

# Cálculo numérico

Métodos de factorización para resolver sistemas de ecuaciones lineales

25 de enero 2015

## Factorización LU

Considera el siguiente ejemplo de factorización LU de una matriz en un sistema de punto flotante con  $\beta = 10$  y  $p = 4$ .

$$A_1 = \begin{bmatrix} 0,001 & 2,000 & 3,000 \\ -1,000 & 3,712 & 4,623 \\ -2,000 & 1,072 & 5,643 \end{bmatrix}.$$

Utilizando el primer renglón como pivote podemos calcular

$$A_2 = \begin{bmatrix} 0,001 & 2,000 & 3,000 \\ 0,000 & 2004. & 3005. \\ 0,000 & 4001. & 6006. \end{bmatrix}.$$

Si ahora utilizamos el segundo renglón como pivote tenemos el resultado

$$A_3 = \hat{U} = \begin{bmatrix} 0,001 & 2,000 & 3,000 \\ 0,000 & 2004. & 3005. \\ 0,000 & 0,000 & 5,000 \end{bmatrix}.$$

El valor correcto de  $u_{33}$  es 5.922 ... por lo tanto tenemos que el valor calculado coincide con el correcto en solamente una cifra.

**Ejercicio.** Aplica el método de eliminación gaussiana con pivoteo, en aritmética decimal de 4 dígitos, a la matriz  $A_1$ . Compara el valor de  $\hat{u}_{33}$  con el valor calculado que de obtendría en aritmética decimal de 16 dígitos. ¿Cuántas cifras correctas tiene ahora  $\hat{u}_{33}$ ?

Para el caso general de una matriz de  $n \times n$ , el proceso anterior se representa simbólicamente en la forma

$$A_1 \sim A_2 \sim \dots \sim A_k = U,$$

en donde ' $\sim$ ' indica equivalencia en el sentido de la solución del sistema  $Ax = b$ .

Observa que: a) NO es posible hacer análisis utilizando el símbolo ' $\sim$ '; b) tampoco existe memoria que registre las operaciones que hemos hecho en  $A_1$ . Estas desventajas se pueden remediar introduciendo el concepto de matriz elemental.

No es difícil arribar a los siguientes casos de matrices elementales, útiles en eliminación gaussiana.

1. Tipo 1. Permutaciones de renglones.
2. Tipo 2. Producto de un escalar por un renglón.
3. Tipo 3. Suma de dos renglones, uno de los cuales está multiplicado por un escalar no nulo.

Estas matrices resultan ser casos particulares en el lado izquierdo de la fórmula de Sherman-Morrison. Específicamente, se trate de modificaciones de rango 1 de la matriz identidad:  $E = I - uv^T$ . Ejemplos:

$$\begin{aligned} E_1 &= \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} = I - uu^T, \quad u = e_1 - e_2. \\ E_2 &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & \alpha & 0 \\ 0 & 0 & 1 \end{bmatrix} = I - (1 - \alpha)e_2e_2^T. \\ E_3 &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \alpha & 0 & 1 \end{bmatrix} = I + \alpha e_3e_1^T. \end{aligned}$$

Observa que, descartando el caso trivial  $\alpha = 0$ , todas las matrices elementales son invertibles. En particular, las matrices de tipo 3 tienen dos propiedades importantes

1.

$$E_3^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -\alpha & 0 & 1 \end{bmatrix}$$

2.

$$E_\alpha E_\beta E_\gamma = \begin{bmatrix} 1 & 0 & 0 & 0 \\ \alpha & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ \beta & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \gamma & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ \alpha & 1 & 0 & 0 \\ \beta & 0 & 1 & 0 \\ \gamma & 0 & 0 & 1 \end{bmatrix}.$$

**Ejercicio.** Prueba las propiedades anteriores para el caso  $\mathbb{R}^{n \times n}$ .

Ahora ambos procesos se pueden representar en la forma siguiente:

1. Eliminación gaussiana:  $E_k \cdots E_2 E_1 A = U$ , en donde *todas* las matrices  $E_i$  son matrices elementales de tipo 3. Por lo tanto

$$A = \underbrace{E_1^{-1} E_2^{-1} \cdots E_k^{-1}}_L U.$$

2. Eliminación gaussiana con pivoteo:

$$E_\ell \cdots E_2 E_1 A = U. \quad (1)$$

Observa que algunas de las matrices son tipo 3, mientras que otras son de tipo 1.

