FALL 2024: MATH 790 DAILY UPDATE

Monday, August 26. We began by reviewing some basic concepts about vector spaces over a field, including spanning sets, linear independence, and bases and how to extend these notions to vector spaces that might not be finite dimensional. In particular, we discussed the following fundamental property:

Exchange Property Let V be a vector space over the field F, $\{v_, \dots, v_t\} \subseteq V$ and set $V_0 := \text{Span}\{v_1, \dots, v_t\}$. Let $\{w_1, \dots, w_s\} \subseteq V_0$ be a linearly independent subset of V_0. Then $s \leq t$ and, after re-indexing, V_0 is spanned by $\{w_1, \dots, w_s, v_{t-s+1}, \dots, v_t\}$.

We also noted that the Exchange Property implies that if V is a finite dimensional vector space, any two bases for V have the same number of elements.

We ended class with a discussion of Zorn's Lemma and how it can be used to prove the following theorem.

Theorem. Let V be a vector space over the field F and suppose $S \subseteq V$ is a linearly independent subset of V. Then there exists a basis $B \subseteq V$ containing S.

The idea of the proof of the theorem was as follows. Apply Zorn's lemma to the collection X of linearly independent subset of V containing S to find a maximally linearly independent subset $B \subseteq V$ containing S. As in the finite dimensional case, one then shows that a maximally linearly independent subset must span V, and hence is a basis for V.

Wednesday, August 28. After recording the fact that any two bases for a vector space V have the same cardinality, 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 $\alpha = \{v_1, \dots, v_n\}$ is a basis for V and $\beta = \{w_1, \dots, w_m\}$ is a basis for W, then we write $[T]_\alpha^\beta$ for the $m \times n$ matrix of T with respect to α and β. If we write a_{ij} for the entries of this matrix, the a_{ij} are obtained from the following formulas: For each $1 \leq j \leq n$, $T(v_j) = a_{1j}w_1 + \cdots + a_{mj}w_m$.

We then proved the following crucial formula:

Crucial Formula. Maintaining the notation above, assume further that $S : W \to U$ is a linear transformation and $\gamma = \{u_1, \dots, u_t\}$ is a basis for U. Then $[ST]_\alpha^\gamma = [S]_\beta^\gamma \cdot [T]_\alpha^\beta$.

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 A is an $n \times m$ matrix, B a $t \times m$ matrix and C a $s \times t$ matrix, then $C(BA) = (CB)A$, i.e., matrix multiplication is associative. This follows from the crucial formual above and the fact that composition of linear transformation is associative.

(ii) If α and β are bases for V then $[I_n]_\beta^\alpha \cdot [I_n]_\alpha^\beta = I_n = [I_n]_\alpha^\beta \cdot [I_n]_\beta^\alpha$ which shows that $[I_n]_\beta^\alpha$ is the inverse of the matrix $[I_n]_\alpha^\beta$. The matrix $[I_n]_\beta^\alpha$ is called the *change of basis matrix*, and is obtained by expressing the basis vectors in β in terms of the basis α.

(iii) Suppose α, β are bases for V and $T : V \to V$ is a linear operator. Then $[T]_\beta^\beta = [I_n]_\alpha^\beta \cdot [T]_\alpha^\alpha \cdot [I_n]_\beta^\alpha$. If we set $A := [T]_\alpha^\alpha$, $B := [T]_\beta^\beta$, and $P := [I_n]_\beta^\alpha$, then we have the formula $B = P^{-1}AP$, which is the form this equation takes in most linear algebra texts.

Friday, August 30. 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 discussed the following properties of the determinant, proving some of them by induction on n, assuming they hold for $n = 1, 2$, and giving indications of proofs of others.

(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 $|B| = -|A|$.
(vi) If two rows of A are the same, then $|A| = 0$.
(vii) If B is obtained from A by adding a multiple of the sth row of A to its rth row, then $|B| = |A|$.

After proving (ii) and (iv) above, we noted that together these properties imply that the determinant is a *multilinear functions* of its rows. We continued 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 $n \times m$ matrix M, EM is the matrix obtained from M by employing the corresponding elementary row operation. Since a sequence of elementary row operation renders M into *reduced row echelon form*, we observed that it follows that there are elementary matrices $E_1, \ldots, E_r$ such that $E_r \cdots E_1 M = M_0$, where M_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. We ended class by noting that the inverse of an elementary matrix is again an elementary matrix (of the same type) and recording the further property:

(viii) For any $n \times n$ matrix A and elementary $n \times n$ matrix E, $|EA| = |E| \cdot |A|$.

Wednesday, September 5. We continued our discussion of properties of the determinant of $n \times n$ matrices. Throughout, $A = (a_{ij})$ is an $n \times n$ matrix with entries in the field F. After reviewing some basic facts about elementary matrices, we continued our discussion of properties of the determinant, proving some of these properties, and sketching proofs of others.

(ix) The follow properties are equivalent for A:
   (a) A is invertible
   (b) $A| \neq 0$
   (c) A is a product of elementary matrices.
(x) $|BA| = |B| \cdot |A|$, for any $n \times n$ matrix B.
(xi) If A is upper or lower triangular, then $|A| = a_{11}a_{22} \cdots a_{nn}$.
(xii) $|A^t| = |A|$.
(xiii) $|A| = \Sigma_{i=1}^{n}(-1)^{i+j}a_{ij}|A_{ij}|$, expansion along the jth column.
(xiv) $|A| = \Sigma_{j=1}^{n}(-1)^{i+j}a_{ij}|A_{ij}|$, expansion along the ith row.

We then turned our attention to the main topic of the course: linear operators on a finite dimensional 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 these vector spaces are isomorphic and have 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 $p(T) \in \mathcal{L}(V, V)$, and that there exists $p(x) \in F[x]$, such that $p(A) = 0$ and $p(T) = 0$, if A is a matrix representing T.

Friday, September 6. 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.

Monday, September 9. After reviewing the Cayley-Hamilton theorem, 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 $\mu_{T,v}(x)(v) = 0$.

