1 A Brief Review of Linear Algebra

1.1. Let $E = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ and M be a 2-by-2 matrix. Show that the matrix EM is obtained from M by interchanging the first and second rows.

By direct calculation,

$$\begin{aligned} EM &= \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix} \\ &= \begin{pmatrix} 0 \cdot M_{11} + 1 \cdot M_{21} & 0 \cdot M_{12} + 1 \cdot M_{22} \\ 1 \cdot M_{11} + 0 \cdot M_{21} & 1 \cdot M_{12} + 0 \cdot M_{22} \end{pmatrix} \\ &= \begin{pmatrix} M_{21} & M_{22} \\ M_{11} & M_{12} \end{pmatrix} \end{aligned}$$

which is simply M with its rows interchanged.

1.2. Let $E = \begin{pmatrix} s_1 & 0 \\ 0 & s_2 \end{pmatrix}$ and M be a 2-by-2 matrix. Show that the matrix EM is obtained from M by multiplying the elements in the first row by s_1 and the elements in the second row by s_2.

By direct calculation,

$$\begin{aligned} EM &= \begin{pmatrix} s_1 & 0 \\ 0 & s_2 \end{pmatrix} \begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix} \\ &= \begin{pmatrix} s_1 \cdot M_{11} + 0 \cdot M_{21} & s_1 \cdot M_{12} + 0 \cdot M_{22} \\ 0 \cdot M_{11} + s_2 \cdot M_{21} & 0 \cdot M_{12} + s_2 \cdot M_{22} \end{pmatrix} \\ &= \begin{pmatrix} s_1 M_{11} & s_1 M_{12} \\ s_2 M_{21} & s_2 M_{22} \end{pmatrix} \end{aligned}$$

which is simply M with its first row multiplied by s_1 and its second row multiplied by s_2.

1.3. Let $E_1 = \begin{pmatrix} 1 & 0 \\ s & 1 \end{pmatrix}$ and M by a 2-by-2 matrix. Show that the matrix $E_1 M$ is obtained from M by multiplying the first row by s and adding it to the second row. Similarly for the matrix $E_2 = \begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix}$: $E_2 M$ is obtained from M by multiplying the second row by s and adding it to the first row.

By direct calculation,

$$\begin{aligned} E_1 M &= \begin{pmatrix} 1 & 0 \\ s & 1 \end{pmatrix} \begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix} \\ &= \begin{pmatrix} 1 \cdot M_{11} + 0 \cdot M_{21} & 1 \cdot M_{12} + 0 \cdot M_{22} \\ s \cdot M_{11} + 1 \cdot M_{21} & s \cdot M_{12} + 1 \cdot M_{22} \end{pmatrix} \\ &= \begin{pmatrix} M_{11} & M_{12} \\ sM_{11} + M_{21} & sM_{12} + M_{22} \end{pmatrix} \end{aligned}$$

which is simply M with its first row multiplied by s and added to its second row. Similarly,

$$\begin{aligned} E_1 M &= \begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix} \begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix} \\ &= \begin{pmatrix} 1 \cdot M_{11} + s \cdot M_{21} & 1 \cdot M_{12} + s \cdot M_{22} \\ 0 \cdot M_{11} + 1 \cdot M_{21} & 0 \cdot M_{12} + 1 \cdot M_{22} \end{pmatrix} \\ &= \begin{pmatrix} sM_{21} + M_{11} & sM_{22} + M_{12} \\ M_{21} & M_{22} \end{pmatrix} \end{aligned}$$

is just M with its second row multiplied by s and added to its first row.

1.4. The three operations effected in exercises 1-3 are known as elementary row operations. The Es defined here are called elementary matrices. How do you effect the corresponding there elementary column operations on an arbitrary 2-by-2 matrix M?

The elementary column operations are the transpose of the elementary row operations. To see this, note that we can enact column operations through

$$(EM^T)^T = (M^T)^T E^T = ME^T$$

In particular, we see that we must act on M from the right with E^T to enact column operations.

