Electronic Transactions on Numerical Analysis.
Volume 7, 1998, pp. 18-39.

ISSN 1068-9613.

ETNA
Kent State University
etna@mcs.kent.edu

HARMONIC RITZ AND LEHMANN BOUNDS*

CHRISTOPHER BEATTIE†

Abstract. This article reviews a variety of results related to optimal bounds for matrix eigenvalues — some results presented here are well-known; others are less known; and a few are new. The focus rests especially on Ritz and harmonic Ritz values, and right- and left-definite variants of Lehmann's optimal bounds. Two new computationally advantageous reformulations of left-definite Lehmann bounds are introduced, together with a discussion indicating why they might be preferable to the cheaper right-definite bounds.

Key words. optimal eigenvalue bounds, Lehmann intervals, harmonic Ritz values.

AMS subject classifications. 65F15, 49R05.

Dedicated to the memory of Friedrich Goerisch.

1. Introduction. Eigenvalue estimates that are optimal in some sense have self-evident appeal and leave estimators with a sense of virtue and economy. It is natural then that ongoing searches for effective strategies for difficult tasks, such as estimating matrix eigenvalues that are situated well into the interior of the spectrum, revisit from time to time methods that are known to yield optimal bounds.

The overall thrust of this work is an elaboration of the obvious assertion that useful information about the eigenvalues of a matrix can be obtained from some of its submatrices — or what amounts to the same thing, from discerning the action of the matrix on vectors in a given subspace. In practice, one has a variety of strategies to select from to generate useful subspaces, and then one may select from among a variety of strategies to determine eigenvalue information from this subspace. While these two processes often blur together, it is useful to separate their effects. Our goal here then, is to consider only how various approaches for extracting eigenvalue estimates compare for a given subspace, without regard to the inherent qualities different subspaces may bring to the approximation process.

In §2, the simplest estimates available from a subspace, the Ritz values, are discussed together with a well-known variant ("harmonic Ritz values") and a new cousin ("dual harmonic Ritz values"). Left- and right-definite variants of Lehmann's bounds are reviewed in §3 while §4 considers reformulations that may be computationally advantageous. §5 addresses the question of how right- and left-definite variants of Lehmann bounds compare with one another while §6 considers how they might compare with standard Ritz estimates. The subspaces considered throughout §2–§5 are arbitrary; only in the last section do we assume that the approximating subspace is a Krylov subspace.

2. Ritz and Related Values. Let $\mathbf{K}$ and $\mathbf{M}$ be $n \times n$ real symmetric positive definite matrices and consider the eigenvalue problem

$$\mathbf{K}\mathbf{x} = \lambda\,\mathbf{M}\mathbf{x}. \tag{2.1}$$

Label the eigenvalues from the edges toward the center (following [16]) as

$$\lambda_1 \le \lambda_2 \le \lambda_3 \le \cdots \le \lambda_{-3} \le \lambda_{-2} \le \lambda_{-1}$$

with labeling inherited by the associated eigenvectors: $\mathbf{x}_1,\ \mathbf{x}_2,\ \ldots,\ \mathbf{x}_{-2},\ \mathbf{x}_{-1}$.

*Received January 15, 1998. Accepted for publication July 15, 1998. Recommended by R. Lehoucq.

†Department of Mathematics, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061 USA. (beattie@calvin.math.vt.edu).

Solutions to (2.1) are evidently eigenvalue/eigenvector pairs of the matrix $\mathbf{M}^{-1}\mathbf{K}$, which is non-symmetric on the face of it. However, $\mathbf{M}^{-1}\mathbf{K}$ is self-adjoint with respect to the both the $\mathbf{M}$-inner product, $\mathbf{x}^t\mathbf{M}\mathbf{x}$, and the $\mathbf{K}$-inner product, $\mathbf{x}^t\mathbf{K}\mathbf{x}$. Denote by $\mathbf{x}^{\mathsf{m}}$ the $\mathbf{M}$-adjoint of a vector $\mathbf{x}$, $\mathbf{x}^{\mathsf{m}} = \mathbf{x}^t\mathbf{M}$, and by $\mathbf{x}^{\mathsf{k}}$ the $\mathbf{K}$-adjoint, $\mathbf{x}^{\mathsf{k}} = \mathbf{x}^t\mathbf{K}$. "Self-adjointness" of $\mathbf{M}^{-1}\mathbf{K}$ amounts to the assertion that for all $\mathbf{x}$ and $\mathbf{y}$, $\mathbf{x}^{\mathsf{m}}(\mathbf{M}^{-1}\mathbf{K}\mathbf{y}) = (\mathbf{M}^{-1}\mathbf{K}\mathbf{x})^{\mathsf{m}}\mathbf{y}$ and $\mathbf{x}^{\mathsf{k}}(\mathbf{M}^{-1}\mathbf{K}\mathbf{y}) = (\mathbf{M}^{-1}\mathbf{K}\mathbf{x})^{\mathsf{k}}\mathbf{y}$. Self-adjointness with respect to the $\mathbf{M}$- and $\mathbf{K}$-inner products implies that the matrix representation of $\mathbf{M}^{-1}\mathbf{K}$ with respect to any $\mathbf{M}$-orthogonal or $\mathbf{K}$-orthogonal basis will be *symmetric*.

For a given subspace $\mathcal{P}$ of dimension $m < n$, the Rayleigh-Ritz method proceeds by selecting a basis for $\mathcal{P}$, say constituting the columns of a matrix $\mathbf{P} \in \mathbb{R}^{n\times m}$, and then considering the (smaller) eigenvalue problem

$$\mathbf{P}^t\mathbf{K}\mathbf{P}\mathbf{y} = \Lambda\, \mathbf{P}^t\mathbf{M}\mathbf{P}\mathbf{y}. \tag{2.2}$$

This will yield m eigenvalues (called *Ritz values*) labeled similarly to $\{\lambda_i\}$ as

$$\Lambda_1 \leq \Lambda_2 \leq \Lambda_3 \leq \cdots \leq \Lambda_{-3} \leq \Lambda_{-2} \leq \Lambda_{-1}$$

with corresponding eigenvectors $\mathbf{y}_1,\ \mathbf{y}_2,\ \ldots\ \mathbf{y}_{-2},\ \mathbf{y}_{-1}$. Vectors in $\mathcal{P}$ given as $\mathbf{u}_k = \mathbf{P}\mathbf{y}_k$ are *Ritz vectors* associated with the Ritz values Λ_k. Since $\{\mathbf{y}_1,\ \mathbf{y}_2,\ \ldots,\ \mathbf{y}_{-2},\ \mathbf{y}_{-1}\}$ are linearly independent, the full set of Ritz vectors evidently forms a basis for $\mathcal{P}$, which is furthermore both $\mathbf{K}$-orthogonal and $\mathbf{M}$-orthogonal and may be presumed to be $\mathbf{M}$-normalized without loss of generality: $\mathbf{u}_i^{\mathsf{k}}\mathbf{u}_j = \mathbf{u}_i^{\mathsf{m}}\mathbf{u}_j = 0$ for $i \neq j$, and $\mathbf{u}_i^{\mathsf{m}}\mathbf{u}_i = 1$.

Harmonic Ritz values [17] result from applying the Rayleigh-Ritz method to the eigenvalue problem

$$\mathbf{K}\mathbf{M}^{-1}\mathbf{K}\mathbf{x} = \lambda\, \mathbf{K}\mathbf{x}, \tag{2.3}$$

which is equivalent to (2.1) — it has the same eigenvalues and eigenvectors. If we use the same subspace $\mathcal{P}$, the harmonic Ritz values are then the eigenvalues of the $m \times m$ problem

$$\mathbf{P}^t\mathbf{K}\mathbf{M}^{-1}\mathbf{K}\mathbf{P}\mathbf{y} = \tilde{\Lambda}\, \mathbf{P}^t\mathbf{K}\mathbf{P}\mathbf{y}, \tag{2.4}$$

yielding

$$\tilde{\Lambda}_1 \leq \tilde{\Lambda}_2 \leq \tilde{\Lambda}_3 \leq \cdots \leq \tilde{\Lambda}_{-3} \leq \tilde{\Lambda}_{-2} \leq \tilde{\Lambda}_{-1}.$$

Just as Ritz values are weighted means of the eigenvalues of the matrix, $\Lambda_i = \sum_{j=1}^n \gamma_{ij}\lambda_j$ with $\gamma_{ij} \geq 0$ and $\sum_{j=1}^n \gamma_{ij} = 1$, harmonic Ritz values are harmonic means of the eigenvalues of the matrix,

$$\frac{1}{\tilde{\Lambda}_i} = \sum_{j=1}^{n} \tilde{\gamma}_{ij}\frac{1}{\lambda_j}.$$

with $\tilde{\gamma}_{ij} \geq 0$ and $\sum_{j=1}^n \tilde{\gamma}_{ij} = 1$. The term "harmonic Ritz value" occasionally has been used in a more general sense that incorporates a shift making it equivalent to a right-definite Lehmann value, $\Lambda^{(R)}$, introduced in the next section. For clarity, the narrower (shiftless) definition is used here.

Quantities which will be introduced here (for lack of a better name) as *dual harmonic Ritz* values result from applying the Rayleigh-Ritz method to the eigenvalue problem

$$\mathbf{M}\mathbf{x} = \lambda\, \mathbf{M}\mathbf{K}^{-1}\mathbf{M}\mathbf{x}, \tag{2.5}$$

which is also equivalent to (2.1), in the sense of having the same eigenvalues and eigenvectors. If we use the same approximating subspace $\mathcal{P}$, the dual harmonic Ritz values are the eigenvalues of the $m \times m$ problem

$$\mathbf{P}^t\mathbf{M}\mathbf{P}\mathbf{y} = \tilde{\tilde{\Lambda}}\, \mathbf{P}^t\mathbf{M}\mathbf{K}^{-1}\mathbf{M}\mathbf{P}\mathbf{y}, \tag{2.6}$$

yielding

$$\tilde{\tilde{\Lambda}}_1 \le \tilde{\tilde{\Lambda}}_2 \le \tilde{\tilde{\Lambda}}_3 \le \cdots \le \tilde{\tilde{\Lambda}}_{-3} \le \tilde{\tilde{\Lambda}}_{-2} \le \tilde{\tilde{\Lambda}}_{-1}.$$

Dual harmonic Ritz values are also harmonic means of the matrix eigenvalues, however with a different weighting than for harmonic Ritz values. Notice that the dual harmonic Ritz problem associated with $\mathcal{P}$ is equivalent to a harmonic Ritz problem associated with $\mathbf{K}^{-1}\mathbf{M}\mathcal{P}$, the subspace that would result after a single step of inverse iteration.

Both harmonic Ritz and dual harmonic Ritz values were known as long as 50 years ago and found to be useful in differential eigenvalue problems — Collatz [2] referred to the harmonic Ritz problem (2.4) as Grammel's equations (citing Grammel's earlier work [8]) and viewed the Rayleigh quotients for the Ritz problem (2.2), the harmonic Ritz problem (2.4), and the dual harmonic Ritz problem (2.6), all as elements of an infinite monotone sequence of "Schwarz quotients" that could be generated iteratively.

As long as $\mathbf{K}$ and $\mathbf{M}$ are positive definite, all three of Ritz, harmonic Ritz, and dual harmonic Ritz values provide "inner" bounds to the "outer" eigenvalues of the pencil $\mathbf{K}-\lambda\mathbf{M}$ (that is, of the problem (2.1)). In comparing the three types of approximations using the same subspace $\mathcal{P}$, harmonic Ritz values provide the best bounds of the three to the upper eigenvalues of (2.1); dual harmonic Ritz values provide the best bounds of the three to the lower eigenvalues. As an example, Figure 1 shows bounds obtained for a sequence of nested Krylov subspaces taken for $\mathcal{P}$, with $\mathbf{K} = diag([1,\, 3,\, 5,\, \ldots,\, 99])$, $\mathbf{M} = \mathbf{I}$, and a starting vector of all ones (the example of [17]).

The following result spells this out and is a special case of what Collatz demonstrated as "monotonicity of Schwartz quotients." The pattern of proof follows Collatz [2].

