

Improvement by Iteration for Compact Operator Equations

By Ian H. Sloan

Abstract. The equation $y = f + Ky$ is considered in a separable Hilbert space H , with K assumed compact and linear. It is shown that every approximation to y of the form $y_{1n} = \sum^n a_{ni} u_i$ (where $\{u_i\}$ is a given complete set in H , and the a_{ni} , $1 \leq i \leq n$, are arbitrary numbers) is less accurate than the best approximation of the form $y_{2n} = f + \sum^n b_{ni} K u_i$, if n is sufficiently large. Specifically it is shown that if y_{1n} is chosen optimally (i.e. if the coefficients a_{ni} are chosen to minimize $\|y - y_{1n}\|$), and if y_{2n} is chosen to be the first iterate of y_{1n} , i.e. $y_{2n} = f + Ky_{1n}$, then $\|y - y_{2n}\| \leq \alpha_n \|y - y_{1n}\|$, with $\alpha_n \rightarrow 0$. A similar result is also obtained, provided the homogeneous equation $x = Kx$ has no nontrivial solution, if instead y_{1n} is chosen to be the approximate solution by the Galerkin or Galerkin-Petrov method. A generalization of the first result to the approximate forms y_{3n}, y_{4n}, \dots obtained by further iteration is also shown to be valid, if the range of K is dense in H .

1. Introduction. This paper is concerned with the approximate solution of the equation

$$(1) \quad y = f + Ky,$$

where y and f belong to a separable Hilbert space H , and K is a compact linear operator in H . The main results can also be extended to the eigenvalue equation $y = \lambda Ky$, by setting $f = 0$ and replacing K by λK . The whole discussion has immediate application to linear integral equations with square-integrable kernels, H then being an L^2 space.

A common way of approaching Eq. (1) in practice is to seek an approximate solution of the form

$$(2) \quad y_{1n} = \sum_{i=1}^n a_{ni} u_i,$$

where $\{u_i\}$ is some suitable complete set of linearly independent elements of H , and the a_{ni} are real or complex coefficients. Indeed, many numerical methods [2], [4] for the approximate solution of integral equations (e.g. Galerkin method, collocation, least squares and variational methods) make use of an approximate solution of the form (2), differing from each other only in the choice of the coefficients.

The main purpose of the present paper is to show that the alternative form

$$(3) \quad y_{2n} = f + \sum_{i=1}^n b_{ni} K u_i$$

Received April 21, 1975; revised December 30, 1975.

AMS (MOS) subject classifications (1970). Primary 45L05, 45B05, 65R05; Secondary 45C05.

Key words and phrases. Integral equation, compact kernel, Galerkin method, Galerkin-Petrov method, eigenvalue problem.

is always capable of better accuracy than (2), if n is sufficiently large. Specifically, let y_{1n}^* and y_{2n}^* denote the best possible approximations of the forms (2) and (3), i.e. y_{1n}^* and y_{2n}^* are the approximations that minimize the respective Hilbert space error norms $\|y - y_{1n}\|$ and $\|y - y_{2n}\|$. Then it is shown in the following section (Theorem 1) that $\|y - y_{2n}^*\| \leq \alpha_n \|y - y_{1n}^*\|$, where $\alpha_n \rightarrow 0$. Indeed, it is shown that even the choice $y_{2n} = f + Ky_{1n}^*$ (which is of the form (3), but not in general optimal) satisfies $\|y - y_{2n}\| \leq \alpha_n \|y - y_{1n}^*\|$, with $\alpha_n \rightarrow 0$. In other words, the best approximation of the form (2) is always improved by iteration (unless, of course, it is already exact), if n is sufficiently large. An explicit expression for α_n is given in the theorem; hence information is available, in principle, on how large n needs to be in any particular case for the improvement to be significant.

An extension of the result to the still more highly iterated approximate forms y_{3n}, y_{4n}, \dots , given by

$$y_{3n} = f + Kf + \sum_{i=1}^n c_{ni} K^2 u_i, \text{ etc.,}$$

is given in Theorem 2. The essential result is that the optimal approximation of the m th kind y_{mn}^* is always improved by a single iteration, if n is sufficiently large, provided (for $m > 1$) that the range of K is dense in H , or what is equivalent, that the null space of K^* is trivial.

