7 Chain rules

Chain rule I

Let $x = x(t)$, $y = y(t)$ and $z = f(x, y)$ be differentiable functions. Then $z = f(x(t), y(t))$ is a differentiable function of t and

$$\frac{dz}{dt} = \frac{\partial z}{\partial x}\frac{dx}{dt} + \frac{\partial z}{\partial y}\frac{dy}{dt},$$

where the ordinary derivatives are evaluated at t and the partial derivatives are evaluated at (x, y).

Proof.

A change of Δt in t produces changes of Δx in x and Δy in y . These, in turn, produce a change of Δz in z. Since $z = f(x, y)$ is differentiable, we have

$$\Delta z = \frac{\partial f}{\partial x}\Delta x + \frac{\partial f}{\partial y}\Delta y + \varepsilon_1 \Delta x + \varepsilon_2 \Delta y,$$

where $\varepsilon_1, \varepsilon_2 \to 0$ as $(\Delta x, \Delta y) \to (0, 0)$. Dividing both sides of this equation by Δt, we have

$$\frac{\Delta z}{\Delta t} = \frac{\partial f}{\partial x}\frac{\Delta x}{\Delta t} + \frac{\partial f}{\partial y}\frac{\Delta y}{\Delta t} + \varepsilon_1 \frac{\Delta x}{\Delta t} + \varepsilon_2 \frac{\Delta y}{\Delta t},$$

If we now let $\Delta t \to 0$, then

$$\Delta x = x(t + \Delta t) - x(t) \to 0,$$

because x is differentiable, therefore, continuous. Similarly,

$$\Delta y \to 0.$$

This, in turn, means that $\varepsilon_1, \varepsilon_2 \to 0$, so

$$\begin{aligned}
\frac{dz}{dt} &= \lim_{\Delta t\to 0} \frac{\Delta z}{\Delta t} \\
&= \frac{\partial f}{\partial x} \lim_{\Delta t\to 0} \frac{\Delta x}{\Delta t} + \frac{\partial f}{\partial y} \lim_{\Delta t\to 0} \frac{\Delta y}{\Delta t} + \lim_{\Delta t\to 0} \varepsilon_1 \lim_{\Delta t\to 0} \frac{\Delta x}{\Delta t} \\
&\qquad + \lim_{\Delta t\to 0} \varepsilon_2 \lim_{\Delta t\to 0} \frac{\Delta y}{\Delta t} \\
&= \frac{\partial z}{\partial x}\frac{dx}{dt} + \frac{\partial z}{\partial y}\frac{dy}{dt} + 0 \cdot \frac{dx}{dt} + 0 \cdot \frac{dy}{dt} \\
&= \frac{\partial z}{\partial x}\frac{dx}{dt} + \frac{\partial z}{\partial y}\frac{dy}{dt}.
\end{aligned}$$

Hence $z = f(x(t), y(t))$ is a differentiable function of t and

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

◀

Example 38. If $z = x^2y + 3xy^4$, where $x = \sin 2t$ and $y = \cos t$, find dz/dt when $t = 0$.

The derivative in Example(38) can be interpreted as the rate of change of z with respect to t as the point (x, y) moves along the curve C with parametric equations $x = \sin 2t$, $y = \cos t$, (See the figure given below.)

In particular, when $t = 0$, the point (x, y) is (0,1) and $dz/dt = 6$ is the rate of change as we move along the curve C through (0,1). If, for instance,

$$z = T(x, y) = x^2 + 3xy^4$$

represents the temperature at the point (x, y), then the composite function $z = T(\sin 2t, \cos t)$ represents the temperature

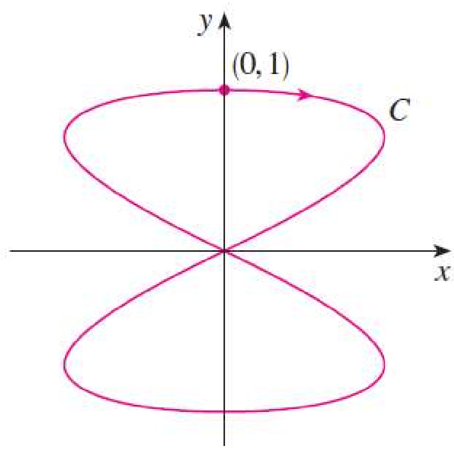


Figure 27: The curve $x = \sin 2t$, $y = \cos t$

at points on C and the derivative dz/dt represents the rate at which the temperature changes along C.

Chain rule II

Let $x = x(s,t)$, $y = y(s,t)$ and $z = f(x,y)$ be differentiable functions. Then $z = f(x(s,t), y(s,t))$ be a differentiable function of s, t. And

$$\frac{\partial z}{\partial s} = \frac{\partial z}{\partial x}\frac{\partial x}{\partial s} + \frac{\partial z}{\partial y}\frac{\partial y}{\partial s}$$
$$\frac{\partial z}{\partial t} = \frac{\partial z}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial z}{\partial y}\frac{\partial y}{\partial t}.$$

Example 39. If $z = e^x \sin y$, where $x = st^2, y = s^2 t$, find $\partial z/\partial s, \partial z/\partial t$.

It is easy to extend the chain rule to the general situation in which a dependent variable z is a function of n intermediate variables $x_1, ..., x_n$ each of which is, in turn, a function of m independent variables $t_1, ..., t_m$.

Example 40. Write out the Chain Rule for the case, where $w = f(x, y, z, t)$ and $x = x(u, v), y = y(u, v), z = z(u, v), t = t(u, v)$.

Example 41. If $g(s, t) = f(s^2 - t^2, t^2 - s^2)$ and f is differentiable, show that g satisfies the equation:

$$t\frac{\partial g}{\partial s} + s\frac{\partial g}{\partial t} = 0.$$

Example 42. If $z = f(x, y)$ and f has continuous second-order partial derivatives and $x = r^2 + s^2, y = 2rs$, find $\partial z/\partial r, \partial^2 z/\partial r^2$.

Implicit Differentiation

(as an application of the chain rule.)

Case I: $F(x, y) = 0$, where $y = f(x)$.

Implicit Function Theorem I

If F is defined on a disk containing (a, b), where $F(a, b) = 0, F_y(a, b) \neq 0$, and F_x, F_y are continuous on the disk, then the equation $F(x, y) = 0$ defines y as a function of x near the point (a, b) and the derivative of this function is given by

$$\frac{dy}{dx} = -\frac{\frac{\partial F}{\partial x}}{\frac{\partial F}{\partial y}} = -\frac{F_x}{F_y}.$$

Proof.

