1.4.4 The Chain Rule

If y is a differentiable function of x *i.e.* $y = f(x)$ and x is a differentiable function of t *i.e.* $x = g(t)$, then y is a function of t

$$y = f(g(t)) \text{ and } \frac{dy}{dt} = \frac{dy}{dx}\frac{dx}{dt}$$

Theorem 1.4.5 *(Chain Rule 1) Let $x(t)$ and $y(t)$ be differentiable functions. Let $f(x, y)$ have continuous first order partial derivatives. Then*

$$\frac{df}{dt} = \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt}.$$

Proof 1.4.4

$$\frac{\Delta f}{\Delta t} = \frac{\partial f}{\partial x}\frac{\Delta x}{\Delta t} + \frac{\partial f}{\partial x}\frac{\Delta y}{\Delta t} + \epsilon_1\frac{\Delta x}{\Delta t} + \epsilon_2\frac{\Delta y}{\Delta t}$$

As $\Delta t \to 0$, we have $\Delta x \to 0, \Delta y \to 0, \epsilon_1 \to 0, \epsilon_2 \to 0$. Then the result follows.

Theorem 1.4.6 *(Chain Rule 2) Let $f(x, y)$ have continuous first order partial derivatives. Suppose $x = x(s, t)$ and $y = y(s, t)$ are functions such that x_s, x_t, y_s and y_t are also continuous. Then*

$$\frac{\partial f}{\partial s} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial s} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial s}, \quad \frac{\partial f}{\partial t} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial t},$$

The chain rule can be used for functions of more than two variables:
Given a function $f(x_1, x_2, ..., x_n)$ defined at points of $\mathbb{R}^n$, consider the values of f along a curve

$$x_1 = x_1(t), x_2 = x_2(t), .., x_n = x_n(t).$$

Here $t \in \mathbb{R}$ is a parameter along the curve (e.g. time or arc length).

Let $w = f(x_1, x_2, ..., x_n)$ function of t. If $f, x_1, x_2, ..., x_n$ are differentiable, then

$$\frac{dw}{dt} = \frac{\partial w}{\partial x_1}\frac{dx}{dt} + ... + \frac{\partial w}{\partial x_n}\frac{dx_n}{dt}$$

where each $\dfrac{\partial w}{\partial x_i}$ is evaluated at $(x_1, x_2, ..., x_n)$.

Example 1.4.11 *Let $z = e^x \sin y, x = st^2, y = s^2 t$. Then*

$$\begin{aligned}
\frac{\partial z}{\partial s} &= (e^x \sin y)t^2 + (e^x \cos y)2st = te^{st^2}(t\sin(s^2t) + 2s\cos(s^2t)). \\
\frac{\partial z}{\partial t} &= (e^x \sin y)2st + (e^x \cos y)s^2 = se^{st^2}(2t\sin(s^2t) + s\cos(s^2t)).
\end{aligned}$$

Example 1.4.12 *Given that $z = f(x, y)$ has continuous second order partial derivatives and that $x = r^2 + s^2, y = 2rs$, find z_{rr}. We have $x_r = 2r, y_r = 2s$. Then*

$$\begin{aligned}
z_r &= 2rz_x + 2sz_y. \\
z_{xr} &= z_{xx}x_r + z_{xy}y_r = 2rz_{xx} + 2sz_{xy}. \\
z_{yr} &= z_{yx}x_r + z_{yy}y_r = 2rz_{yx} + 2sz_{yy}. \\
z_{rr} &= \frac{\partial z_r}{\partial r} = \frac{\partial}{\partial r}(2rz_x + 2sz_y) = 2z_x + 2rz_{xr} + 2sz_{yr}. \\
&= 2z_x + 2r(2rz_{xx} + 2sz_{xy}) + 2s(2rz_{yx} + 2sz_{yy}) \\
&= 2z_x + 4r^2 z_{xx} + 8rsz_{xy} + 4s^2 z_{yy}.
\end{aligned}$$

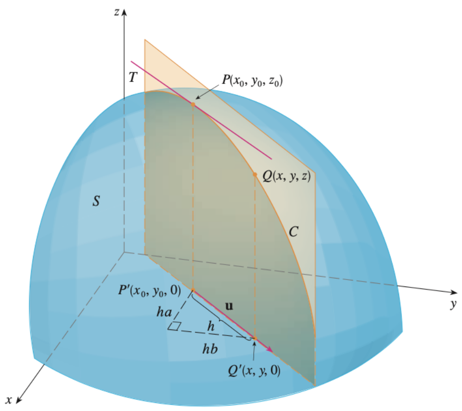
Example 1.4.13 *Find z_x and z_y if $x^3 + y^3 + z^3 + 6xyz = 1$.*

Example 1.4.14 *Given that $w = x^2 + y^2 + z^2$ and $z(x, y)$ satisfies $z^3xy + yz + y^3 = 1$, evaluate $\dfrac{\partial w}{\partial x}$ at $(2, 1, 1)$.*

1.4.5 Directional Derivative

Recall that if $f(x,y)$is a function,then $f_x(x_0,y_0)$is the rate of change in f with respect to change in x, at (x_0,y_0), that is, in the direction **i**. Similarly, $f_y(x_0,y_0)$ is the rate of change at (x_0,y_0) in the direction **j**. How do we find the rate of change of $f(x,y)$ at (x_0,y_0) in the direction of any unit vector **u**?

