C2-4: Computer Arithmetics

IEEE-754 Arithmetic Standard

$$\overline{se_1\cdots e_m b_1\cdots b_n}=\begin{cases}((-1)^s\times\overline{1.b_1\cdots b_n}\times 10^{\overline{e_1\cdots e_m}-\overline{1_{m-2}\cdots 1_0}})_2 & \forall\ \overline{e_1\cdots e_m}\notin\{0,\overline{1_{m-1}\cdots 1_1}\}\\ ((-1)^s\times\overline{0.b_1\cdots b_n}\times 10^{1-\overline{1_{m-2}\cdots 1_0}})_2 & \text{if } \overline{e_1\cdots e_m}=0\\ \pm\infty & \text{if } \overline{e_1\cdots e_m}=\overline{1_{m-1}\cdots 1_1}\wedge\overline{b_1\cdots b_n}=0\\ \text{NaN} & \text{if } \overline{e_1\cdots e_m}=\overline{1_{m-1}\cdots 1_1}\wedge\overline{b_1\cdots b_n}\neq 0\end{cases}$$

$\therefore$ Relative error$=2^{-(n+1)}$, Single-precision float: $(m,n)=(8,23)$, Double-precision float: $(m,n)=(11,52)$

Max denormal$=(\overline{0.1_1\cdots 1_n}\times 10^{1-\overline{1_{m-1}\cdots 1_1}})_2=(1-2^{-n})\times 2^{2-2^{m-1}}$, min positive normal$=(1\times 10^{1-\overline{1_{m-1}\cdots 1_1}})_2=2^{2-2^{m-1}}$

Float Operations

$x\circledast y=fl(fl(x)\star fl(y))$:non-distributive, single-precision implicitly

Computational addition: closure, non-associative, identity, inverse, commutative

Computational subtraction: Addition-analogous, negative skew symmetry $x\circledast y=-(y\circledast x)$

Computational multiplication: Addition-analogoys, non-inverse

Avoided ops: Big terms' sum, Close terms' difference, small terms' denormal product

Kahan's Compensated-Summation Algorithm

```
total,comp=0,0 # 1 float addition, 4 float subtractions vs 1 float addition only per naive summation
for i in range(n):
    comp_x=x_i-comp
    comp_sum=total+comp_x
    comp=(comp_sum-total)-comp_x
    total=comp_sum
return total
```

Error Terminologies

Rounding error:$\because$ finite-precision ops. Truncation error:$\because$ infinite series' clip. Absolute error$=|p-\hat{p}|$. Relative error$=|\frac{p-\hat{p}}{p}|$

C5-7: Matrix Multiplication

$A_{mn}@B_{np}$

```
for i in range(1,m+1): # mnp float mults, m(n-1)p float adds
    for j in range(1,p+1):
        c_{ij}=a_{i1}b_{1j}
        for k in range(2,n+1):c_{ij}=c_{ij}+a_{ik}b_{kj}
return C=(c_{ij})_{mp} # Dense
```

$A_{mn}@U_{nn}$, Upper-Triangular U

```
for i in range(1,n+1): # mn(n+1)/2 float mults, mn(n-1)/2 float adds
    for j in range(1,n+1):
        c_{ij}=a_{i1}b_{1j}
        for k in range(2,j+1):c_{ij}=c_{ij}+a_{ik}b_{kj} # Optimisation
return C=(c_{ij})_{mp} # Dense
```

$L_{nn}@L_{nn}$, Lower-Triangular L

```
for i in range(1,n+1): # \sum_i \sum_j \sum_{k=j}^{i} 1 = n(n+1)(n+2)/6 float mults, (n-1)n(n+1)/6 float adds
    for j in range(1,i+1): # else a_{ij}=0
        c_{ij}=a_{ij}b_{jj} # else b_{kj}=0
        for k in range(j+1,i+1):c_{ij}=c_{ij}+a_{ik}b_{kj}
return C=(c_{ij})_{nn} # Lower-Triangular
```

$T_{nn}@U_{nn}$, Tridiagonal T

```
for i in range(1,n+1): # \sum_i \sum_{j=max{1,i-1}}^{n} \sum_{k=max{1,i-1}}^{min{i+1,j}} 1 = (n+1)(3n-2)/2 float mults, n(n-1) float adds
    for j in range(max{1,i-1},n+1):
        c_{ij}=a_{i max{1,i-1}} b_{max{1,i-1} j}
        for k in range(max{2,i},min{i+2,j+1}):c_{ij}=c_{ij}+a_{ik}b_{kj}
return C=(c_{ij})_{nn} # Upper-Hessenberg
```

$U_{nn}@LH_{nn}$

```
for i in range(1,n+1): # \sum_i \sum_j \sum_{k=max{i,j-1}}^{n} 1=n(n^2+3n-1)/3 float mults, (n-1)n(n+1)/3 float adds
    for j in range(1,n+1):
        c_{ij}=a_{i max{i,j-1}} b_{max{i,j-1} j}
        for k in range(max{i+1,j},n+1):c_{ij}+a_{ik}b_{kj}
return C=(c_{ij})_{nn} # Dense
```

