Chapter 5

Directional Derivatives and the Gradient

A function $f(x)$ of a single independent variable can be characterized at any point x_0 by the slope, in terms of the first derivative $f'(x_0)$, the curvature, in terms of the second derivative $f''(x_0)$, and so on. These quantities enable a function to be visualized in terms of its "steepness" and the "sharpness" of its bends. A function f of two or more independent variables can be similarly characterized in terms of its partial derivatives. But partial derivatives such as f_x and f_y are unnecessarily restrictive in that they represent the rates of change of a function only along the x- and y-axes (Fig. 1.2). In fact, derivatives can be taken along *any* direction in the space of the independent variables. The computation of such directional derivatives is the basis for introducing a new type of derivative for functions of two or more variables, called the **gradient**. This quantity is a *vector* that provides similar information to contour plots, such as isobars on a weather map or a relief map of a mountainous range, but in a much more succinct form. In some respects, the gradient is the natural generalization to higher dimensions of the ordinary derivative in that it determines the tangent plane to the surface of a function (for the case of two independent variables), just as the ordinary derivative is the slope of the tangent line to a function. The gradient has a wide variety of applications, ranging from the calculation the flow of physical quantities such as heat, the solution of certain types of linear equations (the "conjugate gradient" method, and in image processing, where the gradient is used to extract information about the edges in an image, which is especially important in biological and medical imaging.

5.1 Directional Derivatives

We will confine our discussion initially to functions of two independent variables because the various quantities associated with derivatives are easier to visualize than for the case of three independent variables and the results obtained are straightforward to generalize. The task at hand is the calculation of the derivative of a function along a particular direction. This proceeds by specifying the point $\boldsymbol{r}_0$ and the direction along a *unit* vector $\boldsymbol{u}$:

$$\boldsymbol{r}_0 = x_0\,\boldsymbol{i} + y_0\,\boldsymbol{j}\,, \tag{5.1}$$

$$\boldsymbol{u} = a\,\boldsymbol{i} + b\,\boldsymbol{j}\,, \tag{5.2}$$

where the stipulation that $\boldsymbol{u}$ must be a unit vector, $\boldsymbol{u}\cdot\boldsymbol{u} = 1$, means that $a^2 + b^2 = 1$. We now form the vector

$$\begin{aligned}\boldsymbol{r} &= \boldsymbol{r}_0 + \boldsymbol{u}\,s\\ &= (x_0 + as)\,\boldsymbol{i} + (y_0 + bs)\,\boldsymbol{j}\,,\end{aligned} \tag{5.3}$$

where $0 \le s \le 1$ is a parameter, as shown in Fig. 5.1. In terms of Cartesian components, $\boldsymbol{r} = (x, y)$,

$$x = x_0 + as, \tag{5.4}$$

$$y = y_0 + bs\,. \tag{5.5}$$

Thus, for $s = 0$, $\boldsymbol{r} = \boldsymbol{r}_0$, and for $s = 1$, $\boldsymbol{r} = \boldsymbol{r}_0 + \boldsymbol{u}$ (Fig. 5.1).

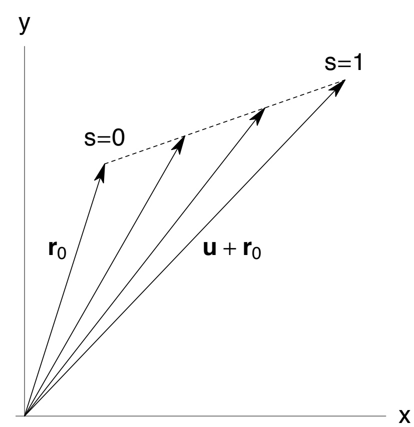

Figure 5.1: The vectors $\boldsymbol{r}_0$ ($s = 0$), $\boldsymbol{r}_0+\boldsymbol{u}$ ($s = 1$), and the those for which $0 \le s \le 1$.

