

## \* 2.1 Vector and Matrix Norms.

Def: A norm is a function  $\|\cdot\|: \mathbb{R}^n \rightarrow \mathbb{R}$  that assigns a real-valued length to each vector. It must satisfy the following conditions:

- 1)  $\|\vec{x}\| > 0$  if  $\vec{x} \neq 0$ , and  $\|0\|=0$  (positive definiteness)
- 2)  $\|\alpha\vec{x}\| = |\alpha| \|\vec{x}\|$  (absolute homogeneity)
- 3)  $\|\vec{x} + \vec{y}\| \leq \|\vec{x}\| + \|\vec{y}\|$  (triangle inequality).

E.g. The Euclidean norm is defined by.

$$\|\vec{x}\|_2 = \left( \sum_{i=1}^n |x_i|^2 \right)^{\frac{1}{2}}$$

Why is it a norm?

1)  $\|0\|_2 = 0 \quad \checkmark$

If  $\vec{x} \neq 0$ , then there exists  $x_j$  s.t.  $x_j \neq 0$

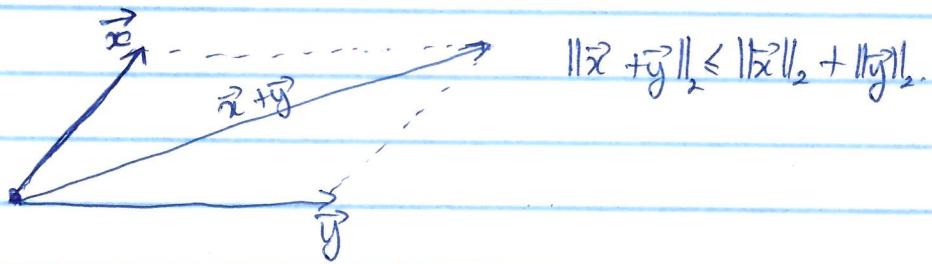
$$\Rightarrow x_j^2 > 0$$

$$\|\vec{x}\|_2 = \sqrt{\sum_{i=1}^n |x_i|^2} > 0$$

2)  $\|\alpha\vec{x}\|_2 = \left( \sum_{i=1}^n (\alpha x_i)^2 \right)^{\frac{1}{2}} = |\alpha| \|\vec{x}\|_2 \quad \checkmark$

3) Will follow the Cauchy-Schwarz inequality.

Geometrically,



31

• Cauchy - Schwarz inequality:

For all  $\vec{x}, \vec{y}$  in  $\mathbb{R}^n$

$$|\langle \vec{x}, \vec{y} \rangle| \leq \|\vec{x}\|_2 \|\vec{y}\|_2.$$

equivalently

$$\left| \sum_{i=1}^n x_i y_i \right| \leq \left( \sum_{i=1}^n |x_i|^2 \right)^{1/2} \left( \sum_{i=1}^n |y_i|^2 \right)^{1/2}.$$

Simple proof:

$$|\langle x, y \rangle| = \|x\|_2 \|y\|_2 |\cos \theta| \leq \|x\|_2 \|y\|_2$$

where  $\theta$  is the angle between  $x$  and  $y$ .

3) Want to show  $\|x + y\|_2 \leq \|x\|_2 + \|y\|_2$ .

$$\|x + y\|_2^2 = \sum_{i=1}^n (x_i + y_i)^2$$

$$= \sum_{i=1}^n (x_i^2 + 2x_i y_i + y_i^2)$$

$$= \sum_{i=1}^n x_i^2 + 2 \sum_{i=1}^n x_i y_i + \sum_{i=1}^n y_i^2$$

$$= \|x\|_2^2 + 2 \sum_{i=1}^n x_i y_i + \|y\|_2^2.$$

$$\begin{aligned} &\stackrel{\text{C-S}}{\leq} \|x\|_2^2 + 2 \|x\|_2 \|y\|_2 + \|y\|_2^2 \\ &= (\|x\|_2 + \|y\|_2)^2. \end{aligned}$$

\*  $p$ -norms

For any  $p \geq 1$ , we define

$$\|x\|_p = \left( \sum_{i=1}^n |x_i|^p \right)^{1/p}$$

For  $p$ -norms, the triangle inequality is called Minkowski Inequality.

E.g. • Euclidean norm is 2-norm.  $\|x\|_2$ .

$$\cdot 1\text{-norm: } \|x\|_1 = \sum_{i=1}^n |x_i|$$

$$\cdot \infty\text{-norm: } \|x\|_\infty = \max_{1 \leq i \leq n} |x_i|.$$

Q: Why do we require  $p \geq 1$ ?

What happens if  $0 \leq p < 1$ ? Condition 3) fails.

E.g. consider  $p = 1/2$ .

$$x = (1, 0) \quad \text{and} \quad y = (0, 1).$$

$$\Rightarrow x + y = (1, 1).$$

$$\|x\|_p = \|x\|_{1/2} = \left( \sum_{i=1}^2 |x_i|^{1/2} \right)^2 = 1.$$

$$\|y\|_p = 1.$$

$$\text{But } \|x+y\|_p = (\sqrt{1+1})^2 = 4.$$

$$\Rightarrow \|x\|_p + \|x+p y\|_p = 2 < \|x+y\|_p.$$

\* A-norm of Vectors.

