CHAPTER 16

Differentiable Functions of Several Variables

§16.1. The Differential and Partial Derivatives

Let $w = f(x,y,z)$ be a function of the three variables x,y,z. In this chapter we shall explore how to evaluate the change in w near a point (x_0,y_0,z_0), and make use of that evaluation. For functions of one variable, this led to the derivative: dw/dx is the rate of change of w with respect to x. But in more than one variable, the lack of a unique independent variable makes this more complicated. In particular, the rates of change may differ, depending upon the direction in which we move. We start by using the one variable theory to define change in w with respect to one variable at a time.

Definition 16.1 *Suppose we are given a function* $w = f(x,y,z)$*. The* **partial derivative** *of* f **with respect to** x *is defined by differentiating* f *with respect to* x*, considering* y *and* z *as being held constant. That is, at a point* (x_0,y_0,z_0)*, the value of the partial derivative with respect to* x *is*

(16.1) $$\frac{\partial f}{\partial x}(x_0,y_0,z_0) = \frac{d}{dx} f(x,y_0,z_0) = \lim_{h\to 0} \frac{f(x_0+h,y_0,z_0) - f(x_0,y_0,z_0)}{h} .$$

Similarly, if we keep x and z constant, we define the **partial derivative** of f **with respect to** y by

(16.2) $$\frac{\partial f}{\partial y} = \frac{d}{dy} f(x_0,y,z_0) ,$$

and by keeping x and y constant, we define the **partial derivative** of f **with respect to** z by

(16.3) $$\frac{\partial f}{\partial z} = \frac{d}{dz} f(x_0,y_0,z) .$$

Example 16.1 Find the partial derivatives of $f(x,y) = x(1+xy)^2$.

Thinking of y as a constant, we have

(16.4) $$\frac{\partial f}{\partial x} = (1+xy)^2 + x(2(1+xy)y) = (1+xy)(1+3xy) .$$

Now, we think of x as constant and differentiate with respect to y:

(16.5) $$\frac{\partial f}{\partial y} = x(2(1+xy)x) = 2x^2(1+xy)\ .$$

Example 16.2 The partial derivatives of $f(x,y,z) = xyz$ are

(16.6) $$\frac{\partial f}{\partial x} = yz\ , \quad \frac{\partial f}{\partial y} = xz\ , \quad \frac{\partial f}{\partial z} = zy\ .$$

Of course, the partial derivatives are themselves functions, and when it is possible to differentiate the partial derivatives, we do so, obtaining higher order derivatives. More precisely, the partial derivatives are found by differentiating the formula for f with respect to the relevant variable, treating the other variable as a constant. Apply this procedure to the functions so obtained to get the **second partial derivatives**:

(16.7) $$\frac{\partial^2 f}{\partial x^2} = \frac{\partial}{\partial x}(\frac{\partial f}{\partial x})\ , \quad \frac{\partial^2 f}{\partial y \partial x} = \frac{\partial}{\partial y}(\frac{\partial f}{\partial x}), \quad \frac{\partial^2 f}{\partial x \partial y} = \frac{\partial}{\partial x}(\frac{\partial f}{\partial y})\ , \quad \frac{\partial^2 f}{\partial y^2} = \frac{\partial}{\partial y}(\frac{\partial f}{\partial y})$$

Example 16.3 Calculate the second partial derivatives of the function in example 1.

We have $f(x,y) = x(1+xy)^2$, and have found

(16.8) $$\frac{\partial f}{\partial x} = (1+xy)(1+3xy)\ , \quad \frac{\partial f}{\partial y} = 2x^2(1+xy)\ .$$

Differentiating these expressions, we obtain

(16.9) $$\frac{\partial^2 f}{\partial x^2} = (1+xy)(3y) + y(1+3xy) = 4y + 6xy^2$$

(16.10) $$\frac{\partial^2 f}{\partial y \partial x} = (1+xy)(3x) + x(1+3xy) = 4x + 6x^2y$$

(16.11) $$\frac{\partial^2 f}{\partial x \partial y} = 4x(1+xy) + 2x^2y = 4x + 6x^2y$$

(16.12) $$\frac{\partial^2 f}{\partial y^2} = 2x^2(x) = 2x^3\ .$$

Notice that the second and third lines are equal. This is a general fact: the mixed partials (the middle terms above) are equal when the second partials are continuous:

(16.13) $$\frac{\partial^2 f}{\partial y \partial x} = \frac{\partial^2 f}{\partial x \partial y}$$

This is not easily proven, but is easily verified by many examples. Thus $\partial^2 f/\partial x\partial y$ can be calculated in whatever is the most convenient order. Finally, we note an alternative notation for partial derivatives;

$$f_x = \frac{\partial f}{\partial x}, \quad f_y = \frac{\partial f}{\partial y}, \quad f_{xx} = \frac{\partial^2 f}{\partial x^2}, \quad f_{xy} = \frac{\partial^2 f}{\partial x\partial y} \quad f_{yy} = \frac{\partial^2 f}{\partial y^2}\text{ , etc.} \tag{16.14}$$

Example 16.4 Let $f(x,y) = y\tan x + x\sec y$. Show that $f_{xy} = f_{yx}$.

We calculate the first partial derivatives and then the mixed partials in both orders:

$$f_x = y\sec^2 x + \sec y\ , \qquad f_y = \tan x + x\sec y\tan y \tag{16.15}$$

$$f_{yx} = \sec^2 x + \sec y\tan y \qquad f_{xy} = \sec^2 x + \sec y\tan y\ . \tag{16.16}$$

The partial derivatives of a function $w = f(x,y,z)$ tell us the rates of change of w in the coordinate directions. But there are many directions at a point on the plane or in space: how do we find these rates in other directions? More generally, if two or three variables are changing, how do we explore the corresponding change in w? The answer to these questions starts with the generalization of the idea of the differential as linear approximation. For a function of one variable, a function $w = f(x)$ is differentiable

if it is can be locally approximated by a linear function

$$w = w_0 + m(x - x_0) \tag{16.17}$$

or, what is the same, the graph of $w = f(x)$ at a point (x_0,y_0) is more and more like a straight line, the closer we look. The line is determined by its slope $m = f'(x_0)$. For functions of more than one variable, the idea is the same, but takes a little more explanation and notation.

Definition 16.2 *Let $w = f(x,y,z)$ be a function defined near the point (x_0,y_0,z_0). We say that f is* **differentiable** *if it can be well- approximated near (x_0,y_0,z_0) by a linear function*

$$w - w_0 = a(x - x_0) + b(y - y_0) + c(z - z_0)\ . \tag{16.18}$$

In this case, we call the linear function the **differential** *of f at (x_0,y_0,z_0), denoted $df((x_0,y_0,z_0)$. It is important to keep in mind that the differential is a function of a vector at the point; that is, of the increments $(x - x_0, y - y_0, z - z_0)$.*

If $f(x,y)$ is a function of two variables, we can consider the **graph** of the function as the set of points (x,y,z) such that $z = f(x,y)$. To say that f is differentiable is to say that this graph is more and more like a plane, the closer we look. This plane, called the **tangent plane** to the graph, is the graph of the approximating linear function, the differential. For a precise definition of what we mean by "well" approximated, see the discussion in section 16.3. The following example illustrates this meaning.

Example 16.5 Let $f(x,y) = x^2 + y$. Find the differential of f at the point (1,3). Find the equation of the tangent plane to the graph of $z = f(x)$ at the point.

