STUDIA MATHEMATICA 147 (1) (2001)

An iterative procedure for solving the Riccati equation $A_2R - RA_1 = A_3 + RA_4R$

by

M. Thamban Nair (Madras)

Abstract. Let X_1 and X_2 be complex Banach spaces, and let $A_1 \in \mathrm{BL}(X_1)$, $A_2 \in \mathrm{BL}(X_2)$, $A_3 \in \mathrm{BL}(X_1, X_2)$ and $A_4 \in \mathrm{BL}(X_2, X_1)$. We propose an iterative procedure which is a modified form of Newton's iterations for obtaining approximations for the solution $R \in \mathrm{BL}(X_1, X_2)$ of the *Riccati equation* $A_2R - RA_1 = A_3 + RA_4R$, and show that the convergence of the method is quadratic. The advantage of the present procedure is that the conditions imposed on the operators A_1, A_2, A_3, A_4 are weaker than the corresponding conditions for Newton's iterations, considered earlier by Demmel (1987), Nair (1989) and Nair (1990) in the context of obtaining error bounds for approximate spectral elements. Also, we discuss an application of the procedure to spectral approximation under perturbations of the operator.

1. Introduction. Let X_1 and X_2 be complex Banach spaces, and let

$$A_1 : X_1 \to X_1, \quad A_2 : X_2 \to X_2, \quad A_3 : X_1 \to X_2, \quad A_4 : X_2 \to X_1$$

be bounded linear operators. We shall specify conditions on the above operators so that the *Riccati equation*

$$A_2R - RA_1 = A_3 + RA_4R$$

has a unique solution $R \in \mathrm{BL}(X_1, X_2)$, and also introduce an iterative procedure for obtaining $R_k \in \mathrm{BL}(X_1, X_2)$, $k = 1, 2, \ldots$, such that

$$\|R - (R_1 + \ldots + R_k)\| \to 0 \quad \text{as } k \to \infty,$$

and

$$\|R - (R_1 + \ldots + R_{k+1})\| \le \alpha \|R - (R_1 + \ldots + R_k)\|^2$$

for some $\alpha > 0$.

The above Riccati equation arises naturally when looking for an invariant subspace of a bounded linear operator having a specified complementary subspace. To see this let $A : X \to X$ be a bounded linear operator on a complex Banach space X and $P_0 : X \to X$ be a (bounded linear) projection

2000 *Mathematics Subject Classification*: 45B05, 47A10, 65F35, 65R20.

operator. Suppose

$$X_1 = P_0(X), \qquad X_2 = (I - P_0)(X).$$

Let $R : X_1 \to X_2$ be a bounded linear operator and $P : X \to X$ be defined by

$$Px = P_0 x + RP_0 x, \qquad x \in X.$$

Then it can be easily seen that P is a projection operator along X_2. Moreover, $P(X)$ is invariant under A if and only if R satisfies the Riccati equation

$$A_2 R - RA_1 = A_3 + RA_4 R, \tag{1.1}$$

where $A_1 : X_1 \to X_1$, $A_2 : X_2 \to X_2$, $A_3 : X_1 \to X_2$, $A_4 : X_2 \to X_1$ are given by

$$A_1 = P_0 A|_{X_1}, \quad A_2 = (I - P_0)A|_{X_2}, \quad A_3 = (I - P_0)A|_{X_1}, \quad A_4 = P_0 A|_{X_2}.$$

(For details one may see [5] or [6].) Thus the problem of finding the invariant subspace $P(X)$ is equivalent to solving the Riccati equation (1.1).

Now suppose that (S_k) is a sequence of bounded linear operators from X_1 to X_2 such that

$$\|R - S_k\| \to 0 \quad \text{as } k \to \infty,$$

where $R : X_1 \to X_2$ satisfies (1.1). Then the operators

$$P_k := P_0 + S_k P_0, \qquad k = 1, 2, \ldots,$$

are projections along X_2, and we have

$$\|P - P_k\| \to 0 \quad \text{as } k \to \infty.$$

Recall from Kato [2] that, for subspaces M_1 and M_2 of X, if we define the *distance from* M_1 *to* M_2 as

$$\operatorname{sep}(M_1, M_2) := \sup\{\operatorname{dist}(x, M_2) : x \in M_1,\ \|x\| = 1\},$$

then $\operatorname{gap}(M_1, M_2)$, the *gap* between M_1 and M_2, is given by

$$\operatorname{gap}(M_1, M_2) = \max\{\operatorname{sep}(M_1, M_2), \operatorname{sep}(M_2, M_1)\}.$$

If we set

$$M_k = P_k(X), \qquad M = P(X),$$

then it follows that

$$\operatorname{sep}(M_k, M) \le \|R - S_k\|, \qquad \operatorname{gap}(M_k, M) \le \|P - P_k\| = \|(R - S_k)P_0\|.$$

Thus the convergence of (S_k) to R implies the convergence of (M_k) to M in the sense of sep and gap.

