KYBERNETIKA — VOLUME 18 (1982), NUMBER 4

QUASI-NEWTON METHODS WITHOUT PROJECTIONS FOR UNCONSTRAINED MINIMIZATION

LADISLAV LUKŠAN

This contribution contains a description of a class of quasi-Newton methods without projections for unconstrained minimization which are modifications of quasi-Newton methods with projections proposed by Davidon. An algorithm which realizes a class of quasi-Newton methods without projections is given and its efficiency is demonstrated on test functions.

1. INTRODUCTION

Recently Davidon [1] has proposed a class of quasi-Newton methods which can find a minimum of the quadratic function

$$F(x) = \tfrac{1}{2}(x - \tilde{x})^{\mathrm{T}} G(x - \tilde{x})$$

after a finite number of iterations with inexact line searches and which are efficient for general unconstrained minimization. Here x is an n-dimensional vector and G is a symmetric positive definite matrix of the order n. Davidon's methods are iterative methods whose iteration has a form

$$(1.1)\qquad \begin{cases} x^+ = x - \varrho H g \\ u^+ = \beta u - \alpha v \\ H^+ = H + \dfrac{1}{b} Pd(Pd)^{\mathrm{T}} - \dfrac{1}{a} PHy(PHy)^{\mathrm{T}} + \\ \qquad + \dfrac{\vartheta b}{a}\left(\dfrac{a}{b} Pd - PHy\right)\left(\dfrac{a}{b} Pd - PHy\right)^{\mathrm{T}} \end{cases}$$

where g is the gradient of the objective function $F(x)$ at the point x, H is a symmetric positive definite matrix of the order n and ϱ is a steplength such that $F(x^+) < F(x)$.

(We use the notation $x^+ = x - \varrho Hg$ instead of the standard notation $x_{k+1} = x_k - \varrho_k H_k g_k$, $k = 1, 2, \ldots$). At the same time

$$\alpha = u^T y .$$
$$\beta = v^T y$$

and

$$a = (PHy)^T H^{-1} PHy$$
$$b = (PHy)^T H^{-1} Pd$$
$$c = (Pd)^T H^{-1} Pd$$

where $d = x^+ - x$, $y = g^+ - g$, $v = d - Hy$ and where P is a projection matrix into the subspace spanned by vectors u and v. H is an arbitrary symmetric positive definite matrix and $u = Hg$ in the first iteration.

The projection matrix P is the essential feature of Davidon's methods but it complicates the algorithm of a class of these methods. We can write the projection matrix P in the form

$$P = V(V^T H^{-1} V)^{-1} V^T H^{-1}$$

where $V = [u, v]$ is a matrix which has two columns u and v. At the same time the matrix recurrence formula in (1.1) can be written in the form

$$H^+ = H + PUA(PU)^T$$

where $U = [d, Hy]$ and A is a 2×2 matrix which must be taken so that the quasi-Newton condition $H^+ y = d$ may hold. If we combine both above expressions we obtain

$$H^+ = H + VBV^T$$

where

$$B = (V^T H^{-1} V)^{-1} VH^{-1} UAU^T H^{-1} V(V^T H^{-1} V)^{-1} .$$

We need not use the last expression for 2×2 matrix B but the matrix B must be taken to keep the quasi-Newton condition $H^+ y = d$ valid. In this way we obtain a one-parameter class of quasi-Newton methods with parameter φ whose iteration has the form

$$\text{(1.2)} \qquad \begin{cases} x^+ = x - \varrho Hg \\ u^+ = \beta u - \alpha v \\ H^+ = H + 1/\beta \, (vv^T - \varphi u^+ (u^+)^T) \end{cases}$$

The matrix recurrence formula in (1.2) is much easier than that in (1.1) but there exists no simple choice of the parametes φ except for the value $\varphi = 0$ which corresponds to a rank – one update in (1.1).

In this paper we are proposing an analysis of the class of quasi-Newton methods without projections whose iteration has the form (1.2). Most attention will be devoted to the choice of the parameter φ. Also several various forms of the quasi-Newton methods without projections will be shown. Moreover, we shall describe an algorithm which combines several quasi-Newton methods without projections and show the results of numerical experiments with this algorithm.

2. ANALYSIS OF QUASI-NEWTON METHODS WITHOUT PROJECTIONS

This section is devoted to the choice of the parameter φ in (1.2). We use notations

$$\varepsilon = u^T H^{-1} u$$
$$\sigma = u^T H^{-1} v$$
$$\tau = v^T H^{-1} v$$

and

$$\gamma = u^T H^{-1} d = \sigma + \alpha$$
$$\delta = v^T H^{-1} d = \tau + \beta$$

and

$$A = \beta^2(\varepsilon\tau - \sigma^2)$$
$$B = \beta\delta(\varepsilon\tau - \sigma^2)$$
$$C = \delta^2(\varepsilon\tau - \sigma^2)$$
$$D = (\beta\sigma - \alpha\tau)^2 = (\delta\sigma - \gamma\tau)^2$$

Positive definiteness of the matrix H implies $\varepsilon > 0$ and $\tau > 0$ for $u \neq 0$ and $v \neq 0$ respectively. Moreover from the Schwartz inequality we obtain $\varepsilon\tau - \sigma^2 \geqq 0$ so that we have $A \geqq 0$ and $C \geqq 0$. Furthermore $D \geqq 0$ as a square of a real number.

First we shall study the conditions for positive definiteness of the matrix H^+.

Lemma 2.1. Let H^+ be a matrix defined in (1.2) where H is a symmetric positive definite matrix of order n and $\beta \neq 0$. Then the matrix $H^{-1/2}H^+H^{-1/2}$ has $n - 2$ unit eigenvalues and each of the other eigenvalues is a solution of the quadratic equation

$$(2.1) \qquad \begin{cases} \lambda^2 - p\lambda + q = 0 \\ \text{where} \\ p = \dfrac{\delta}{\beta} - \dfrac{\varphi}{\beta\tau} \quad (A + D) + 1 \\ q = \dfrac{\delta}{\beta} - \dfrac{\varphi}{\beta\tau} \quad (B + D) \end{cases}$$

Moreover $\lambda_1'/p' \geqq 0$ and $\lambda_2'/p' \geqq 0$ for $\varepsilon\tau - \sigma^2 > 0$ where λ_1', λ_2' and p' are the derivatives of λ_1, λ_2 and p with respect to the parameter φ respectively (λ_1 and λ_2 are the roots of the quadratic equation (2.1)).

