ON THE APPROXIMATION OF SOLUTIONS OF COMPACT OPERATOR EQUATIONS

IOANNIS K. ARGYROS*

Abstract.

We approximate in isolated solution of a compact operator equation using the solutions of a family of collectively compact operator equations.

Key Words and Phrases: Collectively compact, Banach space, quadratic operator.

A. M. S. 1980 Classification codes: 65, 46(B15).

Introduction.

We study the quadratic equation

* Department of Mathematics, New Mexico State University.
Las Cruces, New Mexico 88003.

$$x = y + Q(x) \quad (1)$$

in a Banach space X , where $y \in X$ is fixed and Q is a compact quadratic operator from a subset K of X into X [2], [9]. We assume in the first part of this paper that solutions of the family of quadratic equations

$$x = y + Q_n(x), \quad n = 1, 2, \dots \quad (2)$$

where $\{Q_n\}$, $n = 1, 2, \dots$ are collectively compact and the Q_n converge pointwise to Q are known. We then use these solutions to approximate a solution of (1).

For linear equations, if (1) is nonsingular, (2) is uniquely solvable for sufficiently large n and the approximate solutions converge to the true solution [1].

For nonlinear equations, under certain differentiability conditions on Q and Q_n , $n = 1, 2, \dots$, (2) has a unique solution in a neighborhood of an isolated solution of (1) if n is sufficiently large and these solutions converge to the isolated solution [8].

In [4] use of the Schauder-Leray Theory has been made to extend the above results.

Here we give a simplified proof based on the contraction mapping principle, of the existence, uniqueness and convergence of solutions of (2) in a neighborhood of an isolate solution of (1).

In the second part we show how can we find all the solutions of each one of the equations in (2). In particular when the operator Q_n are of finite rank n , we show that the problem of finding solutions of

(2) is equivalent to solving an algebraic system in $\mathbb{R}^n$ (or $\mathbb{C}^n$).

I. General approximation theory.

Let B be the unique symmetric bilinear operator associated with Q [2], [9]. In case B , or Q are bounded operators on K [9], we can prove the following proposition:

Proposition 1. If either Q or its unique associated symmetric bilinear operator B is bounded, then both are, and

$$||Q|| \leq ||B|| \leq 2||Q||. \quad (3)$$

Moreover, if the parallelogram law holds then $||Q|| = ||B||$.

Proof. Let B be bounded, since $Q(x) = B(x, x)$ we have

$$||Q(x)|| = ||B(x, x)|| \leq ||B|| \cdot ||x|| \cdot ||x|| \quad \text{for all } x \in K$$

so,

$$||Q|| \leq ||B|| \quad (4)$$

that is Q is also bounded.

Conversely: Let Q be bounded, since

$$4B(x, y) = Q(x + y) - Q(x - y)$$

we have

$$\begin{aligned}
4||B(x,y)|| &= ||Q(x+y) - Q(x-y)|| \\
&\leq ||Q(x+y)|| + ||Q(x-y)|| \\
&\leq ||Q|| [||x+y||^2 + ||x-y||^2] \quad (5) \\
&\leq ||Q|| [2(||x||^2 + ||y||^2) + 4||x|| \cdot ||y||]
\end{aligned}$$

so, for $||x||, ||y|| \leq 1$

$$\sup ||B(x,y)|| \leq 2||Q||$$

$$||x|| \leq 1$$

$$||y|| \leq 1$$

that is B is bounded and

$$||B|| \leq 2||Q|| \quad (6)$$

Combining inequalities (4) and (6) we obtain (3).

Finally, if the parallelogram law holds equality is easily achieved if we replace the bracket in (5) by $2(||x||^2 + ||y||^2)$ and proceed as above to obtain $||B|| \leq ||Q||$.

