linearly independent, then $L(\mathbf{d}_1, \ldots, \mathbf{d}_n) = E_n$, and hence $\mathbf{x}_{n+1}$ is a minimum point of f over E_n. This completes the proof.

Generating Conjugate Directions

In the remainder of this section, we describe several methods for generating conjugate directions of quadratic forms. These methods lead naturally to powerful algorithms for minimizing both quadratic and nonquadratic functions. In particular, we discuss the Davidon-Fletcher-Powell method, the conjugate gradient method of Fletcher and Reeves, and the method of Zangwill.

The Method of Davidon-Fletcher-Powell

The method was originally proposed by Davidon [1959] and later developed by Fletcher and Powell [1963]. The Davidon-Fletcher-Powell method is also referred to as the *variable metric method*. This method falls under the general class of *quasi-Newton procedures*, where the search directions are of the form $-\mathbf{D}_j\nabla f(\mathbf{y})$. The gradient direction is thus deflected by premultiplying it by $-\mathbf{D}_j$, where $\mathbf{D}_j$ is an $n \times n$ positive definite symmetric matrix that approximates the inverse of the Hessian matrix. For the purpose of the next step, $\mathbf{D}_{j+1}$ is formed by adding to $\mathbf{D}_j$ two symmetric matrices, each of rank one. Thus, this scheme is sometimes referred to as *rank two correction.*

Summary of the Davidon-Fletcher-Powell Method

We now summarize the Davidon-Fletcher-Powell method for minimizing a differentiable function of several variables. In particular, if the function is quadratic, then, as shown later, the method yields conjugate directions and terminates in one complete iteration, that is, after searching along each of the conjugate directions.

Initialization Step Let $\varepsilon > 0$ be the termination scalar. Choose an initial point $\mathbf{x}_1$ and an initial symmetric positive definite matrix $\mathbf{D}_1$. Let $\mathbf{y}_1 = \mathbf{x}_1$, let $k = j = 1$, and go to the main step.

Main Step

1. If $\|\nabla f(\mathbf{y}_j)\| < \varepsilon$, stop; otherwise let, $\mathbf{d}_j = -\mathbf{D}_j\nabla f(\mathbf{y}_j)$ and let λ_j be an optimal solution to the problem to minimize $f(\mathbf{y}_j + \lambda\mathbf{d}_j)$ subject to $\lambda \geq 0$. Let $\mathbf{y}_{j+1} = \mathbf{y}_j + \lambda_j\mathbf{d}_j$. If $j < n$, go to step 2. If $j = n$, let $\mathbf{y}_1 = \mathbf{x}_{k+1} = \mathbf{y}_{n+1}$, replace k by $k+1$, let $j = 1$, and repeat step 1.
2. Construct $\mathbf{D}_{j+1}$ as follows:

$$\mathbf{D}_{j+1} = \mathbf{D}_j + \frac{\mathbf{p}_j\mathbf{p}_j'}{\mathbf{p}_j'\mathbf{q}_j} - \frac{\mathbf{D}_j\mathbf{q}_j\mathbf{q}_j'\mathbf{D}_j}{\mathbf{q}_j'\mathbf{D}_j\mathbf{q}_j} \tag{8.18}$$

where

$$\mathbf{p}_j = \lambda_j\mathbf{d}_j \tag{8.19}$$

$$\mathbf{q}_j = \nabla f(\mathbf{y}_{j+1}) - \nabla f(\mathbf{y}_j) \tag{8.20}$$

Replace j by $j+1$, and repeat step 1.

8.6.4 Example

Consider the following problem

Minimize $\quad (x_1-2)^4 + (x_1 - 2x_2)^2$

The summary of the computations using the Davidon-Fletcher-Powell method is given in Table 8.13. At each iteration, for $j = 1, 2, \mathbf{d}_j$ is given by $-\mathbf{D}_j\nabla f(\mathbf{y}_j)$, where $\mathbf{D}_1$ is the identity matrix, and $\mathbf{D}_2$ is computed from (8.18) to (8.20). At iteration $k = 1$, we have $\mathbf{p}_1 = (2.7, -1.49)'$ and $\mathbf{q}_1 = (44.73, -22.72)'$ in (8.18). At iteration 2 we have $\mathbf{p}_1 = (-0.1, 0.05)'$ and $\mathbf{q}_1 = (-0.7, 0.8)'$, and finally at iteration 3, we have $\mathbf{p}_1 = (-0.02, 0.02)'$ and $\mathbf{q}_1 = (-0.14, 0.24)'$. The point $\mathbf{y}_{j+1}$ is computed by optimizing along the direction $\mathbf{d}_j$ starting from $\mathbf{y}_j$ for $j = 1, 2$. The procedure is terminated at the point $\mathbf{y}_2 = (2.115, 1.058)'$ in the fourth iteration, since $\|\nabla f(\mathbf{y}_2)\| = 0.006$ is quite small. The path taken by the method is depicted in Figure 8.17.

Lemma 8.6.5 shows that each matrix $\mathbf{D}_j$ is positive definite and $\mathbf{d}_j$ is a direction of descent.

8.6.5 Lemma

Let $\mathbf{y}_1 \in E_n$, and let $\mathbf{D}_1$ be an initial positive definite symmetric matrix. For $j = 1, \ldots, n$, let $\mathbf{y}_{j+1} = \mathbf{y}_j + \lambda_j\mathbf{d}_j$, where $\mathbf{d}_j = -\mathbf{D}_j\nabla f(\mathbf{y}_j)$, and λ_j solves the problem to minimize $f(\mathbf{y}_j + \lambda\mathbf{d}_j)$ subject to $\lambda \geq 0$. Furthermore, for $j = 1, \ldots, n-1$, let $\mathbf{D}_{j+1}$ be given by (8.18) to (8.20). If $\nabla f(\mathbf{y}_j) \neq \mathbf{0}$ for $j = 1, \ldots, n$, then $\mathbf{D}_1, \ldots, \mathbf{D}_n$ are symmetric and positive definite so that $\mathbf{d}_1, \ldots, \mathbf{d}_n$ are descent directions.

Proof

We prove the result by induction. For $j = 1$, $\mathbf{D}_1$ is symmetric and positive definite by assumption. Furthermore, $\nabla f(\mathbf{y}_1)'\mathbf{d}_1 = -\nabla f(\mathbf{y}_1)'\mathbf{D}_1\nabla f(\mathbf{y}_1) < 0$, since $\mathbf{D}_1$ is positive definite. By Theorem 4.2.1, then $\mathbf{d}_1$ is a descent direction. We will assume that the result holds true for $j \leq n-1$, and then show that it holds for $j+1$. Let $\mathbf{x}$ be a nonzero vector in E_n; then by (8.18), we have

$$\mathbf{x}'\mathbf{D}_{j+1}\mathbf{x} = \mathbf{x}'\mathbf{D}_j\mathbf{x} + \frac{(\mathbf{x}'\mathbf{p}_j)^2}{\mathbf{p}_j'\mathbf{q}_j} - \frac{(\mathbf{x}'\mathbf{D}_j\mathbf{q}_j)^2}{\mathbf{q}_j'\mathbf{D}_j\mathbf{q}_j} \tag{8.21}$$

Table 8.13 **Summary of Computations for the Davidon-Fletcher-Powell Method**

Iteration k	$\mathbf{x}_k$ $f(\mathbf{x}_k)$	j	$\mathbf{y}_j$ $f(\mathbf{y}_j)$	$\nabla f(\mathbf{y}_j)$	$\lVert\nabla f(\mathbf{y}_j)\rVert$	$\mathbf{D}_j$	$\mathbf{d}_j$	λ_j	$\mathbf{y}_{j+1}$
1	(0.00, 3.00) (52.00)	1	(0.00, 3.00) (52.00)	(−44.00, 24.00)	50.12	$\begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}$	(44.00, −24.00)	0.062	(2.70, 1.51)
		2	(2.70, 1.51) (0.34)	(0.73, 1.28)	1.47	$\begin{bmatrix}0.25 & 0.38\\0.38 & 0.81\end{bmatrix}$	(−0.67, −1.31)	0.22	(2.55, 1.22)
2	(2.55, 1.22) (0.1036)	1	(2.55, 1.22) (0.1036)	(0.89, −0.44)	0.99	$\begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}$	(−0.89, 0.44)	0.11	(2.45, 1.27)
		2	(2.45, 1.27) (0.0490)	(0.18, 0.36)	0.40	$\begin{bmatrix}0.65 & 0.45\\0.45 & 0.46\end{bmatrix}$	(−0.28, −0.25)	0.64	(2.27, 1.11)
3	(2.27, 1.11) (0.008)	1	(2.27, 1.11) (0.008)	(0.18, −0.20)	0.27	$\begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}$	(−0.18, 0.20)	0.10	(2.25, 1.13)
		2	(2.25, 1.13) (0.004)	(0.04, 0.04)	0.06	$\begin{bmatrix}0.80 & 0.38\\0.38 & 0.31\end{bmatrix}$	(−0.05, −0.03)	2.64	(2.12, 1.05)
4	(2.12, 1.05) (0.0005)	1	(2.12, 1.05) (0.0005)	(0.05, −0.08)	0.09	$\begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}$	(−0.05, 0.08)	0.10	(2.115, 1.058)
		2	(2.115, 1.058) (0.0002)	(0.004, 0.004)	0.006				

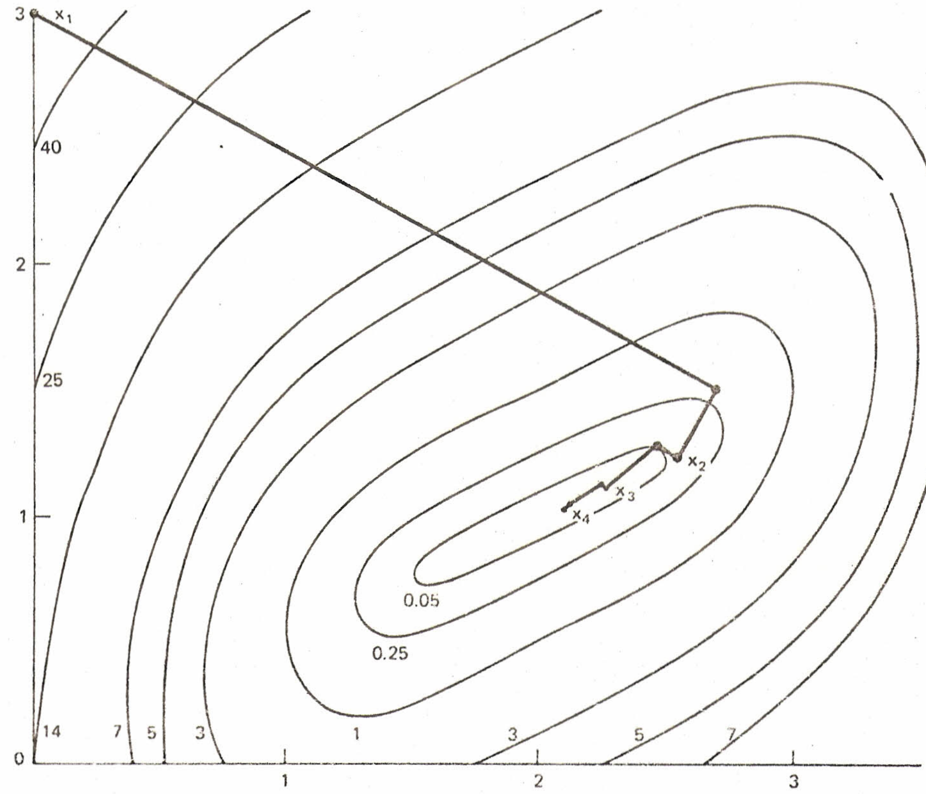

Figure 8.17 Illustration of the Davidon–Fletcher–Powell method.