Suppose that we now wish to find the rate of change of z at (x_0,y_0) in the direction of an arbitrary unit vector $u=(a,b)$. To do this we consider the surface S with the equation $z=f(x,y)$ (the graph of f) and we let $z_0=f(x_0,y_0)$. Then the point $P(x_0,y_0,z_0)$ lies on S. The vertical plane that passes through P in the direction of u intersects S in a curve C. The slope of the tangent line T to C at the point P is the rate of change of z in the direction of u.



If $Q(x,y,z)$ is another point on C and P',Q' are the projections of P,Q onto the $xy-$ plane, then the vector $P'Q'$ is parallel to u and so

$$PQ = hu =< ha, hb >$$

for some scalar h. Therefore $x-x_0=ha, y-y_0=hb,$ so $x=x_0+ha, y=y_0+hb$, and

$$\frac{\Delta z}{h}=\frac{z-z_0}{h}=\frac{f(x_0+ha,y_0+hb)-f(x_0,y_0)}{h}.$$

If we take the limit as $h\to 0$, we obtain the rate of change of z (with respect to distance) in the direction of u, which is called the directional derivative of f in the direction of u.

Definition 1.4.2 *The directional derivative of f at x_0,y_0 in the direction of a unit vector $u=(a,b)$ is*

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

if this limit exists.

Theorem 1.4.7 *If f is a differentiable function of x and y, then f has a directional derivative in the direction of any unit vector $u=(a,b)$ and*

$$D_uf(x,y)=f_x(x,y)a+f_y(x,y)b$$

Proof 1.4.5 *If we define a function g of the single variable h by*

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

then, by the definition of a derivative, we have

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

On the other hand, we can write $g(h)=f(x,y)$, where $x=x_0+ha, y=y_0+hb$, so the Chain Rule gives,

$$g'(h)=\frac{\partial f}{\partial x}\frac{dx}{dh}+\frac{\partial f}{\partial y}\frac{dy}{dh}=f_x(x,y)a+f_y(x,y)b$$

If we now put $h = 0$*, then* $x = x_0, y = y_0$*, and*

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

□

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

$$\begin{aligned} D_u f(x_0, y_0) &= f_x(x_0, y_0)a + f_y(x_0, y_0)b \\ &= < f_x(x_0, y_0), f_y(x_0, y_0) > \cdot < a, b > \\ &= < f_x(x_0, y_0), f_y(x_0, y_0) > \cdot \mathbf{u} \end{aligned}$$

Definition 1.4.3 *If* f *is a function of two variables* x *and* y*, then the gradient of* f *is the vector function* ∇f *defined by*

$$\nabla f(x, y) = < f_x(x_0, y_0), f_y(x_0, y_0) > = \frac{\partial f}{\partial x}\boldsymbol{i} + \frac{\partial f}{\partial y}\boldsymbol{j}$$

Example 1.4.15 *If* $f(x, y) = \sin x + e^{xy}$*, then find* ∇f

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 $v = 2\mathbf{i} + 5\mathbf{j}$.

1.4.6 Maximizing the Directional Derivative

Suppose we have a function f 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 does f change fastest and what is the maximum rate of change? The answers are provided by the following theorem.

Theorem 1.4.8 *Suppose* f *is a differentiable function of two or three variables. The maximum value of the directional derivative* $D_u f(\boldsymbol{x})$ *is* $|\nabla f(\boldsymbol{x})|$ *and it occurs when* u *has the same direction as the gradient vector* $\nabla f(x)$.

Proof 1.4.6

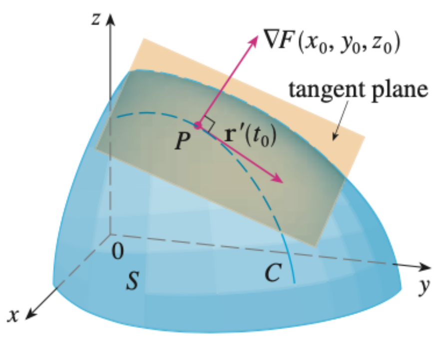
$$D_u f = \nabla f \cdot u = |\nabla f||u|| \cos(\theta) = |\nabla f| \cos(\theta)$$

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

Example 1.4.16 *If* $f(x, y) = xe^y$*, find the rate of change of* f *at the point* $P(2, 0)$ *in the direction from* P *to* $Q(1/2, 2)$*. In what direction does f have the maximum rate of change? What is this maximum rate of change?*

1.4.7 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.



