Chapter 9
Conjugate Direction Methods

Conjugate direction methods can be regarded as being somewhat intermediate between the method of steepest descent and Newton's method. They are motivated by the desire to accelerate the typically slow convergence associated with steepest descent while avoiding the information requirements associated with the evaluation, storage, and inversion of the Hessian (or at least solution of a corresponding system of equations) as required by Newton's method.

Conjugate direction methods invariably are invented and analyzed for the purely quadratic problem

$$\text{minimize } \frac{1}{2}\mathbf{x}^T\mathbf{Q}\mathbf{x} - \mathbf{b}^T\mathbf{x},$$

where $\mathbf{Q}$ is an $n \times n$ symmetric positive definite matrix. The techniques once worked out for this problem are then extended, by approximation, to more general problems; it being argued that, since near the solution point every problem is approximately quadratic, convergence behavior is similar to that for the pure quadratic situation.

The area of conjugate direction algorithms has been one of great creativity in the nonlinear programming field, illustrating that detailed analysis of the pure quadratic problem can lead to significant practical advances. Indeed, conjugate direction methods, especially the method of conjugate gradients, have proved to be extremely effective in dealing with general objective functions and are considered among the best general purpose methods.

9.1 Conjugate Directions

Definition. Given a symmetric matrix $\mathbf{Q}$, two vectors $\mathbf{d}_1$ and $\mathbf{d}_2$ are said to be $\mathbf{Q}$-*orthogonal*, or *conjugate with respect to* $\mathbf{Q}$, if $\mathbf{d}_1^T\mathbf{Q}\mathbf{d}_2 = 0$.

D.G. Luenberger, Y. Ye, *Linear and Nonlinear Programming*, International Series in Operations Research & Management Science 228,
DOI 10.1007/978-3-319-18842-3_9

In the applications that we consider, the matrix $\mathbf{Q}$ will be positive definite but this is not inherent in the basic definition. Thus if $\mathbf{Q} = 0$, any two vectors are conjugate, while if $\mathbf{Q} = \mathbf{I}$, conjugacy is equivalent to the usual notion of orthogonality. A finite set of vectors $\mathbf{d}_0, \mathbf{d}_1, \ldots, \mathbf{d}_k$ is said to be a $\mathbf{Q}$-orthogonal set if $\mathbf{d}_i^T \mathbf{Q} \mathbf{d}_j = 0$ for all $i \neq j$.

Proposition. *If $\mathbf{Q}$ is positive definite and the set of nonzero vectors $\mathbf{d}_0, \mathbf{d}_1, \mathbf{d}_2, \ldots, \mathbf{d}_k$ are $\mathbf{Q}$-orthogonal, then these vectors are linearly independent.*

Proof. Suppose there are constants α_i, $i = 0, 1, 2, \ldots, k$ such that

$$\alpha_0 \mathbf{d}_0 + \cdots + \alpha_k \mathbf{d}_k = 0.$$

Multiplying by $\mathbf{Q}$ and taking the scalar product with $\mathbf{d}_i$ yields

$$\alpha_i \mathbf{d}_i^T \mathbf{Q} \mathbf{d}_i = 0.$$

Or, since $\mathbf{d}_i^T \mathbf{Q} \mathbf{d}_i > 0$ in view of the positive definiteness of $\mathbf{Q}$, we have $\alpha_i = 0$. ∎

Before discussing the general conjugate direction algorithm, let us investigate just why the notion of $\mathbf{Q}$-orthogonality is useful in the solution of the quadratic problem

$$\text{minimize } \frac{1}{2} \mathbf{x}^T \mathbf{Q} \mathbf{x} - \mathbf{b}^T \mathbf{x}, \tag{9.1}$$

when $\mathbf{Q}$ is positive definite. Recall that the unique solution to this problem is also the unique solution to the linear equation

$$\mathbf{Q} \mathbf{x} = \mathbf{b}, \tag{9.2}$$

and hence that the quadratic minimization problem is equivalent to a linear equation problem.

Corresponding to the $n \times n$ positive definite matrix $\mathbf{Q}$ let $\mathbf{d}_0, \mathbf{d}_1, \ldots, \mathbf{d}_{n-1}$ be n nonzero $\mathbf{Q}$-orthogonal vectors. By the above proposition they are linearly independent, which implies that the solution $\mathbf{x}^*$ of (9.1) or (9.2) can be expanded in terms of them as

$$\mathbf{x}^* = \alpha_0 \mathbf{d}_0 + \cdots + \alpha_{n-1} \mathbf{d}_{n-1} \tag{9.3}$$

for some set of α_i's. In fact, multiplying by $\mathbf{Q}$ and then taking the scalar product with $\mathbf{d}_i$ yields directly

$$\alpha_i = \frac{\mathbf{d}_i^T \mathbf{Q} \mathbf{x}^*}{\mathbf{d}_i^T \mathbf{Q} \mathbf{d}_i} = \frac{\mathbf{d}_i^T \mathbf{b}}{\mathbf{d}_i^T \mathbf{Q} \mathbf{d}_i}. \tag{9.4}$$

This shows that the α_i's and consequently the solution $\mathbf{x}^*$ can be found by evaluation of simple scalar products. The end result is

$$\mathbf{x}^* = \sum_{i=0}^{n-1} \frac{\mathbf{d}_i^T \mathbf{b}}{\mathbf{d}_i^T \mathbf{Q} \mathbf{d}_i} \mathbf{d}_i. \tag{9.5}$$

There are two basic ideas imbedded in (9.5). The first is the idea of selecting an orthogonal set of $\mathbf{d}_i$'s so that by taking an appropriate scalar product, all terms on the right side of (9.3), except the ***i*th**, vanish. This could, of course, have been accomplished by making the $\mathbf{d}_i$'s orthogonal in the ordinary sense instead of making them $\mathbf{Q}$-orthogonal. The second basic observation, however, is that by using $\mathbf{Q}$-orthogonality the resulting equation for α_i can be expressed in terms of the known vector $\mathbf{b}$ rather than the unknown vector $\mathbf{x}^*$; hence the coefficients can be evaluated without knowing $\mathbf{x}^*$.

The expansion for $\mathbf{x}^*$ can be considered to be the result of an iterative process of n steps where at the ith step $\alpha_i\mathbf{d}_i$ is added. Viewing the procedure this way, and allowing for an arbitrary initial point for the iteration, the basic conjugate direction method is obtained.

Conjugate Direction Theorem. *Let $\{\mathbf{d}_i\}_{i=0}^{n-1}$ be a set of nonzero $\mathbf{Q}$-orthogonal vectors. For any $\mathbf{x}_0 \in E^n$ the sequence $\{\mathbf{x}_k\}$ generated according to*

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k\mathbf{d}_k, \ k \geqslant 0 \tag{9.6}$$

with

$$\alpha_k = -\frac{\mathbf{g}_k^T\mathbf{d}_k}{\mathbf{d}_k^T\mathbf{Q}\mathbf{d}_k} \tag{9.7}$$

and

$$\mathbf{g}_k = \mathbf{Q}\mathbf{x}_k - \mathbf{b},$$

converges to the unique solution, $\mathbf{x}^*$, of $\mathbf{Q}\mathbf{x} = \mathbf{b}$ after n steps, that is, $\mathbf{x}_n = \mathbf{x}^*$.

Proof. Since the $\mathbf{d}_k$'s are linearly independent, we can write

$$\mathbf{x}^* - \mathbf{x}_0 = \alpha_0\mathbf{d}_0 + \alpha_1\mathbf{d}_1 + \cdots + \alpha_{n-1}\mathbf{d}_{n-1}$$

for some set of α_k's. As we did to get (9.4), we multiply by $\mathbf{Q}$ and take the scalar product with $\mathbf{d}_k$ to find

$$\alpha_k = \frac{\mathbf{d}_k^T\mathbf{Q}(\mathbf{x}^* - \mathbf{x}_0)}{\mathbf{d}_k^T\mathbf{Q}\mathbf{d}_k}. \tag{9.8}$$

Now following the iterative process (9.6) from $\mathbf{x}_0$ up to $\mathbf{x}_k$ gives

$$\mathbf{x}_k - \mathbf{x}_0 = \alpha_0\mathbf{d}_0 + \alpha_1\mathbf{d}_1 + \cdots + \alpha_{k-1}\mathbf{d}_{k-1}, \tag{9.9}$$

and hence by the $\mathbf{Q}$-orthogonality of the $\mathbf{d}_k$'s it follows that

$$\mathbf{d}_k^T\mathbf{Q}(\mathbf{x}_k - \mathbf{x}_0) = 0. \tag{9.10}$$

Substituting (9.10) into (9.8) produces

$$\alpha_k = \frac{\mathbf{d}_k^T\mathbf{Q}(\mathbf{x}^* - \mathbf{x}_k)}{\mathbf{d}_k^T\mathbf{Q}\mathbf{d}_k} = -\frac{\mathbf{g}_k^T\mathbf{d}_k}{\mathbf{d}_k^T\mathbf{Q}\mathbf{d}_k},$$

which is identical with (9.7). ∎

To this point the conjugate direction method has been derived essentially through the observation that solving (9.1) is equivalent to solving (9.2). The conjugate direction method has been viewed simply as a somewhat special, but nevertheless straightforward, orthogonal expansion for the solution to (9.2). This viewpoint, although important because of its underlying simplicity, ignores some of the most important aspects of the algorithm; especially those aspects that are important when extending the method to nonquadratic problems. These additional properties are discussed in the next section.

