

SQUARE ROOTS OF OPERATORS

PAUL R. HALMOS, GÜNTER LUMER, AND JUAN J. SCHÄFFER¹

Introduction. If H is a complex Hilbert space and if A is an operator on H (i.e., a bounded linear transformation of H into itself), under what conditions does there exist an operator B on H such that $B^2=A$? In other words, when does an operator have a square root? The spectral theorem implies that the normality of A is a sufficient condition for the existence of B ; the special case of positive definite operators can be treated by more elementary means and is, in fact, often used as a step in the proof of the spectral theorem. As far as we are aware, no useful necessary and sufficient conditions for the existence of a square root are known, even in the classical case of finite-dimensional Hilbert spaces. The problem of finding some easily applicable conditions is of interest, in part because the use of square roots is frequently a helpful technique in the study of algebraic properties of operators, and in part because of the information that such conditions might yield about the hitherto rather mysterious behavior of non-normal operators.

If a non-zero, 2-rowed square matrix is nilpotent, then its index of nilpotence is equal to 2; this comment shows that no such matrix can have a square root. On the other hand, an elementary computation, based on the Jordan canonical form, shows that every invertible matrix does have a square root. Since the number 0 is known to have a special significance in the formation of square roots, it is not unreasonable to conjecture that its absence from the spectrum of an operator A is sufficient to ensure the existence of a square root of A , or, in other words, that even on not necessarily finite-dimensional Hilbert spaces, every invertible operator has a square root. (This conjecture was first called to our attention by Irving Kaplansky.) The main purpose of this paper is to prove that this conjecture is false. More precisely, we shall describe a small but interesting class of operators, derive a necessary and sufficient condition that an operator in this class have a square root, and achieve our announced purpose by exhibiting a relatively large subclass of invertible operators that do not satisfy the condition. We note in passing that our methods solve the analogous problem for n th roots, $n \geq 2$, and that,

Presented to the Society, September 5, 1952; received by the editors April 4, 1952.

¹ The research on this paper was done while all the authors were in residence at the Instituto de Matemática y Estadística, Universidad de la República, Montevideo, Uruguay.

curiously, the condition for the existence of an n th root, within the class of operators we treat, is the same as that for the special case $n=2$.

Analytic position operators. Let D be a bounded domain (i.e., a bounded, open, and connected subset of the complex plane), let μ be planar Lebesgue measure in D , and let H be the set of all complex-valued functions that are analytic throughout D and square-integrable with respect to μ . In other words, an element of H is an analytic function x of the complex variable t in D and is such that

$$\|x\|^2 = \int_D |x(t)|^2 d\mu(t) < \infty.$$

With respect to the pointwise linear operations, and the inner product defined by

$$(x, y) = \int_D x(t)\bar{y}(t)d\mu(t),$$

the set H is a complex inner product space. If $\lambda \in D$ and $x \in H$, we write

$$v_\lambda(x) = x(\lambda);$$

it is obvious that, for each fixed λ , the functional v_λ is linear.

LEMMA 1. *For each λ in D , the functional v_λ on H is bounded; if, in fact, $a^2(\lambda)$ is the area of the largest open circle C with center at λ that is contained in D , then*

$$|v_\lambda(x)| \leq \frac{1}{a(\lambda)} \|x\|.$$

PROOF. Since

$$\begin{aligned} \|x\|^2 &= \int_D |x(t)|^2 d\mu(t) \geq \int_C |x(t)|^2 d\mu(t) \\ &= \int_C \left| \sum_{n=0}^{\infty} a_n(t - \lambda)^n \right|^2 d\mu(t), \end{aligned}$$

and since the powers of $(t - \lambda)$ constitute an orthogonal set in $L_2(C)$, it follows that

$$\|x\|^2 \geq \int_C |a_0|^2 d\mu(t) = a^2(\lambda) \cdot |x(\lambda)|^2.$$

LEMMA 2. *The inner product space H is a Hilbert space.*

PROOF. If $\{x_n\}$ is a Cauchy sequence in H , then, by Lemma 1,

$$|x_n(\lambda) - x_m(\lambda)| \leq \frac{1}{a(\lambda)} \|x_n - x_m\|$$

for every λ in D . It follows that if D_0 is a compact subset of D , so that $a(\lambda)$ is bounded away from 0 when $\lambda \in D_0$, then the sequence $\{x_n\}$ of functions is uniformly convergent on D_0 . This implies that there exists an analytic function x on D such that $x_n(\lambda) \rightarrow x(\lambda)$ for all λ in D . At the same time, the completeness of $L_2(D)$ implies the existence of a complex-valued, square-integrable, but not necessarily analytic, function y on D such that $x_n \rightarrow y$ in the mean of order 2. It follows that a subsequence of $\{x_n\}$ converges to y almost everywhere and hence that $x = y$ almost everywhere. This implies that x is square-integrable, i.e., that $x \in H$, and hence that H is complete.²