Let the curve C is described by a continuous vector function $r(t) = (x(t), y(t), z(t))$. Let t_0 be the parameter value corresponding to P; that is, $r(t_0) = (x_0, y_0, z_0)$. 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, and z are differentiable functions of t and F is also differentiable, then we can use the Chain Rule to differentiate both sides

$$\frac{\partial F}{\partial x}\frac{dx}{dt} + \frac{\partial F}{\partial y}\frac{dy}{dt} + \frac{\partial F}{\partial z}\frac{dz}{dt} = 0$$

But, since $\nabla F = (F_x, F_y, F_z)$ and $r'(t) = x'(t), y'(t), z'(t)$, therefore in terms of a dot product

$$\nabla F.r'(t) = 0$$

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

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

Then the gradient vector at $P, \nabla F(x_0, y_0, z_0)$, is perpendicular to the tangent vector $r'(t_0)$ to any curve C on S that passes through P. If $F(x_0, y_0, z_0) \neq 0$, it is therefore natural to define the tangent plane to the level surface $F(x, y, z) = k$ at $P(x_0, y_0, z_0)$ as the plane that passes through P and has normal vector $\nabla \mathrm{F}(\mathrm{x}_0, y_0, z_0)$.
Now we can write the equation of this tangent plane as

$$F_x(x_0, y_0, z_0)(x - x_0) + F_y(x_0, y_0, z_0)(y - y_0) + F_z(x_0, y_0, z_0)(z - z_0) = 0$$

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

$$\frac{x - x_0}{f_x(x_0, y_0, z_0)} = \frac{y - y_0}{f_x(x_0, y_0, z_0)} = \frac{z - z_0}{f_x(x_0, y_0, z_0)}$$

Example 1.4.17 *Find the equations of the tangent plane and normal line at the point* $(-2, 1, -3)$ *to the ellipsoid*

$$\frac{x^2}{4} + y^2 + \frac{z^2}{9} = 3$$

1.4.8 The Jacobian matrix

The Jacobian operator is a generalization of the derivative operator to the vector-valued functions. As we have seen earlier, a vector-valued function is a mapping from

$$f : \mathbb{R}^n \to \mathbb{R}^m,$$

hence, now instead of having a scalar value of the function f, we will have a mapping $[x_1, x_2, .., x_n] \to [f_1, f_2, .., f_m]$. Thus, we now need the rate of change of each component of f with respect to each component of the input variable x, this is exactly what is captured by a matrix called Jacobian matrix J.

$$J_p f = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \frac{\partial f_m}{\partial x_2} & \cdots & \frac{\partial f_m}{\partial x_n} \end{pmatrix}$$

The first order approximation of near a point is obtained using the Jacobian matirx:

$$A(x) = f(x_0) + J_p^T(f(x_0))(x - x_0)$$

1.4.9 The Hessian matrix

For a function $f : \mathbb{R}^n \to \mathbb{R}$ that is twice differentiable, the Hessian matrix is the matrix of second derivatives:

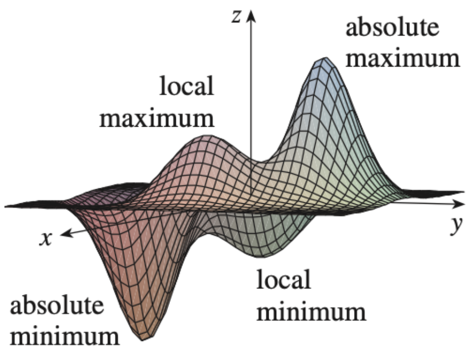
$$H(f) = D^2 f = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \cdots & \frac{\partial f}{\partial x_n \partial x_1} \\ \vdots & \vdots & \vdots \\ \frac{\partial^2 f}{\partial x_1 \partial x_n} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{pmatrix}$$

Remarks: If is twice continuously differentiable, the Hessian matrix is always a symmetric matrix. This is because partial derivatives commute:

$$\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial^2 f}{\partial x_j \partial x_i}$$

1.5 Maximum and Minimum Values

In this section we see how to use partial derivatives to locate maxima and minima of functions of two variables.



There are two points (a, b) where f has a local maximum, that is, where $f(a, b)$ is larger than nearby values of $f(x, y)$. The larger of these two values is the absolute maximum. Likewise, f has two local minima, where $f(a, b)$ is smaller than nearby values. The smaller of these two values is the absolute minimum.

