## Problem 1: 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

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

(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 $1 \times 2$ 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 2: 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 $\dfrac{\partial F}{\partial r}$ 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 $\ell$ (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 $(DF)(r,s) = \begin{bmatrix} \dfrac{\partial F}{\partial r} & \dfrac{\partial F}{\partial s} \end{bmatrix}$, namely

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

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 $2 \times 2$ matrices and then multiply against the $1 \times 2$ matrix, the expression we get for the desired first entry $\dfrac{\partial F}{\partial r}$ 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 $\ell$. 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

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

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.

## Problem 3: Computations with inverses

(a) Which of the following matrices are invertible? For each of the invertible ones, write down the inverse and check that it works by multiplying in both orders to confirm that you get $\mathrm{I}_2$ each time:

$$A = \begin{bmatrix} 1 & 3 \\ 2 & 8 \end{bmatrix}, \quad B = \begin{bmatrix} 2 & 3 \\ 4 & 6 \end{bmatrix}, \quad C = \begin{bmatrix} -1 & -4 \\ 5 & -3 \end{bmatrix}.$$

(b) For the matrices

$$A = \begin{bmatrix} 4 & 1 & -1 \\ 0 & 5 & 1 \\ 0 & 0 & 6 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & -1 & 2 \\ 3 & -3 & 5 \\ 3 & -2 & 2 \end{bmatrix}$$

check that the matrices

$$A' = \begin{bmatrix} 1/4 & -1/20 & 1/20 \\ 0 & 1/5 & -1/30 \\ 0 & 0 & 1/6 \end{bmatrix}, \quad B' = \begin{bmatrix} 4 & -2 & 1 \\ 9 & -4 & 1 \\ 3 & -1 & 0 \end{bmatrix}$$

are respective inverses by multiplying to see that $A'A$ and $AA'$ both equal $\mathrm{I}_3$ and likewise for $B'B$ and $BB'$.

**Solution:**

(a) In the $2 \times 2$ case, invertibility of $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ is exactly the condition that the $ad - bc \neq 0$, in which case the inverse is

$$\frac{1}{ad-bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}.$$

For $A$, we have $1 \cdot 8 - 3 \cdot 2 = 2$, so the inverse is $\frac{1}{2}\begin{bmatrix} 8 & -3 \\ -2 & 1 \end{bmatrix}$. To see that this works, we multiply

$$\frac{1}{2}\begin{bmatrix} 8 & -3 \\ -2 & 1 \end{bmatrix}\begin{bmatrix} 1 & 3 \\ 2 & 8 \end{bmatrix} = \frac{1}{2}\begin{bmatrix} 8-6 & 24-24 \\ -2+2 & -6+8 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

The product in the other order is an extremely similar calculation even down to the level of the arithmetic for finding individual matrix entries (this is special to the $2 \times 2$ case).

For $B$, we have $2 \cdot 6 - 3 \cdot 4 = 0$, so $B$ is not invertible.

For $C$, we have $(-1) \cdot (-3) - (-4) \cdot 5 = 23$, so the inverse is $\frac{1}{23}\begin{bmatrix} -3 & 4 \\ -5 & -1 \end{bmatrix}$. To see that this works, we multiply

$$\frac{1}{23}\begin{bmatrix} -3 & 4 \\ -5 & -1 \end{bmatrix}\begin{bmatrix} -1 & -4 \\ 5 & -3 \end{bmatrix} = \frac{1}{23}\begin{bmatrix} 3+20 & 12-12 \\ 5-5 & 20+3 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

The product in the other order works out very similarly, as it did for $A$.

(b) These are all direct matrix multiplications, the main point of which is to see that computing products such as $A'A$ and $AA'$ in opposite orders involves *very different* calculations at the level of matrix entries, in contrast with the $2 \times 2$ cases in (a).

## Problem 4: Using inversion

(a) Consider the system of equations

$$4x + y - z = 7, \ \ 5y + z = -3, \ \ 6z = 2.$$

Explain why this is the same as $A\mathbf{x} = \mathbf{b}$ for $\mathbf{x} = \begin{bmatrix} x \\ y \\ z \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} 7 \\ -3 \\ 2 \end{bmatrix}$, and $A$ as in Problem 3(b). Then use your knowledge of $A^{-1}$ from there to "solve for $\mathbf{x}$" (hint: multiply both sides of the vector formulation by $A^{-1}$), and check that the solution you obtained really works. Does the method work if the constants on the right side of these equations change?