Since $\mathbf{D}_j$ is a symmetric positive definite matrix, there exists a positive definite symmetric matrix $\mathbf{D}_j^{1/2}$ such that $\mathbf{D}_j = \mathbf{D}_j^{1/2}\mathbf{D}_j^{1/2}$. Let $\mathbf{a} = \mathbf{D}_j^{1/2}\mathbf{x}$ and $\mathbf{b} = \mathbf{D}_j^{1/2}\mathbf{q}_j$. Then $\mathbf{x}'\mathbf{D}_j\mathbf{x} = \mathbf{a}'\mathbf{a}$, $\mathbf{q}_j'\mathbf{D}_j\mathbf{q}_j = \mathbf{b}'\mathbf{b}$, and $\mathbf{x}'\mathbf{D}_j\mathbf{q}_j = \mathbf{a}'\mathbf{b}$. Substituting in (8.21), we get

$$\mathbf{x}'\mathbf{D}_{j+1}\mathbf{x} = \frac{(\mathbf{a}'\mathbf{a})(\mathbf{b}'\mathbf{b}) - (\mathbf{a}'\mathbf{b})^2}{\mathbf{b}'\mathbf{b}} + \frac{(\mathbf{x}'\mathbf{p}_j)^2}{\mathbf{p}_j'\mathbf{q}_j} \tag{8.22}$$

By the Schwartz inequality, $(\mathbf{a}'\mathbf{a})(\mathbf{b}'\mathbf{b}) \geq (\mathbf{a}'\mathbf{b})^2$. Thus, to show that $\mathbf{x}'\mathbf{D}_{j+1}\mathbf{x} \geq 0$, it suffices to show that $\mathbf{p}_j'\mathbf{q}_j > 0$ and that $\mathbf{b}'\mathbf{b} > 0$. From (8.19) and (8.20), it follows that

$$\mathbf{p}_j'\mathbf{q}_j = \lambda_j\mathbf{d}_j'[\nabla f(\mathbf{y}_{j+1}) - \nabla f(\mathbf{y}_j)]$$

The reader may note that $\mathbf{d}_j'\nabla f(\mathbf{y}_{j+1}) = 0$, and by definition, $\mathbf{d}_j = -\mathbf{D}_j\nabla f(\mathbf{y}_j)$. Substituting in the above equation, it follows that

$$\mathbf{p}_j'\mathbf{q}_j = \lambda_j\nabla f(\mathbf{y}_j)'\mathbf{D}_j\nabla f(\mathbf{y}_j) \tag{8.23}$$

Note that $\nabla f(\mathbf{y}_j) \neq \mathbf{0}$ by assumption, and that $\mathbf{D}_j$ is positive definite, so that $\nabla f(\mathbf{y}_j)'\mathbf{D}_j\nabla f(\mathbf{y}_j) > 0$. Furthermore, $\mathbf{d}_j$ is a descent direction and hence $\lambda_j > 0$. Therefore, from (8.23), $\mathbf{p}_j'\mathbf{q}_j > 0$. Furthermore, $\mathbf{q}_j \neq \mathbf{0}$, and hence $\mathbf{b}'\mathbf{b} = \mathbf{q}_j'\mathbf{D}_j\mathbf{q}_j > 0$.

We now show that $\mathbf{x}'\mathbf{D}_{j+1}\mathbf{x} > 0$. By contradiction, suppose that $\mathbf{x}'\mathbf{D}_{j+1}\mathbf{x} = 0$. This is only possible if $(\mathbf{a}'\mathbf{a})(\mathbf{b}'\mathbf{b}) = (\mathbf{a}'\mathbf{b})^2$ and $\mathbf{p}_j'\mathbf{x} = 0$. First, note that $(\mathbf{a}'\mathbf{a})(\mathbf{b}'\mathbf{b}) = (\mathbf{a}'\mathbf{b})^2$ only if $\mathbf{a} = \lambda\mathbf{b}$; that is $\mathbf{D}_j^{1/2}\mathbf{x} = \lambda\mathbf{D}_j^{1/2}\mathbf{q}_j$. Thus $\mathbf{x} = \lambda\mathbf{q}_j$. Since $\mathbf{x} \neq \mathbf{0}$, $\lambda \neq 0$. Now, $0 = \mathbf{p}_j'\mathbf{x} = \lambda\mathbf{p}_j'\mathbf{q}_j$ contradicts the fact that $\mathbf{p}_j'\mathbf{q}_j > 0$ and $\lambda \neq 0$. Therefore $\mathbf{x}'\mathbf{D}_{j+1}\mathbf{x} > 0$, so that $\mathbf{D}_{j+1}$ is positive definite.

Since $\nabla f(\mathbf{y}_{j+1}) \neq \mathbf{0}$ and since $\mathbf{D}_{j+1}$ is positive definite, $\nabla f(\mathbf{y}_{j+1})'\mathbf{d}_{j+1} = -\nabla f(\mathbf{y}_{j+1})'\mathbf{D}_{j+1}\nabla f(\mathbf{y}_{j+1}) < 0$. By Theorem 4.2.1, then, $\mathbf{d}_{j+1}$ is a descent direction.

The Quadratic Case

If the objective function f is quadratic, then by Theorem 8.6.6 below, the directions $\mathbf{d}_1, \ldots, \mathbf{d}_n$ generated by the Davidon-Fletcher-Powell method are conjugate. Therefore, by part 3 of Theorem 8.6.3, the method stops after one complete iteration with an optimal point. Furthermore, the matrix $\mathbf{D}_{n+1}$ obtained at the end of the iteration is precisely the inverse of the Hessian matrix $\mathbf{H}$.

8.6.6 Theorem

Let $\mathbf{H}$ be an $n \times n$ symmetric positive definite matrix, and consider the problem to minimize $f(\mathbf{x}) = \mathbf{c}'\mathbf{x} + \frac{1}{2}\mathbf{x}'\mathbf{H}\mathbf{x}$ subject to $\mathbf{x} \in E_n$. Suppose that the problem is solved by the Davidon-Fletcher-Powell method, starting with an initial point $\mathbf{y}_1$ and a symmetric positive definite matrix $\mathbf{D}_1$. In particular, for $j = 1, \ldots, n$, let λ_j be an optimal solution to the problem to minimize $f(\mathbf{y}_j + \lambda\mathbf{d}_j)$ subject to $\lambda \geq 0$, and let $\mathbf{y}_{j+1} = \mathbf{y}_j + \lambda_j\mathbf{d}_j$, where $\mathbf{d}_j = -\mathbf{D}_j\nabla f(\mathbf{y}_j)$ and $\mathbf{D}_j$ is determined by (8.18) to (8.20). If $\nabla f(\mathbf{y}_j) \neq \mathbf{0}$ for each j, then the directions $\mathbf{d}_1, \ldots, \mathbf{d}_n$ are $\mathbf{H}$-conjugate and $\mathbf{D}_{n+1} = \mathbf{H}^{-1}$. Furthermore, $\mathbf{y}_{n+1}$ is an optimal solution to the problem.

Proof

We first show that for any j with $1 \leq j \leq n$, we must have the following conditions:

1. $\mathbf{d}_1, \ldots, \mathbf{d}_j$ are linearly independent.
2. $\mathbf{d}_i'\mathbf{H}\mathbf{d}_k = 0$ for $i \neq k$; $i, k \leq j$.
3. $\mathbf{D}_{j+1}\mathbf{H}\mathbf{p}_k = \mathbf{p}_k$, or equivalently, $\mathbf{D}_{j+1}\mathbf{H}\mathbf{d}_k = \mathbf{d}_k$ for $1 \leq k \leq j$.

where $\mathbf{p}_k = \lambda_k\mathbf{d}_k$. We prove this result by induction. For $j = 1$, parts 1 and 2 are obvious. To prove part 3, first note that for any k, we have

$$\mathbf{H}\mathbf{p}_k = \mathbf{H}(\lambda_k\mathbf{d}_k) = \mathbf{H}(\mathbf{y}_{k+1} - \mathbf{y}_k) = \nabla f(\mathbf{y}_{k+1}) - \nabla f(\mathbf{y}_k) = \mathbf{q}_k \tag{8.24}$$

In particular, $\mathbf{H}\mathbf{p}_1 = \mathbf{q}_1$. Thus, letting $j = 1$ in (8.18), we get

$$\mathbf{D}_2\mathbf{H}\mathbf{p}_1 = \left[\mathbf{D}_1 + \frac{\mathbf{p}_1\mathbf{p}_1'}{\mathbf{p}_1'\mathbf{q}_1} - \frac{\mathbf{D}_1\mathbf{q}_1\mathbf{q}_1'\mathbf{D}_1}{\mathbf{q}_1'\mathbf{D}_1\mathbf{q}_1}\right]\mathbf{q}_1 = \mathbf{p}_1$$

so that part 3 holds true for $j = 1$.

Now suppose that parts 1, 2, and 3 hold for $j \leq n - 1$. To show that they also hold true for $j + 1$, first recall by part 1 of Theroem 8.6.3 that $\mathbf{d}_i'\nabla f(\mathbf{y}_{j+1}) = 0$ for $i \leq j$. By the induction hypothesis of part 3, $\mathbf{d}_i = \mathbf{D}_{j+1}\mathbf{H}\mathbf{d}_i$ for $i \leq j$. Thus, for $i \leq j$, we have

$$0 = \mathbf{d}_i'\nabla f(\mathbf{y}_{j+1}) = \mathbf{d}_i'\mathbf{H}\mathbf{D}_{j+1}\nabla f(\mathbf{y}_{j+1}) = -\mathbf{d}_i'\mathbf{H}\mathbf{d}_{j+1}$$

In view of the induction hypothesis on part 2, the above equation shows that part 2 also holds for $j + 1$.

Now we show that part 3 holds for $j + 1$. Letting $k \leq j + 1$,

$$\mathbf{D}_{j+2}\mathbf{H}\mathbf{p}_k = \left[\mathbf{D}_{j+1} + \frac{\mathbf{p}_{j+1}\mathbf{p}_{j+1}'}{\mathbf{p}_{j+1}'\mathbf{q}_{j+1}} - \frac{\mathbf{D}_{j+1}\mathbf{q}_{j+1}\mathbf{q}_{j+1}'\mathbf{D}_{j+1}}{\mathbf{q}_{j+1}'\mathbf{D}_{j+1}\mathbf{q}_{j+1}}\right]\mathbf{H}\mathbf{p}_k \tag{8.25}$$

Noting (8.24) and letting $k = j + 1$ in (8.25), it follows that $\mathbf{D}_{j+2}\mathbf{H}\mathbf{p}_{j+1} = \mathbf{p}_{j+1}$ Now let $k \leq j$. Since part 2 holds for $j + 1$,

$$\mathbf{p}_{j+1}'\mathbf{H}\mathbf{p}_k = \lambda_k\lambda_{j+1}\mathbf{d}_{j+1}'\mathbf{H}\mathbf{d}_k = 0 \tag{8.26}$$

Noting the induction hypothesis on part 3, (8.24), and the fact that part 2 holds true for $j + 1$, we get

$$\mathbf{q}_{j+1}'\mathbf{D}_{j+1}\mathbf{H}\mathbf{p}_k = \mathbf{q}_{j+1}'\mathbf{p}_k = \mathbf{p}_{j+1}'\mathbf{H}\mathbf{p}_k = \lambda_{j+1}\lambda_k\mathbf{d}_{j+1}'\mathbf{H}\mathbf{d}_k = 0 \tag{8.27}$$

Substituting (8.26) and (8.27) in (8.25), and noting the induction hypothesis on part 3, we get

$$\mathbf{D}_{j+2}\mathbf{H}\mathbf{p}_k = \mathbf{D}_{j+1}\mathbf{H}\mathbf{p}_k = \mathbf{p}_k$$

Thus part 3 holds for $j + 1$.

