```
Cancellation Error: Subtracting nearly equal numbers Example: p = 0.54617, q = 0.54601, true r = p - q = 0.00016 4-digit: p^* = 0.5462, q^* = 0.5460, r^* = 0.002 (RE=25%) Nested Multiplication: Reduces error
       f(z) = 1.01z^{4} - 4.62z^{3} - 3.11z^{2} + 12.2z - 1.99 = (((1.01z - 4.62)z - 3.11)z +
       12.2)z - 1.99
• P_n(x) = \sum_{k=0}^n \frac{f^{(k)}(a)}{k!} (x-a)^k
• Remainder: R_n(x) = \frac{f^{(n+1)}(\xi)}{(n+1)!}(x-a)^{n+1}, \xi \in (a,x)
     Linear Approx: f(x_0 + h) \approx f(x_0) + hf'(x_0)
    Example: \xi \in (0, \pi/2), \sin \xi \le 1 \Rightarrow R_n \le \frac{(x_0)^n}{n!}
 §1.3 Convergence
    \begin{array}{l} \alpha = \lim_{n \to \infty} (\alpha_n) \\ \text{Rate: } \alpha_n = \alpha + \mathcal{O}(\beta_n) \text{ if } |\alpha_n - \alpha| \leq K |\beta_n| \\ \text{Find largest } p \text{ where } \alpha_n - \alpha = \mathcal{O}(1/n^p) \end{array}
 §1.4 Matrix Operations
• A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix}, B = \begin{bmatrix} b_{11} \\ b_{21} \\ b_{31} \end{bmatrix}
                                                                                             \begin{bmatrix} b_{12} \\ b_{22} \\ b_{32} \end{bmatrix} \\ a_{11}b_{12} + a_{12}b_{22} + a_{13}b_{32} \\ a_{21}b_{12} + a_{22}b_{22} + a_{23}b_{32} \end{bmatrix}
• AB = \begin{bmatrix} a_{11}b_{11} + a_{12}b_{21} + a_{13}b_{31} \\ a_{21}b_{11} + a_{22}b_{21} + a_{23}b_{31} \end{bmatrix}
§6.1 Gaussian Elimination
     \mathcal{O}(n^3) complexity
• Pivoting:

- Partial (PP): Max element in column
            Scaled PP: s_i = \max_j |a_{ij}|, pivot \max(a_{ik}/s_i)

Scaled PP: s<sub>i</sub> = max<sub>j</sub> |a<sub>ij</sub>|, pivot max(a<sub>ik</sub>/s<sub>i</sub>)
Complete (CP): Full matrix search (O(n³))
LU Decomposition: PA = LU through GE steps and LUx = Pb.
LU Algorithm: L = E<sup>-1</sup><sub>n-a,m-b</sub>E<sup>-1</sup><sub>n-a+1,m-b+1</sub> ··· U = GE.
Choleski Factorization: If a matrix is symmetric and positive definite, it may be factored to the form LDL<sup>T</sup>
86.2 Special Matrices

    Inverse Matrix: An inverse matrix of A is A^{-1} such that AA^{-1} = I Properties: (AB)^{-1} = B^{-1}A^{-1}, (A^{-1})^T = (A^T)^{-1} Singular: A matrix is singular iff its det is 0. Diagonal: d_{ij} = 0 for i \neq j: All non-diagonal entries are 0. Symmetric: A = A^T, (AB)^T = B^TA^T

    Bisection:

     Permutation: Row swaps of I_n, PA reorders rows: P^T = P^{-1} Diagonally Dominant: |a_{ii}| > \sum_{j \neq i} |a_{ij}| (nonsingular)
      Positive Definite: x^T Ax > 0 \Rightarrow A = LDL^T, a_{ii} > 0, a_{ij}^2 < a_{ii}a_{jj}
       Minor: M_{ij} is a submatrix of A with the row i deleted and column j deleted.
      Band: an n \times n matrix is a band matrix if p, q \in \mathbb{Z}: 1 \le p, q \le n exist with a_{i_j} = 0 for i + p \le j or j + q \ge i The bandwidth is defined as w = p + q - 1. For
     adiagonal matrix, p=1, q=1, w=1 a diagonal matrix, p=1, q=1, w=1 Tridiagonal: Band with p=2, q=2. It exhibits the following properties -a_{ii}=l_{ii} -a_{i,i+1}=l_{ii}u_{i,i+1}: i=1\cdots n-1 -a_{i:i-1}=l_{i,i-1}: i=2,3,\cdots, n -a_{ii}=l_{i,i-1}u_{i-1,i}+l_{ii}: i=2\cdots n