Proof. If we use expression $H^+ = H + VBV^T$ we obtain

$$H^{-1/2}H^+H^{-1/2} = I + H^{-1/2}VBV^TH^{-1/2}$$

This matrix has $n - 2$ unit eigenvalues which correspond to $n - 2$ linearly independent eigenvectors from the orthogonal complement of the subspace spanned by columns of the matrix $H^{-1/2}V$. Remaining two eigenvectors can be written in the form $H^{-1/2}Vz$ so that the corresponding eigenvalues must satisfy the equation

$$(I + BVH^{-1}V)\, z = \lambda z$$

or

$$\det\left((1 - \lambda) I + BV^{\mathrm{T}}H^{-1}V\right) = 0$$

where I is the unit matrix of order 2. The expression $H^{+} = H + VBV^{\mathrm{T}}$ is equivalent to the matrix recurrence formula in (1.2) only if

$$B = \begin{bmatrix} -\varphi\beta, & \varphi\alpha \\ \varphi\alpha, & (1 - \varphi\alpha^2)/\beta \end{bmatrix}$$

After substituting this matrix into the above determinant equation, we obtain (2.1) after some algebraic manipulations.

Now let $\varepsilon\tau - \sigma^2 > 0$. Then from (2.1) we have $p' = (\mathrm{A} + D)/\beta\tau \neq 0$. Differentiating the quadratic equation (2.1) with respect to the parameter φ we obtain

$$2\lambda\lambda' - p'\lambda - p\lambda' + q' = 0$$

If $2\lambda - p = 0$ further differentiating gives $2\lambda'(\lambda' - p') = 0$ (since $p'' = 0$ and $q'' = 0$ from (2.1)) and $\lambda'_1/p' \geqq 0$ and $\lambda'_2/p' \geqq 0$ hold. If $2\lambda - p \neq 0$ then

$$\lambda' = \frac{p'}{2}\left(1 + \frac{\dfrac{p}{2} - \dfrac{q'}{p'}}{\lambda - \dfrac{p}{2}}\right)$$

After substituting the roots λ_1 and λ_2 of the quadratic equation (2.1) into the last expression we can see that $\lambda'_1/p' \geqq 0$ and $\lambda'_2/p' \geqq 0$ if and only if

$$\left|\frac{p}{2} - \frac{q'}{p'}\right| \leqq \sqrt{\left(\left(\frac{p}{2}\right)^2 - q\right)}$$

which is equivalent to the inequality

$$q(p')^2 - pp'q' + (q')^2 \leqq 0$$

If we use (2.1), we have after some algebraic manipulations, $q(p')^2 - pp'q' + (q')^2 = -(\varepsilon\tau - \sigma^2)\,\mathrm{D}/\beta^2 \leqq 0$ and the Lemma is proved. □

Theorem 2.1. Let H^{+} be a matrix defined in (1.2) where H is a symmetric positive definite matrix of order n and $\beta \neq 0$. When $\varepsilon\tau - \sigma^2 = 0$ then H^{+} is positive definite if and only if $\beta\delta > 0$. When $\varepsilon\tau - \sigma^2 > 0$ then H^{+} is positive definite if and only if $\mathrm{B} + \mathrm{D} > 0$ and $q > 0$, where q is defined by (2.1).

Proof. Lemma 2.1 implies that H^{+} is positive definite if and only if $p > 0$ and $q > 0$. Let $\varepsilon\tau - \sigma^2 = 0$ Then u and v are linearly dependent i.e. $u = \mu v$ for some μ so that $\varepsilon = \mu\sigma$, $\sigma = \mu\tau$, $\alpha = \mu\beta$, $\gamma = \mu\delta$ and therefore $\mathrm{A} + \mathrm{D} = 0$, $\mathrm{B} + \mathrm{D} = 0$.

Thus we have

$$p - 1 = q = \frac{\delta}{\beta}$$

from (2.1) so that H^+ is positive definite if and only if $\beta\delta > 0$. Now let $\varepsilon\tau - \sigma^2 > 0$. Then three cases occur:

(a) If $\mathrm{B} + \mathrm{D} = 0$ then $q > 0$ implies $\beta\delta > 0$ so that $\mathrm{B} + \mathrm{D} \geqq \mathrm{B} > 0$, which is the contradiction.

(b) If $\mathrm{B} + \mathrm{D} < 0$ then $\mathrm{A} + \mathrm{D} > 0$ (since $\mathrm{A} + \mathrm{D} \geqq 0$ and the equality $\mathrm{A} + \mathrm{D} = 0$ implies the equality $\mathrm{B} + \mathrm{D} = 0$) and the derivatives p' and q' have opposite signs. Let $p' > 0$ and $q' < 0$ (the proof for $p' < 0$ and $q' > 0$ is analogous). Then for sufficiently small values of the parameter φ we obtain $p < 0$ so that

$$\lambda_1 = \frac{p}{2} - \sqrt{\left(\left(\frac{p}{2}\right)^2 - q\right)} \leqq \frac{p}{2} < 0$$

and for sufficiently large values of the parameter φ we obtain $q < 0$ so that

$$\lambda_1 = \frac{p}{2} - \sqrt{\left(\left(\frac{p}{2}\right)^2 - q\right)} < \frac{p}{2} - \frac{p}{2} = 0$$

But λ_1 is a monotone function of the parameter φ by Lemma 2.1 so that $\lambda_1 < 0$ for all values of the parameter φ and H^+ cannot be positive definite for any value of the parameter φ.

