JOURNAL OF APPROXIMATION THEORY 25, 233–247 (1979)

Extrapolation and the Method of Regularization for Generalized Inverses

C. W. GROETSCH AND J. T. KING*

Department of Mathematical Sciences, University of Cincinnati, Cincinnati, Ohio 45521

Communicated by Oved Shisha

Received September 26, 1977

Richardson's "extrapolation to the limit" idea is applied to the method of regularization for approximating the generalized inverse of a linear operator in Hilbert space. Uniform error bounds for successive extrapolates are derived for the case of bounded linear operators with closed range. For bounded linear operators with arbitrary range, and for densely defined closed linear operators, pointwise error bounds are derived, assuming certain "smoothness" conditions on the data.

1. INTRODUCTION

Let H_1 and H_2 be Hilbert spaces over the same field of scalars and let $T: H_1 \to H_2$ be a bounded linear operator. We denote the range, nullspace and adjoint of T by $R(T)$, $N(T)$ and T^*, respectively.

We consider, for $b \in H_2$, the operator equation

$$Tx = b. \tag{1.1}$$

One says that $u \in H_1$ is a *least squares solution* of (1.1) if

$$\inf_{x \in H_1} \| Tx - b \| = \| Tu - b \|.$$

It is not difficult to show that u is a least squares solution of (1.1) if and only if

$$T^*Tu = T^*b \tag{1.2}$$

or equivalently

$$Tu = Qb \tag{1.3}$$

where Q is the orthogonal projection of H_2 onto $\overline{R(T)}$, the closure of $R(T)$.

If $R(T)$ is closed, then $Qb \in R(T)$ for all $b \in H_2$ and hence for any $b \in H_2$ the equation (1.1) has at least one least squares solution. Let $V(b)$ denote the

* The work of this author was partially supported by NSF Grant MCS76 24433.

0021-9045/79/030233-15$02.00/0

set of all least squares solutions of (1.1). As T is a bounded linear operator, it follows that $V(b)$ is a closed convex set and hence contains a unique vector of smallest norm. Let $T^\dagger: H_2 \to H_1$ be defined by $T^\dagger b \in V(b)$ and

$$\| T^\dagger b \| = \inf_{x \in V(b)} \| x \|,$$

i.e., $T^\dagger b$ is the minimal norm least squares solution of equation (1.1). The mapping $T^\dagger$ is then a bounded linear operator which is commonly called the Moore–Penrose generalized inverse of T.

If $R(T) \neq \overline{R(T)}$, then for some $b \in H_2$, $Qb \notin R(T)$ and hence $V(b)$ is empty. However, if b is such that $Qb \in R(T)$ then we may proceed as above. Thus, in order to define a generalized inverse as above for operators with arbitrary range, one must restrict the domain of definition to those $b \in H_2$ for which $Qb \in R(T)$. The largest such set is

$$R(T) + R(T)^\perp = \{x + y: x \in R(T), y \in R(T)^\perp\}$$

which will henceforth be denoted by $\mathscr{D}(T^\dagger)$. Then $T^\dagger: \mathscr{D}(T^\dagger) \to H_1$ is defined exactly as above, i.e., $T^\dagger b$ is the minimal norm least squares solution of (1.1). Of course, if $R(T)$ is closed then $\mathscr{D}(T^\dagger) = H_2$ and this coincides with the previous definition. However, if $R(T)$ is not closed, then $\mathscr{D}(T^\dagger)$ is a dense proper subspace of H_2 and $T^\dagger$ is an unbounded linear operator (see e.g. [7], [3]).

Tihonov [11] has introduced the idea of approximately minimizing both the functional $\| Tx - b \|$ and the norm $\| x \|$ by minimizing the functional $f: H_1 \to \mathbb{R}$ given by

$$f(x) = \| Tx - b \|^2 + \beta \| x \|^2$$

where β is a small positive parameter. It is easy to see that this minimization problem always has a unique solution, $u(\beta)$, given by

$$u(\beta) = (T^*T + \beta I)^{-1} T^*b.$$

Therefore minimizing f is equivalent to solving

$$(T^*T + \beta I)\, u(\beta) = T^*b. \tag{1.4}$$

This approach is commonly referred to as Tihonov regularization.

A difficulty with this approach is that for β small, the problem (1.4) becomes ill conditioned and regularization of (1.1) by (1.4) requires the selection of an "optimal" β as a compromise between accuracy and conditioning. This paper is a study of the use of extrapolation in Tihonov's method to obtain greater accuracy while at the same time maintaining moderate conditioning.

Suppose we let $e(\beta) = u(\beta) - u$ where $u(\beta)$ solves (1.4) and $u = T^{\dagger}b$. Moreover, suppose there exists an integer $k \geqslant 1$ and vectors $\{w_j\}_{j=1}^{k}$ such that

$$e(\beta) = \sum_{j=1}^{k} \beta^j w_j + O(\beta^k). \tag{1.5}$$

Such an asymptotic error formula suggests the use of a standard technique in numerical analysis, namely Richardson extrapolation to the limit as $\beta \to 0$.

