MATH 411 LINEAR ALGEBRA

Abstract. This is meant to be a study guide emphasizing the main ideas of the course. It is not meant to be a complete list of everything covered or everything you should know. It should also serve as a list of notations, some of which are hard to find in the book.

1. A note on linear algebra terminology

Some words in linear algebra are different than they are for the same concept in more general abstract algebra. This is a list of some of these more specialize terms:

- span — generate
- spanning set — generating set
- linear transformation — homomorphism
- null space — kernel
- vector space over a field — module over a ring (Section 5)
- direct sum — coproduct

2. Definitions and notation as they are presented in class

- *polynomial*
- *degree*
- *monic* polynomial (leading coefficient is 1)
- *vector space over a field F*, where F might be $\mathbb{R}$ (the real numbers, $\mathbb{C}$ (the complex numbers), $\mathbb{Q}$ (the rational numbers) or $\mathbb{Z}_2 = \{0, 1\}$.
- F^n denotes the vector space of n-tuples of elements from F and $F^{m\times n}$ denotes the set of all $m \times n$ matrices with entries in F (page 29). $\mathcal{C}(X, F)$ denotes the vector space of F-valued functions defined on a set X.
- *subspace*
- *linear combination*
- *linearly dependent* and *linearly independent*
- *spanning set* or generating set. Definition 4.6 uses $\mathrm{S}(T)$ for the subspace spanned by the set T.
- *basis*
- *dimension*, the number of elements in a basis.
- *linear transformation*
- *isomorphism*
- *direct sum* of two subspaces of a vector space.
- *complementary subspace*

- *quotient space* of a vector space *modulo* a subspace.
- *kernel* or *null space* of a linear transformation.
- *rank* of a linear transformation is the dimension of the range.
- *linear functional*
- $L(V, W)$ denotes the vector space of all linear transformations from V into W.
- *dual space* $V^* = L(V, F)$ of a vector space V.
- *dual basis*
- For S a subset of V, the *annihilator* of S is $S^\circ = \{ f \in V^* \mid f(s) = 0 \text{ for all } s \in S \}$
- *inner product, inner product space, Euclidean space, unitary space.*
- The *norm* of a vector v is $\|v\| = \sqrt{(v|v)}$.
- *metric*
- $x \perp y$ means x is *orthogonal* to y. $S^\perp = \{ x \mid x \perp S \}$.
- *orthonormal basis*
- *similar matrices*
- *change of basis matrix*
- *transpose* of a linear transformation
- *adjoint* of a linear transformation
- *permutation*
- *determinant, minor and cofactor* of a matrix
- *classical adjoint* of a matrix
- *eigenvalue, eigenvector, eigenspace*
- *characteristic polynomial, minimal polynomial* of a matrix or a linear transformation.
- A matrix is *diagonalizable* if it is similar to a diagonal matrix. A linear transformation is *diagonalizable* if there is a basis in which its matrix is diagonal.

3. Main theorems as they are presented in class

Theorem 3.1. *Polynomial theorems.*

(1) *Division algorithm for polynomials*
(2) *Euclidean algorithm (for finding greatest common divisors)*
(3) *Unique factorization theorem for polynomials*

Theorem 3.2. *A nonempty subset W of a vector space V over a field F is a subspace of V iff*

$$a_1 w_1 + a_2 w_2 \in W \quad \text{for all } a_i \in F,\ w_i \in W\ .$$

Theorem 3.3. *Any spanning set $\mathcal{S}$ of a vector space V contains a basis.*

Proof. The zero space has the empty set as a basis, so we may assume that $V \neq \{0\}$. Let

$$\mathcal{L} = \{\, K \subset \mathcal{S} \mid K \text{ is linearly independent} \,\}$$

The set $\mathcal{L}$ is nonempty (since sets of one vector are linearly independent) and is partially ordered by set inclusion.

For any totally ordered subset $\{K_\alpha\}$ of $\mathcal{L}$, let $K_0 = \bigcup K_\alpha$. We claim K_0 is an upper bound for the set $\{K_\alpha\}$ in $\mathcal{L}$. It certainly contains all the sets K_α, so we need to show that it is linearly independent. Assume that there is an equation $\sum_{i=1}^n c_i u_i = 0$ with $c_i \in F$ and $u_i \in K_0$. The each u_i is in some set K_{α_i} in the union. Since $\{K_\alpha\}$ is totally ordered, we may order the subscripts so that $K_{\alpha_1} \subset K_{\alpha_2} \subset \cdots \subset K_{\alpha_n}$. Then all $u_i \in K_{\alpha_n}$, which is a linearly independent set. Therefore all $c_i = 0$ as desired.

We now apply Zorn's Lemma to conclude that $\mathcal{L}$ has a maximal element $\mathcal{B}$. We claim that $\mathcal{B}$ is a basis for V. By definition of $\mathcal{L}$, the set $\mathcal{B}$ is contained in the generating set $\mathcal{S}$ and is linearly independent. If $\mathcal{B} = \mathcal{S}$, then it spans V. If not, there is some element $v \in \mathcal{S}$ with $v \notin \mathcal{B}$. If v were not a linear combination of elements in $\mathcal{B}$, then we could make a larger linearly independent set in $\mathcal{L}$, namely $\mathcal{B} \cup \{v\}$, contradicting the maximality of $\mathcal{B}$. Thus any element of the spanning set $\mathcal{S}$ can be written as a linear combination of elements of $\mathcal{B}$, and therefore $\mathcal{B}$ also spans V, hence is a basis. □

Theorem 3.4. *Any linearly independent set can be extended to a basis.*

Theorem 3.5. *The following are equivalent:*

(1) $\mathcal{B}$ *is a basis for* V*.*
(2) $\mathcal{B}$ *is a maximal linearly independent set in* V*.*
(3) $\mathcal{B}$ *is a minimal spanning set for* V*.*

Theorem 3.6. *Any two bases for a vector space* V *have the same cardinality.*

Proof. Let $\mathcal{B} = \{v_i\}_{i \in I}$ and $\mathcal{C} = \{w_j\}_{j \in J}$ be two bases for V. If either I or J is finite, the sets have the same cardinality as in the proof in the book for finite dimensional spaces. Thus we may assume that both sets are infinite. For each $i \in I$, write

$$v_i = a_1 w_{j_1} + \cdots a_r w_{j_r} \qquad (a_i \in F)$$

Every element of $\mathcal{C}$ occurs in such an expression, for otherwise that w which did not occur could be written in terms of the basis $\{v_i\}_{i \in I}$, and in turn back in terms of other w_j's contradicting the linear independence of $\mathcal{C}$.

