# LECTURE APRIL 10: INNER PRODUCT SPACES

## 1. Review of the dot product in Euclidean space

Recall that for two vectors $v = (a, b)$ and $w = (c, d)$ in $\mathbf{R}^2$, the dot product is

$$v \cdot w = ac + bd$$

The result is a scalar!!!!! Sometimes this is called the scalar product, for this reason. In 3D there's a similar formula. There's another formula for dot product:

$$v \cdot w = |v|\,|w| \cos\theta$$

Here, $|v|$ means the length of the vector $v$, namely $|v| = \sqrt{a^2 + b^2}$, and $\theta$ means the angle between the two vectors. This formula is valid in any dimension whatsoever, even $\mathbf{R}^4$!

The length of a vector is already given by the dot product,

$$|v| = \sqrt{v \cdot v}$$

Thus both length and angle can be read off from the dot product. There is another related concept, *orthogonality*: two vectors $v$ and $w$ are orthogonal if

$$v \cdot w = 0.$$

This means that there is an angle of $\pi/2$ between $v$ and $w$.

## 2. Inner product spaces

Given an (abstract) vector space $V$, with scalar field the real numbers, there is no automatic concept of length, angle, or orthogonality.

An *inner product* space is a vector space $V$ together with a function

$$\langle v, w\rangle \colon V \times V \to \mathbf{R}.$$

It has to satisfy:

(1) $\langle v_1 + v_2, w\rangle = \langle v_1, w\rangle + \langle v_2, w\rangle$
(2) $\langle cv, w\rangle = c\langle v, w\rangle$.
(3) $\langle w, v\rangle = \langle v, w\rangle$.
(4) $\langle v, v\rangle \geqslant 0$, with equality if and only if $v = 0$.

These axioms are supposed to capture what the dot product does in Euclidean space. But they apply to many other vector spaces, even ones of infinite dimensions.

You can prove theorems starting from these axioms. For instance, why is $\langle \mathbf{0}, v\rangle = 0$? Here's a one line proof:

$$\langle \mathbf{0}, v\rangle = \langle 0\mathbf{0}, v\rangle = 0\langle \mathbf{0}, v\rangle = 0.$$

Given an inner product space, we can define a notion of *length* or *magnitude* of its vectors, like this:

$$\|v\| = \sqrt{\langle v, v\rangle}.$$

In the asbtract setting of vector spaces, $\|v\|$ is called the *norm* of $v$. We have $\|v\| > 0$ for all nonzero vectors $v$.

We can also say that $v$ and $w$ are orthogonal if

$$\langle v, w\rangle = 0$$

We can also try to define the angle between two nonzero vectors $v, w$ in an inner product space, using the formula

$$\langle v, w\rangle = \|v\| \, \|w\| \cos\theta$$
$$\theta = \cos^{-1}\left(\frac{\langle v, w\rangle}{\|v\| \, \|w\|}\right)$$

In order for this to make sense, we need the expression in parentheses to be between $-1$ and $1$.

**Theorem 2.1** (The Cauchy-Schwarz inequality). *In an inner product space, for all $v, w$ we have*

$$|\langle v, w\rangle| \leqslant \|v\| \, \|w\| \, .$$

## 3. An example: continuous functions on an interval

An example: Let $V$ be the vector space of all continuous functions $f\colon [0, 2\pi] \to \mathbf{R}$. Define an inner product on it by

$$\langle f, g\rangle = \int_0^{2\pi} f(x)g(x)\ dx$$

The last axiom would be verified by saying that

$$\int_0^{2\pi} f(x)^2\ dx \geqslant 0,$$

with equality if and only if $f$ is the zero function.

If $f(x) = \cos(x)$ and $g(x) = \sin(x)$, then $f, g \in V$. We can compute:

$$\langle f, g\rangle = \int_0^{2\pi} \cos(x)\sin(x)dx = \frac{1}{2}\sin(x)^2|_0^{2\pi} = 0$$

In fact any two functions of the form $\cos(mx)$ and $\sin(nx)$ are going to be orthogonal to each other, as long as they are different. This result is very important in Fourier analysis.

The norm of $f$ is

$$\|f\| = \sqrt{\langle f, f\rangle} = \sqrt{\int_0^{2\pi} \cos^2(x)dx} = \sqrt{\pi}$$

## 4. Pythagorean theorem and Cauchy-Schwartz inequality

The Pythagorean theorem says: if $a, b, c$ are the sides of a right triangle, with $c$ the hypoteneuse, then $a^2 + b^2 = c^2$.

In terms of the Euclidean plane $\mathbf{R}^2$, it says: if $v$ and $w$ are vectors which are orthogonal to each other, then

$$|v|^2 + |w|^2 = |v + w|^2 \, .$$

This formula works in $\mathbf{R}^3$ as well.

**Theorem 4.1.** *Let $V$ be an inner product space, and let $v, w \in V$ be orthogonal vectors. Then*

$$\|v\|^2 + \|w\|^2 = \|v + w\|^2 \, .$$

*Proof.* Let $v, w \in V$ be orthogonal vectors. Then $\langle v, w\rangle = 0$. We have

$$\|v + w\|^2 = \langle v + w, v + w\rangle = \langle v, v + w\rangle + \langle w, v + w\rangle = \langle v, v\rangle + \langle v, w\rangle + \langle w, v\rangle + \langle w, w\rangle$$

The inner two terms are 0, and so

$$\|v + w\|^2 = \|v\|^2 + \|w\|^2 \, .$$

$\square$

**Theorem 4.2** (Cauchy-Schwartz inequality)**.** *Let $v, w \in V$ be vectors in an inner product space. Then*

$$|\langle v, w\rangle| \leqslant \|v\| \, \|w\| \, ,$$

*with equality if and only if the two vectors are parallel.*

The theorem means that if

$$c = \frac{\langle v, w\rangle}{\|v\| \, \|w\|},$$

then $|c| \leqslant 1$. In other words, $-1 \leqslant c \leqslant 1$. Then it makes sense to define an angle $\theta$ by $\cos\theta = c$, and then,

$$\langle v, w\rangle = \|v\| \, \|w\| \cos\theta.$$

We consider $\theta$ to be the angle between $v$ and $w$. If $v$ and $w$ are parallel, then $\theta = 0$. If $v$ and $w$ are orthogonal, then $\theta = \pi/2$.

