FALL 2022: MATH 790 DAILY UPDATE

Wednesday, December 7. We continued our discussion of finite matrix groups over $\mathbb{C}$ by noting a sixth property not mentioned in the previous lecture: For each $A_i \in G$, $\text{trace}(A_i^{-1}) = \overline{\text{trace}(A)}$.

We then proved a matrix version of Maschke's Theorem:

Theorem. For a finite matrix group $G \subseteq \text{Gl}_n(\mathbb{C})$, every G-invariant subspace of $\mathbb{C}^n$ is a direct sum of irreducible, G-invariant subspaces.

The key point of the proof is to use the celebrated averaging technique. If $\langle\ ,\ \rangle$ denotes the standard inner product on $\mathbb{C}^n$, one defines a new inner product as follows: $\{v_1, v_2\} := \frac{1}{|G|}\sum_{A\in G}\langle Av_1, Av_2\rangle$. This inner product has the property that $\{Av_1, Av_2\} = \{v_1, v_2\}$ for all $A \in G$ and $v_i \in \mathbb{C}^n$. It follows easily from this that if $U \subseteq \mathbb{C}^n$ is a G-invariant subspace, then $U^\perp$ is also G-invariant, where orthogonality is taken in terms of the new inner product. One then proves Maschke's Theorem by induction on the dimension of the given subspace.

Monday, December 5. We began class by sketching a proof that if V is a vector space over F and $F \subseteq K$ is an extension of fields, then $K \otimes_F V$ has the structure of a vector space over K, with scalar multiplication given by $k \cdot (\alpha \otimes v) := (k\alpha) \otimes v$, for all $k \in K$ and $\alpha \otimes v \in K \otimes V$. Full details that this multiplication is well-defined are given in a handout on tensor products.

We then began a discussion of matrix groups. We defined a *matrix group G (over $\mathbb{C}$)* to be a subset of $\text{Gl}_n(\mathbb{C})$ closed under products and taking inverses. $\text{Gl}_n(\mathbb{C})$ is a matrix group, as are matrices with determinant one and permutation matrices. Our focus will be on finite matrix groups. We then worked through the proofs the following properties of a finite matrix group $G \subseteq \text{Gl}_n(\mathbb{C})$:

(i) If $G = \{A_1, \ldots, A_r\}$, then for any $A \in G$, $= \{AA_1, \ldots, AA_r\}$.
(ii) For each $A \in G$, there exists $d > 0$ such that $A^d = I$.
(iii) Each $A \in G$ is diagonalizable.
(iv) For all $A \in G$, the eigenvalues of A are complex roots of unity.
(v) The trace of any matrix A in G is a sum of complex roots of unity.

The crucial point to the items above is that each A in G satisfies a polynomial of the form $x^d - 1$, which has distinct roots. Therefore each $\mu_A(x)$ has distinct roots and therefore each A is diagonalizable. If $Av = \lambda v$, with $v \neq 0$, then $A^d v = \lambda^d v$, so that $\lambda^d = 1$. Thus, the eigenvalues of A are complex roots of unity.

Friday, December 2. Given vector spaces V and U over the field F, we showed the tensor product $V \otimes_F U$ exists. This was done by first taking a vector space $\mathcal{H}$ having as its basis $\{(v, u) \mid (v, u) \in V \times U\}$ and forming the quotient space $\mathcal{H}/\mathcal{K}$, where $\mathcal{K}$ is the subspace generated by expressions of the form

(i) $(v_1 + v_2, u) - (v_1, u) - (v_2, u)$, for all $v_i \in V, u \in U$.
(ii) $(v, u_1 + u_2) - (v, u_1) - (v, u_2)$, for all $v \in V, u_i \in U$.
(iii) $\lambda(v, u) - (\lambda v, u)$, for all $v \in V, u \in U, \lambda \in F$.
(iv) $\lambda(v, u) - (v, \lambda u)$, for all $v \in V, u \in U, \lambda \in F$.

Upon taking $\phi : V \times U \to \mathcal{H}/\mathcal{K}$ defined by $\phi((v, u)) := (v, u) + \mathcal{K}$, we noted that ϕ is bilinear and the pair $(\mathcal{H}/\mathcal{K}, \phi)$ is a tensor product of V and W and therefore, by our convention, gives $V \otimes_F U$. Thus, we write $v \otimes u$, instead of $(v, u) + \mathcal{K}$. We then noted the crucial point that the elements of $V \otimes_F U$ are linear combinations of expressions of the form $v \otimes u$ and that it is therefore **not** the case that every element in the tensor product is of the form $v \otimes u$.

We ended class by proving the following two facts: (i) $0 \otimes u = 0 = v \otimes 0$ in $V \otimes_F U$; and (ii) If $\{v_\alpha\}_{\alpha\in A}$ is a basis for V and $\{u_\beta\}_{\beta\in B}$ is a basis for U, then $\{v_\alpha \otimes u_\beta\}_{\alpha\in A,\beta\in B}$ is a basis for $V \otimes_F U$. This enabled us to conclude that if V and W are finite dimensional, then $\dim(V \otimes_F U) = \dim(V) \cdot \dim(U)$.

Wednesday, November 30. We began class by recalling the quotient space V/W discussed in the previous lecture as well as some of its properties. We then recalled the statement of the first isomorphism proven at the end of the previous lecture. This was followed by stating the second and third isomorphisms theorems and sketching their proofs. The proofs of these theorems were applications of the First Isomorphisms theorem.

Theorem. Let $U \subseteq W, W_1, W_2$, be subspaces of the vector space V.

(i) Second Isomorphism Theorem: $(W_1 + W_2)/W_2 \cong W_1/(W_1 \cap W_2)$.
(ii) Third Isomorphism Theorem: W/U is a subspace of V/U and $(V/U)/(W/U) \cong V/W$.

After a brief discussion of bilinearity and its relation to familiar products, we defined a tensor product of F-vector spaces V and U to be a pair (P, ϕ) where P is a vector space over F and $\phi : V \times U \to P$ is a bilinear map satisfying: Given a bilinear map $h : V \times U \to W$, where W is a vector space over F, there exists a *unique* linear transformation $T : P \to W$ such that $T\phi = h$. Using the definition, we showed that tensor products are unique in the following sense: If (P', ϕ') is a tensor product of V and U, then there exists an isomorphism $T : P \to P'$ satisfying $T\phi = \phi'$. We therefore noted that the tensor product is denoted $V \otimes_F U$, or just $V \otimes U$, if the field F is understood. We ended class by noting that the construction of the tensor product will follow from the general principle that one can impose relations on elements in a vector space V by forming an appropriate quotient space V/W.

Monday, November 28. For a vector space V and subspace $W \subseteq V$, and $v \in V$, we defined the *coset* $v + W := \{v + w \mid w \in W\}$. We noted that if W is a line through the origin in $V = \mathbb{R}^2$, then $v + W$ is just a translate of W, i.e., a line through v parallel to W. Thus, algebraically, we can regard the abstract coset $v + W$ as a translate of W. We emphasized that $v + W$ is never a subspace unless $v + W = W$, and this latter condition holds if and only if $v \in W$. We also characterized the cosets of W as the equivalence classes resulting from the equivalence relation: $v_1 \sim v_2$ if and only if $v_1 - v_2 \in W$. This then enabled us to prove that V/W, the set of cosets of W, is a vector space under the operations: (i) $(v_1 + W) + (v_2 + W) := (v_1 + v_2) + W$ and (ii) $\lambda \cdot (v + W) := \lambda v + W$, is a vector space called the *quotient space* of V by W, or V *mod* W. The key point was showing the the operations on V/W were well-defined.

We then proved the following two theorems:

Theorem. Let V be a finite dimensional vector space and $W \subseteq V$ a non-zero subspace. If $w_1, \ldots, w_r$ is a basis for W and $w_1, \ldots, w_r, v_1, \ldots, v_s$ is a vector basis for V, then $v_1 + W, \ldots, v_s + W$ is a basis for V/W. In particular, the dimension of V/W is the dimension of V minus the dimension of W.

First Isomorphism Theorem. let $T : V \to U$ be a linear transformation between the vector spaces V and U and W denote the kernel of T. Then V/W is isomorphic to $\text{im}(T)$.

We showed that $\overline{T} : V/W \to \text{im}(T)$ given by $\overline{T}(v + W) := T(v)$ is the required isomorphism, the main point being that $\overline{T}$ is well-defined.

Monday, November 21. We began class by giving examples of and recalling the formula for powers of the Jordan block $J(\lambda, s)$, noting that $J(\lambda, s)^n$ is the $s \times s$ lower triangular matrix whose diagonal entries are λ^n and whose ith subdiagonal (below the main diagonal) consists of $\binom{n}{i}\lambda^{n-i}$. Thus, for example

$$J(\lambda, 3)^n = \begin{pmatrix} \lambda^n & 0 & 0 \\ n\lambda^{n-1} & \lambda & 0 \\ \binom{n}{2}\lambda^{n-2} & n\lambda^{n-1} & \lambda^n \end{pmatrix}.$$

We then calculated e^{Jt} for a Jordan block $J := J(\lambda, n)$, and t an indeterminate, as the $n \times n$ lower triangular matrix whose diagonal entries are $e^{\lambda t}$ and whose ith subdiagonal (below the main diagonal) consists of $\frac{t^i}{i!}e^{\lambda t}$. The matrix e^J is obtained by setting $t = 1$. Thus for example, when $J = J(\lambda, 3)$,

$$e^{Jt} = \begin{pmatrix} e^{\lambda t} & 0 & 0 \\ te^{\lambda t} & e^{\lambda t} & 0 \\ \frac{t^2}{2!}e^{\lambda t} & te^{\lambda t} & e^{\lambda t} \end{pmatrix}.$$

As before, once we know the form e^{Jt} takes for a Jordan block J, we have $e^{At} = Pe^{\tilde{J}t}P^{-1}$, where $A = P\tilde{J}P^{-1}$ and $\tilde{J}$ is the JCF of A.

We finished class by showing that if $\mathbf{X}'(t) = A \cdot \mathbf{X}(t)$ is an $n \times n$ system of first order linear differential equations, where $\mathbf{X}(t) = \begin{pmatrix} x_1(t) \\ \vdots \\ x_n(t) \end{pmatrix}$ and A is an $n \times n$ matrix over $\mathbb{R}$, then the general solution is given by $\mathbf{X}(t) = e^{At} \cdot \vec{\alpha}$, for $\vec{\alpha} \in \mathbb{R}^n$. If $\mathbf{X}(0)$ represents a set of initial conditions, then the solution to the system of equations is given by $\mathbf{X}(t) = e^{At} \cdot \mathbf{X}(0)$.

Friday, November 18. We reviewed the process for finding a p^{th} root of a nonsingular square matrix with entries in $\mathbb{C}$, as well as the inductive procedure for constructing the polynomials $p_n(x)$ given in the previous lecture. We then discussed how taking large powers of an elements in $\mathrm{M}_n(F)$ is straightforward for diagonalizable matrices A, since $A^t = PD(\lambda_1^t, \ldots, \lambda_n^t)P^{-1}$, for P the diagonalizing matrix. We then calculated powers of Jordan blocks, in order to see how to simplify calculating powers of matrices in general.

