Click here to visit folder of solution videos on course Canvas/Panopto page

Problem 1: Distributing money

Five people sit around a circular table, each with some amount of money in their wallets. Simultaneously, each person divides their money into three equal thirds and gives one third to the person on the left, one third to the person on the right, and keeps the remaining third for themselves.

Write down a 5×5 matrix A that describes this operation. In other words, label the people as $\#1$, $\#2$, etc. going around the table in some way (beginning with some choice of "first" person), so person $\#5$ is sitting next to person $\#1$. Then A should take as input a 5-vector $\mathbf{x}$ whose ith entry x_i is the amount of money the ith person has *before* doing this, and $A\mathbf{x}$ should be the 5-vector whose ith entry is the amount of money the ith person has *after* the operation. (The answer is a Markov matrix!)

Solution: If the "input" to the above operation is the vector $\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix}$, then the "output" should be $\begin{bmatrix} \frac{1}{3}x_5 + \frac{1}{3}x_1 + \frac{1}{3}x_2 \\ \frac{1}{3}x_1 + \frac{1}{3}x_2 + \frac{1}{3}x_3 \\ \frac{1}{3}x_2 + \frac{1}{3}x_3 + \frac{1}{3}x_4 \\ \frac{1}{3}x_3 + \frac{1}{3}x_4 + \frac{1}{3}x_5 \\ \frac{1}{3}x_4 + \frac{1}{3}x_5 + \frac{1}{3}x_1 \end{bmatrix}$. In terms of matrix-vector products, we may write this output vector as the product

$$\begin{bmatrix} \frac{1}{3}x_5 + \frac{1}{3}x_1 + \frac{1}{3}x_2 \\ \frac{1}{3}x_1 + \frac{1}{3}x_2 + \frac{1}{3}x_3 \\ \frac{1}{3}x_2 + \frac{1}{3}x_3 + \frac{1}{3}x_4 \\ \frac{1}{3}x_3 + \frac{1}{3}x_4 + \frac{1}{3}x_5 \\ \frac{1}{3}x_4 + \frac{1}{3}x_5 + \frac{1}{3}x_1 \end{bmatrix} = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & 0 & 0 & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 & 0 \\ 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 \\ 0 & 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & 0 & 0 & \frac{1}{3} & \frac{1}{3} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix}. \quad \text{Thus, the desired matrix is } \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & 0 & 0 & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 & 0 \\ 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 \\ 0 & 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & 0 & 0 & \frac{1}{3} & \frac{1}{3} \end{bmatrix}.$$

Problem 2: Migration

Assume there are 3 cities A, B, and C. Assume that in every given year

- 80% of the residents of A stay in A, while 10% move to B and 10% move to C.
- 70% of the residents of B stay in B, while 10% move to A and 20% move to C.
- 60% of the residents of C stay in C, while 10% move to A and 30% move to B.

(We disregard births and deaths, and people moving to or from other locations.)

(a) Write down the 3×3 Markov matrix for this process (where we use a "population vector" in $\mathbf{R}^3$ whose first entry is the population of A, second entry is the population of B, and third entry is the population of C). Why must its columns add up to 1, even if the given percentages for movement among the cities were changed?

(b) Assume that initially each of the cities has 10,000 inhabitants. How many inhabitants does each city have after 1 year?

(c) How many inhabitants does city B have after 2 years?

(d) Now assume that of the 80% of residents who stay in city A, 5% die every year (so 75% remain). Write down the matrix for this new process. Do its columns still add up to 1? If not, where does your argument for column sums in (a) break down in this new setting?

Solution:

(a) The Markov matrix is

$$M = \begin{bmatrix} 0.8 & 0.1 & 0.1 \\ 0.1 & 0.7 & 0.3 \\ 0.1 & 0.2 & 0.6 \end{bmatrix}.$$

The sum of its second column entries accounts for all former residents of city B, so this must add up to $100\% = 1$ even if the percentages for moving from city B were to change. The same reasoning applies to the first and third columns (using cities A and C respectively).

