

Lecture 13 - A

§8.2, 8.3 Power Method

T. Takahashi

Department of Mathematics and Statistics
University of New Mexico



We will answer these questions:

- What are eigenvalues?
- What is the power method and what is it used for?
- How quickly does power method converge?
- How can you modify the power method to find other eigenvalues?



Review Eigenvalues



Eigenvalues, Eigenvectors

Definition: Eigenvalues and Eigenvectors

For a matrix A , the scalar-vector pairs (λ, v) such that

$$Av = \lambda v, \quad v \neq 0$$

are **eigenvalue-eigenvectors**. The pair (λ, v) is called the **eigenpair**.

- For an $n \times n$ matrix, we expect n eigenvalues
- The eigenvalues can be complex even if A contains real numbers
- Typically order them $\lambda_1, \lambda_2, \dots, \lambda_n$ from greatest to least in magnitude

Singularity

One of the equivalent ways to say $n \times n$ matrix is singular is to say it has an eigenvalue $\lambda = 0$.



Finding Eigenvalues and Eigenvectors

Example

Find the eigenvalues and eigenvectors of $A = \begin{bmatrix} 3 & 2 \\ 7 & -2 \end{bmatrix}$

- First, calculate the **characteristic equation** $\det(\lambda I - A) = 0$

$$0 = \det(\lambda I - A) = \begin{vmatrix} \lambda - 3 & -2 \\ -7 & \lambda + 2 \end{vmatrix} = \lambda^2 - \lambda - 6 - 14 = \lambda^2 - \lambda - 20$$

- Find the roots to get the eigenvalues:

$$0 = \lambda^2 - \lambda - 20 = (\lambda - 5)(\lambda + 4) \implies \lambda_1 = 5 \quad \lambda_2 = -4$$



Finding Eigenvalues and Eigenvectors

- Solve the system $(\lambda I - A)v_1 = 0$ with $\lambda_1 = 5$

$$5I - A = \begin{bmatrix} 2 & -2 \\ -7 & 7 \end{bmatrix} \implies v_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

- Solve the system $(\lambda I - A)v_1 = 0$ with $\lambda_1 = -4$

$$-4I - A = \begin{bmatrix} -7 & -2 \\ -7 & -2 \end{bmatrix} \implies v_2 = \begin{bmatrix} -2 \\ 7 \end{bmatrix}$$

- *Note:* Any scalar multiple of an eigenvector is an eigenvector



Computing Eigenvalues and Eigenvectors

Theorem

If v is an eigenvector of A then its eigenvalue is

$$\lambda = \frac{(Av)_i}{v_i}$$

Proof:

- By definition we have $Av = \lambda v$ which means

$$(Av)_i = (\lambda v)_i$$

$$(Av)_i = \lambda v_i$$

$$\implies \lambda = \frac{(Av)_i}{v_i}$$

Key Idea

If we know a particular eigenvector v , then we can find its corresponding eigenvalue λ very easily.

Theorem

If (v, λ) is an eigenpair then

$$A^n v = \lambda^n v,$$

where n is any positive integer.

Proof:

- Start with definition

$$Av = \lambda v$$

- Multiply both sides by A

$$AAv = \lambda Av$$

- Factor

$$A(Av) = (Av)\lambda$$

- By definition we have

$$A^2 v = (\lambda v)\lambda = \lambda^2 v$$



Eigenvalues?

- Find the eigenvalues of

$$A = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$$

- Solve $0 = \det(\lambda I - A)$

$$0 = \det(\lambda I - A) = \begin{vmatrix} \lambda - 2 & 0 \\ 0 & \lambda - 2 \end{vmatrix} = (\lambda - 2)(\lambda - 2) \implies \lambda = 2$$



Eigenvalues?

