

Chapter 9

Differentiation in Several Variables

The most powerful method available for studying a function in several variables is to approximate it locally, near a given point, by an affine function. When this can be done, it provides a wealth of information about the original function. Affine approximation leads to the definition of the *differential* of a function of several variables. The differential of a function F , when it exists, is a matrix of partial derivatives of coordinate functions of F . For this reason, we precede the discussion of the differential with a brief review of partial derivatives.

9.1 Partial Derivatives

In this section, f will be a real valued function defined on an open set in \mathbb{R}^p .

Definition 9.1.1. The partial derivative of f with respect to its j th variable at $x = (x_1, \dots, x_j, \dots, x_p)$ is denoted $\frac{\partial f}{\partial x_j}(x)$ and is defined by

$$\frac{\partial f}{\partial x_j}(x) = \frac{d}{dt} f(x_1, \dots, x_{j-1}, t, x_{j+1}, \dots, x_p) \Big|_{t=x_j},$$

provided this derivative exists.

Thus, the partial derivative of a function f , with respect to its j th variable, at a point x in its domain is obtained by fixing all of the variables of f , except the j th one, at the appropriate values $x_1, \dots, x_{j-1}, x_{j+1}, \dots, x_p$, then differentiating with respect to the remaining variable and evaluating at x_j .

Remark 9.1.2. When it is not necessary to explicitly exhibit the point x at which the partial derivative is being computed (because it is understood from the context or because x is a generic point of the domain of f) we will simply write $\frac{\partial f}{\partial x_j}$ for the partial derivative of x with respect to its j th variable.

Two other notations that are often used for the partial derivative of f with respect to x_j are f_{x_j} and f_j . We won't use these in this text.

Example 9.1.3. Find the partial derivatives of the function

$$f(x_1, x_2, x_3, x_4) = x_1^2 + x_1x_3 - 4x_2^2x_4^3.$$

Solution: To find $\frac{\partial f}{\partial x_1}$, we consider x_2, x_3, x_4 to be fixed constants and we differentiate with respect to the remaining variable and evaluate at x_1 . The result is

$$\frac{\partial f}{\partial x_1} = 2x_1 + x_3.$$

Similarly, we have

$$\begin{aligned}\frac{\partial f}{\partial x_2} &= -8x_2x_4^3, \\ \frac{\partial f}{\partial x_3} &= x_1, \\ \frac{\partial f}{\partial x_4} &= -12x_2^2x_4^2.\end{aligned}$$

Example 9.1.4. Find the partial derivatives of the function

$$f(x, y, z) = z^2 \cos xy.$$

Solution: We have

$$\begin{aligned}\frac{\partial f}{\partial x} &= -yz^2 \sin xy, \\ \frac{\partial f}{\partial y} &= -xz^2 \sin xy, \\ \frac{\partial f}{\partial z} &= 2z \cos xy.\end{aligned}$$

The Partial Derivatives as Limits

If we use the definition of the derivative of a function of one variable as the limit of a difference quotient, the result is

$$\frac{\partial f}{\partial x_j}(x_1, \dots, x_p) = \lim_{h \rightarrow 0} \frac{f(x_1, \dots, x_j + h, \dots, x_p) - f(x_1, \dots, x_j, \dots, x_p)}{h}.$$

The notation involved in this statement becomes much simpler if we note that the point $(x_1, \dots, x_j + h, \dots, x_p)$ may be written as $x + h e_j$, where e_j is the basis vector with 1 in the j th entry and 0 elsewhere. Then,

$$\frac{\partial f}{\partial x_j}(x) = \lim_{h \rightarrow 0} \frac{f(x + h e_j) - f(x)}{h}. \quad (9.1.1)$$

Higher Order Partial Derivatives

The partial derivatives defined so far are *first order* partial derivatives. We define second order partial derivatives of f in the following fashion: for $i, j = 1, \dots, p$ we set

$$\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial}{\partial x_i} \left(\frac{\partial f}{\partial x_j} \right). \quad (9.1.2)$$

The meaning of this is as follows: If the partial derivative $\frac{\partial f}{\partial x_j}$ exists in a neighborhood of a point $x \in \mathbb{R}^p$, then we may attempt to take the partial derivative with respect to x_i of the resulting function at the point x . The result, if it exists, is the right side of the above equation. The expression on the left is the notation that is commonly used for this second order partial derivative. In the case where $i = j$, we modify this notation slightly and write

$$\frac{\partial^2 f}{\partial x_j^2} = \frac{\partial}{\partial x_j} \left(\frac{\partial f}{\partial x_j} \right).$$

A useful way to think of this process is as follows: the expression $\frac{\partial}{\partial x_j}$ is an *operator* – that is, a transformation which takes a function f on an open set U to another function $\frac{\partial f}{\partial x_j}$ on U (provided this derivative exists on U). In fact, this operator is a linear operator (preserves sums and scalar products) because the derivative of a sum is the sum of the derivatives and the derivative of a constant times a function is the constant times the derivative of the function. Such operators may be *composed* – that is, we may first apply one such operator, $\frac{\partial}{\partial x_j}$, to a function and then apply another, $\frac{\partial}{\partial x_i}$, to the result. In fact, we may continue to compose such operators, applying one after another, as long as the resulting function has the appropriate partial derivatives on the given open set. From this point of view, the second order partial derivative of (9.1.2) is just the result of applying to f the second order differential operator

$$\frac{\partial^2}{\partial x_i \partial x_j} = \frac{\partial}{\partial x_i} \circ \frac{\partial}{\partial x_j}.$$

We may, of course, define higher order partial differential operators in an analogous fashion. Given integers j_1, j_2, \dots, j_m between 1 and p , we set

$$\frac{\partial^m}{\partial x_{j_1} \cdots \partial x_{j_m}} = \frac{\partial}{\partial x_{j_1}} \circ \frac{\partial}{\partial x_{j_2}} \circ \cdots \circ \frac{\partial}{\partial x_{j_m}}.$$

The resulting operator is a partial differential operator of *total degree* m .

Example 9.1.5. Find $\frac{\partial^5 f}{\partial x \partial y \partial z \partial y \partial x}$ if $f(x, y, z) = x^2 y^3 z^4 + x^2 + y^4 + xyz$.

Solution: We proceed one derivative at a time:

$$\text{apply } \frac{\partial}{\partial x} : \quad \frac{\partial f}{\partial x} = 2xy^3z^4 + 2x + yz,$$

$$\text{apply } \frac{\partial}{\partial y} : \quad \frac{\partial^2 f}{\partial y \partial x} = 6xy^2z^4 + z,$$

$$\text{apply } \frac{\partial}{\partial z} : \quad \frac{\partial^3 f}{\partial z \partial y \partial x} = 24xy^2z^3 + 1,$$

$$\text{apply } \frac{\partial}{\partial y} : \quad \frac{\partial^4 f}{\partial y \partial z \partial y \partial x} = 48xyz^3,$$

$$\text{apply } \frac{\partial}{\partial x} : \quad \frac{\partial^5 f}{\partial x \partial y \partial z \partial y \partial x} = 48yz^3.$$

Equality of Mixed Partial

It is natural to ask whether or not, in a mixed higher order partial derivative, the order in which the derivatives are taken makes a difference. Some additional calculation using the previous example (Exercise 9.1.5) shows that, at least for the function f of that example, the order in which the five partial derivative operators are applied makes no difference. This is not always the case, but it is the case under rather mild continuity assumptions. When it is the case, we may change the order in which the partial derivatives are taken so as to collect partial derivatives with respect to the same variable together. For example, the 5th order mixed partial derivative of the previous example can be re-written as

$$\frac{\partial^5 f}{\partial x \partial x \partial y \partial y \partial z} = \frac{\partial^5 f}{\partial x^2 \partial y^2 \partial z}.$$

The next theorem tells us when interchanging the order of a mixed partial derivative is legitimate.

Theorem 9.1.6. *Suppose f is a function defined on an open disc $B_r(a, b) \subset \mathbb{R}^2$. Also suppose that both first order partial derivatives exist in $B_r(a, b)$ and that $\frac{\partial^2 f}{\partial y \partial x}$ exists in $B_r(a, b)$ and is continuous at (a, b) . Then $\frac{\partial^2 f}{\partial x \partial y}$ exists at (a, b) and is equal to $\frac{\partial^2 f}{\partial y \partial x}(a, b)$.*

Proof. We introduce a function $\lambda(h, k)$, defined for (h, k) in the disc $B = B_r(0, 0)$, by

$$\lambda(h, k) = f(a + h, b + k) - f(a + h, b) - f(a, b + k) + f(a, b).$$

It follows from the hypotheses of the theorem that the partial derivative of $\lambda(h, k)$ with respect to h exists for all (h, k) in the disc B . If $(h, k) \in B$, the rectangle with vertices $(0, 0)$, $(0, k)$, $(h, 0)$ and (h, k) is also contained in this disc and so the partial derivative of λ with respect to its first variable exists on an open set containing this rectangle.

Now for fixed k ,

$$\lambda(h, k) = g(h) - g(0) \quad \text{where} \quad g(u) = f(a + u, b + k) - f(a + u, b).$$

The function g is differentiable on an open interval containing $[0, h]$, and so we may apply the Mean Value Theorem to g to conclude there is a number $s \in (0, h)$ such that $g(h) - g(0) = hg'(s)$. This means

$$\lambda(h, k) = h \left(\frac{\partial f}{\partial x}(a + s, b + k) - \frac{\partial f}{\partial x}(a + s, b) \right). \quad (9.1.3)$$

Of course, the number s depends on h and k .

Since $\frac{\partial^2 f}{\partial y \partial x}$ exists on B , $\frac{\partial f}{\partial x}$ is a differentiable function of its second variable on B . Hence, we may apply the Mean Value Theorem to this function as well. We conclude that there is a point $t \in (0, k)$ such that

$$\frac{\partial f}{\partial x}(a + s, b + k) - \frac{\partial f}{\partial x}(a + s, b) = k \frac{\partial^2 f}{\partial y \partial x}(a + s, b + t). \quad (9.1.4)$$

Combining (9.1.3) and (9.1.4) yields

$$\frac{1}{hk} \lambda(h, k) = \frac{\partial^2 f}{\partial y \partial x}(a + s, b + t).$$

By hypothesis, the second order partial derivative on the right is continuous at (a, b) . This implies that

$$\lim_{(h,k) \rightarrow (0,0)} \frac{\lambda(h, k)}{hk} = \frac{\partial^2 f}{\partial y \partial x}(a, b).$$

This conclusion uses the fact that the point $(a + s, b + t)$, wherever it is, is at least closer to (a, b) than the point $(a + h, b + k)$.

We complete the proof by noting that the above limit exists independently of how (h, k) approaches $(0, 0)$. In particular, the result will be the same if we first let k approach 0 and then h . However,

$$\begin{aligned} & \lim_{h \rightarrow 0} \lim_{k \rightarrow 0} \frac{1}{hk} \lambda(h, k) \\ &= \lim_{h \rightarrow 0} \lim_{k \rightarrow 0} \frac{1}{h} \left(\frac{f(a + h, b + k) - f(a + h, b)}{k} - \frac{f(a, b + k) - f(a, b)}{k} \right) \\ &= \lim_{h \rightarrow 0} \frac{1}{h} \left(\lim_{k \rightarrow 0} \frac{f(a + h, b + k) - f(a + h, b)}{k} - \lim_{k \rightarrow 0} \frac{f(a, b + k) - f(a, b)}{k} \right) \\ &= \lim_{h \rightarrow 0} \frac{1}{h} \left(\frac{\partial f}{\partial y}(a + h, b) - \frac{\partial f}{\partial y}(a, b) \right) \\ &= \frac{\partial^2 f}{\partial x \partial y}(a, b). \end{aligned}$$

Hence, this second order partial derivative also exists and it equals $\frac{\partial^2 f}{\partial y \partial x}(a, b)$. Note that distributing the limit with respect to k across the difference in the second step above requires that we know the two limits involved exist. This follows from the assumption that $\frac{\partial f}{\partial y}$ exists in $B_r(a, b)$. \square

Obviously, the same result holds, with the same proof, if x and y are reversed in the statement of the above theorem. That is, if we assume either one of the second order mixed partials exists in a neighborhood of (a, b) and is continuous at (a, b) , then the other one also exists at (a, b) and the two are equal at (a, b) .

The following example shows that the continuity of the mixed partial that is assumed to exist is a necessary assumption in the above theorem.

Example 9.1.7. For the function

$$f(x, y) = \begin{cases} \frac{x^3 y - xy^3}{x^2 + y^2} & \text{if } (x, y) \neq (0, 0), \\ 0 & \text{if } (x, y) = (0, 0) \end{cases},$$

show that the first order partial derivatives exist and are continuous everywhere. Then show that the mixed second order partial derivatives $\frac{\partial^2 f}{\partial x \partial y}$ and $\frac{\partial^2 f}{\partial y \partial x}$ exist everywhere, but they are not equal at $(0, 0)$. Why doesn't this contradict the above theorem?

Solution: Except at the point $(0, 0)$ where the denominator vanishes, we may use the standard rules of differentiation to show that

$$\begin{aligned} \frac{\partial f}{\partial x} &= \frac{(3x^2 y - y^3)(x^2 + y^2) - 2x(x^3 y - xy^3)}{(x^2 + y^2)^2}, \\ \frac{\partial f}{\partial y} &= \frac{(x^3 - 3xy^2)(x^2 + y^2) - 2y(x^3 y - xy^3)}{(x^2 + y^2)^2}. \end{aligned} \tag{9.1.5}$$

These expressions may be differentiated again to show that each of the second order partial derivatives also exists, except possibly at $(0, 0)$.

In order to calculate $\frac{\partial f}{\partial x}(0, 0)$ we set $y = 0$ in the expression for f . The resulting function of x is identically 0 and, hence, has derivative 0 with respect to x . Similar reasoning leads to the same conclusion for $\frac{\partial f}{\partial y}(0, 0)$. Since both the expressions in (9.1.5) have limit 0 as $(x, y) \rightarrow (0, 0)$, the first order partial derivatives are continuous everywhere, including at $(0, 0)$, where they both have the value 0.

To calculate $\frac{\partial^2 f}{\partial x \partial y}$, we first note that $\frac{\partial f}{\partial y}(x, 0) = x$, for all x . Hence, $\frac{\partial^2 f}{\partial x \partial y}(0, 0) = 1$.

On the other hand, $\frac{\partial f}{\partial x}(0, y) = -y$, and so $\frac{\partial^2 f}{\partial y \partial x}(0, 0) = -1$.

The two mixed partials are not equal at $(0, 0)$ even though they both exist everywhere. Why doesn't this contradict the previous theorem? It must be the case that neither of these mixed partial derivatives is continuous at $(0, 0)$ – a fact that will be verified in the exercises.

An important hypothesis in many theorems is that a function f belongs to the class $\mathcal{C}^k(U)$ defined below.

Definition 9.1.8. If U is an open subset of \mathbb{R}^p then a function $F : U \rightarrow \mathbb{R}^q$ is said to be \mathcal{C}^k on U if, for each coordinate function f_j of F , all partial derivatives of f_j of total order less than or equal to k exist and are continuous on U .

Functions which are \mathcal{C}^1 on U will be called *smooth* functions on U .

By using Theorem 9.1.6 to interchange pairs of adjacent first order partial differential operators, the following theorem may be proved:

Theorem 9.1.9. *If a real valued function f is \mathcal{C}^k on $U \subset \mathbb{R}^p$ and $m \leq k$, then the m th order partial derivative $\frac{\partial^m f}{\partial x_{j_1} \cdots \partial x_{j_m}}$ is independent of the order in which the first order partial derivatives $\frac{\partial}{\partial j_i}$ are applied.*

Exercise Set 9.1

1. If $f(x, y) = \sqrt{x^2 + y^2}$, find $\frac{\partial f}{\partial x}$ and $\frac{\partial f}{\partial y}$. Are there any points in the plane where they don't exist?
2. If $f(x, y) = xy^2 + xy + y^3$, find all first and second order partial derivatives of f .
3. If $f(x, y) = x \cos y$, find $\frac{\partial f}{\partial x}$, $\frac{\partial f}{\partial y}$, $\frac{\partial^2 f}{\partial x \partial y}$, and $\frac{\partial^2 f}{\partial y \partial x}$.
4. If $f(x, y) = e^{xy} \sin y$, find $\frac{\partial f}{\partial x}$, $\frac{\partial f}{\partial y}$, $\frac{\partial^2 f}{\partial x \partial y}$, and $\frac{\partial^2 f}{\partial y \partial x}$.
5. If f is the function of Example 9.1.5 directly calculate

$$\frac{\partial^5 f}{\partial x^2 \partial y^2 \partial z}.$$

Verify that it is the same as the mixed partial derivative of f calculated in the example.

6. Let $f : \mathbb{R} \rightarrow \mathbb{R}$ be differentiable on \mathbb{R} and define a function $g : \mathbb{R}^2 \rightarrow \mathbb{R}$ by $g(x, y) = f(x + y)$. Use (9.1.1) to show that $\frac{\partial g}{\partial x} = \frac{\partial g}{\partial y}$ on \mathbb{R}^2 .

7. Theorem 9.1.6 is a statement about a function of two variables. Show how it can be applied several times in a step by step procedure to prove that if $U \subset \mathbb{R}^3$ and f is \mathcal{C}^3 on U , then

$$\frac{\partial^3 f}{\partial x \partial y \partial z} = \frac{\partial^3 f}{\partial z \partial y \partial x}.$$

8. If $p > 0$, let f be the function

$$f(x, y) = \begin{cases} \frac{x^2}{(x^2 + y^2)^p} & \text{if } (x, y) \neq (0, 0) \\ 0 & \text{if } (x, y) = (0, 0) \end{cases}$$

For which values of p is $\frac{\partial f}{\partial x}$ continuous at $(0, 0)$?

9. If f is the function of Example 9.1.7, show by direct calculation that $\frac{\partial^2 f}{\partial x \partial y}$ is not continuous at $(0, 0)$. A similar calculation shows that $\frac{\partial^2 f}{\partial y \partial x}$ is not continuous at $(0, 0)$ (you need not do both calculations).

