Exercise 15.3. This exercise illustrates in the 3×3 case how matrix multiplication encodes certain "row operations" on general matrices. Consider the following two 3×3 matrices A and B, and a general 3×3 matrix M:

$$A = \begin{bmatrix} 1 & a & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad M = \begin{bmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{31} & m_{32} & m_{33} \end{bmatrix}.$$

(a) Show that for every 3×3 matrix M, the product AM is obtained from M by adding a times the second row to the first row. How is MA related to M?

(b) Show that for every 3×3 matrix M, the product BM is obtained from M by swapping the first and second rows. How is MB related to M?

These facts for 3×3 matrices work for $n \times n$ matrices similarly (but you are not being asked to show this; it just requires some more dexterity with the notation).

a)

$$AM = \begin{bmatrix} 1 & a & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{31} & m_{32} & m_{33} \end{bmatrix} = \begin{bmatrix} m_{11} + am_{21} & m_{12} + am_{22} & m_{13} + am_{23} \\ m_{21} & m_{22} & m_{23} \\ m_{31} & m_{32} & m_{33} \end{bmatrix}$$

We see that this corresponds with our fact that AM is obtained by adding a times row 2 to row 1.

$$MA = \begin{bmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{31} & m_{32} & m_{33} \end{bmatrix} \begin{bmatrix} 1 & a & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} m_{11} & am_{11} + m_{12} & m_{13} \\ m_{21} & am_{21} + m_{22} & m_{23} \\ m_{31} & am_{31} + m_{32} & m_{33} \end{bmatrix}$$

MA is obtained from M by adding a times the first column to the second column.

b)

$$BM = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{31} & m_{32} & m_{33} \end{bmatrix} = \begin{bmatrix} m_{21} & m_{22} & m_{23} \\ m_{11} & m_{12} & m_{13} \\ m_{31} & m_{32} & m_{33} \end{bmatrix}$$

We see that this corresponds with our fact that BM is obtained by swapping the first and second rows.

$$MB = \begin{bmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{31} & m_{32} & m_{33} \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} m_{12} & m_{11} & m_{13} \\ m_{22} & m_{21} & m_{23} \\ m_{32} & m_{31} & m_{33} \end{bmatrix}$$

MB is obtained from M by swapping the first and second columns.

Exercise 15.8. Let $A = \begin{bmatrix} 2 & 3 & -4 \\ 4 & -1 & 6 \\ 5 & 2 & 1 \end{bmatrix}$ (as in Exercise 15.7).

(a) Find a *non-zero* 3×3 matrix M for which AM is the 3×3 zero matrix (there are many valid solutions). (One approach: because of Exercise 15.7, we know that $A(B - C) = 0$, but $B - C$ only has two columns.)

(b) Using your matrix M from part (a), compute MA. Is it equal to AM? (The answer will depend on your choice of M.)

a) Suppose each column of M is in the form of $c_i = \begin{bmatrix} x_i \\ y_i \\ z_i \end{bmatrix}$.

We wish to find c_i such that Ac_i is the $\vec{0}$ in $\mathbb{R}^3$, for all c_i. Therefore, we can do matrix-vector multiplication with a general vector $\vec{c}$ to get a system of equations

$$\begin{bmatrix} 2 & 3 & -4 \\ 4 & -1 & 6 \\ 5 & 2 & 1 \end{bmatrix}\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \vec{0} \rightarrow \begin{bmatrix} 2x+3y-4z \\ 4x-y+6z \\ 5x+2y+z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

Therefore, our equations are $2x+3y-4z=0$, $4x-y+6z=0$, and $5x+2y+z=0$. We multiply the last equation by -2 and add the LHS and RHS of all to get

$$(2x+3y-4z)+(4x-y+6z)+(-10x-4y-2z) = (0+0+0)$$

This equation simplifies to $-4x-2y=0 \rightarrow -4x = 2y \rightarrow y = -2x$. We substitute this into the first equation to get $2x + 3(-2x) - 4z = 0 \rightarrow -4x-4z=0 \rightarrow z = -x$.

Therefore, for all colums c_i, it can be represented as $\begin{bmatrix} x_i \\ -2x_i \\ -x_i \end{bmatrix}$.

We pick $x_i = 1, 2, 3$ for each of our colums to get $\boxed{M = \begin{bmatrix} 1 & 2 & 3 \\ -2 & -4 & -6 \\ -1 & -2 & -3 \end{bmatrix}}$.

Using this matrix, $AM = \begin{bmatrix} 2 & 3 & -4 \\ 4 & -1 & 6 \\ 5 & 2 & 1 \end{bmatrix}\begin{bmatrix} 1 & 2 & 3 \\ -2 & -4 & -6 \\ -1 & -2 & -3 \end{bmatrix} = \begin{bmatrix} 2-6+4 & 4-12+8 & 6-18+12 \\ 4+2-6 & 8+4-12 & 12+6-18 \\ 5-4-1 & 10-8-2 & 15-12-3 \end{bmatrix}$

$$AM = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

b)

$$MA = \begin{bmatrix} 1 & 2 & 3 \\ -2 & -4 & -6 \\ -1 & -2 & -3 \end{bmatrix}\begin{bmatrix} 2 & 3 & -4 \\ 4 & -1 & 6 \\ 5 & 2 & 1 \end{bmatrix} = \boxed{\begin{bmatrix} 25 & 7 & 11 \\ -50 & -14 & -22 \\ -25 & -7 & -11 \end{bmatrix} \text{ This is not equal to } AM.}$$

