SQUARE ROOTS OF OPERATORS. II

PAUL R. HALMOS AND GÜNTER LUMER

Introduction. It was recently shown[1] that on an infinite-dimensional Hilbert space there exist invertible operators with no square roots. The main purpose of this paper is to show that there are many such operators. More precisely, we shall show that the operators with square roots are not even dense among the invertible operators, or, in other words, that there exists a nonempty open set consisting entirely of invertible operators with no square roots. (The topological terminology refers to the so-called uniform, or norm, topology.) The technique consists in analyzing the reasons that make for the existence of one such operator and proving that, in a sense, those reasons are sufficiently continuous to remain in force for nearby operators. The main result is obtained in the last section; the first three sections are devoted to the appropriate continuity theorems.

Semicontinuity. With every subset E of a complex Banach algebra B (with unit 1) we associate a function Λ as follows. The domain of Λ is the algebra B and the values of Λ are subsets of the complex plane; if $a \epsilon B$, then $\Lambda(a)$ is the set of all those complex numbers λ for which $a - \lambda$ ($= a - \lambda \cdot 1$) belongs to E. The dependence of Λ on E will be indicated, when necessary, by writing Λ_E for Λ. By way of mentioning an example, we observe that if S is the set of all singular ($=$ noninvertible) elements of B, then $\Lambda_S(a)$ is the spectrum of a. Since $\Lambda_E(a)$ is the inverse image of the set E with respect to the (continuous) mapping $\lambda \rightarrow a - \lambda$, it follows that if E is open (or closed), then $\Lambda_E(a)$ is open (or closed) for every a.

Our immediate task is to investigate the continuity properties of functions such as Λ. Since in all cases of interest to us the values of Λ will be either compact sets or open sets, we may formulate the appropriate definitions as follows. A function Λ from B into the set of all open subsets of the complex plane is *lower semicontinuous* if to every element a of B and to every compact set K included in $\Lambda(a)$ there corresponds a positive number δ such that $K \subset \Lambda(b)$ whenever $\|a - b\| < \delta$. A function Λ from B into the set of all compact subsets of the complex plane is *upper semicontinuous* if to every element a of

---

B and to every open set U including $\Lambda(a)$ there corresponds a positive number δ such that $\Lambda(b) \subset U$ whenever $\|a-b\| < \delta$.

In connection with compact (or, more generally, bounded) subsets of the complex plane, we record here a useful item of notation: if Γ is a bounded set of complex numbers, the symbol $|\Gamma|$ will be used to denote $\sup \{|\gamma| : \gamma \epsilon \Gamma\}$. Thus, for instance, since $\Lambda_S(a)$ is the spectrum of a, the symbol $|\Lambda_S(a)|$ denotes the spectral radius of a.

THEOREM 1. *If E is an open subset of B, then Λ_E is lower semicontinuous.*

PROOF. We have already remarked that if E is open, then $\Lambda_E(a)$ is open for every a, so that the assertion of the theorem is meaningful. (Recall that we defined lower semicontinuity for open-valued functions only.) Suppose now that a is a fixed element of B, K is a compact subset of $\Lambda_E(a)$, and λ is an element of K. It follows that $\lambda \epsilon \Lambda_E(a)$, or $a - \lambda \epsilon E$; since E is open, there exists a positive number δ such that $b \epsilon E$ whenever $\|(a-\lambda)-b\| < 2\delta$. If, in particular, $\|a-b\| < \delta$ and $|\lambda - \mu| < \delta$, so that $\|(a-\lambda)-(b-\mu)\| < 2\delta$, then $b - \mu \epsilon E$, or $\mu \epsilon \Lambda_E(b)$. In other words, if U is the open disc with center at λ and radius δ, then $U \subset \Lambda_E(b)$ whenever $\|a-b\| < \delta$. We have proved so far that for every λ in K there exists a positive number $\delta(\lambda)$ such that the open disc $U(\lambda)$ with center at λ and radius $\delta(\lambda)$ is included in $\Lambda_E(b)$ whenever $\|a-b\| < \delta(\lambda)$. The compactness of K implies the existence of a finite set $\{\lambda_1, \cdots, \lambda_n\}$ of points in K such that $K \subset U(\lambda_1) \cup \cdots \cup U(\lambda_n)$. If $\delta = \min \{\delta(\lambda_1), \cdots, \delta(\lambda_n)\}$, and if $\|a-b\| < \delta$, then $\|a-b\| < \delta(\lambda_j)$ and therefore $U(\lambda_j) \subset \Lambda_E(b)$ for all j. It follows that if $\|a-b\| < \delta$, then $K \subset \Lambda_E(b)$, i.e., that Λ_E is lower semicontinuous.