To finish the induction argument, we only need to show that part 1 holds true for $j + 1$. Suppose that $\sum_{i=1}^{j+1} \alpha_i\mathbf{d}_i = \mathbf{0}$. Multiplying by $\mathbf{d}_{j+1}'\mathbf{H}$ and noting that part 2 holds for $j + 1$, it follows that $\alpha_{j+1}\mathbf{d}_{j+1}'\mathbf{H}\mathbf{d}_{j+1} = 0$. By assumption $\nabla f(\mathbf{y}_{j+1}) \neq \mathbf{0}$ and by Lemma 8.6.5 $\mathbf{D}_{j+1}$ is positive definite, so that $\mathbf{d}_{j+1} = -\mathbf{D}_{j+1}\nabla f(\mathbf{y}_{j+1}) \neq \mathbf{0}$. Since $\mathbf{H}$ is positive definite, then $\mathbf{d}_{j+1}'\mathbf{H}\mathbf{d}_{j+1} \neq 0$, and hence $\alpha_{j+1} = 0$. This in turn implies that $\sum_{i=1}^{j} \alpha_i\mathbf{d}_i = \mathbf{0}$, and since $\mathbf{d}_1, \ldots, \mathbf{d}_j$ are linearly independent by the induction hypothesis, $\alpha_i = 0$ for $i = 1, \ldots, j$. Thus $\mathbf{d}_1, \ldots, \mathbf{d}_{j+1}$ are linearly independent and part 1 holds for $j + 1$. Thus parts 1, 2, and 3 hold. In particular, conjugacy of $\mathbf{d}_1, \ldots, \mathbf{d}_n$ follow from parts 1 and 2 by letting $j = n$.

Now let $j = n$ in part 3. Then $\mathbf{D}_{n+1}\mathbf{H}\mathbf{d}_k = \mathbf{d}_k$ for $k = 1, \ldots, n$. If we let $\mathbf{D}$ be the matrix whose columns are $\mathbf{d}_1, \ldots, \mathbf{d}_n$, then $\mathbf{D}_{n+1}\mathbf{H}\mathbf{D} = \mathbf{D}$. Since $\mathbf{D}$ is invertible, then $\mathbf{D}_{n+1}\mathbf{H}=\mathbf{I}$, which is only possible if $\mathbf{D}_{n+1}=\mathbf{H}^{-1}$. Finally, $\mathbf{y}_{n+1}$ is an optimal solution by Theorem 8.6.3.

The Conjugate Gradient Method of Fletcher and Reeves

The conjugate gradient method, credited to Fletcher and Reeves [1964], deflects the direction of steepest descent by adding to it a positive multiple of the direction used in the last step. For the quadratic case, as we will learn, deflecting the steepest descent direction in this fashion produces a set of conjugate directions.

Summary of the Conjugate Gradient Method

A summary of the conjugate gradient method for minimizing a differentiable function is summarized below.

Initialization Step Choose a termination scalar $\varepsilon>0$ and an initial point $\mathbf{x}_1$. Let $\mathbf{y}_1=\mathbf{x}_1$, $\mathbf{d}_1=-\nabla f(\mathbf{y}_1)$, $k=j=1$, and go to the main step.

Main Step

1. If $\|\nabla f(\mathbf{y}_j)\|<\varepsilon$, stop. Otherwise, let λ_j be an optimal solution to the problem to minimize $f(\mathbf{y}_j+\lambda\mathbf{d}_j)$ subject to $\lambda\geq 0$, and let $\mathbf{y}_{j+1}=\mathbf{y}_j+\lambda_j\mathbf{d}_j$. If $j<n$, go to step 2; otherwise, go to step 3.
2. Let $\mathbf{d}_{j+1}=-\nabla f(\mathbf{y}_{j+1})+\alpha_j\mathbf{d}_j$, where $\alpha_j=\dfrac{\|\nabla f(\mathbf{y}_{j+1})\|^2}{\|\nabla f(\mathbf{y}_j)\|^2}$. Replace j by $j+1$, and go to step 1.
3. Let $\mathbf{y}_1=\mathbf{x}_{k+1}=\mathbf{y}_{n+1}$, and let $\mathbf{d}_1=-\nabla f(\mathbf{y}_1)$. Let $j=1$, replace k by $k+1$, and go to step 1.

Note that $\mathbf{d}_{j+1}=(1/\mu_1)(\mu_1[-\nabla f(\mathbf{y}_{j+1})]+\mu_2\mathbf{d}_j)$, where

$$\mu_1=\frac{\|\nabla f(\mathbf{y}_j)\|^2}{\|\nabla f(\mathbf{y}_j)\|^2+\|\nabla f(\mathbf{y}_{j+1})\|^2}\quad\text{and}\quad\mu_2=\frac{\|\nabla f(\mathbf{y}_{j+1})\|^2}{\|\nabla f(\mathbf{y}_j)\|^2+\|\nabla f(\mathbf{y}_{j+1})\|^2}$$

so that $\mathbf{d}_{j+1}$ is essentially a convex combination of the current steepest descent direction and the direction used at the last iteration.

8.6.7 Example

Consider the following problem:

$$\text{Minimze}\quad (x_1-2)^4+(x_1-2x_2)^2$$

The summary of the computations using the method of Fletcher and Reeves is given in Table 8.14. At each iteration $\mathbf{d}_1$ was given by $-\nabla f(\mathbf{y}_1)$, and $\mathbf{d}_2$ was

TABLE 8.14 Summary of Computations for the Method of Fletcher and Reeves

Iteration k	$\mathbf{x}_k$ $f(\mathbf{x}_k)$	j	$\mathbf{y}_j$ $f(\mathbf{y}_j)$	$\nabla f(\mathbf{y}_j)$	$\|\nabla f(\mathbf{y}_j)\|$	α_1	$\mathbf{d}_j$	λ_j	$\mathbf{y}_{j+1}$
1	(0.00, 3.00) 52.00	1	(0.00, 3.00) 52.00	(−44.00, 24.00)	50.12	—	(44.00, −24.00)	0.062	(2.70, 1.51)
		2	(2.70, 1.51) 0.34	(0.73, 1.28)	1.47	0.0009	(−0.69, −1.30)	0.23	(2.54, 1.21)
2	(2.54, 1.21) 0.10	1	(2.54, 1.21) 0.10	(0.87, −0.48)	0.99	—	(−0.87, 0.48)	0.11	(2.44, 1.26)
		2	(2.44, 1.26) 0.04	(0.18, 0.32)	0.37	0.14	(−0.30, −0.25)	0.63	(2.25, 1.10)
3	(2.25, 1.10) 0.008	1	(2.25, 1.10) 0.008	(0.16, −0.20)	0.26	—	(−0.16, 0.20)	0.10	(2.23, 1.12)
		2	(2.23, 1.12) 0.003	(0.03, 0.04)	0.05	0.04	(−0.036, −0.032)	1.02	(2.19, 1.09)
4	(2.19, 1.09) 0.0017	1	(2.19, 1.09) 0.0017	(0.05, −0.04)	0.06	—	(−0.05, 0.04)	0.11	(2.185, 1.094)
		2	(2.185, 1.094) 0.0012	(0.02, 0.01)	0.02				

given by $\mathbf{d}_2 = -\nabla f(\mathbf{y}_2) + \alpha_1 \mathbf{d}_1$, where $\alpha_1 = \|\nabla f(\mathbf{y}_2)\|^2/\|\nabla f(\mathbf{y}_1)\|^2$. Furthermore, $\mathbf{y}_{j+1}$ is obtained by optimizing along $\mathbf{d}_j$, starting from $\mathbf{y}_j$. At iteration 4, the point $\mathbf{y}_2 = (2.185, 1.094)'$, which is very close to the optimal point $(2.00, 1.00)$, is reached. Since the norm of the gradient is equal to 0.02, which is small, we stop here. The progress of the algorithm is shown in Figure 8.18.

The Quadratic Case

If the function f is quadratic, Theorem 8.6.8 below shows that the directions $\mathbf{d}_1, \ldots, \mathbf{d}_n$ generated are indeed conjugate, and hence by Theorem 8.6.3, the conjugate gradient algorithm produces an optimal solution after one complete application of the main step, that is, after, at most, n line searches have been performed.

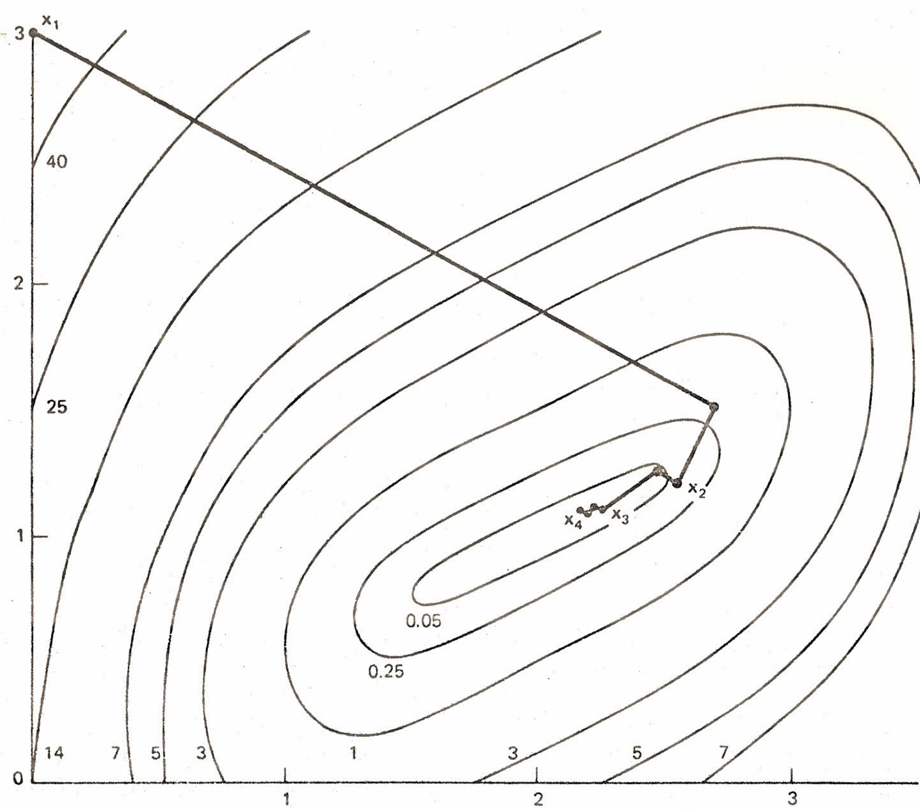

Figure 8.18 Illustration of the method of Fletcher and Reeves.