Exercise 15.11. Let D be a 5×5 diagonal matrix with diagonal entries $d_1, d_2, \ldots, d_5 \in \mathbf{R}$ (from top left to bottom right). Suppose A is a 5×5 matrix whose second and fourth columns, respectively, are $\begin{bmatrix} 2 \\ -1 \\ 0 \\ 3 \\ 4 \end{bmatrix}$ and $\begin{bmatrix} 1 \\ 8 \\ -9 \\ 3 \\ 1 \end{bmatrix}$. For each of the following, compute it in terms of the d_i's or explain why there is not enough information given to do so:

(a) The second column of DA.
(b) The fourth column of AD.

a)

$$DA = \begin{array}{c} D \\ \begin{bmatrix} d_1 & 0 & 0 & 0 & 0 \\ 0 & d_2 & 0 & 0 & 0 \\ 0 & 0 & d_3 & 0 & 0 \\ 0 & 0 & 0 & d_4 & 0 \\ 0 & 0 & 0 & 0 & d_5 \end{bmatrix} \end{array} \begin{array}{c} A \\ \begin{bmatrix} \vdots & 2 & \vdots & 1 & \vdots \\ \vdots & -1 & \vdots & 8 & \vdots \\ \vdots & 0 & \vdots & -9 & \vdots \\ \vdots & 3 & \vdots & 3 & \vdots \\ \vdots & 4 & \vdots & 1 & \vdots \end{bmatrix} \end{array} = \begin{bmatrix} - & 2d_1 & - & 1d_1 & - \\ - & -1d_2 & - & 8d_2 & - \\ - & 0d_3 & - & -9d_3 & - \\ - & 3d_4 & - & 3d_4 & - \\ - & 4d_5 & - & 1d_5 & - \end{bmatrix}$$

The second column of DA is $\boxed{\begin{bmatrix} 2d_1 \\ -d_2 \\ 0 \\ 3d_4 \\ 4d_5 \end{bmatrix}}$.

b)

$$AD = \begin{bmatrix} \vdots & 2 & \vdots & 1 & \vdots \\ \vdots & -1 & \vdots & 8 & \vdots \\ \vdots & 0 & \vdots & -9 & \vdots \\ \vdots & 3 & \vdots & 3 & \vdots \\ \vdots & 4 & \vdots & 1 & \vdots \end{bmatrix} \begin{bmatrix} d_1 & 0 & 0 & 0 & 0 \\ 0 & d_2 & 0 & 0 & 0 \\ 0 & 0 & d_3 & 0 & 0 \\ 0 & 0 & 0 & d_4 & 0 \\ 0 & 0 & 0 & 0 & d_5 \end{bmatrix} = \begin{bmatrix} \vdots & \vdots & \vdots & c_1 & \vdots \\ \vdots & \vdots & \vdots & c_2 & \vdots \\ \vdots & \vdots & \vdots & c_3 & \vdots \\ \vdots & \vdots & \vdots & c_4 & \vdots \\ \vdots & \vdots & \vdots & c_5 & \vdots \end{bmatrix}$$

Let x indicate an unknown value, and c_i indicate the value on the ith row of the fourth column of AD.

$$c_1 = x(0) + 2(0) + x(0) + 1(d_4) + x(0) = d_4$$
$$c_2 = x(0) - 1(0) + x(0) + 8(d_4) + x(0) = 8d_4$$
$$c_3 = x(0) + 0(0) + x(0) - 9(d_4) + x(0) = -9d_4$$
$$c_4 = x(0) + 3(0) + x(0) + 3(d_4) + x(0) = 3d_4$$
$$c_5 = x(0) + 4(0) + x(0) + 1(d_4) + x(0) = d_4$$

The fourth column of AD is $\boxed{\begin{bmatrix} d_4 \\ 8d_4 \\ -9d_4 \\ 3d_4 \\ d_4 \end{bmatrix}}$

Exercise 15.13.

(a) Let $A = \begin{bmatrix} 1 \\ -3 \\ 5 \end{bmatrix}$, $B = \begin{bmatrix} -2 & 1 & 4 \end{bmatrix}$. Compute BA and AB. (Note that the 1×1 matrix BA has as its single entry the dot product, also sometimes called the *inner product*, of the two vectors $\begin{bmatrix} 1 \\ -3 \\ 5 \end{bmatrix}$ and $\begin{bmatrix} -2 \\ 1 \\ 4 \end{bmatrix}$; the 3×3 matrix product AB in the other order is then sometimes called the *outer product* of $\begin{bmatrix} 1 \\ -3 \\ 5 \end{bmatrix}$ and $\begin{bmatrix} -2 \\ 1 \\ 4 \end{bmatrix}$.)

(b) What do you notice about how the rows of AB are related to each other, and likewise for the columns?

