Multivariable Vector-Valued Functions

Contents

31.1 Multivariable Vector-Valued Functions

In *Calculus I*, we studied functions of the form $y = f(x)$, for example $f(x) = x^2$. Such functions, which have a single variable for input and a real number for output, are called **single-variable real-valued** functions. The graph of such a function is a curve in the xy-plane, as seen for example in Figure 1 of this section, where the function is $f(x) = x^2$.

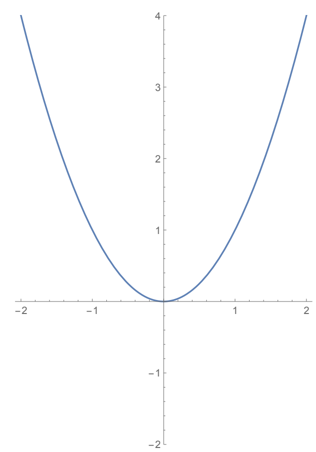


Figure 1: Graph of $f(x) = x^2$

If we use the standard symbol $\mathbb{R}$ to denote the set of real numbers, then functions of this type can be denoted as $f : \mathbb{R} \to \mathbb{R}$, where the name of the function is "f" and where the arrow indicates that the function takes input in $\mathbb{R}$ and gives output in $\mathbb{R}$. This arrow notation, which you might not have used up till now, is widely used in advanced mathematics.

In *Calculus II*, we added a new type of function to our repertoire, including functions of the form $z = f(x, y)$, for example $f(x, y) = x^2 - y^2$, and of the form $w = g(x, y, z)$, for example $g(x, y, z) = 5xy^2z^3$. Such functions, which have more than one variable for input and a real number for output, are called **multivariable real-valued** functions. The graph of a function of the form $z = f(x, y)$ is a surface in $\mathbb{R}^3$, as seen for example in Figure 2 of this section, where the function is $f(x, y) = x^2 - y^2$.

If we use the standard symbol $\mathbb{R}^n$ to denote n-dimensional space with real coordinates, then functions of this type can be denoted as $f : \mathbb{R}^n \to \mathbb{R}$, where, as before, the name of the function is "f" and where the arrow indicates that the function takes input in $\mathbb{R}^n$ and gives output in $\mathbb{R}$. For

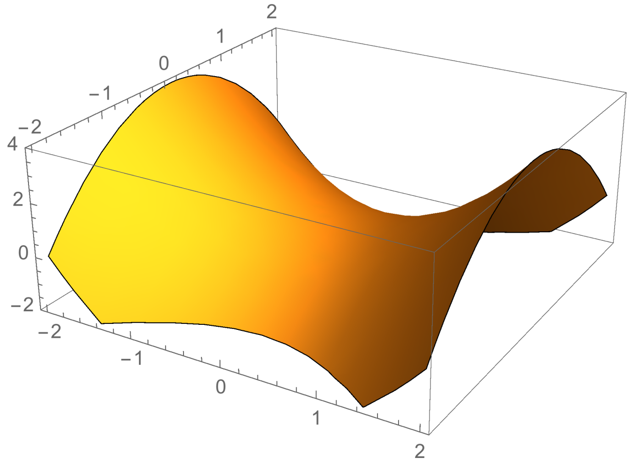


Figure 2: Graph of $f(x, y) = x^2 - y^2$

example, the function $f(x, y) = x^2 - y^2$ is a function $f\colon \mathbb{R}^2 \to \mathbb{R}$, and the function $g(x, y, z) = 5xy^2z^3$ is a function $g\colon \mathbb{R}^3 \to \mathbb{R}$.

Yet a third type of function was discussed earlier in *Vector Calculus*, where we considered functions of the form $\boldsymbol{r}(t) = \begin{bmatrix} f(t) \\ g(t) \end{bmatrix}$, for example $\boldsymbol{r}(t) = \begin{bmatrix} t^2 \\ 3t+1 \end{bmatrix}$, and of the form $\boldsymbol{s}(t) = \begin{bmatrix} f(t) \\ g(t) \\ h(t) \end{bmatrix}$, for example $\boldsymbol{s}(t) = \begin{bmatrix} \cos t \\ \sin t \\ 2t \end{bmatrix}$. Such functions, which have one variable for input and a vector for output, are called **single-variable vector-valued** functions. The functions $f(t)$ and $g(t)$, which are the component functions of $\boldsymbol{r}(t)$, are each a single-variable real-valued function, and similarly for the component functions of $\boldsymbol{s}(t)$.

Graphing single-variable vector-valued functions is different from graphing single-variable real-valued and multivariable real-valued functions. We graph functions of the form $y = f(x)$ and $z = f(x, y)$ by having a separate axis for each of the variables, both the "input" variables and the "output" variable. By contrast, for graphing single-variable vector-valued functions we do not given an axis for the "input" variable, for example the variable t in the function $\boldsymbol{s}(t) = \begin{bmatrix} \cos t \\ \sin t \\ 2t \end{bmatrix}$, but rather we have axes only for the "output" variables (in this example three axes). So, the picture we see of a single-variable vector-valued function is technically not the graph of the function, but rather the image of the function. See Figure 3 of this section for the image of $\boldsymbol{s}(t) = \begin{bmatrix} \cos t \\ \sin t \\ 2t \end{bmatrix}$.

Single-variable vector-valued functions can be denoted as $\boldsymbol{r}\colon \mathbb{R} \to \mathbb{R}^n$, where, as before, the name of the function is "$\boldsymbol{r}$" and where the arrow indicates that the function takes input in $\mathbb{R}$ and gives output in $\mathbb{R}^n$. For example, the function $\boldsymbol{r}(t) = \begin{bmatrix} t^2 \\ 3t+1 \end{bmatrix}$ is a function $\boldsymbol{r}\colon \mathbb{R} \to \mathbb{R}^2$, and the function $\boldsymbol{s}(t) = \begin{bmatrix} \cos t \\ \sin t \\ 2t \end{bmatrix}$ is a function $\boldsymbol{s}\colon \mathbb{R} \to \mathbb{R}^3$.

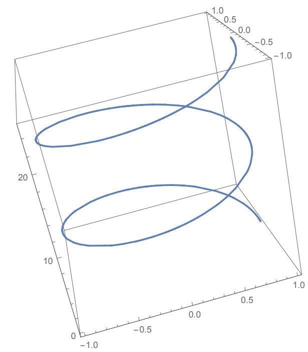


Figure 3: Graph of $s(t)$

We now want to introduce a new type of function that includes, and generalizes, all three of the previous types of functions.

The Definition of Multivariable Vector-Valued Functions

The new type of function we consider, called **multivariable vector-valued** functions, are functions of the form $\boldsymbol{F}\colon \mathbb{R}^n \to \mathbb{R}^m$, where n and m are positive integers.

There is an unfortunate ambiguity in the way we think about functions of the form $\boldsymbol{F}\colon \mathbb{R}^n \to \mathbb{R}^m$ because there is an ambiguity in the way we think about each of $\mathbb{R}^n$ and $\mathbb{R}^m$. Specifically, we can think of each of $\mathbb{R}^n$ and $\mathbb{R}^m$ as consisting of either points or vectors; these two points of view are each useful in different situations. Hence, a function of the form $\boldsymbol{F}\colon \mathbb{R}^n \to \mathbb{R}^m$ could be thought of as taking points to points, or points to vectors, or vectors to points, or vectors to vectors. Again, these different points of view are each useful in different situations. Most often we will think of a function of the form $\boldsymbol{F}\colon \mathbb{R}^n \to \mathbb{R}^m$ as taking points to vectors, though we will feel free to think of such a function in one of the other ways as needed.

Because we will most often think of a function of the form $\boldsymbol{F}\colon \mathbb{R}^n \to \mathbb{R}^m$ as taking points to vectors, we will write such functions with that point of view in mind.

For example, a multivariable vector-valued function $\boldsymbol{G}\colon \mathbb{R}^2 \to \mathbb{R}^3$ might be given by the formula $\boldsymbol{G}(x, y) = \begin{bmatrix} x+y \\ 3x^2 \\ 2x+\sin y \end{bmatrix}$. Then, for example, we see that $\boldsymbol{G}(3, 0) = \begin{bmatrix} 3 \\ 27 \\ 6 \end{bmatrix}$. Observe that the "input" of this function is the point (x, y), and the "output" is the vector $\begin{bmatrix} x+y \\ 3x^2 \\ 2x+\sin y \end{bmatrix}$.

We note that if we wanted to write multivariable vector-valued functions absolutely properly, we would write $\boldsymbol{G}((x, y))$ rather than $\boldsymbol{G}(x, y)$, where the inner parentheses in the notation $\boldsymbol{G}((x, y))$ refer to the point (x, y), and the outer parentheses in $\boldsymbol{G}((x, y))$ are the standard parentheses used in the notation of functions. However, whereas it would be proper to write $\boldsymbol{G}((x, y))$, we will stick with the more standard abbreviation $\boldsymbol{G}(x, y)$.

In general, we have the following definition.

Multivariable Vector-Valued Functions

1. A **multivariable vector-valued function** is a function $\boldsymbol{F}\colon \mathbb{R}^n \to \mathbb{R}^m$ for some positive integers n and m.

2. A function $\boldsymbol{F}\colon \mathbb{R}^n \to \mathbb{R}^m$ has the form

$$\boldsymbol{F}(x_1, x_2, \dots, x_n) = \begin{bmatrix} f_1(x_1,x_2,\dots,x_n) \\ f_2(x_1,x_2,\dots,x_n) \\ \vdots \\ f_m(x_1,x_2,\dots,x_n) \end{bmatrix},$$

   where $f_1, f_2, \dots, f_m\colon \mathbb{R}^n \to \mathbb{R}$ are multivariable real-valued functions. The functions $f_1, f_2, \dots, f_m$ are called the **component functions** of $\boldsymbol{F}$.

3. A function $\boldsymbol{F}\colon \mathbb{R}^2 \to \mathbb{R}^3$ has the form

$$\boldsymbol{F}(x, y) = \begin{bmatrix} P(x,y) \\ Q(x,y) \\ R(x,y) \end{bmatrix}.$$

We note that for a multivariable vector-valued function $\boldsymbol{F}\colon \mathbb{R}^n \to \mathbb{R}^m$, the two numbers n and m can be either equal or unequal. For example, we could have functions $\boldsymbol{F}\colon \mathbb{R}^2 \to \mathbb{R}^2$ and $\boldsymbol{G}\colon \mathbb{R}^3 \to \mathbb{R}^3$. The type of function where $n = m$ is particularly useful, and will be viewed at times as taking points to points and at other times as taking points to vectors.

A single variable is a number in $\mathbb{R}$, and multiple variables are points in $\mathbb{R}^n$ for some positive integer n. Observe that we can, in particular, use the value $n = 1$ when we consider $\mathbb{R}^n$, where $\mathbb{R}^1$ is just another name for $\mathbb{R}$. If we look functions of the form $\boldsymbol{F}\colon \mathbb{R}^n \to \mathbb{R}^m$ in the special case when $n = 1$, then we obtain functions of the form $\boldsymbol{F}\colon \mathbb{R} \to \mathbb{R}^m$, which are just single-variable vector-valued functions. Hence single-variable vector-valued functions are a special case of multivariable vector-valued functions.

Similarly, if we look functions of the form $\boldsymbol{F}\colon \mathbb{R}^n \to \mathbb{R}^m$ in the special case when $m = 1$, then we obtain functions of the form $\boldsymbol{F}\colon \mathbb{R}^n \to \mathbb{R}$, which are just multivariable real-valued functions. Hence multivariable real-valued functions are also a special case of multivariable vector-valued functions.

Clearly, when we use both $n = 1$ and $m = 1$, then functions of the form functions of the form $\boldsymbol{F} \colon \mathbb{R}^n \to \mathbb{R}^m$ become simply $\boldsymbol{F} \colon \mathbb{R} \to \mathbb{R}$, which are single-variable real-valued functions. Hence, multivariable vector-valued functions include all the previous three types of functions we have seen.

Of course, just because multivariable vector-valued functions include all the previous three types of functions does not alone make multivariable vector-valued functions interesting. The reason we study such functions is because they in fact arise is many places in mathematics and its applications, including diverse fields such as physics and economics.

Finally, we need one slight modification of our definition of multivariable vector-valued functions. Consider the function $\boldsymbol{G} \colon \mathbb{R}^2 \to \mathbb{R}^3$ given by the formula $\boldsymbol{G}(x, y) = \begin{bmatrix} x+y \\ 3x^2 \\ 2x+e^y \end{bmatrix}$. Observe that this formula works for every possible point (x, y) in $\mathbb{R}^2$. On the other hand, suppose we wanted to define a function by the formula $\boldsymbol{H}(x, y) = \begin{bmatrix} 3xy \\ \sqrt{y-x} \\ 2x+5y \end{bmatrix}$. Observe that $\boldsymbol{H}(x, y)$ is not defined for all (x, y). Rather, it is defined only when $y - x \geq 0$, meaning when $y \geq x$. Let A be the region of the xy-plane that consists of all points (x, y) that satisfy $y \geq x$; that is, the region A consists of all points (x, y) that are on or above the line $y = x$. Hence, it would not be proper to write $\boldsymbol{H} \colon \mathbb{R}^2 \to \mathbb{R}^3$, because the $\mathbb{R}^2$ in this notation would imply that all of $\mathbb{R}^2$ can be used as input. Rather, we would write $\boldsymbol{H} \colon A \to \mathbb{R}^3$.

In general, multivariable vector-valued functions have the form $\boldsymbol{F} \colon D \to \mathbb{R}^m$, where D is a subset of $\mathbb{R}^n$. The set D is called the **domain** of the function $\boldsymbol{F}$. In a course at this level we assume the domain of a function $\boldsymbol{F} \colon D \to \mathbb{R}^m$ is the largest possible subset of $\mathbb{R}^n$ that allows the formula for $\boldsymbol{F}$ to be defined. In more advanced mathematics courses, the term "domain" is used slightly differently.

Visualizing Multivariable Vector-Valued Functions

As mentioned above, and as seen in Figure 1 and Figure 2 of this section, we can graph functions of the form $y = f(x)$ and $z = f(x, y)$. As seen in Figure 3 of this section, we can also draw the image of functions of the form $\boldsymbol{s}(t) = \begin{bmatrix} f(t) \\ g(t) \\ h(t) \end{bmatrix}$. The question of visualizing other types of multivariable vector-valued functions is trickier.

One type of multivariable vector-valued functions that we can visualize are functions $\boldsymbol{F}\colon \mathbb{R}^2 \to \mathbb{R}^3$, which have the form $\boldsymbol{F}(x, y) = \begin{bmatrix} P(x,y) \\ Q(x,y) \\ R(x,y) \end{bmatrix}$, where $P(x, y)$, $Q(x, y)$ and $R(x, y)$ are multivariable real-valued functions $P, Q, R\colon \mathbb{R}^2 \to \mathbb{R}$. Whereas the image of a function of the form $\boldsymbol{s}(t) = \begin{bmatrix} f(t) \\ g(t) \\ h(t) \end{bmatrix}$ is a curve in $\mathbb{R}^3$, the image of a function of the form $\boldsymbol{F}(x, y) = \begin{bmatrix} P(x,y) \\ Q(x,y) \\ R(x,y) \end{bmatrix}$ is, in general, a surface in $\mathbb{R}^3$, although this surface might have self-intersections. The reason the image of such a function is a surface (except in unusual cases) is that the two independent variables x and y give the image of the function two degrees of freedom, that is, two independent directions of motion. (An exception to that would be, for example, if $\boldsymbol{F}(x, y)$ is a constant function, in which case its image would be a single point.)

For example, we see in Figure 4 of this section the image of the function $\boldsymbol{F}\colon \mathbb{R}^2 \to \mathbb{R}^3$ given by the formula $\boldsymbol{F}(x, y) = \begin{bmatrix} x+3y \\ 2x-y \\ x^2-y^2 \end{bmatrix}$; the surface has been drawn slightly transparently, to help with visualization.

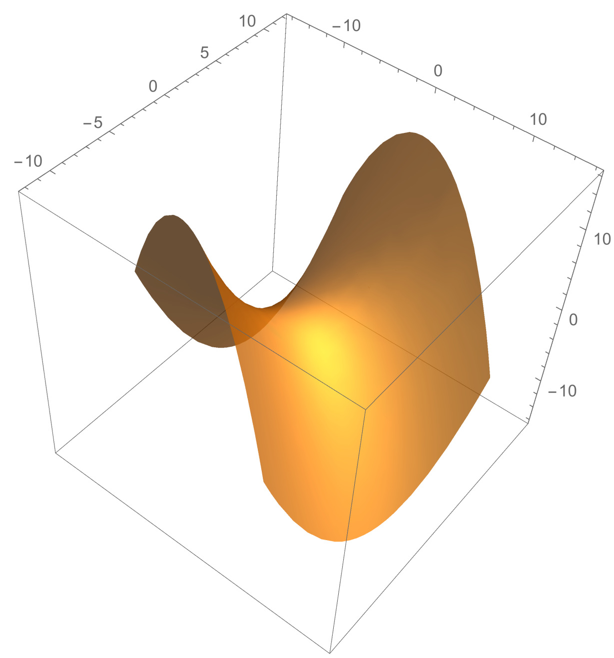


Figure 4: Graph of $\boldsymbol{F}$

One important point to keep in mind regarding the images of such functions is that, in contrast to the graphs of functions of the form $y = f(x)$ and $z = f(x, y)$, the images of functions $\boldsymbol{F}\colon \mathbb{R}^2 \to \mathbb{R}^3$ might have self-intersection. For example, see Figure 5 of this section for the image of the function $\boldsymbol{G}\colon \mathbb{R}^2 \to \mathbb{R}^3$ given by the formula $\boldsymbol{G}(x,y) = \begin{bmatrix} xy \\ x^2 \\ y \end{bmatrix}$.

