AN OPTIMUM CUBICALLY CONVERGENT ITERATIVE METHOD OF INVERTING A LINEAR BOUNDED OPERATOR IN HILBERT SPACE

M. ALTMAN

1. In paper [1] we considered a power series method of inverting a linear bounded operator in Hilbert space. This method is actually an iterative method with the same speed of convergence as a geometric progression. A product of two linear operators we shall call briefly a multiplication. Thus, in general, a power series approximative method has the following two properties:

(1) at each iteration we use one multiplication;

(2) the convergence is linear.

In paper [2] we considered an iterative method of inverting an arbitrary linear bounded operator in a Hilbert space. This method requires two multiplications at each iteration step, and the convergence is quadratic. In the present paper we give an iterative method of inverting an arbitrary linear bounded operator in a Hilbert space. This method requires three multiplications at each iteration step and is cubically convergent. Thus, the quadratically convergent method which requires two multiplications at each iteration step may be called the iterative hyperpower method of order two. Analogously, the cubically convergent iterative method which requires three multiplications at each iteration step may be called the iterative hyperpower method of order three. The following two problems arise now in a natural way:

(1) Is it possible to construct an iterative hyperpower method of any degree?

(2) To give a comparison of the hyperpower methods of different degrees, and to answer the question whether there exists an optimum method. As a criterion for a hyperpower method to be better we can assume the following:

The method I is better than the method II if after some iteration steps using the same amount of multiplications for both methods, the method I gives better accuracy. In this paper we construct a certain class of iterative hyperpower methods and for this class the answers to both questions mentioned above is positive. It turns out that the optimum method of this class is the iterative hyperpower method of degree three.

Let A be a linear (i.e. additive and homogeneous) bounded operator with the domain and the range in a Banach space X.

Let us assume that the operator A is non-singular, i.e. A has an

Received October 5, 1959. Based on research supported by O.N.R., U.S.A.

inverse A^{-1} defined on the space X. Let us suppose that the linear bounded operator R_1 is an approximate reciprocal of A. Suppose also that R_1 satisfies the following condition

$$\| I - AR_1 \| = a < 1 , \tag{1}$$

where I is the identity mapping of X

Let p be any positive integer such that $p \geqq 2$. We shall construct an iterative hyperpower method of degree p with the following property

$$I - AR_{n+1} = (I - AR_n)^p , \tag{2}$$

where (R_n) is the sequence of the approximate inverses of A. It is easy to see that this sequence can be defined as follows

$$R_{n+1} = R_n(I + T_n + T_n^2 + \cdots + T_n^{p-1}) , \tag{3}$$

where

$$T_n = I - AR_n , \qquad n = 1, 2, \cdots . \tag{4}$$

Multiplying both sides in (3) by A we get by (4)

$$AR_{n+1} = (I - T_n)(I + T_n + T_n^2 + \cdots + T_n^{p-1}) = I - T_n^p .$$

Hence we obtain the relationship (2).

Thus, we have the following theorem.

THEOREM 1. *The sequence of the approximate inverses R_n defined by formula* (3) *converges in the norm of operators toward the inverse of the non-singular operator A, provided that R_1 satisfies condition* (1). *The error estimate is given by the formula*

$$\| A^{-1} - R_{n+1} \| \leqq \| A^{-1} \| a^{p^n} \tag{5}$$

or

$$\| A^{-1} - R_{n+1} \| \leqq \| R_1 \| \frac{a^{p^n}}{1 - a} \tag{6}$$

Proof. Formula (2) gives by induction

$$AR_{n+1} = I - T_1^{p^n} . \tag{7}$$

Hence we get by (7)

$$R - R_{n+1} = RT_1^{p^n} \tag{8}$$

or

$$R - R_{n+1} = R_1(I - T_1)^{-1}T_1^{p^n} . \tag{9}$$

Formula (5) follows from (8) and formula (6) follows from (9).

For $p = 2$ formula (3) yields

$$R_{n+1} = R_n(2I - AR_n) . \tag{10}$$

This case was considered in [23]. For $p = 3$ we get

$$R_{n+1} = R_n(I + (I - AR_n) + (I - AR_n)^2) \tag{11}$$

or

$$R_{n+1} = R_n(3I - 3AR_n + (AR_n)^2) \tag{12}$$

Thus, we have a class of methods with the property (2).

