Natalia Lazzati

Mathematics for Economics

Note 2: Differential Calculus

Note 2 is based on de la Fuente (2000, Ch. 4) and Simon and Blume (1994, Ch. 14).

This note introduces the concept of differentiability and discusses some of its main implications. We start with real-valued functions of only one argument, and then extend the notion of differentiability to multivalued functions. The key to the extension lies in the interpretation of differentiability in terms of the existence of a "good" linear approximation to a function around a point. We also show that important aspects of the local behavior of "sufficiently differentiable" functions are captured accurately by linear or quadratic approximations. The material has important applications to optimization and comparative statics.

Differentiability of Univariate Real-Valued Functions

Differentiability and Taylor's Expansion

Let f be a univariate real-valued function, $f : \mathbb{R} \to \mathbb{R}$. The concept of differentiability relates to the notion of slope of a function at a point x. Given a second point y in the domain of f, the difference quotient $[f(y) - f(x)] / (y - x)$ gives the slope of a secant to the function through the points $[x, f(x)]$ and $[y, f(y)]$. As we take points y closer and closer to x, the secant becomes a better approximation to the tangent to the graph of f at the point $[x, f(x)]$, and in the limit the two coincide. Thus, we can define the derivative of f at x as the limit

$$f'(x) = \lim_{y \to x} \frac{f(y) - f(x)}{y - x} = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h} \text{ with } h \in \mathbb{R}$$

whenever it exists, and we can interpret it as the slope of the function at this point.

Definition 1. (Derivative of a univariate function) Let $f : \mathbb{R} \to \mathbb{R}$. We say that f is differentiable at a point x if the following limit exists

$$\lim_{h \to 0} \frac{f(x+h) - f(x)}{h} \text{ with } h \in \mathbb{R}.$$

When it does exist, we say the value of the limit is the derivative of f at x, written $f'(x)$, $Df(x)$ or $D_x f(x)$. If f is differentiable at each point in its domain, we say the function is differentiable.

The value of differentiable functions around some initial point can be nicely approximated through Taylor's expansion.

Theorem 2. (Taylor's formula for univariate functions) Let $f : \mathbb{R} \to \mathbb{R}$ be n times differentiable. For all x and $x+h$ in the real line we have

$$f(x+h) = f(x) + \sum_{k=1}^{n-1} \frac{f^{(k)}(x)}{k!} h^k + R_n(h)$$

where $f^{(k)}(x)$ is the kth derivative of f evaluated at x, and the remainder or error term $R_n(h)$ is of the form

$$R_n(h) = \frac{f^{(n)}(x+\lambda h)}{n!} h^n$$

for some $\lambda \in (0,1)$. That is, remainder has the same form as the other terms except that the nth derivative is evaluated at some point between x and $x+h$.

Proof. Put $y = x+h$ and define the function $F(z)$ for z between x and y by

$$F(z) = f(y) - f(z) - \sum_{k=1}^{n-1} \frac{f^{(k)}(z)}{k!} (y-z)^k .$$

Then the theorem says that for some point $x + \lambda(y-x)$ between x and y

$$F(x) = \frac{f^{(n)}[x+\lambda(y-x)]}{n!} (y-x)^n .$$

First, observe that $F(y) = 0$ and that most terms in

$$\begin{aligned} F'(z) &= -f'(z) - \sum_{k=1}^{n-1} \left[\frac{f^{(k)}(z)}{k!} k (y-z)^{k-1} (-1) + \frac{f^{(k+1)}(z)}{k!} (y-z)^k \right] \\ &= -f'(z) - \sum_{k=1}^{n-1} \left[\frac{f^{(k+1)}(z)}{k!} (y-z)^k - \frac{f^{(k)}(z)}{k!} k (y-z)^{k-1} \right] \end{aligned}$$

cancel, leaving us with

$$F'(z) = -\frac{f^{(n)}(z)}{(n-1)!} (y-z)^{n-1} . \tag{1}$$

Next we define the function

$$G(z) = F(z) - \left(\frac{y-z}{y-x} \right)^n F(x)$$

and observe that G is a continuous function on the open interval (x, y), with

$$G(y) = F(y) - 0 = 0 = F(x) - F(x) = G(x)$$

and

$$G'(z) = F'(z) - n\left(\frac{y-z}{y-x}\right)^{n-1}\frac{-1}{y-x}F(x). \tag{2}$$

