Some comments on Grad, Div, Curl in Chap. 8

- Steepest Descent Method (Gradient Method) for optimization problems
  see next slides

- Some basic formulas for grad, div, curl

$$\underline{\nabla}(fg) = f\,\underline{\nabla}g + g\underline{\nabla}f$$

$$\underline{\nabla}(f/g) = (1/g^2)(g\underline{\nabla}f - f\,\underline{\nabla}g)$$

$$\underline{\nabla}\cdot(f\,\underline{v}) = f\,\underline{\nabla}\cdot\underline{v} = \underline{v}\cdot\underline{\nabla}f \qquad \underline{\nabla}\cdot(\underline{v}\underline{w}) = \underline{w}(\underline{\nabla}\cdot\underline{v}) = \underline{v}\cdot\underline{\nabla}\underline{w}$$

$$\underline{\nabla}\cdot(f\,\underline{\nabla}g) = f\,\underline{\nabla}\cdot\underline{\nabla}g + \underline{\nabla}g\cdot\underline{\nabla}f = f\,\underline{\nabla}^2 g + \underline{\nabla}g\cdot\underline{\nabla}f$$

$$\underline{\nabla}^2 f = \underline{\nabla}\cdot\underline{\nabla}f$$

$$\underline{\nabla}^2(fg) = g\underline{\nabla}^2 f + 2\underline{\nabla}f\cdot\underline{\nabla}g + f\,\underline{\nabla}^2 g$$

$$\underline{\nabla}\times(f\,\underline{v}) = \underline{\nabla}f\times\underline{v} + f\,\underline{\nabla}\times\underline{v}$$

$$\underline{\nabla}\cdot(\underline{u}\times\underline{v}) = \underline{v}\cdot(\underline{\nabla}\times\underline{u}) - \underline{u}\cdot(\underline{\nabla}\times\underline{v})$$

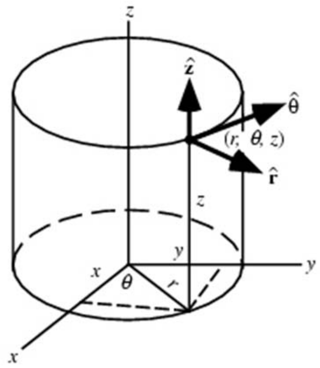
$$\underline{\nabla}\times(\underline{\nabla}f) = \underline{0}$$

$$\underline{\nabla}\cdot(\underline{\nabla}\times\underline{v}) = 0$$

- The form of grad, div, curl in curvilinear coordinates (see Appendix A3.4)

Special cases: cylindrical, spherical coordinates

(a) Cylindrical coordinates:



$$x = r\cos\theta,\ \ y = r\sin\theta,\ \ z = z$$

$$r = \sqrt{x^2 + y^2},\ \ \theta = \tan^{-1}(y/x),\ \ z = z$$

$$\frac{\partial}{\partial x} = (\cos\theta)\frac{\partial}{\partial r} + (-\frac{\sin\theta}{r})\frac{\partial}{\partial\theta} + (0)\frac{\partial}{\partial z}$$

$$\frac{\partial}{\partial y} = (\sin\theta)\frac{\partial}{\partial r} + (\frac{\cos\theta}{r})\frac{\partial}{\partial\theta} + (0)\frac{\partial}{\partial z}$$

$$\frac{\partial}{\partial z} = (0)\frac{\partial}{\partial r} + (0)\frac{\partial}{\partial\theta} + (1)\frac{\partial}{\partial z}$$

$$\underline{\delta}_x = (\cos\theta)\underline{\delta}_r + (-\sin\theta)\underline{\delta}_\theta + (0)\underline{\delta}_z$$

$$\underline{\delta}_y = (\sin\theta)\underline{\delta}_r + (\cos\theta)\underline{\delta}_\theta + (0)\underline{\delta}_z$$

$$\underline{\delta}_z = (0)\underline{\delta}_r + (0)\underline{\delta}_\theta + (1)\underline{\delta}_z$$

$$\underline{\delta}_r = \frac{\partial r}{\partial x}\underline{\delta}_x + \frac{\partial r}{\partial y}\underline{\delta}_y + \frac{\partial r}{\partial z}\underline{\delta}_z = (\cos\theta)\underline{\delta}_x + (\sin\theta)\underline{\delta}_y + (0)\underline{\delta}_z$$

$$\underline{\delta}_\theta = \frac{\partial s_\theta}{\partial x}\underline{\delta}_x + \frac{\partial s_\theta}{\partial y}\underline{\delta}_y + \frac{\partial s_\theta}{\partial z}\underline{\delta}_z = (-\sin\theta)\underline{\delta}_x + (\cos\theta)\underline{\delta}_y + (0)\underline{\delta}_z$$

$$\underline{\delta}_z = (0)\underline{\delta}_x + (0)\underline{\delta}_y + (1)\underline{\delta}_z$$

$$ds_\theta = r\,d\theta$$

$$\frac{\partial}{\partial r}\underline{\delta}_r = 0,\ \frac{\partial}{\partial r}\underline{\delta}_\theta = 0,\ \frac{\partial}{\partial r}\underline{\delta}_z = 0$$

$$\frac{\partial}{\partial \theta}\underline{\delta}_r = \underline{\delta}_\theta,\ \frac{\partial}{\partial \theta}\underline{\delta}_\theta = -\underline{\delta}_r,\ \frac{\partial}{\partial \theta}\underline{\delta}_z = 0$$

$$\frac{\partial}{\partial z}\underline{\delta}_r = 0,\ \frac{\partial}{\partial z}\underline{\delta}_\theta = 0,\ \frac{\partial}{\partial z}\underline{\delta}_z = 0$$

