

Análisis numérico: Tarea 1

1 Método del polinomio característico

Sea A una matriz cuadrada $n \times n$ de un espacio vectorial cerrado y sea $v \in \mathbb{C}^n$ un vector tales que

$$Av = \lambda v, \quad v \neq 0, \quad (1)$$

para un escalar $\lambda \in \mathbb{C}$. A los vectores v que satisfagan la ecuación se les nombrará los vectores propios, eigenvectores o autovectores, correspondientes al valor propio, eigenvalor o autovalor λ .

Se observa que una operación sobre los vectores propios no cambia los ángulos de la matriz A , sino que simplemente escala los resultados. Se desea obtener estos valores. Reorganizando (1):

$$\begin{aligned} Av &= \lambda v \\ \iff Av - \lambda v &= 0 \\ \iff Av - \lambda I v &= 0 \\ \iff (A - \lambda I)v &= 0 \end{aligned}$$

Donde I es la matriz identidad de tamaño n . Como $v \neq 0$, luego se busca una solución no trivial para la matriz $A - \lambda I$.

Observemos que si se desea una solución no trivial para $A - \lambda I$, luego esta debe ser no invertible.

Demostración ($A - \lambda I$ no es invertible): Suponga que $(A - \lambda I)$ es invertible, por lo tanto existe $(A - \lambda I)^{-1}$. Luego,

$$\begin{aligned} v &= Iv \\ &= ((A - \lambda I)^{-1}(A - \lambda I))v \\ &= (A - \lambda I)^{-1}((A - \lambda I)v) \\ &= (A - \lambda I)^{-1}0 \\ &= 0 \end{aligned}$$

Que es una contradicción, dado que $v \neq 0$, luego $(A - \lambda I)$ no es invertible.

Desconocemos tanto v como λ , por lo que inicialmente se usa $\det(A - \lambda I)$ para hallar los valores propios. Como $A - \lambda I$ no es invertible, esto significa que

$$\det(A - \lambda I) = 0 \quad (2)$$

Para A , el método de la expansión de Laplace demuestra que $\det(A - \lambda I_n)$ es un polinomio de grado n . Luego $\det(A - \lambda I) = 0$ tiene n soluciones $\lambda \in \mathbb{C}$ por el teorema fundamental del álgebra.

Se observa que los valores propios de la matriz son idénticos a las raíces del polinomio característico. Para calcular el determinante, requerimos definir *minor* y *cof*(A)

Si A es una matrix $n \times n$, la $\text{minor}(A)_{ij}$ es la determinante de la matriz $n - 1 \times n - 1$ resultante de eliminar la i -ésima fila y la j -ésima columna de A .

Si A es una matrix $n \times n$, el cofactor ij -ésimo, denotado $\text{cof}(A)_{ij}$ es definido como

$$\text{cof}(A)_{ij} = (-1)^{i+j} \text{minor}(A)_{ij}$$

Con esto, podemos definir la determinante como

$$\det(A) = \sum_{j=1}^{\infty} a_{ij} \text{cof}(A)_{ij} = \sum_{i=1}^{\infty} a_{ij} \text{cof}(A)_{ij}$$

$$p_A(\lambda) = \begin{vmatrix} a_{11} - \lambda & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} - \lambda & \dots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & \dots & \dots & a_{nn} - \lambda \end{vmatrix}$$

El cálculo de la determinante resulta en un problema de polinomios (de ahí el nombre de polinomio característico $p_A(\lambda)$). Se llega a un problema de hallar raíces y existen múltiples métodos que lo solucionan con diferente complejidad y velocidad. Unos son analíticos y permiten hallar raíces de manera exacta. Sin embargo, por el teorema de Abel-Ruffini, para matrices con $n \geq 5$ sus raíces no siempre se podrán expresar exactamente en radicales. De todas formas, existen métodos iterativos que ayudan a aproximar su resultado, el más reconocido es el método de Newton, debido a su convergencia cuadrática bajo condiciones específicas.