THEOREM 2.1. *Suppose* $\mathbf{K}$ *and* $\mathbf{M}$ *are positive definite. Then*

$$\begin{array}{rl} \lambda_k \;\le\; \tilde{\tilde{\Lambda}}_k \;\le\; \Lambda_k \;\le\; \tilde{\Lambda}_k & \textit{for} \quad k = 1,\, 2,\, \ldots \\ \tilde{\tilde{\Lambda}}_{-\ell} \;\le\; \Lambda_{-\ell} \;\le\; \tilde{\Lambda}_{-\ell} \;\le\; \lambda_{-\ell} & \textit{for} \quad \ell = 1,\, 2,\, \ldots \end{array}$$

Proof. The min-max characterization yields

$$\begin{aligned} \lambda_k = \min_{\dim \mathcal{S}=k} \max_{\mathbf{x}\in\mathcal{S}} \frac{\mathbf{x}^t\mathbf{K}\mathbf{x}}{\mathbf{x}^t\mathbf{M}\mathbf{x}} &\le \min_{\substack{\dim \mathcal{S}=k\\ \mathcal{S}\subset\mathcal{P}}} \max_{\mathbf{x}\in\mathcal{S}} \frac{\mathbf{x}^t\mathbf{K}\mathbf{x}}{\mathbf{x}^t\mathbf{M}\mathbf{x}} \\ &= \min_{\dim \mathcal{R}=k} \max_{\mathbf{y}\in\mathcal{R}} \frac{\mathbf{y}^t\mathbf{P}^t\mathbf{K}\mathbf{P}\mathbf{y}}{\mathbf{y}^t\mathbf{P}^t\mathbf{M}\mathbf{P}\mathbf{y}} = \Lambda_k, \end{aligned}$$

and likewise,

$$\begin{aligned} \lambda_k = \min_{\dim \mathcal{S}=k} \max_{\mathbf{x}\in\mathcal{S}} \frac{\mathbf{x}^t\mathbf{K}\mathbf{M}^{-1}\mathbf{K}\mathbf{x}}{\mathbf{x}^t\mathbf{K}\mathbf{x}} &\le \min_{\substack{\dim \mathcal{S}=k\\ \mathcal{S}\subset\mathcal{P}}} \max_{\mathbf{x}\in\mathcal{S}} \frac{\mathbf{x}^t\mathbf{K}\mathbf{M}^{-1}\mathbf{K}\mathbf{x}}{\mathbf{x}^t\mathbf{K}\mathbf{x}} \\ &= \min_{\dim \mathcal{R}=k} \max_{\mathbf{y}\in\mathcal{R}} \frac{\mathbf{y}^t\mathbf{P}^t\mathbf{K}\mathbf{M}^{-1}\mathbf{K}\mathbf{P}\mathbf{y}}{\mathbf{y}^t\mathbf{P}^t\mathbf{K}\mathbf{P}\mathbf{y}} = \tilde{\Lambda}_k. \end{aligned}$$

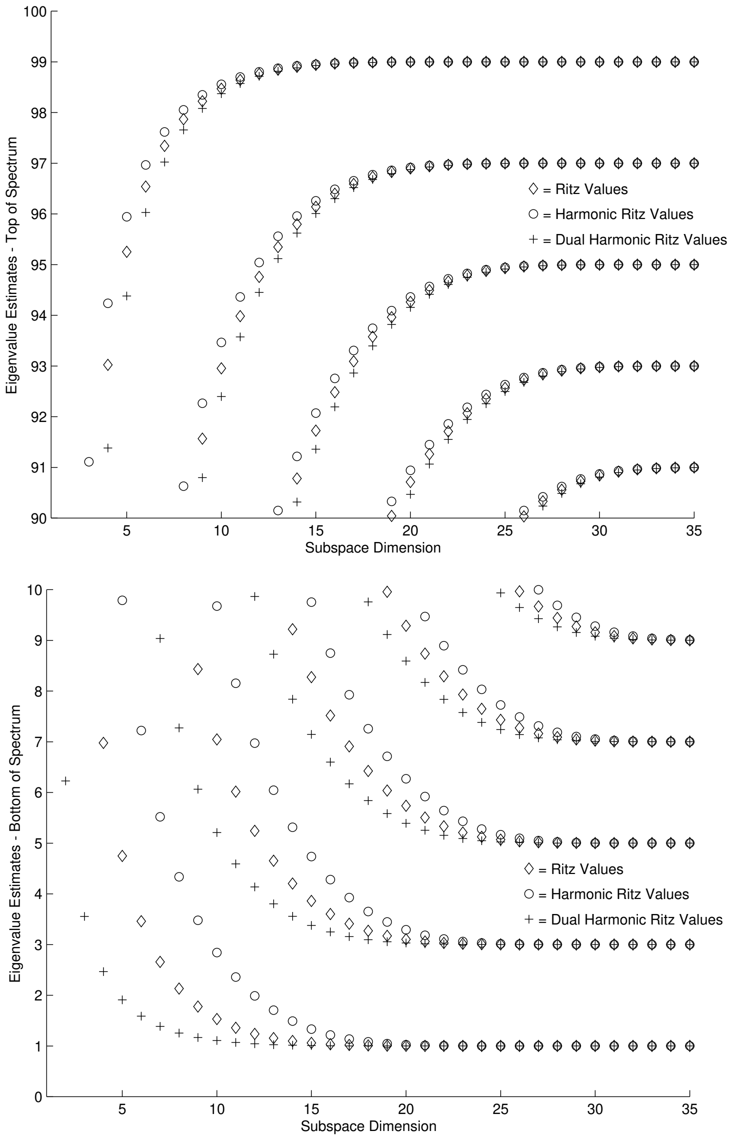

FIG. 2.1. *Comparison of bounds on upper and lower portions of the spectrum.*

A similar argument shows $\lambda_k \leq \tilde{\tilde{\Lambda}}_k$. By repeating the argument for the eigenvalue problem $-\mathbf{K}\mathbf{x} = (-\lambda)\mathbf{M}\mathbf{x}$, one finds $-\lambda_\ell(-\mathbf{K},\ \mathbf{M}) \leq -\Lambda_{-\ell}$ (where $\lambda(\mathbf{A},\ \mathbf{B})$ is used to denote an eigenvalue of the pencil $\mathbf{A} - \lambda\mathbf{B}$). Notice that $-\lambda_\ell(-\mathbf{K},\ \mathbf{M}) = \lambda_{-\ell}(\mathbf{K},\ \mathbf{M})$. Thus, $\Lambda_{-\ell} \leq \lambda_{-\ell}$ and $\tilde{\Lambda}_{-\ell} \leq \lambda_{-\ell}$.

For any $\mathbf{x} \in \mathbb{R}^n$, the Cauchy-Schwarz inequality implies

$$\begin{aligned} (\mathbf{x}^t\mathbf{K}\mathbf{x})^2 &= (\mathbf{x}^t\mathbf{K}\mathbf{M}^{-1/2}\mathbf{M}^{1/2}\mathbf{x})^2 \leq \mathbf{x}^t\mathbf{K}\mathbf{M}^{-1}\mathbf{K}\mathbf{x}\ \mathbf{x}^t\mathbf{M}\mathbf{x} \\ \text{and}\quad (\mathbf{x}^t\mathbf{M}\mathbf{x})^2 &= (\mathbf{x}^t\mathbf{M}\mathbf{K}^{-1/2}\mathbf{K}^{1/2}\mathbf{x})^2 \leq \mathbf{x}^t\mathbf{M}\mathbf{K}^{-1}\mathbf{M}\mathbf{x}\ \mathbf{x}^t\mathbf{K}\mathbf{x}. \end{aligned}$$

Thus,

$$\frac{\mathbf{x}^t\mathbf{M}\mathbf{x}}{\mathbf{x}^t\mathbf{M}\mathbf{K}^{-1}\mathbf{M}\mathbf{x}} \le \frac{\mathbf{x}^t\mathbf{K}\mathbf{x}}{\mathbf{x}^t\mathbf{M}\mathbf{x}} \le \frac{\mathbf{x}^t\mathbf{K}\mathbf{M}^{-1}\mathbf{K}\mathbf{x}}{\mathbf{x}^t\mathbf{K}\mathbf{x}},$$

which then implies for each $k = 1,\ 2,\ \ldots,\ m$

$$\begin{aligned} 0 < \lambda_k \le \tilde{\tilde{\Lambda}}_k &= \min_{\substack{\dim \mathcal{S}=k \\ \mathcal{S}\subset\mathcal{P}}} \max_{\mathbf{x}\in\mathcal{S}} \frac{\mathbf{x}^t\mathbf{M}\mathbf{x}}{\mathbf{x}^t\mathbf{M}\mathbf{K}^{-1}\mathbf{M}\mathbf{x}} \\ &\le \min_{\substack{\dim \mathcal{S}=k \\ \mathcal{S}\subset\mathcal{P}}} \max_{\mathbf{x}\in\mathcal{S}} \frac{\mathbf{x}^t\mathbf{K}\mathbf{x}}{\mathbf{x}^t\mathbf{M}\mathbf{x}} = \Lambda_k \\ &\le \min_{\substack{\dim \mathcal{S}=k \\ \mathcal{S}\subset\mathcal{P}}} \max_{\mathbf{x}\in\mathcal{S}} \frac{\mathbf{x}^t\mathbf{K}\mathbf{M}^{-1}\mathbf{K}\mathbf{x}}{\mathbf{x}^t\mathbf{K}\mathbf{x}} = \tilde{\Lambda}_k \qquad \square \end{aligned}$$

The situation is somewhat different if $\mathbf{K}$ is *indefinite*. The Ritz estimates are still "inner" bounds, that is $\lambda_k \le \Lambda_k$ and $\Lambda_{-\ell} \le \lambda_{-\ell}$. However, both harmonic Ritz and dual harmonic Ritz values now provide "outer" bounds (*lower* bounds) to negative eigenvalues of (2.1) and no simple relationship is known that would predict which of the three bounds is best (essentially owing to there being no simple analog of the Cauchy-Schwarz inequality for indefinite inner products).

Despite the differences in behavior described above, Ritz, harmonic Ritz, and dual harmonic Ritz values each provide *optimal* bounds – obviously each with respect to a slightly different notion of optimality. For the Ritz problem, the matrices $\mathbf{P}^t\mathbf{K}\mathbf{P}$ and $\mathbf{P}^t\mathbf{M}\mathbf{P}$ provide a "sampling" of the full matrices $\mathbf{K}$ and $\mathbf{M}$ on the subspace $\mathcal{P}$. Whatever spectral information about the original eigenvalue problem (2.1) that we are able to deduce by examining the Rayleigh-Ritz problem (2.2) we must draw the same conclusions for *all* matrix pencils that are "aliased" by the Rayleigh-Ritz sampling. Define the following set of such $n \times n$ matrix pairs:

$$\mathcal{C}(\mathcal{P}) = \left\{ (\mathbf{A},\ \mathbf{B}) \;\middle|\; \begin{array}{c} \mathbf{A} \text{ and } \mathbf{B} \text{ are positive definite} \\ \mathbf{P}^t(\mathbf{A}-\mathbf{K})\mathbf{P} = 0 \\ \mathbf{P}^t(\mathbf{B}-\mathbf{M})\mathbf{P} = 0 \end{array} \right\}$$

THEOREM 2.2. *For any choice of positive integers $\nu,\ \pi$ with $\nu + \pi = m$ and any choice of matrix pairs $(\mathbf{A},\ \mathbf{B}) \in \mathcal{C}(\mathcal{P})$*

$$\begin{aligned} \lambda_k(\mathbf{A},\ \mathbf{B}) \le \Lambda_k \quad &\textit{for} \quad k = 1,\ 2,\ \ldots,\ \nu \\ \Lambda_{-\ell} \le \lambda_{-\ell}(\mathbf{A},\ \mathbf{B}) \quad &\textit{for} \quad \ell = 1,\ 2,\ \ldots,\ \pi. \end{aligned}$$

Furthermore, for each index pair $\nu,\ \pi$, there exists a matrix pair $(\hat{\mathbf{A}},\ \hat{\mathbf{B}}) \in \mathcal{C}(\mathcal{P})$ such that

$$\begin{aligned} \lambda_k(\hat{\mathbf{A}},\ \hat{\mathbf{B}}) = \Lambda_k \quad &\textit{for} \quad k = 1,\ 2,\ \ldots,\ \nu \\ \Lambda_{-\ell} = \lambda_{-\ell}(\hat{\mathbf{A}},\ \hat{\mathbf{B}}) \quad &\textit{for} \quad \ell = 1,\ 2,\ \ldots,\ \pi. \end{aligned}$$

So, no better bounds are possible with only the information available to the Rayleigh-Ritz method as described by (2.2).

