Eigenvalue Problem

One of the most widely used procedures to estimate eigenvalues of a matrix A is the power method. Our basic assumptions are that A is a real $n \times n$ matrix, its eigen values $\lambda_1, \lambda_2, \cdots, \lambda_n$ are numbered so that

$$|\lambda_1| > |\lambda_2| \geq |\lambda_3| \geq \cdots \geq |\lambda_n| \quad \text{---(1)}$$

and that A has n linearly independent eigenvectors $\{\underset{\sim}{u}_i\}_{i=1}^{n}$, where $A\underset{\sim}{u}_i = \lambda_i \underset{\sim}{u}_i$, $1 \leq i \leq n$. When the power method works, it gives us an estimate of the dominant eigenvalue, λ_1.

We first need the fact that any $n \times 1$ vector can be expressed as a linear combination of the eigen vectors $\underset{\sim}{u}_1, \underset{\sim}{u}_2, \cdots, \underset{\sim}{u}_n$. To see this, let B be the $(n \times n)$ matrix defined such that the k^{th} column vector of B is $\underset{\sim}{u}_k$, $1 \leq k \leq n$, that is $B = [\underset{\sim}{u}_1, \underset{\sim}{u}_2, \cdots, \underset{\sim}{u}_n]$. Since $\underset{\sim}{u}_i$'s are linearly independent, B is invertible. Now, let $\underset{\sim}{x}_0$ be an arbitrary $(n \times 1)$ vector, and consider the linear system of equations $B\underset{\sim}{x} = \underset{\sim}{x}_0$. Since B is invertible, this system has a unique solution, $\underset{\sim}{x} = (\alpha_1, \alpha_2, \cdots, \alpha_n)^T$, and so $\underset{\sim}{x}_0$ can be uniquely expressed as

$$\underset{\sim}{x}_0 = \alpha_1 \underset{\sim}{u}_1 + \alpha_2 \underset{\sim}{u}_2 + \cdots + \alpha_n \underset{\sim}{u}_n \quad \text{---(2)}$$

Next, observe that if $A\underset{\sim}{u}_i = \lambda_i \underset{\sim}{u}_i$ then $A(A\underset{\sim}{u}_i) = \lambda_i A\underset{\sim}{u}_i$ or $A^2\underset{\sim}{u}_i = \lambda_i^2 \underset{\sim}{u}_i$. Continuing, it follows that $A^3\underset{\sim}{u}_i = \lambda_i^3 \underset{\sim}{u}_i$, or in general

$$A^k \underset{\sim}{u}_i = \lambda_i^k \underset{\sim}{u}_i \quad \text{---(3)}$$

Thus if we form the sequence

$$\underset{\sim}{x}_1 = A\underset{\sim}{x}_0$$

$$\underset{\sim}{x}_2 = A\underset{\sim}{x}_1 \,, \quad (\underset{\sim}{x}_2 = A^2\underset{\sim}{x}_0)$$

$$\underset{\sim}{x}_3 = A\underset{\sim}{x}_2 \,, \quad (\underset{\sim}{x}_3 = A^3\underset{\sim}{x}_0)$$

$$\underset{\sim}{x}_k = A\underset{\sim}{x}_{k-1} \,, \quad (\underset{\sim}{x}_k = A^k\underset{\sim}{x}_0)$$

$$\therefore \underset{\sim}{x}_k = A^k(\alpha_1\underset{\sim}{u}_1 + \alpha_2\underset{\sim}{u}_2 + \cdots + \alpha_n\underset{\sim}{u}_n)$$

$$= \alpha_1(A^k\underset{\sim}{u}_1) + \alpha_2(A^k\underset{\sim}{u}_2) + \cdots + \alpha_n(A^k\underset{\sim}{u}_n)$$

$$= \alpha_1\lambda_1^k\underset{\sim}{u}_1 + \alpha_2\lambda_2^k\underset{\sim}{u}_2 + \cdots + \alpha_n\lambda_n^k\underset{\sim}{u}_n \quad (\text{by (3)}) \qquad \text{—(4)}$$

Recalling (1), it is natural to write $\underset{\sim}{x}_k$ as

$$\underset{\sim}{x}_k = \lambda_1^k\left\{\alpha_1\underset{\sim}{u}_1 + \alpha_2\left(\frac{\lambda_2}{\lambda_1}\right)^k\underset{\sim}{u}_2 + \cdots + \alpha_n\left(\frac{\lambda_n}{\lambda_1}\right)^k\underset{\sim}{u}_n\right\} \qquad \text{—(5)}$$

In (5), we see for large k that the approximation $\underset{\sim}{x}_k \simeq \lambda_1^k\alpha_1\underset{\sim}{u}_1$ should be good since $|\lambda_1| > |\lambda_2| \geq |\lambda_3| \geq \cdots \geq |\lambda_n|$.

