The Power Method for Eigenvalues and Eigenvectors

The spectrum of a square matrix A, denoted by $\sigma(A)$ is the set of all eigenvalues of A. The spectral radius of A, denoted by $\rho(A)$ is defined as:

$$\rho(A) = \max\{|\lambda| : \ \lambda \in \sigma(A)\}$$

An eigenvalue of A that is larger in absolute value than any other eigenvalue is called the dominant eigenvalue; a corresponding eigenvector is called a dominant eigenvector.

The most commonly encountered vector norm (often simply called "the norm" of a vector) is the Euclidean norm, given by

$$\|x\|_2 = \|x\| = \sqrt{x_1^2 + x_2^2 + ... + x_n^2}.$$

The max norm, also known as the infinity norm of a vector is the largest component of the vector in absolute value:

$$\|x\|_{max} = ||x||_\infty = \max\{|x_1|\,,\ |x_2|\,,\ \ldots,\ |x_n|\}.$$

An eigenvector V is said to be normalized if its coordinates are divided by its norm. If the max norm is used, then clearly, the coordinate of largest magnitude is equal to one.

Theorem 1. *If u is an eigenvector of a matrix A, then its corresponding eigenvalue is given by*

$$\lambda = \frac{u^* \cdot Au}{u^* \cdot u},$$

where u^* is the conjugate transpose of u. This quotient is called the Rayleigh quotient.

Proof. Since u is an eigenvector of A, we know that $Au = \lambda u$ and we can write

$$\frac{u^* \cdot Au}{u^* \cdot u} = \frac{\lambda u^* \cdot u}{u^* \cdot u} = \lambda\ \frac{u^* \cdot u}{u^* \cdot u} = \lambda\ \times 1 = \lambda$$

■

Theorem 2. *Let $\lambda_1,\ \lambda_2,\ \ldots\ \lambda_m$ be the m eigenvalues (counted with multiplicity) of the $n \times n$ matrix A and let $v_1, v_2,\ \ldots,\ v_m$ be the corresponding eigenvectors. Suppose that λ_1 is the dominant eigenvalue, so that*

$$|\lambda_1| > |\lambda_2| \geq \cdots \lambda_j \geq \cdots \geq |\lambda_m|.$$

Suppose x_0 is in the subspace generated the eigenvectors, i.e.,

$$x_0 = c_1 v_1 + c_2 v_2 + \cdots + c_m v_m$$

with $c_1 \neq 0$. Then $x_k = \dfrac{A^k x_0}{\|A^k x_0\|}$ converges to (a multiple of) the dominant eigenvector v_1.

Proof. We have

$$\begin{aligned} A^k x_0 &= c_1 A^k v_1 + c_2 A^k v_2 + \cdots + c_m A^k v_m \\ &= c_1 \lambda_1^k v_1 + c_2 \lambda_2^k v_2 + \cdots + c_m \lambda_m^k v_m \\ &= c_1 \lambda_1^k \left[v_1 + \frac{c_2}{c_1} \left(\frac{\lambda_2}{\lambda_1} \right)^k v_2 + \cdots + \frac{c_m}{c_1} \left(\frac{\lambda_m}{\lambda_1} \right)^k v_m \right]. \end{aligned}$$

The expression within brackets converges to v_1 because $|\lambda_j/\lambda_1| < 1$ for $j > 1$. Thus $x_k = \dfrac{A^k x_0}{\|A^k x_0\|}$ converges to a dominant eigenvector. ■

Remark 1. The convergence is geometric, with ratio

$$\left| \frac{\lambda_2}{\lambda_1} \right|.$$

Thus, the method converges slowly if λ_2 is close in magnitude to the dominant eigenvalue λ_1.

Given an $n \times n$ matrix A, for $i = 1, 2, \ldots, n$, define:

$$R_i(A) = \sum_{i=1}^{n} |a_{ij}| \quad \text{and} \quad r_i(A) = R_i(A) - |a_{ii}|.$$

Levy-Deslanque theorem. *If the matrix A is strictly diagonally dominant, that is*

$$|a_{ii}| > r_i(A) \quad \textit{for all} \quad i = 1, 2, \ldots, n.$$

Then A is invertible.

