A. Miller
M542
Spring 2000

1 Linear Transformations

In this section we consider only finite dimensional vector spaces V or W over an arbitrary field $\mathbb{F}$.

Theorem 1.1 *Every linear transformation $L : \mathbb{F}^n \to \mathbb{F}^m$ is determined by an $m \times n$ matrix A:*

$$L(X) = AX$$

for every $X \in \mathbb{F}^n$

Proof:
Given A since

$$A(X+Y) = AX + AY \text{ and } A(aX) = a(AX)$$

it is clear that $L(X) = AX$ is a linear transformation.

For the converse, assume $L : \mathbb{F}^n \to \mathbb{F}^m$ is a linear transformation. The canonical basis for $\mathbb{F}^n$ is the sequence of columns of the $n \times n$ identity matrix, $I_{n\times n}$. So let $e_i = \mathrm{col}_i(I_{n\times n})$. Note for any vector

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = x_1e_1 + x_2e_2 + \cdots + x_ne_n.$$

Let A be the matrix such that

$$L(e_i) = \mathrm{col}_i(A)$$

for every $i = 1, \ldots, n$. Then

$$\begin{aligned} L\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} &= L(x_1e_1 + x_2e_2 + \cdots + x_ne_n) \\ &= x_1L(e_1) + x_2L(e_2) + \cdots + x_nL(e_n) \\ &= x_1\mathrm{col}_1(A) + x_2\mathrm{col}_2(A) + \cdots + x_n\mathrm{col}_n(A) \\ &= A\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \end{aligned}$$

:foorP

Theorem 1.2 *Suppose V and W are vector space over a field $\mathbb{F}$. If $\dim(V) = \dim(W)$, then V is isomorphic to W.*

Proof:
This is true in general, but we only proof it in case the spaces have finite dimension.

Let $v_1, \ldots, v_n$ be a basis for V and $w_1, \ldots, w_n$ be a basis for W. Given any $v \in V$ there exists a unique sequence of scalars $a_1, \ldots, a_n$ such that

$$v = a_1 v_1 + \cdots + a_n v_n$$

They exists because $v_1, \ldots, v_n$ span V and they are unique because

$$v = a_1 v_1 + \cdots + a_n v_n \text{ and } v = b_1 v_1 + \cdots + b_n v_n$$

implies

$$z = (a_1 - b_1)v_1 + \cdots + (a_n - b_n)v_n$$

and so by linear independence $a_1 = b_1, \ldots, a_n = b_n$. Now define $L : V \to W$ by

$$L(a_1 v_1 + \cdots + a_n v_n) = a_1 w_1 + \cdots + a_n w_n$$

Now we check that L is linear. Suppose $v = a_1 v_1 + \cdots + a_n v_n$ and $w = b_1 v_1 + \cdots + b_n v_n$ and a, b are scalars. Then

$$\begin{aligned} L(av + bw) &= L((aa_1 + bb_1)v_1 + \cdots + (aa_n + bb_n)v_n) \\ &= (aa_1 + bb_1)w_1 + \cdots + (aa_n + bb_n)w_n \\ &= a(a_1 w_1) + \cdots a_n w_n) + b(b_1 w_1 + \cdots + b_n w_n) \\ &= aL(v) + bL(w) \end{aligned}$$

Next to see that L is one-to-one and onto, note that for any w there exists unique $a_1, a_2, \ldots, a_n$ such that $w = a_1 w_1 + \cdots + a_n w_n$ and so $v = a_1 v_1 + \cdots + a_n v_n$ is the unique element of V such that $L(v) = w$.
:foorP

Definition 1.3 *For $L : V \to W$ a linear transformation, define the <u>null space</u> (or kernel) of L, $\text{null}(L)$, and the <u>range space</u> of L, $\text{range}(L)$ as follows:*
(a) $\text{null}(L) = \{v \in V : L(v) = z\}$
(b) $\text{range}(L) = \{w \in W :$ *there exists* $v \in V$ *such that* $L(v) = w\}$

Proposition 1.4 $\text{null}(L)$ *and* $\text{range}(L)$ *are subspaces of V and W, respectively.*