Now define a correspondence

$$i \longleftrightarrow \{j_1, \ldots, j_r\} \subset J$$

That is, we have a mapping from I into the collection of finite subsets of J. It follows from set theory that J has cardinality less than or equal to the cardinality of I [Bourbaki, Theory of Sets III.6.4, Prop. 5]. Reverse the roles of I and J to see that $|I| \leq |J|$ and therefore $|I| = |J|$. □

Theorem 3.7. *Any two vector spaces of the same dimension over a field F are isomorphic.*

Theorem 3.8. *If U and W are finite dimensional subspaces of a vector space V, then*

$$\dim(U + W) = \dim U + \dim W - \dim(U \cap W) \ .$$

Theorem 3.9. *If W is a subspace of V, then there is a complementary subspace U; that is $V = U + W$ and $U \cap W = \{0\}$.*

Theorem 3.10. *Let W be a subspace of V with quotient mapping $T\colon V \to V/W$. A subspace U is a complement to W iff $T|_U\colon U \to V/W$ is an isomorphism.*

Theorem 3.11. *A linear transformation $T\colon V \to W$ is one-to-one iff $\ker T = \{0\}$. If $\dim V = \dim W = n$, then the following are equivalent:*

(1) *T is an isomorphism.*
(2) *T is one-to-one.*
(3) $\ker T = \{0\}$.
(4) *T is onto.*

Theorem 3.12 (The "usual" homomorphism results all hold)**.** *Let $T\colon V \to W$ be a linear transformation between vector spaces V and W. Then*

- *The image (or range) of T, $\operatorname{im}(T)$, is a subspace of W.*
- *$V/\ker(T)$ is isomorphic to $\operatorname{im}(T)$.*
- *If T is onto, then there exists a one-to-one correspondence between subspaces of W and subspaces of V containing $\ker(T)$.*

For a finite dimentional vector space V and a linear transformation $T\colon V \to W$, the book uses the terms *nullity* for $\dim \ker T$ and *rank* for $\dim \operatorname{im} T$. Since $V/\ker(T) \cong \operatorname{im}(T)$, we obtain $\dim V - \dim \ker T = \dim \operatorname{im} T$, or $\dim V$ equals the rank plus the nullity of T.

Theorem 3.13 (Linear transformations and matrices)**.** *Let V be a vector space with ordered basis $S = \{v_1, \ldots, v_n\}$ and let W be a vector space with ordered basis $T = \{w_1, \ldots, w_m\}$, $m, n > 0$. Let $L\colon V \to W$ be a linear transformation. Let A be the $m \times n$ matrix with jth column equal to the coordinate vector $[L(v_j)]_T$ of $L(v_j)$ with respect to the basis T. If $w = L(v)$ for some $v \in V$, then $[w]_T = A[v]_S$ and A is the unique matrix with this property.*

Proof. Write $A = [c_{ij}]$ where $[L(v_j)]_T = \begin{bmatrix} c_{1j} \\ \vdots \\ c_{mj} \end{bmatrix}$, or equivalently, $L(v_j) = \sum_{i=1}^m c_{ij} w_i$.

Let $[v]_S = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix}$ and $[w]_T = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix}$. This means $v = \sum_{j=1}^n a_j v_j$, so

$$L(v) = \sum_{j=1}^n a_j L(v_j) = \sum_{j=1}^n a_j \sum_{i=1}^m c_{ij} w_i = \sum_{i=1}^m \left(\sum_{j=1}^n a_j c_{ij} \right) w_i,$$

which must equal $w = \sum_{i=1}^m b_i w_i$. Setting coefficients equal, we obtain $b_i = \sum_{j=1}^n c_{ij} a_j$ for each $i = 1, 2, \ldots, m$. In matrix form, this says

$$\begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix} = [c_{ij}] \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix},$$

or $[w]_T = A[v]_S$, as desired.

To see that A is unique, take $v = v_k$, so that we require $[v]_S = [v_k]_S = e_k$, the standard basis column vector with a 1 in the kth row, and $A[v]_S = \begin{bmatrix} c_{1k} \\ \vdots \\ c_{mk} \end{bmatrix}$, the kth column of A, which is unique since $L(v_k)$ has a unique coordinate vector with respect to the basis T. □

In the context of the previous theorem, Exercise 6(b) on page 55 was in essence asking for the matrix of the identity linear transformation on a certain 3 dimensional vector space of functions with respect to different bases for the domain and range. This is often called a *change of basis matrix* or a *transition matrix* since it carries one basis to another. In that exercise, the basis for the domain was $\{1, \cos x, \sin x\}$ and the basis for the range was $\{1, e^{ix}, e^{-ix}\}$ so the matrix P has as its second column, the coordinate vector of $\cos x = 0 \cdot 1 + \frac{1}{2} e^{ix} + \frac{1}{2} e^{-ix}$, namely $\begin{bmatrix} 0 \\ 1/2 \\ 1/2 \end{bmatrix}$.