10. If f is defined on \mathbb{R}^2 by

$$f(x, y) = \begin{cases} \frac{xy}{x^2 + y^2} & \text{if } (x, y) \neq (0, 0) \\ 0 & \text{if } (x, y) = (0, 0), \end{cases}$$

show that both $\frac{\partial f}{\partial x}$ and $\frac{\partial f}{\partial y}$ exist everywhere, but they are not continuous at $(0, 0)$. In fact, f itself is not continuous at $(0, 0)$ (see Example 8.1.3).

9.2 The Differential

Let f be a real valued function defined on an interval on the line. Recall that the equation of the tangent line to the curve $y = f(x)$, at a point a where f is differentiable, is:

$$y = f(a) + f'(a)(x - a)$$

This is the equation of the line which best approximates the curve when x is near a . The right side, is an affine function,

$$T(x) = f(a) + f'(a)(x - a),$$

of x . What is special about T that makes its graph the line which best approximates the curve $y = f(x)$ near a ? For convenience of notation let $h = x - a$, so that $x = a + h$. Then

$$f(a + h) - T(a + h) = f(a + h) - f(a) - f'(a)h$$

and so

$$\lim_{h \rightarrow 0} \frac{f(a+h) - T(a+h)}{h} = \lim_{h \rightarrow a} \frac{f(a+h) - f(a)}{h} - f'(a) = 0.$$

In other words, not only do f and T have the same value at a , but as h approaches 0, the difference between $f(a+h)$ and $T(a+h)$ approaches zero faster than h does. No affine function other than T has this property (Exercise 9.2.7).

Example 9.2.1. What is the best affine approximation to $f(x) = x^3 - 2x + 1$ at the point $(2, 5)$?

Solution: Here, $a = 2$, $f(a) = 5$, and $f'(a) = f'(2) = 10$, so the best affine approximation to $f(x)$ at $x = 2$ is $T(x) = 5 + 10(x - 2)$.

Affine Approximation in Several Variables

By analogy with the single variable case, if $F : D \rightarrow \mathbb{R}^q$ is a function defined on a subset D of \mathbb{R}^p , then the best affine approximation to F at $a \in D$ would be an affine function $T : \mathbb{R}^p \rightarrow \mathbb{R}^q$ such that $F(a+h) - T(a+h)$ goes to 0 faster than h as the vector h approaches 0. In order for this to make sense at all, a must be a limit point of D and, in fact, we will require that a be an interior point of D . This ensures that there is an open ball, centered at a , which is contained in D . It must also be the case that F and its affine approximation T have the same value at a . However, if T is affine and $T(a) = F(a)$, then T has the form

$$T(x) = F(a) + L(x - a),$$

where L is a linear function from \mathbb{R}^p to \mathbb{R}^q .

A function which has a best affine approximation at a is said to be *differentiable* at a . The precise definition of this concept is as follows:

Definition 9.2.2. Let $F : D \rightarrow \mathbb{R}^q$ be a function with domain $D \subset \mathbb{R}^p$, and let a be an interior point of D . We say that F is differentiable at a if there is a linear function $L : \mathbb{R}^p \rightarrow \mathbb{R}^q$ such that

$$\lim_{h \rightarrow 0} \frac{F(a+h) - F(a) - Lh}{\|h\|} = 0. \quad (9.2.1)$$

In this case, we call the linear function L the *differential* of F at a and denote it by $dF(a)$.

Just as in the single variable case, if F is differentiable, then the function

$$T(x) = F(a) + dF(a)(x - a)$$

is the best affine approximation to $F(x)$ for x near a .

Also, as in the single variable case, differentiability implies continuity. We state this in the following theorem, the proof of which is left to the exercises.

Theorem 9.2.3. *If $F : D \rightarrow \mathbb{R}^q$ is differentiable at $a \in D$, then F is continuous at a .*

Example 9.2.4. Let F be the function from \mathbb{R}^2 to \mathbb{R}^2 defined by

$$F(x, y) = (x^2 + y^2, xy).$$

Show that F is differentiable at $(1, 2)$ and its differential is the linear function with matrix

$$A = \begin{pmatrix} 2 & 4 \\ 2 & 1 \end{pmatrix}.$$

Find the affine function which best approximates F near $(1, 2)$.

Solution: With $a = (1, 2)$ and $h = (x - 1, y - 2) = (s, t)$, we have $F(a) = (5, 2)$ and

$$\begin{aligned} F(a+h) - F(a) - Ah &= ((1+s)^2 + (2+t)^2 - 5 - 2s - 4t, (1+s)(2+t) - 2 - 2s - t) \\ &= (s^2 + t^2, st) \end{aligned}$$

Thus, the error $F(a+h) - F(a) - Ah$ if $F(a+h)$ is approximated by $F(a) + Ah$ is

$$(s^2 + t^2, st).$$

Then,

$$\|F(a+h) - F(a) - Ah\|^2 = (s^2 + t^2)^2 + (st)^2 \leq 2\|h\|^4.$$

This implies,

$$\frac{\|F(a+h) - F(a) - Ah\|}{\|h\|} \leq \sqrt{2}\|h\|,$$

which has limit 0 as $h \rightarrow 0$. This shows that F is differentiable at $(1, 2)$ and that $dF(1, 2) = A$.

The best affine approximation to $F(x, y)$ near $(1, 2)$ is

$$\begin{aligned} T(x, y) &= (5, 2) + \begin{pmatrix} 2 & 4 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} x-1 \\ y-2 \end{pmatrix} \\ &= (5 + 2(x-1) + 4(y-2), 2 + 2(x-1) + (y-2)) \\ &= (-5 + 2x + 4y, -2 + 2x + y). \end{aligned}$$

The Differential Matrix

Let $F : D \rightarrow \mathbb{R}^q$ be a function with $D \subset \mathbb{R}^p$ and a an interior point of D . If F is differentiable at a , then it is easy to compute the matrix (c_{ij}) of its differential $dF(a)$. This is called the *differential matrix* of F at a . As usual, we will tend to ignore the technical difference between the linear function $dF(a)$ and its corresponding matrix (see Remark 8.4.13).

We suppose that $F(x) = (f_1(x), f_2(x), \dots, f_q(x))$, so that f_i is the i th coordinate function of F . For $j = 1, \dots, p$, we apply (9.2.1) in the special case in which h approaches 0 along the line $h = te_j$ – that is, along the j th coordinate

axis. Since the vector expression in 9.2.1 converges to 0, the same thing is true of each of its coordinate functions. This means,

$$\lim_{t \rightarrow 0} \frac{f_i(a + te_j) - f_i(a) - c_{ij}t}{t} = 0,$$

which implies

$$c_{ij} = \lim_{t \rightarrow 0} \frac{f_i(a + te_j) - f_i(a)}{t}.$$

The limit that appears in this equation is just the partial derivative

$$\frac{\partial f_i}{\partial x_j}(a),$$

of f_i with respect to its j th variable at the point a . This is true for each i and each j . Thus, we have proved the following theorem.

Theorem 9.2.5. *If $F : D \rightarrow \mathbb{R}^q$ is differentiable at an interior point a of $D \subset \mathbb{R}^p$, then its differential at a is the linear function $dF(a) : \mathbb{R}^p \rightarrow \mathbb{R}^q$ with matrix*

$$\left(\frac{\partial f_i}{\partial x_j}(a) \right)_{ij} = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(a) & \frac{\partial f_1}{\partial x_2}(a) & \cdots & \frac{\partial f_1}{\partial x_p}(a) \\ \frac{\partial f_2}{\partial x_1}(a) & \frac{\partial f_2}{\partial x_2}(a) & \cdots & \frac{\partial f_2}{\partial x_p}(a) \\ \vdots & \vdots & \cdots & \vdots \\ \frac{\partial f_q}{\partial x_1}(a) & \frac{\partial f_q}{\partial x_2}(a) & \cdots & \frac{\partial f_q}{\partial x_p}(a) \end{pmatrix}. \quad (9.2.2)$$

If F is defined and differentiable at all points of an open set $U \subset \mathbb{R}^p$, then we say that F is differentiable on U . Its differential dF is then a function on U whose values are linear transformations from \mathbb{R}^p to \mathbb{R}^q . Equivalently, its differential matrix dF is a $q \times p$ matrix whose entries are functions on U .

Example 9.2.6. Assuming that the function F of Example 9.2.4 is differentiable everywhere, find its differential matrix. Verify that, at $a = (1, 2)$, it is the matrix A of the example.

Solution The coordinate functions for F are given by $f_1(x, y) = x^2 + y^2$ and $f_2(x, y) = xy$. The point a in this example is $a = (1, 2)$. The partial derivatives of f_1 and f_2 are

$$\begin{aligned} \frac{\partial f_1}{\partial x} &= 2x, & \frac{\partial f_1}{\partial y} &= 2y \\ \frac{\partial f_2}{\partial x} &= y, & \frac{\partial f_2}{\partial y} &= x. \end{aligned}$$

Thus, the differential matrix at a general point (x, y) is

$$\begin{pmatrix} 2x & 2y \\ y & x \end{pmatrix}$$

At the particular point $a = (1, 2)$, this is

$$\begin{pmatrix} 2 & 4 \\ 2 & 1 \end{pmatrix}.$$

This is, indeed, the matrix A of Example 9.2.4.

A Condition for Differentiability

Since the vector function in (9.2.1) has limit 0 if and only if each of its coordinate functions has limit 0, we have the following theorem.

Theorem 9.2.7. *If $D \subset \mathbb{R}^p$ and $F = (f_1, \dots, f_q) : D \rightarrow \mathbb{R}^q$ is a function, then F is differentiable at $a \in D$ if and only if, for each i , the coordinate function f_i is differentiable at a . In this case, the differential matrix dF is the matrix whose i th row is the differential df_i of the coordinate function f_i .*

This result allows us to reduce the proof of following theorem to the case $q = 1$.

Theorem 9.2.8. *Let $F = (f_1, \dots, f_q) : U \rightarrow \mathbb{R}^q$ be a function defined on an open subset U of \mathbb{R}^p . If each first order partial derivative of each coordinate function f_i exists on U , then F is differentiable at each point of U where these partial derivatives are all continuous. Thus, if F is C^1 on all of U , then F is differentiable on all of U .*

Proof. By the previous theorem, it is enough to prove that each of the coordinate functions of F is differentiable at the point in question. Hence, it is enough to prove the theorem in the case $q = 1$. To complete the proof, we will prove the following statement by induction on p : If f is a real valued function defined on an open set $U \subset \mathbb{R}^p$ and each first order partial derivative of f exists on U , then f is differentiable at each point of U where all of these partial derivatives are continuous.

If $p = 1$, then the hypothesis implies, in particular, that f has a derivative at each point of U . For a function of one variable, this means the function is differentiable at each point of U . This completes the base case of the induction argument.

We now assume our statement is true for functions of p variables and let f be a function of $p + 1$ variables. We write points of \mathbb{R}^{p+1} in the form (x, y) with $x \in \mathbb{R}^p$ and $y \in \mathbb{R}$. For some $a = (a_1, \dots, a_p) \in \mathbb{R}^p$ and $b \in \mathbb{R}$ we suppose (a, b) is a point of U at which the first order partial derivatives of f are all continuous.

If $h = (h_1, \dots, h_p) \in \mathbb{R}^p$ and $k \in \mathbb{R}$, then

$$\begin{aligned} f(a+h, b+k) - f(a, b) \\ = f(a+h, b) - f(a, b) + f(a+h, b+k) - f(a+h, b). \end{aligned}$$

If we set $g(x) = f(x, b)$ for x in an appropriate neighborhood of a in \mathbb{R}^p and use the Mean Value Theorem in the last variable on the last two terms above, then

this becomes

$$f(a+h, b+k) - f(a, b) = g(a+h) - g(a) + \frac{\partial f}{\partial y}(a+h, c)k, \quad (9.2.3)$$

for some c between b and $b+k$.

Since g is a function of p variables which satisfies the hypotheses of the theorem, g is differentiable at a by our induction assumption. Hence, $dg(a)$ exists and

$$\lim_{h \rightarrow 0} \frac{g(a+h) - g(a) - dg(a)h}{\|h\|} = 0.$$

Because $\|h\| \leq \|(h, k)\|$ this implies

$$\lim_{(h,k) \rightarrow 0} \frac{g(a+h) - g(a) - dg(a)h}{\|(h, k)\|} = 0. \quad (9.2.4)$$

Since $\frac{\partial f}{\partial y}$ is continuous at (a, b) , $|k| \leq \|(h, k)\|$, and $(a+h, c) \rightarrow (a, b)$ as $(h, k) \rightarrow (0, 0)$, we also have

$$\lim_{(h,k) \rightarrow 0} \frac{1}{\|(h, k)\|} \left(\frac{\partial f}{\partial y}(a+h, c) - \frac{\partial f}{\partial y}(a, b) \right) k = 0. \quad (9.2.5)$$

Let v be the vector whose first p components are the components of $dg(a)$ and whose last component is $\frac{\partial f}{\partial y}(a, b)$. Then, by (9.2.3),

$$\begin{aligned} f(a+h, b+k) - f(a, b) - v \cdot (h, k) \\ = g(a+h) - g(a) - dg(a)h + \left(\frac{\partial f}{\partial y}(a+h, c) - \frac{\partial f}{\partial y}(a, b) \right) k, \end{aligned} \quad (9.2.6)$$

On combining (9.2.4), (9.2.5), and (9.2.6), we conclude that

$$\lim_{(h,k) \rightarrow (0,0)} \frac{f(a+h, b+k) - f(a, b) - v \cdot (h, k)}{\|(h, k)\|} = 0,$$

and, hence, that f is differentiable at (a, b) with differential v . This completes the induction and finishes the proof of the theorem. \square

Example 9.2.9. Show that the function $F: \mathbb{R}^2 \rightarrow \mathbb{R}^3$ defined by

$$F(x, y) = (x e^y, y e^x, xy)$$

is differentiable everywhere, and then find its differential matrix.

Solution: The first order partial derivatives of the coordinate functions of F exist and are continuous everywhere. Hence, F is differentiable everywhere by the previous theorem. Its differential matrix is

$$dF(x, y) = \begin{pmatrix} e^y & x e^y \\ y e^x & e^x \\ y & x \end{pmatrix}.$$

A Function Which is not Differentiable

The existence of the first order partial derivatives is not, by itself, enough to ensure that a function is differentiable. This is demonstrated by the next example.

Example 9.2.10. Show that the function f defined by

$$f(x, y) = \begin{cases} \frac{xy}{x^2 + y^2} & \text{if } (x, y) \neq (0, 0) \\ 0 & \text{if } (x, y) = (0, 0). \end{cases}$$

is not differentiable at $(0, 0)$ even though its first order partial derivatives exist everywhere.

Solution: This is a rational function with a denominator which vanishes only at $(0, 0)$. Hence, its first order partial derivatives exist everywhere except possibly at $(0, 0)$. However f is identically 0 on both coordinate axes (that is, $f(x, 0) = 0 = f(0, y)$). Hence, both $\frac{\partial f}{\partial x}$ and $\frac{\partial f}{\partial y}$ exist at $(0, 0)$ and equal 0. However, f is clearly not differentiable at $(0, 0)$, since it is not even continuous at this point (see Example 8.1.3).

Exercise Set 9.2

1. If $L : \mathbb{R}^p \rightarrow \mathbb{R}^p$ is a linear function, show that $dL = L$. In other words, if L has matrix A , then A is the differential matrix of the linear function $L(x) = Ax$.
2. Find the best affine approximation near $(0, 0)$ to the function $F : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ defined by

$$F(x, y) = (xy - 2x + y + 1, x^2 + y^2 + x - 3y + 6).$$

3. If F is the function of the previous exercise, find the best affine approximation to F near $(1, -1)$.
4. Find the differential matrix for the function $G : \mathbb{R}^+ \times \mathbb{R} \rightarrow \mathbb{R}^3$ defined by

$$G(x, y) = (y \ln x, x e^y, \sin xy).$$

Then find the best affine approximation to G at the point $(1, \pi)$.

5. Find the differential of the real valued function $f(x, y, z) = xy^2 \cos xz$. Then find the best affine approximation to f at the point $(1, 1, \pi/2)$.
6. Find the differential of the curve, $\gamma(t) = (\sin(2\pi t), \cos(2\pi t), t^2)$. Then find the best affine approximation to the curve γ at the point $t = 1$.

7. Prove that if f is a real valued function defined on an open interval containing a and if S is an affine function such that

$$\lim_{h \rightarrow 0} \frac{f(a+h) - S(a+h)}{h} = 0,$$

then $S(a+h) = f(a) + f'(a)h$.

8. Prove that if U is a neighborhood of 0 in \mathbb{R}^p and if $F : U \rightarrow \mathbb{R}^q$ is a function such that $F(0) = 0$, then F is differentiable at 0 with $dF = 0$ if and only if $\lim_{x \rightarrow 0} \|F(x)\|/\|x\| = 0$.
9. Prove Theorem 9.2.3. That is, prove that if a function is differentiable at a point in its domain, then it is continuous at that point.
10. Does the function defined by

$$f(x, y) = \begin{cases} \frac{x^3}{x^2 + y^2} & \text{if } (x, y) \neq (0, 0) \\ 0 & \text{if } (x, y) = (0, 0) \end{cases}$$

have first order partial derivatives at every point of \mathbb{R}^2 ? Is this function differentiable at $(0, 0)$? Give reasons for your answers.

11. If $f : \mathbb{R}^p \rightarrow \mathbb{R}$ is differentiable at $a \in \mathbb{R}^p$, then show that, for each $h \in \mathbb{R}^p$, the function $g : \mathbb{R} \rightarrow \mathbb{R}$ defined by $g(t) = f(a + th)$ has a derivative at $t = 0$. Can you compute it in terms of $df(a)$ and h ?
12. Prove that a function $F : \mathbb{R}^p \rightarrow \mathbb{R}^q$ is affine if and only if it is differentiable everywhere and its differential matrix is constant.