Proof:
$\text{null}(L)$: Suppose $u, v \in \text{null}(L)$ and a, b scalars. Then $L(u) = z$ and $L(v) = z$ and

$$L(au + bv) = aL(u) + bL(v) = az + bz = z + z = z.$$

Hence $au + bv \in \text{null}(L)$

$\text{range}(L)$: Suppose $w_1, w_2 \in \text{range}(L)$ and a_1, a_2 are scalars. Then there exists $v_1, v_2 \in V$ such that $L(v_1) = w_1$ and $L(v_2) = w_2$. Then

$$L(a_1 v_1 + a_2 v_2) = a_1 L(v_1) + a_2 L(v_2) = a_1 w_1 + a_2 w_2$$

and so $a_1 w_1 + a_2 w_2 \in \text{range}(L)$.
:foorP

Theorem 1.5 *Suppose $L : V \to W$ is a linear transformation. Then*

$$\dim(V) = \dim(\text{null}(L)) + \dim(\text{range}(L)).$$

Proof:

Let $u_1, \ldots, u_n$ be a basis for null(L) and $v_1, \ldots, v_m$ be a basis for range(L). Let $w_1, \ldots, w_m$ be in V such that $L(w_i) = v_i$.

Claim. $u_1, \ldots, u_n, w_1, \ldots, w_m$ is a basis for V.

Proof:

First we show they are linearly independent. Suppose

$$c_1u_1 + \cdots + c_nu_n + d_1w_1 + \cdots + d_mw_m = z$$

then applying L we get

$$z = L(z) = L(c_1u_1 + \cdots + c_nu_n) + L(d_1w_1 + \cdots + d_mw_m) = z + d_1v_1 + \cdots + d_mv_m$$

because $c_1u_1 + \cdots + c_nu_n \in \text{null}(L)$. Hence

$$d_1v_1 + \cdots + d_mv_m = z$$

and since the $v's$ are linearly independent we get $d_1 = \cdots = d_m = 0$. Hence $c_1u_1 + \cdots + c_nv_n = z$ and since the u's are linearly independent $c_1 = \cdots = c_n = 0$. So, our set is linearly independent.

Now we must see that they span V. Suppose $v \in V$ is arbitrary and let $L(v) = d_1v_1 + \cdots + d_mv_m$, then set $u = d_1w_1 + \cdots + d_mw_m$ and notice that

$$L(u) = L(d_1w_1 + \cdots + d_mw_m) = d_1L(w_1) + \cdots + d_mL(w_m) = d_1v_1 + \cdots + d_mv_m.$$

It follows that $L(u) = L(v)$ and so $L(u - v) = z$ since L is linear and therefore $v - u \in \text{null}(L)$. Therefore (since the u's are basis for null(L)) there exists c's such that

$$v - u = c_1u_1 + \cdots + c_nu_n$$

and therefore

$$v = u + c_1u_1 + \cdots + c_nu_n = d_1w_1 + \cdots + d_mw_m + c_1u_1 + \cdots + c_nu_n.$$

Thus $v \in span(\{u_1, \ldots, u_n, w_1, \ldots, w_m\})$. This proves the Claim.

:foorP From the Claim it follows that the dimension of V is $n + m$ or in other words dim(null(L)) + dim(range(L)). This proves the Theorem.

:foorP

Theorem 1.6 *Suppose $A, B \in \mathbb{F}^{n\times n}$, then A is similar to B iff there exists a basis $v_1, v_2, \ldots, v_n$ for $\mathbb{F}^{n\times 1}$ such that for every j*

$$\mathbf{A}(v_j) = \sum_{i=1}^{n} entry_{ij}(B)v_i.$$

Furthermore, given such a basis if P is the invertible matrix where $col_j(P) = v_j$ for each j, then P witnesses their similarity, i.e., $A = PBP^{-1}$.