(c) If $\mathrm{B} + \mathrm{D} > 0$ then $\mathrm{A} + \mathrm{D} > 0$ (as in the case (b)) and the conditions $p > 0$ and $q > 0$ are equivalent to the conditions $p(\mathrm{B} + \mathrm{D}) > 0$ and $q(\mathrm{A} + \mathrm{D}) > 0$. But from (2.1) we have

$$p(\mathrm{B} + \mathrm{D}) - q(\mathrm{A} + \mathrm{D}) = \mathrm{C} + \mathrm{D} \geqq 0$$

so that $q > 0$ implies both $p(\mathrm{B} + \mathrm{D}) > 0$ and $q(\mathrm{A} + \mathrm{D}) > 0$ and H^+ is positive definite if and only if $q > 0$. □

Theorem 2.1 gives conditions for positive definiteness of the matrix H^+ in (1.2) which are sufficient for the direction $s^+ = -H^+g^+$ to be a descent one. The parameter φ in (1.2) must be taken in accord with these conditions.

Now we shall study conditioning of the matrix H^+. We shall use the condition number of the matrix $H^{-1/2}H^+H^{-1/2}$ as a measure of conditioning of the matrix H^+.

Lemma 2.2. The roots λ_1 and λ_2 of the quadratic equation (2.1) satisfy inequalities $\lambda_1 \leqq 1 \leqq \lambda_2$ if and only if $\varphi \geqq 0$. Moreover if $\mathrm{B} + \mathrm{D} > 0$ then the quotient λ_2/λ_1 reaches a minimum for

$$\varphi = \frac{\tau^2(\mathrm{D} - \mathrm{B})}{(\mathrm{A} + \mathrm{D})(\mathrm{B} + \mathrm{D})} \tag{2.2}$$

Proof. After substituting the roots λ_1 and λ_2 of the quadratic equation (2.1) into the condition $\lambda_1 \leqq 1 \leqq \lambda_2$ we obtain an equivalent condition

$$\sqrt{\left(\left(\frac{p}{2}\right)^2 - q\right)} \geqq \left|\frac{p}{2} - 1\right|$$

This inequality has the solution $p - q \geqq 1$. But from (2.1) we have $p - q = \varphi(\varepsilon\tau - \sigma^2) + 1$ so that $p - q \geqq 1$ if and only if $\varphi \geqq 0$.

Let $B + D > 0$. Then also $A + D > 0$ and from (2.1) we obtain $p > 0$ and $q > 0$ for the value of the parameter φ given by (2.2). Let's consider only those values of the parameter φ for which $p > 0$ and $q > 0$. Then from the quadratic equation (2.1) we have

$$\frac{\lambda_2}{\lambda_1} = (r + \sqrt{(r^2 - 1)})^2$$

where

$$r = \frac{p}{2\sqrt{q}}$$

and by differentiating we obtain

$$\left(\frac{\lambda_2}{\lambda_1}\right)' = \frac{2r'}{\sqrt{(r^2 - 1)}} (r + \sqrt{(r^2 - 1)})^2$$

Since $r + \sqrt{(r^2 - 1)} \neq 0$, a stationary point of the ratio λ_2/λ_1 is given as a solution of the equation $r' = 0$. It gives the condition

$$2p'q - pq' = 0$$

and if we use (2.1), we obtain (2.2). Moreover

$$\left(\frac{\lambda_2}{\lambda_1}\right)'' = \frac{2r''}{\sqrt{(r^2 - 1)}} (r + \sqrt{(r^2 - 1)})^2$$

for $r' = 0$ so that the stationary point of the ratio λ_2/λ_1 is a minimum if $r'' > 0$. It gives the condition $p'q' > 0$ (since $2p'q - pq' = 0$). But from (2.1) we obtain

$$p'q' = \frac{1}{\beta^2\tau^2} (A + D)(B + D) > 0$$

since $A + D > 0$ and $B + D > 0$. □

Theorem 2.2. Let H^+ be a matrix defined in (1.2) where H is a symmetric positive definite matrix of order n and $\beta \neq 0$. Let $B + D > 0$. Then H^+ is positive definite and the condition number of the matrix $H^{-1/2}H^+H^{-1/2}$ has a minimum if and only if

$$\varphi = \max\left(0, \frac{\tau^2(D - B)}{(A + D)(B + D)}\right)$$

Proof. From the definition of the condition number and from Lemma 2.1 we have

$$K(H^{-1/2}H^{+}H^{-1/2}) = \frac{\max(1, \lambda_2)}{\min(1, \lambda_1)}$$

where λ_1 and λ_2 are the roots of the quadratic equation (2.1). Since we suppose that $\mathrm{B} + \mathrm{D} > 0$ we can use Lemma 2.2 which shows that the quotient λ_2/λ_1 has a minimum for the value of the parameter φ given by (2.2). By the same Lemma $K(H^{-1/2}H^{+}H^{-1/2})$ is equivalent to the quotient λ_2/λ_1 if and only if $\varphi \geqq 0$. Theorem 2.2 then holds for $\mathrm{D} - \mathrm{B} \geqq 0$.

Now if we let $\mathrm{D} - \mathrm{B} < 0$, then $K(H^{-1/2}H^{+}H^{-1/2})$ is not equivalent to the quotient λ_2/λ_1, but the roots λ_1 and λ_2 are monotone functions of the parameter φ and their derivatives have the same signs (see Lemma 2.1). The roots λ_1 and λ_2 move in the same direction when the parameter φ is changed so that $K(H^{-1/2}H^{+}H^{-1/2})$ has a minimum if either $\lambda_1 = 1$ or $\lambda_2 = 1$. This situation appears only when $\varphi = 0$ so that $\varphi = 0$ is an optimal choice of the parameter φ when $\mathrm{D} - \mathrm{B} < 0$. Now $\mathrm{D} - \mathrm{B} < 0$ implies $\mathrm{B} > 0$ so that $\beta\delta > 0$ and if we set $\varphi = 0$ into (2.1) we obtain $q > 0$ so that the conditions for positive definiteness of the matrix H^{+} are not violated. □

Theorem 2.2 gives a special value of the parameter φ which has satisfied the condition for the positive definiteness of the matrix H^{+}. This value will be used in Section 5 for standing the relation between the parameter ϑ in (1.1) and the parameter φ in (1.2).

3. INVERSE FORM OF QUASI-NEWTON METHODS WITHOUT PROJECTIONS

Inverse form of quasi-Newton methods without projections uses the recurrence formula for the matrix $G = H^{-1}$ instead of that for the matrix H in (1.2). We shall use the recurrence formula for the matrix $G = H^{-1}$ for computation of the value $\varepsilon = u^{\mathrm{T}}H^{-1}u$ in Algorithm 6.1.

