Lecturas de Optimización Numérica:

FACTORIZACIONES MATRICIALES, SERIES DE NEUMANN Y DESCOMPOSICIONES ORTOGONALES

Prof. Dr. Fredy Vides
Scientific Computing Innovation Center, UNAH &
Centre for Analysis of Data-Driven Systems
E-mail: fredy.vides@unah.edu.hn

ÍNDICE

OBJETIVOS

1. Estudiar algunas propiedades elementales de las factorizaciones matriciales.
2. Estudiar algunas propiedades elementales de las series matriciales de Neumann.

1. FACTORIZACIONES MATRICIALES

1.1. Factorización LU. Considerando un sistema de ecuaciones lineales representado en forma matricial:

$$Ax = b \tag{1.1}$$

donde $A \in \mathbb{R}^{n\times n}$, $b \in \mathbb{R}^n$ están dados, y $x \in \mathbb{R}^n$ debe determinarse resolviendo el sistema (1.1).

Observación. Como se establece en §6.1 y §6.5 del libro de texto [1], al aplicar un algoritmo de eliminación gaussiana para resolver (1.1) se necesitan $\mathcal{O}(n^3/3)$ operaciones aritméticas.

Notación. En este curso, $\mathbb{GL}_n$, $\mathbb{D}(n)$, $\mathbb{I}(n)$ y $\mathbb{S}(n)$ denotarán los conjuntos de matrices invertibles, diagonales, triangulares inferiores, y triangulares superiores en $\mathbb{R}^{n\times n}$ (en el sentido estudiado en MM 211 vectores y matrices), respectivamente.

Notación: En este curso, al considerar un sistema de ecuaciones $Ax = b$ se asumirá, a menos que se especifique lo contrario, que A,x son compatibles en el producto, y que Ax y b pertenencen al mismo tipo de vectores (matrices).

Propiedad. Si la eliminación gaussiana se puede realizar en el sistema lineal $Ax = b$ sin intercambios de renglones, entonces la matriz A se puede factorizar como el producto de una matriz triangular $L = [l_{jk}] \in \mathbb{I}(n)$ y una matriz triangular superior $U = [u_{jk}] \in \mathbb{S}(n)$, es decir, $A = LU$, donde $l_{jj} = 1$, para $j = 1, \dots, n$.

Consderando el siguiente procedimiento para resolver us sistema $Ax = b$ tal que A puede factorizarse como $A = LU$:

1. Solución del sistema $Ax = b$ por factorización LU:
   a) Definir $y = Ux$ (consideración teórica, no se realiza ninguna operación).
   b) Resolver $Ly = b$ (se requieren $\mathcal{O}(n^2)$ operaciones dado que $L \in \mathbb{I}(n)$).
   c) Resolver $Ux = y$ (se requieren $\mathcal{O}(n^2)$ operaciones dado que $U \in \mathbb{S}(n)$).

Observación. Al aplicar la factorización LU el número de operaciones necesario para resolver el sistema $Ax = b$ se reduce a partir de $\mathcal{O}(n^3/3)$ a $\mathcal{O}(2n^2)$.

2. Matrices Simétricas Positivas Definidas

En la clase de vectores y matrices se estableción que $A \in \mathbb{R}^{n\times n}$ es simétrica si $A^\top = A$.

Definición. Decimos que una matriz simétrica $A \in \mathbb{R}^{n\times n}$ es positiva semi-definida o **SPSD**, si para cada $x \in \mathbb{R}^n$ se cumple que $x^\top Ax \geq 0$. Si $A \in \mathbb{R}^{n\times n}$ es una matriz simétrica que cumple la restricción $x^\top Ax > 0$ para cada $x \in \mathbb{R}^n \setminus \{0\}$, se dice que A es simétrica positiva definida o **SPD**.