Finalmente, para encontrar los vectores propios, se toma $(A - \lambda I)v = 0$ y se reemplaza cada una de las raíces o valores propios hallados $\lambda_1, \lambda_2, \dots, \lambda_n$ del polinomio característico. Reemplazando n veces en la matriz se obtienen los n vectores propios de A , uno para cada valor λ_i . Sin embargo, es de tener en cuenta que esta es una operación que puede ser computacionalmente compleja, se usan métodos como eliminación Gaussiana que son sensibles al valor de las entradas y pueden llegar a afectar la precisión del resultado.

$$\begin{bmatrix} a_{11} - \lambda_i & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} - \lambda_i & \dots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & \dots & \dots & a_{nn} - \lambda_i \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

Se convierte a la matriz aumentada, para operar con eliminación Gaussiana.

$$\left[\begin{array}{cccc|c} a_{11} - \lambda_i & a_{12} & \dots & a_{1n} & 0 \\ a_{21} & a_{22} - \lambda_i & \dots & \vdots & \vdots \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{n1} & \dots & \dots & a_{nn} - \lambda_i & 0 \end{array} \right]$$

Para cada valor propio λ_i , se opera hasta llegar a la forma escalonada de la matriz aumentada, se calculan las variables libres de x_1, \dots, x_n , y se hallan los vectores no nulos que cumplen $Av = \lambda_i v$.

1.1 Error

El método del polinomio característico y la eliminación gaussiana son métodos directos. Estos son aquellos que buscan la respuesta en un número finito de pasos, sin errores de redondeo. Estos métodos cuentan con la desventaja de no ser lo suficientemente eficientes como para trabajar con *millones* de datos, debido a que superarían las capacidades de cómputo disponibles, por lo que se suelen usar preferiblemente métodos iterativos, como el algoritmo de la descomposición QR o el método de las potencias.

Es de tener en cuenta que al programar un método directo en una computadora se debe de tener en cuenta el error atribuido a que se tiene una representación aproximada de los números. Un problema fehaciente en el método narrado es que las matrices son altamente sensibles a pequeños cambios en los datos de entrada.

Esta sensibilidad se puede medir por el número de la condición $\text{cond}(A)$. Su definición es: considere los cambios δA y δb en A y en b y su cambio resultante δz , en la solución z .

$$\text{cond}(A) \equiv \max\left(\frac{\frac{\|\delta z\|}{\|z\|}}{\frac{\|\delta A\|}{\|A\|} + \frac{\|\delta b\|}{\|b\|}}\right) = \max\left(\frac{\text{error relativo del resultado}}{\text{error relativo del input}}\right) \quad (3)$$

De esta fórmula, se puede observar que al trabajar con números pequeños, es decir, que se acercan al epsilon de máquina (la precisión de máquina o el número más pequeño que sea representable en el formato de punto flotante), el número de la condición puede hacerse grande en operaciones necesarias para hallar el espacio nulo. Adicionalmente, como las entradas se truncan o aproximan, las matrices con este tipo de entradas suelen ser representadas erróneamente como si tuvieran rango completo.

Calculando el error en las operaciones de la eliminación Gaussiana en una computadora, Gene H. Golub de la Universidad de Chicago obtuvo dos maneras de acotar los errores relacionados a sus parámetros:

$$\frac{\|\delta v\|}{\|v\|} \leq \frac{\text{cond}(A)}{1 - \text{cond}(A) \frac{\|\delta A\|}{\|A\|}} \frac{\|\delta A\|}{\|A\|} \quad (4)$$

$$\frac{\|\delta A\|_\infty}{\|A\|_\infty} \leq 2n(n+1)Gu \quad (5)$$

Donde

- G es un factor de crecimiento que para la mayoría de las matrices cumple $G \leq k$, donde k es un paso que denota el proceso de hacer todos los elementos de la diagonal correspondiente a una columna específica igual a 0.
- La precisión u , que está relacionada al epsilon de máquina.

Estos dos, al igual que $\text{cond}(A)$, son factores determinantes en la precisión de la solución computada. En general, para obtener un buen resultado, se requiere que $\text{cond}(A)u \leq 1$.

2 Método de las potencias

Se dice que un eigenvalor de una matriz A es el *eigenvalor dominante* de A si su valor absoluto es mayor estricto que los valores absolutos de los eigenvalores restantes. Un eigenvector que corresponda al eigenvalor dominante se denomina *eigenvector dominante* de A .

