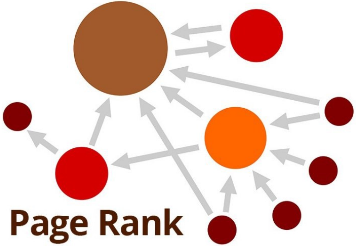
In some applications, knowing the largest and smallest eigenvalues of a matrix A and possibly their corresponding eigenvectors are crucial. For instance

- (Last lecture) Condition number of a matrix A in the linear system $Ax = b$ by $\kappa(A) = \frac{|\lambda_{\max}(A)|}{|\lambda_{\min}(A)|}$. If $\kappa(A)$ was large, the system was unstable/ill-conditioned.
- Page rank



- Principle Component Analysis (PCA)

Q: How can we find the dominant eigenvalues/eigenvectors efficiently?

- Power Method aims to find the dominant eigenvalue of a matrix
- Inverse Power method aims to find the smallest eigenvalue of a matrix.

Recall: Eigenvalue and Eigenvectors

The eigenvalue λ_i and the eigenvector v_i of a matrix $A \in \mathbb{R}^{n \times n}$ satisfy

$$(A - \lambda I_n)v = 0$$

The eigenvalues are the n roots of the characteristic polynomial of degree n given by $p(\lambda) = \det(A - I_n\lambda)$.

Given λ, the eigenvector can be found by solving the linear system

$$(A - \lambda I_n)v = 0$$

If the eigenvector v is known, we can find the eigenvalue:
premultiplying $Av = \lambda v$ by v^T and divinding both sides by $v^T v \neq 0$, we obtain

$$\lambda = \frac{v^T A v}{v^T v} = \frac{v^T A v}{\|v\|_2^2}$$

Motivation Behind Power Method

Given a matrix $A \in \mathbb{R}^{n\times n}$, for any initial point $x^{(0)} \in \mathbb{R}^n$, as q increases

$$A^q x^{(0)} \to \text{ the eigenvector associated to the largest eigenvalue}$$

Note that $A^q = \underbrace{AAA\cdots A}_{q \text{ times}}$

To evaluate

$$A^q x^{(0)} = \underbrace{AA\cdots AAA}_{q \text{ times}} x^{(0)} = \underbrace{AA\cdots AA}_{q-1 \text{ times}}(Ax^{(0)})$$

Let's define

$$x^{(1)} := Ax^{(0)}$$

then

$$A^q x^{(0)} = \underbrace{AA\cdots AA}_{q-1 \text{ times}} x^{(1)} = \underbrace{AA\cdots A}_{q-2 \text{ times}}(Ax^{(1)})$$

Let's define

$$x^{(2)} := Ax^{(1)}$$

then

$$A^{q-1}x^{(1)} = \underbrace{AAAA}_{q-2\ times}\ x^{(2)} = \underbrace{AAAA}_{q-3\ times}(Ax^{(2)})$$

Let's define again

$$x^{(3)} := Ax^{(2)}$$

and continue this process.

In short, we can express an iterative process, given $x^{(0)}$, for $q = 1, 2, 3, \ldots,$

$$x^{(q)} = Ax^{(q-1)}$$

then

$$x^{(q)} = A^q x^{(0)}$$

Example 1

The dominant eigenvalue of the matrix A is $\lambda_1 = 4$ with eigenvector $q^{(1)} = [1, 1]^{\mathsf{T}}$, and the smallest eigenvalue is $\lambda_2 = -1$ with eigenvector $q^{(2)} = [-3, 2]^{\mathsf{T}}$.

