Acta Sci. Math., 43 (1981), 273—280

Contractions with spectral radius one and invariant subspaces

C. FOIAŞ, C. M. PEARCY, and B. SZ.-NAGY

1. Introduction. Let $\mathfrak{H}$ be a separable, complex Hilbert space, and $\mathscr{L}(\mathfrak{H})$ the Banach algebra of (bounded linear) operators on $\mathfrak{H}$. The purpose of this paper is to make some progress on the invariant subspace problem for contraction operators $A\in\mathscr{L}(\mathfrak{H})$ whose spectrum $\sigma(A)$ has at least one point on the unit circle $C=\{\lambda: |\lambda|=1\}$. From this point of view it does not restrict generality to ignore the unitary part of A (if any) and, by virtue of the Riesz decomposition theorem, to assume that $\sigma(A)$ is connected. More precisely, it suffices to consider operators of the following class

(P): The set of all completely nonunitary contractions A in $\mathscr{L}(\mathfrak{H})$ with connected spectrum $\sigma(A)$ containing the point 1.

We shall also have to do with the Banach algebra $H^\infty=H^\infty(D)$ of bounded holomorphic functions u on the open unit disc $D=\{\lambda\in\mathbf{C}: |\lambda|<1\}$, with supremum norm: $\|u\|_\infty=\sup_{\lambda\in D}|u(\lambda)|$. Recall that there is an H^∞-functional calculus for completely nonunitary contractions A so that the operator $u(A)$ is defined for every $u\in H^\infty$ and has various properties reflecting those of A and u. In particular, if $|u(\lambda)|<1$, on D, then $B=u(A)$ is a completely nonunitary contraction also, and we have $v(B)=(v\circ u)(A)$ for every $v\in H^\infty$. (Cf. [9], Chapter III, and in particular Theorem III. 2.1.)

We shall also need the following spectral mapping theorem, which was proved in [6] but not explicitly stated in this form:

Proposition (FM). *Suppose T is a completely nonunitary contraction whose spectrum $\sigma(T)$ contains a point z on the unit circle. Suppose u is a function in H^∞, which has a continuous extension $\hat{u}$ to $D\cup\{z\}$. Then $\hat{u}(z)\in\sigma(u(T))$.*

Also recall that a subset S of D is called *dominating for C* if

$$\sup_{\lambda\in S}|u(\lambda)|=\|u\|_\infty \quad \text{for all} \quad u\in H^\infty,$$

Received June 2, 1980.

and that these subsets S of D can be characterized by the property that almost every point of C is a non-tangential limit point of S; cf. [2]. In analogy with this characterization, we say that a subset S of D is *dominating for some subset* s of the unit circle C if almost every point of s is a non-tangential limit point of S.

Operators with rich spectrum have more chance to have invariant subspaces. In particular, it was proved in [3] that every contraction T for which $\sigma(T)\cap D$ is dominating for C has a non-trivial invariant subspace. Whether contractions with $\sigma(T)\cap D$ dominating a proper subarc of C only, also do the same, is still unknown. Nevertheless, it may be useful to know that the spectrum of every operator of class (P) can be "blown up", in a certain sense, so that it be dominating for a subarc of C.

For any operator $T\in\mathscr{L}(\mathfrak{H})$ let us denote by $\mathscr{W}(T)$ the set of operators which are weak limits of sequences of polynomials of T. Clearly, every invariant or hyperinvariant subspace for T is invariant or hyperinvariant, respectively, for every operator in $\mathscr{W}(T)$. In case T_1, T_2 are such that $T_2\in\mathscr{W}(T_1)$ and $T_1\in\mathscr{W}(T_2)$, we shall call T_1, T_2 $\mathscr{W}$-equivalent: they have the same invariant and hyperinvariant subspaces, respectively. Our main result is the following

Theorem. *For every subarc $E=E_\varepsilon=\{e^{it}: -\varepsilon/2\leqq t\leqq\varepsilon/2\}$ of C, $0<\varepsilon\leqq 2\pi$, there exists a function $g=g_\varepsilon\in H^\infty$, which maps D conformally into itself and is such that for $h=g\circ g$ and for every $A\in$(P)*