Del algebra lineal se sabe que pueden encontrar los eigenvalores de una matriz al resolver su ecuación característica. En los problemas prácticos, este método no es eficaz, para valores grandes de n , las ecuaciones polinómicas como ésta son difíciles y requieren mucho tiempo para resolverse. Además, las técnicas numéricas para aproximar raíces de ecuaciones polinómicas de alto grado son sensibles a errores de redondeo. En esta sección se analiza un método para obtener una aproximación del eigenvalor dominante de una matriz y un eigenvector correspondiente. Como se presenta aquí, el método se puede utilizar solo para encontrar el eigenvalor dominante de una matriz. Aunque esta restricción puede parecer severa, los eigenvalor dominantes son de interés principal en muchas aplicaciones físicas.

2.1 Representación del método de Potencia

El método de potencia para aproximar eigenvalores es iterativo. Primero, suponemos que la matriz A tiene un eigenvalor dominante con eigenvectores dominantes correspondientes. Luego, elegimos una aproximación inicial \mathbf{x}_0 de uno de los eigenvectores dominantes de A . Esta aproximación inicial debe ser un vector *no nulo* en \mathbb{R}^n . Finalmente, formamos la secuencia dada por:

$$\begin{aligned} \mathbf{x}_1 &= A\mathbf{x}_0, \\ \mathbf{x}_2 &= A\mathbf{x}_1 = A(A\mathbf{x}_0) = A^2\mathbf{x}_0, \\ \mathbf{x}_3 &= A\mathbf{x}_2 = A(A^2\mathbf{x}_0) = A^3\mathbf{x}_0, \\ &\vdots \\ \mathbf{x}_k &= A\mathbf{x}_{k-1} = A(A^{k-1}\mathbf{x}_0) = A^k\mathbf{x}_0. \end{aligned}$$

Para potencias grandes de k , y escalando adecuadamente esta secuencia, veremos que obtenemos una buena aproximación del eigenvector dominante de A .

Los cálculos anteriores están produciendo aproximaciones cada vez mejores para un eigenvector dominante de A .

En seguida se muestra en qué forma obtener una aproximación del eigenvalor dominante, una vez que se conoce una aproximación de un eigenvector dominante. Sea λ un eigenvalor de A y \mathbf{x} un eigenvector correspondiente. Si $\langle \cdot, \cdot \rangle$ denota el producto interior (interno), entonces:

$$\frac{\langle \mathbf{x}, A\mathbf{x} \rangle}{\langle \mathbf{x}, \mathbf{x} \rangle} = \frac{\langle \mathbf{x}, \lambda \mathbf{x} \rangle}{\langle \mathbf{x}, \mathbf{x} \rangle} = \frac{\lambda \langle \mathbf{x}, \mathbf{x} \rangle}{\langle \mathbf{x}, \mathbf{x} \rangle} = \lambda.$$

Por lo tanto, si $\tilde{\mathbf{x}}$ es una aproximación para un eigenvector dominante, se puede obtener una aproximación del eigenvalor dominante por medio de:

$$\text{eigenvalor dominante} \approx \frac{\langle \tilde{\mathbf{x}}, A\tilde{\mathbf{x}} \rangle}{\langle \tilde{\mathbf{x}}, \tilde{\mathbf{x}} \rangle}. \quad (1)$$

La razón que se da en (1) se conoce como *cociente de Rayleigh*.

2.2 Pasos del Método de las Potencias

A continuación se resumen los pasos del *método de las potencias con reducción a escala*:

- **Paso 0.** Se selecciona un vector arbitrario diferente de cero, \mathbf{x}_0 .
- **Paso 1.** Se calcula $A\mathbf{x}_0$ y se reduce a fin de obtener la primera aproximación para un eigenvector dominante. Se nombra como \mathbf{x}_1 .
- **Paso 2.** Se calcula $A\mathbf{x}_1$ y se reduce para obtener la segunda aproximación, \mathbf{x}_2 .
- **Paso 3.** Se calcula $A\mathbf{x}_2$ y se reduce para obtener la tercera aproximación, \mathbf{x}_3 .

