PROCEEDINGS OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 69, Number 2, May 1978

GENERALIZATIONS OF TEMPLE'S INEQUALITY

EVANS M. HARRELL II

ABSTRACT. T. Kato's little-known generalization of a classic variational inequality for eigenvalues is extended to the case of normal operators and briefly discussed.

It is usually not possible to evaluate precisely the eigenvalues of the linear operators which occur in realistic models in the physical sciences. It is thus a problem of great practical importance to have formulae for approximate evaluation of eigenvalues and for the errors of those approximations. The most important approximate formula for an eigenvalue is the Rayleigh-Ritz inequality, which gives an upper bound for the lowest eigenvalue of a selfadjoint operator. This is the prototype of a variational estimate, whereby a set of approximate eigenfunctions is guessed at and used to estimate the eigenvalues. The problem of obtaining lower bounds for the lowest eigenvalue of a selfadjoint operator is notoriously more difficult than the discovery of upper bounds, but some methods are widely known, though not so widely as the Rayleigh-Ritz inequality. The best such bound which relies only on the selfadjointness of the operator and the isolation of the lowest eigenvalue from the rest of the spectrum is due to G. Temple [**7** Theorem 1].

The proofs of the Rayleigh-Ritz inequality and Temple's inequality show them to be straightforward applications of the spectral theorem [**5**], [**6**], and similar arguments can extend these inequalities to give useful estimates for any isolated eigenvalue of a selfadjoint operator–without the necessity of first estimating all the lower eigenvalues, as with the min-max principle. It is peculiar and unfortunate that more than two decades elapsed between Temple's original paper and the discovery of the generalization of Temple's inequality to arbitrary eigenvalues, and that this generalization has remained but little known for almost three more decades. In this paper Temple's inequality is generalized still further to the case of normal operators. Its use is not only for numerical computation, but also for the proofs of many abstract theorems about perturbation expansions and convergence of operator-valued functions [**1**], [**2**], [**4**], [**5**].

The classical result is

Received by the editors February 18, 1977 and, in revised form, June 28, 1977.

AMS (MOS) subject classifications (1970). Primary 47A10, 35P15.

Key words and phrases. Temple's inequality, bounds for eigenvalues, normal operators, unitary operators, trial functions.

THEOREM 1 (G. TEMPLE). *Let A be a selfadjoint operator on a Hilbert space with inner product $(\cdot, \cdot)$, linear in the second position and conjugate-linear in the first. Suppose A has a lowest eigenvalue E_0 isolated from the rest of the spectrum, and define*

$$\beta \equiv \inf[\operatorname{sp}(A) \setminus \{E_0\}].$$

Let ψ be a trial function for A, that is, a normalized vector in the domain of A, such that $(\psi, A\psi) < \beta$. Then

$$E_0 \geqslant (\psi, A\psi) - [\|A\psi\|^2 - (\psi, A\psi)^2]/[\beta - (\psi, A\psi)].$$

This inequality is naturally coupled with the Rayleigh-Ritz inequality, viz.,

$$E_0 \leqslant (\psi, A\psi).$$

REMARK. In practice, $\inf[\operatorname{sp}(A) \setminus \{E_0\}]$ may not be known, and so it is common to see this theorem stated for β any number less than or equal to $\inf[\operatorname{sp}(A) \setminus \{E_0\}]$, but greater than E_0. A lower β gives a weaker inequality.

T. Kato [3] found a generalization of Temple's inequality for arbitrary eigenvalues, which has the additional virtue of being symmetric in upper and lower bounds. The proof of the following version of that theorem emphasizes the geometrical idea which is also the key to the generalization.

THEOREM 2 (T. KATO). *Let A be a selfadjoint operator on a Hilbert space, and let ψ be a trial function for A. Define*

$$\eta \equiv (\psi, A\psi) \quad \textit{and} \quad \varepsilon \equiv \|[A - \eta]\psi\|.$$

(Hence $\varepsilon^2 = \|A\psi\|^2 - \eta^2$.) If

$$\varepsilon^2 < (\beta - \eta)(\eta - \alpha) \tag{1}$$

for real numbers $\alpha < \beta$, then $\operatorname{sp}(A) \cap (\alpha, \beta) \neq \varnothing$ (empty set). Moreover, if the only point of the spectrum in the interval (α, β) is an eigenvalue E, then

$$\eta - \varepsilon^2/[\beta - \eta] \leqslant E \leqslant \eta + \varepsilon^2/[\eta - \alpha]. \tag{2}$$