Theorem 8.6.8

Consider the problem to minimize $f(\mathbf{x}) = \mathbf{c}'\mathbf{x} + \frac{1}{2}\mathbf{x}'\mathbf{H}\mathbf{x}$ subject to $\mathbf{x} \in E_n$. Suppose that the problem is solved by the conjugate gradient method, starting with $\mathbf{y}_1$ and $\mathbf{d}_1 = -\nabla f(\mathbf{y}_1)$. In particular, for $j = 1, \ldots, n$, let λ_j be an optimal solution to the problem to minimize $f(\mathbf{y}_j + \lambda \mathbf{d}_j)$ subject to $\lambda \geq 0$. Let $\mathbf{y}_{j+1} = \mathbf{y}_j + \lambda_j \mathbf{d}_j$, and let $\mathbf{d}_{j+1} = -\nabla f(\mathbf{y}_{j+1}) + \alpha_j \mathbf{d}_j$, where $\alpha_j = \|\nabla f(\mathbf{y}_{j+1})\|^2/\|\nabla f(\mathbf{y}_j)\|^2$. If $\nabla f(\mathbf{y}_j) \neq \mathbf{0}$ for $j = 1, \ldots, n$, then the following statements are true:

1. $\mathbf{d}_1, \ldots, \mathbf{d}_n$ are $\mathbf{H}$-conjugate.
2. $\mathbf{d}_1, \ldots, \mathbf{d}_n$ are descent directions.
3. $\alpha_j = \dfrac{\|\nabla f(\mathbf{y}_{j+1})\|^2}{\|\nabla f(\mathbf{y}_j)\|^2} = \dfrac{\mathbf{d}_j'\mathbf{H}\nabla f(\mathbf{y}_{j+1})}{\mathbf{d}_j'\mathbf{H}\mathbf{d}_j}$ for $j = 1, \ldots, n$

Proof

First suppose that parts 1, 2, and 3 hold for j. We show that they also hold for $j+1$. To show that part 1 holds for $j+1$, we first demonstrate that $\mathbf{d}_k'\mathbf{H}\mathbf{d}_{j+1} = 0$ for $k \leq j$. Since $\mathbf{d}_{j+1} = -\nabla f(\mathbf{y}_{j+1}) + \alpha_j \mathbf{d}_j$, noting the induction hypothesis in part 3, and letting $k = j$, we get

$$\mathbf{d}_j'\mathbf{H}\mathbf{d}_{j+1} = \mathbf{d}_j'\mathbf{H}\left[-\nabla f(\mathbf{y}_{j+1}) + \frac{\mathbf{d}_j'\mathbf{H}\nabla f(\mathbf{y}_{j+1})}{\mathbf{d}_j'\mathbf{H}\mathbf{d}_j}\mathbf{d}_j\right] = 0 \tag{8.28}$$

Now let $k < j$. Since $\mathbf{d}_{j+1} = -\nabla f(\mathbf{y}_{j+1}) + \alpha_j \mathbf{d}_j$, and sicne $\mathbf{d}_k'\mathbf{H}\mathbf{d}_j = 0$ by the induction hypothesis of part 1,

$$\mathbf{d}_k'\mathbf{H}\mathbf{d}_{j+1} = -\mathbf{d}_k'\mathbf{H}\nabla f(\mathbf{y}_{j+1}) \tag{8.29}$$

Since $\nabla f(\mathbf{y}_{k+1}) = \mathbf{c} + \mathbf{H}\mathbf{y}_{k+1}$ and $\mathbf{y}_{k+1} = \mathbf{y}_k + \lambda_k \mathbf{d}_k$, note that

$$\begin{aligned}\mathbf{d}_{k+1} &= -\nabla f(\mathbf{y}_{k+1}) + \alpha_k \mathbf{d}_k \\ &= -[\nabla f(\mathbf{y}_k) + \lambda_k \mathbf{H}\mathbf{d}_k] + \alpha_k \mathbf{d}_k \\ &= -[-\mathbf{d}_k + \alpha_{k-1}\mathbf{d}_{k-1} + \lambda_k \mathbf{H}\mathbf{d}_k] + \alpha_k \mathbf{d}_k\end{aligned}$$

By the induction hypothesis of part 2, $\mathbf{d}_k$ is a descent direction, and hence $\lambda_k > 0$. Therefore

$$\mathbf{d}_k'\mathbf{H} = \frac{1}{\lambda_k}\left[-\mathbf{d}_{k+1}' + (1 + \alpha_k)\mathbf{d}_k' - \alpha_{k-1}\mathbf{d}_{k-1}'\right] \tag{8.30}$$

From (8.29) and (8.30), it follows that

$$\begin{aligned}\mathbf{d}_k'\mathbf{H}\mathbf{d}_{j+1} &= -\mathbf{d}_k'\mathbf{H}\nabla f(\mathbf{y}_{j+1}) \\ &= -\frac{1}{\lambda_k}[-\mathbf{d}_{k+1}'\nabla f(\mathbf{y}_{j+1}) + (1 + \alpha_k)\mathbf{d}_k'\nabla f(\mathbf{y}_{j+1}) - \alpha_{k-1}\mathbf{d}_{k-1}'\nabla f(\mathbf{y}_{j+1})]\end{aligned}$$

By part 1 of Theorem 8.6.3, and since $\mathbf{d}_1, \ldots, \mathbf{d}_j$ are assumed conjugate, then $\mathbf{d}_{k+1}^t \nabla f(\mathbf{y}_{j+1}) = \mathbf{d}_k^t \nabla f(\mathbf{y}_{j+1}) = \mathbf{d}_{k-1}^t \nabla f(\mathbf{y}_{j+1}) = 0$. Thus the above equation implies that $\mathbf{d}_k^t \mathbf{H} \mathbf{d}_{j+1} = 0$ for $k < j$. This, together with (8.28), shows that $\mathbf{d}_k^t \mathbf{H} \mathbf{d}_{j+1} = 0$ for all $k \leq j$.

In order to show that $\mathbf{d}_1, \ldots, \mathbf{d}_{j+1}$ are $\mathbf{H}$-conjugate, it thus sufficies to show that they are linearly independent. Suppose that $\sum_{i=1}^{j+1} \gamma_i \mathbf{d}_i = \mathbf{0}$. Then, $\sum_{i=1}^{j} \gamma_i \mathbf{d}_i + \gamma_{j+1}[-\nabla f(\mathbf{y}_{j+1}) + \alpha_j \mathbf{d}_j] = \mathbf{0}$. Multiplying by $\nabla f(\mathbf{y}_{j+1})^t$, and noting part 1 of Theorem 8.6.3, it follows that $\gamma_{j+1} \|\nabla f(\mathbf{y}_{j+1})\|^2 = 0$. Since $\nabla f(\mathbf{y}_{j+1}) \neq \mathbf{0}$, $\gamma_{j+1} = 0$. This implies that $\sum_{i=1}^{j} \gamma_i \mathbf{d}_i = \mathbf{0}$, and in view of conjugacy of $\mathbf{d}_1, \ldots, \mathbf{d}_j$, it follows that $\gamma_1 = \cdots = \gamma_j = 0$. Thus, $\mathbf{d}_1, \ldots, \mathbf{d}_{j+1}$ are linearly independent and $\mathbf{H}$-conjugate, so that part 1 holds true for $j+1$.

Now we show that part 2 holds for $j+1$; that is, $\mathbf{d}_{j+1}$ is a descent direction. Note that $\nabla f(\mathbf{y}_{j+1}) \neq \mathbf{0}$ by assumption, and that $\nabla f(\mathbf{y}_{j+1})^t \mathbf{d}_j = 0$ by part 1 of Theorem 8.6.3. Then $\nabla f(\mathbf{y}_{j+1})^t \mathbf{d}_{j+1} = -\|\nabla f(\mathbf{y}_{j+1})\|^2 + \alpha_j \nabla f(\mathbf{y}_{j+1})^t \mathbf{d}_j = -\|\nabla f(\mathbf{y}_{j+1})\|^2 < 0$. By Theorem 4.1.2, $\mathbf{d}_{j+1}$ is a descent direction.

Now we show that part 3 holds for $j+1$. By letting $k = j+1$ in (8.30), and multiplying by $\nabla f(\mathbf{y}_{j+2})$, it follows that

$$\begin{aligned}\lambda_{j+1} \mathbf{d}_{j+1}^t \mathbf{H} \nabla f(\mathbf{y}_{j+2}) &= [-\mathbf{d}_{j+2}^t + (1+\alpha_{j+1})\mathbf{d}_{j+1}^t - \alpha_j \mathbf{d}_j^t] \nabla f(\mathbf{y}_{j+2}) \\ &= [\nabla f(\mathbf{y}_{j+2})^t + \mathbf{d}_{j+1}^t - \alpha_j \mathbf{d}_j^t] \nabla f(\mathbf{y}_{j+2})\end{aligned}$$

Since $\mathbf{d}_1, \ldots, \mathbf{d}_{j+1}$ are $\mathbf{H}$-conjugate, then by part 1 of Theorem 8.6.3, $\mathbf{d}_{j+1}^t \nabla f(\mathbf{y}_{j+2}) = \mathbf{d}_j^t \nabla f(\mathbf{y}_{j+2}) = 0$. The above equation then implies that

$$\|\nabla f(\mathbf{y}_{j+2})\|^2 = \lambda_{j+1} \mathbf{d}_{j+1}^t \mathbf{H} \nabla f(\mathbf{y}_{j+2}) \tag{8.31}$$

Multiplying $\nabla f(\mathbf{y}_{j+1}) = \nabla f(\mathbf{y}_{j+2}) - \lambda_{j+1} \mathbf{H} \mathbf{d}_{j+1}$ by $\nabla f(\mathbf{y}_{j+1})^t$, and noting that $\mathbf{d}_j^t \mathbf{H} \mathbf{d}_{j+1} = \mathbf{d}_{j+1}^t \nabla f(\mathbf{y}_{j+2}) = \mathbf{d}_j^t \nabla f(\mathbf{y}_{j+2}) = 0$, we get

$$\begin{aligned}\|\nabla f(\mathbf{y}_{j+1})\|^2 &= \nabla f(\mathbf{y}_{j+1})^t [\nabla f(\mathbf{y}_{j+2}) - \lambda_{j+1} \mathbf{H} \mathbf{d}_{j+1}] \\ &= (-\mathbf{d}_{j+1}^t + \alpha_j \mathbf{d}_j^t)[\nabla f(\mathbf{y}_{j+2}) - \lambda_{j+1} \mathbf{H} \mathbf{d}_{j+1}] \\ &= \lambda_{j+1} \mathbf{d}_{j+1}^t \mathbf{H} \mathbf{d}_{j+1}\end{aligned} \tag{8.32}$$

From (8.31) and (8.32), it is obvious that part 3 holds true for $j+1$.

We have thus shown that if parts 1, 2, and 3 hold for j, then they also hold for $j+1$. Note that parts 1 and 2 trivially hold for $j = 1$. In addition, using a similar argument to that used in proving that part 3 holds for $j+1$, it can be easily demonstrated that it holds for $j = 1$. This completes the proof.