Definition 1.5.1 *A function of two variables has a local maximum at (a, b) if $f(x, y) \leq f(a, b)$ when (x, y) is near (a, b). [This means that $f(x, y) \leq f(a, b)$ for all points (x, y) in some disk with center (a, b).] The number $f(a, b)$ is called a local maximum value. If $f(x, y) \geq f(a, b)$ when (x, y) is near (a, b), then f has a local minimum at (a, b) and $f(a, b)$ is a local minimum value.*

If the inequalities in the Definition hold for all points (x, y) in the domain of f, then f has an absolute maximum (or absolute minimum) at (a, b).

Theorem 1.5.1 *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 1.5.1 *Let f has a local maximum (or minimum) at (a, b). Let $g(x) = f(x, b)$.then g has a local maximum (or minimum) at a, so $g'(a) = 0$. But $g'(a) = f_x(a, b)$ and so $f_x(a, b) = 0$. Similarly, by letting a function $G(y) = f(a, y)$, we obtain $f_y(a, b) = 0$.*

A point (a, b) is called a critical point (or stationary point) of f if $f_x(a, b) = 0$ and $f_y(a, b) = 0$,$(\nabla f = 0)$ or if one of these partial derivatives does not exist. Theorem (1.5.1) says that if f has a local maximum or minimum at (a, b), then (a, b) is a critical point of f. However, as in single-variable calculus, not all critical points give rise to maxima or minima. At a critical point, a function could have a local maximum or a local minimum or neither.

Example 1.5.1 *Find the critical point(s) of the following functions*

- $g(x, y) = x^2 + 6xy + 4y^2 + 2x4y$.

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

We need to be able to determine whether or not a function has an extreme value at a critical point. The following test, which is proved at the end of this section, is analogous to the Second Derivative Test for functions of one variable.

Definition 1.5.2 *Second Derivatives 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 $f_y(a, b) = 0$ [that is, (a, b) is a critical point of f]. Let*

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

a If $D > 0$ and $f_{xx}(a, b) > 0$, then $f(a, b)$ is a local minimum.

b If $D > 0$ and $f_{xx}(a, b) < 0$, then $f(a, b)$ is a local maximum.

c If $D < 0$ then $f(a, b)$ is not a local maximum or minimum.

Note:

- In case (c) the point (a, b) is called a saddle point of f and the graph of f crosses its tangent plane at (a, b).
- If $D = 0$, the test gives no information:f could have a local maximum or local minimum at (a, b), or (a, b) could be a saddle point of f.
- To remember the formula for D, it's helpful to write it as a determinant:

$$D = \begin{bmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{bmatrix} = f_{xx}f_{yy} - f_{xy}^2$$

Example 1.5.2 *Find the local maximum and minimum values and saddle points of $f(x, y) = x^4 + y^4 - 4xy + 1$.*

Example 1.5.3 *Find the shortest distance from the point $(1, 0, -2)$ to the plane $x + 2y + z = 4$.*

Example 1.5.4 *A rectangular box without a lid is to be made from $12m^2$ of cardboard. Find the maximum volume of such a box.*

For a function f of one variable, the Extreme Value Theorem says that if f is continuous on a closed interval $[a, b]$, then f has an absolute minimum value and an absolute maximum value. According to the Closed Interval Method we find these by evaluating f not only at the critical numbers but also at the endpoints a and b.

There is a similar situation for functions of two variables. Just as a closed interval contains its endpoints, a closed set in $\mathbb{R}^2$ is one that contains all its boundary points and a bounded set in $\mathbb{R}^2$ is one that is contained within some disk.

Theorem 1.5.2 *(Extreme Value Theorem for Functions of Two Variables)*
If f is continuous on a closed, bounded set D in $\mathbb{R}^2$, then f attains an absolute maximum value $f(x_1, y_1)$ and an absolute minimum value $f(x_2, y_2)$ at some points (x_1, y_1) and (x_2, y_2) in D.

To find the extreme values guaranteed by Theorem (1.5.2), we note that, if f has an extreme value at (x_1, y_1), then (x_1, y_1) is either a critical point of f or a boundary point of D. Thus we have the following extension of the Closed Interval Method.

To find the absolute maximum and minimum values of a continuous function f on a closed, bounded set D:

1. Find the values of f at the critical points of f in D.

2. Find the extreme values of f on the boundary of D.

3. The largest of the values from steps 1 and 2 is the absolute maximum value; the smallest of these values is the absolute minimum value.

Example 1.5.5 *Find the absolute maximum and minimum values of the function* $f(x, y) = x^2 - 2xy + 2y$ *on the rectangle* $D = (x, y)|0 \leq x \leq 3, 0 \leq y \leq 2$.