1. Métodos para resolver sistemas lineales de ecuaciones

1. Métodos directos para resolver sistemas de ecuaciones lineales y algunas factorizaciones matriciales

Los métodos directos permiten obtener, teóricamente, una solución exacta del sistema en un número finito de pasos. Decimos teóricamente porque en la práctica se presentan errores de redondeo en los cálculos. El más conocido de este tipo de métodos es la eliminación Gaussiana. Por otra parte, los métodos iterativos permiten obtener una aproximación de la solución del sistema a partir de una sucesión que converge a dicha solución. La conveniencia de usar uno u otro método depende del tamaño y la estructura de la matriz de coeficientes del sistema.

Método de eliminación de Gauss

Esta técnica para resolver sistemas de ecuaciones lineales fue desarrollada por Carl Friedrich Gauss en su tratado *Theoria motus corporum coelestium in sectionibus conicis solem ambientum* (1809) y la idea básica consiste en transformar el sistema dado en uno equivalente que sea triangular el cual es fácil de resolver.

Comenzamos recordando algunos conceptos familiares de álgebra lineal.

Definición 1.1. Una matriz $A = (a_{ij})$ es *triangular superior* si $a_{ij} = 0$ para $i > j$, es decir, si las entradas que están debajo de la diagonal principal son ceros.

Definición 1.2. Una matriz $A = (a_{ij})$ es *triangular inferior* si $a_{ij} = 0$ para $i < j$, es decir, si las entradas que están encima de la diagonal principal son ceros.

Se puede probar fácilmente (ejercicio) que el producto de dos matrices triangulares superiores de orden n es una también una matriz triangular superior y el producto de dos matrices triangulares inferiores de orden n es una también una matriz triangular inferior.

Usaremos la notación $E_{ij}(k)$ para referirnos a la matriz elemental que se obtiene de la matriz identidad I_n de orden n, al multiplicar la i-ésima fila por k y sumarle el resultado a la j-ésima fila (esta operación la denotamos $kF_i + F_j$). Recordamos que toda matriz elemental cuadrada es invertible y su inversa es una matriz elemental del mismo tipo. En particular,

$$E_{ij}^{-1}(k) = E_{ij}(-k).$$

También recordamos que el resultado de realizar una operación elemental (con filas) sobre una matriz A de orden n es equivalente a multiplicar A por la izquierda con la matriz elemental se obtiene de I_n realizado la misma operación elemental. Por ejemplo,

$$A = \begin{pmatrix} 2 & -4 \\ -6 & 3 \end{pmatrix} \xrightarrow{3F_1 + F_2} \begin{pmatrix} 2 & -4 \\ 0 & -9 \end{pmatrix}$$

La matriz resultante de la operación elemental $3F_1 + F_2$ sobre A es equivalente al producto $E_{12}(3)A$.

Se presenta ahora un ejemplo que será ilustrativo para las ideas generales del método de Gauss y la factorización LU.

Ejemplo 1.1. Considere el sistema lineal

$$\begin{aligned} 2x_1 + x_2 - x_3 &= 3 \\ -4x_1 - 3x_2 + 5x_3 &= 0 \\ 2x_1 + 3x_2 + 2x_3 &= 1 \end{aligned}$$

En forma matricial el sistema puede escribirse como

$$\begin{aligned}
&\underbrace{\begin{pmatrix} 2 & 1 & -1 \\ -4 & -3 & 5 \\ 2 & 3 & 2 \end{pmatrix}}_{=A} \underbrace{\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}}_{=\boldsymbol{x}} = \underbrace{\begin{pmatrix} 3 \\ 0 \\ 1 \end{pmatrix}}_{=\boldsymbol{b}} \\
\xrightarrow{2F_1 + F_2 | - F_1 + F_3} &\underbrace{\begin{pmatrix} 2 & 1 & -1 \\ 0 & -1 & 3 \\ 0 & 2 & 3 \end{pmatrix}}_{=E_{13}(-1)E_{12}(2)A} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 3 \\ 6 \\ -2 \end{pmatrix} \\
\xrightarrow{2F_2 + F_3} &\underbrace{\begin{pmatrix} 2 & 1 & -1 \\ 0 & -1 & 3 \\ 0 & 0 & 9 \end{pmatrix}}_{=E_{23}(2)E_{13}(-1)E_{12}(2)A} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 3 \\ 6 \\ 10 \end{pmatrix}
\end{aligned}$$

Note que el último sistema tiene matriz de coeficientes triangular superior, la llamremos U, y se puede resolver fácilmente en forma regresiva, esto es, de la última ecuación se obtiene el valor de $x_3 = 10/9$, el cual se usa para hallar el valor de $x_2 = -8/3$ en la segunda ecuación y con estos dos valores se encuentra $x_1 = 61/18$ de la primera ecuación. Por otra parte, como

$$E_{23}(2)E_{13}(-1)E_{12}(2)A = U,$$

entonces

$$A = E_{12}^{-1}(2)E_{13}^{-1}(-1)E_{23}^{-1}(2)U = E_{12}(-2)E_{13}(1)E_{23}(-2)U = LU,$$

donde $L = E_{12}(-2)E_{13}(1)E_{23}(-2)$ es triangular inferior por ser producto de matrices triangulares inferiores. Más adelante veremos condiciones para la existencia de este tipo de factorizaciones LU.