Theorem 3.1. Let H^{+} be a matrix defined in (1.2) where H is a symmetric positive definite matrix of order n and $\beta\delta \neq 0$. Let $G = H^{-1}$ and $G^{+} = (H^{+})^{-1}$. Then

$$\left\{\begin{aligned} & G^{+} = G - 1/\delta\,(Gv(Gv)^{\mathrm{T}} - \psi\, z^{+}(z^{+})^{\mathrm{T}}) \\ & \text{where} \\ & z^{+} = \delta Gu - \gamma Gv \end{aligned}\right. \tag{3.1}$$

and where ψ is a free parameter which is related to the parameter φ in (1.2) by the formula $\psi = \varphi/q$ (q is defined by (2.1)). Moreover if u^{+} is the vector determined from (1.2) then $G^{+}u^{+} = z^{+}/q$.

Proof. If we use the expression $H^{+} = H + VBV^{\mathrm{T}}$, we obtain

$$G^{+} = G - GV(B^{-1} + V^{\mathrm{T}}GV)^{-1}\, V^{\mathrm{T}}G$$

by the well-known Voodbury theorem. Let $Q = (B^{-1} + V^{\mathrm{T}}GV)^{-1}$. Then the quasi-Newton condition $G^+d = y$ holds only if

$$Q = \begin{bmatrix} -\psi\delta, & \psi\gamma \\ \psi\gamma, & 1/\delta\,(1 - \psi\gamma^2) \end{bmatrix}$$

for some value of the parameter ψ so that (3.1) is proved. The formula $\psi = \varphi/q$ can be obtained by comparing elements of matrices Q and $(B^{-1} + V^{\mathrm{T}}GV)^{-1}$ (the elements of the matrix B are given in Section 2 and the elements of the matrix $V^{\mathrm{T}}GV$ are numbers ε, σ, σ and τ). The formula $G^+u^+ = z^+/q$ follows from (3.1), from the expression $u^+ = \beta u - \alpha v$, from (2.1) and from the formula $\psi = \varphi/q$. □

The inverse form of quasi-Newton methods without projections based on Theorem 3.1 uses the iteration

$$(3.2)\qquad \begin{cases} Gs &= -g \\ x^+ &= x + \varrho s \\ z^+ &= \delta z - \gamma w \\ G^+ &= G - 1/\delta(ww^{\mathrm{T}} - \psi z^+(z^+)^{\mathrm{T}}) \\ \text{where} & \\ w &= Gv = -(\varrho g + y) \end{cases}$$

and where now $\gamma = z^{\mathrm{T}}d$ and $\delta = w^{\mathrm{T}}d$. G is an arbitrary symmetric positive definite matrix of order n and $z = g$ in the first iteration. We must use values $\varepsilon = z^{\mathrm{T}}G^{-1}z$, $\sigma = z^{\mathrm{T}}G^{-1}w$ and $\tau = w^{\mathrm{T}}G^{-1}w$ instead of those given in Section 2.

4. PRODUCT FORM OF QUASI-NEWTON METHODS WITHOUT PROJECTIONS

Product form of quasi-Newton methods without projections uses recurrence formula for a rectangular matrix S of full rank in the factorization $H = SS^{\mathrm{T}}$. It is advantageous for the minimization with linear constraints (see [6]).

Lemma 4.1. Let

$$H + VBV^{\mathrm{T}} = (I + Vab^{\mathrm{T}}V^{\mathrm{T}}H^{-1})\,H(I + H^{-1}Vba^{\mathrm{T}}V)$$

where a and b are two-dimensional vectors. Then

$$ab^{\mathrm{T}} - ba^{\mathrm{T}} = \begin{bmatrix} 0, & +\sqrt{\varphi} \\ -\sqrt{\varphi}, & 0 \end{bmatrix}$$

Proof. After arranging the expression for $H + VBV^{\mathrm{T}}$ and comparing the coefficients we obtain

$$B = [a, b]\begin{bmatrix} b^{\mathrm{T}}V^{\mathrm{T}}H^{-1}Vb, & 1 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} a^{\mathrm{T}} \\ b^{\mathrm{T}} \end{bmatrix}$$

so that

$$\det B = -(\det [a, b])^2 =$$

$$= - \begin{vmatrix} a_1, & b_1 \\ a_2, & b_2 \end{vmatrix}^2 = -(a_1 b_2 - a_2 b_1)^2$$

But

$$ab^T - ba^T = \begin{bmatrix} 0 & a_1 b_2 - a_2 b_1 \\ a_2 b_1 - a_1 b_2, & 0 \end{bmatrix}$$

and $\det B = -\varphi$ (see the proof of Lemma 2.1) so that the Lemma is proved. □

Theorem 4.1. Let H^+ be a matrix defined in (2.1) where H is a symmetric positive definite matrix of order n and $\beta \neq 0$. Let $\varphi \geqq 0$ in (1.2) and let $\mathrm{B} + \mathrm{D} > 0$ and $q > 0$ (q is defined by (2.1)). Then

$$(4.1) \qquad \begin{cases} H^+ = RHR^T \\ \text{where} \\ R = I + \dfrac{(v + \sqrt{(\varphi)}\, u^+)(\sqrt{(q)}\, w - \sqrt{(\varphi)}\, z^+)^T}{\delta + \beta\sqrt{(q)} + (\beta\sigma - \alpha\tau)\sqrt{(\varphi)}} \end{cases}$$