Proof. The first assertion is a restatement of Theorem 2.1 for the matrix pencil $\mathbf{A} - \lambda\mathbf{B}$. To show optimality, define the matrix of Ritz vectors:

$$\mathbf{U} = [\mathbf{u}_1,\ \mathbf{u}_2,\ \ldots, \mathbf{u}_\nu,\ \mathbf{u}_{-\pi},\ \ldots,\ \mathbf{u}_{-2},\ \mathbf{u}_{-1}].$$

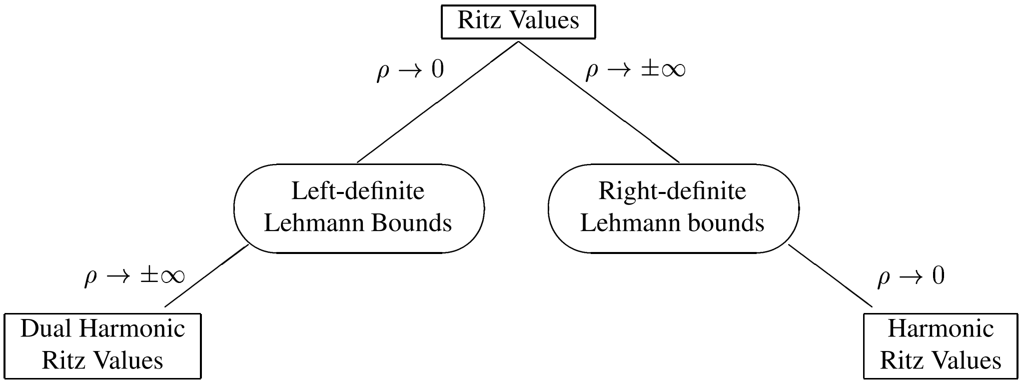

FIG. 2.2. *A taxonomy of eigenvalue estimates.*

Notice that $\mathbf{U}$ is an $\mathbf{M}$-orthonormal basis for $\mathcal{P}$: $\mathbf{U}^t\mathbf{M}\mathbf{U} = \mathbf{I}$. Define also the diagonal matrix of Ritz values

$$\mathbf{D} = \begin{bmatrix} \Lambda_1 & & & & & \\ & \ddots & & & & \\ & & \Lambda_\nu & & & \\ & & & \Lambda_{-\pi} & & \\ & & & & \ddots & \\ & & & & & \Lambda_{-1} \end{bmatrix}$$

and fix $\hat{\Lambda} = \frac{1}{2}(\Lambda_\nu + \Lambda_{-\pi})$. Now, consider

$$\hat{\mathbf{A}} = \mathbf{M}\mathbf{U}\mathbf{D}\mathbf{U}^t\mathbf{M} + \hat{\Lambda}(\mathbf{M} - \mathbf{M}\mathbf{U}\mathbf{U}^t\mathbf{M}) \quad \text{and} \quad \hat{\mathbf{B}} = \mathbf{M}.$$

One may verify that all required conditions are satisfied, in particular

$$(\hat{\mathbf{A}} - \lambda\hat{\mathbf{B}})\mathbf{U} = \mathbf{M}\mathbf{U}(\mathbf{D} - \lambda\mathbf{I}),$$

and for any $\mathbf{v} \in \mathbb{R}^n$ with $\mathbf{v}^t\mathbf{M}\mathbf{U} = 0$,

$$(\hat{\mathbf{A}} - \lambda\hat{\mathbf{B}})\mathbf{v} = \mathbf{M}\mathbf{v}(\hat{\Lambda} - \lambda). \qquad \square$$

A similar construction can be used to show the (analogously defined) optimality of harmonic Ritz values and dual harmonic Ritz values.

As we will see in following sections, Ritz values, harmonic Ritz values, and dual harmonic Ritz values are limiting cases of parameterized families of bounds arising from "left-definite" and "right-definite" Lehmann intervals.

3. Lehmann's Optimal Intervals. Each of the Ritz-related methods discussed above will have certain advantages in estimating the extreme eigenvalues of (2.1). None are particularly effective in estimating interior eigenvalues, however. Usual strategies for obtaining accurate estimates to the eigenvalues of (2.1) lying close to a given value ρ involve a spectral mapping that turns the spectrum "inside out" around ρ — mapping interior eigenvalues in the neighborhood of ρ to extreme eigenvalues that are more accessible. "Shift and invert" strategies typically use the spectral mapping $\lambda \mapsto \frac{1}{\lambda-\rho}$. A variant used especially for buckling problems (where $\mathbf{M}$ may be singular) utilizes instead the spectral mapping $\lambda \mapsto \frac{\lambda}{\lambda-\rho}$. As we shall see, both of these spectral mappings play a fundamental role in the optimal bounds discovered by Lehmann ([11], [12], [13]). The derivation used here is in the spirit of that given by Maehly in [14], and the associated methods are sometimes called Lehmann-Maehly methods.

Fix a scalar ρ that is not an eigenvalue of (2.1) and define the index r to satisfy

$$\lambda_{r-1} < \rho < \lambda_r. \tag{3.1}$$

The *right-definite Lehmann method* follows first from considering the spectral mapping $\lambda \mapsto \frac{1}{\lambda-\rho}$ and an associated eigenvalue problem equivalent to (2.1):

$$\mathbf{M}(\mathbf{K}-\rho\mathbf{M})^{-1}\mathbf{M}\mathbf{x} = \frac{1}{\lambda-\rho}\mathbf{M}\mathbf{x}, \tag{3.2}$$

which has eigenvalues distributed as

$$\frac{1}{\lambda_{r-1}-\rho} \le \frac{1}{\lambda_{r-2}-\rho} \le \cdots \le 0 \le \cdots \le \frac{1}{\lambda_{r+1}-\rho} \le \frac{1}{\lambda_r-\rho}.$$

Notice that eigenvalues of (2.1) flanking ρ are mapped to extremal eigenvalues of (3.2). Now use an m-dimensional subspace $\mathcal{S}$, spanned by the columns of a matrix $\mathbf{S}$ to generate Rayleigh-Ritz estimates for the eigenvalues of (3.2):

$$[\mathbf{S}^t\mathbf{M}(\mathbf{K}-\rho\mathbf{M})^{-1}\mathbf{M}\mathbf{S}]\mathbf{y} = R\,[\mathbf{S}^t\mathbf{M}\mathbf{S}]\mathbf{y}, \tag{3.3}$$

where $\mathbf{S} \in \mathbb{R}^{n\times m}$. Suppose (3.3) has ν negative eigenvalues $R_1 \le \cdots \le R_\nu < 0$ and $\pi = m-\nu$ positive eigenvalues $0 < R_{-\pi} \le \cdots \le R_{-1}$. Regardless of the subspace $\mathcal{S}$ that is chosen, the min-max principle (or Theorem 2.1) guarantees that, for each $k = 1,\ 2,\ \ldots,\nu$ and $\ell = 1,\ 2,\ \ldots,\pi$,

$$\frac{1}{\lambda_{r-k}-\rho} \le R_k \quad \text{and} \quad R_{-\ell} \le \frac{1}{\lambda_{r+\ell-1}-\rho}.$$

Rearrange and introduce

$$\Lambda_{-k}^{(R)} \stackrel{def}{=} \rho + \frac{1}{R_k} \le \lambda_{r-k} \quad \text{and} \quad \lambda_{r+\ell-1} \le \rho + \frac{1}{R_{-\ell}} \stackrel{def}{=} \Lambda_\ell^{(R)} \tag{3.4}$$

for $k = 1,\ 2,\ \ldots,\nu$ and $\ell = 1,\ 2,\ \ldots,\pi$. Notice that labeling of $\Lambda^{(R)}$ is arranged relative to ρ:

$$\ldots\Lambda_{-3}^{(R)} \le \Lambda_{-2}^{(R)} \le \Lambda_{-1}^{(R)} < \rho < \Lambda_1^{(R)} \le \Lambda_2^{(R)} \le \Lambda_3^{(R)}\ldots$$

An equivalent statement combining (3.1) and (3.4) is

> Each of the intervals $[\Lambda_{-k}^{(R)},\ \rho)$ and $(\rho,\ \Lambda_\ell^{(R)}]$ contain respectively at least k and ℓ eigenvalues of (2.1) for $k = 1,\ 2,\ \ldots,\nu$ and $\ell = 1,\ 2,\ \ldots,\pi$.

To avoid the need in (3.3) for solving linear systems having the indefinite coefficient matrix $(\mathbf{K}-\rho\mathbf{M})$, change variables in (3.3) as $\mathbf{P} = (\mathbf{K}-\rho\mathbf{M})^{-1}\mathbf{M}\mathbf{S}$ — which then *implicitly* determines $\mathcal{S}$ via a choice of $\mathcal{P}$. (3.3) can then be rewritten as

$$[\mathbf{P}^t(\mathbf{K}-\rho\mathbf{M})\mathbf{P}]\mathbf{y} = R\,[\mathbf{P}^t(\mathbf{K}-\rho\mathbf{M})\mathbf{M}^{-1}(\mathbf{K}-\rho\mathbf{M})\mathbf{P}]\mathbf{y}, \tag{3.5}$$

When $\dim \mathcal{P} = 1$, (3.4) becomes *Temple's inequality*

$$\rho + \frac{\mathbf{p}^t(\mathbf{K} - \rho\mathbf{M})\mathbf{M}^{-1}(\mathbf{K} - \rho\mathbf{M})\mathbf{p}}{\mathbf{p}^t(\mathbf{K} - \rho\mathbf{M})\mathbf{p}} = \frac{\mathbf{p}^t(\mathbf{K}\mathbf{M}^{-1}\mathbf{K} - \rho\mathbf{K})\mathbf{p}}{\mathbf{p}^t(\mathbf{K} - \rho\mathbf{M})\mathbf{p}} \leq \lambda_{r-1}.$$

Some additional notation will reduce the impending clutter of symbols. Introduce matrices of *Schwarz constants*:

$$\mathbf{H}_0 \stackrel{def}{=} [\mathbf{P}^t\mathbf{K}\mathbf{M}^{-1}\mathbf{K}\mathbf{P}],\ \mathbf{H}_1 \stackrel{def}{=} [\mathbf{P}^t\mathbf{K}\mathbf{P}],\ \text{ and }\ \mathbf{H}_2 \stackrel{def}{=} [\mathbf{P}^t\mathbf{M}\mathbf{P}].$$

Then expanding out the various terms, (3.5) becomes

$$[\mathbf{H}_1 - \rho\mathbf{H}_2]\mathbf{y} = R\,[\mathbf{H}_0 - 2\rho\mathbf{H}_1 + \rho^2\mathbf{H}_2]\mathbf{y} \tag{3.6}$$

which may be rearranged to obtain

$$[\mathbf{H}_0 - \rho\mathbf{H}_1]\mathbf{y} = \Lambda^{(R)}\,[\mathbf{H}_1 - \rho\mathbf{H}_2]\mathbf{y}. \tag{3.7}$$

Notice that (3.7) could be written in terms of the $\mathbf{M}$-inner product as

$$\mathbf{P}^{\mathsf{m}}[(\mathbf{M}^{-1}\mathbf{K})^2 - \rho(\mathbf{M}^{-1}\mathbf{K})]\mathbf{P}\mathbf{y} = \Lambda^{(R)}\,\mathbf{P}^{\mathsf{m}}[(\mathbf{M}^{-1}\mathbf{K}) - \rho\mathbf{I}]\mathbf{P}\mathbf{y} \tag{3.8}$$

or in terms of the $\mathbf{K}$-inner product as

$$\mathbf{P}^{\mathsf{k}}[(\mathbf{M}^{-1}\mathbf{K}) - \rho\mathbf{I}]\mathbf{P}\mathbf{y} = \Lambda^{(R)}\,\mathbf{P}^{\mathsf{k}}[\mathbf{I} - \rho(\mathbf{M}^{-1}\mathbf{K})^{-1}]\mathbf{P}\mathbf{y}. \tag{3.9}$$

The *left-definite Lehmann method* can be obtained by considering the spectral mapping $\lambda \mapsto \frac{\lambda}{\lambda-\rho}$ and an associated eigenvalue problem — also equivalent to (2.1):

$$\mathbf{K}(\mathbf{K} - \rho\mathbf{M})^{-1}\mathbf{K}\mathbf{x} = \frac{\lambda}{\lambda - \rho}\mathbf{K}\mathbf{x} \tag{3.10}$$

which has eigenvalues distributed as

$$\frac{\lambda_{r-1}}{\lambda_{r-1} - \rho} \leq \frac{\lambda_{r-2}}{\lambda_{r-2} - \rho} \leq \cdots < 0 \text{ and } 1 < \cdots \leq \frac{\lambda_{r+1}}{\lambda_{r+1} - \rho} \leq \frac{\lambda_r}{\lambda_r - \rho} \tag{3.11}$$

(as long as both $\mathbf{K}$ and $\mathbf{M}$ are positive definite, no eigenvalue gets mapped into the interval $[0, 1]$). Again the eigenvalues of (2.1) flanking ρ are mapped to extremal eigenvalues of (3.10). Using an m-dimensional subspace $\mathcal{T}$ (spanned by the columns of a matrix $\mathbf{T}$), one may generate Rayleigh-Ritz estimates for the eigenvalues of (3.10):

$$[\mathbf{T}^t\mathbf{K}(\mathbf{K} - \rho\mathbf{M})^{-1}\mathbf{K}\mathbf{T}]\mathbf{y} = L[\mathbf{T}^t\mathbf{K}\mathbf{T}]\mathbf{y}, \tag{3.12}$$

where $\mathbf{T} \in \mathbb{R}^{n\times m}$.