In order to estimate λ_1, we replace k by $(k+1)$ in (5) to obtain $\underset{\sim}{x}_{k+1} \simeq \lambda_1^{k+1}\alpha_1\underset{\sim}{u}_1$, or $\underset{\sim}{x}_{k+1} \simeq \lambda_1\underset{\sim}{x}_k$. This suggests a number of ways to approximate λ_1. For example, $\underset{\sim}{x}_k^T\underset{\sim}{x}_{k+1} \simeq \lambda_1\underset{\sim}{x}_k^T\underset{\sim}{x}_k$ & thus $\lambda_1 \simeq \underset{\sim}{x}_k^T\underset{\sim}{x}_{k+1}/\underset{\sim}{x}_k^T\underset{\sim}{x}_k$. Another common procedure is to divide the largest component of $\underset{\sim}{x}_k$ into the corresponding component of $\underset{\sim}{x}_{k+1}$. If the largest component is the j^{th} component, then this amounts to $\lambda_1 \simeq \underset{\sim}{e}_j^T\underset{\sim}{x}_{k+1}/\underset{\sim}{e}_j^T\underset{\sim}{x}_k$.

Since $\underset{\sim}{x}_k \simeq \lambda_1^k\alpha_1\underset{\sim}{u}_1$ i.e. $\underset{\sim}{x}_k$ differs from the eigen vector $\underset{\sim}{u}_1$ solely by a numerical factor and, consequently, is also an eigen vector corresponding to the same eigenvalue λ_1.

[Proof. $A\underset{\sim}{x}_k = A\cdot\lambda_1^k\alpha_1\underset{\sim}{u}_1 = \lambda_1^k\alpha_1 A\underset{\sim}{u}_1 = \lambda_1^k\alpha_1(\lambda_1\underset{\sim}{u}_1)$

$= \lambda_1\underset{\sim}{x}_k$

$\therefore \underset{\sim}{x}_k$ is an eigen vector corresponding to λ_1]

We wish to be fairly careful about the description of the power method, for this description has two parts, namely the 'practical' part which shows how the method is actually carried out, and the 'theoretical' part which shows why the method works, what can go wrong, and how to interpret the output. First of all, powers of A are never actually computed in the implementation of the power method, nor is it necessary to know the a_i's and the u_i's in (2). We merely guess a vector x_0 & form the sequence $\{x_k\}$ from $x_k = A x_{k-1}$; $k = 1, 2, 3, \cdots$. The equation $x_k = A^k x_0$ is the theoretical basis for the power method, but all we really need is the vector x_k. The estimates for λ_1 can be found by generating the sequence of scalars $\{\beta_k\}$ from the equation

$$\beta_k = v^T x_{k+1} / v^T x_k \quad - - \quad (6)$$

where v is usually x_k or the r^{th} unit vector e_r (where the r^{th} component of x_k is the dominant one). When programming the power method, it is a little easier to generate the estimate β_k from $\beta_k = x_k^T x_{k+1} / x_k^T x_k$ than to include a test at each step to determine the maximum component of x_k. The advantage in finding the maximum component of x_k is that round-off errors will tend to be reduced. Finally as k increases, the vectors x_k are usually growing quite large if $|\lambda_1| > 1$ (or quite small if $|\lambda_1| < 1$), as can be seen from (4). So it is desirable to 'scale' or 'normalize' x_k to keep it within the computational limits. This can be done by dividing x_k by $\|x_k\|$ at each step. We now show that this has no effect on the convergence.

Let x_0 be given by (2), where $\|x_0\| = 1$. Let $y_i = A x_{i-1}$, $\beta_i = v^T y_i / v^T x_{i-1}$, $x_i = y_i / \|y_i\|$, for $i = 1, 2, \cdots$.

Put $\underset{\sim}{w}_0 = \underset{\sim}{x}_0$ & $\underset{\sim}{w}_i = A\underset{\sim}{w}_{i-1}$, $i = 1, 2, \ldots$.