(b) We can calculate this population vector as

$$M \begin{bmatrix} 10000 \\ 10000 \\ 10000 \end{bmatrix} = \begin{bmatrix} 10000 \\ 11000 \\ 9000 \end{bmatrix}.$$

(c) We can calculate this population vector as

$$M \begin{bmatrix} 10000 \\ 11000 \\ 9000 \end{bmatrix},$$

but we only need the second entry, which is 11400.

(d) The new Markov matrix would be

$$M' = \begin{bmatrix} 0.75 & 0.1 & 0.1 \\ 0.1 & 0.7 & 0.3 \\ 0.1 & 0.2 & 0.6 \end{bmatrix}.$$

The first column does not add up to 1 any more because not all citizens of A will survive to the next year: the argument in (a) for column sums being 100% assumes that everyone in a given year contributes to the movement among cities next year (so no new people are added to or removed from the process); this doesn't work when people may exit the process (such as through death).

Problem 3: Some more matrix algebra

Consider the linear transformation $T : \mathbf{R}^3 \to \mathbf{R}^2$ given by projecting a vector $\mathbf{v} \in \mathbf{R}^3$ onto its first two components (viewed as a 2-vector), then reflecting that projection across the line $x + y = 0$ in $\mathbf{R}^2$, and finally adding to this the $45°$ clockwise rotation of the projection of $\mathbf{v}$ onto its last two components. Find the 2×3 matrix A that computes T.

Solution: We first find the 2×3 matrix B representing the composition of projection on the first two components followed by the reflection. We will do this by using that the ith column of B is the effect of the composition on $\mathbf{e}_i$. The line $x+y=0$ is the diagonal going from the upper left to the lower right corner, so the composition sends the 3-vector $\mathbf{e}_1$ to $\begin{bmatrix} 0 \\ -1 \end{bmatrix}$ and sends $\mathbf{e}_2$ to $\begin{bmatrix} -1 \\ 0 \end{bmatrix}$. Finally, $\mathbf{e}_3$ is killed by the projection. Hence, the matrix is

$$B = \begin{bmatrix} 0 & -1 & 0 \\ -1 & 0 & 0 \end{bmatrix}.$$

(You could also have arrived at this by computing matrices for the projection and the reflection and multiplying them in the correct order.)

Now we calculate the matrix C given by composing projection onto the last two components and the rotation. We'll do this by multiplying matrices. The projection matrix is $\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$, and the rotation matrix is $\frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}$, so

$$C = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} 0 & 1 & 1 \\ 0 & -1 & 1 \end{bmatrix}.$$

Altogether, the matrix A is given as

$$A = B + C = \begin{bmatrix} 0 & -1 + \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -1 & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}.$$

Problem 4: Fibonacci Numbers (Extra)

The *Fibonacci numbers* are the sequence of numbers $a_1, a_2, a_3, \ldots$ obtained given by starting with initial values $a_1 = 1$ and $a_2 = 1$ with each successive term being the sum of the two preceding terms; i.e., $a_{n+2} = a_{n+1} + a_n$ for $n \geq 1$.

(a) Write down the first 8 Fibonacci numbers.

(b) Fnd a (non-Markov!) 2×2 matrix M for which $\begin{bmatrix} a_{n+2} \\ a_{n+1} \end{bmatrix} = M \begin{bmatrix} a_{n+1} \\ a_n \end{bmatrix}$ for every $n \geq 1$.

(c) Obtain an expression for a_{n+4} in terms of a_{n+1} and a_n in two ways: first do it by directly feeding the defining formula for the sequence (each term is the sum of the two preceding terms) into itself a few times, and then do it by computing M^3. Check that you get the same expression via each method. (The second method is the more useful approach when a_{n+4} is replaced with a_{n+k} for k much much bigger than 4.)

Using techniques from later in the course, one can study powers of M to find a slick explicit formula for a_n in terms of a_1 and a_2 (which are just numbers) for any choice of those two initial values.

Solution:

(a) The Fibonacci numbers are given by 1,1,2,3,5,8,13,21,....

