CONVERGENCE OF DEGENERATE-KERNEL METHODS

IAN H. SLOAN

Abstract

Additional convergence results are given for the approximate solution in the space $L^2(a, b)$ of Fredholm integral equations of the second kind, $y = f + Ky$, by the degenerate-kernel methods of Sloan, Burn and Datyner. Convergence to the exact solution is proved for a class of these methods (including 'method 2'), under suitable conditions on the kernel K, and error bounds are obtained. In every case the convergence is faster than that of the best approximate solution of the form $y_n = \Sigma^n a_{ni} u_i$, where $u_1, \cdots, u_n$ are the appropriate functions used in the rank-n degenerate-kernel approximation. In addition, the error for method 2 is shown to be relatively unaffected if the integral equation has an eigenvalue near 1.

1. Introduction

This paper obtains convergence results for a recently proposed family of degenerate-kernel methods [8], for the approximate solution in the Hilbert space $L^2(a, b)$ of the Fredholm integral equation of the second kind,

$$y(t) = f(t) + \int_a^b K(t, s) y(s) ds, \qquad a \le t \le b, \tag{1.1}$$

or

$$y = f + Ky, \tag{1.2}$$

where y and f are real or complex functions in $L^2(a, b)$, K is a compact kernel, and 1 is a regular value of the equation, so that the solution y is unique.

The methods of [8] are described below. In a previous paper [7], it has been shown that the simplest of these methods, 'method 1', always converges to the exact result, and furthermore that the rate of convergence is always faster than that of the Galerkin method, even though the computational requirements of the two methods are almost identical. But the arguments of

422

[7] do not apply to 'method 2', numerically the most accurate of the methods tested in [8]. In the present paper, convergence results and error bounds are now obtained for method 2 and other variants of the methods of [8], under suitable conditions on the kernel K. Moreover, the error bound for method 2 confirms a property first observed numerically in [8], that method 2 is relatively unaffected if the integral equation has an eigenvalue near 1.

In the general method of [8], the exact equation (1.2) is approximated by

$$y_n = f + K_n y_n, \tag{1.3}$$

where K_n is a degenerate kernel of rank n, defined by

$$K_n x = \sum_{i,j=1}^{n} K u_i D_{ij}(v_j, x), \qquad x \in L^2(a, b). \tag{1.4}$$

Here both $\{u_i\}$ and $\{v_i\}$ are linearly independent sets in $L^2(a, b)$ which have yet to be specified, and the inner product is defined by

$$(v, x) = \int_a^b \overline{v(s)} x(s) ds, \qquad v, x \in L^2(a, b).$$

Furthermore, D is an $n \times n$ matrix, whose inverse is given by

$$(D^{-1})_{ji} = (v_j, u_i), \qquad j, i = 1, \ldots, n.$$

The solution of (1.3) is then easily found to be

$$y_n = f + \sum_{i=1}^{n} b_{ni} K u_i, \tag{1.5}$$

where the coefficients b_{ni} satisfy

$$\sum_{i=1}^{n} [(v_j, u_i) - (v_j, K u_i)] b_{ni} = (v_j, f), \qquad j = 1, \ldots, n.$$

In the present work, we consider several possible choices for the set $\{v_i\}$, assuming that $\{u_i\}$ is a given linearly independent set which generates the whole space $L^2(a, b)$; i.e. the linear envelope of $\{u_i\}$ is assumed to be dense in $L^2(a, b)$. (The qualitative criterion given in [8] for choosing the set $\{u_i\}$ is that it should be a good basis set for approximating the exact solution y. The error estimates obtained below provide a firm foundation for that criterion.)

The following three particular choices of $\{v_i\}$, listed here in order of increasing complexity, are discussed in this paper:

choice 1: $v_i = u_i$,
choice 2: $v_i = K^* u_i$,
choice 3: $v_i = K^* K u_i$,

where K^* is the adjoint of K, an integral operator with kernel defined by $K^*(s, t) = \overline{K(t, s)}$. The first and second choices correspond to methods 1 and 2 of [8]. The first, included here for completeness, is equivalent in its effect to a single iteration of the Galerkin method [3, §16.1]. The second choice has the advantage mentioned above, that it is relatively unaffected by an eigenvalue of the equation near 1, and though it is harder to use than choice 1, can still be a practical method in some circumstances. [The essential problem in implementing method 2 is the evaluation of the inner products $(v_j, Ku_i) = (K^*u_j, Ku_i)$. If K is symmetric, and if the quantities $Ku_i(t_j)$ have already been evaluated at a set of quadrature points t_j, for use in say the Galerkin method, then with good organization it will often be possible for the required inner products to be evaluated with little extra cost.] The third choice will usually be too cumbersome for practical calculations, but it is of theoretical interest, because as $n \to \infty$ the corresponding error approaches the least possible error for an approximate solution of the form (1.5).

