AN ITERATIVE SOLUTION OF THE QUADRATIC EQUATION IN BANACH SPACE

J. E. McFARLAND[1]

I. **Introduction.** The solutions of the quadratic equation $ax^2+bx+c=0$, where a, b, and c are real numbers, are given in exact form by the quadratic formula. Moreover, if the roots are real and irrational, Newton's method or continued fractions may be used to approximate the solutions.

A generalized form of this problem, the quadratic equation in Banach space, is not at present so easily handled; many questions remain to be answered concerning the nature and number of roots and the best methods for finding them. The generalization of familiar methods has probably contributed the most satisfactory answers to these questions. Newton's method has been generalized to Banach space by Kantorovich [**3**], and in a particular case, this method has been used for the numerical solution of a "quadratic" integral equation of Chandrasekhar [**5**]. In a recent paper, Rall has proposed a quadratic formula in Banach space [**6**].

This paper presents the results of investigations on the continued fraction approach. An iterative process is given for finding roots of the quadratic equation in Banach space. The process is essentially a generalization to Banach space of a periodic continued fraction. A method is also given for extending the iterative process to include a wider range of application.[2]

II. **Basic concepts.** The preliminary considerations necessary for our investigation constitute generally common knowledge of Banach space, linear and bilinear operators in a Banach space. Information about these concepts and illustrative examples are readily available from several sources [**1**; **3**; **6**; **7**]. Most of such knowledge is assumed here; only the especially pertinent facts will be noted.

Let X hereafter denote a Banach space, that is, a complete normed linear space, and θ its null element of addition. If $x\in X$, we shall denote the norm of x by $\|x\|$. A linear operator L in X is a single-

Presented to the Society June 21, 1958; received by the editors January 9, 1958.

[1] This paper was prepared with financial assistance from the Office of Ordnance Research, U. S. Army, under terms of Contract DA-04-200-Ord-177 with Oregon State College.

[2] The author wishes to thank Dr. L. B. Rall who suggested the "continued fraction" approach and Dr. A. T. Lonseth who directed the preparation of this paper.

valued map of X into itself, where Lx shall denote the image of $x \in X$ under the map L. L is (1) additive, (2) continuous, and (3) homogeneous: (1) $L(x+y)=Lx+Ly$ for all $x, y \in X$; (2) if $x_n \to x$ as $n \to \infty$, then $Lx_n \to Lx$ as $n \to \infty$ ($\|x_n - x\| \to 0$ as $n \to \infty$ implies $\|Lx_n - Lx\| \to 0$ as $n \to \infty$); (3) $L(\lambda x)=\lambda(Lx)$ for any complex λ. Moreover, an additive and continuous operator is (4) bounded [**2**, p. 16]: (4) there exists a non-negative real number M such that

$$\|Lx\| \leqq M \cdot \|x\| \text{ for all } x \in X.$$

The norm of L is defined by

$$\|L\| = \text{greatest lower bound } \{M: \|Lx\| \leqq M \cdot \|x\| \text{ for all } x \in X\},$$

so that

$$\|Lx\| \leqq \|L\| \cdot \|x\| \text{ for all } x \in X.$$

If R and S are linear operators in X, then $(R+S)$ and (RS) are again linear operators in X and $\|R+S\| \leqq \|R\| + \|S\|$, $\|RS\| \leqq \|R\| \cdot \|S\|$ [**4**, p. 194]. If there exists a linear operator P in X such that $PL=LP=I$, where $Ix=x$ for all $x \in X$, then P is denoted by $P=L^{-1}$ and is called the inverse of L. A sufficient condition that $(I+L)^{-1}$ exist is that $\|L\| < 1$; in this case, $\|(I+L)^{-1}\| \leqq 1/(1-\|L\|)$ [**3**, p. 24]. With the definitions of the norm and sum of linear operators and of scalar multiplication, the collection of all linear operators in X forms a Banach space [**2**, pp. 32–33]. Denote this space by (X).

A bilinear operator is a linear operator mapping X into (X), that is $(Bx) \in (X)$ for each $x \in X$ [**6**, p. 5]. Thus, for each $x, y \in X$, $Bxy = (Bx)y$ is an element of X. From these facts we deduce that if $x, y, z \in X$, $Bx(y+z)=Bxy+Bxz$ and $B(x+y)z=Bxz+Byz$; moreover, since Bx and B are bounded, $\|Bx\| \leqq \|B\| \cdot \|x\|$ and $\|Bxy\| \leqq \|B\| \cdot \|x\| \cdot \|y\|$ for all $x, y \in X$.

