Juan Carlos Llamas Núñez 3º DG Mat-Inf

Entrega 3

1.- Sea $A \in M_n$ una matriz de diagonal estrictamente dominante

a) Demostrar que si A se descompone en la forma $A = M - N$, siendo $m_{ii} = a_{ii}$ y $m_{ij} n_{ij} = 0$ para $i,j = 1, \dots n$ entonces el método iterativo asociado a tal descomposición de A está bien definido y es convergente

b) Deducir, a partir de a), resultados de convergencia para los métodos de Jacobi y Gauss-Seidel por bloques.

---

a) Vamos a comenzar probando que la matriz M es de diagonal estrictamente dominante.

Escribiendo la relación $A = M - N$ elemento a elemento se tiene que $a_{ij} = m_{ij} - n_{ij}$ $\forall i,j = 1 \dots n$. De la ecuación $m_{ij} n_{ij} = 0$ se sigue que $m_{ij} = 0$ (y entonces $n_{ij} = -a_{ij}$) o que $n_{ij} = 0$ (y entonces $m_{ij} = a_{ij}$). Esto último es lo que sucede en la diagonal, ya que como $m_{ii} = a_{ii}$ $\forall i = 1, \dots, n$, entonces $n_{ii} = 0$ $\forall i = 1, \dots, n$. Intuitivamente, la relación $m_{ij} n_{ij} = 0$ quiere decir que A se puede expresar como "suma disjunta" (en cada $a_{ij} = m_{ij} - n_{ij}$ uno de los sumandos es 0) de M y $-N$.

Volviendo al problema, hemos deducido que $m_{ij} \in \{0, a_{ij}\}$ $\forall i,j = 1 \dots n$, luego $|m_{ij}| \leq |a_{ij}|$ $\forall i,j = 1 \dots n$ y $|m_{ii}| = |a_{ii}|$ $\forall i = 1 \dots n$.

Entonces, dado $i \in \{1, \dots n\}$

$$|m_{ii}| = |a_{ii}| > \sum_{\substack{j=1 \\ j \neq i}}^{n} |a_{ij}| \geq \sum_{\substack{j=1 \\ j \neq i}}^{n} |m_{ij}|$$ donde la primera desigualdad se obtiene de que A es una matriz de diagonal estrictamente dominante

Esto prueba que M es de diagonal estrictamente dominante y, en particular, inversible. Como M es inversible, entonces el método iterativo asociado a la descomposición $A = M - N$ está bien definido ya que

$$Au = b \Leftrightarrow (M-N)u = b \Leftrightarrow Mu = Nu + b \Leftrightarrow u = Bu + c$$

con $B = M^{-1}N$ y $c = M^{-1}b$, donde se necesita que M sea inversible.

Para ver que el método es convergente hay que probar que $\rho(B) < 1$.

Razonando por reducción al absurdo sea $\lambda \in sp(B)$ con $|\lambda| \geq 1$. Como λ es un autovalor de B se tiene que $\det(B - \lambda Id) = 0$, y multiplicando a ambos lados por $\det(M)$ se tiene que:

$$\det(M) \cdot \det(B - \lambda Id) = \det(M[M^{-1}N - \lambda Id]) = \det(N - \lambda M) = 0 \cdot \det(M) = 0.$$

Sea $C = N - \lambda M$. Como $\det(C) = 0$ entonces C no es inversible y, en particular, no es de diagonal estrictamente dominante. Esto quiere decir que $\exists i_0 \in \{1 \dots n\}$ tal que

$$|c_{i_0 i_0}| \leq \sum_{\substack{j=1 \\ j \neq i_0}}^{n} |c_{i_0 j}|.$$

Hemos visto antes que $n_{ii} = 0 \ \forall i = 1 \dots n$ luego $n_{i_0 i_0} = 0$ y