Consider the function

$$z = F(x, y) = F(x, f(x)).$$

By the chain rule, we have

$$\frac{dz}{dx} = \frac{\partial z}{\partial x}\frac{dx}{dx} + \frac{\partial z}{\partial y}\frac{dy}{dx} = F_x(x, y)\frac{dx}{dx} + F_y(x, y)\frac{dy}{dx}.$$

Since $z = F(x, y) = 0$ for all x in the domain of f, we obtain

$$\frac{dz}{dx} = 0$$

and we have

$$F_x(x, y)\frac{dx}{dx} + F_y(x, y)\frac{dy}{dx} = 0.$$

Therefore, if $F_y(x, y) \neq 0$, then

$$\frac{dy}{dx} = -\frac{F_x(x, y)}{F_y(x, y)}.$$

◀

Example 43. Find y' if $x^3 + y^3 = 6xy$.

<u>Case II:</u> $F(x, y, z) = 0$, where $z = f(x, y)$.

Implicit Function Theorem II

If F is defined within a sphere containing (a, b, c), where $F(a, b, c) = 0, F_z(a, b, c) \neq 0$, and F_x, F_y, F_z are continuous inside the sphere, then the equation $F(x, y, z) = 0$ defines z as a function of x, y near the point (a, b, c) and the derivative of this function is differentiable, with partial derivatives given by

$$\frac{\partial z}{\partial x} = -\frac{F_x}{F_z}, \quad \frac{\partial z}{\partial y} = -\frac{F_y}{F_z}.$$

Example 44. Find $\partial z/\partial x$ and $\partial z/\partial y$ if $x^3 + y^3 + z^3 + 6xyz = 1$.

8 Directional derivatives and gradient vector

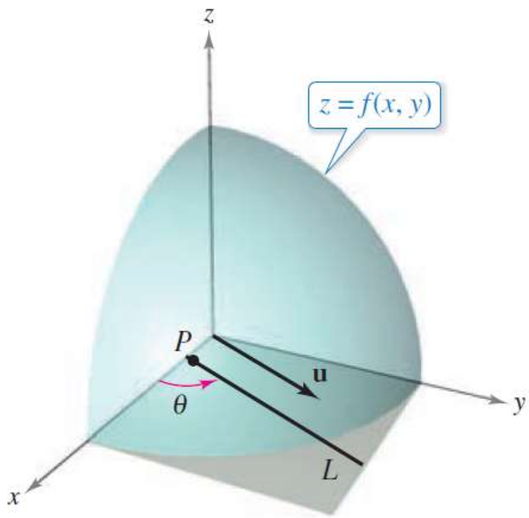
The partial derivatives of a function f tell us the rate of change of f in the direction of the coordinate axes.

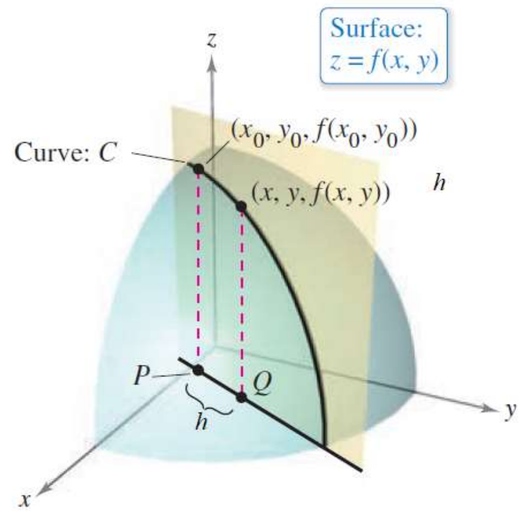
How can we measure the rate of change of f
in other directions?

In order to formally define the derivative in a particular direction of motion, we want to represent the change in f for a given unit change in the direction of motion.

We can represent this unit change in direction with a unit vector, say $u = (a, b)$. This unit vector helps us to "mark off" units on the line. A vector equation for the line through (x_0, y_0) in this direction is

$$v(h) = (x_0 + ha, y_0 + hb).$$





Because u is a unit vector, the value of h is precisely the distance along the line from (x_0, y_0) to (x_0+ha, y_0+hb). Indeed,

$$\begin{aligned}
&\|(x_0 + ha, y_0 + hb) - (x_0, y_0)\| \\
=&\|(ha, hb)\| \\
=&|h|\|(a, b)\| \\
=&|h|. \qquad (\because (a, b) \text{ is a unit vector})
\end{aligned}$$

If we move a distance in the direction of u from a fixed point (x_0, y_0), we then arrive at the new point $(x_0 + ha, y_0 + hb)$. It now follows that the slope of the secant line to the curve on the surface through (x_0, y_0) in the direction of u through the points (x_0, y_0) and $(x_0 + ha, y_0 + hb)$ is

$$m_{\text{sec}} = \frac{f(x_0 + ha, y_0 + hb) - f(x_0, y_0)}{h}. \tag{1}$$

To get the instantaneous rate of change of f in the direction $u = (a, b)$, we must take the limit of the quantity in Equation (1) as $h \to 0$. Doing so results in the formal definition of the directional derivative.

Directional derivatives

The directional derivative of f at (x_0, y_0) in the direction of a unit vector $u = (a, b)$ is

$$D_u f(x_0, y_0) = \lim_{h \to 0} \frac{f(x_0 + ha, y_0 + hb) - f(x_0, y_0)}{h}$$

if this limit exists.

Partial derivatives and directional derivatives

If $u = i = (1, 0)$, then

$$D_i f(x_0, y_0) = \lim_{h \to 0} \frac{f(x_0 + h, y_0) - f(x_0, y_0)}{h} = f_x(x_0, y_0).$$

If $u = j = (0, 1)$, then

$$D_j f(x_0, y_0) = \lim_{h \to 0} \frac{f(x_0, y_0 + h) - f(x_0, y_0)}{h} = f_y(x_0, y_0).$$

Thus,

$$D_i f = f_x, \qquad D_j f = f_y.$$

In other words, the partial derivatives of f with respect to x and y are just special cases of the directional derivative.

Method of finding directional derivatives

It is time consuming to find the directional derivative using the above definition. However, we can find a way to evaluate directional derivatives without resorting to the limit definition.

Theorem 8.1. If f is a differentiable function of x and y, then f has a directional derivative at (x_0, y_0) in the direction of a **unit vector** $u = (a, b)$ and

$$D_u f(x_0, y_0) = f_x(x_0, y_0)a + f_y(x_0, y_0)b. \tag{9}$$

Proof.