In practical situations, what one would initially have is an *approximate spectral subspace* M_0, in the sense that M_0 is a spectral subspace of a perturbed operator A_0. In such a situation, the projection P_0 may be the spectral projection with $M_0 = P_0(X)$ associated with a *spectral set* Λ_0 of A_0.

Then, what one would like to look for is a *spectral subspace* M associated with a spectral set Λ of A.

The application of our results, with suitable conditions on M_0, does in fact give a sequence (M_k) of subspaces and a sequence (Λ_k) of closed subsets of the complex plane defined by

$$\Lambda_k = \sigma(P_0 A|_{M_k}), \qquad k = 1, 2, \ldots,$$

such that

$$\operatorname{sep}(M_k, M) \to 0, \qquad \operatorname{gap}(M_k, M) \to 0,$$
$$\operatorname{sep}(\Lambda_k, \Lambda) := \sup\{\operatorname{dist}(\lambda, \Lambda) : \lambda \in \Lambda_k\} \to 0$$

as $k \to \infty$, where M is the spectral subspace of A associated with a spectral set Λ.

In all the above discussed results, the crucial role is played by the size of the quantity

$$\varepsilon := \frac{\|A_3\| \cdot \|A_4\|}{\delta^2}.$$

Here, $\delta := \operatorname{sep}(A_1, A_2)$ is defined by

$$\operatorname{sep}(A_1, A_2) = \begin{cases} 0 & \text{if } 0 \in \sigma(T) \\ 1/\|T^{-1}\| & \text{otherwise,} \end{cases}$$

where $T : \mathrm{BL}(X_1, X_2) \to \mathrm{BL}(X_1, X_2)$ is defined by

$$T(B) = A_2 B - B A_1, \qquad B \in \mathrm{BL}(X_1, X_2).$$

It is to be remarked that iterative procedures for approximately solving Riccati equation (1.1) and the corresponding problem of approximating an invariant subspace/spectral subspace have been considered by many authors (see e.g. Stewart [9], Demmel [1], Nair [5], [6], [7]). The results in Stewart [9] and Nair [5] are based on a *Piccard-type* iteration, namely,

$$A_2 R_{k+1} - R_{k+1} A_1 = A_3 + R_k A_4 R_k, \qquad R_0 = 0, \tag{1.2}$$

and provide *linear convergence* under the assumption that $\varepsilon < 1/4$. Demmel [1] and Nair [7] use Newton's method, namely,

$$(A_2 - R_k A_4) R_{k+1} - R_{k+1}(A_1 + A_4 R_k) = A_3 - R_k A_4 R_k, \qquad R_0 = 0, \tag{1.3}$$

and obtain *quadratic convergence* whenever $\varepsilon < 1/12$. Note that the quadratic convergence of the iteration in (1.3) is achieved not only by imposing a stronger assumption on the matrices A_1, A_2, A_3, A_4, but also by requiring to solve the equation each time with new coefficient matrices.

In [6], the author considered two iterative procedures which modify the iterations (1.2) and (1.3) and obtained improved estimates together with linear convergence if $\varepsilon < 1/4$ and quadratic convergence in the case of $\varepsilon < (\sqrt{3} - 1)/4$.

The new iterative procedure suggested in this paper is

$$A_2^{(k)} R_{k+1} - R_{k+1} A_1^{(k)} = R_k A_4 R_k, \tag{1.4}$$

where

$$A_1^{(k)} = A_1^{(k-1)} - A_4 R_k, \qquad A_2^{(k)} = A_2^{(k-1)} - R_k A_4,$$

and R_1 is the unique solution of

$$A_2^{(0)} R_1 - R_1 A_1^{(0)} = A_3 \quad \text{with} \quad A_1^{(0)} = A_1,\ A_2^{(0)} = A_2.$$