**Ejercicio.** Prueba que el producto en (1) se puede reorganizar de tal manera que

$$PA = LU,$$

en donde  $P$  es una matriz de permutación.

## Número de condición

Hasta ahora hemos considerado la siguiente definición intuitiva de número de condición de una matriz no singular  $A$

$$\text{cond}(A) = \|A\| \|A^{-1}\|.$$

Con la herramienta que conocemos, estamos en condiciones de presentar una definición más rigurosa. Consideremos el siguiente sistema de ecuaciones lineales parametrizado

$$(A + tE)x(t) = b + te, \quad x(0) = x, \quad (2)$$

en donde  $A$  es una matriz no singular y  $b \neq 0$ ;  $E$  es una matriz arbitraria;  $e$  es un vector arbitrario. Entonces es claro que  $x(t)$  es diferenciable en una vecindad de  $t = 0$ , de hecho podemos obtener una expresión para  $x(t)$

$$x(t) = x(0) + tx'(0) + \frac{t^2}{2!}x''(0) + \cdots + O(t^{n+1}) \quad (3)$$

Utilizando cualquier norma matricial consistente

$$\|x(t) - x(0)\| = \|tx'(0) + O(t^2)\|$$

La derivada la obtenemos de la definición (2)

$$x'(0) = A^{-1}f - A^{-1}Ex(0)$$

por lo tanto

$$\frac{\|x(t) - x(0)\|}{\|x(0)\|} \leq \frac{|t|\|A^{-1}(f - Ex(0))\|}{\|x(0)\|} + O(t^2) \quad (4)$$

$$\leq \underbrace{\|A\| \|A^{-1}\|}_{\text{cond}(A)} \left[ |t| \frac{\|e\|}{\|b\|} + |t| \frac{\|E\|}{\|A\|} \right] + O(t^2). \quad (5)$$

**Ejercicio.** Obtener (5) a partir de (4).

Ejemplo. Matrices bien condicionadas.

Supongamos que  $|\epsilon| \ll 1$

$$A = \begin{bmatrix} \epsilon & -1 \\ 1 & 1 \end{bmatrix} e, \quad A^{-1} = \frac{1}{1+\epsilon} \begin{bmatrix} 1 & -1 \\ 1 & \epsilon \end{bmatrix}.$$

$$\|A\|_1 = 2, \quad \|A\|_\infty = 2, \quad \|A\|_F = \sqrt{3 + \epsilon^2}.$$

$$\|A^{-1}\|_1 = \frac{2}{1+\epsilon}, \quad \|A^{-1}\|_\infty = \frac{2}{1+\epsilon}, \quad \|A^{-1}\|_F = \frac{\sqrt{3 + \epsilon^2}}{1+\epsilon}.$$

Supongamos que  $\epsilon \rightarrow 0$ , entonces

$$\text{cond}_1(A) = 4, \quad \text{cond}_\infty(A) = 4, \quad \text{cond}_F(A) = 3.$$

Ejemplo. Matrices mal condicionadas.

$$B = \begin{bmatrix} 1+\epsilon & 1 \\ 1 & 1 \end{bmatrix}, \quad B^{-1} = \frac{1}{\epsilon} \begin{bmatrix} 1 & -1 \\ -1 & 1+\epsilon \end{bmatrix}.$$

$$\|B\|_1 \approx 2, \quad \|B\|_\infty \approx 2, \quad \|B\|_F = \sqrt{3 + (1+\epsilon)^2} \approx 2.$$

$$\|B^{-1}\|_1 = \frac{2}{\epsilon}, \quad \|B^{-1}\|_\infty = \frac{2}{\epsilon}, \quad \|B^{-1}\|_F = \frac{\sqrt{3 + \epsilon^2}}{\epsilon}.$$