Proof. Suppose $\det(A) = 0$, then for some nonzero vector $u = (u_1, u_2, \ldots, u_n)^t$, $A\,u = \theta$. Now let k be the index where

$$u_k \geq u_i \quad \text{for all} \quad i = 1, 2, \ldots, n\,.$$

Then

$$|a_{kk}|\ ||uk|| = \left| \sum_{i \neq j} a_{kj}\, u_j \right| \leq \sum_{i \neq j} |a_{kj}|\, u_j| \leq r_i(A) \cdot$$

which is a contradiction with $|akk| > r_k(A)$. ■

Gerschgorin's Disks theorem. *The eigenvalues of A lie in the union of the disks*

$$D_i(a_{ii}, r_i(A)), \quad \textit{centered at} \;\; a_{ii} \quad \textit{with the radius} \quad r_i(A).$$

Moreover, the union of any k of these disks that do not intersect the remaining $(n-k)$ contains precisely k (counting multiplicities) of the eigenvalues.

Proof. We only prove the first part.
Let k be an eigenvalue of A, then $\det(A - \lambda_k I_n) = 0$. By the Levy-Deslanque theorem, we conclude that $|\lambda_k - a_{ii}| < r_i(A)$ for at least one i. ■

The fact that A and A^t have the same set of eigenvalues, one can state a similar theorem dealing with columns.

Example 1. Consider the symmetric matrix

$$\begin{pmatrix} 2 & 1 & 2 \\ 1 & 3 & -1 \\ 2 & -1 & 9 \end{pmatrix}$$

The eigenvalues are real and according to the above theorem, there are two eigenvalues between -1 and 5 and the dominant eigenvalue is in $[6, 12]$.

The spectral norm of $m \times n$ matrix A, denoted by $||A||_2$, which is the square root of the maximum eigenvalue of the positive semi-definite matrix A^*A or AA^* (we choose the one in smaller size),

$$||A||_2 = \sqrt{\rho(A^*A)} = \sqrt{\rho(AA^*)} \tag{4}$$

is often referred to as "the" matrix norm. So in order to find the spectral norm of an $m \times n$ matrix A, we need to find the dominant eigenvalue of AA^* or A^*A.

♠ **Power Method.** This method is a simple iterative technique to obtain the dominant eigenvalue and the corresponding dominant eigenvector. It does not compute a matrix decomposition, and hence it can be used when A is a very large sparse matrix. However, it may converge only slowly.

The power iteration algorithm starts with a vector x_0, which may be an approximation to the dominant eigenvector or a random vector. The method is described by the iteration

$$x_{k+1} = \frac{Ax_k}{\|Ab_k\|_\infty}.$$

So, at every iteration, the vector bk is multiplied by the matrix A and normalized by the max norm.

The convergence is guaranteed under the following conditions:

a. A has an eigenvalue that is strictly greater in magnitude than its other eigenvalues;
b. The starting vector x_0 has a nonzero component in the direction of an eigenvector associated with the dominant eigenvalue (see Theorem 2).

Remark 2. It is known that any matrix with positive entries has a unique dominant eigenvalues. Most non-negative matrices with not many zero entries also have a unique dominant eigenvalue. Therefore these types of matrices are good candidates for the power method.

Algorithm. The algorithm for the power method is very simple.

Choose an initial guess x_0, a maximum number of iterations N, and a tolerance Tol;

set $k = 1$;

while $k \le N$;

set $y_k = A\,x_{k-1}$;

set $\alpha_k = \|y_k\|_\infty$;

set $x_k = y_k/\alpha_k$;

if $|\alpha_k - \alpha_{k-1}| \le Tol$, then $k = N + 1$; else, set $k = k + 1$;

end.