Se procede ahora a la descripción del método de Gauss para resolver el sistema lineal de ecuaciones

$$\begin{aligned}
a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + \cdots + a_{1n}x_n &= b_1 \\
a_{21}x_1 + a_{22}x_2 + a_{23}x_3 + \cdots + a_{2n}x_n &= b_2 \\
a_{31}x_1 + a_{32}x_2 + a_{33}x_3 + \cdots + a_{3n}x_n &= b_3 \\
&\vdots \\
a_{n1}x_1 + a_{n2}x_2 + a_{n3}x_3 + \cdots + a_{nn}x_n &= b_n,
\end{aligned} \tag{1}$$

el cual puede escribirse en forma matricial como

$$A\boldsymbol{x} = \boldsymbol{b},$$

donde $A = (a_{ij})$ es una matriz $n \times n$ con entradas reales (o complejas) llamada matriz de coeficientes, $\boldsymbol{b} = (b_1, b_2, \ldots, b_n)^{\mathrm{T}}$ y $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)^{\mathrm{T}}$ es el vector de incógnitas. Suponemos que la matriz de coeficientes es no singular y por tanto, el sistema (16) tiene solución única. Inicialmente se usa la primera ecuación para eliminar la primera incógnita x_1 en las $n-1$ ecuaciones restantes. Más concretamente, para eliminar x_1 en la i-ésima ecuación con $i = 2, \ldots, n$, se multiplica la primera ecuación por $-a_{i1}/a_{11}$ y se le suma a la i-ésima ecuación. En este punto se requiere que $a_{11} \neq 0$, si este no fuese el caso, dado que la matriz de coeficientes es no singular, al menos uno de los elementos de la primera columna es no nulo, así se pueden reordenar las ecuaciones de manera que el coeficiente en la fila 1 y columna 1 sea no nulo.

Este primer paso transforma el sistema original en el sistema equivalente

$$\begin{aligned} a_{11}^{(1)}x_1 + a_{12}^{(1)}x_2 + a_{13}^{(1)}x_3 + \cdots + a_{1n}^{(1)}x_n &= b_1^{(1)} \\ a_{22}^{(2)}x_2 + a_{23}^{(2)}x_3 + \cdots + a_{2n}^{(2)}x_n &= b_2^{(2)} \\ a_{32}^{(2)}x_2 + a_{33}^{(2)}x_3 + \cdots + a_{3n}^{(2)}x_n &= b_3^{(2)} \\ &\vdots \\ a_{n2}^{(2)}x_2 + a_{n3}^{(2)}x_3 + \cdots + a_{nn}^{(2)}x_n &= b_n^{(2)} \end{aligned} \tag{2}$$

donde los nuevos coeficientes están dados por

$$\begin{aligned} a_{1j}^{(1)} &:= a_{1j}, \qquad j = 1, \ldots, n \\ a_{ij}^{(2)} &:= a_{ij}^{(1)} - \frac{a_{i1}^{(1)}a_{1j}^{(1)}}{a_{11}^{(1)}}, \qquad i, j = 2, \ldots, n, \end{aligned}$$

y los nuevos términos del lado derecho están dados por

$$\begin{aligned} b_1^{(1)} &:= b_1, \\ b_i^{(2)} &:= b_i^{(1)} - \frac{a_{i1}^{(1)}b_1^{(1)}}{a_{11}^{(1)}}, \qquad i = 2, \ldots, n. \end{aligned}$$

Obsérvese que el determinante de la matriz de coeficientes del sistema (17) es diferente de cero pues es igual al determinante de la matriz de coeficientes del sistema (16) (con la excepción de un posible cambio de signo en caso de que haya cambiado el orden de las filas). Luego, al calcular el determinante del sistema (17) por cofactores (usando la primera columna) se concluye que el determinante de la submatriz de coeficientes que resulta del sistema (17) al eliminar la primera fila y la primera columna es también diferente de cero. El paso siguiente es usar segunda ecuación del sistema (17) para eliminar la incógnita x_2 en las $n-2$ ecuaciones que están debajo de dicha ecuación. Esto se logra multiplicando la segunda ecuación por $-a_{i2}^{(2)}/a_{22}^{(2)}$ y se le suma a la i-ésima ecuación para cada $i = 3, \cdots, n$. Aquí se ha asumido que $a_{22}^{(2)} \neq 0$, si este no fuese el caso, se realiza un intercambio de filas teniendo en cuenta que al menos uno de coeficientes $a_{i2}^{(2)}$ con $i \in \{2, \cdots n\}$ es no nulo por la observación previa. Este segundo paso transforma el sistema (17) en el sistema equivalente

$$\begin{aligned} a_{11}^{(1)}x_1 + a_{12}^{(1)}x_2 + a_{13}^{(1)}x_3 + \cdots + a_{1n}^{(1)}x_n &= b_1^{(1)} \\ a_{22}^{(2)}x_2 + a_{23}^{(2)}x_3 + \cdots + a_{2n}^{(2)}x_n &= b_2^{(2)} \\ a_{33}^{(3)}x_3 + \cdots + a_{3n}^{(3)}x_n &= b_3^{(3)} \\ &\vdots \\ a_{n3}^{(3)}x_3 + \cdots + a_{nn}^{(3)}x_n &= b_n^{(3)} \end{aligned} \tag{3}$$

con

$$\begin{aligned} a_{ij}^{(3)} &:= a_{ij}^{(2)} - \frac{a_{i2}^{(2)}a_{2j}^{(2)}}{a_{22}^{(2)}}, \qquad i,j = 3,\ldots,n, \\ b_i^{(3)} &:= b_i^{(2)} - \frac{a_{i2}^{(2)}b_2^{(2)}}{a_{22}^{(2)}}, \qquad i = 3,\ldots,n. \end{aligned}$$