Let Q, Q_n be as before and let $B, B_n, n = 1, 2, \dots$ be the symmetric bilinear operators associated with Q and Q_n respectively. Let $x_o \in K$ be a solution of (1) then it can easily be checked that the first and second Fréchet derivatives of $Q_n, n = 1, 2, \dots$ at x_o are

$$Q_n'(x_o) = 2B_n(x_o),$$

$$Q_n''(x_o) = 2B_n, n = 1, 2, \dots \quad [6], [9].$$

Let $I - 2B(x_0) = I - Q'(x_0)$ be nonsingular. The fact that $Q'(x_0)$ exists implies that $Q'(x_0)$ is compact. Let $c > 0$ be a uniform bound on the norm of the Q_n 's, $n = 1, 2, \dots$ that is

$$\|Q_n\| \leq c, \quad n = 1, 2, \dots$$

Then the hypotheses on $\{Q_n\}$; $n = 1, 2, \dots$ implies that $\{Q_n'(x_0)\}$, $n = 1, 2, \dots$ is a sequence of collectively compact linear operators on X and $Q_n'(x_0)$ converges pointwise to $Q'(x_0)$. Moreover by applying the collectively compact approximation theory to $Q'(x_0)$, $\{Q_n'(x_0)\}$, $n = 1, 2, \dots$ we obtain that the linear operators $[I - Q_n'(x_0)]$ are nonsingular for sufficiently large n , say $n \geq N$, and

$$\|[I - Q_n'(x_0)]^{-1}\| \leq d < \infty, \quad n \geq N \quad \text{for some } d > 0.$$

We have now proved the following proposition:

Proposition 2. Assume that $[I - Q'(x_0)]$ is nonsingular and that Q, Q_n are as in the introduction and Proposition 1. Then the linear operators $[I - Q_n'(x_0)]$ are nonsingular with bounded inverses for sufficiently large n .

Definition 1. Let Q, Q_n, B, B_n be as before. Assume that the hypotheses of Proposition 2 are true. Let c, d denote the uniform bounds on the norms of the B_n and $(I - Q_n'(x_0))^{-1}$, $n = 1, 2, \dots$ respectively. Define the real functions $f(r)$ and $g(r)$ by

$$f(r) = 4d c r - 1$$

and

$$g(r) = d c r^2 - r + d c \|x_0\|^2 + d \|y\|$$

Note that

$$f(r) < 0$$

and

$$g(r) \leq 0$$

$$1 - 4d^2 c^2 \|y\| > 0$$

$$\text{if } 1 - 4d^2 c^2 \|x_0\|^2 - 4d^2 c \|y\| > 0 \quad (7)$$

$$\text{and } r_- \leq r < \frac{1}{4dc} \equiv R$$

where r_- is the small root of $g(r) = 0$.

We can now prove the main theorem:

Theorem 1. Let $x_0 \in X$ be an isolated solution of (1) satisfying (7). Moreover assume that the hypotheses of Proposition 2 are true. Then,

(a) there exists $r > 0$, with

$$r_- \leq r < R$$

such that for $n \geq N$, (2) has a unique solution

$$x_n \in \bar{U}(x_0, r) = \{x \in X \mid \|x - x_0\| \leq r\}$$

and

$$(b) \quad \lim_{n \rightarrow \infty} x_n = x_0.$$

Proof. (a) Define the operator T on X by

$$T(x) \equiv x - [I - Q_n'(x_o)]^{-1}[x - y - Q_n(x)].$$

Note that the fixed points of T are solutions of (2).

Claim 1. T is a contraction operator on $\bar{U}(x_o, r)$.

Let $w, v \in \bar{U}(x_o, r)$, then

$$\begin{aligned} \|T(w) - T(v)\| &= \|w - v - [I - Q_n'(x_o)]^{-1}[w - v - Q_n(w) + Q_n(v)]\| \\ &= \|[I - Q_n'(x_o)]^{-1}[B_n(x - x_o) + B_n(v - x_o)](w - v)\| \\ &\leq 4cd r \|w - v\|. \end{aligned}$$

So, T is a contraction if $f(r) < 0$ which is true by hypothesis (7).

Claim 2. T maps $\bar{U}(z, r)$ into $\bar{U}(z, r)$.