Asymptotic Analysis

Big-O Notation:$(\forall f,g:\mathbb{Z}^+\to\mathbb{R}^+)(\exists M\in\mathbb{R}^+,x_0\in\mathbb{R})(\forall x\geq x_0)(|f(x)|\leq Mg(x))\Leftrightarrow f(x)=O(g(x))$

Big-Ω Notation:$(\forall f,g:\mathbb{Z}^+\to\mathbb{R}^+)(\exists M\in\mathbb{R}^+,x_0\in\mathbb{R})(\forall x\geq x_0)(|f(x)|\geq Mg(x))\Leftrightarrow f(x)=\Omega(g(x))\Leftrightarrow g(x)=O(f(x))$

Big-Θ Notation: $(\forall f,g:\mathbb{Z}^+\to\mathbb{R}^+)(\exists M_1,M_2\in\mathbb{R}^+,x_0\in\mathbb{R})(\forall x\geq x_0)(M_1g(x)\leq|f(x)|\leq M_2g(x))\Leftrightarrow f(x)=\Theta(g(x))$

Strassen-Winograd Algorithm

Necessary condition: $(\exists n\in\mathbb{N}_0)(A,B\in Mat_{2^n\times 2^n}(\mathbb{R}))$ (though zero-paddable)

$S_1,S_3,S_5,S_7=A_{21}+A_{22},A_{11}-A_{21},B_{12}-B_{11},B_{22}-B_{12}$

$S_2,S_6=S_1-A_{11},B_{22}-S_5;S_4,S_8=A_{12}-S_2,S_6-B_{21}$

$M_1,M_2,M_3,M_4,M_5,M_6,M_7=S_2S_6,A_{11}B_{11},A-12B_{21},S_3S_7,S_1S_5,S_4B_{22},A_{22}S_8$

$T_1=M_1+M_2;T_2=T_1+M_4$

$\therefore\begin{bmatrix}C_{11}&C_{12}\\C_{21}&C_{22}\end{bmatrix}=\begin{bmatrix}M_2+M_3&T_1+M_5+M_6\\T_2-M_7&T_2+M_5\end{bmatrix}$

Denote float mults and float adds as N_n^M,N_n^A, then $\begin{cases}N_n^M=7N_{n-1}^M\wedge N_0^M=1\Rightarrow N_n^M=7^n\\N_n^A=15(2^{n-1})^2+7N_{n-1}^A\wedge N_0^A=0\Rightarrow N_n^A=5(7^n-4^n)\end{cases}$

Take $m=$row length=column length$=2^n$, then $N_n^A=O(m^{\log_2 7})<O(m^3)$ naively.

C8-12: Matrix Decomposition

Inplace Gaussian Elimination [Total ops: $\frac{n(n+1)}{2}$ float divs, $\frac{(n-1)n(2n+5)}{6}=O(n^3)$ float mults/subs]

Initial state recoverable in final $\begin{bmatrix}\boxed{a_{11}}&a'_{12}\cdots a'_{1n}|b_1\\a_{21}&\boxed{a_{22}}\cdots a'_{2n}|b_2\\\vdots&\\a_{n1}&a_{n2}\cdots\boxed{a_{nn}}|b_n\end{bmatrix}$ by tracing descendingly in $j\in[1,n-1]:R_i\mapsto R'_i+\frac{a_{ij}}{a_{ii}}R_j$

Gaussian Elimination to REF Algorithm

```
for i in range(1,n): # n(n-1)/2 float divs, (n-1)n(n+1)/3 float mults/subs
    for j in range(i+1,n+1): i=column index, j=row index
        m_{ji} = a_{ji}/a_{ii}
        for k in range(i+1,n+2):a_{jk} = a_{jk} - m_{ji}a_{ik}
return A=(a_{ji})_{n×(n+1)}
```

Backward Substitution Algorithm

```
for i in range(n,0,-1): # n float divs, n(n-1)/2 float mults/subs
    x_i = a_{i,n+1} # b_i initially
    for j in range(i+1,n+1):x_i = x_i - a_{ij}x_j
    x_i = x_i/a_{ii}
return x⃗
```

Pivoting & Label Swapping Algorithm

```
j=i
while j<=n and a_{r_j i}==0:j=j+1
if j==n+1:raise Exception("Singular A")
if j>i:r_i,r_j = r_j,r_i # Only swap row labels, not whole content
```

$\therefore$ Limitation: Numerical Instability$\Rightarrow\neq 0$, hence skipped erroneously.

Thomas Tridiagonal Gaussian Elimination Algorithm

Necessary conditions:$|a_{11}|>|a_{21}|\geq 0,\forall i\in[2,n-1](|a_{ii}|>|a_{i+1,i}|+|a_{i,i+1}|,|a_{nn}|>|a_{n-1,n}|$

Mechanism sketch: Provably 0 pivoting, each next iteration alters exactly the next diagonal entry.

```
for i in range(1,n): # 2n-1 float divs, 3n-3=O(n) float mults/subs
    m_i = a_{i+1,i}/a_{ii};a_{i+1,i+1},b_{i+1} = a_{i+1,i+1} - m a_{i,i+1},b_{i+1} - m b_i
for i in range(n,0,-1):x_i = (b_i - a_{i,i+1}x_{i+1})/a_{ii} if i<n else b_n/a_{nn}
return x⃗
```

Partial Pivoting Algorithm

```
j=i # \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} 1 = (n-1)n/2 float comps
for k in range(i+1,n+1):
    if |a_{r_k i}| > |a_{r_j i}|:j=k # Tiebreak=No swap
if a_{r_j i}==0:raise Exception("Singular A")
if j>i:r_i,r_j = r_j,r_i # Only swap row labels, not whole content
```

$\therefore$ Limitation: Miss scaled-up unstable rows.

