Masanori Ozawa; Hiroshi Yanai

Some notes on the quasi-Newton methods

Aplikace matematiky, Vol. 27 (1982), No. 6, 433–445

Persistent URL: http://dml.cz/dmlcz/103990

SOME NOTES ON THE QUASI-NEWTON METHODS

MASANORI OZAWA and HIROSHI YANAI

(Received April 17, 1981)

1. INTRODUCTION

Various algorithms of unconstrained optimization problems are known as members of the Quasi-Newton methods. The main idea of the Quasi-Newton methods is to use conjugate directions associated with the Hessian matrix of the objective function. This idea was first introduced into optimization by Davidon [3]. Many papers followed this pioneering work: Broyden [1], [2], Peason [11], Powell [12], Fletcher [7], and so forth.

These papers developed computational techniques as well as theoretical consideration of their own algorithms. However, so far as the authors know, there are only a few papers which treated the heuristics of various methods of the Quasi-Newton type and/or theoretical relations among them. Yanai [14] tried to organize a class of Quasi-Newton methods as a special case of the Gram-Schmidt orthogonalization method.

This paper also attempts to clarify the heuristics and to organize a class of Quasi-Newton methods by specifying the general solutions of matrix equations.

2. THE UNCONSTRAINED OPTIMIZATION PROBLEM AND THE FUNDAMENTAL IDEA OF QUASI-NEWTON METHODS

Throughout this paper, we consider the minimization problem of the function

$$f(\boldsymbol{x}) = \tfrac{1}{2}\boldsymbol{x}^{\mathsf{T}}A\boldsymbol{x} + \boldsymbol{b}^{\mathsf{T}}\boldsymbol{x} + c\,, \tag{1}$$

where $\boldsymbol{x}, \boldsymbol{b} \in \mathbb{R}^n$, A is an $n \times n$ symmetric positive definite matrix and c is a scalar.

We assume that we can evaluate only $\operatorname{grad} f(\boldsymbol{x})$ corresponding to any given $\boldsymbol{x} \in \mathbb{R}^n$. Besides searching for the minimal point $\bar{\boldsymbol{x}}$, we attempt to specify the matrix A and the vector $\boldsymbol{b}$, which determine the function $f(\boldsymbol{x})$ itself.

If we know the values of grad $f(\boldsymbol{x})$ at several points $\boldsymbol{x}^1, \boldsymbol{x}^2, \ldots$, we can determine all these factors — $\bar{\boldsymbol{x}}$, A and $\boldsymbol{b}$. Indeed, for example, assume that $\boldsymbol{x}^1, \boldsymbol{x}^2, \ldots, \boldsymbol{x}^{n+1}$ are in general position*) and

(2) $$\boldsymbol{g}^i := \operatorname{grad} f(\boldsymbol{x}^i), \quad i = 1, 2, \ldots, n+1 .$$

Since the gradient vector of (1) has the form

(3) $$\boldsymbol{g}^i = A\boldsymbol{x}^i + \boldsymbol{b},$$

we obtain linear relations

(4) $$\boldsymbol{y}^i = A\boldsymbol{z}^i, \quad i = 1, 2, \ldots, n,$$

where

(5) $$\boldsymbol{y}^i = \boldsymbol{g}^{i+1} - \boldsymbol{g}^i$$

and

(6) $$\boldsymbol{z}^i = \boldsymbol{x}^{i+1} - \boldsymbol{x}^i .$$

The relations (4) are combined into

(7) $$Y = AZ,$$

where Y and Z are matrices of the forms:

(8) $$Y = [\boldsymbol{y}^1 \vdots \boldsymbol{y}^2 \vdots \ldots \vdots \boldsymbol{y}^n],$$

(9) $$Z = [\boldsymbol{z}^1 \vdots \boldsymbol{z}^2 \vdots \ldots \vdots \boldsymbol{z}^n].$$

Since $\boldsymbol{x}^1, \boldsymbol{x}^2, \ldots, \boldsymbol{x}^{n+1}$ are in general position, the matrix Z is nonsingular; A is obtained by

(10) $$A = YZ^{-1} .$$

Again by (3), we can evaluate $\boldsymbol{b}$ by any $\boldsymbol{x}^i$ and $\boldsymbol{g}^i$ as

$$\boldsymbol{b} = \boldsymbol{g}^i - A\boldsymbol{x}^i .$$

Using the factors obtained above, we can now evaluate the minimal point $\bar{\boldsymbol{x}}$ as

$$\bar{\boldsymbol{x}} = -A^{-1}\boldsymbol{b} .$$

We have now seen how the matrices A and/or A^{-1} can be determined by the gradients of $f(\boldsymbol{x})$ evaluated at $n+1$ points in general position. In Quasi-Newton methods, however, recurrence relations are constructed to generate sequences of matrices converging to A^{-1} in a finite number of steps:

