The derivative matrix

Riley Levy

August 14, 2018

1 Vectors

A n-dimensional vector is an ordered pair of n real numbers. This is written as

$$(x, y, z, \dots, \omega) = \begin{bmatrix} x \\ y \\ z \\ \vdots \\ \omega \end{bmatrix}$$

The space $\mathbb{R}^n$ of ordered real n-tuples is called the n-dimensional vector space. Vectors will be written in boldface: $\mathbf{x}$, matrices will be written as capitals: A, scalars will be written plainly: k.

Vectors are added by

$$(x_1, x_2, \dots, x_n) + (y_1, y_2, \dots y_n) = (x_1 + y_1, x_2 + y_2, \dots, x_n + y_n)$$

or written as column vectors is a special case of matrix addition

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} + \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} x_1 + y_1 \\ x_2 + y_2 \\ \vdots \\ x_n + y_n \end{bmatrix}$$

You can "scale" or "scalar multiply" a vector via

$$k(\mathbf{x}_1, \mathbf{x}_2, \dots \mathbf{x}_n) = (k\mathbf{x}_1, k\mathbf{x}_2, \dots k\mathbf{x}_n)$$

Geometrically, this stretches or compresses the vector by a factor of k, where k is a real number

You can combine vectors and scalars to get a new vector using these operations. This is called a linear combination. Formally, a linear combination of the vectors $\mathbf{x}_1, \dots, \mathbf{x}_n$ is

$$a_1\mathbf{x}_1 + a_2\mathbf{x}_2 + \dots + a_n\mathbf{x}_n$$

The standard basis of $\mathbb{R}^n$ is the set of vectors

$$\boldsymbol{e}_1 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \boldsymbol{e}_2 = \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}, \dots \boldsymbol{e}_n = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}$$

Any vector in $\mathbb{R}^n$ can be written as a unique linear combination of the standard basis vectors. Consider the example

$$(a, b, c) = a\boldsymbol{e}_1 + b\boldsymbol{e}_2 + c\boldsymbol{e}_3$$

Define the dot product of two vectors by

$$(x_1, x_2, \dots x_n) \cdot (y_1, y_2, \dots y_2) = x_1 y_1 + x_2 y_2 + \dots x_n y_n$$

Note the length of $\mathbf{x}$, $|\mathbf{x}| = \mathbf{x} \cdot \mathbf{x}$.

A function $\mathbf{f} = (f_1, f_2, \dots f_n)$ has derivative

$$\mathbf{f}' = (f_1', f_2', \dots f_n')$$

Problems

1. The function $t \mapsto (\cos t, \sin t)$ traces a unit circle with unit speed. Use the fact the tangent to a circle is perpendicular to the radius to show $\sin' = \cos$ and $\cos' = -\sin$
2. Derive the centripital acceleration of a particle moving in a circle of radius r with angular velocity ω. Its position as a function of time is

$$t \mapsto r(\cos \omega t, \sin \omega t)$$

2 Linear functions

A function $\mathbf{f}$ is said to be linear iff

- $\mathbf{f}$ distributes over vector addition:

$$\mathbf{f}(\mathbf{x} + \mathbf{y}) = \mathbf{f}(\mathbf{x}) + \mathbf{f}(\mathbf{y})$$

- $\mathbf{f}$ commutes with scalar multiplication

$$\mathbf{f}(k\mathbf{x}) = k\mathbf{f}(\mathbf{x})$$

equivalently, f is linear iff it preserves linear combinations:

$$\mathbf{f}(a_1\mathbf{x}_1 + a_2\mathbf{x}_2 + \cdots + a_n\mathbf{x}_n) = a_1\mathbf{f}(\mathbf{x}_1) + a_2\mathbf{f}(\mathbf{x}_2) + \cdots + a_n\mathbf{f}(\mathbf{x}_n)$$

It's not obvious yet but this is the generalization needed to make "local linearity" or "best linear approximation" meaningful in higher dimensions, which we use to make sense of the derivative.

