SOLUTIONS TO ARTIN'S *ALGEBRA*, 2ND ED. CH. 4 – LINEAR OPERATORS

COLIN COMMANS

Contents

1.1

Let A be a $\ell \times m$ matrix and let B be an $n \times p$ matrix. Prove that the rule $M \rightsquigarrow AMB$ defines a linear transformation from the space $F^{m\times n}$ of $m \times n$ matrices to the space $F^{\ell\times p}$.

Solution.

Proof.
Let $\psi(M) = AMB$ and choose $M_1, M_2 \in F^{m\times n}$. Then

$$\psi(M_1 + M_2) = A(M_1 + M_2)B = [AM_1 + AM_2]B = AM_1B + AM_2B = \psi(M_1) + \psi(M_2)$$

and for any $c \in F$

$$\psi(cM_1) = A(cM_1)B = c(AM_1B) = c\psi(M)$$

Thus ψ is a linear transformation. □

1.2

Let $v_1, \dots, v_n$ be elements of a vector space V. Prove that the map $\varphi : F^n \to V$ defined by $\varphi(X) = v_1x_1 + \cdots + v_nx_n$ is a linear transformation.

Solution.

Proof.
Note that for any $X, Y \in F^n$

$$\varphi(X+Y) = \sum_{i=1}^{n} v_i[X+Y]_i = \sum_{i=1}^{n} v_i(x_i + y_i) = \sum_{i=1}^{n} v_ix_i + \sum_{i=1}^{n} v_iy_i = \varphi(X) + \varphi(Y)$$

and for any $c \in F$

$$\varphi(cX) = \sum_{i=1}^{n} v_i[cX]_i = \sum_{i=1}^{n} v_i(cx_i) = c\sum_{i=1}^{n} v_ix_i = c\varphi(X)$$

Thus φ is a linear transformation. □

1.3

Let A be an $m \times n$ matrix. Use the dimension formula to prove that the space of solutions of the linear system $AX = 0$ has dimension at least $n - m$.

Solution.

Proof.
Since the left-multiply map $T_A : F^n \to F^m, \quad X \mapsto AX$ is linear, by the dimension formula we have that

$$\dim(\ker T_A) + \dim(\operatorname{im} T_A) = \dim F^n = n \implies \dim(\ker T_A) = n - \dim(\operatorname{im} T_A)$$

and since $\operatorname{im} T_A \subset F^m$, we have $\dim(\operatorname{im} T_A) \leq \dim(F^m) = m$. Thus the dimension of the space $\{X \mid AX = 0\}$, which is exactly $\dim(\ker T_A)$, is

$$\dim(\ker T_A) = n - \dim(\operatorname{im} T_a) \geq n - m$$

□

1.4

Prove that every $m \times n$ matrix A of rank 1 has the form $A = XY^t$, where X, Y are m- and n-dimensional column vectors. How uniquely determined are these vectors?

Solution.

Proof.
Since A has rank 1, that means the span of the columns is the span of a single vector, say $X \in F^m$. And since each column $a_i \in F^m$ of A is in the column space, there exists $y_1, \ldots, y_n \in F$ such that

$$A = [a_1 \ \ldots \ a_n] = [y_1 X \ \ldots y_n X] = X[y_1 \ \ldots \ y_n] =: XY^t$$

Note that the values $y_1, \ldots, y_n$ are uniquely determined by X (by being a basis), but X is not necessarily unique. Indeed, taking any other nonzero vector in span $\{X\}$, i.e. any cX for nonzero $c \in F$, note that

$$X'Y'^t := (cX)(c^{-1}Y)^t = (cc^{-1})XY^t = XY^t = A$$

so the decomposition is unique up to nonzero scalar multiples. □

1.5

(a) Let U and W be vector spaces over a field F. Show that the two operations $(u, w) + (u', w') = (u + u', w + w')$ and $c(u, w) = (cu, cw)$ on pairs of vectors make the product set $U \times W$ into a vector space. It is called the *product space.*

(b) Let U and W be subspaces of a vector space V. Show that the map $T : U \times W \to V$ defined by $T(u, w) = u + w$ is a linear transformation.

(c) Express the dimension formula for T in terms of the dimensions of subspaces of V.

Solution.

(a) *Proof.*
Clearly $U \times W$ under addition forms an abelian group with identity $(0, 0)$ since U and W are themselves vector spaces. Next note that

$$1(u, w) = (1u, 1w) = (u, w)$$

and

$$a[b(u, w)] = a(bu, bw) = (a[bu], a[bw]) = ([ab]u, [ab]w) = [ab](u, w)$$

Finally, for distributivity we have

$$[a+b](u, w) = ([a+b]u, [a+b]w) = (au+bu, aw+bw) = (au, aw)+(bu, bw) = a(u, w)+b(u, w)$$

and

$$\begin{aligned} a[(u, w) + (u', w')] = a(u + u', w + w') &= (a[u + u'], a[w + w']) \\ &= (au + au', aw + aw') \\ &= (au, aw) + (au', aw') \\ &= a(u, w) + a(u', w') \end{aligned}$$

Thus $U \times W$ with these operations satisfy the vector space axioms. □

(b) *Proof.*
For any $(u_1, w_1), (u_2, w_2) \in U \times W$ we have

$$\begin{aligned} T((u_1, w_1) + (u_2, w_2)) &= T(u_1 + u_2, w_1 + w_2) \\ &= [u_1 + u_2] + [w_1 + w_2] \\ &= [u_1 + w_1] + [u_2 + w_2] \\ &= T(u_1, w_1) + T(u_2, w_2) \end{aligned}$$

and for any $c \in F$

$$T(c(u_1, w_1)) = T(cu_1, cw_1) = cu_1 + cw_1 = c(u_1 + w_1) = cT(u_1, w_1)$$

Therefore T is linear. □

(c) By the dimension formula we have

$$\dim(\ker T) + \dim(\operatorname{im} T) = \dim(U \times W)$$

First note that $\operatorname{im} T = \{u + w \mid u \in U, w \in W\} = U + W$ (see Section 3.6). Furthermore, if $a_1, \ldots, a_k$ is a basis for U and $b_1, \ldots, b_\ell$ is a basis for W, then for any $(u, w) \in U \times W$ we have the unique decomposition

$$(u, w) = \left(\sum_{i=1}^{k} u_i a_i, \sum_{i=1}^{\ell} w_i b_i\right) = \sum_{i=1}^{k} u_i (a_i, 0) + \sum_{i=1}^{\ell} w_i (0, b_i)$$

Hence $\{(a_1, 0), \ldots, (a_k, 0), (0, b_1), \ldots, (0, b_\ell)\}$ is a basis for $U \times W$ and so we have that $\dim(U \times W) = \dim U + \dim W$. Finally, note that

$$(u, w) \in \ker T \implies u + w = 0 \implies u = -w \implies u \in W \implies u \in U \cap W$$

Thus we can define a map $f : \ker T \to U \cap W$ by $f(u, w) = u$. This is a linear map as

$$f((u_1, w_1) + (u_2, w_2)) = f(u_1 + u_2, w_1 + w_2) = u_1 + u_2 = f(u_1, w_1) + f(u_2, w_2)$$

and

$$f(c(u, w)) = f(cu, cw) = cu = cf(u, w)$$

It is also injective since

$$f(u, w) = 0 \implies u = 0 \implies 0 = T(u, w) = u + w = 0 + w = w \implies (u, w) = (0, 0)$$

and it is surjective since for any $v \in U \cap W$, we have $(v, -v) \in \ker T$ and $f(v, -v) = v$. Thus f is an isomorphism, $\ker T$ is isomorphic to $U \cap W$, and in particular $\dim(\ker T) = \dim(U \cap W)$. Therefore the dimension formula says

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

[NB: This is exactly Prop 3.6.6(a).]

2.1

Let A and B be 2×2 matrices. Determine the matrix of the operator $T : M \rightsquigarrow AMB$ on the space $F^{2\times 2}$ of 2×2 matrices, with respect to the basis $(e_{11}, e_{12}, e_{21}, e_{22})$ of $F^{2\times 2}$.

Solution.
We can compute the matrix product by hand:

$$
\begin{aligned}
AMB &= \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} \\
&= \begin{bmatrix} a_{11}m_{11} + a_{12}m_{21} & a_{11}m_{12} + a_{12}m_{22} \\ a_{21}m_{11} + a_{22}m_{21} & a_{21}m_{12} + a_{22}m_{22} \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} \\
&= \begin{bmatrix} b_{11}(a_{11}m_{11} + a_{12}m_{21}) + b_{21}(a_{11}m_{12} + a_{12}m_{22}) & b_{12}(a_{11}m_{11} + a_{12}m_{21}) + b_{22}(a_{11}m_{12} + a_{12}m_{22}) \\ b_{11}(a_{21}m_{11} + a_{22}m_{21}) + b_{21}(a_{21}m_{12} + a_{22}m_{22}) & b_{12}(a_{21}m_{11} + a_{22}m_{21}) + b_{22}(a_{21}m_{12} + a_{22}m_{22}) \end{bmatrix} \\
&= \begin{bmatrix} (b_{11}a_{11})m_{11} + (b_{11}a_{12})m_{21} + (b_{21}a_{11})m_{12} + (b_{21}a_{12})m_{22} & (b_{12}a_{11})m_{11} + (b_{12}a_{12})m_{21} + (b_{22}a_{11})m_{12} + (b_{22}a_{12})m_{22} \\ (b_{11}a_{21})m_{11} + (b_{11}a_{22})m_{21} + (b_{21}a_{21})m_{12} + (b_{21}a_{22})m_{22} & (b_{12}a_{21})m_{11} + (b_{12}a_{22})m_{21} + (b_{22}a_{21})m_{12} + (b_{22}a_{22})m_{22} \end{bmatrix} \\
&= m_{11} \begin{bmatrix} b_{11}a_{11} & b_{12}a_{11} \\ b_{11}a_{21} & b_{12}a_{21} \end{bmatrix} + m_{21} \begin{bmatrix} b_{11}a_{12} & b_{12}a_{12} \\ b_{11}a_{22} & b_{12}a_{22} \end{bmatrix} + m_{12} \begin{bmatrix} b_{21}a_{11} & b_{22}a_{11} \\ b_{21}a_{21} & b_{22}a_{21} \end{bmatrix} + m_{22} \begin{bmatrix} b_{21}a_{12} & b_{22}a_{12} \\ b_{21}a_{22} & b_{22}a_{22} \end{bmatrix}
\end{aligned}
$$

Hence we can write in terms of our basis $(e_{11}, e_{12}, e_{21}, e_{22})$

$$
T(e_{ij}) = \begin{bmatrix} b_{j1}a_{1i} & b_{j2}a_{1i} \\ b_{j1}a_{2i} & b_{j2}a_{2i} \end{bmatrix} = e_{11}(b_{j1}a_{1i}) + e_{12}(b_{j2}a_{1i}) + e_{21}(b_{j1}a_{2i}) + e_{22}(b_{j2}a_{2i}) \leftrightsquigarrow \begin{bmatrix} b_{j1}a_{1i} \\ b_{j2}a_{1i} \\ b_{j1}a_{2i} \\ b_{j2}a_{2i} \end{bmatrix}
$$

Putting all these column vectors together then gives the matrix representation

$$
T \leftrightsquigarrow \begin{bmatrix} b_{11}a_{11} & b_{21}a_{11} & b_{11}a_{12} & b_{21}a_{12} \\ b_{12}a_{11} & b_{22}a_{11} & b_{12}a_{12} & b_{22}a_{12} \\ b_{11}a_{21} & b_{21}a_{21} & b_{11}a_{22} & b_{21}a_{22} \\ b_{12}a_{21} & b_{22}a_{21} & b_{12}a_{22} & b_{22}a_{22} \end{bmatrix} = \begin{bmatrix} a_{11}B^t & a_{12}B^t \\ a_{21}B^t & a_{22}B^t \end{bmatrix}
$$

2.2

Let A be an $n \times n$ matrix, and let V denote the space of n-dimensional *row* vectors. What is the matrix of the linear operator "right multiplication by A" with respect to the standard basis of V?

Solution.
If we write the standard basis of V as $e_1^t, \dots, e_n^t$, then note that

$$e_i^t A = (A^t e_i)^t$$

and since $A^t e_i$ is the ith column of A^t, this means that $e_i^t A$ is the ith row of A and so

$$e_i^t A = [a_{i1} \ \dots \ a_{in}] = a_{i1} e_1^t + \cdots + a_{in} e_n^t \leftrightsquigarrow \begin{bmatrix} a_{i1} \\ \vdots \\ a_{in} \end{bmatrix}$$

with respect to the standard basis. Stacking these as columns gives the matrix representation of $T : X \mapsto XA$ to be

$$T \leftrightsquigarrow \begin{bmatrix} a_{11} & \dots & a_{n1} \\ \vdots & \ddots & \vdots \\ a_{1n} & \dots & a_{nn} \end{bmatrix} = A^t$$

2.3

Find all real 2×2 matrices that carry the line $y = x$ to the line $y = 3x$.

Solution.
We first need to write the two lines as subspaces of $\mathbb{R}^2$. Indeed,

$$y = x \implies \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x \\ x \end{bmatrix} = x \begin{bmatrix} 1 \\ 1 \end{bmatrix} \implies \{(x, y) \in \mathbb{R}^2 \mid y = x\} = \operatorname{span}\left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\} =: A$$

and

$$y = 3x \implies \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x \\ 3x \end{bmatrix} = x \begin{bmatrix} 1 \\ 3 \end{bmatrix} \implies \{(x, y) \in \mathbb{R}^2 \mid y = 3x\} = \operatorname{span}\left\{ \begin{bmatrix} 1 \\ 3 \end{bmatrix} \right\} =: B$$

Thus we need to find all linear transformations $T : \mathbb{R}^2 \to \mathbb{R}^2$ such that $\operatorname{im}(T|_A) \subset B$. Note that any such map forces $T([1 \ \ 1]^t)$ to be an element of B. Hence if we let M_T be the matrix representation of T with respect to the standard basis, we have

$$M_T \begin{bmatrix} 1 \\ 1 \end{bmatrix} := \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} a + b \\ c + d \end{bmatrix} \in B \implies c + d = 3(a + b)$$

Therefore the set of all possible matrices M_T is

$$\left\{ \begin{bmatrix} a & b \\ c & d \end{bmatrix} \middle| \ c + d = 3(a + b) \right\}$$

2.4

Prove Theorem 4.2.10(b) using row and column operations.

Solution.

Proof.
Let A be an $m \times n$ matrix. We first row reduce A to A''' via elementary row operations:

$$A''' = E_k \dots E_1 A$$

Now the number of pivot 1s is exactly the rank of A, so we can permute the columns of A''' via column operations to move all the pivots to the left and create A'' in the form

$$A'' = A''' F_1 \dots F_m = \left[\begin{array}{c|c} I_r & B \\ \hline 0 & 0 \end{array}\right]$$

Finally, we can use more column operations on A'' to clear out the entries of the submatrix B (e.g. to clear out the entry b_{ij}, add $-b_{ij}$ times column i of A'' and add it to column $r+j$ of A''). Doing these operations will clear out B and create A':

$$A' = A'' G_1 \dots G_n = \left[\begin{array}{c|c} I_r & 0 \\ \hline 0 & 0 \end{array}\right]$$

Finally, set $Q = E_1^{-1} \dots E_k^{-1}$ and $P = F_1 \dots F_m G_1 \dots G_n$ and note

$$\left[\begin{array}{c|c} I_r & 0 \\ \hline 0 & 0 \end{array}\right] = A' = A'' G_1 \dots G_n = A''' F_1 \dots F_m G_1 \dots G_n = E_k \dots E_1 A F_1 \dots F_m G_1 \dots G_n = Q^{-1} A P$$

which is our desired decomposition. □

[NB: This exercise more precisely answers the question in Exercise 1.3.4]

2.5

Let A be an $m \times n$ matrix of rank r, let I be a set of r row indices such that the corresponding rows of A are independent, and let J be a set of r column indices such that the corresponding columns of A are independent. Let M denote the $r \times r$ submatrix of A obtained by taking rows from I and columns from J. Prove that M is invertible.

Solution.

Proof.
Suppose that M is not invertible. In particular (e.g. by Thm 1.2.21), there exists a nonzero r-dimensional vector X such that $MX = 0$. We now construct an n-dimension vector Y via:

$$Y_j := \begin{cases} X_j & \text{if } j \in J \\ 0 & \text{if } j \notin J \end{cases}$$

Now note that if we choose an index $i \in I$, then

$$[AY]_i = \sum_{j=1}^{n} a_{ij} Y_j = 0 + \sum_{j \in J} a_{ij} X_j = \sum_{j \in J} m_{ij} X_j = [MX]_i = 0 \qquad (\star)$$

However if we write A row-wise, then

$$AY = \begin{bmatrix} -A_1- \\ \vdots \\ -A_m- \end{bmatrix} Y = \begin{bmatrix} A_1 Y \\ \vdots \\ A_m Y \end{bmatrix} \implies [AY]_i = A_i Y$$

Hence $0 = [AY]_i = A_i Y$ for all $i \in I$, and since $\{A_i \mid i \in I\}$ span the row-space (i.e. each A_k is a linear combination of the $A_i, i \in I$), we have that $A_i Y = 0$ for $i = 1, \ldots, m$ and thus $AY = 0$. Now writing A column-wise,

$$\begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix} = AY = [Ae_1 \ \ldots \ Ae_n] \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix} = \sum_{j=1}^{n} (Ae_j) Y_j = 0 + \sum_{j \in J} (Ae_j) X_j$$

And since $\sum_{j \in J} (Ae_j) X_j$ is a linear combination of r linearly-independent columns of A, this forces $X_j = 0$ for $j = 1, \ldots, r$ which contradicts X being nonzero. Therefore M is invertible.