PROOF. Suppose to the contrary that $\operatorname{sp}(A) \cap (\alpha, \beta) = \varnothing$. Then if $\nu \in \operatorname{sp}(A)$, $|\nu - [\beta + \alpha]/2| \geqslant [\beta - \alpha]/2$. (This is just the statement that the distance from ν to the center of the interval is at least as great as half the width of the interval.)

$$\nu^2 - \eta^2 \geqslant [\beta + \alpha]\nu - \alpha\beta - \eta^2. \tag{3}$$

By the spectral theorem, there is a normalized measure μ_ψ associated with ψ so that $\int_{\operatorname{sp}(A)} \nu \, d\mu_\psi(\nu) = (\psi, A\psi)$ and $\int_{\operatorname{sp}(A)} \nu^2 \, d\mu_\psi(\nu) = (A\psi, A\psi) = \|A\psi\|^2$. If (3) is integrated by this measure, then

$$\|A\psi\|^2 - \eta^2 \geqslant (\beta + \alpha)\eta - \alpha\beta - \eta^2 = (\beta - \eta)(\eta - \alpha),$$

contradicting (1). Thus $\mathrm{sp}(A) \cap (\alpha, \beta) \neq \varnothing$.

(2) follows by noting that (1) still holds when either α is increased to any value less than $\eta - \varepsilon^2/[\beta - \eta]$, or β is decreased to any value greater than $\eta + \varepsilon^2/[\eta - \alpha]$.

REMARK. The Rayleigh-Ritz inequality and Temple's inequality are corollaries of Theorem 2, with $\alpha = \infty$ and $\beta = \inf[\mathrm{sp}(A) \setminus \{E_0\}]$.

Since the spectral theorem holds for all normal operators, Theorem 2 can be generalized by exploiting the same geometrical idea in the complex plane. The most important cases are treated in the following theorem.

THEOREM 3. *Let A be a normal operator on a Hilbert space, and let ψ be a trial function for A. Define η and ε as before, and define*

$$C(\gamma, d) \equiv \{z \in \mathbf{C}\colon |z - \gamma| < d\},$$

the open disc of radius d around the complex point γ.

A. *Suppose that $\eta \notin \mathrm{sp}(A)$. Then for any line in $\mathbf{C}$ containing η, either* (i) *that line contains at least two points of the spectrum of A, one on each side of η, or* (ii) *each open half-plane divided by the line contains part of the spectrum.*

B. *If*

$$\varepsilon^2 < d^2 - |\eta - \gamma|^2, \tag{4}$$

then $\mathrm{sp}(A) \cap C(\gamma, d) \neq \varnothing$.

C. *If it is known that there is only one point, E, of the spectrum in $C(\gamma, d)$, and $\varepsilon < l \equiv d - |\eta - \gamma|$ (which is the distance from η to the edge of $C(\gamma, d)$), then*

$$|E - \eta| \leqslant \varepsilon^2/l, \tag{5}$$

that is, E is contained in the closed disc, $\overline{C(\eta, \varepsilon^2/l)}$.

D. *Suppose that an isolated point E is known to be the only point of* $\mathrm{sp}(A)$ *outside some sector in the complex plane, of arbitrary vertex, orientation, and opening angle $< \pi$, and that η is also outside that sector. Some larger sector with vertex η, which we denote by $\{z\colon \alpha_1 \leqslant \arg(z - \eta) \leqslant \alpha_2\}$, where $\alpha_1 < \alpha_2 < \alpha_1 + 2\pi$, must also contain* $\mathrm{sp}(A) \setminus \{E\}$. *Then*

$$E \in \{z\colon \alpha_1 + \pi \leqslant \arg(z - \eta) \leqslant \alpha_2 + \pi\}, \tag{6}$$

and

$$|E - \eta| \leqslant \varepsilon^2/l', \tag{7}$$

where l' is (less than or) equal to the distance from η to the original sector containing $\mathrm{sp}(A) \setminus \{E\}$.

E. *If A is unitary, and for some $\beta_1 < \beta_2 < \beta_1 + 2\pi$,*

$$(8) \qquad |\eta| \cos((\beta_1 + \beta_2)/2 - \arg(\eta)) > \cos((\beta_2 - \beta_1)/2),$$

then $\mathrm{sp}(A) \cap \{z = e^{i\theta}: \beta_1 < \theta < \beta_2\} \neq \emptyset$. *If it is known that there is exactly one point, E, of* $\mathrm{sp}(A)$ *in the open arc* $\{z = e^{i\theta}: \beta_1 < \theta < \beta_2\}$, *then*

$$\arg(\eta) - \tan^{-1}\left[\frac{(1 - |\eta|^2)\sin(\beta_2 - \arg(\eta))}{2|\eta| - (|\eta|^2 + 1)\cos(\beta_2 - \arg(\eta))}\right] \leqslant \arg(E)$$

$$\leqslant \arg(\eta)\tan^{-1}\left[\frac{(1 - |\eta^2|)\sin(\arg(\eta) - \beta_1)}{2|\eta| - (|\eta|^2 + 1)\cos(\arg(\eta) - \beta_1)}\right].$$