(c) Let $C = \begin{bmatrix} 3 & -2 & 2 \\ -6 & 4 & -4 \\ 12 & -8 & 8 \end{bmatrix}$. Find a 3×1 matrix A and 1×3 matrix B for which $C = AB$. (Matrices arising in this way, as a nonzero column multiplied on the left against a nonzero row, are called "rank 1" matrices and are very important throughout data analysis.)

a) $BA = \begin{bmatrix} -2 & 1 & 4 \end{bmatrix} \begin{bmatrix} 1 \\ -3 \\ 5 \end{bmatrix} = \begin{bmatrix} -2 - 3 + 20 \end{bmatrix} = \boxed{\begin{bmatrix} 15 \end{bmatrix}}$

$AB = \begin{bmatrix} 1 \\ -3 \\ 5 \end{bmatrix} \begin{bmatrix} -2 & 1 & 4 \end{bmatrix} = \boxed{\begin{bmatrix} -2 & 1 & 4 \\ 6 & -3 & -12 \\ -10 & 5 & 20 \end{bmatrix}}$

b) The rows of AB are all scalar multiples of each other. They are all multiples of the matrix B. The colums are also multiples of each other. They are all scalar multiples of the matrix A.

c) $\boxed{A = \begin{bmatrix} 1 \\ -2 \\ 4 \end{bmatrix} \quad B = \begin{bmatrix} 3 & -2 & 2 \end{bmatrix}}$

$AB = \begin{bmatrix} 1 \\ -2 \\ 4 \end{bmatrix} \begin{bmatrix} 3 & -2 & 2 \end{bmatrix} = \begin{bmatrix} 3 & -2 & 2 \\ -6 & 4 & -4 \\ 12 & -8 & 8 \end{bmatrix}$

Exercise 16.1. Two phone companies, let's call them P_1 and P_2, are competing for customers in the Bay Area. Suppose that during each calendar year, 85% of the customers of P_1 stay with P_1 and 15% of P_1's customers switch to P_2, and that during each calendar year, 92% of the customers of P_2 stay with P_2 and 8% of their customers switch to P_1.

(a) For each $n \geq 0$, define the vector $\mathbf{v}_n = \begin{bmatrix} x_n \\ y_n \end{bmatrix}$, where x_n is the number of customers for P_1 in year n, and y_n is the number of customers of P_2 in year n. Explain why

$$x_{n+1} = (0.85)x_n + (0.08)y_n, \quad y_{n+1} = (0.15)x_n + (0.92)y_n,$$

and write a 2×2 Markov matrix M for which $\mathbf{v}_{n+1} = M\mathbf{v}_n$ for all $n \geq 0$.

(b) For the correct Markov matrix M that you should have found in (a), direct computation (which we are not asking you to do here) shows that $M^2 = \begin{bmatrix} .7345 & .1416 \\ .2655 & .8584 \end{bmatrix}$. What does the first row of this matrix tell us concerning the behavior of customers of each of P_1 and P_2 after two years have elapsed? (A customer who began with one of these companies and is with it at the end of two years may have switched to the other company in the middle of this time, but don't try to keep track of that level of detail.)

(c) Again using the correct matrix M from (a), one can show that $M^m \approx \begin{bmatrix} .3478 & .3478 \\ .6522 & .6522 \end{bmatrix}$ for large m. Interpret the meaning of the two numbers appearing in this matrix.

a) We know that each calendar year, 85% of customers in P_1 stay in P_1 and 8% of customers of P_2 move to P_1. Similarly, 92% of customers in P_2 stay in P_2 and 15% of customers in P_1 move to P_2. This change would be represented as the result for the next calendar year. Therefore, if the current year is n, the number of customers in P_1 in the next calendar year is $x_{n+1} = 0.85x_n + 0.08y_n$. Similarly, the number of customers in P_2 in the next calendar year is $y_{n+1} = 0.15x_n + 0.92y_n$.

$$\vec{v}_{n+1} = \begin{bmatrix} x_{n+1} \\ y_{n+1} \end{bmatrix} = \begin{bmatrix} 0.85x_n + 0.08y_n \\ 0.15x_n + 0.92y_n \end{bmatrix} = \begin{bmatrix} 0.85 & 0.08 \\ 0.15 & 0.92 \end{bmatrix}\begin{bmatrix} x_n \\ y_n \end{bmatrix}$$

$$\boxed{M = \begin{bmatrix} 0.85 & 0.08 \\ 0.15 & 0.92 \end{bmatrix}}$$

b) The first row of M^2 tells us that after 2 years, 73.45% of the customers who were originally in P_1 stayed in P_1, and 14.16% of the customers that were originally in P_2 have switched to P_1.

c) This matrix tells us that that the matrix stabilizes. This means after m years, where m is large, regardless of what the initial number of customers in P_1 and P_2 are (as long as there is the same total number of customers), 34.78% of the total number of customers will be in P_1, and 65.22% of the total number of customers will be in P_2.

Exercise 16.3. A population of birds lives on two islands I_1 and I_2. On January 1 of year y, there are n_1 birds on island I_1 and n_2 birds on island I_2. Each spring, 10% of the birds on I_1 move to I_2, and each fall 10% of the birds on I_2 move to I_1. There is no other movement of birds between the islands. Let $\mathbf{p} = \begin{bmatrix} n_1 \\ n_2 \end{bmatrix}$ be the "population vector" whose ith entry is the bird population on island I_i on January 1 of some particular year y.