We have $(x_0,y_0) = (1,3)$, and $z_0 = f(x_0,y_0) = 4$. Express $z - 4$ in terms of $x - 1$ and $y - 3$:

$$z - 4 = x^2 + y - 4 = (1 + (x - 1))^2 + (3 + (y - 3)) - 4 \tag{16.19}$$

(16.20) $$= 1 + 2(x-1) + (x-1)^2 + 3 + (y-3)\ , \quad \text{simplifing to}$$

(16.21) $$z - 4 = 2(x-1) + y - 3 + (x-1)^2\ .$$

Comparing with (16.18), the first two terms give the differential. $(x-1)^2$ is the error in the approximation. The equation of the tangent plane is

(16.22) $$z - 4 = 2(x-1) + y - 3 \quad \text{or} \quad z = 2x + y - 1\ .$$

If we just follow the function along the line where $y = y_0,\ z = z_0$, then (16.18) becomes just $w - w_0 = a(x - x_0)$; comparing this with definition 16.1, we see that a is the derivative of w in the x-direction, that is $a = \partial w/\partial x$. Similarly $b = \partial w/\partial y$ and $c = \partial w/\partial z$. Finally, since the variables $x,\ y,\ z$ are themselves linear, we have that dx *is* $x - x_0$,and so forth. This leads to the following restatement of the definition of differentiability:

Proposition 16.1 *Suppose that* $w = f(x,y,z)$ *is differentiable at* (x_0, y_0, z_0)*. Then*

(16.23) $$dw = \frac{\partial f}{\partial x}dx + \frac{\partial f}{\partial y}dy + \frac{\partial f}{\partial z}dz\ .$$

There are a variety of ways to use formula (16.23), which we now illustrate.

Example 16.6 Let

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

Find the equation of the tangent plane to the graph at the point (2,-1).

At $(x_0, y_0) = (2, -1)$, we have $z_0 = f(x_0, y_0) = 6$. We calculate

(16.25) $$\frac{\partial f}{\partial x} = 2x - y\ , \quad \frac{\partial f}{\partial y} = -x + 3y^2\ ,$$

so, at (2,-1), $\partial f/\partial x = 5,\ \partial f/\partial y = 1$. Substituting these values in (16.18) we obtain

(16.26) $$z - 6 = 5(x-2) + (y+1) \quad \text{or} \quad z = 5x + y - 3\ .$$

An alternative approach is to differentiate equation (16.24) implicitly:

(16.27) $$dz = 2xdx - xdy - ydx + 3y^2dy\ .$$

Evaluating at (2,-1), we have $z_0 = 6$, and $dz = 4dx - 2dy + dx + 3dy$. This is the equation of the tangent plane, with the differentials dx, dy, dz replaced by the increments $x-2, y+1, z-5$:

(16.28) $$z - 6 = 4(x-2) - 2(y+1) + (x-2) + 3(y+1)\ ,$$

which is the same as (16.26).

Example 16.7 Find the equation of the tangent plane to the graph of the function $z = x^2 + xy - y$ at (2,-1, 1).

First, we calculate the differential

(16.29) $$dz = 2xdx + xdy + ydx - dy$$

and then evaluate it at the point:

(16.30) $$dz = 4dx + 2dy - dx - dy = 3dx + dy\ .$$

We now get the equation of the tangent plane by replacing the differentials by the increments:

(16.31) $$z - 1 = 3(x-2) + (y+1) \quad \text{or} \quad z = 3x + y - 4\ .$$

Example 16.8 Find the points at which the graph of $z = f(x,y) = x^2 - 2xy + y$ has a horizontal tangent plane.

The horizontal plane through the point (x_0, y_0, z_0) has the equation $z - z_0 = 0$. Thus our points are those where $df = 0$; i.e., solutions of the pair of equations

(16.32) $$\frac{\partial f}{\partial x} = 0 \qquad \frac{\partial f}{\partial y} = 0\ .$$

Calculating, we get $2x - 2y = 0,\ -2x + 1 = 0$, so $x = 1/2,\ y = 1/2$ and our point is $(1/2, 1/2)$.

Example 16.9 Given the function $z = x^2 - xy + y^3$, in what direction, at the point (1,1,1) is the rate of change of z equal to zero?

The differential of z is $dz = (2x - y)dx + (-x + 3y^2)dy$, so at (1,1,1), we have $dz = dx + 2dy$. This is zero for the direction in which $dx = -2dy$; that is along the line of slope -1/2. Thus the answer is given by a vector in that direction, for example: $-2\mathbf{I} + \mathbf{J}$.

Example 16.10 Suppose that we have designed a cylindrical silo of base radius 6 meters and height 10 meters, and we are asked to increase the radius by .25 m and the height by .2 m. By (approximately) how much do we increase the volume?

The volume of a cylinder of radius r and height h is $V = \pi r^2 h$. To answer this question, we consider the linear approximation of volume, so we take the differential of V:

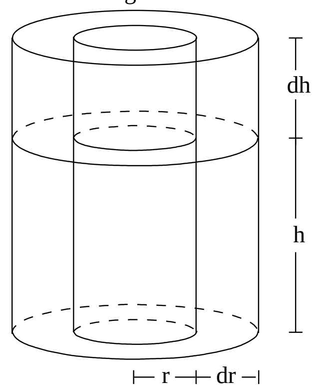
(16.33) $$dV = 2\pi rhdr + \pi r^2 dh\ .$$

Now, in our case $r = 5,\ h = 10,\ dr = .25,\ dh = .2$, so we calculate

(16.34) $$dV = 2\pi(5)(10)(.25) + \pi(5)^2(.2) = \pi(25+5) = 30\pi \quad \text{cubic meters}\ .$$

By looking at figure 1, we can identify the two terms in the increment of volume: the first is the volume of the shell of width dr around the cylinder, and the second is the volume of the cap of height dh. The negligible part is the volume $2\pi drdh$ of the washer at the top of width dr and height dh.

Figure 16.1



Proposition 16.2 *(The Chain Rule). Let $w = f(x,y,z)$ be a function defined in a region R in space. Suppose that γ is a curve in R given parametrically by $x = x(t),\ y = y(t).\ z = z(t)$, with $t = 0$ corresponding to (x_0,y_0,z_0). Then, considering $w = f(x(t),y(t),z(t))$ as a function of t along γ, we have*

$$\frac{dw}{dt} = \frac{\partial w}{\partial x}\frac{dx}{dt} + \frac{\partial w}{\partial y}\frac{dy}{dt} + \frac{\partial w}{\partial z}\frac{dz}{dt} \ . \tag{16.35}$$

That is, the rate of change of w with respect to t along γ is given by (16.35). We shall give an explanation of this formula in section 3.

Example 16.11 Let $w = f(x,y,z) = xy + y^2z$. Consider the curve given parametrically by $x = t,\ y = t^2,\ z = \ln t$. Find dw/dt at $t = 2$.

Differentiating,

$$\frac{\partial w}{\partial x} = y\ , \quad \frac{\partial w}{\partial y} = x + 2yz\ , \quad \frac{\partial w}{\partial z} = y^2\ , \tag{16.36}$$

$$\frac{dx}{dt} = 1\ , \quad \frac{dy}{dt} = 2t\ , \quad \frac{dz}{dt} = \frac{1}{t}\ , \tag{16.37}$$

so

$$\frac{dw}{dt} = y(1) + (x + 2yz)(2t) + y^2(\frac{1}{t})\ . \tag{16.38}$$

At $t = 2$ we calculate $x = 2,\ y = 4$ and $z = \ln 2$, giving

$$\frac{dw}{dt} = 4 + (2 + 8\ln 2)(4) + 16/2 = 20 + 32\ln 2\ . \tag{16.39}$$

Example 16.12 Let $z = x^2 + y^2$. Find the maximum value of z on the ellipse given parametrically by $x = \cos t,\ y = 2\sin t$.

We need to find the points at which $dz/dt = 0$. Now

$$\frac{\partial z}{\partial x} = 2x\ , \quad \frac{\partial z}{\partial y} = 2y\ , \quad \frac{dx}{dt} = -\sin t\ , \quad \frac{dy}{dt} = 2\cos t\ , \tag{16.40}$$

and thus

$$\frac{dz}{dt} = -2x\sin t + 4y\cos t\ . \tag{16.41}$$

Since $x = \cos t$ and $y = 2\sin t$, this gives $dz/dt = -2\sin t\cos t + 8\sin t\cos t$. Set this equal to zero to obtain $6\sin t\cos t = 0$, or $3\sin(2t) = 0$, which has the solutions $t = 0,\ \pm\pi/2,\ \pi$, and $x = \pm 1,\ y = 0$ or $x = 0,\ y = \pm 2$. The corresponding values of z are thus 1 and 4, so 1 is the minimum value and 4 is the maximum value on the ellipse.

