The Power Method for Spectral Radius

Given an $n \times n$ square matrix A, assume that it has eigenvalues

$$|\lambda_1| > |\lambda_2| \geq |\lambda_3| \geq \cdots \geq |\lambda_n| \geq 0$$

with associated eigenvectors $\boldsymbol{v}_1, \ldots \boldsymbol{v}_n$ that are **linearly independent** and **normalized** ($\|\boldsymbol{v}_i\|_\infty = 1$ – to normalize a vector in the l^∞ sense, we divide each component by the maximal entry in absolute value, namely, $\frac{\boldsymbol{v}}{\|\boldsymbol{v}\|_\infty}$). As the $\boldsymbol{v}_i$'s are linearly independent, they span $\mathbb{R}^n$ (linear algebra knowledge), i.e., for any $\boldsymbol{x}^{(0)} \in \mathbb{R}^n$, we can find coefficients β_i's such that

$$\boldsymbol{x}^{(0)} = \sum_{i=1}^{n} \beta_i \boldsymbol{v}_i.$$

Now, let's multiply both sides by A on the left.

$$A\boldsymbol{x}^{(0)} = A\sum_{i=1}^{n} \beta_i \boldsymbol{v}_i = \sum_{i=1}^{n} \beta_i A\boldsymbol{v}_i = \sum_{i=1}^{n} \beta_i \lambda_i \boldsymbol{v}_i.$$

Let's apply A again on both sides and obtain

$$A^2\boldsymbol{x}^{(0)} = \sum_{i=1}^{n} \beta_i \lambda_i A\boldsymbol{v}_i = \sum_{i=1}^{n} \beta_i \lambda_i^2 \boldsymbol{v}_i.$$

After multiplying A k times on the left, we have

$$A^k\boldsymbol{x}^{(0)} = \sum_{i=1}^{n} \beta_i \lambda_i^k \boldsymbol{v}_i.$$

Knowing that λ_1 has the largest magnitude, let's factor it out as in

$$A^k\boldsymbol{x}^{(0)} = \lambda_1^k \sum_{i=1}^{n} \beta_i \left(\frac{\lambda_i}{\lambda_1}\right)^k \boldsymbol{v}_i = \lambda_1^k \beta_1 \boldsymbol{v}_1 + \lambda_1^k \left(\beta_2 \left(\frac{\lambda_2}{\lambda_1}\right)^k \boldsymbol{v}_2 + \cdots + \beta_n \left(\frac{\lambda_n}{\lambda_1}\right)^k \boldsymbol{v}_n\right).$$

As k grows, the contribution from $\left(\frac{\lambda_i}{\lambda_1}\right)^k$ diminishes for $i = 2, 3, \ldots, n$, and the RHS becomes dominated by the leading term ($i = 1$), that is, by defining $\boldsymbol{x}^{(k)} = A^k\boldsymbol{x}^{(0)}$, we have

$$\lim_{k\to\infty} \boldsymbol{x}^{(k)} = \lim_{k\to\infty} A^k\boldsymbol{x}^{(0)} = \lim_{k\to\infty} \lambda_1^k \beta_1 \boldsymbol{v}_1.$$

This is saying that when k is large enough, $A^k\boldsymbol{x}^{(0)}$ is almost parallel to $\boldsymbol{v}_1$.

We observe that the limit on the RHS only converges if $|\lambda_1| < 1$ and $\beta_1 \neq 0$. Meanwhile, the limit on the LHS may incur overflow or underflow since the entries of A can easily blow up from repeated matrix multiplications. To avoid this, we rescale the product $A^k\boldsymbol{x}^{(0)}$ by its infinity norm $\left\|A^{k-1}\boldsymbol{x}^{(0)}\right\|_\infty$, so that we always iterating the multiplication to a unit vector in l^∞.

With the rescaling, define $\boldsymbol{z}^{(k)} = \frac{A^k\boldsymbol{x}^{(0)}}{\left\|A^{k-1}\boldsymbol{x}^{(0)}\right\|_\infty}$, we see that

$$\lim_{k\to\infty} \boldsymbol{z}^{(k)} = \lim_{k\to\infty} \frac{A^k\boldsymbol{x}^{(0)}}{\left\|A^{k-1}\boldsymbol{x}^{(0)}\right\|_\infty} = \lim_{k\to\infty} \frac{\lambda_1^k \beta_1 \boldsymbol{v}_1}{\left\|\lambda_1^{k-1} \beta_1 \boldsymbol{v}_1\right\|_\infty} = \lambda_1 \boldsymbol{v}_1,$$

namely, the rescaled iteration leads to the principal eigenvector corresponding to the spectral radius of A.