Display the dominant eigenvector $u_1 \approx x_k$ and the dominant eigenvalue $\lambda_1 \approx \dfrac{x_k^* \cdot A\,x_k}{x_k^* \cdot x_k}$. ■

Example 2. Calculate seven iterations of the power method to approximate a dominant eigenvector of the matrix

$$A = \begin{pmatrix} 1 & 2 & 0 \\ -2 & 1 & 2 \\ 1 & 3 & 1 \end{pmatrix} \quad \text{with} \quad x_0 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}.$$

Solution Let

$$y_1 = Ax_0 = \begin{pmatrix} 1 & 2 & 0 \\ -2 & 1 & 2 \\ 1 & 3 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 3 \\ 1 \\ 5 \end{pmatrix}.$$

Then $\alpha_1 = \|y_1\|_\infty = 5$; we obtain

$$x_1 = \frac{y_1}{\|y_1\|_\infty} = \frac{1}{5}\begin{pmatrix}3\\1\\5\end{pmatrix} = \begin{pmatrix}0.60\\0.20\\1.00\end{pmatrix}.$$

A second iteration yields

$$y_2 = (1.00,\ 1.00\ 2.20)^t \ \text{ with } \ \alpha_2 = \|y_2\|_\infty = 2.20 \ \text{ and } \ x_2 = (0.45,\ 0.45,\ 1.00)^t.$$

Continuing this process, we obtain the sequence of approximations shown below:

x_0	x_1	x_2	x_3	x_4	x_5	x_6	x_7
1.00	0.60	0.45	0.48	0.51	0.50	0.50	0.50
1.00	0.20	0.45	0.55	0.51	0.49	0.50	0.50
1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	α_1	α_2	α_3	α_4	α_5	α_6	α_7
	5.00	2.20	2.82	3.13	3.02	2.99	3.00

Using the Rayleigh quotient (Theorem 1), we approximate the dominant eigenvalue of A to be 3.00.

♣ **When can the power method fail?** In principle, the approximations x_k approach the the dominant eigenvector eigenvector v_1 while the scaling factors α_k tend to the corresponding eigenvalue λ_1 as the number of iterations k increases. This can fail if there is not a single largest eigenvalue: if $|\lambda_2| = |\lambda_1|$, then in general the power method does not converge at all (see Theorem 2). Since we do not know the eigenvalues of A in advance, we cannot know in advance whether the power method will work or not.

Recall that the eigenvalues of a real matrix A are in general complex, and occur in conjugate pairs (which necessarily have the same absolute value). So if the largest eigenvalue of A is not real (and there is no reason why it should be) then the power method will certainly fail. For this reason it is sensible to apply the power method only to matrices whose eigenvalues are known to be real, for example symmetric matrices. For such matrices we could consider ourselves unlucky if the two largest eigenvalues happened to have the same absolute value, i.e. if $|\lambda_1| = |\lambda_2|$.

As we explained in the proof of Theorem 2, when the power method does converge, the rate at which it does so is determined by the ratio $\frac{|\lambda_1|}{|\lambda_2|}$. Thus if $|\lambda_2|$ is only slightly smaller

than $|\lambda_1|$, then the method converges very slowly, so that a large number of iterations will be required to obtain a decent accuracy.

Another point worth making is that it is not strictly true that the power method necessarily converges to the largest eigenvalue λ_1. This would fail in the exceptional case $c_1 = 0$ (see Theorem 2), that is if the initial guess x_0 contains no component of the dominant eigenvector v_1. In this case, the method will converge to the largest eigenvalue whose eigenvector is contained in the decomposition of x_0. If there are two such eigenvalues with the same absolute value, then the method will not converge at all. Of course it is extremely unlikely in general for c_1 to be zero. In practice we can help to avoid this unfortunate possibility by filling x_0 with random entries rather than something obvious like $x_0 = (1,\ 1,\ \ldots, 1)^t$. Suppose the dominant eigenvector is $v_1 = (1, 0, 0, \ldots, 0, 0)^t$ and we choose our initial vector $x_0 = (0,\ 1, 0, 0 \ldots, 0, 0)^t$, then clearly we shall not be able to obtain the dominant eigenvector.

♠ **Inverse Power Method** This method operates with A^{-1} rather than A. Thus we replace the step $y_k = Ax_{k-1}$ with $y_k = A^{-1}x_{k-1}$ (if A^{-1} is known) or solve the system $Ay_k = x_{k-1}$. Since the eigenvalues of A^{-1} are $1/\lambda_1,\ 1/\lambda_2,\ \cdots 1/\lambda_m$, the inverse power method should converge to $1/\lambda_m$, the largest eigenvalue of A^{-1}. Thus this method gives a way of finding the smallest (in absolute value) eigenvalue of a matrix.