(1) $\sigma(g(A))\cap C=E$,

(2) $\sigma(h(A))\cap D$ *is dominating for the arc E, and*

(3) *in case E_ε is a proper subarc of C (i.e., if $\varepsilon<2\pi$), then A and $g(A)$, as well as $g(A)$ and $h(A)$, are $\mathscr{W}$-equivalent.*

Corollary 1. *There exists a nonconstant function $h\in H^\infty$ such that, for every operator $A\in$(P), $h(A)$ has a nontrivial invariant subspace.*

Proof. Apply (2) with $E_{2\pi}$ and the cited result of [3].

Corollary 2. *If it is true that an operator T has a nontrivial invariant subspace whenever T^2 has one, then every operator $A\in$(P) has a nontrivial invariant subspace.*

Proof. Let g and $h=g\circ g$ be the functions corresponding to E_π. Using the spectral mapping theorem and (1) we infer for $T=h(A)$ that $\sigma(T^2)\cap D=\sigma(T)^2\cap\cap D=(\sigma(T)\cap D)^2$ is dominating for $E_\pi^2=E_{2\pi}$; thus by [3] T^2 has a nontrivial invariant subspace. By assumption this implies the same for T, and by (3), for A also.

The following cunsequence is less immediate.

Corollary 3. *There exists a function $f \in H^\infty$ such that, for every $A \in \mathcal{L}(\mathfrak{H})$ of class* (P) *we have* $\sigma(f(A)) = D^-$ (the closed unit disc).

Proof. Let g be the function corresponding to E_π in the Theorem, and note that $E_\pi \subset \sigma(g(A))$ by (1). Let K be a Cantor set on E_π and let F be a continuous function mapping K onto D^- (cf. [1, Problem 4T]). By the Carleson-Rudin Theorem (cf. [8, p. 81]), there exists a function $k \in H^\infty$, which is continuous on D^- and such that $k|K = F$ and $\|k\|_\infty = \max_K |F| = 1$. Since $|g(\lambda)| < 1$ on D, the operator $T = g(A)$ is a completely nonunitary contraction in $\mathcal{L}(\mathfrak{H})$, and we have $k(T) = (k \circ g)(A)$. Since $K \subset E_\pi \subset \sigma(T)$, it follows from Proposition (FM) that $k(K) \subset \sigma(k(T))$. But we have $k(K) = F(K) = D^-$, and thus, setting $f = k \circ g$, we conclude that $D^- \subset \subset \sigma(f(A))$ ($\subset D^-$ because $\|f\|_\infty \leqq 1$). The proof is complete.

Corollary 4. *If every completely nonunitary contraction in $\mathcal{L}(\mathfrak{H})$, whose spectrum is the closed unit disc has a nontrivial hyperinvariant subspace, then every nonscalar contraction in $\mathcal{L}(\mathfrak{H})$ with spectral radius one has a nontrivial hyperinvariant subspace.*

Proof. Let A be a nonscalar contraction with spectral radius one. If either A has a unitary direct summand or $\sigma(A)$ is disconnected, then A has nontrivial hyperinvariant subspace for trivial reasons. Thus, without loss of generality we may suppose $A \in$(P). By Corollary 3, there exists $f \in H^\infty$ such that $\sigma(f(A)) = D^-$. The result now follows from the hypothesis and the fact that the commutant of A is contained in the commutant of $f(A)$.