Also, methods for selecting or generating sequences of conjugate directions have not yet been presented. Some methods for doing this are discussed in the exercises; while the most important method, that of conjugate gradients, is discussed in Sect. 9.3.

9.2 Descent Properties of the Conjugate Direction Method

We define $\mathcal{B}_k$ as the subspace of E^n spanned by $\{\mathbf{d}_0, \mathbf{d}_1, \ldots, \mathbf{d}_{k-1}\}$. We shall show that as the method of conjugate directions progresses each $\mathbf{x}_k$ minimizes the objective over the k-dimensional linear variety $\mathbf{x}_0 + \mathcal{B}_k$.

Expanding Subspace Theorem. *Let $\{\mathbf{d}_i\}_{i=0}^{n-1}$ be a sequence of nonzero $\mathbf{Q}$-orthogonal vectors in E^n. Then for any $\mathbf{x}_0 \in E^n$ the sequence $\{\mathbf{x}_k\}$ generated according to*

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{d}_k \tag{9.11}$$

$$\alpha_k = -\frac{\mathbf{g}_k^T \mathbf{d}_k}{\mathbf{d}_k^T \mathbf{Q} \mathbf{d}_k} \tag{9.12}$$

has the property that $\mathbf{x}_k$ minimizes $f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T\mathbf{Q}\mathbf{x} - \mathbf{b}^T\mathbf{x}$ on the line $\mathbf{x} = \mathbf{x}_{k-1} + \alpha \mathbf{d}_{k-1}$, $-\infty < \alpha < \infty$, as well as on the linear variety $\mathbf{x}_0 + \mathcal{B}_k$.

Proof. It need only be shown that $\mathbf{x}_k$ minimizes f on the linear variety $\mathbf{x}_0 + \mathcal{B}_k$, since it contains the line $\mathbf{x} = \mathbf{x}_{k-1} + \alpha \mathbf{d}_{k-1}$. Since f is a strictly convex function, the conclusion will hold if it can be shown that $\mathbf{g}_k$ is orthogonal to $\mathcal{B}_k$ (that is, the gradient of f at $\mathbf{x}_k$ is orthogonal to the subspace $\mathcal{B}_k$). The situation is illustrated in Fig. 9.1. (Compare Theorem 2, Sect. 7.5.)

We prove $\mathbf{g}_k \perp \mathcal{B}_k$ by induction. Since $\mathcal{B}_0$ is empty that hypothesis is true for $k = 0$. Assuming that it is true for k, that is, assuming $\mathbf{g}_k \perp \mathcal{B}_k$, we show that $\mathbf{g}_{k+1} \perp \mathcal{B}_{k+1}$. We have

$$\mathbf{g}_{k+1} = \mathbf{g}_k + \alpha_k \mathbf{Q}\mathbf{d}_k, \tag{9.13}$$

and hence

$$\mathbf{d}_k^T \mathbf{g}_{k+1} = \mathbf{d}_k^T \mathbf{g}_k + \alpha_k \mathbf{d}_k^T \mathbf{Q} \mathbf{d}_k = 0 \tag{9.14}$$

by definition of α_k. Also for $i < k$

$$\mathbf{d}_i^T \mathbf{g}_{k+1} = \mathbf{d}_i^T \mathbf{g}_k + \alpha_k \mathbf{d}_i^T \mathbf{Q} \mathbf{d}_k. \tag{9.15}$$

The first term on the right-hand side of (9.15) vanishes because of the induction hypothesis, while the second vanishes by the **Q**-orthogonality of the $\mathbf{d}_i$'s. Thus $\mathbf{g}_{k+1} \perp \mathcal{B}_{k+1}$. ∎

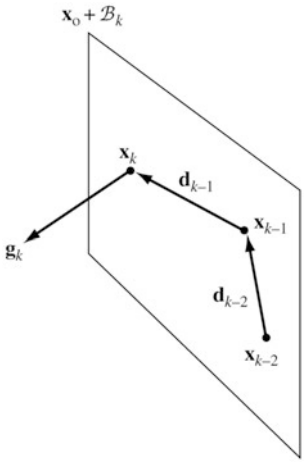

Fig. 9.1 Conjugate direction method

Corollary. *In the method of conjugate directions the gradients* $\mathbf{g}_k$, $k = 0, 1, \ldots, n$ *satisfy*

$$\mathbf{g}_k^T \mathbf{d}_i = 0 \text{ for } i < k.$$

The above theorem is referred to as the Expanding Subspace Theorem, since the $\mathcal{B}_k$'s form a sequence of subspaces with $\mathcal{B}_{k+1} \supset \mathcal{B}_k$. Since $\mathbf{x}_k$ minimizes f over $\mathbf{x}_0 + \mathcal{B}_k$, it is clear that $\mathbf{x}_n$ must be the overall minimum of f.

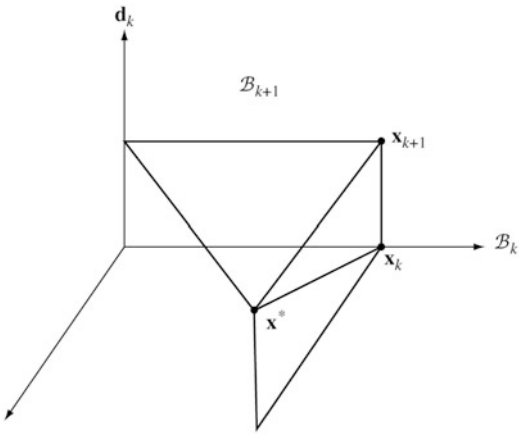

Fig. 9.2 Interpretation of expanding subspace theorem

To obtain another interpretation of this result we again introduce the function

$$E(\mathbf{x}) = \frac{1}{2}(\mathbf{x} - \mathbf{x}^*)^T \mathbf{Q}(\mathbf{x} - \mathbf{x}^*) \tag{9.16}$$

as a measure of how close the vector $\mathbf{x}$ is to the solution $\mathbf{x}^*$. Since $E(\mathbf{x}) = f(\mathbf{x}) + (1/2)\mathbf{x}^{*T}\mathbf{Q}\mathbf{x}^*$ the function E can be regarded as the objective that we seek to minimize.

By considering the minimization of E we can regard the original problem as one of minimizing a generalized distance from the point $\mathbf{x}^*$. Indeed, if we had $\mathbf{Q} = \mathbf{I}$, the generalized notion of distance would correspond (within a factor of two) to the usual Euclidean distance. For an arbitrary positive-definite $\mathbf{Q}$ we say E is a generalized Euclidean metric or distance function. Vectors $\mathbf{d}_i$, $i = 0, 1, \ldots, n-1$ that are $\mathbf{Q}$-orthogonal may be regarded as orthogonal in this generalized Euclidean space and this leads to the simple interpretation of the Expanding Subspace Theorem illustrated in Fig. 9.2. For simplicity we assume $\mathbf{x}_0 = \mathbf{0}$. In the figure $\mathbf{d}_k$ is shown as being orthogonal to $\mathcal{B}_k$ with respect to the generalized metric. The point $\mathbf{x}_k$ minimizes E over $\mathcal{B}_k$ while $\mathbf{x}_{k+1}$ minimizes E over $\mathcal{B}_{k+1}$. The basic property is that, since $\mathbf{d}_k$ is orthogonal to $\mathcal{B}_k$, the point $\mathbf{x}_{k+1}$ can be found by minimizing E along $\mathbf{d}_k$ and adding the result to $\mathbf{x}_k$.

