Calculating the Best Approximate Solution of an Operator Equation

H. Wolkowicz, S. Zlobec

Mathematics of Computation, Volume 32, Issue 144 (Oct., 1978), 1183-1213.

MATHEMATICS OF COMPUTATION, VOLUME 32, NUMBER 144
OCTOBER 1978, PAGES 1183–1213

Calculating the Best Approximate Solution of an Operator Equation*

By H. Wolkowicz** and S. Zlobec***

Abstract. This paper furnishes two classes of methods for calculating the best approximate solution of an operator equation in Banach spaces, where the operator is bounded, linear and has closed range. The best approximate solution can be calculated by an iterative method in Banach spaces stated in terms of an operator parameter. Specifying the parameter yields some new and some old iterative techniques. Another approach is to extend the classical approximation theory of Kantorovich for equations with invertible operators to the singular case. The best approximate solution is now obtained as the limit of the best approximate solutions of simpler equations, usually systems of linear algebraic equations. In particular, a Galerkin-type method is formulated and its convergence to the best approximate solution is established. The methods of this paper can also be used for calculating the best least squares solution in Hilbert spaces or the true solution in the case of an invertible operator.

1. Introduction. A solution of a consistent operator equation

$$Ax = b, \tag{1}$$

where A is a bounded linear operator from a Banach space X into itself and b is an element of X, can be calculated in two ways. One can use a simple iterative scheme set up in X, e.g. Krasnosel'skiĭ et al. [18, Chapter 1], or an extension to Banach spaces of various well-known matrix iterative schemes, as suggested by, e.g. Petryshyn [29], [30] and Kammerer and Nashed [14]. The other way is to approximate the original equation (1) by a sequence of equations

$$\bar{A}\bar{x} = \bar{b}, \tag{2}$$

which are possibly easier to handle, and use appropriate error analysis. The latter approach is generally more successful. One of the first theories which studies the relationship between (1) and (2) was given by Kantorovich [16] and elaborated in the book by Kantorovich and Akilov [17]. Kantorovich's theory has been developed only for consistent equations. In particular, it is concerned with the following problems:

Received October 28, 1976; revised February 28, 1978.
AMS (MOS) subject classifications (1970). Primary 65J05; Secondary 15A09, 41A65.
Key words and phrases. Best approximate solution, inconsistent equation, generalized inverse of an operator, iterative methods, Kantorovich's theory of approximation methods, Galerkin's method.
* Research partly supported by the National Research Council of Canada.

** The contribution of this author is part of his M. Sc. thesis in Applied Mathematics. He is presently at Department of Mathematics, Dalhousie University, Halifax, Nova Scotia, Canada B3H 4H8.
*** On Sabbatical leave at Department of Mathematics, University of Delaware, Newark, Delaware 19711.

(i) Find conditions under which the consistency of (1) implies the consistency of (2).

(ii) If both (1) and (2) are consistent, estimate the distance between their solutions.

(iii) Find conditions under which the solutions of a sequence of approximate equations (2) converge to the solution of the equation (1).

(iv) Estimate the norm of A in terms of the norm of $\bar{A}$ and vice versa.

Kantorovich's approximation theory is rather general and, therefore, it is in principle applicable in many consistent situations, including the study and numerical treatment of infinite systems of linear equations, integral equations, ordinary differential equations and boundary value problems.

Various approximation theories have been recently developed and applied to particular problems by different authors, many of whom use the Kantorovich theory as a starting point. For instance, Thomas [38] refines some of Kantorovich's ideas and applies them to develop an approximation theory for the Nyström method of solving integral equations. Phillips [31] and Prenter [32] formulate approximation theories for the collocation method, while Ikebe [13] works with the Galerkin method. Anselone [1] and Anselone and Moore [2] use the notion of collectively compact operators to formulate a different error analysis. Moore and Nashed [22] further developed the ideas of Anselone and Moore for possibly inconsistent operator equations in Banach spaces. They use the notions of generalized inveses of linear operators on Banach spaces and "best approximate" solutions of linear operator equations. Furthermore, they get, in special cases, some results in the perturbation theory of rectangular matrices obtained earlier by Ben-Israel [6] and Stewart [36].

An approximation theory for general, possibly inconsistent, linear equations in Hilbert spaces has been studied using the classical approach of Kantorovich (rather than the one of Moore and Nashed) by Zlobec [41]. One of the objectives of this paper is to continue the latter approach and formulate Kantorovich's theory for general, possibly inconsistent, linear equations in Banach spaces. The basic idea is here to establish and explore a relationship between best approximate solutions of (1) and (2) and then use this relationship as a source for formulating various specific schemes for calculating the best approximate solution of (1).

In the iterative computation of the best approximate solutions, as well as in Kantorovich's theory for singular equations, we will often use the concept of the generalized inverse of an operator. Some basic results on generalized inverses in Banach spaces are summarized in Section 2. In Section 3 an iterative scheme is set up in Banach spaces for calculating both the best approximate solution and the generalized inverse. This section extends from Hilbert to Banach spaces some results from the book by Ben-Israel and Greville [9, Chapter 8]. In Section 4, conditions for the consistency of $Ax = y$, for every y in a given subspace, are stated in terms of an approximate equation. Various error estimates are obtained as special cases. Kantorovich's theory for general linear equations is formulated in Section 5. The results are formulated in such a way that a comparison with the corresponding results for the nonsingular case from

[17] is easily made. The most important results in this section are Theorem 4, which gives an error estimate, and Theorem 5, which gives conditions for convergence of approximate schemes. Using Kantorovich's theory, in Section 6, a Galerkin-type method for calculating the best approximate solution is stated, and its convergence is established for a class of operator equations in Banach spaces.

Situations where inconsistent linear operator equations arise are numerous and they include: integral equations in the theory of elasticity, potential theory and hydromechanics, e.g. Muskhelishvili [24], the integral formulation of the interior Neumann problem for the Laplacian, e.g. Kammerer and Nashed [14] and Atkinson [4], the eigenvalue problem in the case of a nonhomogeneous integral equation when the associated homogeneous equation has a nontrivial solution, e.g. Kammerer and Nashed [14], and boundary value problems, e.g. Langford [19]. They also appear in the numerical solution of differential equations, for instance in the collocation method when the number of collocation points is bigger than the number of coefficients to be determined, e.g. Krasnosel'skiĭ et al. [18]. If the number of collocation points is smaller than the number of coefficients, then, if consistent, the approximate equation (2) has infinitely many solutions; and one may again be interested in calculating the best approximate solution. Under- and over-determined initial value problems have been studied by Lovass-Nagy and Powers [20]. In the finite dimensional case the under- and over-determined systems appear frequently in statistics, e.g. Rao and Mitra [33]; see also Ben-Israel and Greville [9].

2. **Best Approximate Solutions and Generalized Inverses.** In order to formulate an iterative method for calculating the best approximate solution and develop Kantorovich's theory for general, possibly inconsistent, operator equations in Banach spaces, we employ the following notation and notions in the sequel:

X, Y, $\overline{X}$, $\overline{Y}$ denote real or complex *Banach spaces*,

$l(X, Y)$ the *set of all linear operators from X into Y*,

$l_b(X, Y)$ the *set of all bounded linear operators from X into Y*,

$l(X)$ and $l_b(X)$ the sets $l(X, X)$ and $l_b(X, X)$, respectively.

If $A \in l(X, Y)$, then $\|A\|$ denotes the operator norm of A, i.e. $\|A\| = \sup_{\|x\|=1}\|Ax\|$. Further, if $S \subset X$, then $A|_S$ means *A restricted to S.*

$\sigma(A)$ is the *spectrum* of A and

$\rho(A) = \sup_{\lambda\in\sigma(A)}|\lambda|$ the *spectral radius* of A.

For every $A \in l(X, Y)$, $R(A) = \{y \in Y: y = Ax$ for some $x \in X\}$ is the *range space of A* and $N(A) = \{x \in X: Ax = 0\}$ is the *null space of A*. If X and Y are Hilbert spaces, A^* is the *adjoint of A*. For the above notions and their properties, see e.g. Taylor [37]. For any two operators $A \in l(Y)$ and $B \in l(X)$, we denote $R\{A, B\} = \{Z \in l_b(X, Y): Z = AUB$, where $U \in l(X, Y)\}$, e.g. Ben-Israel [5]. A linear operator $P \in l(X)$, is called a *projection* (of X) if $P^2 = P$. If $R(P) = M$, then we denote P by P_M and call it the *projection of X onto M.* Every projection P_M decomposes X into two *algebraic complements* $M = R(P_M)$ and $N = R(I - P_M)$. This implies $X = M + N$, and we write $N = M^c$. If M and N are both closed, then we say that M has a *topological complement* in X and write

$$X = M \oplus N. \tag{3}$$

For an example of decomposition (3), the reader is referred to Nashed's paper [25, p. 327]. Recall that a closed subspace M of X has a topological complement if, and only if, there exists a continuous projection P_M (of X), e.g. Taylor [37, p. 241]. However, not every closed subspace has a topological complement, as shown by Murray [23] in 1937.

Consider $A \in l_b(X, Y)$. We shall assume that there exist continuous projections, $P_{N(A)} \in l_b(X)$ and $P_{R(A)} \in l_b(Y)$. (In particular, such an A must have a closed range.) $P_{N(A)}$ determines the complement $N(A)^c = (I - P_{N(A)})X$. Similarly, $P_{R(A)}$ determines the complement $R(A)^c = (I - P_{R(A)})Y$. Hence, $X = N(A) \oplus N(A)^c$ and $Y = R(A) \oplus R(A)^c$. When $A \in l_b(X, Y)$ and projections $P_{N(A)} \in l_b(X)$ and $P_{R(A)} \in l_b(Y)$ are given, then the system

$$A^+A = P_{N(A)^c}, \tag{4}$$

$$AA^+ = P_{R(A)}, \tag{5}$$

$$A^+P_{R(A)} = A^+ \tag{6}$$

always has a unique solution $A^+ \in l_b(Y, X)$, called the *generalized inverse* of A (relative to the projections $P_{N(A)}$ and $P_{R(A)}$). The operator A^+ then establishes a one-to-one correspondence between $R(A)$ and $N(A)^c$, i.e. $A^+|_{R(A)} = (A|_{N(A)^c})^{-1}$, e.g. Nashed [25], Kammerer and Plemmons [15]. For every given b in Y, the vector $x^* = A^+b$ is called the *best approximate solution of the equation* $Ax = b$ (relative to $P_{N(A)}$ and $P_{R(A)}$). The vector A^+b is then a unique solution in $N(A)^c$ of the projectional equation $Ax = P_{R(A)}b$.

Remark 1. The term "best approximate" solution is used by Newman and Odell [27] under different circumstances. There $\hat{x}$ is a "best approximate" solution of $Ax = b$, where $A: X \longrightarrow Y$, $b \in Y$ if, for every $x \in X$ with $x \neq \hat{x}$, either

$$\|A\hat{x} - b\| < \|Ax - b\|$$

or

$$\|A\hat{x} - b\| = \|Ax - b\| \quad \text{and} \quad \|\hat{x}\| < \|x\|.$$

(This corresponds to the notion of *best least squares solution* in the case of Hilbert spaces.) In order to avoid possible ambiguity, we shall refer to the above $\hat{x}$ as the "X, Y-best approximate" solution of the equation $Ax = b$. If the norms on X and Y are strictly convex, then an "X, Y-best approximate" solution exists. If they are not strictly convex, then an "X, Y-best approximate" solution may not exist. In order to find $\hat{x}$, we need the notion of an *X-projection* (also called a "metric projection" by Blather, Morris and Wulbert in [11]). Suppose that S is a subspace of X. Then the mapping E_S is the X-projection onto S if, for every $x \in X$, E_Sx solves the minimization problem $\min_{y \in S}\|x - y\|$. In general, the mapping E_S is not linear. An instance in which E_S is linear is when S and S^c have a basis and the norm in X is a "TK norm" with respect to these bases, e.g. Singer [34]. In Hilbert spaces, E_S corresponds to the orthogonal projection P_S. When the "X, Y-best approximate" solution $\hat{x}$ exists, then $\hat{x} = Bb$, where $B = (I - E_{N(A)})A^+E_{R(A)}$ and A^+ is any generalized inverse of A with respect to some $P_{N(A)^c}$ and $P_{R(A)}$. (Note that B need not be linear.) Thus, we see that when $E_{N(A)}$ and $E_{R(A)}$ are linear, one may choose $P_{N(A)^c} = I - E_{N(A)}$ and $P_{R(A)} = E_{R(A)}$ in which case the "X, Y-best approximate" solution $\hat{x} = Bb$ coincides with the "best approximate" solution $x^* = A^+b$. (See Example 3 below.)