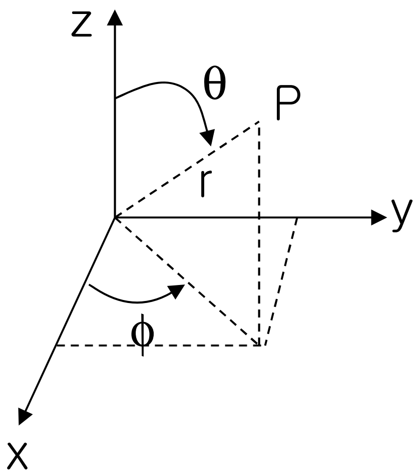
$$\Rightarrow \underline{\nabla} = \underline{\delta}_x \frac{\partial}{\partial x} + \underline{\delta}_y \frac{\partial}{\partial y} + \underline{\delta}_z \frac{\partial}{\partial z}$$

$$= \underline{\delta}_r \frac{\partial}{\partial r} + \underline{\delta}_\theta \frac{1}{r}\frac{\partial}{\partial \theta} + \underline{\delta}_z \frac{\partial}{\partial z}$$

$$\Rightarrow \underline{\nabla}\cdot\underline{v} = \left(\underline{\delta}_r \frac{\partial}{\partial r} + \underline{\delta}_\theta \frac{1}{r}\frac{\partial}{\partial \theta} + \underline{\delta}_z \frac{\partial}{\partial z}\right)\cdot\left(\underline{\delta}_r v_r + \underline{\delta}_\theta v_\theta + \underline{\delta}_z v_z\right)$$

$$= \frac{\partial v_r}{\partial r} + \frac{1}{r}\frac{\partial v_\theta}{\partial \theta} + \frac{v_r}{r} + \frac{\partial v_z}{\partial z}$$

(b) Spherical coordinates:



$$x = r\sin\theta\cos\phi,\ \ y = r\sin\theta\sin\phi,\ \ z = r\cos\theta$$

$$r = \sqrt{x^2+y^2+z^2},\ \ \theta = \tan^{-1}(\sqrt{x^2+y^2}/z),\ \ \phi = \tan^{-1}(y/x)$$

$$\frac{\partial}{\partial x} = (\sin\theta\cos\phi)\frac{\partial}{\partial r} + \left(\frac{\cos\theta\cos\phi}{r}\right)\frac{\partial}{\partial \theta} + \left(-\frac{\sin\phi}{r\sin\theta}\right)\frac{\partial}{\partial \phi}$$

$$\frac{\partial}{\partial y} = (\sin\theta\sin\phi)\frac{\partial}{\partial r} + \left(\frac{\cos\theta\sin\phi}{r}\right)\frac{\partial}{\partial \theta} + \left(\frac{\cos\phi}{r\sin\theta}\right)\frac{\partial}{\partial \phi}$$

$$\frac{\partial}{\partial z} = (\cos\theta)\frac{\partial}{\partial r} + \left(-\frac{\sin\theta}{r}\right)\frac{\partial}{\partial \theta} + (0)\frac{\partial}{\partial \phi}$$

$$\underline{\delta}_r = (\sin\theta\cos\phi)\underline{\delta}_x + (\sin\theta\sin\phi)\underline{\delta}_y + (\cos\theta)\underline{\delta}_z$$
$$\underline{\delta}_\theta = (\cos\theta\cos\phi)\underline{\delta}_x + (\cos\theta\sin\phi)\underline{\delta}_y + (-\sin\theta)\underline{\delta}_z$$
$$\underline{\delta}_\phi = (-\sin\phi)\underline{\delta}_x + (\cos\phi)\underline{\delta}_y + (0)\underline{\delta}_z$$

$$ds_\theta = r\,d\theta,\ \ ds_\phi = r\sin\theta\,d\phi$$

$$\underline{\delta}_x = (\sin\theta\cos\phi)\underline{\delta}_r + (\cos\theta\cos\phi)\underline{\delta}_\theta + (-\sin\phi)\underline{\delta}_\phi$$
$$\underline{\delta}_y = (\sin\theta\sin\phi)\underline{\delta}_r + (\cos\theta\sin\phi)\underline{\delta}_\theta + (\cos\phi)\underline{\delta}_\phi$$
$$\underline{\delta}_z = (\cos\theta)\underline{\delta}_r + (-\sin\theta)\underline{\delta}_\theta + (0)\underline{\delta}_\phi$$

$$\frac{\partial}{\partial r}\underline{\delta}_r = 0,\ \frac{\partial}{\partial r}\underline{\delta}_\theta = 0,\ \frac{\partial}{\partial r}\underline{\delta}_\phi = 0$$
$$\frac{\partial}{\partial \theta}\underline{\delta}_r = \underline{\delta}_\theta,\ \frac{\partial}{\partial \theta}\underline{\delta}_\theta = -\underline{\delta}_r,\ \frac{\partial}{\partial \theta}\underline{\delta}_\phi = 0$$
$$\frac{\partial}{\partial \phi}\underline{\delta}_r = \underline{\delta}_\phi\sin\theta,\ \ \frac{\partial}{\partial \phi}\underline{\delta}_\theta = \underline{\delta}_\phi\cos\theta,\ \ \frac{\partial}{\partial \phi}\underline{\delta}_\phi = -\underline{\delta}_r\sin\theta - \underline{\delta}_\theta\cos\theta$$

$$\Rightarrow \underline{\nabla} = ???$$

$$\Rightarrow \underline{\nabla}\cdot\underline{v} = ???$$

and a_2

$$a_0 = f(0) = 3$$
$$a_0 + 2.618\,034a_1 + 6.854a_2 = f(2.618\,034) = 5.236\,61$$
$$a_1 = f'(0) = -3$$