The Method of Zangwill

We now discuss a method credited to Zangwill [1967] for minimizing a function of several variables. At step j, suppose that the directions $\mathbf{d}_1, \ldots, \mathbf{d}_j$ are available. Zangwill's method generates a new direction $\mathbf{d}_{j+1}$ as follows. Let

$\mathbf{y}_1$ and $\mathbf{z}_1$ be two points such that $\mathbf{z}_1 - \mathbf{y}_1 \notin L(\mathbf{d}_1, \ldots, \mathbf{d}_j)$, where $L(\mathbf{d}_1, \ldots, \mathbf{d}_j)$ is the linear subspace spanned by $\mathbf{d}_1, \ldots, \mathbf{d}_j$. Let $\mathbf{y}_{j+1}$ and $\mathbf{z}_{j+1}$ be produced by minimizing f sequenctially along $\mathbf{d}_1, \ldots, \mathbf{d}_j$ starting from $\mathbf{y}_1$ and $\mathbf{z}_1$, respectively. Then $\mathbf{d}_{j+1}$ is given by

$$\mathbf{d}_{j+1} = \mathbf{z}_{j+1} - \mathbf{y}_{j+1}$$

Summary of the Method of Zangwill

A summary of Zangwill's method for minimizing a function f of several variables is given below. As will be shown later, if the function f is differentiable, then the method converges to a point with zero gradient.

Initialization Step Choose a termination scalar $\varepsilon > 0$, and choose an initial point $\mathbf{x}_1$. Let $\mathbf{y}_1 = \mathbf{x}_1$, let $\mathbf{d}_1 = -\nabla f(\mathbf{y}_1)$, let $k = j = 1$, and go to the main step.

Main Step

1. Let λ_j be an optimal solution to the problem to minimize $f(\mathbf{y}_j + \lambda \mathbf{d}_j)$ subject to $\lambda \in E_1$, and let $\mathbf{y}_{j+1} = \mathbf{y}_j + \lambda_j \mathbf{d}_j$. If $j = n$, go to step 4; otherwise, go to step 2.
2. Let $\mathbf{d} = -\nabla f(\mathbf{y}_{j+1})$, and let $\hat{\mu}$ be an optimal solution to the problem to minimize $f(\mathbf{y}_{j+1} + \mu \mathbf{d})$ subject to $\mu \geq 0$. Let $\mathbf{z}_1 = \mathbf{y}_{j+1} + \hat{\mu} \mathbf{d}$. Let $i = 1$, and go to step 3.
3. If $\|\nabla f(\mathbf{z}_i)\| < \varepsilon$, stop with $\mathbf{z}_i$. Otherwise, let μ_i be an optimal solution to the problem to minimize $f(\mathbf{z}_i + \mu \mathbf{d}_i)$ subject to $\mu \in E_1$. Let $\mathbf{z}_{i+1} = \mathbf{z}_i + \mu_i \mathbf{d}_i$. If $i < j$, replace i by $i+1$, and repeat step 3. Otherwise, let $\mathbf{d}_{j+1} = \mathbf{z}_{j+1} - \mathbf{y}_{j+1}$, replace j by $j+1$, and go to step 1.
4. Let $\mathbf{y}_1 = \mathbf{x}_{k+1} = \mathbf{y}_{n+1}$. Let $\mathbf{d}_1 = -\nabla f(\mathbf{y}_1)$, replace k by $k+1$, let $j = 1$, and go to step 1.

Note that the steepest descent search in step 2 is used to ensure that $\mathbf{z}_1 - \mathbf{y}_1 \notin L(\mathbf{d}_1, \ldots, \mathbf{d}_j)$ for the quadratic case so that finite convergence is guaranteed.

8.6.9 Example

Consider the following problem:

$$\text{Minimize} \quad (x_1 - 2)^4 + (x_1 - 2x_2)^2$$

We solve this problem using Zangwill's method, starting from the point $(0.0, 3.0)$. The results are summarized in Table 8.15. Note that during each iteration, $\mathbf{y}_2$ is obtained from $\mathbf{y}_1$ by optimizing along $\mathbf{d}_1 = -\nabla f(\mathbf{y}_1)$, and $\mathbf{z}_1$ is obtained from $\mathbf{y}_2$ by optimizing along $\mathbf{d} = -\nabla f(\mathbf{y}_2)$. To obtain $\mathbf{z}_2$ from $\mathbf{z}_1$ we optimize along $\mathbf{d}_1$, and to obtain $\mathbf{y}_3$ from $\mathbf{y}_2$ we optimize along $\mathbf{d}_2 = (\mathbf{z}_2 - \mathbf{y}_2)$.

TABLE 8.15 **Summary of Computations for the Method of Zangwill**

Iteration $k=1$	$x_1=(0.00, 3.00)^t$		$f(x_1)=52.00$						
j	y_j	d_j	λ_j	y_{j+1}	d	$\hat{\mu}$	z_1 $f(z_1)$	μ_1	z_2 $f(z_2)$
1	(0.00, 3.00)	(44.00, −24.00)	0.062	(2.70, 1.51)	(−0.73, −1.28)	0.25	(2.52, 1.20) 0.090	−0.0013	(2.46, 1.23) 0.045
2	(2.70, 1.51)	(−0.24, −0.28)	1.50	(2.34, 1.09)	—	—	—	—	—
Iteration $k=2$	$x_2=(2.34, 1.09)^t$		$f(x_2)=0.039$						
j	y_j	d_j	λ_j	y_{j+1}	d	$\hat{\mu}$	z_1 $f(z_1)$	μ_1	z_2 $f(z_2)$
1	(2.34, 1.09)	(−0.48, 0.64)	0.10	(2.29, 1.15)	(−0.08, −0.04)	3.60	(2.00, 1.01) 0.004	—	—

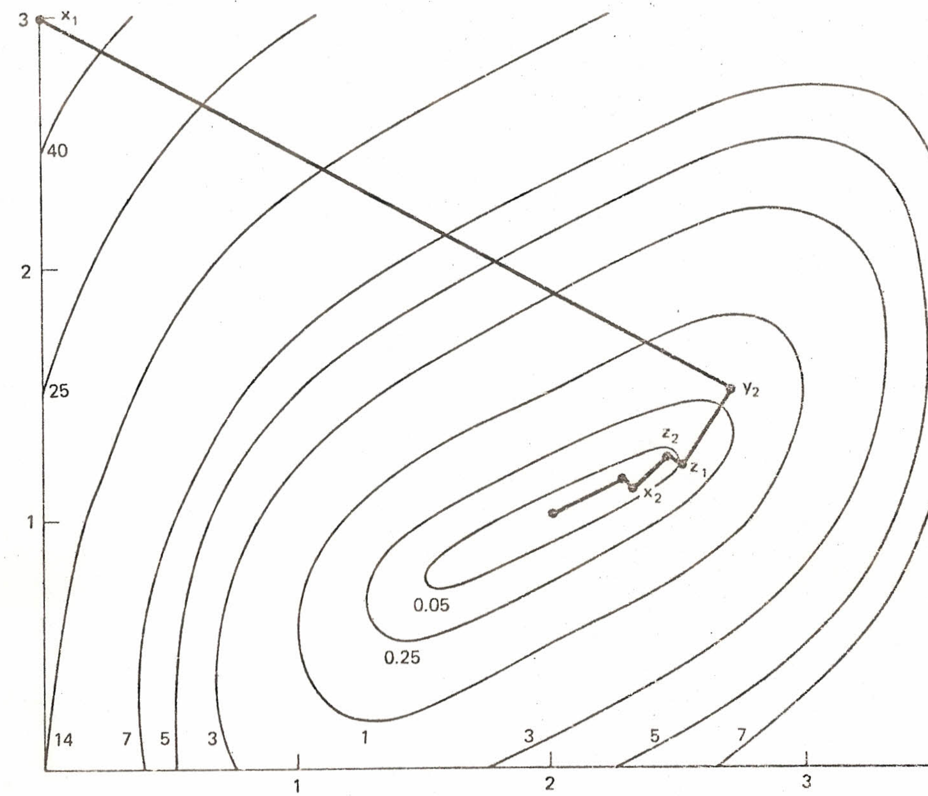

Figure 8.19 Illustration of the method of Zangwill.

The algorithm is terminated during the second iteration when the point $\mathbf{z}_1 = (2.00, 1.01)^t$ is reached, since $\|\nabla f(\mathbf{z}_1)\| = 0.09$ is small. The path taken by the algorithm is shown in Figure 8.19.

The Quadratic Case

If the function f is quadratic, then Zangwill's method generates conjugate directions and from Theorem 8.6.3 will yield an optimal solution in one iteration of the algorithm. The process of generating a new direction is illustrated in Figure 8.20. Given $\mathbf{d}_1$, $\mathbf{y}_2$ and $\mathbf{z}_2$ are produced by minimizing f along $\mathbf{d}_1$ starting from $\mathbf{y}_1$ and $\mathbf{z}_1$, respectively, where $\mathbf{y}_1 - \mathbf{z}_1 \neq \lambda \mathbf{d}_1$ for any $\lambda \in E_1$. Letting $\mathbf{d}_2 = \mathbf{z}_2 - \mathbf{y}_2$, note that $\mathbf{d}_1$ and $\mathbf{d}_2$ are conjugate. In particular, as expected from Theorem 8.6.3, minimizing f starting from either $\mathbf{y}_2$ or $\mathbf{z}_2$ along $\mathbf{d}_2$, leads to the optimal point $\bar{\mathbf{x}}$. Note that if $\mathbf{y}_1 - \mathbf{z}_1 = \lambda \mathbf{d}_1$ for some λ, that is, if $\mathbf{z}_1$ is on the line passing through $\mathbf{y}_1$ along the direction $\mathbf{d}_1$, then $\mathbf{y}_2 = \mathbf{z}_2$ and $\mathbf{d}_2 = \mathbf{0}$. Then $\mathbf{d}_1$ and

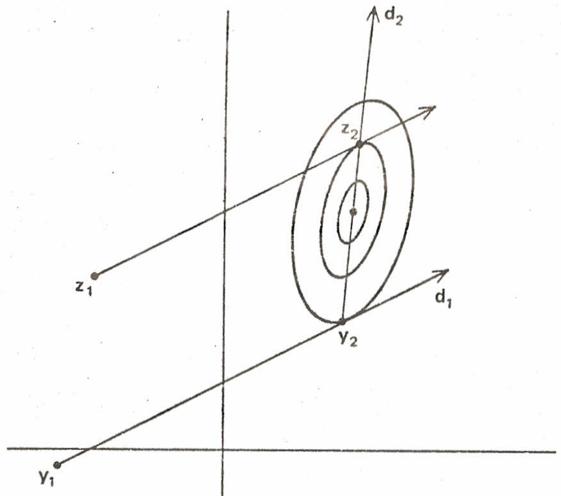

Figure 8.20 Generating conjugate directions by the method of Zangwill.

$\mathbf{d}_2$ are linearly dependent and hence not conjugate. Lemma 8.6.10 below shows that the above method for generating new directions produces a set of conjugate directions.