*Proof.* Let $v, w \in V$ be vectors. If $w = 0$, then the inequality is obvious. So assume that $w \neq 0$, which implies by axiom 4 that $\langle w, w\rangle \neq 0$. Let

$$z = v - \frac{\langle v, w\rangle}{\langle w, w\rangle} w$$

The inner product between $z$ and $w$ is

$$\langle z, w\rangle = \langle v - \frac{\langle v, w\rangle}{\langle w, w\rangle} w, w\rangle = \langle v, w\rangle - \langle \frac{\langle v, w\rangle}{\langle w, w\rangle} w, w\rangle = \langle v, w\rangle - \frac{\langle v, w\rangle}{\langle w, w\rangle}\langle w, w\rangle = 0.$$

So $z$ and $w$ are orthogonal. Then

$$\|v\|^2 = \left\| z + \frac{\langle v, w\rangle}{\langle w, w\rangle} w \right\|^2 .$$

By the Pythagorean theorem, this is

$$\|v\|^2 = \|z\|^2 + \left\| \frac{\langle v, w\rangle}{\langle w, w\rangle} w \right\|^2 \geqslant \left\| \frac{\langle v, w\rangle}{\langle w, w\rangle} w \right\|^2 = \left( \frac{\langle v, w\rangle}{\langle w, w\rangle} \right)^2 \|w\|^2 = \frac{\langle v, w\rangle^2}{\|w\|^2}$$

We get

$$\langle v, w\rangle^2 \leqslant \|v\|^2 \, \|w\|^2 \, .$$

Take the square root of both sides to get

$$|\langle v, w\rangle| \leqslant \|v\| \, \|w\| \, .$$

□

## 5. Complex inner product spaces

We can certainly talk about a vector space over the complex numbers **C**. We can also talk about an inner product on such a vector space.

An *complex inner product* space is a vector space $V$ together with a function

$$\langle v, w\rangle \colon V \times V \to \mathbf{C}.$$

It has to satisfy:

(1) $\langle v_1 + v_2, w\rangle = \langle v_1, w\rangle + \langle v_2, w\rangle$
(2) $\langle cv, w\rangle = c\langle v, w\rangle$.
(3) $\langle w, v\rangle = \overline{\langle v, w\rangle}$.
(4) $\langle v, v\rangle \geqslant 0$, with equality if and only if $v = 0$.

The complex conjugate of $\alpha = a + bi$ is $\overline{\alpha} = a - bi$ (here $a, b \in \mathbf{R}$). If $\alpha = \overline{\alpha}$, then $\alpha \in \mathbf{R}$. Axiom 3 says that for any $v \in V$, we have $\langle v, v\rangle = \overline{\langle v, v\rangle}$, and therefore $\langle v, v\rangle \in \mathbf{R}$.

In the complex Euclidean plane $\mathbf{C}^2$, let's say we have two vectors

$$\begin{aligned} v &= (\alpha_1, \alpha_2) \\ w &= (\beta_1, \beta_2) \end{aligned}$$

with $\alpha_i, \beta_i \in \mathbf{C}$. The inner product is

$$\langle v, w \rangle = \alpha_1 \overline{\beta}_1 + \alpha_2 \overline{\beta}_2$$

Then

$$\langle v, v \rangle = \alpha_1 \overline{\alpha}_1 + \alpha_2 \overline{\alpha}_2 = |\alpha_1|^2 + |\alpha_2|^2 \geqslant 0$$

This verifies axiom 4.

## 6. Orthonormal bases

**Theorem 6.1.** *Let $V$ be an inner product space, and let $v_1, \dots, v_n \in V$ be all nonzero vectors which are orthogonal to each other. Then $v_1, \dots, v_n$ are linearly independent.*

*Proof.* Let $v_1, \dots, v_n \in V$ be nonzero orthogonal vectors. Assume we have a linear combination of them which equals 0:

$$a_1 v_1 + a_2 v_2 + \cdots + a_n v_n = 0$$

For each $j = 1, \dots, n$, we look at

$$\langle a_1 v_1 + a_2 v_2 + \cdots + a_n v_n, v_j \rangle = 0$$

This becomes

$$a_1 \langle v_1, v_j \rangle + a_2 \langle v_2, v_j \rangle + \cdots + a_n \langle v_n, v_j \rangle = 0.$$

We have $\langle v_i, v_j \rangle = 0$ whenever $i \neq j$, so that the only surviving term is

$$a_j \langle v_j, v_j \rangle = 0.$$

Since $v_j \neq 0$, Axiom 4 says that $\langle v_j, v_j \rangle \neq 0$. Therefore $a_j = 0$. We get that $a_1 = a_2 = \cdots = a_n = 0$. □

**Definition 6.2.** *A vector $v$ in an inner product space is* normal *if $\|v\| = 1$. If $v$ is any nonzero vector, its* normalization *if $v/\|v\|$; this is a normal vector. An* orthonormal basis *for an inner product space is a basis consisting of normal vectors which are mutually orthogonal.*