Supongamos que  $\epsilon \rightarrow 0$ , entonces

$$\text{cond}_1(B) \rightarrow \infty, \quad \text{cond}_\infty(B) \rightarrow \infty, \quad \text{cond}_F(B) \rightarrow \infty.$$

Ejemplo. Suma de  $n$  números  $s_n = x_1 + x_2 + \cdots + x_n$

$$\text{cond}(s_n) = \frac{|x_1| + |x_2| + \cdots + |x_n|}{|x_1 + x_2 + \cdots + x_n|}$$

$\text{cond}(\cdot)$  es llamado el número de condición del problema.

Conclusión preliminar

Números de condición *grandes*  $\approx$  singularidad.

## Análisis de estabilidad

Supongamos que resolvemos numéricamente el sistema de ecuaciones lineales

$$Ax = b, \quad b \neq 0, \quad (6)$$

en donde  $A$  es una matriz no singular en  $\mathbb{R}^{n \times n}$ ;  $x, b \in \mathbb{R}^n$ .

El problema que queremos resolver es cuantificar el efecto del error por redondeo en la solución calculada por el método numérico.

Antes de continuar es útil es considerar el caso ideal en el que el único error se debe al redondeo de los datos de entrada  $A, b$ . Es decir que la computadora no introduce error a partir de que los datos se redondean.

Es importante notar que el argumento  $t \rightarrow 0$  es poco satisfactorio para modelar el efecto del redondeo. Por otra parte, suponer la invertibilidad de  $A + E$  es plausible pero no es un argumento matemáticamente sólido. Sin embargo podemos anticipar que

$$\frac{\|x - \hat{x}\|_\infty}{\|x\|_\infty} \lesssim 2\epsilon_M \text{cond}_\infty(A).$$

**Ejercicio.** Verificar la desigualdad anterior.

En los siguientes párrafos vamos a construir una prueba rigurosa del resultado anterior. En primer lugar vamos a probar que  $A - E$  es invertible bajo condiciones poco restrictivas.

**Lema 1.** Supongamos que  $E$  es una matriz de perturbación tal que  $\|E\|_p < 1$ . Entonces:

1.  $I - E$  es invertible.
- 2.

$$(I - E)^{-1} = \sum_{k=0}^{\infty} E^k.$$

- 3.

$$\|(I - E)^{-1}\|_p \leq \frac{1}{1 - \|E\|_p}.$$

**Prueba.**

1. Supongamos que  $I - E$  no es invertible, entonces existe un vector  $y \neq 0$  tal que  $(I - E)y = 0$  y por lo tanto

$$y = Ey, \quad \|y\|_p \leq \|E\|_p \|y\|_p$$

de donde es inmediato concluir la contradicción  $\|E\|_p \geq 1$ .

2. Consideremos la identidad

$$(I + E + \cdots + E^k)(I - E) = I - E^{k+1}$$

y la hipótesis  $\|E\|_p < 1$ . Entonces tenemos que

$$\lim_{k \rightarrow \infty} (I + E + \cdots + E^k)(I - E) = \lim_{k \rightarrow \infty} (I - E^{k+1}) = I$$

de donde, utilizando la invertibilidad de  $(I - E)$ , obtenemos el resultado pedido.

3. Consideremos la serie formada por las potencias de  $\|E\|_p$ , es decir

$$\sum_{k=0}^{\infty} \|E\|_p^k,$$

de la hipótesis  $\|E\|_p < 1$  se sigue que

$$\sum_{k=0}^{\infty} \|E\|_p^k = \frac{1}{1 - \|E\|_p}$$

de donde es fácil obtener el resultado pedido.  $\square$

El resultado anterior se puede extender al caso general; es decir  $A$  una matriz invertible a la que se agrega una matriz de perturbación  $E$ .

