640:350: LINEAR ALGEBRA HIGHLIGHTS OF LECTURES FOR WEEK 13

Sections to be covered: part of 6.3 (definition of adjoint operator, **Theorem 6.10**, its **Corollary**, and **Theorem 6.11**), part of 6.4 (definition of self-adjoint operators, of Hermitian matrices, **Theorem 6.17**), part of 6.5 (definition of unitary/orthogonal operators/matrices, **Theorem 6.18**, **Theorem 6.23** and applications to conic sections), part of 6.6 (definition and property of orthogonal projection).

Much of what is done in **6.3** $-$ **6.6** is a conceptualization and generalization of what one learns about real symmetric and orthogonal matrices in elementary linear algebra courses. Let's first recall a few key facts about these two objects.

Orthogonal matrices: An $n \times n$ matrix A with real entries is called an orthogonal matrix if its columns form an orthonormal basis of $\mathbb{R}^n$. The following are equivalent descriptions for A to be an orthogonal matrix.

(i) $A^tA = I_{n\times n}$.
(ii) A^{-1} exists and equals A^t.
(iii) The rows of A form an orthonormal basis of $\mathbb{R}^n$
(iv) $||A\mathbf{x}|| = ||\mathbf{x}||$ for any vector $\mathbf{x} \in \mathbb{R}^n$.
(v) $A\mathbf{x} \cdot A\mathbf{y} = \mathbf{x} \cdot \mathbf{y}$ for any pairs of vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, where $\cdot$ refers to the standard dot product on $\mathbb{R}^n$.
(vi) If $\{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$ is an orthonormal basis of $\mathbb{R}^n$, then $\{A\mathbf{v}_1, \cdots, A\mathbf{v}_n\}$ is an orthonormal basis of $\mathbb{R}^n$.

Real symmetric matrices: An $n\times n$ matrix A with real entries is called symmetric if $A^t = A$. The following are known about the eigenvalues and eigenspaces of a real symmetric matrix A.

(i) All the eigenvalues of A are real.
(ii) If $\mathbf{x}$ is an eigenvector of A corresponding to eigenvalue λ and $\mathbf{y}$ is an eigenvector of A corresponding to eigenvalue μ, with $\mu \neq \lambda$, then $\mathbf{x} \cdot \mathbf{y} = 0$, i.e., $\mathbf{x} \perp \mathbf{y}$.
(iii) For each eigenvalue λ of A, its eigenspace E_λ equals its generalized eigenspace K_λ.
(iv) There is an orthonormal basis of $\mathbb{R}^n$ consisting of eigenvectors of A.
(v) There is an $n \times n$ orthogonal matrix P such that $P^tAP = P^{-1}AP$ is diagonal.

For an orthogonal matrix A, we are particularly interested in properties (iv) and (v). (iv) says that $L_A : \mathbb{R}^n \mapsto \mathbb{R}^n$ (defined by $L_A(\mathbf{x}) = A\mathbf{x}$) preserves the length (norm) of all vectors, while (v) says that L_A preseves inner product between any vectors. To see how (v) relates to (i), one views $\mathbf{x}$ and $\mathbf{y}$ as column vectors, then

$$\mathbf{x} \cdot \mathbf{y} = \mathbf{x}^t\mathbf{y} = \mathbf{y}^t\mathbf{x}.$$

Thus

$$A\mathbf{x} \cdot A\mathbf{y} = (A\mathbf{y})^tA\mathbf{x} = \mathbf{y}^tA^tA\mathbf{x} = A^tA\mathbf{x} \cdot \mathbf{y}. \tag{1}$$

In order for $A\mathbf{x} \cdot A\mathbf{y} = \mathbf{x} \cdot \mathbf{y}$ for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, we first arbitrarily fix an $\mathbf{x} \in \mathbb{R}^n$ and allow $\mathbf{y}$ vary arbitrarily. Then by (1) and property (e) of **Theorem 6.1** on p. 333, we must have $A^tA\mathbf{x} = \mathbf{x}$. Since this relation holds for every $\mathbf{x} \in \mathbb{R}^n$, by taking $\mathbf{x}$ to be the standard basis vectors $\mathbf{e}_1, \cdots, \mathbf{e}_n$, we see that column j of A^tA must match $\mathbf{e}_j$ for $j = 1, \cdots, n$, which implies that $A^tA = I_{n\times n}$.

Property (ii) for a real symmetric matrix A also follows from this line of argument: Since $A\mathbf{x} = \lambda\mathbf{x}$ and $A\mathbf{y} = \mu\mathbf{y}$, we find

$$\lambda\mathbf{x}\cdot\mathbf{y} = A\mathbf{x}\cdot\mathbf{y} = (A\mathbf{x})^t\mathbf{y} = \mathbf{x}^tA^t\mathbf{y} = \mathbf{x}\cdot A^t\mathbf{y} = \mathbf{x}\cdot A\mathbf{y} = \mathbf{x}\cdot(\mu\mathbf{y}) = \mu\mathbf{x}\cdot\mathbf{y}.$$

Since $\lambda \neq \mu$, we must have $\mathbf{x}\cdot\mathbf{y} = 0$.