As an example, what would an orthonormal basis for $\mathbf{R}^2$ look like? One of the vectors can be $v_1 = (a, b)$, the other $v_2 = (c, d)$. We have $\|v_1\|^2 = a^2 + b^2 = 1$ and $\|v_2\|^2 = c^2 + d^2 = 1$. In polar coordinates: $v_1 = (\cos\theta, \sin\theta)$ for some angle $\theta$. There are two possibilities for $v_2$: either it is $(\cos(\theta + \pi/2), \sin(\theta + \pi/2))$, or $(\cos(\theta - \pi/2), \sin(\theta - \pi/2))$. Thus either $v_2 = (-\sin\theta, \cos\theta)$ or else $v_2 = (\sin\theta, -\cos\theta)$. The change of basis matrix for this basis $\{v_1, v_2\}$ is of the form

$$\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \text{ or } \begin{pmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{pmatrix}$$

The first matrix represents a rotation by the angle $\theta$, and the second matrix represents a reflection through a line, which makes an angle of $\theta/2$ with the $x$-axis. Considered as linear transformations, both of these matrices *preserve lengths and angles*, both only the first one *preserves orientation.*

**Theorem 6.3** (Graham-Schmidt)**.** *Let $V$ be a finite-dimensional inner product space. Then there exists an orthonormal basis for $V$.*

Let's say $V$ has such a basis, call it $\{v_1, \dots, v_n\}$. Then any two vectors in $v$, call them $x$ and $y$, would be linear combinations of the $v$s:

$$\begin{aligned} x &= a_1 v_1 + \cdots + a_n v_n \\ y &= b_1 v_1 + \cdots + b_n v_n \end{aligned}$$

What is $\langle x, y\rangle$? It is:

$$\langle x, y\rangle = \langle a_1 v_1 + \cdots + a_n v_n, b_1 v_1 + \cdots + b_n v_n\rangle = a_1\bar{b}_1 + \cdots + a_n\bar{b}_n$$

This is the ordinary dot product for Euclidean space! In other words,

$$\langle x, y\rangle = [x]_B \cdot [y]_B$$

where $B$ is the orthonormal basis. This means that every inner product space resembles Euclidean space with its dot product.

*Proof.* The proof is an algorithm, the *Graham-Schmidt orthogonalization procedure*. It turns an arbitrary basis $w_1, \dots, w_n$, into an orthonormal basis $v_1, \dots, v_n$. We're going to insert an intermediate step, where we construct first a basis $x_1, \dots, x_n$ which is merely orthogonal (not necessarily orthonormal).

The pattern goes:

$$\begin{aligned}
x_1 &= w_1 \\
x_2 &= w_2 - \frac{\langle w_2, x_1\rangle}{\langle x_1, x_1\rangle} x_1 \\
x_3 &= w_3 - \frac{\langle w_3, x_1\rangle}{\langle x_1, x_1\rangle} x_1 - \frac{\langle w_3, x_2\rangle}{\langle x_2, x_2\rangle} x_2 \\
&\vdots \\
x_n &= w_n - \sum_{i=1}^{n-1} \frac{\langle w_n, x_i\rangle}{\langle x_i, x_i\rangle} x_i
\end{aligned}$$

Let's compute some inner products:

$$\begin{aligned}
\langle x_2, x_1\rangle &= \langle w_2, x_1\rangle - \frac{\langle w_2, x_1\rangle}{\langle x_1, x_1\rangle}\langle x_1, x_1\rangle = 0 \\
\langle x_3, x_1\rangle &= \langle w_3, x_1\rangle - \frac{\langle w_3, x_1\rangle}{\langle x_1, x_1\rangle}\langle x_1, x_1\rangle - \frac{\langle w_3, x_2\rangle}{\langle x_2, x_2\rangle}\langle x_2, x_1\rangle = 0 \\
\langle x_3, x_2\rangle &= \langle w_3, x_2\rangle - \frac{\langle w_3, x_1\rangle}{\langle x_1, x_1\rangle}\langle x_1, x_2\rangle - \frac{\langle w_3, x_2\rangle}{\langle x_2, x_2\rangle}\langle x_2, x_2\rangle = 0 \\
&\vdots
\end{aligned}$$

We find that the $x_1, \dots, x_n$ are all orthogonal to each other. In fact their span is the same as the span of the $w_1, \dots, w_n$ (why?), so that they are a basis. Finally, the orthonormal basis is obtained by normalizing:

$$v_i = x_i / \|x_i\|.$$

□

Let's do an example. Let $V = P_2(\mathbf{R})$, the vector space of polynomials with degree $\leqslant 2$. Let's define an inner product on this by

$$\langle f, g\rangle = \int_{-1}^{1} f(x)g(x)\, dx.$$

Find an orthonormal basis for $V$.

Let's start with the basis $\{1, x, x^2\}$ for $V$. We'll apply Graham-Schmidt to this. It'll be useful to pre-compute the inner products among these vectors:

$$\begin{aligned}
\langle 1, 1 \rangle &= \int_{-1}^{1} 1 \, dx = 2 \\
\langle 1, x \rangle &= \int_{-1}^{1} x \, dx = 0 \\
\langle 1, x^2 \rangle &= \int_{-1}^{1} x^2 \, dx = \frac{2}{3} \\
\langle x, x^2 \rangle &= \int_{-1}^{1} x^3 \, dx = 0
\end{aligned}$$

Applying Graham-Schmidt to $\{1, x, x^2\}$ produces these orthogonal vectors:

$$\begin{aligned}
x_1 &= 1 \\
x_2 &= x - \frac{\langle x, 1 \rangle}{\langle 1, 1 \rangle} 1 = x \\
x_3 &= x^2 - \frac{\langle x^2, 1 \rangle}{\langle 1, 1 \rangle} 1 - \frac{\langle x^2, x \rangle}{\langle x, x \rangle} x = x^2 - \frac{1}{3}
\end{aligned}$$

Normalizing these vectors produces the orthonormal basis

$$\left\{ \frac{1}{\sqrt{2}}, \sqrt{\frac{3}{2}} x, \sqrt{\frac{5}{8}} (3x^2 - 1) \right\}.$$

## 7. Fourier coefficients

Let $V$ be a finite-dimensional inner product space. Let $B = \{v_1, v_2, \ldots, v_n\}$ be an orthonormal basis. This means

$$\langle v_i, v_j \rangle = \delta_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