We can think of an equation of the form $f(x,y,z) = 0$ as defining z **implicitly** as a function of x and y, in the sense that we could solve for z, given specific values of x and y. However, just as in one dimension, we need not solve for z to find the partial derivatives. If we take the differential of the defining equation $f(x,y,z) = 0$ we get

$$f_x dx + f_y dy + f_z dz = 0 \quad \text{so that} \quad dz = -\frac{f_x}{f_z}dx - \frac{f_y}{f_z}dy\ . \tag{16.42}$$

The coefficient of dx is thus $\partial z/\partial x$, and the coefficient of dy is $\partial z/\partial y$. Of course if $f_z = 0$, these are not defined. But if $f_z \neq 0$, then this method works.

Proposition 16.3 *Suppose that f is a differentiable function of (x,y,z) near the point (x_0,y_0,z_0), and that $f_z((x_0,y_0,z_0) \neq 0$. Then the equation $f(x,y,z) = 0$ defines z implicitly as a function of x,y and*

$$\frac{\partial z}{\partial x} = -\frac{f_x}{f_z} \quad \text{and} \quad \frac{\partial z}{\partial x} = -\frac{f_y}{f_z}\ . \tag{16.43}$$

Example 16.13 Given $f(x,y,z) = z^3 + 3xz^2 + y^2z$, find expressions for $\partial z/\partial x$ and $\partial z/\partial y$ where z is defined implicitly as a function of (x,y) by the equation $f(x,y,z) = 5$. Evaluate these at the point (1,1,1).

First we calculate the partial derivatives:

$$\frac{\partial f}{\partial x} = 3z^2\ , \quad \frac{\partial f}{\partial y} = 2yz\ , \quad \frac{\partial f}{\partial z} = 3z^2 + 6zx + y^2, \tag{16.44}$$

so that

$$\frac{\partial z}{\partial x} = -\frac{3z^2}{3z^2 + 6zx + y^2}\ , \quad \frac{\partial z}{\partial y} = -\frac{2yz}{3z^2 + 6zx + y^2}\ . \tag{16.45}$$

The values at (1,1,1) are $\partial z/\partial x = -3/10,\ \partial z/\partial x = -1/10$.

§16.2. Gradients and Vector methods

Let $w = f(x,y,z)$, where f is a differentiable function. To put the formula for the differential,(16.23), in vector form, we introduce the **gradient** of the function f:

(16.46) $$\nabla f = \frac{\partial f}{\partial x}\mathbf{I} + \frac{\partial f}{\partial y}\mathbf{J} + \frac{\partial f}{\partial z}\mathbf{K} ,$$

and the vector differential $d\mathbf{X} = dx\mathbf{I} + dy\mathbf{J} + dz\mathbf{K}$. We interpret $d\mathbf{X}$ as a small change in the vector $\mathbf{X}$. Then (16.17) can be rewritten as

(16.47) $$dw = \frac{\partial f}{\partial x}dx + \frac{\partial f}{\partial y}dy + \frac{\partial f}{\partial z}dz = (\nabla f)\cdot d\mathbf{X} .$$

This leads to the following vectorial form of the chain rule.

Proposition 16.4 *(The Gradient Form of the Chain Rule). Let $w = f(x,y,z)$ be a function defined in a region R in space. Suppose that γ is a curve in R given parametrically by $\mathbf{X} = \mathbf{X}(t)$, with $t = 0$ corresponding to $\mathbf{X_0}$. Then, considering $w = f(\mathbf{X}(t))$ as a function of t along γ, we have*

(16.48) $$\frac{dw}{dt} = (\nabla f)\cdot\frac{d\mathbf{X}}{dt} ,$$

evaluated at $\mathbf{X_0}$.

The partial derivatives tell us the rate of change of the function f in the coordinate directions. Using the gradient, we can calculate the rate of change in any direction.

Definition 16.3 *Let $w = f(x,y,z)$ be differentiable in a neighborhood of $\mathbf{X_0}$. For any vector $\mathbf{V}$, let $\mathbf{X}(t) = \mathbf{X_0} + t\mathbf{V}$ parametrize the line through $\mathbf{X_0}$ in the direction $\mathbf{V}$. The* **derivative of f along V** *is*

(16.49) $$D_{\mathbf{V}}f(\mathbf{X_0}) = \lim_{t\to 0}\frac{f(\mathbf{X_0} + t\mathbf{V}) - f(\mathbf{X_0})}{t} .$$

Propostion 16.5. Given the differentiable function f and a vector $\mathbf{V}$, we have

(16.50) $$D_{\mathbf{V}}f(\mathbf{X_0}) = \nabla f\cdot\mathbf{V} .$$

The right hand side of (16.49) is the derivative of f along the line in the direction of $\mathbf{V}$. That line is parametrized by $\mathbf{X}(t) = \mathbf{X_0} + t\mathbf{V}$, so $d\mathbf{X}/dt = \mathbf{V}$. Now, by the chain rule

(16.51) $$D_{\mathbf{V}}f(\mathbf{X_0}) = \frac{d}{dt}f(\mathbf{X}(t)) = \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt} + \frac{\partial f}{\partial z}\frac{dz}{dt} = \nabla f\cdot\frac{d\mathbf{X}}{dt} = \nabla f\cdot\mathbf{V} .$$

If we replace $\mathbf{V}$ by a unit vector $\mathbf{U}$, then the parameter t represents distance along the line, since $|\mathbf{X}(t) - \mathbf{X_0}| = t|\mathbf{U}| = t$. We say that the line is parametrized by arc length, and refer to $D_{\mathbf{U}}f$ as the **directional derivative** of f in the direction $\mathbf{U}$.

Example 16.14 Let $f(x,y) = x^3 - 3x^2 + xy + 7$ and $\mathbf{U} = 0.6\mathbf{I} - 0.8\mathbf{J}$. Find $D_{\mathbf{U}}f(1,-2)$.

We have $f_x = 3x^2 - 6x + y$, $f_y = x$. Evaluating at (1,-2), we have $\nabla f(1,-2) = -5\mathbf{I} + \mathbf{J}$. Thus

(16.52) $$D_{\mathbf{U}}f(1,-2) = \nabla f(1,-2)\cdot\mathbf{U} = -3 - 0.8 = -3.8 .$$

Example 16.15 For f as above, find the direction $\mathbf{U}$ at (1,-2) in which $D_{\mathbf{U}}f = 0$.

Let $\mathbf{U} = a\mathbf{I} + b\mathbf{J}$. We must solve

$$\nabla f(1,-2)\cdot\mathbf{U} = (-5\mathbf{I} + \mathbf{J}\cdot(a\mathbf{I} + b\mathbf{J}) = -5a + b = 0\ . \tag{16.53}$$

This gives $b = 5a$. Since $\mathbf{U}$ is a unit vector, we have $a^2 + b^2 = 26a^2 = 1$, so $a = 1/\sqrt{26},\ b = 5/\sqrt{26}$ will do Thus

$$\mathbf{U} = \frac{\mathbf{I} + 5\mathbf{J}}{\sqrt{26}}\ . \tag{16.54}$$

We also have the answer $-\mathbf{U}$. Notice that both these vectors are unit vectors in the direction of $\nabla f^{\perp}$.

Example 16.16 Let γ be parametrized by $\mathbf{X}(t) = t^2\mathbf{I} + \ln t\mathbf{J} + t\mathbf{K}$, and let $w = f(x,y) = xyz$. Find dw/dt along γ. What is the rate of change of w with respect to t at the point $t = 2$?

