IMA Journal of Numerical Analysis (1994) **14**, 443–460

On the convergence of conjugate gradient algorithms

R. PYTLAK

Imperial College of Science, Technology and Medicine, Centre for Process Systems Engineering, London SW7 2BT, UK

[Received 1 December 1989 and in revised form 26 June 1993]

In this paper we present a new family of conjugate gradient algorithms. This family originates in the algorithms provided by Wolfe and Lemaréchal for non-differentiable problems. It is shown that the Wolfe–Lemaréchal algorithm is identical to the Fletcher–Reeves algorithm when the objective function is smooth and when line searches are exact. The convergence properties of the new algorithms are investigated. One of them is globally convergent under minimum requirements on the directional minimization.

1. Introduction

In this paper we consider algorithms for the unconstrained minimization problem:

$$\min_{x \in \mathcal{R}^n} f(x). \tag{1.1}$$

In general, we assume that the function f is continuously differentiable, i.e., $f \in C^1$ (however in some cases we will apply a stronger assumption that $f \in C^2$). To solve this problem we may use the conjugate gradient algorithm. The direction at the kth step of this algorithm is determined according to the rule:

$$d_k = -\nabla f(x_k) + t_k d_{k-1} \tag{1.2}$$

where, e.g.,

$$t_k = \frac{\langle \nabla f(x_k) - \nabla f(x_{k-1}), \nabla f(x_k) \rangle}{\| \nabla f(x_{k-1}) \|^2}, \qquad \text{or} \qquad t_k = \frac{\| \nabla f(x_k) \|^2}{\| \nabla f(x_{k-1}) \|^2}, \tag{1.3}$$

see, e.g., Bertsekas (1982), Fletcher (1987), Fletcher and Reeves (1964), Polak and Ribiére (1969); more complicated formulae are also possible Shanno (1978a, b). The first formula in (1.3) is usually called the Polak–Ribiére formula while the second one is the Fletcher–Reeves formula.

The conjugate gradient algorithm is motivated by the fact that if f is a quadratic function and if the minimization in the direction d_k is exact, then a minimum is reached in at most n steps.

In the mid-seventies, new conjugate gradient algorithms were constructed by Wolfe (1975) and Lemaréchal (1975) for solving non-differentiable problems. Their bundle methods take a convex hull of the subgradients from current and previous iterations to define a descent direction. It was proved that if the function is quadratic and the directional minimization is exact, then their algorithms reduce to the conjugate gradient method.

If the function f is differentiable then the Wolfe–Lemaréchal algorithm defines the search direction by

$$d_k = -Nr\{\nabla f(x_k), -d_{k-1}\}, \tag{1.4}$$

where $Nr\{a, b\}$ is defined as the point from a line segment spanned by the vectors a and b which has the smallest norm, i.e.,

$$\|Nr\{a, b\}\| = \min\{\|\lambda a + (1-\lambda)b\| : 0 \leq \lambda \leq 1\}, \tag{1.5}$$

and $\|\cdot\|$ is the Euclidean norm. Let us notice that the operation $Nr\{\cdot, \cdot\}$ can be performed easily. This is a simple univariate quadratic problem with box constraints and can be solved analytically.

The rule (1.4) was mainly motivated by the need of having d_k as a direction of descent (in the non-differentiable case $\nabla f(x_k)$ is substituted by a bundle of subgradients).

Let us consider the problem:

$$\begin{aligned} &\min_{(\mu, d) \in \Re^{n+1}} \{\mu + \tfrac{1}{2}\|d\|^2\} \\ &\text{s.t.} \quad \langle \nabla f(x_k), d\rangle \leq \mu, \\ &\qquad -\langle d_{k-1}, d\rangle \leq \mu. \end{aligned} \tag{1.6}$$

We obtain the solution of this problem by solving its dual:

$$\begin{aligned} \max_{0 \leq \lambda \leq 1} \{-\tfrac{1}{2}\|\lambda \nabla f(x_k) - (1-\lambda)d_{k-1}\|^2\} &= -\tfrac{1}{2}\|Nr\{\nabla f(x_k), -d_{k-1}\}\|^2 \\ &= -\tfrac{1}{2}\|\lambda_k \nabla f(x_k) - (1-\lambda_k)d_{k-1}\|^2. \end{aligned} \tag{1.7}$$

Moreover the optimal value μ_k is

$$\mu_k = -\|d_k\|^2. \tag{1.8}$$

From this we can easily deduce the following properties:

$$\langle \nabla f(x_k), d_k\rangle \leq -\|d_k\|^2, \tag{1.9}$$

$$-\langle d_{k-1}, d_k\rangle \leq -\|d_k\|^2, \tag{1.10}$$

$$\langle \nabla f(x_k), d_k\rangle = -\|d_k\|^2, \qquad \langle d_{k-1}, d_k\rangle = \|d_k\|^2, \qquad \text{if} \quad 0 < \lambda_k < 1, \tag{1.11}$$

because $(\lambda_k, 1-\lambda_k)$ are the Lagrange multipliers for the problem (1.6). From (1.9) we have that if $\|d_k\| \neq 0$, d_k is a direction of descent.

We ask the following questions. If the function f is not quadratic, to which conjugate gradient algorithm, if any, is the Wolfe–Lemaréchal algorithm equivalent? Can we construct other conjugate gradient algorithms based on their scheme?

In order to answer these questions we parametrize the rule (1.4). We introduce the family of algorithms:

$$d_k = -Nr\{\nabla f(x_k), -\beta_k d_{k-1}\} \tag{1.12}$$

and will work out formulae for β_k for which the process $x_{k+1} = x_k + \alpha_k d_k$ becomes the conjugate gradient algorithm.

Let us notice that if $\beta_k = 1$ then we will have the Wolfe–Lemaréchal algorithm. We will show that the Wolfe–Lemaréchal algorithm is equivalent to the Fletcher–Reeves algorithm, provided that the directional minimization is exact. Moreover we will prove that there exists a counterpart of the Polak–Ribiére algorithm, and that it is globally convergent under minimal requirements on directional minimization.

Wolfe and Lemaréchal consider only quadratic functions as far as the differentiable problems were concerned. We use their scheme for general continuously differentiable functions and study the effect of β_k on the performance of the algorithms.