 §Strategies
     RoC With inf limit: set h = 1/n and solve accordingly.
       D \cdot (L+U): given D has ONLY diagonal entries and L+U has NO diagonal entries,
       the resulting matrix A is composed of entries a_{ij} = d_{ii} \cdot (l+u)_{ij}
      Verification of Bisection: To verify bisection can be applied, make sure that f(a) and f(b) are of different signs.
       Error of Bisection: To compute the accuracy of bisection to an \varepsilon, we use
• Failure of Newton's Method: NM Fails if f'(x) = 0 for some x.

• Triangle Inequality: |x+y| \le |x| + |y|

§Key Definitions & Identities

• Continuity: f \in C^n[a, b] reads: the nth derivative of f on [a,b] is continuous.
    Series Expansions
    Series Expansions -e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots \quad (\forall x \in \mathbb{R})
-\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \cdots
-\cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \cdots
-\arctan x = x - \frac{x^3}{3} + \frac{x^5}{5} - \frac{x^7}{7} + \cdots \quad (|x| \le 1)
-\ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \cdots \quad (|x| < 1)
-\frac{1}{1-x} = 1 + x + x^2 + x^3 + x^4 + \cdots \quad |x| < 1

\frac{1}{1+x} = 1 - x + x^2 + x^3 + x^4 - \dots : |x| < 1

Core Identities
-\sin^2 \theta + \cos^2 \theta = 1

-\sin(a \pm b) = \sin a \cos b \pm \cos a \sin b
-\cos(a \pm b) = \cos a \cos b \mp \sin a \sin b
                                                                                                                   Angle Transformations
                                                                                                                  Angle Haistonia to 18 -\sin 2\theta = 2\sin \theta \cos \theta -\cos 2\theta = 2\cos^2 \theta - 1 = 1 - 2\sin^2 \theta -\sin^2 \theta = \frac{1-\cos 2\theta}{2}, \cos^2 \theta = \frac{1+\cos 2\theta}{2}
    Vector Norms
- ||x|| > 0
- ||x|| = 0 \Leftrightarrow x = 0
                                                                                                                  \begin{array}{ll} - & ||\alpha x|| = |\alpha|||x|| \\ - & ||x+y|| \le ||x|| + ||y|| \end{array}
 §Useful Examples
    Suppose g(x) = \frac{5}{x^2} + 2 Show p_n = g(p_{n-1}) will converge to g for \forall p_0[2.5, 3].
      Since this is a decreasing function, the max of g(x) is g(2.5) and the min of g(x)
           First compute the max, g(2.5) = 14/5 < 3
- second compute the min, g(3)=5/9+2>2.5

- Last compute |g'(x)|=-10/x^3\leq \max_{x\in[2.5,3]}|g'(x)|=16/25<1

• Given ||A|| is a natural matrix norm of matrix A. show |\lambda|\leq ||A|| for any nonsingular A and any \lambda of A. ||A||=\max_{||x||=1}||Ax||\geq ||Ax||: x is an e-vec s.t — x—=1=||\lambda x||=|\lambda|||x||=|\lambda|\square

• When performing Jacobi or GS, when computing L+U, flip the signs of all entries.

• To determine convergence for fixed point, compute g'(p_0)\leq 1, which gives a,b. Prove g(x) cts on [a,b], g(x)\in [a,b], g'(x) exists on (a,b), |g'(x)|\leq k: \forall x\in (a,b),0< k<1
     - Second compute the min, g(3) = 5/9 + 2 > 2.5
```

