JOURNAL OF OPTIMIZATION THEORY AND APPLICATIONS: Vol. 43, No. 2, JUNE 1984

Planar Quasi-Newton Algorithms for Unconstrained Saddlepoint Problems[1]

S. S. OREN[2]

Communicated by H. Y. Huang

Abstract. A new class of quasi-Newton methods is introduced that can locate a unique stationary point of an n-dimensional quadratic function in at most n steps. When applied to positive-definite or negative-definite quadratic functions, the new class is identical to Huang's symmetric family of quasi-Newton methods (Ref. 1). Unlike the latter, however, the new family can handle indefinite quadratic forms and therefore is capable of solving saddlepoint problems that arise, for instance, in constrained optimization. The novel feature of the new class is a planar iteration that is activated whenever the algorithm encounters a near-singular direction of search, along which the objective function approaches zero curvature. In such iterations, the next point is selected as the stationary point of the objective function over a plane containing the problematic search direction, and the inverse Hessian approximation is updated with respect to that plane via a new four-parameter family of rank-three updates. It is shown that the new class possesses properties which are similar to or which generalize the properties of Huang's family. Furthermore, the new method is equivalent to Fletcher's (Ref. 2) modified version of Luenberger's (Ref. 3) hyperbolic pairs method, with respect to the metric defined by the initial inverse Hessian approximation. Several issues related to implementing the proposed method in nonquadratic cases are discussed.

Key Words. Nonlinear optimization, quasi-Newton methods, conjugate-directions algorithms, indefinite quadratic forms, saddlepoint problems.

---

[1] An earlier version of this paper was presented at the 10th Mathematical Programing Symposium, Montreal, Canada, 1979.

[2] Associate Professor, Department of Industrial Engineering and Operations Research, University of California, Berkeley, California.

1. Introduction

Variable-metric methods, which are also referred to as quasi-Newton methods, were originally developed by Davidon (Ref. 4) and by Fletcher and Powell (Ref. 5) for solving smooth unconstrained minimization problems of the form: minimize $f(x)$, where $x \in E^n$ and $f \in C^2$.

As many well-known algorithms, quasi-Newton methods are based on a quadratic model of the objective function. In other words, they are obtained by assuming a quadratic function and adapted to handle nonquadratic problems by applying them to local quadratic approximations of the objective. These approximations are obtained from gradient information and updated on each iteration as new gradient measurements become available. Variable metric methods are typically quasi-Newton algorithms that guarantee the positive definiteness of these local quadratic approximations (these two terms, however, are often treated as synonyms). A general class of quasi-Newton methods, which contains as special cases most algorithms of this type, is the so-called Huang's (Ref. 1) symmetric family. An equivalent representation of this family has been also given by Oren and Luenberger (Ref. 6). For the quadratic case,

$$f(x) = (\tfrac{1}{2})x^T F x - b^T x + c, \tag{1}$$

where F is a symmetric $n \times n$ positive-definite matrix, b is an n-vector, and c a scalar, this class can be represented conceptually by the following iterative process.

Algorithm 1.1. Start with an arbitrary point x_0, an $n \times n$ positive-definite symmetric matrix H_0, and the gradient

$$g_0 = \nabla f(x_0)^T = Fx_0 - b.$$

Step 1. If $g_k \neq 0$, then obtain

$$d_k = -H_k g_k. \tag{2}$$

Step 2. Obtain

$$x_{k+1} = x_k + \alpha_k d_k, \tag{3}$$

where

$$\alpha_k = \{\alpha | \nabla f(x_k + \alpha d_k) d_k = 0\} = -(g_k^T d_k)/(d_k^T F d_k). \tag{4}$$

Step 3. Obtain

$$g_{k+1} = \nabla f(x_{k+1})^T = Fx_{k+1} - b, \tag{5}$$

$$p_k = x_{k+1} - x_k = \alpha_k d_k, \tag{6}$$

$$q_k = g_{k+1} - g_k = Fp_k. \tag{7}$$

Step 4. Select arbitrary scalars γ_k and $\varphi_k \neq 1/g_{k+1}^T H_k g_{k+1}$ and obtain

$$H_{k+1} = H_k + (1/p_k^T q_k)(p_k v_k^T + v_k p_k^T + \gamma_k p_k p_k^T) + \varphi_k v_k v_k^T, \tag{8}$$

where

$$v_k = (q_k^T H_k q_k / p_k^T q_k) p_k - H_k q_k. \tag{9}$$

Step 5. Increment k by 1 and return to Step 1.

The parameters φ_k and γ_k in the above algorithm provide the two degrees of freedom available in Huang's (Ref. 1) symmetric family. The class of algorithms described in Ref. 6 differs from the above by having the parameter γ_k multiply H_k rather than $p_k p_k^T$. Specific well-known variable-metric methods, such as the DFP algorithm given in Refs. 4 and 5 or the BFGS algorithm presented, for example, in Ref. 7, can be obtained from this general class by appropriate selections of φ_k and γ_k.

Some of the important well-known properties of Algorithm 1.1 are summarized for reference in the following theorem, without proof.

Theorem 1.1. For a positive-definite, quadratic objective function with Hessian F, Algorithm 1.1 satisfies the following conditions, for $k = 0, 1, \ldots, n-1$.

$$\text{(a)} \quad g_i^T d_j = d_i^T F d_j = 0, \qquad \text{for } i < j \leq k, \tag{10}$$

and

$$g_j^T d_j < 0 \quad \text{if } g_j \neq 0, \qquad \text{for } j \leq k. \tag{11}$$

Thus, Algorithm 1.1 is a *conjugate-direction descent method* and will hence converge to a stationary point in at most n steps.

(b) If $H_0 = I$, then Algorithm 1.1 is a *conjugate-gradient algorithm*, i.e.,

$$\langle d_0, d_1, \ldots, d_k \rangle = \langle g_0, g_1, \ldots, g_k \rangle = \langle g_0, Fg_0, \ldots, F^{k-1} g_0 \rangle^3. \tag{12}$$

$$\text{(c)} \quad H_{k+1} q_j = H_{k+1} F p_j = \omega_j p_j, \qquad \text{for } j \leq k, \tag{13}$$

where

$$\omega_j = \gamma_j + (q_j^T H_j q_j / p_j^T q_j) \tag{14}$$

[3] Throughout this paper, $\langle y_0, \ldots, y_k \rangle$ denotes the subspace spanned by the vectors $y_0, \ldots, y_k$, while $[y_0, \ldots, y_k]$ is the matrix formed by these vectors.

and, if

$$\gamma_j = 1 - (q_j^T H_j q_j / p_j^T q_j), \qquad \text{for } j = 0, 1, \ldots, n-1,$$

then

$$H_n = F^{-1}.$$

(d) Any symmetric rank-two update of the form

$$H_{k+1} = H_k + [H_k q_k, p_k] \Phi [H_k q_k, p_k]^T \tag{15}$$

satisfies (13) if and only if it can be expressed in the form given by (8) and (9).

Proof. The proofs of parts (a), (b), (c) of Theorem 1.1 for special cases of Algorithm 1.1 were given by various authors (e.g., Refs. 7 and 5). Huang (Ref. 1), however, was the first to consider the general case presented above. □

A key assumption in the development of the presently known variable-metric algorithms and in particular those included in Huang's family is that

$$d_k^T F d_k \neq 0, \qquad \text{for all } k.$$

This is guaranteed by assuming definiteness of the quadratic objective function used in the derivation. In fact, Theorem 1.1 holds as long as

$$d_k^T F d_k \neq 0, \qquad \text{for all } k,$$

regardless of whether F is definite or not. However, in the indefinite case, we may encounter situations where, for some k, $d_k^T F d_k$ and hence $p_k^T q_k$ vanish, causing the method to break down; in practice, this will occur even when

$$d_k^T F d_k = \delta \neq 0, \qquad \text{for small } \delta.$$

The definiteness assumption is reasonable if the method is designed for minimization or maximization purposes. In certain applications, however, we may need to find a stationary point of a function where the Hessian matrix is indefinite, i.e., a saddlepoint [see, e.g., Sinclair and Fletcher (Ref. 8)]. The appropriate quadratic model for such a problem is of the form (1), where F is a nonsingular, symmetric indefinite matrix. Locating the stationary point of this quadratic is equivalent to solving the linear system of equations

$$Fx = b.$$

A classic example leading to such a problem is minimizing the quadratic function

$$f(x)=(1/2)x^TGx-h^Tx+c, \tag{16}$$

subject to

$$Ax=a, \tag{17}$$

where G is a symmetric $n\times n$ matrix, h is an n-vector, c is a scalar, A is an $m\times n$ matrix of rank m, and a is an m-vector. For this problem to have a unique solution, it is sufficient that G be positive definite over the subspace $\{y|Ay=0\}$. This solution can be obtained by finding the saddlepoint of the Lagrangian

$$\begin{aligned} L(x,\lambda) &= f(x)+\lambda^T(Ax-a)\\ &= \tfrac{1}{2}[x^T,\ \lambda^T]\begin{bmatrix} G & A^T\\ A & 0\end{bmatrix}\begin{bmatrix} x\\ \lambda\end{bmatrix}-[h^T,a^T]\begin{bmatrix} x\\ \lambda\end{bmatrix}+c. \end{aligned} \tag{18}$$

It is possible, of course, to find the saddlepoint of a nonsingular indefinite quadratic function $f(x)$ by minimizing the squared norm of its gradient $\|\nabla f(x)\|^2$. The Hessian of the new objective is the square of the original Hessian; thus, it is positive definite and standard methods can be applied safely. This approach, however, requires more computational work per iteration as compared to a direct method, due to the squaring of the Hessian. Furthermore, squaring the condition number of the Hessian, implied by this approach, has an adverse effect from a numerical stability point of view.

The alternative approach adopted in this paper addresses explicitly possible indefiniteness of the Hessian by devising algorithms that can handle such cases. Existing work along these lines has concentrated on generalization and alternative implementation of conjugate-gradient algorithms so as to handle indefinite quadratic forms. The first method of this type was proposed by Luenberger (Ref. 3), followed by subsequent methods described by Luenberger (Ref. 9), Paige and Saunders (Ref. 10), Fletcher (Ref. 5), Sinclair and Fletcher (Ref. 8), and Hestenes (Ref. 11). Luenberger's (Ref. 3) method employs the concept of *hyperbolic pairs* to devise an alternate planar iteration that is executed when the regular conjugate-gradient iteration breaks down due to division by zero. Such division by zero occurs when the direction of search d_k is singular, i.e.,

$$d_k^TFd_k=0.$$

In its original form, this method does not adequately cope with the situation when $d_k^TFd_k$ is very small, but not exactly zero. Fletcher (Ref. 2) has shown, however, that this difficulty may be alleviated by performing an iteration which executes simultaneously two regular conjugate-gradient steps, whenever $d_k^T F d_k$ becomes too small. He has further shown that, when $d_k^T F d_k$ is exactly zero, this double iteration reduces to the one proposed by Luenberger (Ref. 3) for this case.