By Rolle's Theorem (see SB, p. 824), there exists some $\lambda \in (0,1)$ such that

$$G'[x + \lambda(y-x)] = 0.$$

Expanding that expression using (1) and (2), we get

$$\begin{aligned}
0 &= G'[x+\lambda(y-x)] = F'[x+\lambda(y-x)] + n\left(\frac{y-x-\lambda(y-x)}{y-x}\right)^{n-1}\frac{1}{y-x}F(x) \\
\Longrightarrow & \frac{f^{(n)}[x+\lambda(y-x)]}{(n-1)!}[y-x-\lambda(y-x)]^{n-1} = n\left(\frac{(1-\lambda)(y-x)}{y-x}\right)^{n-1}\frac{1}{y-x}F(x) \\
\Longrightarrow & \frac{f^{(n)}[x+\lambda(y-x)]}{(n-1)!}[(1-\lambda)(y-x)]^{n-1} = n(1-\lambda)^{n-1}\frac{1}{y-x}F(x) \\
\Longrightarrow & \frac{f^{(n)}[x+\lambda(y-x)]}{n!}(y-x)^{n} = F(x)
\end{aligned}$$

which is the desired result. ■

Taylor's theorem gives us a formula for constructing a polynomial approximation to a differentiable function. If we let $n = 2$, and omitting the remainder, we get

$$f(x+h) \cong f(x) + f'(x)h. \tag{3}$$

The differentiability of f implies that the error term will be small. Hence the linear function in the right-hand side of (3) is guaranteed to be a decent approximation to f near x. Higher-order approximations that use more derivatives will be even better.

The Chain Rule

Let g and h be two real-valued functions on $\mathbb{R}$, the function formed by first applying function g to any number x and then applying function h to the result $g(x)$ is called the composition of functions g and h and is written as

$$f(x) = h[g(x)] \text{ or } f(x) = (h \circ g)(x).$$

Example 1. The functions which describe a firm's behavior, such as its profit function π, are usually written as functions of a firm's output y. If one wants to study the dependence of a firm's profit on the amount of an input x it uses, one must compose the profit function with the firm's production function $y = f(x)$. The latter function tells us how much output y the firm can obtain from x units of the given input. The result is a function

$$F(x) \equiv \pi [f(x)].$$

For instance, if $\pi(y) = -y^4 + 6y^2 - 5$ and $f(x) = 5x^{2/3}$, then

$$F(x) \equiv \pi [f(x)] = -625x^{8/3} + 150x^{4/3} - 5.$$

Notice that we use different names for π and F as their arguments are different. ▲

The derivative of a composite function is obtained as the derivative of the outside function (evaluated at the inside function) times the derivative of the inside function. This general form is called the Chain Rule.

Theorem 3. (Chain Rule for univariate functions) Let g and h be two real-valued differentiable functions on $\mathbb{R}$, and define $f(x) = h[g(x)]$. Then

$$\frac{df}{dx}(x) = h'[g(x)]\, g'(x).$$

Example 2. Consider the model in Example 1. By using the Chain Rule we get

$$\begin{aligned}
\pi'[f(x)]\, f'(x) &= \left\{-4[f(x)]^3 + 12[f(x)]\right\} \frac{10}{3} x^{-1/3} \\
&= \left[-4\left(5x^{2/3}\right)^3 + 12\left(5x^{2/3}\right)\right] \frac{10}{3} x^{-1/3} \\
&= \left(-500x^2 + 60x^{2/3}\right) \frac{10}{3} x^{-1/3} \\
&= -\frac{5000}{3} x^{5/3} + 200x^{1/3}.
\end{aligned}$$

Note that the latter is equivalent to the derivative of F with respect to x

$$F'(x) = -\frac{5000}{3} x^{5/3} + 200x^{1/3}.$$

This result corroborates the claim in Theorem 3. ▲

Differentiability of Multivariate Real-Valued Functions

In this section we study differentiability of functions from $\mathbb{R}^n$ into $\mathbb{R}$. We build this concept on the results derived for real-valued univariate functions.