**Theorem 7.1.** *Let $y \in V$ be any vector, then*

$$y = \sum_{i=1}^{n} \langle y, v_i \rangle v_i$$

*Proof.* We know already that $y$ must be a linear combination of basis vectors:

$$y = \sum_{i=1}^{n} a_i v_i$$

Now consider $\langle y, v_j \rangle$ for each $j = 1, \ldots, n$:

$$\langle y, v_j \rangle = \langle \sum_{i=1}^{n} a_i v_i, v_j \rangle = \sum_{i=1}^{n} a_i \langle v_i, v_j \rangle = \sum_{i=1}^{n} a_i \delta_{ij} = a_j$$

Plug this in to our expression for $y$, and we're done. □

The coefficients $\langle y, v_i \rangle$ are called the *Fourier coefficients* of $y$ with respect to the orthonormal basis $B$.

## 8. Orthogonal complements

**Definition 8.1.** *Let $V$ be an inner product space, and let $S \subset V$ be a subset. I define*

$$S^{\perp} = \left\{ v \in V \;\middle|\; \langle v, s \rangle = 0, \text{ all } s \in S \right\}$$

*This is called the* orthogonal complement *of $S$ in $V$.*

Then $S^{\perp}$ is a subspace of $V$.

**Example 8.2.** $\{0\}^{\perp} = V$. *Also,* $V^{\perp} = \{0\}$.

**Example 8.3.** *Let* $V = \mathbf{R}^2$ *with its dot product, and let* $v = e_1 = (1,0)$. *Then* $\{v\}^{\perp}$ *is the span of* $e_2 = (0,1)$. *In general, if* $(a,b) \neq (0,0)$ *then* $\{(a,b)\}^{\perp}$ *is the span of* $(b,-a)$, *because*

$$(a,b)\cdot(b,-a) = ab - ba = 0.$$

**Example 8.4.** *The orthogonal complement of a line in* $\mathbf{R}^3$ *is a plane.*

In general, if $W \subset V$ is a subspace of a finite-dimensional inner product space, then $\dim W + \dim W^{\perp} = \dim V$.

## 9. The closest vector problem

Given a finite-dimensional inner product space $V$, and two vectors $x, y \in V$, the norm $\|x-y\|$ can be interpreted as the *distance* between $x$ and $y$. If $V = \mathbf{R}^n$ with its dot product, then $\|x-y\|$ is simply the Euclidean distance that you would calculate with the distance formula.

Given a subspace $W \subset V$ and a vector $y \in V$ (probably not in $W$), find the vector $u \in V$ which is *closest* to $y$.

**Theorem 9.1.** *Given a subspace* $W \subset V$ *and a vector* $y \in V$, *there exist unique vectors* $u \in W$ *and* $z \in W^{\perp}$, *such that* $y = u + z$. *Furthermore,* $u$ *is the closest vector in* $W$ *to* $y$.

*Proof.* Let $v_1, \ldots, v_n$ be an orthonormal basis for $W$. Let

$$u = \sum_{i=1}^{n} \langle y, v_i \rangle v_i.$$

Then $u \in W$, because it is a linear combination of the vectors $v_i \in W$. Let $z = y - u$. I claim that $z \in W^{\perp}$. It is enough to check that $\langle z, v_j \rangle = 0$ for all $j$. We have

$$\begin{aligned}
\langle z, v_j \rangle &= \langle y - u, v_j \rangle \\
&= \langle y, v_j \rangle - \langle \sum_{i=1}^{n} \langle y, v_i \rangle v_i, v_j \rangle \\
&= \langle y, v_j \rangle - \sum_{i=1}^{n} \langle y, v_i \rangle \langle v_i, v_j \rangle \\
&= \langle y, v_j \rangle - \sum_{i=1}^{n} \langle y, v_i \rangle \delta_{ij} \\
&= \langle y, v_j \rangle - \langle y, v_j \rangle = 0
\end{aligned}$$

Thus $y = u + z$ for some $u \in W$ and $z \in W^{\perp}$. (Uniqueness left as exercise.)

Why is $u$ the closest to $y$? Suppose $u' \in W$ is any other vector. I claim that $u$ is closer to $y$ than $u'$. The distance between $y$ and $u'$ is

$$\left\|y - u'\right\| = \left\|(y-u) - (u'-u)\right\| = \left\|z - (u'-u)\right\|$$

We have that $u', u \in W$, so $u' - u \in W$. Therefore $z$ is perpendicular to $u' - u$, so by Pythagoras:

$$\left\|y - u'\right\|^2 = \left\|z - (u'-u)\right\|^2 = \|z\|^2 + \left\|u'-u\right\|^2 = \|y-u\|^2 + \left\|u'-u\right\|^2 \geqslant \|y-u\|^2,$$

with equality holding if and only if $u = u'$.

□

## 10. Transposes and adjoints

Remember that for an $m \times n$ matrix $A$, the transpose $A^t$ is defined by swapping the rows and columns of $A$. Thus, $A^t$ is an $n \times m$ matrix.

It's related to the dot product in the following way. If $v = \begin{pmatrix} a \\ b \end{pmatrix}$ and $w = \begin{pmatrix} c \\ d \end{pmatrix}$, then

$$v \cdot w = ac + bd = \begin{pmatrix} a & b \end{pmatrix} \begin{pmatrix} c \\ d \end{pmatrix} = v^t w$$

Recall this property of transposes:

$$(AB)^t = B^t A^t$$

So now say we have a matrix $A$, a vector $v$, a vector $w$, so that $Av$ is defined and in the same vector space as $w$.