For a fixed point (x_0, y_0) let

$$x = x_0 + ha, \qquad y = y_0 + hb.$$

Then differentiating both x and y with respect to h, we obtain

$$x' = a, \qquad y' = b.$$

Let

$$g(h) = f(x_0 + ha, y_0 + hb) = f(x, y).$$

Because f is differentiable, we can apply the Chain Rule to obtain

$$g'(h) = f_x(x, y)x'(h) + f_y(x, y)y'(h) = f_x(x, y)a + f_y(x, y)b.$$

If $h = 0$, then $x = x_0, y = y_0$, and so,

$$g'(0) = f_x(x_0, y_0)a + f_y(x_0, y_0)b.$$

By the definition of $g'(h)$, it is also true that

$$\begin{aligned} g'(0) &= \lim_{h \to 0} \frac{g(h) - g(0)}{h} \\ &= \lim_{h \to 0} \frac{f(x_0 + ha, y_0 + hb) - f(x_0, y_0)}{h} \\ &= D_u f(x_0, y_0). \end{aligned}$$

Therefore, we see that

$$D_u f(x_0, y_0) = f_x(x_0, y_0)a + f_y(x_0, y_0)b. \qquad \blacktriangleleft$$

Remark 1.

To use the theroem, we must have a unit vector in the direction of motion. In the event that we have a direction prescribed by a non-unit vector, we must first scale the vector to have length 1.

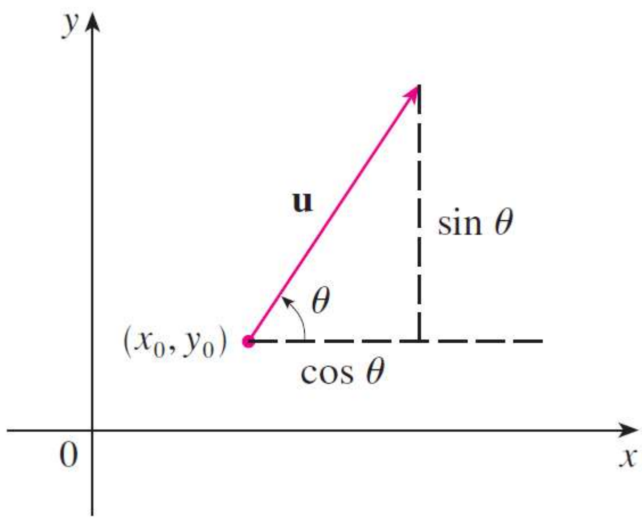


Figure 28: A unit vector $\vec{u} = (a, b) = (\cos\theta, \sin\theta)$

If the unit vector u makes an angle θ with the positive x-axis (as in Figure 28), then we can write $u = (\cos\theta, \sin\theta)$ and Formula (9) becomes

$$D_u f(x_0, y_0) = f_x(x_0, y_0)\cos\theta + f_y(x_0, y_0)\sin\theta.$$

Example 45. Find the directional derivative $D_u f(x, y)$ if

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

and u is the unit vector given by angle $\theta = \pi/6$. What is $D_u f(1, 2)$?

Example 46. Find the directional derivative of

$$f(x,y) = x^2 \sin 2y$$

at $(1, \pi/2)$ in the direction of

$$v = 3i - 4j.$$

The Gradient Vector

Notice that the directional derivative of a differentiable function can be written as the dot product of two vectors:

$$\begin{aligned} D_u f(x,y) &= f_x(x,y)a + f_y(x,y)b \\ &= (f_x(x,y), f_y(x,y)) \cdot (a,b) \\ &= (f_x(x,y), f_y(x,y)) \cdot u. \end{aligned}$$

The first vector in this dot product is called the **gradient** of f and is denoted by

$$\operatorname{grad} f \quad \text{or} \quad \nabla f.$$

Gradient

If f is a function of two variables x and y, then the gradient of f is the vector function ∇f defined by

$$\nabla f(x,y) = (f_x(x,y), f_y(x,y)) = f_x i + f_y j.$$

Example 47. Find the gradient f if

$$f(x,y) = \sin x + e^{xy}.$$

What is $\nabla f(0,1)$?

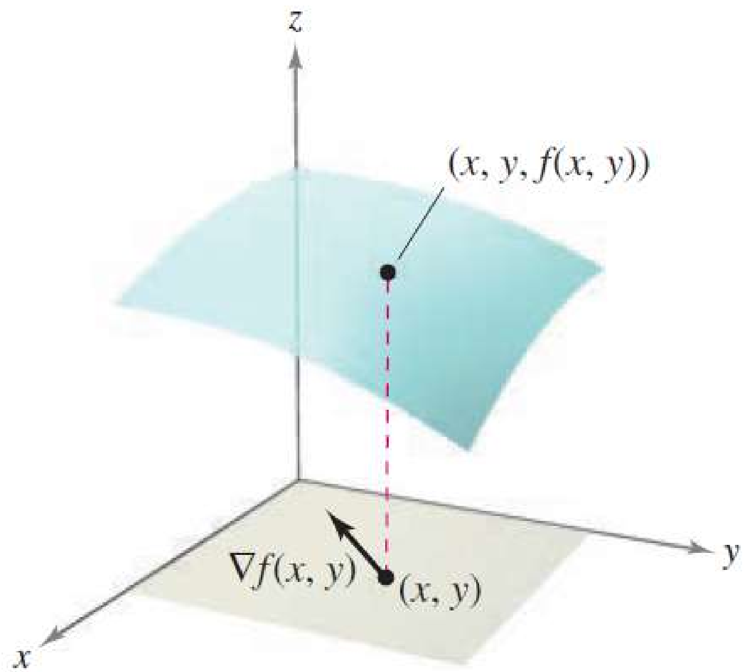


Figure 29: The gradient of f: a vector in the xy- plane.

With this notation for the gradient vector, we can rewrite the expression (9) for the directional derivative of a differentiable function as

$$\boxed{D_u f(x, y) = \nabla f(x, y) \cdot u.}$$

This expresses the directional derivative in the direction of u as the scalar projection of the gradient vector onto u.

Example 48 (Using a gradient vector to find a directional derivative). Find the directional derivative of the function $f(x, y) = x^2y^3 - 4y$ at the point $(2, -1)$ in the direction of the vector $v = 2i + 5j$.