Inplace Gaussian Elimination with Partial Pivoting: Initial state recoverable similarly via Backward Substitution.

First identify row-labels via $\max\{|a_{ij}|\}_{i\notin\{\text{Selected Rows}\}}$ ascendingly in $j\in[1,n-1]$.

Uncompromised Scaled Partial Pivoting Algorithm [Inplace Gaussian Elimination irrecoverable$\Leftarrow$var s_{r_k}]

```
for k in range(i,n+1):# (n-1)(n+2)/2 float divs, \sum_{i=1}^{n-1}(\sum_{k=i}^{n} \sum_{j=i+1}^{n} 1 + \sum_{k=i+1}^{n} 1) = n(n-1)(2n+5)/6 float comps
    s_{r_k} = max{|a_{r_k i}|···|a_{r_k n}|}
j,max_value=i,|a_{r_j i}|/s_{r_j}
for k in range(i+1,n+1):
    r=|a_{r_k i}|/s_{r_k}
    if r>max_value:j,max_value=k,r
if a_{r_j i}==0:raise Exception("Singular A")
if j>i:r_i,r_j = r_j,r_i # Only swap row labels, not whole content
```

Compromised Scaled Partial Pivoting Algorithm [Adopted in MA2213]

```
for k in range(1,n+1):s_{r_k} = max{|a_{r_k 1}|···|a_{r_k n}|} # Once for all
# Then each specific ith pivoting is identical to uncompromised variant.
```

$\therefore$ Total ops: $\frac{(n-1)(n+2)}{2}$ float divs, $\sum_{i=1}^{n-1}\sum_{k=i+1}^{n}1=\frac{3n(n-1)}{2}$ float comps. Inplace Gaussian Elimination irrecoverable

Linear System Sensitivity

Forward error$=|\vec{x}-\hat{\vec{x}}|$. Residual error$=|A\vec{x}-A\hat{\vec{x}}|$. $\forall A\in Mat_{n\times n}(\mathbb{R})(norm(A)=\|A\|=\sqrt{\max\{\lambda|\lambda\in Eigen(A^TA)\}})$

Condition number $\kappa(A)=\|A\|\|A^{-1}\|$ ($\because$ provided exist) $\Rightarrow$Instability rises with $\kappa(A)$ ($\because$Vandemonde A of monomials).

$\therefore|A\vec{x}-A\hat{\vec{x}}|\to 0\Rightarrow\frac{\|\vec{x}-\hat{\vec{x}}\|}{\|\vec{x}\|}\leq\frac{\kappa(A)}{1-\kappa(A)\frac{\|A\|-\|\hat{A}\|}{\|A\|}}(\frac{\|A-\hat{A}\|}{\|A\|}+\frac{\|\vec{b}-\hat{\vec{b}}\|}{\|\vec{b}\|})$

Unique PLDU Factorisation [L Unit Lower-Triangular, Time Complexity=$O(n^2)$ if A predecomposed once]

Mechanism: $A\vec{x}=\vec{b}\Leftrightarrow U\vec{x}=L^{-1}\vec{b}\Rightarrow$Solve RHS by Forward Substitution, then LHS by Backward Substitution.

Forward sub eg: $A=\begin{bmatrix}1&1&0&3\\2&1&-1&1\\3&-1&-1&2\\-1&2&3&-1\end{bmatrix}=\begin{bmatrix}0&0&0&1\\0&1&0&0\\1&0&0&0\\0&0&1&0\end{bmatrix}\begin{bmatrix}1&0&0&0\\2/3&1&0&0\\-1/3&1&1&0\\1/3&4/5&1/5&1\end{bmatrix}\begin{bmatrix}3&-1&-1&2\\0&5/3&-1/3&-1/3\\0&0&3&0\\0&0&0&13/5\end{bmatrix}$

$\vec{b}=\begin{bmatrix}4\\-2\\-5\\0\end{bmatrix}\Leftrightarrow L^{-1}P^T\vec{b}=\begin{bmatrix}1&0&0&0\\2/3&1&0&0\\-1/3&1&1&0\\1/3&4/5&1/5&1\end{bmatrix}^{-1}\begin{bmatrix}-5\\-2\\0\\4\end{bmatrix}=\begin{bmatrix}1&0&0&0\\0&1&0&0\\0&1&1&0\\0&4/5&1/5&1\end{bmatrix}^{-1}\begin{bmatrix}-5\\4/3\\-5/3\\17/3\end{bmatrix}=\begin{bmatrix}-5\\4/3\\-3\\26/5\end{bmatrix}$

$\therefore$ A symmetric$\Rightarrow A=LDU=LDL^T$. If further A positive definite, Cholesky decomposition: $\exists L(A=LL^T)$