Partial and Directional Derivatives

To extend the previous concept of differentiability to functions of n variables we need to specify the direction along which we are approaching x. The problem did not appear before as we can only approach x from either the left or the right, and the derivative of the univariate function f at x is defined as the common value of both one-sided limits whenever they coincide. In multivariate functions however we can approach a point from an infinite number of directions, and therefore we have to specify the one we are considering. This observation leads us to the concept of directional derivative, that we formalize next.

Definition 4. (Directional derivative) The directional derivative of $f : \mathbb{R}^n \to \mathbb{R}$ in the direction of $\mathbf{v}$ at the point $\mathbf{x}$ is defined by

$$\lim_{h \to 0} \frac{f(\mathbf{x} + h\mathbf{v})}{h} \text{ with } h \in \mathbb{R} \text{ and } \|\mathbf{v}\| = 1$$

whenever this limit exists.

Directional derivatives in the direction of the coordinate axes are of special interest. The partial derivative of f with respect to its ith argument is defined as

$$\lim_{h \to 0} \frac{f(\mathbf{x} + h\mathbf{e}_i)}{h} \text{ with } h \in \mathbb{R}$$

where $\mathbf{e}_i$ is a vector whose components are all zero except for the ith one, which is 1.

Definition 5. (Partial derivative) Let f be a multivariate function $f : \mathbb{R}^n \to \mathbb{R}$. The partial derivative of f with respect to its ith argument, x_i, at a point $\mathbf{x}$, is the limit

$$\frac{\partial f(\mathbf{x})}{\partial x_i} = \lim_{h \to 0} \frac{f(x_1, ..., x_i + h, ..., x_n) - f(x_1, ..., x_i, ..., x_n)}{h} \text{ with } h \in \mathbb{R}$$

whenever it exists. [Other usual ways to write $\partial f(\mathbf{x}) / \partial x_i$ are $D_{x_i} f(\mathbf{x})$ and $f_{x_i}(\mathbf{x})$.]

Example 3. Let us consider the function $f(x_1, x_2) = x_1^2 x_2$. Its partial derivatives are

$$\frac{\partial f_1}{\partial x_1}(x_1, x_2) = 2x_1 x_2 \text{ and } \frac{\partial f_1}{\partial x_2}(x_1, x_2) = x_1^2.$$

[**Calculate the partial derivatives for the function** $g(x_1, x_2) = x_1 x_2$.] ▲

Let $f : \mathbb{R}^n \to \mathbb{R}$, then each one of its partial derivatives $\partial f(\mathbf{x}) / \partial x_i$ is also a real-valued function of n variables and the partials of $\partial f(\mathbf{x}) / \partial x_i$ can be defined as before. The partials of $\partial f(\mathbf{x}) / \partial x_i$, for $i = 1, ..., n$, are the second-order partial derivatives of f.

Differentiability and the Taylor's Expansion

The main objective is to define a concept of differentiability for multivariate functions $f : \mathbb{R}^n \to \mathbb{R}$. We will define differentiability in terms of the possibility of approximating the local behavior of f through a linear function. After doing this, we will relate the concept of differentiability to the partial derivatives of f, that we defined in the last sub-section.

Let us think for a moment in the definition of derivative for a function $f : \mathbb{R} \to \mathbb{R}$. To say f is differentiable at x, is the same as to say that there exists a real number a_x, that we define as $f'(x)$, such that

$$\lim_{h \to 0} \frac{f(x+h) - [f(x) + a_x h]}{h} = 0. \tag{4}$$

To interpret this result let us assume that we want to approximate the value of $f(x+h)$ by a linear function. One possibility is to use $f(x) + f'(x)h$. Expression (4) guarantees that the approximation will be good whenever h is small. [If we define $E(h) = f(x+h) - [f(x) + a_x h]$ as the error of the linear approximation, then condition (4) ensures $\lim_{h \to 0} E(h)/h = 0$.]

There is no difficulty in extending this notion of differentiability to mappings from $\mathbb{R}^n$ to $\mathbb{R}$. Before giving a formal definition, we want to emphasize the importance of differentiability for our purposes:

Because differentiable functions admit good linear approximations, so do differentiable models. This gives us a tractable way to analyze them. When we use calculus to study a nonlinear model, we are in fact constructing a linear approximation to it in some neighborhood of a point of interest. The assumption that the behavioral functions in the model are differentiable means that the approximation error is small.

