Scalar Valued Functions of Several Variables; the Gradient Vector

Scalar Valued Functions Let us consider a scalar (i.e., numerical, rather than vector) valued function of n variables:

$$y = \phi(X) = \phi(x_1, x_2, ..., x_n), \ X \in \mathcal{D},$$

where the domain $\mathcal{D}$ is a region in $\mathbf{R}^n$. By this we mean that $\mathcal{D}$ is an *open, connected set.* The term open means that each point $X_0 \in \mathcal{D}$ is the center of a ball $\mathcal{N}(X_0, r) = \{X \,|\|X - X_0\| < r\}$, for some $r > 0$, which still lies in $\mathcal{D}$. The term connected, as we will use it here, means that any two points in $\mathcal{D}$ can be connected by a curve lying in $\mathcal{D}$.

The *partial derivatives* of the function $\phi(X)$ at a point $X_0 \in \mathcal{D}$ are the numbers, which we denote by $\frac{\partial \phi}{\partial x_k}(X_0), \ k = 1, 2, ..., n$, obtained by differentiating the function $\phi(X)$ with respect to the variable x_k at the point X_0, treating the other variables, $x_1, ..., x_{k-1}, x_{k+1}, ..., x_n$, as constant quantities. Taking the components of X_0 to be $x_{0,k}, \ k = 1, 2, ..., n$, these partial derivatives exist just in case the corresponding limits of difference quotients

$$\lim_{x_k \to x_{0,k}} \frac{\phi(x_{0,1}, ..., x_{0,k-1}, x_k, x_{0,k+1}, ..., x_{0,n}) - \phi(x_{0,1}, ..., x_{0,k-1}, x_{0,k}, x_{0,k+1}, ..., x_{0,n})}{x_k - x_{0,k}}$$

exist; the limit values are the partial derivatives we have indicated.

As we allow the point X_0 to range throughout the region $\mathcal{D}$ the relationship thus engendered between X_0 and the partial derivative $\frac{\partial \phi}{\partial x_k}(X_0)$ defines a function; replacing X_0 by the symbol X, we denote this function by $\frac{\partial \phi}{\partial x_k}(X)$. It is also common to see this function written as $\phi_{x_k}(X) = \phi_{x_k}(x_1, x_2, ..., x_n)$. If each of these functions is continuous as a function of $X \in \mathcal{D}$ then we say that $\phi(X)$ is *continuously differentiable* in the region $\mathcal{D}$. In the work to follow we will assume continuous differentiability unless we specifically indicate otherwise.

If we combine the partial derivatives $\frac{\partial \phi}{\partial x_k}(X_0)$, $k = 1, 2, ..., n$ into an n-dimensional vector, in the obvious order, we obtain the *gradient (vector)* of the function $\phi(X)$ at the point X_0. A commonly used notation for the gradient at X_0 is $\nabla\phi(X_0)$. Again letting X_0 range throughout $\mathcal{D}$, and replacing X_0 by X we obtain the vector function

$$\nabla\phi(X) \;=\; \nabla\phi(x_1, x_2, ..., x_n) \;=\; \left(\frac{\partial\phi}{\partial x_1}(X), \frac{\partial\phi}{\partial x_2}(X), ..., \frac{\partial\phi}{\partial x_n}(X)\right).$$

In $\mathbf{R}^3$ we can write

$$\begin{aligned}\nabla\phi(X) \;&=\; \left(\frac{\partial\phi}{\partial x}(x,y,z),\; \frac{\partial\phi}{\partial y}(x,y,z), \frac{\partial\phi}{\partial z}(x,y,z)\right)\\ &=\; \frac{\partial\phi}{\partial x}(x,y,z)\,\mathbf{i} + \frac{\partial\phi}{\partial y}(x,y,z)\,\mathbf{j} + \frac{\partial\phi}{\partial z}(x,y,z)\,\mathbf{k}.\end{aligned}$$

In this way we obtain a vector field in the region $\mathcal{D}$, since the dimension of $\nabla\phi(X)$ is n, the same as the dimension of the independent vector variable X.

Example 1 If, in $\mathbf{R}^2$, we define

$$\phi(x,y) \;=\; x^2y \,+\, \frac{y^3}{3},$$

then the corresponding gradient field is

$$\nabla\phi(x,y) \;=\; \left(\frac{\partial}{\partial x}(x^2y \,+\, \frac{y^3}{3}),\; \frac{\partial}{\partial y}(x^2y \,+\, \frac{y^3}{3})\right) \;=\; \big(2xy,\; x^2 \,+\, y^2\big).$$

The gradient vector at the point $(1,2)$, for example, is $(4,\, 5)$.

