The Power Method

- The power method is the basis of the most widely used method for computing eigenvalues and eigenvectors, i.e., the QR algorithm.

- Best description of the QR algorithm is in Golub/van Loan


- Suppose $A$ is a square matrix and we want to find its eigenvalue with the largest absolute value, i.e., its dominant eigenvalue.

- The basic idea of the power method is to start with a random vector, and to keep multiplying it with $A$. Each multiplication amplifies the component corresponding to the dominant eigenvalue, until eventually it dominates all others.

- To begin with, suppose $A$ has a dominant eigenvalue $\lambda_1$ where

$$|\lambda_1| > |\lambda_2| \geq |\lambda_3| \geq \ldots \geq |\lambda_n|$$

and

$$Ax_i = \lambda_ix_i, \quad i = 1, \ldots, n.$$
Note that the $x_i$ are vectors. $\lambda_1$ is the dominant eigenvalue.

- Here is version 1 of the power method:

1. Pick a random vector

$$q^{(0)} = \sum_{j=1}^{n} \alpha_j x_j \in \mathbb{R}^n$$

(of course we do not know the coefficients $\alpha_j$ but we assume $A$ has a complete set of eigenvectors so we can be sure they exist.)

2. For $k = 0, 1, 2, \ldots$ let

$$q^{(k+1)} = Aq^{(k)}.$$

- Every time we multiply with $A$ we amplify the dominant component:

$$q^{(1)} = Aq^{(0)} = \sum_{j=1}^{n} \alpha_j Ax_j = \sum_{j=1}^{n} \alpha_j \lambda_j x_j.$$

and in general

$$q^{(k)} = \sum_{j=1}^{n} \alpha_j \lambda_j^k x_j.$$

- Eventually the $\lambda_1$ term will dominate the others so that $q^{(k)}$ is a good approximation of the corresponding eigenvector.

- This won’t work! Why not?
We need to normalize!

Here is version 2 of the power method:

1. Pick a random vector

\[ q^{(0)} = \sum_{j=1}^{n} \alpha_j x_j \in \mathbb{R}^n \]

2. For \( k = 0, 1, 2, \ldots \) let

\[ z^{(k+1)} = Aq^{(k)} \]

\[ q^{(k+1)} = \frac{z^{(k+1)}}{\|z^{(k+1)}\|} \]

(for some suitable norm, often \( \| \cdot \|_\infty \))
Estimating the dominant eigenvalue

• Given an approximation \( q \), say, of an eigenvector, how do we approximate the corresponding eigenvalue?

• Idea: Find \( \lambda \) such that

\[
F(\lambda) = \| Aq - \lambda q \|_2^2 = \min
\]

• This is a simple calculus problem. We differentiate \( F \), set the derivative to zero, and solve for \( \lambda \):

\[
F(\lambda) = \| Aq - \lambda q \|_2^2 \\
= (Aq - \lambda q)^T (Aq - \lambda q) \\
= q^T A^T Aq - \lambda q^T (A + A^T) q + \lambda^2 q^T q
\]

and hence

\[
F'(\lambda) = 2\lambda q^T q - q^T (A + A^T) q = 0
\]

which gives

\[
\lambda = \frac{q^T (A + A^T) q}{2q^T q}
\]

• In the special case that \( A \) is symmetric this estimate turns into the Rayleigh Quotient

\[
\lambda = \frac{q^T A q}{q^T q}.
\]
What can go wrong?

- There might be no dominant eigenvalue, i.e.,

\[ |\lambda_1| = |\lambda_2| = \ldots = |\lambda_k| > |\lambda_{k+1}| \quad k > 1. \]

This has several subcases, including:
- \( \lambda_1 \) is an eigenvalue of algebraic multiplicity \( k \). In that case the iteration converges to a particular vector in the space spanned by the eigenvectors corresponding to \( \lambda_1 \).
- \( \lambda_1 = -\lambda_2, \quad k = 2 \). In that case the iteration becomes an oscillation of length 2, and one can work out the values of \( \lambda_1 \) and \( \lambda_2 \).
- \( \lambda_1 = \bar{\lambda}_2, \quad k = 2 \). The two dominant eigenvalues form a conjugate complex pair and the iteration becomes periodic. Again, one could work out the eigenvalues.
- Exercise: Think of other possibilities, e.g., defectiveness, \( k > 2 \).

- One might start with a random vector that has a zero component in the dominant eigenvalue\(^{-1-}\), i.e., \( \alpha_1 = 0 \). In that case, technically, the method converges to \( \lambda_2 \) (provided \( \alpha_2 \neq 0 \) and \( |\lambda_2| > |\lambda_3| \)). It does in exact

\(^{-1-}\) This happened to me the first time I assigned a homework problem where I had the class compute the eigenvalues of a 3 \( \times \) 3 matrix and then run the power method to get the largest of those...
arithmetic. However, in floating point arithmetic, round-off errors make the $\lambda_1$ component non-zero, and so eventually you do get the dominant eigenvector. This is the only case I know where round-off errors actually get you out of trouble.