Of course, in this procedure, the coefficient matrices also have to be computed at each step of the iteration. But the advantage over existing Newton-type methods is that the conditions on the matrices A_1, A_2, A_3, A_4 are weaker. We show, in fact, that the iterative procedure (1.4) is valid for $\varepsilon < 1/4$, and it provides quadratic convergence whenever $\varepsilon < 3/16$. Observe that

$$\frac{1}{12} < \frac{\sqrt{3}-1}{4} < \frac{3}{16} < \frac{1}{4}.$$

2. Basic definitions and preliminary results. Let X, X_1 and X_2 be complex Banach spaces. We denote by $\mathrm{BL}(X_1, X_2)$ the space of all bounded linear operators from X_1 into X_2, and denote $\mathrm{BL}(X, X)$ by $\mathrm{BL}(X)$.

The following result is, by now, well known in the literature. The "if" part was proved by Rosenblum [8] for A_1, A_2 in a Banach algebra, and the "only if" part was proved by Stewart [9] while obtaining error bounds for invariant subspaces of a closed linear operator in a Hilbert space. Stewart's result was extended by Nair [4] (quoted in [5]) to Banach space operators, with an alternate proof.

PROPOSITION 2.1. *For* $A_1 \in \mathrm{BL}(X_1)$, $A_2 \in \mathrm{BL}(X_2)$, *consider the map* $T : \mathrm{BL}(X_1, X_2) \to \mathrm{BL}(X_1, X_2)$ *defined by*

$$T(B) = A_2 B - B A_1, \qquad B \in \mathrm{BL}(X_1, X_2).$$

Then $0 \notin \sigma(T)$ *if and only if* $\sigma(A_1) \cap \sigma(A_2) = \emptyset$.

For $A_1 \in \mathrm{BL}(X_1)$ and $A_2 \in \mathrm{BL}(X_2)$, define

$$\mathrm{sep}(A_1, A_2) = \begin{cases} 0 & \text{if } 0 \in \sigma(T), \\ 1/\|T^{-1}\| & \text{otherwise}, \end{cases}$$

where T is as in Proposition 2.1. Using the standard perturbation theory arguments (cf. Kato [2]), we obtain the following result (see e.g. Nair [4], [5]).

PROPOSITION 2.2. *Let* $A_1, V_1 \in \mathrm{BL}(X_1)$ *and* $A_2, V_2 \in \mathrm{BL}(X_2)$. *Then*

$$\mathrm{sep}(A_1 + V_1, A_2 + V_2) \geq \mathrm{sep}(A_1, A_2) - \|V_1\| - \|V_2\|.$$

In particular, *if*

$$\mathrm{sep}(A_1, A_2) > 0, \qquad \|V_1\| + \|V_2\| < \mathrm{sep}(A_1, A_2),$$

then the operator

$$B \mapsto (A_2 + V_2)B - B(A_1 + V_1), \qquad B \in \mathrm{BL}(X_1, X_2),$$

is invertible.

For closed subsets Λ_1 and Λ_2 of the complex plane, define

$$\mathrm{sep}(\Lambda_1, \Lambda_2) := \sup\{\mathrm{dist}(\lambda, \Lambda_2) : \lambda \in \Lambda_1\}.$$

We note that for $\omega > 0$,

$$\mathrm{sep}(\Lambda_1, \Lambda_2) < \omega \iff \Lambda_1 \subseteq \{z : \mathrm{dist}(z, \Lambda_2) < \omega\}.$$

Let $A \in \mathrm{BL}(X)$ and Λ be a *spectral set* of A, i.e., Λ is a subset of the spectrum $\sigma(A)$ of A such that both Λ and $\sigma(A) \setminus \Lambda$ are closed subsets of the complex plane. Then (cf. Taylor [10]) there exists an open subset Ω of the complex plane such that its boundary Γ consists of a finite number of simple closed contours and

$$\sigma(A) \cap (\Omega \cup \Gamma) = \Lambda.$$

The range of the *spectral projection* (cf. Kato [2], Limaye [3])

$$P_\Lambda = \frac{-1}{2\pi i} \int_\Gamma (A - zI)^{-1}\, dz$$

is called the *spectral subspace* of A associated with the spectral set Λ.

Here is a characterization of a spectral subspace which is found to be useful in spectral approximation (cf. Nair [4], [5]).