También podemos añadir a nuestro repertorio la siguiente identidad sobre dos matrices invertibles  $A$  y  $B$ , las cuales difieren en  $E$ .

$$B^{-1} = A^{-1} - B^{-1}(B - A)A^{-1}. \quad (7)$$

**Ejercicio.** Verificar e interpretar la igualdad anterior.

**Lema 2.** Si  $A$  es invertible y  $\|A^{-1}E\|_p < 1$  entonces  $A + E$  es invertible y además

$$\|(A + E)^{-1} - A^{-1}\|_p \leq \frac{\|E\|_p \|A^{-1}\|_p^2}{1 - \|A^{-1}E\|_p}.$$

**Prueba.** El primer resultado es trivial de probar. El segundo es consecuencia inmediata de la identidad (7).  $\square$

En la siguiente proposición consideramos el sistema  $Ax = b$  y las correspondientes perturbaciones para  $A$  y para  $b$ .

**Lema 3.** Supongamos que  $A$  es invertible y que las perturbaciones  $E$  y  $e$  satisfacen

$$\|E\|_p \leq \epsilon \|A\|_p, \quad \|e\|_p \leq \epsilon \|b\|_p.$$

Si  $r = \epsilon \text{cond}(A) < 1$ , entonces  $A + E$  es no singular y además

$$\frac{\|y\|_p}{\|x\|_p} \leq \frac{1 + r}{1 - r},$$

en donde  $y$ , por el primer resultado, es la solución del sistema perturbado

$$(A + E)y = b + e.$$

**Prueba.**  $\|A^{-1}E\|_p \leq \|A^{-1}\|_p \|E\|_p \leq \epsilon \|A^{-1}\|_p \|A\|_p < 1$ ; utilizando el Lema 2 tenemos que  $A + E$  es invertible.

Para probar el segundo resultado partimos de la igualdad

$$(A + E)y = b + e = Ax + e$$

si multiplicamos ambos lados por  $A^{-1}$

$$(I + A^{-1}E)y = x + A^{-1}e.$$

Notar que hemos probado que  $\|A^{-1}E\|_p < 1$ , por lo tanto  $(I + A^{-1}E)$  es invertible y podemos obtener

$$y = (I + A^{-1}E)^{-1}(x + A^{-1}e)$$

calculando la norma en ambos lados

$$\begin{aligned} \|y\|_p &= \|(I + A^{-1}E)^{-1}(x + A^{-1}e)\|_p \\ &\leq \|(I + A^{-1}E)^{-1}\|_p (\|x\|_p + \|A^{-1}e\|_p) \end{aligned}$$

ahora utilizamos el Lema 1 para acotar el término  $\|(I + A^{-1}E)^{-1}\|$ , es decir

$$\|(I + A^{-1}E)^{-1}\| \leq \frac{1}{1 - \|A^{-1}E\|_p} \leq \frac{1}{1 - r}$$

para el término  $\|A^{-1}e\|$  procedemos como sigue

$$\|A^{-1}e\|_p \leq \|A^{-1}\|_p \|e\|_p \leq \epsilon \|A^{-1}\|_p \|b\|_p$$

De donde es fácil obtener el resultado pedido.  $\square$

Por medio de la teoría anterior podemos formular una cota rigurosa para el error relativo en el siguiente

**Teorema 1.** Si se cumplen las condiciones del Lema 3 entonces

$$\frac{\|y - x\|_p}{\|x\|_p} \leq \frac{2\epsilon}{1 - r} \text{cond}_p(A)$$

**Prueba.** Partimos de

$$y - x = A^{-1}e - A^{-1}Ey$$

calculando la norma en ambos lados de la igualdad anterior

$$\|y - x\|_p = \|A^{-1}e - A^{-1}Ey\|_p \leq \epsilon \|A^{-1}\|_p \|b\|_p + \epsilon \|A^{-1}\|_p \|A\|_p \|y\|_p$$

de donde

$$\begin{aligned}\frac{\|y - x\|_p}{\|x\|_p} &\leq \epsilon \operatorname{cond}_p(A) \frac{\|b\|_p}{\|A\|_p \|x\|_p} + \epsilon \operatorname{cond}_p(A) \frac{\|y\|_p}{\|x\|_p} \\ &\leq \epsilon \operatorname{cond}_p(A) \left(1 + \frac{1+r}{1-r}\right) = \frac{2\epsilon}{1-r} \operatorname{cond}_p(A).\end{aligned}$$