The question is now if there is an optimum method in this class of methods. To compare two methods we shall use the criterion mentioned above, i.e. the method is better if using the same number of multiplications gives a better accuracy.

Let p and q be two different positive integers. Consider the correspondings methods M_p and M_q defined by the formula (3). At each iteration step the method M_p takes p multiplications and the method M_p takes q multiplications in the sense defined above. Suppose that after a certain number of iteration steps which is different for both methods we get the same number of multiplications which is equal to

$$mp = nq . \tag{13}$$

Then in virtue of (5) the accuracy of the methods M_p and M_q is given by the exponents p^m and q^n respectively. Suppose that

$$p^m > q^n .$$

Then we have by (13)

$$p^m > q^s ,$$

where

$$s = \frac{mp}{q} .$$

Hence, we have

$$p^{1/p} > q^{1/q} . \tag{14}$$

The inequality (14) shows that we obtain the optimum method M_p for p such that the function $p^{1/p}\,(p = 2, 3, \cdots)$ achieves its maximum. This is the case when $p = 3$ since the maximum of the function

$$y = x^{1/x} , \quad x > 0$$

is attained at $x = e$.

2. We shall now apply Theorem 1 in order to find the approximate inverse of a linear bounded operator in a Hilbert space. Thus, we suppose that X is a Hilbert space H and A is a non-singular linear bounded operator with the domain and the range in H.

Let us begin with the case when A is a self-adjoint and positive definite operator, or, more precisely

$$A^* = A ,$$

where A^* is the adjoint of A, and A satisfies the condition

$$m(x, x) \leqq (Ax, x) \leqq M(x, x) ,$$

where $0 < m < M$, and m, M are the minimum and maximum eigenvalues of A respectively.

Consider the linear operator.

$$T_\alpha = I - \alpha A , \qquad 0 < \alpha < 2/M .$$

In virtue of the critical value theorem[1] we have

$$\frac{M - m}{M + m} \leqq \| T_\alpha \| = a_\alpha < 1 \quad \text{if} \quad 0 < \alpha < 2/M . \tag{13}$$

The minimum of the norm $\| T_\alpha \|$ is equal to

$$a_c = \| T_{\alpha_c} \| = \frac{M - m}{M + m} \tag{14}$$

and is reached precisely at the critical value α_c of A, i.e. for

$$\alpha = \alpha_c = \frac{2}{M + m} .$$

Thus, we get the following theorem.

THEOREM 2. *Let us suppose that A is a self-adjoint positive defined linear operator. If*

$$R_1 = \alpha I \quad \text{for} \quad 0 < \alpha < 2/M , \tag{15}$$

then the sequence of operators R_n determined by the iterative process in (3) *converges in the norm of the operators toward the inverse of A. The error estimate is given by the following formula*

$$\| A^{-1} - R_n \| \leqq \frac{1}{m} a_\alpha^{p^n} \tag{16}$$

[1] See [1], [2]

or

$$\| A^{-1} - R_n \| \leqq \frac{\alpha}{1 - a_\alpha} a_\alpha^{p^n} , \tag{17}$$

where $a_\alpha = \| T_\alpha \|$. *The convergence is best for the critical value of* A, i.e. *for* $\alpha = \alpha_c = 2/M + m$. *In this case* a_α *in formulae* (16) *and* (17) *should be replaced by* a_c *defined in* (14).

Putting $p = 3$ in Theorem 2 we get the theorem for the optimum method. Thus, we have

COROLLARY 1. *The iterative process defined by the formula* (11) *or* (12) *converges cubically toward the inverse of* A *provided that* R_1 *is defined by* (14). *The error estimate is given by formula* (16) *or* (17), *where* $p = 3$. *The convergence is best for the critical value of* A, i.e. *for* $\alpha = \alpha_c = 2/M + m$. *In this case* a_α *in formulae* (16) *and* (17) *should be replaced by* a_c *defined in* (14).