(11) $$H_{k+1} = \Phi(H_k, H_{k-1}, \ldots, H_0;\ \boldsymbol{g}^{k+1}, \boldsymbol{g}^k, \ldots, \boldsymbol{g}^1),$$

*) $n+1$ vectors in $\mathbb{R}^n$, $\boldsymbol{x}^1, \boldsymbol{x}^2, \ldots, \boldsymbol{x}^{n+1}$ are in *general position* if $\boldsymbol{x}^1 - \boldsymbol{x}^{n+1}, \boldsymbol{x}^2 - \boldsymbol{x}^{n+1}, \ldots, \boldsymbol{x}^n - \boldsymbol{x}^{n+1}$ are linearly independent.

(12) $$H_k \to A^{-1}.$$

The gradient vectors are evaluated at the points given by

(13) $$x^{k+1} = x^k - \mu_k H_k g^k, \quad k = 1, 2, 3, \ldots.$$

We believe that the recursive methods have been introduced in Quasi-Newton methods firstly as a reflection of the *traditional* steepest descent methods. The second reason seems to be the intension to apply the Quasi-Newton methods to non-quadratic objective functions, in which the *local* Hessian matrix is of interest only at the minimal point.

3. ARTIFICIAL CONDITIONS

As we have mentioned in the preceding section, the Quasi-Newton methods are presented by the scheme (11)~(13). *Design* of a Quasi-Newton method is determined by a specification of the transformation Φ in (11). However, for the convenience of the specification, several artifical conditions are introduced in most algorithms existing. The following condition is the most common:

(C−1) $$H_{k+1} y^i = z^i, \quad i = 1, 2, \ldots, k,$$

or in matrix notation,

(C−1′) $$H_{k+1} Y_k = Z_k,$$

where

(14) $$Y_k = [y^1 \vdots y^2 \vdots \ldots \vdots y^k], \quad n \times k,$$

(15) $$Z_k = [z^1 \vdots z^2 \vdots \ldots \vdots z^k], \quad n \times k.$$

This condition implies that the matrix H_{k+1} includes all the information about the objective function obtained so far by evaluating the gradients at $x^1, x^2, \ldots, x^{k+1}$.

By solving the matrix equation (C − 1′) (cf. Appendix B), we obtain as the general form of H_{k+1},

(16) $$H_{k+1} = T_k(I - Y_k(U_k^T Y_k)^{-1} U_k^T) + Z_k(Z_k^T Y_k)^{-1} Z_k^T,$$

where T_k is an arbitrary $n \times n$ matrix, whereas U_k is an $n \times k$ matrix with

(17) $$\det(U_k^T Y_k) \neq 0.$$

Notice also that the first term on the right hand side of (16) is the general solution of the homogeneous equation

(18) $$H_{k+1} Y_k = 0$$

while the second is a special solution of (C − 1′).

Since H_{k+1} converges to a symmetric matrix A^{-1}, it is quite natural to construct the recurrence relation (11) so that H_{k+1}'s are also symmetric:

$$(C-2) \qquad H_{k+1} = H_{k+1}^{\mathsf{T}}, \quad k = 1, 2, \ldots, n.$$

In (16), the second term on the right hand side is symmetric since

$$(20) \qquad Z_k(Z^{\mathsf{T}}Y_k)^{-1} Z_k^{\mathsf{T}} = Z_k(Z_k^{\mathsf{T}}AZ_k)^{-1} Z_k^{\mathsf{T}}.$$

Hence it is only necessary that the first term be also symmetric in order to have a symmetric H_{k+1}. In order to have symmetric H_{k+1}, we put, without loss of generality,

$$(21) \qquad T_k = (I - Y_k(U_k^{\mathsf{T}}Y_k)^{-1} U_k^{\mathsf{T}})^{\mathsf{T}} N_k,$$

where N_k is an arbitrary $n \times n$ symmetric matrix.

In addition to the symmetry of H_{k+1}, we introduce the condition of positive definiteness since A and hence A^{-1} are positive definite. However, since A is positive definite, so is H_{k+1} provided N_k is positive definite (sufficient condition, cf. Appendix C). Thus we introduce

$$(C-3) \qquad N_k : \text{positive definite}, \quad k = 1, 2, \ldots, n.$$

In most of the existing algorithms, however, only a single matrix N is used as N_k:

$$(C-4) \qquad N_k = N, \quad k = 1, 2, 3, \ldots, n,$$

where N is an arbitrary $n \times n$ symmetric positive definite matrix.

