Math 115A - Week 10
Textbook sections: 3.1-5.1
Topics covered:

- Linear functionals
- Adjoints of linear operators
- Self-adjoint operators
- Normal operators
- Stuff about the final

* * * * *

Linear functionals

- Let F be either the real or complex numbers, and let V, W be vector spaces over the field of scalars F. We know what a linear transformation T from V to W is; it is a transformation that takes as input a vector v in V and returns a vector Tv in W, which preserves addition $T(v+v') = Tv + Tv'$ and scalar multiplication $T(cv) = cTv$.

- We now look at some special types of linear transformation, where the input space V or the output space W is very small. We first look at what happens when the input space is just F, the field of scalars.

- **Example** The linear transformation $T : \mathbf{R} \to \mathbf{R}^3$ defined by $Tc := (3c, 4c, 5c)$ is a linear transformation from the field of scalars $\mathbf{R}$ to a vector space $\mathbf{R}^3$.

- Note that the above example can be written as $Tc := cw$, where w is the vector $(3, 4, 5)$ in $\mathbf{R}^3$. The following lemma says, in fact, that all linear transformations from the field of scalars to another vector space are of this form:

- **Lemma 1.** Let $T : F \to W$ be a linear transformation from F to W. Then there is a vector $w \in W$ such that $Tc = cw$ for all $c \in F$.

- **Proof** Since $c = c1$, we have $Tc = T(c1) = c(T1)$. So if we set $w := T1$, then we have $Tc = cw$ for all $c \in F$. $\square$

- Now we look at what happens when the *output* space is the field of scalars.

- **Definition** A *linear functional* on a vector space V is a linear transformation $T : V \to F$ from V to the field of scalars F.

- Thus linear functionals are in some sense the "opposite" of vectors: they "eat" a vector as input and spit out a scalar as output. (They are sometimes called *covectors* or *dual vectors* for this reason; sometimes physicists call them *axial vectors*. Another name used is *1-forms*. In quantum mechanics, one sometimes uses Dirac's "braket" notation, in which vectors are called "kets" and covectors are called "bras").

- **Example 1.** The linear transformation $T : \mathbf{R}^3 \to \mathbf{R}$ defined by $T(x, y, z) := 3x + 4y + 5z$ is a linear functional on $\mathbf{R}^3$. Another example is *altitude*: the linear transformation $A : \mathbf{R}^3 \to \mathbf{R}$ defined by $A(x, y, z) := z$; this takes a vector in three-dimensional space as input and returns its altitude (the z co-ordinate).

- **Example 2 (integration as a linear functional).** The linear transformation $I : C([0, 1]; \mathbf{R}) \to \mathbf{R}$ defined by $If := \int_0^1 f(x)\, dx$ is a linear functional, for instance $I(x^2) = 1/3$.

- **Example 3 (evaluation as a linear functional).** The linear transformation $E : C([0, 1]; \mathbf{R}) \to \mathbf{R}$ defined by $Ef = f(0)$ is a linear functional, thu for instance $E(x^2) = 0$, and $E(e^x) = 1$.

- **Example 4.** Let V be any inner product space, and let w be any vector in V. Then the linear transformation $T : V \to F$ defined by $Tv := \langle v, w\rangle$ is a linear functional on V (this is because inner product is linear in the first variable v). For instance, the linear functional $T(x, y, z) := 3x + 4y + 5z$ in Example 1 is of this type, since $T(x, y, z) = \langle (x, y, z), (3, 4, 5)\rangle$; similarly the altitude function can be written in this form, as $A(x, y, z) = \langle (x, y, z), (1, 0, 0)\rangle$. Also, the integration functional I in Example 2 is also of this form, since $If = \langle f, 1\rangle$.

(As it turns out, the evaluation function E from Example 3 is not of this form, at least on $C([0,1];\mathbf{R})$; but see below.)

- It turns out that on an *finite-dimensional* inner product space V, every linear functional is of the form given in the previous example:

- **Riesz representation theorem.** Let V be a finite-dimensional inner product space, and let $T : V \to F$ be a linear functional on V. Then there is a vector $w \in W$ such that $Tv = \langle v, w \rangle$ for all $v \in V$.

- **Proof.** Let's say that V is n-dimensional. By the Gram-Schmidt orthogonalization process we can find an orthonormal basis $v_1, v_2, \ldots, v_n$ of V. Let v be any vector in V. From the previous week's notes we have the formula

$$v = \langle v, v_1 \rangle v_1 + \ldots + \langle v, v_n \rangle v_n.$$

Applying T to both sides, we obtain

$$Tv = \langle v, v_1 \rangle T v_1 + \ldots + \langle v, v_n \rangle T v_n.$$

Since $Tv_1, \ldots, Tv_n$ are all scalars, and $\langle v, w \rangle c = \langle v, \overline{c} w$ for any scalar c, and we thus have

$$Tv = \langle v, \overline{Tv_1} v_1 + \ldots \overline{Tv_n} v_n \rangle.$$

Thus if we let $w \in V$ be the vector

$$w := \overline{Tv_1} v_1 + \ldots \overline{Tv_n} v_n$$

then we have $Tv = \langle v, w \rangle$ for all $v \in V$, as desired. $\square$

- (Actually, this is only the Riesz representation theorem for finite dimensional spaces. There are more general Riesz representation theorems for such infinite-dimensional spaces as $C([0,1];\mathbf{R})$, but they are beyond the scope of this course).