9.3 The Conjugate Gradient Method

The conjugate gradient method is the conjugate direction method that is obtained by selecting the successive direction vectors as a conjugate version of the successive gradients obtained as the method progresses. Thus, the directions are not specified beforehand, but rather are determined sequentially at each step of the iteration. At step k one evaluates the current negative gradient vector and adds to it a linear combination of the previous direction vectors to obtain a new conjugate direction vector along which to move.

There are three primary advantages to this method of direction selection. First, unless the solution is attained in less than n steps, the gradient is always nonzero and linearly independent of all previous direction vectors. Indeed, the gradient $\mathbf{g}_k$ is orthogonal to the subspace $\mathcal{B}_k$ generated by $\mathbf{d}_0, \mathbf{d}_1, \ldots, \mathbf{d}_{k-1}$. If the solution is reached before n steps are taken, the gradient vanishes and the process terminates—it being unnecessary, in this case, to find additional directions.

Second, a more important advantage of the conjugate gradient method is the especially simple formula that is used to determine the new direction vector. This simplicity makes the method only slightly more complicated than steepest descent.

Third, because the directions are based on the gradients, the process makes good uniform progress toward the solution at every step. This is in contrast to the situation for arbitrary sequences of conjugate directions in which progress may be slight until the final few steps. Although for the pure quadratic problem uniform progress is of no great importance, it is important for generalizations to nonquadratic problems.

Conjugate Gradient Algorithm

Starting at any $\mathbf{x}_0 \in E^n$ define $\mathbf{d}_0 = -\mathbf{g}_0 = \mathbf{b} - \mathbf{Q}\mathbf{x}_0$ and

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{d}_k \tag{9.17}$$

$$\alpha_k = -\frac{\mathbf{g}_k^T \mathbf{d}_k}{\mathbf{d}_k^T \mathbf{Q}\mathbf{d}_k} \tag{9.18}$$

$$\mathbf{d}_{k+1} = -\mathbf{g}_{k+1} + \beta_k \mathbf{d}_k \tag{9.19}$$

$$\beta_k = \frac{\mathbf{g}_{k+1}^T \mathbf{Q}\mathbf{d}_k}{\mathbf{d}_k^T \mathbf{Q}\mathbf{d}_k}, \tag{9.20}$$

where $\mathbf{g}_k = \mathbf{Q}\mathbf{x}_k - \mathbf{b}$.

In the algorithm the first step is identical to a steepest descent step; each succeeding step moves in a direction that is a linear combination of the current gradient and the preceding direction vector. The attractive feature of the algorithm is the simple formulae, (9.19) and (9.20), for updating the direction vector. The method is only slightly more complicated to implement than the method of steepest descent but converges in a finite number of steps.

Verification of the Algorithm

To verify that the algorithm is a conjugate direction algorithm, it is necessary to verify that the vectors $\{\mathbf{d}_k\}$ are $\mathbf{Q}$-orthogonal. It is easiest to prove this by simultaneously proving a number of other properties of the algorithm. This is done in the theorem below where the notation $[\mathbf{d}_0, \mathbf{d}_1, \ldots, \mathbf{d}_k]$ is used to denote the subspace spanned by the vectors $\mathbf{d}_0, \mathbf{d}_1, \ldots, \mathbf{d}_k$.

Conjugate Gradient Theorem. *The conjugate gradient algorithm (9.17)–(9.20) is a conjugate direction method. If it does not terminate at $\mathbf{x}_k$, then*

a) $[\mathbf{g}_0, \mathbf{g}_1, \ldots, \mathbf{g}_k] = [\mathbf{g}_0, \mathbf{Q}\mathbf{g}_0, \ldots, \mathbf{Q}^k\mathbf{g}_0]$

b) $[\mathbf{d}_0, \mathbf{d}_1, \ldots, \mathbf{d}_k] = [\mathbf{g}_0, \mathbf{Q}\mathbf{g}_0, \ldots, \mathbf{Q}^k\mathbf{g}_0]$

c) $\mathbf{d}_k^T \mathbf{Q}\mathbf{d}_i = 0$ *for* $i \leqslant k-1$

d) $\alpha_k = \mathbf{g}_k^T \mathbf{g}_k / \mathbf{d}_k^T \mathbf{Q}\mathbf{d}_k$

e) $\beta_k = \mathbf{g}_{k+1}^T \mathbf{g}_{k+1} / \mathbf{g}_k^T \mathbf{g}_k$.

Proof. We first prove (a), (b) and (c) simultaneously by induction. Clearly, they are true for $k = 0$. Now suppose they are true for k, we show that they are true for $k+1$. We have

$$\mathbf{g}_{k+1} = \mathbf{g}_k + \alpha_k \mathbf{Q}\mathbf{d}_k.$$

By the induction hypothesis both $\mathbf{g}_k$ and $\mathbf{Q}\mathbf{d}_k$ belong to $[\mathbf{g}_0, \mathbf{Q}\mathbf{g}_0, \ldots, \mathbf{Q}^{k+1}\mathbf{g}_0]$, the first by (a) and the second by (b). Thus $\mathbf{g}_{k+1} \in [\mathbf{g}_0, \mathbf{Q}\mathbf{g}_0, \ldots, \mathbf{Q}^{k+1}\mathbf{g}_0]$. Furthermore $\mathbf{g}_{k+1} \notin [\mathbf{g}_0, \mathbf{Q}\mathbf{g}_0, \ldots, \mathbf{Q}^k\mathbf{g}_0] = [\mathbf{d}_0, \mathbf{d}_1, \ldots, \mathbf{d}_k]$ since otherwise $\mathbf{g}_{k+1} = 0$,

because for any conjugate direction method $\mathbf{g}_{k+1}$ is orthogonal to $[\mathbf{d}_0, \mathbf{d}_1, \ldots, \mathbf{d}_k]$. (The induction hypothesis on (c) guarantees that the method is a conjugate direction method up to $\mathbf{x}_{k+1}$.) Thus, finally we conclude that

$$[\mathbf{g}_0, \mathbf{g}_1, \ldots, \mathbf{g}_{k+1}] = [\mathbf{g}_0, \mathbf{Q}\mathbf{g}_0, \ldots, \mathbf{Q}^{k+1}\mathbf{g}_0],$$

which proves (a).

To prove (b) we write

$$\mathbf{d}_{k+1} = -\mathbf{g}_{k+1} + \beta_k \mathbf{d}_k,$$

and (b) immediately follows from (a) and the induction hypothesis on (b).

Next, to prove (c) we have

$$\mathbf{d}_{k+1}^T \mathbf{Q}\mathbf{d}_i = -\mathbf{g}_{k+1}^T \mathbf{Q}\mathbf{d}_i + \beta_k \mathbf{d}_k^T \mathbf{Q}\mathbf{d}_i.$$

For $i = k$ the right side is zero by definition of β_k. For $i < k$ both terms vanish. The first term vanishes since $\mathbf{Q}\mathbf{d}_i \in [\mathbf{d}_1, \mathbf{d}_2, \ldots, \mathbf{d}_{i+1}]$, the induction hypothesis which guarantees the method is a conjugate direction method up to $\mathbf{x}_{k+1}$, and by the Expanding Subspace Theorem that guarantees that $\mathbf{g}_{k+1}$ is orthogonal to $[\mathbf{d}_0, \mathbf{d}_1, \ldots, \mathbf{d}_{i+1}]$. The second term vanishes by the induction hypothesis on (c). This proves (c), which also proves that the method is a conjugate direction method.

To prove (d) we have

$$-\mathbf{g}_k^T \mathbf{d}_k = \mathbf{g}_k^T \mathbf{g}_k - \beta_{k-1} \mathbf{g}_k^T \mathbf{d}_{k-1},$$

and the second term is zero by the Expanding Subspace Theorem.