C13-17: Lagrange & Newton Polynomial Interpolations

Horner's Method for $P_m(x)=\sum_{i=0}^{m} a_i x^i$

$P_m(x)=a_0+x(a_1+\cdots+x(a_{m-1}+xa_m)\cdots)$ ($\because$ Rightfold, m float mults/adds vs 2m float mults, m float adds naively)

Weierstrass Approximation Theorem

$(\forall f\in \mathrm{Cont}[a,b])(\forall \epsilon>0)(\exists \text{polynomial } P(x))(\forall x\in[a,b])(|f(x)-P(x)|<\epsilon)$

Eg: Bernstein expansion $B_n(x)=\sum_{i=0}^{n} f(\frac{i}{n})\binom{n}{i}x^i(1-x)^{n-i}, f\in \mathrm{Cont}[0,1]$

nth-order Lagrange Interpolating Polynomial: $P_n(x)=\sum_{i=0}^{n} f(x_i)L_i(x)=\sum_{i=0}^{n}\frac{f(x_i)}{\prod_{j\neq i}(x_i-x_j)}\prod_{j\neq i}(x-x_j)$

ith Lagrange-basis polynomial $L_i(x)=\prod_{j\neq i}\frac{x-x_j}{x_i-x_j}\Rightarrow L_i(x_k)=\begin{cases}0 & \forall k\neq i\\ 1 & \text{if } k=i\end{cases}$ =Kronecker delta $\phi_i(x_k)=\delta_{ik}$

$\therefore$ Adding 1 node x_{n+1}, $P_{n+1}(x)-P_n(x)=[f(x_{n+1})-P_n(x_{n+1})]\prod_{j=0}^{n}\frac{x-x_j}{x_{n+1}-x_j}\cdots(1)$

kth Divided Difference $f[x_0\cdots x_k]=\sum_{i=0}^{k}\frac{f(x_i)}{\prod_{j\neq i}(x_i-x_j)}$ **i.e. Lagrange Polynomial's weights**

$\therefore$ From (1), nth-order Newton Interpolating Polynomial $P_n(x)=f[x_0]+f[x_0,x_1](x-x_0)+\cdots+f[x_0\cdots x_n](x-x_0)\cdots(x-x_{n-1})$

· Method of Divided Differences (Recursion proven by Ordinary Induction): $f[x_0\cdots x_n]=\frac{f[x_1\cdots x_n]-f[x_0\cdots x_{n-1}]}{x_n-x_0}$

·f invariant on all $(n+1)!$ rearrangements, proven by Strong Induction+Bubble Sort.

Hermite Interpolation [Lagrange Generalisation]

$\because$ Taylor Expansion: $f(x_1)=f(x_0+h)=\sum_{k=0}^{\infty}\frac{f^{(k)}(x_0)}{k!}h^k\Rightarrow f[x_0,x_1]=\sum_{k=1}^{\infty}\frac{f^{(k)}(x_0)}{k!}h^{k-1}$

$\therefore$ Equal spacing: $f[x_0,x_0,x_0]=\lim_{h\to 0}f[x_0,x_0+h,x_0+2h]=\lim_{h\to 0}\frac{\sum_{k=1}^{\infty}\frac{f^{(k)}(x_0+h)}{k!}h^{k-1}-\sum_{k=1}^{\infty}\frac{f^{(k)}(x_0)}{k!}h^{k-1}}{2!h}=\frac{f''(x_0)}{2!}$

Inductively, $f[\overbrace{x_i\cdots x_i}^{j}]=\frac{f^{(j-1)}(x_i)}{(j-1)!}$

Chebyshev Nodes against high-order Runge Phenomenon in scalable $[-1,1]$ **(Runge non-poly eg:** $f=\frac{1}{1+25x^2}$**)**

$\forall i\in[0,n](x_i=\cos(\frac{(i+0.5)\pi}{n+1}))$

$\because$ nth-order Chebyshev Polynomial of 1st kind: $T_n(\cos\theta)=\cos n\theta\Rightarrow T_0=1\wedge T_1=x$

nth-order Chebyshev Polynomial of 2nd kind: $U_n(\cos\theta)=\frac{\sin((n+1)\theta)}{\sin\theta}\Rightarrow U_0=1\wedge U_1=2x$.

$\therefore$ Same Recurrence: $T_{n+1}=2xT_n-T_{n-1}$. Symmetry: $T_n(x)=(-1)^nT_n(-x)$. Chebyshev nodes=Roots-of-unity of $T_n(x)$

· Orthogonality: $\int_{-1}^{1}\frac{T_mT_n}{\sqrt{1-x^2}}dx=\begin{cases}0 & \forall m\neq n\\ \pi & \text{elif } m=0\\ \frac{\pi}{2} & \text{elif } m\neq 0\end{cases}\wedge\int_{-1}^{1}U_mU_n\sqrt{1-x^2}dx=\begin{cases}0 & \forall m\neq n\\ \frac{\pi}{2} & \forall m=n\end{cases}$

· Leading coefficient$=2^{n-1}$ inductively$\Rightarrow(\forall \text{ monic } P_n(x))(\exists\zeta\in[-1,1])(|P_n(\zeta)|\geq\frac{1}{2^{n-1}})$