4. CONSTRUCTION OF THE POINTS $\{x^i\}$

In the preceding sections we assumed that we start with a set of $n + 1$ points $\{x^i\}$ in *general position*. The next lemma gives a sufficient condition that (13) provides such points in a sequence.

Lemma 1. *If* (C−1) *and* (C−2) *are fulfilled, and if*

$$z^i \neq 0, \quad i = 1, 2, \ldots, k \ (\leqq n)$$

and

$$z^{i\mathsf{T}} g^{i+1} = 0, \quad i = 1, 2, \ldots, k \ (\leqq n),$$

then $z^1, z^2, \ldots, z^{k+1}$ *are mutually conjugate with respect to* A:

$$(22) \qquad z^{i\mathsf{T}} A z^{i+l} = 0 \quad \text{for} \quad i < i + l \leqq n,$$

and hence they are linearly independent.

Proof. $z^1, z^2, \ldots, z^{k+1}$ are linearly independent if they are mutually conjugate with respect to a positive definite matrix A. Hence, it suffices to show (22).

On the other hand,

$$z^{i\mathsf{T}} A z^{i+l} = -\mu_{i+l} z^{i\mathsf{T}} g^{i+l} \tag{23}$$

holds for all $i < i + l \leqq n$. In fact, by (13) and (16),

$$z^{i\mathsf{T}} A z^{i+l} = -\mu_{i+l} z^{i\mathsf{T}} A H_{i+l} g^{i+l}. \tag{24}$$

But since

$$y^{i\mathsf{T}} = z^{i\mathsf{T}} A \tag{(4) bis}$$

and

$$y^{i\mathsf{T}} H_{i+l} = z^{i\mathsf{T}} \tag{25}$$

by (C−1) and (C−2), we obtain (23).

Accordingly, the relation (22) is obtained if

$$z^{i\mathsf{T}} g^{i+l} = 0 \tag{26}$$

is proved.

The relation (26) is proved inductively. In fact, for $l = 1$, (26) coincides with the proposition of the lemma.

Assume that (26) and hence (23) hold for $l = 1, 2, \ldots, j-1$. For $l = j$, we have by (3)

$$\begin{aligned} z^{i\mathsf{T}} g^{i+j} &= z^{i\mathsf{T}}(A x^{i+j} + b) = \\ &= z^{i\mathsf{T}}\{A(x^{i+1} + x^{i+2} - x^{i+1} + \ldots + x^{i+j} - x^{j+j-1}) + b\} = \\ &= z^{i\mathsf{T}}(A x^{i+1} + b) + z^{i\mathsf{T}} A(x^{i+2} - x^{i+1}) + \ldots \\ &\quad \ldots + z^{i\mathsf{T}} A(x^{i+j} - x^{i+j-1}). \end{aligned} \tag{27}$$

Hence, this relation is reduced to

$$z^{i\mathsf{T}} g^{i+j} = z^{i\mathsf{T}} g^{i+1} + \sum_{p=1}^{j-1} z^{i\mathsf{T}} A z^{i+p} \tag{28}$$

by (3) and (6).

The first and the second terms on the right hand side of (28) are zero by the proposition of the lemma and the inductive hypothesis, respectively. Thus we have proved (26), which completes the proof of the lemma. Q.E.D.

One of the methods most widely used to construct the sequences of points $\{x^i\}$ in such a manner that they satisfy the conditions mentioned so far ((C−1) ~ (C−3)) is to apply so called linear search at each stage: x^{i+1} is determined as the minimal point of the objective function on the straight line

$$x = x^i - \xi H_i g^i. \tag{29}$$

Thus

$$x^{i+1} = x^i - \mu_i H_i g^i , \tag{30}$$

where μ_i gives the minimum of the function

$$\Psi_i(\xi) := f(x^i - \xi H_i g^i) . \tag{31}$$

In fact, since

$$f_\xi(x^i - \xi H_i g^i) = 0 \tag{32}$$

at $\xi = \mu_i$, we obtain

$$\begin{aligned} 0 &= \operatorname{grad} f(x^i - \mu_i H_i g^i)^\mathsf{T} H_i g^i \\ &= \operatorname{grad} f(x^{i+1})^\mathsf{T} H_i g^i \\ &= g^{i+1\mathsf{T}} H_i g^i . \end{aligned} \tag{33}$$

But since by (30)

$$-\mu_i H_i g^i = x^{i+1} - x^i = z^i , \tag{34}$$

we have

$$z^{i\mathsf{T}} g^{i+1} = 0 . \tag{35}$$

Hence, so far as $\mu_i \neq 0$, $\{x^i\}$ satisfies the condition of Lemma 1.