[NB: Strictly speaking, $(\star)$ is false since the entry a_{ij} is not necessarily the entry m_{ij} for $i \in I, j \in J$. A simple fix is given $i \in I, j \in J$, set $\bar{i}, \bar{j} \in \{1, \ldots, r\}$ such that $m_{\bar{i}\bar{j}} = a_{ij}$. Then $(\star)$ can be replaced with the corrected (albeit slightly more clunky)

$$[AY]_i = \sum_{j=1}^{n} a_{ij} Y_j = 0 + \sum_{j \in J} a_{ij} X_{\bar{j}} = \sum_{j \in J} m_{\bar{i}\bar{j}} X_{\bar{j}} = [MX]_{\bar{i}} = 0$$

An even simpler fix is to just set $m_{ij} := a_{ij}$ and only allowing access to an entry of X and M via I and J, as this also fixes the same nitpick in our construction of Y.]

□

3.1

Determine the dimensions of the kernel and the image of the linear operator T on the space $\mathbb{R}^n$ defined by $T(x_1, \dots, x_n)^t = (x_1 + x_n, x_2 + x_{n-1}, \dots, x_n + x_1)^t$.

Solution.
Note that by the dimension formula we have

$$\dim(\ker T) + \dim(\operatorname{im} T) = \dim(\mathbb{R}^n) = n \implies \dim(\operatorname{im} T) = n - \dim(\ker T)$$

So it suffices to calculate $\dim(\ker T)$. We have

$$\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \in \ker T \iff 0 = T\left(\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}\right) = \begin{bmatrix} x_1 + x_n \\ \vdots \\ x_n + x_1 \end{bmatrix}$$

If n is even,

$$X \in \ker T \iff \begin{cases} x_n = -x_1 \\ x_{n-1} = -x_2 \\ \dots \\ x_{n/2+1} = -x_{n/2} \end{cases} \iff X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_{n/2} \\ -x_{n/2} \\ \vdots \\ -x_2 \\ -x_1 \end{bmatrix} = x_1 \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \\ 0 \\ \vdots \\ 0 \\ -1 \end{bmatrix} + x_2 \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \\ 0 \\ \vdots \\ -1 \\ 0 \end{bmatrix} + \dots + x_{n/2} \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \\ -1 \\ \vdots \\ 0 \\ 0 \end{bmatrix}$$

Thus

$$\ker T = \operatorname{span}\left\{ \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \\ 0 \\ \vdots \\ 0 \\ -1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \\ 0 \\ \vdots \\ -1 \\ 0 \end{bmatrix}, \dots, \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \\ -1 \\ \vdots \\ 0 \\ 0 \end{bmatrix} \right\}$$

and since these vectors are independent, they form a basis of $\ker T$ and so $\dim(\ker T) = n/2$ when n is even.

Now if n is odd,

$$X \in \ker T \iff \begin{cases} x_n = -x_1 \\ x_{n-1} = -x_2 \dots \\ x_{(n+1)/2+1} = -x_{(n-1)/2} \\ x_{(n+1)/2} = 0 \end{cases}$$

$$\iff X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_{(n-1)/2} \\ 0 \\ -x_{(n-1)/2} \\ \vdots \\ -x_2 \\ -x_1 \end{bmatrix} = x_1 \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \\ 0 \\ 0 \\ \vdots \\ 0 \\ -1 \end{bmatrix} + x_2 \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \\ 0 \\ 0 \\ \vdots \\ -1 \\ 0 \end{bmatrix} + \cdots + x_{(n-1)/2} \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \\ 0 \\ -1 \\ \vdots \\ 0 \\ 0 \end{bmatrix}$$

Thus

$$\ker T = \operatorname{span} \left\{ \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \\ 0 \\ 0 \\ \vdots \\ 0 \\ -1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \\ 0 \\ 0 \\ \vdots \\ -1 \\ 0 \end{bmatrix}, \dots, \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \\ 0 \\ -1 \\ \vdots \\ 0 \\ 0 \end{bmatrix} \right\}$$

and since these vectors are independent, they form a basis of $\ker T$ and so $\dim(\ker T) = (n-1)/2$ when n is odd. In summary,

$$\dim(\ker T) = \begin{cases} n/2 & \text{even } n \\ (n-1)/2 & \text{odd } n \end{cases} = \lfloor n/2 \rfloor \implies \dim(\operatorname{im} T) = n - \lfloor n/2 \rfloor = \begin{cases} n/2 & \text{even } n \\ (n+1)/2 & \text{odd } n \end{cases}$$

3.2

(a) Let $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ be a real matrix, with c not zero. Show that using conjugation by elementary matrices, one can eliminate the "a" entry.

(b) Which matrices with $c = 0$ are similar to a matrix in which the "a" entry is zero?

Solution.

(a) *Proof.*
We want to find an invertible matrix P such that

$$P^{-1}AP = \begin{bmatrix} 0 & b' \\ c' & d' \end{bmatrix}$$

One idea is to try to eliminate a via an operation like adding $-\frac{a}{c}$ times the second row to the first row. In this case, we want P^{-1} to be of the form

$$P^{-1} = \begin{bmatrix} 1 & k \\ 0 & 1 \end{bmatrix} \implies P = \begin{bmatrix} 1 & -k \\ 0 & 1 \end{bmatrix}$$

Then

$$\begin{aligned} P^{-1}AP = \begin{bmatrix} 1 & k \\ 0 & 1 \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} 1 & -k \\ 0 & 1 \end{bmatrix} &= \begin{bmatrix} a+kc & b+kd \\ c & d \end{bmatrix} \begin{bmatrix} 1 & -k \\ 0 & 1 \end{bmatrix} \\ &= \begin{bmatrix} a+kc & -ka-k^2c+b+kd \\ c & -kc+d \end{bmatrix} \end{aligned}$$

and indeed setting $k = -\frac{a}{c}$ (as $c \neq 0$) will clear out the "a" entry. □

(b) We are now given the case $A = \begin{bmatrix} a & b \\ 0 & d \end{bmatrix}$ is upper triangular. Note that if we want to eliminate a via conjugation by elementary matrices, swapping and multiplying by scalars will not work (swapping will simply transpose and multiplying will scale only the off-diagonal, neither of which affects a). Also from (a) any P of the form $P = \begin{bmatrix} 1 & -k \\ 0 & 1 \end{bmatrix}$ will give

$$P^{-1}AP = \begin{bmatrix} a+kc & -ka-k^2c+b+kd \\ c & -kc+d \end{bmatrix} = \begin{bmatrix} a & -ka+b+kd \\ 0 & d \end{bmatrix}$$

which fails as well. Hence our last option is an elementary matrix of the form

$$P^{-1} = \begin{bmatrix} 1 & 0 \\ k & 1 \end{bmatrix} \implies P = \begin{bmatrix} 1 & 0 \\ -k & 1 \end{bmatrix}$$

and now

$$P^{-1}AP = \begin{bmatrix} 1 & 0 \\ k & 1 \end{bmatrix} \begin{bmatrix} a & b \\ 0 & d \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -k & 1 \end{bmatrix} = \begin{bmatrix} a & b \\ ka & kb+d \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -k & 1 \end{bmatrix} = \begin{bmatrix} a-kb & b \\ ka & kb+d \end{bmatrix}$$

Hence we can clear out a if and only if $b \neq 0$ (by setting $k = \frac{a}{b}$) or $a = 0$ originally (in which case just take $P = I$).

3.3

Let $T : V \to V$ be a linear operator on a vector space of dimension 2. Assume that T is not multiplication by a scalar. Prove that there is a vector v in V such that $(v, T(v))$ is a basis of V, and describe the matrix T with respect to that basis.

Solution.

Proof.
Suppose otherwise, i.e. for every vector $v \in V$, $T(v)$ and v are not independent. This means for every $v \in V$, there exists a $\lambda \in F$ (assuming V is a vector space over F) such that $T(v) = \lambda v$. In particular, we have $\lambda_1, \lambda_2, \mu \in F$ such that

$$T\left(\begin{bmatrix}1\\0\end{bmatrix}\right) = \begin{bmatrix}\lambda_1\\0\end{bmatrix}, \qquad T\left(\begin{bmatrix}0\\1\end{bmatrix}\right) = \begin{bmatrix}0\\\lambda_2\end{bmatrix}, \qquad T\left(\begin{bmatrix}1\\1\end{bmatrix}\right) = \begin{bmatrix}\mu\\\mu\end{bmatrix}$$

However by linearity we have

$$\begin{bmatrix}\mu\\\mu\end{bmatrix} = T\left(\begin{bmatrix}1\\1\end{bmatrix}\right) = T\left(\begin{bmatrix}1\\0\end{bmatrix}\right) + T\left(\begin{bmatrix}0\\1\end{bmatrix}\right) = \begin{bmatrix}\lambda_1\\0\end{bmatrix} + \begin{bmatrix}0\\\lambda_2\end{bmatrix} = \begin{bmatrix}\lambda_1\\\lambda_2\end{bmatrix} \implies \lambda_1 = \lambda_2 = \mu$$

But now for any $v = [v_1 \;\; v_2]^t \in V$, we have

$$T\left(\begin{bmatrix}v_1\\v_2\end{bmatrix}\right) = v_1 T\left(\begin{bmatrix}1\\0\end{bmatrix}\right) + v_2 T\left(\begin{bmatrix}0\\1\end{bmatrix}\right) = v_1\begin{bmatrix}\mu\\0\end{bmatrix} + v_2\begin{bmatrix}0\\\mu\end{bmatrix} = \begin{bmatrix}\mu v_1\\\mu v_2\end{bmatrix} = \mu\begin{bmatrix}v_1\\v_2\end{bmatrix}$$

and so T is multiplication by a scalar μ, which is a contradiction. □

Now let $v \in V$ be a vector such that $(v, T(v))$ is a basis of V. The matrix representation of T with respect to this basis can be found by computing

$$T(v) = 0v + 1T(v) \qquad \text{and} \qquad T(T(v)) = \alpha v + \beta T(v)$$

Hence the matrix representation with respect to this basis is $\begin{bmatrix}0 & \alpha\\1 & \beta\end{bmatrix}$.

3.4

Let B be a complex $n \times n$ matrix. Prove or disprove: The linear operator T on the space of all $n \times n$ matrices defined by $T(A) = AB - BA$ is singular.

Solution.
T is always singular, as $T(I) = B - B = 0$ and so $\ker T \neq \{0\}$.

4.1

Let T be a linear operator on a vector space V, and let λ be a scalar. The *eigenspace* $V^{(\lambda)}$ is the set of eigenvectors of T with eigenvalue λ, together with 0. Prove that $V^{(\lambda)}$ is a T-invariant subspace.

Solution.

Proof.
First note that $V^{(\lambda)} = \ker(T - \lambda I)$ and so it is a subspace. Now choose $v \in V^{(\lambda)}$. Then by construction $T(v) = \lambda v$. However, note that

$$T(\lambda v) = \lambda T(v) = \lambda(\lambda v) \implies T(v) \in V^{(\lambda)}$$

Thus $V^{(\lambda)}$ is a T-invariant subspace. □

4.2

(a) Let T be a linear operator on a finite-dimensional vector space V, such that T^2 is the identity operator. Prove that for any vector v in V, $v - Tv$ is either an eigenvector with eigenvalue -1, or the zero vector. With notation as in Exercise 4.1, prove that V is the direct sum of the eigenspaces $V^{(1)}$ and $V^{(-1)}$.

(b) Generalize this method to prove that a linear operator T such that $T^4 = I$ decomposes a complex vector space into a sum of four eigenspaces.

Solution.

(a) First we prove that $v - T(v) \in V^{(-1)}$.

Proof.
If $v - T(v)$ is not zero, then

$$T(v - T(v)) = T(v) - T^2(v) = T(v) - v = -1(v - T(v))$$

so it is an eigenvector with eigenvalue -1. □

Next we prove that $V = V^{(1)} \oplus V^{(-1)}$.

Proof.
First note that

$$v \in V^{(1)} \cap V^{(-1)} \iff v = T(v) = -v \iff v = 0$$

Thus $V^{(1)} \cap V^{(-1)} = \{0\}$ so it suffices to show that $V = V^{(1)} + V^{(-1)}$. Indeed, for any $v \in V$, since $v + T(v) \in V^{(1)}$ by $T(v + T(v)) = T(v) + T^2(v) = T(v) + v$, then

$$T(v) \in V^{(1)},\ v - T(v) \in V^{(-1)} \implies v = \tfrac{1}{2}(v - T(v)) + \tfrac{1}{2}(v + T(v)) \in V^{(1)} + V^{(-1)}$$

Therefore $V = V^{(1)} \oplus V^{(-1)}$ by Prop 3.6.6(c). □

(b) *Proof.*
Note that for any $v \in V$,

$$\begin{aligned}
T(v + T(v) + T^2(v) + T^3(v)) = T(v) + T^2(v) + T^3(v) + v \Rightarrow v + T(v) + T^2(v) + T^3(v) \in V^{(1)} \\
T(v - T(v) + T^2(v) - T^3(v)) = T(v) - T^2(v) + T^3(v) - v \Rightarrow v - T(v) + T^2(v) - T^3(v) \in V^{(-1)} \\
T(v - iT(v) - T^2(v) + iT^3(v)) = T(v) - iT^2(v) - T^3(v) + iv \Rightarrow v - iT(v) - T^2(v) + iT^3(v) \in V^{(i)} \\
T(v + iT(v) - T^2(v) - iT^3(v)) = T(v) + iT^2(v) - T^3(v) - iv \Rightarrow v + iT(v) - T^2(v) - iT^3(v) \in V^{(-i)}
\end{aligned}$$

and that

$$\begin{aligned}
&\tfrac{1}{4}[v + T(v) + T^2(v) + T^3(v)] \\
&\quad + \tfrac{1}{4}[v - T(v) + T^2(v) - T^3(v)] \\
&\quad + \tfrac{1}{4}[v - iT(v) - T^2(v) + iT^3(v)] \\
&\quad + \tfrac{1}{4}[v + iT(v) - T^2(v) - iT^3(v)] \\
= &\tfrac{1}{4}[v + v + v + v] \\
&\quad + \tfrac{1}{4}[T(v) - T(v) - iT(v) + iT(v)] \\
&\quad + \tfrac{1}{4}[T^2(v) + T^(v) - T^2(v) - T^2(v)] \\
&\quad + \tfrac{1}{4}[T^3(v) - T^3(v) + iT^3(v) - iT^3(v)] \\
= &v
\end{aligned}$$

Thus $V = V^{(1)} + V^{(-1)} + V^{(i)} + V^{(-i)}$. By Exercise 3.5.3, it suffices to show that

$$V^{(1)} \cap V^{(-1)} = \{0\},\ (V^{(1)} + V^{(-1)}) \cap V^{(i)} = \{0\},\ \text{and } (V^{(1)} + V^{(-1)} + V^{(i)}) \cap V^{(-i)} = \{0\}$$

Indeed,

$$v \in V^{(1)} \cap V^{(-1)} \iff v = T(v) = -v \implies 2v = 0 \implies v = 0$$

So $V^{(1)}, V^{(-1)}$ are independent.
If $v \in (V^{(1)} + V^{(-1)}) \cap V^{(i)}$, then there exists $v_1 \in V^{(1)}, v_2 \in V^{(-1)}$ such that $v = v_1 + v_2$ and so

$$iv = T(v) = T(v_1 + v_2) = v_1 - v_2 \implies i(v_1 + v_2) - (v_1 - v_2) = 0 \implies (i-1)v_1 + (i+1)v_2 = 0$$

which by independence of $V^{(1)}$ and $V^{(-1)}$ forces $v_1 = v_2 = 0$, and hence $v = 0$ and $V^{(1)}, V^{(-1)}, V^{(i)}$ are independent.
Similarly, if $v \in (V^{(1)} + V^{(-1)} + V^{(i)}) \cap V^{(-i)}$, then there exists $v_1 \in V^{(1)}, v_2 \in V^{(-1)}, v_3 \in V^{(i)}$ such that $v = v_1 + v_2 + v_3$ and so

$$\begin{aligned}
-iv = T(v) = T(v_1 + v_2 + v_3) = v_1 - v_2 + iv_3 &\implies -i(v_1 + v_2 + v_3) = v_1 - v_2 + iv_3 \\
&\implies (-i-1)v_1 + (-i+1)v_2 + (-2i)v_3 = 0
\end{aligned}$$

which by independence of $V^{(1)}, V^{(-1)}, V^{(i)}$ forces $v_1 = v_2 = v_3 = 0$, and hence $v = 0$.
Therefore by Exercise 3.5.3 $V = V^{(1)} \oplus V^{(-1)} \oplus V^{(i)} \oplus V^{(-i)}$. □

4.3

Let T be a linear operator on a vector space V. Prove that if W_1 and W_2 are T-invariant subspaces of V, then $W_1 + W_2$ and $W_1 \cap W_2$ are T-invariant.

Solution.

Proof.
(i) Choose $w_1 + w_2 \in W_1 + W_2$. Then $T(w_1) \in W_1$ and $T(w_2) \in W_2$, thus

$$T(w_1 + w_2) = T(w_1) + T(w_2) \in W_1 + W_2$$

Therefore $W_1 + W_2$ is T-invariant.
(ii) Choose $w \in W_1 \cap W_2$. Then

$$\begin{cases} w \in W_1 \implies T(w) \in W_1 \\ w \in W_2 \implies T(w) \in W_2 \end{cases}$$

Thus $T(w) \in W_1 \cap W_2$ and therefore $W_1 \cap W_2$ is T-invariant. □

4.4

A 2×2 matrix A has an eigenvector $v_1 = (1,1)^t$ with eigenvalue 2 and also an eigenvector $v_2 = (1,2)^t$ with eigenvalue 3. Determine A.