THEOREM 2. *If E is a closed subset of B such that $|\Lambda_E(a)| \leqq \|a\|$ for all a in B, then Λ_E is upper semicontinuous.*

PROOF. Since E is closed, so is $\Lambda_E(a)$ for every a in B. Our boundedness assumption implies that Λ_E is compact-valued and hence that the assertion of the theorem is meaningful. Suppose now that a is a fixed element of B and that U is an open set including $\Lambda_E(a)$. If D is the closed disc with center at the origin and radius $1 + \|a\|$, then $K = D - U$ is a compact subset of $\Lambda_{E'}(a)$ (where E' is the complement of E in B). It follows from Theorem 1 that there exists a positive number δ such that $K \subset \Lambda_{E'}(b)$ whenever $\|a-b\| < \delta$; we assume, without any loss of generality, that $\delta < 1$. We have proved that if $\|a-b\| < \delta$, then $\Lambda_E(b) \subset U \cup D'$ (where D' is the complement of D in the complex plane). Since $|\Lambda_E(b)| \leqq \|b\| \leqq \|a\| + \|a-b\| < 1 + \|a\|$

whenever $\|a-b\| < \delta$, it follows that $\Lambda_E(b) \subset D$ and hence that $\Lambda_E(b) \subset U$ whenever $\|a-b\| < \delta$. This proves the upper semicontinuity of Λ_E.[2]

Spectra. In order to apply the semicontinuity results of the preceding section, we proceed to exhibit an important closed subset and an important open subset of the Banach algebra B.

We shall say that an element a of B is a *generalized left divisor of zero* if to every positive number δ there corresponds an element b (necessarily different from 0) such that $\|ab\| < \delta\|b\|$; the set of all generalized left divisors of zero will be denoted by D.[3]

LEMMA 1. *The set D of all generalized left divisors of zero is closed.*

PROOF. If $a_0 \in D'$, then there exists a positive number δ such that $\|a_0 b\| \geq \delta\|b\|$ for all b. It follows that if $\|a_0 - a\| < \delta$, so that $\delta - \|a_0 - a\| > 0$, then

$$\|ab\| \geq \|a_0 b\| - \|(a_0 - a)b\| \geq \delta\|b\| - \|a_0 - a\| \cdot \|b\|$$
$$= (\delta - \|a_0 - a\|)\|b\|,$$

so that $a \in D'$. We have proved, in other words, that the complement of D is open.

It is well known that the set S of all singular elements of B is closed and that, in fact, if a is invertible and $\|a_0 - a\| < \|a^{-1}\|^{-1}$, then a_0 is invertible.[4] We shall make use of this fact in the proof of our next result.

LEMMA 2. *The set $S-D$ of all those singular elements which are not generalized left divisors of zero is open.*[5]

PROOF. It is sufficient to prove that $S-D$ is included in the interior S^0 of S, for then $S-D = S^0 \cap (S-D) = S^0 - D$, and the desired result

---

[2] Theorem 2 is a slight generalization of a well-known result; cf. E. Hille, *Functional analysis and semi-groups*, New York, 1948, p. 118. We shall need Theorem 1 below; we only prove Theorem 2 in order to indicate the amusing duality between what we need and what is already in the literature.

[3] In a complete treatment of this circle of ideas the set D should perhaps be denoted by D_-, in order to be able to indicate by the dual symbol D_+ the similarly defined set of all generalized right divisors of zero. The union of D_+ and D_- is usually called the set of all generalized divisors of zero; cf. Hille, op. cit. p. 94.

[4] Cf. Hille, op. cit. p. 92.