- **Example** Consider the linear functional $T : \mathbf{C}^3 \to \mathbf{C}$ defined by $T(x, y, z) := 3x + iy + 5z$. From the Riesz representation theorem we know that there must be some vector $w \in \mathbf{C}^3$ such that $Tv := \langle v, w \rangle$

for all $v \in \mathbf{C}^3$. In this case we can see what w is by inspection, but let us pretend that we are unable to see this, and instead use the formula in the proof of the Riesz representation theorem. Namely, we know that

$$w := \overline{Tv_1} v_1 + \ldots \overline{Tv_n} v_n$$

whenever $v_1, \ldots, v_n$ is an orthonormal basis for $\mathbf{C}^3$. Thus, using the standard basis $(1,0,0), (0,1,0), (0,0,1)$, we obtain

$$w := \overline{T(1,0,0)}(1,0,0) + \overline{T(0,1,0)}(0,1,0) + \overline{T(0,0,1)}(0,0,1)$$
$$= \overline{3}(1,0,0) + \overline{i}(0,1,0) + \overline{5}(0,0,1) = (3,-i,5).$$

Thus $Tv = \langle v, (3,-i,5) \rangle$, which one can easily check is consistent with our definition of T.

- More generally, we see that any linear functional $T : F^n \to F$ (where $F = \mathbf{R}$ or $\mathbf{C}$) can be written in the form $Tv := \langle v, w \rangle$, where w is the vector

  $$w = (\overline{Te_1}, \overline{Te_2}, \ldots, \overline{Te_n}),$$

  and $e_1, \ldots, e_n$ is the standard basis for F^n. (i.e. the first component of w is $\overline{Te_1}$, etc. For instance, in the previous example $Te_1 = T(1,0,0) = 3$, so the first component of w is $\overline{3} = 3$.

- **Example** Let $P_2(\mathbf{R})$ be the polynomials of degree at most 2, with the inner product

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

  Let $E : P_2(\mathbf{R}) \to \mathbf{R}$ be the evaluation function $E(f) := f(0)$, for instance $E(x^2 + 2x + 3) = 3$. From the Riesz representation theorem we know that $E(f) = \langle f, w \rangle$ for some $w \in P_2(\mathbf{R})$; we now find what this w is. We first find an orthonormal basis for $P_2(\mathbf{R})$. From last week's notes, we know that

  $$v_1 := \frac{1}{\sqrt{2}}; v_2 := \frac{\sqrt{3}}{\sqrt{2}} x; v_3 := \frac{\sqrt{45}}{\sqrt{8}} (x^2 - \frac{1}{3})$$

  is an orthonormal basis for $P_2(\mathbf{R})$. Thus we can compute w using the formula

  $$w = \overline{Tv_1} v_1 + \overline{Tv_2} v_2 + \overline{Tv_3} v_3$$

from the proof of the Riesz representation theorem. Since $Tv_1 = \frac{1}{\sqrt{2}}$, $Tv_2 = 0$, and $Tv_3 = \frac{\sqrt{45}}{\sqrt{8}}(-\frac{1}{3})$, we thus have

$$w = \frac{1}{\sqrt{2}}\frac{1}{\sqrt{2}} + \frac{\sqrt{45}}{\sqrt{8}}(-\frac{1}{3})\frac{\sqrt{45}}{\sqrt{8}}(x^2 - \frac{1}{3})$$

which simplifies to

$$w = \frac{1}{2} - \frac{5}{24}(3x^2 - 1) = \frac{17}{24} - \frac{5}{8}x^2.$$

- It may seem that the vector w that is obtained by the Riesz representation theorem would depend on which orthonormal basis $v_1, \ldots, v_n$ one chooses for V. But it turns out that this is not the case:

- **Lemma 2.** Let $T : V \to \mathbf{R}$ be a linear functional on an inner product space V. Then there can be at most one vector $w \in V$ with the property that $Tv = \langle v, w \rangle$ for all $v \in V$.

- **Proof.** Suppose for contradiction that there were at least two different vectors w, w' in V such that $Tv = \langle v, w \rangle$ and $Tv = \langle v, w' \rangle$ for all $v \in V$. Then we have

  $$\langle v, w - w' \rangle = \langle v, w \rangle - \langle v, w' \rangle = Tv - Tv = 0$$

  for all $v \in V$. In particular, if we apply this identity to the vector $v := w - w'$ we obtain

  $$\|w - w'\|^2 = \langle w - w', w - w' \rangle = 0$$

  which implies that $w - w' = 0$, so that w and w' are not different after all. This contradiction shows that there could only have been one such vector w to begin with, as desired. $\square$

- Another way to view Lemma 2 is the following: if $\langle v, w \rangle = \langle v, w' \rangle$ for all $v \in V$, then w and w' must be equal. (If you like, this is sort of like being able to "cancel" v from both sides of an identity involving an inner product, provided that you know the identity holds for *all* v).

* * * * *

Adjoints

- The Riesz representation theorem allows us to turn linear functionals $T : V \to \mathbf{R}$ into vectors $w \in V$, if V is a finite-dimensional inner product space. This leads us to a useful notion, that of the *adjoint* of a linear operator.

- Let $T : V \to W$ be a linear transformation from one inner product space to another. Then for every vector $w \in W$, we can define a linear functional $T_w : V \to \mathbf{R}$ by the formula

$$T_w v := \langle Tv, w \rangle.$$

- **Example.** If $T : \mathbf{R}^3 \to \mathbf{R}^2$ is the linear transformation

$$T(x, y, z) := (x + 2y + 3z, 4x + 5y + 6z)$$

  and w was the vector $(10, 1) \in \mathbf{R}^2$, then $T_w : \mathbf{R}^3 \to \mathbf{R}$ would be the linear functional

$$T_w(x, y, z) = \langle (x + 2y + 3z, 4x + 5y + 6z), (10, 1) \rangle = 14x + 25y + 36z.$$

- One can easily check that T_w is indeed a linear functional on V:

$$T_w(v+v') = \langle T(v+v'), w \rangle = \langle Tv+Tv', w \rangle = \langle Tv, w \rangle + \langle Tv', w \rangle = T_w v + T_w v'$$

$$T_w(cv) = \langle T(cv), w \rangle = \langle cTv, w \rangle = c\langle Tv, w \rangle = cT_w v.$$

- By the Riesz representation theorem, there must be a vector, called $T^* w \in V$, such that $T_w v = \langle v, T^* w \rangle$ for all $v \in V$, or in other words that

$$\langle Tv, w \rangle = \langle v, T^* w \rangle$$

  for all $w \in W$ and $v \in V$; this is probably the most basic property of T^*. Note that by Lemma 2, there can only be one possible value for $T^* w$ for each w.