To complete our introduction we would like to notice that the new algorithm with the direction finding subproblem (1.12) will have properties similar to (1.9)–(1.11) (which will be extensively used in this paper):

$$\langle \nabla f(x_k), d_k \rangle \leq -\|d_k\|^2, \tag{1.13}$$

$$-\beta_k \cdot \langle d_{k-1}, d_k \rangle \leq -\|d_k\|^2, \tag{1.14}$$

$$\langle \nabla f(x_k), d_k \rangle = -\|d_k\|^2, \qquad \beta_k \langle d_{k-1}, d_k \rangle = \|d_k\|^2, \qquad \text{if} \quad 0 < \lambda_k < 1. \tag{1.15}$$

The organization of this paper is as follows. In Section 2 we state our general algorithm together with its convergence properties. In Section 3 we introduce new conjugate gradient algorithms, while in Section 4 we prove global convergence for one version of our algorithm. In Section 5 we present numerical results of applying this version of our algorithm to some standard test problems.

Finally we remind the reader that we are concerned with functions defined on the Euclidean space $\mathscr{R}^n$, $\|\cdot\|$ is the Euclidean norm, $\langle \cdot, \cdot \rangle$ is a scalar product and rd (a, b) the angle between the two vectors a and b.

2. A general algorithm

Our general algorithm is as follows.

ALGORITHM Parameters: $\mu, \eta \in (0, 1)$, $\eta > \mu$, $\varepsilon > 0$, $\{\beta_k\}_0^\infty$.
Data: x_0.

1. Set $k = 0$
2. Compute:

$$d_k = -\nabla f(x_k). \tag{2.1}$$

If $\|d_k\| = 0$ then STOP, if not go to Step 3.

3. Compute:

$$d_k = -Nr\{\nabla f(x_k), -\beta_k d_{k-1}\}, \tag{2.2}$$

if $\|d_k\| = 0$ then STOP.

4. Find a positive number α_k such that:

$$f(x_k + \alpha_k d_k) - f(x_k) \leqslant -\mu\alpha_k \|d_k\|^2, \tag{2.3}$$

$$\langle \nabla f(x_k + \alpha_k d_k), d_k \rangle \geqslant -\eta \|d_k\|^2. \tag{2.4}$$

5. Substitute $x_k + \alpha_k d_k$ for x_{k+1}, increase k by one, go to Step 3.

The directional minimization is defined by the expressions (2.3)–(2.4). These rules, which lead to inexact minimization, were taken from the algorithms for non-differentiable problems (Bihain (1984), Kiwiel (1985), Lemaréchal (1975), Mifflin (1977)).

REMARK The choice of μ is very important for the performance of the Algorithm. Because quadratic interpolation is usually applied to approximate values of α_k which satisfy (2.3)–(2.4), we should assume that $0 < \mu < 0{\cdot}5$. Otherwise we could miss minimizing points in the case where $g(\alpha) = f(x_k + \alpha d_k)$ are quadratic functions and this would deteriorate the performance of the Algorithm.

A procedure which finds α_k satisfying (2.3)–(2.4), in a finite number of operations, can be easily constructed (Mifflin (1977)). We include considerations devoted to this procedure solely for completeness of this work.

LEMMA 1 There exists a procedure which finds α_k satisfying (2.3)–(2.4) in a finite number of operations, or produces $\alpha_k \to \infty$ such that $f(x_k + \alpha_k d_k) \to -\infty$.

Proof. The proof is based on Mifflin (1977). Let $\varepsilon > 0$ be such that $\eta > \mu + \varepsilon$. Let α^0 be an arbitrary positive number and set $\alpha_N^0 = \infty$, $\alpha_\mu^0 = 0$. The procedure, which we propose, produces a sequence of points $\{\alpha^l\}$ constructed in the following way. If

$$f(x_k + \alpha^l d_k) - f(x_k) \leqslant -(\mu + \varepsilon)\alpha^l \|d_k\|^2 \tag{2.5}$$

we set $\alpha_\mu^{l+1} = \alpha^l$, $\alpha_N^{l+1} = \alpha_N^l$ and $\alpha_\mu^{l+1} = \alpha_\mu^l$, $\alpha_N^{l+1} = \alpha^l$ otherwise. Next, we substitute α^{l+1} by $(\alpha_\mu^l + \alpha_N^l)/2$ if $\alpha_N^{l+1} < \infty$, or by $2\alpha^l$ if $\alpha_N^{l+1} = \infty$.

If for every α^l generated by the procedure we have that α^l satisfies (2.5) then α_N^l will remain equal to infinity and $f(x_k + \alpha^l d_k) \to -\infty$. Therefore let us suppose that there exists α^l such that (2.5) is not fulfilled. This means that at some iteration the bisection procedure has started and we have the sequences $\{\alpha_N^l\}$, $\{\alpha_\mu^l\}$ such that $\alpha_N^l - \alpha_\mu^l \to 0$, $\alpha_N^l \to \hat{\alpha}$, because either α_N^{l+1} or α_μ^{l+1} is set to $(\alpha_N^l + \alpha_\mu^l)/2$ and $\alpha_N^{l+1} - \alpha_\mu^{l+1} = (\alpha_N^l - \alpha_\mu^l)/2$. Moreover, let us assume that for every l, α^l never satisfies simultaneously (2.4) and (2.5) (otherwise we have found the desired coefficient α_k). In this case $\alpha_\mu^l \to \hat{\alpha}$ and $\hat{\alpha}$ satisfies (2.3). If we have (2.4) for α^l infinitely often then the procedure will terminate after a finite number of iterations, because $\alpha^l \to \hat{\alpha}$, f is continuous and (2.3) will have to be satisfied.

Now, let us suppose that (2.4) holds only for the finite number of α^l, thus

$$\langle \nabla f(x_k + \alpha^l d_k), d_k \rangle < -\eta \|d_k\|^2$$

for infinitely many times. This leads to

$$\langle \nabla f(x_k + \hat{\alpha} d_k), d_k \rangle \leqslant -\eta \|d_k\|^2.$$

Because

$$f(x_k + \alpha_N^l d_k) - f(x_k) > -(\mu + \varepsilon)\alpha_N^l \|d_k\|^2, \tag{2.6}$$

thus

$$f(x_k + \alpha_N^l d_k) - f(x_k + \hat{\alpha} d_k) > -(\mu + \varepsilon)(\alpha_N^l - \hat{\alpha}) \|d_k\|^2 \tag{2.7}$$

and

$$-\eta \|d_k\|^2 \geq \lim_{l \to \infty} \langle \nabla f(x_k + \alpha^l d_k), d_k \rangle = \langle \nabla f(x_k + \hat{\alpha} d_k), d_k \rangle$$

$$= \lim_{l \to \infty} \frac{f(x_k + \alpha_N^l d_k) - f(x_k + \hat{\alpha} d_k)}{\alpha_N^l - \hat{\alpha}} \geq -(\mu + \varepsilon) \|d_k\|^2,$$

but this is impossible since $\eta > \mu + \varepsilon$. □

The step length conditions (2.3), (2.4) have been implemented (Pytlak (1989)) and work well in practice.