Continuando de esta manera, se obtiene una sucesión $\mathbf{x}_0, \mathbf{x}_1, \mathbf{x}_2, \dots$, de aproximaciones cada vez mejores para un eigenvector dominante.

Adicionalmente, si se quiere saber las aproximaciones de λ en la k iteración \mathbf{x}_k se calcula

$$\lambda \approx \lambda_k = \frac{\langle \mathbf{x}_k, A\mathbf{x}_k \rangle}{\langle \mathbf{x}_k, \mathbf{x}_k \rangle}.$$

2.3 Teoría del Método de las Potencias

Resultados previos necesarios del álgebra lineal:

Teorema 1. Si A es una matriz de $n \times n$, entonces las proposiciones que siguen son equivalentes:

- A es diagonalizable. Esto es que $P^{-1}AP = D$, donde P es una matriz invertible y D una matriz diagonal.
- A tiene n eigenvectores linealmente independientes.

Teorema 2. Si $S = \{v_1, v_2, \dots, v_n\}$ es un conjunto de n vectores linealmente independientes en un espacio V de dimensión n , entonces S es una base para V .

Ahora hagamos la demostración de que el método de las potencias funciona, Cuando A es una matriz diagonalizable, con un eigenvalor dominante.

Proposición Sea A una matriz diagonalizable de $n \times n$, con un eigenvalor dominante. Si \mathbf{x}_0 es un vector arbitrario diferente de cero en \mathbb{R}^n , entonces el vector

$$A^p \mathbf{x}_0, \quad p \in \mathbb{N}$$

es, una buena aproximación para un eigenvector dominante de A , cuando el exponente p es grande.

Demostración:

Sea A una matriz diagonalizable de $n \times n$. Por el teorema 1, A tiene n eigenvectores linealmente independientes, $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$. Sean $\lambda_1, \lambda_2, \dots, \lambda_n$ los eigenvalores correspondientes, siendo λ_1 el eigenvalor dominante esto es:

$$|\lambda_1| > |\lambda_2| \geq \dots \geq |\lambda_n| \geq 0. \quad (2)$$

Por el teorema 2, los eigenvectores $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$ forman una base para \mathbb{R}^n ; por tanto, se puede expresar un vector arbitrario \mathbf{x}_0 en \mathbb{R}^n en la forma:

$$\mathbf{x}_0 = k_1 \mathbf{v}_1 + k_2 \mathbf{v}_2 + \dots + k_n \mathbf{v}_n. \quad (3)$$

Al multiplicar, desde la izquierda, los dos miembros por A se obtiene:

$$A\mathbf{x}_0 = A(k_1 \mathbf{v}_1 + k_2 \mathbf{v}_2 + \dots + k_n \mathbf{v}_n),$$

$$A\mathbf{x}_0 = k_1 A\mathbf{v}_1 + k_2 A\mathbf{v}_2 + \dots + k_n A\mathbf{v}_n,$$

$$A\mathbf{x}_0 = k_1 \lambda_1 \mathbf{v}_1 + k_2 \lambda_2 \mathbf{v}_2 + \dots + k_n \lambda_n \mathbf{v}_n.$$

Si se multiplica una vez más por A , se obtiene:

$$A^2 \mathbf{x}_0 = A(k_1 \lambda_1 \mathbf{v}_1 + k_2 \lambda_2 \mathbf{v}_2 + \dots + k_n \lambda_n \mathbf{v}_n),$$

$$A^2 \mathbf{x}_0 = k_1 \lambda_1^2 \mathbf{v}_1 + k_2 \lambda_2^2 \mathbf{v}_2 + \dots + k_n \lambda_n^2 \mathbf{v}_n.$$

Continuando, después de p multiplicaciones por A , se obtendrá:

$$A^p \mathbf{x}_0 = k_1 \lambda_1^p \mathbf{v}_1 + k_2 \lambda_2^p \mathbf{v}_2 + \dots + k_n \lambda_n^p \mathbf{v}_n. \quad (4)$$

Puesto que $\lambda_1 \neq 0$ (véase (2)), la ecuación (4) se puede reescribir como:

$$A^p \mathbf{x}_0 = \lambda_1^p \left(k_1 \mathbf{v}_1 + k_2 \left(\frac{\lambda_2}{\lambda_1} \right)^p \mathbf{v}_2 + \dots + k_n \left(\frac{\lambda_n}{\lambda_1} \right)^p \mathbf{v}_n \right). \quad (5)$$