Proof:
The equation $A = PBP^{-1}$ is the same as $AP = PB$. But $col_j(AP) = Acol_j(P)$ and $col_j(PB) = Pcol_j(B)$ and $Pcol_j(B) = \Sigma_{i=1}^n b_{ij} col_i(P)$. Hence

$$\mathbf{A}(col_j(P)) = \Sigma_{i=1}^n b_{ij} col_i(P).$$

:foorP

2 Triangulizability

In this section we consider only square matrices over the field of complex numbers, $\mathbb{C}$. All vector spaces V, W, etc are assumed to be finite dimensional vector spaces over the complex numbers.

Theorem 2.1 *Suppose for every linear transformation $L : V \to V$ that V has a basis $v_1, v_2, \ldots, v_n$ such that*

$$L(v_k) \in \text{span}(\{v_1, \ldots, v_k\})$$

for every k with $1 < k \leq n$. Then every $n \times n$ matrix is similar to to an upper triangular matrix.

Proof:
This follows from Theorem 1.6. Suppose A is an $n \times n$ matrix and $L_A : \mathbb{C}^n \to \mathbb{C}^n$ is the linear transformation determined by A. Then

$$L_A(v_k) = b_{1,k} v_1 + b_{2,k} v_k + \cdots + b_{k,k} v_k$$

means that A is similar to B where

$$\text{col}_k(B) = \begin{bmatrix} b_{1,k} \\ b_{2,k} \\ \vdots \\ b_{k,k} \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

Hence, B is an upper triangular matrix.
:foorP

Theorem 2.2 *Suppose V is a finite dimensional vector space over $\mathbb{C}$ and $L : V \to V$ is a linear transformation. Then there exists a nontrivial $v \in V$ and $\lambda \in \mathbb{C}$ such that $L(v) = \lambda v$.*

Proof:
Any matrix $A \in \mathbb{C}^{n\times n}$ has an eigenvalue because the characteristic polynomial,

$$p(x) = det(A - xI)$$

is a polynomial of degree n and so has a root in $\mathbb{C}$. Hence, there exists $\lambda \in \mathbb{C}$ such that $Av = \lambda v$ has a nontrivial solution for v.

But V is isomorphic to $\mathbb{C}^{n\times 1}$ where $n = dim(V)$. If $\Phi : V \to \mathbb{C}^{n\times 1}$ is an isomorphism, then $\mathbf{A} = \Phi L \Phi^{-1}$ is a linear transformation $\mathbf{A} : \mathbb{C}^{n\times 1} \to \mathbb{C}^{n\times 1}$. So it corresponds to a matrix $A \in \mathbb{C}^{n\times n}$. If $Au = \lambda u$, then

$$\Phi L \Phi^{-1} u = \lambda u \text{ so } L\Phi^{-1}u = \Phi^{-1}\lambda u = \lambda \Phi^{-1} u.$$

So putting $v = \Phi^{-1}u$ gives us $L(v) = \lambda v$.

Alternative proof without using determinants: Consider

$$\mathcal{I} = \{f(x) \in \mathbb{C}[x] : f(A) = 0\}$$

this in an ideal in the ring of polynomials. The vector space of $n \times n$ matrices has dimension n^2. Hence the sequence $\{I, A, \ldots, A^{n^2}\}$ of n^2+1 matrices is linear dependent. It follows $\mathcal{I}$ is a nontrivial ideal. The minimal polynomial $q(x)$ of A is the generator of this ideal. Let $q(x) = (x - \lambda)f(x)$. Since $f(x)$ has degree less than q it must be that some column v of $f(A)$ is nontrivial. But then $(A - \lambda I)v = 0$.
:foorP

Definition 2.3 *If W_1 and W_2 are subspaces of a vector space V such that $W_1 \cap W_2 = \{\mathbf{0}\}$, then define*

$$W_1 \bigoplus W_2 = \{w_1 + w_2 : w_1 \in W_1, w_2 \in W_2\}.$$

Whenever we write $W_1 \bigoplus W_2$ we will be assuming that $W_1 \cap W_2 = \{\mathbf{0}\}$.