Now consider a scalar function $f(x, y)$. Along the direction defined by $\boldsymbol{r}$,

$$f(x, y) = f[x(s), y(s)] = f(x_0 + as, y_0 + bs)\,. \tag{5.6}$$

Thus, the derivative of f with respect to s is, according to the chain rule,

$$\begin{aligned} \frac{df}{ds} &= \frac{\partial f}{\partial x}\frac{dx}{ds} + \frac{\partial f}{\partial y}\frac{dy}{ds} \\ &= \frac{\partial f}{\partial x} a + \frac{\partial f}{\partial y} b\,. \end{aligned} \tag{5.7}$$

We can rewrite this expression in vector notation as

$$\frac{df}{ds} = \underbrace{\left(\frac{\partial f}{\partial x}\,\boldsymbol{i} + \frac{\partial f}{\partial y}\,\boldsymbol{j}\right)}_{\boldsymbol{\nabla} f} \cdot \underbrace{(a\,\boldsymbol{i} + b\,\boldsymbol{j})}_{\boldsymbol{u}} = \frac{\partial f}{\partial x} a + \frac{\partial f}{\partial y} b\,, \tag{5.8}$$

where we have defined the **gradient** of f, denoted by $\boldsymbol{\nabla} f$, as

$$\boxed{\boldsymbol{\nabla} f = \frac{\partial f}{\partial x}\,\boldsymbol{i} + \frac{\partial f}{\partial y}\,\boldsymbol{j}\,.} \tag{5.9}$$

The symbol "$\boldsymbol{\nabla}$" is referred to as "nabla", "grad", or "del". The gradient is a *vector field* calculated from the *scalar* function f. The quantity df/ds is the **directional derivative** of f in the direction of the unit vector $\boldsymbol{u}$, and is written as

$$\boxed{\boldsymbol{\nabla}_{\boldsymbol{u}} f \equiv \boldsymbol{\nabla} f \cdot \boldsymbol{u}\,,} \tag{5.10}$$

which is a *scalar* quantity because it is obtained as the "dot" product of two vectors. This derivative can be written in a form analogous to the ordinary derivative in Sec. 1.1. Using the notation $f(\boldsymbol{r}) \equiv f(x, y)$, we have that

$$\boxed{\left.\boldsymbol{\nabla}_{\boldsymbol{u}} f\right|_{\boldsymbol{r}_0} = \lim_{s \to 0} \left[\frac{f(\boldsymbol{r}_0 + s\boldsymbol{u}) - f(\boldsymbol{r}_0)}{s}\right]\,.} \tag{5.11}$$

Example. Consider the special cases where $\boldsymbol{u} = \boldsymbol{i}$ and $\boldsymbol{u} = \boldsymbol{j}$. From Eq. (5.10), we obtain

$$\boldsymbol{\nabla}_{\boldsymbol{i}} f = \left(\frac{\partial f}{\partial x}\, \boldsymbol{i} + \frac{\partial f}{\partial y}\, \boldsymbol{j} \right) \cdot \boldsymbol{i} = \frac{\partial f}{\partial x}\,, \tag{5.12}$$

$$\boldsymbol{\nabla}_{\boldsymbol{j}} f = \left(\frac{\partial f}{\partial x}\, \boldsymbol{i} + \frac{\partial f}{\partial y}\, \boldsymbol{j} \right) \cdot \boldsymbol{j} = \frac{\partial f}{\partial y}\,. \tag{5.13}$$

Thus, the directional derivatives along the directions of the coordinate axes reduces to the familiar partial derivatives. Along any other directions, the directional derivative is the weighted average of these two derivatives. This explains why $\boldsymbol{u}$ must be a unit vector: the role of $\boldsymbol{u}$ is only to provide the direction in the directional derivative and not to affect its magnitude. ▮