9.3 The Chain Rule

The differential of a function of several variables has properties similar to those of the derivative of a real valued function of a single variable. The simplest of these are stated in the following theorem, whose proof is left to the exercises.

Theorem 9.3.1. *Suppose F and G are functions defined on an open set $U \subset \mathbb{R}^p$, with values in \mathbb{R}^q , and c is a scalar. If F and G are differentiable at a point $x \in U$, then*

- (a) cF is differentiable at x and $d(cF)(x) = cdF(x)$; and
- (b) $F + G$ is differentiable at x and $d(F + G)(x) = dF(x) + dG(x)$.

A result which is more difficult to prove, but is of great importance is the chain rule for functions of several variables. The proof becomes considerably simpler if we reformulate the concept of differentiability in the following way.

An Equivalent Formulation of Differentiability

If f is a real valued function defined on an open interval containing the point $a \in \mathbb{R}$, then we can always express $f(a+h) - f(a)$ for h near but not equal to 0 in the following way:

$$f(a+h) - f(a) = q(h)h, \quad (9.3.1)$$

where $q(h)$ is just the difference quotient

$$q(h) = \frac{f(a+h) - f(a)}{h}.$$

Of course, f is differentiable at a if and only if q has a limit as $h \rightarrow 0$. The derivative is then defined to be this limit. The function q becomes continuous at 0 if it is given the value $f'(a)$ at $h = 0$ and then (9.3.1) holds at $h = 0$ as well as at all nearby points. In fact, the differentiability of f at a is equivalent to the existence of a function q which satisfies (9.3.1) and is continuous at $h = 0$. This suggests the following reformulation of the definition of differentiability.

Theorem 9.3.2. *Let F be a function defined on an open set $U \subset \mathbb{R}^p$ with values in \mathbb{R}^q and let a be a point of U . Then F is differentiable at a if and only if there is a $q \times p$ matrix valued function $Q(h)$, defined in a neighborhood of 0, such that Q is continuous at 0 and $F(a+h) - F(a)$ is the vector-matrix product*

$$F(a+h) - F(a) = Q(h)h$$

for all h in a neighborhood of 0. If this condition holds, then $dF(a) = Q(0)$.

Proof. Suppose a matrix Q with the required properties exists on some neighborhood V of 0. Then, for $h \in V$,

$$\frac{F(a+h) - F(a) - Q(0)h}{\|h\|} = \frac{Q(h)h - Q(0)h}{\|h\|} = \frac{(Q(h) - Q(0))h}{\|h\|}.$$

This expression has norm less than or equal to $\|Q(h) - Q(0)\|$ which converges to 0 as $h \rightarrow 0$, since Q is continuous at 0. Thus, F is differentiable and its differential matrix is $Q(0)$.

Conversely, suppose F is differentiable at a . If we set

$$\epsilon(h) = F(a+h) - F(a) - dF(a)h.$$

Then ϵ is a function on a neighborhood of 0 with values in \mathbb{R}^q and

$$\lim_{h \rightarrow 0} \frac{\epsilon(h)}{\|h\|} = 0.$$

If, when written out in terms of coordinate functions, $\epsilon = (\epsilon_1, \epsilon_2, \dots, \epsilon_q)$, and

$h = (h_1, h_2, \dots, h_p)$, then we define a $q \times p$ matrix $\Delta(h)$ by

$$\Delta(h) = \|h\|^{-2} \begin{pmatrix} \epsilon_1 h_1 & \epsilon_1 h_2 & \cdots & \epsilon_1 h_p \\ \epsilon_2 h_1 & \epsilon_2 h_2 & \cdots & \epsilon_2 h_p \\ \vdots & \vdots & \cdots & \vdots \\ \vdots & \vdots & \cdots & \vdots \\ \epsilon_q h_1 & \epsilon_q h_2 & \cdots & \epsilon_q h_p \end{pmatrix}.$$

This is a matrix valued function of h , defined on a neighborhood of 0, except at 0 itself. Moreover, if we define this function to be 0 when $h = 0$, then it becomes continuous at $h = 0$, since

$$\frac{|\epsilon_i(h)h_j|}{\|h\|^2} \leq \frac{\|\epsilon(h)\|\|h\|}{\|h\|^2} = \frac{\|\epsilon(h)\|}{\|h\|},$$

and this has limit 0 as $h \rightarrow 0$. Note also that if we apply the matrix $\Delta(h)$ to the vector h , the result is

$$\Delta(h)h = \epsilon(h),$$

Thus, if we set

$$Q(h) = dF(a) + \Delta(h),$$

then Q is continuous at $h = 0$, $Q(0) = dF(a)$, and

$$F(a+h) - F(a) = dF(a)h + \epsilon(h) = dF(a)h + \Delta(h)h = Q(h)h.$$

This completes the proof. \square

The Chain Rule

After the above reformulation of differentiability, the chain rule has a simple proof.

Theorem 9.3.3. *Let U and V be open subsets of \mathbb{R}^r and \mathbb{R}^p , respectively, and let $G : U \rightarrow \mathbb{R}^p$ and $F : V \rightarrow \mathbb{R}^q$ be functions with $G(U) \subset V$. Suppose $a \in U$, G is differentiable at a , and F is differentiable at $b = G(a)$. Then $F \circ G$ is differentiable at a and*

$$d(F \circ G)(a) = dF(G(a))dG(a).$$

Proof. By the previous theorem, there are matrix valued functions Q_G and Q_F , defined in neighborhoods of 0 in \mathbb{R}^r and \mathbb{R}^p , respectively, each continuous at 0, with $Q_F(0) = dF(b)$, $Q_G(0) = dG(a)$, and such that

$$G(a+h) - G(a) = Q_G(h)h \quad \text{and} \quad F(b+k) - F(b) = Q_F(k)k$$

for h and k in appropriate neighborhoods of 0. Then, since $G(a) = b$,

$$F \circ G(a+h) - F \circ G(a) = F(b + Q_G(h)h) - F(b) = Q_F(Q_G(h)h)Q_G(h)h.$$

Since Q_G and Q_F are both continuous at 0, we have

$$\lim_{h \rightarrow 0} Q_F(Q_G(h)h)Q_G(h) = Q_F(0)Q_G(0) = dF(b)dG(a) = dF(G(a))dG(a).$$

Thus, if we choose $Q_{F \circ G}(h)$ to be $Q_F(Q_G(h)h)Q_G(h)$, it satisfies the conditions of the previous theorem with F replaced by $F \circ G$ and, hence, by that theorem, $d(F \circ G)(a)$ exists and equals $dF(G(a))dG(a)$. \square

Example 9.3.4. Let $f(x, y)$ be a real valued function of two variables and let

$$\phi(r, s, t) = f(r(s+t), r(s-t)).$$

Find $d\phi(1, 2, 1)$ if $\frac{\partial f}{\partial x}(3, 1) = 4$ and $\frac{\partial f}{\partial y}(3, 1) = -5$.

Solution: The function ϕ is just $f \circ G$, where $G: \mathbb{R}^3 \rightarrow \mathbb{R}^2$ is defined by

$$G(r, s, t) = (r(s+t), r(s-t)).$$

We have $G(1, 2, 1) = (3, 1)$ and

$$dG(1, 2, 1) = \begin{pmatrix} 3 & 1 & 1 \\ 1 & 1 & -1 \end{pmatrix}.$$

Thus, $d\phi(1, 2, 1) = dF(G(1, 2, 1))dG(1, 2, 1)$ is

$$\begin{aligned} & \left(\frac{\partial f}{\partial x}(3, 1), \frac{\partial f}{\partial y}(3, 1) \right) \begin{pmatrix} 3 & 1 & 1 \\ 1 & 1 & -1 \end{pmatrix} \\ &= (4, -5) \begin{pmatrix} 3 & 1 & 1 \\ 1 & 1 & -1 \end{pmatrix} = (7, -1, 9). \end{aligned}$$

Example 9.3.5. If $F(x, y) = (f_1(x, y), f_2(x, y))$ is a differentiable function from \mathbb{R}^2 to \mathbb{R}^2 and we define $G: \mathbb{R}^2 \rightarrow \mathbb{R}^2$ by $G(s, t) = F(s^2 + t^2, s^2 - t^2)$, find an expression for the differential matrix of G in terms of the partial derivatives of f_1 and f_2 .

Solution: The function G is $F \circ H$ where $H(s, t) = (s^2 + t^2, s^2 - t^2)$. The differential matrices of F and H are

$$dF = \begin{pmatrix} \frac{\partial f_1}{\partial x} & \frac{\partial f_1}{\partial y} \\ \frac{\partial f_2}{\partial x} & \frac{\partial f_2}{\partial y} \end{pmatrix} \text{ and } dH = \begin{pmatrix} 2s & 2t \\ 2s & -2t \end{pmatrix}.$$

By the chain rule,

$$\begin{aligned} dG(s, t) &= d(F \circ H)(s, t) = dF(H(s, t))dH(s, t) \\ &= \begin{pmatrix} 2s \left(\frac{\partial f_1}{\partial x} + \frac{\partial f_1}{\partial y} \right) & 2t \left(\frac{\partial f_1}{\partial x} - \frac{\partial f_1}{\partial y} \right) \\ 2s \left(\frac{\partial f_2}{\partial x} + \frac{\partial f_2}{\partial y} \right) & 2t \left(\frac{\partial f_2}{\partial x} - \frac{\partial f_2}{\partial y} \right) \end{pmatrix}, \end{aligned}$$

where the partial derivatives of f_1 and f_2 are to be evaluated at the point $H(s, t) = (s^2 + t^2, s^2 - t^2)$.

Differential of an Inner Product

The following theorem is a nice application of the chain rule.

Theorem 9.3.6. *Suppose F and G are functions defined in a neighborhood of a point $a \in \mathbb{R}^p$ and with values in \mathbb{R}^q . If F and G are both differentiable at a , then $F \cdot G$ is also differentiable at a and*

$$d(F \cdot G)(a) = G(a)dF(a) + F(a)dG(a),$$

where each of the products on the right is the matrix product of a $1 \times q$ times a $q \times p$ matrix.

Proof. Let $H : \mathbb{R}^{2q} \rightarrow \mathbb{R}$ be defined by

$$H(u, v) = u \cdot v,$$

where, if $u = (u_1, \dots, u_q)$ and $v = (v_1, \dots, v_q)$ are vectors in \mathbb{R}^q , then (u, v) denotes the vector $(u_1, \dots, u_q, v_1, \dots, v_q)$ in \mathbb{R}^{2q} .

Now $F \cdot G = H \circ (F, G)$, where (F, G) denotes the function with values in \mathbb{R}^{2q} whose first q coordinate functions are the coordinate functions of F and whose last q coordinate functions are the coordinate functions of G .

The function H is differentiable everywhere because its coordinate functions $u_i v_i$ have continuous partial derivatives everywhere. That is,

$$\frac{\partial u_i v_i}{\partial u_i} = v_i, \quad \frac{\partial u_i v_i}{\partial v_i} = u_i,$$

and all other first order partial derivatives are zero. This means that its differential is the $1 \times 2q$ matrix

$$(v_1, \dots, v_q, u_1, \dots, u_q).$$

Since F and G are differentiable at a , the coordinate functions of both are all differentiable at a . This implies that the function (F, G) is differentiable at a , since each of its coordinate functions is a coordinate function of F or a coordinate function of G . Furthermore,

$$d(F, G)(a) = \begin{pmatrix} dF(a) \\ dG(a) \end{pmatrix},$$

where the matrix on the right has its first q rows the rows of $dF(a)$ and its last q rows the rows of $dG(a)$.

By the chain rule,

$$\begin{aligned} d(F \cdot G)(a) &= dH(F(a), G(a))d(F, G)(a) \\ &= (G(a), F(a)) \begin{pmatrix} dF(a) \\ dG(a) \end{pmatrix} \\ &= G(a)dF(a) + F(a)dG(a). \end{aligned}$$

□

Dependent Variable Notation

A notation that is often used in connection with differentiation and specifically the chain rule is one which emphasizes the variables in a problem, some of which depend on others through functional relations, but which de-emphasizes the functions defining these relations. In this notation, a function F of p variables with values in \mathbb{R}^q determines a vector of q dependent variables

$$u = (u_1, u_2, \dots, u_q)$$

which depend on a vector of p variables

$$x = (x_1, x_2, \dots, x_p)$$

through the relation $u = F(x)$. The differential matrix is then the matrix

$$\left(\frac{\partial u_i}{\partial x_j} \right)_{ij} = \begin{pmatrix} \frac{\partial u_1}{\partial x_1} & \frac{\partial u_1}{\partial x_2} & \dots & \frac{\partial u_1}{\partial x_p} \\ \frac{\partial u_2}{\partial x_1} & \frac{\partial u_2}{\partial x_2} & \dots & \frac{\partial u_2}{\partial x_p} \\ \vdots & \vdots & \dots & \vdots \\ \frac{\partial u_q}{\partial x_1} & \frac{\partial u_q}{\partial x_2} & \dots & \frac{\partial u_q}{\partial x_p} \end{pmatrix}.$$

where $\frac{\partial u_i}{\partial x_j}$ is understood to be the partial derivative $\frac{\partial f_i}{\partial x_j}$ of the i th coordinate function of F evaluated at a generic point x of the domain of F .

Now the variables x_j themselves may depend on a vector of variables

$$t = (t_1, t_2, \dots, t_r)$$

through a function G . The differential matrix for this relationship would be the matrix

$$\left(\frac{\partial x_j}{\partial t_k} \right)_{jk}.$$

Since the variables u_i depend on the variables x_j , which in turn depend on the variables t_k , the variables u_i also depend on the variables t_k (through the function $F \circ G$), and the differential matrix for this relationship is denoted

$$\left(\frac{\partial u_i}{\partial t_k} \right)_{ik}.$$

Using this notation, the chain rule becomes

$$\left(\frac{\partial u_i}{\partial t_k} \right)_{ik} = \left(\frac{\partial u_i}{\partial x_j} \right)_{ij} \left(\frac{\partial x_j}{\partial t_k} \right)_{jk}, \quad (9.3.2)$$

where the expression on the right is the product of the indicated matrices. This product will involve the variables x_j as well as the variables t_k and it is important to remember that the x_j s are themselves functions of the variables t_k .

A Change of Variables

Example 9.3.7. If $u = f(x, y)$ expresses the variable u as a function of Cartesian coordinates (x, y) on an open subset of the plane, what is the relationship between the differential matrix of u as a function of (x, y) and its differential matrix as a function of the corresponding polar coordinates (r, θ) , where $x = r \cos \theta$ and $y = r \sin \theta$.

Solution: The change of coordinate transformation $(x, y) = (r \cos \theta, r \sin \theta)$ has differential matrix

$$\begin{pmatrix} \cos \theta & -r \sin \theta \\ \sin \theta & r \cos \theta \end{pmatrix}.$$

Thus,

$$\begin{pmatrix} \frac{\partial u}{\partial r} & \frac{\partial u}{\partial \theta} \end{pmatrix} = \begin{pmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \end{pmatrix} \begin{pmatrix} \cos \theta & -r \sin \theta \\ \sin \theta & r \cos \theta \end{pmatrix},$$

or

$$\begin{aligned} \frac{\partial u}{\partial r} &= \cos \theta \frac{\partial u}{\partial x} + \sin \theta \frac{\partial u}{\partial y} \\ \frac{\partial u}{\partial \theta} &= -r \sin \theta \frac{\partial u}{\partial x} + r \cos \theta \frac{\partial u}{\partial y}. \end{aligned}$$

Exercise Set 9.3

1. If F is a function from an open subset U of \mathbb{R}^p to \mathbb{R}^q which is differentiable at a and if B is an $r \times q$ matrix, then show that $d(BF)(a) = BdF(a)$. Here, $BF(x)$ is the matrix B applied to the vector $F(x)$ and $BdF(a)$ is the product of the matrix B and the matrix $dF(a)$.
2. If $f(x, y)$ is a differentiable function of $(x, y) \in \mathbb{R}^2$, and $g(t) = f(tx, ty)$, for all $t \in \mathbb{R}$, find $g'(1)$ in terms of the partial derivatives of f .
3. An n -homogeneous function on \mathbb{R}^2 is a function that satisfies $f(tx, ty) = t^n f(x, y)$ for all $t \in \mathbb{R}$ and $(x, y) \in \mathbb{R}^2$. Show that a differentiable function on \mathbb{R}^2 is n -homogeneous if and only if it satisfies the differential equation

$$x \frac{\partial f}{\partial x} + y \frac{\partial f}{\partial y} = nf$$

at each $(x, y) \in \mathbb{R}^2$.

4. If f is a differentiable function on \mathbb{R} and $g(x, y) = f(xy)$, show that

$$x \frac{\partial g}{\partial x} - y \frac{\partial g}{\partial y} = 0.$$

5. If f and g are twice differentiable functions on \mathbb{R} and

$$h(x, y) = f(x - y) + g(x + y),$$

show that h satisfies the wave equation:

$$\frac{\partial^2 h}{\partial x^2} - \frac{\partial^2 h}{\partial y^2} = 0.$$

6. If u is a variable which is a differentiable function of (x, y) in an open set $U \subset \mathbb{R}^2$, if x and y are differentiable functions of $(s, t) \in V$ for an open set $V \subset \mathbb{R}^2$, and if $(x, y) \in U$ whenever $(s, t) \in V$, then use the chain rule to obtain expressions for $\frac{\partial u}{\partial s}$ and $\frac{\partial u}{\partial t}$ on V in terms of the partial derivatives of u with respect to x and y and the partial derivatives of x and y with respect to s and t .
7. Do the preceding exercise in the special case where

$$x = as + bt \quad \text{and} \quad y = cs + dt.$$

for some constants a, b, c, d .

8. If $F(x, y) = (f_1(x, y), f_2(x, y))$ is a differentiable function from \mathbb{R}^2 to \mathbb{R}^2 and we define $G : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ by $G(s, t) = F(st, s+t)$, find an expression for the differential matrix of G in terms of the partial derivatives of f_1 and f_2 .
9. If (x, y, z) are the Cartesian coordinates of a point in \mathbb{R}^3 and the spherical coordinates of the same point are r, θ, ϕ , then

$$x = r \cos \theta \sin \phi, \quad y = r \sin \theta \sin \phi, \quad z = r \cos \phi.$$

Let u be a variable which is a differentiable function of (x, y, z) on \mathbb{R}^3 . Find a formula for the partial derivatives of u with respect to r, θ, ϕ in terms of its partial derivatives with respect to x, y, z .