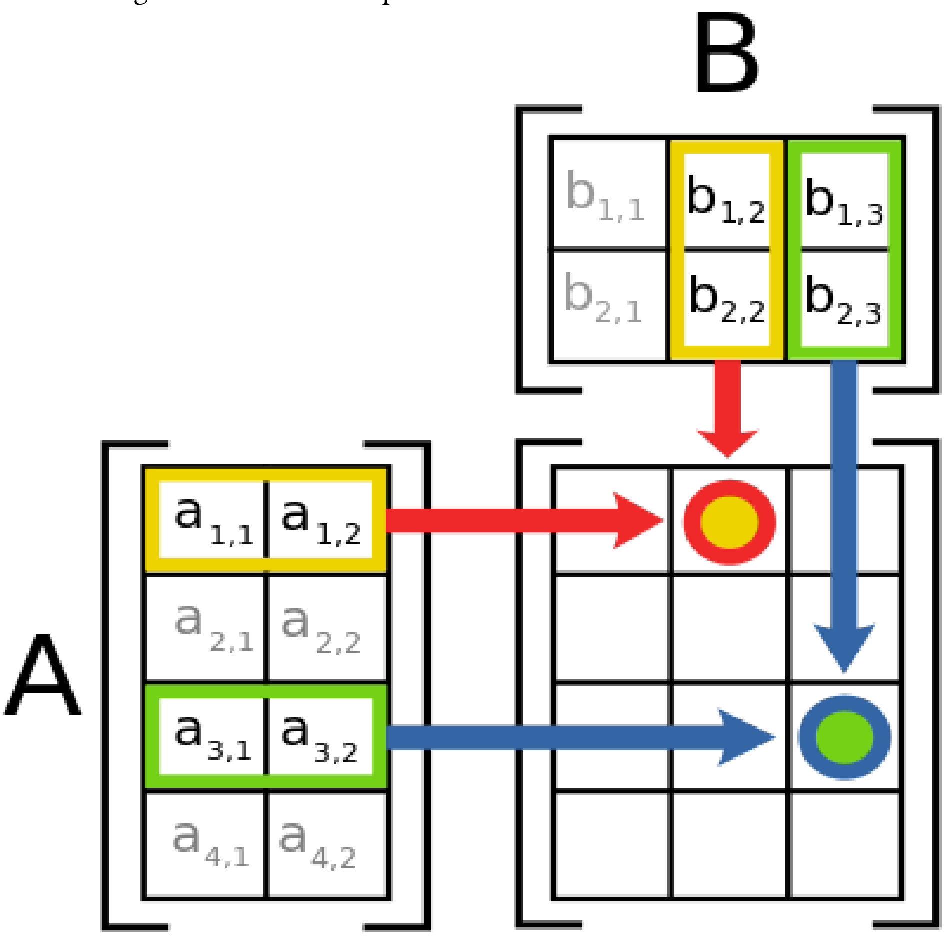
Some examples of linear maps are:

- scalar multiplication
- matrix multiplication
- rotations and dilations centered at the origin
- reflections in lines through the origin
- even d/dx and $\int$ are linear in some sense: they commute with constant multiplication and distribute over addition.

3 Matrices

Matrices are essentially coordinate representations for linear maps (in finite dimensional spaces).

Here's a diagram of matrix multiplication:



Matrix multiplication goes row dot-product column. So the red/yellow circle will have the value $(AB)_{12} = a_1 \cdot b_1$. The green/blue circle has value $(AB)_{33} = a_{31}b_{13} + a_{32}b_{23}$.

In general,

$$(AB)_{ij} = \sum_k a_{ik} b_{kj}$$

The dot product is a special case of matrix multiplication:

$$\mathbf{x} \cdot \mathbf{y} = \mathbf{x}^\top \mathbf{y}$$

where $\mathbf{x}^\top$ is the *transpose* of $\mathbf{x}$:

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}^\top = \begin{bmatrix} x_1 & x_2 & \dots & x_n \end{bmatrix}$$

In general,

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^\top = \begin{bmatrix} a & c \\ b & d \end{bmatrix}$$

The matrix

$$I = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{bmatrix}$$

is called the identity matrix. It is the multiplicative identity for matrices:

$$xI = Ix = x$$

Problems

A square matrix M has an inverse if there is some matrix M^{-1} such that

$$MM^{-1} = M^{-1}M = I$$

1. Show by multiplying with $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ the following do not have inverses: $\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$, $\begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}$, $\begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}$, $\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$.
2. Show you can represent complex numbers with the encoding

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

$$i \mapsto \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$$

More explicitly,

$$(a+bi)(c+di) \mapsto \left(a\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix} + b\begin{bmatrix}0 & -1\\ 1 & 0\end{bmatrix}\right)\left(c\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix} + d\begin{bmatrix}0 & -1\\ 1 & 0\end{bmatrix}\right)$$

gives the correct product. (Hint: Note $\left[\begin{smallmatrix}1 & 0\\ 0 & 1\end{smallmatrix}\right]$ is the identity matrix. Then it suffices to show $\left[\begin{smallmatrix}0 & -1\\ 1 & 0\end{smallmatrix}\right]^2 = -I$. The rest follows by linearity.)