De (2) se deduce que:

$$\frac{\lambda_2}{\lambda_1}, \dots, \frac{\lambda_n}{\lambda_1}$$

son todos menores que uno en valor absoluto; por tanto, $\left(\frac{\lambda_2}{\lambda_1} \right)^p, \dots, \left(\frac{\lambda_n}{\lambda_1} \right)^p$ se aproximan paulatinamente a cero, a medida que p crece, y por (5), la aproximación:

$$A^p \mathbf{x}_0 \approx \lambda_1^p k_1 \mathbf{v}_1 \quad (6)$$

se hace cada vez mejor.

Si $k_1 \neq 0$, entonces $\lambda_1^p k_1 \mathbf{v}_1$ es un múltiplo escalar diferente de cero del eigenvector dominante \mathbf{v}_1 ; por tanto, $\lambda_1^p k_1 \mathbf{v}_1$ también es un eigenvector dominante. Por tanto, por (6), $A^p \mathbf{x}_0$ se convierte cada vez en una estimación mejor de un eigenvector dominante, conforme p crece. Además, si los vectores \mathbf{x}_p se normalizan es claro que $\mathbf{v}_\lambda \approx A^p \mathbf{x}_0$.

□

Nota: Por lo común no es posible decir por la simple observación de \mathbf{x}_0 si $k_1 \neq 0$. Si, por accidente, $k_1 = 0$, el método de las potencias todavía funciona en los problemas prácticos, ya que generalmente los errores de redondeo en las computadoras tienden a hacer que k_1 sea pequeño, pero diferente de cero. Curiosamente, este es uno de los casos en el que los errores ayudan a obtener resultados correctos.

Orden de convergencia: Suponga que

$$\mathbf{x}_k = \frac{A^k \mathbf{x}_0}{\|A^k \mathbf{x}_0\|},$$

es decir, está normalizado. Veamos que la sucesión

$$\mathbf{x}_k \rightarrow t \mathbf{v}_1$$

con un orden de convergencia

$$O\left(\left(\frac{\lambda_1}{\lambda_2}\right)^k\right),$$

donde λ_2 y λ_1 son como en la proposición anterior y $t\mathbf{v}_1 \neq \vec{0}$ es el eigenvector dominante \mathbf{v}_1 escalado en un factor de $t \neq 0$.

$$\begin{aligned}\|\mathbf{x}_k - t\mathbf{v}_1\| &\leq \|\mathbf{x}_k\| + \|t\mathbf{v}_1\| \\ &= 1 + \|t\mathbf{v}_1\|.\end{aligned}\tag{7}$$

Ya que $\|\mathbf{x}_k\| = 1$. Ahora como $t\mathbf{v}_1 \neq \vec{0}$, es claro que $\|t\mathbf{v}_1\| > 0$. Luego, por (2), se tiene que:

$$\frac{|\lambda_1|}{|\lambda_2|} > 1 \quad \text{así} \quad \left(\frac{|\lambda_1|}{|\lambda_2|}\right)^k > 1,$$

y multiplican. Por (7), se tiene:

$$1 + \|t\mathbf{v}_1\| < (1 + \|t\mathbf{v}_1\|) \left(\frac{|\lambda_1|}{|\lambda_2|}\right)^k$$

Por tanto, de la primera desigualdad y de la ultima, se sigue que

$$\|\mathbf{x}_k - \lambda_1 \mathbf{v}_1\| \leq \varepsilon \left(\frac{|\lambda_1|}{|\lambda_2|}\right)^k.$$

donde el escalar $\varepsilon = 1 + \|t\mathbf{v}_1\|$.

3 Método de iteración QR

El método de descomposición QR es uno de los métodos más eficaces para hallar valores propios de una matriz cuadrada. Fue inventado de manera independiente por John G.F. Francis de Reino Unido y por Vera Kublánovskaya de la URSS alrededor de la década de los 50. Este método comienza por realizar una descomposición QR de una matriz (lo que le da el nombre), donde Q es una matriz ortonormal y R es una matriz triangular superior. A partir de esto, se construye una iteración de las matrices resultantes que converge eventualmente a una matriz triangular cuyos elementos de la diagonal serán los valores propios de la matriz inicial. A continuación, se detalla el procedimiento y sus fundamentos matemáticos.

