Review Of Numerical Techniques

CS 471/Math 471
Brian T. Smith, UNM
Andrew C. Pineda, UNM

Introduction to Numerical Techniques

- Recommend you review the material in Ch. 2 of IHPSC, Fosdick et al
- Techniques that provide the foundation for numerical modeling and simulation
  – provides mathematical modeling techniques
  – provides the analysis techniques to ensure that you have valid numerical results

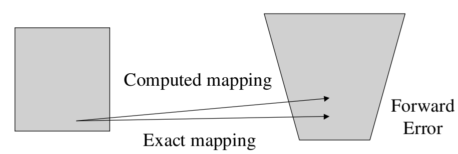
Reference and Topics

– Numerical Computation: Methods, Software, and Analysis, C. W. Ueberhuber, Springer-Verlag, 1997, ISBN 3-540-62058-3 & 62057-5

– Topics covered in this presentation
- Modeling and Errors
- Taylor Series
- Elementary Linear Algebra
- Elementary Numerical Solutions of Ordinary Differential Equations

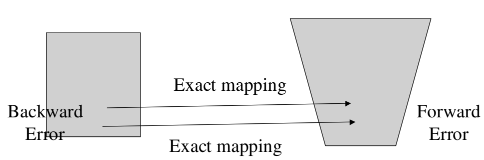
Modeling And Errors

Problem Space Solution Space



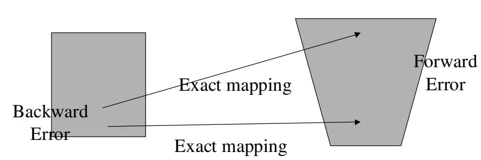
Well-Posed Problems

Problem Space Solution Space



Badly-Posed Problems

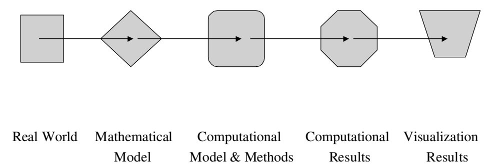
Problem Space Solution Space



Many Models

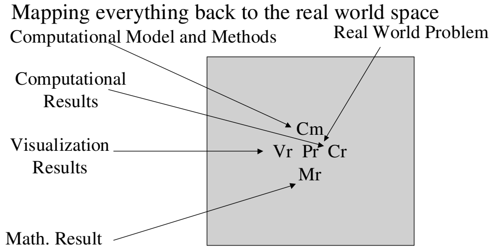
- Real World
  - space of problems to be solved in real world
- Mathematical Model
  - approximate the problem with a simpler, more tractable model
- Computational Model and Method
  - approximate the mathematical model with a discrete model and a method to solve the discrete model
- Computational Results
  - approximate with finite precision arithmetic
- Visualization
  - display the result on a 1200x800 pixel platform

Many Models



Viewing Error In The Real World

Mapping everything back to the real world space



Comments On Real World Picture

- The previous slide locates the real world problem with the problems actually solved clustered around it.
- The differences are the backward approximation errors intrinsic in the modeling and computational processes
- Small differences represent stable results
- Large differences represent unstable results
  - obviously, the goal is stable results

Measuring Error

- Absolute error measurement

$$e(x) = x_{exact} - x_{approx}$$

- Relative error measurement

$$r(x) = e(x) / x_{exact}$$

- We can often estimate an upper bound for r(x), say it is: $|r|_{ub}$

Measuring Error Continued

- If both x_{exact} and x_{approx} are positive, then

$$\frac{x_{approx}}{1+|r|_{ub}} \le x_{exact} \le \frac{x_{approx}}{1-|r|_{ub}}$$

- If both x_{exact} and x_{approx} are negative, then

$$\frac{x_{approx}}{1-|r|_{ub}} \le x_{exact} \le \frac{x_{approx}}{1+|r|_{ub}}$$

Rounding Error Analysis

– How do we do it? A lot of work

- a simple example
  – compute $a^2 - b^2$ assuming a and b are exact
  – use the method (a-b)(a+b)
- error analysis shows:

$$fl((a-b)(a+b)) = ((a-b)(1+\eta_1)(a+b)(1+\eta_2))(1+\eta_3)$$

$$r(x) = \eta_1 + \eta_2 + \eta_3 + O(\varepsilon^2)$$

- bounded by: 3ε approximately
- the wrong way to compute it: $a^2 - b^2$ Why?