Then

$$Av \cdot w = (Av)^t w = v^t A^t w = v \cdot A^t w.$$

Therefore:

$$Av \cdot w = v \cdot A^t w$$

If $V$ is a real inner product space, and $T$ is a linear operator on $V$, an *adjoint* of $T$ is a linear operator $T^*$ satisfying the property:

$$\langle Tv, w \rangle = \langle v, T^* w \rangle, \text{ for all } v, w \in W$$

In real Euclidean space, the adjoint is the transpose.

It's a little more general than this: For two inner product spaces $V$ and $W$, and a linear transformation $T \colon V \to W$, the *adjoint* of $T$ is a linear transformation $T^* \colon W \to V$, satisfying:

$$\langle Tv, w \rangle_W = \langle v, T^* w \rangle_V, \text{ for all } v \in V, w \in W.$$

In the situation of complex Euclidean spaces, $T$ is a matrix $A$ with complex entries, and then

$$A^* = \overline{A}^t.$$

Thus

$$\begin{pmatrix} i & 2i \\ 3 & 4+i \end{pmatrix}^* = \begin{pmatrix} -i & 3 \\ -2i & 4-i \end{pmatrix}$$

Importantly, we always have

$$A^{**} = A.$$

## 11. Least squares regression

Geometrically, you may want to fit a line that matches some data points in the plane. Say, you have points $(x_i, y_i) = (1, 2), (2, 3), (3, 5), (4, 7)$. We want to find the line $y = mx + b$ that fits the data best.

What should this mean? For each $i = 1, 2, 3, 4$, the observed data point is $(x_i, y_i)$, and the predicted data point is $(x_i, mx_i + b)$. The error is $|y_i - (mx_i + b)|$.

We want to minimize the squared errors:

$$\sum_{i=1}^{4} (y_i - (mx_i + b))^2$$

We rewrite this in terms of vectors and their norms. Let

$$A = \begin{pmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \\ 4 & 1 \end{pmatrix}, x = \begin{pmatrix} m \\ b \end{pmatrix}, y = \begin{pmatrix} 2 \\ 3 \\ 5 \\ 7 \end{pmatrix}$$

Then

$$Ax = \begin{pmatrix} m+b \\ 2m+b \\ 3m+b \\ 4m+b \end{pmatrix}$$

$$\|Ax - y\|^2 = \sum_{i=1}^{4} (y_i - (mx_i + b))^2$$

is the squared error. We are given $A$ and $y$, and want to find $x$. This is a bit like trying to solve $Ax = y$, but there might not be an exact solution – that would be if the line fit the data perfectly.

Instead, we are looking for the $x$ that makes $\|Ax - y\|$ smallest.

This is a closest vector problem! Let

$$W = R(A)$$

be the range of the $4 \times 2$ matrix $A$, so that $W$ is a subspace of $\mathbf{R}^4$. Remember that the range $R(A)$ is the set of all vectors $Ax$. We want to find the vector in $W$ which is closest to $y$.

By our theorem about the closest vector, there exist vectors $u \in W$ and $z \in W^\perp$ such that

$$y = u + z,$$

and then $u$ is the closest vector in $W$ to $y$. Since $u \in W = R(A)$, we must have $u = Ax$ for some $x \in \mathbf{R}^2$.

Let's spell out what it means for $z \in W^\perp = R(A)^\perp$. This means, for all $v \in \mathbf{R}^2$, we have $\langle Av, z\rangle = 0$. Therefore $\langle v, A^*z\rangle = 0$ for all $v \in \mathbf{R}^2$. This means that $A^*z \in (\mathbf{R}^2)^\perp = \{0\}$. We have then $A^*z = 0$.

We want to solve for $x$ in

$$y = Ax + z.$$

Multiply both sides by $A^*$:

$$A^*y = A^*Ax + A^*z = A^*Ax.$$

As long as $A^*A$ is invertible, we could solve for $x$:

$$x = (A^*A)^{-1}A^*y.$$

We have

$$A^*A = \begin{pmatrix} 30 & 10 \\ 10 & 4 \end{pmatrix}$$

and

$$(A^*A)^{-1} = \frac{1}{20}\begin{pmatrix} 4 & -10 \\ -10 & 30 \end{pmatrix} = \frac{1}{10}\begin{pmatrix} 2 & -5 \\ -5 & 15 \end{pmatrix}$$

and finally

$$x = (A^*A)^{-1}A^*y = \frac{1}{10}\begin{pmatrix} 2 & -5 \\ -5 & 15 \end{pmatrix}\begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 1 & 1 & 1 \end{pmatrix}\begin{pmatrix} 2 \\ 3 \\ 5 \\ 7 \end{pmatrix} = \begin{pmatrix} 1.7 \\ 0 \end{pmatrix}$$

Therefore the line $y = 1.7x$ best fits the data.

To make this work, we needed $A^*A$ to be an invertible matrix.

**Theorem 11.1.** *Let $A$ be an $m \times n$ matrix, so that $A^*A$ is an $n \times n$ matrix. Then $A^*A$ and $A$ have the same rank.*

As a corollary to this, if $A$ has rank $n$, then $A^*A$ has rank $n$, and so it is invertible.

*Proof.* By the dimension theorem, it suffices to show that the nullities of $A$ and $A^*A$ are the same. In fact we claim that $N(A) = N(A^*A)$.

For $N(A) \subset N(A^*A)$: let $v \in N(A)$, so that $Av = 0$, then $A^*Av = 0$, so $v \in N(A^*A)$.

For $N(A^*A) \subset N(A)$: let $v \in N(A^*A)$, so that $A^*Av = 0$. Then $\langle A^*Av, v\rangle = 0$. Thus $\langle Av, Av\rangle = 0$. By Axiom 4, $Av = 0$, which means that $v \in N(A)$. $\square$

## 12. Normal and self-adjoint operators

Let $F = \mathbf{R}$ or $F = \mathbf{C}$. Remember that for an $m \times n$ matrix $A$ with coefficients in $F$, its adjoint is its conjugate transpose:

$$A^* = \overline{A}^t.$$

(So if $F = \mathbf{R}$, the adjoint is just the transpose.) The adjoint is characterized by the property that

$$Av \cdot w = v \cdot A^*w$$

for any $v \in \mathbf{R}^n$, $w \in \mathbf{R}^m$.