Descomposición QR Sea $A \in \mathbb{R}^{n \times n}$ una matriz de columnas linealmente independientes. La descomposición QR se basa en una versión matricial del método Gram-Schmidt para ortonormalizar un conjunto de vectores. Este método toma un conjunto finito de vectores linealmente independientes en un espacio vectorial de dimensión n , sea este

$$\mathbf{S} = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k\}, \quad k \leq n,$$

para generar a partir de ellos un conjunto ortonormal

$$\mathbf{S}' = \{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_k\},$$

donde:

$$\mathbf{u}'_1 = \mathbf{v}_1, \quad \mathbf{u}'_2 = \mathbf{v}_2 - \text{proj}_{\mathbf{u}'_1}(\mathbf{v}_2), \quad \dots, \quad \mathbf{u}'_k = \mathbf{v}_k - \sum_{j=1}^{k-1} \text{proj}_{\mathbf{u}'_j}(\mathbf{v}_k),$$

tal que:

$$\mathbf{u}_k = \frac{\mathbf{u}'_k}{\|\mathbf{u}'_k\|}.$$

Así, sea $A = [\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_n]$, donde $\mathbf{c}_i \in \mathbb{R}^{n \times 1}$ son linealmente independientes. Usando Gram-Schmidt, tenemos que:

$$\mathbf{u}'_1 = \mathbf{c}_1, \quad \mathbf{u}'_2 = \mathbf{c}_2 - \text{proj}_{\mathbf{u}'_1}(\mathbf{c}_2) = \mathbf{c}_2 - \frac{\langle \mathbf{c}_2, \mathbf{u}'_1 \rangle}{\langle \mathbf{u}'_1, \mathbf{u}'_1 \rangle} \mathbf{u}'_1, \quad \dots$$

$$\mathbf{u}'_k = \mathbf{c}_k - \sum_{j=1}^{k-1} \text{proj}_{\mathbf{u}'_j}(\mathbf{c}_k). \quad (*)$$

tal que

$$\mathbf{u}_k = \frac{\mathbf{u}'_k}{\|\mathbf{u}'_k\|}.$$

Luego, $Q = [\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n]$ es una matriz ortogonal de $n \times n$, y $R \in \mathbb{R}^{n \times n}$. Ahora bien, por (*) tenemos que:

$$\begin{aligned} \mathbf{u}'_k &= \mathbf{c}_k - \frac{\langle \mathbf{c}_k, \mathbf{u}'_1 \rangle}{\langle \mathbf{u}'_1, \mathbf{u}'_1 \rangle} \mathbf{u}'_1 - \dots - \frac{\langle \mathbf{c}_k, \mathbf{u}'_{k-1} \rangle}{\langle \mathbf{u}'_{k-1}, \mathbf{u}'_{k-1} \rangle} \mathbf{u}'_{k-1} \\ &= \mathbf{c}_k - \frac{\langle \mathbf{c}_k, \mathbf{u}'_1 \rangle}{\|\mathbf{u}'_1\|^2} \mathbf{u}'_1 - \dots - \frac{\langle \mathbf{c}_k, \mathbf{u}'_{k-1} \rangle}{\|\mathbf{u}'_{k-1}\|^2} \mathbf{u}'_{k-1} \\ &= \mathbf{c}_k - \frac{\langle \mathbf{c}_k, \mathbf{u}'_1 \rangle}{\|\mathbf{u}'_1\|} \mathbf{u}_1 - \dots - \frac{\langle \mathbf{c}_k, \mathbf{u}'_{k-1} \rangle}{\|\mathbf{u}'_{k-1}\|} \mathbf{u}_{k-1} \\ \implies \mathbf{c}_k &= \mathbf{u}_k \|\mathbf{u}'_k\| + \langle \mathbf{c}_k, \mathbf{u}'_1 \rangle \mathbf{u}_1 + \dots + \langle \mathbf{c}_k, \mathbf{u}'_{k-1} \rangle \mathbf{u}_{k-1}. \end{aligned}$$