$\therefore \underset{\sim}{y}_1 = A\underset{\sim}{x}_0 = \underset{\sim}{w}_1$.

$\underset{\sim}{w}_2 = A\underset{\sim}{w}_1 = A^2\underset{\sim}{x}_0 = A^2\underset{\sim}{w}_0$.

$\underset{\sim}{w}_3 = A\underset{\sim}{w}_2 = A^3\underset{\sim}{w}_0 = A^3\underset{\sim}{x}_0$.

& in general, $\underset{\sim}{w}_i = A^i\underset{\sim}{x}_0 = A^i\underset{\sim}{w}_0$.

Also $\underset{\sim}{y}_2 = A\underset{\sim}{x}_1 = A\,\underset{\sim}{y}_1/\|y_1\| = \dfrac{A^2\underset{\sim}{w}_0}{\|y_1\|}$

$\underset{\sim}{y}_3 = A\underset{\sim}{x}_2 = A\cdot\dfrac{\underset{\sim}{y}_2}{\|y_2\|} = \dfrac{A^3\underset{\sim}{w}_0}{\|y_1\|\cdot\|y_2\|}$

& in general, $\underset{\sim}{y}_i = \dfrac{A^i\underset{\sim}{w}_0}{\|y_1\|\,\|y_2\| \cdots \|y_{i-1}\|}$.

Now $\dfrac{\underset{\sim}{v}^T\underset{\sim}{y}_i}{\underset{\sim}{v}^T\underset{\sim}{x}_{i-1}} = \dfrac{\underset{\sim}{v}^T\dfrac{A^i\underset{\sim}{w}_0}{\|y_1\|\,\|y_2\|\cdots\|y_{i-1}\|}}{\underset{\sim}{v}^T\dfrac{\underset{\sim}{y}_{i-1}}{\|y_{i-1}\|}}$

$= \dfrac{\underset{\sim}{v}^T\dfrac{A^i\underset{\sim}{w}_0}{\|y_1\|\,\|y_2\|\cdots\|y_{i-2}\|}}{\underset{\sim}{v}^T\dfrac{A^{i-1}\underset{\sim}{w}_0}{\|y_1\|\,\|y_2\|\cdots\|y_{i-2}\|}}$

$= \underset{\sim}{v}^TA^i\underset{\sim}{w}_0 \,/\, \underset{\sim}{v}^TA^{i-1}\underset{\sim}{w}_0$

$= \underset{\sim}{v}^T\underset{\sim}{w}_i \,/\, \underset{\sim}{v}^T\underset{\sim}{w}_{i-1}$

which is similar to $\beta_K = \underset{\sim}{v}^T\underset{\sim}{x}_{K+1} / \underset{\sim}{v}^T\underset{\sim}{x}_K$.

Thus, normalizing $\underset{\sim}{x}_i$ in each step does not effect the convergence.

We like to mention here that $\underset{\sim}{x}_{K+1}$ $(= A\underset{\sim}{x}_K)$ should be a finite vector if $\underset{\sim}{x}_K$ is normalized.

In order to study the theoretical convergence of the power method, it is most convenient to consider a variation of the method in which β_K are calculated in a slightly different fashion from the two ways described above. In this variation, we assume $\underset{\sim}{v}$ is a fixed vector

such that $\underset{\sim}{v}^T \underset{\sim}{u}_1 \neq 0$. We then define β_k to be given by $\beta_k = \underset{\sim}{v}^T \underset{\sim}{x}_{k+1} / \underset{\sim}{v}^T \underset{\sim}{x}_k$. If we set $v_j = \underset{\sim}{v}^T \underset{\sim}{u}_j$, equation (5) shows that

$$\beta_k = \lambda_1 \frac{\alpha_1 v_1 + \alpha_2 v_2 \left(\frac{\lambda_2}{\lambda_1}\right)^{k+1} + \cdots + \alpha_n v_n \left(\frac{\lambda_n}{\lambda_1}\right)^{k+1}}{\alpha_1 v_1 + \alpha_2 v_2 \left(\frac{\lambda_2}{\lambda_1}\right)^{k} + \cdots + \alpha_n v_n \left(\frac{\lambda_n}{\lambda_1}\right)^{k}} \qquad \text{——(7)}.$$

The theoretical result (7) reveals that the sequence $\{\beta_k\}$ converges to λ_1 at essentially the same rate as $\left(\frac{\lambda_2}{\lambda_1}\right)^k$ converges to zero.

