Eigenvalues & Eigenvectors

In physics, needs for calculating eigenvalues and eigenvectors appear in widely varying systems.

In importance, perhaps it comes next to Monte Carlo methods. Inversion and eigenvalues are two most important programs of any numerical linear algebra packages like LAPACK.

Standard way to express an eigenvalue problem of a $n \times n$ matrix $\mathbf{A}$ is,

$$\mathbf{A} \cdot \mathbf{x} = \lambda \mathbf{x} \quad \text{or} \quad \mathbf{A}\,\mathbf{x}_\nu = \lambda_\nu \mathbf{x}_\nu, \quad \forall \nu \in \{1, 2, \ldots, n\}$$

where λ_ν, $\mathbf{x}_\nu$ are eigenvalues and eigenvectors respectively. Analytically, the eigenvalue problem is solved by finding the roots of the characteristic polynomial $p_A(\lambda)$ of degree n of $\mathbf{A}$,

$$p_A(\lambda) = \det\left(\mathbf{A} - \lambda_\nu \mathbb{1}\right) = 0$$

The roots may or may not be real or distinct. However, this method is never recommended for numerical determination of λ_ν.

A word of wisdom – *in general e-values and e-vectors of a matrix cannot be determined using a finite number of rational operations. They have to be approximate!*

Solving eigenvalue problems numerically is a difficult undertaking with respect to stability, convergence and computing time.

Generally makes use of various properties of matrix **A**, such as symmetric, Hermitian, normal, unitary, tridiagonal, Hessenberg, sparse etc. and design algorithms accordingly.

Choice of algorithm is also affected by whether need is for all or only a few e-values, need e-vectors or not and so on. Other concerns are degenerate e-values, zero or extreme e-values and many such. There is no one-size-fits-all eigen algorithm.

Important point is that *small changes in* **A** *do not necessarily lead to small changes in the e-values*.

$$\mathbf{A} = \begin{pmatrix} 1 & 1000 \\ 0 & 1 \end{pmatrix} \quad \text{and} \quad \lambda_A = 1,\ 1$$

$$\mathbf{B} = \begin{pmatrix} 1 & 1000 \\ 0.001 & 1 \end{pmatrix} \quad \text{and} \quad \lambda_B = 0,\ 2$$

However, if **A** is symmetric, small changes do not lead to large changes in e-values. Hence, algorithms dealing with symmetric matrices are in general successful.

Power iteration method

An e-value algorithm which returns the largest $|\lambda|$ and the corresponding eigenvector v. It is rather simple and converges slowly. It is used when

1. $n \times n$ matrix $\mathbf{A}$ has n linearly independent e-vectors, and
2. e-values can be ordered in magnitude : $|\lambda_1| > |\lambda_2| > \cdots > |\lambda_n|$. The λ_1 is called dominant e-value and the corresponding e-vector is dominant e-vector of $\mathbf{A}$.

This algorithm was part of the Google's first version of PageRank algorithm (ref. Wikipedia).

Let $\{\mathbf{v}_0, \mathbf{v}_1, \ldots, \mathbf{v}_n\}$ is a set of linearly independent e-vectors. If $\mathbf{x}_0$ be initial guess for dominant e-vector, then

$$\mathbf{x}_0 = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n$$

k repeated multiplication by $\mathbf{A}$ yields,

$$\begin{aligned}
\mathbf{x}_1 &= \mathbf{A}\mathbf{x}_0 = c_1\lambda_1\mathbf{v}_1 + c_2\lambda_2\mathbf{v}_2 + \cdots + c_n\lambda_n\mathbf{v}_n \\
\mathbf{x}_2 &= \mathbf{A}^2\mathbf{x}_0 = c_1\lambda_1^2\mathbf{v}_1 + c_2\lambda_2^2\mathbf{v}_2 + \cdots + c_n\lambda_n^2\mathbf{v}_n \\
&\vdots \\
\mathbf{x}_k &= \mathbf{A}^k\mathbf{x}_0 = c_1\lambda_1^k\mathbf{v}_1 + c_2\lambda_2^k\mathbf{v}_2 + \cdots + c_n\lambda_n^k\mathbf{v}_n
\end{aligned}$$

Dividing the last equation by λ_1^k we get,

$$\frac{\mathbf{x}_k}{\lambda_1^k} = \frac{1}{\lambda_1^k}\mathbf{A}^k\mathbf{x}_0 = c_1\mathbf{v}_1 + \left(\frac{\lambda_2}{\lambda_1}\right)^k c_2\mathbf{v}_2 + \cdots + \left(\frac{\lambda_n}{\lambda_1}\right)^k c_n\mathbf{v}_n \approx c_1\mathbf{v}_1$$