Functions of Three Variables

For functions of three variables we can define directional derivatives in a similar manner. Again $D_u f(x, y, z)$ can be interpreted as the rate of change of the function in the direction of a unit vector $\vec{u}$.

Directional derivative

The directional derivative of f at $\vec{x}_0 = (x_0, y_0, z_0)$ in the direction of a unit vector $\vec{u} = (a, b, c)$ is

$$D_u f(\vec{x}_0) = \lim_{h \to 0} \frac{f(\vec{x}_0 + h\vec{u}) - f(\vec{x}_0)}{h}$$

if this limit exists.

This is reasonable because the vector equation of the line through $\vec{x}_0$ in the direction of the vector $\vec{u}$ is given by

$$\vec{x}_0 + t\vec{u}$$

and so $f(\vec{x}_0 + h\vec{u})$ represents the value of f at a point $\vec{x}_0$ on this line.

If $f(x, y, z)$ is differentiable and $\vec{u} = (a, b, c)$, then we can prove by the same method as in the case of a function of two variables that

$$D_{\vec{u}} f(x, y, z) = f_x(x, y, z)a + f_y(x, y, z)b + f_z(x, y, z)c. \tag{10}$$

For a function f of three variables, the gradient vector is

$$\boxed{\nabla f(x,y,z) = (f_x(x,y,z), f_y(x,y,z), f_z(x,y,z)).}$$

or

$$\boxed{\nabla f = (f_x, f_y, f_z) = \frac{\partial f}{\partial x} i + \frac{\partial f}{\partial y} j + \frac{\partial f}{\partial z} k}$$

Then, just as with functions of two variables, Formula (10) for the directional derivative can be rewritten as

$$\boxed{D_{\vec{u}} f(x,y,z) = \nabla f_x(x,y,z) \cdot \vec{u}.}$$

Example 49. If $f(x,y,z) = x \sin yz$,

(a) find the gradient of f

(b) find the directional derivative of f at (1, 3, 0) in the direction of $v = i + 2j - k$.

Maximizing the Directional Derivative

Suppose we have a function of two or three variables and we consider all possible directional derivatives of f at a given point. These give the rates of change of f in all possible directions. We can then ask the questions:

- In which of these directions f does change fastest or which one is the direction of maximum increase of f and
- What is the maximum rate of change?

The answers are provided by the following theorem.

Theorem 8.2. Suppose f is a differentiable function of two or three variables. Then the maximum value of the directional derivative $D_{\vec{u}f(x)}$ is $\|\nabla f(x)\|$ and it occurs when $\vec{u}$ has the same direction as the gradient vector $\nabla f(x)$.

Proof.

We have

$$\begin{aligned} D_{\vec{u}} f &= \nabla f \cdot \vec{u} \\ &= \|\nabla f\| \|\vec{u}\| \cos\theta \\ &= \|\nabla f\| \cos\theta, \end{aligned}$$

where θ is the angle between ∇f and $\vec{u}$. The maximum value of $\cos\theta$ is 1 and this occurs when $\theta = 0$. Therefore the maximum value of $D_{\vec{u}} f$ is $\|\nabla f\|$ and it occurs when $\theta = 0$, that is, when $\vec{u}$ has the same direction as ∇f. ◀

Example 50. Let $f(x, y) = xe^y$.

(a) find the directional derivative of f at $P(2, 0)$ in the direction from P to $Q(1/2, 2)$.

(b) In what direction does f have the maximum rate of change? What is this maximum rate of change?

Example 51. Suppose that the temperature in degrees Celsius on the surface of a metal plate is given by

$$T(x, y) = 20 - 4x^2 - y^2,$$

where x and y are measured in centimeters. In what direction from $(2, -3)$ does the temperature increase most rapidly? What is this rate of increase?

Properties of the Gradient

We are now in a position to draw some interesting and important conclusions about the gradient.

First, suppose that $\nabla f_P \neq 0$ and let u be a unit vector (see the figure given below).

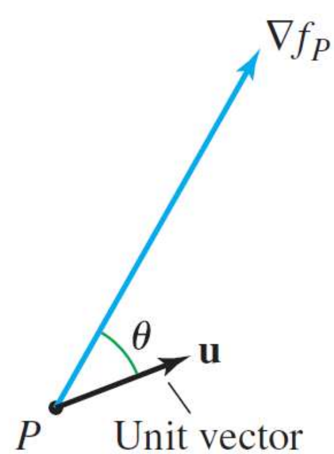


Figure 30: $D_u f(P) = \|\nabla f_P\| \cos\theta$.

We know that

$$D_u f(P) = \nabla f_P \cdot u = \|\nabla f_P\| \cos\theta,$$

where θ is the angle between ∇f_P and u. In other words, The rate of change in a given direction varies with the cosine of the angle θ between the gradient and the direction.

Because the cosine takes values between -1 and 1, we have

$$-\|\nabla f_P\| \leq D_u f(P) \leq \|\nabla f_P\|.$$

Since $\cos\theta = 1$, the maximum value of $D_u f(P)$ occurs for $\theta = 0$ – that is, when u points in the direction of ∇f_P. In other words,

The gradient vector ∇f_P points in the direction of the maximum rate of increase, and this maximum rate is $\|\nabla f_P\|$.

Similarly, f decreases most rapidly in the opposite direction, $-\nabla f_P$, because $\cos\theta = -1$ for $\theta = \pi$. The rate of maximum decrease is $-\|\nabla f_P\|$. The directional derivative is zero in directions orthogonal to the gradient because $\cos(\pi/2) = \theta$.

Another key property is that

Gradient vectors are normal to level curves.

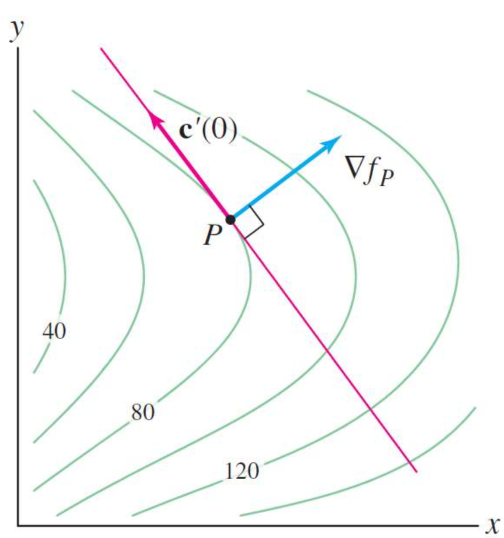


Figure 31: Contour map of $f(x, y)$. The gradient at P is orthogonal to the level curve through P.