- **Example** Continuing the previous example, we see that

$$T_w(x, y, z) = \langle (x, y, z), (14, 25, 36) \rangle$$

and hence by Lemma 2, the only possible choice for $T^* w$ is

$$T^*(10, 1) = T^* w = (14, 25, 36).$$

- Note that while T is a transformation that turns a vector v in V to a vector Tv in W, T^* does the opposite, starting with a vector w in W as input and returning a vector $T^* w$ in V as output. This seems similar to how an inverse T^{-1} of T would work, but it is important to emphasize that T^* is *not* the inverse of T, and it makes sense even when T is not invertible.

- We refer to $T^* : W \to V$ as the *adjoint* of T. Thus when we move an operator T from one side of an inner product to another, we have to replace it with its adjoint. This is similar to how when one moves a scalar from one side of an inner product to another, you have to replace it by its complex conjugate: $\langle cv, w \rangle = \langle v, \bar{c} w \rangle$. Thus the adjoint is like the complex conjugate, but for linear transformations rather than for scalars.

- **Lemma 3.** If $T : V \to W$ is a linear transformation, then its adjoint $T^* : W \to V$ is also a linear transformation.

- **Proof.** We have to prove that $T^*(w + w') = T^* w + T^* w'$ and $T^*(cw) = cT^* w$ for all $w, w' \in W$ and scalars c.

- First we prove that $T^*(w + w') = T^* w + T^* w'$. By definition of T^*, we have

$$\langle v, T^*(w + w') \rangle = \langle Tv, w + w' \rangle$$

for all $v \in V$. But

$$\langle Tv, w + w' \rangle = \langle Tv, w \rangle + \langle Tv, w' \rangle = \langle v, T^* w \rangle + \langle v, T^* w' \rangle = \langle v, T^* w + T^* w' \rangle.$$

Thus we have

$$\langle v, T^*(w + w') \rangle = \langle v, T^* w + T^* w' \rangle$$

for all $v \in V$. By Lemma 2, we must therefore have $T^* w + T^* w' = T^*(w + w')$ as desired.

- Now we show that $T^*(cw) = cT^*w$. We have

$$\langle v, T^*(cw)\rangle = \langle Tv, cw\rangle = \overline{c}\langle Tv, w\rangle = \overline{c}\langle v, T^*w\rangle = \langle v, cT^*w\rangle$$

for all $v \in V$. By Lemma 2, we thus have $T^*(cw) = cT^*w$ as desired. □

- **Example** Let us continue our previous example of the linear transformation $T : \mathbf{R}^3 \to \mathbf{R}^2$ defined by

$$T(x, y, z) := (x + 2y + 3z, 4x + 5y + 6z).$$

Let us work out what $T^* : \mathbf{R}^2 \to \mathbf{R}^3$ is. Let (a, b) be any vector in $\mathbf{R}^2$. Then we have

$$\langle T(x, y, z), (a, b)\rangle = \langle (x, y, z), T^*(a, b)\rangle$$

for all $(x, y, z) \in \mathbf{R}^3$. The left-hand side is

$$\langle (x + 2y + 3z, 4x + 5y + 6z), (a, b)\rangle = a(x + 2y + 3z) + b(4x + 5y + 6z)$$

$$= (a+4b)x+(2a+5b)y+(3a+6b)z = \langle (x, y, z), (a+4b, 2a+5b, 3a+6b)\rangle.$$

Thus we have

$$\langle (x, y, z), (a + 4b, 2a + 5b, 3a + 6b)\rangle = \langle (x, y, z), T^*(a, b)\rangle$$

for all x, y, z; by Lemma 2, this implies that

$$T^*(a, b) = (a + 4b, 2a + 5b, 3a + 6b).$$

- This example was rather tedious to compute. However, things become easier with the aid of orthonormal bases. Recall (from Corollary 7 of last week's notes) that if v is a vector in V and $\beta := (v_1, \ldots, v_n)$ is an orthonormal basis of V, then the column vector $[v]^\beta$ is given by

$$[v]^\beta = \begin{pmatrix} \langle v, v_1\rangle \\ \vdots \\ \langle v, v_n\rangle \end{pmatrix}.$$

Thus the i^{th} row entry of $[v]^\beta$ is just $\langle v, v_i\rangle$.

- Now suppose that $T : V \to W$ is a linear transformation, and $\beta := (v_1, \ldots, v_n)$ is an orthonormal basis of V and $\gamma := (w_1, \ldots, w_m)$ is an orthonormal basis of W. Then $[T]^\gamma_\beta$ is a matrix with m rows and n columns, whose j^{th} column is given by $[Tv_j]^\beta$. In other words, we have

$$[T]^\gamma_\beta = \begin{pmatrix} \langle Tv_1, w_1\rangle & \langle Tv_2, w_1\rangle & \ldots & \langle Tv_n, w_1\rangle \\ \langle Tv_1, w_2\rangle & \langle Tv_2, w_2\rangle & \ldots & \langle Tv_n, w_2\rangle \\ & & \vdots & \\ \langle Tv_1, w_m\rangle & \langle Tv_2, w_m\rangle & \ldots & \langle Tv_n, w_m\rangle \end{pmatrix}.$$

In other words, the entry in the i^{th} row and j^{th} column is $\langle Tv_j, w_i\rangle$.