2. A conformal map. The proofs involve some conformal maps of D and we turn now to some definitions in that area.

A bounded simply connected domain G in $\mathbf{C}$ is called a *Carathéodory domain* if its boundary ∂G coincides with the boundary of the unbounded component of $\mathbf{C} \setminus G^-$ (the bar denoting closure). One knows from [10] that a simply connected domain G in $\mathbf{C}$ is Carathéodory if and only if every Riemann mapping function g of D onto G is a sequential weak* generator for H^∞, i.e. has the property that every function $u \in H^\infty$ is the weak* limit of a sequence $\{p_n \circ g\}$ of polynomials in g (this amounts to saying that the functions $(p_n \circ g)(\lambda)$ are uniformly bounded on D and converge pointwise to $u(\lambda)$ as $n \to \infty$). Hence, from known facts about the H^∞-functional calculus (cf. [9] Theorem III. 2.1) it follows that if G is a Carathéodory domain contained in D and g is a Riemann mapping function of D onto G, then, upon setting $u(\lambda) = \lambda$, we see that every completely nonunitary contraction A in $\mathcal{L}(\mathfrak{H})$ is the limit in the weak operator topology of $\mathcal{L}(\mathfrak{H})$, of a sequence $\{p_n(g(A))\}$ of polynomials in $g(A)$. On the other hand, every function $u(\lambda) = \sum_0^\infty c_k \lambda^k$ in H^∞ is, by Fejér's theorem, the pointwise limit of the bounded sequence

$\{u_n\}$ of polynomials $u_n(\lambda)=\sum_0^\infty \left(1-\frac{n}{k+1}\right) c_k\lambda^k$; and hence $u(A)$ is the weak limit of the sequence $\{u_n(A)\}$ of polynomials of A. We infer that our A and $g(A)$ are $\mathcal{W}$-equivalent.

Now we turn to fix a subarc $E=E_\varepsilon$ of C $(0<\varepsilon\leqq 2\pi)$, centered on the point 1. We associate with E_ε the domain

$$G_\varepsilon = D\diagdown\left[K\cup\left(\bigcup_0^\infty L_n\right)\right],$$

where

$$K=\left\{re^{it}\colon\ 0\leqq r\leqq 1,\ \frac{\varepsilon}{2}\leqq t\leqq 2\pi-\frac{\varepsilon}{2}\right\},$$

$$L_n=\left\{re^{it}\colon\ \frac{2n+1}{2n+5}\leqq r\leqq\frac{2n+2}{2n+6},\ -\frac{n+1}{n+2}\frac{\varepsilon}{2}\leqq(-1)^n t\leqq\frac{\varepsilon}{2}\right\}.$$

For a sketch of G_ε see Figure 1.

Clearly, G_ε is simply connected, and its boundary ∂G_ε is formed by the subarc E_ε of C and by a path J_ε contained in D; J_ε is simple (that is, a Jordan arc) if $\varepsilon<2\pi$,

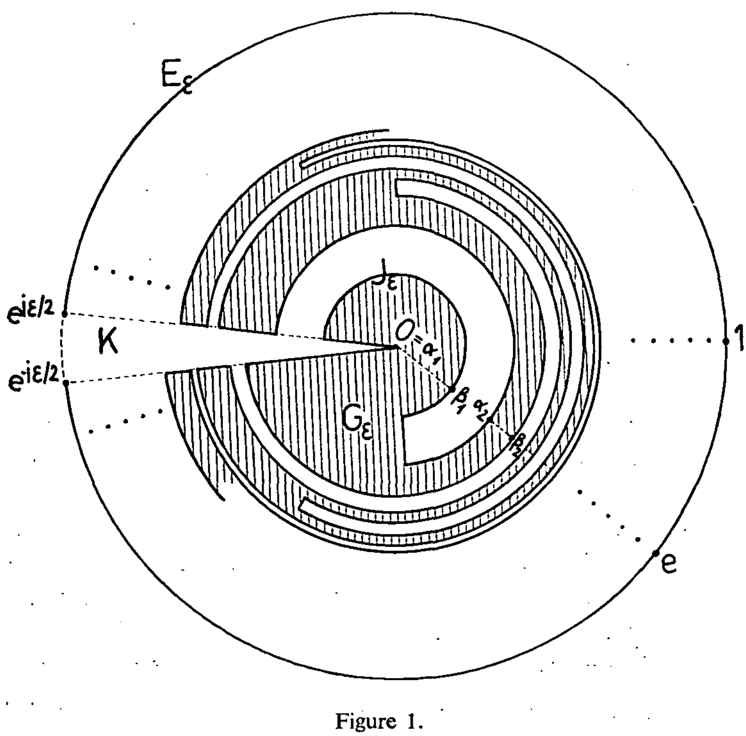

Figure 1.

and has also some overlapping segments if $\varepsilon=2\pi$. Note that if $\varepsilon<2\pi$, G_ε is a Carathéodory domain.