REMARK 1. The convergence of the iterative process is uniform with respect to α for any closed interval contained in the interval $0 < \alpha < 2/M$. Let us observe that α in (15) can be replaced by any number $1/K$, where K is greater than $\| A \|$. However, the convergence is faster when K is smaller. If the operator A is defined by a matrix

$$A = (a_{ij}) \qquad i, j = 1, 2, \cdots, k \tag{18}$$

satisfying the conditions of Theorem 2, then K can be replaced by any of the following numbers

$$\max_i \sum_{j=1}^{k} | a_{ij} | \; ; \quad \max_j \sum_{i=1}^{k} | a_{ij} | \; ; \quad \Big(\sum_{i,j=1}^{k} | a_{ij} |^2 \Big)^{1/2} . \tag{19}$$

However, the convergence is faster when K is smaller.

3. We shall now consider the general case when A is an arbitrary non-singular linear bounded operator in H.

Since the operator AA^* is self-adjoint and positive definite, we have the following inequalities

$$m^2(x, x) \leqq (AA^*x, x) \leqq M^2(x, x) ,$$

where $0 < m^2 < M^2$ and m^2, M^2 are the minimum and the maximum eigenvalues of AA^* respectively.

Let us consider the linear operator

$$T_\alpha = I - \alpha AA^* , \qquad 0 < \alpha < 2/M^2 .$$

Using the same argument as in §2, we get the following inequalities instead of (13).

$$\frac{M^2 - m^2}{M^2 + m^2} \leqq \| T_\alpha \| = a_\alpha < 1 \quad \text{if} \quad 0 < \alpha < 2/M^2 . \tag{20}$$

The minimum of the norm $\| T_\alpha \|$ is reached at

$$\alpha = \alpha_c = \frac{2}{M^2 + m^2}$$

and is equal to

$$a_c = \| T_{\alpha_c} \| = \frac{M^2 - m^2}{M^2 + m^2} . \tag{21}$$

Thus we obtain the following theorem.

THEOREM 3. *If*

$$R_1 = \alpha A^* \quad \text{for} \quad 0 < \alpha < 2/M^2 , \tag{22}$$

then the sequence of operators R_n determined by the iterative process in (3) *converges in the norm of the operators toward the inverse of A. The error estimate is given by the formulae* (16) *or* (17), *where a_α should be replaced by the expression in* (18). *The convergence is best for*

$$\alpha = \alpha_c = \frac{2}{M^2 + m^2} .$$

For the error estimate in this case a_α in formulae (16) *and* (17) *should be replaced by a_c defined in* (21).

Putting $p = 3$ in Theorem 3 we get the theorem for the optimum method in general case. Thus we have

COROLLARY 2. *If R_1 is determined by* (22) *then the iterative process defined in* (11) *or* (12) *converges cubically toward the inverse of A. For the error estimate we have the formulae* (16) *or* (17), *where $p = 3$. The convergence is best for the critical value of AA^*,* i.e. *for*

$$\alpha = \alpha_c = \frac{2}{M^2 + m^2} .$$

In this case a_α in formulae (16) *and* (17) *should be replaced by a_c determined in* (21).

REMARK 2. The convergence of the iterative process defined by Theorem 3 is uniform with respect to α for any closed interval contained in the interval $0 < \alpha < 2/M^2$. Let us remark that α in (22) can be replaced by any number $1/K$, where K is greater than $\| A \|^2$. However, the convergence is faster when K is smaller.

If the operator A is defined by the non-singular matrix in (18), then for K we can take any of the numbers in (19) with the matrix AA^* replacing the matrix A. We can also take for K any of the squared numbers in (19).

The table below shows the difference in rate of convergence between the following three method: I, II, III, where

I is the power series method considered in [1] (see page 52)

II is the quadratically convergent defined in (10)

III is the cubically convergent optimum method defined in (11) or (12).

Number of Iterations			Number of Multiplication			Accuracy ($a < 1$)		
I	II	III	I	II	III	I	II	III
6	3	2	6	6	6	a^6	a^8	a^9
12	6	4	12	12	12	a^{12}	a^{64}	a^{81}
18	9	6	18	18	18	a^{18}	a^{512}	a^{729}
24	12	8	24	24	24	a^{24}	a^{4096}	a^{6561}

REFERENCES

1. M. Altman, *Inversion of non-singular linear bounded operators in Hilbert space with application to matrix calculus.*

2. ———, *Inversion of linear bounded operators in Hilbert space with application to matrix calculus.*

CALIFORNIA INSTITUTE OF TECHNOLOGY, PASADENA,
ACADEMY OF SCIENCES, WARSAW.