- We can apply similar reasoning to the linear transformation $T^* : W \to V$. Then $[T^*]^\beta_\gamma$ is a matrix with n rows and m columns, and the entry in the i^{th} row and j^{th} column is $\langle T^* w_j, v_i\rangle$. But

$$\langle T^* w_j, v_i\rangle = \overline{\langle v_i, T^* w_j\rangle} = \overline{\langle Tv_i, w_j\rangle}.$$

Thus, the matrix $[T^*]^\beta_\gamma$ is given by

$$[T^*]^\beta_\gamma = \begin{pmatrix} \overline{\langle Tv_1, w_1\rangle} & \overline{\langle Tv_1, w_2\rangle} & \ldots & \overline{\langle Tv_1, w_m\rangle} \\ \overline{\langle Tv_2, w_1\rangle} & \overline{\langle Tv_2, w_2\rangle} & \ldots & \overline{\langle Tv_2, w_m\rangle} \\ & & \vdots & \\ \overline{\langle Tv_n, w_1\rangle} & \overline{\langle Tv_n, w_2\rangle} & \ldots & \overline{\langle Tv_n, w_m\rangle} \end{pmatrix}.$$

Comparing this with our formula for $[T]^\gamma_\beta$ we see that $[T^*]^\beta_\gamma$ is the adjoint of $[T]^\gamma_\beta$:

- **Theorem 3.** If $T : V \to W$ is a linear transformation, β is an orthonormal basis of V, and γ is an orthonormal basis of W, then

$$[T^*]^\beta_\gamma = ([T]^\gamma_\beta)^\dagger.$$

- **Example** Let us once again take the example of the linear transformation $T : \mathbf{R}^3 \to \mathbf{R}^2$ defined by

$$T(x, y, z) := (x + 2y + 3z, 4x + 5y + 6z).$$

Let $\beta := ((1,0,0),(0,1,0),(0,0,1))$ be the standard basis of $\mathbf{R}^3$, and let $\gamma := ((1,0),(0,1))$ be the standard basis of $\mathbf{R}^2$. Then we have

$$[T]_\beta^\gamma := \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}$$

(why?). On the other hand, if we write the linear transformation

$$T^*(a,b) = (a+4b, 2a+5b, 3a+6b)$$

in matrix form, we see that

$$[T^*]_\gamma^\beta := \begin{pmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{pmatrix},$$

which is the adjoint of $[T]_\beta^\gamma$. (In this example, the field of scalars is real, and so the complex conjugation aspect of the adjoint does not make an appearance.

- The following corollary connects the notion of adjoint of a *linear transformation* with that of adjoint of a *matrix.*

- **Corollary 4.** Let A be an $m \times n$ matrix with either real or complex entries. Then the adjoint of L_A is $L_{A^\dagger}$.

- **Proof.** Let F be the field of scalars that the entries of A lie in. Then L_A is a linear transformation from F^m to F^n, and $L_{A^\dagger}$ is a linear transformation from F^n to F^m. If we let β be the standard basis of F^m and γ be the standard basis of F^n, then by Theorem 3

$$[L_A^*]_\gamma^\beta = ([L_A]_\beta^\gamma)^\dagger = A^\dagger = [L_{A^\dagger}]_\gamma^\beta$$

and hence $L_A^* = L_{A^\dagger}$ as desired. □

- In particular, we see that

$$\langle Av, w\rangle = \langle v, A^\dagger w\rangle$$

for any $m \times n$ matrix A, any column vector v of length n, and any column vector w of length m.

- **Example** Let A be the matrix

$$A = \begin{pmatrix} 1 & i & 0 \\ 0 & 1+i & 3 \end{pmatrix},$$

so that $L_A : \mathbf{C}^3 \to \mathbf{C}^2$ is the linear transformation defined by

$$L_A \begin{pmatrix} z_1 \\ z_2 \\ z_3 \end{pmatrix} = A \begin{pmatrix} z_1 \\ z_2 \\ z_3 \end{pmatrix} = \begin{pmatrix} z_1 + iz_2 \\ (1+i)z_2 + 3z_3 \end{pmatrix}.$$

Then the adjoint of this transformation is given by $L_{A^\dagger}$, where $A^\dagger$ is the adjoint of A:

$$A^\dagger = \begin{pmatrix} 1 & 0 \\ -i & 1-i \\ 0 & 3 \end{pmatrix},$$

so

$$L_{A^\dagger} \begin{pmatrix} w_1 \\ w_2 \end{pmatrix} = A \begin{pmatrix} w_1 \\ w_2 \end{pmatrix} = \begin{pmatrix} w_1 \\ -iw_1 + (1-i)w_2 \\ 3w_3 \end{pmatrix}.$$

- Some basic properties of adjoints. Firstly, the process of taking adjoints is conjugate linear: if $T : V \to W$ and $U : V \to W$ are linear transformations, and c is a scalar, then $(T+U)^* = T^* + U^*$ and $(cT)^* = \overline{c}T^*$. Let's just prove the second claim, as the first is similar (or can be found in the textbook). We look at the expression $\langle v, (cT)^* w\rangle$ for any $v \in V$ and $w \in W$, and compute:

$$\langle v, (cT)^* w\rangle = \langle cTv, w\rangle = c\langle Tv, w\rangle = c\langle v, T^* w\rangle = \langle v, \overline{c}T^* w\rangle.$$

Since this identity is true for all $v \in V$, we thus have (by Lemma 2) that $(cT)^* w = \overline{c}T^* w$ for all $w \in W$, and so $(cT)^* = \overline{c}T^*$ as desired.

- This argument shows a key trick in understanding adjoints: in order to understand a transformation T or its adjoint, it is often a good idea to start by looking at the expression $\langle Tv, w\rangle = \langle v, T^* w\rangle$ and rewrite it in some other way.

- Some other properties, which we leave as exercises: $(T^*)^* = T$ (i.e. if T^* is the adjoint of T, then T is the adjoint of T^*); the adjoint of the identity operator is again the identity; and if $T : V \to W$ and $S : U \to V$ are linear transformations, then $(TS)^* = S^*T^*$. (This last identity can be verified by playing around with $\langle u, S^*T^*w\rangle$ for $u \in U$ and $w \in W$). If T is invertible, we also have $(T^{-1})^* = (T^*)^{-1}$ (i.e. the inverse of the adjoint is the adjoint of the inverse). This can be seen by starting with the identity $TT^{-1} = T^{-1}T = I$ and taking adjoints of all sides.