Lemma 2.4 *For V a vector space and W_i's subspaces:*
(a) $W_1 \bigoplus W_2$ is a subspace of V
(b) For any $u \in W_1 \bigoplus W_2$, $w_1, w_1' \in W_1$, and $w_2, w_2' \in W_2$, if $u = w_1 + w_2$ and $u = w_1' + w_2'$ then $w_1 = w_1'$ and $w_2 = w_2'$.
(c) If B_1 is a basis for W_1 and B_2 is a basis for W_2, then $B_1 \cup B_2$ is a basis for $W_1 \bigoplus W_2$.
(d) Given $W_1 \bigoplus W_2$ define $P : W_1 \bigoplus W_2 \to W_2$ by $P(w_1+w_2) = w_2$ where $w_2 \in W_2$, and $w_1 \in W_1$, then P is a linear transformation such that $kernel(P) = W_1$ and $P(v) = v$ for all $v \in W_2$. (P is called a projection.)
(e) For any W_1 a subspace of a finite dimensional V there exists W_2 a subspace of V such $V = W_1 \bigoplus W_2$.

Proof:
(b) $W_1 \ni w_1 - w_1' = w_2 - w_2' \in W_2$ so $w_1 - w_1' = w_2 - w_2' - \mathbf{0}$, since $W_1 \cap W_2 = \{\mathbf{0}\}$.
(e) Take any basis $v_1, v_2, \ldots, v_n$ for W_1 extend it to a basis for V say

$$v_1, v_2, \ldots, v_n, u_1, u_2, \ldots, u_m.$$

Let $W_2 = span(\{u_1, u_2, \ldots, u_m\})$.
:foorP

Lemma 2.5 *Suppose $L : V \to V$ is a linear transformation and $W \neq V$ a proper subspace of V. Then there exists $v \in V$ such that $v \notin W$ and $\lambda \in \mathbb{C}$ such that $L(v) - \lambda v \in W$.*

Proof:
Let W' be a subspace such that $V = W \bigoplus W'$ and define $P : V \to W'$ by $P(w + w') = w'$. Then P is a linear transformation. Define $L_0 : W' \to W'$ by $L_0(w) = P(L(w))$ so it is a linear transformation. Using Theorem 2.2 there exists $\lambda \in \mathbb{C}$ and a nontrivial $v \in W'$ such that $L_0(v) = \lambda v$. Since P is the identity on W' we have that $P(\lambda v) = \lambda v$. Consequently

$$P(L(v) - \lambda v) = P(L(v)) - P(\lambda v) = L_0(v) - \lambda v = \mathbf{0}.$$

Since W is the kernel of P we have $L(v) - \lambda v \in W$.
:foorP

Theorem 2.6 *Suppose $L : V \to V$ is a linear transformation. Then V has a basis $v_1, v_2, \ldots, v_n$ such that for each $k = 1, \ldots, n \quad L(v_k) \in span(\{v_1, v_2, \ldots, v_k\})$.*

Proof:
Inductively build this sequence. By Theorem 2.2 we can find v_1 an eigenvector of L and so, $L(v_1) = \lambda_1 v_1 \in span(v_1)$. Given $v_1, v_2, \ldots, v_k$ let $W_k = span\{v_1, v_2, \ldots, v_k\}$. If $W_k \neq V$ apply Lemma 2.5 to get $v_{k+1} \notin W_k$ such that for some λ_{k+1}

$$L(v_{k+1}) - \lambda_{k+1} v_{k+1} \in W_k \text{ and so } L(v_{k+1}) \in span\{v_1, v_2, \ldots, v_{k+1}\}.$$

:foorP

Corollary 2.7 *Every matrix $A \in \mathbb{C}^{n \times n}$ is similar to an upper triangular matrix.*

Definition 2.8 *A sequence $u_1, u_2, \ldots, u_n$ in an inner product space is orthonormal iff for all i, j*

$$\langle u_i, u_j \rangle = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

Proposition 2.9 *Orthonormal sequences are linearly independent.*

Proof:

$$\langle c_1u_1 + \cdots + c_nu_n, u_i\rangle = c_1\langle u_1, u_i\rangle + \cdots c_i\langle u_i, u_i\rangle + \cdots c_n\langle u_n, u_i\rangle = c_i\langle u_i, u_i\rangle = c_i$$