Let g_ε be a conformal mapping function of D onto G_ε, and let $\tilde{g}_\varepsilon$ be its Carathéodory extension to a homeomorphism of D^- onto the prime end compactification of G_ε. (See, e.g., [4], [7, p. 44], and [5].) It is no restriction of generality and so we shall assume that g_ε is normalized in such a way that the point 1 of D^- corresponds under $\tilde{g}_\varepsilon$ to that prime end $\hat{E}_\varepsilon$ of G_ε whose "impression" (see e.g. [5]) is the set E_ε, that is, the prime end determined by the sequence of crosscuts consisting of the segments

$$\left(\frac{2n}{2n+4}, \frac{2n+1}{2n+5}\right) \quad (n = 0, 1, \ldots)$$

of the real line. All the other prime ends of G_ε have one point impressions lying on the path J_ε, every point of J_ε being the impression of just one prime end (even in the case $\varepsilon=2\pi$, because we consider overlapping points of the path $J_{2\pi}$ as different ones).

Stating things slightly differently (cf. [7], pp. 40—44), we have:

a) $\tilde{g}_\varepsilon$ is a homeomorphism of $D^-\setminus 1$ onto $G_\varepsilon \cup J_\varepsilon$,

b) the set of cluster points of all sequences $g_\varepsilon(\lambda_n)$, where $\lambda_n \in D$ and $\lambda_n \to 1$, is exactly the set E_ε,

c) if a sequence $\{\lambda_n\}$ of points of $G_\varepsilon \cup J_\varepsilon$ converges to a point of E_ε then the sequence $\tilde{g}_\varepsilon^{-1}(\lambda_n)$ converges to 1.

In order to deduce one more fact let us consider a point e in the interior of E_ε. Let $l_n=(\alpha_n, \beta_n)$ $(n=1, 2, \ldots)$ be the sequence of the segments of the ray $(0, e)$ in G_ε $(|\alpha_n|<|\beta_n|)$; see Figure 1. Observe from a), b), and c) above and the geometry of the domain G_ε that the endpoints are situated on the path J_ε, at least for n large enough, in the following order:

$$(*) \qquad \ldots, \beta_{n+2}, \alpha_{n+1}, \beta_n, \alpha_{n-1}, \ldots, \beta_{n-1}, \alpha_n, \beta_{n+1}, \alpha_{n+2}, \ldots .$$

The corresponding points $a_n=\tilde{g}_\varepsilon^{-1}(\alpha_n)$, $b_n=\tilde{g}_\varepsilon^{-1}(\beta_n)$ on the open arc $C\setminus\{1\}$ must then be situated in the same order, and by virtue of property c) they must converge in both directions to 1, that is,

$$1 \leftarrow \ldots, b_{n+2}, a_{n+1}, b_n, a_{n-1}, \ldots, b_{n-1}, a_n, b_{n+1}, a_{n+2}, \ldots \to 1$$

as $n\to\infty$. The segments l_n themselves are mapped by g_ε^{-1} on disjoint open Jordan arcs $j_n=g_\varepsilon^{-1}(l_n)$ lying in D and having their endpoints a_n, b_n on C. Each of the closed arcs j_n^- dissects D^- and, again by property c), the convergence $l_n^- \to e$ implies the convergence $j_n^- \to 1$ (in the sense that every open disc centered at 1 contains j_n^- for n sufficiently large). See Figure 2.

We shall refer to the fact $j_n^- \to 1$, just established, as *property* d) of the mapping g_ε.