Interpolation Error

$\because$ Intermediate Value Theorem: $(\forall f\in\mathrm{Cont}[a,b])(\forall c\in\{\min\{f(a),f(b)\},\max\{f(a),f(b)\}\})(\exists x\in[a,b])(f(x)=c)$

$\therefore$ Bolzano's Theorem: $\forall f\in\mathrm{Cont}[a,b](f(a)f(b)<0\Rightarrow\exists x\in[a,b]f(x)=0)$

$\because$ Mean Value Theorem: $(\forall f\in\mathrm{Diff}(a,b)\cap\mathrm{Cont}[a,b])(\exists c\in(a,b))(f'(c)=\frac{f(b)-f(a)}{b-a})$

$\therefore$ Rolle's Theorem: $(\forall f\in\mathrm{Diff}(a,b)\cap\mathrm{Cont}[a,b])(f(a)=f(b)\Rightarrow\exists c\in(a,b)f'(c)=0)$

$\because$ ($\forall$ pairwise-distinct $S=\{x_0\cdots x_n\}$)$(\exists\zeta\in(\min(S),\max(S)))(f[x_0\cdots x_n]=\frac{f^{(n)}(\zeta)}{n!}$ via $\frac{n(n+1)}{2}$ applications of Rolle's Theorem) i.e. Newton Polynomial: $P_n^{(n)}(x)=n!f[x_0\cdots x_n]\Rightarrow$ Find zeros of $g^{(n)}(x)=f^{(n)}(x)-P_n^{(n)}(x)$.

$\therefore$ Adding 1 node x, inductively $\exists\zeta\in(\min(S\cup\{x\}),\max(S\cup\{x\}))(f(x)-P_n(x)=\frac{f^{(n+1)}(\zeta)}{(n+1)!}\prod_{i=0}^{n}(x-x_i))$

$\therefore|f(x)-P_n(x)|\leq\frac{|\max_{\zeta\in(\min(S\cup\{x\}),\max(S\cup\{x\}))}(f^{(n+1)}(\zeta))|}{(n+1)!}|\prod_{i=0}^{n}(x-x_i)|$ i.e. compare $P_n(x),e(x)=\prod(x-x_i)$.

Cubic Spline (n-piecewise) Interpolation [Solve $p^{(k)}(x)$ **in** $S''(x)$**, solve** $S''(x)$ **in divided differences]**

Intended solution $S(x)=\begin{cases}p^{(1)}(x) & \forall x\in[x_0,x_1]\\ \cdots & \\ p^{(n)}(x) & \forall x\in[x_{n-1},x_n]\end{cases}\Rightarrow$Suffice to compute $\forall i\in[0,n]p^{(i)}(x)$

$\therefore$ 4n guiding equations$=\begin{cases}\forall k\in[1,n](p^{(k)}(x_{k-1})=f(x_{k-1})\wedge p^k(x_k)=f(x_k)) & \because \text{Interpolation:}2n \text{ equations}\\ \forall k\in[1,n-1](p'^{(k)}(x_k)=p'^{(k+1)}(x_k)) & \because \text{Continuity:}(n-1) \text{ equations}\\ \forall k\in[1,n-1](p''^{(k)}(x_k)=p''^{(k+1)}(x_k)\Leftrightarrow S''(x_k)=S''(x_{k+1})) & \because \text{Curvature:}(n-1) \text{ equations}\\ \text{Clamped boundary}\Rightarrow p'^{(1)}(x_0)=f'(x_0)\wedge p'^{(n)}(x_n)=f'(x_n) & \because \text{Type constraint:}2 \text{ equations}\\ \text{Natural boundary}\Rightarrow p''^{(1)}(x_0)=p''^{(n)}(x_n)=0 & \because \text{Type constraint:}2 \text{ equations}\end{cases}$

$\therefore p^{(k)}(x)=f[x_{k-1}]+f[x_{k-1},x_k](x-x_{k-1})+(x-x_{k-1})(x-x_k)(\frac{S''(x_k)-S''(x_{k-1})}{6x_k-6x_{k-1}}x+\frac{x_k-2x_{k-1}}{6x_k-6x_{k-1}}S''(x_k)+\frac{2x_k-x_{k-1}}{6x_k-6x_{k-1}}S''(x_{k-1}))$

$\therefore$ Plug continuity: $\forall k\in[1,n-1](6f[x_{k-1},x_k,x_{k+1}]=\frac{x_k-x_{k-1}}{x_{k+1}-x_{k-1}}S''(x_{k-1})+2S''(x_k)+\frac{x_{k+1}-x_k}{x_{k+1}-x_{k-1}}S''(x_{k+1})$
$=\lambda_kS''(x_{k-1})+2S''(x_k)+(1-\lambda_k)S''(x_{k+1}))$.