We then noted that $\deg\mu_{A,v}(x) \leq \deg\mu_A(x) \leq n$. This was followed by noting that $\mu_{A,v}(x)$ divides $\mu_A(x)$, which divides $\chi_A(x)$ in the polynomial ring $F[x]$. For this, we needed to discuss the division algorithm in $F[x]$, namely: Let $0 \neq f(x), g(x) \in F[x]$, 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 division algorithm 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 $\mu_{T,v}(x)$. We also noted that the arguments given can be extended to show:

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 $p(x)$

As a corollary we gave a proof of the following:

Important Fact. $\lambda \in F$ is a root of $\mu_A(x)$ (resp., $\mu_T(x)$) if and only if λ is a root of $\chi_A(x)$ (resp., $\chi_T(x)$).

We ended class by defining *eigenvalues* and *eigenvectors* for both T and A, and noting that $\lambda \in F$ is an eigenvalue of A or T if and only if $\chi_A(\lambda) = 0$ or $\chi_T(\lambda) = 0$.

Wednesday, September 11. We began class by recalling the definitions of eigenvalues and eigenvectors for transformations and matrices, and discussed the floowing proposition, essentially proven last time, with the help of HW 6.

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) $\chi_T(\lambda) = 0$.

We then had a brief digression to discuss direct sums of subspaces. beginning with

Definition. Let $W_1, \dots, W_r$ be subspaces of the vector space V.

(i) $S = W_1 + \cdots + W_r$, the *sum of* $W_1, \dots, W_r$, is the set of all vectors of the form $w_1 + \cdots + w_r$, with each $w_i \in W_i$.
(ii) S is the *direct sum* of $W_1, \dots, W_r$ if $S = W_1 + \cdots + W_r$ and $W_j \cap (\Sigma_{i\neq j} W_i) = 0$, for all $1 \leq j \leq r$.

This definition was followed by noting that $\mathbb{R}^2$ is the direct sum of any two distinct lines through (0,0) and $\mathbb{R}^3$ is the direct sum of any plane through the origin and line through the origin not contained in the plane. This was followed by

Proposition. Let $W_1, \dots, W_r$ be subspaces of the vector space V, whose sum is S.

(i) $W_1 + \cdots + W_t$ is a subspace of V.
(ii) The following conditions are equivalent.
   (a) $S = W_1 \oplus \cdots \oplus W_r$
   (b) Every element $s \in S$ can be written *uniquely* as $s = w_1 + \cdots + w_r$, with each $w_i \in W_i$.
   (c) Whenever $w_1 + \cdots + w_r = 0$, with each $w_i \in W_i$, then $w_i = 0$, for all i.

We then presented the following:

Proposition. Let $T \in \mathcal{L}(V,V)$ and suppose $\lambda_1, \dots, \lambda_r$ are distinct eigenvalues of T. For each $1 \leq i \leq r$, 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 $S := E_{\lambda_1} + \cdots + E_{\lambda_r}$, we have $W = E_{\lambda_1} \oplus \cdots \oplus E_{\lambda_r}$. In particular, if $v_1, \dots, v_r$ are non-zero vectors such that $v_i \in E_{\lambda_i}$, for $1 \leq i \leq r$, then $v_1, \dots, v_r$ are linearly independent. In other words, eigenvectors corresponding to distinct eigenvalues are linearly independent.

We ended class by defining what it means for $T \in \mathcal{L}(V,V)$ or $A \in \mathrm{M}_n(F)$ to be diagonalizable:T is diagonalizable if there exists a basis $B \subseteq V$ such that $[T]_B^B = D$, a diagonal matrix and A is diagonalizable if there exists an invertible $n \times n$ matrix P such that $P^{-1}AP$ is a diagonal matrix.

Friday, September 13. We began class by recalling what it means for $T \in \mathcal{L}(V,V)$ ir $A \in \mathrm{M}_n(F)$ to be diagonalizable.

We then discussed the following observations for $T \in \mathcal{L}(V,V)$ diagonalizable:

(i) T is diagonalizable if and only there is a basis for V consisting of eigenvectors for T.

(ii) If $[T] = D(\lambda_1, \ldots, \lambda_n)$, then $\lambda_1, \ldots, \lambda_n$ are the only eigenvalues of T. Here, we used the following notation: $D(\lambda_1, \ldots, \lambda_n)$ is the $n \times n$ diagonal matrix with $\lambda_1, \ldots, \lambda_n$ down its diagonal.
(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 then presented:

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 $[T]_{B_1}^{B_1} = D(\lambda_1, \ldots, \lambda_n)$ and $[T]_{B_1}^{B_1} = D(\gamma_1, \ldots, \gamma_n)$. Then, after re-indexing, $\lambda_1 = \gamma_1, \ldots, \lambda_n = \gamma_n$.
(ii) There exist distinct $\lambda_1, \ldots, \lambda_r \in F$ such that $\chi_T(x) = (x - \lambda_1)^{e_1} \cdots (x - \lambda_r)^{e_r}$.
(iii) There exist distinct $\lambda_1, \ldots, \lambda_r \in F$ such that $V = E_{\lambda_1} \oplus \cdots \oplus E_{\lambda_r}$.

This was followed by the **important**

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 noted that e is called the *algebraic multiplicity* of λ and $\dim(E_\lambda)$ is called the *geometric multiplicity* of λ, and thus, the proposition above states that for any eigenvalue of T (or A), the geometric multiplicity is always less than to equal to the algebraic multiplicity.

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

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 $\lambda_1, \ldots, \lambda_r$ such that $V = E_{\lambda_1} \oplus \cdots \oplus E_{\lambda_r}$.
(iii) There exist distinct $\lambda_1, \ldots, \lambda_r \in F$ 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 $1 \leq i \leq r$.
(iv) There exist distinct eigenvalues $\lambda_1, \ldots, \lambda_r$ such that $\dim(V) = \dim(E_{\lambda_1}) + \cdots + \dim(E_{\lambda_r})$.

We ended class by noting 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.

Monday, September 16. We spent most of the class discussing and proving the Diagonalizability Theorem stated in the previous lecture. For this, we needed the following lemma.

Lemma. Suppose $V = W_1 \oplus \cdots \oplus W_r$, for subspaces $W_i \subseteq V$. If $B_i \subseteq W_i$ is a basis for W_i, for $1 \leq i \leq r$, then $B := \cup_{i=1}^r B_i$ is a basis for V. Conversely, suppose $B \subseteq V$ is a basis for V and $B = B_1 \cup \cdots \cup B_r$ is a partition of B. Then $V = W_1 \oplus \cdots \oplus W_r$, where each $W_i := \mathrm{Span}(B_i)$.