$c_{i_0 i_0} = -\lambda m_{i_0 i_0} = -\lambda a_{i_0 i_0}$. De esta forma:

$$|c_{i_0 i_0}| = |-\lambda a_{i_0 i_0}| = |\lambda| |a_{i_0 i_0}| \leq \sum_{\substack{j=1 \\ j \neq i_0}}^{n} |c_{i_0 j}| \underset{|\lambda| \geq 1 > 0}{\Leftrightarrow} |a_{i_0 i_0}| \leq \frac{1}{|\lambda|} \sum_{\substack{j=1 \\ j \neq i_0}}^{n} |c_{i_0 j}|$$

Buscamos probar que $\dfrac{|c_{i_0 j}|}{|\lambda|} \leq |a_{i_0 j}| \ \forall j \in \{1 \dots n\} \setminus \{i_0\}$ con lo que tendremos que $|a_{i_0 i_0}| \leq \dfrac{1}{|\lambda|} \sum_{\substack{j=1 \\ j \neq i_0}}^{n} |c_{i_0 j}| \leq \sum_{\substack{j=1 \\ j \neq i_0}}^{n} |a_{i_0 j}|$ y ya habremos terminado, porque esto contradice que A sea de diagonal estrictamente dominante. ¡La contradicción vendrá de haber supuesto que $\rho(B) \geq 1$.

Para probar que $\frac{|c_{i_0 j}|}{|\lambda|} \leq |a_{i_0 j}|$ volvemos a usar la condición $m_{ij} \cdot n_{ij} = 0$ que nos decía que m_{ij} y n_{ij} no podían ser simultáneamente no nulos.

Si $m_{i_0 j} = 0$ entonces $c_{i_0 j} = n_{i_0 j} - \lambda m_{i_0 j} = n_{i_0 j}$ y $a_{i_0 j} = m_{i_0 j} - n_{i_0 j} = -n_{i_0 j}$

luego $\frac{|c_{i_0 j}|}{|\lambda|} = \frac{|n_{i_0 j}|}{|\lambda|} \underset{\substack{\uparrow \\ |\lambda| \geq 1}}{\leq} |n_{i_0 j}| = |-n_{i_0 j}| = |a_{i_0 j}|$.

Si $n_{i_0 j} = 0$ entonces $c_{i_0 j} = n_{i_0 j} - \lambda m_{i_0 j} = -\lambda m_{i_0 j}$ y $a_{i_0 j} = m_{i_0 j} - n_{i_0 j} = m_{i_0 j}$

luego $\frac{|c_{i_0 j}|}{|\lambda|} = \frac{|-\lambda m_{i_0 j}|}{|\lambda|} = |m_{i_0 j}| \leq |m_{i_0 j}| = |a_{i_0 j}|$

Queda probado que $\rho(B) < 1$ y sabemos que esto es una condición necesaria y suficiente para que el método iterativo asociado a la matriz B sea convergente.

b) Sea $A = D - E - F$ la descomposición por bloques de la matriz A con

$$D = \begin{pmatrix} A_{11} & & & 0 \\ & A_{22} & & \\ & & \ddots & \\ 0 & & & A_{pp} \end{pmatrix}, \quad E = \begin{pmatrix} 0 & & & & & \\ -A_{21} & 0 & & & 0 & \\ -A_{31} & -A_{32} & \ddots & & & \\ \vdots & & \ddots & \ddots & & \\ & & & -A_{p-1\,p-2} & 0 & \\ -A_{p1} & \cdots & & & -A_{p\,p-1} & 0 \end{pmatrix}$$ y

$$F = \begin{pmatrix} 0 & -A_{12} & -A_{13} & \cdots & -A_{1,p} \\ & 0 & -A_{23} & \cdots & \vdots \\ & & 0 & \ddots & \vdots \\ & 0 & & \ddots & -A_{p-2,p} \\ & & & 0 & -A_{p-1,p} \\ & & & & 0 \end{pmatrix}.$$

Entonces, para el método de Jacobi por bloques las matrices M y N son $M = D$, $N = E + F$ $(A = M - N)$ y para el método de Gauss-Seidel las matrices M y N son $M = D - E$, $N = F$ $(A = M - N)$.

