Last time we found an eigenvector by iteratively taking powers of A. That was actually

section 6.3.1

Today we'll back up to the beginning of the chapter since the eigenvalue–eigenvector problem is sometimes rushed in Math 340. Also the book has some results which were not part of the usual linear algebra course. Before that, let's look at what we did before in the book and the algorithms that come after that...

This is what we did in class last time (page 118)

```
function NORMALIZED-ITERATION(A)
  v ← ARBITRARY(n)
  for k ← 1, 2, 3, ...
    w ← A v
    v ← w/‖w‖
  return v
```

(b)

Once we found the eigenvector we just divided $A*x./x$ to verify it was an eigenvector and exhibit the eigenvalue...

(Julia element by element division)

If x has any element which is zero (or close to zero), then dividing won't work well for that entry...

```
function RAYLEIGH-QUOTIENT-ITERATION(A, σ)
  v ← ARBITRARY(n)
  for k ← 1, 2, 3, ...
    w ← (A − σ I_{n×n})^{-1} v
    v ← w/‖w‖
    σ ← (vᵀ A v)/‖v‖₂²
  return v
```

better idea is to find λ using a least squares approximation...

We will consider a better way to find eigenvalues using what's called shifted inverse iteration...

6.3.3 Shifting

Suppose λ_2 is the eigenvalue of A with second-largest ma
fastest when $|\lambda_2/\lambda_1|$ is small, since in this case the power
is nearly 1, it may take many iterations before a single
If the eigenvalues of A are $\lambda_1, \ldots, \lambda_n$ with correspo
the eigenvalues of $A - \sigma I_{n\times n}$ are $\lambda_1 - \sigma, \ldots, \lambda_n - \sigma$, sin

$$(A - \sigma I_{n\times n})\vec{x}_i = A\vec{x}_i - \sigma\vec{x}_i = \lambda_i\vec{x}_i -$$

With this idea in mind, one way to make power iterati
such that:

$$\left|\frac{\lambda_2 - \sigma}{\lambda_1 - \sigma}\right| < \left|\frac{\lambda_2}{\lambda_1}\right|.$$

Same thing coming up...

If you want to find all the eigenvectors an even better algorithm is based on QR factorization:

```
function QR-ITERATION(A ∈ R^{n×n})
  for k ← 1, 2, 3, ...
    Q, R ← QR-FACTORIZE(A)
    A ← RQ
  return diag(R)
```

That looks strange because the factors are in the wrong order...

This is a weird algorithm because it repeatedly factors $A = QR$ and then multiplies the factors back together in the wrong order as RQ

I find it absolutely amazing that this iterative scheme can be used to find eigenvalues

Now, let's turn back to the beginning of the chapter and look at some applications and theory before working on any more computation... (also the lab computers are broken today)

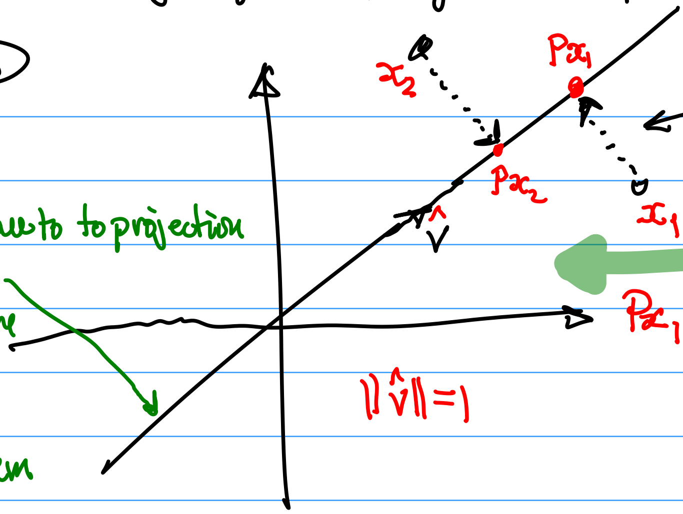
On the first page of the chapter

When A is symmetric, the eigenvectors of A are the critical points of $\vec{x}^\top A\vec{x}$ under the constraint $\|\vec{x}\|_2 = 1$.

what does this mean and where did it come from? (Lagrange multipliers)

Examples of Eigenvalue/Eigenvector problems:

①



$\leftarrow$ minimize the distances. by choosing a good vector v.