$\mathbf{x}_0$ has to be so chosen such that it is not orthogonal to $\mathbf{v}_1$ which ensures that $c_1 \neq 0$. The next $k+1$-th multiplication gives,

$$\frac{\mathbf{x}_{k+1}}{\lambda_1^{k+1}} = \frac{1}{\lambda_1^{k+1}}\mathbf{A}^{k+1}\mathbf{x}_0 = c_1\mathbf{v}_1$$

Let $\mathbf{y}$ be any vector not orthogonal to $\mathbf{v}_1$ (can be identical to $\mathbf{x}_0$), then

$$\begin{aligned}\frac{\mathbf{x}_k \cdot \mathbf{y}}{\lambda_1^k} &= \frac{1}{\lambda_1^k}\mathbf{A}^k\mathbf{x}_0 \cdot \mathbf{y} = c_1\mathbf{v}_1 \cdot \mathbf{y} \\ \frac{\mathbf{x}_{k+1} \cdot \mathbf{y}}{\lambda_1^{k+1}} &= \frac{1}{\lambda_1^{k+1}}\mathbf{A}^{k+1}\mathbf{x}_0 \cdot \mathbf{y} = c_1\mathbf{v}_1 \cdot \mathbf{y}\end{aligned}$$

Equating the two, we obtain the *Rayleigh quotient* formula for determining the largest eigenvalue λ_1,

$$\frac{\mathbf{x}_{k+1} \cdot \mathbf{y}}{\mathbf{x}_k \cdot \mathbf{y}} = \frac{\mathbf{A}^{k+1}\mathbf{x}_0 \cdot \mathbf{y}}{\mathbf{A}^k\mathbf{x}_0 \cdot \mathbf{y}} = \lambda_1$$

Iteration stops when $|\lambda_1^{(k-1)} - \lambda_1^{(k)}| < \epsilon$. Convergence to *true* λ_1 depends on how $(\lambda_2/\lambda_1)^k$ is decreasing with k.

Dominant e-vector $\mathbf{v}_1$ is approximated by

$$\mathbf{v}_1 \approx \mathbf{A}^k \mathbf{x}_0$$

To see this consider the equation for dominant e-value λ_1,

$$\mathbf{A}\,\mathbf{v}_1 = \lambda_1\,\mathbf{v}_1 \quad \Rightarrow \quad \lambda_1 = \frac{\mathbf{A}\mathbf{v}_1 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}$$

If we choose $\mathbf{y} = \mathbf{A}^k \mathbf{x}_0$, then

$$\lambda_1 = \frac{\mathbf{A}^{k+1}\mathbf{x}_0 \cdot \mathbf{y}}{\mathbf{A}^k\mathbf{x}_0 \cdot \mathbf{y}} = \frac{\mathbf{A}\,\mathbf{A}^k\mathbf{x}_0 \cdot \mathbf{A}^k\mathbf{x}_0}{\mathbf{A}^k\mathbf{x}_0 \cdot \mathbf{A}^k\mathbf{x}_0} \equiv \frac{\mathbf{A}\mathbf{v}_1 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}$$

Although it is not recommended, but one can attempt to find non-dominant e-values by defining,

$$\boldsymbol{\mathcal{A}} = \mathbf{A} - \lambda_1 \mathbf{U}_1 \mathbf{U}_1^T, \quad \text{where,} \quad \mathbf{U}_1 = \mathbf{v}_1/|\mathbf{v}_1|$$

which has e-values $0, \lambda_2, \lambda_3, \ldots, \lambda_n$ and e-vectors of $\boldsymbol{\mathcal{A}}$ are e-vectors of $\mathbf{A}$.

Depending on accuracy achieved for λ_1, there can be large errors that can creep into λ_2.

Pressing this method for e-values beyond λ_2 is absolutely inadvisable.

Find the dominant e-value and its corresponding e-vector for

$$\mathbf{A} = \begin{pmatrix} 1 & -1 & 0 \\ -2 & 4 & -2 \\ 0 & -1 & 2 \end{pmatrix}$$

$$\lambda = 5.125,\ 1.637,\ 0.238$$

$$\mathbf{v} = \begin{pmatrix} 0.22 \\ -0.93 \\ 0.30 \end{pmatrix}, \quad \begin{pmatrix} -0.47 \\ 0.30 \\ 0.83 \end{pmatrix}, \quad \begin{pmatrix} 0.75 \\ 0.57 \\ 0.32 \end{pmatrix} \quad \text{(normalized)}$$