(a) Let M be the 2×2 matrix for which the population vector at the end of year y is $M\mathbf{p}$ (disregarding births and deaths of birds). Then M is a product of some of the matrices below:

$$A = \begin{bmatrix} 0.9 & 0.1 \\ 0 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} 0.9 & 0 \\ 0.1 & 1 \end{bmatrix}, \quad C = \begin{bmatrix} 1 & 0 \\ 0.1 & 0.9 \end{bmatrix}, \quad D = \begin{bmatrix} 1 & 0.1 \\ 0 & 0.9 \end{bmatrix}.$$

Find the matrices above for which M is a product of these matrices (remember that order matters), and then compute M.

(b) What is the matrix N for which $N\mathbf{p}$ is the population vector at the end of year $y+1$? Compute it explicitly.

(c) For the correct answer M to (a), $M^n \approx \begin{bmatrix} .5263 & .5263 \\ .4737 & .4737 \end{bmatrix}$ for large n. Interpret the meaning of the common number in the first row and the common number in the second row.

a) In spring, $p_s = \begin{bmatrix} n_{1s} \\ n_{2s} \end{bmatrix} = \begin{bmatrix} 0.9 n_1 + 0 n_2 \\ 0.1 n_1 + 1 n_2 \end{bmatrix}$. In fall, $p_f = \begin{bmatrix} n_{1f} \\ n_{2f} \end{bmatrix} = \begin{bmatrix} 1 n_{1s} + 0.1 n_{2s} \\ 0 n_{1s} + 0.9 n_{2s} \end{bmatrix}$

Put together, the population at the end of the year is $\begin{bmatrix} 1 & 0.1 \\ 0 & 0.9 \end{bmatrix} \begin{bmatrix} 0.9 & 0 \\ 0.1 & 1 \end{bmatrix} \begin{bmatrix} n_1 \\ n_2 \end{bmatrix}$.

Therefore, our matrices are B and D, and $\boxed{M = DB.}$

$$M = \begin{bmatrix} 1 & 0.1 \\ 0 & 0.9 \end{bmatrix} \begin{bmatrix} 0.9 & 0 \\ 0.1 & 1 \end{bmatrix} = \begin{bmatrix} 0.9 + 0.01 & 0 + 0.1 \\ 0 + 0.09 & 0 + 0.9 \end{bmatrix} = \boxed{\begin{bmatrix} 0.91 & 0.1 \\ 0.09 & 0.9 \end{bmatrix}}$$

b) At the end of year n+1, the change from $\vec{p}$ is $\begin{bmatrix} 1 & 0.1 \\ 0 & 0.9 \end{bmatrix} \begin{bmatrix} 0.9 & 0 \\ 0.1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0.1 \\ 0 & 0.9 \end{bmatrix} \begin{bmatrix} 0.9 & 0 \\ 0.1 & 1 \end{bmatrix} \vec{p}$.

This is equivalent to $M^2 \vec{p}$. Therefore,

$$N = M^2 = \begin{bmatrix} 0.91 & 0.1 \\ 0.09 & 0.9 \end{bmatrix} \begin{bmatrix} 0.91 & 0.1 \\ 0.09 & 0.9 \end{bmatrix} = \begin{bmatrix} 0.8281 + 0.009 & 0.091 + 0.09 \\ 0.0819 + 0.081 & 0.009 + 0.81 \end{bmatrix} = \boxed{\begin{bmatrix} 0.8371 & 0.181 \\ 0.1629 & 0.819 \end{bmatrix}}$$

c) This matrix tells us that that the matrix stabilizes. This means after n years, where n is large, regardless of what the initial population of islands I_1 and I_2, (as long as there is the same total population of the islands), 52.63% of the total population of birds will be on I_1, and 47.37% of the total population will be on I_2.

Exercise 16.5. Let A be the Markov matrix $\begin{bmatrix} 1/2 & 3/4 \\ 1/2 & 1/4 \end{bmatrix}$.

(a) For $\mathbf{v} = \begin{bmatrix} 3 \\ 2 \end{bmatrix}$ and $\mathbf{w} = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$, check that $\{\mathbf{v}, \mathbf{w}\}$ is a basis of $\mathbf{R}^2$ with $A\mathbf{v} = \mathbf{v}$ and $A\mathbf{w} = -(1/4)\mathbf{w}$. (How such $\mathbf{v}$ and $\mathbf{w}$ are found will be discussed in Section 23.3.)

(b) The basis property in (a) tells us that any $\mathbf{x} \in \mathbf{R}^2$ can be written as $\mathbf{x} = \alpha\mathbf{v} + \beta\mathbf{w}$ for some scalars α, β. Using the conclusions in (a), show that $A\mathbf{x} = \alpha\mathbf{v} - (\beta/4)\mathbf{w}$, $A^2\mathbf{x} = \alpha\mathbf{v} + (\beta/16)\mathbf{w}$, and $A^3\mathbf{x} = \alpha\mathbf{v} - (\beta/64)\mathbf{w}$.