The obvious limitation is that it generally yields only local results, valid only in some small neighborhood of the initial solution.

The formal definition of differentiability is as follows.

Definition 6. (Differentiability of multivariate functions) A function $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable at $\mathbf{x}$ if there exists a vector $\mathbf{a_x}$ in $\mathbb{R}^n$ such that

$$\lim_{\|\mathbf{h}\|\to 0} \frac{\left| f(\mathbf{x}+\mathbf{h}) - \left[f(\mathbf{x}) + \mathbf{a}_{\mathbf{x}}^T \mathbf{h} \right] \right|}{\|\mathbf{h}\|} = 0 \tag{5}$$

where $\mathbf{h} \in \mathbb{R}^n$ and $\|\cdot\|$ is the Euclidean norm of a vector, $\|\mathbf{x}\| = \sqrt{\sum_{i=1}^{n} (x_i)^2}$. If f is differentiable at every point in its domain, we say that f is differentiable.

If f is differentiable, then we can define its derivative as the function

$$D_{\mathbf{x}} f : \mathbb{R}^n \to \mathbf{a_x}$$

that goes from $\mathbb{R}^n$ to an n-dimensional vector $\mathbf{a_x}$. When it is clear from the context that we are differentiating with respect to the vector $\mathbf{x}$ we will write Df instead of $D_{\mathbf{x}} f$.

It is apparent now that we can, as in the case of a univariate real-valued function, interpret the differential in terms of a linear approximation to $f(\mathbf{x}+\mathbf{h})$. That is, we can consider $f(\mathbf{x}) + \mathbf{a}_{\mathbf{x}}^T \mathbf{h}$ as a linear approximation of $f(\mathbf{x}+\mathbf{h})$. Expression (5) guarantees that the approximation will be good whenever $\mathbf{h}$ is small. [If we define $E(\mathbf{h}) = f(\mathbf{x}+\mathbf{h}) - \left[f(\mathbf{x}) + \mathbf{a}_{\mathbf{x}}^T \mathbf{h} \right]$ as the error of the linear approximation, then condition (4) ensures $\lim_{\|\mathbf{h}\|\to 0} |E(\mathbf{h})| / \|\mathbf{h}\| = 0$.]

The derivative of f at $\mathbf{x}$, $Df(\mathbf{x})$, relates to the partial derivatives of f at $\mathbf{x}$ in a natural way. If $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable at $\mathbf{x}$ then the derivative of f at $\mathbf{x}$ is the vector of partial derivatives

$$Df(\mathbf{x}) = D_{\mathbf{x}} f(\mathbf{x}) = \left[\begin{array}{cccc} \frac{\partial f}{\partial x_1}(\mathbf{x}) & \frac{\partial f}{\partial x_2}(\mathbf{x}) & \ldots & \frac{\partial f}{\partial x_n}(\mathbf{x}) \end{array} \right]$$

that we often call Jacobian of f (evaluated) at $\mathbf{x}$.

Moreover, the differentiability of f is guaranteed if its partial derivatives exist and are smooth. If the n partial derivatives of a function $f : \mathbb{R}^n \to \mathbb{R}$ exist and are themselves continuous functions of $(x_1, ..., x_n)$, we say that f is continuously differentiable or C^1. An important result in calculus states that if f is C^1, then f is differentiable [see de la Fuente (2000), pp. 172-175].

Example 4. Consider the function $f(x_1, x_2) = x_1^2 x_2$ in Example 3. In this case,

$$Df(x_1, x_2) = \begin{pmatrix} 2x_1x_2 & x_1^2 \end{pmatrix}.$$

[**Calculate the same concept for the function** $g(x_1, x_2) = x_1x_2$.] ▲

The partials of $\partial f(\mathbf{x})/\partial x_i$, for $i = 1, ..., n$, are the second-order partial derivative of f that we identify by

$$D^2 f(\mathbf{x}) = D^2_{\mathbf{x}} f(\mathbf{x}) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2}(\mathbf{x}) & \frac{\partial^2 f}{\partial x_2 \partial x_1}(\mathbf{x}) & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_1}(\mathbf{x}) \\ \frac{\partial^2 f}{\partial x_1 \partial x_2}(\mathbf{x}) & \frac{\partial^2 f}{\partial x_2^2}(\mathbf{x}) & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_2}(\mathbf{x}) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_1 \partial x_n}(\mathbf{x}) & \frac{\partial^2 f}{\partial x_2 \partial x_n}(\mathbf{x}) & \cdots & \frac{\partial^2 f}{\partial x_n^2}(\mathbf{x}) \end{pmatrix}.$$