(b) We have $a_{n+2} = a_{n+1} + a_n$, so $\begin{bmatrix} a_{n+2} \\ a_{n+1} \end{bmatrix} = \begin{bmatrix} a_{n+1} + a_n \\ a_{n+1} \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} a_{n+1} \\ a_n \end{bmatrix}$. Hence, the matrix M is

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

(c) From the definition using the preceding two terms,

$$a_{n+4} = a_{n+3} + a_{n+2} = (a_{n+2} + a_{n+1}) + a_{n+2} = 2a_{n+2} + a_{n+1} = 2(a_{n+1} + a_n) + a_{n+1} = 3a_{n+1} + 2a_n.$$

Alternatively, we calculate $M^3 = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 3 & 2 \\ 2 & 1 \end{bmatrix}$. Since we have

$$\begin{bmatrix} a_{n+4} \\ a_{n+3} \end{bmatrix} = M^3 \begin{bmatrix} a_{n+1} \\ a_n \end{bmatrix},$$

we read off $a_{n+4} = 3a_{n+1} + 2a_n$.

Problem 5: Chain Rule I

Define the functions $f : \mathbf{R}^2 \to \mathbf{R}^2$, $g : \mathbf{R}^3 \to \mathbf{R}$, and $h : \mathbf{R} \to \mathbf{R}^2$ by

$$f(y, z) = \begin{bmatrix} yz + e^y \\ \ln(1 + y^2 z^2) \end{bmatrix}, \quad g(r, s, t) = rs + t, \quad h(v) = \begin{bmatrix} ve^{-v} \\ v^2 \end{bmatrix}.$$

(a) Compute all three derivative matrices $(Df)(y,z)$, $(Dg)(r,s,t)$, and $(Dh)(v)$; make sure your matrix has the correct number of rows and columns in each case. Also compute $(Df)(1,1)$.

(b) Compute $(D(h\circ g))(1,-1,1)$ in two ways: the Chain Rule, and by explicit computation of $(h\circ g)(r,s,t)$.

(c) Compute the y-partial derivative of $g(y,f(y,z)) = g(y,f_1(y,z),f_2(y,z))$ in two ways: (i) work out $g(y,f(y,z))$ in terms of y and z, and (ii) use that $g(y,f(y,z)) = (g\circ q)(y,z)$ for $q(y,z) = (y,f(y,z)) = (y,yz+e^y,\ln(1+y^2z^2)) \in \mathbf{R}^3$ and compute $D(g\circ q)$ by the Chain Rule (in which the desired partial derivative is a specific matrix entry).

The first method is certainly easier in this case, so the point is just to see how the Chain Rule organizes the work very differently (its real power is for more complicated situations than this).

Solution:

(a) Computing partial derivatives of component functions, we have

$$(Df)(y,z) = \begin{bmatrix} z+e^y & y \\ 2yz^2/(1+y^2z^2) & 2zy^2/(1+y^2z^2) \end{bmatrix}, \ (Dg)(r,s,t) = \begin{bmatrix} s & r & 1 \end{bmatrix}, \ (Dh)(v) = \begin{bmatrix} (1-v)e^{-v} \\ 2v \end{bmatrix}.$$

In particular, $(Df)(1,1) = \begin{bmatrix} 1+e & 1 \\ 1 & 1 \end{bmatrix}$.

(b) By the Chain Rule,

$$\begin{aligned} (D(h\circ g))(1,-1,1) = (Dh)(g(1,-1,1))\,(Dg)(1,-1,1) &= (Dh)(0)\begin{bmatrix} -1 & 1 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 1 \\ 0 \end{bmatrix}\begin{bmatrix} -1 & 1 & 1 \end{bmatrix} \\ &= \begin{bmatrix} -1 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix}, \end{aligned}$$

whereas a direct calculation of the composite function gives

$$(h\circ g)(r,s,t) = \begin{bmatrix} (rs+t)e^{-rs-t} \\ (rs+t)^2 \end{bmatrix},$$

so this has derivative matrix

$$\begin{bmatrix} (1-(rs+t))se^{-rs-t} & (1-(rs+t))re^{-rs-t} & (1-(rs+t))e^{-rs-t} \\ 2(rs+t)s & 2(rs+t)r & 2(rs+t) \end{bmatrix},$$

and evaluating this at $(r,s,t) = (1,-1,1)$ gives

$$\begin{bmatrix} (1-0)(-1)e^0 & (1-0)1e^0 & (1-0)e^0 \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} -1 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

(agreeing with the previous answer).