- Another useful property is that a matrix has the same rank as its adjoint. To see this, recall that the adjoint of a matrix is the conjugate of its transpose. From Lemma 7 of week 6 notes, we know that a matrix has the same rank as its transpose. It is also easy to see that a matrix has the same rank as its conjugate (this is basically because the conjugate of an elementary matrix is again an elementary matrix, and the conjugate of a matrix in row-echelon form is again a matrix in row echelon form.) Combining these two observations we see that the adjoint of a matrix must also have the same rank. From Theorem 3 (and Lemma 9 of week 6 notes) we see therefore that a linear operator from one finite-dimensional inner product space to another has the same rank as its adjoint.

- In a similar vein, if A is a square matrix with determinant d, then A^* will have determinant $\overline{d}$. (We will only sketch a proof of this fact here: first prove it for elementary matrices, and for diagonal matrices. Then to handle the general case, use Proposition 5 from week 6 notes, as well as the identity $(BA)^\dagger = A^\dagger B^\dagger$).

* * * * *

Normal operators

- Recall that in the Week 7 notes we discussed the problem of whether a linear transformation was diagonalizable, i.e. whether it had a basis of eigenvectors. We did not fully resolve this question, and in fact we will not be able to give a truly satisfactory answer to this question until Math 115B. However, there is a special class of linear transformations (aka operators) for which we can give a good answer - *normal* operators.

- **Definition** Let $T : V \to V$ be a linear transformation on V, so that the adjoint $T^* : V \to V$ is another linear transformation on V. We say that T is *normal* if $TT^* = T^*T$.

- **Example 1** Let $T : \mathbf{R}^2 \to \mathbf{R}^2$ be the linear transformation $T(x, y) := (y, -x)$. Then $T^* : \mathbf{R}^2 \to \mathbf{R}^2$ can be computed to be the linear transformation $T^*(x, y) = (-y, x)$ (why?), and so
$$TT^*(x, y) = T(-y, x) = (x, y)$$
and
$$T^*T(x, y) = T^*(y, -x) = (x, y).$$
Thus $TT^*(x, y)$ and $T^*T(x, y)$ agree for all $(x, y) \in \mathbf{R}^2$, which implies that $TT^* = T^*T$. Thus this transformation is normal.

- **Example 2** Let $T : \mathbf{R}^2 \to \mathbf{R}^2$ be the linear transformation $T(x, y) := (0, x)$. Then $T^*(x, y) = (y, 0)$ (why?). So
$$TT^*(x, y) = T(y, 0) = (0, y)$$
and
$$T^*T(x, y) = T^*(0, x) = (x, 0).$$
So in general $TT^*(x, y)$ and $T^*T(x, y)$ are not equal, and so $TT^* \neq T^*T$. Thus this transformation is not normal.

- In analogy to the above definition, we define a square matrix A to be *normal* if $AA^\dagger = A^\dagger A$. For instance, the matrix
$$\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$$
can easily be checked to be normal, while the matrix
$$\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$$
is not. (Why do these two examples correspond to Examples 1 and 2 above?)

- Another example, easily checked: every diagonal matrix is normal.

  From Theorem 3 we have

- **Proposition 5.** Let $T : V \to V$ be a linear transformation on a finite-dimensional inner product space, and let β be an orthonormal basis. Then $T : V \to V$ is normal if and only if the matrix $[T]_\beta^\beta$ is.

- **Proof.** If T is normal, then $TT^* = T^*T$. Now taking matrices with respect to β, we obtain

$$[T]_\beta^\beta [T^*]_\beta^\beta = [T^*]_\beta^\beta [T]_\beta^\beta.$$

  But by Theorem 3, $[T^*]_\beta^\beta$ is the adjoint of $[T]_\beta^\beta$. Thus $[T]_\beta^\beta$ is normal. This proves the "only if" portion of the Proposition; the "if" part follows by reversing the above steps. □

- Normal transformations have several nice properties. First of all, when T is normal then T and T^* will have the same eigenvectors (but slightly different eigenvalues):

- **Lemma 6.** Let $T : V \to V$ be normal, and suppose that $Tv = \lambda v$ for some vector $v \in V$ and some scalar λ. Then $T^*v = \overline{\lambda} v$.

- Warning: the above lemma is only true for normal operators! For other linear transformations, it is quite possible that T and T^* have totally different eigenvectors and eigenvalues.

- **Proof** To show $T^*v = \overline{\lambda} v$, it suffices to show that $\|T^*v - \overline{\lambda} v\| = 0$, which in turn will follow if we can show that

$$\langle T^*v - \overline{\lambda} v, T^*v - \overline{\lambda} v\rangle = 0.$$

  We expand out the left-hand side as

$$\langle T^*v, T^*v\rangle - \langle \overline{\lambda} v, T^*v\rangle - \langle T^*v, \overline{\lambda} v\rangle + \langle \overline{\lambda} v, \overline{\lambda} v\rangle.$$

  Pulling the λs out and swapping the Ts over, this becomes

$$\langle v, TT^*v\rangle - \overline{\lambda}\langle Tv, v\rangle - \lambda\langle v, Tv\rangle + \lambda\overline{\lambda}\langle v, v\rangle.$$

Since T is normal and $Tv = \lambda v$, we have $T^*v = T^*Tv = \lambda T^*v$. Thus we can rewrite this expression as

$$\overline{\lambda}\langle v, T^*v\rangle - \lambda\overline{\lambda}\langle v, v\rangle - \lambda\overline{\lambda}\langle v, v\rangle + \lambda\overline{\lambda}\langle v, v\rangle.$$

But $\langle v, T^*v\rangle = \langle Tv, v\rangle = \lambda\langle v, v\rangle$. If we insert this in the above expression we then see that everything cancels to zero, as desired. □

- **Lemma 7.** Let $T : V \to V$ be normal, and let v_1, v_2 be two eigenvectors of T with *distinct* eigenvalues λ_1, λ_2. Then v_1 and v_2 must be orthogonal.

- (Compare this with Proposition 6 of the Week 8 notes, which merely asserts that these vectors v_1 and v_2 are linearly independent. Again, we caution that this orthogonality of eigenvectors is only true for normal operators.)

