The Power Method

In this lesson we will present the power method for finding the first eigenvector and eigenvalue of a matrix. Then we will prove the convergence of the method for diagonalizable matrices (if $|\lambda_1| > |\lambda_2|$ where λ_i is the i^{th} largest eigenvalue) and discuss the rate of convergence.

Algorithm 1 The Power Method

Choose a random vector $q^{(0)} \in R^n$
for $k = 1, 2, \ldots$ (while $\|q^{(k-1)} - q^{(k-2)}\| > \epsilon$)
  $z^{(k)} = Aq^{(k-1)}$
  $q^{(k)} = z^{(k)}/\|z^{(k)}\|$
  $\lambda^{(k)} = [q^{(k)}]^T A q^{(k)}$
end

Let us examine the convergence properties of the power iteration. If A is diagonalizable (see appendix for a reminder) then there exist n independent eigenvectors of A. Let $x_1, \ldots x_n$ be these eigenvectors, then $x_1, \ldots x_n$ form a basis of R^n. Hence the initial vector $q^{(0)}$ can be written as:

$$q^{(0)} = a_1 x_1 + a_2 x_2 + \cdots + a_n x_n \quad (1)$$

where $a_1, \ldots, a_n$ are scalars. multiplying both sides of the equation in A^k yields:

$$\begin{aligned} A^k q^{(0)} &= A^k(a_1 x_1 + a_2 x_2 + \cdots + a_n x_n) = a_1 A^k x_1 + a_2 A^k x_2 + \ldots + a_n A^k x_n \quad (2) \\ &= a_1 \lambda_1^k x_1 + a_2 \lambda_2^k x_2 + \ldots + a_n \lambda_n^k x_n = a_1 \lambda_1^k \left(x_1 + \sum_{j=2}^{n} \frac{a_j}{a_1} \left(\frac{\lambda_j}{\lambda_1} \right)^k x_j \right) \end{aligned}$$

If $|\lambda_1| > |\lambda_2| \geq \ldots |\lambda_n|$ then we say that λ_1 is a dominant eigenvalue. In this case $\left(\frac{\lambda_j}{\lambda_1}\right)^k \to 0$ and therefore if $a_1 \neq 0$, $A^k q^{(0)} \to a_1 \lambda_1^k x_1$. The power method normalizes the products $Aq^{(k-1)}$ to avoid overflow/underflow, therefore it converges to x_1 (assuming it has unit norm).

The power method converges if λ_1 is dominant and if $q^{(0)}$ has a component in the direction of the corresponding eigenvector x_1. In practice, the usefulness of the power method depends upon the ration $|\lambda_2|/|\lambda_1|$, since it dictates the rate of convergence. The danger that $q^{(0)}$ is deficient in x_1 ($a_1 = 0$) is a less worrisome matter because if $q^{(0)}$ is chosen randomly the probability for this is 0. Moreover, rounding errors sustained during the iteration typically ensure that the subsequent $q^{(k)}$ have a component in this direction.

If the power method has converged to the dominant eigenvector after k iterations then $[q^{(k)}]^T A q^{(k)} \approx [q^{(k)}]^T \lambda q^{(k)} = \lambda [q^{(k)}]^T q^{(k)} = \lambda \|q^{(k)}\|^2 = \lambda$ ($\|q^{(k)}\|^2 = 1$ because $q^{(k)}$ is normalized in each iteration).

Notice that in each iteration we compute a single matrix-vector multiplication ($O(n^2)$). We never perform matrix-matrix multiplication which requires greater number of operations ($O(n^3)$). If the matrix A is sparse (only a small portion of the entries of A are non-zero), matrix-vector multiplication can be performed very efficiently. Therefore the power method is practical even if n is very large, such as in Google's Page Rank algorithm.

An example for the case that $|\lambda_1| = |\lambda_2|$ and the method does not converge is rotation matrices. Consider a 2×2 rotation matrix U. (reminder: a 2×2 rotation matrix is of the form $\begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}$). U is orthonormal, that is $U^T U = U U^T = I$. let λ be an eigenvalue of U and let x be the corresponding eigenvector.

$$|\lambda|^2 \|x\|^2 = \|\lambda x\|^2 = \|Ux\|^2 = \|x^T U^T U x\|^2 = \|x^T x\|^2 = \|x\|^2 \tag{3}$$

therefore $|\lambda| = 1$. If $U \neq I$ then $x \neq Ux$ for $x \neq 0$ and the power method does not converge.

Appendix

A matrix $A \in R^{n \times n}$ is diagonalizable if there exists an invertible matrix X such that $A = XDX^{-1}$ where D is a diagonal matrix.
claim: A is diagonalizable iff it has n linearly independent eigenvector.
proof: Suppose that A has n linearly independent eigenvectors. Denote these eigenvectors by $x_1 \dots x_n$. Then $x_1 \dots x_n$ are linearly independent iff the rank of the matrix $X = \begin{bmatrix} | & & | \\ x_1 & \cdots & x_n \\ | & & | \end{bmatrix}$ is n iff X is invertible. x_i is an eigenvector

of A, hence $Ax_i = x_i\lambda_i$. Taking the collection of these equations for the n eigenvectors in matrix notation we get:

$$A \begin{bmatrix} | & & | \\ x_1 & \cdots & x_n \\ | & & | \end{bmatrix} = \begin{bmatrix} | & & | \\ x_1\lambda_1 & \cdots & x_n\lambda_n \\ | & & | \end{bmatrix} = \begin{bmatrix} | & & | \\ x_1 & \cdots & x_n \\ | & & | \end{bmatrix} \begin{bmatrix} \lambda_1 & \dots & 0 \\ \vdots & & \\ 0 & \dots & \lambda_n \end{bmatrix}$$

Let $D = \begin{bmatrix} \lambda_1 & \dots & 0 \\ \vdots & & \\ 0 & \dots & \lambda_n \end{bmatrix}$, then the last equation is AX = XD or $A = XDX^{-1}$ and hence A is diagonalizable. The columns of the matrix X are the eigenvectors of A and the entries on the diagonal of D are the corresponding eigenvalues.