Proof. Conditions $\mathrm{B} + \mathrm{D} > 0$ and $q > 0$ imply the positive definiteness of the matrix H^+. Suppose that $H^+ = RHR^T$ where $R = I + Vab^TV^TH^{-1}$. Then $\det R = \sqrt{q}$ since $\det H^+/\det H = q$ by Lemma 2.1. Since

$$(I + Vab^TV^TH^{-1})^{-1} = I - \frac{1}{\sqrt{q}} Vab^TV^TH^{-1}$$

the quasi-Newton condition $H^+y = d$ can be written in the form

$$H(I + H^{-1}Vba^TV)\, y = \left(I - \frac{1}{\sqrt{q}} Vab^TV^TH^{-1}\right) d$$

or

$$Vba^TV^Ty + \frac{1}{\sqrt{q}} Vab^TV^TH^{-1}d = v$$

Multiplying the last equation by the vectors $(H^{-1}d)^T$ and y^T we obtain

$$\mathrm{R} + \frac{1}{\sqrt{q}}\, \mathrm{P} = \delta$$

$$\mathrm{S} + \frac{1}{\sqrt{q}}\, \mathrm{R} = \beta$$

where

$$\begin{bmatrix} \mathrm{P}, & \mathrm{Q} \\ \mathrm{R}, & \mathrm{S} \end{bmatrix} = U^TH^{-1}Vab^TV^TH^{-1}U$$

Now Lemma 4.1 implies

$$Q - R = d^T H^{-1} V(ab^T - ba^T) V^T y =$$

$$= [\gamma, \delta] \begin{bmatrix} 0, & +\sqrt{\varphi} \\ -\sqrt{\varphi}, & 0 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = (\beta\gamma - \alpha\delta)\sqrt{\varphi} = (\beta\sigma - \alpha\tau)\sqrt{\varphi}$$

Moreover det $(U^T H^{-1} V a b^T V^T H^{-1} U) = 0$ since the matrix ab^T has rank 1 so that

$$PS - QR = 0$$

Now we have four equations for four unknowns P, Q, R, S. These equations have a unique solution which can be written in the form

$$\begin{bmatrix} P, & Q \\ R, & S \end{bmatrix} = \frac{1}{\lambda} \begin{bmatrix} \delta + (\beta\sigma - \alpha\tau)\sqrt{\varphi} \\ \beta \end{bmatrix} \begin{bmatrix} \delta\sqrt{q} \\ \beta\sqrt{q} + (\beta\sigma - \alpha\tau)\sqrt{\varphi} \end{bmatrix}^T$$

if $\lambda = \delta + \beta\sqrt{q} + (\beta\sigma - \alpha\tau)\sqrt{\varphi} \neq 0$. Thus we obtain

$$ab^T = (U^T H^{-1} V)^{-1} \begin{bmatrix} P, & Q \\ R, & S \end{bmatrix} (V^T H^{-1} U)^{-1} =$$

$$= \frac{1}{\lambda} \begin{bmatrix} \beta\sqrt{\varphi} \\ 1 - \alpha\sqrt{\varphi} \end{bmatrix} \begin{bmatrix} -\delta\sqrt{\varphi} \\ \sqrt{q} + \gamma\sqrt{\varphi} \end{bmatrix}^T$$

after some algebraic manipulations (since elements of the matrix $U^T H^{-1} V$ are numbers $\gamma, \delta, \alpha, \beta$). After substituting the matrix ab^T in the expression $R = I + Vab^T V^T H^{-1}$ we obtain (4.1). □

Theorem 4.1 can be used for derivation of the product form of quasi-Newton methods without projections. Let $H = SS^T$ and $H^+ = S^+(S^+)^T$. The product form of quasi-Newton methods without projections uses iteration

$$(4.2) \qquad \begin{cases} x^+ = x - \varrho S\tilde{g} \\ \tilde{a} = \tilde{v} + \sqrt{(\varphi)}(\beta\tilde{u} - \alpha\tilde{v}) \\ \tilde{b} = \sqrt{(q)}\,\tilde{v} - \sqrt{(\varphi)}(\delta\tilde{u} - \gamma\tilde{v}) \\ \tilde{u}^+ = \dfrac{1}{q}\left(I + \dfrac{1}{\lambda}\tilde{b}\tilde{a}^T\right)(\delta\tilde{u} - \gamma\tilde{v}) \\ S^+ = S\left(I + \dfrac{1}{\lambda}\tilde{a}\tilde{b}^T\right) \\ \text{where} \\ \lambda = \delta + \beta\sqrt{(q)} + (\beta\sigma - \alpha\tau)\sqrt{\varphi} \end{cases}$$

and where now $\tilde{g} = S^T g$, $\tilde{d} = -\varrho\tilde{g}$, $\tilde{y} = S^T y$ and $\tilde{v} = \tilde{d} - \tilde{y}$. S is an arbitrary nonsingular matrix and $\tilde{u} = \tilde{g}$ in the first iteration. We must use values $\alpha = \tilde{u}^T\tilde{y}$, $\beta = \tilde{v}^T\tilde{y}$, $\gamma = \tilde{u}^T\tilde{d}$, $\delta = \tilde{v}^T\tilde{d}$, $\varepsilon = \tilde{u}^T\tilde{u}$, $\sigma = \tilde{u}^T\tilde{v}$ and $\tau = \tilde{v}^T\tilde{v}$ instead of those given in Section 2.

5. THE RELATION BETWEEN THE QUASI-NEWTON METHODS WITHOUT PROJECTIONS AND THE DAVIDON'S METHODS

The Davidon's methods use matrix recurrence formula (1.1). If P is the unit matrix of order n, then (1.1) defines the classical variable metric methods. The matrix H^+ in (1.1) is positive definite if and only if $b > 0$ and $\vartheta(ac - b^2) + b > 0$. The last inequality serves for determination of the special values of the parameter ϑ. If P is not the unit matrix then $b > 0$ is not satisfied automatically. We shall show below that the condition $b > 0$ is equivalent to the condition $\mathrm{B} + \mathrm{D} > 0$ which is used in Theorem 2.2.

Now we shall study the relation between the parameter ϑ in (1.1) and the parameter φ in (1.2). It enables us to find some special values of the parameter φ.