Furthermore, define $\boldsymbol{y}^{(k)} = \frac{A^k\boldsymbol{x}^{(0)}}{\left\|A^k\boldsymbol{x}^{(0)}\right\|_\infty}$, we find that

$$\lim_{k\to\infty} \boldsymbol{y}^{(k)} = \lim_{k\to\infty} \frac{A^k\boldsymbol{x}^{(0)}}{\left\|A^k\boldsymbol{x}^{(0)}\right\|_\infty} = \lim_{k\to\infty} \frac{\lambda_1^k \beta_1 \boldsymbol{v}_1}{\left\|\lambda_1^k \beta_1 \boldsymbol{v}_1\right\|_\infty} = \boldsymbol{v}_1$$

which shows that the sequence of $\boldsymbol{z}^{(k)}$ (as a result of repeated multiplication of A) converges to the principle eigenvector of A, corresponding to the eigenvalue with maximal magnitude. Know the limits of both $\boldsymbol{x}^{(k)}$ and $\boldsymbol{z}^{(k)}$ determines λ_1 since for k large enough, we must have

$$\lambda_1 \boldsymbol{z}^{(k)} \approx \boldsymbol{x}^{(k)} \implies |\lambda_1| \approx \frac{\left\|\boldsymbol{x}^{(k)}\right\|}{\left\|\boldsymbol{z}^{(k)}\right\|} = \frac{\left\|A^k\boldsymbol{x}^{(0)}\right\|_\infty}{\left\|A^{k-1}\boldsymbol{x}^{(0)}\right\|_\infty} = \frac{\left\|\boldsymbol{x}^{(k)}\right\|_\infty}{\left\|\boldsymbol{x}^{(k-1)}\right\|_\infty},$$

that is, the leading eigenvalue is equal to ratio of the infinity norm of the successive iterates. The method of determining λ_1 is called the **Power Method**.

Example. Let $A = \begin{bmatrix} -2 & -3 \\ 6 & 7 \end{bmatrix}$. We know its eigenvalues $\lambda_1 = 4$ and λ_2, with corresponding eigenvectors $\boldsymbol{v}_1 = (1, -2)^{\mathrm{T}}$ and $\boldsymbol{v}_2 = (1, -1)^{\mathrm{T}}$. Suppose we start with $\boldsymbol{x}^{(0)} = (1, 1)^{\mathrm{T}}$.

$$\begin{aligned}
\boldsymbol{x}_1 = A\boldsymbol{x}_0 &= \begin{bmatrix} -5 \\ 13 \end{bmatrix}, \\
\boldsymbol{x}_2 = A\boldsymbol{x}_1 &= \begin{bmatrix} -29 \\ 61 \end{bmatrix}, \\
\boldsymbol{x}_3 = A\boldsymbol{x}_2 &= \begin{bmatrix} -125 \\ 253 \end{bmatrix}, \\
\boldsymbol{x}_4 = A\boldsymbol{x}_3 &= \begin{bmatrix} -509 \\ 1021 \end{bmatrix}, \\
\boldsymbol{x}_5 = A\boldsymbol{x}_4 &= \begin{bmatrix} -2045 \\ 4093 \end{bmatrix}, \\
\boldsymbol{x}_6 = A\boldsymbol{x}_5 &= \begin{bmatrix} -8189 \\ 16381 \end{bmatrix}.
\end{aligned}$$

By the derivation above, we estimate the eigenvalue with maximal magnitude

$$\begin{aligned}
\lambda_1^{(1)} &= \frac{\|\boldsymbol{x}_2\|_\infty}{\|\boldsymbol{x}_1\|_\infty} = \frac{61}{13} = 4.6923 \\
\lambda_1^{(2)} &= \frac{\|\boldsymbol{x}_3\|_\infty}{\|\boldsymbol{x}_2\|_\infty} = \frac{253}{61} = 4.1475 \\
\lambda_1^{(3)} &= \frac{\|\boldsymbol{x}_4\|_\infty}{\|\boldsymbol{x}_3\|_\infty} = \frac{1021}{253} = 4.03557 \\
\lambda_1^{(4)} &= \frac{\|\boldsymbol{x}_5\|_\infty}{\|\boldsymbol{x}_4\|_\infty} = \frac{4093}{1021} = 4.00881 \\
\lambda_1^{(5)} &= \frac{\|\boldsymbol{x}_6\|_\infty}{\|\boldsymbol{x}_5\|_\infty} = \frac{16381}{4093} = 4.00200.
\end{aligned}$$

and $\boldsymbol{x}_6 = \begin{bmatrix} -8189 \\ 16381 \end{bmatrix}$ which normalizes to $\begin{bmatrix} -0.49908 \\ 1 \end{bmatrix} \approx \boldsymbol{v}_1$.

This examples shows that rescaling is absolutely necessary when we expect the method to converge slowly because the values in the iterates can easily blow up. The derivation made above is useful to give us a sense of how things may converge, but is not useful in practice. Below, we conclude the power method in an algorithm with careful considerations of rescaling.

