Ann. Inst. Statist. Math.
Vol. 42, No. 4, 805–811 (1990)

A NECESSARY AND SUFFICIENT CONVERGENCE CONDITION OF ORTHOMIN(k) METHODS FOR LEAST SQUARES PROBLEM WITH WEIGHT

SHAO LIANG ZHANG[1]* AND YOSHIO OYANAGI[2]

[1]*Doctoral Program in Engineering, University of Tsukuba, Tsukuba, Ibaraki 305, Japan*
[2]*Institute of Information Sciences, University of Tsukuba, Tsukuba, Ibaraki 305, Japan*

(Received August 17, 1989; revised November 28, 1989)

Abstract. A class of Orthomin-type methods for linear systems based on conjugate residuals is extended to a form suitable for solving a least squares problem with weight. In these algorithms a mapping matrix as preconditioner is brought into use. We also give a necessary and sufficient condition for the convergence of the algorithm. Furthermore, we also study the construction of the mapping matrix for which the necessary and sufficient condition holds.

Key words and phrases: Conjugate direction, least square methods with weight, Orthomin method, preconditioning.

1. Introduction

Let A be an $m \times n$ $(m \geq n)$ real inconsistent matrix, x and b real vectors of respective dimensions n and m. The system of linear equations

$$Ax = b \tag{1.1}$$

possibly has no solution. Alternately, it is natural to consider solving the least equares problem with weight of equation (1.1):

$$\|b - A\bar{x}\|_W^2 = \min_{x \in R^n} \|b - Ax\|_W^2 \tag{1.2}$$

where $\bar{x}$ is a solution and the W-norm $\|\cdot\|_W$ is defined by $\|x\|_W = (x, Wx)$ for a symmetric positive definite matrix W. There is an $n \times m$ matrix G such that Gb is always a solution of the W-least squares problem equation (1.2) for any b (Rao and Mitra (1971)). This G is called a generalized

*Now at Inst. of Computational Fluid Dynamics, Hara-machi 2-1-4, Meguro, Tokyo 152, Japan.

inverse of A.

Let $P = AG$, then P has the following properties (Rao and Mitra (1971)):

$$PA = A; \quad (WP)^T = WP; \quad P^2 = P; \quad P^T WA = WA. \tag{1.3}$$

Since $\bar{P} = W^{1/2} P W^{-1/2}$ is a symmetric and semi-positive definite matrix from the above properties, we can write $\bar{P} = QQ^T$ where Q is an $m \times \rho$ matrix (ρ = rank (A)) with orthonormal column vectors.

A common technique for solving the least squares problem is to solve the normal equation by a conjugate gradient method (Hestenes and Stiefel (1952))

$$A^T WAx = A^T Wb. \tag{1.4}$$

However, since the condition number of $A^T A$ is the square of that of A, numerical methods based on $A^T A$ should not be used.

In Oyanagi and Zhang (1987), we presented the Orthomin(k) method of conjugate residual type for the linear least squares problem, and gave a sufficient condition for convergence of the algorithm. Successively, in this paper, we will extend the technique to the least squares problem with weight, and also discuss the necessary convergence condition. In Oyanagi and Zhang (1987), we also pointed out the computational advantages of employing the concept of the mapping matrix B. Incorporating the mapping matrix B into the Orthomin(k) method (Concus and Golub (1976), Vinsome (1976)), we have the following algorithm.