(c) Explain why we should have $A^m\mathbf{x} \approx \alpha\mathbf{v}$ for large m. (Hint: how does $\pm\beta/4^m$ behave as m grows?) This is a special case of a general result (Proposition 24.4.2) that describes high powers of a square matrix in terms of a later concept called "dominant eigenvalue".

a) $\vec{v}$ and $\vec{w}$ are not scalar multiples of each other. Therefore, $\{\vec{v}, \vec{w}\}$ is a basis of $\mathbb{R}^2$ since $\dim(\text{span}(\vec{v}, \vec{w})) = 2$.

$$A\vec{v} = \begin{bmatrix} 1/2 & 3/4 \\ 1/2 & 1/4 \end{bmatrix}\begin{bmatrix} 3 \\ 2 \end{bmatrix} = \begin{bmatrix} 3/2 + 3/2 \\ 3/2 + 1/2 \end{bmatrix} = \begin{bmatrix} 3 \\ 2 \end{bmatrix} = \vec{v}$$

$$A\vec{w} = \begin{bmatrix} 1/2 & 3/4 \\ 1/2 & 1/4 \end{bmatrix}\begin{bmatrix} -1 \\ 1 \end{bmatrix} = \begin{bmatrix} -1/2 + 3/4 \\ -1/2 + 1/4 \end{bmatrix} = \begin{bmatrix} 1/4 \\ -1/4 \end{bmatrix} = -\frac{1}{4}\begin{bmatrix} -1 \\ 1 \end{bmatrix} = \left(-\frac{1}{4}\right)\vec{w}$$

b) $\vec{x} = \alpha\vec{v} + \beta\vec{w}$: ① $A\vec{x} = A(\alpha\vec{v} + \beta\vec{w}) = \alpha(A\vec{v}) + \beta(A\vec{w}) = \alpha(\vec{v}) + \beta(-\frac{1}{4}\vec{w})$

$$A\vec{x} = \alpha\vec{v} - (\beta/4)\vec{w}$$

② $A^2\vec{x} = A(A\vec{x}) = A(\alpha\vec{v} - (\beta/4)\vec{w}) = \alpha(A\vec{v}) - (\beta/4)(A\vec{w})$

$$= \alpha(\vec{v}) - (\beta/4)(-\tfrac{1}{4}\vec{w})$$

$$A^2\vec{x} = \alpha\vec{v} + (\beta/16)\vec{w}$$

③ $A^3\vec{x} = A(A^2\vec{x}) = A(\alpha\vec{v} + (\beta/16)\vec{w}) = \alpha(A\vec{v}) + (\beta/16)(A\vec{w})$

$$= \alpha(\vec{v}) + (\beta/16)(-\tfrac{1}{4}\vec{w})$$

$$A^3\vec{x} = \alpha\vec{v} - (\beta/64)\vec{w}$$

c) For large m, the scalar multiplier of $\vec{w}$ will progressively get smaller until it is inconsequential. In other words, $\pm\beta/4^m$ approaches 0 as m continues to grow. Therefore, for large enough m, $(\pm\beta/4^m)\vec{w} \approx \vec{0}$, so $A^m\vec{x} \approx \alpha\vec{v}$.

Exercise 17.2. Assume that $f : \mathbf{R}^3 \to \mathbf{R}^3$ is a function with $f(1,1,2) = \begin{bmatrix} 2 \\ 2 \\ 1 \end{bmatrix}$ and $(Df)(1,1,2) = \begin{bmatrix} 1 & 3 & 1 \\ 2 & 4 & 3 \\ 1 & 1 & 0 \end{bmatrix}$. Let $g(x,y,z) = \sqrt{x^2+y^2+z^2}$. Calculate $(g \circ f)_y(1,1,2)$. (Hint: this is an entry in a certain derivative matrix.)

$$D(g \circ f)(1,1,2) = D_g(f(1,1,2)) \cdot Df(1,1,2)$$

$$= D_g(2,2,1) \cdot \begin{bmatrix} 1 & 3 & 1 \\ 2 & 4 & 3 \\ 1 & 1 & 0 \end{bmatrix}$$

$$D_g(x,y,z) = \begin{bmatrix} \frac{1}{2}(x^2+y^2+z^2)^{-1/2} \cdot 2x & \frac{1}{2}(x^2+y^2+z^2)^{-1/2} \cdot 2y & \frac{1}{2}(x^2+y^2+z^2)^{-1/2} \cdot 2z \end{bmatrix}$$

$$= \begin{bmatrix} \frac{x}{\sqrt{x^2+y^2+z^2}} & \frac{y}{\sqrt{x^2+y^2+z^2}} & \frac{z}{\sqrt{x^2+y^2+z^2}} \end{bmatrix}$$

$$D_g(2,2,1) = \begin{bmatrix} \frac{2}{3} & \frac{2}{3} & \frac{1}{3} \end{bmatrix}$$

y-component

$$D(g \circ f)(1,1,2) = \begin{bmatrix} 2/3 & 2/3 & 1/3 \end{bmatrix} \begin{bmatrix} 1 & 3 & 1 \\ 2 & 4 & 3 \\ 1 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 7/3 & 5 & 8/3 \end{bmatrix}$$

$$\boxed{(g \circ f)_y(1,1,2) = 5}$$

Exercise 17.3. Let $f : \mathbf{R}^2 \to \mathbf{R}^2$ be the function $f(x, y) = (x^3y^2 - y, xy^3 - x)$, so $f(1, 1) = (0, 0)$. By using the Chain Rule to compute $D(f \circ f)(1, 1)$, give the linear approximation to $(f \circ f)(1 + h, 1 + k)$ for h, k near 0. (If you try to compute $(f \circ f)(x, y) = f(f(x, y))$ explicitly with the aim of directly computing its partial derivatives at $(1, 1)$, you will get a total mess! This illustrates one important role for the multivariable Chain Rule in the machine learning algorithm called backpropagation that is discussed in Appendix G; see in particular the final paragraph of Section G.4.)