8.6.10 Lemma

Let $\mathbf{H}$ be an $n \times n$ symmetric matrix, and let $f(\mathbf{x}) = \mathbf{c}^t\mathbf{x} + \frac{1}{2}\mathbf{x}^t\mathbf{H}\mathbf{x}$. Suppose that $j < n$, and let $\mathbf{d}_1, \ldots, \mathbf{d}_j$ be $\mathbf{H}$-conjugate. Let $\mathbf{y}_1$ be an arbitrary vector in E_n, and let $\mathbf{z}_1$ be such that $\mathbf{z}_1 - \mathbf{y}_1 \notin L(\mathbf{d}_1, \ldots, \mathbf{d}_j)$. For $i = 1, \ldots, j$, let λ_i and μ_i be optimal solutions to the problems to minimize $f(\mathbf{y}_i + \lambda \mathbf{d}_i)$ subject to $\lambda \in E_1$ and to minimize $f(\mathbf{z}_i + \mu \mathbf{d}_i)$ subject to $\mu \in E_1$, and let $\mathbf{y}_{i+1} = \mathbf{y}_i + \lambda_i \mathbf{d}_i$ and $\mathbf{z}_{i+1} = \mathbf{z}_i + \mu_i \mathbf{d}_i$. Now let $\mathbf{d}_{j+1} = \mathbf{z}_{j+1} - \mathbf{y}_{j+1}$. Then $\mathbf{d}_1, \ldots, \mathbf{d}_{j+1}$ are $\mathbf{H}$-conjugate.

Proof

Since $\mathbf{d}_1, \ldots, \mathbf{d}_j$ are $\mathbf{H}$-conjugate, then by part 1 of Theorem 8.6.3, we must have

$$0 = \mathbf{d}_i^t[\nabla f(\mathbf{z}_{j+1}) - \nabla f(\mathbf{y}_{j+1})] = \mathbf{d}_i^t\mathbf{H}(\mathbf{z}_{j+1} - \mathbf{y}_{j+1}) = \mathbf{d}_i^t\mathbf{H}\mathbf{d}_{j+1} \qquad \text{for } i \le j$$

In view of the above equation, since $\mathbf{d}_1, \ldots, \mathbf{d}_j$ are $\mathbf{H}$-conjugate, it only suffices to demonstrate that $\mathbf{d}_1, \ldots, \mathbf{d}_{j+1}$ are linearly independent. Suppose, by contradiction to the desired conclusion, that $\mathbf{d}_{j+1}$ could be represented as $\sum_{i=1}^{j} \gamma_i \mathbf{d}_i$. Therefore $\mathbf{z}_{j+1} - \mathbf{y}_{j+1} = \sum_{i=1}^{j} \gamma_i \mathbf{d}_i$, and since $\mathbf{y}_{j+1} = \mathbf{y}_1 + \sum_{i=1}^{j} \lambda_i \mathbf{d}_i$ and $\mathbf{z}_{j+1} = \mathbf{z}_1 + \sum_{i=1}^{j} \mu_i \mathbf{d}_i$, it follows that

$$\mathbf{z}_1 - \mathbf{y}_1 = \sum_{i=1}^{j} (\gamma_i - \mu_i + \lambda_i)\mathbf{d}_i \in L(\mathbf{d}_1, \ldots, \mathbf{d}_j)$$

which contradicts our assumption. Therefore, $\mathbf{d}_1, \ldots, \mathbf{d}_{j+1}$ are linearly independent, and the proof is complete.

Convergence of Conjugate Direction Methods

As shown in Theorem 8.6.3, if the function under consideration is quadratic, then any conjugate direction algorithm produces an optimal solution in a finite number of steps. We now discuss convergence of these methods if the function is not necessarily quadratic.

In Theorem 7.3.4, we showed that a composite algorithm $\mathbf{A} = \mathbf{CB}$ converges to a point in the solution set Ω if the following properties hold true:

1. $\mathbf{B}$ is closed at points not in Ω.
2. If $\mathbf{y} \in \mathbf{B}(\mathbf{x})$, then $f(\mathbf{y}) < f(\mathbf{x})$ for $\mathbf{x} \notin \Omega$.
3. If $\mathbf{z} \in \mathbf{C}(\mathbf{y})$, then $f(\mathbf{z}) \le f(\mathbf{y})$.
4. The set $\Lambda = \{\mathbf{x}: f(\mathbf{x}) \le f(\mathbf{x}_1)\}$ is compact, where $\mathbf{x}_1$ is the starting point.

For the conjugate direction algorithms discussed in this chapter, the map $\mathbf{B}$ is of the following form. Given $\mathbf{x}$, then $\mathbf{y} \in \mathbf{B}(\mathbf{x})$ means that $\mathbf{y}$ is obtained by minimizing f starting from $\mathbf{x}$ along the direction $\mathbf{d} = -\mathbf{D}\nabla f(\mathbf{x})$, where $\mathbf{D}$ is a specified positive definite matrix. In particular, for the conjugate gradient method and for the method of Zangwill, $\mathbf{D} = \mathbf{I}$. For the Davidon–Fletcher–Powell method, $\mathbf{D}$ is an arbitrary positive definite matrix. Furthermore, starting from the point obtained by applying the map $\mathbf{B}$, the map $\mathbf{C}$ is defined by minimizing the function f along the directions specified by the particular algorithms. Thus the map $\mathbf{C}$ satisfies property 3 above.

Letting $\Omega = \{\mathbf{x}: \nabla f(\mathbf{x}) = \mathbf{0}\}$, we now show that the map $\mathbf{B}$ satisfies properties 1 and 2 above. Let $\mathbf{x} \notin \Omega$, and let $\mathbf{x}_k \to \mathbf{x}$. Furthermore, let $\mathbf{y}_k \in \mathbf{B}(\mathbf{x}_k)$ and let $\mathbf{y}_k \to \mathbf{y}$. We need to show that $\mathbf{y} \in \mathbf{B}(\mathbf{x})$. By definition of $\mathbf{y}_k$, $\mathbf{y}_k = \mathbf{x}_k - \lambda_k \mathbf{D}\nabla f(\mathbf{x}_k)$ for $\lambda_k \ge 0$ such that

$$f(\mathbf{y}_k) \le f[\mathbf{x}_k - \lambda \mathbf{D}\nabla f(\mathbf{x}_k)] \qquad \text{for all } \lambda \ge 0 \qquad (8.33)$$

Since $\nabla f(\mathbf{x}) \ne \mathbf{0}$, then λ_k converges to $\bar{\lambda} = \|\mathbf{y} - \mathbf{x}\| / \|\mathbf{D}\nabla f(\mathbf{x})\| \ge 0$. Therefore $\mathbf{y} = \mathbf{x} - \bar{\lambda}\mathbf{D}\nabla f(\mathbf{x})$. Taking the limit as $k \to \infty$ in (8.33), then $f(\mathbf{y}) \le f[\mathbf{x} - \lambda \mathbf{D}\nabla f(\mathbf{x})]$ for all $\lambda \ge 0$, so that $\mathbf{y}$ is indeed obtained by minimizing f starting from $\mathbf{x}$ in the direction $-\mathbf{D}\nabla f(\mathbf{x})$. Thus $\mathbf{y} \in \mathbf{B}(\mathbf{x})$, and $\mathbf{B}$ is closed. Also part 2 holds by noting that $-\nabla f(\mathbf{x})^t\mathbf{D}\nabla f(\mathbf{x}) < 0$, so that $-\mathbf{D}\nabla f(\mathbf{x})$ is a descent direction. Assuming that the set defined in part 4 is compact, it then follows that the conjugate direction algorithms discussed in this section converge to a point with zero gradient.

Some algorithms cannot naturally be broken into maps **B** and **C** satisfying the above properties. The difficulty here is the absence of a closed map that guarantees strict improvement at points outside the solution set. In this case, since the map **C** is not necessarily closed, overall convergence is not guaranteed. In order to overcome the difficulty, a *spacer step*, using a map **B** satisfying properties 1 and 2 above, is employed at each iteration. Typically, the spacer step involves minimizing the function along the negative gradient direction.

Exercises

8.1 For the uniform search method, the dichotomous search method, the golden section method, and the Fibonacci search method, compute the number of functional evaluations required for $\alpha = 0.1, 0.01, 0.001$, and 0.0001, where α is the ratio of the final interval of uncertainty to the length of the initial interval of uncertainty.

8.2 Suppose that θ is differentiable and let $|\theta'| \le a$. Furthermore, suppose that the uniform search method is used to minimize θ. Let $\hat{\lambda}$ be a grid point such that $\theta(\bar{\lambda}) - \theta(\hat{\lambda}) \ge \varepsilon > 0$, for each grid point $\bar{\lambda} \ne \hat{\lambda}$. If the grid length δ is such that $a\delta \le \varepsilon$, show, without assuming strict quasiconvexity, that no point outside the interval $[\hat{\lambda} - \delta, \hat{\lambda} + \delta]$ could provide a functional value less than $\theta(\hat{\lambda})$.

8.3 Show that the golden section method approaches the method of Fibonacci as the number of functional evaluations n approaches ∞.

8.4 Consider the following definitions.

A function $\theta: E_1 \to E_1$ to be *minimized* is said to be *strongly unimodal* over the interval $[a, b]$ if there exists a $\bar{\lambda}$ that minimizes θ over the interval, and for $\lambda_1, \lambda_2 \in [a, b]$ such that $\lambda_1 < \lambda_2$ we have:

$$\lambda_2 \le \bar{\lambda} \quad \text{implies } \theta(\lambda_1) > \theta(\lambda_2)$$
$$\lambda_1 \ge \bar{\lambda} \quad \text{implies } \theta(\lambda_1) < \theta(\lambda_2)$$

A function $\theta: E_1 \to E_1$ to be minimized is said to be *unimodal* over the interval $[a, b]$ if there exists a $\bar{\lambda}$ that minimizes θ over the interval and for $\lambda_1, \lambda_2 \in [a, b]$ such that $\theta(\lambda_1) \ne \theta(\bar{\lambda})$, $\theta(\lambda_2) \ne \theta(\bar{\lambda})$, and $\lambda_1 < \lambda_2$ we have

$$\lambda_2 \le \bar{\lambda} \quad \text{implies } \theta(\lambda_1) > \theta(\lambda_2)$$
$$\lambda_1 \ge \bar{\lambda} \quad \text{implies } \theta(\lambda_1) < \theta(\lambda_2)$$

a. Show that if θ is strongly unimodal over $[a, b]$, then θ is strongly quasiconvex over $[a, b]$. Conversely, show that if θ is strongly quasiconvex over $[a, b]$ and has a minimum in this interval, then it is strongly unimodal over the interval.

b. Show that if θ is unimodal and continuous over $[a, b]$, then θ is strictly quasiconvex over $[a, b]$. Conversely, show that if θ is strictly quasiconvex over $[a, b]$ and has a minimum in this interval, then it is unimodal over this interval.

8.5 Consider the function f defined by $f(\mathbf{x}) = (x_1^3 + x_2)^2 + 2(x_2 - x_1 - 4)^4$. Given a point $\mathbf{x}_1$ and a nonzero direction vector $\mathbf{d}$, let $\theta(\lambda) = f(\mathbf{x}_1 + \lambda\mathbf{d})$.

a. Obtain an explicit expression for $\theta(\lambda)$.

b. For $\mathbf{x}_1 = (0, 0)'$ and $\mathbf{d} = (1, 1)'$, using the Fibonacci method, find the value of λ that solves the problem to minimize $\theta(\lambda)$ subject to $\lambda \in E_1$.

c. For $\mathbf{x}_1 = (4, 5)'$ and $\mathbf{d} = (1, -2)'$, using the golden section method, find the value of λ that solves the problem to minimize $\theta(\lambda)$ subject to $\lambda \in E_1$.

d. Repeat parts b and c using the interval bisection method.