5.2 Meaning of the Gradient

5.2.1 Magnitude

To provide an interpretation of the gradient, we first observe that, since $|\boldsymbol{u}|$ is a unit vector, we can write the directional derivative as

$$\boldsymbol{\nabla}_{\boldsymbol{u}} f = \boldsymbol{\nabla} f \cdot \boldsymbol{u} = |\boldsymbol{\nabla} f|\, |\boldsymbol{u}| \cos\theta = |\boldsymbol{\nabla} f| \cos\theta\,, \tag{5.14}$$

where θ is the angle between $\boldsymbol{\nabla} f$ and $\boldsymbol{u}$. The maximum value of the right-hand side is obtained for $\theta = 0$, when $\boldsymbol{\nabla} f$ and $\boldsymbol{u}$ are parallel. We conclude that

$$\boldsymbol{\nabla}_{\boldsymbol{u}} f \le |\boldsymbol{\nabla} f|\,, \tag{5.15}$$

i.e. *the absolute value of the gradient of a function is the maximum rate of change of that function.*

5.2.2 Direction

We can proceed further and obtain an interpretation of the *direction* of the gradient. Consider the surface $f(x, y) = \text{constant}$ and a curve $[x(s), y(s)]$

that lies on this surface, so $f[x(s), y(s)] =$ constant for all s. We will take the directional derivative of f at $s = s_0$, where $\boldsymbol{r} = \boldsymbol{r}_0$:

$$\boldsymbol{\nabla}_{\boldsymbol{u}} f\Big|_{\boldsymbol{r}_0} = \boldsymbol{\nabla} f\Big|_{\boldsymbol{r}_0} \cdot \boldsymbol{u}\,, \tag{5.16}$$

where $\boldsymbol{u}$ is a unit vector that is tangent to the surface $f(x, y, z) =$ constant. Then, using the definition of the derivative in Eq. (5.11), we have that

$$\boldsymbol{\nabla}_{\boldsymbol{u}} f\Big|_{\boldsymbol{r}_0} = \lim_{\Delta s \to 0} \left[\frac{f[x(s_0 + \Delta s), y(s_0 + \Delta s)] - f(x_0, y_0)}{\Delta s} \right]. \tag{5.17}$$

Since

$$f[x(s_0 + \Delta s), y(s_0 + \Delta s)] = f(x_0, y_0) = \text{constant}\,, \tag{5.18}$$

we conclude that

$$\boldsymbol{\nabla}_{\boldsymbol{u}} f\Big|_{\boldsymbol{r}_0} = \boldsymbol{\nabla} f\Big|_{\boldsymbol{r}_0} \cdot \boldsymbol{u} = 0\,, \tag{5.19}$$

i.e. *the gradient of f is normal to surfaces of constant f.*

Example. Consider the function

$$f(x, y) = 1 - x^2 - y^2 \tag{5.20}$$

for the ranges of variables $-1 \le x \le 1$ and $-1 \le y \le 1$. This surface, which is a paraboloid, is shown in Fig. 5.2. The gradient of f is defined as

$$\boldsymbol{\nabla} f = \frac{\partial f}{\partial x}\,\boldsymbol{i} + \frac{\partial f}{\partial y}\,\boldsymbol{j}\,. \tag{5.21}$$

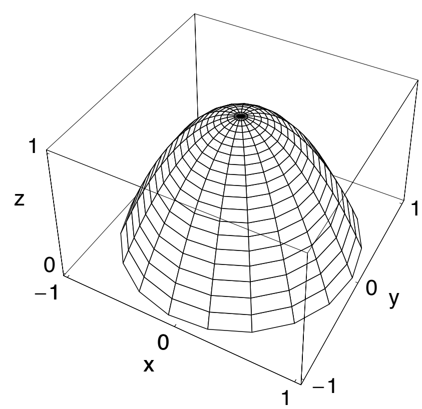

Figure 5.2: The function $z = 1 - x^2 - y^2$ for $-1 \le x \le 1$ and $-1 \le y \le 1$.