Solution. We are given that

$$Av_1 = 2v_1 \implies A\begin{bmatrix}1\\1\end{bmatrix} = \begin{bmatrix}2\\2\end{bmatrix} \quad \text{and} \quad Av_2 = 3v_2 \implies A\begin{bmatrix}1\\2\end{bmatrix} = \begin{bmatrix}3\\6\end{bmatrix}$$

We can find the columns of A by computing Ae_1 and Ae_2:

$$A\begin{bmatrix}1\\0\end{bmatrix} = A\left(\begin{bmatrix}2\\2\end{bmatrix} - \begin{bmatrix}1\\2\end{bmatrix}\right) = A(2v_1 - v_2) = 2Av_1 - Av_2 = 2\begin{bmatrix}2\\2\end{bmatrix} - \begin{bmatrix}3\\6\end{bmatrix} = \begin{bmatrix}1\\-2\end{bmatrix}$$

$$A\begin{bmatrix}0\\1\end{bmatrix} = A\left(\begin{bmatrix}1\\2\end{bmatrix} - \begin{bmatrix}1\\1\end{bmatrix}\right) = A(v_2 - v_1) = Av_2 - Av_1 = \begin{bmatrix}3\\6\end{bmatrix} - \begin{bmatrix}2\\2\end{bmatrix} = \begin{bmatrix}1\\4\end{bmatrix}$$

Therefore

$$A = \begin{bmatrix}1 & 1\\-2 & 4\end{bmatrix}$$

4.5

Find all invariant subspaces of the real linear operator whose matrix is

$$\text{(a)} \begin{bmatrix} 1 & 1 \\ & 1 \end{bmatrix}, \qquad \text{(b)} \begin{bmatrix} 1 & & \\ & 2 & \\ & & 3 \end{bmatrix}$$

Solution.

Denote $A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ and $B = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix}$.

(a) We proceed by dimension:

- Dimension 0: Only $\{0\}$.
- Dimension 1: We need to find a nonzero vector v such that for every $kv \in \operatorname{span}\{v\}$, we have $A(kv) = kAv \in \operatorname{span}\{v\}$. Another way to word this is that v is an eigenvector of A. Note that

$$Av = \lambda v \iff \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} v_1 + v_2 \\ v_2 \end{bmatrix} = \begin{bmatrix} \lambda v_1 \\ \lambda v_2 \end{bmatrix} \iff \begin{cases} \lambda = 1 \\ v_2 = 0 \end{cases}$$

  Hence $\operatorname{span}\left\{[1 \;\; 0]^t\right\}$ is the only A-invariant subspace of dimension 1.

- Dimension 2: Only $\mathbb{R}^2$.

(b) Again we proceed by dimension:

- Dimension 0: Only $\{0\}$.
- Dimension 1: As in (a), it is only the span of a single eigenvector. B has eigenvalues of $1, 2, 3$ with eigenvectors e_1, e_2, e_3 respectively. Hence the B-invariant subspaces are $\operatorname{span}\{e_1\}, \operatorname{span}\{e_2\}, \operatorname{span}\{e_3\}$.
- Dimension 2: The span of any two eigenvectors will work. More generally, if $v_1, \dots, v_k$ are eigenvectors of T with eigenvalues $\lambda_1, \dots, \lambda_k$, then

$$T\left(\sum_{i=1}^{k} \alpha_i v_i\right) = \sum_{i=1}^{k} \alpha_i T(v_i) = \sum_{i=1}^{k} (\alpha_i \lambda_i) v_i \in \operatorname{span}\{v_1, \dots, v_k\}$$

so $\operatorname{span}\{v_1, \dots, v_k\}$ is a T-invariant subspace of dimension k. Hence we have the B-invariant subspaces: $\operatorname{span}\{e_1, e_2\}, \operatorname{span}\{e_2, e_3\}, \operatorname{span}\{e_1, e_3\}$.

To show that these are the only ones, let W be a 2-dimensional B-invariant subspace. Note that we can restrict our original linear operator T to $T|_W : W \to W$, which will have the same matrix representation B with respect to the standard basis. In particular, B being a positive matrix means that $T|_W$ has a real eigenvalue λ, so there exists $w \in W$ such that $Bw = \lambda w$. But since B is the matrix representation of T, w must also be an eigenvector of T and so $\lambda \in \{1, 2, 3\}$ and w is a scalar multiple of e_1, e_2, e_3. Suppose that $w = ke_1$. Then we have $e_1 \in W$. Choose $w' \in W$ not in the span of e_1, i.e. $W = \operatorname{span}\{e_1, w'\}$. If we write $w' = ae_1 + be_2 + ce_3$, note that

$$Bw' = ae_1 + 2be_2 + 3ce_3$$

But by B-invariance we have $Bw \in W = \operatorname{span}\{e_1, w'\}$ so there exists α, β such that

$$Bw' = \alpha e_1 + \beta w' = \alpha e_1 + \beta(ae_1 + be_2 + ce_3) = (\alpha + \beta b)e_1 + \beta b e_2 + \beta c e_3$$

Hence we have system

$$\begin{cases} a = \alpha + \beta b \\ 2b = \beta b \\ 3c = \beta c \end{cases} \implies \begin{cases} \beta = 2 \text{ or } b = 0 \\ \beta = 3 \text{ or } c = 0 \end{cases}$$

Since β cannot be both 2 and 3, we have $b = 0$ or $c = 0$. Furthermore, we cannot have $b = c = 0$ since otherwise $w' = ae_1 \in \operatorname{span}\{e_1\}$ which contradicts our choice of w'. Hence exactly one of b, c is zero and thus $w' \in \operatorname{span}\{e_1, e_2\}$ or $w' \in \operatorname{span}\{e_1, e_3\}$ and $W = \operatorname{span}\{e_1, e_2\}$ or $W = \operatorname{span}\{e_1, e_3\}$. The other cases of $w = ke_2$ and $w = ke_3$ give similar results, therefore we have $W = \operatorname{span}\{e_i, e_j\}$ for $i \neq j$.
- Dimension 3: Only $\mathbb{R}^3$.

4.6

Let P be the real vector space of polynomials $p(x) = a_0 + a_1x + \cdots + a_nx^n$ of degree at most n, and let D denote the derivative $\frac{d}{dx}$, considered as a linear operator on P.

(a) Prove that D is a nilpotent operator, meaning that $D^k = 0$ for sufficiently large k.

(b) Find the matrix of D with respect to a convenient basis.

(c) Determine all D-invariant subspaces of P.

Solution.

(a) *Proof.*

From the power rule $(\frac{d}{dx}(x^k) = kx^{k-1})$ we have that $\dfrac{d^{n+1}}{dx^{n+1}}(x^k) = 0$ for all $0 \leq k \leq n$. Hence for any $p(x) \in P$,

$$D^{n+1}(p(x)) = D^{n+1}(a_0+a_1x+\cdots+a_nx^n) = a_0D^{n+1}(1)+a_1D^{n+1}(x)+\cdots+a_nD^{n+1}(x^n) = 0$$

Therefore $D^{n+1} = 0$ and D is nilpotent. $\square$

(b) If we use the basis

$$1, x, x^2, \ldots, x^n \tag{$\star$}$$

Then we have correspondence

$$a_0 + a_1x + \cdots + a_nx^n \leftrightsquigarrow \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_n \end{bmatrix}$$

and applying D to each basis vector gives

$$D(1) = 0 \leftrightsquigarrow \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix}, \quad D(x) = 1 \leftrightsquigarrow \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix}, \quad D(x^2) = 2x \leftrightsquigarrow \begin{bmatrix} 0 \\ 2 \\ \vdots \\ 0 \\ 0 \end{bmatrix}, \quad \ldots, \quad D(x^n) = nx^{n-1} \leftrightsquigarrow \begin{bmatrix} 0 \\ 0 \\ \vdots \\ n \\ 0 \end{bmatrix}$$

Hence the matrix associated to D with respect to basis $(\star)$ is

$$\begin{bmatrix} 0 & 1 & 0 & \ldots & 0 \\ 0 & 0 & 2 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \ldots & n \\ 0 & 0 & 0 & \ldots & 0 \end{bmatrix}$$

(c) Since taking the derivative of a degree-k polynomial gives one of degree $(k-1)$, any D-invariant subspace with a polynomial p must include all polynomials of degrees less than the degree of p. Hence for $k = n, \ldots, 0, -1$

$$W \text{ is a } D\text{-invariant subspace of } P \iff W = \{p(x) \in P \mid \deg(p) \leq k\} =: P_k$$

where $P_{-1} = \{0\}$ and $P_n = P$.

4.7

Let $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ be a real 2×2 matrix. The condition that a column vector X be an eigenvector for left multiplication by A is that $AX = Y$ be a scalar multiple of X, which means that the slopes $s = x_2/x_1$ and $s' = y_2/y_1$ are equal.

(a) Find the equation in s that expresses this equality.

(b) Suppose that the entries of A are positive real numbers. Prove that there is an eigenvector in the first quadrant and also one in the second quadrant.

Solution.

(a) We can compute

$$AX = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} ax_1 + bx_2 \\ cx_1 + dx_2 \end{bmatrix}$$

Hence

$$s' = s \iff \frac{cx_1 + dx_2}{ax_1 + bx_2} = \frac{x_2}{x_1}$$

If we substitute $x_2 = x_1 s$ into the RHS we get

$$s = \frac{x_2}{x_1} = \frac{cx_1 + dx_1 s}{ax_1 + bx_1 s} = \frac{x_1(c + ds)}{x_1(a + bs)} = \frac{c + ds}{a + bs} \implies c + ds = s(a + bs) = as + bs^2$$

which rearranges to the quadratic $bs^2 + (a - d)s - c = 0$ and is our desired equation in s.

(b) *Proof.*
From (a), finding an eigenvector X amounts to finding an $s = x_2/x_1$ such that the quadratic $f(s) = bs^2 + (a - d)s - c$ has a root. Since we can scale the eigenvector by any scalar, we can simply set $x_1 = 1$ to begin with and thus $s = x_2$ is the only value we vary. Furthermore, we are given that $a, b, c, d > 0$, so

$$\begin{cases} c > 0 \implies f(0) = -c < 0 \\ b > 0 \implies f(s) \to +\infty \text{ as } s \to +\infty \\ b > 0 \implies f(s) \to +\infty \text{ as } s \to -\infty \end{cases}$$

Since $f(s)$ is continuous, by the intermediate value theorem there must exist $t_1 > 0$ and $t_2 < 0$ such that $f(t_1) = f(t_2) = 0$. But now $[1 \;\; t_1]^t$ and $[1 \;\; t_2]^t$ are eigenvectors in the first and second quadrant respectively. □

4.8

Let T be a linear operator on a finite-dimensional vector space for which every nonzero vector is an eigenvector. Prove that T is multiplication by a scalar.

Solution.

Proof.
Let V be a n-dimensional vector space over F with basis $v_1, \ldots, v_n$ and $T : V \to V$ a linear operator. Now by assumption there exists $\lambda_1, \ldots, \lambda_n, \mu \in F$ such that

$$T(v_1) = \lambda_1 v_1, \quad \ldots, \quad T(v_n) = \lambda_n v_n, \quad \text{and} \quad T(v_1 + \cdots + v_n) = \mu(v_1 + \cdots + v_n)$$

But by linearity

$$\mu(v_1 + \cdots + v_n) = T(v_1 + \cdots + v_n) = T(v_1) + \cdots + T(v_n) = \lambda_1 v_1 + \cdots + \lambda_n v_n$$

Hence

$$(\mu - \lambda_1)v_1 + \cdots + \cdots + (\mu - \lambda_n)v_n = 0$$

and the independence of $v_1, \ldots, v_n$ forces $\mu = \lambda_i$ for $i = 1, \ldots, n$. Now for any $v \in V$, there exists $\alpha_1, \ldots, \alpha_n$ such that

$$v = \sum_{i=1}^{n} \alpha_i v_i \implies T(v) = T\left(\sum_{i=1}^{n} \alpha_i v_i\right) = \sum_{i=1}^{n} \alpha_i T(v_i) = \sum_{i=1}^{n} \alpha_i \mu v_i = \mu \sum_{i=1}^{n} \alpha_i v_i = \mu v$$

Hence T is simply multiplication by the scalar μ. □

§5 - The Characteristic Polynomial

5.1

Compute the characteristic polynomials and the complex eigenvalues and eigenvectors of

(a) $\begin{bmatrix} -2 & 2 \\ -2 & 3 \end{bmatrix}$, (b) $\begin{bmatrix} 1 & i \\ -i & 1 \end{bmatrix}$, (c) $\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$

Solution.

(a) We have characteristic polynomial

$$p(t) = \det(tI - A) = \det \begin{bmatrix} t+2 & -2 \\ 2 & t-3 \end{bmatrix} = (t^2 - t - 6) + 4 = t^2 - t - 2 = (t-2)(t+1)$$

Thus we have eigenvalues $\lambda_1 = 2, \lambda_2 = -1$. To find eigenvectors, we have the system

$$2 \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \begin{bmatrix} -2 & 2 \\ -2 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} -2x_1 + 2x_2 \\ -2x_1 + 3x_3 \end{bmatrix} \implies \begin{cases} 2x_1 = -2x_1 + 2x_2 \\ 2x_2 = -2x_1 + 3x_2 \end{cases} \implies x_2 = 2x_1$$

Hence $\begin{bmatrix} 1 \\ 2 \end{bmatrix}$ is an eigenvector with eigenvalue 2. Similarly

$$-1 \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \begin{bmatrix} -2 & 2 \\ -2 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} -2x_1 + 2x_2 \\ -2x_1 + 3x_3 \end{bmatrix} \implies \begin{cases} -x_1 = -2x_1 + 2x_2 \\ -x_2 = -2x_1 + 3x_2 \end{cases} \implies x_2 = \tfrac{1}{2}x_1$$

Hence $\begin{bmatrix} 2 \\ 1 \end{bmatrix}$ is an eigenvector with eigenvalue -1.

(b) We have characteristic polynomial

$$p(t) = \det(tI - B) = \det \begin{bmatrix} t-1 & -i \\ i & t-1 \end{bmatrix} = (t^2 - 2t + 1) - 1 = t(t-2)$$

Thus we have eigenvalues $\lambda_1 = 0, \lambda_2 = 2$. To find eigenvectors, we have the system

$$0 \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 & i \\ -i & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_1 + ix_2 \\ -ix_1 + x_2 \end{bmatrix} \implies \begin{cases} 0 = x_1 + ix_2 \\ 0 = -ix_1 + x_2 \end{cases} \implies x_2 = ix_1$$

Hence $\begin{bmatrix} 1 \\ i \end{bmatrix}$ is an eigenvector with eigenvalue 0. Similarly

$$2 \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 & i \\ -i & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_1 + ix_2 \\ -ix_1 + x_2 \end{bmatrix} \implies \begin{cases} 2x_1 = x_1 + ix_2 \\ 2x_2 = -ix_1 + x_2 \end{cases} \implies x_2 = -ix_1$$

Hence $\begin{bmatrix} 1 \\ -i \end{bmatrix}$ is an eigenvector with eigenvalue 2.

(c) We have characteristic polynomial

$$p(t) = \det(tI - C) = \det \begin{bmatrix} t - \cos\theta & \sin\theta \\ -\sin\theta & t - \cos\theta \end{bmatrix} = (t^2 - 2t\cos\theta + \cos^2\theta) + \sin^2\theta = t^2 - (2\cos\theta)t + 1$$

and the quadratic formula gives eigenvalues

$$\lambda = \frac{2\cos\theta \pm \sqrt{4\cos^2\theta - 4}}{2} = \cos\theta \pm \tfrac{1}{2}\sqrt{-4\sin^2\theta} = \cos\theta \pm i\sin\theta$$

To find eigenvectors, we have the system

$$(\cos\theta + i\sin\theta)\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_1\cos\theta - x_2\sin\theta \\ x_1\sin\theta + x_2\cos\theta \end{bmatrix}$$

$$\implies \begin{cases} x_1\cos\theta + x_1 i\sin\theta = x_1\cos\theta - x_2\sin\theta \\ x_2\cos\theta + x_2 i\sin\theta = x_1\sin\theta + x_2\cos\theta \end{cases} \implies x_2 = -ix_1$$

Hence $\begin{bmatrix} 1 \\ -i \end{bmatrix}$ is an eigenvector with eigenvalue $\cos\theta + i\sin\theta$. Similarly

$$(\cos\theta - i\sin\theta)\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_1\cos\theta - x_2\sin\theta \\ x_1\sin\theta + x_2\cos\theta \end{bmatrix}$$

$$\implies \begin{cases} x_1\cos\theta - x_1 i\sin\theta = x_1\cos\theta - x_2\sin\theta \\ x_2\cos\theta - x_2 i\sin\theta = x_1\sin\theta + x_2\cos\theta \end{cases} \implies x_2 = ix_1$$

Hence $\begin{bmatrix} 1 \\ i \end{bmatrix}$ is an eigenvector with eigenvalue $\cos\theta - i\sin\theta$.

5.2

The characteristic polynomial of the matrix below is $t^3 - 4t - 1$. Determine the missing entries.