The equation $(Ax)(t) = tx(t)$ defines a linear transformation A of H into itself; the boundedness of D implies that A is an operator. We shall call A the *analytic position operator* associated with the domain D . The class of analytic position operators is the class we mentioned in the introduction.³

Spectra. For our purposes we shall have need of an almost complete analysis of the spectrum of an analytic position operator; we devote this brief section to recalling the pertinent facts about spectra. If A is an operator on a Hilbert space H , the *spectrum* of A , in symbols $\Lambda(A)$, is the set of all those complex numbers λ for which $A - \lambda$ is not invertible. A well known and easy geometric argument shows that there are two (exhaustive but not exclusive) ways in which

² The results of Lemmas 1 and 2 are not new; they can be found, for instance, in Stefan Bergman, *Sur les fonctions orthogonales de plusieurs variables complexes avec les applications à la théorie des fonctions analytiques*, Paris, 1947, p. 24. Since, however, we have occasion below to make use of the notation and of the cornerstone (Lemma 1) of the reasoning above, we thought it appropriate to give the proof. It might also be remarked that since our proof makes explicit use of the Riesz-Fischer theorem, instead of proving it in the particular case at hand, it is somewhat simpler, from the point of view of the standard theory of Hilbert spaces, than the more analytic argument given by Bergman.

³ If μ is a compact measure in the complex plane (i.e., a measure with compact support, in the terminology of Bourbaki), and if $(Ax)(t) = tx(t)$ for every complex-valued function x that is square-integrable with respect to μ , then we call A the position operator associated with μ . The adjective *analytic* serves as a reminder that the spectrum (support) of μ is the closure of a bounded open set and that the domain of A consists only of functions analytic in the given open set. The terminology is motivated by an analogy with certain operators considered in quantum mechanics.

$A - \lambda$ can fail to be invertible. One way is that the range of $A - \lambda$, in symbols $R(A - \lambda)$, fails to be dense in H ; if this is the case, we say that λ belongs to the *compression spectrum* of A and we write $\lambda \in \Sigma(A)$. Another way is that, for suitable vectors x , the expression $\|(A - \lambda)x\|$ becomes arbitrarily small in comparison with $\|x\|$; if this is the case, we say that λ belongs to the *approximate point spectrum* of A and we write $\lambda \in \Pi(A)$. A more special way for $A - \lambda$ to fail to be invertible is for λ to belong to the *point spectrum* $\Pi_0(A)$; this means, of course, that $(A - \lambda)x = 0$ for a suitable non-zero vector x . We shall need to make use of the facts that (1) $\Lambda(A)$ is a closed (and in fact compact) subset of the complex plane, (2) $\Lambda(A) = \Sigma(A) \cup \Pi(A)$, and (3) $\Pi_0(A^*) = \Sigma^*(A)$. (In (3), A^* denotes, of course, the adjoint of the operator A , and $\Sigma^*(A)$ denotes the set of all complex numbers of the form $\bar{\lambda}$ with $\lambda \in \Sigma(A)$.)⁴

We shall also need a slightly more deep-lying fact about spectra, a special case of the so-called spectral mapping theorem.⁵ The part of the theorem that is relevant to our work asserts that if p is any polynomial (and, in particular, if $p(\lambda) = \lambda^2$), then $\Sigma(p(A)) = p(\Sigma(A))$, $\Pi(p(A)) = p(\Pi(A))$, and $\Pi_0(p(A)) = p(\Pi_0(A))$. (A symbol such as $p(\Sigma(A))$ denotes the set of all complex numbers of the form $p(\lambda)$ with $\lambda \in \Sigma(A)$.) It is pertinent to remark that the first two of these three equations, together with (2) above, imply that $\Lambda(p(A)) = p(\Lambda(A))$.

The spectrum of an analytic position operator. Suppose now that A is the analytic position operator associated with a domain D ; we propose to establish the connection between D and the spectrum of A . We begin with an auxiliary result.

LEMMA 3. *If $\lambda \in D$ and if y is a function in H such that $y(\lambda) = 0$, then there exists a unique function x in H such that $y(t) = (t - \lambda)x(t)$ for all t in D . If, moreover, δ is a positive number such that the open circle C with center λ and radius δ is contained in D , then*

$$\|y\|^2 \geq \frac{\delta^2}{2} \|x\|^2.$$

PROOF. It is obvious that there exists a unique analytic function x on D such that $y(t) = (t - \lambda)x(t)$ for all t in D . Since the square-integrability of x , i.e., the relation $x \in H$, is a weaker condition than the asserted inequality between $\|x\|$ and $\|y\|$, it remains only to prove that inequality.