The discussion above was followed by the definition of the exponential of a matrix: Given $A \in \mathrm{M}_n(\mathbb{R})$, $e^A := \Sigma_{t=0}^{\infty} \frac{1}{t!} A^t$. We used the singular value decomposition to show that the (i, j) entries of the matrices in the sum defining e^A are absolutely convergent, so the definition of e^A makes sense. We noted that when A is diagonalizable, we obtain $e^A = PD(e^{\lambda_1}, \ldots, e^{\lambda_m})P^{-1}$, as expected. We finished class by noting - but not proving - that if

$$\begin{pmatrix} x_1'(t) \\ x_2'(x) \end{pmatrix} = A \cdot \begin{pmatrix} x_(t) \\ x_2(t) \end{pmatrix},$$

then the solution to the system of linear first order differential equations is given by $\begin{pmatrix} x_1(t) \\ x_2(t) \end{pmatrix} = e^{At} \cdot \begin{pmatrix} x_1(0) \\ x_2(0) \end{pmatrix}$.

Wednesday, November 16. We continued our discussion of finding roots and powers of diagonalizable matrices by finding that the matrix $\begin{pmatrix} -3 & -10 \\ 3 & 8 \end{pmatrix}$ is a square root of $\begin{pmatrix} -21 & -50 \\ 15 & 34 \end{pmatrix}$. We then turned our attention to proving the following theorem:

Theorem. Let A be a nonsingular $n \times n$ matrix over $\mathbb{C}$. The, for $q \geq 2$, there exists an $n \times n$ matrix over $\mathbb{C}$ such that $B^q = A$. In other words, every nonsingular $n \times n$ matrix over $\mathbb{C}$has a q^{th} root.

Before indicating a proof of the theorem, we found three cube roots of the matrix $A = \begin{pmatrix} 37 & -49 \\ 25 & -33 \end{pmatrix}$. We did this by first noting that $\chi_A(x) = (x-2)^2$ and the JCF of A is $J = \begin{pmatrix} 2 & 0 \\ 1 & 2 \end{pmatrix}$. We then set $M := \begin{pmatrix} 0 & 0 \\ \frac{1}{2} & 0 \end{pmatrix}$ and $B_0 := I_2 + \frac{1}{3} \cdot M$, so that $B_0^3 = I_2 + M$. Thus, $2 \cdot B_0^3 = J$. Thus, for $\omega := e^{\frac{2\pi i}{3}}$, the matrices $\sqrt[3]{2} \cdot B_0, \sqrt[3]{2}\omega \cdot B_0, \sqrt[3]{2}\omega^2 \cdot B_0$ are cube roots of J. For $P = \begin{pmatrix} 3 & 7 \\ 2 & 5 \end{pmatrix}$, $A = PJP^{-1}$, so that if B is any one of the three cube roots of J, PBP^{-1} is a cube root of A.

The proof of the theorem relied on the following proposition, whose proof we did not present in class.

Proposition. Fix a positive integer $p \geq 2$. For $n \geq 2$, there exists polynomials $p_n(x) \in \mathbb{Q}[x]$ such that:

(i) degree $p(x) = n - 1$.
(ii) $p_n(x) = p_{n-1}(x) + \alpha_n x^{n-1}$, for $\alpha \in \mathbb{Q}$.
(iii) The constant term of $p_n(x) = 1$.
(iv) $p_n(x)^p = (1 + x) + x^n q_n(x)$, with $q_n(x) \in \mathbb{Q}[x]$.

Proof. Induct on n. It is easy to check that $p_2(x) = 1 + \frac{1}{p}x$ satisfies the conclusions of the lemma. Assume $p_{n-1}(x)$ exists. Write $p_n(x) = p_{n-1}(x) + \alpha_n x^{n-1}$, with α_n to be determined. If we find α_n such that (iv) holds, then statements (i)-(iii) will also hold, by induction. We have

$$\begin{aligned}
p_n(x)^q &= (p_{n-1}(x) + \alpha_n x^{n-1})^p \\
&= p_{n-1}(x)^p + \binom{p}{1} p_{n-1}(x)^{p-1}\alpha_n x^{n-1} + \cdots + \binom{p}{p-1} p_{n-1}(x)\alpha_n^{p-1}x^{(n-1)(p-1)} + \alpha_n^p x^{(n-1)p} \\
&= (1+x) + x^{n-1}q_{n-1}(x) + \binom{p}{1} p_{n-1}(x)^{p-1}\alpha_n x^{n-1} + \cdots + \binom{p}{p-1} p_{n-1}(x)\alpha_n^{p-1}x^{(n-1)(p-1)} + \alpha_n^p x^{(n-1)p}
\end{aligned}$$

Note that the coefficient of x^{n-1} in the last equation above is $\beta + p\alpha_n$, where β is the constant term of $q_{n-1}(x)$, since the constant term of $p_{n-1}(x)$ equals 1. Thus, if we set $\alpha_n = -\frac{\beta}{p}$, the x^{n-1} term drops out from the expression above and all remaining terms, except the terms in $(1+x)$, have degree greater than or equal to n. Thus, we may write $p_x(x)^p = (1+x) + x^n q_n(x)$, as required.

With the proposition in hand, we were able to prove the theorem by first finding a pth root of a single Jordan block $J(\lambda, n)$ by noting that if $M := \lambda^{-1}C$, where C is the companion matrix of x^n and $B_0 := p_n(M)$, with $p_n(x)$ as in the proposition implies that $B_0^p = I_n + M$. Thus, as in the example, $\lambda \cdot B_0^p = J(\lambda, n)$. Therefore, for any ω, pth root of λ, we have $(\omega \cdot B_0)^p = J(\lambda, n)$. We then showed that it follows readily that we can find a pth root B of any matrix J in JCF, so that if $A = PJP^{-1}$, PBP^{-1} is a pth root of A.

Monday, November 14. We finished our discussion concerning the JCF by noting that the JCF of a matrix $A \in \mathrm{M}_n(F)$ or operator $T \in \mathcal{L}(V, V)$ can be found as follows:

(i) First calculate $\chi_A(x) = (x-\lambda_1)^{f_1}\cdots(x-\lambda_r)^{f_r}$.

(ii) For each $1 \le i \le r$, calculate $t_{i,j} := \text{nullity}(A-\lambda_i)^j$ until two consecutive terms $t_{i,j}$ are equal. Set e_i to be the first j such that $t_{i,j} = t_{i,j+1}$.

(iii) For e_i as in (ii), we have $\mu_A(x) = (x-\lambda_1)^{e_1}\cdots(x-\lambda_r)^{e_r}$.

(iv) The Jordan canonical form of A is given by $\tilde{A} = \begin{pmatrix} A_1 & 0 & \cdots & 0 \\ 0 & A_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & A_r \end{pmatrix}$ where each

$A_i = \begin{pmatrix} J(\lambda_i, e_{i,1}) & 0 & \cdots & 0 \\ 0 & J(\lambda_i, e_{i,2}) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & J(\lambda_i, e_{i,s_i}) \end{pmatrix}$, with $s_i := t_{i,1}$ and $e_{i,1} = e_1$ and $t_{i,1} - (t_{i,2} - t_{i,1})$ equals the number of blocks of size one, ..., $(t_{i,j+1} - t_{i,j}) - (t_{j+1} - t_{j+1})$ equals the number of blocks of size j, ..., $(t_{i,e_i} - t_{i,e_i-1})$ is the number of blocsk of size e_i.

We then verified directly the formulas for the number of blocks of a given size for the 9×9 matrix with Jordan blocks $J(\lambda,3), J(\lambda,2), J(\lambda,2), J(\lambda,1), J(\lambda,1)$. This was followed by using the algorithm above to find the JCF of the matrix $A = \begin{pmatrix} 3 & 4 & 2 \\ -2 & -3 & -1 \\ -4 & -4 & -2 \end{pmatrix}$ and also a Jordan basis for the corresponding operator on $\mathbb{R}^3$.

We ended class by beginning a discussion of a new topic, namely finding powers and roots of square matrices. We showed that the problem of calculating powers (over any field) and roots (over $\mathbb{C}$) of diagonalizable matrices is fairly straightforward: For example, over $\mathbb{C}$, if $P^{-1}AP = D(\lambda_1, \ldots, \lambda_n)$ and $\gamma_i \in \mathbb{C}$ satisfy $\gamma_i^c = \lambda_i$, for each i, then $B^c = A$ for $B := PD(\gamma_1, \ldots, \gamma_n)P^{-1}$, and we call B a cth root of A. This is possible, since for any integer $c \ge 2$, and $z \in \mathbb{C}$, z has c distinct cth roots.

Friday, November 11. We reviewed the details of the discussion from the previous lecture in which we were able to describe formulas counting the number of Jordan blocks and the numbers of Jordan blocks of a given size in the JCF of an operator or matrix.

Wednesday, November 9. We began class by writing down all possible JCFs for 3×3 matrices all of whose eigenvalues are in F. We also wrote all possible JCFs 4×4 matrices A satisfying $\chi_A(x) = (x-\lambda_1)^2(x-\lambda_2)^2$. We then proved the following proposition:

Proposition. Let $T \in \mathcal{L}(V,V)$ or $A \in \mathrm{M}_n(F)$ have JCF $\tilde{A} = \begin{pmatrix} A_1 & 0 & \cdots & 0 \\ 0 & A_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & A_r \end{pmatrix}$ where each $A_i = \begin{pmatrix} J(\lambda_i, e_{i,1}) & 0 & \cdots & 0 \\ 0 & J(\lambda_i, e_{i,2}) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & J(\lambda_i, e_{i,s_i}) \end{pmatrix}$. Then for each $1 \le i \le r$, $s_i = \dim(E_{\lambda_i})$.

Thus the number of Jordan blocks associated to λ_i equals the dimension of the corresponding eigenspace E_{λ_i}. This was followed by a discussion of how to calculate the size of the Jordan blocks appearing in each A_i. The discussion was carried our for an operator T with $\mu_T(x) = x^e$. Letting t_i denote the dimension of the kernel of T^i we saw that $t_{i+1} - t_i$ equals the number of Jordan blocks of size greater than i. It follows that the number of Jordan blocks whose size equals i is $(t_i - t_{i-1}) - (t_{i+1} - t_i)$.

Monday, November 7. We began class by reviewing how the JCF of a matrix A (or elements of $\mathcal{L}(V,V)$) with $\mu_A(x) = (x-\lambda)^e$ can be derived from the RCF for a nilpotent matrix or nilpotent transformation. In this case, there is an invertible matrix P such that $P^{-1}AP$ is block diagonal with Jordan blocks of the form $J(\lambda, e_i)$. We then stated and derived the general form of the:

Jordan Canonical Form Theorem. Suppose $\dim(V) < \infty$ and $T \in \mathcal{L}(V,V)$.
Write $\mu_T(x) = (x-\lambda_1)^{e_1} \cdots (x-\lambda_r)^{e_r}$, where each $\lambda_i \in F$. Then there exists a basis $B \subseteq V$, and for each $1 \le i \le r$, $e_i = e_{i1} \ge \cdots \ge e_{is_i}$ such that the matrix of T with respect to B has the form $\begin{pmatrix} A_1 & 0 & \cdots & 0 \\ 0 & A_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & A_r \end{pmatrix}$, where each $A_i = \begin{pmatrix} J(\lambda_i, e_{i,1}) & 0 & \cdots & 0 \\ 0 & J(\lambda_i, e_{i,2}) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & J(\lambda_i, e_{i,s_i}) \end{pmatrix}$.

We also noted that in some presentations, a Jordan block $J(\lambda, e)$ is written as an $e \times e$ matrix with λ down the diagonal and 1s above the diagonal. This was followed by showing that the JCF of the matrix $A = \begin{pmatrix} 2 & 1 & 3 \\ 5 & 3 & 6 \\ -2 & -1 & -2 \end{pmatrix}$ is $\tilde{A} := \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix}$ and $P = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 5 & 3 \\ 0 & -2 & -1 \end{pmatrix}$ satisfies $P^{-1}AP = \tilde{A}$.