REMARKS. Parts C, D, and E of this theorem all give bounds $E = \eta + O(\varepsilon^2)$, as does Theorem 2. Parts C and D are not the strongest possible conclusions under these hypotheses; part C gives the smallest disc that must contain E, and part D gives the smallest circular sector that must contain E. More precise estimates are obtainable using figures contained in, and osculating, this disc and this circular sector. In general, if more is known about the position of $\mathrm{sp}(A)$, more precise estimates can be obtained.

Part D,(6), taken by itself, is the natural generalization of the Rayleigh-Ritz inequality for normal operators, as it says that the isolated eigenvalue is on the far side of the variational estimate, η, from the rest of the spectrum.

The reason that ε does not appear in part E is that ε is uniquely determined by η, when A is unitary; $\|A\psi\|^2 = \|\psi\|^2 = 1$, so

$$\varepsilon^2 = \|A\psi\|^2 - |(\psi, A\psi)|^2 = 1 - |\eta|^2.$$

PROOF. A. The spectral theorem implies that η is in the convex hull of $\mathrm{sp}(A)$, which immediately gives part A.

B. Assume to the contrary that $\mathrm{sp}(A) \cap C(\gamma, d) = \emptyset$. Then if $\nu \in \mathrm{sp}(A)$, $|\nu - \gamma| \geqslant d$, and so by simple algebra,

$$|\nu|^2 - |\eta|^2 \geqslant d^2 + \gamma\bar{\nu} + \bar{\gamma}\nu - |\gamma|^2 - |\eta|^2.$$

If this is integrated by the spectral measure associated with ψ, there results

$$\|A\psi\|^2 - |\eta|^2 \geqslant d^2 - |\eta - \gamma|^2,$$

which contradicts (4). Thus $\mathrm{sp}(A) \cap C(\gamma, d) \neq \emptyset$.

C. For the estimate (5), note that E must be in the intersection of all discs $C(\gamma', d')$ where $\varepsilon^2 < d'^2 - |\eta - \gamma'|^2$, and *a fortiori* in the intersection of all discs where $d' = l/2 + \varepsilon^2/2l$, and $|\gamma' - \eta| < l/2 - \varepsilon^2/2l$. (These are optimal conditions on $C(\gamma', d')$ with d' independent of γ', subject to the restriction that $C(\gamma', d') \subset C(\gamma, d)$. A typical $C(\gamma', d')$ is shown in Figure 1.) The intersection of all such discs is a closed disc centered on η, with radius $l/2 + \varepsilon^2/2l - (l/2 - \varepsilon^2/2l) = \varepsilon^2/l$, i.e., $E \in \overline{C(\eta, \varepsilon^2/l)}$.

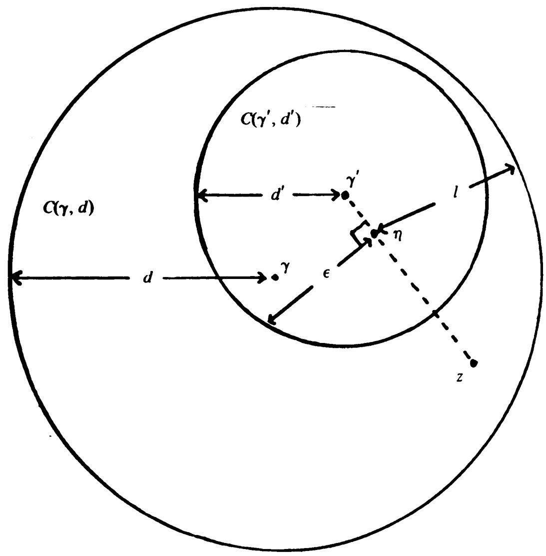

FIGURE 1. Geometry of the proof of Theorem 3C.

The point z cannot be the eigenvalue unless it is contained in all discs $C(\gamma', d')$, where $\varepsilon^2 < d'^2 - |\gamma' - \eta|^2$; and,in particular, in the disc where γ' and d' are chosen so that the near edge of $C(\gamma', d')$ is as far as possible from z, but with $C(\gamma', d') \subset C(\gamma, d)$.