Given a positive definite matrix  $A \in \mathbb{R}^{n \times n}$ , the A-norm on  $\mathbb{R}^n$  is defined by

$$\|x\|_A = \sqrt{x^T A x}$$

\* Matrix Norms:

Def: A matrix norm is a function that assigns to each  $A \in \mathbb{R}^{n \times n}$  a real number  $\|A\|$ , called the norm of  $A$ , such that the following properties hold:

for  $A, B \in \mathbb{R}^{n \times n}$ ,

- 1)  $\|A\| \geq 0$ , and  $\|A\| = 0$  only if  $\|A\| = 0$ .
- 2)  $\|A + B\| \leq \|A\| + \|B\|$ .
- 3)  $\|xA\| = |x| \|A\|$ .
- + 4)  $\|AB\| \leq \|A\| \|B\|$ . (submultiplicativity).

(generalization of the Cauchy-Schwarz inequality).

E.g. The Frobenius norm is defined by

$$\|A\|_F = \sqrt{\sum_{i=1}^n \sum_{j=1}^n |a_{ij}|^2} = \sqrt{\sum_{j=1}^n \|\vec{a}_j\|_2^2}$$

where  $\vec{a}_j$  is the  $j$ th column of  $A$ .

Further

$$1) \|A\|_F \geq 0 \text{ and } \checkmark.$$

$$2) \checkmark.$$

$$3) \|A + B\|_F = \sqrt{\sum_{j=1}^n \|\vec{a}_j + \vec{b}_j\|_2^2}$$

$$\stackrel{\text{triangle ineq.}}{\leq} \sqrt{\sum_{j=1}^n (\|\vec{a}_j\|_2^2 + \|\vec{b}_j\|_2^2)}$$

$$= \sqrt{\sum_j \|\vec{a}_j\|_2^2 + 2 \sum_j \|\vec{a}_j\|_2 \|\vec{b}_j\|_2 + \sum_j \|\vec{b}_j\|_2^2}$$

$$\stackrel{\text{CS ineq}}{\leq} \sqrt{\sum_j \|\vec{a}_j\|_2^2 + 2 \sqrt{\sum_j \|\vec{a}_j\|_2^2} \sqrt{\sum_j \|\vec{b}_j\|_2^2} + \sum_j \|\vec{b}_j\|_2^2}$$

$$\leq \sqrt{(\sum_j \|\vec{a}_j\|_2^2 + \sum_j \|\vec{b}_j\|_2^2)^2} = \|A\|_F + \|B\|_F$$

(34)

4) Let  $C = AB$ .

$$\begin{aligned}
 \|AB\|_F^2 &= \|C\|_F^2 = \sum_i \sum_j |c_{ij}|^2 \\
 &= \sum_i \sum_j |\vec{a}_i^\top \vec{b}_j|^2 \\
 &\leq \sum_i \sum_j \|\vec{a}_i\|_2^2 \|\vec{b}_j\|_2^2 \\
 &= \sum_i \|\vec{a}_i\|_2^2 \left( \sum_j \|\vec{b}_j\|_2^2 \right) \\
 &= \sum_i \|\vec{a}_i\|_2^2 \|B\|_F^2 \\
 &= \|A\|_F^2 \|B\|_F^2.
 \end{aligned}$$

Every vector norm on  $\mathbb{R}^n$  can be used to define a matrix norm on  $\mathbb{R}^{n \times n}$  in a natural way.

Given a vector norm  $\|\cdot\|_v$ , the matrix norm induced by  $\|\cdot\|_v$  is defined by

$$\|A\|_M = \max_{x \neq 0} \frac{\|Ax\|_v}{\|x\|_v}.$$

Thm: The induced norm is a matrix norm.

Some of the most important matrix norms are induced by  $p$ -norms. For  $1 \leq p \leq \infty$ , the matrix  $p$ -norm is defined by

$$\|A\|_p = \max_{x \neq 0} \frac{\|Ax\|_p}{\|x\|_p}$$

E.g. 1)  $1$ -norm:

$$\|A\|_1 = \max_{x \neq 0} \frac{\|Ax\|_1}{\|x\|_1} = \max_{x \in \mathbb{R}^n, \|x\|_1=1} \|Ax\|_1 = \max_{1 \leq j \leq n} \sum_{i=1}^n |a_{ij}|$$

What is it equal to?

- Maximum of  $1$ -norm of column vectors of  $A$ .

2)  $\infty$ -norm:

$$\|A\|_\infty = \max_{\|x\|_\infty=1} \|Ax\|_\infty = \max_{1 \leq i \leq n} \sum_{j=1}^n |a_{ij}|$$

What is it equal to?

- Maximum of  $1$ -norm of row vectors of  $A$ .