10. Suppose U and V are open subsets of \mathbb{R}^p and $F : U \rightarrow V$ has an inverse function $G : V \rightarrow U$. This means $F \circ G(y) = y$ for all $y \in V$ and $G \circ F(x) = x$ for all $x \in U$. Show that, if F is differentiable on U and G is differentiable on V , then $dF(x)$ is non-singular at each $x \in U$, and for each $x \in U$,

$$dF(x)^{-1} = dG(y) \quad \text{where} \quad y = F(x).$$

11. Show that if F is differentiable function on an open set $U \subset \mathbb{R}^p$ with values in \mathbb{R}^q , then the real valued function $\|F(x)\|^2$ on U has zero differential at x if and only if the vector $F(x)$ is orthogonal to each of the columns of $dF(x)$.
12. Prove Theorem 9.3.1.
13. If $f(x, y) = x^2 + y^2$ find a 1×2 matrix valued function Q which satisfies the conclusion of Theorem 9.3.2 for f .

14. In the proof of Theorem 9.3.3, the following fact is used twice: If $A(h)$ is a $q \times p$ matrix whose entries are functions of $h \in \mathbb{R}^p$ and if $A(h)$ is continuous at $h = 0$, then $\lim_{h \rightarrow 0} A(h)h = 0$, where $A(h)h$ is the result of the matrix $A(h)$ acting via vector-matrix product on the vector h . Prove that this limit is 0, as claimed.

9.4 Applications of the Chain Rule

The Gradient

The case $q = 1$ is of special interest in this discussion. In this case, we are dealing with a real valued function f on a domain $D \subset \mathbb{R}^p$. At any point x where the function f is differentiable, its differential matrix is a $1 \times p$ matrix – that is, a row vector

$$df = \left(\frac{\partial f}{\partial x_1}, \dots, \frac{\partial f}{\partial x_p} \right),$$

The resulting vector is called the *gradient* of f at x . It is sometimes denoted ∇f and sometimes denoted $\text{grad } f$.

If $f(x_1, \dots, x_p)$ is the function f with its argument written out in terms of coordinates, then a notation often used for df is

$$df = \frac{\partial f}{\partial x_1} dx_1 + \frac{\partial f}{\partial x_2} dx_2 + \dots + \frac{\partial f}{\partial x_p} dx_p. \quad (9.4.1)$$

The interpretation of this is as follows: It is understood that df and the partial derivatives in this equation are evaluated at some generic point x of the domain of f . For each j , dx_j is the differential of the j th coordinate function x_j on \mathbb{R}^p . As such, it is the linear transformation from \mathbb{R}^p to \mathbb{R} which sends a vector $(v_1, \dots, v_p) \in \mathbb{R}^p$ to its j th component v_j . As a row vector, it is the vector which has 1 as j th component and 0 for all other components. Earlier we called this vector e_j , but in the context of differentials it is common to call it dx_j . Equation 9.4.1 expresses the fact that, for each function f as above, df at a given point is a linear combination of the basis elements dx_j with the coefficients being the corresponding partial derivatives of f at that point.

Example 9.4.1. If $f(x, y, z) = z^2 + \sin xy$, find the gradient of f at a generic point (x, y, z) and at the particular point $(1, 0, 3)$.

Solution: At (x, y, z) the gradient of f is

$$df = (y \cos xy, x \cos xy, 2z).$$

At $(x, y, z) = (1, 0, 3)$ this is the vector $(0, 1, 6)$. In terms of the basis vectors dx, dy, dz , we have

$$df = y \cos xy \, dx + x \cos xy \, dy + 2z \, dz,$$

which, at $(x, y, z) = (1, 0, 3)$ is $dy + 6 \, dz$.

Directional Derivatives

We specify a *direction* in \mathbb{R}^p by specifying a unit vector (vector of length 1) that points in this direction. For example, in \mathbb{R}^2 we may specify a direction by specifying an angle θ relative to the positive x axis, but this is equivalent to specifying the unit vector $(\cos \theta, \sin \theta)$ which points in this direction.

Given a function f , defined on a neighborhood of a point $a \in \mathbb{R}^p$, each first order partial derivative of f at a is defined by restricting f to a line through a parallel to one of the coordinate axes and differentiating the resulting function of one variable. However, there is nothing special about the coordinate axes. We may restrict f to a line in any direction through a and differentiate the resulting function of one variable. This leads to the concept of directional derivative.

Definition 9.4.2. Suppose f is a function defined in a neighborhood of $a \in \mathbb{R}^p$ and u is a unit vector in \mathbb{R}^p . The *directional derivative* of f at a , in the direction u , is defined to be

$$D_u f(a) = \frac{d}{dt} f(a + tu)|_{t=0}$$

If f happens to be differentiable at a , then its directional derivatives all exist and are easily calculated.

Theorem 9.4.3. Suppose f is a function defined in a neighborhood of $a \in \mathbb{R}^p$ and differentiable at a . If u is a unit vector in \mathbb{R}^p , then the directional derivative $D_u f(a)$ exists and

$$D_u f(a) = df(a)u.$$

Proof. If $g : \mathbb{R} \rightarrow \mathbb{R}^p$ is defined by $g(t) = a + tu$, then $dg(t) = g'(t) = u$ and $D_u f(a) = d(f \circ g)(0)$. The chain rule implies that this exists and is equal to $df(a)dg(0) = df(a)u$. \square

The directional derivative $D_u f(a)$ represents the rate of change of f as we pass through a in the direction specified by u . If this is positive, then it represents the rate of *increase* of f in the u direction as we pass through a .

The proof of the following theorem is left to the exercises.

Theorem 9.4.4. Suppose f is a real valued function which is defined and differentiable in a neighborhood of $a \in \mathbb{R}^p$, and suppose that $df(a) \neq 0$. Then the gradient $df(a)$ points in the direction of greatest increase for f at a – that is, $D_u f(a)$ has its maximum value when the unit vector u is a positive scalar multiple of $df(a)$.

Example 9.4.5. If $f(x, y) = 2 - x^2 - y^2$, find the direction of greatest increase of f at $(1, 1)$ and the rate of increase of f in this direction at $(1, 1)$.

Solution: The gradient of f is

$$df(x, y) = (-2x, -2y).$$

At $(1, 1)$ this is

$$df(1, 1) = (-2, -2).$$

A unit vector which points in the same direction is $u = (-1/\sqrt{2}, -1/\sqrt{2})$. The directional derivative in the direction of u is

$$D_u f(1, 1) = df(1, 1) \cdot u = \sqrt{2} + \sqrt{2} = 2\sqrt{2}.$$

The Derivative of a Curve

Another special case of importance in the study of differentials is the case of a curve in \mathbb{R}^q – that is, a function

$$\gamma(t) = (\gamma_1(t), \gamma_2(t), \dots, \gamma_q(t)),$$

defined on an interval $I \subset \mathbb{R}$, with values in \mathbb{R}^q . In this case, the differential matrix $d\gamma$, at an interior point of I is a $q \times 1$ matrix – that is, a column vector. This is the column vector obtained by transposing the vector

$$\gamma'(t) = (\gamma'_1(t), \gamma'_2(t), \dots, \gamma'_q(t)).$$

obtained by differentiating the coordinate functions of γ .

If $a \in I$, the best affine approximation to $\gamma(t)$ for t near a is the function

$$\tau(t) = \gamma(a) + \gamma'(a)(t - a).$$

Assuming $\gamma'(a) \neq 0$, this is a parametric equation for a line through $b = \gamma(a)$ which is parallel to the vector $\gamma'(a)$. If one more restriction on the curve γ is met, this line will be called the *tangent line* to the curve at $\gamma(a)$.

The additional restriction needed on γ is that a is the only point on the interval I at which γ has the value b . Otherwise, the curve crosses itself at b and the tangent line to the curve at b is not well defined – there is a different tangent line for each branch of the curve passing through b (see Figure 9.1). In this case, we will say that b is a *crossing point* for γ . Crossing points can be eliminated by replacing the interval I with a smaller open interval, containing a , but no other points at which γ has the value $\gamma(a)$. In our continuing discussion of curves and their tangent lines, we will assume that $\gamma(a)$ is not a crossing point of γ . This assumption and the assumption that $\gamma'(a) \neq 0$ ensure that γ has a well defined tangent line at $\gamma(a)$.

Note that each point $\tau(t)$ which is on the tangent line and sufficiently close to $\gamma(a)$ determines a parameter value $t \in I$ and this, in turn, determines a point $\gamma(t)$ on the curve. The two points $\gamma(t)$ and $\tau(t)$ differ from one another by

$$\gamma(t) - \gamma(a) - \gamma'(a)(t - a)$$

and the norm of this vector approaches 0 faster than $t - a$ approaches 0 as $t \rightarrow a$. This justifies the claim that the curve γ and the line τ are tangent at the point $\gamma(a)$. Note, however, that this line of reasoning is only valid if $\gamma'(a) \neq 0$, since, otherwise, τ is constant and fails to determine a non-degenerate line.

If $\gamma'(a) \neq 0$, the vector

$$T(a) = \frac{\gamma'(a)}{\|\gamma'(a)\|}$$

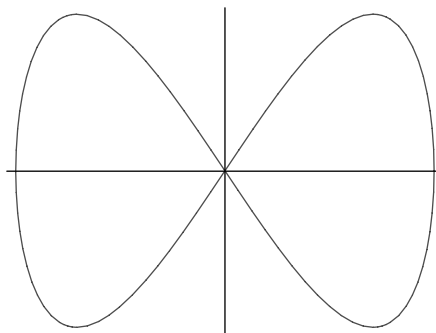


Figure 9.1: Curve With a Crossing Point

is a unit vector (a vector of length one) which is parallel to the tangent line at a . It is called the *tangent vector* to the curve at $\gamma(a)$.

The vector $\gamma'(a)$ is sometimes called the *velocity vector* of the curve at $\gamma(a)$, since it does represent velocity in the case where the curve is describing the motion of a body through space.

Example 9.4.6. The parameterized curve $\gamma(t) = (\cos t, \sin 2t)$, $0 < t < 2\pi$, passes through the origin. At the origin, find its velocity vector, tangent vector, and tangent line. Do the same problem if the domain of γ is restricted to $(0, \pi)$.

Solution: The origin is a crossing point for this curve (see Figure 9.1). The curve passes through the origin when $t = \pi/2$ and when $t = 3\pi/2$. Thus, there is no well defined velocity vector, tangent vector, or tangent line. If we restrict the domain of γ to the interval $(0, \pi)$, then the effect is to choose one branch of the curve and the crossing is eliminated. Then the curve passes through $(0, 0)$ only at $\pi/2$. We have

$$\gamma'(t) = (-\sin t, 2 \cos 2t) \quad \text{and} \quad \gamma'(\pi/2) = (-1, -2).$$

Hence, the velocity vector at $(0, 0)$ is $\gamma'(\pi/2) = (-1, -2)$, the tangent vector at this point is $\frac{\gamma'(\pi/2)}{\|\gamma'(\pi/2)\|} = \left(\frac{-1}{\sqrt{5}}, \frac{-2}{\sqrt{5}} \right)$ and a parametric equation for the tangent line to this curve at $(0, 0)$ is

$$\tau(t) = (0, 0) + (t - \pi/2)(-1, -2) = (\pi/2 - t, \pi - 2t).$$

If we define the domain of γ to be $(\pi, 2\pi)$, then we are choosing the other branch of the curve – the one which passes through $(0, 0)$ at $t = 3\pi/2$. We leave the problem of finding the tangent line to the curve at this point to the exercises.

Higher Dimensional Tangent spaces

The following discussion is a higher dimensional version of the above discussion of curves and tangent lines. Suppose $p < q$, $U \subset \mathbb{R}^p$ is open, and $F : U \rightarrow \mathbb{R}^q$

is a smooth function. Since dF is a $q \times p$ matrix at each point of U and $p < q$, the maximal possible rank of dF is p . Suppose $a \in U$ is a point at which dF has rank p . Then the function

$$\Phi(x) = F(a) + dF(a)(x - a) \quad (9.4.2)$$

is an affine function of rank p (Definition 8.5.8). This implies that its image is a p -dimensional affine subspace of \mathbb{R}^q (a translate of a p -dimensional linear subspace). Each point in this subspace which is sufficiently near $F(a)$ is $\Phi(x)$ for some $x \in U$ and, for such a point, there is a corresponding point $F(x)$ in the image of F . Now Φ is the best affine approximation to F near a and so the norm of

$$F(x) - \Phi(x) = F(x) - F(a) - dF(a)(x - a)$$

approaches 0 faster than $\|x - a\|$ approaches 0 as $x \rightarrow a$. This justifies calling the image of Φ the *tangent space* to the image of F at $F(a)$. At least, this is the case if a is the only point in U at which F has the value $F(a)$ (so that $F(a)$ is not a *crossing point* of F). The situation described in this discussion is important enough to warrant a definition.

A function F , defined on U , is *one to one* if there are no two distinct points of U at which F has the same value.

Definition 9.4.7. With $p < q$, let U be an open subset of \mathbb{R}^p and $F : U \rightarrow \mathbb{R}^q$ be a one to one smooth function on U such that $dF(a)$ has rank p at each point $a \in U$. Then we will call the image S of F a smoothly parameterized p -surface in \mathbb{R}^q and we will say that F is a smooth parameterization of S .

We define the tangent space of S at each $b = F(a) \in S$ to be the affine subspace of \mathbb{R}^q which is the image of the function Φ of (9.4.2).

In the case where $p = q - 1$, a p -surface in \mathbb{R}^q is called a *hypersurface* in \mathbb{R}^q and its tangent space at $b = F(a)$ is its tangent *hyperplane* at b . If $q = 3$ and $p = 2$, then a 2-surface in \mathbb{R}^3 is just a surface and its tangent space at b is its tangent plane at b .

Example 9.4.8. . With $a = r_0 \cos \theta_0$, $b = r_0 \sin \theta_0$, and $r_0 > 0$, find the tangent plane at (a, b, r_0) to the cone in \mathbb{R}^3 parameterized by the function G defined by

$$G(r, \theta) = (r \cos \theta, r \sin \theta, r).$$

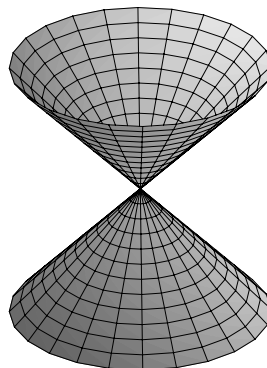
Is there a point on the cone where the tangent plane is not defined?

Solution: The differential dG at (r_0, θ_0) is

$$\begin{pmatrix} \cos \theta_0 & -r_0 \sin \theta_0 \\ \sin \theta_0 & r_0 \cos \theta_0 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} a/r_0 & -b \\ b/r_0 & a \\ 1 & 0 \end{pmatrix}.$$

If $r_0 \neq 0$, this matrix has rank 2. It defines a parameterized plane by

$$\begin{aligned} \Phi(r, \theta) &= \begin{pmatrix} a \\ b \\ r_0 \end{pmatrix} + \begin{pmatrix} a/r_0 & -b \\ b/r_0 & a \\ 1 & 0 \end{pmatrix} \begin{pmatrix} r - r_0 \\ \theta - \theta_0 \end{pmatrix} \quad \text{or} \\ \Phi(r, \theta) &= (ar/r_0 - b(\theta - \theta_0), br/r_0 + a(\theta - \theta_0), r). \end{aligned}$$

Figure 9.2: Cone in \mathbb{R}^3

There is no tangent plane to the curve at the origin. The differential of G at this point has rank 1 rather than rank 2 and the origin is a crossing point, which means that G does not satisfy the conditions of Definition 9.4.7. In fact, it is apparent from Figure 9.2 that there is no parametrization of the cone in a neighborhood of the origin that will make it a smooth p -surface and no reasonable candidate for a tangent plane.

Level Sets

If $F : U \rightarrow \mathbb{R}^d$ is a function defined on an open subset U of \mathbb{R}^a , then a *level set* for F is a set of the form

$$S = \{y \in U : F(y) = c\}$$

where c is a constant vector in \mathbb{R}^d . By subtracting c from F , we can always arrange things so that S is the subset of U defined by the equation $F(y) = 0$.

Under these circumstances, it is often the case that locally (meaning near a given point $b \in S$) S can be represented as a smoothly parameterized surface of some dimension and its tangent space can be realized as the set of solutions y to the equation

$$dF(b)(y - b) = 0.$$

We will learn more about when this is true in the last section of this chapter. For now, we settle for a couple of preliminary results.

Theorem 9.4.9. *With F as above, let V be an open subset of \mathbb{R}^p and $G : V \rightarrow \mathbb{R}^a$ a smooth function such that $G(V)$ is contained in a level set of F . Then*

$$dF(y)dG(x) = 0, \quad \text{where } y = G(x),$$

for each $x \in V$.

Proof. If the image of G lies in a level set of F , then there is a constant $c \in \mathbb{R}^d$ such that

$$(F \circ G)(x) = c \quad \text{for all } x \in V.$$

Then, by the chain rule,

$$0 = d(F \circ G)(x) = dF(G(x))dG(x).$$

□

Example 9.4.10. Show that a curve γ in \mathbb{R}^p of constant norm, $\|\gamma(t)\|$, has its tangent vector orthogonal to its position vector at each point.

Solution: If $\|\gamma(t)\|$ is constant, then so is $\|\gamma(t)\|^2$. This means that γ has its image in a level set of the function $f(x) = \|x\|^2 = x \cdot x$. By the previous theorem, $df(x)d\gamma(t) = 0$ if $x = \gamma(t)$ is a point on the curve. This means that the velocity vector $\gamma'(t)$ is orthogonal to the gradient $2x$ of the function f at each point $x = \gamma(t)$ of the curve (see Exercise 9.4.6). Hence, $\gamma'(t)$ is orthogonal to $\gamma(t)$ at each t . Since the tangent vector $T(t) = \gamma'(t)/\|\gamma'(t)\|$ is a scalar times $\gamma'(t)$, it is also orthogonal to the position vector $\gamma(t)$ for each t .

