Journal of Computational Mathematics, Vol.12, No.2, 1994, 185–194.

A MODIFIED CONJUGATE DIRECTION METHOD FOR COMPUTING THE PSEUDOINVERSE*

Zhao Jin-xi
(Dept. of Math., Nanjing University, Nanjing, China)

Abstract

In this paper we are concerned with the modified conjugate direction method for computing the pseudoinverse by using an orthogonal basis of the range space of A. Numerical results show that the new method retains some main advantages in terms of efficiency and accuracy.

1. Introduction

The method of least squares is a standard tool for solving problems such as control, state evaluation and identification[2]. The linear least square problem is defined as the minimization of the norm of the residual vector

$$\min_x \|Ax - b\|_2^2, \tag{1}$$

where $A \in R^{m\times n}$ with rank k, $b \in R^m$ is a real vector to be approximated, and $x \in R^n$ is a real vector.

Connecting with the linear least squares problem (1), the computation of the pseudoinverse of A is also quite common in this context. A real $n \times m$ matrix G is called the pseudoinverse of A if G satisfies the following conditions:

$$(1) AGA = A, \quad (2) GAG = G, \quad (3)(AG)^T = AG, \quad (4)(GA)^T = GA, \tag{2}$$

and can be written as

$$A^+ = G.$$

Thus, the least squares solution of the minimum norm of problem (1) is

$$x = A^+ b. \tag{3}$$

This solution is unique whether the problem is consistent or not.

In this paper, a class of conjugate direction method for computing the pseudoinverse is considered. The given method requires less computational work and has other advantages.

* Received July 28, 1993.

Throughout this paper, let $R(\cdot), N(\cdot)$ stand for the range space and the null space of a matrix, respectively. $R(x)^{\perp}$ denotes an orthogonal complement of $R(\cdot)$ and $\|\cdot\|$ the Euclidian vector norm.

2. The Case of the Full Rank Matrix

In this section we are concerned only with the simplest case that A is an $m \times m$ symmetric positive definite matrix. Let vectors $p_1, p_2, \cdots, p_m \in R^m$ be mutually conjugate, i.e.

$$p_i^T A p_j = \begin{cases} 0, & i \neq j, \\ d_i \neq 0, & i = j. \end{cases}$$

Using these vectors, we can easily obtain the sequence of matrices

$$G_i = \sum_{j=1}^{i} \frac{p_j p_j^T}{d_j}, \quad i = 1, 2, \cdots, m.$$

Due to the conjugacy of vectors $p_1, p_2, \cdots, p_m$, we have

$$G_i A p_j = p_j, \quad j \leq i, \tag{4}$$

and

$$G_i A p_j = 0, \quad j > i. \tag{5}$$

In particular, for $i = m$, the matrix G_m satisfies

$$G_m A p_j = p_j \tag{6}$$

or

$$(G_m A - I) p_j = 0, \quad j = 1, 2, \cdots, m.$$

Since vectors $p_1, p_2, \cdots, p_m$ are linearly independent, it follows that

$$G_m A = I$$

and

$$G_m = A^{-1}.$$

Observe that the sequence of matrices $G_i, i = 1, 2, \cdots, m$, are generated by the following relation:

$$G_0 = 0, \quad G_i = G_{i-1} + \frac{p_i p_i^T}{d_i}. \tag{7}$$

Summarizing, we have[3]

Theorem 1. *Let $A \in R^{m \times m}$ be a symmetric positive definite matrix. Given a set of vectors $p_1, p_2, \cdots, p_m \in R^m$, which are mutually conjugate, and the matrices $G_0, G_1, \cdots, G_m \in R^{m \times m}$ generated by the recurrence relation (7), then $G_i, i =$*

$1, 2, \cdots, m$, *satisfy the expressions* (4), (5), (6) *and*

$$A^{-1} = G_m = \sum_{i=1}^{m} \frac{p_i p_i^T}{d_i}, \quad d_i = p_i^T A p_i. \tag{8}$$