3. Matrix multiplication is not commutative in general. Check that

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

3.1 Matrix times a vector

If you write a vector as a column matrix, then matrix $\times$ vector products are transformations of the plane, or n-space. Rotations, reflections, dilations are some examples of these transformations.

The number of columns of the matrix must match the dimension of the input vector

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

The number of rows of the matrix matches the dimension of the output. So (left-multiplying by) a 3 $\times$ 2 matrix is a function $\mathbb{R}^3 \to \mathbb{R}^2$.

In general, a matrix with n columns and m rows is a map $\mathbb{R}^n \to \mathbb{R}^m$.

Theorem 1 *The n-th column of a matrix is the image of the n-th standard basis vector.*

Proof. Here is the argument for the basis vector $(1,0,0)$ and a 3 $\times$ 3 matrix, but the other cases are essentially the same.

Consider the product with arbitrary matrix

$$\begin{bmatrix}a_1 & b_1 & c_1\\ a_2 & b_2 & c_2\\ a_3 & b_3 & c_3\end{bmatrix}\begin{bmatrix}1\\ 0\\ 0\end{bmatrix} = \begin{bmatrix}a_1\\ a_2\\ a_3\end{bmatrix}$$

Multiplying by the standard basis vector $(1,0,0)$ pulls out the first column. □

This result gives us a new way to look at matrix multiplication:

Theorem 2 *A matrix times the vector v is a linear combination of the columns with v determining weights:*

$$M\mathbf{v} = \begin{bmatrix}a_1 & b_1 & c_1\\ a_2 & b_2 & c_2\\ a_3 & b_3 & c_3\end{bmatrix}\begin{bmatrix}x\\ y\\ z\end{bmatrix} = x\mathbf{a} + y\mathbf{b} + z\mathbf{c}$$

where $\mathbf{a} = (a_1, a_2, a_3)$,$\mathbf{b} = (b_1, b_2, b_3)$,$\mathbf{c} = (c_1, c_2, c_3)$.

Proof. This follows from theorem 1 and linearity. Note we can decompose the vector into a linear combination of standard basis vectors

$$\mathbf{v} = \begin{bmatrix} x \\ y \\ z \end{bmatrix} = x \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + y \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + z \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

Then by linearity

$$M\mathbf{v} = M \left(x \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + y \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + z \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right) = xM \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + yM \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + zM \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

By theorem 1, this is

$$x \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} + y \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} + z \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} = x\mathbf{a} + y\mathbf{b} + z\mathbf{c}$$

□

So a matrix × a vector can be viewed as the vector determining a linear combination of the matrix's columns. (You can adapt this argument to show that every linear function between finite dimensional vector spaces can be represented as a matrix acting on a vector).

Theorems 1 and 2 tell us how to construct the corresponding matrix for a linear function we have defined by other means. For example, we can calculate the matrix that rotates a point in the plane θ degrees clockwise around the origin. We only need to consider the action on the standard basis vectors, like so

$$\begin{bmatrix} 1 \\ 0 \end{bmatrix} \mapsto \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix}$$
$$\begin{bmatrix} 0 \\ 1 \end{bmatrix} \mapsto \begin{bmatrix} -\sin\theta \\ \cos\theta \end{bmatrix}$$

Hence the matrix

$$R_\theta = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

This gives a quick way to calculate planar rotations.

Problems

1. What is the corresponding matrix for reflection in the line $y = 0$ in the plane? $x = 0$? $x = y$?
2. Rotation and reflection do not commute. Use your answer from problem 1 to find a pair of noncommuting matrices.
3. Show that translation is not linear, and therefore cannot be represented as a matrix as discussed in section 3.1.

4. However, translation in $\mathbb{R}^2$ can be considered the projection of a linear map from $\mathbb{R}^3 \to \mathbb{R}^3$ restricted to the plane given by $z = 1$. Find its matrix. (Hint: encode the vector (x, y) as $(x, y, 1)$)

5. Use the fact that translation and reflection don't commute, along with the answer to problem 4, to find another pair of noncommuting matrices.