3)  $2$ -norm: (spectral norm)

$$\|A\|_2 = \max_{\|x\|_2=1} \|Ax\|_2$$

the largest singular value of  $A$  (will explain in later lectures)

- $D$  is diagonal matrix.

$$\|D\|_2 = \max_{1 \leq i \leq n} |d_{ii}|$$

(26)

E.g.  $A = \begin{bmatrix} 1 & 0 & 1 \\ 2 & 3 & 4 \\ -1 & 1 & 2 \end{bmatrix}$

$\cdot \|A\|_1 ? \quad \sum_{i=1}^n |a_{i1}| = 1+2+1=4.$

$$\sum_{i=1}^n |a_{i2}| = 0+3+1=4.$$

$$\sum_{i=1}^n |a_{i3}| = 1+4+2=\boxed{7}.$$

$$\Rightarrow \|A\|_1 = 7.$$

$\cdot \|A\|_\infty ? \quad \sum_{j=1}^3 |a_{ij}| = 2$

$$\sum_{j=1}^3 |a_{2j}| = \boxed{9}$$

$$\sum_{j=1}^3 |a_{3j}| = 4.$$

$$\|A\|_\infty = 9.$$

(37)

## \*5.2 Basic facts about eigenvalues and eigenvectors

$A \in \mathbb{C}^{n \times n}$  That is,  $A$  is an  $n \times n$  matrix whose entries  $a_{ij}$  are complex numbers.

An Eigenvalue of an  $n \times n$  matrix  $A$  is a (complex) number  $\lambda$  such that there exists a vector  $\vec{v} \in \mathbb{C}^n, \vec{v} \neq 0$ , for which

$$A\vec{v} = \lambda\vec{v}$$

such  $\vec{v}$  is called an eigenvector of  $A$ .

$\Rightarrow \lambda$  is called the eigenvalue of  $A$  associated with the eigenvector  $v$ .

$(\lambda, v)$  is called an eigenpair of  $A$ .

E.g. Given

$$A = \begin{bmatrix} 8 & -1 & -5 \\ -4 & 4 & -2 \\ 18 & -5 & 7 \end{bmatrix}$$

Then

$$\begin{bmatrix} 8 & -1 & -5 \\ -4 & 4 & -2 \\ 18 & -5 & 7 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}$$

$$A \underbrace{\begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}}_{\vec{v}} = 1 \cdot \vec{v}$$

so  $\vec{v}$  is an eigenvector associated with the eigenvalue  $\lambda = 1$ .

E.g. Let  $A = \begin{bmatrix} 3 & 2 \\ 2 & 6 \end{bmatrix}$ . One of its eigenvector is

$\begin{bmatrix} 1/\sqrt{5} \\ 2/\sqrt{5} \end{bmatrix}$  and its corresponding eigenvalue is 7.

$$\begin{bmatrix} 3 & 2 \\ 2 & 6 \end{bmatrix} \begin{bmatrix} 1/\sqrt{5} \\ 2/\sqrt{5} \end{bmatrix} = 7 \begin{bmatrix} 1/\sqrt{5} \\ 2/\sqrt{5} \end{bmatrix}$$

(38)

Thm: Eigenvectors corresponding to distinct eigenvalues are linearly independent.

Corollary: If  $A$  has  $n$  distinct eigenvalues, then  $A$  is non defective.

Q: How many linearly independent eigenvectors can an  $n \times n$  matrix have?

A: anywhere from 1 to  $n$ .

E.g.  $n = 3$ .

$$\begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix} \text{ has } 3 \text{ lin. ind. eigenvectors.}$$

$$\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

$$\begin{bmatrix} 2 & 1 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix} \text{ has } 2 \text{ lin. ind. eigenvectors.}$$

$$\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

$$\begin{bmatrix} 2 & 1 & 0 \\ 0 & 2 & 1 \\ 0 & 0 & 2 \end{bmatrix} \text{ has } 1 \text{ lin. ind. eigenvector.}$$

$$\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$

Remark: If a matrix has an eigenvalue of multiplicity greater than 1, it may be defective or non defective.

(39)

Clearly, any non zero scalar multiple of an eigenvalue eigenvector is also an eigenvector

$$Av = \lambda v \Rightarrow A(\alpha v) = \lambda(\alpha v) \text{ for } \alpha \in \mathbb{C} \setminus \{0\}$$

The set of all eigenvalues of  $A$  is called the spectrum of  $A$ .

Q: How to find eigenpairs of  $A$ ?

Observe that if  $\vec{v}$  is an eigenvector and  $\lambda$  its corresponding eigenvalue, then

$$A\vec{v} = \lambda\vec{v}$$

$$\Leftrightarrow (A - \lambda I)\vec{v} = \vec{0},$$

where  $I$  is the  $n \times n$  identity matrix  
and  $\vec{0}$  is a vector of all 0's.

Fact: (from linear algebra) in order for  $(A - \lambda I)\vec{v} = \vec{0}$  to hold for a vector  $\vec{v} \neq \vec{0}$ ,  $\det(A - \lambda I) = 0$ .

Fact:  $\det(A - \lambda I)$  is an  $n$ -th-degree polynomial in  $\lambda$ , from which we can get the  $n$  values of  $\lambda$  that are the eigenvalues of  $M$ . For any such value, say  $\lambda^*$ , we can solve the equation  $A\vec{x} = \lambda^*\vec{x}$  for  $\vec{x}$  to obtain an eigenvector associated with  $\lambda^*$ .