Solving the three equations simultaneously, we get $a_0 = 3$, $a_1 = -3$ and $a_2 = 1.4722$. The minimum point of the parabolic curve using Eq. (5.31) is given as $\bar{\alpha} = 1.0189$ and $f(\bar{\alpha}) = 0.694\,43$. This estimate can be improved using an iteration as before.

Note that in the preceding an estimate of the minimum point of the function $f(\alpha)$ can be found in only two function evaluations. Since the slope $f'(0) = \mathbf{c}^{(k)} \cdot \mathbf{d}^{(k)}$ is known for multidimensional problems, no additional calculations are required to evaluate it at $\alpha = 0$. ||

5.4 STEEPEST DESCENT METHOD

In the previous section we assumed that a search direction in the design space is known and we tackled the problem of step size determination. In this and subsequent sections we shall address the question of determination of the search direction **d**. *The basic requirement for* **d** *is that the cost function be reduced if we move a small distance along the direction. This will be called the descent direction.*

Several methods are available for determining a descent direction for unconstrained optimization problems. The steepest descent method or the *gradient method* is the simplest, the oldest and probably the best known numerical method for unconstrained optimization. The philosophy of the method, introduced by Cauchy in 1847, is to find the direction **d** at the current iteration in which the cost function $f(\mathbf{x})$ decreases most rapidly, at least locally. It is due to this philosophy that the method is called the steepest descent search technique. Also, properties of the gradient vector are used in the iterative process which is the reason for its alternate name: the *gradient method*. The steepest descent method is a *first-order method* since only the gradient of the cost function is calculated and used to evaluate the search direction. Later, we shall discuss *second-order methods* where Hessian of the function will be used in determining the search direction. We shall first study properties of the gradient vector of a scalar function before stating an algorithm for the method.

5.4.1 Properties of Gradient Vector

The gradient vector of a scalar function $f(x_1, x_2, \ldots, x_n)$ was defined in Chapter 3. Just as a reminder, we define it again as the column vector:

$$\nabla f = \left[\frac{\partial f}{\partial x_1}\ \frac{\partial f}{\partial x_2} \cdots \frac{\partial f}{\partial x_n}\right]^T \equiv \mathbf{c} \tag{5.32}$$

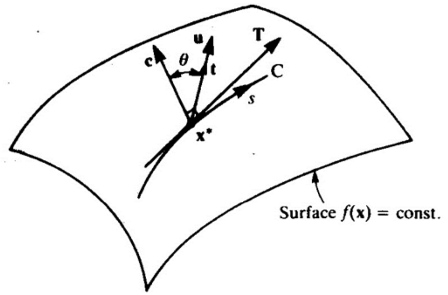


FIGURE 5.11
Gradient vector for the surface $f(\mathbf{x}) = \text{constant}$ at the point $\mathbf{x}^*$.

To simplify the notation, we shall use vector **c** to represent gradient of the scalar function $f(\mathbf{x})$; that is, $c_i = \partial f / \partial x_i$. We shall use a superscript to denote the point at which this vector is calculated, as

$$\mathbf{c}^{(k)} = \mathbf{c}(\mathbf{x}^{(k)}) = \left[\frac{\partial f(\mathbf{x}^{(k)})}{\partial x_i}\right]^T \tag{5.33}$$

The gradient vector has several *properties* that are used in the steepest descent method. Since proofs of the properties are also quite instructive, they are also given.

Property 1. *The gradient vector* **c** *of a function* $f(x_1, x_2, \ldots, x_n)$ *at the point* $\mathbf{x}^* = (x_1^*, x_2^*, \ldots, x_n^*)$ *is orthogonal (normal) to the tangent plane for the surface* $f(x_1, x_2, \ldots, x_n) = \text{constant}$.

This is an important property of the gradient vector shown graphically in Fig. 5.11. It shows the surface $f(\mathbf{x}) = \text{constant}$; $\mathbf{x}^*$ is a point on the surface; C is any curve on the surface through the point $\mathbf{x}^*$; **T** is a vector tangent to the curve C at the point $\mathbf{x}^*$, **u** is any unit vector; and **c** is the gradient vector at $\mathbf{x}^*$. According to the above property, vectors **c** and **T** are normal to each other, i.e. their dot product is zero, $\mathbf{c} \cdot \mathbf{T} = 0$.