An attractive feature of the degenerate-kernel methods of [8] is their good rate of convergence: it turns out that the approximate solution y_n always converges to y faster than the best possible linear combination of the form

$$y'_n = \sum_{i=1}^{n} a_{ni} u_i, \tag{1.6}$$

provided only that $\|K - K_n\| \to 0$. The best approximation of the form (1.6), i.e. the particular y'_n that minimises $\|y - y'_n\|$, is (see [1])

$$y'_n = P_n y, \qquad \qquad \cdot \cdot \tag{1.7}$$

where P_n is the orthogonal projection operator onto the span U_n of $u_1, \cdots, u_n$; thus the least possible error for an approximation of the form (1.6) is $\|y - P_n y\|$. On the other hand, it is shown in [7] (see also section 3) that if $\|K - K_n\| \to 0$ then the error from the degenerate-kernel methods satisfies

$$\|y - y_n\| \leq \beta_n \|y - P_n y\|, \tag{1.8}$$

where, with I denoting the identity,

$$\beta_n = \|(I - K_n)^{-1}\| \, \|K - K_n\|$$

$$\to 0 \text{ as } n \to \infty.$$

Thus y_n converges faster than $P_n y$ to the exact solution y (unless both $P_n y$ and y_n become exact for large n).

The first task with any particular choice of $\{v_i\}$ is therefore to prove that $\|K - K_n\| \to 0$, so that (1.8) applies. In the present paper this is accomplished, under suitable conditions on the kernel K, for a rather general choice of the

$\{v_i\}$, which under suitable conditions includes each of the above three choices of $\{v_i\}$ as a special case. The principal results are stated in the following section, and proved in section 3.

Useful reviews of other methods for solving (1.1) are given by [2] and [5].

2. Principal results

CHOICE 1: $v_i = u_i$

This choice corresponds to method 1 of [8], which is equivalent to a single iteration of the Galerkin method. For this choice the convergence has been proved previously in [7], with no restriction (other than compactness) on K. The error bound (1.8) applies.

CHOICE 2: $v_i = K^*u_i$

This choice corresponds to method 2 of [8]. A sufficient condition for convergence in this case is that K be symmetric and positive. In that case $\|K - K_n\| \to 0$, and the error norm satisfies (1.8), and also the stronger bound

$$\|y - y_n\| \leqq (1 + \gamma_n)\|K - K_n\| \, \|y - P_n y\|, \tag{2.1}$$

where

$$\gamma_n = \|(I - K_n)^{-1}\| \, \|K_n - K_n P_n\|$$
$$\to 0 \text{ as } n \to \infty.$$

CHOICE 3: $v_i = K^*Ku_i$

In this case it is sufficient that neither $Kx = 0$ nor $K^*x = 0$ has a non-trivial solution $x \in L^2(a, b)$. Then $K_n = Q_n K$, where Q_n is the orthogonal projection operator onto the span of $Ku_1, \cdots, Ku_n$, and $\|K - K_n\| \to 0$. The error norm satisfies (1.8), and also the stronger bounds

$$\|y - y_n\| \leqq (1 + \delta_n)\|Ky - Q_n Ky\| \tag{2.2}$$

$$\leqq (1 + \delta_n)\|K - K_n\| \, \|y - P_n y\|, \tag{2.3}$$

where

$$\delta_n = \|(I - K_n)^{-1}\| \, \|K_n - K_n Q_n\|$$
$$\to 0 \text{ as } n \to \infty.$$

It may be noted that for sufficiently large n, the error bound (2.2) approximates with arbitrary accuracy $\|Ky - Q_n Ky\|$, which is the least possible value of $\|y - y_n\|$ if y_n is of the form (1.5).

GENERAL CHOICE: $v_i = A^*SAu_i$

Here A is a bounded linear operator, with the property that neither $Ax = 0$ nor $A^*x = 0$ has a non-trivial solution in $L^2(a, b)$; S is a regular linear operator [3, §16.6]; and K is assumed to be expressible as

$$K = BA,$$

where B is a bounded linear operator, and at least one of A and B is compact.

3. Proofs

We first prove the results for choice 2, that is $v_i = K^*u_i$, for the case in which K is a symmetric, positive operator.