If (3.12) has ν negative eigenvalues $L_1 \leq L_2 \leq \cdots \leq L_\nu < 0$ and $\pi = m - \nu$ positive eigenvalues $1 < L_{-\pi} \leq \cdots \leq L_{-2} \leq L_{-1}$, then regardless of the subspace $\mathcal{T}$ that is chosen, the min-max principle (or again, Theorem 2.1) guarantees that

$$\frac{\lambda_{r-k}}{\lambda_{r-k} - \rho} \leq L_k \text{ and } L_{-\ell} \leq \frac{\lambda_{r+\ell-1}}{\lambda_{r+\ell-1} - \rho} \tag{3.13}$$

or equivalently that

$$\Lambda^{(L)}_{-k} \stackrel{def}{=} \rho - \frac{\rho}{1-L_k} \leq \lambda_{r-k} \text{ and } \lambda_{r+\ell-1} \leq \rho - \frac{\rho}{1-L_{-\ell}} \stackrel{def}{=} \Lambda^{(L)}_{\ell} \tag{3.14}$$

for $k = 1,\ 2,\ \ldots,\nu$ and $\ell = 1,\ 2,\ \ldots,\pi$. Just as for $\Lambda^{(R)}$, the labeling of $\Lambda^{(L)}$ is done relative to ρ:

$$\ldots\Lambda^{(L)}_{-3} \leq \Lambda^{(L)}_{-2} \leq \Lambda^{(L)}_{-1} < \rho < \Lambda^{(L)}_{1} \leq \Lambda^{(L)}_{2} \leq \Lambda^{(L)}_{3} \ldots .$$

An equivalent statement combining (3.1) and (3.14) is

> Each of the intervals $[\Lambda^{(L)}_{-k},\ \rho)$ and $(\rho,\ \Lambda^{(L)}_{\ell}]$ contain respectively at least k and ℓ eigenvalues of (2.1) for $k = 1,\ 2,\ \ldots,\nu$ and $\ell = 1,\ 2,\ \ldots,\pi$.

As before, in order to avoid solving systems with the indefinite coefficient matrix $(\mathbf{K} - \rho\mathbf{M})$, change variables in (3.12) as $\mathbf{P} = (\mathbf{K} - \rho\mathbf{M})^{-1}\mathbf{K}\mathbf{T}$ which then *implicitly* determines $\mathcal{T}$ via a choice of $\mathcal{P}$. (3.12) can then be rewritten as

$$[\mathbf{P}^t(\mathbf{K} - \rho\mathbf{M})\mathbf{P}]\mathbf{y} = L\,[\mathbf{P}^t(\mathbf{K} - \rho\mathbf{M})\mathbf{K}^{-1}(\mathbf{K} - \rho\mathbf{M})\mathbf{P}]\mathbf{y}. \tag{3.15}$$

Introduce

$$\mathbf{H}_3 \stackrel{def}{=} [\mathbf{P}^t\mathbf{M}\mathbf{K}^{-1}\mathbf{M}\mathbf{P}].$$

Then (3.15) becomes

$$[\mathbf{H}_1 - \rho\mathbf{H}_2]\mathbf{y} = L[\mathbf{H}_1 - 2\rho\mathbf{H}_2 + \rho^2\mathbf{H}_3]\mathbf{y} \tag{3.16}$$

which may be rearranged to get

$$[\mathbf{H}_1 - \rho\mathbf{H}_2]\mathbf{y} = \Lambda^{(L)}\,[\mathbf{H}_2 - \rho\mathbf{H}_3]\mathbf{y}. \tag{3.17}$$

Observe that both (3.6) and (3.16) are Hermitian definite pencils with the same left-hand side. By the Sylvester Law of Inertia, they each have the same number of negative (and hence positive) eigenvalues. If a shift of $\rho = 0$ is chosen in (3.7), the harmonic Ritz problem (2.4) is obtained and $\tilde{\Lambda}_\ell = \Lambda^{(R)}_\ell\,|_{\rho=0}$. As $\rho \to \pm\infty$, (3.7) reduces to the Ritz problem (2.2). Similarly, if a shift of $\rho = 0$ is chosen in (3.17), the Ritz problem (2.2) is obtained and $\Lambda_\ell = \Lambda^{(L)}_\ell\,|_{\rho=0}$. As $\rho \to \pm\infty$, (3.17) reduces to the dual harmonic Ritz problem (2.6).

The left- and right-definite Lehmann bounds, $\Lambda^{(L)}$ and $\Lambda^{(R)}$, that are below the parameter ρ are roughly monotone increasing with respect to ρ. Goerisch [4] discovered this for ρ satisfying (3.1) by working from the optimality of the Lehmann bounds (in essence, a larger $\rho < \lambda_r$ places more restrictions on the "aliasing" operators, thus improving the bounds below ρ). As ρ is increased further, the r in (3.1) changes and the labeling of $\Lambda^{(L)}$ and $\Lambda^{(R)}$ shifts. Monotonicity might not hold for each subspace, however Goerisch [4] showed that in the infinite dimensional setting, provided conditions for convergence hold, monotonicity can be guaranteed asymptotically. This more complicated circumstance for ρ bridging across λ_r is discussed in [19].

Notice that (3.17) could be obtained formally from the right-definite method expressed in (3.9) by direct substitution of the $\mathbf{M}$-inner product for the $\mathbf{K}$-inner product.

$$\mathbf{P}^{\mathsf{m}}[(\mathbf{M}^{-1}\mathbf{K}) - \rho\mathbf{I}]\mathbf{P}\mathbf{y} = \Lambda^{(L)}\,\mathbf{P}^{\mathsf{m}}[\mathbf{I} - \rho(\mathbf{M}^{-1}\mathbf{K})^{-1}]\mathbf{P}\mathbf{y}. \tag{3.18}$$

Such a substitution also converts the harmonic Ritz problem into a Ritz problem and the Ritz problem, then into a dual harmonic Ritz problem. This provides some impetus to call the "left-definite Lehmann" method the "harmonic Lehmann" method, but Lehmann himself referred to this method as "left-definite" and besides the correspondences are a bit backward since (right-definite) Lehmann is to Ritz as "dual harmonic Ritz" is to " harmonic Lehmann."

4. Alternative Formulations. Kahan developed a formulation of Lehmann's right-definite method that is particularly well-suited to many computational settings for matrix eigenvalue problems (cf. [16], Chap. 10). We review that development here and extend it to Lehmann's left-definite method in Theorems 4.1 and 4.2. A different tack leads to a re-formulation of right-definite Lehmann bounds as singular values of a related matrix. We will discover that a similar reformulation of left-definite bounds leads to a generalized singular value decomposition of a related pair of matrices.

For a given m-dimensional subspace $\mathcal{P}$, suppose the columns of $\mathbf{Q}_1$ provide an $\mathbf{M}$-orthonormal basis for $\mathcal{P}$: $Ran(\mathbf{Q}_1) = \mathcal{P}$ ("Ran" denotes the range space of a matrix) and $\mathbf{Q}_1^m\mathbf{Q}_1 = \mathbf{Q}_1^t\mathbf{M}\mathbf{Q}_1 = \mathbf{I}$. Define $\mathbf{H}$ from the "residual orthogonality" condition

$$(\mathbf{M}^{-1}\mathbf{K}\mathbf{Q}_1 - \mathbf{Q}_1\mathbf{H})^t\mathbf{M}\mathbf{Q}_1 = 0,$$

so that $\mathbf{H} = \mathbf{Q}_1^t\mathbf{K}\mathbf{Q}_1$ and observe (say, from the Gram-Schmidt process) that there is an upper triangular matrix $\mathbf{C}$ and a matrix $\mathbf{Q}_2$ with $\mathbf{M}$-orthonormal columns, so that

$$\mathbf{Q}_2\mathbf{C} = \mathbf{M}^{-1}\mathbf{K}\mathbf{Q}_1 - \mathbf{Q}_1\mathbf{H}.$$

Pick $\mathbf{Q}_3$ to fill out an $\mathbf{M}$-orthonormal basis for $\mathbb{R}^n$ in conjunction with $\mathbf{Q}_1$ and $\mathbf{Q}_2$. Then with $\mathbf{Q} = [\mathbf{Q}_1\ \mathbf{Q}_2\ \mathbf{Q}_3]$, we have $\mathbf{Q}^t\mathbf{M}\mathbf{Q} = \mathbf{I}$ and

$$\mathbf{M}^{-1}\mathbf{K}\mathbf{Q} = \mathbf{Q}\begin{bmatrix} \mathbf{H} & \mathbf{C}^t & 0 \\ \mathbf{C} & \mathbf{V}_{11} & \mathbf{V}_{21}^t \\ 0 & \mathbf{V}_{21} & \mathbf{V}_{22} \end{bmatrix} \qquad \begin{array}{l} \text{where} \\ \mathbf{H} \text{ is } m \times m \\ \mathbf{V}_{11} \text{ is } k \times k. \end{array}$$

While this shows how $\mathbf{H}$ and $\mathbf{C}$ might be constructed (essentially one step of a block Lanczos process), there may be other situations of interest when $\mathbf{H}$ and $\mathbf{C}$ are known *a priori*. In any case, we assume that the bottom right block 2×2 submatrix, $\mathbf{V}$, is either unknown or at least unpleasant to deal with. With additional unitary massage, $rank(\mathbf{C}) = k$ could be assumed (possibly resulting in a smaller $\mathbf{V}_{11}$), though it isn't necessary in what follows. The situation $rank(\mathbf{C}) = k \ll m \ll n$ is common. What follows is a *deus ex machina* development of Kahan's formulation of Lehmann bounds that offers brevity but little of the insight and revelation that one may find in the excellent discussion of ([16], Chapter 10).

Apply the right-definite Lehmann bounds from (3.5) using $\mathbf{P} = \mathbf{Q}_1$. Then, $(\mathbf{K} - \rho\mathbf{M})\mathbf{P} = \mathbf{Q}_1(\mathbf{H} - \rho\mathbf{I}) + \mathbf{Q}_2\mathbf{C}$ and the right-definite Lehmann problem (3.6) appears as

$$(\mathbf{H} - \rho\mathbf{I})\mathbf{y} = R\ \left[(\mathbf{H} - \rho\mathbf{I})^2 + \mathbf{C}^t\mathbf{C}\right]\mathbf{y}. \tag{4.1}$$

The associated right-definite bound is $\Lambda^{(R)} = \rho + 1/R$ and we may manipulate (4.1) to get an equivalent condition on $\Lambda^{(R)}$:

$$0 = \left[(\mathbf{H} - \rho\mathbf{I})(\mathbf{H} - \Lambda^{(R)}\mathbf{I}) + \mathbf{C}^t\mathbf{C}\right]\mathbf{y}. \tag{4.2}$$

One may recognize that the coefficient matrix of (4.2) is a Schur complement of the $(m + k) \times (m + k)$ matrix

$$\mathbf{Y}(\Lambda^{(R)}) \stackrel{def}{=} \begin{bmatrix} -(\mathbf{H} - \rho\mathbf{I})(\mathbf{H} - \Lambda^{(R)}\mathbf{I}) & \mathbf{C}^t \\ \mathbf{C} & \mathbf{I} \end{bmatrix}.$$