Normalization: 32 bit - Sign bit: 1, Sign exp: 1, Exp: 7, Normalized mantissa: 23.

Absolute Error =  $|p - p^*|$ , Relative Error =  $\frac{|p - p^*|}{|p|}$ 

Significant Digits: RE  $< 5 \times 10^{-t}$ 

fl(x): Machine representation

§1.1 Error Analysis

```
§LA Determinants
2x2: |A| = ad - bc
                                \sum_{j=1}^{n} a_{ij} A_{ij} \quad \text{via} \quad \text{cofactors} \quad A_{ij}
                                                                                              (-1)^{i+j}M_{ij}
• nxn:
               |A|
                                                            – Identical rows: |A| = 0
  Properties
                                                            - |AB| = |A||B|, |A^T| = |A|
- |A^{-1}| = \frac{1}{|A|}
   - Swap rows: |\tilde{A}| = -|A|
  - Scale row: |\tilde{A}| = \lambda |A|
§7 Norms & Eigen
• ||x||_2 = \sqrt{\sum x_i^2}, ||x||_{\infty} = \max |x_i| • ||A||_2 = \sqrt{\rho(A)}
```

 $||A||_{\infty} \stackrel{\bullet}{=} \max_{i} \sum_{j} |a_{ij}| \text{ Basically} \stackrel{\bullet}{\bullet} ||A|| = \max_{||x||=1} ||Ax||$ sum all rows together and deter-

wine the largest one.  $\forall x \in \mathbb{R}^n : ||x||_{\infty} \leq ||x||_2 \leq \sqrt{n}||x||_{\infty}$ A distance between matrices A and B wrt a matrix norm  $||\cdot||$  is ||A - B||

Theorem: For any vector  $x \neq 0$ , matrix Am and abt natural norm  $||\cdot||$  we have  $||Ax|| \le ||A|| \cdot ||x||$ 

 $\begin{aligned} &||Ax|| \leq ||A|| \cdot ||x|| \\ &| & \text{Cauchy-Schwarz: } ||x+y||_2 \leq ||x||_2||y||_2 \\ &| & \text{Eigen: } \lambda \text{ is an eigenvalue if } A\lambda = v \cdot \lambda \\ &| & \text{Finding Eigenthings: } \det(A-\lambda I) = 0 : \forall \lambda. \\ &| & \text{Spectral Radius: } \rho(A) = \max |\lambda_i|, \, \rho(A) \leq ||A|| \\ &| & \text{Theorem: If A is } n \times n: \\ &- ||A||_2 = [\rho(A^tA)]^{1/2} \\ &- \rho(A) \leq ||A|| : \forall ||\cdot|| \end{aligned}$ 

 $\begin{array}{l} \rho(A) \geq ||A||: \, \forall ||\cdot|| \\ \text{Convergent: } \lim_{k \to \infty} A^k = 0 \Leftrightarrow \rho(A) < 1 \\ \text{Matrix Norms: matrix norms have the following properties} \\ - \, ||A|| \geq 0 \text{ with } ||A|| = 0 \leftrightarrow A = 0 \\ - \, ||\alpha A|| = |\alpha| \cdot ||A|| \\ - \, ||A + B|| \leq ||A|| + ||B|| \end{array}$  Iterative Methods

### §Iterative Methods

• General Iteration:  $x^{k+1} = Tx^k + c$ • Jacobi:  $x^{k+1} = D^{-1}(L+U)x^k + D^{-1}b$ • Gauss-Seidel:  $x^{k+1} = (D-L)^{-1}Ux^k + (D-L)^{-1}b$ 

Stein-Rosenberg: For matrices with positive diagonals:  $\rho_{GS} \leq \rho_J < 1$ Speed of Convergence: given matrices  $T_{GS} = (D-L)^{-1}U$  and  $T_J = D^{-1}(L+1)$ U), compare  $\rho$ . The bigger the  $\rho$ , the faster the convergence.

