

Curso de Métodos Numéricos

DEMAT, Universidad de Guanajuato

Clase 12: Cálculo de eigenvectores y eigenvalores

- Introducción.
- Círculos de Gershgorin.
- Método de la potencia.

MAT-251

Dr. Joaquín Peña Acevedo
CIMAT A.C.

e-mail: joaquin@cimat.mx

Eigenvalores y eigenvectores de una matriz

Sea $\mathbf{A} \in \mathbb{R}^{n \times n}$.

Definición

El polinomio $p(\lambda) = \det(\mathbf{A} - \lambda \mathbf{I})$ es llamado el *polinomio característico* de \mathbf{A} . Las raíces $p(\lambda) = 0$ del polinomio son los *eigenvalores* de \mathbf{A} .

Definición

Un vector $\mathbf{v} \neq \mathbf{0}$ que satisface $\mathbf{A}\mathbf{v} = \lambda\mathbf{v}$ es un *eigenvector* de \mathbf{A} . Al par ordenado (λ, \mathbf{v}) se le llama *eigenpar*.

Observaciones:

Eigenvalores y eigenvectores de una matriz

Sea $\mathbf{A} \in \mathbb{R}^{n \times n}$.

Definición

El polinomio $p(\lambda) = \det(\mathbf{A} - \lambda \mathbf{I})$ es llamado el *polinomio característico* de \mathbf{A} . Las raíces $p(\lambda) = 0$ del polinomio son los *eigenvalores* de \mathbf{A} .

Definición

Un vector $\mathbf{v} \neq \mathbf{0}$ que satisface $\mathbf{A}\mathbf{v} = \lambda\mathbf{v}$ es un *eigenvector* de \mathbf{A} . Al par ordenado (λ, \mathbf{v}) se le llama *eigenpar*.

Observaciones:

- Como el grado del polinomio característico es n , entonces hay n eigenvalores asociados a \mathbf{A} .

Eigenvalores y eigenvectores de una matriz

Sea $\mathbf{A} \in \mathbb{R}^{n \times n}$.

Definición

El polinomio $p(\lambda) = \det(\mathbf{A} - \lambda \mathbf{I})$ es llamado el *polinomio característico* de \mathbf{A} . Las raíces $p(\lambda) = 0$ del polinomio son los *eigenvalores* de \mathbf{A} .

Definición

Un vector $\mathbf{v} \neq \mathbf{0}$ que satisface $\mathbf{A}\mathbf{v} = \lambda\mathbf{v}$ es un *eigenvector* de \mathbf{A} . Al par ordenado (λ, \mathbf{v}) se le llama *eigenpar*.

Observaciones:

- Como el grado del polinomio característico es n , entonces hay n eigenvalores asociados a \mathbf{A} .
- Para cada eigenvalor λ , puesto que $\det(\mathbf{A} - \lambda \mathbf{I}) = 0$, la matriz $\mathbf{A} - \lambda \mathbf{I}$ es singular, por lo que podemos hallar un vector $\mathbf{v} \neq \mathbf{0}$ tal que $(\mathbf{A} - \lambda \mathbf{I})\mathbf{v} = \mathbf{0}$.

Supongamos que (λ, \mathbf{v}) es un eigenpar de \mathbf{A} .

Dado $\delta \in \mathbb{R}$, ¿Cuáles son los eigenvalores de $\mathbf{A} + \delta \mathbf{I}$?

Supongamos que (λ, \mathbf{v}) es un eigenpar de \mathbf{A} .

Dado $\delta \in \mathbb{R}$, ¿Cuáles son los eigenvalores de $\mathbf{A} + \delta \mathbf{I}$?

$$(\mathbf{A} + \delta \mathbf{I})\mathbf{v} = \mathbf{A}\mathbf{v} + \delta \mathbf{v} = \lambda \mathbf{v} + \delta \mathbf{v} = (\lambda + \delta)\mathbf{v}$$

Entonces $(\lambda + \delta, \mathbf{v})$ es un eigenpar de $\mathbf{A} + \delta \mathbf{I}$.