La prueba está completa para cualquier norma- $p$ .  $\square$

En el caso particular  $p = \infty$ ,  $r = 1/2$  y  $\epsilon = \epsilon_M$  tenemos que

$$\boxed{\frac{\|x - \hat{x}\|_\infty}{\|x\|_\infty} \leq 4\epsilon_M \operatorname{cond}_\infty(A)}.$$

**Boceto de prueba.** Se trata de rastrear el error por redondeo en cada paso que hace EGp, proyectando dicho error en la matriz original  $A$  en la forma  $A + E$ .

En el paso  $k = 1$  obtenemos los multiplicadores

$$m_{i1} = \operatorname{float}(a_{i1}/a_{11}) = \frac{a_{i1}(1 + \epsilon_{i1})}{a_{11}}, \quad i = 2, \dots, n,$$

en donde  $|\epsilon_{i1}| \leq \epsilon_M$ . Ahora expresamos los multiplicadores en función de valores perturbados de  $a_{i1}$

$$\bar{a}_{i1} = a_{i1}(1 + \epsilon_{i1}) \tag{8}$$

entonces tenemos que

$$m_{i1} = \frac{\bar{a}_{i1}}{a_{11}}, \quad i = 2, \dots, n.$$

Ahora eliminamos los elementos debajo del elemento pivote  $a_{11}$

$$\begin{aligned}a'_{ij} &= \operatorname{float}(a_{ij} - m_{i1}a_{1j}) \\ &= [a_{ij} - m_{i1}a_{1j}(1 + \epsilon_{ij})](1 + \eta_{ij}) \\ &= a_{ij} - m_{i1}a_{1j} + a_{ij}\eta_{ij} - m_{i1}a_{1j}(\epsilon_{ij} + \eta_{ij} + \epsilon_{ij}\eta_{ij}).\end{aligned}$$

Si definimos

$$\bar{a}_{ij} = a_{ij} + a_{ij}\eta_{ij} - m_{i1}a_{1j}(\epsilon_{ij} + \eta_{ij} + \epsilon_{ij}\eta_{ij}), \tag{9}$$

tenemos que

$$a'_{ij} = \bar{a}_{ij} - m_{i1}a_{1j}.$$

Entonces la matriz

$$\left[ \begin{array}{c|ccc} a_{11} & a_{12} & \cdots & a_{1n} \\ \hline m_{21} & a'_{22} & \cdots & a'_{2n} \\ \vdots & \vdots & & \vdots \\ m_{n1} & a'_{n2} & \cdots & a'_{nn} \end{array} \right]$$



que resulta de hacer EG *con error por redondeo* en  $A$ , es la misma que resulta de hacer EG *sin error por redondeo* en

$$\left[ \begin{array}{c|ccc} a_{11} & a_{12} & \cdots & a_{1n} \\ \hline \bar{a}_{21} & \bar{a}_{22} & \cdots & \bar{a}_{2n} \\ \vdots & \vdots & & \vdots \\ \bar{a}_{n1} & \bar{a}_{n2} & \cdots & \bar{a}_{nn} \end{array} \right].$$

Observa con cuidado que, de las expresiones (8) y (9), se puede obtener inmediatamente la primera matriz de error

$$A + E_1 = \left[ \begin{array}{c|ccc} a_{11} & a_{12} & \cdots & a_{1n} \\ \hline a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{array} \right] + \left[ \begin{array}{c|ccc} 0 & 0 & \cdots & 0 \\ \hline a_{21}\epsilon_{21} & & & \\ \vdots & a_{ij}\eta_{ij} - m_{i1}a_{1j}(\epsilon_{ij} + \eta_{ij} + \epsilon_{ij}\eta_{ij}) & & \\ a_{n1}\epsilon_{n1} & & & \end{array} \right].$$

Ahora tratemos de investigar si los elementos de  $E_1$  están acotados.

