571$
$R_3 = 1.333$

(0,0,1.333)

$R_1 = 1.167$
$R_2 = .952$
$R_3 = 0$

(1.167, 0, 1.333)

$R_1 = 0$
$R_2 = 1.119$
$R_3 = -.259$

(1.167, 1.119, 1.333)

$R_1 = -.14025$
$R_2 = 0$
$R_3 = -.383$

∴ 1.333 − .383 = .950

(1.167, 1.119, 0.95)

$R_1 = -.188$
$R_2 = -.109$
$R_3 = 0$

∴ 1.167 − .188 = .979

(.979, 1.119, .950)

$R_1 = 0$
$R_2 = -.136$
$R_3 = .041$

∴ 1.119 − .136 = .983

(.979, .983, .950)

$R_1 = .0169$
$R_2 = 0$
$R_3 = 0.0565$

∴ 0.95 + .0565 = 1.0065

Relaxation Method

- More rapidly convergent than Gauss-Seidel

Transpose all to one side of eq & divide by − largest coeff in each eqn.

$8x_1 + x_2 - x_3 = 8$ $\quad \div$ by -8

$2x_1 + x_2 + 9x_3 = 12$ $\quad \div$ by -9 $\quad$ largest coeff in each eqn.

$x_1 - 7x_2 + 2x_3 = -4$ $\quad \div$ by $-(-7)$

$-1x_1 - .125x_2 + .125x_3 + 1 = 0$

$-.222x_1 - .111x_2 - 1x_3 + 1.333 = 0$

$.143x_1 - x_2 + .286x_3 + .571 = 0$

∘ call RHS R_1 R_2 R_3

interchange R_2 & R_3 so that -1 is on diag

Now pick a trial x_1, x_2, x_3 ~~(1, 1, 1)~~

~~$R_1 = 0$~~

~~$R_2 = 0$~~

~~$R_3 = 0$~~

what if $(0,0,0) \implies$ $R_1 = 1$, $R_3 = 1.333$, $R_2 = .571$

try to reduce largest residual by leaving x_1, x_2 along & change x_3 (relates to R_3) $\Rightarrow x_3 = 1.333$

$\Rightarrow R_3 := 0$

R_1: $\quad -0 - 0 + \frac{1}{8}\left(\frac{4}{3}\right) + 1 = 1.1666 \quad \triangleleft$ highest residual

R_2: $\quad 0 - 0 + \frac{2}{7} \cdot \frac{4}{3} + \frac{4}{7} = \frac{8}{21} + \frac{4}{7} = \frac{20}{21} \cong 0.95$

$R_1 = 1.1666 - x_1 + 0 + \overset{0}{\cancel{}} + \frac{8}{21} + 1 \qquad x_1 = -1.1666$

Southwell showed that it is better to under- or over-relax since residuals never stay at zero but change with each iteration.

If we note the direction of relaxation we can anticipate how to make the change.

Best way is to set

mod Gauss Seidel

$$x_i^{(k+1)} = x_i^{(k)} + \frac{\omega}{a_{ii}}\left[b_i - \sum_{j=1}^{i-1} a_{ij} x_j^{(k+1)} - \sum_{j=i}^{n} a_{ij} x_j^{(k)} \right] \quad j<i$$

where $1 \le \omega \le 2$ overrelaxation factor [illegible] ~~52~~

Residuals 62 Generate the matrix and solve

trig 65 Solve ~~set~~ starting w/ $(P_1\ P_2\ P_3\ P_4)^T$

$$= (400, 300, 200, 100)$$

$$= \begin{bmatrix} -.3 & +.2 & +.1 & 0 & 0 & 0 \\ & & & & & \\ & & & & & \\ & & & & & \end{bmatrix} \begin{pmatrix} 500-P_1 \\ P_1-P_2 \\ P_1-P_3 \\ P_2-P_3 \\ P_2-P_4 \\ P_3-P_4 \\ P_4-0 \end{pmatrix}$$

Solve 57

Simultaneous Non Linear Equations

$$f(x,y)=0$$
$$g(x,y)=0$$

Assume we have an $x_n, y_n \rightarrow f_n$ & g_n.

$f_{n+1} = f(x_n+\Delta x,\, y_n+\Delta y)$

$$0 = f_{n+1} = f_n + \frac{\partial f}{\partial x}\left(\overset{\Delta x}{\cancel{x_{n+1}-x_n}}\right) + \frac{\partial f}{\partial y}\left(\overset{\Delta y}{\cancel{y_{n+1}-y_n}}\right)$$

$$0 = g_{n+1} = g_n + \frac{\partial g}{\partial x}\left(\overset{\Delta x}{\cancel{x_{n+1}-x_n}}\right) + \frac{\partial g}{\partial y}\left(\overset{\Delta y}{\cancel{y_{n+1}-y_n}}\right).$$

$$\begin{aligned} f_x\Delta x + f_y\Delta y &= -f_n = \begin{pmatrix} f_x & f_y \\ g_x & g_y \end{pmatrix}\begin{pmatrix}\Delta x\\ \Delta y\end{pmatrix} = \text{known} \Rightarrow \\ g_x\Delta x + g_y\Delta y &= -g_n \phantom{= \begin{pmatrix} f_x \end{pmatrix}} = \text{known} = \end{aligned} \qquad \begin{aligned}\Delta x &= x_{n+1}-x_n \\ \Delta y &= y_{n+1}-y_n\end{aligned}$$

$$\Delta x = x_{n+1}-x_n = \frac{\begin{vmatrix} -f_n & f_y \\ -g_n & g_y \end{vmatrix}}{\begin{vmatrix} f_x & f_y \\ g_x & g_y \end{vmatrix}_{@\,x_n,y_n}} \; ; \quad \Delta y = y_{n+1}-y_n = \frac{\begin{vmatrix} -f_n & f_y \\ g_n & g_y \end{vmatrix}}{\begin{vmatrix} f_x & f_y \\ g_x & g_y \end{vmatrix}_{@\,x_n,y_n}}$$

now use x_{n+1}, y_{n+1} to find $f_{n+1}, g_{n+1}, f_x, f_y, g_x, g_y$ & repeat.

- denominator is called the Jacobian. $J(f,g)$ $\quad \dfrac{\partial(f,g)}{\partial(x,y)}$
- this system has a unique solution if $J(f,g) \neq 0$ in interval
- converges if f, g and all their derivatives to 2nd order are continuous & bounded
- if x_0, y_0 are sufficiently close to the root. $\begin{pmatrix} f \\ g \end{pmatrix}$
- in the damped Newton $\underset{\sim}{x}_{n+1} = \underset{\sim}{x}_n + \underset{\sim}{h}$ $\begin{pmatrix}\Delta x\\ \Delta y\end{pmatrix}$ if $\|\underset{\sim}{f}(\underset{\sim}{x}_{n+1})\|_2 > \|f(\underset{\sim}{x}_n)\|_2$ then $\underset{\sim}{x}_{n+1} = \underset{\sim}{x}_n + \underset{\sim}{h}/2^i$ $\quad i=1,2,\ldots$ for first i where $\|f(x_{n+1})\| < \|f(x_n)\|$

~~SOLUTION TO <u>LINEAR</u> ALGEBRAIC EQNS~~

- ~~occurs in systems that approximate the partial diff eqns by finite differences~~
  - ~~elasticity, heat transfer, vibrations~~
- ~~we can solve a set of equations by~~

First Central-Difference Expressions $O(\Delta x^2)$

$$
\begin{aligned}
y_i' &= \frac{y_{i+1} - y_{i-1}}{2(\Delta x)} \\
y_i'' &= \frac{y_{i+1} - 2y_i + y_{i-1}}{(\Delta x)^2} \\
y_i''' &= \frac{y_{i+2} - 2y_{i+1} + 2y_{i-1} - y_{i-2}}{2(\Delta x)^3} \\
y_i'''' &= \frac{y_{i+2} - 4y_{i+1} + 6y_i - 4y_{i-1} + y_{i-2}}{(\Delta x)^4}
\end{aligned}
\tag{5-38}
$$

Second Central-Difference Expressions $O(\Delta x^4)$

$$
\begin{aligned}
y_i' &= \frac{-y_{i+2} + 8y_{i+1} - 8y_{i-1} + y_{i-2}}{12(\Delta x)} \\
y_i'' &= \frac{-y_{i+2} + 16y_{i+1} - 30y_i + 16y_{i-1} - y_{i-2}}{12(\Delta x)^2} \\
y_i''' &= \frac{-y_{i+3} + 8y_{i+2} - 13y_{i+1} + 13y_{i-1} - 8y_{i-2} + y_{i-3}}{8(\Delta x)^3} \\
y_i'''' &= \frac{-y_{i+3} + 12y_{i+2} - 39y_{i+1} + 56y_i - 39y_{i-1} + 12y_{i-2} - y_{i-3}}{6(\Delta x)^4}
\end{aligned}
\tag{5-39}
$$

First Forward-Difference Expressions $O(\Delta x)$

$$
\begin{aligned}
y_i' &= \frac{y_{i+1} - y_i}{(\Delta x)} \\
y_i'' &= \frac{y_{i+2} - 2y_{i+1} + y_i}{(\Delta x)^2} \\
y_i''' &= \frac{y_{i+3} - 3y_{i+2} + 3y_{i+1} - y_i}{(\Delta x)^3} \\
y_i'''' &= \frac{y_{i+4} - 4y_{i+3} + 6y_{i+2} - 4y_{i+1} + y_i}{(\Delta x)^4}
\end{aligned}
\tag{5-40}
$$

Second Forward-Difference Expressions $O(\Delta x^2)$

$$
\begin{aligned}
y_i' &= \frac{-y_{i+2} + 4y_{i+1} - 3y_i}{2(\Delta x)} \\
y_i'' &= \frac{-y_{i+3} + 4y_{i+2} - 5y_{i+1} + 2y_i}{(\Delta x)^2} \\
y_i''' &= \frac{-3y_{i+4} + 14y_{i+3} - 24y_{i+2} + 18y_{i+1} - 5y_i}{2(\Delta x)^3} \\
y_i'''' &= \frac{-2y_{i+5} + 11y_{i+4} - 24y_{i+3} + 26y_{i+2} - 14y_{i+1} + 3y_i}{(\Delta x)^4}
\end{aligned}
\tag{5-41}
$$

First Backward-Difference Expressions $O(\Delta x)$

$$
\begin{aligned}
y_i' &= \frac{y_i - y_{i-1}}{(\Delta x)} \\
y_i'' &= \frac{y_i - 2y_{i-1} + y_{i-2}}{(\Delta x)^2} \\
y_i''' &= \frac{y_i - 3y_{i-1} + 3y_{i-2} - y_{i-3}}{(\Delta x)^3} \\
y_i'''' &= \frac{y_i - 4y_{i-1} + 6y_{i-2} - 4y_{i-3} + y_{i-4}}{(\Delta x)^4}
\end{aligned}
\tag{5-42}
$$

Second Backward-Difference Expressions $O(\Delta x^2)$

$$
\begin{aligned}
y_i' &= \frac{3y_i - 4y_{i-1} + y_{i-2}}{2(\Delta x)} \\
y_i'' &= \frac{2y_i - 5y_{i-1} + 4y_{i-2} - y_{i-3}}{(\Delta x)^2} \\
y_i''' &= \frac{5y_i - 18y_{i-1} + 24y_{i-2} - 14y_{i-3} + 3y_{i-4}}{2(\Delta x)^3} \\
y_i'''' &= \frac{3y_i - 14y_{i-1} + 26y_{i-2} - 24y_{i-3} + 11y_{i-4} - 2y_{i-5}}{(\Delta x)^4}
\end{aligned}
\tag{5-43}
$$

EXAMPLE 5-3

In Example 3-4 the Newton-Raphson method was used to determine the output lever angles of a crank-and-lever 4-bar linkage system for each 5° of rotation of the input crank. Now we shall determine the angular velocity and the angular acceleration of the output lever of the same type of mechanism for each 5° of rotation of the input crank, with the latter rotating at a uniform angular velocity of 100 radians/sec.

We can determine the output lever positions ϕ, corresponding to each 5° of crank rotation θ, by utilizing Freudenstein's equation and the Newton-Raphson method, as was done in Example 3-4. Such a set of values, in effect, gives us a series of points on the ϕ versus θ curve, and the ϕ values are stored in memory to provide data for the differentiation processes which follow. The slope of the ϕ-θ curve may be related to the angular velocity of the output lever $d\phi/dt$ if we realize that, with the crank rotating at a constant ω, its angular position is given by

$$\theta = \omega t$$

so that

$$\frac{d\phi}{d\theta} = \frac{1}{\omega}\frac{d\phi}{dt}$$

must be solved. Derivations of the more accurate expressions are not given here, but such expressions for several derivatives are included in the list that follows this discussion. Derivations for the higher derivatives are accomplished with much greater facility and far less labor by using *difference, averaging,* and *derivative* operators. Such a method is outside the scope of this text, but it can be found in various books concerned with numerical analysis.[9]

It has been shown that the central-difference expressions for the various derivatives involve values of the function on both sides of the x value at which the derivative of the function is desired. By utilizing the appropriate Taylor series expansions, one can easily obtain expressions for the derivatives which are entirely in terms of values of the function at x_i and points to the right of x_i. These are known as *forward finite-difference* expressions. In a similar manner, derivative expressions which are entirely in terms of values of the function at x_i and points to the *left* of x_i can be found. These are known as *backward finite-difference* expressions. In numerical differentiation, forward-difference expressions are used when data to the left of a point at which a derivative is desired are not available, and backward-difference expressions are used when data to the right of the desired point are not available. Central-difference expressions, however, are more accurate than either forward- or backward-difference expressions. This can be seen by noting the order of the error in the list of differentiation formulas that follows.

Central-difference expressions with error of order h^2

$$\begin{aligned}
y_i' &= \frac{y_{i+1} - y_{i-1}}{2h} \\
y_i'' &= \frac{y_{i+1} - 2y_i + y_{i-1}}{h^2} \\
y_i''' &= \frac{y_{i+2} - 2y_{i+1} + 2y_{i-1} - y_{i-2}}{2h^3} \\
y_i'''' &= \frac{y_{i+2} - 4y_{i+1} + 6y_i - 4y_{i-1} + y_{i-2}}{h^4}
\end{aligned} \tag{5.119}$$

Central-difference expressions with error of order h^4

$$\begin{aligned}
y_i' &= \frac{-y_{i+2} + 8y_{i+1} - 8y_{i-1} + y_{i-2}}{12h} \\
y_i'' &= \frac{-y_{i+2} + 16y_{i+1} - 30y_i + 16y_{i-1} - y_{i-2}}{12h^2} \\
y_i''' &= \frac{-y_{i+3} + 8y_{i+2} - 13y_{i+1} + 13y_{i-1} - 8y_{i-2} + y_{i-3}}{8h^3} \\
y_i'''' &= \frac{-y_{i+3} + 12y_{i+2} - 39y_{i+1} + 56y_i - 39y_{i-1} + 12y_{i-2} - y_{i-3}}{6h^4}
\end{aligned} \tag{5.120}$$

---

[9] M. G. Salvadori and M. L. Baron, *Numerical Methods in Engineering* (Englewood Cliffs, N.J.: Prentice-Hall, 1961).

Forward-difference expressions with error of order h

$$\begin{aligned}
y_i' &= \frac{y_{i+1} - y_i}{h} \\
y_i'' &= \frac{y_{i+2} - 2y_{i+1} + y_i}{h^2} \\
y_i''' &= \frac{y_{i+3} - 3y_{i+2} + 3y_{i+1} - y_i}{h^3} \\
y_i'''' &= \frac{y_{i+4} - 4y_{i+3} + 6y_{i+2} - 4y_{i+1} + y_i}{h^4}
\end{aligned} \tag{5.121}$$

Forward-difference expressions with error of order h^2

$$\begin{aligned}
y_i' &= \frac{-y_{i+2} + 4y_{i+1} - 3y_i}{2h} \\
y_i'' &= \frac{-y_{i+3} + 4y_{i+2} - 5y_{i+1} + 2y_i}{h^2} \\
y_i''' &= \frac{-3y_{i+4} + 14y_{i+3} - 24y_{i+2} + 18y_{i+1} - 5y_i}{2h^3} \\
y_i'''' &= \frac{-2y_{i+5} + 11y_{i+4} - 24y_{i+3} + 26y_{i+2} - 14y_{i+1} + 3y_i}{h^4}
\end{aligned} \tag{5.122}$$

Backward-difference expressions with error of order h

$$\begin{aligned}
y_i' &= \frac{y_i - y_{i-1}}{h} \\
y_i'' &= \frac{y_i - 2y_{i-1} + y_{i-2}}{h^2} \\
y_i''' &= \frac{y_i - 3y_{i-1} + 3y_{i-2} - y_{i-3}}{h^3} \\
y_i'''' &= \frac{y_i - 4y_{i-1} + 6y_{i-2} - 4y_{i-3} + y_{i-4}}{h^4}
\end{aligned} \tag{5.123}$$

Backward-difference expressions with error of order h^2

$$\begin{aligned}
y_i' &= \frac{3y_i - 4y_{i-1} + y_{i-2}}{2h} \\
y_i'' &= \frac{2y_i - 5y_{i-1} + 4y_{i-2} - y_{i-3}}{h^2} \\
y_i''' &= \frac{5y_i - 18y_{i-1} + 24y_{i-2} - 14y_{i-3} + 3y_{i-4}}{2h^3} \\
y_i'''' &= \frac{3y_i - 14y_{i-1} + 26y_{i-2} - 24y_{i-3} + 11y_{i-4} - 2y_{i-5}}{h^4}
\end{aligned} \tag{5.124}$$

EGM 6422

LECTURE 5/6

EGM 6422

LECTURES 5/6 PART 2

must be solved. Derivations of the more accurate expressions are not given here, but such expressions for several derivatives are included in the list that follows this discussion. Derivations for the higher derivatives are accomplished with much greater facility and far less labor by using *difference, averaging,* and *derivative* operators. Such a method is outside the scope of this text, but it can be found in various books concerned with numerical analysis.[9]

It has been shown that the central-difference expressions for the various derivatives involve values of the function on both sides of the x value at which the derivative of the function is desired. By utilizing the appropriate Taylor series expansions, one can easily obtain expressions for the derivatives which are entirely in terms of values of the function at x_i and points to the right of x_i. These are known as *forward finite-difference* expressions. In a similar manner, derivative expressions which are entirely in terms of values of the function at x_i and points to the *left* of x_i can be found. These are known as *backward finite-difference* expressions. In numerical differentiation, forward-difference expressions are used when data to the left of a point at which a derivative is desired are not available, and backward-difference expressions are used when data to the right of the desired point are not available. Central-difference expressions, however, are more accurate than either forward- or backward-difference expressions. This can be seen by noting the order of the error in the list of differentiation formulas that follows.

Central-difference expressions with error of order h^2

$$\begin{aligned}
y_i' &= \frac{y_{i+1} - y_{i-1}}{2h} \\
y_i'' &= \frac{y_{i+1} - 2y_i + y_{i-1}}{h^2} \\
y_i''' &= \frac{y_{i+2} - 2y_{i+1} + 2y_{i-1} - y_{i-2}}{2h^3} \\
y_i'''' &= \frac{y_{i+2} - 4y_{i+1} + 6y_i - 4y_{i-1} + y_{i-2}}{h^4}
\end{aligned} \tag{5.119}$$

Central-difference expressions with error of order h^4

$$\begin{aligned}
y_i' &= \frac{-y_{i+2} + 8y_{i+1} - 8y_{i-1} + y_{i-2}}{12h} \\
y_i'' &= \frac{-y_{i+2} + 16y_{i+1} - 30y_i + 16y_{i-1} - y_{i-2}}{12h^2} \\
y_i''' &= \frac{-y_{i+3} + 8y_{i+2} - 13y_{i+1} + 13y_{i-1} - 8y_{i-2} + y_{i-3}}{8h^3} \\
y_i'''' &= \frac{-y_{i+3} + 12y_{i+2} - 39y_{i+1} + 56y_i - 39y_{i-1} + 12y_{i-2} - y_{i-3}}{6h^4}
\end{aligned} \tag{5.120}$$

---

[9] M. G. Salvadori and M. L. Baron, *Numerical Methods in Engineering* (Englewood Cliffs, N.J.: Prentice-Hall, 1961).

Forward-difference expressions with error of order h

$$\begin{aligned}
y_i' &= \frac{y_{i+1} - y_i}{h} \\
y_i'' &= \frac{y_{i+2} - 2y_{i+1} + y_i}{h^2} \\
y_i''' &= \frac{y_{i+3} - 3y_{i+2} + 3y_{i+1} - y_i}{h^3} \\
y_i'''' &= \frac{y_{i+4} - 4y_{i+3} + 6y_{i+2} - 4y_{i+1} + y_i}{h^4}
\end{aligned} \tag{5.121}$$

Forward-difference expressions with error of order h^2

$$\begin{aligned}
y_i' &= \frac{-y_{i+2} + 4y_{i+1} - 3y_i}{2h} \\
y_i'' &= \frac{-y_{i+3} + 4y_{i+2} - 5y_{i+1} + 2y_i}{h^2} \\
y_i''' &= \frac{-3y_{i+4} + 14y_{i+3} - 24y_{i+2} + 18y_{i+1} - 5y_i}{2h^3} \\
y_i'''' &= \frac{-2y_{i+5} + 11y_{i+4} - 24y_{i+3} + 26y_{i+2} - 14y_{i+1} + 3y_i}{h^4}
\end{aligned} \tag{5.122}$$

Backward-difference expressions with error of order h

$$\begin{aligned}
y_i' &= \frac{y_i - y_{i-1}}{h} \\
y_i'' &= \frac{y_i - 2y_{i-1} + y_{i-2}}{h^2} \\
y_i''' &= \frac{y_i - 3y_{i-1} + 3y_{i-2} - y_{i-3}}{h^3} \\
y_i'''' &= \frac{y_i - 4y_{i-1} + 6y_{i-2} - 4y_{i-3} + y_{i-4}}{h^4}
\end{aligned} \tag{5.123}$$

Backward-difference expressions with error of order h^2

$$\begin{aligned}
y_i' &= \frac{3y_i - 4y_{i-1} + y_{i-2}}{2h} \\
y_i'' &= \frac{2y_i - 5y_{i-1} + 4y_{i-2} - y_{i-3}}{h^2} \\
y_i''' &= \frac{5y_i - 18y_{i-1} + 24y_{i-2} - 14y_{i-3} + 3y_{i-4}}{2h^3} \\
y_i'''' &= \frac{3y_i - 14y_{i-1} + 26y_{i-2} - 24y_{i-3} + 11y_{i-4} - 2y_{i-5}}{h^4}
\end{aligned} \tag{5.124}$$

First Central-Difference Expressions $O(\Delta x^2)$

$$\begin{aligned}
y_i' &= \frac{y_{i+1} - y_{i-1}}{2(\Delta x)} \\
y_i'' &= \frac{y_{i+1} - 2y_i + y_{i-1}}{(\Delta x)^2} \\
y_i''' &= \frac{y_{i+2} - 2y_{i+1} + 2y_{i-1} - y_{i-2}}{2(\Delta x)^3} \\
y_i'''' &= \frac{y_{i+2} - 4y_{i+1} + 6y_i - 4y_{i-1} + y_{i-2}}{(\Delta x)^4}
\end{aligned} \tag{5-38}$$

Second Central-Difference Expressions $O(\Delta x^4)$

$$\begin{aligned}
y_i' &= \frac{-y_{i+2} + 8y_{i+1} - 8y_{i-1} + y_{i-2}}{12(\Delta x)} \\
y_i'' &= \frac{-y_{i+2} + 16y_{i+1} - 30y_i + 16y_{i-1} - y_{i-2}}{12(\Delta x)^2} \\
y_i''' &= \frac{-y_{i+3} + 8y_{i+2} - 13y_{i+1} + 13y_{i-1} - 8y_{i-2} + y_{i-3}}{8(\Delta x)^3} \\
y_i'''' &= \frac{-y_{i+3} + 12y_{i+2} - 39y_{i+1} + 56y_i - 39y_{i-1} + 12y_{i-2} - y_{i-3}}{6(\Delta x)^4}
\end{aligned} \tag{5-39}$$

irst Forward-Difference Expressions $O(\Delta x)$

$$\begin{aligned}
y_i' &= \frac{y_{i+1} - y_i}{(\Delta x)} \\
y_i'' &= \frac{y_{i+2} - 2y_{i+1} + y_i}{(\Delta x)^2} \\
y_i''' &= \frac{y_{i+3} - 3y_{i+2} + 3y_{i+1} - y_i}{(\Delta x)^3} \\
y_i'''' &= \frac{y_{i+4} - 4y_{i+3} + 6y_{i+2} - 4y_{i+1} + y_i}{(\Delta x)^4}
\end{aligned} \tag{5-40}$$

cond Forward-Difference Expressions $O(\Delta x^2)$

$$\begin{aligned}
{}_i' &= \frac{-y_{i+2} + 4y_{i+1} - 3y_i}{2(\Delta x)} \\
{}' &= \frac{-y_{i+3} + 4y_{i+2} - 5y_{i+1} + 2y_i}{(\Delta x)^2} \\
{}' &= \frac{-3y_{i+4} + 14y_{i+3} - 24y_{i+2} + 18y_{i+1} - 5y_i}{2(\Delta x)^3} \\
{}' &= \frac{-2y_{i+5} + 11y_{i+4} - 24y_{i+3} + 26y_{i+2} - 14y_{i+1} + 3y_i}{(\Delta x)^4}
\end{aligned} \tag{5-41}$$

First Backward-Difference Expressions $O(\Delta x)$

$$\begin{aligned}
y_i' &= \frac{y_i - y_{i-1}}{(\Delta x)} \\
y_i'' &= \frac{y_i - 2y_{i-1} + y_{i-2}}{(\Delta x)^2} \\
y_i''' &= \frac{y_i - 3y_{i-1} + 3y_{i-2} - y_{i-3}}{(\Delta x)^3} \\
y_i'''' &= \frac{y_i - 4y_{i-1} + 6y_{i-2} - 4y_{i-3} + y_{i-4}}{(\Delta x)^4}
\end{aligned} \tag{5-42}$$

Second Backward-Difference Expressions $O(\Delta x^2)$

$$\begin{aligned}
y_i' &= \frac{3y_i - 4y_{i-1} + y_{i-2}}{2(\Delta x)} \\
y_i'' &= \frac{2y_i - 5y_{i-1} + 4y_{i-2} - y_{i-3}}{(\Delta x)^2} \\
y_i''' &= \frac{5y_i - 18y_{i-1} + 24y_{i-2} - 14y_{i-3} + 3y_{i-4}}{2(\Delta x)^3} \\
y_i'''' &= \frac{3y_i - 14y_{i-1} + 26y_{i-2} - 24y_{i-3} + 11y_{i-4} - 2y_{i-5}}{(\Delta x)^4}
\end{aligned} \tag{5-43}$$

EXAMPLE 5-3

In Example 3-4 the Newton-Raphson method was used to determine the output lever angles of a crank-and-lever 4-bar linkage system for each 5° of rotation of the input crank. Now we shall determine the angular velocity and the angular acceleration of the output lever of the same type of mechanism for each 5° of rotation of the input crank, with the latter rotating at a uniform angular velocity of 100 radians/sec.

We can determine the output lever positions ϕ, corresponding to each 5° of crank rotation θ, by utilizing Freudenstein's equation and the Newton-Raphson method, as was done in Example 3-4. Such a set of values, in effect, gives us a series of points on the ϕ versus θ curve, and the ϕ values are stored in memory to provide data for the differentiation processes which follow. The slope of the ϕ-θ curve may be related to the angular velocity of the output lever $d\phi/dt$ if we realize that, with the crank rotating at a constant ω, its angular position is given by

$$\theta = \omega t$$

so that

$$\frac{d\phi}{d\theta} = \frac{1}{\omega}\frac{d\phi}{dt}$$

must be solved. Derivations of the more accurate expressions are not given here, but such expressions for several derivatives are included in the list that follows this discussion. Derivations for the higher derivatives are accomplished with much greater facility and far less labor by using *difference, averaging,* and *derivative* operators. Such a method is outside the scope of this text, but it can be found in various books concerned with numerical analysis.[9]

It has been shown that the central-difference expressions for the various derivatives involve values of the function on both sides of the x value at which the derivative of the function is desired. By utilizing the appropriate Taylor series expansions, one can easily obtain expressions for the derivatives which are entirely in terms of values of the function at x_i and points to the right of x_i. These are known as *forward finite-difference* expressions. In a similar manner, derivative expressions which are entirely in terms of values of the function at x_i and points to the *left* of x_i can be found. These are known as *backward finite-difference* expressions. In numerical differentiation, forward-difference expressions are used when data to the left of a point at which a derivative is desired are not available, and backward-difference expressions are used when data to the right of the desired point are not available. Central-difference expressions, however, are more accurate than either forward- or backward-difference expressions. This can be seen by noting the order of the error in the list of differentiation formulas that follows.

Central-difference expressions with error of order h^2

$$y_i' = \frac{y_{i+1} - y_{i-1}}{2h}$$

$$y_i'' = \frac{y_{i+1} - 2y_i + y_{i-1}}{h^2}$$

$$y_i''' = \frac{y_{i+2} - 2y_{i+1} + 2y_{i-1} - y_{i-2}}{2h^3} \qquad (5.119)$$

$$y_i'''' = \frac{y_{i+2} - 4y_{i+1} + 6y_i - 4y_{i-1} + y_{i-2}}{h^4}$$

Central-difference expressions with error of order h^4

$$y_i' = \frac{-y_{i+2} + 8y_{i+1} - 8y_{i-1} + y_{i-2}}{12h}$$

$$y_i'' = \frac{-y_{i+2} + 16y_{i+1} - 30y_i + 16y_{i-1} - y_{i-2}}{12h^2}$$

$$y_i''' = \frac{-y_{i+3} + 8y_{i+2} - 13y_{i+1} + 13y_{i-1} - 8y_{i-2} + y_{i-3}}{8h^3} \qquad (5.120)$$

$$y_i'''' = \frac{-y_{i+3} + 12y_{i+2} - 39y_{i+1} + 56y_i - 39y_{i-1} + 12y_{i-2} - y_{i-3}}{6h^4}$$

---

[9] M. G. Salvadori and M. L. Baron, *Numerical Methods in Engineering* (Englewood Cliffs, N.J.: Prentice-Hall, 1961).

Forward-difference expressions with error of order h

$$y_i' = \frac{y_{i+1} - y_i}{h}$$

$$y_i'' = \frac{y_{i+2} - 2y_{i+1} + y_i}{h^2}$$

$$y_i''' = \frac{y_{i+3} - 3y_{i+2} + 3y_{i+1} - y_i}{h^3} \qquad (5.121)$$

$$y_i'''' = \frac{y_{i+4} - 4y_{i+3} + 6y_{i+2} - 4y_{i+1} + y_i}{h^4}$$

Forward-difference expressions with error of order h^2

$$y_i' = \frac{-y_{i+2} + 4y_{i+1} - 3y_i}{2h}$$

$$y_i'' = \frac{-y_{i+3} + 4y_{i+2} - 5y_{i+1} + 2y_i}{h^2}$$

$$y_i''' = \frac{-3y_{i+4} + 14y_{i+3} - 24y_{i+2} + 18y_{i+1} - 5y_i}{2h^3} \qquad (5.122)$$

$$y_i'''' = \frac{-2y_{i+5} + 11y_{i+4} - 24y_{i+3} + 26y_{i+2} - 14y_{i+1} + 3y_i}{h^4}$$

Backward-difference expressions with error of order h

$$y_i' = \frac{y_i - y_{i-1}}{h}$$

$$y_i'' = \frac{y_i - 2y_{i-1} + y_{i-2}}{h^2}$$

$$y_i''' = \frac{y_i - 3y_{i-1} + 3y_{i-2} - y_{i-3}}{h^3} \qquad (5.123)$$

$$y_i'''' = \frac{y_i - 4y_{i-1} + 6y_{i-2} - 4y_{i-3} + y_{i-4}}{h^4}$$

Backward-difference expressions with error of order h^2

$$y_i' = \frac{3y_i - 4y_{i-1} + y_{i-2}}{2h}$$

$$y_i'' = \frac{2y_i - 5y_{i-1} + 4y_{i-2} - y_{i-3}}{h^2}$$

$$y_i''' = \frac{5y_i - 18y_{i-1} + 24y_{i-2} - 14y_{i-3} + 3y_{i-4}}{2h^3} \qquad (5.124)$$

$$y_i'''' = \frac{3y_i - 14y_{i-1} + 26y_{i-2} - 24y_{i-3} + 11y_{i-4} - 2y_{i-5}}{h^4}$$

First Central-Difference Expressions $O(\Delta x^2)$

$$\begin{aligned}
y_i' &= \frac{y_{i+1} - y_{i-1}}{2(\Delta x)} \\
y_i'' &= \frac{y_{i+1} - 2y_i + y_{i-1}}{(\Delta x)^2} \\
y_i''' &= \frac{y_{i+2} - 2y_{i+1} + 2y_{i-1} - y_{i-2}}{2(\Delta x)^3} \\
y_i'''' &= \frac{y_{i+2} - 4y_{i+1} + 6y_i - 4y_{i-1} + y_{i-2}}{(\Delta x)^4}
\end{aligned} \tag{5-38}$$

Second Central-Difference Expressions $O(\Delta x^4)$

$$\begin{aligned}
y_i' &= \frac{-y_{i+2} + 8y_{i+1} - 8y_{i-1} + y_{i-2}}{12(\Delta x)} \\
y_i'' &= \frac{-y_{i+2} + 16y_{i+1} - 30y_i + 16y_{i-1} - y_{i-2}}{12(\Delta x)^2} \\
y_i''' &= \frac{-y_{i+3} + 8y_{i+2} - 13y_{i+1} + 13y_{i-1} - 8y_{i-2} + y_{i-3}}{8(\Delta x)^3} \\
y_i'''' &= \frac{-y_{i+3} + 12y_{i+2} - 39y_{i+1} + 56y_i - 39y_{i-1} + 12y_{i-2} - y_{i-3}}{6(\Delta x)^4}
\end{aligned} \tag{5-39}$$

First Forward-Difference Expressions $O(\Delta x)$

$$\begin{aligned}
y_i' &= \frac{y_{i+1} - y_i}{(\Delta x)} \\
y_i'' &= \frac{y_{i+2} - 2y_{i+1} + y_i}{(\Delta x)^2} \\
y_i''' &= \frac{y_{i+3} - 3y_{i+2} + 3y_{i+1} - y_i}{(\Delta x)^3} \\
y_i'''' &= \frac{y_{i+4} - 4y_{i+3} + 6y_{i+2} - 4y_{i+1} + y_i}{(\Delta x)^4}
\end{aligned} \tag{5-40}$$

Second Forward-Difference Expressions $O(\Delta x^2)$

$$\begin{aligned}
y_i' &= \frac{-y_{i+2} + 4y_{i+1} - 3y_i}{2(\Delta x)} \\
y_i'' &= \frac{-y_{i+3} + 4y_{i+2} - 5y_{i+1} + 2y_i}{(\Delta x)^2} \\
y_i''' &= \frac{-3y_{i+4} + 14y_{i+3} - 24y_{i+2} + 18y_{i+1} - 5y_i}{2(\Delta x)^3} \\
y_i'''' &= \frac{-2y_{i+5} + 11y_{i+4} - 24y_{i+3} + 26y_{i+2} - 14y_{i+1} + 3y_i}{(\Delta x)^4}
\end{aligned} \tag{5-41}$$

First Backward-Difference Expressions $O(\Delta x)$

$$\begin{aligned}
y_i' &= \frac{y_i - y_{i-1}}{(\Delta x)} \\
y_i'' &= \frac{y_i - 2y_{i-1} + y_{i-2}}{(\Delta x)^2} \\
y_i''' &= \frac{y_i - 3y_{i-1} + 3y_{i-2} - y_{i-3}}{(\Delta x)^3} \\
y_i'''' &= \frac{y_i - 4y_{i-1} + 6y_{i-2} - 4y_{i-3} + y_{i-4}}{(\Delta x)^4}
\end{aligned} \tag{5-42}$$

Second Backward-Difference Expressions $O(\Delta x^2)$

$$\begin{aligned}
y_i' &= \frac{3y_i - 4y_{i-1} + y_{i-2}}{2(\Delta x)} \\
y_i'' &= \frac{2y_i - 5y_{i-1} + 4y_{i-2} - y_{i-3}}{(\Delta x)^2} \\
y_i''' &= \frac{5y_i - 18y_{i-1} + 24y_{i-2} - 14y_{i-3} + 3y_{i-4}}{2(\Delta x)^3} \\
y_i'''' &= \frac{3y_i - 14y_{i-1} + 26y_{i-2} - 24y_{i-3} + 11y_{i-4} - 2y_{i-5}}{(\Delta x)^4}
\end{aligned} \tag{5-43}$$

EXAMPLE 5-3

In Example 3-4 the Newton-Raphson method was used to determine the output lever angles of a crank-and-lever 4-bar linkage system for each 5° of rotation of the input crank. Now we shall determine the angular velocity and the angular acceleration of the output lever of the same type of mechanism for each 5° of rotation of the input crank, with the latter rotating at a uniform angular velocity of 100 radians/sec.

We can determine the output lever positions ϕ, corresponding to each 5° of crank rotation θ, by utilizing Freudenstein's equation and the Newton-Raphson method, as was done in Example 3-4. Such a set of values, in effect, gives us a series of points on the ϕ versus θ curve, and the ϕ values are stored in memory to provide data for the differentiation processes which follow. The slope of the ϕ-θ curve may be related to the angular velocity of the output lever $d\phi/dt$ if we realize that, with the crank rotating at a constant ω, its angular position is given by

$$\theta = \omega t$$

so that

$$\frac{d\phi}{d\theta} = \frac{1}{\omega}\frac{d\phi}{dt}$$

First Central-Difference Expressions $O(\Delta x^2)$

$$y_i' = \frac{y_{i+1} - y_{i-1}}{2(\Delta x)}$$

$$y_i'' = \frac{y_{i+1} - 2y_i + y_{i-1}}{(\Delta x)^2}$$

$$y_i''' = \frac{y_{i+2} - 2y_{i+1} + 2y_{i-1} - y_{i-2}}{2(\Delta x)^3} \qquad (5\text{-}38)$$

$$y_i'''' = \frac{y_{i+2} - 4y_{i+1} + 6y_i - 4y_{i-1} + y_{i-2}}{(\Delta x)^4}$$

Second Central-Difference Expressions $O(\Delta x^4)$

$$y_i' = \frac{-y_{i+2} + 8y_{i+1} - 8y_{i-1} + y_{i-2}}{12(\Delta x)}$$

$$y_i'' = \frac{-y_{i+2} + 16y_{i+1} - 30y_i + 16y_{i-1} - y_{i-2}}{12(\Delta x)^2}$$

$$y_i''' = \frac{-y_{i+3} + 8y_{i+2} - 13y_{i+1} + 13y_{i-1} - 8y_{i-2} + y_{i-3}}{8(\Delta x)^3} \qquad (5\text{-}39)$$

$$y_i'''' = \frac{-y_{i+3} + 12y_{i+2} - 39y_{i+1} + 56y_i - 39y_{i-1} + 12y_{i-2} - y_{i-3}}{6(\Delta x)^4}$$

First Forward-Difference Expressions $O(\Delta x)$

$$y_i' = \frac{y_{i+1} - y_i}{(\Delta x)}$$

$$y_i'' = \frac{y_{i+2} - 2y_{i+1} + y_i}{(\Delta x)^2}$$

$$y_i''' = \frac{y_{i+3} - 3y_{i+2} + 3y_{i+1} - y_i}{(\Delta x)^3} \qquad (5\text{-}40)$$

$$y_i'''' = \frac{y_{i+4} - 4y_{i+3} + 6y_{i+2} - 4y_{i+1} + y_i}{(\Delta x)^4}$$

econd Forward-Difference Expressions $O(\Delta x^2)$

$$_i' = \frac{-y_{i+2} + 4y_{i+1} - 3y_i}{2(\Delta x)}$$

$$'' = \frac{-y_{i+3} + 4y_{i+2} - 5y_{i+1} + 2y_i}{(\Delta x)^2}$$

$$' = \frac{-3y_{i+4} + 14y_{i+3} - 24y_{i+2} + 18y_{i+1} - 5y_i}{2(\Delta x)^3} \qquad (5\text{-}41)$$

$$' = \frac{-2y_{i+5} + 11y_{i+4} - 24y_{i+3} + 26y_{i+2} - 14y_{i+1} + 3y_i}{(\Delta x)^4}$$

First Backward-Difference Expressions $O(\Delta x)$

$$y_i' = \frac{y_i - y_{i-1}}{(\Delta x)}$$

$$y_i'' = \frac{y_i - 2y_{i-1} + y_{i-2}}{(\Delta x)^2}$$

$$y_i''' = \frac{y_i - 3y_{i-1} + 3y_{i-2} - y_{i-3}}{(\Delta x)^3} \qquad (5\text{-}42)$$

$$y_i'''' = \frac{y_i - 4y_{i-1} + 6y_{i-2} - 4y_{i-3} + y_{i-4}}{(\Delta x)^4}$$

Second Backward-Difference Expressions

$$y_i' = \frac{3y_i - 4y_{i-1} + y_{i-2}}{2(\Delta x)}$$

$O(\Delta x^2)$

$$y_i'' = \frac{2y_i - 5y_{i-1} + 4y_{i-2} - y_{i-3}}{(\Delta x)^2}$$

$$y_i''' = \frac{5y_i - 18y_{i-1} + 24y_{i-2} - 14y_{i-3} + 3y_{i-4}}{2(\Delta x)^3} \qquad (5\text{-}43)$$

$$y_i'''' = \frac{3y_i - 14y_{i-1} + 26y_{i-2} - 24y_{i-3} + 11y_{i-4} - 2y_{i-5}}{(\Delta x)^4}$$

EXAMPLE 5-3

In Example 3-4 the Newton-Raphson method was used to determine the output lever angles of a crank-and-lever 4-bar linkage system for each 5° of rotation of the input crank. Now we shall determine the angular velocity and the angular acceleration of the output lever of the same type of mechanism for each 5° of rotation of the input crank, with the latter rotating at a uniform angular velocity of 100 radians/sec.

We can determine the output lever positions ϕ, corresponding to each 5° of crank rotation θ, by utilizing Freudenstein's equation and the Newton-Raphson method, as was done in Example 3-4. Such a set of values, in effect, gives us a series of points on the ϕ versus θ curve, and the ϕ values are stored in memory to provide data for the differentiation processes which follow. The slope of the ϕ-θ curve may be related to the angular velocity of the output lever $d\phi/dt$ if we realize that, with the crank rotating at a constant ω, its angular position is given by

$$\theta = \omega t$$

so that

$$\frac{d\phi}{d\theta} = \frac{1}{\omega}\frac{d\phi}{dt}$$

must be solved. Derivations of the more accurate expressions are not given here, but such expressions for several derivatives are included in the list that follows this discussion. Derivations for the higher derivatives are accomplished with much greater facility and far less labor by using *difference, averaging,* and *derivative* operators. Such a method is outside the scope of this text, but it can be found in various books concerned with numerical analysis.[9]

It has been shown that the central-difference expressions for the various derivatives involve values of the function on both sides of the x value at which the derivative of the function is desired. By utilizing the appropriate Taylor series expansions, one can easily obtain expressions for the derivatives which are entirely in terms of values of the function at x_i and points to the right of x_i. These are known as *forward finite-difference* expressions. In a similar manner, derivative expressions which are entirely in terms of values of the function at x_i and points to the *left* of x_i can be found. These are known as *backward finite-difference* expressions. In numerical differentiation, forward-difference expressions are used when data to the left of a point at which a derivative is desired are not available, and backward-difference expressions are used when data to the right of the desired point are not available. Central-difference expressions, however, are more accurate than either forward- or backward-difference expressions. This can be seen by noting the order of the error in the list of differentiation formulas that follows.

Central-difference expressions with error of order h^2

$$
\begin{aligned}
y'_i &= \frac{y_{i+1} - y_{i-1}}{2h} \\
y''_i &= \frac{y_{i+1} - 2y_i + y_{i-1}}{h^2} \\
y'''_i &= \frac{y_{i+2} - 2y_{i+1} + 2y_{i-1} - y_{i-2}}{2h^3} \\
y''''_i &= \frac{y_{i+2} - 4y_{i+1} + 6y_i - 4y_{i-1} + y_{i-2}}{h^4}
\end{aligned}
\tag{5.119}
$$

Central-difference expressions with error of order h^4

$$
\begin{aligned}
y'_i &= \frac{-y_{i+2} + 8y_{i+1} - 8y_{i-1} + y_{i-2}}{12h} \\
y''_i &= \frac{-y_{i+2} + 16y_{i+1} - 30y_i + 16y_{i-1} - y_{i-2}}{12h^2} \\
y'''_i &= \frac{-y_{i+3} + 8y_{i+2} - 13y_{i+1} + 13y_{i-1} - 8y_{i-2} + y_{i-3}}{8h^3} \\
y''''_i &= \frac{-y_{i+3} + 12y_{i+2} - 39y_{i+1} + 56y_i - 39y_{i-1} + 12y_{i-2} - y_{i-3}}{6h^4}
\end{aligned}
\tag{5.120}
$$

---

[9] M. G. Salvadori and M. L. Baron, *Numerical Methods in Engineering* (Englewood Cliffs, N.J.: Prentice-Hall, 1961).

Forward-difference expressions with error of order h

$$
\begin{aligned}
y'_i &= \frac{y_{i+1} - y_i}{h} \\
y''_i &= \frac{y_{i+2} - 2y_{i+1} + y_i}{h^2} \\
y'''_i &= \frac{y_{i+3} - 3y_{i+2} + 3y_{i+1} - y_i}{h^3} \\
y''''_i &= \frac{y_{i+4} - 4y_{i+3} + 6y_{i+2} - 4y_{i+1} + y_i}{h^4}
\end{aligned}
\tag{5.121}
$$

Forward-difference expressions with error of order h^2

$$
\begin{aligned}
y'_i &= \frac{-y_{i+2} + 4y_{i+1} - 3y_i}{2h} \\
y''_i &= \frac{-y_{i+3} + 4y_{i+2} - 5y_{i+1} + 2y_i}{h^2} \\
y'''_i &= \frac{-3y_{i+4} + 14y_{i+3} - 24y_{i+2} + 18y_{i+1} - 5y_i}{2h^3} \\
y''''_i &= \frac{-2y_{i+5} + 11y_{i+4} - 24y_{i+3} + 26y_{i+2} - 14y_{i+1} + 3y_i}{h^4}
\end{aligned}
\tag{5.122}
$$

Backward-difference expressions with error of order h

$$
\begin{aligned}
y'_i &= \frac{y_i - y_{i-1}}{h} \\
y''_i &= \frac{y_i - 2y_{i-1} + y_{i-2}}{h^2} \\
y'''_i &= \frac{y_i - 3y_{i-1} + 3y_{i-2} - y_{i-3}}{h^3} \\
y''''_i &= \frac{y_i - 4y_{i-1} + 6y_{i-2} - 4y_{i-3} + y_{i-4}}{h^4}
\end{aligned}
\tag{5.123}
$$

Backward-difference expressions with error of order h^2

$$
\begin{aligned}
y'_i &= \frac{3y_i - 4y_{i-1} + y_{i-2}}{2h} \\
y''_i &= \frac{2y_i - 5y_{i-1} + 4y_{i-2} - y_{i-3}}{h^2} \\
y'''_i &= \frac{5y_i - 18y_{i-1} + 24y_{i-2} - 14y_{i-3} + 3y_{i-4}}{2h^3} \\
y''''_i &= \frac{3y_i - 14y_{i-1} + 26y_{i-2} - 24y_{i-3} + 11y_{i-4} - 2y_{i-5}}{h^4}
\end{aligned}
\tag{5.124}
$$

First Central-Difference Expressions $O(\Delta x^2)$

$$y_i' = \frac{y_{i+1} - y_{i-1}}{2(\Delta x)}$$

$$y_i'' = \frac{y_{i+1} - 2y_i + y_{i-1}}{(\Delta x)^2}$$

$$y_i''' = \frac{y_{i+2} - 2y_{i+1} + 2y_{i-1} - y_{i-2}}{2(\Delta x)^3} \tag{5-38}$$

$$y_i'''' = \frac{y_{i+2} - 4y_{i+1} + 6y_i - 4y_{i-1} + y_{i-2}}{(\Delta x)^4}$$

Second Central-Difference Expressions $O(\Delta x^4)$

$$y_i' = \frac{-y_{i+2} + 8y_{i+1} - 8y_{i-1} + y_{i-2}}{12(\Delta x)}$$

$$y_i'' = \frac{-y_{i+2} + 16y_{i+1} - 30y_i + 16y_{i-1} - y_{i-2}}{12(\Delta x)^2}$$

$$y_i''' = \frac{-y_{i+3} + 8y_{i+2} - 13y_{i+1} + 13y_{i-1} - 8y_{i-2} + y_{i-3}}{8(\Delta x)^3} \tag{5-39}$$

$$y_i'''' = \frac{-y_{i+3} + 12y_{i+2} - 39y_{i+1} + 56y_i - 39y_{i-1} + 12y_{i-2} - y_{i-3}}{6(\Delta x)^4}$$

First Forward-Difference Expressions $O(\Delta x)$

$$y_i' = \frac{y_{i+1} - y_i}{(\Delta x)}$$

$$y_i'' = \frac{y_{i+2} - 2y_{i+1} + y_i}{(\Delta x)^2}$$

$$y_i''' = \frac{y_{i+3} - 3y_{i+2} + 3y_{i+1} - y_i}{(\Delta x)^3} \tag{5-40}$$

$$y_i'''' = \frac{y_{i+4} - 4y_{i+3} + 6y_{i+2} - 4y_{i+1} + y_i}{(\Delta x)^4}$$

Second Forward-Difference Expressions $O(\Delta x^2)$

$$y_i' = \frac{-y_{i+2} + 4y_{i+1} - 3y_i}{2(\Delta x)}$$

$$y_i'' = \frac{-y_{i+3} + 4y_{i+2} - 5y_{i+1} + 2y_i}{(\Delta x)^2}$$

$$y_i''' = \frac{-3y_{i+4} + 14y_{i+3} - 24y_{i+2} + 18y_{i+1} - 5y_i}{2(\Delta x)^3} \tag{5-41}$$

$$y_i'''' = \frac{-2y_{i+5} + 11y_{i+4} - 24y_{i+3} + 26y_{i+2} - 14y_{i+1} + 3y_i}{(\Delta x)^4}$$

First Backward-Difference Expressions $O(\Delta x)$

$$y_i' = \frac{y_i - y_{i-1}}{(\Delta x)}$$

$$y_i'' = \frac{y_i - 2y_{i-1} + y_{i-2}}{(\Delta x)^2}$$

$$y_i''' = \frac{y_i - 3y_{i-1} + 3y_{i-2} - y_{i-3}}{(\Delta x)^3} \tag{5-42}$$

$$y_i'''' = \frac{y_i - 4y_{i-1} + 6y_{i-2} - 4y_{i-3} + y_{i-4}}{(\Delta x)^4}$$

Second Backward-Difference Expressions $O(\Delta x^2)$

$$y_i' = \frac{3y_i - 4y_{i-1} + y_{i-2}}{2(\Delta x)}$$

$$y_i'' = \frac{2y_i - 5y_{i-1} + 4y_{i-2} - y_{i-3}}{(\Delta x)^2}$$

$$y_i''' = \frac{5y_i - 18y_{i-1} + 24y_{i-2} - 14y_{i-3} + 3y_{i-4}}{2(\Delta x)^3} \tag{5-43}$$

$$y_i'''' = \frac{3y_i - 14y_{i-1} + 26y_{i-2} - 24y_{i-3} + 11y_{i-4} - 2y_{i-5}}{(\Delta x)^4}$$

EXAMPLE 5-3

In Example 3-4 the Newton-Raphson method was used to determine the output lever angles of a crank-and-lever 4-bar linkage system for each 5° of rotation of the input crank. Now we shall determine the angular velocity and the angular acceleration of the output lever of the same type of mechanism for each 5° of rotation of the input crank, with the latter rotating at a uniform angular velocity of 100 radians/sec.

We can determine the output lever positions ϕ, corresponding to each 5° of crank rotation θ, by utilizing Freudenstein's equation and the Newton-Raphson method, as was done in Example 3-4. Such a set of values, in effect, gives us a series of points on the ϕ versus θ curve, and the ϕ values are stored in memory to provide data for the differentiation processes which follow. The slope of the ϕ-θ curve may be related to the angular velocity of the output lever $d\phi/dt$ if we realize that, with the crank rotating at a constant ω, its angular position is given by

$$\theta = \omega t$$

so that

$$\frac{d\phi}{d\theta} = \frac{1}{\omega}\frac{d\phi}{dt}$$

EGM 6422

LECTURE 8/9

SIMULTANEOUS NON LINEAR EQUATION

$$f(x,y)=0$$
$$g(x,y)=0$$

WANT x, y

→ Ⓐ IF I HAVE $x_0, y_0 \Rightarrow f(x_0,y_0)=f_0$

$$g(x_0,y_0)=g_0$$

$$x_1 = x_0+\Delta x$$
$$y_1 = y_0+\Delta y$$

$$f(x_1,y_1)=0$$
$$g(x_1,y_1)=0$$

$$0 = f(x_1,y_1) = f(x_0+\Delta x, y_0+\Delta y) = f(x_0,y_0) + \frac{\partial f}{\partial x}\Delta x + \frac{\partial f}{\partial y}\Delta y + \text{h.o.t.}$$
$$0 = g(x_1,y_1) = g(x_0+\Delta x, y_0+\Delta y) = g(x_0,y_0) + \frac{\partial g}{\partial x}\Delta x + \frac{\partial g}{\partial y}\Delta y + \text{h.o.t.}$$

$$-f_0 = \frac{\partial f}{\partial x}\Delta x + \frac{\partial f}{\partial y}\Delta y$$
$$-g_0 = \frac{\partial g}{\partial x}\Delta x + \frac{\partial g}{\partial y}\Delta y$$

neglect h.o.t.
$O(\Delta x^2, \Delta y^2)$

$$\Delta x = \frac{\begin{vmatrix} -f_0 & \frac{\partial f}{\partial y} \\ -g_0 & \partial g/\partial y \end{vmatrix}}{\begin{vmatrix} \frac{\partial f}{\partial x} & \frac{\partial f}{\partial y} \\ \frac{\partial g}{\partial x} & \frac{\partial g}{\partial y} \end{vmatrix}}$$

- all f, g are eval. at x_0, y_0
- all derivs are eval. at x_0, y_0

$$\Delta y = \frac{\begin{vmatrix} \partial f/\partial x & -f_0 \\ \partial g/\partial x & -g_0 \end{vmatrix}}{\begin{vmatrix} \frac{\partial f}{\partial x} & \partial f/\partial y \\ \partial g/\partial x & \partial g/\partial y \end{vmatrix}}$$

$$\begin{vmatrix} A & B \\ C & D \end{vmatrix} = AD - BC$$

all fns & derivs are evaluated at x_0, y_0

$x_1 = x_0 + \Delta x$

$y_1 = y_0 + \Delta y$

check if $|f(x_1,y_1)| < \epsilon$

$|g(x_1,y_1)| < \epsilon$ } determined by user $0.0001 = \epsilon$

IF YES THEN PRINT & END

IF NO THEN

Ⓐ $x_0 := x_1$

$y_0 := y_1$

$f(x,y) = 2xy - y^3$

$\frac{\partial f}{\partial x} = 2y \qquad \frac{\partial f}{\partial y} = 2x - 3y^2$

$g(x,y) = e^{xy}$

$\frac{\partial g}{\partial x} = e^{xy} \cdot y \qquad \frac{\partial g}{\partial y} = e^{xy} \cdot x$

Δx small

& Δy small

$g(x_0,y_0) \qquad g(x_0+\Delta x, y_0) \qquad g(x_0, y_0+\Delta y)$

$$\frac{\partial g}{\partial x} \cong \frac{g(x_0+\Delta x, y_0) - g(x_0,y_0)}{\Delta x}$$

$$\partial g/\partial y \cong [g(x_0, y_0+\Delta y) - g(x_0,y_0)]/\Delta y$$

A $f(x_0) = f(x_0)$

B $f(x_0+\Delta x) = f(x_0) + f'(x_0)\Delta x + f''(x_0)\frac{\Delta x^2}{2!} + f'''(x_0)\Delta x^3/3! + \cdots$

C $f(x_0-\Delta x) = f(x_0) - f'(x_0)\Delta x + f''(x_0)\Delta x^2/2! - f'''(x_0)\Delta x^3/3! + \cdots$

D $f(x_0+2\Delta x) + f(x_0) + f'(x_0)2\Delta x + f''(x_0)4\Delta x^2/2! + f'''(x_0)8\Delta x^3/3!$

$$\frac{A f(x_0) + B f(x_0+\Delta x) + C f(x_0-\Delta x)}{(B-C)\Delta x} = \frac{(A+B+C)f(x_0)}{(B-C)\Delta x} + \frac{(B-C)f'(x_0)\Delta x}{(B-C)\Delta x} + \frac{(B+C)\frac{\Delta x^2}{2!}f''(x_0)}{(B-C)\Delta x} + \frac{(B-C)f'''(x_0)\frac{\Delta x^3}{3!}}{(B-C)\Delta x} + \cdots$$

if want $f'(x_0)$ with error Δx^2 then $\Rightarrow \frac{(B+C)}{B-C}\Delta x = 0$ $\frac{(B+C)\frac{\Delta x^2}{2!}}{B-C} = 0 \Rightarrow B = -C$

$$\Rightarrow \frac{-(B+C)f(x_0)}{(B-C)\Delta x} + \frac{B}{(B-C)\Delta x}f(x_0+\Delta x) + \frac{C\,f(x_0-\Delta x)}{(B-C)\Delta x} = f'(x_0) + \frac{\Delta x^2}{3!}f'''(x_0) \text{ error}$$

$$\Rightarrow \overset{\Vert}{0} \qquad \frac{1}{2\Delta x}f_1 \qquad -\frac{1}{2}\frac{f_{-1}}{\Delta x}$$

if want $f'(x_0)$ with error Δx then $\frac{B-C}{B-C} = 1$ & $\frac{(B+C)}{B-C}\frac{\Delta x}{2}$ can be anything

choose simplest namely let $C = 0$ $B = 1$ or
$C = 1$ $B = 0$

$C = 0$ $B = 1$ $\frac{f(x_0+\Delta x) - f(x_0)}{\Delta x} = f'(x_0) + f''(x_0)\frac{\Delta x}{2}$

$C = 1$ $B = 0$ $\frac{-f(x_0) + f(x_0-\Delta x)}{\Delta x} = f'(x_0) - f''(x_0)\frac{\Delta x}{2}$

$$A f(x_0) + B f(x_0+\Delta x) + D f(x_0+2\Delta x) = (A+B+D)f(x_0) + (B+2D)f'(x_0)\Delta x + (B+4D)f''(x_0)\Delta x^2/2! + (B+8D)f'''(x_0)\Delta x^3/3! + \cdots$$

$$\frac{-(B+D)f(x_0) + B f(x_0+\Delta x) + D f(x_0+2\Delta x)}{(B+2D)\Delta x} = f'(x_0) + \frac{(B+4D)}{B+2D}f''(x_0)\frac{\Delta x}{2!} + \frac{B+8D}{B+2D}f'''(x_0)\frac{\Delta x^2}{3!}$$

want $f'(x_0)$ to $O(\Delta x^2)$ $\Rightarrow \frac{(B+4D)}{B+2D}\frac{\Delta x}{2!} = 0$ $B = -4D$

$$-\frac{(-3D)}{-2D}f_0 \quad \frac{-4D}{-2D}f_1 \quad + \frac{D}{-2D}f_2 = f'(x_0) + \frac{4D}{-2D}f'''(x_0)\frac{\Delta x^2}{3!}$$

$$\frac{-\frac{3}{2}f_0 + \frac{4}{2}f_1 - \frac{1}{2}f_2}{\Delta x} = f'(x_0)$$

FINITE DIFFERENCE FORMS

Given x_0 & Δx $\quad f(x)$

$f'(x_0)$

$$f(x_0+\Delta x) = f(x_1) = f(x_0) + f'(x_0)\Delta x + \underline{f''(x_0)\frac{\Delta x^2}{2!}} + \cdots$$

Forward difference formula

$$\frac{f(x_1) - f(x_0)}{\Delta x} = f'(x_0) + \vdots\, \underbrace{f''(x_0)\frac{\Delta x}{2!} + \cdots}_{O(\Delta x)}$$

$$f(x_0-\Delta x) = f(x_{-1}) = f(x_0) - f'(x_0)\Delta x + f''(x_0)\frac{\Delta x^2}{2!} + \cdots$$

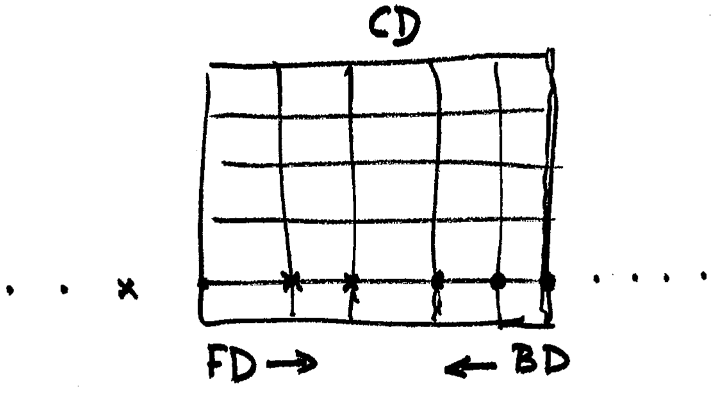
Backward Difference formula

$$\frac{f(x_0) - f(x_{-1})}{\Delta x} = f'(x_0) - \vdots\, \underbrace{f''(x_0)\cdot\frac{\Delta x}{2!} + \cdots}_{O(\Delta x)}$$

$$f(x_1) + f(x_{-1}) = 2f'(x_0)\Delta x + 2f'''(x_0)\frac{\Delta x^3}{3!} + \text{h.o.t..}$$

Centered Difference method

$$\frac{f(x_1) - f(x_{-1})}{2\Delta x} = f'(x_0) + \vdots\, \underbrace{2f'''(x_0)\frac{\Delta x^2}{2!} + \text{h.o.t.}}_{O(\Delta x^2)}$$



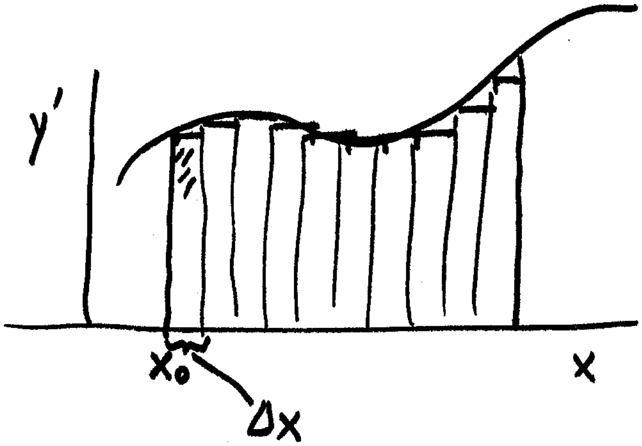
SOLUTION OF ODEs

$$y' = \frac{dy}{dx} = f(x,y)$$

$$y(x=x_0) = y_0$$

$$\int_{y_0}^{y} d\bar{y} = \int_{x_0}^{x} f(\bar{x},y)\,d\bar{x}$$

$$y = y_0 + \int_{x_0}^{x} f(\bar{x},y)\,d\bar{x}$$



Euler's Method

$$\longrightarrow y = y_0 + f(x_0,y_0)\Delta x = y_0 + y'\Big|_{x=x_0}\Delta x + \underbrace{\underline{\underline{y''\frac{\Delta x^2}{2!}+\dots}}}_{\text{error}}$$

$\Delta x = x_1 - x_0$ then $y \cong y_1$

$x_0 := x_1$

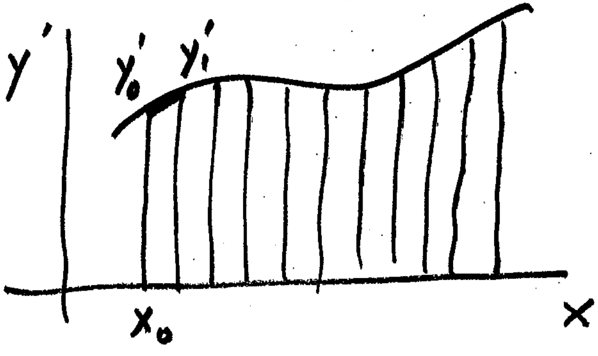
$y_0 := y_1$

Local error = per step error $O(\Delta x^2)$

global error $O(\Delta x)$

$O(\Delta x^2)$ per step

Modified Euler's Method.



$$y_{1_c} = y_0 + \frac{y_0' + y_1'}{2}\Delta x$$

$y_0' = f(x_0,y_0)$

$y_1' = f(x_1,y_1)$ problem

$y_{1_p} \cong y_0 + f(x_0,y_0)\Delta x$

$x_1 = x_0 + \Delta x$

$y_0' = y'\big|_{x=x_0}$

$y_1' = y'\big|_{x=x_0+\Delta x=x_1}$

L.E. $O(\Delta x^3)$

G.E. $O(\Delta x^2)$

Taylor's Algorithm of order k

$$y_1 = y_0 + \Delta x \, T_k(x_0, y_0) \qquad y_1 \neq y(x_1)$$

$$T_k(x,y) = f(x,y) + \frac{\Delta x}{2!} f'(x,y) + \frac{\Delta x^2}{3!} f''(x,y) + \dots + \frac{\Delta x^{k-1}}{k!} f^{(k-1)}(x,y)$$

$k=1 \quad T_1(x,y) = f(x,y)$ Euler's Method

$k=2 \quad T_2(x,y) = f(x,y) + \frac{\Delta x}{2!} f'(x,y)$ Modified Euler

$$f'(x,y) = \frac{\partial f}{\partial x} + \frac{\partial f}{\partial y} \cdot y' = \frac{\partial f}{\partial x} + \frac{\partial f}{\partial y} \cdot f$$

APP 7. PG 143-144

APP8 PG 144-145

Euler's method (Taylor's Algorithm of order 1)

$$y' = f(x,y).$$

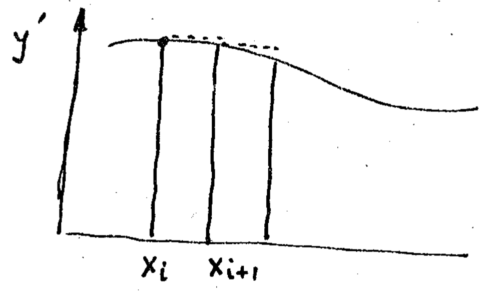
$$\int_{y_0}^{y_1} dy = \int_{x_0}^{x_1} y'\,dx = \int_{x_0}^{x_1} f(x,y)\,dx$$

ASSUME $f(x,y) \cong$ const.

RECTANGULAR AREA UNDER y' vs. X CURVE

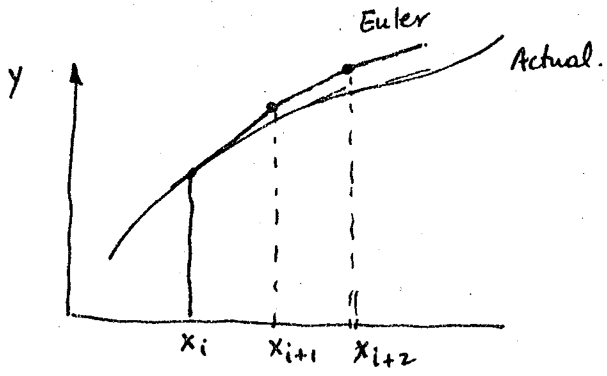
$$y_1 = y_0 + f(x_0,y_0)\,\Delta x = y_0 + y'\Big|_{x=x_0}\Delta x$$

$$\therefore \quad y_{i+1} = y_i + y_i'\,\Delta x \qquad y_i' = f(x_i,y_i).$$



SELF STARTING.

ASSUMES $y_i' =$ const.



Error exists since

$$y_1 = y_0 + y_0'\Delta x + y''\frac{\Delta x^2}{2} + \dots$$

local error is $O(\Delta x)^2$

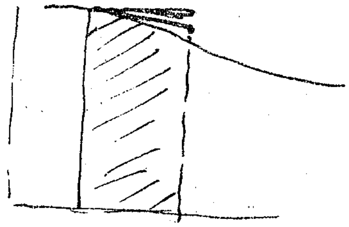
global error is $O(\Delta x)$

Modified Euler method. (Taylor's Algorithm of order 2)

let $\tilde{y}_{i+1} = y_i + y_i'\Delta x$ let $\tilde{y}_{i+1}$ be the predicted y_{i+1}.

to find the corrected y_{i+1} define $\tilde{y}'_{i+1} = f(x_{i+1}, \tilde{y}_{i+1}) = f(x_i,y_i) + \frac{\partial f}{\partial x}\Delta x + \frac{\partial f}{\partial y} y_i'\Delta x$

and then take $\dfrac{\tilde{y}'_{i+1} + y_i'}{2}\Delta x + y_i = y_{i+1}$



this has ~~an~~ local error of $O(\Delta x^3)$

global error of $O(\Delta x^2)$

Modified Euler is self starting

General Taylor Algorith of order k.

$$T_k(x,y) = f(x,y) + \frac{\Delta x}{2!} f'(x,y) + \frac{\Delta x^2}{3!} f''(x,y) + \dots + \frac{\Delta x^{k-1}}{k!} f^{(k-1)}(x,y)$$

if we now chose a step size $h = (b-a)/N$

- let $x_0 = a$ $\quad x_i = a + \Delta x h$ $\quad y_i \doteq y(x_i)$ $\quad x_N = b$
- Then $y_{i+1} = y_i + \Delta x \, T_k(x_i, y_i)$ $\qquad i = 0, 1, 2, \dots, N-1$

$\therefore$ for $k=1$ $\quad T_1 = f(x,y)$

$k=2$ $\quad T_2 = f(x,y) + \frac{\Delta x}{2!} \underbrace{f'(x,y)}_{df/dx}$

where $f' = f_x + f_y \cdot f = \frac{\partial f}{\partial x} + \overset{\curvearrowright}{f} \cdot \frac{\partial f}{\partial y} = \frac{\partial f}{\partial x} + \frac{\partial f}{\partial y}\frac{dy}{dx}$

THE REASON WHY NONE ARE USED ABOVE $k=2$ IS THE REQUIREMENT OF DEFINING THE DERIVATIVES OF f

error for $k^{\underline{th}}$ order is $\frac{\Delta x^{k+1}}{(k+1)!} f^{(k)}(x)$ or $\frac{\Delta x^{k+1}}{(k+1)!} y^{k+1}$

WHAT IF you had $y'' = f(x, y, y')$

HOW WOULD THE MODIFIED EULER METHOD LOOK.

$$y' = p$$
$$y'' = p' = f(x,y,p)$$
$$y' = g(x,y,p) = p$$

let
$$\left.\begin{aligned} \tilde{p} &= p_0 + f\,\Delta x \\ \tilde{y} &= y_0 + g\,(\Delta x) \\ \tilde{x} &= x_0 + \Delta x \end{aligned}\right\} \rightarrow \text{find} \quad \begin{aligned} f(\tilde{x}, \tilde{y}, \tilde{p}) &= \tilde{p}' \\ g(\tilde{x}, \tilde{y}, \tilde{p}) &= \tilde{y}' \end{aligned}$$

then
$$p_{i+1} = p_i + \frac{p_i' + \tilde{p}_{i+1}'}{2} \Delta x$$
$$y_{i+1} = y_i + \frac{y_i' + \tilde{y}_{i+1}'}{2} \Delta x$$

Non-self starting modified Euler

Multistep

$$\tilde{y}_{i+1} = P(y_{i+1}) = y_{i-1} + 2y_i' \Delta x$$

where $\frac{dy}{dt} \cong \frac{y_{i+1} - y_{i-1}}{2\Delta}$ $O(\Delta x^2)$

Then $P(y_{i+1}') = f(x_{i+1}, \tilde{y}_{i+1}) = \tilde{y}_{i+1}'$

then $y_{i+1} = y_i + \frac{\Delta}{2}\left[y_i' + \tilde{y}_{i+1}'\right]$

To start: use self starting modified Euler since it has same error per step

---

For Runge Kutta

$$y_{i+1} = y_i + a_1 k_1 + a_2 k_2$$

where

$$k_1 = h\, f(x,y)$$

$$k_2 = h\, f(x + p_1 h,\ y + q_1 k_1)$$

$$y_{i+1} = y_i + p_1 h \frac{\partial f}{\partial x} + q_1 \overset{hf}{k_1} \frac{\partial f}{\partial y} + \frac{1}{2}\left[p_1 q_1 h k \frac{\partial f}{\partial x}\frac{\partial f}{\partial y} + (p_1 h)^2 \frac{\partial^2 f}{\partial x^2} + (q_1 k)^2 \frac{\partial^2 f}{\partial y^2}\right]$$

$$a_2 k_2 = a_2 h\left\{f(x,y) + p_1 h \frac{\partial f}{\partial x} + q_1 k_1 \frac{\partial f}{\partial y} + \dots\right\}$$

$$a_1 k_1 \approx a_1 h f$$

$\therefore\ a_2 = \frac{1}{2}\quad p_1 = 1\quad q_1 = 1$ $\qquad \therefore\ (a_1 + a_2) h f$

but $y_{i+1} = y_i + \underset{f}{\frac{dy}{dx}} \cdot \Delta x + \frac{d^2y}{dx^2} \cdot \frac{\Delta x^2}{2} + \dots$ by taylor series

$$\left[\frac{\partial f}{\partial x} + \frac{\partial f}{\partial y} f\right] \frac{\Delta x^2}{2}$$

$$\frac{d}{dx} f(x,y)$$

$\therefore\ (a_1 + a_2) h f = f \Delta x$ or $a_1 + a_2 = 1$

$$\frac{\Delta x^2}{2} f_x = p_1 h$$

Runge-Kutta

$$y_{n+1} = y_n + a_1 k_1 + a_2 k_2$$

where $k_1 = hf(x_n, y_n) = hf_n$

$$k_2 = h f(x_n + \alpha h, y_n + \beta k_1)$$

$f(x+\Delta x, y+\Delta y) = f(x,y) + \frac{\partial f}{\partial x}\Delta x + \frac{\partial f}{\partial y}\Delta y + hot$

$\Delta x = \alpha h \quad \Delta y = \beta k_1$

$$y_{n+1} = y_n + a_1\left[hf_n\right] + a_2 h\left\{f_n + \frac{\partial f}{\partial x}\Big|_n \alpha h + \frac{\partial f}{\partial y}\Big|_n \beta k_1\right\}\Big]$$

error is $O(\text{stepsize}^3)$

$$= y_n + \underbrace{h(a_1+a_2)}\, f_n + \underbrace{a_2\alpha}\, h^2 f_x + \underbrace{a_2\beta}\, h^2 f_n \frac{\partial f}{\partial y} + \dots$$

TAYLOR SERIES

$$y(x_n+\Delta x) \cong y_{n+1} = y_n + \underbrace{\frac{dy}{dx}}_{f} h + \frac{d}{dx}\left(\frac{dy}{dx}\right)\frac{h^2}{2} + \dots$$

$\Delta x = h = \text{stepsize}$

$$\frac{d}{dx} f(x,y)$$

$$\frac{\partial f}{\partial x} + \frac{\partial f}{\partial y}\overset{f}{\cancel{\frac{dy}{dx}}}$$

now $\frac{d^3}{dx^3}y = \frac{d}{dx}\left\{f_x + f_y f\right\} = \left(\frac{\partial}{\partial x} + y'\frac{\partial}{\partial y}\right)($

$= f_{xx} + 2f_{xy}f + f_{yy}f^2 + f_y f_x + f$

$\therefore \quad a_1 + a_2 = 1 \qquad a_2\alpha = 1/2 \qquad a_2\beta = 1/2$

INFINITE SOL'NS

could have picked $a_1 = 3/4 \quad a_2 = 1/4 \quad \alpha = \beta = 2$

min bound to error: $\beta = 3/2 \quad \alpha = 3/2 \quad a_2 = \frac{1}{3} \quad a_1 = 2/3$

$a_2 = 2 \quad a_1 = -1 \quad \beta = \frac{1}{4} \quad \alpha = 1/4$

midpoint method: $a_1 = 0 \quad a_2 = 1 \quad \alpha = 1/2 \quad \beta = 1/2$

ONE CASE: if $a_1 = a_2 = 1/2 \quad \alpha = \beta = 1$

get modified euler

$$y_{n+1} = y_n + \frac{1}{2}(k_1 + k_2)$$

$$k_1 = hf_n = h y_n'$$

$$k_2 = h f(x_n + h, y_n + k_1)$$

$$= h f(\underbrace{x_n + h}_{x_{n+1}}, \underbrace{y_n + hf_n}_{\tilde{y}_{n+1}}) = h\tilde{y}_{n+1}'$$

in general $y_{n+1} = y_n + \sum_{i=1}^{m} a_i k_i$

$$k_i = h f\left(x_n + \alpha_{i-1} h, y_n + \sum_{j=1}^{i-1} \beta_{i-1\,j} k_j\right)$$

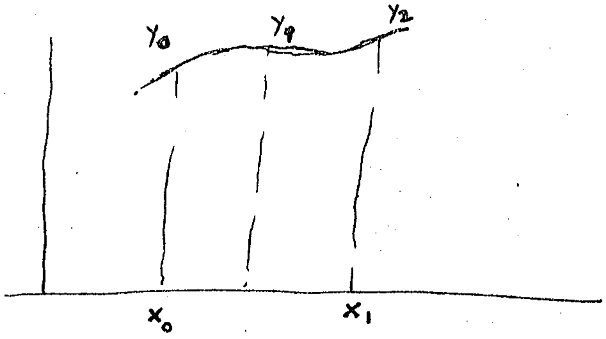
Pick a_i's α_i's & β_{ij}'s in such a manner to meet some $O(h^p)$

The runge kutta technique

assume $y_{i+1} = y_i + \sum a_j k_j$ such that it matches

~~such that~~ $y_{i+1} = y_i + y_i' \Delta x + y_i'' \frac{\Delta x^2}{2}$

$y' = f(x,y)$ $\qquad y_{i+1} = y_i + \int f(x,y)\,dx$



Using $\int f(x,y)\,dx$ as being simpson's rule $^{*}\ \frac{1}{6}\left[k_0 + 4k_1 + k_2\right]$

† * $k_0 = hf(x_0, y_0)$

† * $k_1 = hf(x_0 + \frac{\Delta x}{2},\ y_0 + \frac{k_0}{2})$ $\qquad \Delta y = k_0 \Delta x$

$k_3 = hf(x_0 + \Delta x,\ y_0 + k_0)$

$k_2 = hf(x_0 + \Delta x\ \ y_0 + k_3)$

~~this is $O(\Delta x^3)$~~ this is of local error $O(\Delta x^4)$ $\qquad$ global $O(\Delta x^3)$

if however we defined * $k_2 = hf\{x_0 + \Delta x,\ y_0\ (k_0 - 2k_1)\}$ Wolford original edition

then the results are of local error $O(\Delta x^4)$. $\qquad$ global $O(\Delta x^3)$

Using † $\frac{1}{6}\left[k_0 + 2k_1 + 2k_2 + k_3\right]$

† $k_2 = hf(x_0 + \frac{\Delta x}{2},\ y_0 + \frac{k_1}{2})$

† $k_3 = hf(x_0 + \Delta x,\ y_0 + k_2)$ $\qquad$ local error $O(\Delta x^5)$

global error $O(\Delta x^4)$

To determine accuracy, we may divide stepsize in half & redo calculations to the same x_n

If results are negligible continue to march ; otherwise reduce stepsize in half & redo calculations

1) expensive in computational effort
2) not a good idea to constantly change stepsize

Runge-Kutta-Fehlberg (RKF) provides w/ same computations $O(h^5)$ & $O(h^6)$ methods & also the error estimate as well.

$$k_1 = hf(x_n, y_n)$$

Ch 6 | p 344 7th Ed.

$$k_2 = h\,f(x_n + h/4, y_n + k_1/4)$$

$$k_3 = h\,f(x_n + 3h/8,\ y_n + \tfrac{3}{32}(k_1 + 3k_2))$$

$$k_4 = h\,f(x_n + 12h/13,\ y_n + \tfrac{1}{2197}(1932k_1 - 7200k_2 + 7296k_3))$$

$$k_5 = hf(x_n + h, y_n + \frac{439k_1}{216} - 8k_2 + 3680k_3/513 - 845k_4/4104)$$

$$k_6 = h\,f(x_n + h/2,\ y_n - 8k_1/27 + 2k_2 - 3544k_3/2565 + 1859k_4/4104 - 11k_5/40)$$

$$\hat{y}_{n+1} = y_n + (25k_1/216 + 1408k_3/2565 + 2197k_4/4104 - k_5/5) \qquad \text{local err } O(h^5)$$

$$y_{n+1} = y_n + (16k_1/135 + 6656k_3/12825 + 28561k_4/56430 - 9k_5/50 + 2k_6/55) \qquad \text{local err } O(h^6)$$

$$E = k_1/360 - 128k_3/4275 - 2197k_4/75240 + k_5/50 + 2k_6/55$$

We compare y_{n+1} using two different orders RK formulas instead of halving stepsize that use same k's

Also we get error estimate at same time

Intro to Numerical Computations by Yakowitz & Szidarovszky

*7.7
ADAPTIVE STEP-SIZE SELECTION AND ERROR CONTROL

Up to this point we have not discussed how the step size h of the preceding methods is to be chosen. Obviously, there is a trade-off to be made: If the step size is too small, then computer time is needlessly wasted and accumulation of arithmetic roundoff errors can become a hazard. A large step size invites large truncation error associated with higher-order terms neglected in the construction of the methods. For simplicity, our developments will be concerned only with Runge–Kutta rules.

Techniques for automatic step-size selection are based on estimating the local error at each step and then choosing the step size to keep this estimated error within some tolerance bound. Thus step-size selection hinges on estimation of the *local error*, which at the jth step is defined to be

$$\hat{y}(x_{j+1}) - y_{j+1}.$$

Here y_{j+1} is, of course, the computed approximation of $y(x_{j+1})$, and $\hat{y}(x_{j+1})$ we define to be the exact value at x_{j+1} of the differential equation solution that passes through the point (x_j, y_j). That is, $\hat{y}(x)$ solves the initial-value problem

$$\hat{y}' = f(x, \hat{y}), \qquad \hat{y}(x_j) = y_j.$$

In contrast to local errors, the *global error* at x_{j-1} is defined to be

$$y(x_{j-1}) - y_{j-1},$$

where $y(x)$ is the exact solution of the original initial-value problem (7.3). Figure 7.7 illustrates the relationships between $y(x)$, $\hat{y}(x)$, and local and global errors. Intuitively, the local error is the additional truncation error arising from inexact solution at a given step. The global error gives the accumulated total error propagating from the entire sequence of steps.

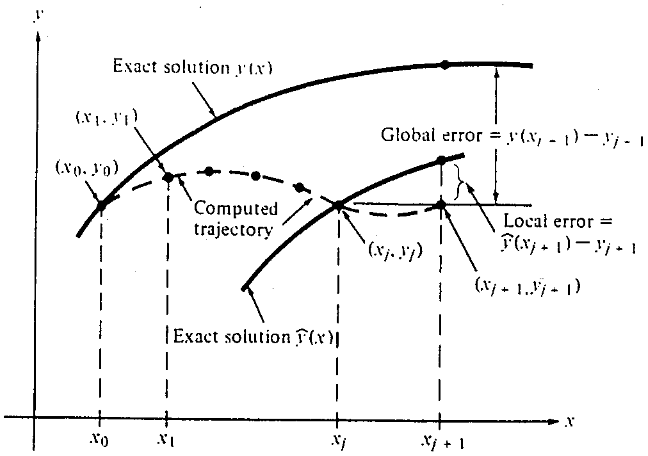


FIGURE 7–7 Relationship Between $y(x)$, $\hat{y}(x)$, Local and Global Errors

Assume that some Runge–Kutta procedure has been selected. We let $y_0, y_1, y_2, \ldots$ denote the computed solution values at the arguments $x_0, x_1, x_2, \ldots$. The local error estimation techniques at each stage apply a higher-order technique to compute an additional approximation, say z_{j+1}, of $y(x_{j+1})$. Since a higher-order technique is used, if the solution is "well behaved" and the step size h is small enough that neglected terms really are negligible, then one may anticipate that the local error of the higher-order method is much less than that of the selected Runge–Kutta procedure. That is,

$$|\hat{y}(x_{j+1}) - z_{j+1}| << |\hat{y}(x_{j+1}) - y_{j+1}|. \tag{7.51}$$

If the approximation above indeed holds, then

$$z_{j+1} - y_{j+1} \simeq \hat{y}(x_{j+1}) - y_{j+1}. \tag{7.52}$$

and we take $z_{j+1} - y_{j+1}$ as the estimate of local error.

Of course, computation of z_{j+1} is typically more expensive than that of y_{j+1} itself, since z_{j+1} must be more accurate. Here, as in other walks of life, information must be paid for. A popular idea toward making this expense as small as possible has been offered by Fehlberg (1964). For a given order, say $p + 1$, the corresponding member of the Fehlberg family computes z_{j+1} with a minimum number of function calls, according to the limitations in Table 7.10, and then provides the pth-order estimate y_{j+1} without any additional function calls. A particularly popular Fehlberg rule is given in Table 7.21, which gives a fifth-order estimate z_{j+1} for a fourth-order rule y_{j+1}.

Subroutine RKF (Table 7.22) implements a single step of this Runge–Kutta–Fehlberg formula, outputting y_{j+1} and z_{j+1} as the parameters YOUT and ZOUT. In view of (7.52), the difference of these values provides a local error estimate. Subroutine ARUKU (Table 7.23) utilizes RKF to update the step size as the computation progresses. If the absolute value of ZOUT-YOUT is less than the user-specified value TOL (for tolerance), the value ZOUT is accepted for y_{j+1}, and a larger step size (by a factor of 3) is chosen for the next step. Otherwise, h is reduced by a factor of 10, and the computation is repeated from the same condition x_j and y_j. Strictly speaking, YOUT, rather than ZOUT, should be chosen for y_{j+1}, but since in principle the higher-order estimate ZOUT should be more accurate, and since it is available, we adopt the pragmatic viewpoint that it should be used. The reader will note that AKUKU is an obvious modification of subroutine ASIMP for adaptive quadrature (Section 3.8.2).

TABLE 7.21 Runge–Kutta–Fehlberg Formula

$$k_1 = f(x_j, y_j)$$

$$k_2 = f\left(x_j + \frac{1}{4}h,\, y_j + \frac{1}{4}hk_1\right)$$

$$k_3 = f\left(x_j + \frac{3}{8}h,\, y_j + h\left(\frac{3}{32}k_1 + \frac{9}{32}k_2\right)\right)$$

$$k_4 = f\left(x_j + \frac{12}{13}h,\, y_j + h\left(\frac{1932}{2197}k_1 - \frac{7200}{2197}k_2 + \frac{7296}{2197}k_3\right)\right)$$

$$k_5 = f\left(x_j + h,\, y_j + h\left(\frac{439}{216}k_1 - 8k_2 + \frac{3680}{513}k_3 - \frac{845}{4104}k_4\right)\right)$$

$$k_6 = f\left(x_j + \frac{1}{2}h,\, y_j + h\left(-\frac{8}{27}k_1 + 2k_2 - \frac{3544}{2565}k_3 + \frac{1859}{4104}k_4 - \frac{11}{40}k_5\right)\right)$$

$$y_{j+1} = y_j + h\left(\frac{25}{216}k_1 + \frac{1408}{2565}k_3 + \frac{2197}{4104}k_4 - \frac{1}{5}k_5\right)$$

$$z_{j+1} = y_j + h\left(\frac{16}{135}k_1 + \frac{6656}{12825}k_3 + \frac{28561}{56430}k_4 - \frac{9}{50}k_5 + \frac{2}{55}k_6\right)$$

TABLE 7.22 Subroutine RKF for the Runge–Kutta–Fehlberg Formula

```
      SUBROUTINE RKF(XI,YI,H,YOUT,ZOUT)
C
C  ************************************************************
C  *  FUNCTION: A CALL TO THIS SUBROUTINE COMPUTES ONE STEP OF*
C  *            THE SOLUTION AND A GUESS OF THE ERROR FOR A   *
C  *            DIFFERENTIAL EQUATION Y'=F(X,Y) WITH INITIAL  *
C  *            VALUES XI, YI. THIS SOLUTION IS OBTAINED      *
C  *            USING A 4-TH ORDER RUNGE-KUTTA FEHLBERG STEP  *
C  *            METHOD IMBEDDED IN A 5-TH ORDER STEP SOLUTION *
C  *  USAGE:                                                  *
C  *       CALL SEQUENCE: CALL RKF(XI,YI,H,YOUT,ZOUT)         *
C  *       EXTERNAL FUNCTIONS/SUBROUTINES: FUNCTION F(U,V)    *
C  *  PARAMETERS:                                             *
C  *       INPUT:                                             *
C  *          XI=INDEPENDENT VARIABLE INITIAL VALUE           *
C  *          YI=DEPENDENT VARIABLE INITIAL VALUE             *
C  *           H=INTERVAL STEP SIZE                           *
C  *      OUTPUT:                                             *
C  *         YOUT=4-TH ORDER SOLUTION ESTIMATE                *
C  *         ZOUT=5-TH ORDER SOLUTION ESTIMATE                *
C  *              (ZOUT-YOUT=LOCAL ERROR ESTIMATE)            *
C  ************************************************************
C
      REAL K1,K2,K3,K4,K5,K6
      K1=F(XI,YI)
      U=XI+0.25*H
      V=YI+0.25*H*K1
      K2=F(U,V)
      U=XI+(3./8.)*H
      V=YI+H*((3./32.)*K1+(9./32.)*K2)
      K3=F(U,V)
      U=XI+H*(12./13.)
      V=YI+(H/2197.)*(1932.*K1-7200.*K2+7296.*K3)
      K4=F(U,V)
      U=XI+H
      V=YI+H*((439./216.)*K1-8.*K2+(3680./513.)*K3-
     1(845./4104.)*K4)
      K5=F(U,V)
      U=XI+0.5*H
      V=-(8./27.)*K1+2.*K2-(3544./2565.)*K3+
     1(1859./4104.)*K4-(11./40.)*K5
      V=YI+H*V
      K6=F(U,V)
      YOUT=(25./216.)*K1+(1408./2565.)*K3+
     1(2197./4104.)*K4-K5/5.
      YOUT=YI+H*YOUT
      ZOUT=(16./135.)*K1+(6656./12825.)*K3+
     1(28561./56430.)*K4-(9./50.)*K5
      ZOUT=ZOUT+(2./55.)*K6
      ZOUT=YI+H*ZOUT
      RETURN
      END
```

TABLE 7.23 Subroutine AKUKU for the Adaptive Runge–Kutta Method

```
      SUBROUTINE ARUKU(X,Y,B,M,TOL)
C
C   ************************************************************
C   *  FUNCTION: THIS SUBROUTINE COMPUTES THE SOLUTION OF A     *
C   *            DIFFERENTIAL EQUATION BY ADAPTIVELY CHOOSING   *
C   *            THE STEP SIZE TO LIMIT THE LOCAL ERROR EST-    *
C   *            IMATE WITHIN A GIVEN TOLERANCE. A 4-TH ORDER   *
C   *            RUNGE-KUTTA-FEHLBERG METHOD IS USED            *
C   *  USAGE:                                                   *
C   *       CALL SEQUENCE: CALL ARUKU(X,Y,B,M,TOL)              *
C   *       EXTERNAL FUNCTIONS/SUBROUTINES:                     *
C   *                        SUBROUTINE RKF(XI,YI,H,YOUT,ZOUT)  *
C   *  PARAMETERS:                                              *
C   *       INPUT:                                              *
C   *        X(1)=INDEPENDENT VARIABLE INITIAL VALUE            *
C   *        Y(1)=DEPENDENT VARIABLE INITIAL VALUE              *
C   *            B=SOLUTION INTERVAL ENDPOINT (LAST X VALUE)    *
C   *            M=MAXIMUM NUMBER OF ITERATIONS                 *
C   *      OUTPUT:                                              *
C   *            X=M BY 1 ARRAY OF INDEPENDENT VARIABLE VALUES  *
C   *            Y=M BY 1 ARRAY OF DEPENDENT VARIABLE SOLUTION  *
C   *              VALUES                                       *
C   ************************************************************
C
      DIMENSION X(M),Y(M)
C     *** INITIALIZATION ***
      H=.10E-02
      I=1
      N=0
C     *** COMPUTE SOLUTION ITERATIVELY ***
      DO WHILE(X(I).LE.B)
         N=N+1
         CALL RKF(X(I),Y(I),H,YOUT,ZOUT)
         ERR=ZOUT-YOUT
C     *** TEST IF THE NUMBER OF ITERATIONS EXCEEDED ***
         IF(N.GT.M) THEN
            WRITE(6,1)
    1       FORMAT(1X,'PROGRAM STOPPED TOO MANY ITERATIONS')
            STOP
         END IF
C     *** TEST STEP SIZE ***
         IF(ABS(ERR).LT.TOL) THEN
            I=I+1
            X(I)=X(I-1)+H
            H=3.0*H
            Y(I)=ZOUT
         ELSE
            H=H/10.0
         END IF
      END DO
      M=I
      H=B-X(I-1)
      X(M)=X(I-1)+H
      CALL RKF(X(I-1),Y(I-1),H,Y(M),ZOUT)
      RETURN
      END
```

We serve notice that the code ARUKU is intended only to illustrate the principles of automatic error control. It is inefficient and does not have the safeguards of a professional differential equation program package. More will be said about this matter after the following computational example.