¿Cuáles son los eigenvalores de \mathbf{A}^{-1} ?

Supongamos que (λ, \mathbf{v}) es un eigenpar de \mathbf{A} .

Dado $\delta \in \mathbb{R}$, ¿Cuáles son los eigenvalores de $\mathbf{A} + \delta \mathbf{I}$?

$$(\mathbf{A} + \delta \mathbf{I})\mathbf{v} = \mathbf{A}\mathbf{v} + \delta \mathbf{v} = \lambda \mathbf{v} + \delta \mathbf{v} = (\lambda + \delta)\mathbf{v}$$

Entonces $(\lambda + \delta, \mathbf{v})$ es un eigenpar de $\mathbf{A} + \delta \mathbf{I}$.

¿Cuáles son los eigenvalores de \mathbf{A}^{-1} ?

Tenemos que $\mathbf{A}\mathbf{v} = \lambda \mathbf{v}$. Entonces

$$\mathbf{A}^{-1}\mathbf{A}\mathbf{v} = \lambda \mathbf{A}^{-1}\mathbf{v} \quad \Rightarrow \quad \frac{1}{\lambda} \mathbf{v} = \mathbf{A}^{-1}\mathbf{v}$$

Así, $(1/\lambda, \mathbf{v})$ es un eigenpar de \mathbf{A}^{-1} .

Proposición

Sea \mathbf{A} una matriz simétrica de tamaño n . Entonces, contando multiplicidades, \mathbf{A} tiene n eigenvalores reales λ_i , $i = 1, \dots, n$, y

$$\mathbf{A}\mathbf{v}_i = \lambda_i\mathbf{v}_i, \quad \mathbf{v}_i^\top \mathbf{v}_j = 0 \quad i \neq j.$$

Si todos los eigenvalores son distintos, los eigenvectores son únicos, excepto por un factor de escala.

La proposición anterior garantiza la descomposición

$$\mathbf{A} = \mathbf{V}\mathbf{D}\mathbf{V}^\top.$$

donde $\mathbf{V} = [\mathbf{v}_1 \cdots \mathbf{v}_n]$ es una matriz unitaria, con los eigenvectores como columnas, y $\mathbf{D} = \text{diag}(\lambda_1, \dots, \lambda_n)$ una matriz diagonal con los eigenvalores correspondientes.

Esta es la *descomposición espectral* de la matriz simétrica \mathbf{A} .

$$AV = VD$$

Solución de $\mathbf{Ax} = \mathbf{b}$ con una matriz simétrica \mathbf{A}

Dada la descomposición espectral de una matriz \mathbf{A} simétrica,

$$\mathbf{A} = \mathbf{VDV}^T,$$

si los eigenvalores son diferentes de cero, \mathbf{D} es invertible y se puede resolver el sistema $\mathbf{Ax} = \mathbf{b}$ mediante

$$\mathbf{VDV}^T \mathbf{x} = \mathbf{b} \quad \Rightarrow \quad \mathbf{x} = \mathbf{VD}^{-1} \mathbf{V}^T \mathbf{b}$$

Proposición 1

Sea \mathbf{A} una matriz simétrica. Si todos los eigenvalores son positivos, la matriz \mathbf{A} es definida positiva.

Solución de $\mathbf{Ax} = \mathbf{b}$ con una matriz simétrica \mathbf{A}

Dada la descomposición espectral de una matriz \mathbf{A} simétrica,

$$\mathbf{A} = \mathbf{VDV}^T,$$

si los eigenvalores son diferentes de cero, \mathbf{D} es invertible y se puede resolver el sistema $\mathbf{Ax} = \mathbf{b}$ mediante

$$\mathbf{VDV}^T \mathbf{x} = \mathbf{b} \quad \Rightarrow \quad \mathbf{x} = \mathbf{VD}^{-1} \mathbf{V}^T \mathbf{b}$$

Proposición 1

Sea \mathbf{A} una matriz simétrica. Si todos los eigenvalores son positivos, la matriz \mathbf{A} es definida positiva.