1) $f \circ f(1,1) = f(f(1,1)) = f(0,0) = (0,0)$

2) $$Df\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} 3x^2y^2 & 2yx^3 - 1 \\ y^3 - 1 & 3y^2x \end{bmatrix}$$

$$D(f \circ f)\left(\begin{bmatrix} 1 \\ 1 \end{bmatrix}\right) = Df\left(f\left(\begin{bmatrix} 1 \\ 1 \end{bmatrix}\right)\right) \cdot Df\left(\begin{bmatrix} 1 \\ 1 \end{bmatrix}\right)$$
$$= Df(0,0) \cdot Df(1,1)$$
$$= \begin{bmatrix} 0 & -1 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} 3 & 1 \\ 0 & 3 \end{bmatrix} = \begin{bmatrix} 0 & -3 \\ -3 & -1 \end{bmatrix}$$

3) $$(f \circ f)\left(\begin{bmatrix} 1+h \\ 1+k \end{bmatrix}\right) \approx (f \circ f)\left(\begin{bmatrix} 1 \\ 1 \end{bmatrix}\right) + D(f \circ f)\left(\begin{bmatrix} 1 \\ 1 \end{bmatrix}\right) \cdot \left(\begin{bmatrix} 1+h \\ 1+k \end{bmatrix} - \begin{bmatrix} 1 \\ 1 \end{bmatrix}\right)$$
$$= \begin{bmatrix} 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 & -3 \\ -3 & -1 \end{bmatrix} \begin{bmatrix} h \\ k \end{bmatrix}$$

$$\boxed{(f \circ f)(1+h, 1+k) = \begin{bmatrix} -3k \\ -3h - k \end{bmatrix}}$$

Exercise 17.5. Let $F : \mathbf{R}^2 \to \mathbf{R}$ be a function.

(a) Consider a level curve $F(x, y) = c$ defining y implicitly as a function $y(x)$ of x. By using the Chain Rule for $F(x, y(x)) = (F \circ h)(x)$ with $h(x) = (x, y(x))$ to compute the x-derivative of $F(x, y(x))$, deduce that wherever F_y is non-vanishing we have

$$\frac{dy}{dx} = -\frac{F_x(x, y)}{F_y(x, y)} = -\frac{\partial F/\partial x}{\partial F/\partial y}.$$

(This is a universal answer to *all* "implicit differentiation" problems in single-variable calculus!)

a) $F(x,y) = c \to F(x, y(x)) = c$

$(F \circ h)(x) = c \longrightarrow D(F \circ h)(x) = Dc = 0$

$D(F \circ h)(x) = DF(h(x)) \cdot Dh(x)$

$DF(h(x)) = DF(x,y) = \begin{bmatrix} F_x(x,y) & F_y(x,y) \end{bmatrix} = \begin{bmatrix} \partial F/\partial x & \partial F/\partial y \end{bmatrix}$

$Dh(x) = \begin{bmatrix} 1 \\ \frac{dy}{dx} \end{bmatrix}$

$D(F \circ h)(x) = \begin{bmatrix} F_x(x,y) & F_y(x,y) \end{bmatrix} \begin{bmatrix} 1 \\ \frac{dy}{dx} \end{bmatrix} = F_x(x,y) + F_y(x,y)\left(\frac{dy}{dx}\right)$

$F_x(x,y) + F_y(x,y)\left(\frac{dy}{dx}\right) = Dc = 0 \longrightarrow F_y(x,y)\left(\frac{dy}{dx}\right) = -F_x(x,y)$

If $F_y(x,y) \neq 0$, then $\boxed{\frac{dy}{dx} = \frac{-F_x(x,y)}{F_y(x,y)} = \frac{-\partial F/\partial x}{\partial F/\partial y}}$

(b) Use (a) to calculate the slope of the tangent line to the level curve $2x^3y - y^5x = 1$ at $(1, 1)$. This curve is shown in Figure 17.5.1 below (though the picture isn't used in the solution).

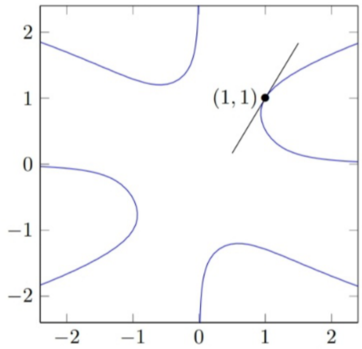

FIGURE 17.5.1. The curve $2x^3y - y^5x = 1$ (with 4 "parts") and its tangent line at $(1, 1)$.