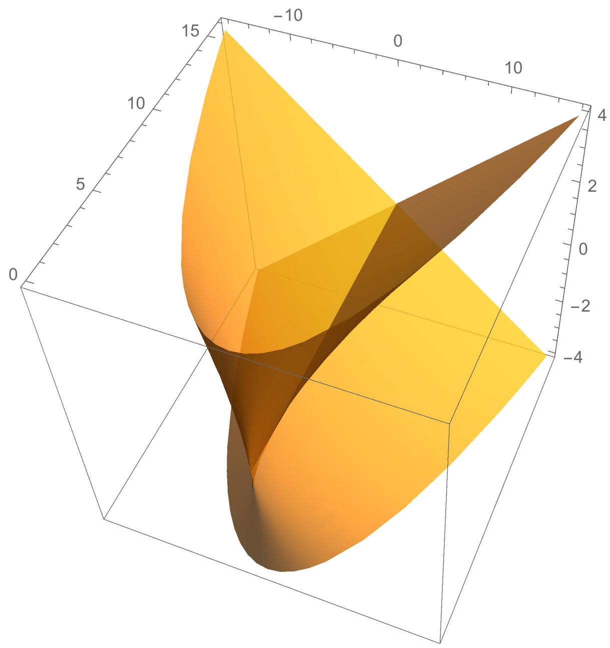


Figure 5: Graph of $\boldsymbol{G}$

One thing we can do to help us visualize multivariable vector-valued functions is to use what we know about the images of single-variable vector-valued functions. The idea is that for any function $\boldsymbol{F}\colon \mathbb{R}^n \to \mathbb{R}^m$, we can select any one of the coordinates in $\mathbb{R}^n$, and fix the values of all the other coordinates, and then think of the resulting function as a function $\mathbb{R} \to \mathbb{R}^m$.

For example, consider the function $\boldsymbol{F}\colon \mathbb{R}^2 \to \mathbb{R}^3$ given by the formula $\boldsymbol{F}(x,y) = \begin{bmatrix} x+3y \\ 2x-y \\ x^2-y^2 \end{bmatrix}$. Let c be a real number. We can then make a new function $\boldsymbol{F}_{xc}\colon \mathbb{R} \to \mathbb{R}^3$ by holding the variable y in $\boldsymbol{F}(x,y)$ to the fixed valued c, obtaining the formula $\boldsymbol{F}_{xc}(t) = \boldsymbol{F}(t,c) = \begin{bmatrix} t+3c \\ 2t-c \\ t^2-c^2 \end{bmatrix}$. For example, we have $\boldsymbol{F}_{x2}(t) = \begin{bmatrix} t+6 \\ 2t-2 \\ t^2-4 \end{bmatrix}$. We can then plot the image of this new single-variable vector-valued function. The image of this function will be a curve in $\mathbb{R}^3$, and this curve will sit inside the image of the original function $\boldsymbol{F}$. Similarly, we can make another new single-variable vector-valued function $\boldsymbol{F}_{dy}\colon \mathbb{R} \to \mathbb{R}^3$ by holding the variable x in $\boldsymbol{F}(x,y)$ to the fixed valued d, obtaining the formula $\boldsymbol{F}_{dy}(t) = \boldsymbol{F}(d,t) = \begin{bmatrix} d+3t \\ 2d-t \\ d^2-t^2 \end{bmatrix}$.

These single-variable vector-valued functions obtained by holding all but one of the variables fixed are called **parameter curves**.

See Figure 6 of this section for the image of the function $\boldsymbol{F}$ together with the images of six parameter curves, which are $\boldsymbol{F}_{x0}$, $\boldsymbol{F}_{x1}$ and $\boldsymbol{F}_{x2}$ in red, and $\boldsymbol{F}_{0y}$, $\boldsymbol{F}_{1y}$ and $\boldsymbol{F}_{2y}$ in blue.

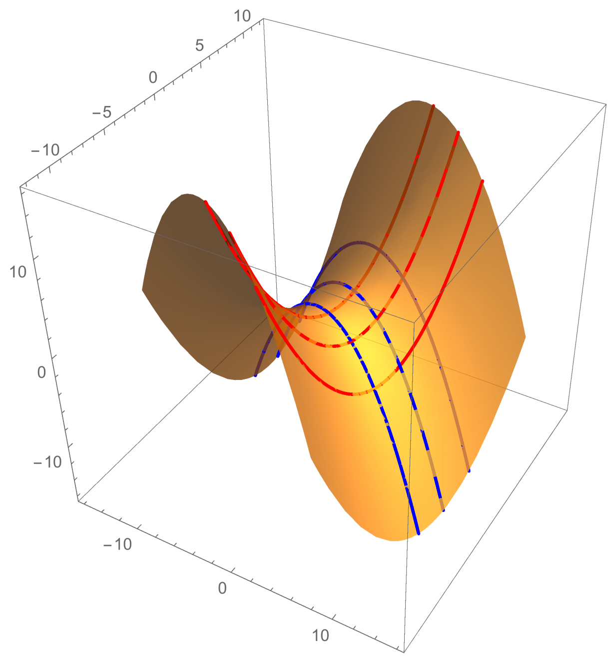


Figure 6: Graph of $\boldsymbol{F}$ with some parameter curves

It might seem uninteresting to draw the parameter curves for the above function $\boldsymbol{F}$, because we already saw the actual image of the this function, which was a surface. However, parameter curves become more useful for functions that we cannot easily visualize. For example, consider the function $\boldsymbol{H}\colon \mathbb{R}^3 \to \mathbb{R}^3$ given by the formula $\boldsymbol{H}(x, y, z) = \begin{bmatrix} x+3y+4z \\ 2x-y+z \\ x^2-y^2-z^2 \end{bmatrix}$. We cannot visualize the image of this function, because it would be some sort of folded-over version of $\mathbb{R}^3$ sitting inside of $\mathbb{R}^3$. However, to get some sort of visual idea of the function, we can still graph parameter curves for this function. For example, we can make a new function $\boldsymbol{H}_{xcd}\colon \mathbb{R} \to \mathbb{R}^3$ by holding the variables y and z in $\boldsymbol{H}(x, y, z)$ to the fixed valued c and d, obtaining the formula $\boldsymbol{H}_{xcd}(t) = \boldsymbol{H}(t, c, d) = \begin{bmatrix} t+3c+4d \\ 2t-c+d \\ t^2-c^2-d^2 \end{bmatrix}$. For example, we have $\boldsymbol{H}_{x12}(t) = \begin{bmatrix} t+11 \\ 2t-1 \\ t^2-5 \end{bmatrix}$. Even though we cannot visualize the original function directly, we can still plot the image of this parameter curve. See Figure 7 of this section for the images of a few of the parameter curves for $\boldsymbol{H}$, including $\boldsymbol{H}_{x12}$. In this figure, the curves in red are obtained by holding y and z fixed; the curves in blue are obtained by holding z and z fixed; and the curves in green are obtained by holding x and y fixed.

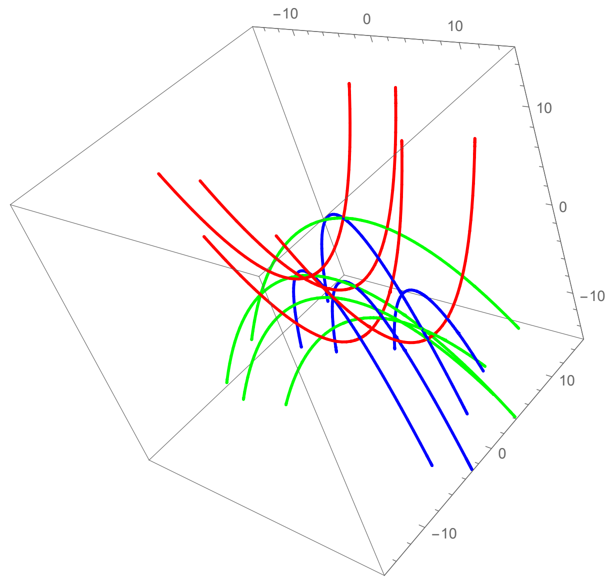


Figure 7: Graph of some parameter curves of $\boldsymbol{H}$

Limits of Multivariable Vector-Valued Functions

Limits are the essential idea that make calculus distinct from precalculus, and limits are present—either explicitly or implicitly—in everything we do in calculus. For example, the definition of derivatives and partial derivatives are given directly in terms of limits. However, because limits are somewhat tricky to work with, various convenient rules for calculating derivatives have been found, for example the Product Rule and the Quotient Rule. Of course, these rules are proved using limits, so whenever these rules are used, limits are implicitly used as well, though we do not have to deal with the limits explicitly in such situations.

In *Calculus I*, we saw limits of single-variable real-valued functions, that is, limits of the form $\lim_{x \to c} f(x)$. In *Calculus II*, we saw limits of multivariable real-valued functions, that is, limits of the form $\lim_{(x,y) \to (a,b)} f(x, y)$. In our current context, where we are interested in limits of multivariable vector-valued functions, there is really very little that is new, similarly to what we saw when we discussed limits of single-variable vector-valued functions.

For limits of multivariable vector-valued functions, let us start with the particular case of a function $\boldsymbol{G}\colon \mathbb{R}^2 \to \mathbb{R}^3$, which has the form $\boldsymbol{G}(x, y) = \begin{bmatrix} P(x,y) \\ Q(x,y) \\ R(x,y) \end{bmatrix}$, where where $P, Q, R\colon \mathbb{R}^2 \to \mathbb{R}$ are the component functions of $\boldsymbol{G}$. Let (a, b) be a point in $\mathbb{R}^2$. We want to find $\lim_{(x,y) \to (a,b)} \boldsymbol{G}(x, y)$. It turns out (though it takes some work to prove it rigorously, which we will not do), that to compute the limit of $\boldsymbol{G}$, we simply compute the limits of each of the component functions separately, and since those are multivariable real-valued functions, we compute those limits just as we did previously. That is, we have

$$\lim_{(x,y)\to(a,b)} \boldsymbol{G}(x,y) = \begin{bmatrix} \lim\limits_{(x,y)\to(a,b)} P(x,y) \\ \lim\limits_{(x,y)\to(a,b)} Q(x,y) \\ \lim\limits_{(x,y)\to(a,b)} R(x,y) \end{bmatrix},$$

provided that all three limits $\lim\limits_{(x,y)\to(a,b)} P(x,y)$, and $\lim\limits_{(x,y)\to(a,b)} Q(x,y)$ and $\lim\limits_{(x,y)\to(a,b)} R(x,y)$ exist. If even one of these three limits (of $P(x,y)$, and $Q(x,y)$ and $R(x,y)$) does not exist, then $\lim\limits_{(x,y)\to(a,b)} \boldsymbol{G}(x,y)$ does not exist.

The above limit can be written without using name of the function $\boldsymbol{G}$, by simply writing

$$\lim_{(x,y)\to(a,b)} \begin{bmatrix} P(x,y) \\ Q(x,y) \\ R(x,y) \end{bmatrix} = \begin{bmatrix} \lim\limits_{(x,y)\to(a,b)} P(x,y) \\ \lim\limits_{(x,y)\to(a,b)} Q(x,y) \\ \lim\limits_{(x,y)\to(a,b)} R(x,y) \end{bmatrix},$$

again, provided all three limits exist.

Example 1

Let $\boldsymbol{F}\colon \mathbb{R}^2 \to \mathbb{R}^3$ be given by the formula $\boldsymbol{F}(x,y) = \begin{bmatrix} 3x+2y \\ x^2y^3 \\ 2xe^y \end{bmatrix}$. Find $\lim\limits_{(x,y)\to(1,0)} \boldsymbol{F}(x,y)$, if the limit exists.

SOLUTION The component functions of $\boldsymbol{F}$ are the three multivariable real-valued functions $P, Q, R\colon \mathbb{R}^2 \to \mathbb{R}$ of $\boldsymbol{F}$ given by the formulas $P(x,y) = 3x+2y$, and $Q(x,y) = x^2y^3$ and $R(x,y) = 2xe^y$. We observing that each of these component functions is continuous, and hence we see that

$$\begin{aligned} \lim_{(x,y)\to(1,0)} P(x,y) &= \lim_{(x,y)\to(1,0)} (3x+2y) = 3\cdot 1 + 2\cdot 0 = 3 \\ \lim_{(x,y)\to(1,0)} Q(x,y) &= \lim_{(x,y)\to(1,0)} x^2y^3 = 1^2\cdot 0^3 = 0 \\ \lim_{(x,y)\to(1,0)} R(x,y) &= \lim_{(x,y)\to(1,0)} 2xe^y = 2\cdot 1\cdot e^0 = 2. \end{aligned}$$

We deduce that $\lim_{(x,y)\to(1,0)} \boldsymbol{F}(x,y)$ exists, and that

$$\lim_{(x,y)\to(1,0)} \boldsymbol{F}(x,y) = \begin{bmatrix} \lim_{(x,y)\to(1,0)} (3x+2y) \\ \lim_{(x,y)\to(1,0)} x^2y^3 \\ \lim_{(x,y)\to(1,0)} 2xe^y \end{bmatrix} = \begin{bmatrix} 3 \\ 0 \\ 2 \end{bmatrix}.$$

More generally, we have the following definition.

Multivariable Vector-Valued Functions: Limits

Let $\boldsymbol{F}\colon \mathbb{R}^n \to \mathbb{R}^m$ be a function, and let $(a_1, a_2, \ldots, a_n) \in \mathbb{R}^n$. Suppose that $\boldsymbol{F}$ is given by the formula

$$\boldsymbol{F}(x_1, x_2, \ldots, x_n) = \begin{bmatrix} f_1(x_1,x_2,\ldots,x_n) \\ f_2(x_1,x_2,\ldots,x_n) \\ \vdots \\ f_m(x_1,x_2,\ldots,x_n) \end{bmatrix},$$

where $f_1, f_2, \ldots, f_m\colon \mathbb{R}^n \to \mathbb{R}$ are the component functions of $\boldsymbol{F}$.

1. The **limit** of $\boldsymbol{F}(x_1, x_2, \ldots, x_n)$ as $(x_1, x_2, \ldots, x_n)$ goes to $(a_1, a_2, \ldots, a_n)$ is given by

$$\lim_{(x_1,x_2,\ldots,x_n)\to(a_1,a_2,\ldots,a_n)} \boldsymbol{F}(x_1, x_2, \ldots, x_n) = \begin{bmatrix} \lim_{(x_1,x_2,\ldots,x_n)\to(a_1,a_2,\ldots,a_n)} f_1(x_1,x_2,\ldots,x_n) \\ \lim_{(x_1,x_2,\ldots,x_n)\to(a_1,a_2,\ldots,a_n)} f_2(x_1,x_2,\ldots,x_n) \\ \vdots \\ \lim_{(x_1,x_2,\ldots,x_n)\to(a_1,a_2,\ldots,a_n)} f_m(x_1,x_2,\ldots,x_n) \end{bmatrix},$$

provided that all of the limits of the form $\lim_{(x_1,x_2,\ldots,x_n)\to(a_1,a_2,\ldots,a_n)} f_i(x_1, x_2, \ldots, x_n)$ exist. If even one of the limits $\lim_{(x_1,x_2,\ldots,x_n)\to(a_1,a_2,\ldots,a_n)} f_i(x_1, x_2, \ldots, x_n)$ does not exist, then $\lim_{(x_1,x_2,\ldots,x_n)\to(a_1,a_2,\ldots,a_n)} \boldsymbol{F}(x_1, x_2, \ldots, x_n)$ does not exist.

2. The above limit can be written without using name of the function

$\boldsymbol{F}$ by simply writing

$$\lim_{(x_1,x_2,\ldots,x_n)\to(a_1,a_2,\ldots,a_n)} \begin{bmatrix} f_1(x_1,x_2,\ldots,x_n) \\ f_2(x_1,x_2,\ldots,x_n) \\ \vdots \\ f_m(x_1,x_2,\ldots,x_n) \end{bmatrix} = \begin{bmatrix} \lim\limits_{(x_1,x_2,\ldots,x_n)\to(a_1,a_2,\ldots,a_n)} f_1(x_1,x_2,\ldots,x_n) \\ \lim\limits_{(x_1,x_2,\ldots,x_n)\to(a_1,a_2,\ldots,a_n)} f_2(x_1,x_2,\ldots,x_n) \\ \vdots \\ \lim\limits_{(x_1,x_2,\ldots,x_n)\to(a_1,a_2,\ldots,a_n)} f_m(x_1,x_2,\ldots,x_n) \end{bmatrix},$$

again, provided that all of the limits of the form $\lim\limits_{(x_1,x_2,\ldots,x_n)\to(a_1,a_2,\ldots,a_n)} f_i(x_1, x_2, \ldots, x_n)$ exist.

Example 2

Find $\lim\limits_{(x,y)\to(0,0)} \begin{bmatrix} \frac{\sin(x^2+y^2)}{x^2+y^2} \\ \frac{x^2-y^2}{x^2+y^2} \end{bmatrix}$, if the limit exists.

SOLUTION First, we consider the limit $\lim\limits_{(x,y)\to(0,0)} \frac{\sin(x^2+y^2)}{x^2+y^2}$. Using polar coordinates, we know $r^2 = x^2 + y^2$. Clearly, as (x, y) gets closer and closer to $(0, 0)$, then r gets closer and closer to 0. Hence, using l'Hôpital's Rule, we see that

$$\begin{aligned} \lim_{(x,y)\to(0,0)} \frac{\sin(x^2+y^2)}{x^2+y^2} &= \lim_{r\to 0} \frac{\sin(r^2)}{r^2} \\ &= \lim_{r\to 0} \frac{2r\cos(r^2)}{2r} \\ &= \lim_{r\to 0} \cos(r^2) = \cos(0^2) = 1. \end{aligned}$$

Next, we consider the limit $\lim\limits_{(x,y)\to(0,0)} \frac{x^2-y^2}{x^2+y^2}$. Let us see what happens as (x, y) approaches $(0, 0)$ from different directions. First, suppose that (x, y) approaches $(0, 0)$ along the x-axis, which means that we set $y = 0$ and then take the limit, which yields

$$\begin{aligned} \lim_{(x,0)\to(0,0)} \frac{x^2-y^2}{x^2+y^2} &= \lim_{(x,0)\to(0,0)} \frac{x^2-0^2}{x^2+0^2} \\ &= \lim_{(x,0)\to(0,0)} 1 = 1. \end{aligned}$$

Second, suppose that (x, y) approaches $(0, 0)$ along the y-axis, which means that we set $x = 0$ and then take the limit, which yields

$$\begin{aligned}\lim_{(0,y)\to(0,0)} \frac{x^2 - y^2}{x^2 + y^2} &= \lim_{(0,y)\to(0,0)} \frac{0^2 - y^2}{0^2 + y^2}\\ &= \lim_{(x,0)\to(0,0)} -1 = -1.\end{aligned}$$

Because we obtained different numbers as (x, y) approached $(0, 0)$ from different directions, we deduce that $\lim\limits_{(x,y)\to(0,0)} \frac{x^2-y^2}{x^2+y^2}$ does not exist.

