Eigenvalues and Eigenvectors

Let $A \in \mathbb{C}^{n\times n}$, the **eigenvalue problem** consists in finding a scalar λ (real or complex) and a nonnull vector $\mathbf{x}$ such that

$$A\mathbf{x} = \lambda\mathbf{x} \qquad (1)$$

Any such λ is called an **eigenvalue** of A, while $\mathbf{x}$ is the associated **eigenvector**. The latter is not unique; indeed all multiples of $\mathbf{x}$, $\alpha\mathbf{x}$ with $\alpha \neq 0$, are also eigenvectors associated with λ.

Should $\mathbf{x}$ be known, λ can be recovered by using the **Rayleigh quotient**

$$\mathbf{x}^H A\mathbf{x}/||\mathbf{x}||^2 = \bar{\mathbf{x}}^T A\mathbf{x}/||\mathbf{x}||^2.$$

The eigenvalues of A are the roots of the **characteristic polynomial** of A:

$$p_A(\lambda) = \det(A - \lambda I).$$

A $n \times n$ matrix has exactly n eigenvalues (real or complex), not necessarily distinct. A matrix $A \in \mathbb{C}^{n\times n}$ is said to be **diagonalizable** if there exists a nonsingular matrix $U \in \mathbb{C}^{n\times n}$ such that

$$U^{-1}AU = \Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_n). \qquad (2)$$

The columns of U are the eigenvectors of A. If A is diagonal or triangular, the λ's are its diagonal entries; otherwise, if A is a general large matrix, seeking the zeros of p_A is hard.

The power method

Let $A \in \mathbb{R}^{n\times n}$, so with real entries, and assume that its eigenvalues are ordered

$$|\lambda_1| = |\lambda_{\max}| > |\lambda_2| \geq |\lambda_3| \geq \ldots \geq |\lambda_n| = |\lambda_{\min}|.$$

The **power method** can find the greater eigenvalue $\lambda_{\max} = \lambda_1$ of a non-singular matrix A.

Let $\mathbf{x}^{(0)} \in \mathbb{C}^n$ be given and set $\mathbf{y}^{(0)} = \mathbf{x}^{(0)}/||\mathbf{x}^{(0)}||$, for $k = 1, 2, \ldots$ compute

$$\boxed{\begin{aligned} \mathbf{x}^{(k)} &= A\mathbf{y}^{(k-1)} \\ \mathbf{y}^{(k)} &= \frac{\mathbf{x}^{(k)}}{||\mathbf{x}^{(k)}||} \\ \lambda^{(k)} &= (\mathbf{y}^{(k)})^H A\mathbf{y}^{(k)} \end{aligned}}$$

until $|\lambda^{(k)} - \lambda^{(k-1)}| < \varepsilon|\lambda^{(k)}|$ (stopping condition), where ε is the desired tolerance.

By induction on k one can check that

$$\mathbf{y}^{(k)} = \frac{\mathbf{x}^{(k)}}{||\mathbf{x}^{(k)}||} = \frac{A\mathbf{y}^{(k-1)}}{||A\mathbf{y}^{(k-1)}||} = \ldots = \frac{A^k\mathbf{y}^{(0)}}{||A^k\mathbf{y}^{(0)}||}, \quad k \geq 1.$$

This relation explains the role played by the powers of A in the method: the term $\mathbf{y}^{(k)}$ can also be expressed as a power, thus the name of the method.

This method generates a sequence of unitary vectors $\{\mathbf{y}^{(k)}\}$ such that for $k \to \infty$ they align in the direction of the eigenvector $\mathbf{x}_1$. In all cases, we have that $\lambda^{(k)} \to \lambda_{\max} = \lambda_1$ for $k \to \infty$.

Convergence

Since the eigenvectors $\mathbf{x}_1, \ldots, \mathbf{x}_n$ of A are linearly independent, these eigenvectors form a basis for $\mathbb{C}^n$. Thus the vectors $\mathbf{x}^{(0)}$ and $\mathbf{y}^{(0)}$ can be written as

$$\mathbf{x}^{(0)} = \sum_{i=1}^{n} \alpha_i \mathbf{x}_i, \qquad \mathbf{y}^{(0)} = \beta^{(0)} \sum_{i=1}^{n} \alpha_i \mathbf{x}_i, \qquad \text{with } \beta^{(0)} = 1/||\mathbf{x}^{(0)}|| \text{ and } \alpha_i \in \mathbb{C}.$$