Proof.

To prove this, suppose that P lies on the level curve $f(x, y) = k$. We parametrize this level curve by a path $c(\mathrm{t})$ such that $c(0) = P$ and $c'(0) \neq 0$ (this is possible whenever $\nabla f_P \neq 0$). Then $f(c(t)) = k$ for all t , so by the Chain Rule,

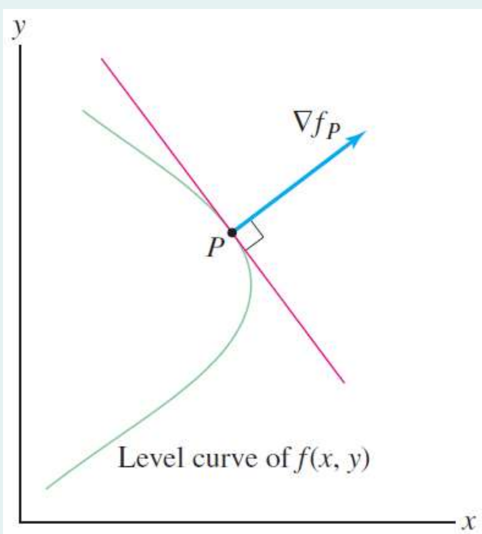
$$\nabla f_P \cdot c'(0) = \frac{d}{dt} f(c(0)) = \frac{d}{dt} k = 0.$$

This proves that ∇f_P is orthogonal to $c'(0)$, and since $c'(0)$ is tangent to the level curve, we conclude that ∇f_P is normal to the level curve. ◀

For functions of three variables, a similar argument shows that ∇f_P is normal to the level surface $f(x, y, z) = k$ through P.

Graphical insight

At each point P, there is a unique direction in which $f(x, y)$ increases most rapidly (per unit distance). This chosen direction is perpendicular to the level curves and that it is specified by the gradient vector.



For most functions, however, the direction of maximum rate of increase varies from point to point.

In summary,

- $D_u f(P) = \nabla f_P \cdot u = \|\nabla f_P\| \cos\theta$.
  That is, the rate of change in a given direction varies with the cosine of the angle θ between the gradient and the direction.
- The gradient vector ∇f_P points in the direction of the maximum rate of increase, and this maximum rate is $\|\nabla f_P\|$.
- The gradient vector $-\nabla f_P$ points in the direction of the maximum rate of decrease, and this maximum rate of decrease is $-\|\nabla f_P\|$.
- Gradient vector ∇f_P is normal to level curve (or surface) of f at P.

Tangent Planes to Level Surfaces

Suppose S is a surface with equation $F(x, y, z) = k$, that is, it is a level surface of a function F of three variables, and let $P(x_0, y_0, z_0)$ be a point on S. Let C be any curve that lies on the surface S and passes through the point P. The curve C is described by a continuous vector function $\vec{r}(t) = (x(t), y(t), z(t))$. Let t_0 be the parameter value corresponding to P; that is, $\vec{r}(t_0) = (x_0, y_0, z_0)$. Since C lies on s, any point $(x(t), y(t), z(t))$ must satisfy the equation of s, that is,

$$f(x(t), y(t), z(t)) = K$$

If x, y, z are differentiable functions of t and F is also differentiable, then we can use the Chain Rule to differentiate its both sides to obtain

$$\nabla F \cdot \vec{r}'(t) = 0.$$

In particular, when $t = t_0$ we have $\vec{r}(t_0) = (x_0, y_0, z_0)$, so

$$\nabla F(x_0, y_0, z_0) \cdot \vec{r}'(t_0) = 0.$$

This equation says that the gradient vector $\nabla F(x_0, y_0, z_0)$ at P is perpendicular to the tangent vector $\vec{r}'(t_0)$ to any curve C on S that passes through P. (See the figure given below.)

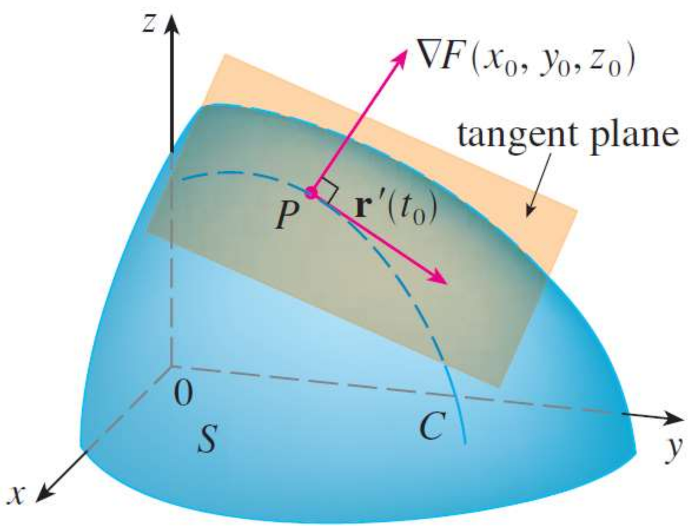


Figure 32

If $\nabla F(x_0, y_0, z_0) \neq 0$, it is therefore natural to define the **tangent plane to the level surface** $F(x_0, y_0, z_0) = k$ at $P(x_0, y_0, z_0)$ as the plane that passes through P and has normal vector $\nabla F(x_0, y_0, z_0)$. Using the standard equation of a plane, we can write the equation of this tangent plane as

$$\begin{aligned} F_x(x_0, y_0, z_0)(x - x_0) + F_y(x_0, y_0, z_0)(y - y_0) \\ + F_z(x_0, y_0, z_0)(z - z_0) = 0. \end{aligned} \tag{11}$$

The normal line to S at P is the line passing through P and perpendicular to the tangent plane. The direction of the normal line is therefore given by the gradient vector $\nabla F(x_0, y_0, z_0)$ and so, its symmetric equations are

$$\boxed{\frac{x - x_0}{F_x(x_0, y_0, z_0)} = \frac{y - y_0}{F_y(x_0, y_0, z_0)} = \frac{z - z_0}{F_z(x_0, y_0, z_0)}}$$

In the special case in which the equation of a surface S is of the form $z = f(x, y, z)$ (that is, is the graph of a function f of two variables), we can rewrite the equation as

$$F(x, y, z) = f(x, y, z) - z = 0$$

and regard S as a level surface (with $k = 0$) of F. Then

$$\begin{aligned} F_x(x_0, y_0, z_0) &= f_x(x_0, y_0) \\ F_y(x_0, y_0, z_0) &= f_y(x_0, y_0) \\ F_z(x_0, y_0, z_0) &= -1 \end{aligned}$$

so Equation (11) becomes

$$\boxed{f_x(x_0, y_0)(x - x_0) + f_y(x_0, y_0)(y - y_0) - (z - z_0) = 0.}$$

Thus our new, more general, definition of a tangent plane is consistent with the definition that was given earlier.