We conclude that $\lim\limits_{(x,y)\to(0,0)} \begin{bmatrix} \frac{\sin(x^2+y^2)}{x^2+y^2} \\ \frac{x^2-y^2}{x^2+y^2} \end{bmatrix}$ does not exist, because the limit of one of the component functions does not exist.

Finally, we note that the notion of continuity works for multivariable vector-valued functions just as it does for single-variable vector-valued functions, again by looking at one component at a time. For example, consider a function $\boldsymbol{G}\colon \mathbb{R}^2 \to \mathbb{R}^3$, which has the form $\boldsymbol{G}(x, y) = \begin{bmatrix} P(x,y) \\ Q(x,y) \\ R(x,y) \end{bmatrix}$. Then the function $\boldsymbol{G}(x, y)$ is continuous if and only if all three of $P(x, y)$, $Q(x, y)$ and $R(x, y)$ are continuous, where the continuity of all three is as discussed for multivariable real-valued functions.

Example 3

Let $\boldsymbol{F}\colon \mathbb{R}^2 \to \mathbb{R}^3$ be given by the formula $\boldsymbol{F}(x, y) = \begin{bmatrix} 5x+y^2 \\ \sin(xy) \\ xe^y \end{bmatrix}$. Is $\boldsymbol{F}$ continuous?

SOLUTION To figure out if $\boldsymbol{F}$ is continuous, we need to figure out if each of the component functions $P(x, y) = 5x + y^2$, and $Q(x, y) = \sin(xy)$ and $R(x, y) = xe^y$ are continuous. We know that polynomials in each of x and y are continuous, so each of $5x$ and y^2 are continuous, and we we know that sums and products of continuous functions are continuous, and hence $P(x, y) = 5x + y^2$ and xy are both continuous. We know that $\sin x$ and e^y are continuous, and we know that compositions of continuous functions are continuous, and hence $Q(x, y) = \sin(xy)$ and $R(x, y) = xe^y$ are continuous. We conclude that $\boldsymbol{F}$ is continuous.

SUMMARY

Multivariable Vector-Valued Functions

1. A **multivariable vector-valued function** is a function $\boldsymbol{F}\colon \mathbb{R}^n \to \mathbb{R}^m$ for some positive integers n and m.

2. A function $\boldsymbol{F}\colon \mathbb{R}^n \to \mathbb{R}^m$ has the form

$$\boldsymbol{F}(x_1, x_2, \ldots, x_n) = \begin{bmatrix} f_1(x_1,x_2,\ldots,x_n) \\ f_2(x_1,x_2,\ldots,x_n) \\ \vdots \\ f_m(x_1,x_2,\ldots,x_n) \end{bmatrix},$$

   where $f_1, f_2, \ldots, f_m\colon \mathbb{R}^n \to \mathbb{R}$ are multivariable real-valued functions. The functions $f_1, f_2, \ldots, f_m$ are called the **component functions** of $\boldsymbol{F}$.

3. A function $\boldsymbol{F}\colon \mathbb{R}^2 \to \mathbb{R}^3$ has the form

$$\boldsymbol{F}(x, y) = \begin{bmatrix} P(x,y) \\ Q(x,y) \\ R(x,y) \end{bmatrix}.$$

Multivariable Vector-Valued Functions: Limits

Let $\boldsymbol{F}\colon \mathbb{R}^n \to \mathbb{R}^m$ be a function, and let $(a_1, a_2, \ldots, a_n) \in \mathbb{R}^n$. Suppose that $\boldsymbol{F}$ is given by the formula

$$\boldsymbol{F}(x_1, x_2, \ldots, x_n) = \begin{bmatrix} f_1(x_1,x_2,\ldots,x_n) \\ f_2(x_1,x_2,\ldots,x_n) \\ \vdots \\ f_m(x_1,x_2,\ldots,x_n) \end{bmatrix},$$

where $f_1, f_2, \ldots, f_m\colon \mathbb{R}^n \to \mathbb{R}$ are the component functions of $\boldsymbol{F}$.

1. The **limit** of $\boldsymbol{F}(x_1, x_2, \ldots, x_n)$ as $(x_1, x_2, \ldots, x_n)$ goes to

$(a_1, a_2, \ldots, a_n)$ is given by

$$\lim_{(x_1,x_2,\ldots,x_n)\to(a_1,a_2,\ldots,a_n)} \boldsymbol{F}(x_1, x_2, \ldots, x_n) = \begin{bmatrix} \lim_{(x_1,x_2,\ldots,x_n)\to(a_1,a_2,\ldots,a_n)} f_1(x_1,x_2,\ldots,x_n) \\ \lim_{(x_1,x_2,\ldots,x_n)\to(a_1,a_2,\ldots,a_n)} f_2(x_1,x_2,\ldots,x_n) \\ \vdots \\ \lim_{(x_1,x_2,\ldots,x_n)\to(a_1,a_2,\ldots,a_n)} f_m(x_1,x_2,\ldots,x_n) \end{bmatrix},$$

provided that all of the limits of the form $\lim_{(x_1,x_2,\ldots,x_n)\to(a_1,a_2,\ldots,a_n)} f_i(x_1, x_2, \ldots, x_n)$ exist. If even one of the limits $\lim_{(x_1,x_2,\ldots,x_n)\to(a_1,a_2,\ldots,a_n)} f_i(x_1, x_2, \ldots, x_n)$ does not exist, then $\lim_{(x_1,x_2,\ldots,x_n)\to(a_1,a_2,\ldots,a_n)} \boldsymbol{F}(x_1, x_2, \ldots, x_n)$ does not exist.

2. The above limit can be written without using name of the function $\boldsymbol{F}$ by simply writing

$$\lim_{(x_1,x_2,\ldots,x_n)\to(a_1,a_2,\ldots,a_n)} \begin{bmatrix} f_1(x_1,x_2,\ldots,x_n) \\ f_2(x_1,x_2,\ldots,x_n) \\ \vdots \\ f_m(x_1,x_2,\ldots,x_n) \end{bmatrix} = \begin{bmatrix} \lim_{(x_1,x_2,\ldots,x_n)\to(a_1,a_2,\ldots,a_n)} f_1(x_1,x_2,\ldots,x_n) \\ \lim_{(x_1,x_2,\ldots,x_n)\to(a_1,a_2,\ldots,a_n)} f_2(x_1,x_2,\ldots,x_n) \\ \vdots \\ \lim_{(x_1,x_2,\ldots,x_n)\to(a_1,a_2,\ldots,a_n)} f_m(x_1,x_2,\ldots,x_n) \end{bmatrix},$$

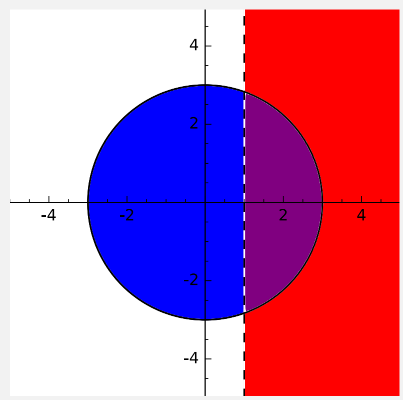
again, provided that all of the limits of the form $\lim_{(x_1,x_2,\ldots,x_n)\to(a_1,a_2,\ldots,a_n)} f_i(x_1, x_2, \ldots, x_n)$ exist.

EXAMPLES

Example 4

Let $\boldsymbol{H}\colon \mathbb{R}^2 \to \mathbb{R}^2$ be given by the formula $\boldsymbol{H}(x, y) = \begin{bmatrix} \sqrt{9-x^2-y^2} \\ \ln(x-1) \end{bmatrix}$. Find and sketch the domain of $\boldsymbol{H}$.

SOLUTION The domain of $\boldsymbol{H}$ is the set of all points (x, y) in $\mathbb{R}^2$ for which both $\sqrt{9 - x^2 - y^2}$ and $\ln(x - 1)$ are defined. That is, the domain is the set of all points (x, y) in $\mathbb{R}^2$ for which both inequalities $9 - x^2 - y^2 \geq 0$ and $x - 1 > 0$ hold, which is the same as $x^2 + y^2 \leq 3^2$ and $x > 1$. The solution of $x^2 + y^2 \leq 3^2$ is the interior of the circle of radius 3 centered at the origin together with the circle, shown below in blue; the solution of $x > 1$ is the half-plane that is to the right of the vertical line $x = 1$, not including the line, shown below in red. The domain of $\boldsymbol{H}$ is the intersection of these two regions, which is shown below in purple, where the part of the boundary of the region that is solid is included in the region, and the part of the boundary that is dashed is not included in the region.



EXERCISES

Basic Exercises

1–3 ■ Find and sketch the domain of each function.

1. Let $\boldsymbol{F}\colon \mathbb{R}^2 \to \mathbb{R}^2$ be given by the formula
$$\boldsymbol{F}(x, y) = \begin{bmatrix} \sqrt{x-y^2} \\ \sqrt{9-x} \end{bmatrix}.$$

2. Let $\boldsymbol{G}\colon \mathbb{R}^2 \to \mathbb{R}^3$ be given by the formula
$$\boldsymbol{G}(x, y) = \begin{bmatrix} \ln(x+y) \\ \sqrt{x^2 y} \\ \sin(xy) \end{bmatrix}.$$

31.2 Derivatives

In *Calculus I*, we saw how to take the derivative of single-variable real-valued functions. For example, if $f\colon \mathbb{R} \to \mathbb{R}$ is given by the formula $f(x) = x^2$, then the derivative of this function is given by the formula $f'(x) = 2x$. If we do not need the formula for the function, but only the name of the function, then the original function is f, and the derivative is the function f'. We also used the Leibniz notation for the derivative of this function, which is written $\dfrac{df}{dx}$ or $\dfrac{dy}{dx}$. Whatever notation for the derivative is used, the main point is that we start with a function $f\colon \mathbb{R} \to \mathbb{R}$ and then arrive at a new function, the derivative of f, which is useful in the study of the original function.

Of course, the way we compute f' is via the limit definition, which is given by the formula

$$f'(x) = \lim_{h\to 0} \frac{f(x+h) - f(x)}{h},$$

for those values of x for which the limit exists. In practice, once we prove various computational rules such as the Product Rule, the Quotient Rule and the Chain Rule (proved using the limit definition of derivatives), we don't subsequently need to use the rather cumbersome limit definition to find derivatives in practice.

In *Calculus II*, we saw how to take the partial derivatives of multivariable real-valued functions. For example, if $f\colon \mathbb{R}^2 \to \mathbb{R}$ is given by the formula $f(x, y) = 3x^2y$, then the partial derivatives of this function are given by the formulas $f_x(x, y) = 6xy$ and $f_y(x, y) = 3x^2$. We also used the Leibniz notation for the partial derivative of this function, which are written $\frac{\partial f}{\partial x}$ or $\frac{\partial z}{\partial x}$, and $\frac{\partial f}{\partial y}$ or $\frac{\partial z}{\partial y}$.

The important point to note here is that in contrast to what we saw for single-variable real-valued functions, for multivariable real-valued functions we start with one function, for example $f\colon \mathbb{R}^2 \to \mathbb{R}$, and we obtained not one single thing that is the derivative of f, but rather two partial derivatives, each of which is also a function $\mathbb{R}^2 \to \mathbb{R}$.

Recall that the partial derivatives of a multivariable real-valued function are also defined by a limit definition, though in practice we compute them by holding all but one of the variables as if they were constants, and then using the rules for computing derivatives of single-variable real-valued functions.

We previously saw how to take the derivative of single-variable vector-valued functions. For example, if $\boldsymbol{r}\colon \mathbb{R} \to \mathbb{R}^2$ is given by the formula $\boldsymbol{r}(t) = \begin{bmatrix} t^2 \\ 3t+1 \end{bmatrix}$, then the derivative of this function is given by the formula $\boldsymbol{r}'(t) = \begin{bmatrix} 2t \\ 3 \end{bmatrix}$. We can view the derivative in this case in two ways, either as a single vector-valued function, or as a vector containing a number of derivatives of single-variable real-valued functions, namely, the derivatives of the component functions of $\boldsymbol{r}$.

We now turn to the derivative of the new type of function we are considering in this chapter, namely, multivariable vector-valued functions.

The Derivative of Multivariable Vector-Valued Functions

Suppose, for example, that $\boldsymbol{G}\colon \mathbb{R}^2 \to \mathbb{R}^3$ is given by the formula $\boldsymbol{G}(x, y) = \begin{bmatrix} 5x+y \\ 3x^2 \\ 2x+e^y \end{bmatrix}$. Rather than attempting to take the derivative via limits, we now use partial derivatives, taking our inspiration from the way we found derivatives for both multivariable real-valued functions and single-variable vector-valued functions. The function $\boldsymbol{G}$ is defined by three component functions, which we can write as $\boldsymbol{G}(x, y) = \begin{bmatrix} P(x,y) \\ Q(x,y) \\ R(x,y) \end{bmatrix}$, where $P(x, y) = 5x + y$ and $Q(x, y) = 3x^2$ and $R(x, y) = 2x + e^y$. Each of these three functions has two partial derivatives, so that we obtain a total of six partial derivatives, which are

$$\begin{aligned} \frac{\partial P}{\partial x} &= 5 \quad \text{and} \quad \frac{\partial P}{\partial y} = 1 \\ \frac{\partial Q}{\partial x} &= 6x \quad \text{and} \quad \frac{\partial Q}{\partial y} = 0 \\ \frac{\partial R}{\partial x} &= 2 \quad \text{and} \quad \frac{\partial R}{\partial y} = e^y. \end{aligned} \tag{1}$$

It would be possible simply to deal with each of these six partial derivatives separately, but it turns out that we can arrange them in a particularly convenient and useful way, by making use of matrices.

Observe that in the notation $\begin{bmatrix} P(x,y) \\ Q(x,y) \\ R(x,y) \end{bmatrix}$, the "input" variables x and y in each of P, Q and R are written horizontally, in a row, whereas the "output" variables P, Q and R themselves are written vertically, in a column. We will arrange our six partial derivatives by maintaining the original horizontal and vertical arrangement. That is, for the function $P(x, y)$, which is in the top column of the output, we will find its two partial derivatives, which are $\frac{\partial P}{\partial x}$ and $\frac{\partial P}{\partial y}$, and write them horizontally, in that order. That will be the first row in our matrix. Next, for the function $Q(x, y)$, which is in the second column of the output, we will find its two partial derivatives, which are $\frac{\partial Q}{\partial x}$ and $\frac{\partial Q}{\partial y}$, and write them horizontally, in that order, forming the second row of our matrix. Finally, for the function $R(x, y)$, which is in the third column of the output, we will find its two partial derivatives, which are $\frac{\partial R}{\partial x}$ and $\frac{\partial R}{\partial y}$, and write them horizontally, in that order, forming the third row of our matrix. The matrix of partial derivatives is then

$$\begin{bmatrix} \frac{\partial P}{\partial x} & \frac{\partial P}{\partial y} \\ \frac{\partial Q}{\partial x} & \frac{\partial Q}{\partial y} \\ \frac{\partial R}{\partial x} & \frac{\partial R}{\partial y} \end{bmatrix}.$$

This matrix is called the **derivative** of $\boldsymbol{G}$, and is denoted $D\boldsymbol{G}$; if we need to specify the input variables, we would write $D\boldsymbol{G}(x, y)$, which could also be written as $D\boldsymbol{G}(p)$, where p denotes a point in $\mathbb{R}^2$.

Example 1

Let $\boldsymbol{G} \colon \mathbb{R}^2 \to \mathbb{R}^3$ be given by the formula $\boldsymbol{G}(x, y) = \begin{bmatrix} 5x+y \\ 3x^2 \\ 2x+e^y \end{bmatrix}$. Find the derivative of $\boldsymbol{G}$.

SOLUTION Using the partial derivatives we computed in Equation (1) of this section, we see that

$$D\boldsymbol{G}(x, y) = \begin{bmatrix} \frac{\partial P}{\partial x} & \frac{\partial P}{\partial y} \\ \frac{\partial Q}{\partial x} & \frac{\partial Q}{\partial y} \\ \frac{\partial R}{\partial x} & \frac{\partial R}{\partial y} \end{bmatrix} = \begin{bmatrix} 5 & 1 \\ 6x & 0 \\ 2 & e^y \end{bmatrix}.$$

Now let us look at the general case.

Multivariable Vector-Valued Functions: Derivatives

1. Let $\boldsymbol{F} \colon \mathbb{R}^n \to \mathbb{R}^m$ be a multivariable vector-valued function. Suppose that $\boldsymbol{F}(x_1, x_2, \ldots, x_n)$ is given by the formula

$$\boldsymbol{F}(x_1, x_2, \ldots, x_n) = \begin{bmatrix} f_1(x_1,x_2,\ldots,x_n) \\ f_2(x_1,x_2,\ldots,x_n) \\ \vdots \\ f_m(x_1,x_2,\ldots,x_n) \end{bmatrix}.$$

The **derivative** (also called the **Jacobian matrix**) of $\boldsymbol{F}$ is the $m \times n$ matrix

$$DF(x_1, x_2, \ldots, x_n) = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \frac{\partial f_m}{\partial x_2} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix}. \tag{2}$$

2. If the point $(x_1, x_2, \ldots, x_n)$ is abbreviated by p, then the derivative of $\boldsymbol{F}$ is also denoted $D\boldsymbol{F}(p)$; it is also written $D\boldsymbol{F}_p$ or $\boldsymbol{F}'(p)$. When only the name of the derivative is needed, without listing the variables, it is written $D\boldsymbol{F}$.