$$A = \begin{bmatrix} 1 & 3 \\ 2 & 2 \end{bmatrix}.$$

Let $x^{(0)} = [-5, 5]^{\mathsf{T}}$:

$$x^{(1)} = Ax^{(0)} = \begin{bmatrix} 1 & 3 \\ 2 & 2 \end{bmatrix} \begin{bmatrix} -5 \\ 5 \end{bmatrix} = \begin{bmatrix} 10 \\ 0 \end{bmatrix}$$

$$x^{(2)} = Ax^{(1)} = \begin{bmatrix} 1 & 3 \\ 2 & 2 \end{bmatrix} \begin{bmatrix} 10 \\ 0 \end{bmatrix} = \begin{bmatrix} 10 \\ 20 \end{bmatrix}$$

$$x^{(3)} = Ax^{(2)} = \begin{bmatrix} 1 & 3 \\ 2 & 2 \end{bmatrix} \begin{bmatrix} 10 \\ 20 \end{bmatrix} = \begin{bmatrix} 70 \\ 60 \end{bmatrix}$$

$$x^{(4)} = Ax^{(3)} = \begin{bmatrix} 1 & 3 \\ 2 & 2 \end{bmatrix} \begin{bmatrix} 70 \\ 60 \end{bmatrix} = \begin{bmatrix} 250 \\ 260 \end{bmatrix}$$

What we see is that multiplying a matrix with a random vector resulted in moving the vector close to the dominant eigenvector of A.

Power Method for Coding

Given a matrix $A \in \mathbb{R}^{n\times n}$ with eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ by its *dominant eigenvalue* we mean an eigenvalue λ_i such that $|\lambda_i| > |\lambda_j| \qquad j \neq i$ if it exists.

Power Method is suitable for computation of the dominant eigenvalue and the corresponding eigenvector.

Starting with an arbitrary initial guess $x^{(0)}$ of the dominant eigenvector, we construct a sequence of vectors

$$\{x^{(k)} : k \geq 0\}$$

by the following iterative process

$$x^{(k)} = Ax^{(k-1)}, \qquad k \geq 1$$

Note that

$$x^{(k)} = Ax^{(k-1)} = A^2x^{(k-2)} = A^3x^{(k-3)} = \cdots = A^kx^{(0)}$$

Q: Why $x^{(k)}$ is converging to the dominant eigenvector?

Let assume that $v_1, \ldots, v_n$ are eigenvectors of A, then $x^{(0)}$ be the representation of the initial guess as a linear combination of the eigenvectors

$$x^{(0)} = \sum_{i=1}^{n} c_i v_i$$

$$\begin{aligned}
x^{(k)} &= A^k x^{(0)} \\
&= A^k \sum_{i=1}^{n} c_i v_i \\
&= A^k (c_1 v_1 + c_2 v_2 + \cdots + c_n v_n) \\
&= c_1 A^k v_1 + c_2 A^k v_2 + \cdots + c_n A^k v_n) \\
&= c_1 \lambda_1^k v_1 + c_2 \lambda_2^k v_2 + \cdots + c_n \lambda_n^k v_n) \\
&= \lambda_1^k (c_1 v_1 + c_2 \left(\frac{\lambda_2}{\lambda_1}\right)^k v_2 + \cdots + c_n \left(\frac{\lambda_n}{\lambda_1}\right)^k v_n
\end{aligned}$$

Since

$$|\lambda_i/\lambda_1| < 1, \quad \rightarrow \lim_{k\to\infty} |\lambda_i/\lambda_1| = 0$$

hence the direction of $x^{(k)}$ tends to that of v_1 as $k \to \infty$, assuming $c_1 \neq 0$.

Also,

$$\mu^{(k)} = \frac{(x^{(k)})^T A x^{(k)}}{(x^{(k)})^T (x^{(k)})}$$

we have $\mu^{(k)} \to \lambda_1, k \to \infty$.

To ensure the convergence of $x^{(k)}$ to a nonzero vector of bounded length, we scale $x^{(k)}$ at each iteration

$$v^{(k)} = \frac{x^{(k)}}{\|x^{(k)}\|_2}, \quad k \geq 0$$

Also,

$$\mu^{(k)} = \frac{(v^{(k)})^T A v^{(k)}}{(v^{(k)})^T v^{(k)}} = (v^{(k)})^T A v^{(k)}$$

Power Method

Given $x^{(0)}$, k positive integer

For $i = 1, \ldots, k$ do

Step 1. $v^{(i-1)} = \frac{x^{(i-1)}}{\|x^{(i-1)}\|_2}$

Step 2. $x^{(i)} = Av^{(i-1)}$

Step 3. $\mu_i = {v^{(i-1)}}^{\mathsf{T}} A v^{(i-1)} = {v^{(i-1)}}^{\mathsf{T}} x^{(i)}$

End For

Then the approximated eigenvector is

$$u^{(k)} = \frac{x^{(k)}}{\|x^{(k)}\|}$$

The approximated eigenvalue is μ_k

Spectral gap

Definition

Assume that the dominant and the second dominant eigenvalues of the matrix A be λ_1 and λ_2. The distance between λ_1 and λ_2, that is $|\lambda_1 - \lambda_2|$ is called spectral gap.