Let $w \in \bar{U}(x_o, r)$ then,

$$\begin{aligned} \|T(w) - x_o\| &\leq r \text{ if} \\ \|T(w) - x_o\| &= \|w - [I - Q_n'(x_o)]^{-1}[w - y - Q_n(w)] - x_o\| \\ &= \|[I - Q_n'(x_o)]^{-1}[Q_n(w - x_o) - Q_n(x_o) + y]\| \\ &\leq cdr^2 + cd\|x_o\|^2 + d\|y\| \leq r, \text{ or } g(r) \leq 0 \end{aligned}$$

which is true if (7) holds. Part (a) now follows from the contraction mapping principle [7], [9].

(b) we have,

$$\begin{aligned}
 (I - Q'_n(x_0))(x_n - x_0) &= (I - 2B_n(x_0))(x_n - x_0) \\
 &= B_n(x_n, x_n) - B(x_0, x_0) - 2B_n(x_0)(x_n - x_0) \\
 &= B_n(x_0, x_0) - B(x_0, x_0) + B_n(x_n - x_0, x_n - x_0) \\
 &= Q_n(x_0) - Q(x_0) + Q_n(x_n - x_0).
 \end{aligned}$$

Therefore,

$$||x_n - x_0|| \leq d ||Q_n(x_0) - Q(x_0)|| + dc ||x_n - x_0||^2$$

or

$$\begin{aligned}
 ||x_n - x_0|| &\leq \frac{d}{1 - dc} ||Q_n(x_0) - Q(x_0)|| \\
 &\leq \frac{d}{1 - dc} ||Q_n(x_0) - Q(x_0)|| \rightarrow 0 \quad \text{as } n \rightarrow \infty.
 \end{aligned}$$

That is $\lim_{n \rightarrow \infty} x_n = x_0$ and the theorem is proved.

In practice many examples of interest are such that the operators Q_n in (2) are bounded and of finite rank, therefore compact [7]. The evaluation of the parameter c is of considerable difficulty even then.

Here is an example that illustrates how the Q_n 's, $n = 1, 2, \dots$ may look like and how we evaluate c .

Example 1. Consider the Banach space $X = C[0, 1]$, the space of continuous functions on $[0, 1]$ equipped with the max-norm. Define the finite rank

operator Q on X by

$$Q(x)(s) = \int_0^1 (s-t)x^2(t)dt, \quad 0 \leq s \leq 1, \quad \text{rank}(Q) = 2.$$

Now,

$$|Q(x)(s)| = \left| s \int_0^1 x^2(t)dt - \int_0^1 tx^2(t)dt \right| \leq \|x\|^2 \left(1 + \frac{1}{2}\right), \quad \text{for all } x,$$

$$\max_s \|Q(x)\| \leq \frac{3}{2} \|x\|^2,$$

so

$$\|Q\| \leq \frac{3}{2}.$$

We can do better here: take a fixed x , $\|x\| \leq 1$, then

$$\|Q(x)\| = \max_s |sa - b|,$$

$$a = \int_0^1 x^2(t)dt \geq 0 \quad \text{and} \quad b = \int_0^1 tx^2(t)dt \geq 0.$$

Note, that $ab = \int_0^1 (1-t)x^2(t)dt \geq 0$, so

$$\|Q(x)\| = \max\{b, a-b\}.$$

But, $tx^2(t) \leq t$ for all $x^2(t) \leq 1$, that is

$$b \leq \int_0^1 t dt = \frac{1}{2}$$

and

$$(1-t)x^2(t) \leq 1-t \quad \text{for all } x^2(t) \leq 1$$

so,

$$a - b \leq \int_0^1 (1 - t) dt = \frac{1}{2}$$

and

$$||Q|| = \sup_{||x|| \leq 1} ||Q(x)|| \leq \max\{\frac{1}{2}, \frac{1}{2}\} = \frac{1}{2}.$$