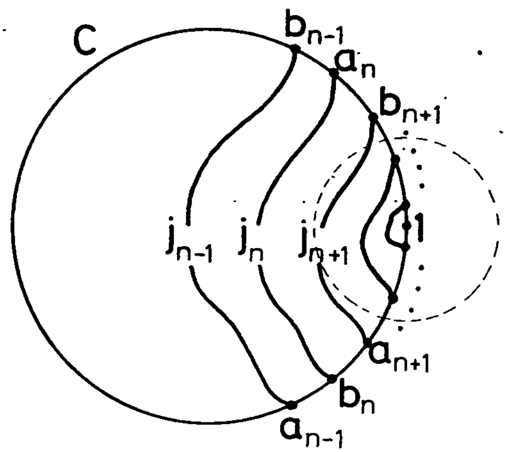

Figure 2.

3. Proof of the Theorem.

Let us consider the conformal mapping functions g_ε $(0<\varepsilon\leqq 2\pi)$ introduced above and let A be an operator of class (P). We show that $E_\varepsilon\subset\sigma(g_\varepsilon(A))$.

Suppose, to the contrary, that there is a point $e\in E_\varepsilon$ which is *not* in $\sigma(g_\varepsilon(A))$. Since $\sigma(g_\varepsilon(A))$ is compact, there is a neighborhood N of e such that $\sigma(g_\varepsilon(A))\cap\cap N=\emptyset$ and we can change e on E_ε, if necessary, so that it remains in N and be different from the endpoints of E_ε. The segments l_n on the ray $(0,e)$, considered in the preceding Section, will be contained in N, with their endpoints α_n and β_n, for n large enough, say $n\geqq n_0$, and hence $\sigma(g_\varepsilon(A))\cap l_n^-=\emptyset$. Furthermore, we may suppose that n_0 has been chosen large enough that the endpoints α_n,β_n appear in the order $(*)$ for $n>n_0$.

By virtue of [6], Corollary 3.1, we have $u(\sigma(A)\cap D)\subset\sigma(u(A))$ for every $u\in H^\infty$, so we infer that

$$g_\varepsilon(\sigma(A)\cap D)\cap l_n=\emptyset\quad(n\geqq n_0),$$

and because $g_\varepsilon^{-1}(l_n)=j_n$, it follows that

$$\sigma(A)\cap j_n=(\sigma(A)\cap D)\cap j_n=\emptyset\quad(n\geqq n_0).$$

Moreover, since $a_n,b_n\in C\diagdown\{1\}$ for all n, it follows from property a) above of g_ε that $\tilde g_\varepsilon$ is continuous at a_n and b_n, and since $\tilde g_\varepsilon(a_n)=\alpha_n$, $\tilde g_\varepsilon(b_n)=\beta_n$, we know from Proposition (FM) and the fact that $\alpha_n,\beta_n\in N$ for $n\geqq n_0$, that neither a_n nor b_n can belong to $\sigma(A)$ for such n. Thus

$$\sigma(A)\cap j_n^-=\emptyset\quad(n\geqq n_0).$$

Since $\sigma(A)$ is connected and since $j_1^-\to 1$ by property d) above, we conclude that $\sigma(A)$ consists of the single point 1.

But this implies by [9], Chapter VI, that the characteristic function $\Theta_A(\lambda)$ of A is a contractive, operator valued, analytic function on $D^-\setminus\{1\}$, unitary valued on $C\setminus\{1\}$, and, moreover, $\Theta_A(\lambda)^{-1}$ exists for every $\lambda\in D^-\setminus\{1\}$ and is an analytic function on D. From the analyticity of $\Theta_A(\lambda)^{-1}$ it follows that $\|\Theta_A(\lambda)^{-1}\|$ is subharmonic on D. Moreover, it is continuous on $D^-\setminus\{1\}$, satisfies $\|\Theta_A(\lambda)^{-1}\|\geqq \geqq\|\Theta_A(\lambda)\Theta_A(\lambda)^{-1}\|=\|I\|=1$, and is equal to 1 on $C\setminus\{1\}$.