• Error:  $||x-x^k|| \leq \frac{||T||^k}{1-||T||} ||x^1-x^0||$ • Stopping:  $\frac{||x^k-x^{k-1}||}{||x^k||} < \varepsilon$ 

### §2 Nonlinear Equations

- While  $f(p_n) \neq 0$  or < T:  $p_n = \frac{a_1 + b_1}{2}$ - Error:  $\frac{b_n - a_n}{2} < T$ ,  $p = a + \frac{b - a}{2}$ 

• Fixed-Point:

 $p_n = g(p_{n-1})$ , converges if  $|g'(x)| \le K < 1$ Algorithm: For  $i < N_0$ :  $p = g(p_0)$ , check  $|p - p_0| < T$ 

A fixed point is defined as a point in which p = f(p)

Newton:  $- p_n = p_{n-1} - \frac{f(p_{n-1})}{f'(p_{n-1})}$ Newton:

Quadratic convergence if  $f'(p) \neq 0$ 

Secant:  $- p_{n+1} = p_n - \frac{f(p_n)(p_n - p_{n-1})}{f(p_n) - f(p_{n-1})}$ 

- Approx derivative:  $\frac{f(p_{n-1}) - f(p_{n-2})}{p_{n-1} - p_{n-2}}$ 

Algorithm: Store  $q_0 = f(p_0)$ ,  $q_1 = f(p_1)$ , SET  $p = p_1 - \frac{q_1(p_1 - p_0)}{q_1 - q_0}$  IF STOP-PING CONDITION: RETURN p; i++,  $p_0=p_1,q_0=q_1,p_1=p,q_1=f(p)$  ENDWHILE OUTPUT FAILURE.

§Theorems

Bisection: Suppose  $f \in C[a,b]: f(a) \cdot f(b) < 0$ . Bisection generates  $\{p_n\}$  approximating a zero p with  $|p_n - p| \le \frac{b-a}{2n}: n \ge 1$ Fixed Point: If  $g \in C[a,b], g([a,b]) \subseteq [a,b]$  g has a fixed point in [a,b], addi-

tionally if  $|g'| \le K < 1$ , then the fixed point is unique. **Fixed Point Theorem**: Let  $g \in C[a,b]$  and  $g(x) \in [a,b] : \forall x \in [a,b]$ . Suppose

as well that g' exists on (a,b) and positive K < 1 exists with  $|g'(x)| \leq K : \forall x \in \mathcal{C}$ (a,b). Then for any number  $p_0 \in [a,b]$ h the sequence defined by  $p_n = g(p_{n-1})$ :  $n \ge 1$  converges to the unique point  $p \in [a,b]$ Corollary: If g satisfies the hypothesis of the above theorem,  $|p_n - p| \le 1$ 

 $k^n \max(p_0 - a, b - p_0)$  and  $|p_n - p| \le \frac{k^n}{1 - k} |p_1 - p_0| : \forall n > 1$ 

Newton: For f ∈ C²[a, b] with simple root, ∃δ > 0 : p<sub>0</sub> ∈ [p − δ, p + δ] converges.
Matrix Invertibility: |A| ≠ 0 ⇔ unique solution Ax = b ⇔ A⁻¹ exists

• Taylor: With  $R_n(x) \Rightarrow f(x) = P_n(x) + R_n(x)$ ,  $R_n(x) = \frac{f^{n+1}(\xi)}{(n+1)!}(x - f(x))$  $(x_0)^{n+1}$ : $\xi \epsilon(x,x_0)$ 

Existence of Inverse: if A is square, detA ≠ 0 ↔ Ax = 0 has soln x = 0 ↔ Ax = b has a unique soln for any n-vector b. ↔ A<sup>-1</sup> exists.
Diagonally Dominant Matrices: dd matrices are nonsingular. A being dd

means LU can be performed without P. A matrix is positive definite if  $x^t Ax > 0$ . PD matrices are nonsingular,  $\forall i = 1, \dots, n : a_{ii} > 0, \max |a_{kj}| > \max |a_{ii}|,$ 

Convergence of DD: If A is strictly DD, Jacobi and GS converge to the unique

• Positive Definitive Check: A matrix is PD if the leading principle submatrix determinants are positive.