Repitiendo este proceso, se logra transformar el sistema original en el sistema (triangular) equivalente

$$\begin{aligned} a_{11}^{(1)}x_1 + a_{12}^{(1)}x_2 + a_{13}^{(1)}x_3 + \cdots + a_{1n}^{(1)}x_n &= b_1^{(1)} \\ a_{22}^{(2)}x_2 + a_{23}^{(2)}x_3 + \cdots + a_{2n}^{(2)}x_n &= b_2^{(2)} \\ a_{33}^{(3)}x_3 + \cdots + a_{3n}^{(3)}x_n &= b_3^{(3)} \\ &\vdots \\ a_{nn}^{(n)}x_n &= b_n^{(n)} \end{aligned} \tag{4}$$

donde

$$\begin{aligned} a_{ij}^{(m+1)} &:= a_{ij}^{(m)} - \frac{a_{im}^{(m)}a_{mj}^{(m)}}{a_{mm}^{(m)}}, \qquad i,j = m+1,\ldots,n, \\ b_i^{(m+1)} &:= b_i^{(m)} - \frac{a_{im}^{(m)}b_m^{(m)}}{a_{mm}^{(m)}}, \qquad i = m+1,\ldots,n, \end{aligned}$$

para $m = 1,\ldots,n-1$. Las entradas $a_{mm}^{(m)}$ reciben el nombre de *pivotes* y su elección merece particular atención para evitar pérdida de precisión en los cálculos computacionales. Más concretamente, para controlar la influencia de errores de redondeo, se debe buscar que el cociente $a_{im}^{(m)}/a_{mm}^{(m)}$ sea lo más pequeño posible, lo cual se logra reordenando las filas y columnas de manera que el elemento pivote $a_{mm}^{(m)}$ sea la entrada de mayor valor absoluto en la $(n-m+1)\times(n-m+1)$ submatriz resultante del m-ésimo paso de eliminación. Este proceso se conoce como *pivotado completo*; cuando la entrada de mayor valor absoluto se busca intercambiando solo filas (ecuaciones) o solo columnas, el procedimiento recibe el nombre de *pivotado parcial*. Veamos un ejemplo que ilustra los problemas de inestabilidad que se pueden presentar en una máquina con el proceso de eliminación Gaussiana cuando no se realiza el proceso de pivotado.

Ejemplo 1.2. Considere el sistema

$$\begin{pmatrix} 10^{-10} & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

Suponga que el problema se va a resolver con una máquina cuyos cálculos se desarrollan con una aritmética de punto flotante de 10^{-8}.

Si se multiplica la primera fila por -10^{10} y se le suma a la segunda, obtenemos

$$\begin{pmatrix} 10^{-10} & 1 \\ 0 & 1-10^{10} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1 \\ -10^{10} \end{pmatrix}$$

pero, el número $1-10^{10}$ no será representado en forma exacta; este será redondeado a -10^{10}. En consecuencia, el sistema queda

$$\begin{pmatrix} 10^{-10} & 1 \\ 0 & -10^{10} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1 \\ -10^{10} \end{pmatrix}$$

cuya solución es $(0,1)^{\mathrm{T}}$, pero la solución correcta del sistema original es $(-1,1)^{\mathrm{T}}$.

Veamos ahora que ocurre si hacemos el proceso de eliminación de Gauss pero intercambiando filas, resulta el sistema equivalente

$$\begin{pmatrix} 1 & 1 \\ 10^{-10} & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

Multiplicando la primera fila por -10^{-10} y sumándosela a la segunda, obtenemos

$$\begin{pmatrix} 1 & 1 \\ 0 & 1-10^{-10} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

y con el redondeo queda

$$\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

cuya solución es $(-1, 1)^{\mathrm{T}}$, que corresponde a solución correcta del sistema original.

En este ejemplo, la primera forma de resolver el sistema era un algoritmo inestable, en el sentido de que amplificaba los errores conduciendo a resultados incorrectos, sin embargo, esto se pudo resolver con el proceso de pivotado en el segundo algoritmo.

El sistema triangular (19) se resuelve fácilmente despejando x_n de la última ecuación, luego se reemplaza dicho valor en la penúltima ecuación para encontrar x_{n-1} y así sucesivamente procediendo en forma regresiva hasta hallar x_1 de la primera ecuación. El método de eliminación Gaussiana sin pivotar se describe en el siguiente algoritmo [4].

Algoritmo (Eliminacion Gaussiana sin pivotar)**.**

1. Eliminación progresiva

    for $m = 1, ..., n-1$
        for $i = m+1, ..., n$
            for $j = m+1, ..., n$
                $a_{ij} := a_{ij} - \dfrac{a_{im}a_{mj}}{a_{mm}}$
            end
            $b_i := b_i - \dfrac{a_{im}b_m}{a_{mm}}$
        end
    end

2. Sustitución regresiva

    for $m = n, n-1, \ldots, 1$
        $x_m := b_m$
        for $j = m+1, ..., n$
            $x_m := x_m - a_{mj}x_j$
        end
        $x_m := \dfrac{x_m}{a_{mm}}$
    end

Factorización LU

Se dice que una matriz A tiene una factorización o descomposición LU si existen matrices $L = (l_{ii})$ triangular inferior y $U = (u_{ii})$ matriz triangular superior tales que $A = LU$.