Notación. En este curso también utilizaremos la notación $A > 0$ para denotar que la matriz $A \in \mathbb{R}^{n\times n}$ es SPD, las matrices SPD también serán llamadas definidas positivas (**DP**) o positivas definidas (**PD**) en este curso.

Definición. Una **primera submatriz principal** de una matriz A es una matriz A_k de la forma

$$A_k = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1k} \\ a_{21} & a_{22} & \cdots & a_{2k} \\ \vdots & \vdots & & \vdots \\ a_{k1} & a_{k2} & \cdots & a_{kk} \end{bmatrix}$$

para algunas $1 \leq k \leq n$.

Propiedades. Si $A \in \mathbb{R}^{n\times n}$ es una matriz SPD:

1. $A \in \mathbb{GL}_n$;
2. $a_{ii} > 0$, para cada $i = 1, 2, \cdots, n$;
3. $\max_{1\leq k,j\leq n} |a_{kj}| \leq \max_{1\leq i\leq n} |a_{ii}|$;
4. $(a_{ij})^2 < a_{ii}a_{jj}$ para cada $i \neq j$

Propiedad. Dada una matriz $A \in \mathbb{R}^{n\times n}$, las siguientes proposiciones son equivalentes:

1. A es SPD;
2. $\det(A_k) > 0$ para cada una de sus primeras submatrices principales;

3. Se puede realizar eliminación gaussiana sin interacambios de renglones en el sistema $Ax = b$ con todos los elementos pivote positivos.
4. A se puede factorizar como $A = LDL^\top$, donde $L \in \mathbb{I}(n)$, $D \in \mathbb{D}(n)$, con $d_{jj} > 0$ y $l_{jj} = 1$, para cada $1 \leq j \leq n$.
5. A se puede factorizar en la forma $A = LL^\top$, donde $L = [l_{jk}] \in \mathbb{I}(n)$ y $l_{jj} \neq 0$ para cada $1 \leq j \leq n$.

Ejemplo computacional. Consideremos la matriz:

$$U = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{bmatrix}$$

Es claro que la matriz:

$$A = U^\top U = \begin{bmatrix} 1 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix}$$

es SPD dado que para cada $x \in \mathbb{R}^3 \backslash \{0\}$:

$$x^\top Ax = x^\top U^\top Ux = (Ux)^\top (Ux) = (Ux) \cdot (Ux) = \|Ux\|_2^2 > 0$$

dado que $Ux \neq 0$, debido a que $x \neq 0$, y que $U \in \mathbb{GL}_n$ como consecuencia de que $\det(U) = 1 > 0$. Podemos ahora calcular las determinantes de las primeras submatrices principales A_1, A_2, A_3

$$A_1 = \begin{bmatrix} 1 \end{bmatrix}, A_1 = \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix}, A_3 = A$$

con las siguientes secuencias de comandos:

```
>> A1=A(1,1)
A1 =  1
>> A2=A(1:2,1:2)
A2 =

   1  -1
  -1   2

>> A3=A(1:3,1:3)
A3 =

   1  -1   0
  -1   2  -1
   0  -1   2
>> det(A1)
ans =  1
>> det(A2)
ans =  1
>> det(A3)
ans =  1
```

Es posible observar que todos los determinantes son positivos, como lo establecen las propiedades previas.

3. Factorización de Cholesky

Propiedad. Si $A \in \mathbb{R}^{n\times n}$ es una matriz simétrica, tal que se puede aplicar eliminación gaussiana sin intercambios de renglones al sistema $Ax = b$ correspondiente. Entonces A se puede factorizar como $A = LDL^\top$, donde $L \in \mathbb{I}(n)$, $D \in \mathbb{D}(n)$ con $l_{jj} = 1$, para cada $1 \leq j \leq n$.