Entonces en el caso mínimos cuadrados lineales, la solución se obtiene resolviendo el sistema $\mathbf{A}^T \mathbf{Ax} = \mathbf{A}^T \mathbf{b}$. Si se tiene la descomposición espectral de $\mathbf{A}^T \mathbf{A}$, se puede resolver el sistema y determinar si la matriz es definida positiva.

Radio espectral de la matriz

El conjunto $\sigma(\mathbf{A})$ de todos los eigenvalores distintos de \mathbf{A} se llama el *espectro* de \mathbf{A} .

El *radio espectral* $\rho(\mathbf{A})$ de una matriz \mathbf{A} se define como

$$\rho(\mathbf{A}) = \max_{\lambda \in \sigma(\mathbf{A})} |\lambda|.$$

Proposición

Sea \mathbf{A} una matriz $n \times n$. Entonces $\rho(\mathbf{A}) \leq \|\mathbf{A}\|$ para cualquier norma natural.

Radio espectral de la matriz

El conjunto $\sigma(\mathbf{A})$ de todos los eigenvalores distintos de \mathbf{A} se llama el *espectro* de \mathbf{A} .

El *radio espectral* $\rho(\mathbf{A})$ de una matriz \mathbf{A} se define como

$$\rho(\mathbf{A}) = \max_{\lambda \in \sigma(\mathbf{A})} |\lambda|.$$

Proposición

Sea \mathbf{A} una matriz $n \times n$. Entonces $\rho(\mathbf{A}) \leq \|\mathbf{A}\|$ para cualquier norma natural.

Para todo eigenpar (λ, \mathbf{v}) de \mathbf{A} , con $\|\mathbf{v}\| = 1$, tenemos

$$|\lambda| = |\lambda| \|\mathbf{v}\| = \|\mathbf{A}\mathbf{v}\| \leq \|\mathbf{A}\|$$

En particular se cumple para $\rho(\mathbf{A})$.

De esta forma, todos los eigenvalores están dentro de una vecindad del origen de radio $\|\mathbf{A}\|$.

Círculo de Gershgorin

Si λ es un eigenvalor de una matriz $\mathbf{A} = [a_{ij}]$ de tamaño n , hay un entero $i \in \{1, 2, \dots, n\}$ tal que $|\lambda - a_{ii}| \leq \sum_{j \neq i}^n |a_{ij}|$.

Supongamos que $\mathbf{v} = (v_1, \dots, v_n)^T$ es un eigenvector de \mathbf{A} , $\mathbf{A}\mathbf{v} = \lambda\mathbf{v}$, y que está escalado de modo que $1 = \|\mathbf{v}\|_\infty = |v_i|$ para algún índice i . Sea \mathbf{e}_i el vector canónico i -ésimo.

Teorema de Gershgorin

Sea \mathbf{A} una matriz cuadrada arbitraria. Los eigenvalores λ de $\mathbf{A} = [a_{ij}]$ están localizados en la unión de n discos definidos por

$$|\lambda - a_{ii}| \leq r_i \quad \text{donde} \quad r_i = \sum_{\substack{j=1 \\ j \neq i}}^n |a_{ij}| \quad i = 1, \dots, n.$$

Teorema de Gershgorin

Sea \mathbf{A} una matriz cuadrada arbitraria. Los eigenvalores λ de $\mathbf{A} = [a_{ij}]$ están localizados en la unión de n discos definidos por

$$|\lambda - a_{ii}| \leq r_i \quad \text{donde} \quad r_i = \sum_{\substack{j=1 \\ j \neq i}}^n |a_{ij}| \quad i = 1, \dots, n.$$

Lo que nos dice el teorema es que todos los eigenvalores de \mathbf{A} están contenidos en la unión C_r de los n círculos con centro en a_{ii} y radio r_i .