♠ **Spectral Shift** This method uses the fact that the eigenvalues of $A - \alpha I_n$ are

$$\lambda_1 - \alpha\,,\ \lambda_2 - \alpha\,,\ \ldots\,,\ \lambda_m - \alpha\,.$$

Thus having computed the eigenvalue λ_1, we start the method over again using the shifted matrix $A - \lambda_1 I_n$. This reduces the eigenvalue we have already determined to zero, and the method now converges to the largest in absolute value of

$$\lambda_2 - \lambda_1\,,\ \lambda_3 - \lambda_1\,,\ \cdots, \lambda_m - \lambda_1\,.$$

A spectral shift always enables us to find at least one more eigenvalue than the dominant one. However it is not clear how it could be implemented in general to find all the eigenvalues of an $n \times n$ matrix A. This method works only for symmetric matrices, but

is somewhat more subtle since it shifts only one eigenvalue at a time to zero, leaving the others unchanged.

♠ **Symmetric Power Method** This method is often used to approximate the dominant eigenvalue and eigenvector of a symmetric matrix. The vector norm used in this method is the Euclidean norm (L^2-norm) instead of the max norm.

Algorithm. Choose a maximum number of iterations N and a tolerance Tol; then start with an initial guess x_0;

set $k = 1$;

while $k \le N$;

set $y_k = A\,x_{k-1}$;

set $\alpha_k = \|y_k\|_2$;

set $x_k = y_k/\alpha_k$;

if $|\alpha_k - \alpha_{k-1}| \le Tol$, then $k = N + 1$; else, set $k = k + 1$;

end.

Display the dominant eigenvector $u_1 \approx x_k$ and the dominant eigenvalue $\lambda_1 \approx \dfrac{x_k^* \cdot A\,x_k}{x_k^* \cdot x_k}$. ■

Example 3. With a tolerance of $Tol = 0.001$, use the symmetric power method to approximate a dominant eigenvector of the matrix

$$A = \begin{pmatrix} 4 & 1 & 1 \\ 1 & 3 & 2 \\ 1 & 2 & 3 \end{pmatrix} \quad \text{with} \quad x_0 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}.$$

Solution. Let

$$y_1 = Ax_0 = \begin{pmatrix} 4 & 1 & 1 \\ 1 & 3 & 2 \\ 1 & 2 & 3 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 4 \\ 1 \\ 1 \end{pmatrix}.$$

Then $\mu_1 = x_0^t y_1 = 4$ and $\alpha_1 = y_1\|_2 = \sqrt{18}$; we obtain

$$x_1 = \frac{y_1}{\|y_1\|_2} = \frac{1}{\sqrt{18}} \begin{pmatrix} 4 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0.942809042 \\ 0.235702260 \\ 0.235702260 \end{pmatrix}.$$

A second iteration yields

$$\mu = 5.000000000 \quad \text{and} \quad x_2 = \begin{pmatrix} 0.816496581 \\ 0.408248290 \\ 0.408248290 \end{pmatrix}.$$

Continuing this process, we obtain the sequence of approximations shown below:

Iteration	eigenvalue	eigenvector		
1	4.000000000	0.942809042	0.235702260	0.235702260
2	5.000000000	0.816496581	0.408248290	0.408248290
3	5.666666667	0.710669055	0.497468338	0.497468338
4	5.909090909	0.816496581	0.408248290	0.408248290
5	5.976744186	0.612836480	0.558762673	0.558762673
6	5.994152047	0.595247159	0.568190470	0.568190470
7	5.998535871	0.586335581	0.572804760	0.572804760
8	5.999633834	0.581851940	0.575086220	0.575086220
9	5.999908450	0.579603333	0.576220434	0.576220434
10	5.999977112	0.578477355	0.576785901	0.576785901
11	5.999994278	0.577913950	0.577068222	0.577068222

We conclude that the dominant eigenvalue and eigenvector are $\lambda_1 = 6$ and $v_1 = (1, 1, 1)^t$, respectively. To obtain the same accuracy with the power method we need 19 iterations.