Claramente los elementos de la primera columna de  $E_1$  satisfacen

$$|e_{i1}| \leq |a_{i1}|\epsilon_M, \quad \frac{|e_{i1}|}{|a_{i1}|} \leq \epsilon_M.$$

Los elementos de la submatriz que resulta de quitar el primer renglón y la primera columna de  $E$

$$e_{ij} = a_{ij}\eta_{ij} - m_{i1}a_{1j}(\epsilon_{ij} + \eta_{ij} + \epsilon_{ij}\eta_{ij})$$

satisfacen

$$\begin{aligned} |e_{ij}| &\leq |a_{ij}\eta_{ij}| + |m_{i1}a_{1j}|(|\epsilon_{ij}| + |\eta_{ij}| + |\epsilon_{ij}\eta_{ij}|) \\ &\leq |a_{ij}|\epsilon_M + 3|a_{1j}|\epsilon_M \end{aligned}$$

o bien

$$\frac{|e_{ij}|}{a} \leq 4\epsilon_M, \quad a = \max_{1 \leq i, j \leq n} |a_{ij}|.$$

Finalmente, los coeficientes de  $E_1$  después del primer paso de EGp satisfacen

$$\frac{|e_{ij}|}{a} \leq 4\epsilon_M, \quad 1 \leq i, j \leq n.$$

**Ejercicio.** Es fácil probar que

$$\frac{\|E\|_1}{\|A\|_1} \leq 4(n-1)\epsilon_M$$

**Ejercicio.** Construye un contraejemplo que pruebe que la definición

$$\|A\| = \max_{1 \leq i, j \leq n} |a_{ij}|$$

no satisface las propiedades de una norma consistente.

## Factorización QR

Ahora podemos proceder en forma análoga para describir al método de factorización QR.

Considera un vector arbitrario no nulo  $a$  que deseamos transformar en el vector  $q$

$$a = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}, \quad q = \begin{bmatrix} \pm \|a\|_2 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad (10)$$

Observa con cuidado que queremos comprimir  $n - 1$  pasos de eliminación gaussiana en un único paso. Observa también que  $\|a\|_2 = \|q\|_2$ .

El problema de obtener la transformación que logra  $a \sim q$  se resuelve construyendo una reflexión de Householder.

La resolución de sistemas de ecuaciones lineales mediante factorización ortogonal se base en dos propiedades elementales:

1. Si  $Q$  es una matriz ortogonal de  $n \times n$ , entonces el sistema  $Qx = b$  se puede resolver trivialmente en la forma

$$x = Q^T b$$

2. Si  $R$  es una matriz triangular superior de  $n \times n$ , con elementos no nulos en la diagonal principal, entonces el sistema  $Rx = b$  se puede resolver por sustitución hacia atrás.

Una pregunta natural es la siguiente: ¿es posible representar a cualquier matriz  $A$ , no singular, en la forma  $A = QR$ , en donde  $Q$  y  $R$  tienen las propiedades descritas en 1) y 2)? La respuesta es claramente afirmativa. Por lo menos el algoritmo de Gram-Schmidt es capaz de construir una base ortonormal a partir de cualquier base de  $\mathbb{R}^n$ . Las columnas de  $Q$  forman la base nueva, mientras que la matriz  $R$  representa la relación que existe entre ambas bases. Claramente  $R$  es una matriz invertible.