Lemma 5.1. Let H be a symmetric positive definite matrix of order n and let $\varepsilon\tau - \sigma^2 > 0$. Then

$$a = \frac{\mathrm{A} + \mathrm{D}}{\tau(\varepsilon\tau - \sigma^2)}$$

$$b = \frac{\mathrm{B} + \mathrm{D}}{\tau(\varepsilon\tau - \sigma^2)}$$

$$c = \frac{\mathrm{C} + \mathrm{D}}{\tau(\varepsilon\tau - \sigma^2)}$$

and

$$ac - b^2 = \frac{\mathrm{D}}{\varepsilon\tau - \sigma^2}$$

Proof. By definition (see Section 1) we have

$$\begin{bmatrix} c, & b \\ b, & a \end{bmatrix} = (PU)^{\mathrm{T}} H^{-1} PU = U^{\mathrm{T}} H^{-1} V (V^{\mathrm{T}} H^{-1} V)^{-1} V^{\mathrm{T}} H^{-1} U =$$

$$= \begin{bmatrix} \gamma, & \delta \\ \alpha, & \beta \end{bmatrix} \begin{bmatrix} \varepsilon, & \sigma \\ \sigma, & \tau \end{bmatrix}^{-1} \begin{bmatrix} \gamma, & \alpha \\ \delta, & \beta \end{bmatrix}$$

The matrix $V^{\mathrm{T}}H^{-1}V$ is nonsingular since $\varepsilon\tau - \sigma^2 > 0$. If we use definitions of A, B, C and D, we prove the Lemma after some algebraic manipulations. □

Lemma 5.1 implies that, provided $\varepsilon\tau - \sigma^2 > 0$, $b > 0$ if and only if $\mathrm{B} + \mathrm{D} > 0$.

Theorem 5.1. Let H be a symmetric positive definite matrix of order n and let $\varepsilon\tau - \sigma^2 > 0$. Let $\beta \neq 0$ and $\mathrm{B} + \mathrm{D} \neq 0$. Then the recurrence formulae (1.1) and (1.2) give the same matrix H^+ if and only if

$$\varphi = \frac{\tau^2 \mathrm{D}}{(\mathrm{A} + \mathrm{D})(\mathrm{B} + \mathrm{D})}(1 - \beta\vartheta) \tag{5.1}$$

Proof. If we compare the expression $H^+ = H + PUA(PU)^T$ with the matrix recurrence formula in (1.1), we observe that the elements of the matrix A are linear functions of the parameter ϑ. Since elements of the matrix B depend linearly on elements of the matrix A, we can suppose that the parameter φ is a linear function of the parameter ϑ i.e.

$$\varphi = \mu\vartheta + \nu$$

for some μ and ν. If we compare the optimally conditioned quasi-Newton method without projections which is given by Theorem 2.2 with the optimally conditioned Davidon's method proposed in [1], we obtain equations

$$\mu \frac{1}{b - a} + \nu = 0$$

$$\mu \frac{c - b}{ac - b^2} + \nu = \frac{\tau^2(D - B)}{(A + D)(B + D)}$$

whose solution gives (5.1). □

Now we can determine values of the parameters φ and ψ in (1.2) and (3.2) which correspond to standard values of the parameter ϑ. These values are given in Table 1.

Table 1.

	Parameter ϑ	Parameter φ	Parameter ψ
1	0	$\frac{\tau^2 D}{(A + D)(B + D)}$	$\frac{\tau^2 D}{(B + D)^2}$
2	$\frac{1}{b}$	$\frac{\tau^2 D}{(B + D)^2}$	$\frac{\tau^2 D}{(B + D)(C + D)}$
3	$\frac{1}{b + a}$	$\frac{2\tau^2 D}{(A + B + 2D)(B + D)}$	$\frac{2\tau^2 D}{(B + D)(B + C + 2D)}$
4	$\frac{a - b}{ac - b^2}$	$\frac{\tau^2}{B + D}$	$\frac{\tau^2}{B + D}$
5	$\frac{c - b}{ac - b^2}$	$\frac{\tau^2(D - B)}{(A + D)(B + D)}$	$\frac{\tau^2(D - B)}{(B + D)(C + D)}$
6	$\frac{1}{b - a}$	0	0

First five values in the Table 1 satisfy the condition $q > 0$ which is necessary for the positive definiteness of the matrix H^+. The last value satisfies this condition only if $\beta\delta > 0$.

6. THE IMPLEMENTATION OF QUASI-NEWTON METHODS WITHOUT PROJECTIONS

There are three possible ways of implementing quasi-Newton methods without projections. We can use the basic iteration (1.2) or the inverse form iteration (3.2) or the product form iteration (4.2). In this section we shall describe an algorithm which uses safeguarded basic iteration. Let us state several notes first:

1. If the search direction $s = -Hg$ does not satisfy the condition $-s^{\mathrm{T}}g \geqq \varepsilon_0 \|s\| \|g\|$ for a small positive number ε_0, the algorithm must be restarted.
2. The steplength ϱ is chosen by the standard procedure so that the condition $(1 - \varepsilon_1) \varrho s^{\mathrm{T}} g \leqq F^+ - F \leqq \varepsilon_1 \varrho s^{\mathrm{T}} g$ holds. The safeguarded cubic interpolation with the initial estimate $\varrho = \min(1, 4(\tilde{F} - F)/s^{\mathrm{T}}g)$ is used, where $\tilde{F}$ is a lower bound of a minimum value of the objective function.
3. If $\mathrm{B} + \mathrm{D} \leqq 0$ the basic iteration cannot be used without sacrificing the desired positive definiteness of the matrix H^+. In this case we use another matrix recurrence formula

$$H^+ = H + \frac{2dd^{\mathrm{T}}}{y^{\mathrm{T}}d} - \frac{(d + Hy)(d + Hy)^{\mathrm{T}}}{y^{\mathrm{T}}(d + Hy)} \tag{6.1}$$

given in [3].

4. The selection of the quasi-Newton method without projections is controlled by the value of the integer m. If $m = 1$, $m = 2$, $m = 3$ or $m = 4$, the values of the parameter φ from the m-th row in the Table 1 are used respectively. If $m = 5$, the optimal value given by Theorem 2.2 is used. If $m = 6$ we use the value $\varphi = 0$ if $\beta\delta > 0$, and the value from the fifth row in the Table 1 otherwise.
5. Let $h = \|d\|$. Then $\|d\| = O(h)$, $\|y\| = O(h)$ and $\|v\| = O(h)$. Assume also that $\|u\| = O(h)$. Then all values α, β, γ, δ, ε, σ and τ are $O(h^2)$, the value $\varepsilon\tau - \sigma^2 = O(h^4)$ and the values A, B, C and D are $O(h^8)$. Therefore some scaling of these values must be used.

Now we are in a position to describe the complete algorithm.