[5] It is relevant to observe that $D \subset S$, i.e., that an invertible element cannot be a generalized left divisor of zero. Indeed, if a is invertible, then $\|b\| = \|a^{-1}ab\| \leq \|a^{-1}\| \cdot \|ab\|$, and therefore $\|ab\| \geq \delta\|b\|$, with $\delta = \|a^{-1}\|^{-1}$, for all b. Lemma 2 occurs also in the work of C. E. Rickart, *The singular elements of a Banach algebra*, Duke Math. J. vol. 14 (1947) pp. 1063–1077.

follows from Lemma 1. By complementation the assertion $S-D \subseteq S^0$ becomes the assertion that the closure of the set S' of invertible operators is included in $S' \cup D$; it is therefore sufficient to prove that if a_0 is in the closure of S' but not in D, then a_0 is in S'. Since a_0 is not in D, it follows that $\|a_0 b\| \geq \delta \|b\|$ for some positive δ and all b in B. Since a_0 is in the closure of the set of invertible elements, it follows that there is an invertible element a such that $\|a_0 - a\| < \delta/2$. A simple chain of inequalities (cf. the proof of Lemma 1) implies that $\|ab\| \geq (\delta/2)\|b\|$ for all b, and hence, with $b = a^{-1}$, that $\delta/2 \leq \|a^{-1}\|^{-1}$. From the relation $\|a_0 - a\| < \|a^{-1}\|^{-1}$ we may now conclude that a_0 is invertible; the proof of the lemma is complete.[6]

The Banach algebra of greatest interest to us will be the algebra B of all (bounded) operators on a (complex) Hilbert space H. If B is such an operator algebra, then, in addition to the spectrum $\Lambda_S(a)$ and the related construct $\Lambda_D(a)$, it is also profitable to consider, for each operator a, the approximate point spectrum $\Pi(a)$. By definition a complex number λ belongs to $\Pi(a)$ if and only if to every positive number δ there corresponds a vector x of H (necessarily different from 0) such that $\|(a - \lambda)x\| < \delta \|x\|$. There is a striking similarity between this definition and the definition of a generalized left divisor of zero. There is, in fact, more than a similarity; it is easy to prove that $\Lambda_D(a) = \Pi(a)$ for every a.[7]

If, for the sake of harmony with customary notation, we write $\Sigma(a) = \Lambda_S(a)$, then the principal result that we shall need below may be formulated as follows.

THEOREM 3. *If B is the algebra of all operators on a Hilbert space, then the mapping $a \rightarrow \Sigma(a) - \Pi(a)$ is lower semicontinuous.*

PROOF. We know that $\Sigma(a) - \Pi(a) = \Lambda_S(a) - \Lambda_D(a) = \Lambda_{S-D}(a)$; the proof is completed by an application of Lemma 2 and Theorem 1.

Multiplicity. From now on we assume that the algebra B we are working with is the algebra of all operators on a Hilbert space H. If a is such an operator and if λ is a complex number, we denote by $m(a, \lambda)$ the (geometric) multiplicity of $\bar{\lambda}$ (= the complex conjugate of λ) as a proper value of a^*, i.e., $m(a, \lambda)$ is the dimension of the null space of $a^* - \bar{\lambda}$. (Thus, for example, if $\bar{\lambda}$ is not a proper value of a^*, then $m(a, \lambda) = 0$.) In the course of the discussion that follows, repeated use will be made of the fact that $m(a, \lambda) = \dim R^{\perp}(a - \lambda)$

---

[6] In both statement and proof, Lemma 2 is a slight improvement on known results; cf. Hille, op. cit. p. 108.

[7] Cf. Paul R. Halmos, *Commutators of operators*, II, Amer. J. Math. vol. 76 (1954) pp. 191–198.

(where R denotes range).

Since the values of the function m are cardinal numbers, it is not reasonable to expect it to be continuous. There do, in fact, exist operators a_0 (and even normal ones) and complex numbers λ_0, such that $m(a_0, \lambda_0) = 0$ and such that λ_0 is a limit point of numbers λ for which $m(a_0, \lambda)$ is positive, or even infinite. In other words, m is not upper semicontinuous in its second argument. Similar examples show that m is also not lower semicontinuous in its second argument, and others that m is neither upper nor lower semicontinuous in its first argument. Nevertheless, we shall show in this section that m is continuous, jointly in its two arguments, on a large and useful set of pairs (a, λ).

LEMMA 3. *If q is a bounded, linear, one-to-one transformation from a Hilbert space M into a Hilbert space N, then* dim $M \leq$ dim N.