Let $\beta_0, \ldots, \beta_k$ distinct values of the parameter β, and choose coefficients $a_0^{(k)}, \ldots, a_k^{(k)}$ so that

$$\sum_{i=0}^{k} a_i^{(k)} = 1 \qquad \text{and} \qquad \sum_{i=0}^{k} a_i^{(k)} \beta_i^{\,j} = 0, \qquad 1 \leqslant j \leqslant k. \tag{1.6}$$

The system of equations in (1.6) is a Vandermonde system and therefore the coefficients $a_0^{(k)}, \ldots, a_k^{(k)}$ are uniquely determined. Let $\bar{\beta} = \max_{0 \leqslant i \leqslant k} \beta_i$, then it follows readily from (1.5) and (1.6) that

$$\sum_{i=0}^{k} a_i^{(k)} u(\beta_i) - u = O(\bar{\beta}^k). \tag{1.7}$$

This suggests the following approximation to $T^{\dagger}b$

$$\sum_{i=0}^{k} a_i^{(k)} (T^*T + \beta_i I)^{-1}\, T^*b = \sum_{i=0}^{k} a_i^{(k)} u(\beta_i). \tag{1.8}$$

Moreover (1.7) indicates that reasonable accuracy and conditioning may be achieved by a "moderate" choice of $\bar{\beta}$.

In this paper we analyze approximations to $T^{\dagger}b$ of the type given by (1.8). It is shown that such approximations converge to $T^{\dagger}b$ and estimates of the rate of convergence are given.

2. Operators with Closed Range

Throughout this section we assume that $T\colon H_1 \to H_2$ is a bounded linear operator with closed range and we denote $R(T^*)$ by $\mathscr{H}$. Then $\mathscr{H}$ is a closed subspace of H_1 (see e.g. [12]) and is therefore a Hilbert space. Let $\tilde{T}$ denote the restriction of T^*T to $\mathscr{H}$. Then $\tilde{T}\colon \mathscr{H} \to \mathscr{H}$ is a positive definite invertible operator (see e.g. [6], [3]). Moreover, it is not difficult to see that $T^{\dagger} = \tilde{T}^{-1}T^*$. Clearly in approximating $\tilde{T}^{-1}$ one obtains corresponding approximations to

$T^{\dagger}$. The following general result on such approximations is a consequence of the spectral radius formula and the spectral mapping theorem.

THEOREM 2.1 [3]. *If* $\{S_t(x)\}$ *is a family of continuous real valued functions on* $(0, \| T \|^2]$, *then*

$$\| T^{\dagger} - S_t(\tilde{T}) T^* \| \leqslant \sup_{\lambda \in \sigma(\tilde{T})} | \lambda S_t(\lambda) - 1 | \, \| T^{\dagger} \| \tag{2.1}$$

where $\sigma(\tilde{T})$ *is the spectrum of* $\tilde{T}$. *Consequently, if* $\lim_{t \to 0} S_t(x) = x^{-1}$ *uniformly on compact subsets of* $(0, \| T \|^2]$, *then*

$$\lim_{t \to 0} S_t(\tilde{T}) T^* = T^{\dagger}$$

where the convergence is in the operator norm.

Let $k \geqslant 0$ be an integer, and $\beta_0, \ldots, \beta_k$ distinct positive numbeıs. Choose $a_0^{(k)}, \ldots, a_k^{(k)}$ so that (1.6) is satisfied. We define

$$S_\beta{}^k(x) = \sum_{i=0}^{k} a_i^{(k)} (x + \beta_i)^{-1} \tag{2.2}$$

and we shall apply Theorem 2.1 to the operators $S_\beta{}^k(\tilde{T})$, $k = 0, 1, 2, \ldots$. We have, using (1.6),

$$xS_\beta{}^k(x) - 1 = - \sum_{i=0}^{k} a_i^{(k)} \beta_i (x + \beta_i)^{-1}$$

$$= - \prod_{i=0}^{k} (x + \beta_i)^{-1} \sum_{i=0}^{k} a_i^{(k)} \beta_i \prod_{j \neq i} (x + \beta_j). \tag{2.3}$$

LEMMA 2.1. *If* $\beta_0, \ldots, \beta_k$ *are distinct positive numbers and* $a_0^{(k)}, \ldots, a_k^{(k)}$ *satisfy* (1.6), *then*

$$\sum_{i=0}^{k} a_i^{(k)} \beta_i \prod_{j \neq i} (x + \beta_j) = \prod_{i=0}^{k} \beta_i . \tag{2.4}$$