Algorithm 6.1.

Step 1: Determine the initial vector x and compute values $F := F(x)$ and $g := g(x)$. Set $k := 0$.

Step 2: Test for convergence. If the termination criteria are satisfied (for example if $\|g\|$ is sufficiently small) then stop.

Step 3: In the first iteration (when $k = 0$) go to step 4 else go to step 6.

Step 4: Restart. Determine a symmetric positive definite matrix H of order n (for example $H := I$, where I is the unit matrix of order n). Store diagonal elements G_{ii} of the matrix $G = H^{-1}$ when the first derivatives of the objective function are not given analytically.

Step 5: Set $u := Hg$, $z := g$ and $s := -u$. Set $l := 0$ and go to step 20.

Step 6: Set $s := g - g_1$, $v := x - x_1 - Hs$ and $w := -\varrho g_1 - s$ and compute $\tau := v^T w$. If $\tau \leqq 0$ go to step 4 else go to step 7.

Step 7: Compute $\varepsilon := u^T z$. If $\varepsilon \leqq 0$ go to step 8 else go to step 9.

Step 8: If $l = 0$ go to step 4 else set $l := 0$, $u := Hg_1$, $z := g_1$ and go to step 7.

Step 9: Set $\lambda := \sqrt{(\tau/\varepsilon)}$, $u := \lambda u$, $z := \lambda z$, $\varepsilon := \tau$ and compute $\alpha := s^T u$, $\beta := s^T v$ and $\sigma := u^T w$. If $\beta = 0$ go to step 18 else go to step 10.

Step 10: Set $\alpha := \alpha/\tau$, $\beta := \beta/\tau$, $\sigma := \sigma/\tau$, $\gamma := \alpha + \sigma$, $\delta := \beta + 1$ and $\omega := 1 - \sigma^2$. If $\omega \leqq 0$ go to step 8 else go to step 11.

Step 11: Set $\mathrm{A} := \beta^2 \omega$, $\mathrm{B} := \beta\delta\omega$ and $\mathrm{D} := (\beta\sigma - \alpha)^2$. If $\mathrm{B} + \mathrm{D} \leqq 0$ go to step 12 else go to step 13.

Step 12: If $l = 0$ go to step 18 else set $l := 0$, $u := Hg_1$, $z := g_1$ and go to step 7.

Step 13: Select the value of the parameter φ from Table 1 according to the value of the integer m (see note 4 above). If either $\varphi < 0$ or $\varphi > 10^4$ go to step 14 else go to step 15.

Step 14: If $\beta\delta \leqq 0$ go to step 12 else set $\varphi := 0$ and go to step 15.

Step 15: Set $q := (\delta - \varphi(\mathrm{B} + \mathrm{D}))/\beta$. If $q \leqq 0$ go to step 4 else go to step 16.

Step 16: Set $l := 1$. Set $u := \beta u - \alpha v$, $z := (\delta z - \gamma w)/q$, $\beta := \tau\beta$ and $H := H + (vv^T - \varphi uu^T)/\beta$. Set $\psi := \varphi/q$, $\delta := \tau\delta$ and adjust the diagonal elements G_{ii} of the matrix $G = H^{-1}$ by the rule $G_{ii} := G_{ii} - (w_i^2 - \psi z_i^2)/\delta$, when the first derivatives of the objective function are not given analytically.

Step 17: Set $s := -Hg$ and go to step 20.

Step 18: Set $v := x - x_1$ and $w := Hs$. Compute $\sigma := s^T v$ and $\tau := s^T w$. If either $\sigma \leqq 0$ or $\tau \leqq 0$ go to step 4 else go to step 19.

Step 19: Set $w := v + w$ and $H := H + 2vv^T/\sigma - ww^T/(\sigma + \tau)$. Set $w := s - \varrho g_1$, compute $\varepsilon := -\varrho v^T g_1$ and adjust the diagonal elements G_{ii} of the matrix $G = H^{-1}$ by the rule $G_{ii} := G_{ii} + 2s_i^2/\sigma - w_i^2/(\sigma + \varepsilon)$, when the first derivatives of the objective function are not given analytically. Go to step 5.

Step 20: If $-s^T g \geqq 10^{-3}\|s\| \|g\|$ go to step 21 else go to step 4.

Step 21: Set $x_1 := x$, $g_1 := g$ and $F_1 := F$. Use a standard procedure to determine the steplength ϱ so that $0{\cdot}99\varrho s^T g_1 \leqq F - F_1 \leqq 0{\cdot}01\varrho s^T g_1$ holds, where F and g are the new values $F := F(x)$ and $g := g(x)$ at the point $x := x_1 + \varrho s$. (These values are determined in the present step by use of a standard procedure).

Step 22: Set $k := k + 1$ and go to step 2.

The diagonal elements G_{ii} of the matrix $G = H^{-1}$ adjusted in the steps 4, 16 and 19 of Algorithm 6.1 serve to the determination of optimal differences for computation of the first derivatives of the objective function by the method of Stewart [7]. Algorithm 6.1 uses three integers k, l, m. Here k is an iteration count, l is a working integer which indicates that the basic iteration was successful and m is a controlling parameter for the choice of a quasi-Newton method without projections specified by user. In the step 21 of Algorithm 6.1 we can use any procedure for the determination of the steplength (see note 2 above).

Algorithm 6.1 uses one symmetric matrix H of order n and 9 n-dimensional vectors $x, x_1, g, g_1, s, u, v, w$ and z, i.e. it needs approximately $n(n + 19)/2$ words in computer storage. Each basic iteration of Algorithm 6.1 uses two matrix by vector products, 9 scalar by vector products, 6 inner products and two symmetric tensor products i.e. approximately $n(3n + 16)$ multiplications.

7. NUMERICAL EXPERIMENTS

Efficiency of Algorithm 6.1 was tested by means of 18 examples proposed in [5]. Results of these tests are shown in Table 2.

Table 2.