Hence, if for $n\geqq n_0$, we denote by D_n^- the part of D^- bounded by j_n^- and that arc (a_n, b_n) on C which does not contain the point 1, we shall have

$$D_n^-\subset D_{n+1}^-\subset\dots,\quad\text{and}\quad \bigcup_{n_0}^{\infty} D_n^-=D^-\setminus\{1\},$$

For each $n\geqq n_0$, the maximum of $\|\Theta_A(\lambda)^{-1}\|$ on D_n^- will be attained for at least one point $z_n\in j_n$ (apply the maximum principle for subharmonic functions). Because $\zeta_n=g_\varepsilon(z_n)$ lies on $g_\varepsilon(j_n)=l_n$ we have $\zeta_n\to e$ as $n\to\infty$. Since $l_n^-\subset N$ for $n\geqq n_0$, we also know that, for such n, $(\zeta_n-g_\varepsilon(A))^{-1}$ exists and that $(\zeta_n-g_\varepsilon(A))^{-1}\to (e-g_\varepsilon(A))^{-1}$ as $n\to\infty$. In particular, then, there exists a positive number M such that $\|(\zeta_n-g_\varepsilon(A))^{-1}\|\leqq M$ for $n\geqq n_0$. Furthermore, we may factor $\zeta_n-g_\varepsilon(\lambda)$ as

$$\zeta_n-g_\varepsilon(\lambda)=(\lambda-z_n)(1-\bar z_n\lambda)^{-1}k_n(\lambda),\quad n\geqq n_0,$$

and it is obvious that the k_n belong to H^∞ and satisfy $\|k_n\|_\infty\leqq 2$ for all $n\geqq n_0$. Thus, from [9], Proposition VI. 4.2, we have, for $n\geqq n_0$,

$$\|\Theta_A(z_\varepsilon)^{-1}\|=\|(1-\bar z_n A)(A-z_n)^{-1}\|=\|k_n(A)(\zeta_n-g_\varepsilon(A))^{-1}\|\leqq 2M.$$

But this clearly implies, by the way the z_n were chosen, that $\|\Theta_A(\lambda)^{-1}\|$ is bounded on the open unit disc D, and that implies, in turn, by [9], Theorem IX.1.2, that A is similar to some unitary operator U. Then $\sigma(U)=\sigma(A)=\{1\}$, so U must be the identity operator, which implies the same for A. But this contradicts the fact that A is completely nonunitary.

This contradiction proves that $\sigma(g_\varepsilon(A))\supset E_\varepsilon$. Let us add that (if $\varepsilon<2\pi$) we have $\|(g_\varepsilon-a)^{-1}\|_\infty\leqq[\text{dist}\,(a, E_\varepsilon)]^{-1}$ for $a\in C\setminus E_\varepsilon$, and hence $\sigma(g_\varepsilon(A))\cap C=E_\varepsilon$.

Recall also that if $\varepsilon<2\pi$, then G_ε is a Carathéodory domain so that, in this case, $g_\varepsilon(A)$ is $\mathscr{W}$-equivalent with A.

We apply Proposition (FM) to the case $T=g_\varepsilon(A)$, $u=g_\varepsilon$, and any point $e\in E_\varepsilon\setminus 1$. This is possible since g_ε can be extended continuously to $D\cup\{e\}$ by defining $\hat g_\varepsilon(e)=\gamma$, where γ is the impression (on J_ε) of $\hat g_\varepsilon(e)$. As e runs over $E_\varepsilon\setminus\{1\}$, γ runs over J_ε so we infer by Proposition (FM) that $J_\varepsilon\subset\sigma(g_\varepsilon(g_\varepsilon(A)))$. Since J_ε obviously is dominating for E_ε, so does $\sigma(h_\varepsilon(A))\cap D$, where $h_\varepsilon=g_\varepsilon\circ g_\varepsilon$. Moreover, in case $\varepsilon<2\pi$ we know that $A\in\mathscr{W}(g_\varepsilon(A))$ and by the same reason $g_\varepsilon(A)\in \in\mathscr{W}(h_\varepsilon(A))$, and on the other hand $h_\varepsilon(A)\in\mathscr{W}(A)$, so we infer that every invariant (hyperinvariant) subspace for $h_\varepsilon(A)$ is invariant (hyperinvariant) for A, and conversely.