Hence, (4.2) has a non-trivial solution if and only if $\mathbf{Y}(\Lambda^{(R)})$ is singular. Suppose that neither ρ nor $\Lambda^{(R)}$ are eigenvalues of $\mathbf{H}$ for the time being and define

$$\begin{aligned} \mathbf{L}_1 &\stackrel{def}{=} \begin{bmatrix} \mathbf{I} & 0 \\ \mathbf{C}(\mathbf{H}-\rho\mathbf{I})^{-1}(\mathbf{H}-\Lambda^{(R)}\mathbf{I})^{-1} & \mathbf{I} \end{bmatrix}, \\ \mathbf{L}_2 &\stackrel{def}{=} \begin{bmatrix} \mathbf{I} & 0 \\ \mathbf{C}(\mathbf{H}-\Lambda^{(R)}\mathbf{I})^{-1} & \mathbf{I} \end{bmatrix}, \\ \text{and} \quad \mathbf{D}(\Lambda^{(R)}) &\stackrel{def}{=} \begin{bmatrix} -(\mathbf{H}-\rho\mathbf{I})^{-1} & 0 \\ 0 & (\rho-\Lambda^{(R)})\mathbf{I} \end{bmatrix}. \end{aligned}$$

Then

$$\mathbf{L}_2\mathbf{D}(\Lambda^{(R)})\mathbf{L}_1\mathbf{Y}(\Lambda^{(R)})\mathbf{L}_1^t\mathbf{L}_2^t = \begin{bmatrix} \mathbf{H}-\Lambda^{(R)}\mathbf{I} & \mathbf{C}^t \\ \mathbf{C} & \rho\mathbf{I}+\mathbf{C}(\mathbf{H}-\rho\mathbf{I})^{-1}\mathbf{C}^t-\Lambda^{(R)}\mathbf{I} \end{bmatrix}.$$

Thus $\Lambda^{(R)}$ is an eigenvalue of the $(m+k)\times(m+k)$ matrix

$$\begin{bmatrix} \mathbf{H} & \mathbf{C}^t \\ \mathbf{C} & \rho\mathbf{I}+\mathbf{C}(\mathbf{H}-\rho\mathbf{I})^{-1}\mathbf{C}^t \end{bmatrix} \tag{4.3}$$

if and only if either $\mathbf{D}(\Lambda^{(R)})$ is singular or $\mathbf{Y}(\Lambda^{(R)})$ is singular, which is to say, if and only if either $\Lambda^{(R)}$ is a right-definite Lehmann bound satisfying (4.2) or $\Lambda^{(R)} = \rho$ (which will occur with multiplicity k). A limiting argument can be mustered to handle the exceptional cases where either ρ or $\Lambda^{(R)}$ are eigenvalues of $\mathbf{H}$. In situations where either the smaller eigenvalues of (2.1) are of interest or $\|\mathbf{C}\|$ is much smaller than $\|\mathbf{H}\|$, finding the eigenvalues of (4.3) is likely to yield substantially more accurate results for $\Lambda^{(R)}$ then a direct attack on (4.1). A similar formulation for left-definite Lehmann problems will be described below.

Consider the application of the left-definite problem (3.16) with $\mathbf{P} = \mathbf{Q}_1$. Note that $\mathbf{K}\mathbf{Q}_1 = \mathbf{Q}_1\mathbf{H} + \mathbf{Q}_2\mathbf{C}$ implies that

$$\mathbf{K}^{-1}\mathbf{Q}_1 = \mathbf{Q}_1\mathbf{H}^{-1} - \mathbf{K}^{-1}\mathbf{Q}_2\mathbf{C}\mathbf{H}^{-1}$$

so then

$$\mathbf{Q}_1^t\mathbf{K}^{-1}\mathbf{Q}_1 = \mathbf{H}^{-1} + \mathbf{H}^{-1}\mathbf{C}^t\mathbf{W}\mathbf{C}\mathbf{H}^{-1} \tag{4.4}$$

where $\mathbf{W} = \mathbf{Q}_2^t\mathbf{K}^{-1}\mathbf{Q}_2$ has been introduced. (3.16) becomes

$$(\mathbf{H}-\rho\mathbf{I})\mathbf{y} = L\,\left[(\mathbf{H}-\rho\mathbf{I}) - \rho(\mathbf{I}-\rho(\mathbf{H}^{-1}+\mathbf{H}^{-1}\mathbf{C}^t\mathbf{W}\mathbf{C}\mathbf{H}^{-1}))\right]\mathbf{y}. \tag{4.5}$$

The associated left-definite bound is $\Lambda^{(L)} = -\rho L/(1-L)$ and we may manipulate (4.5) to get an equivalent condition on $\Lambda^{(L)}$:

$$0 = \left[(\mathbf{H}-\rho\mathbf{I})(\mathbf{H}-\Lambda^{(L)}\mathbf{I})\mathbf{H} + \rho\Lambda^{(L)}\mathbf{C}^t\mathbf{W}\mathbf{C}\right]\mathbf{y}. \tag{4.6}$$

Equation (4.6) has a non-trivial solution if and only if the $(m+k)\times(m+k)$ matrix

$$\hat{\mathbf{Y}}(\Lambda^{(L)}) \stackrel{def}{=} \begin{bmatrix} -(\mathbf{H}-\rho\mathbf{I})(\mathbf{H}-\Lambda^{(L)}\mathbf{I})\mathbf{H} & \Lambda^{(L)}\mathbf{C}^t \\ \rho\mathbf{C} & \mathbf{W}^{-1} \end{bmatrix}$$

is singular. Suppose that neither ρ nor $\Lambda^{(L)}$ are eigenvalues of $\mathbf{H}$, and define

$$\begin{aligned}
\mathbf{F} &\stackrel{def}{=} (\mathbf{H}-\rho\mathbf{I})^{-1}(\mathbf{H}-\Lambda^{(L)}\mathbf{I})^{-1}\mathbf{H}^{-1} \\
\hat{\mathbf{L}}_1 &\stackrel{def}{=} \begin{bmatrix} \mathbf{I} & 0 \\ \rho\mathbf{C}\mathbf{F} & \mathbf{I} \end{bmatrix}, \quad \hat{\mathbf{U}}_1 \stackrel{def}{=} \begin{bmatrix} \mathbf{I} & \Lambda^{(L)}\mathbf{F}\mathbf{C}^t \\ 0 & \mathbf{I} \end{bmatrix}, \\
\hat{\mathbf{L}}_2 &\stackrel{def}{=} \begin{bmatrix} \mathbf{I} & 0 \\ \mathbf{C}(\mathbf{H}-\Lambda^{(L)}\mathbf{I})^{-1} & \mathbf{I} \end{bmatrix}, \\
\text{and} \quad \hat{\mathbf{D}}(\Lambda^{(L)}) &\stackrel{def}{=} \begin{bmatrix} -(\mathbf{H}-\rho\mathbf{I})^{-1}\mathbf{H}^{-1} & 0 \\ 0 & (\rho-\Lambda^{(L)})/\rho\mathbf{I} \end{bmatrix}.
\end{aligned}$$

Then

$$\hat{\mathbf{L}}_2\hat{\mathbf{D}}(\Lambda^{(L)})\hat{\mathbf{L}}_1\hat{\mathbf{Y}}(\Lambda^{(L)})\hat{\mathbf{U}}_1\hat{\mathbf{L}}_2^t = \begin{bmatrix} \mathbf{H}-\Lambda^{(L)}\mathbf{I} & \mathbf{C}^t \\ \mathbf{C} & \frac{\rho-\Lambda^{(L)}}{\rho}\mathbf{N}_1 + \frac{\Lambda^{(L)}}{\rho}\mathbf{N}_2 \end{bmatrix}, \tag{4.7}$$

where $\mathbf{N}_1 = \mathbf{W}^{-1} + \mathbf{C}\mathbf{H}^{-1}\mathbf{C}^t$ and $\mathbf{N}_2 = \mathbf{C}(\mathbf{H}-\rho\mathbf{I})^{-1}\mathbf{C}^t$. Thus $\Lambda^{(L)}$ is an eigenvalue of an auxiliary $(m+k)\times(m+k)$ matrix pencil — not unlike the right-definite case. This matrix pencil will be *definite* when $\mathbf{N}_1 - \mathbf{N}_2$ is positive-definite, which in turn can be guaranteed when the $(r-1)st$ Ritz value is a sufficiently accurate approximation to λ_{r-1}:

THEOREM 4.1. *Suppose ρ is not an eigenvalue of (2.1). Each interval $[\Lambda_{-i}^{(L)},\ \rho)$ and $(\rho,\ \Lambda_j^{(L)}]$ contains respectively at least i and j eigenvalues of (2.1), where*

$$0 < \Lambda_{-\nu}^{(L)} \le \cdots \le \Lambda_{-2}^{(L)} \le \Lambda_{-1}^{(L)} < \rho < \Lambda_1^{(L)} \le \Lambda_2^{(L)} \le \dots$$

are the positive eigenvalues of the $(m+k)\times(m+k)$ matrix pencil

$$\begin{bmatrix} \mathbf{H} & \mathbf{C}^t \\ \mathbf{C} & \mathbf{N}_1 \end{bmatrix} - \Lambda^{(L)} \begin{bmatrix} \mathbf{I} & 0 \\ 0 & \mathbf{M}_1 \end{bmatrix}, \tag{4.8}$$

$$\begin{aligned}
\textit{where} \quad \mathbf{M}_1 &= \frac{1}{\rho}(\mathbf{N}_1 - \mathbf{N}_2) \\
\mathbf{N}_1 &= \mathbf{W}^{-1} + \mathbf{C}\mathbf{H}^{-1}\mathbf{C}^t, \quad \textit{and} \\
\mathbf{N}_2 &= \mathbf{C}(\mathbf{H}-\rho\mathbf{I})^{-1}\mathbf{C}^t.
\end{aligned}$$

ρ is an eigenvalue of (4.8) with multiplicity k. If the Ritz value $\Lambda_{r-1} < \rho$, then $\mathbf{M}_1$ is positive definite and (4.8) is a Hermitian definite pencil.

Proof. The first assertion follows immediately from (4.7), since then $\Lambda^{(L)}$ is an eigenvalue of (4.8) if and only if either $\hat{\mathbf{D}}(\Lambda^{(L)})$ is singular or $\hat{\mathbf{Y}}(\Lambda^{(L)})$ is singular. As before a limiting argument handles the exceptional cases where either ρ or $\Lambda^{(L)}$ are eigenvalues of $\mathbf{H}$.

For the second statement, note that $\Lambda_{r-1} < \rho$ implies from the way that r was chosen in (3.1) that $\mathbf{H}-\rho\mathbf{I}$ has precisely $r-1$ negative eigenvalues. Note then that $\mathbf{N}_1 - \mathbf{N}_2$ is positive-definite if and only if the matrix

$$\begin{bmatrix} \frac{1}{\rho}(\mathbf{H}-\rho\mathbf{I})\mathbf{H} & 0 \\ 0 & \mathbf{N}_1 - \mathbf{N}_2 \end{bmatrix} \tag{4.9}$$

has precisely $r-1$ negative eigenvalues. Define

$$\begin{aligned}
\tilde{\mathbf{L}}_1 &= \begin{bmatrix} \mathbf{I} & 0 \\ \rho\mathbf{C}\mathbf{H}^{-1}(\mathbf{H}-\rho\mathbf{I})^{-1} & \mathbf{I} \end{bmatrix}, \\
\tilde{\mathbf{L}}_2 &= \begin{bmatrix} \mathbf{I} & -\mathbf{C}^t\mathbf{W} \\ 0 & \mathbf{I} \end{bmatrix}, \quad \text{and } \tilde{\mathbf{D}} = \begin{bmatrix} \rho\mathbf{H}^{-1} & 0 \\ 0 & \mathbf{I} \end{bmatrix}.
\end{aligned}$$

and calculate with $\tilde{\mathbf{F}} = \tilde{\mathbf{D}}\tilde{\mathbf{L}}_2\tilde{\mathbf{L}}_1$

$$\tilde{\mathbf{F}} \begin{bmatrix} \frac{1}{\rho}(\mathbf{H}-\rho\mathbf{I})\mathbf{H} & 0 \\ 0 & \mathbf{N}_1 - \mathbf{N}_2 \end{bmatrix} \tilde{\mathbf{F}}^t = \begin{bmatrix} \rho(\mathbf{I}-\rho\mathbf{P}^t\mathbf{K}^{-1}\mathbf{P}) & 0 \\ 0 & \mathbf{W}^{-1} \end{bmatrix}. \tag{4.10}$$

Suppose (4.9) had more than $r-1$ negative eigenvalues. Then (4.10) has more than $r-1$ negative eigenvalues and therefore $\mathbf{I}-\rho\mathbf{P}^t\mathbf{K}^{-1}\mathbf{P}$ has more than $r-1$ negative eigenvalues. Equivalently, this means that $\mathbf{P}^t\mathbf{K}^{-1}\mathbf{P}$ has r or more eigenvalues *above* $1/\rho$. Since the eigenvalues of $\mathbf{P}^t\mathbf{K}^{-1}\mathbf{P}$ provide inner bounds to the outer eigenvalues of $\mathbf{K}^{-1}$, this implies in turn that $\mathbf{K}^{-1}$ must have r or more eigenvalues *above* $1/\rho$. But this contradicts the choice of ρ made in (3.1). □