3. If $\boldsymbol{F}(x, y) = \begin{bmatrix} P(x,y) \\ Q(x,y) \\ R(x,y) \end{bmatrix}$, then

$$D\boldsymbol{F}(x, y) = \begin{bmatrix} \frac{\partial P}{\partial x} & \frac{\partial P}{\partial y} \\ \frac{\partial Q}{\partial x} & \frac{\partial Q}{\partial y} \\ \frac{\partial R}{\partial x} & \frac{\partial R}{\partial y} \end{bmatrix}.$$

Of course, not every multivariable vector-valued function has a derivative, because the components of such a function are multivariable real-valued functions, and not every multivariable real-valued function has partial derivatives. For example, the function $\boldsymbol{F}\colon \mathbb{R}^2 \to \mathbb{R}^2$ given by the formula $\boldsymbol{F}(x, y) = \begin{bmatrix} 5x+|y| \\ 3x^2y \end{bmatrix}$ does not have all possible partial derivatives when $y = 0$.

It might seem reasonable to call a multivariable vector-valued function "differentiable" if all its components have all their partial derivatives, but we cannot use that term this way, because it has another meaning, as mentioned briefly in an Optional Section of this chapter, where the derivative of multivariable vector-valued function via limits is discussed.

Observe that if $\boldsymbol{F}\colon \mathbb{R}^n \to \mathbb{R}^m$ is a multivariable vector-valued function, then $D\boldsymbol{F}$ is an $m \times n$ matrix, where each entry in this matrix is a multivariable real-valued function. For example, for the function $\boldsymbol{G}\colon \mathbb{R}^2 \to \mathbb{R}^3$ discussed in Example 1 of this section, we saw that the derivative is a 3×2 matrix.

The obvious question now arises as to why we have arranged the partial derivatives into a matrix in this way, as opposed to arranging the partial derivatives in some other way, or simply treating the partial derivatives separately. The reason to arrange the partial derivatives in a matrix in this particular way is that doing so will allow us to use matrix multiplication, and other aspects of matrices, to provide some formulas that are very nice analogs of what we saw in *Calculus I* for single-variable real-valued functions. Most immediately, we will see examples of the use of matrix multiplication in this context in Section 31.3, where we discuss linear approximations, and in Section 31.4, where we discuss the Chain Rule.

While the notion of the derivative of a multivariable vector-valued function in general is new, we observe that for single-variable real-valued, multivariable real-valued and single-variable vector-valued functions, our matrix approach is really just another way of writing the familiar derivatives of those previous types of functions.

First, suppose $f\colon \mathbb{R} \to \mathbb{R}$ is a single-variable real-valued function. The derivative of f at a number x is just a single number $f'(x)$. On the other hand, we can think of $\mathbb{R}$ as $\mathbb{R}^1$, and so we can think of $f\colon \mathbb{R}^1 \to \mathbb{R}^1$ as a multivariable vector-valued function. Viewed that way, the function f has a derivative which is a 1×1 matrix, namely, the matrix $Df(x) = \left[\, \frac{\partial f}{\partial x} \,\right] = [\, f(x) \,]$. Of course, a 1×1 matrix is a silly thing, and we will just take the normal derivatives of single-variable real-valued functions, but it is important to note that other than the extra square brackets, there is no difference between studying the function $f\colon \mathbb{R} \to \mathbb{R}$ as a single-variable real-valued function and the function $f\colon \mathbb{R}^1 \to \mathbb{R}^1$ as a multivariable vector-valued function.

Next, suppose $f\colon \mathbb{R}^n \to \mathbb{R}$ is a multivariable real-valued function. When we first encountered such functions, we simply looked at each of the n partial derivatives separately. On the other hand, we can think of $f\colon \mathbb{R}^n \to \mathbb{R}^1$ as a multivariable vector-valued function. Viewed that way, the function f has a derivative which is a $1 \times n$ matrix, which is a row matrix, namely, the matrix

$$Df(x_1, x_2, \ldots, x_n) = \left[\, \frac{\partial f}{\partial x_1} \; \frac{\partial f}{\partial x_2} \; \cdots \; \frac{\partial f}{\partial x_n} \,\right].$$

Hence, we see that viewing a $f\colon \mathbb{R}^n \to \mathbb{R}$ as a multivariable vector-valued function simply assembles the partial derivatives as a row matrix. In subsequent sections, we will define and make use of the gradient of such a function, which is just the transpose of $Df(x_1, x_2, \ldots, x_n)$, and hence is a column vector.

Finally, suppose $\boldsymbol{r}\colon \mathbb{R} \to \mathbb{R}^n$ is a single-variable vector-valued function. Suppose that this function is written as a column vector via component functions as $\boldsymbol{r}(t) = \begin{bmatrix} f_1(t) \\ \vdots \\ f_n(t) \end{bmatrix}$. Then the derivative of this function is the column vector $\boldsymbol{r}'(t) = \begin{bmatrix} f_1'(t) \\ \vdots \\ f_n'(t) \end{bmatrix}$. Once again, we can think of we can think of $\boldsymbol{r}\colon \mathbb{R}^1 \to \mathbb{R}^n$ as a multivariable vector-valued function. Viewed that way, the function f has a derivative which is an $n \times 1$ matrix, namely, the matrix $D\boldsymbol{r}(t) = \begin{bmatrix} \frac{\partial f_1}{\partial t} \\ \vdots \\ \frac{\partial f_n}{\partial t} \end{bmatrix} = \begin{bmatrix} f_1'(t) \\ \vdots \\ f_n'(t) \end{bmatrix}$. Of course, an $n \times 1$ matrix is the same as a column vector, and so we see that viewing a $\boldsymbol{r}\colon \mathbb{R} \to \mathbb{R}^n$ as a multivariable vector-valued function just gives us the derivative we already knew, except for thinking of it as a matrix rather than a column vector.

All told, if $\boldsymbol{F}\colon \mathbb{R}^n \to \mathbb{R}^m$ and $n \neq 1$ and $m \neq 1$, then the derivative $D\boldsymbol{F}$ that we have currently defined is something genuinely new, and if either $n = 1$ or $m = 1$, or both, then $D\boldsymbol{F}$ is just a slightly different way of writing the derivatives we are already familiar with. Hence, what we are considering at present incorporates as special cases everything we have seen up till now regarding derivatives.

Finally, we note that for a function of the form $\boldsymbol{F}\colon \mathbb{R}^n \to \mathbb{R}^m$, the columns of the derivative matrix of $\boldsymbol{F}$ are tangent vectors to parameter curves of the function, as defined in Section 31.1.

Basic Properties of the Derivative

In *Calculus I*, we saw a number of convenient rules for finding the derivatives of single-variable real-valued functions, including the Product Rule, the Quotient Rule and the Chain Rule. Some, but not all, of these rules have analogs for multivariable vector-valued functions. For example, because we cannot divide vectors, there is no analog of the Quotient Rule for multivariable vector-valued functions. We can take the dot product of multivariable vector-valued functions, just as we did for single-variable vector-valued functions, and there is a version of the Product Rule for multivariable vector-valued functions, though it is a bit more complicated than the version we saw for single-variable vector-valued functions, and we will not state it. There is a very nice version of the Chain Rule in our present context, which is discussed in Section 31.4. For now, we state the analogs of the most basic properties of derivatives of single-variable real-valued functions.

Basic Rules for Derivatives

Let $\boldsymbol{F}, \boldsymbol{G}\colon \mathbb{R}^n \to \mathbb{R}^m$ be functions, let c be a real number, and let p be a point in $\mathbb{R}^n$. Suppose that $\boldsymbol{F}$ and $\boldsymbol{G}$ are differentiable. Then

1. $D(\boldsymbol{F}+\boldsymbol{G})(p) = D\boldsymbol{F}(p) + D\boldsymbol{G}(p)$;
2. $D(\boldsymbol{F}-\boldsymbol{G})(p) = D\boldsymbol{F}(p) - D\boldsymbol{G}(p)$;
3. $D(c\boldsymbol{F})(p) = cD\boldsymbol{F}(p)$.

The Jacobian Determinant

Of the many things that can be done with matrices, recall determinants of square matrices, which are very useful in a variety of situations. In our present content, observe that while in general the derivative of a multivariable vector-valued function $\boldsymbol{F}\colon \mathbb{R}^n \to \mathbb{R}^m$ is not a square matrix when $n \neq m$, it will be a square matrix when $n = m$, and in that case we can take the determinant of the derivative.

Suppose we have a multivariable vector-valued function $\boldsymbol{F}\colon \mathbb{R}^n \to \mathbb{R}^n$. The **Jacobian determinant** of $\boldsymbol{F}$, is defined to be $\det D\boldsymbol{F}$. If p is a point in $\mathbb{R}^n$, the Jacobian determinant of $\boldsymbol{F}$ at p is denoted $\det D\boldsymbol{F}(p)$.

Example 2

Let $\boldsymbol{F}\colon \mathbb{R}^2 \to \mathbb{R}^2$ be given by the formula $\boldsymbol{F}(x,y) = \begin{bmatrix} 3x^2y \\ 5x+y^3 \end{bmatrix}$. Find the Jacobian determinant of $\boldsymbol{F}$.

SOLUTION We compute $D\boldsymbol{F}(x,y) = \begin{bmatrix} 6xy & 3x^2 \\ 5 & 3y^2 \end{bmatrix}$, and hence $\det D\boldsymbol{F}(x,y) = 6xy \cdot 3y^2 - 3x^2 \cdot 5 = 18xy^3 - 15x^2$.

We mention that there is some slight confusion regarding the term "Jacobian" in this context. Suppose $\boldsymbol{F}\colon \mathbb{R}^n \to \mathbb{R}^n$ is a multivariable vector-valued function. In some books, the derivative $D\boldsymbol{F}$ is called the "Jacobian matrix" of $\boldsymbol{F}$, which sounds very similar to the "Jacobian determinant" of $\boldsymbol{F}$. Even worse, some books use the word "Jacobian" without the associated word "matrix" or "determinant," and it is then unclear which of the two uses of the word "Jacobian" is meant. To avoid confusion, we will not use the word "Jacobian" by itself, and will say "Jacobian determinant" when we mean that, and will say "derivative" when we mean that.

There is also some confusion regarding the notation for the Jacobian determinant. First, rather than writing $\det\begin{bmatrix} a & b \\ c & d \end{bmatrix}$, some people use the old fashioned notation $\begin{vmatrix} a & b \\ c & d \end{vmatrix}$. Specifically, suppose we have $\boldsymbol{F}(x, y) = \begin{bmatrix} P(x,y) \\ Q(x,y) \end{bmatrix}$. Then $D\boldsymbol{F}(x, y) = \begin{bmatrix} \frac{\partial P}{\partial x} & \frac{\partial P}{\partial y} \\ \frac{\partial Q}{\partial x} & \frac{\partial Q}{\partial y} \end{bmatrix}$, and whereas we would then write the Jacobian determinant of $\boldsymbol{F}$ as $\det D\boldsymbol{F}(x, y) = \det \begin{bmatrix} \frac{\partial P}{\partial x} & \frac{\partial P}{\partial y} \\ \frac{\partial Q}{\partial x} & \frac{\partial Q}{\partial y} \end{bmatrix}$, the people who use the alternative notation for determinants would write the Jacobian determinant of $\boldsymbol{F}$ as $\begin{vmatrix} \frac{\partial P}{\partial x} & \frac{\partial P}{\partial y} \\ \frac{\partial Q}{\partial x} & \frac{\partial Q}{\partial y} \end{vmatrix}$. Because writing $\begin{vmatrix} \frac{\partial P}{\partial x} & \frac{\partial P}{\partial y} \\ \frac{\partial Q}{\partial x} & \frac{\partial Q}{\partial y} \end{vmatrix}$ by hand can look very similar to $\begin{bmatrix} \frac{\partial P}{\partial x} & \frac{\partial P}{\partial y} \\ \frac{\partial Q}{\partial x} & \frac{\partial Q}{\partial y} \end{bmatrix}$, we recommend using the notation $\det \begin{bmatrix} \frac{\partial P}{\partial x} & \frac{\partial P}{\partial y} \\ \frac{\partial Q}{\partial x} & \frac{\partial Q}{\partial y} \end{bmatrix}$, which is unambiguous.

Additionally, rather than writing $\boldsymbol{F}$ as $\begin{bmatrix} P(x,y) \\ Q(x,y) \end{bmatrix}$, where x and y are the independent variables, it is sometimes convenient to write x and y as the dependent variables, and u and v as the independent variables, and one then writes $x = g(u, v)$ and $y = h(u, v)$, which we would write as $\boldsymbol{T}(u, v) = \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} g(u,v) \\ h(u,v) \end{bmatrix}$. The Jacobian determinant of $\boldsymbol{T}$ is then $\det \begin{bmatrix} \frac{\partial g}{\partial u} & \frac{\partial g}{\partial v} \\ \frac{\partial h}{\partial u} & \frac{\partial h}{\partial v} \end{bmatrix}$. Sometimes, however, functions are not given names, so that rather than writing $x = g(u, v)$, we simply think of x as a function of u and v, which we could write $x = x(u, v)$, and similarly for y. The Jacobian determinant of x and y as functions of u and v would then be written, without function names and with the old fashioned notation for the determinant, as $\begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix}$. Moreover, instead of writing the Jacobian determinant in this situation in either of the above notations, some people use the old-fashioned notation $\frac{\partial(x,y)}{\partial(u,v)}$, which does not even need the name of the function. Similar notation is used with more variables.

We summarize these ideas as follows.

The Jacobian Determinant

1. Let $\boldsymbol{F}\colon \mathbb{R}^n \to \mathbb{R}^n$ be a function. The determinant of the derivative of $\boldsymbol{F}$ is called the **Jacobian determinant** (or just the **Jacobian**) of the function, and is denoted

$$\det D\boldsymbol{F}(x_1, x_2, \ldots, x_n),$$

or similarly if a different notation for the derivative is used.

2. If the vector $\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$ is abbreviated by p, then the Jacobian determinant of $\boldsymbol{F}$ is also denoted $\det D\boldsymbol{F}(p)$. When only the name of the

Jacobian determinant is needed, without listing the variables, it is written $\det DF$.

3. If $F(u,v) = \begin{bmatrix} P(u,v) \\ Q(u,v) \end{bmatrix}$, then the Jacobian determinant of F is sometimes denoted $\begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix}$ or $\frac{\partial(x,y)}{\partial(u,v)}$.

Geometric Meaning of the Jacobian Determinant

Let $H\colon \mathbb{R}^2 \to \mathbb{R}^2$ be the linear map given by the formula $H(x,y) = \begin{bmatrix} ax+by \\ cx+dy \end{bmatrix}$. The two basis vectors $e_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $e_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ in $\mathbb{R}^2$ form a square, as seen in Figure 1 of this section. We can then look at the parallelogram formed by the two vectors $H(e_1)$ and $H(e_2)$, which is seen in Figure 2 for the specific example where $H(x,y) = \begin{bmatrix} 2x+y \\ x+y \end{bmatrix}$; this new parallelogram is the image result of applying the function H to the original square. In general, we ask what is the relation between the area of the original square, which is 1, and the area of the new parallelogram.

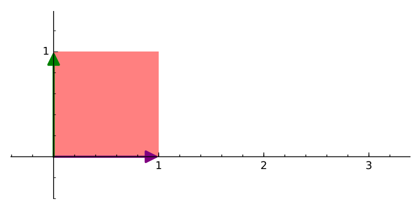


Figure 1: Square formed by the vectors e_1 and e_2

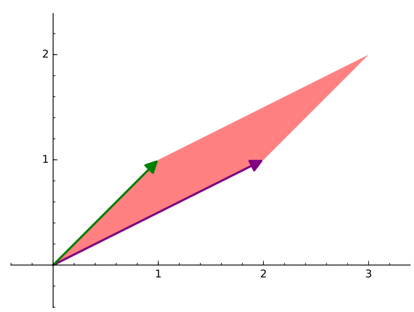


Figure 2: Parallelogram formed by the vectors $F(e_1)$ and $F(e_2)$

Observe that $H(e_1) = \begin{bmatrix} a \\ c \end{bmatrix}$ and $H(e_2) = \begin{bmatrix} b \\ d \end{bmatrix}$. As we saw in our discussion of the cross product, the area of the area of the parallelogram formed by the two vectors $H(e_1)$ and $H(e_2)$ is $|\det\begin{bmatrix} a & b \\ c & d \end{bmatrix}|$. Hence, we see that applying the linear map H to the original square results in multiplying the area by $|\det\left[\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right]|$. We note that the matrix $\left[\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right]$ should look familiar in this context, because it is precisely $D\boldsymbol{H}$.

More generally, let $\boldsymbol{F}\colon \mathbb{R}^n \to \mathbb{R}^n$ be a function, and let p be a point in $\mathbb{R}^n$. Of course, the function $\boldsymbol{F}$ need not be a linear map. However, as we will see in more detail in Section 31.3, the matrix $D\boldsymbol{F}(p)$ is the matrix of the linear map that best approximates $\boldsymbol{F}$ at the point p. Then, analogously to what we saw above for the linear map $\boldsymbol{H}$, it turns out more generally that if we were to take a small region A in $\mathbb{R}^n$ that is near p, then the area of $\boldsymbol{F}(A)$ is approximately equal to $D\boldsymbol{F}(p)$ times the area of A. The smaller the region, and the closer to p, the better the approximation. The proof of this fact is not simple, and it requires certain hypotheses on the function $\boldsymbol{F}$, but the inutive idea is simply that we use the matrix $D\boldsymbol{F}(p)$, and that matrix multiplication is a linear map.

This geometric way of thinking about $D\boldsymbol{F}(p)$ will be used when we discuss the change of variable formula for double and triple integrals.