Since K is positive and compact, there exists a positive, compact operator $A = A^*$, such that

$$K = AA. \tag{3.1}$$

Hence equation (1.4) defining K_n can be written as

$$
\begin{aligned}
K_n x &= \sum_{i,j=1}^{n} AAu_i D_{ij}(AAu_j, x) \\
&= A \sum Au_i D_{ij}(Au_j, Ax) \\
&= A\Gamma_n Ax, \qquad x \in L^2(a, b),
\end{aligned}
\tag{3.2}
$$

where Γ_n is the operator defined by

$$\Gamma_n x = \sum w_i D_{ij}(w_j, x), \qquad x \in L^2(a, b), \tag{3.3}$$

with

$$w_i = Au_i, \qquad i = 1, \cdots, n.$$

Furthermore, the $n \times n$ matrix D is defined by

$$(D^{-1})_{ji} = (AAu_j, u_i) = (Au_j, Au_i) = (w_j, w_i). \tag{3.4}$$

The functions $w_1, \cdots, w_n$ are necessarily linearly independent, because there is no non-trivial solution of $Ax = 0$, since A is positive. Hence the matrix D^{-1} in (3.4), being a Gram matrix [4, page 59], is non-singular. The operator Γ_n given by (3.3) is therefore well-defined, and is easily seen to be simply the orthogonal projection onto the n-dimensional subspace AU_n spanned by $w_1, \cdots, w_n$. [To prove this, it is only necessary to express Γ_n in terms of an orthonormal basis $w_1', \cdots, w_n'$ of AU_n, in which case (3.3) becomes $\Gamma_n x = \sum_{i=1}^{n}(w_i', x)w_i'$.]

Moreover, the subspaces $\{AU_n\}$ form an ultimately dense sequence [3, page 201] in $L^2(a, b)$. [For if the contrary is supposed, then there exists a non-zero $x \in L^2$ such that $0 = (x, w_i) = (x, Au_i) = (Ax, u_i)$ for all $i \geqq 1$, which implies the contradiction $Ax = 0$.] Since Γ_n is an orthogonal projection onto a member of an ultimately dense sequence, and since A is compact, it follows [6, page 204] that $\|A - \Gamma_n A\| \to 0$. Therefore

$$\|K - K_n\| = \|AA - A\Gamma_n A\| \leqq \|A\| \|A - \Gamma_n A\| \to 0$$

as $n \to \infty$. This completes the first part of the proof for the case of choice 2.

Since K_n converges in norm to K, it follows by a standard argument (see for example [2, page 15, Theorem 5]) that for n sufficiently large, a bounded inverse $(I - K_n)^{-1}$ exists, which converges in norm to $(I - K)^{-1}$. Consequently, the solution of (1.3) can be written explicitly as

$$y_n = (I - K_n)^{-1}f, \qquad \cdot$$

if n is sufficiently large. Together with the analogous expression for y, namely

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

this yields

$$y - y_n = (I - K_n)^{-1}(K - K_n)(I - K)^{-1}f$$

$$= (I - K_n)^{-1}(K - K_n)y, \tag{3.5}$$

which leads to an obvious error bound,

$$\|y - y_n\| \leqq \|(I - K_n)^{-1}\| \|K - K_n\| \|y\|,$$

thus proving the convergence of y_n to y.

The much smaller error bound (1.8) is obtained, as in [7], by exploiting the relation

$$K_n u_i = Ku_i, \qquad i = 1, \cdots, n,$$

which is easily seen from the definition (1.4) to be valid for any choice of the $\{v_i\}$ for which D^{-1} is non-singular. An equivalent relation is

$$K_n P_n = KP_n, \tag{3.6}$$

where P_n is the orthogonal projection operator onto the span U_n of $u_1, \cdots, u_n$. It follows that (3.5) can be rewritten as

$$y - y_n = (I - K_n)^{-1}(K - K_n)(y - P_n y), \tag{3.7}$$

thus the error norm also satisfies

$$\|y - y_n\| \leqq \|(I - K_n)^{-1}\| \|K - K_n\| \|y - P_n y\|, \tag{3.8}$$

which is equivalent to (1.8).