Rounding Errors Continued

- Supposed we tried the computation $a^2 - b^2$
- The results assuming the usual floating point analysis model yields:

$$fl(a^2 - b^2) = (a^2(1+\eta_1) - b^2(1+\eta_2))(1+\eta_3)$$

$$= ((a^2 - b^2) + a^2\eta_1 - b^2\eta_2) + (a^2 - b^2 + a^2\eta_1 - b^2\eta_2)\eta_3$$

$$\frac{fl(a^2 - b^2) - (a^2 - b^2)}{a^2 - b^2} = \eta_3 + \frac{a^2\eta_1 - b^2\eta_2}{a^2 - b^2}(1+\eta_3)$$

Rounding Errors Continued

– Cancellation of significant digits occurs and is potentially damaging when:

- a subtraction occurs between two nearly operands
- the operands are imprecise, that is, are the result of computations that have rounding errors in them

– The computation $a^2 - b^2$ is a good and simple example

– See Section 2.2 for a different way of describing the cancellation problem

Taylor's Series

– Is an approximation technique, and is only one of many possible techniques

– Assumes an arbitrary function f(x) which is continuous and differentiable to some number of derivatives

– Essence: to determine a simple approximation, a polynomial in this case, that matches the function f(x) and its derivatives at a point x_0

Taylor Series Continued

- Constant approximation: match the $f(x)$ at x_0

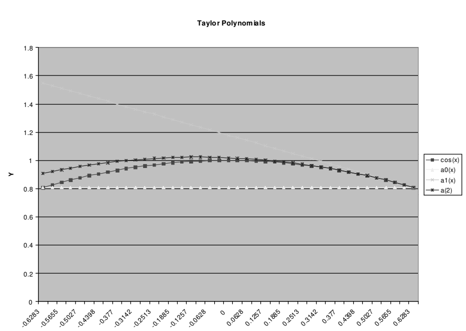
$$a_0(x) = f(x_o)$$

- Linear approx: match $f(x), f'(x)$ at x_0

$$a_1(x) = f(x_0) + (x - x_0) f'(x_0)$$

- Quadratic approx: match $f(x), f'(x), f''(x)$ at x_0

$$a_2(x) = f(x_0) + (x - x_0) f'(x_0) + (x - x_0)^2 f''(x_0)/2!$$



General Form and Approximation Error

- General form:

$$a_n(x) = f(x_0) + (x - x_0) f'(x_0) + (x - x_0)^2 f''(x_0)/2! + \ldots + (x - x_0)^n f^{(n)}(x_0)/n!$$

- Error term

$$f^{(n+1)}(\xi)(x - x_0)^{n+1}/(n+1)!$$

where ξ is in the interval (x, x_0)

10/10/1997 na_intro 19

Bounding The Truncation Error

- To determine how accurate your Taylor polynomial is, we must bound this error term, say as follows:

$$\max_{\xi \in (x, x_0)} |f^{(n+1)}(\xi)(x - x_0)^{n+1}/(n+1)!|$$

10/10/1997 na_intro 20

Linear Algebra

- Vector x -- traditionally a column
  - x^T is a row vector
  - $\| x \|_2$ is the length of x
  - this length is defined as $(\sum_{i=1}^{n} x_i^2)^{1/2}$
  - usual vector operations, inner products
  - linear combination of vectors $x^{(1)}, x^{(2)}, \ldots, x^{(k)}$

$$Xa = \sum_{i=1}^{k} a_i x^{(i)}$$

10/10/1997 na_intro 21

Special Vectors

- Unit vectors $e^{(j)}$

$$e_i^{(j)} = \begin{cases} 1 & (i = j) \\ 0 & (i \neq j) \end{cases}$$

- Orthogonal vectors $x^T y = 0$

10/10/1997 na_intro 22

Vector Norms

- A vector norm measures the length of a vector
  - has to satisfy 4 properties
    - (1) ||x|| > 0 if x /= 0
    - (2) ||x|| = 0 if and only if x = 0
    - (3) If s is a scalar, ||sx|| = |s| ||x||
    - (4) If y is a vector, then ||x + y || <= ||x|| + ||y||
- There are many vector norms: e.g.. p-norms
  - for any integer p

$$\| x \|_p = (\sum_{i=1}^{n} x_i^p)^{1/p}$$

10/10/1997 na_intro 23

Matrices

- Matrix A of order n: $A = (a_c^{(1)}, a_c^{(2)}, \ldots, a_c^{(n)})$ or

$$A = \begin{pmatrix} a_{11}\ a_{12} \ldots a_{1n} \\ a_{21}\ a_{22} \ldots a_{2n} \\ . \ \ . \ \ . \ \ . \ \ . \\ a_{n1}\ a_{n2} \ldots a_{nn} \end{pmatrix}$$