Proof. By (1.6), $a = [a_0^{(k)}, \ldots, a_k^{(k)}]^T \in \mathbb{R}^{k+1}$ satisfies $\Lambda a = e_1$ where $e_1 = [1, 0, \ldots, 0]^T \in \mathbb{R}^{k+1}$ and $\Lambda = (\lambda_{ij})$ with $\lambda_{ij} = \beta_j^{i-1}$, $1 \leqslant i, j \leqslant k+1$. Let

$$p(x) = \sum_{i=0}^{k} a_i^{(k)} \beta_i \prod_{j \neq i} (x + \beta_j)$$

and note that for $0 \leqslant \ell \leqslant k$

$$p(-\beta_\ell) = a_\ell^{(k)} \beta_\ell \prod_{j \neq \ell} (\beta_j - \beta_\ell).$$

Therefore $p(-\beta_\ell) = \prod_{j=0}^{k} \beta_j$ if and only if

$$a_\ell^{(k)} = \prod_{j \neq \ell} \frac{\beta_j}{\beta_j - \beta_\ell}, \qquad \ell = 0, \ldots, k. \tag{2.5}$$

By Cramer's rule we have $a_\ell^{(k)} = \det \Lambda_\ell / \det \Lambda$ where Λ_ℓ is the matrix Λ with the ℓth column replaced by e_1. But

$$\begin{aligned} \det \Lambda &= \prod_{k \geqslant i > j \geqslant 0} (\beta_i - \beta_j) \\ &= \prod_{j > \ell} (\beta_j - \beta_\ell) \prod_{j < \ell} (\beta_\ell - \beta_j) \prod_{\substack{k \geqslant i > j \geqslant 0 \\ i, j \neq \ell}} (\beta_i - \beta_j) \end{aligned}$$

and therefore (setting $\beta_\ell = 0$)

$$\det \Lambda_\ell = \prod_{j > \ell} (\beta_j) \prod_{j < \ell} (-\beta_j) \prod_{\substack{k \geqslant i > j \geqslant 0 \\ i, j \neq \ell}} (\beta_i - \beta_j),$$

from which it follows that $a_\ell^{(k)}$ is given by (2.5). As p is a polynomial of degree k which agrees with $\prod_{i=0}^{k} \beta_i$ at $k + 1$ distinct points, the result follows.

It is convenient to think of β_i as $\gamma_i \beta$, where $\beta > 0$ is fixed and $\gamma_0, \ldots, \gamma_k$ are distinct positive numbers. A typical choice for γ_i is $\gamma_i = 2^{-i}$, $0 \leqslant i \leqslant k$.

THEOREM 2.2. *Suppose $\gamma_0, \ldots, \gamma_k \in (0, 1]$ are distinct and $\beta_i = \gamma_i \beta$ where $\beta > 0$. If $\{S_\beta^k(x)\}$ are defined by* (2.2), *then*

$$\| T^\dagger - S_\beta^k(\hat{T}) T^* \| \leqslant \| T^\dagger \| \beta^{k+1} \prod_{i=0}^{k} \gamma_i (\| T^\dagger \|^{-2} + \gamma_i \beta)^{-1}$$

and hence

$$\lim_{\beta \to 0} S_\beta^k(\hat{T}) T^* = T^\dagger$$

where the convergence is in the operator norm.

Proof. It is easy to see that $\lambda \geqslant \| T^\dagger \|^{-2}$ for $\lambda \in \sigma(\hat{T})$ (see e.g. [3]). The result now follows by (2.3), (2.4) and Theorem 2.1.

As a result of Theorem 2.2 it follows that for any integer $k \geqslant 0$ and distinct $\beta_0, \ldots, \beta_k$, say $\beta_i = \gamma_i \beta$ ($\beta > 0$, fixed), we have for each $b \in H_2$

$$\lim_{\beta \to 0} u_\beta^{(k)} = T^\dagger b$$

where

$$u_\beta^{(k)} \equiv u_\beta^{(k)}(\beta_0, \ldots, \beta_k) = \sum_{i=0}^{k} a_i^{(k)} u(\beta_i)$$

and

$$(T^* T + \beta_i I) u(\beta_i) = T^* b.$$

If $\gamma_i = 2^{-i}$, $0 \leqslant i \leqslant k$, and $\beta_i = \gamma_i \beta$ with $\beta > 0$ fixed we can give $u_\beta^{(j)}$ explicitly by the recursion relations

$$u_\beta^{(1)}(\beta_0, \beta_1) = \frac{2u(\beta_1) - u(\beta_0)}{2 - 1}$$

and for $2 \leqslant j \leqslant k$

$$u_\beta^{(j)}(\beta_0, ..., \beta_j) = \frac{2^j u_\beta^{(j-1)}(\beta_1, ..., \beta_j) - u_\beta^{(j-1)}(\beta_0, ..., \beta_{j-1})}{2^j - 1}. \tag{2.6}$$

We conclude this section with a pointwise relative error estimate. First we need the following.