For the various practical methods of the first kind, the solutions y_{1n} that arise are of course generally not optimal. For each such method, an interesting practical question is suggested by Theorem 1, namely: is it true that y_{1n} is necessarily improved by an iteration, if n is sufficiently large? In the present work (Theorem 3) this question is answered in the affirmative for the Galerkin-Petrov method, thereby extending a result previously obtained [6], in the course of a study of degenerate-kernel methods for integral equations, for the ordinary Galerkin (or Bubnov-Galerkin) method.

The result for the Galerkin methods, that iteration necessarily improves the accuracy if n is sufficiently large, is of practical importance, because the required iteration can be carried out without extra work, since the Galerkin methods already require the calculation of the quantities $Ku_i, i = 1, \dots, n$. The fact that iteration of the Galerkin solution is sometimes beneficial has been observed previously [2], on the basis of practical experience. It is apparent from some recent calculations [7] that the benefit obtainable by the iteration can be very striking indeed, even in cases where $\|K\| \gg 1$.

2. Principal Result. Let U_n be the subspace of H spanned by u_1, \dots, u_n , and let P_n be the orthogonal projection operator onto U_n . Then P_n has the properties $P_n^2 = P_n = P_n^*$, $\|P_n\| = 1$. The sequence $\{u_i\}$ is assumed to be complete in H , hence $P_n x \rightarrow x$ for all $x \in H$.

The best approximation to y of the form (2), i.e. the element $y_{1n} \in U_n$ that minimizes $\|y - y_{1n}\|$, is [1]

$$(4) \quad y_{1n}^* = P_n y.$$

Similarly, since $y - y_{2n}$ can be written, with the aid of Eqs. (1) and (3), as

$$y - y_{2n} = Ky - \sum_{i=1}^n b_{ni}Ku_i,$$

it is seen that $\|y - y_{2n}\|$ is minimized by choosing $\sum b_{ni}Ku_i$ to be the orthogonal projection of Ky onto KU_n (the span of Ku_1, \dots, Ku_n). Thus, the best approximation to y of the form (3) is

$$(5) \quad y_{2n}^* = f + P_{2n}Ky,$$

where P_{2n} is the orthogonal projection operator onto KU_n .

THEOREM 1. *If $y_{2n} = f + Ky_{1n}^*$, then*

$$\|y - y_{2n}^*\| \leq \|y - y_{2n}\| \leq \alpha_n \|y - y_{1n}^*\|,$$

where

$$\alpha_n = \|K - KP_n\| \rightarrow 0.$$

Proof. The first inequality follows from the optimal nature of y_{2n}^* . To prove the second inequality, note that

$$(6) \quad y - y_{2n} = (f + Ky) - (f + Ky_{1n}^*) = K(y - P_n y) = (K - KP_n)(y - P_n y),$$

giving

$$\|y - y_{2n}\| \leq \|K - KP_n\| \|y - y_{1n}^*\|.$$

The proof is completed by the standard result [5, p. 204] $\|K - KP_n\| \rightarrow 0$. (This follows, for example, by noting that

$$\|K - KP_n\| = \|K^* - P_n K^*\| = \sup_{x \in H, \|x\|=1} \|K^*x - P_n K^*x\|;$$

hence if the contrary $\|K - KP_n\| \not\rightarrow 0$ is supposed, then there exists $\epsilon > 0$ and a sequence $\{x_n\} \subset H$ such that $\|K^*x_n - P_n K^*x_n\| > \epsilon$, $\|x_n\| = 1$. Since K^* is compact, $\{K^*x_n\}$ may be assumed to converge to an element $v \in H$, which implies

$$\|K^*x_n - P_n K^*x_n\| \leq \|(I - P_n)(K^*x_n - v)\| + \|v - P_n v\| \rightarrow 0,$$

a contradiction.) Q.E.D.

The key to the proof of Theorem 1 is the factorization carried out in Eq. (6), by exploiting the projection property $P_n^2 = P_n$.

3. Generalization of Theorem 1. Let y_{mn} be an expression of the form

$$(7) \quad y_{mn} = \sum_{j=0}^{m-2} K^j f + \sum_{i=1}^n a_{mni} K^{m-1} u_i, \quad m \geq 1,$$

where the a_{mni} are numbers; and let P_{mn} denote the orthogonal projection operator onto $K^{m-1}U_n$ (the span of $K^{m-1}u_1, \dots, K^{m-1}u_n$), so that $P_{mn}^2 = P_{mn} = P_{mn}^*$ and $\|P_{mn}\| = 1$.