Así, cada \mathbf{c}_n puede expresarse como combinación lineal de los vectores linealmente independientes \mathbf{u}_i , $i = 1, \dots, n$, y así:

$$A = [\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_n] = [\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n] \begin{bmatrix} \|\mathbf{u}_1\| & \langle \mathbf{c}_2, \mathbf{u}_1 \rangle & \langle \mathbf{c}_3, \mathbf{u}_1 \rangle & \dots & \langle \mathbf{c}_n, \mathbf{u}_1 \rangle \\ 0 & \|\mathbf{u}_2\| & \langle \mathbf{c}_3, \mathbf{u}_2 \rangle & \dots & \langle \mathbf{c}_n, \mathbf{u}_2 \rangle \\ 0 & 0 & \|\mathbf{u}_3\| & \dots & \langle \mathbf{c}_n, \mathbf{u}_3 \rangle \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \|\mathbf{u}_n\| \end{bmatrix}.$$

Notemos que la segunda matriz es una matriz triangular superior con su diagonal positiva. Esta es la matriz R , lo cual completa la descomposición de A de la forma buscada.

Iteración Una vez obtenida la descomposición, hacemos $A = \mathbf{A}_0$, $Q = \mathbf{Q}_0$ y $R = \mathbf{R}_0$, tal que $\mathbf{A}_1 = \mathbf{R}_0 \mathbf{Q}_0$. A la matriz \mathbf{A}_1 le aplicamos el método de descomposición nuevamente tal que $\mathbf{A}_1 = \mathbf{Q}_1 \mathbf{R}_1$, y hacemos $\mathbf{A}_2 = \mathbf{R}_1 \mathbf{Q}_1$. Iteramos este procedimiento sucesivamente, tal que $\mathbf{A}_m = \mathbf{R}_{m-1} \mathbf{Q}_{m-1}$. Notemos, además, que para cualquier m tendremos que:

$$\mathbf{A}_{m+1} = \mathbf{R}_m \mathbf{Q}_m = \mathbf{I} \mathbf{R}_m \mathbf{Q}_m = \mathbf{Q}_m^T \mathbf{Q}_m \mathbf{R}_m \mathbf{Q}_m = \mathbf{Q}_m^T \mathbf{A}_m \mathbf{Q}_m = \mathbf{Q}_m^{-1} \mathbf{A}_m \mathbf{Q}_m. \quad (**)$$

Esto nos indica que \mathbf{A}_m y \mathbf{A}_{m+1} son matrices semejantes y por tanto tienen los mismos valores y vectores propios. Así, es fácil notar que \mathbf{A}_m será semejante a la matriz original A para todo m .

3.1 Convergencia

Bajo ciertas condiciones para A , que no son especialmente restrictivas, la iteración converge a una matriz triangular superior denominada la forma Schur de A , cuyos valores propios están en la diagonal de dicha matriz. La convergencia de este método fue probada por James Hardy Wilkinson. La convergencia de la iteración no es obvia, ni la demostración trivial, por lo que dejaremos acá las hipótesis necesarias y, en rasgos generales, en qué consiste la prueba; para conocer la prueba formal en detalle se referencia al autor mencionado en su artículo "Convergence of the LR, QR, and Related Algorithms" o en su libro "The Algebraic Eigenvalue Problem".

Decir que A converge a su forma Schur es decir que la iteración converge a UTU^{-1} , donde U es una matriz unitaria y T una matriz triangular superior cuyos elementos de la diagonal son los valores propios de A . Para esto, debemos pedirle a la matriz que sus valores propios sean de módulo distinto, es decir, que $|\lambda_1| > |\lambda_2| > \dots > |\lambda_n|$.