To every bilinear B there correspond two bilinear operators: B^*, the permutation of B, defined by $B^*xy, =Byx$ for all $x, y \in X$; and $\overline{B}$, the mean of B, defined by

$$\overline{B}xy = \frac{1}{2}\{Bxy + Byx\} = \frac{1}{2}\{B + B^*\}xy,$$

for all $x, y \in X$. A bilinear operator B such that $B = B^* = \overline{B}$ is said to be symmetric.

III. **Iterative solution of the quadratic equation.** Consider the quadratic equation

$$Bxx + Ax = y, \tag{3.1}$$

where B is a bilinear operator in X, A is a linear operator in X, and y is a given element of X. We seek solutions $x\in X$ satisfying (3.1). Rall has shown that since

$$Bxx + Ax = B^*xx + Ax = \overline{B}xx + Ax,$$

there is no loss of generality in the assumption that B is symmetric [6, pp. 7–8]. This assumption will be made throughout the remainder of this paper.

We shall investigate the iterative procedure given by

$$\begin{aligned} &F_0 = z, \\ &F_{n+1} = (A + BF_n)^{-1}y, \qquad n = 0, 1, 2, \cdots, \end{aligned} \tag{3.2}$$

where $z\in X$. Note that if $F_n\in X$, then $(A+BF_n)\in(X)$, so that if $F_n\in X$ and $(A+BF_n)^{-1}$ exists, it follows that $F_{n+1}\in X$. Since $F_0 = z\in X$, we shall say that F_{n+1} is defined if $(A+BF_k)^{-1}$ exists for $k=0, 1, 2, \cdots, n$. If F_n is defined for $n=0, 1, 2, \cdots$, (3.2) will be said to be defined. Before proceeding, let us justify the investigation of (3.2).

THEOREM 1. *If* (3.2) *is defined and if there exists an* $x\in X$ *such that* $F_n\to x$ *as* $n\to\infty$, *then* x *is a solution of* (3.1).

PROOF. Since $x\in X$, $(A+Bx)\in(X)$. Consider $[(A+Bx)F_{n+1}-y]$. For $n\geqq 1$,

$$\begin{aligned} (A + Bx)F_{n+1} - y &= (A + Bx)(A + BF_n)^{-1}y - y \\ &= [(Bx)(A + BF_n)^{-1} + A(A + BF_n)^{-1} - I]y \\ &= (Bx - BF_n)(A + BF_n)^{-1}y \\ &= (Bx - BF_n)F_{n+1}. \end{aligned}$$

Therefore,

$$\|(A + Bx)F_{n+1} - y\| \leqq \|B\|\cdot\|x - F_n\|\cdot\|F_{n+1}\|.$$

But $\|x-F_n\|\to 0$ as $n\to\infty$ and $\{F_n\}$ is a bounded sequence, so

$$\|(A + Bx)F_{n+1} - y\| \to 0 \text{ as } n \to \infty$$

that is, $(A+Bx)F_{n+1}\to y$ as $n\to\infty$. Since $(A+Bx)$ is continuous, we have

$$(A + Bx)x = y$$

or

$$Ax + Bxx = y.$$

Hence, (3.2) provides an approximate solution to (3.1) if (3.2) is defined and $\{F_n\}$ converges in X. The following theorem gives sufficient conditions that (3.2) be defined.

THEOREM 2. *If*

$$A^{-1} \textit{ exists, and if} \tag{3.3}$$

$$0 < \|A^{-1}\| \cdot \|B\| \cdot \|w\| \leqq \frac{1}{4}, \quad \textit{where} \quad w = A^{-1}y, \tag{3.4}$$

and

$$\begin{aligned} &\frac{1 - (1 - 4\|A^{-1}\| \cdot \|B\| \cdot \|w\|)^{1/2}}{2} \\ &\qquad \leqq \|A^{-1}Bz\| \leqq \frac{1 + (1 - 4\|A^{-1}\| \cdot \|B\| \cdot \|w\|)^{1/2}}{2}, \end{aligned} \tag{3.5}$$

then (3.2) *is defined.*

PROOF. Now $(A+BF_n)^{-1}$ exists if $(I+A^{-1}BF_n)^{-1}$ exists, since $(A+BF_n)^{-1}=(I+A^{-1}BF_n)^{-1}A^{-1}$. We shall show the existence of $(I+A^{-1}BF_n)^{-1}$ for all n. First note that if F_n is defined and $\|A^{-1}BF_n\| <1$, then $(I+A^{-1}BF_n)^{-1}$ exists and $F_{n+1}=(I+A^{-1}BF_n)^{-1}w$ is defined. Hence since $\|A^{-1}BF_0\|=\|A^{-1}Bz\|<1$, the proof follows by induction on the inequality $\|A^{-1}BF_n\| \leqq \|A^{-1}Bz\|$. For $n=0$, equality holds. Assume $\|A^{-1}BF_k\| \leqq \|A^{-1}Bz\|$ for $k=0, 1, 2, \cdots, n$. Then

