UNIVERSITAT POLITÈCNICA DE VALÈNCIA

Cálculo de autovalores

Damián Ginestar Peiró

Departamento de Matemática Aplicada

Universidad Politécnica de Valencia

Curso 2011-2012

Índice

Preliminares

- Dada una matriz $A \in \mathbb{R}^{n \times n}$. Si $Ax = \lambda x$ para $\lambda \in \mathbb{C}$ y $x \neq 0$, entonces λ es autovalor de A y x su correspondiente autovector.
- Los autovalores de A son las raíces del polinomio característico de A

$$P(\lambda) = \det(\lambda I - A) = 0$$

- Dos matrices A y B se dice que son similares si existe una matriz K regular tal que

$$B = K^{-1}AK$$

Si

$$\begin{aligned} AX &= X\Lambda \\ K^{-1}AKK^{-1}X &= K^{-1}X\Lambda \\ BK^{-1}X &= K^{-1}X\Lambda \end{aligned}$$

Las matrices A y B tienen los mismos autovalores

Preliminares

- Si A es una matriz real y simétrica entonces se puede disgonalizar mediante una transformación ortogonal

$$K^T A K = D$$

  y los autovalores de A son reales.
- Los autovectores correspondientes a autovalores distintos son ortogonales.

Preliminares

Teorema (transformaciones de Householder)

Sean x e y dos vectores con la misma norma (norma 2). Entonces existe una matriz ortogonal y simétrica, P, tal que

$$y = Px$$

donde

$$P = \left(I - 2ww^T\right) \ , \quad w = \frac{x-y}{\|x-y\|_2}$$

Prueba:

Se cumple $w^T w = 1$ y

$$y = x + cw$$

con $c = -\|x-y\|_2$. Además

Preliminares

$$w^T\left(x+\frac{1}{2}cw\right)=w^T\left(x-\frac{1}{2}(x-y)\right)=\frac{1}{2}w^T(x+y)$$

$$=\frac{1}{2}\left(w^Tx+w^T\right)=\frac{1}{2\left\|x-y\right\|_2}\left((x-y)^Tx+(x-y)^Ty\right)$$

$$\frac{1}{2\left\|x-y\right\|_2}\left(x^Tx-y^Tx+x^Ty-y^Ty\right)=0$$

Por otra parte

$$w^Tx+\frac{1}{2}cw^Tw=0\ ,\ \rightarrow w^Tx+\frac{1}{2}c=0\ ,\ \rightarrow c=-2w^Tx$$

con lo que

$$y=x-2w^Txw=x-2ww^Tx=\left(I-2ww^T\right)x$$

Preliminares

Así, $P = \left(I - 2ww^T\right) = P^T$.
Además

$$\begin{aligned} PP^T &= \left(I - 2ww^T\right)\left(I - 2ww^T\right) \\ &= I - 4ww^T + 4ww^Tww^T = I \end{aligned}$$

Corolario

Se puede construir una transformación de Householder de forma que

$$P_k x = P_k \begin{pmatrix} x_1 \\ \vdots \\ x_k \\ x_{k+1} \\ x_{k+2} \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} x_1 \\ \vdots \\ x_k \\ -S \\ 0 \\ \vdots \\ 0 \end{pmatrix} = y$$

Preliminares

S ha de ser tal que $\|x\|_2 = \|y\|_2$,

$$S^2 = x_{k+1}^2 + x_{k+2}^2 + \cdots + x_n^2$$

Los errores de truncamiento se propagan menos si se elige

$$S = \operatorname{sign}(x_{n+1}) \left(x_{k+1}^2 + x_{k+2}^2 + \cdots + x_n^2 \right)^{1/2}$$

La transformación de Householder

$$P = I - 2ww^T$$

con

$$w = \frac{1}{R}(x - y) = \frac{1}{R}(0, \ldots, 0, x_{k+1} + S, x_{k+2}, \ldots, x_n)^T$$

Preliminares

R debe cumplir

$$\begin{aligned} R^2 &= (x_{k+1} + S)^2 + x_{k+2}^2 + \cdots + x_n^2 \\ &= 2x_{k+1}S + S^2 + x_{k+1}^2 + \cdots + x_n^2 \\ &= 2x_{k+1}S + S^2 \end{aligned}$$