Proof. To show this, we take any curve C on the surface $f(x_1, x_2, \ldots, x_n) = \text{constant}$, as shown in Fig. 5.11. Let the curve pass through the point $\mathbf{x}^* = (x_1^*, x_2^*, \ldots, x_n^*)$. Also, let s be any parameter along C. Then a unit tangent vector **T** along C at the point $\mathbf{x}^*$ is given as

$$\mathbf{T} = \left[\frac{\partial x_1}{\partial s}\ \frac{\partial x_2}{\partial s} \cdots \frac{\partial x_n}{\partial s}\right]^T \tag{a}$$

Since $f(\mathbf{x}) = \text{constant}$, the derivative of f along the curve C is zero, i.e.

$$\frac{df}{ds} = 0 \quad \text{(directional derivative of } f\text{)}$$

Or, using the chain rule of differentiation

$$\frac{df}{ds} = \frac{\partial f}{\partial x_1}\frac{\partial x_1}{\partial s} + \cdots + \frac{\partial f}{\partial x_n}\frac{\partial x_n}{\partial s} = 0 \tag{b}$$

Writing Eq. (b) in the vector form after identifying $\partial f/\partial x_i$ and $\partial x_i/\partial s$ (from Eq. a) as components of the gradient and the unit tangent vectors, we obtain

$$\mathbf{c}\cdot\mathbf{T}=0, \quad \text{or} \quad \mathbf{c}^T\mathbf{T}=0$$

Since the dot product of the gradient vector **c** with the tangential vector **T** is zero, the vectors are normal to each other. But **T** is any tangent vector at $\mathbf{x}^*$, so **c** is orthogonal to the tangent plane for the surface $f(\mathbf{x})=$ constant at the point $\mathbf{x}^*$. ||

Property 2. *The second property is that gradient represents a direction of maximum rate of increase for the function $f(\mathbf{x})$ at the point $\mathbf{x}^*$.*

Proof. To show this, let **u** be a unit vector in any direction that is not tangent to the surface. This is shown in Fig. 5.11. Let t be a parameter along **u**. The derivative of $f(\mathbf{x})$ in the direction **u** at the point $\mathbf{x}^*$ (i.e. directional derivative of f) is given as

$$\frac{df}{dt}=\lim_{\varepsilon\to 0}\frac{f(\mathbf{x}+\varepsilon\mathbf{u})-f(\mathbf{x})}{\varepsilon} \tag{c}$$

where ε is a small number. Using Taylor series expansion

$$f(\mathbf{x}+\varepsilon\mathbf{u})=f(\mathbf{x})+\varepsilon\left[u_1\frac{\partial f}{\partial x_1}+u_2\frac{\partial f}{\partial x_2}+\ldots+u_n\frac{\partial f}{\partial x_n}\right]+0(\varepsilon^2)$$

where u_i are components of the unit vector **u** and $0(\varepsilon^2)$ are terms of order ε^2. Rewriting the foregoing equation,

$$f(\mathbf{x}+\varepsilon\mathbf{u})-f(\mathbf{x})=\varepsilon\sum_{i=1}^{n}u_i\frac{\partial f}{\partial x_i}+0(\varepsilon^2) \tag{d}$$

Substituting Eq. (d) into Eq. (c) and taking the indicated limit, we get

$$\frac{df}{dt}=\sum_{i=1}^{n}u_i\frac{\partial f}{\partial x_i}=\mathbf{c}\cdot\mathbf{u}=\mathbf{c}^T\mathbf{u} \tag{e}$$

Using the definition of the dot product for Eq. (e),

$$\frac{df}{dt}=\|\mathbf{c}\|\;\|\mathbf{u}\|\cos\theta \tag{f}$$

where θ is the angle between the **c** and **u** vectors. The right-hand side of Eq. (f) will have extreme value when $\theta=0$ or 180°. When $\theta=0$, vector **u** is along **c** and $\cos\theta=1$. Therefore, from Eq. (f), df/dt represents the maximum rate of increase for $f(\mathbf{x})$ when $\theta=0$. Similarly, when $\theta=180°$, vector **u** points in the negative **c** direction. Therefore, from Eq. (f), df/dt represents the maximum rate of decrease for $f(\mathbf{x})$ when $\theta=180°$. ||

According to the foregoing property of the gradient vector, if we need to move away from the surface $f(\mathbf{x})=$ constant, the function increases most rapidly along the gradient vector compared to a move in any other direction. In Fig. 5.11, a small move along the direction **c** will result in a larger increase in the function compared to a similar move along the direction **u**. Of course, any small move along the direction **T** results in no change in the function, since **T** is tangent to the surface.

Property 3. *The maximum rate of change of $f(\mathbf{x})$ at any point $\mathbf{x}^*$ is the magnitude of the gradient vector.*

Proof. Since **u** is a unit vector, the maximum value of df/dt from Eq. (f) is given as

$$\max\left|\frac{df}{dt}\right|=\|\mathbf{c}\|$$

However, for $\theta=0$, **u** is in the direction of the gradient vector. Therefore, the magnitude of the gradient represents the maximum rate of change for the function $f(\mathbf{x})$. ||

These properties show that gradient vector at any point $\mathbf{x}^*$ represents a direction of maximum increase in the function $f(\mathbf{x})$ and the rate of increase is the magnitude of the vector. Gradient is therefore called a direction of steepest ascent for the function $f(\mathbf{x})$.

Example 5.8 Verification of properties of the gradient vector. Verify the properties of the gradient vector for the function $f(\mathbf{x})=25x_1^2+x_2^2$ at the point $\mathbf{x}^{(0)}=(0.6, 4)$.