We then recorded the following corollary to the Diagonalizability Theorem, assuming the fact that, in general, V splits into a direct sum of the spaces $\ker(p_i(T)^{f_i})$, where $\mu_T(x) = p_1(x)^{f_i} \cdots p_r(x)^{f_i}$, with each $p_i(x)$ irreducible over F.

Corollary. Suppose $T \in \mathcal{L}(V, V)$. Then T is diagonalizable if and only if $\mu_T(x) = (x - \lambda_1) \cdots (x - \lambda_r)$, for distinct $\lambda_1, \ldots, \lambda_r \in F$.

We ended class by discussing the dot product over $\mathbb{R}^2$ and $\mathbb{C}^2$ as a means of motivating inner product spaces. We noted that trying to extend the usual definition of the dot product over $\mathbb{R}$ to a dot product over $\mathbb{C}$, without using complex conjugates, leads to the unworkable conclusion that a dot product of non-zero vectors can be zero. More details will follow in the next lecture.

Wednesday, September 18. 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

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

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, verifying some of the axioms in cases 2 and 4 below.

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 defined over F:

(a) For $v \in V$, $v = \vec{0}$ if and only if $\langle v, w\rangle = 0$, for all $w \in W$.
(b) For fixed $v, v' \in V$, $v = v'$ if and only if $\langle v, w\rangle = \langle v', w\rangle$ for all $w \in V$.
(c) For $T_1, T_2 \in \mathcal{L}(V, V)$, $T_1 = T_2$ if and only if $\langle T_1(v), w\rangle = \langle T_2(v), w\rangle$, for all $v, w \in V$.
(d) For $v \in V$, $||v||$, the *length* of $v \in V$, or the *norm* of v, is the real number $||v|| = \sqrt{\langle v, v\rangle}$.
(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 called the *orthogonal complement* of W.

Friday, September 20. 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

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

Monday, September 23. The class did group work on a selected set of problems from the homework.

Wednesday, September 25. We began a discussion of the **very important**:

Spectral Theorem for Real Symmetric Matrices. Let $A \in \mathrm{M}_n(F)$ be a symmetric matrix, i.e., $A = A^t$. Then there exists an orthogonal matrix $P \in \mathrm{M}_n(\mathbb{R})$ such that $P^{-1}AP$ is a diagonal matrix. In other words, symmetric matrices over $\mathbb{R}$ are *orthogonally diagonalizable.* Conversely, if $A \in \mathrm{M}_n(\mathbb{R})$ is orthogonally diagonalizable, then A is symmetric.

We first gave a proof the theorem for 2×2 matrices. A key fact was that if $v_1, v_2 \in \mathbb{R}^2$ are eigenvectors of A corresponding to distinct eigenvectors, then $\langle v_1, v_2 \rangle = 0$. We then discussed (and proved, where relevant) the following comments:

Comments. 1. The Fundamental Theorem of Algebra states that every $p(x)$ in $\mathbb{R}[x]$ or $\mathbb{C}[x]$ splits over $\mathbb{C}$, i.e., there exist distinct $\lambda_1, \ldots, \lambda_r \in \mathbb{C}$, and integers $e_i > 0$ such that $p(x) = (x - \lambda_1)^{e_1} \cdots (x - \lambda_r)^{e_r}$. Thus, for any $A \in \mathrm{M}_n(\mathbb{R})$, $\chi_A(x)$ has all of its roots in $\mathbb{C}$ (and possibly in $\mathbb{R}$).

2. We recalled that $P \in \mathrm{M}_n(\mathbb{R})$ is *orthogonal* if and only if its columns (resp., rows) form an orthonormal basis for $\mathbb{R}^n$. We noted that is is equivalent to saying that P is invertible and $P^{-1} = P^t$.

3. If $A \in \mathrm{M}_n(\mathbb{R})$ is symmetric, then $\langle Av, w \rangle = \langle v, Aw \rangle$, for all column vectors $v, w \in \mathbb{R}^n$

4. If $A \in \mathrm{M}_n(\mathbb{R})$ is symmetric, and $\lambda_1, \lambda_2 \in \mathbb{R}$ are distinct eigenvalues, then every vector in E_{λ_1} is orthogonal to every vector in E_{λ_2}.

5. If $P \in \mathrm{M}_n(\mathbb{R})$ is orthogonal, then P^{-1} is orthogonal.

Friday, September 27. We began class by continuing wth the comments from the previous lecture.

Comments continued. 6. Suppose $P, Q \in \mathrm{M}_n(\mathbb{R})$ are orthogonal matrices. Then PQ is an orthogonal matrix.

7. Let Q_0 be an $(n-1) \times (n-1)$ orthogonal matrix over $\mathbb{R}$. Then the $n \times n$ matrix

$$Q = \left(\begin{array}{c|c} 1 & \mathbf{0} \\ \hline \mathbf{0} & Q_0 \end{array}\right)$$

is orthogonal, where the top bold 0 is a row of $n-1$ zeros and the second bold 0 is a column of $n-1$ zeros.

We then presented the following two propositions:

Proposition A. If $A \in \mathrm{M}_n(\mathbb{R})$ is symmetric, then all of its eigenvalues are in $\mathbb{R}$.

Proposition B. Suppose $A \in \mathrm{M}_n(\mathbb{R})$ is symmetric and $Av = \lambda v$, for $\lambda \in \mathbb{R}$ and $\vec{0} \neq v \in \mathbb{R}^n$. Set $W := \mathrm{Span}\{v\}$. Then $Av \in W^\perp$, for all $v \in W^\perp$.