	Quasi-Newton methods without projections					
	$m = 1$	$m = 2$	$m = 3$	$m = 4$	$m = 5$	$m = 6$
1	45–51	45–50	46–51	48–53	45–51	45–51
2	45–89	31–68	26–65	45–118	26–58	60–93
3	34–41	36–49	35–44	28–38	41–52	40–52
4	106–127	48–64	50–72	44–67	53–72	47–62
5	39–43	37–41	50–56	36–41	37–41	35–39
6	42–45	47–50	44–47	78–88	41–45	62–67
7	A	35–47*	57–68*	52–59*	37–41*	A
8	6–13	6–13	6–13	6–13	6–13	6–13
9	20–28	17–25	17–26	18–30	17–25	17–26
10	194–260	118–157	131–173	112–177	133–187	136–190
11	A	105–141	108–144	136–182	83–120	87–125
12	123–185	107–237	100–194	110–298	122–176	122–188
13	110–128	78–116	82–103	104–149	84–99	198–245
14	A	106–152	100–125	135–276	123–149	176–221
15	183–213	93–109	162–181	148–196	141–159	114–127
16	42–93	47–112	42–96	41–93	43–93	41–91
17	45–83	44–83	45–83	50–89	44–82	46–85
18	4–8	4–8	4–8	4–8	4–8	4–8

Each column in the Table 2 corresponds to a value of the integer m (choice of quasi-Newton method without projections). Each row in the Table 2 corresponds to one example (our numbering 1–18 is identical with [5]). A pair of values in the Table 2 which are separated by a stroke are the number of iterations and the number of function evaluations. An asterisk in the row 7 shows that an alternative local minimum was found (instead of global minimum). The letter A indicates that 300 iterations did not suffice to find a minimum.

To compare known methods for the unconstrained minimization Table 3 has been set. Columns of Table 3 correspond to the new method with $m = 5$, the original algorithm of Davidon [1] and the classical variable metric methods which use $P = I$ in (1.1). (The DFP method, the BFGS method and the method of Hoshino use the

Table 3.

	New method $m = 5$	Davidon [1]	DFP	BFGS	Hoshino [3]
1	45—51	50—55	217—232	69—75	56—62
2	26—58	37—94	31—72	32—74	26—66
3	41—52	41—57	34—40	36—49	34—43
4	53—72	48—78	103—125	48—63	55—73
5	37—41	34—40	102—119	50—56	50—56
6	41—45	39—42	A	50—53	80—86
7	37—41*	43—55*	A	39—45*	47—54*
8	6—13	8—22	6—13	6—13	6—13
9	17—25	18—34	79—93	20—29	25—33
10	133—187	149—261	A	107—146	117—140
11	83—120	95—146	A	188—224	227—263
12	122—176	122—220	101—174	103—242	98—196
13	84—99	78—123	268—295	108—144	133—148
14	123—149	101—171	A	114—163	135—159
15	141—159	151—181	A	132—147	136—148
16	43—93	46—289	48—85	42—91	48—87
17	44—82	63—304	51—88	47—84	48—85
18	4—8	5—12	2—7	2—7	2—7

Table 4.

	Quasi-Newton methods without projections					
	$m = 1$	$m = 2$	$m = 3$	$m = 4$	$m = 5$	$m = 6$
1	68—212	59—186	68—210	65—200	49—152	41—127
2	39—162	42—215	35—154	49—212	37—147	35—153
3	69—240	53—192	43—146	51—205	52—177	35—116
4	95—515	49—297	62—371	58—328	64—374	71—409
5	41—214	45—240	37—193	44—239	32—170	35—185
6	68—363	56—301	54—283	55—307	45—241	52—275
7	A	38—306*	A	A	A	A
8	7—63	7—62	7—63	7—62	7—63	7—62
9	19—158	17—143	19—148	18—139	18—137	19—130

values of the parameter ϑ from the first, the second and the third row in the Table 1 respectively). The meaning of numbers in the Table 3 is the same as in the Table 2. Finally Table 4 contains results for finite difference versions of the quasi-Newton methods without projections when the first derivatives were computed by the method of Stewart [7]. Only first 9 tests are given in the Table 4. The meaning of numbers in the Table 4 is the same as in the Table 2 again.

The same termination criteria, namely $\|g_k\| \leqq 10^{-8}$ or $F_k \leqq 10^{-16}$ or $\|x_k - x_{k-1}\| \leqq 10^{-8}$ and $\|x_{k-1} - x_{k-2}\| \leqq 10^{-8}$ were used for all methods in Tables 2–4.

8. CONCLUSION

The numerical experiments show the high efficiency of the new algorithm. It has about 50% less storage requirements and uses about 30% less number of multiplications than the original algorithm of Davidon. The new algorithm also allows to use the Stewart's method for optimal choice of differences and it is very effective when analytical derivatives are not given. The new algorithm has been implemented in the software package for optimization and nonlinear approximation SPONA (see [4]) as program POPT 49. Moreover the inverse form of quasi-Newton methods without projections has been implemented as program POPT 50. It uses updating the Choleski decomposition of the matrix G and the double dog-leg strategy proposed in [2].

(Received October 20, 1981.)

REFERENCES

[1] W. C. Davidon: Optimally conditioned optimization algorithms without line searches. Math. Programming *9* (1975), 1, 1–30.

[2] J. E. Dennis, H. H. W. Mei: An Unconstrained Optimization Algorithm which Uses Function and Gradient Values. Res. Rept. No. TR 75–246, Dept. of Computer Sci., Cornell University, Ithaca 1975.

[3] S. Hoshino: A formulation of variable metric methods. J. Inst. Math. Appl. *10* (1972), 3 394–403.

[4] L. Lukšan: Software package for optimization and nonlinear approximation. Proc. of 2nd IFAC/IFIP Symposium on software for computer control, Prague 1979.

[5] L. Lukšan: New combined method for unconstrained minimization. Computing *28* (1982), 2, 155–169.

[6] L. Lukšan: Quasi-Newton methods without projections for linearly constrained minimization. Kybernetika *18* (1982), 4, 307–319.

[7] G. W. Stewart: A modification of Davidon's minimization method to accept difference approximation of derivatives. J. Assoc. Comput. Mach. *14* (1967), 1, 72–83.

Ing. Ladislav Lukšan, CSc., Středisko výpočetní techniky ČSAV (General Computing Centre — Czechoslovak Academy of Sciences), Pod vodárenskou věží 4, 182 07 Praha 8. Czechoslovakia.