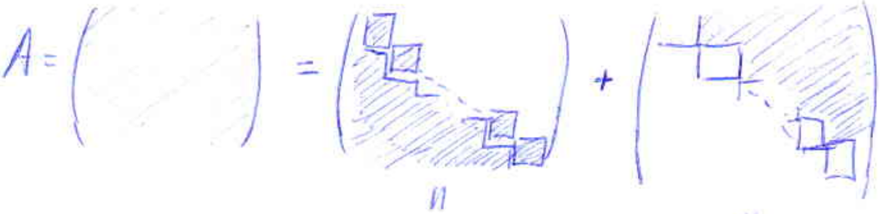
En las hipótesis de que A es una matriz de diagonal estrictamente dominante y según lo probado en a), para garantizar la convergencia de los métodos de Jacobi y Gauss-Seidel por bloques será suficiente probar que $m_{ii} = a_{ii}$ y $m_{ij} n_{ij} = 0$ $\forall i,j = 1 \dots n$ con las matrices M y N definidas antes.

Intuitivamente, $m_{ii} = a_{ii}$ porque en ambas descomposiciones, tanto en el método de Jacobi como en Gauss-Seidel, la matriz D, que contiene la diagonal de A, forma parte de la matriz M y E y F no tienen ningún elemento de la diagonal. Por otro lado, $m_{ij} n_{ij} = 0$ porque cada bloque aparece únicamente en una de las matrices D, E o F. Por tanto, dada una entrada (i,j), el elemento asociado pertenecerá a uno de los bloques. Si ese bloque es el de una de las matrices que forman la descomposición de M entonces el bloque correspondiente de la matriz N será el bloque formado por ceros y $n_{ij} = 0 \Rightarrow n_{ij} \cdot m_{ij} = 0$. Lo análogo sucede cuando el bloque pertenece a la matriz $N \Rightarrow m_{ij} = 0 = m_{ij} n_{ij}$.
Esto ocurre porque en la descomposición, la matriz A se puede descomponer como superposición sin intersección de las matrices M y $-N$.

Jacobi

$D = M \qquad -N = -E - F$

Gauss-Seidel



$D - E = M \qquad -N = -F$

2.- Se considera la matriz tridiagonal

$$A = \begin{pmatrix} 2+\alpha_1 & -1 & & & & \\ -1 & 2+\alpha_2 & -1 & & & \\ & -1 & 2+\alpha_3 & -1 & & \\ & & -1 & \ddots & -1 & \\ & & & -1 & 2+\alpha_{n-1} & -1 \\ & & & & -1 & 2+\alpha_n \end{pmatrix} \quad \text{donde } \alpha_i \geq 0 \quad i = 1, 2, \dots n$$

a) Demostrar por inducción que para cada $K \in \{1, 2, \dots n\}$ se verifica que $\delta_K > \delta_{K-1} > \dots > \delta_2 > \delta_1 > \delta_0 = 1$. (Indicación: Utilizar el apartado a) del Problema 7 de la Hoja 3). Deducir que la matriz A es definida positiva.

b) Para cada $\beta \geq 0$ se considera la descomposición $A = M_\beta - N_\beta$ donde $N_\beta = \mathrm{diag}(\beta - \alpha_1, \beta - \alpha_2, \dots \beta - \alpha_n)$. Encontrar valores del parámetro β para los cuales el método iterativo asociado a esta descomposición $M-N$ de A sea convergente.

---

a) Vamos a probar el apartado a) del Problema 7 de la Hoja 3. En él se nos dice que dada una matriz tridiagonal B con

$$B = \begin{pmatrix} b_1 & c_1 & & 0 \\ a_2 & \ddots & \ddots & \\ & \ddots & \ddots & c_{n-1} \\ 0 & & a_n & b_n \end{pmatrix}$$

y si llamamos δ_K al menor de orden K de B para $k = 1, \dots n$ y por convenio $\delta_0 = 1$, entonces hay que probar $\delta_K = b_K \delta_{K-1} - a_K c_{K-1} \delta_{K-2} \quad \forall K \in \{2 \dots n\}$