Finally, take $x(t) = 1$ for all t , $0 \leq t \leq 1$, so $||x|| = 1$

and

$$||Q(x)|| = \sup_s \left| s \int_0^1 dt - \int_0^1 t dt \right| = \sup_s \left| s - \frac{1}{2} \right| = \frac{1}{2}$$

that is

$$||Q|| = \frac{1}{2}.$$

The problem of evaluating the x_n 's $n = 1, 2, \dots$ whose existence in $\bar{U}(x_o, z)$ is guaranteed by Theorem 1 may be as difficult as that of evaluating solutions of (1). In part II we show however, how to find all the solutions of (2) when the Q_n 's, $n = 1, 2, \dots$ are of finite rank. In particular we show that the problem of solving (2) is equivalent to solving a quadratic system in $\mathbb{R}^n$ or $(\mathbb{Q}^n)$. Assuming then that the algebraic system can be solved, we can select for each n , $n = N, N+1, \dots$, a solution x_n such that $x_n \in \bar{U}(x_o, r)$ and then apply Theorem 1.

From now on we use Q instead of Q_n in (2) for simplicity.

II. The solutions of the approximating equations.

Assume that Q is of finite rank $v = \dim(\text{span}(\text{Rang}(Q)))$ in (2) and set $z = x - y$ to obtain

$$z = Q(z + y)$$

The above equation implies that the problem of solving the auxiliary equation can be translated to a finite dimensional one since z must lie in $\text{rang}(Q)$.

Definition 2. Let A denote the set of all bounded quadratic operators Q in X such that Q has finite rank. Denote by E , the set of all bounded quadratic functionals f on X .

Let $f \in E$, $e \in X$; the operator $f \otimes e : X \rightarrow X$ to $f(x)e \in X$ is bounded quadratic operator of rank one. Thus

$$Q = \sum_{i=1}^n f_i \otimes e_i \in A$$

for any $f_i \in E$, $i = 1, 2, \dots, n$, $e_i \in X$, $i = 1, 2, \dots, n$.

Note that if $Q : X \rightarrow Y$ is a bounded quadratic operator and $L : Y \rightarrow Z$ is a bounded linear operator, then $L \circ Q : X \rightarrow Z$ is a bounded quadratic operator. (Q and L need not be a finite rank).

Definition 3. Denote by $E \otimes X$ the vector subspace generated in the space of all bounded quadratic operators by the set $\{Q \in A / Q = f \otimes e, f \in E, e \in X\}$ so $Q \in E \otimes X$ if and only if

$$Q = \sum_{i=1}^n f_i \otimes e_i.$$

Theorem 2. $A = E \otimes X$.

Proof. Let $\{e_1, \dots, e_n\}$ be a basis for $\text{rang}(Q)$ and choose g_i such that $g_i(e_j) = \delta_{ij}$, $i, j = 1, 2, \dots, n$. Since $\text{rang}(Q)$ is finite dimensional, the $\{g_i\}$, $i = 1, 2, \dots, n$ functionals are bounded and by the Hahn-Banach theorem they can be extended to bounded linear functionals on X without increasing their norms. Let

$$f_i = g_i \circ Q, \quad i = 1, 2, \dots, n.$$

Then the f_i , $i = 1, 2, \dots, n$ are bounded quadratic functionals and

$$Q = \sum_{i=1}^n f_i \otimes e_i.$$

Definition 4. Let f_i^* , $i = 1, 2, \dots, n$ denote the symmetric bilinear functionals associated with the f_i , $i = 1, 2, \dots, n$ where the f_i are bounded quadratic functionals, given by

$$f_i^*(x, y) = \frac{1}{4}(f_i(x + y) - f_i(x - y)).$$

Denote by C' the matrix of the linear transformation $2B(y)(o)$ restricted to $\text{rang}(Q)$ relative to the basis $e_1, \dots, e_n$. Define the $n \times n$ matrix C , by

$$C = I - C',$$

$$\underline{\ell} = \begin{bmatrix} \ell_1 \\ \vdots \\ \ell_n \end{bmatrix}, \quad \text{by } \ell_i = f_i(y), \quad i = 1, 2, \dots, n,$$