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

Solution.
If we write our matrix as

$$A = \begin{bmatrix} 0 & 1 & 2 \\ 1 & 1 & 0 \\ 1 & x & y \end{bmatrix}$$

We get characteristic polynomial

$$\begin{aligned} p(t) = \det(tI - A) = \det \begin{bmatrix} t & -1 & -2 \\ -1 & t-1 & 0 \\ -1 & -x & t-y \end{bmatrix} \\ &= -2(x + t - 1) + (t - y)(t^2 - t - 1) \\ &= t^3 + (-1 - y)t^2 + (-3 + y)t + (2 - 2x + y) \end{aligned}$$

So for it to equal $t^3 - 4t - 1$, we need $-1 - y = 0 \implies y = -1$ and

$$-1 = 2 - 2x + y = 1 - 2x \implies x = 1$$

Thus the missing entries are $x = 1$ and $y = -1$.

Another quicker approach follows when one can note from the characteristic polynomial that the trace of A must be zero, hence $y = -1$ follows immediately. The characteristic polynomial also says that $\det A = 1$, and with $y = -1$ we have $\det A = 2x - 1$ and thus $x = 1$.

5.3

What complex numbers might be eigenvalues of a linear operator such that
(a) $T^r = I$, (b) $T^2 - 5T + 6I = 0$?

Solution.
First note for any linear operator T, if v is an eigenvector with eigenvalue λ, then

$$T^k(v) = T^{k-1}(\lambda v) = T^{k-2}(\lambda^2 v) = \cdots = \lambda^k v$$

Thus for any "polynomial" linear operator $a_n T^n + \cdots + a_1 T + a_0 I$, we have

$$(a_n T^n + \cdots + a_1 T + a_0 I)(v) = a_n T^n(v) + \cdots + a_1 T(v) + a_0 I(v) = (a_n \lambda^n + \cdots + a_1 \lambda + a_0)v$$

With this fact we can now answer the exercise.

(a) If $T^r = I$ for some r, then any eigenvector v with eigenvalue λ will have

$$v = I(v) = T^r(v) = \lambda^r v \implies \lambda^r = 1$$

Thus any eigenvalue λ is an r-th root of unity (assuming r is the smallest such value where $T^r = I$).

(b) If $T^2 - 5T + 6I = 0$, then any eigenvector v with eigenvalue λ will have

$$0 = (T^2 - 5T + 6I)(v) = (\lambda^2 - 5\lambda + 6)v \implies 0 = \lambda^2 - 5\lambda + 6 = (\lambda - 3)(\lambda - 2)$$

Thus any eigenvalue is either $\lambda = 3$ or $\lambda = 2$.

5.4

Find a recursive relation for the characteristic polynomial of the $k \times k$ matrix

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

and compute the polynomial for $k \leq 5$.

Solution.
Denote A_k to be the $k \times k$ matrix above and $p_k(t)$ its characteristic polynomial. Then

$$\begin{aligned} p_k(t) = \det(tI - A_k) &= \det \begin{bmatrix} t & -1 & & & & \\ -1 & t & -1 & & & \\ & -1 & t & -1 & & \\ & & . & . & . & \\ & & & . & . & -1 \\ & & & & -1 & t \end{bmatrix} \\ &= t \det \underbrace{\begin{bmatrix} t & -1 & & & \\ -1 & t & -1 & & \\ & . & . & . & \\ & & . & . & -1 \\ & & & -1 & t \end{bmatrix}}_{tI - A_{k-1}} + 1 \det \begin{bmatrix} -1 & & & & \\ -1 & t & -1 & & \\ & . & . & . & \\ & & . & . & -1 \\ & & & -1 & t \end{bmatrix} + 0 + \dots \\ &= tp_{k-1}(t) + 1 \left(-1 \det \underbrace{\begin{bmatrix} t & -1 & & \\ . & . & . & \\ & . & . & -1 \\ & & -1 & t \end{bmatrix}}_{tI - A_{k-2}} + 0 + \dots \right) \\ &= tp_{k-1}(t) - p_{k-2}(t) \end{aligned}$$

where we expanded along the first column in the first determinant computation and then expanded along the first row in the second determinant computation. Thus we have

- $p_1(t) = \det \begin{bmatrix} t \end{bmatrix} = t$
- $p_2(t) = \det \begin{bmatrix} t & -1 \\ -1 & t \end{bmatrix} = t^2 - 1$
- $p_3(t) = tp_2(t) - p_1(t) = t(t^2 - 1) - t = t^3 - 2t$
- $p_4(t) = tp_3(t) - p_2(t) = t(t^3 - 2t) - (t^2 - 1) = t^4 - 3t^2 + 1$
- $p_5(t) = tp_4(t) - p_3(t) = t(t^4 - 3t^2 + 1) - (t^3 - 2t) = t^5 - 4t^3 + 3t$

5.5

Which real 2×2 matrices have real eigenvalues? Prove that the eigenvalues are real if the off-diagonal entries have the same sign.

Solution.
Let $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ be a real 2×2 matrix. Then it has characteristic polynomial

$$p(t) = \det(tI - A) = \det \begin{bmatrix} t-a & -b \\ -c & t-d \end{bmatrix} = (t-a)(t-d) - bc = t^2 - (a+d)t + (ad - bc)$$

The roots of this quadratic are the eigenvalues of A, so we get real eigenvalues λ_1, λ_2 if and only if the discriminant of $p(t)$ is ≥ 0, i.e.

$$\begin{aligned} \lambda_1, \lambda_2 \in \mathbb{R} &\iff (a+d)^2 - 4(ad - bc) \geq 0 \\ &\iff (a^2 + 2ad + d^2) - 4ad + 4bc \geq 0 \\ &\iff (a^2 - 2ad + d^2) + 4bc \geq 0 \\ &\iff (a-d)^2 \geq -4bc \end{aligned}$$

Thus the set of all matrices with real eigenvalues is

$$\left\{ \begin{bmatrix} a & b \\ c & d \end{bmatrix} \middle| (a-d)^2 \geq -4bc \right\}$$

In particular, if the off-diagonal entries of A have the same sign, then $bc \geq 0 \implies -4bc \leq 0$ and $(a-d)^2$ is always nonnegative, hence

$$(a-d)^2 \geq 0 \geq -4bc$$

and A has real eigenvalues.

5.6

Let V be a vector space with basis $(v_0, \dots, v_n)$ and let $a_0, \dots, a_n$ be scalars. Define a linear operator T on V by the rules $T(v_i) = v_{i+1}$ if $i < n$ and $T(v_n) = a_0v_0 + a_1v_1 + \cdots + a_nv_n$. Determine the matrix of T with respect to the given basis, and the characteristic polynomial of T.

Solution.
Let A be the matrix representation of T with respect to $v_1, \dots, v_n$. Then applying T to each basis vector gives

$$T(v_0) = v_1 \leftrightsquigarrow \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad T(v_1) = v_2 \leftrightsquigarrow \begin{bmatrix} 0 \\ 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}, \quad \dots, \quad T(v_{n-1}) = v_n \leftrightsquigarrow \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}$$

and finally

$$T(v_n) = a_0v_0 + a_1v_1 + \cdots + a_nv_n \leftrightsquigarrow \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}$$

Thus stacking these as columns of A gives

$$A = \begin{bmatrix} 0 & 0 & \dots & 0 & a_0 \\ 1 & 0 & \dots & 0 & a_1 \\ 0 & 1 & \dots & 0 & a_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & a_n \end{bmatrix}$$

We then have characteristic polynomial (via expanding across the first row)

$$\begin{aligned} p(t) = \det(tI - A) &= \det \begin{bmatrix} t & 0 & \dots & 0 & 0 & -a_0 \\ -1 & t & \dots & 0 & 0 & -a_1 \\ 0 & -1 & \ddots & 0 & 0 & -a_2 \\ \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & -1 & t & -a_{n-1} \\ 0 & 0 & \dots & 0 & -1 & -a_n \end{bmatrix} \\ &= t \det \begin{bmatrix} t & \dots & 0 & 0 & -a_1 \\ -1 & \ddots & 0 & 0 & -a_2 \\ \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & \dots & -1 & t & -a_{n-1} \\ 0 & \dots & 0 & -1 & -a_n \end{bmatrix} + (-1)^{n+1} a_0 \det \begin{bmatrix} -1 & t & \dots & 0 & 0 \\ 0 & -1 & \ddots & 0 & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \dots & -1 & t \\ 0 & 0 & \dots & 0 & -1 \end{bmatrix} \\ &=: tp_1(t) + (-1)^{n+1} a_0 (-1)^n \\ &= tp_1(t) - a_0 \end{aligned}$$

where $p_i(t)$ is the characteristic polynomial of our original problem but on the vector space $V' = \text{span}\{v_i, v_{i+1}, \ldots, v_n\}$ and scalars $a_i, a_{i+1}, \ldots, a_n$. In this case the above shows the recursive relation $p_i(t) = tp_{i+1}(t) - a_i$. Thus if we start from $i = n$, we can work back down to our desired $p(t) = p_0(t)$ via

$$\begin{aligned} p_n(t) &= \det[t - a_n] = t - a_n \\ p_{n-1}(t) &= tp_n(t) - a_{n-1} = t(t - a_n) - a_{n-1} = t^2 - a_n t - a_{n-1} \\ p_{n-2}(t) &= tp_{n-1}(t) - a_{n-2} = t(t^2 - a_n t - a_{n-1}) - a_{n-2} = t^3 - a_n t^2 - a_{n-1} t - a_{n-2} \\ &\ldots \\ p(t) = p_0(t) &= t^{n+1} - a_n t^n - a_{n-1} t^{n-1} - \cdots - a_1 t - a_0 \end{aligned}$$

5.7

Do A and A^t have the same eigenvectors? the same eigenvalues?

Solution.
They do not (in general) have the same eigenvectors, e.g. from Exercise 5.1 $\begin{bmatrix} -2 & 2 \\ -2 & 3 \end{bmatrix}$ has eigenvector $[1 \;\; 2]^t$ with eigenvalue 2 but

$$\begin{bmatrix} -2 & 2 \\ -2 & 3 \end{bmatrix}^t \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} -2 & -2 \\ 2 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} -6 \\ 8 \end{bmatrix} \neq \lambda \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

so $[1 \;\; 2]^t$ is not an eigenvector of A^t. However, we claim that A and A^t do share eigenvalues.

Proof.
Let λ be an eigenvalue of A. Then in particular $p(\lambda) = 0$ for the characteristic polynomial of A, i.e. $\det(\lambda I - A) = 0$. But $\det(B) = \det(B^t)$ (from Corollary 1.4.15(b)) implies that

$$0 = \det(\lambda I - A) = \det((\lambda I - A)^t) = \det((\lambda I)^t - A^t) = \det(\lambda I - A^t)$$

Hence λ is an eigenvalue of A^t. The same argument shows that any eigenvalue of A^t is an eigenvalue of A, therefore the two matrices have the same eigenvalues. $\square$

5.8

Let $A = (a_{ij})$ be a 3×3 matrix. Prove that the coefficient of t in the characteristic polynomial is the sum of the symmetric 2×2 minors

$$\det \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} + \det \begin{bmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{bmatrix} + \det \begin{bmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{bmatrix}$$

Solution.

Proof.
Write

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

Then we have characteristic polynomial

$$\begin{aligned} \det(tI - A) &= \det \begin{bmatrix} t - a_{11} & -a_{12} & -a_{13} \\ -a_{21} & t - a_{22} & -a_{23} \\ -a_{31} & -a_{32} & t - a_{33} \end{bmatrix} \\ &= (t - a_{11}) \det \begin{bmatrix} t - a_{22} & -a_{23} \\ -a_{32} & t - a_{33} \end{bmatrix} \\ &\quad + a_{21} \det \begin{bmatrix} -a_{12} & -a_{13} \\ -a_{32} & t - a_{33} \end{bmatrix} \\ &\quad - a_{31} \det \begin{bmatrix} -a_{12} & -a_{13} \\ t - a_{22} & -a_{23} \end{bmatrix} \\ &= (t - a_{11})[(t - a_{22})(t - a_{33}) - a_{23}a_{32}] \\ &\quad + a_{21}[-a_{12}(t - a_{33}) - a_{13}a_{32}] \\ &\quad - a_{31}[a_{12}a_{23} + a_{13}(t - a_{22})] \\ &= [t^3 - (a_{11} + a_{22} + a_{33})t^2 + (a_{22}a_{33} - a_{23}a_{32} + a_{11}a_{22} + a_{11}a_{33})t + k_1] \\ &\quad + [-a_{12}a_{21}t + k_2] \\ &\quad + [-a_{13}a_{31}t + k_3] \\ &= t^3 - (\operatorname{tr} A)t^2 + [a_{22}a_{33} - a_{23}a_{32} + a_{11}a_{22} + a_{11}a_{33} - a_{12}a_{21} - a_{13}a_{31}]t + \det A \end{aligned}$$

And indeed the coefficient of the t term is equal to

$$\begin{aligned} &(a_{11}a_{22} - a_{12}a_{21}) + (a_{22}a_{33} - a_{23}a_{32}) + (a_{11}a_{33} - a_{13}a_{31}) \\ &\qquad = \det \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} + \det \begin{bmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{bmatrix} + \det \begin{bmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{bmatrix} \end{aligned}$$

□

5.9

Consider the linear operator of left multiplication by an $m \times m$ matrix A on the space $F^{m\times m}$ of all $m \times m$ matrices. Determine the trace and determinant of this operator.

Solution.
Denote this operator $L_A : X \mapsto AX$. First note that if $v \in F^m$ is an eigenvector of A with eigenvalue λ, then for any $w \in F^m$ we have

$$L_A(vw^t) = A(vw^t) = (Av)w^t = (\lambda v)w^t = \lambda(vw^t)$$

Also note that we can construct m linearly independent matrices starting from v, e.g. $ve_1^t, \dots, ve_m^t$. Hence for every eigenvalue λ of A, we get m linearly independent matrices with that eigenvalue (not counting multiplicity). Therefore

$$\operatorname{tr}(A) = \sum_{\lambda \text{ eigenvalue}} \lambda \implies \operatorname{tr}(L_A) = \sum_{\lambda \text{ eigenvalue}} m\lambda = m \operatorname{tr}(A)$$

and

$$\det(A) = \prod_{\lambda \text{ eigenvalue}} \lambda \implies \det(L_A) = \prod_{\lambda \text{ eigenvalue}} \lambda^m = (\det A)^m$$

5.10

Let A and B be $n \times n$ matrices. Determine the trace and the determinant of the operator on the space $F^{n\times n}$ defined by $M \rightsquigarrow AMB$.

Solution.
Denote this operator $T : M \mapsto AMB$. Like in Exercise 5.9, note that if v is an eigenvector of A with eigenvalue λ and w is an eigenvector of B^t with eigenvalue μ, then

$$T(vw^t) = A(vw^t)B = (Av)(w^tB) = (Av)(B^tw)^t = (\lambda v)(\mu w)^t = (\lambda\mu)vw^t$$

Therefore we have (not counting multiplicity)

$$\operatorname{tr}(T) = \sum_{\lambda \text{ eigenvalue } A} \sum_{\mu \text{ eigenvalue } B} \lambda\mu = \left(\sum_{\lambda \text{ eigenvalue } A} \lambda\right)\left(\sum_{\mu \text{ eigenvalue } B} \mu\right) = (\operatorname{tr} A)(\operatorname{tr} B)$$

and

$$\det(T) = \prod_{\lambda \text{ eigenvalue } A} \prod_{\mu \text{ eigenvalue } B} \lambda\mu = \left(\prod_{\lambda \text{ eigenvalue } A} \lambda^n\right)\left(\prod_{\mu \text{ eigenvalue } B} \mu^n\right) = (\det A)^n(\det B)^n$$

6.1

Let A be an $n \times n$ matrix whose characteristic polynomial factors into linear factors: $p(t) = (t - \lambda_1) \dots (t - \lambda_n)$. Prove that $\operatorname{tr} A = \lambda_1 + \dots + \lambda_n$, that $\det A = \lambda_1 \dots \lambda_n$.

Solution.

Proof.
By Corollary 4.6.4(b), there is a matrix $P \in GL_n(F)$ such that $B := P^{-1}AP$ is upper triangular. By Prop 4.5.11, the characteristic polynomial of B is the same as A, i.e. $p(t)$, and by Corollary 4.5.9 the eigenvalues of B are the roots of $p(t)$, thus B has eigenvalues $\lambda_1, \dots, \lambda_n$. Now since B is upper triangular, its eigenvalues must be its diagonal entries by Prop 4.5.10 and so by definition

$$\operatorname{tr} B = \lambda_1 + \dots + \lambda_n$$

However (e.g. from Exercise 1.M.3) we have

$$\operatorname{tr} A = \operatorname{tr}(P^{-1}AP) = \operatorname{tr} B = \lambda_1 + \dots + \lambda_n$$

Furthermore, the determinant of an upper triangular matrix is the product of the diagonal entries (e.g. from Exercise 1.4.6) and so $\det B = \lambda_1 \dots \lambda_n$ and again

$$\det A = \det(P^{-1}AP) = \det B = \lambda_1 \dots \lambda_n$$

□

6.2

Suppose that a complex $n \times n$ matrix A has distinct eigenvalues $\lambda_1 \dots \lambda_n$, and let $v_1, \dots, v_n$ be eigenvectors with these eigenvalues.

(a) Show that every eigenvector is a multiple of one of the vectors v_i.

(b) Show how one can recover the matrix from the eigenvalues and eigenvectors.

Solution.