How smooth is a level set for a smooth function $F : U \rightarrow \mathbb{R}^d$? Does it have a tangent space at some or all of its points? If so, does it resemble a curved version of its tangent space?

By Definition 9.4.7, In order for a level set S for F to have a tangent space at a point $b \in S$, there must be a neighborhood of b in which S is a smoothly parameterized p -surface. That is, near b , S must be the image of a smooth function $G : V \rightarrow \mathbb{R}^q$, with V an open subset of \mathbb{R}^p , and the rank of dG equal to p (the maximal rank possible) at each $a \in V$. Then the image of the affine function $\Phi(x) = b + dG(a)(x - a)$ is a p dimensional affine subspace of \mathbb{R}^q (The tangent space to S at $b = G(a)$). Also, by the previous theorem

$$0 = dF(b)dG(a)(x - a) = dF(b)(\Phi(x) - b)$$

This means that the image of $\Phi - b$ is a linear subspace of $K = \ker dF(b)$. Hence, K has dimension at least p and it has dimension exactly p if and only if the image of $\Phi - b$ is equal to K . The dimension of K is p if and only if the rank of $dF(b)$ is $q - p$. Hence, we have proved:

Theorem 9.4.11. *With F as above and S a level set of F containing the point b , if in some neighborhood of b the space S is a smoothly parameterized p surface, and if $dF(b)$ has rank $q - p$ then the tangent space to S at b is the set of solutions y to the equation $dF(b)(y - b) = 0$. If the rank of $dF(b)$ is less than $q - p$, then the set of solutions to this equation contains the tangent space to S at b as a proper subset.*

Example 9.4.12. If $f(x, y, z) = x^2 + y^2 - z^2$ and $S = \{(x, y, z) : f(x, y, z) = 0\}$, show that at every point (a, b, c) on S , except at the origin, S is a smoothly parameterized 2-surface with tangent space defined in terms of the kernel of df

as in the previous theorem. Give the resulting equation for the tangent space. Then show that all of this fails at the origin.

Solution: The surface S is the same as the parameterized surface of Example 9.4.8 and Figure 9.2. By that example, S is a smoothly parameterized 2 surface near each such point except the origin. At $(a, b, c) \neq (0, 0, 0)$, df is $(2a, 2b, 2c)$. This has rank $1 = 3 - 2$. Therefore, by the previous theorem, S has a tangent space given by

$$2a(x - a) + 2b(y - b) + 2c(z - c) = 0.$$

At 0 df is the 0 matrix. Hence, the kernel of $df(0)$ is all of \mathbb{R}^3 . Since S is the cone of Example 9.4.8, it is a 2 dimensional surface and it does not seem reasonable for it to have a 3-dimensional tangent space at a point. The problem is that S is not a smoothly parameterized surface in a neighborhood of the origin and, hence, does not have a tangent space there in the sense we are using the term in this text.

When can a level set of a function $F : U \rightarrow \mathbb{R}^d$ be represented as a smoothly parameterized p -surface where $q - p$ is the rank of $dF(b)$? That is the subject of the implicit function theorem discussed in the last section of this chapter. At this point, it is not clear that a level set of a smooth function F has a smooth parameterization near any of its points.

For some level sets the construction of a smooth parameterization of the right dimension is easy. This is true of a level set which arises as the graph of a function, as the next example shows.

Example 9.4.13. Show that if g is a smooth real valued function defined on \mathbb{R}^2 , then each level set of the function $f : \mathbb{R}^3 \rightarrow \mathbb{R}$ defined by $f(x, y, z) = z - g(x, y)$ may be represented as a parameterized 2-surface.

Solution: Choose $G(x, y) = (x, y, g(x, y) + c)$. This is a smooth function from \mathbb{R}^2 to \mathbb{R}^3 with differential of rank 2 at each point and image equal to the level set $S = \{(x, y, z) : f(x, y, z) = c\}$.

Exercise Set 9.4

1. If $f(x, y, z) = x \sin z + y \cos z$ at each $(x, y, z) \in \mathbb{R}^3$, then find the gradient df of f at any point (x, y, z) . What is $df(1, 2, \pi/4)$?
2. For the function $f(x, y) = x^2 + y^3 + xy$, find the gradient at the point $(1, 1)$, the direction of greatest ascent of f at this point, and a direction in which the rate of increase of this function is 0 (the answers to the last two questions should be unit vectors).
3. Find a parametric equation for the tangent line to the curve

$$\gamma(t) = (t^3, 1/t, e^{2t-2})$$

at the point where $t = 1$.

4. For the curve γ of Example 9.4.6, find a parametric equation of the tangent line to this curve at $(0, 0)$ if the domain of $\gamma(t)$ is $\{t : \pi < t < 2\pi\}$.
5. Prove Theorem 9.4.4
6. Show that the gradient at $x \in \mathbb{R}^p$ of the function $g(x) = x \cdot x$ is the vector $2x$.
7. Let $\gamma : \mathbb{R} \rightarrow \mathbb{R}^p$ be a curve which passes through the origin in \mathbb{R}^p at a point where its velocity vector is non-zero (that is, assume $\gamma(t_0) = 0$ and $\gamma'(t_0) \neq 0$ at some point $t_0 \in \mathbb{R}$). Prove that there is an interval I centered at t_0 such that $\|\gamma(t)\|$ is decreasing for $t < t_0$ and increasing for $t > t_0$. Hint: $\|\gamma\|$ is increasing (decreasing) wherever $\|\gamma\|^2 = \gamma \cdot \gamma$ is increasing (decreasing).
8. Find the tangent space at $(2, 4, 1)$ for the parameterized surface in \mathbb{R}^3 parameterized by the function $G : U \rightarrow \mathbb{R}^3$, where

$$U = \{(u, v) \in \mathbb{R}^2 : u > 0, v > 0\} \quad \text{and} \quad G(u, v) = (uv, u^2, v^2).$$
9. If a surface in \mathbb{R}^3 is defined by the equation $z = g(x, y)$, where g is a differentiable function of (x, y) in an open set U , find the equation for the tangent plane to this surface at a point (a, b, c) on the surface.
10. Find an equation for the tangent plane to the surface $z = x^2 \sin y + 2x$ at the point $(1, 0, 2)$. Also find parametric equations for a line which passes through this point and is perpendicular to the tangent plane.
11. Find the equation for the tangent plane to the cone $z = x^2 + y^2$ at the point $(1, 2, 5)$.
12. Show that for each point (a, b, c) on the surface $x^2 + y^2 + z^2 = 1$, there is a neighborhood of (a, b, c) in which the surface may be represented as a smoothly parameterized 2-surface. Hence, there is a tangent plane to this surface at every point.
13. Find an equation for the tangent plane to the surface of the previous problem at each point (a, b, c) on the surface.
14. Find an equation for the tangent plane to the surface $x^2 + y^2 - z^2 = 1$ at each point (a, b, c) on the surface.

9.5 Taylor's Formula

In this section we discuss Taylor's formula in several variables and some of its applications.

The Formula

If a and x are points of \mathbb{R}^p , then a parameterized line passing through a and x is given by

$$\gamma(t) = a + t(x - a)$$

Note $\gamma(0) = a$ and $\gamma(1) = x$. The *line segment* joining a to x is the closed interval $[a, x]$ on this line defined by

$$[a, x] = \{a + t(x - a) : t \in [0, 1]\}.$$

Let f be a real valued function defined on an open subset $U \subset \mathbb{R}^p$ and suppose that all partial derivatives of f through degree n exist on U and are themselves differentiable on U . If $a, x \in U$ and the line segment joining a to x is contained in U , then we set $h = x - a$ and define a function g on an open interval I containing $[0, 1]$ by

$$g(t) = f(a + th).$$

The function g is $n + 1$ times differentiable on I (by the chain rule) and so g satisfies Taylor's Formula (Theorem 6.5.3):

$$g(t) = g(0) + g'(0)t + \frac{g''(0)}{2}t^2 + \cdots + \frac{g^{(n)}(0)}{n!}t^n + R_n(t), \quad (9.5.1)$$

Where

$$R_n(t) = \frac{g^{(n+1)}(s)}{(n+1)!}t^{n+1} \quad (9.5.2)$$

for some s between 0 and t .

Since $g(1) = f(a + h)$, to get a formula for $f(a + h)$ we need only set $t = 1$ in the above formula and then find expressions for the functions $g^k(0)$ and $g^{(n)}(c)$ in terms of f and its derivatives. This is not difficult for the first few terms:

$$\begin{aligned} g(0) &= f(a) \\ g'(0) &= df(a)h = \sum_{j=1}^p \frac{\partial f}{\partial x_j}(a)h_j \\ g''(0) &= h \cdot d^2f(a)h = \sum_{i=1}^p \sum_{j=1}^p \frac{\partial^2 f}{\partial x_i \partial x_j}(a)h_i h_j \end{aligned} \quad (9.5.3)$$

Here we have used $d^2f(a)$ to stand for the matrix

$$\left(\frac{\partial^2 f}{\partial x_i \partial x_j}(a) \right)_{ij}.$$

If we apply this matrix to h , the result is a vector of length p and we may take the inner product of h with this vector. The result is the formula for $g''(0)$ in (9.5.3).

The k th derivative of g at 0 is

$$g^{(k)}(0) = \sum_{i_1=1}^p \cdots \sum_{i_k=1}^p \frac{\partial^k f}{\partial x_{i_1} \cdots \partial x_{i_k}}(a) h_{i_1} \cdots h_{i_k}. \quad (9.5.4)$$

We may think of this as a k dimensional array (a tensor of rank k)

$$d^k f(a) = \left(\frac{\partial^k f}{\partial x_{i_1} \cdots \partial x_{i_k}}(a) \right),$$

applied k times to the vector h . Here applying a tensor of rank k to a vector h yields a tensor of rank $k - 1$ in the same way applying a matrix (tensor of rank 2) to a vector produces a vector (a tensor of rank 1). Thus, applying the tensor $d^k f(a)$ to the vector h produces the tensor of rank $k - 1$:

$$d^k f(a)h = \left(\sum_{i_k=1}^p \frac{\partial^k f}{\partial x_{i_1} \cdots \partial x_{i_k}}(a) h_{i_k} \right).$$

This has rank $k - 1$ because we have summed over the index i_k , and so the result is no longer a function of this index. If we repeat this k times, we obtain the number (tensor of rank 0) expressed in (9.5.4). This is the result of applying $d^k f(a)$ a total of k times to the vector h and, hence, we will denote it by $d^k f(a)h^k$. Note, in particular, that $d^2 f(a)h^2$ is just $h \cdot d^2 f(a)h$.

If we use this notation for the derivatives of g in (9.5.1) and (9.5.2) the result is :

$$f(a+h) = f(a) + df(a)h + \frac{1}{2}d^2 f(a)h^2 + \cdots + \frac{1}{n!}d^n f(a)h^n + R_n, \quad (9.5.5)$$

where

$$R_n = \frac{1}{(n+1)!}d^{n+1} f(c)h^{n+1}, \quad (9.5.6)$$

for some point c on the line segment joining a to $a+h$. This is Taylor's formula in several variables. Expressed in terms of the variable $x = a+h$ (so that $h = x - a$), this becomes the formula of the following theorem.

Theorem 9.5.1. *Let f be a real valued function defined on an open set $U \subset \mathbb{R}^p$ and suppose all partial derivatives of f through degree n exist and are differentiable on U . If $a, x \in U$ and U contains the line segment $[a, x]$, then*

$$f(x) = f(a) + df(a)(x-a) + \frac{1}{2}d^2 f(a)(x-a)^2 + \cdots + \frac{1}{n!}d^n f(a)(x-a)^n + R_n,$$

where

$$R_n = \frac{1}{(n+1)!}d^{n+1} f(c)(x-a)^{n+1},$$

for some point c on the line segment $[a, x]$.

Example 9.5.2. Find the degree $n = 2$ Taylor's formula for $f(x, y) = \ln(x + y)$ at the point $a = (0, 1)$.

Solution: We will need expressions for all partial derivatives of f through degree 3. However, these are easy to calculate because each n th order partial derivative of f is just the n th derivative of \ln evaluated at $x + y$. Thus, $f(0, 1) = 0$, all first order partial derivatives of f are $(x + y)^{-1}$, which is 1 at $(0, 1)$. The second degree partial derivatives are all equal to $-(x + y)^{-2}$, which is -1 at $(x, y) = (0, 1)$. Each third degree partial derivative is $2(x + y)^{-3}$. Thus, the degree 2 Taylor formula for f is

$$\begin{aligned} \ln(x + y) &= (1, 1) \begin{pmatrix} x \\ y - 1 \end{pmatrix} \\ &\quad - \frac{1}{2}(x, y - 1) \cdot \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} x \\ y - 1 \end{pmatrix} + R_2 \\ &= x + y - 1 - \frac{1}{2}(x + y - 1)^2 + R_2, \end{aligned}$$

where

$$R_2 = \frac{1}{3c^3}(x + y - 1)^3,$$

for some c between 1 and $x + y$. Here the expression in parentheses is the result of applying the rank three tensor which is 1 in every entry three times to the vector $(x, y - 1)$. The result is $(x + y - 1)^3$.

The Mean Value Theorem

The Mean Value Theorem for a real valued function on an open subset of \mathbb{R}^p is a special case of Taylor's formula. In fact, if we apply Theorem 9.5.1 in the case $n = 0$, it yields:

$$f(x) = f(a) + R_0,$$

where

$$R_0 = df(c)(x - a)$$

for some c on the line segment joining a to x . Thus, we have proved,

Theorem 9.5.3. *If f is a differentiable real valued function on $B_r(a) \subset \mathbb{R}^p$, then for $x \in B_r(a)$ we have*

$$f(x) - f(a) = df(c)(x - a)$$

for some point c on the line segment joining a to x .

As is the case for functions of one variable, the several variable mean value theorem has a host of applications. We point out two of these in the following corollaries, the proofs of which are left to the exercises.

Definition 9.5.4. A subset $A \subset \mathbb{R}^p$ is said to be *convex* if, for each pair of points $x, y \in A$, the line segment $[x, y]$ is also contained in A .

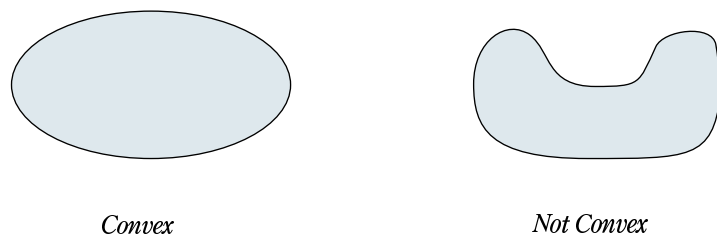


Figure 9.3: Convex and Nonconvex Sets

Figure 9.3 illustrates examples of a convex set and a set which is not convex.

Corollary 9.5.5. *Suppose U is an open convex set and f is a differentiable real valued function on U . If there is a number $M > 0$ such that $\|df(x)\| \leq M$ for all $x \in U$, then*

$$|f(x) - f(y)| \leq M\|x - y\|$$

for all $x, y \in U$.

Corollary 9.5.6. *Let U be a connected open subset of \mathbb{R}^p and f a differentiable function on U . If $df(x) = 0$ for all $x \in U$, then f is a constant.*

Max and Min

We know that if f is a real valued function of one variable, defined on an interval I , which has a local maximum or minimum at an interior point a of I , then either $f'(a)$ fails to exist or $f'(a) = 0$. We now discuss the several variable analogue of this result.

A function defined on a subset $D \subset \mathbb{R}^p$ is said to have a *local maximum* at $a \in D$ if there is a ball $B_r(a)$, centered at a , such that

$$f(x) \leq f(a) \quad \text{for all } x \in D \cap B_r(a).$$

If a is an interior point of D , then r may be chosen so that $D_r(a) \subset D$ and then this inequality holds for all $x \in B_r(a)$. The concept of local minimum is defined in the same way, but with the inequality reversed.

Theorem 9.5.7. *If f is a function defined on $D \subset \mathbb{R}^n$ and if f has a local maximum or a local minimum at an interior point $a \in D$ at which f is differentiable, then $df(a) = 0$.*

Proof. Given any unit vector u , the function $g(t) = f(a + tu)$ is defined for all real numbers t in an open interval containing 0 and it has a local maximum (or minimum) at $t = 0$. By the chain rule, g is differentiable at 0 and its derivative at 0 is the directional derivative $df(a) \cdot u$ of f at a in the direction u . Since the derivative of g at 0 must be 0, we conclude that $df(a) \cdot u = 0$ for all unit vectors u and, hence, $df(a) = 0$. \square

This theorem does not tell us that a function must have a local max or min at a point where df is 0. However, for functions of one variable, the second derivative test does give conditions that ensure that a local max or a local min occurs at a .

The second derivative test for functions of one variable says that if f is a real valued function on an interval I , then f has a local maximum at $a \in I$ if $f'(a) = 0$ and $f''(a) < 0$. It has a local minimum at a if $f'(a) = 0$ and $f''(a) > 0$. The analogue of this in several variables will be presented below, but it requires the concept of a *positive definite* matrix.

Definition 9.5.8. A $p \times p$ matrix A is said to be *positive definite* if $h \cdot Ah > 0$ for every non-zero vector $h \in \mathbb{R}^p$. It is *negative definite* if $h \cdot Ah < 0$ for every non-zero vector $h \in \mathbb{R}^p$.

Note that, in checking to see if a matrix is positive definite, we only need to check that $u \cdot Au > 0$ for every unit vector u in \mathbb{R}^p . This is because, if h is any non-zero vector, then $u = h/\|h\|$ is a unit vector and $h \cdot Ah = \|h\|^2 u \cdot Au$, which is positive if and only if $u \cdot Au$ is positive.