EXAMPLE 7.17

By means of the calling program given in Table 7.24, the automatic step-size routine ARUKU is called on to solve the differential equation

$$y' = y, \qquad y(0) = 1 \tag{7.53}$$

over the interval [0, 1]. We chose this over our "usual" differential equation because in the present case it is easy to compute the exact local error $\hat{y}(x_{j+1}) - y_{j+1}$ and thereby see how well the RKF error estimator is doing. Specifically, the solution of (7.53) that passes through points (x_j, y_j) is

$$\hat{y}(x) = y_j \exp (x - x_j),$$

and if h is the current step size, then the exact local error is given by

$$y_j \exp (h) - \text{YOUT}.$$

TABLE 7.24 Calling Program for the Subroutine ARUKU

```
C     PROGRAM RKFMETH
C
C     ******************************************************************
C     THIS PROGRAM WILL SET UP AND SOLVE NUMERICALLY THE DIFFERENTIAL
C     EQUATION Y'=Y WITH THE INITIAL CONDITION Y(0.)=1.
C     THE SOLUTION IS OBTAINED USING THE AUTOMATIC STEPSIZE ROUTINE
C     USING 4-TH ORDER RUNGE-KUTTA FEHLBERG METHOD
C     CALLS:   ARUKU,RKF (BOTH MODIFIED FOR DOUBLE PRECISION)
C     OUTPUT:
C              X(I)=VALUE OF X FOR I=1,...  (MAX=50)
C              Y(I)=APPROXIMATED VALUE OF Y AT X(I)
C     ******************************************************************
C
      IMPLICIT DOUBLE PRECISION(A-H,O-Z)
      DIMENSION X(50),Y(50)
C
C     *** FIRST, THE INITIAL CONDITION AND ENDPOINT ARE ESTABLISHED***
C     *** THE MAXIMUM NUMBER OF ITERATIONS IS SET TO 50              ***
C
      X(1)=0.D0
      Y(1)=1.D0
      B=1.D0
      MAX=50
      TOL=1.D-4
C
C     *** SUBROUTINE ARUKU WILL APPROXIMATE THE SOLUTION             ***
C     *** RETURNING AT MOST 50 VALUES IN ARRAYS X AND Y              ***
C
      CALL ARUKU(X,Y,B,MAX,TOL)
      WRITE(10,*)(X(I),Y(I),I=1,50)
      STOP
      END
C
C     *** FUNCTION F SPECIFIES THE DIFFERENTIAL EQUATION. IT IS      ***
C     *** CALLED BY SUBROUTINE RKF                                   ***
C
      FUNCTION F(X,Y)
      IMPLICIT DOUBLE PRECISION(A-H,O-Z)
      F=Y
      RETURN
      END
```

Intro to Numerical Computations by Yakowitz & Szidarovszky

*7.7
ADAPTIVE STEP-SIZE SELECTION AND ERROR CONTROL

Up to this point we have not discussed how the step size h of the preceding methods is to be chosen. Obviously, there is a trade-off to be made: If the step size is too small, then computer time is needlessly wasted and accumulation of arithmetic roundoff errors can become a hazard. A large step size invites large truncation error associated with higher-order terms neglected in the construction of the methods. For simplicity, our developments will be concerned only with Runge–Kutta rules.

Techniques for automatic step-size selection are based on estimating the local error at each step and then choosing the step size to keep this estimated error within some tolerance bound. Thus step-size selection hinges on estimation of the *local error*, which at the jth step is defined to be

$$\hat{y}(x_{j+1}) - y_{j+1}.$$

Here y_{j+1} is, of course, the computed approximation of $y(x_{j+1})$, and $\hat{y}(x_{j+1})$ we define to be the exact value at x_{j+1} of the differential equation solution that passes through the point (x_j, y_j). That is, $\hat{y}(x)$ solves the initial-value problem

$$\hat{y}' = f(x, \hat{y}), \qquad \hat{y}(x_j) = y_j.$$

In contrast to local errors, the *global error* at x_{j-1} is defined to be

$$y(x_{j-1}) - y_{j-1},$$

where $y(x)$ is the exact solution of the original initial-value problem (7.3). Figure 7.7 illustrates the relationships between $y(x)$, $\hat{y}(x)$, and local and global errors. Intuitively, the local error is the additional truncation error arising from inexact solution at a given step. The global error gives the accumulated total error propagating from the entire sequence of steps.

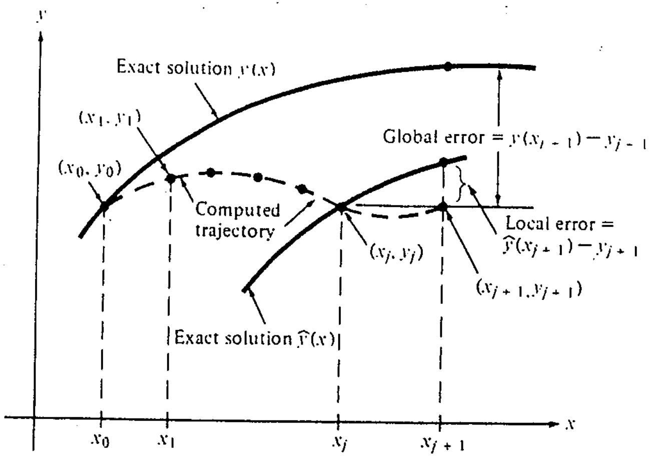


FIGURE 7–7 Relationship Between $y(x)$, $\hat{y}(x)$, Local and Global Errors

Assume that some Runge–Kutta procedure has been selected. We let $y_0, y_1, y_2, \ldots$ denote the computed solution values at the arguments $x_0, x_1, x_2, \ldots$. The local error estimation techniques at each stage apply a higher-order technique to compute an additional approximation, say z_{j+1}, of $y(x_{j+1})$. Since a higher-order technique is used, if the solution is "well behaved" and the step size h is small enough that neglected terms really are negligible, then one may anticipate that the local error of the higher-order method is much less than that of the selected Runge–Kutta procedure. That is,

$$|\hat{y}(x_{j+1}) - z_{j+1}| << |\hat{y}(x_{j+1}) - y_{j+1}|. \tag{7.51}$$

If the approximation above indeed holds, then

$$z_{j+1} - y_{j+1} \simeq \hat{y}(x_{j+1}) - y_{j+1}. \tag{7.52}$$

and we take $z_{j+1} - y_{j+1}$ as the estimate of local error.

Of course, computation of z_{j+1} is typically more expensive than that of y_{j+1} itself, since z_{j+1} must be more accurate. Here, as in other walks of life, information must be paid for. A popular idea toward making this expense as small as possible has been offered by Fehlberg (1964). For a given order, say $p + 1$, the corresponding member of the Fehlberg family computes z_{j+1} with a minimum number of function calls, according to the limitations in Table 7.10, and then provides the pth-order estimate y_{j+1} without any additional function calls. A particularly popular Fehlberg rule is given in Table 7.21, which gives a fifth-order estimate z_{j+1} for a fourth-order rule y_{j+1}.

Subroutine RKF (Table 7.22) implements a single step of this Runge–Kutta–Fehlberg formula, outputting y_{j+1} and z_{j+1} as the parameters YOUT and ZOUT. In view of (7.52), the difference of these values provides a local error estimate. Subroutine ARUKU (Table 7.23) utilizes RKF to update the step size as the computation progresses. If the absolute value of ZOUT-YOUT is less than the user-specified value TOL (for tolerance), the value ZOUT is accepted for y_{j+1}, and a larger step size (by a factor of 3) is chosen for the next step. Otherwise, h is reduced by a factor of 10, and the computation is repeated from the same condition x_j and y_j. Strictly speaking, YOUT, rather than ZOUT, should be chosen for y_{j+1}, but since in principle the higher-order estimate ZOUT should be more accurate, and since it is available, we adopt the pragmatic viewpoint that it should be used. The reader will note that AKUKU is an obvious modification of subroutine ASIMP for adaptive quadrature (Section 3.8.2).

TABLE 7.21 Runge–Kutta–Fehlberg Formula

$$k_1 = f(x_j, y_j)$$

$$k_2 = f\left(x_j + \frac{1}{4}h,\ y_j + \frac{1}{4}hk_1\right)$$

$$k_3 = f\left(x_j + \frac{3}{8}h,\ y_j + h\left(\frac{3}{32}k_1 + \frac{9}{32}k_2\right)\right)$$

$$k_4 = f\left(x_j + \frac{12}{13}h,\ y_j + h\left(\frac{1932}{2197}k_1 - \frac{7200}{2197}k_2 + \frac{7296}{2197}k_3\right)\right)$$

$$k_5 = f\left(x_j + h,\ y_j + h\left(\frac{439}{216}k_1 - 8k_2 + \frac{3680}{513}k_3 - \frac{845}{4104}k_4\right)\right)$$

$$k_6 = f\left(x_j + \frac{1}{2}h,\ y_j + h\left(-\frac{8}{27}k_1 + 2k_2 - \frac{3544}{2565}k_3 + \frac{1859}{4104}k_4 - \frac{11}{40}k_5\right)\right)$$

$$y_{j+1} = y_j + h\left(\frac{25}{216}k_1 + \frac{1408}{2565}k_3 + \frac{2197}{4104}k_4 - \frac{1}{5}k_5\right)$$

$$z_{j+1} = y_j + h\left(\frac{16}{135}k_1 + \frac{6656}{12825}k_3 + \frac{28561}{56430}k_4 - \frac{9}{50}k_5 + \frac{2}{55}k_6\right)$$

TABLE 7.22 Subroutine RKF for the Runge–Kutta–Fehlberg Formula

```
      SUBROUTINE RKF(XI,YI,H,YOUT,ZOUT)
C
C  ***********************************************************
C  *   FUNCTION: A CALL TO THIS SUBROUTINE COMPUTES ONE STEP OF*
C  *             THE SOLUTION AND A GUESS OF THE ERROR FOR A   *
C  *             DIFFERENTIAL EQUATION Y'=F(X,Y) WITH INITIAL  *
C  *             VALUES XI, YI. THIS SOLUTION IS OBTAINED      *
C  *             USING A 4-TH ORDER RUNGE-KUTTA FEHLBERG STEP  *
C  *             METHOD IMBEDDED IN A 5-TH ORDER STEP SOLUTION *
C  *   USAGE:                                                  *
C  *        CALL SEQUENCE: CALL RKF(XI,YI,H,YOUT,ZOUT)         *
C  *        EXTERNAL FUNCTIONS/SUBROUTINES: FUNCTION F(U,V)    *
C  *   PARAMETERS:                                             *
C  *        INPUT:                                             *
C  *           XI=INDEPENDENT VARIABLE INITIAL VALUE           *
C  *           YI=DEPENDENT VARIABLE INITIAL VALUE             *
C  *            H=INTERVAL STEP SIZE                           *
C  *       OUTPUT:                                             *
C  *         YOUT=4-TH ORDER SOLUTION ESTIMATE                 *
C  *         ZOUT=5-TH ORDER SOLUTION ESTIMATE                 *
C  *              (ZOUT-YOUT=LOCAL ERROR ESTIMATE)             *
C  ***********************************************************
C
      REAL K1,K2,K3,K4,K5,K6
      K1=F(XI,YI)
      U=XI+0.25*H
      V=YI+0.25*H*K1
      K2=F(U,V)
      U=XI+(3./8.)*H
      V=YI+H*((3./32.)*K1+(9./32.)*K2)
      K3=F(U,V)
      U=XI+H*(12./13.)
      V=YI+(H/2197.)*(1932.*K1-7200.*K2+7296.*K3)
      K4=F(U,V)
      U=XI+H
      V=YI+H*((439./216.)*K1-8.*K2+(3680./513.)*K3-
     1(845./4104.)*K4)
      K5=F(U,V)
      U=XI+0.5*H
      V=-(8./27.)*K1+2.*K2-(3544./2565.)*K3+
     1(1859./4104.)*K4-(11./40.)*K5
      V=YI+H*V
      K6=F(U,V)
      YOUT=(25./216.)*K1+(1408./2565.)*K3+
     1(2197./4104.)*K4-K5/5.
      YOUT=YI+H*YOUT
      ZOUT=(16./135.)*K1+(6656./12825.)*K3+
     1(28561./56430.)*K4-(9./50.)*K5
      ZOUT=ZOUT+(2./55.)*K6
      ZOUT=YI+H*ZOUT
      RETURN
      END
```

TABLE 7.23 Subroutine AKUKU for the Adaptive Runge–Kutta Method

```
      SUBROUTINE ARUKU(X,Y,B,M,TOL)
C
C   *********************************************************
C   * FUNCTION: THIS SUBROUTINE COMPUTES THE SOLUTION OF A    *
C   *           DIFFERENTIAL EQUATION BY ADAPTIVELY CHOOSING  *
C   *           THE STEP SIZE TO LIMIT THE LOCAL ERROR EST-   *
C   *           IMATE WITHIN A GIVEN TOLERANCE. A 4-TH ORDER  *
C   *           RUNGE-KUTTA-FEHLBERG METHOD IS USED           *
C   * USAGE:                                                  *
C   *      CALL SEQUENCE: CALL ARUKU(X,Y,B,M,TOL)             *
C   *      EXTERNAL FUNCTIONS/SUBROUTINES:                    *
C   *                     SUBROUTINE RKF(XI,YI,H,YOUT,ZOUT)   *
C   * PARAMETERS:                                             *
C   *      INPUT:                                             *
C   *       X(1)-INDEPENDENT VARIABLE INITIAL VALUE           *
C   *       Y(1)-DEPENDENT VARIABLE INITIAL VALUE             *
C   *          B-SOLUTION INTERVAL ENDPOINT (LAST X VALUE)    *
C   *          M-MAXIMUM NUMBER OF ITERATIONS                 *
C   *     OUTPUT:                                             *
C   *          X-M BY 1 ARRAY OF INDEPENDENT VARIABLE VALUES  *
C   *          Y-M BY 1 ARRAY OF DEPENDENT VARIABLE SOLUTION  *
C   *            VALUES                                       *
C   *********************************************************
C
      DIMENSION X(M),Y(M)
C     *** INITIALIZATION ***
      H=.10E-02
      I=1
      N=0
C     *** COMPUTE SOLUTION ITERATIVELY ***
      DO WHILE(X(I).LE.B)
         N=N+1
         CALL RKF(X(I),Y(I),H,YOUT,ZOUT)
         ERR=ZOUT-YOUT
C     *** TEST IF THE NUMBER OF ITERATIONS EXCEEDED ***
         IF(N.GT.M) THEN
            WRITE(6,1)
    1       FORMAT(1X,'PROGRAM STOPPED TOO MANY ITERATIONS')
            STOP
         END IF
C     *** TEST STEP SIZE ***
         IF(ABS(ERR).LT.TOL) THEN
            I=I+1
            X(I)=X(I-1)+H
            H=3.0*H
            Y(I)=ZOUT
         ELSE
            H=H/10.0
         END IF
      END DO
      M=I
      H=B-X(I-1)
      X(M)=X(I-1)+H
      CALL RKF(X(I-1),Y(I-1),H,Y(M),ZOUT)
      RETURN
      END
```

We serve notice that the code ARUKU is intended only to illustrate the principles of automatic error control. It is inefficient and does not have the safeguards of a professional differential equation program package. More will be said about this matter after the following computational example.

EXAMPLE 7.17

By means of the calling program given in Table 7.24, the automatic step-size routine ARUKU is called on to solve the differential equation

$$y' = y, \qquad y(0) = 1 \tag{7.53}$$

over the interval [0, 1]. We chose this over our "usual" differential equation because in the present case it is easy to compute the exact local error $\hat{y}(x_{j+1}) - y_{j+1}$ and thereby see how well the RKF error estimator is doing. Specifically, the solution of (7.53) that passes through points (x_j, y_j) is

$$\hat{y}(x) = y_j \exp(x - x_j),$$

and if h is the current step size, then the exact local error is given by

$$y_j \exp(h) - \text{YOUT}.$$

TABLE 7.24 Calling Program for the Subroutine ARUKU

```
C      PROGRAM RKFMETH
C
C      ***********************************************************
C      THIS PROGRAM WILL SET UP AND SOLVE NUMERICALLY THE DIFFERENTIAL
C      EQUATION Y'=Y WITH THE INITIAL CONDITION Y(0.)=1.
C      THE SOLUTION IS OBTAINED USING THE AUTOMATIC STEPSIZE ROUTINE
C      USING 4-TH ORDER RUNGE-KUTTA FEHLBERG METHOD
C      CALLS:   ARUKU,RKF (BOTH MODIFIED FOR DOUBLE PRECISION)
C      OUTPUT:
C               X(I)=VALUE OF X FOR I=1,...  (MAX=50)
C               Y(I)=APPROXIMATED VALUE OF Y AT X(I)
C      ***********************************************************
C
       IMPLICIT DOUBLE PRECISION(A-H,O-Z)
       DIMENSION X(50),Y(50)
C
C      *** FIRST, THE INITIAL CONDITION AND ENDPOINT ARE ESTABLISHED***
C      *** THE MAXIMUM NUMBER OF ITERATIONS IS SET TO 50            ***
C
       X(1)=0.DO
       Y(1)=1.DO
       B=1.DO
       MAX=50
       TOL=1.D-4
C
C      *** SUBROUTINE ARUKU WILL APPROXIMATE THE SOLUTION           ***
C      *** RETURNING AT MOST 50 VALUES IN ARRAYS X AND Y            ***
C
       CALL ARUKU(X,Y,B,MAX,TOL)
       WRITE(10,*)(X(I),Y(I),I=1,50)
       STOP
       END
C
C      *** FUNCTION F SPECIFIES THE DIFFERENTIAL EQUATION. IT IS    ***
C      *** CALLED BY SUBROUTINE RKF                                 ***
C
       FUNCTION F(X,Y)
       IMPLICIT DOUBLE PRECISION(A-H,O-Z)
       F=Y
       RETURN
       END
```

Multi-Step Methods

Take advantage of previous info.

- construct polynomial that approximates the deriv function & extrapolate it to the next interval
- most methods are based on equal spacing

start w/ $dy/dx = f(x,y)$ & $\int$ between x_n, x_{n+1}

see next page **Adams Method** $y_{n+1} - y_n = \int_{x_n}^{x_{n+1}} f(x,y)\,dx$ let $f(x) = ax^2+bx+c$

Newton – Gregory $f(x_*, y_*) = f_n + s\left[f_n - f_{n-1}\right] + \frac{(s+1)s}{2}\left[f_n - 2f_{n-1} + f_{n-2}\right]$

backward 2nd degree poly through 3 pts f_n, f_{n-1}, f_{n-2} $s = \frac{x - x_n}{h}$

this gives global $O(h^3)$ error &

$$y_{n+1} = y_n + \frac{h}{12}\left[23 f_n - 16 f_{n-1} + 5 f_{n-2}\right] + O(h^4)$$

better results if N-G 4 pt poly $f_n, f_{n-1}, f_{n-2}, f_{n-3}$ assume $f = ax^3+bx^2+cx+d$

where $f_n + s(f_n - f_{n-1}) + \frac{(s+1)s}{2}\left[f_n - 2f_{n-1} + f_{n-3}\right] + \frac{(s+2)s(s+1)}{6}\left[f_n - 3f_{n-1} + 3f_{n-2} + f_{n-3}\right.$

$$y_{n+1} = y_n + \frac{h}{24}\left[55 f_n - 59 f_{n-1} + 37 f_{n-2} - 9 f_{n-3}\right] + O(h^5)$$

Milne Method predicts y_{n+1} by extrapolating values for the deriv. Different than Adams in that it integrates over several intervals.

$$\therefore \int_{x_{n-3}}^{x_{n+1}} \frac{dy}{dx}\,dx = \int_{x_{n-3}}^{x_{n+1}} f(x,y)\,dx$$

1) knowing y at $x_n, x_{n-1}, x_{n-2}, x_{n-3}$ equispaced use a quadratic eqn for $f(x,y)$

2) using $x_n\ x_{n-1}\ x_{n-2}$ and get interpolating poly on $f(x,y)$ then integrate we can get $\underset{\text{Predictor}}{y_{n+1} - y_{n-3}} = \frac{4h}{3}\left(2f_n - f_{n-1} + 2f_{n-2}\right) + O(h^5)$

using this y_{n+1_p} in $f(x,y)$

then

$$\int_{x_{n-1}}^{x_{n+1}} \frac{dy}{dx}\,dx \quad \text{gives} \quad y_{n+1\ \text{corrected}} - y_{n-1} = \frac{h}{3}\left(\overset{f(x_{n+1}, y_{n+1_p})}{f_{n+1}} + 4f_n + f_{n-1}\right) + O(h^5)$$

4) here $f(x,y)$ is a quadratic poly using x_{n+1}, x_n, x_{n-1}

accurate since error of $(y_{n+1\,c} - y_{n+1\,p})$ = local error $= \frac{29}{90}\, y^{V}(\xi)\cdot h^5$

$\Rightarrow$ It is subject to instability if $\int_{n-p}^{n+1}$ m is odd & m=p $h\int_{n-p}^{n+1}\sum_{k=0}^{m} (-1)^k \binom{-s}{k} \Delta^k f_{n-k}\, ds$ since, mth difference vanishes

Table 6.1 ADAMS FORMULAS

Order of formula	Coefficient of h	Coefficients of						Local truncation error E_T
		f_i	f_{i-1}	f_{i-2}	f_{i-3}	f_{i-4}	f_{i-5}	
1	1	1						$\frac{1}{2}h^2y''(\xi)$
2	$\frac{1}{2}$	3	−1					$\frac{5}{12}h^3y'''(\xi)$
3	$\frac{1}{12}$	23	−16	5				$\frac{3}{8}h^4y''''(\xi)$
4	$\frac{1}{24}$	55	−59	37	−9			$\frac{251}{720}h^5y^{\text{V}}(\xi)$
5	$\frac{1}{720}$	1901	−2774	2616	−1274	251		$\frac{475}{1440}h^6y^{\text{VI}}(\xi)$
6	$\frac{1}{1440}$	4277	−7923	9982	−7298	2877	−475	$\frac{19087}{60480}h^7y^{\text{VII}}(\xi)$

represents most of the local truncation error. From the definition of the third backward difference given in Section C.2 of Appendix C we can write

$$h(\tfrac{3}{8}\nabla^3 f_i) = \tfrac{3}{8}h(f_i - 3f_{i-1} + 3f_{i-2} - f_{i-3}). \tag{6.114}$$

Backward finite-difference expressions for derivatives are given by Equations (5.123). From the third of Equations (5.123) we deduce that

$$y''''_i = \frac{y'_i - 3y'_{i-1} + 3y'_{i-2} - y'_{i-3}}{h^3} + O(h)$$

or

$$h^3y''''_i = y'_i - 3y'_{i-1} + 3y'_{i-2} - y'_{i-3} + h^3[O(h)]. \tag{6.115}$$

From Equation (6.115) we determine that

$$(f_i - 3f_{i-1} + 3f_{i-2} - f_{i-3}) = h^3y''''_i - O(h^4). \tag{6.116}$$

Combining Equations (6.116) and (6.114),

$$h(\tfrac{3}{8}\nabla^3 f_i) = \tfrac{3}{8}h^4y''''_i - O(h^5) = \tfrac{3}{8}h^4y''''(\xi)$$

where the fourth derivative of y is evaluated at some unknown x value ξ in the range of x values spanned by the one-step application of the third-order Adams formula. Thus, the local truncation error of the third-order method is

$$E_T = \tfrac{3}{8}h^4y''''(\xi) \tag{6.117}$$

as shown in Table 6.1. In similar fashion we can show that

$$h(\tfrac{1}{2}\nabla f_i) = \tfrac{1}{2}h^2y''(\xi)$$

$$h(\tfrac{5}{12}\nabla^2 f_i) = \tfrac{5}{12}h^3y'''(\xi)$$

$$\vdots$$

for the other local truncation errors shown in Table 6.1.

Carrying out the integration indicated, the general formula is

$$y_{i+1} = y_i + h(f_{i+1} - \tfrac{1}{2}\nabla f_{i+1} - \tfrac{1}{12}\nabla^2 f_{i+1} - \tfrac{1}{24}\nabla^3 f_{i+1} - \tfrac{19}{720}\nabla^4 f_{i+1} - \tfrac{27}{1440}\nabla^5 f_{i+1} - \tfrac{863}{60{,}480}\nabla^6 f_{i+1} - \cdots). \tag{6.137}$$

To obtain the third-order Adams–Moulton corrector formula from this general formula, Equation (6.137) is truncated to three terms following y_i, which yields

$$y_{i+1} = y_i + h(-f_{i+1} - \tfrac{1}{2}\nabla f_{i+1} - \tfrac{1}{12}\nabla^2 f_{i+1}).$$

Then substituting the backward differences as given in Section C.2 in Appendix C into the above, we find

$$y_{i+1} = y_i + \frac{h}{12}(-f_{i-1} + 8f_i + 5f_{i+1}) \tag{6.138}$$

which is identical with Equation (6.129) derived previously.

The fourth-order Adams–Moulton corrector formula is found by truncating Equation (6.137) to four terms following y_i, yielding

$$y_{i+1} = y_i + h(f_{i+1} - \tfrac{1}{2}\nabla f_{i+1} - \tfrac{1}{12}\nabla^2 f_{i+1} - \tfrac{1}{24}\nabla^3 f_{i+1}).$$

Substituting the backward differences from Section C.2 of Appendix C into the above gives the corrector equation

$$y_{i+1} = y_i + \frac{h}{24}(f_{i-2} - 5f_{i-1} + 19f_i + 9f_{i+1}) \tag{6.139}$$

which is identical with Equation (6.134) found previously.

Using the terms in Equation (6.137), we can obtain Adams–Moulton corrector formulas of orders 1 through 7, or of orders 1 through 6 plus a local truncation error expression for each of the six. These expressions are given in Table 6.2. In the local truncation-error expressions ξ is an unknown x value in the range of x values spanned in the one-step application of a particular Adams–Moulton formula. While explicit numerical values cannot be found for these expressions because the derivatives in them cannot be evaluated, they are useful for comparing the truncation errors of the various Adams–Moulton formulas, and for comparing the truncation errors of these formulas with those of other

Table 6.2 ADAMS–MOULTON FORMULAS AND LOCAL TRUNCATION-ERROR EXPRESSIONS

Order of formula	Coefficient of h	Coefficients of f_{i+1}	f_i	f_{i-1}	f_{i-2}	f_{i-3}	f_{i-4}	Local truncation error E_T
1	1	1						$-\frac{1}{2}h^2y''(\xi)$
2	$\frac{1}{2}$	1	1					$-\frac{1}{12}h^3y'''(\xi)$
3	$\frac{1}{12}$	5	8	−1				$-\frac{1}{24}h^4y''''(\xi)$
4	$\frac{1}{24}$	9	19	−5	1			$-\frac{19}{720}h^5y^{\mathrm{V}}(\xi)$
5	$\frac{1}{720}$	251	646	−264	106	−19		$-\frac{27}{1440}h^6y^{\mathrm{VI}}(\xi)$
6	$\frac{1}{1440}$	475	1427	−798	482	−173	27	$-\frac{863}{60480}h^7y^{\mathrm{VII}}(\xi)$

Adams-Moulton does not have instability problems of milne, almost as efficient

1) take a cubic polynom through x_{n-3} x_{n-2} x_{n-1}, x_n & integrate from x_n to x_{n+1}

$$\int_{x_n}^{x_{n+1}} y'\,dx = \int f(x,y)\,dx \qquad \text{let } f = ax^3+bx^2+cx+d$$

where $f(x,y)$ is same as milne yields

$$y_{n+1_p} - y_n = h\left[f_n + \frac{1}{2}(f_n - f_{n-1}) + \frac{5}{12}(f_n - 2f_{n-1} + f_{n-2}) + \frac{3}{8}(f_n - 3f_{n-1} + 3f_{n-2} - f_{n-3})\right]$$

using y_{n+1_p} in $f(x,y)$ to get f_{n+1}

now use cubic poly to approx $f(x,y)$ using $x_{n-2}, x_{n-1}, x_n, x_{n+1}$
& integrate from x_n to x_{n+1}

$$y_{n+1_c} - y_n = h\left(f_{n+1} - \frac{1}{2}[f_{n+1} - f_n] - \frac{1}{12}[f_{n+1} - 2f_n + f_{n-1}] - \frac{1}{24}[f_{n+1} - 3f_n + 3f_{n-1} - f_{n-2}]\right) + O(h^5)$$

$$\Rightarrow \quad y_{n+1_p} = y_n + \frac{h}{24}(55f_n - 59f_{n-1} + 37f_{n-2} - 9f_{n-3})$$

$$y_{n+1_c} = y_n + \frac{h}{24}(9f_{n+1} + 19f_n - 5f_{n-1} + f_{n-2}) \quad \text{where } f_{n+1} = f(x_{n+1}, y_{n+1_p})$$

$(y_c - y_p)\cdot 10^N < 14.2$ if N is the no. of decimal places we want

Adams-Moulton & Milne are more efficient per step than RK & RKF (2 fn eval) vs 4 & 6 with same error locally.

However Multistep are difficult to adapt to change in step size.

Carrying out the integration indicated, the general formula is

$$y_{i+1} = y_i + h(f_{i+1} - \tfrac{1}{2}\nabla f_{i+1} - \tfrac{1}{12}\nabla^2 f_{i+1} - \tfrac{1}{24}\nabla^3 f_{i+1} - \tfrac{19}{720}\nabla^4 f_{i+1} - \tfrac{27}{1440}\nabla^5 f_{i+1} - \tfrac{863}{60{,}480}\nabla^6 f_{i+1} - \cdots). \tag{6.137}$$

To obtain the third-order Adams–Moulton corrector formula from this general formula, Equation (6.137) is truncated to three terms following y_i, which yields

$$y_{i+1} = y_i + h(-f_{i+1} - \tfrac{1}{2}\nabla f_{i+1} - \tfrac{1}{12}\nabla^2 f_{i+1}).$$

Then substituting the backward differences as given in Section C.2 in Appendix C into the above, we find

$$y_{i+1} = y_i + \frac{h}{12}(-f_{i-1} + 8f_i + 5f_{i+1}) \tag{6.138}$$

which is identical with Equation (6.129) derived previously.

The fourth-order Adams–Moulton corrector formula is found by truncating Equation (6.137) to four terms following y_i, yielding

$$y_{i+1} = y_i + h(f_{i+1} - \tfrac{1}{2}\nabla f_{i+1} - \tfrac{1}{12}\nabla^2 f_{i+1} - \tfrac{1}{24}\nabla^3 f_{i+1}).$$

Substituting the backward differences from Section C.2 of Appendix C into the above gives the corrector equation

$$y_{i+1} = y_i + \frac{h}{24}(f_{i-2} - 5f_{i-1} + 19f_i + 9f_{i+1}) \tag{6.139}$$

which is identical with Equation (6.134) found previously.

Using the terms in Equation (6.137), we can obtain Adams–Moulton corrector formulas of orders 1 through 7, or of orders 1 through 6 plus a local truncation error expression for each of the six. These expressions are given in Table 6.2. In the local truncation-error expressions ξ is an unknown x value in the range of x values spanned in the one-step application of a particular Adams–Moulton formula. While explicit numerical values cannot be found for these expressions because the derivatives in them cannot be evaluated, they are useful for comparing the truncation errors of the various Adams–Moulton formulas, and for comparing the truncation errors of these formulas with those of other

Table 6.2 ADAMS–MOULTON FORMULAS AND LOCAL TRUNCATION-ERROR EXPRESSIONS

Order of formula	Coefficient of h	Coefficients of						Local truncation error E_T
		f_{i+1}	f_i	f_{i-1}	f_{i-2}	f_{i-3}	f_{i-4}	
1	1	1						$-\frac{1}{2}h^2y''(\xi)$
2	$\frac{1}{2}$	1	1					$-\frac{1}{12}h^3y'''(\xi)$
3	$\frac{1}{12}$	5	8	−1				$-\frac{1}{24}h^4y''''(\xi)$
4	$\frac{1}{24}$	9	19	−5	1			$-\frac{19}{720}h^5y^{\text{V}}(\xi)$
5	$\frac{1}{720}$	251	646	−264	106	−19		$-\frac{27}{1440}h^6y^{\text{VI}}(\xi)$
6	$\frac{1}{1440}$	475	1427	−798	482	−173	27	$-\frac{863}{60480}h^7y^{\text{VII}}(\xi)$

EGM 6422

LECTURE 9/10

MULTISTEP METHODS

$$y' = f(x,y)$$

ADAMS METHOD

PASS A QUAD. THRU $x_i, f_i = f(x_i, y_i)$
x_{i-1}, f_{i-1}
x_{i-2}, f_{i-2}

$$f(x,y) = f_i + s(f_i - f_{i-1}) + \frac{(s+1)s}{2}(f_i - 2f_{i-1} + f_{i-2})$$

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

$$\int_{x_i}^{x_{i+1}} \frac{dy}{dx}\,dx = \int_{x_i}^{x_{i+1}} f(x,y)\,dx$$

$$y_{i+1} = y_i + \frac{\Delta x}{12}\left[23 f_i - 16 f_{i-1} + 5 f_{i-2}\right] + O(\Delta x^4$$

G.E. $O(\Delta x^3)$

NEWTON-GREGORY

CUBIC (x_i, f_i) (x_{i-1}, f_{i-1}) (x_{i-2}, f_{i-2})

(x_{i-3}, f_{i-3})

$$f(x,y) = f_i + s(f_i - f_{i-1}) + \frac{(s+1)s}{2}(f_i - 2f_{i-1} + f_{i-2})$$

$$+ \frac{(s+2)(s+1)s}{6}(f_i - 3f_{i-1} + 3f_{i-2} - f_{i-3})$$

$$y_{i+1} = y_i + \frac{\Delta x}{24}\left[55f_i - 59f_{i-1} + 37f_{i-2} - 9f_{i-3}\right]$$

$$+ \text{L.E. } O(\Delta x^5)$$

MILNE'S METHOD

$$\int_{y_{i-3}}^{y_{i+1}} \frac{dy}{dx}\,dx = \int f(\bar{x},y)\,d\bar{x}$$

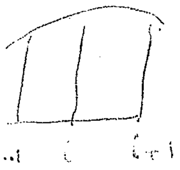
DEFINE $F(x,y)$ using $i, i-1, i-2, i-3$

use this formula & integrate over $i, i-1, i-2, {}^{i-3}$

so that $y_{i+1_p} = y_{i-3} + \frac{4\Delta x}{3}(2f_i - f_{i-1} + 2f_{i-2}]$

now define $f_{i+1} = f(x_{i+1}, y_{i+1p})$

$$\int_{x_{i-1}}^{x_{i+1}} \frac{dy}{dx}\,dx = \int_{x_{i-1}}^{x_{i+1}} f \; dx$$



let f be quadratic using i+1, i, i-1

SIMPSON'S RULE

$$y_{i+1_c} = y_{i-1} + \frac{\Delta x}{3}\left(f_{i+1} + 4f_i + f_{i-1}\right)$$

ADAMS MOULTON METHOD

$y' = f(x,y)$

CUBIC THRU $i, i-1, i-2, i-3$

& INTEGRATES IT OVER i TO $i+1$

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

$$\Longrightarrow y_{i+1_p}$$

DFN CUBIC over $i+1, i, i-1, i-2$ & INTEGRATE OVER i TO $i+1$

$$y_{i+1_p} = y_i + \frac{\Delta x}{24}\left(55f_i - 59f_{i-1} + 37f_{i-2} - 9f_{i-3}\right)$$

$$y_{i+1_c} = y_i + \frac{\Delta x}{24}\left(9f_{i+1} + 19f_i - 5f_{i-1} + f_{i-2}\right)$$

GE $O(\Delta x^4)$

Table 6.1 ADAMS FORMULAS

Order of formula	Coefficient of h	Coefficients of f_i	f_{i-1}	f_{i-2}	f_{i-3}	f_{i-4}	f_{i-5}	Local truncation error E_T
1	1	1						$\frac{1}{2}h^2y''(\xi)$
2	$\frac{1}{2}$	3	-1					$\frac{5}{12}h^3y'''(\xi)$
3	$\frac{1}{12}$	23	-16	5				$\frac{3}{8}h^4y''''(\xi)$
4	$\frac{1}{24}$	55	-59	37	-9			$\frac{251}{720}h^5y^{\text{V}}(\xi)$
5	$\frac{1}{720}$	1901	-2774	2616	-1274	251		$\frac{475}{1440}h^6y^{\text{VI}}(\xi)$
6	$\frac{1}{1440}$	4277	-7923	9982	-7298	2877	-475	$\frac{19087}{60480}h^7y^{\text{VII}}(\xi)$

represents most of the local truncation error. From the definition of the third backward difference given in Section C.2 of Appendix C we can write

$$h(\tfrac{3}{8}\nabla^3 f_i) = \tfrac{3}{8}h(f_i - 3f_{i-1} + 3f_{i-2} - f_{i-3}). \tag{6.114}$$

Backward finite-difference expressions for derivatives are given by Equations (5.123). From the third of Equations (5.123) we deduce that

$$y''''_i = \frac{y'_i - 3y'_{i-1} + 3y'_{i-2} - y'_{i-3}}{h^3} + O(h)$$

or

$$h^3y''''_i = y'_i - 3y'_{i-1} + 3y'_{i-2} - y'_{i-3} + h^3[O(h)]. \tag{6.115}$$

From Equation (6.115) we determine that

$$(f_i - 3f_{i-1} + 3f_{i-2} - f_{i-3}) = h^3y''''_i - O(h^4). \tag{6.116}$$

Combining Equations (6.116) and (6.114),

$$h(\tfrac{3}{8}\nabla^3 f_i) = \tfrac{3}{8}h^4y''''_i - O(h^5) = \tfrac{3}{8}h^4y''''(\xi)$$

where the fourth derivative of y is evaluated at some unknown x value ξ in the range of x values spanned by the one-step application of the third-order Adams formula. Thus, the local truncation error of the third-order method is

$$E_T = \tfrac{3}{8}h^4y''''(\xi) \tag{6.117}$$

as shown in Table 6.1. In similar fashion we can show that

$$h(\tfrac{1}{2}\nabla f_i) = \tfrac{1}{2}h^2y''(\xi)$$

$$h(\tfrac{5}{12}\nabla^2 f_i) = \tfrac{5}{12}h^3y'''(\xi)$$

$$\vdots$$

for the other local truncation errors shown in Table 6.1.

nomial to ap-
$_{i+1}$ instead of
own values of
y determining
ying Equation

ard integration
obtained most
or x_i. The x_i's
the u's, where

f_i in Equation

(6.135)

n as

of Equation
tor equation

(6.136)

Carrying out the integration indicated, the general formula is

$$y_{i+1} = y_i + h(f_{i+1} - \tfrac{1}{2}\nabla f_{i+1} - \tfrac{1}{12}\nabla^2 f_{i+1} - \tfrac{1}{24}\nabla^3 f_{i+1} - \tfrac{19}{720}\nabla^4 f_{i+1}$$
$$- \tfrac{27}{1440}\nabla^5 f_{i+1} - \tfrac{863}{60{,}480}\nabla^6 f_{i+1} - \cdots). \tag{6.137}$$

To obtain the third-order Adams–Moulton corrector formula from this general formula, Equation (6.137) is truncated to three terms following y_i, which yields

$$y_{i+1} = y_i + h(-f_{i+1} - \tfrac{1}{2}\nabla f_{i+1} - \tfrac{1}{12}\nabla^2 f_{i+1}).$$

Then substituting the backward differences as given in Section C.2 in Appendix C into the above, we find

$$y_{i+1} = y_i + \frac{h}{12}(-f_{i-1} + 8f_i + 5f_{i+1}) \tag{6.138}$$

which is identical with Equation (6.129) derived previously.

The fourth-order Adams–Moulton corrector formula is found by truncating Equation (6.137) to four terms following y_i, yielding

$$y_{i+1} = y_i + h(f_{i+1} - \tfrac{1}{2}\nabla f_{i+1} - \tfrac{1}{12}\nabla^2 f_{i+1} - \tfrac{1}{24}\nabla^3 f_{i+1}).$$

Substituting the backward differences from Section C.2 of Appendix C into the above gives the corrector equation

$$y_{i+1} = y_i + \frac{h}{24}(f_{i-2} - 5f_{i-1} + 19f_i + 9f_{i+1}) \tag{6.139}$$

which is identical with Equation (6.134) found previously.

Using the terms in Equation (6.137), we can obtain Adams–Moulton corrector formulas of orders 1 through 7, or of orders 1 through 6 plus a local truncation error expression for each of the six. These expressions are given in Table 6.2. In the local truncation-error expressions ξ is an unknown x value in the range of x values spanned in the one-step application of a particular Adams–Moulton formula. While explicit numerical values cannot be found for these expressions because the derivatives in them cannot be evaluated, they are useful for comparing the truncation errors of the various Adams–Moulton formulas, and for comparing the truncation errors of these formulas with those of other

Table 6.2 ADAMS–MOULTON FORMULAS AND LOCAL TRUNCATION-ERROR EXPRESSIONS

Order of formula	Coefficient of h	Coefficients of						Local truncation error E_T
		f_{i+1}	f_i	f_{i-1}	f_{i-2}	f_{i-3}	f_{i-4}	
1	1	1						$-\frac{1}{2}h^2y''(\xi)$
2	$\frac{1}{2}$	1	1					$-\frac{1}{12}h^3y'''(\xi)$
3	$\frac{1}{12}$	5	8	−1				$-\frac{1}{24}h^4y''''(\xi)$
4	$\frac{1}{24}$	9	19	−5	1			$-\frac{19}{720}h^5y^{\mathrm{V}}(\xi)$
5	$\frac{1}{720}$	251	646	−264	106	−19		$-\frac{27}{1440}h^6y^{\mathrm{VI}}(\xi)$
6	$\frac{1}{1440}$	475	1427	−798	482	−173	27	$-\frac{863}{60480}h^7y^{\mathrm{VII}}(\xi)$

~~Under some conditions numerical methods can lead to instabilities~~

~~example~~ But for $y' + 5y = 5$ $\quad y = Ce^{-5x} + 1$

~~Suppose~~

∴ we introduce an extra solution. if $C_2 = 0$ No problem

but round off & inexact starting values will cause problems.

∴ error between true & diff $= C_2 e^{5x_i} (-1)^i$

will dominate as $i \to \infty$ further the error will oscillate in sign.

However runge-kutta types do not exhibit numerical instabilities when step size is small. Multistep methods can be unstable

- To determine find roots of characteristic eqn: one root tends to sol. if the extraneous roots satisfy $|\beta_i| < 1$ STRONGLY STABLE

1) ~~Show that~~ Find the general solution ~~of~~ to $y_{n+2} + 4hy_{n+1} - y_n = 2h$

- what differential equation does this represent? $\underline{y' + 2y = 1}$
- is this solution stable

$$y_{i+1} - 2y_i + y_{i-1} = \frac{M}{EI} h^2 = Dh^2$$

the homog $\beta^{i-1}[\beta^2 - 2\beta + 1] = 0$ $\qquad \beta_1 = 1 \qquad \beta_1^n = 1^n$

for a double root, $\qquad \beta_2^n = n\beta_1^n \qquad \beta_2^n = n$

∴ $y_i = c_1 i + c_2$

for particular soln

$Y_i = C \qquad C - 2C + C = 0$

$Y_i = Ci \qquad C(i+1) - 2Ci + C(i-1) = 0$

$Y_i = Ci^2 \qquad C(i+1)^2 - 2C(i)^2 + C(i-1)^2 = \cancel{Ci^2} + \cancel{2Ci} + C - \cancel{2Ci^2} + \cancel{Ci^2} - \cancel{2Ci} + C = 2C = Dh^2$

$Dh^2 = 2c \qquad c = \frac{D}{2}h^2$

$$\therefore \quad y_i = \frac{D}{2}h^2 i^2 + c_1 i + c_2 \qquad\qquad x_i = hi$$
$$\frac{D}{2}x_i^2 + c_1 x_i + c_2 .$$

$$y'' = \frac{M}{EI} \rightarrow \quad y' = \frac{Mx}{EI} + c_1$$
$$y = \frac{Mx^2}{2EI} + c_1 x + c_2 .$$

if roots to char eqn include complex roots (conj) $a \pm bi$

$$\Rightarrow \quad \beta_1 = re^{i\theta} \qquad \beta_2 = re^{-i\theta}$$

$$\bar{c}_1\beta_1^n + \bar{c}_2\beta_2^n = r^n\left[c_1 \cos n\theta + c_2 \sin n\theta\right]$$

$c_1 = \bar{c}_1 + \bar{c}_2$
$c_2 = (\bar{c}_1 - \bar{c}_2)i$
$r = \sqrt{a^2 + b^2}$
$\theta = \tan^{-1}(b/c)$

extrapolation to the limit

suppose we know y_N using h $\qquad y_N = y(b)$ as $h \to 0$

$\hat{y}_N$ using $2h$

and suppose y_N has an error of $O(h^4)$

$$y(b) = y_N + c_1 h^4 + o(h^4)$$
$$y(b) = \hat{y}_N + 16c_1 h^4 + o(h^4)$$

$$0 = y_N - \hat{y}_N - 15c_1 h^4 + o(h^4) \quad \Rightarrow \quad c_1 h^4 = \frac{y_N - \hat{y}_N}{15} + o(h^4)$$

what if we uses $h/2 \Rightarrow c_1(h/2)^4 = \frac{\tilde{y}_N - y_N}{15} + o(h^4$

and $\frac{y_N - \hat{y}_N}{\tilde{y}_N - y_N} = 16$ as a check ; $\Rightarrow y^{(1)}_N(b) = y_N + \frac{y_N - \hat{y}_N}{15}$ is better

solutions to difference equations

when $y' + 5y = 5$ is replaced by $y_{i+1} + 10y_i h - y_{i-1} = 10h$ using $y' = \frac{y_{i+1} - y_{i-1}}{2h}$

this is a difference equation. As $h \to 0$ the difference equation should lead to the solution to the original equation.

In general the solution to homog. linear diff eqn is in the form.

$$y_i = C\beta^i \Rightarrow y_{i+1} + 10y_i h - y_{i-1} = 0 \Rightarrow C\beta^{i-1}\left[\beta^2 + 10h\beta - 1\right] = 0$$

$$\beta = \frac{-10h \pm \sqrt{(10h)^2 - 4(1)(-1)}}{2} = -5h \pm \sqrt{1 + (5h)^2}$$

$$= -5h \pm \left(1 + \tfrac{1}{2}(5h^2) + \dots\right) = \begin{matrix} 1 - 5h \\ -1 - 5h \end{matrix} + O(h^2)$$

$$\therefore \quad y_i = C_1(1-5h)^i + C_2(-1)^i(1+5h)^i \qquad \text{if } x_i = ih$$

$$= C_1\left(1 - \frac{5x_i}{i}\right)^i + C_2(-1)^i\left(1 + \frac{5x_i}{i}\right)^i$$

$$\lim_{i\to\infty}\left(1 + \frac{5x_i}{i}\right)^i = \lim_{i\to\infty}\left(1 + \overset{\epsilon}{\frac{5x_i}{i}}\right)^{\frac{i}{5x_i}\cdot 5x_i} = e^{5x_i} \qquad e = \lim_{\epsilon\to 0}(1+\epsilon)^{1/\epsilon}$$

$$\lim_{i\to\infty}\left(1 - \frac{5x_i}{i}\right)^i = \lim_{i\to\infty}\left(1 - \frac{5x_i}{i}\right)^{-\frac{i}{5x_i}\cdot -5x_i} = e^{-5x_i}$$

$$\therefore \quad y_i = C_1 e^{-5x_i} + C_2 e^{+5x_i}(-1)^i$$

for the particular $y_{i+1} + 10y_i h - y_{i-1} = 10h$ const let $y_i^p = C$

$$C + 10hC - C = 10h \qquad \therefore \; C = 1$$

$\therefore$

$$y_i = C_1 e^{-5x_i} + C_2 e^{+5x_i}(-1)^i + 1$$

special equation $y' = \lambda y$, λ constant, is usually considered sufficient, however, to give an indication of the stability of a method.

We consider first the Adams-Bashforth fourth-order method. If in (8.47) we set $f(x, y) = \lambda y$ we obtain

$$y_{n+1} - y_n - \frac{h\lambda}{24}(55y_n - 59y_{n-1} + 37y_{n-2} - 9y_{n-3}) = 0 \qquad (8.73)$$

The characteristic equation for this difference equation is

$$\beta^4 - \beta^3 - \frac{h\lambda}{24}(55\beta^3 - 59\beta^2 + 37\beta - 9) = 0$$

The roots of this equation are of course functions of $h\lambda$. It is customary to write the characteristic equation in the form

$$\rho(\beta) + h\lambda\sigma(\beta) = 0 \qquad (8.74)$$

where $\rho(\beta)$ and $\sigma(\beta)$ are polynomials defined by

$$\rho(\beta) = \beta^4 - \beta^3$$

$$\sigma(\beta) = -\frac{1}{24}(55\beta^3 - 59\beta^2 + 37\beta - 9)$$

We see that as $h \to 0$, (8.74) reduces to $\rho(\beta) = 0$, whose roots are $\beta_1 = 1$, $\beta_2 = \beta_3 = \beta_4 = 0$. For $h \neq 0$, the general solution of (8.73) will have the form

$$y_n = c_1\beta_1^n + c_2\beta_2^n + c_3\beta_3^n + c_4\beta_4^n$$

where the β_i are solutions of (8.74). It can be shown that β_1^n approaches the desired solution of $y' = \lambda y$ as $h \to 0$ while the other roots correspond to extraneous solutions. Since the roots of (8.74) are continuous functions of h, it follows that for h small enough, $|\beta_i| < 1$ for $i = 2, 3, 4$, and hence from the definition of stability that the Adams-Bashforth method is strongly stable. All multistep methods lead to a characteristic equation in the form (8.74) whose left-hand side is sometimes called the **stability polynomial**. The definition of stability can be recast in terms of the stability polynomial. A method is **strongly stable** if all the roots of $\rho(\beta) = 0$ have magnitude less than one except for the simple root $\beta = 1$.

We investigate next the stability properties of Milne's method (8.64*b*) given by

$$y_{n+1} = y_{n-1} + \frac{h}{3}(f_{n+1} + 4f_n + f_{n-1}) \qquad (8.75)$$

Again setting $f(x, y) = \lambda y$ we obtain

$$y_{n+1} - y_{n-1} - \frac{h\lambda}{3}(y_{n+1} + 4y_n + y_{n-1}) = 0$$

and its characteristic equation becomes

$$\rho(\beta) + h\lambda\sigma(\beta) = 0 \qquad (8.76)$$

with

$$\rho(\beta) = \beta^2 - 1$$

$$\sigma(\beta) = \beta^2 + 4\beta + 1$$

This time $\rho(\beta) = 0$ has the roots $\beta_1 = 1$, $\beta_2 = -1$, and hence by the definition above, Milne's method is not strongly stable. To see the implications of this we compute the roots of the stability polynomial (8.76). For h small we have

$$\beta_1 = 1 + \lambda h + \mathcal{O}(h^2)$$

$$\beta_2 = -(1 - \lambda h/3) + \mathcal{O}(h^2) \qquad (8.77)$$

Hence the general solution of (8.75) is

$$y_n = c_1(1 + \lambda h + \mathcal{O}(h^2))^n + c_2(-1)^n(1 - \lambda h/3 + \mathcal{O}(h^2))^n$$

If we set $n = x_n/h$ and let $h \to 0$, this solution approaches

$$y_n = c_1 e^{\lambda x_n} + c_2(-1)^n e^{-\lambda x_n/3} \qquad (8.78)$$

In this case stability depends upon the sign of λ. If $\lambda > 0$ so that the desired solution is exponentially increasing, it is clear that the extraneous solution will be exponentially decreasing so that Milne's method will be stable. On the other hand if $\lambda < 0$, then Milne's method will be unstable since the extraneous solution will be exponentially increasing and will eventually swamp the desired solution. Methods of this type whose stability depends upon the sign of λ for the test equation $y' = \lambda y$ are said to be **weakly stable**. For the more general equation $y' = f(x, y)$ we can expect weak stability from Milne's method whenever $\partial f/\partial y < 0$ on the interval of integration.

In practice all multistep methods will exhibit some instability for some range of values of the step h. Consider, for example, the Adams-Bashforth method of order 2 defined by

$$y_{n+1} = y_n + \frac{h}{2}\{3f_n - f_{n-1}\}$$

If we apply this method to the test equation $y' = \lambda y$, we will obtain the difference equation

$$y_{n+1} - y_n - \frac{h\lambda}{2}\{3y_n - y_{n-1}\} = 0$$

and from this the stability polynomial

$$\beta^2 - \beta - \frac{h\lambda}{2}\{3\beta - 1\}$$

or the equation

$$\beta^2 - \left(1 + \frac{3h\lambda}{2}\right)\beta + \frac{h\lambda}{2} = 0$$

what about $y' = f(x,y)$

$$\frac{y_{i+1} - y_i}{\Delta x} = f(x_i, y_i)$$

$$y_{i+1} = y_i + \Delta x\, f_i$$

normally we won't solve so we use $y_{i+1} = y_i + \Delta x \cdot \lambda y_i$ where $f_i \cong \lambda y_i$

for Adams - Bashforth $y_{i+1} = y_i + \frac{\Delta x \lambda}{24}\left(55 y_i - 59 y_{i-1} + 37 y_{i-2} - 9 y_{i-3}\right)$

if $y_i = \beta^i$ then $0 = \beta^{i+1} - \beta^i - \frac{\Delta x \lambda}{24}\left(55\beta^i - 59\beta^{i-1} + 37\beta^{i-2} - 9\beta^{i-3}\right)$

1
1 1
1 2 1
1 3 3 1
1 4 6 4 1

or $\underbrace{\beta^4 - \beta^3}_{} - \frac{\Delta x \lambda}{24}\left(55\beta^3 - 59\beta^2 + 37\beta - 9\right) = 0$

$1 + 4\epsilon - (1 + 3\epsilon) - \frac{\Delta x}{24}\lambda\left(55 \cdot 1^{(1+\epsilon)} - 59 \cdot 1^{(1+\epsilon)} + 37 \cdot 1^{(1+\epsilon)} - 9\right)$ $\quad -(24 + 84\epsilon)\frac{\Delta x \lambda}{24} + \epsilon \cong 0$ $\quad \epsilon\left(1 - \frac{84\Delta x\lambda}{24}\right) = 24\frac{\Delta x \lambda}{24}$

$\rho(\beta) + \Delta x \lambda\, \sigma(\beta) = 0$; as $\Delta x \to 0$ this reduces to $\rho(\beta) = 0$

$\therefore\ \beta^3(\beta - 1) = 0$ or $\beta_{1,2,3} = 0$ & $\beta_4 = 1$ $\quad \beta$

$$\therefore\ y_i = c_1 \beta_1^i + c_2 i \beta_2^i + c_3 i^2 \beta_3^i + c_4 \beta_4^i$$

as $h \to 0$ $y_i \to$ solution to $y' = \lambda y$ & other solutions are extraneous

Since the original equat. is a continuous function of Δx then it follows that for small Δx $|\beta_i| < 1$ for $i = 1, 2, 3$ & the method is stable

What about Milne's method $y_{n+1} = y_{n-1} + \frac{h}{3}\left(f_{n+1} + 4 f_n + f_{n-1}\right)$

if $f(x,y) = \lambda y$ then we get

$$(\beta^2 - 1) - \frac{\lambda h}{3}\left(\beta^2 + 4\beta + 1\right) = 0$$

$$\beta = \frac{+\frac{4\lambda h}{3} \pm \sqrt{\left(\frac{4\lambda h}{3}\right)^2 - 4\left(1 + \frac{\lambda h}{3}\right)\left(\frac{\lambda h}{3} - 1\right)}}{2\left(1 - \frac{\lambda h}{3}\right)}$$

KEEP TERMS
UP TO h
ONLY

$\lim_{\epsilon \to 0} (1+\epsilon)^{1/\epsilon} = e$

$\beta_1 = 1 + \lambda h$
$\beta_2 = -\left(1 - \lambda h/3\right)$

and the solution is $y_n \to c_1 e^{\lambda x_n} + c_2 (-1)^n e^{-\lambda x_n/3}$

$(1 + \lambda h)^{\frac{i x_i \lambda}{x_i \lambda}} \Rightarrow e^{\lambda x_i}$

$\frac{\lambda x_i}{i}$

does have the disadvantage of not being self-starting, though, as we shall see later, this is not a serious disadvantage.

The development of Milne's method begins by dividing the area under a given portion of a curve $y = f(x)$ into 4 Δx-width strips, as shown in Fig. 6-27. The true area under this portion of the curve is then approximated by considering the area of these 4 strips under a second-degree parabola having 3 coordinates in common with the actual curve, as indicated by the dashed line in Fig. 6-27. The crosshatched area is the approximate area obtained. To determine an expression for this crosshatched area in terms of Δx and the appropriate y ordinates, it is convenient to consider the 4 strips as centered on the y axis, as shown in Fig. 6-28. This arrangement does not compromise the generality of the results obtained, and it has the advantage of simplifying the intermediat expressions involved in determining the desired form of the expressio for the area of 4 such strips. The crosshatched area of Fig. 6-28 is given b

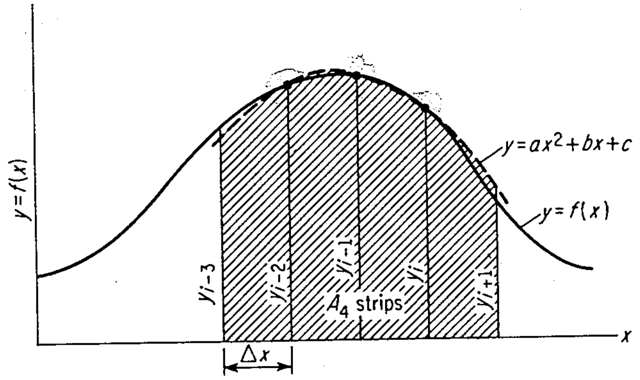


FIG. 6-27. Area in 4 strips under a curve approximated by area under a second-degree parabola.

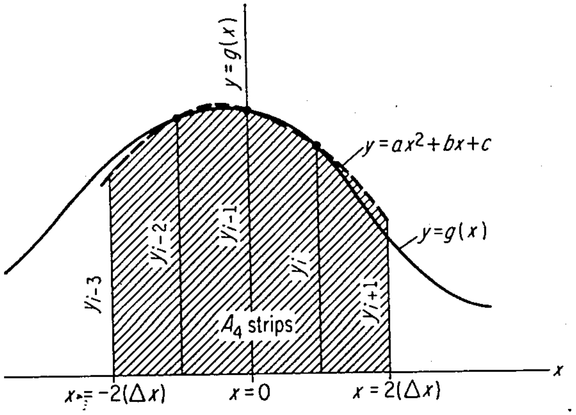


FIG. 6-28. Area of 4 strips centered on y axis.

$$A_{4\,\text{strips}} = \int_{-2(\Delta x)}^{2(\Delta x)} (ax^2 + bx + c)\,dx \qquad \text{(6-83}$$

Integrating Eq. 6-83 and substituting the limits gives

$$A_{4\,\text{strips}} = \frac{16}{3}\,a(\Delta x)^3 + 4c(\Delta x) \qquad \text{(6-84}$$

The constants a and c are determined in the manner explained on p. 286 The appropriate expressions are

$$a = \frac{y_i - 2y_{i-1} + y_{i-2}}{2(\Delta x)^2}$$

$$c = y_{i-1} \qquad \text{(6-85}$$

Substituting Eq. 6-85 into Eq. 6-84 yields, for the area of the 4 strips ir terms of Δx and the y ordinates shown,

$$A_{4\,\text{strips}} = \frac{4}{3}(\Delta x)[2y_i - y_{i-1} + 2y_{i-2}] \qquad \text{(6-86}$$

This expression will be used later as part of the predictor equation.

Let us consider the application of Milne's method in integrating a first-order differential equation of the form

$$y' = f(x, y) \qquad \text{(6-87)}$$

where the value of y is known for $x = 0$. This technique consists, basically, of obtaining approximate values of y by the use of a *predictor* equation and then correcting these values by the iterative use of a corrector equation. Milne's predictor equation

$$P(y_{i+1}) = y_{i-3} + \frac{4}{3}(\Delta x)[2y_i' - y_{i-1}' + 2y_{i-2}'] \qquad \text{(6-88)}$$

utilizes the area of 4 strips under a parabolic approximation of a curve (see Eq. 6-86) to provide a predicted value for the successive y ordinates. Milne's *corrector* equation

$$C(y_{i+1}) = y_{i-1} + \frac{\Delta x}{3}[y_{i-1}' + 4y_i' + P(y_{i+1}')] \qquad \text{(6-89)}$$

provides corrected y values by using Simpson's rule for determining the area of 2 strips under a curve (see p. 286).

Assuming that the resulting y and y' curves of Eq. 6-87 have the general form of the curves shown in Fig. 6-29, the first step is to obtain a predicted value of y_4. Utilizing Eq. 6-88 with $i = 3$,

$$P(y_4) = y_0 + \frac{4}{3}(\Delta x)[2y_3' - y_2' + 2y_1'] \qquad \text{(6-90)}$$

EGN 6422

ADV. ANALYSIS OF
MECH. SYSTEMS

MILNES

$$y_{i+1} = y_{i-1} + \frac{\Delta x}{3}\left(f_{\underline{i+1}} + 4f_i + f_{i-1}\right)$$

$$f(x,y) \cong \lambda y$$

$$y_{i+1} = y_{i-1} + \frac{\lambda \Delta x}{3}\left(y_{i+1} + 4y_i + y_{i-1}\right)$$

$$y_i = \mathcal{C}\beta^i$$

$$\mathcal{C}\beta^{i-1}\left[\left(\beta^2-1\right) - \frac{\lambda\Delta x}{3}\left(\beta^2+4\beta+1\right)\right] = 0$$

$$= \beta^2\left(1-\frac{\lambda\Delta x}{3}\right) - \frac{4\lambda\Delta x}{3}\beta - \left(1+\frac{\lambda\Delta x}{3}\right)$$

$$\beta = \frac{\frac{4\lambda\Delta x}{3} \pm \sqrt{\left(\frac{4\lambda\Delta x}{3}\right)^2 + 4\left(1-\left(\frac{\lambda\Delta x}{3}\right)^2\right)}}{2\left(1-\frac{\lambda\Delta x}{3}\right)}$$

$$\beta_1 \cong 1 + \lambda\Delta x \qquad x_i = i\Delta x$$

$$\beta_2 \cong -\left(1 - \frac{\lambda\Delta x}{3}\right)$$

$$\frac{1}{1-x} = 1 + x + x^2 + \dots$$

$$\sqrt{1-x} \approx 1 - \frac{x}{2}$$

$$y_i = \mathcal{C}_1 e^{\lambda x_i} + \mathcal{C}_2 (-1)^i e^{-\lambda x_i/3}$$

$$y' = f(x,y) \cong \lambda y \qquad y' = \lambda y \qquad y = Ce^{sx} \quad y' = sCe^{sx} \Rightarrow \lambda = s$$

$$y = \mathcal{C}e^{\lambda x}$$

$$\left.\begin{array}{l} \text{if } \lambda > 0 \longrightarrow \text{actual soln} \\ \quad\ \lambda < 0 \longrightarrow \text{diverge} \end{array}\right\} \text{weakly stable}$$

---

if $|\beta_i| < 1 \Longrightarrow$ strongly stable numerical techn

if $|\beta_i| > 1 \Longrightarrow$ weakly stable, possibly unstable

$$\Sigma \underline{F} = m\ddot{y}$$

$$\ddot{y} = f(t, y, \dot{y})$$

NEED 2 CONDITIONS

$y(t=t_0) = y_0$

$\frac{dy}{dt}(t=t_0) = p(t=t_0) = p_0$

LET $\frac{dy}{dt} = \dot{y} = p$

THEN $\ddot{y} = \dot{p} = f(t, y, p) = \frac{dp}{dt}$

$$y_{i+1} = y_i + \int_{t_i}^{t_{i+1}} p\,dt$$

$$p_{i+1} = p_i + \int_{t_i}^{t_{i+1}} f(t, y, p)\,dt$$

$$\boxed{\begin{array}{l} p = p_i \quad @\, t_i \\ f(t_i, y_i, p_i) \end{array}}$$

Runge-Kutta technique

$$y_{i+1} = y_i + \frac{1}{6}\left[q_0 + 2q_1 + 2q_2 + q_3\right]$$

GE $O(\Delta t^4)$

$$p_{i+1} = p_i + \frac{1}{6}\left[k_0 + 2k_1 + 2k_2 + k_3\right]$$

$$\begin{tikzcd}
k_0 = \Delta t\, f(t_i, y_i, p_i) \arrow[r] & q_0 = \Delta t \cdot p_i \arrow[dl] \\
k_1 = \Delta t\, f\left(t_i + \frac{\Delta t}{2}, y_i + \frac{q_0}{2}, p_i + \frac{k_0}{2}\right) \arrow[r] & q_1 = \Delta t\left(p_i + \frac{k_0}{2}\right) \arrow[dl] \\
k_2 = \Delta t\, f\left(t_i + \frac{\Delta t}{2}, y_i + \frac{q_1}{2}, p_i + \frac{k_1}{2}\right) \arrow[r] & q_2 = \Delta t\left(p_i + \frac{k_1}{2}\right) \arrow[dl] \\
k_3 = \Delta t\, f\left(t_i + \frac{\Delta t}{2}, y_i + q_2, p_i + k_2\right) \arrow[r] & q_3 = \Delta t\left(p_i + k_2\right)
\end{tikzcd}$$

EXTRAPOLATION TO THE LIMIT

$\Delta x \qquad y' = f(x,y) \qquad y_0, y_1, y_2, \ldots, y_N$

$2\Delta x \qquad\qquad\qquad\qquad y_0 \quad y_1 \cdots, y_{N/2} = \hat{y}_N$

if we used a numerical technique that has an error of $O(\Delta x^4)$

$$y = y_N + C_1 \Delta x^4 + \overset{o}{\bullet}(\Delta x^4) \qquad \frac{y - (y_N + C_1 \Delta x^4)}{\Delta x^4} = \underset{\uparrow\,\bullet\,(\text{ZERO})}{\underline{0}} + O(\Delta x^5)$$

$$y = \hat{y}_N + C_1 (2\Delta x)^4 + o(\Delta x^4)$$

$$0 = y_N - \hat{y}_N - 15 C_1 \Delta x^4 + o(\Delta x^4)$$

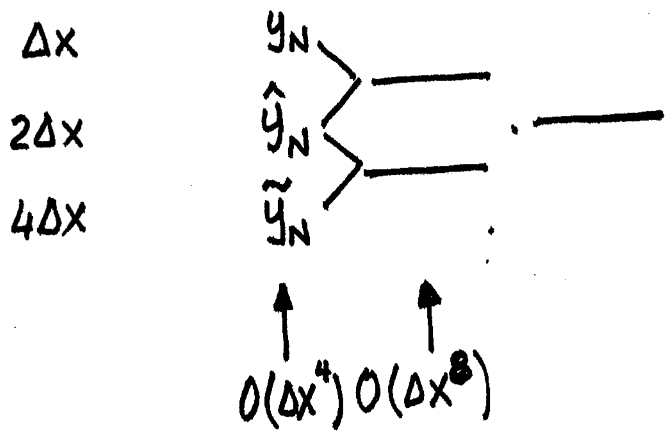
$$C_1 \Delta x^4 = \frac{y_N - \hat{y}_N}{15} + o(\Delta x^4)$$

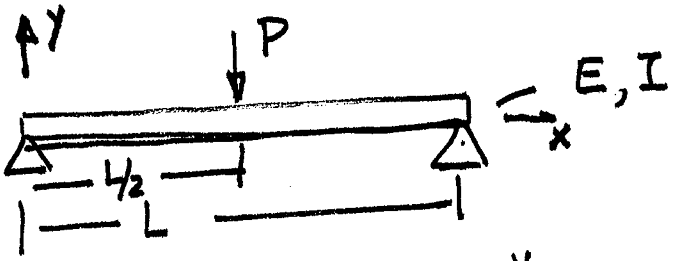
$$y = y_N + \frac{y_N - \hat{y}_N}{15} = \text{m.a.} + \frac{\text{m.a.} - \text{l.a.}}{15 \leftarrow 2^4 - 1}$$

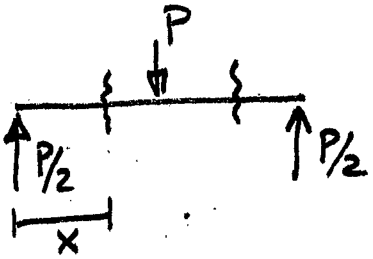
m.a. more accurate

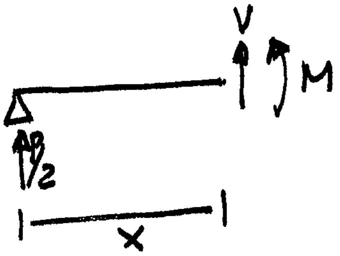
l.a. less accurate

$$\text{M. Euler } O(\Delta x^2) \qquad \text{m.a.} + \frac{\text{m.a.} - \text{l.a.}}{2^2 - 1}$$



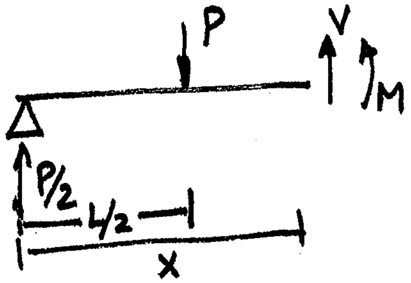






$V = -P/2 \qquad \Sigma F_y = 0$

$M = \frac{P}{2}x \qquad x \leq \frac{L}{2}$

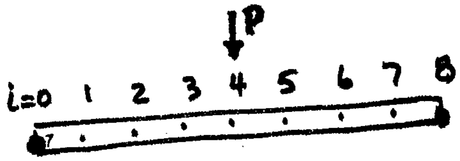


$V = +P/2 \qquad \Sigma F_y = 0$

$M = \frac{P}{2}x - P(x - L/2) \qquad x \geq L/2$

$$EI \frac{d^2w}{dx^2} = M \quad \Rightarrow \quad w_i'' = \frac{M_i}{(EI)_i} \qquad w(x) \qquad x_i = i\Delta x$$

$$\downarrow$$



$\Delta x = \frac{L}{8} \qquad w_0 = 0 \qquad w_8 = 0$

$$\frac{w_{i+1} - 2w_i + w_{i-1}}{\Delta x^2} \cong w_i'' \qquad O(\Delta x^2)$$

$i=1 \qquad \frac{w_2 - 2w_1 + \overset{0}{\cancel{w_0}}}{\Delta x^2} = \left(\frac{M}{EI}\right)_1$

$i=2 \qquad \frac{w_3 - 2w_2 + w_1}{\Delta x^2} = \left(\frac{M}{EI}\right)_2$

$\vdots$

$i=7 \qquad \frac{\overset{0}{\cancel{w_8}} - 2w_7 + w_6}{\Delta x^2} = \left(\frac{M}{EI}\right)_7$

$$\begin{bmatrix} -2 & 1 & 0 & \cdots & & 0 \\ 1 & -2 & 1 & 0 & \cdots & 0 \\ 0 & 1 & -2 & 1 & 0 & 0 \\ 0 & 0 & & \ddots & 0 & 0 \\ 0 & & & 0 & 1 & -2 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ w_3 \\ \vdots \\ w_7 \end{bmatrix} = \begin{bmatrix} (M/EI)_1 \Delta x^2 \\ (M/EI)_2 \Delta x^2 \\ \vdots \\ (M/EI)_7 \Delta x^2 \end{bmatrix}$$

$A\underline{w} = \underline{b}$ *Implicit Numerical Technique*

WHY PDEs? IMPORTANCE?

Physical processes vary with time & location

how will these processes vary with time & space

what drives these processes

where these processes will be in the future & what will
happen at some future location

Processes are described by differential equations

if processes depend on 2 or more independent variables

then they are governed by partial differential equations

Many processes in natural envrions are governed by 2nd order
PDEs

$$\frac{\partial^2 u}{\partial x^2} = \frac{1}{c^2}\frac{\partial^2 u}{\partial t^2}$$ wave eqn.

u - displacement
c - bar velocity = $\sqrt{E/\rho}$

$$\frac{\partial T}{\partial t} = \alpha\left(\frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2}\right)$$ heat eqn in 2-D

T - temp
α - thermal diff.
$\alpha = \frac{k}{c\rho}$

Fick's Law of Diff. $\frac{\partial C_A}{\partial t} = D_{AB}\nabla^2 C_A$

$$\frac{\partial^2 \phi}{\partial x^2} + \frac{\partial^2 \phi}{\partial y^2} = \nabla^2\phi = 0$$ potential flow

$\underline{V} = \nabla\phi$

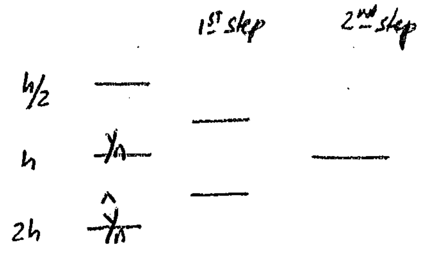
$$\nabla\cdot\underline{V} = 0$$

$$y(b) = y_N + C_1 h^4 + O(h^4)$$
$$y(b) = \hat{y}_N + C_1 \cdot 16 h^4 + O(h^4)$$

$$0 = (y_N - \hat{y}_N) - 15 C_1 h^4 + O(h^4)$$

$$\therefore \quad C_1 = \frac{y_N - \hat{y}_N}{15 h^4} \quad \text{or} \quad C_1 h^4 = \frac{y_N - \hat{y}_N}{15}$$

$$y_B = y_N + \frac{y_N - \hat{y}_N}{15} + O(h^4)$$
$$y_B = \hat{y}_N + \frac{y_N - \hat{y}_N}{15} \cdot 16 + O(h^4)$$



2nd step
$$y_B = y_N + \frac{y_N - \hat{y}_N}{15} + D h^6$$
$$y_B = \hat{y}_N + \frac{y_N - \hat{y}_N}{15} 16 + D h^6 \cdot 64$$

$$\ddot{y} = f(t, y, \dot{y}) \qquad \text{let } \dot{y} = p$$
$$\dot{p} = f(t, y, p) \quad \Longrightarrow \quad p_{i+1} = p_i + \int_{t_i}^{t_{i+1}} f(t, y, p)\,dt$$
$$\dot{y} = p \qquad\qquad y_{i+1} = y_i + \int_{t_i}^{t_{i+1}} p\,dt$$

$$\int p\,dt = \frac{1}{6}\left[q_0 + 2q_1 + 2q_2 + q_3\right]$$
$$\int f\,dt = \frac{1}{6}\left[k_0 + 2k_1 + 2k_2 + k_3\right]$$

x → y
p → v

$$k_0 = \Delta t\, f(t_i, y_i, p_i) \longrightarrow q_0 = \Delta t\, p_i$$
$$k_1 = \Delta t\, f\left(t_i + \frac{\Delta t}{2}, y_i + \frac{q_0}{2}, p_i + \frac{k_0}{2}\right) \qquad q_1 = \Delta t\left[p_i + \frac{k_0}{2}\right)$$
$$k_2 = \Delta t\, f\left(t_i + \frac{\Delta t}{2}, y_i + q_1/2, p_i + \frac{k_1}{2}\right) \qquad q_2 = \Delta t\left(p_i + \frac{k_1}{2}\right)$$
$$k_3 = \Delta t\, f(t_i + \Delta t, y_i + q_2, p_i + k_2 \qquad q_3 = \Delta t\,(p_i + k_2)$$

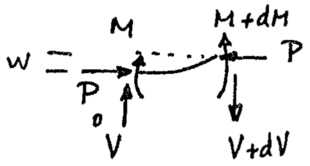
NUMERICAL SIMULATION OF DYNAMIC CRACK BRANCHING AND CRACK CURVING IN FINITE TWO-DIMENSIONAL BODIES

by

Dr. Cesar Levy
Department of Mechanical Engineering
School of Engineering
College of Engineering and Applied Science
(305)554-3439

Math/Sci/Engineering Proposal

Shooting Method



$\Sigma F_x = 0$

$+\downarrow \; \Sigma F_y = V + dV - V = 0$

$+\circlearrowleft \; \Sigma M_0 = + Pw + M + dM - M + (V + dV)\, dx = 0$

$$\therefore \quad Pw + dM - Vdx = 0$$

$$\therefore \quad dM = Vdx - Pw$$

$$M = \int V dx - \int P w' dx$$

$$M = -EIw'' - Pw$$

$\rightarrow EIw'' + Pw = 0$

with $P/_{EI} = \lambda^2$ $\quad w'' + \lambda^2 w = 0$

at ends $\quad w(0) = 0 \quad w(L) = 0$ $\qquad$ using shooting since $w'(0)$ is unknown

$$\frac{EI(w_{i+1} + w_{i-1} - 2w_i)}{\Delta x^2} + Pw_i = 0$$

$$w_{i+1} + (\lambda \Delta x^2 - 2) w_i + w_{i-1} = 0$$

$$\begin{bmatrix} \lambda\Delta x^2 - 2 & 1 & & \\ 1 & (\;) & 1 & \\ & 1 & (\;) & 1 \\ & & \ddots\; 1 & (\;) \end{bmatrix} \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{pmatrix} = \begin{pmatrix} 0 \end{pmatrix}$$

LESSON #4

- PHYSICAL PROCESSES NORMALLY VARY WITH TIME AND LOCATION
- TO UNDERSTAND THESE PROCESSES
  - IT WOULD BE NICE TO KNOW HOW THEY VARY IN TIME & SPACE
  - WHAT DRIVES THESE PROCESSES (HOW PROCESSES DEPEND ON SYSTEM PARAM
  - WHERE THESE PROCESSES WILL BE AT SOME FUTURE TIME OR WHAT WILL HAPPEN AT SOME FUTURE LOCATION
- IT TURNS OUT THAT EQNS THAT DESCRIBE THESE PROCESSES ARE GENERALLY DIFFERENTIAL EQUATIONS
- WHEN THESE PROCESSES DEPEND ON THE VARIATION OF TWO OR MORE QUANTITIES, THEN THESE PROCESSES ARE GOVERNED BY PARTIAL DIFFERENTIAL EQUATIONS
- MANY PROCESSES IN NATURE ARE DESCRIBED BY 2nd ORDER P.D.E.
- EXAMPLES VIBRATIONS OF A ROD. (LONGITUDINAL VIBRATIONS)

$$\frac{\partial^2 u}{\partial x^2} = \frac{1}{c^2}\frac{\partial^2 u}{\partial t^2}$$

$c = \sqrt{E/\rho}$ BAR VELOCITY

u - LONGITUDINAL DISPLACEMENT

HEAT TRANSFER

$$\frac{\partial T}{\partial t} = \alpha\left(\frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2}\right)$$

$\frac{k}{c\rho} = \alpha$ THERMAL DIFFUSIVITY

T TEMPERATURE

k - heat conduction (thermal conductivity

c - heat capacity

ρ - density

fick's law of diffusion $\frac{\partial c_A}{\partial t} = D_{AB}\nabla^2 c_A$

C_A - Concentration of species A

D - diffusivity of A into B

POTENTIAL FLOW

$$\frac{\partial^2 \phi}{\partial x^2} + \frac{\partial^2 \phi}{\partial y^2} = \nabla^2\phi = 0$$

LAPLACE'S EQN

ϕ POTENTIAL $\underset{\sim}{V} = \nabla\phi$

STEADY STATE INCOMPRESSIBLE IRROTATIONAL FLOW CONTINUITY EQN IS $\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} = 0$

$v = \frac{\partial \phi}{\partial y}$ $u = \frac{\partial \phi}{\partial x}$

IN THESE CASES u, T, ϕ FIELD VARIABLE DEPENDENT VARIABLE

x, y, t SPACE COORDINATES OR TIME INDEPENDENT VARIABLES

$$\frac{\partial}{\partial x} u(x,t) = \lim_{\Delta x \to 0} \frac{u(x+\Delta x,t) - u(x,t)}{\Delta x}$$

- THESE THREE EQNS CAN DESCRIBE MANY SYSTEMS IN MECH. ENG.
  - WAVE EQUATION ARISES FROM ACOUSTICS, VIBRATIONS, SHALLOW-WATER WAVE THEORY
  - HEAT EQN ARISES IN HEAT TRANSFER & ONE-DIMENSIONAL DIFFUSION PROBLEMS
  - LAPLACE'S EQN ARISES IN POTENTIAL FLOW THEORY, STEADY STATE HEAT TRANSFER, TORSION OF A BAR, STEADY STATE VIB. OF A MEMBRANE
- DO ALL THREE HAVE ANY COMMON IDEAS ?
- HOW CAN THEY BE SOLVED ? WHAT METHODS EXIST TO SOLVE THE EQNS
- HOW CAN I DERIVE THE MATHEMATICAL EQN ?
- WHAT IS A WELL POSED PROBLEM - CAN I FIND A UNIQUE SOLUTION

- CHARACTERIZATION & CLASSIFICATION
  - MOST GENERAL 2^{nd} ORDER PDE OF A FN $u(x,y)$

$$\varphi(x,y,u,u_x,u_y,u_{xx},u_{xy},u_{yy}) = 0 \qquad u_{xx} = \frac{\partial^2 u}{\partial x^2}$$

  - LINEAR WITH RESPECT TO HIGHEST DERIVATIVE

$$a\,u_{xx} + b\,u_{xy} + c\,u_{yy} + F(x,y,u,u_x,u_y) = 0$$

  - HERE a, b, c are fns of x, y only.
  - if $F(x,y,u,u_x,u_y)$ quasilinear

    if $F(x,y,u,u_x,u_y) = d\,u_x + e\,u_y + fu + g$ LINEAR

    if $g = 0$ THEN IT IS HOMOGENEOUS

LINEAR $2^{\underline{nd}}$ order PDE HOMOGENEOUS

$$a u_{xx} + b u_{yx} + c u_{yy} + d u_x + e u_y + f u = 0 \qquad (1)$$

- IF ALONG SOME CURVE $\varphi(x,y) =$ constant the equation $a\left(\frac{dy}{dx}\right)^2 - b\left(\frac{dy}{dx}\right) + c=$

THEN

$$\frac{dy}{dx} = \frac{b \pm \sqrt{b^2-4ac}}{2a}$$ IS THE SLOPE OF THE CURVE φ IN THE x,y plane

φ is the characteristic

- SO WHAT

IF $b^2-4ac > 0$ $\frac{dy}{dx}$ has 2 values, real & (1) HYPERBOLIC TYPE

$b^2-4ac = 0$ $\frac{dy}{dx}$ has 1 value, real & (1) PARABOLIC TYPE

$b^2-4ac < 0$ $\frac{dy}{dx}$ has 2 complex values & (1) ELLIPTIC TYPE

- CHARACTERIZATION (CLASSIFICATION) OF EQN. IS DETERMINED BY
  $a u_{xx} + b u_{xy} + c u_{yy}$ portion of PDE

- $\frac{\partial^2 u}{\partial x^2} - \frac{1}{c^2}\frac{\partial^2 u}{\partial t^2} = 0$ $a=1$ $b=0$ $c \to \frac{-1}{c^2}$ $d=e=f=0$

$t \to y$

$$\frac{dt}{dx} = \frac{0 \pm \sqrt{0 - 4 \cdot 1 \cdot \frac{-1}{c^2}}}{2 \cdot 1} = \pm\frac{1}{c}$$ HYPERBOLIC

if $c =$ constant $t = \pm \frac{1}{c}x +$ constant $\therefore$ $t + \frac{1}{c}x = C_1 = \varphi_1(t,x)$

$t - \frac{1}{c}x = C_2 = \varphi_2(t,x)$

- $a^2\frac{\partial^2 T}{\partial x^2} - \frac{\partial T}{\partial t} = 0$ $a \to a^2$ $b=0$ $c=0$ $d=0$ $e=-1$ $f=0$

$t \to y$

$$\frac{dt}{dx} = \frac{0 \pm \sqrt{0 - 4a^2 \cdot 0}}{2a^2} = 0$$ PARABOLIC

$t =$ constant $= \varphi_1(t,x)$

- $$\frac{\partial^2 \varphi}{\partial x^2} + \frac{\partial^2 \varphi}{\partial y^2} = 0 \qquad a=1 \quad b=0 \quad c=1 \quad d=e=f=0$$

$$\frac{dy}{dx} = \frac{0 \pm \sqrt{0 - 4.1.1}}{2 \cdot 1} = \pm i \qquad \text{ELLIPTIC}$$

ELLIPTIC EQUATIONS ⟷ LAPLACE TYPE

PARABOLIC " ⟷ HEAT EQUATION

HYPERBOLIC " ⟷ WAVE EQUATION

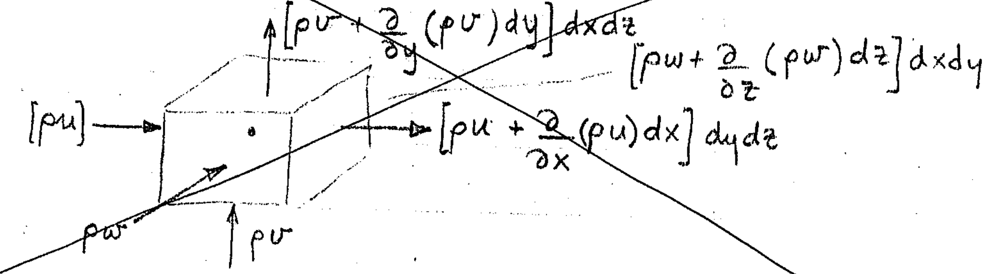
- WHY ARE THESE CURVES $\varphi(x,y)$ important? WHEN WE STUDY HYPERBOLIC EQUATIONS
  - WILL FIND $u = \varphi_1(x+ct) + \varphi_2(x-ct)$
  - $\frac{dt}{dx} = \frac{1}{c} \rightarrow c = \frac{dx}{dt}$ determines speed at which information travels along characteristics
  - $x+ct$, $x-ct$ are characteristic lines along which information travels
  - NOTICE THAT FOR $c < \infty$ INFO TRAVELS AT FINITE SPEEDS
- FOR PARABOLIC EQNS
  - $\frac{dt}{dx} = 0 \Rightarrow \frac{dx}{dt} = \infty \Rightarrow c = \infty$ INFO TRAVELS AT ∞ SPEEDS

—— ∘ —— ∘ —— ∘ —— LESSON # 5

~~• TALK ABOUT ELLIPTIC EQUATIONS FIRST & INVESTIGATE – FLUID DYNAMICS~~

~~LOOK AT AN ELEMENTAL VOLUME WHERE MASS MOVES THROUGH FIXED VOLUME~~

~~FLUID DYNAMICS~~



~~u, v, w are the velocity components of the fluid stream into this fixed volume~~

~~ρ (velocity) · area it crosses = mass flow rate = $\rho \cdot \frac{\text{volume}}{\Delta t} = \frac{\text{mass}}{\Delta t}$~~

EGN 6422

ADV. ANALYSIS OF MECH. SYSTEMS

10/4/05

$$\Sigma \underline{F} = m\ddot{y}$$

$$\ddot{y} = f(t, y, \dot{y})$$

$$\frac{dy}{dt} = \dot{y} = p$$

$$\ddot{y} = \dot{p} = f(t, y, p) = \frac{dp}{dt}$$

$$y_{i+1} = y_i + \int_{t_i}^{t_{i+1}} p\,dt$$

$$p_{i+1} = p_i + (\int_{t_i}^{t_{i+1}} f(t,y,p)\,dt$$

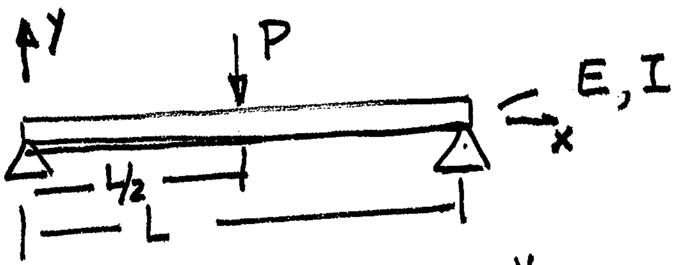
$$\boxed{\begin{array}{l} p = p_i \quad @\,t_i \\ f(t_i, y_i, p_i) \end{array}}$$

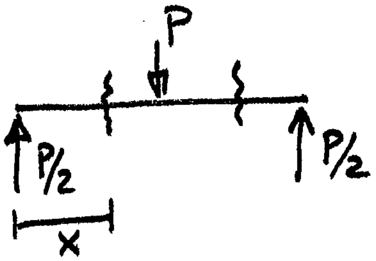
Runge-Kutta technique

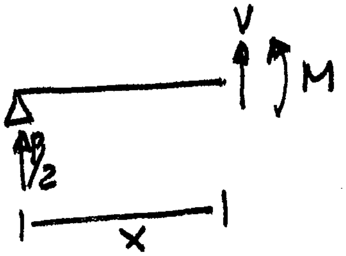
$$\longrightarrow y_{i+1} = y_i + \frac{1}{6}\left[q_0 + 2q_1 + 2q_2 + q_3\right] \qquad \text{GE } O(\Delta t^4)$$

$$\longrightarrow p_{i+1} = p_i + \frac{1}{6}\left[k_0 + 2k_1 + 2k_2 + k_3\right]$$

$$\begin{tikzcd}
k_0 = \Delta t\, f(t_i, y_i, p_i) \arrow[r] & q_0 = \Delta t \cdot p_i \arrow[dl] \\
k_1 = \Delta t\, f(t_i + \frac{\Delta t}{2}, y_i + \frac{q_0}{2}, p_i + \frac{k_0}{2}) \arrow[r] & q_1 = \Delta t\,(p_i + \frac{k_0}{2}) \arrow[dl] \\
k_2 = \Delta t\, f(t_i + \frac{\Delta t}{2}, y_i + \frac{q_1}{2}, p_i + \frac{k_1}{2}) \arrow[r] & q_2 = \Delta t\,(p_i + \frac{k_1}{2}) \arrow[dl] \\
k_3 = \Delta t\, f(t_i + \frac{\Delta t}{2}, y_i + q_2, p_i + k_2) \arrow[r] & q_3 = \Delta t\,(p_i + k_2)
\end{tikzcd}$$

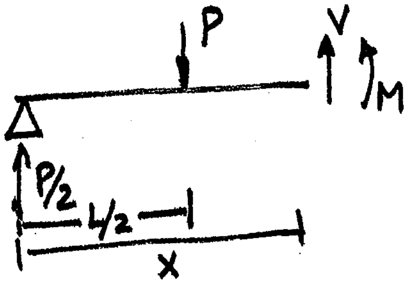






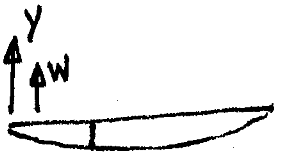
$V = -P/2 \qquad \Sigma F_y = 0$

$M = \frac{P}{2} x$



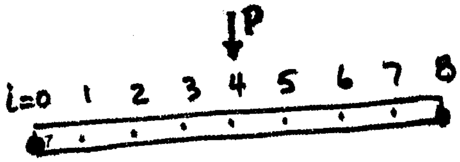
$V = -P/2 \qquad \Sigma F_y = 0$

$M = \frac{P}{2}x - P(x - L/2)$



$$EI \frac{d^2w}{dx^2} = M \implies w_i'' = \frac{M_i}{(EI)_i} \qquad w(x)$$

↓



$\Delta x = \frac{L}{8} \qquad w_0 = 0 \qquad w_8 = 0$

$$\frac{w_{i+1} - 2w_i + w_{i-1}}{\Delta x^2} \cong w_i'' \qquad O(\Delta x^2)$$

$i=1 \qquad \frac{w_2 - 2w_1 + \overset{0}{\cancel{w_0}}}{\Delta x^2} = \left(\frac{M}{EI}\right)_1$

$i=2 \qquad \frac{w_3 - 2w_2 + w_1}{\Delta x^2} = \left(\frac{M}{EI}\right)_2$

$\vdots$

$i=7 \qquad \frac{\overset{0}{\cancel{w_8}} - 2w_7 + w_6}{\Delta x^2} = \left(\frac{M}{EI}\right)_7$

$$\begin{bmatrix} -2 & 1 & 0 & \cdots & \cdots & 0 \\ 1 & -2 & 1 & 0 & \cdots & 0 \\ 0 & 1 & -2 & 1 & 0 & 0 \\ 0 & 0 & & \ddots & 0 & 0 \\ 0 & & & 0 & 1 & -2 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ w_3 \\ \vdots \\ w_7 \end{bmatrix} = \begin{bmatrix} (M/EI)_1 \Delta x^2 \\ (M/EI)_2 \Delta x^2 \\ \vdots \\ (M/EI)_7 \Delta x^2 \end{bmatrix}$$

$A\underline{w} = \underline{b}$ *Implicit Numerical Techniqu*

WHY PDEs ? IMPORTANCE ?