This concludes the proof of the Theorem.

4. Remarks.

(1) If we modify the domain G_ε by taking, say, $G_\varepsilon^* = G_\varepsilon \setminus \{re^{it}: r \leqq 1/10\}$, then the corresponding functions g_ε^* and h_ε^* will satisfy the inequalities $|1/g_\varepsilon^*| \leqq 10$, $|1/h_\varepsilon^*| \leqq 10$ on D, and these imply that $g_\varepsilon^*(A)$ and $h_\varepsilon^*(A)$ are invertible (with inverses bounded by 10). Theorem and its Corollaries obviously hold for these functions also.

(2) The techniques utilized above actually allow one to prove a fairly general spectral mapping for conformal mappings. For a statement see *Abstracts Amer. Math. Soc.*, 81T-47-427, 1981.

(3) Using the Theorem of this paper and another conformal mapping, one can prove an analog of Corollary 2 in which the square roots are replaced by inverses; for a precise statement see the same *Abstracts*, 81T-47-428, 1981.

(4) It is easy to see that the invariant subspace problem for the class of operators A in $\mathscr{L}(\mathfrak{H})$ for which some two of the numbers $r(A)$ (the spectral radius of A), $w(A)$ (the numerical radius of A), and $\|A\|$ (the norm of A) coincide reduces easily to the same problem for the smaller class for which $r(A) = \|A\|$, so the results of this paper actually apply to this larger class.

(5) It is also easy to see (*via* Cayley transforms) that the invariant subspace problem for accretive quasinilpotent operators reduces to the problem for contractions with spectral radius one.

References

[1] A. Brown and C. Pearcy, *Introduction to operator theory*. I. *Elements of functional analysis*, Springer-Verlag (New York, 1977).

[2] L. Brown, A. Shields and K. Zeller, On absolutely convergent exponential sums, *Trans. Amer. Math. Soc.*, **96** (1960), 162—183.

[3] S. Brown, B. Chevreau and C. Pearcy, Contractions with rich spectrum have invariant subspaces, *J. Operator Theory*, **1** (1979), 123—136.

[4] C. Carathéodory, Über die Begrenzung einfach zusammenhängender Gebiete, *Math. Ann.*, **73** (1913), 323—370.

[5] E. F. Collingwood and G. Piranian, The mapping theorems of Carathéodory and Lindelöf, *J. de Math.*, **43** (1964), 187—199.

[6] C. Foiaş and W. Mlak, The extended spectrum of completely non-unitary contractions and the spectral mapping theorem, *Studia Math.*, **26** (1966), 239—245.

[7] G. M. Goluzin, *Geometric theory of functions of a complex variable* (translated from the Russian), A.M.S. (Providence, 1969).

[8] P. R. Halmos, *A Hilbert space problem book*, Van Nostrand (Princeton, 1967).

[9] B. Sz.-Nagy and C. Foiaş, *Harmonic analysis of operators on Hilbert space*, North Holland (Amsterdam, 1970).

[10] D. Sarason, Weak star generators of H^∞, *Pacific J. Math.*, **17** (1966), 519—528.

(C. F.)
DEPARTMENT OF MATH.
INDIANA UNIVERSITY
BLOOMINGTON, IN 47401, USA
and
UNIVERSITÉ DE PARIS-SUD,
91405 ORSAY, FRANCE

(C. M. P.)
DEPARTMENT OF MATH.
UNIVERSITY OF MICHIGAN
ANN ARBOR, MI, 48109, USA

(B. SZ.-N.)
BOLYAI INSTITUTE
UNIVERSITY SZEGED
6720 SZEGED
HUNGARY