In fact, let

$$P = [p_1, p_2, \cdots, p_m] \in R^{m \times m}.$$

Due to the conjugacy of vectors $p_1, p_2, \cdots, p_m$, we have

$$P^T A P = \operatorname{diag}(d_1, d_2, \cdots, d_m).$$

Since vectors $p_1, p_2, \cdots, p_m$ are also linearly independent, it follows that

$$(P^T A P)^{-1} = P^{-1} A^{-1} P^{-T} = \operatorname{diag}(d_1, d_2, \cdots, d_m)^{-1}.$$

It is chvious seen that A^{-1} satisfies

$$A^{-1} = P \operatorname{diag}(d_1, d_2, \cdots, d_m)^{-1} P^T = \sum_{i=1}^{m} \frac{p_i p_i^T}{d_i}. \tag{9}$$

This is the explicit formula generated by the vectors $p_1, p_2, \cdots, p_m$, which are mutually conjugate, for finding the inverse of A.

If the matrix A is singular or non-square, then there is no inverse matrix in the general sense. But there always exists the pseudoinverse A^+ defined by condition (2).

Here, we shall consider the case of full column rank, i.e., $A \in R^{m \times n}$ and rank $(A) = n$. Since the matrix $A \in R^{m \times n}$ is of full column rank, we can find an orthogonal basis of $R(A)$ and assume it to be of the following form:

$$c_1 = Ap_1, c_2 = Ap_2, \cdots, c_n = Ap_n.$$

In other words, the vectors $p_1, p_2, \cdots, p_n$ are mutually conjugate with respect to $A^T A$, i.e.,

$$p_i^T A^T A p_j = \begin{cases} 0, & i \neq j, \\ \|c_i\|^2, & i = j. \end{cases} \tag{10}$$

Since $A^T A$ is nonsingular, it follows from the results described above that

$$(A^T A)^{-1} = \sum_{i=1}^{n} \frac{p_i p_i^T}{\|c_i\|^2}.$$

On the other hand, the pseudoinverse A^+ can be denoted as

$$A^+ = (A^T A)^{-1} A^T.$$

Hence the pseudoinverse $A^\dagger$ generated by the vectors $p_1, p_2, \cdots, p_n$ is

$$A^+ = \sum_{i=1}^{n} \frac{p_i p_i^T A^T}{\|c_i\|^2} = \sum_{i=1}^{n} \frac{p_i c_i^T}{\|c_i\|^2}.$$

Based on the above discussion, we have

Theorem 2. *Let $A \in R^{m\times n}$ be of full column rank, i.e., rank(A)=n, and vectors $p_1, p_2, \cdots, p_n$ be mutually conjugate with respect to A^TA. Then the pseudoinverse generated by $p_1, p_2, \cdots, p_n$ can be expressed as*

$$A^+ = \sum_{i=1}^{n} \frac{p_i c_i^T}{\|c_i\|^2}. \tag{11}$$

If we set

$$C = [\frac{c_1}{\|c_1\|}, \cdots, \frac{c_n}{\|c_n\|}] \in R^{m\times n}$$

and

$$P = [\frac{p_1}{\|c_1\|}, \cdots, \frac{p_n}{\|c_n\|}] \in R^{n\times n},$$

then A^+ can be denoted as

$$A^+ = PC.$$

That is, if we have obtained an orthogonal basis of $R(A)$, $c_1 = Ap_1, \cdots, c_n = Ap_n$, then the pseudoinverse can be generated by the following recurrence relation:

$$G_0 = 0,$$
$$G_i = G_{i-1} + \frac{p_i c_i^T}{\|c_i\|^2}, \quad i = 1, 2, \cdots, n,$$
$$A^+ = G_n.$$

We can also make an analogous discussion when A is of full row rank and obtain the following result.

