Lecture 27

Wednesday, March 11, 2020

Last time we defined the adjoint operator of a linear operator on an inner product space. Specifically, if V is an inner product space and $f: V \to V$ is a linear map then the adjoint operator $f^*: V \to V$ is defined as a linear map satisfying

$$(f(u), v) = (u, f^*(v)) \qquad \forall\, u, v \in V.$$

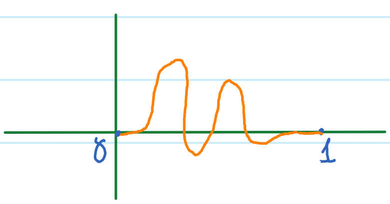
The adjoint of an operator is a "dual" or "companion" operator of its: applying f on the first argument of the inner product is the same as applying f^* on the second argument of the inner product. Practices like this is quite common in math. For example, when taking the integration of uv', one can place the derivative upon u instead of v at the cost of switching the sign and some simple extra terms:

$$\int_a^b uv'\,dx = u(b)v(b) - u(a)v(a) + \int_a^b -u'v\,dx.$$

This was known as Integration by Parts.

<u>Ex</u>: Consider a vector space

$V = \{u : [0,1] \to \mathbb{R}$ smooth, u is equal to zero near 0 and $1\}$.



This is an inner product with the inner product given by

$$(u, v) = \int_0^1 u(x)v(x)\,dx.$$

The differential operator $D: V \to V$, $D(u) = u'$ is a linear map on V. What is its adjoint operator D^*?

By the definition of adjoint operator,

$$(D(u), v) = (u, D^*(v)) \qquad \forall\, u, v \in V.$$

This equation can be rewritten as

$$\int_0^1 u'v\,dx = \int_0^1 u\,D^*(v)\,dx. \qquad (*)$$

$$\text{LHS} = \int_0^1 u'v\,dx = \underbrace{u(x)\,v(x)\Big|_0^1}_{\substack{=0 \text{ because} \\ u \text{ and } v \text{ vanish} \\ \text{at the endpoints}}} - \int_0^1 u v'\,dx = \int_0^1 u(-v')\,dx.$$

$$\text{RHS} = \int_0^1 u\, D^*(v)\,dx.$$

For LHS to be equal to RHS, we must have $D^*(v) = -v'$!

Ex: Let $f: F^n \to F^n$ (where $F = \mathbb{R}$ or $\mathbb{C}$) be a linear map given by $f(x) = Ax$, where A is an $n \times n$ matrix. Find f^*.

We know that f^* is also a map from F^n to F^n. Thus, it is given by matrix multiplication $f^*(x) = Bx$ where B is an $n\times n$ matrix to be found. By the definition of adjoint operators,

$$(f(x), y) = (x, f^*(y)) \qquad \forall\ x, y \in \mathbb{R}^n.$$

In other words,

$$(Ax, y) = (x, By) \qquad \forall\ x, y \in \mathbb{R}^n. \qquad (**)$$

Recall that the inner product on F^n is given by

$$(u, v) = \left(\begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}, \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} \right) = u_1\bar{v}_1 + u_2\bar{v}_2 + \dots + u_n\bar{v}_n = v^* u.$$

Therefore, $(**)$ can be written as

$$y^* A x = (By)^* x.$$

Recall that the matrix $C^* = \overline{C^T}$ is obtained by first taking the transpose of C, then taking the complex conjugate of each entry. C^* is called the conjugate transpose (also called adjoint) of matrix C.

There are two useful propertres of conjugate transpose of a matrix which are quite easy to verify.

$$(C^*)^* = C,$$
$$(CD)^* = D^* C^*.$$

We see that conjugate transpose behaves like taking inverse or taking transpose. With these properties, the equation in orange can be rewritten as

$$y^* A x = y^* B^* x \qquad \forall x, y \in F^n.$$

For this to be true, we must have $A = B^*$. In other word, $B = A^*$. We conclude that $f^*(x) = A^* x$.

In this case, the identity $(f(x), y) = (x, f^*(y))$ becomes

$$(Ax, y) = (x, A^* y) \qquad \forall x, y \in F^n,$$

which is a useful identity.