Como $\sigma(\mathbf{A}^T) = \sigma(\mathbf{A})$, entonces la unión C_c de los círculos definidos por

$$|\lambda - a_{jj}| \leq c_j \quad \text{donde} \quad c_j = \sum_{\substack{i=1 \\ i \neq j}}^n |a_{ij}| \quad j = 1, \dots, n.$$

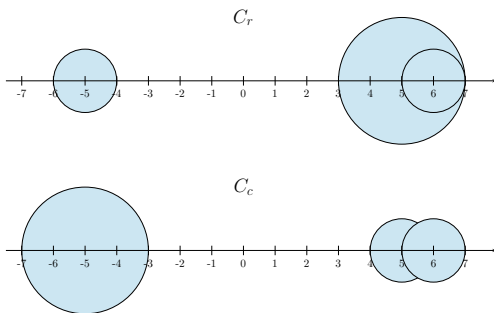
contienen a los eigenvalores de \mathbf{A} .

En resumen, los eigenvalores de \mathbf{A} están contenidos en $C_r \cap C_c$.

Ejemplo de los círculos de Gershgorin (I)

$$\mathbf{A} = \begin{bmatrix} 5 & 1 & -1 \\ 0 & 6 & 1 \\ 1 & 0 & -5 \end{bmatrix} \Rightarrow \|\mathbf{A}\|_{\infty} = \max_{1 \leq i \leq n} \sum_{j=1}^n |a_{ij}| = 7.$$

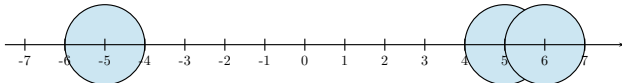
Como $|\lambda| \leq \|\mathbf{A}\|_{\infty}$, entonces todos los eigenvalores de \mathbf{A} están contenidos en un círculo centrado en 0 y radio 7. Con los círculos, se tiene



Ejemplo de los círculos de Gershgorin (II)

Si calculamos la intersección, tenemos

$$C_r \cap C_c$$



Para la matriz dada se tiene que

$$\sigma(\mathbf{A}) = \{5, (1 \pm 5\sqrt{5})/2\} \approx \{5, 6.0902, -5.0902\}.$$

Método de la potencia

- Este método puede encontrar el eigenvalor más grande en valor absoluto y su correspondiente eigenvector.

El algoritmo es el siguiente:

Método de la potencia

Dado un vector inicial \mathbf{v}^0 y fijando $k = 0$, repetimos hasta convergencia los siguientes pasos:

$$\begin{aligned}\mathbf{y}^{k+1} &= \mathbf{A}\mathbf{v}^k \\ \mathbf{v}^{k+1} &= \mathbf{y}^{k+1} / \|\mathbf{y}^{k+1}\| \\ \lambda_{k+1} &= (\mathbf{v}^{k+1})^T \mathbf{A} \mathbf{v}^{k+1} \\ k &= k + 1\end{aligned}$$

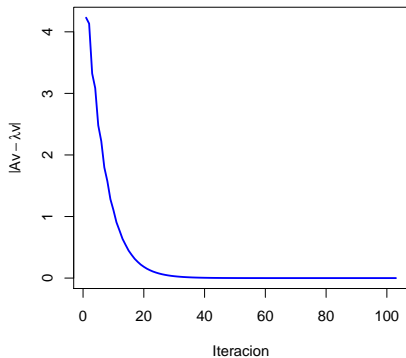
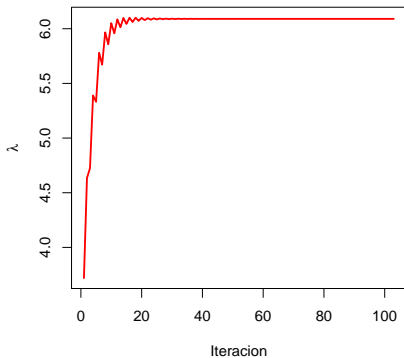
El criterio de convergencia puede ser que el residual sea menor que una cierta tolerancia, $\|\mathbf{y}^k - \lambda_k \mathbf{v}^k\| < \epsilon$.