PROPOSITION 2.3. *For $A \in \mathrm{BL}(X)$, let $P \in \mathrm{BL}(X)$ be a projection whose range $M = P(X)$ is invariant under A. Then M is a spectral subspace of A if and only if*

$$\sigma(PA|_{P(X)}) \cap \sigma((I-P)A|_{(I-P)(X)}) = \emptyset,$$

and in that case the associated spectral set is $\sigma(PA|_{P(X)})$ and

$$\sigma(A) = \sigma(PA|_{P(X)}) \cup \sigma((I-P)A|_{(I-P)(X)}).$$

3. Existence of the solution. Let $A_1 \in \mathrm{BL}(X_1)$, $A_2 \in \mathrm{BL}(X_2)$, $A_3 \in \mathrm{BL}(X_1, X_2)$ and $A_4 \in \mathrm{BL}(X_2, X_1)$. Assume that

$$\delta := \mathrm{sep}(A_1, A_2) > 0.$$

Let

$$\gamma = \|A_3\|, \qquad \eta = \|A_4\|, \qquad \varepsilon = \eta\gamma/\delta^2.$$

In the following we shall make use of the function

$$g(t) = \begin{cases} 1 & \text{if } t = 0, \\ (1 - \sqrt{1-4t})/(2t) & \text{if } 0 < t \le 1/4. \end{cases}$$

It follows that $1 \le g(t) \le 2$ for $0 \le t \le 1/4$, and $s = g(t)$ satisfies the relation

$$ts^2 - s + 1 = 0.$$

THEOREM 3.1. *If $\varepsilon < 1/4$, then the Riccati equation*

$$A_2R - RA_1 = A_3 + RA_4R$$

has a unique solution $R \in \mathrm{BL}(X_1, X_2)$ which satisfies the relation

$$\|R\| \le \frac{\gamma}{\delta} g(\varepsilon).$$

Proof. By the assumption $\delta := \mathrm{sep}(A_1, A_2) > 0$, the operator T on $\mathrm{BL}(X_1, X_2)$ defined by

$$T(B) = A_2B - BA_1, \quad B \in \mathrm{BL}(X_1, X_2),$$

is invertible. Let

$$\mathcal{D} = \left\{ B \in \mathrm{BL}(X_1, X_2) : \|B\| \le \frac{\gamma}{\delta} g(\varepsilon) \right\}.$$

For $B \in \mathcal{D}$, let

$$F(B) = T^{-1}(A_3 + BA_4B).$$

Then we have

$$\begin{aligned} \|F(B)\| &\le \|T^{-1}\|(\|A_3\| + \|A_4\| \cdot \|B\|^2) \le \frac{1}{\delta}\left[\gamma + \eta\left(\frac{\gamma}{\delta} g(\varepsilon)\right)^2\right] \\ &= \frac{\gamma}{\delta}(1 + \varepsilon(g(\varepsilon))^2) = \frac{\gamma}{\delta} g(\varepsilon). \end{aligned}$$

Thus F maps the complete metric space $\mathcal{D}$ into itself. Therefore, the proof will be completed once it is shown that $F : \mathcal{D} \to \mathcal{D}$ is a contraction. For this, let $B_1, B_2 \in \mathcal{D}$ and observe that

$$\begin{aligned} F(B_1 - B_2) &= T^{-1}(B_1A_4B_1 - B_2A_4B_2) \\ &= T^{-1}[B_1A_4(B_1 - B_2) + (B_1 - B_2)A_4B_2]. \end{aligned}$$

Therefore,

$$\begin{aligned} \|F(B_1 - B_2)\| &\le \|T^{-1}\|(\|B_1A_4\| + \|A_4B_2\|)\|B_1 - B_2\| \\ &\le \frac{1}{\delta}\eta(\|B_1\| + \|B_2\|)\|B_1 - B_2\| \le 2\varepsilon g(\varepsilon)\|B_1 - B_2\|. \end{aligned}$$

Since $2\varepsilon g(\varepsilon) = 1 - \sqrt{1 - 4\varepsilon} < 1$, it follows that F is a contraction mapping on $\mathcal{D}$. ∎

4. The iterative procedure. We keep the notation of the previous sections and the assumption $\delta > 0$. Recall that then the operator T is invertible. We also observe that if $\gamma = 0$, then $R = 0$ is the unique solution of the Riccati equation $A_2R - RA_1 = A_3 + RA_4R$. Therefore, hereafter, we assume that $\gamma > 0$.

First we prove a technical result.