Cálculo de autovalores

- Se puede utilizar para calcular los autovalores de A que son las raíces de su polinomio característico

$$P(\lambda) = \det(\lambda I - A) = 0$$

- Los autovalores de A se tienen que aproximar para $n \geq 5$ (Teorema de Abel)
- El método es inestable numéricamente, ya que pequeñas variaciones en los coeficientes del polinomio carcaterístico puede dar lugar a un gran cambio en sus raíces.
  Ejemplo: $p(\lambda) = \lambda^{16}$, $q(\lambda) = \lambda^{16} - 10^{-16}$. Las raíces de p son todas cero. Las raíces de q son tales que $|\lambda| = 0{,}1$.

Cálculo de autovalores

¿Cómo cambian los autovalores al cambiar los elementos de matriz?

Teorema

Sea $A \in \mathbb{R}^{n\times n}$ y $Ax_j = \lambda_j x_j$ de forma que los autovectores $X = (x_1, x_2, \ldots, x_n)$ son linealmente independientes. Si $E \in \mathbb{R}^{n\times n}$ y λ' es un autovalor de $A' = A + E$, entonces A tiene un autovalor de forma que

$$\left|\lambda' - \lambda_i\right| \leq \mathrm{cond}_2(X)\,\|E\|_2$$

donde

$$\mathrm{cond}_2(X) = \|X\|_2 \left\|X^{-1}\right\|_2$$

Método de la potencia

Definición

Si λ es un autovalor de A que es más grande en valor absoluto que cualquier otro autovalor, se dice que es el autovalor dominante de A. a su correspondiente autovector se le llama el autovector dominante

- El método de la potencia es un método iterativo que permite aproximar el autovector dominante de una matriz A

Método de la potencia

Supongamos que los autovectores de $A \in \mathbb{R}^{n\times n}$, $x_1, \ldots\ , x_n$ son linealmente independientes. Entonces forman base de $\diagdown^n$ y un vector inicial v_0 se expresa

$$v_0 = \sum_{j=1}^{n} \alpha_j x_j$$

Entonces

$$Ax_j = \lambda_j x_j\ , \quad A^k x_j = \lambda_j^k x_j$$

y, por tanto,

$$v_k = A^k v_0 = \sum_{j=1}^{n} \alpha_j \lambda_j^k x_j$$

Método de la potencia

Supongamos también que

$$|\lambda_1| > |\lambda_2| \geq |\lambda_3| \geq \cdots \geq |\lambda_n|$$

Entonces

$$\frac{v_k}{\lambda_1^k} = \frac{A^k v_0}{\lambda_1^k} = \alpha_1 x_1 + \sum_{j=2}^{n} \alpha_j \left(\frac{\lambda_j}{\lambda_1} \right)^k x_j$$

si $\alpha_1 \neq 0$

$$\lim_{k \to \infty} \frac{v_k}{\lambda_1^k} = \alpha_1 x_1$$

Método de la potencia

Algoritmo de la potencia

$u = v_0$

Para $k = 0, 1, 2, \ldots$

$v = Au$

$[m, j] = \max(\text{abs}(v))$

$u = \frac{v}{v(j)}$

Test de parada

end para

- La velocidad de convergencia del método depende de la magnitud del cociente λ_2/λ_1. Cuanto más cercano a 1 sea el cociente más lenta será la convergencia.

Deflacción

Supongamos que se conoce un autovalor λ_1 y su correspondiente autovector x_1 de una matriz A. Además se elige x_1 de forma que $\|x_1\| = 1$. Sea

$$H = I - 2ww^T$$

la transformación de Householder que hace

$$Hx_1 = e_1$$

donde e_1 es el primer vector de la base canónica. Como $H^{-1} = H$, se tiene

$$He_1 = x_1$$

Entonces

$$HAHe_1 = HAx_1 = \lambda_1 Hx_1 = \lambda_1 e_1$$

que es la primera columna de la matrix HAH

Deflacción

Así

$$HAH = \begin{pmatrix} \lambda_1 & a^T \\ 0 & A_1 \end{pmatrix}$$

Como

$$\det(\lambda I - A) = \det(\lambda I - HAH) = (\lambda - \lambda_1) \det(\lambda I - A_1)$$

los autovalores de A_1 son los autovalores $\lambda_2, \lambda_3, \dots, \lambda_n$ de la matrix A. De este modo, se puede aplicar el método de la potencia a A_1 para obtener λ_2 y volver a aplicar la deflacción para calcular λ_3 y así sucesivamente