LEMMA 2.2. *If $v \in \overline{R(T)}$, then $\| T^\dagger v \| \geqslant \| T \|^{-1} \| v \|$.*

Proof. For $v \in \overline{R(T)}$, we have $Qv = v$ and therefore

$$\| T \| \| T^\dagger v \| \geqslant \| TT^\dagger v \| = \| Qv \| = \| v \|.$$

THEOREM 2.3. *Suppose $\gamma_0, ..., \gamma_k \in (0, 1]$ are distinct and $\beta_i = \gamma_i \beta$, $\beta > 0$. If $\{S_\beta^k(x)\}$ are defined by (2.2) and $b \in H_2$, then*

$$\frac{\| T^\dagger b - S_\beta^k(\tilde{T}) T^* b \|}{\| T^\dagger b \|} \leqslant \kappa(T) \beta^{k+1} \prod_{i=0}^{k} \gamma_i (\| T^\dagger \|^{-2} + \gamma_i \beta)^{-1},$$

where $\kappa(T) = \| T \| \cdot \| T^\dagger \|$.

Proof. First note that $T^* Qb = T^* b$ and $T^\dagger Qb = T^\dagger b$. We then have

$$\begin{aligned} \|(T^\dagger - S_\beta^k(\tilde{T}) T^*)b \| &= \|(T^\dagger - S_\beta^k(\tilde{T}) T^*) Qb \| \\ &\leqslant \| T^\dagger - S_\beta^k(\tilde{T}) T^* \| \cdot \| Qb \|, \end{aligned}$$

so that

$$\frac{\|(T^\dagger - S_\beta^k(\tilde{T}) T^*)b \|}{\| T^\dagger b \|} \leqslant \frac{\| Qb \|}{\| T^\dagger Qb \|} \| T^\dagger - S_\beta^k(\tilde{T}) T^* \|.$$

But $Qb \in R(T)$, so the result follows from Lemma 2.2 and Theorem 2.2.

3. OPERATORS WITH ARBITRARY RANGE

If $R(T)$ is not closed, then $T^\dagger$ is unbounded and hence it cannot be the uniform limit of a family of bounded linear operators. Therefore for the nonclosed range case the best one can expect to obtain is pointwise convergent approximations to $T^\dagger$ and corresponding pointwise error estimates.

In this section $\tilde{T}$ will denote the restriction of T^*T to $\overline{R(T^*)}$. We denote by Q the projection of H_2 onto $\overline{R(T)}$.

In [2] (see also [3] and [8]) it is shown that if $\{S_t(x)\}$ is a family of continuous real valued functions on $(0, \| T \|^2]$ which converge pointwise to x^{-1} and if $\{xS_t(x)\}$ is uniformly bounded, then

$$T^\dagger b = \lim_{t \to 0} S_t(\tilde{T})\, T^* b$$

for each $b \in \mathscr{D}(T^\dagger)$. It is easy to see that the family of functions defined by (2.2) with $\beta_i = \gamma_i \beta$ (where $\gamma_i \in (0, 1]$) possess the required properties and hence the method of the previous section will converge pointwise to $T^\dagger$ for all $b \in \mathscr{D}(T^\dagger)$.

In this section we will obtain error bounds for certain methods of the type considered in the previous section. However we shall require certain "smoothness" assumptions on the vector $b \in \mathscr{D}(T^\dagger)$. We require throughout this section that $b \in \mathscr{D}(T^\dagger)$ satisfies $Qb \in R(TT^*)$.

We note that if $u = T^\dagger b$, then $Tu = Qb \in R(TT^*)$. Hence for some $z \in H_2$ we have $TT^*z = Tu$. Clearly we may choose $z \in N(T^*)^\perp$. But T is one-to-one on $\overline{R(T^*)} = N(T)^\perp$, and hence $u = T^*z$. Thus the assumption $Qb \in R(TT^*)$ implies that $T^\dagger b = T^*z$ for some $z \in N(T^*)^\perp$.

LEMMA 3.1. *Let $\{S_\beta{}^k(x)\}$ be defined by* (2.2) *and suppose that $z \in N(T^*)^\perp$ is such that $T^*z = T^\dagger b \equiv u$, then there exists a constant C_k such that*

$$\| z - S_\beta{}^k(\hat{T})\, Qb \| \leqslant C_k \| z \|$$

where $\hat{T}$ is the restriction of TT^* to $\overline{R(T)}$.