Theorem 3. *Let $A \in R^{m\times n}$ be of full row rank, and $s_i = A^T p_i, i = 1, 2, \cdots, m$, be an orthogonal basis of $R(A^T)$. Then the pseudoinverse of A can be written as*

$$A^+ = \sum_{i=1}^{m} \frac{s_i p_i^T}{\|s_i\|^2}. \tag{12}$$

From the above discussion, it is seen that, if an orthogonal basis of $R(A)$, $c_i = Ap_i, i = 1, 2, \cdots, n$, has been obtained, then the recurrence expression of the pseudoinverse A^+ generated by $p_1, p_2, \cdots, p_n$ has a special form. The problem to be solved here is how to obtain the mutually conjugate vectors $p_1, p_2, \cdots, p_n$ for a given matrix A of full column rank.

In particular, we can choose a set of vectors $v_1, v_2, \cdots, v_n \in R^n$ which are linearly independent. Then, begin with vectors $Av_1, \cdots, Av_n$ to determine the vectors $c_1, \cdots, c_n$ and $p_1, \cdots, p_n$ by the formulas

$$c_1 = Av_1, \quad p_1 = v_1,$$
$$c_2 = Av_2 - \alpha_{21}c_1, \quad p_2 = v_2 - \alpha_{21}p_1,$$
$$c_3 = Av_3 - \alpha_{31}c_1 - \alpha_{32}c_2, \quad p_3 = v_3 - \alpha_{31}p_1 - \alpha_{32}p_2$$

and at the i-th step,

$$c_i = Av_i - \sum_{j=1}^{i-1} \alpha_{ij} c_j, \quad p_i = v_i - \sum_{j=1}^{i-1} \alpha_{ij} p_j, \quad i = 1, 2, \cdots, n, \tag{13}$$

where, for $j = 1, 2, \cdots, i-1$,

$$\alpha_{ij} = \frac{v_i^T A^T c_j}{\|c_j\|^2}.$$

It is easy to verify that the vectors $c_1, c_2, \cdots, c_n \in R^m$ are mutually orthogonal and $p_1, \cdots, p_n \in R^n$ are mutually conjugate with respect to $A^T A$. When a set of orthogonormal vectors $c_1, c_2, \cdots, c_n$ is found, the method discussed above is closely related to classical Gram-Schmidt(CGM) orthogonalization. In the special case $v_i = e_i, i = 1, 2, \cdots, n$, it is algoritmically equivalent to this, the only difference is that $P = [p_1, p_2, \cdots, p_n]$ is explicitly computed. The classical Gram-Schmidt orthogonalization and its modification have been studied extensively[1]. It is well known in general that the modified Gram-Schmidt (MGS) orthogonalization has better numerical results than that of the CMS and is of great importance in improving computational accuracy. Based on the MGS for computing an orthonormal basis of $R(A)$ and vectors $p_1, p_2, \cdots, p_n$ which are mutually conjugate, We consider the following scheme.

The complete algorithm is then

1. Give $A \in R^{m \times n}$ with $\text{rank}(A) = n$ and select n vectors $v_1, v_2, \cdots, v_n \in R^n$ which are linearly independent.

2. Form

$$C = [c_1, c_2, \cdots, c_n] = [Av_1, Av_2, \cdots, Av_n] \in R^{m \times n},$$

$$P = [p_1, p_2, \cdots, p_n] = [v_1, v_2, \cdots, v_n] \in R^{n \times n},$$

$$G = 0 \in R^{n \times m}.$$

3. For $i = 1 : n$ do

$c_i = C(:, i)$; $p_i = P(:, i)$; $\alpha_i = \text{sqrt}\,(c_i' * c_i)$; $c_i = c_i / \alpha_i$; $p_i = p_i / \alpha_i$; $G = G + p_i * c_i'$;

For $k = i + 1 : n$ do

$\beta = c_i * C(:, k)$; $C(:, k) = C(:, k) - \beta * c_i$; $P(:, k) = P(:, k) - \beta * p_i$

end;

end.