Algunas Observaciones:

1. Por propiedades previas de matrices SPD, y por la propiedad anterior, tenemos que para cada $A \in \mathbb{R}^{n\times n}$ SPD, $A = LDL^\top \implies$

$$\det(A) = \det(LDL^\top) = \det(L)\det(D)\det(L^\top) = 1 \cdot \det(D) \cdot 1 = \det(D)$$

2. Dado que $L \in \mathbb{GL}_n$ como consecuencia de que $\det(L) = 1 \neq 0$, $A = LDL^\top \implies$

$$L^{-1}A(L^{-1})^\top = D$$

de manera que $D = L^{-1}A(L^{-1})^\top$ es claramente SPD, dado que para cada $x \in \mathbb{R}^n \backslash \{0\}$:

$$x^\top Dx = x^\top L^{-1}A(L^{-1})^\top x = ((L^{-1})^\top x)^\top A((L^{-1})^\top x) > 0$$

dado que A es SPD, y dado que $(L^{-1})^\top x \neq 0$ siempre que $x \neq 0$ debido que $L \in \mathbb{GL}_n$.

Por propiedades de matrices SPD $D = L_D L_D^\top$ para $L_D \in \mathbb{I}(n)$, y por las ecuaciones previas, tenemos entonces que,

$$A = LL_D L_D^\top L^\top = LL_D(LL_D)^\top$$

si definimos $L_A = LL_D$ tenemos entonces que.

$$A = L_A L_A^\top$$

Definición. Dada una matriz $A\mathbb{R}^{n\times n}$ SPD, la factorización $LL^\top = A$ con $L \in \mathbb{I}(n)$, recibe el nombre de factorización de **Cholesky** de A.

4. Series de Neumann

Consideremos una norma matricial (inducida) $\|\cdot\| : \mathbb{R}^{n\times n} \to \mathbb{R}$ arbitraria.
Dada un matriz $X \in \mathbb{R}^{n\times n}$, sea $S_N(X)$ la matriz definida por la expresión:

$$S_N(X) = \sum_{k=0}^{N} X^k$$

Definición. Dada una matriz $A \in \mathbb{R}^{n\times n}$, se define la serie matricial de Neumann $\mathbb{N}(A)$ de A como la serie matricial dada por la expresión:

$$\mathbb{N}(A) = \sum_{k=0}^{\infty} A^k$$

Si suponemos que $\|A\| < 1$ (donde I denota la identidad compatible en la suma con A), tenemos que:

$$(I-A)\left(\sum_{k=0}^{N} A^k\right) = (I-A)(I+A+A^2+\cdots+A^{N-1}+A^N)$$
$$= I - A^{N+1}$$

$\Longrightarrow$

$$\left\|(I-A)\left(\sum_{k=0}^{N} A^k\right) - I\right\| = \|I - A^{N+1} - I\|$$
$$= \|-A^{N+1}\|$$
$$\leq \|A\|^{N+1}\xrightarrow[N\to\infty]{}0$$

$\Longrightarrow S_N(A) \to \mathbb{N}(A) = (I-A)^{-1}$ ($S_N(A)$ converge a $\mathbb{N}(A) = (I-A)^{-1}$ con respecto a la distancia inducida por $\|\cdot\|$). Consideremos ahora una matriz $A \in \mathbb{R}^{n\times n}$ tal que $\|I-A\| < 1$. Por las observaciones previas, tenemos que.

$$S_N(I-A) \to \mathbb{N}(A) = (I-(I-A))^{-1} = A^{-1}$$

Consideremos ahora $A, B \in \mathbb{GL}_n$ tales que $\|I-BA\| < 1$ y $\|I-AB\| < 1$, tenemos entonces que:

$$S_N(I-AB) \to (AB)^{-1} = B^{-1}A^{-1}$$

$\Longrightarrow$

$$BS_N(I-AB) \to BB^{-1}A^{-1} = A^{-1}$$

De forma similar se cumple también que:

$$S_N(I-BA) \to (BA)^{-1} = A^{-1}B^{-1}$$

$\Longrightarrow$

$$S_N(I-BA)B \to A^{-1}B^{-1}B = A^{-1}$$

Las expresiones normadas anteriores nos permiten pensar en B como una aproximación de A^{-1}, dado que en el caso exacto $B = A^{-1}$, se tendría que $\|I-BA\| = \|I-I\| = \|\mathbf{0}\| = 0 < 1$, y de forma similar $\|I-AB\| = \|I-I\| = \|\mathbf{0}\| = 0 < 1$. De manera que si consideramos el sistema de ecuaciones $Ax = b$, $y = BAx = Bb$ sería una primer aproximación de la solución x del sistema, y podemos refinar (mejorar la precisión) de la aproximación $y \approx x$ utilizando el esquema iterativo:

$$S_N(I-BA)y = S_N(I-BA)Bb \to A^{-1}B^{-1}Bb = A^{-1}b = x$$

Ejercicio para el lector: Verificar en detalle que $S_N(I-BA)y \to x$ con respecto a la distancia inducida por la norma vectorial correpondiente (que a su vez induce la norma vectorial en consideración), es decir, demostrar que:

$$\lim_{N\to\infty} \|S_N(I-BA)y - x\| = 0$$

bajo las hipótesis consideradas en los argumentos anteriores.

Ejercicio para el lector: Probar o refutar que para cada $A, B \in \mathbb{R}^{n\times n}$ y cada entero $k \geq 0$:

$$A(I-BA)^k = (I-AB)^k A$$

Ejercicio para el lector: Dadas $A, B \in \mathbb{GL}_n$ y un vector $b \in \mathbb{R}^n$, diseñar un algoritmo iterativo que permita calcular las siguientes expresiones en términos de productos matriz vector, evitando en la medida de lo posible productos matriz-matriz explícitamente calculados:

$$\begin{aligned} y_N &= S_N(I - BA)Bb \\ z_N &= BS_N(I - AB)b \end{aligned}$$

5. Prinicipios de Métodos de Gradiente Conjugado: Un Enfoque Iterativo

Consideremos una matriz $A \in \mathbb{R}^{n\times n}$ arbitraria y un vector arbitrario $b \in \mathbb{R}^n$, y consideremos el funcional cuadrático:

$$q_{A,b}(x) = x^\top Ax - 2x^\top b$$

Propiedad. Si $A \in \mathbb{R}^{n\times n}$ es SPD, resolver el sistema $Ax = b$ equivale a resolver el siguiente problema de programación cuadrática.

$$\min_{x\in\mathbb{R}^n} q_{A,b}(x) \tag{5.1}$$

La solución de los problemas cuadráticos de la forma (5.1) correspondientes a un sistema $Ax = b$ con matriz de coeficientes A SPD, pueden resolverse implementando métodos de gradiente conjugado que producen una secuencia $\{x_k\}_{k\geq 0} \subset \mathbb{R}^n$ tal que $x_k \to x = A^{-1}b$ y donde cada elemento x_k de la sucesión está dado por la expresión:

$$x_k = x_{k-1} + \alpha_{k-1}p_{k-1},\ k \geq 1$$

donde $\alpha_k \in \mathbb{R}$ y $p_k \in \mathbb{R}^n$ se calculan de acuerdo a criterios de optimzación que serán estudiados en detalle más adelante, y $x_0 \in \mathbb{R}^n$ es un elemento que se utiliza para inicializar el esquema iterativo.