the block of matrices $\underline{\underline{C}}, \underline{\underline{C}} = \begin{bmatrix} C_1 \\ \vdots \\ C_n \end{bmatrix}$ by $C_i = \{c_i^{jk}\}$. where

$$c_i^{jk} = f_i^*(d_j, d_k), \quad i, j, k = 1, 2, \dots, n.$$

Define $\underline{\underline{v}}$ by $\underline{\underline{v}} = C^{-1} \underline{\underline{e}}$ if $|C| \neq 0$ and the block of matrices

$$\underline{\underline{M}} = \begin{bmatrix} M_1 \\ \vdots \\ M_n \end{bmatrix} \quad \text{with } M_k = |C|^{-1} M_k', \quad k = 1, 2, \dots, n \text{ is the } n \times n \text{ matrix which}$$

results from the determinant of the matrix C if we replace the k^{th} column by

$$\begin{bmatrix} C_1 \\ \vdots \\ C_n \end{bmatrix}. \quad \text{Define } \underline{\underline{CM}} \text{ by } \begin{bmatrix} CM_1 \\ \vdots \\ CM_n \end{bmatrix}.$$

Note that $M_k', \quad k = 1, 2, \dots, n$ is indeed a $n \times n$ matrix. For the case $n = 2$,

$$M_1' = \begin{vmatrix} C_1 & c_{12} \\ C_2 & c_{22} \end{vmatrix} = c_{22}C_1 - c_{12}C_2.$$

$$M_2' = \begin{vmatrix} c_{11} & C_1 \\ c_{21} & C_2 \end{vmatrix} = c_{11}C_2 - c_{21}C_1.$$

Theorem 3. The point $w \in X$ is a solution of (2) if and only if

$$w = y + \sum_{i=1}^n \xi_i e_i$$

where the vector $\underline{\xi} = \begin{bmatrix} \xi_1 \\ \vdots \\ \xi_n \end{bmatrix} \in R^n$ (or $\mathbb{C}^n$) is a solution of

$$\underline{x} = \underline{\ell} + C' \underline{x} + \underline{x}^+ \underline{C} \underline{x} \quad \text{in } R^n \text{ (or } \mathbb{C}^n \text{).} \quad (8)$$

Moreover, if $|C| = |I - C'| \neq 0$ the Cramer's rule transforms the above to

$$\underline{x} = \underline{v} + \underline{x}^+ \underline{M} \underline{x} \quad \text{in } R^n \text{ (or } \mathbb{C}^n \text{),} \quad (9)$$

Proof. Assume that (2) has a solution $w \in X$. Then

$$\begin{aligned} w &= y + Q(w) \\ &= y + \sum_{i=1}^n f_i(w) e_i. \end{aligned}$$

Apply $f_1, f_2, \dots, f_n$ in turn to this vector identity to obtain for $p = 1, 2, \dots, n$

$$\begin{aligned} f_p(w) &= f_p(y + \sum_{k=1}^n f_k(w) e_k) \\ &= f_p(y) + \sum_{k=1}^n f_k^2(w) f_p(e_k) + 2 \sum_{i=1}^n f_k(w) f_p^*(y, e_k) \end{aligned}$$

$$= 2 \sum_{i \neq j}^n f_i(w) f_j(w) f_p^*(e_i, e_j).$$

Letting

$$f_i(w) = x_i, \quad i = 1, 2, \dots, n$$

and writing these equations in vector form, we obtain

$$\underline{x} = \underline{\ell} + C' \underline{x} + \underline{x}^{+r} \underline{Cx}$$

or

$$\underline{Cx} = \underline{\ell} + \underline{x}^{+r} \underline{Cx}.$$

Since $|C| \neq 0$, we obtain (9) by composing both sides of the above equation by C^{-1} .

Conversely, given (9), assume (8) has a solution vector $\xi = \begin{bmatrix} \xi_1 \\ \vdots \\ \xi_n \end{bmatrix}$.