1.5. Let $e = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$. Show that multiplying a 3-by-3 matrix M from the right with e interchanges two columns in M and that multiplying M from the left with e interchanges two rows in M.

Denote $M = \begin{pmatrix} \overrightarrow{M}_1 & \overrightarrow{M}_2 & \overrightarrow{M}_3 \end{pmatrix}$, where each $\overrightarrow{M}_i$ is a column vector. Then

$$Me = \begin{pmatrix} \overrightarrow{M}_1 & \overrightarrow{0} & \overrightarrow{0} \end{pmatrix} + \begin{pmatrix} \overrightarrow{0} & \overrightarrow{0} & \overrightarrow{M}_2 \end{pmatrix} + \begin{pmatrix} \overrightarrow{0} & \overrightarrow{M}_3 & \overrightarrow{0} \end{pmatrix} = \begin{pmatrix} \overrightarrow{M}_1 & \overrightarrow{M}_3 & \overrightarrow{M}_2 \end{pmatrix},$$

i.e. postmultiplying by e exchanges two columns in M.

Now let's change our labeling so that $\overrightarrow{M}_i^T$ refers to a row in M. That is

$$M = \begin{pmatrix} \overrightarrow{M}_1^T \\ \overrightarrow{M}_2^T \\ \overrightarrow{M}_3^T \end{pmatrix}.$$

Premultiplying M by e gives

$$eM = \begin{pmatrix} \overrightarrow{M}_1^T \\ \overrightarrow{0}^T \\ \overrightarrow{0}^T \end{pmatrix} + \begin{pmatrix} \overrightarrow{0}^T \\ \overrightarrow{0}^T \\ \overrightarrow{M}_2^T \end{pmatrix} + \begin{pmatrix} \overrightarrow{0}^T \\ \overrightarrow{M}_3^T \\ \overrightarrow{0}^T \end{pmatrix} = \begin{pmatrix} \overrightarrow{M}_1^T \\ \overrightarrow{M}_3^T \\ \overrightarrow{M}_2^T \end{pmatrix},$$

i.e. it exchanges two rows in M.

1.6. Generalize the elementary row and column operations to n-by-n matrices.

To exchange two rows, simply exchange the desired two rows in the identity matrix before right-multiplying M. You can pull off a similar trick with columns: just switch the desired two columns in the identity before left-multiplying by M.

Scaling rows and columns is easily achieved by placing scaling factors on the diagonal of an empty matrix. Multiplying by M on the right scales the nth row by the nth value, while multiplying by M on the left scales the nth column correspondingly.

Now consider a 'near-identity matrix' with some off-diagonal element given by s. If this is in the E_{ij}th spot, left-multiplying by M will add s times the jth column to the ith column. By contrast, right-multiplying by M will add s times the jth row to the ith row.

1.7. Let $c = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$ and d be a 3-by-3 diagonal matrix. Show that the similarity transformation $c^{-1}dc$ permutes the diagonal elements of d cyclically.

To start,

$$c^{-1} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix},$$

which shows us that

$$\begin{aligned} c^{-1}dc &= \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} d_1 & 0 & 0 \\ 0 & d_2 & 0 \\ 0 & 0 & d_3 \end{pmatrix} \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \\ &= \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 0 & d_1 \\ d_2 & 0 & 0 \\ 0 & d_3 & 0 \end{pmatrix} \\ &= \begin{pmatrix} d_2 & 0 & 0 \\ 0 & d_3 & 0 \\ 0 & 0 & d_1 \end{pmatrix} \end{aligned}$$

Repeating this yields

$$\begin{aligned} c^{-1}(c^{-1}dc)c &= \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} d_2 & 0 & 0 \\ 0 & d_3 & 0 \\ 0 & 0 & d_1 \end{pmatrix} \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \\ &= \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 0 & d_2 \\ d_3 & 0 & 0 \\ 0 & d_1 & 0 \end{pmatrix} \\ &= \begin{pmatrix} d_3 & 0 & 0 \\ 0 & d_1 & 0 \\ 0 & 0 & d_2 \end{pmatrix} \end{aligned}$$

and

$$
\begin{aligned}
c^{-1}(c^{-1}c^{-1}dcc)c &= \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} d_3 & 0 & 0 \\ 0 & d_1 & 0 \\ 0 & 0 & d_2 \end{pmatrix} \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \\
&= \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 0 & d_3 \\ d_1 & 0 & 0 \\ 0 & d_2 & 0 \end{pmatrix} \\
&= \begin{pmatrix} d_1 & 0 & 0 \\ 0 & d_2 & 0 \\ 0 & 0 & d_3 \end{pmatrix}
\end{aligned}
$$