At the first step the power method gives

$$\mathbf{x}^{(1)} = A\mathbf{y}^{(0)} = \beta^{(0)} A \sum_{i=1}^{n} \alpha_i \mathbf{x}_i = \beta^{(0)} \sum_{i=1}^{n} \alpha_i \lambda_i \mathbf{x}_i \quad \text{and, similarly,}$$

$$\mathbf{y}^{(1)} = \beta^{(1)} \sum_{i=1}^{n} \alpha_i \lambda_i \mathbf{x}_i, \quad \beta^{(1)} = \frac{1}{||\mathbf{x}^{(0)}|| \cdot ||\mathbf{x}^{(1)}||}.$$

At a given step k we will have

$$\mathbf{y}^{(k)} = \beta^{(k)} \sum_{i=1}^{n} \alpha_i \lambda_i^k \mathbf{x}_i, \qquad \beta^{(k)} = \frac{1}{||\mathbf{x}^{(0)}|| \cdots ||\mathbf{x}^{(k)}||} = \frac{1}{\prod_{i=0}^{k} ||\mathbf{x}^{(i)}||}$$

And therefore

$$\mathbf{y}^{(k)} = \lambda_1^k \beta^{(k)} \left(\alpha_1 \mathbf{x}_1 + \sum_{i=2}^{n} \alpha_i \frac{\lambda_i^k}{\lambda_1^k} \mathbf{x}_i \right).$$

Since $|\lambda_i/\lambda_1| < 1$ for $i = 2, \ldots, n$, the vector $\mathbf{y}^{(k)}$ tends to align along the same direction as the eigenvector $\mathbf{x}_1$ when k tends to $+\infty$, provided $\alpha_1 \neq 0$.

The inverse power method

The **inverse power method** can find the smaller eigenvalue of a non-singular matrix A. It is like the previous method, but if A is non-singular we can use A^{-1}, which eigenvalues are the reciprocal of those of A, to obtain the eigenvalue of A with minimum modulus.

Let $\mathbf{x}^{(0)} \in \mathbb{C}^n$ be given and set $\mathbf{y}^{(0)} = \mathbf{x}^{(0)}/||\mathbf{x}^{(0)}||$, for $k = 1, 2, \ldots$ compute

$$\boxed{\begin{aligned} \mathbf{x}^{(k)} &= A^{-1}\mathbf{y}^{(k-1)} \\ \mathbf{y}^{(k)} &= \frac{\mathbf{x}^{(k)}}{||\mathbf{x}^{(k)}||} \\ \mu^{(k)} &= (\mathbf{y}^{(k)})^H A^{-1} \mathbf{y}^{(k)} \end{aligned}}$$

until $|\mu^{(k)} - \mu^{(k-1)}| < \varepsilon |\mu^{(k)}|$, where ε is the desired tolerance.

If A admits n linearly independent eigenvectors, and if also the eigenvalue $\lambda_{\min} = \lambda_n$ of minimum modulus is distinct from the others, then

$$\lim_{k \to \infty} \mu^{(k)} = 1/\lambda_{\min},$$

At each step k we have to solve a linear system of the form $A\mathbf{x}^{(k)} = \mathbf{y}^{(k-1)}$, and we can use for example the LU factorization or the Cholesky factorization.

The power method with shift

The power method with shift can find the eigenvalue of A near to a given number μ.

Define $A_\mu = A - \mu I$, whose eigenvalues are $\lambda(A_\mu) = \lambda(A) - \mu$. In order to approximate λ_μ, we can at first approximate the eigenvalue of minimum length of A_μ with the inverse power method. Let $\mathbf{x}^{(0)} \in \mathbb{C}^n$ be given and set $\mathbf{y}^{(0)} = \mathbf{x}^{(0)}/||\mathbf{x}^{(0)}||$, for $k = 1, 2, \dots$ compute

$$\boxed{\begin{aligned} \mathbf{x}^{(k)} &= A_\mu^{-1}\mathbf{y}^{(k-1)} \\ \mathbf{y}^{(k)} &= \frac{\mathbf{x}^{(k)}}{||\mathbf{x}^{(k)}||} \\ \lambda_\mu^{(k)} &= \frac{1}{(\mathbf{y}^{(k)})^H A_\mu^{-1}\mathbf{y}^{(k)}} \end{aligned}}$$

until $|\lambda_\mu^{(k)} - \lambda_\mu^{(k-1)}| < \varepsilon|\lambda_\mu^{(k)}|$, where ε is the desired tolerance. The searched eigenvalue of A is approximated by $\lambda = \lambda_\mu + \mu$.