Finally, to prove (e) we note that $\mathbf{g}_{k+1}^T \mathbf{g}_k = 0$, because $\mathbf{g}_k \in [\mathbf{d}_0, \ldots, \mathbf{d}_k]$ and $\mathbf{g}_{k+1}$ is orthogonal to $[\mathbf{d}_0, \ldots, \mathbf{d}_k]$. Thus since

$$\mathbf{Q}\mathbf{d}_k = \frac{1}{\alpha_k}(\mathbf{g}_{k+1} - \mathbf{g}_k),$$

we have

$$\mathbf{g}_{k+1}^T \mathbf{Q}\mathbf{d}_k = \frac{1}{\alpha_k} \mathbf{g}_{k+1}^T \mathbf{g}_{k+1}. \blacksquare$$

Parts (a) and (b) of this theorem are a formal statement of the interrelation between the direction vectors and the gradient vectors. Part (c) is the equation that verifies that the method is a conjugate direction method. Parts (d) and (e) are identities yielding alternative formulae for α_k and β_k that are often more convenient than the original ones.

9.4 The C–G Method as an Optimal Process

We turn now to the description of a special viewpoint that leads quickly to some very profound convergence results for the method of conjugate gradients. The basis of the viewpoint is part (b) of the Conjugate Gradient Theorem. This result tells us

the spaces $\mathcal{B}_k$ over which we successively minimize are determined by the original gradient $\mathbf{g}_0$ and multiplications of it by $\mathbf{Q}$. Each step of the method brings into consideration an additional power of $\mathbf{Q}$ times $\mathbf{g}_0$. It is this observation we exploit.

Let us consider a new general approach for solving the quadratic minimization problem. Given an arbitrary starting point $\mathbf{x}_0$, let

$$\mathbf{x}_{k+1} = \mathbf{x}_0 + P_k(\mathbf{Q})\mathbf{g}_0, \tag{9.21}$$

where P_k is a polynomial of degree k. Selection of a set of coefficients for each of the polynomials P_k determines a sequence of $\mathbf{x}_k$'s. We have

$$\begin{aligned}\mathbf{x}_{k+1} - \mathbf{x}^* &= \mathbf{x}_0 - \mathbf{x}^* + P_k(\mathbf{Q})\mathbf{Q}(\mathbf{x}_0 - \mathbf{x}^*) \\ &= [\mathbf{I} + \mathbf{Q}P_k(\mathbf{Q})](\mathbf{x}_0 - \mathbf{x}^*),\end{aligned} \tag{9.22}$$

and hence

$$\begin{aligned}E(\mathbf{x}_{k+1}) &= \frac{1}{2}(\mathbf{x}_{k+1} - \mathbf{x}^*)^T\mathbf{Q}(\mathbf{x}_{k+1} - \mathbf{x}^*) \\ &= \frac{1}{2}(\mathbf{x}_0 - \mathbf{x}^*)^T\mathbf{Q}[\mathbf{I} + \mathbf{Q}P_k(\mathbf{Q})]^2(\mathbf{x}_0 - \mathbf{x}^*).\end{aligned} \tag{9.23}$$

We may now pose the problem of selecting the polynomial P_k in such a way as to minimize $E(\mathbf{x}_{k+1})$ with respect to all possible polynomials of degree k. Expanding (9.21), however, we obtain

$$\mathbf{x}_{k+1} = \mathbf{x}_0 + \gamma_0\mathbf{g}_0 + \gamma_1\mathbf{Q}\mathbf{g}_0 + \cdots + \gamma_k\mathbf{Q}^k\mathbf{g}_0, \tag{9.24}$$

where the γ_i's are the coefficients of P_k. In view of

$$\mathcal{B}_{k+1} = [\mathbf{d}_0,\ \mathbf{d}_1,\ \ldots,\ \mathbf{d}_k] = [\mathbf{g}_0,\ \mathbf{Q}\mathbf{g}_0,\ \ldots,\ \mathbf{Q}^k\mathbf{g}_0],$$

the vector $\mathbf{x}_{k+1} = \mathbf{x}_0 + \alpha_0\mathbf{d}_0 + \alpha_1\mathbf{d}_1 + \ldots + \alpha_k\mathbf{d}_k$ generated by the method of conjugate gradients has precisely this form; moreover, according to the Expanding Subspace Theorem, the coefficients γ_i determined by the conjugate gradient process are such as to minimize $E(\mathbf{x}_{k+1})$. Therefore, the problem posed of selecting the optimal P_k is solved by the conjugate gradient procedure.

The explicit relation between the optimal coefficients γ_i of P_k and the constants α_i, β_i associated with the conjugate gradient method is, of course, somewhat complicated, as is the relation between the coefficients of P_k and those of P_{k+1}. The power of the conjugate gradient method is that as it progresses it successively solves each of the optimal polynomial problems while updating only a small amount of information.

We summarize the above development by the following very useful theorem.

Theorem 1. *The point $\mathbf{x}_{k+1}$ generated by the conjugate gradient method satisfies*

$$E(\mathbf{x}_{k+1}) = \min_{P_k} \frac{1}{2}(\mathbf{x}_0 - \mathbf{x}^*)^T \mathbf{Q}[\mathbf{I} + \mathbf{Q}P_k(\mathbf{Q})]^2(\mathbf{x}_0 - \mathbf{x}^*), \tag{9.25}$$

where the minimum is taken with respect to all polynomials P_k of degree k.

Bounds on Convergence

To use Theorem 1 most effectively it is convenient to recast it in terms of eigenvectors and eigenvalues of the matrix $\mathbf{Q}$. Suppose that the vector $\mathbf{x}_0 - \mathbf{x}^*$ is written in the eigenvector expansion

$$\mathbf{x}_0 - \mathbf{x}^* = \xi_1\mathbf{e}_1 + \xi_2\mathbf{e}_2 + \cdots + \xi_n\mathbf{e}_n,$$

where the $\mathbf{e}_i$'s are normalized eigenvectors of $\mathbf{Q}$. Then since $\mathbf{Q}(\mathbf{x}_0 - \mathbf{x}^*) = \lambda_1\xi_1\mathbf{e}_1 + \lambda_2\xi_2\mathbf{e}_2 + \ldots + \lambda_n\xi_n\mathbf{e}_n$ and since the eigenvectors are mutually orthogonal, we have

$$E(\mathbf{x}_0) = \frac{1}{2}(\mathbf{x}_0 - \mathbf{x}^*)^T \mathbf{Q}(\mathbf{x}_0 - \mathbf{x}^*) = \frac{1}{2}\sum_{i=1}^{n} \lambda_i\xi_i^2, \tag{9.26}$$

where the λ_i's are the corresponding eigenvalues of $\mathbf{Q}$. Applying the same manipulations to (9.25), we find that for *any* polynomial P_k of degree k there holds

$$E(\mathbf{x}_{k+1}) \leqslant \frac{1}{2}\sum_{i=1}^{n}[1 + \lambda_i P_k(\lambda_i)]^2\lambda_i\xi_i^2.$$

It then follows that

$$E(\mathbf{x}_{k+1}) \leqslant \max_{\lambda_i}[1 + \lambda_i P_k(\lambda_i)]^2 \frac{1}{2}\sum_{i=1'}^{n} \lambda_i\xi_i^2,$$

and hence finally

$$E(\mathbf{x}_{k+1}) \leqslant \max_{\lambda_i}[1 + \lambda_i P_k(\lambda_i)]^2 E(\mathbf{x}_0).$$

We summarize this result by the following theorem.

Theorem 2. *In the method of conjugate gradients we have*

$$E(\mathbf{x}_{k+1}) \leqslant \max_{\lambda_i}[1 + \lambda_i P_k(\lambda_i)]^2 E(\mathbf{x}_0) \tag{9.27}$$

for any polynomial P_k of degree k, where the maximum is taken over all eigenvalues λ_i of $\mathbf{Q}$.

This way of viewing the conjugate gradient method as an optimal process is exploited in the next section. We note here that it implies the far from obvious fact that every step of the conjugate gradient method is at least as good as a steepest descent

step would be from the same point. To see this, suppose $\mathbf{x}_k$ has been computed by the conjugate gradient method. From (9.24) we know $\mathbf{x}_k$ has the form

$$\mathbf{x}_k = \mathbf{x}_0 + \bar{\gamma}_0\mathbf{g}_0 + \bar{\gamma}_1\mathbf{Q}\mathbf{g}_0 + \cdots + \bar{\gamma}_{k-1}\mathbf{Q}^{k-1}\mathbf{g}_0.$$

Now if $\mathbf{x}_{k+1}$ is computed from $\mathbf{x}_k$ by steepest descent, then $\mathbf{x}_{k+1} = \mathbf{x}_k - \alpha_k\mathbf{g}_k$ for some α_k. In view of part (a) of the Conjugate Gradient Theorem $\mathbf{x}_{k+1}$ will have the form (9.24). Since for the conjugate direction method $E(\mathbf{x}_{k+1})$ is lower than any other $\mathbf{x}_{k+1}$ of the form (9.24), we obtain the desired conclusion.