SUMMARY

Multivariable Vector-Valued Functions: Derivatives

1. Let $\boldsymbol{F}\colon \mathbb{R}^n \to \mathbb{R}^m$ be a multivariable vector-valued function. Suppose that $\boldsymbol{F}(x_1, x_2, \ldots, x_n)$ is given by the formula

$$\boldsymbol{F}(x_1, x_2, \ldots, x_n) = \begin{bmatrix} f_1(x_1,x_2,\ldots,x_n) \\ f_2(x_1,x_2,\ldots,x_n) \\ \vdots \\ f_m(x_1,x_2,\ldots,x_n) \end{bmatrix}.$$

The **derivative** (also called the **Jacobian matrix**) of $\boldsymbol{F}$ is the $m \times n$ matrix

$$D\boldsymbol{F}(x_1, x_2, \ldots, x_n) = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \frac{\partial f_m}{\partial x_2} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix}.$$

2. If the point $(x_1, x_2, \dots, x_n)$ is abbreviated by p, then the derivative of $\boldsymbol{F}$ is also denoted $D\boldsymbol{F}(p)$; it is also written $D\boldsymbol{F}_p$ or $\boldsymbol{F}'(p)$. When only the name of the derivative is needed, without listing the variables, it is written $D\boldsymbol{F}$.

3. If $\boldsymbol{F}(x,y) = \begin{bmatrix} P(x,y) \\ Q(x,y) \\ R(x,y) \end{bmatrix}$, then

$$D\boldsymbol{F}(x,y) = \begin{bmatrix} \frac{\partial P}{\partial x} & \frac{\partial P}{\partial y} \\ \frac{\partial Q}{\partial x} & \frac{\partial Q}{\partial y} \\ \frac{\partial R}{\partial x} & \frac{\partial R}{\partial y} \end{bmatrix}.$$

Basic Rules for Derivatives

Let $\boldsymbol{F}, \boldsymbol{G} \colon \mathbb{R}^n \to \mathbb{R}^m$ be functions, let c be a real number, and let p be a point in $\mathbb{R}^n$. Suppose that $\boldsymbol{F}$ and $\boldsymbol{G}$ are differentiable. Then

1. $D(\boldsymbol{F} + \boldsymbol{G})(p) = D\boldsymbol{F}(p) + D\boldsymbol{G}(p)$;
2. $D(\boldsymbol{F} - \boldsymbol{G})(p) = D\boldsymbol{F}(p) - D\boldsymbol{G}(p)$;
3. $D(c\boldsymbol{F})(p) = cD\boldsymbol{F}(p)$.

The Jacobian Determinant

1. Let $\boldsymbol{F} \colon \mathbb{R}^n \to \mathbb{R}^n$ be a function. The determinant of the derivative of $\boldsymbol{F}$ is called the **Jacobian determinant** (or just the **Jacobian**) of the function, and is denoted

$$\det D\boldsymbol{F}(x_1, x_2, \dots, x_n),$$

or similarly if a different notation for the derivative is used.

2. If the vector $\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$ is abbreviated by p, then the Jacobian determinant of $\boldsymbol{F}$ is also denoted $\det D\boldsymbol{F}(p)$. When only the name of the Jacobian determinant is needed, without listing the variables, it is written $\det D\boldsymbol{F}$.

3. If $\boldsymbol{F}(u,v) = \begin{bmatrix} P(u,v) \\ Q(u,v) \end{bmatrix}$, then the Jacobian determinant of $\boldsymbol{F}$ is sometimes denoted $\begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix}$ or $\frac{\partial(x,y)}{\partial(u,v)}$.

EXAMPLES

Example 3

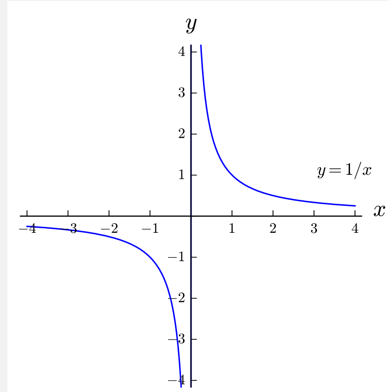
Let $\boldsymbol{H}\colon \mathbb{R}^3 \to \mathbb{R}^3$ be given by the formula $\boldsymbol{H}(x,y,z) = \begin{bmatrix} 3x+2y-z \\ x^2+y^3 \\ 4xz^2 \end{bmatrix}$. Find the Jacobian determinant of $\boldsymbol{H}$.

SOLUTION We compute $D\boldsymbol{H}(x,y,z) = \begin{bmatrix} 3 & 2 & -1 \\ 2x & 3y^2 & 0 \\ 4z^2 & 0 & 8xz \end{bmatrix}$. Hence, by expanding along the bottom row of this matrix, we deduce that $\det D\boldsymbol{H}(x,y,z) = 4z^2 \cdot (0-(-3y^2)) - 0 \cdot (0-(-2x)) + 8xz \cdot (9y^2 - 4x) = 12y^2z^2 + 72xy^2z - 32x^2z$.

Example 4

Let $\boldsymbol{F}\colon \mathbb{R}^2 \to \mathbb{R}^2$ be given by the formula $\boldsymbol{F}(x,y) = \begin{bmatrix} 6x+y^2 \\ 3x^2+2y \end{bmatrix}$. Find and plot all points (x,y) in $\mathbb{R}^2$ for which $\det D\boldsymbol{F}(x,y) = 0$.

SOLUTION We compute $D\boldsymbol{F}(x,y) = \begin{bmatrix} 6 & 2y \\ 6x & 2 \end{bmatrix}$. Hence $\det D\boldsymbol{F}(x,y) = 6\cdot 2 - 6x \cdot 2y = 12 - 12xy$. Therefore $\det D\boldsymbol{F}(x,y) = 0$ yields $12 - 12xy = 0$, which is the same as $xy = 1$, which in turn is the equivalent to $y = \frac{1}{x}$. Hence, plotting all points for which $\det D\boldsymbol{F}(x,y) = 0$ is the same as drawing the graph of $y = \frac{1}{x}$, which is shown below.



COMMON MISTAKES

The derivative of a multivariable vector-valued function of the form $\boldsymbol{F}\colon \mathbb{R}^n \to \mathbb{R}^m$ is an $m \times n$ matrix. A very common mistake when finding the derivative of such a function is to make an $n \times m$ matrix rather than an $m \times n$ matrix. Additionally, even if an $m \times n$ matrix is correctly used, a common mistake is to arrange the partial derivatives in the wrong order within the matrix. The only correct way to form the derivative of a multivariable vector-valued function is by placing the partial derivatives as stated in the text, namely, by having each row in the matrix correspond to a component function (which yields m rows), and having the partial derivatives in each row correspond to the variables in $\boldsymbol{F}$ in the given order (which yields n columns).

EXERCISES

Basic Exercises

1–3 ■ Find the derivative of each function.

1. Let $\boldsymbol{F}\colon \mathbb{R}^3 \to \mathbb{R}^2$ be given by the formula $\boldsymbol{F}(x,y,z) = \begin{bmatrix} 3x^2+yz \\ xz^3+2y^5 \end{bmatrix}$.

2. Let $\boldsymbol{G}\colon \mathbb{R}^2 \to \mathbb{R}^4$ be given by the formula $\boldsymbol{G}(x,y) = \begin{bmatrix} e^x+\sin y \\ \ln(xy) \\ \sqrt{x} \\ x^3y \end{bmatrix}$.

3. Let $\boldsymbol{H}\colon \mathbb{R}^4 \to \mathbb{R}^3$ be given by the formula $\boldsymbol{H}(x,y,z,w) = \begin{bmatrix} x+2y+3z+4w \\ xy \\ 6zw \end{bmatrix}$.

4–6 ■ Find the Jacobian determinant of each function.

4. Let $\boldsymbol{F}\colon \mathbb{R}^2 \to \mathbb{R}^2$ be given by the formula $\boldsymbol{F}(x,y) = \begin{bmatrix} x^2y \\ 3x-y^2 \end{bmatrix}$.

5. Let $\boldsymbol{G}\colon \mathbb{R}^3 \to \mathbb{R}^3$ be given by the formula $\boldsymbol{G}(x,y,z) = \begin{bmatrix} 3x+2y+z \\ e^{5x+y} \\ yz \end{bmatrix}$.

6. Let $\boldsymbol{H}\colon \mathbb{R}^4 \to \mathbb{R}^4$ be given by the formula $\boldsymbol{H}(x,y,z,w) = \begin{bmatrix} 4x-3y+2z-w \\ 3y+zw \\ 4zw \\ w^3 \end{bmatrix}$.

---

7. Let $\boldsymbol{K}\colon \mathbb{R}^2 \to \mathbb{R}^2$ be given by the formula $\boldsymbol{K}(x,y) = \begin{bmatrix} x^2-12x-y^2 \\ xy \end{bmatrix}$. Find and plot all points (x,y) in $\mathbb{R}^2$ for which $\det D\boldsymbol{K}(x,y) = 0$.

31.3 Linear Approximations & Differentials

In *Calculus I*, we saw the formula for the tangent line for a single-variable real-valued function $y = f(x)$ at a point $x = a$, which is

$$L(x) = f(a) + f'(a)(x - a). \tag{1}$$

This formula was used to find approximate values for functions that are difficult to compute. Similarly, in *Calculus II*, we saw the formula for the tangent plane for a multivariable real-valued function $z = f(x, y)$ at a point $(x, y) = (a, b)$, which is

$$L(x, y) = f(a, b) + f_x(a, b)(x - a) + f_y(a, b)(y - b). \tag{2}$$

We now want the analog for multivariable vector-valued functions of the above formulas for the tangent line and the tangent plane. In contrast to tangent lines and tangent planes, which are easy to visualize, in the case of multivariable vector-valued functions there is no nice visualization of the analog of tangent lines and tangent planes. However, while we have no nice visualization, it is nonetheless possible to find a formula that is analogous to the formulas in Equation (1) and Equation (2) of this section, and that can similarly be used to approximate the original function.

The Linear Approximation of a Multivariable Vector-Valued Function

In the following discussion, we note that a point $\mathbb{R}^n$ can also be thought of as corresponding to a vector from the origin to the point. For example, if $p = (a, b)$ is a point in $\mathbb{R}^2$, we can also think of this point as corresponding to the vector $\left[\begin{smallmatrix} a \\ b \end{smallmatrix}\right]$. In principle, the name of this vector should not be p, because points and vectors are theoretically different things, but because the point (a, b) gives rise to the vector $\left[\begin{smallmatrix} a \\ b \end{smallmatrix}\right]$, and the other way around, we will avoid cumbersome notation and use the notation p for both (a, b) and $\left[\begin{smallmatrix} a \\ b \end{smallmatrix}\right]$. Whether p means the point or the vector can be understood from the context.

We now return to our question of the linear approximation of a Multivariable Vector-Valued Function. The key to finding analogs in our current context of Equation (1) and Equation (2) of this section is to observe in Equation (2) that the expression $f_x(a,b)(x-a)+f_y(a,b)(y-b)$ can be thought of as a product of the row matrix $\begin{bmatrix} f_x(a,b) & f_y(a,b) \end{bmatrix}$ and the column matrix $\begin{bmatrix} x-a \\ y-b \end{bmatrix}$. If we write $p = (a,b)$, then the row matrix $\begin{bmatrix} f_x(a,b) & f_y(a,b) \end{bmatrix}$ is just $Df(p)$. If we write $v = (x,y)$, then the column vector $\begin{bmatrix} x-a \\ y-b \end{bmatrix}$ is just $v - p$, where we think of v and p here as vectors. We can then rewrite Equation (2) as $L(v) = f(p) + Df(p)(v-p)$. This last formula looks completely analogous to Equation (1), and, even better, it holds for all multivariable vector-valued functions, as we now state.

Linear Approximations for Multivariable Vector-Valued Functions

Let $\boldsymbol{F}\colon \mathbb{R}^n \to \mathbb{R}^m$ be a function, and let p be a vector in $\mathbb{R}^n$. The **linear approximation** of $\boldsymbol{F}$ at p is the function $\boldsymbol{L}\colon \mathbb{R}^n \to \mathbb{R}^m$ given by the formula

$$\boldsymbol{L}(v) = \boldsymbol{F}(p) + D\boldsymbol{F}(p)(v-p). \tag{3}$$

Example 1

Let $\boldsymbol{F}\colon \mathbb{R}^2 \to \mathbb{R}^2$ be given by the formula $\boldsymbol{F}(x,y) = \begin{bmatrix} y+\sqrt{x} \\ x+\sqrt[3]{y} \end{bmatrix}$. Find the linear approximation of $\boldsymbol{F}$ at $p = (9,8)$.

SOLUTION We compute $D\boldsymbol{F}(x,y) = \begin{bmatrix} \frac{1}{2\sqrt{x}} & 1 \\ 1 & \frac{1}{2y^{\frac{2}{3}}} \end{bmatrix}$. Then $\boldsymbol{F}(9,8) = \begin{bmatrix} 11 \\ 11 \end{bmatrix}$ and $D\boldsymbol{F}(9,8) = \begin{bmatrix} \frac{1}{6} & 1 \\ 1 & \frac{1}{12} \end{bmatrix}$, and we deduce that

$$\begin{aligned} \boldsymbol{L}(x,y) &= \boldsymbol{F}(9,8) + D\boldsymbol{F}(9,8)\left(\begin{bmatrix} x \\ y \end{bmatrix} - \begin{bmatrix} 9 \\ 8 \end{bmatrix}\right) \\ &= \begin{bmatrix} 11 \\ 11 \end{bmatrix} + \begin{bmatrix} \frac{1}{6} & 1 \\ 1 & \frac{1}{12} \end{bmatrix}\begin{bmatrix} x-9 \\ y-8 \end{bmatrix} = \begin{bmatrix} \frac{1}{6}x + y + \frac{3}{2} \\ x + \frac{1}{12}y + \frac{4}{3} \end{bmatrix}. \end{aligned}$$

The value of the linear approximation is, as its name suggests, in calculating the approximate value of a function that is hard to calculate directly. Suppose we have a multivariable vector-valued function $\boldsymbol{F}\colon \mathbb{R}^n \to \mathbb{R}^m$, and we wish to find an approximate value for $\boldsymbol{F}(v)$ for some point v in $\mathbb{R}^n$, due to the fact that $\boldsymbol{F}(v)$ is hard to calculate directly. As was the case with linear approximations for single-variable real-valued and multivariable real-valued functions, here too our ability to make use of Equation (3) to approximate $\boldsymbol{F}(v)$ depends upon finding a convenient point p such that on the one hand p is very close to v (the closer p is to v the better the approximation will be), and such that on the other hand it is easy to compute both $\boldsymbol{F}(p)$ and $D\boldsymbol{F}(p)$. Of course, in practice, it is not always possible to find such a point p, but it is possible in enough cases that this very easy to use method of approximation is of value.

Example 2

Let $\boldsymbol{F}\colon \mathbb{R}^2 \to \mathbb{R}^2$ be given by the formula $\boldsymbol{F}(x, y) = \begin{bmatrix} y+\sqrt{x} \\ x+\sqrt[3]{y} \end{bmatrix}$. Use the linear approximation to compute an approximate value for $\boldsymbol{F}(9.1, 7.8)$.

SOLUTION We saw in Example 1 of this section that the linear approximation of $\boldsymbol{F}$ at $p = (9, 8)$ is

$$\boldsymbol{L}(x, y) = \begin{bmatrix} 11 \\ 11 \end{bmatrix} + \begin{bmatrix} \frac{1}{6} & 1 \\ 1 & \frac{1}{12} \end{bmatrix} \begin{bmatrix} x - 9 \\ y - 8 \end{bmatrix}.$$

Then

$$\boldsymbol{F}(9.1, 7.8) \approx \boldsymbol{L}(9.1, 7.8) = \begin{bmatrix} 11 \\ 11 \end{bmatrix} + \begin{bmatrix} \frac{1}{6} & 1 \\ 1 & \frac{1}{12} \end{bmatrix} \begin{bmatrix} 9.1 - 9 \\ 7.8 - 8 \end{bmatrix} = \frac{1}{60} \begin{bmatrix} 649 \\ 665 \end{bmatrix}.$$

The Differential of a Multivariable Vector-Valued Function

In *Calculus I*, we saw an alternative way of formulating the idea of linear approximation, which was via differentials. Consider first a single-variable real-valued function $y = f(x)$ at a point $x = a$. Suppose we are able to compute $f(a)$, but we then want to change the value of x by Δx, where Δx is used to denote some small change in the value of x. That is, we want to find $f(a + \Delta x)$. The question then arise, if we change the value of x by Δx, how much will the value of y change. We let this change in y be denoted Δy, so that we have $\Delta y = f(a + \Delta x) - f(a)$. If Δx is close to zero, we saw that

$$\begin{aligned} \Delta y &= f(a + \Delta x) - f(a) \\ &\approx L(a + \Delta x) - f(a) = f(a) + f'(a)((a + \Delta x) - a) - f(a) \\ &= f'(a)\Delta x. \end{aligned} \tag{4}$$

We now rewrite Equation (4) in a different notation. We let $dx = \Delta x$, and we let

$$dy = f'(a)dx.$$

With this new notation, we can rewrite Equation (4) as $\Delta y \approx dy$. Hence, if we want to find an approximate value for Δy, we can compute dy, which is often much easier to compute.

Similarly, in *Calculus II*, we saw the same idea in the context of multivariable real-valued functions, where we want to approximate the function $z = f(x, y)$ at the point $(x, y) = (a, b)$. In this case, we change the value of x by Δx, and the value of y by Δy, and we want to compute the change in z, which is defined to be $\Delta z = f(a + \Delta x, b + \Delta y) - f(a, b)$. As before, if Δx and Δy are close to zero, then it is seen that $\Delta z \approx f_x(a, b)\Delta x + f_y(a, b)\Delta y$. In this case, we let $dx = \Delta x$ and $dy = \Delta y$, and we let

$$dz = f_x(a, b)dx + f_y(a, b)dy = \frac{\partial z}{\partial x}dx + \frac{\partial z}{\partial y}dy,$$

which is the same as $dz = L(a + \Delta x, b + \Delta y) - L(a, b)$, where we observe that $L(a, b) = f(a, b)$. It is then seen that $\Delta z \approx dz$.