(c) We explictly compute

$$g(y,f(y,z)) = g(y,yz+e^y,\ln(1+y^2z^2)) = y(yz+e^y) + \ln(1+y^2z^2) = y^2z + ye^y + \ln(1+y^2z^2),$$

and the y-partial derivative is

$$2yz + ye^y + e^y + 2yz^2/(1+y^2z^2).$$

For the second method, this partial derivative is the first entry of the 1×2 derivative matrix of $g\circ q : \mathbf{R}^2 \to \mathbf{R}$. We'll express the full derivative matrix as a product of explicit matrices and then extract the desired matrix entry:

$$\begin{aligned} (D(g\circ q))(y,z) &= (Dg)(q(y,z))\,(Dq)(y,z) \\ &= (Dg)(y,yz+e^y,\ln(1+y^2z^2)) \begin{bmatrix} 1 & 0 \\ z+e^y & y \\ 2yz^2/(1+y^2z^2) & 2zy^2/(1+y^2z^2) \end{bmatrix} \\ &= \begin{bmatrix} yz+e^y & y & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ z+e^y & y \\ 2yz^2/(1+y^2z^2) & 2zy^2/(1+y^2z^2) \end{bmatrix}. \end{aligned}$$

The first entry of this product matrix is

$$(yz + e^y) + y(z + e^y) + 2yz^2/(1 + y^2z^2) = 2yz + e^y + ye^y + 2yz^2/(1 + y^2z^2),$$

which is the same as what we got the first way.

Problem 6: Chain Rule II

For a function $F(x, y)$, suppose $x = G(v, w)$ and $y = H(v, w)$ as expressed as functions of v and w, and that $v = k(r, s)$ and $w = \ell(r, s)$ are expressed as functions of r and s. Then $F(x, y)$ may be regarded as a function of r and s alone via such repeated substitutions. Explicitly:

$$F(x, y) = F(G(v, w), H(v, w)) = F(G(k(r, s), \ell(r, s)), H(k(r, s), \ell(r, s))) = \left(F \circ \begin{bmatrix} G \\ H \end{bmatrix} \circ \begin{bmatrix} k \\ \ell \end{bmatrix}\right)(r, s).$$

This comes up *all the time*: chains of dependencies of collections of variable on other collections of variables and so on.

Find an expression for $\frac{\partial F}{\partial r}$ in terms of partial derivatives of the functions: F (with respect to x and y), G and H (with respect to v and w), and k and ℓ (with respect to r and s).

Solution: We can multiply out matrices and extract a desired entry, or we can build up in stages from r back to (v, w) back to (x, y). First, we do the matrix method. We have to multiply three derivative matrices, corresponding to the three-fold composition:

$$F \circ \begin{bmatrix} G \\ H \end{bmatrix} \circ \begin{bmatrix} k \\ \ell \end{bmatrix}.$$

The Chain Rule gives a three-fold matrix product for computing $(DF)(r, s) = \begin{bmatrix} \frac{\partial F}{\partial r} & \frac{\partial F}{\partial s} \end{bmatrix}$, namely

$$(DF)(r, s) = (DF)(x, y) \begin{bmatrix} (DG)(v, w) \\ (DH)(v, w) \end{bmatrix} \begin{bmatrix} (Dk)(r, s) \\ (D\ell)(r, s) \end{bmatrix} = \begin{bmatrix} \frac{\partial F}{\partial x} & \frac{\partial F}{\partial y} \end{bmatrix} \begin{bmatrix} \frac{\partial G}{\partial v} & \frac{\partial G}{\partial w} \\ \frac{\partial H}{\partial v} & \frac{\partial H}{\partial w} \end{bmatrix} \begin{bmatrix} \frac{\partial k}{\partial r} & \frac{\partial k}{\partial s} \\ \frac{\partial \ell}{\partial r} & \frac{\partial \ell}{\partial s} \end{bmatrix}.$$

There are two ways to multiply out the 3-fold product (which give the same final output, since matrix multiplication is associative), and depending on how you carry it out you might arrive at a slightly different-looking expression, but when everything is expanded out it always gives the same final result (as it must).