Algorithm

In practice, we compute $\boldsymbol{x}^{(k)}$ iteratively. Given an initial guess $\boldsymbol{x}$,

(1) Find the first index p such that $|x_p| = \|\boldsymbol{x}\|_\infty$.
   - (a) Rescale $\boldsymbol{x} = \boldsymbol{x}/x_p$.
   - (b) Begins iteration $\boldsymbol{y} = A\boldsymbol{x}$.
   - (c) Set $\mu = y_p$. Find the first index p such that $|y_p| = \|\boldsymbol{y}\|_\infty$.
     - (i) If $y_p = 0$, then output eigenvector $\boldsymbol{x}$ and an eigenvalue of 0. We should choose a different initial guess.
     - (ii) Otherwise, set
$$\mathit{error} = \|\boldsymbol{x} - \boldsymbol{y}/y_p\|_\infty$$
       and $\boldsymbol{x} = \boldsymbol{y}/y_p$ and return to step 3 until $\mathit{error} < \mathit{tol}$.
   - (d) Output $(\mu, \boldsymbol{x})$ as the eigenvalue-eigenvector pair.

Convergence

Using the final result of the derivation, recall that $\boldsymbol{y}^{(k)} = \frac{A^k\boldsymbol{x}^{(0)}}{\left\|A^k\boldsymbol{x}^{(0)}\right\|_\infty}$ and we know it converges to the principal eigenvector $\boldsymbol{v}_1$. Indeed,

$$
\begin{aligned}
\left\|\boldsymbol{y}^{(k)} - \boldsymbol{v}_1\right\| &= \left\|\frac{A^k\boldsymbol{x}^{(0)}}{\left\|A^k\boldsymbol{x}^{(0)}\right\|_\infty} - \boldsymbol{v}_1\right\| \\
&= \left\|\frac{\sum_{i=1}^n \beta_i\lambda_i^k\boldsymbol{v}_i}{\sum_{i=1}^n \beta_i\lambda_i^k} - \boldsymbol{v}_1\right\| \\
&= \left\|\frac{\lambda_1^k\sum_{i=1}^n \beta_i\left(\frac{\lambda_i}{\lambda_1}\right)^k\boldsymbol{v}_i}{\lambda_1^k\sum_{i=1}^n \beta_i\left(\frac{\lambda_i}{\lambda_1}\right)^k} - \boldsymbol{v}_1\right\| \\
&= \left\|\frac{\sum_{i=1}^n \beta_i\left(\frac{\lambda_i}{\lambda_1}\right)^k\boldsymbol{v}_i}{\sum_{i=1}^n \beta_i\left(\frac{\lambda_i}{\lambda_1}\right)^k} - \boldsymbol{v}_1\right\| \\
&= \left\|\frac{\beta_1\boldsymbol{v}_1 + \sum_{i=2}^n \beta_i\left(\frac{\lambda_i}{\lambda_1}\right)^k\boldsymbol{v}_i}{\beta_1 + \sum_{i=2}^n \beta_i\left(\frac{\lambda_i}{\lambda_1}\right)^k} - \boldsymbol{v}_1\right\| \\
&= \left\|\frac{\sum_{i=2}^n \beta_i\left(\frac{\lambda_i}{\lambda_1}\right)^k\boldsymbol{v}_i - \sum_{i=2}^n \beta_i\left(\frac{\lambda_i}{\lambda_1}\right)^k\boldsymbol{v}_1}{\beta_1 + \sum_{i=2}^n \beta_i\left(\frac{\lambda_i}{\lambda_1}\right)^k}\right\| \\
&\le \left\|\frac{\sum_{i=2}^n \beta_i\left|\frac{\lambda_i}{\lambda_1}\right|^k(\boldsymbol{v}_i - \boldsymbol{v}_1)}{\beta_1}\right\| \\
&\le \left\|\left|\frac{\lambda_2}{\lambda_1}\right|^k\frac{\sum_{i=2}^n \beta_i(\boldsymbol{v}_i - \boldsymbol{v}_1)}{\beta_1}\right\|, \quad \text{since} \quad \lambda_i \le \lambda_2, \quad i = 2,\dots,n. \\
&\le \left|\frac{\lambda_2}{\lambda_1}\right|^k\left\|\frac{\sum_{i=2}^n \beta_i}{\beta_1}\right\|(\|\boldsymbol{v}_i\| + \|\boldsymbol{v}_1\|) \\
&= C\left|\frac{\lambda_2}{\lambda_1}\right|^k
\end{aligned}
$$

where we call $C = 2\left|\frac{\sum_{i=2}^n \beta_i}{\beta_1}\right|$. Thus, the rate of convergence depends on the ratio of $\frac{\lambda_2}{\lambda_1}$ (in fact the rate is $O\left(\left|\frac{\lambda_2}{\lambda_1}\right|^k\right)$). If λ_2 is smaller but very close to λ_1 in magnitude, then the convergence is expected to be slow.