E.g.  $A = \begin{bmatrix} 3 & 2 \\ 2 & 6 \end{bmatrix}$

then  $A - \lambda I = \begin{bmatrix} 3 - \lambda & 2 \\ 2 & 6 - \lambda \end{bmatrix}$

$$\det(A - \lambda I) = (3 - \lambda)(6 - \lambda) - 4.$$

(40)

Set  $\det(A - \lambda I) = 0$   
 $\lambda^2 - 9\lambda + 14 = 0.$

$\lambda = 7$  or  $\lambda = 2$

$\uparrow$  called the first principal eigenvalue or the largest eigenvalue in magnitude.

We solve

$$\begin{bmatrix} 3 & 2 \\ 2 & 6 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = 7 \begin{bmatrix} x \\ y \end{bmatrix}$$

$$\Rightarrow \begin{cases} 3x + 2y = 7x \\ 2x + 6y = 7y \end{cases}$$

$$\Rightarrow y = 2x.$$

Set  $x = 1$ , we get  $y = 2$ .

$\Rightarrow$  A possible eigenvector is

$$\begin{bmatrix} 1 \\ 2 \end{bmatrix}, \text{ corresponding}$$

to the eigenvalue 7. But we prefer a unit eigenvector.

~~Step 1~~

$$\Rightarrow \begin{bmatrix} 1/\sqrt{5} \\ 2/\sqrt{5} \end{bmatrix}$$

Similarly,  $\begin{bmatrix} 2/\sqrt{5} \\ -1/\sqrt{5} \end{bmatrix}$  is an eigenvector associated with unit

the eigenvalue 2.

Q: Is there a faster way to find eigenpairs of A?

(41)

\* Simple eigenvalue/eigenvector relationships:

1)  $A$  is nonsingular if and only if  $0$  is not an eigenvalue of  $A$ .

2) ~~Eigenvalues and eigenvectors of  $\bar{A}$ :~~

2) Eigenvalues and eigenvectors of powers of  $A$ :

$$Av = \lambda v \Rightarrow AAv = A\lambda v$$

$$\bar{A}^2 v = \lambda(Av)$$

$$= \lambda^2 v.$$

In general, the eigenvalues of  $A^k$  are  $\lambda^k$ , and the eigenvectors are the same as those of  $A$ .

3) The eigenvalues of a triangular matrix are its diagonal entries.

The eigenvalues of a block triangular matrix

$$A = \begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1k} \\ A_{21} & \ddots & \cdots & A_{2k} \\ \vdots & & \ddots & \vdots \\ & & & A_{kk} \end{bmatrix}$$

are the union of the eigenvalues of the diagonal submatrices  $A_{11}, A_{22}, \dots, A_{kk}$ .

- A matrix  $A$  is called defective (or not semisimple) if it does not have a set of  $n$  linearly independent eigenvectors.

$A$  has  $n$  linearly independent eigenvectors  $\Rightarrow A$  is non defective or semisimple.

### 5.3 The power method.

The power method is an algorithm for computing the principal or dominant eigenvalue and its corresponding eigenvector.

Thm: Let  $A$  be an  $n \times n$  nondefective matrix with  $\{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n\}$  a set of  $n$  linearly independent eigenvectors and  $\{\lambda_1, \lambda_2, \dots, \lambda_n\}$  their associated eigenvalues, respectively. That is,

$$A\vec{v}_j = \lambda_j \vec{v}_j, \quad 1 \leq j \leq n.$$

Suppose that

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

Given an initial vector  $\vec{x}^{(0)}$ , suppose that

$$\vec{x}^{(0)} = \alpha_1 \vec{v}_1 + \alpha_2 \vec{v}_2 + \dots + \alpha_n \vec{v}_n$$

and  $\alpha_1 \neq 0$ . Define the sequence of vectors

$$\vec{x}^{(k)} = \frac{A\vec{x}^{(k-1)}}{\|A\vec{x}^{(k-1)}\|_\infty} = \frac{A^k \vec{x}^{(0)}}{\|A^k \vec{x}^{(0)}\|_\infty}, \quad k = 1, 2, 3, \dots$$

Then

$$\vec{x}^{(k)} \rightarrow \alpha_1 \vec{v}_1 / \|\alpha_1 \vec{v}_1\|_\infty \text{ as } k \rightarrow \infty$$

and  $\frac{\vec{x}^{(k)} * A\vec{x}^{(k)}}{\vec{x}^{(k)} * \vec{x}^{(k)}}$   $\rightarrow \lambda_1$  as  $k \rightarrow \infty$ .

In other words, as  $k$  becomes very large,  $\vec{x}^{(k)}$  is close to the eigenvector  $\frac{\alpha_1 \vec{v}_1}{\|\alpha_1 \vec{v}_1\|_\infty}$  associated to the dominant eigenvalue  $\lambda_1$ .