Suppose that Y (but not necessarily X) is a Hilbert space, $A \in l_b(X, Y)$, A has closed range and $X = N(A) \oplus N(A)^c$. Then one may choose $P_{R(A)} = E_{R(A)}$, which is now the orthogonal projection on $R(A)$, i.e. $R(A)^c = R(A)^\perp$, and write $Y = R(A) \oplus R(A)^c$. If A^+ is the generalized inverse of A with respect to $P_{N(A)^c}$ and $P_{R(A)}$, the best approximate solution $x^* = A^+b$ is the unique *least squares solution* of $Ax = b$ in $N(A)^c$, i.e. x^* solves the problem

$$\min_{x \in X} \|Ax - b\|; \tag{7}$$

and among all solutions of (7), it is the only one in $N(A)^c$. If both X and Y are Hilbert spaces and $A \in l_b(X, Y)$ has closed range, we may choose $P_{R(A)} = E_{R(A)}$ and $P_{N(A)^c} = I - E_{N(A)}$. These are now the orthogonal projections, i.e. $N(A)^c = R(A^*)$, $R(A)^c = N(A^*)$ and $\hat{x} = Bb = A^+b = x^*$ is the *best least squares solution* of the equation $Ax = b$. This means that x^* is the only solution in $N(A)^c$ of the minimization problem (7), and among all solutions of (7) it is the unique one of smallest norm. For a detailed discussion of the generalized inverse and best least squares solution in Hilbert spaces, the reader is referred to the book by Ben-Israel and Greville [9].

3. An Iterative Method for Calculating the Best Approximate Solution.

In order to calculate the best approximate solution x^* of the operator equation

$$Ax = b, \tag{1}$$

where $A \in l_b(X, Y)$, $P_{N(A)^c} \in l_b(X)$ and $P_{R(A)} \in l_b(Y)$, one can use the following iterative scheme:

$$x_{k+1} = x_k - BAx_k + Bb, \qquad k = 0, 1, \ldots, \tag{8}$$

where

$$B \in R\{P_{N(A)^c}, P_{R(A)}\}.$$

This scheme has been suggested for calculating the best least squares solution in Hilbert spaces in [41]; see also [9, p. 356].

THEOREM 1. *Let $A \in l_b(X, Y)$, $b \in Y$, $P_{N(A)^c} \in l_b(X)$, $P_{R(A)} \in l_b(Y)$, and $B \in R\{P_{N(A)^c}, P_{R(A)}\}$ be given. Then the sequence $\{x_k\}$, generated by* (8), *converges to the best approximate solution x^* of $Ax = b$ for all $x_0 \in N(A)^c$ if $\rho(P_{N(A)^c} - BA) < 1$.*

Proof. We find that

$$\begin{aligned}
x_{k+1} - x^* &= (I - BA)x_k + Bb - x^*, \quad \text{by (8)} \\
&= (I - BA)x_k + BP_{R(A)}b - x^*, \quad \text{since } B \in R\{P_{N(A)^c}, P_{R(A)}\} \\
&= (I - BA)(x_k - x^*), \quad \text{since } P_{R(A)}b = Ax^* \\
&= (I - BA)^{k+1}(x_0 - x^*), \quad \text{by iteration} \\
&= (P_{N(A)^c} - BA)^{k+1}(x_0 - x^*), \quad \text{since } x_0 \text{ and } x^* \text{ are in } N(A)^c. \qquad (9)
\end{aligned}$$

But $\rho(P_{N(A)^c} - BA) = \lim_{n\to\infty} \sup \|(P_{N(A)^c} - BA)^n\|^{1/n} < 1$, by the property of ρ (e.g. Taylor [37]) and the assumption. Therefore, there exists a real number s and a positive integer n_0 such that

$$\|(P_{N(A)^c} - BA)^n\|^{1/n} \leqslant s < 1, \quad \text{for all } n \geqslant n_0.$$

Hence, $\|(P_{N(A)^c} - BA)^n\| \leqslant s^n \longrightarrow 0$ as $n \longrightarrow \infty$. This implies $(P_{N(A)^c} - BA)^n \longrightarrow 0$ as $n \longrightarrow \infty$. Thus, x_k converges to x^*, by (9). □

Remark 2. Necessary conditions for convergence of x_k to x^*, for every $x_0 \in N(A)^c$, are $\rho(P_{N(A)^c} - BA) \leqslant 1$ (not $\rho(P_{N(A)^c} - BA) < 1$!) and $P_{N(A)^c} - BA$ has no eigenvalue λ such that $|\lambda| = 1$.

Proof. When $x_k \longrightarrow x^*$ for any $x_0 \in N(A)^c$, then

$$(P_{N(A)^c} - BA)^k(x_0 - x^*) \longrightarrow 0 \quad \text{as } k \longrightarrow \infty,$$

by (9). Hence, for every $x \in X$,

$$\sup_{k\geqslant 1} \|(P_{N(A)^c} - BA)^k x\| < \infty.$$

Now, by the Banach-Steinhaus theorem, there exists $M > 0$ such that $\|(P_{N(A)^c} - BA)^k\| \leqslant M$, $k = 1, 2, \ldots$. But

$$\begin{aligned}
[\rho(P_{N(A)^c} - BA)]^k &= \rho[(P_{N(A)^c} - BA)^k], \quad \text{by the spectral mapping theorem} \\
&\leqslant \|(P_{N(A)^c} - BA)^k\| \leqslant M.
\end{aligned}$$

Therefore, $\rho(P_{N(A)^c} - BA) \leqslant 1$. In order to prove the second necessary condition, we observe that for $0 \neq x \in X$ and $|\lambda| = 1$, such that $(P_{N(A)^c} - BA)x = \lambda x$, $x \in N(A)^c$ and $(P_{N(A)^c} - BA)^k x = \lambda^k x \not\longrightarrow 0$ as $k \longrightarrow \infty$ contradicting $x_k \longrightarrow x^*$, by (9). □

The above remark is demonstrated by the operator $A \in l_b(l_1)$ defined on $x = (x_i)$ by $(Ax)_1 = 0$ and $(Ax)_i = (1 - 2^{-i})x_i$, $i = 2, 3, \ldots$. If $N(A)^c = R(A)$, $R(A)^c = N(A)$ and $B = 2A$, then

$$\rho(P_{N(A)^c} - BA) = \sup_{i\geqslant 2} \{|-1 + 2^{2-i} - 2^{1-2i}|\} = 1.$$

But for every $x \in N(A)^c$,

$$((I - BA)^k x)_i = \begin{cases} 0, & \text{if } i = 1, \\ (-1 + 2^{2-i} - 2^{1-2i})^k x_i, & \text{if } i = 2, 3, \ldots, \end{cases}$$

which implies convergence.

Remark 3. It is a consequence of Remark 2 that Theorem 1 gives not only a sufficient but also a necessary condition for convergence, if X is finite dimensional. In the case of a Hilbert space X and a normal operator $T = P_{N(A)^c} - BA$, one can show that $T^n x \longrightarrow 0$ if, and only if, $\rho(T) \leqslant 1$ and the spectrum of T has no mass on the unit circle $|z| = 1$. (This follows from the spectral theorem and the Lebesgue dominated convergence theorem.) Specifying $P_{N(A)^c} = P_{R(A^*)}$ and $B = \alpha A^*$ one now establishes the following characterization of convergence: The sequence $\{x_k\}$, $x_0 \in R(A^*)$, converges to the best least squares solution of $Ax = b$ if, and only if,

$$0 < \alpha \leqslant 2/\|A^*A\|$$

and $\|A^*A\|$ is not an eigenvalue of A^*A if $\alpha = 2/\|A^*A\|$. We recall that Petryshyn in [28] gives only the sufficient condition $0 < \alpha < 2/\|A^*A\|$.

Specifying B in (8), one obtains various iterative schemes for computing the best approximate solution. In particular, if one splits $A = M + N$, where $N(A) \subset N(M)$ and $R(A) = R(M)$, chooses M^+ with respect to the continuous projections $P_{R(M)}$ and $P_{N(M)^c}$ such that $P_{R(M)} = P_{R(A)}$ and $N(M)^c \subset N(A)^c$ and specifies $B = \omega M^+$, $\omega \neq 0$, then (8) becomes

$$x_{k+1} = [(1 - \omega)I - \omega M^+N]x_k + \omega M^+b, \qquad x_0 \in N(M)^c. \tag{10}$$

Further, for $\omega = 1$, (10) becomes

$$x_{k+1} = -M^+Nx_k + M^+b, \qquad x_0 \in N(M)^c. \tag{11}$$

If both A and M are invertible, then (10) and (11) become, respectively,

$$x_{k+1} = [(1 - \omega)I - \omega M^{-1}N]x_k + \omega M^{-1}b, \qquad x_0 \in X, \tag{12}$$

and

$$x_{k+1} = -M^{-1}Nx_k + M^{-1}b, \qquad x_0 \in X. \tag{13}$$

The scheme (12) has been studied by Petryshyn [29], who calls it the "Extrapolated Jacobi Method". The scheme (13) is the well-known Jacobi method. Other methods can be obtained by the splitting $A = D + S + Q$ with $B = (\omega^{-1}D + S)^+$, $\omega \neq 0$, where $\omega^{-1}D + S \in l_b(X, Y)$ and $P_{R(\omega^{-1}D+S)} = P_{R(A)}$, $N(\omega^{-1}D + S)^c \subset N(A)^c$. Then (8) becomes

$$\begin{aligned} x_{k+1} &= (D + \omega S)^+[(1 - \omega)D - \omega Q]x_k + \omega(D + \omega S)^+b, \\ &\qquad\qquad x_0 \in N(\omega^{-1}D + S)^c. \end{aligned} \tag{14}$$

If both A and $D + \omega S$ are invertible, the scheme (14) becomes

$$x_{k+1} = (D + \omega S)^{-1}[(1 - \omega)D - \omega Q]x_k + \omega(D + \omega S)^{-1}b, \qquad x_0 \in X, \tag{15}$$

which is known as the "Successive Over-Relaxation Method" (abbreviated SOR method). Specifying $\omega = 1$ in (15), one obtains

$$x_{k+1} = -(D + S)^{-1}Qx_k + (D + S)^{-1}b. \qquad x_0 \in X, \tag{16}$$

which is known as the Gauss-Seidel method. In the case of an $n \times n$ invertible matrix $A = (a_{ij})$, one frequently specifies

$$D = M = (a_{ii}), \quad i = 1, \dots, n,$$

$$S = (a_{ij}), \quad i > j, i = 1, \dots, n, j = 1, \dots, n-1, \text{ zero otherwise,}$$

$$Q = A - D - S,$$

$$N = A - M$$

in (12), (13), (15) and (16). Properties of these schemes for systems with invertible matrices (and linear operators) have been extensively studied, see, e.g. Varga [39], [40] and Petryshyn [28], [29]. Scheme (11) for systems with singular matrices has been studied by Berman and Plemmons [10].

Remark 4. If in the splitting $A = M + N$, where $N(A) \subset N(M)$ and $R(A) = R(M)$, one also requires that $N(A)^c \subset N(M)^c$, then the splitting $A = M + N$ with $N(A) = N(M)$, $N(A)^c = N(M)^c$ and $R(A) = R(M)$ is obtained. The latter, when applied to matrices, is called a *proper splitting*, by Berman and Plemmons. Note that this proper splitting, in the case of Hilbert spaces and the usual (orthogonal) complements, reduces to $A = M + N$, where $N(A) = N(M)$ and $R(A) = R(M)$. The proper splittings are not only useful in iterative calculation of the best least squares solution, but they also play an important role in the Kantorovich approximation theory (see Section 6).

One can slightly modify (8) in order to compute the generalized inverse A^+.

THEOREM 2. *Let $A \in l_b(X, Y)$, $P_{N(A)} \in l_b(X)$ and $P_{R(A)} \in l_b(Y)$. If $B \in R\{P_{N(A)^c}, P_{R(A)}\}$, then the sequence $\{X_k\}$, generated by*

$$X_{k+1} = X_k - BAX_k + B, \quad k = 0, 1, 2, \dots, \tag{17}$$

converges to A^+, the generalized inverse of A relative to $P_{N(A)}$ and $P_{R(A)}$, for all $X_0 \in R\{P_{N(A)^c}, P_{R(A)}\}$ if $\rho(P_{N(A)^c} - BA) < 1$.

Proof. Here

$$\begin{aligned} X_{k+1} - A^+ &= X_k - BAX_k + B - A^+ \\ &= (P_{N(A)^c} - BA)X_k + BP_{R(A)} - P_{N(A)^c}A^+ \\ &= (P_{N(A)^c} - BA)(X_k - A^+) = (P_{N(A)^c} - BA)^{k+1}(X_0 - A^+). \end{aligned}$$

The rest of the proof is analogous to the proof of Theorem 1. (Note that the proof remains valid for $X_0 \in R\{P_{N(A)^c}, I\}$.) □

If (17) is modified as follows:

$$X_{k+1} = X_k - BX_k + B, \quad k = 0, 1, 2, \dots; \tag{18}$$

and if one chooses

$$X_0 \in R\{P_{R(A)}, P_{R(A)}\}, \quad B \in R\{P_{R(A)}, P_{R(A)}\}, \tag{19}$$

then the sequence generated has the property

$$X_{k+1} - P_{R(A)} = (P_{R(A)} - B)^{k+1}(X_0 - P_{R(A)}), \quad k = 0, 1, 2, \ldots.$$

Hence one concludes that whenever $\rho(P_{R(A)} - B) < 1$, the sequence $\{X_k\}$, generated by (18) and (19), converges to the projection $P_{R(A)}$. If X and Y are Hilbert spaces and $P_{R(A)}$ is the orthogonal projection on $R(A)$, then one can choose

$$X_0 = AZ_1A^* \quad \text{and } B = AZ_2A^*,$$

where $Z_1, Z_2 \in l_b(X, X)$. In particular, one can specify $Z_1 = I$, $Z_2 = \alpha I$, where α is a real parameter with the property $\rho(P_{R(A)} - \alpha AA^*) < 1$.