Theorem 3.14 (§3.5, Theorem 15)**.** *Let V be a vector space of dimension n, with basis $\{x_1, \ldots x_n\}$. Then V^* has a basis $\{f_1, \ldots f_n\}$ with the property that $f_j(x_i) = \delta_{ij}$. This is called the* dual basis *to $\{x_1, \ldots x_n\}$. In particular, $\dim V^* = n$.*

Corollary 3.15. *If $\dim V = n$, then $\dim V^{**} = n$ also, and so V is isomorphic to V^{**}. In fact, there is a canonical isomorphism $v \mapsto L_v$ where $L_v(f) = f(v)$ for any $v \in V,\ f \in V^*$.*

Proposition 3.16. *Let V be a vector space and $S \subseteq V$.*

(1) *S° is a subspace of V^*.*
(2) *If $\dim V = n$ and U is a subspace of dimension m, then $\dim U^\circ = n - m$. Furthermore, $U^{\circ\circ} = U^{**}$ is canonically isomorphic to U via the double dual isomorphism.*

Proposition 3.17. *Assume an arbitrary vector space $V = U \oplus W$. Then $V^* = U^\circ \oplus W^\circ$, $U^* \cong W^\circ$ and $W^* \cong U^\circ$.*

Inner Products (See Curtis, Section 32.)

For this section, our field F can only be the real numbers $\mathbb{R}$ or the complex numbers $\mathbb{C}$. We work mainly with two types of examples. The standard inner product on F^n is defined by

$$(v|w) = \sum_{i=1}^{n} a_i \overline{b_i}, \quad \text{where } v = (a_1, \ldots, a_n),\ w = (b_1, \ldots, b_n). \tag{3.1}$$

Our other main example is in working with spaces of integrable functions. For example, if V denotes the vector space of all polynomials with complex coefficients (or all continuous complex-valued functions on [0,1]), we can define an inner product by

$$(f|g) = \int_0^1 f(t)\overline{g(t)}\, dt\ .$$

In either case, the norm of a vector v is defined as $\|v\| = \sqrt{(v|v)}$ and is a REAL number.

Proposition 3.18. *Some useful formulas. Let x, y, z be in an inner product space V, $a \in F$.*

(1) $(x|ay) = \bar{a}(x|y)$.
(2) $(x|y + z) = (x|y) + (x|z)$.
(3) $(x|y) = Re(x|y) + iRe(x|iy)$, *where Re denotes real part.*

Theorem 3.19. *In any inner product space,*

$$\|x \pm y\|^2 = \|x\|^2 \pm 2Re(x|y) + \|y\|^2$$

For a real inner product space, this gives the law of cosines

$$\|x - y\|^2 = \|x\|^2 + \|y\|^2 - 2\|x\|\|y\| \cos\theta$$

Corollary 3.20. *For a Euclidean space, $(x|y) = \frac{1}{4}(\|x + y\|^2 - \|x - y\|^2)$.*

Corollary 3.21. *For a unitary space,*
$(x|y) = \frac{1}{4}(\|x + y\|^2 - \|x - y\|^2 + i\|x + iy\|^2 - i\|x - iy\|^2)$.

Corollary 3.22 (Parallelogram Law). $\|x + y\|^2 + \|x - y\|^2 = 2\|x\|^2 + 2\|y\|^2$.

Theorem 3.23. *On a finite dimensional inner product space V, each linear functional arises as the inner product with a fixed vector in V. That is, given $f \in V^*$, there exists $u_0 \in V$ such that $f(v) = (v|u_0)$ for all $v \in V$.*

Example 3.24. This theorem fails for infinite dimensional spaces. As an example take the space V of all infinite tuples $(a_1, a_2, a_3, \dots)$ in which all but finitely many a_i are zero. This is clearly isomorphic to the space of polynomials, but we are not yet ready to describe an appropriate basis for polynomials (something that is a major topic in Math 407–408). Instead we extend the usual ideas from F^n to V: we have a basis $\{e_i\}_{i=1}^{\infty}$, where e_i is the tuple with 1 in the i-th place and zeros elsewhere. And for an inner product we use equation (3.1) as an infinite sum since only finitely many terms can be nonzero. Thus $(e_i|e_j) = \delta_{ij}$. Now let $f\colon V \to F$ be defined by taking every e_i to 1. Assume that $f(v) = (v|y)$ for all $v \in V$. Write $y = \sum a_i e_i$, $a_i \in F$. Then $1 = f(e_j) = (e_j|\sum a_i e_i) = \sum \overline{a_i}(e_j|e_i) = \overline{a_j}$, so $a_j = 1$ for all j. But the infinite tuple of all 1's is not in V, so we have a contradiction.

Theorem 3.25. *For any inner product space,*

(1) $\|av\| = |a|\|v\|$.
(2) $|(x|y)| \le \|x\|\|y\|$ *(Cauchy-Schwarz inequality)*
(3) $\|x + y\| \le \|x\| + \|y\|$ *(triangle inequality)*

Definition 3.26. A *metric* on a vector space V is a mapping $d\colon V \times V \to \mathbb{R}$ such that $d(x, y) = d(y, x)$, $d(x, y) \ge 0$ and equals 0 iff $x = y$, and $d(x, z) \le d(x, y) + d(y, z)$. Note that $d(x, y) = \|x - y\|$ defines a metric with the special property that $d(x + z, y + z) = d(x, y)$; that is, translation of two vectors preserves the distance between them. [If V is complete in this metric, we say it is a *Hilbert space*.]

If $(x|y) = 0$, we write $x \perp y$ and say x is *orthogonal to* y. Note that $x \perp y \iff y \perp x$, $0 \perp x$ for all x, and $x \perp x \iff x = 0$.