To use (16.48), we calculate

$$\nabla f = yz\mathbf{I} + xz\mathbf{J} + xy\mathbf{K}\ , \qquad \frac{d\mathbf{X}}{dt} = 2t\mathbf{I} + \frac{1}{t}\mathbf{J} + \mathbf{K}\ , \tag{16.55}$$

so that

$$\frac{dw}{dt} = (\nabla f)\cdot\frac{d\mathbf{X}}{dt} = 2tyz + \frac{xz}{t} + xy = 3t^2(\ln t + 1)\ , \tag{16.56}$$

since $x = t^2$, $y = \ln t$ and $z = t$. At $t = 2$, we get $dz/dt = 12(\ln 2 + 1)$.

Example 16.17 Let $\mathbf{X}(t) = \cos t\mathbf{I} + \sin t\mathbf{J}$ parametrize the unit circle, and let $f(x,y) = x^2 + 2xy$. Find the maximum value of f on the unit circle.

The function $z = f(\mathbf{X}(t))$ has a maximum when $dz/dt = 0$. We calculate:

$$\nabla f = (2x + 2y)\mathbf{I} + 2x\mathbf{J} = 2((\cos t + \sin t)\mathbf{I} + 2\cos t\mathbf{J}\ , \tag{16.57}$$

$$\frac{d\mathbf{X}}{dt} = -\sin t\mathbf{I} + \cos t\mathbf{J}\ , \tag{16.58}$$

so that

$$\frac{dz}{dt} = (\nabla f)\cdot\frac{d\mathbf{X}}{dt} = 2((\cos t + \sin t)(-\sin t) + 2\cos^2 t\ . \tag{16.59}$$

To solve $dz/dt = 0$ we use double angle formulas:

$$2((\cos t + \sin t)(-\sin t) + 2\cos^2 t = -2\cos t\sin t + 2(\cos^2 t - \sin^2 t) = -\sin(2t) + 2\cos(2t) \tag{16.60}$$

which is zero when $\tan(2t) = 2$, or $t = 31.7°, 211.7°$. The corresponding values of $x = \cos t,\ y = \sin t$ are $x = \pm.526,\ y = \pm.851$. Calculating the values of z at these points gives the maximum 1.172.

For a function $w = f(x,y,z)$ of three variables defined near the point $\mathbf{X_0} :\ (x_0,y_0,z_0)$, let $w_0 = f(x_0,y_0,z_0)$. The equation $w = w_0$ is the level surface S of w at (x_0,y_0,z_0). For f differentiable at a point

$\mathbf{X_0}$, the fact that f can be approximated by a linear function implies that the surface S looks more and more like a plane, the closer we look. This plane, given by the equation $df(\mathbf{X_0}) = 0$, is the **tangent plane** to S at $\mathbf{X_0}$. We now note that the gradient of f is the normal to this surface, and points in the direction of maximum increase of f.

Proposition 16.5 *Let f be a function differentiable in a neighborhood of the point $\mathbf{X_0}$.*
a) $\nabla f(\mathbf{X_0})$ points in the direction of maximum increase of the function f at $\mathbf{X_0}$.
b) $\nabla f(\mathbf{X_0})$ is the normal to the tangent plane of the level set of f through $\mathbf{X_0}$.

To show a), start with a unit vector $\mathbf{U}$. From (16.50) we have

(16.61) $$D_{\mathbf{U}}(\mathbf{X_0}) = \nabla f \cdot \mathbf{U} = |\nabla f| \cos\beta$$

where β is the angle between ∇f and $\mathbf{U}$ (since $|\mathbf{U}| = 1$). This takes its greatest value when $\cos\beta = 0$, that is $\mathbf{U} = \nabla f$. To show b), let $\mathbf{V}$ be a vector on the tangent plane. By definition, $df(\mathbf{X_0})(\mathbf{V}) = 0$, so

(16.62) $$\nabla f(\mathbf{X_0}) \cdot \mathbf{V} = df(\mathbf{X_0})(\mathbf{V}) = 0\ .$$

Thus $\nabla f(\mathbf{X_0})$ is orthogonal to every vector in the tangent plane, so can be taken to be its normal. Now, a point $\mathbf{X}$ lies in the tangent plane if and only if the vector $\mathbf{X} - \mathbf{X_0}$ lies on the tangent plane, or

(16.63) $$\nabla f(\mathbf{X_0}) \cdot (\mathbf{X} - \mathbf{X_0}) = 0\ ,$$

which is thus the equation of the tangent plane.

Example 16.18 Let $f(x,y) = x^3 + 3x^2y^2 + 2y$. Find the equation of the line tangent to the curve $f(x,y) = 9$ at (2,-1).

The above discussion for three dimensions holds just as well in two dimensions. Thus, by proposition 16.6, the normal to the tangent line to the curve is ∇f.We calculate $\nabla f = (3x^2 + 6xy^2)\mathbf{I} + (6x^2y + 2)\mathbf{J}$; which at $x = 2$, $y = -1$ is the vector $24\mathbf{I} - 22\mathbf{J}$. Now, the equation of the tangent line is given by (16.63), where $\mathbf{X_0} = 2\mathbf{I} - \mathbf{J}$ is the vector to the point (2,-1):

(16.64) $$(24\mathbf{I} - 22\mathbf{J}) \cdot ((x-2)\mathbf{I} + (y+1)\mathbf{J}) = 0 \quad \text{or} \quad 24(x-2) - 22(y+1) = 0\ ,$$

which simplifies to $24x - 22y = 70$.

Example 16.19 Let $f(x,y,z) = xyz$. Find the gradient of f. Find the equation of the tangent plane to the level surface $f(x,y,z) = 2$ at the point $\mathbf{X_0} : (1,2,1)$.

We calculate:

(16.65) $$\nabla f = \frac{\partial f}{\partial x}\mathbf{I} + \frac{\partial f}{\partial y}\mathbf{J} + \frac{\partial f}{\partial z}\mathbf{K} = yz\mathbf{I} + xz\mathbf{J} + xy\mathbf{K}\ .$$

At $\mathbf{X_0}$, $\nabla f = 2\mathbf{I} + \mathbf{J} + 2\mathbf{K}$, so the equation of the tangent plane is $\nabla f \cdot (\mathbf{X} - \mathbf{X_0}) = 0$:

(16.66) $$2(x-1) + (y-2) + 2(z-1) = 0 \quad \text{or} \quad 2x + y + 2z = 5\ .$$

Example 16.20 Let $w = x + xy - yz^2$. Find the equation of the tangent plane to the surface $w = 2$ at (3,1,2).

We calculate $\nabla w = (1+y)\mathbf{I} + (x - z^2)\mathbf{J} - 2zy\mathbf{K}$. At the given point (3,1,2), $\nabla w = 2\mathbf{I} - \mathbf{J} - 6\mathbf{K}$. This is the normal to the tangent plane, at $\mathbf{X_0} = 3\mathbf{I} + \mathbf{J} + 2\mathbf{K}$, so the equation of that plane is

(16.67) $$\nabla w \cdot (\mathbf{X} - \mathbf{X_0}) = 2(x-3) - (y-1) - 6(z-2) = 0\ ,$$

or $2x-y-6z+7=0$.

Example 16.21 Let S be the sphere $x^2+y^2+z^2=a^2,\ a>0$. Show that at any point $\mathbf{X}$ on the sphere, the vector $\mathbf{X}$ is orthogonal to the sphere.
Let $w=x^2+y^2+z^2$, so that S is the level set $w=a^2$. Then ∇w is normal to S at $\mathbf{X}=x\mathbf{I}+y\mathbf{J}+z\mathbf{K}$. But

$$\nabla w = 2x\mathbf{I}+2y\mathbf{J}+2z\mathbf{K}=2\mathbf{X}\ . \tag{16.68}$$

Example 16.22 Let S_1 be the sphere $x^2+y^2+z^2=4$ and S_2 the cylinder $x^2+y^2=1$. Let $\mathbf{X}=x\mathbf{I}+y\mathbf{J}+z\mathbf{K}$ be a point on the curve γ of intersection of the surfaces S_1 and S_2. Find a vector tangent to γ at $\mathbf{X}$.