PROOF. If dim $N <$ dim M, then there is no loss of generality in assuming that $N \subset M$, i.e., that q is an operator on the Hilbert space M such that $q(M) \subset N$; excluding trivial cases, we assume also that dim N is infinite. Let N_0 and M_0 be orthonormal bases of N and M, respectively. If $y \in N_0$, then q^*y can be expanded in terms of countably many elements of M_0; the assumed inequality between the cardinal numbers of N_0 and M_0 implies the existence of an element x_0 in M_0 such that $(x_0, q^*y) = 0$ for all y in N_0. Since $(x_0, q^*y) = (qx_0, y)$, it follows that qx_0 is orthogonal to N_0 and therefore to N. Since, however, $q(M) \subset N$, it follows that $qx_0 = 0$. We have proved that if dim $N <$ dim M, then q has a nontrivial null-space, and, therefore, is not one-to-one.[8]

LEMMA 4. *If M and N are (closed) subspaces of H such that $M \cap N^{\perp} = 0$, then* dim $M \leq$ dim N.

PROOF. If q denotes the projection on N, then, of course, the projection on $N^{\perp}$ is $1 - q$. Since $M \cap N^{\perp} = 0$, it follows that if $x \in M$, $x \neq 0$, then x does not belong to $N^{\perp}$, i.e., $(1-q)x \neq x$, and therefore $qx \neq 0$. In other words, the restriction of q to M is a bounded, linear, one-to-one transformation from M into N; the conclusion follows from Lemma 3.

LEMMA 5. *If a is an operator and δ is a positive number such that $\|ax\| \geq \delta\|x\|$ for all x in H, then $R(a) \cap R^{\perp}(b) = R^{\perp}(a) \cap R(b) = 0$ whenever $\|a-b\| < \delta$.*

REMARK. The assumption that a is bounded from below implies

---

[8] We are indebted to Mr. Guido Weiss for this proof; our original proof of Lemma 3 used more technical machinery.

that $R(a)$ is closed.[9] If a is bounded from below by δ and $\|a-b\| < \delta$, then b is bounded from below (by $\delta - \|a-b\|$), and therefore $R(b)$ is closed.

PROOF. If $y \in R(a) \cap R^{\perp}(b)$, then $y = ax$ and $ax \perp bx$; if, on the other hand, $y \in R^{\perp}(a) \cap R(b)$, then $y = bx$, and, again, $ax \perp bx$. In either case, the orthogonality of ax and bx implies that $\|ax\| \leq \|ax - bx\|$ and hence that $\delta\|x\| \leq \|a-b\| \cdot \|x\|$. Since the last term is (strictly) less than $\delta\|x\|$, unless $x = 0$, it follows that, in both cases, $x = y = 0$.

THEOREM 4. *If a_0 is an operator and if the complex number λ_0 does not belong to the approximate point spectrum of a_0, then the multiplicity function m is continuous at the pair (a_0, λ_0). In other words, if $\lambda_0 \in' \Pi(a_0)$, then there exists a positive number δ such that $m(a, \lambda) = m(a_0, \lambda_0)$ whenever $\|a - a_0\| < \delta$ and $|\lambda - \lambda_0| < \delta$.*

PROOF. Since, by hypothesis, 0 does not belong to the approximate point spectrum of $a_0 - \lambda_0$, there exists a positive number δ such that $\|(a_0 - \lambda_0)x\| \geq 2\delta\|x\|$ for all x in H. It follows from Lemma 5 that if $\|a_0 - a\| < \delta$ and $|\lambda_0 - \lambda| < \delta$, so that $\|(a_0 - \lambda_0) - (a - \lambda)\| < 2\delta$, then $R(a_0 - \lambda_0) \cap R^{\perp}(a - \lambda) = R^{\perp}(a_0 - \lambda_0) \cap R(a - \lambda) = 0$. Applying Lemma 4, with $M = R^{\perp}(a - \lambda)$ and $N = R^{\perp}(a_0 - \lambda_0)$, and using the fact that $R(a_0 - \lambda_0)$ is closed (so that $N^{\perp} = R(a_0 - \lambda_0)$), we conclude that $m(a, \lambda) \leq m(a_0, \lambda_0)$. A similar application of Lemma 4, with the roles of M and N interchanged, concludes the proof of the theorem.