6. In problem 2 (page 4), you encoded complex numbers as 2×2 matrices. What does left-multiplying by this matrix do to a vector?

3.2 Matrix times a matrix

Now we have a geometric interpretation of a matrix $\times$ a vector as a linear map acting on the vector and a way to find the associated matrix to a linear map we have a non-matrix description of, such as rotation.

But we have still been looking at matrix $\times$ a matrix as a purely symbolic computation. It, in fact, has a related meaning.

You can think of a matrix $\times$ a matrix as a *composition of functions*, that may eventually act on a vector. This works because matrix multiplication is associative:

$$(AB)\mathbf{x} = A(B\mathbf{x})$$

where A, B are matrices and $\mathbf{x}$ is a vector.

Alternately, we can understand the product

$$AB$$

by considering it the accumulation of A's action on each column of B:

$$A\begin{bmatrix}\mathbf{b}_1 & \mathbf{b}_2 & \mathbf{b}_3 & \dots & \mathbf{b}_n\end{bmatrix} = \begin{bmatrix}A\mathbf{b}_1 & A\mathbf{b}_2 & A\mathbf{b}_3 & \dots & A\mathbf{b}_n\end{bmatrix}$$

This is probably clearest when considering an example like the following:

$$\begin{bmatrix}a_1 & a_2 \\ b_1 & b_2 \\ c_1 & c_2\end{bmatrix}\begin{bmatrix}x_1 & y_1 \\ x_2 & y_2\end{bmatrix} = \begin{bmatrix}\mathbf{a}\cdot\mathbf{x} & \mathbf{a}\cdot\mathbf{y} \\ \mathbf{b}\cdot\mathbf{x} & \mathbf{b}\cdot\mathbf{y} \\ \mathbf{c}\cdot\mathbf{x} & \mathbf{c}\cdot\mathbf{y}\end{bmatrix}$$

What is the link between the two models given for matrix $\times$ a matrix?

4 Derivatives

In this section we will use the symbols $\mathbf{e}_n$ to represent the n-th standard basis element.

4.1 Partial derivatives

Consider the function $\mathbf{f}(x_1, x_2, \dots x_n) : \mathbb{R}^n \to \mathbb{R}^m$. We define the partial derivatives:

$$\begin{aligned}
\frac{\partial \mathbf{f}}{\partial x_1} &= \lim_{h\to 0} \frac{\mathbf{f}(x_1+h, x_2, \dots x_n) - \mathbf{f}(x_1, x_2, \dots x_n)}{h} = \left(t \mapsto \mathbf{f}(t, x_2, \dots, x_n)\right)' \\
\frac{\partial \mathbf{f}}{\partial x_2} &= \lim_{h\to 0} \frac{\mathbf{f}(x_1, x_2+h, \dots x_n) - \mathbf{f}(x_1, x_2, \dots x_n)}{h} = \left(t \mapsto \mathbf{f}(x_1, t, \dots x_n)\right) \\
&\vdots \\
\frac{\partial \mathbf{f}}{\partial x_n} &= \lim_{h\to 0} \frac{\mathbf{f}(x_1, x_2, \dots x_n+h) - \mathbf{f}(x_1, x_2, \dots x_n)}{h} = \left(t \mapsto \mathbf{f}(x_1, x_2, \dots, t)\right)'
\end{aligned}$$

effectively holding all the variables you're not differentiating with respect to constant. The partial derivative tells us how $\mathbf{f}$ changes when we travel small distances parallel to the axes. Note that $\mathbf{f}$ may be a vector, in which case $\partial f/\partial x_i$ is a vector as well.

Problems

1. Find the partial derivatives of $(x, y) \mapsto x^2 + y^2$.
2. Generalize the results of problem 1 to arbitrary dimensions. (Hint: use symmetry)
3. Find all partial derivatives of $\mathbf{x} \mapsto |\mathbf{x}|$ where x is a vector in $\mathbb{R}^n$. (Hint: use the chain rule to reduce this to problem item 2)
4. Find all partial derivatives of

$$(x, y, z) \mapsto \left(\frac{x}{1-z}, \frac{y}{1-z}\right)$$

4.2 Directional derivatives

We can generalize the idea of taking a derivative *along a direction* as follows:

Consider $\mathbf{f}(\mathbf{x}) : \mathbb{R}^n \mapsto \mathbb{R}^m$

$$\partial_{\mathbf{v}} \mathbf{f} = \lim_{t\to 0} \frac{\mathbf{f}(\mathbf{x} + \mathbf{v}t) - \mathbf{f}(\mathbf{x})}{t} = \left(t \mapsto \mathbf{f}(\mathbf{x} + t\mathbf{v})\right)'$$

where $t \in \mathbb{R}$.