Positive Definitive Check: A matrix is PD iff it may be factored into  $LL^T$  Non-singularity Check: A matrix A has an inverse iff det  $A \neq 0$ . Determinant of Triangular Matrices: The determinant of a triangular matrix

• Bisection (THM1):  $\forall n \geq 1$ :  $b_n - a_n = (b-a) \cdot \frac{1}{2^{n-1}} : p\epsilon(a_n, b_n)$ . Since  $p_n=\frac{1}{2}(a_n+b_n): \forall n\geq 1, |p_n-p|\leq \frac{1}{2}(b_n-a_n)=\frac{b-a}{2n}.$   $\Box$ • **Fixed Point**: Part i: If  $g\in [a,b], g(x)\in [a,b]: \forall x\in [a,b]$  then g(x) has a fixed

point in [a,b]: If g(a) = a or g(b) = b, g has a fixed point at an endpoint. Suppose for contra-

diction that it does not. g(a) > a and g(b) < b. Define h(x) = g(x) - x. Then h is cts on [a,b] and h(a) = g(a) - a > 0 and h(b) = g(b) - b < 0 IVT states that  $\exists p \in (a,b) : h(p) = 0$  Thus  $g(p) - p = 0 \Rightarrow p$  is a fixed point of g. Part ii: Suppose as well  $|g'(x)| \le k < 1 : \forall x \in (a, b)$  and that  $p, q \in [a, b] : p \ne q$ . By MVT,  $\exists \zeta : \frac{g(p) - g(q)}{p - q} = g'(\zeta)$ .  $|p - q| = |g(p) - g(q)| = |g'(\zeta)||p - q| \le g'(\zeta)$ k|p-q| < |p-q| contradiction.

# §2: Error Analysis and Acelerating Convergence

Basic Methods

- Newton's Method: Quadratic convergence if  $f'(p) \neq 0$ . Iteration:
- $x_{n+1} = x_n f(x_n)/f'(x_n).$
- Secant Method: Superlinear convergence (order ≈ 1.618). Uses two previous
- False Position: Bracketing method combining bisection and secant approaches. Algorithm: set  $i = 2, q_0 = f(p_0), q_1 = f(p_1)$ . While  $i \le N_0$ , set  $p = p_1$  $q_1(\frac{p_1-p_0}{q_1-q_0}).$  If  $|p-p_1|~<~TOL$  ret p. i++, q~=~f(p). If  $q~\times~q_1~<~0$  then  $p_0 = p_1, q_0 = q_1$ . Set  $p_1 - p, q_1 = q$ . Endwhile fail.
- Newton's Improved Method:  $p_{n+1} = p_n \frac{f(p_n)f'(p_n)}{f'(p_n)^2 f(p_n)f''(p_n)}$

- Convergence Analysis
   A soln p of f(x) = 0 is a zero of multiplicity m of f if for  $x \neq p$  we can write  $f(x) = (x p)^m q(x)$  where  $\lim_{x \to p} q(x) \neq 0$ . (Simple zeros are multiplicity 1).
- Order of convergence  $\alpha$ :  $\lim \frac{|p_{n+1}-p|}{|p_n-p|^{\alpha}} = \lambda$ Linear  $(\alpha = 1)$ , Quadratic  $(\alpha = 2)$
- Fixed-point: Linear if  $g'(p) \neq 0$ , quadratic if g'(p) = 0 and g'' bounded.

### Special Cases

- Multiple roots: Modify Newton's using  $\mu(x) = f(x)/f'(x)$

 $\hat{p}_n = p_n - \frac{(p_{n+1} - p_n)^2}{(p_{n+2} - p_{n+1}) - (p_{n+1} - p_n)}$  Steffensen's: Combines Aitken's with fixed-point, achieves quadratic convergence  $p_{n+1} = p_n - \frac{[g(p_n) - p_n]^2}{g(g(p_n)) - 2g(p_n) + p_n}$  Polynomial Methods