Proposition 3.27. *If $x \perp T$, then $x \perp \mathrm{S}(T)$ for any subset $T \subseteq V$.*

Theorem 3.28 (Gram-Schmidt)**.** *Every finite dimensional inner product space has an orthonormal basis.*

Proposition 3.29. *If V is an inner product space and W is a finite dimensional subspace, then $V = W \oplus W^{\perp}$ and $W = W^{\perp\perp}$.*

Back to general vector spaces – no inner product, any field

The book writes $L(V, W)$ for the set of all linear transformations from a vector space V to a vector space W over the same field F. We have already noted that this is a vector space. Section 3.2 of the book discusses the *algebra* of linear transformations without telling you what an algebra is. With $L(V, V)$ as an example, we make the definition:

Definition 3.30. An *algebra* A over a field F is a vector space over F that has a multiplication operation [composition for $L(V,V)$] that satisfies

(1) the multiplication is associative and distributes over addition (making A a *ring*).
(2) $a(xy) = (ax)y = x(ay)$ for all $a \in F$, $x, y \in A$.

Example 3.31. Some other examples besides $L(V,V)$ include $\mathbb{C}$ as a vector space over $\mathbb{R}$ and the polynomial ring $F[x]$. These last two have a commutative multiplication, but the composition in $L(V,V)$ is not commutative. In fact, you showed for homework that you may even have $TU = 0$ but $UT \neq 0$. Another example is $F^{n\times n}$, the set of $n \times n$ matrices over F since we can add them, multiply them and also have scalar multiplication.

We have already looked at the connection between linear transformations and matrices, namely, once you pick bases for the domain and range, you have a unique matrix which will let you do the computations. The next theorem says that this correspondence makes $L(V,V)$ and $F^{n\times n}$ isomorphic as algebras (that is, the vector space isomorphism also preserves the multiplication).

Theorem 3.32. *Let U, V, W be finite dimensional vector spaces over a field F. Let $\{u_i\}, \{v_j\}$ and $\{w_k\}$ be bases for these spaces, respectively. For any linear transformation T between two of these spaces, we write* $\operatorname{mat} T$ *for the matrix of T with respect to the appropriate bases.*

(1) *If $S, T \in L(V,W)$, then* $\operatorname{mat}(S+T) = \operatorname{mat} S + \operatorname{mat} T$.
(2) *If $S \in L(V,W)$ and $a \in F$, then* $\operatorname{mat}(aS) = a \operatorname{mat} S$.
(3) *If $S \in L(U,V)$, $T \in L(V,W)$, then* $\operatorname{mat}(TS) = \operatorname{mat} T \cdot \operatorname{mat} S$.

Proposition 3.33. *For a finite dimensional vector space V and $T \in L(V,V)$, there exists $q > 0$ such that $V = T^q(V) \oplus \ker T^q$.*

We say that T is *nilpotent* on the subspace $\ker T^q$. This just means that some power of T is zero. Since no nonzero element of $T^q(V)$ is in the kernel of T, the linear transformation T is 1–1 on $T^q(V)$.

Lets take a closer look at the ideas of 1–1 and onto. We have seen that a linear transformation on a finite dimensional vector space is one-to-one iff it is onto iff it has an inverse. We have also seen examples on homework of linear transformations with an inverse on one side but not the other (such as differentiation of polynomials). Let $T \in L(V,W)$. We say T is *left invertible* if there exists $S \in L(W,V)$ such that $ST = I_V$, the identity on V. We say T is *right invertible* if there exists $R \in L(W,V)$ such that $TR = I_W$, the identity on W. We only say that T is *invertible* if it has both a left and a right inverse, and then the inverses are equal and denoted by T^{-1}.

Proposition 3.34. *If T is left invertible, then T is one-to-one. If T is right invertible, then T is onto.*

Change of basis matrix

We have seen examples of how this works, to change the basis with respect to which the matrix of a linear transformation is computed. We take a closer look. For now, assume that $T \in L(V, V)$, so we are working only with one vector space. Let $\{x_i\}$ and $\{y_i\}$ be two bases and let A and B be the matrices with respect to these bases, respectively. Let $P = (p_{ij})$ be the matrix of the identity linear transformation expressed with respect to $\{y_i\}$ for the domain and $\{x_i\}$ for the image. This means the i-th column of the matrix is the coordinate vector for y_i as linear combination of the x_j basis: $y_i = \sum_k p_{ki} x_k$. As we have noted before, this allows us to compute $B = P^{-1}AP$. We check this a different way; notice that we can also think of P as the matrix of a linear transformation S with respect to $\{x_i\}$ (for both domain and range) satisfying $S(x_i) = y_i$ for each i (because that is what the columns represent in this basis). Thinking of it this way, A and P are both written with respect to the same basis $\{x_i\}$, so AP takes the i-th coordinate vector $\begin{bmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix}$ to $T(y_i)$ written with respect to the basis $\{x_i\}$. Multiplying by P^{-1} now converts the coordinate vector of $T(y_i)$ into the a coordinate vector with respect to $\{y_i\}$, which is what B is supposed to compute. This can all be written out algebraically as in section 3.4 of the book.

We say two matrices A, B are *similar* if they are related by $B = P^{-1}AP$ for some nonsingular matrix P.

Proposition 3.35. *Similarity of matrices is an equivalence relation: that is,*

- *any matrix is similar to itself (reflexivity)*
- *If A is similar to B, then B is similar to A (symmetry)*
- *If A is similar to B and B is similar to C, then A is similar to C (transitivity)*

What we have seen is that a given linear transformation is associated with an equivalence class of similar matrices, and they all represent the linear transformation with respect to different bases.

Transposes

Let $T \colon V \to W$ be a linear transformation. We define the *transpose* of T to be the function $T^t \colon W^* \to V^*$ defined as follows: if $f \in W^*$, set $T^t(f)(v) = f(T(v))$ for any $v \in V$.

Proposition 3.36. *T^t is a linear transformation, so is in $L(W^*, V^*)$.*

Example 3.37. If $T = 0$, then $T^t = 0$. If $V = W$ and I is the identity, then I^t is the identity in $L(V^*, V^*)$.