The iterative scheme (8) can be used to calculate the best approximate solution of the equation (1) in abstract spaces. However, in many situations, it is actually used to calculate the best approximate solution of an approximate equation (2), which is frequently a more manageable finite system of linear algebraic equations. Both cases will be demonstrated. First we use scheme (8).

Example 1. Let us calculate the best least squares solution x^* of the inconsistent equation $Ax(s) = (I - K)x(s) = b(s)$, where

$$Kx(s) = \frac{2}{\pi}\int_0^\pi (\sin s \sin \xi + \tfrac{1}{2} \cos s \cos \xi)x(\xi)\, d\xi$$

and $b(s) = s$. The operator K is chosen from [35]. The problem will be solved in $X = Y = L_2[0, \pi]$ using the iterative scheme (8).

We choose $B = \alpha A^*$, where $0 < \alpha < 2 = 2/\|A^*A\|$. This guarantees here that $\rho(P_{R(A)} - \alpha AA^*) < 1$, and the scheme (8) is convergent for every choice of $x_0(s)$ in $N(A)^c = R(A^*)$. For $x_0(s) = 0$ one finds:

$$\begin{aligned} x_1(s) &= \alpha(s - 2 \sin s + 2(\cos s)/\pi) \\ &= \alpha_1^{(1)}s - \alpha_2^{(1)}\sin s + \alpha_3^{(1)}\cos s, \end{aligned}$$

where $\alpha_1^{(1)} = \alpha$, $\alpha_2^{(1)} = 2\alpha$ and $\alpha_3^{(1)} = 2\alpha/\pi$,

$$\begin{aligned} x_2(s) &= [(1 - \alpha)\alpha_1^{(1)} + \alpha]\, s - [(1 - \alpha)\alpha_2^{(1)} + 2\alpha] \sin s \\ &\quad + \left[\left(1 - \frac{\alpha}{4}\right)\alpha_3^{(1)} - \frac{3\alpha}{\pi}\,\alpha_1^{(1)} + \frac{2}{\pi}\,\alpha\right] \cos s \\ &= \alpha_1^{(2)}s - \alpha_2^{(2)}\sin s + \alpha_3^{(2)}\cos s, \end{aligned}$$

where $\alpha_1^{(2)} = (1 - \alpha)\alpha_1^{(1)} + \alpha$, $\alpha_2^{(2)} = (1 - \alpha)\alpha_2^{(1)} + 2\alpha$, $\alpha_3^{(2)} = (1 - \alpha/4)\alpha_3^{(1)} - (3\alpha/\pi)\alpha_1^{(1)} + (2/\pi)\alpha$. In general, if

$$x_k(s) = \alpha_1^{(k)}s - \alpha_2^{(k)}\sin s + \alpha_3^{(k)}\cos s,$$

then

$$\begin{aligned} x_{k+1}(s) &= [(1 - \alpha)\alpha_1^{(k)} + \alpha]\, s - [(1 - \alpha)\alpha_2^{(k)} + 2\alpha] \sin s \\ &\quad + \left[\left(1 - \frac{\alpha}{4}\right)\alpha_3^{(k)} - \frac{3\alpha}{\pi}\,\alpha_1^{(k)} + \frac{2}{\pi}\,\alpha\right] \cos s \\ &= \alpha_1^{(k+1)}s - \alpha_2^{(k+1)}\sin s + \alpha_3^{(k+1)}\cos s. \end{aligned}$$

Since the iterative schemes

$$\alpha_1^{(k+1)} = (1-\alpha)\alpha_1^{(k)} + \alpha,$$

$$\alpha_2^{(k+1)} = (1-\alpha)\alpha_2^{(k)} + 2\alpha,$$

$$\alpha_3^{(k+1)} = \left(1-\frac{\alpha}{4}\right)\alpha_3^{(k)} - \frac{3\alpha}{\pi}\,\alpha_1^{(k)} + \frac{2}{\pi}\alpha$$

are convergent themselves with the solutions $\alpha_1 = 1$, $\alpha_2 = 2$ and $\alpha_3 = -4/\pi$, respectively, one concludes that

$$x^*(s) = s - 2\sin s - 4(\cos s)/\pi$$

is the best least squares solution.

Using (17) with B defined by

$$Bx(s) = x(s) - \frac{2}{\pi}\int_0^\pi \sin s \sin \xi x(\xi)\,d\xi,$$

one can show that X_k converges to A^+, which is here

$$A^+x(s) = x(s) - \frac{2}{\pi}\int_0^\pi (\sin s \sin \xi - \cos s \cos \xi)x(\xi)\,d\xi.$$

The best least squares solution of the equation introduced in Example 1 will be calculated in Example 4 via Kantorovich's approximation theory. We conclude this section by demonstrating how the iterative scheme (8) can be applied to matrices.

Example 2. Calculate the best least squares solution, using the iterative scheme (8), of the inconsistent system

$$\begin{aligned} x_1 \quad\quad + 3x_3 &= 1, \\ -x_1 + x_2 \quad\quad &= 1, \\ x_1 - x_2 \quad\quad &= 1, \\ x_2 + x_3 &= 0. \end{aligned}$$

Here

$$A = \begin{pmatrix} 1 & 0 & 3 \\ -1 & 1 & 0 \\ 1 & -1 & 0 \\ 0 & 1 & 1 \end{pmatrix}, \qquad b = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 0 \end{pmatrix}.$$

Specifying

$$B = \alpha A^t, \quad \text{where } \alpha = 2/\text{trace } A^tA$$

one obtains

$$B = \frac{1}{8}\begin{pmatrix} 1 & -1 & 1 & 0 \\ 0 & 1 & -1 & 1 \\ 3 & 0 & 0 & 1 \end{pmatrix}.$$

It is easy to verify that for the above choice of B and α, $\rho(P_{R(A^t)} - BA) < 1$, whenever rank $A > 1$. One can start iterating from $x^0 = 0$, in which case the following numerical results are obtained:

k	x_1^k	x_2^k	x_3^k
1	0.2500000	0.0000000	0.2500000
2	0.2500000	0.0625000	0.3125000
3	0.2656250	0.0625000	0.3281250
4	0.2656250	0.0664063	0.3320313
5	0.2666016	0.0664063	0.3330018
6	0.2666016	0.0666504	0.3332520
7	0.2666626	0.0666504	0.3333130
8	0.2666626	0.0666657	0.3333282
9	0.2666664	0.0666657	0.3333321
10	0.2666664	0.0666666	0.3333330
11	0.2666667	0.0666666	0.3333333

The eleventh approximation x^{11} gives the best least squares solution correct to six decimal places:

$$x_1^* = 4/15, \quad x_2^* = 1/15, \quad x_3^* = 1/3.$$

Example 3. In this example we apply scheme (8) to solve the problem $\min_{x \in X} \|Ax - b\|$, where X is a two-dimensional Banach space of scalars

$$x = \begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix}, \quad A = \begin{pmatrix} 2 & 1 \\ 0 & 0 \end{pmatrix} \quad \text{and} \quad b = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

If we specify the norm

$$\|x\| = |\xi_1| + |\xi_2| + \max\{|\xi_1|, |\xi_2|, |\xi_1 + \xi_2|\}$$

(which is a TK norm with respect to the basis $\left(\begin{smallmatrix}1\\0\end{smallmatrix}\right), \left(\begin{smallmatrix}0\\1\end{smallmatrix}\right)$), then $E_{R(A)}$ is linear and equal to $\left(\begin{smallmatrix}1 & 0\\0 & 0\end{smallmatrix}\right)$. Thus, in order to solve the problem, we must choose $P_{R(A)} = E_{R(A)}$. Since $N(A)^c$ can be arbitrarily chosen, let it be $N(A)^c = R(A)$. Hence

$$P_{N(A)^c} = \begin{pmatrix} 1 & \frac{1}{2} \\ 0 & 0 \end{pmatrix}.$$

Further, we choose

$$B = \tfrac{1}{4} P_{N(A)^c} P_{R(A)} = \begin{pmatrix} \tfrac{1}{4} & 0 \\ 0 & 0 \end{pmatrix} \quad \text{and} \quad x_0 = 0.$$

The sequence

$$x_1 = \frac{1}{4}\begin{pmatrix} 1 \\ 0 \end{pmatrix}, \qquad x_2 = \frac{3}{8}\begin{pmatrix} 1 \\ 0 \end{pmatrix}, \ldots$$

converging to $x^* = \tfrac{1}{2}\binom{1}{0}$ is obtained.

4. Condition for the Consistency of the Exact Equation. One of the most useful results in the formulation of the classical Kantorovich theory is a lemma which gives a condition for the consistency of the exact equation (1), e.g. [17, p. 543]. This lemma will now be extended so that it also applies to singular equations.

LEMMA 1. *Let $V \in l_b(X, Y)$, E a closed subspace of X and F any subspace of Y containing $V(E)$. If there exists, for every $y \in F$, an $\hat{x} \in E$ such that*

$$\|V\hat{x} - y\| \leqslant q\|y\| \quad \textit{and} \quad \|\hat{x}\| \leqslant \alpha\|y\|, \tag{20}$$

where $q < 1$ and α are constants, then the equation

$$Vx = y \tag{21}$$

has, for every $y \in F$, a solution $x \in E$ satisfying

$$\|x\| \leqslant \alpha\|y\|/(1 - q). \tag{22}$$

Proof. Similarly to the proof in [17], we will construct an exact solution of (21) by recursion. Take an arbitrary $y \in F$. Set $y_1 = y$. By hypothesis, an $\hat{x}_1 \in E$ exists such that

$$\|V\hat{x}_1 - y_1\| \leqslant q\|y_1\|, \qquad \|\hat{x}_1\| \leqslant \alpha\|y_1\|. \tag{23}$$

Denote

$$y_2 = y_1 - V\hat{x}_1. \tag{24}$$

Clearly $y_2 \in F$, since $y_1 \in F$ and F is a subspace containing $V(E)$. We now apply the condition (20) to y_2. This implies the existence of $\hat{x}_2 \in E$ such that

$$\|V\hat{x}_2 - y_2\| \leqslant q\|y_2\| = q\|y_1 - V\hat{x}_1\|, \quad \text{by (24)}$$

$$\leqslant q^2\|y_1\|, \quad \text{by (23)}.$$

Also, $\|\hat{x}_2\| \leqslant \alpha\|y_2\| \leqslant \alpha q\|y_1\|$. Continuing this process, sequences $\{y_k\}$ and $\{\hat{x}_k\}$ are obtained such that

$$y_{k+1} = y_k - V\hat{x}_k, \qquad k = 1, 2, \ldots, \tag{25}$$

and

$$\|y_k\| \leqslant q^{k-1}\|y_1\|, \qquad \|\hat{x}_k\| \leqslant \alpha q^{k-1}\|y_1\|. \tag{26}$$

By iteration, (24) and (25) give

$$y_{k+1} = y_1 - V(\hat{x}_1 + \hat{x}_2 + \cdots + \hat{x}_k), \qquad k = 1, 2, \ldots . \tag{27}$$

Using the second inequality in (26), and recalling that $y_1 = y$, one obtains

$$\left\| \sum_{k=1}^{\infty} \hat{x}_k \right\| \leqslant \sum_{k=1}^{\infty} \|\hat{x}_k\| \leqslant \alpha \sum_{k=1}^{\infty} q^{k-1} \|y\|.$$

Since $q < 1$, the series $\Sigma_{k=1}^{\infty} \hat{x}_k$ is convergent. Hence $x \triangleq \Sigma_{k=1}^{\infty} \hat{x}_k$ belongs to E, since E is a closed subspace. Furthermore,

$$\|x\| \leqslant \alpha \sum_{k=1}^{\infty} q^{k-1} \|y\| = \frac{\alpha}{1-q} \|y\|.$$

So taking limits in (27) gives

$$\lim_{k\to\infty} y_k = \lim_{k\to\infty} (y_1 - V(\hat{x}_1 + \hat{x}_2 + \cdots + \hat{x}_k))$$

$$= y_1 - Vx, \quad \text{by the continuity of } V.$$

Also, since $q < 1$, we see from (26) that $\lim_{k\to\infty} y_k = 0$. Thus $Vx = y$. We have shown that $x \in E$ is a solution of (21) and it satisfies (22). □

Note that part of the conclusion of Lemma 1 is actually that $F = V(E)$. Lemma 1 has been proved in [17] in the special case when $E = X$ and $F = Y$. The above result will be used in the next section in the approximation theory. However, Lemma 1 is of an independent interest; and in the remainder of this section we will show how, using the lemma, one can establish some new and some well-known estimates related to the equation (1).

PROPOSITION 1. *Let $M \in l_b(X, Y)$ and let $N(M)$ and $R(M)$ have topological complements. Denote by M^+ the generalized inverse of M with respect to these topological complements. Consider*

$$A = M + N \tag{28}$$

such that $A \in l_b(X, Y)$ and $\|NM^+\| < 1$. If

$$R(N) \subset R(M), \tag{29}$$

then the equation

$$Ax = y \tag{30}$$

has, for every $y \in R(M)$, a solution $x^* \in N(M)^c$. Also,

$$\|x^*\| \leqslant \frac{\|M^+\|}{1 - \|NM^+\|} \|y\| \tag{31}$$

and $R(A) = R(M)$. In addition to the above assumptions, if

$$N(M)^c \cap N(A) = \{0\}, \tag{32}$$

then

$$\|A^+\| \leqslant \frac{\|M^+\|}{1 - \|NM^+\|}, \tag{33}$$

where A^+ denotes the generalized inverse of A with respect to $P_{N(A)^c} = P_{N(M)^c}$ and $P_{R(A)} = P_{R(M)}$.