To investigate the convergence of the Algorithm we begin by providing a lemma.

LEMMA 2 If the direction d_k is determined by (2.2) and the step-size coefficient α_k satisfies (2.3)–(2.4) then:

$$\lim_{k \to \infty} \|d_k\| = 0 \tag{2.8}$$

or

$$\lim_{k \to \infty} f(x_k) = -\infty. \tag{2.9}$$

Proof. Let us assume that (2.9) is not true. Because $\{f(x_k)\}_0^\infty$ is non-increasing and bounded from below, it has to be convergent, so that we have

$$f(x_k + \alpha_k d_k) - f(x_k) \xrightarrow[k \to \infty]{} 0. \tag{2.10}$$

(2.10) and (2.3) imply (from the theorem on three sequences) that

$$\mu \alpha_k \|d_k\|^2 \xrightarrow[k \to \infty]{} 0. \tag{2.11}$$

This is not equivalent to (2.8), thus let us suppose that there exists a set of natural numbers K such that

$$\lim_{k \to \infty,\, k \in K} \|d_k\| = \|\bar{d}\| \neq 0, \tag{2.12}$$

where, in general, we assume that $\|\bar{d}\|$ can be infinite. Then, from (2.11) and (2.12) we have

$$\lim_{k \to \infty,\, k \in K} \alpha_k \|d_k\| = 0, \tag{2.13}$$

so that

$$\lim_{k \to \infty,\, k \in K} \|x_{k+1} - x_k\| = 0.$$

Using this, (2.4) and (1.13) we get

$$-\eta\,\|d_k\|^2 \leqslant \langle \nabla f(x_{k+1}),\, d_k\rangle \leqslant -\tfrac{1}{2}(1+\eta)\,\|d_k\|^2 \tag{2.14}$$

for $k \geqslant k_1$.

Obviously this is impossible ($\eta \in (0, 1)$), therefore (2.8) is true. □

Let us look at the specification of the direction d_k. In general we have the following formula for d_k:

$$d_k = -(1-\lambda_k)\nabla f(x_k) + \lambda_k \beta_k d_{k-1} \tag{2.15}$$

where $0 \leqslant \lambda_k \leqslant 1$.

The condition (2.8) is not equivalent to the condition: $\lim_{k\to\infty} \|\nabla f(x_k)\| = 0$. This is due to the additional vector $\beta_k d_{k-1}$ in the formula (2.15).

It can happen that (2.8) holds because the vectors d_{k-1} are not appropriately scaled by the β_k. The Wolfe sequence is the example of this situation. If $\beta_k = 1$ then it can be shown that $\|d_k\| \leqslant \chi\,\|d_{k-1}\|$, $\chi \in (0, 1)$ (see Mifflin (1977), Wolfe (1975)).

Moreover we can have

$$\lim_{k\in K,\, k\to\infty} \operatorname{rd}\left(-\nabla f(x_k),\, d_{k-1}\right) = \pi$$

for a certain sequence $\{\nabla f(x_k)\}_{k\in K}$ and inexact line search.

In each of these situations we shall have (2.8). Thus in order to prove the convergence of the Algorithm we have to exclude these situations.

THEOREM 1 Let us assume that $\{\beta_k\}$ is such that

$$\liminf_{k\to\infty}(\beta_k\,\|d_{k-1}\|) \geqslant \nu_1 \liminf_{k\to\infty} \|\nabla f(x_k)\| \tag{2.16}$$

where ν_1 is some positive constant. If there exists a number ν_2 such that $\nu_2 \in (0, 1)$,

$$\langle \nabla f(x_k),\, d_{k-1}\rangle \leqslant \nu_2\,\|\nabla f(x_k)\|\,\|d_{k-1}\|, \qquad \text{whenever} \quad \lambda_k \in (0, 1) \tag{2.17}$$

then $\lim_{k\to\infty} f(x_k) = -\infty$, or every cluster point $\bar{x}$ of the sequence $\{x_k\}_0^\infty$ generated by the Algorithm is such that $\nabla f(\bar{x}) = 0$.

Proof.

Case (a) Let us suppose that for infinitely often $k \in K_1$, $\lambda_k \in (0, 1)$, thus

$$\langle \nabla f(x_k),\, d_k\rangle = -\|d_k\|^2 \qquad \text{and} \qquad \beta_k\langle d_{k-1},\, d_k\rangle = \|d_k\|^2. \tag{2.18}$$

Moreover let us assume that

$$x_k \xrightarrow[k\to\infty,\, k\in K_1]{} \bar{x}, \qquad \nabla f(\bar{x}) \neq 0.$$

From this it follows that

$$\lim_{k\to\infty,\, k\in K_1} \|\nabla f(x_k)\| \neq 0.$$

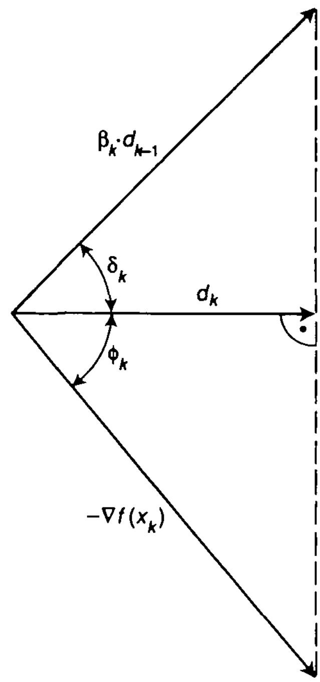

FIG. 1.