(b) Consider the system of equations

$$3x + 2y = 7, \ \ 2x + y = -3, \ \ x + y = 2.$$

Explain why this is the same as $M\mathbf{x} = \mathbf{b}$ for $\mathbf{x} = \begin{bmatrix} x \\ y \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} 7 \\ -3 \\ 2 \end{bmatrix}$, and $M = \begin{bmatrix} 3 & 2 \\ 2 & 1 \\ 1 & 1 \end{bmatrix}$.

Verify that $M'M = \mathrm{I}_2$ for $M' = \begin{bmatrix} 0 & 1 & -1 \\ 2 & -3 & 0 \end{bmatrix}$, and (via multiplication on the left by $M'$) that if $M\mathbf{x} = \mathbf{b}$ then $\mathbf{x} = M'\mathbf{b}$. But compute $M'\mathbf{b}$ explicitly and check that it does *not* actually satisfy the given system of equations. What went wrong?

**Solution:**

(a) Multiplying the matrix-vector product $A\mathbf{x}$ produces the given system of equations entry by entry, so $A\mathbf{x} = \mathbf{b}$ at the level of equating corresponding entries is indeed the given system of equations. Multiplying both sides by $A^{-1}$ gives $A^{-1}(A\mathbf{x}) = A^{-1}\mathbf{b}$, with the left side equal to $(A^{-1}A)\mathbf{x} = \mathrm{I}_3\mathbf{x} = \mathbf{x}$, so we get $\mathbf{x} = A^{-1}\mathbf{b}$. Explicitly, this is

$$\begin{bmatrix} 1/4 & -1/20 & 1/20 \\ 0 & 1/5 & -1/30 \\ 0 & 0 & 1/6 \end{bmatrix} \begin{bmatrix} 7 \\ -3 \\ 2 \end{bmatrix} = \begin{bmatrix} 2 \\ -2/3 \\ 1/3 \end{bmatrix}.$$

Plugging this into the given system of equations indeed works out:

$$8 - 2/3 - 1/3 = 7, \;\; 5(-2/3) + 1/3 = -9/3 = -3, \;\; 6/3 = 2.$$

The same method works in general, because for $\mathbf{x} = A^{-1}\mathbf{b}$ for any $\mathbf{b} \in \mathbf{R}^3$ we have $A\mathbf{x} = A(A^{-1}\mathbf{b}) = (AA^{-1})\mathbf{b} = \mathrm{I}_3\mathbf{b} = \mathbf{b}$. Note that to "derive" the solution we used that $A^{-1}A = \mathrm{I}_3$ but to check it works we used the equation $AA^{-1} = \mathrm{I}_3$ involving multiplication of $A$ and $A^{-1}$ in the *other* order.

(b) The translation into matrix language goes as in (a), and that $M'M = \mathrm{I}_2$ is a direct calculation. We compute $M'\mathbf{b} = \begin{bmatrix} -5 \\ 23 \end{bmatrix}$, which fails *all* of the equations in the system we're given.

The problem is that from the condition $\mathbf{x} = M'\mathbf{b}$ we calculate $M\mathbf{x} = M(M'\mathbf{b}) = (MM')\mathbf{b}$ and $MM'$ might not equal $\mathrm{I}_3$ (so $(MM')\mathbf{b}$ has no reason to equal $\mathbf{b}$). In general, when proceeding from an assumption that there is a solution to deduce what the solution must be, unless all steps were reversible (which isn't the case here: multiplying by $M'$ on the left winds up not being undone by multiplication by $M$ on the left, for example: $MM' \neq \mathrm{I}_3$) we *have to check* if what we obtain actually works. It could fail, as happened here: ultimately it was our *assumption* that there is a solution that was false.

Explicit calculation shows $MM'$ is very different from $\mathrm{I}_3$:

$$MM' = \begin{bmatrix} 4 & -3 & -3 \\ 2 & -1 & -2 \\ 2 & -2 & -1 \end{bmatrix}.$$

 