The calculation of $\mathbf{W} = \mathbf{Q}_2^t\mathbf{K}^{-1}\mathbf{Q}_2$ involves the solution of k linear systems each of the form $\mathbf{Kx} = \mathbf{b}$. If these systems are solved inexactly (one rarely has other options), reasonable concerns arise about the integrity of the resulting bounds. Rigorous inclusion intervals can be maintained if the approximate calculation of $\mathbf{W}$ can be made to have the effect of replacing $\mathbf{W}$ with a matrix $\hat{\mathbf{W}} \geq \mathbf{W}$ (i.e., so that $\hat{\mathbf{W}} - \mathbf{W}$ is positive definite). To see this, observe that with the replacement of $\hat{\mathbf{W}}$ for $\mathbf{W}$ (4.5) becomes

$$(\mathbf{H}-\rho\mathbf{I})\hat{\mathbf{y}} = \hat{L}\ \left[(\mathbf{H}-\rho\mathbf{I}) - \rho(\mathbf{I}-\rho(\mathbf{H}^{-1}+\mathbf{H}^{-1}\mathbf{C}^t\hat{\mathbf{W}}\mathbf{C}\mathbf{H}^{-1}))\right]\hat{\mathbf{y}}. \tag{4.11}$$

The right-hand side of (4.5) has been replaced with a larger right-hand side in (4.11). The left hand side remains the same, so (4.11) and (4.5) will have the same numbers of positive (π) and negative (ν) eigenvalues. The min-max characterization then may be used to show that

$$\begin{array}{rl} L_k \leq \hat{L}_k < 0 & \text{for} \quad k = 1,\ 2,\ \ldots,\ \nu \\ 0 < \hat{L}_{-\ell} \leq L_{-\ell} & \text{for} \quad \ell = 1,\ 2,\ \ldots,\ \pi. \end{array}$$

The inequalities of (3.13) remain valid if $\hat{L}_k$ replaces L_k and $\hat{L}_{-\ell}$ replaces $L_{-\ell}$. Likewise if we define $\hat{\Lambda}^{(L)}_{\pm i} = -\rho\hat{L}_{\mp i}/(1-\hat{L}_{\mp i})$, the usual labeling is retained

$$\ldots\hat{\Lambda}^{(L)}_{-3} \leq \hat{\Lambda}^{(L)}_{-2} \leq \hat{\Lambda}^{(L)}_{-1} < \rho < \hat{\Lambda}^{(L)}_{1} \leq \hat{\Lambda}^{(L)}_{2} \leq \hat{\Lambda}^{(L)}_{3}\ \ldots,$$

and $\hat{\Lambda}^{(L)}_{-k} \leq \Lambda^{(L)}_{-k}$ for each $k = 1, \ldots, \nu$. The situation regarding the positively indexed $\hat{\Lambda}^{(L)}$ that yield bounds above ρ is slightly more complicated since it may occur that $\hat{L}_{-\ell} < 1 < L_{-\ell}$ which would then imply that $\hat{\Lambda}^{(L)}_{\ell} < 0$. In effect, $\hat{\Lambda}^{(L)}_{\ell}$ has "wrapped around" the point at infinity, yielding only trivial bounds for $\lambda_{r+\ell-1}$. Nontrivial bounds are retained whenever $\hat{\Lambda}^{(L)}_{\ell} > 0$, however.

Now, much the same development that yielded Theorem 4.1 may be followed with $\hat{\mathbf{W}}$ replacing $\mathbf{W}$. This is summarized as

THEOREM 4.2. *Suppose ρ is not an eigenvalue of (2.1). Each interval $[\hat{\Lambda}^{(L)}_{-i},\ \rho)$ and $(\rho,\ \hat{\Lambda}^{(L)}_{j}]$ contains respectively at least i and j eigenvalues of (2.1), where*

$$0 < \hat{\Lambda}^{(L)}_{-\nu} \leq \cdots \leq \hat{\Lambda}^{(L)}_{-2} \leq \hat{\Lambda}^{(L)}_{-1} < \rho < \hat{\Lambda}^{(L)}_{1} \leq \hat{\Lambda}^{(L)}_{2} \leq \ldots$$

are the positive eigenvalues of the $(m+k)\times(m+k)$ *matrix pencil*

$$\left[\begin{array}{cc}\mathbf{H} & \mathbf{C}^t\\ \mathbf{C} & \hat{\mathbf{N}}_1\end{array}\right]-\Lambda^{(L)}\left[\begin{array}{cc}\mathbf{I} & 0\\ 0 & \hat{\mathbf{M}}_1\end{array}\right], \tag{4.12}$$

$$\begin{aligned} \textit{where}\quad \hat{\mathbf{M}}_1 &= \frac{1}{\rho}(\hat{\mathbf{N}}_1-\mathbf{N}_2)\\ \hat{\mathbf{N}}_1 &= \hat{\mathbf{W}}^{-1}+\mathbf{C}\mathbf{H}^{-1}\mathbf{C}^t,\\ \mathbf{N}_2 &= \mathbf{C}(\mathbf{H}-\rho\mathbf{I})^{-1}\mathbf{C}^t,\end{aligned}$$

and $\hat{\mathbf{W}}$ *is any positive-definite matrix satisfying* $\hat{\mathbf{W}}\geq\mathbf{W}=\mathbf{Q}_2^t\mathbf{K}^{-1}\mathbf{Q}_2$. *Also,* ρ *is an eigenvalue of (4.12) with multiplicity* k.

Goerisch ([4],[6],[7]) discovered this critical approximation step for the original left-definite Lehmann formulation (3.16) and developed a very flexible framework for applying the approach in a PDE setting. He called it the $\{\mathcal{X},\, b, T\}$ method (referring to an auxiliary vector space $\mathcal{X}$, an auxiliary bilinear form b, and an auxiliary linear operator T that he introduces) but most others refer to this approach simply as the Lehmann-Goerisch method. To give a simple example, suppose a lower bound to $\mathbf{K}$ is known: $\kappa\|\mathbf{x}\|^2\leq\mathbf{x}^t\mathbf{K}\mathbf{x}$, and suppose we have obtained an approximate solution $\mathbf{Z}_2$ to the matrix equation $\mathbf{K}\mathbf{Z}=\mathbf{Q}_2$. Let $\mathbf{R}=\mathbf{Q}_2-\mathbf{K}\mathbf{Z}_2$ be the associated residual matrix. Then one may verify that

$$\begin{aligned}\mathbf{W}=\mathbf{Q}_2^t\mathbf{K}^{-1}\mathbf{Q}_2 &= \mathbf{R}^t\mathbf{K}^{-1}\mathbf{R}+\mathbf{Z}_2^t\mathbf{R}+\mathbf{Q}_2^t\mathbf{Z}_2\\ &\leq\frac{1}{\kappa}\mathbf{R}^t\mathbf{R}+\mathbf{Z}_2^t\mathbf{R}+\mathbf{Q}_2^t\mathbf{Z}_2\stackrel{def}{=}\hat{\mathbf{W}}.\end{aligned}$$

Note that $\hat{\mathbf{W}}$ contains the nominal estimate of $\mathbf{W}$, $\mathbf{Q}_2^t$"$\mathbf{K}^{-1}\mathbf{Q}_2$" $=\mathbf{Q}_2^t\mathbf{Z}_2$, together with correction terms that can be made small by solving $\mathbf{K}\mathbf{Z}=\mathbf{Q}_2$ more accurately while ensuring that in any case $\hat{\mathbf{W}}\geq\mathbf{W}$.

The foregoing development sought to reformulate the original Lehmann problems (and Goerisch's refinements) as bordered matrix eigenvalue problems since computational approaches to resolving such problems can take advantage of this structure more easily than that of the original Lehmann problems. Modern computing methodology now includes accurate and efficient approaches for calculating singular value decompositions of arbitrary matrices. This emerging capacity has shifted the focus of many problem formulations toward computational tasks involving the SVD or its generalizations and the calculation of Lehmann bounds are appropriately considered among them.[1]

To see first how this works for right-definite Lehmann bounds, suppose ρ is not an eigenvalue of (2.1) and consider a right-definite bound $\Lambda^{(R)}<\rho$ that satisfies (4.2). Define $\mu=(\rho+\Lambda^{(R)})/2$ and $\sigma=(\rho-\Lambda^{(R)})/2$, so that the interval $[\Lambda^{(R)},\,\rho)$ can be represented as $[\mu-\sigma,\mu+\sigma)$. Rewriting (4.2) in terms of μ and σ and simplifying yields

$$[(\mathbf{H}-\mu\mathbf{I})^2-\sigma^2\mathbf{I}+\mathbf{C}^t\mathbf{C}]\mathbf{y}=0.$$

That is, σ is a singular value of the $(m+k)\times m$ matrix

$$\left[\begin{array}{c}\mathbf{H}-\mu\mathbf{I}\\ \mathbf{C}\end{array}\right].$$

This leaves unspecified the association between the singular value indices and the Lehmann bound indices — an argument similar to the one outlined below for left-definite problems can be followed or one may consult Lehmann's original article [13].

[1]My thanks to Professor Beresford Parlett for impressing this point upon me.

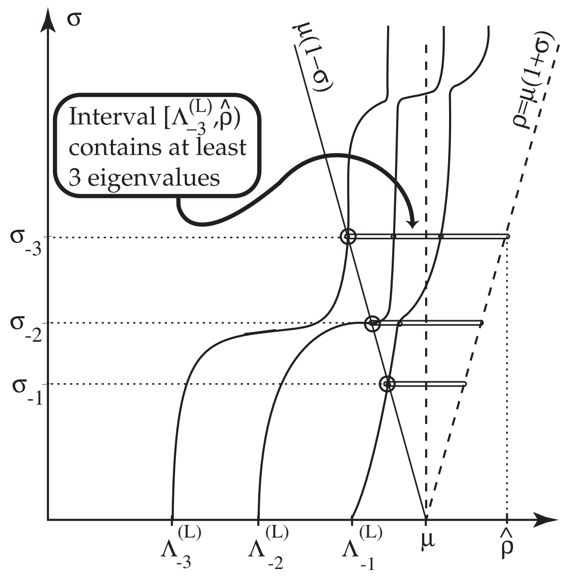

FIG. 4.1. *Connecting the singular value index and the Lehmann bound index.*

Recall that $\sigma > 0$ is a generalized singular value [18] of the pair $(\mathbf{S},\, \mathbf{T})$ if σ^2 is an eigenvalue of the Hermitian definite pencil $\mathbf{S}^t\mathbf{S} - \lambda\mathbf{T}^t\mathbf{T}$.

THEOREM 4.3. *Let $\mathbf{H}$ and $\hat{\mathbf{W}} \geq \mathbf{W}$ be decomposed as $\hat{\mathbf{W}} = \mathbf{L}_1\mathbf{L}_1^t$ and $\mathbf{H} = \mathbf{L}_2\mathbf{L}_2^t$. Suppose μ is a fixed real scalar such that the assignment of $\rho = \mu$ in Theorem 4.2 leads to $\nu > 0$, i.e., to nontrivial bounds in the interval $(0,\, \mu)$. If*

$$0 < \sigma_{-1} \leq \sigma_{-2} \leq \sigma_{-3} \leq \dots$$

denote the increasingly ordered generalized singular values of

$$\left(\begin{bmatrix} \frac{1}{\mu}(\mathbf{H} - \mu\mathbf{I}) \\ \mathbf{L}_1^t\mathbf{C}\mathbf{L}_2^{-t} \end{bmatrix},\, \begin{bmatrix} \mathbf{I} \\ \mathbf{L}_1^t\mathbf{C}\mathbf{L}_2^{-t} \end{bmatrix} \right), \tag{4.13}$$

then each of the intervals $[\mu(1-\sigma_{-i}),\ \mu(1+\sigma_{-i}))$ contains at least i eigenvalues of (2.1).