Ejemplo A

Usando la matriz **A** del ejemplo anterior e inicializando con

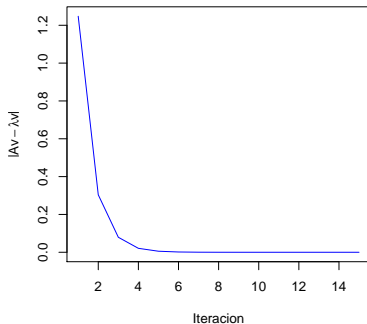
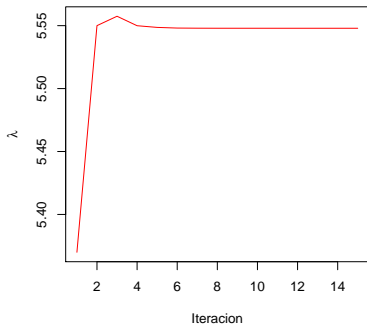
$$\mathbf{v}^0 = (1, 1, 1)^T$$

se obtiene el valor 6.090170 en 103 iteraciones con $\|\mathbf{A}\mathbf{v} - \lambda\mathbf{v}\| \approx 6.28 \times 10^{-8}$.



Ejemplo B

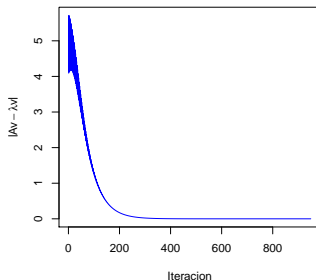
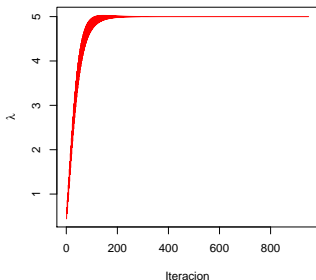
$$\mathbf{A} = \begin{bmatrix} 5.0 & -0.10 & 0.9 & 1.00 & 0.40 \\ -0.5 & 1.45 & -0.05 & 0.00 & 0.25 \\ 0.2 & 0.05 & 1.13 & 0.10 & 0.35 \\ 1.6 & -0.25 & 0.5 & 1.00 & 0.30 \\ 1.4 & 0.40 & 0.2 & 0.25 & -0.80 \end{bmatrix}$$



$\lambda = 5.547928$, $\|\mathbf{A}\mathbf{v} - \lambda\mathbf{v}\| \approx 1.04 \times 10^{-8}$ en 15 iteraciones

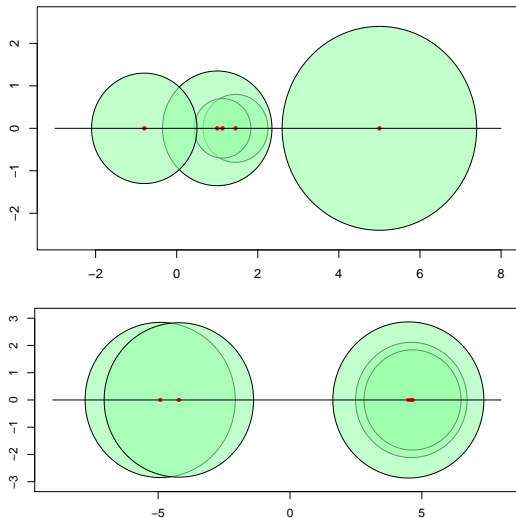
Ejemplo C

$$\mathbf{A} = \begin{bmatrix} 4.6023708 & -0.6484326 & 2.6800333 & 0.1378698 & 0.3655997 \\ -0.3480484 & -4.9229298 & 0.0876574 & -1.2066205 & -1.2046782 \\ 1.0992412 & 0.0206325 & -4.2138133 & -0.3166074 & -1.3973391 \\ 0.5966447 & 2.3193512 & -2.5578133 & 4.6455689 & -0.1206493 \\ -0.0702810 & -1.8568523 & 0.7597306 & -0.1774737 & 4.4888034 \end{bmatrix}$$