As this is the matrix we began with, we see that repeatedly applying $c^{-1}dc$ yields a cycle of permutations.

1.8. Show that the determinant of antisymmetric n-by-n matrices vanishes if n is odd.

Since all antisymmetric matrices satisfy $M^T = -M$, we have $\det M^T = -\det M$ for n odd and $\det M^T = \det M$ for n even. This follows from viewing M^T as an operation that multiplies each column by -1, and hence $\det M^T = (-1)^n \det M$.

Using the fact that $\det M = \det M^T$ for all matrices, we see that an antisymmetric odd-dimensional matrix must have $\det M = -\det M$. This is only possible if $\det M$ vanishes.

1.9. Diagonalize the matrix $M = \begin{pmatrix} a & 1 \\ 0 & b \end{pmatrix}$. Show that M is not diagonalizable if $a = b$.

The eigenvalues of M are a and b (to see this, consider the invariant nature of eigenvalues under the transpose operation), and so its eigenvectors are

$$\eta_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \text{and} \quad \eta_2 = \begin{pmatrix} 1 \\ b-a \end{pmatrix}$$

We could go through the process of performing a similarity transformation, but it's more useful to think of diagonalization as a change of basis to these eigenvectors. In that case, our matrix can be expressed as

$$M_{\text{diag}} = \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix},$$

where $x_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ corresponds to η_1 in our new basis and $x_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ corresponds to η_2.

Clearly, when $a = b$, we have $\eta_1 = \eta_2$. This degeneracy makes diagonalization impossible. To see this explicitly, remember that the similarity transformation we need to perform to diagonalize our matrix is $S^{-1}MS$ where

$$S = \begin{pmatrix} 1 & 1 \\ 0 & b-a \end{pmatrix}.$$

When $a = b$ this has no inverse.

1.10. Write the polynomial equation satisfied by the eigenvalues of a matrix M that is diagonalizable as $\lambda^n + \sum_{k=1}^n c_k \lambda^{n-k} = 0$. Show that the matrix M satisfies the equation $M^n + \sum_{k=1}^n c_k M^{n-k} = 0$.

> It is easiest to exponentiate M after it has been diagonalized. Since $S^{-1}MS = D$, we have $M = SDS^{-1}$. Exponentiating this gives $M^k = SDS^{-1}S \cdots SDS^{-1} = SD^kS^{-1}$. Inserting this into our matrix's characteristic equation yields
>
> $$SD^nS^{-1} + \sum_{k=1}^{n} c_k SD^{n-k}S^{-1} = S(D^n + \sum_{k=1}^{n} c_k D^{n-k})S^{-1} = S \cdot 0 \cdot S^{-1} = 0,$$
>
> where the middle equality follows from the fact that each entry in D is one of M's eigenvalues, and so D^n satisfies $D^n + \sum_{k=1}^n c_k D^{n-k} = 0$.

1.11. Show by explicit computation that the eigenvalues of the traceless matrix $M = \begin{pmatrix} c & a-b \\ a+b & -c \end{pmatrix}$ with a, b, c real have the form $\lambda = \pm w$, with w either real or imaginary. Show that for $b = 0$, the eigenvalues are real. State the theorem that this result verifies.