Theorem 3.38. *Let $S, T \in L(V, W)$, $a \in F$. Then $(S + T)^t = S^t + T^t$ and $(aT)^t = aT^t$. Thus the set of all transposes $\mathcal{T} = \{\, T^t \mid T \in L(V, W) \,\}$ is a subspace of $L(W^*, V^*)$. If also, $W = V$, then $(ST)^t = T^t S^t$, so $\mathcal{T}$ is a subalgebra of $L(V^*, V^*)$.*

Recall that we have canonical injections $V \to V^{**}$ and $W \to W^{**}$, which are isomorphisms when the spaces are finite dimensional. We shall denote these mappings by N_V and N_W. As a reminder, $N_V(v) = L_v$ where $L_v(f) = f(v)$ for all $v \in V$, $f \in V^*$. These provide a nice correspondence between T and T^{tt}.

Proposition 3.39. *Let $T\colon V \to W$ be a linear transformation. Then $T^{tt} N_V = N_W T$ as linear transformations from V to W^{**}.*

Proposition 3.40. *Let $T\colon V \to W$ be a linear transformation. Then $(\operatorname{im} T)^\circ = \ker T^t$.*

Theorem 3.41 (Matrix of the transpose)**.** *If $T \in L(V, V)$ and V is finite dimensional with basis $v_1, \dots, v_n$, then the matrix of T^t with respect to the dual basis is the transpose of the matrix of T with respect to the given basis.*

Adjoints: for inner product spaces. See page 283 of textbook.

Theorem 3.42. *Let V be a finite dimensional inner product space and let $T \in L(V, V)$. Then there exists a unique linear transformation $T^* \in L(V, V)$ such that*

$$(Tv|x) = (v|T^*x) \qquad \text{for all } x, v \in V.$$

The linear transformation T^* of the previous theorem is called the *adjoint* of T.

Proposition 3.43. *For a finite dimensional inner product space V and $S, T \in L(V, V)$,*

(1) $(T^*x|v) = (x, Tv)$ *for all* $x, v \in V$;
(2) $T = T^{**}$;
(3) $(S + T)^* = S^* + T^*$;
(4) $(ST)^* = T^* S^*$;
(5) $(aT)^* = \bar{a} T^*$ *for* $a \in F$.

Proposition 3.44. *With respect to an orthonormal basis for V, the matrix of T^* is the conjugate transpose of the matrix of T.*

Chapter 9 of the book does much more with inner product spaces and special classes of linear transformations. We need to move on to other topics (like determinants), but here are a few of the main ideas:

A *Unitary linear transformation* (in the complex case) or *orthogonal linear transformation* (in the real case) is one that satisfies $T^*T = I$. In the real case this means the rows of the matrix of T are orthonormal. The importance of such linear transformations is that they preserve length, so for example, rotations and reflections are examples. On $\mathbb{R}^2$, an orthogonal linear transformation is either a rotation or a reflection. A major theorem is that any distance preserving function on $\mathbb{R}^n$ is a composition of an orthogonal linear transformation and a translation.

A *self-adjoint* linear transformation is one satisfying $T = T^*$. These are also called *symmetric* if $F = \mathbb{R}$ or *Hermitian* if $F = \mathbb{C}$. The corresponding matrix is symmetric or Hermitian as well. Such matrices are used to generalize the idea of inner product to bilinear or Hermitian forms (Chapter 8). If S is an invertible linear transformation, $T = S^*S$ is *positive* in the sense that $(Tx|x) > 0$ for any $x \neq 0$. This gives rise to a new inner product $[x|y] = (Tx|y)$. On a unitary space, all inner products arise in this way.

A *normal* linear transformation is one satisfying $T^*T = TT^*$. Examples are self-adjoint and unitary linear transformations. These linear transformations have special properties for their eigenvalues, something we still need to discuss.

Determinants

We need a preliminary result from the Permutation notes on the class web page:

Theorem 3.45. *Let S_n be the set of all permutations on n elements. There exists a unique nontrivial function* $\operatorname{sgn}\colon S_n \to \{\pm 1\}$ *with the property that* $\operatorname{sgn}(\sigma\tau) = \operatorname{sgn}(\sigma)\operatorname{sgn}(\tau)$ *for all $\sigma, \tau \in S_n$. Furthermore,* $\operatorname{sgn}(\tau) = -1$ *iff τ can be written as a product of an odd number of transpositions.*

There are numerous ways in which a determinant can be defined. They are all useful, so we need to show they are all equivalent. Let $A = (a_{ij})$ be an $n \times n$ matrix.

(1) We will show that there is a unique alternating multilinear form defined on the columns of A with the property that its value on the identity matrix is 1.
(2) We will see that such a function has the formula
$$\det A = \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) a_{\sigma(1)1} a_{\sigma(2)2} \cdots a_{\sigma(n)n}$$
(3) Computationally, row operations can be used to make a diagonal matrix with the same determinant, and then the determinant is the product of the elements on the main diagonal.
(4) Expansion by minors, along any row or column, gives another way to compute a determinant: $\det A = \sum_{i=1}^n a_{ij}(-1)^{i+j} \det A(i|j)$ is the expansion along the j-th column.
(5) The *classical adjoint* of A, $\operatorname{adj} A$, yields another way to compute determinants via the formula $(\operatorname{adj} A) \cdot A = (\det A) \cdot I$. The classical adjoint is the transpose

of the matrix of cofactors. This also gives a formula for the inverse of a matrix: $A^{-1} = \frac{1}{\det A}\operatorname{adj} A$. This is not computationally efficient for matrices with numeric entries, but can be very useful if the entries are symbolic since row operations create messy denominators.