Proof. We have

$$z - S_\beta{}^k(\hat{T})\, Qb = z - \sum_{i=0}^{k} a_i^{(k)}(\hat{T} + \beta_i I)^{-1}\, Qb$$

$$= z - \sum_{i=0}^{k} a_i^{(k)}(\hat{T} + \beta_i I)^{-1}\, Tu.$$

But $Tu = TT^*z = \hat{T}z$ as $z \in \overline{R(T)} = N(T^*)^\perp$. Therefore

$$z - S_\beta{}^k(\hat{T})\, Qb = \left(I - \sum_{i=0}^{k} a_i^{(k)}(\hat{T} + \beta_i I)^{-1}\, \hat{T}\right) z$$

$$= \sum_{i=0}^{k} a_i^{(k)} \beta_i (\hat{T} + \beta_i I)^{-1}\, z$$

and hence

$$\| z - S_\beta{}^k(\hat{T}) Qb \| \leqslant \sum_{i=0}^{k} | a_i^{(k)} | \beta_i \|(\hat{T} + \beta_i I)^{-1} \| \| z \|$$

$$\leqslant \sum_{i=0}^{k} | a_i^{(k)} | \| z \|,$$

which completes the proof.

LEMMA 3.2. *Let $S_\beta{}^k(x)$ be defined by* (2.2) *and assume that $Qb \in R(\hat{T})$. If the error vector $e_\beta^{(k)}$ is defined by*

$$e_\beta^{(k)} = T^\dagger b - S_\beta{}^k(\hat{T}) T^* b,$$

then

$$\| e_\beta^{(k)} \|^2 \leqslant \| z - S_\beta{}^k(\hat{T}) Qb \| \| Te_\beta^{(k)} \|,$$

where $z \in N(T^*)^\perp$ satisfies $T^* z = T^\dagger b$.

Proof. We note that $T^* b = T^* Qb$. Therefore

$$S_\beta{}^k(\hat{T}) T^* b = S_\beta{}^k(\hat{T}) T^* Qb$$

and hence

$$e_\beta^{(k)} = T^* z - S_\beta{}^k(\hat{T}) T^* Qb.$$

It is easy to see that $T^* S_\beta{}^k(\hat{T}) = S_\beta{}^k(\tilde{T}) T^*$, indeed it is enough to note that

$$(\tilde{T} + \beta I) T^* = T^*(\hat{T} + \beta I)$$

and hence $T^*(\hat{T} + \beta I)^{-1} = (\tilde{T} + \beta I)^{-1} T^*$. It follows that

$$e_\beta^{(k)} = T^*(z - S_\beta{}^k(\hat{T}) Qb).$$

Therefore

$$\| e_\beta^{(k)} \|^2 = \langle T^*(z - S_\beta{}^k(\hat{T}) Qb), e_\beta^{(k)} \rangle$$

$$= \langle z - S_\beta{}^k(\hat{T}) Qb, Te_\beta^{(k)} \rangle$$

and the lemma follows.

We are now in a position to prove an error estimate with minimal assumptions on b.

THEOREM 3.1. *Suppose $Qb \in R(\hat{T})$, then*

$$\| e_\beta^{(0)} \| = \| T^\dagger b - S_\beta{}^0(\tilde{T}) T^* b \| \leqslant \beta^{1/2} \| z \|,$$

where $T^* z = T^\dagger b$, $z \in N(T^*)^\perp$.

Proof. Let $u = T^{\dagger}b$ and $u(\beta) = S_{\beta}^{0}(\hat{T})\, T^{*}b$. Since $T^{*}Tu = T^{*}b$, we have

$$(T^{*}T + \beta I)(u - u(\beta)) = \beta u.$$

It follows that

$$\| Te_{\beta}^{(0)} \|^{2} + \beta \| e_{\beta}^{(0)} \|^{2} = \beta\langle u, e_{\beta}^{(0)}\rangle = \beta\langle T^{*}z, e_{\beta}^{(0)}\rangle,$$

so that

$$\| Te_{\beta}^{(0)} \|^{2} \leqslant \beta\langle z, Te_{\beta}^{(0)}\rangle \leqslant \beta \| z \| \, \| Te_{\beta}^{(0)} \|$$

and hence

$$\| Te_{\beta}^{(0)} \| \leqslant \beta \| z \|,$$

from which the result follows by using Lemmas 3.1 and 3.2.

We now obtain estimates similar to those of Theorem 2.3, however we need to make additional "smoothness" assumptions on b.

LEMMA 3.3. *If $Qb \in R(\hat{T}^{k})$, then for distinct $\beta_0, \dots, \beta_k$ there exist coefficients $a_0^{(k)}, \dots, a_k^{(k)}$ such that*

$$\| Te_{\beta}^{(k)} \| \leqslant C_k \bar{\beta}^{k} \| z_k \|$$

where C_k is given in Lemma 3.1, $\hat{T}^{k} z_k = Qb$ and $\bar{\beta} = \max_{1 \leqslant i \leqslant k} \beta_i$.