By repeated iteration of Eq. (1), the exact solution y can be written in a form similar to (7), namely

$$(8) \quad y = \sum_{j=0}^{m-2} K^j f + K^{m-1} y, \quad m \geq 1.$$

Hence, $y - y_{mn}$ can be written

$$y - y_{mn} = K^{m-1} y - \sum_{i=1}^n a_{mni} K^{m-1} u_i,$$

from which it follows that the approximation to y of the form (7) that minimizes $\|y - y_{mn}\|$ is

$$(9) \quad y_{mn}^* = \sum_{j=0}^{m-2} K^j f + P_{mn} K^{m-1} y.$$

THEOREM 2. *If $K^* x = 0$ has no nontrivial solution in H , then*

$$\|y - y_{m+1,n}^*\| \leq \|y - (f + Ky_{mn}^*)\| \leq \alpha_{mn} \|y - y_{mn}^*\|, \quad m \geq 1,$$

where

$$\alpha_{mn} = \|K - KP_{mn}\| \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

Proof. It follows from Eq. (9) that

$$(10) \quad f + Ky_{mn}^* = \sum_{j=0}^{m-1} K^j f + KP_{mn} K^{m-1} y.$$

Since this is of the form (7) with m replaced by $m + 1$, and since $y_{m+1,n}^*$ is the best possible approximation to y of that form, the first inequality of the theorem is obvious. Furthermore, it follows from (10), and from (8) with m replaced by $m + 1$, that

$$(11) \quad \begin{aligned} y - (f + Ky_{mn}^*) &= K^m y - KP_{mn} K^{m-1} y = (K - KP_{mn})(K^{m-1} y - P_{mn} K^{m-1} y) \\ &= (K - KP_{mn})(y - y_{mn}^*), \end{aligned}$$

hence

$$\|y - (f + Ky_{mn}^*)\| \leq \|K - KP_{mn}\| \|y - y_{mn}^*\|.$$

It only remains to prove that $\|K - KP_{mn}\| \rightarrow 0$ as $n \rightarrow \infty$. Since the null space of K^* is trivial, it follows that the null space of $(K^m)^* = (K^*)^m$ is trivial; and hence that the set $\{K^m u_1, K^m u_2, \dots\}$ is complete in H (since otherwise there exists a nonzero element $x \in H$ satisfying $0 = (x, K^m u_i) = (K^{m*} x, u_i)$ for each $i \geq 1$, which because $\{u_i\}$ is complete yields the contradiction $K^{m*} x = 0$). Thus $P_{mn} x \rightarrow x$ as $n \rightarrow \infty$ for each $x \in H$. The required result $\|K - KP_{mn}\| \rightarrow 0$ then follows exactly as for $\|K - KP_n\|$ in the previous section. Q.E.D.

Again the key step in the proof is a factorization, this time carried out in (11) and based on the property $P_{mn}^2 = P_{mn}$.

4. Galerkin and Galerkin-Petrov methods. The simple Galerkin (or Bubnov-Galerkin) method [3, p. 223] makes use of an approximate solution of the form (2), with the coefficients being fixed by requiring the residual $y_{1n} - f - Ky_{1n}$ to be orthogonal to U_n (the span of u_1, \dots, u_n), or equivalently, by requiring that

$$(12) \quad P_n(y_{1n} - f - Ky_{1n}) = 0,$$

where P_n is the orthogonal projection operator onto U_n . Since $P_n y_{1n} = y_{1n}$, it follows that y_{1n} satisfies an equation of the second kind,

$$(13) \quad y_{1n} = P_n f + P_n K y_{1n}.$$

The more general Galerkin-Petrov method [3, p. 223] again makes use of an approximate solution of the form (2), but it uses a second set of linearly independent functions $\{v_i\}$ to determine the coefficients: Eq. (12) is replaced by

$$(14) \quad Q_n(y_{1n} - f - Ky_{1n}) = 0,$$

where Q_n is the orthogonal projection operator onto V_n , the span of v_1, \dots, v_n . Obviously, the Galerkin-Petrov method includes the ordinary Galerkin method as a special case.