On the other hand, $\{x^i\}$ are not always in general position. Indeed, for example, if it happened that we arrived at the minimal point of the objective function with x^j $(j < n)$, all the subsequent points x^{j+1}, x^{j+2}, ... would be located at the same point and $\mu_i = 0$ for $i = j+1, j+2, \dots$.

If this is not the case, however, we obtain non-zero μ_i's and hence non-zero z^i's. Hence z^i's are linearly independent by Lemma 1, accordingly, $\{x^i\}$ are in general position.

In what follows in this paper, we consider only algorithms involving successive linear minimum searchings.

5. CONSTRUCTION OF THE RECURSIVE ALGORITHMS

We now proceed to the construction of the algorithms. In Sec. 3, we have established the recurrence relation,

$$H_{k+1} = (I - Y_k(U_k^\mathsf{T} Y_k)^{-1} U_k^\mathsf{T})^\mathsf{T} N_k (I - Y_k(U_k^\mathsf{T} Y_k)^{-1} U_k^\mathsf{T}) + Z_k(Z_k^\mathsf{T} Y_k)^{-1} Z_k^\mathsf{T} . \tag{36}$$

If we introduce the notation

$$P_{k+1} := I - Y_k(U_k^\mathsf{T} Y_k)^{-1} U_k^\mathsf{T} , \tag{37}$$

$$Q_{k+1} := Z_k(Z_k^\mathsf{T} Y_k)^{-1} Z_k^\mathsf{T} , \tag{38}$$

the equation (36) reduces to

$$H_{k+1} = P_{k+1}^{\mathsf{T}} N_k P_{k+1} + Q_{k+1} . \tag{39}$$

Moreover, as is easily shown (cf. Appendix D), the matrices P_{k+1} and Q_{k+1} are also determined recursively by the relations

$$P_{i+1} = P_i - \frac{P_i y^i u^{i\mathsf{T}} P_i}{u^{i\mathsf{T}} P_i y^i}, \quad i = 1, 2, \ldots, k, \tag{40}$$

$$Q_{i+1} = Q_i + \frac{z^i z^{i\mathsf{T}}}{z^i y^i}, \quad i = 1, 2, \ldots, k, \tag{41}$$

where $P_1 = I$, $Q_1 = 0$ and u^i is the i-th column vector of U. The recurrence relation (39) is reduced to

$$H_{k+1} = H_k - \frac{H_k y^k y^{k\mathsf{T}} H_k}{y^{k\mathsf{T}} H_k y^k} + \frac{z^k z^{k\mathsf{T}}}{z^{k\mathsf{T}} y^k} + \tag{42}$$

$$+ y^{k\mathsf{T}} H_k y^k \left(\frac{H_k y^k}{y^{k\mathsf{T}} H_k y^k} - \frac{P_k^{\mathsf{T}} u_k}{u^{k\mathsf{T}} P_k y^k} \right) \left(\frac{H_k y^k}{y^{k\mathsf{T}} H_k y^k} - \frac{P_k^{\mathsf{T}} u^k}{u^{k\mathsf{T}} P_k y^k} \right)^{\mathsf{T}} .$$

Indeed by (39), we have

$$H_k y^k = (P_k^{\mathsf{T}} N_k P_k + Q_k)\, y^k = P_k^{\mathsf{T}} N_k P_k y^k + Z_{k-1} (Z_{k-1}^{\mathsf{T}} Y_{k-1})^{-1} Z_{k-1}^{\mathsf{T}} y^k . \tag{43}$$

However, since

$$z^{i\mathsf{T}} y^k = z^{i\mathsf{T}} A z^k = 0 \quad \text{for} \quad i = 1, 2, \ldots, k-1 \tag{44}$$

by (4) and Lemma 1, we obtain

$$H_k y^k = P_k^{\mathsf{T}} N_k P_k y^k . \tag{45}$$

Substituting (40), (41) and (45) into (39), we obtain

$$H_{k+1} = \left(I - \frac{P_k^{\mathsf{T}} u^k y^{k\mathsf{T}}}{u^{k\mathsf{T}} P_k y^k} \right) H_k \left(I - \frac{y^k u^{k\mathsf{T}} P_k}{u^{k\mathsf{T}} P_k y^k} \right) + \frac{z^k z^{k\mathsf{T}}}{z^{k\mathsf{T}} y^k}, \tag{46}$$

which is equivalent to (42), which was the relation to be proved.