Proof. We will show that the assumptions of Lemma 1 are satisfied with $V = A$, $E = N(M)^c$, $F = R(M)$, $q = \|NM^+\|$ and $\alpha = \|M^+\|$. Choose an arbitrary $y \in R(M)$ and let $\hat{x} = M^+y$. Then

$$\begin{aligned}\|A\hat{x} - y\| &= \|(M + N)\hat{x} - y\| \\ &= \|(M + N)\hat{x} - MM^+y\|, \quad \text{since } y \in R(M) \\ &\leqslant \|NM^+\| \, \|y\|,\end{aligned}$$

and $\|\hat{x}\| = \|M^+y\| \leqslant \|M^+\| \, \|y\|$. Thus (20) holds. Now we apply Lemma 1 to the equation (30) to conclude that for every $y \in R(M)$, the equation $Ax = y$ has a solution $x^* \in N(M)^c$ satisfying (31). This implies that

$$R(M) \subset R(A). \tag{34}$$

However,

$$R(A) \subset R(M), \quad \text{by (28) and (29).} \tag{35}$$

Now (34) and (35) imply

$$R(A) = R(M). \tag{36}$$

Since $N(M)^c$ is isomorphic to $R(M)$ via M, it is also isomorphic via A, by (36) and the assumption (32). Thus, one can choose $N(A)^c = N(M)^c$. This determines A^+ with respect to the topological complements $N(M)^c$ and $R(M)$. Now, since $x^* \in N(A)^c = N(M)^c$, we have $x^* = A^+y$, and (33) follows from (31). □

COROLLARY 1. *Let $H \in l_b(X)$ with $\|H\| < 1$ and $P \in l_b(X)$ such that $P^2 = P$. If $R(H) \subset R(P)$, then the equation $(P + H)x = y$ has a solution $x^* \in R(P)$ for each $y \in R(P)$. Also $\|x^*\| \leqslant \|y\|/(1 - \|H\|)$, $R(P + H) = R(P)$ and $\|(P + H)^+\| \leqslant 1/(1 - \|H\|)$, where $(P + H)^+$ denotes the generalized inverse with respect to $P_{N(P+H)^c} = P$ and $P_{R(P+H)} = P$.*

Proof. We will show that the hypotheses of Proposition 1 are satisfied with $M = P$ and $N = H$. Clearly, $P \in l_b(X)$ has topological complements $N(P)^c = R(P)$ and $R(P)^c = N(P)$. So P^+ $(= P)$ is the generalized inverse of P with respect to these complements. Also, $P + H \in l_b(X)$, since $\|H\| < 1$. Take an arbitrary $y \in R(P)$. Then $\hat{x} \triangleq P^+y$ (see the proof of Proposition 1, where $M = P$) is equal to y, i.e. $\hat{x} = y$, since $P = P^+$. Therefore, the assumption $\|HP^+\| < 1$, in Proposition 1, can be replaced by $\|H\| < 1$. Also, the assumption (32), which reads here

$$N(P)^c \cap N(P + H) = \{0\} \tag{37}$$

is satisfied. If (37) were not true, there would exist an $x \neq 0$ such that both

$$x \in N(P)^c = R(P) \quad \text{and} \quad x \in N(P + H).$$

Hence $(P + H)x = x + Hx = 0$, which contradicts the assumption $\|H\| < 1$. □

COROLLARY 2 (BEN-ISRAEL [6]). *Let H be an $n \times n$ real matrix, $\|H\| < 1$ and L be a subspace of R^m such that $R(H) \subset L$. Then*

$$\|(P_L + H)^+\| \leqslant 1/(1 - \|H\|).$$

Proof. Specify $P = P_L$ in Corollary 1. □

COROLLARY 3 (KANTOROVICH AND AKILOV [17, p. 172]). *Let $M \in l_b(X)$ and suppose that $M^{-1} \in l_b(X)$ exists. If $N \in l_b(X)$ and $\|NM^{-1}\| < 1$, then $(M + N)^{-1}$ exists and*

$$\|(M + N)^{-1}\| \leqslant \|M^{-1}\|/(1 - \|NM^{-1}\|).$$

Proof. Apply Proposition 1 to the case when $X = Y$ and M^{-1} exists. □

When $M = I$, Corollary 3 becomes the classical result of Banach, e.g. [17]. □

5. Kantorovich's Theory for Singular Equations. Let $\widetilde{X}$ and $\widetilde{Y}$ be closed subspaces of the Banach spaces X and Y, respectively. Further, let $\widetilde{X}$ and $\widetilde{Y}$ be isomorphic via mappings J_0 and H_0 to the Banach spaces $\overline{X}$ and $\overline{Y}$, respectively. Suppose also that J and H are linear extensions of J_0 and H_0 to all of X and Y, respectively. Such extensions always exist, for we may take $J = J_0 P_{\widetilde{X}}$ and $H = H_0 P_{\widetilde{Y}}$. In many practical situations, $\overline{X}$ and $\overline{Y}$ are chosen to be finite dimensional.

Consider the following two equations:

$$Ax = b, \tag{1}$$

where $A: X \longrightarrow Y$, $b \in Y$, and

$$\overline{A}\overline{x} = \overline{b}, \tag{2}$$

where $\overline{A}: \overline{X} \longrightarrow \overline{Y}$, $\overline{b} \in \overline{Y}$. We shall refer to (1) as the "exact" equation and to (2) as its "approximate" equation. We assume that $A \in l_b(X, Y)$, $\overline{A} \in l_b(\overline{X}, \overline{Y})$ and that the following decompositions are possible:

$$X = N(A) \oplus N(A)^c, \qquad \overline{X} = N(\overline{A}) \oplus N(\overline{A})^c,$$

$$Y = R(A) \oplus R(A)^c, \qquad \overline{Y} = R(\overline{A}) \oplus R(\overline{A})^c.$$

The symbol $\oplus$ is here used to indicate that all eight complements are necessarily closed. Denote by $A^+ \in l_b(Y, X)$ the generalized inverse of A relative to the continuous projections $P_{N(A)}$ and $P_{R(A)}$, and by $\overline{A}^+ \in l_b(\overline{Y}, \overline{X})$ the generalized inverse of $\overline{A}$ relative to the continuous projections $P_{N(\overline{A})}$ and $P_{R(\overline{A})}$. Let us denote by $x^* = A^+b$ and $\overline{x}^* = \overline{A}^+\overline{b}$, the best approximate solutions of the equations (1) and (2), respectively.

In the sequel we will state results relating the exact and approximate equations when some or all of the following conditions are satisfied:

(I) The operator A is represented as $A = M + N$, where M is bounded and $X = N(M) \oplus N(M)^c$, $N(A)^c \subset N(M)^c$ and $Y = R(M) \oplus R(M)^c$. (In this situation M^+ denotes the generalized inverse of M, relative to the continuous projections $P_{N(M)^c}$ and $P_{R(M)}$.)

(II) J_0 maps $N(M)^c \cap \widetilde{X}$ into $N(\overline{A})^c \subset \overline{X}$.

(III) H has the property $\bar{A}^+\bar{b} = \bar{A}^+ HP_{R(A)}b$.

(IV) $\|\bar{A}J_0\tilde{x} - HA\tilde{x}\| \leq \epsilon\|\tilde{x}\|$ for some constant $\epsilon \geq 0$ and all $\tilde{x} \in N(M)^c \cap \tilde{X}$.

(V) For every $x \in N(A)^c$ there is a $u \in N(M)^c \cap \tilde{X}$ such that $\|Mu - P_{R(M)}Nx\| \leq \eta_1\|x\|$ for some constant $\eta_1 \geq 0$.

(VI) There exists a vector $v \in N(M)^c \cap \tilde{X}$ such that $\|Mv - P_{R(M)}P_{R(A)}b\| \leq \eta_2\|P_{R(A)}b\|$ for some constant $\eta_2 \geq 0$.

THEOREM 3. (*Conditions for the solvability of the approximate equation.*) *Let the conditions* (I), (IV) *and* (V) *be satisfied. In addition, suppose that*

(A) $M^+: \tilde{Y} \longrightarrow \tilde{X}$,

(B) $H_0^{-1}HR(A) \subset R(A)$,

(C) $J(N(M)^c)$ *is closed and*

(D) $R(\bar{A}) \subset HR(A)$.

If

$$q = [\epsilon(1 + \eta_1\|M^+\|) + \eta_1\|HA\|\,\|M^+\|]\,\|A^+H_0^{-1}\| < 1, \tag{38}$$

then the equation

$$\bar{A}\bar{x} = HP_{R(A)}b \tag{39}$$

has a solution $\bar{x}^* \in J(N(M)^c)$ *for every* $b \in Y$. *Furthermore*

$$\|\bar{x}^*\| \leq \alpha\|HP_{R(A)}b\|/(1 - q), \tag{40}$$

where

$$\alpha = \|J_0\|(1 + \eta_1\|M^+\|)\|A^+H_0^{-1}\|. \tag{41}$$

Proof. It is sufficient to show that the conditions (20) of Lemma 1 are satisfied for the equation (39) with $E = J(N(M)^c)$, $F = HR(A)$, $V = \bar{A}$, $y = HP_{R(A)}b$ and α and q as in (38) and (41). First, consider the equation $Ax = H_0^{-1}HP_{R(A)}b$ and its solution $x_0 = A^+H_0^{-1}HP_{R(A)}b$. Denote $z = Mx_0 - H_0^{-1}HP_{R(A)}b$. Since $x_0 \in N(A)^c$ and $N(A)^c \subset N(M)^c$, by condition (I), we find that

$$x_0 = M^+z + M^+H_0^{-1}HP_{R(A)}b \tag{42}$$

and also

$$Ax_0 = H_0^{-1}HP_{R(A)}b \tag{43}$$

by definition of x_0 and condition (B). Therefore, $z = Mx_0 - Ax_0 = -Nx_0$, since $A = M + N$.

Now, for $x = -x_0$, condition (V) implies that there exists $u \in N(M)^c \cap \tilde{X}$ such that

$$\|Mu - P_{R(M)}N(-x_0)\| = \|Mu - P_{R(M)}z\| \leq \eta_1\|x_0\| \tag{44}$$

for some $\eta_1 \geq 0$. Denote $\tilde{x} = u + M^+H_0^{-1}HP_{R(A)}b$. Note that $\tilde{x} \in N(M)^c \cap \tilde{X}$, by conditions (V) and (A). We will now show that $J_0\tilde{x}$ is the required element $\hat{x}$ of E in Lemma 1. First,

$$\|\bar{A}J_0\tilde{x} - HP_{R(A)}b\| = \|\bar{A}J_0\tilde{x} - HH_0^{-1}HP_{R(A)}b\|$$
$$= \|\bar{A}J_0\tilde{x} - HAx_0\|, \quad \text{by condition (B) and definition of } x_0$$
$$\leqslant \|\bar{A}J_0\tilde{x} - HA\tilde{x}\| + \|HA\tilde{x} - HAx_0\|, \quad \text{by the triangular inequality}$$
$$\leqslant \epsilon\|\tilde{x}\| + \|HA\|\,\|\tilde{x} - x_0\|, \quad \text{by condition (IV).}$$

Since

$$\|\tilde{x} - x_0\| = \|u + M^+H_0^{-1}HP_{R(A)}b - M^+z - M^+H_0^{-1}HP_{R(A)}b\|,$$
$$\text{by definition of } \tilde{x} \text{ and } x_0$$
$$= \|u - M^+z\| = \|M^+Mu - M^+P_{R(M)}z\|$$
$$\leqslant \eta_1\|M^+\|\,\|x_0\|, \quad \text{by (44)} \tag{45}$$
$$\leqslant \eta_1\|M^+\|\,\|A^+H_0^{-1}\|\,\|HP_{R(A)}b\|, \quad \text{by definition of } x_0 \text{ and}$$
$$\|\tilde{x}\| \leqslant \|x_0\| + \|\tilde{x} - x_0\|$$
$$\leqslant (1 + \eta_1\|M^+\|)\|x_0\|, \quad \text{by (45)}$$
$$\leqslant (1 + \eta_1\|M^+\|)\|A^+H_0^{-1}\|\,\|HP_{R(A)}b\|,$$

the above inequality gives

$$\|\bar{A}J_0\tilde{x} - HP_{R(A)}b\| \leqslant [\epsilon(1 + \eta_1\|M^+\|)\|A^+H_0^{-1}\|$$
$$+ \eta_1\|HA\|\,\|M^+\|\,\|A^+H_0^{-1}\|]\,\|HP_{R(A)}b\|$$
$$= q\|HP_{R(A)}b\|, \tag{46}$$

where q is the constant defined by (38). Also

$$\|J_0\tilde{x}\| \leqslant \|J_0\|\,\|\tilde{x}\| \leqslant \alpha\|HP_{R(A)}b\|, \tag{47}$$

where α is defined by (41). The inequalities (46) and (47) correspond to the assumptions (20) of Lemma 1 with $V = \bar{A}$, $\hat{x} = J_0\tilde{x}$ and $y = HP_{R(A)}b$. Conditions (C) and (D) guarantee that $E = J(N(M)^c)$ be a closed subspace of X and $F = HR(A)$ be a subspace of Y containing $V(E) = \bar{A}(J(N(M)^c))$. All conditions of Lemma 1 are now satisfied, and the conclusions of Theorem 3 follow. □

COROLLARY 4. *Let* $A \in l_b(X, Y)$ *and suppose that the conditions* (I), (IV), (V), (A), (B), (C) *and* (D) *are satisfied. If* $q < 1$ *and* $\bar{A}$ *satisfies the condition*

$$J(N(M)^c) \subset N(\bar{A})^c \tag{E}$$

then

$$\|\bar{A}^+\| \leqslant \alpha\|P_{R(\bar{A})}\|/(1 - q),$$

where q *and* α *are as in Theorem* 3.