Friday, November 4. We began class with a couple of observation that enabled us to prove the following uniqueness theorem:

Uniqueness of the Rational Canonical Form. Let $T \in \mathcal{L}(V,V)$ and suppose there exist bases for V leading to the following invariant factor rational canonical forms for T:

$$B = \begin{pmatrix} B_1 & 0 & \cdots & 0 \\ 0 & B_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & B_r \end{pmatrix} \quad \text{and} \quad C = \begin{pmatrix} C_1 & 0 & \cdots & 0 \\ 0 & C_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & C_s \end{pmatrix}$$

where $B_i = C(f_i(x))$, with $f_r(x) | \cdots | f_1(x) = \mu_T(x)$ and $C_i = C(g_i(x))$ with $g_s(x) | \cdots | g_1(x) = \mu_T(x)$. Then $r = s$ and each $B_i = C_i$.

The proof proceeded roughly as follows: By definition, $C_1 = B_1$. Since the matrices B and C are similar, $f_2(B)$ and $f_2(C)$ are similar. We have

$$f_2(B) = \begin{pmatrix} f_2(B_1) & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \end{pmatrix} \quad \text{and} \quad f_2(C) = \begin{pmatrix} f_2(B_1) & 0 & \cdots & 0 \\ 0 & f_2(C_2) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & f_2(C_s) \end{pmatrix}$$

Since the ranks of these latter matrices are the same, we must have $f_2(C_2) = 0$. Thus, $g_2(x)|f_2(x)$. By symmetry, $f_2(x)|g_2(x)$. Thus, $f_2(x) = g_2(x)$, so $B_2 = C_2$. One continues in a similar fashion to show $B_i = C_i$, for all i, and in particular, $r = s$.

We then began a discussion of the *Jordan canonical form*, by first looking at the elementary divisor rational canonical form of a nilpotent matrix. From there we were able to see that if $\mu_T(x) = (x-\lambda)^e$, then there exists a basis B for V such that $[T]_B^B = \begin{pmatrix} J(\lambda, e_1) & 0 & \cdots & 0 \\ 0 & J(\lambda, e_2) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & J(\lambda, e_r) \end{pmatrix}$, $e_1 \geq \cdots \geq e_r$ and $J(\lambda, e_i) := C(x^{e_1}) + \lambda I_{e_i}$ is an $e_i \times e_i$ *Jordan block* associated with λ.

Wednesday, November 2. We started class by restating the elementary divisor rational canonical form theorem, and then proceeded to calculate the elementary divisor RCF $\tilde{A}$ of the matrix $A = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$. We then informally discussed how the invariant factors of $T \in \mathcal{L}(V, V)$ can be recovered from its elementary divisors. We also calculated P such that $P^{-1}AP = \tilde{A}$. We then considered the two 3×3 matrices having the same characteristic polynomial and calculated both forms of the RCF for the first of these and the invariant factor form of the RCF for the second, noting - as expected - that the RCFs for the matrices were different, even though they had the same characteristic polynomial.

Monday, October 31. We continued our discussion of the Rational Canonical Form Theorem, first stating the matrix form, of the invariant factor version of the theorem:

Rational Canonical Form Theorem for Matrices. Let $A \in \mathrm{M}_n(F)$. Then there exist $f_1(x), \ldots, f_t(x)$ in $F[x]$ and an invertible matrix $P \in \mathrm{M}_n(F)$ such that,

(i) $f_1(x)|f_2(x)|\cdots|f_t(x) = \mu_T(x)$

(ii) $P^{-1}AP$ has the form $\begin{pmatrix} A_1 & 0 & \cdots & 0 \\ 0 & A_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & A_t \end{pmatrix}$, where each $A_i = C(f_i(x))$, the companion matrix of $f_i(x)$.

We followed this by calculating the rational canonical form $\tilde{A}$ for the matrix $A = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$ and an invertible matrix P such that $P^{-1}AP = \tilde{A}$.

We then stated and proved the elementary divisor version of the RCFT

Elementary Divisor Form of the Rational Canonical Form Theorem. Suppose $\dim(V) < \infty$ and $T \in \mathcal{L}(V, V)$. Write $\mu_T(x) = p_1(x)^{e_1} \cdots p_r(x)^{e_r}$, where each $p_i(x)$ is irreducible over F. Then there exists a basis $B \subseteq V$, and for each $1 \leq i \leq r$, $e_i = e_{i1} \geq \cdots \geq e_{is_i}$ such that the matrix of T with respect to B has the form $\begin{pmatrix} A_1 & 0 & \cdots & 0 \\ 0 & A_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & A_r \end{pmatrix}$, where each $A_i = \begin{pmatrix} C(p_i(x)^{e_{i1}}) & 0 & \cdots & 0 \\ 0 & C(p_i(x)^{e_{i2}}) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & C(p_i(x)^{e_{i,s_i}}) \end{pmatrix}$. The polynomials $\{p_i(x)^{e_{ij}}\}$ are called the *elementary divisors* of T.

We ended class with the following important theorem, which follows from the elementary divisor form of the RCFT:

Theorem. Suppose $\dim(V) < \infty$ and $T \in \mathcal{L}(V,V)$. Then $\mu_T(x)$and $\chi_T(x)$ have the same irreducible factors. In particular, with the notation as in the previous theorem, $\chi_T(x) = p_1(x)^{f_1} \cdots p_r(x)^{f_r}$, where each $f_i = e_{i1} + \cdots + e_{is_i}$.

Friday, October 28. We began class by carefully going through the example provided at the end of the previous lecture showing that a cyclic subspace of V need not have a T-invariant complement. We then showed that if V is finite dimensional and $T \in \mathcal{L}(V,V)$, then V is a direct sum of cyclic subspaces. This was stated in a way that enabled us to immediately deduce the following form of the:

Rational Canonical Forma Theorem. Suppose V is a finite dimensional vector space over the field F and $T \in \mathcal{L}(V,V)$. Then there exist $f_1(x), \ldots, f_t(x) \in F[x]$ and a basis $B \subseteq V$ such that:

(i) $f_1(x)|f_2(x)|\cdots|f_t(x) = \mu_T(x)$

(ii) The matrix of T with respect to B has the form $\begin{pmatrix} A_1 & 0 & \cdots & 0 \\ 0 & A_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & A_t \end{pmatrix}$, where each $A_i = C(f_i(x))$, the companion matrix of $f_i(x)$.

We ended class by finding the rational canonical form for the transformation $T : \mathbb{R}^3 \to \mathbb{R}^3$ satisfying $[T]_E^E = \begin{pmatrix} -2 & 0 & 0 \\ -1 & -4 & -1 \\ 2 & 4 & 0 \end{pmatrix}$, where $E = \{e_1, e_2, e_3\}$ is the standard basis for $\mathbb{R}^3$. We found that for the basis $\{e_1, \begin{pmatrix} 0 \\ -1 \\ 2 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}\}$, the matrix of T with respect to B is $\begin{pmatrix} 0 & -4 & 0 \\ 1 & -4 & 0 \\ 0 & 0 & -2 \end{pmatrix}$, the rational canonical form associated to T.

Wednesday, October 28. We began class by recalling that our immediate goal is the following: Given a finite dimensional vector space V and $T \in \mathcal{L}(V,V)$, we can write V as a direct sum of cyclic subspaces with respect to T. We also noted that our previous lecture established the existence of a maximal vector of V with respect to T. We then presented the following:

Key Theorem. Let V be a finite dimensional vector space and $T \in \mathcal{L}(V,V)$. Suppose $v \in V$ is a maximal vector. Then there exists a T-invariant subspace $U \subseteq V$ such that $V = \langle T, v\rangle \oplus U$.

The proof we gave of this theorem is a transcription to notation used in our class of a very nice proof due to M. Geck, which in turn was based upon a proof given by H.G. Jacob.

Proof of the Key Theorem. Suppose $n = \dim(V)$, $d := \deg(\mu_T(x))$ and $v \in V$ is a maximal vector. Thus, $v, T(v), \ldots, T^{d-1}(v)$ is a basis for $\langle T, v\rangle$. Extend these vectors to a basis B for V. For $u \in V$, we let u_d denote the coefficient of $T^{d-1}(v)$ when we write u in terms of the basis B. In our matrix notation, u_d is the dth coordinate of the column vector $[u]_B \in F^n$, which will write as $([u]_B)_d$. Now set

$$U := \{u \in V \mid T^j(u)_d = 0, \text{ for all } 0 \le j \le d-1\}.$$

We show that this U works in the following steps.

(1) U is a subspace of V: Take $u_1, u_2 \in U$, $\lambda \in F$ and $0 \le j \le d-1$,

$$T^j(\lambda u_1 + u_2)_d = (\lambda T^j(u_1) + T^j(u_2))_d = \lambda T^j(U_1)_d + T^j(u_2)_d = 0 + 0 = 0,$$

which shows that U is a subspace.

(2) $\langle T, v\rangle \cap U = 0$: Suppose $u = \alpha_0 v + \alpha_1 T(v) + \cdots + \alpha_{d-1}T^{d-1}(v) \in \langle T, v\rangle \cap U$. Since $u \in U$, $\alpha_{d-1} = 0$. Thus, $u = \alpha_0 v + \alpha_1 T(v) + \cdots + \alpha_{d-2}T^{d-2}(v)$. The coefficient of $T^{d-1}(v)$ in $T(u)$ is α_{d-2}. Since $u \in U$, it follows that $\alpha_{d-2} = 0$. Continuing in this way, one shows that each $\alpha_j = 0$, so that $u = 0$, as required.

(3) $V = \langle T, v\rangle \oplus U$: Since $\langle T, v\rangle \cap U = 0$, it suffices to show that $V = W + U$. We also have,

$$\dim(\langle T, v\rangle + U) = \dim(\langle T, v\rangle) + \dim(U),$$

since $\langle T, v\rangle \cap U = 0$. Now, $\dim(\langle T, v\rangle) = d$. We claim $\dim(U) \geq n - d$. If the claim holds, then

$$\dim(\langle T, v\rangle + U) = d + \dim(U) \geq d + (n - d) \geq n,$$

from which it follows that $\dim(\langle T, v\rangle + U) = n$, so $V = \langle T, v\rangle + U$, and thus, $V = \langle T, v\rangle \oplus U$. For the claim, if we set $A := [T]_B^B$, it follows that $A^j = [T^j]_B^B$, for $0 \leq j \leq d-1$. Thus, if $u \in U$, then $0 = T^j(u)_d = ([T^j(u)]_B)_d = (A^j \cdot [u]_B)_d$. Thus, $u \in U$ if and only if the dth row of A^j times $[u]_B$ is zero for $0 \leq j \leq d-1$. It follows that $u \in U$ if and only if $[u]_B$ is in the solution space of a system of d equations in n unknowns. Since the latter must have dimension at least $n - d$, it follows that $\dim(U) \geq n - d$, as required.

(4) U is T-invariant: Take $u \in U$. We must show $T(u) \in U$, i.e., $T^j(T(u)))_d = 0$, for $0 \leq j \leq d-1$. For $0 \leq j \leq d-2$, this follows because $u \in U$. On the other hand, v is a maximal vector, so $r := \dim(\langle T, u\rangle) \leq d$. Thus, we may write $T^{d-1}(T(u)) = T^d(u) = \alpha_0 u + \cdots + \alpha_{r-1}T^{r-1}(u)$, so that

$$T^d(u)_d = \alpha_0 u_d + \cdots + \alpha_{r-1}T^{r-1}(u)_d = 0 + \cdots + 0 = 0,$$

which shows $T(u) \in U$, and thus completes the proof of the key theorem.