Typically when some information about the eigenvalue structure of $\mathbf{Q}$ is known, that information can be exploited by construction of a suitable polynomial P_k to use in (9.27). Suppose, for example, it were known that $\mathbf{Q}$ had only $m < n$ distinct eigenvalues. Then it is clear that by suitable choice of P_{m-1} it would be possible to make the mth degree polynomial $1 + \lambda P_{m-1}(\lambda)$ have its m zeros at the m eigenvalues. Using that particular polynomial in (9.27) shows that $E(\mathbf{x}_m) = 0$. Thus the optimal solution will be obtained in at most m, rather than n, steps. More sophisticated examples of this type of reasoning are contained in the next section and in the exercises at the end of the chapter.

9.5 The Partial Conjugate Gradient Method

A collection of procedures that are natural to consider at this point are those in which the conjugate gradient procedure is carried out for $m + 1 < n$ steps and then, rather than continuing, the process is restarted from the current point and $m + 1$ more conjugate gradient steps are taken. The special case of $m = 0$ corresponds to the standard method of steepest descent, while $m = n - 1$ corresponds to the full conjugate gradient method. These *partial conjugate gradient* methods are of extreme theoretical and practical importance, and their analysis yields additional insight into the method of conjugate gradients. The development of the last section forms the basis of our analysis.

As before, given the problem

$$\text{minimize } \frac{1}{2}\mathbf{x}^T\mathbf{Q}\mathbf{x} - \mathbf{b}^T\mathbf{x}, \tag{9.28}$$

we define for any point $\mathbf{x}_k$ the gradient $\mathbf{g}_k = \mathbf{Q}\mathbf{x}_k - \mathbf{b}$. We consider an iteration scheme of the form

$$\mathbf{x}_{k+1} = \mathbf{x}_k + P^k(\mathbf{Q})\mathbf{g}_k, \tag{9.29}$$

where P^k is a polynomial of degree m. We select the coefficients of the polynomial P^k so as to minimize

$$E(\mathbf{x}_{k+1}) = \frac{1}{2}(\mathbf{x}_{k+1} - \mathbf{x}^*)^T\mathbf{Q}(\mathbf{x}_{k+1} - \mathbf{x}^*), \tag{9.30}$$

where $\mathbf{x}^*$ is the solution to (9.28). In view of the development of the last section, it is clear that $\mathbf{x}_{k+1}$ can be found by taking $m+1$ conjugate gradient steps rather than explicitly determining the appropriate polynomial directly. (The sequence indexing is slightly different here than in the previous section, since now we do not give separate indices to the intermediate steps of this process. Going from $\mathbf{x}_k$ to $\mathbf{x}_{k+1}$ by the partial conjugate gradient method involves m other points.)

The results of the previous section provide a tool for convergence analysis of this method. In this case, however, we develop a result that is of particular interest for $\mathbf{Q}$'s having a special eigenvalue structure that occurs frequently in optimization problems, especially, as shown below and in Chap. 12, in the context of penalty function methods for solving problems with constraints. We imagine that the eigenvalues of $\mathbf{Q}$ are of two kinds: there are m large eigenvalues that may or may not be located near each other, and $n-m$ smaller eigenvalues located within an interval $[a,\ b]$. Such a distribution of eigenvalues is shown in Fig. 9.3.

As an example, consider as in Sect. 8.3 the problem on E^n

$$\begin{aligned}&\text{minimize } \frac{1}{2}\mathbf{x}^T\mathbf{Q}\mathbf{x} - \mathbf{b}^T\mathbf{x}\\ &\text{subject to } \mathbf{c}^T\mathbf{x} = 0,\end{aligned}$$

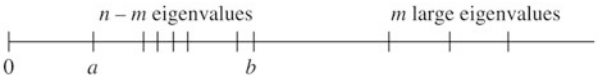

Fig. 9.3 Eigenvalue distribution

where $\mathbf{Q}$ is a symmetric positive definite matrix with eigenvalues in the interval $[a,\ A]$ and $\mathbf{b}$ and $\mathbf{c}$ are vectors in E^n. This is a constrained problem but it can be approximated by the unconstrained problem

$$\text{minimize } \frac{1}{2}\mathbf{x}^T\mathbf{Q}\mathbf{x} - \mathbf{b}^T\mathbf{x} + \frac{1}{2}\mu(\mathbf{c}^T\mathbf{x})^2,$$

where μ is a large positive constant. The last term in the objective function is called a *penalty term*; for large μ minimization with respect to $\mathbf{x}$ will tend to make $\mathbf{c}^T\mathbf{x}$ small.

The total quadratic term in the objective is $\frac{1}{2}\mathbf{x}^T(\mathbf{Q}+\mu\mathbf{c}\mathbf{c}^T)\mathbf{x}$, and thus it is appropriate to consider the eigenvalues of the matrix $\mathbf{Q}+\mu\mathbf{c}\mathbf{c}^T$. As μ tends to infinity it can be shown (see Chap. 13) that one eigenvalue of this matrix tends to infinity and the other $n-1$ eigenvalues remain bounded within the original interval $[a,\ A]$.

As noted before, if steepest descent were applied to a problem with such a structure, convergence would be governed by the ratio of the smallest to largest eigenvalue, which in this case would be quite unfavorable. In the theorem below it is stated that by successively repeating $m+1$ conjugate gradient steps the effects of the

m largest eigenvalues are eliminated and the rate of convergence is determined as if they were not present. A computational example of this phenomenon is presented in Sect. 13.5. The reader may find it interesting to read that section right after this one.

Theorem (Partial Conjugate Gradient Method). *Suppose the symmetric positive definite matrix* $\mathbf{Q}$ *has* $n-m$ *eigenvalues in the interval* $[a,\ b]$, $a > 0$ *and the remaining* m *eigenvalues are greater than* b. *Then the method of partial conjugate gradients, restarted every* $m+1$ *steps, satisfies*

$$E(\mathbf{x}_{k+1}) \leqslant \left(\frac{b-a}{b+a}\right)^2 E(\mathbf{x}_k). \tag{9.31}$$

(The point $\mathbf{x}_{k+1}$ *is found from* $\mathbf{x}_k$ *by taking* $m+1$ *conjugate gradient steps so that each increment in* k *is a composite of several simple steps.)*

Proof. Application of (9.27) yields

$$E(\mathbf{x}_{k+1} \leqslant \max_{\lambda_i}[1+\lambda_i P(\lambda_i)]^2 E(\mathbf{x}_k) \tag{9.32}$$

for any mth-order polynomial P, where the λ_i's are the eigenvalues of $\mathbf{Q}$. Let us select P so that the $(m+1)$th-degree polynomial $q(\lambda) = 1 + \lambda P(\lambda)$ vanishes at $(a+b)/2$ and at the m large eigenvalues of $\mathbf{Q}$. This is illustrated in Fig. 9.4. For this choice of P we may write (9.32) as

$$E(\mathbf{x}_{k+1}) \leqslant \max_{a\leqslant\lambda_i\leqslant b}[1+\lambda_i P(\lambda_i)]^2 E(\mathbf{x}_k).$$

Since the polynomial $q(\lambda) = 1 + \lambda P(\lambda)$ has $m+1$ real roots, $q'(\lambda)$ will have m real roots which alternate between the roots of $q(\lambda)$ on the real axis. Likewise, $q''(\lambda)$ will have $m-1$ real roots which alternate between the roots of $q'(\lambda)$. Thus, since $q(\lambda)$ has no root in the interval $(-\infty,\ (a+b)/2)$, we see that $q''(\lambda)$ does not change sign in that interval; and since it is easily verified that $q''(0) > 0$ it follows that $q(\lambda)$ is convex for $\lambda < (a+b)/2$. Therefore, on $[0,\ (a+b)/2]$, $q(\lambda)$ lies below the line $1 - [2\lambda/(a+b)]$. Thus we conclude that

$$q(\lambda) \leqslant 1 - \frac{2\lambda}{a+b}$$

on $[0,\ (a+b)/2]$ and that

$$q'\left(\frac{a+b}{2}\right) \geqslant -\frac{2}{a+b}.$$

We can see that on $[(a+b)/2,\ b]$

$$q(\lambda) \geqslant 1 - \frac{2\lambda}{a+b},$$

since for $q(\lambda)$ to cross first the line $1-[2\lambda/(a+b)]$ and then the λ-axis would require at least two changes in sign of $q''(\lambda)$, whereas, at most one root of $q''(\lambda)$ exists to the left of the second root of $q(\lambda)$. We see then that the inequality

$$|1 + \lambda P(\lambda)| \leqslant |1 - \frac{2\lambda}{a+b}|$$

is valid on the interval $[a,\ b]$. The final result (9.31) follows immediately. ∎