Let $w_1=x^2+y^2+z^2,\ w_2=x^2+y^2$, so that γ is the intersection of the level sets $w_1=4$ and $w_2=1$. Then ∇w_1 and ∇w_2 are both orthogonal to the tangent to γ, so $\nabla w_1\times\nabla w_2$ points in the direction of the tangent to γ. We calculate:

$$\nabla w_1\times\nabla w_2 = 2(x\mathbf{I}+y\mathbf{J}+z\mathbf{K})\times 2(x\mathbf{I}+y\mathbf{J})=4(-yz\mathbf{I}+zx\mathbf{J})\ . \tag{16.69}$$

In the above, we have considered a surface as a graph or as a level set of a function. Surfaces can also be given **parametrically**. Let u and v be the variables in a region R of the plane, and let $\mathbf{X}(u,v)=x(u,v)\mathbf{I}+y(u,v)\mathbf{J}+z(u,v)\mathbf{K}$ be a vector-valued function on R. Then the set of values of $\mathbf{X}(u,v)$, as (u,v) ranges over R describes a surface in space.

Example 16.23 Consider the function $\mathbf{X}(u,v)=(u-v)\mathbf{I}+(u+v)\mathbf{J}+uv\mathbf{K}$ defined in (u,v) space. In coordinates, this is given by the equations

$$x=u-v\quad y=u+v\quad z=uv\ . \tag{16.70}$$

We can solve for u and v in terms of x and y;

$$u=\frac{x+y}{2}\quad v=\frac{-x+y}{2}\ ; \tag{16.71}$$

putting these in the formula for z we have

$$z=uv=\frac{x+y}{2}\frac{-x+y}{2}=\frac{-x^2+y^2}{4}\ , \tag{16.72}$$

so the surface is the hyperbolic paraboloid $4z=y^2-x^2$.

Now, in general it may not be so easy (or simple) to realize a parametric surface as a level set; however, we can use the parametric equations to , for example, find the tangent plane to the surface at a point.

Proposition 16.6 *Let $\mathbf{X}(u,v)=x(u,v)\mathbf{I}+y(u,v)\mathbf{J}+z(u,v)\mathbf{K}$ be a vector-valued function defined on a region in R-space. Define*

$$\mathbf{X_u}=\frac{\partial x}{\partial u}\mathbf{I}+\frac{\partial y}{\partial u}\mathbf{J}+\frac{\partial z}{\partial u}\mathbf{K}\ , \tag{16.73}$$

(16.74)
$$\mathbf{X_v} = \frac{\partial x}{\partial v}\mathbf{I} + \frac{\partial y}{\partial v}\mathbf{J} + \frac{\partial z}{\partial v}\mathbf{K} \ .$$

a) The vector $\mathbf{X_u} \times \mathbf{X_v}$ *is normal to the surface.*
b) If $w = f(x,y,z)$ *is a function defined near the surface, we can consider it as a function of* u *and* v *by writing* $w = f(x(u,v),y(u,v),z(u,v))$*. Then*

(16.75)
$$\frac{\partial w}{\partial u} = \nabla w \cdot \mathbf{X_u} \quad \frac{\partial w}{\partial v} = \nabla w \cdot \mathbf{X_v} \ .$$

To see this, fix a point (u_0,v_0). If we set $v = v_0$ and let u vary, we get the curve C given parameterically by

(16.76)
$$\mathbf{X}(u) = x(u,v_0)\mathbf{I} + y(u,v_0)\mathbf{J} + z(u,v_0)\mathbf{K} \ .$$

The tangent vector to this curve is $\mathbf{X_u}$, and since the curve lies in the surface, its tangent vector lies in the tangent plane. Similarly, considering the curve $u = u_0$, we see that the vector $\mathbf{X_v}$ also lies in the tangent plane. Thus $\mathbf{X_u} \times \mathbf{X_v}$ is normal to the tangent plane. Part b) follows directly from the chain rule, applied to the curves $u = u_0$ and $v = v_0$.

Example 16.24 Find the equation of the tangent plane to the surface of example 23 at the point (-2,4,3).

From (16.71), this point corresponds to the values $u = 1, \ v = 3$. Now, we differentiate the function defining the surface, obtaining

(16.77)
$$\mathbf{X}_u = \mathbf{I} + \mathbf{J} + v\mathbf{K} \ , \quad \mathbf{X}_v = -\mathbf{I} + \mathbf{J} + u\mathbf{K} \ .$$

The values at $u = 1, \ v = 3$ are $\mathbf{X_u} = \mathbf{I} + \mathbf{J} + 3\mathbf{K}$, $\mathbf{X_v} = -\mathbf{I} + \mathbf{J} + \mathbf{K}$. Thus, a normal to the tangent plane is $\mathbf{N} = \mathbf{X_u} \times \mathbf{X_v} = -2\mathbf{I} - 4\mathbf{J} + 2\mathbf{K}$, and the equation of the tangent plane is

(16.78)
$$-2(x+2) - 4(y-4) + 2(z-3) = 0 \quad \text{or} \quad z = x + 2y - 3$$

Example 16.25 Consider the surface given parametrically by

(16.79)
$$\mathbf{X}(u,v) = 3\cos u \cos v\mathbf{I} + 4\cos u \sin v\mathbf{J} + 5\sin u\mathbf{K} \ .$$

Find the normal to the tangent plane ant the point corresponding to $u = \pi/3, \ v = \pi/6$.

Differentiate:

(16.80)
$$\mathbf{X_u} = -3\sin u \cos v\mathbf{I} - 4\sin u \sin v\mathbf{J} + 5\cos u\mathbf{K} \ ,$$

(16.81)
$$\mathbf{X_v} = -3\cos u \sin v\mathbf{I} + 4\cos u \cos v\mathbf{J} \ .$$

Evaluating at the given point, we have

(16.82)
$$\mathbf{X_u} = -3\frac{\sqrt{3}}{2}\frac{\sqrt{3}}{2}\mathbf{I} - 4\frac{\sqrt{3}}{2}\frac{1}{2}\mathbf{J} + 5\frac{1}{2}\mathbf{K} = -\frac{9}{4}\mathbf{I} - \sqrt{3}\mathbf{J} + \frac{5}{2}\mathbf{K} \ ,$$

(16.83)
$$\mathbf{X_v} = -3\frac{1}{2}\frac{1}{2}\mathbf{I} + 4\frac{1}{2}\frac{\sqrt{3}}{2}\mathbf{J} = -\frac{3}{4}\mathbf{I} + \sqrt{3}\mathbf{J} \ .$$

A normal to the plane is $\mathbf{N} = \mathbf{X_u} \times \mathbf{X_v} = -(5/2)\sqrt{3}\mathbf{I} + (15/8)\mathbf{J} + 3\sqrt{3}\mathbf{K}$.

§16.3. Theoretical considerations

In order to make the intuitive concept of linear approximation, as used above, more precise we start with the idea of closeness in the space itself. We measure the "nearness" of two points by the length of the line segment joining the points. Thus, in vectorial terms, the **distance** between $\mathbf{X}$ and $\mathbf{X_0}$ is $|\mathbf{X}-\mathbf{X_0}|$, that is, the square root of the sum of the squares of the components. We define limits in terms of this distance.