10/10/1997 na_intro 24

Matrix Properties

– Rank: number r of linearly independent columns
  - that is, the largest integer r such that

$$\sum_{\substack{i \in (1,n) \\ r\ values}} a_c^{(i)} x_i \neq 0$$

  - same as the number of linearly independent rows

– Orthogonal matrix Q: $Q^T Q = I = QQ^T$

Special Matrices

- Identity I:
  $I = (\delta_{jk}), \delta_{jj} = 1, \delta_{jk} = 0\ for\ \ j \neq k$
- Diagonal D:
  $D = (d_{jk}), d_{jj}\ \ arbitary\ , d_{jk} = 0\ for\ \ j \neq k$
- Upper triangular U:
  $U = (u_{ij}), u_{ij} = 0\ for\ \ i > j, others\ \ arbitrary$
- Lower triangular L: $l_{ij} = 0\ for\ \ i < j$

Special Matrices Continued

- Symmetric matrix S:
  $S = (s_{ij}),\quad s_{ij} = s_{ji}$ (for all i,j)
- Normal matrix N: $N^T N = N\ N^T$
  - is a matrix with a full set of eigenvectors
  - similar to a diagonal matrix
- Defective matrix: fewer than n eigenvectors

Matrix Properties

- Determinant of a singular matrix is zero
- det(A*B) = det(A)*det(B)
  $(A * B)^T = B^T * A^T,\ \det(A^T) = \det(A)$
- Provided A and B are non-singular,
  $(A * B)^{-1} = B^{-1} * A^{-1}$

  $\det(A^{-1}) = 1 / \det(A)$
- For orthogonal Q, ||Q|| = 1, ||Qx|| = ||x||

Solving Linear Systems

- Systems Ux=b and Lx=b are easy to solve
  - where U is upper triangular and L is lower triangular
- The determinant of U or L is the product of the diagonal elements
  - If A = PLU (LU factorization), then
    det(A) = det(P)*det(L)*det(U)
    det(P) = 1

Analysis Of Linear System Solutions

- For a nonsingular, define $r = b - A\tilde{x}$

$$\| \tilde{x} - x \| \leq \kappa(A) \| x \| \| r \| / \| b \|$$

$$\kappa(A) = \| A \| \| A^{-1} \|$$

- For the computed $\tilde{x}$

$$\| x - \tilde{x} \| \leq \| x \| \varepsilon\ \kappa(A)\, 1.01\ (n^3 + 3n^2) \max_{i,j,k} | a_{ij}^{(k)} | / \| A \|$$

Computing The Condition Number

- Condition number is: $\kappa(A) = \| A \| \| A^{-1} \|$
- Need to estimate the norm of A inverse
- Use the fact that: $\| A^{-1} \| = \max_{b \neq 0} \frac{\| A^{-1} b \|}{\| b \|}$
- Use a special algorithm to select b and iterate (at most twice) with it using the LU factorization of A

Vector Norms

- Maximum element norm (infinity norm)
  $\| x \|_\infty = \max_{1 \le i \le n} | x_i |$
- Maximum sum norm (1-norm)
  $\| x \|_1 = \sum_{i=1}^{n} | x_i |$
- Euclidean norm (2-norm)
  $\| x \|_2 = (\sum_{i=1}^{n} x_i^2)^{1/2}$

Matrix Norms

– Same properties as vector norms
  - maximum row norm (1-norm)
    $\| A \|_1 = \max_{1 \le j \le n} \{ \sum_{i=1}^{n} | a_{ij} | \}$
  - maximum column norm (infinity norm)
    $\| A \|_\infty = \max_{1 \le j \le n} \{ \sum_{i=1}^{n} | a_{ij} | \}$
  - spectral norm (2-norm)
    $\| A \|_2 =$ maximum magnitude of all eigenvalues of A

Consistent Vector And Matrix Norms

- For consistent matrix and vector norms
  ||Ax|| <= ||A|| ||x||
- The following pairs are consistent norms
  - max. element vector norm and max. matrix row norm
  - max. sum vector norm and max. matrix column norm
  - Euclidean vector norm and matrix spectral norm

Special Matrix Transformations

– Given a vector x with $x_1 \neq 0$, there exists a lower triangular matrix L such that $Lx_1 = x_1 e_1$
  - the form of L is:
$$L = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ -x_2/x_1 & 1 & \cdots & 0 \\ \vdots & & & \\ -x_n/x_1 & 0 & \cdots & 1 \end{pmatrix}$$
  - this elementary transformation is used to form the LU factorization of a matrix (Gaussian elimination)
  - row pivoting is necessary to maintain stability

Special Matrix Transformations Continued

– Given a vector $x (x \neq 0)$, there exists an orthogonal matrix H such that $Hx = \| x \|_2 \; e_1$
  H is of the form $I - \alpha \, w \, w^T$ where
  $\sigma = \pm \| x \|_2$, sign is that of x_1
  $w = x + \sigma \; e_1$
  $\alpha^{-1} = \| w \|_2^2 / 2 = w^T w / 2$
  - this orthogonal transformation is used to compute the QR factorization of a matrix
  - orthogonal transformations are norm preserving
  - column pivoting is used to improve the stability