Deflacción

Otra técnica de deflacción es la deflacción de Wielandt, que se basa en lo siguiente.
Supongamos que $\lambda_1, \lambda_2, \cdots, \lambda_n$ son los autovalores de A con los autovectores $x_1, x_2, \ldots, x_n$. Si la multiplicidad de λ_1 es 1 y x es un vector que satisface $x^T x_1 = 1$. Entonces la matriz

$$B = A - \lambda_1 x_1 x^T$$

tiene los autovalores $0, \lambda_2, \cdots, \lambda_n$ y sus autovectores son $x_1, w_2, \ldots, w_n$ con

$$x_j = (\lambda_j - \lambda_1) w_j + \lambda_1 \left(x^T w_j \right) x_1$$

(La prueba: ejercicio)

Método de la potencia inversa

El método de la potencia inversa se basa en las siguientes propiedades:

- Dada una matriz A con uno de sus autovalores λ y x su correspondiente autovector. Para una constante α se cumple

$$(A - \alpha I)\,x = (\lambda - \alpha)\,x$$

- Dada una matriz A con uno de sus autovalores λ y x su correspondiente autovector. Si $\lambda \neq \alpha$ se tiene que $\frac{1}{\lambda-\alpha}$ es autovalor de $(A - \alpha I)^{-1}$ y su correspondiente autovector es x.
- Si se conoce una aproximación α de uno de los autovalores de A. Se aplica el método de la potencia a la matriz $(A - \alpha I)^{-1}$ para obtener el autovalor dominante μ y su correspondiente autovector x

Método de la potencia inversa

- El autovalor λ de la matriz A es

$$\lambda = \frac{1}{\mu} + \alpha$$

- Hay que tener en cuenta que al implementar el método de la potencia inversa hay que calcular productos

$$(A - \alpha I)^{-1} v = y$$

Estos productos se calculan resolviendo los sistemas

$$(A - \alpha I) y = v$$

Método QR

Este método se basa en el teorema de Schur.

Teorema de la forma real de Schur

Dada una matriz $A \in \mathbb{R}^{n \times n}$ existe una matriz ortogonal U de forma que

$$U^{\dagger} A U = T$$

donde T es una matriz casi-triangular

(Matriz casi-triangular: es una matriz triangular a bloques, con bloques 1×1 o bloques 2×2 en la diagonal) **Algoritmo** QR

$$
\begin{aligned}
& A_1 = A \\
& \text{para } k = 1, 2, \ldots \\
& \quad Q_k R_k = A_k \text{ (Factorización } QR\text{)} \\
& \quad A_{k+1} = R_k Q_k
\end{aligned}
$$

Método QR

¿Cómo funciona el algoritmo?

- $A_2 = R_1 Q_1$ es similar a A_1 ya que
  $A_2 = Q_1^T A_1 Q_1 = Q_1^T Q_1 R_1 Q_1 = R_1 Q_1$
- $A_3 = R_2 Q_2 = Q_2^T A_2 Q_2 = Q_2^T Q_1^T A_1 Q_1 Q_2$
- $A_{k+1} = U_k^T A U_k$, donde $U_k = Q_1 \cdots Q_k$. Así, $\lim_{k\to\infty} U_k = U$, donde $U^T A U = T$ (descomposición de Schur)
- Si A es simétrica, T es una matriz diagonal y las columnas de U son los autovectores de A.

Método QR

- Si A tiene los autovalores complejos entonces T es una matriz triangular superior por bloques con bloques 1×1 o bloques 2×2 en la diagonal.
- Los bloques 1×1 están asociados con los autovalores reales los bloques 2×2 con los autovalores complejos.