Definition 16.4 *The* **ball** *of radius c centered at* $\mathbf{X_0}$ *(denoted* $B(\mathbf{X_0},c)$*) is the set of all points of distance less than c from* $\mathbf{X_0}$*. A* **neighborhood** *of* $\mathbf{X_0}$ *is any set which contains some ball centered at* $\mathbf{X_0}$*.*

Definition 16.5 *Suppose that f is a function defined in a neighborhood of* $\mathbf{X_0}$*. We say that*

(16.84)
$$\lim_{\mathbf{X}\to\mathbf{X_0}} f(\mathbf{X}) = L$$

if we can insure that $|f(\mathbf{X})-L|$ *can be made as small as we please by taking* $\mathbf{X}$ *close enough to* $\mathbf{X_0}$*. We say that f is* **continuous** *at* $\mathbf{X_0}$ *if*

(16.85)
$$\lim_{\mathbf{X}\to\mathbf{X_0}} f(\mathbf{X}) = f(\mathbf{X_0}) \ .$$

Just as in one variable, we are assured that all functions which can be expressed by polynomials in the coordinates are continuous.

Definition 16.6 *A* **linear function** *is a function of the form* $L(\mathbf{X}) = \mathbf{C}\cdot\mathbf{X}$*, for some vector* $\mathbf{C}$*. In coordinates we have* $L(x,y,z) = ax+by+cz$*, where we have written* $\mathbf{C} = a\mathbf{I}+b\mathbf{J}+c\mathbf{K}$ *and* $\mathbf{X} = x\mathbf{I}+y\mathbf{J}+z\mathbf{K}$*. Its level surface through* $\mathbf{X_0}$ *is the plane* $a(x-x_0)+b(y-y_0)+c(z-z_0) = 0$*, or* $\mathbf{C}\cdot(\mathbf{X}-\mathbf{X_0}) = 0$*.*

Now we define differentiability at $\mathbf{X_0}$ of a function f: that it can be well- approximated by a linear function. This is the direct generalization of the definition of the derivative in one dimension.

Definition 16.7 *Suppose that f is a function defined in a neighborhood of* $\mathbf{X_0}$*. We say that f is* **differentiable** *at* $\mathbf{X_0}$ *if there is a linear function L such that*

(16.86)
$$\lim_{\mathbf{X}\to\mathbf{X_0}} \frac{|f(\mathbf{X})-f(\mathbf{X_0})-L(\mathbf{X}-\mathbf{X_0})|}{|\mathbf{X}-\mathbf{X_0}|} = 0 \ .$$

In this case, we call L the **differential** *of f at* $\mathbf{X_0}$*, denoted* $df(\mathbf{X_0})$*. We can write* $L\cdot(\mathbf{X}-\mathbf{X_0}) = \nabla f\cdot(\mathbf{X}-\mathbf{X_0})$*, where* ∇f *is the* **gradient** *of f.*

Now, as we have seen, the calculation of differentials amounts to calculating partial derivatives. To see this in terms of the above definition, let's look at the situation in two variables, writing $\mathbf{X} = x\mathbf{I}+y\mathbf{J}$. Suppose that f is differentiable at $\mathbf{X_0}$, and its differential there is $L(x-x_0,y-y_0) = a(x-x_0)+b(y-y_0)$. First we see what happens to equation (16.86) along the line $y = y_0$. Then $\mathbf{X}-\mathbf{X_0} = (x-x_0)\mathbf{I}$, and we

get

(16.87)
$$\lim_{x\to x_0}\left|\frac{f(x,y_0)-f(x_0,y_0)-a(x-x_0)}{x-x_0}\right| = \left|\frac{f(x,y_0)-f(x_0,y_0}{x-x_0})-a\right| = 0 \ ,$$

or

$$\frac{\partial f}{\partial x} = \lim_{x \to x_0} \frac{f(x,y_0) - f(x_0,y_0)}{x - x_0} = a \; . \tag{16.88}$$

In the same way, we see that $b = \partial f / \partial y$.

Now we turn to an argument for the chain rule in two dimensions.

The Chain Rule. Let $w = f(x,y)$ be a differentiable function defined in a region R. Suppose that γ is a differentiable curve in R given parametrically by $\mathbf{X} = \mathbf{X}(t)$, with $t = 0$ corresponding to $\mathbf{X_0}$. Then, considering $w = f(\mathbf{X}(t))$ as a function of t along γ, we have

$$\frac{dw}{dt} = (\nabla f) \cdot \frac{d\mathbf{X}}{dt} \; . \tag{16.89}$$

evaluated at $\mathbf{X_0}$. We start with the definition of differentiability. Let

$$\eta(t) = f(\mathbf{X}(t)) - f(\mathbf{X_0}) - \mathbf{L} \cdot (\mathbf{X}(\mathbf{t}) - \mathbf{X_0}) \; . \tag{16.90}$$

where we have written $\mathbf{L}$ for the gradient of w evaluated at $\mathbf{X_0}$. By (16.86),

$$\lim_{\mathbf{X}(\mathbf{t}) \to \mathbf{X_0}} \frac{|\eta(t)|}{|\mathbf{X}(\mathbf{t}) - \mathbf{X_0}|} = 0 \; . \tag{16.91}$$

Now, by continuity $\mathbf{X}(t) \to \mathbf{X_0}$ as $t \to 0$, and thus

$$\lim_{t \to 0} \left| \frac{\eta(t)}{t} \right| = \lim_{\mathbf{X}(t) \to \mathbf{X_0}} \frac{|\eta(t)|}{|\mathbf{X}(t) - \mathbf{X_0}|} \lim_{t \to 0} \frac{|\mathbf{X}(t) - \mathbf{X_0}|}{|t|} = 0 \; , \tag{16.92}$$

by (16.91), and the assumption of differentiability of $\mathbf{X}(t)$, which assure that the second limit on the right exists. Now, by the definition of η:

$$f(\mathbf{X}(t)) - f(\mathbf{X_0}) = \mathbf{L} \cdot (\mathbf{X}(\mathbf{t}) - \mathbf{X_0}) + \eta(t) \; . \tag{16.93}$$

Divide by t, and take the limit as $t \to 0$:

$$\frac{dw}{dt} = \lim_{t \to 0} \frac{f(\mathbf{X}(t)) - f(\mathbf{X_0})}{t} = \lim_{t \to 0} \mathbf{L} \cdot \frac{\mathbf{X}(\mathbf{t}) - \mathbf{X_0}}{t} + \lim_{t \to 0} \frac{\eta(t)}{t} = \mathbf{L} \cdot \frac{d\mathbf{X}}{dt} \; . \tag{16.94}$$

§16.4. Optimization

Now we turn to the technique for finding maxima and minima of a function $z = f(x,y)$ of two variables.

Definition 16.8 *If $f(x_0,y_0) \geq f(x,y)$ is at least as large as its value at all nearby points, we say that f has a* **local maximum** *at (x_0,y_0). More precisely, if, for some $a > 0$ we have*

$$f(x_0,y_0) \geq f(x,y) \tag{16.95}$$

for all (x,y) *within a distance* a *of* (x_0,y_0)*, then* (x_0,y_0) *is a local maximum point for* f*. Similarly, if instead we have*

$$f(x_0,y_0) \leq f(x,y) \tag{16.96}$$

for all (x,y) *sufficiently close to* (x_0,y_0)*, then* (x_0,y_0) *is a* **local minimum** *point for* f.

The first derivative test for functions of one variable gives us the following criterion:

Proposition 16.7 *Suppose that* $\mathbf{X_0}$ *is a local maximum (or minimum) for* f*. Then* $\nabla f = 0$.

To see this, pick a vector $\mathbf{V}$ and consider the line given by the equation $\mathbf{X}(t) = \mathbf{X_0} + t\mathbf{V}$. Then $f(\mathbf{X}(t))$ has a maximum at $t = 0$, so

$$\nabla f \cdot \mathbf{V} = \frac{d}{dt} f(\mathbf{X}(t))\big|_0 = 0 \ . \tag{16.97}$$

This can only be true for all $\mathbf{V}$ if $\nabla f = 0$.

Definition 16.9 *If* $\nabla f(x_0,y_0) = 0$ *we say that* (x_0,y_0) *is a* **critical point**.