Definición 1.3. Dada una matriz $A \in \mathbb{R}^{n\times n}$, la *submatriz principal* de orden m, para $1 \leq m \leq n$, es la matriz formada por la primeras m filas y m columnas de A.

Definición 1.4. Dada una matriz $A \in \mathbb{R}^{n\times n}$, la *menor principal* de orden m, para $1 \leq m \leq n$, es el determinante de la submatriz principal de order m.

El siguiente resultado da condiciones suficientes y necesarias para la existencia y unicidad de una factorización LU para una matriz cuadrada.

Teorema 1.1. Sea $A \in \mathbb{R}^{n\times n}$. La factorización LU de A con $l_{ii} = 1$ para $i = 1, \ldots, n$ existe y es única si y solo si las submatrices principales A_i de A de orden $i = 1, \ldots, n-1$ son no singulares.

Demostración.
Suponga que las submatrices principales A_i de A de orden $i = 1, \ldots, n-1$ son no singulares y se probará, por inducción sobre n (el orden de A), que la factorización LU de A con $l_{ii} = 1$ para $i = 1, \ldots, n$ existe y es única. Para $n = 1$ el resultado es obvio tomando $L = (1)$ y $U = (a_{11})$. Suponga que el resultado vale para matrices de tamaño $n-1$ y sea $A_n = (a_{ij}) \in \mathbb{R}^{n\times n}$ una matriz cuyas submatrices principales de orden $i = 1, \ldots, n-1$ son no singulares. A_n puede escribirse en bloques como

$$A_n = \begin{pmatrix} A_{n-1} & \boldsymbol{c} \\ \boldsymbol{d}^{\mathrm{T}} & a_{nn} \end{pmatrix}, \tag{5}$$

donde $\boldsymbol{c} = (a_{1n}, a_{2n}, \ldots, a_{(n-1)n})^{\mathrm{T}}$, $\boldsymbol{d}^{\mathrm{T}} = (a_{n1}, a_{n2}, \ldots, a_{n(n-1)})$ y A_{n-1} es la submatriz principal que se obtiene de A_n eliminando la enésima fila y la enésima columna. Puesto que por hipótesis las submatrices principales A_i de A de orden $i = 1, \ldots, n-1$ son no singulares, entonces las submatrices principales de su submatriz A_{n-1} son no singulares. Luego, por hipótesis de inducción sobre A_{n-1}, existe una matriz triangular inferior L_{n-1} con unos en la diagonal principal y una matriz triangular superior U_{n-1} tal que

$$A_{n-1} = L_{n-1}U_{n-1},$$

y esta descomposición es única. Como A_{n-1} es no singular, entonces los factores L_{n-1} y U_{n-1} también los son. Ahora, la expresión de A_n en bloques dada por (20) sugiere que en la descomposición de A_n como un producto L_nU_n, los factores deben elegirse de manera que las primeras $(n-1)$ filas y columnas de L_n y U_n sean las matrices L_{n-1} y U_{n-1} respectivamente. Así, se busca una factorización única de A_n de la forma

$$A_n = L_nU_n = \begin{pmatrix} L_{n-1} & \mathbf{0} \\ \boldsymbol{v}^{\mathrm{T}} & 1 \end{pmatrix} \begin{pmatrix} U_{n-1} & \boldsymbol{w} \\ \mathbf{0}^{\mathrm{T}} & u_{nn} \end{pmatrix} \tag{6}$$

donde el vector columna $\mathbf{0}$, el vector fila $\boldsymbol{v}^{\mathrm{T}}$, vector fila $\mathbf{0}^{\mathrm{T}}$ y el vector columna $\boldsymbol{w}$ tienen todos $(n-1)$ componentes y u_{nn} es el elemento en la última fila y última columna de U_n. El próximo paso es encontrar los vectores $\boldsymbol{v}^{\mathrm{T}}$ y $\boldsymbol{w}$, y el escalar u_{nn} para los cuales la factorización (21) es válida. Con este objetivo en mente, se procede a realizar la multiplicación en bloques (recuerde que este proceso de multiplicación obedece la mismas reglas de la multiplicación matricial ordinaria tomando los bloques como si fueran escalares y que los bloques tengan las dimensiones adecuadas para realizar los productos entre ellos). De esta forma se obtiene

$$A_n = \begin{pmatrix} L_{n-1}U_{n-1} & L_{n-1}\boldsymbol{w} \\ \boldsymbol{v}^{\mathrm{T}}U_{n-1} & \boldsymbol{v}^{\mathrm{T}}\boldsymbol{w} + u_{nn} \end{pmatrix} \tag{7}$$

Igualando los bloques correspondientes en los lados derechos de las ecuaciones (20) y (21), resulta $A_{n-1} = L_{n-1}U_{n-1}$ que era una ecuación ya conocida y también resultan las ecuaciones

$$L_{n-1}\boldsymbol{w} = \boldsymbol{c} \tag{8}$$

$$\boldsymbol{v}^{\mathrm{T}}U_{n-1} = \boldsymbol{d}^{\mathrm{T}} \tag{9}$$

$$\boldsymbol{v}^{\mathrm{T}}\boldsymbol{w} + u_{nn} = a_{nn}. \tag{10}$$

Puesto que L_{n-1} es no singular, se sigue que el sistema lineal (23) determina un valor único para el vector $\boldsymbol{w}$, y como U_{n-1} también es no singular, el vector $\boldsymbol{v}^{\mathrm{T}}$ está univocamente determinado por (24). En consecuencia, de (25), se tiene que u_{nn} está dado en forma única por $u_{nn} = a_{nn} - \boldsymbol{v}^{\mathrm{T}}\boldsymbol{w}$. Esto completa la prueba por inducción.