Polynomial Methods

- Horner's method: Efficient evaluation (n mults/adds) Algorithm: ex: evaluate  $P(x) = \sum_{i=0}^{n} a_i x^i$  and derivative at  $x_0$ . Input  $n, a_j, x_0 : 0 \le j \le n$ . Output:  $Y = P(x_0), z = P'(x_0).$  Set  $y = a_n, z = a_n.$  For j = n-1, n-2, ..., 1 set  $y = x_0y + a_j, z = x_0z + y.$  Endfor set  $y - x_0y + a_0$  output y.z. Deflation: Find roots sequentially via  $P(x) \approx Q(x)(x - x_0)$  Müller's method: Quadratic interpolation using 3 points
- Fundamental thm of alg: If P(x) has a degree  $n \ge 1$ , P(x) has at least one root.
- Cor: there also exists unique constants  $x_1, ... x_k$  such that  $\sum_{i=1}^k m_i = n, P(x) = \sum_{i=1}^k m_i = n$  $a_n \times \prod_{i=1}^k (x - x_i)^{m_i}$

• Cor: these functions are unique. Weierstrass:  $\forall f$  cts on [a,b],  $\forall \varepsilon > 0$ ,  $\exists$  polynomial p(x) with  $|f(x) - p(x)| < \varepsilon$ 

Lagrange Interpolation (unique!):

$$P(x) = \sum_{m=0}^{N} f(x_m) L_m(x)$$
, where  $L_m(x) = \prod_{\substack{k=0 \ k \neq m}}^{N} \frac{x - x_k}{x_m - x_k}$ 

Interpolation Error:  $f(x) - P(x) = \frac{f^{(n+1)}(\xi(x))}{(n+1)!} \prod_{k=0}^{n} (x - x_k)$  for  $f \in C^{n+1}[a, b]$ Newton's Divided Differences:

$$f[x_i] = f(x_i), \quad f[x_i, \dots, x_{i+k}] = \frac{f[x_{i+1}, \dots, x_{i+k}] - f[x_i, \dots, x_{i+k-1}]}{x_{i+k} - x_i}$$

 $P_n(x) = \sum_{k=0}^n f[x_0, \dots, x_k] \prod_{j=0}^{k-1} (x - x_j)$ 

Finite Differences (step h): Forward:  $\Delta f_i = f_{i+1} - f_i$ ; Backward:  $\nabla f_i = f_i - f_{i-1}$ Newton's Formulas: Forward:  $P(x) = \sum_{k=0}^{n} {s \choose k} \Delta^k f_0$  where  $x = x_0 + sh$ Hermite: Given  $(x_j, f(x_j), f'(x_j))$ , unique degree  $\leq 2n + 1$  polynomial:

 $H(x) = \sum_{j=0}^{n} f(x_j) H_j(x) + \sum_{j=0}^{n} f'(x_j) \hat{H}_j(x)$  error:  $f(x) - H(x) = \frac{(x-x_0)^2 \cdots (x-x_n)^2}{(2n+2)!} f^{(2n+2)}(\xi)$ 

**Cubic Splines:**  $S(x_j) = f(x_j)$  and  $S \in C^2[a, b]$ ; Boundary: Clamped (S' at endpoints) or Natural (S'' = 0 at endpoints) Error:  $\max |f(x) - S(x)| \le \frac{5M}{384}h^4$ , where  $h = \max(x_{j+1} - x_j), M = \max|f^{(4)}|$ Numerical Differentiation:

- Numerical Differentiation:

   Forward/Backward (O(h)):  $f'(x_0) \approx \frac{f(x_0 \pm h) f(x_0)}{h}$  Centered  $(O(h^2))$ :  $f'(x_0) \approx \frac{f(x_0 + h) f(x_0 h)}{2h}$  Second Derivative:  $f''(x_0) \approx \frac{f(x_0 h) 2f(x_0) + f(x_0 + h)}{h^2}$

# Richardson Extrapolation: $N_{j+1}(h) = N_j(h/2) + \frac{N_j(h/2) - N_j(h)}{2^j - 1}$

# Numerical Integration:

- Trapezoid Rule  $(O(h^2))$ :  $\int_a^b f(x)dx \approx \frac{h}{2}[f(a) + f(b)]$  Simpson's Rule  $(O(h^4))$ :  $\int_a^b f(x)dx \approx \frac{h}{3}[f(a) + 4f\left(\frac{a+b}{2}\right) + f(b)]$
- Composite errors: Trapezoid  $-\frac{(b-a)h^2}{12}f''(\mu)$ ; Simpson  $-\frac{(b-a)h^4}{180}f^{(4)}(\mu)$ Romberg:  $R_{k,j} = R_{k,j-1} + \frac{R_{k,j-1} R_{k-1,j-1}}{4^{j-1}-1}$ , error  $O(h^{2j})$