If you look closely at the formula in (a), the minus sign is interesting because it shows that in the context of implicit functions if one tries to think about $\partial f/\partial u$ as if it were a fraction (which it is not, despite the notation) then one arrives at incorrect conclusions: consider "cancelling the ∂F's" on the right side of the formula in (a). A more striking sign occurs in Exercise 17.6(b).

b) $F(x,y) = 2x^3y - y^5x = 1 \to \begin{cases} F_x = 6x^2y - y^5 \\ F_y = 2x^3 - 5xy^4 \end{cases}$

$\left.\frac{dy}{dx}\right|_{(1,1)} = \left.\frac{-F_x(x,y)}{F_y(x,y)}\right|_{(1,1)} = \left.\frac{-(6x^2y - y^5)}{2x^3 - 5xy^4}\right|_{(1,1)} = \left.\frac{y^5 - 6x^2y}{2x^3 - 5xy^4}\right|_{(1,1)}$

$= \frac{(1)^5 - 6(1)^2(1)}{2(1)^3 - 5(1)(1)^4} = \frac{1-6}{2-5} = \frac{-5}{-3} = \boxed{\frac{5}{3}}$

Exercise 17.8. Let $f : \mathbf{R}^n \to \mathbf{R}$ and let $\mathbf{v} \in \mathbf{R}^n$ be a unit vector. The *directional derivative* of f in direction $\mathbf{v}$ at a point $\mathbf{a} \in \mathbf{R}^n$ is defined as

$$(D_{\mathbf{v}} f)(\mathbf{a}) = \frac{d}{dt}\Big|_{t=0} f(\mathbf{a} + t\mathbf{v}) \in \mathbf{R}$$

(the rate of change of f at time $t = 0$ when viewed as a function on the line through $\mathbf{a}$ in the direction of $\mathbf{v}$ that is parameterized at unit speed in that direction).

(a) By writing $f(\mathbf{a} + t\mathbf{v})$ as $(f \circ g)(t)$ for $g : \mathbf{R} \to \mathbf{R}^n$ defined by $g(t) = \mathbf{a} + t\mathbf{v}$, use the Chain Rule to show that $(D_{\mathbf{v}} f)(\mathbf{a}) = ((Df)(\mathbf{a}))\mathbf{v}$ (a matrix-vector product).

(b) You are in a hilly region where the height is described by $f(x, y) = \sin(\pi x y)$. You are at $\mathbf{a} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$.

What is the slope in direction directly northeast? (Hint: use the unit direction vector $\begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}$.)

a) $(D_v f)(a) = \frac{d}{dt}\Big|_{t=0} f(\vec{a} + t\vec{v})$

$= \frac{d}{dt}\Big|_{t=0} f(g(t))$

$= D(f \circ g)(0)$

$= Df(g(0)) \cdot Dg(0)$

$= Df(\vec{a}) \cdot [\vec{v}]$

$g(0) = \vec{a} + 0\vec{v} = \vec{a}$

$D_g(t) = [\vec{v}] \rightarrow D_g(0) = [\vec{v}]$

$$\boxed{(D_v f)(a) = ((Df)(\vec{a}))\vec{v}}$$

b) $\vec{v} = \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}$, $\vec{a} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$

$Df(x, y) = [\pi y \cos(\pi x y) \quad \pi x \cos(\pi x y)]$

$Df(\vec{a}) = Df(1, 2) = [\pi(2)\cos(\pi(1)(2)) \quad \pi(1)\cos(\pi(1)(2))]$

$= [2\pi \quad \pi]$

Slope $= (D_v f)(a) = ((Df)(\vec{a}))\vec{v} = [2\pi \quad \pi]\begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix} = [2\pi/\sqrt{2} + \pi/\sqrt{2}] = \boxed{\frac{3\pi}{\sqrt{2}}}$

Exercise 17.12. For a function $f : \mathbf{R}^2 \to \mathbf{R}$, define $h(r,\theta) = f(r\cos\theta, r\sin\theta)$. (Informally, h is "f written in polar coordinates".)

$f(x, y)$

(a) Use the Chain Rule in either the matrix form or (17.1.5) or (17.1.6) to show that

$$\frac{\partial h}{\partial r} = \cos\theta\,\frac{\partial f}{\partial x} + \sin\theta\,\frac{\partial f}{\partial y}, \quad \frac{\partial h}{\partial \theta} = -r\sin\theta\,\frac{\partial f}{\partial x} + r\cos\theta\,\frac{\partial f}{\partial y}.$$

By abuse of notation, $\partial h/\partial r$ and $\partial h/\partial\theta$ are usually written as $\partial f/\partial r$ and $\partial f/\partial\theta$ respectively.

(b) For $r > 0$, use the formulas in (a) to find functions $g_1(r,\theta)$ and $g_2(r,\theta)$ for which

$$\frac{\partial f}{\partial y} = g_1(r,\theta)\frac{\partial h}{\partial\theta} + g_2(r,\theta)\frac{\partial h}{\partial r}.$$

(Hint: use that $\sin^2\theta + \cos^2\theta = 1$.) Using the abuse of notation mentioned at the end of (a), the right side is usually written as $g_1(r,\theta)\,\partial f/\partial\theta + g_2(r,\theta)\,\partial f/\partial r$.

a) Suppose $x = r\cos\theta$ and $y = r\sin\theta$.