Bajo estas hipótesis, para cada paso de la descomposición $\mathbf{A}_m = \mathbf{Q}_m \mathbf{R}_m$ y de la iteración $\mathbf{A}_m = \mathbf{R}_{m-1} \mathbf{Q}_{m-1}$, se puede definir $\mathbf{U}_m = \mathbf{U}_{m-1} \mathbf{Q}_m$, donde $\mathbf{U}_0 = I$, tal que al iterar el algoritmo obtendremos que $U = \lim_{m \rightarrow \infty} \mathbf{U}_m$ y $T = \lim_{m \rightarrow \infty} \mathbf{A}_m$, donde T es la matriz triangular superior con los valores propios en la diagonal y los elementos debajo de esta convergen a cero con orden

$$|a_{ij}^{(k)}| = \mathcal{O}(|\lambda_i/\lambda_j|^k), \quad i > j.$$

De esta forma obtenemos los valores propios $\lambda_1, \lambda_2, \dots, \lambda_n$ de la matriz A .

Vectores propios Similar al método anterior, para encontrar los vectores propios de cada valor propio λ_i , se toma $(A - \lambda_i I)v = 0$ y se despeja v .

3.2 Error

Notemos que este método se basa en multiplicaciones matriciales, de forma que cualquier error inicial en la descomposición QR y sus iteraciones subsiguientes puede afectar la convergencia final; para matrices simétricas, el método QR converge cuadráticamente, convergencia que puede ser más lenta si la matriz no es simétrica. Por otra parte, la precisión de los vectores propios depende de qué tan separados estén los valores propios; en caso de que los valores propios estén muy cerca entre ellos, las perturbaciones pequeñas en las matrices producto de las iteraciones pueden llevar a errores significativos en los resultados. Notemos, además, que al hallar los vectores propios correspondientes a cada valor se debe usar la eliminación Gaussiana, de forma que los errores señalados en el método del polinomio característico, que también hace uso de la eliminación Gaussiana, vuelven a presentarse para este método a la hora de hallar los vectores. Por otra parte, este método, aunque en general es exitoso y bastante eficaz en aproximarse a los valores propios de una matriz, entre mayor sea la dimensión de esta, mayor número de iteraciones requiere el método, llegando incluso a necesitar más de 1000 iteraciones, lo que aumenta su complejidad computacional.

4 Conclusiones

Dependiendo del problema que necesites solucionar asociado a encontrar valores y vectores propios de una matriz particular, la eficiencia en cada uno de los métodos varía. El polinomio característico tiene la ventaja de entregar la solución exacta de los valores propios y vectores propios en matrices pequeñas, sin embargo para matrices más grandes este método puede ser lento, a diferencias de los otros dos métodos que son más eficientes computacionalmente. El método de las potencias posee la ventaja de ser computacionalmente eficiente y fácil de implementar, es útil para matrices apropiadas de todos los tamaños, sin embargo sólo halla la aproximación del valor propio dominante y su respectivo vector propio dominante. a diferencia de los otros dos métodos. Para tener certeza de que el método de las potencias y el método descomposición QR funcione apropiadamente se requieren condiciones que se desconocen a priori. En la práctica se suele hacer unas cuantas iteraciones para observar empíricamente el comportamiento del método. En la práctica se suele utilizar el método de descomposición QR cuando se requiere hallar todos los valores propios de una matriz grande, ya que el método de las potencias solo encuentra el valor propio dominante y el método del polinomio característico no es eficiente.

5 Referencias

1. Hardy, James. (1988) The Algebraic Eigenvalue problem. Oxford University Press. Oxford.
2. Hardy, James. (1965). Convergence of the LR, QR and Related Algorithms. The Computer Journal, 8(1), pp. 77-84.
3. Anton, H. (1976). Introducción al Álgebra Lineal. 3ra ed. Limusa.
4. Interactive Linear Algebra: The Characteristic Polynomial
5. Kernel (linear algebra) - Wikipedia
6. SymPy Documentation
7. Eigenvalues - Purdue University
8. Relationship between Null Space and Determinant - Math Stack Exchange
9. Characteristic Polynomial - Georgia Tech
10. Lecture 11 - Ohio University
11. Notes 9 - University of Chicago
12. QR Algorithm Invalid Results - Math Stack Exchange
13. Convergence of Unshifted QR Algorithm Proof - Math Stack Exchange
14. Convergence and Eigenpairs in the QR Algorithm - Math Stack Exchange
15. QR Algorithm Convergence Property - Math Stack Exchange

16. Convergence of QR Algorithm to Upper Triangular Matrix - Math Stack Exchange
17. Schur Decomposition - Wikipedia
18. Power Iteration - Wikipedia