Because of this, the equalities (2.18), and since by Lemma 2 $\lim_{k\to\infty} \|d_k\| = 0$ (due to our assumption, (2.12) does not hold), we have

$$\lim_{k\to\infty,\, k\in K_1} \cos \mathrm{rd}\,(-\nabla f(x_k), d_k) = \lim_{k\to\infty,\, k\in K_1} \cos \phi_k$$

$$= \lim_{k\to\infty,\, k\in K_1} \left\langle -\frac{\nabla f(x_k)}{\|\nabla f(x_k)\|}, \frac{d_k}{\|d_k\|} \right\rangle = \lim_{k\to\infty,\, k\in K_1} \frac{\|d_k\|}{\|\nabla f(x_k)\|} = 0. \tag{2.19}$$

$$\lim_{k\to\infty,\, k\in K_1} \cos \mathrm{rd}\,(\beta_k d_{k-1}, d_k) \equiv \lim_{k\to\infty,\, k\in K_1} \cos \delta_k = \lim_{k\to\infty,\, k\in K_1} \frac{\|d_k\|}{\|d_{k-1}\|\, \beta_k} = 0. \tag{2.20}$$

Since (2.19), (2.20) are satisfied: $\phi_k \to \pi/2$, $\delta_k \to \pi/2$. Let us consider the angle $\phi_k + \delta_k$. From the usual calculus it follows that

$$\lim_{k\to\infty,\, k\in K_1} \cos(\phi_k + \delta_k) = \lim_{k\in\infty,\, k\in K_1} \cos \phi_k \cos \delta_k - \lim_{k\in\infty,\, k\in K_1} \sin \phi_k \sin \delta_k = -1 \tag{2.21}$$

(see also Fig. 1), but this implies that

$$\lim_{k\to\infty,\, k\in K_1} \left\langle \frac{\nabla f(x_k)}{\|\nabla f(x_k)\|}, \frac{d_{k-1}}{\|d_{k-1}\|} \right\rangle = 1.$$

This contradicts our assumption (2.17) hence we conclude that $\nabla f(\bar{x}) = 0$.

Case (b) Now let us consider the case $\lambda_k = 1$. If it occurs infinitely often for

$k \in K_2$ and

$$x_k \xrightarrow[k\to\infty,\, k\in K_2]{} \bar{x}, \qquad \text{with} \quad \nabla f(\bar{x}) \neq 0$$

we have that there is a v_4 such that

$$\liminf_{k\to\infty,\, k\in K_2} \|\nabla f(x_k)\| = v_4 > 0 \tag{2.22}$$

and by assumption $\lambda_k = 1$, and (2.16)

$$\liminf_{k\to\infty,\, k\in K_2} (\beta_k \|d_{k-1}\|) \geq v_1 v_4 > 0. \tag{2.23}$$

But

$$\lim_{k\to\infty,\, k\in K_2} \|d_k\| = 0 = \lim_{k\to\infty,\, k\in K_2} (\beta_k \|d_{k-1}\|) \geq v_1 v_4 > 0$$

and this is impossible.

Case (c) If we have the case $\lambda_k = 0$ for $k \in K_3$ then $-d_k = \nabla f(x_k)$ for $k \in K_3$. If

$$x_k \xrightarrow[k\to\infty,\, k\in K_3]{} \bar{x}, \qquad \nabla f(\bar{x}) \neq 0$$

then

$$\lim_{k\to\infty,\, k\in K_3} \|\nabla f(x_k)\| > 0$$

but this is a contradiction to $\lim_{k\to\infty} \|d_k\| = 0$. This completes our proof. □

The condition (2.17) is independent of the choice of the sequence $\{\beta_k\}_0^\infty$ and is connected with the directional minimization. Therefore it is not surprising that in some important situations it can be substituted, by other, more easily verifiable conditions.

LEMMA 3 Let $\{x_k\}_0^\infty$ be generated by the Algorithm, where β_k satisfies (2.16). Then

$$\lim_{k\to\infty} f(x_k) = -\infty \qquad \text{or} \qquad \lim_{k\to\infty} \|\nabla f(x_k)\| = 0 \tag{2.24}$$

if one of the following conditions holds:

(i) there exists $\eta \in (0, 1)$ such that

$$|\langle \nabla f(x_{k+1}), d_k \rangle| \leq \eta \|d_k\|^2, \tag{2.25}$$

(ii)

$$\lim_{k\to\infty} \|x_{k+1} - x_k\| = 0, \tag{2.26}$$

(iii) the function f is locally uniformly convex, i.e. there exists an increasing function d from $\mathscr{R}^+$ ($\mathscr{R}^+ = \{x \in \mathscr{R} : x \geq 0\}$) into $\mathscr{R}$ such that:

$$d(0) = 0, \qquad d(t) > 0 \qquad \text{if} \quad t > 0,$$

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y) - \lambda(1-\lambda) d(\|x - y\|),$$
$$\forall x, \quad y \in \mathscr{R}^n, \quad \forall \lambda \in [0, 1]. \tag{2.27}$$

This condition is satisfied in particular if f is strongly convex, in that case $d(t) = \alpha t^2$, $\alpha > 0$.

Proof. First of all we assume that there exists $M > -\infty$ such that

$$f(x_k) \geq M, \qquad \forall k, \tag{2.28}$$

otherwise we would have our thesis.

Let us assume that the case (i) occurs, and

$$\lim_{k \to \infty, k \in K} \|\nabla f(x_k)\| = \alpha > 0 \tag{2.29}$$

(where, in general, we assume that α can be equal to $+\infty$) for some infinite set K. Then from Theorem 1 we shall obtain that for every $v \in (0, 1)$ we will have

$$\langle \nabla f(x_k), d_{k-1} \rangle > v \, \|\nabla f(x_k)\| \, \|d_{k-1}\|$$

(contradiction to the condition (2.17), provided that $0 < \lambda_k < 1$). Because $\|d_k\| \to_{k \to \infty} 0$ (Lemma 2) and since (2.25) we can write

$$\langle \nabla f(x_k), d_{k-1} \rangle > \eta \, \|d_{k-1}\|^2$$

which contradicts our assumption (2.25). Therefore our assumption (2.29) has not been valid.

For part (ii) of the lemma, the proof is carried out along the same lines as for part (i). Firstly we assume that for some infinite set K we have (2.29). Therefore for every $v \in (0, 1)$ we have

$$\langle \nabla f(x_k), d_{k-1} \rangle > v \, \|\nabla f(x_k)\| \, \|d_{k-1}\|.$$

But for sufficiently large $k \in K$, since (2.26), we will achieve the relation (from Theorem 1, and (1.13))

$$0 < \alpha v \leq \lim_{k \to \infty, k \in K} \left\langle \nabla f(x_k), \frac{d_{k-1}}{\|d_{k-1}\|} \right\rangle \leq \lim_{k \to \infty, k \in K} (-\|d_{k-1}\|) = 0. \tag{2.30}$$

This is impossible, thus (2.29) cannot happen.