The partial derivative is a special case of the directional derivative:

$$\frac{\partial}{\partial x_n} = \partial_{\mathbf{e}_n}$$

4.3 Derivative matrix

In the previous section, the way $\partial_{\boldsymbol{v}}$ depends on $\boldsymbol{v}$ is not clear. But because of local linearity, when a function is differentiable we expect that $\partial_{\boldsymbol{v}}\mathbf{f}(\boldsymbol{x})$ is linear in $\boldsymbol{v}$. This is usually how differentiability of $\mathbf{f}$ is defined.

But then

$$\boldsymbol{v} \mapsto \partial_{\boldsymbol{v}}\mathbf{f}(\boldsymbol{x}) : \mathbb{R}^n \to \mathbb{R}^m$$

is a linear map between finite dimensional vector spaces, so we can calculate its matrix. We do this by looking at its image on a standard basis:

$$\partial_{\boldsymbol{e}_i}\mathbf{f}(\boldsymbol{x}) = \left.\frac{\partial \mathbf{f}}{\partial x_i}\right|_x$$

We make each

$$\frac{\partial \mathbf{f}}{\partial x_i}$$

the column of the matrix:

$$f'(x) = Df(x) = \begin{bmatrix} \frac{\partial f}{\partial x_1} & \frac{\partial f}{\partial x_2} & \cdots & \frac{\partial f}{\partial x_n} \end{bmatrix}$$

and

$$\mathbf{f}'(\boldsymbol{x})\boldsymbol{v} = D\mathbf{f}(\boldsymbol{x})(\boldsymbol{v}) = \partial_{\boldsymbol{v}}\mathbf{f}(\boldsymbol{x})$$

The derivative matrix is sometimes called the *derivative operator* (especially in functional analysis) or the *total derivative* (especially in geometry).

The D used here is the notation Spivak uses in *Calculus on Manifolds* The notation df is also common. Using $df|_x$ to indicate the derivative is evaluated at the point x is common, especially in geometry.

4.4 Chain rule

Armed with the definition of a derivative matrix, we can now state the chain rule

$$(\mathbf{f} \circ \mathbf{g})'(\boldsymbol{x}) = \mathbf{f}'\big(\mathbf{g}(\boldsymbol{x})\big)\mathbf{g}'(\boldsymbol{x})$$

or, in Spivak's notation,

$$D(\mathbf{f} \circ \mathbf{g})(\boldsymbol{x}) = D\mathbf{f}\big(\mathbf{g}(\boldsymbol{x})\big) \circ D\mathbf{g}(\boldsymbol{x})$$

This is almost identical to the one dimensional case:

$$(f \circ g)'(x) = f'\big(g(x)\big)g(x)$$

The differences are

- The functions $\mathbf{f}$,$\mathbf{g}$ now take in vectors and output vectors.
- The variable $\boldsymbol{x}$ is now a vector.

- The multiplication between $f'\big(g(x)\big)$ and $g'(x)$ is now matrix multiplication (or composition of linear maps). This operation *is not commutative* in general.

We can derive the chain rule for partial derivatives as a special case:

$$\frac{\partial}{\partial x_n}(\mathbf{f} \circ \mathbf{g}) = (\mathbf{f} \circ \mathbf{g})'(\mathbf{x})\mathbf{e}_n = \mathbf{f}'\big(\mathbf{g}(\mathbf{x})\big)\mathbf{g}'(\mathbf{x})\mathbf{e}_n = \mathbf{f}'\big(\mathbf{g}(\mathbf{x})\big)\frac{\partial \mathbf{g}}{\partial x_n}(\mathbf{x})$$

and

$$\mathbf{f}'\big(\mathbf{g}(x)\big)\frac{\partial \mathbf{g}}{\partial x_n} = \sum_i \frac{\partial \mathbf{f}}{\partial x_i}\big(\mathbf{g}(\mathbf{x})\big)\frac{\partial g_i}{\partial x_n}(\mathbf{x})$$

where $\mathbf{g} = (g_1, g_2, \ldots, g_n)$.