Example 52.

> Find the equations of the tangent plane and normal line at the point $(-2, 1, -3)$ to the ellipsoid
>
> $$\frac{x^2}{4} + y^2 + \frac{z^2}{9} = 3.$$

9 Maximum and minimum values

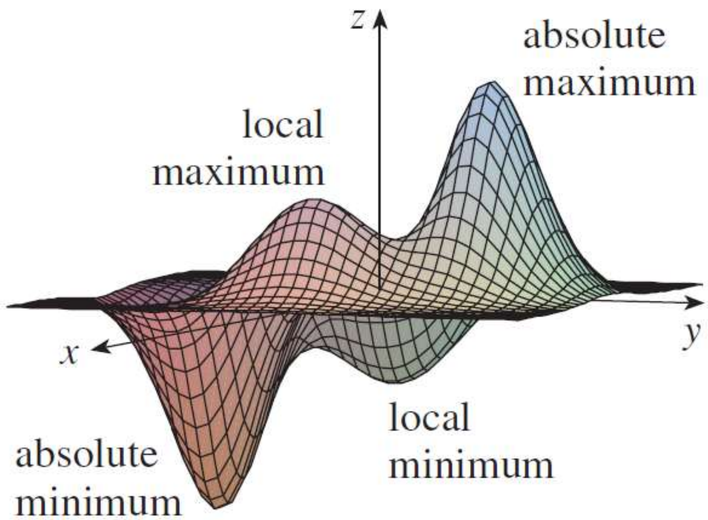


Figure 33

Local extrema

Let f be a function of two variables x and y.

- The function f has a **local maximum** at a point (a, b) provided that $f(x, y) \leq f(a, b)$ for all points (x, y) in some disk with center at (a, b). In this situation we say that $f(a, b)$ is a **local maximum value**.
- The function f has a **local minimum** at a point (a, b) provided that $f(x, y) \geq f(a, b)$ for all points (x, y) in some disk with center at (a, b). In this situation we say that $f(a, b)$ is a **local minimum value**.

We use the term **extremum point** to refer to any point (a, b) at which f has a maximum or minimum. In addition, the function value $f(a, b)$ at an extremum is called an **extremal value**.

Absolute extrema

Let f be a function of two variables x and y.

- An **absolute maximum point** is a point (a, b) for which $f(x, y) \leq f(a, b)$ for all points (x, y) in the domain of f. The value of f at an absolute maximum point is the **maximum value** of f.

- An **absolute minimum point** is a point such that $f(x, y) \geq f(a, b)$ for all points (x, y) in the domain of f. The value of f at an absolute minimum point is the **maximum value** of f.

Critical points

A **critical point** (a, b) of a function $f = f(x, y)$ is a point in the domain of f at which one of the following is true:

1. $f_x(a, b) = 0$ and $f_y(a, b) = 0$,
2. $f_x(a, b)$ or $f_y(a, b)$ fails to exist.

Stationary points

A **stationary point** (a, b) of a function $f = f(x, y)$ is a point in the domain of f at which $f_x(a, b) = f_y(a, b) = 0$.

Theorem 9.1 (Fermat)**.** If f has a local maximum or minimum at (a, b) and the first-order partial derivatives of f exist there, then $f_x(a, b) = 0$ and $f_y(a, b) = 0$.

Proof.

Let $g(x) = f(x, b)$. If f has a local maximum (or minimum) at (a, b), then g has a local maximum (or minimum) at a, so by Fermat's Theorem for functions of one variable,

$$g'(a) = 0.$$

But

$$g'(a) = f_x(a, b)$$

and so

$$f_x(a, b) = 0.$$

Similarly, by applying Fermat's Theorem to the function $G(y) = f(a, y)$, we obtain

$$f_y(a, b) = 0.$$

◀

This theorem says that if f has a local maximum or minimum at (a, b), then (a, b) is a critical point of f. At a critical point, a function could have a local maximum or a local minimum or neither.

Geometric interpretation of Fermat's theorem

If we put $f_x(a, b) = 0$ and $f_y(a, b) = 0$ in the equation of a tangent plane:

$$z = f(x_0, y_0) + f_x(x_0, y_0)(x - x_0) + f_y(x_0, y_0)(y - y_0),$$

we get

$$z = z_0.$$

Thus the geometric interpretation of Fermat's theorem is as follows:

Geometric interpretation of Fermat's theorem

If the graph of f has a tangent plane at a local maximum or minimum, then the tangent plane must be horizontal.

Example 53. Investigate the critical points of the function:

$$f(x, y) = x^2 + y^2 - 2x - 6y + 14.$$

Solution. We have

$$f_y(x, y) = 2x - 2 \qquad f_y(x, y) = 2y - 6.$$

These partial derivatives are equal to 0 when $x = 1$ and y=3, so the only critical point is (1,3). By completing the square, we find that

$$f(x, y) = 4 + (x - 1)^2 + (y - 3)^2.$$

Since $(x - 1)^2, (y - 3)^2 \geq 0$, we have $f(x, y) \geq 4$ for all values of x, y. Therefore $f(1, 3) = 4$ is a local minimum, and in fact it is the absolute minimum of f. This can be confirmed geometrically from the graph of f which is the elliptic paraboloid with vertex shown in the figure geven below.

Example 54.

A function with no extreme values: Investigate the extreme values of $f(x, y) = xy$.