- Find the eigenvalues of

$$A = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$$

- Solve $0 = \det(\lambda I - A)$

$$0 = \det(\lambda I - A) = \begin{vmatrix} \lambda - 2 & 0 \\ 0 & \lambda - 2 \end{vmatrix} = (\lambda - 2)(\lambda - 2) \implies \lambda = 2$$

- Next, find the eigenvalues of

$$B = \begin{bmatrix} 2 & -1 \\ 0 & 2 \end{bmatrix}$$

- Solve $0 = \det(\lambda I - A)$

$$0 = \det(\lambda I - A) = \begin{vmatrix} \lambda - 2 & 1 \\ 0 & \lambda - 2 \end{vmatrix} = (\lambda - 2)(\lambda - 2) \implies \lambda = 2 \text{ same}$$



Eigenvectors?

- Find the eigenvectors of

$$A = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}, \quad \lambda = 2$$

- Solve $(\lambda I - A)v = 0$

$$\begin{bmatrix} \lambda - 2 & 0 \\ 0 & \lambda - 2 \end{bmatrix} v = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} v \implies \text{choose } v_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, v_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$



Eigenvectors?

- Find the eigenvectors of

$$A = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}, \quad \lambda = 2$$

- Solve $(\lambda I - A)v = 0$

$$\begin{bmatrix} \lambda - 2 & 0 \\ 0 & \lambda - 2 \end{bmatrix} v = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} v \implies \text{choose } v_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, v_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

- What about

$$B = \begin{bmatrix} 2 & 1 \\ 0 & 2 \end{bmatrix}, \quad \lambda = 2?$$

- Solve $(\lambda I - A)v = 0$

$$\begin{bmatrix} \lambda - 2 & -1 \\ 0 & \lambda - 2 \end{bmatrix} v = \begin{bmatrix} 0 & -1 \\ 0 & 0 \end{bmatrix} v \implies \text{only have } v = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$



Diagonalizable

Definition: Diagonalizable

When a matrix A has a full set of independent eigenvectors it is said to be **diagonalizable**. Otherwise, the matrix is **deficient**.

Matrix A - Diagonalizable

$$A = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}, \quad \lambda = 2, \quad v_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad v_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

Matrix B - Deficient

$$B = \begin{bmatrix} 2 & 1 \\ 0 & 2 \end{bmatrix}, \quad \lambda = 2, \quad v = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$



Power Method



The Largest Eigenvalue

- Suppose that A is $n \times n$ and has real eigenvalues

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

- Assume A is diagonalizable
- In other words, $\{v_1, v_2, \dots, v_n\}$ forms a basis for $\mathbb{R}^n$.

Question

How can we find the dominant eigenvector v_1 ?



Finding the Largest Eigenvalue

Part 1

- Start with a guess $x^{(0)}$. Since $\{v_1, v_2, \dots, v_n\}$ is a basis, we write

$$x^{(0)} = c_1 v_1 + \dots + c_n v_n$$

- For simplicity, assume all $c_j = 1$:

$$x^{(0)} = v_1 + \dots + v_n$$

- Now consider the sequence $x^{(i)} = Ax^{(i-1)}$:

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

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

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

$$\vdots$$

$$x^{(k+1)} = Ax^{(k)}$$



Finding the Largest Eigenvalue

Part 2

- With recursion, we can deduce that

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

- Using the property $A^n v = \lambda^n v$, we get

$$\begin{aligned} x^{(k)} &= A^k x^{(0)} \\ &= A^k (v_1 + \cdots + v_n) \\ &= \lambda_1^k v_1 + \lambda_2^k v_2 + \cdots + \lambda_n^k v_n. \end{aligned}$$

- Factor out λ_1^k :

$$x^{(k)} = \lambda_1^k \left(v_1 + \left(\frac{\lambda_2}{\lambda_1} \right)^k v_2 + \cdots + \left(\frac{\lambda_n}{\lambda_1} \right)^k v_n \right)$$

Question

What happens to $x^{(k)}$ if we let $k \rightarrow \infty$?