Lagrange interpolation—an application of linear algebra to polynomial approximation

Problem 3.46. Given distinct numbers $\alpha_1, \alpha_2, \ldots, \alpha_n \in F$, and a set of arbitrary numbers $c_1, \ldots, c_n \in F$, find $p(x) \in F[x]$ of degree at most $n-1$ such that

$$p(\alpha_i) = c_i, \qquad (i = 1, \ldots, n)$$

Theorem 3.47. *Define*

$$\pi_i(x) = \frac{\prod_{j \neq i}(x - \alpha_j)}{\prod_{j \neq i}(\alpha_i - \alpha_j)}$$

Then

(1) $\pi_i(\alpha_j) = \delta_{ij}$, *for* $1 \leq i, j \leq n$.
(2) *The set* $\mathcal{B} = \{\pi_1, \ldots, \pi_n\}$ *forms a basis for* $P_n = \{\, p(x) \in F[x] \mid \deg p(x) \leq n-1 \,\}$.
(3) *The polynomial* $p(x) = \sum_{i=1}^{n} c_i \pi_i(x)$ *satisfies the Lagrange Interpolation problem 3.46.*
(4) *The change of basis transformation from the standard basis to the basis* $\mathcal{B}$ *is given by*

$$\begin{bmatrix} 1 & \alpha_1 & \ldots & \alpha_1^{n-1} \\ 1 & \alpha_2 & \ldots & \alpha_2^{n-1} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & \alpha_n & \ldots & \alpha_n^{n-1} \end{bmatrix}$$

The matrix above is called a *Vandermonde matrix*. It is clearly nonsingular since it transforms a basis. In fact, this matrix form is rather common and has a nice formula for its determinant.

Theorem 3.48. *For any* $\alpha_1, \alpha_2, \ldots, \alpha_n \in F$, *the Vandermonde determinant*

$$\begin{vmatrix} 1 & \alpha_1 & \ldots & \alpha_1^{n-1} \\ 1 & \alpha_2 & \ldots & \alpha_2^{n-1} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & \alpha_n & \ldots & \alpha_n^{n-1} \end{vmatrix} = \prod_{j > i}(\alpha_j - \alpha_i).$$

Canonical forms of matrices Our goal in this part is to begin with a linear transformation and find a special basis in which the matrix has a particularly nice form. The most common of these are the rational canonical form and the Jordan canonical form. The latter one is as close as possible to being a diagonal matrix, so one thing we will discover is what property T must have for it to have an associated diagonal matrix.

We shall need to break V up into a direct sum. Recall that we have already seen that $V = W_1 \oplus W_2$ iff $W_1 \cap W_2 = \{0\}$ and $V = W_1 + W_2$ (that is, the two subspaces generate V). For larger direct sums, we need a bit more.

Theorem 3.49. *Let $W_1, \dots, W_n$ be subspaces of V and let $V = W_1 + \cdots + W_n = \{\sum w_i \mid w_i \in W_i\}$. The following are equivalent:*

1. *(Independence) $\sum w_i = 0$, with $w_i \in W_i$ implies each $w_i = 0$.*
2. *For each i, $W_i \cap (\sum_{j \neq i} W_j) = \{0\}$.*
3. *A union of bases for the subspaces W_i forms a basis of V.*

When these conditions hold, we say that V is the direct sum of the subspaces W_i and write $V = W_1 \oplus W_2 \oplus \cdots \oplus W_n$.

Corollary 3.50. *If $V = W_1 \oplus W_2 \oplus \cdots \oplus W_n$, then $\dim V = \sum(\dim W_i)$.*

Definition 3.51. If $T \in L(V, V)$ and M is a subspace of V, we say that M *is invariant under* T if $T(M) \subseteq M$.

Proposition 3.52. *If M is invariant under T, then there exists a basis of V in which the matrix of T has the form $\begin{bmatrix} A & B \\ 0 & C \end{bmatrix}$. If the complement N of M in V (so $V = M \oplus N$) is also invariant, then $B = 0$ in the form above and we have a block diagonal matrix.*

Corollary 3.53. *If $V = W_1 \oplus \cdots \oplus W_n$ and each W_i is invariant under T, we can write the matrix of T as $\begin{bmatrix} A_1 & & 0 \\ & \ddots & \\ 0 & & A_n \end{bmatrix}$, where A_i is the matrix of T restricted to a basis for W_i.*

Definition 3.54. If $V = M \oplus N$, then the function E defined as $E(x + y) = x$ for $x \in M,\ y \in N$, is call the *projection on M along N*.

Check that E is a linear transformation.

Proposition 3.55. *A linear transformation $E \in L(V, V)$ is a projection iff $E^2 = E$.*

Corollary 3.56. *If E is a projection, then $V = \operatorname{im} E \oplus \ker E$. With respect to a basis $\{x_1, \dots, x_n\}$ where $x_1, \dots, x_r$ forms a basis for $\operatorname{im} E$ and $x_{r+1}, \dots, x_n$ forms a basis for $\ker E$, the matrix of E has the form $\begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}$.*

Proposition 3.57. *E is a projection iff $I - E$ is a projection.*

Theorem 3.58. *If $V = W_1 \oplus \cdots \oplus W_n$, then there exist $E_i \in L(V,V)$, $i = 1, \ldots, n$, such that $E_i^2 = E_i$, $I = \sum E_i$ and $E_i E_j = 0$ if $i \neq j$. Conversely, if $E_i E_j = \delta_{ij}$ and $I = \sum E_i$, then $V = \operatorname{im} E_1 \oplus \cdots \oplus \operatorname{im} E_n$.*

Definition 3.59. An element $\lambda \in F$ is an *eigenvalue* or *characteristic value* for a linear transformation $T \in L(V,V)$ provided that $T - \lambda I$ is not invertible. Equivalently, there exists $v \neq 0$ in V such that $Tv = \lambda v$. The vector v is called an *eigenvector* or *characteristic vector* of T associated with λ. The subspace $\{v \in V \mid Tv = \lambda v\}$ is called the *eigenspace* or *characteristic space* associated with λ. We make similar definitions for a matrix A.