If we first multiply the two 2×2 matrices and then multiply against the 1×2 matrix, the expression we get for the desired first entry $\frac{\partial F}{\partial r}$ is

$$\frac{\partial F}{\partial r} = \frac{\partial F}{\partial x}\left(\frac{\partial G}{\partial v}\frac{\partial k}{\partial r} + \frac{\partial G}{\partial w}\frac{\partial \ell}{\partial r}\right) + \frac{\partial F}{\partial y}\left(\frac{\partial H}{\partial v}\frac{\partial k}{\partial r} + \frac{\partial H}{\partial w}\frac{\partial \ell}{\partial r}\right).$$

If instead we multiply the first two matrices in the 3-fold product and then multiply that against the last matrix we get the expression

$$\frac{\partial F}{\partial r} = \left(\frac{\partial F}{\partial x}\frac{\partial G}{\partial v} + \frac{\partial F}{\partial y}\frac{\partial H}{\partial v}\right)\frac{\partial k}{\partial r} + \left(\frac{\partial F}{\partial x}\frac{\partial G}{\partial w} + \frac{\partial F}{\partial y}\frac{\partial H}{\partial w}\right)\frac{\partial \ell}{\partial r}.$$

These two outcomes are seen to coincide upon multiplying everything out.

Here is the "non-matrix" approach via repeated substitution (it is really just another way of expressing one of the above two matrix calculations, as we'll see at the end). That is, first considering F as a function of (v, w) (through the implicit dependence of (x, y) on (v, w)) which in turn is a function of (r, s) we have

$$\frac{\partial F}{\partial r} = \frac{\partial F}{\partial v}\frac{\partial v}{\partial r} + \frac{\partial F}{\partial w}\frac{\partial w}{\partial r}$$

where v and w as functions of r are respectively really k and ℓ. Hence, really the equation is

$$\frac{\partial F}{\partial r} = \frac{\partial F}{\partial v}\frac{\partial k}{\partial r} + \frac{\partial F}{\partial w}\frac{\partial \ell}{\partial r}.$$

Next, we have to take care of $\partial F/\partial v$ and $\partial F/\partial w$. These are given via the dependency through (x, y):

$$\frac{\partial F}{\partial v} = \frac{\partial F}{\partial x}\frac{\partial x}{\partial v} + \frac{\partial F}{\partial y}\frac{\partial y}{\partial v}$$

with x and y as functions of (v, w) really via G and H respectively, so the equation is more properly written as

$$\frac{\partial F}{\partial v} = \frac{\partial F}{\partial x}\frac{\partial G}{\partial v} + \frac{\partial F}{\partial y}\frac{\partial H}{\partial v}.$$

The case of the w-partial comes out identically upon replacing v with w everywhere.

Putting it all together, we have

$$\frac{\partial F}{\partial r} = \left(\frac{\partial F}{\partial x}\frac{\partial G}{\partial v} + \frac{\partial F}{\partial y}\frac{\partial H}{\partial v}\right)\frac{\partial k}{\partial r} + \left(\frac{\partial F}{\partial y}\frac{\partial G}{\partial w} + \frac{\partial F}{\partial y}\frac{\partial H}{\partial w}\right)\frac{\partial \ell}{\partial r}.$$

This is literally the same as what we got by the second way of organizing the 3-fold matrix multiplication. And if you look at things closely, you'll see that this allegedly "non-matrix" approach is really exactly the same algebraic work as that which arose in the second way of carrying out the 3-fold matrix multiplication.