In general, if we have two inner product spaces $V$ and $W$ over $F$ ($F = \mathbf{R}$ or $F = \mathbf{C}$), and a linear transformation $T\colon V \to W$, we say that a linear transformation $T^*\colon W \to V$ is the *adjoint* of $T$ is characterized by the equation

$$\langle Tv, w\rangle_W = \langle v, T^*w\rangle_V.$$

In today's lecture we focus on the case that $A$ is a square matrix. We're going to see how *inner product spaces* interact with the ideas from the previous chapter about *eigenvectors and eigenvalues*.

**Definition 12.1.** *Let $T$ be a linear operator on a vector space $V$. We say that $T$ is* diagonalizable *if there exists a basis for $V$ consisting of eigenvectors for $T$.*

If $V$ is finite-dimensional, this is the same as asking that there exist a basis $B$ for $V$ such that $[T]_B$ is a diagonal matrix.

In the case that $T$ comes from a square matrix $A$, it means that there exists an invertible matrix $Q$ such that $Q^{-1}AQ$ is diagonal.

Now suppose that $V$ is an inner product space, and $T\colon V \to V$ is a linear operator. We have already seen (Gram-Schmidt!) that (if $V$ is finite-dimensional) $V$ has a basis consisting of orthonormal vectors.

An orthonormal basis for $V$ is a basis $v_1, \ldots, v_n$ such that

$$\langle v_i, v_j\rangle = \delta_{ij}$$

**Definition 12.2.** *Let $T$ be a linear operator on an inner product space space $V$. We say that $T$ is* orthogonally diagonalizable *if there exists an orthonormal basis for $V$ consisting of eigenvectors for $T$.*

This means that there exists an orthonormal basis $B$ for $V$ such that $[T]_B$ is diagonal.

If $T$ comes from a square matrix $A$, then for $T$ to be orthogonally diagonalizable, it would mean that there exists an invertible matrix $Q$ with $Q^{-1}AQ$ diagonal, but require that $Q$ have orthonormal columns. This means that $QQ^* = I$. We call such a $Q$ an *orthogonal* matrix.

**Definition 12.3.** *Let $T\colon V \to V$ be a linear operator on a finite-dimensional inner product space $V$. We say that $T$ is* normal *if $T$ and $T^*$ commute: $TT^* = T^*T$.*

We also say that a square matrix $A$ is normal if $AA^* = A^*A$.

How could we ever find such a matrix?

**Definition 12.4.** *A matrix $A$ is* self-adjoint *if $A^* = A$.*

Obviously, a self-adjoint matrix is normal.

If the scalar field is $\mathbf{R}$, then a self-adjoint matrix is one satisfying $A^t = A$. This is called a *symmetric matrix.*

An example would be the matrix

$$A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Geometrically, this is a reflection about $y = x$. The characteristic polynomial is $t^2 - 1 = (t-1)(t+1)$. So the eigenvalues are $1, -1$.

The vector $v_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ is an eigenvector with value 1, and $v_2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$ is an eigenvector with value $-1$. Notice that $v_1 \cdot v_2 = 0$, so these are orthogonal. An orthonormal basis consisting of eigenvectors is

$$\frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

Another example is

$$A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

Geometrically, this is rotation by $\pi/2$. This matrix is anti-symmetric: $A^t = A^* = -A$. But this is enough to conclude that $A$ is normal, since $A$ and $-A$ always commute. The characteristic polynomial is $t^+1 = (t-i))(t+i)$, which only has complex roots. So as a real matrix, this is normal but not diagonalizable. As a complex matrix though, it is orthogonally diagonalizable.

We have

$$\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ i \end{pmatrix} = \begin{pmatrix} -i \\ 1 \end{pmatrix} = -i \begin{pmatrix} 1 \\ i \end{pmatrix}$$

The two eigenvectors for $A$ are $v_1 = \begin{pmatrix} 1 \\ i \end{pmatrix}$ and $v_2 = \begin{pmatrix} 1 \\ -i \end{pmatrix}$, with eigenvalues $-i$ and $i$, respectively. Then $v_1$ and $v_2$ are orthogonal.

**Theorem 12.5.** *Let $T$ be a linear operator on an inner product space $V$ which is normal. For any $v \in V$, we have*

$$\|T(v)\| = \|T^*(v)\|$$

*Proof.* Let $v \in V$. We have

$$\|T(v)\|^2 = \langle T(v), T(v) \rangle = \langle v, T^*T(v) \rangle = \langle v, TT^*(v) \rangle = \langle T^*(v), T^*(v) \rangle = \|T^*(v)\|^2$$

□

**Theorem 12.6.** *Let $T$ be a linear operator on an inner product space which is normal. Let $v$ be an eigenvector of $T$, with value $\lambda$. Then $v$ is also an eigenvector of $T^*$, with value $\overline{\lambda}$.*

*Proof.* Let $v$ be an eigenvector of $T$ with value $\lambda$. Then $Tv = \lambda v$. We have $Tv - \lambda v = 0$, and so if $U = T - \lambda I$, then $Uv = 0$. I claim that $U$ is also a normal matrix (exercise), and so

$$0 = \|U(v)\| = \|U^*(v)\|,$$

so that $U^*(v) = 0$. We have $U^* = T^* - \overline{\lambda} I$. The fact that $U^*(v) = 0$ means that $T^*(v) = \overline{\lambda} v$, so that $v$ is an eigenvector for $T^*$ with value $\overline{\lambda}$. □

We have two big theorems about normal operators. The first concerns complex scalars, and the second concerns real scalars.