Proof. We need to show that

$$\|\bar{A}^+\bar{y}\| \leqslant \frac{\alpha}{1 - q}\|P_{R(\bar{A})}\|\,\|\bar{y}\|, \quad \text{for every } \bar{y} \in \bar{Y}.$$

Let $\bar{x}^*$ denote the solution in $J(N(M)^c)$ of the equation $\bar{A}\bar{x} = P_{R(\bar{A})}\bar{y}$. Such an $\bar{x}^*$ exists, by Theorem 3. In fact, by condition (E), $\bar{x}^* = \bar{A}^+\bar{y}$, i.e. $\bar{x}^*$ is a unique best approximate solution in $N(\bar{A})^c$ of the equation $\bar{A}\bar{x} = \bar{y}$. But $P_{R(\bar{A})}\bar{y} = HP_{R(A)}b$ for some $b \in Y$, by condition (D). So $\bar{x}^*$ is also a solution of the equation $\bar{A}\bar{x} = HP_{R(A)}b$ and it satisfies

$$\|\bar{x}^*\| \leqslant \frac{\alpha}{1-q}\|HP_{R(A)}b\|, \quad \text{by (40).}$$

Therefore,

$$\|\bar{A}^+\bar{y}\| \leqslant \frac{\alpha}{1-q}\|HP_{R(A)}b\| = \frac{\alpha}{1-q}\|P_{R(\bar{A})}\bar{y}\|$$

$$\leqslant \frac{\alpha}{1-q}\|P_{R(\bar{A})}\|\|\bar{y}\|. \quad \square$$

If $X = Y$, $\tilde{X} = \tilde{Y} = \bar{X} = \bar{Y}$, $M = I$ and A^{-1} exists then the conditions (A), (B), (C), (D), and (I) are trivially satisfied, while (IV), (V) and (VI) reduce to

(IV$'$) $\|\bar{A}\tilde{x} - HA\tilde{x}\| \leqslant \epsilon\|\tilde{x}\|$ for some constant $\epsilon \geqslant 0$ and all $\tilde{x} \in \tilde{X}$.

(V$'$) For every $x \in X$ there is a $u \in \tilde{X}$ such that $\|u - Nx\| \leqslant \eta_1\|x\|$ for some constant $\eta_1 \geqslant 0$.

(VI$'$) There exists a vector $v \in \tilde{X}$ such that $\|v - b\| \leqslant \eta_2\|b\|$ for some constant $\eta_2 \geqslant 0$.

In the nonsingular case, Theorem 3 reduces to the following result of Kantorovich and Akilov [17, p. 545].

COROLLARY 5. *Let $A \in l_b(X)$ have an inverse and let the conditions* (IV$'$) *and* (V$'$) *be satisfied. If*

$$q = [\epsilon(1 + \eta_1) + \eta_1\|HA\|]\,\|A^{-1}\| < 1,$$

then the equation $\bar{A}\bar{x} = \bar{b}$ has a solution $\bar{x}^* \in \bar{X}$ for every $\bar{b} \in \bar{X}$. Also

$$\|\bar{x}^*\| \leqslant \alpha\|\bar{b}\|/(1-q),$$

where $\alpha = (1 + \eta_1)\|A^{-1}\|$.

An estimate for the norm of the generalized inverse $\bar{A}^+$ was obtained using Theorem 3. In the nonsingular case, Corollary 4 gives the following result of Kantorovich and Akilov [17, p. 546].

COROLLARY 6. *Let the hypotheses and the notation of Corollary* 5 *hold and let $\bar{A}$ satisfy the condition*:

(E$'$) *"The existence of a solution of the equation $\bar{A}\bar{x} = \bar{b}$ for every $\bar{b} \in \bar{X}$ implies its uniqueness".*

Then

$$\|\bar{A}^{-1}\| \leqslant \alpha/(1-q).$$

Proof. Condition (E$'$) and Corollary 5 imply the existence of $\bar{A}^{-1}$. The result now follows from Corollary 4, since the conditions (D) and (E) are satisfied when $\bar{A}^{-1}$ and A^{-1} exist. $\square$

The following theorem estimates the distance between the best approximate solution $\bar{x}^*$ of the approximate equation and the best approximate solution x^* of the exact equation.

THEOREM 4. (*Estimate of the distance between best approximate solutions.*) *Consider the equation* $Ax = b$ *and its approximate equation* $\bar{A}\bar{x} = \bar{b}$. *If the condition* (I)–(VI) *are satisfied, then*

(48) $$\|x^* - J_0^{-1}\bar{x}^*\| \leqslant p\|x^*\|,$$

where

(49) $$p = \epsilon(1 + c)\|J_0^{-1}\|\|\bar{A}^+\| + c(1 + \|J_0^{-1}\bar{A}^+HA\|)$$

and

(50) $$c = \min\{1, (\eta_1 + \eta_2\|A\|)\|M^+\|\}.$$

Proof. First we show that x^* can be approximated by an $\tilde{x} \in N(M)^c \cap \tilde{X}$ to the order of $\eta_1 + \eta_2$. We know, by conditions (V) and (VI), that there exist u and v in $N(M)^c \cap \tilde{X}$ such that

(51) $$\|Mu - P_{R(M)}Nx^*\| \leqslant \eta_1\|x^*\|$$

and

(52) $$\|Mv - P_{R(M)}P_{R(A)}b\| \leqslant \eta_2\|P_{R(A)}b\|.$$

Denote $\tilde{x} = M^+(Mv - Mu)$. Clearly, $\tilde{x} \in N(M)^c \cap \tilde{X}$. We now show that x^* can be approximated by $\tilde{x}$ to the order of $\eta_1 + \eta_2$.

$$\begin{aligned}
\|x^* - \tilde{x}\| &= \|M^+Mx^* - \tilde{x}\|, \quad \text{by condition (I)}\\
&= \|-M^+Nx^* + M^+(M+N)x^* - M^+(Mv - Mu)\|, \\
&\qquad\qquad \text{by definition of } \tilde{x}\\
&= \|-M^+Nx^* + M^+P_{R(A)}b - M^+(Mv - Mu)\|, \\
&\qquad\qquad \text{since } A = M + N \text{ and } Ax^* = P_{R(A)}b\\
&\leqslant \|M^+\|(\|Mu - P_{R(M)}Nx^*\| + \|Mv - P_{R(M)}P_{R(A)}b\|), \\
&\qquad\qquad \text{since } M^+ = M^+P_{R(M)}\\
&\leqslant \|M^+\|(\eta_1\|x^*\| + \eta_2\|P_{R(A)}b\|), \quad \text{by (51) and (52)}\\
&\leqslant \|M^+\|(\eta_1 + \eta_2\|A\|)\|x^*\|, \quad \text{since } \|P_{R(A)}b\| = \|Ax^*\| \leqslant \|A\|\|x^*\|.
\end{aligned}$$

Hence, we conclude that there exists an $\tilde{x} \in N(M)^c \cap \tilde{X}$ such that

(53) $$\|x^* - \tilde{x}\| \leqslant c\|x^*\|,$$

where $c = \min\{1, (\eta_1 + \eta_2\|A\|)\|M^+\|\}$. (Note that $c \leqslant 1$, since we can always choose $\tilde{x} = 0$ in (53).)

Let us now prove (48). Denote $\bar{x}_0 = \bar{A}^+ HA\tilde{x}$. Then

$$\|x^* - J_0^{-1}\bar{x}^*\| \leqslant \|x^* - \tilde{x}\| + \|\tilde{x} - J_0^{-1}\bar{x}_0\| + \|J_0^{-1}\bar{x}_0 - J_0^{-1}\bar{x}^*\|. \tag{54}$$

The first term on the right-hand side $\|x^* - \tilde{x}\|$ is estimated by (53). The two remaining terms will now be estimated. First

$$\begin{aligned}
\tilde{x} - J_0^{-1}\bar{x}_0 &= \tilde{x} - J_0^{-1}\bar{A}^+ H\bar{A}\tilde{x}, \quad \text{by definition of } \bar{x}_0 \\
&= J_0^{-1}(J_0 - \bar{A}^+ HA)\tilde{x}, \quad \text{since } \tilde{x} \in \tilde{X} \\
&= J_0^{-1}\bar{A}^+(\bar{A}J_0 - HA)\tilde{x}, \\
&\qquad \text{since } \bar{A}^+\bar{A}J_0\tilde{x} = P_{N(\bar{A})^c} J_0\tilde{x} = J_0\tilde{x}, \text{ by condition (II).}
\end{aligned}$$

Hence,

$$\begin{aligned}
\|\tilde{x} - J_0^{-1}\bar{x}_0\| &\leqslant \|J_0^{-1}\bar{A}^+\|\,\|(\bar{A}J_0 - HA)\tilde{x}\| \\
&\leqslant \epsilon\|J_0^{-1}\bar{A}^+\|\,\|\tilde{x}\|, \quad \text{by condition (IV)} \\
&\leqslant \epsilon\|J_0^{-1}\bar{A}^+\|(\|x^*\| + \|x^* - \tilde{x}\|), \quad \text{by the triangle inequality} \\
&\leqslant \epsilon(1 + c)\|J_0^{-1}\bar{A}^+\|\,\|x^*\|, \quad \text{by (53).}
\end{aligned}$$

The third term is estimated as follows:

$$\begin{aligned}
\|J_0^{-1}\bar{x}_0 - J_0^{-1}\bar{x}^*\| &= \|J_0^{-1}(\bar{A}^+ HA\tilde{x} - \bar{A}^+\bar{A}\bar{x}^*)\|, \quad \text{by definition of } \bar{x}_0 \\
&= \|J_0^{-1}(\bar{A}^+ HA\tilde{x} - \bar{A}^+ P_{R(\bar{A})}\bar{b})\| \\
&= \|J_0^{-1}(\bar{A}^+ HA\tilde{x} - \bar{A}^+ HP_{R(A)}b)\|, \quad \text{by Condition (III)} \\
&\leqslant \|J_0^{-1}\bar{A}^+ HA\|\,\|\tilde{x} - x^*\|, \quad \text{since } P_{R(A)}b = Ax^* \\
&\leqslant c\|J_0^{-1}\bar{A}^+ HA\|\,\|x^*\|, \quad \text{by (53).}
\end{aligned}$$

After substituting the above estimates into (54), the conclusion follows. □

Remark 5. It may happen that we could approximate x^* by some $\tilde{x} \in N(M)^c \cap \tilde{X}$ directly. Then we no longer need the conditions (V) and (VI), and we can set $c = \min\{1, \|x^* - \tilde{x}\|\}$ in (49).

Remark 6. If $p < 1$, we can write the estimate (48) as follows

$$\|x^* - J_0^{-1}\bar{x}^*\| \leqslant p\|J_0^{-1}\bar{x}^*\|/(1 - p).$$

This is true, since

$$\|x^* - J_0^{-1}\bar{x}^*\| \leqslant p\|x^*\| \leqslant p(\|J_0^{-1}\bar{x}^*\| + \|x^* - J_0^{-1}\bar{x}^*\|).$$

Remark 7. If X and Y are Hilbert spaces and $N(A)^c = R(A^*)$, $R(A)^c = N(A^*)$, then Theorem 4 reduces to a result obtained by Zlobec in [41]. However, the Hilbert space version of Theorem 4 is proved there under slightly different assumptions.

Remark 8. It may happen that one cannot satisfy condition (III) but that a constant η_3 such that

$$\|\bar{A}^+\bar{b} - \bar{A}^+ HP_{R(A)}b\| \leqslant \eta_3\|x^*\|$$

is found. In this case the constant p in (48) is different. Now

$$\|J_0^{-1}\bar{x}_0 - J_0^{-1}\bar{x}^*\| \leqslant (c\|J_0^{-1}\bar{A}^+HA\| + \eta_3\|J_0^{-1}\|)\|x^*\|,$$

and hence

$$p = [\epsilon(1+c)\|\bar{A}^+\| + \eta_3]\|J_0^{-1}\| + c(1 + \|J_0^{-1}\bar{A}^+HA\|).$$

In the nonsingular case, we get the following result from [17, p. 547]. (Recall that in our setting $x^* \in N(A)^c$. Also, as in [17], we specify $\bar{b} = Hb$.)