Property (iii) for a real symmetric matrix A can be argued as follows. It suffices to prove that $N((A-\lambda I)^2) \subset N(A-\lambda I)$. Take a vector $\mathbf{x} \in N((A-\lambda I)^2)$. Then $(A-\lambda I)^2\mathbf{x} = 0$, which implies $(A-\lambda I)^t(A-\lambda I)\mathbf{x} = 0$, since $(A-\lambda I)^t = A - \lambda I$ for a real symmetric A. Then by the same computations used in proving (1) applied to $A - \lambda I$ now,

$$(A-\lambda I)^t(A-\lambda I)\mathbf{x}\cdot\mathbf{x} = (A-\lambda I)\mathbf{x}\cdot(A-\lambda I)\mathbf{x} = 0,$$

from which we conclude that $(A-\lambda I)\mathbf{x} = 0$, namely, $\mathbf{x} \in N(A-\lambda I)$ — this is the same proof as for Lemm 2 on p. 362, replacing the A there by $A - \lambda I$.

Sections **6.3 − 6.6** generalize the notions and ideas here to a general inner product space. Let's first see how to generalize these notions to $\mathbb{C}^n$ with the standard inner product: $\langle\mathbf{x},\mathbf{y}\rangle = \mathbf{x}^t\bar{\mathbf{y}} = \bar{\mathbf{y}}^t\mathbf{x}$. Then, we still have

$$\langle A\mathbf{x},\mathbf{y}\rangle = (A\mathbf{x})^t\bar{\mathbf{y}} = \mathbf{x}^tA^t\bar{\mathbf{y}} = \mathbf{x}^t\overline{(\overline{A}^t\mathbf{y})} = \mathbf{x}^t\overline{A^*\mathbf{y}} = \langle\mathbf{x},A^*\mathbf{y}\rangle. \tag{2}$$

This is the defining property for A^*.

From this we see that, if $\langle A\mathbf{x},A\mathbf{y}\rangle = \langle\mathbf{x},\mathbf{y}\rangle$ for all $\mathbf{x},\mathbf{y} \in \mathbb{C}^n$, then $\langle\mathbf{x},A^*A\mathbf{y}\rangle = \langle\mathbf{x},\mathbf{y}\rangle$ for all $\mathbf{x},\mathbf{y} \in \mathbb{C}^n$, from which we imply that $A^*A = I_{n\times n}$. Thus the generalization of an orthogonal $n\times n$ real valued matrix is a complex valued matrix A satisfying $A^*A = I_{n\times n}$. Such a matrix is called a unitary matrix, in analogy with complex numbers z having unit norm when satisfing $z^*z = 1$. Checking the algebra, one finds that $A^*A = I_{n\times n}$ is equivalent to saying that the columns of A form an orthonormal basis for $\mathbb{C}^n$, and the other properties generalize in a straight forward fashion. For instance, (ii) changes to $A^{-1} = A^*$. See **Theorem 6.18** for more details.

An $n\times n$ matrix A is called self-adjoint (or Hermitian) if $A^* = A$. From the defining property (2) of A^*, this means

$$\langle A\mathbf{x},\mathbf{y}\rangle = \langle\mathbf{x},A\mathbf{y}\rangle$$

for all $\mathbf{x},\mathbf{y} \in \mathbb{C}^n$. For a Hermitian matrix A, properties of real symmetric matrices generalize over in a straight forward fashion, simply changing the orthogonal matrix P in (v) into a unitary matrix P. For instance, in proving (iii) for a Hermitian matrix A, let λ be an eigenvalue of A, then λ is real by (i), and $(A-\lambda I)^* = A - \overline{\lambda}I = A - \lambda I$, since $A^* = A$, and $\overline{\lambda} = \lambda$. Thus

$$\langle(A-\lambda I)\mathbf{x},(A-\lambda I)\mathbf{y}\rangle = \langle\mathbf{x},(A-\lambda I)^*(A-\lambda I)\mathbf{y}\rangle = \langle\mathbf{x},(A-\lambda I)^2\mathbf{y}\rangle.$$

So if $\mathbf{y} \in N((A-\lambda I)^2)$, we would have

$$\langle(A-\lambda I)\mathbf{x},(A-\lambda I)\mathbf{y}\rangle = 0$$

for all $\mathbf{x} \in V$. In particular, taking $\mathbf{x} = \mathbf{y}$, we find

$$\langle(A-\lambda I)\mathbf{y},(A-\lambda I)\mathbf{y}\rangle = 0,$$

which implies that $(A-\lambda I)\mathbf{y} = \mathbf{0}$, i.e., $\mathbf{y} \in N(A-\lambda I)$.

Our next task is to generalize these notions to a general inner product space over $\mathbb{C}$. The questions you should keep in mind are: (a) how to define the notion of an adjoint operator,

(ii) how to define the notion of a unitary operator, and (c) how to define the notion of a self-adjoint operator.

The discussions above suggest that for each linear operator $T : V \mapsto V$ of an inner product space V, the adjoint $T^* : V \mapsto V$ should be defined similar to property (2):

$$\langle T(\mathbf{x}), \mathbf{y} \rangle = \langle \mathbf{x}, T^*(\mathbf{y}) \rangle, \quad \text{for all } \mathbf{x}, \mathbf{y} \in V.$$

Theorems 6.8, 6.9 provide the proper argument to make such a definition possible. **Theorem 6.10** says that the matrix representation $[T^*]_\beta$ of T^* in *an orthonormal basis* β of V relates to $[T]_\beta$ through the relation $[T^*]_\beta = [T]_\beta^*$. **Theorem 6.11** says that the $*$ operation on linear operators is very much like that of complex conjugation operation on complex numbers.