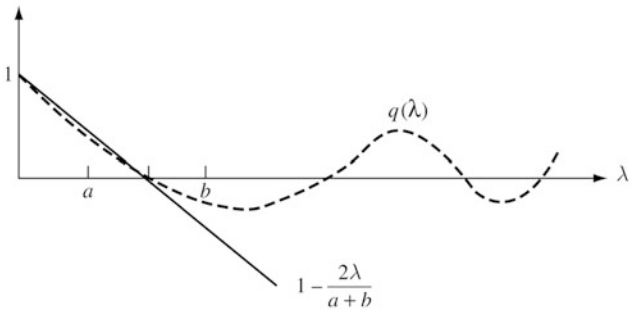

Fig. 9.4 Construction for proof

In view of this theorem, the method of partial conjugate gradients can be regarded as a generalization of steepest descent, not only in its philosophy and implementation, but also in its behavior. Its rate of convergence is bounded by exactly the same formula as that of steepest descent but with the largest eigenvalues removed from consideration. (It is worth noting that for $m = 0$ the above proof provides a simple derivation of the Steepest Descent Theorem.)

9.6 Extension to Nonquadratic Problems

The general unconstrained minimization problem on E^n

$$\text{minimize } f(\mathbf{x})$$

can be attacked by making suitable approximations to the conjugate gradient algorithm. There are a number of ways that this might be accomplished; the choice depends partially on what properties of f are easily computable. We look at three methods in this section and another in the following section.

Quadratic Approximation

In the quadratic approximation method we make the following associations at $\mathbf{x}_k$:

$$\mathbf{g}_k \leftrightarrow \nabla f(\mathbf{x}_k)^T, \qquad \mathbf{Q} \leftrightarrow \mathbf{F}(\mathbf{x}_k),$$

and using these associations, reevaluated at each step, all quantities necessary to implement the basic conjugate gradient algorithm can be evaluated. If f is quadratic, these associations are identities, so that the general algorithm obtained by using them is a generalization of the conjugate gradient scheme. This is similar to the philosophy underlying Newton's method where at each step the solution of a general problem is approximated by the solution of a purely quadratic problem through these same associations.

When applied to nonquadratic problems, conjugate gradient methods will not usually terminate within n steps. It is possible therefore simply to continue finding new directions according to the algorithm and terminate only when some termination criterion is met. Alternatively, the conjugate gradient process can be interrupted after n or $n+1$ steps and restarted with a pure gradient step. Since $\mathbf{Q}$-conjugacy of the direction vectors in the pure conjugate gradient algorithm is dependent on the initial direction being the negative gradient, the restarting procedure seems to be preferred. We always include this restarting procedure. The general conjugate gradient algorithm is then defined as below.

Step 1. Starting at $\mathbf{x}_0$ compute $\mathbf{g}_0 = \nabla f(\mathbf{x}_0)^T$ and set $\mathbf{d}_0 = -\mathbf{g}_0$.
Step 2. For $k = 0, 1, \ldots, n-1$:

(a) Set $\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{d}_k$ where $\alpha_k = \frac{-\mathbf{g}_k^T \mathbf{d}_k}{\mathbf{d}_k^T \mathbf{F}(\mathbf{x}_k)\mathbf{d}_k}$.
(b) Compute $\mathbf{g}_{k+1} = \nabla f(\mathbf{x}_{k+1})^T$.
(c) Unless $k = n-1$, set $\mathbf{d}_{k+1} = -\mathbf{g}_{k+1} + \beta_k \mathbf{d}_k$ where

$$\beta_k = \frac{\mathbf{g}_{k+1}^T \mathbf{F}(\mathbf{x}_k)\mathbf{d}_k}{\mathbf{d}_k^T \mathbf{F}(\mathbf{x}_k)\mathbf{d}_k}$$

and repeat (a).

Step 3. Replace $\mathbf{x}_0$ by $\mathbf{x}_n$ and go back to Step 1.

An attractive feature of the algorithm is that, just as in the pure form of Newton's method, no line searching is required at any stage. Also, the algorithm converges in a finite number of steps for a quadratic problem. The undesirable features are that $\mathbf{F}(\mathbf{x}_k)$ must be evaluated at each point, which is often impractical, and that the algorithm is not, in this form, globally convergent.

Line Search Methods

It is possible to avoid the direct use of the association $\mathbf{Q} \leftrightarrow \mathbf{F}(\mathbf{x}_k)$. First, instead of using the formula for α_k in Step 2(a) above, α_k is found by a line search that minimizes the objective. This agrees with the formula in the quadratic case. Second, the formula for β_k in Step 2(c) is replaced by a different formula, which is, however, equivalent to the one in 2(c) in the quadratic case.

The first such method proposed was the *Fletcher–Reeves method*, in which Part (e) of the Conjugate Gradient Theorem is employed; that is,

$$\beta_k = \frac{\mathbf{g}_{k+1}^T \mathbf{g}_{k+1}}{\mathbf{g}_k^T \mathbf{g}_k}.$$

The complete algorithm (using restarts) is:

Step 1. Given $\mathbf{x}_0$ compute $\mathbf{g}_0 = \nabla f(\mathbf{x}_0)^T$ and set $\mathbf{d}_0 = -\mathbf{g}_0$.
Step 2. For $k = 0, 1, \ldots, n-1$:

(a) Set $\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{d}_k$ where α_k minimizes $f(\mathbf{x}_k + \alpha \mathbf{d}_k)$.
(b) Compute $\mathbf{g}_{k+1} = \nabla f(\mathbf{x}_{k+1})^T$.
(c) Unless $k = n-1$, set $\mathbf{d}_{k+1} = -\mathbf{g}_{k+1} + \beta_k \mathbf{d}_k$ where

$$\beta_k = \frac{\mathbf{g}_{k+1}^T \mathbf{g}_{k+1}}{\mathbf{g}_k^T \mathbf{g}_k}.$$

Step 3. Replace $\mathbf{x}_0$ by $\mathbf{x}_n$ and go back to Step 1.

Another important method of this type is the *Polak–Ribiere method*, where

$$\beta_k = \frac{(\mathbf{g}_{k+1} - \mathbf{g}_k)^T \mathbf{g}_{k+1}}{\mathbf{g}_k^T \mathbf{g}_k}$$

is used to determine β_k. Again this leads to a value identical to the standard formula in the quadratic case. Experimental evidence seems to favor the Polak–Ribiere method over other methods of this general type.

Convergence

Global convergence of the line search methods is established by noting that a pure steepest descent step is taken every n steps and serves as a spacer step. Since the other steps do not increase the objective, and in fact hopefully they decrease it, global convergence is assured. Thus the restarting aspect of the algorithm is important for global convergence analysis, since in general one cannot guarantee that the directions $\mathbf{d}_k$ generated by the method are descent directions.

The local convergence properties of both of the above, and most other, nonquadratic extensions of the conjugate gradient method can be inferred from the quadratic analysis. Assuming that at the solution, $\mathbf{x}^*$, the matrix $\mathbf{F}(\mathbf{x}^*)$ is positive definite, we expect the asymptotic convergence rate per step to be at least as good as steepest descent, since this is true in the quadratic case. In addition to this bound on the single step rate we expect that the method is of order two with respect to each complete cycle of n steps. In other words, since one complete cycle solves a quadratic problem exactly just as Newton's method does in one step, we expect that

for general nonquadratic problems there will hold $|\mathbf{x}_{k+n} - \mathbf{x}^*| \leqslant c|\mathbf{x}_k - \mathbf{x}^*|^2$ for some c and $k = 0,\ n,\ 2n,\ 3n,\ \ldots$. This can indeed be proved, and of course underlies the original motivation for the method. For problems with large n, however, a result of this type is in itself of little comfort, since we probably hope to terminate in fewer than n steps. Further discussion on this general topic is contained in Sect. 10.4.