COROLLARY 7. *Let the conditions* (IV′), (V′) *and* (VI′) *be satisfied and let* A^{-1} *and* $\bar{A}^{-1}$ *exist. Then*

$$\|x^* - \bar{x}^*\| \leqslant p\|x^*\|,$$

where x^* *is the solution of the exact equation* (1), $\bar{x}^*$ *is the solution of the approximate equation* (2) *and*

$$p = 2\epsilon\|\bar{A}^{-1}\| + (\eta_1 + \eta_2\|A\|)(1 + \|\bar{A}^{-1}HA\|).$$

Proof. Specify $M = I$, $X = Y$ and $\widetilde{X} = \widetilde{Y} = \bar{X} = \bar{Y}$ in Theorem 4. □

Our next result gives conditions for convergence of approximation schemes. Suppose that the exact equation $Ax = b$ is approximated by a sequence of equations $\bar{A}_n\bar{x} = \bar{b}_n$, $n = 1, 2, \ldots$, rather than by a single equation. This determines a sequence of the spaces $\widetilde{X}_n$, $\widetilde{Y}_n$, $\bar{X}_n$, $\bar{Y}_n$, the operators $\bar{A}_n$, $(J_0)_n$, $(H_0)_n$, J_n, H_n and the constants ϵ_n, $(\eta_1)_n$, $(\eta_2)_n$, c_n, q_n, α_n, p_n, $n = 1, 2, \ldots$. For the sake of notational simplicity these indices will generally be omitted in the sequel. The following theorem gives conditions for the convergence of the sequence $\bar{x}_n^*$, the best approximate solution of $\bar{A}_n\bar{x}_n = \bar{b}_n$, $n = 1, 2, \ldots$, to x^*, the best approximate solution of the exact equation.

THEOREM 5. (*Convergence of the best approximate solutions.*) *Consider the equation* $Ax = b$ *and a sequence of approximate equations* $\bar{A}\bar{x} = \bar{b}$. *Suppose that for each* $n = 1, 2, \ldots$ *the conditions*:

(i) (I)–(VI) *and* (A), (B), (C), (D) *and* (E),

(ii) $\sup_n\|J_0\| < \infty$, $\sup_n\|H_0^{-1}\| < \infty$ *and* $\sup_n\|P_{R(\bar{A})}\| < \infty$,

(iii) $\lim_{n\to\infty} \epsilon\|J_0^{-1}\| = 0$, $\lim_{n\to\infty} \eta_1\|J_0^{-1}\|\|H\| = 0$, $\lim_{n\to\infty} \eta_2\|J_0^{-1}\|\|H\| = 0$

are satisfied. Then $\lim_{n\to\infty} \eta_1 = 0$, $\lim_{n\to\infty} \eta_2 = 0$ *and the sequence of best approximate solutions of* $\bar{A}\bar{x} = \bar{b}$ *converges to the best approximate solution* x^* *of* $Ax = b$, *i.e.*

$$\lim_{n\to\infty} \|x^* - J_0^{-1}\bar{x}^*\| = 0.$$

More precisely,

$$\|x^* - J_0^{-1}\bar{x}^*\|$$
$$\leqslant [\epsilon c_1\|J_0^{-1}\| + \eta_1(c_2 + c_3\|J_0^{-1}\|\,\|H\|) + \eta_2(c_4 + c_5\|J_0^{-1}\|\|H\|)]\,\|x^*\|,$$

where c_1 *to* c_5 *are some constants independent of the index* n.

Proof. Since $H_0H_0^{-1} = I$, it follows that $\|H_0^{-1}\|\|H_0\| \geqslant 1$, $n = 1, 2, \ldots$. Hence

$$\inf_n \|H_0\| > 0 \tag{55}$$

using the second assumption in (ii). Similarly, one concludes that

$$\inf_n \|J_0^{-1}\| > 0 \tag{56}$$

using the first assumption in (ii). Also $\|H\| \geqslant \|H_0\|$, since $H_0 = H|_{\widetilde{Y}}$. Therefore, by (55),

$$\inf_n \|H\| > 0. \tag{57}$$

From (56), (57) and condition (iii), we conclude that, in particular,

$$\lim_{n\to\infty} \epsilon = 0, \quad \lim_{n\to\infty} \eta_1 = 0, \quad \lim_{n\to\infty} \eta_2 = 0, \quad \lim_{n\to\infty} \eta_1\|H\| = 0. \tag{58}$$

Recall the constant q introduced in Theorem 3:

$$q = [\epsilon(1 + \eta_1\|M^+\|) + \eta_1\|HA\|\,\|M^+\|]\,\|A^+H_0^{-1}\|.$$

For sufficiently large n, using (58), one has $q < ½$ and for such values of n Theorem 3 is applicable. But we can also apply Corollary 4 to obtain

$$\|\bar{A}^+\| \leqslant \alpha\|P_{R(\bar{A})}\|/(1 - q) \leqslant 2\alpha\|P_{R(\bar{A})}\|.$$

Since $\alpha = \|J_0\|(1 + \eta_1\|M^+\|)\|A^+H_0^{-1}\|$, one concludes, using the third assumption in (ii), that $\|\bar{A}^+\|$ is bounded independently of the index n, i.e.

$$\sup_n \|\bar{A}^+\| = s < \infty. \tag{59}$$

The desired estimate now follows for sufficinetly large n:

$$\|x^* - J_0^{-1}\bar{x}^*\| \leqslant [2\epsilon\|J_0^{-1}\|\,\|\bar{A}^+\| + (\eta_1 + \eta_2\|A\|)(1 + \|J_0^{-1}\bar{A}^+HA\|)\|M^+\|]\,\|x^*\|,$$

by (49) and (50)

$$\leqslant [\epsilon c_1\|J_0^{-1}\| + \eta_1(c_2 + c_3\|J_0^{-1}\|\,\|H\|) + \eta_2(c_4 + c_5\|J_0^{-1}\|\,\|H\|)]\,\|x^*\|,$$

where $c_1 = 2s$, $c_2 = \|M^+\|$, $c_3 = s\|A\|\,\|M^+\|$, $c_4 = \|A\|\,\|M^+\|$, $c_5 = s\|A\|^2\|M^+\|$, and s is defined by (59). The right-hand side in the above inequality tends to zero when $n \longrightarrow \infty$, by (iii) and (58). □

A corresponding result in the nonsingular case is given in [17, p. 549] as follows.

COROLLARY 8. *Consider the equation $Ax = b$ and a sequence of approximate equations $\bar{A}\bar{x} = \bar{b}$. Assume that A^{-1} exists and that $\bar{A}$ satisfies condition* (E′) *for each $n = 1, 2, \ldots$. Assume further that for each $n = 1, 2, \ldots$ the conditions* (IV′), (V′), (VI′) *are satisfied, and that*

$$\lim_{n\to\infty} \epsilon = 0, \quad \lim_{n\to\infty} \eta_1\|H\| = 0 \quad \textit{and} \quad \lim_{n\to\infty} \eta_2\|H\| = 0.$$

Then the approximate equations are consistent for sufficiently large n and the sequence of approximate solutions converges to the exact solution x^* of $Ax = b$, i.e.

$$\lim_{n\to\infty} \|x^* - \bar{x}^*\| = 0.$$

More precisely,

$$\|x^* - \bar{x}^*\| \leqslant [\epsilon c_1 + \eta_1(c_2 + c_3\|H\|) + \eta_2(c_4 + c_5\|H\|)]\,\|x^*\|,$$

where c_1 to c_5 are some constants independent of the index n.

Proof. Set $X = Y$, $\bar{X} = \bar{Y} = \widetilde{X} = \widetilde{Y}$ and $M = I$ in Theorem 3. Then $J_0 = H_0 = I$ and $\|H\| = \|J\| \geqslant 1$. Furthermore, from Corollary 3, we see that $R(\bar{A}) = \bar{X}$. So condition (E) is satisfied and $\|P_{R(\bar{A})}\| = 1$. Conditions (i)–(iii) of Theorem 5 hold and the result follows. □

Remark 9. Constants ϵ, η_1, η_2 and c_1 to c_5 in Theorem 5 reduce, in the nonsingular case, to the corresponding constants in Corollary 8.

Let us recall that Corollary 4 gave us an estimate for $\|\bar{A}^+\|$ in terms of $\|A^+\|$ via constants α and q. Our last result gives us the reverse estimate.

THEOREM 6. (*Estimate for the norm of the generalized inverse.*) *Let $A \in l_b(X, Y)$ and $\bar{A} \in l_b(\bar{X}, \bar{Y})$. If the conditions* (I)–(V) *are satisfied and*

$$r = [\epsilon(1 + \eta_1\|M^+\|)\|J_0^{-1}\|\|\bar{A}^+\| + \eta_1\|M^+\|(1 + \|J_0^{-1}\bar{A}^+HA\|)] < 1,$$

then

$$\|A^+\| \leqslant (1 - r)^{-1}\,[\|J_0^{-1}\bar{A}^+H\| + \|M^+\|(1 + \epsilon\|J_0^{-1}\|\|\bar{A}^+\| + \|J_0^{-1}\bar{A}^+HA\|)].$$

Proof. Take $\hat{x} \in X$, $\hat{x} \notin N(A)$. Then $x^* = P_{N(A)^c}\hat{x}$ is clearly the best approximate solution of the equation

$$Ax = A\hat{x}\,. \tag{60}$$

By condition (V), there exists a $u \in N(M)^c \cap \widetilde{X}$ such that

$$\|Mu - P_{R(M)}N(-P_{N(A)^c}x)\| \leqslant \eta_1\|P_{N(A)^c}\hat{x}\|. \tag{61}$$

Now

$$\begin{aligned}\|x^* - u\| &= \|P_{N(A)^c}\hat{x} - u\|\\ &= \|M^+MP_{N(A)^c}\hat{x} - M^+Mu\|, \quad \text{by condition (I)}\\ &\leqslant (\|MP_{N(A)^c}\hat{x} + NP_{N(A)^c}\hat{x}\| + \|P_{R(M)}NP_{N(A)^c}\hat{x} + Mu\|)\|M^+\|\\ &\leqslant \left(\eta_1 + \frac{\|Ax^*\|}{\|P_{N(A)^c}\hat{x}\|}\right)\|P_{N(A)^c}\hat{x}\|\,\|M^+\|, \quad \text{by (61).}\end{aligned}$$

Apply Remark 5 to the exact equation (60) and its approximate equation $\bar{A}\bar{x} = HA\hat{x}$, with

$$c = \min\{1, \|M^+\|(\eta_1 + \|A\hat{x}\|/\|P_{N(A)^c}\hat{x}\|)\} \quad \text{in (49).}$$

Then

$$\|x^* - J_0^{-1}\bar{x}^*\| = \|P_{N(A)^c}\hat{x} - J_0^{-1}\bar{A}^+HA\hat{x}\|$$

$$\leqslant \left\{\epsilon\left[1 + \|M^+\|\left(\eta_1 + \frac{\|A\hat{x}\|}{\|P_{N(A)^c}\hat{x}\|}\right)\right]\|J_0^{-1}\|\|A^+\|\right.$$

$$\left. + \|M^+\|\left(\eta_1 + \frac{\|A\hat{x}\|}{\|P_{N(A)^c}\hat{x}\|}\right)(1 + \|J_0^{-1}\bar{A}^+HA\|)\right\}\|P_{N(A)^c}\hat{x}\|,$$

by (48) and (49)

$$\leqslant [\epsilon(1 + \eta_1\|M^+\|)\|J_0^{-1}\|\|\bar{A}^+\| + \eta_1\|M^+\|(1 + \|J_0^{-1}\bar{A}^+HA\|)]\|P_{N(A)^c}\hat{x}\| \tag{62}$$

$$+ \|M^+\|(1 + \epsilon\|J_0^{-1}\|\|\bar{A}^+\| + \|J_0^{-1}\bar{A}^+HA\|)\|A\hat{x}\|.$$

Now

$$\|P_{N(A)^c}\hat{x}\| \leqslant \|P_{N(A)^c}\hat{x} - J_0^{-1}\bar{A}^+HA\hat{x}\| + \|J_0^{-1}\bar{A}^+H\|\|A\hat{x}\|,$$

by the triangle inequality

$$\leqslant r\|P_{N(A)^c}\hat{x}\|$$

$$+ [\|J_0^{-1}\bar{A}^+H\| + \|M^+\|(1 + \epsilon\|J_0^{-1}\|\|\bar{A}^+\| + \|J_0^{-1}\bar{A}^+HA\|)]\|A\hat{x}\|,$$

by (62) and the definition of r.

Hence

$$\|A\hat{x}\| \geqslant \frac{1 - r}{\|J_0^{-1}\bar{A}^+H\| + \|M^+\|(1 + \epsilon\|J_0^{-1}\|\|\bar{A}^+\| + \|J_0^{-1}\bar{A}^+HA\|)}\|P_{N(A)^c}\hat{x}\|$$

$$= t\|P_{N(A)^c}\hat{x}\|,$$

where t denotes the coefficient of $\|P_{N(A)^c}\hat{x}\|$. Take an arbitrary $0 \neq y \in R(A)$. Then there exists an $\hat{x} \in X$, $\hat{x} \notin N(A)$ such that $y = A\hat{x}$. Hence $A^+y = P_{N(A)^c}\hat{x}$. Furthermore, by the above inequality,

$$\|A^+y\| = \|P_{N(A)^c}\hat{x}\| \leqslant t^{-1}\|A\hat{x}\| = t^{-1}\|y\|,$$

which gives the desired estimate for A^+. Note that here $t > 0$, since $r < 1$. If $y = 0$, the above inequality is trivially satisfied. □

Remark 10. If $A \in l_b(X)$ and in addition $X = Y$, $\bar{X} = \bar{Y} = \widetilde{X} = \widetilde{Y}$, then the above result reduces to the bound for a left inverse of A given in [17, p. 550].