Now let us assume that case (iii) occurs. Because f is uniformly convex and differentiable, we have (see Lemaréchal (1975))

$$\begin{aligned} f(y) &\geq f(x) + \langle \nabla f(x), y - x \rangle + d(\|y - x\|) \\ d(\|y - x\|) &\to 0 \qquad \text{when} \quad \|y - x\| \to 0. \end{aligned} \tag{2.31}$$

Let us suppose that there exists subsequence $\{\nabla f(x_k)\}_{k \in K}$ such that (2.29) is satisfied. In this case, in view of condition (2.16),

$$\langle \nabla f(x_k), x_{k+1} - x_k \rangle = -\alpha_k \, \|d_k\|^2, \tag{2.32}$$

since $0 < \lambda_k < 1$. Thus, from (2.28), we have

$$f(x_{k+1}) - f(x_k) \xrightarrow[k \to \infty]{} 0,$$

and from (2.3)

$$\alpha_k \, \|d_k\|^2 \xrightarrow[k \to \infty]{} 0.$$

Thus it follows, from (2.31), (2.32) and the definition of $d(\cdot)$, that

$$\|x_{k+1} - x_k\| \xrightarrow[k \in K]{} 0.$$

This, as in the proof of part (ii), leads to a contradiction. □

Lemma 3 indicates that the assumption (2.17) is necessary because of the rather inexact directional minimization. Here we must also mention that condition (2.26) plays an important role in conjugate gradient algorithms which are based on the Polak–Ribiére rule (recent results have shown that this assumption is not necessary for the Fletcher–Reeves algorithm, see Al-Baali (1985), Powell (1977, 1986), Shanno (1978b)).

3. New conjugate gradient algorithms

Now we shall give examples of sequences $\{\beta_k\}_0^\infty$ which assure that under exact directional minimization the directions $\{d_k\}_0^\infty$ generated by the Algorithm are conjugate, provided that f is quadratic.

For this purpose we consider the simple direction finding quadratic problem:

$$\min_{0 \leq \lambda \leq 1} \|\lambda(-\beta_k d_{k-1}) + (1-\lambda)\nabla f(x_k)\|. \tag{3.1}$$

We can solve this problem analytically:

$$\|d_k\|^2 = \min_{0 \leq \lambda \leq 1} \left(\beta_k^2 \lambda^2 \|d_{k-1}\|^2 - 2\lambda(1-\lambda)\beta_k \langle \nabla f(x_k), d_{k-1}\rangle + (1-\lambda)^2 \|\nabla f(x_k)\|^2\right). \tag{3.2}$$

So, if $\langle \nabla f(x_k), d_{k-1}\rangle = 0$ we obtain

$$\lambda_{\min} = \frac{\|\nabla f(x_k)\|^2}{\beta_k^2 \|d_{k+1}\|^2 + \|\nabla f(x_k)\|^2}. \tag{3.3}$$

If we let

$$\beta_k = \frac{\gamma_k}{\sqrt{1-\gamma_k^2}} \frac{\|\nabla f(x_k)\|}{\|d_{k-1}\|}, \qquad \gamma_k \in (0, 1) \tag{3.4}$$

then, after some calculations, we obtain

$$\|d_k\|^2 = \gamma_k^2 \|\nabla f(x_k)\|^2. \tag{3.5}$$

We notice that γ_k has an important meaning. We can show that the angle between d_k and $\nabla f(x_k)$ has a cosine:

$$\cos \mathrm{rd}\,(d_k, \nabla f(x_k)) = \frac{\langle d_k, \nabla f(x_k)\rangle}{\|d_k\| \, \|\nabla f(x_k)\|} = -\gamma_k. \tag{3.6}$$

We shall use this to derive formulae for the sequence $\{\beta_k\}_0^\infty$. Let

$$\gamma_k = \frac{\|\nabla f(x_k)\|}{\sqrt{t_k^2 \|d_{k-1}\|^2 + \|\nabla f(x_k)\|^2}} \tag{3.7}$$

where

$$t_k = \frac{\|\nabla f(x_k)\|^2}{\|\nabla f(x_{k-1})\|^2} \tag{3.8}$$

or

$$t_k = \frac{\langle \nabla f(x_k) - \nabla f(x_{k-1}), \nabla f(x_k)\rangle}{\|\nabla f(x_{k-1})\|^2} \tag{3.9}$$

and

$$\underline{d_{k-1}} = \frac{1}{\gamma_{k-1}^2} d_{k-1}. \tag{3.10}$$

These formulae are derived under the hypothesis that the directional minimization is exact. It can be shown that the cosine of the angle between d_k and $\nabla f(x_k)$ coincides with that of the Fletcher–Reeves (3.8) and the Polak–Ribiére algorithms (3.9), respectively.

THEOREM 2 If directional minimization is exact, $\{\beta_k\}_0^\infty$ is defined by (3.4) in which γ_k is determined by (3.7), (3.10) and one of the expressions (3.8) or (3.9) then

(i) the directions $\{d_k\}_0^\infty$ are identical to those generated by the conjugate gradient algorithm,

(ii) if t_k is given by (3.9) and $\lim_{k\to\infty} \|x_{k+1} - x_k\| = 0$ we also have:

$$\liminf_{k\to\infty} \|\nabla f(x_k)\| = 0. \tag{3.11}$$

Proof. We have

$$d_0 = -\nabla f(x_0),$$

$$d_{k+1} = \gamma_{k+1}^2(t_{k+1}\underline{d_k} - \nabla f(x_{k+1})), \qquad \underline{d_k} = \frac{1}{\gamma_k^2} d_k,$$

$$k = 0, 1, \ldots, \qquad \gamma_0 = 1. \tag{3.12}$$

Moreover $\langle \nabla f(x_k), \underline{d_{k-1}}\rangle = 0$.