A Notational Convention Vectors can be written in row form: $(x_1, x_2, ..., x_n)$, or in column form:

$$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}.$$

These vectors are said to be **transposes** of each other; if the row vector is designated as X, then the corresponding column vector is designated as X^*, and vice versa. If X and Y are n-dimensional column and row vectors, respectively, it is common to write

$$X \cdot Y \;=\; Y^*(X) \;=\; (\,y_1 \quad y_2 \quad ... \quad y_n\,) \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \;=\; \sum_{k=1}^{n} y_k\, x_k.$$

If W is an n-dimensional row vector then

$$W\,X \;\equiv\; \sum_{k=1}^{n} w_k\, x_k .$$

Gradients are usually written as **row** vectors; thus the preceding implies that

$$\nabla\phi(X_0)\,X \;\equiv\; \sum_{k=1}^{n} \frac{\partial\phi}{\partial x_k}(X_0)\,x_k .$$

We will use the notation repeatedly in the sequel.

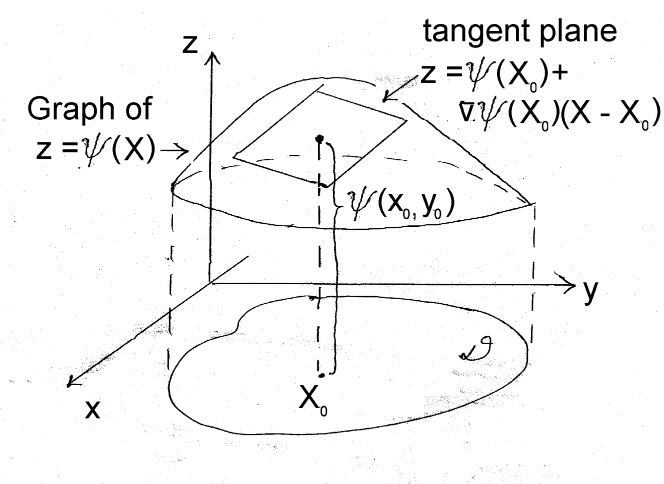
First Order Linear Approximation to a Scalar Function

If $\phi(X)$ is a continuously differentiable function of X in the region $\mathcal{D}$ which contains the point X_0 the gradient $\nabla\phi(X_0)$ can be used to approximate values of the function $\phi(X)$ at points $X \in \mathcal{D}$ lying close to X_0. The formula for this approximation relationship is

$$\phi(X) \approx \phi(X_0) \,+\, \nabla\phi(X_0)\;(X \,-\, X_0) \;=\; \sum_{k=1}^{n} \frac{\partial\phi}{\partial x_k}(X_0)\;(x_k \,-\, x_{0,k})\,.$$

The right hand side defines a *linear function* of the vector variable X, which we call the *linear approximation to $\phi(X)$ at the point X_0*. We will denote this linear function by $\mathcal{L}_{\phi,X_0}(X)$.

Figure 1



This relationship can be illustrated for $n = 2$, $\phi = \phi(x,y)$ by noting that in this case the graph of $z = \mathcal{L}_{\phi,x_0,y_0}(x,y)$ is a plane in $\mathbf{R}^3$ which is tangent to the graph of $z = \phi(x,y)$ at the point $(x_0, y_0, \phi(x_0,y_0))$.

Example 2 Let $w = x^2 + y^3 + z^4$. Taking the "base point" $X_0 = (x_0, y_0, z_0) = (1,1,1)$ we have $w_0 \equiv \phi(1,1,1) = 3$. Now

$$\nabla\phi(x,y,z) = \begin{pmatrix} 2x & 3y & 4z \end{pmatrix}; \qquad \nabla\phi(1,1,1) = \begin{pmatrix} 2 & 3 & 4 \end{pmatrix}.$$

Accordingly, the linear approximation to $\phi(x,y,z)$ as given by the above formula is

$$w = 3 + \begin{pmatrix} 2 & 3 & 4 \end{pmatrix} \begin{pmatrix} x-1 \\ y-1 \\ z-1 \end{pmatrix} = \begin{array}{c} 3 + 2(x-1) + 3(y-1) + 4(z-1) = \\ \\ -6 + 2x + 3y + 4z \end{array}.$$

The linear approximation can be used to estimate the value of ϕ at points X near the "base point" X_0 (with some degree of error corresponding to the term $o\Big(\|X - X_0\|\Big)$). Thus, in the example just studied, if we take $(x,y,z) = (1.1,\ .9,\ 1.05)$, the linear approximation gives the value

$$\mathcal{L}_{\phi,(1,1,1)}(1.1,\ .9,\ 1.05) = -6 + 2(1.1) + 3(.9) + 4(1.05) = 3.1,$$

whereas the actual value is

$$\phi(1.1,\ .9,\ 1.05) = (1.1)^2 + (.9)^3 + (1.05)^4 = 3.1545.$$

Basis for the Approximation; Error We want to see why the linear approximation, involving the gradient vector, works as it does, and to obtain a bound on the error incurred in use of this approximation. For clarity and brevity we carry out the calculations for the case $n = 2$; the argument in a larger number of variables is essentially identical.