Proof. Suppose $\hat{\rho} > \mu$ is not an eigenvalue of (2.1) and consider an associated left-definite bound $\Lambda^{(L)}(\hat{\rho}) < \hat{\rho}$ that satisfies both (4.6) and $\mu = (\hat{\rho} + \Lambda^{(L)})/2$. Since each $\Lambda^{(L)}(\hat{\rho})$ is (essentially) monotone increasing in $\hat{\rho}$ (see below), this can always be done. Define $\sigma = (\hat{\rho} - \Lambda^{(L)})/2\mu$; so that the interval $[\Lambda^{(L)},\, \hat{\rho})$ can be represented as $[\mu(1-\sigma), \mu(1+\sigma))$. Rewriting (4.6) in terms of μ and σ, pre- and post-multiplying the coefficient matrix by $\mathbf{L}_2^{-1}$ and $\mathbf{L}_2^{-t}$, respectively, and then simplifying yields

$$[\frac{1}{\mu^2}(\mathbf{H} - \mu\mathbf{I})^2 + \mathbf{L}_2^{-1}\mathbf{C}^t\hat{\mathbf{W}}\mathbf{C}\mathbf{L}_2^{-t} - \sigma^2(\mathbf{I} + \mathbf{L}_2^{-1}\mathbf{C}^t\hat{\mathbf{W}}\mathbf{C}\mathbf{L}_2^{-t}]\mathbf{z} = 0$$

for $\mathbf{z} = \mathbf{L}_2^t\mathbf{y}$. That is, σ is a generalized singular value of (4.13), $[\mu(1-\sigma),\ \mu(1+\sigma))$ is a Lehmann interval, and so contains at least one eigenvalue of (2.1).

The proof that $[\mu(1-\sigma_{-i}),\ \mu(1+\sigma_{-i}))$ contains at least i eigenvalues of (2.1) is only outlined here and rests essentially on showing that the picture in Figure 4.1 represents the general situation. Consider the Lehmann-Goerisch bounds $\Lambda_{-1}^{(L)}, \Lambda_{-2}^{(L)}, \ldots$ as functions of $\rho = \mu(1+\sigma)$ and observe from (4.6) that $\Lambda^{(L)} \neq \rho$ for all values of ρ. Hence the $\Lambda^{(L)}$-curves in Figure 4.1 never cross the line $\rho = \mu(1+\sigma)$. Although the $\Lambda^{(L)}$-curves might not be monotone increasing in ρ in a strict sense, it is straightforward to show that they may never decrease more rapidly than with a slope of -1, so in particular each $\Lambda^{(L)}$-curve intersects the line $\mu(1-\sigma)$ exactly once. Thus if $\hat{\rho} = \mu(1+\sigma_{-3})$, for example, then $[\Lambda_{-3}^{(L)}(\hat{\rho}),\, \hat{\rho}) \equiv [\mu(1-\sigma_{-3}),\, \mu(1+\sigma_{-3}))$ contains at least 3 eigenvalues of (2.1). □

Although this hasn't been done here, it is plausible that the technical assumption on μ could be disposed of if Lehmann bounds *above* ρ were to be considered as well in the proof (requiring monotonicity with respect to ρ of Lehmann bounds above ρ as well).

5. A Left-Right Comparison. For the general eigenvalue problem (2.1), application of either right- or left-definite Lehmann bounds involve solving linear systems having either $\mathbf{M}$ (for right-definite problems) or $\mathbf{K}$ (for left-definite problems) as a coefficient matrix. If one system is very much simpler than the other (e.g., if $\mathbf{M}$ is diagonal) one may feel compelled to choose the simpler path. But is there a difference in accuracy ? Goerisch and coworkers in Braunschweig and Clausthal (see for example, [5] and [6]) have observed that for many applications in PDE settings, left-definite Lehmann bounds often were superior to right-definite bounds — even if an extra level of approximation is included as described in Theorem 4.2. Along similar lines, Knyazev [10] has produced error estimates for Lehmann methods that suggest left-definite bounds might be better than right-definite bounds asymptotically.

We explore this issue here. Define

$$\mathbf{J}_0 \stackrel{def}{=} \mathbf{H}_0 - \rho\mathbf{H}_1,\ \mathbf{J}_1 \stackrel{def}{=} \mathbf{H}_1 - \rho\mathbf{H}_2,\ \text{and}\ \mathbf{J}_2 \stackrel{def}{=} \mathbf{H}_2 - \rho\mathbf{H}_3.$$

The matrix pencils associated with (3.6) and (3.16) may be written as

$$\mathbf{J}_1 - R(\mathbf{J}_0 - \rho\mathbf{J}_1) \tag{5.1}$$

and

$$\mathbf{J}_1 - L(\mathbf{J}_1 - \rho\mathbf{J}_2) \tag{5.2}$$

for right-definite and left-definite problems, respectively.

The following lemma and theorem incorporate some unpublished results of Goerisch[2].

LEMMA 5.1. *Let* $\mathbf{G} = \begin{bmatrix} \mathbf{J}_0 & \mathbf{J}_1 \\ \mathbf{J}_1 & \mathbf{J}_2 \end{bmatrix} \in \mathbb{R}^{2m\times 2m}$. $\mathbf{G}$ *has no more than* $r-1$ *negative eigenvalues.*

Proof. Suppose that $\mathbf{G}$ has r or more negative eigenvalues. Then there is an r-dimensional subspace $\mathcal{Z}$ of $\mathbb{R}^{2m}$ such that $\mathbf{z}^t\mathbf{G}\mathbf{z} < 0$ for all $\mathbf{z} \in \mathcal{Z}$ with $\mathbf{z} \neq 0$. Define the linear mapping $T : \mathcal{Z} \to \mathbb{R}^n$ by

$$T(\mathbf{z}) = \sum_{i=1}^{m} z_i\mathbf{K}\mathbf{p}_i + \sum_{i=1}^{m} z_{i+m}\mathbf{M}\mathbf{p}_i.$$

Elementary manipulations verify that for $\mathbf{z} \in \mathcal{Z}$ with $\mathbf{z} \neq 0$,

$$\mathbf{z}^t\mathbf{G}\mathbf{z} = T(\mathbf{z})^t\mathbf{M}^{-1}T(\mathbf{z}) - \rho T(\mathbf{z})^t\mathbf{K}^{-1}T(\mathbf{z}) < 0. \tag{5.3}$$

[2] Friedrich Goerisch died suddenly in 1995 after a brief illness. The loss of his passion and insight is still deeply felt among his colleagues and friends.

In particular, this means that $T(\mathbf{z}) = 0$ implies that $\mathbf{z} = 0$, so $null(T) = 0$ and $rank(T) = \dim \mathcal{Z} = r$.

Since $\mathbf{K}$ is positive-definite $\mathbf{u}^t\mathbf{K}^{-1}\mathbf{u} > 0$ for all $\mathbf{u} \in \mathbb{R}^n$ so (5.3) implies

$$\mathbf{u}^t\mathbf{M}^{-1}\mathbf{u}/\mathbf{u}^t\mathbf{K}^{-1}\mathbf{u} < \rho$$

for all $\mathbf{u} \in Ran(T)$ with $\mathbf{u} \neq 0$.

Now λ is an eigenvalue of (2.1) if and only if it is also an eigenvalue of $\mathbf{M}^{-1}\mathbf{v} = \lambda\mathbf{K}^{-1}\mathbf{v}$, so by the min-max principle

$$\lambda_r = \min_{\dim \mathcal{P}=r} \max_{\mathbf{u}\in\mathcal{P}} \frac{\mathbf{u}^t\mathbf{M}^{-1}\mathbf{u}}{\mathbf{u}^t\mathbf{K}^{-1}\mathbf{u}} \leq \max_{\mathbf{u}\in Ran(T)} \frac{\mathbf{u}^t\mathbf{M}^{-1}\mathbf{u}}{\mathbf{u}^t\mathbf{K}^{-1}\mathbf{u}} < \rho$$

which contradicts $\lambda_{r-1} < \rho < \lambda_r$. Thus, $\dim \mathcal{Z} < r$. □

THEOREM 5.2. *If the harmonic Ritz value $\tilde{\Lambda}_{r-1}$ from (2.4) satisfies $\tilde{\Lambda}_{r-1} < \rho$ then left-definite Lehmann bounds will be uniformly better than right-definite Lehmann bounds:*

$$\Lambda_{-k}^{(R)} \leq \Lambda_{-k}^{(L)} \leq \lambda_{r-k} \quad \textit{for } k = 1,\ \ldots,\ r-1, \tag{5.4}$$

$$\lambda_{r+\ell-1} \leq \Lambda_{\ell}^{(L)} \leq \Lambda_{\ell}^{(R)} \quad \textit{for } \ell = 1,\ \ldots,\ m-r+1. \tag{5.5}$$

Proof. To show that (5.4) and (5.5) are true,i it is sufficient to show that $L_k \leq 1 + \rho R_k$ for $k = 1,\ 2,\ \ldots, r-1$ and that $1 + \rho R_{-\ell} \leq L_{-\ell}$ for $\ell = 1,\ 2,\ \ldots, m-r+1$. From (5.1), one finds that $1 + \rho R_k$ and $1 + \rho R_{-\ell}$ are eigenvalues of

$$\mathbf{J}_0 - (1 + \rho R)(\mathbf{J}_0 - \rho\mathbf{J}_1). \tag{5.6}$$

Since $\Lambda_{r-1} \leq \tilde{\Lambda}_{r-1} < \rho$, both $\mathbf{J}_0$ and $\mathbf{J}_1$ have $r-1$ negative eigenvalues. This implies that both (5.1) and (5.2) have $r-1$ negative eigenvalues. Premultiplication of (5.6) by $\mathbf{J}_1\mathbf{J}_0^{-1}$ yields an equivalent matrix pencil:

$$\mathbf{J}_1 - (1 + \rho R)(\mathbf{J}_1 - \rho\mathbf{J}_1\mathbf{J}_0^{-1}\mathbf{J}_1).$$

Consider

$$\mathbf{G} = \begin{bmatrix} \mathbf{J}_0 & \mathbf{J}_1 \\ \mathbf{J}_1 & \mathbf{J}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{I} & 0 \\ \mathbf{J}_1\mathbf{J}_0^{-1} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{J}_0 & 0 \\ 0 & \mathbf{J}_2 - \mathbf{J}_1\mathbf{J}_0^{-1}\mathbf{J}_1 \end{bmatrix} \begin{bmatrix} \mathbf{I} & \mathbf{J}_0^{-1}\mathbf{J}_1 \\ 0 & \mathbf{I} \end{bmatrix}.$$

By the lemma and the Sylvester law of inertia, $\mathbf{J}_0 \oplus \mathbf{J}_2 - \mathbf{J}_1\mathbf{J}_0^{-1}\mathbf{J}_1$ can have no more than $r-1$ negative eigenvalues. Since $\mathbf{J}_0$ has exactly $r-1$ eigenvalues by hypothesis, $\mathbf{J}_2 - \mathbf{J}_1\mathbf{J}_0^{-1}\mathbf{J}_1$ must be positive semi-definite and

$$0 < \mathbf{x}^t(\mathbf{J}_1 - \rho\mathbf{J}_2)\mathbf{x} \leq \mathbf{x}^t(\mathbf{J}_1 - \rho\mathbf{J}_1\mathbf{J}_0^{-1}\mathbf{J}_1)\mathbf{x}$$

for all nontrivial $\mathbf{x}$. Hence, for $k = 1,\ 2,\ \ldots,\ r-1$,

$$\begin{aligned} 1 + \rho R_k &= \min_{\dim \mathcal{S}=k} \max_{\mathbf{x}\in\mathcal{S}} \frac{\mathbf{x}^t\mathbf{J}_1\mathbf{x}}{\mathbf{x}^t(\mathbf{J}_1 - \rho\mathbf{J}_1\mathbf{J}_0^{-1}\mathbf{J}_1)\mathbf{x}} \\ &\geq \min_{\dim \mathcal{S}=k} \max_{\mathbf{x}\in\mathcal{S}} \frac{\mathbf{x}^t\mathbf{J}_1\mathbf{x}}{\mathbf{x}^t(\mathbf{J}_1 - \rho\mathbf{J}_2)\mathbf{x}} = L_k, \end{aligned}$$

and for $\ell = 1,\ 2,\ \ldots,\ m - r + 1$,

$$\begin{aligned}-(1+\rho R_{-\ell}) &= \min_{\dim \mathcal{S}=\ell} \max_{\mathbf{x}\in\mathcal{S}} \frac{-\mathbf{x}^t \mathbf{J}_1 \mathbf{x}}{\mathbf{x}^t(\mathbf{J}_1 - \rho \mathbf{J}_1 \mathbf{J}_0^{-1} \mathbf{J}_1)\mathbf{x}} \\ &\geq \min_{\dim \mathcal{S}=\ell} \max_{\mathbf{x}\in\mathcal{S}} \frac{-\mathbf{x}^t \mathbf{J}_1 \mathbf{x}}{\mathbf{x}^t(\mathbf{J}_1 - \rho \mathbf{J}_2)\mathbf{x}} = -L_{-\ell}.\end{aligned}$$

Since there will be subspaces of dimension up to $r-1$ for which $\mathbf{x}^t\mathbf{J}_1\mathbf{x} < 0$ and subspaces of dimension up to $m-r+1$ for which $\mathbf{x}^t\mathbf{J}_1\mathbf{x} > 0$, we may restrict ourselves to $\mathbf{x}$ for which the numerators in the above expressions are strictly negative with no loss of generality. □

6. A Ritz-Lehmann Comparison. One may hope that the role spectral mapping played in the derivation of both left- and right-definite variants of Lehmann's method might lead to significant improvements beyond the straightforward application of the Rayleigh-Ritz method. Indeed, spectral mapping has been used for some time with Lanczos methods (e.g., [3]) with sometimes spectacular effect and so encouraged, some have considered the use of right-definite Lehmann bounds using Krylov subspaces generated in the course of an ordinary Lanczos process (e.g., [15] and [17]). By and large, results along these lines have been disappointing when compared with what "shift-and-invert" methods offer (albeit at a much higher price). One may instead seek to compare the expected outcomes of Lehmann methods with those of Rayleigh-Ritz methods. Observe that each method makes optimal use of the information required in the sense that no better bounds are possible with the infomation used, so in a certain manner of speaking we are really comparing the utility of various types of information in extracting eigenvalue information.