Let $w \in X$ be defined as

$$w = y + \sum_{i=1}^n \xi_i e_i.$$

Apply $f_1, f_2, \dots, f_n$ in turn to this vector identify to obtain for $p = 1, 2, \dots, n$,

$$f_p(w) = f_p(y) + \sum_{k=1}^n \xi_k^2 f_p(e_k) + 2 \sum_{k=1}^n \xi_k f_p^*(y \cdot e_k)$$

$$+ 2 \sum_{i \neq j}^n \xi_i \xi_j f_p^*(e_i, e_j),$$

or in matrix notation,

$$\underline{f(w)} = \underline{\ell} + C' \underline{\xi} + \underline{\xi}^{+r} \underline{C} \underline{\xi}.$$

Now since $\underline{\xi}$ satisfies (8) we have

$$\underline{\xi} = \underline{\ell} + C' \underline{\xi} + \underline{\xi}^{+r} \underline{C} \underline{\xi}$$

Comparing the last two equations, we get

$$\xi_i = f_i(w), \quad i = 1, 2, \dots, n,$$

so

$$w = y + \sum_{i=1}^n f_i(w) e_i,$$

or

$$w = y + Q(w).$$

Therefore, w is a solution of (2) and the theorem is proved.

Example 2. Let $X = C[0,1]$ and consider the equation

$$x(s) = s + s \int_0^1 x^2(t) dt$$

where $s \in [0,1]$. This equation is of the form (2), with $\text{rank}(Q) = 1$,

$$y(s) = s$$

$$e = s, \text{ and}$$

$$e(s) = \int_0^1 x^2(t) dt.$$

Using the formula,

$$f^*(v, w) = \frac{1}{4}(f(v + w) - f(v - w)),$$

we have

$$C = 1 - 2f^*(y, e) = 1 - 2\frac{1}{4} \int_0^1 4s^2 ds = \frac{1}{3}$$

$$\underline{\ell} = f(y) = f(s) = \int_0^1 s^2 ds = \frac{1}{3}$$

$$\underline{C} = f(e) = f(s) = \int_0^1 s^2 ds = \frac{1}{3}$$

$$\underline{v} = 3 \cdot \frac{1}{3} = 1$$

$$\underline{M} = 3 \cdot \frac{1}{3} = 1.$$

Therefore, (8) becomes

$$\xi = 1 + \xi^2 \quad \text{in } \mathbb{C} \quad \text{with solutions } \frac{1 + i\sqrt{3}}{2};$$

since $x = y + \xi e$, we finally have

$$x(s) = \left(\frac{3 + i\sqrt{3}}{2}\right)s.$$

REFERENCES.

- [1] Anselone, P. M., Collectively Compact Operator Approximation Theory, Prentice-Hall, Englewood Cliffs, N. J., 1971.
- [2] Argyros, I. K., Quadratic equations and applications to Chandrasekhar's and related equations., Bull. Austral. Math Soc., Vol. 32 (1985), 275-292.
- [3] ---, On a contraction theorem and applications. Proceedings of Symposium in Pure Math., A.M.S., Vol. 45, 1(1986), 51-53.
- [4] Atkinson, K. E., The numerical evaluation of fixed points for completely continuous operators. SIAM J. Num. Anal. 10(1973), 799-807.
- [5] Halmos, P., Finite dimensional vector spaces. D. Van. Nostrand. 1958.
- [6] Kelly, C. T., Approximation of solutions of some quadratic integral equations in Transport theory, J. Integ. Eq. 4, (1982), 221-237.
- [7] Krasnoleskii, M. A., Topological Methods in the theory of Nonlinear integral equations, McMillan, N. Y., 1964.
- [8] Moore, R. H., Approximation to Nonlinear operator equations and Newton's method. Numer. Math. 12 (1968), 23-29.
- [9] Rall, L.B., Computational solutions of nonlinear operator equations, Pergamon Press, 1978.