6. VARIOUS ALGORITHMS

We are now ready to present various algorithms with linear minimization described in Sec. 4 by specifying the general matrix recurrence relation (42) given in Sec. 5. The recurrence relation (42) was established under the conditions (C−1), (C−3) and is specified by giving vectors u^k. (We also assume (C−4) for the sake of convenience.) Although the vectors u^k's are arbitrary, several forms are preferred for the convenience in establishing real algorithms. In particular, the following two choices are among the most frequeny used:

(A) select such $\boldsymbol{u}^k$ that satisfies

$$P_k^{\mathsf{T}} \boldsymbol{u}^k = H_k y^k;$$

(B) select such $\boldsymbol{u}^k$ that satisfies

$$P_k^{\mathsf{T}} \boldsymbol{u}^k = z^k.$$

Proofs of existence of vectors satisfying conditions (A) and/or (B) are given in Appendix E.

In what follows in this section, we present several real algorithms:

1°) Davidon-Fletcher-Poweli Method [3], [7]:

$$H_{k+1} = H_k - \frac{H_k y^k y^{k\mathsf{T}} H_k}{y^{k\mathsf{T}} H_k y^k} + \frac{z^k z^{k\mathsf{T}}}{z^{k\mathsf{T}} y^k}. \tag{47}$$

This algorithm is obtained by setting

$$P_k^{\mathsf{T}} \boldsymbol{u}^k = H_k y^k$$

in (42).

2°) Broyden-Fletcher-Goldfarb-Shanno Method [2], [6], [9], [13]:

$$H_{k+1} = \left(I - \frac{z^k y^{k\mathsf{T}}}{z^{k\mathsf{T}} y^k}\right) H_k \left(I - \frac{y^k z^{k\mathsf{T}}}{z^{k\mathsf{T}} y^k}\right) + \frac{z^k z^{k\mathsf{T}}}{z^{k\mathsf{T}} y^k}. \tag{48}$$

This algorithm is obtained by setting

$$P_k^{\mathsf{T}} \boldsymbol{u}^k = z^k$$

in (46).

3°) One Parameter Method by Broyden [2]:

$$H_{k+1} = H_k - \frac{H_k y^k y^{k\mathsf{T}} H_k}{y^{k\mathsf{T}} H_k y^k} + \frac{z^k z^{k\mathsf{T}}}{z^{k\mathsf{T}} z^k} + \tag{49}$$

$$+ \alpha y^{k\mathsf{T}} H_k y^k \left(\frac{H_k y^k}{y^{k\mathsf{T}} H_k y^k} - \frac{z^k}{z^{k\mathsf{T}} y^k}\right) \left(\frac{H_k y^k}{y^{k\mathsf{T}} H_k y^k} \frac{z^k}{z^{k\mathsf{T}} y^k}\right)^{\mathsf{T}},$$

where

$$\alpha := \left(\frac{(1 - \lambda)\, z^{k\mathsf{T}} y^k}{y^{k\mathsf{T}} H_k y^k + (1 - \lambda)\, z^{k\mathsf{T}} y^k}\right)^2. \tag{50}$$

In this algorithm, the vector $\boldsymbol{u}^k$ is selected so that

$$P_k^{\mathsf{T}} \boldsymbol{u}^k = \lambda H_k y^k + (1 - \lambda)\, z^k,$$

namely, the conditions (A) and (B) are "mixed".

4°) Broyden's Rank-One Method [1]:

$$H_{k+1} = H_k + \frac{(z^k - H_k y^k)(z^k - H_k y^k)^{\mathsf{T}}}{(z^k - H_k y^k)^{\mathsf{T}} y^k}. \tag{51}$$

This is a special case of 3°), in which α is selected so that

$$\alpha := \left(\frac{(1-\lambda)\, z^{k\mathsf{T}} y^k}{y^{k\mathsf{T}} H_k y^k + (1-\lambda)\, z^{k\mathsf{T}} y^k}\right)^2 = \frac{z^{k\mathsf{T}} y^k}{z^{k\mathsf{T}} y^k - y^{k\mathsf{T}} H_k y^k}.$$

APPENDIX

Throughout this Appendix, we denote:

$$R(A) := \{x \mid x \in \mathbb{R}^n,\ x = Ay \text{ for some } y \in \mathbb{R}^m\}, \tag{A1}$$

$$N(A) := \{x \mid x \in \mathbb{R}^m,\ Ax = 0\}, \tag{A2}$$

where A is an $n \times m$ matrix.

