Hello and welcome to class!

Last time

We discussed the "least squares" problem, and then considered notions of distance and orthogonality in a general setting, as captured by inner products.

This time

We meet the spectral theorem, and the singular value decomposition. This is the last class on linear algebra. Next week, we begin differential equations!

Review: Inner products

Definition

An inner product on a vector space V is a map

$$\langle \cdot , \cdot \rangle : V \times V \to \mathbb{R}$$

which is distributive, commutative, and positive.

Review: Inner product properties

Distributivity (aka "bilinearity")

$$\langle a\mathbf{v} + b\mathbf{v}', c\mathbf{w} + d\mathbf{w}' \rangle = ac\langle \mathbf{v}, \mathbf{w} \rangle + bc\langle \mathbf{v}', \mathbf{w} \rangle + ad\langle \mathbf{v}, \mathbf{w}' \rangle + bd\langle \mathbf{v}', \mathbf{w}' \rangle$$

Commutativity

$$\langle \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{w}, \mathbf{v} \rangle$$

Positivity

$$\langle \mathbf{v}, \mathbf{v} \rangle \geq 0, \qquad \text{with equality only when } \mathbf{v} = 0$$

Review: Inner products on $\mathbb{R}^n$

Last time, we saw that distributivity implies that any inner product on $\mathbb{R}^n$ is given by a matrix:

$$\langle \mathbf{v}, \mathbf{w} \rangle = \mathbf{v}^T A \mathbf{w} \qquad\qquad A_{i,j} = \langle \mathbf{e}_i, \mathbf{e}_j \rangle$$

When does a matrix define an inner product?

For any matrix A, consider the formula $\langle \mathbf{v}, \mathbf{w} \rangle = \mathbf{v}^T A \mathbf{w}$.

This always satisfies the distributivity axiom.

It satisfies commutativity iff A is symmetric, i.e. $A^T = A$.

When does it satisfy positivity?
(last time: when A is 2x2, positive iff the eigenvalues are)

Eigenvectors of symmetric matrices

Eigenvectors with different eigenvalues are linearly independent.

For a symmetric matrix, eigenvectors with different eigenvalues are in fact orthogonal:

Indeed, since $M^T = M$, for any vector $\mathbf{v}$, we have $\mathbf{v}^T M = (M\mathbf{v})^T$.
If $\mathbf{v}_i, \mathbf{v}_j$ are eigenvectors with eigenvalues λ_i, λ_j:

$$\mathbf{v}_i^T M \mathbf{v}_j = \lambda_j \mathbf{v}_i^T \mathbf{v}_j$$

$$\mathbf{v}_i^T M \mathbf{v}_j = (M\mathbf{v}_i)^T \mathbf{v}_j = \lambda_i \mathbf{v}_i^T \mathbf{v}_j$$

So if $\lambda_i \neq \lambda_j$, then $\mathbf{v}_i^T \mathbf{v}_j = 0$.

The spectral theorem

Theorem

A symmetric matrix M has all real eigenvalues and an orthonormal basis of eigenvectors.

Proof. Given any real basis of eigenvectors, those of different eigenvalues are already orthogonal; and if some eigenspace is of dimension > 1, we may use Gram-Schmidt to replace our basis with an orthonormal one. So it is enough to find a basis of real eigenvectors. Our first task is to produce a single such vector.

The spectral theorem

Certainly M has a complex eigenvector $\mathbf{v}$ with complex eigenvalue λ. Let $\dagger$ denote the operation of taking transpose and complex conjugate.

Then

$$\lambda \mathbf{v}^\dagger \mathbf{v} = \mathbf{v}^\dagger M \mathbf{v} = (M\mathbf{v})^\dagger \mathbf{v} = \overline{\lambda} \mathbf{v}^\dagger \mathbf{v}$$

Since $\mathbf{v}^\dagger \mathbf{v}$ is the sum of the squares of the lengths of the entries of $\mathbf{v}$, it is positive, hence nonzero, hence $\lambda = \overline{\lambda}$ is a real number.

The spectral theorem

Now we know there is a real unit eigenvector $\mathbf{v}$ with the real eigenvalue λ. Consider the orthogonal complement $\mathbf{v}^\perp$. This is preserved by M: if $\mathbf{w}^T\mathbf{v} = 0$, then

$$(M\mathbf{w})^T\mathbf{v} = \mathbf{w}^T M\mathbf{v} = \lambda\mathbf{w}^T\mathbf{v} = 0$$

Pick an orthonormal basis $\mathbf{w}_2, \ldots, \mathbf{w}_n$ of $\mathbf{v}^\perp$. In the basis $\mathbf{v}, \mathbf{w}_2, \mathbf{w}_3, \ldots, \mathbf{w}_n$, the matrix M takes the shape

$$M = \begin{bmatrix} \mathbf{v}^T M\mathbf{v} & 0 & \cdots & 0 \\ 0 & \mathbf{w}_2^T M\mathbf{w}_2 & \cdots & \mathbf{w}_2^T M\mathbf{w}_n \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \mathbf{w}_n^T M\mathbf{w}_2 & \cdots & \mathbf{w}_n^T M\mathbf{w}_n \end{bmatrix}$$

The lower-right block is a smaller symmetric matrix; we are done by induction.

The spectral theorem

Theorem

A symmetric matrix M has all real eigenvalues and an orthonormal basis of eigenvectors.