6. Prinicipios de Despomposiciones ortogonales y Descomposición en valores singulares

6.1. Descomposición QR.

Iniciamos esta sección recordando que por el teorema de ortogonalizacón de Gram-Schmidt dados $x_1, \ldots, x_m \in \mathbb{R}^n$ linealmente independientes, existen $q_1, \ldots, q_m \in \mathbb{R}^n$ ortonormales tales que:

$$\text{gen}\,\{x_1, \ldots, x_m\} = \text{gen}\,\{q_1, \ldots, q_m\}$$

Además si $X, Q \in \mathbb{R}^{n\times m}$ son las matrices definidas por las expresiones:

$$X = \begin{bmatrix} | & & | \\ x_1 & \cdots & x_m \\ | & & | \end{bmatrix}$$

$$Q = \begin{bmatrix} | & & | \\ q_1 & \cdots & q_m \\ | & & | \end{bmatrix}$$

Se cumple que:

$$\begin{aligned} Q^\top Q &= I_m \\ Q^\top X &= R \\ QR &= X \end{aligned}$$

donde I_m representa la matriz identidad en $\mathbb{R}^{m\times m}$ y $R \in \mathbb{S}(m)$. La descomposición ortogonal $X = QR$ recibe el nombre de descomposición QR de X.

Notación. Una matriz $U \in \mathbb{R}^{n\times m}$ que cumple $U^\top U = I_m$ se denominará **matriz ortogonal** en este curso.

Notación. El conjunto de matrices ortogonales en $\mathbb{R}^{n\times m}$, será denotado por $\mathbb{O}(n,m)$, en el caso de las matrices ortogonales en $\mathbb{R}^{n\times n}$, en algunas ocasiones se escribirá $\mathbb{O}(n)$ en lugar de $\mathbb{O}(n,n)$.

Ejercicio para el lector: Aplicar el teorema de ortogonalización de Gram-Schmidt para verificar que toda matriz $X \in \mathbb{R}^{n\times m}$ con columnas linealmente independientes tiene una factorización QR.

Preguna para el lector: Es posible que toda matriz $X \in \mathbb{R}^{n\times n}$ tenga una factorización QR?

6.1.1. Matrices de Permutación. Un tipo especial de matrices ortogonales en $\mathbb{R}^{n\times n}$, son las matrices de permutación, las cuales se obtienen a partir de la matriz identidad $I \in \mathbb{R}^{n\times n}$ permutando sus columnas (o renglones).

Ejemplos. En $\mathbb{R}^{4\times 4}$ las siguientes matrices son ejemplos de matrices de permutación:

$$Q_1 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix},\ Q_2 = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix},\ Q_3 = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix},\ Q_4 = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

Ejercicio para el lector. Calcular el número de matrices de permutación en $\mathbb{R}^{n\times n}$.

6.2. Descomposición en valores singulares. En este sección consideraremos por primera vez uno de los teoremas fundamentales del análisis matricial numérico.

Propiedad (Teorema fundamental de descomposición en valores singulares *SVD*) Si $A \in \mathbb{R}^{m\times n}$, entonces existen matrices ortogonales $U \in \mathbb{R}^{m\times m}$ y $V \in \mathbb{R}^{n\times n}$ tales que:

$$U^\top AV = \Sigma = \mathrm{diag}(\sigma_1,\ldots,\sigma_p) \in \mathbb{R}^{m\times n},\ \ p = \min\{m,n\}$$

donde $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_p \geq 0$.

Donde $\Sigma = \mathrm{diag}(\sigma_1,\ldots,\sigma_p) \in \mathbb{R}^{m\times n}$ es una matriz en $\mathbb{R}^{m\times n}$ determinada por la expresión:

$$\Sigma_{jk} = \begin{cases} \sigma_j, & j = k \\ 0, & j \neq k \end{cases}, 1 \leq j \leq p$$

REFERENCIAS

[1] R. L. Burden, D. J. Faires, A. M. Burden. (2017). Análisis Numérico. 10a Ed. Cengage Learning Editores.
[2] Golub, G. H., Van Loan C. F (1996). Matrix Computations (3aEd.). The Johns Hopkins University Press.
[3] Quarteroni A., Saleri F., Gervasio P. (2014). Scientific computing with MATLAB and Octave (Textbook).
[4] D. G. Luenberger, Y. Ye. (2016). Linear and Nonlinear Programming. 4a Ed. Springer International Publishing Switzerland.