PROPOSITION 4.1. *Let $r_1 = \gamma/\delta$ and $\varepsilon < 1/4$. Then the relations*

$$2\eta r_k + 2\eta(r_1 + \ldots + r_k) < \delta,$$
$$r_{k+1} := \frac{\eta r_k^2}{\delta - 2\eta(r_1 + \ldots + r_k)} < \frac{r_k}{2}$$

hold iteratively.

Proof. Note that $4\eta r_1 = 4\varepsilon\delta < \delta$, so that r_2 is well defined and

$$r_2 = r_1\left(\frac{\eta r_1}{\delta - 2\eta r_1}\right) < \frac{r_1}{2}.$$

Thus the result is proved for $k = 1$. Assume the result for some $k = n - 1$, $n \in \{2, 3, \ldots\}$. Then for $k = n$ we have

$$\begin{aligned} 2\eta r_k + 2\eta(r_1 + \ldots + r_k) &< 2\eta r_1\left(\frac{1}{2^{k-1}} + 1 + \frac{1}{2} + \frac{1}{2^2} + \ldots + \frac{1}{2^{k-1}}\right) \\ &= 4\eta r_1 = 4\varepsilon\delta < \delta. \end{aligned}$$

Hence r_{k+1} is well defined and

$$r_{k+1} = \frac{\eta r_k^2}{\delta - 2\eta(r_1 + \ldots + r_k)} = r_k\left(\frac{\eta r_k}{\delta - 2\eta(r_1 + \ldots + r_k)}\right) < \frac{r_k}{2}. \ \blacksquare$$

THEOREM 4.2. *Let $\varepsilon < 1/4$, $A_1^{(0)} = A_1$, $A_2^{(0)} = A_2$ and (r_k) be the sequence of nonnegative numbers defined in Proposition 4.1. Let R_1 be the unique element in $\mathrm{BL}(X_1, X_2)$ such that*

$$A_2^{(0)}R_1 - R_1A_1^{(0)} = A_3.$$

If $\varepsilon < 1/4$, then the following hold iteratively.

(i) *If $A_1^{(k)} = A_1^{(k-1)} - A_4R_k$ and $A_2^{(k)} = A_2^{(k-1)} - R_kA_4$, then*

$$\delta_k := \mathrm{sep}(A_1^{(k)}, A_2^{(k)}) > \delta - 2\eta(r_1 + \ldots + r_k) > 0.$$

(ii) *Let R_{k+1} be the unique element in $\mathrm{BL}(X_1, X_2)$ such that*

$$A_2^{(k)}R_{k+1} - R_{k+1}A_1^{(k)} = R_kA_4R_k.$$

Then

$$\|R_{k+1}\| \le \frac{\eta}{\delta - 2\eta(r_1 + \ldots + r_k)}\|R_k\|^2 \le r_{k+1}.$$

In particular,

$$\|R_{k+1}\| \le \frac{\eta}{\delta(1 - 4\varepsilon)}\|R_k\|^2.$$

Proof. Clearly,

$$\|R_1\| = \|T^{-1}(A_3)\| \leq \frac{\gamma}{\delta}.$$

Then, by Propositions 2.2 and 4.1,

$$\delta_1 := \operatorname{sep}(A_1^{(1)}, A_2^{(1)}) \geq \operatorname{sep}(A_1^{(0)}, A_2^{(0)}) - 2\eta r_1 = \delta - 2\eta r_1 > 0.$$

Hence there exists a unique $R_2 \in \mathrm{BL}(X_1, X_2)$ such that

$$A_2^{(1)} R_2 - R_2 A_1^{(1)} = R_1 A_4 R_1,$$

and it satisfies

$$\|R_2\| \leq \frac{\|A_4\| \cdot \|R_1\|^2}{\delta_1} \leq \frac{\eta}{\delta - 2\eta r_1} \|R_1\|^2 \leq \frac{\eta r_1^2}{\delta - 2\eta r_1} = r_2.$$

Assume the result for all integers up to $k-1$, $k \geq 2$. Then for k, define

$$A_1^{(k)} = A_1^{(k-1)} - A_4 R_k, \qquad A_2^{(k)} = A_2^{(k-1)} - R_k A_4.$$

Then $\|R_k\| \leq r_k$, and by Propositions 2.2 and 4.1,

$$\delta_k := \operatorname{sep}(A_1^{(k)}, A_2^{(k)}) \geq \delta_{k-1} - 2\eta r_k \geq \delta - 2\eta(r_1 + \ldots + r_k) > 0.$$