Solution. Figure 5.12 shows in the x_1-x_2 plane the iso-cost contours of value 25 and 100 for the function f. The value of the function at (0.6, 4) is $f(0.6, 4)=25$. The gradient of the function at (0.6, 4) is given as

$$\mathbf{c}=\nabla f(0.6, 4)=(\partial f/\partial x_1, \partial f/\partial x_2)$$
$$=(50x_1, 2x_2)=(30, 8)$$
$$\|\mathbf{c}\|=\sqrt{30*30+8*8}=31.048\,35$$

Therefore a unit vector along the gradient is given as

$$\mathbf{C}=\mathbf{c}/\|\mathbf{c}\|=(0.966\,235, 0.257\,663)$$

Using the given function, a vector tangent to the curve at the point (0.6, 4) is given as

$$\mathbf{t}=(-4, 15)$$

This vector is obtained by using the equation for the curve $25x_1^2+x_2^2=25$ and writing the tangent vector as $(\partial x_1/\partial s, \partial x_2/\partial s)$ where s is a parameter along the curve. The unit tangent vector is

$$\mathbf{T}=\mathbf{t}/\|\mathbf{t}\|=(-0.257\,663, 0.966\,235)$$

Property 1. If gradient is normal to the tangent, then $\mathbf{C}\cdot\mathbf{T}=0$. This is indeed true for the preceding data. We can also use the condition that if two lines are orthogonal, then $m_1m_2=-1$, where m_1 and m_2 are the slopes of the two lines. To calculate slope of the tangent we use the equation for the curve $25x_1^2+x_2^2=25$, or

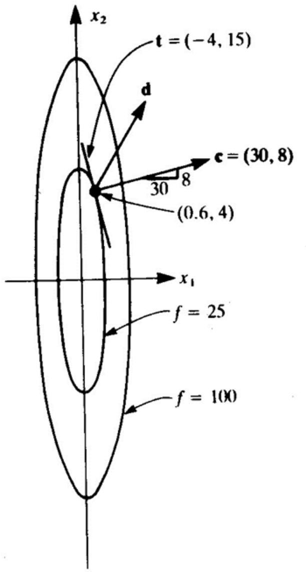


FIGURE 5.12
Iso-cost contours of function $f = 25x_1^2 + x_2^2$ for $f = 25$ and 100.

$x_2 = 5\sqrt{1 - x_1^2}$. Therefore, the slope of the tangent at the point (0.6, 4) is given as

$$m_1 = dx_2/dx_1 = -5x_1/\sqrt{1 - x_1^2} = -\tfrac{1}{3.75}$$

The slope of the gradient vector is $m_2 = \frac{30}{8} = 3.75$. Thus $m_1 m_2$ is, indeed, −1, and the two lines are normal to each other.

Property 2. Consider any arbitrary direction

$$\mathbf{d} = (0.501\,034,\ 0.865\,430)$$

at the point (0.6, 4) as shown in Fig. 5.12. If **C** is the direction of steepest ascent, then the function should increase more rapidly along **C** than along **d**. Let us choose a step size $\alpha = 0.1$ and calculate two points, one along **C** and the other along **d** as

$$\mathbf{x}^{(1)} = \mathbf{x}^{(0)} + \alpha \mathbf{C}$$

$$= \begin{bmatrix} 0.6 \\ 4.0 \end{bmatrix} + 0.1 \begin{bmatrix} 0.966\,235 \\ 0.257\,663 \end{bmatrix} = \begin{bmatrix} 0.696\,623\,5 \\ 4.025\,766\,3 \end{bmatrix}$$

$$\mathbf{x}^{(2)} = \mathbf{x}^{(0)} + \alpha \mathbf{d}$$

$$= \begin{bmatrix} 0.6 \\ 4.0 \end{bmatrix} + 0.1 \begin{bmatrix} 0.501\,034 \\ 0.865\,430 \end{bmatrix} = \begin{bmatrix} 0.650\,103\,4 \\ 4.085\,430\,0 \end{bmatrix}$$

Now, we calculate the function at these points and compare their values:

$$f(\mathbf{x}^{(1)}) = 28.3389$$
$$f(\mathbf{x}^{(2)}) = 27.2566$$

Since $f(\mathbf{x}^{(1)}) > f(\mathbf{x}^{(2)})$, the function increases more rapidly along **C** than along **d**.

Property 3. If the magnitude of the gradient vector represents the maximum rate of change of $f(\mathbf{x})$, then $(\mathbf{c} \cdot \mathbf{c}) > (\mathbf{c} \cdot \mathbf{d})$. $(\mathbf{c} \cdot \mathbf{c}) = 964.0$ and $(\mathbf{c} \cdot \mathbf{d}) = 21.9545$. Therefore, the gradient vector satisfies this property also.

Note that the last two properties are valid only in a local sense, i.e. only in a small neighborhood of the point at which the gradient is evaluated. ||

5.4.2 Steepest Descent Algorithm

The properties of the gradient vector can be used to define an iterative algorithm for the unconstrained optimization problems. The direction of maximum decrease in the cost function is the negative of its gradient at the given point **x**. Any small move in the negative gradient direction will result in the maximum local rate of decrease in the cost function. The negative gradient vector then represents a *direction of steepest descent* for the cost function. This result is summarized in the following theorem.

Theorem 5.1 Steepest descent direction. Let $f(\mathbf{x})$ be a differentiable function with respect to **x**. The direction of steepest descent for $f(\mathbf{x})$ at any point is

$$\mathbf{d} = -\mathbf{c}, \quad \text{or} \quad d_i = -c_i = \frac{-\partial f}{\partial x_i}; \quad i = 1 \text{ to } n \qquad (5.34) \; ||$$

Equation (5.34) gives a direction of change in the design space for use in Eq. (5.4). Based on the preceding discussion, the *steepest descent algorithm* is stated as follows.

Step 1. Estimate a starting design $\mathbf{x}^{(0)}$ and set the iteration counter $k = 0$. Select a convergence parameter $\varepsilon > 0$.