Thus, to find the local maxima or minima of a function in a given region, one must look among the critical points.

Example 16.26 Find the critical points of the function $f(x,y) = x^3 + xy + y^2 - x$.

We calculate the components of the gradient:

$$\frac{\partial f}{\partial x} = 3x^2 + y - 1 \ , \quad \frac{\partial f}{\partial y} = x + 2y \ . \tag{16.98}$$

Now, we set these equal to zero and solve. The second equation gives $x = -2y$; substituting that in the first gives $12y^2 + y - 1 = 0$, which has the roots

$$y = \frac{-1 \pm \sqrt{1+48}}{24} \ , \quad \text{or} \quad y = \frac{1}{4}, \ -\frac{1}{3} \ . \tag{16.99}$$

Thus, the critical points are $(-1/2, 1/4)$, $(2/3, -1/3)$. But now, how can we tell whether or not we have a local maximum or a local minimum at either of these points? In fact, we may have neither; there is a third possibility: that along certain lines through the critical point, the value is a local maximum, and along other lines, the value is a local minimum. Such a point is a **saddle point**.

Example 16.27 Let $z = x^2 - y^2$. Then the origin is a critical point for z. Since $z = x^2$ along the line $y = 0$, z has a minimum at the origin on this line, but on the line $x = 0$, we have $z = -y^2$ which has a maximum at the origin along this line.

We distinguish among these points by using the second derivative test in one variable. In order to make clear what the criterion is, we first consider the case of a quadratic function.

Example 16.28 Let z be a quadratic function of the variables u, v: $z = au^2 + 2buv + cv^2$. The origin is a critical point. By completing the square we can discover what kind of critical point it is:

$$z = a(u^2 + 2\frac{b}{a}uv + \frac{b^2}{a^2}v^2) + (c - \frac{b^2}{a})v^2 = a(u + \frac{b}{a}v)^2 + \frac{ac-b^2}{a}v^2 \ . \tag{16.100}$$

Thus if both terms have positive coefficients, the origin is a minimum; if both terms have negative coefficients, the origin is a maximum, and if the signs differ, the origin is a saddle point. We call the expression $D = ac - b^2$ the **discriminant** of the quadratic function defining z. Notice that if $D > 0$, that the coefficients of (16.102) have the same sign, and if also $a > 0$, we have a minimum. and if $a < 0$, a maximum. If $D < 0$, the coefficients have different signs and we have a saddle point.

This example leads us directly to the general criterion by applying the second derivative test along each line through the critical point. For the function $z = f(x,y)$, let $f_{xx},\ f_{x,y},\ f_{yy}$ represent the second partial derivatives of f.

Proposition 16.8 *Suppose that $\nabla f(x_0,y_0) = 0$, that is (x_0,y_0) is a critical point. Then (evaluating at (x_0,y_0)):*

If $D = f_{xx}f_{yy} - (f_{xy})^2 < 0$ at a point (x_0,y_0), the f has a saddle point there.

If $D = f_{xx}f_{yy} - (f_{xy})^2 > 0$ and $f_{xx} > 0$, at a point (x_0,y_0), then f has a local minimum there.

If $D = f_{xx}f_{yy} - (f_{xy})^2 > 0$ and $f_{xx} < 0$, at a point (x_0,y_0), then f has a local maximum there.

If $D = 0$, we can conclude nothing. We note that when $D > 0$ the second derivative along all lines has the same sign, so we could check whether f_{yy} is greater or less than 0 instead, if that were easier.

To see this, choose a vector $\mathbf{V} = u\mathbf{I} + v\mathbf{J}$ and consider the function $f_{\mathbf{V}}(t) = f(\mathbf{X_0} + t\mathbf{V}) = f(x_0 + tu, y_0 + tv)$. Differentiating we find, by the chain rule,

$$\frac{d}{dt}f_{\mathbf{V}} = uf_x + vf_y\ , \quad \frac{d^2}{dt^2}f_{\mathbf{V}} = \frac{d}{dt}(uf_x + vf_y) = u\frac{df_x}{dt} + v\frac{df_y}{dt}\ . \tag{16.101}$$

We compute the second derivative by applying the chain rule to the functions $f_x,\ f_y$:

$$\frac{d^2}{dt^2}f_{\mathbf{V}} = u(uf_{xx} + vf_{yx}) + v(uf_{xy} + vf_{yy}) = u^2 f_{xx} + 2uvf_{xy} + v^2 f_{yy}\ . \tag{16.102}$$

If this is positive, then the function $f_{\mathbf{V}}$ has a minimum; that is the function f has a minimum along the line in the direction of $\mathbf{V}$. If this holds for all directions $\mathbf{V}$; that is, for all values of $u,\ v$, then f has a local minimum at (x_0,y_0). But, referring back to example 16.28, this is true if $D > 0,\ f_{xx} > 0$. Similarly, if $D < 0,\ f_{xx} < 0$, then f has a maximum along all lines through (x_0,y_0), so f has a local maximum there. However, if D and f_{xx} have different signs, then f has a local maximum in some directions, and a local minimum in others, so we have a saddle point.

Example 16.29 We continue with example 16.26. We found critical points at $P(-1/2, 1/4)$, $Q(2/3, -1/3)$. Differentiating the first partials (see (16.98)), we get

$$f_{xx} = 6x\ , \quad f_{xy} = 1\ , \quad f_{yy} = 2\ . \tag{16.103}$$

Thus

$$\text{at } P: \quad D = 6(-\frac{1}{2})(2) - 1^2 = -7\ , \quad \text{and at } Q: \quad D = 6(\frac{2}{3})(2) - 1^2 = 7\ , \tag{16.104}$$

so P is a saddle point, and since $f_{xx} = 4 > 0$, Q is a local minimum.

Example 16.30 Let $f(x,y) = x^2 + 2y^4 + xy + 4x + 2y$. Find the local maxima and minima of z. Does f have a global maximum or minimum?

First we find the critical points:

$$f_x = 2x + y + 4\ , \quad f_y = 8y^3 + x + 2\ . \tag{16.105}$$

To find the points where both are zero, we obtain $x = -8y^3 - 2$ from the second equation. Putting this in the first,we get

$$2(-8y^3 - 2) + y + 4 = 0\ , \quad \text{or} \quad -16y^3 + y = 0\ . \tag{16.106}$$

This has the solutions $y = 0, \pm 1/4$, so the critical points are $P(-2,0),\ Q(-17/8,1/4), R(-15/8,1/4)$. We now calculate the second derivatives:

$$f_{xx} = 2\ , \quad f_{xy} = 1\ , \quad f_{yy} = 24y^2\ . \tag{16.107}$$

Then $D = 48y^2 - 1$, which is positive at all of these points. Since f_{xx} is everywhere positive, these are all local minima. To determine the global minimum, we evaluate: $f(P) = -4,\ f(Q) = -4.0078,\ f(R) = -4.0078$. Thus the global minimum is -4.0078, attained at both Q and R. Everywhere else the function has a direction in which it is increasing, so it has no global maximum.

Notice, in these problems we have to solve several equations simultaneously, and usually they are not linear. There are no universal algorithms for solving such systems of equations, and we have to follow our intuition. Usually the technique of substitution works (although in the above problem, with other constants the cubic equation in (16.106) would be much more difficult). So, in general the procedure to follow is to look at the given equations to see if, in one of the equations one of the variables can be easily written in terms of the other. If so, substitute that expression in the other equation.