Physical processes vary with time & location

how will these processes vary with time & space

what drives these processes

where these processes will be in the future & what will happen at some future location

Processes are described by differential equations

if processes depend on 2 or more independent variables then they are governed by partial differential equations

Many processes in natural environs are governed by 2nd order PDEs

$$\frac{\partial^2 u}{\partial x^2} = \frac{1}{c^2}\frac{\partial^2 u}{\partial t^2}$$ wave eqn.

u - displacement

c - bar velocity $= \sqrt{E/\rho}$

$$\frac{\partial T}{\partial t} = \alpha\left(\frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2}\right)$$ heat eqn in 2-D

T - temp

α - thermal diff.

$\alpha = \frac{k}{c\rho}$

Fick's Law of Diff. $\frac{\partial C_A}{\partial t} = D_{AB}\nabla^2 C_A$

$$\frac{\partial^2 \phi}{\partial x^2} + \frac{\partial^2 \phi}{\partial y^2} = \nabla^2 \phi = 0$$ potential flow

$\underline{V} = \nabla \phi$

$$\nabla \cdot \underline{V} = 0$$

$$\frac{\partial u(x,y)}{\partial x} = \lim_{\Delta x \to 0} \frac{u(x+\Delta x, y) - u(x,y)}{\Delta x}$$

WAVE EQN - APPEARS IN ACOUSTICS, VIBS, SHALLOW WATER WAVE THEORY

HEAT EQN - " IN HEAT TRANSFER, DIFFUSION PROBLEMS

LAPLACES EQ - " IN POTENTIAL FLOW THEORY, STEADY STATE HEAT TRANS, TORSION OF BAR, STEADY STATE VIB. OF MEMBRANE

WHAT'S COMMON TO ALL
HOW CAN THEY BE SOLVED
WHAT IS A WELL POSED PROBLEM

CHARACTERIZATION & CLASSIFICATION

GENERAL 2nd ORDER PDE

$$\varphi(x, y, u, u_x, u_y, u_{xx}, u_{xy}, u_{yy}) = 0 \qquad u(x,y)$$

$$u_{xx} \equiv \frac{\partial^2 u}{\partial x^2}$$

THE EQ. IS LINEAR WRT ITS HIGHEST DERIVS.

$$\underline{A u_{xx} + B u_{xy} + C u_{yy}} + F(x, y, u, u_x, u_y) = 0$$

$\overline{A}, B, C$ are fns of x & y ~~alone~~ only

when $F(x, y, u, u_x, u_y)$ then equation is quasi linear

IF $F(x, y, u, u_x, u_y) = \underbrace{D u_x + E u_y + F u + G}_{\text{are fns of } x \,\&\, y \text{ only}}$ then equation is linear

IF $G = 0$ then equation homogeneous

2nd order linear PDE

$$A u_{xx} + B u_{xy} + C u_{yy} + D u_x + E u_y + \mathcal{F} u + G = 0$$

IF ALONG SOME CURVE $\varphi(x,y) = \text{const}$ the equation $A(y')^2 - B(y') + C = 0$

THEN

$$\frac{dy}{dx} = y' = \frac{B \pm \sqrt{B^2 - 4AC}}{2A} \quad \text{\& } \varphi \text{ is called the characteristic}$$

$B^2 - 4AC > 0$ PDE is hyperbolic & 2 characteristics exist ; $\frac{dy}{dx}$ is real

$B^2 - 4AC = 0$ PDE is parabolic & 1 real characteristic exists

$B^2 - 4AC < 0$ PDE is elliptic & no real characteristics exist

— - — - — - —

Wave Eqn

$$\frac{\partial^2 u}{\partial x^2} - \frac{1}{c^2}\frac{\partial^2 u}{\partial t^2} = 0 \qquad A = 1 \quad B = 0 \quad C = -\frac{1}{c^2}$$

$$t \to y \qquad \frac{dy}{dx} \to \underset{\substack{\uparrow \\ \text{slowness}}}{\frac{dt}{dx}} = \frac{0 \pm \sqrt{0 - 4 \cdot 1 \cdot \left(-\frac{1}{c^2}\right)}}{2 \cdot 1} = \pm \frac{1}{c}$$

$$\frac{dx}{dt} = \text{velocity} = \frac{1}{\text{slowness}} = \pm c \qquad \text{in magnitude } |c| < \infty$$

$$\begin{aligned} x &= +ct + \mathcal{C}_1 \qquad & x - ct &= \mathcal{C}_1 \\ x &= -ct + \mathcal{C}_2 \qquad & x + ct &= \mathcal{C}_2 \end{aligned} \Bigg\} \quad \begin{aligned} \varphi_1(x,t) &= x - ct \\ \varphi_2(x,t) &= x + ct \end{aligned}$$

1-D Ht equation - PARABOLIC

$$\frac{\partial T}{\partial t} = r\frac{\partial^2 T}{\partial x^2} \qquad A = r \quad B = 0 \quad C = 0$$

$t \rightarrow y$

$T \rightarrow u$

$$\frac{dy}{dx} \rightarrow \frac{dt}{dx} = \frac{B \pm \sqrt{B^2 - 4AC}}{2A} = 0$$

$\downarrow$

slowness

velocity $\frac{dx}{dt} = \infty$

$$t = \text{const} = \int \frac{dt}{dx}\,dx = \int 0\,dx$$

$$t = 0 + \mathcal{C}$$

$$\varphi(x,t) = t$$

2D · steady state heat equation - ELLIPTIC.

$$\frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2} = 0 \qquad A = 1,\ B = 0,\ C = 1$$

$T \rightarrow u$

$$\frac{dy}{dx} = \frac{B \pm \sqrt{B^2 - 4AC}}{2A} = \pm i$$

WHY PDEs? IMPORTANCE?

Physical processes vary with time & location

how will these processes vary with time & space

what drives these processes

where these processes will be in the future & what will happen at some future location

Processes are described by differential equations

if processes depend on 2 or more independent variables then they are governed by partial differential equations

Many processes in natural environs are governed by 2nd order PDEs

$$\frac{\partial^2 u}{\partial x^2} = \frac{1}{c^2}\frac{\partial^2 u}{\partial t^2}$$ wave eqn.

u - displacement

c - bar velocity $= \sqrt{E/\rho}$

$$\frac{\partial T}{\partial t} = \alpha\left(\frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2}\right)$$ heat eqn in 2-D

T - temp

α - thermal diff.

$\alpha = \frac{k}{c\rho}$

Fick's Law of Diff. $\frac{\partial C_A}{\partial t} = D_{AB}\nabla^2 C_A$

$$\frac{\partial^2 \phi}{\partial x^2} + \frac{\partial^2 \phi}{\partial y^2} = \nabla^2 \phi = 0$$ potential flow

$\underline{V} = \nabla \phi$

$\nabla \cdot \underline{V} = 0$

$$\frac{\partial u(x,y)}{\partial x} = \lim_{\Delta x \to 0} \frac{u(x+\Delta x, y) - u(x,y)}{\Delta x}$$

WAVE EQN - APPEARS IN ACOUSTICS, VIBS, SHALLOW WATER WAVE THEORY

HEAT EQN - " IN HEAT TRANSFER, DIFFUSION PROBLEMS

LAPLACES EQ - ' IN POTENTIAL FLOW THEORY, STEADY STATE HEAT TRANS, TORSION OF BAR, STEADY STATE VIB. OF MEMBRANE

WHAT'S COMMON TO ALL
HOW CAN THEY BE SOLVED
WHAT IS A WELL POSED PROBLEM

CHARACTERIZATION & CLASSIFICATION

GENERAL 2nd ORDER PDE

$$\varphi(x, y, u, u_x, u_y, u_{xx}, u_{xy}, u_{yy}) = 0 \qquad u(x,y)$$

$$u_{xx} \equiv \frac{\partial^2 u}{\partial x^2}$$

THE EQ. IS LINEAR WRT ITS HIGHEST DERIVS.

$$\underline{Au_{xx} + Bu_{xy} + Cu_{yy}} + F(x,y,u,u_x,u_y) = 0$$

$\overline{A}$, B, C are fns of x & y ~~alone~~ only

when $F(x,y,u,u_x,u_y)$ then equation is quasi linear

IF $F(x,y,u,u_x,u_y) = \underbrace{Du_x + Eu_y + \mathcal{F}u + G}_{\text{are fns of } x \,\&\, y \text{ only}}$ then equation is linear

IF $G=0$ then equation homogeneous

2nd order linear PDE

$$A u_{xx} + B u_{xy} + C u_{yy} + D u_x + E u_y + F u + G = 0$$

IF ALONG SOME CURVE $\varphi(x,y) = \text{const}$ the equation $A(y')^2 - B(y') + C = 0$

THEN

$$\frac{dy}{dx} = y' = \frac{B \pm \sqrt{B^2 - 4AC}}{2A}$$ & φ is called the characteristic

$B^2 - 4AC > 0$ PDE is hyperbolic & 2 characteristics exist ; $\frac{dy}{dx}$ is real

$B^2 - 4AC = 0$ PDE is parabolic & 1 real characteristic exists

$B^2 - 4AC < 0$ PDE is elliptic & no real characteristics exist

Wave Eqn

$$\frac{\partial^2 u}{\partial x^2} - \frac{1}{c^2}\frac{\partial^2 u}{\partial t^2} = 0 \qquad A = 1 \quad B = 0 \quad C = -\frac{1}{c^2}$$

$t \to y$

$$\frac{dy}{dx} \to \underset{\overset{\uparrow}{\text{slowness}}}{\frac{dt}{dx}} = \frac{0 \pm \sqrt{0 - 4 \cdot 1 \cdot \left(-\frac{1}{c^2}\right)}}{2 \cdot 1} = \pm\frac{1}{c}$$

$$\frac{dx}{dt} = \text{velocity} = \frac{1}{\text{slowness}} = \pm c$$ in magnitude $|c| < \infty$

$$\begin{matrix} x = +ct + \mathcal{C}_1 & & x - ct = \mathcal{C}_1 \\ x = -ct + \mathcal{C}_2 & & x + ct = \mathcal{C}_2 \end{matrix} \left.\vphantom{\begin{matrix} x \\ x \end{matrix}}\right\} \begin{matrix} \varphi_1(x,t) = x - ct \\ \varphi_2(x,t) = x + ct \end{matrix}$$

1-D Ht equation - PARABOLIC

$$\frac{\partial T}{\partial t} = r\frac{\partial^2 T}{\partial x^2} \qquad A = r \quad B = 0 \quad C = 0$$

$t \to y$

$T \to u$

$$\frac{dy}{dx} \to \frac{dt}{dx} = \frac{B \pm \sqrt{B^2 - 4AC}}{2A} = 0$$

$\downarrow$

slowness

velocity $\frac{dx}{dt} = \infty$

$$t = \text{const} = \int \frac{dt}{dx}\,dx = \int 0\,dx$$

$$t = 0 + C$$

$$\varphi(x,t) = t$$

2D steady state heat equation - ELLIPTIC.

$$\frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2} = 0 \qquad A = 1,\ B = 0,\ C = 1$$

$T \to u$

$$\frac{dy}{dx} = \frac{B \pm \sqrt{B^2 - 4AC}}{2A} = \pm i$$

seek a function $u(x, t)$, which is defined in the interior of a rectangle $ABCD$. This region is already determined by the statement of the problem, since the course of the heat propagation in the rod $0 \leqq x \leqq l$ during the time interval $t \leqq t = T$, in which the heat behavior of the boundary is known, was already investigated. Let $t_0 = 0$; we assume that $u(x, t)$ satisfies the heat-conduction equation only for $0 < x < l, 0 < t \leqq T$, i.e., not for $t = 0$ (the side AB) or for $x = 0, x = l$ (the sides AD and BC). For $t = 0$, as well as $x = 0$ and $x = l$, the value of this function is given directly by the initial and boundary conditions. To require that the heat-conduction equation, for example, be satisfied also for $t = 0$ would imply that the derivative $\varphi'' = u_{xx}(x, 0)$ in this equation exists. Therefore, the generality of the physical phenomena to be investigated is limited, and thus the basic functions which do not satisfy this requirement are eliminated from consideration. The condition (3-1.3) loses its meaning when it is not required that $u(x, t)$ in the region $0 \leqq x \leqq l, 0 \leqq t \leqq T$ (i.e., in the closed rectangle $ABCD$) be continuous or this requirement must be replaced by another appropriate assumption.[46] To understand the significance of this requirement we consider the function $v(x, t)$ defined by the following conditions:

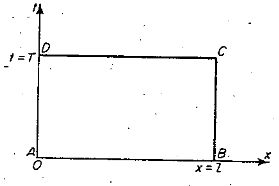


FIG. 37.

$$v(x, t) = C\,, \qquad 0 < x < l\,, \qquad 0 < t \leqq T$$
$$v(x, 0) = \varphi(x)\,, \qquad 0 \leqq x \leqq l$$
$$v(0, t) = \mu_1(t)\,, \qquad v(l, t) = \mu_2(t)\,, \qquad 0 \leqq t \leqq T$$

where C is an arbitrary constant. The function v obviously satisfies both condition (3-1.2) and the boundary conditions. However, this function in no case describes the course of the heat distribution in the rod with an initial temperature $\varphi(x) \neq C$ and boundary temperatures $\mu_1(t) \neq C$ and $\mu_2(t) \neq C$, since it is discontinuous for $t = 0, x = 0, x = l$.

The continuity of $u(x, t)$ for $0 < x < l, 0 < t < T$ directly follows in that $u(x, t)$ satisfies the differential equation. Therefore, the requirement that $u(x, t)$ be continuous in $0 \leqq x \leqq l, 0 \leqq t \leqq T$, is based essentially only on those points at which the boundary and the initial values are prescribed. In the following, by a solution of the equation which satisfies the boundary conditions, we shall always mean a function which satisfies the requirements (3-1.1), (3-1.2), and (3-1.3) and hence not repeat these each time, unless there are special conditions.

Correspondingly, this is the case for other boundary-value problems, in particular for problems of an infinite rod and problems without initial conditions.

---

[46] Later, boundary-value problems with discontinuous boundary and initial conditions will be considered. For these, the problems will be properly defined so that the boundary conditions are fulfilled.

Partial Diff. Eqns. of Math. Physics
by Tykhonov & Samarski

For problems with several independent geometric variables the above statements remain valid. In these problems, an initial temperature and boundary conditions determined on the surface of the body are prescribed for $t = 0$. We can also investigate problems for infinite domains.

With regard to all the problems discussed, the following problems exist[47]:

1. Are the solutions of the problems discussed uniquely determined?
2. Does a solution exist?
3. Do the solutions depend continuously on the auxiliary conditions?

If a problem admits of many solutions, then we naturally cannot speak of "the solution of the problem," and we must first prove the uniqueness. In practice, the second question above is the most important, since generally in proving the existence of a solution, we simultaneously find methods for its calculation.

As noted earlier (see Section 2-2 §3) we speak of a physically determined process when a small change in the initial or boundary conditions causes a small change in the solution. In the following, it will be shown that heat propagation is determined physically by the initial and boundary conditions, i.e., a small change in the initial or boundary conditions implies a small change in the solution.

5. The principle of the maximum

In the following we shall investigate differential equations with constant coefficients,

$$v_t = a^2 v_{xx} + \beta v_x + \gamma v\,. \tag{3-1.34}$$

As already shown, these equations, by the substitution of

$$v = e^{\mu x + \lambda t} u \quad \text{with} \quad \mu = -\frac{\beta}{2a^2}\,, \quad \lambda = \gamma - \frac{\beta^2}{4a^2}$$

can be brought to the form

$$u_t = a^2 u_{xx}\,. \tag{3-1.35}$$

The solutions of this equation have the following properties which will be denoted as the principle of the maximum.

A function $u(x, t)$ defined and continuous in the closed region $0 \leqq t \leqq T$, $0 \leqq x \leqq l$ and satisfying the heat-conduction equation

$$u_t = a^2 u_{xx} \tag{3-1.35}$$

in the region $0 < t < T, 0 < x < l$ assumes its maximum or minimum at the initial moment $t = 0$ or at the boundary points $x = 0$ or $x = l$.

Before we prove this, note that the function $u(x, t) = \text{const.}$ obviously satisfies the heat-conduction equation and assumes a maximum (minimum) at each point. However, this does not contradict our assertion, because it means only that when a maximum (minimum) is assumed in the interior of the region it is also (but not only) assumed for $t = 0$ or for $x = 0$ or $x = l$.

---

[47] *Cf.* Section 2-2.

The physical significance of this statement is immediately clear: if the temperature on the boundary and at the initial moment does not exceed a value M, then in the interior of the body no temperature higher than M can be attained. We shall limit ourselves to the proof of the statement of the maximum and give an indirect proof. We shall designate by M the maximum value of $u(x, t)$ for $t = 0$ $(0 \leqq x \leqq l)$ or for $x = 0$ or $x = l$ $(0 \leqq t \leqq T)$ and assume that the function $u(x, t)$ assumes its maximum at an interior point (x_0, t_0), $(0 < x_0 < l, 0 < t_0 \leqq T)$:[48]

$$u(x_0, t_0) = M + \varepsilon .$$

We now compare the signs in Eq. (3-1.35) at the point (x_0, t_0). Since the function at (x_0, t_0) assumes its maximum,[49] then necessarily

$$\frac{\partial u}{\partial x}(x_0, t_0) = 0 \qquad \text{and} \qquad \frac{\partial^2 u}{\partial x^2}(x_0, t_0) \leqq 0 . \tag{3-1.36}$$

Also, since $u(x_0, t)$ for $t = t_0$ has a maximum,[50] then

$$\frac{\partial u}{\partial t}(x_0, t_0) \geqq 0 . \tag{3-1.37}$$

By comparison of the signs on the left and right sides of (3-1.35) it follows that both sides can be different. These considerations, however, still do not prove the correctness of our theorem; since the right and the left sides can simultaneously equal zero, it would signify no contradiction. We bring forth this consideration simply to emphasize the fundamental concepts of our proof. For the completion of the proof we shall seek more than one point (x_1, t_1) at which $\partial^2 u/\partial x^2 \leqq 0$ and $\partial u/\partial t > 0$. Therefore, we consider the auxiliary function

$$v(x, t) = u(x, t) + k(t_0 - t) , \tag{3-1.38}$$

where k is a constant. Obviously then

$$v(x_0, t_0) = u(x_0, t_0) = M + \varepsilon$$

and

$$k(t_0 - t) \leqq kT .$$

---

[48] If the continuity of $u(x, t)$ were assumed in the bounded region $0 \leqq x \leqq l, 0 \leqq t \leqq T$, then the function $u(x, t)$ could not exceed its maximum, and further considerations would be contradictory. On the basis of the theorem that every continuous function in a bounded region attains its maximum, then (a) the function $u(x, t)$ attains a maximum within or on the boundaries which will be denoted by M; (b) if $u(x, t)$ also were to exceed M only at a point, then a point (x_0, t_0) would exist at which the function $u(x, t)$ assumes a maximum which is larger than M: $u(x_0, t_0) = M + \varepsilon$ $(\varepsilon > 0)$, where $0 < x_0 < l, 0 < t_0 \leqq T$.

[49] As is known from analysis, for the existence of a relative minimum of a function $f(x)$ at an interior point x_0 of an interval $(0, l)$, the conditions

$$\left.\frac{\partial f}{\partial x}\right|_{x=x_0} = 0, \left.\frac{\partial^2 f}{\partial x^2}\right|_{x=x_0} > 0$$

are sufficient. If, therefore, at the point x_0 the function $f(x)$ has a maximum value, then (a) $f'(x_0) = 0$, and (b) $f''(x_0) > 0$ cannot hold; therefore $f''(x_0) \leqq 0$.

[50] Obviously, $\partial u/\partial t = 0$, in case $t_0 < T$, whereas for $t_0 = T$, then $\partial u/\partial t = 0$ must hold.

We now select $k > 0$ so that $kT < \varepsilon/2$, i.e., let $k < \varepsilon/2T$; then the maximum of $v(x, t)$ for $t = 0$ or for $x = 0, x = l$ does not exceed the value $M + \varepsilon/2$, i.e.,

$$v(x, t) \leqq M + \frac{\varepsilon}{2} \qquad \text{for } t = 0 \text{ or } x = 0, x = l , \tag{3-1.39}$$

since for this argument the first summand of (3-1.38) is not larger than M, and the second is not larger than $\varepsilon/2$.

Now, $v(x, t)$ is a continuous function. Thus a point (x_1, t_1) exists at which it assumes its maximum. Then we have

$$v(x_1, t_1) \geqq v(x_0, t_0) = M + \varepsilon .$$

Therefore, $t_1 > 0$ and $0 < x_1 < l$, since for $t = 0$ or $x = 0, x = l$ the inequality (3-1.39) is valid. It follows that

$$v_{xx}(x_1, t_1) = u_{xx}(x_1, t_1) \leqq 0$$

and

$$v_t(x_1, t_1) = u_t(x_1, t_1) - k \geqq 0 \qquad \text{or} \qquad u_t(x_1, t_1) \geqq k > 0 .$$

By comparison of the signs on the right and the left sides in (3-1.35) at the point (x_1, t_1) we conclude that Eq. (3-1.35) at the point (x_1, t_1) cannot be satisfied, since the quantities on the right and left sides have different signs. Therefore, the first part of our proposition is proved. The statement for the minimum can be proved analogously, and it is sufficient to apply the first part to $u_1 = -u$.

6. The uniqueness theorem

We turn now to a series of consequences of the principle of the maximum. First, we prove the uniqueness theorem for the first boundary-value problem. If the functions $u_1(x, t)$ and $u_2(x, t)$, which are defined and continuous in a region $0 \leqq x \leqq l, 0 \leqq t \leqq T$, and which satisfy the heat-conduction equation

$$u_t = a^2 u_{xx} + f(x, t) \qquad \text{for} \qquad 0 < x < l, t > 0 \tag{3-1.35'}$$

as well as the same initial and boundary conditions

$$u_1(x, 0) = u_2(x, 0) = \varphi(x)$$
$$u_1(0, t) = u_2(0, t) = \mu_1(t)$$
$$u_1(l, t) = u_2(l, t) = \mu_2(t) ,$$

then necessarily[51]

$$u_1(x, t) \equiv u_2(x, t) .$$

For the proof of this theorem we consider the function

---

[51] Previously this theorem was refined and the continuity requirement at $t = 0$ was dropped.

$$v(x, t) = u_2(x, t) - u_1(x, t) .$$

Since $u_1(x, t)$ and $u_2(x, t)$ for

$$0 \leqq x \leqq l , \quad 0 \leqq t \leqq T$$

are continuous, their difference $v(x, t)$ in the same region is continuous. Further, $v(x, t)$ as the difference of two solutions of the heat-conduction equation for $0 < x < l, t > 0$ is similarly a solution of the heat-conduction equation in that region. Consequently, the principle of the maximum can also be applied to this function, and the maximum and the minimum of $v(x, t)$ for $t = 0$ or $x = 0$ or $x = l$ is assumed. According to the hypothesis we obtain

$$v(x, 0) = 0 , \quad v(0, t) = 0 , \quad v(l, t) = 0 .$$

Therefore, also

$$v(x, t) \equiv 0 ,$$

i.e.,

$$u_1(x, t) \equiv u_2(x, t) ,$$

from which the uniqueness of the solution of the first boundary-value problem follows.

We shall now prove a series of direct conclusions from the principle of the maximum. In the following discussion we shall refer to "the solution of the heat-conduction equation," instead of enumerating the properties of the function in detail which also satisfy the initial and boundary conditions.

1. If two solutions $u_1(x, t)$ and $u_2(x, t)$ of the heat-conduction equation satisfy the conditions

$$u_1(x, 0) \leqq u_2(x, 0) , \quad u_1(0, t) \leqq u_2(0, t) , \quad u_1(l, t) \leqq u_2(l, t) ,$$

then

$$u_1(x, t) \leqq u_2(x, t)$$

for all $0 \leqq x \leqq l, 0 \leqq t \leqq T$.

The difference $v(x, t) = u_2(x, t) - u_1(x, t)$ satisfies the conditions on which the principal of the maximum is based; also

$$v(x, 0) \geqq 0 \quad v(0, t) \geqq 0 \quad v(l, t) \geqq 0 .$$

Therefore

$$v(x, t) \geqq 0 \quad \text{for} \quad 0 < x < l, 0 < t \leqq T ,$$

since $v(x, t)$ in the region

$$0<x<l , \quad 0<t\leqq T$$

would otherwise have a negative value.

2. If three solutions

$$u(x, t) , \quad \underline{u}(x, t) , \quad \bar{u}(x, t)$$

of the heat-conduction equation satisfy the conditions

$$\underline{u}(x, t) \leqq u(x, t) \leqq \bar{u}(x, t) \quad \text{for} \quad t = 0 , \quad x = 0 , \quad x = l ,$$

then this inequality is fulfilled for all x in $0 \leqq x \leqq l$ and all t in $0 \leqq t \leqq T$.

This assertion represents an application of conclusion (1) to the functions

$$u(x, t) , \quad \bar{u}(x, t) \quad \text{and} \quad u(x, t) , \quad \underline{u}(x, t) .$$

3. If, for two solutions $u_1(x, t)$ and $u_2(x, t)$ of the heat conduction equation, the inequality

$$| u_1(x, t) - u_2(x, t) | \leqq \varepsilon , \quad \text{for} \quad t = 0 , \quad x = 0 , \quad x = l$$

is valid, then

$$| u_1(x, t) - u_2(x, t) | \leqq \varepsilon$$

for all x, t in

$$0 \leqq x \leqq l , \quad 0 \leqq t \leqq T$$

is satisfied.

This assertion results from conclusion (2), when we apply the heat-conduction equation to the solutions

$$\underline{u}(x, t) = - \varepsilon$$
$$u(x, t) = u_1(x, t) - u_2(x, t)$$
$$\bar{u}(x, t) = \varepsilon .$$

The question regarding the continuous dependence of the solution of the first boundary-value problem on the initial and boundary conditions is answered completely by conclusion (3). To understand this, we consider a solution $u(x, t)$ which satisfies other initial and boundary conditions, instead of the solution of the heat-conduction equation which corresponds to the initial and boundary conditions

$$u(x, 0) = \varphi(x) , \quad u(0, t) = \mu_1(t) , \quad u(l, t) = \mu_2(t) .$$

Let these be given by functions $\varphi^*(x)$, $\mu_1^*(t)$ and $\mu_2^*(t)$ which differ by less than ε from the functions $\varphi(x)$, $\mu_1(t)$, and $\mu_2(t)$:

$$| \varphi(x) - \varphi^*(x) | \leqq \varepsilon , \quad | \mu_1(t) - \mu_1^*(t)| \leqq \varepsilon , \quad | \mu_2(t) - \mu_2^*(t) | \leqq \varepsilon .$$

However, the function $u_1(x, t)$ according to conclusion (3) differs by less than ε from the function $u(x, t)$:

$$| u(x, t) - u_1(x, t) | \leqq \varepsilon .$$

Here the principle of the physical determination of a problem arises directly.

We have investigated in detail the question of the uniqueness and the physical determination of a problem in the case of the first boundary-value problem for a bounded interval. The uniqueness theorem for the first boundary-value problem for a two- or three-dimensional bounded region can be proven by a verbatim repetition of these deliberations.

Similar questions arise in the investigation of other problems, an entire

1. A circular tube of radius R. From symmetry only the first quadrant is needed for the numerical computation. The boundary conditions at the two straight edges of the fan-shaped domain are that the variations of velocity normal to the edges are zero. To handle the curved boundary, the method of Program 2.9 may be used.

Compare the numerical result with the analytical solution for Poiseuille flow (Batchelor, 1967, p. 180) that

$$u = -\frac{1}{4\mu}\frac{dp}{dx}(R^2 - r^2)$$

$$Q = -\frac{\pi R^4}{8\mu}\frac{dp}{dx}$$

2. A triangular tube whose two slant walls make 45° angles with the third.

3. A rectangular tube containing a square inner tube.

3.6 Explicit Methods for Solving Parabolic Partial Differential Equations—Generalized Rayleigh Problem

In studying the development of a boundary layer on a body moving through an incompressible fluid, Rayleigh (1911) considered the unsteady motion of an infinitely extended fluid in response to an infinite flat plate suddenly set in motion along its own plane. If the plate is normal to the y-axis and the motion is in the x direction, the continuity equation (3.1.6) is satisfied automatically and the incompressible Navier-Stokes equation (3.1.7) is simplified to

$$\text{forward diff} \quad \frac{\partial u}{\partial t} = \nu \frac{\partial^2 u}{\partial y^2} \quad \text{central diff} \tag{3.6.1}$$

Sometimes this equation, governing arbitary unsteady planar fluid motions, is expressed in terms of vorticity $\zeta(= -\partial u/\partial y)$ in the form

$$\frac{\partial \zeta}{\partial t} = \nu \frac{\partial^2 \zeta}{\partial y^2} \tag{3.6.2}$$

which describes the diffusion of vorticity through a one-dimensional space. According to the discussions of Section 2.9, both (3.6.1) and (3.6.2) are classified as parabolic partial differential equations. Here we will construct a numerical scheme for solving (3.6.1), examine its computational stability, and then apply it to a particular physical problem.

In numerical computations the space coordinates must be finite. Let us assume that the fluid above the plate at $y = 0$ is bounded below a finite depth that is divided into $m - 1$ equally spaced intervals of size h. If the time axis is divided into steps of size τ, a grid system is formed, as shown in Fig. 3.6.1. To approximate (3.6.1) by a finite difference equation at the grid point (i, j), the second-order spatial derivative is replaced by the central-difference formula (2.2.9) and the time derivative is replaced by the forward-difference formula (2.2.6). After rearrangement the equation has the final form

$$u_{i,j+1} = u_{i,j} + R(u_{i-1,j} - 2u_{i,j} + u_{i+1,j}) \qquad R = \nu\frac{\Delta t}{\Delta x^2} \tag{3.6.3}$$

in which $R = \nu\tau/h^2$ is a dimensionless parameter. The equation states that the solution at a certain height at time interval τ later can be computed based on the present informations at the local and two neighboring stations. For given boundary conditions expressed as known functions of time, the solution at time level t_2 is computed explicitly from the initial condition at t_1 by using (3.6.3). Repeating the procedure for the successive time steps, the solution at any desired time level can be obtained. For this property the method in which (3.6.3) is applied is called an *explicit method* for solving the parabolic equation (3.6.1).

Playing a similar role as the Courant number C in (2.12.4), the parameter R in (3.6.3) cannot be arbitrarily chosen, and the limitation imposed on its magnitude is to be determined from a stability analysis of the numerical

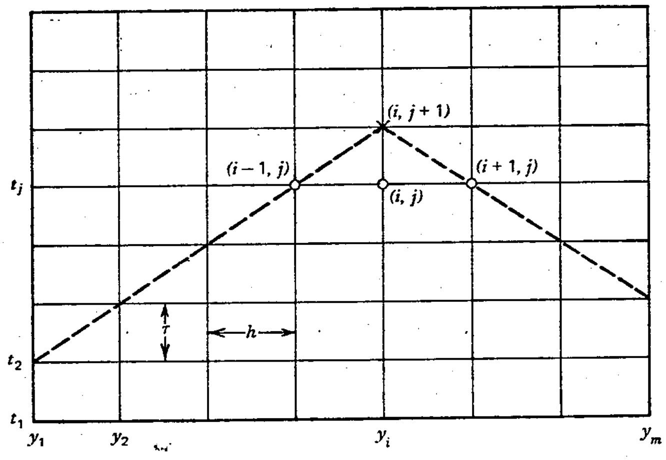


FIGURE 3.6.1 An explicit method for solving parabolic equations.

scheme. Following the technique illustrated in Section 2.12, we assume

$$u_{i,j} = U_j e^{Iikh} \tag{3.6.4}$$

and obtain, after substituting into (3.6.3),

$$U_{j+1} = [1 - 2R(1 - \cos kh)]U_j \tag{3.6.5}$$

The quantity contained within the brackets is the amplification factor λ. If $|\lambda| > 1$, $|U_{j+1}| > |U_j|$ and the amplitude of the solution becomes unbounded as $j \to \infty$. This is called an unstable situation. Thus, for stability we require $\lambda^2 \leqslant 1$ or, consequently, after expanding the left-hand side,

$$R \leqslant \frac{1}{1 - \cos kh}$$

Since the lowest value of the expression on the right-hand side is 1/2 when $\cos kh = -1$, the stability criterion derived for (3.6.3) is

$$\frac{\nu\tau}{h^2} \leqslant \frac{1}{2} \tag{3.6.6}$$

When the upper limiting value is used for this parameter, (3.6.3) has a particularly simple form.

$$u_{i,j+1} = \tfrac{1}{2}(u_{i-1,j} + u_{i+1,j}) \tag{3.6.7}$$

This is called the *Bender-Schmidt recurrence equation*, which determines the solution at (y_i, t_{j+1}) as the average of the values right and left of y_i at a time t_j. However, more accurate results are obtained by using (3.6.3) for $R < 1/2$.

The differential equation (3.6.1) and its finite-difference approximation (3.6.3) apply to any unsteady planar flows bounded by two parallel infinite plates performing arbitrary parallel motions along their own planes. One of the plates may be replaced by a free surface. Furthermore, with modifications to suit cylindrical coordinates, the resulting equations apply to flows between concentric cylinders. Solving for the velocity and the related fields of these flows may be classified as the generalized Rayleigh problem.

For illustrative purposes we consider water contained between two originally stationary flat plates separated by a distance of 1 m. At an initial instant $t = 0$, the upper plate has suddenly acquired a constant speed $u_0(\doteq 1$ m/s) while the lower plate is kept stationary all the time. The sudden motion of the upper plate creates a sharp velocity change there, forming a concentrated vortex sheet right below the plate. The vorticity is diffused downward, according to (3.6.2), into a region practically free of vorticity, and the velocity is redistributed accordingly. We like to find numerically the velocity distribution across the channel at different times.

In terms of the notations of Fig. 3.6.1, the initial velocity distribution is

$$u_{i,1} = 0 \qquad \text{for} \quad i = 1, 2, \ldots, m-1$$
$$u_{m,1} = u_0$$

and the boundary conditions are

$$u_{i,j} = 0 \qquad \text{for} \quad j > 1$$
$$u_{m,j} = u_0 \qquad \text{for} \quad j > 1$$

For water $\nu = 1 \times 10^{-6}$ m²/s, approximately. If the space between plates is divided into 20 equal intervals, $m = 21$ and $h = 0.05$ m. Let us choose $R = 1/4$, which determines the time interval $\tau = Rh^2/\nu$, or 625 s. This time step size seems to be rather large. But it is a reasonable size for a laminar shear flow in which vorticity or velocity gradient is diffused purely by intermolecular activities characterized by a small kinematic viscosity.

In the form shown in (3.6.3), the velocity field is a two-dimensional array. This is not necessary, however, in programming for the computation. In Program 3.5 we use one-dimensional arrays UOLD(I) and UNEW(I) to denote $u_{i,j}$ and $u_{i,j+1}$, respectively. Immediately after completing the computation for one time level, we store the values of UNEW under the name UOLD, print them only at certain selected time levels, and then repeat the same computation to find UNEW at the next time level. This is an often used technique when the size of the field length of a program becomes too large to be handled by the register of a computer.

Because of the large amount of output data, the solution for u is printed every 20 time steps and plotted every 40 time steps, with only the first five curves shown in Fig. 3.6.2. The selective printing and plotting are achieved by assigning these two time-step values to variable names NPRINT and NPLOT, respectively, and doing what we did in Program 2.10—using the special property of integer division.

The output shows that a velocity discontinuity cannot exist in a viscous fluid and is smoothed out immediately by viscous diffusion. As time progresses the velocity profile approaches a linear distribution that varies from 0 at the lower plate to 1 m/s at the upper and corresponds to the solution for the Couette flow between two parallel plates in a steady shear motion.

```
C                          ***** PROGRAM 3.5 *****
C
C        EXPLICIT METHOD IS USED TO FIND THE UNSTEADY VELOCITY
C        DISTRIBUTION IN WATER CONTAINED BETWEEN A STATIONARY
C        LOWER PLATE AND AN UPPER PLATE IN AN IMPULSIVE MOTION
C
      DIMENSION UOLD(21),UNEW(21),Y(21),NSCALE(4),
     A          LABEL(3),PCHAR(5),TPLOT(5)
      REAL NU
```

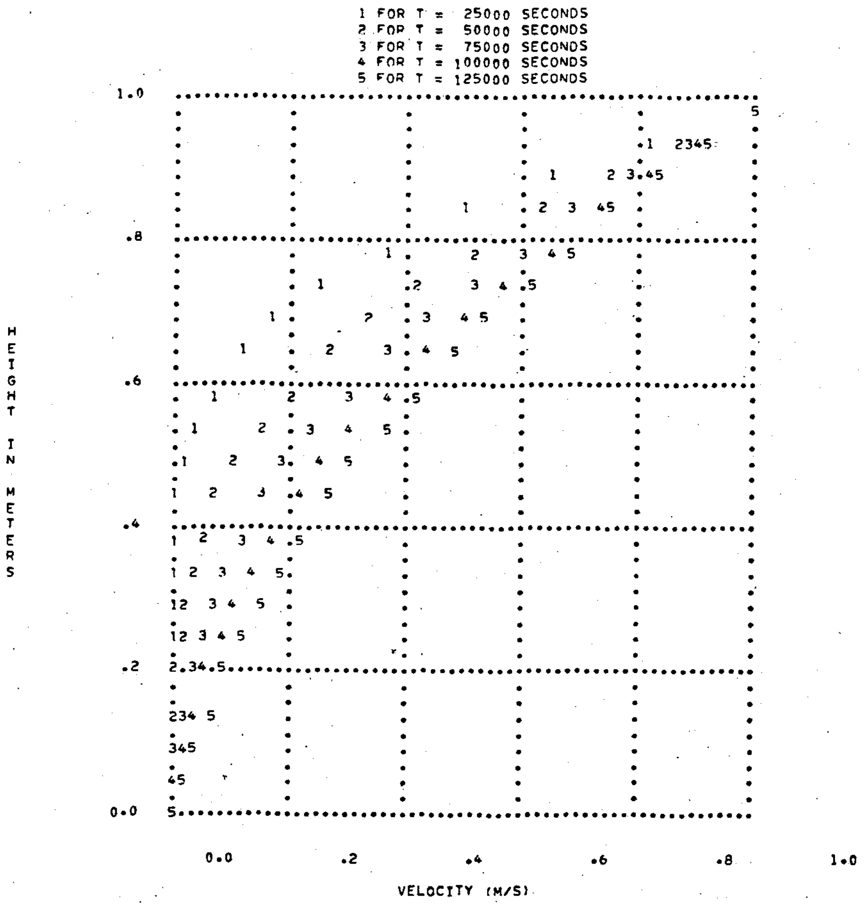


FIGURE 3.6.2 Velocity distribution at different times.

Problem 3.10 Assign a value to R that is greater that 0.5, and then run Program 3.5 to watch the growth of the solution to some unrealistic magnitudes. The result proves the validity of the stability criterion (3.6.6).

Problem 3.11 Find the velocity distribution at increasing times in the originally stationary fluid around a circular cylinder with a free surface (Fig. 3.4.1) after the cylinder is suddenly given a rotation of tangential speed v_b at the surface. As time approaches a very large value, the solution should approach that for Problem 3.6.

Problem 3.12 Find the velocity distribution in the channel described in Program 3.5 with the upper plate replaced by one oscillating at the speed $u_0 \sin \omega t$, where $u_0 = 1$ m/s and $\omega = 1/1000\ \mathrm{s}^{-1}$.

Problem 3.13 In approximating the differential equation (3.6.1) by the finite-difference equation (3.6.3), we used forward difference in time and central difference in space, so that the truncation error of the approximation is $O(\tau, h^2)$. An improved explicit scheme with truncation error of $O(\tau^2, h^2)$ can be constructed by also using central difference in time and by replacing the second $u_{i,j}$ on the right side of (3.6.3) by the time-average $(u_{i,j-1} + u_{i,j+1})/2$. Thus

$$\frac{u_{i,j+1} - u_{i,j-1}}{2\tau} = \nu \frac{u_{i-1,j} - u_{i,j-1} - u_{i,j+1} + u_{i+1,j}}{h^2}$$

$$(1+2R)u_{i,j+1} = 2R(u_{i-1,j} + u_{i+1,j}) + (1-2R)u_{i,j-1} \qquad (3.6.8)$$

This is called the ***DuFort–Frankel formula,*** which involves three time levels, as does the formula (2.12.13) derived for hyperbolic differential equations.

Show that (3.6.8) is an unconditionally stable numerical scheme.

Let us take a close look at the explicit formula (3.6.3). The solution at the grid pont $(i, j+1)$ is computed, using this formula, from the solutions evaluated at three grid points $(i-1, j)$, (i, j), and $(i+1, j)$. These values, in turn, are computed from solutions in their neighborhood at the previous time step. In this way we can trace out the region of dependence of the point $(i, j+1)$, which is confined between the two dashed lines shown in Fig. 3.6.1. It means that the disturbance created at any other height in the fluid reaches the height y_i with a finite speed h/τ. This contradicts the real situation in an incompressible fluid, in which a disturbance at any point is felt immediately by all parts throughout the fluid. Thus, to improve the accuracy of (3.6.3), we may reduce the size of τ or the value of R. In so doing the dashed lines will approach the horizontal grid line passing through $(i, j+1)$ and, in the meantime, more time steps will be needed in the computation to reach the same time level. The improved accuracy in the explicit method is therefore obtained at the expense of an increased amount of computer time.

3.7 Implicit Methods for Solving Parabolic Partial Differential Equations—Starting Flow in a Channel

The deficiency associated with the explicit methods that the solution computed at one point is not affected immediately by the conditions at all other points in the fluid can be avoided by devising an alternative numerical scheme for solving the same diffusion equation (3.6.1). If we still use centered difference in space but use backward, instead of forward, difference in time, a finite-difference equation is obtained at (i, j) of the form

$$\frac{1}{\tau}(u_{i,j}-u_{i,j-1})=\frac{\nu}{h^2}(u_{i-1,j}-2u_{i,j}+u_{i+1,j}) \tag{3.7.1}$$

It becomes, after regrouping,

$$Ru_{i-1,j}-(1+2R)u_{i,j}+Ru_{i+1,j}=-u_{i,j-1} \tag{3.7.2}$$

where $R=\nu\tau/h^2$ is the same dimensionless parameter as that defined in Section 3.6.

The slight modification in approximating the time derivative causes a radical change in the procedure for obtaining a solution. Suppose the solution at $t=t_j$ is to be computed based on the solution known at the previous time step $t=t_{j-1}$; (3.7.2) shows that every three neighboring unknown values are interrelated through this linear algebraic equation. Applying (3.7.2) at all grid points interior to the boundaries at one time level gives a system of simultaneous equations that can be solved for all the unknowns at that time instant. In this way the velocities at different heights are not independent of one another; a change at one point will be felt immediately by all other points. Thus this numerical scheme is more sound than the explicit scheme on physical grounds. By using the numerical scheme the solution can no longer be computed explicitly as before, so (3.7.2) is called a formula for the *implicit method.*

The computational stability of (3.7.2) can be examined again with von Neumann's stability analysis by assuming the form already shown in (3.6.4) for the numerical solution. It can easily be verified that the resulting relationship from that analysis is

$$U_j=\frac{1}{1+2R(1-\cos kh)}U_{j-1}=\lambda U_{j-1} \tag{3.7.3}$$

As $\cos kh$ varies from -1 to $+1$, the value of the amplification factor λ changes from $1/(1+4R)$ to 1 and can never exceed 1. Therefore this numerical scheme is stable for all positive values of R.

Although for computational stability there is no restriction on the magnitude of R as long as it is positive, a smaller value of R results in a more accurate numerical solution. The reason for this is that after multiplying (3.7.1) through by τ, the truncated higher-order terms on the right-hand side are all multiplied by R.

We now apply the implicit method to solve a problem concerning the development of a channel flow caused by the application of a constant pressure gradient. The initially stationary incompressible fluid contained between two parallel infinite plates is set in motion by a suddenly imposed pressure gradient dp/dx along the channel. Simplified for the present geometry, the equation of motion (3.1.7) becomes

$$\frac{\partial u}{\partial t}=-\frac{1}{\rho}\frac{dp}{dx}+\nu\frac{\partial^2 u}{\partial y^2} \tag{3.7.4}$$

If the distance between plates is $2L$ and the origin of the coordinate system is placed at the middle of the channel, the boundary and initial conditions are

$$u=0 \quad \text{at} \quad y=\pm L \quad \text{for all } t \tag{3.7.5}$$

$$u=0 \quad \text{at} \quad t=0 \quad \text{for } -L\leqslant y\leqslant L \tag{3.7.6}$$

We know that as time increases, the velocity profile will approach its steady-state parabolic distribution

$$u_s=-\frac{1}{2\mu}\frac{dp}{dx}(L^2-y^2) \tag{3.7.7}$$

which is a particular solution to (3.7.4) satisfying the boundary conditions (3.7.5). By introducing the dimensionless variables

$$T=\frac{t}{L^2/\nu}, \quad Y=\frac{y}{L}, \quad U=u\Big/\left(-\frac{L^2}{2\mu}\frac{dp}{dx}\right) \tag{3.7.8}$$

and a dimensionless velocity difference

$$W=(u_s-u)\Big/\left(-\frac{L^2}{2\mu}\frac{dp}{dx}\right)=(1-Y^2)-U \tag{3.7.9}$$

the governing equation (3.7.4) is simplified to

$$\frac{\partial W}{\partial T}=\frac{\partial^2 W}{\partial Y^2} \tag{3.7.10}$$

with boundary and initial conditions

$$W=0 \quad \text{at} \quad Y=\pm 1 \quad \text{for all } T \tag{3.7.11}$$

$$W=1-Y^2 \quad \text{at} \quad T=0 \quad \text{for } -1\leqslant Y\leqslant 1 \tag{3.7.12}$$

The implicit numerical scheme for solving (3.7.10) is, according to (3.7.2),

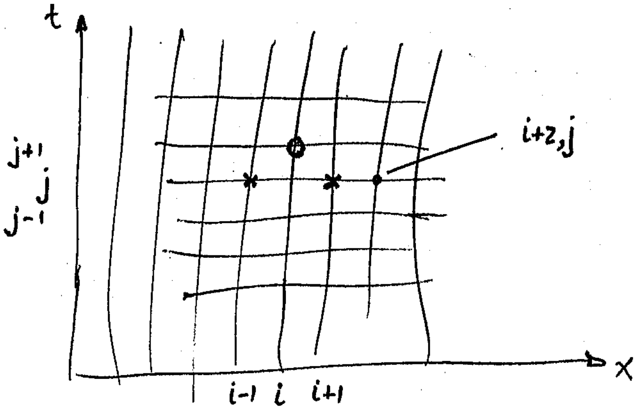
$$R\,W_{i-1,j}-(1+2R)W_{i,j}+R\,W_{i+1,j}=-W_{i,j-1} \tag{3.7.13}$$

$$\dot{T} X = \alpha T X''$$

$$\frac{\dot{T}}{\alpha T} = \frac{X''}{X} = -k^2$$

$$\dot{T} + \alpha k^2 T = 0 \qquad T = e^{-\alpha k^2 t}$$

$$X'' + k^2 X = 0 \qquad X = A \sin kx + B \cos kx \Rightarrow e^{ikx}$$



EGM 6422

LECTURE 11/12

$$\frac{\partial u}{\partial t} = \alpha \frac{\partial u}{\partial x^2}$$ one dim heat equation

EXPLICIT SCHEME $$u_{i,j+1} = u_{i,j} + \overset{C}{\boxed{\frac{\alpha \Delta t}{\Delta x^2}}}\left[u_{i+1,j} - 2u_{i,j} + u_{i-1,j}\right]$$

U^j is the amplitude at t_j,

NEUMANN STABILITY ANALYSIS let $u_{ij} = U^j e^{Iik\Delta x}$ $I = \sqrt{-1}$

$$e^{Iik\Delta x}\left\{ U^{j+1} = U^j + U^j C\left[e^{Ik\Delta x} - 2 + e^{-Ik\Delta x}\right]\right\}$$

$$U^{j+1} = U^j \underbrace{\left\{1 + 2C\left[\cos k\Delta x - 1\right]\right\}}_{A} = U^0 A^{j+1}$$

k is sep var c.

$i\Delta x = x_i$

$e^{I\zeta} + e^{-I\zeta} = 2\cos\zeta$

note U^{j+1} depends on U^j. If we go back U^{j+1} depends on U^0. Any error in U^0 is amplified when $|A|>1$ & dies if $|A|<1$

if $|A|$ is >1 ⟹ solution becomes unbounded

FOR BOUNDEDNESS let $$\left\{1 + 2C\left[\cos k\Delta x - 1\right]\right\}^2 \leq 1$$

$$4C[\quad] + (2C)^2[\quad]^2 \leq 0$$

$$4C\underbrace{\left[\cos k\Delta x - 1\right]}_{-2 \leq [\;] \leq 0}\underbrace{\left\{1 + C\left[\cos k\Delta x - 1\right]\right\}}_{\text{must be } \geq 0} \leq 0$$

$$C \leq \frac{1}{1 - \cos k\Delta x}$$

$1 + C[\cos k\Delta x - 1] \geq 0$

$1 - C[1 - \cos k\Delta x] \geq 0$

$\frac{1}{[1-\cos k\Delta x]} \geq C$

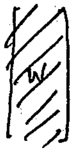
$$C \in [\overset{\cos k\Delta x = -1}{\downarrow}\tfrac{1}{2} \; \& \; \overset{\cos k\Delta x = 1}{\downarrow}\infty]$$ take lower limit $C \leq \frac{1}{2}$

IF $C = \frac{1}{2}$ $$u_{i,j+1} = \frac{1}{2}\left[u_{i+1,j} + u_{i-1,j}\right]$$ Bender-Schmidt condition

- when <u>deriv</u> b.c. are used $C < \frac{1}{2}$
- discont. in the initial conditions can cause errors in your solution

explicit scheme

$$A = \left[\; C \quad 1-2C \quad C \;\right]$$



time j
space i

$$W^j = \begin{bmatrix} w_{i-1,j} \\ w_{i,j} \\ w_{i+1,j} \end{bmatrix}$$

$$\lambda_n = 1 - 4C \sin^2 \frac{n\pi}{2(N+1)}$$

N = # of data points

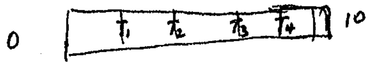
$$W^{j+1} = AW^j = A^{j+1}W^{(0)}$$

error introduced at $W^{(0)}$ will die out if eigenvalues of A are <1

since $\underset{\sim}{W}^{(0)} = \underset{\sim}{W}^{(0)}_{act} + \underset{\sim}{\epsilon}$

$$W^{j+1} = A^{j+1}W^{(0)} + A^{j+1}\epsilon$$

if $\|A\| < 1$ then $\underset{\sim}{A}^{j+1}\underset{\sim}{\epsilon} \leq \|A\|^{j+1}\|\epsilon\|$



$$-0 \cdot R \quad 1+2RT_1 \quad -RT_2 = T_1^{\circ}$$
$$-RT_1 + (1+2R)T_2 - RT_3 = T_2^{\circ}$$
$$-RT_2 + (1+2R)T_3 - RT_4 = T_3^{\circ}$$
$$-RT_3 + (1+2R)T_4 - R \cdot 10 = T_4^{\circ}$$

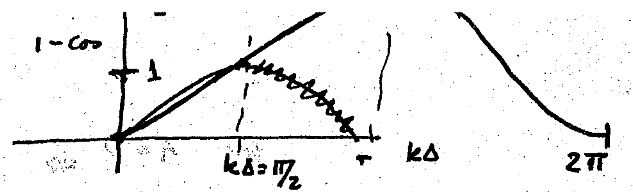
$$\begin{bmatrix} 1+2R & -R & & \\ -R & 1+2R & -R & \\ & -R & 1+2R & -R \\ & & -R & 1+2R \end{bmatrix} \begin{bmatrix} T_1 \\ T_2 \\ T_3 \\ T_4 \end{bmatrix} = \begin{pmatrix} RT_1^{\circ} \\ T_2^{\circ} \\ T_3^{\circ} \\ T_4^{\circ} + 10R \end{pmatrix}$$

implicit BD

$$w^j = Aw^{j+1} = \begin{bmatrix} -C & 1+2C & -C \\ & & \end{bmatrix} w^{j+1}$$

$$\therefore \quad w^{(0)} = A^{j+1} w^{j+1} \quad \rightarrow \quad w^{j+1} = (A^{-1})^{j+1} w^{(0)}$$

λ_n here is $\dfrac{1}{1 + 4C \sin^2 \dfrac{n\pi}{2(N+1)}}$



for solution C can vary from $\frac{1}{2}$ to ∞

for stability $\frac{\alpha \Delta t}{\Delta x^2} \leq \frac{1}{2}$ conditionally stable since it depends on $\underline{\underline{\Delta t}}$

- more accurate results are found for $\frac{\alpha \Delta t}{\Delta x^2} \leq \frac{1}{\text{[illegible]}}$

- in fact modal analysis can be used to show that for $\frac{\alpha \Delta t}{\Delta x^2} < \frac{1}{4}$

$$\lambda_n = 1 - 4C \sin^2 \frac{n\pi}{2(N+1)} \qquad n = 1, 2, \ldots N \quad \text{for } \lambda_n < 1 \quad C < \frac{1}{4}$$

method. is stable non-oscillatory

- what about <u>backward difference in time</u> implicit scheme eval @ i, j+1

$$\cancel{u_{i,j} = u_{i,j} + \frac{\alpha \Delta t}{\Delta x^2} \left[u_{i\,j+1} - 2u_{i,j} + u_{i\,j-1} \right]}$$

$\left.\frac{\partial T}{\partial t}\right|_{ij+1} = \alpha \left.\frac{\partial^2 T}{\partial x^2}\right|_{ij+1}$

$T_{ij} = T_{ij+1} - \Delta t \frac{\partial T}{\partial t}_{ij+1}$

$u_{ij+1}(1+2R) - u_{i+1j+1}R - R\, u_{i-1\,j+1} = u_{ij}$

$$u_{i,j+1} = u_{ij} + \frac{\alpha \Delta t}{\Delta x^2} \left[u_{i+1,j+1} - 2u_{i\,j+1} + u_{i-1\,j+1} \right]$$

$O(\Delta t, \Delta x^2)$

Using $u_{ij} = U^j e^{I i k \Delta x}$

$$e^{I i k \Delta x} \left\{ U^{j+1} = U^j + C \left[U^{j+1} e^{I k \Delta x} - 2U^{j+1} + U^{j+1} e^{-I k \Delta x} \right] \right\}$$

$$U^{j+1} = U^j + C U^{j+1} \left[2(\cos k\Delta x - 1) \right]$$

$$U^{j+1} \left\{ 1 - 2C(\cos k\Delta x - 1) \right\} = U^j \qquad \therefore \quad U^{j+1} = R U^j$$

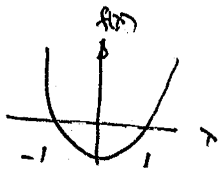
$$R = \frac{1}{1 - 2C\left[\cos k\Delta x - 1\right]} \leq 1$$

this number varies from $\frac{1}{1+4C} \leq R \leq 1$

this is stable for any $C \geqslant 0$ since it is stable irrespective of Δt
it is unconditionally stable – but take C small since error is $O(\Delta t)$

for $A = 0 \Rightarrow C = 0$ impossible since $C = \frac{\alpha \Delta t}{\Delta x^2} > 0$

$A = 0$ eqn $= \lambda^2 - 1$



as $A \uparrow$ curve moves to right both roots are real

and $\underbrace{\text{one root}}_{\text{goes to zero}}$ $|\lambda| < 1$ but $\underbrace{\text{the other } |\lambda| > 1}_{\text{goes to } \infty}$

$$\therefore \quad U^{j+1} = [\underbrace{c_1 \lambda_1^{j+1}}_{\text{can go to zero}} + \underbrace{c_2 \lambda_2^{j+1}}_{\text{goes to infinity}}] U^0$$

LET'S LOOK AT THE CENTERED DIFF IN TIME (LEAPFROG OR RICHARDSON'S)

$$\frac{\partial u}{\partial t}\Big|_{x_i,t_j} = \alpha \frac{\partial^2 u}{\partial x^2}\Big|_{x_i,t_j} \qquad O(\Delta t^2, \Delta x^2)$$

$$u_{ij+1} = u_{ij-1} + \frac{2\alpha\Delta t}{\Delta x^2}\left[u_{i+1j} - 2u_{i,j} + u_{i+1j}\right] \qquad \text{let } u_{ij} = U^j e^{Iik\Delta x}$$

$$e^{Iik\Delta x}\Big\{ U^{j+1} = U^{j-1} + 2c\left[e^{Ik\Delta x}U^j - 2U^j + U^j e^{-Ik\Delta x}\right]\Big\}$$

$$4c\left[U^j(\cos k\Delta x - 1)\right] \qquad A = -4c(\cos k\Delta x - 1)$$

if we assume $U^j = \lambda U^{j-1}$ $\qquad U^{j+1} = \lambda U^j = \lambda^2 U^{j-1} = \lambda^{j+1} U^0$

$$e^{Iik\Delta x} U^{j-1}\left[\lambda^2 = \lambda\{4c(\cos k\Delta x - 1)\} + 1\right] \text{ or } \lambda^2 + 4c(1-\cos k\Delta x)\lambda - 1 = 0$$

$$\lambda^2 = 1 - A\lambda \Rightarrow \lambda^2 + A\lambda - 1 = 0 \qquad A = 4c(1-\cos k\Delta x)$$

$$0 \le A \le 8c$$

$$\lambda = \frac{-A \pm \sqrt{A^2+4}}{2} \qquad \lambda = \frac{-A + \sqrt{A^2 + 4}}{2}$$

$$= -\left(\frac{A}{2}\right) \pm \sqrt{1 + (A/2)^2} \cong -\left(\frac{A}{2}\right) \pm \left[1 + \frac{1}{2}\left(\frac{A}{2}\right)^2 + \dots\right] \qquad \lambda = \frac{-A - \sqrt{A^2+4}}{2} \geq 1$$

$|\lambda| < 1$ cannot be satisfied $\quad \dfrac{2}{-A - \sqrt{A^2+4}}$

no matter what A is
unless A is imag

Du Fort - Frankel Scheme explicit scheme $O(\Delta t^2, \Delta x^2)$

$$\frac{u_{ij+1} - u_{ij-1}}{2\Delta t} = \alpha\left[\frac{u_{i-1j} - u_{ij-1} - u_{ij+1} + u_{i+1j}}{\Delta x^2}\right]$$

eval $\frac{\partial u}{\partial t}\Big|_{i,j}$ $\qquad$ u_{ij} replaced by $\frac{u_{i,j-1} + u_{i,j+1}}{2}$

$$u_{ij+1} = u_{ij-1} + \frac{2\alpha\Delta t}{\Delta x^2}\left[u_{i-1j} - 2u_{ij} + u_{i+1j}\right]$$

HW: show it is unconditionally stable.

why does this work since $w^{(1)} = A\left[w^{(0)} + e^{(0)}\right]$

$$w^{(n)} = Aw^{(n-1)} + \dots = A^n e^{(0)}$$

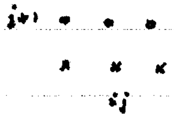
$$\|A^n e^{(0)}\| \le \|A^n\|\,\|e^{(0)}\| \le e^{(0)} \text{ if } \|A^n\| < 1$$

requires the eigenvalues of $A \le 1$

Crank - Nicholson we consider $\frac{\partial u}{\partial t}\left(t_j + \frac{\Delta t}{2}\right) = \frac{u_{i,j+1} - u_{i,j}}{2\left(\Delta t/2\right)} + O\left(\frac{\Delta t}{2}\right)^2$

if we do that then we must take $\frac{\partial^2 u}{\partial x^2}$ also at $t_j + \frac{\Delta t}{2}$

but we define $\frac{\partial^2 u}{\partial x^2}\left(t_j + \frac{\Delta t}{2}\right) = \frac{1}{2}\left[\frac{\partial^2 u}{\partial x^2}(t_j + \Delta t) + \frac{\partial^2 u}{\partial x^2}(t_j)\right]$



then
$$\frac{u_{i,j+1} - u_{i,j}}{\Delta t} = \frac{\alpha}{2}\left[\underbrace{\frac{u_{i+1,j+1} - 2u_{i,j+1} + u_{i-1,j+1}}{\Delta x^2}}_{CD\ @\ t_j+\Delta t} + \underbrace{\frac{u_{i+1,j} - 2u_{i,j} + u_{i-1,j}}{\Delta x^2}}_{CD\ @\ t_j}\right]$$

implicit since information at j+1 depends upon information at adjacent points at j+1 as well

then if $C = \frac{\alpha \Delta t}{\cancel{2}\Delta x^2}$ we get $-C u_{i-1,j+1} + (2+2C) u_{i,j+1} - C u_{i+1,j+1} =$

$$C u_{i-1,j} + (2-2C) u_{i,j} + C u_{i+1,j}$$

$$\begin{bmatrix} -C & 2+2C & -C \end{bmatrix} \begin{bmatrix} u_{i-1} \\ u_i \\ u_{i+1} \end{bmatrix}_{j+1} = \begin{bmatrix} C & 2-2C & C \end{bmatrix} \begin{bmatrix} u_{i-1} \\ u_i \\ u_{i+1} \end{bmatrix}_j$$

or

$$A w^{j+1} = B w^j$$

if $C = 1$ then $\begin{bmatrix} -1 & 4 & -1 \end{bmatrix} \begin{bmatrix} u_{i-1} \\ u_i \\ u_{i+1} \end{bmatrix}_{j+1} = \begin{bmatrix} 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} u_{i-1} \\ u_i \\ u_{i+1} \end{bmatrix}_j$

$$w^{j+1} = A^{-1} B w^j$$

λ_n for $A^{-1}B$ is $\frac{2 - 4C \sin^2(n\pi/2(N-1))}{2 + 4C \sin^2(n\pi/2(N-1))}$ ≤ 1 for any $C \geq 0$

< 1 for any $C > 0$

EGM 6422

ADV. ANAL. OF MECH. SYS.

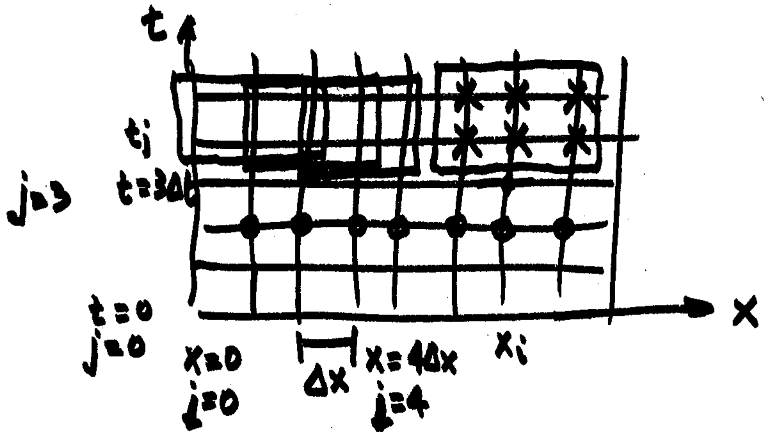
10/13/05

CRANK-NICHOLSON SCHEME

$$\frac{\partial u}{\partial t} = \alpha \frac{\partial^2 u}{\partial x^2} \qquad t = t_j + \frac{\Delta t}{2} \;\&\; x_i$$

CD in time   CD in space $O(\Delta t^2, \Delta x^2)$

$$\frac{u_{ij+1} - u_{ij}}{2\left(\Delta t/2\right)} = \frac{\alpha}{2}\left[\frac{u_{i+1j+1} - 2u_{ij+1} + u_{i-1j+1}}{\Delta x^2} + \frac{u_{i+1j} - 2u_{ij} + u_{i-1j}}{\Delta x^2}\right.$$



$$R = \frac{\alpha \Delta t}{\Delta x^2}$$

$$\begin{bmatrix} 2+2R & -R & 0 & \cdots & & 0 \\ -R & 2+2R & R & 0 & \cdots & 0 \\ & \ddots & \ddots & \ddots & & \\ & & \ddots & \ddots & \ddots & R \\ & & & & -R & 2+2R \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_N \end{bmatrix}_{j+1} = \begin{bmatrix} 2-2R & C & 0 & \cdots & & 0 \\ R & 2-2R & R & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \ddots & & \\ & & \ddots & \ddots & \ddots & R \\ & & & & R & 2-2R \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_N \end{bmatrix}_{j} + BC$$

$$A\,\underset{\sim}{u}^{(j+1)} = B\,\underset{\sim}{u}^{(j)}$$

$$\underset{\sim}{u}^{(j+1)} = A^{-1}B\,\underset{\sim}{u}^{(j)} \longrightarrow = \left(A^{-1}B\right)^{j+1}\underset{\sim}{u}^{(0)}$$

$$\lambda_n \text{ for } A^{-1}B = \frac{2 - 4R\sin^2\left(n\pi/2(N-1)\right)}{2 + 4R\sin^2\left(n\pi/2(N-1)\right)} \leq 1$$

LAX EQUIVALENCE THEOREM

how do know that the numerical scheme converges to the solution of the PDE as Δt & $\Delta x \to 0$

States that if a linear difference equation is consistent with a properly posed linear Initial Value Problem (PDE+IC+BC), then stability is necessary & sufficient condition for convergence.

<u>Well Posed Problem</u>

1) solution is unique if it exists
2) solution depends continuously on the initial data
3) soln always exists for initial data that is arbitrarily close to initial data for which no solution exists.

let u_{ij} the numerical solution

U_{ij} the exact solution

$$\left.\right\} e_{ij} = U_{ij} - u_{ij}$$

$$u_{ij} = U_{ij} - e_{ij}$$

Example: the explicit scheme $O(\Delta t, \Delta x^2)$

$$u_{ij+1} = R(u_{i+1j} + u_{i-1j}) + (1-2R)u_{ij}$$

$$e_{ij+1} = R(e_{i+1j} + e_{i-1j}) + (1-2R)e_{ij} - R\left[U_{i+1j} + U_{i-1j}\right] - (1-2R)U_{ij} + U_{ij+1}$$

$$U_{i+1j} = U(x_{i+1}, t_j) = U_{ij} + \left.\frac{\partial U}{\partial x}\right|_{ij} \Delta x + \frac{\partial^2 U}{\partial x^2}(\xi_1, t_j)\frac{\Delta x^2}{2} \qquad x_i \leq \xi_1 < x_{i+1}$$

$$U_{i-1j} = U(x_{i-1}, t_j) = U_{ij} - \left.\frac{\partial U}{\partial x}\right|_{ij} \Delta x + \frac{\partial^2 U}{\partial x^2}(\xi_2, t_j)\frac{\Delta x^2}{2} \qquad x_{i-1} < \xi_2 < x_i$$

$$U_{ij+1} = U(x_i, t_{j+1}) = U_{ij} + \frac{\partial U}{\partial t}(x_i, \eta) \cdot \Delta t \qquad t_j < \eta \leq t_{j+1}$$

$$e_{ij+1} = R(e_{i+1j} + e_{i-1j}) + (1-2R)e_{ij} + R\left[\frac{\partial^2 U}{\partial x^2}(\ )\frac{\Delta x^2}{2} + \frac{\partial^2 U}{\partial x^2}(\ )\right]\frac{\Delta x^2}{2} + \Delta t \frac{\partial U}{\partial t}(\ )$$

$$R = \alpha\frac{\Delta t}{\Delta x^2}$$

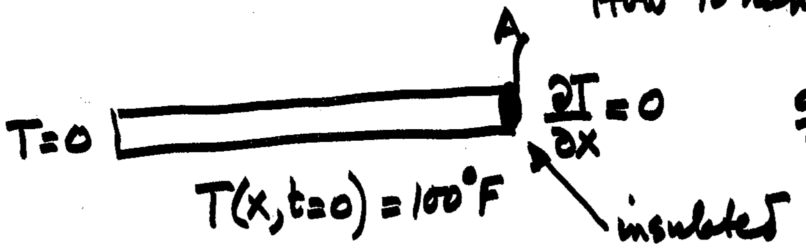
$$e_{ij+1} = R(e_{i+1j} + e_{i-1j}) + (1-2R)e_{ij} + \Delta t \left\{\frac{-\alpha}{2}\left[\frac{\partial^2 U}{\partial x^2}(\ ) + \frac{\partial^2 U}{\partial x^2}(\ )\right] + \frac{\partial U}{\partial t}(\ )\right\}$$

if the errors $\to 0$ as $\Delta x, \Delta t \to 0$

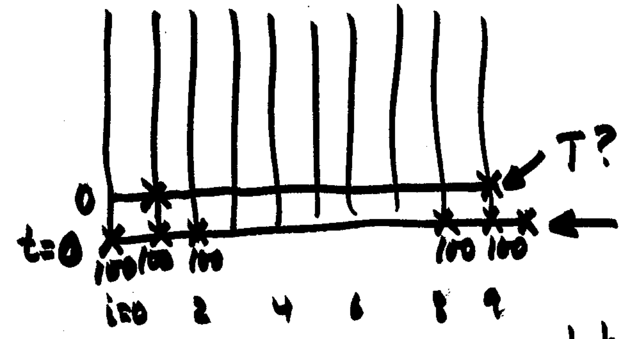
$$0 = -\alpha\frac{\partial^2 U}{\partial x^2} + \frac{\partial U}{\partial t}$$

Assignment show that L.E.T also works for the <u>implicit scheme</u>
Due on 10/20

How to handle derivative bc's



$$q = -kA\frac{\partial T}{\partial x}$$



$$\frac{\partial T}{\partial t} = \alpha\frac{\partial^2 T}{\partial x^2}$$

explicit scheme $O(\Delta t, \Delta x^2)$

$$T_{i1} = R(T_{i+1,0} + T_{i-1,0}) + (1-2R)T_{i0}$$

let $i=1$ $\quad T_{11} = R(\underset{100}{T_{20}} + \underset{100}{T_{00}}) + (1-2R)\underset{100}{T_{10}}$

$$T_{91} = R(\underset{100}{T_{10,0}} + \underset{100}{T_{8,0}}) + (1-2R)\underset{100}{T_{90}}$$

$O(\Delta x^2)$

$$\left.\frac{\partial T}{\partial x}\right|_{ij} = \frac{T_{i+1j} - T_{i-1j}}{2\Delta x} = 0 \Rightarrow T_{i+1j} = T_{i-1j} \Rightarrow T_{10,0} = T_{8,0}$$

$$= \frac{q}{-kA} \qquad + \frac{q \cdot 2\Delta x}{-kA} \qquad + \frac{q \cdot 2\Delta x}{-kA}$$

3 TIME LEVEL SCHEME FOR $\frac{\partial u}{\partial t} = \alpha \frac{\partial^2 u}{\partial x^2}$

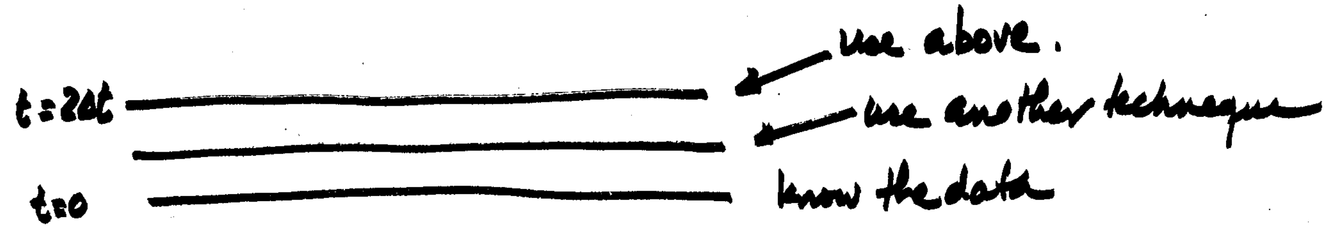
ONLY DO IT TO ACHIEVE AN ADVANTAGE OVER THE 2 TIME LEVEL SCHEME

1) SMALLER LOCAL TRUNCATION ERROR
2) GREATER STABILITY
3) TRANSFORMATION OF A NONLINEAR PROB TO A LINEAR PROBLEM

$$\frac{\partial u}{\partial t} = \alpha \frac{\partial^2 u}{\partial x^2}\Bigg|_{x_i, t_{j+1}} \qquad \begin{array}{l} \text{BD } O(\Delta t^2) \text{ in } t \\ \text{CD } O(\Delta x^2) \text{ in } x \end{array}$$

$$\left[\frac{3}{2}\left(u_{ij+1} - u_{ij}\right) - \frac{1}{2}\left(u_{ij} - u_{ij-1}\right)\right]/\Delta t = \alpha \frac{\left(u_{i+1\,j+1} - 2u_{ij+1} + u_{i-1\,j+1}\right)}{\Delta x^2}$$

- THIS MAINTAINS DISCONT. IN INITIAL DATA
- THIS IS VERY GOOD IN THE CASE OF DATA THAT RAPIDLY VARIES IN X



2-D heat eqn. $\frac{\partial u}{\partial t} = \alpha\left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2}\right)$ $\quad i \rightarrow x$, $j \rightarrow y$, $k \rightarrow t$ $\quad$ FD in t, CD in x & y

Explicit
$$\frac{u_{ijk+1} - u_{ijk}}{\Delta t} = \alpha\left\{\frac{u_{i+1jk} - 2u_{ijk} + u_{i-1jk}}{\Delta x^2} + \frac{u_{ij+1k} - 2u_{ijk} + u_{ij-1k}}{\Delta y^2}\right\}$$

error $O(\Delta t, \Delta x^2, \Delta y^2)$

$$\alpha \Delta t\left\{\frac{1}{\Delta x^2} + \frac{1}{\Delta y^2}\right\} \leq \frac{1}{2}$$

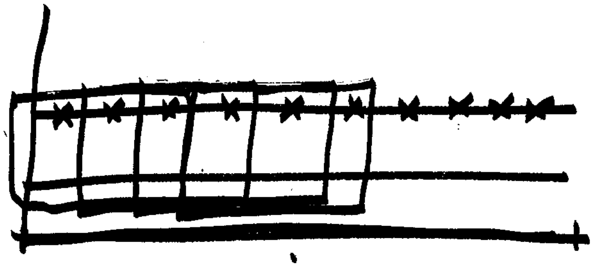
CRANK-NICHOLSON 2-D

$$\frac{u_{ijk+1}-u_{ijk}}{\Delta t} = \frac{\alpha}{2}\left\{ \nabla^2 u\Big|_{ijk} + \nabla^2 u\Big|_{ijk+1} \right\}$$

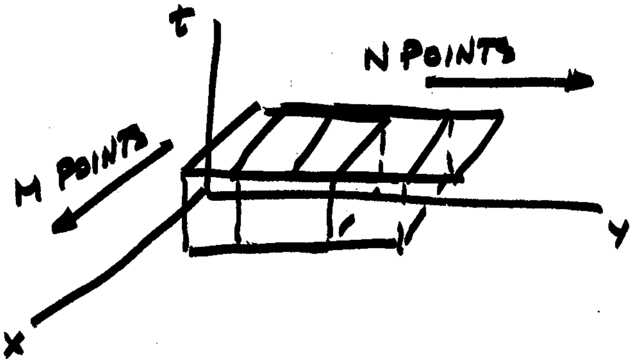
$$= \frac{\alpha}{2}\left\{ \frac{u_{i+1jk}-2u_{ijk}+u_{i-1jk}}{\Delta x^2} + \frac{u_{ij+1k}-2u_{ijk}+u_{ij-1k}}{\Delta y^2} + \frac{u_{i+1jk+1}-2u_{ijk+1}+u_{i-1jk+1}}{\Delta x^2} + \frac{u_{ij+1k+1}-2u_{ijk+1}+u_{ij-1k+1}}{\Delta y^2} \right\}$$

$$O(\Delta t^2, \Delta x^2, \Delta y^2)$$

C-N 1D



M POINTS → M-1 eqs to be solved



$(M-1)(N-1)$ points

Peaceman & Rashford in 1955 Alternating Direction Implicit Method.
(ADI)

READ § 8.2 in chapter 8

Do # 28, 29, 34 10/20 THURS + L.E.T.

Lax Equivalence Theorem

1 Solution is unique when it exists
2 initial & bdry data is given

Stability & Convergence

let $u_{i,j}$ be the numerical
$U_{i,j}$ be the exact

$$\left.\begin{matrix} u_{i,j} \text{ be the numerical} \\ U_{i,j} \text{ be the exact} \end{matrix}\right\} \quad e_{i,j} = U_{i,j} - u_{i,j}$$

for the explicit method $u_{i,j+1} = c\left(u_{i+1,j} + u_{i-1,j}\right) + (1-2c)\,u_{i,j}$

$$u_{i,j} = U_{i,j} - e_{i,j}$$

then
$$e_{i,j+1} = C\left(e_{i+1,j} + e_{i-1,j}\right) + (1-2C)e_{i,j} - C\left(U_{i+1,j} + U_{i-1,j}\right) - (1-2C)U_{i,j} + U_{i,j+1}$$

$$U_{i+1,j} = U_{i,j} + \left.\frac{\partial U}{\partial x}\right|_{i,j} \Delta x + \frac{\Delta x^2}{2}\frac{\partial^2 U}{\partial x^2}(\xi_1, t_j) \qquad x_i \leq \xi \leq x_{i+1}$$

$$U_{i-1,j} = U_{i,j} - \left.\frac{\partial U}{\partial x}\right|_{i,j} \Delta x + \frac{\Delta x^2}{2}\frac{\partial^2 U}{\partial x^2}(\xi_2, t_j)$$

$$U_{i,j+1} = U_{i,j} + \left.\frac{\partial U}{\partial t}\right|_{\eta,j} \Delta t \qquad t_j \leq \eta_j \leq t_{j+1}$$

$$e_{i,j+1} = C\left(e_{i+1,j} + e_{i-1,j}\right) + (1-2C)e_{i,j} + \Delta t \underbrace{\left[\frac{\partial U(x_i,\eta)}{\partial t} - \alpha\frac{\partial^2 U}{\partial x^2}(\xi, t_j)\right]}$$

if the errors $\rightarrow 0$ as $\Delta x, \Delta t \rightarrow 0$ then $\rightarrow \dfrac{\partial U}{\partial t} - \alpha\dfrac{\partial^2 U}{\partial x^2} = 0$

Lax Equiv Theorem states that if a linear difference equation is consistent w/ a properly posed linear IVP then stability is the necessary & sufficient cond. for convergence

Well posed problem
1) Solution unique if it exists
2) Solution depends continuously on initial data
3) Solution always exists for initial data that is arbitrarily close to initial data for which no solution exists

HERE $\underset{\sim}{A}$ is known $\underset{\sim}{b}$ is known. If $\underset{\sim}{A}^{-1}$ exists then $\underset{\sim}{T}$ is known

lets look at parabolic equation $\frac{\partial^2 T}{\partial x^2} = \frac{1}{\alpha}\frac{\partial T}{\partial t}$

solution at time $t+\Delta t$ is dependent on t

$$T(x_i, t_j) = T_{IJ}$$

FORWARD DIFFERENCE $\frac{\partial T}{\partial t} = \frac{T_{I,J+1} - T_{I,J}}{\Delta t}$ order (Δt)

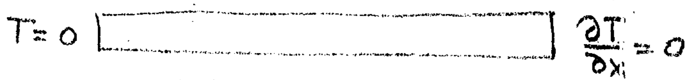
$$\frac{\partial^2 T}{\partial x^2} = \frac{T_{I+1,J} - 2T_{I,J} + T_{I-1,J}}{\Delta x^2} \quad \text{order } (\Delta x)^2$$

$$\Rightarrow \quad T_{I,J+1} = T_{I,J} + \frac{\alpha \Delta t}{\Delta x^2}\left[T_{I+1,J} - 2T_{I,J} + T_{I-1,J}\right]$$

for $\frac{\alpha\Delta t}{\Delta x^2} \leq 0.25$ solution is stable & nonoscillatory

$\frac{\alpha\Delta t}{\Delta x^2} \leq 0.5$ solution is stable

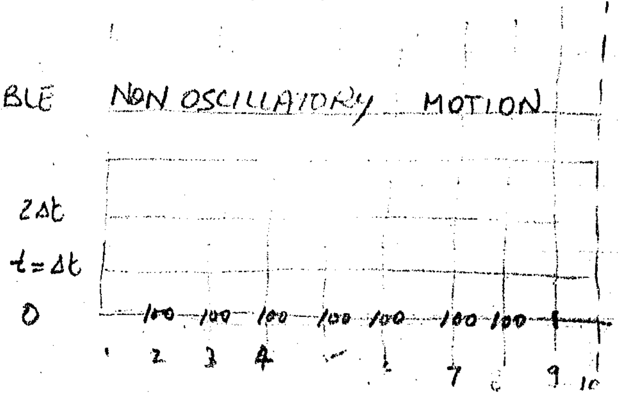
LET'S LOOK AT A ROD



$T(x, t=0) = 100$ °F

$\rho = 168$ lb/ft³
$c = .212$ BTU/lb°F
$k = .0370$ BTU/sec ft°F

FOR STABLE NON OSCILLATORY MOTION



FOR $\alpha = \frac{k}{c\rho} = .00104$ ft²/s
THERMAL DIFFUSIVITY

$\Delta x = 1/12$ ft

TAKE $\Delta t = 1$ sec

$\frac{\alpha\Delta t}{\Delta x^2} = .15$

$$\gamma(\tilde{\alpha}-1)^2 + (1-\beta)(\tilde{\alpha}-1) + 1(1-\beta)$$

let $\tilde{\alpha}-1 = u$

$$u^2 + \frac{(1-\beta)}{\gamma}u + \frac{(1-\beta)}{\gamma} = 0$$

$$-\frac{(1-\beta)}{2\gamma} \pm \frac{\sqrt{\underline{(1-\beta)^2} - 4\,\underline{(1-\beta)}\,\gamma}}{2\gamma}$$

$$\pm\sqrt{\underline{1-\beta}\left[\underline{1-\beta} - 4\gamma\right]}$$

$$\gamma = \frac{\tilde{\alpha}(1-\beta)}{(\tilde{\alpha}-1)^2}$$

AT $X=0$ $t=0$ take T TO BE AVE OF $T(x, t=0) = 100$ and $T(x=0, t) = 0$
or <u>50°F</u>

@ $x = l$ $\frac{\partial T}{\partial x} = 0$ $T_{8,j} = T_{10,j}$

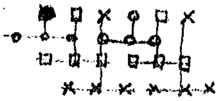
1. use $T_{I,J+1} = T_{I,J} + \frac{\alpha \Delta t}{\Delta x^2} \left[T_{I+1,J} - 2 T_{I,J} + T_{I-1,J} \right]$

FOR PTS 2 - 8

2. use $T_{8,J} = T_{10,J}$ set up an extra colum of imaginary pts at station 10.

3. $T_{I,J} = 100°$ FOR $\begin{array}{l} I \geq 2 \\ J = 0 \end{array}$ FOR $I = 1$ $T_{1,J=0} = \frac{0+100}{2}$

4. Evaluate across x at constant t then start again at next t



IF $-a \frac{\partial T}{\partial x} = b(T - \overline{T})$ @ RHS

$$-a \left[\frac{T_{10,J} - T_{8,j}}{2\Delta x} \right] = b \left(T_{9,J} - \overline{T} \right)$$

$$T_{10,j} = \frac{2\Delta x b}{a} \left(T_{9,j} - \overline{T} \right) \underset{\text{step}}{+} T_{8,j}$$

NOTE DATA AT J+1 ~~pt~~ step depends on data at J'th ~~pt~~ - marching technique Need 2 arrays T1, T

T1 is an array of temp at J+1 step

T " " " " " " J step

problems w/ explicit method

FOR non oscill let $T(I,J) = \tilde{\alpha}^I \beta^J$ $\gamma = \frac{\alpha \Delta t}{\Delta x^2}$

$$\tilde{\alpha}^I \beta^J \left[\beta - 1 - \gamma \left(\tilde{\alpha} - 2 + \frac{1}{\tilde{\alpha}} \right) \right] = 0$$

or $\gamma(\tilde{\alpha}^2 - 2\tilde{\alpha} + 1) + \tilde{\alpha} - \tilde{\alpha}\beta = 0$

$$\tilde{\alpha} = \frac{(\beta - 1 + 2\gamma) \pm \sqrt{(\beta - 1 + 2\gamma)^2 - 4\gamma^2}}{2\gamma}$$

for $\tilde{\alpha}$ to be real. $[(\beta - 1) + 2\gamma]^2 - 4\gamma^2 \geq 0$ $(\beta - 1)[\beta - 1 + 4\gamma] \geq 0$

$$(1-y)^2 - 4\beta(1-y) + 4\beta^2 - 4\beta^2 = \quad (1-y)\left[(1-y) - 4\beta\right]$$

$$x = \frac{y-1+2\beta \pm \sqrt{[(1-y)-2\beta]^2-4\beta^2}}{2\beta}$$

$$(1-y)^2-4\beta(1-y)+\cancel{4\beta^2}-\cancel{4\beta^2}$$

x is real if $(1-y)[1-y-4\beta] > 0$

x is imag if $(1-y)[1-y-4\beta] < 0$

cut off is if $y=1$ or $\frac{1-y}{4} = \beta$

for conv.

$$0 \leq y \leq 1 \implies \frac{1}{4} \geq \beta > 0$$

$y=\frac{1}{2}$ $\quad \frac{-\frac{1}{2}+\frac{1}{4}}{\frac{1}{4}} = -1$

$1-y$

$x = \frac{-1+\frac{1}{2}}{\frac{1}{2}} = -1$

$$x = \frac{y-1+2\beta}{2\beta}$$

$y=0 \quad \beta=\frac{1}{4}$

$y=1 \quad \beta=0$

$$= \frac{-1+\frac{1}{2}}{\frac{1}{2}} = -1$$

$y=0 \quad \beta=\frac{1}{4}$

$y=1 \quad \beta=0$

$$x = \frac{-(1-y)\cancel{y-1}+\frac{1-y}{2}}{\frac{1-y}{2}} = \frac{\frac{1-y}{2}}{\frac{1-y}{2}}\frac{[1-2]}{1} = -1$$

let $x=e^{i\theta}$

$$A+iB = \sqrt{A^2+B^2}\,e^{i\theta}.$$

$$\frac{[(y-1)+2\beta]^2}{4\beta^2} + \frac{(1-y)^2+4\beta(-1+y)}{4\beta^2} \quad -1$$

$$\cancel{2}(y-1)^2+\cancel{8}\beta(y-1)+\cancel{4}\beta^2-4\beta^2-4\beta^2$$

$$(y-1)^2+4\beta(y-1)+2\beta^2-4\beta^2=0$$

$$\tan\theta = \frac{B}{A} = \frac{\sqrt{}}{y-1+2\beta}$$

$$2(y-1)\sqrt{[y-1+4\beta]}$$

$\boxed{e^{i\pi/2}=1}$ $\quad \theta=\pi/2$

$\frac{1-y}{4} = \beta.$ for $\quad |x|=1$

$y=1 \quad \beta=0$

$y=0 \quad \beta=\frac{1}{4}$

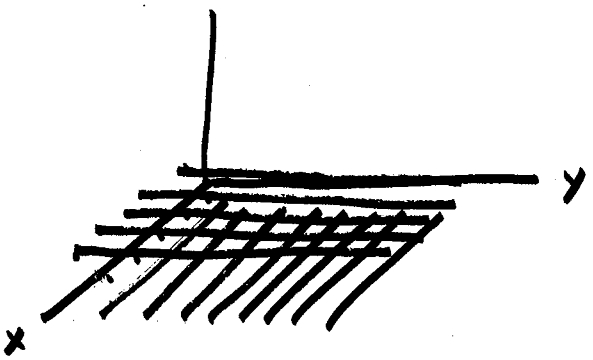
EGM 6422

LECTURE 13/14

ADI method solves the CN eqns in two steps

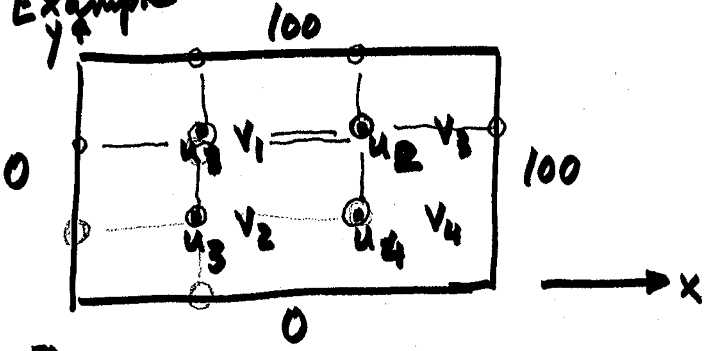
t = Δt 1. Assume x is known & y is unknown

t = 2Δt 2. " y is known & x is unknown



method works well for rectangular regions & every other time step is accurate

Example



i → x
j → y
k → t

KNOWN

$$R\,u_{i+1\,j}^{k} + (1-2R)u_{ij}^{k} + R\,u_{i-1\,j}^{k} = -R\,u_{ij+1}^{k+1} + (1+2R)u_{ij}^{k+1} - R\,u_{ij-1}^{k+1}$$

$$-R\,u_{i+1\,j}^{k+2} + (1+2R)u_{ij}^{k+2} - R\,u_{i-1\,j}^{k+2} = R\,u_{ij+1}^{k+1} + (1-2R)u_{ij}^{k+1} + R\,u_{ij-1}^{k+1}$$

UNKNOWN KNOWN Known

@ k+1 level

@ v_1 $-R\cdot 100 + (1+2R)v_1 - R v_2 = R u_2 + (1-2R)u_1 + 0$

@ v_3 $-R\cdot 100 + (1+2R)v_3 - R v_4 = R\,100 + (1-2R)u_2 + R u_1$

@ v_2 $-R\,v_1 + (1+2R)v_2 - R\cdot 0 = R\cdot u_4 + (1-2R)u_3 + R\cdot 0$

@ v_4 $-R\,v_3 + (1+2R)v_4 - R\cdot 0 = R\cdot 100 + (1-2R)u_4 + R\cdot u_3$

UNKNOWN KNOWN

@ k+2

@ u_1 $-R u_2 + (1+2R)u_1 - R\cdot 0 = R\cdot 100 + (1-2R)v_1 + R v_2$

@ u_2 $-R\,100 + (1+2R)u_2 - R u_1 = R\cdot 100 + (1-2R)v_3 + R v_4$

@ u_3 $-R u_4 + (1+2R)u_3 - R\cdot 0 = R v_1 + (1-2R)v_2 + R\cdot 0$

@ u_4 $-R\cdot 100 + (1+2R)u_4 - R u_3 = R v_3 + (1-2R)v_4 + R\cdot 0$

READ § 8.2 in chapter 8

Do # 28, 29, 34 10/20 THURS + L.E.T.