[[A]]

$$R(I - S^{\mathsf{T}}) = N(A^{\mathsf{T}}), \tag{A3}$$

where A is an $n \times m$ matrix with rank $(A) = m$ and S is the projection matrix into $R(A)$ in the direction of the normal vector of $R(B)$ in which B is an $n \times m$ matrix with det $(B^{\mathsf{T}}A) \neq 0$.

Proof. Since the projection matrix S is given by

$$S = A(B^{\mathsf{T}}A)^{-1} B^{\mathsf{T}}, \tag{A4}$$

we obtain

$$A^{\mathsf{T}}(I - S^{\mathsf{T}}) = A^{\mathsf{T}} - A^{\mathsf{T}}B(A^{\mathsf{T}}B)^{-1} A^{\mathsf{T}} = A^{\mathsf{T}} - A^{\mathsf{T}} = 0,$$

which implies

$$R(I - S^{\mathsf{T}}) \subseteq N(A^{\mathsf{T}}). \tag{A5}$$

On the other hand, for all $x \in N(A^{\mathsf{T}})$ we have

$$x = x - B(A^{\mathsf{T}}B)^{-1} A^{\mathsf{T}}x$$

since $A^{\mathsf{T}}x = 0$. Hence

$$x = (I - B(A^{\mathsf{T}}B)^{-1} A^{\mathsf{T}})\, x = (I - S^{\mathsf{T}})\, x,$$

which implies

$$N(A^{\mathsf{T}}) \subseteq R(I - S^{\mathsf{T}}). \tag{A6}$$

By (A5) and (A6), we obtain

$$R(I - S^{\mathsf{T}}) = N(A^{\mathsf{T}}),$$

which was the relation to be proved.

[[B]]

If A is an $n \times m$ matrix with rank $(A) = m$, the general $n \times l$ matrix solution of the matrix equation

$$A^{\mathsf{T}}X = 0 \tag{A7}$$

is given by

$$X = (I - B(A^{\mathsf{T}}B)^{-1} A^{\mathsf{T}}) N, \tag{A8}$$

where B is an arbitrary $n \times m$ matrix with $\det(A^{\mathsf{T}}B) \neq 0$ and N is an arbitrary $n \times l$ matrix.

Proof. Since the column vectors of the $n \times l$ matrix solution X are in $N(A^{\mathsf{T}})$ and

$$R(I - S^{\mathsf{T}}) = N(A^{\mathsf{T}})$$

by the preceding lemma, the general $n \times l$ matrix solution is given by

$$X = (I - S^{\mathsf{T}}) N,$$

where N is an arbitrary $n \times m$ matrix.

However, since

$$S = A(B^{\mathsf{T}}A)^{-1} B^{\mathsf{T}},$$

we obtain (A8), which was the relation to be proved.

[[C]]

If N_k is positive definite in (21) and $Z_k \neq 0$ in (16) then H_{k+1} is also positive definite in (16).

Proof. The quadratic form $\boldsymbol{a}^{\mathsf{T}}H_{k+1}\boldsymbol{a}$ defined for $\boldsymbol{a} \in \mathbb{R}^n$ is represented as a sum of two quardatic forms as

$$\boldsymbol{a}^{\mathsf{T}}H_{k+1}\boldsymbol{a} = \boldsymbol{a}^{\mathsf{T}}(I - Y_k(U_k^{\mathsf{T}}Y_k)^{-1} U_k^{\mathsf{T}})^{\mathsf{T}} N_k(I - Y_k(U_k^{\mathsf{T}}Y_k)^{-1} U_k^{\mathsf{T}}) \boldsymbol{a} + \\ + \boldsymbol{a}^{\mathsf{T}}Z_k(Z_k^{\mathsf{T}}Y_k)^{-1} Z_k^{\mathsf{T}}\boldsymbol{a} \tag{A9}$$

by (36). These two quadratic forms can be regarded as quadratic forms defined for

$$(I - Y_k(U_k^{\mathsf{T}}Y_k)^{-1} U_k^{\mathsf{T}}) \boldsymbol{a}$$

and

$$Z_k^{\mathsf{T}}\boldsymbol{a},$$

respectively. Since N_k is positive definite and

$$Z_k^{\mathsf{T}}Y_k = Z_k^{\mathsf{T}}AZ_k$$

with a positive definite A, it is clear that these two quadratic forms are at least positive semi-definite. Hence it is only necessary to prove that if $\boldsymbol{a} \neq 0$ then $\boldsymbol{a}^{\mathrm{T}} H_{k+1} \boldsymbol{a} \neq 0$.