⁴ For a treatment of the elementary properties of spectra see, for instance, Paul R. Halmos, *Introduction to Hilbert space and the theory of spectral multiplicity*, New York, 1951, p. 50.

⁵ See Einar Hille, *Functional analysis and semi-groups*, New York, 1948, p. 123.

We observe first of all (cf. the proof of Lemma 1) that $\int_C (t-\lambda)^n (\bar{t}-\bar{\lambda})^m d\mu(t) = 0$ unless $m=n$ ($n, m=0, 1, 2, \dots$), and that

$$\int_C |(t-\lambda)^n|^2 d\mu(t) = \frac{\pi \delta^{2n+2}}{n+1};$$

these facts are the results of an easily verified computation. If x is expanded in powers of $(t-\lambda)$, i.e.,

$$x(t) = \sum_{n=0}^{\infty} a_n (t-\lambda)^n, \quad t \in C,$$

then it follows that

$$\int_C |x(t)|^2 d\mu(t) = \sum_{n=0}^{\infty} |a_n|^2 \cdot \frac{\pi \delta^{2n+2}}{n+1}$$

and

$$\int_C |(t-\lambda)x(t)|^2 d\mu(t) = \sum_{n=0}^{\infty} |a_n|^2 \cdot \frac{\pi \delta^{2n+4}}{n+2}.$$

Consequently

$$\begin{aligned} \int_C |(t-\lambda)x(t)|^2 d\mu(t) &= \sum_{n=0}^{\infty} \frac{\delta^2(n+1)}{n+2} \cdot |a_n|^2 \cdot \frac{\pi \delta^{2n+2}}{n+1} \\ &\geq \frac{\delta^2}{2} \int_C |x(t)|^2 d\mu(t). \end{aligned}$$

The desired result now follows from the relations

$$\begin{aligned} \|y\|^2 &= \int_D |(t-\lambda)x(t)|^2 d\mu(t) \\ &\geq \frac{\delta^2}{2} \int_C |x(t)|^2 d\mu(t) + \delta^2 \int_{D-C} |x(t)|^2 d\mu(t) \\ &\geq \frac{\delta^2}{2} \|x\|^2. \end{aligned}$$

THEOREM 1. *If A is the analytic position operator associated with a domain D , and if $\lambda \in D$, then λ is not an approximate proper value of A , but $\bar{\lambda}$ is a simple proper value of A^* , and therefore, in particular, $D \subset \Sigma(A) - \Pi(A)$.*

PROOF. Lemma 1 implies that

$$\|(A - \lambda)x\|^2 \geq \frac{\delta^2}{2} \|x\|^2$$

for all x in H and hence that λ is not an approximate proper value of A . Lemma 1 implies also that a necessary and sufficient condition that a function y belong to $R(A - \lambda)$ is that $v_\lambda(y) = 0$. Since the range of an operator is dense in the orthogonal complement of the null space of its adjoint, it follows that $A^*z = \bar{\lambda}z$ if and only if z is orthogonal to the null space of v_λ . Since, by the standard Riesz representation of bounded linear functionals in terms of inner products, the dimension of the orthogonal complement of the null space of v_λ is equal to 1, the proof of the theorem is complete.

On the basis of this theorem it is easy to determine the various parts of the spectrum completely. For our purposes, however, it is sufficient to know Theorem 1 and the additional, global fact that $\Lambda(A) = \overline{D}$. Indeed, since $D \subset \Lambda(A)$ by Theorem 1, and since $\Lambda(A)$ is closed, it follows that $\overline{D} \subset \Lambda(A)$. If, on the other hand, λ does not belong to \overline{D} , and if

$$f_\lambda(t) = \frac{1}{t - \lambda},$$

then f_λ is a bounded analytic function on D and consequently the equation $(A_\lambda x)(t) = f_\lambda(t)x(t)$ defines an operator A_λ on H that is easily seen to be a two-sided inverse of $A - \lambda$. In other words $\Lambda(A)$ does not contain any point in the complement of \overline{D} , so that $\Lambda(A) = \overline{D}$.

Square roots of analytic position operators. In order to state the main theorem of this paper, it is convenient to introduce a new notation; if D is a domain, we shall write \sqrt{D} for the (open) set of all complex numbers λ such that $\lambda^2 \in D$.