This is an $n \times n$ matrix, often called Hessian matrix of f (evaluated) at $\mathbf{x}$. If all these n^2 partial derivatives exist and are themselves continuous functions of $(x_1, ..., x_n)$, we say that f is twice continuously differentiable or C^2. Young's theorem states that if f is twice continuously differentiable, then $D^2 f(\mathbf{x})$ is a symmetric matrix.

Theorem 7. **(Young's theorem)** Assume a function $f : \mathbb{R}^n \to \mathbb{R}$ is C^2 on an open subset $U \subseteq \mathbb{R}^n$. Then, for all $\mathbf{x} \in U$ and all $i, j = 1, ..., n$,

$$\frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{x}) = \frac{\partial^2 f}{\partial x_j \partial x_i}(\mathbf{x}).$$

[**Check Young's theorem with the function** $f_1(x_1, x_2) = x_1^2 x_2$ **in Example 3.**]

Taylor's formula can be generalized to the case of a real-valued function of n variables. Because the notation gets messy, and we will only use the simplest case, we will state the following theorem for the case of a first order approximation with a quadratic form remainder.

Theorem 8. **(Taylor's formula for multivariate functions)** Let $f : \mathbb{R}^n \to \mathbb{R}$ be a C^2 function. If $\mathbf{x}$ and $\mathbf{x} + \mathbf{h}$ are in $\mathbb{R}^n$, then

$$f(\mathbf{x} + \mathbf{h}) = f(\mathbf{x}) + Df(\mathbf{x})\mathbf{h} + (1/2)\mathbf{h}^T D^2 f(\mathbf{x} + \lambda \mathbf{h})\mathbf{h}$$

for some $\lambda \in (0, 1)$.

We will use Theorem 8 to prove the sufficient conditions for maxima in optimization problems.

The Chain Rule

In many cases we are interested in the derivatives of composite functions. The following result says that the composition of differentiable functions is differentiable, and its derivative is the product of the derivatives of the original functions.

Theorem 9. (Chain rule for multivariate functions) Let g and h be two functions with

$$g : \mathbb{R}^n \rightarrow \mathbb{R} \text{ and } h : \mathbb{R} \rightarrow \mathbb{R}$$

and define $f(\mathbf{x}) = h[g(\mathbf{x})]$ or $f(\mathbf{x}) = (h \circ g)(\mathbf{x})$ with $f : \mathbb{R}^n \rightarrow \mathbb{R}$. If g and h are differentiable, then $f = h \circ g$ is differentiable and

$$Df(\mathbf{x}) = Dh[g(\mathbf{x})]\, Dg(\mathbf{x}).$$

Proof. See de la Fuente (2000), pp. 176-178. (They provide a proof for a result that is more general than the statement in Theorem 9.) ■

The next example sheds light on the implementation of the last result.

Example 5. Let $f(x_1, x_2) = (x_1 x_2)^2$. We want to use the Chain Rule to find $\partial f / \partial x_1$ and $\partial f / \partial x_2$. To this end, let us define

$$g(x_1, x_2) = x_1 x_2 \text{ and } h(y) = y^2.$$

Notice that $f(x_1, x_2) = (h \circ g)(x_1, x_2)$. By the Chain Rule we have that

$$\begin{aligned} Df(x_1, x_2) &= Dh[g(x_1, x_2)]\, Dg(x_1, x_2) \\ &= 2g(x_1, x_2) \begin{pmatrix} x_2 & x_1 \end{pmatrix} \\ &= 2x_1 x_2 \begin{pmatrix} x_2 & x_1 \end{pmatrix} \\ &= \begin{pmatrix} 2x_1 x_2^2 & 2x_1^2 x_2 \end{pmatrix}. \end{aligned}$$