$\therefore$ If clamped, WLOG $(x_{-1}=x_0,x_n=x_{n+1})\Rightarrow(\lambda_0,\lambda_n)=(0,1)\Rightarrow\begin{cases}6f[x_0,x_0,x_1]=2S''(x_0)+S''(x_1)\\ 6f[x_{n-1},x_n,x_n]=S''(x_{n-1})+2S''(x_n)\end{cases}$

$$\therefore \text{Clamped cond:}\begin{bmatrix}2 & 1-\lambda_0 & 0 & \cdots & 0 & 0 & 0\\ \lambda_1 & 2 & 1-\lambda_1 & \cdots & 0 & 0 & 0\\ 0 & 0 & 0 & \cdots & \lambda_{n-1} & 2 & 1-\lambda_{n-1}\\ 0 & 0 & 0 & \cdots & 0 & \lambda_n & 2\end{bmatrix}\begin{bmatrix}S''(x_0)\\ S''(x_1)\\ \cdots\\ S''(x_{n-1})\\ S''(x_n)\end{bmatrix}\overset{\text{Thomas}}{\widehat{=}}6\begin{bmatrix}f[x_0,x_0,x_1]\\ f[x_0,x_1,x_2]\\ \cdots\\ f[x_{n-2},x_{n-1},x_n]\\ f[x_{n-1},x_n,x_n]\end{bmatrix}$$

$$\therefore \text{Elif natural,}\begin{bmatrix}2 & 1-\lambda_1 & 0 & \cdots & 0 & 0 & 0\\ \lambda_2 & 2 & 1-\lambda_2 & \cdots & 0 & 0 & 0\\ 0 & 0 & 0 & \cdots & \lambda_{n-2} & 2 & 1-\lambda_{n-2}\\ 0 & 0 & 0 & \cdots & 0 & \lambda_{n-1} & 2\end{bmatrix}\begin{bmatrix}S''(x_1)\\ S''(x_2)\\ \cdots\\ S''(x_{n-2})\\ S''(x_{n-1})\end{bmatrix}\overset{\text{Thomas}}{\widehat{=}}6\begin{bmatrix}f[x_0,x_1,x_2]\\ f[x_1,x_2,x_3]\\ \cdots\\ f[x_{n-3},x_{n-2},x_{n-1}]\\ f[x_{n-2},x_{n-1},x_n]\end{bmatrix}$$

$(\forall f\in\mathrm{Diff}[x_0,x_n]\cap\deg(f)=4)(\forall\text{clamped cubic spline } S)(|f(x)-S(x)|\leq\frac{5\max_{x\in[x_0,x_n]}|f^{(4)}(x)|}{384}\max_{i\in[1,n]}\{(x_i-x_{i-1})^4\})$

C18: Least-Squares Approximation/Linear Regression

Degree-1 Regression

$E(a_0,a_1)=\sum_{i=1}^{n}[y_i-(a_0+a_1x_i)]^2$ attains global minimum at $(a_0,a_1)=(\frac{\sum x_i^2\sum y_i-\sum x_i\sum x_iy_i}{n\sum x_i^2-(\sum x_i)^2},\frac{n\sum x_iy_i-\sum x_i\sum y_i}{n\sum x_i^2-(\sum x_i)^2})$

$E(\vec{x})=\|A\vec{x}-\vec{y}\|^2=\vec{x}^TA^TA\vec{x}-2\vec{y}^TA\vec{x}+\vec{y}^T\vec{y}=\sum_j\sum_k x_jx_k\sum_i a_i^{j+k}-2\sum_j x_j\sum_i y_ia_i^j+\sum_i y_i^2\Rightarrow\frac{\partial E(\vec{x})}{\partial\vec{x}}=2A^T(A\vec{x}-\vec{y})$

$\because\frac{\partial}{\partial x_i}\vec{x}^TA^TA\vec{x}=\sum_k(A^TA)_{ik}x_k+\sum_j x_j(A^TA)_{ji}\Rightarrow\frac{\partial}{\partial\vec{x}}\vec{x}^TA^TA\vec{x}=2A^TA\vec{x}.\ |\frac{\partial}{\partial x_i}2\vec{y}^TA\vec{x}=2(\vec{y}^TA)_i\Rightarrow\frac{\partial}{\partial\vec{x}}2\vec{y}^TA\vec{x}=2A^T\vec{y}$

Gram-Schmidt Orthogonalisation Process

Necessary condition: Linearly-independent $\vec{p_0}\cdots\vec{p_m}\in\mathbb{R}^{(n+1)\times 1}$, else SVD first

Mechanism:$\forall i\in[0,m](\vec{v_i}=\vec{p_i}-\sum_{j=0}^{i-1}\frac{\vec{v_j}\cdot\vec{p_i}}{\|\vec{v_j}\|^2}\vec{v_j})$. Optional normalisation $\forall i\in[0,m](\vec{v_i}=\frac{\vec{v_i}}{\|\vec{v_i}\|})$ incorporable therein.

Sufficiency condition: $\mathrm{Span}\{\vec{v_0}\cdots\vec{v_m}\}=\mathrm{Span}\{\vec{p_0}\cdots\vec{p_m}\}$ both of max dim$=m+1$.

