Lecture 22: Orthogonal projection and adjoints

Last lecture Phil talked about inner products and orthogonality. This leads to the idea of an orthogonal projection.

Definition 0.1. *Let V be a finite-dimensional (complex) inner product space. Letting W be a subspace of V, we can write each vector v uniquely as $v = w_1 + w_2$, where $w_1 \in W$ and $w_2 \in W^\perp$. Then define the orthogonal projection onto W as*

$$P_W(v) = w_1 \ .$$

There are many ways to define the orthogonal projection. In some sense it is the "closest" vector to v in w (we will see this soon). Here are some simple properties.

- P_W is linear.
- $P_W^2 = P_W$
- $P_{W^\perp} = I - P_W$.
- For all $v_1, v_2 \in V$, $\langle P_W(v_1), v_2\rangle = \langle v_1, P_W(v_2)\rangle$. This says P_W can be moved "to the other side" in the inner product. Formally this means P_W is its own adjoint (defined soon).

  Proof.

  $$\langle P_W(v_1), v_2\rangle = \langle P_W(v_1), P_W(v_2)\rangle + \langle P_W(v_1), P_{W^\perp}(v_2)\rangle = \langle P_W(v_1), P_W(v_2)\rangle \ .$$

  By the same argument, $\langle v_1, P_W(v_2)\rangle = \langle P_W(v_1), P_W(v_2)\rangle$. □

- For all $v \in V$, $w \in W$,

  $$\|v - P_W(v)\| \leq \|v - w\| \text{ with equality iff } w = P_W(v) \ .$$

  Here the norm comes from the inner product. This says that $P_W(v)$ is the unique closest vector to v in W.

  Proof. Note first that for any $w \in W$ and $w' \in W^\perp$, the *Pythagoras theorem* holds:

  $$\|w + w'\|^2 = \langle w + w', w + w'\rangle = \langle w, w\rangle + 2\langle w, w'\rangle + \langle w', w'\rangle = \|w\|^2 + \|w'\|^2 \ .$$

  Now

  $$\|v - w\|^2 = \|P_W(v) - w + P_{W^\perp}(v)\|^2 = \|P_W(v) - w\|^2 + \|P_{W^\perp}(v)\|^2 \ .$$

  This is at least $\|P_{W^\perp}(v)\|^2 = \|v - P_W(v)\|^2$ and they are equal if and only if $P_W(v) = w$. □

Projection onto a vector. If w is a nonzero vector, we can define $W = \text{Span}(w)$ and consider the orthogonal projection onto W. Let $\{w_1, \ldots, w_{n-1}\}$ be an orthonormal basis for $W^\perp$ (which exists by Gram-Schmidt) and normalize w to be $w' = w/\|w\|$. Then we can write an arbitrary $v \in V$ in terms of the basis $\{w_1, \ldots, w_{n-1}, w'\}$ as

$$v = \langle v, w_1 \rangle w_1 + \cdots + \langle v, w_{n-1} \rangle w_{n-1} + \langle v, w' \rangle w' \ .$$

This gives a formula for $P_W(v) = \langle v, w' \rangle w'$. Rewriting in terms of w, we get

$$P_W(v) = \frac{\langle v, w \rangle}{\|w\|^2} w \ .$$

Since the orthogonal projection onto a subspace was defined without reference to a basis, we see that this does not depend on the choice of w!

We then define the orthogonal projection onto the vector w to be P_W, where $W = \text{Span}(\{w\})$. That is, $P_w(v) = \frac{\langle v, w \rangle}{\|w\|^2} w$.

Adjoint

Before we saw that an orthogonal projection can be moved to the other side of the inner product. This motivates us to look at which operators can do this. Given any linear T, we can define another linear transformation which acts "on the other side" of the inner product.

Theorem 0.2 (Existence of adjoint). *Let $T : V \to V$ be linear and V a finite-dimensional complex inner product space. There exists a unique linear $T^* : V \to V$ such that for all $v, w \in V$,*

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

T^* is called the adjoint of T.

Proof. For the proof, we need a lemma.