It turns out that if a matrix is positive definite, then all nearby matrices are also positive definite. We will prove this using the concept of *operator norm* for a matrix (Definition 8.4.9). Recall that $\|Ax\| \leq \|A\|\|x\|$ if x is a vector in \mathbb{R}^p , A is a $p \times p$ matrix, and $\|A\|$ is the operator norm of A .

Lemma 9.5.9. *If A is a positive definite $p \times p$ matrix, then there is a positive number m such that if B is any $p \times p$ matrix with $\|B - A\| < m/2$, then $u \cdot Bu \geq m/2$ for all unit vectors $u \in \mathbb{R}^p$ and, hence, B is also positive definite.*

Proof. The set of all unit vectors u is a closed bounded subset of \mathbb{R}^p . It is, therefore, compact. The function $g(u) = u \cdot Au$ is a continuous real valued function on this set and, hence, by Corollary 8.2.5, it takes on a minimum value m . Since $u \cdot Au > 0$ for all such u , we conclude that $m > 0$. Now it follows from the Cauchy-Schwarz inequality that

$$u \cdot (A - B)u \leq \|u\| \|(A - B)u\| \leq \|u\|^2 \|A - B\| = \|A - B\|.$$

This implies

$$u \cdot Bu = u \cdot Au - u \cdot (A - B)u \geq m - \|A - B\| \quad (9.5.7)$$

for all unit vectors u . Hence, if $\|A - B\| < m/2$, then $u \cdot Bu > m/2$ for all unit vectors u , which implies that B is positive definite. \square

Theorem 9.5.10. *Let f be a real valued function defined on a neighborhood of $a \in \mathbb{R}^p$. Suppose the second order partial derivatives of f exist in this neighborhood and are continuous at a . If $df(a) = 0$ and $d^2f(a)$ is positive definite, then f has a local minimum at a . If $df(a) = 0$ and $d^2f(a)$ is negative definite, then f has a local maximum at a .*

Proof. We use Taylor's formula with $n = 1$. Since, $df(a) = 0$, It tells us that there is an $r > 0$ such that, for each $h \in B_r(0)$,

$$f(a+h) = f(a) + h \cdot d^2f(c)h, \quad (9.5.8)$$

for some c on the line segment joining a to $a+h$.

Assume $d^2f(a)$ is positive definite. By the previous lemma, there is an $m > 0$ such that if

$$\|d^2f(a) - d^2f(c)\| < m/2, \quad (9.5.9)$$

then $d^2f(c)$ is also positive definite.

Since the second order partial derivatives of f are continuous at a and since $\|c-a\| \leq \|h\|$, it follows from Theorem 8.4.11 that we can ensure (9.5.9) holds by choosing $\|h\|$ sufficiently small. Hence, there is an $\delta > 0$, with $\delta \leq r$, such that $\|h\| < \delta$ implies that $d^2f(c)$ is positive definite for all c on the line segment joining a to h . By 9.5.8, this implies that $f(a+h) > f(a)$. Thus, f has a local minimum at a in this case.

The case where $d^2f(a)$ is negative definite follows from the above by simply applying the above result to $-f$. \square

Max/Min for Functions of 2 Variables

Let f be a function of 2 variables with second order partial derivatives which are defined in a neighborhood of $(x_0, y_0) \in \mathbb{R}^2$ and continuous at this point. The matrix d^2f has the form

$$\begin{pmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial y \partial x} & \frac{\partial^2 f}{\partial y^2} \end{pmatrix}.$$

Since the second order partial derivatives are continuous at (x_0, y_0) , the cross partials are equal and so this matrix is symmetric (meaning it is its own transpose) at (x_0, y_0) . There is a simple criteria for a symmetric 2×2 matrix to be positive definite. This is described in the next theorem, the proof of which is left to the exercises.

Theorem 9.5.11. Let $A = \begin{pmatrix} a & b \\ b & c \end{pmatrix}$ be a symmetric 2×2 matrix and let $\Delta = ac - b^2$ be its determinant. Then

- (a) A is positive definite if and only if $\Delta > 0$ and $a > 0$;
- (b) A is negative definite if and only if $\Delta > 0$ and $a < 0$;
- (c) if $\Delta < 0$, then there are vectors $u, v \in \mathbb{R}^2$ with $u \cdot Au > 0$ and $v \cdot Av < 0$.

For a function f on \mathbb{R}^2 , a point where the expression $u \cdot d^2f(a)u$ is positive for some unit vectors u and negative for others is called a *saddle point*. At such

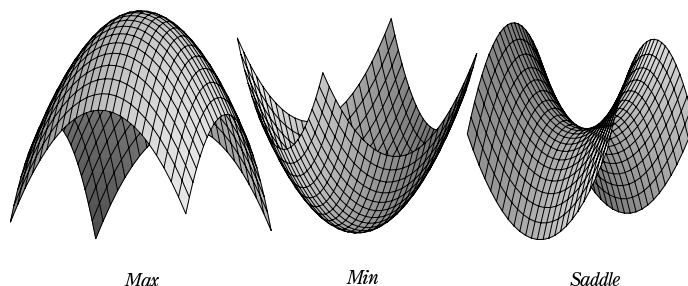


Figure 9.4: Surfaces With Max, Min, and Saddle Points.

a point, there will exist lines through a along which f has a local maximum at a and other lines through a along which f has a local minimum at a .

The previous theorem has the following corollary, the proof of which is also left to the exercises.

Corollary 9.5.12. *Let f be a function of 2 variables with second order partial derivatives which are defined in a neighborhood of $(x_0, y_0) \in \mathbb{R}^2$ and continuous at this point. Let $\Delta = \frac{\partial^2 f}{\partial x^2} \frac{\partial^2 f}{\partial y^2} - \left(\frac{\partial^2 f}{\partial x \partial y} \right)^2$ evaluated at (x_0, y_0) . Then*

- (a) f has a local minimum at (x_0, y_0) if $\Delta > 0$ and $\frac{\partial^2 f}{\partial x^2} > 0$ at (x_0, y_0) ;
- (b) f has a local maximum at (x_0, y_0) if $\Delta > 0$ and $\frac{\partial^2 f}{\partial x^2} < 0$ at (x_0, y_0) ;
- (c) if $\Delta < 0$, then f has a saddle point at x_0, y_0 .

Example 9.5.13. Find all points where the function $f(x, y) = x^2 + xy + y^2 - 2x - 4y + 1$ has a local maximum and all points where it has a local minimum.

Solution: We have $df(x, y) = (2x + y - 2, x + 2y - 4)$. Thus, the only point at which $df(x, y) = 0$ is the point $a = (0, 2)$. This is the only possible point at which a local max or min can occur. The second differential $d^2f(x, y)$ is the constant matrix

$$d^2f(x, y) = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}.$$

This has determinant $\Delta = 3$. By the previous corollary, we conclude that $(0, 2)$ is a point at which a local minimum occurs and there is no local maximum.

Example 9.5.14. Find all points where the function

$$f(x, y) = x^2 + 3xy + y^2 - x - 4y + 5$$

has a local maximum, minimum, or saddle.

Solution: We have $df(x, y) = (2x + 3y - 1, 3x + 2y - 4)$. Thus, the only point at which $df(x, y) = 0$ is the point $a = (2, -1)$. This is the only possible point at which a local max or min can occur. The second differential $d^2f(x, y)$ is the constant matrix

$$d^2f(x, y) = \begin{pmatrix} 2 & 3 \\ 3 & 2 \end{pmatrix}.$$

This has determinant $\Delta = -5$. Thus, $(2, -1)$ is a saddle point for f .

Lagrange Multipliers

Suppose U is an open subset of \mathbb{R}^q and $f : U \rightarrow \mathbb{R}$ and $G : U \rightarrow \mathbb{R}^d$ are differentiable functions. The subject of Lagrange multipliers concerns the problem of finding points of local maximum or local minimum of f subject to the constraint that $G(x) = 0$. That is, we wish to find the points of local maximum and local minimum of f considered as a function on the level set $G(x) = 0$ for G . The following theorem applies to this problem. Its proof uses a corollary of the Implicit Function Theorem which will be proved at the end of this chapter.

Theorem 9.5.15. *With U , F and G as above, suppose that dG has rank d on U and S is the level set $S = \{x \in U : G(x) = 0\}$. If b is a point of relative max or min for f on S , then there is a linear transformation $\Lambda : \mathbb{R}^d \rightarrow \mathbb{R}$ such that $df(b) = \Lambda dG(b)$.*

Proof. In Corollary 9.7.3 we will prove that, under the above conditions, S is a smoothly parameterized p surface in a neighborhood of each point of S . We will assume this result here and we may as well assume that U is the neighborhood. Then S is the image of a one-to-one differentiable function $H : V \rightarrow U$ with $\text{Rank } dH = p$ on V . Furthermore, $dG(H(a)) \circ dH(a) = 0$ for each $a \in V$. Thus, if $a \in V$ and $b = H(a)$, then the kernel of $dG(b)$ contains the image of $dH(a)$. However, the kernel of $dG(b)$ has dimension $q - d = p$ as does the image of $dH(a)$. It follows that the two subspaces of \mathbb{R}^q are equal.

Since f has a local max or min on S at b , $f \circ H$ has a local maximum on V at a . This implies $df(b)dH(a) = d(f \circ H) = 0$. Since $dG(b)$ has rank d , its image is all of \mathbb{R}^d . Thus, for each $y \in \mathbb{R}^d$, there is an $x \in \mathbb{R}^q$ such that $dG(b)x = y$. We then set $\Lambda(y) = df(x)$. If x_1 is another vector in \mathbb{R}^q with $dG(b)x_1 = y$, then $x - x_1 \in \ker dG(b) = \Im dH(b)$ and so $df(b)(x - x_1) = 0$. This means $df(b)x$ is the same vector no matter which vector x is chosen with $dG(b)x = y$. Thus, $\Lambda(y)$ is well defined by the condition

$$\Lambda(y) = df(x) \quad \text{whenever} \quad dG(b)x = y. \quad (9.5.10)$$

For vectors $y_1, y_2 \in \mathbb{R}^d$ we may choose x_1, x_2 such that $dG(b)x_i = y_i$. Then, $dG(b)(x_1 + x_2) = dG(b)x_1 + dG(b)x_2 = y_1 + y_2$ and so

$$\Lambda(y_1 + y_2) = df(x_1 + x_2) = df(x_1) + df(x_2) = \Lambda(y_1) + \Lambda(y_2).$$

A similar argument shows that $\Lambda(kx) = k\Lambda(x)$ if k is a scalar. Thus, Λ is a linear transformation. By (9.5.10) Λ satisfies $df(b) = \Lambda dG(b)$. \square

The above result looks less mysterious if we write it out in terms of the coordinate functions of G . If $G = (g_1, \dots, g_d)$, then S is the surface of vectors $x \in \mathbb{R}^q$ which satisfy the constraints

$$g_1(x) = 0, \dots, g_d(x) = 0. \quad (9.5.11)$$

The theorem says that, if b is a point of S on which f has a local max or min on S , then there is a vector $\Lambda = (\lambda_1, \dots, \lambda_d)$ such that

$$\frac{\partial f}{\partial x_k}(b) = \sum_{j=1}^d \lambda_j \frac{\partial f_j}{\partial x_k}(b) \quad \text{for } k = 1, \dots, q. \quad (9.5.12)$$

Thus, to find candidates for points on S where a local max or min could occur, one should simultaneously solve the equations (9.5.11) and (9.5.12) for $x_1, \dots, x_q, \lambda_1, \dots, \lambda_d$. Note, this system of equations has $d + q$ equations and $d + q$ unknowns. The components $\lambda_1, \dots, \lambda_d$ of Λ are called *Lagrange multipliers*.

Example 9.5.16. Find where the function $f(x, y, z) = 2xy + z$ attains its maximum and minimum values on $S = \{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 + z^2 = 1\}$.

Solution: Since the unit sphere in \mathbb{R}^3 is compact and f is continuous, there are points on S where f attains its maximum and minimum values. We use the method of Lagrange multipliers, as described in the previous theorem to obtain candidates for these points. Here, $d = 1$ and $q = 3$ in (9.5.11) and (9.5.12).

With $g(x, y, z) = x^2 + y^2 + z^2 - 1$, we must solve the system of equations:

$$g(x, y, z) = 0, \quad \frac{\partial f}{\partial x} = \lambda \frac{\partial g}{\partial x}, \quad \frac{\partial f}{\partial y} = \lambda \frac{\partial g}{\partial y}, \quad \frac{\partial f}{\partial z} = \lambda \frac{\partial g}{\partial z}.$$

These are the equations

$$x^2 + y^2 + z^2 = 1, \quad 2x = 2\lambda y, \quad 2y = 2\lambda x, \quad 1 = 2\lambda z.$$

The second and third equations yield $x = \lambda^2 x$ and $y = \lambda^2 y$. These hold if and only if $x = y = 0$ or $\lambda = \pm 1$. But $\lambda = \pm 1$ implies $x = \pm y$ and, together with the fourth equation, implies $z = \pm 1/2$. This, and the first equation imply $x = \pm\sqrt{3/8}$, $y = \pm\sqrt{3/8}$. Thus, the solutions of the above system of equations are

$$(0, 0, 1), \quad (\sqrt{3/8}, \sqrt{3/8}, 1/2), \quad (-\sqrt{3/8}, -\sqrt{3/8}, 1/2), \\ (-\sqrt{3/8}, \sqrt{3/8}, -1/2), \quad \text{and} \quad (\sqrt{3/8}, -\sqrt{3/8}, -1/2).$$

The values of f at these five points are, respectively, $1, 5/4, 5/4, -5/4, -5/4$. Thus, f has maximum value $5/4$ on S which is attained at $(\sqrt{3/8}, \sqrt{3/8}, 1/2)$, and at $(-\sqrt{3/8}, -\sqrt{3/8}, 1/2)$, while the minimum value is $-5/4$ attained at $(-\sqrt{3/8}, \sqrt{3/8}, -1/2)$ and $(\sqrt{3/8}, -\sqrt{3/8}, -1/2)$.

Exercise Set 9.5

1. Find the degree $n = 2$ Taylor's formula for $f(x, y) = x^2 + xy$ at the point $a = (1, 2)$.
2. Find the degree $n = 2$ Taylor's formula for $f(x, y) = e^{xy}$ at the point $a = (0, 0)$.
3. Suppose $a \in \mathbb{R}^p$ and f is a real valued function whose second order partial derivatives all exist and are continuous on $B_r(a)$. Also, suppose that the operator norm $\|d^2 f(x)\|$ of the matrix $d^2 f(x)$ is bounded by M on $B_r(a)$. Prove that

$$|f(x) - f(a) - df(a)(x - a)| \leq M\|x - a\|^2$$

for all $x \in B_r(a)$.

4. Prove Corollary 9.5.5.
5. Prove Corollary 9.5.6.
6. Show that the following form of the Mean Value Theorem is not true: If $F : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ is a differentiable function and $a, b \in \mathbb{R}^2$, then there is a c on the line segment joining a to b such that $F(b) - F(a) = dF(c)(b - a)$. The problem here is that F is vector valued, not real valued.
7. Show that the following version of the Mean Value Theorem for vector valued functions is true: If U is an open set in \mathbb{R}^p containing the line segment joining a to b and if $F : U \rightarrow \mathbb{R}^q$ is a differentiable function on U , then, for each vector $u \in \mathbb{R}^q$, there is a point c on the line segment joining a to b such that

$$u \cdot (F(b) - F(a)) = u \cdot dF(c)(b - a).$$

8. Find all points of relative maximum and relative minimum and all saddle points for
9. Find all points of relative maximum and relative minimum and all saddle points for

$$f(x, y) = 1 - 2x^2 - 2xy - y^2.$$

$$f(x, y) = y^3 + y^2 + x^2 - 2xy - 3y.$$

10. Prove Theorem 9.5.11.
11. Prove Corollary 9.5.12.
12. Show that it is possible for a function to have a relative minimum or maximum or a saddle at a point where both df and $d^2 f$ are 0.
13. Use the Lagrange multiplier method to find the maximal and minimal values of $f(x, y, z) = x - 2y + 3z$ on the sphere $x^2 + y^2 + z^2 = 1$.

9.6 The Inverse Function Theorem

If f is a real valued function of one variable which is \mathcal{C}^1 on an open interval containing a and if $f'(a) \neq 0$, then $f'(a)$ is either positive or negative. Because f' is continuous, $f'(x)$ will have the same sign as $f'(a)$ for all x in some neighborhood of a . This implies that f is strictly monotone in a neighborhood of a and, hence, has an inverse function. This inverse function is differentiable at $b = f(a)$ and $(f^{-1})'(b) = 1/f'(a)$ (Theorem 4.2.9). In this section we will prove an analogous result for a vector valued function F of several variables.

The condition that $f'(a) \neq 0$ is replaced in several variables by the condition that $dF(a)$ is a non-singular matrix (a matrix for which there is an inverse matrix). The conclusion that f is strictly monotone in some open interval containing a is replaced by the conclusion that F is a one to one function in some neighborhood of a in \mathbb{R}^p .

A function $F : V \rightarrow W$ is *one to one* on V if $x = y$ whenever $x, y \in V$ and $F(x) = F(y)$. It is *onto* W if every $u \in W$ is $F(x)$ for some $x \in V$.

Definition 9.6.1. With F as above, we will say that F has a smooth local inverse near a if there are neighborhoods V of a and W of $F(a)$ such that F is a one to one function from V onto W and the function $F^{-1} : W \rightarrow V$, defined by $F^{-1}(u) = x$ if $F(x) = u$, is smooth on W .

In what follows (until the proof of the Inverse Function Theorem is complete), U will be an open subset of \mathbb{R}^p , $F : U \rightarrow \mathbb{R}^p$ a smooth (that is, \mathcal{C}^1) function on U . We will prove that F has a smooth local inverse near any point $a \in U$ at which its differential is non-singular.

The proof involves three steps: (1) we show F is one-to-one in a neighborhood of a ; (2) we show F maps this neighborhood onto an open set; (3) we show the resulting inverse function is smooth and we calculate its differential.