Miscellaneous methods for improving accuracy

(i) ***Reduction of the local truncation error*** (Douglas' equations)

All derivatives can be expressed exactly in terms of infinite series of forward, backward, or central-differences. For example

$$\frac{\partial^2 U}{\partial x^2}=\frac{1}{h^2}(\delta_x^2 U-\tfrac{1}{12}\delta_x^4 U+\tfrac{1}{90}\delta_x^6 U+\cdots) \qquad (2.35)$$

where the subscript x denotes differencing in the x-direction and the central-differences are defined by

$$\delta_x U_{i,j}=U_{i+\frac{1}{2},j}-U_{i-\frac{1}{2},j}$$

and

$$\delta_x^2 U_{i,j}=\delta_x(\delta_x U_{i,j})=U_{i+1,j}-2U_{i,j}+U_{i-1,j}, \text{ etc.} \qquad (2.36)$$

In the approximation methods already considered the right-hand side of (2.35) has been truncated after the first term. If it is truncated after two or more terms the accuracy of the approximation method will always be improved but this normally increases the number of unknowns in an implicit method and complicates the boundary procedure. For equations involving second-order derivatives however it is possible to eliminate the fourth-order central-differences yet leave the number of unknowns unchanged. For example, if the equation

$$\frac{\partial U}{\partial t}=\frac{\partial^2 U}{\partial x^2}$$

is approximated at the point $(i, j+\frac{1}{2})$ by

$$\frac{1}{k}(u_{i,j+1}-u_{i,j})=\frac{1}{2}\left\{\left(\frac{\partial^2 u}{\partial x^2}\right)_{i,j+1}+\left(\frac{\partial^2 u}{\partial x^2}\right)_{i,j}\right\}$$

$$=\frac{1}{2h^2}(\delta_x^2-\tfrac{1}{12}\delta_x^4+\tfrac{1}{90}\delta_x^6+\ldots)(u_{i,j+1}+u_{i,j}),$$

then the terms involving δ_x^4 can be eliminated by operating on both sides with $(1+\frac{1}{12}\delta_x^2)$. This gives that

$$(1+\tfrac{1}{12}\delta_x^2)(u_{i,j+1}-u_{i,j})=\tfrac{1}{2}r(\delta_x^2 u_{i,j+1}+\delta_x^2 u_{i,j})+O(\delta_x^6)$$

which can be written as

$$\{1+(\tfrac{1}{12}-\tfrac{1}{2}r)\delta_x^2\}u_{i,j+1}=\{1+(\tfrac{1}{12}+\tfrac{1}{2}r)\delta_x^2\}u_{i,j}$$

when terms of order δ_x^6 are neglected. By (2.36) it follows that the differential equation at the point $(i, j+\frac{1}{2})$ can be approximated by the implicit algebraic equation

$$(1-6r)u_{i-1,j+1}+(10+12r)u_{i,j+1}+(1-6r)u_{i+1,j+1}$$
$$=(1+6r)u_{i-1,j}+(10-12r)u_{i,j}+(1+6r)u_{i+1,j}, \qquad (2.37)$$

where $r=k/h^2$. The resulting tridiagonal system of equations can be solved by the algorithm on page 23 and requires exactly the same amount of arithmetic as the Crank–Nicolson method. Whereas, however, the local truncation error (see Chapter 3) of the Crank–Nicolson equation is $O(h^2)+O(k^2)$, it is $O(h^4)+O(k^2)$ for the Douglas equation. As proved in Chapter 3, exercise 11, the equations are stable and consistent for all positive r. The numerical solution of example 2.1 by equations (2.37) for $h=0{\cdot}1$ and $r=1$ at $t=0{\cdot}1$ is compared with the Crank–Nicolson solution for $r=1$ in Table 2.17.

TABLE 2.17

$x=$	0·1	0·2	0·3	0·4	0·5
Solution of P.D.E.	0·0934	0·1776	0·2444	0·2873	0·3021
Douglas solution	0·0941	0·1789	0·2463	0·2895	0·3044
C–N solution	0·0948	0·1803	0·2482	0·2918	0·3069

An explicit difference equation of h^4 accuracy is developed in exercise 12.

(ii) ***Use of three time-level difference equations***

The finite-difference approximation of a parabolic equation needs only two time-levels. Three (or more) time-level schemes can be constructed but naturally this is done only to achieve some advantage over two-level schemes, such as a smaller local truncation error, greater stability, or the transformation of a non-linear problem to a linear one as is demonstrated further on in this chapter. For example, the three-level difference equation

Backward diff at t_{j+1}, x_i $O(\Delta t^2)$

$$\frac{3}{2}\frac{(u_{i,j+1}-u_{i,j})}{\Delta t}-\frac{1}{2}\frac{(u_{i,j}-u_{i,j-1})}{\Delta t}=\frac{(u_{i+1,j+1}-2u_{i,j+1}+u_{i-1,j+1})}{\Delta x^2}$$

approximating $\partial U/\partial t=\partial^2 U/\partial x^2$ has a truncation error of the same order as the Crank–Nicolson equation, namely $O(\Delta t^2)+O(\Delta x^2)$, but is

$$\frac{\partial U}{\partial t}=\frac{\partial^2 U}{\partial x^2}\bigg|_{x_i, t_{j+1}}$$

the better one to use when the initial data is discontinuous or varies very rapidly with x. The Crank–Nicolson approximation should be used when the initial data and its derivatives are continuous (reference 31). In order to solve the first set of equations for $u_{i,2}$ it is necessary to calculate a solution along the first time-level by some other method, it being assumed that the initial data along $t=0$ are known. This first time-level solution must be of the same accuracy as that given by the three-levels equation. A three-level variation of the Douglas equation is

$$\frac{1}{12\cancel{k}\Delta t}\{\tfrac{3}{2}(u_{i+1,j+1}-u_{i+1,j})-\tfrac{1}{2}(u_{i+1,j}-u_{i+1,j-1})\}$$

$$+\frac{5}{6\cancel{k}\Delta t}\{\tfrac{3}{2}(u_{i,j+1}-u_{i,j})-\tfrac{1}{2}(u_{i,j}-u_{i,j-1})\}$$

$$+\frac{1}{12\cancel{k}\Delta t}\{\tfrac{3}{2}(u_{i-1,j+1}-u_{i-1,j})-\tfrac{1}{2}(u_{i-1,j}-u_{i-1,j-1})\}$$

$$=\frac{1}{\cancel{h}^2\Delta x}(u_{i+1,j+1}-2u_{i,j+1}+u_{i-1,j+1}),$$

and like the Douglas equation its truncation error is $O(k^2)+O(h^4)$. A number of such schemes for constant and variable coefficients and for one and two space dimensions are discussed in references 23 and 31.

(iii) *Deferred correction method*

In this method the approximating difference equations are solved as usual. Their solution is then used to calculate a correction term, at each mesh point of the solution domain, which is added to the approximating difference equation at each mesh point. The corrected equations are then re-solved and the process repeated if necessary. The correction terms are numbers obtained by differencing the numerical solution in either the x-direction or the t-direction, or both directions. One method for deriving a correction term for the Crank–Nicolson equations is given in Chapter 3, exercise 16, but the following is better as it is based on a general result. Define the averaging operator μ by $\mu f_{j+\frac{1}{2}}=\frac{1}{2}(f_j+f_{j+1})$ and use the following results which are proved in most introductory books to numerical analysis, namely,

$$k\frac{\partial}{\partial t}\equiv 2\sinh^{-1}(\tfrac{1}{2}\delta_t)\text{ and }\mu_t\equiv(1+\tfrac{1}{4}\delta_t^2)^{\frac{1}{2}}.$$

From these it follows that

$$\tfrac{1}{2}k\left\{\left(\frac{\partial U}{\partial t}\right)_{i,j}+\left(\frac{\partial U}{\partial t}\right)_{i,j+1}\right\}=k\mu_t\left(\frac{\partial U}{\partial t}\right)_{i,j+\frac{1}{2}}$$
$$=\{(1+\tfrac{1}{4}\delta_t^2)^{\frac{1}{2}}\,2\sinh^{-1}(\tfrac{1}{2}\delta_t)\}\,U_{i,j+\frac{1}{2}}.$$

The expansion of the right-hand side into positive powers of δ_t leads to

$$\tfrac{1}{2}k\left\{\left(\frac{\partial U}{\partial t}\right)_{i,j}+\left(\frac{\partial U}{\partial t}\right)_{i,j+1}\right\}=(\delta_t+\tfrac{1}{12}\delta_t^3-\tfrac{1}{120}\delta_t^5+\ldots)\,U_{i,j+\frac{1}{2}}$$

which can be rearranged as

$$\delta_t U_{i,j+\frac{1}{2}}=\tfrac{1}{2}k\left\{\left(\frac{\partial U}{\partial t}\right)_{i,j}+\left(\frac{\partial U}{\partial t}\right)_{i,j+1}\right\}+C_tU_{i,j+\frac{1}{2}},$$

giving

$$U_{i,j+1}-U_{i,j}=\tfrac{1}{2}k\frac{\partial}{\partial t}(U_{i,j}+U_{i,j+1})+C_tU_{i,j+\frac{1}{2}},\qquad(2.38)$$

where

$$C_t\equiv-\tfrac{1}{12}\delta_t^3+\tfrac{1}{120}\delta_t^5+\ldots.$$

Equation (2.38) is a general result relating the value of a continuous function at the $(j+1)$th time-level to its value at the jth time-level in terms of first time-derivatives and central-differences in the t-direction. For the equation

$$\frac{\partial U}{\partial t}=\frac{\partial^2 U}{\partial x^2}$$

it follows that

$$\frac{\partial}{\partial t}\equiv\frac{\partial^2}{\partial x^2}.$$

Hence by (2.38),

$$U_{i,j+1}-U_{i,j}=\tfrac{1}{2}k\frac{\partial^2}{\partial x^2}(U_{i,j}+U_{i,j+1})+C_tU_{i,j+\frac{1}{2}}.$$

Using equation (2.35) it is seen that

$$U_{i,j+1}-U_{i,j}=\frac{1}{2}\frac{k}{h^2}(\delta_x^2-\tfrac{1}{12}\delta_x^4+\tfrac{1}{90}\delta_x^6+\ldots)(U_{i,j+1}+U_{i,j})+C_tU_{i,j+\frac{1}{2}}$$
$$=\tfrac{1}{2}r(\delta_x^2U_{i,j+1}+\delta_x^2U_{i,j})+C\qquad(2.39)$$

Crank-Nichloson in 2-D

$$\frac{u_{i,j,n+1} - u_{i,j,n}}{\Delta t} = \frac{\alpha}{2}\left\{ \nabla^2 u\Big|_{i,j,n} + \nabla^2 u\Big|_{i,j,n+1} \right\}$$

$$= \frac{\alpha}{2}\left\{ \frac{u_{i+1,j,n} - 2u_{i,j,n} + u_{i-1,j,n}}{\Delta x^2} + \frac{u_{i,j+1,n} - 2u_{i,j,n} + u_{i,j-1,n}}{\Delta y^2} + \frac{u_{i+1,j,n+1} + \ldots}{\Delta x^2} \ldots \right\}$$

better than the explicit method

$$\frac{u_{i,j,n+1} - u_{i,j,n}}{\delta t} = \alpha\left\{ \frac{u_{i-1,j,n} - 2u_{i,j,n} + u_{i+1,j,n}}{(\Delta x)^2} + \frac{u_{i,j-1,n} - 2u_{i,j,n} + u_{i,j+1,n}}{(\Delta y)^2} \right.$$

since $\Delta t \alpha \left\{ \frac{1}{(\Delta x)^2} + \frac{1}{(\Delta y)^2} \right\} \leq \frac{1}{2}$ for conv. requires Δt to be small

$$O(\Delta t^2, \Delta x^2, \Delta y^2)$$

$C\text{-}N_{2\text{-}D}$ problem: must solve $(M-1)(N-1)$ equations

better to use Alternating direction - implicit scheme (ADI)

leads to a solution of M-1 independent systems of equations

each system contains N-1 unknowns.

Define co-ordinates, (x, y, t), of the mesh points of the solution domain by

$$x = i\,\delta x, \quad y = j\,\delta y, \quad t = n\,\delta t,$$

where i, j, n are positive integers, and denote the values of u at these mesh points by

$$u(i\,\delta x, j\,\delta y, n\,\delta t) = u_{i,j,n}.$$

The explicit finite-difference representation of equation (2.26),

$$\frac{u_{i,j,n+1} - u_{i,j,n}}{\delta t} = \frac{\kappa}{(\delta x)^2}(u_{i-1,j,n} - 2u_{i,j,n} + u_{i+1,j,n}) + \frac{\kappa}{(\delta y)^2}(u_{i,j-1,n} - 2u_{i,j,n} + u_{i,j+1,n}),$$

appears attractively simple but is computationally laborious because the condition for its validity, which is

$$\kappa\left(\frac{1}{(\delta x)^2} + \frac{1}{(\delta y)^2}\right)\delta t \leqslant \tfrac{1}{2},$$

necessitates extremely small values for δt. For most problems it is an impractical method.

The Crank–Nicolson method, namely

$$\frac{u_{i,j,n+1} - u_{i,j,n}}{\delta t} = \frac{\kappa}{2}\left\{\left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2}\right)_{i,j,n} + \left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2}\right)_{i,j,n+1}\right\},$$

is valid for all values of δx, δy and δt, but requires the solution of $(M-1)(N-1)$ simultaneous algebraic equations for each step forward in time, where $N\,\delta x = a$, $M\,\delta y = b$. Unlike the one-dimensional case they cannot be solved by a simple recursive process. For large values of M and N they would often be solved iteratively.

Peaceman and Rachford, 1955, reference 30, put forward the following method and showed, for a typical problem with a rectangular region in the $x-y$ plane, that it involved about twenty-five times less work than the explicit method and about seven times less work than the Crank–Nicolson method.

Assume the solution is known for time $t = n\,\delta t$. Their method consists of replacing only one of the second-order derivatives, $\partial^2 u/\partial x^2$ say, by an implicit difference approximation in terms of unknown pivotal values of u from the $(n+1)$th time-level, the other second-order derivative, $\partial^2 u/\partial y^2$, being replaced by an explicit finite-difference approximation. Application of the corresponding finite-difference equation to each of the $(N-1)$ mesh points along a row parallel to Ox (Fig. 2.8), then gives $(N-1)$ equations for the $(N-1)$ unknown values of u at these mesh points for time $t = (n+1)\,\delta t$. When there are $(M-1)$ rows parallel to Ox the advancement of the solution over the whole rectangle to the $(n+1)$th time-step involves the solution of $(M-1)$ independent systems of equations, each system containing $(N-1)$ unknowns. The solution of these systems is much easier than the solution of the $(N-1)\times(M-1)$ equations associated with fully implicit methods.

The advancement of the solution to the $(n+2)$th time-level is then achieved by replacing $\partial^2 u/\partial y^2$ by an implicit finite-difference approximation and $\partial^2 u/\partial x^2$ by an explicit one, and writing down the finite-difference equation corresponding to each mesh point along columns parallel to Oy. This gives $(N-1)$ independent systems of equations, each system involving $(M-1)$ unknowns.

The time interval δt must be the same for each advancement.

Provided the solution for successive time-steps is derived by alternating between rows and columns as described above the method is valid for all ratios of $\delta t/(\delta x)^2$ and $\delta t/(\delta y)^2$. Each step on its own is unstable and unilateral repetition leads to an unacceptable growth of errors.

ADI
alternating-direction
implicit scheme

The detail is as follows. (Fig. 2.8.) The equation,

implicit in x

let $\frac{k\delta t}{\delta x^2} = r = \frac{k\delta t}{\delta y^2}$

$$\frac{u_{i,j,n+1} - u_{i,j,n}}{\kappa\,\delta t} = \frac{u_{i-1,j,n+1} - 2u_{i,j,n+1} + u_{i+1,j,n+1}}{(\delta x)^2} + \frac{u_{i,j-1,n} - 2u_{i,j,n} + u_{i,j+1,n}}{(\delta y)^2},$$

explicit in y

$$-r u_{i-1,j}^{(n+1)} + (1+2r)u_{i,j}^{(n+1)} - r u_{i+1,j}^{(n+1)} = r u_{i,j-1}^{n} + (1-2r)u_{i,j}^{(n)} + r u_{i,j+1}^{(n)}$$

is used to advance the solution from the nth to the $(n+1)$th time-step, and the equation

explicit in x

$$\frac{u_{i,j,n+2} - u_{i,j,n+1}}{\kappa\,\delta t} = \frac{u_{i-1,j,n+1} - 2u_{i,j,n+1} + u_{i+1,j,n+1}}{(\delta x)^2} + \frac{u_{i,j-1,n+2} - 2u_{i,j,n+2} + u_{i,j+1,n+2}}{(\delta y)^2},$$

implicit in y

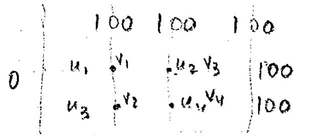
$$-r u_{i,j-1}^{(n+2)} + (1+2r)u_{i,j}^{(n+2)} - r u_{i,j+1}^{(n+2)} = r u_{i-1,j}^{(n+1)} + (1-2r)u_{i,j}^{(n+1)} + r u_{i+1,j}^{(n)}$$

for advancement from the $(n+1)$th to the $(n+2)$th time-step.

At present very little is known about the conditions under which this method is valid, and superior to others, for non-rectangular

ADI method

Peaceman & Rachford



$\uparrow j$
$\rightarrow i$

redo this page

$$\cancel{u_{ij}^{k+1}}\quad r\,u_{ij-1}^{k} + (1-2r)\overset{0}{u}{}_{i,j}^{k} + r\,u_{ij+1}^{k} \overset{\text{known}}{=} \cancel{100} - r\,u_{i-1j}^{k+1} + (1+2r)\overset{\text{Unkn}}{u}{}_{i,j}^{k+1} - r\,u_{i+1j}^{k+1}$$

$$-r\,u_{i+1,j}^{k+2} + (1+2r)u_{i,j}^{k+2} + r\,u_{i-1,j}^{k+2}\ \text{Unk.} = r\,u_{ij+1}^{k+1} + (1-2r)u_{i,j}^{k+1}\ \cancel{(1+2r)} + r\,u_{ij-1}^{k+1}\ \text{known}$$

(

$$\begin{matrix} 1+2r & -r & \\ -r & 1+2r & -r \end{matrix} \begin{bmatrix} V_1 \\ V_2 \end{bmatrix} = \begin{bmatrix} \\ \\ \end{bmatrix}$$

$$\underbrace{-r\,u_{i+1j}^{k+1} + (1+2r)u_{ij}^{k+1} - r\,u_{i-1j}^{k+1}}_{\text{Unknown}} = \underbrace{r\,u_{ij-1}^{k} + (1-2r)u_{ij}^{k} + r\,u_{ij+1}^{k}}_{\text{knowns}}$$

at k+1 level

@ V_1 $\quad -r\overset{0}{\cancel{100}} + (1+2r)u_1 - r\,u_2 = r\,V_2 + (1-2r)V_1 + \cancel{0}\,100$

@ V_3 $\quad -r\overset{u_1}{\cancel{100}} + (1+2r)u_2 - r\overset{100}{\cancel{V_4}} = r\,100 + (1-2r)V_3 + r\,V_4$

@ V_2 $\quad -r\,u_4 + (1+2r)u_3 - r\cdot 0 = r\,V_1 + (1-2r)V_2 + r\cdot 0$

@ V_4 $\quad -r\,u_3 + (1+2r)u_4 - r\,100 = r\cdot\overset{0}{\cancel{100}} + (1-2r)V_4 + r\,V_3$

$$\underbrace{+r\,u_{i+1j}^{k+1} + (1+2r)u_{ij}^{k+1} + r\,u_{i-1j}^{k+1}}_{\text{unknown}} = \underbrace{-r\,u_{ij+1}^{k+2} + (1+2r)u_{ij}^{k+2} + r\,u_{ij-1}^{k+2}}_{\text{known}}$$

at k+2 level

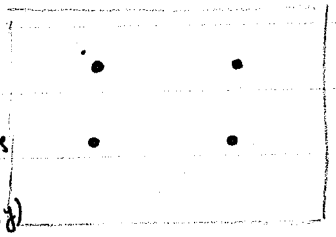
@ u_1 $\quad -r\,u_2 + (1+2r)u_1 - r\cdot 0 = r\cdot 100 + (1-2r)V_1 + r\,V_2$

@ u_2 $\quad -r\cdot 100 + (1+2r)u_2 - r\,u_1 = r\cdot 100 + (1-2r)V_3 + r\,V_4$

@ u_3 $\quad -r\,u_4 + (1+2r)u_3 - r\cdot 0 = r\,V_1 + (1-2r)V_2 + r\cdot 0$

@ u_4 $\quad -r(100) + (1+2r)u_4 - r\,u_3 = r\,V_3 + (1-2r)V_4 + r\cdot 0$

~~pick~~
~~initially data~~
~~is 0°C~~
initial data for points comes from IC.
$u(x,y,t=0) = u_0(x,y)$



$r = \alpha\Delta t/\Delta x^2 \quad \text{if} = .1$

initially ~~u_1, u_2, u_3, u_4~~ $= 0°C$
V_1, V_2, V_3, V_4

$(1/r+2)u_1 - u_2 = 100 + (1/r-2)V_1 + V_2$

$12u_1 - u_2 = 100 + 8V_1 + V_2 = 100$

$u_1 \cong \frac{100}{12} = 8.25$

$-u_1 + (1/r+2)u_2 \cong 200 + (1/r-2)V_3 + V_4$

$12u_2 \cong 208.25 \quad u_2 \cong 17.35$

- Must employ same Δt between any 2 steps – only info every 2nd time step is used
- each step is unstable
- works quite well for rectangular regions loses some of its power in nonrectangular regions
- IN 3-D has problems for some values of r

LECTURE 15/16

EGM 6422

EGM 6422

ADV. ANAL. OF MECH. SYSTEMS

10/18/05

ADI works well for rectangular regions

must use same Δt for each sweep ↕ ↔

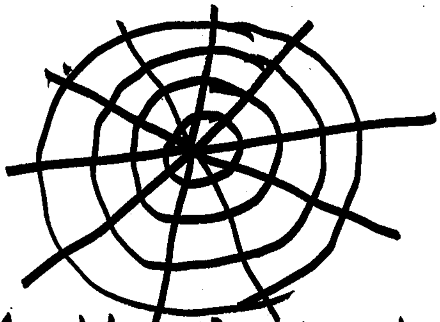
in 3-D some problems exist for certain values of r

Problem is solved on pg 495-497 in your books

Suppose we want to solve $\frac{\partial u}{\partial t} = \alpha\left(\nabla^2 u\right)$ where $\nabla^2 u$ is in (r,θ)

$$\frac{\partial u}{\partial t} = \alpha\left(\frac{\partial^2 u}{\partial r^2} + \underline{\underline{\frac{1}{r}\frac{\partial u}{\partial r}}} + \frac{1}{r^2}\frac{\partial^2 u}{\partial \theta^2}\right) \quad (1)$$ problem at $r=0$

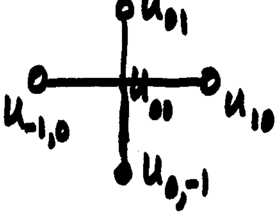
* $r_i = i\Delta r$
  $\theta_j = j\Delta\theta$
  $t_k = k\Delta t$



at the center there is no problem if $\nabla^2 u$ is expressed in cartesian coordinate

$$\nabla^2 u_{ij}^k = \frac{u_{i+1j}^k - 2u_{ij}^k + u_{i-1j}^k}{\Delta x^2} + \frac{u_{ij+1}^k - 2u_{ij}^k + u_{ij-1}^k}{\Delta y^2} \quad (2)$$

use normal eq (1) everywhere except at origen
use equation (2) at the origin

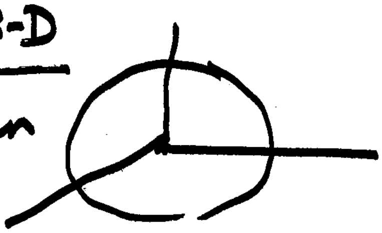


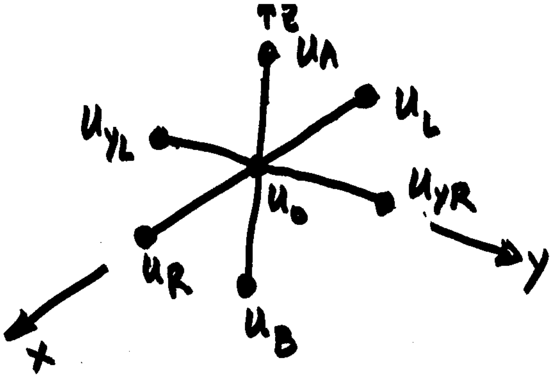
if Δx & Δy are equal

$$\nabla^2 u = \frac{\Sigma \text{ of surrounding pts} - 4u_{00}}{\Delta x^2}$$

suppose we want to solve it in 3-D

we still have problem at origen

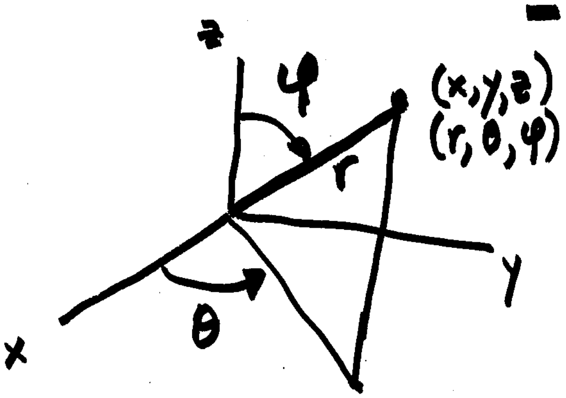




$$\nabla^2 u = \frac{\left[u_A + u_B + u_L + u_R + u_{yR} + u_{yL} - 6u_0\right]}{\Delta x^2}$$

if $\Delta x = \Delta y = \Delta z$

$$\nabla^2 u(r,\theta,\varphi) = \frac{\partial^2 u}{\partial r^2} + \frac{2}{\underline{r}}\frac{\partial u}{\partial r} + \frac{1}{\underline{r^2}}\frac{\partial^2 u}{\partial \theta^2} + \frac{\cot\phi}{\underline{\underline{r}}}\frac{\partial u}{\partial \theta} + \frac{1}{\underline{r^2\sin^2\theta}}\frac{\partial^2 u}{\partial \phi^2}$$



$r_i = i\Delta r$
$\theta_j = j\Delta\theta$
$\varphi_k = k\Delta\varphi$
$t_m = m\Delta t$

Suppose you had a 2-D (r,θ) problem

Suppose the problem had circular symmetry (things don't depend on θ)

$$\nabla^2 u = \frac{\partial^2 u}{\partial r^2} + \frac{1}{r}\frac{\partial u}{\partial r} + \cancel{\frac{1}{r^2}\frac{\partial^2 u}{\partial \theta^2}}$$

if at $r=0$ $\frac{\partial u}{\partial r} = 0$ (symmetry w/r to the origin)

$$u = u(r,t) \qquad \frac{\partial u^*}{\partial r}(r,t) = \underbrace{\frac{\partial u}{\partial r}(r=0,t)}_{=0} + \frac{\partial^2 u}{\partial r^2}(r=0,t)\overset{r}{\cancel{\partial r}} + \dots$$

$$\frac{1}{r}\frac{\partial u}{\partial r} = \frac{\partial^2 u}{\partial r^2}(r=0,t)$$

$$\nabla^2 u(r=0,t) = 2\frac{\partial^2 u}{\partial r^2}(r=0,t)$$ only for circular symmetry & sym wrt the origin

Suppose you had a 3-D (r,θ,φ) problem.

suppose problem had spherical symmetry-no (θ,φ) dependence & suppose that the problem was symmetric wrt origin

$$\nabla^2 u = \frac{\partial^2 u}{\partial r^2} + \frac{2}{r}\frac{\partial u}{\partial r} + \frac{1}{r^2}\cancel{\frac{\partial^2 u}{\partial \theta^2}} + \frac{\cot\theta}{r}\cancel{\frac{\partial u}{\partial \theta}} + \frac{1}{r^2\sin^2\theta}\cancel{\frac{\partial^2 u}{\partial \varphi^2}}$$

$\cong 3\frac{\partial^2 u}{\partial r^2}(r=0,t)$ only for problems w/ spherical symmetry & symmetric about the origin.

---

suppose I can run the same problem w/ Δx_1 & Δt_1 & $\Delta x_2, \Delta t_2$

$\frac{\partial u}{\partial t} = \alpha \frac{\partial^2 u}{\partial x^2}$ $\qquad u_{ij} = u(x_i, t_j)$

$\bar{u}_{ij}$ - solution with $\Delta x_1, \Delta t_1$

$\hat{u}_{ij}$ - solution with $\Delta x_2, \Delta t_2$

U_{ij} - actual

$U_{ij} - \bar{u}_{ij} = A\Delta t_1 + B\Delta x_1^2 + C\Delta t_1^2 + D\Delta x_1^4 + \ldots$ (1) for explicit method

$U_{ij} - \hat{u}_{ij} = A\Delta t_2 + B\Delta x_2^2 + C\Delta t_2^2 + D\Delta x_2^4 + \ldots$ (2)

$(1)\cdot\Delta x_2^2 \mp (2)\cdot\Delta x_1^2$

$$U_{ij} = \frac{\bar{u}_{ij}\Delta x_2^2 - \hat{u}_{ij}\Delta x_1^2}{\Delta x_2^2 - \Delta x_1^2} + A\left\{\frac{\Delta t_1\Delta x_2^2 - \Delta t_2\Delta x_1^2}{\Delta x_2^2 - \Delta x_1^2}\right\} + C\left\{\frac{\Delta t_1^2\Delta x_2^2 - \Delta t_2^2\Delta x_1^2}{\Delta x_2^2 - \Delta x_1^2}\right\}$$

if $\Delta x_2 = 2\Delta x_1$, $\Delta t_2 = 2\Delta t_1$ $\qquad U_{ij} = (4\bar{u}_{ij} - \hat{u}_{ij})/3 + \frac{2}{3}\Delta t_1 \cdot A$

C-N

$$U_{ij} = \hat{u}_{ij} + \frac{\lambda^2}{1-\lambda^2}\left(\hat{u}_{ij} - \bar{u}_{ij}\right)$$

if $\Delta t_2 = \lambda \Delta t_1$

$\Delta x_2 = \lambda \Delta x_1$

$$U - u = A\Delta t^2 + B\Delta x^2 + C\Delta t^4 + D\Delta x^4 + \ldots$$

Suppose we had <u>non linear</u> type parabolic equations

1) difficulties in solving non linear difference equations

One way to solve is by means of linearization

1) Newton's Method.

2) Richtmeyer Type Linearization.

$$U^2 = (V+\Delta V)^2 = V^2 + 2V\Delta V + \underset{\text{neglect}}{\underline{\underline{\Delta V^2}}}$$

$$U^2 + pU = (V+\Delta V)^2 + p(V+\Delta V) = V^2 + 2V\Delta V + \Delta V^2 + p(V+\Delta V$$

= ~~[illegible]~~ $+ (2V+p)\Delta V + V^2 + pV$

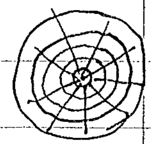
IF you use parabolic equation in 2 or more dimensions

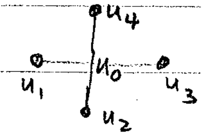
CYLINDRICAL or spherical

$u(r,\theta,z) \quad \nabla^2 u = \frac{\partial^2 u}{\partial r^2} + \frac{1}{r}\frac{\partial u}{\partial r} + \frac{1}{r^2}\frac{\partial^2 u}{\partial \theta^2} + \frac{\partial^2 u}{\partial z^2}$

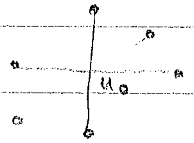
$u(r,\theta,\varphi) \quad = \frac{\partial^2 u}{\partial r^2} + \frac{2}{r}\frac{\partial u}{\partial r} + \frac{1}{r^2}\frac{\partial^2 u}{\partial \theta^2} + \frac{\cot\phi}{r}\frac{\partial u}{\partial \theta} + \frac{1}{r^2\sin^2\theta}\frac{\partial^2 u}{\partial \phi^2}$

Problem is at the origin since $\frac{1}{r}$ terms



CYLIND. at the origin better to use $\nabla^2 u =$  $= \frac{u_1+u_2+u_3+u_4-4u_0}{(\Delta r)^2}$

or $\frac{4(u_m - u_0)}{\Delta r^2}$ where $u_m = \frac{1}{4}\sum_{i=1}^{4} u_i$ that is Σ of all pts at a distance Δr away from origin in a circle.

SPHERICAL  $\frac{6(u_M - u_0)}{\Delta r^2}$ where $u_M = \frac{1}{6}\sum u_i$ that are a distance Δr away in a sphere.

IF circular symmetry $\frac{\partial^2 u}{\partial \theta^2} = 0$ & $\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial r^2} + \frac{1}{r}\frac{\partial u}{\partial r} \quad = \nabla^2 u$

IF we assume that $\frac{\partial u}{\partial r} = 0$ at $r=0$ if problem is symmetric wrt origin

then $u'(r) = \underset{0}{\cancel{u'(0)}} + r u''(0) + \frac{1}{2} r^2 u'''(0) \ldots$ or $\frac{1}{r}\frac{\partial u}{\partial r} = \frac{\partial^2 u}{\partial r^2}$ at origin

$\therefore \quad \frac{\partial u}{\partial t} = 2\frac{\partial^2 u}{\partial r^2}$ at origin only

IF spherical symmetry (ie problem independent of θ, φ) $\Rightarrow$ then $\frac{\partial u}{\partial t} = 3\frac{\partial^2 u}{\partial r^2}$ in spherical symmetry

- Show that the Dufort Frankel Scheme is stable
- Do problem #3 and Problem #21

Due in 9 aug.

5 explit ← 8.31

8.33 → 12 C-N

26 ADI ← 8.34

Li
Trujillo
Wang
Chen
Ayala
Ivan
Chinnapan

- Show that each step in the ADI scheme is unstable, but that

if $r = \frac{\alpha \Delta t_1}{\Delta x_1^2}$ $\quad \therefore \quad \Delta x_1^2 = \frac{\alpha \Delta t_1}{r}$, $\Delta x_2^2 = \frac{\alpha \Delta t_2}{r}$

$$A \left\{ \frac{\Delta t_1 \Delta x_2^2 - \Delta t_2 \Delta x_1^2}{\Delta x_2^2 - \Delta x_1^2} \right\} = A \frac{\alpha}{c} \frac{\Delta t_1 \Delta t_2 - \Delta t_2 \Delta t_1}{\Delta x_2^2 - \Delta x_1^2} = 0$$

$$C \left\{ \frac{\Delta t_1^2 \Delta x_2^2 - \Delta t_2^2 \Delta x_1^2}{\Delta x_2^2 - \Delta x_1^2} \right\} = \frac{\alpha C}{r} \left\{ \frac{\Delta t_1^2 \Delta t_2 - \Delta t_2^2 \Delta t_1}{\frac{\alpha}{r}\left[\Delta t_2 - \Delta t_1\right]} \right\}$$

$$C \Delta t_1 \Delta t_2$$

suppose you are capable of running the program with multiple step sizes

$\Delta t_1, \Delta x_1 \qquad \Delta t_2, \Delta x_2$ let $\bar{u}_{ij}$ = numerical based on $\Delta t_1, \Delta x_1$

then $\hat{u}_{ij}$ = " " " $\Delta t_2, \Delta x_2$

$$U(x_i,t_j) - \bar{u}_{ij} = A\Delta t_1 + B\Delta x_1^2 + C\Delta t_1^2 + D\Delta x_1^4 + \dots$$ FOR EXPLICIT METHOD

$$U(x_i,t_j) - \hat{u}_{ij} = A\Delta t_2 + B\Delta x_2^2 + C\Delta t_2^2 + D\Delta x_2^4 + \dots$$

MULT (1) by Δx_2^2 & (2) by Δx_1^2 & subtract

$$U(\Delta x_2^2 - \Delta x_1^2) - \bar{u}_{ij}(\Delta x_2^2) + \hat{u}_{ij}(\Delta x_1^2) = A(\Delta t_1 \Delta x_2^2 - \Delta t_2 \Delta x_1^2) + C(\Delta t_1^2 \Delta x_2^2 - \Delta t_2^2 \cdot \Delta x_1^2)$$

$$+ D(\qquad)$$

or $$U = \frac{\bar{u}_{ij}\Delta x_2^2 - \hat{u}_{ij}\Delta x_1^2}{\Delta x_2^2 - \Delta x_1^2} + A\left\{\frac{\Delta t_1 \Delta x_2^2 - \Delta t_2 \Delta x_1^2}{\Delta x_2^2 - \Delta x_1^2}\right\} + B\{0\} + C\left\{\frac{\Delta t_1^2 \Delta x_2^2 - \Delta t_2^2}{\Delta x_2^2 - \Delta x_1^2}\right.$$

This is the improvement if 2nd term is small

IF $\Delta x_2 = 2\Delta x_1$ & $\Delta t_2 = 2\Delta t_1$ $$U = \frac{4\bar{u}_{ij} - \hat{u}_{ij}}{3} + \frac{2}{3}\Delta t_1 \cdot A + \dots$$

CAN GET RID OF A IF YOU HAD A 3rd Δt_3 & Δx_3

FOR CRANK NICHOLSON $$U - u = A\Delta t^2 + B\Delta x^2 + C\Delta t^4 + D\Delta x^4 + \dots$$

$$U = \hat{u}_{ij} + \frac{\lambda^2}{1-\lambda^2}(\hat{u}_{ij} - \bar{u}_{ij}) + \dots$$ leave it alone

if $\Delta t_2 = \lambda\Delta t_1 \qquad \Delta x_2 = \lambda\Delta x_1$

what if we have non linear parabolic equations

1) difficulties are with non-linear ~~parabolic~~ difference equations. One way is to somehow linearize

a) Newton's method

~~let $\tilde{u}_{i,j+1} = u_{i,j} + \frac{\partial u_{i,j}}{\partial t}\Delta t$~~ $\qquad u_{i,j+1} = V_i + \epsilon_i$

b) Richtmyer Linearization

LECTURE 16/17

EGM 6422

Newton linearization

suppose from the differential equation we can get the diffence equations so that the diffeunce equations are in the form

$$f_i(u_1, u_2, \ldots, u_N) = 0 \qquad i = 1, \ldots, N$$

then let V_i be a known approx. to u_i for $i = 1, 2, \ldots, N$ so that $f_i(u_1, V_2, \ldots, V_N) \neq 0$

let $u_i = V_i + \epsilon_i$ then

$$f_i(u_1, u_2, \ldots, u_N) = f_i(V_1 + \epsilon_1, V_2 + \epsilon_2, \ldots, V_N + \epsilon_N) = f_i(V_1, V_2, V_3, \ldots, V_N) + \left[\frac{\partial f_i}{\partial V_1}\Delta V_1 + \ldots \frac{\partial f_i}{\partial V_N}\Delta V_N\right]$$

if $u_1, u_2, \ldots, u_N$ are chosen to make $f_i(\quad) = 0$ then

$$-f_i(V_1, V_2, \ldots, V_N) = \frac{\partial f_i}{\partial V_1}\Delta V_1 + \frac{\partial f_i}{\partial V_2}\Delta V_2 + \ldots \frac{\partial f_i}{\partial V_N}\Delta V_N \qquad \text{where } \Delta V_i = u_i - V_i$$

this leads to N equations in N unknowns $\{\Delta V_1, \ldots, \Delta V_N\} = \{\epsilon_1, \epsilon_2, \ldots, \epsilon_N\}$

for example $\frac{\partial U}{\partial t} = \alpha \frac{\partial^2 U^2}{\partial x^2}$ let's use CD at $i, j+\frac{1}{2}$ (CN with $O(\Delta t^2, \Delta x^2)$)

or

$$\frac{u_{ij+1} - u_{ij}}{\Delta t} = \frac{\alpha}{2\Delta x^2}\left[(u_{i-1j+1}^2 - 2u_{ij+1}^2 + u_{i+1j+1}^2) + (u_{i-1j}^2 + 2u_{ij}^2 + u_{i+1j}^2)\right]$$

$\therefore$ let $u_{ij+1} = V_i$ so if $\frac{\Delta x^2}{\alpha \Delta t} = p$

$$V_{i-1}^2 - 2(V_i^2 + pV_i) + V_{i+1}^2 + \{u_{i-1j}^2 - 2(u_{ij}^2 - pu_{ij}) + u_{i+1j}^2\} \overset{=0}{=} f_i(V_{i-1}, V_i, V_{i+1})$$

let V_i be an approximation to $u_{ij+1} = U_i$. Therefore $f_i(V_{i-1}, V_i, V_{i+1}) \neq 0$

then for $\frac{\partial f_i}{\partial V_{i-1}} = 2V_{i-1}$ $\quad \frac{\partial f_i}{\partial V_i} = -2(2V_i + p)$ $\quad \frac{\partial f_i}{\partial V_{i+1}} = 2V_{i+1}$

$$\therefore 2V_{i-1}\Delta V_{i-1} - 2(2V_i + p)\Delta V_i + 2V_{i+1}\Delta V_{i+1} = -f_i(V_1, \ldots, V_N)$$
$$= -\left[V_{i-1}^2 - 2(V_i^2 + pV_i) + V_{i+1}^2 + \{u_{i-1j}^2 - 2(u_{ij}^2 - pu_{ij}) + u_{i+1j}^2\}\right]$$

So for example start $V_i = u_{ij}$; solve for ΔV_i's & check if $\|\underset{\sim}{\Delta V}\| < \epsilon$

if no, then $V_i = V_i + \Delta V_i$

where

$$C=\tfrac{1}{2}r\{(-\tfrac{1}{12}\delta_x^4U_{i,j+1}+\tfrac{1}{90}\delta_x^6U_{i,j+1}+\ldots)$$
$$+(-\tfrac{1}{12}\delta_x^4U_{i,j}+\tfrac{1}{90}\delta_x^6U_{i,j}+\ldots)\}+(-\tfrac{1}{12}\delta_t^3U_{i,j+\frac{1}{2}}+\tfrac{1}{120}\delta_t^5U_{i,j+\frac{1}{2}}).$$

The first approximation to the solution of (2.39) would be found by putting $C=0$ and solving the Crank–Nicolson equations

$$u_{i,j+1}-u_{i,j}=\tfrac{1}{2}r(\delta_x^2u_{i,j+1}+\delta_x^2u_{i,j})$$

over the solution domain $[0<x<1]\times[0<t\leqslant T]$, say. The correction term at each mesh point would then be calculated from a truncated approximation to C such as

$$C'=-\frac{r}{24}(\delta_x^4u_{i,j+1}+\delta_x^4u_{i,j})-\tfrac{1}{12}\delta_t^3u_{i,j+\frac{1}{2}}.$$

This method, in effect, includes higher-order difference terms in the approximations to the derivatives but keeps the matrix of coefficients of the approximation equations tridiagonal which allows the algorithm on page 23 to be used.

(iv) *Richardson's deferred approach to the limit*

For this method two or more solutions approximating the problem must be known for two or more different mesh sizes and the difference between the solution of the partial differential equation and the solution of the approximating equations must be known as a function of the mesh lengths.

Let U represent the solution of the differential equation and $u_{r,s}$ represent the solution of the finite-difference equations for a mesh of size (h_r, k_s). Now assume, for example, that the discretization error

$$U-u(h,k)=Ak+Bh^2+Ck^2+Dh^4+\ldots,$$

as it is for the classical explicit method for finite t.

If two solutions $u_{1,1}$ and $u_{2,2}$ are known then

$$U-u_{1,1}=Ak_1+Bh_1^2+Ck_1^2+Dh_1^4+\ldots$$

and

$$U-u_{2,2}=Ak_2+Bh_2^2+Ck_2^2+Dh_2^4+\ldots.$$

Hence B (or A, but not both) can be eliminated to give that

$$U=\frac{1}{h_2^2-h_1^2}(h_2^2u_{1,1}-h_1^2u_{2,2})+A\frac{k_1h_2^2-k_2h_1^2}{h_2^2-h_1^2}+\ldots$$

If the term involving A is negligible then U is an improvement on $u_{1,1}$ and $u_{2,2}$. For the special case $h_2=2h_1$, $k_2=2k_1$,

$$U=\tfrac{1}{3}(4u_{1,1}-u_{2,2})+\tfrac{2}{3}kA+\ldots.$$

If three different solutions are known then A and B can be eliminated. For the Crank–Nicolson equations,

$$U-u(h,k)=Ah^2+Bk^2+Ch^4+\ldots.$$

If $h_2=\lambda h_1$ and $k_2=\lambda k_1$ it is easily shown that

$$U=u_{2,2}+\frac{\lambda^2}{1-\lambda^2}(u_{2,2}-u_{1,1})+O(h^4).$$

Solution of non-linear parabolic equations

There is no difficulty in formally applying finite-difference methods to non-linear parabolic equations. The difficulties are associated with the difference equations themselves. If they are linear they can usually be solved quite easily, although we still have the problem of determining the conditions that must be satisfied for stability and convergence because the coefficients of the unknowns will be functions of the solution at earlier time-levels. If they are non-linear we have also the problem of their solution. Direct methods, in general, are difficult, so they are usually solved iteratively after being linearized in some way. Taylor's expansion provides a standard way of doing this and the method is usually referred to as Newton's method. As will be seen later this is probably not the best method for parabolic equations but is included for completeness.

Linearization by Newton's method

Let

$$f_i(u_1,u_2,\ldots,u_N)=0,\quad i=1(1)N, \tag{2.40}$$

represent N equations in the N dependent variables $u_1, u_2, \ldots, u_N$.

Let V_i be a known approximation to the exact solution value u_i, $i=1(1)N$.

Put $u_i = V_i + \epsilon_i$ and substitute into equation (2.40). Then by Taylor's expansion to first-order terms in ϵ_i, $i = 1(1)N$,

$$f_i(V_1, V_2, \ldots, V_N) + \left[\frac{\partial f_i}{\partial u_1}\epsilon_1 + \frac{\partial f_i}{\partial u_2}\epsilon_2 + \ldots + \frac{\partial f_i}{\partial u_N}\epsilon_N\right]_{u_i=V_i} = 0,$$
$$i = 1(1)N. \qquad (2.41)$$

The subscript notation on the second bracket indicates that the dependent variables $u_1, u_2, \ldots, u_N$ appearing in the coefficients of $\epsilon_1, \epsilon_2, \ldots, \epsilon_N$ are replaced by $V_1, V_2, \ldots, V_N$ respectively after the differentiations. Equation (2.41) represents N linear equations for the N unknowns $\epsilon_1, \epsilon_2, \ldots, \epsilon_N$ because $V_1, V_2, \ldots, V_N$ are known. When the ϵ's have been calculated the process is repeated, the starting values of the dependent variables for the next iteration being $(V_i + \epsilon_i)$, $i = 1(1)N$. This process of successive approximations is continued until the u_i's have been found to the required degree of accuracy, such as $|\epsilon_i| < 10^{-8}$, $i = 1(1)N$. Some numerical results for a particular problem are given on page 55.

Example 2.7

The function U satisfies the non-linear equation

$$\frac{\partial U}{\partial t} = \frac{\partial^2 U^2}{\partial x^2}, \quad 0 < x < 1,$$

the initial condition $U = 4x(1-x)$, $0 < x < 1$, $t = 0$, and the boundary conditions $U = 0$ at $x = 0$ and 1, $t \geqslant 0$.

If the equation is approximated at the point $\{ih, (j+\frac{1}{2})k\}$ in the $x-t$ plane by the difference scheme

$$\frac{1}{k}\delta_t u_{i,j+\frac{1}{2}} = \frac{1}{2h^2}(\delta_x^2 u_{i,j+1}^2 + \delta_x^2 u_{i,j}^2),$$

use Newton's method to derive a set of linear equations giving an improved value $(V_i + \epsilon_i)$ to the approximate value V_i at the mesh points defined by $x_i = \frac{1}{6}i$, $i = 1(1)5$, $t = k = \frac{1}{36}$.

If the V_i are taken equal to the initial values at $x_i = \frac{1}{6}i$, $i = 1(1)5$, $t = 0$, show that these equations reduce to

$$-19\epsilon_1 + 8\epsilon_2 + \tfrac{14}{9} = 0,$$
$$5\epsilon_1 - 25\epsilon_2 + 9\epsilon_3 - \tfrac{22}{9} = 0,$$

and

$$16\epsilon_2 - 27\epsilon_3 - \tfrac{34}{9} = 0.$$

The approximation equation is

$$\frac{1}{k}(u_{i,j+1} - u_{i,j}) = \frac{1}{2h^2}\{(u_{i-1,j+1}^2 - 2u_{i,j+1}^2 + u_{i+1,j+1}^2)$$
$$+ (u_{i-1,j}^2 - 2u_{i,j}^2 + u_{i+1,j}^2)\}.$$

Put $p = h^2/k$ and denote $u_{i,j+1}$ by u_i. The equation can then be written as

$$u_{i-1}^2 - 2(u_i^2 + pu_i) + u_{i+1}^2 + \{u_{i-1,j}^2 - 2(u_{i,j}^2 - pu_{i,j}) + u_{i+1,j}^2\}$$
$$= 0 \equiv f_i(u_{i-1}, u_i, u_{i+1}).$$

By equation (2.41),

$$\left[\frac{\partial f_i}{\partial u_{i-1}}\epsilon_{i-1} + \frac{\partial f_i}{\partial u_i}\epsilon_i + \frac{\partial f_i}{\partial u_{i+1}}\epsilon_{i+1}\right]_{u_i=V_i} + f_i(V_{i-1}, V_i, V_{i+1}) = 0,$$

hence

$$2V_{i-1}\epsilon_{i-1} - 2(2V_i + p)\epsilon_i + 2V_{i+1}\epsilon_{i+1} + [\{V_{i-1}^2 - 2(V_i^2 + pV_i) + V_{i+1}^2\}$$
$$+ \{u_{i-1,j}^2 - 2(u_{i,j}^2 - pu_{i,j}) + u_{i+1,j}^2\}] = 0, \quad (2.42)$$

where V_i is an approximation to $u_{i,j+1}$.
The problem is symmetric with respect to $x = \frac{1}{2}$. When the V_i are taken equal to $u_{i,0}$, equation (2.42) for $j = 0$ reduces to

$$2u_{i-1,0}\epsilon_{i-1} - 2(2u_{i,0} + p)\epsilon_i + 2u_{i+1,0}\epsilon_{i+1} + \{2u_{i-1,0}^2 - 4u_{i,0}^2 + 2u_{i+1,0}^2\} = 0.$$

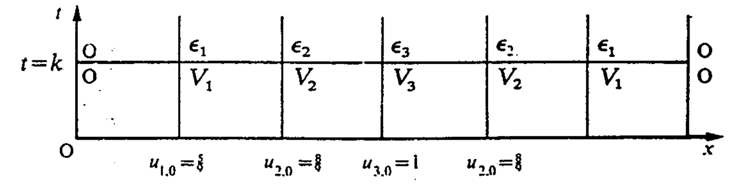
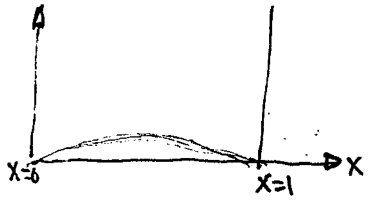


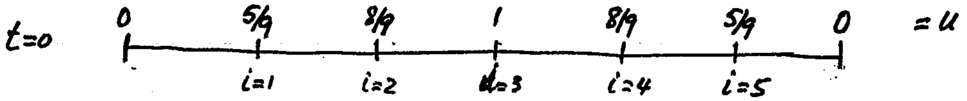
Fig. 2.9

This gives the equations quoted for $p=1$ & the initial values indicated for Fig 2.9 since $\epsilon_0 = 0$.



let $u(x=0,t)=0$ & $u(x=1,t)=0$ & $u(x,t=0) = 4x(1-x)$

let $x=\frac{1}{6}i \quad i=0,\dots,6$



$$\overset{0}{\cancel{2V_0\epsilon_0}} - 2(2V_1+p)\epsilon_1 + 2V_2\epsilon_2 = -\left[\left\{\overset{0}{\cancel{V_0^2}} - 2(V_1^2+pV_1) + V_2^2\right\} + \left\{0^2 - 2(5/9^2 - p5/9) + 8/9^2\right\}\right]$$

$$2V_1\epsilon_1 - 2(2V_2+p)\epsilon_2 + 2V_3\epsilon_3 = -\left[\left\{V_1^2 - 2(V_2^2+pV_2) + V_3^2\right\} + \left\{5/9^2 - 2(8/9^2 - p8/9) + 1^2\right\}\right]$$

$$2V_2\epsilon_2 - 2(2V_3+p)\epsilon_3 + 2V_4\epsilon_4 = -\left[\left\{V_2^2 - 2(V_3^2+pV_3) + V_4^2\right\} + \left\{8/9^2 - 2(1^2 - p\cdot 1) + 8/9^2\right\}\right]$$

$$2V_3\epsilon_3 - 2(2V_4+p)\epsilon_4 + 2V_5\epsilon_5 = -\left[\left\{V_3^2 - 2(V_4^2+pV_4) + V_5^2\right\} + \left\{1^2 - 2(8/9^2 - p8/9) + 5/9^2\right\}\right]$$

$$2V_4\epsilon_4 - 2(2V_5+p)\epsilon_5 + \overset{0}{\cancel{2V_6\epsilon_6}} = -\left[\left\{V_4^2 - 2(V_5^2+pV_5) + \overset{0}{\cancel{V_6^2}}\right\} + \left\{8/9^2 - 2(5/9^2 - p5/9) + 0^2\right\}\right]$$

initially take on RHS $V_1 = 5/9 \quad V_2 = 8/9 \quad V_3 = 1 \quad V_4 = 8/9 \quad V_5 = 5/9$, solve for V_i's.

in all cases V_0 & $V_6 = 0$ since $u(0,t) = u(1,t) = 0$

$P = \frac{\Delta x^2}{\alpha\Delta t}$

IF $p=1$

$$-2\left(\tfrac{10}{9}+1\right)\epsilon_1 + \tfrac{16}{9}\epsilon_2 = -\left[\left\{-2\left(\tfrac{25}{81} + 1\cdot 5/9\right) + \tfrac{64}{81}\right\} + \left\{0^2 - 2\left(\tfrac{25}{81} - 1\cdot 5/9\right) + \tfrac{64}{81}\right\}\right]$$

$$2(5/9)\epsilon_1 - 2\left(\tfrac{16}{9}+1\right)\epsilon_2 + 2\epsilon_3 = -\left[\left\{\tfrac{25}{81} - 2\left(\tfrac{64^2}{81} + 1\cdot\tfrac{8}{9}\right) + 1\right\} + \left\{\tfrac{25}{81} - 2\left(\tfrac{64}{81} - 1\cdot 8/9\right) + 1\right\}\right]$$

⋮

check to see if $\|\Delta \underline{V}\| = \|\underset{\sim}{\epsilon}\| < \epsilon$

you define

$V_0 = 0$

$V_1 = V_1 + \epsilon_1 = 5/9 + \epsilon_1$

$V_2 = V_2 + \epsilon_2 = 8/9 + \epsilon_2$

$V_3 = V_3 + \epsilon_3 = 1 + \epsilon_3$

⋮

V_5

V_6

Check to see if $|f_i(V_{i-1}, V_i, V_{i+1})| < \epsilon$ for $i = 1,\dots,N$

Richtmeyer's linearization used for $\frac{\partial U}{\partial t} = \alpha \frac{\partial^2 U^m}{\partial x^2}$ evaluated at x_i, t_j

Uses $\frac{1}{\Delta t}(u_{i,j+1} - u_{i,j}) = \alpha \left[\theta \frac{\partial^2}{\partial x^2}(u^m_{i,j+1}) + (1-\theta)\frac{\partial^2}{\partial x^2}(u^m_{i,j}) \right]$ implicit weigh average

IF $m = 1$ & $\theta = \frac{1}{2}$ C-N

$$u^m_{i,j+1} = u^m_{i,j} + \Delta t \frac{\partial}{\partial t} u^M_{i,j} + \dots$$ from Taylor series

$$u^m_{i,j} + \Delta t \frac{\partial u^m_{i,j}}{\partial u_{i,j}} \underbrace{\frac{\partial u_{i,j}}{\partial t}} + \dots$$ using chain rule

$$u^m_{i,j} + \Delta t \, m \, u^{m-1}_{i,j} \frac{1}{\Delta t}(u_{i,j+1} - u_{i,j}) + \dots$$

$\therefore$

$$u^m_{i,j+1} = u^m_{i,j} + m u^{m-1}_{i,j}(u_{i,j+1} - u_{i,j})$$ replace unknown $u^m_{i,j+1}$ by approximatio linear in $u_{i,j+1}$

let $w_i = u_{i,j+1} - u_{i,j}$

$$\frac{1}{\Delta t} w_i = \alpha \left[\theta \frac{\partial^2}{\partial x^2}(u^m_{i,j} + m u^{m-1}_{i,j} w_i) + (1-\theta)\frac{\partial^2}{\partial x^2}(u^m_{i,j}) \right]$$

$$= \alpha \left\{ \frac{\partial^2}{\partial x^2}(u^m_{i,j}) + \theta m \frac{\partial^2}{\partial x^2}(u^{m-1}_{i,j} w_i) \right\}$$

$$= \frac{\alpha}{\Delta x^2}\left[\theta m \left\{ u^{m-1}_{i-1,j} w_{i-1} - 2u^{m-1}_{i,j} w_i + u^{m-1}_{i+1,j} w_{i+1} \right\} + \left\{ u^m_{i-1,j} - 2u^m_{i,j} + u^m_{i+1,j} \right\} \right.$$

put this into matrix form & solve for the w_i's

then $u_{i,j+1} = u_{i,j} + w_i$

IF $m = 2$

$$\begin{bmatrix} 1 + 2mC\theta u^{m-1}_{1,j} & -mC\theta u^{m-1}_{2,j} & \\ -mC\theta u^{m-1}_{1,j} & 1 + 2mC\theta u^{m-1}_{2,j} & -mC\theta u^{m-1}_{3,j} \\ & \ddots & \ddots \end{bmatrix} \begin{pmatrix} w_1 \\ w_2 \\ \vdots \end{pmatrix} = \text{RHS}$$

$$\boxed{C = \frac{\alpha \Delta t}{\Delta x^2}}$$

$$\text{RHS} = C\{ u^m_{i-1,j} - 2u^m_{i,j} + u^m_{i+1,j} \}$$

for 1st line $-2C(u^m_{1,j}) + C u^m_{2,j}$

when $m = 1 \Rightarrow \frac{1}{\Delta t} w_i = \alpha \left[\theta \frac{\partial^2}{\partial x^2}(u_{i,j+1}) + (1-\theta)\frac{\partial^2}{\partial x^2}(u_{i,j}) \right]$

if $\theta = 1/2$ we get crank nicholson

scheme unconditionally stable for $\theta \geq 1/4$

for $\theta < 1/4$ $\frac{\Delta t}{\Delta x^2} > \frac{1}{2C \, u_{ij_{max}} (1 - 4\theta)}$

where

$$C=\tfrac{1}{2}r\{(-\tfrac{1}{12}\delta_x^4U_{i,j+1}+\tfrac{1}{90}\delta_x^6U_{i,j+1}+\ldots)$$
$$+(-\tfrac{1}{12}\delta_x^4U_{i,j}+\tfrac{1}{90}\delta_x^6U_{i,j}+\ldots)\}+(-\tfrac{1}{12}\delta_t^3U_{i,j+\frac{1}{2}}+\tfrac{1}{120}\delta_t^5U_{i,j+\frac{1}{2}}).$$

The first approximation to the solution of (2.39) would be found by putting $C=0$ and solving the Crank–Nicolson equations

$$u_{i,j+1}-u_{i,j}=\tfrac{1}{2}r(\delta_x^2u_{i,j+1}+\delta_x^2u_{i,j})$$

over the solution domain $[0<x<1]\times[0<t\leq T]$, say. The correction term at each mesh point would then be calculated from a truncated approximation to C such as

$$C'=-\frac{r}{24}(\delta_x^4u_{i,j+1}+\delta_x^4u_{i,j})-\tfrac{1}{12}\delta_t^3u_{i,j+\frac{1}{2}}.$$

This method, in effect, includes higher-order difference terms in the approximations to the derivatives but keeps the matrix of coefficients of the approximation equations tridiagonal which allows the algorithm on page 23 to be used.

(iv) ***Richardson's deferred approach to the limit***

For this method two or more solutions approximating the problem must be known for two or more different mesh sizes and the difference between the solution of the partial differential equation and the solution of the approximating equations must be known as a function of the mesh lengths.

Let U represent the solution of the differential equation and $u_{r,s}$ represent the solution of the finite-difference equations for a mesh of size (h_r, k_s). Now assume, for example, that the discretization error

$$U-u(h,k)=Ak+Bh^2+Ck^2+Dh^4+\ldots,$$

as it is for the classical explicit method for finite t.

If two solutions $u_{1,1}$ and $u_{2,2}$ are known then

$$U-u_{1,1}=Ak_1+Bh_1^2+Ck_1^2+Dh_1^4+\ldots$$

and

$$U-u_{2,2}=Ak_2+Bh_2^2+Ck_2^2+Dh_2^4+\ldots.$$

Hence B (or A, but not both) can be eliminated to give that

$$U=\frac{1}{h_2^2-h_1^2}(h_2^2u_{1,1}-h_1^2u_{2,2})+A\frac{k_1h_2^2-k_2h_1^2}{h_2^2-h_1^2}+\ldots$$

If the term involving A is negligible then U is an improvement on $u_{1,1}$ and $u_{2,2}$. For the special case $h_2=2h_1$, $k_2=2k_1$,

$$U=\tfrac{1}{3}(4u_{1,1}-u_{2,2})+\tfrac{2}{3}kA+\ldots.$$

If three different solutions are known then A and B can be eliminated. For the Crank–Nicolson equations,

$$U-u(h,k)=Ah^2+Bk^2+Ch^4+\ldots.$$

If $h_2=\lambda h_1$ and $k_2=\lambda k_1$ it is easily shown that

$$U=u_{2,2}+\frac{\lambda^2}{1-\lambda^2}(u_{2,2}-u_{1,1})+O(h^4).$$

Solution of non-linear parabolic equations

There is no difficulty in formally applying finite-difference methods to non-linear parabolic equations. The difficulties are associated with the difference equations themselves. If they are linear they can usually be solved quite easily, although we still have the problem of determining the conditions that must be satisfied for stability and convergence because the coefficients of the unknowns will be functions of the solution at earlier time-levels. If they are non-linear we have also the problem of their solution. Direct methods, in general, are difficult, so they are usually solved iteratively after being linearized in some way. Taylor's expansion provides a standard way of doing this and the method is usually referred to as Newton's method. As will be seen later this is probably not the best method for parabolic equations but is included for completeness.

Linearization by Newton's method

Let

$$f_i(u_1,u_2,\ldots,u_N)=0,\quad i=1(1)N, \tag{2.40}$$

represent N equations in the N dependent variables $u_1, u_2, \ldots, u_N$.

Let V_i be a known approximation to the exact solution value u_i, $i=1(1)N$.

Put $u_i = V_i + \epsilon_i$ and substitute into equation (2.40). Then by Taylor's expansion to first-order terms in ϵ_i, $i = 1(1)N$,

$$f_i(V_1, V_2, \ldots, V_N) + \left[\frac{\partial f_i}{\partial u_1}\epsilon_1 + \frac{\partial f_i}{\partial u_2}\epsilon_2 + \ldots + \frac{\partial f_i}{\partial u_N}\epsilon_N\right]_{u_i = V_i} = 0, \quad i = 1(1)N. \tag{2.41}$$

The subscript notation on the second bracket indicates that the dependent variables $u_1, u_2, \ldots, u_N$ appearing in the coefficients of $\epsilon_1, \epsilon_2, \ldots, \epsilon_N$ are replaced by $V_1, V_2, \ldots, V_N$ respectively after the differentiations. Equation (2.41) represents N linear equations for the N unknowns $\epsilon_1, \epsilon_2, \ldots, \epsilon_N$ because $V_1, V_2, \ldots, V_N$ are known. When the ϵ's have been calculated the process is repeated, the starting values of the dependent variables for the next iteration being $(V_i + \epsilon_i)$, $i = 1(1)N$. This process of successive approximations is continued until the u_i's have been found to the required degree of accuracy, such as $|\epsilon_i| < 10^{-8}$, $i = 1(1)N$. Some numerical results for a particular problem are given on page 55.

Example 2.7

The function U satisfies the non-linear equation

$$\frac{\partial U}{\partial t} = \frac{\partial^2 U^2}{\partial x^2}, \quad 0 < x < 1,$$

the initial condition $U = 4x(1-x)$, $0 < x < 1$, $t = 0$, and the boundary conditions $U = 0$ at $x = 0$ and 1, $t \geqslant 0$.

If the equation is approximated at the point $\{ih, (j+\frac{1}{2})k\}$ in the $x-t$ plane by the difference scheme

$$\frac{1}{k}\delta_t u_{i,j+\frac{1}{2}} = \frac{1}{2h^2}(\delta_x^2 u_{i,j+1}^2 + \delta_x^2 u_{i,j}^2),$$

use Newton's method to derive a set of linear equations giving an improved value $(V_i + \epsilon_i)$ to the approximate value V_i at the mesh points defined by $x_i = \frac{1}{6}i$, $i = 1(1)5$, $t = k = \frac{1}{36}$.

If the V_i are taken equal to the initial values at $x_i = \frac{1}{6}i$, $i = 1(1)5$, $t = 0$, show that these equations reduce to

$$-19\epsilon_1 + 8\epsilon_2 + \tfrac{14}{9} = 0,$$

$$5\epsilon_1 - 25\epsilon_2 + 9\epsilon_3 - \tfrac{22}{9} = 0,$$

and

$$16\epsilon_2 - 27\epsilon_3 - \tfrac{34}{9} = 0.$$

The approximation equation is

$$\frac{1}{k}(u_{i,j+1} - u_{i,j}) = \frac{1}{2h^2}\{(u_{i-1,j+1}^2 - 2u_{i,j+1}^2 + u_{i+1,j+1}^2) + (u_{i-1,j}^2 - 2u_{i,j}^2 + u_{i+1,j}^2)\}.$$

Put $p = h^2/k$ and denote $u_{i,j+1}$ by u_i. The equation can then be written as

$$u_{i-1}^2 - 2(u_i^2 + pu_i) + u_{i+1}^2 + \{u_{i-1,j}^2 - 2(u_{i,j}^2 - pu_{i,j}) + u_{i+1,j}^2\} = 0 \equiv f_i(u_{i-1}, u_i, u_{i+1}).$$

By equation (2.41),

$$\left[\frac{\partial f_i}{\partial u_{i-1}}\epsilon_{i-1} + \frac{\partial f_i}{\partial u_i}\epsilon_i + \frac{\partial f_i}{\partial u_{i+1}}\epsilon_{i+1}\right]_{u_i = V_i} + f_i(V_{i-1}, V_i, V_{i+1}) = 0,$$

hence

$$2V_{i-1}\epsilon_{i-1} - 2(2V_i + p)\epsilon_i + 2V_{i+1}\epsilon_{i+1} + [\{V_{i-1}^2 - 2(V_i^2 + pV_i) + V_{i+1}^2\} + \{u_{i-1,j}^2 - 2(u_{i,j}^2 - pu_{i,j}) + u_{i+1,j}^2\}] = 0, \tag{2.42}$$

where V_i is an approximation to $u_{i,j+1}$.

The problem is symmetric with respect to $x = \frac{1}{2}$. When the V_i are taken equal to $u_{i,0}$, equation (2.42) for $j = 0$ reduces to

$$2u_{i-1,0}\epsilon_{i-1} - 2(2u_{i,0} + p)\epsilon_i + 2u_{i+1,0}\epsilon_{i+1} + \{2u_{i-1,0}^2 - 4u_{i,0}^2 + 2u_{i+1,0}^2\} = 0.$$

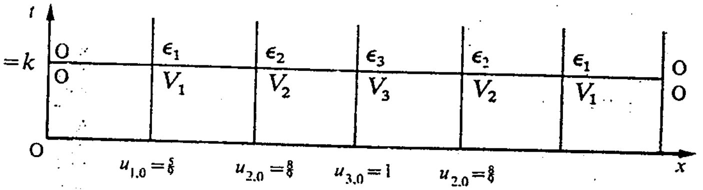


Fig. 2.9

This gives the equations quoted for $p=1$ and the initial values indicated in Fig. 2.9 since $\epsilon_0=0$.

Richtmyer's linearization method

Richtmyer, reference 31, considers the equation

$$\frac{\partial U}{\partial t}=\frac{\partial^2 U^m}{\partial x^2}, \quad m \text{ a positive integer } \geqslant 2,$$

which he approximates by the implicit weighted average difference scheme

$$\frac{1}{k}(u_{i,j+1}-u_{i,j})=\frac{1}{h^2}[\theta\,\delta_x^2(u_{i,j+1}^m)+(1-\theta)\,\delta_x^2(u_{i,j}^m)]. \qquad (2.43)$$

By Taylor's expansion about the point (i, j),

$$u_{i,j+1}^m=u_{i,j}^m+k\frac{\partial u_{i,j}^m}{\partial t}+\dots$$

$$=u_{i,j}^m+k\frac{\partial u_{i,j}^m}{\partial u_{i,j}}\frac{\partial u_{i,j}}{\partial t}+\dots.$$

Hence to terms of order k,

$$u_{i,j+1}^m=u_{i,j}^m+mu_{i,j}^{m-1}(u_{i,j+1}-u_{i,j}),$$

a result which replaces the non-linear unknown $u_{i,j+1}^m$ by an approximation linear in $u_{i,j+1}$.

Putting $\omega_i=u_{i,j+1}-u_{i,j}$ in equation (2.43) leads to

$$\frac{1}{k}\,\omega_i=\frac{1}{h^2}[\theta\,\delta_x^2(u_{i,j}^m+mu_{i,j}^{m-1}\,\omega_i)+(1-\theta)\,\delta_x^2u_{i,j}^m]$$

$$=\frac{1}{h^2}[m\theta\,\delta_x^2u_{i,j}^{m-1}\,\omega_i+\delta_x^2u_{i,j}^m]$$

$$=\frac{1}{h^2}[m\theta(u_{i-1,j}^{m-1}\,\omega_{i-1}-2u_{i,j}^{m-1}\,\omega_i+u_{i+1,j}^{m-1}\,\omega_{i+1}) + (u_{i-1,j}^m-2u_{i,j}^m+u_{i+1,j}^m)],$$

which gives a set of linear equations for the ω_i. The solution at the $(j+1)$th time-level is obtained from $u_{i,j+1}=\omega_i+u_{i,j}$. For known boundary values at $x=0$ and 1, where $Nh=1$ and $r=k/h^2$, the equations in matrix form are for m=2

$$\begin{bmatrix} (1+4r\theta u_{1,j}) & -2r\theta u_{2,j} & & & \\ -2r\theta u_{1,j} & (1+4r\theta u_{2,j}) & -2r\theta u_{3,j} & & \\ & \cdot & & & \\ & & \cdot & & \\ & & & \cdot & \\ & & -2r\theta u_{N-3,j} & (1+4r\theta u_{N-2,j}) & -2r\theta u_{N1,j} \\ & & & -2r\theta u_{N-2,j} & (1+4r\theta u_{N-1,j}) \end{bmatrix} \begin{bmatrix} \omega_1 \\ \omega_2 \\ \\ \\ \\ \omega_{N-2} \\ \omega_{N-1} \end{bmatrix}$$

$$= \begin{bmatrix} -2ru_{1,j} & ru_{2,j} & & \\ ru_{1,j} & -2ru_{2,j} & ru_{3,j} & \\ & \cdot & & \\ & & \cdot & \\ & & \cdot & \\ & & ru_{N-2,j} & -2ru_{N-1,j} \end{bmatrix} \begin{bmatrix} u_{1,j} \\ u_{2,j} \\ \\ \\ \\ u_{N-1,j} \end{bmatrix} + \begin{bmatrix} ru_{0,j}^2+2r\theta u_{0,j}(u_{0,j+1}-u_{0,j}) \\ 0 \\ 0 \\ \cdot \\ \cdot \\ 0 \\ ru_{N,j}^2+2r\theta u_{N,j}(u_{N,j+1}-u_{N,j}) \end{bmatrix} \qquad (2.44)$$

These are easily solved by the algorithm on page 23. The stability of this scheme is considered in Chapter 3, exercise 12, and numerical results for a particular problem are given on page 55.

A three time-level method

Lees, reference 19, considered the non-linear equation

$$b(U)\frac{\partial U}{\partial t}=\frac{\partial}{\partial x}\left\{a(U)\frac{\partial U}{\partial x}\right\}, \quad a(U)>0, \quad b(U)>0, \qquad (2.45)$$

and investigated a difference scheme that

(i) achieved linearity in the unknowns $u_{i,j+1}$ by evaluating all coefficients of $u_{i,j+1}$ at a time-level of known solution values,

(ii) preserved stability by averaging $u_{i,j}$ over three time-levels, and

(iii) maintained accuracy by using central-difference approximations.

EGM 6422

LECTURE 18/19

ELLIPTIC PDES

$$\left.\begin{array}{l}\nabla^2\varphi = 0 \quad (\text{LAPLACE'S EQN}) \\ \nabla^2\varphi = f(x,y)\end{array}\right\} \varphi(x,y)$$

2-D FLUID MOTION φ - velocity vector $= u\underline{i} + v\underline{j}$

φ - stream fn.

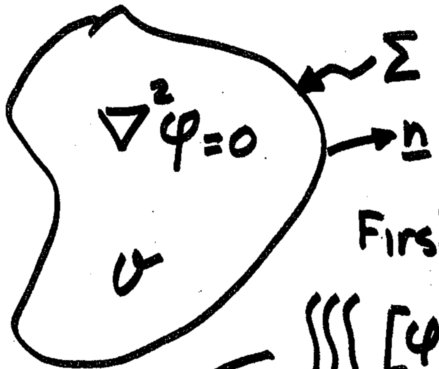
φ - Temperature

φ - torsion fn. $f = -2G\alpha$

$$\underset{\substack{\text{shear}\\\text{modulus}}}{G} \qquad \underset{\substack{\text{angle of}\\\text{twist}}}{\alpha}$$

for solution of laplace's eqn. φ - harmonic fn.

1) If $\varphi(x,y)$ is a harmonic fn. in the region $\mathcal{V}$ which is bdd by the surface Σ, then



$$\oiint_\Sigma \frac{\partial\varphi}{\partial n}\, d\sigma = 0$$

First Green's Theorem φ_1, φ_2 are fns of x,y,z

$$\iiint_{\mathcal{V}} [\varphi_1 \nabla^2\varphi_2 + \nabla\varphi_1 \cdot \nabla\varphi_2]\, d\mathcal{V} = \oiint_\Sigma \underline{n} \cdot \nabla\varphi_2\, d\sigma$$

$$\nabla\varphi = \underline{i}\frac{\partial\varphi}{\partial x} + \underline{j}\frac{\partial\varphi}{\partial y} + \underline{k}\frac{\partial\varphi}{\partial z} \quad ; \text{ if } \varphi_1 = 1 \quad \varphi_2 = \varphi$$

$$\iiint_{\mathcal{V}} \underset{\overset{\|}{0}}{\nabla^2\varphi}\, d\mathcal{V} = \oiint_\Sigma \underline{n} \cdot \nabla\varphi\, d\sigma$$

$$0 = \oiint_\Sigma \underline{n} \cdot \nabla\varphi\, d\sigma = \oiint_\Sigma \frac{\partial\varphi}{\partial n}\, d\sigma$$

if $\nabla^2\varphi = f$ $\qquad \iiint \nabla^2\varphi \, dv = \oiint_\Sigma \frac{\partial\varphi}{\partial n} d\sigma$

$$\iiint f \, dv = \oiint_\Sigma \frac{\partial\varphi}{\partial n} d\sigma$$

if $\nabla^2\varphi = 0$ & $\left.\frac{\partial\varphi}{\partial n}\right|_\Sigma = g$ (2nd type of BC)

$$\iiint_V \nabla^2\varphi \, dv = \oiint_\Sigma \frac{\partial\varphi}{\partial n} d\sigma$$

$$\overset{\shortparallel}{0} \qquad = \oiint_\Sigma g \, d\sigma$$ there are no source terms within the volume.

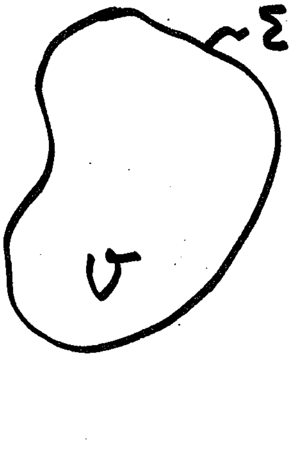
3 types of BC's $\quad \left.\varphi\right|_\Sigma = g$ (Dirichlet BC - 1st type)

$\left.\frac{\partial\varphi}{\partial n}\right|_\Sigma = g$ (Neumann BC - 2nd type)

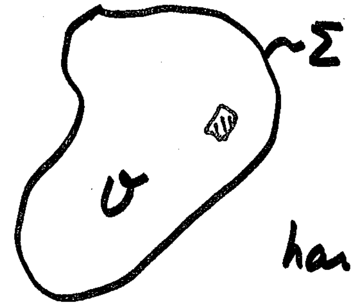
nonunique <u>unless</u> $\oiint_\Sigma g \, d\sigma = 0$

$\left.\left(a\varphi + b\frac{\partial\varphi}{\partial n}\right)\right|_\Sigma = g$ (Robin, Churchill - 3rd type)

2) IF a fn $\varphi(x,y)$ is continuous and defined in the region $V+\Sigma$ and φ satisfies $\nabla^2\varphi = 0$ in the interior of V, then it assumes its maximum & minimum on the boundary Σ: (principle of the maximum).



3) if u & U are continuous in $V+\Sigma$ and are harmonic ($\nabla^2 u = \nabla^2 U = 0$) in V and $u \le U$ on Σ, then $u \le U$ for all points in V.



4) if u and U are continuous in $V+\Sigma$ and are harmonic in V for which $|u| \le U$ on Σ then $|u| \le U$ for all points in V.

5) If we have 3 harmonic fns. $-U, u, U$ so that

$$-U \le u \le U \text{ on } \Sigma, \text{ then } -U \le u \le U \text{ in } V.$$

this is an outcome of ④

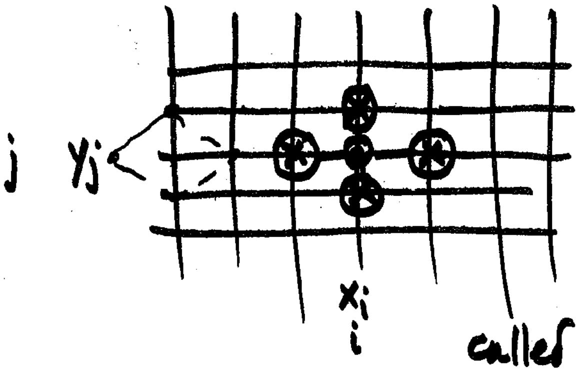
$$\nabla^2 T = f \quad (\text{S.S. heat transfer problems})$$

● $T(x,y) = T$

& let's write the FD equation at x_i, y_j $\quad x_i = i\Delta x,\ y_j = j\Delta y$

& CD in x & y. Look at $\Delta x = \Delta y = h$

$$T_{ij}^{(n+1)} = \frac{1}{4}\left(T_{i+1j}^{(n)} + T_{i-1j}^{(n)} + T_{ij+1}^{(n)} + T_{ij-1}^{(n)} - h^2 f_{ij}\right) \qquad f_{ij} = f(x_i, y_j)$$



To solve: first must assume some initial values on the interior

: iterate one pt. at a time

: iteration will converge

called Liebmann method if $f_{ij} = 0$

Jacobi type iteration

speed up convergence: 1) replace immediately

$$T_{ij}^{(n+1)} = \frac{1}{4}\left(T_{i-1j}^{(n+1)} + T_{i+1j}^{(n)} + T_{ij+1}^{(n)} + T_{ij-1}^{(n+1)} - h^2 f_{ij}\right)$$ Gauss-Seidel type method

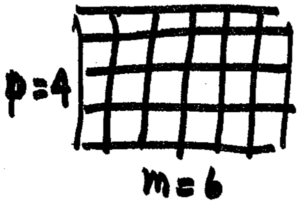
2) use successive over relaxation Southwell's method.

$$T_{ij}^{(n+1)} = T_{ij}^{(n)} + \frac{\omega}{4}\left[T_{i-1j}^{(n+1)} + T_{i+1j}^{(n)} + T_{ij-1}^{(n+1)} + T_{ij+1}^{(n)} - 4T_{ij}^{(n)} - h^2 f_{ij}\right]$$

GS eq = Southwell eqn when $\omega = 1$

$$\omega = \frac{8 - 4\sqrt{4-\alpha^2}}{\alpha^2} \qquad \alpha = \cos(\pi/m) + \cos(\pi/p)$$

$1 \leq \omega \leq 2$ m & p are the # of intervals in the horiz & vert. side of a rectangular region



To prove convergence of the method $e_{ij}^{(n)} = \underset{\uparrow\text{numerical}}{T_{ij}^{(n)}} - \overset{\swarrow\text{actual}}{T_{ij}}$

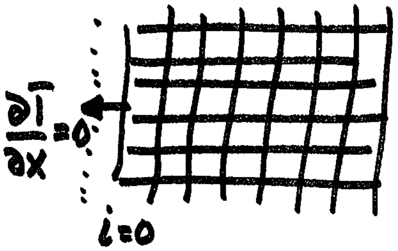
for Liebmann method.

$$e_{ij}^{(n+1)} = \frac{1}{4}\left[e_{ij+1}^{(n)} + e_{ij-1}^{(n)} + e_{i+1j}^{(n)} + e_{i-1j}^{(n)}\right] \overbrace{- Lu_{ij} + h^2 f_{ij}}^{=0}$$

$$|e_{ij}^{(n+1)}| \leq \frac{1}{4}\left\{|e_{ij+1}^{(n)}| + |e_{ij-1}^{(n)}| + |e_{i+1j}^{(n)}| + |e_{i-1j}^{(n)}|\right\} \leq \frac{1}{4} \cdot 4 \max e^{(n)}$$

$|e_{ij}^{(n+1)}| \leq e_{max}^{(n)}$ this shows error doesn't grow but is not a proof of convergence

if $e^{(n+1)} = Ke^{(n)} = K^2 e^{(n-1)} = \ldots . e^{(0)} K^{(n+1)}$ the for convergence $|K| < 1$. if $K = \left(1 - \frac{1}{4^m}\right)$ m = # of pts in one direction then we get convergence



$$T_{ij} = \frac{1}{4}\left[T_{i+1j} + T_{i-1j} + T_{ij+1} + T_{ij-1} - h^2 f_{ij}\right]$$

$$O(\Delta x^2, \Delta y^2)$$

$$\left.\frac{\partial T}{\partial x}\right|_{i=0,j} = CD \text{ in } x = \frac{T_{1j} - T_{-1j}}{2\Delta x} = 0$$

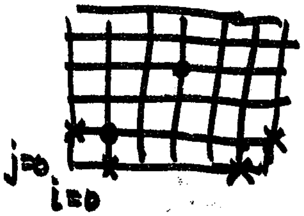
$$T_{1j} = T_{-1j} \qquad \left(T_{-1j} = \overset{*}{T_{1j}} - 2q\Delta x\right)^{=q}$$

$$T_{0j} = \frac{1}{4}\left[T_{1j} + T_{-1j} + T_{0j+1} + T_{0j-1} - h^2 f_{0j}\right]$$

$$= \frac{1}{4}\left[2T_{1j} + T_{0j+1} + T_{0j-1} - h^2 f_{0j}\right]$$

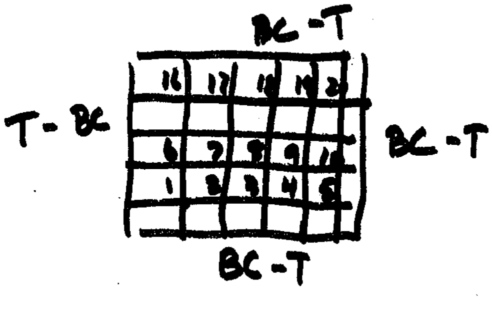
The other way to solve is by matrix methods.

$$\nabla^2 T_{ij} = 0 \qquad \left[T_{i+1j} + T_{i-1j} + T_{ij+1} + T_{ij-1} - 4T_{ij}\right] = h^2 f_{ij}$$



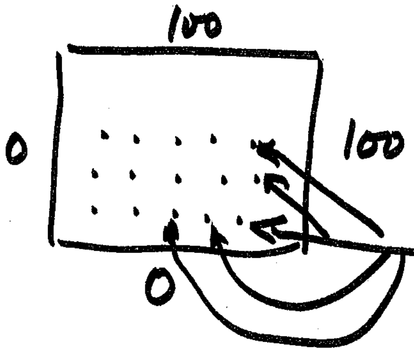
$$\left[T_{21} + \overset{BC}{T_{01}} + T_{12} + \overset{BC}{T_{10}} - 4T_{11}\right] = h^2 f_{11} \qquad i=1, j=1$$

$$\left[T_{31} + T_{11} + T_{22} + \overset{BC}{T_{20}} - 4T_{21}\right] = h^2 f_{21} \qquad i=2, j=1$$



$$\begin{bmatrix} -4 & 1 & \cdots & 1 & & & \\ 1 & -4 & 1 & \cdots & 1 & & \\ & 1 & -4 & 1 & \cdots & 1 & \\ & & \ddots & \ddots & \ddots & & \ddots \\ 1 & & & 1 & -4 & 1 & & 1 \\ & 1 & & & \ddots & \ddots & \ddots & \\ & & 1 & & & \ddots & \ddots & \\ & & & 1 & \cdots & 1 & 1 & -4 \end{bmatrix} \begin{pmatrix} T_1 \\ T_2 \\ \vdots \\ \vdots \\ T_{20} \end{pmatrix} = \begin{pmatrix} h^2 f_{11} + BC + BC \\ h^2 f_{21} + BC \\ \vdots \\ h^2 f_{51} + BC + BC \\ h^2 f_{12} + BC \\ \vdots \\ h^2 f_{54} + BC + BC \end{pmatrix}$$

how do you start the iterative process for Liebmann or G-S or Southwell



$$\Sigma(100+100+0+0) = \frac{200}{4} = 50$$

EGM 6422

LECTURE 19/20

EGM 6422

LECTURE 18/19

EGM 6422

ADV. ANAL. OF MECH. SYS.

10/25/05

ELLIPTIC PDE's

$$\left.\begin{array}{l}\nabla^2\varphi = 0 \quad (\text{LAPLACE's EQN}) \\ \nabla^2\varphi = f(x,y)\end{array}\right\} \varphi(x,y)$$

2-D FLUID MOTION φ - velocity vector $= u\underline{i} + v\underline{j}$

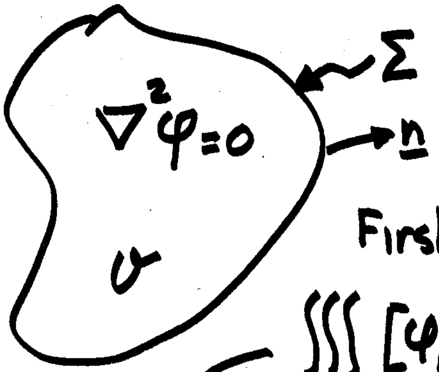
φ - stream fn.

φ - Temperature

φ - torsion fn. $f = -2\underset{\substack{\uparrow \\ \text{shear} \\ \text{modulus}}}{G}\underset{\substack{\nwarrow \\ \text{angle of} \\ \text{twist}}}{\alpha}$

for solution of laplace's eqn. φ - harmonic fn.

1) If $\varphi(x,y)$ is a harmonic fn. in the region υ which is bdd by the surface Σ, then



$$\oiint_\Sigma \frac{\partial\varphi}{\partial n}\, d\sigma = 0$$

First Green's Theorem φ_1, φ_2 are fns of x, y, z

$$\iiint_\upsilon [\varphi_1\nabla^2\varphi_2 + \nabla\varphi_1\cdot\nabla\varphi_2]\, d\upsilon = \oiint_\Sigma \underline{n}\cdot\nabla\varphi_2\, d\sigma$$

$\nabla\varphi = \underline{i}\frac{\partial\varphi}{\partial x} + \underline{j}\frac{\partial\varphi}{\partial y} + \underline{k}\frac{\partial\varphi}{\partial z}$; if $\varphi_1 = 1$ $\varphi_2 = \varphi$

$$\iiint_\upsilon \underset{\substack{\| \\ 0}}{\nabla^2\varphi}\, d\upsilon = \oiint_\Sigma \underline{n}\cdot\nabla\varphi\, d\sigma$$

$$0 = \oiint_\Sigma \underline{n}\cdot\nabla\varphi\, d\sigma = \oiint_\Sigma \frac{\partial\varphi}{\partial n}\, d\sigma$$

if $\nabla^2\varphi = f$ $$\iiint \nabla^2\varphi \, dV = \oiint_\Sigma \frac{\partial\varphi}{\partial n} d\sigma$$

$$\iiint f \, dV = \oiint_\Sigma \frac{\partial\varphi}{\partial n} d\sigma$$

---

if $\nabla^2\varphi = 0$ & $\left.\frac{\partial\varphi}{\partial n}\right|_\Sigma = g$ (2nd type of BC)

$$\iiint_V \nabla^2\varphi \, dV = \oiint_\Sigma \frac{\partial\varphi}{\partial n} d\sigma$$

$$\overset{\shortparallel}{0} \quad = \oiint_\Sigma g \, d\sigma$$ there are no source terms within the volume.

3 types of BC's $\left.\varphi\right|_\Sigma = g$ (Dirichlet BC - 1st type)

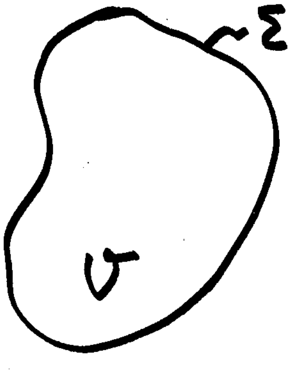
$\left.\frac{\partial\varphi}{\partial n}\right|_\Sigma = g$ (Neumann BC - 2nd type)

nonunique <u>unless</u> $\oiint_\Sigma g \, d\sigma = 0$

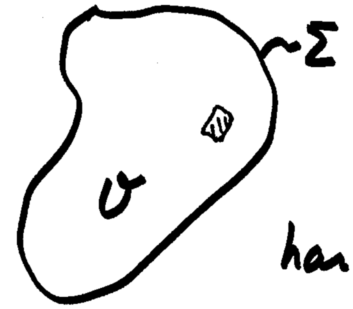
$\left.\left(a\varphi + b\frac{\partial\varphi}{\partial n}\right)\right|_\Sigma = g$ (Robin, Churchill - 3rd type)

---

2) IF a fn $\varphi(x,y)$ is continuous and defined in the region $V+\Sigma$ and φ satisfies $\nabla^2\varphi = 0$ in the interior of V, then it assumes its maximum & minimum on the boundary Σ: (principle of the maximum).



3) if u & U are continuous in $\mathcal{V}+\Sigma$ and are harmonic $(\nabla^2 u = \nabla^2 U = 0)$ in $\mathcal{V}$ and $u \le U$ on Σ, then $u \le U$ for all points in $\mathcal{V}$.



4) if u and U are continuous in $\mathcal{V}+\Sigma$ and are harmonic in $\mathcal{V}$ for which $|u| \le U$ on Σ then $|u| \le U$ for all points in $\mathcal{V}$.