- **Proof.** We have $Tv_1 = \lambda_1 v_1$ and $Tv_2 = \lambda_2 v_2$. By Lemma 6 we thus have $T^*v_1 = \overline{\lambda_1}v_1$ and $T^*v_2 = \overline{\lambda_2}v_2$. Thus

$$\lambda_1\langle v_1, v_2\rangle = \langle Tv_1, v_2\rangle = \langle v_1, T^*v_2\rangle = \langle v_1, \overline{\lambda_2}v_2\rangle = \lambda_2\langle v_1, v_2\rangle.$$

Since $\lambda_1 \neq \lambda_2$, this means that $\langle v_1, v_2\rangle = 0$, and so v_1 and v_2 are orthogonal as desired. □

- This lemma tells us that most linear transformations will not be normal, because in general the eigenvectors corresponding to different eigenvalues will not be orthogonal. (Take for instance the matrix involved in the Fibonacci rabbit example).

- In the other direction, if we have an orthonormal basis of eigenvectors, then the transformation must be normal:

- **Lemma 8.** Let $T : V \to V$ be a linear transformation o an inner product space V, and let β be an orthonormal basis which consists entirely of eigenvectors of T. Then T is normal.

- Compare this lemma to Lemma 2 of Week 7 notes, which sais that if you have a basis of eigenvectors (not necessarily orthonormal), then T is diagonalizable.

- **Proof.** From Lemma 2 of Week 7 notes, we know that the matrix $[T]_\beta^\beta$ is diagonal. But all diagonal matrices are normal (why?), and so $[T]_\beta^\beta$ is normal. By Proposition 5 we thus see that T is normal. □

- We now come to an important theorem, that the converse of Lemma 8 is also true:

- **Spectral theorem for normal operators** Let $T : V \to V$ be a normal linear transformation on a complex finite dimensional inner product space V. Then there is an orthonormal basis β consisting entirely of eigenvectors of T. In particular, T is diagonalizable.

- Thus normal linear transformations are precisely those diagonalizable linear transformations which can be diagonalized using orthonormal bases (as opposed to just being plain diagonalizable, using bases which might not be orthonormal).

- There is also a spectral theorem for normal operators on infinite dimensional inner product spaces, but it is beyond the scope of this course.

- **Proof** Let the dimension of V be n. We shall prove this theorem by induction on n.

- First consider the base case $n = 1$. Then one can pick any orthonormal basis β of V (which in this case will just be a single unit vector), and the vector v in this basis will automatically be an eigenvector of T (because in a one-dimensional space every vector will be a scalar multiple of v). So the spectral theorem is trivially true when $n = 1$.

- Now suppose inductively that $n > 1$, and that the theorem has already been proven for dimension $n - 1$. Let $f(\lambda)$ be the characteristic polynomial of T (or of any matrix representation $[T]_\beta^\beta$ of T; recall that any two such matrix representations are similar and thus have the same characteristic polynomial). From the fundamental theorem of algebra, we know that this characteristic polynomial splits over the complex numbers. Hence there must be at least one root of this polynomial, and hence T has at least one (complex) eigenvalue, and hence at least one eigenvector.

- So now let us pick an eigenvector v_1 of T with eigenvalue λ_1, thus $Tv_1 = \lambda_1 v_1$ and $T^* v_1 = \overline{\lambda_1} v_1$ by Lemma 6. We can normalize v_1 to have length 1, so $\|v_1\| = 1$ (remember that if you multiply an eigenvector by a non-zero scalar you still get an eigenvector, so it's safe to normalize eigenvectors). Let $W := \{cv_1 : c \in \mathbf{C}\}$ denote the span of this eigenvector, thus W is a one-dimensional space. Let $W^\perp := \{v \in V : v \perp v_1\}$ denote the orthogonal complement of W; this is thus an $n-1$ dimensional space.

- Now we see what T and T^* do to $W^\perp$. Let w be any vector in $W^\perp$, thus $w \perp v_1$, i.e. $\langle w, v_1 \rangle = 0$. Then

  $$\langle Tw, v_1 \rangle = \langle w, T^* v_1 \rangle = \langle w, \overline{\lambda_1} v_1 \rangle = \lambda_1 \langle w, v_1 \rangle = 0$$

  and similarly

  $$\langle T^* w, v_1 \rangle = \langle w, T v_1 \rangle = \langle w, \lambda_1 v_1 \rangle = \overline{\lambda_1} \langle w, v_1 \rangle = 0.$$

  Thus if $w \in W^\perp$, then Tw and $T^* w$ are also in $W^\perp$. Thus T and T^* are not only linear transformations from V to V, they are also linear transformations from $W^\perp$ to $W^\perp$. Also, we have

  $$\langle Tw, w' \rangle = \langle w, T^* w' \rangle$$

  for all $w, w' \in W^\perp$, because every vector in $W^\perp$ is a vector in V, and we already have this property for vectors in V. Thus T and T^* are still adjoints of each other even after we restrict the vector space from the n-dimensional space V to the $n-1$-dimensional space $W^\perp$.

- We now apply the induction hypothesis, and find that $W^\perp$ enjoys an orthonormal basis of eigenvectors of T. There are $n-1$ such eigenvectors, since $W^\perp$ is $n-1$ dimensional. Now v_1 is normalized and is orthogonal to all the vectors in this basis, since v_1 lies in W and all the other vectors lie in $W^\perp$. Thus if we add v_1 to this basis we get a new collection of n orthonormal vectors, which automatically form a basis by Corollary 5 of Week 9 notes. Each of these vectors is an eigenvector of T, and so we are done. $\square$

- **Example** The linear transformation $T : \mathbf{R}^2 \to \mathbf{R}^2$ defined by $T(x, y) := (y, -x)$ that we discussed earlier is normal, but not diagonalizable (its characteristic polynomial is $\lambda^2 + 1$, which doesn't split over the reals). This does not contradict the spectral theorem because that only concerns complex inner product spaces. If however we consider the complex linear transformation $T : \mathbf{C}^2 \to \mathbf{C}^2$ defined by $T(z, w) := (w, -z)$, then we can find an orthonormal basis of eigenvectors, namely

$$v_1 := \frac{1}{\sqrt{2}}(1, i); \quad v_2 := \frac{1}{\sqrt{2}}(1, -i)$$

(Exercise: cover up the above line and see if you can find these eigenvectors on your own). Indeed, you can check that v_1 and v_2 are orthonormal, and that $Tv_1 = iv_1$ and $Tv_2 = -iv_2$. Thus we can diagonalize T using an orthonormal basis, to become the diagonal matrix $\mathrm{diag}(i, -i)$.