(43)

$$\text{Pf: } \vec{x}^{(k)} \neq \cancel{A\vec{x}^{(k-1)}} = \cancel{A^2\vec{x}^{(k-2)}} = \dots = \cancel{A^k\vec{x}^{(0)}}$$

$$A^k \vec{x}^{(0)} = \alpha_1 A^k \vec{v}_1 + \alpha_2 A^k \vec{v}_2 + \dots + \alpha_n A^k \vec{v}_n \\ = \alpha_1 \lambda_1^k \vec{v}_1 + \alpha_2 \lambda_2^k \vec{v}_2 + \dots + \alpha_n \lambda_n^k \vec{v}_n.$$

$$= \lambda_1^k \left[ \alpha_1 \vec{v}_1 + \alpha_2 \left( \frac{\lambda_2}{\lambda_1} \right)^k \vec{v}_2 + \dots + \alpha_n \left( \frac{\lambda_n}{\lambda_1} \right)^k \vec{v}_n \right].$$

$$\vec{x}^{(k)} = \frac{A\vec{x}^{(k-1)}}{\|A\vec{x}^{(k-1)}\|_\infty} \\ = \frac{A \frac{A\vec{x}^{(k-2)}}{\|A\vec{x}^{(k-2)}\|_\infty}}{\|A \frac{A\vec{x}^{(k-2)}}{\|A\vec{x}^{(k-2)}\|_\infty}\|_\infty}$$

$$= \frac{A^2\vec{x}^{(k-2)}}{\|A^2\vec{x}^{(k-2)}\|_\infty} \\ = \frac{A^2\vec{x}^{(k-2)}}{\|A^2\vec{x}^{(k-2)}\|_\infty}.$$

$$\Rightarrow \vec{x}^{(k)} = \dots = \frac{A^k \vec{x}^{(0)}}{\|A^k \vec{x}^{(0)}\|_\infty}.$$

$$= \lambda_1^k \left[ \alpha_1 \vec{v}_1 + \alpha_2 \left( \frac{\lambda_2}{\lambda_1} \right)^k \vec{v}_2 + \dots + \alpha_n \left( \frac{\lambda_n}{\lambda_1} \right)^k \vec{v}_n \right] \\ \| \lambda_1^k \left[ \alpha_1 \vec{v}_1 + \alpha_2 \left( \frac{\lambda_2}{\lambda_1} \right)^k \vec{v}_2 + \dots + \alpha_n \left( \frac{\lambda_n}{\lambda_1} \right)^k \vec{v}_n \right] \|_\infty$$

Since  $\left| \frac{\lambda_2}{\lambda_1} \right| < 1 \Rightarrow \left| \frac{\lambda_2}{\lambda_1} \right|^k \rightarrow 0$  as  $k \rightarrow \infty$ .

$\Rightarrow \vec{x}^{(k)} \rightarrow \frac{\alpha_1 \vec{v}_1}{\|\alpha_1 \vec{v}_1\|_\infty} = \vec{v}_1$  eigenvector associated with  $\lambda_1$ .

$$\Rightarrow \frac{\vec{x}^{(k)*} A \vec{x}^{(k)}}{\vec{x}^{(k)*} \vec{x}^{(k)}} \rightarrow \frac{\vec{v}_1^* A \vec{v}_1}{\vec{v}_1^* \vec{v}_1} = \lambda_1.$$

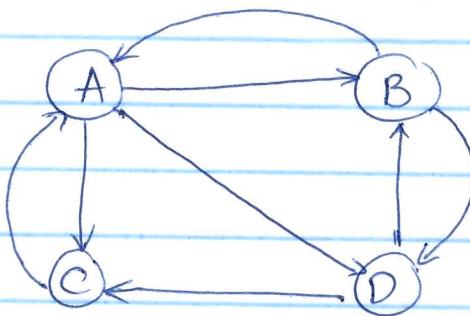
(44)

Remark: After finding the dominant eigenpair of A we can also use the method to find the next largest eigenvalue in magnitude (but I won't discuss it).

(45)

think of the Web as a directed graph  
 pages = nodes/vertices.  
 links = directed edges.

E.g.



Suppose a random surfer starts at a page A. There are links to B, C, and D, so this surfer will next be at each of those pages with probability  $\frac{1}{3}$  and has zero probability of being at A.

A random surfer at B has at next step probability  $\frac{1}{2}$  of being at A,  $\frac{1}{2}$  of being at D, and 0 of being at B or C.

In general, we can define a transition matrix of the Web to describe what happens to random surfers after one step. The element  $m_{ij}$  in row  $i$  and column  $j$  has value  $\frac{1}{k}$  if page  $j$  has  $k$  links out, and one of them is to page  $i$ . Otherwise,  $m_{ij} = 0$ .

E.g.

$$M = \begin{matrix} & \begin{matrix} A & B & C & D \end{matrix} \\ \begin{matrix} A \\ B \\ C \\ D \end{matrix} & \begin{bmatrix} 0 & \frac{1}{2} & 1 & 0 \\ \frac{1}{3} & 0 & 0 & \frac{1}{2} \\ \frac{1}{3} & 0 & 0 & \frac{1}{2} \\ \frac{1}{3} & \frac{1}{2} & 0 & 0 \end{bmatrix} \end{matrix}$$

Suppose we start at a random surfer at any of the  $n$  pages of the Web with equal probability.

(46)

Then the initial vector  $v^{(0)}$  will have  $1/n$  for each component. Then  $\Rightarrow v^{(1)} = Mv^{(0)}$  will tell us what is the probability that the surfer will visit <sup>a page</sup>, after his first visit. And  $v^{(2)} = Mv^{(1)}$  is the probability of the second visit.

. The principal eigenvalue of  $M$  is 1.

$\Rightarrow$  The principal eigenvector of  $M$  tell us where the surfer is most likely to be after a long time.