We finished class by showing that if $T : \mathbb{R}^3 \to \mathbb{R}^3$ is the linear transformation defined by $A = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}$, and $v := \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$, then v is not a maximal vector for T and $\langle T, v\rangle$ does not have a T-invariant complement as a subspace of $\mathbb{R}^3$.

Monday, October 24. For a finite dimensional vector space V, $0 \neq v \in V$ and $T \in \mathcal{L}(V, V)$, we continued our discussion of the cyclic subspace $\langle T, v\rangle$, keeping in mind that our ultimate goal, the Rational Canonical Form theorem, requires us to show that V is a direct sum of cyclic subspaces. In particular, we finished the proof of the proposition started at the end of the previous lecture. We then defined $v \in V$ to be a *maximal vector* of V (with respect to T) if $\deg(\mu_{T,v}(x)) = \deg(\mu_T(x))$, or equivalently, $\dim(\langle T, v\rangle) = \deg(\mu_T(x)))$. These conditions mean that v is a maximal vector if and only if $\dim(\langle T, v\rangle)$ is maximal among the dimensions of cyclic subspaces of R. We then proved the following:

Theorem. Let V be a finite dimensional vector space and $T \in \mathcal{L}(V, V)$. Then T admits a maximal vector.

The proof began by writing $\mu_T(x) = p_1(x)^{e_1} \cdots p_r(x)^{e_r}$, with each $p_i(x)$ irreducible over F and and decomposing $V = W_1 \oplus \cdots \oplus W_r$, with $W_i = \ker(p_i(T)^{e_i})$. We then showed that: (i) each W_i admits a maximal vector w_i for $T|_{W_i}$ and (ii) $v := w_1 + \cdots + w_r$ is a maximal vector for V. The proofs relied on facts that $p_i(x)^{e_i}$ is the minimal polynomial for $T|_{W_i}$ and $\mu_{T|_{W_i}, w_i}$ divides the minimal polynomial of $T|_{W_i}$.

Friday, October 21. We began class by stating that our next goal is the Rational Canonical Form theorem which states that given an operator T on the finite dimensional vector space V, there exists a basis $B \subseteq V$ such that $[T]_B^B$ is in block diagonal form, with each block a companion matrix. This was followed by recalling the primary decomposition theorem from last time and the general fact that if our finite dimensional vector space V can be written $V = W_1 \oplus \cdots \oplus W_s$ as a direct sum of T-invariant subspaces, then if $B_i \subseteq W_i$ is a basis, and $B = B_1 \cup \cdots \cup B_r$, $[T]_B^B$ is block diagonal, with blocks $[T|_{W_i}]_{B_i}^{B_i}$. We then used the primary decomposition to prove the following theorem,

Theorem. Let V be a finite dimensional vector space and $T \in \mathcal{L}(V, V)$. Then T is diagonalizable if and only if there exist distinct $\lambda_1, \ldots, \lambda_r \in F$ such that $\mu_T(x) = (x - \lambda_1) \cdots (x - \lambda_r)$.

We then noted that if V happens to have a basis of the form $B := \{v, T(v), \ldots, T^{n-1}(v)\}$, for some $v \in V$, then $[T]_B^B$ is a companion matrix. This points the way to the following strategy for proving the Rational Canonical Form theorem, namely decompose $V = W_1 \oplus \cdots \oplus W_s$ where each W_s is T-invariant and has a basis for the form $B_i := \{v_i, T(v_i), \ldots, T^{n_i-1}(v_i)\}$, for some $v_i \in W_i$. This then lead to the definition of the

$\langle T, v\rangle$, the *cyclic subspace generated by* v, defined as the subspace of V generated by the set $\{T^j(v) \mid j \geq 0\}$. We finished class by proving the first three statements in the following proposition.

Proposition. For $T \in \mathcal{L}(V, V)$ and $0 \neq v \in V$, suppose $e \geq 1$ is the degree of $\mu_{T,v}(x)$.

(i) $\langle T, v\rangle$ is a T-invariant subspace of V.
(ii) $\{v, T(v), \dots, T^{e-1}(v)\}$ is a basis for $\langle T, v\rangle$.
(iii) $\dim(\langle T, v\rangle) = e$.
(iv) $\mu_{T,v}(x) = \mu_{T|_{\langle T,v\rangle}}$.

Wednesday, October 19. We began class by reviewing some of the facts regarding factorization in $F[x]$ discussed in the last lecture. This was followed by a proof that the division algorithm holds in $F[x]$ and a generalized version of Bezout's Principle in the following form: *If $f_1(x), \dots, f_n(x) \in F[x]$ have no common divisor, then there exist $a_1(x), \dots, a_n(x) \in F[x]$ such that $1 = a_1(x)f_1(x) + \cdots + a_n(x)f_n(x)$.* This fact played a key role in the proof of the following *primary decomposition* theorem:

Theorem. Let V be a finite dimensional vector space over the field F and $T \in \mathcal{L}(V, V)$. Factor the minimal polynomial of T as $\mu_T(x) = p_1(x)^{e_1} \cdots p_r(x)^{e_r}$, where each $p_i(x) \in F[x]$ is irreducible, and set $W_i := \text{kernel}(p_i(T)^{e_i})$. Then:

(i) $V = W_1 \oplus \cdots \oplus W_r$.
(ii) Each W_i is T-invariant.
(iii) $p_i(x)^{e_i}$ is the minimal polynomial of $T|_{W_i}$.

We ended class by noting that a crucial consequence of the theorem is the following observation: Preserving the notation in the theorem, let $B_i \subseteq W_i$ be a basis for W_i, so that $B = B_1 \cup \cdots \cup B_r$ is a basis for B. If we write $A = [T]_B^B$ and $A_i = [T|_{W_i}]_{B_i}^{B_i}$, then:

$$A = \begin{pmatrix} A_1 & 0 & \cdots & 0 \\ 0 & A_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & A_k \end{pmatrix}$$

is block diagonal. Thus, if we can put each A_i into a particular form, then A will be a block diagonal matrix consisting of blocks of a particular form.

Monday, October 17. We began class by reviewing the statements of the two versions of the Singular Value Theorem presented in the previous lecture. We then stated, but did not prove, the following facts associated to the Singular Value Theorem. In the statements of the facts below, $\sigma_1, \geq \cdots \geq \sigma_r > 0$ are the singular values of the real $m \times n$ matrix A:

(i) $\sigma_1 = \max\{||A \cdot v|| \mid v \in \mathbb{R}^n \text{ and } ||v|| \leq 1\}$.
(ii) Given a system of equations $A \cdot X = \mathbf{b}$, the minimum value of $||A \cdot \mathbf{x}_0 - \mathbf{b}||$ is obtained when $\mathbf{x}_0 = A^\dagger \cdot \mathbf{b}$, where $A^\dagger = P\Sigma^{-1}Q^*$ is the *pseudo-inverse* of A. Here Σ^{-1} means the $n \times m$ matrix with $\frac{1}{\sigma_1}, \dots, \frac{1}{\sigma_r}, 0, \dots, 0$ down its main diagonal and zeros elsewhere.
(iii) Consider the systems of equations $A \cdot X = \mathbf{b}$ and $A \cdot X = \mathbf{b}_0$, with $||\mathbf{b} - \mathbf{b}_0||$ small. If $\mathbf{x}$ and $\mathbf{x}_0$ are solutions to these systems, then it need not be the case that $||\mathbf{x} - \mathbf{x}_0||$ is comparably small. However, if $\frac{\sigma_1}{\sigma_r}$ is sufficiently small, then generally the two solutions are close to one another. $\frac{\sigma_1}{\sigma_r}$ is called the *condition number* of A.

We then noted that our next goal is to present the Rational and Jordan canonical forms for linear operators acting on a finite dimensional vector space. Until further notice, our underlying field F will be an arbitrary field. We then had a lengthy discussion about factorization properties in the ring $F[x]$ of polynomials with coefficients in F. The underlying theme was that familiar properties holding in $\mathbb{Z}$ also hold in $F[x]$ because the properties in question follow in $\mathbb{Z}$ from the division algorithm. Since $F[x]$ also has a division algorithm, the same proofs work in the latter setting. Thus our discussion verified the following properties:

(i) Every non-constant polynomial in $F[x]$ can be written as a product of irreducible polynomials in $F[x]$.

(ii) Given non-constant polynomials $f(x), g(x) \in F[x]$, the *greatest common divisor* $d(x)$ of $f(x), g(x)$ exists, where $d(x)$ denotes the monic polynomial in $F[x]$ of largest degree dividing both $f(x)$ and $g(x)$.

(iii) For $d(x)$ as in (ii), $d(x)$ is the last non-zero remainder, when we iterate the division algorithm on $f(x), g(x)$ as follows: Assuming $\deg(g(x)) \geq \deg(f(x))$, write $g(x) = f(x)h(x) + r(x)$, with $r(x) = 0$ or $\deg(r(x)) < \deg(f(x))$. If $r(x) = 0$, $d(x) = f(x)$. Otherwise, write $f(x) = r(x)h_2(x) + r_2(x)$, where $r_2(x) = 0$ or $r_2(x)$ has degree less than $r(x)$. If the former, $r(x)$ is the GCD of $f(x), g(x)$. If the latter, continue the algorithm by dividing $r(x)$ by $r_2(x)$ Do this until we achieve a last non-zero remainder, which we noted was $d(x)$.

(iv) Bezout's Principle: With the notation in (ii) and (iii), there exist $a(x), b(x) \in F[x]$ such that $d(x) = a(x)f(x) + b(x)g(x)$.

We also noted that the factorization in (i) is unique, up to order of irreducible factors and multiplication by elements of F. This is left as a homework problem.

Friday, October 14. We presented the Singular Value Theorem in the following forms:

Singular Value Theorem for Linear Transformations. Let $T \in \mathcal{L}(V, W)$, where V and W are finite dimensional inner product spaces, with $\dim(V) = n$ and $\dim(W) = m$. Then there exist orthonormal bases $B_V \subseteq V$ and $B_W \subseteq W$, $r > 0$, and real numbers $\sigma_1 \geq \cdots \sigma_r > 0$ such that $[T]_{B_V}^{B_W} = \Sigma$, where Σ is an $m \times n$ diagonal matrix whose main diagonal entries are $\sigma_1, \ldots, \sigma_r, 0, \ldots, 0$, where the number of zeros down the main diagonal equals $\min\{n, m\} - r$. The real numbers $\sigma_1, \ldots, \sigma_r$ are called the *singular values* of T.

Singular Value Theorem for Matrices. Let A be an $m \times n$ matrix over $F = \mathbb{R}$ or $\mathbb{C}$. Then there exist a unitary matrix $Q \in M_m(F)$ and a unitary matrix $P \in M_n(F)$ such that $Q^*AP = \Sigma$, where Σ is an $m \times n$ diagonal matrix with $\sigma_1, \ldots, \sigma_r, 0, \ldots, 0$ down its main diagonal. Here r is the rank of A. Moreover, $\sigma_1 \geq \cdots \geq \sigma_r$ are positive real numbers called the *singular values* of A.