Separamos los casos $k=2$ y $k\geq 3$. (para que haya "espacio suficiente" para desarrollar los determinantes y todas las matrices que pongamos tengan dimensión $k\times k$ con $k\geq 1$).

$$\delta_2 = \begin{vmatrix} b_1 & c_1 \\ a_2 & b_2 \end{vmatrix} = b_1 b_2 - c_1 a_2 = b_2 \delta_1 - a_2 c_1 \delta_0$$

Dado $k\geq 3$

$$\delta_k = \det \underbrace{\begin{pmatrix} b_1 & c_1 & & & & \\ a_2 & b_2 & c_2 & & & \\ & \ddots & \ddots & \ddots & & \\ & & a_{k-2} & b_{k-2} & c_{k-2} & \\ & & & a_{k-1} & b_{k-1} & c_{k-1} \\ & & & & a_k & b_k \end{pmatrix}}_{M_k} = b_k \det \underbrace{\begin{pmatrix} b_1 & c_1 & & & \\ a_2 & b_2 & c_2 & & \\ & \ddots & \ddots & \ddots & \\ & & a_{k-2} & b_{k-2} & c_{k-2} \\ & & & a_{k-1} & b_{k-1} \end{pmatrix}}_{M_{k-1}}$$

$$- c_{k-1} \det \underbrace{\begin{pmatrix} b_1 & c_1 & & & \\ a_2 & b_2 & c_2 & & \\ & \ddots & \ddots & \ddots & \\ & & a_{k-2} & b_{k-2} & c_{k-2} \\ & & 0 & 0 & a_k \end{pmatrix}}_{M_{k-1}} = b_k \delta_{k-1} - c_{k-1} a_k \det \underbrace{\begin{pmatrix} b_1 & c_1 & & \\ a_2 & b_2 & \ddots & \\ & \ddots & \ddots & c_{k-3} \\ & & a_{k-2} & b_{k-2} \end{pmatrix}}_{M_{k-2} \text{ (Como poco } M_1)} =$$

$$= b_k \delta_{k-1} - c_{k-1} a_k \delta_{k-2}$$

Aplicando este resultado a la matriz A de nuestro problema, como $a_k = -1 \ \forall k = 2, \dots n$, $c_k = -1 \ \forall k = 1 \dots n-1$ y $b_k = 2 + \alpha_k \ \forall k = 1 \dots n$ obtenemos que $\delta_k = (2+\alpha_k)\delta_{k-1} - (-1)(-1)\delta_{k-2} = (2+\alpha_k)\delta_{k-1} - \delta_{k-2} \quad \forall k \in \{2, \dots n\}$

Probemos por inducción que para cada $k \in \{1 \dots n\}$ se verifica que

$$\delta_k > \delta_{k-1} > \dots > \delta_1 > \delta_0 = 1.$$

El caso base es $k=1$ y se obtiene directamente que

$$\delta_1 = 2 + \alpha_1 \geq 2 > 1 = \delta_0.$$

Supuesto probado para $k \in \{1, \dots n-1\}$ que $\delta_k > \delta_{k-1} > \dots > \delta_1 > \delta_0 = 1$ veamos que

$$\delta_{k+1} > \delta_k > \delta_{k-1} > \dots > \delta_1 > \delta_0 = 1.$$

(por HI se tienen el resto de desigualdades)

Únicamente hace falta probar que $\delta_{k+1} > \delta_k$ así que aplicamos la definición recursiva de δ_{k+1} porque $k+1 \in \{2, \dots n\}$

$$\Longrightarrow \delta_{k+1} = (2+\alpha_k)\delta_k - \delta_{k-1} \underset{\delta_k > \delta_{k-1}}{\overset{HI}{>}} 2\delta_k + \alpha_k \delta_k - \delta_k = \delta_k + \alpha_k \delta_k \underset{}{\overset{\alpha_k, \delta_k \geq 0}{\geq}} \delta_k$$