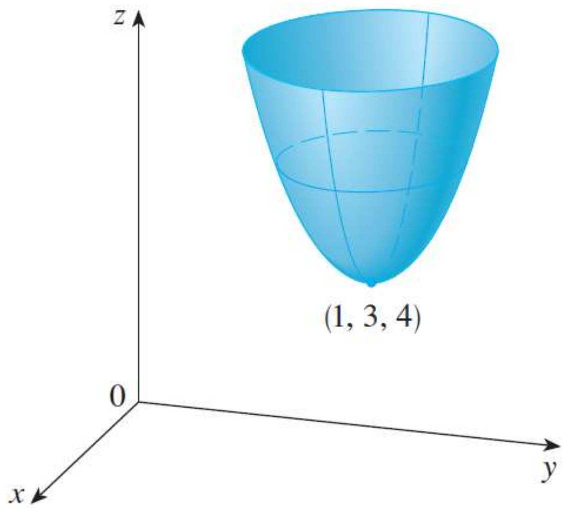


Figure 34: The paraboloid $f(x, y) = x^2 + y^2 - 2x - 6y + 14$

This example illustrates the fact that a function need not have a maximum or minimum value at a critical point. Figure 35 shows how this is possible. The graph of f is the hyperbolic paraboloid $z = xy$, which has a horizontal tangent plane ($z = 0$) at the origin. You can see that $f(0, 0) = 0$ is a maximum in the direction of the line $y = x$ but a minimum in the direction of the line $y = -x$. Near the origin the graph has the shape of a saddle and so the point $(0, 0)$is called a *saddle point* of f.

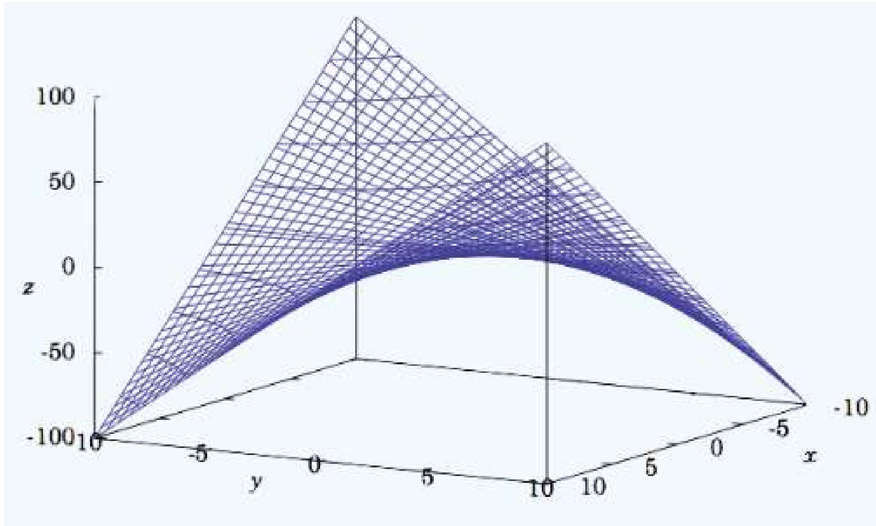


Figure 35: The hyperbolic paraboloid $f(x, y) = xy$

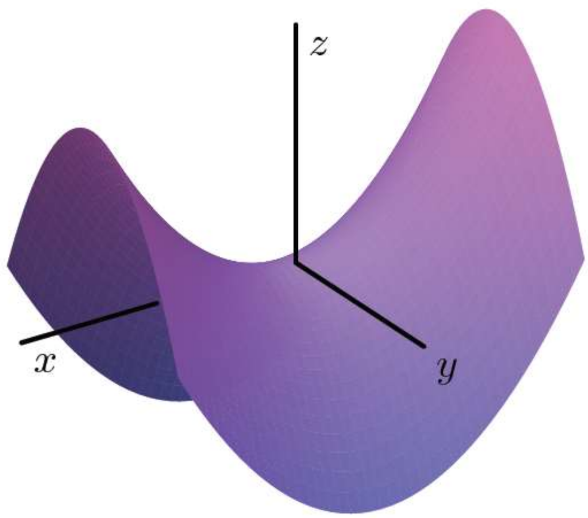


Figure 36: (0,0): saddle point

Saddle points

Given the function $z = f(x, y)$, the point $(a, b, f(a, b))$ is a **saddle point** if there are two distinct vertical planes through this point such that the intersection of the surface with one of the planes has a relative maximum at (a, b) and the intersection wtih the other has a relative minimum at (a, b).

In other words, the point $(a, b, f(a, b))$ is a saddle point if both

$f_x(a,b) = 0$ and $f_y(a,b) = 0$, but f does not have a local extremum at (a,b).

The Second Derivative Test.
Suppose the second partial derivatives of f are continuous on a disk with center (a,b) and suppose that $f_x(a,b) = 0$ and $f_y(a,b) = 0$. Let D be the quantity defined by

$$D = f_{xx}(a,b)f_{yy}(a,b) - f_{xy}(a,b)^2.$$

1. If $D > 0$ and $f_{xx}(a,b) < 0$, then f has a local maximum at (a,b).
2. If $D > 0$ and $f_{xx}(a,b) > 0$, then f has a local minimum at (a,b).
3. If $D < 0$, then f has a saddle point at (a,b).
4. If $D = 0$, then this test yields no information about what happens at (a,b).

The quantity D is called the **discriminant** of the function f at (a,b).

Remark

It's helpful to write D as a determinant:

$$D = \begin{vmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{vmatrix} = f_{xx}f_{yy} - (f_{xy})^2.$$

The matrix

$$\begin{pmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{pmatrix}$$

is called the **Hessian matrix** of f.

Example 55.