It is well known that, in spite of the theoretical equivalence of quasi-Newton and conjugate-gradient methods for definite quadratic functions, the former class of methods is superior to the latter when applied to nonquadratic function minimization. Recent results by Shanno (Ref. 12) shed some light on the above observation by showing that the conjugate-gradient method may be viewed as a memoryless quasi-Newton method. This lack of memory manifests itself in higher sensitivity to line-search accuracy and nonquadratic nature of the objective function. Thus, for small and moderate size problems, where the storage and overhead requirements for matrix updates are not substantial, quasi-Newton methods are commonly preferred. Although no empirical support is yet available, it is conceivable that the above considerations will apply to saddlepoint problems as well. It seems, therefore, worthwhile to develop generalized quasi-Newton methods that can handle indefinite Hessians. Such methods will have the added benefit of producing an estimate of the inverse Hessian which is often needed for various applications. In the example discussed above, for instance, the inverse Hessian of the Lagrangian may be required for sensitivity analysis (see, for example, Ref. 13).

In this paper, we generalize Huang's symmetric family of quasi Newton algorithms so as to handle indefinite Hessians. For this purpose, we shall assume, in our derivations and analysis of the proposed method, the indefinite quadratic objective function given by (1), where F is a nonsingular indefinite matrix. For notational convenience, we use F explicitly in many of the results presented throughout this paper. It should be understood, however, that the proposed algorithms are intended, like conventional quasi-Newton methods, to be implemented without explicit knowledge of the Hessian. In discussing the adaptation of these methods to nonquadratic cases, we will hence show how the various quantities involving F can be obtained in terms of gradients of the objective function and by means of conventional line searches.

2. Possible Breakdown in Huang's Class

The concept of conjugacy that is normally used with respect to positive-definite matrices may be extended to symmetric nonsingular matrices as well. The following definition provides such a generalization along with other concepts that are commonly used in the analysis of indefinite quadratic forms.

Definition 2.1. Given a symmetric nonsingular $n \times n$ matrix Q,

(a) two nonzero vectors $x, y \in E^n$ are said to be conjugate with respect to Q if $x^T Q y = 0$;

(b) a nonzero vector $x \in E^n$ is said to be singular with respect to Q if $x^T Q x = 0$;

(c) a pair of vectors $\{x, y\}$ is said to be a *hyperbolic pair* with respect to Q if x and y are both singular with respect to Q and $x^T Q y \neq 0$.

As pointed out earlier, Algorithm 1.1 can break down when applied to an indefinite quadratic form, since it may generate a singular direction of search d_k, along which the function has zero curvature, i.e.,

$$d_k^T F d_k = p_k^T q_k = 0.$$

The difficulties arise in calculating the stepsize α_k and in updating H_k. Equation (4) for calculating α_k implies that x_{k+1} be a stationary point along the line $x_k + \alpha d_k$. Clearly, if d_k is singular, the one-dimensional function

$$f(\alpha) = f(x_k - \alpha d_k)$$

is linear in α and therefore has no stationary points (except for the degenerate case, where $\alpha = 0$ is stationary). If we attempt to resolve this problem by selecting some arbitrary $\alpha_k > 0$, then a further difficulty resulting from the singularity of d_k arises in updating H_k. As will be shown below, using an update of the type given by (15), when d_k is singular, will either fail to produce an H_{k+1} satisfying condition (13) or d_{k+1} will not be conjugate to d_k unless the two search directions are parallel.

Proposition 2.1. Let H_{k+1} be obtained from $H_k q_k$ and p_k using (15), where

$$q_k = F p_k \quad \text{and} \quad p_k^T q_k = 0.$$

Then, the condition

$$H_{k+1} q_k = \omega_k p_k \tag{19}$$

implies that either

$$H_k q_k = \omega_k p_k$$

and

$$H_{k+1} = H_k + \sigma_k p_k p_k^T \tag{20}$$

or

$$q_k^T H_k q_k \neq 0$$

and

$$H_{k+1} = H_k - (1/q_k^T H_k q_k)(H_k q_k - \omega_k p_k)(H_k q_k - \omega_k p_k)^T + \sigma_k p_k p_k^T, \quad (21)$$

for some arbitrary scalar σ_k.

Proof. For convenience, we suppress the subscript k and denote H_{k+1} by H^*. From (13), (15), and the fact that

$$p^T q = 0,$$

we have

$$H^* Fp = H^* q = Hq + [Hq, p]\Phi[q^T Hq, 0]^T = \omega p, \quad (22)$$

which can be rewritten as

$$Hq\{1 + \Phi_{11}(q^T Hq)\} + p\{\Phi_{21}(q^T Hq) - \omega\} = 0. \quad (23)$$

If Hq and p are collinear, then (15) may be collapsed to a rank-one correction of the form given by (20). Furthermore, since

$$q^T Hq = p^T Fp = 0,$$

(23) implies

$$Hq = \omega p.$$

If, on the other hand, Hq and p are linearly independent, then (23) can be satisfied only if $q^T Hq \neq 0$, while

$$\Phi_{11} = -1/(q^T Hq), \quad (24)$$

$$\Phi_{21} = \omega/(q^T Hq). \quad (25)$$

Substituting (24) and (25) into (15) yields (21), with

$$\sigma = \Phi_{22}.$$

□

Proposition 2.2. Let H_k, p_k, q_k, d_k, g_k be defined as in Algorithm 1.1, where

$$p_k^T q_k = 0,$$

and let H_{k+1} satisfy (15) and (19). Then, either d_{k+1} is parallel to p_k or

$$d_{k+1}^T F p_k \neq 0.$$

Proof. From (2) and (7), we have

$$d_{k+1} = -H_{k+1} g_{k+1} = -H_{k+1} q_k - H_{k+1} g_k. \quad (26)$$

But, since H_{k+1} satisfies (19), we obtain, by (6),

$$d_{k+1} = -\omega_k p_k - H_{k+1} g_k. \tag{27}$$

If $H_k q_k$ is collinear with p_k, then, according to Proposition 2.1, H_{k+1} is determined by (20); and, since

$$H_k g_k = -(1/\alpha_k) p_k,$$

one can easily show that d_{k+1} is collinear with p_k. On the other hand, if $H_k q_k$ and p_k are independent, then H_{k+1} is determined by (21). From (25) and (21), it then follows that d_{k+1} is in the subspace spanned by p_k and $H_k q_k$, i.e.,

$$d_{k+1} = \xi p_k + \zeta H_k q_k, \tag{28}$$

for some scalars ξ and ζ. Thus, by the singularity of p_k, we have

$$d_{k+1}^T F p_k = \xi p_k^T q_k + \zeta q_k^T H_k q_k = \zeta q_k^T H_k q_k, \tag{29}$$

implying that either

$$d_{k+1}^T F p_k \neq 0$$

or

$$\zeta = 0 \quad \text{and} \quad d_{k+1} = \xi p_k. \qquad \square$$

3. Planar Iteration

Propositions 2.1 and 2.2 above assume that $p_k^T q_k$ is exactly zero. In practice, however, the method will break down even when $p^T q_k$ is not zero but very small. In fact, as $p_k^T q_k$ tends to zero, the stepsize α_k becomes arbitrarily large and the next search direction d_{k+1} gets arbitrarily close to p_k. Following Fletcher's (Ref. 2) approach, this difficulty can be alleviated by combining the two subsequent problematic iterations corresponding to the search directions d_k and d_{k+1} into one planar iteration with respect to the subspace $\langle d_k, d_{k+1}\rangle$. In view of the following proposition, however, this is equivalent to performing a planar iteration over the subspace $\langle p_k, H_k q_k\rangle$.

Proposition 3.1. Let H_k, p_k, q_k, d_k, g_k be defined as in Algorithm 1.1, and assume

$$p_k^T q_k \neq 0.$$

Then, assuming $\alpha_k \neq 0$, there holds

$$\langle d_k, d_{k+1}\rangle = \langle p_k, H_k q_k\rangle.$$

Proof. Without loss of generality, we may assume $k=0$. Then, since

$$p_k^T q_k \neq 0,$$

the results of Theorem 1.1 hold for $k=0$ and $k=1$, regardless of whether F is definite or not. By (13), we have

$$\omega_k p_k = H_{k+1} q_k = H_{k+1} g_{k+1} - H_{k+1} g_k \in \langle d_{k+1}, H_{k+1} g_k \rangle. \tag{30}$$

Hence,

$$-d_{k+1} = H_{k+1} g_{k+1} \in \langle p_k, H_{k+1} g_k \rangle. \tag{31}$$

But, by (8) and (9),

$$H_{k+1} g_k \in \langle H_k g_k, p_k, H_k q_k \rangle = \langle p_k, H_k q_k \rangle, \tag{32}$$

implying that

$$\langle d_k, d_{k+1} \rangle \subseteq \langle p_k, H_k q_k \rangle. \tag{33}$$

Since we assumed

$$d_k^T F d_k \neq 0$$

and, by (10),

$$d_k^T F d_{k+1} = 0,$$

the vectors d_k and d_{k+1} are linearly independent, implying identity of the subspaces in (33). □

The basic idea underlying the proposed algorithm is to perform a *planar iteration* with respect to the subspace $\langle p_k, H_k q_k \rangle$ whenever a singular or near-singular direction d_k is encountered. In this planar iteration, the point x_{k+2} and the inverse Hessian approximation H_{k+2} are generated directly from x_k and H_k, bypassing the problematic computation of x_{k+1} and H_{k+1}. Specifically, the point x_{k+2} is selected to be a stationary point of the objective function over the linear manifold $x_k + \langle p_k, H_k q_k \rangle$, i.e.,

$$g_{k+2} \perp \langle p_k, H_k q_k \rangle.$$

Defining

$$p_{k+1} = x_{k+2} - x_k \quad \text{and} \quad q_{k+1} = g_{k+2} - g_k,$$

we then update H_k so that the subspace $\langle p_{k+1}, p_k \rangle$ becomes an eigenspace of $H_{k+2} F$. The planar update is analogous to the linear update in Algorithm 1.1 being based on the vectors p_{k+1}, p_k and their respective approximations $H_k q_{k+1}$ and $H_k q_k$. This update is performed via a rank-three modification of H_k, employing the three vectors $\{H_k q_{k+1}, p_{k+1}, p_k\}$ that span

$\langle H_k q_{k+1}, p_{k+1}, H_k q_k, p_k \rangle$. It should be emphasized that, by performing such a planar iteration, we eliminate the need to search along d_k, so that the stepsize α_k used to determine p_k and q_k may be any arbitrary nonzero scalar.

The main virtue of the above planar iteration is that it is well defined even when $p_k^T q_k$ approaches zero, causing d_k and d_{k+1} to become collinear and the computation of x_{k+1} and H_{k+1} to break down. Furthermore, like Fletcher's (Ref. 2) method, this planar iteration reduces, in the limit, when $p_k^T q_k$ is exactly zero, to a hyperbolic-pairs-type method. This follows from the fact that, when p_k is singular, and

$$q_k^T H_k q_k \neq 0,$$

then the vectors p_k and

$$\hat{p}_k = H_k q_k - (q_k^T H_k F H_k q_k / 2 q_k^T H_k q_k) p_k$$

form a hyperbolic pair spanning $\langle p_k, H_k q_k \rangle$. The justification for the latter assumption and the precise relationship between our method and the algorithms proposed in Refs. 2 and 3 will be addressed later in the paper.