For the proofs, we first noted that the concept of adjoint can be extended to $T \in \mathcal{L}(V, W)$, namely, there exists $T^* \in \mathcal{L}(W, V)$ such that $\langle T(v), w \rangle_W = \langle v, T^*(w) \rangle_V$, for all $v \in V$ and $w \in W$. We also noted that T^* has most of the familiar properties of T^* when T is a linear operator. The key idea behind the proof was to use the fact that T^*T is a self-adjoint operator and therefore orthogonally diagonalizable. The non-zero eigenvalues $\lambda_1, \ldots, \lambda_r$ of T^*T are all positive real numbers and we take $\sigma_i = \sqrt{\lambda_i}$, for $1 \leq i \leq r$. If $B_V = \{v_1, \ldots, v_n\}$ is the orthonormal basis of eigenvectors of T^*T and $B_W = \{u_1, \ldots, u_m\}$ is the orthonormal basis of F^m obtained by extending $\frac{1}{\sigma_1}Tv_1, \ldots, \frac{1}{\sigma_r}Tv_r$ to an orthonormal basis of F^m, then $[T]_{B_V}^{B_W} = \Sigma$. For the matrix version, P is the matrix whose columns are the v_i and Q is the matrix whose columns are the u_j.

We ended class by finding the required P, Q and Σ for $A = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 0 & 1 \end{pmatrix}$.

Wednesday, October 12. We began class by recalling the three spectral theorems presented thus far: (i) Self adjoint operators are orthogoanlly diagonalizable; (ii) A linear operator on a real inner product space is self-adjoint if and only if it is orthogonaliy diagonalizable; (iii) A linear operator on a complex inner product space is normal if and only if it is orthogonally diagonalizable.

This leaves open the question of characterizing normal operators on a real inner product space that are normal, but not self-adjoint. For this, we first noted that $A = \begin{pmatrix} \alpha & \beta \\ -\beta & \alpha \end{pmatrix} \in \mathrm{M}_n(\mathbb{R})$ is a normal matrix. This then lead to the following proposition, whose proof was a straight forward calculation, using the fact that if T is a normal operator on an inner product space, $||T(v)|| = ||T^*(v)||$, for all $v \in V$.

Proposition. Let $T \in \mathcal{L}(V, V)$, where V is a two-dimensional inner product space over $\mathbb{R}$. Then the following are equivalent:

(i) T is normal, but not self-adjoint.

(ii) There exists an orthonormal basis $B \subseteq V$ such that $[T]_B^B = \begin{pmatrix} \alpha & \beta \\ -\beta & \alpha \end{pmatrix}$, with $\alpha, \beta \in \mathbb{R}$ and $\beta > 0$.

We then gave the matrix version: if $A \in \mathrm{M}_2(\mathbb{R})$ is not symmetric, then A is normal if and only if there exists an orthogonal matrix $Q \in \mathrm{M}_n(\mathbb{R})$ such that $Q^tAQ = \begin{pmatrix} \alpha & \beta \\ -\beta & \alpha \end{pmatrix}$, with $\alpha, \beta \in \mathbb{R}$ and $\beta > 0$.

This was followed by stating (but not proving) the following theorem:

Theorem. Let V be a finite dimension inner product space over $\mathbb{R}$ and $T \in \mathcal{L}(V,V)$ such that T is not self-adjoint. Then T is a normal operator if and only if there exists an orthonormal basis $B \subseteq V$ such that $[T]_B^B$ is block diagonal, with blocks $D, A_1, \ldots, A_t$, where D is a diagonal matrix with entries in $\mathbb{R}$ and each $A_i = \begin{pmatrix} \alpha_i & \beta_i \\ -\beta_i & \alpha_i \end{pmatrix}$, with $\alpha_i, \beta_i \in \mathbb{R}$ and $\beta_i > 0$.

Note that it may be the case that D does not appear if T does not have any real eigenvalues. We ended class with a discussion using the Fundamental Theorem of Algebra to show that any non-constant polynomial in $\mathbb{R}[x]$ can be written as a product of linear and quadratic polynomials with coefficients in $\mathbb{R}$, and in particular, an irreducible polynomial with coefficients in $\mathbb{R}$ has degree less than or equal to two.

Friday, October 7. We finished the proof of the Complex Spectral Theorem begun in the previous lecture and followed this by stating and proving the matrix version of this theorem:

Complex Spectral Theorem for Matrices. Let $A \in \mathrm{M}_n(\mathbb{C})$. Then A is normal if and only if there exists a unitary matrix $Q \in \mathrm{M}_n(\mathbb{C})$ such that $Q^*AQ = D$, a diagonal matrix.

Wednesday, October 5. We began class by recalling what it means for $T \in \mathcal{L}(V,V)$ to be a normal operator or $A \in \mathrm{M}_n(F)$ to be a normal matrix. We then proved the following proposition which illustrates why normal operators are interesting:

Proposition. Let V be a finite dimensional inner product space over $\mathbb{C}$ and $T \in \mathcal{L}(V,V)$. Then T is a normal operator if and only if $||T(v)|| = ||T^*(v)||$, for all $v \in V$.

We then moved on to proving two lemmas that are key for the proof of the Complex Spectral Theorem.

Lemma 1. Let $T \in \mathcal{L}(V,V)$ be a normal operator on the finite dimensional complex inner product space. Then there exists $v \in V$, a common eigenvector for T and T^*. Moreover, if $T(v) = \lambda v$, then $T^*(v) = \overline{\lambda} v$.

Lemma 2. Let $T \in \mathcal{L}(V,V)$ be a normal operator on the finite dimensional complex inner product space and $v \in V$, a common eigenvector for T and T^*. Set $W := \langle v \rangle$. Then $W^\perp$ is both T and T^* invariant. Consequently, $T_{|_{W^\perp}}$ is a normal operator on $W^\perp$.

We followed these lemmas by the statement of:

Complex Spectral Theorem. Let $T \in \mathcal{L}(V,V)$ be a normal operator on the finite dimensional complex inner product space. Then T is a normal operator if and only if T is orthogonally diagonalizable.

Using the lemmas above, we showed the only if direction of the theorem. The point being: If T is normal, we take W as in Lemma 2. By induction on $\dim(V)$, there is an orthonormal basis of $W^\perp$ consisting of eigenvalues of $T_{|_{W^\perp}}$. These vectors together with $\frac{1}{||v||} \cdot v$ form an orthonormal basis for V consisting of eigenvectors of T, so that T is orthogonally diagonalizable.

Monday, October 3. We began class by recalling that if $A \in \mathrm{M}_n(\mathbb{R})$ is a symmetric matrix, then there exists an orthogonal matrix $Q \in \mathrm{M}_n(\mathbb{R})$ such that $Q^TAQ = D$, a diagonall matrix, while if $A \in \mathrm{M}_n(\mathbb{C})$, there exists a unitary matrix $Q \in \mathrm{M}_n(\mathbb{C})$ such that $Q^*AQ = D$, a diagonal matrix. We then worked an example, where beginning with the symmetric matrix $A = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 2 & 0 \\ 1 & 0 & 1 \end{pmatrix}$, we found an orthogonal 3×3 matrix Q with entries in $\mathbb{R}$ such that $Q^tAQ = D(2,2,0)$. We then showed that for operators defined over $\mathbb{R}$ or matrices in $\mathrm{M}_n(\mathbb{R})$, the converse to the main theorem discussed in the previous two lectures holds, yielding the following:

Real Spectral Theorem. Let V be a finite dimensional inner product space over $\mathbb{R}$ and $A \in \mathrm{M}_n(\mathbb{R})$. Then:

(i) T is self-adjoint if and only if T is orthogonally diagonalizable.

(ii) A is symmetric if and only if there exists an orthogonal matrix $Q \in \mathrm{M}_n(\mathbb{R})$ such that $Q^tAQ = D$, a diagonal matrix.

We then mentioned that the the theorem above does not hold as stated when the underlying field is $\mathbb{C}$, and that a weaker notion is required:

Definition. When $F = \mathbb{C}$ and $T \in \mathcal{L}(V, V)$ or $A \in \mathrm{M}_n(\mathbb{C})$, we say T is a *normal operator* if $T^*T = TT^*$ and A is a *normal matrix* if $A^*A = AA^*$.

We then noted that the Complex Spectral Theorem is obtained by replacing the self-adjoint condition in the theorem above by the normal condition.

We finished class by considering the matrix $A = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$. This matrix can be considered as an element of $\mathrm{M}_n(\mathbb{R})$ or $\mathrm{M}_n(\mathbb{C})$. Its characteristic polynomial is $x^2 + 1$. Thus, A has no eigenvalue over $\mathbb{R}$ and distinct eigenvalues over $\mathbb{C}$. Therefore, over $\mathbb{R}$, A is not even diagonalizable, let alone orthogonally diagonalizable - and of course A is not symmetric. We noted that A is a normal matrix, and we found a 2×2 unitary matrix Q over $\mathbb{C}$ such that $Q^*AQ = D$, a diagonal matrix. In other words, the normal matrix A is orthogonally diagonalizable over $\mathbb{C}$.

Friday, September 30. We began by giving a detailed review of the proof of the spectral theorem for self-adjoint matrices from the previous lecture. We then gave a proof of the matrix analogue of this theorem, namely, if $A \in \mathrm{M}_n(F)$ and $A = A^*$, then there exists a matrix $Q \in \mathrm{M}_n(F)$, whose columns form an orthonormal basis for F^n, and a diagonal matrix $D \in \mathrm{M}_n(F)$ such that $D = Q^{-1}AQ$. We then noted that $Q^{-1} = Q^*$. This led to the following definitions for a matrix Q whose columns form an orthonormal basis: If $Q \in \mathrm{M}_n(\mathbb{C})$, Q is said to be *unitary*, while if $Q \in \mathrm{M}_n(\mathbb{R})$, Q is said to be *orthogonal*.

We ended class by noting that a special case of the spectral theorem under discussion is the following theorem:

Theorem. Let $A \in \mathrm{M}_n(\mathbb{R})$ be a symmetric matrix. Then the eigenvalues of A belong to $\mathbb{R}$ and there exists an orthogonal matrix $Q \in \mathrm{M}_n(\mathbb{R})$ such that Q^tAQ is a diagonal matrix.

Wednesday, September 28. We continued our discussion of the adjoint of a linear operator T acting on a finite dimensional inner product space V by showing first that the adjoint of T is unique, in that T^* is the only element of $\mathcal{L}(V, V)$ satisfying $\langle T(v), w\rangle = \langle v, T^*(w)\rangle$, for all $v, w \in V$. We then defined the adjoint A^* of $A \in \mathrm{M}_n(F)$ as the conjugate transpose of A and showed that if B is an orthonormal basis for V, then if $A = [T]^B_B$, then $A^* = [T^*]^B_B$ and cautioned that this relation does not hold if B is not an orthonormal basis.

We then verified the following properties for $T, T_1, T_2 \in \mathcal{L}(V, V)$ and $\lambda \in F$:

(i) $(T_1 + T_2)^* = T_1^* + T_2^*$.
(ii) $(\lambda T)^* = \overline{\lambda} T^*$.
(iii) $(T^*)^* = T$.
(iv) $(T_1T_2)^* = T_2^*T_1^*$.
(v) $\langle T^*(v), w\rangle = \langle v, T(w)\rangle$, for all $v, w \in V$.

This was followed defining $T \in \mathcal{L}(V, V)$ to be *self adjoint* if $T = T^*$ and $A \in \mathrm{M}_n(F)$ to be self adjoint if $A^* = A$. We noted that a self adjoint real matrix is called *symmetric* while a self adjoint complex matrix is said to be *Hermetian*. We finished class by stating and proving the first of our *spectral theorems*:

First Spectral Theorem. Let V be a finite dimensional inner product space and $T \in \mathcal{L}(V, V)$ be self-adjoint. Then:

(i) Every eigenvalue of T is a real number.
(ii) There exists an orthonormal basis basis $B \subseteq V$ consisting of eigenvectors for T. In this case we say that T is *orthogonally diagonalizable*.