We then presented the proof of the Spectral Theorem for Real Symmetric Matrices. The proof was by induction on n, the case $n = 1$ being trivial. We then found an eignevalue $\lambda \in \mathbb{R}$ and a unit eigenvector $u \in \mathbb{R}^n$. For $W := \mathrm{Span}\{u\}$, $\mathbb{R}^n = W \oplus W^\perp$. Taking an orthonormal basis for V by extending u to an orthonormal basis for $W^\perp$gave rise to an orthogonal matrix P such that $P^{-1}AP = \tilde{B}$, where

$$\tilde{B} = \left(\begin{array}{c|c} \lambda & \mathbf{0} \\ \hline \mathbf{0} & B \end{array}\right).$$

Upon showing that B is an $(n-1) \times (n-1)$ symmetric matrix, by induction, we have an orthogonal $(n-1) \times (n-1)$ matrix such that $Q_0^{-1} B Q_0 = D$, an $(n-1) \times (n-1)$ diagonal matrix. By Comment 7,

$$Q = \left(\begin{array}{c|c} 1 & \mathbf{0} \\ \hline \mathbf{0} & Q_0 \end{array}\right)$$

is orthogonal, and an easy calculation shows that

$$(PQ)^{-1} A (PQ) = \left(\begin{array}{c|c} \lambda & \mathbf{0} \\ \hline \mathbf{0} & D \end{array}\right),$$

which completed the proof.

We ended class by considering an arbitrary inner product space V over $\mathbb{R}$ and defining $T \in \mathcal{L}(V, V)$ to be a *symmetric linear operator* if and only if $\langle T(v), w \rangle = \langle v, T(w) \rangle$, for all $v, w \in V$. This was motivated by the observation that $A \in \mathrm{M}_n(\mathbb{R})$ is symmetric *if and only if* $\langle Av, w \rangle = \langle v, Aw \rangle$, for all $v, w \in \mathbb{R}^n$.

Monday, September 30. We began class by proving the converse of the spectral theorem for symmetric matrices, namely, if $A \in \mathrm{M}_n(\mathbb{R})$ is orthogonally diagonalizable, then A is symmetric. We then turned to stating and proving this theorem for $T \in \mathcal{L}(V,V)$, recalling that T is symmetric (by definition) if for all $v, w \in V$, we have $\langle T(v), w\rangle = \langle v, T(w)\rangle$.

Spectral Theorem for Symmetric Linear Operators. Let V be a finite dimensional inner product space over $\mathbb{R}$ and $T \in \mathcal{L}(V,V)$. If T is symmetric, then T is orthogonally diagonalizable, i.e., there exists an orthonormal basis for V consisting of eigenvectors for T. Conversely, if T is orthogonally diagonalizable, then T is symmetric.

The proof consisted of transcribing the theorem for T to the setting over matrices over $\mathbb{R}$ and back again. This was facilitated by the following two observations.

Observations. Let V be an inner product space over $\mathbb{R}$ and $E \subseteq V$ an orthonormal basis.

(i) $\langle v, w\rangle = \langle [v]_E, [w]_E\rangle$, for all $v, w \in V$. The first inner product in V, the second in $\mathbb{R}^n$.
(ii) T is symmetric if and only if $[T]_E^E$ is symmetric.

Wednesday, October 2. We began class by asking what form the Spectral Theorem for Symmetric Matrices over $\mathbb{R}$ might take over $\mathbb{C}$. We noted that symmetry might not be the right property by exhibiting a 2×2 symmetric matrix over $\mathbb{C}$ that does not satisfy the key property $\langle Av, w\rangle = \langle v, Aw\rangle$ used in the real case. After defining the concept of the adjoint of $A \in \mathrm{M}_n(\mathbb{C})$, namely A^* is the conjugate transpose of A, we saw that $\langle Av, w\rangle = \langle v, A^*w\rangle$, for all $v, w \in \mathbb{C}$, so that A is *self-adjoint*, i.e., $A^* = A$, if and only if $\langle Av, w\rangle = \langle v, Aw\rangle$, for all $v, w \in \mathbb{C}^n$. We then gave the following definition:

Definition. $Q \in \mathrm{M}_n(\mathbb{C})$ is said to be *unitary* if its columns form an orthonormal basis for $\mathbb{C}^n$, or equivalently, $Q^{-1} = Q^*$.

We discussed how the unitary property replaces the orthogonal property when moving the discussion from $\mathbb{R}$ to $\mathbb{C}$. We then made two key observations that were analogous to key steps in the proof of the real spectral theorem:

Observations. 1. If $A \in \mathrm{M}_n(\mathbb{C})$ is self-adjoint, then all of its eigenvalues are real.

2. If v is an eigenvector of A with value λ, and we set $W := \mathrm{Span}\{v\}$, then $Aw \in W^\perp$, for all $w \in W^\perp$, provided A is self-adjoint.

We then showed how using the two observations above in exactly the same way they were used in the real case, one could derive:

First Spectral Theorem over $\mathbb{C}$. Suppose $A \in \mathrm{M}_n(\mathbb{C})$. Then A is self-adjoint if and only if A has real eigenvalues and there exists a unitary matrix $Q \in \mathrm{M}_n(\mathbb{C})$ such that $Q^*AQ = D$, a diagonal matrix. In other words, A is self-adjoint if and only if it has real eigenvalues and is *unitarily diagonalizable*.

We ended class by noting that if we want a condition on $A \in \mathrm{M}_n(\mathbb{C})$ that is equivalent to being unitarily diagonalizable, then one needs something slightly weaker that self-adjointness. This lead to defining a matrix $A \in \mathrm{M}_n(\mathbb{C})$ to be *normal* if $AA^* = A^*A$. We ended class by noting that $A = \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix}$ is a normal matrix that is not self-adjoint.

Friday, October 4. After reviewing the definitions of normal matrix and unitary matrix, we stated the

Second Spectral Theorem over $\mathbb{C}$. The matrix $A \in \mathrm{M}_n(\mathbb{C})$ is normal if and only if it is unitarily diagonalizable, i.e., A is normal if and only if there exists a unitary $Q \in \mathrm{M}_n(\mathbb{C})$ such that Q^*AQ is a diagonal matrix,

We then showed directly that the normal matrix $A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ is unitarily diagonalizable. This was followed by discussions and proofs of the following two lemmas and key propositions:

Lemma 1. Given $B, C \in \mathrm{M}_n(\mathbb{C})$, if $\langle v, Bw\rangle = \langle v, Cw\rangle$, for all $v, w \in \mathbb{C}^n$, then $B = C$.