- If the spectral gap is large, then $|\lambda_2/\lambda_1|$ is small, in this case the power method converge quickly.
- If the spectral gap is small, then $|\lambda_2/\lambda_1|$ is close to one, the power method converges slowly.

Inverse Power Method

Theorem. If the eigenvalues of the matrix $A \in \mathbb{R}^{n \times n}$ are denoted by

$$\lambda_1, \lambda_2, \ldots, \lambda_n$$

The eigenvalues of the A^{-1} (if inverse exists) are

$$\lambda_1^{-1}, \lambda_2^{-1}, \ldots, \lambda_n^{-1}.$$

The eigenvectors of A^{-1} are the same as A.

Proof. $Av = \lambda v$ implies that $v = \lambda A^{-1} v$, therefore,

$$A^{-1} v = \lambda^{-1} v.$$

Note that eigenvectors are unchanged.

- The largest eigenvalue of A^{-1} is the reciprocal of the smallest magnitude eigenvalue of A.

- Applying power method to A^{-1}, then inverting the resulting, gives the smallest magnitude eigenvalue of A.

- Note that in practice with large systems, $x^{(i)} = A^{-1}x^{(i-1)}$ can be written equivalently as

$$Ax^{(i)} = x^{(i-1)}$$

  and is solved by Gaussian Elimination, or similar methods we learnt.

Inverse Power Method: to find smallest eigenvalue of A

Given $x^{(0)}$, given k a positive integer.

For $i = 1, \ldots, k$ do

Step 1. $v^{(i-1)} = \frac{x^{(i-1)}}{\|x^{(i-1)}\|_2}$

Step 2. Solve the system $Ax^{(i)} = v^{(i-1)}$ for $x^{(i)}$

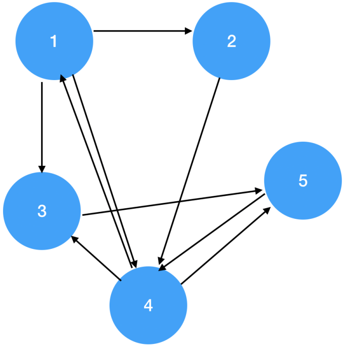
Step 3. $\lambda_i = {v^{(i-1)}}^{\mathsf{T}} A^{-1} v^{(i-1)} = {v^{(i-1)}}^{\mathsf{T}} x^{(i)}$

End For

The largest eigenvalue of A^{-1} is approximately λ_k, and its corresponding eigenvector is $v^{(k)} = \frac{x^{(k)}}{\|x^{(k)}\|_2}$. Then $\frac{1}{\lambda_k}$ gives the smallest eigenvalue of A.

Application: Page Rank

The original google algorithm to rank webpages, it was named after Larry Page one of the founders of google. For example consider five webpage that they have links to each others, as follows.



Imagine a web surfer moves randomly from page to page by uniformly (random variable) clicking a link. Suppose that the web surfer surfs for a very long time, and let X denote the random variable for the page that surfer ends on.

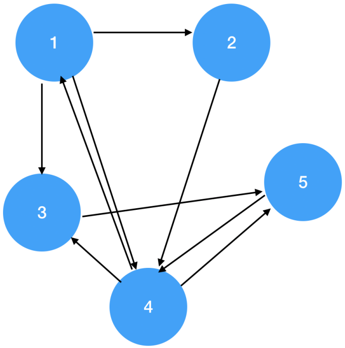
Let's define

$$P(X = i) = r_i$$

where r_i is the probability of the page $X = i$, such that

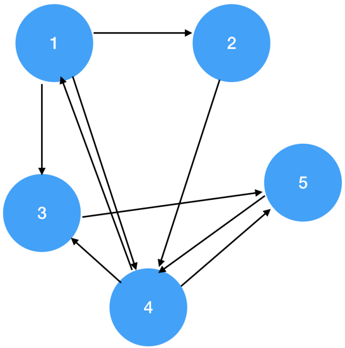
$$r_1 + r_2 + r_3 + r_4 + r_5 = 1$$

The page with the highest probability is the page with the highest rank.
Q: From the given graph, how can we determine the top page with the highest rank?