Hence there exists a unique $R_{k+1} \in \mathrm{BL}(X_1, X_2)$ such that

$$A_2^{(k)} R_{k+1} - R_{k+1} A_1^{(k)} = R_k A_4 R_k.$$

Then

$$\begin{aligned} \|R_{k+1}\| &\leq \frac{\|A_4\| \cdot \|R_k\|^2}{\delta_k} \leq \frac{\eta}{\delta - 2\eta(r_1 + \ldots + r_k)} \|R_k\|^2 \\ &\leq \frac{\eta r_k^2}{\delta - 2\eta(r_1 + \ldots + r_k)} = r_{k+1}. \end{aligned}$$

Since $r_{j+1} < r_j/2$ for $j = 0, 1, 2, \ldots$, we have

$$2\eta(r_1 + \ldots + r_k) = 4\eta r_1 \left(1 - \frac{1}{2^k}\right) \leq 4\varepsilon\delta,$$

so that

$$\|R_{k+1}\| \leq \frac{\eta}{\delta(1 - 4\varepsilon)} \|R_k\|^2. \blacksquare$$

Next we obtain an estimate for the error $\|R - (R_1 + \ldots + R_k)\|$, and also specify an additional condition under which $\|R - (R_1 + \ldots + R_k)\| \to 0$ as $k \to \infty$. For this purpose let $R^{(0)} = R$ and for $k \in \{1, 2, \ldots\}$, let $R^{(k)} = R^{(k-1)} - R_k$, i.e.,

$$R^{(k)} = R - (R_1 + \ldots + R_k).$$

Theorem 4.3. *Let $\varepsilon < 1/4$. Then, for $k = 1, 2, \ldots$,*

$$\|R^{(k)}\| \leq \frac{\eta}{\delta(1 - 4\varepsilon)} \|R^{(k-1)}\|^2.$$

Moreover,

$$\|R^{(k)}\| \leq \frac{\gamma}{\delta} g(\varepsilon) \beta^{2^k - 1}, \qquad \beta = \frac{\varepsilon}{1 - 4\varepsilon} g(\varepsilon).$$

If, in addition, $\varepsilon < 3/16$*, then*

$$\|R^{(k)}\| \to 0 \qquad \text{as } k \to \infty.$$

Proof. First we note that the Riccati equation

$$A_2R - RA_1 = A_3 + RA_4R$$

takes the form

$$A_2^{(k)}R^{(k)} - R^{(k)}A_1^{(k)} = R_kA_4R_k + R^{(k)}A_4R^{(k)}$$

for $k = 1, 2, \ldots,$ where $A_1^{(k)}$ and $A_2^{(k)}$ are defined iteratively as in Theorem 4.2. Thus, considering the the map

$$T_k : B \mapsto A_2^{(k)}B - BA_1^{(k)}, \qquad B \in \mathrm{BL}(X_1, X_2),$$

for $k = 1, 2, \ldots,$ it follows that

$$R^{(k)} = T_k^{-1}(R_kA_4R_k + R^{(k)}A_4R^{(k)}) = R_{k+1} + T_k^{-1}(R^{(k)}A_4R^{(k)}),$$

so that

$$\begin{aligned} R^{(k)} &= R^{(k-1)} - R_k = [R_k + T_{k-1}^{-1}(R^{(k-1)}A_4R^{(k-1)})] - R_k \\ &= T_{k-1}^{-1}(R^{(k-1)}A_4R^{(k-1)}). \end{aligned}$$

Hence, from the estimate for $\mathrm{sep}(A_1^{(k)}, A_2^{(k)})$ from Theorem 4.2,

$$\|R^{(k)}\| \le \|T_{k-1}^{-1}\| \cdot \|A_4\| \cdot \|R^{(k-1)}\|^2 \le \frac{\eta}{\delta(1-4\varepsilon)}\|R^{(k-1)}\|^2.$$

Now let

$$\alpha = \frac{\eta}{\delta(1-4\varepsilon)}, \qquad r = \frac{\gamma}{\delta}g(\varepsilon).$$

Recall from Theorem 3.1 that

$$\|R^{(0)}\| = \|R\| \le \frac{\gamma}{\delta}g(\varepsilon) = r.$$

Then from the relation $\|R^{(k)}\| \le \alpha\|R^{(k-1)}\|^2$ it follows that

$$\|R^{(k)}\| \le r\beta^{2^k-1}, \qquad \beta := \alpha r = \frac{\varepsilon}{1-4\varepsilon}g(\varepsilon),$$

for $k = 0, 1, 2, \ldots$ Note that $\beta < 1$ if and only if $\varepsilon < 3/16$. ∎