1)
$$\frac{\partial h}{\partial r} = \frac{\partial}{\partial r} f(r\cos\theta, r\sin\theta) = \frac{\partial}{\partial r} f(x,y)$$
$$= \sum_{k=1}^{2} \frac{\partial f}{\partial x_k}\frac{\partial x_k}{\partial r}$$
$$= \frac{\partial f}{\partial x}\frac{\partial x}{\partial r} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial r}$$
$$= \frac{\partial}{\partial r}(r\cos\theta)\frac{\partial f}{\partial x} + \frac{\partial}{\partial r}(r\sin\theta)\frac{\partial f}{\partial y}$$
$$\boxed{\frac{\partial h}{\partial r} = \cos\theta\frac{\partial f}{\partial x} + \sin\theta\frac{\partial f}{\partial y}}$$

2)
$$\frac{\partial h}{\partial \theta} = \frac{\partial}{\partial \theta} f(r\cos\theta, r\sin\theta) = \frac{\partial}{\partial \theta} f(x,y)$$
$$= \sum_{k=1}^{2} \frac{\partial f}{\partial x_k}\frac{\partial x_k}{\partial \theta}$$
$$= \frac{\partial f}{\partial x}\frac{\partial x}{\partial \theta} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial \theta}$$
$$= \frac{\partial}{\partial \theta}(r\cos\theta)\frac{\partial f}{\partial x} + \frac{\partial}{\partial \theta}(r\sin\theta)\frac{\partial f}{\partial y}$$
$$\boxed{\frac{\partial h}{\partial \theta} = -r\sin\theta\frac{\partial f}{\partial x} + r\cos\theta\frac{\partial f}{\partial y}}$$

b)
$$\frac{\partial h}{\partial r} = \cos\theta\frac{\partial f}{\partial x} + \sin\theta\frac{\partial f}{\partial y}$$
$$\cos\theta\frac{\partial f}{\partial x} = \frac{\partial h}{\partial r} - \sin\theta\frac{\partial f}{\partial y}$$
$$\frac{\partial f}{\partial x} = \frac{1}{\cos\theta}\frac{\partial h}{\partial r} - \frac{\sin\theta}{\cos\theta}\frac{\partial f}{\partial y}$$

$$\frac{\partial h}{\partial \theta} = -r\sin\theta\frac{\partial f}{\partial x} + r\cos\theta\frac{\partial f}{\partial y}$$
$$r\sin\theta\frac{\partial f}{\partial x} = r\cos\theta\frac{\partial f}{\partial y} - \frac{\partial h}{\partial \theta}$$
$$\frac{\partial f}{\partial x} = \frac{r\cos\theta}{r\sin\theta}\frac{\partial f}{\partial y} - \frac{1}{r\sin\theta}\frac{\partial h}{\partial \theta}$$

$$\frac{1}{\cos\theta}\frac{\partial h}{\partial r} - \frac{\sin\theta}{\cos\theta}\frac{\partial f}{\partial y} = \frac{r\cos\theta}{r\sin\theta}\frac{\partial f}{\partial y} - \frac{1}{r\sin\theta}\frac{\partial h}{\partial \theta}$$
$$\frac{1}{\cos\theta}\frac{\partial h}{\partial r} + \frac{1}{r\sin\theta}\frac{\partial h}{\partial \theta} = \frac{\cos\theta}{\sin\theta}\frac{\partial f}{\partial y} + \frac{\sin\theta}{\cos\theta}\frac{\partial f}{\partial y}$$
$$\frac{\partial f}{\partial y}\left(\frac{\cos\theta}{\sin\theta} + \frac{\sin\theta}{\cos\theta}\right) = \frac{1}{r\sin\theta}\frac{\partial h}{\partial \theta} + \frac{1}{\cos\theta}\frac{\partial h}{\partial r}$$
$$\frac{\partial f}{\partial y}\left(\frac{\cos^2\theta}{\sin\theta\cos\theta} + \frac{\sin^2\theta}{\sin\theta\cos\theta}\right) = \frac{1}{r\sin\theta}\frac{\partial h}{\partial \theta} + \frac{1}{\cos\theta}\frac{\partial h}{\partial r}$$
$$\frac{\partial f}{\partial y}\left(\frac{1}{\sin\theta\cos\theta}\right) = \frac{1}{r\sin\theta}\frac{\partial h}{\partial \theta} + \frac{1}{\cos\theta}\frac{\partial h}{\partial r}$$
$$\frac{\partial f}{\partial y} = \frac{\sin\theta\cos\theta}{r\sin\theta}\frac{\partial h}{\partial \theta} + \frac{\sin\theta\cos\theta}{\cos\theta}\frac{\partial h}{\partial r}$$
$$\frac{\partial f}{\partial y} = \frac{\cos\theta}{r}\frac{\partial h}{\partial \theta} + \sin\theta\frac{\partial h}{\partial r}$$

$$\boxed{g_1(r,\theta) = \frac{\cos\theta}{r}} \qquad \boxed{g_2(r,\theta) = \sin\theta}$$