Assume on the contrary that there exists such a non-zero vector $\boldsymbol{a} \in \mathbb{R}^n$ that $\boldsymbol{a}^{\mathrm{T}} H_{k+1} \boldsymbol{a} = 0$.

By positive semi-definiteness of H_{k+1}, this is possible only if

(A10) $$(I - Y_k (U_k^{\mathrm{T}} Y_k)^{-1} U_k^{\mathrm{T}})\, \boldsymbol{a} = 0$$

and

(A11) $$Z_k^{\mathrm{T}} \boldsymbol{a} = 0\,.$$

However, since

$$R(I - Y_k (U_k^{\mathrm{T}} Y_k)^{-1} U_k^{\mathrm{T}}) = N(Y_k^{\mathrm{T}})$$

by [[A]], we have

(A12) $$\boldsymbol{a} \in R(Y_k)$$

if $\boldsymbol{a}$ satisfies (A10).

On the other hand, (A12) implies that there exists such a non-zero vector $\boldsymbol{b}$ that

$$\boldsymbol{a} = Y_k \boldsymbol{b}\,.$$

Hence $Z_k^{\mathrm{T}} \boldsymbol{a}$ is given by

$$Z_k^{\mathrm{T}} \boldsymbol{a} = Z_k^{\mathrm{T}} Y_k \boldsymbol{b} = Z_k^{\mathrm{T}} A Z_k \boldsymbol{b}\,.$$

But, since A is positive definite and $Z_k \neq 0$,

$$Z_k^{\mathrm{T}} \boldsymbol{a} \neq 0\,.$$

This implies that it is impossible for both the relations (A10) and (A11) to hold simultaneously.

Consequently, we have shown that

$$\boldsymbol{a}^{\mathrm{T}} H_{k+1} \boldsymbol{a} \neq 0$$

for all $\boldsymbol{a} \neq 0$.

[[D]]

Given two $n \times i$ matrices

$$Y_i = [\boldsymbol{y}^1 \vdots \boldsymbol{y}^2 \vdots \ldots \vdots \boldsymbol{y}^i]\,,$$
$$U_i = [\boldsymbol{u}^1 \vdots \boldsymbol{u}^2 \vdots \ldots \vdots \boldsymbol{u}^i]$$

with $\det(U_i^{\mathrm{T}} Y_i) \neq 0$ for $i = 1, 2, \ldots, k$ $(\leqq n)$, the matrix

(A13) $$P_{k+1} := I - Y_k (U_k^{\mathrm{T}} Y_k)^{-1} U_k^{\mathrm{T}}$$

is obtained by k-iterated calculations of the recurrence relation

$$\text{(A14)} \qquad P_{i+1} = P_i - \frac{P_i y^i u^{i\mathsf{T}} P_i}{u^{i\mathsf{T}} P_i y^i}, \quad i = 1, 2, \ldots, k,$$

$$P_1 = I .$$

Proof. The proof is by induction on k.

$k = 1$: For $k = 1$, since $P_1 = I$, we have by (A14)

$$P_2 = P_1 - \frac{P_1 y^1 u^{1\mathsf{T}} P_1}{u^{1\mathsf{T}} P_1 y^1} = I - \frac{y^1 u^{1\mathsf{T}}}{u^{1\mathsf{T}} y^1},$$

which is exactly the relation (A13) for $k = 1$.

$k = 2, 3, \ldots, m$: We assume that the statement of the lemma holds for $k = 2, 3, \ldots$ $\ldots, m$.

$k = m + 1$: Denote

$$\begin{bmatrix} A & b \\ c^\mathsf{T} & d \end{bmatrix} := (U_{m+1}^\mathsf{T} Y_{m+1})^{-1} = \begin{bmatrix} U_m^\mathsf{T} Y_m & U_m^\mathsf{T} y^{m+1} \\ u^{m+1\mathsf{T}} Y_m & u^{m+1\mathsf{T}} y^{m+1} \end{bmatrix}^{-1}.$$

Then A, b, c^T and d are given by the identities

$$A = (U_m^\mathsf{T} Y_m)^{-1} + \frac{1}{s} (U_m^\mathsf{T} Y_m)^{-1} U_m^\mathsf{T} y^{m+1} u^{m+1\mathsf{T}} Y_m (U_m^\mathsf{T} Y_m)^{-1},$$

$$b = -\frac{1}{s} (U_m^\mathsf{T} Y_m)^{-1} U_m^\mathsf{T} y^{m+1},$$

$$c^\mathsf{T} = -\frac{1}{s} u^{m+1\mathsf{T}} Y_m (U_m^\mathsf{T} Y_m)^{-1},$$

$$d = \frac{1}{s},$$

where

$$s := u^{m+1\mathsf{T}} y^{m+1} - u^{m+1\mathsf{T}} Y_m (U_m^\mathsf{T} Y_m)^{-1} U_m^\mathsf{T} y^{m+1} = \det (U_{m+1}^\mathsf{T} Y_{m+1}) .$$