Implementing the planar iteration outlined above hinges, of course, on the existence of the stationary point x_{k+2} and the linear independence of the vectors p_{k+1} and p_k. Furthermore, as we shall see later, implementing the planar update will require that the matrix $[q_{k+1}, q_k]^T [p_{k+1}, p_k]$ be nonsingular. In the following proposition, we establish conditions that will ensure satisfaction of these requirements for the quadratic case. For notational convenience, we omit the subscripts k in the remainder of this section using * and ** to replace the subscripts $k+1$ and $k+2$.

Proposition 3.2. Let F and H be symmetric, nonsingular $n \times n$ matrices and let $g, p, q \in E^n$ be nonzero vectors such that

$$p = -\alpha H g, \qquad \text{for some scalar } \alpha > 0,$$

$$q = Fp,$$

and

$$(p^T q)^2 \neq -\alpha (p^T g)(q^T H q). \tag{34}$$

Then,

$$(q^T H F H q)(p^T q) \neq (q^T H q)^2, \tag{35}$$

if and only if there exist a nonzero vector $p^* \in \langle p, Hq \rangle$ which is linearly independent of p and satisfies the relation

$$g + F p^* \perp \langle p, Hq \rangle.$$

Furthermore, if such a vector exists, then the matrix $[p^*, p]^T F[p^*, p]$ is nonsingular.

Proof. Let

$$p^* = \xi p + \zeta Hq. \tag{36}$$

Then, the orthogonality condition implies that ξ and ζ satisfy the system of equations

$$\begin{bmatrix} p^T q & q^T Hq \\ q^T Hq & q^T HFHq \end{bmatrix} \begin{bmatrix} \xi \\ \zeta \end{bmatrix} = \begin{bmatrix} -p^T g \\ -q^T Hg \end{bmatrix}. \tag{37}$$

Since

$$q^T Hg = -(1/\alpha) p^T q,$$

condition (34) implies that the first column on the left of (37) and the right-hand side of that system are linearly independent. Thus, a solution to (37) must have $\zeta \neq 0$. Furthermore, such a solution exists if and only if the determinant of the 2×2 matrix in (37) does not vanish, i.e., (35) holds. If such a solution exists, then we can also write

$$[p^*, p]^T = \begin{bmatrix} \zeta & \xi \\ 0 & 1 \end{bmatrix} [Hq, p]^T, \tag{38}$$

so that, by (35) and (38),

$$\det\{[p^*, p]^T F[p^*, p]\} = \zeta^2[(q^T HFHq)(p^T q) - (q^T Hq)^2] \neq 0. \tag{39}$$

□

Assuming that Hq and p are linearly independent, one could easily show that, for positive-definite H and F, (34) and (35) follow from the Cauchy–Schwartz inequality. In general, however, if F and H might be indefinite, (34) and (35) could be violated, even when Hq and p are linearly independent. This will not happen when p_k is singular, since, as will be shown later, we can assume (in the quadratic case) that

$$(p^T g)(q^T Hq) \neq 0, \qquad \text{whenever } p^T q = 0.$$

However, when applying planar iterations in situation where the search direction d_k is nearly-singular, one may encounter difficulties unless near-singularity is properly characterized. In particular, the planar iteration will break down if we apply it on an iteration k at which Algorithm 1.1 would generate a singular subsequent search direction d_{k+1}. To clarify this statement, suppose that

$$p_k^T q_k \neq 0,$$

but the vector

$$v = (q_k^T H_k q_k / p_k^T q_k) p_k - H_k q_k$$

is singular, i.e.,

$$v^T F v = 0.$$

Clearly,

$$v^T q = 0,$$

therefore, in view of Proposition 3.2, v is parallel to the following search direction d_{k+1} generated by Algorithm 1.1. It can be shown that, in this case, either there is no stationary point over the linear manifold $x_k + \langle p_k, H_k q_k \rangle$ or, if there is one, it is along the line $x_k + \langle p_k \rangle$. This observation is consistent with Proposition 3.2, since the singularity of v implies violation of (35). Clearly, in such a case, the planar iteration will break down and its should, therefore, not be used, regardless of how small is $p_k^T q_k$. We again emphasize, however, that the above cannot occur if

$$p_k^T q_k = 0.$$

In view of Proposition 3.2, we could use (34) and (35) as a test to assure that the planar iteration is well defined. This would detect, for instance, problematic situations like the one described above. Unfortunately, computing the quantity $q_k^T H_k F H_k q_k$ used in (35) is costly. It would either require explicit use of F (which is typically unavailable) or approximating $F H_k q_k$ based on the difference between g_k and an additional gradient evaluated at some point along the line $x_k + \langle H_k q_k \rangle$. In the following proposition, we introduce a criterion which is based only on available information and, if satisfied, it guarantees (34) and (35).

Proposition 3.3. Let p, q, g, F, H, α be as in Proposition 3.2, and let

$$\sigma = (|q^T H q| / \|q\| \, \|Hq\|)\{\min[|\alpha p^T g|, |q^T H q|]\}. \tag{40}$$

Assuming that

$$(p^T g)(q^T H q) \neq 0,$$

then, conditions (34) and (35) are satisfied if

$$|p^T q| \leq \sigma \epsilon \quad \text{and} \quad \epsilon < \min[1, 1/\|FH\|].$$

Proof. By the Cauchy–Schwartz inequality,

$$|q^T H q| / \|Hq\| \, \|q\| \leq 1.$$

Thus, the given conditions imply that

$$|p^T q| < \min[|\alpha p^T g|, |q^T Hq|],$$

which in turn implies (34). Similarly,

$$|q^T HFHq| \, |p^T q| \leq \|Hq\| \, \|q\| \, \|FH\| \sigma\epsilon < (q^T Hq)^2,$$

implying (35). □

The assumption that

$$(p^T g)(q^T Hq) \neq 0$$

will be justified in the context of a generalization of Algorithm 1.1 which incorporates planar iterations. Specifically, it will be shown that, for a quadratic case and a positive-definite initial matrix H_0, $(p^T g)$ and $(q^T Hq)$ do not vanish when

$$|p^T q| \leq \sigma\epsilon.$$

The criterion defined by Proposition 3.3 may be regarded as a relative measure of near-singularity of the search direction. If we assume that $\|FH\|$ is uniformly bounded above, then Proposition 3.3 guarantees the existence of a sufficiently small fixed ϵ, so that the planar iteration is well defined whenever

$$|p^T q| \leq \sigma\epsilon.$$

4. Planar Rank-Three Updates

In this section, we develop rank-three updates of the form described in the preceding section, to be used in planar quasi-Newton iterations. We shall first present a general class of such updates and then highlight some special cases of interest. For ease of later reference, we shall again express the results in terms of the matrix H and the vectors p, p^*, q, q^*. It should be pointed out, however, that no assumption is being made here as to the source of these quantities or the relations among them, other than those that are explicitly stated in the theorem. In particular, we like to emphasize that the following results do not depend on the objective function being quadratic or on $x + p^*$ being a stationary point over the linear manifold $x + \langle Hq, p\rangle$.

Theorem 4.1. Let H be a symmetric $n \times n$ matrix, and let

$$P = [p^*, p], \qquad Q = [q^*, q]$$

be $n\times 2$ matrices such that Q^TP is nonsingular and the subspace $\langle p^*, p\rangle$ is identical to $\langle Hq, p\rangle$. Then, a matrix H^{**} will satisfy

$$H^{**}=H+[Hq^*, p^*, p]\Phi[Hq^*, p^*, p]^T, \tag{41}$$

$$H^{**}Q=P(Q^TP)^{-1}\Psi, \tag{42}$$

for some symmetric 3×3 matrix Φ and 2×2 matrix Ψ, if

$$H^{**}=H^{**}(\varphi)=H+P(Q^TP)^{-1}[v, 0]^T+[v, 0](P^TQ)^{-1}P^T + P\Gamma P^T+\varphi vv^T, \tag{43}$$

for some scalar φ, where

$$v=P(Q^TP)^{-1}Q^THq^*-Hq^*, \tag{44}$$

$$\Gamma=(Q^TP)^{-1}(\Psi-Q^THQ)(P^TQ)^{-1}. \tag{45}$$

Furthermore, if the vectors $\{Hq^*, p^*, p\}$ are linearly independent, then, for given Ψ and Φ_{11} (the upper left element of Φ), (43), (44), (45), with $\varphi=\Phi_{11}$, define the unique H^{**} satisfying (41) and (42).

Proof. Clearly, the update defined by (43) and (44) is of the general form specified by (41). To prove that (42) is also satisfied, we first note that

$$Q^Tv=Q^TP(Q^TP)^{-1}Q^THq^*-Q^THq^*=O. \tag{46}$$

Thus, by (43), (44), (45),

$$\begin{aligned} H^{**}Q&=HQ+[v, O]+P\Gamma(P^TQ)\\ &=[P(Q^TP)^{-1}Q^THq^*, Hg]-P(Q^TP)^{-1}(Q^THQ)+P(Q^TP)^{-1}\Psi. \end{aligned} \tag{47}$$

We note, however, that $P(Q^TP)^{-1}Q^T$ is a projection operator projecting onto the subspace $\langle p^*, p\rangle$ and along the orthogonal complement of $\langle q^*, q\rangle$. Thus, since

$$Hq\in\langle p^*, p\rangle,$$

there holds that

$$Hq=P(Q^TP)^{-1}Q^{-1}Hg. \tag{48}$$

Consequently,

$$[P(Q^TP)^{-1}Q^THq^*, Hq]=P(Q^TP)^{-1}Q^THQ, \tag{49}$$

so that (47) reduces to (42).

To prove the second part of the theorem, we show by construction that, given Ψ and Φ_{11}, Eqs. (41) and (42) define a unique H^{**} depending linearly on the scalar parameter Φ_{11}. First, by substituting (41) into (42), using (48), and partitioning the matrix Φ, we obtain

$$[Hq^*, p^*, p]W = 0, \tag{50}$$

where

$$W = \left[\begin{array}{c|c} 1 & O \\ \hline O & (Q^TP)^{-1}Q^THq \end{array}\right] + \left[\begin{array}{c|c} \Phi_{11} & \Phi_{12}^T \\ \hline \Phi_{12} & \Phi_{22} \end{array}\right] \left[\begin{array}{c} q^{*T}HQ \\ \hline P^TQ \end{array}\right] - \left[\begin{array}{c} O \\ \hline (Q^TP)^{-1}\Psi \end{array}\right]. \tag{51}$$

Since $[Hq^*, p^*, p]$ is of full rank, (50) is satisfied if and only if

$$W = O.$$

Thus, by (51), we obtain

$$[1, O] + \Phi_{11}q^{*T}HQ + \Phi_{12}^T(P^TQ) = O, \tag{52}$$

$$[O, (Q^TP)^{-1}Q^THq] + \Phi_{12}q^{*T}H[q^*, q] + \Phi_{22}(P^TQ) - (Q^TP)^{-1}\Psi = O. \tag{53}$$

Since (P^TQ) is nonsingular, Eqs. (52) and (53) have a unique solution $\{\Phi_{12}, \Phi_{22}\}$ depending linearly on the parameters Φ_{11} and Ψ. Thus, since H^{**} in (41) is linear in Φ, it follows that, for given Ψ, (41) and (42) define a unique class of matrices $H^{**}(\Phi_{11})$ depending linearly on the scalar parameter Φ_{11}. On the other hand, we have shown that the class $H^{**}(\varphi)$, defined by (43), (44), (45), also satisfies (50) and (42). Hence, the two classes are identical. Furthermore, since

$$Hq^* \notin \langle p^*, p\rangle,$$

while the coefficient of the term Hq^*q^TH in (43) and (44) is φ, it also follows that

$$\varphi = \Phi_{11}. \qquad \square$$

A more explicit representation of the general rank-three update defined by (43) and (44) is given in the following lemma.