Hemos probado que $\forall k \in \{1 \dots n\}$ se tiene que

$$\delta_k > \delta_{k-1} > \dots > \delta_1 > \delta_0 = 1.$$

En particular, tomando $k=n$ obtenemos que

$$\delta_n > \delta_{n-1} > \dots > \delta_1 > \delta_0 = 1 > 0.$$

Esto quiere decir que todos los menores principales de A son estrictamente mayores que 0, de donde se deduce que A es definida positiva (es hermítica porque es una matriz real simétrica).

b) El Teorema 5.2 nos dice que si A es una matriz hermítica definida positiva y que se descompone como $A = M - N$ siendo M una matriz inversible, entonces si la matriz (hermítica) $M^* + N$ es definida positiva entonces $\rho(M^{-1}N) < 1$ y el método iterativo asociado a la matriz $B = M^{-1}N$ converge.

Ya hemos probado que A es una matriz hermítica así que veamos para qué valores de β se tiene que M_β es inversible y $M_\beta^* + N_\beta$ es definida positiva.

$$A = M_\beta - N_\beta \Rightarrow M_\beta = A + N_\beta = \begin{pmatrix} 2+\alpha_1 & -1 & & & \\ -1 & 2+\alpha_2 & -1 & & \\ & -1 & \ddots & \ddots & \\ & & \ddots & \ddots & -1 \\ & & & -1 & 2+\alpha_n \end{pmatrix} + \begin{pmatrix} \beta-\alpha_1 & & & \\ & \beta-\alpha_2 & & \\ & & \ddots & \\ & & & \beta-\alpha_n \end{pmatrix} =$$

$$= \begin{pmatrix} 2+\beta & -1 & & & \\ -1 & 2+\beta & -1 & & \\ & -1 & \ddots & \ddots & \\ & & \ddots & \ddots & -1 \\ & & & -1 & 2+\beta \end{pmatrix}.$$ Por tanto, podemos ver M_β como un caso particular de la matriz A para la cual los α_i son todos $\alpha_i = \beta \geq 0$ $\forall i = 1 \dots n$.

Por tanto, todo lo que era cierto para la matriz A lo será para la matriz M_β, en particular que es inversible porque todos sus menores principales (en concreto el de orden n) son positivos (no nulos), luego $\det(M_\beta) \neq 0$.

Además, M_β es una matriz real simétrica y por tanto hermítica, luego $M_\beta^t = M_\beta$. Por tanto, M_β es inversible $\forall \beta \geq 0$ y no tenemos, de momento, ninguna restricción adicional sobre el parámetro.

Veamos para qué valores de β se tiene que $M_\beta^t + N_\beta = M_\beta + N_\beta$ es definida positiva.

$$M_\beta + N_\beta = \begin{pmatrix} 2+\beta & -1 & & \\ -1 & 2+\beta & \ddots & \\ & \ddots & \ddots & -1 \\ & & -1 & 2+\beta \end{pmatrix} + \begin{pmatrix} \beta-\alpha_1 & & & \\ & \beta-\alpha_2 & & \\ & & \ddots & \\ & & & \beta-\alpha_n \end{pmatrix} = \begin{pmatrix} 2+2\beta-\alpha_1 & -1 & & \\ -1 & 2+2\beta-\alpha_2 & \ddots & \\ & \ddots & \ddots & -1 \\ & & -1 & 2+2\beta-\alpha_n \end{pmatrix}$$

Nótese que esta matriz es también un caso particular de la matriz A cuando $2\beta-\alpha_i \geq 0 \ \forall i=1\dots n \iff \beta \geq \frac{\alpha_i}{2} \ \forall i=1\dots n \iff \beta \geq \frac{\max_{i\in\{1\dots n\}} \alpha_i}{2}$.

Concluimos que si $\beta \geq \frac{\max_{i\in\{1\dots n\}}\{\alpha_i\}}{2}$ entonces la matriz $M_\beta + N_\beta$ es definida positiva y el método iterativo asociado a la descomposición $A = M_\beta - N_\beta$ es convergente.