Proposition 3.60. *λ is an eigenvalue of A iff* $\det(A - \lambda I) = 0$.

The polynomial $f_A(x) = \det(xI - A)$ in the variable x is called the *characteristic polynomial* of A.

Proposition 3.61. *Similar matrices have the same characteristic polynomial.*

Definition 3.62. Let $T \in L(V,V)$ for some finite dimensional vector space V. The *characteristic polynomial* $f_T(x)$ of T is the characteristic polynomial of the matrix of T with respect to any basis.

Proposition 3.63. *If $\lambda_1, \ldots, \lambda_r \in F$ are the distinct eigenvalues of a linear transformation $T \in L(V,V)$ with associated nonzero eigenvectors $x_1, \ldots, x_r$, then the set $\{x_1, \ldots, x_r\}$ is linearly independent. If $r = n = \dim V$, then the matrix of T with respect to this basis is diagonal with the eigenvalues on the diagonal.*

Lemma 3.64. *Let $C = B(\lambda)(A - \lambda I)$ where C has constant entries while the entries of $B(\lambda)$ and $(A - \lambda I)$ are polynomials in λ. Then $C = 0$.*

Theorem 3.65 (Cayley-Hamilton Theorem)**.** *Any matrix satisfies its characteristic polynomial. That is, $f_A(A) = 0$ for any $n \times n$ matrix A.*

It follows that any linear transformation satisfies its characteristic polynomial. We saw at the beginning of the semester that any two polynomials $f(x), g(x)$ can have their greatest common divisor written in the form $d(x) = u(x)f(x) + v(x)g(x)$. If $f(T) = 0 = g(T)$, then $d(T) = 0$. It follows that there is a polynomial of smallest possible degree $m(x)$ that T satisfies. Such polynomials will differ only by a constant multiple, so we may assume that $m(x)$ is monic. This unique monic polynomial is called the *minimal polynomial* of T. Using the division algorithm, we see that $m(x)$ divides $f_T(x)$. Thus the roots of $m(x)$ are all roots of $f_T(x)$. The converse is also true!

Proposition 3.66. *If λ is an eigenvalue of T, then $m(\lambda) = 0$.*

Theorem 3.67. *Let $m(x) \in F[x]$ be the minimal polynomial of $T \in L(V,V)$. Factor $m(x) = p_1(x)^{e_1} \cdots p_s(x)^{e_s}$ where each $p_i(x)$ is irreducible over the field F. Let $T_i =$*

$p_i(T)^{e_i}$, again a linear transformation from V to V. Then $V = \ker(T_1) \oplus \cdots \oplus \ker(T_s)$ is an invariant direct sum decomposition of V.

Theorem 3.68. *T is diagonalizable (i.e., there exists a basis in which the matrix of T is diagonal) if and only if $m(x) = \prod_{i=1}^{s}(x - \lambda_i)$, where the λ_i are all distinct and in F.*

Example 3.69. Let $F = \mathbb{R}$. The matrix $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ has minimal polynomial $x-1$ and is clearly diagonal. The matrix $\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ has minimal polynomial $(x-1)^2$ so is not similar to a diagonal matrix. The matrix $\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$ has characteristic polynomial x^2+1. The roots are distinct in $\mathbb{C}$, so it must also be the minimal polynomial; since the roots are not in $\mathbb{R}$, it is not similar to a diagonal real matrix, but is similar to $\begin{bmatrix} i & 0 \\ 0 & -i \end{bmatrix}$.

Definition 3.70. A linear transformation $T \in L(V,V)$ is *nilpotent* of index q if T^q is the zero linear transformation, but $T^{q-1} \neq 0$.

For example, on the space of polynomials of degree at most n, the derivative is nilpotent of index $n+1$.

Lemma 3.71. *If $T \in L(V,V)$ is nilpotent, then there exists a basis in which the matrix $A = (a_{ij})$ of T is strictly upper triangular ($a_{ij} = 0$ if $i \geq j$.)*

Theorem 3.72. *If the minimal polynomial of a linear transformation T is $m(x) = \prod_1^s (x-\lambda_i)^{e_i}$ (so all eigenvalues are in F), then there exists a basis in which the matrix of T consists of s blocks down the diagonal, the ith block being an upper triangular $e_i \times e_i$ matrix with λ_i in each diagonal position.*

Theorem 3.73. *If T is nilpotent of index q, then V can be written in the form $V = S(v, Tv, T^2v, \ldots, T^{q-1}v) \oplus K$, where K is an invariant subspace and v is any vector not in $\ker T^{q-1}$.*

Corollary 3.74. *If T is nilpotent of index q on an n-dimensional vector space V, then there exists a basis of the form*

$$\{v_1, Tv_1, \ldots, T^{q-1}v_1; v_2, Tv_2, \ldots, T^{q_2-1}v_2; \ldots; v_s, Tv_s, \ldots, T^{q_s-1}v_s\},$$

where $q \geq q_2 \geq \cdots \geq q_s$ and $n = q + q_2 + \cdots + q_s$.

With respect to the basis given in the Corollary, the matrix of T consists of s blocks of size $q_i \times q_i$ each of which is entirely zero except for 1's down the subdiagonal (see Lemma 25.10 of the text). It is more common to reverse the order of the basis so that the 1's are on the superdiagonal. Note that the characteristic polynomial of T is x^n and the minimal polynomial is x^q, which can be seen either from the matrix or from the definition of nilpotent of index q.