Remark. Denote the sequences of approximations of R obtained by our procedure (1.4) and the Newton's iteration (1.3) by (S_k) and $(\widetilde{S}_k)$ respectively. Then the error bound obtained in Theorem 4.3 based on (1.4) can be written using the order function as

$$\|R - S_k\| = O(\beta^{2^k-1}), \qquad \beta = \frac{\varepsilon}{1-4\varepsilon}g(\varepsilon),$$

whereas the result obtained by Demmel [1] for $\varepsilon < 1/12$ based on (1.3) is (cf. Nair [6], Theorem 4.2)

$$\|R - \widetilde{S}_k\| = O(\mu^{2^k-1}), \qquad \mu = \frac{3}{2}\varepsilon g(\varepsilon).$$

We may observe that $\beta < \mu$ whenever $\varepsilon < 1/12$. Thus, our result not only requires weaker assumptions on the coefficient operators, but also the estimate obtained is better for those values of ε for which Demmel's result is applicable.

5. Application to spectral variation. Let $A : X \to X$ be a bounded linear operator on a complex Banach space X and M_0 be a closed subspace of X. We would like to impose conditions on M_0 such that it is *close to* a spectral subspace, and then obtain a sequence (M_k) of closed subspaces of X which converges to a spectral subspace in the sense of gap.

For M_0 to be close to a spectral subspace of A we assume the following.

Assumption 1. *There exists a closed subspace N_0 such that*

$$X = M_0 \oplus N_0. \tag{5.1}$$

Let $P_0 : X \to X$ be the projection onto M_0 along N_0 and let

$$\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$$

be the matrix representation of A with respect to the decomposition (5.1), i.e., if we take

$$Q_1 = P_0, \qquad Q_2 = I - P_0, \qquad X_1 = Q_1(X), \qquad X_2 = Q_2(X),$$

the operators $A_{ij} : X_j \to X_i$ are defined by

$$A_{ij}x = Q_iAx, \qquad x \in X_j.$$

Recall that

• M_0 is invariant under A if and only if $A_{21} = 0$, and

• M_0 is a spectral subspace of A if and only if $\sigma(A_{11}) \cap \sigma(A_{22}) = \emptyset$ and $A_{21} = 0$.

Assumption 2.

$$\sigma(A_{11}) \cap \sigma(A_{22}) = \emptyset. \tag{5.2}$$

This assumption is equivalent to the condition

$$\delta := \operatorname{sep}(A_{11}, A_{22}) > 0.$$

It is to be mentioned that the condition (5.2) is satisfied if, for example, A is an "approximation" of another operator A_0, and M_0 is a spectral subspace of A_0 with corresponding spectral projection P_0. To see this suppose that

$$A = A_0 + V$$

and let

$$\begin{bmatrix} A_{11}^{(0)} & A_{12}^{(0)} \\ A_{21}^{(0)} & A_{22}^{(0)} \end{bmatrix}, \qquad \begin{bmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{bmatrix}$$

be the matrix representations (w.r.t. the decomposition (5.1)) of A_0 and V respectively. Since P_0 is a spectral projection of A_0, we have

$$A_{12}^{(0)} = 0 = A_{21}^{(0)} \quad \text{and} \quad \delta_0 := \operatorname{sep}(A_{11}^{(0)}, A_{22}^{(0)}) > 0.$$

Thus

$$\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} = \begin{bmatrix} A_{11}^{(0)} + V_{11} & V_{12} \\ V_{21} & A_{22}^{(0)} + V_{22} \end{bmatrix}.$$

By Proposition 2.2, it follows that $\delta \geq \delta_0 - \|V_{11}\| - \|V_{22}\|$.

Hence Assumption 2 is satisfied if

$$\|V_{11}\| + \|V_{22}\| < \delta_0.$$

Note that if the last inequality is satisfied, then

$$\varepsilon = \frac{\|A_{21}\| \cdot \|A_{12}\|}{\delta^2} \leq \frac{\|V_{21}\| \cdot \|V_{12}\|}{(\delta_0 - \|V_{11}\| - \|V_{22}\|)^2}.$$

If we use the notation

$$A_1 = A_{11}, \qquad A_2 = A_{22}, \qquad A_3 = -A_{21}, \qquad A_4 = A_{12},$$

then the definitions of δ and ε in this section agree with those in the previous section.