Resta probar la recíproca, esto es, que si la factorización LU de A con unos en la diagonal principal de L, existe y es única, entonces las submatrices principales A_i de A de orden $i = 1, \ldots, n-1$ son no singulares. Para esto, se consideran dos casos, cuando A es no singular y cuando A es singular. Veamos el primer caso. Si A es no singular entonces

$$0 \neq \det(A) = \det(L)\det(U) = \det(U) = u_{11}u_{22}\cdots u_{nn}$$

Ahora, para cada $i = 1, \ldots, n$, la submatriz principal A_i de A de orden i puede escribirse como

$$A_i = L_iU_i = \begin{pmatrix} L_{i-1} & \mathbf{0} \\ \boldsymbol{l}^{\mathrm{T}} & 1 \end{pmatrix} \begin{pmatrix} U_{i-1} & \boldsymbol{u} \\ \mathbf{0}^{\mathrm{T}} & u_{ii} \end{pmatrix} \tag{11}$$

con $\boldsymbol{l}^{\mathrm{T}} = (l_{i1}, l_{i2}, \ldots, l_{i(i-1)})$ y $\boldsymbol{u} = (u_{1i}, u_{2i}, \ldots, u_{(i-1)i})^{\mathrm{T}}$. Luego,

$$det(A_i) = \det(L_i)\det(U_i) = \det(U_i) = u_{11}u_{22}\cdots u_{ii}$$

y como $u_{11}u_{22}\cdots u_{ii}\cdots u_{nn} \neq 0$, se sigue que $det(A_i) = u_{11}u_{22}\cdots u_{ii} \neq 0$, es decir, cada A_i es no singular.

Ahora suponga que A es singular. Entonces,

$$0 = \det(A) = u_{11}u_{22}\cdots u_{nn},$$

lo cual implica que al menos una de las entradas diagonales de U es cero. Sea $k \in \{1, \ldots, n\}$ el menor índice para el cual $u_{kk} = 0$. Si $k < n$, de la descomposición (26), se deduce que la factorización se puede desarrollar sin problema hasta el paso $(k+1)$. A partir de este paso y debido a que U_k es singular, no se tiene la existencia y unicidad del vector $\boldsymbol{l}^{\mathrm{T}}$, y por ende, tampoco se tendría la existencia y unicidad de la factorización lo que contradice la hipótesis. Luego, $k = n$ y así $u_{kk} \neq 0$ para cada $k \in \{1, \ldots, n-1\}$. Por lo tanto, para cada $k \in \{1, \ldots, n-1\}$ $det(A_k) = \det(L_k)\det(U_k) = \det(U_k) = u_{11}u_{22}\cdots u_{kk} \neq 0$, es decir, todas las submatrices principales A_k son no singulares para $k = 1, \ldots, n-1$.

□

Del anterior teorema se concluye que, si alguna submatriz principal A_i de A con $i \in \{1, \ldots, n-1\}$ es singular, entonces o la factorización LU de A no existe o no es única. Considere, por ejemplo, la matriz (no singular)

$$A = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}$$

Claramente no es posible encontrar un valor no nulo de u_{11} tal que

$$\begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ l_{21} & 1 \end{bmatrix} \begin{bmatrix} u_{11} & u_{12} \\ 0 & u_{22} \end{bmatrix}.$$

Si se busca resolver el sistema $A\boldsymbol{x} = \boldsymbol{b}$, donde A es una matriz no singular cuya factorización LU existe, el sistema se puede escribir como

$$LU\boldsymbol{x} = \boldsymbol{b}.$$

Tomando $\boldsymbol{y} = U\boldsymbol{x}$, el sistema original se ha reducido a resolver dos sistemas lineales cuyas matrices de coeficientes son triangulares y este tipo de sistemas son más fáciles de resolver. Primero se resuelve para $\boldsymbol{y}$ el sistema triangular inferior

$$L\boldsymbol{y} = \boldsymbol{b},$$

usando sustitución progresiva, es decir, si

$$\begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ l_{21} & 1 & 0 & \cdots & 0 \\ l_{31} & l_{32} & 1 & \cdots & \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ l_{n1} & l_{n2} & l_{n3} & & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ \vdots \\ b_n \end{bmatrix},$$

entonces

$$y_1 = b_1 \qquad \text{y} \qquad y_i = b_i - \sum_{k=1}^{i-1} l_{ik} y_k \qquad \text{para} \qquad i = 1, \ldots, n. \tag{12}$$

Como $\boldsymbol{y}$ ya es conocido, el sistema triangular superior $U\boldsymbol{x} = \boldsymbol{y}$, o bien

$$\begin{bmatrix} u_{11} & u_{12} & u_{13} & \cdots & u_{1n} \\ 0 & u_{22} & u_{23} & \cdots & u_{2n} \\ 0 & 0 & u_{33} & \cdots & u_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & u_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{bmatrix},$$

se resuelve por sustitución regresiva

$$x_n = y_n/u_{nn} \qquad \text{y} \qquad x_i = \frac{1}{u_{ii}} \left(y_i - \sum_{k=i+1}^{n} u_{ik} x_k \right), \tag{13}$$

para $i = n-1, n-2, \ldots, 1$.