[**Check the result directly, by differentiating** $f(x_1, x_2)$ **with respect to** x_1 **and** x_2.] ▲

Differentiability of Functions from $\mathbb{R}^n$ into $\mathbb{R}^m$

We now turn to the general case where $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^m$ is a function of n variables whose value is a vector of m elements. As we learned in Note 1, we can think of the mapping $\mathbf{f}$ as a vector of component functions f_i, each of which is a real-valued function of n variables

$$\mathbf{f} = (f_1, f_2, ..., f_m)^T \text{ with } f_i : \mathbb{R}^n \to \mathbb{R} \text{ for } i = 1, ..., m.$$

Thinking in this way, the extension is trivial (although the notation becomes messier!).

Definition 10. (Differentiability of functions of $\mathbb{R}^n$ to $\mathbb{R}^m$) A function $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^m$ is differentiable at $\mathbf{x}$ if there exists a matrix $A_\mathbf{x}$ such that

$$\lim_{\|\mathbf{h}\| \to 0} \frac{\|\mathbf{f}(\mathbf{x}+\mathbf{h}) - [\mathbf{f}(\mathbf{x}) + A_\mathbf{x}\mathbf{h}]\|}{\|\mathbf{h}\|} = 0 \tag{6}$$

where $\mathbf{h} \in \mathbb{R}^n$ and $\|\cdot\|$ is the Euclidean norm of a vector, $\|\mathbf{x}\| = \sqrt{\sum_{i=1}^n (x_i)^2}$. If $\mathbf{f}$ is differentiable at every point in its domain, we say that $\mathbf{f}$ is differentiable.

If $\mathbf{f}$ is differentiable, then we can define its derivative as the function

$$D\mathbf{f} = D_\mathbf{x}\mathbf{f} : \mathbb{R}^n \to A_\mathbf{x}$$

that goes from $\mathbb{R}^n$ to an $m \times n$ matrix $A_\mathbf{x}$.

Here again, the derivative of $\mathbf{f}$ at $\mathbf{x}$, $D\mathbf{f}(\mathbf{x})$, relates to the partial derivatives of $\mathbf{f}$ at $\mathbf{x}$. If $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^m$ is differentiable at $\mathbf{x}$, then the derivative of $\mathbf{f}$ at $\mathbf{x}$ is the matrix of partial derivatives

$$D\mathbf{f}(\mathbf{x}) = D_\mathbf{x}\mathbf{f}(\mathbf{x}) = \begin{pmatrix} Df_1(\mathbf{x}) \\ Df_2(\mathbf{x}) \\ \vdots \\ Df_m(\mathbf{x}) \end{pmatrix} = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(\mathbf{x}) & \frac{\partial f_1}{\partial x_2}(\mathbf{x}) & ... & \frac{\partial f_1}{\partial x_n}(\mathbf{x}) \\ \frac{\partial f_2}{\partial x_1}(\mathbf{x}) & \frac{\partial f_2}{\partial x_2}(\mathbf{x}) & ... & \frac{\partial f_2}{\partial x_n}(\mathbf{x}) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1}(\mathbf{x}) & \frac{\partial f_m}{\partial x_2}(\mathbf{x}) & ... & \frac{\partial f_m}{\partial x_n}(\mathbf{x}) \end{pmatrix}.$$

that we often call Jacobian matrix of $\mathbf{f}$ (evaluated) at $\mathbf{x}$. If the partial derivatives of the component functions $f_1, f_2, ..., f_m$ exist and are continuous, then $\mathbf{f}$ is differentiable [see de la Fuente (2000), pp. 172-175].