(a) *Proof.*
Let w be an eigenvector of A, say $Aw = \mu w$ for some $\mu \in \mathbb{C}$. By Prop 4.5.14, A has at most n eigenvalues so by distinctness we have $\mu = \lambda_i$ for some $1 \leq i \leq n$. Note that by Prop 4.6.5 we have that $v_1, \dots, v_n$ forms a basis of $\mathbb{C}^n$, thus there exists unique $\alpha_1, \dots, \alpha_n \in \mathbb{C}$ such that $w = \sum \alpha_j v_j$. Thus

$$\mu w = \mu \left(\sum_{j=1}^{n} \alpha_j v_j \right) = \sum_{j=1}^{n} (\mu \alpha_j) v_j \tag{$\dagger$}$$

However as μ is an eigenvalue, we also have

$$\mu w = Aw = A \left(\sum_{j=1}^{n} \alpha_j v_j \right) = \sum_{j=1}^{n} \alpha_j (A v_j) = \sum_{j=1}^{n} \alpha_j (\lambda_j v_j) = \sum_{j=1}^{n} (\lambda_j \alpha_j) v_j \tag{$\star$}$$

Hence from ($\dagger$) and ($\star$) we get

$$0 = \mu w - \mu w = \sum_{j=1}^{n} (\mu \alpha_j) v_j - \sum_{j=1}^{n} (\lambda_j \alpha_j) v_j = \sum_{j=1}^{n} (\mu - \lambda_j) \alpha_j v_j$$

Now by linear independence this forces $(\mu - \lambda_j)\alpha_j = 0$ for all $j = 1, \dots, n$. However since $\mu = \lambda_i$, by distinctness we have $(\mu - \lambda_j) \neq 0$ for all $j \neq i$ and so $\alpha_j = 0$ for $j \neq i$. Therefore

$$w = \alpha_1 v_1 + \cdots + \alpha_i v_i + \cdots + \alpha_n v_n = 0 + \cdots + \alpha_i v_i + \cdots + 0 = \alpha_i v_i$$

and w is a multiple of v_i. □

(b) *Proof.*
It suffices to recover the jth column of A, i.e. calculate Ae_j. Note that by Prop 4.6.5 we have that $v_1, \dots, v_n$ forms a basis of $\mathbb{C}^n$, thus there exists unique $\alpha_1, \dots, \alpha_n \in \mathbb{C}$ such that $e_j = \sum \alpha_i v_i$. Then

$$Ae_j = A \left(\sum_{i=1}^{n} \alpha_i v_i \right) = \sum_{i=1}^{n} \alpha_i (A v_i) = \sum_{i=1}^{n} \alpha_i (\lambda_i v_i) = \sum_{i=1}^{n} (\alpha_i \lambda_i) v_i$$

which are all known scalars and vectors. □

6.3

Let T be a linear operator that has two linearly independent eigenvectors with the same eigenvalue λ. Prove that λ is a multiple root of the characteristic polynomial of T.

Solution.

Proof.
Let v, w be linearly independent vectors such that $T(v) = \lambda v$ and $T(w) = \lambda w$. Extend v, w to a basis $v, w, b_1, \ldots, b_n$ and let A be the matrix of T with respect to this basis. Then it will have the form

$$A = \left[\begin{array}{cc|c} \lambda & 0 & \multirow{2}{*}{0} \\ 0 & \lambda & \\ \hline \multicolumn{2}{c|}{0} & B \end{array}\right]$$

where B is a $n \times n$ matrix. Now by definition the characteristic polynomial $p(t)$ of T is

$$\begin{aligned} p(t) = \det(tI - A) &= \det \left[\begin{array}{cc|c} t-\lambda & 0 & \multirow{2}{*}{0} \\ 0 & t-\lambda & \\ \hline \multicolumn{2}{c|}{0} & tI_n - B \end{array}\right] \\ &= (t-\lambda)\det \left[\begin{array}{c|c} t-\lambda & 0 \\ \hline 0 & tI_n - B \end{array}\right] \\ &= (t-\lambda)^2 \det(tI_n - B) \end{aligned}$$

Therefore λ is a multiple root of $p(t)$. □

6.4

Let $A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$. Find a matrix P such that $P^{-1}AP$ is diagonal, and find a formula for the matrix A^{30}.

Solution.
We first find an eigenbasis. The eigenvalues of A are 3 and 1, which have corresponding eigenvectors $v = [1 \ \ 1]^t$ and $w = [1 \ -1]^t$. Hence our basechange matrix is

$$P = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \quad \text{and} \quad P^{-1} = -\frac{1}{2}\begin{bmatrix} -1 & -1 \\ -1 & 1 \end{bmatrix}$$

and indeed

$$P^{-1}AP = -\frac{1}{2}\begin{bmatrix} -1 & -1 \\ -1 & 1 \end{bmatrix}\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} = -\frac{1}{2}\begin{bmatrix} -3 & -3 \\ -1 & 1 \end{bmatrix}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} = \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix}$$

is diagonal. Now $(P^{-1}AP)^{30} = P^{-1}A^{30}P$ means

$$\begin{aligned} A^{30} = P(P^{-1}AP)^{30}P^{-1} &= -\frac{1}{2}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}\begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix}^{30}\begin{bmatrix} -1 & -1 \\ -1 & 1 \end{bmatrix} \\ &= -\frac{1}{2}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}\begin{bmatrix} 3^{30} & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} -1 & -1 \\ -1 & 1 \end{bmatrix} \\ &= -\frac{1}{2}\begin{bmatrix} 3^{30} & 1 \\ 3^{30} & -1 \end{bmatrix}\begin{bmatrix} -1 & -1 \\ -1 & 1 \end{bmatrix} \\ &= \frac{1}{2}\begin{bmatrix} 3^{30}+1 & 3^{30}-1 \\ 3^{30}-1 & 3^{30}+1 \end{bmatrix} \end{aligned}$$

6.5

In each case, find a complex matrix P such that $P^{-1}AP$ is diagonal.

(a) $\begin{bmatrix} 1 & i \\ -i & 1 \end{bmatrix}$, (b) $\begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$, (c) $\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$

Solution.

(a) We have eigenvalues 0 and 2 with eigenvectors $[1 \;\; i]^t$ and $[1 \;\; -i]^t$ respectively (see Exercise 5.1(b)). These vectors give an eigenbasis with basechange matrix

$$P = \begin{bmatrix} 1 & 1 \\ i & -i \end{bmatrix}$$

and indeed

$$P^{-1}AP = \begin{bmatrix} 0 & 0 \\ 0 & 2 \end{bmatrix}$$

(b) Note that the characteristic polynomial of the matrix is $p(t) = t^3 - 1$, so our eigenvalues are the cube roots of unity, i.e.

$$\lambda_1 = 1, \quad \lambda_2 = e^{2\pi/3i} = -\frac{1}{2} + \frac{1}{\sqrt{3}}i, \quad \lambda_3 = e^{4\pi/3i} = -\frac{1}{2} - \frac{1}{\sqrt{3}}i$$

These have corresponding eigenvectors

$$v_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \quad v_2 = \begin{bmatrix} 1 \\ \lambda_3 \\ \lambda_2 \end{bmatrix}, \quad v_3 = \begin{bmatrix} 1 \\ \lambda_2 \\ \lambda_3 \end{bmatrix}$$

This gives an eigenbasis with basechange matrix

$$P = \begin{bmatrix} 1 & 1 & 1 \\ 1 & \lambda_3 & \lambda_2 \\ 1 & \lambda_2 & \lambda_3 \end{bmatrix}$$

(c) We have eigenvalues $\cos\theta + i\sin\theta$ and $\cos\theta - i\sin\theta$ with eigenvectors $[1 \;\; -i]^t$ and $[1 \;\; i]^t$ respectively (see Exercise 5.1(c)). These vectors give an eigenbasis with basechange matrix

$$P = \begin{bmatrix} 1 & 1 \\ -i & i \end{bmatrix} \quad \text{and} \quad P^{-1} = \frac{1}{2i}\begin{bmatrix} i & -1 \\ i & 1 \end{bmatrix} = \frac{1}{2}\begin{bmatrix} 1 & i \\ 1 & -i \end{bmatrix}$$

and indeed

$$P^{-1}AP = \begin{bmatrix} \cos\theta + i\sin\theta & 0 \\ 0 & \cos\theta - i\sin\theta \end{bmatrix}$$

6.6

Suppose that A is diagonalizable. Can the diagonalization be done with a matrix P in the special linear group?

Solution.
We claim yes.

Proof.
By assumption there exists an invertible $n \times n$ matrix P such that $P^{-1}AP$ is diagonal. Now set

$$Q := \frac{1}{\sqrt[n]{\det P}}P$$

First note that $Q \in SL_n(\mathbb{C})$ as

$$\det Q = \det(\tfrac{1}{\sqrt[n]{\det P}}P) = (\tfrac{1}{\sqrt[n]{\det P}})^n \cdot \det P = \tfrac{1}{\det P} \cdot \det P = 1$$

Also, we have $Q^{-1} = (\sqrt[n]{\det P})P^{-1}$ since

$$\left(\frac{1}{\sqrt[n]{\det P}}P\right)\left((\sqrt[n]{\det P})P^{-1}\right) = \frac{\sqrt[n]{\det P}}{\sqrt[n]{\det P}}PP^{-1} = I$$

Finally,

$$Q^{-1}AQ = \left((\sqrt[n]{\det P})P^{-1}\right)A\left(\frac{1}{\sqrt[n]{\det P}}P\right) = \frac{\sqrt[n]{\det P}}{\sqrt[n]{\det P}}P^{-1}AP = P^{-1}AP$$

which is diagonal. □

6.7

Prove that if A and B are $n \times n$ matrices and A is nonsingular, then AB is similar to BA.

Solution.

Proof.

$$A^{-1}(AB)A = BA$$

□

6.8

A linear operator T is nilpotent if some positive power T^k is zero. Prove that T is nilpotent if and only if there is a basis of V such that the matrix of T is upper triangular, with diagonal entries zero.

Solution.

Proof.
$\Longrightarrow$: Suppose that T is nilpotent, i.e. T^k is zero for $k > 0$.
First note that T is not invertible since for any matrix representation A, we have

$$0 = \det(A^k) = (\det A)^k$$

and so $\det A = 0$. By rank-nullity, if $\dim V = n$, then we have $\dim \ker T > 0$ and $\dim \operatorname{im} T < n$.
We now induct on n.
Base Case: $n = 1$: Then we have T is just multiplication by a scalar, which must be zero as it is nilpotent, i.e. $k = 1$. Then any matrix representation of T is just zero.
IH: Assume the result holds for any dimension less than n. If we set $W := \operatorname{im} T$, then from the reasoning above we have that $m := \dim W < n$. Furthermore, note that for any $v \in W \subset V$, we clearly have $T(v) \in W$ and so W is T-invariant. Thus the restriction $T|_W$ is a linear operator, and it is nilpotent as $(T|_W)^k$ is zero. Therefore the IH applies and we have a basis $b_1, \ldots, b_m$ of W such that the matrix of $T|_W$ is strictly upper triangular, i.e.

$$T(b_1) = 0 \text{ and } T(b_i) \in \operatorname{span}\{b_1, \ldots, b_{i-1}\} \text{ for } i = 2, \ldots, m$$

Now extend this to a basis $b_1, \ldots, b_m, b_{m+1}, \ldots, b_n$ of V. Indeed, now note for $j = m+1, \ldots, n$ we have

$$T(b_j) \in \operatorname{im} T = W = \operatorname{span}\{b_1, \ldots, b_m\} \subset \operatorname{span}\{b_1, \ldots, b_m, \ldots, b_{j-1}\}$$

Therefore the matrix with respect to $b_1, \ldots, b_n$ is strictly upper triangular.

$\Longleftarrow$: Suppose that $b_1, \ldots, b_n$ is a basis of V such that the matrix with respect to this basis is strictly upper triangular. We claim that T^n is zero.
Set $W_0 := \{0\}$ and $W_i := \operatorname{span}\{b_1, \ldots, b_i\}$ for $i = 1, \ldots, n$. Then note that since the matrix is strictly upper triangular,

$$T(b_1) = 0 \text{ and } T(b_i) \in \operatorname{span}\{b_1, \ldots, b_{i-1}\} = W_{i-1} \text{ for } i = 2, \ldots, n$$

Thus we have $T(W_i) \subset W_{i-1}$, which after repeatedly applying T gives the chain

$$T^n(W_n) \subset T^{n-1}(W_{n-1}) \subset \cdots \subset T(W_1) \subset W_0 = \{0\}$$

But $W_n = \operatorname{span}\{b_1, \ldots, b_n\} = V$ and so $T^n(V) \subset \{0\}$ implies $T^n(v) = 0$ for every $v \in V$, i.e. T^n is zero and therefore T is nilpotent. □

6.9

Find all real 2×2 matrices such that $A^2 = I$, and describe geometrically the way they operate by left multiplication on $\mathbb{R}^2$.

Solution.
First note that $A^2 = I$ implies that $\det A = \pm 1$. Also, we must have $A^{-1} = A$ and so

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \pm 1 \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

If $\det A = 1$, then this forces $b = c = 0$ and $a = d = 1$, i.e. $A = I$, or $a = d = -1$, i.e. $A = -I$. If $\det A = -1$, this forces $d = -a$ and $bc = 1 - a^2$. Therefore

$$\left\{ A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \;\middle|\; A^2 = I \right\} = \{I, -I\} \cup \left\{ \begin{bmatrix} a & b \\ c & -a \end{bmatrix} \;\middle|\; bc = 1 - a^2 \right\}$$

Geometrically, I is the identity transformation and $-I$ is a rotation of 180 degrees.

We claim all other involutory matrices represent a reflection. Given A, note that $\det A = -1$ and $\operatorname{tr} A = 0$ forces its characteristic polynomial to be $p(t) = t^2 - 1$, hence A has eigenvalues $1, -1$. If v_1 has eigenvalue 1 and v_2 has eigenvalue -1, then for any vector $av_1 + bv_2 \in \mathbb{R}^2$, we have

$$A(av_1 + bv_2) = aA(v_1) + bA(v_2) = av_1 - bv_2$$

Hence the line $\operatorname{span}\{v_1\}$ is our line of reflection.

6.10

Let M be a matrix made up of two diagonal blocks: $M = \begin{bmatrix} A & 0 \\ 0 & D \end{bmatrix}$. Prove that M is diagonalizable if and only if A and D are diagonalizable.

Solution.

Proof.
$\Longrightarrow$: Suppose that M (acting on the vector space V) is diagonalizable. Then by Prop 4.4.8(b) there is a basis of eigenvectors $v_1, \ldots, v_n, v_{n+1}, \ldots, v_{n+m} \in V$ (where A is $n \times n$ and D is $m \times m$). Note that if we write these eigenvectors in block form, i.e.

$$v_i = \begin{bmatrix} x_i \\ y_i \end{bmatrix}$$

Then

$$\lambda_i v_i = M v_i \implies \begin{bmatrix} \lambda_i x_i \\ \lambda_i y_i \end{bmatrix} = \begin{bmatrix} A & 0 \\ 0 & D \end{bmatrix} \begin{bmatrix} x_i \\ y_i \end{bmatrix} = \begin{bmatrix} A a_i \\ D b_i \end{bmatrix}$$

so x_i is either zero or an eigenvector of A, and y_i is either zero or an eigenvector of D. If A acts on vectors in W, then we can define the projection map via

$$\pi : V \to W \quad v_i \mapsto x_i$$

and extending linearly. This map is surjective (since for any $w \in W$, we have $[w \;\; 0]^t \in V$ a linear combination of the v_i's) and so the spanning $v_1, \ldots, v_{n+m}$ must map to a spanning set of W. Thus we have $W = \operatorname{span}\{x_1, \ldots, x_{n+m}\}$. Now removing all zero vectors and linearly dependent vectors will form an eigenbasis for A, which by Prop 4.4.8(b) means A is diagonalizable. The same argument projecting to the y_i's shows that D is also diagonalizable.

$\Longleftarrow$: Suppose that A and D are diagonalizable, i.e. there exists P, Q such that $P^{-1}AP$ and $Q^{-1}DQ$ are diagonal. Now note that

$$\begin{bmatrix} P^{-1} & 0 \\ 0 & Q^{-1} \end{bmatrix} \begin{bmatrix} A & 0 \\ 0 & D \end{bmatrix} \begin{bmatrix} P & 0 \\ 0 & Q \end{bmatrix} = \begin{bmatrix} P^{-1}AP & 0 \\ 0 & Q^{-1}DQ \end{bmatrix}$$

is diagonal and that

$$\begin{bmatrix} P & 0 \\ 0 & Q \end{bmatrix} \begin{bmatrix} P^{-1} & 0 \\ 0 & Q^{-1} \end{bmatrix} = \begin{bmatrix} PP^{-1} & 0 \\ 0 & QQ^{-1} \end{bmatrix} = I \implies \begin{bmatrix} P & 0 \\ 0 & Q \end{bmatrix}^{-1} = \begin{bmatrix} P^{-1} & 0 \\ 0 & Q^{-1} \end{bmatrix}$$

So for $R = \begin{bmatrix} P & 0 \\ 0 & Q \end{bmatrix}$, we have $R^{-1}MR$ is diagonal and therefore M is diagonalizable. □

6.11

Let $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ be a 2×2 matrix with eigenvalue λ.

(a) Show that unless it is zero, the vector $(b, \lambda - a)^t$ is an eigenvector.

(b) Find a matrix P such that $P^{-1}AP$ is diagonal, assuming that $b \neq 0$ and that A has distinct eigenvalues.

Solution.