Step 2. Calculate the gradient of $f(\mathbf{x})$ at the point $\mathbf{x}^{(k)}$ as $\mathbf{c}^{(k)} = \nabla f(\mathbf{x}^{(k)})$. Calculate $\|\mathbf{c}^{(k)}\|$. If $\|\mathbf{c}^{(k)}\| < \varepsilon$, then stop the iterative process as $\mathbf{x}^* = \mathbf{x}^{(k)}$ is a minimum point. Otherwise, go to Step 3.

Step 3. Let the search direction at the current point $\mathbf{x}^{(k)}$ be $\mathbf{d}^{(k)} = -\mathbf{c}^{(k)}$.

Step 4. Calculate a step size α_k to minimize $f(\mathbf{x}^{(k)} + \alpha \mathbf{d}^{(k)})$. A one-dimensional search is used to determine α_k.

Step 5. Update the design as $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \alpha_k \mathbf{d}^{(k)}$. Set $k = k + 1$ and go to Step 2.

The basic idea of the steepest descent method is quite simple. We start with an initial estimate for the minimum design. The direction of steepest descent is computed at that point. If the direction is nonzero, we move as far as possible along it to reduce the cost function. At the new design point we calculate the steepest descent direction again and repeat the entire process.

Note that since $\mathbf{d} = -\mathbf{c}$, *the descent condition of Inequality (5.8) is automatically satisfied as* $\mathbf{c} \cdot \mathbf{d} = -\|\mathbf{c}\|^2 < 0$.

It is interesting to note that the successive directions of steepest descent are normal to one another, i.e.

$$\mathbf{c}^{(k)} \cdot \mathbf{c}^{(k+1)} = 0 \qquad (5.35)$$

This can be shown quite easily by using the necessary conditions to determine the optimum step size. In Step 4 of the algorithm, it is required to compute α_k to minimize $f(\mathbf{x}^{(k)} + \alpha\mathbf{d}^{(k)})$. The necessary condition for this is $df/d\alpha = 0$. Using the chain rule of differentiation, we get

$$\frac{df(\mathbf{x}^{(k+1)})}{d\alpha} = \left[\frac{\partial f(\mathbf{x}^{(k+1)})}{\partial \mathbf{x}}\right]^T \frac{\partial \mathbf{x}^{(k+1)}}{\partial \alpha}$$

which gives

$$\mathbf{c}^{(k+1)} \cdot \mathbf{d}^{(k)} = 0, \quad \text{or} \quad \mathbf{c}^{(k+1)} \cdot \mathbf{c}^{(k)} = 0 \qquad (5.36)$$

since

$$\mathbf{c}^{(k+1)} = \frac{\partial f(\mathbf{x}^{(k+1)})}{\partial \mathbf{x}} \quad \text{and} \quad \frac{\partial \mathbf{x}^{(k+1)}}{\partial \alpha} = \frac{\partial}{\partial \alpha}(\mathbf{x}^{(k)} + \alpha\mathbf{d}^{(k)}) = \mathbf{d}^{(k)}$$

In the two-dimensional case, $\mathbf{x} = (x_1, x_2)$. Figure 5.13 is a view of the design variable space. The closed curves in the figure are contours of the cost function $f(\mathbf{x})$. The figure shows several steepest descent directions that are orthogonal to each other.

5.4.2.1 LINE SEARCH TERMINATION CRITERION. The numerical methods of one-dimensional minimization are often used to perform line search in multidimensional problems. Many times the numerical methods will give an approximate or inexact value of the step size. Thus, line search termination criterion is useful to decide the accuracy of a numerical method in the step size calculation. For the exact value of the step size, Eq. (5.36) must hold, i.e.

$$\mathbf{c}^{(k+1)} \cdot \mathbf{d}^{(k)} = 0 \qquad (5.37)$$

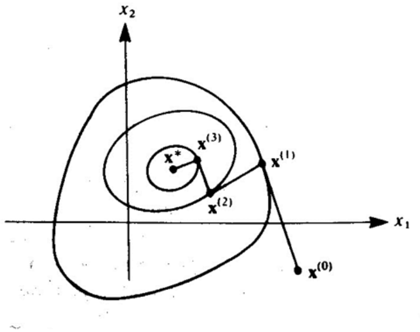


FIGURE 5.13
Orthogonal steepest descent paths.

where $\mathbf{c}^{(k+1)}$ is the gradient at $\mathbf{x}^{(k+1)}$ and $\mathbf{d}^{(k)}$ is the direction of travel in the previous iteration. Due to round off and truncation errors in computer calculations, the line search termination criterion of Eq. (5.37) cannot be satisfied precisely; nevertheless, it gives an indication of the accuracy in numerical evaluation of the step size. Note that the line search termination criterion does not depend on how the direction of descent $\mathbf{d}$ is calculated.

Example 5.9 Use of steepest descent algorithm. Minimize $f(x_1, x_2) = x_1^2 + x_2^2 - 2x_1x_2$ using the steepest descent method starting from the point $(1, 0)$.

Solution. To solve the problem, we follow steps of the steepest descent algorithm.