Find the local maximum and minimum values and saddle points of

$$f(x, y) = x^4 + y^4 - 4xy + 1.$$

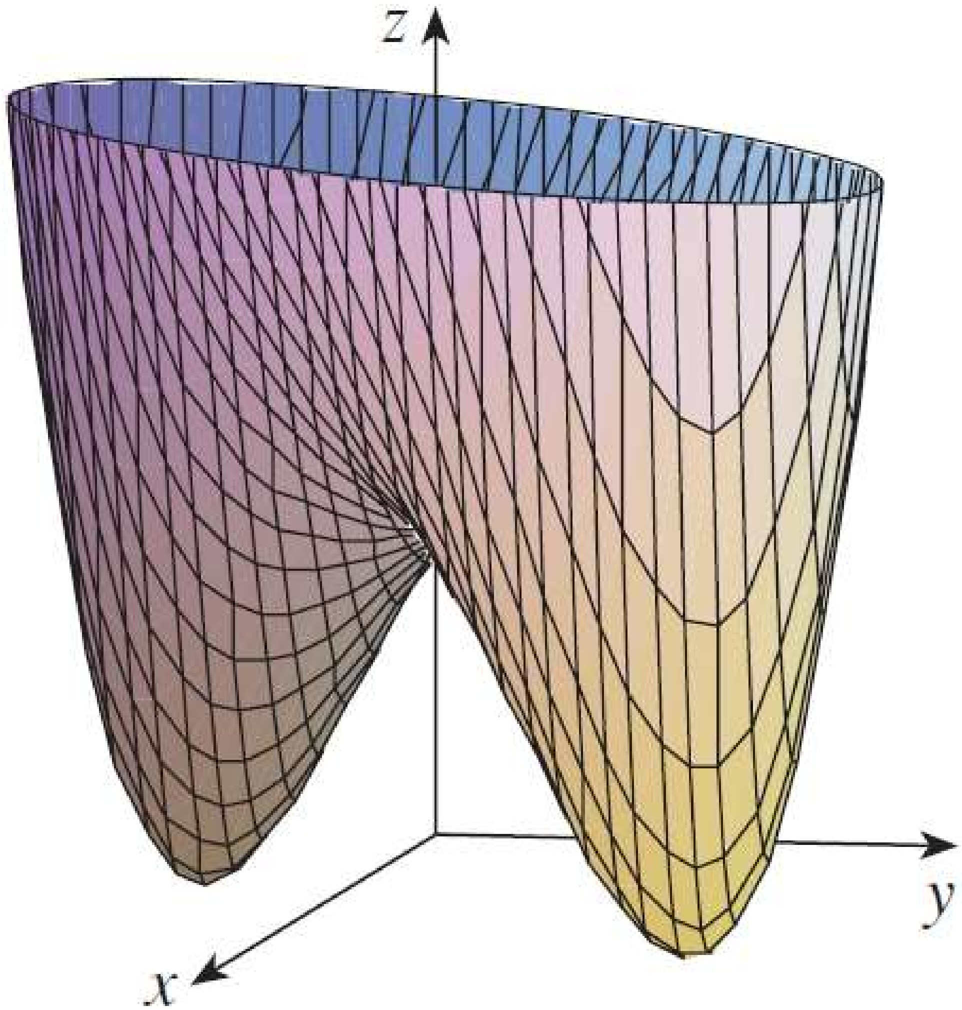


Figure 37: $f(x, y) = x^4 + y^4 - 4xy + 1$

Example 56.

Find all local maxima and minima of

$$f(x,y) = (x^2+y^2)e^{-(x^2+y^2)}.$$

Solution. First find the critical points, i.e. where $\nabla f = 0$. Since

$$\frac{\partial f}{\partial x} = 2x(1-(x^2+y^2))e^{-(x^2+y^2)}$$
$$\frac{\partial f}{\partial y} = 2y(1-(x^2+y^2))e^{-(x^2+y^2)},$$

then the critical points are $(0,0)$ and all points (x,y) on the unit circle $x^2+y^2=1$.

Now, the second-order partial derivatives are:

$$\frac{\partial^2 f}{\partial x^2} = 2[1-(x^2+y^2)-2x^2-2x^2(1-(x^2+y^2))]e^{-(x^2+y^2)}$$
$$\frac{\partial^2 f}{\partial y^2} = 2[1-(x^2+y^2)-2y^2-2y^2(1-(x^2+y^2))]e^{-(x^2+y^2)}$$
$$\frac{\partial^2 f}{\partial y\partial x} = -4xy[2-(x^2+y^2)]e^{-(x^2+y^2)}$$

At $(0,0)$, we have $D = 4 > 0$ and $\frac{\partial^2 f}{\partial x^2}(0,0) = 2 > 0$, so $(0,0)$ is a local minimum. However, for points (x,y) on the unit circle $x^2+y^2=1$, we have

$$D = (-4x^2e^{-1})(-4y^2e^{-1}) - (-4xye^{-1})^2 = 0$$

and so the test fails.

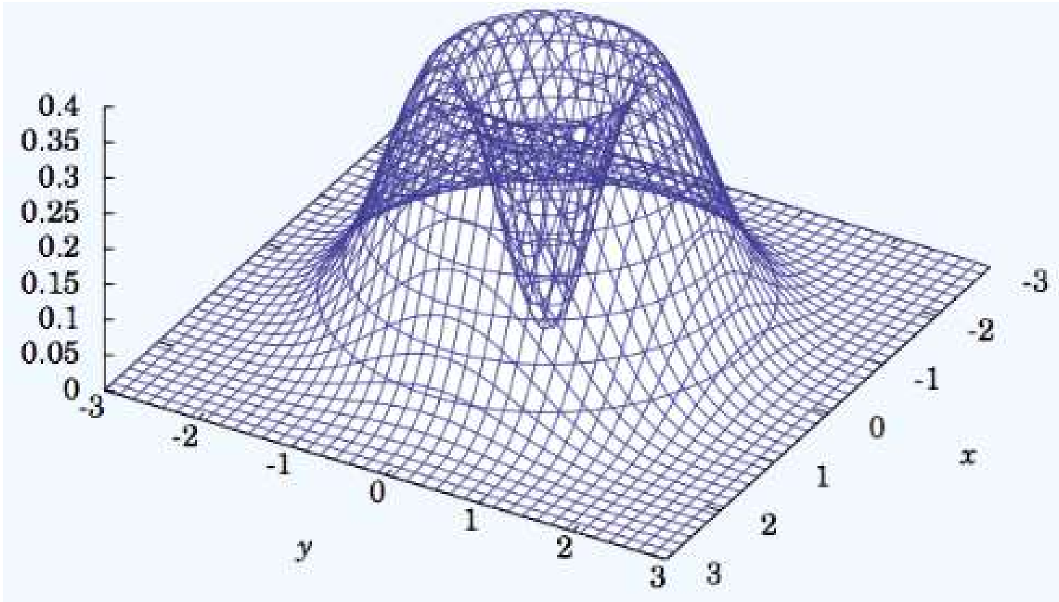


Figure 38: $f(x, y) = (x^2 + y^2)e^{-(x^2+y^2)}$

If we look at the graph of $f(x, y)$, as shown in the above figure, it looks like we might have a local maximum for (x, y) on the unit circle $x^2 + y^2 = 1$. If we switch to using polar coordinates (r, θ) instead of (x, y) in $\mathbb{R}^2$, where $r^2 = x^2 + y^2$, then we see that we can write $f(x, y)$ as a function $g(r)$ of the variable r alone:

$$g(r) = r^2 e^{-r^2}.$$

Then

$$g'(r) = r^2(1 - r^2)e^{-r^2},$$

so it has a critical point at $r = 1$, and we can check that

$$g''(1) = -4e^{-1} < 0,$$

so the Second Derivative Test from single-variable calculus says that $r = 1$ is a local maximum. But $r = 1$ corresponds to the unit circle $x^2 + y^2 = 1$. Thus, the points (x, y) on the unit circle $x^2 + y^2 = 1$ are local maximum points for f .