8.6 Find the minimum of $e^{-\lambda} + \lambda^2$ by each of the following procedures.

a. Golden section method.

b. Dichotomous search method.

c. Newton's method.

d. Bisecting search method.

8.7 Consider the problem to minimize $f(\mathbf{x}+\lambda\mathbf{d})$ subject to $\lambda \in E_1$. Show that a necessary condition for a minimum at $\bar{\lambda}$ is that $\mathbf{d}'\nabla f(\mathbf{y})=0$, where $\mathbf{y}=\mathbf{x}+\bar{\lambda}\mathbf{d}$. Under what assumptions is this condition sufficient for optimality?

8.8 Consider the problem to minimize $f(\mathbf{x}+\lambda\mathbf{d})$ subject to $\mathbf{x}+\lambda\mathbf{d}\in S$ and $\lambda\geq 0$, where S is a compact convex set, and f is a convex function. Furthermore, suppose that $\mathbf{d}$ is an improving direction. Show that an optimal solution $\bar{\lambda}$ is given by $\bar{\lambda}=\text{minimum } (\lambda_1,\lambda_2)$, where λ_1 satisfies $\mathbf{d}'\nabla f(\mathbf{x}+\lambda_1\mathbf{d})=0$, and $\lambda_2=\text{maximum }\{\lambda:\mathbf{x}+\lambda\mathbf{d}\in S\}$.

8.9 Consider the problem to minimize $3\lambda-2\lambda^2+\lambda^3+2\lambda^4$ subject to $\lambda\geq 0$.

a. Write a necessary condition for a minimum. Can you make use of this condition to find the global minimum?

b. Is the function strictly quasiconvex over the region $\{\lambda:\lambda\geq 0\}$? Apply the Fibonacci search method to find the minimum.

c. Apply both the bisecting search method and Newton's method, to the above problem, starting from $\lambda_1=6$.

8.10 In Section 8.2 we described Newton's method for finding a point where the derivative of a function vanishes.

a. Show how the method can be used to find a point where the value of a continuously differentiable function is equal to zero. Illustrate the method for $\theta(\lambda)=\lambda^3-\lambda$ starting from $\lambda_1=5$.

b. Will the method converge for any starting point? Prove or give a counterexample.

8.11 Show how the line search procedures of Section 8.1 can be used to find a point where a given function assumes the value zero. Illustrate by the function θ defined by $\theta(\lambda)=\lambda^2-3\lambda+2$.
(*Hint:* Consider the absolute value function $\hat{\theta}=|\theta|$).

8.12 In Section 8.2 we discussed the bisecting search method for finding a point where the derivative of a pseudoconvex function vanishes. Show how the method can be used to find a point where a function is equal to zero. Explicitly state the assumptions that the function needs to satisfy. Illustrate by the function θ defined by $\theta(\lambda)=\lambda^3-\lambda$ defined on the interval [0.5, 10.0].

8.13 It can be verified that, in Example 9.2.3, for a given value of μ, if $\mathbf{x}_\mu=(x_1,x_2)'$, then x_1 satisfies

$$2(x_1-2)^3+\frac{\mu x_1(8x_1^2-6x_1+1)}{4+\mu}=0$$

For $\mu=0.1$, 1.0, 10.0, and 100.0, find the value of x_1, satisfying the above equation using a suitable procedure.

8.14 Let $\theta:E_1\to E_1$ and suppose that we have the three points (λ_1,θ_1), (λ_2,θ_2), and (λ_3,θ_3), where $\theta_j=\theta(\lambda_j)$ for $j=1,2,3$. Show that the quadratic curve q passing through these points is given by

$$q(\lambda)=\frac{\theta_1(\lambda-\lambda_2)(\lambda-\lambda_3)}{(\lambda_1-\lambda_2)(\lambda_1-\lambda_3)}+\frac{\theta_2(\lambda-\lambda_1)(\lambda-\lambda_3)}{(\lambda_2-\lambda_1)(\lambda_2-\lambda_3)}+\frac{\theta_3(\lambda-\lambda_1)(\lambda-\lambda_2)}{(\lambda_3-\lambda_1)(\lambda_3-\lambda_2)}$$

Furthermore, show that the derivative of q vanishes at the point $\bar{\lambda}$ given by

$$\bar{\lambda}=\frac{1}{2}\,\frac{b_{23}\theta_1+b_{31}\theta_2+b_{12}\theta_3}{a_{23}\theta_1+a_{31}\theta_2+a_{12}\theta_3}$$

where $a_{ij}=\lambda_i-\lambda_j$ and $b_{ij}=\lambda_i^2-\lambda_j^2$. Find the quadratic curve passing through the points (1, 3), (2, 1), and (4, 6), and compute $\bar{\lambda}$.

8.15 Let $\theta:E_1\to E_1$, and suppose that we have the three points (λ_1,θ_1), (λ_2,θ_2), (λ_3,θ_3), where $\theta_j=\theta(\lambda_j)$. Furthermore, suppose that $\lambda_1<\lambda_2<\lambda_3$, $\theta_1\geq\theta_2$, and $\theta_2\leq\theta_3$. Utilizing Exercise 8.14 above, compute a minimum point $\bar{\lambda}$ of the quadratic form passing through the points (λ_1,θ_1), (λ_2,θ_2), and (λ_3,θ_3). If $\theta(\bar{\lambda})>\theta(\lambda_2)$, let $\bar{\lambda}_1=\lambda_1$, $\bar{\lambda}_2=\lambda_2$, and $\bar{\lambda}_3=\bar{\lambda}$. If, on the other hand, $\theta(\bar{\lambda})\leq\theta(\lambda_2)$, let $\bar{\lambda}_1=\lambda_2$, $\bar{\lambda}_2=\bar{\lambda}$, and $\bar{\lambda}_3=\lambda_3$. The process is repeated by letting $\lambda_1=\bar{\lambda}_1$, $\lambda_2=\bar{\lambda}_2$, and $\lambda_3=\bar{\lambda}_3$ and fitting a new quadratic form through the points (λ_1,θ_1), (λ_2,θ_2), and (λ_3,θ_3).

a. Propose a method for finding λ_1, λ_2, λ_3 such that $\lambda_1<\lambda_2<\lambda_3$, $\theta_1\geq\theta_2$, and $\theta_2\leq\theta_3$.

b. Show that if θ is strictly quasiconvex, then the new interval of uncertainty indeed contains the minimum.

c. Use the procedure described in this exercise to minimize $3\lambda-2\lambda^2+\lambda^3+2\lambda^4$ over $\lambda\geq 0$.

(We have described in Exercises 8.14 and 8.15 a derivative-free line search method using *quadratic fit.*)

8.16 Consider the problem to minimize $(x_1^3-x_2)^2+2(x_2-x_1)^4$. Solve the problem using each of the following methods. Do the methods converge to the same point? If not, explain.

a. The cyclic coordinate method.
b. The method of Hooke and Jeeves.
c. The method of Rosenbrock.
d. The method of Zangwill.
e. The method of steepest descent.
f. The method of Fletcher and Reeves.
g. The method of Davidon–Fletcher–Powell.

8.17 Consider the problem to minimize $(1-x_1)^2+5(x_2-x_1^2)^2$. Starting from the point (2, 0), solve the problem by the following procedures.

a. The cyclic coordinate method.
b. The method of Hooke and Jeeves.
c. The method of Rosenbrock.
d. The method of Davidon–Fletcher–Powell.
e. The method of Zangwill.

8.18 Solve the problem to maximize $x_1+2x_2+5x_1x_2-x_1^2+3x_2^2$ by the method of Hooke and Jeeves.

8.19 Consider the model $y=\alpha+\beta x+\gamma x^2+\varepsilon$, where x is the independent variable, y is the observed dependent variable, α, β, and γ are unknown parameters, and ε is a random component representing the experimental error. The following table gives the values of x and the corresponding values of y. Formulate the problem of finding the best estimates of α, β, and γ as an unconstrained optimization problem, by minimizing:

a. The sum of squared errors.
b. The sum of the absolute values of the errors.
c. The maximum absolute value of the error.

For each case, find α, β, and γ by a suitable method.

x	0	1	2	3	4	5
y	2	2	−12	−27	−60	−90

8.20 Let $f: E_n \to E_1$ be differentiable at $\mathbf{x}$ and let the vectors $\mathbf{d}_1, \ldots, \mathbf{d}_n$ in E_n be linearly independent. Suppose that the minimum of $f(\mathbf{x}+\lambda \mathbf{d}_j)$ over $\lambda \in E_1$ occurs at $\lambda = 0$, for $j = 1, \ldots, n$. Show that $\nabla f(\mathbf{x}) = \mathbf{0}$. Does this imply that f has a local minimum at $\mathbf{x}$?

8.21 Suppose that $\mathbf{x}_k$ and $\mathbf{x}_{k+1}$ are two consecutive points generated by the steepest descent method. Show that $\nabla f(\mathbf{x}_k)'\nabla f(\mathbf{x}_{k+1}) = 0$.

8.22 Consider the following problem.

$$\begin{aligned} &\text{Minimize} && x_1 + x_2 \\ &\text{subject to} && x_1^2 + x_2^2 = 4 \\ &&& -2x_1 - x_2 \le 4 \end{aligned}$$

a. Formulate the Lagrangian dual problem by incorporating both constraints into the objective function via the Lagrangian multipliers u_1 and u_2.
b. Using a suitable unconstrained optimization method, compute the gradient of the dual function θ at the point $(1, 2)$.
c. Starting from the point $\bar{\mathbf{u}} = (1, 2)'$, carry one iteration of the steepest ascent method of the dual problem. In particular, solve the following problem, where $\mathbf{d} = \nabla \theta(\bar{u})$.

$$\begin{aligned} &\text{Maximize} && \theta(\bar{\mathbf{u}} + \lambda \mathbf{d}) \\ &\text{subject to} && \bar{u}_2 + \lambda d_2 \ge 0 \\ &&& \lambda \ge 0 \end{aligned}$$

8.23 Suppose that f is twice continuously differentiable and that the Hessian matrix is invertible everywhere. Given $\mathbf{x}_k$, let $\mathbf{x}_{k+1} = \mathbf{x}_k + \lambda_k \mathbf{d}_k$, where $\mathbf{d}_k = -\mathbf{H}(\mathbf{x}_k)^{-1}\nabla f(\mathbf{x}_k)$ and λ_k is an optimal solution to the problem to minimize $f(\mathbf{x}_k + \lambda \mathbf{d}_k)$ subject to $\lambda \in E_1$. Show that this modification of Newton's method converges to a point in the solution set $\Omega = \{\bar{\mathbf{x}} : \nabla f(\bar{\mathbf{x}})'\mathbf{H}(\bar{\mathbf{x}})^{-1}\nabla f(\bar{\mathbf{x}}) = 0\}$. Illustrate by minimizing $(x_1 - 2)^4 + (x_1 - 2x_2)^2$ starting from the point $(-2, 3)$.

8.24 Let $\mathbf{H}$ be a symmetric $n \times n$ matrix, and let $\mathbf{d}_1, \ldots, \mathbf{d}_n$ be a set of characteristic vectors of $\mathbf{H}$. Show that $\mathbf{d}_1, \ldots, \mathbf{d}_n$ are $\mathbf{H}$-conjugate.