6. Galerkin's Method for Best Approximate Solutions. In this section we will use Kantorovich's theory to prove that a Galerkin type method, when applied to a certain kind of, possibly inconsistent, operator equation, produces the best approximate solution. This solution is obtained as the limit of a sequence of best approximate solutions of, possibly inconsistent, systems of linear algebraic equations. In the case of Hilbert space, another method is suggested by Nashed [26]. Unlike our approach he finds the best least squares solution by applying Galerkin's method, with a suitably chosen basis, to the *consistent* equations $A^*Ax = A^*b$ and $Ax = P_{R(A)}b$, rather than to $Ax = b$.

Consider an equation $Ax = b$ in a separable Banach space X, where $A \in l_b(X)$ and $b \in X$, not necessarily $b \in R(A)$, are given. We assume that $A = I + N$, where N is compact (which implies that $N(A)$ is finite dimensional) and $R(A)$ is closed. Further we assume that $R(A) = N(A)^c$. Denote by $\{\varphi_i: i = 1, \ldots, m\}$ a basis of $N(A)$, $m = \dim N(A)$, and by $\{\psi_i: i = 1, 2, \ldots\}$ a basis of $R(A)$. It is assumed that $R(A)$ has a countable basis. Then every $x \in X$ can be written as

$$x = \sum_{i=1}^{m} c_i(x)\varphi_i + \sum_{i=1}^{\infty} d_i(x)\psi_i,$$

where $c_i(x)$, $i = 1, \ldots, m$, and $d_i(x)$, $i = 1, 2, \ldots$, are some coefficients which depend on x. The above situation occurs, for instance, in $X = C[0, 1]$ with A a Fredholm integral operator of the second kind with a continuous and symmetric kernel. The best approximate solution of such an equation can be calculated by Galerkin's method as follows: For sufficiently large n solve the system of n linear algebraic equations in n unknowns.

$$\sum_{j=1}^{n} d_i(A\psi_j)\xi_j = d_i(b), \qquad i = 1, \ldots, n. \tag{63}$$

We will show that, for sufficiently large n, the system (63) is consistent, and that the sequence of solutions $\bar{x} = (\xi_j)$ converges to the best approximate solution of $Ax = b$, with respect to $P_{N(A)^c} = P_{R(A)}$ and $P_{R(A)}$, when $n \longrightarrow \infty$. (Note that in this situation both A and A^+ leave $R(A)$ invariant.) In order to prove the consistency of (63), for large n, we will use a result from Krasnosel'skiĭ et al. [18, p. 212] which is stated here as the following lemma.

LEMMA 2. *Let $T \in l_b(X)$ be compact and let $\{P_n: n = 1, 2, \ldots\}$ be a sequence of projections in $l_b(X)$, where X is a Banach space. If $P_n \longrightarrow I$ strongly, i.e., for every $x \in X$,*

$$\|P_n x - x\| \longrightarrow 0 \quad \text{as } n \longrightarrow \infty,$$

then $\|(I - P_n)T\| \longrightarrow 0$ as $n \longrightarrow \infty$.

In our situation we specify

$$Y = X, \quad \tilde{Y} = \tilde{X} = \operatorname{span}\{\psi_1, \ldots, \psi_n\} \quad \text{and} \quad P_{\tilde{Y}} = P_{\tilde{X}},$$

which is defined by

$$P_{\tilde{X}}x = P_{\tilde{X}}\left(\sum_{i=1}^{m} c_i(x)\varphi_i + \sum_{i=1}^{\infty} d_i(x)\psi_i\right) = \sum_{i=1}^{n} d_i(x)\psi_i.$$

Further, J and H are defined by

$$Jx = Hx = \begin{pmatrix} d_1(x) \\ \vdots \\ d_n(x) \end{pmatrix},$$

while

$$J_0 = H_0 = J|_{\tilde{X}}, \quad \bar{b} = Hb \quad \text{and} \quad \bar{A} = HAJ_0^{-1}.$$

Note that $P_{\widetilde{X}} = H_0^{-1}H$ and $\overline{X} = \overline{Y}$ is the space of all n-tuples. The norm in $\overline{X}$ and $\overline{Y}$ is defined by

$$(64) \qquad \|\overline{x}\| = \|Jx\| = \sup_{k=1,\dots,n} \left\| \sum_{i=1}^{k} d_i(x)\psi_i \right\|.$$

We will first show that the system (63) is consistent for large n and then that all conditions of Theorem 5 are satisfied.

Matrix $\overline{A} = (\overline{a}_{ij})$, $\overline{a}_{ij} = d_i(A\psi_j)$, $i, j = 1, \dots, n$, has the inverse if, and only if, $P_{\widetilde{X}}A|_{\widetilde{X}}$ has the inverse. By Lemma 2, where $X = R(A)$, $I = I|_{R(A)}$, $P_n = P_{\widetilde{X}}|_{R(A)}$ and $T = N|_{R(A)}$,

$$\|(I|_{R(A)} - P_{\widetilde{X}}|_{R(A)})N|_{R(A)}\| = \|(P_{\widetilde{X}} - I)P_{R(A)}N|_{R(A)}\| \longrightarrow 0$$

as $n \longrightarrow \infty$. Since $(A|_{R(A)})^{-1}$ is bounded (by the assumptions on A), this further implies that

$$\|(P_{\widetilde{X}} - I)P_{R(A)}N|_{R(A)}\| \|(A|_{R(A)})^{-1}\| < 1$$

for sufficiently large n. Now, by specifying $X = Y = R(A)$,

$$M = A|_{R(A)} \quad \text{and} \quad N = (P_{\widetilde{X}} - I)P_{R(A)}N|_{R(A)}$$

in Proposition 1, we conclude that $M + N$ is invertible, which is here

$$\begin{aligned} A|_{R(A)} + (P_{\widetilde{X}} - I)P_{R(A)}N|_{R(A)} &= I|_{R(A)} + P_{\widetilde{X}}N|_{R(A)} \\ &= P_{\widetilde{X}}A|_{\widetilde{X}} \quad \text{when restricted to } \widetilde{X}. \end{aligned}$$

Therefore, $\overline{A}$ is invertible, which implies that the system (63) is consistent for large n.

Let us now show that all assumptions of Theorem 5 are satisfied for sufficiently large n.

Condition I. Since $M = I$, this condition is obviously satisfied.

Condition II. We know that $\overline{A}$ is invertible, so this condition, for large n, reduces to $J_0: \widetilde{X} \longrightarrow \overline{X}$, which is always satisfied.

Condition III. Since $\overline{A}$ is invertible, the condition becomes $\overline{b} = HP_{R(A)}b$, which is satisfied by our construction of H and $\overline{b}$.

Condition IV. One can specify $\epsilon = 0$, because $\overline{A} = HAJ_0^{-1}$.

Condition V. For an arbitrary $x \in N(A)^c$, take $u = P_{\widetilde{X}}Nx$. Then

$$\begin{aligned} \|Mu - P_{R(M)}Nx\| &= \|(P_{\widetilde{X}} - P_{R(A)})Nx\|, \quad \text{since } M = I \text{ and } N(A)^c = R(A) \\ &\leqslant \|(P_{\widetilde{X}} - P_{R(A)})N|_{R(A)}\| \, \|x\|. \end{aligned}$$

So, one can specify $\eta_1 = \|(P_{\widetilde{X}} - P_{R(A)})N|_{R(A)}\|$.

Condition VI. Take $v = P_{\widetilde{X}}P_{R(A)}b$. Then

$$\|Mv - P_{R(M)}P_{R(A)}b\| = \|(P_{\widetilde{X}} - P_{R(A)})P_{R(A)}b\|.$$

Therefore, one can choose

$$\eta_2 = \frac{1}{\|P_{R(A)}b\|}\|(P_{\widetilde{X}} - P_{R(A)})P_{R(A)}b\|.$$

If $P_{R(A)}b = 0$, then set $\eta_2 = 0$.

Condition A. Since $M = I$ and $\widetilde{X} = \widetilde{Y}$, the condition is satisfied.

Condition B. By our construction of H and H_0, $H_0^{-1}H = P_{\widetilde{X}}$. Since $\widetilde{X} \subset R(A)$, one concludes that $H_0^{-1}HR(A) \subset R(A)$.

Condition C. $J(N(M)^c) = \overline{X}$, which is closed.

Condition D. Since $R(\overline{A}) = \overline{X}$, this condition is satisfied by the construction of H.

Condition E. For large n, $\overline{A}$ is invertible and $N(\overline{A})^c = \overline{X}$, so the condition is satisfied.

In order to prove conditions (ii) of Theorem 5 we proceed as follows: Define a linear mapping T from $R(A)$ into the space of sequences $Tx = (d_i(x))$, $i = 1, 2, \ldots$, such that $\Sigma_{i=1}^{\infty} d_i(x)\psi_i$ is an element in $R(A)$. It is shown in [21, p. 135] that T is a linear bijection and that T is bounded and has the bounded inverse T^{-1}, if the norm in the sequence space is defined by

$$\|Tx\| = \sup_k \left\| \sum_{i=1}^{k} d_i(x)\psi_i \right\|. \tag{65}$$

Since $\|J_0x\| \leq \|Tx\|$ for every $x \in \widetilde{X} \subset R(A)$, where the norms are taken as in (64) and (65), respectively, one concludes that $\|J_0\| \leq \|T\| < \infty$, regardless of n. Hence $\sup_n\|J_0\| < \infty$. Space $\overline{X}$ is homeomorphic with the subspace of the above sequence space consisting of all sequences with zero components from $(n + 1)$st on. Therefore, $H_0^{-1}\overline{x} = T^{-1}\overline{x}$ for all $\overline{x} \in \overline{X}$. Hence, $\|H_0^{-1}\| = \|T^{-1}|_{\overline{X}}\| \leq \|T^{-1}\| < \infty$, regardless of n; and one concludes that $\sup_n\|H_0^{-1}\| < \infty$. Since $\overline{A}$ is invertible for large n, $P_{R(\overline{A})} = I$; and thus, $\sup_n\|P_{R(\overline{A})}\| < \infty$.

Finally, the conditions (iii) are satisfied, since $\epsilon = 0$, $\eta_1 \longrightarrow 0$ and $\eta_2 \longrightarrow 0$ as $n \longrightarrow \infty$, by Lemma 2, while $J_0 = H_0$; and thus $\sup_n\|J_0^{-1}\| = \sup_n\|H_0^{-1}\| < \infty$, and $\|Hx\| \leq \|Tx\|$ for every $x \in R(A)$, $Hx = 0$ for $x \in R(A)^c$, by the construction of H and T, which implies $\sup_n\|H\| \leq \|T\| < \infty$, regardless of n.

All the conditions of Theorem 5 are satisfied; and one concludes that $\lim_{n\to\infty}\|x^* - J_0^{-1}\overline{x}^*\| = 0$, where x^* is the best approximate solution of $Ax = b$; and $\overline{x}^*$ is the exact solution (for large n) of the approximate equation (63).

The best approximate solution of $Ax = b$ can also be calculated by solving systems of linear algebraic equations (63) in the case of a proper splitting $A = M + N$ if, in addition to the proper splitting, $M^+\colon \widetilde{X} \longrightarrow \widetilde{X}$ for sufficiently large n. All the conditions of Theorem 5 are still satisfied. The only modification is that u and v in Conditions V and VI are taken as follows: $u = M^+P_{\widetilde{X}}Nx$ and $v = M^+P_{\widetilde{X}}P_{R(A)}b$. Here $\widetilde{X}$ is still $\mathrm{span}\{\psi_1, \ldots, \psi_n\}$ in $R(A)$. In fact, this requirement on $\widetilde{X}$ can be relaxed. One can choose $\widetilde{X} = \mathrm{span}\{\tau_1, \ldots, \tau_n\}$, where $\{\tau_1, \ldots, \tau_n\}$ is an arbitrary set of linearly independent vectors in X provided that $P_{\widetilde{X}}P_{R(A)} = P_{R(A)}P_{\widetilde{X}}$ for sufficiently large n and $\tau_1, \ldots, \tau_n, \tau_{n+1}, \ldots$ is a basis of X. However, with this arbitrary construction of $\widetilde{X}$, the system (63) may be inconsistent for sufficiently large n, in which case the best approximate solution $\overline{x}^* = \overline{A}^+\overline{b}$ is obtained. Now one can show, using Lemma 2, that

$$J_0(N(A)^c \cap \widetilde{X}) = N(\overline{A})^c \cap \overline{X} \quad \text{and} \quad H_0(N(A) \cap \widetilde{X}) = N(\overline{A}) \cap \overline{X}$$

for sufficiently large n. These relations imply that the only conditions which need verification, i.e. Conditions II and III, are also satisfied.

A Galerkin method for calculating the best approximate solution of $Ax = b$ can be formulated as follows:

(i) Find a proper splitting $A = M + N$ with N compact.

(ii) Find a basis $\{\tau_1, \tau_2, \dots\}$ of X such that

$$M^+ : \widetilde{X} \longrightarrow \widetilde{X} \quad \text{and} \quad P_{\widetilde{X}} P_{R(A)} = P_{R(A)} P_{\widetilde{X}} \tag{66}$$

for sufficiently large n, where $\widetilde{X} = \operatorname{span}\{\tau_1, \dots, \tau_n\}$.