Finding the Largest Eigenvalue

Part 3

- Notice that the fractions are all less than 1

$$x^{(k)} = \lambda_1^k \left(v_1 + \underbrace{\left(\frac{\lambda_2}{\lambda_1} \right)^k}_{<1} v_2 + \cdots + \left(\frac{\lambda_n}{\lambda_1} \right)^k v_n \right)$$

- Therefore, in the limit

$$\begin{aligned} \lim_{k \rightarrow \infty} x^{(k)} &= \lim_{k \rightarrow \infty} \lambda_1^k \left(v_1 + \left(\frac{\lambda_2}{\lambda_1} \right)^k v_2 + \cdots + \left(\frac{\lambda_n}{\lambda_1} \right)^k v_n \right) \\ &= \lim_{k \rightarrow \infty} \lambda_1^k (v_1 + 0 + \cdots + 0) \\ &= \lim_{k \rightarrow \infty} \lambda_1^k v_1 \end{aligned}$$

- That means $x^{(k)}$ approximates the eigenvector v_1



The Power Method

Method: Power Method

- **Given:** Given A with independent eigenvectors
- **Output:** Largest Eigenvalue λ_1
- **Idea:** Extremely simple and efficient method to find v_1

How it works:

- Start with any non-zero initial guess
- Keep multiplying the vector by A
- Eventually, the vector becomes parallel to v_1

If we know an eigenvector, then finding its eigenvalue is nearly free



The Power Method

Memory efficient implementation:

```
1 initialize old
2 for k=1:kmax
3     new = A*old           % get next vector
4     lam = new(1)/old(1)   % estimate eigenvalue
5     old = new             % update
```

Getting the Eigenvalue

Since $x^{(k+1)} = Ax^{(k)}$ and $\lambda = (Av)_i/v_i$, we have the formula

$$\lambda \approx \frac{(Ax^{(k)})_1}{x_1^{(k)}} = \frac{x_1^{(k+1)}}{x_1^{(k)}}$$



Power Method by Hand

Example

Approximate the largest eigenvalue of $A = \begin{bmatrix} 3 & 2 \\ 7 & -2 \end{bmatrix}$ with two iterations of the power method using the initial guess $x^{(0)} = [1 \ 0]^T$.

- $x^{(0)} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$
- $x^{(1)} = \begin{bmatrix} 3 & 2 \\ 7 & -2 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 3 \\ 7 \end{bmatrix}$
- $x^{(2)} = \begin{bmatrix} 3 & 2 \\ 7 & -2 \end{bmatrix} \begin{bmatrix} 3 \\ 7 \end{bmatrix} = \begin{bmatrix} 23 \\ 7 \end{bmatrix}$
- $\lambda \approx 23/3 = 7\frac{2}{3}$ (true answer is $\lambda = 5$)



Normalization



Continue Iterating

Question

What happens if we kept going?

- $x^{(0)} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$
- $x^{(1)} = \begin{bmatrix} 3 & 2 \\ 7 & -2 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 3 \\ 7 \end{bmatrix}$
- $x^{(2)} = \begin{bmatrix} 3 & 2 \\ 7 & -2 \end{bmatrix} \begin{bmatrix} 3 \\ 7 \end{bmatrix} = \begin{bmatrix} 23 \\ 7 \end{bmatrix}$
- $x^{(3)} = \begin{bmatrix} 83 \\ 147 \end{bmatrix} \dots$
- $x^{(10)} = \begin{bmatrix} 318383 \\ 252847 \end{bmatrix}, \quad \lambda \approx 4.5883$

What is happening?