The argument in the previous paragraph is not restricted to a particular choice of $\{v_i\}$. A stronger error bound can be obtained for the case $v_j = K^*u_j$ by taking into account a special property of K_n, namely

$$P_nK_n = P_nK, \tag{3.9}$$

which holds in addition to (3.6), and which follows easily from (1.4) in this case. Thus the operator $K - K_n$ can be written as

$$K - K_n = (I - P_n)(K - K_n).$$

To profit from this relation, we use it in (3.7), together with the identity

$$(I - K_n)^{-1} = I + (I - K_n)^{-1}K_n, \tag{3.10}$$

to obtain

$$y - y_n = [I + (I - K_n)^{-1}(K_n - K_nP_n)](K - K_n)(y - P_ny).$$

This leads immediately to the error bound

$$\|y - y_n\| \leqq (1 + \gamma_n)\|K - K_n\| \|y - P_ny\|, \tag{3.11}$$

where

$$\gamma_n = \|(I - K_n)^{-1}\| \|K_n - K_nP_n\|. \tag{3.12}$$

The first factor in (3.12) has a finite limit, since

$$\|(I - K_n)^{-1}\| \to \|(I - K)^{-1}\| \quad \text{as } n \to \infty,$$

and the second factor approaches zero, since

$$\|K_n - K_nP_n\| = \|K_n - KP_n\|$$
$$\leqq \|K_n - K\| + \|K - KP_n\|,$$

in which each term approaches zero. Thus $\gamma_n \to 0$ as $n \to \infty$, and the results stated in the previous section for choice 2 are now proved.

The results stated for choice 3 can be proved by a similar argument. Indeed, all but the last part of the above proof can be generalized to the case introduced at the end of the previous section, in which $K = BA$, where A and B are bounded operators, at least one of which is compact, and $v_i = A^*SAu_i$, where S is a regular operator. In this case K_n can be written as $K_n = B\Gamma_nA$, where Γ_n is a projection (uniformly bounded, but not necessarily orthogonal) onto AU_n, and the limit $\|K - K_n\| \to 0$ follows from either $\|A - \Gamma_nA\| \to 0$ or $\|B - B\Gamma_n\| \to 0$, depending on whether A or B is compact; we omit the details. The three special choices 1, 2 and 3 can be obtained by taking respectively $B = K$, $A = I$ for choice 1, $B = A = K^{1/2}$ for choice 2, and $B = I$, $A = K$ for choice 3, together with $S = I$ in each case.

4. Numerical example

The various choices of $\{v_i\}$ are tested in this section on the integral equation

$$y(t) = t^2 + \lambda \int_0^1 K(t, s)y(s)ds, \qquad 0 \leqq t \leqq 1,$$

where

$$K(t, s) = \begin{cases} t(2 - s), & t \leqq s, \\ s(2 - t), & s \leqq t. \end{cases}$$

The exact solution of this equation is known analytically from the equivalent boundary-value problem

$$y'' + 2\lambda y = 2, \qquad y(0) = 0, \qquad y(1) + y'(1) = 3.$$

The set $\{u_i\}$ is chosen (as in [7]) to be the complete orthogonal set

$$u_i(t) = P_{i-1}(2t - 1), \qquad i \geqq 1,$$

where P_j is the Legendre polynomial of degree j.

Results are given here for two values of λ, the first of which is $\lambda = -\lambda_1$, where λ_1 is the smallest characteristic value of the kernel, $\lambda_1 = 2.057929\cdots$. This choice is considered a fair one, as in [7], in that the kernel is neither very large nor very small — in fact $\|K\| = 1$. (As is evident from the numerical results of [8], smaller values of $|\lambda|$ are more favourable to the degenerate-kernel methods, and larger values of $|\lambda|$ less favourable.)

TABLE 1

Error norms $\|y - y_n\|$, with $\lambda = -\lambda_1 = -2.057929\cdots$

n	$\|y - P_n y\|$	choice 1 ('method 1')	choice 2 ('method 2')	choice 3	least possible
1	2.2–1	4.9–2	3.59–2	3.573–2	3.573–2
2	9 0–2	1.1–2	5.22–3	5.082–3	5.082–3
3	8.6–3	5.0–4	2.64–4	2.532–4	2.532–4
4	1.9–3	6.6–5	3 39–5	3.141–5	3.141–5
5	1.1–4	2.4–6	1.32–6	1.202–6	1.202–6
6	1.6–5	2.6–7	1.38–7	1.224–7	1.224–7
7	6.6–7	7.7–9	4.24–9	3.732–9	3.732–9
8	7.5–8	6.7–10	3.68–10	3.168–10	3.168–10

Table 1 gives error norms for $\lambda = -\lambda_1$, for each of the three choices of $\{v_i\}$ considered in this paper, and for values of n from 1 to 8. Also given are the values of $\| y - P_n y \|$, which is the least possible error norm for an approximation of the form (1.6). The choice 1 error norms in Table 1, as known already from [7], are much smaller than $\| y - P_n y \|$, and approach zero much faster as $n \to \infty$, in conformity with condition (1.8).

The present interest in Table 1 is mainly in choices 2 and 3. The first point of note is that their error norms are smaller than those for choice 1, by roughly a factor of two.

