- Test out using the power method to compute $\|A\|_2$.

- Trying to maximize $\|Ax\|_2$ subject to the condition $\|x\|_2 = 1$.

$$\sqrt{Ax\cdot Ax} = \sqrt{x^TA^TAx} = \sqrt{x^TBx}$$

From last time... (the notes posted online from last week are now corrected)

Wrong

$$\cancel{Ax = Ac_1x_1 + Ac_2x_2 + \cdots + Ac_nx_n}$$
$$\cancel{= c_1\lambda_1x_1 + c_2\lambda_2x_2 + \cdots + c_n\lambda_nx_n}$$

$$\|Ax\|^2 = (c_1^2\lambda_1)^{\cancel{2}} + (c_2^2\lambda_2)^{\cancel{2}} + \cdots + (c_n^2\lambda_n)^{\cancel{2}}$$

$Bx_1 = \lambda_1x_1$, $Bx_2 = \lambda_2x_2$, ..., $Bx_n = \lambda_nx_n$

where the x_i's form an orthonormal basis of $\mathbb{R}^n$.

Given $x \in \mathbb{R}^n$ there is c_i's such that

$$x = c_1x_1 + c_2x_2 + \cdots + c_nx_n$$

Then

$$\|Ax\|_2^2 = x^TBx$$

$$= (c_1x_1 + c_2x_2 + \cdots + c_nx_n)^T B (c_1x_1 + c_2x_2 + \cdots + c_nx_n)$$

$$= (c_1x_1 + c_2x_2 + \cdots + c_nx_n)^T (c_1\lambda_1x_1 + c_2\lambda_2x_2 + \cdots + c_n\lambda_nx_n)$$

Since B is symmetric, we already chose the x_i's to be an orthonormal basis of eigenvectors.

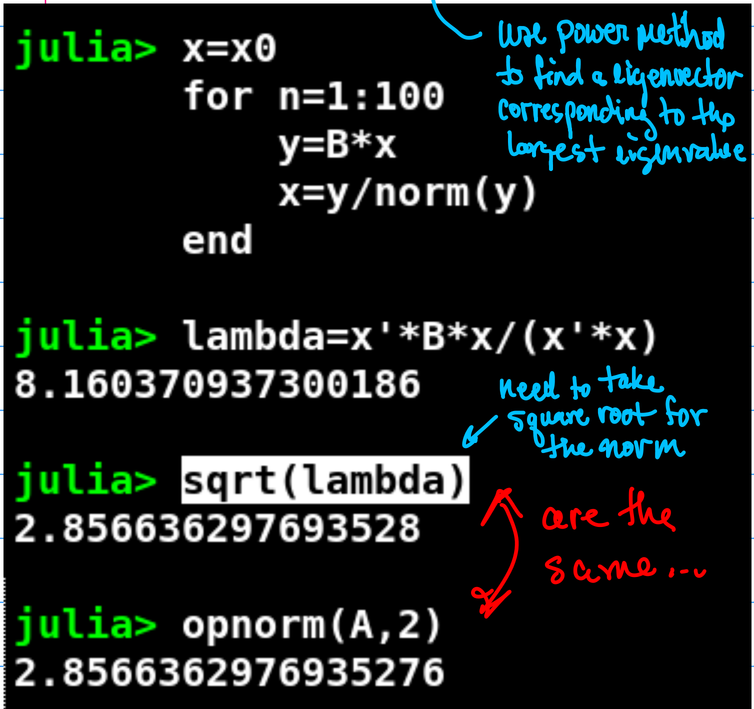
$$= c_1^2\lambda_1 + c_2^2\lambda_2 + \cdots + c_n^2\lambda_n = \sum_{k=1}^{n} c_k^2\lambda_k$$

$$\|A\|_2 = \max\{\|Ax\|_2 : \|x\|_2 = 1\}$$

$$= \sqrt{\max\{\|Ax\|_2^2 : \|x\|_2 = 1\}}$$

$$= \sqrt{\max\left\{\sum_{k=1}^{n} c_k^2\lambda_k : \sum_{k=1}^{n} c_k^2 = 1\right\}}$$

$$= \sqrt{\max\{\lambda_k : k = 1, \ldots, n\}} = \max\{\sqrt{\lambda_k} : k = 1, \ldots, n\}$$



```
julia> x=x0
       for n=1:100
           y=B*x
           x=y/norm(y)
       end

julia> lambda=x'*B*x/(x'*x)
8.160370937300186

julia> sqrt(lambda)
2.856636297693528

julia> opnorm(A,2)
2.8566362976935276
```

use power method to find a eigenvector corresponding to the largest eigenvalue

need to take square root for the norm

are the same...

The singular value decomposition is based on the $B = A^T A$ idea...