Lemma 2. Suppose $B \in \mathrm{M}_n(\mathbb{C})$ and $W \subseteq \mathbb{C}^n$ is a B-invariant subspace, i.e., $Bw \in W$, for all $w \in W$. Then B has an eigenvector in W.

Key Proposition 1. Suppose $A \in \mathrm{M}_n(\mathbb{C})$ is normal. Then A, A^* have a common eigenvector $v \in \mathbb{C}^n$, and if $Av = \lambda v$, then $A^*v = \overline{\lambda}v$.

Key Proposition 2. Suppose $0 \neq v \in \mathbb{C}^n$, $Av = \lambda v$, $A^*v = \overline{\lambda}v$, for $\lambda \in \mathbb{C}$. Set $W := \mathrm{Span}\{v\}$. Then $W^\perp$ both A-invariant and A^*-invariant.

Monday, October 7. The purpose of today's lecture was to present a proof of the Second Spectral Theorem for Complex Matrices as stated in the lecture of Friday, October 4. The main point was that the proof of this theorem was essentially the same as the proof of the spectral theorem over $\mathbb{R}$, once one has the Key Proposition 2 from the previous lecture in hand. Nevertheless, we went through the details, to emphasize the similarity with the proof for real matrices, as given in the lecture of Friday, September 27.

The proof was by induction on n, the case $n = 1$ being trivial. We then found an eigenvalue $\lambda \in \mathbb{C}$ and a unit eigenvector $u \in \mathbb{R}^n$, so that for $W := \mathrm{Span}\{u\}$, $\mathbb{C}^n = W \oplus W^\perp$ and $W^\perp$ is A-invariant. Taking an orthonormal basis for $\mathbb{C}^n$ by extending u to an orthonormal basis for $W^\perp$gave rise to a unitary matrix P such that $P^*AP = \tilde{B}$, where

$$\tilde{B} = \left(\begin{array}{c|c} \lambda & \mathbf{0} \\ \hline \mathbf{0} & B \end{array}\right).$$

Upon showing that B is an $(n-1)\times(n-1)$ normal matrix, by induction, we have a unitary $(n-1)\times(n-1)$ matrix Q_0 such that $Q_0^*BQ_0 = D$, an $(n-1)\times(n-1)$ diagonal matrix. As in Comment 7 from September 27,

$$Q = \left(\begin{array}{c|c} 1 & \mathbf{0} \\ \hline \mathbf{0} & Q_0 \end{array}\right)$$

is unitary, and an easy calculation shows that

$$(PQ)^*A(PQ) = \left(\begin{array}{c|c} \lambda & \mathbf{0} \\ \hline \mathbf{0} & D \end{array}\right),$$

which completed the proof.

Wednesday, October 9. We presented and proved:

Second Spectral Theorem for Linear Transformation over $\mathbb{C}$. Suppose V is a finite dimensional inner product space over $\mathbb{C}$ and $T \in \mathcal{L}(V, V)$. Then T is normal if and only if there exists an orthonormal basis $B \subseteq V$ consisting of eigenvalues of T, i.e., $[T]_B^B$ is a diagonal matrix.

For this we needed to define the adjoint T^*, for $T \in \mathcal{L}(V, V)$. We opted for the following approach. Fix an orthonormal basis $B \subseteq V$ and set $A := [T]_B^B$. Define T^* by the equation $[T^*]_B^B = A^*$. We then established the following facts:

(i) $[T(v)]_B = A^* \cdot [v]_B$, for all $v \in V$.
(ii) T^* is independent of the orthonormal basis B.
(iii) $\langle T(v), w\rangle = \langle v, T^*(w)\rangle$, for all $v, w \in V$.
(iv) T is normal if and only if $[T]_B^B$ is normal, for any orthonormal basis $B \subseteq V$.

We these facts in hand, we were able to establish the second spectral theorem for $T \in \mathcal{L}(V, V)$ by transcribing the statement of the theorem to matrices, applying conclusion of the theorem for matrices, and then transcribing back to T.

Friday, October 11. We began class by showing that if V is a real inner product space, then $T \in \mathcal{L}(V, V)$ is normal if and only if $||T(v)|| = ||T^t(v)||$, for all $v \in V$, where T^t is just T^*, but written as a transpose, since there is no conjugation over $\mathbb{R}$.

We then proceeded to discuss a spectral theorem-like result for normal matrices over $\mathbb{R}$. We first showed by direct claculation that if A is a 2×2 normal matrix over $\mathbb{R}$ that is not symmetric, then $A = \begin{pmatrix} \alpha & \beta \\ -\beta & \alpha \end{pmatrix}$, for some $\alpha, \beta \in \mathbb{R}$. We then stated the general case:

Normal Matrix Theorem over $\mathbb{R}$. Let $A \in \mathrm{M}_n(\mathbb{R})$, and assume A is not symmetric. Then A is normal if and only there exists an orthogonal matrix P such that P^tAP is block diagonal, with blocks $D, A_1, \ldots, A_r$, 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 $r \geq 1$.

While we did not give a complete proof, we outlined how an inductive proof would proceed, along similar lines to the inductive proofs for the various spectral theorems. Assuming that A is normal, the crucial point in this case, is rather than starting with a one dimensional eigenspace of $\mathbb{R}^n$ and passing to its orthogonal complement, one starts with a two dimensional subspace and passes to its orthogonal complement. The two dimensional subspace arises as the nullspace of $q(A)$, where $q(x)$ is a degree two irreducible factor of $\mu_A(x)$. Such a factor necessarily exists if A is not symmetric.

Wednesday, October 16. 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 $\sigma_1 \geq \cdots \geq \sigma_r > 0$ 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 of the statement for linear transformations, 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 also noted that the matrix form of the theorem is often presented as the:

Singular Value Decomposition. 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 $A = Q\Sigma P^*$, 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.

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

Friday, October 18. 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 $\frac{1}{\sigma_1}, \ldots, \frac{1}{\sigma_r}, 0, \ldots, 0$ 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.

Monday, October 21. 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 of the division algorithm and a generalized version of Bezout's Principle in the following form: *If $f_1(x), \ldots, f_n(x) \in F[x]$ have no common divisor, then there exist $a_1(x), \ldots, a_n(x) \in F[x]$ 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. 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 then noted a crucial consequence of the theorem is the following:

Corollary. 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.. We also noted that theorem fills in the missing part of the corollary to the diagonalization theorem from the lecture of September 16.