Let $\phi(x,y)$ be continuously differentiable in a region $\mathcal{D}$, let $X_0 = (x_0,y_0)$ be a point in $\mathcal{D}$ and let $\mathcal{N}(X_0, r)$ be a disc of positive radius, r, centered at X_0, which still lies in $\mathcal{D}$. Let $X = (x,y)$ be a point in $\mathcal{N}(X_0,r)$. We consider the difference

$$\phi(X) - \phi(X_0) = \phi(x,y) - \phi(x_0,y_0) = \phi(x,y) - \phi(x,y_0) + \phi(x,y_0) - \phi(x_0,y_0).$$

Applying the **Mean Value Theorem** twice, we have

$$\phi(x,y) - \phi(x,y_0) + \phi(x,y_0) - \phi(x_0,y_0) = \frac{\partial\phi}{\partial y}(x,\eta)\,(y-y_0) + \frac{\partial\phi}{\partial x}(\xi,y_0)\,(x-x_0),$$

where ξ lies between x_0 and x and η lies between y_0 and y. Then, clearly,

$$\phi(x,y) - \phi(x_0,y_0) = \frac{\partial\phi}{\partial y}(x_0,y_0)\,(y-y_0) + \frac{\partial\phi}{\partial x}(x_0,y_0)\,(x-x_0)$$

$$+ \left(\frac{\partial\phi}{\partial y}(x,\eta) - \frac{\partial\phi}{\partial y}(x_0,y_0)\right)(y-y_0) + \left(\frac{\partial\phi}{\partial x}(\xi,y_0) - \frac{\partial\phi}{\partial x}(x_0,y_0)\right)(x-x_0)$$
$$= \nabla\phi(x_0,y_0)\,(X-X_0) + \gamma(X_0,X)\,(X-X_0),$$

where, in the error term $\gamma(X_0,X)\,(X-X_0)$,

$$\gamma(X_0,X) = \left(\ \tfrac{\partial\phi}{\partial y}(x,\eta) - \tfrac{\partial\phi}{\partial y}(x_0,y_0) \quad \tfrac{\partial\phi}{\partial x}(\xi,y_0) - \tfrac{\partial\phi}{\partial x}(x_0,y_0)\ \right).$$

Applying the Schwarz inequality to the error term we have

$$|\gamma(X_0,X)\,(X-X_0)| \le \|\gamma(X_0,X)\|\,\|X-X_0\|.$$

Because $\phi(X)$ is continuously differentiable, $\lim_{X\to X_0}\|\gamma(X_0,X)\| =$

$$\lim_{X\to X_0}\|(\ \tfrac{\partial\phi}{\partial y}(x,\eta) - \tfrac{\partial\phi}{\partial y}(x_0,y_0) \quad \tfrac{\partial\phi}{\partial x}(\xi,y_0) - \tfrac{\partial\phi}{\partial x}(x_0,y_0)\ )\| = 0.$$

This means that $|\gamma(X_0,X)\,(X-X_0)| \le \|\gamma(X_0,X)\|\,\|X-X_0\|$ has the property

$$\lim_{X\to X_0}\frac{|\gamma(X_0,X)\,(X-X_0)|}{\|X-X_0\|} = 0,$$

a relationship which we express by saying that

$$\phi(X) = \phi(X_0) + \nabla\phi(X_0)\,(X-X_0) + o\,(\|X-X_0\|)\,;$$

i.e., the error in the approximation tends to zero, as $\|X-X_0\| \to 0$, more rapidly than any multiple of $\|X-X_0\|$.

It should be noted that the error estimate just obtained depends critically on the continuity of the partial derivatives of the function $\phi(X)$. Without that property one can show, with appropriate examples, that such an estimate need not hold. The approximation relationship holds in exactly the same way in $\mathbf{R}^n$, for a general positive integer n, as we have shown it to hold for $n = 2$.