If we generate the sequence $\{\beta_k\}$ by $\beta_k = \underset{\sim}{x}_k^T \underset{\sim}{x}_{k+1} / \underset{\sim}{x}_k^T \underset{\sim}{x}_k$ where $\underset{\sim}{x}_k$ has been normalized at each step, then also $v_j = \underset{\sim}{x}_k^T \underset{\sim}{u}_j$ are finite quantities and also converges to λ_1, since the analysis is similar to that in (7).

When β_k is calculated by the formula

$$\beta_k = \underset{\sim}{l}_j^T \underset{\sim}{x}_{k+1} / \underset{\sim}{l}_j^T \underset{\sim}{x}_k$$

the convergence of $\{\beta_k\}$ to λ_1 can be shown in the following manner.

We have by (4)

$$\underset{\sim}{x}_k = \alpha_1 \lambda_1^k \underset{\sim}{u}_1 + \alpha_2 \lambda_2^k \underset{\sim}{u}_2 + \cdots + \alpha_n \lambda_n^k \underset{\sim}{u}_n$$

$$= \sum_{j=1}^{n} \alpha_j \lambda_j^k \underset{\sim}{u}_j \qquad \cdots (4').$$

Let $\underset{\sim}{x}_k = (x_1^k, x_2^k, \ldots, x_n^k)^T$ and choose a basis $\{\underset{\sim}{l}_1', \underset{\sim}{l}_2', \ldots, \underset{\sim}{l}_n'\}$ in E_n.

Decomposing the eigen vectors $\underset{\sim}{u}_j$ into the vectors of the basis, we have

$$\underset{\sim}{u}_j = \sum_{i=1}^{n} u_i^j \underset{\sim}{l}_i', \text{ where } \underset{\sim}{u}_j = (u_1^j, u_2^j, \ldots, u_n^j). \qquad \text{——(8)}.$$

Substituting (8) into (4'), we obtain

$$\underset{\sim}{x}_k = \sum_{j=1}^{n} \alpha_j \lambda_j^k \sum_{i=1}^{n} u_i^j \underset{\sim}{l}_i'$$

or changing the order of summation

$$\underset{\sim}{x}_k = \sum_{i=1}^{n} \underset{\sim}{l}_i' \sum_{j=1}^{n} \alpha_j u_i^j \lambda_j^k \quad \text{---(9)}$$

Let the i^{th} component of $\underset{\sim}{x}_k$ is the largest component.

Equating the i^{th} coordinate of $\underset{\sim}{x}_k$ from (9), we have

$$x_i^k = \sum_{j=1}^{n} \alpha_j u_i^j \lambda_j^k \quad \text{---(9')}$$

Similarly, $x_i^{k+1} = \sum_{j=1}^{n} \alpha_j u_i^j \lambda_j^{k+1} \quad \text{---(9'')}$

Dividing the second sum by the first, we get

$$\beta_k = \frac{x_i^{k+1}}{x_i^k} = \frac{\alpha_1 u_i^1 \lambda_1^{k+1} + \alpha_2 u_i^2 \lambda_2^{k+1} + \cdots + \alpha_n u_i^n \lambda_n^{k+1}}{\alpha_1 u_i^1 \lambda_1^k + \alpha_2 u_i^2 \lambda_2^k + \cdots + \alpha_n u_i^n \lambda_n^k} \quad \text{---(10)}$$

Suppose that $\alpha_1 \neq 0$ & $u_i^1 \neq 0$. This can be achieved by appropriate choice of the initial vector $\underset{\sim}{x}_0$.

Transform expression (10) as follows:

$$\beta_k = \frac{x_i^{k+1}}{x_i^k} = \lambda_1 \frac{1 + \frac{\alpha_2 u_i^2}{\alpha_1 u_i^1}\left(\frac{\lambda_2}{\lambda_1}\right)^{k+1} + \cdots + \frac{\alpha_n u_i^n}{\alpha_1 u_i^1}\left(\frac{\lambda_n}{\lambda_1}\right)^{k+1}}{1 + \frac{\alpha_2 u_i^2}{\alpha_1 u_i^1}\left(\frac{\lambda_2}{\lambda_1}\right)^{k} + \cdots + \frac{\alpha_n u_i^n}{\alpha_1 u_i^1}\left(\frac{\lambda_n}{\lambda_1}\right)^{k}} \quad \text{---(11)}$$

The theoretical result (11) reveals that the sequence $\{\beta_k\}$ also converges to λ_1.