8.25 Let $\mathbf{a}_1, \ldots, \mathbf{a}_n$ be a set of linearly independent vectors in E_n, and let $\mathbf{H}$ be an $n \times n$ symmetric positive definite matrix.

a. Show that the vectors $\mathbf{d}_1, \ldots, \mathbf{d}_n$ defined below are $\mathbf{H}$-conjugate.

$$\mathbf{d}_k = \begin{cases} \mathbf{a}_k & \text{if } k = 1 \\ \mathbf{a}_k - \displaystyle\sum_{i=1}^{k-1} \left[\frac{\mathbf{d}_i'\mathbf{H}\mathbf{a}_k}{\mathbf{d}_i'\mathbf{H}\mathbf{d}_i}\right]\mathbf{d}_i & \text{if } k \ge 2 \end{cases}$$

b. Suppose that $\mathbf{a}_1, \ldots, \mathbf{a}_n$ are the unit vectors in E_n, and let $\mathbf{D}$ be the matrix whose columns are the vectors $\mathbf{d}_1, \ldots, \mathbf{d}_n$, defined in part a above. Show that $\mathbf{D}$ is upper triangular with all diagonal elements equal to one.
c. Illustrate by letting $\mathbf{a}_1 = (1, 0, 0)'$, $\mathbf{a}_2 = (1, -1, 4)'$, $\mathbf{a}_3 = (2, -1, 6)'$, and

$$\mathbf{H} = \begin{bmatrix} 1 & 0 & -1 \\ 0 & 3 & 4 \\ -1 & 4 & 2 \end{bmatrix}$$

d. Illustrate by letting $\mathbf{a}_1$, $\mathbf{a}_2$, and $\mathbf{a}_3$ be the unit vectors in E_3 and $\mathbf{H}$ as given in part c above.

8.26 Consider the quadratic form $f(\mathbf{x}) = \mathbf{c}'\mathbf{x} + \frac{1}{2}\mathbf{x}'\mathbf{H}\mathbf{x}$, where $\mathbf{H}$ is a symmetric $n \times n$ matrix. In many applications, it is desirable to obtain separability in the variables, by eliminating the cross-product terms. This could be done by rotating the axes as follows. Let $\mathbf{D}$ be an $n \times n$ matrix whose columns $\mathbf{d}_1, \ldots, \mathbf{d}_n$ are $\mathbf{H}$-conjugate. Letting $\mathbf{x} = \mathbf{D}\mathbf{y}$, verify that the quadratic form is equivalent to $\sum_{j=1}^n \alpha_j y_j + \frac{1}{2}\sum_{j=1}^n \beta_j y_j^2$, where $(\alpha_1, \ldots, \alpha_n) = \mathbf{c}'\mathbf{D}$, and $\beta_j = \mathbf{d}_j'\mathbf{H}\mathbf{d}_j$ for $j = 1, \ldots, n$. Furthermore, translating and rotating the axes could be accomplished by the transformation $\mathbf{x} = \mathbf{D}\mathbf{y} + \mathbf{z}$, where $\mathbf{z}$ is any vector satisfying $\mathbf{H}\mathbf{z} + \mathbf{c} = \mathbf{0}$, that is $\nabla f(\mathbf{z}) = \mathbf{0}$. In this case, show that the quadratic form is equivalent to $(\mathbf{c}'\mathbf{z} + \frac{1}{2}\mathbf{z}'\mathbf{H}\mathbf{z}) + \frac{1}{2}\sum_{j=1}^n \beta_j y_j^2$. Use the result of this exercise to draw accurate contours of the quadratic form $2x_1 - 4x_2 + x_1^2 + 2x_1x_2 + 3x_2^2$.

8.27 Consider the problem to maximize $-x_1^2 - x_2^2 + x_1x_2 - x_1 + 2x_2$. Starting from the origin, solve the problem by the Davidon–Fletcher–Powell method, with $\mathbf{D}_1$ as the identity. Also solve the problem by the conjugate gradient method. Note that the two procedures generate identical set of directions. Show that, in general, if $\mathbf{D}_1 = \mathbf{I}$, then the two methods are identical for quadratic functions.

8.28 Consider the problem to minimize $x_1^2 + x_2^2$ subject to $x_1 + x_2 - 2 = 0$.

a. Find the optimal solution to this problem, and verify optimality by the Kuhn–Tucker conditions.
b. One approach to solve the problem is to transform it into a problem of the form to minimize $x_1^2 + x_2^2 + \mu(x_1 + x_2 - 2)^2$, where $\mu > 0$ is a large scalar. Solve the unconstrained problem for $\mu = 10$ by the conjugate gradient method, starting from the origin.

8.29 Solve the problem to minimize $x_1 + 2x_2^2 + e^{x_1^2 + x_2^2}$, starting with the point $(1, 0)$ and using both the conjugate gradient method and the method of Zangwill.

8.30 Consider the following problem:

$$\text{Minimize } x_1^2 + 2x_1x_2 + 2x_2^2 + x_3^2 - x_2x_3 + x_1 + 3x_2 - x_3$$

Using Exercise 8.25 or any other method, generate three conjugate directions.

Starting from the origin, solve the problem by minimizing along these directions.

8.31 Consider the system of simultaneous equations

$$h_i(\mathbf{x}) = 0 \qquad \text{for } i = 1, \ldots, l$$

a. Show how to solve the above system by unconstrained optimization techniques.
(*Hint:* Consider the problem to minimize $\sum_{i=1}^{l} |h_i(\mathbf{x})|^p$, where p is a positive integer.)

b. Solve the following system

$$(x_1-2)^4+(x_1-2x_2)^2-5=0$$
$$x_1^2-x_2=0$$

8.32 Consider the problem to minimize $f(\mathbf{x})$ subject to $h_i(\mathbf{x})=0$ for $i=1,\ldots,l$. A point $\mathbf{x}$ is said to be a Kuhn–Tucker point if there exists a vector $\mathbf{v} \in E_l$ such that

$$\nabla f(\mathbf{x}) + \sum_{i=1}^{l} v_i \nabla h_i(\mathbf{x}) = \mathbf{0}$$

$$h_i(\mathbf{x}) = 0 \qquad \text{for } i = 1, \ldots, l$$

a. Show how to solve the above system using unconstrained optimization techniques.
(*Hint:* See Exercise 8.31 above.)

b. Find the Kuhn–Tucker point for the following problem.

Minimize $(x_1-2)^4+(x_1-2x_2)^2$

subject to $x_1^2-x_2=0$

8.33 Consider the problem to minimize $f(\mathbf{x})$ subject to $g_i(\mathbf{x}) \le 0$ for $i=1,\ldots,m$.

a. Show that the Kuhn–Tucker conditions hold at a point $\mathbf{x}$ if there exist u_i and s_i for $i=1,\ldots,m$ such that

$$\nabla f(\mathbf{x}) + \sum_{i=1}^{m} u_i^2 \nabla g_i(\mathbf{x}) = \mathbf{0}$$

$$g_i(\mathbf{x}) + s_i^2 = 0 \qquad \text{for } i = 1, \ldots, m$$

$$u_i s_i = 0 \qquad \text{for } i = 1, \ldots, m$$

b. Show that unconstrained optimization techniques could be used to find a solution to the above system.
(*Hint:* See Exercise 8.31 above.)

c. Use a suitable unconstrained optimization technique to find a Kuhn–Tucker point to the following problem:

Minimize $2x_1^2+x_2^2-2x_1x_2+2x_1+6$

subject to $-2x_1-x_2+3 \le 0$

8.34 A problem of the following structure frequently arises in the context of solving a more general nonlinear programming problem:

Minimize $f(\mathbf{x})$

subject to $a_i \le x_i \le b_i \qquad$ for $i=1,\ldots,m$

a. Investigate appropriate modifications of the unconstrained optimization methods discussed in this chapter so that the lower and upper bounds on the variables could be handled.

b. Use the results of part a to solve the following problem:

Minimize $(x_1-2)^4+(x_1-2x_2)^2$

subject to $3 \le x_1 \le 5$

$2 \le x_2 \le 6$

8.35 Consider the following method of *parallel tangents* credited to Shah, Buehler, and Kempthorne [1964] for minimizing a differentiable function f of several variables.

Initialization Step

Choose a termination scalar $\varepsilon > 0$, and choose a starting point $\mathbf{x}_1$. Let $\mathbf{y}_0 = \mathbf{x}_1$, $k = j = 1$, and go to the main step.

Main Step

1. Let $\mathbf{d} = -\nabla f(\mathbf{x}_k)$ and let $\hat{\lambda}$ be an optimal solution to the problem to minimize $f(\mathbf{x}_k+\lambda\mathbf{d})$ subject to $\lambda \ge 0$. Let $\mathbf{y}_1 = \mathbf{x}_k + \hat{\lambda}\mathbf{d}$. Go to step 2.
2. Let $\mathbf{d} = -\nabla f(\mathbf{y}_j)$, and let λ_j be an optimal solution to the problem to minimize $f(\mathbf{y}_j+\lambda\mathbf{d})$ subject to $\lambda \ge 0$. Let $\mathbf{z}_j = \mathbf{y}_j + \lambda_j\mathbf{d}$, and go to step 3.
3. Let $\mathbf{d} = \mathbf{z}_j - \mathbf{y}_{j-1}$, and let μ_j be an optimal solution to the problem to minimize $f(\mathbf{z}_j+\mu\mathbf{d})$ subject to $\mu \in E_1$. Let $\mathbf{y}_{j+1} = \mathbf{z}_j + \mu_j\mathbf{d}$. If $j<n$, replace j by $j+1$, and go to step 2. If $j=n$, go to step 4.
4. Let $\mathbf{x}_{k+1} = \mathbf{y}_{n+1}$. If $\|\mathbf{x}_{k+1}-\mathbf{x}_k\| < \varepsilon$, stop. Otherwise, let $\mathbf{y}_0 = \mathbf{x}_{k+1}$, replace k by $k+1$, let $j=1$, and go to step 1.

Using Theorem 7.3.4 show that the method converges. Solve the following problems using the method of parallel tangents.

a. Minimize $x_1^2+x_2^2+2x_1x_2-2x_1-6x_2$.
b. Minimize $x_1^2+x_2^2-2x_1x_2-2x_1-x_2$.
(Note that the optimal solution for this problem is unbounded.)
c. Minimize $(x_1-2)^2+(x_1-2x_2)^2$.

8.36 Let $f: E_n \to E_1$ be differentiable. Consider the following procedure for minimizing f.

Initialization Step

Choose a termination scalar $\varepsilon > 0$ and an initial step size $\Delta > 0$. Let m be a positive integer denoting the number of allowable failures before reducing step size. Let $\mathbf{x}_1$ be the starting point and let the current upper bound on the optimal objective value, $\text{UB} = f(\mathbf{x}_1)$. Let $\nu = 0$, let $k = 1$, and go to the main step.

Main Step

1. Let $\mathbf{d}_k = -\nabla f(\mathbf{x}_k)$, and let $\mathbf{x}_{k+1} = \mathbf{x}_k + \Delta\mathbf{d}_k$. If $f(\mathbf{x}_{k+1}) < \text{UB}$, let $\nu = 0$, $\hat{\mathbf{x}} = \mathbf{x}_{k+1}$, $\text{UB} = f(\hat{\mathbf{x}})$, and go to step 2. If, on the other hand, $f(\mathbf{x}_{k+1}) \ge \text{UB}$, then replace ν