§16.4.1 The Method of Lagrange Multipliers

Let C be a curve in the plane, not going through the origin. Let's find the point on C which is closest to the origin. This amounts to finding the minimum value of $f(x,y) = x^2 + y^2$ on the curve C. If C is given parametrically by the equations $x = x(t),\ y = y(t)$, we know what to do: differentiate $f(x(t),y(t))$ and set the derivative equal to zero. But, if the curve is given implicitly by an equation $g(x,y) = c$, we don't want to solve the equation explicitly, and we don't have to. Looking at the condition

$$\frac{d}{dt} f(x(t),y(t) = 0 \quad \text{as} \quad \nabla f \cdot \frac{d\mathbf{X}}{dt} = 0 \tag{16.108}$$

we see that the requirement is that ∇f is orthogonal to the tangent to the curve at the minimizing point. But ∇g is orthogonal to its level set C everywhere, so at the minimizing point we have that ∇f and ∇g are collinear; that is, they are multiples of each other. Thus, we can solve the problem by fiinding the solution of the system

$$\nabla f = \lambda \nabla g\ , \quad g(x,y) = c\ . \tag{16.109}$$

This gives three scalar equations in three unknowns, which, in principle, can be solved. Of course the value of λ is not of interest, but is useful as an auxiliary to finding the values of $x,\ y$.

Example 16.31 Find the point on the line $3x - 2y = 1$ which is closest to the point (4,7).

Given the constraint $g(x,y) = 3x - 2y = 1$, we want to minimize $f(x,y) = (x-4)^2 + (y-7)^2$. The gradients are

$$\nabla f = 2(x-4)\mathbf{I} + 2(y-7)\mathbf{J} \quad \text{and} \quad \nabla g = 3\mathbf{I} - 2\mathbf{J} \ . \tag{16.110}$$

These gradients are collinear at the minimizing point, so we have to solve the equations

$$2(x-4) = 3\lambda \ , \quad 2(y-7) = 2\lambda \quad \text{and} \quad 3x - 2y = 1 \ . \tag{16.111}$$

We can eliminate λ from the first two equations:

$$4(x-4) = 6\lambda = 6(y-7) \quad \text{so that} \quad 4x - 6y = -26 \ . \tag{16.112}$$

Now we have simultaneous linear equations in x and y which we can solve, getting the point (16,47/2). We note that the Lagrangian equations (16.109) just say that the line from this point to (4,7) has to be orthogonal to the given line; something we knew from geometry.

Example 16.32 Find the maximum value of $f(x,y) = xy$ on the ellipse $x^2 + 4y^2 = 1$.

Let $g(x,y) = x^2 + 4y^2$. We calculate the gradients: $\nabla f = y\mathbf{I} + x\mathbf{J}$ and $\nabla g = 2x\mathbf{I} + 8y\mathbf{J}$. At the point on the ellipse at which we have the maximum, we have ∇f orthogonal to the tangent to the ellipse, so is collinear with ∇g. Thus we have the equation $\nabla f = \lambda \nabla g$ for some λ. This gives the scalar equations

$$y = 2\lambda x \ , \quad x = 8\lambda y \ , x^2 + 8y^2 = 1 \ . \tag{16.113}$$

We can eliminate λ by dividing the first equation by the second:

$$\frac{y}{x} = \frac{2\lambda x}{8\lambda y} = \frac{x}{4y} \quad \text{giving} \quad x^2 = 4y^2 \ . \tag{16.114}$$

Substituting that in the last equation gives $4y^2 + 4y^2 = 1$, so that $y = \pm 1/(2\sqrt{2})$. Then

$$x^2 = 4y^2 = \frac{4}{8} \quad \text{so that} \quad x = \pm \frac{1}{\sqrt{2}} \ . \tag{16.115}$$

The possible values of $f(x,y) = xy$ at these points are $\pm 1/4$, so the maximum value of f is 1/4, and its minimum is $-1/4$.

The parameter λ, called the **Lagrange multiplier**, serves the purpose of finding a relation between x and y which is a consequence of the optimization. The value of λ is not important, but in some cases it may make the problem easier to first determine λ.

To summarize: given the problem: minimize (or maximize) a function $f(x,y)$ subject to a constraint $g(x,y) = c$. We observe that the chain rule tells us that, at the optimizing point, ∇f is orthogonal to the tangent to the level set of g. But so is ∇g, so we must have $\nabla f = \lambda \nabla g$ for some λ. Solve this equation in conjunction with $g(x,y) = c$ to find the point. This method (of Lagrange multipliers) works in three dimensions as well.

Proposition 16.9 *Suppose that $w = f(x,y,z)$ is a differentiable function, and we wish to find its maxima and minima subject to a constraint $g(x,y,z) = c$. At an optimizing point P there is a λ such that*

$$\nabla f(P) = \lambda \nabla g(P) \ , \quad g(x,y,z) = c \ . \tag{16.116}$$

These equations give a system of four equations in four unknowns which, in typical circumstances, have only a finite number of solutions. The maximum (minimum) of the function must occur at one of these points.
To see why this is true, we follow the two dimensional argument. Let S be the level surface $g(x,y,z) = c$. Let C be a curve through P lying in the surface S. Then f is optimized along C, so that the derivative of f along the curve is zero at P. But this just says that $\nabla f(P)$ is orthogonal to the tangent to the curve. Since every vector in the tangent plane to S is the tangent vector to such a curve, $\nabla f(P)$ is orthogonal to the tangent plane to S. But so is $\nabla g(P)$, so $\nabla f(P)$ and $\nabla g(P)$ must be colinear.

Example 16.33 Find the point on the plane $2x+3y+z=1$ closest to the point $(1,-1,0)$.

Here the constraint is $g(x,y,z) = 2x+3y+z = 1$ and the function to be minimized is $f(x,y,z) = (x-1)^2+(y+1)2+z^2$. Taking the gradients and introducing the Lagrange multipier, we are led to the equations

(16.117) $$2(x-1) = 2\lambda \ , \quad 2(y+1) = 3\lambda \ , 2z = \lambda \ , \quad 2x+3y+z = 1 \ .$$

We use the first three equations to express the variables in terms of λ, and then use the last to solve for λ:

(16.118) $$x = \lambda+1 \ , \quad y = \frac{3\lambda-2}{2} \ , \quad z = \frac{\lambda}{2} \ ,$$

so that

(16.119) $$2(\lambda+1)+3\frac{3\lambda-2}{2}+\frac{\lambda}{2} = 1 \ .$$

This gives $\lambda = 1/7$. Substituting into equations 16.118), we find the desired point to be (1/7, -11/14, 1/14).

Example 16.34 Farmer Brown wishes to enclose a rectangular coop of 1000 square feet. He will build three sides of brick, costing \$25 per linear foot, and the fourth of chain link fence, at \$ 15 per linear foot. What should the dimensions be to minimize the cost?

Let x and y be the dimensions of the coop, where x represents the sides, both of which are to be of brick. The constraint is $g(x,y) = xy = 1000$, and the cost function is $C = 25(2x+y)+15y$. We have $\nabla C = 50\mathbf{I}+40\mathbf{J}$, and $\nabla g = y\mathbf{I}+x\mathbf{J}$. The equations to solve are:

(16.120) $$50 = \lambda y \ , \quad 40 = \lambda x \ , xy = 1000 \ ,$$

so

(16.121) $$1000 = xy = (\frac{50}{\lambda})(\frac{40}{\lambda}) \ ,$$

or $\lambda^2 = (50)(40)/1000 = 2$, giving $\lambda = \sqrt{2}$. Then $x = 40/\sqrt{2} = 20\sqrt{2}$, $y = 50/\lambda = 25\sqrt{2}$.

Many problems involve finding the maximum or minimum of a function of many variables subject to many constraints The technique of Lagrange multipliers works in this general context, but - of course - is much more difficult to employ. To give a sense of the general procedure, we state the proposition in the case of a function of three variables with two constraints.

Proposition 16.10 *To find the extreme values of a function* $f(x,y,z)$ *subject to two constraints (say along a curve),* $g(x,y,z) = c,\ h(x,y,z) = d$*, we have to solve the five equations in the five unknowns* x, y, z, λ, μ*:*

$$\nabla f(P) = \lambda \nabla g(P) + \mu \nabla h\,, \quad g(x,y,z) = c, \quad h(x,y,z) = d\ . \tag{16.122}$$