(a) *Proof.*
Let μ be the other complex eigenvalue of A (possibly $\mu = \lambda$). Then we have

$$\begin{cases} \lambda + \mu & = \operatorname{tr} A = a + d \\ \lambda\mu & = \det A = ad - bc \end{cases}$$

Now

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} b \\ \lambda - a \end{bmatrix} = \begin{bmatrix} ab + \lambda b - ab \\ bc + \lambda d - ad \end{bmatrix} = \begin{bmatrix} \lambda b \\ \lambda d - (ad - bc) \end{bmatrix} = \begin{bmatrix} \lambda b \\ \lambda d - \lambda\mu \end{bmatrix} = \lambda \begin{bmatrix} b \\ d - \mu \end{bmatrix} = \lambda \begin{bmatrix} b \\ \lambda - a \end{bmatrix}$$

Therefore $[b \quad \lambda - a]^t$ is an eigenvector with eigenvalue λ. □

(b) From (a), we have eigenvectors $[b \quad \lambda - a]^t$ and $[b \quad \mu - a]^t$. Thus we have an eigenbasis with basechange matrix

$$P = \begin{bmatrix} b & b \\ \lambda - a & \mu - a \end{bmatrix} \quad \text{and} \quad P^{-1} = \frac{1}{b(\mu - \lambda)} \begin{bmatrix} \mu - a & -b \\ a - \lambda & b \end{bmatrix}$$

Note that P is invertible since $b \neq 0$ and $\lambda \neq \mu$ by assumption. Now indeed

$$\begin{aligned} P^{-1}AP &= \frac{1}{b(\mu - \lambda)} \begin{bmatrix} \mu - a & -b \\ a - \lambda & b \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} b & b \\ \lambda - a & \mu - a \end{bmatrix} \\ &= \frac{1}{b(\mu - \lambda)} \begin{bmatrix} \mu - a & -b \\ a - \lambda & b \end{bmatrix} \begin{bmatrix} \lambda b & \mu b \\ \lambda(\lambda - a) & \mu(\mu - a) \end{bmatrix} \\ &= \frac{1}{b(\mu - \lambda)} \begin{bmatrix} \lambda\mu b - \lambda^2 b & 0 \\ 0 & -\lambda\mu b + \mu^2 b \end{bmatrix} \\ &= \frac{1}{b(\mu - \lambda)} \begin{bmatrix} \lambda b(\mu - \lambda) & 0 \\ 0 & \mu b(\mu - \lambda) \end{bmatrix} \\ &= \begin{bmatrix} \lambda & 0 \\ 0 & \mu \end{bmatrix} \end{aligned}$$

N.B. I use the convention of $k \times k$ Jordan blocks being written as

$$J_k(\lambda) := \begin{bmatrix} \lambda & 1 & & \\ & \lambda & \ddots & \\ & & \ddots & 1 \\ & & & \lambda \end{bmatrix}$$

i.e. the transpose of Artin's J_λ, and the Jordan form being written as

$$J = \begin{bmatrix} J_{k_1}(\lambda_1) & & \\ & \ddots & \\ & & J_{k_\ell}(\lambda_\ell) \end{bmatrix} =: J_{k_1}(\lambda_1) \oplus \cdots \oplus J_{k_\ell}(\lambda_\ell)$$

I believe this is the more common practice, and the results in Artin about Jordan blocks and Jordan form—as written like this—still hold, as the only difference in the proofs is how the basis gets ordered, see Exercise 7.8.

7.1

Determine the Jordan form of the matrix $\begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}$.

Solution.
Note that we have characteristic polynomial

$$p(t) = \det \begin{bmatrix} t-1 & -1 & 0 \\ 0 & t-1 & 0 \\ 0 & -1 & t-1 \end{bmatrix} = (t-1)^3$$

Hence we have eigenvalue $\lambda = 1$ with multiplicity 3. Next note that

$$\operatorname{im}(A - \lambda I) = \operatorname{im} \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} = \operatorname{span} \left\{ \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \right\}$$

and so by rank-nullity we have $\ker(A - \lambda I)$ has dimension $3 - 1 = 2$, i.e. the space of eigenvectors with eigenvalue $\lambda = 1$ is two-dimensional, i.e. $\lambda = 1$ has two Jordan blocks. Since $\lambda = 1$ is our only eigenvalue, our 3×3 Jordan matrix J must consist of these two blocks. This forces

$$J = J_2(1) \oplus J_1(1) = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

7.2

Prove that $A = \begin{bmatrix} 1 & 1 & 1 \\ -1 & -1 & -1 \\ 1 & 1 & 1 \end{bmatrix}$ is an *idempotent* matrix, i.e., that $A^2 = A$, and find its Jordan form.

Solution.

Proof.
We can prove this via a direct computation of A^2, but note that

$$A = \begin{bmatrix} 1 & 1 & 1 \\ -1 & -1 & -1 \\ 1 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \end{bmatrix} =: vw^t$$

Hence

$$A^2 = (vw^t)(vw^t) = v(w^t v)w^t = v(1)w^t = vw^t = A$$

□

To find its Jordan form, we have characteristic polynomial

$$p(t) = \det \begin{bmatrix} t-1 & -1 & -1 \\ 1 & t+1 & 1 \\ -1 & -1 & t-1 \end{bmatrix} = t^2(t-1)$$

Hence we have eigenvalues $\lambda_1 = 1$ with multiplicity 1 and $\lambda_2 = 0$ with multiplicity 2. Now note that

$$\operatorname{im}(A - \lambda_2 I) = \operatorname{im} A = \operatorname{span} \left\{ \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} \right\}$$

and so by rank-nullity $\ker(A - \lambda_2 I)$ has dimension $3 - 1 = 2$ and $\lambda_2 = 0$ has two Jordan blocks. Since we need a 1×1 Jordan block for $\lambda_1 = 1$ this forces the Jordan form of A to be

$$J = J_1(1) \oplus J_1(0) \oplus J_1(0) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

7.3

Let V be a complex vector space of dimension 5, and let T be a linear operator on V whose characteristic polynomial is $(t-\lambda)^5$. Suppose that the rank of the operator $T-\lambda I$ is 2. What are the possible Jordan forms for T?

Solution.
By rank-nullity $\ker(T-\lambda I)$ has dimension $5-2=3$, so λ must have 3 Jordan blocks. Since the Jordan form is a 5×5 matrix and λ is our only eigenvalue (as the characteristic polynomial is $(t-\lambda)^5$)), we need to partition 5 into a sum of three positive integers, which can only be done by

$$5 = 3+1+1 = 2+2+1$$

Hence the possible Jordan forms for T is either

$$J_3(\lambda)\oplus J_1(\lambda)\oplus J_1(\lambda) = \begin{bmatrix} \lambda & 1 & & & \\ & \lambda & 1 & & \\ & & \lambda & & \\ & & & \lambda & \\ & & & & \lambda \end{bmatrix}$$

or

$$J_2(\lambda)\oplus J_2(\lambda)\oplus J_1(\lambda) = \begin{bmatrix} \lambda & 1 & & & \\ & \lambda & & & \\ & & \lambda & 1 & \\ & & & \lambda & \\ & & & & \lambda \end{bmatrix}$$

7.4

(a) Determine all possible Jordan forms for a matrix whose characteristic polynomial is $(t+2)^2(t-5)^3$.

(b) What are the possible Jordan forms for a matrix whose characteristic polynomial is $(t+2)^2(t-5)^3$, when space of eigenvectors with eigenvalue -2 is one-dimensional, and the space of eigenvectors with eigenvalue 5 is two-dimensional?

Solution.

(a) All that we are given is that we have eigenvalues $\lambda_1 = -2$ with multiplicity 2 and $\lambda_2 = 5$ with multiplicity 3. Then for λ_1 we can have either 1 or 2 Jordan blocks and for λ_2 we can have either 1, 2, or 3 Jordan blocks, giving six possible Jordan forms in total (up to reordering); namely:

1. $J_2(-2) \oplus J_3(5)$
2. $J_1(-2) \oplus J_1(-2) \oplus J_3(5)$
3. $J_2(-2) \oplus J_2(5) \oplus J_1(5)$
4. $J_1(-2) \oplus J_1(-2) \oplus J_2(5) \oplus J_1(5)$
5. $J_2(-2) \oplus J_1(5) \oplus J_1(5) \oplus J_1(5)$
6. $J_1(-2) \oplus J_1(-2) \oplus J_1(5) \oplus J_1(5) \oplus J_1(5)$

(b) If $\dim\ker(T - \lambda_1 I) = 1$ and $\dim\ker(T - \lambda_2 I) = 2$, then this forces $\lambda_1 = -2$ to have 1 Jordan block and $\lambda_2 = 5$ to have 2 Jordan blocks. This then forces the Jordan form to be $J_2(-2) \oplus J_2(5) \oplus J_1(5)$.

7.5

What is the Jordan form of a matrix A all of whose eigenvectors are multiples of a single vector?

Solution.
Suppose every eigenvector is in span $\{v\}$ for some vector v. If w is an eigenvector with eigenvalue λ, then we can write $w = kv$ and note that

$$A(kv) = Aw = \lambda w = \lambda(kv) \implies Av = \lambda v$$

so v is an eigenvector with eigenvalue λ, and for any eigenvector $w' = \ell v$, note that

$$Aw' = A(\ell v) = \ell(Av) = \ell(\lambda v) = \lambda(\ell v) = \lambda w'$$

Thus every eigenvector of A has eigenvalue λ. If A is $n \times n$, this means that it has characteristic polynomial $p(t) = (t - \lambda)^n$. Furthermore, we have $\ker(A - \lambda I) = \text{span}\,\{v\}$ is one-dimensional, so λ must have 1 Jordan block and the Jordan form of A must be simply $J_n(\lambda)$.

7.6

Determine all invariant subspaces of a linear operator whose Jordan form consists of one block.

Solution.
Let $T : V \to V$ be a linear operator with Jordan form $J_n(\lambda)$. We can then find a basis $b_1, \ldots, b_n$ of V so that $Tb_1 = \lambda b_1$ and $Tb_i = \lambda b_i + b_{i-1}$ for $2 \leq i \leq n$.
(This is precisely the vector space formulation of saying that the matrix representation of T is the Jordan block $J_n(\lambda)$)
Note that

$$\begin{cases} Tb_1 = \lambda b_1 \\ Tb_i = \lambda b_i + b_{i-1} \end{cases} \implies \begin{cases} (T - \lambda I)b_1 = 0 \\ (T - \lambda I)b_i = b_{i-1} \end{cases} \implies \begin{cases} (T - \lambda I)^n = 0, \ (T - \lambda I)^{n-1} \neq 0 \\ \text{span}\,\{b_1, \ldots, b_k\} = \ker(T - \lambda I)^k \end{cases}$$

Now define

$$V_k := \ker(T - \lambda I)^k = \text{span}\,\{b_1, \ldots, b_k\}$$

and set $V_0 = \{0\}$. We claim $V_0, V_1, \ldots, V_n = V$ are all the invariant subspaces of T.

- First, V_k is T-invariant since for any $v \in V_k = \text{span}\,\{b_1, \ldots, b_k\}$,

$$\begin{aligned} T(v) &= T(\alpha_1 b_1 + \alpha_2 b_2 + \cdots + \alpha_k b_k) \\ &= \alpha_1 T(b_1) + \alpha_2 T(b_2) + \cdots + \alpha_k T(b_k) \\ &= \alpha_1(\lambda b_1) + \alpha_2(\lambda b_2 + b_1) + \cdots + \alpha_k(\lambda b_k + b_{k-1}) \in \text{span}\,\{b_1, \ldots, b_k\} = V_k \end{aligned}$$

- Next, let W be a nonzero T-invariant subspace and let $k > 0$ be the smallest number such that for every $w \in W$ we have $w = \alpha_1 b_1 + \cdots + \alpha_k b_k$, i.e. $W \subset V_k$ and $W \not\subset V_{k-1}$. In particular there then exists $w' \in W$ such that $w' = \alpha_1 b_1 + \cdots + \alpha_k b_k$ and $\alpha_k \neq 0$. However $b_1, \ldots, b_{k-1} \in \ker(T - \lambda I)^{k-1}$ means that

$$(T - \lambda I)^{k-1} w' = (T - \lambda I)^{k-1}(\alpha_1 b_1 + \cdots + \alpha_k b_k) = (T - \lambda I)^{k-1}(\alpha_k b_k) = \alpha_k b_1$$

  Furthermore, note for every $w \in W$ that

$$(T - \lambda I)w = T(w) - \lambda w \in W$$

  since $T(w) \in W$ by assumption. Thus W is also $(T - \lambda I)$-invariant, and so we have $(T - \lambda I)^{k-1} w' = \alpha_k b_1 \in W$ which implies $b_1 \in W$. A similar process shows $(T - \lambda I)^{k-2} w'$ is in both W and $\text{span}\,\{b_1, b_2\}$, which independence then forces $b_2 \in W$. Continuing inductively, we get $b_1, \ldots, b_k \in W$ and so $V_k = \text{span}\,\{b_1, \ldots, b_k\} \subset W$ and $W = V_k$.

7.7

Is every complex square matrix A such that $A^2 = A$ diagonalizable?

Solution.
We claim yes.

Proof.
Let λ be an eigenvalue of A and let v be an eigenvector. Then note

$$Av = \lambda v \implies A(Av) = \begin{cases} A(\lambda v) = \lambda^2 v \\ A^2 v = Av = \lambda v \end{cases} \implies \lambda^2 v = \lambda v \implies (\lambda^2 - \lambda)v = 0 \implies \lambda(\lambda - 1) = 0$$

Thus all the eigenvalues of A are either 0 or 1.
Now let v be a generalized eigenvector of A, i.e. $(A - \lambda I)^k v = 0$. By Corollary 4.7.13, it suffices to show that v is an eigenvector. If $k = 1$, we are done. Otherwise, $k \geq 2$ and we consider cases:

- $\lambda = 0$: Then $A - \lambda I = A$ and

$$0 = A^k v = A^{k-2}(A^2 v) = A^{k-2}(Av) = A^{k-1} v$$

  Repeating $(k-1)$ times then gives $Av = 0$ and v is an eigenvector with eigenvalue 0.

- $\lambda = 1$: Then $A - \lambda I = A - I$ and

$$\begin{aligned} 0 &= (A - I)^k v \\ &= (A - I)^{k-2}(A - I)^2 v \\ &= (A - I)^{k-2}(A^2 - 2A + I)v \\ &= (A - I)^{k-2}(A - 2A + I)v \\ &= (A - I)^{k-2}(I - A)v \\ &= -(A - I)^{k-2}(A - I)v \\ &= -(A - I)^{k-1} v \end{aligned}$$

  Repeating $(k-1)$ times then gives $(A - I)v = \pm 0 = 0$ and so v is an eigenvector.

Therefore in both cases v is an eigenvector and so A is diagonalizable. $\square$

7.8

Is every complex square matrix A similar to its transpose?

Solution.
We claim yes.

Proof.
Note that

$$B_k^{-1}J_k(\lambda)B_k := \begin{bmatrix} 0 & \dots & 1 \\ \vdots & \iddots & \vdots \\ 1 & \dots & 0 \end{bmatrix}^{-1} \begin{bmatrix} \lambda & 1 & & \\ & \lambda & \ddots & \\ & & \ddots & 1 \\ & & & \lambda \end{bmatrix} \begin{bmatrix} 0 & \dots & 1 \\ \vdots & \iddots & \vdots \\ 1 & \dots & 0 \end{bmatrix} = \begin{bmatrix} \lambda & & & \\ 1 & \ddots & & \\ & \ddots & \lambda & \\ & & 1 & \lambda \end{bmatrix}$$

i.e. a Jordan block $J_k(\lambda)$ is similar to its transpose. This means—via using a block matrix with these B_k's as blocks—that any matrix in Jordan form

$$J = \begin{bmatrix} J_{k_1}(\lambda_1) & & \\ & \ddots & \\ & & J_{k_\ell}(\lambda_\ell) \end{bmatrix}$$

must also be similar to its transpose J^t, which is a form using $J_k(\lambda)^t$ as its blocks (i.e. what Artin calls "Jordan form"; in particular note that this justifies our convention from the top of the chapter to be able to use Artin's results).

Finally, putting A into Jordan form J is simply a change of basis, so we have $J = P^{-1}AP$. Then

$$J^t = (P^{-1}AP)^t = P^tA^t(P^t)^{-1}$$

and so A^t is similar to J^t. Combining this all together, we have A similar to J, J similar to J^t, and J^t similar to A^t. Therefore by transitivity A is similar to A^t. □

7.9

Find a 2×2 matrix with entries in $\mathbb{F}_p$ that has a power equal to the identity and an eigenvalue in $\mathbb{F}_p$, but is not diagonalizable.

Solution.
Consider the 2×2 matrix in $\mathbb{F}_p$

$$J_2(1) = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} =: \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} = I + N$$

Then note that via $N^2 = 0$ and binomial expansion we have

$$J_2(1)^p = (I + N)^p = \sum_{i=0}^{p} \binom{p}{i} N^i I^{p-i} = \binom{p}{0} I + \binom{p}{1} N + 0 = I + pN$$

but since we are in $\mathbb{F}_p$ we have $pN \equiv 0$ and so $J_2(1)^p \equiv I$.

Furthermore, $J_2(1)$ has characteristic polynomial $p(t) = (t-1)^2$, and thus it has eigenvalue $1 \in \mathbb{F}_p$.

However, by Corollary 4.7.13 $J_2(1)$ is not diagonalizable since it is already in Jordan form and is a 2×2 block.

M.1

Let $v = (a_1, \dots, a_n)$ be a real row vector. We may form the $n! \times n$ matrix M whose rows are obtained by permuting the entries of v in all possible ways. The rows can be listed in an arbitrary order. Thus if $n = 3$, M might be

$$\begin{bmatrix} a_1 & a_2 & a_3 \\ a_1 & a_3 & a_2 \\ a_2 & a_3 & a_1 \\ a_2 & a_1 & a_3 \\ a_3 & a_1 & a_2 \\ a_3 & a_2 & a_1 \end{bmatrix}$$

Determine the possible ranks that such a matrix could have.