Numerical results show that generally the algorithm above has good numerical behavior and requires substantially less work.

3. Rank Deficient Problems

It is necessary to consider the case of the rank deficient matrix for computing a pseudoinverse, that is, the matrix $A \in R^{m \times n}$ and $\text{rank}(A) = r < \min(m, n)$.

First, we have

Lemma 1. *If $A \in R^{m\times n}$ and* rank $(A) = r$, *then there always exist r linearly independent vectors $p_1, \cdots p_r$, such that*

$$p_i^T A^T A p_j = \begin{cases} 0, & i \neq j, \\ \neq 0, & i = j. \end{cases}$$

That is, there always exist r mutually conjugate vectors $p_1, \cdots, p_r$ with respect to $A^T A$.

Proof. Inasmuch as rank $(A) = r$, there always exist r mutually orthogonal vectors $c_1, \cdots, c_r$ to form an orthogonal basis of $R(A)$.

On the other hand, since

$$c_i \in R(A), \quad i = 1, 2, \cdots, r, \tag{14}$$

there exist vectors $p_i \in R^n, i = 1, 2, \cdots, r$, such that

$$c_i = Ap_i, i = 1, 2, \cdots, r. \tag{15}$$

Hence, it follows from the orthogonality of $c_i, i = 1, 2, \cdots, r$, that the vectors $p_1, p_2, \cdots, p_r$ are mutually conjugate and linearly independent.

Suppose that a set of mutual conjugate vectors $p_1, \cdots, p_k, (k < r)$, and $c_1 = Ap_1, \cdots, c_k = Ap_k$ have been obtained at the k-th step. Then the key point is how to generate a new vector p_{k+1}, which is conjugate with $p_1, p_2, \cdots, p_k$. For illustration, set

$$G_k = \sum_{j=1}^{k} \frac{p_i c_i^T}{\|c_i\|^2}. \tag{16}$$

Here, we can give the following theorem.

Theorem 4. *Assume that vectors $c_1 = Ap_1, \cdots, c_k = Ap_k, (k < r)$, are mutually orthogonal. The matrices $G_1, \cdots, G_k$ can be generated as in expression* (16) *and take*

$$B_k = I_{n\times n} - G_k A. \tag{17}$$

Then for any vector $u \in R^n$, the vector

$$p_{k+1} = B_k u \tag{18}$$

is conjugate with $p_1, p_2, \cdots, p_k$.

Proof. Inasmuch as

$$B_k = I_{n\times n} - G_k A = I_{n\times n} - \sum_{i=1}^{k} \frac{p_i c_i^T A}{\|c_i\|^2},$$

it follows from the conjugacy of the vectors $p_1, \cdots, p_k$ that

$$p_j^T A^T A B_k u = p_j^T A^T A u - \sum_{i=1}^{k} \frac{p_j^T A^T A p_i c_i^T A u}{\|c_i\|^2} = 0, \quad 1 \leq j \leq k.$$

Thus, the vector $p_{k+1} = B_k u$ is mutually conjugate with the vectors $p_1, p_2, \cdots, p_k$.

By the above theorem, for any vector $u \in R^n$, a vector

$$p_{k+1} = B_k u$$

is always mutually conjugate with respect to vectors $p_1, \cdots, p_k$. But it is not guaranteed that the vectors $p_1, \cdots, p_{k+1}$ are linearly independent. We ask whether, for $k \leq r = \text{rank}(A)$, the vectors $p_1, \cdots, p_{k+1}$ are mutually conjugate and linearly independent. In other words, we have a great deal of freedom in the choice of a vector $u \in R^n$, but it must satisfy

$$c_{k+1} = AB_k u \neq 0.$$

In this respect, we have the following theorem.