* * * * *

Self-adjoint operators

- To summarize the previous section: in the world of complex inner product spaces, normal linear transformations (aka normal operators) are the best kind of linear transformations: they are not only diagonalizable, but they are diagonalizable using the best kind of basis, namely an orthonormal basis. However, there is a subclass of normal transformations which are even better: the *self-adjoint* transformations.

- **Definition** A linear transformation $T : V \to V$ on a finite-dimensional inner product space V is said to be *self-adjoint* if $T^* = T$, i.e. T is its own adjoint. A square matrix A is said to be *self-adjoint* if $A^\dagger = A$, i.e. A is its own adjoint.

- **Example.** The linear transformation $T : \mathbf{R}^2 \to \mathbf{R}^2$ defined by $T(x, y) := (y, -x)$ is normal, but not self-adjoint, because its adjoint $T^*(x, y) = (-y, x)$ is not the same as T. However, the linear transformation $T : \mathbf{R}^2 \to \mathbf{R}^2$ defined by $T(x, y) = (y, x)$ is self-adjoint, because its adjoint is given by $T^*(x, y) = (y, x)$ (why?), and this is the same as T.

- **Example.** The matrix

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

is normal, but not self-adjoint, because its adjoint

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

is not the same as A. However, the matrix

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

is normal, but not self-adjoint, because its adjoint

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

is the same as A. (Why does this example correspond to the preceding one? It is easy to check, using Proposition 5, that a linear transformation is self-adjoint if and only if its matrix in some orthonormal basis is self-adjoint).

- **Example.** Every real diagonal matrix is self-adjoint, but any other type of diagonal matrix is not (e.g. $\text{diag}(2+i, 4+3i)$ has an adjoint of $\text{diag}(2-i, 4-3i)$ and is hence not self-adjoint, though it is still normal).

- It is clear that all self-adjoint linear transformations are normal, since if $T^* = T$ then T^*T and TT^* are both equal to T^2 and are hence equal to each other. Similarly, every self-adjoint matrix is normal. However, not every normal matrix is self-adjoint, and not every normal linear transformation is self-adjoint; see the above examples.

- A self-adjoint transformation over a complex inner product space is sometimes known as a *Hermitian* transformation. A self-adjoint transformation over a real inner product space is known as a *symmetric* transformation. Similarly, a complex self-adjoint matrix is known as a *Hermitian* matrix, while a real self-adjoint matrix is known as a *symmetric* matrix. (A matrix is *symmetric* if $A^t = A$. When the matrix is real, the transpose A^t is the same as the adjoint, thus self-adjoint and symmetric ahve the same meaning for real matrices, but not for complex matrices).

- **Example** The matrix
$$A := \begin{pmatrix} 0 & i \\ -i & 0 \end{pmatrix}$$
is its own adjoint (why?), and is hence Hermitian, but it is not symmetric, since it is not its own transpose. Note that every real symmetric matrix is automatically Hermitian, because every real matrix is also a complex matrix (with all the imaginary parts equal to 0).

- From the spectral theorem for normal matrices, we know that any Hermitian operator on a complex inner product space has an orthonormal basis of eigenvectors. But we can say a little bit more:

- **Theorem 9** All the eigenvalues of a Hermitian operator are real.

- **Proof.** Let λ be an eigenvalue of a Hermitian operator T, thus $Tv = \lambda v$ for some non-zero eigenvector v. But then by Lemma 6, $T^* v = \overline{\lambda} v$. But since T is Hermitian, $T = T^*$, and hence $\lambda v = \overline{\lambda} v$. Since v is non-zero, this means that $\lambda = \overline{\lambda}$, i.e. λ is real. Thus all the eigenvalues of T are real. □

- A similar line of reasoning shows that all the eigenvalues of a Hermitian matrix are real.

- **Corollary 10.** The characteristic polynomial of a Hermitian matrix splits over the reals.

- **Proof.** We know already from the Fundamental Theorem of Algebra that the characteristic polynomial splits over the complex numbers. But since the matrix is Hermitian, every root of the characteristic polynomial must be real. Thus the polynomial must split over the reals. □

- We can now prove

- **Spectral theorem for self-adjoint operators** Let T be a self-adjoint linear transformation on an inner product space V (which can be either real or complex). Then there is an orthonormal basis of V which consists entirely of eigenvectors of V, with *real* eigenvalues.

- **Proof.** We repeat the proof of the Spectral theorem for normal operators, i.e. we do an induction on the dimension n of the space V. When $n = 1$ the claim is again trivial (and we use the fact that every Lemma 9 to make sure the eigenvalue is real). Now suppose inductively that $n > 1$ and the claim has already been proven for $n - 1$.

  From Corollary 10 we know that T has at least one *real* eigenvalue. Thus we can find a real λ_1 and a non-zero vector v_1 such that $Tv_1 = \lambda_1 v_1$. We can then normalize v_1 to have unit length. We now repeat the rest of the proof of the spectral theorem for normal operators, to obtain the same conclusion except that the eigenvalues are now real. $\square$

- Notice one subtle difference between the spectral theorem for self-adjoint operators and the spectral theorem for normal operators: the spectral theorem for normal operators requires the inner product space to be complex, but the one for self-adjoint operators does not. In particular, every symmetric operator on a real vector space is diagonalizable.