Lemma 4.1. The rank-three update defined by (43) and (44) is equivalent to

$$H^{**} = H + a(p^Tq)(p^*v^T + vp^{*T}) - a(p^{*T}q)(pv^T + vp^T) + [p^*, p]\Gamma[p^*, p]^T + \varphi vv^T, \tag{54}$$

where

$$v = bp^* + cp - Hq^*, \tag{55}$$

$$a = 1/\{(q^{*T}p^*)(q^Tp) - (q^{*T}p)(p^{*T}q)\}, \tag{56}$$

$$b = a\{(q^Tp)(q^{*T}Hq^*) - (q^{*T}p)(q^THq^*)\}, \tag{57}$$

$$c = a\{(q^{*T}p^*)(q^THq^*) - (p^{*T}q)(q^{*T}Hq^*)\}. \tag{58}$$

Proof. The result is obtained by substituting, in (43) and (44),

$$Q = [q^*, q], \qquad P = [p^*, p],$$

and the relation

$$(Q^TP)^{-1} = a\begin{bmatrix} q^Tp & -q^{*T}p \\ -p^{*T}q & q^{*T}p^* \end{bmatrix}, \tag{59}$$

where

$$a = 1/\det(Q^TP). \qquad \square$$

Further simplification of the above class of rank-three updates may be obtained by assuming specific relationships among the vectors p^*, p, q^*, q. The assumptions in the following corollary correspond to the case where these vectors are generated by the planar iteration outlined in the previous section, while the search direction p is singular.

Corollary 4.1. Let H, p^*, p, q^*, q be as in Theorem 4.1. In addition, assume that

$$p = -\alpha Hg, \qquad \text{for some scalar } \alpha > 0,$$

and vector g such that

$$[p^*, p]^T(g + q^*) = 0 \quad \text{and} \quad q^Tp = 0.$$

Then, the update defined by (43) and (44) is equivalent to

$$H^{**} = H + (1/q^{*T}p)(pv^T + vp^T) + [p^*, p]\Gamma[p^*, p]^T + \varphi vv^T, \tag{60}$$

where

$$v = (q^{*T}Hg^*/q^{*T}p)p - Hq^*. \tag{61}$$

Proof. Since

$$Hq \in \langle p^*, p\rangle,$$

we have

$$qH^T(g + q^*) = 0. \tag{62}$$

Thus,

$$q^T H q^* = -q^T H g = (1/\alpha) q^T p = 0. \tag{63}$$

Also since the matrix $Q^T P$ is assumed to be nonsingular and

$$p^T q = 0,$$

the terms $p^{*T} q$ and $q^{*T} p$ must not vanish. Substituting the above into (56), (57), (58) results in

$$a = -1/\{(q^{*T} p)(p^{*T} q)\}, \qquad b = 0, \qquad c = (q^{*T} H q^*)/(q^{*T} p),$$

which lead to (60) and (61). □

If the vectors q^*, q, p^*, p are generated by a planar iteration and the objective function is quadratic with Hessian F, then

$$[q^*, q] = F[p^*, p].$$

In that case, (42) implies that the subspace $\langle p^*, p\rangle$ is an eigenspace of $H^{**}F$. Of particular interest is the case where each of the vectors p^* and p is an eigenvector of $H^{**}F$, which corresponds to the matrix $(Q^T P)^{-1}\Psi$ in (42) being diagonal. The matrix H^{**} will then satisfy the relations

$$H^{**} q^* = \omega^* p^*, \tag{64}$$

$$H^{**} q = \omega p. \tag{65}$$

Premultiplying (64) and (65) by q^T and q^{*T} reveals that the eigenvalues ω and ω^* are related through

$$\omega^*(p^{*T} q) = \omega(p^T q^*). \tag{66}$$

It thus follows that, for the quadratic case, if p^* and p become eigenvalues of $H^{**}F$, they must be either conjugate with respect to the Hessian F or they must have the same corresponding eigenvalues.

To obtain the subclass of rank-three updates that satisfy (64) and (65), we have to calculate the parameter matrix Γ, appearing in (43), by substituting into (45) the relation

$$(Q^T P)^{-1}\Psi = (\omega/p^{*T} q)\,\mathrm{diag}(p^T q^*, p^{*T} q). \tag{67}$$

Clearly, the only free parameter remaining in the matrix Γ that results from the above substitution is the eigenvalue ω. Thus, the class of rank-three updates satisfying (64) and (65) is a two-parameter class like Huang's symmetric rank-two family. As we shall see below, the updates described above will again reduce to a relatively simple form for the special case considered in Corollary 4.1.

Corollary 4.2. Let H, p^*, p, q^*, q be as in Corollary 4.1. Then, the update defined by (60) and (61), with Γ given by

$$\Gamma=\{\tau/\{(p^{*T}p)(q^{*T}p)\}\}\begin{bmatrix}0 & q^{*T}p\\ q^{*T}p & p^{*T}q^*\end{bmatrix}-\{1/(p^{*T}q)^2\}$$

$$\times\begin{bmatrix}q^THq & 0\\ 0 & q^{*T}Hq^*\end{bmatrix}, \tag{68}$$

$$\tau=\omega+\{(q^{*T}p^*)(q^THq)/\{(q^{*T}p)(p^{*T}q)\}\}, \tag{69}$$

for some scalar ω will satisfy

$$H^{**}q^*=\omega(p^Tq^*/p^{*T}q)p^*, \tag{70}$$

$$H^{**}q=\omega p. \tag{71}$$

Proof. The proof is obtained by substituting (59), (63), (66), (67) into (45). □

For the case

$$[q^*, q]=F[p^*, p],$$

(e.g., quadratic objective function), Eqs. (68) and (69) may be simplified somewhat more, due to the fact that

$$p^Tq^*=p^{*T}q.$$

5. Planar Quasi-Newton Algorithm

The results and concepts presented above will be incorporated now in a conceptual iterative procedure that extends Huang's symmetric class by performing a planar iteration whenever a near-singular search direction is encountered. The algorithm will be described in terms of quantities involving only gradient information, so that it can be implemented without explicit use of the Hessian. This is particularly important for potential extension to nonquadratic problems.

Algorithm 5.1. Start with an arbitrary point x_0, an $n\times n$ symmetric positive-definite matrix H_0, and the gradient

$$g_0=\nabla f(x_0)^T.$$

Step 1. (*i*) If $g_k\neq 0$, then obtain

$$d_k=-H_kg_k. \tag{72}$$

(*ii*) Select some arbitrary stepsize $\alpha_k > 0$, then set

$$p_k = \alpha_k d_k,$$

and obtain

$$g_{k+1} = \nabla f(x_k + p_k)^T, \tag{73}$$

$$q_k = g_{k+1} - g_k, \tag{74}$$

$$\sigma_k = (|q_k^T H_k q_k| / \|q_k\| \, \|H_k q_k\|)\{\min[|\alpha_k p_k^T g_k|, |q_k^T H_k q_k|]\}. \tag{75}$$

(*iii*) For some predetermined small $\varepsilon \geq 0$, if

$$|p_k^T q_k| > \varepsilon \sigma_k$$

[Case A], then execute Steps 2 to 5 of Algorithm 1.1, resetting p_k, q_{k+1}, q_k. Otherwise [Case B], execute a planar iteration as follows.

Step 2. Obtain

$$x_{k+2} \in x_k + \langle p_k, H_k q_k \rangle, \tag{76}$$

such that

$$\nabla f(x_{k+2}) \perp \langle p_k, H_k q_k \rangle. \tag{77}$$

Step 3. Set

$$p_{k+1} = x_{k+2} - x_k, \tag{78}$$

$$g_{k+2} = \nabla f(x_{k+2})^T, \tag{79}$$

$$q_{k+1} = g_{k+2} - g_k. \tag{80}$$

Step 4. Set

$$P_k = [p_{k+1}, p_k], \qquad Q_k = [q_{k+1}, q_k],$$

and select an arbitrary symmetric 2×2 matrix Γ_k and some scalar

$$\varphi_{k+1} \neq -1/g_{k+2} H_k g_{k+2}.$$

Then, obtain H_{k+2} using the rank-three update

$$H_{k+2} = H_k + P_k R_k [v_k, O]^T + [v_k, O] R_k^T P_k^T + P_k \Gamma_k P_k^T + \varphi_{k+1} v_k v_k^T, \tag{81}$$

where

$$v_k = P_k R_k Q_k^T H_k q_{k+1} - H_k q_{k+1}, \tag{82}$$

$$R_k = (Q_k^T P_k)^{-1}. \tag{83}$$

Step 5. Increment k by 2, and return to Step 1.

Step 2 in Case B of Algorithm 5.1 above calls for the location of a stationary point on a two-dimensional linear manifold. Similar requirements are present in Fletcher's (Ref. 2) algorithm and in Luenberger's (Ref. 3) method. For the quadratic case, this step can be performed analytically leading to a closed-form expression for p_{k+1} as a linear combination of p_k and $H_k q_k$. In the following proposition, we present such an expression for the general case, where p_k is near singular, and for the special case where

$$p_k^T q_k = 0.$$

These expressions are equivalent to the ones obtained in Refs. 2 and 3. Again, for convenience, we suppress the subscript k and replace $k+1$ and $k+2$ by * and **, respectively.

Proposition 5.1. Let H, p, p^*, q, g, α be as in Case B of Algorithm 5.1, for a quadratic objective function having a nonsingular Hessian F. Then, assuming that (34) and (35) hold and that

$$q^T H q \neq 0,$$

we have

$$p^* = \{1/[\alpha(1-rs)]\}\{(r+ts)p - (t+r^2)Hq\}, \tag{84}$$

where

$$r = (p^T q)/(q^T H q), \tag{85}$$

$$s = (q^T HFHq)/(q^T Hq), \tag{86}$$

$$t = (\alpha p^T g)/(q^T H q). \tag{87}$$

If in addition

$$p^T q = 0,$$

then

$$p^* = (p^T g / q^T Hq)\{(q^T HFHq / q^T Hq)p - Hq\}. \tag{88}$$

Proof. The results follow from (36), (37), and the fact that

$$p = -\alpha H g. \qquad \square$$

Evaluating (84) or (88) requires the computation of $q^T HFHq$. To avoid the explicit use of F, one can obtain the vector FHq by using another gradient along the line $x_k + \beta H_k q_k$ and obtaining (exactly in the quadratic case and approximately for a nonquadratic function)

$$FH_k q_k = (1/\beta)\{\nabla f(x_k + \beta H_k q_k) - \nabla f(x_k)\}, \tag{89}$$

for some $\beta \neq 0$. For the special case where

$$p^T q = 0,$$

the following proposition suggests an alternative way for obtaining p^*.