Hence if $c_1u_1 + \cdots + c_nu_n = \mathbf{0}$, then

$$0 = \langle \mathbf{0}, u_i\rangle = \langle c_1u_1 + \cdots + c_nu_n, u_i\rangle = c_i$$

and therefore $c_i = 0$ for all i.
:foorP

Theorem 2.10 *(Gram-Schmidt Orthogonalization Process). If $v_1, v_2, \ldots, v_n$ and linear independent, then there exists $u_1, \ldots, u_n$ an orthonormal sequence such that for every $k = 1, \ldots, n$*

$$span(\{v_1, \ldots, v_k\}) = span(\{u_1, \ldots, u_k\}).$$

Proof:
Set $u_1 = \frac{1}{||v_1||}v_1$. Given $u_1, u_2, \ldots, u_k$ we will construct u_{k+1}. Let $c_i = \langle v_{k+1}, u_i\rangle$ for each $i = 1, \ldots, k$ and set $w = v_{k+1} - (c_1u_1 + \cdots + c_ku_k)$.

1. $w \neq \mathbf{0}$
2. $\langle w, u_i\rangle = 0$ all $i = 1, \ldots, k$
3. $w \in span\{v_1, \ldots, v_{k+1}\}$
4. $v_{k+1} \in span\{u_1, \ldots, u_k, w\}$
5. $span\{v_1, \ldots, v_{k+1}\} = span\{u_1, \ldots, u_k, w\}$

Then set

$$u_{k+1} = \frac{1}{||w||}w.$$

:foorP

Theorem 2.11 *Suppose V is a finite dimensional inner product space and $L : V \to V$ is a linear transformation, then V has an orthonormal basis $u_1, u_2, \ldots, u_n$ such that for each $k = 1, \ldots, n$*

$$L(u_k) \in span(\{u_1, u_2, \ldots, u_k\}).$$

Proof:
Use Theorem 2.6 to get $v_1, \ldots, v_n$ and then apply the Gram-Schmidt orthogonalization process to get $u_1, u_2, \ldots, u_n$.
:foorP

Corollary 2.12 *(Schur) For every matrix $A \in \mathbb{C}^{n\times n}$ there exists a unitary matrix P (ie $P^{-1} = P^*$ the conjugate transpose) such that $P^{-1}AP$ is an upper triangular matrix.*

Proof:
A matrix P is unitary iff $col_1(P), \ldots, col_n(P)$ is an orthonormal basis.
:foorP

Corollary 2.13 *If $A = A^*$ then A is (unitarily) similar to a diagonal matrix all of whose entries are real. Hence all the eigenvalues of A are real.*

Proof:
Let $P^*AP = U$ where U is upper triangular. Then

$$U^* = (P^*AP)^* = P^*A^*P^{**} = P^*AP = U.$$

Hence $U^* = U$ and so U is diagonal and all of its entries are real.
:foorP

3 Jordan Normal Form

In this section all vector spaces V, W, etc., are assumed to be finite dimensional vector spaces over an algebraically closed field $\mathbb{F}$, e.g., the complex numbers.

Definition 3.1 *$\langle v_1, v_2, \ldots, v_n\rangle$ is an L-shifting sequence iff $v_1 \neq \mathbf{0}, L(v_1) = \mathbf{0}$ and $L(v_{k+1}) = v_k$ for each $k = 1, 2, \ldots, n-1$.*

Definition 3.2 *For $W \subseteq V \;\; L(W) = \{L(v) : v \in W\}$. It is the same as the range of L when $W = V$.*

Theorem 3.3 *An L-shifting sequence $\langle v_1, v_2, \ldots, v_n\rangle$, is linearly independent. Also if $W = span(\{v_1, v_2, \ldots, v_n\})$, then $L(W) \subseteq W$.*