Note that, due to projection onto the line this is not the same as the usual least squares problem

assume $\hat{v} \in \mathbb{R}^m$ unit vector

and $x_i \in \mathbb{R}^m$ so this picture takes place in $\mathbb{R}^m$, even though it only looks two-dimensional in the notes.

$$\begin{aligned} Px_1 &= (x_1 \cdot \hat{v})\hat{v} \\ &= (x_1^T \hat{v})\hat{v} \\ &= (\hat{v}^T x_1)\hat{v} \\ &= \boxed{\underbrace{\hat{v} v^T}_{\text{projection matrix.}} x_1} \end{aligned}$$

$$J = \sum_{i=1}^{n} \| x_i - (x_i \cdot v) v \|^2$$

Question: minimize J over all vectors v such that $\|v\| = 1$.

$\leftarrow$ constrained optimization problem...

Lagrange multipliers...

write the constrain in the form "something" $= 0$

$$\|v\|^2 - 1 = 0$$

$$\mathcal{L} = J + \lambda(\|v\|^2 - 1)$$ $\leftarrow$ differentiate this and set equal to zero

$$\mathcal{L} = \sum_{i=1}^{n} \| x_i - (x_i \cdot v) v \|^2 + \lambda(\|v\|^2 - 1)$$

$$= \sum_{i=1}^{n} \| x_i - \underbrace{v v^T}_{} x_i \|^2 + \lambda(\|v\|^2 - 1)$$

We worked with this in class, but it's easier to keep things as the dot products, so I have simplified the notes (see green at top of next page).

work on one term out of the sum

$V \cdot V = 1$

$$\| x_i - (x_i \cdot v) v \|^2 = (x_i - (x_i \cdot v) v) \cdot (x_i - (x_i \cdot v) v)$$

$$= x_i \cdot x_i - (x_i \cdot \underbrace{(x_i \cdot v)}_{v \cdot x_i} v) - (x_i \cdot v) v \cdot x_i + (x_i \cdot v) \underbrace{v \cdot (x_i \cdot v) v}$$

$$= \|x_i\|^2 - 2(x_i \cdot v)^2 + (x_i \cdot v)^2 = \|x_i\|^2 - (x_i \cdot v)^2$$

we want $\nabla_v \mathcal{L}$ the gradient of $\mathcal{L}$ with respect to v so compute partial derivatives

$$\frac{\partial}{\partial v_k} \left(\|x_i\|^2 - (x_i \cdot v)^2 \right)$$

$$\frac{\partial}{\partial v_k} \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} \leftarrow k\text{th} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \\ 0 \end{bmatrix} = e_k$$

$$= -2 (x_i \cdot v) \frac{\partial}{\partial v_k} (x_i \cdot v)$$

$$= -2 (x_i \cdot v)\, x_i \cdot \frac{\partial v}{\partial v_k}$$

$$= -2 (x_i \cdot v) \boxed{x_i \cdot e_k}$$

$$= -2\, e_k^T \boxed{x_i x_i^T} v$$

index k from the partial derivates

what is this? Matrix

$x_i x_i^T \in \mathbb{R}^{m \times m}$

$m \times 1 \quad 1 \times m$

identity a matrix

Gradient

$$\nabla_v \left(\|x_i\|^2 - (x_i \cdot v)^2 \right) = \begin{bmatrix} \frac{\partial}{\partial v_1} (\text{---}\prime\prime\text{---}) \\ \vdots \\ \frac{\partial}{\partial v_m} (\text{---}\prime\prime\text{---}) \end{bmatrix}$$