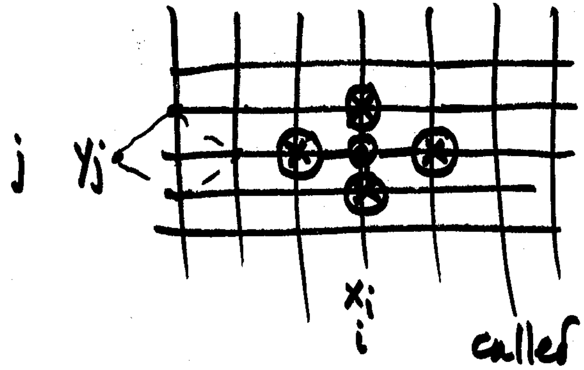
5) If we have 3 harmonic fns. $-U, u, U$ so that $-U \le u \le U$ on Σ, then $-U \le u \le U$ in $\mathcal{V}$.

this is an outcome of ④

$\nabla^2 T = f$ (S.S. heat transfer problems)

- $T(x,y) = T$
- let's write the FD equation at x_i, y_j $\quad x_i = i\Delta x,\ y_j = j\Delta y$
- CD in x & y. Look at $\Delta x = \Delta y = h$

$$T_{ij}^{(n+1)} = \frac{1}{4}\left(T_{i+1j}^{(n)} + T_{i-1j}^{(n)} + T_{ij+1}^{(n)} + T_{ij-1}^{(n)} - h^2 f_{ij}\right) \qquad f_{ij} = f(x_i, y_j)$$



To solve: first must assume some initial values on the interior
: iterate one pt. at a time
: iteration will converge

called Liebmann method if $f_{ij}=0$
Jacobi type iteration

speed up convergence: 1) replace immediately

$$T_{ij}^{(n+1)} = \frac{1}{4}\left(T_{i-1j}^{(n+1)} + T_{i+1j}^{(n)} + T_{ij+1}^{(n)} + T_{ij-1}^{(n+1)} - h^2 f_{ij}\right)$$

Gauss-Seidel type method

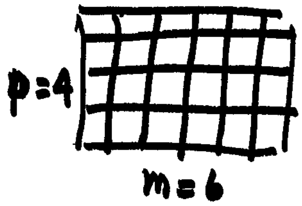
2) use successive over relaxation Southwell's method.

$$T_{ij}^{(n+1)} = T_{ij}^{(n)} + \frac{\omega}{4}\left[T_{i-1j}^{(n+1)} + T_{i+1j}^{(n)} + T_{ij-1}^{(n+1)} + T_{ij+1}^{(n)} - 4T_{ij}^{(n)} - h^2 f_{ij}\right]$$

GS eq = Southwell eqn when $\omega = 1$

$$\omega = \frac{8 - 4\sqrt{4-\alpha^2}}{\alpha^2} \qquad \alpha = \cos(\pi/m) + \cos(\pi/p)$$

$1 \leq \omega \leq 2$ m & p are the # of intervals in the horiz & vert. side of a rectangular region



To <u>prove convergence</u> of the method $e_{ij}^{(n)} = \underset{\uparrow \text{numerical}}{T_{ij}^{(n)}} - \overset{\swarrow \text{actual}}{T_{ij}}$

for Liebmann method.

$$e_{ij}^{(n+1)} = \frac{1}{4}\left[e_{ij+1}^{(n)} + e_{ij-1}^{(n)} + e_{i+1j}^{(n)} + e_{i-1j}^{(n)}\right]$$

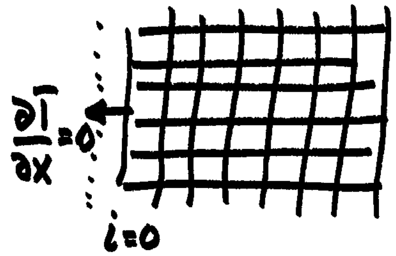
$$|e_{ij}^{(n+1)}| \leq \frac{1}{4}\left\{|e_{ij+1}^{(n)}| + |e_{ij-1}^{(n)}| + |e_{i+1j}^{(n)}| + |e_{i-1j}^{(n)}|\right\} \leq \frac{1}{4} \cdot 4 \max e^{(n)}$$

$$|e_{ij}^{(n+1)}| \leq e_{\max}^{(n)}$$

this shows error doesn't grow but is not a proof of convergence

if $e^{(n+1)} = Ke^{(n)} = K^2 e^{(n-1)} = \ldots . e^{(0)} K^{(n+1)}$ the for convergence $|K| < 1$. if $K = \left(1 - \frac{1}{4^m}\right)$ m = # of pts in one direction then we get convergence

IF we have Neumann Type BCs



$$T_{ij} = \frac{1}{4}\left[T_{i+1j} + T_{i-1j} + T_{ij+1} + T_{ij-1} - h^2 f_{ij}\right]$$

$$\left.\frac{\partial T}{\partial x}\right|_{i=0,j} = \text{CD in } x = \frac{T_{1j} - T_{-1j}}{2\Delta x} = 0 \qquad O(\Delta x^2; \Delta y^2)$$

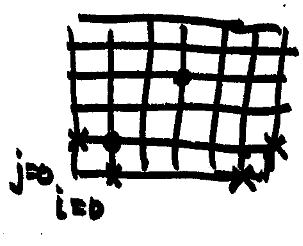
$$T_{1j} = T_{-1j} \qquad \left(T_{-1j} = T_{1j}^{*} - 2q\Delta x\right)^{=q}$$

$$T_{0j} = \frac{1}{4}\left[T_{1j} + T_{-1j} + T_{0j+1} + T_{0j-1} - h^2 f_{0j}\right]$$

$$= \frac{1}{4}\left[2T_{1j} + T_{0j+1} + T_{0j-1} - h^2 f_{0j}\right]$$

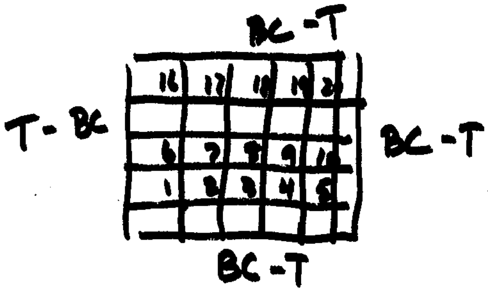
The other way to solve is by matrix methods.

$$\nabla^2 T_{ij} = 0 \qquad \left[T_{i+1j} + T_{i-1j} + T_{ij+1} + T_{ij-1} - 4T_{ij}\right] = h^2 f_{ij}$$



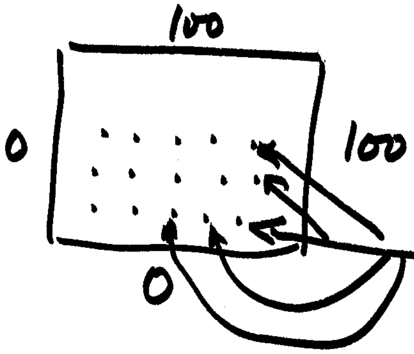
$$\left[T_{21} + \overset{BC}{T_{01}} + T_{12} + \overset{BC}{T_{10}} - 4T_{11}\right] = h^2 f_{11} \qquad i=1, j=1$$

$$\left[T_{31} + T_{11} + T_{22} + \overset{BC}{T_{20}} - 4T_{21}\right] = h^2 f_{21} \qquad i=2, j=1$$



$$\begin{bmatrix} -4 & 1 & \cdots & 1 & & & \\ 1 & -4 & 1 & \cdots & 1 & & \\ & 1 & -4 & 1 & \cdots & 1 & \\ 1 & & \ddots & \ddots & \ddots & & \ddots \\ & 1 & & 1 & -4 & 1 & 1 \\ & & 1 & & \ddots & \ddots & \ddots \\ & & & 1 \cdots 1 & & 1 & -4 \end{bmatrix} \begin{pmatrix} T_1 \\ T_2 \\ \vdots \\ T_{20} \end{pmatrix} = \begin{pmatrix} h^2 f_{11} + BC + BC \\ h^2 f_{21} + BC \\ \vdots \\ h^2 f_{51} + BC + BC \\ h^2 f_{12} + BC \\ \vdots \\ h^2 f_{54} + BC + BC \end{pmatrix}$$

how do you start the iterative process for Liebmann or G-S or Southwell



$$\Sigma(100+100+0+0) = \frac{200}{4} = 50$$

where μ and ν are continuous functions, represents a harmonic function. That is, since the integrands including their derivatives are continuous throughout with the exception of the boundary Σ, the derivatives of arbitrary order of (4-2.12) can be formed by differentiation under the integral signs. Moreover, since the functions

$$\frac{1}{r_{MP}} \quad \text{and} \quad \frac{\partial}{\partial n_P}\left(\frac{1}{r_{MP}}\right) = \frac{\partial}{\partial \xi}\left(\frac{1}{r}\right)\cos\alpha_P + \frac{\partial}{\partial \eta}\left(\frac{1}{r}\right)\cos\beta_P + \frac{\partial}{\partial \zeta}\left(\frac{1}{r}\right)\cos\gamma_P$$

satisfy Laplace's differential equation with respect to the variable points $M(x, y, z)$, the functions (4-2.12) by the generalized superposition principle (see the lemma on page 63) similarly satisfy Laplace's differential equation with respect to x, y, z.

Hence we arrive at an important conclusion: Every harmonic function, in the interior of the region in which it is harmonic, possesses derivatives of all orders.[79] Further we note that a harmonic function in a region T at each point M_0 in T is analytic (i.e., it can be developed in a power series). This assertion also results from the integral representation (4-2.11).

Corresponding formulas hold also for harmonic functions of two independent variables. Let S be any region in the x, y plane and C its boundary curve. Further let n be the direction of the exterior normal (with respect to S) to this curve. Then if $v = \ln(1/r_{M_0P})$ is introduced where $r_{M_0P} = \sqrt{(x-x_0)^2+(y-y_0)^2}$ is the distance from a point M_0 in the interior of S, then, by similar considerations as above, instead of (4-2.10) we arrive at the expression

$$2\pi u(M_0) = -\int_C\left[u\frac{\partial}{\partial n}\left(\ln\frac{1}{r_{M_0P}}\right) - \ln\frac{1}{r_{M_0P}}\frac{\partial u}{\partial n}\right]ds_P - \int_S\int \ln\frac{1}{r}\,\Delta u\,dS\,, \qquad (4\text{-}2.12')$$

where M_0 is an arbitrary fixed point in the region S.

For a harmonic function $u(M)$ it follows that

$$u(M_0) = \frac{1}{2\pi}\int_C\left[\ln\frac{1}{r}\frac{\partial u}{\partial n} - u\frac{\partial}{\partial n}\left(\ln\frac{1}{r}\right)\right]ds\,. \qquad (4\text{-}2.12'')$$

2. Some fundamental properties of harmonic functions

Harmonic functions possess the following important properties:

1. If $v(M)$ is a harmonic function in a region T which is bounded by the surface Σ, then

$$\iiint \nabla\cdot\nabla v\,d\tau = \iint_S \nabla v\cdot n\,d\sigma = \int_S\int \frac{\partial v}{\partial n}\,d\sigma = 0\,, \qquad \nabla v\cdot n\,d\sigma \qquad (4\text{-}2.13)$$

where S is an arbitrary closed surface which lies entirely in the region T.

If in the first Green's formula (4-2.3) an arbitrary harmonic function v ($\Delta v = 0$) is introduced and $u \equiv 1$ then formula (4-2.13) follows. From (4-2.13) we conclude that the second boundary-value problem ($\Delta u = 0$ in T; $\partial u/\partial n = f|_\Sigma$) possesses a solution only if

$$\int_\Sigma\int f\,d\sigma = 0\,. \qquad \text{divergence free.}$$

This property of a harmonic function can be interpreted as a condition that no sources exist in the region T.

2. If $u(M)$ is a harmonic function in a region T, and M_0 is a point lying in the interior of T, then

$$u(M_0) = \frac{1}{4\pi a^2}\int_{\Sigma_a}\int u\,d\sigma\,, \qquad (4\text{-}2.14)$$

where Σ_a is a spherical surface with radius a and center point M_0 which lies entirely in the region T. This property is stated by the average-value theorem, which reads:

Theorem. The value of a harmonic function at any point M_0 is equal to the average value of the function on an arbitrary spherical surface Σ_a with center at M_0, provided Σ_a lies entirely in the region in which $u(M)$ is harmonic.

For the proof of this proposition we apply formula (4-2.11) to a sphere K_a with center M_0 and surface Σ_a:

$$4\pi u(M_0) = -\int_{\Sigma_a}\int\left[u\frac{\partial}{\partial n}\left(\frac{1}{r}\right) - \frac{1}{r}\frac{\partial u}{\partial n}\right]d\sigma\,. \qquad 2\pi u(M_0) = \int_C \ln\frac{1}{r}\frac{\partial u}{\partial n} - u\frac{\partial}{\partial n}((\ldots$$

If we bear in mind that

$$\frac{1}{r} = \frac{1}{a} \quad \text{on } \Sigma_a \quad \text{and} \quad \int_{\Sigma_a}\int \frac{\partial u}{\partial n}\,d\sigma = 0 \qquad + \frac{u}{r}$$

as well as

$$\frac{\partial}{\partial n}\left(\frac{1}{r}\right)\bigg|_{\Sigma_a} = \frac{\partial}{\partial r}\left(\frac{1}{r}\right)\bigg|_{r=a} = -\frac{1}{a^2}$$

(the direction of the exterior normal to Σ_a coincides with the direction of increasing radius) then we obtain

$$u(M_0) = \frac{1}{4\pi a^2}\int_{\Sigma_a}\int u\,d\sigma\,,$$

which was to be proved.[80]

For two independent variables the analogous theorem is valid:

$$\iiint_V [\varphi_1\nabla^2\varphi_2 + \nabla\varphi_1\cdot\nabla\varphi_2]\,d\tau = \oint\!\!\oint_S \underline{n}\cdot\varphi_1\nabla\varphi_2\,d\sigma$$

$$\text{if } \varphi_1 = 1 \qquad \iiint \nabla^2\varphi_2\,d\tau = \oint\!\!\oint_S \underline{n}\cdot\nabla\varphi_2\,d\sigma$$

---

[79] If for a function u which is harmonic in T the condition that it and its first derivatives are continuous on the boundary Σ is not fulfilled, the theorem still remains valid. By this we mean that each point M is enclosed by a region, including its boundary, lying in the interior of T.

[80] For the proof of this theorem we have used Eq. (4-2.13). This, however, assumes that the derivatives also exist on the spherical surface. If a function $u(M)$, which is continuous in the closed region $T + \Sigma$, satisfies the equation $\Delta u = 0$ only for interior points of T, then this conclusion would not be correct for a sphere Σ_{a_0}, which borders the boundary Σ. However, if the theorem for each $a < a_0$ is true, then as $a \to a_0$ we obtain

$$u(M_0) = \frac{1}{4\pi a_0^2}\int_{\Sigma_{a_0}}\int u(M)\,d\sigma\,.$$

Solutions w dirichlet BC's are unique
Solutions w Neuman BC's are not unique

if Bdry data is discontinuous they are <u>not</u> propagated into the solution.

$$u(M_0) = \frac{1}{2\pi a}\int_{C_a} u\,ds\,, \qquad (4\text{-}2.15)$$

where C_a is a circle of radius a about the center M_0.

3. If a function $u(M)$ is continuous and defined in a closed region $T + \Sigma$ and satisfies Laplace's differential equation $\Delta u = 0$ in the interior of T, then it assumes its maximum and its minimum on the boundary Σ (principle of the maximum).

If the function $u(M)$ were to assume its maximum at an interior point M_0 in T, then $u_0 = u(M_0) \geqq u(M)$ for each M in T.

Now enclose the point M_0 with a sphere of radius ρ whose surface Σ_ρ lies entirely within T. Since by hypothesis $u(M_0)$ is the maximum of the function $u(M)$ in $T + \Sigma$, then $u|_\Sigma \leqq u(M_0)$. Therefore, by use of the average-value formula (4-2.14), provided that everywhere under the integral signs we replace $u(M)$ by $u(M_0)$, we obtain

$$u(M_0) = \frac{1}{4\pi\rho^2}\int_{\Sigma_\rho}\!\!\int u(M)\,d\sigma_M \leqq \frac{1}{4\pi\rho^2}\int_{\Sigma_\rho}\!\!\int u(M_0)\,d\sigma = u(M_0)\,. \qquad (4\text{-}2.16)$$

Now if we assume that at least one point M exists on Σ_ρ such that the inequality $u(M) < u(M_0)$ is valid, then obviously in the last formula the inequality sign must hold, which in turn implies a contradiction; consequently on the entire surface Σ_ρ we must have $u(M) \equiv u(M_0)$.

If ρ_0^m is the minimal distance of the point M_0 from the boundary Σ_ρ, then $u(M) \equiv u(M_0)$ is also valid for points belonging to $\Sigma_{\rho_0^m}$. Hence, because of continuity, it also follows that at those points M^* which belong to the intersection of $\Sigma_{\rho_0^m}$ and Σ the relation $u(M^*) = u(M_0)$ is valid. Therefore our theorem is proved; and the last conclusion shows that the maximum $u(M_0)$ is assumed at least at one point on the boundary.

It is easily seen that $u(M) \equiv u(M_0)$ must be valid in the entire region when the region T is connected and if at least at one interior point M_0 the maximum is assumed.

To demonstrate the above, we select another arbitrary point $M^{(0)}$ in T and connect it with M_0 by the polygonal line L (Figure 46), whose length is designated by l. Let M_1 be the last current point of L through the spherical surface $\Sigma_{\rho_0^m}$. At this point, then, $u(M_1) = u(M_0)$ is still valid. Now we enclose this point by the spherical surface $\Sigma_{\rho_1^m}$, where ρ_1^m is the minimal distance of the point M_1 from the boundary. We obtain another such point M_2 as the last current point of L through the spherical surface $\Sigma_{\rho_1^m}$. By this procedure, we arrive after a finite number of steps (the number p of necessary steps is certainly not larger than $1/\rho^{(m)}$ if $\rho^{(m)}$ denotes the minimum distance between L and Σ) at a spherical surface which contains the point $M^{(0)}$. From this it follows that $u(M^{(0)}) = u(M_0)$. Because of the arbitrary choice of the point $M^{(0)}$ and the continuity of $u(M)$ in the closed region $T + \Sigma$ we conclude that $u(M) \equiv u(M_0)$ holds everywhere (including the boundary). Of all harmonic functions, therefore, only the constant functions can assume their maximum at an interior point.

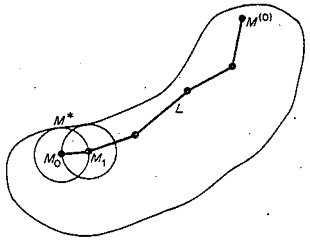


FIG. 46.

The corresponding statement is also true for the minimum.

Conclusion 1:

If u and U are continuous in $T + \Sigma$ and are harmonic functions in T for which

$$u \leqq U \quad \text{on } \Sigma$$

then also $u \leqq U$ holds at all points in the interior of T.

The function $U - u$ is therefore continuous in $T + \Sigma$, harmonic in T; hence

$$U - u \geqq 0 \quad \text{on } \Sigma\,.$$

Consequently, according to the principle of the maximum we must have

$$U - u \geqq 0$$

at all points in the interior of T—precisely our assertion.

Conclusion 2:

If u and U are continuous in the region $T + \Sigma$ and are harmonic functions in T for which

$$|u| \leqq U \quad \text{on } \Sigma$$

then also

$$|u| \leqq U$$

at all points in the interior of T.

From the above assumptions it follows that the three harmonic functions $-U$, u, and U satisfy the relation

$$-U \leqq u \leqq U \quad \text{on } \Sigma\,.$$

However, it follows by a twice-repeated application of Conclusion 1 that

$$-U \leqq u \leqq U$$

at all points in the interior of T, or

$$|u| \leqq U$$

in the interior of T.

3. Uniqueness and stability of the solution of the first boundary-value problem

Let us consider a region T which is bounded by the surface Σ. Then the first boundary-value problem for Laplace's differential equation in the region T reads as follows:

arise from equilib + steady state problems.

NUMERICAL SOLUTIONS OF ELLIPTIC PDE'S

LOOK AT 2-D FLUID MOTION $\underset{\sim}{V}$ - veloc vector $u\underset{\sim}{i} + v\underset{\sim}{j} = \nabla \times \underset{\sim}{\psi}$

$$\nabla^2 \psi = f(x,y) \qquad \text{if } \underset{\sim}{V} = \nabla \times \underset{\sim}{\psi}$$

ψ - is the stream fn. f is the vorticity vector. IF NO VORTICITY

IRROTATIONAL FLOW IS

$$\nabla^2 \psi = 0$$

$\nabla^2 T = 0$ is steady state heat equation; if heat source exists $\nabla^2 T = f(x,y)$. also torsion in a rectangular bar gives rise to $\nabla^2 \phi + 2G\alpha = 0$

torsion function angle of twist

using centred difference $T_{xx} = \dfrac{T_{i+1,j} - 2T_{i,j} + T_{i-1,j}}{\Delta x^2}$ $O(\Delta x^2)$ is error

$$T_{yy} = \frac{T_{i,j+1} - 2T_{ij} + T_{ij-1}}{(\Delta y)^2} \qquad O(\Delta y^2) \text{ is error}$$

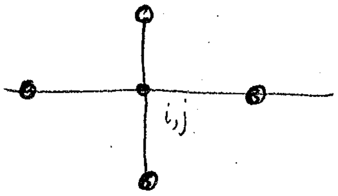
if $\Delta x = \Delta y = h$

$$T_{ij} = \frac{1}{4}\left(T_{i-1,j} + T_{i+1,j} + T_{ij-1} + T_{ij+1} - \cancel{4} h^2 f_{ij}\right)$$

$f_{ij} = f(x_i, y_j)$ $\qquad T_{ij} = T(x_i, y_j)$

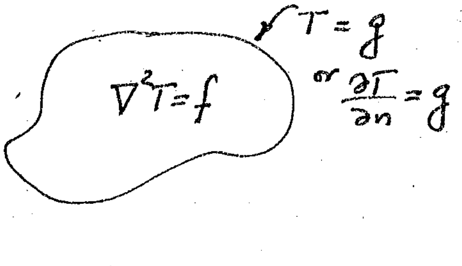
$x_i = i\Delta x = ih$ $\qquad y_j = j\Delta y = jh$

$$\nabla^2 T = \frac{1}{6h^2} \cdot \begin{vmatrix} 1 & 4 & 1 \\ 4 & -20 & 4 \\ 1 & 4 & 1 \end{vmatrix} T_{i,j}$$



error is $O(h^2)$

local error is $O(h^6)$ if $\Delta x = \Delta y$



Liebmann is iterative scheme

Jacobi is non replacing iterative

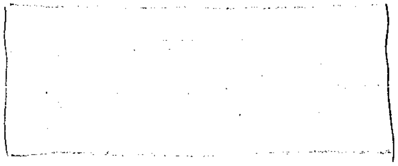
As you remember relaxation method

1) Assume some initial values
2) iterate one pt at a time
3) iterations converge $\qquad T_{ij}^{(n+1)} = \frac{1}{4}\left(T_{i-1,j}^{(n)} + T_{i+1,j}^{(n)} + T_{ij-1}^{(n)} + T_{ij+1}^{(n)} - \cancel{4} h^2 f_{ij}\right)$

Jacobi | Liebmann's method

Liebmann's Method

$$T_{ij}^{(n+1)} = \cancel{T_{ij}^{(n)}} \ \frac{1}{4}\left[T_{ij+1}^{n} + T_{ij-1}^{n} + T_{i+1j}^{n} + T_{i-1j}^{n}\right]$$



$$T_{ij}^{n+1} = \frac{1}{4}\left(T_{i-1,j}^{(n+1)} + T_{i+1,j}^{n} + T_{ij-1}^{(n+1)} + T_{ij+1}^{(n)} - h^2 f_{ij}\right)$$

Gauss-Seidel is replacing-iterative / Southwell is relaxation method.

to speed up convergence 1) replace immediately

2) use successive over relaxation

$$T_{ij}^{(n+1)} \underset{n}{=} T_{ij}^{(n)} + \frac{\omega}{4}\left[T_{i-1,j}^{(n+1)} + T_{i+1,j}^{(n)} + T_{i,j-1}^{(n+1)} + T_{ij+1}^{(n)} - 4T_{ij}^{(n)} - h^2 f_{ij}\right]$$

if $\omega = \dfrac{8 - 4\sqrt{4-\alpha^2}}{\alpha^2}$ $\qquad \alpha = \cos(\pi/m) + \cos(\pi/n)$

$1 \leq \omega \leq 2$

- where m & n are the # of intervals in the horiz & vertical side of a rectangular region & $\Delta x = \Delta y = h$

how do we prove convergence define $e_{ij}^{(n)} = T_{ij}^{(n)} - T_{ij}$

actual

at n^{th} step.

<u>Look at Liebmann's method:</u>

Look at $e_{ij}^{n+1} = \frac{1}{4}\left[e_{ij+1}^{n} + e_{ij-1}^{n} + e_{i+1,j}^{n} + e_{i-1,j}^{n}\right]$

$$|\quad| \leq \frac{1}{4}\left[|\quad| + |\quad| + |\quad| + |\quad|\right] \leq \frac{1}{4} \cdot 4 \max e^{n}$$

$$e^{n+1} \leq e^{n}$$

- this however doesn't prove convergence—only that the errors do not ↑

$$\text{if} \quad e^{n+1} = Ke^{n} = K^2 e^{n-1} = K\left[Ke^{n-1}\right] = \cdots = K^{n+1} e^{0}$$

if $K < 1$ we get convergence.

$$K = \left(1 - \frac{1}{4^m}\right) \qquad , \quad m = \text{\# of pts in one directi}$$

THE ABOVE ASSUMES THAT BC IS GIVEN ON T

WHAT IF BC IS GIVEN ON $\dfrac{\partial T}{\partial x}$ or $\dfrac{\partial T}{\partial y}$.

① USE CENTRAL DIFF SO THAT $\dfrac{\partial T}{\partial x} = 0 \Longrightarrow T_{i+1,j} = T_{i-1,j}$

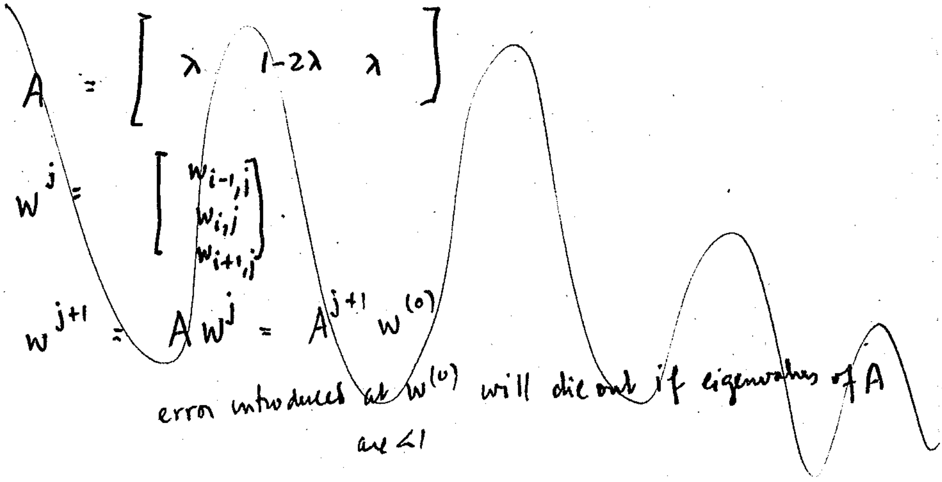
② place an extra column past actual boundary of region explicit scheme

$$A = \begin{bmatrix} \lambda & 1-2\lambda & \lambda \end{bmatrix}$$

$$w^j = \begin{bmatrix} w_{i-1,j} \\ w_{i,j} \\ w_{i+1,j} \end{bmatrix}$$

$$w^{j+1} = A w^j = A^{j+1} w^{(0)}$$

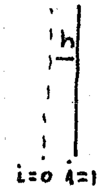
error introduced at $w^{(0)}$ will die out if eigenvalues of A are <1



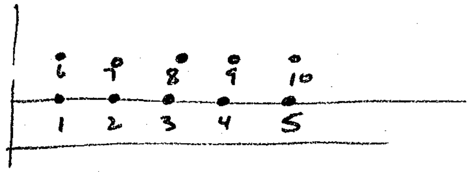
and solve for the values of T on the boundary.

ex: $T_{ij} = \frac{1}{4}\left[\underline{\underline{2T_{i+1j}}} + T_{ij-1} + T_{ij+1} \right]$ $\quad T_{i-1,j} = T_{i+1,j}$

↖ pick the pt inside region



OR SOLVE THIS VIA MATRICES



$$4T_1 - \cancel{4}(T_2 + BC_1 + BC_2 + T_6) = 0$$
$$4T_2 - (T_1 + T_3 + T_7 + BC) = 0$$

$$\begin{bmatrix} 4 & -1 & \cdot & \cdot & \cdot & -1 & \cdot & \cdot & \\ -1 & 4 & -1 & \cdot & \cdot & \cdot & -1 & \cdot & \\ \cdot & -1 & 4 & -1 & \cdot & \cdot & \cdot & -1 & \\ & & & & & & & & \end{bmatrix}$$

~~LET'S LOOK AT $\frac{\partial u}{\partial t} = \alpha \frac{\partial^2 u}{\partial x^2}$ one dim heat equation~~

~~EXPLICIT SCHEME $u_{i,j+1} = u_{i,j} + \underbrace{\boxed{\left(\frac{\alpha \Delta t}{\Delta x^2}\right)}}^{C}\left[u_{i+1,j} - 2u_{i,j} + u_{i-1,j}\right]$~~

~~U^j is the amplitude at t_j,~~

~~NEUMANN STABILITY ANALYSIS let $u_{ij} = U^j e^{Ik\Delta x}$~~ $\quad I = \sqrt{-1}$

~~$e^{Ik\Delta x}\left\{ U^{j+1} = U^j + U^j C\left[e^{Ik\Delta x} - 2 + e^{-Ik\Delta x}\right]\right\}$~~

~~$U^{j+1} = U^j \left\{ \underbrace{1 + 2C[\cos k\Delta x - 1]}_{A} \right\} = U^0 A^{j+1}$~~

k is sep var con

$i\Delta x = x_i$

$e^{I\xi} + e^{-I\xi} = 2\cos\xi$

~~if $|A|$ is >1 ⇒ solution becomes unbounded~~

~~FOR BOUNDEDNESS let $\{1 + 2C[\cos k\Delta x - 1]\}^2 \leq 1$~~

~~$4C[\quad] + (2C)^2[\quad]^2 \leq 0$~~

~~$4C[\quad]\{1 + C[\quad]\} \leq 0$~~

~~$C \leq \frac{1}{1 - \cos k\Delta x}$~~

~~backwards diff scheme~~

$$A = \begin{bmatrix} -\lambda & 1+2\lambda & -\lambda \\ & & \end{bmatrix}$$

$$w^{j+1} = \begin{bmatrix} w_{i+1,j+1} \\ w_{i,j+1} \\ w_{i+1,j+1} \end{bmatrix}$$

$$w^{j} = A w^{j+1}$$

$$w^{(0)} = A^{j+1} w^{j+1} \qquad w^{j+1} (A^{j+1})^{-1} = w^{(0)}$$

Boundary Conditions: Temperature dirichlet

$\frac{\partial T}{\partial n}$ Neumann

$A\frac{\partial T}{\partial n} = B(T-C)$ Robin

Note that plate ~~conductivity~~ physical properties doesn't play a part

however if k, A are fns(x,y) $\frac{1}{AC\rho}\left[\frac{\partial}{\partial x}\left(kA\frac{\partial T}{\partial x}\right) + \frac{\partial}{\partial y}\left(kA\frac{\partial T}{\partial y}\right)\right] = 0$

- For better results - obtain more equations by using more points (make $\Delta x, \Delta y$ smaller)
- We get banded matrices

$$\begin{bmatrix} -4 & 1 & & 1 & & & \\ 1 & -4 & 1 & & 1 & & \\ & 1 & -4 & 1 & & 1 & \\ & & 1 & -4 & 1 & & 1 \\ 1 & & & \ddots & \ddots & \ddots & \\ & 1 & & & \ddots & \ddots & \ddots \\ & & 1 & & & \ddots & \ddots \\ & & & \ddots & & & \ddots \end{bmatrix} \begin{pmatrix} T_1 \\ T_2 \\ \vdots \\ T_n \end{pmatrix} = \overset{bc's}{\begin{pmatrix} \vdots \\ \vdots \end{pmatrix}}$$

Better to solve by iteration Liebmann's Method.

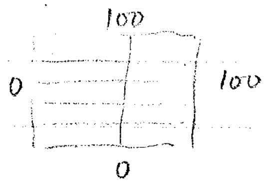
- starting w/ some initial approx. to T

use

$$T_{ij} = \frac{1}{4}\left(T_{i+1,j} + T_{i-1,j} + T_{i,j+1} + T_{i,j-1}\right)$$

- then replace new values by old values

Guidelines to determine initial T_{ij} choose some average between the boundaries



$\frac{200}{4} = 50$ at interior pts

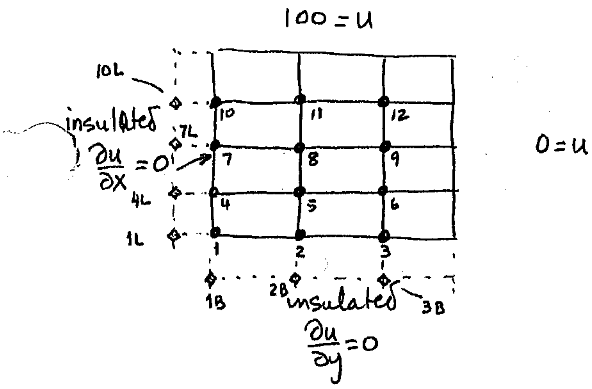
since T_{ij} is average of surrounding pts

cannot be > values at bdry

Liebmann's method takes care of storage but slowly converges.

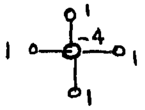
Southwell takes care of convergence - immediately replace

Implicit method.

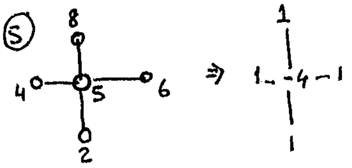


$$\nabla^2 u = 0 \Rightarrow u_{i+1j} + u_{i-1j} - 4u_{ij} + u_{ij+1} + u_{ij-1} = 0 \quad \text{IF } \Delta x = \Delta y$$

@ pt 5, 6, 8, 9, 11, 12



@ ⑤

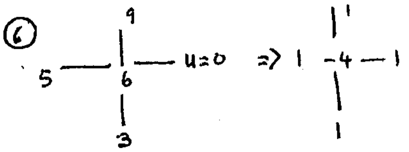


node nos.

coeff of nodes

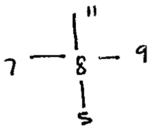
$$u_4 + u_8 + u_6 + u_2 - 4u_5 = 0$$

@ ⑥



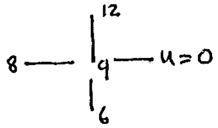
$$u_5 + u_9 + 0 + u_3 - 4u_6 = 0$$

@ ⑧



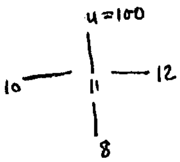
$$u_7 + u_{11} + u_9 + u_5 - 4u_8 = 0$$

@ ⑨



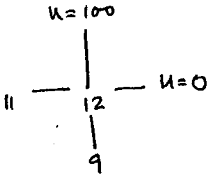
$$u_8 + u_{12} + 0 + u_6 - 4u_9 = 0$$

@ ⑪



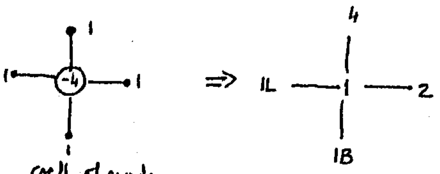
$$u_{10} + 100 + u_{12} + u_8 - 4u_{11} = 0$$

@ ⑫



$$u_{11} + 100 + 0 + u_9 - 4u_{12} = 0$$

@ ①



coeff of nodes

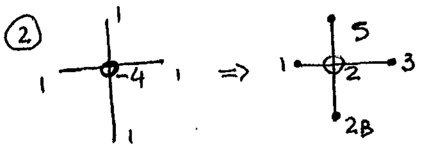
node no.

using $\left.\frac{\partial u}{\partial x}\right|_{node_1} = 0 \Rightarrow \frac{u_2 - u_{1L}}{2\Delta x} = 0 \Rightarrow \underbrace{u_2 = u_{1L}}$

using $\left.\frac{\partial y}{\partial y}\right|_{node_1} = 0 \Rightarrow \frac{u_4 - u_{1B}}{2\Delta y} = 0 \Rightarrow \underbrace{u_4 = u_{1B}}$

$$\overbrace{u_{1L} + u_4 + u_2 + u_{1B}} - 4u_1 = 0 \Rightarrow 2u_4 + 2u_2 - 4u_1 = 0$$

@ ②

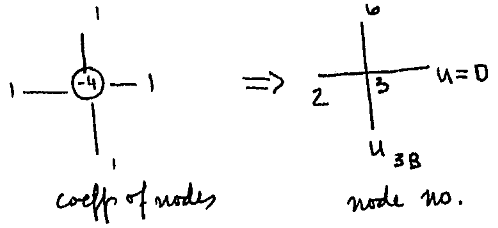


using $\frac{\partial u}{\partial y} = \frac{u_5 - u_{2B}}{2\Delta y} = 0 \Rightarrow u_5 = u_{2B}$

$$0 = u_1 + u_5 + u_3 + u_{2B} - 4u_2 \Rightarrow u_1 + 2u_5 + u_3 - 4u_2 = 0$$

"Geometrical, Mechanical, and Structural Adaptation of Mouse Femora Exposed to Different Loadings," *ASCE's Journal of Engineering Mechanics* (in press).	Levy, C.	Gordon, K.	M. Perl	Refereed Paper
Website http://allstar.eng.fiu.edu/ created as part of ALLSTAR project	Levy, Y.	Levy, C.	Jeffries, J.,	Other

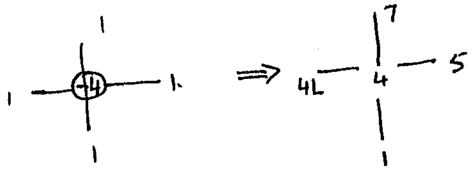
@ ③



using $\frac{\partial u}{\partial y} = \frac{u_6 - u_{3B}}{2\Delta y} = 0 \Rightarrow u_{3B} = u_6$

$u_2 + u_6 + 0 + u_{3B} - 4u_3 = 0 \Rightarrow u_2 + 2u_6 - 4u_3 = 0$

@ ④



using $\frac{\partial u}{\partial x} = 0 \quad \frac{u_5 - u_{4L}}{2\Delta x} = 0 \Rightarrow u_{4L} = u_5$

$u_{4L} + u_7 + u_5 + u_1 - 4u_4 = 0 \Rightarrow 2u_5 + u_7 + u_1 - 4u_4 = 0$

node 7 & node 10 use same method to get

$2u_8 + u_{10} + u_4 - 4u_7 = 0$

&

$2u_{11} + 100 + u_7 - 4u_{10} = 0$

$$\left[\begin{array}{ccc|ccc|ccc|ccc}
-4 & 2 & 0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & -4 & 1 & 0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & -4 & 0 & 0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 \\
\hline
1 & 0 & 0 & -4 & 2 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 1 & -4 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 1 & -4 & 0 & 0 & 1 & 0 & 0 & 0 \\
\hline
0 & 0 & 0 & 1 & 0 & 0 & -4 & 2 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 1 & -4 & 1 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & -4 & 0 & 0 & 1 \\
\hline
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & -4 & 2 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & -4 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & -4
\end{array}\right]
\begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \\ u_5 \\ u_6 \\ u_7 \\ u_8 \\ u_9 \\ u_{10} \\ u_{11} \\ u_{12} \end{bmatrix}
=
\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ -100 \\ -100 \\ -100 \end{bmatrix}
\quad \Leftarrow \; M\underline{z} = \underline{t}$$

$$\left[\begin{array}{c|c|c|c}
A & 2I & 0 & 0 \\
\hline
I & A & I & 0 \\
\hline
0 & I & A & I \\
\hline
0 & 0 & I & A
\end{array}\right]
\begin{bmatrix} \underline{p} \\ \underline{q} \\ \underline{r} \\ \underline{s} \end{bmatrix}
=
\begin{bmatrix} \underline{0} \\ \underline{0} \\ \underline{0} \\ \underline{b} \end{bmatrix}$$

$\underline{p} = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix} \qquad \underline{0} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$

$\underline{q} = \begin{bmatrix} u_4 \\ u_5 \\ u_6 \end{bmatrix} \qquad \underline{b} = \begin{bmatrix} -100 \\ -100 \\ -100 \end{bmatrix}$

$\underline{r} = \begin{bmatrix} u_7 \\ u_8 \\ u_9 \end{bmatrix} \qquad A = \begin{bmatrix} -4 & 2 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & -4 \end{bmatrix}$

$\underline{s} = \begin{bmatrix} u_{10} \\ u_{11} \\ u_{12} \end{bmatrix}$

TABLE 2

List of Accepted and Printed Publications

From July 1, 1996, to June 30, 1997

Title and Journal/Proceedings with Date	Primary FIU Author	Other FIU Authors	Non-FIU Authors	<u>Category</u> Review Article Book Chapter Paper Abstract Other
"Vibration Characteristics of a Partially Covered Double Sandwich Cantilever Beam," *AIAA Journal,* Vol. **34**, pp. 2622-2626 (1996). Also on CD-ROM *AIAA Journal on Disc*, Vol. **2** **(1)** (1996).	Chen Q.	Levy, C.		Reviewed Paper
"Interaction Effects on the 3-D Stress Intensity Factor of Combined Arrays of Radial and Longitudinal Coplanar Cracks in an Internally Pressurized Thick-Walled Cylinder," *Journal of Pressure Vessel Technology,* Vol. **119**, (1997).	Levy, C.	J. Wang	M. Perl	Reviewed Paper
"Erosion Geometry Effects on the Stress Intensity Factors of a_t Crack Emanating from an Erosion in a Pressurized Autofrettaged Thick-Wall Cylinder," 1997 PVP Conference, Orlando, FL, July, 1997 (*in press*).	Levy, C.	H. Fang	M. Perl	Refereed Proceeding Paper
"Experimental Analysis of Smart Structure with Damping Treatment and SMA," Proc. of the MRS Symp., Boston, MA, 2-6 Dec. 1996, Vol. **459**, pp. 163-168.	Chen, Q.	Ma, J. Levy, C.		refereed proceeding paper

$A\underline{p} + 2I\underline{q} = \underline{0}$ solve $A\underline{p} = -2I\underline{q}$

$\underline{p} + A\underline{q} + \underline{r} = \underline{0}$ $\underline{r} = -\underline{p} - A\underline{q}$ $A\underline{r} = -A\underline{p} - A^2\underline{q} = (2I - A^2)\underline{q}$

$\underline{q} + A\underline{r} + \underline{s} = \underline{0}$ $\underline{s} = -\underline{q} - A\underline{r}$ $\underline{s} = -\underline{q} - (2I - A^2)\underline{q} = -(3I - A^2)\underline{q}$

$\underline{r} + A\underline{s} = \underline{b}$ $A\underline{r} + A^2\underline{s} = A\underline{b}$ $(2I - A^2)\underline{q} - (3A^2 - A^4)\underline{q} = A\underline{b}$

$(A^4 - 4A^2 + 2I)\underline{q} = A\underline{b}$

solve

$A\underline{r} = (2I - A^2)\underline{q}$ ③ for $\underline{r}$

$\underline{s} = (A^2 - 3I)\underline{q}$ ② for $\underline{s}$

$(A^4 - A^2 + 2I)\underline{q} = A\underline{b}$ ① for $\underline{q}$

$A\underline{p} = -2I\underline{q}$ ④ for $\underline{p}$

In general, let

$$M = LU = \begin{bmatrix} \beta_1 & 0 & 0 & 0 \\ \alpha_2 & \beta_2 & 0 & 0 \\ 0 & \alpha_3 & \beta_3 & 0 \\ 0 & 0 & \alpha_4 & \beta_4 \end{bmatrix} \begin{bmatrix} I_1 & \gamma_1 & 0 & 0 \\ 0 & I_2 & \gamma_2 & 0 \\ 0 & 0 & I_3 & \gamma_3 \\ 0 & 0 & 0 & I_4 \end{bmatrix} = \begin{bmatrix} A & 2I & 0 & 0 \\ I & A & I & 0 \\ 0 & I & A & I \\ 0 & 0 & I & A \end{bmatrix}$$

where $\beta_1 I_1 = A$ $\beta_1 = A$

$\beta_1\gamma_1 = 2I$ $\gamma_1 = 2\beta_1^{-1} = 2A^{-1}$

$\alpha_2 I_1 = I$ $\alpha_2 = I$

$\alpha_2\gamma_1 + \beta_2 I_2 = A$ $\beta_2 = A - \alpha_2\gamma_1$ or $\beta_1\beta_2 = \beta_1 A - 2I = A^2 - 2I$ $\beta_2 = A - 2A^{-1}$

$\beta_2\gamma_2 = I$ $\gamma_2 = \beta_2^{-1} = (A - 2A^{-1})^{-1}$

$\alpha_3 I_2 = I$ $\alpha_3 = I$

$\alpha_3\gamma_2 + \beta_3 I_3 = A$ $\beta_3 = A - \alpha_3\gamma_2$ $\beta_2\beta_3 = \beta_2 A - I$

$\beta_3\gamma_3 = I$ $\gamma_3 = \beta_3^{-1}$

$\alpha_4 I_3 = I$ $\alpha_4 = I$

$\alpha_4\gamma_3 + \beta_4 I_4 = A$ $\beta_4 = A - \alpha_4\gamma_3$ $\beta_3\beta_4 = \beta_3 A - I$

Now let $M\underline{z} = \underline{t}$ or $LU\underline{z} = \underline{t}$ or $U\underline{z} = \underline{y}$ & $L\underline{y} = \underline{t}$

$$L\underline{y} = \underline{t} \qquad \begin{bmatrix} \beta_1 & 0 & 0 & 0 \\ \alpha_2 & \beta_2 & 0 & 0 \\ 0 & \alpha_3 & \beta_3 & 0 \\ 0 & 0 & \alpha_4 & \beta_4 \end{bmatrix} \begin{bmatrix} \underline{y}_1 \\ \underline{y}_2 \\ \underline{y}_3 \\ \underline{y}_4 \end{bmatrix} = \begin{bmatrix} \underline{0} \\ \underline{0} \\ \underline{0} \\ \underline{b} \end{bmatrix}$$

$\beta_1\underline{y}_1 = \underline{0}$ $A\underline{y}_1 = \underline{0} \rightarrow \underline{y}_1 = \underline{0}$

$\alpha_2\underline{y}_1 + \beta_2\underline{y}_2 = \underline{0} \Rightarrow$ solve $\beta_2\underline{y}_2 = -\alpha_2\underline{y}_1$ for $\underline{y}_2$

$\alpha_3\underline{y}_2 + \beta_3\underline{y}_3 = \underline{0} \Rightarrow$ solve $\beta_3\underline{y}_3 = -\alpha_3\underline{y}_2$ for $\underline{y}_3$

$\alpha_4\underline{y}_3 + \beta_4\underline{y}_4 = \underline{b} \Rightarrow$ solve $\beta_4\underline{y}_4 = -\alpha_4\underline{y}_3 + \underline{b}$ for $\underline{y}_4$

10.) Books/Chapters (provide complete citation)

none

knowing $\begin{pmatrix} \underline{y}_1 \\ \underline{y}_2 \\ \underline{y}_3 \\ \underline{y}_4 \end{pmatrix}$ we now solve $U\underline{z} = \underline{y}$

$$\begin{bmatrix} I_1 & \gamma_1 & 0 & 0 \\ 0 & I_2 & \gamma_2 & 0 \\ 0 & 0 & I_3 & \gamma_3 \\ 0 & 0 & 0 & I_4 \end{bmatrix} \begin{pmatrix} \underline{p} \\ \underline{q} \\ \underline{r} \\ \underline{s} \end{pmatrix} = \begin{pmatrix} \underline{y}_1 \\ \underline{y}_2 \\ \underline{y}_3 \\ \underline{y}_4 \end{pmatrix}$$

Backward Subst.

$I_1\underline{p} + \gamma_1\underline{q} = \underline{y}_1$ $\quad \underline{p} = \underline{y}_1 - \gamma_1\underline{q}$ or $\beta_1\underline{p} = \beta_1\underline{y}_1 - 2\underline{q}$ ③ for $\underline{p}$

$I_2\underline{q} + \gamma_2\underline{r} = \underline{y}_2$ $\quad \underline{q} = \underline{y}_2 - \gamma_2\underline{r}$ or $\beta_2\underline{q} = \beta_2\underline{y}_2 - \underline{r}$ ② for $\underline{q}$

$I_3\underline{r} + \gamma_3\underline{s} = \underline{y}_3$ $\quad \underline{r} = \underline{y}_3 - \gamma_3\underline{s}$ or $\beta_3\underline{r} = \beta_3\underline{y}_3 - \underline{s}$ solve for $\underline{r}$

$I_4\underline{s} = \underline{y}_4$ $\quad \underline{s} = \underline{y}_4$ ①

For forward subst. $\quad A\underline{y}_1 = \underline{0} \quad \underline{y}_1 = \underline{0}$

$\beta_2\underline{y}_2 = -\alpha_2\underline{y}_1 \Rightarrow \beta_1\beta_2\underline{y}_2 = -\beta_1\underline{y}_1$ or $(\beta_1 A - 2I)\underline{y}_2 = -\beta_1\underline{y}_1$

$\beta_3\underline{y}_3 = -\alpha_3\underline{y}_2 \Rightarrow \beta_2\beta_3\underline{y}_3 = -\beta_2\underline{y}_2$ or $(\beta_2 A - I)\underline{y}_3 = -\beta_2\underline{y}_2$

$\beta_4\underline{y}_4 = -\alpha_4\underline{y}_3 + \underline{b} \Rightarrow \beta_3\beta_4\underline{y}_4 = -\beta_3\underline{y}_3 + \beta_3\underline{b}$ or $(\beta_3 A - I)\underline{y}_4 = -\beta_3\underline{y}_3 + \beta_3\underline{b}$

α_s are never computed ; γ_s are not computed ; only β_s are computed.

use LU decomp. to solve for β_{i+1}
how? use LU decomp on β_i
then use $L_i U_i \beta_{i+1} = $ rhs columns

for example $\beta_1\beta_2 = \beta_1 A - 2I \Rightarrow A\beta_2 = A^2 - 2I$

$(3\times3)\,(\beta_{21}\ \beta_{22}\ \beta_{23}) = (3\times3) \Rightarrow \beta_2$

cols. of β_2 $\quad$ cols of $\beta_1 A - 2I$

$\beta_2\beta_3 = \beta_2 A - I \Rightarrow \beta_3$

$\beta_3\beta_4 = \beta_3 A - I \Rightarrow \beta_4$

---

knowing β_i's use them to find $\underset{\sim}{y}_i$'s

$\beta_2\underline{y}_2 = -\alpha_2\underline{y}_1 = -\underline{y}_1 \Rightarrow \underline{y}_2$

$\beta_3\underline{y}_3 = -\underline{y}_2 \Rightarrow \underline{y}_3$

$\beta_4\underline{y}_4 = -\underline{y}_3 + \underline{b} \Rightarrow \underline{y}_4$

---

knowing $\underline{y}$ solve the eqs to find $\underline{p}, \underline{q}, \underline{r}, \underline{s}$

from $U\underline{z} = \underline{y}$

$I_4\underline{s} = \underline{y}_4$ $\quad \to \quad s = y_4$

$I_3\underline{r} + \gamma_3\underline{s} = \underline{y}_3 \quad \therefore \quad \underline{r} = \underline{y}_3 - \gamma_3\underline{s}$ $\quad \xrightarrow{\times \text{ by } \beta_3} \beta_3\underline{r} = \beta_3\underline{y}_3 - s$

$I_2\underline{q} + \gamma_2\underline{r} = \underline{y}_2 \quad \therefore \quad \underline{q} = \underline{y}_2 - \gamma_2\underline{r}$ $\quad \xrightarrow{\times \text{ by } \beta_2} \beta_2\underline{q} = \beta_2\underline{y}_2 - \underline{r}$

$I_1\underline{p} + \gamma_1\underline{q} = \underline{y}_1 \quad \therefore \quad \underline{p} = \underline{y}_1 - \gamma_1\underline{q}$ $\quad \xrightarrow{\times \text{ by } \beta_1} \beta_1\underline{p} = \beta_1\underline{y}_1 - \underline{q}$

~~go to top of page for backward substitution~~

Case 1 Coverage Length Effects

Figure 2 shows the results of system loss factor versus coverage ratio. For the parameters given, the maximum loss factor for the first mode occurs when coverage length is 0.7L. From theoretical result, the maximum value occurs at about 0.6L. This confirms the result from previous theories that there exists an optimal coverage ratio which makes the loss factor a maximum. Also, the theoretically generated loss factor is smaller than the experimental result. The main reason for the difference may be that in the experiments epoxy adhesives are used to connect the damping layers with the beam and constraining layers. These adhesive layers will dissipate energy but such effects were not considered in the theoretical analysis. The maximum error for system loss factor between theory and experiment is about 12%.

The effects of coverage length on the natural frequency of the system are shown in Figure 3. An increase in coverage length causes an increase in the natural frequencies. This tendency confirms the numerical results obtained from the mathematical model. In addition, the theoretical results give larger natural frequencies compared to the experimental results. The difference in natural frequency between experiment and numerical method is no more than 5%.

Case 2 End Mass Effects

Figure 4 compares system loss factor with different mass ratios. The trends in both the theoretical and experimental results show that an increase of the mass ratio (the ratio of end mass to the mass of the whole beam) will decrease the system loss factor. Also, the system loss factors obtained by experiments are larger than those obtained by theoretical analysis. The primary explanation is that in the experiments epoxy adhesives were used to connect the damping layers with the beam and constraining layers. These adhesive layers will dissipate energy and

Remember $Ax = b$

$$[L+D+U]x = b$$

or $(\overset{L}{\cancel{\;}}+D)\cancel{x}^{(n+1)} = b - \overset{U}{\cancel{\;}}x^{(n)}$

$$x^{(n+1)} = D^{-1}\left[b - \overset{L}{\cancel{\;}}x^{(n+1)} - \overset{U}{\cancel{\;}}x^{(n)}\right]$$

$= D^{-1}b - D^{-1}\overset{L}{\cancel{\;}}x^{(n+1)} - D^{-1}\overset{U}{\cancel{\;}}x^{(n)} - D^{-1}Dx^{(n)} + x^{(n)}$
$= D^{-1}(b - \cancel{\;}x^{(n+1)} - (D+\overset{L}{\cancel{\;}})x^{(n)}) + x^{(n)}$

$$D^{-1} = \begin{bmatrix} 1/a_{11} & & \\ & \ddots & \\ & & 1/a_{nn} \end{bmatrix}$$

S.O.R defined $x^{(n+1)} = (1-\omega)\,x^{(n)} + \omega\,x^{(n+1)}$ from

for S.O.R. $x_i^{(n+1)} = x_i^{(n)} + \frac{\omega}{a_{ii}}\left(b_i - \sum_{j=1}^{i-1} a_{ij}x_j^{(n+1)} - \sum_{j=i}^{n} a_{ij}\,x_j^{(n)}\right)$

Gauss Seidel $G_{GS} = -(L+D)^{-1}U$
$\underline{x}^{(n+1)} = -G\,\underline{x}^{(n)} + (L+D)^{-1}\underline{b}$

This is ~~[illegible]~~ Gauss-Seidel

$$x^{(n+1)} = (1-\omega)\,x^{(n)} + \omega\,x^{(n+1)}$$

from

For Jacobi $Ax = b$

$$x^{(n+1)} = D^{-1}\left(b - [L+U]\,x^{(n)}\right) = D^{-1}\left(b - Ax^{(n)}\right) + x^{(n)}$$

or $\underline{x}^{(n+1)} = (1-\omega)\underline{x}^{(n)} + \omega D^{-1}(b - [L+U]\underline{x}^{(n)}) = (1+\omega)\underline{x}^{(n)} + \omega D^{-1}(b - A\underline{x}^{(n)}) + \omega\underline{x}^{(n)} = x^{(n)} + \omega D^{-1}(b - A\underline{x}^{(n)})$

$$\underline{x_i^{(n+1)} = \frac{\omega}{a_{ii}}\left(b_i - \sum a_{ij}x_j^{(n)}\right) + \underline{x}_i^{(n)}}$$

$G_J = -D^{-1}(L+U)$
$\underline{x}^{(n+1)} = G\,\underline{x}^{(n)} + D^{-1}\underline{b}$

Actually it is better to use G-S + S.O.R.

$$x^{(k+1)} = (D+\omega L)^{-1}\left[(1-\omega)D - \omega U\right]x^{(k)} + \omega(D+\omega L)^{-1}b$$

start w/ $Dx = Dx - \omega\left[b - (L+D+U)x\right]$

then $[D+\omega L]\,x^{(k+1)} = [(1-\omega)D - \omega U]\,x^{(k)} + \omega b$

$$G_\omega = [D+\omega L]^{-1}\,[(1-\omega)D - \omega U]$$

all in the form $\underset{\sim}{x}^{(n+1)} = G\,\underset{\sim}{x}^{(n)} + \underset{\sim}{c}$

for convergence $\underset{\sim}{e}^{n+1} = \underset{\sim}{x}^{n+1} - \underset{\sim}{x} = G\underset{\sim}{e}^{(n)} = \dots G^{n}\underset{\sim}{e}^{(0)}$ since $\underset{\sim}{x} = G\underset{\sim}{x} + \underset{\sim}{c}$

for $\underset{\sim}{e}^{n+1} \to 0$ as $n \to \infty \implies$ ~~$G^{n+1} \to 0$~~ $|\lambda_i| < 1$

if $\underset{\sim}{G}$ has eigenvectors $\underset{\sim}{v}_s$ then $\underset{\sim}{e}^{(0)} = \sum c_i \underset{\sim}{v}_i$

then $G\underline{e}^{(0)} = \sum c_i G \underset{\sim}{v}_i = \sum c_i \lambda_i \underset{\sim}{v}_i$

$\therefore G^{n}\underline{e}^{(0)} = \underset{\sim}{e}^{(n+1)} = \sum c_i \lambda_i^{n+1} \underset{\sim}{v}_i \longrightarrow 0$ only if $|\lambda_i| < 1$

From Ames Book

p 172 • by choosing $\omega = \dfrac{2}{1+\sqrt{1-\mu_1^2}}$ where μ_1 is the spectral radius Jacobi iteration matrix ie $-D^{-1}(L+U)$

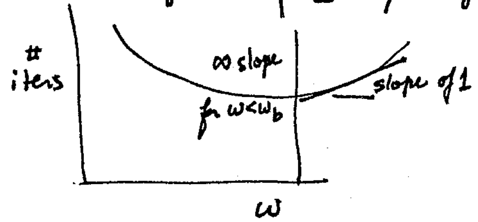
then the S.O.R. method will increase convergence speed

p 170 • for gauss seidel method $^{(\omega=1)}$ $\lambda_{GS} = \lambda_J^2$ if consistent ordering & if matrix has property "A" see pg 168

p 172 • better to overestimate ω than to underestimate (Young, 1954)

$\lambda - \omega\mu_i\lambda^{1/2} + \omega - 1 = 0$ relates λ eigenvalue of G_ω to μ_i eigenvalue of G_J

p 173 • can try by changing ω & plot to find W_{opt}. (Bellman, 1961)

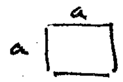


p178 Dirichlet 5 pt molecule has $\omega_{opt} = \dfrac{2}{1+h} \cong 2-2h$ for $0<x<\pi$ $0<y<\pi$

p179 Dirichlet 9 pt molecule $W_{opt} \underset{\text{Garabedian}}{\cong} 2 - 1.442\sqrt{2}h = 2 - 2.04h$ h is stepsize

p 179 Dirichlet 9 pt formula for laplace eqn. $\omega_{opt} = 2 - 2.116\dfrac{\pi h}{a} + 2.24\left(\dfrac{\pi h}{a}\right)^2$

Vooren & Vliegenhart



p 180 for unequal mesh spacing 9 pt $0<x<a$ $0<y<b$

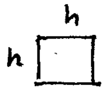
$$\lambda_J = \frac{1}{h^2+k^2}\left(k^2\cos\frac{\pi h}{a} + h^2\cos\frac{\pi k}{b}\right)$$

$\Delta x = h$
$\Delta y = k$

$$\begin{matrix} 1 & 4 & 1 \\ 4 & -20 & 4 \\ 1 & 4 & 1 \end{matrix} + O(h^6)$$

$$\lambda_{GS} = \lambda_J^2$$

Convergence rates for $\nabla^2 u = 0$ w/ Dirichlet BC



using 5 pt formula

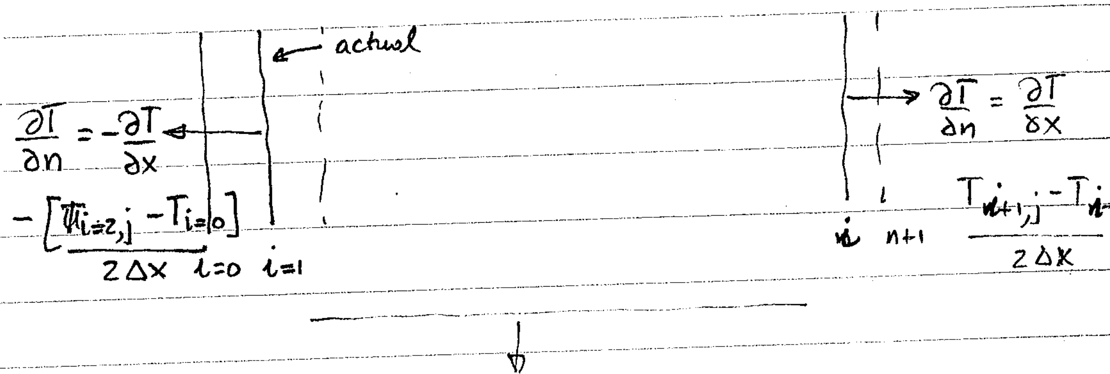
Pg 197	Point Iterative Methods - solve for unknowns pt. by pt.	Jacobi	$\frac{1}{2}h^2$	line iterative schemes - solve for unknowns along a line	Jacobi	h^2	Pg 201
		G-S	h^2		G-S	$2h^2$	
		Optimum SOR	$2h$		SOR	$2\sqrt{2}h$	

property (A) a matrix has property A if we can find a permutation matrix π, ∋. $\pi A \pi^T = \begin{bmatrix} D_1 & F \\ G & D_2 \end{bmatrix}$ where D_1, D_2 are square diag matrices & F & G are rectangular matrices

EGM 6422
LECTURE 21/22

How to handle deriv BC's

as before define a new row of pts



$$\frac{\partial T}{\partial n} = -\frac{\partial T}{\partial x}$$

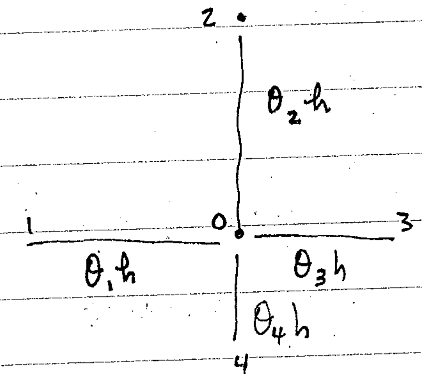
$$-\frac{\left[T_{i=2,j} - T_{i=0}\right]}{2\Delta x}$$

$$\frac{\partial T}{\partial n} = \frac{\partial T}{\partial x}$$

$$\frac{T_{i+1,j} - T_{i-1,j}}{2\Delta x}$$

$$\frac{\partial T}{\partial n} = -\frac{\partial T}{\partial y}$$

irregular BCs



$$\left.\frac{\partial T}{\partial x}\right|_{1-0} = \frac{T_0 - T_1}{\theta_1 h} \qquad \left.\frac{\partial T}{\partial x}\right|_{3-0} = \frac{T_3 - T_0}{\theta_3 h}$$

$$\frac{\partial}{\partial x}\left(\frac{\partial T}{\partial x}\right) = \frac{\frac{T_3 - T_0}{\theta_3 h} - \frac{T_0 - T_1}{\theta_1 h}}{\frac{1}{2}(\theta_1 + \theta_3) h}$$

$$= \frac{2}{h^2}\left\{\frac{T_1 - T_0}{\theta_1(\theta_1 + \theta_3)} + \frac{T_3 - T_0}{\theta_3(\theta_1 + \theta\right.$$

also $$\frac{\partial^2}{\partial y^2} = \frac{2}{h^2}\left\{\frac{T_2 - T_0}{\theta_2(\theta_2 + \theta_4)} + \frac{T_4 - T_0}{\theta_4(\theta_2 + \theta_4)}\right\}$$

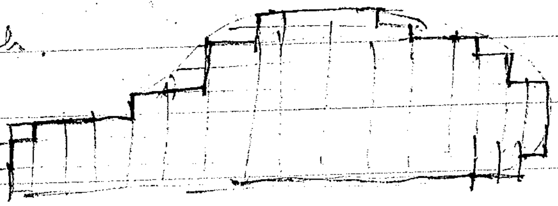
each operator is $O(h)$

we use this for $\nabla^2 T$ near a bdry

$$\frac{2}{h^2}\left[\frac{T_1}{\theta_1(\theta_1 + \theta_3)} + \frac{T_2}{\theta_2(\theta_2 + \theta_4)} + \frac{T_3}{\theta_3(\theta_1 + \theta_3)} + \frac{T_4}{\theta_4(\theta_2 + \theta_4)} - \left(\frac{1}{\theta_1\theta_3} + \frac{1}{\theta_2\theta_4}\right)T\right.$$

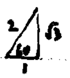
this is $O(h)$

- Another way is to make the mesh very fine & to approximate the bdry by the mesh.



~~$\sin(45+60) = \sin 45\cos 60 + \cos 45 \sin 60$~~

~~$= \frac{\sqrt{2}}{2}\cdot\frac{\sqrt{3}}{2} + \frac{\sqrt{2}}{2}\cdot\frac{\sqrt{3}}{2}$~~



~~$\frac{\sqrt{2}(\sqrt{3}+1)}{4}$~~

~~$\cos(45+60) = \cos 45\cos 60 - \sin 45 \sin 60$~~

~~$\frac{\sqrt{2}}{2}\cdot\frac{\sqrt{3}}{2} - \frac{\sqrt{2}}{2}\cdot\frac{\sqrt{3}}{2}$~~

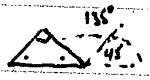
~~$-\frac{\sqrt{2}(\sqrt{3}-1)}{4}$~~

~~$\sin(45+30) = \sin 45\cos 30 + \cos 45 \sin 30$~~

~~$\frac{\sqrt{3}}{2}\cdot\frac{\sqrt{2}}{2} + \frac{1}{2}\cdot\frac{\sqrt{2}}{2} = \frac{\sqrt{2}}{2}(\sqrt{3}+1)$~~

~~$\cos(45+30) = \cos 45\cos 30 - \sin 180 \sin 30$~~

~~$-\frac{\sqrt{3}}{2}$~~



$\sin(45+30) = S45C30 + C45S30 =$

$\sqrt{2}/2 \cdot \frac{\sqrt{3}}{2} + \frac{\sqrt{2}}{2}\cdot\frac{1}{2} = \frac{\sqrt{2}}{4}(\sqrt{3}+1)$

$\cos(45+30) = C45C30 - S45S30$

$\frac{\sqrt{2}}{2}\cdot\frac{\sqrt{3}}{2} - \frac{\sqrt{2}}{2}\cdot\frac{1}{2} = \frac{\sqrt{2}}{4}(\sqrt{3}-1)$

$\sin(45+60) = S45C60 + C45S60$

$\frac{\sqrt{2}}{2}\cdot\frac{1}{2} + \frac{\sqrt{2}}{2}\cdot\frac{\sqrt{3}}{2} = \frac{\sqrt{2}}{4}(\sqrt{3}+1)$

$\cos(45+60) = C45C60 - S45S60$

$\frac{\sqrt{2}}{2}\cdot\frac{1}{2} - \frac{\sqrt{2}}{2}\cdot\frac{\sqrt{3}}{2} = -\frac{\sqrt{2}}{4}(\sqrt{3}-1)$

0 75, ~~150~~ 105

~~$\frac{\partial^2 u}{\partial a^2} = \frac{\partial^2 u}{\partial x^2}$~~

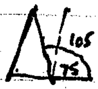
~~$\frac{\partial^2 u}{\partial b^2} = \frac{2(\sqrt{3}+1)^2}{16} u_{xx} + 2u_{xy}\frac{2}{16}\cdot 2 + u_{yy}\frac{2}{16}(\sqrt{3}-1)^2$~~

$\frac{2(4+2\sqrt{3})}{16}$ $\frac{2}{16}(2-2\sqrt{3})$

~~$\frac{\partial^2 u}{\partial c^2} = \frac{3}{4}u_{xx} + 2u_{xy}\frac{1}{2}\cdot -\frac{\sqrt{3}}{2} + u_{yy}\cdot\frac{1}{4}$~~

~~$\frac{\partial^2 u}{\partial b^2} + \frac{\partial^2 c}{\partial c^2} = \left(\frac{20}{16} + \frac{4\sqrt{3}}{16}\right)u_{xx} + \frac{8(1-\sqrt{3})}{16}u_{xy} + \left(\frac{8}{16} - \frac{4\sqrt{3}}{16}\right)u_{yy}$~~

$A+B+C =$



For ~~[illegible]~~: 75, 75, 30

$\frac{\partial^2 u}{\partial a^2} = \frac{\partial^2 u}{\partial x^2}$

$\frac{\partial^2 u}{\partial b^2} = \frac{2}{16}(\sqrt{3}-1)^2 u_{xx} + \frac{2}{16}\cdot 2u_{xy} + \frac{2}{16}(\sqrt{3}+1)^2 u_{yy}$

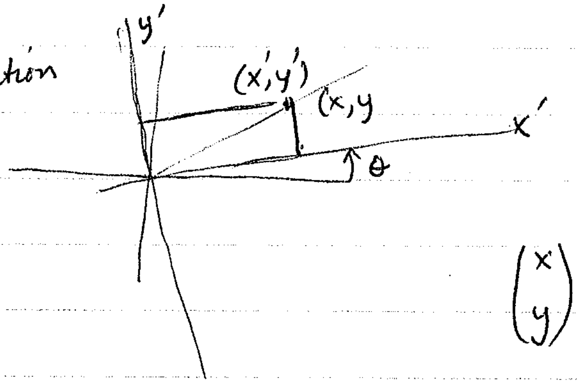
$\frac{\partial^2 u}{\partial c^2} = \frac{2}{16}(\sqrt{3}-1)^2 u_{xx} - \frac{2}{16}\cdot 2u_{xy} + \frac{2}{16}(\sqrt{3}+1)^2 u_{yy}$

$u_{bb} + u_{cc} = \frac{4}{16}(2-2\sqrt{3})u_{xx} + \frac{4}{16}(4+2\sqrt{3})u_{yy}$

$u_{bb} + u_{cc} + \frac{4}{16}(2+4\sqrt{3})u_{aa} = \frac{4}{16}(4+2\sqrt{3})\nabla^2 u$ ✓

Sometimes we cannot fit regular mesh but we can an equispaced triangular networks

Use rotation matrix



$$x = x'\cos\theta - y'\sin\theta \qquad x' = x\cos\theta + y s$$

$$y = x'\sin\theta + y'\cos\theta \qquad y' = -x s\theta + y c$$

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} x' \\ y' \end{pmatrix}; \quad \begin{pmatrix} \cos\theta & +\sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}$$

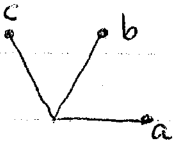
thus $$\frac{\partial u}{\partial x'} = \frac{\partial u}{\partial x}\frac{\partial x}{\partial x'} + \frac{\partial u}{\partial y}\frac{\partial y}{\partial x'} = u_{,x}\cos\theta + u_{,y}\sin\theta$$

$$\frac{\partial}{\partial x'}\left(\frac{\partial u}{\partial x'}\right) = \frac{\partial}{\partial x}\left(\frac{\partial x}{\partial x'}\right) + \frac{\partial}{\partial y}\left(\frac{\partial y}{\partial x'}\right)$$

$$= \frac{\partial}{\partial x}\left(\frac{\partial u}{\partial x'}\right)\cos\theta + \frac{\partial}{\partial y}\left(\frac{\partial u}{\partial x'}\right)\sin\theta$$

$$= u_{xx}\cos^2\theta + 2u_{xy}\sin\theta\cos\theta + u_{yy}\sin^2\theta$$

For equispaced



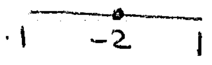
$$\theta = 0 \qquad \frac{\partial^2 u}{\partial a^2} = u_{xx}$$

$$\theta = 60^\circ \qquad \frac{\partial^2 u}{\partial b^2} = \frac{1}{4}u_{xx} + \frac{\sqrt{3}}{2}u_{xy} + \frac{3}{4}u_{yy}$$

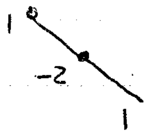
$$\theta = 120^\circ \qquad \frac{\partial^2 u}{\partial c^2} = \frac{1}{4}u_{xx} - \frac{\sqrt{3}}{2}u_{xy} + \frac{3}{4}u_{yy}$$

$$\left(\frac{\partial^2 u}{\partial a^2} + \frac{\partial^2 u}{\partial b^2} + \frac{\partial^2 u}{\partial c^2}\right) = \frac{3}{2}\nabla^2 u$$

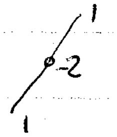
$$\frac{\partial^2 u}{\partial a^2} =$$



$$\frac{\partial^2 u}{\partial c^2}$$



$$\frac{\partial^2 u}{\partial b^2}$$



$$\therefore \nabla^2\Psi = \frac{2}{3h^2}\left\{\begin{matrix} & 1 & & 1 & \\ 1 & & -6 & & 1 \\ & 1 & & 1 & \end{matrix}\right\}T_{ij} \qquad \text{this is } O(h^2)$$

$\nabla^2\Psi = f$ RHS is f_{ij}

what if we wanted to get better results using finer mesh

if $T_{ij} - \bar{T}_{ij} = A h_1^p$ error based on leading term of operator

$T_{ij} - \hat{T}_{ij} = A h_2^P$] then $T_{ij} = \dfrac{h_2^P \bar{T}_{ij} - h_1^P \hat{T}_{ij}}{h_2^P - h_1^P}$

it is of dubious use near curved boundaries, near corners w/ interior angles > 180° & near boundaries where function values are <u>not</u> smooth.

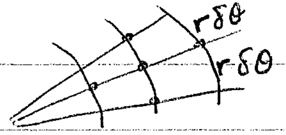
if p is unknown solve 3 solutions with $h_1 = h$ $\bar{T}_{ij}$

$h_2 = \frac{1}{2}h$ $\hat{T}_{ij}$

$h_3 = \frac{1}{4}h$ $\tilde{T}_{ij}$

then $$\frac{\hat{T}_{ij} - \bar{T}_{ij}}{\tilde{T}_{ij} - \hat{T}_{ij}} = 2^P \implies p$$

what about Polar coords. $\frac{\partial^2 u}{\partial r^2} + \frac{1}{r}\frac{\partial u}{\partial r} + \frac{1}{r^2}\frac{\partial^2 u}{\partial \theta^2}$



$$\nabla^2 T = \frac{u_{i+1,j} - 2u_{ij} + u_{i-1,j}}{\Delta r^2} + \frac{1}{i\Delta r}\frac{u_{i+1,j} - u_{i-1,j}}{2\Delta r} + \frac{u_{i,j+1} - 2u_{i,j} + u_{i,}}{(i\Delta r)^2 (\Delta\theta)^2}$$

Ordering of mesh → go row by row $\left[\begin{array}{cccc} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 9 & 10 & 11 & 12 \end{array}\right]$ etc

<u>DO 13</u> not iterative but by matrix sol

repeat 13 using Liebmann & Southwell for $\omega = 1.2, 1.4$

51

ADI method may also be used here. In 3-D as well where we do it in x, then y, then z. We ignore the intermediate steps

In 2-D 2nd, 4th, 6th etc are correct ... ignore 1, 3, 5, 7...

in 3D 3rd, 6th, 9th etc are correct ... ignore 1, 2, 4, 5, 7, 8 etc...

either use heat equation forms & let $t \rightarrow \infty$ or

$$\nabla^2 u = 0 = \frac{u_{i-1,j}^{(n+1)} - 2u_{i,j}^{(n+1)} + u_{i+1,j}^{(n+1)}}{\Delta x^2} + \frac{u_{i,j-1}^{(n)} + 2u_{i,j}^{(n)} + u_{i,j+1}^{(n)}}{\Delta y^2} \quad \text{impl in x} \quad \text{expl in y}$$

$$\Rightarrow \quad \frac{u_{i-1,j}^{(n+1)} - 2u_{i,j}^{(n+1)} + u_{i+1,j}^{(n+1)}}{\Delta x^2} + \frac{u_{i,j-1}^{(n+2)} - 2u_{i,j}^{(n+2} + u_{i,j+1}^{(n+2}}{\Delta y^2} \quad \text{expl. in x} \quad \text{impl in y}$$

Can use letting r be very large in

$$-r u_{i-1j}^{(n+1)} (1+2)u_{i,j}^{(n+1)} - r u_{i+1,j}^{(n+1)} = r u_{i,j-1}^{(n)} + (1-2r) u_{i,j}^{(n)} + (u_{ij+1}^{(n}$$

Then

$$-r u_{ij-1}^{(n+2)} + (1+2r) u_{ij}^{(n+2)} - r u_{ij+1}^{(n+2)} = r u_{i-1j}^{(n+1} + (1-2r) u_{ij}^{(n+1)} + r u_{i+1j}^{(n+1)}$$

reduce to
$$-u_{i-1,j}^{n+1} + 2u_{i,j}^{n+1} - u_{i+1j}^{n+1} = u_{ij+1}^{n} - 2u_{ij}^{n} + u_{ij+1}^{n}$$
$$-u_{ij-1}^{n+2} + 2u_{ij}^{n+2} + u_{ij+1}^{n+2} = u_{i-1j}^{n+1} + 2u_{ij}^{(n+1)} + u_{i+1j}^{(n+1)}$$

The implementation of the iteration

In practice, the iteration defined by equation (5.57), namely

$$(\mathbf{A}+\mathbf{N})\mathbf{u}^{(n+1)}=(\mathbf{A}+\mathbf{N})\mathbf{u}^{(n)}+(\mathbf{q}-\mathbf{A}\mathbf{u}^{(n)}),$$

where $(\mathbf{A}+\mathbf{N})=\bar{\mathbf{L}}\bar{\mathbf{U}}$, is dealt with as follows.

Let $\mathbf{d}^{(n)}=\mathbf{u}^{(n+1)}-\mathbf{u}^{(n)}$ and $\mathbf{R}^{(n)}=\mathbf{q}-\mathbf{A}\mathbf{u}^{(n)}$.

Then by (5.57) a complete cycle of the iteration consists of the solution of

$$\bar{\mathbf{L}}\bar{\mathbf{U}}\mathbf{d}^{(n)}=\mathbf{R}^{(n)}, \tag{5.64}$$

followed by

$$\mathbf{u}^{(n+1)}=\mathbf{u}^{(n)}+\mathbf{d}^{(n)},$$

which is the iterative refinement described on page 226. Equation (5.64) would, of course, be solved by the forward and backward substitutions

$$\bar{\mathbf{L}}\mathbf{y}^{(n)}=\mathbf{R}^{(n)}$$

and

$$\bar{\mathbf{U}}\mathbf{d}^{(n)}=\mathbf{y}^{(n)}.$$

An additional acceleration parameter ω can also be introduced into the procedure by replacing (5.64) with

$$\bar{\mathbf{L}}\bar{\mathbf{U}}\mathbf{d}^{(n)}=\omega\mathbf{R}^{(n)},$$

as in reference 10. Further details concerning the calculation of α and the solution of the equations are given in Stone's paper, reference 32. His results indicate that the method is economical arithmetically in relation to older methods and that its rate of convergence is much less sensitive to the choice of iteration parameters than are the SOR and ADI methods.

Two recent direct methods

A method for 'variables separable' equations

The following method which depends upon the differential equation being 'variables separable', although that is not immediately obvious, was first proposed by Hockney, 1966, reference 17, who

considered the problem

$$\frac{\partial^2 U}{\partial x^2}+\frac{\partial^2 U}{\partial y^2}=g(x,y)\quad (x,y)\in D,$$
$$U=0,\qquad (x,y)\in C,$$

where C is the boundary of the rectangular domain $D=\{(x,y):0<x<a,\ 0<y<b\}$. Using Fig. 5.9 and a square mesh, the five-point difference equations approximating this problem may be written in partitioned form as

$$\begin{bmatrix}\mathbf{B} & \mathbf{I} & & & \\ \mathbf{I} & \mathbf{B} & \mathbf{I} & & \\ & \mathbf{I} & \mathbf{B} & \mathbf{I} & \\ & & \cdot & \cdot & \cdot \\ & & & \mathbf{I} & \mathbf{B}\end{bmatrix}\begin{bmatrix}\mathbf{u}_1\\ \mathbf{u}_2\\ \mathbf{u}_3\\ \\ \mathbf{u}_M\end{bmatrix}=\begin{bmatrix}\mathbf{b}_1\\ \mathbf{b}_2\\ \mathbf{b}_3\\ \\ \mathbf{b}_M\end{bmatrix} \tag{5.65}$$

where $\mathbf{u}_r$ is the vector of mesh values along $y=rh$, $r=1(1)M$, $\mathbf{b}_r$ is a known vector corresponding to $\mathbf{u}_r$ and the $N\times N$ matrix $\mathbf{B}$ is

$$\begin{bmatrix}-4 & 1 & & \\ 1 & -4 & 1 & \\ & \cdot & \cdot & \cdot \\ & & 1 & -4\end{bmatrix},$$

where N is the number of mesh points along a row parallel to Ox. By (5.65)

$$\begin{aligned}\mathbf{B}\mathbf{u}_1+\mathbf{u}_2&=\mathbf{b}_1\\ \mathbf{u}_1+\mathbf{B}\mathbf{u}_2+\mathbf{u}_3&=\mathbf{b}_2\\ \cdot\quad\cdot\quad\cdot&\\ \mathbf{u}_{M-1}+\mathbf{B}\mathbf{u}_M&=\mathbf{b}_M.\end{aligned} \tag{5.66}$$

Let $\mathbf{q}_r$ be an eigenvector of $\mathbf{B}$ corresponding to the eigenvalue λ_r. Then

$$\mathbf{B}\mathbf{q}_r=\lambda_r\mathbf{q}_r,\quad r=1(1)M,$$

and this set of equations can be written as

$$\mathbf{B}\mathbf{Q}=\mathbf{Q}\operatorname{diag}(\lambda_1,\lambda_2,\ldots,\lambda_M),$$

where $\mathbf{Q}$ is the modal matrix $[\mathbf{q}_1\,\mathbf{q}_2\ldots\mathbf{q}_M]$. But $\mathbf{B}$ is symmetric. therefore the eigenvectors $\mathbf{q}_r$, $r=1(1)M$, can be normalized so that $\mathbf{Q}^T\mathbf{Q}=\mathbf{I}$. Hence $\mathbf{Q}^T\mathbf{B}\mathbf{Q}=\operatorname{diag}(\lambda_1,\lambda_2,\ldots,\lambda_M)=\mathbf{\Lambda}$, say.

Let

$$\bar{\mathbf{u}}_r = \mathbf{Q}^T\mathbf{u}_r \text{ and } \bar{\mathbf{b}}_r = \mathbf{Q}^T\mathbf{b}_r, \tag{5.67}$$

from which it follows that

$$\mathbf{u}_r = \mathbf{Q}\bar{\mathbf{u}}_r \text{ and } \mathbf{b}_r = \mathbf{Q}\bar{\mathbf{b}}_r. \tag{5.68}$$

Substituting from (5.68) into (5.66) and premultiplying throughout with $\mathbf{Q}^T$ leads to the equations

$$\begin{aligned} \mathbf{\Lambda}\bar{\mathbf{u}}_1 + \bar{\mathbf{u}}_2 &= \bar{\mathbf{b}}_1 \\ \bar{\mathbf{u}}_1 + \mathbf{\Lambda}\bar{\mathbf{u}}_2 + \bar{\mathbf{u}}_3 &= \bar{\mathbf{b}}_2 \\ \cdot \quad \cdot \quad &\cdot \\ \bar{\mathbf{u}}_{M-1} + \mathbf{\Lambda}\bar{\mathbf{u}}_M &= \bar{\mathbf{b}}_M. \end{aligned} \tag{5.69}$$

Denote the ith components of $\bar{\mathbf{u}}_r$ and $\bar{\mathbf{b}}_r$ by $\bar{u}_{i,r}$ and $\bar{b}_{i,r}$ respectively and select the ith row of each of the equations (5.69). This gives the tridiagonal system of equations

$$\begin{aligned} \lambda_i\bar{u}_{i,1} + \bar{u}_{i,2} &= \bar{b}_{i,1} \\ \bar{u}_{i,1} + \lambda_i\bar{u}_{i,2} + \bar{u}_{i,3} &= \bar{b}_{i,2} \\ \bar{u}_{i,2} + \lambda_i\bar{u}_{i,3} + \bar{u}_{i,4} &= \bar{b}_{i,3} \\ \cdot \quad \cdot \quad &\cdot \\ \bar{u}_{i,M-1} + \lambda_i\bar{u}_{i,M} &= \bar{b}_{i,M}, \end{aligned} \tag{5.70}$$

for $\bar{u}_{i,r}$, $r = 1(1)M$. All the components of $\bar{\mathbf{u}}_r$, $r = 1(1)M$ can clearly be found by solving N such sets of equations for $\bar{u}_{i,r}$, $i = 1(1)N$. The procedure is therefore:

(i) Calculate the eigenvalues and eigenvectors of **B**. (These are well known for the problem considered. See page 113.)

(ii) Compute $\bar{\mathbf{b}}_r = \mathbf{Q}^T\mathbf{b}_r$.

(iii) Solve equations (5.70), which is easily done.

(iv) Calculate $\mathbf{u}_r = \mathbf{Q}_r\bar{\mathbf{u}}_r$.

This method has been extended to more general self-adjoint 'variables separable' elliptic equations, to problems with derivative boundary conditions, and with irregular boundaries, see references 4 and 5, but research on the method is still relatively recent.

George's dissection method

As mentioned previously the standard Gauss elimination method for solving equations with a large band-width coefficient matrix **A** is inefficient in the sense that zero elements within the band are replaced by non-zero elements that have to be stored in the computer and used at subsequent stages of the elimination.

George, 1973, reference 15, by a combination of analysis, graph theory and intuition, formulated an ordering of the equations that gave substantial reductions in the 'fill-up', the computer storage required and in the volume of the arithmetic of the elimination process. His ordering has since been proved to be virtually optimal. For the five-point difference approximation of a Dirichlet elliptic problem defined over a rectangular region, and with a mesh giving $N\times N$ equations $\mathbf{Au}=\mathbf{b}$, the number of non-zero elements in the final upper triangular matrix $\mathbf{U}$ of $\mathbf{A}=\mathbf{LU}$ is $O(N^2 \log_2 N)$ and the volume of associated arithmetic is $O(N^3)$. The corresponding figures for the natural reading order of the mesh points is $O(N^3)$ and $O(N^4)$. For large N the savings in storage and effort are clearly considerable. A simpler but less efficient ordering has also been given by George (1972) in reference 14.

EXERCISES AND SOLUTIONS

1. The function ϕ satisfies the equation

$$\frac{\partial^2\phi}{\partial x^2}+\frac{\partial^2\phi}{\partial y^2}+2=0$$

at every point inside the square bounded by the straight lines $x=\pm1$, $y=\pm1$, and is zero on the boundary. Calculate a finite-difference solution using a square mesh of side $\frac{1}{2}$. (The non-dimensional form of the torsion problem for a solid elastic cylinder with a square cross-section.)

Assuming the discretization error is proportional to h^2 calculate an improved value of ϕ at the point $(0, 0)$. (The analytical solution value is $0{\cdot}589$.)

Solution

Because of the symmetry there are only three unknowns; ϕ_1 at $(0,0)$, ϕ_2 at $(\frac{1}{2},0)$, ϕ_3 at $(\frac{1}{2},\frac{1}{2})$. The equations are $8\phi_2-8\phi_1+1=0$, $4\phi_3+2\phi_1-8\phi_2+1=0$ and $4\phi_2-8\phi_3+1=0$, giving $\phi_1=0{\cdot}562$, $\phi_2=0{\cdot}438$ and $\phi_3=0{\cdot}344$ to $3D$.

Case (*ii*) $1-4r^2\sin^2\dfrac{k\pi}{N}<0.$

Then

$$|\lambda|^2=\frac{1}{(2r+1)^2}\left\{\left(2r\cos\frac{k\pi}{N}\right)^2+4r^2\sin^2\frac{k\pi}{N}-1\right\}$$

$$=\frac{4r^2-1}{4r^2+4r+1}<1 \quad \text{since} \quad r>0.$$

Therefore the equations are unconditionally stable for all positive r.

Brief introduction to the analytical solution of homogeneous finite-difference equations

Linear equations with constant coefficients

Consider the difference equation

$$u_{j+2}+au_{j+1}+bu_j=0,\ j=0,1,2,\cdots, \tag{3.28}$$

where a and b are real constants.

Assume that

$$u_j=Am^j$$

is a solution, where A and m are non-zero constants. Substitution into (3.28) shows that m is a root of the quadratic equation

$$m^2+am+b=0. \tag{3.29}$$

Case (*i*) Roots real and distinct, $m=m_1$ and $m=m_2$, say. One solution is $u_j=Am_1^j$ and another is $u_j=Bm_2^j$ where A and B are arbitrary constants. As equation (3.28) is linear in u its general solution is

$$u_j=Am_1^j+Bm_2^j.$$

Case (*ii*) Repeated roots, $m=m_1$ twice, say. Clearly one solution is $u_j=Am_1^j$.

Put $u_j=m_1^jf(j)$. Substitution into (3.28) and the use of $a=-2m_1$, $b=m_1^2$ leads to

$$f(j+2)-2f(j+1)+f(j)=0.$$

By inspection it is seen that $f(j)=j$ satisfies this equation. Therefore a second solution of (3.28) is $u_j=Bjm_1^j$. Hence the solution of equation (3.28) in this case is

$$u_j=(A+Bj)m_1^j.$$

Case (*iii*) Complex roots.

Because a and b are real the roots of (3.29) will be conjugate complex numbers, $m_1=re^{i\theta}$ and $m_2=re^{-i\theta}$, say, where $i=\sqrt{(-1)}$.

Hence $a=-r(e^{i\theta}+e^{-i\theta})=-2r\cos\theta$ and $b=r^2$. As in Case (i) the solution of (3.28) is

$$u_j=Ar^je^{ij\theta}+Br^je^{-ij\theta}=r^j\{(A+B)\cos j\theta+i(A-B)\sin j\theta\}.$$

Since A and B are arbitrary constants and $r=b^{\frac{1}{2}}$, this can be written as

$$u_j=b^{\frac{1}{2}j}(C\cos j\theta+D\sin j\theta),$$

where C and D are arbitrary constants and $\cos\theta=-a/2r=-a/2\sqrt{b}$. Methods for deriving particular integrals for non-homogeneous difference equations are given in 'Finite Difference Equations' by H. Levy and F. Lessman. (Pitman).

The eigenvalues and vectors of a common tridiagonal matrix

Let

$$\mathbf{A}=\begin{bmatrix} a & b & & & & \\ c & a & b & & & \\ & c & a & b & & \\ & & & \cdot & & \\ & & & & \cdot & \\ & & & & c & a \end{bmatrix}$$

be a square matrix of order N, where a, b, and c may be real or complex numbers.

Let λ represent an eigenvalue of $\mathbf{A}$ and $\mathbf{v}$ the corresponding eigenvector with components $v_1, v_2, \ldots, v_N$. Then the eigenvalue equation $\mathbf{A}\mathbf{v}=\lambda\mathbf{v}$ gives

$$(a-\lambda)v_1+bv_2=0$$

$$cv_1+(a-\lambda)v_2+bv_3=0$$

$$\cdots$$

$$cv_{j-1}+(a-\lambda)v_j+bv_{j+1}=0$$

and

$$cv_{N-1}+(a-\lambda)v_N=0.$$

If we define $v_0 = v_{N+1} = 0$ then these N equations can be combined into the single difference equation

$$cv_{j-1} + (a-\lambda)v_j + bv_{j+1} = 0, \quad j = 1(1)N. \tag{3.30}$$

As shown, previously, the solution of (3.30) is

$$v_j = Bm_1^j + Cm_2^j, \tag{3.31}$$

where B and C are arbitrary constants and m_1, m_2 are the roots of the equation

$$c + (a-\lambda)m + bm^2 = 0. \tag{3.32}$$

(It is proved later that the roots cannot be equal.)
By equation (3.31) it follows, since $v_0 = v_{N+1} = 0$, that,

$$0 = B + C$$

and

$$0 = Bm_1^{N+1} + Cm_2^{N+1}.$$

Hence

$$\left(\frac{m_1}{m_2}\right)^{N+1} = 1 = e^{i2s\pi}, \quad s = 1(1)N,$$

where $i = \sqrt{(-1)}$. Therefore

$$\frac{m_1}{m_2} = e^{i2s\pi/(N+1)}. \tag{3.33}$$

By equation (3.32),

$$m_1 m_2 = \frac{c}{b},$$

and elimination of m_2 between (3.33) and (3.34) leads to

$$m_1 = \left(\frac{c}{b}\right)^{\frac{1}{2}} e^{is\pi/(N+1)}.$$

Similarly,

$$m_2 = \left(\frac{c}{b}\right)^{\frac{1}{2}} e^{-is\pi/(N+1)}.$$

Again, by equation (3.32),

$$m_1 + m_2 = (\lambda - a)/b,$$

giving that

$$\lambda = a + b\left(\frac{c}{b}\right)^{\frac{1}{2}}\left(e^{is\pi/(N+1)} + e^{-is\pi/(N+1)}\right).$$

Hence the N eigenvalues are given by

$$\lambda_s = a + 2b\left(\frac{c}{b}\right)^{\frac{1}{2}} \cos\frac{s\pi}{N+1}, \quad s = 1(1)N.$$

The jth component of the eigenvector is

$$v_j = Bm_1^j + Cm_2^j = B\left(\frac{c}{b}\right)^{\frac{1}{2}j}\left(e^{ijs\pi/N+1} - e^{-ijs\pi/N+1}\right)$$
$$= 2iB\left(\frac{c}{b}\right)^{\frac{1}{2}j} \sin\frac{js\pi}{N+1},$$

so the eigenvector $\mathbf{v}_s$ corresponding to λ_s can be taken as

$$\mathbf{v}_s^T = \left\{\left(\frac{c}{b}\right)^{\frac{1}{2}} \sin\frac{s\pi}{N+1}, \frac{c}{b}\sin\frac{2s\pi}{N+1}, \left(\frac{c}{b}\right)^{\frac{3}{2}} \times \sin\frac{3s\pi}{N+1}, \ldots, \left(\frac{c}{b}\right)^{\frac{N}{2}} \sin\frac{Ns\pi}{N+1}\right\}.$$

It is easily shown that the roots of equation (3.32) cannot be equal because if we assume $m_1 = m_2$ the solution of (3.32) is then

$$v_j = (B + Cj)m_1^j$$

and $v_0 = v_{N+1} = 0$ implies that $B = C = 0$, giving $\mathbf{v} = 0$, which is not possible.