We now want the analog for multivariable vector-valued functions of the above formulas for differentials. Let $\boldsymbol{F}\colon \mathbb{R}^n \to \mathbb{R}^m$ be a function, and let p be a point in $\mathbb{R}^n$. We want to approximate the function $z = \boldsymbol{F}(v)$ at the point $v = p$ in $\mathbb{R}^n$. In this case, we change the value of v by Δv, and we want to compute the change in z, which is defined to be $\Delta z = \boldsymbol{F}(p + \Delta v) - \boldsymbol{F}(p)$. As before, if Δv is close to the origin, then we use Equation (3) of this section to obtain

$$\begin{aligned}\Delta z &= \boldsymbol{F}(p + \Delta v) - \boldsymbol{F}(p)\\ &\approx \boldsymbol{L}(p + \Delta v) - \boldsymbol{F}(p)\\ &= \boldsymbol{F}(p) + D\boldsymbol{F}(p)((p + \Delta v) - p) - \boldsymbol{F}(p)\\ &= D\boldsymbol{F}(p)\Delta v.\end{aligned} \tag{5}$$

We let $dv = \Delta v$, and we let

$$dz = D\boldsymbol{F}(p)dv, \tag{6}$$

which, by Equation (5), is the same as $dz = \boldsymbol{L}(p + \Delta v) - \boldsymbol{L}(p)$, where we observe that $\boldsymbol{L}(p) = \boldsymbol{F}(p)$. Once again, it is then seen that $\Delta z \approx dz$.

When working with multivariable vector-valued functions, it is common not to assign a single letter, such as "z," to denote the value of $\boldsymbol{F}(v)$. As such, we can rewrite $\Delta z = \boldsymbol{F}(p+\Delta v)-\boldsymbol{F}(p)$ as $\Delta\boldsymbol{F}(p) = \boldsymbol{F}(p+\Delta v)-\boldsymbol{F}(p)$, and we can rewrite Equation (6) as

$$d\boldsymbol{F}(p) = D\boldsymbol{F}(p)dv.$$

If Δv is close to the origin, then $\Delta\boldsymbol{F}(p) \approx d\boldsymbol{F}(p)$. Hence, if we want to find an approximate value for $\Delta\boldsymbol{F}(p)$, we can compute $d\boldsymbol{F}(p)$, which is often much easier to compute.

The above ideas can be summarized as follows.

Differentials for Multivariable Vector-Valued Functions

Let $\boldsymbol{F}\colon \mathbb{R}^n \to \mathbb{R}^m$ be a function, and let p be a point in $\mathbb{R}^n$. Let $z = \boldsymbol{F}(v)$.

1. Let Δv be a point in $\mathbb{R}^n$ that is close to the origin. The **change** in $\boldsymbol{F}$ at p, denoted Δz or $\Delta\boldsymbol{F}(p)$, is defined by

$$\Delta z = \Delta\boldsymbol{F}(p) = \boldsymbol{F}(p + \Delta v) - \boldsymbol{F}(p)$$

2. Let $dv = \Delta v$. The **differential** of $\boldsymbol{F}$ at p, denoted dz or $d\boldsymbol{F}(p)$, is defined by

$$dz = d\boldsymbol{F}(p) = D\boldsymbol{F}(p)dv.$$

3. If Δv is close to the origin, then $d\boldsymbol{F}(p)$ can be used as an approximation to $\Delta\boldsymbol{F}(p)$

Example 3

Let $\boldsymbol{F}\colon \mathbb{R}^2 \to \mathbb{R}^2$ be given by the formula $\boldsymbol{F}(x, y) = \begin{bmatrix} y+\sqrt{x} \\ x+\sqrt[3]{y} \end{bmatrix}$. Find $d\boldsymbol{F}(p)$.

SOLUTION We saw in Example 1 of this section that $D\boldsymbol{F}(x, y) = \begin{bmatrix} \frac{1}{2\sqrt{x}} & 1 \\ 1 & \frac{1}{2y^{\frac{2}{3}}} \end{bmatrix}$. If we write $dv = \begin{bmatrix} dx \\ dy \end{bmatrix}$, then

$$dz = d\boldsymbol{F}(p) = D\boldsymbol{F}(x, y)dv = \begin{bmatrix} \frac{1}{2\sqrt{x}} & 1 \\ 1 & \frac{1}{2y^{\frac{2}{3}}} \end{bmatrix} \begin{bmatrix} dx \\ dy \end{bmatrix} = \begin{bmatrix} \frac{1}{2\sqrt{x}}dx + dy \\ dx + \frac{1}{2y^{\frac{2}{3}}}dy \end{bmatrix}.$$

Is important to note that using differentials to approximate certain values is not really any different from using the linear approximation; it is just a different way of writing it, and it emphasizes a different way of thinking about it: with the linear approximation, we think of the tangent line as being close to the original function at the point of tangency, whereas with differentials, we think of approximating a change in the value of the "output" variable when we make a small change in the value of the "input" variable.

SUMMARY

Linear Approximations for Multivariable Vector-Valued Functions

Let $\boldsymbol{F}\colon \mathbb{R}^n \to \mathbb{R}^m$ be a function, and let p be a vector in $\mathbb{R}^n$. The **linear approximation** of $\boldsymbol{F}$ at p is the function $\boldsymbol{L}\colon \mathbb{R}^n \to \mathbb{R}^m$ given by the formula

$$\boldsymbol{L}(v) = \boldsymbol{F}(p) + D\boldsymbol{F}(p)(v - p).$$

Differentials for Multivariable Vector-Valued Functions

Let $\boldsymbol{F}\colon \mathbb{R}^n \to \mathbb{R}^m$ be a function, and let p be a point in $\mathbb{R}^n$. Let $z = \boldsymbol{F}(v)$.

1. Let Δv be a point in $\mathbb{R}^n$ that is close to the origin. The **change** in $\boldsymbol{F}$ at p, denoted Δz or $\Delta\boldsymbol{F}(p)$, is defined by

$$\Delta z = \Delta\boldsymbol{F}(p) = \boldsymbol{F}(p + \Delta v) - \boldsymbol{F}(p)$$

2. Let $dv = \Delta v$. The **differential** of $\boldsymbol{F}$ at p, denoted dz or $d\boldsymbol{F}(p)$, is defined by

$$dz = d\boldsymbol{F}(p) = D\boldsymbol{F}(p)dv.$$

3. If Δv is close to the origin, then $d\boldsymbol{F}(p)$ can be used as an approximation to $\Delta\boldsymbol{F}(p)$

EXAMPLES

Example 4

Let $\boldsymbol{F}\colon \mathbb{R}^2 \to \mathbb{R}^2$ be a function. Suppose that $\boldsymbol{F}(1,2) = \left[\begin{smallmatrix} 5 \\ 7 \end{smallmatrix}\right]$ and that $D\boldsymbol{F}(1,2) = \left[\begin{smallmatrix} 3 & 1 \\ 4 & 5 \end{smallmatrix}\right]$. Estimate the value of $\boldsymbol{F}(1.02, 1.97)$.

SOLUTION Using Equation (3) of this section, we see that

$$\begin{aligned} \boldsymbol{F}(1.02, 1.97) \approx \boldsymbol{L}(1.02, 1.97) &= \boldsymbol{F}(1,2) + D\boldsymbol{F}(1,2)\left(\begin{bmatrix} 1.02 \\ 1.97 \end{bmatrix} - \begin{bmatrix} 1 \\ 2 \end{bmatrix}\right) \\ &= \begin{bmatrix} 5 \\ 7 \end{bmatrix} + \begin{bmatrix} 3 & 1 \\ 4 & 5 \end{bmatrix}\begin{bmatrix} 0.02 \\ -0.03 \end{bmatrix} = \begin{bmatrix} 5.03 \\ 6.93 \end{bmatrix}. \end{aligned}$$

COMMON MISTAKES

There is a question regarding the exact meaning of the symbols dx, dy, dz, etc. Some people think of these symbols as denoting an "infinitesimally small" changes in these variables, whatever infinitesimally small might mean. From a historical perspective, the idea of infinitesimally small changes in variables was indeed used in the early period of the development of Calculus, and for some people that approach is still intuitively appealing. However, it is important to note that in the modern approach to calculus (now over 150 years old), there are no such things as infinitesimally small real numbers (as can be seen in a course on real analysis). Hence, it would be a mistake to think that dx, dy, dz, etc., are actually infinitesimally small numbers, even though thinking about them that way might be convenient or intuitively appealing. Unfortunately, it takes more advanced mathematics than is available to us here to give a rigorous definition of what such things really are, though there is indeed such a definition. At the same time, there now exists a completely rigorous way to define infinitesimally small numbers, not as part of the real numbers (which is not possible) but as some other kind of number that can be added on to the real numbers; the study of that approach, which would take us very far afield, is called non-standard analysis.

EXERCISES

Basic Exercises

1–2 ■ For each function, use the linear approximation to compute the requested approximation.

1. Let $\boldsymbol{F}\colon \mathbb{R}^2 \to \mathbb{R}^2$ be given by the formula $\boldsymbol{F}(x, y) = \begin{bmatrix} \sqrt{xy} \\ \frac{y}{x} \end{bmatrix}$. Compute an approximate value for $\boldsymbol{F}(1.9, 18.2)$.

2. Let $\boldsymbol{G}\colon \mathbb{R}^3 \to \mathbb{R}^2$ be given by the formula $\boldsymbol{G}(x, y)z = \begin{bmatrix} e^{x+2y+3z} \\ \sqrt[3]{xyz} \end{bmatrix}$. Compute an approximate value for $\boldsymbol{G}(1.1, 0.9, -1.1)$.

31.4 The Chain Rule

In *Calculus I*, we saw a number of methods for finding the derivatives of single-variable real-valued functions, for example the Product Rule, the Quotient Rule and the Chain Rule. Whereas the Product Rule does not have a simple analog for multivariable vector-valued functions, and the Quotient Rule has no analog at all for such functions, it turns out that the Chain Rule has a very nice analog for multivariable vector-valued functions, as we will now see. There are two (equivalent) ways to state the Chain Rule for multivariable vector-valued functions, one approach using matrices and matrix multiplication, and the other not using matrices. We will start with the matrix formulation, because the formula for the Chain Rule for multivariable vector-valued functions in the matrix approach looks just like the Chain Rule formula from *Calculus I*, and because it shows the power of matrices; the matrix approach to the Chain Rule is the "right way to do it." We will then show the non-matrix formulation, which, while not as aesthetically appealing as the matrix approach, is useful in practice.

The version of the Chain Rule from *Calculus I* is used to find the derivatives of functions such as $h\colon \mathbb{R} \to \mathbb{R}$ given by the formula $h(x) = \sin(x^2 + 7)$. The idea is to decompose this type of function into the composition of two simpler functions, where one of the simpler functions is "inside" the other function. For example, for the function $h(x) = \sin(x^2 + 7)$, we define new functions $f\colon \mathbb{R} \to \mathbb{R}$ and $g\colon \mathbb{R} \to \mathbb{R}$ given by the formulas $f(x) = \sin x$ and $g(x) = x^2 + 7$, and we then see that $h(x) = f(g(x))$. Therefore, in order to find the derivative of $h(x)$, we need to find the derivative of $f(g(x))$, and the latter type of expression is precisely the subject of the Chain Rule, which says that

$$[f(g(x))]' = f'(g(x))g'(x).$$

For example, we see that $[\sin(x^2 + 7)]' = \cos(x^2 + 7) \cdot 2x$.

We now want to rewrite the formula for the Chain Rule for single-variable real-valued functions using a slightly different notation, which you might or might not have seen previously. Whether or not you have already seen this notation, it is very important to become familiar with this notation, because it is used extensively in advanced mathematics.

Composition of Functions

Rather than starting with a single function, for example $h\colon \mathbb{R} \to \mathbb{R}$ given by the formula $h(x) = \sin(x^2 + 7)$, and then decomposing it by writing $h(x) = f(g(x))$, where $f : \mathbb{R} \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$ are given by the formulas $f(x) = \sin x$ and $g(x) = x^2 + 7$, we now want to start with two functions such as f and g and then combine them.

For example, let $f : \mathbb{R} \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$ be given by the formulas $f(x) = x^2$ and $g(x) = x + 3$. We then want to construct a new function, which will be given by the formula $f(g(x)) = (x + 3)^2$. It would be very convenient to have a name for this combined function. We could call this function by a new letter, for example writing $k\colon \mathbb{R} \to \mathbb{R}$ as $k(x) = (x+3)^2$, but calling the function that results from combining f and g by an arbitrary name such as k is confusing. It would be much better to give the new function a name that reflects that it is made up of f and g, and that is what we now state.

Let $f : \mathbb{R} \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$ be functions. The **composition** of f and g is the function $f \circ g : \mathbb{R} \to \mathbb{R}$ given by the formula

$$(f \circ g)(x) = f(g(x)). \tag{1}$$

For example, let $f : \mathbb{R} \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$ be given by the formulas $f(x) = x^2$ and $g(x) = x + 3$. Then $(f \circ g)(x) = f(g(x)) = (x + 3)^2$, and $(g \circ f)(x) = g(f(x)) = x^2 + 3$. Observe that $f \circ g$ is not the same as $g \circ f$.

The reader who is encountering the notation $f \circ g$ for the first time might find it necessary to get used to the fact that it is "backwards" from what might be expected, because $f \circ g$ means doing g first and then f even though we generally read from left to right in English. Think of "$\circ$" as meaning "following." We will stick with the "$\circ$" notation in spite of any slight confusion it might cause at first, because it is extremely widespread, and because the reader will find that it works well once she is used to it.

It is important to look carefully at the definition of composition. The notation "$f \circ g$" is the name of one single-variable real-valued function, which we constructed out of the two functions f and g. In Equation (1), we say how to compute this function, which means, we say what this function does to each real number x. Specifically, this equation says that the value of the function $f \circ g$ at the number x, which is denoted $(f \circ g)(x)$, is computed by doing $f(g(x))$.

Observe that the parentheses in the expression "$(f \circ g)(x)$" need to be used exactly as written. It would not be correct to write "$f \circ g(x)$," because $\circ$ is an operation that combines two functions, whereas "$f(x)$" is not a function but a single number.

Composition of functions works for multivariable vector-valued functions just as it does for single-variable real-valued functions, with one caveat, which is that the "output" of the first function has to be the "input" of the second function in order to form a composition of two functions.

For example, if $\boldsymbol{F}\colon \mathbb{R}^3 \to \mathbb{R}^4$ and $\boldsymbol{G}\colon \mathbb{R}^2 \to \mathbb{R}^3$ are functions then we can form the composition $\boldsymbol{F} \circ \boldsymbol{G}\colon \mathbb{R}^2 \to \mathbb{R}^4$, because the "output" of $\boldsymbol{G}$ and the "input" of $\boldsymbol{F}$ are both $\mathbb{R}^3$, but we cannot form the composition $\boldsymbol{G} \circ \boldsymbol{F}$, because the "output" of $\boldsymbol{F}$ is not the same as the "input" of $\boldsymbol{G}$.

More generally, we have the following definition.

Composition of Functions

Let $\boldsymbol{G}\colon \mathbb{R}^n \to \mathbb{R}^k$ and $\boldsymbol{F}\colon \mathbb{R}^k \to \mathbb{R}^m$ be functions. The **composition** of $\boldsymbol{F}$ and $\boldsymbol{G}$ is the function $\boldsymbol{F} \circ \boldsymbol{G}\colon \mathbb{R}^n \to \mathbb{R}^m$ given by the formula

$$(\boldsymbol{F} \circ \boldsymbol{G})(p) = \boldsymbol{F}(\boldsymbol{G}(p)). \tag{2}$$

Example 1

let $\boldsymbol{F}\colon \mathbb{R}^3 \to \mathbb{R}^4$ and $\boldsymbol{G}\colon \mathbb{R}^2 \to \mathbb{R}^3$ be given by the formulas

$$\boldsymbol{F}(x, y, z) = \begin{bmatrix} 5x+y \\ 3z \\ 2xz \\ y-z \end{bmatrix} \quad \text{and} \quad \boldsymbol{G}(x, y) = \begin{bmatrix} xy \\ x+2y \\ x-y \end{bmatrix}.$$

Find the formula for $(\boldsymbol{F} \circ \boldsymbol{G})(x, y)$.

SOLUTION The function $\boldsymbol{F} \circ \boldsymbol{G}\colon \mathbb{R}^2 \to \mathbb{R}^4$ is given by the formula

$$(\boldsymbol{F} \circ \boldsymbol{G})(x, y) = \begin{bmatrix} 5(xy)+(x+2y) \\ 3(x-y) \\ 2(xy)(x-y) \\ (x+2y)-(x-y) \end{bmatrix} = \begin{bmatrix} x+2y+5xy \\ 3x-2y \\ 2x^2y-2xy^2 \\ 3y \end{bmatrix}.$$

Now that we have the notion of composition of functions, we can restate the the Chain Rule for single-variable real-valued functions as

$$[(f \circ g)(x)]' = f'(g(x))g'(x). \tag{3}$$

For example, suppose we want to find the derivative of the function $p\colon \mathbb{R} \to \mathbb{R}$ be given by the formula $p(x) = \cos(x^2)$. We then define functions $h\colon \mathbb{R} \to \mathbb{R}$ and $k\colon \mathbb{R} \to \mathbb{R}$ by the formulas $h(x) = \cos x$ and $k(x) = x^2$, and we observe that $p = h \circ k$. We then use the Chain Rule to compute

$$[\cos(x^2)]' = p'(x) = [(h \circ k)(x)]' = h'(k(x))k'(x) = -\sin(x^2) \cdot 2x.$$

Of course, in practice, when someone wants to compute $[\cos(x^2)]'$, she would not given names to the functions h and k, and would simply write $[\cos(x^2)]' = -\sin(x^2) \cdot 2x$, but we wanted to emphasize that the Chain Rule is really about the composition of functions.

We now turn to the analog of Equation (3) for multivariable vector-valued functions.

The Chain Rule via Matrix Multiplication

Let $\boldsymbol{G}\colon \mathbb{R}^n \to \mathbb{R}^k$ and $\boldsymbol{F}\colon \mathbb{R}^k \to \mathbb{R}^m$ be functions. We want to find a formula for $D(\boldsymbol{F} \circ \boldsymbol{G})(p)$ that is analogous to Equation (3) of this section.