The spectral theorem: reformulation

Recall that a square matrix O is said to be orthogonal if $O^T = O^{-1}$; equivalently, if the columns are orthonormal; equivalently, if the rows are orthonormal.

Theorem

If M is symmetric, there is an orthogonal matrix O and a real diagonal matrix D with

$$M = ODO^{-1} = ODO^T$$

Proof: the columns of O are orthonormal eigenvectors of M.

The spectral theorem: reformulation

Theorem

If M is symmetric, there is an orthonormal basis $\mathbf{v}_i$ and real numbers λ_i such that

$$M = \lambda_1 \mathbf{v}_1 \mathbf{v}_1^T + \cdots + \lambda_n \mathbf{v}_n \mathbf{v}_n^T$$

Proof: the $\mathbf{v}_i$ are the eigenvectors of M and the λ_i are their eigenvalues.

Note by orthonormality, $(\mathbf{v}_i \mathbf{v}_i^T)^2 = \mathbf{v}_i \mathbf{v}_i^T$, and $(\mathbf{v}_i \mathbf{v}_i^T)(\mathbf{v}_j \mathbf{v}_j^T) = 0$ when $i \neq j$.

What does positive mean?

Suppose M is a symmetric matrix. When is it true that:

$$\mathbf{w}^T M \mathbf{w} \geq 0, \qquad \text{with equality only when } \mathbf{w} = 0$$

Claim: if and only if M has only positive eigenvalues.

Take an orthonormal basis $\mathbf{v}_i$ of eigenvectors for M, with eigenvalues λ_i. Note $\mathbf{v}_i^T M \mathbf{v}_i = \lambda_i \mathbf{v}_i^T \mathbf{v}_i = \lambda_i$. So if any of the λ_i is ≤ 0, then certainly M is not positive.

On the other hand, expand any $\mathbf{w} = (\sum w_i \mathbf{v}_i)$. Then

$$\mathbf{w}^T M \mathbf{w} = (\sum w_i \mathbf{v}_i)^T M (\sum w_i \mathbf{v}_i) = \sum w_i^2 \lambda_i$$

This is certainly positive if the λ_i are positive and $\mathbf{w} \neq 0$.

Square-roots of positive matrices

Positive numbers have square-roots.

So do positive symmetric matrices: if $A = ODO^{-1}$ with D having all non-negative entries, then we can write $\sqrt{D}$ for the diagonal matrix whose entries are the square-roots of D's entries, and:

$$\sqrt{A} := O\sqrt{D}O^{-1}$$

Note that $\sqrt{A}$ is again symmetric and positive.

Stretching and shrinking

How big or small can be $\frac{|A\mathbf{w}|}{|\mathbf{w}|}$?

If A is symmetric, then we take a basis of orthonormal eigenvectors $\mathbf{v}_i$ with eigenvalues λ_i and write:

$$\mathbf{w} = \sum w_i \mathbf{v}_i$$

$$A\mathbf{w} = \sum w_i \lambda_i \mathbf{v}_i$$

From this it is not hard to see

$$\min(|\lambda_i|) \leq \frac{|A\mathbf{w}|}{|\mathbf{w}|} = \sqrt{\frac{\sum w_i^2 \lambda_i^2}{\sum w_i^2}} \leq \max(|\lambda_i|)$$

Stretching and shrinking

What about for matrices A in general? Maybe not even square?

$$|A\mathbf{v}|^2 = (A\mathbf{v})^T(A\mathbf{v}) = \mathbf{v}^T A^T A\mathbf{v}$$

Note the matrix $A^T A$ is (square and) symmetric. It's also non-negative: $\mathbf{v}^T A^T A\mathbf{v} = |A\mathbf{v}|^2 \geq 0$.

So it has a symmetric non-negative square-root $B = \sqrt{A^T A}$. So we reduced the problem to the symmetric case, since

$$|A\mathbf{v}|^2 = \mathbf{v}^T A^T A\mathbf{v} = \mathbf{v}^T B^2\mathbf{v} = \mathbf{v}^T B^T B\mathbf{v} = |B\mathbf{v}|^2$$

Singular values

Let $\mathbf{v}_i$ be an orthonormal basis of eigenvectors for $A^T A$, with eigenvalues λ_i. We order them so that $\lambda_1 \geq \lambda_2 \geq \cdots$.

As we have seen, all $\lambda_i \geq 0$. We write $\sigma_i = \sqrt{\lambda_i}$. The σ_i are called the singular values of A. Note:

$$(A\mathbf{v}_i) \cdot (A\mathbf{v}_j) = \mathbf{v}_j^T A^T A\mathbf{v}_i = 0 \qquad\qquad i \neq j$$

$$(A\mathbf{v}_i) \cdot (A\mathbf{v}_i) = \mathbf{v}_i^T A^T A\mathbf{v}_i = \sigma_i^2 \mathbf{v}_i^T \mathbf{v}_i = \sigma_i^2$$

Singular value decomposition

So the $\mathbf{v}_i$ are an orthonormal basis whose images are also orthogonal. We rescale the images to the orthonormal

$$\mathbf{u}_i := \frac{1}{\sigma_i} A\mathbf{v}_i$$

We extend the $\mathbf{u}_i$ to an orthonormal basis. The matrices U, V whose columns are the basis vectors $\mathbf{u}_i$, $\mathbf{v}_i$ are orthogonal: $U^T = U^{-1}$ and $V^T = V^{-1}$.

The matrix $\Sigma := U^T AV$ is diagonal, with entries σ_i.

$$A = U\Sigma V^T$$