An analytical solution of the classical explicit approximation to $\partial U/\partial t = \partial^2 U/\partial x^2$

Consider the equation

$$\frac{\partial U}{\partial t} = \frac{\partial^2 U}{\partial x^2}, \quad 0 < x < 1,$$

where $U = 0$ at $x = 0$ and 1, $t > 0$, and U is known when $t = 0$, $0 \leqslant x \leqslant 1$.

The classical explicit approximation to the differential equation is

$$u_{i,j+1} = ru_{i-1,j} + (1-2r)u_{i,j} + ru_{i+1,j}, \tag{3.35}$$

change of the momentum of the object struck during the time of the stroke. Let the change in velocity of the point in the interval Δx be equal to v, where v is the initial velocity. Under the assumption that the initial velocity v is constant in Δx, then we obtain the change of momentum,

$$\rho v \Delta x = I ,$$

where ρ is the linear density of the string. Consequently, we must solve the wave equation with the initial velocity,

$$\psi_{\text{int}} = v = \frac{I}{\rho \Delta x} \qquad x, x + \Delta x ,$$

$$\psi_{\text{ext}} = 0 \qquad x, x + \Delta x ,$$

for the initial displacement.

The lengthening obtained by the action of the impulse can be described by a trapezoid whose lower base equals $(2at) + \Delta x$ and whose upper base equals $(2at) - \Delta x$, for $t > \Delta x/2a$. Obviously, the quantity $I/\Delta x = I_0$ can be interpreted as the impulse density. As $\Delta x \to 0$ the following results for the form of the displacement: the lengthening is equal to zero everywhere outside the interval $(x - at,\ x + at)$, and inside it is equal to $1/2a \cdot 1/\rho$. Loosely speaking, one can say that the displacement is produced by the point impulse I.

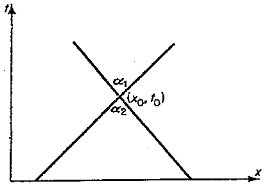


FIG. 10.

We consider now the x, t phase plane (Figure 10) and place the two characteristics through (x_0, t_0):

$$x - at = x_0 - at_0$$
$$x + at = x_0 + at_0 .$$

They determine two angles α_1 and α_2, the so-called upper and lower characteristic angles at the point (x_0, t_0).

The action of a point impulse at the point (x_0, t_0) produces a lengthening which in the interior of the above characteristic angles equals $1/2a \cdot 1/\rho$ and outside the interval equals zero.

Of interest to us now is the region in which the solution is uniquely defined by the initial conditions when these are prescribed in a given interval PQ of the lines $t = 0$.

Formula (2-2.9) shows that it suffices for the determination of the function u at any point $M(x, t)$ of the x, t phase plane (Figure 7) when the initial conditions in the interval PQ are known. Thus, P, Q are the points of the x axis with the coordinates $x - at$ and $x + at$. The segments MP and MQ of the characteristics passing through the point M and the segment PQ of the x axis form a triangle MPQ called the characteristic triangle of the point M.

If the initial conditions are not given on the entire line $-\infty < x < \infty$ but are given only in a fixed interval PQ, then these initial conditions define the solution uniquely within the characteristic triangle which has the interval PQ as a base.

3. Stability of the solution

The solution of Eq. (2-2.1) is uniquely determined by the initial conditions (2-2.2). We shall prove that between this solution and the initial conditions exists a continuous dependence and, in fact, we have the theorem:

For each time interval $0 \leqq t \leqq t_0$ and for arbitrary ε there exists a number $\delta(\varepsilon, t_0)$ such that two solutions $u_1(x, t)$ and $u_2(x, t)$ of Eq. (2-2.1) differ from each other by an amount less than ε:

$$| u_1(x, t) - u_2(x, t) | < \varepsilon , \qquad 0 \leqq t \leqq t_0 ,$$

provided that the initial values

$$u_1(x, 0) = \varphi_1(x) \qquad \text{and} \qquad u_2(x, 0) = \varphi_2(x)$$
$$\frac{\partial u_1}{\partial t}(x, 0) = \psi_1(x) \qquad\qquad \frac{\partial u_2}{\partial t}(x, 0) = \psi_2(x)$$

differ from each other by an amount less than δ:

$$| \varphi_1(x) - \varphi_2(x) | < \delta , \qquad | \psi_1(x) - \psi_2(x) | < \delta . \qquad (2\text{-}2.11)$$

The proof of this theorem is surprisingly simple. The functions $u_1(x, t)$ and $u_2(x, t)$ are linked t the initial values by formula (2-2.9), so that

$$| u_1(x, t) - u_2(x, t) | \leqq \frac{| \varphi_1(x + at) - \varphi_2(x + at) |}{2}$$
$$+ \frac{| \varphi_1(x - at) - \varphi_2(x - at) |}{2} + \frac{1}{2a} \int_{x-at}^{x+at} | \psi_1(\alpha) - \psi_2(\alpha) | \, d\alpha ,$$

whereas on the basis of the inequality (2-2.11), there follows

$$| u_1(x, t) - u_2(x, t) | \leqq \frac{\delta}{2} + \frac{\delta}{2} + \frac{1}{2a} \delta \cdot 2at \leqq \delta(1 + t_0) .$$

Hence our assertion is proved if we take

$$\delta = \frac{\varepsilon}{1 + t_0} .$$

Every physically defined process must be capable of description through functions which depend continuously on those initial conditions determining the process. If the solution of a boundary-value problem depends continuously on the initial conditions, then one also says that the boundary-value problem is well set or the solution is stable.

If this continuous dependence did not exist, there could be two essentially different processes corresponding to practically the same set of initial conditions (whose difference lies within the limits of the accuracy of measurement); that is, the solution would not be stable. It cannot be asserted that such processes are determined by the initial conditions (in a physical sense). From the above theorem, it follows that the vibrations of a string are deter-

[redacted]

TRO TO COMPUTATIONAL
FLUID · MECHANICS , C.Y. CHOW
JOHN WILEY PUBLISHERS

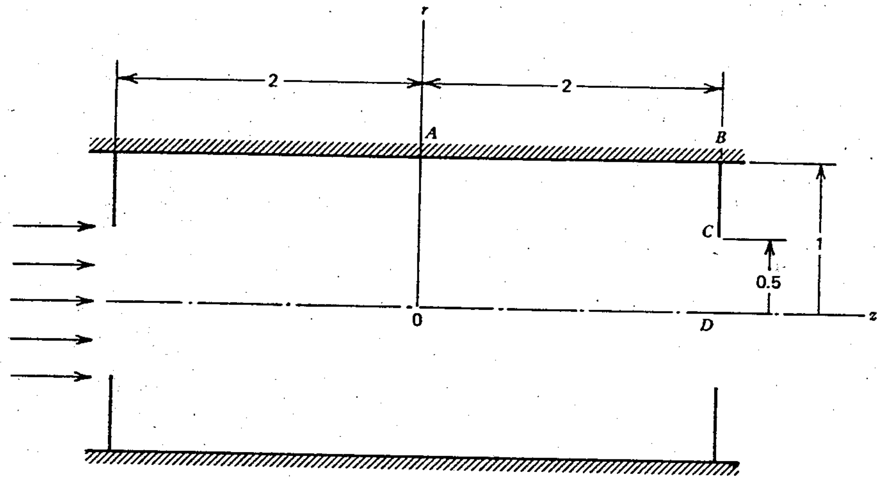


FIGURE 2.11.9 Axisymmetric flow through a tube containing repeated partitions.

metric tube containing repeated partitions (see Fig. 2.11.9). Choose a square mesh of size $h = 0.1$ for numerical computation.

Around the boundary of this region, $\psi = 0$ along OD, $\psi = 1$ along ABC, and $\partial\psi/\partial x = 0$ along both OA and CD. The derivative boundary conditions are the result of symmetry of stream function about these two sections.

2.12 Numerical Solution of Hyperbolic Partial Differential Equations

Problems concerning wave motions in fluid mechanics are governed by hyperbolic partial differential equations. One example mentioned in Section 2.9 is the supersonic flow past a thin body whose governing equation is (2.9.2). Another commonly cited example is the propagation of a one-dimensional sound wave of small amplitude, described by (see Liepmann and Roshko, 1957, p. 68)

$$\frac{\partial^2 u}{\partial t^2} = a^2 \frac{\partial^2 u}{\partial x^2} \qquad (2.12.1)$$

in which t is time, x is the coordinate in the direction of wave propagation, a is the speed of sound treated as constant in the linearized analysis, and u is the fluid speed. It can be shown that density, pressure, and temperature are all governed by equations of the same form.

In this section a numerical technique is developed for solving (2.12.1) to find u at any time $t > 0$ in the spatial domain $0 \leqslant x \leqslant L$, provided that the initial conditions of u are given at $t = 0$ and are expressed in the following form, with functions f and g to be specified for a particular problem.

$$u(x, 0) = f(x) \qquad (2.12.2)$$

$$\frac{\partial u}{\partial t}(x, 0) = g(x) \qquad (2.13.3)$$

Boundary conditions are to be specified at both ends of the gaseous domain, say within a channel of constant cross-sectional area. If one end of the channel is enclosed by a rigid wall, then u must be zero there at all times. On the other hand, at an end that opens to the atmosphere, the pressure there must be a constant or, alternatively, $\partial u/\partial x$ must vanish at that section.

To solve this mixed initial-boundary-value problem numerically, we divide the spatial range of the domain into small intervals of length h and the time axis into intervals of size τ. The total number of vertical grid lines is m, whereas that of horizontal grid lines can be as many as needed in a particular computation. Lines and points in the grid system are named according to Fig. 2.12.1.

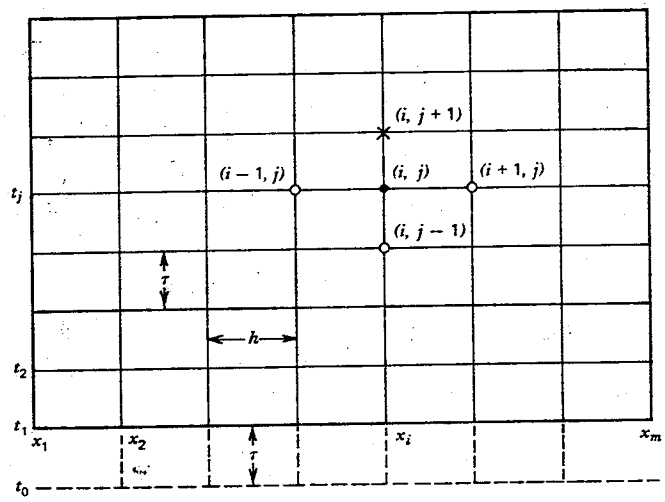


FIGURE 2.12.1 Grid system for numerical computation.

A difference equation can be derived following exactly the same procedure as that used to obtain the numerical scheme (2.10.2) for solving the Poisson equation. Using the central-difference formula to approximate the derivatives in (2.12.1), we obtain, after regrouping,

$$u_{i,j+1} = 2u_{i,j} + C^2(u_{i-1,j} - 2u_{i,j} + u_{i+1,j}) - u_{i,j-1} \tag{2.12.4}$$

where C is a dimensionless parameter called the *Courant number*, defined by

$$\boxed{C \equiv \frac{a\tau}{h}} \tag{2.12.5}$$

(2.12.4) computes the solution at a certain time level based on the solutions at two previous time levels.

Actually, various numerical schemes can be constructed for solving the same partial differential equation by using different finite-difference approximations. The applicability of a numerical scheme is determined by whether it is stable, that is, whether the numerical solution grows and becomes unbounded after repeatedly applying the scheme. In a way we have proved in Section 2.10 that Richardson's iterative formula is stable, and so is Liebmann's. Here we will use a different approach to find the conditions under which (2.12.4) is computationally stable. It turns out that the stability of this numerical scheme is determined by the magnitude of C.

Following *von Neumann's stability analysis*, we assume that time and space variables are separable and that the solution to (2.12.4) can be expanded in the form of a Fourier series. A representative Fourier component may be written

$$u_{i,j} = U_j \, e^{Iikh} \tag{2.12.6}$$

where U_j is the amplitude at t_j of the wave component whose wave number is k, and $I = \sqrt{-1}$. Similarly,

$$u_{i,j\pm1} = U_{j\pm1} \, e^{Iikh}, \qquad u_{i\pm1,j} = U_j \, e^{I(i\pm1)kh}$$

Substituting this into (2.12.4) gives, after canceling the common factor e^{Iikh},

$$U_{j+1} = 2U_j + C^2 U_j(e^{-Ikh} + e^{Ikh} - 2) - U_{j-1}$$

By using the identity that $(e^{I\theta} + e^{-I\theta})/2 = \cos\theta$, it becomes

$$U_{j+1} = AU_j - U_{j-1} \tag{2.12.7}$$

where $A \equiv 2[1 - C^2(1 - \cos kh)]$. By introduction of an *amplification factor* λ such that

$$U_j = \lambda U_{j-1} \qquad \text{and} \qquad U_{j+1} = \lambda U_j = \lambda^2 U_{j-1} \tag{2.12.8}$$

(2.12.7) is reduced to

$$\lambda^2 - A\lambda + 1 = 0 \tag{2.12.9}$$

whose roots are

$$\lambda = \frac{A}{2} \pm \sqrt{\left(\frac{A}{2}\right)^2 - 1} \tag{2.12.10}$$

For $|A| \geq 2$, the roots are real, but their magnitudes are $|\lambda| \geq 1$; for $|A| < 2$ the magnitudes of the complex roots are less than 1. An inspection of (2.12.8) concludes that the amplitude grows indefinitely with increasing time unless $|\lambda| \leq 1$. Thus the inequality that $|A| \leq 2$ or, $A^2 \leq 4$, determines the condition for stability; that is, for stability,

$$[1 - C^2(1 - \cos kh)]^2 \leq 1$$

After expanding the left side and rearranging, we obtain

$$C^2 \leq \frac{2}{1 - \cos kh}$$

When $\cos kh$ varies from -1 to $+1$, the function on the right-hand side varies from 1 to infinity, of which the lowest value is chosen to insure stability. Therefore the stability criterion for the numerical scheme (2.12.4) is $C^2 \leq 1$, or

$$\frac{a\tau}{h} \leq 1 \tag{2.12.11}$$

To arrive at this expression, we have used the fact that each of the three variables on its left is positive. This relationship implies that τ and h cannot be chosen independently.

If the Courant number is chosen to be

$$\frac{a\tau}{h} = 1 \tag{2.12.12}$$

(2.12.4) takes an especially simple form.

$$u_{i,j+1} = u_{i-1,j} + u_{i+1,j} - u_{i,j-1} \tag{2.12.13}$$

As shown in Fig. 2.12.1, this equation states that the value of u at a grid point marked by a cross is computed from the values already computed at three circled grid points at two previous time steps. The numerical scheme (2.12.13) is commonly referred to as the *leapfrog method*. It will now be proved that this numerical method actually gives the exact solution to the differential equation (2.12.1).

It is well known that the solution to (2.12.1) satisfying initial conditions (2.12.2) and (2.12.3) is

$$u(x, t) = \tfrac{1}{2}[f(x + at) + f(x - at)] + \frac{1}{2a}\int_{x-at}^{x+at} g(v)\, dv \tag{2.12.14}$$

which can be verified by substituting it back into each of those equations. The

solution may be written in the simpler functional form

$$u(x,t) = F(x-at) + G(x+at) \tag{2.12.15}$$

where the functions F and G represent simple waves propagating without changing shape along the positive and negative x directions at constant speed a. The lines of slope $dx/dt = \pm a$ in the x-t plane, which trace the progress of the waves, are called the characteristics of the wave equation (see Liepmann and Roshko, 1957, p. 69).

When applied at a grid point (x_i, t_j), (2.12.15) becomes

$$u_{i,j} = F(x_i - at_j) + G(x_i + at_j)$$

From Fig. 2.12.1 we have

$$x_i = x_1 + (i-1)h \qquad \text{and} \qquad t_j = t_1 + (j-1)\tau$$

so that

$$u_{i,j} = F(\alpha + ih - ja\tau) + G(\beta + ih + ja\tau)$$

in which $\alpha \equiv (x_1 - h) - a(t_1 - \tau)$ and $\beta \equiv (x_1 - h) + a(t_1 - \tau)$. With $a\tau = h$, obtained from the condition that $C = 1$, it reduces to

$$u_{i,j} = F[\alpha + (i-j)h] + G[\beta + (i+j)h]$$

According to this relation, the right-hand side of (2.12.13) is rewritten

$$\begin{aligned} u_{i-1,j} + u_{i+1,j} - u_{i,j-1} &= F[\alpha + (i-j-1)h] + G[\cancel{\beta + (i+j-1)h}] \\ &\quad + F[\cancel{\alpha + (i-j+1)h}] + G[\beta + (i+j+1)h] \\ &\quad - F[\cancel{\alpha + (i-j+1)h}] - G[\cancel{\beta + (i+j-1)h}] \\ &= F[\alpha + (i-j-1)h] + G[\beta + (i+j+1)h] \end{aligned}$$

which is exactly $u_{i,j+1}$ or the left-hand side of (2.12.13). It follows that exact solution is computed for (2.12.1) by the leapfrog scheme (2.12.13).

Having introduced the concept of characteristics, we are now in a position to interpret the physical meaning of the stability criterion (2.12.11) by use of Fig. 2.12.2. An examination of (2.12.4) reveals that the solution at the grid point P is influenced by the solution at each of the grid points at previous time steps contained within two diagonals PQ and PR of slope $(dx/dt)_n = \pm h/\tau$. Thus the region $PQRP$ is the domain of dependence of point P in the numerical computation. If Pq and Pr are the backward characteristics of slope $(dx/dt)_c = \pm a$ passing through P, and if $a < h/\tau$ or, equivalently, if $|(dx/dt)_c| < |(dx/dt)_n|$, these lines will lie between PQ and PR, as shown in the figure. However, from the theory of characteristics, it is known that point P can receive signals only from the region $PqrP$, which is its physical domain of dependence. In the present case of $a\tau/h < 1$, in which the computational domain of dependence contains the physical domain of dependence, all the

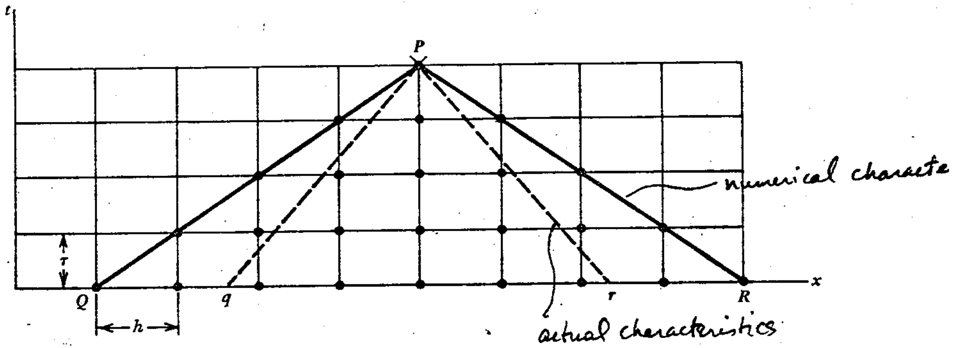


FIGURE 2.12.2 Physical interpretation of stability criterion (2.12.11).

information required to determine the condition at P is included in the computation so that the numerical scheme is stable. The result is inaccurate because of the inclusion of some unnecessary information originating from the region between PQ and Pq and the region between PR and Pr. If $a\tau/h > 1$, the characteristics Pq and Pr would be drawn outside of PQ and PR. In this case only a part of the needed information is used to determine the solution at P, and the computation is unstable. It becomes obvious that when $a\tau/h = 1$ (i.e. when the computational and physical domains of dependence coincide), the numerical solution is the exact solution.

In using the formula (2.12.13) information is needed at two previous time steps. It cannot be used directly at the initial stage to compute the solution at t_2, since conditions are specified only at the initial instant $t_1 = 0$. To help start the numerical procedure, we construct in Fig. 2.12.1 a row of fictitious grid points at $t_0 = t_1 - \tau$, and then rewrite the initial conditions (2.12.2) and (2.12.3) in index notation:

$$u_{i,1} = f_i \tag{2.12.16}$$

$$u_{i,0} = u_{i,2} - 2\tau g_i \tag{2.12.17}$$

in which f_i and g_i represent, respectively, $f(x_i)$ and $g(x_i)$. In obtaining the second expression we have approximated $\partial u/\partial t$ by the central-difference form (2.2.8). For $j = 1$ and with substitution from the preceding equations, (2.12.13) becomes

$$\begin{aligned} u_{i,2} &= u_{i-1,1} + u_{i+1,1} - u_{i,0} \\ &= f_{i-1} + f_{i+1} - u_{i,2} + 2\tau g_i \end{aligned}$$

or

$$u_{i,2} = \tfrac{1}{2}(f_{i-1} + f_{i+1}) + \tau g_i, \qquad i = 2, \cdots, m-1 \tag{2.12.18}$$

This is called the starting formula for (2.12.13).

EGM 6422

LECTURE 12/13

what about the wave equation $\frac{\partial^2 u}{\partial t^2} = \alpha^2 \frac{\partial^2 u}{\partial x^2}$

using centered diff in both dir given

$$\frac{u_{ij+1} - 2u_{ij} + u_{ij-1}}{\Delta t^2} = \alpha^2 \frac{u_{i+1j} - 2u_{ij} + u_{i-1,j}}{\Delta x^2}$$

c^2 — (courant number c)

or $$u_{ij+1} = \alpha^2 \frac{\Delta t^2}{\Delta x^2}\left[u_{i+1j} - 2u_{ij} + u_{i-1j}\right] + 2u_{ij} - u_{ij-1}$$

$$= 2(1-c^2)u_{ij} + c^2 u_{i+1j} + c^2 u_{i-1j} - u_{ij-1}$$

$$\begin{array}{cc} \begin{bmatrix} c^2 & 2(1-c^2) & c^2 \\ & & \\ & & \end{bmatrix} \begin{bmatrix} u_{i-1} \\ u_i \\ u_{i+1} \end{bmatrix} & - \begin{bmatrix} u_i \end{bmatrix} \\ j & j-1 \\ & \text{timestep} \end{array}$$

$$W^{j+1} = \underset{\sim}{A} W^{(j)} - W^{(j-1)}$$

requires info at 2 time levels to get at 3rd time level

if again we let $u_{ij} = U^j e^{Ik\Delta x}$ and factor out $e^{Ik\Delta x}$

$$U^{j+1} = 2U^j + c^2 U^j\left[e^{Ik\Delta x} + e^{-Ik\Delta x} - 2\right] - U^{j-1}$$

$$= U^j \cdot \left\{2 - 2c^2 + 2c^2 \cos k\Delta x\right\} - U^{j-1}$$

$$U^{j+1} = U^j \cdot \underbrace{2\left\{1 - c^2(1 - \cos k\Delta x)\right\}}_{A} - U^{j-1}$$

now assume $U^{j+1} = \lambda U^j$ $U^{j+1} = \lambda^2 U^{j-1}$ $U^j = \lambda U^{j-1}$

$U^{j-1}(\lambda^2 - \lambda A + 1) = 0$ $\lambda = \frac{A \pm \sqrt{A^2 - 4}}{2}$

for $|A| < 2$ then $|\lambda| < 1$ so not initial error will grow

change of the momentum of the object struck during the time of the stroke. Let the change in velocity of the point in the interval Δx be equal to v, where v is the initial velocity. Under the assumption that the initial velocity v is constant in Δx, then we obtain the change of momentum,

$$\rho v \Delta x = I,$$

where ρ is the linear density of the string. Consequently, we must solve the wave equation with the initial velocity,

$$\psi_{\text{int}} = v = \frac{I}{\rho \Delta x} \qquad x, x + \Delta x,$$

$$\psi_{\text{ext}} = 0 \qquad x, x + \Delta x,$$

for the initial displacement.

The lengthening obtained by the action of the impulse can be described by a trapezoid whose lower base equals $(2at) + \Delta x$ and whose upper base equals $(2at) - \Delta x$, for $t > \Delta x/2a$. Obviously, the quantity $I/\Delta x = I_0$ can be interpreted as the impulse density. As $\Delta x \to 0$ the following results for the form of the displacement: the lengthening is equal to zero everywhere outside the interval $(x - at, x + at)$, and inside it is equal to $1/2a \cdot 1/\rho$. Loosely speaking, one can say that the displacement is produced by the point impulse I.

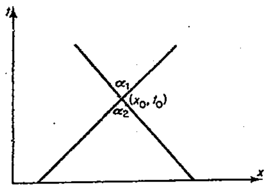


FIG. 10.

We consider now the x, t phase plane (Figure 10) and place the two characteristics through (x_0, t_0):

$$x - at = x_0 - at_0$$
$$x + at = x_0 + at_0 .$$

They determine two angles α_1 and α_2, the so-called upper and lower characteristic angles at the point (x_0, t_0).

The action of a point impulse at the point (x_0, t_0) produces a lengthening which in the interior of the above characteristic angles equals $1/2a \cdot 1/\rho$ and outside the interval equals zero.

Of interest to us now is the region in which the solution is uniquely defined by the initial conditions when these are prescribed in a given interval PQ of the lines $t = 0$.

Formula (2-2.9) shows that it suffices for the determination of the function u at any point $M(x, t)$ of the x, t phase plane (Figure 7) when the initial conditions in the interval PQ are known. Thus, P, Q are the points of the x axis with the coordinates $x - at$ and $x + at$. The segments MP and MQ of the characteristics passing through the point M and the segment PQ of the x axis form a triangle MPQ called the characteristic triangle of the point M.

If the initial conditions are not given on the entire line $-\infty < x < \infty$ but are given only in a fixed interval PQ, then these initial conditions define the solution uniquely within the characteristic triangle which has the interval PQ as a base.

3. Stability of the solution

The solution of Eq. (2-2.1) is uniquely determined by the initial conditions (2-2.2). We shall prove that between this solution and the initial conditions exists a continuous dependence and, in fact, we have the theorem:

For each time interval $0 \leqq t \leqq t_0$ and for arbitrary ε there exists a number $\delta(\varepsilon, t_0)$ such that two solutions $u_1(x, t)$ and $u_2(x, t)$ of Eq. (2-2.1) differ from each other by an amount less than ε:

$$| u_1(x, t) - u_2(x, t) | < \varepsilon , \qquad 0 \leqq t \leqq t_0 ,$$

provided that the initial values

$$u_1(x, 0) = \varphi_1(x) \qquad \text{and} \qquad u_2(x, 0) = \varphi_2(x)$$
$$\frac{\partial u_1}{\partial t}(x, 0) = \psi_1(x) \qquad\qquad \frac{\partial u_2}{\partial t}(x, 0) = \psi_2(x)$$

differ from each other by an amount less than δ:

$$| \varphi_1(x) - \varphi_2(x) | < \delta , \qquad | \psi_1(x) - \psi_2(x) | < \delta . \tag{2-2.11}$$

The proof of this theorem is surprisingly simple. The functions $u_1(x, t)$ and $u_2(x, t)$ are linked t the initial values by formula (2-2.9), so that

$$| u_1(x, t) - u_2(x, t) | \leqq \frac{| \varphi_1(x + at) - \varphi_2(x - at) |}{2}$$
$$+ \frac{| \varphi_1(x - at) - \varphi_2(x - at) |}{2} + \frac{1}{2a} \int_{x-at}^{x+at} | \psi_1(\alpha) - \psi_2(\alpha) | \, d\alpha ,$$

whereas on the basis of the inequality (2-2.11), there follows

$$| u_1(x, t) - u_2(x, t) | \leqq \frac{\delta}{2} + \frac{\delta}{2} + \frac{1}{2a} \delta \cdot 2at \leqq \delta(1 + t_0) .$$

Hence our assertion is proved if we take

$$\delta = \frac{\varepsilon}{1 + t_0} .$$

Every physically defined process must be capable of description through functions which depend continuously on those initial conditions determining the process. If the solution of a boundary-value problem depends continuously on the initial conditions, then one also says that the boundary-value problem is well set or the solution is stable.

If this continuous dependence did not exist, there could be two essentially different processes corresponding to practically the same set of initial conditions (whose difference lies within the limits of the accuracy of measurement); that is, the solution would not be stable. It cannot be asserted that such processes are determined by the initial conditions (in a physical sense). From the above theorem, it follows that the vibrations of a string are deter-

if $|A| < 2$ $\quad \lambda = \frac{A \pm Bi}{2} = \frac{A \pm i\sqrt{4-A^2}}{2} \qquad |\ \ |^2 = \frac{A^2 + (\sqrt{4-A^2})^2}{4} = 1$

real part ≤ 1

$$|A| < 2 \Rightarrow 4\left\{1 - c^2(1-\cos k\Delta x)\right\}^2 \le 4 \quad \Rightarrow \quad \left\{1 - C^2(1-\cos k\Delta x)\right\}^2 \le 1$$

$$1 - 2C^2(1-\cos\ ) + C^4(1-\cos)^2 \le 1$$

$$\underbrace{C^2(1-\cos)}_{+}\left[-2 + C^2(1-\cos\ )\right] \le 0$$

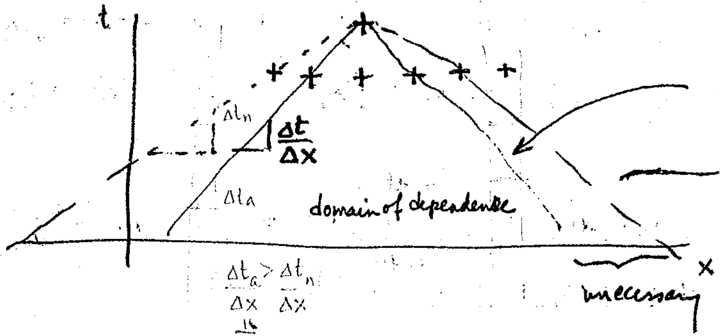
$$C^2 \le \frac{2}{1-\cos k\Delta x}$$

$C^2 \le 1$ to ∞ for $-1 \le \cos k\Delta x \le 1$

choose $C^2 \le 1$ for stability $\Rightarrow -1 \le C \le 1$

<u>TAKE</u> $\frac{\alpha\Delta t}{\Delta x} \le 1$ $\qquad$ take $\frac{\alpha\Delta t}{\Delta x} = 1$

$\Rightarrow$ as long as we stay within the physical characteristics we are stable but not accurate



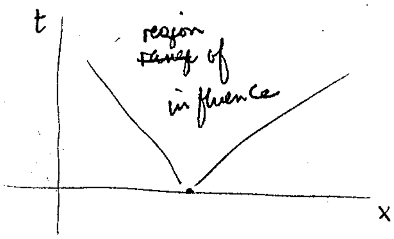
$\frac{\Delta t}{\Delta x} = \pm\frac{1}{\alpha} \qquad \alpha\frac{\Delta t}{\Delta x} = \pm 1$

$\frac{\alpha\Delta t}{\Delta x} < 1$

contains info that is ~~inaccurate~~ unnecessary info originating from between.

IF $\frac{\alpha\Delta t}{\Delta x} < 1$ computational domain of dep. contains physical domain of dep., is stable but is <u>inaccurate</u> 'cause it contains more info than nec.

IF $\frac{\alpha\Delta t}{\Delta x} > 1$ computation domain is smaller than phys. domain ~~&~~ numerical scheme is unstable.



let $\frac{\partial^2 u}{\partial t^2} = a^2 \frac{\partial^2 u}{\partial x^2}$

$$\frac{dt}{dx} = \pm\frac{1}{a} \qquad \alpha t \pm x = C_1 , C_2$$

let $u(x,t) = F(x-at) + G(x+at)$

$$u(x,0) = F(x) + G(x) = f(x)$$

$$\frac{\partial u}{\partial t}(x,0) = \left(\frac{\partial F}{\partial t} + \frac{\partial G}{\partial t}\right)\bigg|_{t=0} = -aF'(x) + aG'(x) = g(x)$$

$$F'(x) = G'(x) - \frac{1}{a}g(x)$$

$$\left.\begin{aligned} F(x) &= G(x) - \frac{1}{a}\int^x g(\sigma)d\sigma \\ F(x) &= -G(x) + f(x) \end{aligned}\right\} \quad \begin{aligned} 2F(x) &= f(x) - \frac{1}{a}\int^x g(\sigma)d\sigma \\ F(x) &= \frac{1}{2}f(x) - \frac{1}{2a}\int^x g(\sigma)d\sigma \end{aligned}$$

$$G'(x) = F'(x) + \frac{1}{a}g(x)$$

$$\left.\begin{aligned} G(x) &= F(x) + \int^x \frac{1}{a}g(\sigma)d\sigma \\ G(x) &= -F(x) + f(x) \end{aligned}\right\} \quad \begin{aligned} 2G(x) &= f(x) + \frac{1}{a}\int^x g(\sigma)d\sigma \\ G(x) &= \frac{1}{2}f(x) + \frac{1}{2a}\int^x g(\sigma)d\sigma \end{aligned}$$

$$F(x-at) = \frac{1}{2}f(x-at) - \frac{1}{2\alpha}\int^{x-at} g(\sigma)d\sigma = \frac{1}{2}f(x-at) + \frac{1}{2a}\int_{x-at} g(\sigma)d\sigma$$

$$G(x+at) = \frac{1}{2}f(x+at) + \frac{1}{2\alpha}\int^{x+at} g(\sigma)d\sigma$$

$$u(x,t) = \frac{1}{2}\left[f(x-at) + f(x+at)\right] + \frac{1}{2\alpha}\int_{x-at}^{x+at} g(\sigma)d\sigma$$

if $\frac{a\Delta t}{\Delta x} = 1$ then $u_{i,j+1} = u_{i+1,j} + u_{i-1,j} - u_{i,j-1}$ Leapfrog method gives exact soln.

when ~~since~~ $j=0 \Rightarrow t=0$ initial condition : we need to define $u_{i,-1}$

to do this use $\frac{\partial u}{\partial t}(x, t=0) = g(x)$

use $\frac{\partial u}{\partial t} = \frac{u_{i,1} - u_{i,-1}}{2\Delta t} = g(x_i)$ since this form of $\frac{\partial u}{\partial t}$ is of same order as that of eqns.

$$\therefore \quad u_{i,-1} = u_{i,1} - 2\Delta t\, g(x_i)$$

and at $t=0$

$$u_{i,1} = u_{i+1,0} + u_{i-1,0} \overset{-u_{i,-1}}{-u_{i,1}} + 2\Delta t\, g(x_i)$$

$$\text{or} \quad u_{i,1} = \frac{1}{2}\left(u_{i+1,0} + u_{i-1,0}\right) + \Delta t\, g(x_i)$$

for $j>0$ then we revert to $u_{i,j+1} = u_{i+1,j} + u_{i-1,j} - u_{i,j-1}$

A better approx for $u_{i,-1}$ is to use D'alembert Soln.

$$u(x_i,t_j) = \frac{1}{2}\left[f(x_i+at_j) + f(x_i-at_j)\right] + \frac{1}{2a}\int_{x_i-at_j}^{x_i+at_j} g(\sigma)\,d\sigma$$

now. $u(x_i,\Delta t) = u_{i,1} = \frac{1}{2}\left[f(x_i+a\Delta t) + f(x_i-a\Delta t)\right] + \frac{1}{2a}\int_{x_i-a\Delta t}^{x_i+a\Delta t} g(\sigma)\,d\sigma$

but since $\frac{a\Delta t}{\Delta x} = 1$ this is

$$x_i + a\Delta t = x_i + \Delta x = x_{i+1} + \cancel{O(\Delta t)}\ a.0$$
$$x_i - a\Delta t = x_i - \Delta x = x_{i-1} \cancel{+O(\Delta t)}\ a.0$$

$$u_{i,1} = \frac{1}{2}\left[\underset{u_{i+1,0}}{f(x_{i+1})} + \underset{u_{i-1,0}}{f(x_{i-1})}\right] + \frac{1}{2a}\int_{x_{i-1}}^{x_{i+1}} g(\sigma)\,d\sigma$$

but $u_{i+1,0} = F(x_{i+1}) + G(x_{i+1}) = F(x_{i+1}) + G(x_i$ [illegible] $a\Delta t)$ since $x_{i+1} = x_i + \Delta x = x_i - a\Delta t$ for $G(x_{i+1})$

and $u_{i-1,0} = F(x_{i-1}) + G(x_{i-1})$

$$F(x_{i+1}) + F(x_{i-1}) = F(x_{i+1}) + G(x_{i+1}) + F(x_{i-1}) - G(x_{i+1}) = u_{i+1,0} + u_{i-1,0}$$

$F(x) = \frac{1}{2}f(x) + \frac{1}{2a}\int^x g(\sigma)\,d\sigma$

$G(x) = \frac{1}{2}f(x) - \frac{1}{2a}\int^x g(\sigma)\,d\sigma$

$\therefore u_{i+1,0} = f(x_{i+1}) + \cancel{\frac{1}{2a}\int^{x_{i+1}} g(\sigma)\,d\sigma}$

$u_{i-1,0} = f(x_{i-1}) - \cancel{\frac{1}{2a}\int^{x_{i-1}} g(\sigma)\,d\sigma}$

$$\therefore \quad u_{i,1} = \tfrac{1}{2}\left[u_{i+1,0} + u_{i-1,0}\right] + \frac{1}{2a}\int_{x_{i-1}}^{x_{i+1}} g(\sigma)\,d\sigma$$

same as before IFF* $g(x) = \text{constant}.$ $\Rightarrow \frac{1}{2a}\int_{x_{i-1}}^{x_{i+1}} g(\sigma)\,d\sigma = \frac{1}{2a} g(x)\underbrace{[x_{i+1} - x_{i-1}]}_{2\Delta x} = g(x)\underbrace{\boxed{\frac{\Delta x}{a}}}_{\Delta t}$

since $g(x)$ is const, take it at $x = x_i$

Solution by characteristics

$$\text{if} \quad a u_{xx} + b u_{xt} + c u_{tt} + d u_x + e u_t + f u + g = 0 \qquad (1)$$

$$\text{then} \quad \frac{dt}{dx} = \frac{b \pm \sqrt{b^2 - 4ac}}{2a} \qquad \text{describes characteristic}$$

and if we let $p = \frac{\partial u}{\partial x}$ & $q = \frac{\partial u}{\partial t}$ $\qquad$ $p = p(x,t)$, $q = q(x,t)$ since $u = u(x,t)$

$$\text{then} \quad dp = \frac{\partial^2 u}{\partial x^2}dx + \frac{\partial^2 u}{\partial x \partial t}dt \qquad \Rightarrow \qquad u_{xx} = \frac{dp - u_{xt}\,dt}{dx}$$

$$dq = \frac{\partial^2 u}{\partial x \partial t}dx + \frac{\partial^2 u}{\partial t^2}dt \qquad\qquad u_{tt} = \frac{dq - u_{xt}\,dx}{dt}$$

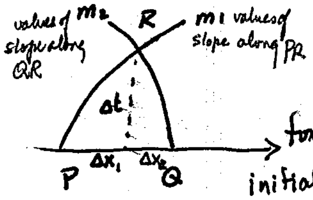
now put into (1) to get

$$\text{thus} \quad u_{xt}\left[a\left(\frac{dt}{dx}\right)^2 - b\left(\frac{dt}{dx}\right) + c\right] - \left[a\frac{dp}{dx}\frac{dt}{dx} + c\frac{dq}{dx} + \underbrace{\{du_x + eu_t + fu + g\}}_{K}\frac{dt}{dx}\right] = 0$$

if we pick the curve such that $\frac{dt}{dx} = \frac{b \pm \sqrt{b^2 - 4ac}}{2a}$ this is same

$$\frac{dt}{dx} = m_1, m_2$$

as saying (1) has become $a\frac{dp}{dx}\frac{dt}{dx} + c\frac{dq}{dx} + K\frac{dt}{dx} = 0.$



for same Δt on the characteristics $\qquad \Delta t = m_1 \Delta x_1 \qquad$ & $\Delta t = m_2 \Delta x_2$

initial data curve

using Euler

solve for x & t
assuming m_1 & m_2 = const

$$\left.\begin{aligned} &\therefore \quad a\,\Delta p_{PR}\, m_1 + c\,\Delta q_{PR} + K\Delta t = 0 \\ &\& \quad a\,\Delta p_{QR}\, m_2 + c\,\Delta q_{QR} + K\Delta t = 0 \end{aligned}\right\} \text{ given } \Delta t$$

solve for Δp & Δq

$p_{new} = p_{old} + \Delta p/2$
$q_{new} = q_{old} + \Delta q/2$

Finally since $u = u(x,t)$ $\qquad du = \frac{\partial u}{\partial x}dx + \frac{\partial u}{\partial t}dt$

$$du = p\,dx + q\,dt$$

$$\text{then} \quad \Delta u = p\Delta x + q\,\Delta t$$

we can improve by using $m_{PR} = [m_1(P) + m_1(R)]/2$ & $[m_2(Q) + m_2(R)]/2$

EGM 6422

LECTURE 23/24

EGM 6422

LECTURE 24/25

$$u_{i+1,0}\begin{cases}\frac{1}{2}f(x_{i+1}) + \frac{1}{2a}\int^{x_{i+1}} g\,dx = F(x_{i+1})\\ \frac{1}{2}f(x_{i-1}) - \frac{1}{2a}\int^{x_{i-1}} g\,dx = G(x_{i+1}) = G(x_i+\Delta x) = G(x_i-a\Delta t)\end{cases}$$

$$u_{i-1,0}\begin{cases}\frac{1}{2}f(x_{i-1}) + \frac{1}{2a}\int^{x_{i-1}} g\,dx = F(x_{i-1})\\ \frac{1}{2}f(x_{i+1}) - \frac{1}{2a}\int^{x_{i+1}} g\,dx = G(x_{i-1}) = G(x_i-\Delta x) = G(x_i+a\Delta t)\end{cases}$$

$$f(x_{i+1}) + f(x_{i-1})$$

11. N.T. Asnani and B.C. Nakra, Vibration analysis of multilayered beams involving visco-elastic layers, *J.Instr. Engr., India*, **50**, 187-193 (1970).

12. S. He, and M.D. Rao, Damping of laminated composite beams with multiple visco-elastic layers, *Proc. 2nd Int. Cong. on Recent Developments in Air- and Structure- Borne Sound and Vibration*, 265-270 (1992).

13. M.D. Rao and S. He, Dynamic analysis and design of laminated composite beams with multiple damping layers, *AIAA Journal*, **31(4)**, 736-745 (1993).

14. C. Levy and Q. Chen, Vibration analysis of the partially covered double sandwich-type cantilever beam, *J. Sound and Vib.*, **177(1)**, 15-29 (1994).

15. H. Zhou and M.D. Rao, Viscoelastic analysis of bounded tubular joints under torsion, *Int. J. Solids Struct.*, **30(16)**, 2199-2211 (1993).

16. S. He and M.D. Rao, Vibration analysis of adhesively bonded lap joint, Pt. I: Theory, Pt. II: Numerical solution, *J. Sound & Vib.*, **152(3)**, 405-425 (1992).

17. Q. Chen and C. Levy, Vibration analysis of a partially covered, double sandwich cantilever beam with concentrated at free end, *Int. J. Solids Struct.*, **31(17)**, 2377-2391 (1994).

18. Q. Chen and C. Levy, Vibration characteristics of partially covered double-sandwich cantilever beam, *AIAA Journal*, **34(12)**, 2622-2626 (1996).

The finite-difference method, with

$$\frac{Tg(\Delta x)^2}{w(\Delta t)^2} = 1,$$

will be found to be the equivalent of integration along the characteristics, lending further support to the likelihood that it will give exact answers.

EXAMPLE 9.3 Solve

$$\frac{\partial^2 u}{\partial t^2} = 2\frac{\partial^2 u}{\partial x^2} - 4,$$

with initial conditions

$$\begin{aligned} u &= 12x && \text{for} \quad 0 \le x \le 0.25, \\ u &= 4 - 4x && \text{for} \quad 0.25 \le x \le 1.0, \\ \frac{\partial u}{\partial t} &= 0 && \text{for} \quad 0 \le x \le 1.0; \end{aligned}$$

$\frac{\partial u}{\partial x} = 12$

$\frac{\partial u}{\partial x} = -4$

boundary conditions are $u = 0$ at $x = 0$ and at $x = 1.0$.

Putting the equation into the standard form,

$$a\frac{\partial^2 u}{\partial x^2} + b\frac{\partial^2 u}{\partial x\,\partial t} + c\frac{\partial^2 u}{\partial t^2} + e = 0,$$

gives $a = -2$, $b = 0$, $c = 1$, and $e = 4$. (The equation is linear since a, b, c, and e are independent of u, u_x, and u_t.)

The slopes of the characteristics are the roots of

$$-2m^2 + 1 = 0,$$

$$m = \pm\frac{\sqrt{2}}{2},$$

$m = \frac{b \pm \sqrt{b^2-4ac}}{2a}$

$m = \frac{0 \pm \sqrt{0^2-4(-2)\cdot 1}}{2(-2)} = \mp\frac{\sqrt{2}}{2} = \frac{\Delta t}{\Delta x}$

so the characteristic curves are straight lines in the xt-plane, as shown in Fig. 9.5. Consider points P, Q, and R—(0.25, 0), (0.75, 0), and (0.5, 0.1768)—and solve Eq. (9.16), which is

$$am_{\text{av}}\,\Delta p + c\,\Delta q + e\,\Delta t = 0.$$

$t_R - t_p = m(\quad x_P)$

$t_R - t_a = m(\quad -x_a)$

Along $P \to R$:

$$-2\left(\frac{\sqrt{2}}{2}\right)\Delta p + \Delta q + 4\Delta t = 0,$$

$$-\sqrt{2}(p_R - p_P) + (q_R - q_P) + 4(0.1768) = 0;$$

Along $Q \to R$:

$$-2\left(\frac{-\sqrt{2}}{2}\right)\Delta p + \Delta q + 4\Delta t = 0,$$

$$\sqrt{2}(p_R - p_Q) + (q_R - q_Q) + 4(0.1768) = 0.$$

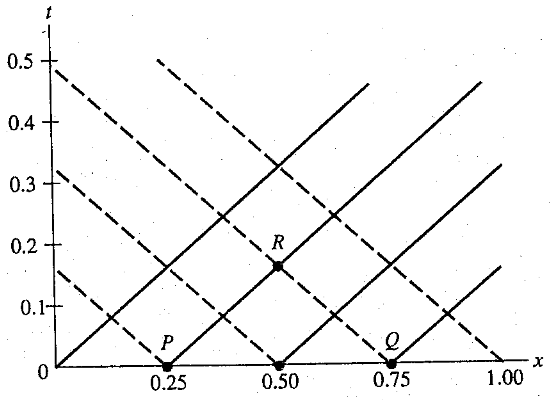


Figure 9.5

Using

$$p_P = \left(\frac{\partial u}{\partial x}\right)_P = -4,^{*} \qquad q_P = \left(\frac{\partial u}{\partial t}\right)_P = 0, \qquad p_Q = \left(\frac{\partial u}{\partial x}\right)_Q = -4,$$

$$q_Q = \left(\frac{\partial u}{\partial t}\right)_Q = 0,$$

we find $p_R = -4, q_R = -\sqrt{2}/2$ by solving the equations $P \rightarrow R$ and $Q \rightarrow R$ simultaneously.

Now we evaluate u at point R through its change along $P \rightarrow R$:

$$\Delta u = p_{\text{av}} \Delta x + q_{\text{av}} \Delta t = \overset{\frac{p_R + p_P}{2}}{-4}(0.25) + \overset{(q_P + q_R)/2}{\left(\frac{0 - \sqrt{2}/2}{2}\right)}(0.1768)$$

$$= -1.0625;$$

$$u_R = \underset{u_P + \Delta u}{3 + (-1.0625)} = 1.9375.$$

(If we compute through evaluating Δu along $Q \rightarrow R$, we get the same result.)

For this simple problem, the finite-difference method is much simpler, and we expect it to give the same results. Following the procedure of Section 9.2,[†] we compute with $\Delta x = 0.25$, $\Delta t = \Delta x/\sqrt{c} = 0.1768$, and obtain Table 9.3. The circled value agrees exactly with that calculated by the method of characteristics.

---

[*] The gradient has a discontinuity at $x = 0.25$. The value of $\partial u/\partial x$ for points to the right of P applies for the region PRQ.

[†] The algorithm is $u_i^{j+1} = (u_{i+1}^j + u_{i-1}^j) - u_i^{j-1} - 4(\Delta t)^2$ with $\Delta t = \Delta x/\sqrt{2}$. For the first time step, $u_i^1 = \frac{1}{2}(u_{i+1}^0 + u_{i-1}^0) - \frac{1}{2}(4)(\Delta t)^2$.

Table 9.3

x	0	0.25	0.5	0.75	1.0
$u(t = 0)$	0.0	3.0	2.0	1.0	0.0
$u(t = 0.1768)$	0.0	0.9375	$\boxed{1.9375}$	0.9375	0.0
$u(t = 0.3535)$	0.0	−1.1875	−0.2500	0.8125	0.0
$u(t = 0.5303)$	0.0	−1.3125	−2.4375	−1.3125	0.0

EXAMPLE 9.4 Solve

$$\frac{\partial^2 u}{\partial t^2} = (1 + 2x)\frac{\partial^2 u}{\partial x^2}$$

over (0, 1) with fixed boundaries and the initial conditions

$$u(x,0) = 0, \qquad \frac{\partial u}{\partial t}(x,0) = x(1 - x).$$

For this problem, $a = -(1 + 2x), b = 0, c = +1, e = 0$. Then $am^2 + bm + c = 0$ gives

$$m = \pm\sqrt{\frac{1}{(1 + 2x)}}.$$

The characteristic curves are found by solving the differential equations $dt/dx = \sqrt{1/(1 + 2x)}$ and $dt/dx = -\sqrt{1/(1 + 2x)}$. Integrating* from the initial point x_0 and t_0, we have

$$t = t_0 + \sqrt{1 + 2x} - \sqrt{1 + 2x_0} \quad \text{from } m_+,$$
$$t = t_0 - \sqrt{1 + 2x} + \sqrt{1 + 2x_0} \quad \text{from } m_-.$$

Figure 9.6 shows several of the characteristic curves. We select two points on the initial curve for $t = 0$, at $P = (0.25, 0)$ and $Q = (0.75, 0)$, whose characteristics intersect at point R. Solving for the intersection, we find $R = (0.4841, 0.1782)$.

We now solve Eq. (9.6) to obtain $p = \partial u/\partial x$ and $q = \partial u/\partial t$ at R:

At point P: $x = 0.25, \quad t = 0, \quad u = 0, \quad p = \left(\frac{\partial u}{\partial x}\right)_P = 0,$

$$q = \left(\frac{\partial u}{\partial t}\right)_P = x - x^2 = 0.1875,$$

$$m = \sqrt{\frac{1}{(1 + 2x)}} = 0.8165,$$

$$a = -(1 + 2x) = -1.5, \quad b = 0, \quad c = +1, \quad e = 0.$$

$t = 0 + \sqrt{1+2x} - \sqrt{1+2(.25)}$

$t = 0 - \sqrt{1+2x} + \sqrt{1+2(.75)}$

$\downarrow$

$t = 0.1782$

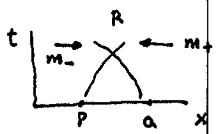
$x = .4841$

At point Q: $x = 0.75, \quad t = 0, \quad u = 0, \quad p = 0, \quad q = x - x^2 = 0.1875,$

$$m = -\sqrt{\frac{1}{(1+2x)}} = -0.6325,$$

$$-(1+2x) = a = -2.5, \quad b = 0, \quad c = +1, \quad e = 0.$$

At point R: $x = 0.4841, \quad t = 0.1783,$



$$m_+ = \sqrt{\frac{1}{(1+2x)}} = 0.7128,$$

$$m_- = -\sqrt{\frac{1}{(1+2x)}} = -0.7128,$$

$$a = -1.9682, \quad b = 0, \quad c = 1, \quad e = 0.$$

Equation (9.16) becomes, when we use average values for a and m,

$$P \to R: \quad \overset{\frac{a_P+a_R}{2}}{-1.7341}(\overset{\frac{m_P+m_R}{2}}{0.7646})(p_R - \overset{p_P}{0}) + \overset{(c_P+c_R)/2}{(1)}(q_R - \overset{q_P}{0.1875}) = 0;$$

$$Q \to R: \quad \overset{(a_Q+a_R)/2}{-2.2341}(\overset{(m_Q+m_R)/2}{-0.6726})(p_R - \overset{p_Q}{0}) + \overset{(c_Q+c_R)/2}{(1)}(q_R - \overset{q_Q}{0.1875}) = 0.$$

Solving simultaneously, we get $p_R = 0, q_R = 0.1875$.

We calculate the change in u along the characteristics:

$$P \to R: \quad \Delta u = \underset{\frac{p_P+p_R}{2}}{0}(\overset{\Delta x = (.4841 - .25) = x_R - x_P}{0.2341}) + 0.1875(\overset{\Delta t = (.1783 - 0) = t_R - t_P}{0.1783}) = 0.0334,$$

$$Q \to R: \quad \Delta u = \underset{(p_Q+p_R)/2}{0}(\overset{(.4841 - .75) = x_R - x_Q}{-0.2659}) + 0.1875(\overset{t_R - t_Q}{0.1783}) = 0.0334.$$

$$u_R = \underset{u_P \,+\, \Delta u_{P\to R} \,\equiv\, u_Q + \Delta u_{Q \to R}}{0 + 0.0334} = 0.0334.$$

Figure 9.6 gives the results at several other intersections of characteristics. Students should verify these results to be sure they understand the method of characteristics.

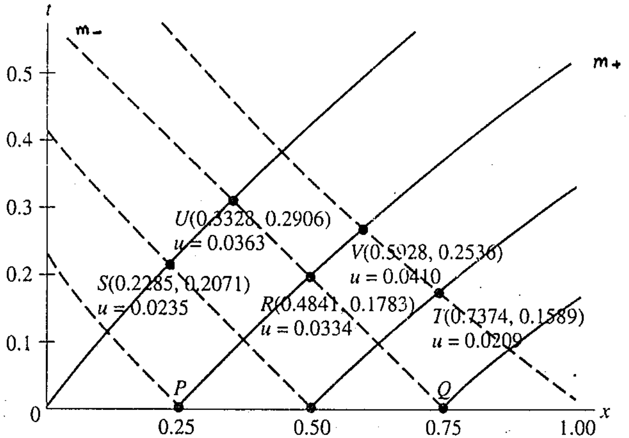


Figure 9.6

EXAMPLE 9.5 Solve the quasilinear equation, with conditions as shown, by numerical integration along the characteristics. (This might be a vibrating string with tension related to the displacement u and subject to an external lateral force.)

$$\frac{\partial^2 u}{\partial x^2} - u\frac{\partial^2 u}{\partial t^2} + (1 - x^2) = 0, \qquad u(x,0) = x(1-x), \qquad u_t(x,0) = 0, = q$$

$$u(0,t) = 0, \qquad u(1,t) = 0. \tag{9.18}$$

We will advance the solution beyond the start from P, at $x = 0.2, t = 0$, and Q, at $x = 0.4, t = 0$, to one new point R. Comparing Eq. (9.18) to the standard form,

$$au_{xx} + bu_{xt} + cu_{tt} + e = 0,$$

we have $a = 1, b = 0, c = -u, e = 1 - x^2$. We first compute u, p, and q at points P and Q,

$$u = x(1-x) = x - x^2$$

(from the initial conditions), so

$$u_P = 0.2(1 - 0.2) = 0.16,$$
$$u_Q = 0.4(1 - 0.4) = 0.24.$$

Also,

$$p = \frac{\partial u}{\partial x} = 1 - 2x$$

(by differentiating the initial conditions), so

$$p_P = 1 - 2(0.2) = 0.6,$$
$$p_Q = 1 - 2(0.4) = 0.2;$$

and

$$q = \frac{\partial u}{\partial t} = 0$$

(from the initial conditions), so

$$q_P = 0,$$
$$q_Q = 0.$$

To locate point R, we need the slope m of the characteristic. Using $am^2 - bm + c = 0$, we get

$$m = \frac{b \pm \sqrt{b^2 - 4ac}}{2a},$$

$$m = \frac{\pm\sqrt{4u}}{2} = \pm\sqrt{u}.$$

Since m depends on the solution u, we will need to find point R through the

predictor–corrector approach. In the first trial, use the initial values over the whole arc; that is, take $m_+ = +m_P$ and $m_- = -m_Q$:

$$m_+ = \sqrt{u_P} = \sqrt{0.16} = 0.4,$$
$$m_- = -\sqrt{u_Q} = -\sqrt{0.24} = -0.490.$$

We now estimate the coordinates of R by solving simultaneously

$$t_R = m_+(x_R - x_P) = 0.4(x_R - 0.2),$$
$$t_R = m_-(x_R - x_Q) = -0.490(x_R - 0.4).$$

These give

$$x_R = 0.310, \qquad t_R = 0.044.$$

We write Eq. (9.16) along each characteristic, again using the initial values of m, since m at R is still unknown:

$$am\,\Delta p + c\,\Delta q + e\,\Delta t = 0, \quad x_P^2+x_R^2 \qquad [(1+x_P^2)+(1-x_R^2)]/2$$

$$(1)\overset{m_+}{(0.4)}(p_R - 0.6) + (-0.16)(q_R - 0) + \left(1 - \frac{0.04 + 0.096}{2}\right)(0.044) = 0, \qquad [(1-x_Q^2)+(1-x_R^2)]/2$$

$$(1)(-0.490)(p_R - 0.2) + (-0.24)(q_R - 0) + \left(1 - \frac{\overset{x_Q^2+x_R^2}{0.16 + 0.096}}{2}\right)(0.044) = 0.$$

In these equations we used the arithmetic average of x^2 in the last terms. Solving simultaneously, we get

$$p_R = 0.399, \qquad q_R = -0.246.$$

As a first approximation for u at R, then,

$$\Delta u = p\,\Delta x + q\,\Delta t,$$
$$u_R - \overset{u_P}{0.16} = \frac{\overset{p_P}{0.6} + \overset{p_R}{0.399}}{2}\overset{x_R - x_P}{(0.310 - 0.2)} + \frac{\overset{q_P + q_R}{0 - 0.246}}{2}\overset{t_R - t_P}{(0.044 - 0)},$$
$$u_R = 0.2095.$$

The last computation was along PR, using average values of p and q. We could have alternatively proceeded along QR. If this is done,

$$u_R - 0.24 = \frac{\overset{p_Q}{0.2} + \overset{p_R}{0.399}}{2}(\overset{x_R}{0.310} \overset{- x_Q}{- 0.4}) + \frac{\overset{q_Q + q_R}{0 - 0.246}}{2}\overset{t_R - t_Q}{(0.044 - 0)},$$
$$u_R = 0.2076. \qquad u_R = \frac{u_{RP}+u_{RQ}}{2}$$

The two values should be close to each other. Let us use the average value, 0.2086, as our initial estimate of u_R. We now repeat the work. In getting the coordinates of R, we now use average values of the slopes,

$$t_R = \frac{\overset{m_P}{0.4} + \overset{m_R}{\sqrt{0.2086}}}{2}(x_R - 0.2),$$
$$t_R = \frac{\overset{m_Q}{-0.490} - \overset{m_R}{\sqrt{0.2086}}}{2}(x_R - 0.4),$$
$$x_R = 0.305, \qquad t_R = 0.045;$$

$a m \Delta p + c \Delta q + e \Delta t$

$$(1)\left(\frac{\overset{m_P}{0.4} + \overset{m_R}{\sqrt{0.2086}}}{2}\right)(p_R - 0.6) - \left(\frac{\overset{\overset{c_P}{m_Q}}{0.16} + \overset{c_R}{0.2086}}{2}\right)(q_R - 0)$$

$$+ \left(1 - \frac{\overset{x_P^2}{0.04} + \overset{x_R^2}{0.0930}}{2}\right)(0.045) = 0,$$

$$(1)\left(\frac{\overset{m_Q}{-0.490} - \overset{m_R}{\sqrt{0.2086}}}{2}\right)(p_R - 0.2) - \left(\frac{\overset{c_Q}{0.24} + \overset{c_R}{0.2086}}{2}\right)(q_R - 0)$$

$$+ \left(1 - \frac{\overset{x_Q^2}{0.16} + \overset{x_R^2}{0.0930}}{2}\right)(0.045) = 0,$$

$$p_R = 0.398, \qquad q_R = -0.242;$$

$$u_R = \overset{u_P}{0.16} + \frac{\overset{p_P}{0.6} \overset{+}{+} \overset{p_R}{0.398}}{2}(0.305 - 0.2) + \frac{\overset{q_P}{0} \overset{+}{-} \overset{q_R}{0.242}}{2}(0.045 - 0),$$

$$u_R = 0.2071 \quad \text{(along } PR\text{)};$$

$$u_R = \overset{u_Q}{0.24} + \frac{\overset{p_Q}{0.2} \overset{+}{+} \overset{p_R}{0.398}}{2}(0.305 - 0.4) + \frac{\overset{q_Q}{0} \overset{+}{-} \overset{q_R}{0.242}}{2}(0.045 - 0),$$

$$u_R = 0.2063 \qquad \text{(along } QR\text{)}.$$

$u_R = \frac{u_{RP} + u_{RQ}}{2} = .2067$

The average value is 0.2067.

Another round of calculations gives $u_R = 0.2066$, which checks the previous value sufficiently. This method is, of course, very tedious by hand.

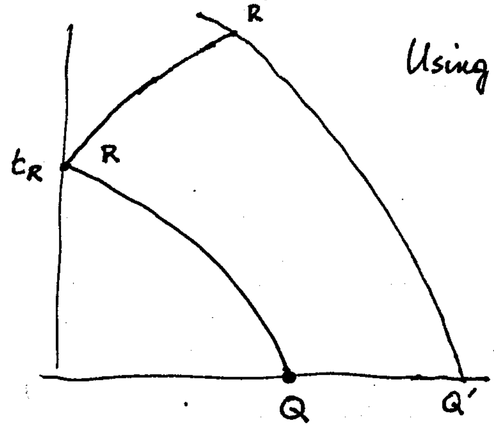
9.5 The Wave Equation in Two Space Dimensions

The finite-difference method can be applied to hyperbolic partial-differential equations in two or more space dimensions. A typical problem is the vibrating membrane. Consider a thin flexible membrane stretched over a rectangular frame and set to vibrating. A development analogous to that for the vibrating string gives

$$\boxed{\frac{\partial^2 u}{\partial t^2} = \frac{Tg}{w}\left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2}\right),}$$

in which u is the displacement, t is the time, x and y are the space coordinates, T is the uniform tension per unit length, g is the acceleration of gravity, and w is the weight per unit area. For simplification, let $Tg/w = \alpha^2$. Replacing each derivative by its central-difference approximation, and using $h = \Delta x = \Delta y$, gives (we recognize the Laplacian on the right-hand side)

$$\frac{u_{i,j}^{k+1} - 2u_{i,j}^{k} + u_{i,j}^{k-1}}{(\Delta t)^2} = \alpha^2 \frac{u_{i+1,j}^{k} + u_{i-1,j}^{k} + u_{i,j+1}^{k} + u_{i,j-1}^{k} - 4u_{i,j}^{k}}{h^2}. \qquad (9.19)$$



Using BC's in solving for data at boundaries

$\frac{dt}{dx} = m_-$; let's assume $u(0,t)=0$ here, i.e., u_R

$$t_R - \overset{0}{t_Q} = m_-\,(\overset{0}{x_R} - \overset{\checkmark}{x_Q}) \quad (1) \text{ this gives } t_R$$

$$\overset{=0}{u_R} - \overset{\checkmark}{u_Q} = \frac{1}{2}(p_R + \overset{\checkmark}{p_Q})(\overset{0}{x_R} - x_Q) + \frac{1}{2}(q_R + \overset{\checkmark}{q_Q})(t_R - \overset{0}{t_Q}) \quad (2)$$

$$0 = \left(\frac{\overset{\sim}{a_R} + \overset{\checkmark}{a_Q}}{2}\right) m_-\,(p_R - \overset{\checkmark}{p_Q}) + \left(\frac{\overset{\sim}{c_R} + \overset{\checkmark}{c_Q}}{2}\right)(q_R - \overset{\checkmark}{q_Q})$$

$$+ \left(\frac{\overset{\sim}{K_R} + \overset{\checkmark}{K_Q}}{2}\right)(\underset{\substack{\uparrow \\ \text{from (1)}}}{t_R} - \overset{0}{t_Q}) \quad (3)$$

from the eqs (2) and (3) then we can find p_R + q_R $\sim$ values are assumed

To continue : note that we have found t_R , $\overset{BC}{u_R=0}$, p_R, q_R : we can use R + Q' to find R' in the usual manner

EGM 6422

LECTURE 26/27

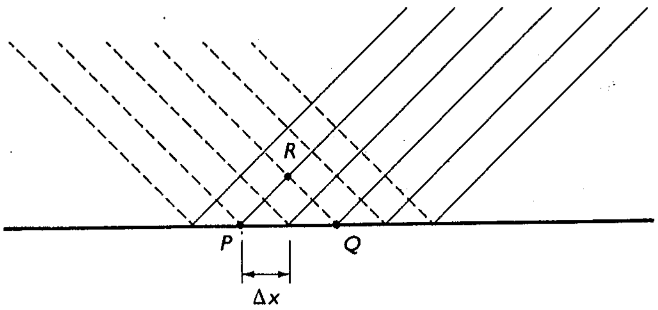


Figure 9.4

EXAMPLE 1 Solve

$$\frac{\partial^2 u}{\partial t^2} = 2\frac{\partial^2 u}{\partial x^2} - 4,$$

with initial conditions

$$u = 12x \quad \text{for} \quad 0 \le x \le 0.25,$$

$$u = 4 - 4x \quad \text{for} \quad 0.25 \le x \le 1.0,$$

$$\frac{\partial u}{\partial t} = 0 \quad \text{for} \quad 0 \le x \le 1.0;$$

boundary conditions are $u = 0$ at $x = 0$ and at $x = 1.0$.

Putting the equation into the standard form,

$$a\frac{\partial^2 u}{\partial x^2} + b\frac{\partial^2 u}{\partial x\,\partial t} + c\frac{\partial^2 u}{\partial t^2} + e = 0,$$

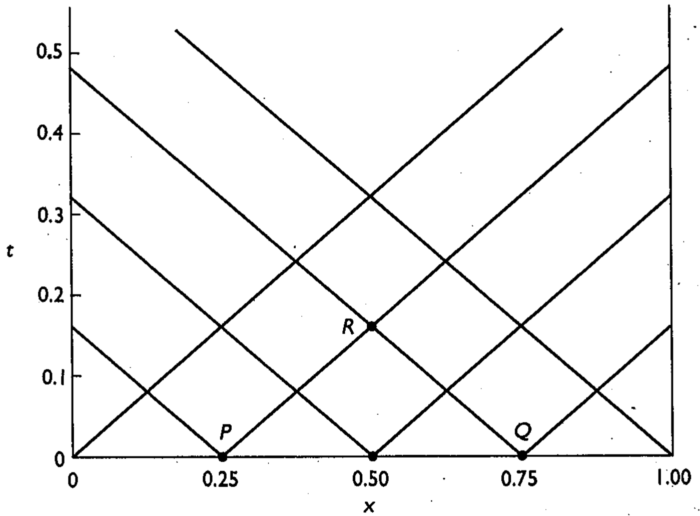


Figure 9.5

gives $a = -2$, $b = 0$, $c = 1$, and $e = 4$. (The equation is linear since a, b, c, and e are independent of u, u_x, and u_t.)

The slopes of the characteristics are the roots of

$$-2m^2 + 1 = 0,$$
$$m = \pm\frac{\sqrt{2}}{2},$$

so the characteristic curves are straight lines in the xt-plane, as shown in Fig. 9.5. Consider points P, Q, and R—(0.25, 0), (0.75, 0), and (0.5, 0.1768)—and solve Eq. (9.16), which is

$$am_{\text{av}}\,\Delta p + c\,\Delta q + e\,\Delta t = 0.$$

$$\boxed{\begin{array}{ll} \text{Along } P \to R: & -2\left(\dfrac{\sqrt{2}}{2}\right)\Delta p + \Delta q + 4\Delta t = 0, \\ & -\sqrt{2}(p_R - p_P) + (q_R - q_P) + 4(0.1768) = 0; \\ \text{Along } Q \to R: & -2\left(\dfrac{-\sqrt{2}}{2}\right)\Delta p + \Delta q + 4\Delta t = 0, \\ & \sqrt{2}(p_R - p_Q) + (q_R - q_Q) + 4(0.1768) = 0. \end{array}}$$

Using

$$p_P = \left(\frac{\partial u}{\partial x}\right)_P = -4,^{*} \qquad q_P = \left(\frac{\partial u}{\partial t}\right)_P = 0, \qquad p_Q = \left(\frac{\partial u}{\partial x}\right)_Q = -4,$$
$$q_Q = \left(\frac{\partial u}{\partial t}\right)_Q = 0,$$

we find $p_R = -4$, $q_R = -\sqrt{2}/2$ by solving the equations $P \to R$ and $Q \to R$ simultaneously.

Now we evaluate u at point R through its change along $P \to R$:

$$\begin{aligned} \Delta u &= p_{\text{av}}\,\Delta x + q_{\text{av}}\,\Delta t = -4(0.25) + \left(\frac{0 - \sqrt{2}/2}{2}\right)(0.1768) \\ &= -1.0625; \\ u_R &= 3 + (-1.0625) = 1.9375. \end{aligned}$$

(If we compute through evaluating Δu along $Q \to R$, we get the same result.)

---

*The gradient has a discontinuity at $x = 0.25$. The value of $\partial u/\partial x$ for points to the right of P applies for the region PRQ.

Table 9.4

x	0	0.25	0.5	0.75	1.0
$u(t = 0)$	0.0	3.0	2.0	1.0	0.0
$u(t = 0.1768)$	0.0	0.9375	$\boxed{1.9375}$	0.9375	0.0
$u(t = 0.3535)$	0.0	−1.1875	−0.2500	0.8125	0.0
$u(t = 0.5303)$	0.0	−1.3125	−2.4375	−1.3125	0.0

For this simple problem, the finite-difference method is much simpler, and we expect it to give the same results. Following the procedure of Section 9.1* we compute with $\Delta x = 0.25$, $\Delta t = \Delta x/\sqrt{c} = 0.1768$, and obtain Table 9.4. The circled value agrees exactly with that calculated by the method of characteristics. ■

EXAMPLE 2 Solve

$$\frac{\partial^2 u}{\partial t^2} = (1 + 2x)\frac{\partial^2 u}{\partial x^2}$$

over (0, 1) with fixed boundaries and the initial conditions:

$$u(x, 0) = 0, \qquad \frac{\partial u}{\partial t}(x, 0) = x(1 - x).$$

For this problem, $a = -(1 + 2x)$, $b = 0$, $c = 1$, $e = 0$. Then $am^2 + bm + c = 0$ gives

$$m = \pm\sqrt{\frac{1}{(1 + 2x)}}.$$

The characteristic curves are found by solving the differential equations $dt/dx = \sqrt{1/(1 + 2x)}$ and $dt/dx = -\sqrt{1/(1 + 2x)}$. Integrating† from the initial point x_0 and t_0, we have

$$t = t_0 + \sqrt{1 + 2x} - \sqrt{1 + 2x_0} \quad \text{from } m_+,$$
$$t = t_0 - \sqrt{1 + 2x} + \sqrt{1 + 2x_0} \quad \text{from } m_-.$$

Figure 9.6 shows several of the characteristic curves. We select two points on the initial curve for $t = 0$, at $P = (0.25, 0)$ and $Q = (0.75, 0)$, whose characteristics intersect at point R. Solving for the intersection, we find $R = (0.4841, 0.1782)$.

---

*The algorithm is $u_i^{j+1} = (u_{i+1}^j + u_{i-1}^j) - u_i^{j-1} - 4(\Delta t)^2$ with $\Delta t = \Delta x/\sqrt{2}$. For the first time step, $u_i^1 = \frac{1}{2}(u_{i+1}^0 + u_{i-1}^0) - \frac{1}{2}(4)(\Delta t)^2$.

†In this example, the integration methods of calculus are easy to use. We could use a numerical method if they were not.

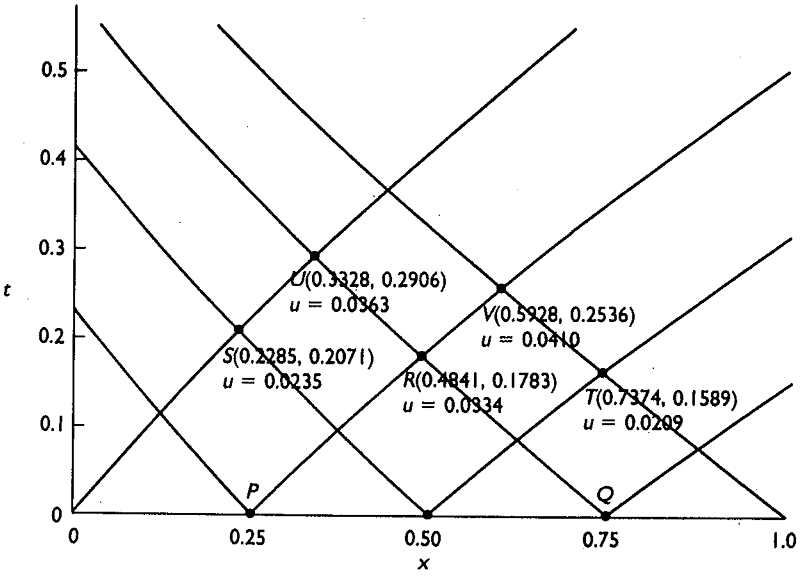


Figure 9.6

We now solve Eq. (9.16) to obtain $p = \partial u/\partial x$ and $q = \partial u/\partial t$ at R:

$$\begin{aligned}
\text{At point } P: \quad & x = 0.25, \quad t = 0, \quad u = 0, \quad p = \left(\frac{\partial u}{\partial x}\right)_P = 0, \\
& q = \left(\frac{\partial u}{\partial t}\right)_P = x - x^2 = 0.1875, \\
& m = \sqrt{1/(1 + 2x)} = 0.8165, \\
& a = -(1 + 2x) = -1.5, \quad b = 0, \quad c = 1, \quad e = 0. \\
\text{At point } Q: \quad & x = 0.75, \quad t = 0, \quad u = 0, \quad p = 0, \quad q = x - x^2 = 0.1875, \\
& m = -\sqrt{1/(1 + 2x)} = -0.6325, \\
& a = -2.5, b = 0, \quad c = 1, \quad e = 0. \\
\text{At point } R: \quad & x = 0.4841, \quad t = 0.1783, \\
& m_+ = \sqrt{1/(1 + 2x)} = 0.7128, \\
& m_- = -\sqrt{1/(1 + 2x)} = -0.7128, \\
& a = -1.9682, \quad b = 0, \quad c = 1, \quad e = 0.
\end{aligned}$$

Equation (9.16) becomes, when we use average values for a and m,

$$P \to R: \qquad -1.7341(0.7646)(p_R - 0) + (1)(q_R - 0.1875) = 0;$$

$$Q \to R: \qquad -2.2341(-0.6726)(p_R - 0) + (1)(q_R - 0.1875) = 0.$$

Solving simultaneously, we get $p_R = 0$, $q_R = 0.1875$.

We calculate the change in u along the characteristics:

$$P \to R: \qquad \Delta u = 0(0.2341) + 0.1875(0.1783) = 0.0334,$$

$$Q \to R: \qquad \Delta u = 0(-0.2659) + 0.1875(0.1783) = 0.0334,$$

$$u_R = 0 + 0.0334 = 0.0334.$$

Figure 9.6 gives the results at several other intersections of characteristics. Students should verify these results to be sure they understand the method of characteristics. ▪

EXAMPLE 3 Solve the quasilinear equation, with conditions as shown, by numerical integration along the characteristics. (This might be a vibrating string with tension related to the displacement u and subject to an external lateral force.)

$$\frac{\partial^2 u}{\partial x^2} - u\frac{\partial^2 u}{\partial t^2} + (1 - x^2) = 0, \qquad u(x, 0) = x(1 - x), \qquad u_t(x, 0) = 0,$$

$$u(0, t) = 0, \qquad u(1, t) = 0. \tag{9.18}$$

We will advance the solution beyond the start from P, at $x = 0.2$, $t = 0$, and Q, at $x = 0.4$, $t = 0$, to one new point R. Comparing Eq. (9.18) to the standard form,

$$au_{xx} + bu_{xt} + cu_{tt} + e = 0,$$

we have $a = 1$, $b = 0$, $c = -u$, $e = 1 - x^2$. We first compute u, p, and q at points P and Q

$$u = x(1 - x)$$

(from the initial conditions), so

$$u_P = 0.2(1 - 0.2) = 0.16,$$

$$u_Q = 0.4(1 - 0.4) = 0.24.$$

Also,

$$p = \frac{\partial u}{\partial x} = 1 - 2x\ .$$

(by differentiating the initial conditions), so

$$p_P = 1 - 2(0.2) = 0.6,$$

$$p_Q = 1 - 2(0.4) = 0.2;$$

and

$$q = \frac{\partial u}{\partial t} = 0$$

(from the initial conditions), so

$$q_P = 0,$$
$$q_Q = 0.$$

To locate point R, we need the slope m of the characteristic. Using $am^2 - bm + c = 0$, we get

$$m = \frac{b \pm \sqrt{b^2 - 4ac}}{2a},$$
$$m = \frac{\pm\sqrt{4u}}{2} = \pm\sqrt{u}.$$

Since m depends on the solution u, we will need to find point R through the predictor–corrector approach. In the first trial, use the initial values over the whole arc; that is, take $m_+ = +m_P$ and $m_- = -m_Q$:

$$m_+ = \sqrt{u_P} = \sqrt{0.16} = 0.4,$$
$$m_- = \sqrt{u_Q} = -\sqrt{0.24} = -0.490.$$

We now estimate the coordinates of R by solving simultaneously

$$t_R = m_+(x_R - x_P) = 0.4(x_R - 0.2),$$
$$t_R = m_-(x_R - x_Q) = -0.490(x_R - 0.4).$$

These give

$$x_R = 0.310, \qquad t_R = 0.044.$$

We write Eq. (9.16) along each characteristic, again using the initial values of m, since m at R is still unknown:

$$am\,\Delta p + c\,\Delta q + e\,\Delta t = 0,$$
$$(1)(0.4)(p_R - 0.6) + (-0.16)(q_R - 0) + \left(1 - \frac{0.04 + 0.096}{2}\right)(0.044) = 0,$$
$$(1)(-0.490(p_R - 0.2) + (-0.24)(q_R - 0) + \left(1 - \frac{0.16 + 0.096}{2}\right)(0.044) = 0.$$

In these equations we used the arithmetic average of x^2 in the last terms. Solving simultaneously, we get

$$p_R = 0.399, \qquad q_R = -0.246.$$

As a first approximation for u at R, then,

$$\begin{aligned}
\Delta u &= p\,\Delta x + q\,\Delta t,\\
u_R - 0.16 &= \frac{0.6 + 0.399}{2}(0.310 - 0.2) + \frac{0 - 0.246}{2}(0.044 - 0),\\
u_R &= 0.2095.
\end{aligned}$$

The last computation was along PR, using average values of p and q. We could have alternatively proceeded along QR. If this is done,

$$\begin{aligned}
u_R - 0.24 &= \frac{0.2 + 0.399}{2}(0.310 - 0.4) + \frac{0 - 0.246}{2}(0.044 - 0),\\
u_R &= 0.2076.
\end{aligned}$$

The two values should be close to each other. Let us use the average value, 0.2086, as our initial estimate of u_R. We now repeat the work. In getting the coordinates of R, we now use average values of the slopes,

$$t_R = \frac{0.4 + \sqrt{0.2086}}{2}(x_R - 0.2),$$

$$t_R = \frac{-0.490 - \sqrt{0.2086}}{2}(x_R - 0.4),$$

$$x_R = 0.305, \qquad t_R = 0.045;$$

$$(1)\left(\frac{0.4 + \sqrt{0.2086}}{2}\right)(p_R - 0.6) - \left(\frac{0.16 + 0.2086}{2}\right)(q_R - 0) + \left(1 - \frac{0.04 + 0.0930}{2}\right)(0.045) = 0,$$

$$(1)\left(\frac{-0.490 - \sqrt{0.2086}}{2}\right)(p_R - 0.2) - \left(\frac{0.24 + 0.2086}{2}\right)(q_R - 0) + \left(1 - \frac{0.16 + 0.0930}{2}\right)(0.045) = 0,$$

$$p_R = 0.398, \qquad q_R = -0.242;$$

$$u_R = 0.16 + \frac{0.6 + 0.398}{2}(0.305 - 0.2) + \frac{0 - 0.242}{2}(0.045 - 0),$$

$$u_R = 0.2071 \qquad (\text{along } PR);$$

$$u_R = 0.24 + \frac{0.2 + 0.398}{2}(0.305 - 0.4) + \frac{0 - 0.242}{2}(0.045 - 0),$$

$$u_R = 0.2063 \qquad (\text{along } QR).$$

The average value is 0.2067.

Another round of calculations gives $u_R = 0.2066$, which checks the previous value sufficiently. This method is, of course, very tedious by hand. ∎

FD method easier to program but discontinuities tend to be defused
method of characteristics keep discontinuities but are harder to program

also $a_{av}\, m_{av}\, \Delta p + c_{av}\, \Delta q + K_{av}\, \Delta t = 0$

where $a_{av} = \frac{1}{2}[\ddot{a}(R) + a(P)]$ $\quad m_{av} = [m_1(R) + m_1(P)]\frac{1}{2}$

$c_{av} = \frac{1}{2}[c(R) + c(P)]$ $\quad$ & $K_{av} = [K(P) + K(R)]\frac{1}{2}$

and then repeat to get better approx after we use

$$\Delta u = \frac{1}{2}[p(R) + p(P)]\,\Delta x + \frac{1}{2}[q(R) + q(P)]\,\Delta t$$

if the initial data lies along a characteristic - the initial data will have to solve the PDE. Solution along the characteristic will be unique but not anywhere else.

in 2-D space wave eqn is

$$\frac{\partial^2 u}{\partial t^2} = a^2\left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2}\right)$$

let's use CD in time & CD in space with $\Delta x = \Delta y$

$$u_{i,j,k+1} = \frac{a^2\Delta t^2}{\Delta x^2}\left\{\begin{matrix} & 1 & \\ 1 & 0 & 1 \\ & 1 & \end{matrix}\right\} u_{i,j,k} + \left(2 - 4a^2\frac{\Delta t^2}{\Delta x^2}\right) u_{i,j,k} - u_{i,j,k-1}$$

if $\frac{a^2\Delta t^2}{\Delta x^2} = \frac{1}{2}$ $\quad u_{i,j,k+1} = \frac{1}{2}\left\{\begin{matrix} & 1 & \\ 1 & 0 & 1 \\ & 1 & \end{matrix}\right\} u_{i,j,k} - u_{i,j,k-1}$

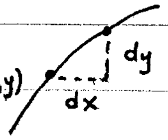
at initial step $k=0$ $\quad u_{i,j,1} = \frac{1}{4}\left\{\begin{matrix} & 1 & \\ 1 & 0 & 1 \\ & 1 & \end{matrix}\right\} u_{i,j,0} + \Delta t\, g(x_i, y_j)$

$\frac{a^2\Delta t^2}{\Delta x^2} = \frac{1}{2}$ is max value for stability; however results are not exact & need smaller time steps to process.

IF we have a first order eqn.

$$(1) \quad a\frac{\partial U}{\partial x} + b\frac{\partial U}{\partial y} = c \qquad \text{where} \quad \begin{array}{l} a = a(x,y,U) \text{ only} \\ b = b(x,y,U) \text{ only} \\ c = c(x,y,U) \text{ only} \end{array}$$

we ask can we find a ~~direction~~ curve such that the expression can be integrated independent of the partial derivatives



$$dU = \frac{\partial U}{\partial x}dx + \frac{\partial U}{\partial y}dy$$

solve (1) for ~~and~~ $\frac{\partial U}{\partial x} = \left(c - b\frac{\partial U}{\partial y}\right)/a$

$$\therefore \quad dU = \left(c - b\frac{\partial U}{\partial y}\right)/a \; dx + \frac{\partial U}{\partial y}dy$$

$$= \frac{c}{a}dx + \frac{\partial U}{\partial y}\left(dy - \frac{b\,dx}{a}\right)$$

$$\frac{\partial U}{\partial y} = \cancel{\left(\frac{dU - \frac{\partial U}{\partial x}dx}{dy}\right)} = \frac{c - a\frac{\partial U}{\partial x}}{b}$$

$$dU = \frac{\partial U}{\partial x}dx + \left(\frac{0 - a\,\partial U/\partial x}{b}\right)dy$$

$$dU = \frac{c}{b}dy + \left(dx - \frac{a\,dy}{b}\right)\frac{\partial U}{\partial x}$$

if the curve $f(x,y) = \text{const}$ can be picked so that $dy - \frac{b\,dx}{a} = 0$ then dU is not involving $\frac{\partial U}{\partial x}, \frac{\partial U}{\partial y}$

then $dU = \frac{c}{a}dx$ & $U(x,y)$ can be found from $\int \frac{c}{a}dx + \mathcal{E}$

what is $f(x,y)$? pick it so that $\frac{dy}{dx} = \frac{b}{a}$ & $\frac{dU}{dx} = \frac{c}{a}$

$$\frac{dU}{dy} = \frac{c}{b}$$

example consider $\overset{a}{y}\frac{\partial U}{\partial x} + \overset{b}{1}\frac{\partial U}{\partial y} = \overset{c}{2}$ then $\frac{dy}{dx} = \frac{1}{y}$ or $x - \frac{y^2}{2} = \text{Const.}$

curve or the characteristic

& $U = \int \frac{2}{y}dx$ ~~or~~ $\underset{\text{since } \frac{dy}{dx} = \frac{1}{y} \Rightarrow y\,dy = dx}{\int 2\,dy} = 2y + f(x) \quad \therefore \quad U = 2y + f(x)$

let $U = U_R$ be given along $y = 0 \Rightarrow \mathcal{E} = U_R \quad \therefore \quad U = 2y + U_R$

Numerically $\Delta y \cdot a = \Delta x \cdot b$ if x_R is known (from stepsize) & x_i & y_i are known (initial conditions) this gives y_R

& $a\,\Delta U = \Delta x \cdot c$

$$(y_R - y_i)a = (x_R - x_i)\cdot b$$

then $(U_R - U_i)a = (x_R - x_i)\cdot c$

check $(U_R - U_i)b = (y_R - y_i)c$

Lax - Wendroff explicit Scheme used in fluid flow problems

$$\frac{\partial \underline{U}}{\partial t} + \frac{\partial \underline{F}(\underline{U})}{\partial x} = 0$$ conservation equation

in one D $\frac{\partial U}{\partial t} + a\frac{\partial U}{\partial x} = 0 \Rightarrow \frac{\partial}{\partial t} = -a\frac{\partial}{\partial x}$ on U

now using Taylor Series $U_{i,j+1} = U_{i,j} + \Delta t \left.\frac{\partial U}{\partial t}\right|_{i,j} + \frac{\Delta t^2}{2}\left.\frac{\partial^2 U}{\partial t^2}\right|_{i,j} + \dots = U_{i,j} + \Delta t\left(-a\frac{\partial U}{\partial x}\right) + \frac{\Delta t^2}{2}\left(a^2\frac{\partial^2 U}{\partial x^2}\right.$

replace by centered Diff on $\frac{\partial}{\partial x}$ & $\frac{\partial^2}{\partial x^2}$

$$U_{i,j+1} = U_{i,j} + \frac{a\Delta t}{2\Delta x}\left(U_{i+1,j} - U_{i-1,j}\right) + \frac{\Delta t^2 a^2}{2\Delta x^2}\left(U_{i+1,j} - 2U_{ij} + U_{i-1,j}\right.$$

provides an explicit scheme

Δt is kept small when x changes occur rapidly with time

(Smith - Num Solns of PDE)

For final use example given on pg 166 - 168 $u_{xx} - u^2 u_{tt} = 0$ $u = 0.2 + 5x^2$ & $\frac{\partial u}{\partial t} = 3x$ @ $t=0$ $0 \le x \le 1$

nonlin characteristic prob. use pts $x_p = 0.2,\ 0.3 = x_Q$

EGM 6422

LECTURE 27/28

the region is defined solely by limiting values for the variables themselves, this is not necessary.

In Excel, we invoke `Tools/Solver;` in Quattro Pro, we use `Tools/Numeric Tools/Optimizer.` In dialog boxes, we enter the cell numbers for variables and the function, together with constraints that define the region. Clicking on `Solve` produces the solution and we can get the successive iterations if we want to see them. Options that are available include Gradient, Conjugate, and Newton.

7.3 Linear Programming

A widely used technique for maximizing the profits or minimizing the costs is *linear programming*. It is often used in business to determine those decisions that will increase profitability. It has other business applications, such as finding the optimal schedule for an outside salesman to visit his customers.

The word *programming* here does not mean a computer program in the ordinary sense (although computers are nearly always used to solve the problems). It refers instead to a systematic procedure, one that is based on solving set linear equations. Linear programming is *linear* in that the function whose optimum is sought is a linear combination of two or more (often many) independent variables. The solution is subject to a number of constraints, and these are themselves always a linear combination of the variables. A constraint, for example, might be how a limited resource will be utilized by several competing potential applications.

A Simple Problem

We begin with a simple problem with just two variables, but this will illustrate the method and introduce some of the many special terms of linear programming. The problem is to maximize $f(x_1, x_2) = 5x_1 + 8x_2$,

subject to:

$$x_1 + 3x_2 \le 12,$$
$$3x_1 + 2x_2 \le 15,$$
$$x_1, x_2 \ge 0.$$

Think of a company that is to manufacture two products. The amount of each is measured by x_1 and x_2. $f(x_1, x_2)$ is the *objective function*. This function, $f(x_1, x_2)$, determines the manufacturing profit. The larger the values for x_1 and x_2, the greater the profit. The coefficients are the profit per unit of product.

However, it is not possible to manufacture any desired quantity of these products, for there is a limited amount of two necessary resources. (These might be available employees, critically important parts, machine availability, or the like.) The *constraint relations* show how each of the resources is used up in the manufacturing process. The coefficients

lves, this is not

`ols/Numeric`
ariables and the
ve produces the
Options that are

e costs is *linear*
hat will increase
l schedule for an

e ordinary sense
fers instead to a
Linear program-
nbination of two
number of con-
ariables. A con-
veral competing

trate the method
he problem is to

ach is measured
, determines the
rofit. The coeffi-

se products, for
vailable employ-
straint relations
The coefficients

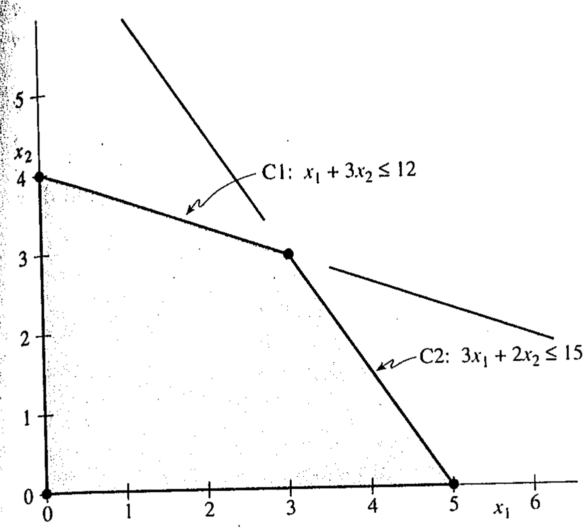


Figure 7.9
The feasible region

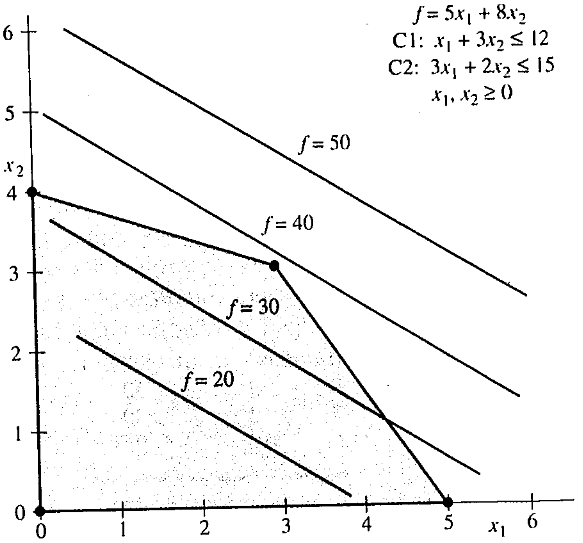


Figure 7.10
Objective function values are superimposed on the feasible region

in the constraint relations represent the required amount of the resource used per unit amount of the product.