Stabler provably-equivalent mechanism:$\forall i\in[0,m]\vec{p_{i,j}}=\begin{cases}\vec{p_i} & \text{if } j=0\\ \vec{p_{i,j-1}}-\frac{\vec{v_{j-1}}\cdot\vec{p_{i,j-1}}}{\|\vec{v_{j-1}}\|^2}\vec{v_{j-1}} & \forall j\in[1,i]\end{cases}$ s.t.$\vec{v_i}=\vec{p_{i,i}}$

Proof Sketch for $i=2$: $\vec{p_{2,2}}=\vec{p_{2,1}}-\frac{\vec{v_1}\cdot\vec{p_{2,1}}}{\|v_1\|^2}\vec{v_1}\wedge\vec{p_0}\cdot\vec{p_{2,1}}=0\Rightarrow\vec{p_0}\cdot\vec{p_{2,2}}=0-\frac{\vec{v_1}\cdot\vec{p_{2,1}}}{\|v_1\|^2}(\vec{p_0}\cdot\vec{v_1})=0$

QR Factorisation for general $A_{(n+1)\times(m+1)}$

Necessary condition: Linearly-independent columns$\Leftrightarrow$ A full column rank, else SVD first

Mechanism:1.Orthonormalise A's column set $\{\vec{p_0}\cdots\vec{p_m}\}\mapsto\{\vec{v_0}\cdots\vec{v_m}\}$ forming Q's columns ($\because Q^TQ=I_{m+1}$)

$$2.\forall i\in[0,m](\vec{p_i}=\sum_{j=0}^{i-1}\frac{\vec{v_j}\cdot\vec{p_i}}{\|\vec{v_j}\|}\vec{v_j}+\|\vec{v_i}\|\vec{v_i})\Rightarrow R_{(m+1)\times(m+1)}=\begin{bmatrix}\|\vec{v_0}\| & \vec{v_0}\cdot\vec{p_1} & \cdots & \vec{v_0}\cdot\vec{p_m}\\ 0 & \|\vec{v_1}\| & \cdots & \vec{v_1}\cdot\vec{p_m}\\ \cdots & & & \\ 0 & 0 & \cdots & \|\vec{v_m}\|\end{bmatrix}$$

Least-Squares Solution $\vec{u}$ **to** $A\vec{x}=\vec{b}$: $\forall\vec{v}\in\mathbb{R}^{(n+1)\times 1}\left\|A\vec{u}-\vec{b}\right\|\leq\left\|A\vec{v}-\vec{b}\right\|$

$\because\left\|A\vec{v}-\vec{b}\right\|$ minimised iff $\vec{b}=\mathrm{proj}_A(\vec{b})\Leftrightarrow A\vec{u}-\vec{b}\perp\mathrm{Col}(A)\Leftrightarrow A^T(A\vec{u}-\vec{b})=\vec{0}$

$\therefore\vec{u}=(A^TA)^{-1}A^T\vec{b}=R^{-1}Q^T\vec{b}$ ($\because$ R invertible) $=\vec{b}-\sum_{i=0}^{m}\frac{\vec{b}\cdot\vec{v_i}}{\|\vec{v_i}\|^2}\vec{v_i}=\|\vec{n}\|$, min dist$=\|\vec{n}\|^2$

Legendre Orthonormal Polynomials $\int_{-1}^{1}P_mP_ndx=0\forall m\neq n\Leftrightarrow P_n=\frac{\frac{d^n}{dx^n}(x^2-1)^n}{2^nn!}$ **via Gram-Schmidt Process**

$P_0=1,P_1=x,P_2=\frac{3}{2}x^2-\frac{1}{2},P_3=\frac{5}{2}x^3-\frac{3}{2}x,P_4=\frac{35}{8}x^4-\frac{15}{4}x^2+\frac{3}{8},P_5=\frac{63}{8}x^5-\frac{35}{4}x^3+\frac{15}{8}$

$P_6=\frac{231x^6}{16}-\frac{315x^4}{16}+\frac{105x^2}{16}-\frac{5}{16},P_7=\frac{429x^7}{16}-\frac{693x^5}{16}+\frac{315x^3}{16}-\frac{35x}{16},P_8=\frac{6435x^8}{128}-\frac{3003x^6}{32}+\frac{3465x^4}{64}-\frac{315x^2}{32}+\frac{35}{128}$

Recall Rodrigue's formula above. Alternatively, $\frac{1}{\sqrt{1-2xt+t^2}}=\sum_{n=0}^{\infty}P_nt^n$ generating function.

C19-21: Numerical Integration [Newton-Cotes Quadratures]

Closed Quadrature:$\int_a^b f(x)dx=\frac{b-a}{n}\sum_{i=0}^{n}f(x_i)\frac{(-1)^{n-i}}{i!(n-1)!}\int_0^n\frac{\prod_{t=0}^{n}(s-t)}{s-i}ds+C\begin{cases}(b-a)^{n+2}f^{(n+1)}(\zeta) & \forall 2\nmid n\\ (b-a)^{n+3}f^{(n+2)}(\zeta) & \forall 2\mid n\end{cases}$

$n=1$(Trapezoidal, just 2 ends): $\frac{b-a}{2}[f(a)+f(b)]-\frac{(b-a)^3f^{(2)}(\zeta)}{12}$ | $n=2$(Simpson): $\frac{b-a}{6}[f(a)+4f(\frac{a+b}{2})+f(b)]-\frac{(b-a)^5f^{(4)}(\zeta)}{2880}$

$n=3$(Simpson 3/8): $\frac{b-a}{8}[f(a)+3f(\frac{2a+b}{3})+3f(\frac{a+2b}{3})+f(b)]-\frac{(b-a)^5f^{(4)}(\zeta)}{6480}$