$$\begin{aligned} \|A^{-1}BF_{n+1}\| &= \|A^{-1}B(I + A^{-1}BF_n)^{-1}w\| \\ &\leqq \frac{\|A^{-1}\| \cdot \|B\| \cdot \|w\|}{1 - \|A^{-1}BF_n\|} \\ &\leqq \frac{\|A^{-1}\| \cdot \|B\| \cdot \|w\|}{1 - \|A^{-1}Bz\|}, \end{aligned}$$

so $\|A^{-1}BF_{n+1}\| \leqq \|A^{-1}Bz\|$ if

$$(\|A^{-1}Bz\|)^2 - \|A^{-1}Bz\| + \|A^{-1}\| \cdot \|B\| \cdot \|w\| \leqq 0,$$

which is true by (3.4) and (3.5).

If (3.4) and (3.5) are strengthened slightly, (3.2) may be shown to converge, and an error bound is obtained.

THEOREM 3. *If*

$$A^{-1} \textit{ exists, and if} \tag{3.6}$$

$$0 < \|A^{-1}\| \cdot \|B\| \cdot \|w\| \leqq \delta < 1/4, \tag{3.7}$$

and if

$$\frac{1-(1-4\delta)^{1/2}}{2} \leqq \|A^{-1}Bz\| < \frac{1}{2}, \tag{3.8}$$

then there exists an $x \in X$ *such that* $F_n \to x$ *as* $n \to \infty$, *and*

$$\|x - F_n\| \leqq \frac{\beta^n}{1-\beta}\|(A + Bz)^{-1}y - z\|, \tag{3.9}$$

where

$$\beta = \frac{\|A^{-1}Bz\|}{1 - \|A^{-1}Bz\|} < 1. \tag{3.10}$$

Proof. By Theorem 2, (3.2) is defined. Consider $(F_{n+1} - F_n)$. For $n \geqq 1$,

$$\begin{aligned} F_{n+1} - F_n &= (A + BF_n)^{-1}y - (A + BF_{n-1})^{-1}y \\ &= (A + BF_n)^{-1}(BF_{n-1} - BF_n)(A + BF_{n-1})^{-1}y \\ &= (I + A^{-1}BF_n)^{-1}A^{-1}B(F_{n-1} - F_n)F_n \\ &= (I + A^{-1}BF_n)^{-1}A^{-1}BF_n(F_{n-1} - F_n), \end{aligned}$$

so

$$\begin{aligned} \|F_{n+1} - F_n\| &\leqq \frac{\|A^{-1}BF_n\|}{1 - \|A^{-1}BF_n\|}\|F_n - F_{n-1}\| \\ &\leqq \frac{\|A^{-1}Bz\|}{1 - \|A^{-1}Bz\|}\|F_n - F_{n-1}\| = \beta \cdot \|F_n - F_{n-1}\|. \end{aligned} \tag{3.11}$$

Induction on (3.11) gives

$$\|F_{n+1} - F_n\| \leqq \beta^n \cdot \|F_1 - F_0\|. \tag{3.12}$$

Using (3.12), it follows that if $n \geqq 0$, $p = 1, 2, \cdots$,

$$\|F_{n+p} - F_n\| \leqq \sum_{k=n}^{n+p-1} \|F_{k+1} - F_k\| \leqq \sum_{k=n}^{n+p-1} \beta^k \cdot \|F_1 - F_0\|,$$

and since $0 < \beta < 1$,

$$\|F_{n+p} - F_n\| \leqq \sum_{k=n}^{\infty} \beta^k \cdot \|F_1 - F_0\| = \frac{\beta^n}{1-\beta}\|F_1 - F_0\|. \tag{3.13}$$

Hence $\{F_n\}$ is a Cauchy sequence, and there exists an $x \in X$ such that $F_n \to x$ as $n \to \infty$. Letting $p \to \infty$ in (3.13) we obtain

$$\|x - F_n\| \leqq \frac{\beta^n}{1-\beta}\cdot\|F_1 - F_0\|.$$

In the application of (3.2) to many particular problems, the exact values of $\|A\|$ and $\|B\|$ are not known, but upper bounds for them can be calculated. In this case the validation of the left-hand inequalities of (3.7) and (3.8) might be considered a rather troublesome task. In the case of (3.7), a clarification resolves this difficulty. If (3.6) holds, then $\|A^{-1}\|>0$, so $\|A^{-1}\|\cdot\|B\|\cdot\|w\|=0$ if and only if $\|w\|=0$ or $\|B\|=0$. Since A^{-1} exists, $\|w\|=\|A^{-1}y\|=0$ if and only if $y=\theta$ [6, p. 5], in which case (3.2) gives the particular root $x=\theta$ for (3.1). If $\|B\|=0$, then $Bxy=\theta$ for all $x, y\in X$, so (3.1) becomes the linear equation $Ax=y$.