Cuando se dispone de la factorización LU de la matriz de coeficientes del sistema $A\boldsymbol{x} = \boldsymbol{b}$ y se usan las fórmulas (27)-(28) para resolver los dos sistemas triangulares $L\boldsymbol{y} = \boldsymbol{b}$ y $U\boldsymbol{x} = \boldsymbol{y}$ se requieren n^2 multiplicaciones/divisiones y $n^2 - n$ sumas/sustracciones, lo cual resulta relativamente más económico que la eliminanación Gaussiana en la cual se realizan aproximadamente $n^3/3$ multiplicaciones/divisiones y $n^3/3$ sumas/sustracciones.

Factorización LDU

En la descomposión LU de la matriz A descrita en el Teorema (1.5) la matriz L tiene unos en la diagonal principal mientras que para la matriz U eso no se tiene. Esto puede remediarse reescribiendo U de la siguiente manera

$$\begin{bmatrix} u_{11} & u_{12} & \cdots & u_{1n} \\ 0 & u_{22} & \cdots & u_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \vdots & u_{nn} \end{bmatrix} = \begin{bmatrix} u_{11} & 0 & \cdots & 0 \\ 0 & u_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \vdots & u_{nn} \end{bmatrix} \begin{bmatrix} 1 & u_{12}/u_{11} & \cdots & u_{1n}/u_{11} \\ 0 & 1 & \cdots & u_{2n}/u_{22} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \vdots & 1 \end{bmatrix}$$

Así, se obtiene la factorización $A = LDU$, donde L y U son matrices triangulares inferior y superior respectivamente, con unos en la diagonal principal y $D = diag(u_{11}, u_{22}, \ldots, u_{nn})$ es la matriz diagonal de pivotes.

Factorización de Cholesky

Para el caso A simétrica y definida positiva, se puede obtener una factorización LU con $U = L^{\mathrm{T}}$ teniendo L entradas diagonales positivas como lo indica el siguiente resultado debido a André-Louis Cholesky (1875-1918). La prueba se puede obtener directamente del Teorema de factorización LU teniendo en cuenta que si una matriz A es simétrica y definida positiva, entonces las submatrices principales de A son definidas positivas y por tanto, no singulares. Sin embargo, optaremos por imitar la prueba del teorema anterior para obtener también por inducción las entradas de L.

Teorema 1.2 (Factorización de Cholesky)**.** Si $A \in \mathbb{R}^{n\times n}$ es simétrica y definida positiva, entonces existe una única matriz triangular inferior L con entradas diagonales positivas tal que

$$A = LL^{\mathrm{T}} \tag{14}$$

Las entradas l_{ij} de L vienen dadas por

$$l_{11} = \sqrt{a_{11}},$$

y para $i = 2, \ldots, n$

$$l_{ij} = \frac{a_{ij} - \displaystyle\sum_{m=1}^{j-1} l_{im} l_{jm}}{l_{jj}}, \qquad j = 1, \ldots, i-1 \tag{15}$$

$$l_{ii} = \left(a_{ii} - \sum_{m=1}^{i-1} l_{im}^2 \right)^{1/2} \tag{16}$$

Demostración.
Nótese inicialmente que el hecho de que A sea definida positiva implica que sus entradas diagonales son positivas. En efecto, para cada $i = 1, \ldots, n$, $a_{ii} = \boldsymbol{e}_i^{\mathrm{T}} A \boldsymbol{e}_i > 0$, donde $\boldsymbol{e}_i$ es el vector cuya $i-$ésima entrada es uno y el resto de las entradas son nulas. Con esta observación se procede a la prueba del teorema por inducción sobre n.

Para $n = 1$ el resultado es obvio tomando $L = (\sqrt{a_{11}})$. Suponga que el resultado vale para matrices de tamaño $n-1$ y sea $A = A_n \in \mathbb{R}^{n \times n}$ simétrica y definida positiva. A_n puede escribirse como

$$A_n = \begin{pmatrix} A_{n-1} & \boldsymbol{v} \\ \boldsymbol{v}^{\mathrm{T}} & a_{nn} \end{pmatrix}, \tag{17}$$

donde A_{n-1} es la submatriz principal que se obtiene de A_n eliminando la enésima fila y la enésima columna, $\boldsymbol{v} = (a_{1n}, a_{2n}, \ldots, a_{(n-1)n})^{\mathrm{T}}$. Puesto que A_n es simétrica y definida positiva, la submatriz principal A_{n-1} también lo es. En efecto, la simetría es clara. Sea $\boldsymbol{w} = (w_1, w_2, \ldots, w_{n-1})^{\mathrm{T}}$ un elemento cualquiera no nulo de $\mathbb{R}^{n-1}$ y tome $\boldsymbol{u} = (w_1, w_2, \ldots, w_{n-1}, 0)^{\mathrm{T}}$. Luego, $\boldsymbol{w}^{\mathrm{T}} A_{n-1} \boldsymbol{w} = \boldsymbol{u}^{\mathrm{T}} A_n \boldsymbol{u} > 0$, lo cual implica que A_{n-1} es definida positiva. Aplicando la hipótesis de inducción a A_{n-1}, existe una matriz triangular inferior L_{n-1} con entradas diagonales positivas tal que