$n=4$(Boole): $\frac{b-a}{90}[7f(a)+32f(\frac{3a+b}{4})+12f(\frac{a+b}{2})+32f(\frac{a+3b}{4})+7f(b)]-\frac{(b-a)^7f^{(6)}(\zeta)}{1935360}$

$n=5$: $\frac{b-a}{288}[19f(a)+75f(\frac{4a+b}{5})+50f(\frac{3a+2b}{5})+50f(\frac{2a+3b}{5})+75f(\frac{a+4b}{5})+19f(b)]-\frac{11(b-a)^7f^{(6)}(\zeta)}{37800000}$

$n=6$: $\frac{b-a}{840}[41f(a)+216f(\frac{5a+b}{6})+27f(\frac{2a+b}{3})+272f(\frac{a+b}{2})+27f(\frac{a+2b}{3})+216f(\frac{a+5b}{6})+41f(b)]-\frac{(b-a)^9f^{(8)}(\zeta)}{1567641600}$

$n=7$: $\frac{b-a}{17280}[751f(a)+3577f(\frac{6a+b}{7})+1323f(\frac{5a+2b}{7})+2989f(\frac{4a+3b}{7})+2989f(\frac{3a+4b}{7})+1323f(\frac{2a+5b}{7})+3577f(\frac{a+6b}{7})+$
$751f(b)]-\frac{167(b-a)^9f^{(8)}(\zeta)}{426924691200}$

Open Quadrature:$\int_a^b f(x)dx=\frac{b-a}{n+2}\sum_{i=0}^{n}f(x_i)\frac{(-1)^{n-i}}{i!(n-1)!}\int_0^{n+2}\frac{\prod_{t=1}^{n+1}(s-t)}{s-i-1}ds+C\begin{cases}(b-a)^{n+2}f^{(n+1)}(\zeta) & \forall 2\nmid n\\ (b-a)^{n+3}f^{(n+2)}(\zeta) & \forall 2\mid n\end{cases}$

$n=0$(Midpoint, 1 inode): $(b-a)f(\frac{a+b}{2})+\frac{(b-a)^3f^{(2)}(\zeta)}{24}$ | $n=1$: $\frac{b-a}{2}[f(\frac{2a+b}{3})+f(\frac{a+2b}{3})]+\frac{(b-a)^3f^{(2)}(\zeta)}{36}$

$n=2$(Milne): $\frac{b-a}{3}[2f(\frac{3a+b}{4})-f(\frac{a+b}{2})+2f(\frac{a+3b}{4})]+\frac{7(b-a)^5f^{(4)}(\zeta)}{23040}$

$n=3$: $\frac{b-a}{24}[11f(\frac{4a+b}{5})+f(\frac{3a+2b}{5})+f(\frac{2a+3b}{5})+11f(\frac{a+4b}{5})]+\frac{19(b-a)^5f^{(4)}(\zeta)}{90000}$

$n=4$: $\frac{b-a}{20}[11f(\frac{5a+b}{6})-14f(\frac{2a+b}{3})+26f(\frac{a+b}{2})-14f(\frac{a+2b}{3})+11f(\frac{a+5b}{6})]-\frac{41(b-a)^7f^{(6)}(\zeta)}{39191040}$

$n=5$: $\frac{b-a}{1440}[611f(\frac{6a+b}{7})-453f(\frac{5a+2b}{7})+562f(\frac{4a+3b}{7})+562f(\frac{3a+4b}{7})-453f(\frac{2a+5b}{7})+611f(\frac{a+6b}{7})]-\frac{751(b-a)^7f^{(6)}(\zeta)}{1016487360}$

Stirling's Approximation:$\lim_{n\to\infty}\frac{n!}{\sqrt{2n\pi}(n/e)^n}=1$

$\therefore n\uparrow\vee b-a\downarrow\Rightarrow$Accuracy$\uparrow$. Specifically, Stirling's series:$n!\approx\sqrt{2n\pi}(\frac{n}{e})^n(1+\frac{1}{12n}+\frac{1}{288n^2}-\frac{139}{51840n^3}-\frac{571}{2488320n^4}+\cdots)$

3 Composite Newton-Cotes Formulae

Composite Trapezoidal: $\int_a^b f(x)dx=\frac{b-a}{n}[\frac{1}{2}f(x_0)+\sum_{i=1}^{n-1}f(x_i)+\frac{1}{2}f(x_n)]-\frac{(b-a)^3}{12n^2}f^{(2)}(\zeta)$. Necessity: $f\in\mathrm{Diff}(a,b)$

Composite Simpson: $\int_a^b f(x)dx=\frac{b-a}{3n}[f(x_0)+2\sum_{i=1}^{\frac{n}{2}-1}f(x_{2i})+4\sum_{i=1}^{\frac{n}{2}}f(x_{2i-1})+f(x_n)]-\frac{(b-a)^5}{180n^4}f^{(4)}(\zeta)$, $\boxed{2\mid n}$

Composite Midpoint: $\int_a^b f(x)dx=\frac{b-a}{n}\sum_{i=1}^{n}f(\frac{x_{i-1}+x_i}{2})+\frac{(b-a)^3}{24n^2}f^{(2)}(\zeta)$

Node count$=n+1$, degree of accuracy (precision)$=n$ if odd else $n+1$