Some restriction is necessary on the choice of $\{v_i\}$ in the Galerkin-Petrov method, to avoid the possibility of $Q_n y_{1n}$ vanishing, or becoming very small. A suitable restriction [3, p. 224], assumed in the present work, is that the aperture $\theta(U_n, V_n)$ between U_n and V_n , which is defined by [3, p. 205]

$$\theta(U_n, V_n) = \|P_n - Q_n\|,$$

should be strictly less than 1 in the limit,

$$(15) \quad \overline{\lim}_{n \rightarrow \infty} \theta(U_n, V_n) < 1.$$

It follows easily from this condition and from the completeness of the set $\{u_i\}$ that the set $\{v_i\}$ is necessarily complete.

The following theorem asserts that the approximate solution by the Galerkin-Petrov method is necessarily improved by an iteration for n sufficiently large, provided that 1 is not an eigenvalue of K .

THEOREM 3. *If the homogeneous equation $x = Kx$ has only the trivial solution, and if y_{1n} is the approximate solution of (1) by the Galerkin-Petrov method and if y_{2n} is defined by $y_{2n} = f + Ky_{1n}$, then*

$$\|y - y_{2n}\| \leq \beta_n \|y - y_{1n}^*\| \leq \beta_n \|y - y_{1n}\|,$$

where $\beta_n \rightarrow 0$ as $n \rightarrow \infty$.

An explicit expression for β_n is given by Eq. (26).

Proof. Since K is compact and does not have 1 as an eigenvalue, the solution of Eq. (1) is unique and exists for any $f \in H$, and can be written

$$(16) \quad y = (I - K)^{-1}f,$$

where I is the identity operator, and $(I - K)^{-1}$ is a bounded linear operator in H .

The first step in the proof is to obtain from (14) an equation of the second kind for y_{1n} , similar to the Galerkin equation (13). From the condition (15), it follows [3, p. 206] that for n sufficiently large, say $n \geq n_0$, Q_n maps U_n one-one onto V_n .

Thus, if \bar{Q}_n denotes the operator Q_n with its domain restricted to U_n , the inverse \bar{Q}_n^{-1} exists for $n \geq n_0$. Furthermore, the inverses are uniformly bounded,

$$(17) \quad \|\bar{Q}_n^{-1}\| \leq M, \quad n \geq n_0.$$

By applying \bar{Q}_n^{-1} to Eq. (14), and using $Q_n y_{1n} = \bar{Q}_n y_{1n}$, we obtain the desired equation

$$(18) \quad y_{1n} = \Pi_n f + \Pi_n K y_{1n}, \quad n \geq n_0,$$

where

$$(19) \quad \Pi_n = \bar{Q}_n^{-1} Q_n.$$

In particular, if $V_n = U_n$, then $\Pi_n = Q_n = P_n$, and (18) reduces to the Galerkin equation (13).

In the general case, it follows from (19) that the range of Π_n is U_n , and that $\Pi_n^2 = \Pi_n$, so that Π_n is a projection operator (in general not orthogonal) onto U_n . From (17), $\|\Pi_n\|$ is uniformly bounded,

$$(20) \quad \|\Pi_n\| \leq M, \quad n \geq n_0.$$

It also follows from (19) that $\Pi_n P_n = P_n$ and that $Q_n \Pi_n = Q_n$; hence $I - \Pi_n$ can be written in either of the forms

$$(21) \quad I - \Pi_n = (I - \Pi_n)(I - P_n) = (I - Q_n)(I - \Pi_n).$$

Since $y_{2n} = f + K y_{1n}$, where y_{1n} satisfies (18), it follows immediately that $y_{1n} = \Pi_n y_{2n}$, hence y_{2n} satisfies

$$(22) \quad y_{2n} = f + K \Pi_n y_{2n},$$

an equation of the second kind with compact kernel $K \Pi_n$.