$$A_{n-1} = L_{n-1} L_{n-1}^{\mathrm{T}},$$

y las entradas l_{ij} de L_{n-1}, $i, j = 1, 2 \ldots, n-1$ vienen dadas por

$$l_{11} = \sqrt{a_{11}}.$$

Para $i = 2, \ldots, n-1$

$$l_{ij} = \frac{a_{ij} - \sum_{m=1}^{j-1} l_{im} l_{jm}}{l_{jj}}, \qquad j = 1, \ldots, i-1 \tag{18}$$

$$l_{ii} = \left(a_{ii} - \sum_{m=1}^{i-1} l_{im}^2 \right)^{1/2}. \tag{19}$$

Se busca una factorización única de A_n de la forma

$$A_n = \begin{pmatrix} L_{n-1} & \boldsymbol{O}_{(n-1)\times 1} \\ \boldsymbol{y}^{\mathrm{T}} & \beta \end{pmatrix} \begin{pmatrix} L_{n-1}^{\mathrm{T}} & \boldsymbol{y} \\ \boldsymbol{O}_{1\times(n-1)} & \beta \end{pmatrix} \tag{20}$$

Haciendo el producto de las matrices por bloques e igualando con A_n (ver (32)), resulta

$$L_{n-1}\boldsymbol{y} = \boldsymbol{v} \qquad \text{y} \qquad \boldsymbol{y}^{\mathrm{T}}\boldsymbol{y} + \beta^2 = a_{nn}. \tag{21}$$

Puesto que L_{n-1} es no singular, se sigue que $\boldsymbol{y}$ está unívocamente determinada. Además,

$$0 < \det(A_n) = \beta \det(L_{n-1})\beta \det(L_{n-1}^{\mathrm{T}}) = \beta^2(\det(L_{n-1}))^2,$$

y como $det(L_{n-1}) > 0$, se sigue que β es un real no nulo, de hecho, $\beta = \sqrt{a_{nn} - \boldsymbol{y}^{\mathrm{T}}\boldsymbol{y}}$. Tomando

$$L_n := \begin{pmatrix} L_{n-1} & \boldsymbol{O}_{(n-1)\times 1} \\ \boldsymbol{y}^{\mathrm{T}} & \beta \end{pmatrix}$$

se tiene la primera parte del teorema y las fórmulas (30)-(31) para $i = 1, \ldots, n-1$. Las fórmulas completas hasta $i = n$ resultan de (36):

$$\begin{pmatrix} l_{11} & 0 & 0 & \cdots & 0 \\ l_{21} & l_{22} & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ l_{(n-1)1} & l_{(n-1)2} & l_{(n-1)3} & \cdots & l_{(n-1)(n-1)} \end{pmatrix} \begin{pmatrix} l_{n1} \\ l_{n2} \\ \vdots \\ l_{n(n-1)} \end{pmatrix} = \begin{pmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{(n-1)n} \end{pmatrix},$$

y

$$\beta = \sqrt{a_{nn} - \boldsymbol{y}^{\mathrm{T}}\boldsymbol{y}} = \sqrt{a_{nn} - \sum_{m=1}^{n-1} l_{nm}^2}.$$

□

La factorización de Cholesky es muy conveniente para matrices simétricas y definidas positivas, entre otras, por las siguientes razones:

- Para n grande, el número de operaciones requeridas es aproximadamente $n^3/6$, esto es la mitad del número de operaciones requeridas en la factorización LU de una matriz no simétrica.
- No se requiere pivoteo para la estabilidad numérica.

Ejemplo 1.3. Use el método de factorización de Cholesky para resolver el sistema

$$\begin{aligned} x_1 - 2x_2 + 2x_3 &= 4 \\ -2x_1 + 5x_2 - 3x_3 &= -7 \\ 2x_1 - 3x_2 + 6x_3 &= 10 \end{aligned}$$

Solución.

En forma matricial el sistema puede escribirse como

$$\underbrace{\begin{pmatrix} 1 & -2 & 2 \\ -2 & 5 & -3 \\ 2 & -3 & 6 \end{pmatrix}}_{=A} \underbrace{\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}}_{=\boldsymbol{x}} = \underbrace{\begin{pmatrix} 4 \\ -7 \\ 10 \end{pmatrix}}_{=\boldsymbol{b}}$$

La matriz de coeficientes A es simétrica y definida positiva (verifíquelo). Luego, por el Teorema de Descomposición de Cholesky, existe una única matriz triangular inferior L con entradas diagonales positivas tal que

$$A = LL^{\mathrm{T}} \tag{22}$$

Las entradas l_{ij} de L vienen dadas por (ver (34))

$$l_{11} = \sqrt{a_{11}} = 1.$$

Para $i = 2$:

$$\begin{aligned} l_{21} &= \frac{a_{21} - \sum_{m=1}^{0} l_{2m} l_{1m}}{l_{11}} = \frac{-2-0}{1} = -2 \\ l_{22} &= \left(a_{22} - \sum_{m=1}^{1} l_{2m}^2 \right)^{1/2} = \left(5 - (-2)^2\right)^{1/2} = 1. \end{aligned}$$