The λ gets more accurate but the vector keeps growing

Overflow or Underflow

- Eventually, the vector will **overflow**

```
1 x =  
2  
3   Inf  
4   Inf
```

- The multiplication causes the magnitude of the vector to grow each time

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

- On the other hand, if $\|A\| < 1$, then

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

would cause the vector to shrink and eventually **underflow**



Scale the Vectors

- We need to iterate to improve
- But iteration leads to overflow/underflow

The Fix

- ▶ Any constant multiple of an eigenvector is still an eigenvector
- ▶ We can scale down $x^{(k)}$ so it doesn't overflow



Definition: Normalization

Normalizing a vector v means scaling it in a way that makes its norm equal to one, e.g.

$$\|cv\| = 1.$$

- Example:

$$v = \begin{bmatrix} 3 \\ 4 \end{bmatrix}$$

- We can use the 2-norm

$$\|v\|_2 = \sqrt{3^2 + 4^2} = 5$$

- Scale down by $c = 1/5$

$$cv = \frac{1}{5} \begin{bmatrix} 3 \\ 4 \end{bmatrix} = \begin{bmatrix} 3/5 \\ 4/5 \end{bmatrix}.$$



1 2-Norm

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

2 Infinity norm

$$\|x\|_\infty = \max(|x_1|, |x_2|, \dots, |x_n|)$$

3 1-Norm

$$\|x\|_1 = |x_1| + |x_2| + \cdots + |x_n|$$

- Suppose all elements $x_i > 0$
- Then normalizing

$$\frac{x}{\|x\|_1}$$

causes the new vector to have elements that add up to 1



Need for Normalization

- Starting with $x^{(0)}$, if we iterate $x^{(i)} = Ax^{(i-1)}$ infinitely...

$$\begin{aligned}\lim_{k \rightarrow \infty} x^{(k)} &= \lim_{k \rightarrow \infty} \lambda_1^k \left(v_1 + \left(\frac{\lambda_2}{\lambda_1} \right)^k v_2 + \cdots + \left(\frac{\lambda_n}{\lambda_1} \right)^k v_n \right) \\ &= \lim_{k \rightarrow \infty} \lambda_1^k (v_1 + 0 + \cdots + 0) \\ &= \lim_{k \rightarrow \infty} \lambda_1^k v_1\end{aligned}$$

- If $|\lambda_1| < 1$, then the entire vector converges to zero
- If $|\lambda_1| > 1$, then the vector diverges

Key Idea

- Simple method $x^{(i)} = Ax^{(i-1)}$ will diverge or give us garbage.
- But, scaling $x^{(i)} := \frac{Ax^{(i-1)}}{\|Ax^{(i-1)}\|}$ fixes both issues!



The Power Method (with normalization)

New algorithm with normalization

```
1 initialize old
2 for k=1:kmax
3     new = A*old
4     lam = new(1)/old(1)
5     old = new/norm(new,1)    % scale down with 1-norm
```

- Add one line of code to scale down the vector
- Avoid underflow and overflow
- This is the standard power method



Convergence Rate

Theorem

Let A have eigenvalues ordered as

$$|\lambda_1| > |\lambda_2| \geq \cdots \geq |\lambda_n|.$$

Then the power method converges linearly, and the error satisfies

$$E_k = \mathcal{O}\left(\left|\frac{\lambda_2}{\lambda_1}\right|^k\right).$$

- The convergence speed depends on the eigenvalue gap
- Larger separation between $|\lambda_1|$ and $|\lambda_2|$ gives faster convergence
- Power method is linearly convergent (not fast, but cheap)



Idea Behind the Convergence Rate

- After k iterations,

$$x^{(k)} = \lambda_1^k \left(c_1 v_1 + \left(\frac{\lambda_2}{\lambda_1} \right)^k c_2 v_2 + \cdots + \left(\frac{\lambda_n}{\lambda_1} \right)^k c_n v_n \right).$$

- Since the error is $E = |x^{(k)} - v_1|$ and $v_1 \approx v_1 \lambda_1^k$, asymptotically we have

$$E \leq M \left(\frac{\lambda_2}{\lambda_1} \right)^k$$

since λ_2/λ_1 is the most dominant fraction

Key Idea

Convergence is always linear. The convergence rate depends on the ratio between the largest and second largest eigenvectors.