Problems

1. Calculate the matrix of derivatives of the map

$$f(x, y) = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

2. Find the matrix of partial derivatives of the map $\mathbf{x} \mapsto \mathbf{x} \cdot \mathbf{x}$ using problem 2 (page 8).

3. Calculate $\left(x \mapsto \sqrt{x \cdot x}\right)'$ using the chain rule for derivative matrices and the result from problem 2 on page 8.

4. The function

$$\mathbf{f}(\mathbf{x}) = \left(\mathbf{x}, \sqrt{1 - \mathbf{x} \cdot \mathbf{x}}\right)$$

projects a vector onto the surface of a unit sphere centered at 0. Find its derivative.

4.5 Derivative matrices in their own right

Although partial derivatives give a useful way to compute the matrix derivative, we haven't actually given a careful definition of this. Spivak's *Calculus on manifolds* provides one. Here is that same definition, rewritten to use asymptotic notation:

We say

$$a = b + o(h)$$

when

$$\lim_{h \to 0} \frac{a - b}{h} = 0$$

You can think of $o(h)$ as absorbing all terms that are *sublinear* in h (approach 0 faster than h, in the sense made precise above). This notation makes it convenient to work with local linearity.

Now we define the derivative of $\mathbf{f}$ at $\mathbf{x}$ as the unique linear function $\mathbf{f}'(\mathbf{x})$ such that

$$\mathbf{f}(\mathbf{x} + \mathbf{h}) = \mathbf{f}(\mathbf{x}) + \mathbf{f}'(\mathbf{x})\mathbf{h} + o(|\mathbf{h}|)$$

This makes computing some derivatives easier. For example, we can compute the derivative of $f(x) = x \cdot x$ quickly:

$$(x + h) \cdot (x + h) = \underbrace{x \cdot x}_{f(x)} + \underbrace{2x \cdot}_{f'} h + \underbrace{h \cdot h}_{o(|h|)}$$

So $\mathbf{f}'(x) = 2x^\top$.

It also helps us prove

Theorem 3 (chain rule)

$$(\mathbf{f} \circ \mathbf{g})'(x) = \mathbf{f}'\big(\mathbf{g}(x)\big)\mathbf{g}'(x)$$

Proof.

$$\begin{aligned}
\mathbf{f}(\mathbf{g}(x + h)) &= \mathbf{f}\Big(\mathbf{g}(x) + \mathbf{g}'(x)h + o(|h|)\Big) \\
&= \mathbf{f}\big(\mathbf{g}(x)\big) + \mathbf{f}'\big(\mathbf{g}(x)\big)\Big(\mathbf{g}'(x)h + o(|h|)\Big) + \underbrace{o\big(|\mathbf{g}'(x)h| + |h|\big)}_{o(|h)} \\
&= \mathbf{f}(\mathbf{g}(x)) + \mathbf{f}'\big(\mathbf{g}(x)\big)\mathbf{g}'(x)h + \underbrace{\mathbf{f}'\big(\mathbf{g}(x)\big)o(|h|) + o(|h|)}_{o(|h|)} \\
&= (\mathbf{f} \circ \mathbf{g})(x) + \underbrace{\mathbf{f}'\big(\mathbf{g}(x)\big)\mathbf{g}'(x)}_{(\mathbf{f}\circ\mathbf{g})'} h + o(|h|)
\end{aligned}$$

□

Problems

1. Suppose $\mathbf{f}$ is a linear function. What is $\mathbf{f}'$?

2. Prove the product rule

$$(\mathbf{f}(x)g(x))' = \mathbf{f}'(x)g(x) + \mathbf{f}(x)g'(x)$$

where g is a scalar. But g' may not be, so order matters. (Hint: expand $\mathbf{f}(x + h)g(x + h)$ using local linearity, analogous to the proof of the chain rule.)

3. Derive the quotient rule

$$\left(\frac{\mathbf{f}(x)}{g(x)}\right)' = \frac{\mathbf{f}'(x)g(x) - \mathbf{f}(x)g'(x)}{(g(x))^2}$$

where g is a scalar.

4. Using the quotient rule, find

$$\left(x \mapsto \frac{x}{1 - x \cdot x}\right)'$$