(1.5)

$r_0 = b - Ax_0, \quad p_0 = Br_0$

for $i = 0$ until convergence

$\quad \alpha_i = (r_i, Ap_i)_W / (Ap_i, Ap_i)_W,$

$\quad x_{i+1} = x_i + \alpha_i p_i, \quad r_{i+1} = r_i - \alpha_i Ap_i$

$\quad$ for $j = 0$ to min $(k-1, i)$

$\quad\quad \beta_{i,i-j} = -(ABr_i, Ap_{i-j})_W / (Ap_{i-j}, Ap_{i-j})_W,$

$\quad\quad p_{i+1} = Br_{i+1} + \sum_{j=i-k+1}^{i} \beta_{i,j} p_j.$

Here x_0 is an arbitrary initial guess, α_i is so determined as to minimize the W-norm of new residual $\|r_i - \alpha Ap_i\|_W$ as a function of α along the direction p_i called a correction vector, and the choice of p_{i+1} is r_{i+1} plus a linear combination of former correction vectors, $p_{i-1}, p_{i-2}, \ldots, p_{i-k+1}$. This means that p_{i+1} are so chosen as to make p_{i+1} $A^T WA$-orthogonal to only previous k vectors $\{p_j\}_{i-k+1}^{i}$. We will call this algorithm CR-WLS(k) method. The number k may be $0, 1, 2, \ldots$, depending on the characteristic of the problem and the computer used. The work vector necessary to implement CR-WLS(k) is x, r, ABr and k sets of p and Ap.

The residual and the correction vectors obey the following relations due to the construction of the correction vectors p_i (Eisenstat *et al.* (1983)).

THEOREM 1.1.

$$(Ap_i, Ap_j)_W = 0 \qquad |i-j| \le k, \quad i \ne j, \tag{1.6a}$$

$$(r_i, Ap_j)_W = 0 \qquad 0 < i-j < k, \tag{1.6b}$$

$$(r_i, Ap_i)_W = (r_i, ABr_i)_W, \tag{1.6c}$$

$$(r_i, ABr_j)_W = 0 \qquad 0 < i-j < k, \tag{1.6d}$$

$$(r_i, Ap_j)_W = (r_{i-k}, Ap_j)_W \qquad 0 \le i-j \le k. \tag{1.6e}$$

In Section 2 we give a necessary and sufficient convergence condition and the rate of decrease of the residuals (Theorem 2.2). In Section 3 we present an analysis of the choice of an appropriate mapping matrix B. Section 4 is the conclusion.

2. A necessary and sufficient condition for convergence

In this section, we will give a necessary and sufficient condition for the convergence of the CR-WLS(k) method. We present an error bound as the result of convergence.

Now we will present the following result, which gives an error bound of CR-WLS(k) (Oyanagi and Zhang (1987)).

THEOREM 2.1. *Let $\{r_i\}$ be a sequence of residuals in algorithm* (1.5), *then the following inequality holds*:

$$\frac{\|r_{i+1} - \bar{r}\|_W^2}{\|r_i - r\|_W^2} \le 1 - \frac{\lambda_{\min}^2(M)}{\lambda_{\max}(M)\lambda_{\min}(M) + \rho(R)^2}, \tag{2.1}$$

provided

(a) $AB(I-P) = 0$ *and*

(b) $M \equiv Q^T(W^{1/2}ABW^{-1/2} + W^{-1/2}B^TA^TW^{1/2})Q/2$ *is positive or negative definite*,

where $\bar{r} = b - A\bar{x}$, $R = Q^T(W^{1/2}ABW^{-1/2} - W^{-1/2}B^TA^TW^{1/2})Q/2$, $\lambda_{\min}$ *and*

$\lambda_{\max}$ *are the maximum and minimum eigenvalues and* $\rho(R)$ *is the spectral radius of* R.

The proof of Theorem 2.1 is given in Oyanagi and Zhang (1987). The theorem shows that the CR-WLS(k) method is at least linearly convergent where (a) and (b) are the sufficient condition for the convergence.

Now, we will prove that (a) and (b) are also necessary by constructing counterexamples. First we will present two lemmas.

LEMMA 2.1. $\bar{x}$ *is a solution of W-least squares problem* (1.2), *if and only if* $\bar{x}$ *is a solution of the normal equation* (1.4).

LEMMA 2.2. *Suppose that the convergence condition* (a) *and* (b) *hold. Let* $(r_i, Ap_i)_W = 0$, *then* $A^T W r_i = 0$.