D. (6) is a consequence of part A. Denote by P the point of the original sector containing $\mathrm{sp}(A) \setminus \{E\}$ nearest to η. Then choose γ to be collinear with η and P, and on the far side of η from P, so that $|\eta - \gamma| > (\varepsilon^2/l' - l')/2$, and choose $d = |\eta - \gamma| + l'$. The disc $C(\gamma, d)$ does not intersect the original sector, and $\varepsilon^2 < d^2 - |\eta - \gamma|^2$. Therefore by part B, $E \in C(\gamma, d) \subset C(\eta, r)$, for some $r > \varepsilon^2/l'$. If $|\eta - \gamma|$ is decreased to $(\varepsilon^2/l' - l')/2$, r can be taken arbitrarily close to ε^2/l', giving (7).

E. The first claim of part E is also a consequence of part A, because (8) is equivalent to the statement that η is on the same side of the chord from $e^{i\beta_1}$ to $e^{i\beta_2}$ as the arc $\{z = e^{i\theta}\colon \beta_1 < \theta < \beta_2\}$. (Recall that the spectrum of a unitary operator is contained in the unit circle in the complex plane.)

As in Theorem 2, the best bounds on E are obtained by alternatively increasing β_1 and decreasing β_2 as much as possible subject to the restriction (8). This amounts to finding the sector (6) of part D, and seeing what arc it subtends. A schoolchild's exercise in geometry, using the law of sines, shows that the chord containing η and intercepting the unit circle α radians in one direction from the point $\eta/|\eta|$, also intercepts the unit circle at

$$\tan^{-1}\Big((1 - |\eta|^2)\sin(\alpha)/\big[2|\eta| - (|\eta|^2 + 1)\cos(\alpha)\big]\Big)$$

radians in the other direction from $\eta/|\eta|$, which gives (9).

At least two improvements on this theorem could be made: to treat the case where there are n isolated eigenvalues in some region, as was done in the selfadjoint case in [**3**]; and to get inequivalent variational bounds on the eigenvalues by applying these theorems to $f(A)$, for various functions f. (The functions most likely to be known are the polynomials of A, the inverse and resolvent of A, and e^{-tA} and e^{itA}.) As these are straightforward, and are best tailored to the specific problems they may be used in, I shall leave them to be calculated as they are needed.

A useful companion for Theorem 3, which may be of great antiquity, at least for the selfadjoint case, is the following estimate of the accuracy of the trial function.

LEMMA 4. *Let E be an eigenvalue of a normal operator A, such that E is isolated from the rest of the spectrum by a distance $d > 0$. Let ψ be a trial function for A satisfying $\|[A - E]\psi\| \leqslant d' < d$. Let P be the orthogonal projection onto the eigenspace of E. Then*

$$1 \geqslant (\psi, P\psi) = \|P\psi\|^2 \geqslant 1 - (d'/d)^2.$$

PROOF. $A - E = [A - E][1 - P]$. Therefore

$$d' \geqslant \|[A - E]\psi\| = \|[A - E][1 - P]\psi\| \geqslant d\|[1 - P]\psi\|.$$

Squaring, and using $\|[1 - P]\psi\|^2 = 1 - \|P\psi\|^2$, one finds

$$\|P\psi\|^2 \geqslant 1 - (d'/d)^2.$$

The other inequality is trivial.

REFERENCES

1. W. M. Greenlee, *Perturbation of bounded states of the radial equation by potentials with a repulsive hard core–first order asymptotics*, J. Functional Analysis (to appear).

2. E. Harrell, *Singular perturbation potentials*, Ann. Physics **105** (1977), 379–405.

3. T. Kato, *On the upper and lower bounds of eigenvalues*, J. Phys. Soc. Japan **4** (1949), 334–339.

4. ———, *On the convergence of the perturbation method*. I, Progr. Theoret. Phys. **4** (1949), 514–523; II.1, Ibid. **5** (1950), 95–101; II.2, Ibid. **5** (1950), 207–212.

5. ———, *Perturbation theory for linear operators*, Grundlehren Math. Wiss., Band **132**, Springer-Verlag, New York, 1966.

6. M. Reed and B. Simon, *Methods of modern mathematical physics*. I: *Functional analysis*, Academic Press, New York, 1972.

7. G. Temple, *The theory of Rayleigh's principle as applied to continuous systems*, Proc. Roy. Soc. London Ser. A **119** (1928), 276–293.

DEPARTMENT OF MATHEMATICS, MASSACHUSETTS INSTITUTE OF TECHNOLOGY, CAMBRIDGE, MASSACHUSETTS 02139