The ideas behind the proof are the following: Induct on the dimension of V, the case the $\dim(V) = 1$ being clear. For the general case, the *Fundamental Theorem of Algebra* guarantees that $\chi_T(x)$ splits over $\mathbb{C}$. An easy calculation shows (i) in the theorem above. For (ii), one takes λ an eigenvalue of T and E

its eigenspace. Self-adjointness shows that $E^\perp$ is T-invariant and that $T|_{E^\perp}$ is self adjoint. By induction, $E^\perp$ has an orthonormal basis B_2 consisting of eigenvectors for $T|_{E^\perp}$ (and hence eigenvectors for T). An orthonormal basis for E consists of eigenvectors for T, and thus $B_1 \cup B_2$ is the required orthonormal basis for V consisting of eigenvectors for T.

Monday, September 26. We began class with the following results, which enabled us to establish the existence of an *adjoint* T^* for any $T \in \mathcal{L}(V, V)$.

Propositions. For the finite dimensional vector space V, and $0 \neq v$, let $\hat{v} \in V^{**}$ be the linear functional on V^* given by $\hat{v}(f) = f(v)$, for all $f \in V^*$. Then

(i) The map $\phi : V \to V^{**}$ given by $\phi(v) = \hat{v}$ is an isomorphism of vector spaces.
(ii) If V is an inner product space, then given any $f \in V^*$, there exists a unique $v_0 \in V$ such that $f(v) = \langle v, v_0 \rangle$, for all $v \in V$.

We then gave the following:

Theorem-Definition. Let V be a finite dimensional inner product space and $T \in \mathcal{L}(V, V)$. Then there exists a unique element $T^* \in \mathcal{L}(V, V)$ satisfying $\langle T(v), w \rangle = \langle v, T^*(w) \rangle$, for all $v, w \in V$. We call T^* the *adjoint* of T.

The proof of the existence of T^* proceeded as follows: Fix $w \in V$. Then $\phi_w : V \to F$ given by $\phi_w(v) = \langle T(v), w \rangle$, for all $v \in V$ defines an element of V^*. Thus, by the proposition above, there exists a unique $w_0 \in V$ such that $\phi_w = \langle -, w_0 \rangle$. If we define $T^*(w) := w_0$, then $\langle T(v), w \rangle = \langle v, T^*(w) \rangle$, for all $v \in V$. Doing this for each $w \in V$ gives the required T^*. The proof of uniqueness follows easily from basic properties of the inner product.

Friday, September 23. We began class with an example motivating the concept of the *adjoint* of a linear operator $T \in \mathcal{L}(V, V)$. Starting with an orthonormal basis u_1, u_2 for $\mathbb{C}^2$ as an inner product space, with the usual Euclidean inner product, we then took A to be the matrix whose columns are u_1 and u_2. We then set A^* to be the matrix obtained from A by taking the conjugate of the entries of A^t, the transpose of A, thereby obtaining the *conjugate transpose of* A. We then showed that for all $v, w \in \mathbb{C}^2$, $\langle Av, w \rangle = \langle v, A^* w \rangle$. We then noted that we will soon show that if V is a finite dimensional vector space and $T \in \mathcal{L}(V, V)$, there exists $T^* \in \mathcal{L}(V, V)$ such that $\langle T(v), w \rangle = \langle v, T^*(w) \rangle$,for all $v, w \in V$.

For a vector space V, we then defined V^*, the dual space of V, as the set of all *linear functionals* on V, i.e., the set of linear transformations from V to F. If $B = \{v_1, \dots, v_n\} \subseteq V$ is a basis, we defined the *dual basis* $B^* := \{v_1^*, \dots, v_n^*\} \subseteq V^*$ where the v_i^* are defined by the formulas $v_i^*(v_j) = \delta_{ij}$, for all $1 \leq i, j \leq n$, where δ_{ij} is the *Kroenecker delta*, whose value is 1, when $i = j$ and 0 when $i \neq j$. We then proved:

Theorem. The dual basis B^* is a basis for V^*.

It follows from this theorem that $\dim(V^*) = n = \dim\,(V)$. We ended class by noting that if $v \in V$, then there is a canonical element $\hat{v} \in V^{**}$, the *double dual of* V, i.e., the dual space of V^*, defined by $\hat{v}(f) := f(v)$, for all $f \in V^*$.

Wednesday, September 21. We continued our discussion of diagonalizability, for V an n-dimensional vector space over the field F and $T \in \mathcal{L}(V, V)$. Using the proposition from the end of the previous class as a starting point, we first noted that if $\chi_T(x)$ splits as a product of linear polynomials, this is not enough to insure that T is diagonalizable. For example, if $T : \mathbb{R}^2 \to \mathbb{R}^2$, and $[T]_B^B = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ for some basis B, then $\chi_T(x) = (x-1)^2$, but T is not diagonalizable. This was followed by the observation that if T has n *distinct eigenvalues*, then T is diagonalizable. We then proved the following proposition.

Proposition. Let $T \in \mathcal{L}(V, V)$ and assume $\chi_T(x) = (x - \lambda)^e p(x)$, for $\lambda \in F$, $p(x) \in F[x]$ and $p(\lambda) \neq 0$. Then $\dim(E_\lambda) \leq e$.

We were then able to prove the main theorem concerning diagonalizability.

Theorem. Let $T \in \mathcal{L}(V, V)$, where V has dimension n. The following are equivalent:

(i) T is diagonalizable.

(ii) There exist distinct eigenvalues $\lambda_1, \dots, \lambda_r$ such that $V = E_{\lambda_1} \oplus \cdots \oplus E_{\lambda_r}$.
(iii) There exist distinct $\lambda_1, \dots, \lambda_r \in F$ such that $\chi_T(x) = (x-\lambda_1)^{e_1} \cdots (x-\lambda_r)^{e_r}$ where $\dim(E_{\lambda_i}) = e_i$, for all $1 \le i \le r$.
(iv) There exist distinct eigenvalues $\lambda_1, \dots, \lambda_r$ such that $\dim(V) = \dim(E_{\lambda_1}) + \cdots + \dim(E_{\lambda_r})$.

Monday, September 19. We began by noting the following important property: Let V be a finite dimensional vector space, $\lambda \in F$ and $T \in \mathcal{L}(V, V)$. Then λ is a root of $\chi_T(x)$ if and only if λ is a root of $\mu_T(x)$.

We then proceeded with the following definitions.

Definitions. Let V have dimension n, $T \in \mathcal{L}(V, V)$ and $A \in \mathrm{M}_n(F)$. Then:
(i) T is *diagonalizable* if and only if there exists a basis B for V such that $[T]_B^B$ is a diagonal matrix.
(ii) A is *diagonalizable* if and only if there exists an invertible $P \in \mathrm{M}_n(F)$ such that $P^{-1}AP$ is a diagonal matrix.

We then noted the equivalence of the following statements for $T \in \mathcal{L}(V, V)$:
(i) T is diagonalizable if and only there is a basis for V consisting of eigenvectors for T.
(ii) T is diagonalizable if and only if some matrix representing T is diagonalizable if and only if every matrix representing T is diagonalizable.

We finished class by proving the following proposition, where, in the statement below, we use the following notation: $D(\lambda_1, \dots, \lambda_n)$ is the $n \times n$ diagonal matrix with $\lambda_1, \dots, \lambda_n$ down its diagonal.

Proposition. Let V have dimension n and suppose $T \in \mathcal{L}(V, V)$ is diagonalizable.
(i) Suppose $B_1, B_2 \subseteq V$ are bases such that $[T]_{B_1}^{B_1} = D(\lambda_1, \dots, \lambda_n)$ and $[T]_{B_1}^{B_1} = D(\gamma_1, \dots, \gamma_n)$. Then, after re-indexing, $\lambda_1 = \gamma_1, \dots, \lambda_n = \gamma_n$.
(ii) There exist distinct $\lambda_1, \dots, \lambda_r \in F$ such that $\chi_T(x) = (x-\lambda_1)^{e_1} \cdots (x-\lambda_r)^{e_r}$.
(iii) There exist distinct $\lambda_1, \dots, \lambda_r \in F$ such that $V = E_{\lambda_1} \oplus \cdots \oplus E_{\lambda_r}$.

Friday, September 16. We began class with an example illustrating one of our main goals for the semester, namely the spectral theorem which states that *a real symmetric matrix is orthogonally diagonalizable.* For the example we took $A = \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix}$ and noted that A was diagonalizable since it eigenvalues 0, 5 are distinct. We then noticed that any eigenvector for 0 is orthogonal to any eigenvector for 5. We then noted that taking an eigenvector u_1 of length one for 0 and an eigenvector u_2 of length one for 5, the matrix P whose columns are u_1, u_2 is an *orthogonal matrix* such that $P^{-1}AP = \begin{pmatrix} 0 & 0 \\ 0 & 5 \end{pmatrix}$.

We followed the discussion by defining *eigenvalues* and *eigenvectors* for $T \in V$ and $A \in \mathrm{M}_n(F)$: If $\lambda \in F$, then λ is an eigenvalue of T if $T(v) = \lambda v$, some $v \neq 0$, which we call an eigenvector of λ with (with respect to T). Similarly, we defined λ to be an eigenvalue of A if there exists $0 \neq v_0 \in F^n$ such that $A \cdot v_0 = \lambda v_0$, in which case v_0 is an eigenvector for λ (with respect to A).

We then proved the following two propositions.

Proposition. Let $T \in \mathcal{L}(V, V)$ and $\lambda \in F$. The following are equivalent:
(i) λ is an eigenvalue of T.
(ii) λ is an eigenvalue for any matrix A representing T.
(iii) $\chi_A(\lambda) = 0$ for any matrix representing T.
(iv) $\chi_T(\lambda) = 0$.

Proposition. Let $T \in \mathcal{L}(V, V)$ and suppose $\lambda_1, \dots, \lambda_r$ are distinct eigenvalues of T. For each $1 \le i \le r$, set $E_{\lambda_i} := \{v \in V \mid T(v) = \lambda_i v\}$. Then:
(i) Each E_{λ_i} is a subspace of V called the *eigenspace of* λ_i.
(ii) Upon setting $W := E_{\lambda_1} + \cdots + E_{\lambda_r}$, we have $W = E_{\lambda_1} \oplus \cdots \oplus E_{\lambda_r}$.
(iii) If $v_1, \dots, v_r$ are non-zero vectors such that $v_i \in E_{\lambda_i}$, for $1 \le i \le r$, then $v_1, \dots, v_r$ are linearly independent. In other words, eigenvectors corresponding to distinct eigenvalues are linearly independent.

We were running out of time at the end of class during the discussion of the proof of (iii), so here is how (iii) follows from (ii): Suppose $\alpha_1 v_1 + \cdots + \alpha_r v_r = 0$. Then each $\alpha_i v_i \in E_{\lambda_i}$. If we set $w_i = \lambda_i v_i$ we have $w_1 + \cdots + w_r = 0$, Since the sum in (ii) is direct, and each $w_i \in E_{\lambda_i}$, we must have each $w_i = 0$. Thus, $\alpha_i v_i = 0$, for all i and therefore $\alpha_i = 0$, for all i. Hence $v_1, \ldots, v_r$ are linearly independent.

Wednesday, September 14. We gave a detailed proof of the fact that if V is a finite dimensional inner product space and $W \subseteq V$ is a subspace, then every vector $v \in V$ can be written uniquely in the form $v = w + w'$, with $w \in W$ and $w' \in W^{\perp}$. This was done by using the fact that an orthonormal basis for W extends to an orthonormal basis for V. We then defined w to be the *orthogonal projection of v onto* W and $W^{\perp}$ to be the *orthogonal complement* of W. The decomposition of V in terms of W and $W^{\perp}$ led to the following definition.