THEOREM 2. *If A is the analytic position operator associated with a domain D , then a necessary and sufficient condition that A have a square root is that \sqrt{D} be disconnected.*

PROOF. Let U be a component of \sqrt{D} and write $V = U \cup (-U)$, where $-U$ denotes the set of all complex numbers of the form $-\lambda$ with $\lambda \in U$. Since D is open, so also are V and $W = D - V$. Since V is the complete square root of V^2 , and since V and W are disjoint, it follows that V^2 and W^2 are disjoint, where V^2 , for instance, denotes the set of all complex numbers of the form λ^2 with $\lambda \in V$. Since an analytic function (and in particular the function f defined by $f(\lambda) = \lambda^2$) is an open mapping, V^2 and W^2 are open, and, clearly, $V^2 \cup W^2 = D$. The connectedness of D implies that W^2 is empty and hence that W

is empty. This proves that if \sqrt{D} is not connected, then it consists of exactly two open components: U and its reflection $-U$. The correspondence f (where $f(\lambda) = \lambda^2$) is one-to-one between U and D , and therefore the inverse correspondence g between D and U is a (single-valued) bounded analytic function on D . The equation $(Bx)(t) = g(t)x(t)$ defines, therefore, an operator B on H ; evidently $B^2 = A$.

Suppose now that \sqrt{D} is connected; the same is then true of the set $\sqrt{D} - \{0\}$ obtained by deleting the origin from \sqrt{D} , in case it happened to belong to \sqrt{D} . We shall derive a contradiction from the assumption that there exists an operator B on H such that $B^2 = A$. Since $D \subset \Sigma(A) - \Pi(A)$, it follows from the spectral mapping theorem that no point of \sqrt{D} can belong to $\Pi(B)$. Since $\Lambda^2(B) = \Lambda(A)$, it follows that if $\lambda \in D$, then at least one of the two square roots of λ belongs to $\Lambda(B)$ and hence, in view of the preceding sentence, to $\Sigma(B)$. If both the square roots of a non-zero number λ in D belonged to $\Sigma(B)$, then their conjugates would belong to $\Pi_0(B^*)$ and this would contradict the fact that $\bar{\lambda}$ is a simple proper value of A^* . What we have proved so far may be formulated as follows: if $\lambda \in \sqrt{D} - \{0\}$, then exactly one of the two numbers λ and $-\lambda$ belongs to $\Lambda(B)$. The set $(\sqrt{D} - \{0\}) \cap \Lambda(B)$ is a closed subset of $\sqrt{D} - \{0\}$ in the relative topology; the homeomorphism $\lambda \rightarrow -\lambda$ carries it onto its relative complement in $\sqrt{D} - \{0\}$, which is therefore also closed. Since $\sqrt{D} - \{0\}$ is connected, we have reached the desired contradiction, and the proof of the theorem is complete.⁶

Conclusion. (a) The condition of Theorem 2 is of a mixed algebraic and topological nature. It is not difficult, however, to replace it by a purely topological condition. Elementary reasoning, based on the information that the proof of Theorem 2 yields about the structure of \sqrt{D} , implies that \sqrt{D} is connected if and only if D contains a Jordan curve surrounding the origin. Less elementary, topological, considerations can be used to show also that the latter condition is satisfied if and only if the origin does not belong to the infinite component of the complement of D .⁷ Both these conditions may be expressed, in intuitive terms, by saying that D surrounds the origin.

⁶ A sufficient condition for the existence of square roots of more general operators is given by Hille (op. cit., p. 276); in the special case of analytic position operators, however, the condition of Theorem 2 is considerably less restrictive than Hille's.

⁷ The sufficiency of this condition follows also from Aurel Wintner, *On the logarithms of bounded matrices*, Amer. J. Math. vol. 74 (1952) pp. 360-364. Wintner's work appeared after this paper had been submitted for publication. Wintner, incidentally, also raises the question (answered in (c) below) of the existence of an invertible operator without a square root.

(b) For reasons of notational convenience we treated the case of square roots only; exactly the same method, however, solves the problem of n th roots for any $n = 2, 3, 4, \dots$. The principal theorem in this case is obtained from Theorem 2 by replacing \sqrt{D} by $\sqrt[n]{D}$. Since the purely topological reformulation of this condition, along the lines of (a) above, is independent of n , it follows that a necessary and sufficient condition that an analytic position operator have an n th root ($n = 2, 3, 4, \dots$) is that it have a square root.

(c) What we now know about the spectra and the square roots of analytic position operators makes it very easy to construct an example of the type mentioned in the introduction, i.e., an example of an invertible operator without a square root. Indeed, the analytic position operator associated with a domain D is invertible if and only if 0 does not belong to \overline{D} and it has a square root if and only if \sqrt{D} is disconnected. The problem becomes then the construction of a domain D far from and surrounding the origin. This problem has many obvious solutions; one, for example, is an annulus with center at the origin.

UNIVERSITY OF CHICAGO AND
UNIVERSITY OF MONTEVIDEO