> The characteristic equation of M is given by
>
> $$-(c+\lambda)(c-\lambda) - (a+b)(a-b) = \lambda^2 - c^2 + b^2 - a^2 = 0$$
>
> i.e. $\lambda = \pm\sqrt{a^2 + c^2 - b^2}$. This value is real when $b^2 \leq a^2 + b^2$ and imaginary otherwise. When $b = 0$, we have $\lambda = \pm\sqrt{a^2 + c^2}$. This makes sense: eigenvalues give the values by which our matrix stretches the plane. M rotates our standard Cartesian unit vectors while stretching by a in one direction and c in another. Our eigenvalues are simply the total distance stretched. We have just verified the Pythagorean theorem.

1.12. Let the matrix in exercise 11 be complex. Find its eigenvalues. When do they become real?

> Our characteristic equation does not change, and so $\lambda = \pm\sqrt{a^2 + c^2 - b^2}$. From this, it is clear that we obtain real eigenvalues when $a^2 - b^2 + c^2$ lands on the positive half of the real line.

1.13. There were of course quite a few nineteenth-century theorems about matrices. Given A and B an m-by-n matrix and an n-by-m matrix, respectively, Sylvester's theorem states that $\det(I_m + AB) = \det(I_n + BA)$. Prove this for A and B square and invertible.

> A particularly nice way to prove this involves finding two matrices M and N such that $\det MN = \det(I + AB)$ and $\det NM = \det(I + BA)$. Then the proof follows from the fact that $\det MN = \det M \det N = \det N \det M = \det NM$.
>
> Consider the matrices
>
> $$M = \begin{pmatrix} 1 & -A \\ B & 1 \end{pmatrix} \quad \text{and} \quad N = \begin{pmatrix} 1 & A \\ 0 & 1 \end{pmatrix},$$
>
> where each 1 denotes a fitting identity submatrix. We have $\det MN = \det(I + AB)$ and $\det NM = \det(I + BA)$, and so $\det(I_m + AB) = \det(I_n + BA)$. As indicated in the punchline of the last sentence, this proof works for all matrices that are compatible, not just those that are square and invertible.

1.14. Show that $(M^{-1})^T = (M^T)^{-1}$ and $(M^{-1})^* = (M^*)^{-1}$. Thus, $(M^{-1})^\dagger = (M^\dagger)^{-1}$.

Starting with an identity, we see

$$\begin{aligned} MM^{-1} &= I \\ (MM^{-1})^T &= I \\ (M^{-1})^T M^T &= I \\ (M^{-1})^T &= (M^T)^{-1} \end{aligned}$$

Similarly,

$$\begin{aligned} MM^{-1} &= I \\ M^*(M^{-1})^* &= I \\ (M^{-1})^* &= (M^*)^{-1} \end{aligned}$$

As $M^\dagger = M^{*T}$, the above results show $(M^{-1})^\dagger = (M^\dagger)^{-1}$.

1.15. The n-by-n matrix M defined by $M_{ij} = x_j^{i-1}$ plays an important role in random matrix theory, for example. Show that $\det M$, known as the Vandermonde determinant, is equal to (up to an overall sign) $\prod_{i<j}(x_i - x_j)$.

Consider the Vandermonde matrix

$$M = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \\ \vdots & \vdots & \ddots & \vdots \\ x_1^{n-1} & x_2^{n-1} & \cdots & x_n^{n-1} \end{pmatrix}.$$

Let us diagonalize this matrix by use of elementary column operations. Subtracting the first column from all others, we see $\det M = \det A$, where A is the $(n-1)$-by-$(n-1)$ matrix given by

$$A = \begin{pmatrix} x_2 - x_1 & \cdots & x_n - x_1 \\ \vdots & \ddots & \vdots \\ x_2^{n-1} - x_1^{n-1} & \cdots & x_n^{n-1} - x_1^{n-1} \end{pmatrix}.$$

From this, it is clear that

$$A_{ij} = (x_{j+1}^i - x_1^i) = (x_{j+1} - x_1) \sum_{k=0}^{i-1} x_{j+1}^k x_1^{i-1-k}.$$