Proposition 5.2. Let H, F, p, p^*, q, g be as in Proposition 5.1, and assume

$$p^T q = 0.$$

Then,

$$p^* = \beta^* Hq + \alpha^* p, \tag{90}$$

where

$$\beta^* = \arg\min_\beta \{[\nabla f(x + \beta Hq)]p\}^2, \tag{91}$$

$$\alpha^* = \arg\min_\alpha \{[\nabla f(x + \beta^* Hq + \alpha p)]Hq\}^2. \tag{92}$$

Proof. Since f is quadratic with Hessian F and

$$g = \nabla f(x)^T,$$

(91) implies that

$$[g + \beta^* FHg]^T p = 0. \tag{93}$$

Using the fact that

$$q = Fp,$$

this yields

$$\beta^* = -(p^T g)/(q^T Hq). \tag{94}$$

Similarly, (92) implies that

$$[g + \beta^* FHq + \alpha^* Fp]^T Hq = 0. \tag{95}$$

But Hg is parallel to p. Hence,

$$g^T Hq = 0;$$

consequently, by (95),

$$\alpha^* = -\beta^* (q^T HFHq)/(q^T Hq). \tag{96}$$

Substituting (94) and (96) into (90) results in (88). □

In view of Proposition 5.2, if p_k is singular, the point x_{k+2} in Case B of Algorithm 5.1 can be obtained by performing two line searches, as

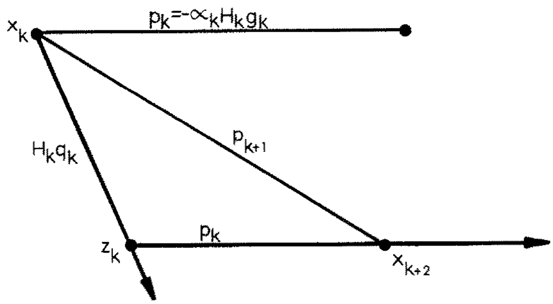

Fig. 1. Two-step implementation of the planar search.

illustrated in Fig. 1. First, the point z_k is obtained by minimizing, along the line $x_k+\beta H_k q_k$, the squared inner product of the gradient $\nabla f(z_k)$ with the vector p_k. Then, the point x_{k+2} is obtained by minimizing, along the line $z_k+\alpha p_k$, the squared inner product of the gradient $\nabla f(x_{k+1})$ with the vector $H_k q_k$.

Although the result of Proposition 5.2 depends on p_k being exactly singular, the above two-step implementation of the planar search might be a reasonable approximation even when p_k is near-singular. In fact, one could show that the norm of the error in p_{k+1} introduced by using such an approximate plane-search procedure is of the order of the parameter ϵ in Algorithm 5.1. The most attractive feature of the above two-step implementation of the planar search is its easy extension to nonquadratic problems. We will further elaborate on this issue when we discuss such extensions.

6. Quadratic Termination and Conjugacy Properties of Planar Quasi-Newton Algorithms

In the following theorem, we prove the basic conjugacy property of the search directions and search planes generated by Algorithm 5.1 when applied to a quadratic objective function. These results are proved by induction, simultaneously with other auxiliary results used in the proof.

Theorem 6.1. Let the sequences $\{x_k\}$ $\{g_k\}$, $\{d_k\}$, $\{p_k\}$, $\{q_k\}$, $\{H_k\}$, $\{P_k\}$, $\{Q_k\}$, $\{R_k\}$, $\{\alpha_k\}$, $\{\sigma_k\}$, $\{\varphi_k\}$, $\{\gamma_k\}$, $\{\Gamma_k\}$ be defined by Algorithms 1.1 and 5.1 for a quadratic objective function having a nonsingular Hessian F, and assume

$$\epsilon<\min[1, 1/\|FH_k\|], \qquad \text{for all } k.$$

Let $\mathscr{I}$ be the set of indices corresponding to the search directions activating Case B of the algorithm, i.e.,

$$\mathscr{I} = \{i \| p_i^T q_i| \leq \epsilon\sigma_i, g_i \neq 0, p_i \neq 0\},$$

and

$$\mathscr{J} = \{j | j-1 \in \mathscr{I}\}.$$

Then, the following results hold for $k = 0, 1, \ldots, n-1$.

(a) $p_j^T g_j = p_i^T F p_j = 0$, for $0 \leq i < j \leq k$, if $j \notin \mathscr{J}$, and for $0 \leq i < j-1 \leq k-1$, if $j \in \mathscr{J}$. (97)

(b) For any $u \in E^n$,

$$(H_j - H_0)u \in \langle p_0, \ldots, p_j \rangle, \qquad \text{for } j \in \{j | 0 \leq j \leq k, j \notin \mathscr{J}\}. \tag{98}$$

(c) $g_k^T H_j g_k = g_k^T H_0 g_k$, for $j \in \{j | 0 \leq j \leq k-1, j \notin \mathscr{J}\}$; (99)

and, if $k \notin \mathscr{J}$, then

$$g_k^T H_k g_k = (1 + \varphi_{k-1} g_k^T H_0 g_k) g_k^T H_0 g_k. \tag{100}$$

(d) If $k \notin \mathscr{J}$ and $g_k \neq 0$, then

$$g_k^T p_k \neq 0; \tag{101}$$

and, if $k \in \mathscr{J}$, then

$$q_{k-1}^T H_{k-1} q_{k-1} \neq 0. \tag{102}$$

(e) If $k \notin \mathscr{I}$, then

$$H_{k+1} F p_i = \omega_i p_i, \qquad \text{for } i \in \{i | 0 \leq i \leq k,\ i \notin \mathscr{I},\ i \notin \mathscr{J}\}, \tag{103}$$

$$\omega_i = \gamma_i + (q_i^T H_i q_i / p_i^T q_i), \tag{104}$$

and

$$H_{k+1} F [p_{i+1}, p_i] = [p_{i+1}, p_i] \Omega_i, \qquad \text{for } i \in \{i | 0 \leq i \leq k-1, i \in \mathscr{I}\}, \tag{105}$$

$$\Omega_i = \Gamma_i R_i^{-1} + R_i Q_i H_i Q_i. \tag{106}$$

Proof. We prove the parts of the theorem simultaneously by induction. First, we show that the theorem holds for $k = 0$ and, if $0 \in \mathscr{I}$, then it holds for $k \leq 1$.

The above is clearly true for parts (a), (b), (c). Furthermore, by the positive definiteness of H_0, it also holds for (d). Finally, if $0 \notin \mathscr{I}$, then part (e) holds for $k = 0$ by Theorem 1.1, while, if $0 \in \mathscr{I}$, then it holds for $k \leq 1$ by Theorem 4.1.

We shall assume now by induction that the theorem holds up to k and further assume that $k \notin \mathcal{I}$. Then, we complete the proof by showing that the theorem holds for $k+1$ if $k+1 \notin \mathcal{I}$ and it holds for $k+2$ if $k+1 \in \mathcal{I}$.

(a) Clearly, for a quadratic objective function with Hessian F,

$$q_i = FP_i, \qquad \text{for all } i.$$

Thus,

$$g_{k+1} = g_{i+1} + \sum_{\substack{j=i+1 \\ j \notin \mathcal{I}}}^{k} Fp, \qquad \text{for } i \leq k-1, i \notin \mathcal{I}. \tag{107}$$

Therefore, by (4), (77), (97),

$$p_i^T g_{k+1} = p_i^T g_{i+1} = 0, \qquad \text{for } i \in \{i | 0 \leq i \leq k, \ i \notin \mathcal{I}\}. \tag{108}$$

Since $i \in \mathcal{I}$ implies $i+1 \notin \mathcal{I}$, we may rewrite (105) for $i \in \mathcal{I}$ by replacing i with $i+1$. Consequently, by (76) through (80) and (97),

$$p_i^T g_{k+1} = p_i^T g_{i+2} = 0, \qquad \text{for } i \in \{0 \leq i \leq k-1, i \notin \mathcal{I}\}. \tag{109}$$

Since $k \notin \mathcal{I}$, we may combine (108) and (109) obtaining

$$p_i^T g_{k+1} = 0, \qquad \text{for } 0 \leq i \leq k. \tag{110}$$

Again, since $k \notin \mathcal{I}$, then by (72)

$$p_i^T F d_{k+1} = -p_i^T F H_{k+1} g_{k+1}, \qquad \text{for all } i. \tag{111}$$

Hence, by (103), (110), (111),

$$p_i^T F p_{k+1} = -\alpha_{k+1} \omega_i p_i^T g_{k+1} = 0, \qquad \text{for } i \in \{i | 0 \leq i \leq k, \ i \notin \mathcal{I}, i \notin \mathcal{J}\}, \tag{112}$$

while, by (105), (111), (112),

$$\begin{aligned}[p_{i+1}, p_i]^T F p_{k+1} &= -\alpha_{k+1} [p_{i+1}, p_i]^T F H_{k+1} g_{k+1} \\ &= -\alpha_{k+1} \Omega_i^T [p_{i+1}, p_i]^T g_{k+1} = 0, \\ &\quad \text{for } i \in \{i | 0 \leq i \leq k-1, i \in \mathcal{I}\}.\end{aligned} \tag{113}$$

Since $k \notin \mathcal{I}$, we may again combine (113) and (112) obtaining

$$p_i^T F p_{k+1} = 0, \qquad \text{for } 0 < i \leq k. \tag{114}$$

Equations (110) and (114) prove (a) for $k+1$ in both cases $k+1 \in \mathcal{I}$ and $k+1 \notin \mathcal{I}$. For the case $k+1 \in \mathcal{I}$, we also have, by (107), (110), (114),

$$p_i^T g_{k+2} = p_i^T (g_{k+1} + F p_{k+1}) = 0, \qquad \text{for } 0 \leq i \leq k \text{ and } k+1 \in \mathcal{I}. \tag{115}$$

Furthermore, from (76) and (77), there exist scalars ξ and η (depending on k) such that

$$p_i^T F p_{k+2} = \xi p_i^T F H_{k+1} F p_{k+1} + \eta p_i^T F p_{k+1}), \qquad \text{for } 0 \le i \le k. \tag{116}$$

From (110), (105), and the assumption that $k \notin \mathcal{I}$, it follows, however, that

$$H_{k+1} F p_i \in \langle p_0, \ldots, p_k \rangle.$$

Thus, by (114) and (116),

$$p_i^T F p_{k+2} = 0, \qquad \text{for } 0 \le i \le k \text{ and } k+1 \in \mathcal{I}. \tag{117}$$

Equations (115) and (117) prove (a) for $k+2$ in the case $k+1 \in \mathcal{I}$.