Proof. There exist $\{z_i\}_{i=1}^{k}$ such that $\hat{T}z_1 = Qb$ and $\hat{T}z_i = z_{i-1}$ for $2 \leqslant i \leqslant k$. It is readily seen that for any $\beta > 0$,

$$(\hat{T} + \beta I)\left(\sum_{\ell=1}^{k} (-\beta)^{\ell-1} z_{\ell}\right) = Qb + (-\beta)^{k} z_k . \tag{3.3}$$

We recall that $(\hat{T} + \beta I)(u - u(\beta)) = \beta u$, so that

$$(\hat{T} + \beta I)\, T(u - u(\beta)) = \beta Tu = \beta Qb. \tag{3.4}$$

It follows from (3.3) and (3.4) that

$$(\hat{T} + \beta I)\left(T(u - u(\beta)) - \sum_{\ell=1}^{k} (-1)^{\ell-1} \beta^{\ell} z_{\ell}\right) = (-\beta)^{k+1} z_k$$

so that

$$T(u - u(\beta)) - \sum_{\ell=1}^{k} (-1)^{\ell-1} \beta^{\ell} z_{\ell} = (-\beta)^{k+1} (\hat{T} + \beta I)^{-1} z_k . \tag{3.5}$$

Choose coefficients $a_0^{(k)},..., a_k^{(k)}$ satisfying (1.6), it then follows from (3.5) that

$$Te_\beta^{(k)} = \sum_{i=0}^{k} a_i^{(k)}(-\beta_i)^{k+1}(\hat{T} + \beta_i I)^{-1} z_k .$$

Therefore

$$\| Te_\beta^{(k)} \| \leqslant \sum_{i=0}^{k} | a_i^{(k)} | \beta_i^{k+1} \|(\hat{T} + \beta_i I)^{-1} \| \cdot \| z_k \|$$

and hence

$$\| Te_\beta^{(k)} \| \leqslant \sum_{i=0}^{k} | a_i^{(k)} | \beta_i^{k} \| z_k \|,$$

and the result follows.

THEOREM 3.2. *Suppose $Qb \in R(\hat{T}^k)$ and that $\gamma_0 ,..., \gamma_k \in (0, 1]$ are distinct. If $\beta_i = \gamma_i \beta$ with $\beta > 0$, then*

$$\| e_\beta^{(k)} \| \leqslant C_k \beta^{k-1/2} \| z_k \|,$$

where $C_k = \sum_{i=0}^{k} | a_i^{(k)} |$ and $\hat{T}^k z_k = Qb$.

Proof. By virtue of (3.3) we have, for any $\beta > 0$,

$$\sum_{\ell=1}^{k} (-\beta)^{\ell-1} z_\ell = (\hat{T} + \beta I)^{-1} Qb + (-\beta)^k (\hat{T} + \beta I)^{-1} z_k$$

and hence setting β equal to β_i successively and using the definition of $a_0^{(k)},..., a_k^{(k)}$, we obtain

$$z_1 - S_\beta^{k}(\hat{T})\, Qb = \sum_{i=0}^{k} a_i^{(k)}(-\beta_i)^k (\hat{T} + \beta_i I)^{-1} z_k .$$

Therefore

$$\begin{aligned} \| z_1 - S_\beta^{k}(\hat{T})\, Qb \| &\leqslant \sum_{i=0}^{k} | a_i^{(k)} | \beta_i^{k} \|(\hat{T} + \beta_i I)^{-1} \| \cdot \| z_k \| \\ &\leqslant C_k \| z_k \| \max_{0 \leqslant i \leqslant k} \beta_i^{k-1} \end{aligned}$$

which, when combined with Lemmas 3.2 and 3.3, proves the theorem.

Our final theorem shows that if we are willing to make stronger assumptions regarding the vector b, then a somewhat better error bound can be obtained.

THEOREM 3.3. *Suppose $Qb \in R(\tilde{T}^k T)$ and that $\gamma_0, ..., \gamma_k \in (0, 1]$ are distinct. If $\beta_i = \gamma_i \beta$, with $\beta > 0$, then*

$$\| e_\beta^{(k)} \| \leqslant C_k \beta^k \| w_k \|,$$

where $C_k = \sum_{i=0}^k | a_i^{(k)} |$ and $Qb = T\tilde{T}^k w_k$.

Proof. First note that $\tilde{T}^k T = T\tilde{T}^k$ and therefore our assumption on Qb gives vectors $\{w_i\}_{i=1}^k$ such that

$$Qb = T\tilde{T}w_1 ; \qquad \tilde{T}w_i = w_{i-1}, \qquad 2 \leqslant i \leqslant k.$$

If we let $u = T^\dagger b$, then it follows that

$$Tu = Qb = T\tilde{T}w_1,$$

and hence $u = \tilde{T}w_1$. Given $\beta > 0$, denote $(\tilde{T} + \beta I)^{-1} T^* b$ by $u(\beta)$. It then follows that

$$(\tilde{T} + \beta I)(u - u(\beta)) = \beta u$$

and hence

$$\begin{aligned}(\tilde{T} + \beta I)\left(u - u(\beta) + \sum_{i=1}^k (-\beta)^i w_i\right) &= \beta u + \sum_{i=1}^k (-\beta)^i \tilde{T}w_i + \sum_{i=1}^k (-\beta)^{i+1} w_i \\ &= (-1)^k \beta^{k+1} w_k .\end{aligned}$$