Hence

$$I - Y_{m+1}(U_{m+1}^\mathsf{T} Y_{m+1})^{-1} U_{m+1}^\mathsf{T} = I - Y_m (U_m^\mathsf{T} Y_m)^{-1} U_m^\mathsf{T} - \frac{(I - Y_m (U_m^\mathsf{T} Y_m)^{-1} U_m^\mathsf{T})\, y^{m+1} u^{m+1\mathsf{T}} (I - Y_m (U_m^\mathsf{T} Y_m)^{-1} U_m^\mathsf{T})}{u^{m+1\mathsf{T}} (I - Y_m (U_m^\mathsf{T} Y_m)^{-1} U_m^\mathsf{T})\, y^{m+1}},$$

which coincides with P_{m+2}.

Hence

$$I - Y_{m+1}(U_{m+1}^\mathsf{T} Y_{m+1})^{-1} U_{m+1}^\mathsf{T} = P_{m+2},$$

which was the relation to be proved.

References

[1] *C. G. Broyden:* Quasi-Newton Methods and heir Application to Function Minimisation. Mathematics of Computation, Vol. 21, pp. 368—381, (1967).

[2] *C. G. Broyden:* The Convergence of a Class of Double-Rank Minimisation Algorithms. Journal of the Institute of Mathematics and its Applications, Vol. 6, pp. 79—90, 222—231 (1970).

[3] *W. C. Davidon:* Variable Metric Method for Minimization, Argonne National Laboratory Rept. ANL-5990 (Rev.) (1959).

[4] *W. C. Davidon:* Optimally Conditioned Optimization Algorithms without Line Searches. Mathematical Programming, Vol. 9, pp. 1—30, (1975).

[5] *J. E. Dennis, J. J. Moré:* Quasi-Newton Methods, Motivation and Theory. SIAM Review, Vol. 19, pp. 46—89, (1977).

[6] *R. Fletcher:* A New Approach to Variable Metric Algorithms. The Computer Journal, Vol. 13, pp. 317—322, (1970).

[7] *R. Fletcher, M. J. D. Powell:* A Rapidly Convergent Descent Method for Minimization. The Computer Journal, Vol. 6, pp. 163—168, (1963).

[8] *R. Fletcher, C. M. Reeves:* Function Minimisation by Conjugate Gradients. The Computer Journal, Vol. 7, pp. 149—154, (1964).

[9] *D. Goldfarb:* A Family of Variable-Metric Methods Derived by Variational Means. Mathematics of Computation, Vol. 24, pp. 23—26, (1970).

[10] *H. Y. Huang:* A Unified Approach to Quadratically Convergent Algorithms for Function Minimisation, Journal of Optimization Theory and Applications, Vol. 5, pp. 405—423, (1970).

[11] *J. D. Pearson:* Variable Metric Methods of Minimisation. The Computer Journal, Vol. 12, pp. 171—179, (1969).

[12] *M. J. D. Powell:* An Efficient Method of Finding the Minimum of a Function of Several Variables without Calculating Derivatives. The Computer Journal, Vol. 7, pp. 155—162, (1964).

[13] *D. F. Shanno:* Conditioning of Quasi-Newton Methods for Function Minimization. Mathematics of Computation, Vol. 24, pp. 647—657, (1970).

[14] *H. Yanai:* On Conjugate Direction Methods. Seminar Report Vol. 190, Institute for Mathematical Sciences, Kyoto Univ., (1973). (In Japanese)

Souhrn

NĚKOLIK POZNÁMEK O KVAZI-NEWTONOVÝCH METODÁCH

MASANORI OZAWA, HIROSHI YANAI

Přehledná poznámka, jejímž cílem je vyšetřit heuristiku a přirozené vztahy ve třídě kvazi-Newtonových metod v optimizačních problémech. Je dokázáno, že jistý speciální algoritmus této třídy je určen, jestliže charakterizujeme jisté parametry (skalární nebo maticové) v obecném řešení maticové rovnice.

Authors' address: Prof. *Masanori Ozawa,* Prof. *Hiroshi Yanai,* Department of Administration Engineering, Faculty of Engineering, Keio University 3-14-1 Hiyoshi, Yokohama, 223, Japan.