The partial derivatives of f are

$$\frac{\partial f}{\partial x} = -2x\,, \quad \frac{\partial f}{\partial y} = -2y\,, \tag{5.22}$$

so the gradient is

$$\boldsymbol{\nabla} f = -2x\,\boldsymbol{i} - 2y\,\boldsymbol{j}\,. \tag{5.23}$$

According to the discussion in this section, the gradient represents the maximum rate of change of f and point in the direction normal to surfaces of constant f. In this case, the "surfaces" of constant f are curves in the x-y plane. For $f = z_0$, these curves are given by

$$x^2 + y^2 = 1 - z_0\,, \tag{5.24}$$

which are circles of radius $\sqrt{1-z_0}$ centered at the origin. As z_0 increases from 0 to 1, the radii of the circles decreases from 1 to 0, i.e. the height of f above the x-y plane increases toward the origin, which is the maximum of f, as shown in Fig. 5.2. These contours are shown in Fig. 5.3, together with the gradient calculated in Eq. (5.22). The gradient is seen to be a radial vector field that points toward the origin, i.e. along the direction of the maximum rate of change of f. Also evident is that the gradient vectors are normal to the circles of constant f.

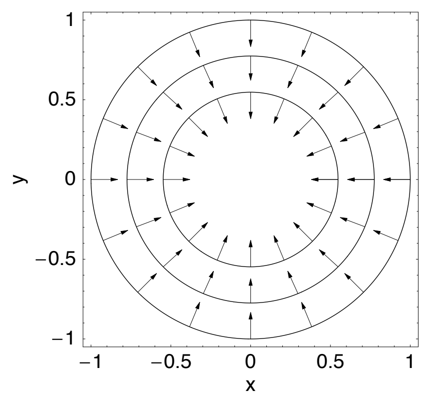

Figure 5.3: The contours of constant f in Eq. (5.23) and the gradient field in Eq. (5.23).

Having computed the gradient of f, we can now determine its directional derivative. For any unit vector $\boldsymbol{u} = a\,\boldsymbol{i} + b\,\boldsymbol{j}$, where $a^2 + b^2 = 1$, we have from Eq. (5.10),

$$\boldsymbol{\nabla_u} f = (-2x\,\boldsymbol{i} - 2y\,\boldsymbol{j})\cdot(a\,\boldsymbol{i} + b\,\boldsymbol{j}) = -2ax - 2yb\,. \tag{5.25}$$

The calculation of this derivative requires the specification of a point (x, y) and a direction (a, b). *All* unit vectors can be written in terms of polar coordinates as

$$\boldsymbol{u} = \cos\phi\,\boldsymbol{i} + \sin\phi\,\boldsymbol{j}\,, \tag{5.26}$$

where $0 \le \phi < 2\pi$, in which case the directional derivative becomes

$$\boldsymbol{\nabla_u} f = -2x\cos\phi - 2y\sin\phi\,. \tag{5.27}$$

For example, at the point $(\frac{1}{2}, \frac{1}{2})$,

$$\boldsymbol{\nabla_u} f\Big|_{(\frac{1}{2},\frac{1}{2})} = -\cos\phi - \sin\phi\,. \tag{5.28}$$

The maximum of the directional derivative is obtained for $\frac{5}{4}\pi$, where the sine and cosine function both have the values $-\frac{1}{2}\sqrt{2}$:

$$\boldsymbol{\nabla}_{\boldsymbol{u}} f\Big|_{(\frac{1}{2},\frac{1}{2})} = -\cos\left(\tfrac{5}{4}\pi\right) - \sin\left(\tfrac{5}{4}\pi\right) = \sqrt{2}\,, \tag{5.29}$$

which is equal to the magnitude of the gradient. The minimum value is obtained for $\frac{1}{4}\pi$, where the sine and cosine function both have the values $\frac{1}{2}\sqrt{2}$:

$$\boldsymbol{\nabla}_{\boldsymbol{u}} f\Big|_{(\frac{1}{2},\frac{1}{2})} = -\cos\left(\tfrac{5}{4}\pi\right) - \sin\left(\tfrac{5}{4}\pi\right) = -\sqrt{2}\,, \tag{5.30}$$

which is also equal to the magnitude of the gradient. The directional derivative vanishes for $\phi = \frac{3}{4}\pi$ and $\phi = \frac{7}{4}\pi$, which is the tangential direction to the contours of constant f. ∎