We ended class by stating the major theorem that is our next goal:

Rational Canonical Form Theorem via Elementary Divisors. 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.

Wednesday, October 23. We began class by stating both forms of the Rational canonical Form theorem, first the elementary divisor form, as stated in the previous lecture, and then the:

Rational Canonical Form Theorem via Invariant Factors. 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 $f_i(x)$.

The polynomials $f_1(x), \ldots, f_t(x)$ are called the *invariant factors* of T.

We then noted that for $T \in \mathcal{L}(V, V)$, if V has a basis of the form $B := \{v, T(v), \ldots, T^{n-1}(v)\}$, for some $v \in V$, with $n = \dim(V)$, then $[T]_B^B = C(\mu_{T,v}(x))$, the companion matrix of $\mu_{T,v}(x)$. This was followed by defining the *T-cyclic subspace of V generated by v* as $\text{Span}\{v, T(v), T^2(v), \ldots\}$, which we denoted by $\langle T, v\rangle$. This in turn led to the:

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) $B := \{v, T(v), \ldots, T^{e-1}(v)\}$ is a basis for $\langle T, v\rangle$.
(iii) $\dim\langle T, v\rangle = e$.
(iv) $\mu_{T,v}(x) = \mu_{T|_{\langle T,v\rangle}}(x)$.
(v) $[T|_{\langle T,v\rangle}]_B^B = C(\mu_{T,v}(x))$.

We ended class with the following observation, which essentially follows from work done in this lecture and the previous one.

Observation. For a finite dimensional vector space V and $T \in \mathcal{L}(V, V)$, the following are equivalent:

(i) V is a direct sum of T-cyclic subspaces.

(ii) There exists a basis $B \subseteq V$ such that 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 $f_i(x)$.

Friday, October 25. We continued our preliminary discussions of the Rational Canonical Form theorem, which requires us to show that V is a direct sum of cyclic subspaces. 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 $T|_{W_i}$.

We then discussed how the following theorem is key to proving the RCF theorem.

Key Theorem. Let T be a linear operator on the finite dimensional vector space V and $v \in V$ a maximal vector with respect to T. Then $\langle T, v\rangle$ has a T-invariant complement. That is, there exists a T-invariant subspace $U \subseteq V$ such that $V = \langle T, v\rangle \bigoplus U$.

We did not prove the Key Theorem, but ended class with the following example showing that, in general, a T-cyclic subspace need not admit a T-invariant complement.

Example. Let $T : \mathbb{R}^3 \to \mathbb{R}^3$ be the linear transformation $T : \mathbb{R}^3 \to \mathbb{R}^3$ such that

$$[T]_E^E = A = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix},$$

where E is the standard basis of $\mathbb{R}^3$ and set $v := \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$. 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 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 $\dim(U) \ge n - d$. If the claim holds, then

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

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 $0 \le j \le d-1$. 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 $0 \le j \le d-1$. 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 $\dim(U) \ge n - d$, 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 then used the key theorem and induction to show that if V is a finite dimensional vector space and $T \in \mathcal{L}(V, V)$, then there exist $v_1 \ldots, v_t \in V$ such that

$$V = \langle T, v_1\rangle \oplus \cdots \oplus \langle T, v_t\rangle,$$

where upon setting $f_i(x) = \mu_{T,v_i}(x)$, $f_t(x) \mid \cdots \mid f_1(x) = \mu_T(x)$.

This led immediately to

Rational Canonical Form Theorem via Invariant factors. 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 $R := \begin{pmatrix} A_1 & 0 & \cdots & 0 \\ 0 & A_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & A_t \end{pmatrix}$, where each $A_i = C(f_i(x))$, the companion matrix of $f_i(x)$. R is the rational canonical form of T.

We ended class with the matrix from of the theorem:

RCF for matrices. Let $A \in \mathrm{M}_n(F)$. Then there exists an invertible $n \times n$ matrix P, where $P^{-1}AP = R$, for R as above, with $f_1(x)|f_2(x)|\cdots|f_t(x) = \mu_T(x)$.

Wednesday, October 30. We began class by restating the invariant factor form of both the operator and matrix forms of the Rational Canonical Form theorem. We then spent the rest of the class looking at the rational canonical forms for non-diagonalizable matrices in the two and three dimensional cases.

When A is a 2×2 non-diagonalizable matrix, we noted that $\chi_A(x) = \mu_A(x) = x^2 + ax + b$ which immediately implies that the RCF is $R := \begin{pmatrix} 0 & -b \\ 1 & -a \end{pmatrix}$ and that $F^2 = \mathrm{Span}\{v, Av\}$, for a vector $v \in F^2$ maximal with respect to A. We noted if P is the matrix whose columns re v, Av, then P is the change of basis matrix, i.e., $P^{-1}AP = R$. Finally, we observed that if $\mu_A(x)$ is irreducible, v can be any non-zero vector in F^2, while if $\mu_A(x)$ is reducible, it must have the form $\mu_A(x) = (x - \lambda)^2$ and v can be any vector not in E_λ.

When A is a non-diagonalizable 3×3 matrix, we noted that either $\mu_A(x) = x^3 + ax^2 + bx + c$, a degree three polynomial, or $\mu_A(x) = (x - \lambda)^2$, for some $\lambda \in F$. In the first case, the rational canonical form is $R = \begin{pmatrix} 0 & 0 & -c \\ 1 & 0 & -b \\ 0 & 1 & -a \end{pmatrix}$, while in the second case $R = \begin{pmatrix} 0 & -\lambda^2 & 0 \\ 1 & 2\lambda & 0 \\ 0 & 0 & \lambda \end{pmatrix}$. In the second case, the change of basis matrix P will have columns v, Av, w, where w is an eigenvector for A and v is not an eigenvector for A. In the first case, there exists $v \in F^3$ such that the change of basis matrix will have columns v, Av, A^2v, with v a maximal vector for A. To find v: If $\mu_A(x)$ is irreducible, any non-zero $v \in F^3$ works. Otherwise: (i) $\mu_A(x) = p(x)(x-\lambda)$, with $p(\lambda) \neq 0$ and either $p(x)$ is irreducible or $p(x) = (x-\lambda_1)^2$, or (ii) $\mu_A(x) = (x-\lambda)^3$. In the first of these cases, the primary decomposition theorem gives $F^3 = W \bigoplus E_\lambda$. W is two-dimensional, so the degree two case can be used to find $w_1 \in W$ a maximal vector for A restricted to W. Taking $w_2 \in E_\lambda$, then $v = w_1 + w_2$ is the required vector. If $\mu_A(x) = (x - \lambda)^3$, one takes v any vector not in the null space of $(A - \lambda I_3)^2$.