Una segunda pregunta, de orden práctico, es si el algoritmo de Gram-Schmidt resulta ser *numéricamente estable*. La respuesta es, desafortunadamente, negativa. Por esta razón recurrimos al concepto geométrico de reflector elemental o *transformación de Householder*.

$$H = I - 2 \frac{uu^T}{u^T u},$$

Ahora podemos formular el problema (10) en términos de una transformación de Householder. Es decir, construir un vector  $u$  tal que

$$Ha = q,$$

No es difícil demostrar que  $u$  está dado por la siguiente expresión:

$$u = a \mp \|a\|_2 e_1, \quad (11)$$

en donde el signo corresponde al de la primera componente de  $a$ .

**Ejercicio.** a) Deriva la expresión (11); b) ¿Cuál es la justificación para la elección del signo en el cálculo de  $u$ ?

La factorización QR de una matriz cuadrada no singular de  $n \times n$  se puede obtener fácilmente mediante la aplicación consecutiva de  $n - 1$  transformaciones de Householder

$$\underbrace{H_{n-1} \cdots H_2 H_1}_{Q^T} A = R, \quad A = QR,$$

en donde  $Q$  es ortogonal y  $R$  es triangular superior.

## Factorización de Cholesky

En diversas aplicaciones la matriz de coeficientes  $A$  es simétrica positiva definida (spd). Por ejemplo, en optimización de funciones de varias variables,  $A$  está asociada con la Hessiana de la función objetivo.

Las propiedades de las matrices spd facilitan el cálculo de la factorización LU, La simetría hace posible reducir la complejidad del algoritmo debido a que sólo es necesario almacenar la mitad de la información. Otro ahorro se debe a que la fase de pivoteo se puede omitir sin comprometer la estabilidad numérica del algoritmo.

La factorización de Cholesky tiene la forma

$$A = LL^T,$$

en donde  $L$  es una matriz triangular inferior con entradas positivas en la diagonal principal.

La siguiente es una versión que requiere de calcular  $n$  raíces cuadradas. Esta ligera desventaja se puede corregir reorganizando la factorización en la forma

$$A = LDL^T$$

en donde  $D$  es una matriz diagonal con elementos positivos en la diagonal;  $L$  es triangular inferior.

## Ejercicios prácticos

1. Escribe una realización en MATLAB de la factorización de Cholesky.

```

desde k=1 hasta n
    a(k,k) = sqrt(a(k,k))
    desde i=k+1 hasta n
        a(i,k) = a(i,k)/a(k,k)
    fin
    desde j=k+1 hasta n
        desde i=j hasta n
            a(i,j) = a(i,j) - a(i,k)*a(j,k)
        fin
    fin

```

```

        fin
    fin
fin

```

2. En problemas de mínimos cuadrados es necesario resolver sistemas de ecuaciones lineales inconsistentes de la forma

$$Ax = b,$$

en donde  $A$  es una matriz rectangular de  $m \times n$  con  $m \gg n$ . El problema anterior, al no tener solución, se reformula como

$$\text{minimizar}_{x \in \mathbb{R}^n} \|r\|_2^2, \quad r = Ax - b,$$

Si las columnas de  $A$  forman un conjunto LI, entonces el problema anterior tiene solución única. La solución se obtiene comúnmente mediante las llamadas *ecuaciones normales*

$$A^T A x = A^T b. \quad (12)$$

Formar  $A^T A$  tiene serias desventajas de tipo computacional y numérico. En particular, el número de condición de  $A^T A$  es aproximadamente *el cuadrado* del número de condición de  $A$ ,

La forma apropiada de resolver el problema anterior es mediante la factorización QR de la matriz  $A$ . Si  $A = QR$ , entonces tenemos que de (12) obtenemos

$$Rx = Q^T b.$$

Observa con cuidado que el método no necesita calcular  $A^T A$  en forma explícita.

Prueba tu programa con el problema del censo (1900-2000) para estimar la población en los EUA en 2010. Utiliza los datos sin escalar y ajusta: a) una recta; b) una cuadrática. Utiliza el escalamiento sugerido en `minsq.m` y compara las predicciones.

3. Utiliza el programa `qrsteps.m` para obtener la factorización QR de las matrices del ejercicio anterior.