- **Example** The matrix
$$A := \begin{pmatrix} 0 & i \\ -i & 0 \end{pmatrix}$$
  is Hermitian, and thus so is the linear transformation $L_A : \mathbf{C}^2 \to \mathbf{C}^2$, which is given by
$$L_A \begin{pmatrix} z \\ w \end{pmatrix} = \begin{pmatrix} iw \\ -iz \end{pmatrix}.$$
  By the spectral theorem, $\mathbf{C}^2$ must have an orthonormal basis of eigenvectors with real eigenvalues. One such basis is
$$v_1 := \begin{pmatrix} 1/\sqrt{2} \\ i/\sqrt{2} \end{pmatrix}; \quad v_2 := \begin{pmatrix} 1/\sqrt{2} \\ -i/\sqrt{2} \end{pmatrix};$$
  one can verify that v_1 and v_2 are an orthonormal basis for the complex two-dimensional inner product space $\mathbf{C}^2$, and that $L_A v_1 = -v_1$ and $L_A v_2 = +v_2$. Thus L_A can be diagonalized using an orthonormal basis to give the matrix $\mathrm{diag}(+1, -1)$. Note that while the eigen*values* of L_A are real, the eigen*vectors* are still complex. The spectral theorem says nothing as to how real or complex the eigenvectors are (indeed,

in many inner product spaces, such a question does not really make sense).

- Self-adjoint operators are thus the very best of all operators: not only are they diagonalizable, with an orthonormal basis of eigenvectors, the eigenvalues are also real. (Conversely, it is easy to modify Lemma 8 to show that any operator with these properties is necessarily self-adjoint). Fortunately, self-adjoint operators come up all over the place in real life. For instance, in quantum mechanics, almost all the linear transformations one sees there are Hermitian (this is basically because while quantum mechanics uses complex inner product spaces, the quantities we can actually observe in physical reality must be real-valued).

* * * * *

Stuff about the final

- The final is three hours long, and will be held in the usual classroom (MS 5127), on Tuesday, Dec 10, from 8am-11am.

- The final will be eight to ten questions. They will be of varying difficulty and length, and so they will have different point values assigned to them. It is probably best for you to read all the questions at the beginning before deciding which one to do first; it may not necessarily be a good idea to do the questions in order.

- The final will be comprehensive, covering everything from first week to last week. Because we did not have time to cover the Week 10 notes, this means that everything from Week 1 to Week 9 will be covered. Because the midterm already tested the material from Weeks 1-5, there will be more of an emphasis on the Weeks 6-10 material, however it may still be worthwhile to review the Weeks 1-5 material, because much of the Weeks 6-10 material depends indirectly on the Weeks 1-5 material.

- The questions will be more or less evenly split between three types of questions. The first are *computational questions* - in which you have to compute things like orthonormal bases, characteristic polynomials, eigenvectors and eigenvalues, null spaces and ranges, and so forth. It is probably a good idea to review all the computational questions in the homework, midterm, and in the practice exams, and also the examples given in the class notes and in the textbook. (If you find yourself remembering how to do these problems from before, then perhaps you could look at other such questions in the textbook instead). It is probably best if you attempt these questions without looking at any solutions or without too much help from your friends. (You can certainly use your textbook and notes, though, and any index cards you have prepared).

- The second type of question are "Find"-type questions - questions in which you have to find some object (a matrix, or a basis, or a linear transformation) - which satisfies various properties. These type of questions are trickier than computational questions because they often do not fit any pattern or match any previous problem you have seen. They usually require that you have a good understanding of what all the concepts mean (for instance, you'll have extreme difficulty finding an orthonormal basis with various properties if you do not know what it means for a basis to be orthonormal, and how to check if it is). Many of the theorems in the notes may help give you some clues as to how to find these types of objects; you can put these theorems on your index cards. If you don't know what to do, you can start with trial and error - just try to find an object which satisfies some, if not all, of the properties requested, and then try to modify your guess so that more and more of the properties are satisfied, until you have all of the properties that the question asked for. (Thus you should be able to give examples of linear transformations, examples of orthonormal bases, etc.)

- The third type of question are the *short proof* questions - these are often set in some abstract vector space or inner product space, with some objects (linear transformations, vectors, etc.) assumed to have certain properties, and your job is to deduce other properties of these objects. *You will not be asked to re-prove any of the theorems in the textbook or notes* - you can take these theorems as given, without any need to prove them again. These questions are usually considered the trickiest of all, although often once you see how to do them, they are usually quite short, and do not involve much computation. You should go over the proof-type questions in the homework and in the practice exams, preferably without any outside assistance; the proof-type questions in the final will be of a similar nature.

- There are two practice finals on the class web page. Neither of them are completely indicative of what type of questions you will get on the final, because they both contain questions which require material not covered in this course, or are at a higher level of difficulty than what this course expects. However, they both contain questions which are similar to what will be in the actual final. Please read the annotations on the web page next to the hyperlinks to these finals before using them.

- You may bring *two* 5x8 index cards into the exam - the index card that you used for the first midterm, and an additional 5x8 index card. I will distribute these cards in class, or you can bring your own. It is up to you what to put on these cards; for instance, if you have difficulty remembering the exact definition of some concept, or on how to compute a certain quantity, or on what the exact theorems and relationships are between, say, orthonormal bases and inner products, I would write on the card whatever it would take to address these difficulties - this is more likely to be useful than merely copying down everything from the notes and textbook. One thing to try is to go over the homework and practice exams, and make a note of every time you wish you had remembered some concept or theorem or trick when doing a problem. Then put all those notes into your card.

- Remember, you have three hours to finish, and only ten questions. There is not nearly as much time pressure as there is in a 50-minute midterm, and it is in fact common for people to finish early. So take your time, and try to be careful; there is no need to rush things. There will not be much time-consuming computation in the questions in the final, even in the "computational" questions. A minute spent thinking about what a problem is asking, and how best to go about it, can save ten or twenty minutes of wasted effort.

- Good luck!

* * * * *