Friday, November 1. We discussed homework problems in preparation for Exam 2.

Monday, November 4. We began a discussion and proof 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. The points of the proof for this case was to apply the RCF theorem to $S := T - \lambda I$.

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 $1 \leq i \leq r$, $e_i = e_{i1} \geq \cdots \geq e_{is_i}$ 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 $A_i = \begin{pmatrix} J(\lambda_i, e_{i,1}) & 0 & \cdots & 0 \\ 0 & J(\lambda_i, e_{i,2}) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & J(\lambda_i, e_{i,s_i}) \end{pmatrix}$.

The point of the proof was to use the primary decomposition of V and apply the case of one irreducible factor to each T restricted to a primary component.

We then wrote 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 ended class by proving 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 $1 \leq i \leq r$, $s_i = \dim(E_{\lambda_i})$.

Thus the number of Jordan blocks associated to λ_i equals the dimension of the corresponding eigenspace E_{λ_i}.

Wednesday, November 6. Given $\mu_A(x) = (x-\lambda_1)^{e_1} \cdots (x-\lambda_r)^{e_r}$, we reviewed the fact that 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, where A_i is the corresponding block as given in the previous lecture. 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 $(t_i - t_{i-1}) - (t_{i+1} - t_i)$.

We then had a 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 $1 \leq i \leq r$, calculate $t_{i,j} := \mathrm{nullity}(A-\lambda_i)^j$ 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 $\mathbb{R}^3$.

Friday, November 8. 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 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 $c \geq 2$, and $z \in \mathbb{C}$, z has c distinct cth roots.

We continued our discussion of finding roots and powers of diagonalizable matrices by finding pth roots of $A \begin{pmatrix} -21 & -50 \\ 15 & 34 \end{pmatrix}$. If we fix c a pth root of 4 and d a pth root of 9, then the matrices $H = \begin{pmatrix} 6c - 5d & 10 - 10d \\ -3c + 3d & -5c + 6d \end{pmatrix}$ were shown to be pth roots of A.

We then turned our attention to discussing and ultimately 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.

We began by looking at a 2×2 Jordan block $J = \begin{pmatrix} \lambda & 0 \\ 1 & \lambda \end{pmatrix}$, with $0 \neq \lambda \in \mathbb{C}$. We then set $M := \begin{pmatrix} 0 & 0 \\ \frac{1}{\lambda} & 0 \end{pmatrix}$ and $B_0 := I_2 + \frac{1}{n} \cdot M$, so that $B_0^n = I_2 + M = \begin{pmatrix} 1 & 0 \\ \frac{1}{\lambda} & 1 \end{pmatrix}$. Thus, $\lambda B_0^n = J$. It followed that for any $\omega \in \mathbb{C}$ with $\omega^n = \lambda$, $B^n = J$, for $B = \omega B_0$. We ended class by applying this technique to find three cube roots of the matrix $A = \begin{pmatrix} 37 & -49 \\ 25 & 33 \end{pmatrix}$.

Monday, November 11. We continued our of the following theorem:

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

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.

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 13. We began class by 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 then showed 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 $\mathbf{X}(t) = e^{At} \cdot \mathbf{X}(0)$.

This was followed by beginning a discussion of quotient spaces. We began with a subspace W of V, and showed that the following relation is an equivalence relation: For $v_1, v_2 \in V$, $v_1 \equiv v_2 \bmod W$ if and only if $v_1 - v_2 \in W$. We then noted that the equivalence class of $v \in V$ as $v + W := \{v + w \mid w \in W\}$. We then discussed what it means for operations on equivalence classes to be well defined.

Friday, November 15. We continued our introductory discussion of quotient spaces. We began by recalling that 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 vector space over F and $W \subseteq V$ a non-zero subspace. If $\{w_\beta\}_{\beta \in B}$ is a basis for W and $\{v_\alpha + W\}_{\alpha \in A}$ is a vector basis for V/W, then $B := \{v_\alpha\}_\alpha \cup \{w_\beta\}_\beta$ is a basis for V. In particular, V is finite dimensional if and only if W and V/W are finite dimensional, in which case $\dim V = \dim(V/W) + \dim W$.

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 18. We began calss by proving the:

Second Isomorphism Theorem. Let $U \subseteq W \subseteq V$ be subspaces. Then W/U is a subspace of V/U and $(V/U)/(W/U) \cong V/W$.

We then began discussion of tensor products. We followed our tensor product handout. Suppose V is a vector space over F and K is a field containing F. Is there a natural way to extend the scalars of V from F

to K, i.e., is there a way to make V into a vector space over K? If we regard K as a vector space over F, then one way to extend scalars is to take the tensor product of K with V. Of course, there are more direct ways of doing this in the context of vector spaces over a field, but the tensor product appears throughout mathematics and can be quite subtle when the objects in question are not vector spaces. For example, the tensor product of non-zero objects over $\mathbb{Z}$ can be zero!

If V and W are vector spaces the tensor product will be a vector space over F generated by vectors that can be written as $v \otimes w$, with $v \in V$ and $w \in W$, where such expressions satisfy the following *bilinear relations*:

$$(\lambda v_1 + v_2) \otimes w = \lambda(v_1 \otimes w) + v_2 \otimes w \quad \text{and} \quad v \otimes (\lambda w_1 + w_2) = \lambda(v \otimes w_1) + v \otimes w_2,$$

for all $v_i \in V$, $w_i \in W$ and $\lambda \in F$. The basic idea of the construction is to start with a large vector space with basis elements consisting of the pairs $(v, w) \in V \times W$ and then impose the required bilinear relations by modding out the subspace generated by the corresponding bilinear expressions.