Scaling and Partial Methods

Convergence of the partial conjugate gradient method, restarted every $m + 1$ steps, will in general be linear. The rate will be determined by the eigenvalue structure of the Hessian matrix $\mathbf{F}(\mathbf{x}^*)$, and it may be possible to obtain fast convergence by changing the eigenvalue structure through scaling procedures. If, for example, the eigenvalues can be arranged to occur in $m + 1$ bunches, the rate of the partial method will be relatively fast. Other structures can be analyzed by use of Theorem 2, Sect. 9.4, by using $\mathbf{F}(\mathbf{x}^*)$ rather than Q.

*9.7 *Parallel Tangents

In early experiments with the method of steepest descent the path of descent was noticed to be highly zig-zag in character, making slow indirect progress toward the solution. (This phenomenon is now quite well understood and is predicted by the convergence analysis of Sect. 8.2.) It was also noticed that in two dimensions the solution point often lies close to the line that connects the zig-zag points, as illustrated in Fig. 9.5. This observation motivated the *accelerated gradient method* in which a complete cycle consists of taking two steepest descent steps and then searching along the line connecting the initial point and the point obtained after the two gradient steps. The method of parallel tangents (PARTAN) was developed through an

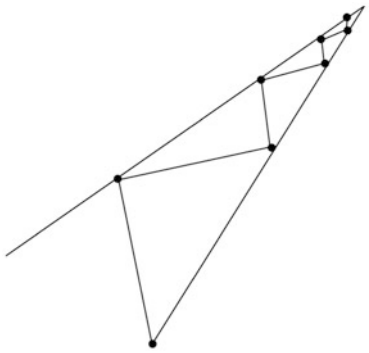

Fig. 9.5 Path of gradient method

attempt to extend this idea to an acceleration scheme involving all previous steps. The original development was based largely on a special geometric property of the tangents to the contours of a quadratic function, but the method is now recognized as a particular implementation of the method of conjugate gradients, and this is the context in which it is treated here.

The algorithm is defined by reference to Fig. 9.6. Starting at an arbitrary point $\mathbf{x}_0$ the point $\mathbf{x}_1$ is found by a standard steepest descent step. After that, from a point $\mathbf{x}_k$ the corresponding $\mathbf{y}_k$ is first found by a standard steepest descent step from $\mathbf{x}_k$, and then $\mathbf{x}_{k+1}$ is taken to be the minimum point on the line connecting $\mathbf{x}_{k-1}$ and $\mathbf{y}_k$. The process is continued for n steps and then restarted with a standard steepest descent step.

Notice that except for the first step, $\mathbf{x}_{k+1}$ is determined from $\mathbf{x}_k$, not by searching along a single line, but by searching along two lines. The direction $\mathbf{d}_k$ connecting two successive points (indicated as dotted lines in the figure) is thus determined only indirectly. We shall see, however, that, in the case where the objective function is quadratic, the $\mathbf{d}_k$'s are the same directions, and the $\mathbf{x}_k$'s are the same points, as would be generated by the method of conjugate gradients.

PARTAN Theorem. *For a quadratic function, PARTAN is equivalent to the method of conjugate gradients.*

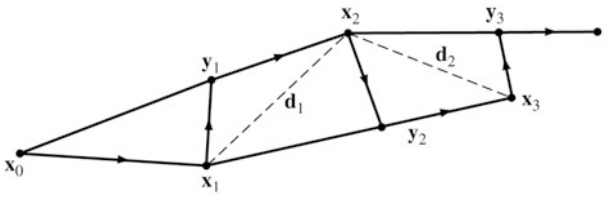

Fig. 9.6 PARTAN

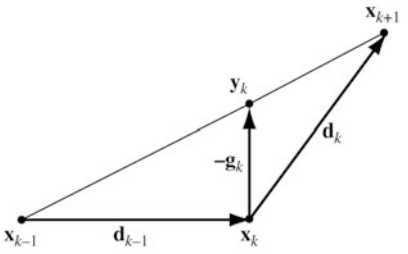

Fig. 9.7 One step of PARTAN

Proof. The proof is by induction. It is certainly true of the first step, since it is a steepest descent step. Suppose that $\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_k$ have been generated by the conjugate gradient method and $\mathbf{x}_{k+1}$ is determined according to PARTAN. This single

step is shown in Fig. 9.7. We want to show that $\mathbf{x}_{k+1}$ is the same point as would be generated by another step of the conjugate gradient method. For this to be true $\mathbf{x}_{k+1}$ must be that point which minimizes f over the plane defined by $\mathbf{d}_{k-1}$ and $\mathbf{g}_k = \nabla f(\mathbf{x}_k)^T$. From the theory of conjugate gradients, this point will also minimize f over the subspace determined by $\mathbf{g}_k$ and all previous $\mathbf{d}_i$' s. Equivalently, we must find the point $\mathbf{x}$ where $\nabla f(\mathbf{x})$ is orthogonal to both $\mathbf{g}_k$ and $\mathbf{d}_{k-1}$. Since $\mathbf{y}_k$ minimizes f along $\mathbf{g}_k$, we see that $\nabla f(\mathbf{y}_k)$ is orthogonal to $\mathbf{g}_k$. Since $\nabla f(\mathbf{x}_{k-1})$ is contained in the subspace $[\mathbf{d}_0, \mathbf{d}_1, \ldots, \mathbf{d}_{k-1}]$ and because $\mathbf{g}_k$ is orthogonal to this subspace by the Expanding Subspace Theorem, we see that $\nabla f(\mathbf{x}_{k-1})$ is also orthogonal to $\mathbf{g}_k$. Since $\nabla f(\mathbf{x})$ is linear in $\mathbf{x}$, it follows that at every point $\mathbf{x}$ on the line through $\mathbf{x}_{k-1}$ and $\mathbf{y}_k$ we have $\nabla f(\mathbf{x})$ orthogonal to $\mathbf{g}_k$. By minimizing f along this line, a point $\mathbf{x}_{k+1}$ is obtained where in addition $\nabla f(\mathbf{x}_{k+1})$ is orthogonal to the line. Thus $\nabla f(\mathbf{x}_{k+1})$ is orthogonal to both $\mathbf{g}_k$ and the line joining $\mathbf{x}_{k-1}$ and $\mathbf{y}_k$. It follows that $\nabla f(\mathbf{x}_{k+1})$ is orthogonal to the plane. ∎

There are advantages and disadvantages of PARTAN relative to other methods when applied to nonquadratic problems. One attractive feature of the algorithm is its simplicity and ease of implementation. Probably its most desirable property, however, is its strong global convergence characteristics. Each step of the process is at least as good as steepest descent; since going from $\mathbf{x}_k$ to $\mathbf{y}_k$ is exactly steepest descent, and the additional move to $\mathbf{x}_{k+1}$ provides further decrease of the objective function. Thus global convergence is not tied to the fact that the process is restarted every n steps. It is suggested, however, that PARTAN should be restarted every n steps (or $n + 1$ steps) so that it will behave like the conjugate gradient method near the solution.

An undesirable feature of the algorithm is that two line searches are required at each step, except the first, rather than one as is required by, say, the Fletcher–Reeves method. This is at least partially compensated by the fact that searches need not be as accurate for PARTAN, for while inaccurate searches in the Fletcher–Reeves method may yield nonsensical successive search directions, PARTAN will at least do as well as steepest descent.

9.8 Exercises

1. Let $\mathbf{Q}$ be a positive definite symmetric matrix and suppose $\mathbf{p}_0, \mathbf{p}_1, \ldots, \mathbf{p}_{n-1}$ are linearly independent vectors in E^n. Show that a Gram–Schmidt procedure can be used to generate a sequence of $\mathbf{Q}$-conjugate directions from the $\mathbf{p}_i$'s. Specifically, show that $\mathbf{d}_0, \mathbf{d}_1, \ldots, \mathbf{d}_{n-1}$ defined recursively by

$$\mathbf{d}_0 = \mathbf{p}_0$$
$$\mathbf{d}_{k+1} = \mathbf{p}_{k+1} - \sum_{i=0}^{k} \frac{\mathbf{p}_{k+1}^T \mathbf{Q} \mathbf{d}_i}{\mathbf{d}_i^T \mathbf{Q} \mathbf{d}_i} \mathbf{d}_i$$

form's a $\mathbf{Q}$-conjugate set.