Theorem 5. *Let*

$$C = [\frac{c_1}{\|c_1\|}, \cdots, \frac{c_k}{\|c_k\|}] \in R^{m \times k}$$

be a column orthonormal matrix and $c_i = Ap_i, i = 1, 2, \cdots, k$. *If we take a vector* $u \in R^n$ *such that*

$$Au \in R(C)^{\perp},$$

then the vectors $p_1, \cdots, p_k, p_{k+1} = B_k u$ *are linearly independent and mutually conjugate.*

Proof. The conjugacy has been proved as above. Since

$$AB_k u = \Big(I_{n\times n} - \sum_{i=1}^{k} \frac{c_i c_i^T}{\|c_i\|^2}\Big) Au,$$

we have

$$c_{k+1} = Ap_{k+1} = AB_k u \neq 0$$

as long as

$$Au \notin N\Big(I_{n\times n} - \sum_{i=1}^{k} \frac{c_i c_i^T}{\|c_i\|^2}\Big).$$

Inasmuch as

$$N\Big(I_{n\times n} - \sum_{i=1}^{k} \frac{c_i c_i^T}{\|c_i\|^2}\Big) = R(C),$$

if

$$Au \in R(C)^{\perp},$$

we can have

$$c_{k+1} = B_k u \neq 0.$$

From subsequent discussions, we can see that the orthogonal projector

$$\sum_{i=1}^{k} \frac{c_i c_i^T}{\|c_i\|^2}$$

is of simple form. Hence it is not hard to find a vector $u \in R^n$ such that

$$Au \in R(C)^{\perp}.$$

Naturally, we want to ask whether or not the expression (11) for the pseudoinverse of full rank can be extend to the rank-deficient case. In general, the conclusion is negative. To see why, we need the following further discussion.

Let $c_1, c_2, \cdots, c_r$ be an orthogonal basis of $R(A)$ obtained by the orthogonalizing process and

$$c_i = Ap_i, \quad i = 1, 2, \cdots, r.$$

Observe that the expression of p_i is not unique for the given vector c_i. Take

$$P = [\frac{p_1}{\|c_1\|}, \cdots, \frac{p_r}{\|c_r\|}] \in R^{n\times r} \tag{19}$$

and

$$C = [\frac{c_1}{\|c_1\|}, \cdots, \frac{c_r}{\|c_r\|}] \in R^{m\times r}. \tag{20}$$

Then the column orthonormal matrix C can be denoted as

$$C = AP. \tag{21}$$

Since a set of vectors $c_1, \cdots, c_r$ is an orthogonal basis of $R(A)$, the matrix A may be written as

$$A = CD, \tag{22}$$

where $D \in R^{r\times n}$ and $\mathrm{rank}(D) = r$. Then we can prove the following lemma.

Lemma 2. *If $P \in R^{n\times r}$ and $D \in R^{r\times n}$ are defined as above, then*

$$DP = I_{r\times r}.$$

Proof. It follows from the relations (21) and (22) that

$$C = CDP.$$

Since C is a column orthonormal matrix, we have the conclusion of the lemma.

$$DP = I_{r\times r}.$$

We can also see that the matrices P and D satisfy

$$DPD = D, \quad PDP = P, \quad (DP)^T = DP, \quad (PD)^2 = PD,$$

i.e., $P \in R^{n\times r}$ is a $\{1,2,3\}$[1)] inverse matrix (see [6]) of $D \in R^{r\times n}$. Based on the above lemma, it is easy to prove the following theorem.

Theorem 6. *Assume that $A \in R^{m\times n}$, $\mathrm{rank}(A) = r < \min(m, n)$ and*

$$c_1 = Ap_1, \cdots, c_r = Ap_r$$

are an orthogonal basis of $R(A)$. Then the matrix

$$B = \sum_{i=1}^{r} \frac{p_i c_i^T}{\|c_i\|^2} \tag{23}$$

---

1) G is an $\{i, j, k\}$ inverse matrix if only the i, j, k-th conditions in (2) are satisfied by G.

satisfies

$$BAB = B, \quad ABA = A, \quad (BA)^2 = BA, \quad (AB)^T = AB,$$

i.e., the matrix $B \in R^{n\times m}$ *defined by* (23) *is a* $\{1,2,3\}$ *inverse matrix of* $A \in R^{m\times n}$.