Now consider the directions generated by the conjugate gradient method (see (1.2)–(1.3)).

$$\bar{d}_0 = -\nabla f(x_0),$$

$$\bar{d}_{k+1} = t_{k+1}\bar{d}_k - \nabla f(x_{k+1}), \qquad k = 0, 1, \ldots,$$

and $\langle \nabla f(x_k), \bar{d}_{k-1}\rangle = 0$. It is evident that the directions $\{\underline{d_k}\}_0^\infty$ fulfill these conditions and because $\gamma_k^2 > 0$, $d_k = \gamma_k^2 \underline{d_k}$. We have thus proved assertion (i). Regarding (ii) we recall (3.6). If

$$\liminf_{k\to\infty} \|\nabla f(x_k)\| > 0$$

then it can be shown that $\liminf_{k\to\infty} \gamma_k > 0$. But that together with condition (2.26) leads to a contradiction (Lemma 3(ii)), hence $\liminf_{k\to\infty} \|\nabla f(x_k)\| = 0$. □

From (3.7)–(3.10), (3.4) it is evident that we can deliver many formulae for β_k, those described by (3.4) and (3.7)–(3.10) are only examples. Another formula is obtained by setting

$$t_k = \frac{\langle \nabla f(x_k) - \nabla f(x_{k-1}), \nabla f(x_k) \rangle}{\langle \nabla f(x_k) - \nabla f(x_{k-1}), d_{k-1} \rangle} \tag{3.13}$$

in (3.7) (this is the Hestenes–Stiefel formula which is equivalent to the Polak–Ribiére formula only when directional minimization is exact, see Shanno (1978*a*)).

If we define β_k by (3.4) with γ_k given by (3.7) we obtain

$$\beta_k = \frac{\| \nabla f(x_k) \|^2 \gamma_{k-1}^2}{|t_k| \, \| d_{k-1} \|^2}. \tag{3.14}$$

If we take t_k defined by (3.8) and recall (3.5) (which is valid when directional minimization is exact) we obtain the Wolfe–Lemaréchal formula

$$\beta_k = 1. \tag{3.15}$$

Similarly we can obtain the fomula (see (3.9))

$$\beta_k = \frac{\| \nabla f(x_k) \|^2}{| \langle \nabla f(x_k) - \nabla f(x_{k-1}), \nabla f(x_k) \rangle |}. \tag{3.16}$$

As (3.15) can be regarded as the Fletcher–Reeves formula, (3.16) can be treated as the Polak–Ribiére formula. (3.14) with γ_{k-1} and t_k from (3.7)–(3.10) can be interpreted similarly.

Obviously formula (3.15) does not satisfy the assumptions of our theorem concerning the convergence of the Algorithm, because

$$\beta_k \| d_{k-1} \| \xrightarrow[k \to \infty]{} 0.$$

In the case of the Wolfe–Lemaréchal formula the method of assuring global convergence reduces to restarting the Algorithm whenever $\|d_k\| \leq \delta_k$, where $\delta_k \to_{k \to \infty} 0^+$.

4. The globally convergent conjugate gradient algorithm

In this section we examine the Algorithm with the sequence $\{\beta_k\}_0^\infty$ defined by (3.16). To prove global convergence results we have to assume that there exists $L < +\infty$ such that:

$$\| \nabla f(x) - \nabla f(y) \| \leq L \, \| x - y \| \qquad \forall x, y \in \mathfrak{R}^n. \tag{4.1}$$

We can prove the following theorem.

THEOREM 3 If the function f satisfies condition (4.1) then the Algorithm gives

$$\lim_{k \to \infty} f(x_k) = -\infty, \qquad \text{or} \qquad \lim_{k \to \infty} \| \nabla f(x_k) \| = 0 \tag{4.2}$$

provided that:

(i) β_k is given by (3.16),
(ii) there exists $M < +\infty$ such that $\alpha_k \leq M$, $\forall k$.

Proof. We have for formula (3.16):

$$\beta_k \|d_{k-1}\| = \frac{\|\nabla f(x_k)\|^2 \|d_{k-1}\|}{|\langle \nabla f(x_k) - \nabla f(x_{k-1}), \nabla f(x_k)\rangle|}$$
$$\geq \frac{\|\nabla f(x_k)\|^2 \|d_{k-1}\|}{L\|\nabla f(x_k)\| \|\alpha_{k-1}\| \|d_{k-1}\|} \geq \frac{\|\nabla f(x_k)\|}{LM}. \tag{4.3}$$

If $f(x_k) \geq \bar{L} > -\infty$ then, because $\|x_k - x_{k-1}\| \leq M \|d_{k-1}\|$ and $\|d_k\| \to_{k\to\infty} 0$ (Lemma 2), we know from Lemma 3(ii) that only violation of the condition (2.16) can destroy our convergence results (4.2). But (4.3) assures us that condition (2.16) is fulfilled. □

This convergence result is not as restrictive as other results for conjugate gradient methods (especially for the Polak–Ribiére methods). Shanno (1978b) proved that $(2.26) \Rightarrow \lim\inf_{k\to\infty} \|\nabla f(x_k)\| = 0$ (if $f(x) \geq \bar{L} > -\infty, \forall x$) for his algorithm which reduces to the Polak–Ribiére algorithm when directional minimization is exact.

Condition (ii) in Theorem 3 is very useful in proving convergence of our Algorithm (with the rule (3.16)) for problems with strictly convex functions.

THEOREM 4 If the function f is twice continuously differentiable and there exist $+\infty > M > 0$, $+\infty > m > 0$ such that:

$$M\|u\|^2 \geq u^{\mathrm{T}} \nabla^2 f(x) u \geq m\|u\|^2, \qquad \forall x, u, \tag{4.4}$$

then the sequence $\{x_k\}_0^\infty$ generated by the Algorithm with β_k calculated in accordance with (3.16) converges to the minimizer of f.

Proof. Assume first that the direction of descent satisfies the condition:

$$\langle \nabla f(x_k), d_k\rangle = -\|d_k\|^2, \tag{4.5}$$

then we will show that there exists $M_\alpha^1 < M_\alpha^2 < \infty$ such that $\alpha_k^1 \geq M_\alpha^1$ and $\alpha_k^2 \leq M_\alpha^2$, where α_k^1, α_k^2 are those values of α which satisfy relations (2.3) and (2.4), respectively. Because f is twice continuously differentiable we have from Taylor's theorem:

$$\langle \nabla f(x_k + \alpha_k d_k) - \nabla f(x_k), \alpha_k d_k\rangle$$
$$= \int_0^1 \langle \alpha_k d_k^{\mathrm{T}} \nabla^2 f(x_k + \lambda\alpha_k d_k), \alpha_k d_k\rangle \, \mathrm{d}\lambda \leq M\alpha_k^2 \|d_k\|^2.$$

Thus

$$\langle \nabla f(x_k + \alpha_k d_k), d_k\rangle \leq \langle \nabla f(x_k), d_k\rangle + M\alpha_k \|d_k\|^2 = \|d_k\|^2(M\alpha_k - 1),$$

so that if

$$\langle \nabla f(x_k + \alpha_k^1 d_k), d_k\rangle \geq -\eta \|d_k\|^2$$

then $\alpha_k^1 \geq M_\alpha^1 = (1-\eta)/M$. In order to show that there exists M_α^2 we have to use Taylor's theorem again to get

$$f(x_k + \alpha d_k) = f(x_k) + \alpha \langle \nabla f(x_k), d_k \rangle + \tfrac{1}{2}(\alpha)^2 d_k^T \nabla^2 f(\xi_k) d_k$$

for some ξ_k. Therefore, if

$$f(x_k + \alpha_k^2 d_k) - f(x_k) \leq -\mu \alpha_k^2 \|d_k\|^2$$

then $\alpha_k^2 \leq M_k^2 = 2(1-\mu)/m$. Thus, in the first case, there exists $M_\alpha = M_\alpha^2 < \infty$ such that $\alpha_k \leq M_\alpha$ for all k for which (4.5) is fulfilled, thus from Theorem 3 (because condition (4.1) is obviously satisfied) we have our thesis.