Taking inner products with $u - u(\beta) + \sum_{i=1}^k (-\beta)^i w_i$, we have

$$\left\| T\left(u - u(\beta) + \sum_{i=1}^k (-\beta)^i w_i\right)\right\|^2 + \beta \left\| u - u(\beta) + \sum_{i=1}^k (-\beta)^i w_i \right\|^2 = (-1)^k \beta^{k+1} \left\langle w_k , u - u(\beta) + \sum_{i=1}^k (-\beta)^i w_i \right\rangle.$$

Therefore for any $\beta > 0$,

$$\left\| u - u(\beta) + \sum_{i=1}^k (-\beta)^i w_i \right\| \leqslant \beta^k \| w_k \|. \tag{3.6}$$

If we choose $a_0^{(k)}, ..., a_k^{(k)}$ to satisfy (1.6), then

$$e_\beta^{(k)} = u - \sum_{\ell=0}^k a_\ell^{(k)} u(\beta_\ell) = \sum_{\ell=0}^k a_\ell^{(k)} \left(u - u(\beta_\ell) + \sum_{i=1}^k (-\beta_\ell)^i w_i\right).$$

Setting $\beta = \beta_0, \ldots, \beta_\ell$ successively in (3.6) then gives

$$\| e_\beta^{(k)} \| \leqslant \sum_{\ell=0}^{k} | a_\ell^{(k)} | \, \beta_\ell{}^k \| w_k \|$$

and the result follows.

4. Unbounded Operators

Beutler and Root [1] have used the (unextrapolated) method of regularization to approximate the generalized inverse of a densely defined, closed unbounded linear operator. In this section we give a brief account of how the results on extrapolation in the previous section may be extended to the case of densely defined closed linear operators.

A generalized inverse for unbounded operators between Hilbert spaces was apparently first given by Tseng in 1949 (see [5] and [7] for discussions of the history of this topic). The generalized inverse of a densely defined closed linear operator $T: \mathscr{D}(T) \to H_2$ is the linear operator with domain

$$\mathscr{D}(T^\dagger) = R(T) + R(T)^\perp$$

defined for $b \in \mathscr{D}(T^\dagger)$ as before by $T^\dagger b = u$, where u is the minimal norm solution of the equation

$$Tu = Qb,$$

where Q is the projection of H_2 onto $\overline{R(T)}$. We note that this equation has solutions for each $b \in \mathscr{D}(T^\dagger)$ and that the set of solutions is closed and convex since T is a closed linear operator.

A careful reading of the previous section reveals that in order to extend the results to densely defined closed linear operators we need only the following facts which are required in Lemmas 3.1, 3.2 and 3.3 respectively:

$$Qb \in R(TT^*) \text{ implies } T^\dagger b = T^* z, \text{ some } z \in \overline{R(T)}; \tag{4.1}$$

$$\text{for } \beta > 0, \ (T^*T + \beta I)^{-1} T^* \subseteq T^*(TT^* + \beta I)^{-1}; \tag{4.2}$$

and

$$(TT^* + \beta I)^{-1} \text{ is a bounded linear operator.} \tag{4.3}$$

As for (4.1), if $Qb \in R(TT^*)$, then as before there is a $z \in \overline{R(T)}$ such that $T^\dagger b - T^* z \in N(T)$. But $R(T^\dagger) = \mathscr{D}(T) \cap N(T)^\perp$ (see [5], [6]) and hence $T^\dagger b = T^* z$.

It is well-known that if $A: \mathscr{D}(A) \to H_2$ is a densely defined closed linear operator, then

$$(AA^* + I)^{-1}: H_2 \to \mathscr{D}(AA^*)$$

exists as a bounded linear operator (see [10]). The assertion in (4.3) follows easily from this fact. Finally, if $b \in \mathscr{D}(T^*)$, then

$$\begin{aligned} T^*b &= T^*(TT^* + \beta I)(TT^* + \beta I)^{-1}\, b \\ &= (T^*T + \beta I)\, T^*(TT^* + \beta I)^{-1}\, b \end{aligned}$$

which establishes (4.2).