Condition (3.8) may be completely eliminated with a particular choice of z in (3.2). If $z=A^{-1}y=w$, and if (3.6) and (3.7) are satisfied, it may be proved that (3.2) is defined and converges to an element $x\in X$, where

$$\|x - F_n\| \leqq \frac{(4\delta)^n}{1-4\delta}\|F_1 - F_0\|.$$

The existence and convergence proofs are similar to Theorems 2 and 3, using in this case the fact that $\|A^{-1}BF_n\|\leqq 1/2$ for all n.[3] Although $z=w$ is certainly a convenient choice, it may not be the best choice. If it is possible to choose z so that (3.8) is satisfied and z is reasonably close to the actual solution of (3.1), much quicker convergence is obtained as indicated by (3.9).

Mention has been made in the introduction that (3.2) is a continued fraction approach to the solution of (3.1). To make this fact more apparent, let X be the space of real numbers with $\|x\|=|x|$, and consider the quadratic equation

$$ax^2 + bx + c = 0, \tag{3.14}$$

where a, b, and c are real numbers. Now a is a bilinear operator, $\|a\|=|a|$, and b is a linear operator with $b^{-1}=1/b$ and $\|b^{-1}\|=1/|b|$. Thus, according to Theorem 3, if $b\neq 0$, $a\neq 0$, $c\neq 0$, and $b^2-4|ac|>0$, a root x of (3.14) as given by (3.2) is the continued fraction

$$x = \frac{(-c)}{b\ +}\ \frac{a(-c)}{b\ +}\ \frac{a(-c)}{b\ +}\cdots.$$

IV. **A method for extending the process.** Suppose (3.1) fails to satisfy the conditions of Theorem 3. One might still hope to find

[3] The bilinear operator B is not assumed symmetric in these proofs.

(3.2) applicable. For some iterative processes, such a hope is realized with a more judicious "guess" with which to start the iteration. However, according to (3.6) and (3.7) we cannot assure ourselves of such food fortune with (3.2). An extension is possible, nevertheless.

Consider first the quadratic equation

$$Bxx = y, \tag{4.1}$$

an equation where (3.2) is not directly applicable. If we make the substitution $x=u+v$, u, $v\in X$, since B is symmetric we obtain

$$Buu + 2(Bv)u = y - Bvv, \tag{4.2}$$

a quadratic of form (3.1). Thus, if we can choose a $v\in X$ such that $(Bv)^{-1}$ exists and $0<\|(Bv)^{-1}\|\cdot\|B\|\cdot\|(Bv)^{-1}y-v\|<1$, then (3.2) may be applied to (4.2) to find a solution u. The solution of (4.1) is then given by $x=u+v$. A known approximation to a root of (4.1) would be the logical choice for v.

The same approach may be used for (3.1). In this case the transformed equation becomes

$$Buu + (2Bv + A)u = y - Av - Bvv. \tag{4.3}$$

A solution u of (4.3) can be obtained by (3.2) if v is so chosen that $(2Bv+A)^{-1}$ exists and

$$0 < \|(2Bv + A)^{-1}\|\cdot\|B\|\cdot\|(2Bv + A)^{-1}(y - Av - Bvv)\| < 1/4.$$

Once again, a known approximation to a root of the original Equation (3.1) would be the logical choice for v.

Bibliography

1. S. Banach, *Théorie des opérations linéaires*, Monografje Matematyczne, vol. 1, Warsaw, 1932.

2. E. Hille, *Functional analysis and semi-groups*, Amer. Math. Soc. Colloquium Publications, New York, 1948.

3. L. V. Kantorovich, *Functional analysis and applied mathematics*, Tr. by C. D. Benster, Los Angeles, Institute for Numerical Analysis, 1952.

4. A. T. Lonseth, *The propagation of error in linear problems*, Trans. Amer. Math. Soc. vol. 6 (1947) pp. 193–212.

5. L. B. Rall, *An application of Newton's method to the solution of a non-linear integral equation* (U. S. Army Off. Ord. Research. Project: Numerical Solution of Integral Equations. Technical Report No. 7). Unpublished mimeographed report, Dept. of Math., Oregon State College.

6. ———, *The quadratic formula in Banach space*, (Tech. Report No. 10 of the same series as [5]), Unpublished mimeographed report, Dept. of Math., Oregon State College.

7. A. C. Zaanen, *Linear analysis*, New York, Interscience, 1953.

Oregon State College