One to One

The next theorem shows that our function F is necessarily one to one on some open ball centered at a point where dF is non-singular. In fact, it shows much more.

Theorem 9.6.2. *If $a \in U$ and $dF(a)$ is non-singular, then there is an open ball $B_r(a)$, centered at a , and a positive number M such that :*

(a) *the matrix $dF(x)$ is non-singular for all $x \in B_r(a)$;*

(b) *$\|x - y\| \leq M\|F(x) - F(y)\|$ for all $x, y \in B_r(a)$,*

(c) *the function F is one to one on $B_r(a)$.*

Proof. Let B be an inverse matrix for $dF(a)$. Then $d(BF)(a) = BdF(a) = I$, where I is the $p \times p$ identity matrix (Exercise 9.3.1).

Let $G(x) = BF(x)$. Note that $dG(a) = I$, which is positive definite (since $u \cdot Iu = \|u\|^2 = 1$ for every unit vector $u \in \mathbb{R}^p$). Hence, by Lemma 9.5.9, there is an $m > 0$ such that $dG(x)$ is also positive definite and, in fact,

$$m/2 \leq u \cdot dG(x)u \quad \text{whenever} \quad \|dG(x) - dG(a)\| < m/2$$

and u is a unit vector in \mathbb{R}^p .

The partial derivatives of the coordinate functions of F are all continuous and so the same thing is true of G . It follows from Theorem 8.4.11 that, given $m > 0$, there is an r such that $B_r(a) \subset U$ and

$$\|dG(x) - dG(a)\| < m/2 \quad \text{whenever} \quad \|x - a\| < r.$$

Thus,

$$u \cdot dG(x)u \geq m/2 \tag{9.6.1}$$

for all $x \in B_r(a)$ and all unit vectors $u \in \mathbb{R}^p$. In particular, $dG(x)$ is positive definite and, hence, non-singular, for all $x \in B_r(a)$. Since $dF(x) = B^{-1}dG(x)$, this matrix is also non-singular for all $x \in B_r(a)$. This proves part (a).

Given two distinct points $x, y \in B_r(a)$, we set $k = \|y - x\| \neq 0$ and $u = (y - x)/k$. Then u is a unit vector and the function ϕ , defined by,

$$\phi(t) = u \cdot G(x + tu).$$

is a real valued differentiable function on an open interval containing $[0, k]$.

By the Mean Value Theorem, there is an $s \in [0, k]$ at which

$$k\phi'(s) = \phi(k) - \phi(0).$$

By the chain rule, $k\phi'(s) = ku \cdot dG(x + su)u$ and $\phi(k) - \phi(0) = u \cdot (G(y) - G(x))$. Thus,

$$ku \cdot dG(c)u = u \cdot (G(y) - G(x)),$$

where $c = x + su$. Then, by (9.6.1),

$$\begin{aligned} mk/2 \leq ku \cdot dG(c)u &= u \cdot (G(y) - G(x)) \\ &\leq \|u\| \|G(y) - G(x)\| \leq \|B\| \|F(y) - F(x)\|, \end{aligned} \tag{9.6.2}$$

which, since $k = \|y - x\|$, implies

$$\|y - x\| \leq \frac{2\|B\|}{m} \|F(x) - F(y)\|.$$

This concludes the proof of part (b) if we set $M = 2\|B\|/m$.

Part (c) – that F is one to one on $B_r(a)$ – follows immediately from part (b) which shows that, for $x, y \in B_r(a)$, $x = y$ whenever $F(x) = F(y)$. \square

Open Mapping Theorem

An open map is a function F such that $F(V)$ is open whenever V is open.

Theorem 9.6.3. *With F as above, if dF is non-singular at every point of an open subset V of U , then $F : V \rightarrow \mathbb{R}^p$ is an open map.*

Proof. Given $a \in V$, set $b = F(a)$. We will show that $F(V)$ contains an open ball centered at b . If we can do this for every $a \in V$, then $F(V)$ is open. The same argument can be applied to every open subset of V and, hence, we may conclude that F is an open map.

The fact that $dF(a)$ is non-singular implies there is a open ball $B_r(a) \subset V$ for which the conclusions of the previous theorem hold. We will show that the image of this ball contains an open ball $B_\delta(b)$

Let r_1 be a positive number less than r . Then part (b) of the previous theorem implies that there is a positive number M such that

$$\|x - y\| \leq M\|F(x) - F(y)\| \quad \text{for all } x, y \in \overline{B}_{r_1}(a).$$

Since $b = F(a)$, this implies, in particular, that

$$\|F(x) - b\| \geq \frac{r_1}{M} \quad \text{whenever } \|x - a\| = r_1. \quad (9.6.3)$$

We set $\delta = \frac{r_1}{2M}$ and let v be any element of $B_\delta(b)$. If

$$g(x) = \|F(x) - v\| \quad \text{for } x \in \overline{B}_{r_1}(a),$$

then our objective is to show that $g(u) = 0$ for some u in this ball.

We will first show that g takes on its minimum value at an interior point of $\overline{B}_{r_1}(a)$. It does take on a minimum value, since g is a continuous function on the compact set $\overline{B}_{r_1}(a)$ (Corollary 8.2.5). Thus, we need to show that it does not take on this minimum at a boundary point of \overline{B}_{r_1} .

If x is a boundary point of \overline{B}_{r_1} , then $\|x - a\| = r_1$ and (9.6.3) applies. Also, $v \in B_\delta(b)$ means $\|b - v\| < \frac{r_1}{2M}$. Thus,

$$g(x) = \|F(x) - v\| \geq \|F(x) - b\| - \|b - v\| \geq \frac{r_1}{2M} = \delta$$

on the boundary of \overline{B}_{r_1} .

Since $g(a) = \|F(a) - v\| = \|b - v\| < \delta$, the function $g(x)$ does not achieve its minimum value on the boundary of $\overline{B}_{r_1}(a)$. Hence, it must achieve its minimum value at a point u in the open ball $B_{r_1}(a)$. Then $g^2(x) = (F(x) - v) \cdot (F(x) - v)$ has a local minimum at u and, hence, its differential vanishes at u , by Theorem 9.5.7. By Theorem 9.3.6, its differential is $2(F(x) - v)dF(x)$. This expression vanishes at u if and only if $F(u) - v$ is orthogonal to all the columns of $dF(u)$. Since $dF(u)$ is non-singular, by Theorem 9.6.2 part (a), this can happen only if $F(u) - v = 0$. Hence, we have shown that each $v \in B_\delta(b)$ is the image under F of some $u \in B_r(a)$, as required. \square

The Inverse Function and its Differential

With F as above, if F is one-to-one with a non-singular differential on an open subset V of U then $\phi(V) = W$ is also open, by the previous theorem. In this situation, F has an inverse function $F^{-1} : W \rightarrow V$ defined by the condition that, for each $y \in W$, $F^{-1}(y)$ is the unique $x \in V$ such that $F(x) = y$.

Theorem 9.6.4. *With F , V and W as above, the inverse function $F : W \rightarrow V$ is a smooth function on W with differential given by*

$$dF^{-1}(b) = (dF(a))^{-1} = (dF(F^{-1}(b)))^{-1} \quad (9.6.4)$$

for each $b \in W$. Here $a = F^{-1}(b) \in V$.

Proof. For a , we choose r as in Theorem 9.6.2 and we choose it small enough that $B_r(a) \subset V$. Then $F(B_r(a))$ is also open, by the previous theorem.

If $y \in F(B_r(a))$ and $x = F^{-1}(y)$, then $x \in B_r(a)$. By the choice of r , the inequality in part (b) of Theorem 9.6.2 holds for x and a and says that

$$\|F^{-1}(x) - F^{-1}(a)\| = \|x - a\| \leq M\|y - b\|.$$

This implies that F^{-1} is continuous at b . We calculate the differential of F^{-1} at b as follows:

The fact that F is differentiable at a means that if we set

$$\epsilon(x) = F(x) - F(a) - dF(a)(x - a), \quad (9.6.5)$$

then

$$\lim_{x \rightarrow a} \frac{\epsilon(x)}{\|x - a\|} = 0.$$

If we apply the matrix $(dF(a))^{-1}$ to both sides of (9.6.5) and use $a = F^{-1}(b)$, $x = F^{-1}(y)$, the result is

$$dF(a)^{-1}\epsilon(y) = (dF(a))^{-1}(y - b) - (F^{-1}(y) - F^{-1}(b)),$$

or

$$F^{-1}(y) - F^{-1}(b) - dF(a)^{-1}(y - b) = -dF(a)^{-1}\epsilon(x).$$

If we set $K = \|(dF(a))^{-1}\|$, then

$$\frac{\|F^{-1}(y) - F^{-1}(b) - (dF(a))^{-1}(y - b)\|}{\|y - b\|} \leq \frac{K\|\epsilon(x)\|}{\|y - b\|} \leq \frac{KM\|\epsilon(x)\|}{\|x - a\|}.$$

Since F^{-1} is continuous at b , $x = F^{-1}(y)$ approaches $a = F^{-1}(b)$ as y approaches b . Thus, the right side of the above inequality approaches 0 as $y \rightarrow b$. By definition, this means that F^{-1} is differentiable at b and

$$dF^{-1}(b) = (dF(a))^{-1} = (dF(F^{-1}(b)))^{-1}.$$

The partial derivatives of the coordinate functions of F^{-1} are the entries of its differential matrix dF^{-1} , which we just concluded is given by (9.6.4). Since,

F^{-1} is continuous on W , the entries of $dF(x)$ (the partial derivatives of the coordinate functions of F) are continuous on V , and the determinant of $dF(x)$ is continuous and non-vanishing on V , we conclude that the partial derivatives of the coordinate functions of F^{-1} are continuous on W . This means that F^{-1} is \mathcal{C}^1 , as claimed. This completes the proof. \square

The Inverse Function Theorem

The proof of the Inverse Function Theorem is now just a matter of combining the previous three theorems.

Theorem 9.6.5. *Let U be an open subset of \mathbb{R}^p and $F : U \rightarrow \mathbb{R}^p$ a smooth function. If $a \in U$ and $\det dF(a) \neq 0$, then F has a smooth local inverse function near a , with differential given by (9.6.4).*

Proof. By Theorem 9.6.2, F is one-to-one with a non-singular differential in an open ball $B_r(a)$. By Theorem 9.6.3, the image of $B_r(a)$ under F is an open set W . Then F has an inverse function $F^{-1} : W \rightarrow B_r(a)$ and, by Theorem 9.6.4, the inverse function is smooth with differential as claimed. \square

Example 9.6.6. Find all points $a = (r, \theta) \in \mathbb{R}^2$ such that the polar change of coordinates function

$$F(r, \theta) = (r \cos \theta, r \sin \theta)$$

has a smooth local inverse near a . Find the inverse and its differential near one such point

Solution: The differential of F is

$$dF(r, \theta) = \begin{pmatrix} \cos \theta & -r \sin \theta \\ \sin \theta & r \cos \theta \end{pmatrix}.$$

The determinant of this matrix is r , and so dF is non-singular everywhere except at $r = 0$. By the previous theorem, this implies that F has a smooth local inverse near each $a = (r, \theta)$ with $r \neq 0$.

If we choose the point $a = (1, 0)$, then $F(a) = (1, 0)$. If V is the neighborhood of a defined by

$$V = \{(r, \theta) : r > 0, -\pi/2 < \theta < \pi/2\},$$

and W is the neighborhood of $b = F(a)$ defined by

$$W = \{(x, y) : x > 0\},$$

then

$$F^{-1}(x, y) = \left(\sqrt{x^2 + y^2}, \tan^{-1}(y/x) \right) \quad (9.6.6)$$

defines the inverse function $F^{-1} : W \rightarrow V$.

The inverse matrix $(dF(r, \theta))^{-1}$ of the differential matrix $dF(r, \theta)$ of F is

$$\begin{pmatrix} \cos \theta & \sin \theta \\ -r^{-1} \sin \theta & r^{-1} \cos \theta \end{pmatrix}.$$

By the previous theorem, this is the differential of the inverse function F^{-1} at the point $(x, y) = F(r, \theta)$. If express r and θ in terms of x and y using (9.6.6), we obtain

$$dF^{-1}(x, y) = \begin{pmatrix} \frac{x}{\sqrt{x^2 + y^2}} & \frac{y}{\sqrt{x^2 + y^2}} \\ -\frac{y}{x^2 + y^2} & \frac{x}{x^2 + y^2} \end{pmatrix}.$$

Note that the function F of the above example is definitely not one to one on all of \mathbb{R}^2 or on $\{(r, \theta) \in \mathbb{R}^2 : r \neq 0\}$ and so, as a function with either of these sets as domain, it does not have an inverse function. It is only when we restrict the domain of F to a set like the set V in the above example that it has an inverse function. What are some other sets V with the property that the restriction of F to the set V has an inverse function? This question is left to the exercises.

Exercise Set 9.6

1. According to the Inverse Function Theorem, near which points of \mathbb{R} does the sin function have a smooth local inverse function? According to this theorem, what is the derivative of the inverse function when it exists?
2. Show that the function $F : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ defined by $F(x, y) = (x^2 + y^2, xy)$ has a smooth local inverse near points (x, y) where $x \neq \pm y$. On the set $\{(x, y) : -x < y < x\}$ find the inverse function F^{-1} and identify its domain. Calculate the differential of this inverse function (1) directly, and (2) by using the Inverse Function Theorem. Verify that the two methods give the same answer.
3. Near which points of \mathbb{R}^3 does the spherical change of coordinates function

$$F(\rho, \theta, \phi) = (\rho \cos \theta \sin \phi, \rho \sin \theta \sin \phi, \rho \cos \phi)$$

have a smooth local inverse? What is the differential of the local inverse at those points, where it exists? To avoid tedious computation, express this in terms of (r, θ, ϕ) rather than in terms of the image variables $(x, y, z) = F(r, \theta, \phi)$.

4. Show that the system of equations

$$\begin{aligned} x &= u^4 - u + uv + v^2 \\ y &= \cos u + \sin v \end{aligned}$$

can be solved for (u, v) as a smooth function F of (x, y) , in some neighborhood of $(0, 0)$, in such a way that $(u, v) = (0, 0)$ when $(x, y) = (0, 1)$. What is the differential of the resulting function F at $(0, 1)$?

5. Find a smooth local inverse function near $(1, \pi/2)$ for the function F of Example 9.6.6.

6. Find a smooth local inverse function near $(1, 2\pi)$ for the function F of Example 9.6.6. Note that this is different from the inverse function found in the example, even though the point $b = F(a)$ is the same in both cases.
7. Show that if U is a convex open subset of \mathbb{R}^p and $F : U \rightarrow \mathbb{R}^p$ is a \mathcal{C}^1 function on U with a differential dF which is positive definite at every point of U , then F is one to one. Hint: examine the role played by the function ϕ in the proof of Theorem 9.6.2.
8. Show by example that the result of the previous problem is not true if U is only assumed to be connected, rather than convex. Hint: try the function $F(x, y) = (x^2 - y^2, 2xy)$ on $\mathbb{R}^2 \setminus \{0\}$.
9. Show that if $F = (f_1, f_2) : \mathbb{R}^3 \rightarrow \mathbb{R}^2$ is a \mathcal{C}^1 function and a is a point of \mathbb{R}^3 at which dF has rank 2, then there is a \mathcal{C}^1 function $f_3 : \mathbb{R}^3 \rightarrow \mathbb{R}$ such that $\Phi = (f_1, f_2, f_3) : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ has a \mathcal{C}^1 inverse function near a .
10. Show that the condition that $dF(a)$ be non-singular is necessary in the Inverse Function Theorem, by showing that if a function F from a neighborhood of a in \mathbb{R}^p to \mathbb{R}^p is differentiable at a and has an inverse function at a which is differentiable at $F(a)$, then $dF(a)$ is non-singular.
11. Let $\gamma : I \rightarrow \mathbb{R}^3$ be a smooth parameterized curve, defined on an open interval I , and let t_0 be a point of I with $\gamma'(t_0) \neq 0$. Prove that there are neighborhoods $U \subset I$ of t_0 and V of $\gamma(t_0)$ and a pair f, g of \mathcal{C}^1 functions defined in V such that the image of U under γ is the set of solutions in V of the system of equations $f(x, y, z) = 0, g(x, y, z) = 0$. Hint: show that there is a \mathcal{C}^1 function F from a neighborhood of $(t_0, 0, 0)$ in \mathbb{R}^3 to \mathbb{R}^3 with $F(t, 0, 0) = \gamma(t)$ and with $dF(t_0, 0, 0)$ non-singular. Then apply the Inverse Function Theorem to F . The functions f and g are then two of the coordinate functions of F^{-1} .
12. If $F : \mathbb{R}^p \rightarrow \mathbb{R}^p$ is a \mathcal{C}^1 function, what can you say about F at a point of \mathbb{R}^p where $\|F\|$ has a local minimum? How about a point where $\|F\|$ has a local maximum?

9.7 The Implicit Function Theorem

In this section we continue to develop consequences of the Inverse Function Theorem. The most notable of these is the Implicit Function Theorem. First we interpret the Inverse Function Theorem in the context of local systems of coordinates

Local Systems of Coordinates

Let F be a smooth function defined on an open subset U of \mathbb{R}^p which has values in \mathbb{R}^p and which has a smooth local inverse near a point $a \in U$. Then there is a

neighborhood V of a and a neighborhood W of $b = F(a)$ such that $F : V \rightarrow W$ is one to one and onto and has a smooth inverse function $G = F^{-1} : W \rightarrow V$.

We define a change of coordinates for points in V as follows: If

$$F = (f_1, f_2, \dots, f_p),$$

then we define new coordinates (u_1, u_2, \dots, u_p) for a point $x = (x_1, x_2, \dots, x_p)$ in V by setting

$$u_i = f_i(x_1, x_2, \dots, x_p) \quad \text{for } i = 1, \dots, p.$$

These new coordinates u_1, \dots, u_p are smooth functions of the old coordinates x_1, \dots, x_p and, similarly, the old coordinates are smooth functions of the new coordinates since

$$x_j = g_j(u_1, u_2, \dots, u_p) \quad \text{for } j = 1, \dots, p,$$

where g_j is the j th coordinate function of the inverse function G .