- Given any matrix $A \in \mathbb{R}^n$ then $B = A^T A$ is symmetric and positive semidefinite.

This is a way to use the spectral theorem with any matrix

- Then there is an eigenbasis of orthonormal vectors of B.

Spectral theorem

$$B x_1 = \lambda_1 x_1 , \ B x_2 = \lambda_2 x_2 , \ \ldots , \ B x_n = \lambda_n x_n$$

eigenvectors

$$x_i \cdot x_j = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

orthonormal

- The thing that can be done with this idea is called the singular value decomposition of A.

$$U = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} \in \mathbb{R}^{n \times n}$$

orthogonal matrix since orthonormal columns and U is square means that

in other words

$U^T U = I$ and since square then $U^{-1} = U^T$.

Note $$BU = B\left[\begin{array}{c|c|c|c} x_1 & x_2 & \cdots & x_n \end{array}\right]$$

definition of matrix multiplication

the fact that the x_i's are eigenvectors

$$= \left[\begin{array}{c|c|c|c} Bx_1 & Bx_2 & \cdots & Bx_n \end{array}\right] = \left[\begin{array}{c|c|c|c} \lambda_1 x_1 & \lambda_2 x_2 & \cdots & \lambda_n x_n \end{array}\right]$$

$$= \left[\begin{array}{c|c|c|c} x_1 & x_2 & \cdots & x_n \end{array}\right] \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix} \leftarrow D$$

multiply on the right by a diagonal matrix rescales the columns...

Therefore $$BU = UD$$

But we're really interested in the original matrix A

$$\underbrace{A^T A}\, U = UD$$

$$\underline{AU} = \left[\begin{array}{c|c|c|c} Ax_1 & Ax_2 & \cdots & Ax_n \end{array}\right] \leftarrow$$

The columns of AU are orthogonal...

(but not unit vectors)

$$Ax_i \cdot Ax_j = x_i^T A^T A x_j = x_i^T B x_j = x_i^T \lambda_j x_j$$

$$= \lambda_j \underbrace{x_i^T \cdot x_j}_{\text{ortho-normal}} = \begin{cases} \lambda_j & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

Since $\|Ax_i\|_2 = \sqrt{Ax_i \cdot Ax_i} = \sqrt{\lambda_i}$

we can normalize these columns by dividing by $\sqrt{\lambda_i}$

$$V = \left[\frac{Ax_1}{\sqrt{\lambda_1}} \;\middle|\; \frac{Ax_2}{\sqrt{\lambda_2}} \;\middle|\; \cdots \;\middle|\; \frac{Ax_n}{\sqrt{\lambda_n}} \right] \leftarrow \text{orthonormal columns} \in \mathbb{R}^{n \times n}$$

$V^T V = I$ and $V^T = V^{-1}$ i.e. V is an orthogonal matrix ...

$$AU = V \begin{bmatrix} \sqrt{\lambda_1} & 0 & \cdots & 0 \\ 0 & \sqrt{\lambda_2} & & 0 \\ & & \ddots & \vdots \\ 0 & 0 & \cdots & \sqrt{\lambda_n} \end{bmatrix} \leftarrow \Sigma$$

diagonal matrix

$$\underset{\substack{\uparrow \\ \text{orthogonal matrix}}}{AU} = \underset{\substack{\uparrow \\ \text{orthogonal matrix}}}{V}\Sigma \leftarrow \text{diagonal matrix}$$

$$A = V \cdot \Sigma\, U^T \leftarrow$$ This factorization is called the singular value decomposition.

This factorization is important because it works for every matrix and because orthogonal and diagonal matrices have a natural geometric interpretation

- We think of V and U being things like reflections and rotations which don't change the size of vectors or the angles between them

- Then the diagonal matrix Σ contains all the scaling information in A.

Separating out the scaling information is important in many data analysis and machine learning applications.

For example, writing Σ as the sum

$$\Sigma = \begin{bmatrix} \sqrt{\lambda_1} & & & \\ & 0 & & \\ & & \ddots & \\ & & & 0 \end{bmatrix} + \begin{bmatrix} 0 & & & \\ & \sqrt{\lambda_2} & & \\ & & 0 & \\ & & & \ddots \\ & & & & 0 \end{bmatrix} + \cdots + \begin{bmatrix} 0 & & & \\ & \ddots & & \\ & & \ddots & \\ & & & \sqrt{\lambda_n} \end{bmatrix}$$

separates the scaling operations into pieces. One can then consider the parts where the λ_i's are largest to be the most important contributions to the matrix A.

Please read Chapter 7.2 on applications of the singular value decomposition in our textbook for next time. Note, there is no class on Thursday due to the holiday (Veterans Day) so try to read all of 7.2 in preparation for Nov 16 next week.