5. Numerical Illustration

As a simple illustration of the extrapolation procedure we compute the generalized inverse of the matrix

$$T = \begin{bmatrix} -1 & 0 & 1 & 2 \\ -1 & 1 & 0 & -1 \\ 0 & -1 & 1 & 3 \\ 0 & 1 & -1 & -3 \\ 1 & -1 & 0 & 1 \\ 1 & 0 & -1 & -2 \end{bmatrix}$$

The exact generalized inverse, to seven decimal places is

$$T^\dagger = \begin{bmatrix} -.1470588 & -.1764705 & .0294117 & -.0294117 & .1764705 & .1470588 \\ .0784313 & .1274509 & -.0490196 & .0490196 & -.1274509 & -.0784313 \\ .0686274 & .0490196 & .0196078 & -.0196078 & -.0490196 & -.0686274 \\ .0588235 & -.0294117 & .0882352 & -.0882352 & .0294117 & -.0588235 \end{bmatrix}$$

(see Ben Noble [9]).

Using $\beta = .1$, $\beta_i = 2^{-i}\beta$ and the Cholesky decomposition method of computing $(T^*T + \beta I)^{-1}$, the following results are obtained (correct digits are underlined).

No extrapolation:

$$\begin{bmatrix} -.\underline{14}47045 & -.\underline{17}35493 & 0.\underline{02}88447 & -.\underline{02}88447 & 0.\underline{17}35493 & 0.\underline{14}47045 \\ 0.\underline{07}69193 & 0.\underline{12}54747 & -.\underline{04}85553 & 0.\underline{04}85553 & -.\underline{12}54747 & -.\underline{07}69193 \\ 0.\underline{06}77852 & 0.\underline{04}80746 & 0.\underline{019}7105 & -.\underline{019}7105 & -.\underline{04}80746 & -.\underline{06}77852 \\ 0.\underline{058}6510 & -.\underline{029}3255 & 0.\underline{08}79765 & -.\underline{08}79765 & 0.\underline{029}3255 & -.\underline{058}6510 \end{bmatrix}$$

One extrapolation:

$$\begin{bmatrix} -.1458720 & -.1749978 & 0.0291258 & -.0291258 & 0.1749978 & 0.1458720 \\ 0.0776689 & 0.1264547 & -.0487858 & 0.0487858 & -.1264547 & -.0776689 \\ 0.0682030 & 0.0485431 & 0.0196599 & -.0196599 & -.0485431 & -.0682030 \\ 0.0587371 & -.0293685 & 0.0881057 & -.0881057 & 0.0293685 & -.0587371 \end{bmatrix}$$

Two extrapolations:

$$\begin{bmatrix} -.1470587 & -.1764704 & 0.0294117 & -.0294117 & 0.1764704 & 0.1470587 \\ 0.0784313 & 0.1274509 & -.0490195 & .0490195 & -.1274509 & -.0784313 \\ 0.0686274 & 0.0490195 & 0.0196078 & -.0197068 & -.0490195 & -.0686274 \\ 0.0588235 & -.0294117 & 0.0882352 & -.0882352 & 0.0294117 & -.0588235 \end{bmatrix}$$

After three extrapolations the approximation agrees with the true generalized inverse to seven decimal places.

These results were obtained by recomputing the Cholesky decomposition for each distinct parameter value. An iterative method for solving several linear systems which depend on a parameter is given in [4].

References

1. F. J. Beutler and W. J. Root, The operator pseudoinverse in control and systems identification, *in* "Generalized Inverses and Applications" (M. Z. Nashed Ed.), Academic Press, New York, 1976.
2. C. W. Groetsch, Representations of the generalized inverse, *J. Math. Anal. Appl.* **49** (1975), 154–157.
3. C. W. Groetsch, "Generalized Inverses of Linear Operators: Representation and Approximation," Dekker, New York, 1977.
4. J. T. King and S. M. Serbin, Computational experiments and techniques for the penalty method with extrapolation, *Math. Comp.* **32** (1978), 111–126.
5. L. J. Lardy, A series representation for the generalized inverse of a closed linear operator, *Atti Accad. Naz. Lincei Rend. Cl. Sci. Fis. Mat. Natur. Ser. VIII* **58** (1975), 151–157.
6. M. Z. Nashed, Steepest descent for singular linear operator equations, *SIAM J. Numer. Anal.* **7** (1970), 358–362.
7. M. Z. Nashed, Generalized inverses, normal solvability and iteration for singular equations, *in* "Nonlinear Functional Analysis" (L. B. Rall, Ed.), Academic Press, New York, 1971.
8. M. Z. Nashed, Perturbations and approximations for generalized inverses and linear operator equations, *in* "Generalized Inverses and Applications" (M. Z. Nashed, Ed.), Academic Press, New York, 1976.

9. B. Noble, "A Method for Computing the Generalized Inverse of a Matrix," MRC Technical Summary Report No. 644, April 1966.
10. F. Riesz and B. Sz.-Nagy, "Functional Analysis," Ungar, New York, 1956.
11. A. N. Tihonov, Regularization and incorrectly posed problems, *Soviet Math. Dokl.* 4 (1963), 1624–1627.
12. K. Yosida, "Functional Analysis," 2nd ed., Springer–Verlag, Berlin/New York, 1968.