Para $i = 3$:

$$l_{31} = \frac{a_{31} - \sum_{m=1}^{0} l_{3m} l_{1m}}{l_{11}} = \frac{2-0}{1} = 2$$

$$
\begin{aligned}
l_{32} &= \frac{a_{32} - \displaystyle\sum_{m=1}^{1} l_{3m} l_{2m}}{l_{22}} = \frac{-3-(2)(-2)}{1} = 1 \\
l_{33} &= \left(a_{33} - \sum_{m=1}^{2} l_{3m}^2 \right)^{1/2} = \left(6-(2)^2-(1)^2\right)^{1/2} = 1.
\end{aligned}
$$

Así,

$$
L = \begin{pmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 2 & 1 & 1 \end{pmatrix} \qquad \text{y} \qquad L^T = \begin{pmatrix} 1 & -2 & 2 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}
$$

Luego, el sistema puede escribirse como

$$
\begin{pmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 2 & 1 & 1 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} 4 \\ -7 \\ 10 \end{pmatrix} \tag{23}
$$

con

$$
\boldsymbol{y} = L^T \boldsymbol{x}. \tag{24}
$$

El sistema triangular (38) se puede resolver muy fácilmente en forma progresiva, obteniéndose $\boldsymbol{y} = (4,1,1)^{\mathrm{T}}$. Finalmente, con el vector $\boldsymbol{y}$ encontrado, se resuelve el sistema (39):

$$
\begin{pmatrix} 1 & -2 & 2 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 4 \\ 1 \\ 1 \end{pmatrix},
$$

cuya solución es $\boldsymbol{x} = (2,0,1)^{\mathrm{T}}$.

Ejercicio 1.1.

1. Considere la matriz tridiagonal de tamaño

$$
T = \begin{pmatrix}
\beta_1 & \gamma_1 & 0 & 0 & \cdots & 0 \\
\alpha_1 & \beta_2 & \gamma_2 & 0 & \cdots & 0 \\
0 & \alpha_2 & \beta_3 & \gamma_3 & \cdots & 0 \\
\vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\
0 & 0 & 0 & \alpha_{n-2} & \beta_{n-1} & \gamma_{n-1} \\
0 & 0 & 0 & 0 & \alpha_{n-1} & \beta_n
\end{pmatrix}
$$

a) Si T tiene una factorización LU, verifique que estos factores están dados por

$$L = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ \alpha_1/\pi_1 & 1 & 0 & \cdots & 0 \\ 0 & \alpha_2/\pi_2 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \alpha_{n-1}/\pi_{n-1} & 1 \end{pmatrix}, \quad U = \begin{pmatrix} \pi_1 & \gamma_1 & 0 & \cdots & 0 \\ 0 & \pi_2 & \gamma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \pi_{n-1} & \gamma_{n-1} \\ 0 & 0 & 0 & 0 & \pi_n \end{pmatrix}$$

donde los valores π's están dados por la fórmula de recursión

$$\pi_1 = \beta_1 \qquad \text{and} \qquad \pi_{i+1} = \beta_{i+1} - \frac{\alpha_i \gamma_i}{\pi_i}.$$

b) Escriba un código para la descomposición descrita en el punto anterior.

2. Sea $A \in \mathbb{R}^{n\times n}$. Pruebe que si $A = LL^T$, donde L es una matriz triangular inferior con entradas diagonales positivas (*descomposición de Cholesky*), entonces A es simétrica y definida positiva.

3. Considere la matriz
$$A = \begin{pmatrix} 4 & -1 & 0 \\ -1 & 4 & -1 \\ 0 & -1 & 4 \end{pmatrix}.$$
Verifique que A es simétrica y definida positiva. Luego, halle la descomposición de Cholesky de A.

4. Considere el sistema $A\boldsymbol{u} = \boldsymbol{b}$, dado en (6) tomando una distribución uniforme de cargas, esto es, $\boldsymbol{b} = (1, 1, \ldots, 1)^\top \in \mathbb{R}^n$.

   a) Use la descomposición (7) para probar que la matriz A es simtrica y definida positiva.

   b) Tome $n = 1000$ y resuelva el sistema usando eliminación Gaussiana y el método de Cholesky. Compare los tiempo de cómputo y las soluciones obtenidas.

Bibliografía

[1] K. Atkinson and W. Han. *Theoretical Numerical Analysis*, Springer, New York, Third Edition, 2009.

[2] R. Bellman. *Introducción al Análisis Matricial*, Editorial Reverté, 1965.

[3] R. Horn and C. Johnson. *Matrix Analysis*, Cambridge University Press, 1985.

[4] R. Kress. *Numerical Analysis*, Springer, New York, 1998.

[5] : Kurmanbek, Y. Erlangga, Y. Amanbek. *Explicit inverse of near Toeplitz pentadiagonal matrices related to higher order difference operators.* Results in Applied Mathematics, vol. 11 (2021).

[6] D. Moursund and C. Duris *Elementary Theory and Application of Numerical Analysis*, McGraw-Hill, New York, 1967.

[7] A. Quarteroni, R. Sacco and F. Saleri. *Numerical Mathematics*, Springer, Second Edition,2007.

[8] Y. Saad. *iterative Methods for Sparse Linear Systems*, SIAM, Second Edition, 2003.

[9] J. C. Strikwerda *Finite Difference Schemes and Partial differential Equations*, SIAM, Second Edition, 2004.