Lemma 0.3 (Riesz representation theorem). *Let V a finite-dimensional inner product space. For each $f \in V^*$ there exists a unique $w_f \in V$ such that for all $v \in V$,*

$$f(v) = \langle v, w_f \rangle \ .$$

Proof. Recall the map $R_{\langle \cdot, \cdot \rangle} : V \to V^*$ given by $R_{\langle \cdot, \cdot \rangle}(w)(v) = \langle v, w \rangle$. Since the inner product is a sesquilinear form, this map is linear. In fact, the inner product has rank equal to $\dim V$ when viewed as a sesquilinear form, since only the zero vector is in its nullspace. Thus the rank of $R_{\langle \cdot, \cdot \rangle}$ is $\dim V$ and it is an isomorphism. This means that given $f \in V^*$ there exists a unique $w_f \in V$ such that $R_{\langle \cdot, \cdot, \rangle}(w_f) = f$. In other words, a unique $w_f \in V$ such that for all $v \in V$,

$$\langle v, w_f \rangle = R_{\langle \cdot, \cdot \rangle}(w_f)(v) = f(v) \ .$$

□

We now use the lemma. Given $T : V \to V$ linear and $w \in V$, define a function $f_{T,w} : V \to \mathbb{C}$ by

$$f_{T,w}(v) = \langle T(v), w \rangle \ .$$

This is a linear functional, since it equals $R_w \circ T$. So by Riesz, there exists a unique $\hat{w}$ such that

$$\langle T(v), w \rangle = f_{T,w}(v) = \langle v, \hat{w} \rangle \text{ for all } v \in V \ .$$

We define $T^*(w) = \hat{w}$.

By definition $T^*(w)$ satisfies $\langle T(v), w \rangle = \langle v, T^*(w) \rangle$ for all $v, w \in V$. We must simply show it is linear. So given $c \in \mathbb{C}$ and $v, w_1, w_2 \in V$,

$$\begin{aligned}\langle v, T^*(cw_1 + w_2) \rangle &= \langle T(v), cw_1 + w_2 \rangle = \overline{c} \langle T(v), w_1 \rangle + \langle T(v), w_2 \rangle \\ &= \overline{c} \langle v, T^*(w_1) \rangle + \langle v, T^*(w_2) \rangle \\ &= \langle v, cT^*(w_1) + T^*(w_2) \rangle \ .\end{aligned}$$

By uniqueness, $T^*(cw_1 + w_2) = cT^*(w_1) + T^*(w_2)$. □

The adjoint has many interesting properties. Some simple ones you can verify:

- $(T + S)^* = T^* + S^*$
- $(TS)^* = S^* T^*$.
- $(cT)^* = \overline{c} T^*$.

These can be seen to follow from the next property.

Proposition 0.4. *Let $T : V \to V$ be linear with B an orthonormal basis. Then*

$$[T^*]_B^B = \overline{([T]_B^B)}^t \ .$$

Proof. Write $B = \{v_1, \ldots, v_n\}$ and use orthonormality to express

$$\begin{aligned}T^*(v_j) &= \langle T^*(v_j), v_1 \rangle v_1 + \cdots + \langle T^*(v_j), v_n \rangle v_n \\ &= \overline{\langle T(v_1), v_j \rangle} v_1 + \cdots + \overline{\langle T(v_n), v_j \rangle} v_n \ .\end{aligned}$$

This means the (i, j)-th entry of $[T^*]_B^B$ is $\overline{\langle T(v_i), v_j \rangle}$.

On the other hand, we can write

$$T(v_j) = \langle T(v_j), v_1 \rangle v_1 + \cdots + \langle T(v_j), v_n \rangle v_n \ ,$$

so the (i, j)-th entry of $[T]_B^B$ is $\langle T(v_j), v_i \rangle$. □

Many properties of linear transformations are defined in terms of their adjoints.

Definition 0.5. *Let V be an inner product space and $T : V \to V$ be linear. Then T is*

- *self-adjoint if $T = T^*$;*
- *skew-adjoint if $T = -T^*$;*
- *unitary if T is invertible and $T^{-1} = T^*$;*
- *normal if $TT^* = T^*T$.*