Gaussian Quadrature: Uses Legendre polynomial roots as nodes, exact for degree  $\leq 2n-1$  Scale to [a,b] via  $\int_a^b f(x)dx = \frac{b-a}{2} \int_{-1}^1 f\left(\frac{(b-a)t+a+b}{2}\right)dt$ 

# §ODE Initial Value Problems

- Basic Problem:  $y'(t) = f(t, y), \ y(a) = \alpha$  Lipschitz Condition:  $|f(t, y_1) f(t, y_2)| \le L|y_1 y_2|$  Existence/uniqueness guaranteed when  $|\frac{\partial f}{\partial y}| \le L$  over convex domain D

Numerical Methods • Euler's Method:  $w_{i+1} = w_i + hf(t_i, w_i)$  [Error: O(h)]

Methods:

- Taylor Methods:  $w_{i+1} = w_i + hT^{(n)}(t_i, w_i)$  where  $T^{(n)} = f + \frac{h}{2}f' + \cdots + \frac{h}{2}f' + \cdots$  $\frac{h^{n-1}}{n!}f^{(n-1)}$  ${\bf Midpoint} \ \ ({\bf RK2}) \ \ w_{i+1} \ \ = \ \ w_i \ +$ Runge-Kutta

$$hf(t_i + \frac{h}{2}, w_i + \frac{h}{2}f_i)$$

Modified Euler (RK2)  $w_{i+1} = w_i +$  $\frac{h}{2}(f_i + f(t_{i+1}, w_i + hf_i))$ 

$$k_1 = hf(t_i, w_i)$$

$$k_2 = h f(t_i + \frac{h}{2}, w_i + \frac{k_1}{2})$$

$$k_2 = h f(t_i + \frac{1}{2}, w_i + \frac{1}{2})$$

• Classical RK4: 
$$k_3 = hf(t_i + \frac{h}{2}, w_i + \frac{k_1}{2})$$
  
•  $k_4 = hf(t_i + \frac{h}{2}, w_i + \frac{k_2}{2})$   
•  $k_4 = hf(t_i + h, w_i + k_3)$   
•  $w_{i+1} = w_i + \frac{1}{6}(k_1 + 2k_2 + 2k_3 + k_4)$ 

Error Analysis

- Local truncation error:  $\tau_{i+1} = y(t_{i+1}) w_{i+1}$  given  $w_i = y(t_i)$  Global truncation error: Accumulated error across all steps
- For Euler:  $|\tau_i| \leq \frac{h^2}{2} M$  (local), O(h) (global) where  $M = \max |y''|$
- For RK4: O(h<sup>5</sup>) (local), O(h<sup>4</sup>) (global)

- Well-posed problem requires: unique solution exists + small input changes  $\rightarrow$  small output changes
- Step size formula:  $h < \frac{2\varepsilon}{M(b-a)}$  for error  $\varepsilon$ , where  $M = \max |y''|$
- Example:  $y' = y \cos t$  has Lipschitz constant L = 1 since  $\left| \frac{\partial f}{\partial y} \right| = |\cos t| \le 1$
- Chapter 3: Interpolation Lagrange Interpolation
- $\Delta^k$ :  $\Delta^2 f_i = \Delta(\Delta(f_i)) = \Delta(f_{i+1} f_i) = f_{i+2} 2f_{i+1} + f_i$  Error:  $|f P| \le \frac{\prod |(x x_k)|}{(n+1)!} \max |f^{(n+1)}|$

- Avoid high-degree polynomials for non-smooth f; use **Chebyshev nodes** to mitigate Runge's phenomenon.