In the last column of Table 1 are given the least possible values of $\| y - y_n \|$ for y_n of the form (1.5), namely $\| Ky - Q_n Ky \|$, where Q_n is the orthogonal projection onto the span of $Ku_1, \cdots, Ku_n$. It is evident that the error norms for choice 3 are very close indeed to these optimal error norms, even for small values of n: indeed, the last two columns agree to four significant figures. The close agreement for larger values of n is of course expected, in that the choice 3 error bound given by (2.2) approximates the optimal error norm with arbitrary accuracy, if n is sufficiently large. The error norms for the more practical method 2 are seen to be only about 10% larger than the optimal norms.

TABLE 2

Error norms $\| y - y_n \|$, with $\lambda = \lambda_1 + 0.001 = 2.058929 \cdots$

n	$\| y - P_n y \|$	choice 1 ('method 1')	choice 2 ('method 2')	choice 3	least possible
1	$3.2 + 2$	$8.2 + 2$	$8.5\ +2$	$1.2\ \ +3$	$3.64\ \ +1$
2	$1.3 + 2$	$8.4 + 2$	$1.3\ \ +3$	$1.6\ \ +2$	7.50
3	$1.4 + 1$	$1.2 + 3$	$1.1\ \ +1$	$4.64\ -1$	$3.53\ -1$
4	2.9	$2.1 + 1$	$3.23 - 1$	$4.869 - 2$	$4.835 - 2$
5	$1.9 - 1$	$8.6 - 2$	$2.26 - 3$	$1.873 - 3$	$1.873 - 3$
6	$2.5 - 2$	$1.6 - 3$	$2.18 - 4$	$1.922 - 4$	$1.922 - 4$
7	$1.2 - 3$	$1.4 - 5$	$7.15 - 6$	$6.138 - 6$	$6.138 - 6$
8	$1.2 - 4$	$1.1 - 6$	$5.87 - 7$	$5.033 - 7$	$5.033 - 7$

Table 2 contains the error norms for a second value of λ, this one chosen to be very close to a characteristic value; specifically, $\lambda = \lambda_1 + 10^{-3} = 2.058929 \cdots \cdots$. The equation in this case corresponds physically to a forced vibration at a frequency very close to a natural frequency of the system, thus the amplitude is large — in fact $\| y \| \simeq 821$. Such a situation is unfavourable for any numerical method. Nevertheless, for choices 2 and 3 the error bounds

(2.1) and (2.3) indicate that the relative errors should still be similar to those in Table 1, if n is sufficiently large.

It is evident from Table 2 that when n is small all three of the methods show the large errors to be expected from the near instability of the system. However, in conformity with (2.2), the choice 3 error norms recover very rapidly as n increases, and agree qualitatively with the optimal error norms for $n \geqq 3$, and to four significant figures for $n \geqq 5$. The similar recovery of the choice 2 error norms, anticipated from (2.1), is slightly less rapid, but the pattern observed in Table 1 is fully established for $n \geqq 6$.

An interesting observation from Table 2 is that even the choice 1 error norms recover in a similar (though slower) manner as n increases, so that for large n the observed error norms are far smaller than the theoretical bounds given by (1.8). However, it is believed that the good behaviour of the error norms for choice 1 in this example is a consequence of the particular choice of the subspaces $\{U_n\}$, and therefore that it may not extend to all other choices of the subspaces, whereas the good behaviour for choice 2, if K is symmetric and positive, is guaranteed by (2.1).

Acknowledgements

References

[1] N. I. Akhiezer and I. M. Glazman, *Theory of linear operators in Hilbert space*, Volume 1 (Frederick Ungar, 1961), Chapter 1.

[2] K. E. Atkinson, *A survey of numerical methods for the solution of Fredholm integral equations of the second kind* (SIAM, 1976).

[3] M. A. Krasnosel'skii, G. M. Vainikko, P. P. Zabreiko, Ya. B. Rutitskii and V. Ya. Stetsenko, *Approximate solution of operator equations* (Wolters–Noordhoff, 1972).

[4] S. G. Mikhlin, *Variational methods in mathematical physics* (Pergamon, 1964).

[5] S. G. Mikhlin and K. L. Smolitskiy, *Approximate methods for solution of differential and integral equations* (American Elsevier, 1967).

[6] F. Riesz and B. Sz.-Nagy, *Functional analysis* (Frederick Ungar, 1955).

[7] I. H. Sloan, 'Error analysis for a class of degenerate-kernel methods', *Numer. Math.*, 25 (1976), 231–238.

[8] I. H. Sloan, B. J. Burn and N. Datyner, 'A new approach to the numerical solution of integral equations', *J. Computational Physics* 18 (1975), 92–105.

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