Definition. Let $W_1, \ldots, W_t$ be subspaces of the vector space V.

(i) $W_1 + \cdots + W_t$, the *sum of* $W_1, \ldots, W_t$, is the set of all vectors of the form $w_1 + \cdots + w_t$, with each $w_i \in W_i$.

(ii) V is the *direct sum* of $W_1, \ldots, W_t$ if $V = W_1 + \cdots + W_t$ and every $v \in V$ can be written *uniquely* as $v = w_1 + \cdots + w_t$, for $w_i \in W_i$. The uniqueness statement means that if we also have $v = w_1' + \cdots + w_t'$, with each $w_i' \in W_i$, then $w_i = w_i'$, for all $1 \leq i \leq t$. In this case we write $V = W_1 \oplus \cdots \oplus W_t$.

This definition was followed by the following proposition.

Proposition. Let $W_1, \ldots, W_t$ be subspaces of the vector space V.

(i) $W_1 + \cdots + W_t$ is a subspace of V.

(ii) $V = W_1 \oplus \cdots \oplus W_t$ if and only if $V = W_1 + \cdots + W_t$ and $W_j \cap (\Sigma_{i \neq j} W_i) = 0$, for all $1 \leq j \leq t$.

We then noted that the discussion of $W^{\perp}$ at the start of class shows $V = W \oplus W^{\perp}$, when V is a finite dimensional inner product space. We also showed that if $W \subseteq \mathbb{R}^3$ is a plane through the origin and $L \subseteq \mathbb{R}^3$ is a line through the origin not contained in W, then $\mathbb{R}^3 = W \oplus L$.

Monday, September 12. Throughout today's lecture, V denoted an inner product space with $F = \mathbb{R}$ or $\mathbb{C}$. We began with the observation that if $v_1, \ldots, v_n$ are mutually orthogonal vectors, then they are linearly independent over F. We then noted that a partial converse is given by:

Gram-Schmidt Orthogonalization. Let $v_1, \ldots, v_n$ be linear independent vectors in the inner product space V. Then there exist $w_1, \ldots, w_n \in U := \langle v_1, \ldots, v_n \rangle$ such that $w_1, \ldots, w_n$ are mutually orthogonal vectors and $\langle w_1, \ldots, w_n \rangle = U$.

The proof proceeded by induction on n, using the observation that if $w_1, \ldots, w_{i-1}$ have been constructed so that the conclusion of the theorem applies to $w_1, \ldots, w_{i-1} \in \langle v_1, \ldots, v_{i-1} \rangle$, then for

$$w_i := v_i - \frac{\langle v_i, w_1 \rangle}{\langle w_1, w_1 \rangle} w_1 - \cdots - \frac{\langle v_i, w_{i-1} \rangle}{\langle w_{i-1}, w_{i-1} \rangle} w_{i-1},$$

$w_1, \ldots, w_i \in \langle v_1, \ldots, v_i \rangle$ satisfy the conclusion of the theorem.

We then defined an *orthonormal* system of vectors to be an orthogonal set of vectors having length one. It followed from the theorem above that if V is an inner product space, and $W \subseteq V$ is a finite dimensional subspace, then W has an *orthonormal basis*. We noted that if $u_1, \ldots, u_n$ is an orthonormal basis for V, then any $v \in V$ can be written as

$$v = \langle v, u_1 \rangle \cdot u_1 + \cdots + \langle v, u_n \rangle \cdot u_n.$$

This was followed by the observation that if $u_1, \ldots, u_r$ is an orthonormal basis for W, then this basis can be extended to an orthonormal basis $u_1, \ldots, u_r, u_{r+1}, \ldots, u_n$ for V. We then noted that $W^{\perp} = \langle u_{r+1}, \ldots, u_n \rangle$ and every v in V can be written *uniquely* as $v = w + w'$, where $w \in W$ and $w' \in W^{\perp}$.

Friday, September 9. We began class by defining the concept of *inner product space*: An inner product space is a vector space V over $F = \mathbb{R}$ or $\mathbb{C}$ together with a function $\phi : V \times V \to F$ satisfying:

(i) $\phi(v, v) \in \mathbb{R}$, for all $v \in V$.

(ii) $\phi(v, v) \geq 0$ for all $v \in V$ and $\phi(v, v) = 0$ if and only if $v = 0$.

(iii) $\phi(w, v) = \overline{\phi(v, w)}$, for all $v, w \in V$.

(iv) $\phi(\lambda v, w) = \lambda\phi(v, w)$ and $\phi(v, \lambda w) = \overline{\lambda}\phi(v, w)$, for all $v, w \in V$ and $\lambda \in F$.
(v) $\phi(\lambda_1 v_1 + \lambda_2 v_2, w) = \lambda_1\phi(v_1, w) + \lambda_2\phi(v_2, w)$, for all $v_i, w_i \in V$ and $\lambda_i \in F$.

Here the overline denotes complex conjugate. Moreover, it follows from properties (iii)-(v) that

$$\begin{aligned}\phi(v, \lambda_1 w_1 + \lambda_2 w_2) &= \overline{\phi(\lambda_1 w_1 + \lambda_2 w_2, v)} \\ &= \overline{\lambda_1\phi(w_1, v) + \lambda_2\phi(w_2, v)} \\ &= \overline{\lambda_1} \cdot \overline{\phi(w_1, v)} + \overline{\lambda_2} \cdot \overline{\phi(w_2, v)} \\ &= \overline{\lambda_1}\phi(v, w_2) + \lambda_2\phi(v, w_2).\end{aligned}$$

for all $v_i, w_i \in V$ and $\lambda_i \in F$.

Henceforth we agreed to write $\langle v, w\rangle$, instead of $\phi(v, w)$. We then gave the following examples of inner product spaces.

Examples. 1. $V = \mathbb{R}^n$, and $\langle v, w\rangle := \alpha_1\beta_1 + \cdots + \alpha_n\beta_n$, for $v = \begin{pmatrix}\alpha_1\\ \vdots \\ \alpha_n\end{pmatrix}$ and $w = \begin{pmatrix}\beta_1\\ \vdots \\ \beta_n\end{pmatrix}$ defines an inner product.

2. 1. $V = \mathbb{C}^n$, and $\langle v, w\rangle := \alpha_1\overline{\beta_1} + \cdots + \alpha_n\overline{\beta_n}$, for $v = \begin{pmatrix}\alpha_1\\ \vdots \\ \alpha_n\end{pmatrix}$ and $w = \begin{pmatrix}\beta_1\\ \vdots \\ \beta_n\end{pmatrix}$ defines an inner product.
3. Letting P_n denote the vector space of polynomials of degree less than or equal to n with coefficients in F, $\langle f, g\rangle := \int_0^1 f(x)\overline{g(x)}\,dx$ defines an inner product, for all $f(x), g(x) \in P_n$.
4. Let $V := \mathrm{M}_n(F)$. Then $\langle A, B\rangle := \mathrm{tr}(A^t \cdot \overline{B})$ defines an inner product, for all $A, B \in \mathrm{M}_n(F)$.

We followed the examples with the observations and definitions below, concerning an inner product space V:

(a) For fixed $v, v' \in V$, $v = v'$ if and only if $\langle v, w\rangle = \langle v', w\rangle$ for all $w \in V$.
(b) For $v \in V$, $||v||$, the *length* of $v \in V$ or the *norm* of v, is the real number $||v|| = \sqrt{\langle v, v\rangle}$.
(c) For $v \in V$, $||v|| \geq 0$ and $||v|| = 0$ if and only if $v = 0$.
(d) For $v \in V$ and $\lambda = a + bi \in F$, $||\lambda v|| = |\lambda| \cdot ||v||$, where $|\lambda| = \sqrt{a^2 + b^2}$. In particular, if $\lambda = \frac{1}{||v||}$, then $||\lambda v|| = 1$.
(e) Vectors $v, w \in V$ are *orthogonal* if $\langle v, w\rangle = 0$.
(f) For a subspace $W \subseteq V$, $W^\perp := \{u \in V \mid \langle w, u\rangle = 0, \text{ for all } w \in W\}$ is a subspace of V satisfying $W \cap W^\perp = (0)$.

Wednesday, September 7. We began class by discussing the three polynomials that will play important roles in the main theorems we study this semester.

Definition. Let V be an n-dimensional vector space over the field F, v a non-zero vector in V or a column vector in F^n, $A \in \mathrm{M}_n(F)$ and $T \in \mathcal{L}(V, V)$.

(i) The *characteristic polynomial* of A or T: $\chi_A(x) := |xI_n - A|$ and $\chi_T(x) := \chi_A(x)$, for any matrix A representing T. By the Cayley-Hamilton theorem, $\chi_A(A) = 0$ and $\chi_T(T) = 0$.
(ii) The *minimal polynomial* of A or T: $\mu_A(x)$ is the monic polynomial of least degree in $F[x]$ such that $\mu_A(A) = 0$, Similarly, $\mu_T(x)$ is the monic polynomial of least degree such that $\mu_T(T) = 0$.
(iii) The *minimal polynomial of A or T with respect to v:* $\mu_{A,v}(x)$ is the monic polynomial of least degree in $F[x]$ such that $\mu_{A,v}(x) \cdot v = 0$ and $\mu_{T,v}(x)$ is the monic polynomial of least degree such that $\mu_{T,v}(x)(v) = 0$.

This was followed by:

Proposition. In the notation above, suppose $p(x) \in F[x]$.

(i) If $p(A) = 0$, then $\mu_A(x)$ divides $p(x)$ in $F[x]$. Similarly, if $p(T) = 0$, then $\mu_T(x)$ divides $p(x)$.
(ii) If $p(A) \cdot v = 0$, then $\mu_{A,v}(x)$ divides $p(x)$ in $F[x]$. Similarly, if $p(T)(v) = 0$, then $\mu_{T,v}(x)$ divides $p(x)$

Before proving the proposition, we reviewed the *division algorithm* in $F[x]$: Let $f(x), g(x) \in F[x]$, then there exist *unique* $h(x), r(x) \in F[x]$ such that $g(x) = f(x)h(x) + r(x)$, where $r(x) = 0$ or the degree of $r(x)$ is less than the degree of $f(x)$. We then used the proposition to show that $\mu_A(x)$ is the *unique* monic polynomial of least degree such that $\mu_A(A) = 0$. Similarly, $\mu_{A,v}(x)$ is the *unique* monic polynomial of least degree such that $\mu_{A,v}(A) \cdot v = 0$. Similar uniqueness properties hold for $\mu_T(x)$ and $\mu_{T,v}(x)$.

As a means of leading into inner product spaces, we ended class by recalling the basic properties of the dot product of vectors from $\mathbb{R}^2$. We also observed that the usual definition of the dot product needs to be modified when considering vectors in $\mathbb{C}^2$, namely, if $v = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}$ and $w = \begin{pmatrix} \gamma \\ \delta \end{pmatrix}$ are vectors in $\mathbb{C}^2$, then defining $v \cdot w := \alpha\overline{\gamma} + \beta\overline{\delta}$ insures that $v \cdot v = 0$ if and only if $v = 0$. Here $\overline{\alpha}$ denotes the complex conjugate of α.

Friday, September 2. Today's lecture was devoted to a proof of the important:

Cayley-Hamilton Theorem. Let $A \in \mathrm{M}_n(F)$ and set $\chi_A(x) := |xI_n - A|$, the *characteristic polynomial* of A. Then $\chi_A(A) = 0$. Moreover, if $T \in \mathcal{L}(V, V)$, then $\chi_T(T) = 0$, where $\chi_T(x) = \chi_A(x)$, for any $A \in \mathrm{M}_n(F)$ representing T.