Proof:

$$L(c_1v_1 + c_2v_2 + \cdots + c_nv_n) = c_2v_1 + c_3v_2 + \cdots + c_nv_{n-1}$$

So $L(W) \subseteq W$. Continuing to apply L we get

$$L^{n-1}(c_1v_1 + c_2v_2 + \cdots + c_nv_n) = c_nv_1.$$

So if $c_1v_1 + c_2v_2 + \cdots + c_nv_n = \mathbf{0}$, then $c_nv_1 = \mathbf{0}$ and so $c_n = 0$. Similarly $L^{n-2}(c_1v_1 + c_2v_2 + \cdots + c_{n-1}v_{n-1} = c_{n-1}v_1$ and so $c_{n-1} = 0$. Continuing this way we see that $c_i = 0$ for all i.
:foorP

Theorem 3.4 *Suppose $L : V \to V$ is a linear transformation and let $W_1 = \{v \in V : \exists n \; L^n(v) = \mathbf{0}\}$. Then $L(W_1) \subseteq W_1$ and there exists W_2 with $L(W_2) \subseteq W_2$ such that $V = W_1 \bigoplus W_2$.*

Proof:
There must be some n such that $W_1 = kernel(L^n)$, because no L-shifting sequence can be bigger than the dimension of V. Let $W_2 = range(L^n)$.

$L(W_1) \subseteq W_1$ since $u \in W_1$ implies $L^n(u) = \mathbf{0}$ implies $L^{n-1}(L(u)) = \mathbf{0}$ implies $L(u) \in W_1$.

$L(W_2) \subseteq W_2$ since $u \in W_2$ implies there exists v such that $L^n(v) = u$ so $L(u) = L^n(L(v))$ which implies $L(u) \in W_2$.

Since $dim(V) = dim(kernel(L^n)) + dim(range(L^n))$ it is enough to see that $W_1 \cap W_2 = \{\mathbf{0}\}$. So suppose $v \in W_1 \cap W_2$. Then for some $u \in V$ $L^n(u) = v$ and $L^n(v) = \mathbf{0}$. Hence $L^{n+n}(u) = \mathbf{0}$, and so $u \in W_1$ so $v = L^n(u) = \mathbf{0}$.
:foorP

Definition 3.5 *Define a linear transformation $L : V \to V$ to be nilpotent iff for every $v \in V$ there exists n such that $L^n(v) = \mathbf{0}$.*

Theorem 3.6 *Suppose $L : V \to V$ is a nilpotent linear transformation. Then there exists W_1, W_2 such that $L(W_1) \subseteq W_1$, $L(W_2) \subseteq W_2$, $V = W_1 \bigoplus W_2$ and W_1 has an L-shifting sequence for a basis.*

Proof:
Let n be the largest such that there exists an L-shifting sequence, say $\langle v_1, v_2, \ldots, v_n \rangle$. This means that for every $v \in V$ $L^n(v) = \mathbf{0}$. Let $W_1 = span\{v_1, v_2, \ldots, v_n\}$. We will first need to prove the following Claim:

Claim. Suppose $L(W) \subseteq W$ and $W_1 \cap W = \{\mathbf{0}\}$, then either $W_1 \bigoplus W = V$ or there exists $u \notin W_1 \bigoplus W$ such that $L(u) \in W$.

Proof of Claim: First note that there is a vector $v \notin W_1 \bigoplus W$ such that $L(v) \in W_1 \bigoplus W$. To see this, start with $v_0 \notin W_1 \bigoplus W$, and iteratively apply L to get $v_{k+1} = L(v_k)$. Since $v_n = L^n(v_0) = \mathbf{0} \in W_1 \bigoplus W$ there must be some k such that $v_k \notin W_1 \bigoplus W$, but $v_{k+1} \in W_1 \bigoplus W$.

So let $v \notin W_1 \bigoplus W$ and $L(v) \in W_1 \bigoplus W$ so

$$L(v) = c_1 v_1 + \ldots + c_n v_n + w$$

for some $w \in W$. It must be that $c_n = 0$, since applying the linear transformation L^{n-1} we get

$$L^{n-1}(L(v)) = L^{n-1}(c_1 v_1 + \ldots + c_n v_n) + L^{n-1}(w) = c_n v_1 + L^{n-1}(w)$$

and $L^{n-1}(L(v)) = L^n(v) = \mathbf{0}$ and $L^{n-1}(w) \in W$ implies $c_n v_1 \in W$ hence $c_n = 0$. Consequently

$$L(v) = c_1 v_1 + \cdots + c_{n-1} v_{n-1} + w.$$

Let

$$u = v - (c_1 v_2 + \cdots + c_{n-1} v_n).$$

Then $u - v \in W_1 \bigoplus W$, and since $v \notin W_1 \bigoplus W$ we have $u \notin W_1 \bigoplus W$. But $L(u) = w \in W_2$ as was required to prove the Lemma.