By subtracting the constant b from F , if necessary, we may assume that $F(a) = 0$ and W is a neighborhood of 0. This just makes the point a the origin in the new coordinate system.

A coordinate hyperplane (intersected with W) in the new coordinates is a set of the form

$$H_i = \{u \in W : u_i = 0\}.$$

In the original coordinates, this is the set

$$\{x \in V : f_i(x) = 0\}.$$

This means that the level set $\{x \in V : f_i(x) = 0\}$ for the function f_i looks like a smoothly deformed hyperplane (intersected with V). Similarly, the subset obtained by setting k of the coordinates $\{u_1, \dots, u_p\}$ equal to zero is a $p - k$ dimensional subspace of \mathbb{R}^p . In the old coordinates this looks like a smoothly deformed $p - k$ subspace intersected with V . If $k = p - 1$ the result is a line through the origin in the new coordinates and a curve through a in the old coordinates.

Parameterizing a Curve

A key question raised in in the last subsection of Section 9.4 is: when does a level set for a smooth function from one Euclidean space to another locally have a smooth parameterization and, hence, a tangent space at each of its points? The following example gives an answer to this question in the case of a level set for a real valued function on \mathbb{R}^2 . The method used in this example is a model for the proof of the Implicit Function Theorem, which will be proved next.

Example 9.7.1. Show that if $f : \mathbb{R}^2 \rightarrow \mathbb{R}^1$ is a smooth function and (a, b) is a point of \mathbb{R}^2 such that $f(a, b) = 0$ and $df(a, b) \neq 0$, then there is a neighborhood V of (a, b) in which $S = \{(x, y) : f(x, y) = 0\}$ is the image of a smooth parameterized curve. Find the tangent line to this curve at (a, b) .

Solution: Since $df(a, b) \neq 0$, either $\frac{\partial f}{\partial x}$ or $\frac{\partial f}{\partial y}$ is non-zero at (a, b) . Assume $\frac{\partial f}{\partial y}(a, b) \neq 0$ (the analysis in the other case is the same, but with the roles of x and y reversed). We define a function $H : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ by

$$H(x, y) = (x, f(x, y)).$$

The differential matrix of this function is

$$\begin{pmatrix} 1 & 0 \\ \frac{\partial f}{\partial x} & \frac{\partial f}{\partial y} \end{pmatrix}.$$

which has determinant $\frac{\partial f}{\partial y}$. Since $\frac{\partial f}{\partial y}(a, b) \neq 0$, this matrix is non-singular at (a, b) . Hence, there is a neighborhood V of (a, b) , a neighborhood W of $(a, 0)$, and a smooth inverse function $H^{-1} : W \rightarrow V$ for H . We have

$$H^{-1}(x, 0) = (k(x), g(x)),$$

for some smooth real valued functions k, g , defined for all x with $(x, 0) \in W$. Then,

$$(x, 0) = H \circ H^{-1}(x, 0) = (k(x), f(k(x), g(x))) \quad \text{whenever } (x, 0) \in W.$$

It follows that $k(x) = x$ and $f(x, g(x)) = 0$ for all such x . On the other hand, if $(x, y) \in V$ and $f(x, y) = 0$, then $H(x, y) = (x, 0)$ and so

$$(x, y) = H^{-1} \circ H(x, y) = H^{-1}(x, 0) = (x, g(x)).$$

Thus, $y = g(x)$. We conclude that, for $(x, y) \in V$, $f(x, y) = 0$ if and only if $y = g(x)$. Since, $(a, b) \in V$ and $f(a, b) = 0$, this means, in particular, that $g(a) = b$. Thus, we have proved that, near (a, b) , S is the graph of the smooth function g and

$$\gamma(x) = (x, g(x))$$

is a smooth parameterization of S near (a, b) .

The tangent line to S at (a, b) is given parametrically by

$$\begin{aligned} \tau(x) &= (a, b) + \gamma'(a, b)(x - a) \\ &= (a, b) + (1, g'(a))(x - a) = (x, b + g'(a)(x - a)), \end{aligned}$$

where, since $f(x, g(x)) = 0$, the chain rule tells us that

$$g' = - \left(\frac{\partial f}{\partial y} \right)^{-1} \frac{\partial f}{\partial x}.$$

The tangent line can also be described as the set of all (x, y) such that $(x - a, y - b)$ is orthogonal to the gradient of f at (a, b) – that is, all solutions to the equation

$$\frac{\partial f}{\partial x}(a, b)(x - a) + \frac{\partial f}{\partial y}(a, b)(y - b) = 0.$$

The Implicit Function Theorem

The proof of the Implicit Function Theorem follows exactly the same pattern as the solution to the preceding exercise.

The Implicit Function Theorem provides the answer to a very simple question: When can an equation of the form

$$F(x, y) = 0$$

be solved for y as a function of x ? That is, when can we find a function g such that $F(x, g(x)) = 0$? We note several things about this problem:

1. The problem makes perfectly good sense if F is a real valued function of 2 real variables (as in the previous example), but it also makes sense if F is a vector valued function of variables x and y which are also vectors.
2. As was the case with the Inverse Function Theorem, we might expect that there are local solutions to this problem for (x, y) near a point (a, b) where $F(a, b) = 0$, even though global solutions may not be possible.
3. Whether such a local solution is possible near a given point may depend on conditions on the differential matrix of F at the point.

In the statement and the proof of the Implicit Function Theorem, we will need to deal with certain submatrices of the full differential matrix of a function F . In this regard, the following notation will be useful. If f_1, f_2, \dots, f_k are smooth functions defined on an open set U in some Euclidean space \mathbb{R}^d (these may be some or all of the coordinate functions of a vector function F defined on U) and if y_1, \dots, y_m are some of the coordinates describing points in \mathbb{R}^d , then we set

$$\frac{\partial(f_1, \dots, f_k)}{\partial(y_1, \dots, y_m)} = \begin{pmatrix} \frac{\partial f_1}{\partial y_1} & \frac{\partial f_1}{\partial y_2} & \dots & \frac{\partial f_1}{\partial y_m} \\ \frac{\partial f_2}{\partial y_1} & \frac{\partial f_2}{\partial y_2} & \dots & \frac{\partial f_2}{\partial y_m} \\ \dots & \dots & \dots & \dots \\ \frac{\partial f_k}{\partial y_1} & \frac{\partial f_k}{\partial y_2} & \dots & \frac{\partial f_k}{\partial y_m} \end{pmatrix}$$

If $F = (f_1, \dots, f_q) : U \rightarrow \mathbb{R}^q$ is a function on a subset U of \mathbb{R}^p with the coordinates in \mathbb{R}^p labeled x_1, \dots, x_p , then $\frac{\partial(f_1, \dots, f_q)}{\partial(x_1, \dots, x_p)}$ is just another notation for dF . However, we will want to use this notation in cases where only some of the coordinate functions and/or some of the variables of F are used.

In the following theorem, \mathbb{R}^{p+q} will be identified with $\mathbb{R}^p \times \mathbb{R}^q$ and points in this space will be expressed in the form $(x, y) = (x_1, \dots, x_p, y_1, \dots, y_q)$.

Theorem 9.7.2. Let $U \subset \mathbb{R}^{p+q}$ be open, let $F = (f_1, \dots, f_q) : U \rightarrow \mathbb{R}^q$ be a smooth function, and let (a, b) be a point of U with $F(a, b) = 0$. Also, suppose the square matrix

$$\frac{\partial(f_1, \dots, f_q)}{\partial(y_1, \dots, y_q)}$$

is non-singular. Then there are neighborhoods $V \subset U$ of (a, b) and A of a and a smooth function $G : A \rightarrow \mathbb{R}^q$ such that $(x, G(x)) \in V$ for all $x \in A$, $G(a) = b$, and

$$F(x, y) = 0 \quad \text{for } x, y \in V \quad \text{if and only if } y = G(x).$$

Furthermore the differential of G on A is given by

$$dG = \frac{\partial(g_1, \dots, g_q)}{\partial(x_1, \dots, x_p)} = - \left(\frac{\partial(f_1, \dots, f_q)}{\partial(y_1, \dots, y_q)} \right)^{-1} \frac{\partial(f_1, \dots, f_q)}{\partial(x_1, \dots, x_p)}. \quad (9.7.1)$$

Proof. We will prove this by applying the Inverse Function Theorem to another function H , constructed from F . We define $H : U \rightarrow \mathbb{R}^p \times \mathbb{R}^q$ by

$$H(x, y) = (x, F(x, y)). \quad \text{for } (x, y) \in U.$$

The function H is \mathcal{C}^1 on U because F is \mathcal{C}^1 . The differential of H is

$$dH = \begin{pmatrix} 1 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 0 & \cdots & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 1 & 0 & \cdots & 0 \\ \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_p} & \frac{\partial f_1}{\partial y_1} & \cdots & \frac{\partial f_1}{\partial y_q} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \frac{\partial f_q}{\partial x_1} & \cdots & \frac{\partial f_q}{\partial x_p} & \frac{\partial f_q}{\partial y_1} & \cdots & \frac{\partial f_q}{\partial y_q} \end{pmatrix}$$

with an identity matrix in the upper left $p \times p$ block and a 0 matrix in the upper right $p \times q$ block. The bottom q rows form the differential matrix dF for F . The determinant of dH is just the determinant of the lower right $q \times q$ block – that is, the determinant of $\frac{\partial(f_1, \dots, f_q)}{\partial(y_1, \dots, y_q)}$. This determinant is non-zero at (a, b) by hypothesis. Hence, dH also has a non-zero determinant at (a, b) and is, therefore, non-singular at this point.

By the Inverse Function Theorem (Theorem 9.6.5) there are neighborhoods $V \subset U$ of (a, b) and W of $H(a, b)$ such that H has a smooth inverse function $H^{-1} : W \rightarrow V$. We have

$$H^{-1}(x, 0) = (K(x), G(x)),$$

for some smooth functions K and G , defined on $A = \{x \in \mathbb{R}^p : (x, 0) \in W\}$ with values in \mathbb{R}^q . The set A is open because it is the inverse image of W under the continuous function $x \rightarrow (x, 0) : \mathbb{R}^p \rightarrow \mathbb{R}^p \times \mathbb{R}^q$. Furthermore,

$$(x, 0) = H \circ H^{-1}(x, 0) = (K(x), F(K(x), G(x))) \quad \text{whenever } x \in A.$$

Thus, $K(x) = x$ and $F(x, G(x)) = 0$ for all $x \in A$. On the other hand, if $(x, y) \in V$ and $F(x, y) = 0$, then $H(x, y) = (x, 0)$ and so

$$(x, y) = H^{-1} \circ H(x, y) = H^{-1}(x, 0) = (x, G(x)).$$

Thus, $y = G(x)$. We conclude that if $(x, y) \in V$, then $F(x, y) = 0$ if and only if $y = G(x)$. This applies, in particular, when $(x, y) = (a, b)$ and so $G(a) = b$.

If we take the differential of both sides of the equation $F(x, G(x)) = 0$ the result is

$$\frac{\partial(f_1, \dots, f_q)}{\partial(x_1, \dots, x_p)} + \frac{\partial(f_1, \dots, f_q)}{\partial(y_1, \dots, y_q)} \frac{\partial(g_1, \dots, g_q)}{\partial(x_1, \dots, x_p)} = 0.$$

On solving this for $\frac{\partial(g_1, \dots, g_q)}{\partial(x_1, \dots, x_p)}$, we obtain (9.7.1). □

The Implicit Function Theorem leads to conditions under which a level set of a function has a smooth parameterization and, hence, a tangent space. This is the issue raised at the end of Section 9.4. This is also a key issue in the hypotheses of the theorem concerning the method of Lagrange Multipliers (Theorem 9.5.15).

Corollary 9.7.3. *Let $U \subset \mathbb{R}^d$ be an open set and $F : U \rightarrow \mathbb{R}^q$ a smooth function. Suppose $c \in U$, $F(c) = 0$, and $dF(c)$ has rank q . Then there is a neighborhood V of c , $V \subset U$, such that the level set $S = \{u \in V : F(u) = 0\}$ is a smooth p -surface, where $p = d - q$. That is, S has a smooth parameterization of dimension p . Hence, S has a tangent space at each point of S . Furthermore, the tangent space at c is the set of solutions u to the equation*

$$dF(c)(u - c) = 0.$$

Proof. Since $dF(c)$ has rank q , there is a $q \times q$ submatrix of the $q \times d$ matrix $dF(c)$ which is non-singular. By rearranging the variables in F , if necessary, we may assume that the last q columns of dF form a non-singular matrix. With $p = d - q$, we may represent \mathbb{R}^d as $\mathbb{R}^p \times \mathbb{R}^q$ and label the variables by $(x, y) = (x_1, \dots, x_p, y_1, \dots, y_q)$, as in the preceding theorem. Then the hypotheses of that theorem are satisfied, with $c = (a, b)$.

By the Implicit Function Theorem, there are neighborhoods V of $c = (a, b)$ and A of a and a smooth function $G : A \rightarrow \mathbb{R}^q$ with $(x, G(x)) \in V$ for all $x \in A$ and such that $F(x, y) = 0$ for $(x, y) \in V$ if and only if $y = G(x)$.

Thus, $S = \{u = (x, y) \in V : F(u) = 0\}$ is the graph of the smooth function G . Then the function $H(x) = (x, G(x))$ is a smooth parameterization of S . □

Example 9.7.4. For the system of equations

$$\begin{aligned}u^2 + v^2 - x &= 0 \\ u + v + y &= 0,\end{aligned}$$

find the points on the solution set S at which it may not be possible to solve for u and v as smooth functions of x and y in some neighborhood of the point.

Solution According to the Implicit Function Theorem, there will be smooth solutions in a neighborhood of any point where the following matrix is non-singular:

$$\frac{\partial(f_1, f_2)}{\partial(u, v)} = \begin{pmatrix} 2u & 2v \\ 1 & 1 \end{pmatrix},$$

where $f_1(x, y, u, v) = u^2 + v^2 - x$ and $f_2(x, y, u, v) = u + v + y$. This matrix is singular only when $u = v$. This happens at a point on S if and only if $u = v$ and $y^2 = 2x$.

Recall that the kernel of an affine transformation $L : \mathbb{R}^p \rightarrow \mathbb{R}$ of rank 1 is a hyperplane in \mathbb{R}^p . The Implicit Function Theorem allows us to draw a similar conclusion for functions which are not affine.

Example 9.7.5. For the equation

$$x^2 + y^2 + z^3 = 0,$$

at which points on its solution set S can we be assured that there is a neighborhood of the point in which S is a smoothly parameterized surface? Find an equation of the tangent space at each such point.

Solution: By the corollary to the Implicit Function Theorem, there will be a smooth parameterization of S in a neighborhood of any point at which df has rank 1, where $f(x, y, z) = x^2 + y^2 + z^3$. Since

$$df(x, y, z) = (2x, 2y, 3z^2),$$

the only point at which such a parameterization may not be possible is the origin.

At any point (a, b, c) which is not the origin, an equation for the tangent space is

$$df(a, b, c)(x - a, y - b, z - c) = 0,$$

or

$$2a(x - a) + 2b(y - b) + 3c^2(z - c) = 0.$$

Exercise Set 9.7

1. Are there any points on the graph of the equation $x^3 + 3xy^2 + 2y^3 = 1$ where it may not be possible to solve for y as a smooth function of x in some neighborhood of the point?

2. Can the equation $xz + yz + \sin(x + y + z) = 0$ be solved, in a neighborhood of $(0, 0, 0)$ for z as a smooth function $z = g(x, y)$ of (x, y) , with $g(0, 0) = 0$?
3. Find $\frac{\partial(f_1, f_2)}{\partial(u, v)}$ if

$$\begin{aligned} f_1(x, y, u, v) &= u^2 + v^2 + x^2 + y^2 \\ f_2(x, y, u, v) &= xu + yv + x - y. \end{aligned}$$

At which points (x, y, u, v) is this matrix non-singular?

4. Show that the system of equations

$$\begin{aligned} u^2 + v^2 + 2u - xy + z &= 0 \\ u^3 + \sin v - xu + yv + z^2 &= 0 \end{aligned}$$

has a solution for (u, v) as a smooth function of (x, y, z) , in some neighborhood of $(0, 0, 0)$, with the property that $(u, v) = (0, 0)$ when $(x, y, z) = (0, 0, 0)$.

5. Show that the system of equations

$$\begin{aligned} u^3 + x^2v^2 - 2y + w &= 0 \\ v^3 + y^2u^2 - 2x + w &= 0 \\ w^2 + wx - y^2 &= 0 \end{aligned}$$

has a solution for u, v, w as functions of (x, y) in a neighborhood of the point $(x, y, u, v, w) = (1, 1, 1, 1, 0)$ with $u(1, 1) = 1, v(1, 1) = 1, w(1, 1) = 0$.

6. For the equation $xy + yz + xz = 1$, at which points on the solution set S is there a neighborhood in which S is a smooth 2 surface? At each such point (a, b, c) , find an equation of the tangent plane.
7. For the system of equations

$$\begin{aligned} x^2 + y^2 - z^2 &= 0 \\ x + y + z &= 0, \end{aligned}$$

at which points of the solution set S is there a neighborhood in which S is a smooth curve? At each such point, find an equation of the tangent line.

8. For the system of equations

$$\begin{aligned} x^2 + y^2 + u^2 - 3v &= 1 \\ 2x + xy - y + 3u^2 - 9v &= 0, \end{aligned}$$

find all points on the solution set S for which there is a neighborhood in which S is a smooth 2 surface.

9. If $F(x, y, u, v) = (x e^u + y e^u, xv + yu) \in \mathbb{R}^2$, find those points (x, y, u, v) at which the level set of F , containing this point, is a smooth 2-surface in a neighborhood of the point.
10. If $F : \mathbb{R}^p \rightarrow \mathbb{R}^q$ is a smooth function and dF has rank q at a certain point $a \in \mathbb{R}^p$, prove that there is a neighborhood of a in which dF has rank q .