The proof of the theorem relied on several ancillary notions and results. First, we defined the *companion matrix* $C(f(x))$ associated to $f(x) \in F[x]$ as follows: Given $f(x) = x^s + \alpha_{s-1}x^{s-1} + \cdots + \alpha_1 x + \alpha_0$, then

$$C(f(x)) = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 & -\alpha_0 \\ 1 & 0 & 0 & \cdots & 0 & -\alpha_1 \\ 0 & 1 & 0 & \cdots & 0 & -\alpha_2 \\ \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 1 & -\alpha_{s-1} \end{pmatrix}.$$

Thus, for example, if $f(x) = x^3 + \alpha_2 x^2 + \alpha_1 x + \alpha_0$, then $C(f(x)) = \begin{pmatrix} 0 & 0 & -\alpha_0 \\ 1 & 0 & -\alpha_1 \\ 0 & 1 & -\alpha_2 \end{pmatrix}$. We then noted that $\chi_{C(f(x))}(x) = f(x)$, i.e., the characteristic polynomial of $C(f(x))$ is $f(x)$, and left its proof as an exercise.

Then given $0 \neq v \in F^n$, we defined $\mu_{A,v}(x)$ to be the monic polynomial $p(x) \in F[x]$ of least degree such that $p(A)v = 0$. It followed from this that if we write $\mu_{A,v}(x) = x^s + \alpha_{s-1}x^{s-1} + \cdots + \alpha_0$, then:

(i) $v, Av, \ldots, A^{s-1}v$ are linearly independent in F^n

(ii) $A^s v = -\alpha_0 v - \alpha_1 Av \cdots - \alpha_{s-1}A^{s-1}v$.

Our final preliminary result was that if $A = \begin{pmatrix} B & C \\ 0 & D \end{pmatrix}$ is a block matrix, where A is $n \times n$, B is $s \times s$, C is $s \times r$ and D is $r \times r$, where $n = s + r$, then $|A| = |B| \cdot |D|$.

We then proceeded with the proof of the Cayley-Hamilton Theorem. Here is the path we followed. Take $0 \neq v \in F^n$, where F^n is the vector space of column vectors of length n with entries in F. Suppose

$$\mu_{A,v}(x) = x^s + \alpha_{s-1}x^{s-1} + \cdots + \alpha_0,$$

so that $v, Av, \ldots, A^{s-1}v$ are linearly independent over F. Extend these elements to a basis $\mathcal{B}$ for F^n. Define $T : F^n \to F^n$ by $T(w) := Aw$, for all $w \in F^n$. Note that the matrix of T with respect to the standard basis for F^n is just A. Now, $[T]_{\mathcal{B}}^{\mathcal{B}} := E = \begin{pmatrix} B & C \\ 0 & D \end{pmatrix}$ is a block matrix as above, where $B = C(\mu_{A,v}(x))$. Note that $\chi_A(x) = \chi_T(x) = \chi_E(x)$, since A and E are two matrices representing T. Thus,

$$\chi_A(x) = \chi_{C(\mu_{A,v}(x))}(x) \cdot \chi_D(x).$$

Therefore,

$$\chi_A(A)v = \chi_D(A)\mu_{A,v}(A)v = \chi_D(A)0 = 0.$$

Since this is true for all $v \in F^n$, it follows that $\chi_A(A) = 0$, which gives the Cayley-Hamilton Theorem.

Wednesday, August 31. We finished our discussion concerning properties of $|A|$ by showing:

(i) If A is upper or lower triangular, then $|A|$ is the product of the diagonal entries of A.

(ii) $|A^t| = |A|$.
(iii) For a fixed value of $1 \leq j \leq n$, $|A| = \Sigma_{i=1}^n (-1)^{i+j} a_{ij} |A_{ij}|$, expansion along the jth column.
(iv) For a fixed value of $1 \leq i \leq n$, $|A| = \Sigma_{j=1}^n (-1)^{i+j} a_{ij} |A_{ij}|$, expansion along the ith row.

We then turned our attention to the main topic of the course: linear operators on a finite dimension vector space and their corresponding matrices. We will write $\mathrm{M}_n(F)$ to denote the $n \times n$ matrices over the field F and $\mathcal{L}(V, V)$ for the set of linear operators on the vector space V. We noted that both of these sets are vector spaces of dimension n^2 over F, assuming the dimension of V is n. Finally, we noted that if $A \in \mathrm{M}_n(F)$, $T \in \mathcal{L}(V, V)$, and $p(x) \in F[x]$, the ring of polynomials with coefficients in F, then $p(A) \in \mathrm{M}_n(F)$ and $p(T) \in \mathcal{L}(V, V)$.

Monday, August 29. We continued our discussion and proofs of the properties of $|A|$, the determinant of the $n \times n$ matrix A. In particular, we established the following properties:

(i) If two rows of A are equal, then $|A| = 0$.
(ii) If B is obtained from A by adding a multiple of one row of A to another row of A, then $|B| = |A|$.
(iii) If $E_1, \ldots, E_r$ are elementary matrices, then $|E_r \cdots E_1 A| = |E_r| \cdots |E_1| \cdot |A|$.
(iv) If B is an $n \times n$ matrix, then $|BA| = |B| \cdot |A|$.

We also recalled the important fact from a first course in linear algebra that the matrix A is invertible, if and only if by a sequence of elementary row operations, the reduced row echelon form of A is I_n. This process of using row operations also shows that the inverse of A is the product of the elementary matrices corresponding to the row operations used in this process. This lead to a proof of the following:

Proposition. Let A be an $n \times n$ matrix. The following are equivalent:

(i) A is invertible, i.e., A^{-1} exists.
(ii) $|A| \neq 0$.
(iii) A is a product of elementary matrices.

Friday, August 26. We began with an $n \times n$ matrix $A = (a_{ij})$ with entries in the field F. We defined the determinant of A inductively in terms of the Laplace expansion along the first column of A:

$$\det(A) = |A| = \sum_{i=1}^{n} (-1)^{i+1} a_{i1} \cdot |A_{i1}|,$$

where A_{i1} denotes the matrix obtained from A by deleting its ith row and first column. With this definition we proved the following properties of the determinant by induction on n, assuming they hold for $n = 1, 2$:

(i) $|I_n| = 1$, where I_n is the $n \times n$ identity matrix.
(ii) If B is obtained from A by multiplying one of its rows by $\lambda \in F$, then $|B| = \lambda \cdot |A|$.
(ii) If a row of A consists entirely of zeros, then $|A| = 0$.
(iv) If for some $1 \leq k \leq n$, and all $1 \leq i \leq n$, with $i \neq k$, the ith rows of A, B, C are the same, while the kth row of C is the sum of the kth rows of A and B, then $|C| = |A| + |B|$.
(v) If B is obtained from A by interchanging two rows, then $|B| = -|A|$.

After proving (ii) and (iv) above, we noted that together these properties imply that the determinant is a *multplinear functions* of its rows. We ended class by recalling the three familiar elementary row operations: (i) multiplying a row by a non-zero element of F; (ii) interchanging two rows; (iii) adding a multiple of one row to another row and defined an *elementary matrix* to be a matrix obtained by applying an elementary row operation to I_n. We noted (as an exercise) if E is an elementary row operation, then for any $n \times n$ matrix A, EA is the matrix obtained from A by employing the corresponding elementary row operation. Since a sequence of elementary row operation renders A into *reduced row echelon form*, we observed that it follows that there are elementary matrices $E_1, \ldots, E_r$ such that $E_r \cdots E_1 A = A_0$, where A_0 is in reduced row echelon form, i.e., the leading entry of each row is 1; the entries above and below each leading 1 are 0s; if the ith row and jth row of A_0 are not zero, and $i < j$, then the leading entry for the jth row is to the right of leading entry of the ith row; all rows consisting entirely of 0s are at the bottom of the matrix A_0. In particular, A_0 is upper triangular.

Wednesday, August 24. We stated and proved a number of formulas involving representing linear transformations as matrices. The essential point was to establish some notation that (hopefully) makes remembering these standard formulas easy – or at least easy to recover. The notation we used was the following: Let V and W be finite dimensional vector spaces over the field F and $T : V \to W$ a linear transformation. If $B_V = \{v_1, \ldots, v_n\}$ is a basis for V and $B_W = \{w_1, \ldots, w_m\}$ is a basis for W, then we write $[T]_{B_V}^{B_W}$ for the matrix of T with respect to these bases. We then proved that if $S : W \to U$ is a linear transformation and $B_U = \{u_1, \ldots, u_p\}$ is a basis for U, then $[ST]_{B_V}^{B_U} = [S]_{B_W}^{B_U} \cdot [T]_{B_V}^{B_W}$. The notation here is suggestive and makes it easy to remember the relations between the various maps, matrices and bases. Some corollaries of this formula presented in class are:

(i) If B_1 and B_2 are bases for V then $I_n = [I_n]_{B_2}^{B_1} \cdot [I_n]_{B_1}^{B_2}$, which shows that $[I_n]_{B_2}^{B_1}$ is the inverse of the matrix $[I_n]_{B_1}^{B_2}$.

(ii) Suppose B_V^1 and B_V^2 are bases for V and B_W^1 and B_W^2 are bases for W. Then

$$[T]_{B_V^2}^{B_W^2} = [I]_{B_W^1}^{B_W^2} \cdot [T]_{B_V^1}^{B_W^1} \cdot [I]_{B_V^2}^{B_V^1}.$$

This is the standard *change of basis* formula. If we write $B = [T]_{B_V^2}^{B_W^2}$, $A = [T]_{B_V^1}^{B_W^1}$, $P = [I]_{B_V^2}^{B_V^1}$, and $Q = [I]_{B_W^2}^{B_W^1}$, we get the familiar looking version $B = Q^{-1}AP$.

(iii) As a special case of (ii), if we take $V = W$, then (using the same notation), we have $B = P^{-1}AP$.

Monday, August 22. We discussed an overview of topics to be covered in the course and reviewed the basic notions of linear independence, spanning sets and bases. Here we did not require the ambient vector space to be finite dimensional. We then defined the ingredients in Zorn's Lemma and stated Zorn's Lemma: *If a partially order set X has the property that every chain has an upper bound, then X has a maximal element.* We then proved the following theorem:

Theorem. Let V be a vector space over the field F and $S \subseteq V$ a linearly independent subset. Then:

(i) There exists a subset $T \subseteq V$ containing S such that T is maximal among linearly independent sets of vectors containing S.

(ii) For T as in (i), T spans V.

The idea of the proof is the following: Take X to be the set of linearly independent subsets of V containing S. Partially order X as follows: for $T_1, T_2 \in T$, write $T_1 \leq T_2$ if and only if $T_1 \subseteq T_2$. Then X is a partially ordered set. We then showed that if C is a chain in X, then $\tilde{C} := \bigcup\{C_\alpha \mid C_\alpha \in C\}$ is a linearly independent subset of X. Thus $\tilde{C} \in X$ is an upper bound for C. Thus, every chain has an upper bound, so X has a maximal element T - which by definition is a maximally linearly independent subset of X containing S. We then noted that if T is a linearly independent subset of V and $v \notin \langle T \rangle$, then $T \cup \{v\}$ is linearly independent. This implies that if T is maximally linearly independent, $\langle T \rangle = V$. An immediate corollary is that any vector space has a basis, and in fact, the proof shows that any linearly independent subset of V can be extended to a basis.

As an example, we noted that the ring of polynomials $\mathbb{R}[x]$ over $\mathbb{R}$ forms a vector space over $\mathbb{R}$ and $\{1, x, x^2, \ldots, \}$ is a basis for $\mathbb{R}[x]$.