Since the condition

$$(BA)^T = BA \tag{24}$$

is not satisfied in the general case, the matrix B *is not the pseudoinverse defined by* (2) *when* $r < \min(m,n)$.

On the other hand, we can now define a matrix

$$P^c = [p_1^c, \cdots, p_r^c] = D^+ \in R^{n\times r}, \tag{25}$$

as a special choice of P, where $D \in R^{r\times n}$ is given by (22) and D^+ is the pseudoinverse[1)] of D. Obviously, the matrix D is of row full rank. Then we have.

Lemma 3. *The vectors* $p_1^c, \cdots, p_r^c$ *defined by* (25) *are mutually conjugate with respect to* A^TA.

Proof. Since

$$A = CD$$

and

$$P^c = [p_1^c, \cdots, p_r^c] = D^+,$$

it follows that

$$(P^c)^T A^T A P^c = (D^T)^+ D^T C^T C D D^+ = I_{r\times r}.$$

That is, the vectors $p_1^c, \cdots, p_r^c$ are mutually conjugate with respect to A^TA.

Then we can have the following theorem.

Theorem 7. *Assume* $A \in R^{m\times n}$, $\mathrm{rank}(A) = r$. *If the matrices* D, P^c, C *are defined by* (22), (25) *and* (21), *respectively, then the pseudoinverse of* A *can be expressed as*

$$A^+ = \sum_{i=1}^{r} p_i^c c_i^T. \tag{26}$$

Proof. This theorem is proved by direct verification.

4. Numerical Experiment

The algorithm described in the previous sections has been implemented by using MATLAB, and has been run on the VAX computer of the University of Trento in Italy. The considered pseudoinverse problem has the following data.

$$A = [\max(i,j)] \in R^{15\times 10}.$$

Algorithm.

Alg.1. The CGS to form an orthonormal basis of $R(A)$;

1) The Computed Pseudoinverse is donoted by G.

Alg.2. The given algorithm in the paper;

Alg.3. The algorithm PINV of the MATLAB.

Numerical results are given in the following table.

Table. The Accuracy and Flops[1)]

Algor.	$\|GAG-G\|$	$\|AGA-A\|$	$\|(AG)^T-AG\|$	$\|(GA)^T-GA\|$	Flops
Alg.1	.2478E-08	.2046E-08	.1314E-12	.5288E-08	7373
Alg.2	.1246E-13	.9720E-12	.2766E-13	.2086E-12	8670
Alg.3	.4561E-13	.2196E-12	.1192E-12	.3641E-13	31407

From the above table, we can see that the method obtained in this paper is more efficient than that of the classical Gram-Schmidt orthogonalization. it needs less work than that of the PINV algorithm in MATLAB and preserves good numerical behavior.

Acknowledgement. The author would like to thank Prof. A. Tognoli for his help and encouragement when he visited the University of Trento, Italy.

References

[1] Å. Björck, Least squares methods, in: P.G. Ciarlet and J.L. Lions, ed., Handbook of Numerical Analysis, Vol. I, 1990.

[2] G.H. Golub and C.F. Van Loan, Matrix Computations, Second edition, The John Hopkins university Press, Baltimore, 1989.

[3] M. Hestenes, Conjugate Direction Methods in Optimization, Springer-Verlay, 1980.

[4] J. Zhao, Huang's method for solution of consistent linear equations and its generalization, *J. Numer. Math. of Chinese Univer.*, 3(1981).

[5] J. Zhao, The study of recurrence methods for solving ill-conditioned linear systems, Ph. D. thesis, Nanjing University, China, 1987.

[6] A. Ben-Israel and T.N.E. Greville, Generalized Inverse: Theory and Applications, New York, 1974.