The intuition behind PageRank is that the more likely a surfer is to be at a page, the more important the page is.

E.g. For  $M$  as the previous example

$$\begin{bmatrix} 1/4 \\ 1/4 \\ 1/4 \\ 1/4 \end{bmatrix}, \begin{bmatrix} 9/24 \\ 5/24 \\ 5/24 \\ 5/24 \end{bmatrix}, \begin{bmatrix} 15/48 \\ 11/48 \\ 11/48 \\ 11/48 \end{bmatrix}, \dots, \begin{bmatrix} 3/9 \\ 2/9 \\ 2/9 \\ 2/9 \end{bmatrix}$$

(47)

The power iteration method in your textbook:

Start with an initial vector:  ~~$\alpha$~~   $\cdot q^{(0)}$

$$\text{Define: } q^{(k+1)} = \frac{Aq^{(k)}}{\lambda_{k+1}}$$

where  $\lambda_{k+1}$  is a scaling factor.

We will take  $\lambda_{k+1}$  to be the entry of  $Aq^{(k)}$  that is largest in absolute value.

$$\text{E.g. } A = \begin{bmatrix} 9 & 1 \\ 1 & 2 \end{bmatrix}$$

$$\text{Start with } q^{(0)} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

$$\cdot \text{ 1st iteration: } Aq^{(0)} = \begin{bmatrix} 10 \\ 3 \end{bmatrix} \Rightarrow \lambda_1 = 10.$$

$$q^{(1)} = \frac{Aq^{(0)}}{\lambda_1} = \begin{bmatrix} 1 \\ 0.3 \end{bmatrix}.$$

$$\cdot \text{ 2nd iteration: } Aq^{(1)} = \begin{bmatrix} 9 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ 0.3 \end{bmatrix} = \begin{bmatrix} 9.3 \\ 1 + 0.6 \end{bmatrix} = \begin{bmatrix} 9.3 \\ 1.6 \end{bmatrix}$$

$$\Rightarrow \lambda_2 = 9.3$$

$$\Rightarrow q^{(2)} = \frac{Aq^{(1)}}{\lambda_2} = \begin{bmatrix} 1 \\ 0.572034 \end{bmatrix}.$$

$j$	$\lambda_j$	$q^{(j)}$
8	9.140066	$[1, 0.140057]^T$
9	9.140057	$[1, 0.140055]^T$
10	9.140055	$[1, 0.140055]^T$
	$\lambda_1 = 9.14005$	$v_1 = \begin{bmatrix} 1 \\ 0.140055 \end{bmatrix}$

(48)

## Convergence Analysis of Iterative Methods.

Two questions regarding the Jacobi and Gauss-Seidel Method:

1. When will each of these methods work? That is, under what conditions will they produce a sequence of approximations  $x^{(0)}$ ,  $x^{(1)}$ ,  $x^{(2)}$ , ... that converges to a true solution  $x^*$ ?
2. When they do work, how quickly will the approximations approach the true solution? That is, what will the rate of convergence be?

Def: Given an  $n \times n$  nonsingular matrix  $A$ , a splitting of  $A$  is a decomposition  $A = M - N$ , where  $M$  is nonsingular. The matrix  $M$  is called the splitting matrix.

E.g.  $M = \begin{bmatrix} a_{11} & & & \\ & a_{22} & & \\ & & \ddots & \\ & & & a_{nn} \end{bmatrix}$  . the diagonal matrix whose entries are the same as those of  $A$ .  
 main diagonal

Goal: Solve  $Ax = b$  for  $x$ .

We have

$$(M - N)x = b$$

$$Mx = Nx + b.$$

$$\bar{M}^{-1} Mx = \bar{M}^{-1}(Nx + b).$$

$$x = \bar{M}^{-1}Nx + \bar{M}^{-1}b.$$

In general, an iterative method that finds the solution to  $Ax = b$  takes the form:

$$Mx^{(k+1)} = Nx^{(k)} + b$$

or  $x^{(k+1)} = \bar{M}^{-1}Nx^{(k)} + \bar{M}^{-1}b$

(49)

E.g. Jacobi:  $M = D$  and  $N = D - A$ .  
 (Exercise 8.2.4).

Gauss-Seidel:

$M =$  lower triangular part of  $A$ .  
 $-N =$  strictly upper triangular part of  $A$ .  
 (Exercise 8.2.12)

Ideally, we want  $M \approx A$  and "easy to solve".  
 and  $N \approx 0$ .

Suppose  $x^*$  is the true solution to  $Ax = b$ .

Let  $e^{(k)} = x^* - x^{(k)}$  be the difference between  
 the true solution and its approximant after  $k$  iterations.

Fact:  $x^{(k)} \xrightarrow[\text{as } k \rightarrow \infty]{} x^*$  if and only if  $e^{(k)} \xrightarrow[\text{as } k \rightarrow \infty]{} 0$ .