Other Eigenvalues



Smallest Eigenvalue

Question

How can we find the smallest eigenvalue of A .

Theorem

If λ is an eigenvalue of A , then $\frac{1}{\lambda}$ is an eigenvalue of A^{-1} .

Proof:

- Suppose $Av = \lambda v$ for some eigenvector v .
- Then, multiplying both sides by A^{-1} gives $A^{-1}Av = A^{-1}\lambda v$.
- Simplifying, we get $v = \lambda A^{-1}v$ or

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

- By definition, $\frac{1}{\lambda}$ is an eigenvalue of A^{-1} .



Reciprocals

- Suppose the eigenvalues of A are

$$\lambda_1 = 10, \quad \lambda_2 = 5, \quad \lambda_3 = 2$$

- Then the eigenvalues of A^{-1} must be

$$\lambda_1 = \frac{1}{2}, \quad \lambda_2 = \frac{1}{5}, \quad \lambda_3 = \frac{1}{10}$$



Reciprocals

- Suppose the eigenvalues of A are

$$\lambda_1 = 10, \quad \lambda_2 = 5, \quad \lambda_3 = 2$$

- Then the eigenvalues of A^{-1} must be

$$\lambda_1 = \frac{1}{2}, \quad \lambda_2 = \frac{1}{5}, \quad \lambda_3 = \frac{1}{10}$$

- Notice that the *smallest* eigenvalue of A is the reciprocal of the *largest* eigenvalue of A^{-1}

$$2 \quad \text{v.s.} \quad \frac{1}{2}$$

Key Idea

To find the smallest eigenvalue of A , do power method on A^{-1}



Method: Inverse Power Method

- **Given:** Matrix A with independent eigenvectors
- **Output:** Smallest eigenvalue in magnitude λ_n
- **Idea:** Simply do power method on A^{-1}

Note:

- You are never supposed to compute A^{-1}
- Compute a factorization instead



Inverse Power Method Algorithm

Algorithm:

```
1 factorize A
2 initialize old
3 for k=1 to kmax
4     new = solve(A, old)           % Solve system A * new = old
5     lam = new(1) / old(1)         % Estimate smallest eigenvalue
6     old = new / max(new)         % Normalize vector
7 end
8
9 return 1/lam
```

- Note: Instead of directly inverting A , we solve $A \mathbf{new} = \mathbf{old}$ at each iteration, which is more computationally efficient.
- This yields the smallest eigenvalue of A , with $\lambda_{\min} \approx \frac{1}{\text{lam}}$.



Remarks on the Inverse Power Method

Why would you want to know λ_n ?

Condition Number

Suppose $\lambda_1, \lambda_2, \dots, \lambda_n$ are the eigenvalues of A from greatest to least in magnitude. Then the condition number is

$$\kappa(A) = \frac{|\lambda_1|}{|\lambda_n|}.$$

And the convergence rate?

Convergence Rate Formula

For error E_k at iteration k , we have:

$$E_k = \mathcal{O} \left(\left| \frac{\lambda_{n-1}}{\lambda_n} \right|^k \right).$$

Other Eigenvalues

- Power method finds λ_1
- Inverse power method finds λ_n
- There are still $n - 2$ other eigenvalues

$$\lambda_2, \lambda_3, \dots, \lambda_{n-1}$$

Question

How do you find the other eigenvalues?



Other Eigenvalues

- Power method finds λ_1
- Inverse power method finds λ_n
- There are still $n - 2$ other eigenvalues

$$\lambda_2, \lambda_3, \dots, \lambda_{n-1}$$

Question

How do you find the other eigenvalues?

- Homework!



- Why do we care about the dominant eigenvector?
- What is a Markov chain?
- How are Markov chains and the power method used together?