(b) For the case $k \notin \mathcal{I}$, H_{k+1} is determined by (8) and (9). Thus, using (98), we get

$$(H_{k+1} - H_0) u = (H_{k+1} - H_k) u + (H_k - H_0) u \in \langle H_k q_k, p_0, \ldots, p_k \rangle. \tag{118}$$

Therefore, in view of Proposition 3.1,

$$(H_{k+1} - H_0) u \in \langle p_0, \ldots, p_{k+1} \rangle. \tag{119}$$

For the case $k \in \mathcal{I}$, H_{k+1} is defined by (81) through (83). Using (98), we now have

$$\begin{aligned}(H_{k+1} - H_0) u &= (H_{k+1} - H_{k-1}) u + (H_{k-1} - H_0) u \\ &\in \langle H_{k-1} q_k, p_0, \ldots, p_k \rangle \subseteq \langle H_{k-1} g_{k+1}, p_k, p_{k-1} \rangle.\end{aligned} \tag{120}$$

However, in view of (77) and (72),

$$d_{k+1} = -H_{k+1} g_{k+1} = -(1 + \varphi_k g_{k+1}^T H_{k-1} g_{k+1}) H_{k-1} g_{k+1} + \mu_k p_k + \eta_k p_{k-1}, \tag{121}$$

for some scalars μ_k and η_k. Since the choice of φ_k is restricted so that the coefficient of $H_{k-1} g_{k+1}$ in (121) does not vanish, (121) implies that

$$\langle H_{k-1} g_{k+1}, p_k, p_{k-1} \rangle \subseteq \langle p_{k-1}, p_k, p_{k+1} \rangle; \tag{122}$$

therefore, (120) is equivalent to (119). The above proves (b) for $k+1$. If however $k+1 \in \mathcal{I}$, then (98) does not have to hold for $k+2$, so the proof holds by default for $k+2$ as well.

(c) From (98), we have

$$(H_j - H_0) g_{k+1} \in \langle p_0, \ldots, p_j \rangle, \qquad \text{for } j \in \{j \mid 0 \le j \le k,\ j \notin \mathcal{I}\}. \tag{123}$$

Thus, by (110),

$$g_{k+1}^T (H_j - H_0) g_{k+1} = 0, \qquad \text{for } j \in \{j \mid 0 \le j < k+1,\ j \notin \mathcal{I}\}, \tag{124}$$

proving (99) for $k+1$.

If $k+1 \in \mathcal{I}$, then again, by (98) and (119),

$$(Hj - H_0)g_{k+2} \in \langle p_0, \ldots, p_j \rangle, \qquad \text{for } j \in \{j | 0 \le j \le k+1, j \notin \mathcal{I}\}. \tag{125}$$

Thus, by (115), we get

$$g_{k+2}^T(H_j - H_0)g_{k+2} = 0, \qquad \text{for } j \in \{j | 0 \le j < k+2, j \notin \mathcal{I}\} \text{ and } k+1 \in \mathcal{I}. \tag{126}$$

To prove (100), consider first the case $k \notin \mathcal{I}$. Then, by (8), (9), (2), (4),

$$H_{k+1}g_{k+1} = (1 + \varphi_k g_{k+1}^T H_k g_{k+1})H_k g_{k+1} + \mu_k p_k, \tag{127}$$

for some scalar μ_k. Premultiplying (127) by g_{k+1}^T, while using (110) and (124), yields

$$g_{k+1}^T H_{k+1} g_{k+1} = (1 + \varphi_k g_{k+1}^T H_0 g_{k+1}) g_{k+1}^T H_0 g_{k+1}. \tag{128}$$

Likewise, for the case $k \in \mathcal{I}$, premultiplying (121) by g_{k+1}^T, while using (110) and (125), yields again (128). This completes the proof for $k+1$. If however $k+1 \in \mathcal{I}$, then the proof holds by default for $k+2$ as well.

(d) The restrictions on φ_k in Step 4 of Algorithms 1.1 and 5.1, along with (124), imply that

$$1 + \varphi_k g_{k+1}^T H_0 g_{k+1} \neq 0.$$

Thus, by (72), (128), and the positive definiteness of H_0,

$$g_{k+1}^T d_{k+1} = -g_{k+1}^T H_{k+1} g_{k+1} < 0, \qquad \text{if } g_{k+1} \neq 0. \tag{129}$$

In view of (129),

$$\alpha_{k+1} \neq 0.$$

Therefore, since

$$p_{k+1} = \alpha_{k+1} d_{k+1},$$

(129) proves (102) for $k+1$. Furthermore, if $k+1 \in \mathcal{I}$, then the proof extends by default to $k+2$.

The assumption that $k \notin \mathcal{I}$ also implies $k+1 \notin \mathcal{I}$, so (102) extends by default to $k+1$. If, however, $k+1 \in \mathcal{I}$, then it remains to show that (102) holds for $k+2$. By (98), we have

$$(H_{k+1} - H_0)q_{k+1} \in \langle p_0, \ldots, p_{k+1} \rangle. \tag{130}$$

Now, suppose that

$$q_{k+1}^T H_{k+1} q_{k+1} = 0.$$

Then,

$$\sigma_{k+1} = 0;$$

and, since $k+1 \in \mathscr{I}$, this implies

$$p_{k+1}^T q_{k+1} = 0.$$

Consequently, from (130) and (114), we get

$$q_{k+1}^T H_0 q_{k+1} = q_{k+1}^T H_{k+1} q_{k+1} = 0. \tag{131}$$

But (131) contradicts the fact that H_0 is positive definite and

$$p_{k+1} \neq 0.$$

Hence, (102) holds for $k+2$, when $k+1 \in \mathscr{I}$.

(e) First, we consider the case $k+1 \notin \mathscr{I}$. In this case, H_{k+2} is determined by (8) and (9). From (103) and (105), we have

$$H_{k+1} F p_i \in \langle p_0, \ldots, p_k \rangle. \tag{132}$$

Thus, by (9) and (114), we obtain

$$v_{k+1}^T F p_i = -q_{k+1} H_{k+1} F p_i = 0, \qquad \text{for } 0 \leq i \leq k. \tag{133}$$

Consequently, (8), (114), (133) yield

$$H_{k+2} F p_i = H_{k+1} F p_i, \qquad \text{for } 0 \leq i \leq k, \tag{134}$$

which implies that (103) and (105) also hold with H_{k+1} replaced by H_{k+2}. On the other hand, since by (9)

$$v_{k+1}^T q_{k+1} = 0,$$

(8) yields

$$H_{k+2} F p_{k+1} = H_{k+2} q_{k+1} = H_{k+1} q_{k+1} + v_{k+1} + \gamma_{k+1} p_{k+1} = \omega_{k+1} p_{k+1}, \tag{135}$$

where ω_{k+1} is given by (104) with $i = k+1$. Thus, (134) and (135) prove part (e) for $k+1$ in the case $k+1 \notin \mathscr{I}$.

Next, we consider the case $k+1 \in \mathscr{I}$. Here, H_{k+2} is undefined and H_{k+3} is defined by (81) through (83). The nonsingularity of R_{k+1} and the existence of such an H_{k+3} follows from Propositions 3.2 and 3.3 and from part (d) of this theorem. Since (132) still holds, we obtain, by (117),

$$q_{k+2}^T H_{k+1} F p_i = 0, \qquad \text{for } 0 \leq i \leq k. \tag{136}$$

Consequently, from (81), (82), (114), (136), we get

$$H_{k+3} F p_i = H_{k+1} F p_i, \qquad \text{for } 0 \leq i \leq k. \tag{137}$$

Hence, (103) and (105) hold with H_{k+1} replaced by H_{k+3}, i.e.

$$H_{k+3} F p_i = \omega_i p_i, \qquad \text{for } i \in \{i | 0 \leq i \leq k,\ i \notin \mathscr{I},\ i \notin \bar{\mathscr{I}}\}, \tag{138}$$

$$H_{k+3} F [p_{i+1}, p_i] = [p_{i+1}, p_i] \Omega_i, \qquad \text{for } i \in \{i | 0 \leq i \leq k-1,\ i \in \mathscr{I}\}. \tag{139}$$

Clearly, if $k+1 \in \mathcal{I}$, then (138) extends by default to the range $0 \le i \le k+2$ and, by construction of the rank-three update, (139) extends to $0 \le i \le k+1$. Thus, for $k+1 \in \mathcal{I}$, we have proved that (e) holds up to $k+2$. □

Corollary 6.1. For a nonsingular quadratic objective function, the sequence of points $\{x_k\}$ generated by Algorithm 5.1 is such that every point x_k is the unique stationary point of the objective function over the linear manifold $x_0 + \langle p_0, \ldots, p_{k-1} \rangle$.

Proof. The points x_k are defined by Algorithm 5.1 only for $k-1 \notin \mathcal{I}$, where the index set $\mathcal{I}$ is as in Theorem 6.1. Clearly, all these points are such that

$$x_k \in x_0 + \langle p_0, \ldots, p_{k-1} \rangle.$$

Furthermore, by Theorem 6.1(a), $\nabla f(x_k)$ is orthogonal to the subspace $\langle p_0, \ldots, p_{k-1} \rangle$, which proves the result. □

Corollary 6.2. For a nonsingular quadratic objective function, the sequence of points $\{x_k\}$ generated by Algorithm 5.1 is such that, for all $k = 0, 1, \ldots, n-1$, if

$$\nabla f(x_k) \neq 0,$$

then the vectors $p_0, \ldots, p_k$ are linearly independent. Furthermore, if $p_k \in \mathcal{I}$ (the index sets $\mathcal{I}$ and $\mathcal{J}$ are as in Theorem 6.1), then the vectors $p_0, \ldots, p_{k+1}$ are linearly independent.

Proof. Assume

$$g_i \neq 0, \qquad \text{for } i = 0, \ldots, k,\ i \notin \mathcal{J}.$$

Then, by Theorem 6.1(d), the corresponding vectors p_i are nonzero; and, if $k \in \mathcal{I}$, then p_{k+1} is also nonzero. Let

$$M_k = [p_0, p_1, \ldots, p_k],$$

and assume that $k \notin \mathcal{I}$. Suppose that there exist a vector $a \in E^n$ such that

$$M_k a = 0.$$

Then, the vector a must also satisfy

$$M_k^T F M_k a = 0,$$

where F is the Hessian matrix of the objective function. Let P_i be as in

Theorem 6.1. Then, it follows from Theorem 6.1(a) that

$$\det(M_k^T F M_k) = \prod_{\substack{0\le i\le k\\ i\notin\mathcal{I},\, i\notin\mathcal{J}}} (p_i^T F p_i) \cdot \prod_{\substack{0\le i\le k-1\\ i\in\mathcal{I}}} \{\det(P_i^T F P_i)\}. \tag{140}$$

The first product in (140) does not vanish, by the definition of $\mathcal{I}$ and $\mathcal{J}$, while $\det(P_i^T F P_i)$ does not vanish for $i\in\mathcal{I}$, by Proposition 3.2, Proposition 3.3, and Theorem 6.1(d). Thus, the matrix $M_k^T F M_k$ is nonsingular, which implies that $a=0$, and therefore M_k has full rank. For the case $k\in\mathcal{I}$, Proposition 3.2, Proposition 3.3, and Theorem 6.1(d) guarantee that the right-hand side of (140) does not vanish for $0\le i\le k+1$. Hence, by the same argument given above, M_{k+1} has full rank. □

Corollary 6.3. Algorithm 5.1 will converge to the unique stationary point of a nonsingular n-dimensional quadratic function in at most n steps.