PROOF. From the property (1.6c) of Theorem 1.1,

$$(r_i, ABr_i)_W = (r_i, Ap_i)_W = 0 .$$

We have $r_i W^{1/2} Q Q^T W^{1/2} A B W^{-1/2} Q Q^T W^{1/2} r_i = 0$ by (a) and (1.3).
From (b), we have $Q^T W^{1/2} r_i = 0$. Hence,

$$A^T W r_i = A^T W P r_i = A^T W^{1/2} Q Q^T W^{1/2} r_i = 0 . \qquad \square$$

Lemma 2.2 implies that the algorithm (1.5) will no longer improve x_i if $\alpha_i = 0$ at the i-th step, in which case the x_i is a solution of the W-least squares problem (1.2).

In order to prove that the conditions (a) and (b) are necessary, we will present two counterexamples such that $\alpha_i = 0$ but $A^T M r_i \neq 0$, when either (a) or (b) does not hold.

Example 2.1. Suppose that the convergence condition (b) does not hold. Then, there is a vector $c \neq 0 \in R^n$, such that $c^T Q^T W^{1/2} A B Q W^{-1/2} c = 0$.

Since Q is a full-rank $m \times p$ matrix, the vector

$$e = W^{-1/2} \bar{P} Q c = W^{-1/2} Q c$$

is non-zero. Hence, $A^T W e \neq 0$ by definition of $\bar{P}$ and non-singularity of W.

We consider the following least squares problem with weight which has a particular right-hand side:

$$Ax = e .$$

Let the CR-WLS(k) algorithm start with an initial guess $x_0 = 0$. Then,

$r_0 = e - Ax_0 = e$, $p_0 = Br_0 = Be$ and $\alpha_0 = 0$, while $A^T W r_0 = A^T W e \neq 0$, which means r_0 is an not an optimal residual. Since $p_1 = 0$, we cannot continue the iteration. In this case, the CR-WLS(k) algorithm does not give the correct solution.

LEMMA 2.3. *Let C be an* $m \times m$ *matrix. C is skew-symmetric if and only if*

$$x^T C x = 0 \quad \textit{for every} \quad x \in R^n .$$

Example 2.2. Suppose that the convergence condition (a) does not hold. Then it is easy to prove that $WAB(I-P)$ is not a skew-symmetric matrix using the relation $P^T WA = WA$. There exists a vector $x \neq 0$ such that $x^T WAB(I-P)x \neq 0$. Note that $A^T W x \neq 0$. Since the vector x is decomposed $x = Qy + z$, where $y \in R^p$ and $z \in \text{Ker}(Q^T)$, we have $x^T WAB \cdot (I-P)x = (Qy)^T WABz \neq 0$. If we choose $\delta = -(Qy)^T WABQy / (Qy)^T WABz$, then $x_\delta^T WABx_\delta = 0$, and $A^T W x_\delta \neq 0$, for $x_\delta = Qy + \delta z$.

Consider the W-least squares problem

$$Ax = x_\delta$$

which is solved by the CR-WLS(k) algorithm starting with the initial guess $x_0 = 0$. At the zero-th iteration, we will have

$$\alpha_0 = 0 \quad \text{by relation (1.6c), but} \quad A^T W r_0 \neq 0 .$$

Hence, a counterexample of convergence can be given analogously as in the previous example. In conclusion, we have proven the following main result.

THEOREM 2.2. *The* CR-WLS(k) *algorithm converges for any right-hand side and for any initial guess, if and only if the conditions* (a) *and* (b) *hold.*

The proof is an immediate consequence of Theorem 2.1, Lemma 2.2, Example 2.1 and Example 2.2.

3. Discussion about the choice of mapping matrix B

In the previous section we did not specify the mapping matrix B. In this section, we will discuss how to choose an appropriate mapping matrix B in order to make CR-WLS(k) converge quickly.