**Theorem 12.7.** *Let $T$ be a linear operator on a finite-dimensional complex inner product space $V$. Then $T$ is orthogonally diagonalizable if and only if it is normal.*

Remember that normal means that $T$ and $T^*$ commute, and orthogonally diagonalizable means that $V$ has a orthonormal basis consisting of eigenvectors for $T$.

*Proof.* For the "if" direction, assume that $T$ is normal." Let's first observe what happens in the case that $\dim V = 2$. Since we're over $\mathbf{C}$, there exists one eigenvector $v_1$ for $T$, say with value $\lambda_1$. Normalize this so that $\|v_1\| = 1$. Now consider $\{v_1\}^\perp$: this has to be one-dimensional, with normal vector $v_2$. By the previous theorem we have $T^*v_1 = \overline{\lambda}_1 v_1$, and so

$$0 = \overline{\lambda}_1\langle v_1, v_2\rangle = \langle T^*v_1, v_2\rangle = \langle v_1, Tv_2\rangle,$$

so that $Tv_2 \in \{v_1\}^\perp$ lies in the span of $v_2$: $Tv_2 = \lambda_2 v_2$. We now have a basis for $V$ consisting of eigenvectors, namely $\{v_1, v_2\}$.

Let's also do the case $\dim V = 3$ (the general case is by induction). Once again let $v_1 \in V$ be a normalized eigenvector, with value $\lambda_1$. Now consider $W = \{v_1\}^\perp$, a 2-dimensional inner product space. Given $w \in W$, we have

$$0 = \overline{\lambda}_1\langle v_1, w\rangle = \langle T^*v_1, w\rangle = \langle v_1, Tw\rangle,$$

so that $Tw \in W$ again. Thus $T$ stabilizes $W$. We can now apply the 2-dimensional case to $W$ to get a basis $\{v_2, v_3\}$ of $W$ consisting of orthonormal eigenvectors for $T$. Then $\{v_1, v_2, v_3\}$ is a basis of $V$ consisting of orthonormal eigenvectors for $T$.

For the converse of the theorem, suppose $B$ is an orthonormal basis consisting of eigenvectors for $T$. Then $[T]_B$ is a diagonal matrix. Then so is $[T^*]_B = [T]_B^*$. We know that diagonal matrices commute with each other, and therefore so do $T$ and $T^*$.

□

The last theorem concerns *real* inner product spaces. Recall that an operator $T$ is *self-adjoint* if $T = T^*$. (For a real matrix $A$ to be self-adjoint, it just means that $A = A^t$, so that $A$ is symmetric about its diagonal. Such matrices are often called "symmetric".)

**Theorem 12.8.** *Let $T$ be a linear operator on a finite-dimensional real inner product space $V$. Then $T$ is orthogonally diagonalizable if and only if it is self-adjoint.*

Remember that self-adjoint means that $T = T^*$. For a real matrix $A$ to be self-adjoint, it means that $A = A^t$, so that $A$ is an $n \times n$ symmetric matrix.

**Example 12.9.** *Let $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ be a symmetric $2 \times 2$ matrix. Then $b = c$, so $A = \begin{pmatrix} a & b \\ b & d \end{pmatrix}$. Assume that $A$ is real, so that $a, b, d \in \mathbf{R}$. The theorem says that $A$ is diagonalizable, let's verify this.*

*The characteristic polynomial of $A$ is*

$$t^2 - (a + d)t + (ad - b^2)$$

*To investigate the nature of the roots, I look at the discriminant of this polynomial:*

$$(a + d)^2 - 4(ad - b^2) = a^2 + 2ad + d^2 - 4ad + 4b^2 = a^2 - 2ad + d^2 + 4b^2 = (a - d)^2 + (2b)^2 \geqslant 0.$$

*The discriminant is positive as long as $a-d$ and $2b$ are not both zero. In that case, the roots of this polynomial are distinct and real, and therefore $A$ is diagonalizable. Otherwise, $a = d$ and $b = 0$, but then $A = \begin{pmatrix} a & 0 \\ 0 & a \end{pmatrix}$ is already diagonal.*

*Proof.* For the "if" direction, assume that $T$ is self-adjoint. After choosing a basis for $V$, let's represent $T$ as an $n \times n$ matrix $A$, with real entries. Thus $A = A^t$ is symmetric. It will be enough to show that there is an orthonormal basis for $\mathbf{R}^n$ consisting of eigenvectors for $A$.

Even though $A$ has real entries, it still makes sense to apply $A$ to vectors in $\mathbf{C}^n$, thereby getting a linear operator on $\mathbf{C}^n$. So we may talk about a complex eigenvector $v \in \mathbf{C}^n$ with eigenvalue $\lambda \in \mathbf{C}$. We have

$$\lambda\langle v, v\rangle = \langle Av, v\rangle = \langle v, A^*v\rangle = \langle v, Av\rangle = \overline{\langle Av, v\rangle} = \overline{\lambda\langle v, v\rangle} = \overline{\lambda}\langle v, v\rangle.$$

Since $\langle v, v\rangle \neq 0$, we must have $\lambda = \overline{\lambda}$, so that $\lambda \in \mathbf{R}$. We have shown that all eigenvalues of $A$ are real; that is, the characteristic polynomial of $A$ has only real roots.

Let $\lambda_1$ be one of those roots, and let $v_1 \in \mathbf{R}^n$ be an eigenvector: $Av_1 = \lambda_1 v_1$. Let $W = \{v_1\}^\perp$. Then $A$ preserves $W$ (same proof as before, uses the fact that $A$ is normal). Proceed inductively to find an orthonormal basis of eigenvectors for $A$.

For the "only if" direction: Assume that $V$ has an orthonormal basis $B$ consisting of eigenvectors of $T$. Then $[T]_B$ is a diagonal matrix. Such matrices are always self-adjoint: $[T]_B = [T]_B^*$. This implies that $T = T^*$, so that $T$ is self-adjoint. □

## 13. Othogonal and Unitary operators

Let's review some of our vocab words from this chapter:

(1) $T$ is normal if $TT^* = T^*T$.
(2) $T$ is self-adjoint if $T = T^*$.
(3) $T$ is orthogonal (in the real case) or unitary (in the complex case) if $TT^* = I$.