Let us assume that condition (4.5) is fulfilled only finitely many times. Thus $d_k = \beta_k d_{k-1}$ for $k \geq k_1$. Because $\{x : f(x) \leq f(x_0)\}$ is a closed bounded set (Lemma 12.8 in Pshenichnyi and Danilin (1978)), it is easy to show that

$$\sum_{k=k_1}^{\infty} \|x_{k+1} - x_k\| < \infty,$$

thus

$$\sum_{k=k_1}^{\infty} \alpha_k \beta_k \|d_{k-1}\| < \infty,$$

and

$$\lim_{k\to\infty} \alpha_k \beta_k \|d_{k-1}\| = 0.$$

But

$$\alpha_k \beta_k \|d_{k-1}\| = \frac{\|\nabla f(x_k)\|^2 \alpha_k}{|\langle \nabla f(x_k) - \nabla f(x_{k-1}), \nabla f(x_k) \rangle|} \|d_{k-1}\| \geq \frac{\alpha_k}{\alpha_{k-1}} \frac{\|\nabla f(x_k)\|}{L}.$$

Therefore either

$$\lim_{k\to\infty} \|\nabla f(x_k)\| = 0, \qquad \text{or} \qquad \lim_{k\to\infty} \frac{\alpha_k}{\alpha_{k-1}} = 0$$

and from the D'Alambert theorem it follows that $\sum_{k=k_1}^{\infty} \alpha_k < \infty$. Therefore there exists $M < \infty$ such that $\alpha_k \leq M$, $\forall k \geq k_1$, and on the basis of Theorem 3 we have proved our thesis. □

The algorithm has been presented under the silent assumption that at every iteration we have

$$\langle \nabla f(x_{k+1}), d_k \rangle \neq \|\nabla f(x_{k+1})\| \, \|d_k\|.$$

If this condition is not satisfied then d_k, $\nabla f(x_{k+1})$ are linearly dependent, so that the linear combination of these vectors will always be a vector collinear with $\nabla f(x_{k+1})$. Of course we can avoid this situation by simple modification of the directional minimization procedure.

We end this section by discussing the use of the Armijo step-size rule in the Algorithm, i.e. when the coefficient α_k in (2.3), (2.4) is determined acccording to:

$$\alpha_k = \arg\max \{\tau^i : f(x_k + \tau^i d_k) - f(x_k) \leq -\mu \tau^i \|d_k\|^2, \quad \tau \in (0, 1), \quad i = 0, 1, \ldots\}. \tag{4.6}$$

If the Armijo step size rule is applied one can prove that one of the conditions (2.8), (2.9) will hold. This can be demonstrated in the same way as Lemma 2 (some obvious modifications can be found in the proof of Theorem 4.13 (in Bertsekas (1982)). Moreover, because (2.26) is fulfilled, and $\{\beta_k\}$ satisfies condition (2.16), the global convergence follows from Theorem 3 (instead of using Theorem 1 we can rely on Lemma 3, because of (2.26)).

The new algorithm does not require exact directional minimization and is globally convergent without any additional assumptions except (4.1). It seems to be the first such algorithm. Yet we do not recommend the use of (4.6) in the Algorithm, because the conjugate gradient algorithm with this directional minimization can exhibit slow convergence.

5. Numerical experiments

Our Algorithm has been implemented on a SUN SPARC1 workstation in FORTRAN using *double precision* accuracy. In order to give an idea of the behaviour of our Algorithm with the formula for β_k given by (3.16) we have run it on several standard test problems.

1. Rosenbrock function (Rosenbrock (1960)):

$$f = 100(x_2 - x_1^2)^2 + (1 - x_1)^2, \qquad x \in \mathfrak{R}^2,$$

2. Extended Rosenbrock function (Shanno (1978a)):

$$f = \sum_{i=2}^{10} \{100(x_i - x_{i-1}^2)^2 + (1 - x_i)^2\}, \qquad x \in \mathfrak{R}^{10},$$

3. Powell function (Powell (1962)):

$$f = (x_1 + 10x_2)^2 + 5(x_3 - x_4)^2 + (x_2 - 2x_3)^4 + 10(x_1 - x_4)^4, \qquad x \in \mathfrak{R}^4,$$

4. Cube function (Himmeblau (1972)):

$$f = 100(x_2 - x_1^3)^2 + (1 - x_1)^2, \qquad x \in \mathfrak{R}^2,$$

5. Beale function (Himmeblau (1972)):

$$f = \sum_{i=1}^{3} (c_i - x_1(1 - x_2^i))^2, \qquad c_1 = 1{\cdot}5, \qquad c_2 = 2{\cdot}25,$$

$$c_3 = 2{\cdot}625, \qquad x \in \mathfrak{R}^2,$$

6. Wood function (Colville 1968)):

$$f = 100(x_2 - x_1^2)^2 + (1 - x_1)^2 + 90(x_4 - x_3^2)^2 + (1 - x_3)^2$$
$$+ 10{\cdot}1\{(x_2 - 1)^2 + (x_4 - 1)^2\} + 19{\cdot}8(x_2 - 1)(x_4 - 1), \qquad x \in \mathfrak{R}^4,$$

7. Watson function (Le (1985)):

$$f = \sum_{i=1}^{30} \left\{ \sum_{j=1}^{10} (j - 1)x_j y_i^{j-2} - \left(\sum_{j=1}^{10} x_j y_i^{j-1} \right)^2 - 1 \right\}^2,$$
$$y_i = (i - 1)/29, \qquad x \in \mathfrak{R}^{10},$$