Solution.
Note that each row must have the same sum $s := a_1 + \cdots + a_n$. Furthermore, the rank of M is the same as the rank of M^t by Thm 4.2.14 so it suffices to investigate the independence of the rows of M (or equivalently, work with M^t). Consider the collection for any $x \in \mathbb{R}$

$$W_x = \{w \in \mathbb{R}^n \mid w_1 + \cdots + w_n = x\}$$

Note that W_0 is an $(n-1)$-dimensional subspace, and that $\operatorname{im} M^t \subset \operatorname{span}\{W_s\}$.
Now we consider cases:

- If v is a constant vector $a_1 = \cdots = a_n = a$ for some $a \in \mathbb{R}$, then every row of M is simply v. Thus we have M rank 0 if $a = 0$ and M rank 1 if $a \neq 0$.

- If v is not constant, then we claim that M has rank $n-1$ or n. To see this, as v is not constant and the rows of M contain all permutations of v, there exists $a_p \neq a_q$ and rows r, r' of M such that we can go from r to r' by simply swapping a_p and a_q, e.g.

$$r = [a_1 \quad \cdots \quad a_p \quad \cdots \quad a_q \quad \cdots \quad a_n]$$

$$r' = [a_1 \quad \cdots \quad a_q \quad \cdots \quad a_p \quad \cdots \quad a_n]$$

However, in this case now we have

$$r - r' = [0 \quad \cdots \quad a_p - a_q \quad \cdots \quad -(a_p - a_q) \quad \cdots \quad 0] = (a_p - a_q)[0 \quad \cdots \quad 1 \quad \cdots \quad -1 \quad \cdots \quad 0]$$

and then thinking of these rows as vectors in $\mathbb{R}^n$, we can write $r - r' = (a_p - a_q)(e_i - e_j)$ (if a_p is the ith entry of r and the jth entry of r') and we have

$$\begin{cases} a_p - a_q \neq 0 \\ r, r' \in \operatorname{im} M^t \end{cases} \implies e_i - e_j \in \operatorname{im} M^t$$

But since M contains all permutations of v's entries, we can find rows r and r' as above where the a_p and a_q are anywhere we would like. More formally, for any $i \neq j$ we can find rows r and r' such that $r_i = r'_j = a_p$, $r_j = r'_i = a_q$, and $r_k = r'_k$ for $k \neq i, j$.

The above reasoning then shows that $e_i - e_j \in \operatorname{im} M^t$ for any $i \neq j$. And since the $(e_i - e_j)$'s generate W_0, we have that

$$W_0 \subset \operatorname{im} M^t \implies n - 1 = \dim W_0 \leq \dim(\operatorname{im} M^t) = \operatorname{rank}(M)$$

Finally, if $s = 0$, then $\operatorname{im} M^t \subset \operatorname{span}\{W_s\}$ forces $\operatorname{im} M^t = W_0$ and so M has rank $n - 1$. If $s \neq 0$, then we have $v \notin W_0$ and thus W_0 is a strict subset of $\operatorname{im} M^t$ and $\operatorname{rank}(M) > n-1$. But M having n columns forces $\operatorname{rank}(M) \leq n$, so we must have $\operatorname{rank}(M) = n$.

Therefore the only relevant factors are whether $a_1 = \cdots = a_n$ and $a_1 + \cdots + a_n = 0$, or in terms of the ones vector $\mathbf{1} \in \mathbb{R}^n$, whether $v = a\mathbf{1}$ and $v^t\mathbf{1} = 0$. In summary:

- If $v = 0\mathbf{1}$, then $\operatorname{rank}(M) = 0$.
- If $v = a\mathbf{1}$ with $a \neq 0$, then $\operatorname{rank}(M) = 1$.
- If $v \neq a\mathbf{1}$ and $v^t\mathbf{1} = 0$, then $\operatorname{rank}(M) = n - 1$.
- If $v \neq a\mathbf{1}$ and $v^t\mathbf{1} \neq 0$, then $\operatorname{rank}(M) = n$.

This exhausts all possibilities and therefore are all of the possible ranks of M.

M.2

Let A be a complex $n \times n$ matrix with n distinct eigenvalues $\lambda_1, \dots, \lambda_n$. Assume that λ_1 is the largest eigenvalue, that is, that $|\lambda_1| > |\lambda_i|$ for all $i > 1$.

(a) Prove that for most vectors X, the sequence $X_k = \lambda_1^{-k} A^k X$ converges to an eigenvector Y with eigenvalue λ_1, and describe precisely what the conditions on X are for this to be true.

(b) Prove the same thing without assuming that the eigenvalues $\lambda_1, \dots, \lambda_n$ are distinct.

Solution.

(a) *Proof.*
Since the eigenvalues are distinct, we can find an eigenbasis $b_1, \dots, b_n$ (such that $Ab_i = \lambda_i b_i$), so we can write

$$X = \alpha_1 b_1 + \cdots + \alpha_n b_n$$

Now note each term of our sequence looks like

$$\begin{aligned} X_k &= \frac{1}{\lambda_1^k} A^k X \\ &= \frac{1}{\lambda_1^k} A^k (\alpha_1 b_1 + \alpha_2 b_2 + \cdots + \alpha_n b_n) \\ &= \frac{1}{\lambda_1^k} (\alpha_1 A^k b_1 + \alpha_2 A^k b_2 + \cdots + \alpha_n A^k b_n) \\ &= \frac{1}{\lambda_1^k} (\alpha_1 \lambda_1^k b_1 + \alpha_2 \lambda_2^k b_2 + \cdots + \alpha_n \lambda_n^k b_n) \\ &= \frac{1}{\lambda_1^k} \lambda_1^k (\alpha_1 b_1 + \alpha_2 \tfrac{\lambda_2^k}{\lambda_1^k} b_2 + \cdots + \alpha_n \tfrac{\lambda_n^k}{\lambda_1^k} b_n) \\ &= \alpha_1 b_1 + \alpha_2 (\tfrac{\lambda_2}{\lambda_1})^k b_2 + \cdots + \alpha_n (\tfrac{\lambda_n}{\lambda_1})^k b_n \end{aligned}$$

and for every $i > 1$ we have

$$|\lambda_1| > |\lambda_i| \implies \left| \frac{\lambda_i}{\lambda_1} \right| < 1 \implies \lim_{k\to\infty} \left| \frac{\lambda_i}{\lambda_1} \right|^k = 0 \implies \lim_{k\to\infty} \left(\frac{\lambda_i}{\lambda_1} \right)^k = 0$$

Thus we have $X_k \to \alpha_1 b_1$ as $k \to \infty$, and indeed $\alpha_1 b_1$ is an eigenvector with eigenvalue λ_1 as long as α_1 is nonzero. Hence the sequence converges to an eigenvector if $\alpha_1 \neq 0$. ☐

(b) Note we still have $|\lambda_1| > |\lambda_i|$ for $i > 1$, so we can put A into Jordan form $J = PAP^{-1}$ to look like

$$J = \begin{bmatrix} \lambda_1 & & & \\ & J_{k_2}(\mu_2) & & \\ & & \ddots & \\ & & & J_{k_\ell}(\mu_\ell) \end{bmatrix}$$

where for each Jordan block, $\mu_i \in \{\lambda_2, \ldots, \lambda_n\}$. Now (from P) we have a basis of generalized eigenvectors $b_1, \ldots, b_n$ such that b_1 is an eigenvector with λ_1. We then write

$$X = \alpha_1 b_1 + \cdots + \alpha_n b_n$$

and can write this as a vector v with respect to our basis $b_1, \ldots, b_n$ and with blocks:

$$v = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix} =: \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_\ell \end{bmatrix}$$

where each v_i is of length k_i (and $v_1 = \alpha_1$) in order to perform block multiplication with J. Also note that $J^n = PA^nP^{-1}$, so we have A^n acting on X is the same as J^n acting on v and

$$X_n = \frac{1}{\lambda_1^n} A^n X = \frac{1}{\lambda_1^n} J^n v = \frac{1}{\lambda_1^n} \begin{bmatrix} \lambda_1^n & & & \\ & J_{k_2}(\mu_2)^n & & \\ & & \ddots & \\ & & & J_{k_\ell}(\mu_\ell)^n \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_\ell \end{bmatrix} = \frac{1}{\lambda_1^n} \begin{bmatrix} \lambda_1^n v_1 \\ J_{k_2}(\mu_2)^n v_2 \\ \vdots \\ J_{k_\ell}(\mu_\ell)^n v_\ell \end{bmatrix}$$

Thus it suffices to show for $i > 1$ that $\lambda_1^{-n} J_{k_i}(\mu_i)^n v_i \to 0$ as $n \to \infty$, i.e. work each block separately. Choose $i > 1$ and, to clean up notation, set $\mu := \mu_i$ and $J_{k_i}(\mu_i) =: J_k(\mu) = \mu I + N$. If $\mu = 0$ then we are done as $J_k(0) = N$ is nilpotent. Otherwise, we have via binomial expansion

$$\begin{aligned} \lambda_1^{-n} J_k(\mu)^n = \frac{1}{\lambda_1^n}(\mu I + N)^n &= \frac{1}{\lambda_1^n} \sum_{j=0}^{n} \binom{n}{j} \mu^{n-j} N^j \\ &\overset{(\star)}{=} \frac{1}{\lambda_1^n} \sum_{j=0}^{k} \binom{n}{j} \mu^{n-j} N^j \\ &= \sum_{j=0}^{k} \binom{n}{j} \mu^{-j} \left(\frac{\mu}{\lambda_1}\right)^n N^j \end{aligned}$$

where $(\star)$ follows from $N^j = 0$ for $j \geq k$ (the size of our Jordan block $J_k(\mu)$), and since we are sending $n \to \infty$ we may assume that $n \geq k$ and so our sum is finite. Note that

$$|\lambda_1| > |\mu| \implies \left|\frac{\mu}{\lambda_1}\right| < 1 \implies \lim_{n\to\infty} \left|\frac{\mu}{\lambda_1}\right|^n = 0 \implies \lim_{n\to\infty} \left(\frac{\mu}{\lambda_1}\right)^n = 0$$

and since $\binom{n}{j}$ is a polynomial in n of degree j while $(\mu/\lambda_1)^n$ is exponential, we have that

$$\lim_{n\to\infty} \binom{n}{j} \left(\frac{\mu}{\lambda_1}\right)^n = 0$$

Thus

$$\lim_{n\to\infty} \lambda_1^{-n} J_k(\mu)^n = \lim_{n\to\infty} \sum_{j=0}^{k} \binom{n}{j} \mu^{-j} \left(\frac{\mu}{\lambda_1}\right)^n N^j = \sum_{j=0}^{k} \mu^{-j} 0 N^j = 0$$

and $\lambda_1^{-n} J_{k_i}(\mu_i)^n v_i \to 0$ as $n \to \infty$. Therefore

$$\lim_{n\to\infty} X_n = \lim_{n\to\infty} \frac{1}{\lambda_1^n} \begin{bmatrix} \lambda_1^n v_1 \\ J_{k_2}(\mu_2)^n v_2 \\ \vdots \\ J_{k_\ell}(\mu_\ell)^n v_\ell \end{bmatrix} = \begin{bmatrix} v_1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \alpha_1 b_1 + 0 b_2 + \cdots + 0 b_n$$

and $X_n \to \alpha_1 b_1$ as $n \to \infty$, which is an eigenvector with eigenvalue λ_1 for $\alpha_1 \neq 0$.

M.3

Compute the largest eigenvalue of the matrix $\begin{bmatrix} 3 & 1 \\ 3 & 4 \end{bmatrix}$ to three-place accuracy, using a method based on Exercise M.2.

Solution.
Note that the sequence in Exercise M.2 of $X_k = \lambda_1^{-k} A^k X$ requires knowledge of the largest eigenvalue λ_1. However, the sequence $Y_k = A^k X = AY_{k-1}$ will still tend towards the direction of an eigenvector, but also increase/decrease in magnitude exponentially depending on whether $|\lambda_1|$ is greater/less than 1. Hence if we keep normalizing the vector after multiplying by A, we will tend towards a unit eigenvector Z with eigenvalue λ_1. Our complete method then is to first choose an arbitrary vector Z_0 and then compute the sequence of vectors

$$Z_k = \frac{AZ_{k-1}}{\|AZ_{k-1}\|}$$

Now note that

$$Z_k \overset{k\to\infty}{\longrightarrow} Z \implies \|AZ_k\| \overset{k\to\infty}{\longrightarrow} \|AZ\| = \|\lambda_1 Z\| = |\lambda_1|\|Z\| = |\lambda_1|$$

For our example, we have $A = \begin{bmatrix} 3 & 1 \\ 3 & 4 \end{bmatrix}$ and we choose $Z_0 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$.
Now with the help of a Python program we compute

$$\begin{aligned}
\|AZ_0\| &\approx 4.2426 \\
\|AZ_1\| &\approx 5.7008 \\
\|AZ_2\| &\approx 5.4926 \\
\|AZ_3\| &\approx 5.3675 \\
\|AZ_4\| &\approx 5.3238 \\
\|AZ_5\| &\approx 5.3095 \\
\|AZ_6\| &\approx 5.3049 \\
\|AZ_7\| &\approx 5.3034 \\
\|AZ_8\| &\approx 5.3029 \\
\|AZ_9\| &\approx 5.3027 \\
\|AZ_{10}\| &\approx 5.3027
\end{aligned}$$

Thus after the fourth decimal place stabilizes, we can confirm the first three decimal places are correct and so $|\lambda_1| \approx 5.3027$. Indeed, solving the characteristic polynomial directly gives

$$\lambda_1 = \frac{7 + \sqrt{13}}{2} \approx 5.30277538$$

M.4

If $X = (x_1, x_2, \dots)$ is an infinite real row vector and $A = (a_{ij}), 0 < i, j < \infty$ is an infinite real matrix, one may or may not be able to define the matrix product XA. For which A can one define right multiplication on the space $\mathbb{R}^\infty$ of all infinite row vectors? On the space

$$Z = \{(a) \in \mathbb{R}^\infty \mid a_n = 0 \text{ for all but finitely many } n\}?$$

Solution.
We can define XA for all $X \in \mathbb{R}^\infty$ as long as each column of A has finitely-many nonzero entries, and we can define XA for all $X \in Z$ for any A.

M.5

Let $\varphi : F^n \to F^m$ be left multiplication by an $m \times n$ matrix A.

(a) Prove that the following are equivalent:

- A has a right inverse, a matrix B such that $AB = I$.
- φ is surjective.
- The rank of A is m.

(b) Prove that the following are equivalent:

- A has a left inverse, a matrix B such that $BA = I$.
- φ is injective.
- The rank of A is n.

Solution.

(a) *Proof.*

(1) $\implies$ (2): Suppose that $AB = I$. Then for any $y \in F^m$, we have

$$\varphi(By) = A(By) = (AB)y = Iy = y \implies y \in \operatorname{im}\varphi$$

and thus φ is surjective.

(2) $\implies$ (3): Suppose that φ is surjective. Then $\operatorname{im}\varphi = F^m$ and

$$\operatorname{rank}(A) = \dim(\operatorname{im} A) = \dim(\operatorname{im}\varphi) = \dim(F^m) = m$$

(3) $\implies$ (1): Suppose that $\operatorname{rank}(A) = m$. Then $\operatorname{im} A \subset F^m$ and $\dim(\operatorname{im} A) = \dim(F^m) = m$ forces $\operatorname{im} A = F^m$ by Prop 3.4.23. Hence for every standard basis vector $e_i \in F^m$, there exists $b_i \in F^n$ such that $Ab_i = e_i$. Setting $B := [b_1 \quad \dots \quad b_m]$,

$$AB = A[b_1 \quad \dots \quad b_m] = [Ab_1 \quad \dots \quad Ab_m] = [e_1 \quad \dots \quad e_m] = I$$

and A has a right inverse. □

(b) *Proof.*

Define $\psi : F^m \to F^n$ by $\psi(y) = A^t y$. Then note

$$A \text{ has a left inverse} \iff BA = I \iff A^t B^t = I \iff A^t \text{ has a right inverse}$$

and

$$\varphi \text{ is injective} \iff \operatorname{null}(A) = \{0\} \overset{(\star)}{\iff} \operatorname{rank}(A^t) = n \iff \operatorname{im}(A^t) = F^n \iff \psi \text{ surjective}$$

and

$$\operatorname{rank}(A) = n \iff \operatorname{rank}(A^t) = n$$

Therefore the equivalence follows from (a) applied to the rightmost statements above.
To elaborate on $(\star)$, it follows from the dimension formula that $n = \operatorname{rank}(A) + \operatorname{nullity}(A)$, so

$$\operatorname{null}(A) = \{0\} \iff \operatorname{nullity}(A) = 0 \iff \operatorname{rank}(A^t) = \operatorname{rank}(A) = n$$

□

M.6

Without using the characteristic polynomial, prove that a linear operator on a vector space of dimension n can have at most n distinct eigenvalues.

Solution.

Proof.
Suppose otherwise, i.e. there exists a linear operator $T : V \to V$ with $\dim V = n$ such that T has distinct eigenvectors $\lambda_1, \ldots, \lambda_r$ for $r > n$. Then if $v_1, \ldots, v_r \in V$ are corresponding eigenvectors, by Prop 4.6.5 we have that $v_1, \ldots, v_r$ are linearly independent. However, if we take a basis $b_1, \ldots, b_n$ of V, then

$$r = |\{v_1, \ldots, v_r\}| > |\{b_1, \ldots, b_n\}| = n$$

which contradicts Prop 3.4.21(c). Therefore every linear operator $T : V \to V$ can have at most n distinct eigenvalues. □

M.7

Let T be a linear operator on a vector space V. Let K_r and W_r denote the kernel and image, respectively, of T^r.