Notice that each of the constraints is linear and that the objective function is also a linear combination. The last inequality is a special one; while not a constraint in the same sense as the others, it forces the solution to have only nonnegative values for the variables. This is common because it is impossible to make a neqative quantity of product.

We will first solve this graphically; this will introduce the topic and help to define a number of special terms. A plot of the constraints in Figure 7.9 shows the *feasible region*, the possible production quantities of product 1 and product 2. (We have scaled the numbers to make them small. The actual quantities might be 100 or 1000 times as great.) The region is bordered by the heavy lines.

Observe that the feasible region is bounded by the x_1, x_2 axes (from the nonnegativity condition) and by two intersecting lines. There are four vertices to the polygonal region, including one at the origin and two on the axes.

In Figure 7.10, we redraw the feasible region and superimpose on it a number of lines defined by setting the objection relation equal to several values.

have same slope

Because the objective function is linear, the lines for $f(x_1, x_2)$ are parallel. The larger the value assigned to the function, the farther from the origin the line lies. Some of the lines do not fall within the feasible region—we cannot achieve that much profit. Some lie within the region but represent choices that give less profit than the maximum. Points on such lines within the region are *feasible solutions*. There is one line (not drawn) that would show the maximum; it would just touch the feasible region. In this example, it will touch at

the point (3, 3). A different objective function whose slope is different might touch the region at a different vertex. The important conclusion from this is that the *optimal solution* will always fall at one of the vertices of the region.

The four vertices of the region in Figure 7.10 (we include the origin) are called *basic feasible solutions*. It is then clear that one way to solve this linear programming problem is to find values for x_1 and x_2 at the vertices and from these compute the values for the objective function at each vertex of the region (more commonly called *corner points*), and then select the point where it is a maximum. For our example, these values are

x_1	x_2	$f(x_1, x_2)$
0	0	0
0	4	32
3	3	39
5	0	25

This confirms the fact that the optimal value for the two products is three units and three units, respectively.

Examining Figure 7.10 suggests several other possibilities:

1. If the objective function had different coefficients, the objective function lines might be parallel to one of the constraints and one of these lines will coincide with an edge of the region. In that case, there are multiple optimal solutions. Any combination of choices for x_1 and x_2 that lie on that edge give the same profit.
2. There could be a third constraint and this can have different possible effects:
   a. It could lie totally outside the feasible region and thus not limit the amounts to be produced. We would call this a redundant constraint.
   b. It could coincide with one of the previous constraints. This too is redundant; the region is not affected.
   c. It could lie partially within the region. This would decrease the area of the feasible polygon and might create additional corner points.
3. The graphical method for solving a linear programming problem is fine if there are only two variables. It could be applied (with difficulty) to three variables, but more than three is virtually impossible. We need to find a different way to solve linear programming problems because some applications have hundreds of variables.

The Simplex Method

Even though we have already solved our example, we will use it to introduce the *simplex method,* which is most frequently used for linear programming. We repeat the problem: Maximize

$$f(x_1, x_2) = 5x_1 + 8x_2,$$

subject to

$$\begin{aligned} x_1 + 3x_2 &\le 12, \\ 3x_1 + 2x_2 &\le 15, \\ x_1, x_2 &\ge 0. \end{aligned}$$

The simplex method solves the problem through solving a set of equations that represent the constraints. "But our constraints aren't equations, they are inequalities," you say. That is a good observation. We need to change inequalities to equalities. This can be done by a simple device: We add another variable to the constraint, a quantity called a *slack variable.* This measures the amount of the resource not utilized; it takes up the "slack." Call the slack variable for the first constraint x_3, and that for the second, x_4. Our problem then becomes to maximize

$$f(x_1, x_2) = 5x_1 + 8x_2 + 0x_3 + 0x_4$$

subject to

$$\begin{aligned} x_1 + 3x_2 + x_3 &= 12, \\ 3x_1 + 2x_2 + x_4 &= 15, \\ x_1, x_2, x_3, x_4 &\ge 0. \end{aligned}$$

We have expanded the objective function to include the slack variables. They contribute nothing to profits, of course.

In matrix form, the constraint equations are

$$\begin{bmatrix} 1 & 3 & 1 & 0 \\ 3 & 2 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 12 \\ 15 \end{bmatrix}.$$

This system is underdetermined; there are only two equations but four variables. Still, we can solve this if we first assign values to two of the variables and move these terms to the right-hand side. Observe that adding the slacks to the system expanded the matrix of constraint coefficients to include an identity matrix.

Let us assign zero to both x_1 and x_2. The system is reduced to

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 12 \\ 15 \end{bmatrix},$$

where the solution is obvious: $x_3 = 12$, $x_4 = 15$. "Of course," you say, "if neither product is made, the entire amount of both of the resources is unused. The slacks measure that." The important result is that we have values for x_1 and x_2 at a corner point, a basic feasible solution to the problem, though surely not the optimum.

In the terminology of linear programming, what we have just done is to cause two variables to leave the system and two to enter. The ones that leave are x_1 and x_2; the ones that enter are x_3 and x_4.

Suppose we allow a new variable to enter the system, replacing one that is already there. So, one of x_3 or x_4 must leave. In effect, we are exchanging a current variable for one not yet in the system.

Which of x_1 or x_2 should we select to enter? Looking at the objective function, we see that the profit from one unit of x_2 is 8, while one unit of x_1 returns only 5; x_2 is the better choice. (You may want to see whether the other choice ends up at the same final answer.) So, x_2 is to enter the system. Now we must decide which of x_3 and x_4 should leave. We answer the question by trying both possibilities:

If x_3 leaves, the variables in the equations are x_2 and x_4, and the system becomes

$3x_2+0x_4=12$
$2x_2+1x_4=15$

$$\begin{bmatrix}3 & 0\\ 2 & 1\end{bmatrix}\begin{bmatrix}x_2\\ x_4\end{bmatrix}=\begin{bmatrix}12\\ 15\end{bmatrix}, \quad \text{solution: } \begin{aligned} x_2 &= 4,\\ x_4 &= 7.\end{aligned}$$

$f(x_1,x_2,x_3,x_4)=5x_1+8x_2+0x_3+0x_4$
$=32$

If x_4 leaves instead, we have

$3x_2+x_3=12$ $x_3=-10.5$
$2x_2+0x_3=15$ $x_2=7.5$

$$\begin{bmatrix}3 & 1\\ 2 & 0\end{bmatrix}\begin{bmatrix}x_2\\ x_3\end{bmatrix}=\begin{bmatrix}12\\ 15\end{bmatrix}, \quad \text{solution: } \begin{aligned} x_2 &= \cancel{6.}\ 7.5\\ x_3 &= \cancel{-6.}\ -10.5\end{aligned}$$

Only the first is acceptable: the second violates the nonnegativity condition. The variables now present are x_2 and x_4. Remembering that x_1 is still zero but now x_2 is 4; we have moved from our initial basic feasible solution, (0, 0) to another basic feasible solution, (0, 4). At this point, the value of the objective function is 32.

We proceed in similar fashion to allow x_1 to enter. x_4 will have to leave. The variables present are the non-slacks, x_1 and x_2. We need to solve:

$$\begin{bmatrix}1 & 3\\ 3 & 2\end{bmatrix}\begin{bmatrix}x_1\\ x_2\end{bmatrix}=\begin{bmatrix}12\\ 15\end{bmatrix}, \quad \text{solution: } \begin{aligned} x_1 &= 3,\\ x_2 &= 3.\end{aligned}$$

We have moved to another basic feasible solution (3, 3), where the value of the objective function is 39. In this problem, we know that this must be the optimum point because removing either x_1 or x_2 can only reduce the profit.

(What we have done is to solve for the intersection of the two constraints, a corner point.)

Variations of the Problem

Even this simple example can illustrate how some variants to the problem affect it.

1. What if the lower limit of one of the variables is something other than zero? This would have to be a positive quantity. It also would have to be small enough to lie within the feasible region, or else we would say the problem is *infeasible:* No solution is possible. This is also true if both variables have lower limits other than zero. Having lower limits other than zero will reduce the area of the feasible region. The initial basic feasible solution would still be at one of the lower-limit points. If the nonegativity constraint were replaced by $x_1+x_2 \geq$ [illegible]

this would chop off a triangle from the lower-left part of the feasible region. We would have to include this inequality in the matrix of constraints. With a greater than or equal relation, the slack variable is subtracted to give the constraint equation.

2. What if additional greater than or equal constraints are included? We just include these with a subtracted slack variable. It is then possible to have a diamond-shaped feasible region.
3. What if the lines for the objective function are parallel to one of the constraints? One of these lines would then coincide with an edge of the region, and any point on this edge is optimal; there is then an infinity of optimal points, all with the same value for the objective function.
4. What if the objective function has a positive slope? (This would mean that one of the products incurs a loss rather than a profit but that, while unlikely, could happen.) The objective function lines would then intersect the constraints. For a region like that of Figure 7.10, the optimum would still occur at a corner point. The simplex method will still find it.
5. What if we want to minimize an objective function? (The coefficients then would represent unit costs rather than unit profits.) The simplex procedure works exactly the same—we just maximize the negative of the objective function.
6. Can we use the simplex method to solve a problem where either the objective function or a constraint is discontinuous? No, the requirement of linearity is absolute.

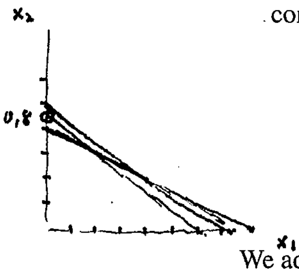
Another Example

We now present a slightly more complex problem that will show how the simplex method works when there are more than two constraints. It often occurs that there are more constraints than variables. The example still has only two variables, so it could be solved graphically or by computing a list of function values are the corners. Here is our example:

Maximize

$$f(x_1, x_2) = 8x_1 + 9x_2,$$

constraints:



$$\text{C1: } 2x_1 + 4x_2 \leq 32,$$
$$\text{C2: } 3x_1 + 4x_2 \leq 36,$$
$$\text{C3: } 6x_1 + 4x_2 \leq 60,$$
$$x_1, x_2 \geq 0.$$

We add slacks x_3, x_4, x_5 to the three constraints. In matrix form we have:

$$f = 8x_1 + 9x_2 + 0x_3 + 0x_4 + 0x_5,$$

$$\begin{bmatrix} 2 & 4 & 1 & 0 & 0 \\ 3 & 4 & 0 & 1 & 0 \\ 6 & 4 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} 32 \\ 36 \\ 60 \end{bmatrix},$$

$$x_1, x_2, x_3, x_4, x_5 \geq 0.$$

We begin as is customary with a basic feasible solution at the origin, (0, 0), where $f = 0$. We improve the solution, by bringing in a new variable to replace one of x_3, x_4, or x_5. Our best choice of the variable to bring into the solution is x_2. We need to see which of the current variables is to leave, so we try each in turn.

If x_3 leaves and x_2 enters, the variables in the solution are x_2, x_4, and x_5. We solve:

$$\begin{bmatrix} 4 & 0 & 0 \\ 4 & 1 & 0 \\ 4 & 0 & 1 \end{bmatrix} \underset{\sim}{x} = \begin{bmatrix} 32 \\ 36 \\ 60 \end{bmatrix}, \quad \text{whose solution is} \begin{cases} x_2 = 8 \\ x_4 = 4, \\ x_5 = 28 \end{cases} \leftarrow \text{corner pt } (0,8)$$

which we can accept; the nonnegativity condition holds. $f(x_1, x_2) = 72$

Let us see if any of the other choices is acceptable. If x_4 leaves instead of x_3, the variables in the solution will be x_2, x_3, and x_5. We solve:

$$\begin{bmatrix} 4 & 1 & 0 \\ 4 & 0 & 0 \\ 4 & 0 & 1 \end{bmatrix} x = \begin{bmatrix} 32 \\ 36 \\ 60 \end{bmatrix}, \quad \text{whose solution is} \begin{cases} x_2 = 9 \\ x_3 = -4. \leftarrow \times \\ x_5 = 24 \end{cases}$$

This is not acceptable. What if we let x_5 leave instead of x_3? The variables will be x_2, x_3, and x_4. We solve:

$$\begin{bmatrix} 4 & 1 & 0 \\ 4 & 0 & 1 \\ 4 & 0 & 0 \end{bmatrix} x = \begin{bmatrix} 32 \\ 36 \\ 60 \end{bmatrix}, \quad \text{whose solution is} \begin{cases} x_2 = 15 \\ x_3 = -28, \leftarrow \times \\ x_4 = -24 \swarrow \end{cases}$$

which is also not acceptable. With variables x_2, x_4, and x_5 in the system, the value for x_2 is 8, x_1 is zero. At (0, 8), f is 72.

We hope to improve the solution by replacing x_3 or x_4. We don't want to put x_5 back in, so we let x_1 replace either x_3 or x_4. If we replace x_3, we will have x_1, x_2, and x_4. We solve:

$$\begin{bmatrix} 2 & 4 & 0 \\ 3 & 4 & 1 \\ 6 & 4 & 0 \end{bmatrix} x = \begin{bmatrix} 32 \\ 36 \\ 60 \end{bmatrix}, \quad \text{whose solution is} \begin{cases} x_1 = 7 \\ x_2 = 4.5, \quad \times \\ x_4 = -3 \swarrow \end{cases}$$

which we must reject. We try the other choice, giving the variables as x_1, x_2, and x_3. We solve:

$$\begin{bmatrix} 2 & 4 & 1 \\ 3 & 4 & 0 \\ 6 & 4 & 0 \end{bmatrix} x = \begin{bmatrix} 32 \\ 36 \\ 60 \end{bmatrix}, \quad \text{whose solution is} \begin{cases} x_1 = 8 \\ x_2 = 3. \\ x_3 = 4 \end{cases} \Big\} \leftarrow \text{corner pt } (8,3)$$

We can accept this. So, we have moved from (0, 8) to (8, 3), where the value of f is 91.

Can we improve further? The only possibility is to put x_5 back in, replacing x_3. With variables x_1, x_2, and x_5, we solve:

$$\begin{bmatrix} 2 & 4 & 0 \\ 3 & 4 & 0 \\ 6 & 4 & 1 \end{bmatrix} = \begin{bmatrix} 32 \\ 36 \\ 60 \end{bmatrix}, \quad \text{whose solution is } \begin{cases} x_1 = 4 \\ x_2 = 6. \\ x_5 = 12 \end{cases} \Big) \leftarrow \text{corner pt} \ (4,6)$$

At (4, 6), $f = 86$, and we do not increase the value of f. It seems that the optimum is at (8, 3), where $f = 91$.

There is one more corner point that we could test; it is at (10, 0), where $f = 80$, less than that at other corners.

$$\begin{bmatrix} 2 & 1 & 0 \\ 3 & 0 & 1 \\ 6 & 0 & 0 \end{bmatrix} \begin{pmatrix} x_1 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 32 \\ 36 \\ 60 \end{pmatrix} \qquad \begin{matrix} x_1 = 10 \\ x_3 = 12 \\ x_4 = 6 \end{matrix} \leftarrow \text{corner pt} \ (10,0) \quad f(x_1,x_2) = 80$$

Are There More Efficient Ways?

$$\begin{bmatrix} 2 & 1 & 0 \\ 3 & 0 & 0 \\ 6 & 0 & 1 \end{bmatrix} \begin{pmatrix} x_1 \\ x_3 \\ x_5 \end{pmatrix} = \begin{pmatrix} 32 \\ 36 \\ 60 \end{pmatrix} \qquad \begin{matrix} x_1 = 12 \\ x_3 = 8 \\ x_5 = -12 \end{matrix} \quad X$$

We have used a procedure that would most clearly show the basic principle behind the simplex method. This is perhaps not the most efficient. We solved the examples in this way to emphasize that we move from one basic feasible solution to another where the objective function is improved. We did this by replacing one current variable with another. Selecting the variable to enter was easy: We chose the one that would contribute most to the objective function, the one with the larger unit profit. We selected which variable would leave by examining whether the nonnegativity constraints were violated. This examination was done by computing the amounts of the present variables that would remain in the solution when the new variable entered; if any of these were negative, we rejected it.

An alternative procedure sets up a *simplex tableau*. In using this tableau, all of candidates for leaving the basis are tested simultaneously, rather than individually as we have done. The tableau is modified at each iteration by doing the equivalent of a Gauss–Jordan reduction. This may require fewer arithmetic operations but what is happening to the variables is not seen as clearly.

Every linear programming problem has another problem called its *dual* and the solution to the dual problem is the same as for the *primal problem*. The dual may require less effort to solve than the primal, and solving it will be more efficient. We discuss the dual to a primal problem later in this section.

A problem with many variables and many constraints can be solved in the same way as we have described but doing it would be painfully slow. The use of a computer program is essential and there are many available. We can even use the Excel or Quattro Pro spreadsheet programs. Here is how Quattro Pro solves a linear programming problem. We use the last example as an illustration.

Using Quattro Pro

We restate the problem:

$$\text{Maximize } f(x_1, x_2) = 8x_1 + 9x_2,$$

Eq. No.	x_1	x_2	x_3
1	−1	−0.125	0.125
2	0.143	−1	0.286
3	−0.222	−0.111	−1

x_1	R_1	x_2	R_2	x_3	R_3
0	~~1000~~	0	~~571~~	0	~~1333~~
	+167		+381		
	~~1167~~		~~952~~	+1333	~~0~~
			+167		−259
+1167	~~0~~		~~1119~~		~~−259~~
	−140				−124
	~~−140~~	+1119	~~0~~		~~−383~~
	− 48		− 109		
	~~−189~~		~~−109~~	−383	~~0~~
			− 27		+ 42
−189	~~0~~		~~−136~~		~~42~~
	+ 17				+ 15
	~~17~~	−136	~~0~~		~~57~~
	+ 7		+ 16		
	~~24~~		~~16~~	+57	~~0~~
			+ 3		− 5
+24	~~0~~		~~19~~		~~−5~~
	− 2				− 2
	~~−2~~	+19	~~0~~		~~−7~~
	− 1		− 2		
	~~−3~~		~~−2~~	−7	~~0~~
			0		+ 1
−3	~~0~~		~~−2~~		~~1~~
	0				0
	~~0~~	−2	~~0~~		~~1~~
	0		0	+1	0
999		1000		1001	
Check residuals:	1		−1		0

Figure 2.2 Solving a set of linear equations by relaxation

We make three double columns, one for each variable and for the residual of the equation in which that variable appears with −1 coefficient. The initial x values and the initial residuals are entered as the first row of the table. It is convenient to work entirely with integers by multiplying the initial x-values and residuals by 1000, and then to scale down the solution by dividing by 1000 at the end of the computations. We avoid fractions; if a fractional change in a variable is needed to relax to zero, we only relax to near zero.

In Fig. 2.2, we set down the increments to the x's but record the cumulative effect on the residuals. (The old values of the residuals are crossed out when replaced by

a new value.) When the residuals are zero, we add the various increments to the initial value to get the final value. In this example, round-off errors cause an error of one in the third decimal.

It is important to make a final check by recomputing residuals at the end of the calculation to check for mistakes in arithmetic. The method is not usually programmed because searching on the computer for the largest residual is slow, adding enough execution time that the acceleration gives no net benefit. The search can be done rapidly by scanning the residuals in a hand calculation, however.

Southwell and his co-workers observed, for many situations, that relaxing the residuals to zero was less efficient than relaxing beyond zero (*overrelaxing*) or relaxing short of zero (*underrelaxing*). The reason this strategy is an improved one is that a zero residual doesn't stay zero; relaxing the residual of another equation affects the first residual, so it is appropriate to anticipate and allow for this by an appropriate under- or overrelaxation.

Table 2.1 shows that a significant improvement in the speed of convergence is obtained if R_1 is underrelaxed by 10% and R_3 is underrelaxed by 25%. Unfortunately, the optimum degree of under- or overrelaxation is not easily determined. In many problems, acceleration is obtained by overrelaxing rather than underrelaxing.

Table 2.1 Accelerated solution of linear equations by relaxation

Eq. No.	x_1	x_2	x_3
1	−1	−0.125	0.125
2	0.143	−1	0.286
3	−0.222	−0.111	−1

x_1	R_1	x_2	R_2	x_3	R_3
0	~~1000~~	0	~~571~~	0	~~1333~~
	+125		+286		
	~~1125~~		~~857~~	+1000	~~333~~
			+144		−225
+1013	~~112~~		~~1001~~		~~108~~
	−125				−111
	~~−13~~	+1001	~~0~~		~~−3~~
			−2		+3
−12	~~−1~~		~~−2~~		~~0~~
	+0				+0
	~~−1~~	−2	~~0~~		~~0~~
			+0		+0
−1	0		0		0
1000		999		1000	

Even though Southwell's relaxation method is not often used today, there is one aspect of it that has influence on the iterative solution of linear equations by computer. In using the Gauss–Seidel method, we can speed up the convergence by

"overrelaxation," that is, by making the residuals go to the other side of zero instead of just relaxing to zero as in the first example. We can apply this technique to Gauss–Seidel iteration by modifying the algorithm.

The standard relationship for Gauss–Seidel iteration for the set of equations $Ax = b$, for variable x_i, can be written

$$\boxed{x_i^{(k+1)} = \frac{1}{a_{ii}}\left(b_i - \sum_{j=1}^{i-1} a_{ij}x_j^{(k+1)} - \sum_{j=i+1}^{n} a_{ij}x_j^{(k)}\right), \qquad (2.50)}$$

where the superscript $(k + 1)$ indicates that this is the $(k + 1)$st iterate. On the right side we use the most recent estimates of the x_j, which will be either $x_j^{(k)}$ or $x_j^{(k+1)}$.

An algebraically equivalent form for Eq. (2.50) is

$$\boxed{x_j^{(k+1)} = x_i^{(k)} + \frac{1}{a_{ii}}\left(b_i - \sum_{j=1}^{i-1} a_{ij}x_j^{(k+1)} - \sum_{j=i}^{n} a_{ij}x_j^{(k)}\right),}$$

because $x_i^{(k)}$ is both added to and subtracted from the right side. In this form, we see that Gauss–Seidel and Southwell's relaxation can have identical arithmetic: The term we add to $x_i^{(k)}$ to get $x_i^{(k+1)}$ is exactly the increment that relaxes the residual to zero. (Of course, we apply the relaxation to the x_i's in a different sequence in the two methods.) Overrelaxation can be applied to Gauss–Seidel if we will add to $x_i^{(k)}$ some multiple of the second term. It can be shown that this multiple should never be more than 2 in magnitude (to avoid divergence), and the optimum overrelaxation factor lies between 1.0 and 2.0. Our iteration equations take this form, where w is the *overrelaxation factor*:

$$\boxed{x_i^{(k+1)} = x_i^{(k)} + \frac{w}{a_{ii}}\left(b_i - \sum_{j=1}^{i-1} a_{ij}x_j^{(k+1)} - \sum_{j=i}^{n} a_{ij}x_j^{(k)}\right). \qquad (2.51)}$$

Table 2.2 shows how the convergence rate is influenced by the value of w for the system

$$\begin{bmatrix} -4 & 1 & 1 & 1 \\ 1 & -4 & 1 & 1 \\ 1 & 1 & -4 & 1 \\ 1 & 1 & 1 & -4 \end{bmatrix} x = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix},$$

starting with an initial estimate of $x = 0$. The exact solution is

$$x_1 = -1, \qquad x_2 = -1, \qquad x_3 = -1, \qquad x_4 = -1.$$

Relaxation Method

- More rapidly convergent than Gauss-Seidel

$8x_1 + x_2 - x_3 = 8$ ÷ by -8

$2x_1 + x_2 + 9x_3 = 12$ ÷ by -9

$x_1 - 7x_2 + 2x_3 = -4$ ÷ by $-(-7)$

Transpose all to one side of eq & divide by − largest coeff in each eqn.

largest coeff in each eqn.

$-1x_1 - .125x_2 + .125x_3 + 1 = 0$

$-.222x_1 - .111x_2 - 1x_3 + 1.333 = 0$

$.143x_1 - x_2 + .286x_3 + .571 = 0$

∘ call RHS R_1 R_2 R_3

interchange R_2 & R_3 so that -1 is on diag

Now pick a trial x_1, x_2, x_3 ~~(1,1,1)~~

~~$R_1 = 0$~~

~~$R_2 = 0$~~

~~$R_3 = 0$~~

what if $(0,0,0) \Rightarrow$

$R_1 = 1$

$R_3 = 1.333$

$R_2 = .571$

try to reduce largest residual by leaving x_1, x_2 alone & change x_3 (relates to R_3) $\Rightarrow x_3 = 1.333$

$\Rightarrow R_3 := 0$

R_1: $-0 - 0 + \frac{1}{8}\left(\frac{4}{3}\right) + 1 = 1.1666$ ◁ highest residual

R_2: $0 - 0 + \frac{2}{7} \cdot \frac{4}{3} + \frac{4}{7} = \frac{8}{21} + \frac{4}{7} = \frac{20}{21} \cong 0.95$

$R_1 = 1.1666 - x_1 + 0 + \overset{0}{} + \frac{8}{21} + 1 \qquad x_1 = -1.1666$

Southwell showed that it is better to under- or over-relax since residuals never stay at zero but change with each iteration.

If we note the direction of relaxation we can anticipate how to make the change.

Best way is to set

mod gauss seidel

$$x_i^{(k+1)} = x_i^{(k)} + \frac{\omega}{a_{ii}}\left[b_i - \sum_{j=1}^{i-1} a_{ij} x_{j\,j<i}^{(k+1)} - \sum_{j=i}^{n} a_{ij} x_j^{(k)} \right]$$

where $1 \le \omega \le 2$ overrelaxation factor

$-1x_1 - .125x_2 + .125x_3 + 1 = R_1$

$.143x_1 - 1x_2 + .286x_3 + .571 = R_2$

$-.222x_1 - .111x_2 - 1x_3 + 1.333 = R_3$

(0,0,0)

$R_1 = 1$
$R_2 = .571$
$R_3 = 1.333$

(0,0, 1.333)

$R_1 = 1.167$
$R_2 = .952$
$R_3 = 0$

($\overset{1.167,\ 0,\ 1.333}{\text{[illegible]}}$)

$R_1 = 0$
$R_2 = 1.119$
$R_3 = -.259$

(1.167, 1.119, 1.333)

$R_1 = -.14025$
$R_2 = 0$
$R_3 = -.383$

∴ 1.333 − .383 = .950

(1.167, $\overset{1.119}{\text{[illegible]}}$, 0.95)

$R_1 = -.188$
$R_2 = -.109$
$R_3 = 0$

∴ 1.167 − .188 = .979

(.979, 1.119, .950)

$R_1 = 0$
$R_2 = -.136$
$R_3 = .041$

∴ 1.119 − .136 = .983

(.979, .983, .950)

$R_1 = .0169$
$R_2 = 0$
$R_3 = 0.0565$

∴ 0.95 + .0565 = 1.0065

Jacobi Method

$$A x = b$$

$$[L + D + U] x = b$$

$$[L + U] X + D X = b \quad \text{thus}$$

$$X = D^{-1} b - D^{-1} [L + U] X$$

$$X = D^{-1} \{ b - [L + U] X \} \qquad \text{here } D = \begin{bmatrix} a_{11} & & 0 \\ & a_{22} & \\ 0 & & \ddots \; a_{nn} \end{bmatrix}$$

$$x_i^{(n+1)} = \frac{b_i}{a_{ii}} - \sum_{\substack{j=1 \\ i \neq j}}^{n} \frac{a_{ij}}{a_{ii}} x_i^{(n)} = x_i^{(n)} + \frac{b_i}{a_{ii}} - \sum_{j=1}^{n} \frac{a_{ij}}{a_{ii}} x_i^{(n)}$$

condition for convergence is $|a_{ii}| > \sum_{\substack{j=1 \\ i \neq j}}^{n} |a_{ij}|$

normally written as $\quad x^{(n+1)} = x^{(n)} + D^{-1} (b - A x^{(n)})$

$$D^{-1} b - D^{-1} (L + D + U) x^{(n)}$$

$$\cancel{x^{(n)}} + D^{-1} b - \{ D^{-1} (L + U) x^{(n)} + \cancel{x^{(n)}} \}$$

Gauss - Seidel

$$X^{(n+1)} = x^{(n)} + (L+D)^{-1} \{ b - A x^{(n)} \}$$

$$= X^{(n)} + (L+D)^{-1} b - [L+D]^{-1} [L+D] x^{(n)} - (L+D)^{-1} U x^{(n)}$$

$$X^{(n+1)} = x^{(n)} + (L+D)^{-1} b - x^{(n)} - (L+D)^{-1} U x^{(n)} \qquad \checkmark\checkmark$$

$$= (L+D)^{-1} [b - U x^{(n)}] \qquad \text{or} \qquad (L+D) x^{(n+1)} = b - U x^{(n)}$$

$$D x^{(n+1)} = b - L x^{(n+1)} - U x^{(n)}$$

$$x^{(n+1)} = D^{-1} [b - L x^{(n+1)} - U x^{(n)}]$$

$$A x = b$$

$$(L + D + U) - (L + D)$$

$$(L+D) \underset{\sim}{x} + U \underset{\sim}{x} = b$$

$$\underset{\sim}{x} = (L+D)^{-1} \underset{\sim}{b} - (L+D)^{-1} U \underset{\sim}{x}$$

$$= (L+D)^{-1} [\underset{\sim}{b} - A \underset{\sim}{x}] + \underset{\sim}{x}$$

$$\underset{\sim}{x}^{(n+1)} = \underset{\sim}{x}^{(n)} + (L+D)^{-1} [\underset{\sim}{b} - A \underset{\sim}{x}^{(n)}]$$

$$(L+D) \underset{\sim}{x}^{(n+1)} = (L+D) \underset{\sim}{x}^{(n)} + b - A \underset{\sim}{x}^{(n)}$$

$$= - U \underset{\sim}{x}^{(n)} + b \qquad \text{or } x^{(n+1)} = [- L x^{(n+1)} - U x^{(n)} + b] / a_{ii}$$

Figure 2.6 *(continued)*

```
    AFTER ITERATION NUMBER 4   X AND F VALUES ARE

          0.56551       2.25006

          0.38258      -0.51029

    AFTER ITERATION NUMBER 5   X AND F VALUES ARE

          0.26959       2.23942

          0.08768      -0.07001

    AFTER ITERATION NUMBER 6   X AND F VALUES ARE

          0.20694       2.22739

          0.00407      -0.00252

    AFTER ITERATION NUMBER 7   X AND F VALUES ARE

          0.20434       2.22671

          0.00000       0.00000

    AFTER ITERATION NUMBER 8   X AND F VALUES ARE

          0.20434       2.22671

          0.00000       0.00000


   THE X-VALUES ARE:

         0.204337

         2.226712
```

Exercises

Section 2.2

1. Given the matrices A, B, and the vectors x, y,

$$A = \begin{bmatrix} 2 & -1 & 3 & -4 \\ 0 & 3 & 4 & -1 \\ 2 & 5 & 5 & 4 \end{bmatrix}, \quad x = \begin{bmatrix} 1 \\ -3 \\ 8 \\ -2 \end{bmatrix},$$

$$B = \begin{bmatrix} 1 & -1 & 6 & -4 \\ 3 & 3 & 4 & 3 \\ 4 & 0 & 2 & -5 \end{bmatrix}, \quad y = \begin{bmatrix} -1 \\ 2 \\ -3 \\ 2 \end{bmatrix},$$

a. Find $3A, 2A + 4B, 2x - 3y$.

▶**b.** Find $A - B, Ax, By$.

▶**c.** Find x^Ty, xy^T.

d. Find B^T.

2. Given the matrices

$$A = \begin{bmatrix} -3 & 1 & -2 \\ 2 & 3 & 0 \\ -1 & 2 & 3 \end{bmatrix}, \quad B = \begin{bmatrix} 3 & -2 & 5 \\ 2 & -4 & 1 \\ -4 & 1 & 6 \end{bmatrix},$$

▶**a.** Find BA, B^3, AA^T.

b. Find $\det(A), \det(B)$.

c. A square matrix can always be expressed as a sum of an upper-triangular matrix U and a lower-triangular matrix L. Find two different combinations of L's and U's such that $A = L + U$.

3. Given the matrices

$$A = \begin{bmatrix} 1 & -2 & 2 \\ 3 & 1 & 1 \\ 2 & 0 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} -1 & -2 & 4 \\ 1 & 3 & -5 \\ 2 & 4 & -7 \end{bmatrix},$$

$$C = \begin{bmatrix} 2 & 1 & 2 \\ 1 & 0 & 4 \\ 1 & 1 & 3 \end{bmatrix},$$

a. Show that $AB = BA = I$, where I is the 3×3 identity matrix. We shall later identify B as the inverse of A.

b. Show that $AI = IA = A$.

c. Show that $AC \neq CA$ and also that $BC \neq CB$. In general, matrices do not commute under multiplication.

▸**d.** A square matrix can be expressed also as the sum of an upper-triangular matrix, a diagonal matrix, and a lower-triangular matrix. Express A as $L + D + U$.

4. Let

$$A = \begin{bmatrix} -2 & 8 \\ -1 & 7 \end{bmatrix}, \quad B = \begin{bmatrix} 8 & -1 & 1 \\ -2 & 2 & -1 \\ -2 & 4 & -3 \end{bmatrix}$$

a. Find the characteristic polynomials of both A and B.

▸**b.** Find the eigenvalues of both A and B.

5. Write as a set of equations:

$$\begin{bmatrix} 2 & 1 & 1 & -2 \\ 4 & 0 & 2 & 1 \\ 3 & 2 & 2 & 0 \\ 1 & 3 & 2 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 0 \\ 8 \\ 7 \\ 3 \end{bmatrix}.$$

▸**6.** Write in matrix form:

$$\begin{aligned} 3x - 6y + 2z &= 15, \\ -4x + y - z &= -2, \\ x - 3y + 7z &= 22. \end{aligned}$$

Section 2.3

7. a. Solve by back-substitution:

$$\begin{aligned} 2x_1 - 3x_2 + x_3 &= -11, \\ 4x_2 - 3x_3 &= -10, \\ 2x_3 &= 4. \end{aligned}$$

▸**b.** Solve by forward-substitution:

$$\begin{aligned} 5x_3 &= 10, \\ 3x_2 - 3x_3 &= 3, \\ 2x_1 - x_2 + 2x_3 &= 7. \end{aligned}$$

✓**8.** Solve the set of equations in Exercise 5.

✓**9.** Solve the set of equations in Exercise 6.

10. Solve the following (given as the augmented matrix):

$$\left[\begin{array}{ccc|c} 1 & 1 & -2 & 3 \\ 4 & -2 & 1 & 5 \\ 3 & -1 & 3 & 8 \end{array}\right].$$

▸**11.** Show that the following does not have a solution:

$$\begin{aligned} 3x_1 + 2x_2 - x_3 - 4x_4 &= 10, \\ x_1 - x_2 + 3x_3 - x_4 &= -4, \\ 2x_1 + x_2 - 3x_3 &= 16, \\ -x_2 + 8x_3 - 5x_4 &= 3. \end{aligned}$$

12. If the right-hand side of Exercise 11 is $(2, 3, 1, 3)^T$, show that there are an infinite number of solutions.

▸**13.** Show that the set of left-hand sides of Exercise 11 are not independent vectors.

Section 2.4

✓**14.** Using Gaussian elimination with partial pivoting and back-substitution,

a. Solve the equations of Exercise 5.

b. Using part (a), find the determinant of the coefficient matrix.

c. What is the LU decomposition of the coefficient matrix? (Rows may have been interchanged.)

▸**15.** Using Gaussian elimination with partial pivoting and back-substitution,

a. Solve the equations of Exercise 10.

b. Using part (a), find the determinant of the coefficient matrix.

c. What is the LU decomposition of the coefficient matrix? (Rows may have been interchanged.)

▸**16. a.** Solve the system

$$\begin{aligned} 2.51x_1 + 1.48x_2 + 4.53x_3 &= 0.05, \\ 1.48x_1 + 0.93x_2 - 1.30x_3 &= 1.03, \\ 2.68x_1 + 3.04x_2 - 1.48x_3 &= -0.53, \end{aligned}$$

by Gaussian elimination, carrying just three significant digits and chopping. Do not interchange rows. Observe that there is a small divisor in reducing the third equation.

b. Repeat part (a), but now use partial pivoting. Observe that there are no small divisors.

c. Substitute each set of answers into the original equations and observe that the left- and right-hand sides match much better with the answers to part (b). The solution, correct to six digits, is

$$x_1 = 1.45310, \quad x_2 = -1.58919, \quad x_3 = -0.27489.$$

17. Solve the systems of Exercises 5, 10, and 11 by the Gauss–Jordan method.

18. Augment the coefficient matrix with all three of the right-hand sides and get all three solutions simultaneously, given

$$A = \begin{bmatrix} 4 & 0 & -1 & 3 \\ 2 & 1 & -2 & 0 \\ 0 & 3 & 2 & -2 \\ 1 & 1 & 0 & 5 \end{bmatrix}, \quad b_1 = \begin{bmatrix} 0 \\ 1 \\ 4 \\ -2 \end{bmatrix},$$

$$b_2 = \begin{bmatrix} 0 \\ 0 \\ -2 \\ 4 \end{bmatrix}, \quad b_3 = \begin{bmatrix} 7 \\ -1 \\ 4 \\ -2 \end{bmatrix}.$$

19. a. For a general $n \times n$ matrix, show that steps 1–4 of Gaussian elimination take at most $n(n-1)(2n-1)/6 + n(n-1)$ multiplications/divisions. You will need to know that

$$1 + 2 + 3 + \cdots + n = \frac{n(n+1)}{2},$$

$$1^2 + 2^2 + 3^2 + \cdots + n^2 = \frac{n(2n+1)(n+1)}{6}.$$

b. Show also that the back-substitution part of Gaussian elimination takes $n(n+1)/2$ multiplications/divisions.

▸**c.** Verify that Gauss–Jordan takes about 50% more operations than Gaussian elimination for the case of three equations. In this, add the number of adds, subtracts, multiplies, and divides.

▸**20.** Suppose we want to solve the system $Az = b$, where the a_{ij}, z_i, and b_i are complex numbers.

a. Show that this can be done using only real arithmetic. (*Hint:* A can be written as $B + Ci$.)

b. If one solves the system in a computer using a language that permits complex numbers, compare the amount of storage space needed compared to the amount if done as in part (a).

21. a. Show that the system

$$\begin{bmatrix} 2 + 2i & -1 + 2i \\ -3i & 3 - 2i \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = \begin{bmatrix} 2 + 2i \\ 1 - 3i \end{bmatrix}$$

can be written as

$$\begin{bmatrix} 2 & -1 & -2 & -2 \\ 0 & 3 & 3 & 2 \\ 2 & 2 & 2 & -1 \\ -3 & -2 & 0 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \\ 2 \\ -3 \end{bmatrix}.$$

▸**b.** Solve the system of part (a), then find z_1 and z_2.

Section 2.5

▸**22.** Use Crout reduction to solve Exercise 10.

23. Use Crout reduction to solve Exercise 11.

24. Repeat Exercise 16, but this time use Crout reduction.

25. Suppose that we do not know all of the three right-hand sides of Exercise 18 in advance.

a. Solve $Ax = b_1$ by Gaussian elimination, getting the LU decomposition. Then use the LU to solve with the other two right-hand sides.

b. Repeat part (a), this time using Crout reduction.

26. Exercise 19 shows that the number of multiplications/divisions to solve a system of n equations is $O(n^3)$. If we already have the LU decomposition of the coefficient matrix, find how many multiplications/divisions are required to first get $Ly = b'$ and then solve from $Ux = b'$. Make a table that compares the total number of multiplications/divisions in the two cases, if $n = 5, 10, 20, 100$.

Section 2.6

27. Which of these matrices are singular?

a.
$$A = \begin{bmatrix} 3 & 2 & -1 \\ 0 & -1 & 4 \\ 6 & 3 & 2 \end{bmatrix}$$

b.
$$B = \begin{bmatrix} 1 & 0 & -2 & 3 \\ 3 & 1 & 1 & 4 \\ -1 & 0 & 2 & -1 \\ 4 & 2 & 6 & 0 \end{bmatrix}$$

c.
$$C = \begin{bmatrix} 1 & 0 & -2 & 3 \\ 3 & 1 & 1 & 4 \\ -1 & 0 & 2 & -1 \\ 4 & 3 & 6 & 0 \end{bmatrix}$$

28. a. Find values of x and y that make A singular:

$$A = \begin{bmatrix} 4 & 1 & 0 \\ 2 & -1 & 2 \\ x & y & -1 \end{bmatrix}.$$

b. Find values for x and y that make A nonsingular.

29. Matrix A in Exercise 27 is singular. Do its rows form linearly independent vectors? Find the values for the a_i in Eq. (2.28).

▸**b.** Repeat part (a) for the elements of A considered as column vectors.

30. Do these sets of equations have a solution? Find a solution if it exists.

a.
$$\begin{cases} 3x - 2y + z = 2, \\ x - 3y + z = 5, \\ x + y - z = -5, \\ 3x + z = 0. \end{cases}$$

▸**b.**
$$\begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 1 \end{bmatrix} x = \begin{bmatrix} 1 \\ -2 \\ 0 \\ 4 \end{bmatrix}.$$

c.
$$\begin{bmatrix} 2 & 1 & 3 \\ -1 & 0 & 2 \\ 6 & 2 & 2 \end{bmatrix} x = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}.$$

d.
$$\begin{bmatrix} 2 & 1 & 3 \\ -1 & 0 & 2 \\ 6 & 2 & 2 \end{bmatrix} x = \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix}.$$

▸**31.** The Hilbert matrix is a classic case of the pathological situation called "ill-conditioning." The 4 × 4 Hilbert matrix is

$$H = \begin{bmatrix} 1 & \frac{1}{2} & \frac{1}{3} & \frac{1}{4} \\ \frac{1}{2} & \frac{1}{3} & \frac{1}{4} & \frac{1}{5} \\ \frac{1}{3} & \frac{1}{4} & \frac{1}{5} & \frac{1}{6} \\ \frac{1}{4} & \frac{1}{5} & \frac{1}{6} & \frac{1}{7} \end{bmatrix}.$$

For the system $Hx = b$, with $b^T = [25/12, 77/60, 57/60, 319/420]$, the exact solution is $x^T = [1, 1, 1, 1]$.

a. Show that the matrix is ill-conditioned by showing that it is nearly singular.

b. Using only three significant digits (chopped) in your arithmetic, find the solution to $Hx = b$. Explain why the answers are so poor.

c. Using only three significant digits, but rounding, again find the solution and compare it to that obtained in part (b).

d. Using five significant digits in your arithmetic, again find the solution and compare it to those found in parts (b) and (c).

Section 2.7

32. Find the determinant of the matrix

$$\begin{bmatrix} 0 & 1 & -1 \\ 3 & 1 & -4 \\ 2 & 1 & 1 \end{bmatrix}$$

by row operations to make it (a) upper-triangular, (b) lower-triangular.

33. Find the determinant of the matrix

$$\begin{bmatrix} 1 & 4 & -2 & 1 \\ -1 & 2 & -1 & 1 \\ 3 & 3 & 0 & 4 \\ 4 & -4 & 2 & 3 \end{bmatrix}.$$

34. Invert the coefficient matrix in Exercise 5, and then use the inverse to generate the solution.

35. If the constant vector in Exercise 5 is changed to one with components (1, 2, 4, 2), what is now the solution? Observe that the inverse obtained in Exercise 34 gives the answer readily.

36. Attempt to find the inverse of the coefficient matrix in Exercise 11. Note that a singular matrix has no inverse.

37. a. Find the determinant of the Hilbert matrix in Exercise 31. A small value of the determinant (when the matrix has elements of the order of unity) indicates ill-conditioning.

▸b. Find the inverse of the Hilbert matrix in Exercise 31. The inverse of an ill-conditioned matrix has some very large elements in comparison to the elements of the original matrix.

▸38. Both Gaussian elimination and the Gauss–Jordan method can be adapted to invert a matrix. In Section 2.7, we say "it is more efficient to use the Gaussian elimination algorithm." Verify this for the specific case of a 3 × 3 matrix by counting arithmetic operations for each method.

Section 2.8

39. For each of the following, evaluate the norms $\|*\|_p, p = 1, 2,$ and ∞. (For a matrix, the 2-norm is the largest magnitude eigenvalue.)

a. $x = [1.15, -2.3, 19.1, 2.0]$

b. $y = [-2, 1, 0, 7, -11]$

c. $$A = \begin{bmatrix} -2 & 8 & 0 \\ -1 & 7 & 0 \\ 0 & 3 & 2 \end{bmatrix}$$

▸d. $$B = \begin{bmatrix} 8 & -2 & 1 \\ -2 & 2 & -1 \\ -2 & 4 & -3 \end{bmatrix}$$

e. Find the norms of $B^2, A + B, AB$.

f. Does the triangle inequality of Eq. (2.29) hold for $A + B$? for $x + y$?

▸40. Find the ∞-norm of the Hilbert matrix of Exercise 31.

41. Find the ∞-norm of the inverse of the Hilbert matrix of Exercise 31.

Section 2.9*

42. Consider the system $Ax = b$, where

$$A = \begin{bmatrix} 3.01 & 6.03 & 1.99 \\ 1.27 & 4.16 & -1.23 \\ 0.987 & -4.81 & 9.34 \end{bmatrix}, \quad b = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}.$$

▸a. Using double precision (or a calculator with 10 or more digits of accuracy), solve for x.

b. Solve the system using three-digit (chopped) arithmetic for each arithmetic operation; call this solution $\bar{x}$.

c. Compare x and $\bar{x}$, and compute $e = x - \bar{x}$. What is $\|e\|_2$?

d. Is the system ill-conditioned? What evidence is there to support your conclusion?

43. Repeat Exercise 42, but change the element a_{33} to -9.34.

▸44. Suppose, in Exercise 42, that uncertainties of measurement give slight changes in some of the elements of A. Specifically, suppose a_{11} is 3.00 instead of 3.01 and a_{31} is 0.99 instead of 0.987. What change does this cause in the solution vector (using precise arithmetic)?

45. Compute the residuals for the imperfect solutions in 42(b) and 43(b). Use double precision in this computation.

46. What are the condition numbers of the coefficient matrices in Exercises 42 and 43? Use the 1-norms.

47. Verify Eq. (2.37), using the results in Exercises 42 and 43.

▸48. Verify Eq. (2.37), using the results in Exercises 43, 45, and 46.

49. Verify Eq. (2.38) for the results of Exercise 44.

50. Apply iterative improvement to the imperfect solution of Exercise 42.

▸51. Apply iterative improvement to the imperfect solution of Exercise 43.

Section 2.10

52. Solve Exercise 6 by the Jacobi method, beginning with the initial vector (0, 0, 0). Compare the rate of convergence when Gauss–Seidel is used with the same starting vector.

53. Solve Exercise 6 by Gauss–Seidel iteration, beginning with approximate solution (2, 2, −1).

▸54. The pair of equations

$$x_1 + 2x_2 = 3,$$
$$3x_1 + x_2 = 4,$$

---

* In certain exercises (42, 43, 44, 50, 51), imperfect solutions will result because low-precision arithmetic is used when the condition number is large. This exaggerates the condition number problem.

can be rearranged to give $x_1 = 3 - 2x_2, x_2 = 4 - 3x_1$. Apply the Jacobi method to this rearrangement, beginning with a vector very close to the solution $x^{(1)} = (1.01, 1.01)^T$, and observe divergence. Now apply Gauss–Seidel. Which method diverges more rapidly?

▸**55.** Solve the system

$$\begin{aligned} 9x + 4y + z &= -17, \\ x - 2y - 6z &= 14, \\ x + 6y &= 4, \end{aligned}$$

a. Using the Gauss–Jacobi method.

b. Using the Gauss–Seidel method. How much faster is the convergence than in part (a)?

Section 2.11

▸**56.** Beginning with (0, 0, 0), use relaxation to solve the system

$$\begin{aligned} 6x_1 - 3x_2 + x_3 &= 11, \\ 2x_1 + x_2 - 8x_3 &= -15, \\ x_1 - 7x_2 + x_3 &= 10. \end{aligned}$$

57. Solve the system in Exercise 55 by relaxation.

58. Relaxation is especially well adapted to problems like Exercise 18. Solve by the relaxation method, starting with the vector [1, 1, 1, 1] which one obtains by inspection. for b, only

Section 2.12

▸**59.** Find the two intersections nearest the origin of the two curves $x^2 + x - y^2 = 1$ and $y - \sin x^2 = 0$.

60. Solve the system

$$\begin{aligned} x^2 + y^2 + z^2 &= 9, \\ xyz &= 1, \\ x + y - z^2 &= 0, \end{aligned}$$

by Newton's method to obtain the solution near (2.5, 0.2, 1.6).

▸**61.** Solve by using Newton's method:

$$\begin{aligned} x^3 + 3y^2 &= 21, \\ x^2 + 2y + 2 &= 0. \end{aligned}$$

Make sketches of the graphs to locate approximate values of the intersections.

62. Apply Eq. (2.58) to compute partials and solve this system by Newton's method:

$$\begin{aligned} xyz - x^2 + y^2 &= 1.34, \\ xy - z^2 &= 0.09, \\ e^x - e^y + z &= 0.41. \end{aligned}$$

There should be a solution near (1, 1, 1).

63. At the end of Section 2.12, it is suggested that it would be more efficient to avoid recomputing the partials at each step of Newton's method for a nonlinear system, doing it only after each nth step when there are n equations. Redo Exercises 59 and 62 using this modification. Compare the rate of convergence with that when the partials are recomputed at each step.

Section 2.13

64. Given matrix A, write the permutation matrix that does the following interchanges.

$$A = \begin{bmatrix} 4 & 7 & 3 & -2 \\ 0 & 0 & 2 & 1 \\ -6 & 1 & 1 & 0 \\ 1 & 0 & 1 & 8 \end{bmatrix}.$$

a. Row 3 with row 1

▸**b.** Row 1 with row 4

▸**c.** Column 2 with column 1

▸**d.** Row 2 with row 4 and column 4 with column 2 simultaneously

65. Confirm that $P^{-1} = P$ for each permutation matrix of Exercise 64.

66. Confirm Eq. (2.61) by first computing $H = H_3P_3H_2P_2H_1P_1$ and then multiplying this times A.

67. Repeat Exercise 66, this time using

$$H = H_3H_2H_1P_3P_2P_1.$$

Section 2.14

(Note that for Exercises 68–76, the matrix A can indicate either the coefficient matrix or the augmented matrix.)

68. Use MAPLE to solve Exercise 10.

a. Use linsolve.

b. Use gaussjord(A) and gausselim(A).

69. Use MAPLE to solve Exercise 2, parts (a) and (b).

70. Use MAPLE to solve Exercise 3.

71. Use MAPLE to solve the system of Exercise 10
  a. through reduced row echelon, rref(A).
  b. using the inverse matrix.
  c. with linsolve(A, b).

72. Use MAPLE to solve Exercise 4.

73. Use MAPLE to get the exact solution to the system given in Exercise 31.

74. Do Exercise 37 with MAPLE.

75. Plot the curves of Exercise 59 to get the approximations, then refine with fsolve.

76. Solve Exercise 61 using MAPLE.

Section 2.15

77. Show that solving an $n \times n$ system by Gaussian elimination requires these numbers of steps:
Making upper-triangular:

$(2n^3 + 3n^2 - 5n)/6$ multiplications/divisions
$(n^3 - n)/3$ additions/subtractions

Back-substitution:

$(n^2 + n)/2$ multiplications/divisions
$(n^2 - n)/2$ additions/subtractions

For a total of

$(n^3 + 3n^2 - n)/3$ multiplications/divisions
$(2n^3 + 3n^2 - 5n)/6$ additions/subtractions

78. Find the equivalent number of operations (as in Exercise 77) for the Gauss–Jordan method.

79. Develop an algorithm for inverting an $n \times n$ nonsingular matrix by parallel processing using approximately n^2 processors.

80. The final algorithm developed in Section 2.15 used $n^2 + n$ processors. Show how this can further be improved so that only $(n + 1)(n - 1) = n^2 - 1$ processors are needed.

81. Develop an algorithm for solving a system of n linear equations by Jacobi iteration using n^2 processors.

Applied Problems and Projects

82. In considering the movement of space vehicles, it is frequently necessary to transform coordinate systems. The standard inertial coordinate system has the N-axis pointed north, the E-axis pointed east, and the D-axis pointed toward the center of the earth. A second system is the vehicle's local coordinate system (with the i-axis straight ahead of the vehicle, the j-axis to the right, and the k-axis downward). We can transform the vector whose local coordinates are (i, j, k) to the inertial system by multiplying by transformation matrices:

$$\begin{bmatrix} n \\ e \\ d \end{bmatrix} = \begin{bmatrix} \cos a & -\sin a & 0 \\ \sin a & \cos a & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \cos b & 0 & \sin b \\ 0 & 1 & 0 \\ -\sin b & 0 & \cos b \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos c & -\sin c \\ 0 & \sin c & \cos c \end{bmatrix} \begin{bmatrix} i \\ j \\ k \end{bmatrix}.$$

Transform the vector $[2.06, -2.44, -0.47]^T$ to the inertial system if $a = 27°$, $b = 5°, c = 72°$.

83. a. Exercise 31 shows the pattern for Hilbert matrices. Find the condition number of the 9×9 Hilbert matrix.
  b. Suppose we have a system of nine linear equations whose coefficients are the 9×9 Hilbert matrix. Find the right-hand side (the b-vector) if the solution vector has ones for all components. Now increase the value of the first component of the b-vector by 1% and find the solution to the perturbed system. Which component of the solution vector is most changed?

84. Electrical engineers often must find the currents flowing and voltages existing in a complex resistor network. Here is a typical problem.

Seven resistors are connected as shown, and voltage is applied to the circuit at points 1 and 6 (see Fig. 2.7). You may recognize the network as a variation on a Wheatstone bridge.

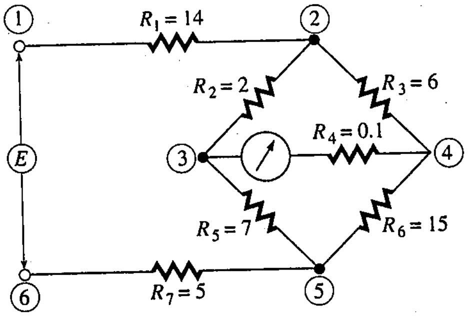


Figure 2.7

While we are especially interested in finding the current that flows through the ammeter, the computational method can give the voltages at each numbered point (these are called *nodes*) and the current through each of the branches of the circuit. Two laws are involved:

Kirchhoff's law: The sum of all currents flowing into a node is zero.

Ohm's law: The current through a resistor equals the voltage across it divided by its resistance.

We can set up eleven equations using these laws and from these solve for eleven unknown quantities (the four voltages and seven currents). If $V_1 = 5$ volts and $V_6 = 0$ volts, set up the eleven equations and solve to find the voltage at each other node and the currents flowing in each branch of the circuit.

85. Mass spectrometry analysis gives a series of peak height readings for various ion masses. For each peak, the height h_j is contributed to by the various constituents. These make different contributions c_{ij} per unit concentration p_i so that the relation

$$h_j = \sum_{i=1}^{n} c_{ij} p_i$$

holds, with n being the number of components present. Carnahan (1964) gives the values shown in Table 2.3 for c_{ij}:

Table 2.3

Peak number	Component				
	CH_4	C_2H_4	C_2H_6	C_3H_6	C_3H_8
1	0.165	0.202	0.317	0.234	0.182
2	27.7	0.862	0.062	0.073	0.131
3		22.35	13.05	4.420	6.001
4			11.28	0	1.110
5				9.850	1.684
6					15.94

If a sample had measured peak heights of $h_1 = 5.20$, $h_2 = 61.7$, $h_3 = 149.2$, $h_4 = 79.4$, $h_5 = 89.3$, and $h_6 = 69.3$, calculate the values of p_i for each component. The total of all the p_i values was 21.53.

86. The truss in Section 2.1 is called *statically determinate* because nine linearly independent equations can be established to relate the nine unknown values of the tensions in the members. If an additional cross brace is added, as sketched in Fig. 2.8, we have ten unknowns but still only nine equations can be written; we now have a statically *indeterminate* system. Consideration of the stretching or compression of the members permits a solution, however. We need to solve a set of equations that gives the displacements x of each pin, which is of the form $ASA^Tx = P$. We then get the tensions f by matrix multiplication: $SA^Tx = f$. The necessary matrices and vectors are

$$A = \begin{bmatrix} 0.7071 & 0 & 0 & -1 & -0.8660 & 0 & 0 & 0 & 0 & 0 \\ 0.7071 & 0 & 1 & 0 & 0.5 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & -0.8660 \\ 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & -0.5 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0.7071 & 0.5 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & -0.7071 & 0.8660 \\ 0 & 0 & 0 & 0 & 0.8660 & 1 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & -0.5 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0.7071 & 0 \end{bmatrix}.$$

S is a diagonal matrix with values (from upper left to lower right) of

$$\begin{matrix} 4255, & 6000, & 6000, & 3670, & 3000, \\ 3670, & 6000, & 6000, & 4255, & 3000. \end{matrix}$$

(These quantities are the values of aE/L, where a is the cross-sectional area of a member, E is the Young's modulus for the material, and L is the length.)

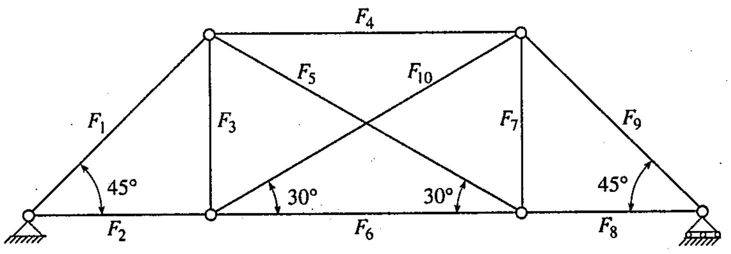


Figure 2.8

Solve the system of equations to determine the values of f for each of three loading vectors:

$$P_1 = [0, -1000, 0, 0, 500, 0, 0, -500, 0]^T,$$
$$P_2 = [1000, 0, 0, -500, 0, 1000, 0, -500, 0]^T,$$
$$P_3 = [0, 0, 0, -500, 0, 0, 0, -500, 0]^T.$$

87. For turbulent flow of fluids in an interconnected network (see Fig. 2.9), the flow rate V from one node to another is about proportional to the square root of the difference in pressures at the nodes. (Thus fluid flow differs from flow of electrical current in a network in that nonlinear equations result.) For the conduits in Fig. 2.9, find the pressure at each node. The values of b represent conductance factors in the relation $v_{ij} = b_{ij}(p_i - p_j)^{1/2}$.

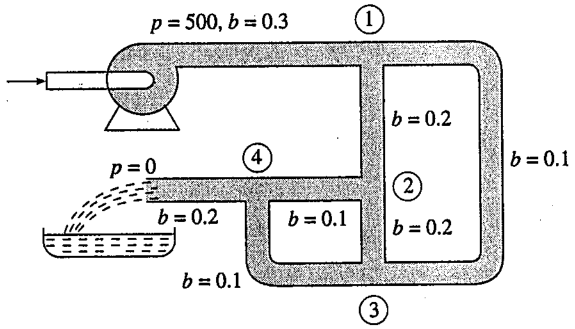


Figure 2.9

These equations can be set up for the pressures at each node:

$$\begin{aligned}
\text{At node 1: } & 0.3\sqrt{500 - p_1} = 0.2\sqrt{p_1 - p_2} + 0.1\sqrt{p_1 - p_3};\\
\text{node 2: } & 0.2\sqrt{p_1 - p_2} = 0.1\sqrt{p_2 - p_4} + 0.2\sqrt{p_2 - p_3};\\
\text{node 3: } & 0.1\sqrt{p_1 - p_3} + 0.2\sqrt{p_2 - p_3} = 0.1\sqrt{p_3 - p_4};\\
\text{node 4: } & 0.1\sqrt{p_2 - p_4} + 0.1\sqrt{p_3 - p_4} = 0.2\sqrt{p_4 - 0}.
\end{aligned}$$

88. Translate each of the programs of this chapter to another computer language.

89. Implement the solutions to Exercise 55 with a spreadsheet.

Intro to Numerical Computations by Yakowitz & Szidarovszky

*7.7
ADAPTIVE STEP-SIZE SELECTION AND ERROR CONTROL

Up to this point we have not discussed how the step size h of the preceding methods is to be chosen. Obviously, there is a trade-off to be made: If the step size is too small, then computer time is needlessly wasted and accumulation of arithmetic roundoff errors can become a hazard. A large step size invites large truncation error associated with higher-order terms neglected in the construction of the methods. For simplicity, our developments will be concerned only with Runge–Kutta rules.

Techniques for automatic step-size selection are based on estimating the local error at each step and then choosing the step size to keep this estimated error within some tolerance bound. Thus step-size selection hinges on estimation of the *local error*, which at the jth step is defined to be

$$\hat{y}(x_{j-1}) - y_{j-1}.$$

Here y_{j-1} is, of course, the computed approximation of $y(x_{j-1})$, and $\hat{y}(x_{j-1})$ we define to be the exact value at x_{j-1} of the differential equation solution that passes through the point (x_j, y_j). That is, $\hat{y}(x)$ solves the initial-value problem

$$\hat{y}' = f(x, \hat{y}), \qquad \hat{y}(x_j) = y_j.$$

In contrast to local errors, the *global error* at x_{j-1} is defined to be

$$y(x_{j-1}) - y_{j-1},$$

where $y(x)$ is the exact solution of the original initial-value problem (7.3). Figure 7.7 illustrates the relationships between $y(x)$, $\hat{y}(x)$, and local and global errors. Intuitively, the local error is the additional truncation error arising from inexact solution at a given step. The global error gives the accumulated total error propagating from the entire sequence of steps.

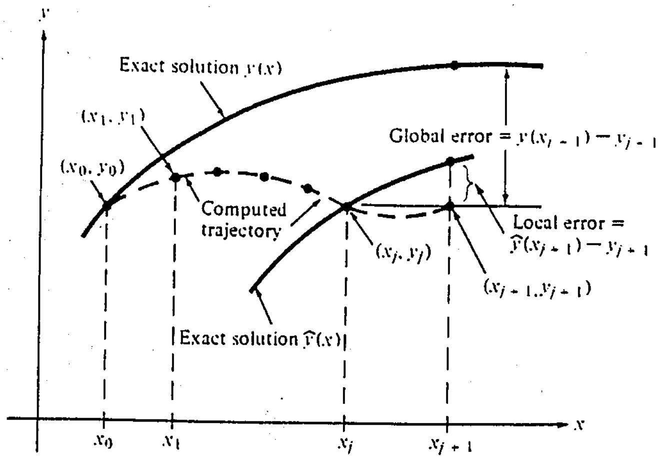


FIGURE 7–7 Relationship Between $y(x)$, $\hat{y}(x)$, Local and Global Errors

Assume that some Runge–Kutta procedure has been selected. We let y_0, y_1, y_2, . . . denote the computed solution values at the arguments x_0, x_1, x_2, ... The local error estimation techniques at each stage apply a higher-order technique to compute an additional approximation, say z_{j+1}, of $y(x_{j+1})$. Since a higher-order technique is used, if the solution is "well behaved" and the step size h is small enough that neglected terms really are negligible, then one may anticipate that the local error of the higher-order method is much less than that of the selected Runge–Kutta procedure. That is,

$$|\hat{y}(x_{j+1}) - z_{j+1}| << |\hat{y}(x_{j+1}) - y_{j+1}|. \tag{7.51}$$

If the approximation above indeed holds, then

$$z_{j+1} - y_{j+1} \simeq \hat{y}(x_{j+1}) - y_{j+1}. \tag{7.52}$$

and we take $z_{j+1} - y_{j+1}$ as the estimate of local error.

Of course, computation of z_{j+1} is typically more expensive than that of y_{j+1} itself, since z_{j+1} must be more accurate. Here, as in other walks of life, information must be paid for. A popular idea toward making this expense as small as possible has been offered by Fehlberg (1964). For a given order, say $p + 1$, the corresponding member of the Fehlberg family computes z_{j+1} with a minimum number of function calls, according to the limitations in Table 7.10, and then provides the pth-order estimate y_{j+1} without any additional function calls. A particularly popular Fehlberg rule is given in Table 7.21, which gives a fifth-order estimate z_{j+1} for a fourth-order rule y_{j+1}.

Subroutine RKF (Table 7.22) implements a single step of this Runge–Kutta–Fehlberg formula, outputting y_{j+1} and z_{j+1} as the parameters YOUT and ZOUT. In view of (7.52), the difference of these values provides a local error estimate. Subroutine ARUKU (Table 7.23) utilizes RKF to update the step size as the computation progresses. If the absolute value of ZOUT-YOUT is less than the user-specified value TOL (for tolerance), the value ZOUT is accepted for y_{j+1}, and a larger step size (by a factor of 3) is chosen for the next step. Otherwise, h is reduced by a factor of 10, and the computation is repeated from the same condition x_j and y_j. Strictly speaking, YOUT, rather than ZOUT, should be chosen for y_{j+1}, but since in principle the higher-order estimate ZOUT should be more accurate, and since it is available, we adopt the pragmatic viewpoint that it should be used. The reader will note that AKUKU is an obvious modification of subroutine ASIMP for adaptive quadrature (Section 3.8.2).

TABLE 7.21 Runge–Kutta–Fehlberg Formula

$$k_1 = f(x_j, y_j)$$

$$k_2 = f\left(x_j + \frac{1}{4}h, y_j + \frac{1}{4}hk_1\right)$$

$$k_3 = f\left(x_j + \frac{3}{8}h, y_j + h\left(\frac{3}{32}k_1 + \frac{9}{32}k_2\right)\right)$$

$$k_4 = f\left(x_j + \frac{12}{13}h, y_j + h\left(\frac{1932}{2197}k_1 - \frac{7200}{2197}k_2 + \frac{7296}{2197}k_3\right)\right)$$

$$k_5 = f\left(x_j + h, y_j + h\left(\frac{439}{216}k_1 - 8k_2 + \frac{3680}{513}k_3 - \frac{845}{4104}k_4\right)\right)$$

$$k_6 = f\left(x_j + \frac{1}{2}h, y_j + h\left(-\frac{8}{27}k_1 + 2k_2 - \frac{3544}{2565}k_3 + \frac{1859}{4104}k_4 - \frac{11}{40}k_5\right)\right)$$

$$y_{j+1} = y_j + h\left(\frac{25}{216}k_1 + \frac{1408}{2565}k_3 + \frac{2197}{4104}k_4 - \frac{1}{5}k_5\right)$$

$$z_{j+1} = y_j + h\left(\frac{16}{135}k_1 + \frac{6656}{12825}k_3 + \frac{28561}{56430}k_4 - \frac{9}{50}k_5 + \frac{2}{55}k_6\right)$$

TABLE 7.22 Subroutine RKF for the Runge–Kutta–Fehlberg Formula

```
      SUBROUTINE RKF(XI,YI,H,YOUT,ZOUT)
C
C  **********************************************************
C  *  FUNCTION: A CALL TO THIS SUBROUTINE COMPUTES ONE STEP OF*
C  *            THE SOLUTION AND A GUESS OF THE ERROR FOR A   *
C  *            DIFFERENTIAL EQUATION Y'=F(X,Y) WITH INITIAL  *
C  *            VALUES XI, YI. THIS SOLUTION IS OBTAINED      *
C  *            USING A 4-TH ORDER RUNGE-KUTTA FEHLBERG STEP  *
C  *            METHOD IMBEDDED IN A 5-TH ORDER STEP SOLUTION *
C  *  USAGE:                                                 *
C  *       CALL SEQUENCE: CALL RKF(XI,YI,H,YOUT,ZOUT)        *
C  *       EXTERNAL FUNCTIONS/SUBROUTINES: FUNCTION F(U,V)   *
C  *  PARAMETERS:                                            *
C  *       INPUT:                                            *
C  *          XI=INDEPENDENT VARIABLE INITIAL VALUE          *
C  *          YI=DEPENDENT VARIABLE INITIAL VALUE            *
C  *           H=INTERVAL STEP SIZE                          *
C  *      OUTPUT:                                            *
C  *        YOUT=4-TH ORDER SOLUTION ESTIMATE                *
C  *        ZOUT=5-TH ORDER SOLUTION ESTIMATE                *
C  *             (ZOUT-YOUT=LOCAL ERROR ESTIMATE)            *
C  **********************************************************
C
      REAL K1,K2,K3,K4,K5,K6
      K1=F(XI,YI)
      U=XI+0.25*H
      V=YI+0.25*H*K1
      K2=F(U,V)
      U=XI+(3./8.)*H
      V=YI+H*((3./32.)*K1+(9./32.)*K2)
      K3=F(U,V)
      U=XI+H*(12./13.)
      V=YI+(H/2197.)*(1932.*K1-7200.*K2+7296.*K3)
      K4=F(U,V)
      U=XI+H
      V=YI+H*((439./216.)*K1-8.*K2+(3680./513.)*K3-
     1(845./4104.)*K4)
      K5=F(U,V)
      U=XI+0.5*H
      V=-(8./27.)*K1+2.*K2-(3544./2565.)*K3+
     1(1859./4104.)*K4-(11./40.)*K5
      V=YI+H*V
      K6=F(U,V)
      YOUT=(25./216.)*K1+(1408./2565.)*K3+
     1(2197./4104.)*K4-K5/5.
      YOUT=YI+H*YOUT
      ZOUT=(16./135.)*K1+(6656./12825.)*K3+
     1(28561./56430.)*K4-(9./50.)*K5
      ZOUT=ZOUT+(2./55.)*K6
      ZOUT=YI+H*ZOUT
      RETURN
      END
```

TABLE 7.23 Subroutine AKUKU for the Adaptive Runge–Kutta Method

```
      SUBROUTINE ARUKU(X,Y,B,M,TOL)
C
C  **********************************************************
C  *  FUNCTION: THIS SUBROUTINE COMPUTES THE SOLUTION OF A    *
C  *            DIFFERENTIAL EQUATION BY ADAPTIVELY CHOOSING  *
C  *            THE STEP SIZE TO LIMIT THE LOCAL ERROR EST-   *
C  *            IMATE WITHIN A GIVEN TOLERANCE. A 4-TH ORDER  *
C  *            RUNGE-KUTTA-FEHLBERG METHOD IS USED           *
C  *  USAGE:                                                  *
C  *       CALL SEQUENCE: CALL ARUKU(X,Y,B,M,TOL)             *
C  *       EXTERNAL FUNCTIONS/SUBROUTINES:                    *
C  *                         SUBROUTINE RKF(XI,YI,H,YOUT,ZOUT) *
C  *  PARAMETERS:                                             *
C  *       INPUT:                                             *
C  *        X(1)=INDEPENDENT VARIABLE INITIAL VALUE           *
C  *        Y(1)=DEPENDENT VARIABLE INITIAL VALUE             *
C  *           B=SOLUTION INTERVAL ENDPOINT (LAST X VALUE)    *
C  *           M=MAXIMUM NUMBER OF ITERATIONS                 *
C  *      OUTPUT:                                             *
C  *           X=M BY 1 ARRAY OF INDEPENDENT VARIABLE VALUES  *
C  *           Y=M BY 1 ARRAY OF DEPENDENT VARIABLE SOLUTION  *
C  *             VALUES                                       *
C  **********************************************************
C
      DIMENSION X(M),Y(M)
C     *** INITIALIZATION ***
      H=.10E-02
      I=1
      N=0
C     *** COMPUTE SOLUTION ITERATIVELY ***
      DO WHILE(X(I).LE.B)
         N=N+1
         CALL RKF(X(I),Y(I),H,YOUT,ZOUT)
         ERR=ZOUT-YOUT
C     *** TEST IF THE NUMBER OF ITERATIONS EXCEEDED ***
         IF(N.GT.M) THEN
            WRITE(6,1)
1           FORMAT(1X,'PROGRAM STOPPED TOO MANY ITERATIONS')
            STOP
         END IF
C     *** TEST STEP SIZE ***
         IF(ABS(ERR).LT.TOL) THEN
            I=I+1
            X(I)=X(I-1)+H
            H=3.0*H
            Y(I)=ZOUT
         ELSE
            H=H/10.0
         END IF
      END DO
      M=I
      H=B-X(I-1)
      X(M)=X(I-1)+H
      CALL RKF(X(I-1),Y(I-1),H,Y(M),ZOUT)
      RETURN
      END
```

We serve notice that the code ARUKU is intended only to illustrate the principles of automatic error control. It is inefficient and does not have the safeguards of a professional differential equation program package. More will be said about this matter after the following computational example.

EXAMPLE 7.17

By means of the calling program given in Table 7.24, the automatic step-size routine ARUKU is called on to solve the differential equation

$$y' = y, \qquad y(0) = 1 \tag{7.53}$$

over the interval [0, 1]. We chose this over our "usual" differential equation because in the present case it is easy to compute the exact local error $\hat{y}(x_{j+1}) - y_{j+1}$ and thereby see how well the RKF error estimator is doing. Specifically, the solution of (7.53) that passes through points (x_j, y_j) is

$$\hat{y}(x) = y_j \exp(x - x_j),$$

and if h is the current step size, then the exact local error is given by

$$y_j \exp(h) - \text{YOUT}.$$

TABLE 7.24 Calling Program for the Subroutine ARUKU

```
C       PROGRAM RKFMETH
C
C       ******************************************************************
C       THIS PROGRAM WILL SET UP AND SOLVE NUMERICALLY THE DIFFERENTIAL
C       EQUATION Y'=Y WITH THE INITIAL CONDITION Y(0.)=1.
C       THE SOLUTION IS OBTAINED USING THE AUTOMATIC STEPSIZE ROUTINE
C       USING 4-TH ORDER RUNGE-KUTTA FEHLBERG METHOD
C       CALLS:   ARUKU,RKF (BOTH MODIFIED FOR DOUBLE PRECISION)
C       OUTPUT:
C                X(I)=VALUE OF X FOR I=1,...  (MAX=50)
C                Y(I)=APPROXIMATED VALUE OF Y AT X(I)
C       ******************************************************************
C
        IMPLICIT DOUBLE PRECISION(A-H,O-Z)
        DIMENSION X(50),Y(50)
C
C       *** FIRST, THE INITIAL CONDITION AND ENDPOINT ARE ESTABLISHED***
C       *** THE MAXIMUM NUMBER OF ITERATIONS IS SET TO 50              ***
C
        X(1)=0.D0
        Y(1)=1.D0
        B=1.D0
        MAX=50
        TOL=1.D-4
C
C       *** SUBROUTINE ARUKU WILL APPROXIMATE THE SOLUTION             ***
C       *** RETURNING AT MOST 50 VALUES IN ARRAYS X AND Y              ***
C
        CALL ARUKU(X,Y,B,MAX,TOL)
        WRITE(10,*)(X(I),Y(I),I=1,50)
        STOP
        END
C
C       *** FUNCTION F SPECIFIES THE DIFFERENTIAL EQUATION. IT IS      ***
C       *** CALLED BY SUBROUTINE RKF                                   ***
C
        FUNCTION F(X,Y)
        IMPLICIT DOUBLE PRECISION(A-H,O-Z)
        F=Y
        RETURN
        END
```

Table 6.1 ADAMS FORMULAS

Order of formula	Coefficient of h	Coefficients of						Local truncation error E_T
		f_i	f_{i-1}	f_{i-2}	f_{i-3}	f_{i-4}	f_{i-5}	
1	1	1						$\frac{1}{2}h^2y''(\xi)$
2	$\frac{1}{2}$	3	-1					$\frac{5}{12}h^3y'''(\xi)$
3	$\frac{1}{12}$	23	-16	5				$\frac{3}{8}h^4y''''(\xi)$
4	$\frac{1}{24}$	55	-59	37	-9			$\frac{251}{720}h^5y^{\text{V}}(\xi)$
5	$\frac{1}{720}$	1901	-2774	2616	-1274	251		$\frac{475}{1440}h^6y^{\text{VI}}(\xi)$
6	$\frac{1}{1440}$	4277	-7923	9982	-7298	2877	-475	$\frac{19087}{60480}h^7y^{\text{VII}}(\xi)$

represents most of the local truncation error. From the definition of the third backward difference given in Section C.2 of Appendix C we can write

$$h(\tfrac{3}{8}\nabla^3 f_i) = \tfrac{3}{8}h(f_i - 3f_{i-1} + 3f_{i-2} - f_{i-3}). \tag{6.114}$$

Backward finite-difference expressions for derivatives are given by Equations (5.123). From the third of Equations (5.123) we deduce that

$$y''''_i = \frac{y'_i - 3y'_{i-1} + 3y'_{i-2} - y'_{i-3}}{h^3} + O(h)$$

or

$$h^3y''''_i = y'_i - 3y'_{i-1} + 3y'_{i-2} - y'_{i-3} + h^3[O(h)]. \tag{6.115}$$

From Equation (6.115) we determine that

$$(f_i - 3f_{i-1} + 3f_{i-2} - f_{i-3}) = h^3y''''_i - O(h^4). \tag{6.116}$$

Combining Equations (6.116) and (6.114),

$$h(\tfrac{3}{8}\nabla^3 f_i) = \tfrac{3}{8}h^4y''''_i - O(h^5) = \tfrac{3}{8}h^4y''''(\xi)$$

where the fourth derivative of y is evaluated at some unknown x value ξ in the range of x values spanned by the one-step application of the third-order Adams formula. Thus, the local truncation error of the third-order method is

$$E_T = \tfrac{3}{8}h^4y''''(\xi) \tag{6.117}$$

as shown in Table 6.1. In similar fashion we can show that

$$h(\tfrac{1}{2}\nabla f_i) = \tfrac{1}{2}h^2y''(\xi)$$

$$h(\tfrac{5}{12}\nabla^2 f_i) = \tfrac{5}{12}h^3y'''(\xi)$$

$$\vdots$$

for the other local truncation errors shown in Table 6.1.

Carrying out the integration indicated, the general formula is

$$y_{i+1} = y_i + h(f_{i+1} - \tfrac{1}{2}\nabla f_{i+1} - \tfrac{1}{12}\nabla^2 f_{i+1} - \tfrac{1}{24}\nabla^3 f_{i+1} - \tfrac{19}{720}\nabla^4 f_{i+1} - \tfrac{27}{1440}\nabla^5 f_{i+1} - \tfrac{863}{60{,}480}\nabla^6 f_{i+1} - \cdots). \qquad (6.137)$$

To obtain the third-order Adams–Moulton corrector formula from this general formula, Equation (6.137) is truncated to three terms following y_i, which yields

$$y_{i+1} = y_i + h(-f_{i+1} - \tfrac{1}{2}\nabla f_{i+1} - \tfrac{1}{12}\nabla^2 f_{i+1}).$$

Then substituting the backward differences as given in Section C.2 in Appendix C into the above, we find

$$y_{i+1} = y_i + \frac{h}{12}(-f_{i-1} + 8f_i + 5f_{i+1}) \qquad (6.138)$$

which is identical with Equation (6.129) derived previously.

The fourth-order Adams–Moulton corrector formula is found by truncating Equation (6.137) to four terms following y_i, yielding

$$y_{i+1} = y_i + h(f_{i+1} - \tfrac{1}{2}\nabla f_{i+1} - \tfrac{1}{12}\nabla^2 f_{i+1} - \tfrac{1}{24}\nabla^3 f_{i+1}).$$

Substituting the backward differences from Section C.2 of Appendix C into the above gives the corrector equation

$$y_{i+1} = y_i + \frac{h}{24}(f_{i-2} - 5f_{i-1} + 19f_i + 9f_{i+1}) \qquad (6.139)$$

which is identical with Equation (6.134) found previously.

Using the terms in Equation (6.137), we can obtain Adams–Moulton corrector formulas of orders 1 through 7, or of orders 1 through 6 plus a local truncation error expression for each of the six. These expressions are given in Table 6.2. In the local truncation-error expressions ξ is an unknown x value in the range of x values spanned in the one-step application of a particular Adams–Moulton formula. While explicit numerical values cannot be found for these expressions because the derivatives in them cannot be evaluated, they are useful for comparing the truncation errors of the various Adams–Moulton formulas, and for comparing the truncation errors of these formulas with those of other

Table 6.2 ADAMS–MOULTON FORMULAS AND LOCAL TRUNCATION-ERROR EXPRESSIONS

Order of formula	Coefficient of h	Coefficients of						Local truncation error E_T
		f_{i+1}	f_i	f_{i-1}	f_{i-2}	f_{i-3}	f_{i-4}	
1	1	1						$-\frac{1}{2}h^2y''(\xi)$
2	$\frac{1}{2}$	1	1					$-\frac{1}{12}h^3y'''(\xi)$
3	$\frac{1}{12}$	5	8	−1				$-\frac{1}{24}h^4y''''(\xi)$
4	$\frac{1}{24}$	9	19	−5	1			$-\frac{19}{720}h^5y^{\mathrm{V}}(\xi)$
5	$\frac{1}{720}$	251	646	−264	106	−19		$-\frac{27}{1440}h^6y^{\mathrm{VI}}(\xi)$
6	$\frac{1}{1440}$	475	1427	−798	482	−173	27	$-\frac{863}{60480}h^7y^{\mathrm{VII}}(\xi)$

does have the disadvantage of not being self-starting, though, as we shall see later, this is not a serious disadvantage.

The development of Milne's method begins by dividing the area under a given portion of a curve $y = f(x)$ into 4 Δx-width strips, as shown in Fig. 6-27. The true area under this portion of the curve is then approximated by considering the area of these 4 strips under a second-degree parabola having 3 coordinates in common with the actual curve, as indicated by the dashed line in Fig. 6-27. The crosshatched area is the approximate area obtained. To determine an expression for this cross-hatched area in terms of Δx and the appropriate y ordinates, it is convenient to consider the 4 strips as centered on the y axis, as shown in Fig. 6-28. This arrangement does not compromise the generality of the results obtained, and it has the advantage of simplifying the intermediate expressions involved in determining the desired form of the expression for the area of 4 such strips. The crosshatched area of Fig. 6-28 is given by

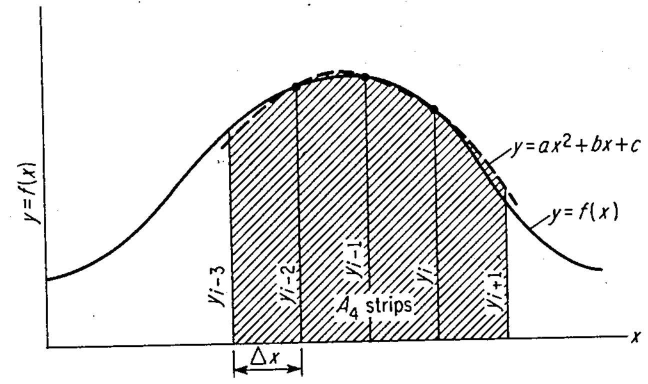


FIG. 6-27. Area in 4 strips under a curve approximated by area under a second-degree parabola.

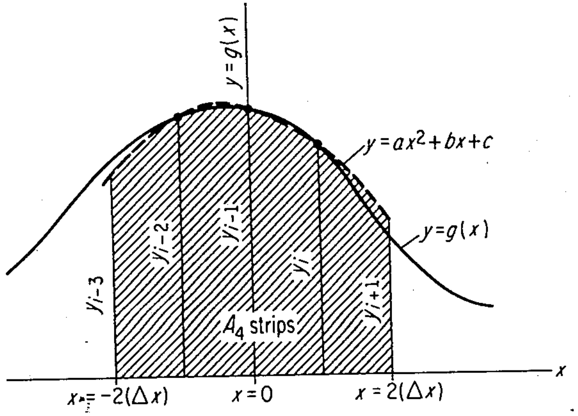


FIG. 6-28. Area of 4 strips centered on y axis.

$$A_{4\,\text{strips}} = \int_{-2(\Delta x)}^{2(\Delta x)} (ax^2 + bx + c)dx \tag{6-83}$$

Integrating Eq. 6-83 and substituting the limits gives

$$A_{4\,\text{strips}} = \frac{16}{3}a(\Delta x)^3 + 4c(\Delta x) \tag{6-84}$$

The constants a and c are determined in the manner explained on p. 286. The appropriate expressions are

$$a = \frac{y_i - 2y_{i-1} + y_{i-2}}{2(\Delta x)^2}$$
$$c = y_{i-1} \tag{6-85}$$

Substituting Eq. 6-85 into Eq. 6-84 yields, for the area of the 4 strips in terms of Δx and the y ordinates shown,

$$A_{4\,\text{strips}} = \frac{4}{3}(\Delta x)[2y_i - y_{i-1} + 2y_{i-2}] \tag{6-86}$$

This expression will be used later as part of the predictor equation.

Let us consider the application of Milne's method in integrating a first-order differential equation of the form

$$y' = f(x, y) \tag{6-87}$$

where the value of y is known for $x = 0$. This technique consists, basically, of obtaining approximate values of y by the use of a *predictor* equation and then correcting these values by the iterative use of a corrector equation. Milne's predictor equation

$$P(y_{i+1}) = y_{i-3} + \frac{4}{3}(\Delta x)[2y_i' - y_{i-1}' + 2y_{i-2}'] \tag{6-88}$$

utilizes the area of 4 strips under a parabolic approximation of a curve (see Eq. 6-86) to provide a predicted value for the successive y ordinates. Milne's *corrector* equation

$$C(y_{i+1}) = y_{i-1} + \frac{\Delta x}{3}[y_{i-1}' + 4y_i' + P(y_{i+1}')] \tag{6-89}$$

provides corrected y values by using Simpson's rule for determining the area of 2 strips under a curve (see p. 286).

Assuming that the resulting y and y' curves of Eq. 6-87 have the general form of the curves shown in Fig. 6-29, the first step is to obtain a predicted value of y_4. Utilizing Eq. 6-88 with $i = 3$,

$$P(y_4) = y_0 + \frac{4}{3}(\Delta x)[2y_3' - y_2' + 2y_1'] \tag{6-90}$$

17. (For Ex. 1): y(0.1) = 1.115895, y(0.5) = 2.027337, using h = 0.1. The error estimates are all less than 2E-7.

(For Ex. 6): y(1) = 1.414217, using h = 0.5; error estimate is 6E-6. (The analytical result is 1.414214, actual error is 3E-6.)

(For Ex. 9): y(0.5) = 3.4429728, using h = 0.25; error estimate is -4.2E-6. With h = 0.5, y(0.5) = 3.443255; error estimate is -9.7E-5.

18.

x	y(computed)	y(analytical)
0.2	-0.781269	-0.781269
0.4	-0.529680	-0.529680
0.6	-0.251188	-0.251188
0.8	0.049329	0.049329
1.0	0.367879	0.367879

19. a) Using RKF with h = 0.5, v(6.0) = 104.149. The error estimate is -6.4E-6. Each figure in the answer is presumably correct.

b) With RK-Merson and h = 0.5, v(6.0) = 104.149.

20. a) -0.28326 b) -0.28387

c) -0.28396 (To get the value with a quartic, add (251/720)hΔ^4f to the results from a cubic.)

21.

t	0.8	1.0	1.2
y	2.0155	2.2843	2.5246
analytical	2.0145	2.2817	2.5199

22. Exact results are obtained because dy/dt is a quadratic.

23.

x	0.8	1.0	1.2	1.4
y	2.20794	2.19060	2.21327	2.29194

24. y(1.2) = 2.5199 versus 2.5199 (analytical).

25. y(4) = 4.2256 (predicted), y(4) = 4.1147 (corrected). The error estimate is +0.0038 so y has 3-digit accuracy. (The actual error is -0.0084). The origianl data must be correct to at least 3-digits.

26. By RKF:

x	0.2	0.4	0.6
y	1.06267	1.24601	1.51691

By Milne:

x	0.8	1.0
y	1.77973	1.89205

27.

x	0.8	1.0	1.2	1.6	2.0
y	2.3164	2.3780	2.4350	2.5381	2.6294
est error	<5E-5	<5E-5	<5E-5	6E-5	<5E-5

(h was increased at x = 1.2).

28. Substitute the values.

29. Divide: (1+2r)/(1-r) and (-1+2r)/(1-r).

30. Substitute definitions of the differences in terms of f-values and collect terms.

31.

x	0.8	1.0	1.2
y	2.0145	2.2817	2.5199

(These match the analytical results.)

32.

x	0.8	1.0
y	1.77999	1.89541

(These are less accurate than by Milne's method.)

33. Using Runge-Kutta:

x	0	0.2	0.4	0.6
y	0	0.00040	0.000640	0.03247

Using Adams-Moulton:

x	0.8	1.0
y	0.10343	0.25757
error	-3E-5	-1.5E-5 (need to decrease h)

Interpolated values (Eq. 5.23):

x	0.5	0.6	0.7	0.8
y	0.01586	0.03247	0.06038	0.10343

Continuing with Adams-Moulton:

x	0.9	1.0	1.1	1.2	1.3	1.4
y	0.16687	0.25712	0.38356	0.55797	0.79932	1.13878
error	-1E-5	-1E-5	-3E-5	-5E-5	-10E-5	-21E-5

(Probably one should decrease h again at x = 1.3.)

WHY PDEs ? IMPORTANCE ?

Physical processes vary with time & location

how will these processes vary with time & space

what drives these processes

where these processes will be in the future & what will happen at some future location

Processes are described by differential equations

if processes depend on 2 or more independent variables then they are governed by partial differential equations

Many processes in natural environs are governed by 2nd order PDEs

$$\frac{\partial^2 u}{\partial x^2} = \frac{1}{c^2}\frac{\partial^2 u}{\partial t^2}$$ wave eqn.

u - displacement

c - bar velocity $= \sqrt{E/\rho}$

$$\frac{\partial T}{\partial t} = \alpha\left(\frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2}\right)$$ heat eqn in 2-D

T - temp

α - thermal diff.