Reuse computations with **Neville's Method**: 
$$P_{i,j}(x) = \frac{(x-x_i)P_{i+1,j}-(x-x_j)P_{i,j-1}}{x_j-x_i}$$

# Hermite Interpolation

- Error term:  $\frac{(x-x_0)^2 \cdots (x-x_n)^2}{(2n+2)!} \max |f^{(2n+2)}|$
- Handle derivatives via **divided differences** with repeated nodes:  $z_{2i} = z_{2i+1}$
- Warning: Noisy derivatives  $\Rightarrow$  amplified errors.

• Cubic splines: Solve tridiagonal system (O(n) ops) with:

$$h_{i-1}c_{i-1} + 2(h_{i-1} + h_i)c_i + h_ic_{i+1} = \frac{3}{h_i}(a_{i+1} - a_i) - \frac{3}{h_{i-1}}(a_i - a_{i-1})$$

- Natural splines (S'' = 0): Stable but less accurate. Clamped splines: Require f'(a), f'(b) but higher accuracy.
- Error:  $O(h^4)$  for  $f \in C^4$ ,  $O(h^2)$  for linear splines. Numerical Differentiation

- Centered difference:  $f'(x_0) \approx \frac{f(x_0+h)-f(x_0-h)}{2h}$ , error  $O(h^2)$  Optimal h: Balance truncation  $(\propto h^2)$  and roundoff  $(\propto h^{-1})$
- Richardson Extrapolation:  $N_j(h) = N_{j-1}(h/2) + \frac{N_{j-1}(h/2) N_{j-1}(h)}{2^j}$
- Trapezoidal & Simpson's Rules

   Trapezoidal error:  $-\frac{(b-a)}{12}h^2f''(\mu)$  Simpson's error:  $-\frac{(b-a)}{180}h^4f^{(4)}(\mu)$
- Romberg Integration: Accelerate Trapezoidal Rule via:

$$R_{k,j} = R_{k,j-1} + \frac{R_{k,j-1} - R_{k-1,j-1}}{4^{j-1} - 1}$$

- Stop when  $|R_{k,k} R_{k-1,k-1}| < \epsilon$ . Adaptive Quadrature
- Error estimate:  $\frac{1}{15}|S(a,b) S(a,c) S(c,b)|, c = (a+b)/2$
- Subdivide intervals where error  $> \epsilon/2$ .

## Gaussian Quadrature

- Nodes: Roots of Legendre polynomials  $P_n(x)$ . Weights:  $c_i = \int_{-1}^1 \prod_{j \neq i} \frac{x x_j}{x_i x_j} dx$
- Transform to [-1,1]:  $x = \frac{(b-a)t + (a+b)}{2}$ ,  $dx = \frac{b-a}{2}dt$
- Exact for polynomials of degree  $\leq 2n-1$ . General Tips

- Monitor error terms:  $\propto h^k f^{(k)}$
- For oscillatory f, use splines or piecewise methods. High-degree polynomials: Unstable for noisy/non-smooth data.
- Symmetry: Exploit in Gaussian quadrature and even/odd functions.
- Thm 2.8: Let p be a soln of the eq x = g(x) and suppose g'(p) = 0 and g'' is cts and strictly bounded by M on an interval I containing p. Then  $\exists \delta > 0$  such that  $p_0 \in [p \delta, p + \delta]$  the seq:  $p_n = g(p_{n-1})$ :  $n \ge 1$  converges at least quadratically to p. Moreover, for large n,  $|p_{n+1}-p|<\frac{M}{2}|p_n-p|^2$
- Thm 2.10:  $f \in C'[a,b]$  has a simple zero at p in (a,b) iff f(p) = 0,  $f'(p) \neq 0$ Thm 2.11: The function  $f \in C^m[a,b]$  has a zero of multiplicity m at p iff  $0 = f(p) = f'(p) = f''(p) = \cdots = f^{(m-1)}(p)$
- **Thm 3.3**: Suppose  $x_0, x_1, ..., x_n$  are distinct numbers in [a,b] and  $f \in C^{n+1}[a,b]$

then for each x in [a, b], a number  $\xi(x)$  in (a,b) exists with  $f(x) = P(x) + \frac{f^{(n+1)}(\xi(x))}{(n+1)!} \prod_{i=0}^{n} (x-x_i) \text{ with P(x) being the nth Lagrange inter-}$ polating polynomial