Suppose $\varepsilon < 1/4$, and R and R_k are as in the last section. Define

$$P = P_0 + RP_0, \qquad M = P(X),$$

and for $k = 1, 2, \ldots$, let

$$P_k = P_0 + (R_1 + \ldots + R_k)P_0, \qquad M_k = P_k(X).$$

Then, by Theorem 4.3, taking $M_k = P_k(X)$, we have

$$\operatorname{sep}(M_k, M) \leq \frac{\gamma}{\delta} g(\varepsilon) \beta^{2^k - 1}, \qquad \operatorname{gap}(M_k, M) \leq \|P_0\| \frac{\gamma}{\delta} g(\varepsilon) \beta^{2^k - 1},$$

for $k = 1, 2, \ldots$ If, in addition, $\varepsilon < 3/16$, then

$$\operatorname{sep}(M_k, M) \to 0, \qquad \operatorname{gap}(M_k, M) \to 0 \qquad \text{as } k \to \infty.$$

It can be seen that

$$PA|_{P(X)} = P(A_{11} + A_{12}R)|_{P(X)},$$
$$(I - P)A|_{(I-P)(X)} = (I - P)(A_{22} + RA_{12})|_{(I-P)(X)},$$

so that

$$\sigma(PA|_{P(X)}) = \sigma(A_{11} + A_{12}R), \qquad \sigma((I - P)A|_{(I-P)(X)}) = \sigma(A_{22} + RA_{12}).$$

Hence, by Proposition 2.3, $M = P(X)$ is a spectral subspace of A with the associated spectral set $\Lambda := \sigma(A_{11} + A_{12}R)$ provided

$$\sigma(A_{11} + A_{12}R) \cap \sigma(A_{22} + RA_{12}) = \emptyset.$$

But this is true because

$$\operatorname{sep}(A_{11} + A_{12}R, A_{22} + RA_{12}) \geq \delta - 2\eta\|R\| \geq \delta(1 - 2\varepsilon g(\varepsilon)) > 0.$$

Next let

$$B = A_{11} + A_{12}R, \qquad B_k = A_{11} + A_{12}S_k,$$

where $S_k = R_1 + \ldots + R_k$, $k = 1, 2, \ldots$ Then we have

$$\|B - B_k\| \leq \|R - (R_1 + \ldots + R_k)\| \to 0 \qquad \text{as } k \to \infty.$$

Thus, by the upper semicontinuity of the spectrum (cf. Kato [2], Chapter IV, Remark 3.3), it follows that

$$\operatorname{sep}(\sigma(B_k), \sigma(B)) \to 0 \qquad \text{as } k \to \infty.$$

Acknowledgements. I thank the referees for many useful comments on the earlier versions of this paper.

References

[1] J. W. Demmel, *Three methods for refining invariant subspaces*, Computing 38 (1987), 43–57.

[2] T. Kato, *Perturbation Theory for Linear Operators*, 2nd ed., Springer, 1976.

[3] B. V. Limaye, *Spectral Perturbation and Approximation with Numerical Experiments*, Proc. Centre for Mathematical Analysis, Australian National Univ., Vol. 13, 1987.

[4] M. T. Nair, *Approximation and localization of eigenelements*, Ph.D. thesis, I.I.T. Bombay, 1984.

[5] —, *Approximation of spectral sets and spectral subspaces in Banach spaces*, J. Indian Math. Soc. 54 (1989), 1–14.

[6] —, *On iterative refinements for spectral sets and spectral subspaces*, Numer. Funct. Anal. Optim. 10 (1989), 1019–1037.

[7] —, *Computable error estimates for Newton's iterations for refining invariant subspaces*, Indian J. Pure Appl. Math. 21 (1990), 1049–1054.

[8] M. Rosenblum, *On the operator equation $BX - XA = Q$*, Duke Math. J. 23 (1956), 263–270.

[9] G. W. Stewart, *Error bounds for approximate invariant subspaces of closed operators*, SIAM J. Numer. Anal. 8 (1971), 796–808.

[10] A. E. Taylor, *Spectral theory of closed distributive operators*, Acta Math. 84 (1950), 189–224.

Department of Mathematics
Indian Institute of Technology Madras
Chennai 600 036, India
E-mail: mtnair@iitm.ac.in

Received July 29, 1998
Revised version February 22, 2001 (4157)