The value of the shift can be modified during the iterations, by setting $\mu = \lambda^{(k)}$. This yields a faster convergence; however the computational cost grows substantially since now at each iteration the matrix A_μ does change and the LU factorization has to be performed at each iteration.

How to compute the shift

We need to locate (more or less accurately) the eigenvalues of A in the complex plane.

Let $A \in \mathbb{C}^{n\times n}$ be a square matrix of dimension n. The **Gershgorin circles** $C_i^{(r)}$ and $C_i^{(c)}$ associated with its i-th row and i-th column are respectively defined as

$$\mathcal{R}_i = C_i^{(r)} = \left\{ z \in \mathbb{C} : |z - a_{ii}| \le \sum_{j=1, j\neq i}^{n} |a_{ij}| \right\},$$

$$\mathcal{C}_i = C_i^{(c)} = \left\{ z \in \mathbb{C} : |z - a_{ii}| \le \sum_{j=1, j\neq i}^{n} |a_{ji}| \right\}.$$

$\mathcal{R}_i$ is called the i-th **row circle** and $\mathcal{C}_i$ the i-th **column circle**.

First Gershgorin theorem. All the eigenvalues of a given matrix $A \in \mathbb{C}^{n\times n}$ belong to the region of the complex plane which is the intersection of the two regions formed respectively by the union of the row circles $\mathcal{S}_\mathcal{R} = \bigcup_{i=1}^n \mathcal{R}_i$ and the union of the column circles $\mathcal{S}_\mathcal{C} = \bigcup_{i=1}^n \mathcal{C}_i$:

$$\forall \lambda \in \sigma(A), \quad \lambda \in \mathcal{S}_\mathcal{R} \bigcap \mathcal{S}_\mathcal{C}.$$

Moreover, should m row circles (or column circles), with $1 \le m \le n$, be disconnected from the union of the remaining $n - m$ circles, then their union contains exactly m eigenvalues.

There is no guarantee that a circle should contain eigenvalues, unless it is isolated from the others. The information provided by Gershgorin circles are in general quite coarse, thus the previous result can provide only a preliminary guess of the shift.

Remark. Note that all the eigenvalues of a strictly diagonally dominant matrix are non-null.

The QR method

We will see some iterative techniques for simultaneously approximating all the eigenvalues of a given matrix A. The basic idea consists of reducing A, by means of suitable similarity transformations, into a form for which the calculation of the eigenvalues is easier than on the starting matrix.

If A and B are similar $P^{-1}AP = B$ then $\lambda_A = \lambda_B$

$$BP^{-1}\mathbf{x} = P^{-1}A\mathbf{x} = \lambda P^{-1}\mathbf{x}$$

A method to compute all the eigenvalues of A is transforming it in a similar diagonal / triangular matrix.

The QR method uses repeatedly the QR factorization to compute λ.

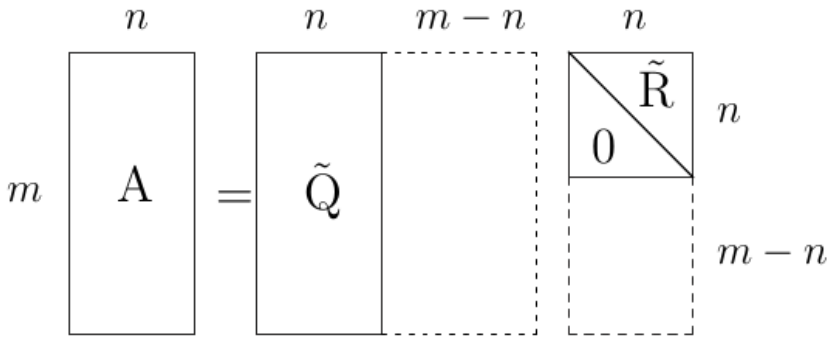

Fig. 3.1. The reduced factorization. The matrices of the QR factorization are drawn in dashed lines

$$\implies Q^{(k+1)}R^{(k+1)} = A^{(k)} \implies A^{(k+1)} = R^{(k+1)}Q^{(k+1)}$$

$A^{(k)}$ and $A^{(k+1)}$ are similar and the rate of decay to zero of lower triangular coefficients in $A^{(k)}$ depends on $\max_i |\lambda_{i+1}/\lambda_i|,\ \forall i$. If A is symmetric, $A^{(k)}$ for $k \to \infty$ is diagonal.