If we examine the right-hand side of Equation (3), we see that the reason we can multiply $f'(g(x))$ and $g'(x)$ is that for each real number x, the quantities $f'(g(x))$ and $g'(x)$ are also real numbers, and we know that real numbers can be multiplied. The analogs for multivariable vector-valued functions of $f'(g(x))$ and $g'(x)$ would be $D\boldsymbol{F}(\boldsymbol{G}(p))$ and $D\boldsymbol{G}(p)$. What are these two quantities? They are matrices, and, more specifically, we know that $D\boldsymbol{F}(\boldsymbol{G}(p))$ is an $m \times k$ matrix and $D\boldsymbol{G}(p)$ is a $k \times n$ matrix. Recalling the definition of matrix multiplication, we see from the dimensions of the matrices $D\boldsymbol{F}(\boldsymbol{G}(p))$ and $D\boldsymbol{G}(p)$ that they can in fact be multiplied. Indeed, not only can these two matrices be multiplied, but taking their product is exactly what we need for the Chain Rule for multivariable vector-valued functions, which we now state.

The Chain Rule via Matrix Multiplication

Let $\boldsymbol{G}\colon \mathbb{R}^n \to \mathbb{R}^k$ and $\boldsymbol{F}\colon \mathbb{R}^k \to \mathbb{R}^m$ be functions, and let p be a point in $\mathbb{R}^n$. Suppose that $\boldsymbol{F}$ and $\boldsymbol{G}$ are differentiable. Then

$$D(\boldsymbol{F} \circ \boldsymbol{G})(p) = D\boldsymbol{F}(\boldsymbol{G}(p))\, D\boldsymbol{G}(p), \tag{4}$$

where the multiplication is matrix multiplication.

We will not prove Equation (4), but let us look at an example, to see that it really works.

Example 2

Let $\boldsymbol{F}\colon \mathbb{R}^3 \to \mathbb{R}^2$ and $\boldsymbol{G}\colon \mathbb{R}^2 \to \mathbb{R}^3$ be given by the formulas

$$\boldsymbol{F}(x,y,z) = \begin{bmatrix} 2xy \\ y-z \end{bmatrix} \quad \text{and} \quad \boldsymbol{G}(x,y) = \begin{bmatrix} x-4y \\ xy \\ x+y \end{bmatrix}.$$

Find each of $D\boldsymbol{F}(\boldsymbol{G}(x,y))$ and $D\boldsymbol{G}(x,y)$ and $D(\boldsymbol{F}\circ\boldsymbol{G})(x,y)$ directly, and verify the The Chain Rule via Matrix Multiplication for this example.

SOLUTION The function $\boldsymbol{F}\circ\boldsymbol{G}\colon \mathbb{R}^2 \to \mathbb{R}^2$ is given by the formula

$$(\boldsymbol{F}\circ\boldsymbol{G})(x,y) = \begin{bmatrix} 2(x-4y)(xy) \\ (xy)-(x+y) \end{bmatrix} = \begin{bmatrix} 2x^2y-8xy^2 \\ xy-x-y \end{bmatrix}.$$

Next, we compute the derivatives of our three functions, which are

$$\begin{aligned} D\boldsymbol{F}(x,y,z) &= \begin{bmatrix} 2y & 2x & 0 \\ 0 & 1 & -1 \end{bmatrix} \\ D\boldsymbol{G}(x,y) &= \begin{bmatrix} 1 & -4 \\ y & x \\ 1 & 1 \end{bmatrix} \\ D(\boldsymbol{F}\circ\boldsymbol{G})(x,y) &= \begin{bmatrix} 4xy-8y^2 & 2x^2-16xy \\ y-1 & x-1 \end{bmatrix}. \end{aligned}$$

We then have

$$D\boldsymbol{F}(\boldsymbol{G}(x,y)) = \begin{bmatrix} 2(xy) & 2(x-4y) & 0 \\ 0 & 1 & -1 \end{bmatrix} = \begin{bmatrix} 2xy & 2x-8y & 0 \\ 0 & 1 & -1 \end{bmatrix}.$$

Finally, we compute

$$\begin{aligned} D\boldsymbol{F}(\boldsymbol{G}(x,y))\,D\boldsymbol{G}(x,y) &= \begin{bmatrix} 2xy & 2x-8y & 0 \\ 0 & 1 & -1 \end{bmatrix}\begin{bmatrix} 1 & -4 \\ y & x \\ 1 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 2xy\cdot 1+(2x-8y)\cdot y+0\cdot 1 & 2xy\cdot(-4)+(2x-8y)\cdot x+0\cdot 1 \\ 0\cdot 1+1\cdot y-1\cdot 1 & 0\cdot(-4)+1\cdot x-1\cdot 1 \end{bmatrix} \\ &= \begin{bmatrix} 4xy-8y^2 & 2x^2-16xy \\ y-1 & x-1 \end{bmatrix} \\ &= D(\boldsymbol{F}\circ\boldsymbol{G})(x,y). \end{aligned}$$

Hence, we see that the The Chain Rule via Matrix Multiplication works for this example.

The Chain Rule without Matrices

Whereas the nicest-looking, and most general, formulation of the Chain Rule for multivariable vector-valued functions is via matrix multiplication, it is also possible to state and use specific cases of this Chain Rule without matrices. This non-matrix formulation of the Chain Rule will be written entirely using Leibniz notation for derivatives. To start, let us recall the statement of the Chain Rule for single-variable real-valued functions in Leibniz Notation, which is as follows.

Let $f\colon \mathbb{R} \to \mathbb{R}$ and $g\colon \mathbb{R} \to \mathbb{R}$ be functions. We write $y = f(u)$ and $u = g(x)$. Suppose that f and g are differentiable. Then

$$\frac{dy}{dx} = \frac{dy}{du}\frac{du}{dx}. \tag{5}$$

We make a few observations about this formulation of the Chain Rule for single-variable real-valued functions. First, the formula in Equation (5) is completely equivalent to the formula in Equation (3); only the notation has changed. Second, observe that in Equation (5), we did not make use of the names of the function f and g, but rather, all we used was the names of the variables. Third, whereas it appears as if we could prove Equation (5) by simply canceling the two appearances of du in the right-hand side of the equation, such canceling, which useful mnemonically, is not actually a valid thing to do. We will soon see why it is not valid.

To see the non-matrix version of the Chain Rule for multivariable vector-valued functions, we first look at various special cases, starting with functions of the form $f\colon \mathbb{R}^2 \to \mathbb{R}$ and $\boldsymbol{r}\colon \mathbb{R} \to \mathbb{R}^2$, and with the composition $f \circ \boldsymbol{r}$. Let t be a real number. By Equation (4) of this section we see that

$$D(f \circ \boldsymbol{r})(t) = Df(\boldsymbol{r}(t))\, D\boldsymbol{r}(t). \tag{6}$$

To write out the matrices in the above equation, let us first write $z = f(x, y)$ and $r(t) = \left[\begin{smallmatrix} x(t) \\ y(t) \end{smallmatrix}\right]$. Then Equation (6) can be rewritten as

$$\begin{bmatrix} \dfrac{dz}{dt} \end{bmatrix} = \begin{bmatrix} \dfrac{\partial z}{\partial x} & \dfrac{\partial z}{\partial y} \end{bmatrix} \begin{bmatrix} \frac{dx}{dt} \\ \frac{dy}{dt} \end{bmatrix}.$$

Multiplying the matrices on the right-hand side of this equation gives

$$\begin{bmatrix} \dfrac{dz}{dt} \end{bmatrix} = \begin{bmatrix} \dfrac{\partial z}{\partial x}\dfrac{dx}{dt} + \dfrac{\partial z}{\partial y}\dfrac{dy}{dt} \end{bmatrix}.$$

Of course, using 1×1 matrices is silly, and we simply equate the entries in these two matrices to obtain

$$\frac{dz}{dt} = \frac{\partial z}{\partial x}\frac{dx}{dt} + \frac{\partial z}{\partial y}\frac{dy}{dt}. \tag{7}$$

This last equation is the Chain Rule in the particular case of a function of the form $z = f(x, y)$, where each of x and y are functions of a single variable t.

Observe that in the above version of the Chain Rule, we write single-variable derivatives $\frac{dz}{dt}$, $\frac{dx}{dt}$ and $\frac{dy}{dt}$, whereas we write partial derivatives $\frac{\partial z}{\partial x}$ and $\frac{\partial z}{\partial y}$. It is important to keep track of which functions have one variable, which then use the notation for single-variable derivatives, and which functions have more than one variable, which then use the notation for partial derivatives.

Example 3

Let $z = x^2 + y^3$, and $x = \sqrt{t}$ and $y = \sin t$. Find $\frac{dz}{dt}$.

SOLUTION First, we compute the various derivatives (some partial and some regular) that appear in the right-hand side of Equation (7) of this section, obtaining $\frac{\partial z}{\partial x} = 2x$, and $\frac{\partial z}{\partial y} = 3y^2$, and $\frac{dx}{dt} = \frac{1}{2\sqrt{t}}$ and $\frac{dy}{dt} = \cos t$. Next, using Equation (7), we compute

$$\begin{aligned}\frac{dz}{dt} &= \frac{\partial z}{\partial x}\frac{dx}{dt} + \frac{\partial z}{\partial y}\frac{dy}{dt}\\ &= 2x \cdot \frac{1}{2\sqrt{t}} + 3y^2 \cdot \cos t\\ &= 2\sqrt{t} \cdot \frac{1}{2\sqrt{t}} + 3(\sin t)^2 \cdot \cos t\\ &= 1 + 3\sin^2 t \cos t.\end{aligned}$$

Observe in Example 3 of this section, that after using Equation (7), which yielded an expression that contained all three variables x, y and t, we then rewrote each of x and y in terms of t, so that the final answer involves only the variable t. It would not have been proper to leave the answer involving all of x, y and t, because we were asked to find $\frac{dz}{dt}$, which means that we are thinking of z as a function of t, and hence this derivative should be a function of t.

We will skip the details, but the same type of argument as above in the case of $w = f(x, y, z)$, where each of x, y and z are functions of the single variable t, yields the equation

$$\frac{dw}{dt} = \frac{\partial w}{\partial x}\frac{dx}{dt} + \frac{\partial w}{\partial y}\frac{dy}{dt} + \frac{\partial w}{\partial z}\frac{dz}{dt}.$$

Clearly, the same idea holds for functions of more than three variables.

Next, we look at at functions of the form $f : \mathbb{R}^2 \to \mathbb{R}$ and $\boldsymbol{G} : \mathbb{R}^2 \to \mathbb{R}^2$, with the composition $f \circ \boldsymbol{G}$. We write $z = f(x, y)$ and $\boldsymbol{G}(x, y) = \begin{bmatrix} x(s,t) \\ y(s,t) \end{bmatrix}$.

Then by Equation (4) of this section we see that

$$D(f \circ \boldsymbol{G})(x, y) = Df(\boldsymbol{G}(x, y))\, D\boldsymbol{G}(x, y),$$

which, when written as matrices, yields,

$$\begin{bmatrix} \frac{\partial z}{\partial s} & \frac{\partial z}{\partial t} \end{bmatrix} = \begin{bmatrix} \frac{\partial z}{\partial x} & \frac{\partial z}{\partial y} \end{bmatrix} \begin{bmatrix} \frac{dx}{ds} & \frac{dy}{ds} \\ \frac{dx}{dt} & \frac{dy}{dt} \end{bmatrix}.$$

Multiplying the matrices on the right-hand side of this equation gives

$$\begin{bmatrix} \frac{\partial z}{\partial s} & \frac{\partial z}{\partial t} \end{bmatrix} = \begin{bmatrix} \frac{\partial z}{\partial x}\frac{\partial x}{\partial s} + \frac{\partial z}{\partial y}\frac{\partial y}{\partial s} & \frac{\partial z}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial z}{\partial y}\frac{\partial y}{\partial t} \end{bmatrix},$$

and then equating corresponding entries of these matrices yields

$$\frac{\partial z}{\partial s} = \frac{\partial z}{\partial x}\frac{\partial x}{\partial s} + \frac{\partial z}{\partial y}\frac{\partial y}{\partial s} \quad \text{and} \quad \frac{\partial z}{\partial t} = \frac{\partial z}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial z}{\partial y}\frac{\partial y}{\partial t}. \tag{8}$$

This last equation is the Chain Rule in the particular case of a function of the form $z = f(x, y)$, where each of x and y are functions of two variables s and t.

We can now see that the wish of beginning calculus students to cancel the two appearances of du in the right-hand side of Equation (5) is not a valid thing to do. If such a cancellation were valid, then it would plausibly also be valid to cancel the various instances of ∂x and ∂y in either of the equations of Equation (8), but doing so in the left-hand equation would yield $\frac{\partial z}{\partial s} = \frac{\partial z}{\partial s} + \frac{\partial z}{\partial s}$, which, upon cancelling, yields $1 = 1+1$, and that is clearly not possible. Hence, this type of canceling is not valid.

The same type of argument as above in the case of $w = f(x, y, z)$, where each of x, y and z are functions of two variables s and t, yields the equations

$$\frac{\partial w}{\partial s} = \frac{\partial w}{\partial x}\frac{\partial x}{\partial s} + \frac{\partial w}{\partial y}\frac{\partial y}{\partial s} + \frac{\partial w}{\partial z}\frac{\partial z}{\partial s} \quad \text{and} \quad \frac{\partial w}{\partial t} = \frac{\partial w}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial w}{\partial y}\frac{\partial y}{\partial t} + \frac{\partial w}{\partial z}\frac{\partial z}{\partial t}.$$

Clearly, the same idea holds for functions of more than three variables.

Example 4

Let $z = 5x^2y$, and $x = 3s + t$ and $y = \sin(st)$. Find $\frac{\partial z}{\partial s}$ and $\frac{\partial z}{\partial t}$.

SOLUTION First, we compute the various derivatives (some partial and some regular) that appear in the right-hand side of Equation (8) of this section, obtaining $\frac{\partial z}{\partial x} = 10xy$, and $\frac{\partial z}{\partial y} = 5x^2$, and $\frac{\partial x}{\partial s} = 3$, and $\frac{\partial x}{\partial t} = 1$, and $\frac{\partial y}{\partial s} = t\cos(st)$ and $\frac{dy}{dt} = s\cos(st)$. Next, using Equation (8), we compute

$$\begin{aligned}\frac{\partial z}{\partial s} &= \frac{\partial z}{\partial x}\frac{\partial x}{\partial s} + \frac{\partial z}{\partial y}\frac{\partial y}{\partial s}\\ &= 10xy \cdot 3 + 5x^2 \cdot t\cos(st)\\ &= 10(3s+t)\sin(st)\cdot 3 + 5(3s+t)^2 \cdot t\cos(st)\\ &= 30(3s+t)\sin(st) + 5t(3s+t)^2\cos(st)\end{aligned}$$

and

$$\begin{aligned}\frac{\partial z}{\partial t} &= \frac{\partial z}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial z}{\partial y}\frac{\partial y}{\partial t}\\ &= 10xy \cdot 1 + 5x^2 \cdot s\cos(st)\\ &= 10(3s+t)\sin(st) + 5s(3s+t)^2\cos(st).\end{aligned}$$

Observe in Example 4 of this section, that similarly to Example 3 of this section, after using Equation (8), which yielded an expression that contained all three variables x, y, s and t, we then rewrote each of x and y in terms of s and t, so that the final answer involves only the variable s and t.

Formulas such as those in Equation (8) are very convenient to use in practice, and, as we now mention, they can be worked out in any given situation without the use of matrices. Let us look at the particular case of a function of the form $z = f(x, y)$, where each of x and y are functions of two variables s and t. We can write the relation between these variables via a **tree diagram**, which, in this particular situation, is seen in Figure 1 of this section.

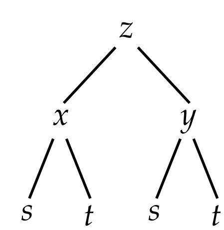


Figure 1: Tree diagram

Suppose we want to find the formula for $\dfrac{\partial z}{\partial s}$. We then consider the tree diagram in Figure 1, and we see that to go from z to s, there are two routes, one of which is via x and the other via y. These two routes tell us that in the Chain Rule for this particular situation, we have two terms, which are $\dfrac{\partial z}{\partial x}\dfrac{\partial x}{\partial s}$, representing the route via x, and $\dfrac{\partial z}{\partial y}\dfrac{\partial y}{\partial s}$, representing the route via y. We then add these terms, and obtain the equation on the left-hand side of Equation (8) of this section. Clearly, we can use the analogous method in other cases.

Finally, we can write out the general version of the Chain Rule without matrices as follows.

The Chain Rule Without Matrices

Let $z = f(x_1, \ldots, x_n)$, and where each of $x_1, \ldots, x_n$ is a function of $t_1, \ldots, t_m$. Then

$$\frac{\partial z}{\partial t_i} = \frac{\partial z}{\partial x_1}\frac{\partial x_1}{\partial t_i} + \frac{\partial z}{\partial x_2}\frac{\partial x_2}{\partial t_i} + \cdots + \frac{\partial z}{\partial x_n}\frac{\partial x_n}{\partial t_i}, \tag{9}$$

for each $i = 1, 2, \ldots, m$.

Ultimately, there is one Chain Rule for multivariable vector-valued functions, whether we write it via matrices or via equations such as Equation (8) of this section. Each of these formulations of the Chain Rule is useful in certain situations.

SUMMARY

Composition of Functions

Let $\boldsymbol{G}\colon \mathbb{R}^n \to \mathbb{R}^k$ and $\boldsymbol{F}\colon \mathbb{R}^k \to \mathbb{R}^m$ be functions. The **composition** of $\boldsymbol{F}$ and $\boldsymbol{G}$ is the function $\boldsymbol{F} \circ \boldsymbol{G}\colon \mathbb{R}^n \to \mathbb{R}^m$ given by the formula

$$(\boldsymbol{F} \circ \boldsymbol{G})(p) = \boldsymbol{F}(\boldsymbol{G}(p)).$$

The Chain Rule via Matrix Multiplication

Let $\boldsymbol{G}\colon \mathbb{R}^n \to \mathbb{R}^k$ and $\boldsymbol{F}\colon \mathbb{R}^k \to \mathbb{R}^m$ be functions, and let p be a point in $\mathbb{R}^n$. Suppose that $\boldsymbol{F}$ and $\boldsymbol{G}$ are differentiable. Then

$$D(\boldsymbol{F} \circ \boldsymbol{G})(p) = D\boldsymbol{F}(\boldsymbol{G}(p))\, D\boldsymbol{G}(p),$$

where the multiplication is matrix multiplication.