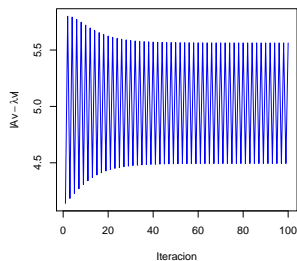
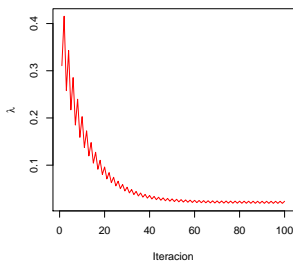
$\lambda = 5.0$, $\|A\mathbf{v} - \lambda\mathbf{v}\| \approx 4.62 \times 10^{-8}$ en 949 iteraciones

Comparación de los ejemplos B y C



Ejemplo D (I)

$$\mathbf{A} = \begin{bmatrix} 4.5986464 & -0.6556048 & 2.7087409 & 0.1380393 & 0.3692491 \\ -0.3513251 & -5.0234452 & 0.0894931 & -1.2194504 & -1.2172484 \\ 1.1110924 & 0.0219267 & -4.3105139 & -0.3201127 & -1.4125578 \\ 0.6010043 & 2.3446976 & -2.5872029 & 4.6477059 & -0.1220816 \\ -0.0718948 & -1.8765426 & 0.7680468 & -0.1796849 & 4.4876068 \end{bmatrix}$$

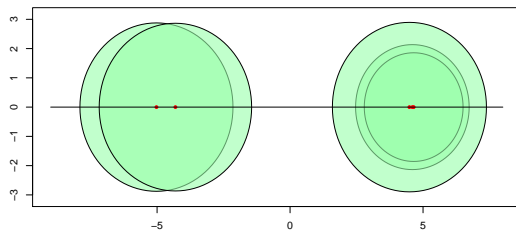


Para $k > 1000$ iteraciones tenemos que

Ejemplo D (II)

$$\lambda_k \approx \begin{cases} 0.023417 & k \text{ par} \\ 0.018916 & k \text{ impar} \end{cases}$$

$$\|\mathbf{A}\mathbf{v} - \lambda\mathbf{v}\|_2 \approx \begin{cases} 5.5630 & k \text{ par} \\ 4.4939 & k \text{ impar} \end{cases}$$



Ejemplo D (III)

Los eigenvalores de las matrices de los ejemplos anteriores se muestran en la siguiente tabla:

Ejemplo	Eigenvalores:
----------------	----------------------

B	−0.9419583, 0.6370984, 1.0520125, 1.4849196, 5.5479279
C	−4.9000000, −4.5000000, 4.4000000, 4.6000000, 5.0000000
D	−5.0000000, −4.6000000, 4.4000000, 4.6000000, 5.0000000

Convergencia del método de la potencia (I)

Supongamos que $(\lambda_i, \mathbf{v}_i)$ es un eigenpar de $\mathbf{A} \in \mathbb{R}^{n \times n}$ con

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

Dado $\mathbf{x}^0 \in \mathbb{R}^n$, se debe tener que $\mathbf{x}^0 = \sum_{i=1}^n \beta_i \mathbf{v}_i$. Así

$$\mathbf{A}^k \mathbf{x}^{(0)} = \sum_{i=1}^n \beta_i \lambda_i^k \mathbf{v}_i.$$

$$\frac{1}{\lambda_1^k} \mathbf{A}^k \mathbf{x}^{(0)} = \beta_1 \mathbf{v}_1 + \sum_{i=2}^n \beta_i \frac{\lambda_i^k}{\lambda_1^k} \mathbf{v}_i.$$

Como $|\lambda_1| > |\lambda_i|$ para $i = 2, \dots, n$, tenemos que

$$\frac{\lambda_i^k}{\lambda_1^k} \longrightarrow 0 \quad \text{si} \quad k \longrightarrow \infty$$

Entonces

$$\frac{1}{\lambda_1^k} \mathbf{A}^k \mathbf{x}^{(0)} = \beta_1 \mathbf{v}_1 + \epsilon^{(k)} \quad \text{con} \quad \epsilon^{(k)} \longrightarrow 0 \quad \text{si} \quad k \longrightarrow \infty$$