5.3 The Gradient in Three Dimensions

All of the discussion in Secs. 5.1 and 5.2 can be generalized to functions of three independent variables with only minor modifications necessitated by the added dimension. We only quote the main results here and leave the derivations as an exercise. The gradient of a function of $f(x,y,z)$ of three independent variables is

$$\boxed{\boldsymbol{\nabla} f = \frac{\partial f}{\partial x}\,\boldsymbol{i} + \frac{\partial f}{\partial y}\,\boldsymbol{j} + \frac{\partial f}{\partial z}\,\boldsymbol{k}\,.} \tag{5.31}$$

The directional derivative at a point (x_0, y_0, z_0) in the direction of the unit vector

$$\boldsymbol{u} = a\,\boldsymbol{i} + b\,\boldsymbol{j} + c\,\boldsymbol{k}\,, \tag{5.32}$$

where $a^2 + b^2 + c^2 = 1$, is

$$\boxed{\boldsymbol{\nabla}_{\boldsymbol{u}} f = \boldsymbol{\nabla} f \cdot \boldsymbol{u}\,.} \tag{5.33}$$

The properties of the gradient are the same as for functions of two independent variables, namely, that the magnitude of the gradient of f at a point is the maximum rate of change of f at that point, and the gradient points in the direction of the maximum rate of change of f, normal to surfaces of constant f.

Example. Consider the function

$$f(x, y, z) = x^2 + y^2 + z^2 \,. \tag{5.34}$$

The surfaces of constant f are concentric spheres, as shown in Fig. 5.4. The partial derivatives of f are

$$\frac{\partial f}{\partial x} = 2x\,, \quad \frac{\partial f}{\partial y} = 2y\,, \quad \frac{\partial f}{\partial z} = 2z\,, \tag{5.35}$$

so the gradient of f is

$$\boldsymbol{\nabla} f = 2x\,\boldsymbol{i} + 2y\,\boldsymbol{j} + 2z\,\boldsymbol{k}\,. \tag{5.36}$$

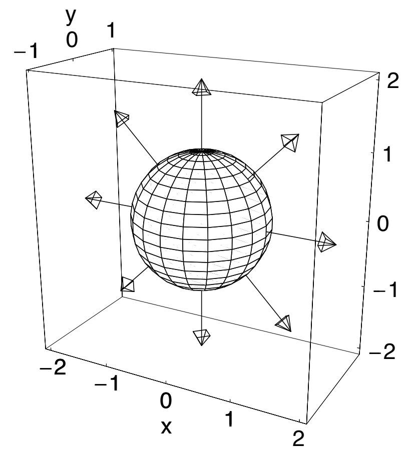

Figure 5.4: The surface $x^2 + y^2 + z^2 =$ constant shown together with the gradient at several points.

Several of these vectors are plotted in Fig. 5.4. These vectors are normal to the spherical surface and point away from the origin, because points on spherical surfaces with increasing radii are further away from the origin.

To calculate the directional derivative, we observe that any unit vector $\boldsymbol{u}$ in three dimensions can be expressed as

$$\boldsymbol{u} = \sin\theta\cos\phi\,\boldsymbol{i} + \sin\theta\sin\phi\,\boldsymbol{j} + \cos\theta\,\boldsymbol{k}\,, \tag{5.37}$$

where $0 \le \phi < 2\pi$ and $0 \le \theta \le \pi$. Geometrically, theses vectors emanate from the origin with endpoints on the surface of a unit sphere. Thus,

$$\boldsymbol{\nabla}_{\boldsymbol{u}} f = 2x\sin\theta\cos\phi + 2y\sin\theta\sin\phi + 2z\cos\theta\,. \tag{5.38}$$

For example, at $(1, 0, 0)$,

$$\boldsymbol{\nabla}_{\boldsymbol{u}} f\Big|_{(1,0,0)} = 2\sin\theta\cos\phi\,. \tag{5.39}$$

This derivative has a maximum value for $\phi = 0$ and $\theta = \frac{1}{2}\pi$, i.e. for $\boldsymbol{u}$ along the x-axis (as shown in Fig. 5.4), and vanishes if $\phi = \frac{1}{2}\pi$ or $\phi = \frac{3}{2}\pi$ and $\theta = 0$ or $\theta = \pi$. These four unit vectors point along the positive and negative y- and z-axes, i.e. normal to the gradient. ∎