Ejemplo

$$A = \begin{pmatrix} 0{,}9501 & 0{,}8913 & 0{,}8212 & 0{,}9218 \\ 0{,}2311 & 0{,}7621 & 0{,}4447 & 0{,}7382 \\ 0{,}6068 & 0{,}4565 & 0{,}6154 & 0{,}1763 \\ 0{,}4860 & 0{,}0185 & 0{,}7919 & 0{,}4057 \end{pmatrix}$$

$$A_{14} = \left(\begin{array}{r|rr|r} 2{,}3230 & 0{,}0471 & -0{,}3924 & -0{,}6505 \\ -0{,}0000 & 0{,}1302 & 0{,}3613 & 0{,}1594 \\ -0{,}0000 & -0{,}5863 & 0{,}0527 & -0{,}2578 \\ 0{,}0000 & 0{,}0000 & -0{,}0000 & 0{,}2275 \end{array}\right)$$

$\lambda_1 = 2{,}3230,\ \lambda_2 = 0{,}0914 + 0{,}4586\,i\ \lambda_2 = 0{,}0914 - 0{,}4586\,i,\ \lambda_4 = 0{,}2275$

Método QR

- Aunque el método QR tiene buenas propiedades de convergencia, requiere la factorización QR de una matriz, que tiene un coste $O\left(n^3\right)$, teniendo el algoritmo global un coste $O\left(n^4\right)$.
- Para reducir este problema, la matriz A se pasa a una forma Hessenberg superior.
- Una matriz es Hessenberg superior si

$$a_{ij} = 0 \text{ si } \; i > j + 1$$

Para dimensión 5

$$\begin{pmatrix} x & x & x & x & x \\ x & x & x & x & x \\ 0 & x & x & x & x \\ 0 & 0 & x & x & x \\ 0 & 0 & 0 & x & x \end{pmatrix}$$

Método QR

Para pasar A a la forma Hessenberg, se construyen $H_1, H_2, \dots, H_{n-2}$ transformaciones de Householder de forma que

$$H_{n-2}H_{n-1}\cdots H_1 A H_1 H_2 \cdots H_{n-2}$$

es similar a la matriz A y tiene estructura Hessenberg superior.

Método QR

Ejemplo $A \in \mathbb{R}^{4\times 4}$

$$A = \begin{pmatrix} x & x & x & x \\ x & x & x & x \\ x & x & x & x \\ x & x & x & x \end{pmatrix} \xrightarrow{H_1 A} \begin{pmatrix} x & x & x & x \\ x & x & x & x \\ 0 & x & x & x \\ 0 & x & x & x \end{pmatrix} \xrightarrow{H_1 A H_1} \begin{pmatrix} x & x & x & x \\ x & x & x & x \\ 0 & x & x & x \\ 0 & x & x & x \end{pmatrix}$$

$$\xrightarrow{H_2 H_1 A H_1} \begin{pmatrix} x & x & x & x \\ x & x & x & x \\ 0 & x & x & x \\ 0 & 0 & x & x \end{pmatrix} \xrightarrow{H_2 H_1 A H_1 H_2} \begin{pmatrix} x & x & x & x \\ x & x & x & x \\ 0 & x & x & x \\ 0 & 0 & x & x \end{pmatrix}$$

$$H_1 = \begin{pmatrix} 1 & 0 \\ 0 & \hat{H}_1 \end{pmatrix}, H_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \hat{H}_2 \end{pmatrix}, \quad \hat{H}_k = I - 2\hat{u}_k \hat{u}_k^T, \hat{u}_k^T \hat{u}_k = 1$$

Método QR

- La transformación a la forma Hessenberg y el algoritmo QR se combinan de forma que el nuevo algoritmo tiene coste $O\left(n^3\right)$ global.
- También se usan 'shifts' similares al método de la potencia inversa combinados con el algoritmo QR para acelerar su convergencia.
- En cada iteración se elige un escalar α_k y se factoriza $A_k - \alpha_k I$ como producto $Q_k R_k$. Entonces,

$$A_{k+1} = R_k Q_k + \alpha_k I$$

- $A_{k+1} = Q_k^T A_k R_k$ es similar a A_k, ya que

$$Q_k^T A_k Q_k = Q_k^T \left(Q_k R_k + \alpha_k I\right) Q_k = R_k Q_k + \alpha_k I = A_{k+1}$$