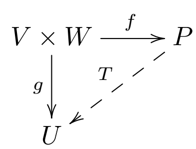
We begin with the formal definition of the tensor product. This definition is expressed in terms of a universal property the tensor product enjoys in relation to certain commutative diagrams. While this definition is very abstract, it is the principal tool for developing properties of the tensor product. In fact, the construction of the tensor product plays a somewhat minor role in this regard.

Definitions. Suppose V and W are vector spaces over the field F.

1. A *bilinear map* on $V \times W$ is a function $f : V \times W \to P$, where P is a vector space over F, satisfying:
   (i) $f(v_1 + v_2, w) = f(v_1, w) + f(v_2, w)$, for all $v_i \in V$ and $w \in W$.
   (ii) $f(v, w_1 + w_2) = f(v, w_1) + f(v, w_2)$, for all $v \in V$ and $w_i \in W$.
   (iii) $f(\lambda v, w) = \lambda f(v, w) = f(v, \lambda w)$, for all $v \in V, w \in W, \lambda \in F$.

In other word, for any fixed $v_0 \in V$, $f(v_0, w)$ is a linear operator on W, and for any fixed $w_0 \in W$, $f(v, w_0)$ is a linear operator on V.

2. A *tensor product* of V and W consists of a pair (P, f), where P is a vector space over F and $f : V \times W \to P$ is a bilinear map such that given a vector space U and bilinear map $g : V \times W \to U$, there exists a unique linear transformation $T : P \to U$ such that $T \circ f = g$. Diagrammatically, we may represent this condition as follows:

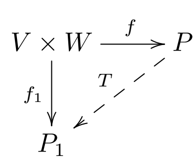


where the filled in arrows f and g are indicating maps that are given and the dotted arrow T indicates the map that results from invoking the definition. The condition $T \circ f = g$ is often expressed by saying that the diagram above is a *commutative diagram.*

Let us assume temporarily that tensor products exist. We will show how to derive some basic properties of the tensor product using the definition above. We begin with the uniqueness of the tensor product. It is not difficult to show that if (P, f) is a tensor product of V and W and $\alpha : P \to P_1$ is an isomorphism of vector spaces, then for $f_1 := \alpha \circ f$, (P_1, f_1) is also a tensor product of V and W. Our first proposition shows that this is the only way of creating another tensor product of V and W. In other words, tensor products are unique up to isomorphism. Thus, we will refer to the resulting vector space as *the* tensor product of V and W.

Proposition 1. *Let (P, f) and (P_1, f_1) be tensor products of the vectors spaces V and W. Then there exists an isomorphism $T : P \to P_1$ such that $f_1 = T \circ f$.*

Proof. Using the definition of tensor product twice, we have the following commutative diagrams



and

$$\begin{array}{ccc} V \times W & \xrightarrow{f_1} & P_1 \\ {\scriptstyle f}\downarrow & \swarrow_{S} & \\ P & & \end{array}$$

with induced linear transformations $T : P \to P_1$ and $S : P_1 \to P$ satisfying $f_1 = T \circ f$ and $f = S \circ f_1$. Thus, $f = S \circ (T \circ f) = (ST) \circ f$, which means we have a commutative diagram

$$\begin{array}{ccc} V \times W & \xrightarrow{f} & P\ . \\ {\scriptstyle f}\downarrow & \swarrow_{ST} & \\ P & & \end{array}$$

But this diagram also commutes if we replace ST by the identity map Id on P. Since by definition, there is a unique diagonal map making this diagram commute, we must have $ST = Id$. In exactly the same way, we see that TS is the identity on P_1. This means that T is an isomorphism, and since $f_1 = T \circ f$, the proof is complete. □

Thus, once we show that tensor products exist, they are unique up to isomorphism. For each V and W, let us choose a representative of the isomorphism class of tensor products and denote it by $V \otimes W$.

We derive one more property of the tensor product before establishing the existence of tensor products.

Proposition 2. *Given vector spaces V and W over F, $V \otimes W \cong W \otimes V$.*

Proof. The proof is similar the the proof of the previous proposition. Let $f : V \times W \to V \otimes W$ and $h : W \times V \to W \otimes V$ be the given bilinear maps. Define $g : V \times W \to W \otimes V$ by $g(v, w) := h(w, v)$. If we show that g is bilinear, then we have a commutative diagram

$$\begin{array}{ccc} V \times W & \xrightarrow{f} & V \otimes W \\ {\scriptstyle g}\downarrow & \swarrow_{T} & \\ W \otimes V & & \end{array}$$

with a linear transformation $T : V \otimes W \to W \otimes V$ such that $f \circ T = g$. To see that g is bilinear, note that for $\lambda \in F$, $v_1, v_2 \in V$ and $w \in W$,

$$g(\lambda v_1 + v_2, w) = h(w, \lambda v_1 + v_2) = \lambda h(w, v_1) + h(w, v_2) = \lambda g(v_1, w) + g(v_2, w).$$

The proof that g is linear in its second variable is similar. Now, by symmetry, we also have a bilinear map $k : W \times V \to V \otimes W$ and commutative diagram

$$\begin{array}{ccc} W \times V & \xrightarrow{h} & W \otimes V \\ {\scriptstyle k}\downarrow & \swarrow_{S} & \\ V \otimes W & & \end{array}$$

with a linear transformation $S : W \otimes V \to V \otimes W$ such that $S \circ h = k$. Here, $k(w, v) = f(v, w)$. Now, let $(v, w) \in V \times W$. Then,

$$ST \circ f(v, w) = Sg(v, w) = Sh(w, v) = k(w, v) = f(v, w),$$

and hence the diagram

$$\begin{array}{ccc} V \times W & \xrightarrow{f} & V \otimes W \\ {\scriptstyle f}\downarrow & \swarrow_{ST} & \\ V \otimes W & & \end{array}$$

is commutative, i.e., $f = ST \circ f$. But the diagram is also commutative if we replace the linear transformation ST by the identity map Id on $V \otimes W$. By the uniqueness of the induced maps, ST is the identity on $V \otimes W$. Similarly, TS is the identity on $W \otimes V$. This shows $V \otimes W \cong W \otimes V$, as required. $\square$