Observe that since  $x^*$  is the true solution to  $Ax = b$ ,  
 $Ax^* = b$ .

$$(M - N)x^* = b$$

$$Mx^* = Nx^* + b$$

$$\Rightarrow M(x^* - x^{(k+1)}) = N(x^* - x^{(k)}) + b - b$$

$$Me^{(k+1)} = Ne^{(k)}$$

$$e^{(k+1)} = \underbrace{M^{-1}Ne^{(k)}}_G$$

$$\Rightarrow e^{(k+1)} = Ge^{(k)} \quad \text{for } k=0, 1, \dots$$

$$(*) \rightarrow \boxed{e^{(k)} = G^k e^{(0)}}$$

Remark:  $e^{(0)}$  is our initial error. (its size depends on the initial guess)

(\*) shows that regardless of what the initial guess was,  
 $e^{(k)} \rightarrow 0$  if  $G^k \rightarrow 0$

$\Rightarrow$  the convergence theory of iterative methods resembles the analysis of the power method for computing a dominant eigenvector.

Suppose that  $G$  is semisimple (i.e. it has  $n$  lin. ind eigenvectors)  $v_1, \dots, v_n$ . and let  $\lambda_1, \dots, \lambda_n$  be corresponding eigenvalues, i.e.  $Gv_j = \lambda_j v_j$ ,  $1 \leq j \leq n$ .

We have

$$e^{(0)} = c_1 v_1 + c_2 v_2 + \dots + c_n v_n.$$

$$Ge^{(0)} = c_1 Gv_1 + c_2 Gv_2 + \dots + c_n Gv_n$$

$$e^{(1)} = c_1 \lambda_1 v_1 + c_2 \lambda_2 v_2 + \dots + c_n \lambda_n v_n.$$

$$\Rightarrow e^{(k)} = c_1 \lambda_1^k v_1 + c_2 \lambda_2^k v_2 + \dots + c_n \lambda_n^k v_n.$$

For some vector norm  $\|\cdot\|$ .

$$\|e^{(k)}\| \leq |c_1| |\lambda_1|^k \|v_1\| + |c_2| |\lambda_2|^k \|v_2\| + \dots + |c_n| |\lambda_n|^k \|v_n\|.$$

Fact:  $|\lambda_i|^k \rightarrow 0$  if and only if  $|\lambda_i| < 1$ .

So  $\|e^{(k)}\| \rightarrow 0$  for every initial guess  $x^{(0)}$   
 if and only if  $\max_i |\lambda_i| < 1$ .

51

Let  $\sigma(G)$  be the spectrum of  $G$ , i.e. the set of all eigenvalues of  $G$ .

The spectral radius  $\rho(G)$  is defined by.

$$\rho(G) = \max_{\lambda \in \sigma(G)} |\lambda|.$$

$\Rightarrow$  The iterations converge (for any starting vector) if and only if  $\rho(G) < 1$ .

\* How about the rate of convergence?

Def: A sequence  $(x^{(j)})$  that converges to  $x$  is said to converge linearly if there is a number  $r$  satisfying  $0 < r < 1$  such that

$$\lim_{j \rightarrow \infty} \frac{\|x^{(j+1)} - x\|}{\|x^{(j)} - x\|} = r.$$

(In other words, for large  $j$   $\|x^{(j+1)} - x\| \approx r \|x^{(j)} - x\|$ .)  
r is called the convergence ratio.

Suppose  $1 > |\lambda_1| > |\lambda_2| \geq |\lambda_3| \geq \dots \geq |\lambda_n| > 0$ .

Then  $|\lambda_1| = \rho(G)$

and  $e^{(k)} \approx c_1 \lambda_1^k v_1$ .

$$\Rightarrow \frac{\|e^{(k+1)}\|}{\|e^{(k)}\|} \approx |\lambda_1| = \rho(G) \quad \text{for large } k,$$

$\Rightarrow$  The convergence is linear with convergence ratio  $\rho(G)$ .  
The smaller  $\rho(G)$  is, the faster the iterations converge.

(52)

## 8.4 Steepest Descent. Methods

$A$  is symmetric and positive definite.

The problem of solving  $Ax = b$  can be reformulated as a minimization problem.

$\Rightarrow$  proceed to study methods for solving the minimization problem.

Let  $\mathcal{F}: \mathbb{R}^n \rightarrow \mathbb{R}$  be defined by

$$(*) \quad f(y) = \frac{1}{2} y^T A y - y^T b.$$

Thm: Let  $A \in \mathbb{R}^{n \times n}$  p.d., let  $b \in \mathbb{R}^n$ , and define  $\mathcal{F}$  as in (\*). Then there is exactly one  $x \in \mathbb{R}^n$  for which

$$\mathcal{F}(x) = \min_y \mathcal{F}(y)$$

and this  $x$  is the solution of  $Ax = b$ .

To minimize  $\mathcal{F}(y)$ ,

Pf: Let  $\hat{x}$  be the solution of  $Ax = b$ , that is  $A\hat{x} = b$ . Then

$$\begin{aligned} f(y) &= \frac{1}{2} y^T A y - y^T b = \frac{1}{2} y^T A y - y^T A \hat{x} \\ &= \frac{1}{2} y^T A y - y^T A \hat{x} + \frac{1}{2} \hat{x}^T A \hat{x} - \frac{1}{2} \hat{x}^T A \hat{x} \\ &= \frac{1}{2} (y - \hat{x})^T A (y - \hat{x}) - \frac{1}{2} \hat{x}^T A \hat{x}. \end{aligned}$$

Since  $\hat{x}^T A \hat{x}$  is independent of  $y$ , this is minimized when  $y = \hat{x}$ . ( $(y - \hat{x})^T A (y - \hat{x}) > 0$  if  $y - \hat{x} \neq 0$ )  
 $= 0$  if  $y - \hat{x} = 0$ .)