2. Suppose the $\mathbf{p}_i$'s in Exercise 1 are generated as *moments* of $\mathbf{Q}$, that is, suppose $\mathbf{p}_k = \mathbf{Q}^k\mathbf{p}_0$, $k = 1, 2, \ldots, n-1$. Show that the corresponding $\mathbf{d}_k$'s can then be generated by a (three-term) recursion formula where $\mathbf{d}_{k+1}$ is defined only in terms of $\mathbf{Qd}_k, \mathbf{d}_k$ and $\mathbf{d}_{k-1}$.
3. Suppose the $\mathbf{p}_k$'s in Exercise 1 are taken as $\mathbf{p}_k = \mathbf{e}_k$ where $\mathbf{e}_k$ is the kth unit coordinate vector and the $\mathbf{d}_k$'s are constructed accordingly. Show that using $\mathbf{d}_k$'s in a conjugate direction method to minimize $(1/2)\mathbf{x}^T\mathbf{Qx} - \mathbf{b}^T\mathbf{x}$ is equivalent to the application of Gaussian elimination to solve $\mathbf{Qx} = \mathbf{b}$.
4. Let $f(\mathbf{x}) = (1/2)\mathbf{x}^T\mathbf{Qx} - \mathbf{b}^T\mathbf{x}$ be defined on E^n with $\mathbf{Q}$ positive definite. Let $\mathbf{x}_1$ be a minimum point of f over a subspace of E^n containing the vector $\mathbf{d}$ and let $\mathbf{x}_2$ be the minimum of f over another subspace containing $\mathbf{d}$. Suppose $f(\mathbf{x}_1) < f(\mathbf{x}_2)$. Show that $\mathbf{x}_1 - \mathbf{x}_2$ is $\mathbf{Q}$-conjugate to $\mathbf{d}$.
5. Let $\mathbf{Q}$ be a symmetric matrix. Show that any two eigenvectors of $\mathbf{Q}$, corresponding to distinct eigenvalues, are $\mathbf{Q}$-conjugate.
6. Let $\mathbf{Q}$ be an $n \times n$ symmetric matrix and let $\mathbf{d}_0, \mathbf{d}_1, \ldots, \mathbf{d}_{n-1}$ be $\mathbf{Q}$-conjugate. Show how to find an $\mathbf{E}$ such that $\mathbf{E}^T\mathbf{QE}$ is diagonal.
7. Show that in the conjugate gradient method $\mathbf{Qd}_{k-1} \in \mathcal{B}_{k+1}$.
8. Derive the rate of convergence of the method of steepest descent by viewing it as a one-step optimal process.
9. Let $P^k(\mathbf{Q}) = c_0 + c_1\mathbf{Q} + c_2\mathbf{Q}^2 + \cdots + c_m\mathbf{Q}^m$ be the optimal polynomial in (9.29) minimizing (9.30). Show that the c_i's can be found explicitly by solving the vector equation

$$-\begin{bmatrix} \mathbf{g}_k^T\mathbf{Qg}_k & \mathbf{g}_k^T\mathbf{Q}^2\mathbf{g}_k & \cdots & \mathbf{g}_k^T\mathbf{Q}^{m+1}\mathbf{g}_k \\ \mathbf{g}_k^T\mathbf{Q}^2 g_k & \mathbf{g}_k^T\mathbf{Q}^3\mathbf{g}_k & \cdots & \mathbf{g}_k^T\mathbf{Q}^{m+2}\mathbf{g}_k \\ \vdots & & & \\ \mathbf{g}_k^T\mathbf{Q}^{m+1}\mathbf{g}_k & \cdots & & \mathbf{g}_k^T\mathbf{Q}^{2m+1}\mathbf{g}_k \end{bmatrix} \begin{bmatrix} c_0 \\ c_1 \\ \vdots \\ c_m \end{bmatrix} = \begin{bmatrix} \mathbf{g}_k^T\mathbf{g}_k \\ \mathbf{g}_k^T\mathbf{Qg}_k \\ \vdots \\ \mathbf{g}_k^T\mathbf{Q}^m\mathbf{g}_k \end{bmatrix}$$

   Show that this reduces to steepest descent when $m = 0$.
10. Show that for the method of conjugate directions there holds

$$E(\mathbf{x}_k) \leqslant 4\left(\frac{1 - \sqrt{\gamma}}{1 + \sqrt{\gamma}}\right)^{2k} E(\mathbf{x}_0),$$

    where $\gamma = a/A$ and a and A are the smallest and largest eigenvalues of $\mathbf{Q}$. *Hint*: In (9.27) select $P_{k-1}(\lambda)$ so that

$$1 + \lambda P_{k-1}(\lambda) = \frac{T_k\left(\frac{A+a-2\lambda}{A-a}\right)}{T_k\left(\frac{A+a}{A-a}\right)},$$

where $T_k(\lambda) = \cos(k \arccos \lambda)$ is the kth Chebyshev polynomial. This choice gives the minimum maximum magnitude on $[a, A]$. Verify and use the inequality

$$\frac{(1-\gamma)^k}{(1+\sqrt{\gamma})^{2k} + (1-\sqrt{\gamma})^{2k}} \leqslant \left(\frac{1-\sqrt{\gamma}}{1+\sqrt{\gamma}}\right)^k.$$

11. Suppose it is known that each eigenvalue of $\mathbf{Q}$ lies either in the interval $[a, A]$ or in the interval $[a+\Delta, A+\Delta]$ where a, A, and Δ are all positive. Show that the partial conjugate gradient method restarted every two steps will converge with a ratio no greater than $[(A-a)/(A+a)]^2$ no matter how large Δ is.
12. Modify the first method given in Sect. 9.6 so that it is globally convergent.
13. Show that in the purely quadratic form of the conjugate gradient method $\mathbf{d}_k^T \mathbf{Q} \mathbf{d}_k = -\mathbf{d}_k^T \mathbf{Q} \mathbf{g}_k$. Using this show that to obtain $\mathbf{x}_{k+1}$ from $\mathbf{x}_k$ it is necessary to use $\mathbf{Q}$ only to evaluate $\mathbf{g}_k$ and $\mathbf{Q}\mathbf{g}_k$.
14. Show that in the quadratic problem $\mathbf{Q}\mathbf{g}_k$ can be evaluated by taking a unit step from $\mathbf{x}_k$ in the direction of the negative gradient and evaluating the gradient there. Specifically, if $\mathbf{y}_k = \mathbf{x}_k - \mathbf{g}_k$ and $\mathbf{p}_k = \nabla f(\mathbf{y}_k)^T$, then $\mathbf{Q}\mathbf{g}_k = \mathbf{g}_k - \mathbf{p}_k$.
15. Combine the results of Exercises 13 and 14 to derive a conjugate gradient method for general problems much in the spirit of the first method of Sect. 9.6 but which does not require knowledge of $\mathbf{F}(\mathbf{x}_k)$ or a line search.

References

9.1–9.3 For the original development of conjugate direction methods, see Hestenes and Stiefel [H10] and Hestenes [H7], [H9]. For another introductory treatment see Beckman [B8]. The method was extended to the case where Q is not positive definite, which arises in constrained problems, by Luenberger [L9], [L11].

9.4 The idea of viewing the conjugate gradient method as an optimal process was originated by Stiefel [S10]. Also see Daniel [D1] and Faddeev and Faddeeva [F1].

9.5 The partial conjugate gradient method presented here is identical to the so-called s-step gradient method. See Faddeev and Faddeeva [F1] and Forsythe [F14]. The bound on the rate of convergence given in this section in terms of the interval containing the $n-m$ smallest eigenvalues was first given in Luenberger [L13]. Although this bound cannot be expected to be tight, it is a reasonable conjecture that it becomes tight as the m largest eigenvalues tend to infinity with arbitrarily large separation.

9.6 For the first approximate method, see Daniel [D1]. For the line search methods, see Fletcher and Reeves [F12], Polak and Ribiere [P5], and Polak [P4]. For proof of the n-step, order two convergence, see Cohen [C4]. For a survey of computational experience of these methods, see Fletcher [F9].

9.7 PARTAN is due to Shah, Buehler, and Kempthome [S2]. Also see Wolfe [W5].

9.8 The approach indicated in Exercises 1 and 2 can be used as a foundation for the development of conjugate gradients; see Antosiewicz and Rheinboldt [A7], Vorobyev [V6], Faddeev and Faddeeva [F1], and Luenberger [L8]. The result stated in Exercise 3 is due to Hestenes and Stiefel [H10]. Exercise 4 is due to Powell [P6]. For the solution to Exercise 10, see Faddeev and Faddeeva [F1] or Daniel [D1].