(iii) Calculate $\bar{A} = HAJ_0^{-1}$ and $\bar{b} = Hb$. The elements of $\bar{A} = (a_{ij})$ are determined by $a_{ij} = e_i(A\tau_j)$, where $e_i(x)$ is the ith coefficient of x in the expansion $x = \Sigma_{i=1}^{\infty} e_i(x)\tau_i$, while the elements of $\bar{b} = (b_i)$ are determined by $b_i = e_i(b)$.

(iv) Calculate the best approximate solution of $\bar{A}\bar{x} = \bar{b}$, i.e. $\bar{x}^* = \bar{A}^+\bar{b}$.

If A is written as $A = I + N$, in which case we may not have a proper splitting, then the basis $\{\tau_1, \tau_2, \dots\}$ must be chosen as a basis of $R(A)$. The conditions (66) are then redundant, and $\bar{A}$ is invertible for sufficiently large n.

EXAMPLE 4. The best least squares solution of the equation $Ax = b$ from Example 1 will now be calculated using Galerkin's method.

First, the operator A can be written as $A = M + N$, where

$$Mx(s) = x(s) - \left(x(s), \sqrt{\frac{2}{\pi}} \sin s\right)\sqrt{\frac{2}{\pi}} \sin s,$$

$$Nx(s) = -\frac{1}{2}\left(x(s), \sqrt{\frac{2}{\pi}} \cos s\right)\sqrt{\frac{2}{\pi}} \cos s.$$

Here $(\cdot\,,\cdot)$ denotes the inner product in $L_2[0, \pi]$. Since M is the orthogonal projection on $(\operatorname{span}\{\sin s\})^\perp$ and $(Nx(s), \sin s) = 0$ for every $x \in L_2[0, \pi]$, one concludes that $R(N) \subset (\operatorname{span}\{\sin s\})^\perp = R(M)$. Furthermore

$$\|N\| \leqslant \left(\int_0^\pi \int_0^\pi \left(\frac{1}{\pi} \cos t \cos \xi\right)^2 dt\, d\xi\right)^{1/2} = \frac{1}{2} < 1,$$

which implies that $A = M + N$ is a proper splitting, by Corollary 1.

Second, we choose the following basis of X:

$$\sqrt{\frac{2}{\pi}} \cos s, \sqrt{\frac{2}{\pi}} \sin s, \sqrt{\frac{2}{\pi}} \sin 3s, \sqrt{\frac{2}{\pi}} \sin 5s, \dots. \tag{67}$$

The conditions (66) are now satisfied for every n.

Third, we calculate $\bar{A}$ and $\bar{b}$ for $n = 1, 2, \dots$.

$$a_{11} = \left(\sqrt{\frac{2}{\pi}} \cos s, A\left(\sqrt{\frac{2}{\pi}} \cos s\right)\right) = \frac{1}{2},$$

since $A(\cos s) = ½\cos s$

$$b_1 = \left(s, \sqrt{\frac{2}{\pi}} \cos s\right) = -2\sqrt{\frac{2}{\pi}}.$$

Thus, $\bar{A}\bar{x} = \bar{b}$ for $n = 1$ is given by $½\bar{x} = -2\sqrt{2/\pi}$, which gives $\bar{x}^* = -4\sqrt{2/\pi}$. For

$n = 2$, the system (63) is

$$\begin{pmatrix} \frac{1}{2} & 0 \\ 0 & 0 \end{pmatrix} \bar{x} = \begin{pmatrix} -2\sqrt{2/\pi} \\ \pi\sqrt{2/\pi} \end{pmatrix}$$

and its best least squares solution is

$$\bar{x}^* = \begin{pmatrix} -4\sqrt{2/\pi} \\ 0 \end{pmatrix}.$$

For $n = 3$, the system (63) becomes

$$\begin{pmatrix} \frac{1}{2} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \bar{x} = \begin{pmatrix} -2\sqrt{2/\pi} \\ \pi\sqrt{2/\pi} \\ (\pi/3)\sqrt{2/\pi} \end{pmatrix}$$

with the best least squares solution

$$\bar{x}^* = \begin{pmatrix} -4\sqrt{2/\pi} \\ 0 \\ (\pi/3)\sqrt{2/\pi} \end{pmatrix}.$$

At the nth step ($n \geqslant 3$), we obtain

$$\bar{A} = \begin{pmatrix} \frac{1}{2} & & & & 0 \\ & 0 & & & \\ & & 1 & & \\ & & & 1 & \\ & & & & \ddots & \\ 0 & & & & & 1 \end{pmatrix}, \quad \bar{b} = \sqrt{\frac{2}{\pi}} \begin{pmatrix} -2 \\ \pi \\ \pi/3 \\ \pi/5 \\ \vdots \\ \pi/(2n-3) \end{pmatrix};$$

and the best least squares solution is

$$\bar{x}^* = \sqrt{2/\pi} \begin{pmatrix} -4 \\ 0 \\ \pi/3 \\ \pi/5 \\ \vdots \\ \pi/(2n-3) \end{pmatrix}.$$

Hence

$$J_0^{-1}\bar{x}^* = \frac{2}{\pi}\left(-4\cos s + \frac{\pi}{3}\sin 3s + \frac{\pi}{5}\sin 5s + \cdots + \frac{\pi}{2n-3}\sin(2n-3)\right).$$

Since the coefficients $(x^*(s), \tau_i)$, $i = 1, \ldots, n$, of the function $x^*(s) = s - 2 \sin s - 4(\cos s)/\pi$, in the basis (67), are

$$-4\sqrt{\frac{2}{\pi}}, \quad 0, \quad \frac{\pi}{3}\sqrt{\frac{2}{\pi}}, \quad \frac{\pi}{5}\sqrt{\frac{2}{\pi}}, \quad \ldots, \quad \frac{\pi}{2n-3}\sqrt{\frac{2}{\pi}}$$

for every n, we conclude that $J_0^{-1}\overline{x}^* \longrightarrow x^*(s)$, i.e. $x^*(s)$ is the best least squares solution of $Ax = b$. The same result has been obtained in Example 1 using iteration in $L_2[0, 1]$.

Let us note that the Kantorovich theory for singular equations leads to various methods for calculating the best approximate solution, the Galerkin method described above being only one of them. By weakening our present conditions (as in Remark 8) one obtains different types of convergence schemes, e.g. based on the fact that $J_0^{-1}N(\overline{A}_n)^c \longrightarrow N(A)^c$ (possibly $J_0^{-1}N(\overline{A}_n)^c \not\subset N(A)^c$), $H_0^{-1}R(\overline{A}_n) \longrightarrow R(A)$ (possibly $H_0^{-1}R(\overline{A}_n) \not\subset R(A)$), etc.

Acknowledgement. The authors are indebted to the referee for his cogent comments.

Department of Mathematics
McGill University
Montreal, Quebec, Canada H3A 2K6

1. P. M. ANSELONE, *Collectively Compact Approximation Theory*, Prentice-Hall, Englewood Cliffs, N.J., 1971.

2. P. M. ANSELONE & R. H. MOORE, "Approximate solutions of integral and operator equations," *J. Math. Anal. Appl.*, v. 9, 1964, pp. 268–277.

3. K. E. ATKINSON, "The solution of non-unique linear integral equations," *Numer. Math.*, v. 10, 1967, pp. 117–124.

4. K. E. ATKINSON, *A Survey of Numerical Methods for the Solution of Fredholm Integral Equations of the Second Kind*, SIAM, Philadelphia, Pa., 1976.

5. A. BEN-ISRAEL, "On direct sum decompositions of Hestenes algebras," *Israel J. Math.*, v. 2, 1964, pp. 50–54.

6. A. BEN-ISRAEL, "On error bounds for generalized inverses," *SIAM J. Numer. Anal.*, v. 3, 1966, pp. 585–592.

7. A. BEN-ISRAEL, "A note on an iterative method for generalized inversion of matrices," *Math. Comp.*, v. 20, 1966, pp. 439–440.

8. A. BEN-ISRAEL & DAN COHEN, "On iterative computation of generalized inverses and associated projections," *J. SIAM Numer. Anal.*, v. 3, 1966, pp. 410–419.

9. A. BEN-ISRAEL & T. N. E. GREVILLE, *Generalized Inverses, Theory and Applications*, Interscience, New York, 1974.

10. A. BERMAN & R. J. PLEMMONS, "Cones and iterative methods for best least squares solutions of linear systems," *SIAM J. Numer. Anal.*, v. 11, 1974, pp. 145–154.

11. J. BLATHER, P. D. MORRIS & D. E. WULBERT, "Continuity of set-valued metric projection," *Math. Ann.*, v. 178, 1968, pp. 12–24.

12. C. W. GROETSCH, *Computational Theory of Generalized Inverses of Bounded Linear Operators: Representation and Approximation*, Dekker, New York, 1977.

13. Y. IKEBE, *The Galerkin Method for the Numerical Solution of Fredholm Integral Equations of the Second Kind*, Rept. CNA-S, Univ. of Texas, Austin, Texas, 1970.

14. W. J. KAMMERER & M. Z. NASHED, "On the convergence of the conjugate gradient method for singular linear operator equations," *SIAM J. Numer. Anal.*, v. 9, 1972, pp. 165–181.

15. W. J. KAMMERER & R. J. PLEMMONS, "Direct iterative methods for least squares solutions to singular operator equations," *J. Math. Anal. Appl.*, v. 49, 1975, pp. 512–526.

16. L. V. KANTOROVICH, "Functional analysis and applied mathematics," *Uspehi Mat. Nauk*, v. 3, 1948, pp. 89–195. (Russian)

17. L. V. KANTOROVICH & G. P. AKILOV, *Functional Analysis in Normed Spaces*, English transl., Pergamon Press, Oxford, 1964.

18. M. A. KRASNOSEL'SKIĬ ET AL., *Approximate Solution of Operator Equations*, Noordhoff, Groningen, 1972.

19. W. F. LANGFORD, "The generalized inverse of an unbounded linear operator with unbounded constraints," *SIAM J. Math. Anal.* (To appear.)

20. V. LOVASS-NAGY & D. L. POWERS, "On under- and over-determined initial value problems," *Internat. J. Control*, v. 19, 1974, pp. 653–656.

21. L. A. LIUSTERNIK & V. I. SOBOLEV, *Elements of Functional Analysis*, Ungar, New York, 1961.

22. R. H. MOORE & M. Z. NASHED, "Approximations to generalized inverses of linear operators," *SIAM J. Appl. Math.*, v. 27, 1974, pp. 1–16.

23. F. J. MURRAY, "On complementary manifolds and projections in spaces L_p and l_p," *Trans. Amer. Math. Soc.*, v. 41, 1937, pp. 138–152.

24. N. I. MUSKHELISHVILI, *Singular Integral Equations*, Noordhoff, Groningen, 1946.

25. M. Z. NASHED, "Generalized inverses, normal solvability, and iteration for singular operator equations," in *Nonlinear Functional Analysis and Applications* (L. B. RALL, editor), Academic Press, New York, 1971, pp. 311–359.

26. M. Z. NASHED, "Perturbations and approximations for generalized inverses and linear operator equations," in *Generalized Inverses and Applications* (M. Z. NASHED, editor), Academic Press, New York, 1976, pp. 325–396.

27. T. G. NEWMAN & P. L. ODELL, "On the concept of a $p - q$ generalized inverse of a matrix," *SIAM J. Appl. Math.*, v. 17, 1969, pp. 520–525.

28. W. V. PETRYSHYN, "On generalized inverses and on the uniform convergence of $(I - \beta K)^n$ with application to iterative methods," *J. Math. Anal. Appl.*, v. 18, 1967, pp. 417–439.

29. W. V. PETRYSHYN, "On the generalized overrelaxation method for operator equations," *Proc. Amer. Math. Soc.*, v. 14, 1963, pp. 917–924.

30. W. V. PETRYSHYN, "On the extrapolated Jacobi or simultaneous displacements method in the solution of matrix and operator equations, *Math. Comp.*, v. 19, 1965, pp. 37–56.

31. J. L. PHILLIPS, *Collocation as a Projection Method for Solving Integral and Other Operator Equations*, Thesis, Purdue University, 1969.

32. P. M. PRENTER, "Collocation method for the numerical solution of integral equations," *SIAM J. Numer. Anal.*, v. 10, 1973, pp. 570–581.

33. C. R. RAO & S. K. MITRA, *Generalized Inverse of Matrices and its Application*, Wiley, New York, 1971.

34. I. SINGER, *Bases in Banach Spaces*, I, Springer-Verlag, New York, 1970.

35. I. STAKGOLD, *Boundary Value Problems of Mathematical Physics*, Vol. I, Macmillan Series in Advanced Mathematica and Theoretical Physics, Macmillan, New York, 1969.

36. G. W. STEWART, "On the continuity of the generalized inverse," *SIAM J. Appl. Math.*, v. 17, 1969, pp. 33–45.

37. A. E. TAYLOR, *Introduction to Functional Analysis*, Wiley, New York, 1958.

38. K. S. THOMAS, "On the approximate solution of operator equations," *Numer. Math.*, v. 23, 1975, pp. 231–239.

39. R. S. VARGA, *Extensions of the Successive Overrelaxation Theory with Applications to Finite Element Approximations*, Department of Mathematics, Kent State University, Kent, Ohio.

40. R. S. VARGA, *Matrix Iterative Analysis*, Prentice-Hall, Englewood Cliffs, N.J., 1962.

41. S. ZLOBEC, "On computing the best least squares solutions in Hilbert spaces", *Rend. Circ. Mat. Palermo Ser.* II, v. 25, 1976, pp. 1–15.