Proof. By Corollary 6.2, either $g_k=0$, for some $k<n$, or the vectors $p_0,\ldots,p_{n-1}$ are linearly independent. If the latter occurs, then

$$g_n=0,$$

by Corollary 4.1. □

Corollary 6.4. The sequences $\{H_k\}, \{p_k\}, \{q_k\}$ generated by Algorithm 5.1 for a nonsingular quadratic objective function are such that, for $k=0,1,\ldots n-1$, either

$$g_{k+1}=0$$

or else: if $k\notin\mathcal{J}$ ($\mathcal{J}$ as in Theorem 6.1), then the vectors H_k, q_k, p_k are linearly independent; and, if $k\in\mathcal{J}$, then

$$H_{k-1}q_k\notin\langle p_k, p_{k-1}\rangle.$$

Proof. Since H_0 is positive definite, then, by Theorem 6.1(c),

$$g_{k+1}^T H_k g_{k+1} = g_{k+1}^T H_0 g_{k+1} > 0, \qquad \text{for } k\notin\mathcal{J}, \tag{141}$$

$$g_{k+1}^T H_{k-1} g_{k+1} = g_{k+1}^T H_0 g_{k+1} > 0, \qquad \text{for } k\in\mathcal{J}. \tag{142}$$

Suppose that

$$k\notin\mathcal{J} \quad\text{and}\quad H_k q_k\in\langle p_k\rangle.$$

Then, by (72) and (36),

$$H_k g_{k+1}\in\langle p_k\rangle.$$

However, since g_{k+1} is orthogonal to p_k, this would imply

$$g_{k+1}^T H_k g_{k+1}=0,$$

contradicting (141). Similarly, if

$$k \in \mathcal{J} \quad \text{and} \quad H_{k-1}q_k \in \langle p_k, p_{k-1}\rangle,$$

then, by (72) and (80),

$$H_{k-1}g_{k+1} \in \langle p_k, p_{k-1}\rangle.$$

However, since g_{k+1} is orthogonal to the subspace $\langle p_k, p_{k-1}\rangle$, this would imply that

$$g_{k+1}^T H_{k-1} g_{k+1} = 0,$$

contradicting (142). □

The results presented in this section characterize the general properties of Algorithm 5.1 which do not depend on the choice of parameters in the updates. In particular, it was shown that Algorithm 5.1 generates search directions or search planes that are mutually conjugate with respect to the Hessian of the objective function. Algorithms of this type are in a sense a generalization of the well-known class of *conjugate-direction algorithms* for function minimization. We note from Corollaries 4.1 and 6.1 that, like the latter, Algorithm 5.1 possesses the quadratic termination property and the points that it generates are the stationary points over the expanding linear manifold spanned by the previous search directions and search planes. To distinguish such generalized methods from conventional conjugate direction algorithms, we shall refer to them as *planar conjugate direction algorithms*.

The results of Theorem 4.1, together with Corollary 6.4, imply the existence and uniqueness of the rank-three update defined by (81) through (83), once the parameters φ_{k+1} and Γ_k have been selected. This result is analogous to that stated in Theorem 1.1(d) for the rank-two update. An important question, yet to be addressed, concerns the effect of specific parameter selections on the algorithm performance. For the quadratic case, this issue is resolved (at least theoretically) in the following section, where it is shown that, under perfect line and plane searches, all the methods represented by Algorithm 5.1 (corresponding to different parameter selections) are equivalent.

7. Relationship to the Conjugate-Gradient Algorithm and Newton's Method

In the next theorem, we show that Algorithm 5.1 satisfies a relation equivalent to (12) with respect to the metric H_0. Thus, by analogy to conventional terminology, we shall refer to it as a *planar scaled conjugate-gradient algorithm*. An immediate consequence of this characteristic is that, for $H_0 = I$, Algorithm 5.1 becomes a *planar conjugate gradient-algorithm*, which is equivalent to the methods described in Refs. 2 and 11. The invariance of the sequence $\{x_k\}$ with respect to the parameters of the updates also follows from this theorem.

Theorem 7.1. Let the sequences $\{d_k\}$, $\{p_k\}$, $\{g_k\}$, $\{H_k\}$ be as in Theorem 6.1. Then, for $k = 0, 1, \ldots, n-1$, there holds that

$$\begin{aligned}\langle p_0, \ldots, p_k\rangle &= \langle H_0 g_0, \ldots, H_0 g_k\rangle \\ &= \langle H_0 g_0, (H_0 F) H_0 g_0, \ldots, (H_0 F)^k H_0 g_0\rangle. \end{aligned} \tag{143}$$

Proof. We prove the result by induction. Clearly, (143) holds for $k = 0$. Furthermore, if $0 \in \mathscr{I}$, where $\mathscr{I}$ is the index set defined in Theorem 6.1, then, by (76) and (78),

$$\langle p_0, p_1\rangle \subseteq \langle p_0, H_0 q_0\rangle \subseteq \langle H_0 g_0, H_0 F H_0 g_0\rangle; \tag{144}$$

and, since p_0 and p_1 are linearly independent, by Corollary 6.2, the three subspaces in (144) must be equal. Thus, if $0 \in \mathscr{I}$, then (143) holds for $k \leq 1$.

As in Theorem 6.1, we assume by induction that (143) holds up to k, where $k \notin \mathscr{I}$. Then, we complete the proof by showing that (143) holds for $k+1$ and, if $k+1 \in \mathscr{I}$, it also holds for $k+2$. By Theorem 6.1(b), we have

$$H_{k+1} g_{k+1} = -d_{k+1} \in \langle p_0, \ldots, p_k, H_0 g_{k+1}\rangle. \tag{145}$$

Thus, from (143) and (145), we get

$$\langle p_0, \ldots, p_{k+1}\rangle \subseteq \langle H_0 g_0, \ldots, H_0 g_{k+1}\rangle. \tag{146}$$

On the other hand, by (143), we also have

$$H_0 g_k \in \langle H_0 g_0, \ldots, (H_0 F)^k H_0 g_0\rangle$$

and, likewise,

$$H_0 q_k = H_0 F p_k \in \langle (H_0 F) H_0 g_0, \ldots, (H_0 F)^{k+1} H_0 g_0\rangle.$$

Therefore,

$$H_0 g_{k+1} = H_0 q_k + H_0 g_k \in \langle H_0 g_0, (H_0 F) H_0 g_0, \ldots, (H_0 F)^{k+1} H_0 g_0\rangle \tag{147}$$

and, by (146) and (143),

$$\langle H_0 g_0, \ldots, H_0 g_{k+1}\rangle \subseteq \langle H_0 g_0, (H_0 F) H_0 g_0, \ldots, (H_0 F)^{k+1} H_0 g_0\rangle. \tag{148}$$

If

$$g_{k+1} = 0,$$

then

$$p_{k+1}=0 \quad \text{and} \quad H_0Fp_k=H_0g_k,$$

so by (143)

$$(H_0F)^{k+1}H_0g_0 \in \langle H_0Fp_0, \ldots, H_0Fp_k\rangle$$
$$\subseteq \langle H_0g_0, (H_0F)H_0g_0, \ldots, (H_0F)^kH_0g_0\rangle. \tag{149}$$

This implies that (143) also holds for $k+1$ since the $(k+1)$th vector in each of the three spanning sets in (143) depends on the previous k vectors that span equal subspaces. On the other, if

$$g_{k+1}\neq 0,$$

then the vectors $p_0, \ldots, p_{k+1}$ are linearly independent by Corollary 6.2, so they span a $(k+2)$-dimensional subspace. But, since the dimension of

$$\langle H_0g_0, \ldots, H_0g_{k+1}\rangle \quad \text{and} \quad \langle H_0g_0, (H_0F)H_0g_0, \ldots, (H_0F)^{k+1}H_0g_0\rangle$$

is at most $k+2$, the inclusion signs in (146) and (148) may be replaced by equalities. The above proves (143) for $k+1$ in general. If however $k+1\in\mathcal{J}$, then, by (76) and (78), we have

$$p_{k+2}\in\langle H_{k+1}q_{k+1}, p_{k+1}\rangle \in \langle p_0, \ldots, p_{k+1}, H_0g_k, H_0g_{k+2}\rangle. \tag{150}$$

Thus, by (147), we get

$$\langle p_0, \ldots, p_{k+2}\rangle \subseteq \langle H_0g_0, \ldots, H_0g_{k+2}\rangle. \tag{151}$$

But, by (146) and (148),

$$H_0g_{k+2}=H_0g_k+H_0Fp_{k+1}\in\langle H_0g_0, (H_0F)H_0g_0, \ldots, (H_0F)^{k+2}H_0g_0\rangle, \tag{152}$$

implying

$$\langle H_0g_0, \ldots, H_0g_{k+2}\rangle \subseteq \langle H_0g_0, H_0FH_0g_0, \ldots, (H_0F)^{k+2}H_0g_0\rangle. \tag{153}$$

Again, since $p_0, \ldots, p_{k+2}$ are linearly independent by Corollary 6.2, the subspace on the left in (151) and (153) has the maximal dimension of the subspaces on the right. Consequently, the three subspaces in (151) and (153) must be identical, proving (143) for $k+2$ when $k+1\in\mathcal{J}$. □

Corollary 7.1. Algorithm 5.1, with $H_0=I$, is a *planar conjugate-gradient algorithm*; i.e., when applied to a quadratic objective function with nonsingular Hessian F, then, for $k=0, 1, \ldots, n-1$, either

$$\nabla f(x_k)=0$$

or the following holds:

(a) $p_j^T F p_i = g_j^T p_i = g_j^T g_i = 0, \qquad \text{for } 0 \le i < j-1 \le k-1;$

and, if p_{j-1} does not trigger a planar iteration ($j \notin \mathscr{I}$,), then also

$$p_j^T F p_{j-1} = g_j^T p_{j-1} = g_j^T g_{j-1} = 0, \qquad \text{for } 0 \le j \le k.$$

(b) $\langle p_0, p_1, \ldots, p_k \rangle = \langle g_0, g_1, \ldots, g_k \rangle = \langle g_0, F g_0, \ldots, F^k g_0 \rangle.$

Proof. The proof follows from Theorems 6.1(a) and 7.1. □

Corollary 7.2. For a quadratic objective function, given fixed initial values x_0, H_0, and $\mathscr{I}$ as in Theorem 6.1, the sequence $\{x_k\}_{k \notin \mathscr{I}}$ generated by Algorithm 5.1 is unique and independent of the choice of the parameters $\{\varphi_k\}$, $\{\gamma_k\}$, $\{\Gamma_k\}$.