1. The starting design is estimated as $\mathbf{x}^{(0)} = (1, 0)$.
2. $\mathbf{c}^{(0)} = (2x_1 - 2x_2, 2x_2 - 2x_1) = (2, -2)$; $\|\mathbf{c}^{(0)}\| = 2\sqrt{2} \neq 0$.
3. Set $\mathbf{d}^{(0)} = -\mathbf{c}^{(0)} = (-2, 2)$.
4. Calculate α to minimize $f(\mathbf{x}^{(0)} + \alpha\mathbf{d}^{(0)})$ where $\mathbf{x}^{(0)} + \alpha\mathbf{d}^{(0)} = (1 - 2\alpha, 2\alpha)$

$$f(\mathbf{x}^{(0)} + \alpha\mathbf{d}^{(0)}) = (1-2\alpha)^2 + (2\alpha)^2 - 2(1-2\alpha)(2\alpha)$$
$$= 16\alpha^2 - 8\alpha + 1 \equiv f(\alpha)$$

Since this is a simple function of α, we can use necessary and sufficient conditions to solve for the optimum step length. In general, numerical one-dimensional search will have to be used to calculate α. Using the analytic approach to solve for α, we get

$$\frac{df(\alpha)}{d\alpha} = 0; \qquad 32\alpha - 8 = 0 \text{ or } \alpha_0 = 0.25$$

$$\frac{d^2 f(\alpha)}{d\alpha^2} = 32 > 0.$$

Therefore, the sufficiency condition for minimum is satisfied.

5. Updating the design $(\mathbf{x}^{(0)} + \alpha_0\mathbf{d}^{(0)})$:

$$x_2^{(1)} = 1 - 0.25(2) = 0.5$$
$$x_2^{(1)} = 0 + 0.25(2) = 0.5$$

Solving for $\mathbf{c}^{(1)}$ from the expression in Step 2, we see that $\mathbf{c}^{(1)} = (0, 0)$ which satisfies the stopping criterion. Therefore, $(0.5, 0.5)$ is a minimum point for the given problem. ||

The preceding problem is quite simple and an optimum point is obtained in only one iteration. This is because the condition number of the Hessian of the cost function is one (condition number is a scalar associated with the given matrix; refer to Section B.8 in Appendix B). In such a case, the steepest descent method will converge in just one iteration with any starting point. In general, the algorithm will require several iterations before an acceptable optimum is reached.

Example 5.10 Use of steepest descent algorithm. Minimize $f(x_1, x_2, x_3) = x_1^2 + 2x_2^2 + 2x_3^2 + 2x_1x_2 + 2x_2x_3$ using the steepest descent method with a starting design as (2, 4, 10). Select the convergence parameter ε as 0.005. Perform a line search by Golden Section search with initial step length $\delta = 0.05$ and an accuracy of 0.0001.

Solution.

1. Let $\mathbf{c} = \nabla f = (2x_1 + 2x_2, 4x_2 + 2x_1 + 2x_3, 4x_3 + 2x_2)$. Now, $\mathbf{c}^{(0)} = (12, 40, 48)$ and $\|\mathbf{c}^{(0)}\| = \sqrt{4048} = 63.6 > \varepsilon$.
2. $\mathbf{d}^{(0)} = -\mathbf{c}^{(0)} = (-12, -40, -48)$.
3. Calculate α_0 by Golden Section search to minimize $f(\mathbf{x}^{(0)} + \alpha\mathbf{d}^{(0)})$; $\alpha_0 = 0.1587$.
4. Update the design as $\mathbf{x}^{(1)} = \mathbf{x}^{(0)} + \alpha_0\mathbf{d}^{(0)}$:

$$\mathbf{x}^{(1)} = (0.0956, -2.348, 2.381)$$

5. $\mathbf{c}^{(1)} = (-4.5, -4.438, 4.828)$, $\|\mathbf{c}^{(1)}\| = 7.952 > \varepsilon$.

Note that $\mathbf{c}^{(1)} \cdot \mathbf{d}^{(0)} = 0$ which verifies the line search termination criterion. The steps in steepest descent algorithm should be repeated until the convergence criterion is satisfied. Appendix D contains the computer program and user supplied subroutines FUNCT and GRAD to implement steps of the steepest descent algorithm. The iterative history for the problem with the program is given in Table 5.3. The optimum cost function value is 0.0 and the optimum point is (0, 0, 0). Note that a large number of iterations and function evaluations are needed to reach the optimum. ||

The method of steepest descent is quite simple and robust (it is convergent). However, it has several drawbacks. These are:

1. Even if convergence of the method is guaranteed, a large number of iterations may be required for the minimization of even positive definite quadratic forms, i.e. the method can be quite slow to converge to the minimum point.
2. Information calculated at the previous iterations is not used. Each iteration is started independent of others, which is inefficient.
3. Only first-order information about the function is used at each iteration to determine the search direction. This is one reason that convergence of the method is slow. It can further deteriorate if only inaccurate line search is used. Moreover, the rate of convergence depends on the condition number of the Hessian of the cost function at the optimum point. If the condition number is large, the rate of convergence of the method is slow.
4. Practical experience with the method has shown that a substantial decrease in the cost function is achieved in the initial few iterations and the cost function decreases quite slowly in later iterations.
5. The direction of steepest descent (direction of most rapid decrease in the cost function) may be good in a local sense (in a small neighborhood) but not in a global sense.