The identity can be written as

$$I = \begin{bmatrix} e_1^T \\ e_2^T \\ \vdots \\ e_m^T \end{bmatrix}$$

rows of the identity matrix

$$= \begin{bmatrix} -2 e_1^T (x_i x_i^T) v \\ \vdots \\ -2 e_m^T (x_i x_i^T) v \end{bmatrix} = -2 I \underbrace{(x_i x_i^T) v}$$

factor this part out

Therefore

$$\nabla_v \| x_i - (x_i \cdot v) v \|^2 = -2 (x_i x_i^T) v .$$

$$\mathcal{L} = \sum_{i=1}^{n} \| x_i - (x_i \cdot v) v \|^2 + \lambda \left(\|v\|^2 - 1 \right)$$

still need to take the gradient of this term

$$\nabla_v \|v\|^2 = \begin{bmatrix} \frac{\partial}{\partial v_1} \|v\|^2 \\ \vdots \\ \frac{\partial}{\partial v_m} \|v\|^2 \end{bmatrix} = \begin{bmatrix} \frac{\partial}{\partial v_1} (v \cdot v) \\ \vdots \\ \frac{\partial}{\partial v_m} (v \cdot v) \end{bmatrix}$$

$$= \begin{bmatrix} e_1 \cdot v + v \cdot e_1 \\ \vdots \\ e_m \cdot v + v \cdot e_m \end{bmatrix} = 2 \begin{bmatrix} e_1^T v \\ \vdots \\ e_m^T v \end{bmatrix} = 2 I v = 2v$$

again the identity matrix

factor the v out (and the 2.

minimize

$$\nabla \mathcal{L} = \boxed{\sum_{i=1}^{n} -2 (x_i x_i^T) v} + \lambda 2v$$

add up the gradients of all the terms in the sum

$$= -2 \underbrace{\left(\sum_{i=1}^{n} x_i x_i^T \right)}_{A} v + 2\lambda v$$

factor out the 2 on the right

factor out the v on the left

$$A = \sum_{i=1}^{n} x_i x_i^T$$

Let A be the sum of all the matrices from each of the terms in the gradient

$$= -2Av + 2\lambda v = 0$$

or

$$\boxed{Av = \lambda v}$$

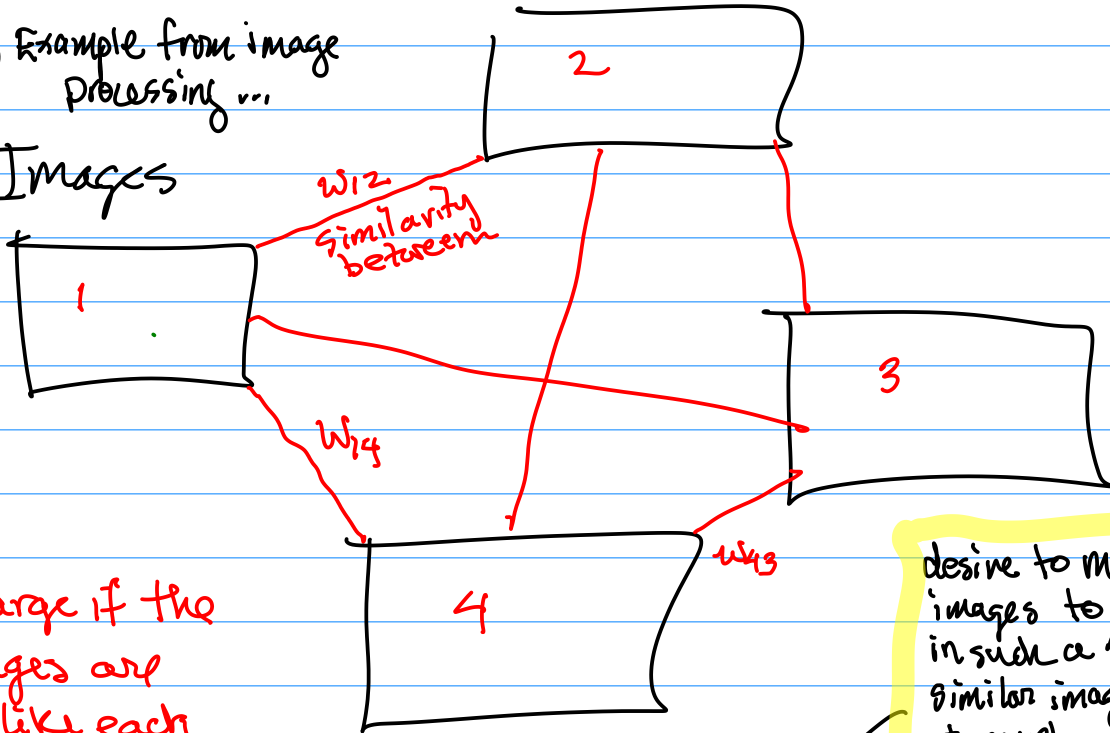
eigenvalue/eigenvector problem...