$\alpha = \frac{k}{c\rho}$

Fick's Law of Diff. $\frac{\partial C_A}{\partial t} = D_{AB}\nabla^2 C_A$

$$\frac{\partial^2 \phi}{\partial x^2} + \frac{\partial^2 \phi}{\partial y^2} = \nabla^2\phi = 0$$ potential flow

$\underline{V} = \nabla\phi$

$\nabla\cdot\underline{V} = 0$

$$\frac{\partial u(x,y)}{\partial x} = \lim_{\Delta x \to 0} \frac{u(x+\Delta x, y) - u(x,y)}{\Delta x}$$

WAVE EQN - APPEARS IN ACOUSTICS, VIBS, SHALLOW WATER WAVE THEORY

HEAT EQN - " IN HEAT TRANSFER, DIFFUSION PROBLEMS

LAPLACES EQ - ' IN POTENTIAL FLOW THEORY, STEADY STATE HEAT TRANS.
TORSION OF BAR, STEADY STATE VIB. OF MEMBRANE

WHAT'S COMMON TO ALL
HOW CAN THEY BE SOLVED
WHAT IS A WELL POSED PROBLEM

CHARACTERIZATION & CLASSIFICATION

GENERAL 2nd ORDER PDE

$$\varphi(x, y, u, u_x, u_y, u_{xx}, u_{xy}, u_{yy}) = 0 \qquad u(x,y)$$

$$u_{xx} \equiv \frac{\partial^2 u}{\partial x^2}$$

THE EQ. IS LINEAR WRT ITS HIGHEST DERIVS.

$$A u_{xx} + B u_{xy} + C u_{yy} + F(x, y, u, u_x, u_y) = 0$$

A, B, C are fns of x & y ~~alone~~ only

when $F(x, y, u, u_x, u_y)$ then equation is quasi linear

IF $F(x, y, u, u_x, u_y) = \underbrace{D u_x + E u_y + F u + G}_{\text{are fns of } x \,\&\, y \text{ only}}$ then equation is linear

IF $G = 0$ then equation homogeneous

2$^{\underline{nd}}$ order linear PDE

$$A u_{xx} + B u_{xy} + C u_{yy} + D u_x + E u_y + F u + G = 0$$

IF ALONG SOME CURVE $\varphi(x,y) = \text{const}$ the equation $A(y')^2 - B(y') + C = 0$

THEN

$$\frac{dy}{dx} = y' = \frac{B \pm \sqrt{B^2-4AC}}{2A}$$ & φ is called the characteristic

$B^2-4AC > 0$ PDE is hyperbolic & 2 characteristics exist ; $\frac{dy}{dx}$ is real

$B^2-4AC = 0$ PDE is parabolic & 1 real characteristic exists

$B^2-4AC < 0$ PDE is elliptic & no real characteristics exist

Wave Eqn

$$\frac{\partial^2 u}{\partial x^2} - \frac{1}{c^2}\frac{\partial^2 u}{\partial t^2} = 0 \qquad A=1 \quad B=0 \quad C=-\frac{1}{c^2}$$

$t \to y$

$$\frac{dy}{dx} \to \underset{\substack{\uparrow \\ \text{slowness}}}{\frac{dt}{dx}} = \frac{0 \pm \sqrt{0-4\cdot1\cdot\left(-\frac{1}{c^2}\right)}}{2\cdot1} = \pm\frac{1}{c}$$

$$\frac{dx}{dt} = \text{velocity} = \frac{1}{\text{slowness}} = \pm c \qquad \text{in magnitude } |c| < \infty$$

$$\begin{aligned} x &= +ct + \mathscr{C}_1 & \quad x - ct &= \mathscr{C}_1 \\ x &= -ct + \mathscr{C}_2 & \quad x + ct &= \mathscr{C}_2 \end{aligned} \Bigg\} \quad \begin{aligned} \varphi_1(x,t) &= x - ct \\ \varphi_2(x,t) &= x + ct \end{aligned}$$

$$\frac{\partial T}{\partial t} = r\frac{\partial^2 T}{\partial x^2} \qquad A = r \quad B = 0 \quad C = 0$$

$t \rightarrow y$

$T \rightarrow u$

$$\frac{dy}{dx} \rightarrow \frac{dt}{dx} = \frac{B \pm \sqrt{B^2 - 4AC}}{2A} = 0$$

$\downarrow$

slowness

velocity $\frac{dx}{dt} = \infty$

$$t = \text{const} = \int \frac{dt}{dx}\,dx = \int 0\,dx$$

$$t = 0 + \mathcal{C}$$

$$\varphi(x,t) = t$$

2D steady state heat equation - ELLIPTIC.

$$\frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2} = 0 \qquad A = 1,\ B = 0,\ C = 1$$

$T \rightarrow u$

$$\frac{dy}{dx} = \frac{B \pm \sqrt{B^2 - 4AC}}{2A} = \pm i$$

LETS LOOK AT $\frac{\partial u}{\partial t} = \alpha \frac{\partial u}{\partial x^2}$ one dim heat equation

EXPLICIT SCHEME $u_{i,j+1} = u_{i,j} + \overset{C}{\boxed{\frac{\alpha \Delta t}{\Delta x^2}}} \left[u_{i+1,j} - 2u_{i,j} + u_{i-1,j} \right]$

U^j is the amplitude at t_j,

NEUMANN STABILITY ANALYSIS let $u_{ij} = U^j e^{Iik\Delta x}$ $I = \sqrt{-1}$

$$e^{Iik\Delta x} \left\{ U^{j+1} = U^j + U^j C \left[e^{Ik\Delta x} - 2 + e^{-Ik\Delta x} \right] \right\}$$

$$U^{j+1} = U^j \left\{ \underbrace{1 + 2C\left[\cos k\Delta x - 1\right]}_{A} \right\} = U^0 A^{j+1}$$

k is sep var con

$i \Delta x = x_i$

$e^{I\zeta} + e^{-I\zeta} = 2\cos\zeta$

note U^{j+1} depends on U^j. If we go back U^{j+1} depends on U^0. Any error in U^0 is amplified when $|A|>1$ & dies if $|A|<1$

if $|A|$ is >1 ⟹ solution becomes unbounded

FOR BOUNDEDNESS let $\left\{ 1 + 2C\left[\cos k\Delta x - 1\right] \right\}^2 \leq 1$

$$4C[\quad] + (2C)^2[\quad]^2 \leq 0$$

$$4C \underbrace{\left[\cos k\Delta x - 1\right]}_{-2 \leq [\ ] \leq 0} \left\{ \underbrace{1 + C\left[\cos k\Delta x - 1\right]}_{\text{must be} \geq 0} \right\} \leq 0$$

$$C \leq \frac{1}{1 - \cos k\Delta x}$$

$1 + C\left[\cos k\Delta x - 1\right] \geq 0$

$1 - C\left[1 - \cos k\Delta x\right] \geq 0$

$\frac{1}{[1 - \cos k\Delta x]} \geq C$

$$C \in [\overset{\cos k\Delta x = -1}{\tfrac{1}{2}} \ \& \ \overset{\cos k\Delta x = 1}{\infty}]$$ take lower limit $C \leq \frac{1}{2}$

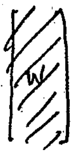
Bender-Schmidt condition

IF $C = \frac{1}{2}$ $u_{i,j+1} = \frac{1}{2}\left[u_{i+1,j} + u_{i-1,j}\right]$

- when deriv b.c. are used $C < \frac{1}{2}$
- discont. in the initial conditions can cause errors in your solution

explicit scheme

$$A = \left[C \quad 1-2C \quad C \right]$$



time j
space i

$$w^j = \begin{bmatrix} w_{i-1,j} \\ w_{i,j} \\ w_{i+1,j} \end{bmatrix}$$

$\lambda_n = 1 - 4C \sin^2 \frac{n\pi}{2(N+1)}$

N = # of data points

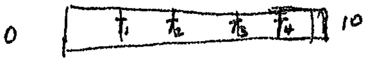
$$w^{j+1} = A w^j = A^{j+1} w^{(0)}$$

error introduced at $w^{(0)}$ will die out if eigenvalues of A are < 1

since $\underset{\sim}{w}^{(0)} = \underset{\sim}{w}^{(0)}_{act} + \underset{\sim}{\epsilon}$

$$w^{j+1} = A^{j+1} w^{(0)} + A^{j+1} \epsilon$$

if $\|A\| < 1$ then $\underset{\sim}{A}^{j+1} \underset{\sim}{\epsilon} \leq \|A\|^{j+1} \|\epsilon\|$



$$-0\cdot R \quad 1+2RT_1 \quad -RT_2 \qquad = T_1^{0}$$
$$-RT_1 + (1+2R)T_2 - RT_3 \quad = T_2^{0}$$
$$-RT_2 + (1+2R)T_3 - RT_4 = T_3^{0}$$
$$-RT_3 + (1+2R)T_4 - R\cdot 10 = T_4^{0}$$

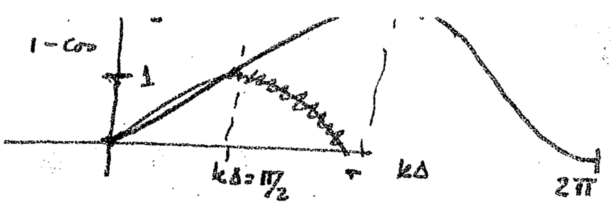
$$\begin{bmatrix} 1+2R & -R & & \\ -R & 1+2R & -R & \\ & -R & 1+2R & -R \\ & & -R & 1+2R \end{bmatrix} \begin{pmatrix} T_1 \\ T_2 \\ T_3 \\ T_4 \end{pmatrix} = \begin{pmatrix} RT_1^{0} \\ T_2^{0} \\ T_3^{0} \\ T_4^{0}+10R \end{pmatrix}$$

implicit BD

$$w^{j} = Aw^{j+1} = \begin{bmatrix} -C & 1+2C & -C \\ & & \end{bmatrix} W^{j+1}$$

$$\therefore \quad W^{(0)} = A^{j+1} w^{j+1} \quad \rightarrow \quad w^{j+1} = (A^{-1})^{j+1} w^{(0)}$$

λ_n here is $\dfrac{1}{1+4C\sin^2\dfrac{n\pi}{2(N+1)}}$



for solution C can vary from $\frac{1}{2}$ to ∞

for stability $\frac{\alpha\Delta t}{\Delta x^2} \leq \frac{1}{2}$ conditionally stable since it depends on $\underline{\Delta t}$

- more accurate results are found for $\frac{\alpha\Delta t}{\Delta x^2} \leq \frac{1}{[illegible]}$
- in fact modal analysis can be used to show that, for $\frac{\alpha\Delta t}{\Delta x^2} < \frac{1}{4}$

$$\lambda_n = 1 - 4C\sin^2\frac{n\pi}{2(N+1)} \qquad n = 1, 2, \ldots N \quad \text{for } \lambda_n < 1 \quad C < \frac{1}{4}$$

method is stable non-oscillatory

- What about <u>backward difference in time</u> implicit scheme eval @ i, j+1

~~$U_{i,j} = U_{i,j} + \frac{\alpha\Delta t}{\Delta x^2}\left[U_{i,j+1} - 2U_{i,j} + U_{i,j-1}\right]$~~

$$\left.\frac{\partial T}{\partial t}\right|_{ij+1} = \alpha\left.\frac{\partial^2 T}{\partial x^2}\right|_{ij+1}$$

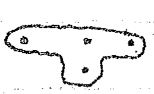
$$T_{ij} = T_{ij+1} - \Delta t\frac{\partial T}{\partial t}_{ij+1}$$

$$U_{ij+1}(1+2R) - U_{i+1j+1}R - RU_{i-1j+1} = U_{ij}$$

$$U_{i,j+1} = U_{ij} + \frac{\alpha\Delta t}{\Delta x^2}\left[U_{i+1,j+1} - 2U_{i,j+1} + U_{i-1\,j+1}\right]$$

$$O(\Delta t, \Delta x^2)$$

Using $U_{ij} = U^j e^{Iik\Delta x}$



$$e^{Iik\Delta x}\left\{U^{j+1} = U^j + C\left[U^{j+1}e^{Ik\Delta x} - 2U^{j+1} + U^{j+1}e^{-Ik\Delta x}\right]\right\}$$

$$U^{j+1} = U^j + CU^{j+1}\left[2(\cos k\Delta x - 1)\right]$$

$$U^{j+1}\left\{1 - 2C(\cos k\Delta x - 1)\right\} = U^j \qquad \therefore \quad U^{j+1} = RU^j$$

$$R = \frac{1}{1 - 2C\left[\cos k\Delta x - 1\right]} \leq 1$$

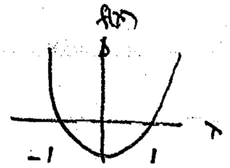
this number varies from $\frac{1}{1+4C} \leq R \leq 1$

this is stable for any $C \geq 0$ since it is stable irrespective of Δt

it is unconditionally stable – but take C small since error is $O(\Delta t)$

for $A = 0 \Rightarrow C = 0$ impossible since $C = \frac{\alpha \Delta t}{\Delta x^2} > 0$

$A = 0$ eqn $= \lambda^2 - 1$



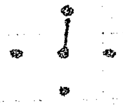
as $A \uparrow$ curve moves to right both roots are real

and $\underbrace{\text{one root } |\lambda|<1}_{\text{goes to zero}}$ but the $\underbrace{\text{other } |\lambda|>1}_{\text{goes to } \infty}$

$$\therefore \quad U^{j+1} = [\underbrace{c_1\lambda_1^{j+1}}_{\text{can go to zero}} + \underbrace{c_2\lambda_2^{j+1}}_{\text{goes to infinity}}]U^0$$

LET'S LOOK AT THE CENTERED DIFF IN TIME (LEAPFROG OR RICHARDSON'S)

$$\left.\frac{\partial u}{\partial t}\right|_{x_i,t_j} = \alpha \left.\frac{\partial^2 u}{\partial x^2}\right|_{x_i,t_j} \qquad O(\Delta t^2, \Delta x^2)$$



$$u_{ij+1} = u_{ij-1} + \frac{2\alpha\Delta t}{\Delta x^2}\left[u_{i+1j} - 2u_{i,j} + u_{i-1j}\right] \qquad \text{let } u_{ij} = U^j e^{Iik\Delta x}$$

$$e^{Iik\Delta x}\left\{ U^{j+1} = U^{j-1} + 2c\left[e^{Ik\Delta x}U^j - 2U^j + U^j e^{-Ik\Delta x}\right]\right\}$$

$$4c\left[U^j\left(\cos k\Delta x - 1\right)\right] \qquad A = -4c\,(\cos k\Delta x - 1)$$

if we assume $U^j = \lambda U^{j-1}$ $\qquad U^{j+1} = \lambda U^j = \lambda^2 U^{j-1} = \lambda^{j+1}U^0$

$$e^{Iik\Delta x}\,U^{j-1}\left[\lambda^2 = \lambda\{4c(\cos k\Delta x - 1)\} + 1\right] \quad \text{or} \quad \lambda^2 + 4c(1-\cos k\Delta x)\lambda - 1 = 0$$

$$\lambda^2 = 1 \mp A\lambda \quad \Rightarrow \quad \lambda^2 + A\lambda - 1 = 0 \qquad A = 4c\,(1-\cos k\Delta x)$$

$$0 \le A \le 8c$$

$$\lambda = \frac{-A \pm \sqrt{A^2+4}}{2} \qquad \lambda = \frac{-A + \sqrt{A^2 + 4}}{2}$$

$$= -\left(\frac{A}{2}\right) \pm \sqrt{1 + (A/2)^2} \cong -\left(\frac{A}{2}\right) \pm \left[1 + \frac{1}{2}\left(\frac{A}{2}\right)^2 + \dots\right] \qquad \lambda = \frac{-A - \sqrt{A^2+4}}{2} \geq 1$$

$|\lambda| < 1$ cannot be satisfied

$$\frac{-A - \sqrt{A^2+4}}{2}$$

no matter what A is

unless A is imag

Du Fort - Frankel Scheme explicit scheme $O(\Delta t^2, \Delta x^2)$

$$\frac{u_{ij+1} - u_{ij-1}}{2\Delta t} = \alpha\left[\frac{u_{i-1j} - u_{ij-1} - u_{ij+1} + u_{i+1j}}{\Delta x^2}\right]$$

eval $\left.\frac{\partial u}{\partial t}\right|_{i,j}$

u_{ij} replaced by $\frac{u_{i,j-1} + u_{i,j+1}}{2}$

$$u_{ij+1} = u_{ij-1} + \frac{2\alpha\Delta t}{\Delta x^2}\left[u_{i-1j} - 2u_{ij} + u_{i+1j}\right]$$

HW: show it is unconditionally stable.

why does this work since $w^{(1)} = A\left[w^{(0)} + e^{(0)}\right]$

$$w^{(n)} = Aw^{(n-1)} + \dots = A^n e^{(0)}$$

$$\|A^n e^{(0)}\| \le \|A^n\|\,\|e^{(0)}\| \le e^{(0)} \text{ if}$$

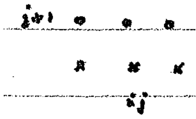
$$\|A^n\| < 1$$

requires the eigenvalues of $A \le 1$

Crank - Nicholson we consider $\frac{\partial u}{\partial t}\left(t_j+\frac{\Delta t}{2}\right) = \frac{u_{i,j+1}-u_{i,j}}{2\,(\Delta t/2)} + O\left(\frac{\Delta t}{2}\right)^2$

if we do that then we must take $\frac{\partial^2 u}{\partial x^2}$ also at $t_j+\frac{\Delta t}{2}$

but we define $\frac{\partial^2 u}{\partial x^2}\left(t_j+\frac{\Delta t}{2}\right) = \frac{1}{2}\left[\frac{\partial^2 u}{\partial x^2}(t_j+\Delta t) + \frac{\partial^2 u}{\partial x^2}(t_j)\right]$



then $$\frac{u_{i,j+1}-u_{i,j}}{\Delta t} = \frac{\alpha}{2}\left[\underbrace{\frac{u_{i+1,j+1}-2u_{i,j+1}+u_{i-1,j+1}}{\Delta x^2}}_{CD\ @\ t_j+\Delta t} + \underbrace{\frac{u_{i+1,j}-2u_{i,j}+u_{i-1,j}}{\Delta x^2}}_{CD\ @\ t_j}\right]$$

implicit since information at j+1 depends upon information at adjacent points at j+1 as well

then if $C = \frac{\alpha\Delta t}{\cancel{2}\Delta x^2}$ we get $-Cu_{i-1,j+1} + (2+2C)u_{i,j+1} - Cu_{i+1,j+1} =$

$$Cu_{i-1,j} + (2-2C)u_{i,j} + Cu_{i+1,j}$$

$$\begin{bmatrix} -C & 2+2C & -C \end{bmatrix} \begin{bmatrix} u_{i-1} \\ u_i \\ u_{i+1} \end{bmatrix}_{j+1} = \begin{bmatrix} C & 2-2C & C \end{bmatrix} \begin{bmatrix} u_{i-1} \\ u_i \\ u_{i+1} \end{bmatrix}_j$$

or

$$A w^{j+1} = B w^j$$

if $C=1$ then $\begin{bmatrix} -1 & 4 & -1 \end{bmatrix} \begin{bmatrix} u_{i-1} \\ u_i \\ u_{i+1} \end{bmatrix}_{j+1} = \begin{bmatrix} 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} u_{i-1} \\ u_i \\ u_{i+1} \end{bmatrix}_j$

$$w^{j+1} = A^{-1}Bw^j$$

λ_n for $A^{-1}B$ is $\frac{2-4C\sin^2(n\pi/2(N-1))}{2+4C\sin^2(n\pi/2(N-1))}$ ≤ 1 for any $C \geq 0$

< 1 for any $C > 0$

seek a function $u(x, t)$, which is defined in the interior of a rectangle $ABCD$. This region is already determined by the statement of the problem, since the course of the heat propagation in the rod $0 \leqq x \leqq l$ during the time interval $t \leqq t = T$, in which the heat behavior of the boundary is known, was already investigated. Let $t_0 = 0$; we assume that $u(x, t)$ satisfies the heat-conduction equation only for $0 < x < l, 0 < t \leqq T$, i.e., not for $t = 0$ (the side AB) or for $x = 0, x = l$ (the sides AD and BC). For $t = 0$, as well as $x = 0$ and $x = l$, the value of this function is given directly by the initial and boundary conditions. To require that the heat-conduction equation, for example, be satisfied also for $t = 0$ would imply that the derivative $\varphi'' = u_{xx}(x, 0)$ in this equation exists. Therefore, the generality of the physical phenomena to be investigated is limited, and thus the basic functions which do not satisfy this requirement are eliminated from consideration. The condition (3-1.3) loses its meaning when it is not required that $u(x, t)$ in the region $0 \leqq x \leqq l, 0 \leqq t \leqq T$ (i.e., in the closed rectangle $ABCD$) be continuous or this requirement must be replaced by another appropriate assumption.[46] To understand the significance of this requirement we consider the function $v(x, t)$ defined by the following conditions:

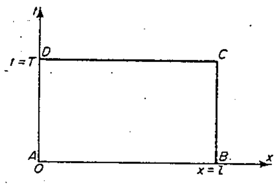


FIG. 37.

$$v(x, t) = C\,, \qquad 0 < x < l\,, \qquad 0 < t \leqq T$$
$$v(x, 0) = \varphi(x)\,, \qquad 0 \leqq x \leqq l$$
$$v(0, t) = \mu_1(t)\,, \qquad v(l, t) = \mu_2(t)\,, \qquad 0 \leqq t \leqq T$$

where C is an arbitrary constant. The function v obviously satisfies both condition (3-1.2) and the boundary conditions. However, this function in no case describes the course of the heat distribution in the rod with an initial temperature $\varphi(x) \neq C$ and boundary temperatures $\mu_1(t) \neq C$ and $\mu_2(t) \neq C$, since it is discontinuous for $t = 0, x = 0, x = l$.

The continuity of $u(x, t)$ for $0 < x < l, 0 < t < T$ directly follows in that $u(x, t)$ satisfies the differential equation. Therefore, the requirement that $u(x, t)$ be continuous in $0 \leqq x \leqq l, 0 \leqq t \leqq T$, is based essentially only on those points at which the boundary and the initial values are prescribed. In the following, by a solution of the equation which satisfies the boundary conditions, we shall always mean a function which satisfies the requirements (3-1.1), (3-1.2), and (3-1.3) and hence not repeat these each time, unless there are special conditions.

Correspondingly, this is the case for other boundary-value problems, in particular for problems of an infinite rod and problems without initial conditions.

---

[46] Later, boundary-value problems with discontinuous boundary and initial conditions will be considered. For these, the problems will be properly defined so that the boundary conditions are fulfilled.

Partial Diff. Eqns. of Math. Physics
by Tykhonov & Samarski

For problems with several independent geometric variables the above statements remain valid. In these problems, an initial temperature and boundary conditions determined on the surface of the body are prescribed for $t = 0$. We can also investigate problems for infinite domains.

With regard to all the problems discussed, the following problems exist[47]:

1. Are the solutions of the problems discussed uniquely determined?
2. Does a solution exist?
3. Do the solutions depend continuously on the auxiliary conditions?

If a problem admits of many solutions, then we naturally cannot speak of "the solution of the problem," and we must first prove the uniqueness. In practice, the second question above is the most important, since generally in proving the existence of a solution, we simultaneously find methods for its calculation.

As noted earlier (see Section 2-2 §3) we speak of a physically determined process when a small change in the initial or boundary conditions causes a small change in the solution. In the following, it will be shown that heat propagation is determined physically by the initial and boundary conditions, i.e., a small change in the initial or boundary conditions implies a small change in the solution.

5. The principle of the maximum

In the following we shall investigate differential equations with constant coefficients,

$$v_t = a^2 v_{xx} + \beta v_x + \gamma v\,. \tag{3-1.34}$$

As already shown, these equations, by the substitution of

$$v = e^{\mu x + \lambda t} u \quad \text{with} \quad \mu = -\frac{\beta}{2a^2}\,, \quad \lambda = \gamma - \frac{\beta^2}{4a^2}$$

can be brought to the form

$$u_t = a^2 u_{xx}\,. \tag{3-1.35}$$

The solutions of this equation have the following properties which will be denoted as the principle of the maximum.

A function $u(x, t)$ defined and continuous in the closed region $0 \leqq t \leqq T$, $0 \leqq x \leqq l$ and satisfying the heat-conduction equation

$$u_t = a^2 u_{xx} \tag{3-1.35}$$

in the region $0 < t < T, 0 < x < l$ assumes its maximum or minimum at the initial moment $t = 0$ or at the boundary points $x = 0$ or $x = l$.

Before we prove this, note that the function $u(x, t) =$ const. obviously satisfies the heat-conduction equation and assumes a maximum (minimum) at each point. However, this does not contradict our assertion, because it means only that when a maximum (minimum) is assumed in the interior of the region it is also (but not only) assumed for $t = 0$ or for $x = 0$ or $x = l$.

---

[47] *Cf.* Section 2-2.

The physical significance of this statement is immediately clear: if the temperature on the boundary and at the initial moment does not exceed a value M, then in the interior of the body no temperature higher than M can be attained. We shall limit ourselves to the proof of the statement of the maximum and give an indirect proof. We shall designate by M the maximum value of $u(x, t)$ for $t = 0$ $(0 \leqq x \leqq l)$ or for $x = 0$ or $x = l$ $(0 \leqq t \leqq T)$ and assume that the function $u(x, t)$ assumes its maximum at an interior point (x_0, t_0), $(0 < x_0 < l, 0 < t_0 \leqq T)$:[48]

$$u(x_0, t_0) = M + \varepsilon .$$

We now compare the signs in Eq. (3-1.35) at the point (x_0, t_0). Since the function at (x_0, t_0) assumes its maximum,[49] then necessarily

$$\frac{\partial u}{\partial x}(x_0, t_0) = 0 \qquad \text{and} \qquad \frac{\partial^2 u}{\partial x^2}(x_0, t_0) \leqq 0 . \tag{3-1.36}$$

Also, since $u(x_0, t)$ for $t = t_0$ has a maximum,[50] then

$$\frac{\partial u}{\partial t}(x_0, t_0) \geqq 0 . \tag{3-1.37}$$

By comparison of the signs on the left and right sides of (3-1.35) it follows that both sides can be different. These considerations, however, still do not prove the correctness of our theorem; since the right and the left sides can simultaneously equal zero, it would signify no contradiction. We bring forth this consideration simply to emphasize the fundamental concepts of our proof. For the completion of the proof we shall seek more than one point (x_1, t_1) at which $\partial^2 u/\partial x^2 \leqq 0$ and $\partial u/\partial t > 0$. Therefore, we consider the auxiliary function

$$v(x, t) = u(x, t) + k(t_0 - t) , \tag{3-1.38}$$

where k is a constant. Obviously then

$$v(x_0, t_0) = u(x_0, t_0) = M + \varepsilon$$

and

$$k(t_0 - t) \leqq kT .$$

---

[48] If the continuity of $u(x, t)$ were assumed in the bounded region $0 \leqq x \leqq l$, $0 \leqq t \leqq T$, then the function $u(x, t)$ could not exceed its maximum, and further considerations would be contradictory. On the basis of the theorem that every continuous function in a bounded region attains its maximum, then (a) the function $u(x, t)$ attains a maximum within or on the boundaries which will be denoted by M; (b) if $u(x, t)$ also were to exceed M only at a point, then a point (x_0, t_0) would exist at which the function $u(x, t)$ assumes a maximum which is larger than M: $u(x_0, t_0) = M + \varepsilon$ $(\varepsilon > 0)$, where $0 < x_0 < l$, $0 < t_0 \leqq T$.

[49] As is known from analysis, for the existence of a relative minimum of a function $f(x)$ at an interior point x_0 of an interval $(0, l)$, the conditions

$$\left.\frac{\partial f}{\partial x}\right|_{x=x_0} = 0, \left.\frac{\partial^2 f}{\partial x^2}\right|_{x=x_0} > 0$$

are sufficient. If, therefore, at the point x_0 the function $f(x)$ has a maximum value, then (a) $f'(x_0) = 0$, and (b) $f''(x_0) > 0$ cannot hold; therefore $f''(x_0) \leqq 0$.

[50] Obviously, $\partial u/\partial t = 0$, in case $t_0 < T$, whereas for $t_0 = T$, then $\partial u/\partial t = 0$ must hold.

We now select $k > 0$ so that $kT < \varepsilon/2$, i.e., let $k < \varepsilon/2T$; then the maximum of $v(x, t)$ for $t = 0$ or for $x = 0, x = l$ does not exceed the value $M + \varepsilon/2$, i.e.,

$$v(x, t) \leqq M + \frac{\varepsilon}{2} \qquad \text{for } t = 0 \text{ or } x = 0, x = l , \tag{3-1.39}$$

since for this argument the first summand of (3-1.38) is not larger than M, and the second is not larger than $\varepsilon/2$.

Now, $v(x, t)$ is a continuous function. Thus a point (x_1, t_1) exists at which it assumes its maximum. Then we have

$$v(x_1, t_1) \geqq v(x_0, t_0) = M + \varepsilon .$$

Therefore, $t_1 > 0$ and $0 < x_1 < l$, since for $t = 0$ or $x = 0$, $x = l$ the inequality (3-1.39) is valid. It follows that

$$v_{xx}(x_1, t_1) = u_{xx}(x_1, t_1) \leqq 0$$

and

$$v_t(x_1, t_1) = u_t(x_1, t_1) - k \geqq 0 \qquad \text{or} \qquad u_t(x_1, t_1) \geqq k > 0 .$$

By comparison of the signs on the right and the left sides in (3-1.35) at the point (x_1, t_1) we conclude that Eq. (3-1.35) at the point (x_1, t_1) cannot be satisfied, since the quantities on the right and left sides have different signs. Therefore, the first part of our proposition is proved. The statement for the minimum can be proved analogously, and it is sufficient to apply the first part to $u_1 = -u$.

6. The uniqueness theorem

We turn now to a series of consequences of the principle of the maximum. First, we prove the uniqueness theorem for the first boundary-value problem. If the functions $u_1(x, t)$ and $u_2(x, t)$, which are defined and continuous in a region $0 \leqq x \leqq l$, $0 \leqq t \leqq T$, and which satisfy the heat-conduction equation

$$u_t = a^2 u_{xx} + f(x, t) \qquad \text{for} \qquad 0 < x < l, t > 0 \tag{3-1.35'}$$

as well as the same initial and boundary conditions

$$\begin{aligned} u_1(x, 0) &= u_2(x, 0) = \varphi(x) \\ u_1(0, t) &= u_2(0, t) = \mu_1(t) \\ u_1(l, t) &= u_2(l, t) = \mu_2(t) , \end{aligned}$$

then necessarily[51]

$$u_1(x, t) \equiv u_2(x, t) .$$

For the proof of this theorem we consider the function

---

[51] Previously this theorem was refined and the continuity requirement at $t = 0$ was dropped.

$$v(x, t) = u_2(x, t) - u_1(x, t) .$$

Since $u_1(x, t)$ and $u_2(x, t)$ for

$$0 \leqq x \leqq l , \qquad 0 \leqq t \leqq T$$

are continuous, their difference $v(x, t)$ in the same region is continuous. Further, $v(x, t)$ as the difference of two solutions of the heat-conduction equation for $0 < x < l$, $t > 0$ is similarly a solution of the heat-conduction equation in that region. Consequently, the principle of the maximum can also be applied to this function, and the maximum and the minimum of $v(x, t)$ for $t = 0$ or $x = 0$ or $x = l$ is assumed. According to the hypothesis we obtain

$$v(x, 0) = 0 , \qquad v(0, t) = 0 , \qquad v(l, t) = 0 .$$

Therefore, also

$$v(x, t) \equiv 0 ,$$

i.e.,

$$u_1(x, t) \equiv u_2(x, t) ,$$

from which the uniqueness of the solution of the first boundary-value problem follows.

We shall now prove a series of direct conclusions from the principle of the maximum. In the following discussion we shall refer to "the solution of the heat-conduction equation," instead of enumerating the properties of the function in detail which also satisfy the initial and boundary conditions.

1. If two solutions $u_1(x, t)$ and $u_2(x, t)$ of the heat-conduction equation satisfy the conditions

$$u_1(x, 0) \leqq u_2(x, 0) , \qquad u_1(0, t) \leqq u_2(0, t) , \qquad u_1(l, t) \leqq u_2(l, t) ,$$

then

$$u_1(x, t) \leqq u_2(x, t)$$

for all $0 \leqq x \leqq l, 0 \leqq t \leqq T$.

The difference $v(x, t) = u_2(x, t) - u_1(x, t)$ satisfies the conditions on which the principal of the maximum is based; also

$$v(x, 0) \geqq 0 \qquad v(0, t) \geqq 0 \qquad v(l, t) \geqq 0 .$$

Therefore

$$v(x, t) \geqq 0 \qquad \text{for} \qquad 0 < x < l,\ 0 < t \leqq T ,$$

since $v(x, t)$ in the region

$$0 < x < l , \qquad 0 < t \leqq T$$

would otherwise have a negative value.

2. If three solutions

$$u(x, t) , \qquad \underline{u}(x, t) , \qquad \bar{u}(x, t)$$

of the heat-conduction equation satisfy the conditions

$$\underline{u}(x, t) \leqq u(x, t) \leqq \bar{u}(x, t) \qquad \text{for} \qquad t = 0 , \qquad x = 0 , \qquad x = l ,$$

then this inequality is fulfilled for all x in $0 \leqq x \leqq l$ and all t in $0 \leqq t \leqq T$.

This assertion represents an application of conclusion (1) to the functions

$$u(x, t) , \qquad \bar{u}(x, t) \qquad \text{and} \qquad u(x, t) , \qquad \underline{u}(x, t) .$$

3. If, for two solutions $u_1(x, t)$ and $u_2(x, t)$ of the heat conduction equation, the inequality

$$| u_1(x, t) - u_2(x, t) | \leqq \varepsilon , \qquad \text{for} \qquad t = 0 , \qquad x = 0 , \qquad x = l$$

is valid, then

$$| u_1(x, t) - u_2(x, t) | \leqq \varepsilon$$

for all x, t in

$$0 \leqq x \leqq l , \qquad 0 \leqq t \leqq T$$

is satisfied.

This assertion results from conclusion (2), when we apply the heat-conduction equation to the solutions

$$\underline{u}(x, t) = - \varepsilon$$
$$u(x, t) = u_1(x, t) - u_2(x, t)$$
$$\bar{u}(x, t) = \varepsilon .$$

The question regarding the continuous dependence of the solution of the first boundary-value problem on the initial and boundary conditions is answered completely by conclusion (3). To understand this, we consider a solution $u(x, t)$ which satisfies other initial and boundary conditions, instead of the solution of the heat-conduction equation which corresponds to the initial and boundary conditions

$$u(x, 0) = \varphi(x) , \qquad u(0, t) = \mu_1(t) , \qquad u(l, t) = \mu_2(t) .$$

Let these be given by functions $\varphi^*(x)$, $\mu_1^*(t)$ and $\mu_2^*(t)$ which differ by less than ε from the functions $\varphi(x)$, $\mu_1(t)$, and $\mu_2(t)$:

$$| \varphi(x) - \varphi^*(x) | \leqq \varepsilon , \qquad | \mu_1(t) - \mu_1^*(t) | \leqq \varepsilon , \qquad | \mu_2(t) - \mu_2^*(t) | \leqq \varepsilon .$$

However, the function $u_1(x, t)$ according to conclusion (3) differs by less than ε from the function $u(x, t)$:

$$| u(x, t) - u_1(x, t) | \leqq \varepsilon .$$

Here the principle of the physical determination of a problem arises directly.

We have investigated in detail the question of the uniqueness and the physical determination of a problem in the case of the first boundary-value problem for a bounded interval. The uniqueness theorem for the first boundary-value problem for a two- or three-dimensional bounded region can be proven by a verbatim repetition of these deliberations.

Similar questions arise in the investigation of other problems, an entire

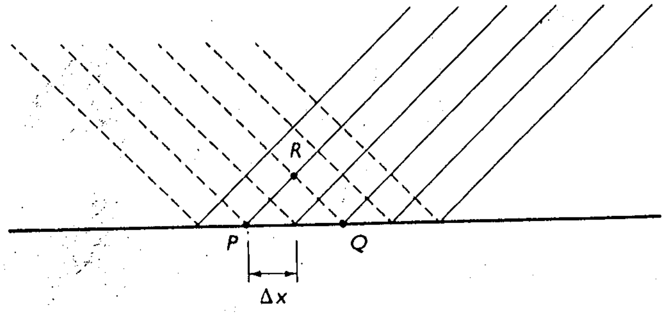


Figure 9.4

EXAMPLE 1 Solve

$$\frac{\partial^2 u}{\partial t^2} = 2\frac{\partial^2 u}{\partial x^2} - 4,$$

with initial conditions

$$u = 12x \quad \text{for} \quad 0 \le x \le 0.25,$$
$$u = 4 - 4x \quad \text{for} \quad 0.25 \le x \le 1.0,$$
$$\frac{\partial u}{\partial t} = 0 \quad \text{for} \quad 0 \le x \le 1.0;$$

boundary conditions are $u = 0$ at $x = 0$ and at $x = 1.0$.

Putting the equation into the standard form,

$$a\frac{\partial^2 u}{\partial x^2} + b\frac{\partial^2 u}{\partial x\,\partial t} + c\frac{\partial^2 u}{\partial t^2} + e = 0,$$

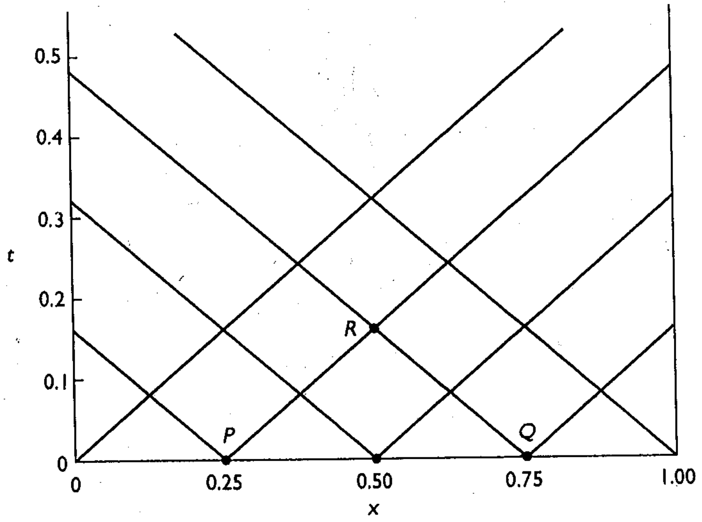


Figure 9.5

gives $a = -2$, $b = 0$, $c = 1$, and $e = 4$. (The equation is linear since a, b, c, and e are independent of u, u_x, and u_t.)

The slopes of the characteristics are the roots of

$$-2m^2 + 1 = 0,$$

$$m = \pm\frac{\sqrt{2}}{2},$$

so the characteristic curves are straight lines in the xt-plane, as shown in Fig. 9.5. Consider points P, Q, and R—(0.25, 0), (0.75, 0), and (0.5, 0.1768)—and solve Eq. (9.16), which is

$$am_{\text{av}}\,\Delta p + c\,\Delta q + e\,\Delta t = 0.$$

Along $P \to R$:
$$-2\left(\frac{\sqrt{2}}{2}\right)\Delta p + \Delta q + 4\Delta t = 0,$$
$$-\sqrt{2}(p_R - p_P) + (q_R - q_P) + 4(0.1768) = 0;$$

Along $Q \to R$:
$$-2\left(\frac{-\sqrt{2}}{2}\right)\Delta p + \Delta q + 4\Delta t = 0,$$
$$\sqrt{2}(p_R - p_Q) + (q_R - q_Q) + 4(0.1768) = 0.$$

Using

$$p_P = \left(\frac{\partial u}{\partial x}\right)_P = -4,^{*} \qquad q_P = \left(\frac{\partial u}{\partial t}\right)_P = 0, \qquad p_Q = \left(\frac{\partial u}{\partial x}\right)_Q = -4,$$
$$q_Q = \left(\frac{\partial u}{\partial t}\right)_Q = 0,$$

we find $p_R = -4$, $q_R = -\sqrt{2}/2$ by solving the equations $P \to R$ and $Q \to R$ simultaneously.

Now we evaluate u at point R through its change along $P \to R$:

$$\Delta u = p_{\text{av}}\,\Delta x + q_{\text{av}}\,\Delta t = -4(0.25) + \left(\frac{0 - \sqrt{2}/2}{2}\right)(0.1768)$$
$$= -1.0625;$$
$$u_R = 3 + (-1.0625) = 1.9375.$$

(If we compute through evaluating Δu along $Q \to R$, we get the same result.)

---

*The gradient has a discontinuity at $x = 0.25$. The value of $\partial u/\partial x$ for points to the right of P applies for the region PRQ.

Table 9.4

x	0	0.25	0.5	0.75	1.0
$u(t = 0)$	0.0	3.0	2.0	1.0	0.0
$u(t = 0.1768)$	0.0	0.9375	$\boxed{1.9375}$	0.9375	0.0
$u(t = 0.3535)$	0.0	−1.1875	−0.2500	0.8125	0.0
$u(t = 0.5303)$	0.0	−1.3125	−2.4375	−1.3125	0.0

For this simple problem, the finite-difference method is much simpler, and we expect it to give the same results. Following the procedure of Section 9.1* we compute with $\Delta x = 0.25$, $\Delta t = \Delta x/\sqrt{c} = 0.1768$, and obtain Table 9.4. The circled value agrees exactly with that calculated by the method of characteristics. ■

EXAMPLE 2 Solve

$$\frac{\partial^2 u}{\partial t^2} = (1 + 2x)\frac{\partial^2 u}{\partial x^2}$$

over (0, 1) with fixed boundaries and the initial conditions:

$$u(x, 0) = 0, \qquad \frac{\partial u}{\partial t}(x, 0) = x(1 - x).$$

For this problem, $a = -(1 + 2x)$, $b = 0$, $c = 1$, $e = 0$. Then $am^2 + bm + c = 0$ gives

$$m = \pm\sqrt{\frac{1}{(1 + 2x)}}.$$

The characteristic curves are found by solving the differential equations $dt/dx = \sqrt{1/(1 + 2x)}$ and $dt/dx = -\sqrt{1/(1 + 2x)}$. Integrating† from the initial point x_0 and t_0, we have

$$t = t_0 + \sqrt{1 + 2x} - \sqrt{1 + 2x_0} \quad \text{from } m_+,$$
$$t = t_0 - \sqrt{1 + 2x} + \sqrt{1 + 2x_0} \quad \text{from } m_-.$$

Figure 9.6 shows several of the characteristic curves. We select two points on the initial curve for $t = 0$, at $P = (0.25, 0)$ and $Q = (0.75, 0)$, whose characteristics intersect at point R. Solving for the intersection, we find $R = (0.4841, 0.1782)$.

---

*The algorithm is $u_i^{j+1} = (u_{i+1}^j + u_{i-1}^j) - u_i^{j-1} - 4(\Delta t)^2$ with $\Delta t = \Delta x/\sqrt{2}$. For the first time step, $u_i^1 = \frac{1}{2}(u_{i+1}^0 + u_{i-1}^0) - \frac{1}{2}(4)(\Delta t)^2$.

†In this example, the integration methods of calculus are easy to use. We could use a numerical method if they were not.

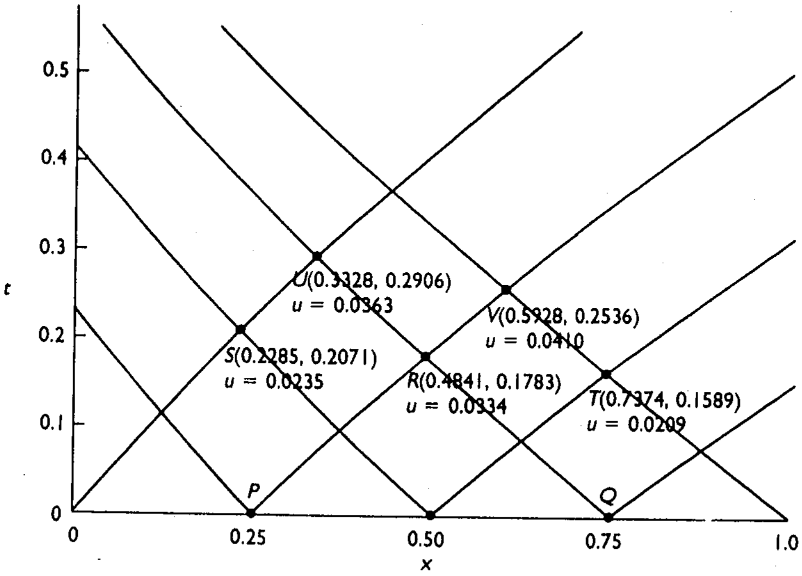


Figure 9.6

We now solve Eq. (9.16) to obtain $p = \partial u/\partial x$ and $q = \partial u/\partial t$ at R:

At point P: $x = 0.25,\quad t = 0,\quad u = 0,\quad p = \left(\dfrac{\partial u}{\partial x}\right)_P = 0,$

$$q = \left(\frac{\partial u}{\partial t}\right)_P = x - x^2 = 0.1875,$$

$$m = \sqrt{1/(1 + 2x)} = 0.8165,$$

$$a = -(1 + 2x) = -1.5,\quad b = 0,\quad c = 1,\quad e = 0.$$

At point Q: $x = 0.75,\quad t = 0,\quad u = 0,\quad p = 0,\quad q = x - x^2 = 0.1875,$

$$m = -\sqrt{1/(1 + 2x)} = -0.6325,$$

$$a = -2.5, b = 0,\quad c = 1,\quad e = 0.$$

At point R: $x = 0.4841,\quad t = 0.1783,$

$$m_+ = \sqrt{1/(1 + 2x)} = 0.7128,$$

$$m_- = -\sqrt{1/(1 + 2x)} = -0.7128,$$

$$a = -1.9682,\quad b = 0,\quad c = 1,\quad e = 0.$$

Equation (9.16) becomes, when we use average values for a and m,

$$P \to R: \qquad -1.7341(0.7646)(p_R - 0) + (1)(q_R - 0.1875) = 0;$$
$$Q \to R: \qquad -2.2341(-0.6726)(p_R - 0) + (1)(q_R - 0.1875) = 0.$$

Solving simultaneously, we get $p_R = 0$, $q_R = 0.1875$.

We calculate the change in u along the characteristics:

$$P \to R: \qquad \Delta u = 0(0.2341) + 0.1875(0.1783) = 0.0334,$$
$$Q \to R: \qquad \Delta u = 0(-0.2659) + 0.1875(0.1783) = 0.0334,$$
$$u_R = 0 + 0.0334 = 0.0334.$$

Figure 9.6 gives the results at several other intersections of characteristics. Students should verify these results to be sure they understand the method of characteristics. ∎

EXAMPLE 3 Solve the quasilinear equation, with conditions as shown, by numerical integration along the characteristics. (This might be a vibrating string with tension related to the displacement u and subject to an external lateral force.)

$$\frac{\partial^2 u}{\partial x^2} - u\frac{\partial^2 u}{\partial t^2} + (1 - x^2) = 0, \qquad u(x, 0) = x(1 - x), \qquad u_t(x, 0) = 0,$$
$$u(0, t) = 0, \qquad u(1, t) = 0. \tag{9.18}$$

We will advance the solution beyond the start from P, at $x = 0.2$, $t = 0$, and Q, at $x = 0.4$, $t = 0$, to one new point R. Comparing Eq. (9.18) to the standard form,

$$au_{xx} + bu_{xt} + cu_{tt} + e = 0,$$

we have $a = 1$, $b = 0$, $c = -u$, $e = 1 - x^2$. We first compute u, p, and q at points P and Q

$$u = x(1 - x)$$

(from the initial conditions), so

$$u_P = 0.2(1 - 0.2) = 0.16,$$
$$u_Q = 0.4(1 - 0.4) = 0.24.$$

Also,

$$p = \frac{\partial u}{\partial x} = 1 - 2x$$

(by differentiating the initial conditions), so

$$p_P = 1 - 2(0.2) = 0.6,$$
$$p_Q = 1 - 2(0.4) = 0.2;$$

and

$$q = \frac{\partial u}{\partial t} = 0$$

(from the initial conditions), so

$$q_P = 0,$$
$$q_Q = 0.$$

To locate point R, we need the slope m of the characteristic. Using $am^2 - bm + c = 0$, we get

$$m = \frac{b \pm \sqrt{b^2 - 4ac}}{2a},$$
$$m = \frac{\pm\sqrt{4u}}{2} = \pm\sqrt{u}.$$

Since m depends on the solution u, we will need to find point R through the predictor–corrector approach. In the first trial, use the initial values over the whole arc; that is, take $m_+ = +m_P$ and $m_- = -m_Q$:

$$m_+ = \sqrt{u_P} = \sqrt{0.16} = 0.4,$$
$$m_- = \sqrt{u_Q} = -\sqrt{0.24} = -0.490.$$

We now estimate the coordinates of R by solving simultaneously

$$t_R = m_+(x_R - x_P) = 0.4(x_R - 0.2),$$
$$t_R = m_-(x_R - x_Q) = -0.490(x_R - 0.4).$$

These give

$$x_R = 0.310, \qquad t_R = 0.044.$$

We write Eq. (9.16) along each characteristic, again using the initial values of m, since m at R is still unknown:

$$am\,\Delta p + c\,\Delta q + e\,\Delta t = 0,$$
$$(1)(0.4)(p_R - 0.6) + (-0.16)(q_R - 0) + \left(1 - \frac{0.04 + 0.096}{2}\right)(0.044) = 0,$$
$$(1)(-0.490(p_R - 0.2) + (-0.24)(q_R - 0) + \left(1 - \frac{0.16 + 0.096}{2}\right)(0.044) = 0.$$

In these equations we used the arithmetic average of x^2 in the last terms. Solving simultaneously, we get

$$p_R = 0.399, \qquad q_R = -0.246.$$

As a first approximation for u at R, then,

$$\Delta u = p\,\Delta x + q\,\Delta t,$$

$$u_R - 0.16 = \frac{0.6 + 0.399}{2}(0.310 - 0.2) + \frac{0 - 0.246}{2}(0.044 - 0),$$

$$u_R = 0.2095.$$

The last computation was along PR, using average values of p and q. We could have alternatively proceeded along QR. If this is done,

$$u_R - 0.24 = \frac{0.2 + 0.399}{2}(0.310 - 0.4) + \frac{0 - 0.246}{2}(0.044 - 0),$$

$$u_R = 0.2076.$$

The two values should be close to each other. Let us use the average value, 0.2086, as our initial estimate of u_R. We now repeat the work. In getting the coordinates of P, we now use average values of the slopes,

$$t_R = \frac{0.4 + \sqrt{0.2086}}{2}(x_R - 0.2),$$

$$t_R = \frac{-0.490 - \sqrt{0.2086}}{2}(x_R - 0.4),$$

$$x_R = 0.305, \qquad t_R = 0.045;$$

$$(1)\left(\frac{0.4 + \sqrt{0.2086}}{2}\right)(p_R - 0.6) - \left(\frac{0.16 + 0.2086}{2}\right)(q_R - 0) + \left(1 - \frac{0.04 + 0.0930}{2}\right)(0.045) = 0,$$

$$(1)\left(\frac{-0.490 - \sqrt{0.2086}}{2}\right)(p_R - 0.2) - \left(\frac{0.24 + 0.2086}{2}\right)(q_R - 0) + \left(1 - \frac{0.16 + 0.0930}{2}\right)(0.045) = 0,$$

$$p_R = 0.398, \qquad q_R = -0.242;$$

$$u_R = 0.16 + \frac{0.6 + 0.398}{2}(0.305 - 0.2) + \frac{0 - 0.242}{2}(0.045 - 0),$$

$$u_R = 0.2071 \quad \text{(along } PR\text{)};$$

$$u_R = 0.24 + \frac{0.2 + 0.398}{2}(0.305 - 0.4) + \frac{0 - 0.242}{2}(0.045 - 0),$$

$$u_R = 0.2063 \quad \text{(along } QR\text{)}.$$

The average value is 0.2067.

Another round of calculations gives $u_R = 0.2066$, which checks the previous value sufficiently. This method is, of course, very tedious by hand. ■

where δ_x^2 is the usual central difference. For j odd, Eqn. (2-265) becomes

$$U_{i,j+2} = U_{i,j} + 2r\,\delta_x^2 U_{i,j+1}\,.$$

Thus, Gordon's method is made up of a series of familiar calculations.

Another series of asymmetric approximations for the diffusion equation, $u_t = u_{xx}$, was introduced by Saul'yev [85] followed by variations due to Larkin [86], Barakat and Clark [87], and Liu [88], and related *group explicit* methods by Abdullah [89], Evans and Abdullah [90], and Evans [91]. In each case, the result is an approximation or pair of approximations that are explicit and are unconditionally stable.

The basic ideas are easy to describe. Saul'yev replaces $u_x|_{i-1/2,j}$ in

$$u_{xx}|_{i,j} = \frac{u_x|_{i+1/2,j} - u_x|_{i-1/2,j}}{h}$$

by $u_x|_{i-1/2,j+1}$, and uses the central difference approximations

$$u_x|_{i+1/2,j} = \frac{U_{i+1,j} - U_{i,j}}{h}\,,$$

$$u_x|_{i-1/2,j+1} = \frac{U_{i,j+1} - U_{i-1,j+1}}{h}$$

for u_x and a forward difference for u_t to obtain the algorithm (Saul'yev A), which has been solved for $U_{i,j+1}$:

$$(1+r)U_{i,j+1} = U_{i,j} + r(U_{i-1,j+1} - U_{i,j} + U_{i+1,j})\,. \qquad (2\text{-}266)$$

The asymmetric computational molecule for Eqn. (2-266) is shown in Fig. 2-16.

An analogous form (Saul'yev B) (Problem 2-58) is

$$(1+r)U_{i,j+1} = U_{i,j} + r(U_{i+1,j+1} - U_{i,j} + U_{i-1,j})\,, \qquad (2\text{-}267)$$

with the computational molecule shown in Fig. 2-17.

It's important to note that Saul'yev A is explicit if the evaluation begins at the left-hand boundary ($x = 0$) and moves to the right, so that only the single value $U_{i,j+1}$ is unknown. A similar explicit nature holds for Saul'yev B if the calculation proceeds to the left from the right-hand boundary (say, $x = 1$). It is interesting to note that if $r = 1$ then Saul'yev A simplifies to

$$U_{i,j+1} = \tfrac{1}{2}(U_{i-1,j+1} + U_{i+1,j})\,,$$

and Saul'yev B to

$$U_{i,j+1} = \tfrac{1}{2}(U_{i+1,j+1} + U_{i-1,j})$$

—i.e., the value of $U_{i,j}$ is not needed in the computation!

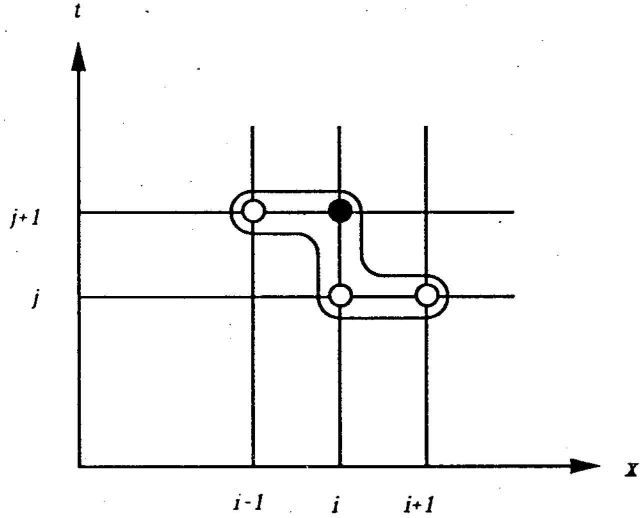


Figure 2-16 Molecule for Eqn. (2-266)

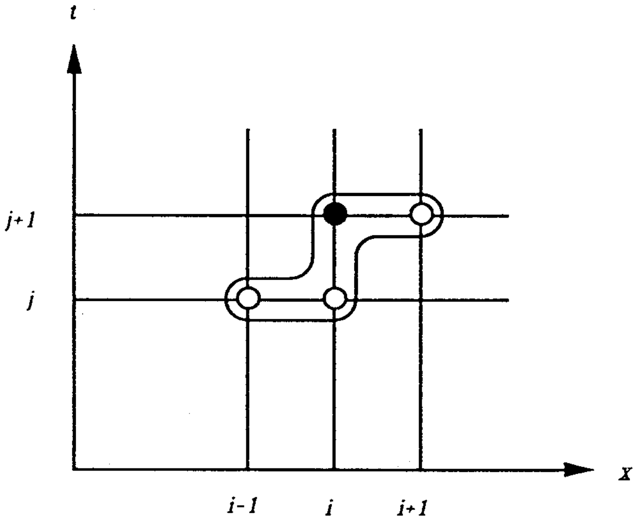


Figure 2-17 Molecule for Eqn. (2-267)

The options open to using the Saul'yev schemes are:

(i) Use Eqn. (2-266) only and proceed line by line in the t direction, but *always* from the left boundary to the right on a line.
(ii) Use Eqn. (2-267) only and proceed line by line in the t direction, but *always* from the right boundary to the left on a line.

As the arrows indicate, the wave equation propagates disturbances in opposite directions, with reflections occurring at fixed ends, with a reversal of sign on reflection. Stability is demonstrated because the original error does not grow in size.

9.4 METHOD OF CHARACTERISTICS

The properties of the solution to the wave equation are further elucidated by considering the "characteristic curves" of the equation. This will also permit us to extend our numerical method to more general hyperbolic equations.

Consider the second-order partial-differential equation in two variables x and t:

$$\boxed{au_{xx} + bu_{xt} + cu_{tt} + e = 0. \qquad (9.14)}$$

Here we have used the subscript notation to represent partial derivatives. The coefficients a, b, c, and e may be functions of x, t, u_x, u_t, and u, so the equation is very general.* We take $u_{xt} = u_{tx}$. To facilitate manipulations, let

$$p = \frac{\partial u}{\partial x} = u_x, \qquad q = \frac{\partial u}{\partial t} = u_t.$$

Write out the differentials of p and q:

$$dp = \frac{\partial p}{\partial x}\,dx + \frac{\partial p}{\partial t}\,dt = u_{xx}\,dx + u_{xt}\,dt,$$

$$dq = \frac{\partial q}{\partial x}\,dx + \frac{\partial q}{\partial t}\,dt = u_{xt}\,dx + u_{tt}\,dt.$$

Solving these last equations for u_{xx} and u_{tt}, respectively, we have

$$u_{xx} = \frac{dp - u_{xt}\,dt}{dx} = \frac{dp}{dx} - u_{xt}\frac{dt}{dx},$$

$$u_{tt} = \frac{dq - u_{xt}\,dx}{dt} = \frac{dq}{dt} - u_{xt}\frac{dx}{dt}.$$

Substituting in Eq. (9.14) and rearranging, we obtain

$$-au_{xt}\frac{dt}{dx} + bu_{xt} - cu_{xt}\frac{dx}{dt} + a\frac{dp}{dx} + c\frac{dq}{dt} + e = 0.$$

---

*When the coefficients are independent of the function u or its derivatives, it is linear. If they are functions of u, u_x, or u_t (but not u_{xx} or u_{tt}), it is called quasilinear.

Now multiplying by $-dt/dx$, we get

$$\boxed{u_{xt}\left[a\left(\frac{dt}{dx}\right)^2 - b\left(\frac{dt}{dx}\right) + c\right] - \left[a\frac{dp}{dx}\frac{dt}{dx} + c\frac{dq}{dx} + e\frac{dt}{dx}\right] = 0.}$$

Suppose, in the xt-plane, we define curves such that the first bracketed expression is zero. On such curves, the original differential equation is equivalent to setting the second bracketed expression equal to zero. That is, if

$$\boxed{am^2 - bm + c = 0, \qquad (9.15)}$$

where $m = dt/dx$, then the solution to the original equation (Eq. (9.14)) can be found by solving

$$\boxed{am\,dp + c\,dq + e\,dt = 0. \qquad (9.16)}$$

We have elected to write Eq. (9.16) in the form of differentials. It will be seen that we reduce the original problem, which is a second-order partial-differential equation, to solving a pair of first-order equations of the form of Eq. (9.16).

The curves whose slope m is given by Eq. (9.15) are called the *characteristics* of the differential equation. Since the equation is a quadratic, it may have one, two, or no real solutions, depending on the value of $b^2 - 4ac$. The value of this discriminant is the usual basis for classifying partial-differential equations. If

$$b^2 - 4ac < 0,$$

the equation is called *elliptic*, and there are no (real) characteristics. If $b^2 - 4ac = 0$, there is a single characteristic at any point, and the equation is termed *parabolic*. When $b^2 - 4ac > 0$, at every point there will be a pair of characteristic curves whose slopes are given by the two distinct, real roots of Eq. (9.15). Such equations are called *hyperbolic*, and our present discussion considers only this type.

Along the characteristics, the solution has special and desirable properties. For example, discontinuities in the initial conditions are propagated along them. On the characteristic curves, the numerical solution can be developed in the general case also.

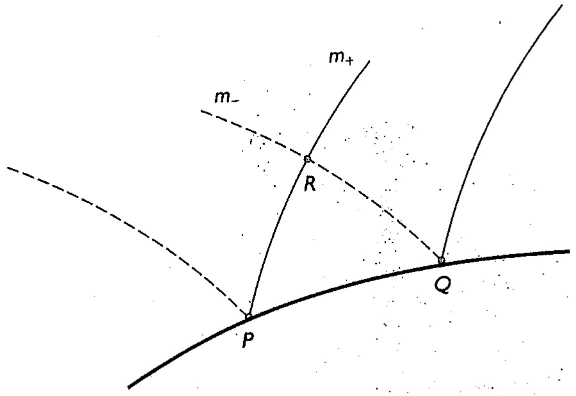


Figure 9.3

We will now outline a method of solving equations of the form of Eq. (9.14) by numerical integration along the characteristics.* We visualize the initial conditions as specifying the function u on some curve in the tx-plane,† as well as its normal derivative. Consider two points, P and Q on this initial curve (Fig. 9.3). When Eq. (9.14) is hyperbolic, there are two characteristic curves through each point. The rightmost curve through P intersects the leftmost curve through Q, and these curves are such that their slopes are given by the appropriate roots of Eq. (9.15). Call m_+ the values of the slope on curve PR, and m_- the values on curve QR.

Since these curves are characteristics, the solution to the problem can be found by solving Eq. (9.16) along them.

Our procedure will be to first find point R (perhaps only as a first approximation if a, b, or c involve the unknown function u). This is done by solving the equation

$$\boxed{\Delta t = \left(\frac{dt}{dx}\right)_{\text{av}} \Delta x = m_{\text{av}}\,\Delta x, \qquad \textbf{(9.17)}}$$

applied over the arcs PR and QR simultaneously. When dt/dx is a function of x and/or t, it may be possible to integrate Eq. (9.17) analytically. When dt/dx varies with u, we will use a procedure resembling the Euler predictor–corrector method, by predicting with m_{av} taken as equal to m_+ at P or m_- at Q to start the solution. We correct by using the arithmetic average of m at the endpoints of each arc as soon as the value of m at R can be evaluated.

We then integrate Eq. (9.16) in the form

$$\boxed{a_{\text{av}}\, m_{\text{av}}\, \Delta p + c_{\text{av}}\, \Delta q + e_{\text{av}}\, \Delta t = 0,}$$

---

*We discuss only the solution of hyperbolic differential equations by the method of characteristics, but the technique can also be applied to parabolic equations.

†This curve must not itself be one of the characteristics, or advancing the solution is impossible.

starting first from P and then from Q, using the appropriate values of m for each. This will estimate p and q at point R. Finally we evaluate the function u at R from

$$du = \frac{\partial u}{\partial x}\,dx + \frac{\partial u}{\partial t}\,dt,$$

used in the form

$$\boxed{\Delta u = p_{\text{av}}\,\Delta x + q_{\text{av}}\,\Delta t.}$$

The equation for Δu can be applied to either the change along

$$P \rightarrow R \qquad \text{or} \qquad Q \rightarrow R;$$

when these do not give the same result, we will *average* the two values. It may be necessary to iterate this procedure to get improved values at R. Average values are used for all varying quantities. The calculations are repeated for a second point S that is the intersection of chararacteristics through Q and another initial point W, and then continued in a like manner to evaluate u throughout the region in the xt-plane as desired. We illustrate with three examples, first with dt/dx equal to a constant, then with dt/dx varying with x, and finally a more complex example with dt/dx varying with u.

For the simple wave equation,

$$u_{tt} = c^2 u_{xx},$$

the slopes of the characteristics are

$$m = \pm\frac{1}{c},$$

and the characteristics are the lines

$$t = \pm\frac{1}{c}(x - x_i).$$

Consider the curves from points P and Q, taken on the line for $t = 0$ (Fig. 9.4). The network of points used in the explicit finite-difference method of Section 9.1 are seen to be the intersections of characteristics through pairs of points spaced $2\Delta x$ apart. The finite-difference method, with

$$\frac{Tg(\Delta x)^2}{w(\Delta t)^2} = 1,$$

will be found to be the equivalent of integration along the characteristics, lending further support to the likelihood that it will give exact answers.

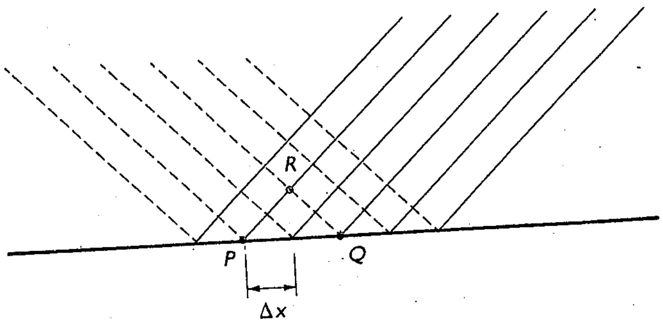


Figure 9.4

EXAMPLE 1 Solve

$$\frac{\partial^2 u}{\partial t^2} = 2\frac{\partial^2 u}{\partial x^2} - 4,$$

with initial conditions

$$u = 12x \quad \text{for} \quad 0 \le x \le 0.25,$$
$$u = 4 - 4x \quad \text{for} \quad 0.25 \le x \le 1.0,$$
$$\frac{\partial u}{\partial t} = 0 \quad \text{for} \quad 0 \le x \le 1.0;$$

boundary conditions are $u = 0$ at $x = 0$ and at $x = 1.0$.

Putting the equation into the standard form,

$$a\frac{\partial^2 u}{\partial x^2} + b\frac{\partial^2 u}{\partial x\,\partial t} + c\frac{\partial^2 u}{\partial t^2} + e = 0,$$

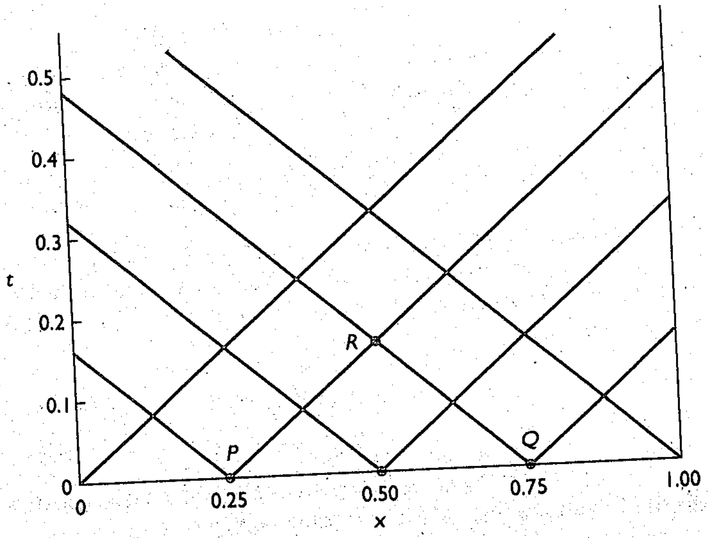


Figure 9.5

gives $a = -2$, $b = 0$, $c = 1$, and $e = 4$. (The equation is linear since a, b, c, and e are independent of u, u_x, and u_t.)

The slopes of the characteristics are the roots of

$$-2m^2 + 1 = 0,$$
$$m = \pm\frac{\sqrt{2}}{2},$$

so the characteristic curves are straight lines in the xt-plane, as shown in Fig. 9.5. Consider points P, Q, and R—(0.25, 0), (0.75, 0), and (0.5, 0.1768)—and solve Eq. (9.16), which is

$$am_{\text{av}}\,\Delta p + c\,\Delta q + e\,\Delta t = 0.$$

$$\boxed{\begin{array}{ll} \text{Along } P \to R: & -2\left(\frac{\sqrt{2}}{2}\right)\Delta p + \Delta q + 4\Delta t = 0, \\ & -\sqrt{2}(p_R - p_P) + (q_R - q_P) + 4(0.1768) = 0; \\ \text{Along } Q \to R: & -2\left(\frac{-\sqrt{2}}{2}\right)\Delta p + \Delta q + 4\Delta t = 0, \\ & \sqrt{2}(p_R - p_Q) + (q_R - q_Q) + 4(0.1768) = 0. \end{array}}$$

Using

$$p_P = \left(\frac{\partial u}{\partial x}\right)_P = -4,^* \qquad q_P = \left(\frac{\partial u}{\partial t}\right)_P = 0, \qquad p_Q = \left(\frac{\partial u}{\partial x}\right)_Q = -4,$$
$$q_Q = \left(\frac{\partial u}{\partial t}\right)_Q = 0,$$

we find $p_R = -4$, $q_R = -\sqrt{2}/2$ by solving the equations $P \to R$ and $Q \to R$ simultaneously.

Now we evaluate u at point R through its change along $P \to R$:

$$\Delta u = p_{\text{av}}\,\Delta x + q_{\text{av}}\,\Delta t = -4(0.25) + \left(\frac{0 - \sqrt{2}/2}{2}\right)(0.1768)$$
$$= -1.0625;$$
$$u_R = 3 + (-1.0625) = 1.9375.$$

(If we compute through evaluating Δu along $Q \to R$, we get the same result.)

---

*The gradient has a discontinuity at $x = 0.25$. The value of $\partial u/\partial x$ for points to the right of P applies for the region PRQ.

Table 9.4

x	0	0.25	0.5	0.75	1.0
$u(t = 0)$	0.0	3.0	2.0	1.0	0.0
$u(t = 0.1768)$	0.0	0.9375	1.9375	0.9375	0.0
$u(t = 0.3535)$	0.0	−1.1875	−0.2500	0.8125	0.0
$u(t = 0.5303)$	0.0	−1.3125	−2.4375	−1.3125	0.0

For this simple problem, the finite-difference method is much simpler, and we expect it to give the same results. Following the procedure of Section 9.1[*] we compute with $\Delta x = 0.25$, $\Delta t = \Delta x/\sqrt{c} = 0.1768$, and obtain Table 9.4. The circled value agrees exactly with that calculated by the method of characteristics. ▪

EXAMPLE 2 Solve

$$\frac{\partial^2 u}{\partial t^2} = (1 + 2x)\frac{\partial^2 u}{\partial x^2}$$

over (0, 1) with fixed boundaries and the initial conditions:

$$u(x, 0) = 0, \qquad \frac{\partial u}{\partial t}(x, 0) = x(1 - x).$$

For this problem, $a = -(1 + 2x)$, $b = 0$, $c = 1$, $e = 0$. Then $am^2 + bm + c = 0$ gives

$$m = \pm\sqrt{\frac{1}{(1 + 2x)}}.$$

The characteristic curves are found by solving the differential equations $dt/dx = \sqrt{1/(1 + 2x)}$ and $dt/dx = -\sqrt{1/(1 + 2x)}$. Integrating[†] from the initial point x_0 and t_0, we have

$$t = t_0 + \sqrt{1 + 2x} - \sqrt{1 + 2x_0} \quad \text{from } m_+,$$
$$t = t_0 - \sqrt{1 + 2x} + \sqrt{1 + 2x_0} \quad \text{from } m_-.$$

Figure 9.6 shows several of the characteristic curves. We select two points on the initial curve for $t = 0$, at $P = (0.25, 0)$ and $Q = (0.75, 0)$, whose characteristics intersect at point R. Solving for the intersection, we find $R = (0.4841, 0.1782)$.

---

[*] The algorithm is $u_i^{j+1} = (u_{i+1}^j + u_{i-1}^j) - u_i^{j-1} - 4(\Delta t)^2$ with $\Delta t = \Delta x/\sqrt{2}$. For the first time step, $u_i^1 = \frac{1}{2}(u_{i+1}^0 + u_{i-1}^0) - \frac{1}{2}(4)(\Delta t)^2$.

[†] In this example, the integration methods of calculus are easy to use. We could use a numerical method if they were not.

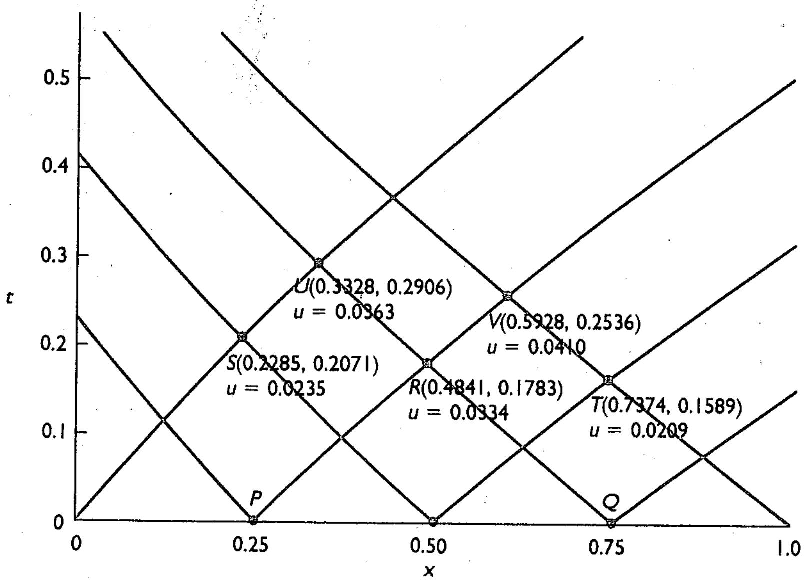


Figure 9.6

We now solve Eq. (9.16) to obtain $p = \partial u/\partial x$ and $q = \partial u/\partial t$ at R:

At point P: $x = 0.25, \quad t = 0, \quad u = 0, \quad p = \left(\dfrac{\partial u}{\partial x}\right)_P = 0,$

$$q = \left(\frac{\partial u}{\partial t}\right)_P = x - x^2 = 0.1875,$$

$$m = \sqrt{1/(1 + 2x)} = 0.8165,$$

$$a = -(1 + 2x) = -1.5, \quad b = 0, \quad c = 1, \quad e = 0.$$

At point Q: $x = 0.75, \quad t = 0, \quad u = 0, \quad p = 0, \quad q = x - x^2 = 0.1875,$

$$m = -\sqrt{1/(1 + 2x)} = -0.6325,$$

$$a = -2.5, b = 0, \quad c = 1, \quad e = 0.$$

At point R: $x = 0.4841, \quad t = 0.1783,$

$$m_+ = \sqrt{1/(1 + 2x)} = 0.7128,$$

$$m_- = -\sqrt{1/(1 + 2x)} = -0.7128,$$

$$a = -1.9682, \quad b = 0, \quad c = 1, \quad e = 0.$$

Equation (9.16) becomes, when we use average values for a and m,

$$P \to R: \qquad -1.7341(0.7646)(p_R - 0) + (1)(q_R - 0.1875) = 0;$$
$$Q \to R: \qquad -2.2341(-0.6726)(p_R - 0) + (1)(q_R - 0.1875) = 0.$$

Solving simultaneously, we get $p_R = 0$, $q_R = 0.1875$.

We calculate the change in u along the characteristics:

$$P \to R: \qquad \Delta u = 0(0.2341) + 0.1875(0.1783) = 0.0334,$$
$$Q \to R: \qquad \Delta u = 0(-0.2659) + 0.1875(0.1783) = 0.0334,$$
$$u_R = 0 + 0.0334 = 0.0334.$$

Figure 9.6 gives the results at several other intersections of characteristics. Students should verify these results to be sure they understand the method of characteristics. ◼

EXAMPLE 3 Solve the quasilinear equation, with conditions as shown, by numerical integration along the characteristics. (This might be a vibrating string with tension related to the displacement u and subject to an external lateral force.)

$$\frac{\partial^2 u}{\partial x^2} - u\frac{\partial^2 u}{\partial t^2} + (1 - x^2) = 0, \qquad u(x, 0) = x(1 - x), \qquad u_t(x, 0) = 0,$$
$$u(0, t) = 0, \qquad u(1, t) = 0. \qquad \textbf{(9.18)}$$

We will advance the solution beyond the start from P, at $x = 0.2$, $t = 0$, and Q, at $x = 0.4$, $t = 0$, to one new point R. Comparing Eq. (9.18) to the standard form,

$$au_{xx} + bu_{xt} + cu_{tt} + e = 0,$$

we have $a = 1$, $b = 0$, $c = -u$, $e = 1 - x^2$. We first compute u, p, and q at points P and Q

$$u = x(1 - x)$$

(from the initial conditions), so

$$u_P = 0.2(1 - 0.2) = 0.16,$$
$$u_Q = 0.4(1 - 0.4) = 0.24.$$

Also,

$$p = \frac{\partial u}{\partial x} = 1 - 2x$$

(by differentiating the initial conditions), so

$$p_P = 1 - 2(0.2) = 0.6,$$
$$p_Q = 1 - 2(0.4) = 0.2;$$

and

$$q = \frac{\partial u}{\partial t} = 0$$

(from the initial conditions), so

$$q_P = 0,$$
$$q_Q = 0.$$

To locate point R, we need the slope m of the characteristic. Using $am^2 - bm + c = 0$, we get

$$m = \frac{b \pm \sqrt{b^2 - 4ac}}{2a},$$
$$m = \frac{\pm\sqrt{4u}}{2} = \pm\sqrt{u}.$$

Since m depends on the solution u, we will need to find point R through the predictor–corrector approach. In the first trial, use the initial values over the whole arc; that is, take $m_+ = +m_P$ and $m_- = -m_Q$:

$$m_+ = \sqrt{u_P} = \sqrt{0.16} = 0.4,$$
$$m_- = \sqrt{u_Q} = -\sqrt{0.24} = -0.490.$$

We now estimate the coordinates of R by solving simultaneously

$$t_R = m_+(x_R - x_P) = 0.4(x_R - 0.2),$$
$$t_R = m_-(x_R - x_Q) = -0.490(x_R - 0.4).$$

These give

$$x_R = 0.310, \qquad t_R = 0.044.$$

We write Eq. (9.16) along each characteristic, again using the initial values of m, since m at R is still unknown:

$$am\,\Delta p + c\,\Delta q + e\,\Delta t = 0,$$
$$(1)(0.4)(p_R - 0.6) + (-0.16)(q_R - 0) + \left(1 - \frac{0.04 + 0.096}{2}\right)(0.044) = 0,$$
$$(1)(-0.490(p_R - 0.2) + (-0.24)(q_R - 0) + \left(1 - \frac{0.16 + 0.096}{2}\right)(0.044) = 0.$$

In these equations we used the arithmetic average of x^2 in the last terms. Solving simultaneously, we get

$$p_R = 0.399, \qquad q_R = -0.246.$$

As a first approximation for u at R, then,

$$\Delta u = p\,\Delta x + q\,\Delta t,$$

$$u_R - 0.16 = \frac{0.6 + 0.399}{2}(0.310 - 0.2) + \frac{0 - 0.246}{2}(0.044 - 0),$$

$$u_R = 0.2095.$$

The last computation was along PR, using average values of p and q. We could have alternatively proceeded along QR. If this is done,

$$u_R - 0.24 = \frac{0.2 + 0.399}{2}(0.310 - 0.4) + \frac{0 - 0.246}{2}(0.044 - 0),$$

$$u_R = 0.2076.$$

The two values should be close to each other. Let us use the average value, 0.2086, as our initial estimate of u_R. We now repeat the work. In getting the coordinates of R, we now use average values of the slopes,

$$t_R = \frac{0.4 + \sqrt{0.2086}}{2}(x_R - 0.2),$$

$$t_R = \frac{-0.490 - \sqrt{0.2086}}{2}(x_R - 0.4),$$

$$x_R = 0.305, \qquad t_R = 0.045;$$

$$(1)\left(\frac{0.4 + \sqrt{0.2086}}{2}\right)(p_R - 0.6) - \left(\frac{0.16 + 0.2086}{2}\right)(q_R - 0) + \left(1 - \frac{0.04 + 0.0930}{2}\right)(0.045) = 0,$$

$$(1)\left(\frac{-0.490 - \sqrt{0.2086}}{2}\right)(p_R - 0.2) - \left(\frac{0.24 + 0.2086}{2}\right)(q_R - 0) + \left(1 - \frac{0.16 + 0.0930}{2}\right)(0.045) = 0,$$

$$p_R = 0.398, \qquad q_R = -0.242;$$

$$u_R = 0.16 + \frac{0.6 + 0.398}{2}(0.305 - 0.2) + \frac{0 - 0.242}{2}(0.045 - 0),$$

$$u_R = 0.2071 \quad \text{(along } PR\text{)};$$

$$u_R = 0.24 + \frac{0.2 + 0.398}{2}(0.305 - 0.4) + \frac{0 - 0.242}{2}(0.045 - 0),$$

$$u_R = 0.2063 \quad \text{(along } QR\text{)}.$$

The average value is 0.2067.

Another round of calculations gives $u_R = 0.2066$, which checks the previous value sufficiently. This method is, of course, very tedious by hand. ■

(iii) Alternate from line to line by first using Saul'yev A and then B, or the reverse. This is related to *alternating direction methods* to be discussed later.

(iv) Use Saul'yev A and Saul'yev B on the same line and average the results for the final answer (A first, and then B). This is equivalent to introducing the dummy variables $P_{i,j}$ and $Q_{i,j}$ such that

$$(1+r)P_{i,j+1} = U_{i,j} + r(P_{i-1,j+1} - U_{i,j} + U_{i+1,j}), \quad (2\text{-}268a)$$

$$(1+r)Q_{i,j+1} = U_{i,j} + r(Q_{i+1,j+1} - U_{i,j} + U_{i-1,j}). \quad (2\text{-}268b)$$

and

$$U_{i,j+1} = \tfrac{1}{2}(P_{i,j+1} + Q_{i,j+1}). \quad (2\text{-}269)$$

This averaging method has some computational advantage because of the possibility of truncation error cancellation.

Levy

6 $y^2 = x^2 + 1$ soln to $y' = \frac{x}{y}$ $y(0) = 1$

$h = .1$ $y_1(1) = 1.3855$ actual $y(1) = \sqrt{2} = 1.414$ 2.03% error

$h = .2$ $\tilde{y}_1(1) = 1.3550$ 4.2% error

$y_{1_{ACT}} = y_1 + 2(\tilde{y}_1 - y_1) = 1.4160$ error -0.13%

11 $y' = x + y + xy$ $y(x=.1) = 1.11590$ RK using $\Delta x = 0.1$ 4 Func EV

5 $y(x=.1) = 1.1140$ euler $\Delta x = 0.01$ 10 Fun EV

7 $y(x=.1) = 1.11587$ m.E $\Delta x = 0.025$ 8 Func EV

#23 $y' = y - x^2$ $y(0) = 1$ actual is $y_H = De^x$

$y_P = Ax^2 + Bx + C$

$y' - y + x^2 = 2Ax + B - Ax^2 - Bx - C + x^2 = 0$

$A = 1$ $B = 2$ $C = 2$

$y_T = De^x + x^2 + 2x + 2$

$y_H(0) = 0$ $D + 2 = 1$ $D = -1$

$y_T = -e^x + x^2 + 2x + 2$

$y_p(0.8) = 2.0146$ $y_c = 2.0145$

$y_P(1.0) = 2.2819$ $y_c = 2.2818$

$y_p(1.2) = 2.5202$ $y_c = 2.5200$ error $4 \times 10^{-3}\%$

actual is 2.5199

54 $y''' + ty'' - ty' - 2y = t$ $y(0) = y''(0) = 0$ $y'(0) = 1$

$y' = p$

$y'' = p' = q$

$y''' = p'' = q' = -tq + tp + 2y + t$

$y_{.2} = .20027$ $y_{.4} = .40422$ $y_{.6} = .62099$

$y' = P_{.2} = 1.00531$ $y' = p_{.4} = 1.04180$ $y' = p_{.6} = 1.13801$

$y'' = q_{.2} = .07926$ $y'' = q_{.4} = .30950$ $y'' = q_{.6} = .67328$

Summer 1997

- Part I of Final Examination due 18 July 1997 at noon in EAS 3442

- You may use only your notes and your book

- Please sign : I will not give or receive any help on this examination. I will fail this examination if I do

NAME

Signature

40 pts Problem 1 : (a) Do problem 105 Chapter 4 pg 391 using simpson's rule (here $\Delta x = \Delta y = .2$

(b) Use linear interpolation to find values for $\Delta x = \Delta y = 0.1$ then find $\iint U\,dx\,dy$

(c) use Romberg to get a better results

30 pts Problem 2: Do problem 80 Chapter 5 pg 463

Use 20 steps between $t = 0$ and $t = 0.2$ seconds using a 4th order RK method

30 pts Problem 3: Do problem 76 Chapter 6 pg 539 by replacing the second derivative by finite differences.

(a) let 4 inch dimension be vertical

(b) let 4 inch dimension be horizontal

Exercise Set 10.4 (page 557)

1 **a)** With $\mathbf{x}^{(0)} = (0,0)^t$, $\mathbf{x}^{(5)} = (0.9960763, 0.9973286)^t$

b) With $\mathbf{x}^{(0)} = (0,10)^t$, $\mathbf{x}^{(17)} = (1.537896, 10.96678)^t$

c) With $\mathbf{x}^{(0)} = (1,1)^t$, $\mathbf{x}^{(2)} = (0.5014280, 0.8698339)^t$

d) With $\mathbf{x}^{(0)} = (-0.1, 0.1)^t$, $\mathbf{x}^{(3)} = (-0.2795486, 0.4641872)^t$

3 **a)** With $\mathbf{x}^{(0)} = (0,0,0)^t$, $\mathbf{x}^{(4)} = (1.073144, 1.073144, 0.8044867)^t$

b) With $\mathbf{x}^{(0)} = (0,0,0)^t$, $\mathbf{x}^{(2)} = (-0.02005700, 0.09019657, 0.9946810)^t$

c) With $\mathbf{x}^{(0)} = (0,0,0)^t$, $\mathbf{x}^{(18)} = (1.038654, 1.079038, 0.9289789)^t$

d) With $\mathbf{x}^{(0)} = (0,0,0)^t$, $\mathbf{x}^{(7)} = (0.4730910, 0.01915232, -0.5231886)^t$

5 **a)** With $\mathbf{x}^{(0)} = (0.1, 0.1)^t$, $\mathbf{x}^{(8)} = (5.343082, -0.6262875)^t$ and $\mathbf{G}(\mathbf{x}^{(8)}) = 0.006995494$

b) With $\mathbf{x}^{(0)} = (0,0)^t$, $\mathbf{x}^{(13)} = (0.6157412, 0.3768953)^t$ and $\mathbf{G}(\mathbf{x}^{(13)}) = 0.1481574$

c) With $\mathbf{x}^{(0)} = (0,0,0)^t$, $\mathbf{x}^{(5)} = (-0.6633785, 0.3145720, 0.5000740)^t$ and $G(\mathbf{x}^{(5)}) = 0.6921548$

d) With $\mathbf{x}^{(0)} = (1,1,1)^t$, $\mathbf{x}^{(4)} = (0.04022273, 0.01592477, 0.01594401)^t$ and $G(\mathbf{x}^{(4)}) = 1.010003$

CHAPTER 11

Exercise Set 11.1 (page 565)

1 **a)**

i	x_i	w_{1i}
1	1.047197	−0.533308
2	2.094394	−1.51942

b)

i	x_i	w_{1i}
1	0.7853982	0.1631159
2	1.570796	−0.7662053
3	2.356194	−1.246696

3 **a)**

i	x_i	w_{1i}	w_{2i}
5	1.25	0.1676179	1.656001
10	1.50	0.4581901	0.8016986
15	1.75	0.6077718	0.4406008
20	2.00	0.6931460	0.2610475

b)

i	x_i	w_{1i}	w_{2i}
5	1.0	0.00865076	−0.04179500
10	2.0	0.00007484	−0.00036134
15	3.0	0.00000065	0.00000313
20	4.0	0.00000001	0.00000003

c)

i	x_i	w_{1i}	w_{2i}
3	0.3	0.7833204	−1.800761
6	0.6	0.6023521	0.2968196
9	0.9	0.8568906	1.305988

Exercise Set 12.3 (page 632)

1 a)

i	j	x_i	t_j	w_{ij}
1	1	0.5	0.05	0.632952
2	1	1.0	0.05	0.895129
3	1	1.5	0.05	0.632952
1	2	0.5	0.1	0.566574
2	2	1.0	0.1	0.801256
3	2	1.5	0.1	0.566574

b)

i	j	x_i	t_j	w_{ij}
1	1	1/3	0.05	1.59728
2	1	2/3	0.05	−1.59728
1	2	1/3	0.1	1.47300
2	2	2/3	0.1	−1.47300

3 a) For $h = 0.1$ and $k = 0.01$:

i	j	x_i	t_j	$w_{i,j}$
4	50	0.4	0.5	-9.3352×10^8
10	50	1.0	0.5	-9.1860×10^8
17	50	1.7	0.5	2.6047×10^8

b) For $h = 0.1$ and $k = 0.005$:

i	j	x_i	t_j	$w_{i,j}$
4	100	0.4	0.5	0.1757581
10	100	1.0	0.5	0.2990169
17	100	1.7	0.5	0.1357617

c)

i	j	x_i	t_j	$w_{i,j}$
4	10	0.8	0.4	1.166142
8	10	1.6	0.4	1.252404
12	10	2.4	0.4	0.4681804
16	10	3.2	0.4	−0.1027628

d)

i	j	x_i	t_j	$w_{i,j}$
2	10	0.2	0.4	0.3921147
4	10	0.4	0.4	0.6344550
6	10	0.6	0.4	0.6344550
8	10	0.8	0.4	0.3921148

5 a) For $h = 0.1$ and $k = 0.01$:

i	j	x_i	t_j	$w_{i,j}$
4	50	0.4	0.5	2.3541×10^{-9}
10	50	1.0	0.5	1.7610×10^{-17}
17	50	1.7	0.5	-3.8090×10^{-9}

b) For $h = 0.1$ and $k = 0.01$:

i	j	x_i	t_j	$w_{i,j}$
4	50	0.4	0.5	0.1770914
10	50	1.0	0.5	0.3012839
17	50	1.7	0.5	0.1367806

Ch 2

✓ #30 x HW 1 base on 5 pts for each a, b, c, d do these matrix eq have sol

#23 x 10 pts Crout to solve LU

#14c x 10 pts LU

#52 x 20 pts 10/Jacobi 10/Gauss Seidel , iterative methods

✓ #32 (a) x 5 pts. determinant

39 Y 3 pts for each 18 pts find norms

42 Y 20 pts / 5 each solve system to evaluate error

58 Y only for b, Y 10 pts. relaxation methods

62 10 pts. z soln of non lin eqs.

Ch. 4 36 & 42 & 48 z trap, simpson's rule, romberg, ~~RK y'''~~

Ch 5 6, 11, 54, 23 10 each euler, RK, milne, RK y'''

Ch 6 1 & 5 Prove Leap Frog/Dufort Frankel 1 5a 5b 5c 5d

Ch 7 #16, #26 check using S.O.R. $w = 1 - 1.5\ (.1)$ 10 / 20 pts

Ch 8 #5, #12, #26 30 each

Ch 9 #2 15 pts

PROBLEM 3. A large gun with supporting base weighs 3900 lbs. It is to be designed so that its spring-dashpot shock absorber system must be **critically damped**. If the damping constant is 4925 lb-sec/ft, determine:

1) The initial recoil velocity, given that the gun recoils a maximum of 5 feet from its initial position of x(t=0) = 0.

2) How long does it take to reach its maximum recoil distance.

$c = c_c = 4925$ lb-sec/ft

$= 2\sqrt{km} = 2\sqrt{\frac{kW}{g}}$; $W = 3900$ lb, $m = 121.12$ slugs

$k = \frac{c_c^2 g}{4W} = 50067$ lb/ft

$x = (C_1 + C_2 t) e^{-\omega_n t}$

$\omega_n = \sqrt{\frac{k}{m}} = 20.33$ rad/s

$x(t=0) = 0 = C_1$

$\dot{x}(t) = [C_2 - \omega_n (C_1 + C_2 t)] e^{-\omega_n t}$

$\dot{x}(t=0) = C_2 - \omega_n C_1 = C_2$

$\dot{x} = 0 \Rightarrow t^* = \frac{1}{\omega_n} - C_1/C_2 = \frac{1}{\omega_n}$

$t^* = \frac{1}{20.33} = 0.0492$ sec

$x(t^*) = 5 = C_2/\omega_n e^{-1}$

$\Rightarrow C_2 = 5\omega_n e$

$C_2 = 276.33$ ft/s

$\dot{x}(t=0) = C_2 = 276.33$ ft/s

<u>Version 2</u>

Same

$W = 7800$ lb $m = 242.24$ slugs

$k = 25033$ lb/ft

Same

$\omega_n = \sqrt{\frac{k}{m}} = 10.17$ rad/s

Same

same

Same

Same

$t^* = \frac{1}{10.17} = 0.0984$ sec

Same

Same

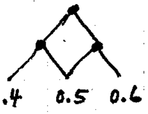
$C_2 = 138.17$ sec

$\dot{x}(t=0) = 138.17$ sec

1 DO IN CHAPTER 9 PROBLEM ~~[illegible]~~ 13 ON PG 621

DETERMINE CHARACTERISTICS

FIND $u(x,t)$



2. DO IN CHAPTER 7 PROB 22 ON PG 539

ASSUME $\Delta x = \Delta y = 0.5$ in

3. DO PROBLEM # 3 CHAPTER 8 ON PG 588

using $\rho = 168$ lb_m/ft^3

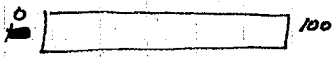
$c = 0.212$ $BTU/lb_m\,°F$

$k = 0.0370$ $BTU/sec\,°F\,ft$

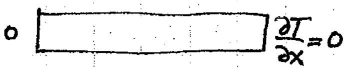
$\alpha = \dfrac{k}{c\rho}$

take $\Delta x = 1$ in

initially



at $t \leq 0$



@ $t=0$

use $\dfrac{\alpha \Delta t}{\Delta x^2} =$ ~~[illegible]~~ then 1/4

④ What can you say about the stability of following numerical scheme to solve $\dfrac{\partial T}{\partial t} = \alpha \dfrac{\partial^2 T}{\partial x^2}$

$$(1+R)T_{i\,j+1} = T_{ij} + R\,(T_{i+1\,j+1} - T_{ij} + T_{i-1\,j}) \qquad R = \frac{\alpha \Delta t}{\Delta x^2}$$

what is its drawbacks (self start, non self start, etc.)