- However, the main problem with the power method is that it converges only slowly if $|\lambda_1/\lambda_2|$ is close to 1, i.e., the dominance is weak.

**Shift of Origin**

- Shift of origin means that we apply the power method to a matrix $B$ of the form

$$B = A - \mu I$$

for some scalar $\mu$.

- The power method will converge to the dominant eigenvalue $\sigma$ of $B$. The eigenvalues of $B = A - \mu I$ are of course $\lambda_i - \mu$, and one can thus obtain the eigenvalue $\lambda = \mu + \sigma$ of $A$.

**Inverse Iteration**

- The basic idea is to apply the power method to $A^{-1}$. Of course, we don’t actually invert $A$. Instead we solve a linear system, using a suitable factorization such as the PLU or QR factorization of $A$.

- Here is **version 3** of the power method:
1. Pick a random vector

\[ q^{(0)} = \sum_{j=1}^{n} \alpha_j x_j \in \mathbb{R}^n \]

2. For \( k = 0, 1, 2, \ldots \):

Solve \( Az^{(k+1)} = q^{(k)} \)

set \( q^{(k+1)} = \frac{z^{(k+1)}}{\|z^{(k+1)}\|} \)

- Assuming that

\[ |\lambda_n| < |\lambda_{n-1}| \leq \cdots \leq |\lambda_1| \]

this will converge to \( 1/\lambda_n \) from which we can compute the smallest eigenvalue.

- Note that here we have a typical case where we solve many linear systems with the same coefficient matrix, and where we know the new right hand side only after we solve the previous system.

**Inverse Iteration and Shift of Origin**

- Inverse Iteration and Shift of Origin can be combined. We apply the power method to the matrix

\[ B = (A - \mu I)^{-1}. \]
• The eigenvalues of $B$ are

$$\eta = \frac{1}{\lambda - \mu} \iff \lambda = \mu + \frac{1}{\eta}$$

• Thus we can find the eigenvalue that is closest to our shift $\mu$.

• Again, we do not actually invert $A - \mu I$.

• Here is version 4 of the power method:

1. Pick a random vector

$$q^{(0)} = \sum_{j=1}^{n} \alpha_j x_j \in \mathbb{R}^n$$

2. For $k = 0, 1, 2, \ldots$:

Solve $(A - \mu I)z^{(k+1)} = q^{(k)}$

set $q^{(k+1)} = \frac{z^{(k+1)}}{\|z^{(k+1)}\|}$

• Here is an interesting complication. We want $A - \mu I$ to be well conditioned so that we can solve the linear system accurately. On the other hand, we want $\mu$ close to $\lambda$, for fast convergence. If $\mu$ actually was an eigenvalue then $A - \mu I$ would be singular and we could not solve the linear system. So the closer $\mu$ is to an eigenvalue, the more ill-conditioned is the linear system.

**Stopping**

• Suppose some version of the power method gives a unit vector \( \hat{x} \) which approximates an eigenvector and an approximation \( \hat{\lambda} \) of the corresponding eigenvalue. A reasonable criterion is to stop when

\[
\| A\hat{x} - \hat{\lambda}\hat{x} \| < \epsilon
\]

for a suitable tolerance \( \epsilon \) which depends on the problem.

**Squaring** \( A \)

• Carrying out \( n \) steps of the power method requires \( n^3 \) operations. Usually the number of iteration will be less than \( n \). But here is an interesting speculation. Suppose we contemplate iterating many more than \( n \) steps. The eigenvalues of \( A^2 \) are the squares of those of \( A \). Squaring takes \( n^3 \) operations, the same as \( n \) steps of the power method. However, squaring \( A^2 \) again also only takes \( n^3 \) operations,
but generates a matrix whose eigenvalues are the fourth power of those of \( A \). Squaring \( k \) times requires \( kn^3 \) operations, but generates the matrix \( A^{2^k} \) whose eigenvectors are the \( 2^k \)-th powers of those of \( A \). Multiplying with \( A^{2^k} \) is equivalent to \( 2^k \) steps of the power method. Carrying out that many steps with the ordinary power method would require \( 2^k n^3 \) steps as opposed to \( kn^3 \) for the squaring method. So it appears that repeated matrix squaring may be a good way to get the dominant eigenvalue. Of course we do have to worry about floating point overflows and underflows, and the need to incorporate a suitable scaling procedure.