General conditions which guarantee success of Gauss Elimination without zeros arising on the diagonal are thus valuable: simply applicable sufficient conditions are particularly valuable:

<u>Define</u> $A \in \mathbb{R}^{n\times n}$ is strictly row diagonally dominant (SRDD) if

$$|a_{ii}| > \sum_{\substack{j=1\\ j\neq i}}^{n} |a_{ij}| \quad \text{for every } i$$

<u>Theorem</u> if Gauss Elimination is applied to a SRDD matrix then no zeros arise on the diagonal during the elimination process

<u>Proof</u> SRDD for $i=1 \Rightarrow a_{11} \neq 0$ so 1^{st} column can be eliminated to give a matrix of the form

$$\left[\begin{array}{c|ccc} a_{11} & a_{12} & \cdots & a_{1n} \\ \hline 0 & & & \\ 0 & & B & \\ \vdots & & & \\ 0 & & & \end{array}\right]$$

where the $(i-1)^{th}$ row of B is

$$a_{i2} - \frac{a_{i1}}{a_{11}} a_{12} \ \dots \ a_{ij} - \frac{a_{i1}}{a_{11}} a_{1j} \ \dots \ a_{in} - \frac{a_{i1}}{a_{11}} a_{1n}.$$

Thus B would be SRDD if

$$\left|a_{ii} - \frac{a_{i1}}{a_{11}} a_{1i}\right| > \sum_{\substack{j=2\\ j\neq i}}^{n} \left|a_{ij} - \frac{a_{i1}}{a_{11}} a_{1j}\right| \quad \text{for each } i$$

and because

$$|x|-|y| \leq |x-y| \leq |x|+|y|, \ \forall x,y.$$

a sufficient condition for this is certainly

$$|a_{ii}| - \left|\frac{a_{i1}}{a_{11}}\right| |a_{1i}| > \sum_{\substack{j=2\\ j\neq i}}^{n} |a_{ij}| + \left|\frac{a_{i1}}{a_{11}}\right| |a_{1j}| \quad \text{for each } i \quad — (*)$$

for each i

$$\Leftrightarrow \underbrace{|a_{ii}| - \sum_{\substack{j=2\\ j\neq i}}^{n} |a_{ij}|}_{> |a_{i1}| \text{ by SRDD of } A \text{ in row } i} > \left|\frac{a_{i1}}{a_{11}}\right| \underbrace{\left[|a_{1i}| + \sum_{\substack{j=2\\ j\neq i}}^{n} |a_{1j}|\right]}_{= \sum_{j=2}^{n} |a_{1j}| < |a_{11}| \text{ by SRDD of } A \text{ in row } 1}$$

So $(*)$ is true if

$$|a_{i1}| \geq \left|\frac{a_{i1}}{a_{11}}\right| |a_{11}| \quad \text{each } i$$

which is certainly true.
Hence B is SRDD.
Thus SRDD is preserved under 1^{st} stage of G.E. Inductively apply the same argument to B to complete the proof

<u>Remark</u>: it can similarly be proved that if A is S Column DD (SCDD) then B as defined in the above proof satisfies this property also.

<u>Corollary</u> For an SCDD matrix, all of the multipliers in GE are of absolute value less than 1.

<u>Note</u> that the linear system derived from central difference replacement of $u'' + pu' + qu = f$ is SRDD and SCDD if $p=0$ and $q>0$ for example.
However for $p = 0 = q$ (where G.E. works as we have seen) we have

$$|a_{ii}| = \sum_{\substack{j=1\\ j\neq i}}^{n} |a_{ij}| \quad \text{in}$$

all but the first and last rows

$$\left(\begin{bmatrix} -2 & 1 & & \\ 1 & -2 & 1 & \\ & \ddots & \ddots & 1 \\ & & 1 & -2 \end{bmatrix}\right)$$

and RDD $\left(|a_{ii}| \geq \sum_{\substack{j=1\\ j\neq i}}^{n} |a_{ij}| \text{ all } i\right)$ is not sufficient to get non-zero diagonals

e.g. $\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \xrightarrow{G.E} \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}$

<u>B: Error Analysis</u>

: consider just the model problem $u'' = f$, $u(a) = \alpha$, $u(b) = \beta$

We compute u_j's to satisfy

$$① - \frac{u_{j+1} - 2u_j + u_{j-1}}{h^2} - f(x_j) = 0$$

$$j = 1, \dots, n, \ u_0 = \alpha, \ u_{n+1} = \beta$$

and we have seen for the exact solution $u(x)$

$$\frac{u(x_{j+1}) - 2u(x_j) + u(x_{j-1})}{h^2} = u''(x_j) + \boxed{\frac{h^2}{24}\left[u^{iv}(\xi^+) + u^{iv}(\xi^-)\right]}$$

$$\xi^- \in (x_{j-1}, x_j), \ \xi^+ \in (x_j, x_{j+1})$$

call this τ_j; note
$$\tau_j \leq M h^2$$
for some bounded constant M provided u has 4 derivatives on (a,b)

$$② - \frac{u(x_{j+1}) - 2u(x_j) + u(x_{j-1})}{h^2} - f(x_j) = \tau_j$$

$$j = 1, \dots, n, \ u(x_0) = \alpha, \ u(x_{n+1}) = \beta$$

so if $e_j = u(x_j) - u_j$ we have ② − ①:

$$\frac{e_{j+1} - 2e_j + e_{j-1}}{h^2} = \tau_j$$

$$j = 1, \dots, n, \ e_0 = 0 = e_n$$

<u>Define</u> τ_j is the <u>local truncation error</u>: it is the quantity by which the exact solution of the ODE does not satisfy the finite difference replacement
we let $\tau = \max_j |\tau_j| \leq M h^2$

<u>Lemma</u> (Maximum Principle)

if $\phi_{j+1} - 2\phi_j + \phi_{j-1} \geq 0 \quad j = 1, \dots, n \quad — (*)$

then for each j

$$\phi_j \leq \max\{0, \phi_0, \phi_{n+1}\}$$

i.e. the maximum is taken on the boundary (or is 0)

<u>Proof</u> Assume result not true: then $\exists\, k \in \{1, \dots, n\}$ with $\phi_k > 0$,

$$\phi_k = \max_{j=1,\dots,n} \phi_j, \ \phi_k > \max\{\phi_0, \phi_{n+1}\} \quad — (+)$$

then by $(*)$ since $\phi_{k+1} + \phi_{k-1} \geq 2\phi_k$ we must have $\phi_{k+1} = \phi_{k-1} = \phi_k$ else ϕ_k is not a max!

Applying again this argument with $j = k+1, k-1$ in $(*)$ and continuing

$$\Rightarrow \phi_0 = \phi_1 = \dots = \phi_k = \dots = \phi_{n+1}$$

$\Rightarrow\Leftarrow$

Hence the largest value of the 'mesh function' ϕ_j is taken on the boundary if