TABLE 1

Problem	ITER	IFUN	IGRAD	Le (1985)	Shanno (1978a)
Rosenbrock					
(−1.2,1.)	38	127	56	—	—
Extended Rosenbrock					
(−1.2,1,...,1.)	59	130	59	(22, 138)	F
Powell					
(−3,,−1.,0.,1.)	64	156	65	—	(64, 160)
Cube					
(−1.2,1.)	42	106	43	(13, 169)	—
Beale					
(0.,0.)	26	69	27	(7, 60)	—
Wood					
(−3.,1.,−3.,1.)	106	248	108	—	(113, 195)
(−3.,−1.,−3.,−1.)	67	148	67	(20, 134)	(101, 235)
(−1.2,1.,−1.2,1.)	145	353	152	—	(93, 219)
(−1.2,1.,1.2,1.)	98	232	99	—	(48, 118)
Watson					
(0.,0. ,..., 0.)	18	39	18	(383, 3416)	F
Oren–Spedicato					
(1.,1. ,..., 1.)	20	46	20	(15, 99)	—

8. Oren–Spedicato function (Spedicato (1978)):

$$f = \left\{ \sum_{i=1}^{20} i x_i^2 + c \right\}^r, \qquad c = 0, \qquad r = 2, \qquad x \in \mathfrak{R}^{20}.$$

We use the procedure described in Le (1985). Every time this procedure finished the conditions (2.3), (2.4) were checked with parameters: $\mu = 0{\cdot}0001$, $\eta = 0{\cdot}9$. If these conditions were not satisfied, we applied the algorithm described in the proof of Lemma 1.

The results of our computations are shown in Table 1, where ITER means the number of iterations, IFUN the number of function evaluations and IGRAD the number of gradient evaluations. Our stopping criterion was

$$|\nabla_{x_i} f(x_k)| \leq 10^{-5}, \qquad i = 1, 2, \ldots, n.$$

The numbers in the first column are initial points for our Algorithm.

In the last two columns we show the results (if available) for two efficient methods based on the conjugate gradient approach (which use the same stopping criterion). The method described in Le (1985) requires as many gradient calculations as the number of iterations, while for the method in Shanno (1978a) this number is equal to the number of function evaluations. The first number in the brackets is the number of iterations, the second is the number of function evaluations. The symbol 'F' denotes that the method failed to converge. It should be noted that the method presented in Shanno (1978a) requires two additional vectors to calculate a direction of descent. We should mention that both these methods were favourable compared to other conjugate gradient algorithms.

A comparison with existing techniques gives rise to many questions: computers and compiler systems on which the other methods were run are not fully comparable, different line searches and restarting procedures were used, etc. Therefore the main reason for presenting the numerical results of the new method was to show that the method (without any restarting procedure) is a promising alternative.

Acknowledgement

I wish to thank the two anonymous referees for their very helpful suggestions which significantly improved this paper.

References

Al-Baali, M. 1985 Descent property and global convergence of the Fletcher–Reeves method with inexact line search. *IMA J. Numer. Anal.* **5**, 121–124.

Bertsekas, D. P. 1982 *Constrained Optimization and Lagrange Multipliers Methods.* New York: Academic.

Bihain, A. 1984 Optimization of upper semidifferentiable functions. *J. Optimiz. Theory Appl.* **44**, 545–568.

Colville, A. R. 1968 A comparative study of nonlinear programming codes. *IBM New York Scientific Centre Report* 320–2949.

Fletcher, R. 1987 *Practical Methods of Optimization.* Chichester: Wiley.

Fletcher, R., & Reeves, C. M. 1964 Function minimization by conjugate gradients. *Comput. J.* **7**, 149–154.

Himmeblau, D. H. 1972 *Applied Nonlinear Programming.* New York: McGraw Hill.

Kiwiel, K. 1985 *Methods of Descent for Nondifferentiable Optimization* (*Springer Lecture Notes in Mathematics 1133*). Berlin: Springer.

Le, D. 1985 A fast and robust unconstrained optimization method requiring minimum storage. *Math. Program.* **32**, 41–68.

Lemaréchal, C. 1975 An extension of Davidon methods to nondifferentiable problems. *Mathematical Programming Study 3* (M. L. Balinski and P. Wolfe, eds). Amsterdam: North-Holland, pp. 95–109.

Mifflin, R. 1977 An algorithm for constrained optimization with semismooth functions. *Math. Oper. Res.* **2**, 191–207.

Polak, E., Mayne, D. Q., & Wardi, Y. 1983 On the extension of constrained optimization algorithms from differentiable to nondifferentiable problems. *SIAM J. Control Optimiz.* **21**, 179–203.

Polak, E., & Ribiére, C. 1969 Note sur la convergence de methods de directions conjugres. *Revue Francaise Inf. Recherche Operationalle* **16**, 35–43.

Powell, M. J. D. 1962 An iterative method for finding stationary values of a function of several variables. *Comput. J.* **5**, 147–151.

Powell, M. J. D. 1977 Restart procedures for the conjugate gradient method. *Math. Program.* **12**, 241–254.

Powell, M. J. D. 1986 Convergence properties of algorithms for nonlinear optimization. *SIAM Rev.* **28**, 487–500.

Pshenichnyi, B. N., & Danilin, Yu. M. 1978 *Numerical Methods in Extremal Problems.* Moscow: Mir.

Pytlak, R. 1989 Numerical experiment with new conjugate direction methods for nondifferentiable optimization. *Proc. 28th IEEE Conf. for Decision and Control* (*12–15 December, Tampa*). New York: IEEE, pp 2457–2462.

ROSENBROCK, H. H. 1960 An automatic method for finding the greatest or least value of a function. *Comput. J.* **3**, 175–184.

SHANNO, D. F. 1978a Conjugate gradient methods with inexact searches. *Math. Oper. Res.* **3**, 244–256.

SHANNO, D. F. 1978b On the convergence of a new conjugate gradient algorithm. *SIAM J. Numer. Anal.* **15**, 1247–1257.

SPEDICATO, E. 1978 Computational experience with quasi-Newton algorithms for minimization problems of moderately large size. *Towards Global Optimization 2* (L. C. W. Dixon and G. P. Szego, eds). Amsterdam: North-Holland, pp 209–219.

WOLFE, P. 1975 A method of conjugate subgradients for minimizing nondifferentiable functions. *Mathematical Programming Study 3* (M. L. Balinski and P. Wolfe, eds). Amsterdam: North-Holland, pp 145–173.