This proves the Claim.

To prove theorem use the same sort of argument as Theorem 2.6 using the lemma at each induction step.
:foorP

Definition 3.7 *The shift matrix S is the square matrix such that entry$_{i,i+1}S = 1$ for each i and all other entries of S are* 0.

Definition 3.8 *Matrices of the form $J = \lambda I + S$ are called Jordan block matrices.*

Example:

$$J = \begin{bmatrix} \lambda & 1 & 0 & 0 \\ 0 & \lambda & 1 & 0 \\ 0 & 0 & \lambda & 1 \\ 0 & 0 & 0 & \lambda \end{bmatrix} \quad \lambda I = \begin{bmatrix} \lambda & 0 & 0 & 0 \\ 0 & \lambda & 0 & 0 \\ 0 & 0 & \lambda & 0 \\ 0 & 0 & 0 & \lambda \end{bmatrix} \quad S = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Note that $\lambda \in \mathbb{F}$ can be anything including 0 and J can be 1×1. Now suppose that $A \in \mathbb{F}^{4\times 4}$ and v_1, v_2, v_3, v_4 is a basis for $\mathbb{F}^{4\times 1}$ such that

$$Av_1 = \lambda v_1,\ Av_2 = \lambda v_2 + v_1,\ Av_3 = \lambda v_3 + v_2,\ Av_4 = \lambda v_4 + v_3.$$

Then by Theorem 1.6 A is similar to J. Note also that v_1, v_2, v_3, v_4 is a B-shifting sequence for $B = A - \lambda I$, that is $Bv_1 = \mathbf{0}$,$Bv_2 = v_1$, $Bv_3 = v_2$, and $Bv_4 = v_3$.

Theorem 3.9 *(Jordan normal form) Every square matrix A is similar to a matrix in the block diagonal form:*

$$\begin{bmatrix} J_1 & 0 & \dots & 0 \\ 0 & J_2 & \dots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & J_n \end{bmatrix}$$

where each J_i is a Jordan block matrix.

Proof:
A basis is called a Jordan basis for L iff it can be broken up into blocks $B_1, B_2, \dots, B_n$ such that for each i there exists λ_i such that B_i is an $L - \lambda_i I$-shifting sequence. To prove our result it suffices to show every L has a Jordan basis.

Say that $L : V \to V$ is decomposable iff it is possible to find $W_1 \bigoplus W_2 = V$ (each W_i nontrivial) such that $L(W_1) \subseteq W_1$ and $L(W_2) \subseteq W_2$. Otherwise L is indecomposable.

Suppose for contradiction that $L : V \to V$ is a linear transformation which does not have a Jordan basis, and suppose that the dimension of V is as small as possible. We will derive a contradiction.

Note that L must be indecomposable, since Jordan bases for the restrictions to W_1 and W_2 would exist and give a Jordan basis for L.

Note also that for any $\lambda \in \mathbb{F}$ L has a Jordan basis iff $L - \lambda I$ has a Jordan basis. (Replace each λ_i by $\lambda_i - \lambda$, then $Lv_i = \lambda_i v_i + v_{i-1}$ iff $(L - \lambda I)v_i = (\lambda_i - \lambda)v_i + v_{i-1}$.) Also L is indecomposable iff $L - \lambda I$ is indecomposable. This means that without loss we may assume that $\lambda = 0$ is an eigenvalue of L, since if λ is any eigenvalue of L then 0 is an eigenvalue of $L - \lambda I$. Theorem 3.4 implies that L must be nilpotent. Finally Theorem 3.6 implies that L has basis which corresponds to a shift matrix. This is a contradiction which proves the theorem.
:foorP