The most trivial choice would be $B = G$, where G is a generalized inverse of A. In this case, in the first step of (1.5), x_1 gives one of the least squares solutions. This choice is unrealistic, since if we knew G, we would simply compute Gb without applying any iterative method.

There is a certain trade-off between the number of iterations and the complexity of computing Br. The more the B resembles G, the faster the method will converge. On the other hand, the cost to compute Br at each iteration will become large if B is made close to G.

The CR-WLS(k) algorithm covers a wide class of methods. They differ in the choice of the mapping matrix B and the parameter k. The particular choice of B critically depends on the application and cannot be discussed in general. First we discuss what is meant by the convergence condition (a) $ABP = AB$ and how to construct B satisfying the condition (Oyanagi and Zhang (1987)).

THEOREM 3.1. *$AB(I-P)=0$ holds, if and only if there exists an $m \times m$ matrix D such that $AB = DAA^T$.*

We will give here a few general comments concerning the choice of B.

The simplest choice of B which automatically satisfies the two conditions (a) and (b) in Theorem 2.1 is $B = DA^T$ where D is an appropriate $n \times n$ matrix. In this case M is symmetric and the convergence rate is controlled by $1 - \lambda_{\min}(M)/\lambda_{\max}(M)$. In practical cases, A is a large sparse matrix, so that multiplying A^T from the right will not be too time-consuming. The matrix D should not have too complex a structure. If the symmetric part of D is positive definite, the condition (b) is also satisfied. The user will obtain variant algorithms by choosing different D's.

We have to make the condition number of M in Theorem 2.1 as small as possible. The extreme choice would be to set D equal to a generalized inverse of (A^TWA). In this case, B is a generalized inverse of A. If the columns of A are approximately orthogonal, we may take D as the inverse of the diagonal part of (A^TWA). Incomplete Cholesky decomposition (Meijerink and van der Vorst (1977)) of (A^TWA) will also be applicable.

Using Theorem 2.1, we have the result that if B is chosen to make AB symmetric, and M has dense eigenvalues, the convergence of CR-WLS(k) is fast.

4. Conclusion

We have given a necessary and sufficient condition for convergence of the conjugate residual type Orthomin(k) method which is extended to the W-least squares problem. This method has computational advantages by virtue of the mapping matrix B which we introduced.

Several numerical tests have been performed and we found that the algorithm works well for data smoothing problems by discrete splines, which will be discussed elsewhere.

REFERENCES

Concus, P. and Golub, G. H. (1976). A Generalized conjugate gradient method for nonsymmetric systems of linear equations, *Lecture Notes in Economics and Mathematical Systems*, (eds. R. Glowinski and J. L. Lions), Vol. 134, 56–65, Springer, Berlin.

Eisenstat, S. L., Elman, H. C. and Schults, M. H. (1983). Variational iterative methods for nonsymmetric systems of linear equations, *SIAM J. Numer. Anal.*, **20**, 345–357.

Hestenes, M. R. and Stiefel, E. (1952). Methods of conjugate gradient for solving linear systems, *J. Res. Nat. Bur. Standards*, **49**, 409.

Meijerink, J. A. and van der Vorst, H. Z. (1977). An iterative solution method for linear systems of which the coefficient matrix is a symmetric M-matrix, *Math. Comp.*, **31**, 148–162.

Oyanagi, Y. and Zhang, S. L. (1987). Conjugate Residual Method for Least Squares Problems, *First IASC World Conference on Computational Statistics and Data Analysis*, 167–174.

Rao, C. R. and Mitra, S. K. (1971). *Generalized Inverse of Matrices and Its Applications*, Wiley, New York.

Vinsome, P. K. W. (1976). ORTHOMIN—an iterative method for solving sparse sets of simultaneous linear equations, *Fourth Symposium on Reservoir Simulation, Society of Petroleum Engineers of AIME*, 149–159.