TABLE 5.3
Optimum solution for Example 5.10 with steepest descent program

No.	x_1	x_2	x_3	f(x)	α	‖c‖
1	2.000 00E + 00	4.000 00E + 00	1.000 00E + 01	3.320 00E + 02	1.587 18E − 01	6.362 39E + 01
2	9.538 70E − 02	−2.348 71E + 00	2.381 55E + 00	1.075 03E + 01	3.058 72E − 01	7.959 22E + 00
3	1.473 84E + 00	−9.903 42E − 01	9.045 62E − 01	1.059 36E + 00	1.815 71E − 01	2.061 42E + 00
4	1.298 26E + 00	−1.134 77E + 00	6.072 28E − 01	6.737 57E − 01	6.499 89E − 01	8.139 10E − 01
5	1.085 73E + 00	−6.615 14E − 01	5.036 40E − 01	4.585 34E − 01	1.905 88E − 01	1.217 29E + 00
6	9.240 28E − 01	−7.630 36E − 01	3.718 42E − 01	3.172 18E − 01	5.880 53E − 01	5.631 54E − 01
7	7.346 84E − 01	−4.922 94E − 01	3.946 01E − 01	2.240 07E − 01	1.938 77E − 01	8.193 77E − 01
8	6.406 97E − 01	−5.484 01E − 01	2.794 74E − 01	1.589 46E − 01	5.725 54E − 01	3.991 41E − 01
9	5.350 08E − 01	−3.461 39E − 01	2.673 96E − 01	1.133 73E − 01	1.946 60E − 01	5.775 45E − 01
10	4.614 78E − 01	−3.890 14E − 01	1.939 50E − 01	8.091 74E − 02	5.697 67E − 01	2.848 37E − 01
11	3.789 02E − 01	−2.493 07E − 01	1.952 19E − 01	5.783 10E − 02	1.946 01E − 01	4.118 95E − 01
12	3.284 64E − 01	−2.786 95E − 01	1.402 91E − 01	4.131 41E − 02	5.720 88E − 01	2.033 39E − 01
13	2.715 19E − 01	−1.772 81E − 01	1.381 32E − 01	2.949 40E − 02	1.944 39E − 01	2.947 10E − 01
14	2.348 72E − 01	−1.987 04E − 01	9.963 96E − 02	2.104 99E − 02	5.726 50E − 01	1.451 12E − 01
15	1.934 49E − 01	−1.266 69E − 01	9.898 06E − 02	1.502 33E − 02	1.944 57E − 01	2.104 30E − 01
16	1.674 77E − 01	−1.418 72E − 01	7.125 41E − 02	1.071 97E − 02	5.717 77E − 01	1.035 80E − 01
17	1.381 96E − 01	−9.039 73E − 02	7.052 67E − 02	7.653 62E − 03	1.945 34E − 01	1.500 77E − 01
18	1.195 99E − 01	−1.012 63E − 01	5.081 80E − 02	5.463 41E − 03	5.708 32E − 01	7.395 59E − 02
19	9.866 59E − 02	−6.460 51E − 02	5.039 27E − 02	3.901 56E − 03	1.945 71E − 01	1.070 16E − 01
20	8.541 14E − 02	−7.232 89E − 02	3.631 34E − 02	2.786 93E − 03	5.721 47E − 01	5.280 78E − 02
21	7.044 12E − 02	−4.608 67E − 02	3.597 26E − 02	1.989 46E − 03	1.943 20E − 01	7.653 97E − 02
22	6.097 61E − 02	−5.162 11E − 02	2.592 30E − 02	1.419 91E − 03	5.743 72E − 01	3.766 50E − 02
23	5.022 96E − 02	−3.284 70E − 02	2.566 47E − 02	1.012 41E − 03	1.942 24E − 01	5.469 15E − 02
24	4.347 74E − 02	−3.680 93E − 02	1.848 53E − 02	7.219 38E − 04	5.744 09E − 01	2.685 89E − 02
25	3.581 70E − 02	−2.341 88E − 02	1.830 00E − 02	5.148 07E − 04	1.943 79E − 01	3.901 03E − 02
26	3.099 71E − 02	−2.624 87E − 02	1.317 57E − 02	3.670 49E − 04	5.714 30E − 01	1.916 85E − 02
27	2.557 04E − 02	−1.673 48E − 02	1.305 84E − 02	2.621 04E − 04	1.944 75E − 01	2.776 37E − 02
28	2.213 38E − 02	−1.874 14E − 02	9.409 25E − 03	1.871 31E − 04	5.725 54E − 01	1.368 15E − 02
29	1.824 92E − 02	−1.193 96E − 02	9.321 00E − 03	1.335 49E − 04	1.944 75E − 01	1.983 48E − 02
30	1.579 51E − 02	−1.337 52E − 02	6.714 12E − 03	9.530 54E − 05	5.718 73E − 01	9.766 61E − 03
31	1.302 74E − 02	−8.524 32E − 03	6.653 48E − 03	6.804 63E − 05	1.944 39E − 01	1.415 34E − 02
32	1.127 62E − 02	−9.547 93E − 03	4.793 62E − 03	4.856 93E − 05	5.726 20E − 01	6.970 01E − 03
33	9.296 91E − 03	−6.082 44E − 03	4.748 60E − 03	3.466 11E − 05	1.942 83E − 01	1.010 55E − 02

Optimum design variables: 8.047 87E − 03, −6.813 19E − 03, 3.421 74E − 03.
Optimum cost function value: 2.473 47E − 05.
Norm of gradient at optimum: 4.970 71E − 03.
Total no. of function evaluations: 753.

The methods discussed in later sections try to overcome some of these difficulties.

5.4.3* Scaling of Design Variables

The rate of convergence of the steepest descent method is at the most linear even for a quadratic cost function. It is possible to accelerate the rate of convergence of the steepest descent method by scaling the design variables.