Some terminology:

- If $f^* = f^{-1}$ then f is said to be unitary. This definition is equivalent to the definition we gave earlier: $(f(u), f(v)) = (u, v)$ for all $u, v \in V$.
- If $f^* = f$ then f is said to be Hermittian or self-adjoint.
- If $f^* = -f$ then f is said to be skew-Hermittian.

People also use similar terminology for matrices. Let $A \in M_{n\times n}(F)$.

- If $A^* = A^{-1}$ then A is said to be unitary.
- If $A^* = A$ then A is said to be Hermittian or self-adjoint.
- If $A^* = -A$ then A is said to be skew-Hermittian.

* Singular value decomposition:

We know that a square matrix may be diagonalizable. The diagonalizability of a matrix makes certain calculations simple. For example, if we write $A = PDP^{-1}$ then

$$A^n = PD^nP^{-1}$$

for any power n. However, not all square matrices are diagonalizable (although most matrices are).

There is another type of decomposition, namely

$$A = PDQ^* = PDQ^{-1}$$

where P and Q are unitary matrices, and D is a diagonal matrix with real entries. This is known as singular value decomposition. Unlike diagonalization, <u>all</u> matrices (including non-square matrices) have a singular value decomposition.

Singular value decomposition plays an important role in data compression and machine learning. SVD is different from diagonalization in that P and Q are not necessarily the same. If P and Q in the SVD happen to be equal to each other then SVD is also diagonalization.

Given a matrix A, how do we find P, Q, D? In general, Q and D are simple to find. P is a little bit tricky to find in some cases. From the formula

$$A = PDQ^*$$

we take the conjugate transpose of both sides

$$A^* = (PDQ^*)^* = (Q^*)^* D^* P^* = QDP^*.$$

Then

$$A^*A = (QDP^*)(PDQ^*) = QD^2Q^* = QD^2Q^{-1}.$$

This is exactly the diagonalization of A^*A. In other words, when we diagonalize matrix A^*A and find

$$A^*A = NEN^{-1}$$

then $Q = N$ and $D^2 = E$. If

$$E = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix}$$

then

$$D = \begin{bmatrix} \sqrt{\lambda_1} & & & \\ & \sqrt{\lambda_2} & & \\ & & \ddots & \\ & & & \sqrt{\lambda_n} \end{bmatrix}.$$

The values $\sigma_k = \sqrt{\lambda_k}$ are called singular values of matrix A.

Thus, Q and D are found by diagonalizing matrix A^*A. This is the procedure one is familiar with in Linear Algebra I. Now how do we find P? From the equation

$$A = PDQ^* = PDQ^{-1}$$

we have

$$AQ = PD.$$

If D is invertible (i.e. all the entries on the diagonal are nonzero) then

$$P = AQD^{-1}.$$

If D is not invertible then the procedure is a little more involved.

<u>Ex</u>: Find a singular value decomposition of matrix

$$A = \begin{bmatrix} 2 & 2 \\ -1 & 1 \end{bmatrix}.$$

First, we compute A^*A.

$$A^*A = \begin{bmatrix} 2 & -1 \\ 2 & 1 \end{bmatrix}\begin{bmatrix} 2 & 2 \\ -1 & 1 \end{bmatrix} = \begin{bmatrix} 5 & 3 \\ 3 & 5 \end{bmatrix}$$

This matrix has two eigenvalues $\lambda_1 = 8$ and $\lambda_2 = 2$. Corresponding eigenvectors are $v_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ and $v_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$.

To form matrix Q from v_1 and v_2, one needs to rescale v_1 and v_2 so that their norms are equal to 1.

$$v_1' = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad v_2' = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ -1 \end{bmatrix}.$$

Thus, $$Q = \begin{bmatrix} | & | \\ v_1' & v_2' \\ | & | \end{bmatrix} = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}.$$

Matrix D is $$D = \begin{bmatrix} \sqrt{\lambda_1} & 0 \\ 0 & \sqrt{\lambda_2} \end{bmatrix} = \begin{bmatrix} \sqrt{8} & 0 \\ 0 & \sqrt{2} \end{bmatrix}.$$

This happens to be an invertible matrix, so P is simple to find:

$$P = AQD^{-1} = \begin{bmatrix} 2 & 2 \\ -1 & 1 \end{bmatrix} \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} \sqrt{8} & 0 \\ 0 & \sqrt{2} \end{bmatrix}^{-1}$$

$$= \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}.$$