I found this example interesting, because usually one thinks of solving the normal equations $A^TAx = A^Tx$ when solving a least squares problem and that's either Gaussian elimination of by a QR factorization. In this case there is a projection onto the vector v and that leads to an eigenvalue-eigenvector problem...

② The second example is from differential equations, and this was discussed in Math285 ODE which many of you have already taken or will.

③ Example from image processing...

Images



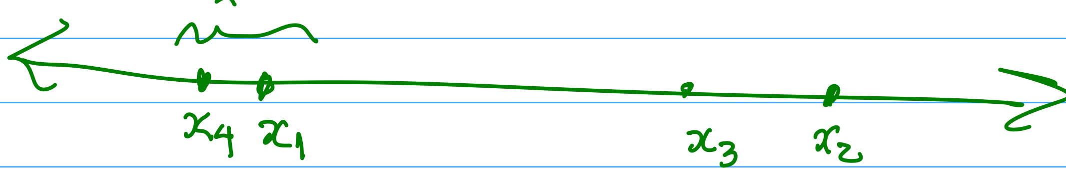
$w_{ij} > 0$

w_{ij} is large if the the images are more like each other and small if very different

desire to map the images to a line in such a way that similar images correspond to nearby points on the line.

assign a value in $\mathbb{R}$ so that

means image 1 is quite similar to image 4



Since images aren't simalar to each other on a linear scale, so approximation needs to be made in assigning the values x_k and that can be optimized...

$$E = \sum_{i,j} w_{ij} (x_i - x_j)^2$$

minimize E....

Take all $x_i =$ const...

constraints: $\|x\|^2 = 1$ $\quad \Sigma x_i = 0$

That's enough examples. Now let's look at the next section in the book. This is a review of the theory in Math 330 which may or may not have been taught in that class. Even if it was taught that was some time ago, so lets carefully look at what's in the chapter...

<u>Theorem</u> every $A \in \mathbb{R}^{n \times n}$ (or $\mathbb{C}^{n \times n}$) has at least one eigenvalue...

note if A is symmetric the eigenvalue can be used to reduce the dimension of the matrix in which case one can apply the same theorem again to get another eigenvalue and eigenvector

pick any $x \neq 0$ and $x \in \mathbb{R}^n$...

$$\underbrace{x \quad Ax \quad A^2x \quad \cdots \quad A^kx}$$

if k is large enough, then these vectors can't all be linearly independent...

Choose k to be the least such that the vectors are linearly dependent...

(For example $k=n$ gives $n+1$ vectors and those couldn't all be linearly independent

Thus

$$c_0x + c_1Ax + \cdots + c_kA^kx = 0$$

note $c_k \neq 0$ by choice of k

and not all the c_i's are zero...

$$p(t) = \frac{c_0}{c_k} + \frac{c_1}{c_k}t + \cdots + \frac{c_k}{c_k}t^k$$

for convenience divide out by c_k so that $p(t)$ is a monic polynomial.

Factor $p(t)$ by the Fundamental theorem of algebra...

$$p(t) = (t-\alpha_1)(t-\alpha_2)\cdots(t-\alpha_k)$$

for some $\alpha_k \in \mathbb{C}$

* the fundamental theorem of algebra only works for complex numbers.

Therefore

$$\underbrace{(A-\alpha_1 I)(A-\alpha_2 I)\cdots(A-\alpha_k I)}x = 0$$

since $x \neq 0$ then nullspace of this product is not empty... so one of the factors is not invertible and has a non-empty nullspace...

For example, suppose

$$(A - \alpha_j I) \text{ is not invertible...}$$

That means α_j is and eigenvalue and the eigenvector is any non-zero element of

$$\mathrm{Null}(A - \alpha_j I) \ldots$$

This shows every matrix $A \in \mathbb{R}^n$ has at least one eigenvalue and eigenvector. Unfortunately some matrices have one one. Symmetric matrices always have n eigenvectors though the eigenvalues may not be unique.