Think about an operator $T\colon \mathbf{R}^2 \to \mathbf{R}^2$ geometrically, as moving points in spaces. Let's ask the question: is length preserved? Is angle preserved? Is area preserved? Is orientation preserved?

As far as the last two questions go, the answer involves the determinant: The area gets stretched out by a factor of $|\det T|$, and orientation goes with the sign of $\det T$.

**Definition 13.1.** *Let $T$ be a linear operator on an inner product space $V$ (assume finite dimensional). Suppose that for all $v \in V$ we have*

$$\|T(v)\| = \|v\|.$$

*Then we call $T$* orthogonal *if $V$ is real, and* unitary *if $V$ is complex.*

Thus rotations and reflections are orthogonal, but dilations and shears are not.

**Theorem 13.2.** *The following are equivalent.*

(1) *For all $v, w \in V$, we have*
$$\langle T(v), T(w)\rangle = \langle v, w\rangle$$
*(angle-preserving).*
(2) *$T$ is orthogonal or unitary (length-preserving): for all $v \in V$ we have*
$$\|T(v)\| = \|v\|.$$
(3) *$TT^* = I$.*

*Proof.* Let's start with (1) implies (2). Assume that $T$ is angle-preserving. Then for any $v \in V$, we have

$$\|T(v)\| = \sqrt{\langle T(v), T(v)\rangle} = \sqrt{\langle v, v\rangle} = \|v\|.$$

Next we'll do (2) implies (3). Assume that $T$ is length-preserving. This means that

$$\langle v, v\rangle = \langle T(v), T(v)\rangle = \langle v, T^*T(v)\rangle$$

This means that

$$\langle v, T^*T(v)\rangle - \langle v, v\rangle = \langle v, T^*T(v) - v\rangle = \langle v, (T^*T - I)v\rangle = 0$$

Let $U = T^*T - I$. Then $\langle v, U(v)\rangle = 0$ for all $v$. We also know that

$$U^* = (T^*T - I)^* = (T^*T)^* - I^* = T^*T^{**} - I = T^*T - I = U.$$

Therefore $U$ is self-adjoint. By our previous theorem, we know that there's a basis for $V$ consisting of eigenvectors for $U$. Let $v$ be one of those basis vectors, so that $Uv = \lambda v$. We have

$$0 = \langle v, U(v)\rangle = \langle v, \lambda v\rangle = \overline{\lambda}\langle v, v\rangle$$

Since $v \neq 0$, neither is $\langle v, v\rangle$ and therefore $\lambda = 0$. So $U(v) = 0$. Since this is true for all vectors in a basis for $v$, we must have that $U = 0$ identically. Thus $T^*T = I$, and so $T^* = T^{-1}$, which also implies $TT^* = I$.

For (3) implies (1), assume that $TT^* = I$. Therefore $T^*T = I$. We have

$$\langle T(v), T(w)\rangle = \langle T^*T(v), w\rangle = \langle v, w\rangle.$$

Therefore $T$ is angle-preserving. $\square$

**Example 13.3.** *Let $A$ be a $2\times 2$ orthogonal or unitary matrix, $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$. We would have*

$$AA^* = \begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} \overline{a} & \overline{c} \\ \overline{b} & \overline{d} \end{pmatrix} = \begin{pmatrix} |a|^2 + |b|^2 & a\overline{c} + b\overline{d} \\ \overline{a}c + \overline{b}d & |c|^2 + |d|^2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$

*so that*

$$\begin{aligned} |a|^2 + |b|^2 &= 1 \\ |c|^2 + |d|^2 &= 1 \\ a\overline{c} + b\overline{d} &= 0 \end{aligned}$$

*So if I let $v = (a, b)$ and $w = (c, d)$, these equations say that*

$$\begin{aligned} \|v\| &= 1 \\ \|w\| &= 1 \\ \langle v, w\rangle &= 0 \end{aligned}$$

*Thus the rows of $A$ are an orthonormal basis for $F^2$. The same is true for the columns of $A$.*

**Example 13.4.** *Let $A$ be a $2\times 2$ orthogonal (real) matrix, so that $AA^t = I$. Then $A$ falls into one of two categories:*

$$A = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$

*(rotation by $\theta$), or else:*

$$A = \begin{pmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{pmatrix}$$

*(reflection through the line which makes an angle of $\theta/2$ with the $x$-axis).*

Notice that in the first case, $\det A = 1$, and in the second case, $\det A = -1$. This matrix is not diagonalizable. The roots of the characteristic polynomial are complex, they are $e^{i\theta}, e^{-i\theta}$.

In the second case, consider $v = \begin{pmatrix} \cos(\theta/2) \\ \sin(\theta/2) \end{pmatrix}$. We have

$$Av = \begin{pmatrix} \cos(\theta)\cos(\theta/2) + \sin(\theta)\sin(\theta/2) \\ \sin(\theta)\cos(\theta/2) - \cos(\theta)\sin(\theta/2) \end{pmatrix} = \begin{pmatrix} \cos(\theta/2) \\ \sin(\theta/2) \end{pmatrix} = v.$$

If $w = \begin{pmatrix} \sin(\theta/2) \\ -\cos(\theta/2) \end{pmatrix}$, then $Aw = -w$. We have that $v$ and $w$ are an orthonormal basis for $\mathbf{R}^2$, consisting of eigenvectors for $A$.

We conclude that $A$ is a reflection through the line containing the vector $v$. We also conclude that $A$ is orthogonally similar to the matrix $\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$.

**Theorem 13.5.** *The eigenvalues of an orthogonal or unitary matrix all have absolute value 1.*