The Chain Rule Without Matrices

Let $z = f(x_1, \ldots, x_n)$, and where each of $x_1, \ldots, x_n$ is a function of $t_1, \ldots, t_m$. Then

$$\frac{\partial z}{\partial t_i} = \frac{\partial z}{\partial x_1}\frac{\partial x_1}{\partial t_i} + \frac{\partial z}{\partial x_2}\frac{\partial x_2}{\partial t_i} + \cdots + \frac{\partial z}{\partial x_n}\frac{\partial x_n}{\partial t_i},$$

for each $i = 1, 2, \ldots, m$.

EXAMPLES

Example 5

The height and width of a rectangle are each changing as a function of time. Suppose that at a certain instant of time, the height is 3 inches and is increasing at a rate of 2 inches per second, and the width is 5 inches and is decreasing at a rate of 1 inch per second. Find the rate of change of the area of the rectangle at that instant of time.

SOLUTION Let h denote the height of the rectangle and let w denote the width of the rectangle. The area of the rectangle as a function of height and width is the function $A\colon \mathbb{R}^2 \to \mathbb{R}$ given by the formula $A(h, w) = hw$. Then $\frac{\partial A}{\partial h} = w$ and $\frac{\partial A}{\partial w} = h$, and hence at the given instant of time we have $\frac{\partial A}{\partial h} = 5$ and $\frac{\partial A}{\partial w} = 3$.

Let t denote the time. We think of h and w as functions of t. The information in the problem says that at the given instant of time, we have $\frac{dh}{dt} = 2$ and $\frac{dw}{dt} = -1$.

By Equation (7) of this section we then see that at the given instant of time, we have

$$\frac{dA}{dt} = \frac{\partial A}{\partial h}\frac{dh}{dt} + \frac{\partial A}{\partial w}\frac{dw}{dt} = 5 \cdot 2 + 3 \cdot (-1) = 7.$$

COMMON MISTAKES

When using the Chain Rule, whether in the matrix formulation or the non-matrix formulation, the final answer should be expressed in the variables in the domain of $\boldsymbol{F} \circ \boldsymbol{G}$. A common mistake is to mix different types of variables. For example, when using formulas such as $\frac{\partial z}{\partial r} = \frac{\partial z}{\partial x}\frac{\partial x}{\partial r} + \frac{\partial z}{\partial y}\frac{\partial y}{\partial r}$ and $\frac{\partial z}{\partial s} = \frac{\partial z}{\partial x}\frac{\partial x}{\partial s} + \frac{\partial z}{\partial y}\frac{\partial y}{\partial s}$, the final answer should be expressed in terms of the variables r and s only, with all instances of x and y expressed in terms of r and s.

EXERCISES

Basic Exercises

1–3 ■ For each pair of functions $\boldsymbol{F}\colon \mathbb{R}^2 \to \mathbb{R}^4$ and $\boldsymbol{G}\colon \mathbb{R}^3 \to \mathbb{R}^2$, find the formula for $(\boldsymbol{F} \circ \boldsymbol{G})(x, y, z)$.

1. $\boldsymbol{F}(x,y) = \begin{bmatrix} 2x-y \\ xy \\ x^2 \\ y^2 \end{bmatrix}$ and $\boldsymbol{G}(x,y,z) = \begin{bmatrix} 5x+3z \\ z-y \end{bmatrix}$.

2. $\boldsymbol{F}(x,y) = \begin{bmatrix} 5xy \\ x-y \\ 4y \\ 6x \end{bmatrix}$ and $\boldsymbol{G}(x,y,z) = \begin{bmatrix} \sin(x+3) \\ \cos(y-z) \end{bmatrix}$.

3. $\boldsymbol{F}(x,y) = \begin{bmatrix} e^y \\ \sin(3x) \\ x+2y \\ xy \end{bmatrix}$ and $\boldsymbol{G}(x,y,z) = \begin{bmatrix} 3xy \\ 5yz \end{bmatrix}$.

4–6 ■ For each pair of functions $\boldsymbol{F}\colon \mathbb{R}^3 \to \mathbb{R}^2$ and $\boldsymbol{G}\colon \mathbb{R}^2 \to \mathbb{R}^3$, find each of $D\boldsymbol{F}(\boldsymbol{G}(x,y))$ and $D\boldsymbol{G}(x,y)$ and $D(\boldsymbol{F} \circ \boldsymbol{G})(x,y)$ directly, and verify the The Chain Rule via Matrix Multiplication for these functions.

4. $\boldsymbol{F}(x,y,z) = \begin{bmatrix} 2x-3y+z \\ xy+yz \end{bmatrix}$ and $\boldsymbol{G}(x,y) = \begin{bmatrix} 5x \\ 8y \\ y-x \end{bmatrix}$.

5. $\boldsymbol{F}(x,y,z) = \begin{bmatrix} xyz \\ x-y+z \end{bmatrix}$ and $\boldsymbol{G}(x,y) = \begin{bmatrix} x+3 \\ y+1 \\ x+y \end{bmatrix}$.

6. $\boldsymbol{F}(x, y, z) = \begin{bmatrix} e^{2x+z} \\ y^2+xz \end{bmatrix}$ and $\boldsymbol{G}(x, y) = \begin{bmatrix} y-x \\ 4x \\ xy \end{bmatrix}$.

7–9 ■ For each set of functions, find $\frac{dz}{dt}$.

7. $z = x^4 + 5y$, and $x = 3t + 1$ and $y = t^2$.

8. $z = 3xy^2$, and $x = \tan t$ and $y = t^3$.

9. $z = \sin(xy)$, and $x = 2t^5$ and $y = t + 2$.

10–12 ■ For each set of functions, find $\frac{\partial z}{\partial s}$ and $\frac{\partial z}{\partial t}$.

10. $z = 3x + y^2$, and $x = s + 2t$ and $y = s - t$.

11. $z = x^2 y$, and $x = s^2 t^3$ and $y = 2t$.

12. $z = \cos(2x + 3y)$, and $x = s^3$ and $y = st$.

13. The height and radius of a cylinder are each changing as a function of time. Suppose that at a certain instant of time, the height is 3 inches and is decreasing at a rate of 5 inches per second, and the radius is 2 inches and is increasing at a rate of 2 inches per second. Find the rate of change of the volume of the cylinder at that instant of time. Is the volume increasing or decreasing at that instant of time?

14. The length, width and height of a box are each changing as a function of time. Suppose that at a certain instant of time, the length is 4 inches and is increasing at a rate of 2 inches per second, the width is 3 inches and is decreasing at a rate of 6 inches per second, and the height is 2 inches and is increasing at a rate of 1 inch per second. Find the rate of change of the volume of the box at that instant of time. Is the volume increasing or decreasing at that instant of time?

Optional: Limit Definition of the Derivative for Multivariable Vector-Valued Functions

In Section 31.2 we saw the definition of the derivative of multivariable vector-valued functions. Specifically, if we have a differentiable function $\boldsymbol{F}: \mathbb{R}^n \to \mathbb{R}^m$, then the derivative of this function is the matrix $D\boldsymbol{F}$, which has the various partial derivatives of the components of $\boldsymbol{F}$ as the entries in the matrix. One the one hand, this definition of the derivative is very convenient to compute, and is very useful in solving a variety of mathematical problems. On the other hand, this definition is not entirely satisfactory from a conceptual point of view, because it is not defined as a limit analogous to the limit definition of the derivative of single-variable real-valued functions. That definition, as we recall, is that if $f: \mathbb{R} \to \mathbb{R}$ is a function, and a is a real number, then the derivative of f at a is defined to be

$$f'(a) = \lim_{h \to 0} \frac{f(a+h) - f(a)}{h}, \tag{1}$$

provided that the limit exists. Recall too that partial derivatives are defined via similar limits.

The question arises, can the derivative of multivariable vector-valued functions also be defined via some sort of limit? Of course, it would be very troubling if such a limit definition existed but it were not equivalent to the definition of the derivative of multivariable vector-valued functions that we have already seen. We will now see that a limit definition of the derivative of multivariable vector-valued functions is indeed possible, and it yields the same matrix we are already familiar with, though we will skip the rigorous details (which are not trivial).

Let $\boldsymbol{F}: \mathbb{R}^n \to \mathbb{R}^m$ be a multivariable vector-valued function, and let p be a point in $\mathbb{R}^n$. If we want to find the derivative of $\boldsymbol{F}$ at p via a limit, the first thing one might attempt is an exact copy of the right-hand side of Equation (1), but with the single-variable real-valued function f replaced by the multivariable vector-valued function $\boldsymbol{F}$, and with the real numbers a and h and 0 replaced by p and k and $\mathit{0}$ in $\mathbb{R}^n$, which are now thought of as vectors. Such a substitution would yield

$$\lim_{k\to\mathit{0}} \frac{\boldsymbol{F}(p+k) - \boldsymbol{F}(p)}{k}.$$

There is nothing wrong with the numerator of the above fraction, because we can add and subtract vectors in $\mathbb{R}^n$, but we have a very serious problem with the denominator, because we cannot divide by vectors in $\mathbb{R}^n$. So, the above limit simply makes no sense.

To solve our problem, we rearrange Equation (1) into a form that is more useful to our purposes. First, we bring $f'(a)$ to the left-hand side of the equation, and bring it into the limit, which is possible because $f'(a)$ does not depend upon h, and we obtain

$$\lim_{h\to 0}\left[\frac{f(a+h) - f(a)}{h} - f'(a)\right] = 0.$$

Next, we take common denominator, yielding

$$\lim_{h\to 0} \frac{f(a+h) - [f(a) + f'(a)h]}{h} = 0.$$

Finally, we observe that whether or not this fraction goes to zero as h goes to zero does not change if we modify it by a plus or minus sign, and hence we deduce that the limit will not change if we introduce an absolute value in the denominator, obtaining

$$\lim_{h\to 0} \frac{f(a+h) - [f(a) + f'(a)h]}{|h|} = 0. \tag{2}$$

Equation (2) is equivalent to Equation (1). Hence, we could use Equation (2) to define the derivative of the function f at the point a. That is, the function f is differentiable at a if there exists a number, denoted $f'(a)$, such that Equation (2) holds. Of course, using Equation (2) instead of Equation (1) is much less convenient and intuitively meaningful if we are interested only in the derivatives of single-variable real-valued functions, but, unlike Equation (1), it turns out that Equation (2) can be generalized to multivariable vector-valued functions, as we now see.

As before, let $\boldsymbol{F}\colon \mathbb{R}^n \to \mathbb{R}^m$ be a multivariable vector-valued function, and let p be a point in $\mathbb{R}^n$. We now want to consider the analog of Equation (2), but, again with the single-variable real-valued function f replaced by the multivariable vector-valued function $\boldsymbol{F}$, with the real numbers a and h and 0 replaced by p and k and $\mathit{0}$ in $\mathbb{R}^n$ (again thought of as vectors), and with the absolute value of h, denoted $|h|$, replace by the length of k, denoted $|k|$. Such a substitution would yield

$$\lim_{k \to 0} \frac{\boldsymbol{F}(p+k) - [\boldsymbol{F}(p) + D\boldsymbol{F}(p)\, k]}{|k|} = 0. \tag{3}$$

In contrast with our previous, and unsuccessful, attempt to formulate the analog of Equation (1) for the function $\boldsymbol{F}$, Equation (3) makes good sense, for the following reasons. First, we note that $D\boldsymbol{F}(p)$ is an $m \times n$ matrix, and that k, which is a column vector, can be thought of an an $n \times 1$ matrix, so that it makes sense to take the product $D\boldsymbol{F}(p)\, k$, which is a column vector in $\mathbb{R}^m$. Second, we note that because $\boldsymbol{F}(p+k)$ and $\boldsymbol{F}(p)$ are vectors in $\mathbb{R}^m$, the addition and subtraction in the numerator of Equation (3) make sense. Finally, we observe that whereas k is a vector, and hence we cannot divide by it, there is no problem dividing by the number $|k|$. Putting these observations together, we see that Equation (3) is a reasonable criterion we could impose upon the function $\boldsymbol{F}$.

Not only does Equation (3) makes sense, but it can be used to define the derivative of $\boldsymbol{F}$ via limits. That is, we say that the function $\boldsymbol{F}$ is **differentiable** at p if there exists an $m \times n$ matrix A such that

$$\lim_{k \to 0} \frac{\boldsymbol{F}(p+k) - [\boldsymbol{F}(p) + A\, k]}{|k|} = 0.$$

The above definition, while nicely analogous to the limit definition of the single-variable real-valued functions, raises a number of questions. Suppose that the function $\boldsymbol{F}$ in the above definition is differentiable at p. Is the matrix A the same as the matrix of partial derivatives $D\boldsymbol{F}(p)$ that was defined in Equation (2) of Section 1? Indeed, if $\boldsymbol{F}$ is differentiable, does that guarantee that all the components of $\boldsymbol{F}$ have all their partial derivatives, which is needed to form the matrix $D\boldsymbol{F}(p)$? And, conversely, if all the components of $\boldsymbol{F}$ have all their partial derivatives, does that guarantee that $\boldsymbol{F}$ is differentiable in the sense of the limit in Equation (3)?

It turns out that things work out reasonably nicely, as we will now state, though without giving proofs. First, if $\boldsymbol{F}$ is differentiable in the sense of the limit in Equation (3), then it is always the case that all the components of $\boldsymbol{F}$ have all their partial derivatives, and, further, the matrix A is precisely the matrix of partial derivatives $D\boldsymbol{F}(p)$ that was defined in Equation (2) of Section 1. For the converse, we need a slight additional hypothesis, namely, that all the components of $\boldsymbol{F}$ have continuous partial derivatives on some neighborhood of p; if that is true, then $\boldsymbol{F}$ is differentiable in the sense of the limit in Equation (3).

We will not go into further theoretical details, but let us look at the relation of the limit definition and the matrix of partial derivatives via an example.

Suppose that $\boldsymbol{F}\colon \mathbb{R}^2 \to \mathbb{R}^3$ is given by the formula $\boldsymbol{F}(x,y) = \left[\begin{smallmatrix} x^2 \\ 3y \\ 5xy \end{smallmatrix}\right]$. Then

$$D\boldsymbol{F}(x,y) = \begin{bmatrix} 2x & 0 \\ 0 & 3 \\ 5y & 5x \end{bmatrix}.$$

Let $p = \left[\begin{smallmatrix} x \\ y \end{smallmatrix}\right]$ and $k = \left[\begin{smallmatrix} s \\ t \end{smallmatrix}\right]$. Then the left-hand side of Equation (3) becomes

$$\begin{aligned}
&\lim_{k\to 0} \frac{\boldsymbol{F}(p+k) - [\boldsymbol{F}(p) + D\boldsymbol{F}(p)\,k]}{|k|} \\
&= \lim_{\left[\begin{smallmatrix} s \\ t \end{smallmatrix}\right]\to\left[\begin{smallmatrix} 0 \\ 0 \end{smallmatrix}\right]} \frac{\begin{bmatrix} (x+s)^2 \\ 3(y+t) \\ 5(x+s)(y+t) \end{bmatrix} - \left(\begin{bmatrix} x^2 \\ 3y \\ 5xy \end{bmatrix} + \begin{bmatrix} 2x & 0 \\ 0 & 3 \\ 5y & 5x \end{bmatrix}\begin{bmatrix} s \\ t \end{bmatrix}\right)}{\sqrt{s^2+t^2}} \\
&= \lim_{\left[\begin{smallmatrix} s \\ t \end{smallmatrix}\right]\to\left[\begin{smallmatrix} 0 \\ 0 \end{smallmatrix}\right]} \frac{\begin{bmatrix} x^2 + 2xs + s^2 - x^2 - 2xs - 0 \\ 3y + 3t - 3y - 0 - 3t \\ 5xy + 5xt + 5ys + 5st - 5xy - 5yt - 5xy \end{bmatrix}}{\sqrt{s^2+t^2}} \\
&= \lim_{\substack{s\to 0 \\ t\to 0}} \begin{bmatrix} \frac{s^2}{\sqrt{s^2+t^2}} \\ 0 \\ \frac{5st}{\sqrt{s^2+t^2}} \end{bmatrix} = \lim_{\substack{s\to 0 \\ t\to 0}} \begin{bmatrix} \frac{s}{\sqrt{1+\frac{t^2}{s^2}}} \\ 0 \\ \frac{5t}{\sqrt{1+\frac{t^2}{s^2}}} \end{bmatrix}.
\end{aligned} \tag{4}$$

To complete this calculation, we observe that no matter what non-zero values of s and t we substitute in the expression $\frac{1}{\sqrt{1+\frac{t^2}{s^2}}}$, the denominator will always be at least 1, and therefore the whole fraction will be no greater than 1. It then follows that

$$\lim_{\substack{s\to 0 \\ t\to 0}} \frac{s}{\sqrt{1+\frac{t^2}{s^2}}} = \lim_{\substack{s\to 0 \\ t\to 0}} s\frac{1}{\sqrt{1+\frac{t^2}{s^2}}} = 0,$$

and similarly

$$\lim_{\substack{s\to 0 \\ t\to 0}} \frac{5t}{\sqrt{1+\frac{t^2}{s^2}}} = \lim_{\substack{s\to 0 \\ t\to 0}} 5t\frac{1}{\sqrt{1+\frac{t^2}{s^2}}} = 0.$$

We deduce that the last expression in Equation (4) equals the zero vector, which is what we wanted to verify.

The above example is not a proof that the limit definition of $D\boldsymbol{F}(p)$ given in Equation (3) is the same as the matrix of partial derivatives given in Equation (2), but the example shows that the equivalence of these definitions, which can be proved, is at least plausible.