5.4 Physical Applications of the Gradient*

There are several physical situations that are formulated in terms of the gradient. Perhaps the most familiar application of the gradient is in classical mechanics. Given an energy-conserving potential $V(\boldsymbol{r})$ of a particle as a function of position $\boldsymbol{r}$, the force $\boldsymbol{F}$ acting on the particle is the negative gradient of the potential:

$$\boxed{\boldsymbol{F} = -\boldsymbol{\nabla} V\,.} \tag{5.40}$$

The minus sign indicates that the forces acting on the particle point in the direction of decreasing potential energy. The motion of the particle is obtained by solving Newton's second law of motion:

$$\boxed{m\frac{d^2\boldsymbol{r}}{dt^2} = -\boldsymbol{\nabla} V\,,} \tag{5.41}$$

where m is the mass of the particle. The existence of a potential associated with a force was discussed in Sec. 4.3.

The electrostatic force $\boldsymbol{E}$ is conservative, so the work done on a particle depends only on the initial and final position of the particle, and not on the path followed. as discussed in Sec. 4.3, with each conservative force, a potential energy can be associated. For the electrostatic force, the associated potential is calculated from

$$\boxed{\boldsymbol{E} = -\boldsymbol{\nabla}\phi\,.} \tag{5.42}$$

In most metals and semiconductors, the relationship between the electrical current density $\boldsymbol{j}$ and the applied electric field $\boldsymbol{E}$ is given by **Ohm's law**:

$$\boldsymbol{j} = \sigma \boldsymbol{E}\,, \tag{5.43}$$

where σ is the electrical conductivity. By using Eq. (5.42), we can write this relation as

$$\boxed{\boldsymbol{j} = -\sigma \boldsymbol{\nabla} \phi\,.} \tag{5.44}$$

As the following discussion shows, there are several phenomena that are described by equations of this form.

The relationship between the heat flow $\boldsymbol{q}$ in the presence of variations of temperature T is expressed in terms of **Fourier's law**:

$$\boxed{\boldsymbol{q} = -C \boldsymbol{\nabla} T\,,} \tag{5.45}$$

where C is the coefficient of thermal conductivity. The coefficient C may vary with the temperature, and certainly varies from one substance to another, but it is always a *positive* constant. This of course makes intuitive sense, at least if the molecular concept of temperature is invoked; the heat (kinetic energy at the microscopic scale) tends to flow from regions of high concentration of internal energy to regions of low internal energy, which is consistent with the statement above that the heat flow is directed in the direction of the gradient of the temperature, which defines its maximum rate of change.

Similar concepts apply to particle diffusion. The current of particles $\boldsymbol{j}$ in the presence of a varying concentration $\boldsymbol{c}$ of the particles is given by **Fick's law**:

$$\boxed{\boldsymbol{j} = -D \boldsymbol{\nabla} c\,,} \tag{5.46}$$

where D is the diffusion coefficient. The negative sign indicates that transport of material is from high to low concentrations, so that any variations in the concentrations tend to be smoothed out.

5.5 Summary

This chapter has introduced a derivative operation on scalar functions with several independent variables called the gradient, and denoted by $\boldsymbol{\nabla}$:

$$\boldsymbol{\nabla} f = \frac{\partial f}{\partial x}\,\boldsymbol{i} + \frac{\partial f}{\partial y}\,\boldsymbol{j} + \frac{\partial f}{\partial z}\,\boldsymbol{k}\,. \tag{5.47}$$

The gradient is a *vector* quantity calculated from a *scalar* function. The gradient is an intrinsic property of the function has the magnitude of the maximum rate of change of f and points in the direction normal to surfaces of constant f. The gradient is used in the following applications:

1. Identify the magnitude and direction of the maximum rate of change of a function at a given point.
2. Calculate the directional derivative at a point in a direction specified by a unit vector.
3. Find the normal to a surface as a specified point.
4. Determine the tangent plane to a surface at a given point.

Examples of these applications are provided in Classwork 6 and Problem Set 6.