Convergencia del método de la potencia (II)

Sea \mathbf{u} un vector tal que $\mathbf{u}^\top \mathbf{v}_1 \neq 0$. Entonces

$$\frac{\mathbf{u}^\top \mathbf{A}^{k+1} \mathbf{x}^{(0)}}{\mathbf{u}^\top \mathbf{A}^k \mathbf{x}^{(0)}} = \frac{\lambda_1^{k+1} (\beta_1 \mathbf{u}^\top \mathbf{v}_1 + \mathbf{u}^\top \boldsymbol{\epsilon}^{(k+1)})}{\lambda_1^k (\beta_1 \mathbf{u}^\top \mathbf{v}_1 + \mathbf{u}^\top \boldsymbol{\epsilon}^{(k)})} \longrightarrow \lambda_1 \quad \text{si } k \longrightarrow \infty$$

Puesto que $\mathbf{A} \mathbf{v}_1 = \lambda_1 \mathbf{v}_1$, entonces

$$\lambda_1 = \frac{\mathbf{v}_1^\top \mathbf{A} \mathbf{v}_1}{\mathbf{v}_1^\top \mathbf{v}_1},$$

una elección natural para el vector \mathbf{u} es

$$\mathbf{u} = \mathbf{A}^k \mathbf{x}^{(0)}$$

De esta forma tenemos que

$$\frac{(\mathbf{A}^k \mathbf{x}^{(0)})^\top \mathbf{A}^{k+1} \mathbf{x}^{(0)}}{(\mathbf{A}^k \mathbf{x}^{(0)})^\top \mathbf{A}^k \mathbf{x}^{(0)}} = \frac{(\mathbf{y}^{(k)})^\top \mathbf{A} \mathbf{y}^{(k)}}{(\mathbf{y}^{(k)})^\top \mathbf{y}^{(k)}} = \frac{(\mathbf{y}^{(k)})^\top \mathbf{A} \mathbf{y}^{(k)}}{\|\mathbf{y}^{(k)}\|^2} = \left(\frac{\mathbf{y}^{(k)}}{\|\mathbf{y}^{(k)}\|} \right)^\top \mathbf{A} \frac{\mathbf{y}^{(k)}}{\|\mathbf{y}^{(k)}\|}$$

Convergencia del método de la potencia (III)

$$\frac{(\mathbf{A}^k \mathbf{x}^{(0)})^\top \mathbf{A}^{k+1} \mathbf{x}^{(0)}}{(\mathbf{A}^k \mathbf{x}^{(0)})^\top \mathbf{A}^k \mathbf{x}^{(0)}} = (\mathbf{v}^{(k)})^\top \mathbf{A} \mathbf{v}^{(k)}$$

Las suposiciones importantes para que método converja son

- Hay un eigenvalor dominante, es decir,

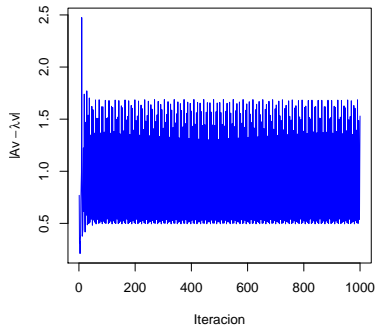
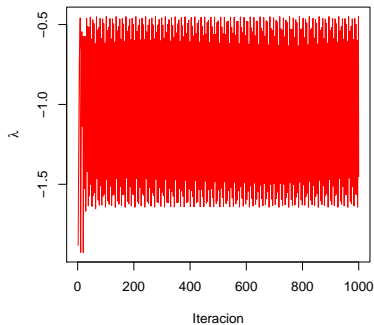
$$|\lambda_1| > |\lambda_i| \quad \text{para } i = 2, \dots, n.$$

- El vector inicial $\mathbf{x}^{(0)}$ no puede ser ortogonal a \mathbf{v}_1 .