Proposition 3.75. *If T is nilpotent of index q on V and $V = V_1 \oplus \cdots \oplus V_r = U_1 \oplus \cdots \oplus U_s$, both in decreasing order of dimension and with bases of the form $v, Tv, \ldots, T^k v$, then $r = s$ and $\dim U_i = \dim V_i$ for each i.*

Theorem 3.76 (Jordan Canonical Form)**.** *A matrix A with all eigenvalues in F is similar to a unique (up to order of the blocks) matrix in Jordan form; the Jordan form consists of a block matrix where each block has an eigenvalue repeated down the diagonal, 1's separated by single 0's on the subdiagonal (or superdiagonal) and zeros elsewhere.*

Example 3.77. (1) Let

$$A = \begin{bmatrix} -1 & 2 & 2 \\ 2 & 2 & 2 \\ -3 & -6 & -6 \end{bmatrix}$$

Then $\det(xI - A) = x(x+2)(x+3)$ is the characteristic polynomial and also the minimal polynomial since the roots are distinct. The Jordan Canonical Form of the matrix is

$$\begin{bmatrix} 0 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & -3 \end{bmatrix}$$

(2) Let

$$A = \begin{bmatrix} 1 & 1 & -1 \\ -1 & 3 & -1 \\ -1 & 2 & 0 \end{bmatrix}$$

Then $\det(xI - A) = (x-1)^2(x-2)$ is the characteristic polynomial, so the minimal polynomial is either $(x-1)(x-2)$ or $(x-1)^2(x-2)$. We can check that $(A - I)(A - 2I) \neq 0$, so the latter polynomial works. Thus the Jordan Canonical Form of the matrix is

$$\begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

(3) Let

$$A = \begin{bmatrix} 1 & 0 & -1 & 1 & 0 \\ -4 & 1 & -3 & 2 & 1 \\ -2 & -1 & 0 & 1 & 1 \\ -3 & -1 & -3 & 4 & 1 \\ -8 & -2 & -7 & 5 & 4 \end{bmatrix}$$

Then $\det(xI - A) = (x-2)^5$ is the characteristic polynomial, so $A - 2I$ is nilpotent. Since $(A - 2I)^2 \neq 0$ and $(A - 2I)^3 = 0$, the index $q = 3$ and the minimal polynomial is $(x-2)^3$. We now almost know the Jordan Canonical Form, namely

$$B = \begin{bmatrix} 2 & 1 & 0 & 0 & 0 \\ 0 & 2 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 2 & ? \\ 0 & 0 & 0 & 0 & 2 \end{bmatrix}$$

We now follow all those theorems on nilpotent linear transformations. Let T be the linear transformation with matrix $A - 2I$. Choose some v such that $(A-2I)^2 v \neq 0$, say $v = (0,0,0,1,0)^t$. Our first three basis elements will be v, Tv, T^2v or $(0,0,0,1,0)^t, (1,2,1,2,5)^t, (0,0,1,1,1)^t$. For the complementary subspace K, we need a vector v_4 such that $v_4 \notin S(v, Tv, T^2v)$ and $Tv_4 \neq 0$; $v_4 = (-1,0,1,0,0)^t$ works and then $v_5 = Tv_4 = (0,1,0,0,1)^t$ completes the basis. The index $q = 2$ for T on the subspace K and the question mark in B is a 1. If there were no such v_4, that is, if $q = 1$, then the question mark would be a 0. The matrix for T is just B without the 2's on the diagonal, that is, $B - 2I$.

Rational Canonical Form. We next look at a canonical form which does not require that the eigenvalues be in the field F. Chapter 7 includes a lot of abstract algebra for polynomial rings which is normally done in a first year graduate algebra class [Fundamental Theorem of finitely generated torsion modules over a principal ideal domain]. For that reason, I will present the main theorem without proof; we know enough to understand the statement of the theorem.

Again, V is a finite dimensional vector space and $T \in L(V,V)$.

Theorem 3.78. *Let the minimal polynomial of T be $m_T(x) = \sum_0^r c_i x^i$. Assume that there exists a vector $v \in V$ such that V is spanned by the vectors $T^k v$, $k = 0, 1, \ldots$. Then $\{v, Tv, \ldots, T^{r-1}v\}$ is a basis for V in which the matrix of T becomes*

$$C(m_T(x)) = \begin{bmatrix} 0 & 0 & \ldots & 0 & -c_0 \\ 1 & 0 & \ldots & 0 & -c_1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \ldots & 1 & -c_{r-1} \end{bmatrix},$$

called the companion matrix of m_T.

Theorem 3.79. *Assume the minimal polynomial of T is $m_T(x) = p(x)^e$, where $p(x)$ is a monic irreducible polynomial over F. Then the matrix of T with respect to an appropriate basis consists of blocks down the diagonal of the form $C(p(x)^{e_i})$, $i = 1, \ldots, r$, where $e = e_1 \geq e_2 \geq \cdots \geq e_r$.*

Theorem 3.80 (Rational Canonical Form)**.** *Factor the minimal polynomial $m_T(x) = \prod_1^s p_i(x)^{e_i}$ into powers of distinct monic irreducible polynomials. Then there exists a* *basis of V such that the matrix of T has blocks down the diagonal in the form of the previous theorem.*

Example 3.81. Let

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

Then $\det(xI - A) = (x-1)(x+2)^2$ is the characteristic polynomial. One can check that $A^2 + A - 2I = 0$, so that $m_A(x) = (x-1)(x+2)$. We have $e_1 = e_2 = 1$ so the Rational Canonical form of A is

$$\begin{bmatrix} C(x+2) & 0 & 0 \\ 0 & C(x+2) & 0 \\ 0 & 0 & C(x+1) \end{bmatrix} = \begin{bmatrix} -2 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

As another example, look at

Example 3.82. Let

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

Then $\det(xI - A) = x^2 + 1$ is the characteristic polynomial and the eigenvalues are not real. The Rational Canonical form (over $\mathbb{R}$) of A is

$$C(x^2+1) = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$$