Proof. The result follows from Corollary 4.1 and the uniqueness of the subspaces $\langle p_0, \ldots, p_k \rangle$, which is implied by Theorem 7.1. □

The results of Corollaries 7.1 and 7.2 indicate that, for a quadratic objective function, Algorithm 5.1 with $H_0 = I$ is equivalent [i.e., generates the same sequence $\{x_k\}$] to Luenberger's (Ref. 3) and Fletcher's (Ref. 2) generalizations of the conjugate-gradient method. This is true for any values of the parameters used in the updates. It should be noted however that, although the parameters $\{\varphi_k\}$, $\{\gamma_k\}$, $\{\Gamma_k\}$ do not affect the values of x_k for $k \notin \mathscr{I}$, they might affect the index set $\mathscr{I}$ itself. Furthermore, these parameters determine the inverse Hessian approximations H_k. In the following proposition, it is shown that, for certain selections of the parameters γ_k and Γ_k, the sequence of matrices $\{H_k\}$ generated by Algorithm 5.1 converges to a scalar multiple of the inverse Hessian F^{-1}. The subclass of Algorithm 5.1 corresponding to this parameter selection will therefore have the additional capability of generating a Newton step at the end of its n-step cycle. In view of Corollary 6.3, this capability seems redundant, since the algorithm is expected to terminate by its nth iteration. Nevertheless, in practice, this extra feature might have important implications, which will be further discussed in the conclusion section.

Corollary 7.3. Let the sequences $\{H_k\}$ and $\{x_k\}$ be as in Theorem 6.1, with $\{\gamma_k\}$ and $\{\Gamma_k\}$ selected such that

$$\gamma_k = \omega - (q_k^T H_k q_k / p_k^T q_k), \qquad \text{for } k \notin \mathscr{I} \cup \mathscr{I},$$

$$\Gamma_k = \{\omega \begin{bmatrix} (p_k^T q_{k+1}/p_{k+1}^T g_k) & 0 \\ 0 & 1 \end{bmatrix} + R_k Q_k^T H_k Q_k\} R_k, \qquad \text{for } k \notin \mathscr{I},$$

with some fixed $\omega \neq 0$. Assuming

$$\nabla f(x_k) \neq 0, \qquad \text{for } k < n,$$

then

$$H_n = \omega F^{-1}.$$

Proof. By Theorem 6.1(e) and (67), we have

$$H_n F p_i = \omega p_i, \qquad \text{for } i = 0, \ldots, n-1.$$

Thus, if we define

$$M = [p_0, \ldots, p_{n-1}],$$

we may write

$$H_n F M = \omega M.$$

However, since M is nonsingular (by Corollary 6.2), the result follows from multiplying the last equation by $M^{-1}F^{-1}$. □

8. Adaptation to Nonquadratic Problems

Like most conventional quasi-Newton algorithms, Algorithm 5.1 can be easily adapted to handle nonquadratic problems. As usual, in such extensions, the successive approximations to the solution $\{x_k\}$ are determined by search procedures while the updates of the deflection matrices $\{H_k\}$ are based on gradients evaluated in the process. One element which complicates this process in our case is the plane search which, as one may expect, is considerably more involved and more expensive than a simple line search. Such searches can be implemented in a variety of ways that, like line searches, involve tradeoffs between speed of convergence, accuracy, and information requirements. One possible approach is to approximate the plane search with a line search that becomes exactly equivalent to it when the objective function is quadratic. In view of Proposition 5.1, this can be done by redefining the point x_{k+2} in Case B of Algorithm 5.1 as a stationary point of the objective along the line $x_k + \rho d_{k+1}$, where

$$d_{k+1} = (r_k + t_k s_k) p_k - (t_k + r_k^2) H_k q_k,$$

with r_k, s_k, t_k defined by (85) through (87). As pointed out earlier, evaluating s_k requires an additional gradient evaluation along the line $x_k + \beta H_k q_k$.

Current experience with quasi-Newton algorithms for minimization indicates that such methods are not too sensitive to line-search accuracy (at least not as much as conjugate-gradient methods). This gives reason to believe that approximating the plane search with a line search might be appropriate in the context of a quasi-Newton method.

Another point to be considered is the fact that planar iterations are intended for situations where the search direction is nearly singular. In other words, we would normally expect the trigger parameter ϵ to be very small. On the other hand, as shown in Corollary 4.1 and in Propositions 5.1 and 5.2, the update and plane-search procedures get considerably simplified when p_k is exactly singular. In particular, we have shown in Proposition 5.2 that the plane search can be reduced in this case to two line searches. The relative errors introduced in the update and the plane search by applying this simplified iteration when p_k is not singular are of the same order as ϵ. Thus, for sufficiently small ϵ, one might be able to use the simplified planar update given by (54) and (55) and the two-step implementation of the plane search, defined in Proposition 5.2, without adversely affecting the algorithm performance.

An unsettled issue which arises when applying Algorithm 5.1 to nonquadratic saddlepoint problems concerns the global convergence of the new method. The difficulty follows from the fact that Algorithm 5.1 is not a descent method with respect to the objective function. To guarantee convergence of the method on nonquadratic functions, we need a descent function that will monotonically decrease over the sequence of points generated by the algorithm. A natural candidate for this purpose is the norm of the gradient $\|\nabla f(x)\|$. Unfortunately, this is not a valid descent function. Consider for example

$$f(x, y) = x^4 - 3x^3 + 3x^2 - y^2.$$

This function has a saddlepoint with a nonsingular Hessian at the origin, while, along the line $y = 0$, its gradient norm $\|\nabla f(x, y)\|$ has local minima at $x = 0$ and $x = 1$ and a local maximum at $x = \frac{1}{2}$. Thus, nothing can be said at this point about the global convergence of Algorithm 5.1 on nonquadratic saddlepoint problems.

9. Conclusions

The analysis and results presented in this paper focus on the motivation and the theoretical properties of the new class of quasi-Newton methods that can handle indefinite Hessians. For this purpose, we have considered the quadratic case and analyzed the new family of algorithms in its most general form. The free parameters of the family appear in the updates for the deflection matrix H_k, having two free parameters for a regular iteration and four parameters in a planar iteration. Another parameter is ϵ, which determines when a planar iteration should be performed. As shown by our analysis, for fixed x_0 and H_0, the sequence of points generated by the new class of algorithms, for a quadratic objective, is unique. Thus, from a theoretical point of view, it might appear that the choice of parameters in the updates has no effect on the performance of the algorithm. It should be noted, however, that the choice of parameters does affect the sequence of matrices H_k. We have shown, for instance, that, for specific choices of γ_k and Γ_k, the matrices H_k will approximate the inverse Hessian and converge to it after n steps.

Judging from current experience with algorithms from Huang's class, one would expect that, in implementing algorithms from the new family for nonquadratic problems (with imperfect line or plane searches), the choice of parameters will affect performance. Such differences in performance are commonly attributed to the characteristics of the deflection matrices which affect the algorithm numerical stability and their sensitivity to the line-search accuracy and the quadratic nature of the objective function. It seems advantageous, for instance, to select γ_k and Γ_k such that H_k will approximate the inverse Hessian, as indicated above. In this case, the nth step becomes a Newton step, which can compensate to some extent for error that has been accumulated as a result of imperfection of the preceding conjugate-direction steps. The remaining parameter φ_k can be chosen to keep H_k away from singularity in the same way that H_k is maintained positive definite by restricting the range of φ_k in Huang's class. Again, based on past experience, the choice of φ_k is likely to have considerable influence on the stability and overall performance of algorithms belonging to the new class. Finally, one may also expect that the trigger parameter ϵ, which determines how often planar iterations are performed, will also have in practice a noticeable effect on the algorithm performance.

Examining the above issues and the effects of the approximations proposed in the previous section requires further analysis and extensive numerical experiments. Such investigations, however, were out of the scope of this paper. In order to assess the practical value of the proposed method for solving general saddlepoint problems, one would also have to resolve the global convergence issue mentioned earlier.

As a final note, we like to point out that, although the planar iteration was primarily designed to address difficulties associated with indefinite Hessians, it may also prove useful in minimization problems in which the Hessian is ill conditioned. By employing planar iterations (within the framework of Algorithm 5.1), one may avoid difficulties arising in conventional quasi-Newton methods when successive search directions become nearly dependent. The aforementioned global convergence issue is easily resolved when applying Algorithm 5.1 to minimization problems. Clearly, in that case, the objective can serve as a descent functional. Thus, to assure global convergence, it is sufficient to maintain positive definiteness of H, which can be done by appropriate selection of the parameter φ_k in the rank-two and rank-three updates.

References

1. HUANG, H. Y., *Unified Approach to Quadratically Convergent Algorithms for Function Minimization*, Journal of Optimization Theory and Application, Vol. 5, pp. 405–423, 1970.
2. FLETCHER, R., *Conjugage Gradient Methods for Indefinite Systems*, Proceedings of the 1975 Dundee Numerical Analysis Conference, Springer-Verlag, Berlin, Germany, pp. 73–89, 1975.
3. LUENBERGER, D. G., *Hyperbolic Pairs in the Method of Conjugate Gradients*, SIAM Journal on Applied Mathematics, Vol. 17, pp. 1263–1267, 1969.
4. DAVIDON, W. C., *Variable-Metric Methods for Minimization*, AEC, Research and Development Report No. ANL-5990 (Revised), 1959.
5. FLETCHER, R., and POWELL, M. J. D., *A Rapidly Convergent Descent Method for Minimization*, Computer Journal, Vol. 6, pp. 163–168, 1963.
6. OREN, S. S., and LUENBERGER, D. G., *Self-Scaling Variable Metric (SSVM) Algorithms, Part 1: Criteria and Sufficient Conditions for Scaling a Class of Algorithms*, Management Science, Vol 20, pp. 845–862, 1974.
7. BROYDEN, C. G., *The Convergence of a Class of Double-Rank Minimization Algorithms, Part 2: The New Algorithm*, Journal of the Institute for Mathematics and Applications, Vol. 6, pp. 222–231, 1970.
8. SINCLAIR, J. E., and FLETCHER, R., *A New Method of Saddle-Point Location for the Calculation of Defect Migration Energies*, Journal of Physics, Communications: Solid State Physics, Vol. 7, pp. 864–870, 1974.
9. LUENBERGER, D. G., *The Conjugate Residual Method for Constrained Minimization Problems*, SIAM Journal on Numerical Analysis, Vol. 7, pp. 390–398, 1970.
10. PAIGE, C. C., and SAUNDERS, M. A., *Solution of Sparse Indefinite Systems of Linear Equations*, SIAM Journal on Numerical Analysis, Vol. 12, pp. 617–629, 1975.
11. HESTENES, M. R., *Conjugate Directions Methods in Optimization*, Optimization Techniques: Part 1, Edited by J. Stoer, Springer-Verlag, Berlin, Germany, pp. 8–27, 1978.
12. SHANNO, D. F., *Conjugate Gradient Methods with Inexact Searches*, Mathematics of Operations Research, Vol. 3, pp. 244–256, 1978.
13. BUYS, J. D., and GONIN, R., *The Use of Augmented Lagrangian Functions for Sensitivity Analysis in Nonlinear Programming*, Mathematical Programming, Vol. 12, pp. 281–284, 1977.