Hay que notar otro aspecto importante para el método de la potencia, que se ilustra en el siguiente ejemplo.

Ejemplo E (I)

$$\mathbf{A} = \begin{bmatrix} 0.3 & -0.2 & -0.5 & -0.7 & -0.3 \\ 0.1 & -0.3 & 0.4 & 0.4 & -0.3 \\ -2.9 & -0.6 & -1.1 & -0.9 & 1.0 \\ -1.4 & 0.0 & 0.6 & -0.5 & -0.3 \\ 0.8 & -1.5 & -0.6 & 1.2 & -0.7 \end{bmatrix}$$

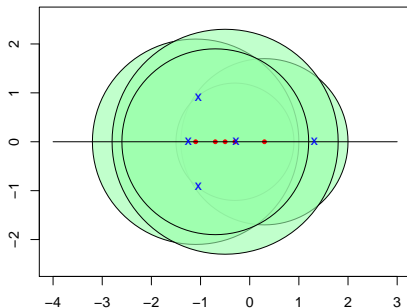


Ejemplo E (II)

$$\|\mathbf{A}\mathbf{v} - \lambda\mathbf{v}\| \approx 1.1 \quad \text{en 1000 iteraciones}$$

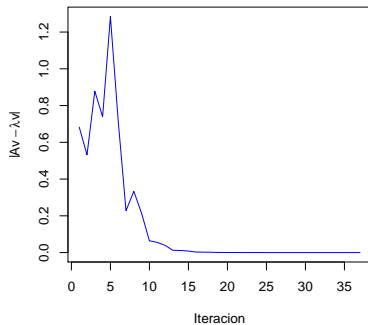
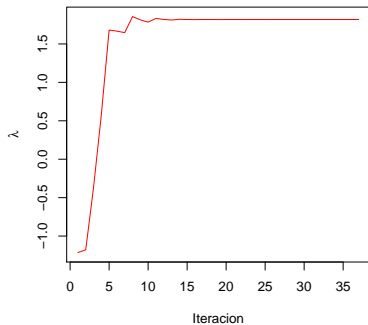
Los eigenvalores del matriz anterior son

$$1.317638, -1.047000 + 0.914528i, -1.047000 - 0.914528i, \\ -1.248337, -0.275302$$



Ejemplo F (I)

$$\mathbf{A} = \begin{bmatrix} 0.8 & -0.2 & -0.5 & -0.7 & -0.3 \\ 0.1 & 0.2 & 0.4 & 0.4 & -0.3 \\ -2.9 & -0.6 & -0.6 & -0.9 & 1.0 \\ -1.4 & 0.0 & 0.6 & 0.0 & -0.3 \\ 0.8 & -1.5 & -0.6 & 1.2 & -0.2 \end{bmatrix}$$

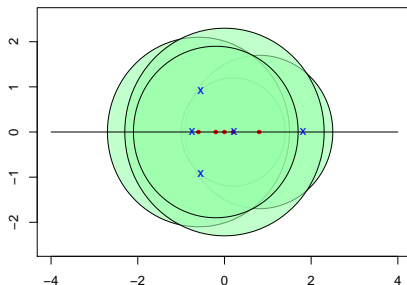


Ejemplo F (II)

$\lambda = 1.817638$, $\|\mathbf{A}\mathbf{v} - \lambda\mathbf{v}\| \approx 4.68 \times 10^{-8}$ en 37 iteraciones

Los eigenvalores del matriz anterior son

1.817638 , $-0.547000 + 0.914528i$, $-0.547000 - 0.914528i$
 -0.748337 , 0.224698



Ventajas y desventajas del método de la potencia (I)

- La principal ventaja del método es que sólo requiere productos matriz por vector para calcular el eigenpar $(\lambda_1, \mathbf{v}_1)$.
- La desventaja es que sólo permite estimar un eigenpar y sólo si $|\lambda_1| > |\lambda_2|$. Además, la velocidad de convergencia depende de que tan cerca está el valor $\left| \frac{\lambda_1}{\lambda_2} \right|$ de 1.