THEOREM 5. *If a_0 is an operator and if K is a compact subset of the complement of $\Pi(a_0)$, then there exists a positive number δ such that $m(a, \lambda) = m(a_0, \lambda)$ whenever $\lambda \in K$ and $\|a - a_0\| < \delta$.*

PROOF. By Theorem 4, to each λ_0 in K there corresponds a positive number $\delta(\lambda_0)$ such that $m(a, \lambda) = m(a_0, \lambda_0)$ whenever $\|a - a_0\| < \delta(\lambda_0)$ and λ belongs to the open disc $U(\lambda_0)$ with center at λ_0 and radius $\delta(\lambda_0)$. The compactness of K implies the existence of a finite set $\{\lambda_1, \cdots, \lambda_n\}$ in K such that $K \subset U(\lambda_1) \cup \cdots \cup U(\lambda_n)$. If $\delta = \min\{\delta(\lambda_1), \cdots, \delta(\lambda_n)\}$, and if $\|a - a_0\| < \delta$, then $\|a - a_0\| < \delta(\lambda_j)$ for all j. It follows that if $\lambda \in U(\lambda_j)$, then $m(a, \lambda) = m(a_0, \lambda_j)$; since this applies, in particular, to a_0 itself, so that $m(a_0, \lambda) = m(a_0, \lambda_j)$, it follows that $m(a, \lambda) = m(a_0, \lambda)$ whenever $\lambda \in U(\lambda_j)$. Since, finally, this is true for all j, it follows that $m(a, \lambda) = m(a_0, \lambda)$ whenever $\lambda \in K$.

Application. The following lemma presents in a distilled form those properties of an operator which were used in (I) to prove that certain operators have no square roots.

---

[9] See Paul R. Halmos, *Introduction to Hilbert space*, New York, 1951, p. 37.

LEMMA 6. *If a is an operator and K a compact subset of $\Sigma(a) - \Pi(a)$ such that* (i) $0 \; \epsilon' \; K$, (ii) $\sqrt{K}$ *is connected,*[10] *and* (iii) $m(a, \lambda) = 1$ *whenever* $\lambda \; \epsilon \; K$, *then a has no square roots.*

PROOF. Suppose, on the contrary, that $b^2 = a$, and write

$$L = \sqrt{K} \cap \Sigma(b).$$

Clearly L and $-L$ (= the image of L under the mapping $\lambda \rightarrow -\lambda$) are closed subsets of $\sqrt{K}$, and, since

$$K \subset \Sigma(a) = (\Sigma(b))^2,$$

it follows that $L \cup (-L) = \sqrt{K}$. Since K is disjoint from $\Pi(a)$, it follows that $\sqrt{K}$ is disjoint from $\Pi(b)$ and hence that $L \subset \Sigma(b) - \Pi(b)$. Consequently, if $\lambda \; \epsilon \; L$, then $\bar{\lambda}^2$ is a proper value of a^*; since $\lambda^2 \; \epsilon \; K$, it follows (by (iii)) that $m(a, \lambda^2) = 1$ and hence that $-\lambda$ cannot belong to L (recall that, by (i), $\lambda \neq -\lambda$). In other words, L and $-L$ are disjoint; since (by (ii)) this is a contradiction, the proof is complete.

THEOREM 6. *There exists an invertible operator a, and there exists a positive number δ, such that if $\|a - b\| < \delta$, then b is invertible and b does not have a square root.*

PROOF. Let a be an invertible operator (e.g., one of the analytic position operators constructed in (I)), such that there exists a compact subset K of $\Sigma(a) - \Pi(a)$ with the properties (i), (ii), and (iii) described in Lemma 6. Since the set of invertible operators is open, every operator b sufficiently near to a is itself invertible. Since $\Sigma(b) - \Pi(b)$ depends semicontinuously on b (Theorem 3), it follows that if b is sufficiently near to a, then $K \subset \Sigma(b) - \Pi(b)$. Since, finally, the multiplicity function m is equicontinuous on K (Theorem 5), it follows that if b is sufficiently near to a, then $m(b, \lambda) = 1$ for all λ in K. In other words, there exists a positive number δ such that if $\|a - b\| < \delta$, then b is invertible, $K \subset \Sigma(b) - \Pi(b)$, and $m(b, \lambda) = 1$ for all λ in K. Lemma 6, applied to b in place of a, implies that b has no square root.

It is perhaps worth remarking that all our considerations remain valid, with only inessential changes, for nth roots in place of square roots.

UNIVERSITY OF CHICAGO AND
UNIVERSITY OF MONTEVIDEO

---

[10] The symbol $\sqrt{K}$ denotes the complete inverse image of the set K with respect to the mapping $\lambda \rightarrow \lambda^2$.