Proposition (Chain Rule) Suppose $X(t)$ is a continuously differentiable curve in $\mathbf{R}^n$ and $\phi(X)$ is a continuously differentiable function of $X \in \mathcal{D} \subset \mathbf{R}^n$. Then

$$\frac{d}{dt}\phi(X(t)) = \nabla\phi(X(t))\, X^{'}(t).$$

Fix a value of t, say t_0. Then $X(t) = X(t_0) + X^{'}(t_0)(t-t_0) + o\,(|t-t_0|)$ as $t \to t_0$. Using the formula for the first order approximation to $\phi(X)$ at $X = X(t_0)$ we find that

$$\begin{aligned}
\phi(X(t)) &= \phi(X(t_0)) + \nabla\phi(X(t_0))\,(X(t) - X(t_0)) + o\,(\|X(t) - X(t_0)\|) \\
&= \phi(X(t_0)) + \nabla\phi(X(t_0))\,(X^{'}(t_0)(t-t_0) + o\,(|t-t_0|) - X(t_0)) \\
&\quad + o\left(\left\|\nabla\phi(X(t_0))\,(X^{'}(t_0)(t-t_0) + o\,(|t-t_0|)\right\|\right) \\
&= \phi(X(t_0) + \nabla\phi(X(t_0))\,(X^{'}(t_0)(t-t_0) + o\,(|t-t_0|)\,.
\end{aligned}$$

Dividing by $t - t_0$ we then have

$$\frac{\phi(X(t)) - \phi(X(t_0))}{t - t_0} = \nabla\phi(X(t_0))\, X^{'}(t_0) + \frac{o\,(|t-t_0|)}{t - t_0}.$$

Letting $t \to t_0$ the difference quotient on the left approaches $\frac{d}{dt}\phi(X(t))\Big|_{t=t_0}$ while the expression at the right approaches $\nabla\phi(X(t_0))\, X^{'}(t_0)$. Thus

$$\frac{d\phi}{dt}(X(t))\Bigg|_{t=t_0} = \nabla\phi(X(t_0))\, X^{'}(t_0).$$

Since t_0 could be any value of t, the result follows.

An important special case occurs when we take $X(t)$ to be the straight line

$$X(t) = X_0 + t\,U, \qquad \|U\| = 1.$$

We then obtain, since $\frac{dX}{dt} \equiv U$,

$$\frac{d\phi}{dt}(X_0) = \frac{d\phi}{dt}(X(0)) = \nabla\phi(X_0)\, U \equiv \frac{dX}{dU}(X_0).$$

This is called the *directional derivative* of $\phi(X)$ in the direction U at the point X_0. It represents the slope of the graph of $\phi(X)$ in the direction corresponding to the unit vector U. Using the Schwarz inequality again we obtain

$$\left|\frac{d\phi}{dt}(X_0)\right| \;=\; |\nabla\phi(X_0)\,U| \;\leq\; \|\nabla\phi(X_0)\|\,\|U\|$$

with the inequality holding strictly unless U is collinear with $\nabla\phi(X_0)$. The two unit vectors collinear with $\nabla\phi(X_0)$ are $U_\pm \;=\; \pm\frac{\nabla\phi(X_0)^*}{\|\nabla\phi(X_0)\|}$, provided $\nabla\phi(X_0)^*$ is not the zero (column) vector. For those the corresponding directional derivatives are

$$\nabla\phi(X_0)\,U_\pm \;=\; \pm\nabla\phi(X_0)\,\frac{\nabla\phi(X_0)^*}{\|\nabla\phi(X_0)\|} \;=\; \pm\,\|\nabla\phi(X_0)\|\,.$$

The gradient vectors U_+, U_- thus correspond to the extreme values of the directional derivative $\frac{\partial\phi}{\partial U}(X_0)$; the "uphill" direction U_+ yields its maximum and the "downhill" direction U_- yields its minimum, those maximum and minimum values being $\|\nabla\phi(X_0)\|$ and $-\,\|\nabla\phi(X_0)\|$, respectively.

Suggested Exercises

1. Find the gradients, as vector functions of X, for

$$\phi(X) = \phi(x,\,y) = \sin(x^2+y^2)+2x-3y; \qquad \psi(X) = \psi(x,\,y,\,z) = x^2+2\,x\,y^2+y^4-z^2.$$

2. Find the linear approximation to $\psi(x,\,y,\,z)$, as defined in 2.,at the point $X = \begin{pmatrix} -1 \\ 1 \\ 2 \end{pmatrix}$. Use the linear approximation to estimate the values of ϕ at the points $\begin{pmatrix} -1.2 \\ 1.2 \\ 2.3 \end{pmatrix}$ and $\begin{pmatrix} -1.4 \\ 1.4 \\ 2.5 \end{pmatrix}$. Where is the approximation most effective? Why?

3. Let $\phi(x,\,y)$ be as defined in 1. Find all unit vectors U such that $\frac{\partial\phi}{\partial U}(-1,\,2) \;=\; 0$.

4. Let $X(t)$ be the curve described by $X(t) \;=\; \begin{pmatrix} t+1 \\ -t^3 \\ t^2-t \end{pmatrix}$ and let $\psi(X)$ be as defined in 1. Compute $\frac{d}{dt}\psi(X(t))$ as a function of t and give its value when $t \;=\; 1$.