Dividing each column by its associated $(x_{j+1} - x_1)$, we find $\det M = \det A = \prod_{j=2}^n (x_j - x_1) \det B$, where

$$B = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ \sum_{k=0}^{1} x_2^k x_1^{1-k} & \sum_{k=0}^{1} x_3^k x_1^{1-k} & \cdots & \sum_{k=0}^{1} x_n^k x_1^{1-k} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{k=0}^{n-2} x_2^k x_1^{n-2-k} & \sum_{k=0}^{n-2} x_3^k x_1^{n-2-k} & \cdots & \sum_{k=0}^{n-2} x_n^k x_1^{n-2-k} \end{pmatrix}$$

Defining $a_{ij} = \sum_{k=0}^{i-1} x_{j+1}^k x_1^{i-1-k}$, we see $a_{ij} = x_{j+1}^{i-1} + x_1 \sum_{k=0}^{i-2} x_{j+1}^k x_1^{i-2-k} = x_{j+1}^{i-1} + x_1 a_{i-1,j}$. Thus we can rewrite our new matrix as

$$B = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1,n-1} \\ x_2 + x_1 a_{11} & x_3 + x_1 a_{12} & \cdots & x_n + x_1 a_{1,n-1} \\ \vdots & \vdots & \ddots & \vdots \\ x_2^{n-2} + x_1 a_{n-2,1} & x_3^{n-2} + x_1 a_{n-2,2} & \cdots & x_n^{n-2} + x_1 a_{n-2,n-1} \end{pmatrix}.$$

Noting that $B_{ij} = a_{ij}$, we can start at the bottom of this matrix and subtract x_1 times the $(i-1)$th row from the ith row to determine B'

$$B' = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ x_2 & x_3 & \cdots & x_n \\ \vdots & \vdots & \ddots & \vdots \\ x_2^{n-2} & x_3^{n-2} & \cdots & x_n^{n-2} \end{pmatrix},$$

which is a $(n-1) \times (n-1)$ Vandermonde matrix with $\det B' = \det B$. We can continue in this fashion until we reach a 1×1 matrix, at which point we see the determinant of A is given by

$$\det A = \prod_{i<j} (x_i - x_j).$$

1.16. Let the Cartesian coordinates of the three vertices of a triangle be given by (x_i, y_i) for $i = 1, 2, 3$. Show that the area of a triangle is given by

$$\text{Area of triangle} = \frac{1}{2} \det \begin{pmatrix} x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \\ 1 & 1 & 1 \end{pmatrix}$$

Interestingly, this expression is the beginning of the notion of projective geometry and plays a crucial role in recent development in theoretical physics (see, for example, Henriette Elvang and You-tin Huang, Scattering Amplitudes in Gauge Theory and Gravity, p. 203).

Consider the bounding box of an arbitrary triangle with one vertex being shared between the two objects. Label the coordinates of this vertex (x_1, y_1). Take the other two vertices to be (x_2, y_2) and (x_3, y_3). Obviously, these lie on the edges of the bounding box.

We will arbitrarily take the second point to be further from the first in the x-coordinate, with the third point being further from the first in the y-coordinate. The argument can be repeated for different variations.

The total area of our triangle is given by the area of its bounding box minus the area of its surrounding triangles. In our case, these quantities are given by

$$\begin{aligned}
A &= (x_2 - x_1)(y_3 - y_1) = x_2y_3 - x_2y_1 - x_1y_3 + x_1y_1 \\
B &= \frac{1}{2}(x_3 - x_1)(y_3 - y_1) = \frac{1}{2}(x_3y_3 - x_3y_1 - x_1y_3 + x_1y_1) \\
C &= \frac{1}{2}(x_2 - x_3)(y_3 - y_2) = \frac{1}{2}(x_2y_3 - x_2y_2 - x_3y_3 + x_3y_2) \\
D &= \frac{1}{2}(x_2 - x_1)(y_2 - y_1) = \frac{1}{2}(x_2y_2 - x_2y_1 - x_1y_2 + x_1y_1)
\end{aligned}$$

and so our triangle's area is

$$A - B - C - D = \frac{1}{2}(x_1(y_2 - y_3) - x_2(y_1 - y_3) + x_3(y_1 - y_2)) = \frac{1}{2}\begin{vmatrix} x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \\ 1 & 1 & 1 \end{vmatrix}.$$