We need to define the **adjacency matrix** corresponding to the graph

$$A = \begin{pmatrix} 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$



Convert A to **transition probability matrix**:

$$A = \begin{pmatrix} 0 & 1/3 & 1/3 & 1/3 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 1/3 & 0 & 1/3 & 0 & 1/3 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

The transition probability matrix has the following properties:

- row adds to 1
- all elements are nonnegative
- A^T has the max eigenvalue equal to 1, and all the other eigenvalues lie in $(-1, 1)$ and can be imaginary.
- Let v denotes the eigenvector associated eigenvalue 1 then

  $$\frac{v}{\sum_{i=1}^{n} v_i}$$

  gives the vector for probability of different pages, such that

  $$P(X = i) = r_i = \frac{v_i}{\sum_{i=1}^{n} v_i}.$$

Command Window

```
>> A=[0,1/3,1/3,1/3,0;0 0 0 1 0;0 0 0 0 1; 1/3 0 1/3 0 1/3; 0 0 0 1 0]

A =

         0    0.3333    0.3333    0.3333         0
         0         0         0    1.0000         0
         0         0         0         0    1.0000
    0.3333         0    0.3333         0    0.3333
         0         0         0    1.0000         0

>> A'

ans =

         0         0         0    0.3333         0
    0.3333         0         0         0         0
    0.3333         0         0    0.3333         0
    0.3333    1.0000         0         0    1.0000
         0         0    1.0000    0.3333         0
```

```
>> [V,D]=eig(A')

V =

  Columns 1 through 4

   0.2402 + 0.0000i  -0.0000 + 0.0000i  -0.5000 + 0.0000i  -0.2500 - 0.3536i
   0.0801 + 0.0000i   0.7071 + 0.0000i   0.5000 + 0.0000i  -0.0833 + 0.2357i
   0.3203 + 0.0000i   0.0000 + 0.0000i  -0.0000 + 0.0000i  -0.3333 - 0.1179i
   0.7206 + 0.0000i  -0.0000 + 0.0000i   0.5000 + 0.0000i   0.7500 + 0.0000i
   0.5604 + 0.0000i  -0.7071 + 0.0000i  -0.5000 + 0.0000i  -0.0833 + 0.2357i

  Column 5

  -0.2500 + 0.3536i
  -0.0833 - 0.2357i
  -0.3333 + 0.1179i
   0.7500 + 0.0000i
  -0.0833 - 0.2357i
```

```
D =

  Columns 1 through 4

   1.0000 + 0.0000i   0.0000 + 0.0000i   0.0000 + 0.0000i   0.0000 + 0.0000i
   0.0000 + 0.0000i  -0.0000 + 0.0000i   0.0000 + 0.0000i   0.0000 + 0.0000i
   0.0000 + 0.0000i   0.0000 + 0.0000i  -0.3333 + 0.0000i   0.0000 + 0.0000i
   0.0000 + 0.0000i   0.0000 + 0.0000i   0.0000 + 0.0000i  -0.3333 + 0.4714i
   0.0000 + 0.0000i   0.0000 + 0.0000i   0.0000 + 0.0000i   0.0000 + 0.0000i

  Column 5

   0.0000 + 0.0000i
   0.0000 + 0.0000i
   0.0000 + 0.0000i
   0.0000 + 0.0000i
  -0.3333 - 0.4714i
```

```
>> V(:,1)

ans =

    0.2402
    0.0801
    0.3203
    0.7206
    0.5604

>> V(:,1)/sum(V(:,1))

ans =

    0.1250
    0.0417
    0.1667
    0.3750
    0.2917
```