Eigenvalue Problem

- For any matrix, there exists a scalar λ and a vector x such that $Ax = \lambda x$
- For a matrix of order n, there are exactly n λs and at most n independent vectors x
  - the λs may be multiple
  - when multiple, there may be fewer independent eigenvectors than the multiplicity

Eigenproblems Continued

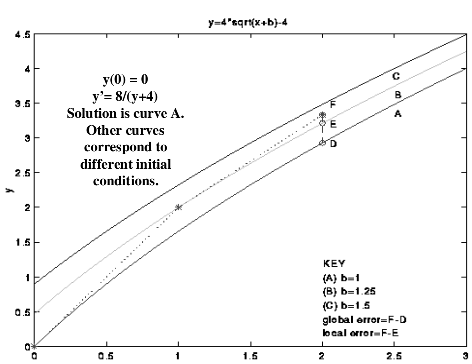
- A normal matrix has a full set of eigenvectors
  - a symmetric matrix is normal
  - a complex Hermitian matrix is normal
- Stability of the eigenvalues
  - for a symmetric (or Hermitian) matrix, the eigenvalues are stable (small changes in matrix, maintaining symmetry, causes small changes in the eigenvalues
  - for a general matrix, the eigenvalues are unstable when the eigenvalues are near multiple

Solving Ordinary Differential Equations

- 1st order ordinary differential equation problem
  - given f(x,y), find a solution y such that $y' = f(x, y)$
    - there are many solutions (all differing by a constant)
    - a unique solution is proved if an initial value for y is also provided
- 2nd order ordinary differential equation problem
  - given $f(x, y, y')$, find y such that $y'' = f(x, y, y')$
  - there are many solutions as before
  - the solution becomes unique if two initial conditions are provided or two boundary conditions

Graph Of A Solution Technique

- Try computing the solution as $y_{i+1} = y_i + hf(x_i, y_i)$ with y(0) = c



Analysis Of Results -- Errors

- Study the model problem $y' = y$, y(0) = 1
- The computed solution is:

$$y(0) = y_0 = 1$$
$$y_1 = y_0 + hf(x_0, y_0) = 1 + h$$
$$y_2 = y_1 + hf(x_1, y_1) = (1+h) + h(1+h) = (1+h)^2$$
$$. \quad . \quad .$$
$$y_n = (1+h)^n = 1 + nh + n(n-1)h/2 + O(h^3)$$

Analysis Continued

- The exact solution is $y(x) = e^x$ for y(0)=1
- The error in the computed solution is:

$$\begin{aligned} Error &= e^{nh} - (1+h)^n = nh^2/2 + \cdots \\ &= hx_n/2 + \cdots = O(h) \end{aligned}$$

- Thus, the global error is O(h)
- Local error is the error of a single step, namely, the Taylor series error $O(h^2)$

Solving 2nd Order ODEs

- Given $f(x,y,z)$ find a solution y to $y'' = f(x, y, y')$ with initial or boundary conditions for y and y'
  - introduce a new variable z defined as y'. Then

    $y' = z$

    $z' = f(x, y, z)$
  - using Euler's method on each equation above

    $y_0 = y(0) \quad z_0 = y'(0)$

    $x_i = ih \quad y_i = y(x_i) \quad z_i = z(x_i)$

    $y_{i+1} = y_i + hz_i$

    $z_{i+1} = z_i + hf(x_i, y_i, z_i)$

A Different Approach To Approximation

- Instead of Euler (truncated Taylor series), approximate the derivatives in the equation with finite differences
  - 1st derivative using a forward difference:
  - 1st derivative using a backward difference
  - 1st derivative using a central difference
  - 2nd derivative using forward difference

Approximations

- Forward difference is O(h):

  $$y'(x) \approx (y(x+h) - y(x))/h$$
- Backward difference is O(h):

  $$y'(x) \approx (y(x) - y(x-h))/h$$
- Central difference is $O(h^2)$:

  $$y'(x) \approx (y(x+h) - y(x-h))/2h$$
- 2nd derivative difference is $O(h^2)$:

  $$y''(x) \approx (y(x+h) - 2y(x) + y(x-h))/h^2$$

Using These Approximations

- Using the forward difference on $y' = f(x,y)$ gives the Euler method
- Using the backward difference on $y' = f(x,y)$ gives an implicit method (y_{i+1} in terms of y_{i+1})
- Using the 2nd derivative approximation on

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

  $y_{i+1} = 2y_i - y_{i-1} + h^2 f(x_i, y_i, (y_i - y_{i-1})/h))$
  - Need an 3 approximation for Try:

    $y_1 = y_0 + hy'_0 + h^2 f(x_0, y_0, y'_0)/2$