Example 6. Consider the functions $f_1(x_1, x_2) = x_1^2 x_2$ and $f_2(x_1, x_2) = \ln(x_1 + x_2)$. Its partial derivatives are

$$\frac{\partial f_1}{\partial x_1}(x_1, x_2) = 2x_1 x_2,\ \frac{\partial f_1}{\partial x_2}(x_1, x_2) = x_1^2,\ \frac{\partial f_2}{\partial x_1}(x_1, x_2) = \frac{1}{x_1 + x_2} \text{ and } \frac{\partial f_2}{\partial x_2}(x_1, x_2) = \frac{1}{x_1 + x_2}.$$

Therefore,

$$Df_1(x_1, x_2) = \begin{pmatrix} 2x_1 x_2 & x_1^2 \end{pmatrix} \text{ and } Df_2(x_1, x_2) = \begin{pmatrix} \frac{1}{x_1+x_2} & \frac{1}{x_1+x_2} \end{pmatrix}.$$

If we let $\mathbf{f} = (f_1, f_2)^T$, then

$$D\mathbf{f}(\mathbf{x}) \begin{pmatrix} 2x_1 x_2 & x_1^2 \\ \frac{1}{x_1+x_2} & \frac{1}{x_1+x_2} \end{pmatrix}.$$

[**Calculate the same concepts for the functions** $g_1(x_1, x_2) = x_1 x_2$ **and** $g_2(x_1, x_2) = \ln(2x_1 + 3x_2)$.] ▲

The following result states again that the composition of differentiable functions is differentiable, and its derivative is the product of the derivatives of the original functions.

Theorem 9. (Chain rule for multivariate functions) Let $\mathbf{g}$ and $\mathbf{h}$ be two functions with

$$\mathbf{g} : \mathbb{R}^n \to \mathbb{R}^m \text{ and } \mathbf{h} : \mathbb{R}^m \to \mathbb{R}^p$$

and define $\mathbf{f}(\mathbf{x}) = \mathbf{h}[\mathbf{g}(\mathbf{x})]$ or $\mathbf{f}(\mathbf{x}) = (\mathbf{h} \circ \mathbf{g})(\mathbf{x})$ with $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^p$. If $\mathbf{g}$ and $\mathbf{h}$ are differentiable, then $\mathbf{f} = \mathbf{h} \circ \mathbf{g}$ is differentiable and

$$D\mathbf{f}(\mathbf{x}) = D\mathbf{h}[\mathbf{g}(\mathbf{x})]\, D\mathbf{g}(\mathbf{x}).$$

Proof. See de la Fuente (2000), pp. 176-178. ■

The last example applies the Chain Rule to a case where $n = m = 3$ and $p = 1$.

Example 7. Let $l = xy^2 z$, with

$$x = r + t,\ y = s \text{ and } z = s + t.$$

We want to use the Chain Rule to address $\partial l / \partial r$, $\partial l / \partial s$ and $\partial l / \partial t$.

Let us define $\mathbf{g} = (g_1, g_2, g_3)^T$, with $g_i : \mathbb{R}^3 \to \mathbb{R}$ for $i = 1, 2, 3$, as follows

$$\begin{aligned} g_1(r,s,t) &= r+t \\ g_2(r,s,t) &= s \\ g_3(r,s,t) &= s+t. \end{aligned}$$

Then $\mathbf{g} : \mathbb{R}^3 \to \mathbb{R}^3$. In addition, let $h : \mathbb{R}^3 \to \mathbb{R}$ be defined as $h(x,y,z) = xy^2z$. Then $l = f(r,s,t) = h[\mathbf{g}(r,s,t)]$. Notice that

$$\begin{aligned} D\mathbf{g}(r,s,t) &= \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix}, \mathbf{g}(r,s,t) = \begin{pmatrix} r+t & s & s+t \end{pmatrix} \\ Dh(x,y,z) &= \begin{pmatrix} y^2z & 2xyz & xy^2 \end{pmatrix}. \end{aligned}$$

It follows that

$$Dh[\mathbf{g}(r,s,t)] = \begin{bmatrix} s^2(s+t) & 2(r+t)s(s+t) & (r+t)s^2 \end{bmatrix}.$$

By the Chain Rule

$$\begin{aligned} Df(r,s,t) &= Dh[\mathbf{g}(r,s,t)]\,D\mathbf{g}(r,s,t) \\ &= \begin{bmatrix} s^2(s+t) & 2(r+t)s(s+t) & (r+t)s^2 \end{bmatrix} \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix} \\ &= \begin{bmatrix} s^2(s+t) & s(r+t)(3s+2t) & s^2(r+s+2t) \end{bmatrix}. \end{aligned}$$

[**Check the result directly, by obtaining an expression for** $l = f(r,s,t)$ **in terms of** r, s **and** t **and taking the partial derivatives.**] ▲