(53)

## \* Descent Methods:

The basic idea behind a descent method:

at  $k$ th iteration: we have  $x^{(k)}$  an approximation to  $x^*$ .

at  $(k+1)$ th iteration: find a better approximation  $x^{(k+1)}$ .

That is we find  $x^{(k+1)}$  s.t.  $f(x^{(k)}) \geq f(x^{(1)}) \geq \dots \geq f(x^{(k)}) \geq f(x^{(k+1)})$

$\Rightarrow$  descent method:

$\Rightarrow$  pick a search direction  $p^{(k)}$  and let

$$\vec{x}^{(k+1)} = \vec{x}^{(k)} + \alpha_k \vec{p}^{(k)}$$

scalar

In other words, search for a minimum along a line defined by the current iterate,  $x^{(k)}$ , and the search direction  $p^{(k)}$ , and pick  $\alpha_k$  so that  $f(x^{(k+1)}) \leq f(x^{(k)})$ .

Find  $x^{(k+1)}$  s.t.  $f(x^{(k+1)}) = f(x^{(k)}) + \alpha_k p^{(k)}$ .

$$f(x^{(k+1)}) = \min_{\alpha_k} f(x^{(k)} + \alpha_k p^{(k)})$$

$$= \frac{1}{2} (x^{(k)} + \alpha_k p^{(k)})^T A (x^{(k)} + \alpha_k p^{(k)}) - (x^{(k)} + \alpha_k p^{(k)})^T b$$

$$= \left( \frac{1}{2} x^{(k)T} A x^{(k)} \right) + \alpha_k p^{(k)T} A x^{(k)} + \frac{1}{2} \alpha_k^2 p^{(k)T} A p^{(k)}$$

$$= f(x^{(k)}) - x^{(k)T} b - \alpha_k p^{(k)T} b$$

$$= f(x^{(k)}) + \frac{1}{2} \alpha_k^2 p^{(k)T} A p^{(k)} + \alpha_k p^{(k)T} (A x^{(k)} - b)$$

$$= f(x^{(k)}) + \frac{1}{2} \alpha_k^2 p^{(k)T} A p^{(k)} - p^{(k)T} r^{(k)} \alpha_k$$

where  $r^{(k)} = b - A x^{(k)}$ , the residual.

This  $\Leftrightarrow$  is a quadratic equation in  $\alpha_k$  since this is the only free variable).

$$0 = \frac{df(x^{(k)} + \alpha_k p^{(k)})}{d\alpha_k} = \alpha_k p^{(k)T} A p^{(k)} - p^{(k)T} r^{(k)}$$

$$\Rightarrow \alpha_k = \frac{p^{(k)T} r^{(k)}}{p^{(k)T} A p^{(k)}}$$

$$\text{and } x^{(k+1)} = x^{(k)} + \alpha_k p^{(k)}$$

$\Rightarrow$  How to pick the search direction  $p^{(k)}$ ?

\* Steepest Descent:

For a function  $f: \mathbb{R}^n \rightarrow \mathbb{R}$  that we are trying to minimize, for a given  $x$ , the direction in which the function most rapidly increases in value at  $x$  is given by the gradient,  $\nabla f(x)$ . Thus, the direction in which the function decreases rapidly is  $-\nabla f(x)$ .

For our function

$$f(x) = \frac{1}{2} x^T A x - x^T b,$$

$$\nabla f(x) = Ax - b \quad (\text{check!}).$$

$$\Rightarrow -\nabla f(x) = -Ax + b = b - Ax.$$

$$\text{Recall } r^{(k)} = b - Ax^{(k)} = -\nabla f(x^{(k)}).$$

pick the search direction  $p^{(k)} = r^{(k)} = b - Ax^{(k)}$ .

$$\Rightarrow x^{(k+1)} = x^{(k)} + \alpha_k p^{(k)} = x^{(k)} - \alpha_k \nabla f(x^{(k)}),$$

where  $\alpha_k = \frac{p^{(k)T} r^{(k)}}{p^{(k)T} A p^{(k)}}.$  | Note that:  
 $r^{(k+1)} = b - Ax^{(k+1)} = \underbrace{b - Ax^{(k)}}_{r^{(k)}} - \alpha_k \underbrace{Ap^{(k)}}_{q^{(k)}} = \underbrace{r^{(k)}}_{- \alpha_k q^{(k)}}$

\* Pseudo code:

Given  $A, b, x^{(0)}$ .

$$r^{(0)} = b - Ax^{(0)}, \quad p^{(0)} = r^{(0)}.$$

for  $k = 0, 1, \dots$

$$q^{(k)} = Ap^{(k)}$$

$$\alpha_k = \frac{p^{(k)T} r^{(k)}}{p^{(k)T} q^{(k)}}.$$

$$x^{(k+1)} = x^{(k)} + \alpha_k p^{(k)}$$

$$r^{(k+1)} = r^{(k)} - \alpha_k q^{(k)}$$

$$p^{(k+1)} = r^{(k+1)}$$

end for.