The operator $K \Pi_n$ converges in norm to K , because by virtue of (21) and (20),

$$\|K - K \Pi_n\| = \|(K - K Q_n)(I - \Pi_n)\| \leq \|K - K Q_n\| (1 + M),$$

in which $\|K - K Q_n\| \rightarrow 0$ by the same argument as used for $\|K - K P_n\|$ in Section 2. It follows that $(I - K \Pi_n)^{-1}$ exists as a bounded linear operator for n sufficiently large, and that it converges in norm to the bounded operator $(I - K)^{-1}$. The solution of (22) can therefore be written, for n sufficiently large, as

$$(23) \quad y_{2n} = (I - K \Pi_n)^{-1} f.$$

It follows from (16) and (23) that

$$\begin{aligned} y - y_{2n} &= [(I - K)^{-1} - (I - K \Pi_n)^{-1}] f = (I - K \Pi_n)^{-1} (K - K \Pi_n) (I - K)^{-1} f \\ &= (I - K \Pi_n)^{-1} (K - K \Pi_n) y. \end{aligned}$$

On using $\Pi_n P_n = P_n$, the last factor can be written as

$$(24) \quad (K - K\Pi_n)y = (K - K\Pi_n)(y - P_n y);$$

hence we obtain

$$(25) \quad \|y - y_{2n}\| \leq \|(I - K\Pi_n)^{-1}\| \|K - K\Pi_n\| \|y - P_n y\| = \beta_n \|y - y_{1n}^*\|,$$

where

$$(26) \quad \beta_n = \|(I - K\Pi_n)^{-1}\| \|K - K\Pi_n\|.$$

The limit $\beta_n \rightarrow 0$ follows from $\|(I - K\Pi_n)^{-1}\| \rightarrow \|(I - K)^{-1}\|$ and the result $\|K - K\Pi_n\| \rightarrow 0$ proved above.

The remaining inequality $\|y - y_{1n}^*\| \leq \|y - y_{1n}\|$ follows from the optimal property of y_{1n}^* . Q.E.D.

The key to the rapid convergence established for the Galerkin-Petrov method is the factorization carried out in (24).

The bound on $\|y - y_{2n}\|$ given by (25) may be very large if K has an eigenvalue close to 1, so that it may occasionally be useful to note that even in the worst case $\|y - y_{2n}\|$ is also bounded by

$$\|y - y_{2n}\| = \|(f + Ky) - (f + Ky_{1n})\| \leq \|K\| \|y - y_{1n}\|.$$

However, it has been found in practical calculations [7] that the iterated Galerkin approximation retains a marked superiority over the Galerkin approximation itself, even in cases where there is an eigenvalue of K very near to 1.

5. Acknowledgements. I am indebted to Professor A. N. Mitra and the Department of Physics and Astrophysics at Delhi University for their hospitality at the time this research was carried out, and to the Leverhulme Foundation for the award of a fellowship.

Department of Applied Mathematics
University of New South Wales
Kensington, N.S.W. 2033, Australia

1. N. I. AHIEZER & I. M. GLAZMAN, *Theory of Linear Operators in Hilbert Space*, vol. 1, GITTL, Moscow, 1950; English transl., Ungar, New York, 1961, Chapter 1. MR 13, 358; 41 #9015a.

2. C. T. H. BAKER, "Expansion methods," *Numerical Solution of Integral Equations*, L. M. Delves and J. Walsh (Editors), Clarendon Press, Oxford, 1974, Chapter 7.

3. M. A. KRASNOSEL'SKIĬ, G. M. VAĬNIKKO, P. P. ZABREĬKO, Ja. B. RUTICKIĬ & V. Ja. STECENKO, *Approximate Solution of Operator Equations*, "Nauka", Moscow, 1969; English transl., Wolters-Noordhoff, Groningen, 1972. MR 41 #4271.

4. S. G. MIHLIN & H. L. SMOLICKIĬ, *Approximate Methods for Solution of Differential and Integral Equations*, "Nauka", Moscow, 1965; English transl. by Scripta Technica, Modern Analytic and Computational Methods in Science and Mathematics, vol. 5, American Elsevier, New York, 1967, Chapter 4. MR 33 #855; 36 #1108.

5. F. RIESZ & B. SZ.-NAGY, *Functional Analysis*, Akad. Kiadó, Budapest, 1952, 2nd ed., 1953; English transl. of 2nd ed., Ungar, New York, 1955. MR 17, 175.

6. I. H. SLOAN, "Error analysis for a class of degenerate-kernel methods," *Numer. Math.*, v. 25, 1976, pp. 231-238.

7. I. H. SLOAN, B. J. BURN & N. DATYNER, "A new approach to the numerical solution of integral equations," *J. Computational Phys.*, v. 18, 1975, pp. 92-105.