(a) Show that $K_1 \subset K_2 \subset \dots$ and that $W_1 \supset W_2 \supset \dots$.

(b) The following conditions might or might not hold for a particular value of r:

$$(1)\ K_r = K_{r+1}, \quad (2)\ W_r = W_{r+1}, \quad (3)\ W_r \cap K_1 = \{0\}, \quad (4)\ W_1 + K_r = V$$

Find all implications among the conditions (1)–(4) when V is finite-dimensional.

(c) Do the same thing when V is infinite-dimensional.

Solution.

(a) *Proof.*
Choose $v \in K_r$. Then $T^r(v) = 0$ and

$$T^{r+1}(v) = T(T^r(v)) = T(0) = 0 \implies v \in K_{r+1}$$

Thus $K_r \subset K_{r+1}$.
Now choose $w \in W_r$. Then we have $T^r(v) = w$ for some $v \in V$. Note that

$$T^{r-1}(T(v)) = T^r(v) = w \implies w \in W_{r-1}$$

Thus $W_r \supset W_{r-1}$. □

(b) Let $\dim V = n$. We claim all four conditions are equivalent.

Proof.
(1) $\iff$ (2): For any r, $T^r : V \to V$ is a linear operator and so by rank-nullity we have

$$\dim K_r + \dim W_r = n \tag{$\star$}$$

Then we have equivalences

$$\begin{aligned} K_r = K_{r+1} &\iff \dim K_r = \dim K_{r+1} \\ &\overset{(\star)}{\iff} (n - \dim W_r) = (n - \dim W_{r+1}) \\ &\iff \dim W_r = \dim W_{r+1} \\ &\overset{(\dagger)}{\iff} W_r = W_{r+1} \end{aligned}$$

where $(\dagger)$ comes from (a) and Prop 3.4.23. Thus we have (1) $\iff$ (2).

(1) $\implies$ (3): Suppose that $K_r = K_{r+1}$. Choose $w \in W_r \cap K_1$. Then $T(w) = 0$ and there exists $v \in V$ such that $T^r(v) = w$. Now note

$$T^{r+1}(v) = T(T^r(v)) = T(w) = 0 \implies v \in K_{r+1} = K_r \implies w = T^r(v) = 0$$

Thus $W_r \cap K_1 = \{0\}$.

(3) $\implies$ (4): Suppose that $W_r \cap K_1 = \{0\}$. Choose $w \in W_r \cap K_r$. Then $T^r(w) = 0$ and there exists $v \in V$ such that $T^r(v) = w$. Now

$$\begin{aligned} \begin{cases} T(T^{r-1}(w)) = T^r(w) = 0 \\ T^r(T^{r-1}(v)) = T^{r-1}(T^r(v)) = T^{r-1}(w) \end{cases} &\implies T^{r-1}(w) \in W_r \cap K_1 = \{0\} \\ &\implies T^{r-1}(w) = 0 \\ &\implies w \in W_r \cap K_{r-1} \end{aligned}$$

The same argument shows $w \in W_r \cap K_{r-2}$ and so inductively we have

$$w \in W_r \cap K_1 = \{0\} \implies w = 0 \implies W_r \cap K_r = \{0\}$$

Thus by Prop 3.6.6(a),

$$n \overset{(\star)}{=} \dim W_r + \dim K_r = \dim(W_r \cap K_r) + \dim(W_r + K_r) = 0 + \dim(W_r + K_r)$$

and since $W_r \subset W_1$,

$$\begin{aligned} W_r + K_r \subset W_1 + K_r \subset V &\implies n = \dim(W_r + K_r) \leq \dim(W_1 + K_r) \leq \dim(V) = n \\ &\implies \dim(W_1 + K_r) = n \end{aligned}$$

and finally by Prop 3.4.23 we have $W_1 + K_r = V$.

(4) $\implies$ (2): Suppose that $W_1 + K_r = V$. Then it suffices to show that $W_r \subset W_{r+1}$. Choose $w \in W_r$, i.e. $w = T^r(v)$ for some $v \in V = W_1 + K_r$. Then there exists $T(v') \in W_1$ and $v'' \in K_r$ such that $v = T(v') + v''$. Now note

$$w = T^r(v) = T^r(T(v') + v'') = T^r(T(v')) + T^r(v'') = T^{r+1}(v') + 0 \implies w \in W_{r+1}$$

Hence $W_r \subset W_{r+1}$. □

(c) We claim (1) $\iff$ (3) and (2) $\iff$ (4) are the only implications.

Proof.
(1) $\implies$ (3): See proof in (b).
(3) $\implies$ (1): Suppose that $W_r \cap K_1 = \{0\}$. Assume that $K_r \neq K_{r+1}$. As $K_r \subset K_{r+1}$, this means that there exists $v \in K_{r+1}$ such that $v \notin K_r$. But now for $w = T^r(v) \in W_r$, note

$$T(w) = T(T^r(v)) = T^{r+1}(v) = 0 \implies w \in K_1 \cap W_r = \{0\} \implies w = T^r(v) = 0$$

Hence $v \in K_r$, which is a contradiction and so $K_r = K_{r+1}$.

(2) $\implies$ (4): Suppose that $W_r = W_{r+1}$. Choose $v \in V$. Note that $T^r(v) \in W_r = W_{r+1}$, so there exists $v' \in V$ such that $T^{r+1}(v') = T^r(v)$. Now

$$T^{r+1}(v') = T^r(v) \implies 0 = T^r(v - T(v')) \implies (v - T(v')) \in K_r$$

and since $T(v') \in W_1$, we have

$$v = T(v') + (v - T(v')) \in W_1 + K_r$$

Thus $V = W_1 + K_r$.
(4) $\implies$ (2): See proof in (b).

Finally, it suffices to show (1) $\not\Longrightarrow$ (2) and (2) $\not\Longrightarrow$ (1):
(1) $\not\Longrightarrow$ (2): Consider the right-shift operator

$$R : \mathbb{R}^\infty \to \mathbb{R}^\infty, \quad (a_1, a_2, a_3, \ldots,) \mapsto (0, a_1, a_2, \ldots,)$$

Note that $K_1 = K_2 = \{0\}$, so (1) holds for $r = 1$. However, (2) fails since

$$W_1 = \{(0, a_2, a_3, \ldots,)\} \neq \{(0, 0, a_3, \ldots,)\} = W_2$$

(2) $\not\Longrightarrow$ (1): Consider the left-shift operator

$$L : \mathbb{R}^\infty \to \mathbb{R}^\infty, \quad (a_1, a_2, a_3, \ldots,) \mapsto (a_2, a_3, \ldots,)$$

Note that $W_1 = W_2 = \mathbb{R}^\infty$, so (2) holds for $r = 1$. However, (1) fails since

$$K_1 = \{(a_1, 0, 0, \ldots,)\} \neq \{(a_1, a_2, 0, \ldots,)\} = K_2$$

□

M.8

Let T be a linear operator on a finite-dimensional complex vector space V.

(a) Let λ be an eigenvalue of T, and let V_λ be the set of generalized eigenvectors, together with the zero vector. Prove that V_λ is a T-invariant subspace of V. (This subspace is called a *generalized eigenspace.*)

(b) Prove that V is the direct sum of its generalized eigenspaces.

Solution.

(a) *Proof.*
Let $C = T - \lambda I$. Then from Exercise M.7(a) we have

$$\ker C \subset \ker C^2 \subset \dots$$

and since V is finite-dimensional, this chain must stabilize eventually, i.e. $\ker C^k = \ker C^{k+1}$ for some k. Thus

$$V_\lambda = \{v \in V \mid C^i(v) = 0 \text{ for some } i = 1, 2, \dots\} = \bigcup_{i \geq 1} \ker C^i = \ker C^k$$

and so V_λ is a subspace. Also note for any $v \in V$ that

$$C(T(v)) = T^2(v) - \lambda T(v) = T((T - \lambda I)(v)) = T(C(v))$$

Thus C and T commute. Now choose $v \in V_\lambda$, i.e. $(T - \lambda I)^k(v) = C^k(v) = 0$. Then

$$C^k(T(v)) = T(C^k(v)) = T(0) = 0 \implies T(v) \in V_\lambda$$

Thus V_λ is T-invariant. $\square$

(b) *Proof.*

We induct on $n := \dim V$.

If $n = 0$, no generalized eigenvectors exist and the result holds vacuously.

Now assume the result holds for all complex vector spaces of dimension $< n$.

Since V is a vector space over $\mathbb{C}$, T has an eigenvalue λ and $V_\lambda \neq \{0\}$. Then from the proof of (a), there exists k such that $\ker(T - \lambda I)^k = \ker(T - \lambda I)^{k+1}$. Now we have

$$V_\lambda = \ker(T - \lambda I)^k \quad \text{and} \quad W := \operatorname{im}(T - \lambda I)^k$$

Then from Exercise M.7(b) and its proof,

$$\ker(T - \lambda I)^k = \ker(T - \lambda I)^{k+1} \implies V_\lambda \cap \operatorname{im}(T - \lambda I) = \{0\} \implies V_\lambda \cap W = \{0\}$$

and by rank-nullity

$$n = \dim V = \dim V_\lambda + \dim W$$

Thus

$$\dim(V_\lambda + W) = (\dim V_\lambda + \dim W) - \dim(V_\lambda \cap W) = n - 0 = n \implies V_\lambda + W = V$$

and so by Prop 3.6.6(c) we have $V = V_\lambda \oplus W$.

Furthermore, if $W = \{0\}$ then we are done. Otherwise, note

$$(T - \lambda I)^k(v) \in W \implies T((T - \lambda I)^k(v)) = (T - \lambda I)^k(T(v)) \in W$$

i.e. W is T-invariant and

$$V_\lambda \neq \{0\} \implies \dim V_\lambda > 0 \implies \dim W = n - \dim V_\lambda < n$$

Thus the IH holds for $T|_W : W \to W$ and we can write

$$W = W_{\mu_1} \oplus \cdots \oplus W_{\mu_r}$$

We now need to show

- The distinct eigenvalues of T are $\lambda, \mu_1, \ldots, \mu_r$: First, for distinctness note that if $\lambda = \mu_i$, then there exists nonzero $w \in W$ such that

$$T(w) = T|_W(w) = \mu_i w = \lambda w \implies w \in V_\lambda \implies w \in V_\lambda \cap W = \{0\}$$

  which is a contradiction. Next, every eigenvalue of $T|_W$ is an eigenvalue of T, so we need to show that all of the eigenvalues of T are in the list above. Let $\mu \neq \lambda$ be an eigenvalue of T, i.e. $T(v) = \mu v$. We can write $v = v' + w$ for $v' \in V_\lambda$ and $w \in W$. Now note

$$\begin{aligned} V_\lambda \text{ and } W \text{ are } T\text{-invariant} &\implies V_\lambda \text{ and } W \text{ are } (T - \mu I)\text{-invariant} \\ &\implies (T - \mu I)v' \in V_\lambda \text{ and } (T - \mu I)w \in W \end{aligned}$$

  so we have

$$0 = (T - \mu I)v = (T - \mu I)v' + (T - \mu I)w \in V_\lambda + W$$

which by independence forces $(T-\mu I)v' = (T-\mu I)w = 0$. Specifically, $Tv' = \mu v'$ and more generally for any polynomial $p(x)$, we have that $p(T)v' = p(\mu)v'$ (e.g. $(T^2 - T + I)(v') = \mu^2 v' - \mu v' + v'$). Hence for the polynomial $p(x) = (x-\lambda)^k$, we have $p(T) = (T-\lambda I)^k$ and $p(\mu) = (\mu - \lambda)^k$. But

$$v' \in V_\lambda \implies 0 = (T-\lambda I)^k v' = p(T)v' = p(\mu)v' = (\mu-\lambda)^k v'$$

and $\mu \neq \lambda$ forces $v' = 0$. Thus $v = w \in W$ and so v is an eigenvector of T_W with eigenvalue μ, i.e. $\mu = \mu_i$ for some i.

- Each W_{μ_i} is a generalized eigenspace of V: We have $W_{\mu_i} \subset V_{\mu_i}$, so choose $v \in V_{\mu_i}$, i.e. $(T-\mu_i I)^\ell v = 0$. By the same argument as above, we can write $v = v' + w$ for $v' \in V_\lambda$ and $w \in W$ and

$$\begin{aligned} 0 = (T-\mu_i I)^\ell v = (T-\mu_i I)^\ell v' + (T-\mu_i I)^\ell w \in V_\lambda + W &\implies (T-\mu_i I)^\ell v' = 0 \\ &\implies v' = 0 \\ &\implies v = w \in W \\ &\implies (T|_W - \mu_i I)^\ell v = 0 \\ &\implies v \in \ker(T|_W - \mu_i I)^\ell \\ &\implies v \in W_{\mu_i} \end{aligned}$$

Thus $W_{\mu_i} = V_{\mu_i}$.

Therefore we have

$$V = V_\lambda \oplus W = V_\lambda \oplus (W_{\mu_1} \oplus \cdots \oplus W_{\mu_r}) = V_\lambda \oplus (V_{\mu_1} \oplus \cdots \oplus V_{\mu_r}) = V_\lambda \oplus V_{\mu_1} \oplus \cdots \oplus V_{\mu_r}$$

and V is the direct sum of its generalized eigenspaces. □

M.9

Let V be a finite-dimensional vector space. A linear operator $T : V \to V$ is called a *projection* if $T^2 = T$ (not necessarily an "orthogonal projection"). Let K and W be the kernel and image of a linear operator T. Prove

(a) T is a projection onto W if and only if the restriction of T to W is the identity map.

(b) If T is a projection, then V is the direct sum $W \oplus K$.

(c) The trace of a projection is equal to its rank.

Solution.

(a) *Proof.*
$\Longrightarrow$: Suppose T is a projection onto W. Now choose $w \in W$. Then there exists $v \in V$ such that $T(v) = w$. Now

$$T|_W(w) = T(w) = T(T(v)) = T^2(v) = T(v) = w$$

Thus $T|_W = I$.

$\Longleftarrow$: Suppose that $T|_W = I$. Now choose $v \in V$. Then

$$T(v) \in W \implies T^2(v) = T(T(v)) = T|_W(T(v)) = I(T(v)) = T(v)$$

Thus $T^2 = T$ and T is a projection onto W. ◻

(b) *Proof.*
Let T be a projection. Then

$$\operatorname{im} T^2 = \operatorname{im} T \overset{(\star)}{\implies} \operatorname{im} T \cap \ker T = \{0\} \text{ and } \operatorname{im} T + \ker T = V \overset{(\dagger)}{\implies} V = \operatorname{im} T \oplus \ker T = W \oplus K$$

where $(\star)$ is via Exercise M.7(b) and $(\dagger)$ is via Prop 3.6.6(c). ◻

(c) *Proof.*
Let $w_1, \dots, w_m$ be a basis of W and $k_1, \dots, k_\ell$ be a basis of K. From (b), this means that $w_1, \dots, w_m, k_1, \dots, k_\ell$ forms a basis of V. If we write $w_i = T(v_i)$ for $v_i \in V$, then note

$$T(w_i) = T(T(v_i)) = T^2(v_i) = T(v_i) = w_i = 0w_1 + \cdots + 1w_i + \cdots + 0w_m + 0k_1 + \cdots + 0k_\ell$$

and

$$T(k_i) = 0 = 0w_1 + \cdots + 0w_i + \cdots + 0w_m + 0k_1 + \cdots + 0k_\ell$$

Thus the matrix representation of T with respect to this basis is

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

and therefore $\operatorname{tr} T = m = \dim W = \dim \operatorname{im} T = \operatorname{rank} T$. ◻

M.10

Let A and B be $m \times n$ and $n \times m$ real matrices.

(a) Prove that if λ is a nonzero eigenvalue of the $m \times m$ matrix AB then it is also an eigenvalue of the $n \times n$ matrix BA. Show by example that this need not be true if $\lambda = 0$.

(b) Prove that $I_m - AB$ is invertible if and only if $I_n - BA$ is invertible.

Solution.

(a) *Proof.*
Let λ be a nonzero eigenvalue of AB and let v be an eigenvector, i.e. $v \neq 0$ and $ABv = \lambda v$. Note that if Bv is zero, then

$$Bv = 0 \implies ABv = 0 \implies \lambda v = 0$$

which is a contradiction since λ and v are both assumed to be nonzero. Hence $Bv \neq 0$. Then

$$BA(Bv) = B(ABv) = B(\lambda v) = \lambda Bv$$

and Bv is an eigenvector of BA with eigenvalue λ.

For a counterexample, consider

$$A = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 1 & 0 \end{bmatrix}$$

Then note

$$ABe_2 = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} = 0e_2$$

so 0 is an eigenvalue of AB. However,

$$BA = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \end{bmatrix}$$

and so 0 is not an eigenvalue of BA. □

(b) *Proof.*
Note that $\det(I_m - AB) = 0$ if and only if 1 is an eigenvalue of AB. Furthermore, the result in (a) is an if and only if simply by making the symmetric argument for the reverse. Therefore

$$\begin{aligned} \det(I_m - AB) \neq 0 &\iff 1 \text{ not eigenvalue of } AB \\ &\overset{\text{(a)}}{\iff} 1 \text{ not eigenvalue of } BA \\ &\iff \det(I_n - BA) \neq 0 \end{aligned}$$

i.e. $I_m - AB$ is invertible if and only if $I_n - BA$ is invertible. □