Zimmerman [19] proved that the error in left-definite Lehmann bounds is no worse than proportional to the error in Ritz bounds and may be smaller. Thus, left-definite Lehmann bounds carry the potential of greater accuracy than Ritz bounds. We probably shouldn't expect them to be much better, though. In [10], Knyazev states that eigenvector approximations provided by either the right- or left-definite variants of Lehmann's method will asymptotically approach the corresponding Ritz vectors as they close upon the true eigenvectors. Thus, Lehmann methods appear to recover invariant subspace information with about the same efficiency as Rayleigh-Ritz methods.

It is important to note that Lehmann methods provide eigenvalue *bounds* that often are difficult to obtain in other ways. For example, Behnke [1] combined right-definite Lehmann methods with interval techniques in order to deduce guaranteed bounds to matrix eigenvalue problems and his approach appears to be competitive with the best known interval algorithms for this problem.

For the remainder of this section, we will consider the application of a left-definite Lehmann method within a Lanczos process for resolving a large-scale matrix eigenvalue problem. Since left-definite Lehmann methods are known to be superior to right-definite Lehmann methods (at least to the extent claimed in Section 5), one may seek to improve upon the results of Morgan [15] by using left-definite Lehmann-Goerisch bounds as formulated in Theorem 4.2.

Specifically, let $\mathbf{M} = \mathbf{I}$ in (2.1) and let $\mathbf{T}$ be a tridiagonal matrix that is similar to $\mathbf{K}$ – so that $\mathbf{K} = \mathbf{Q}\mathbf{T}\mathbf{Q}^t$ for some $n \times n$ unitary matrix $\mathbf{Q}$. For any index $1 \leq \ell \leq n$, let $\mathbf{T}_\ell$ denote

ETNA
Kent State University
etna@mcs.kent.edu

the ℓth principal submatrix of $\mathbf{T}$:

$$\mathbf{T}_\ell = \begin{bmatrix} \alpha_1 & \beta_1 & & & & \\ \beta_1 & \alpha_2 & \beta_2 & & & \\ & \beta_2 & \alpha_3 & \ddots & & \\ & & \ddots & \ddots & & \\ & & & & & \beta_{\ell-1} \\ & & & & \beta_{\ell-1} & \alpha_\ell \end{bmatrix},$$

and define $\mathbf{V}$ via a partitioning of $\mathbf{T}$ as

$$\mathbf{T} = \begin{bmatrix} \mathbf{T}_\ell & \beta_\ell \mathbf{e}_\ell \mathbf{e}_1^t \\ \beta_\ell \mathbf{e}_1 \mathbf{e}_\ell^t & \mathbf{V} \end{bmatrix}.$$

Let $\mathbf{Q}_\ell$ denote a matrix containing the first ℓ columns of $\mathbf{Q}$: $\mathbf{Q}_\ell = [\mathbf{q}_1, \ \dots \ \mathbf{q}_\ell]$.

The Lanczos algorithm builds up the matrices $\mathbf{T}$ and $\mathbf{Q}$ one column at a time starting with the vector $\mathbf{q}_1$. Only information on the action of $\mathbf{K}$ on selected vectors in $\mathbb{R}^n$ is used. Different choices for $\mathbf{q}_1$ produce distinct outcomes for $\mathbf{T}$, if all goes well. Extracting useful information when not all goes well is fundamental to modern approaches – a discussion may be found in [16].

At the ℓth step, the basic Lanczos recursion appears as

$$\mathbf{K}\mathbf{Q}_\ell = \mathbf{Q}_\ell \mathbf{T}_\ell + \beta_\ell \mathbf{q}_{\ell+1} \mathbf{e}_\ell^t.$$

In exact arithmetic, the first ℓ steps yields a matrix $\mathbf{Q}_\ell$ that satisfies $\mathbf{Q}_\ell^t \mathbf{Q}_\ell = \mathbf{I}$ and

$$Ran(\mathbf{Q}_\ell) = \text{span}\{\mathbf{q}_1, \mathbf{A}\mathbf{q}_1, \dots, \mathbf{A}^{\ell-1}\mathbf{q}_1\} = \mathcal{K}_\ell(\mathbf{A}, \mathbf{q}_1),$$

a Krylov subspace of order ℓ. The application of Theorem 4.2 is straightforward:

THEOREM 6.1. *Let $\mathbf{M} = \mathbf{I}$ and suppose ρ is not an eigenvalue of (2.1). Each interval $[\Lambda_{-i}^{(L)}, \ \rho)$ and $(\rho, \ \Lambda_j^{(L)}]$ contains respectively at least i and j eigenvalues of the matrix $\mathbf{K}$, where*

$$0 < \Lambda_{-\nu}^{(L)} \cdots \le \Lambda_{-2}^{(L)} \le \Lambda_{-1}^{(L)} < \rho < \Lambda_1^{(L)} \le \Lambda_2^{(L)} \le \dots$$

are the positive eigenvalues of the tridiagonal matrix pencil

$$\begin{bmatrix} \mathbf{T}_\ell & \beta_k \mathbf{e}_k \\ \beta_k \mathbf{e}_k^t & \omega_{k+1}^{-1} + \beta_k^2 \mathbf{e}_k^t \mathbf{T}_k^{-1} \mathbf{e}_k \end{bmatrix} - \Lambda^{(L)} \begin{bmatrix} \mathbf{I} & 0 \\ 0 & (\rho\omega_{k+1})^{-1} - \beta_k^2 \delta_{k+1}(\rho) \end{bmatrix}, \tag{6.1}$$

where ω_{k+1} is any number that satisfies

$$\begin{aligned} \omega_{k+1} &\ge \mathbf{q}_{k+1}^t \mathbf{K}^{-1} \mathbf{q}_{k+1} \\ \text{and} \quad \delta_{k+1}(\rho) &= \mathbf{e}_k^t \mathbf{T}_k^{-1} (\mathbf{T}_k - \rho)^{-1} \mathbf{e}_k. \end{aligned}$$

Note that ρ is a simple eigenvalue of (6.1)

We apply this directly to the numerical example considered in [17] and in Section 1. Figure 2 shows the convergence history both for Ritz bounds and for left-definite Lehmann bounds, for the seventh through tenth eigenvalues of the matrix. We also apply a shift and invert Lanczos method using the spectral transformation $\lambda \mapsto \frac{\lambda}{\lambda - \rho}$. A few features are apparent. The first is that the Lehmann bounds aren't nearly as good as the shift and invert

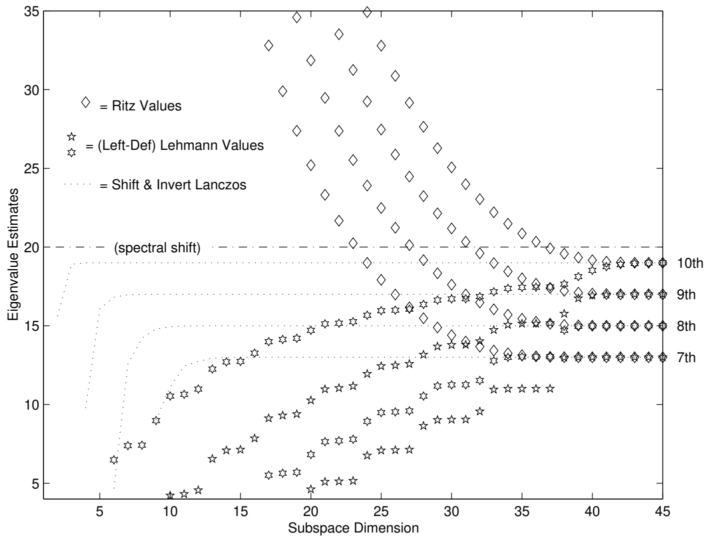

FIG. 6.1. *Convergence of Ritz and Lehmann bounds using Krylov subspaces vs. Shift & invert Lanczos with same starting vector.*

bounds to which they are closely related. Paige, Parlett, and van der Vorst [17] observed this disappointing behaviour for right-definite Lehmann methods (in their context, harmonic Ritz on a shifted matrix) — the left-definite Lehmann method does not fare much better. Knyazev's observations [10] relating convergence of Lehmann eigenvectors to Ritz vectors suggest that spectral information for interior matrix eigenvalues will not be picked up any more rapidly with Lehmann methods than for Ritz methods. This is in stark contrast with shift and invert strategies which will produce approximate eigenvectors that are rapidly drawn into invariant subspaces associated with eigenvalues close to ρ.

The second observation is that, nonetheless, the Lehmann bounds do appear to approach the exact eigenvalues at a rate comparable to that of the Ritz bounds — consistent with the results of Zimmerman discussed above. Furthermore, one can see that the Lehmann bounds appear to pass through a series of stagnation points en route to their limit, and the farther they lie from ρ, the more abrupt the transition between stagnation points. These stagnation points appear to be close to the exact matrix eigenvalues.

The following simple Bauer-Fike style perturbation result lends some insight to this behaviour.

THEOREM 6.2. *Let $\Lambda^{(L)}$ be any left-definite Lehmann bound and denote with Λ_i the Ritz values from (2.2). Then*

$$\min_i \left(\frac{|\Lambda_i - \rho|}{\rho}\right)\left(\frac{|\Lambda_i - \Lambda^{(L)}|}{\Lambda^{(L)}}\right)\Lambda_i \quad \leq \quad \|\mathbf{W}\|\,\|\mathbf{C}\|^2. \tag{6.2}$$

Proof. If either $(\mathbf{H} - \rho\mathbf{I})$ or $(\mathbf{H} - \Lambda^{(L)}\mathbf{I})$ is singular then (6.2) holds trivially. Suppose then that $(\mathbf{H} - \rho\mathbf{I})$ and $(\mathbf{H} - \Lambda^{(L)}\mathbf{I})$ are nonsingular. Rearrange the expression (4.6) to get

$$\mathbf{y} = -\rho\Lambda^{(L)}(\mathbf{H} - \rho\mathbf{I})^{-1}(\mathbf{H} - \Lambda^{(L)}\mathbf{I})^{-1}\mathbf{H}^{-1}\mathbf{C}^t\mathbf{W}\mathbf{C}\mathbf{y}.$$

Take norms on each side and simplify:

$$1 \leq \rho\Lambda^{(L)}\|(\mathbf{H} - \rho\mathbf{I})^{-1}(\mathbf{H} - \Lambda^{(L)}\mathbf{I})^{-1}\mathbf{H}^{-1}\|\,\|\mathbf{W}\|\,\|\mathbf{C}\|^2. \tag{6.3}$$

Then notice that

$$\begin{aligned}\|(\mathbf{H}-\rho\mathbf{I})^{-1}(\mathbf{H}-\Lambda^{(L)}\mathbf{I})^{-1}\mathbf{H}^{-1}\| &= \max_i \left(\frac{1}{|\Lambda_i-\rho|}\right)\left(\frac{1}{|\Lambda_i-\Lambda^{(L)}|}\right)\frac{1}{\Lambda_i} \\ &= 1/\min_i(|\Lambda_i-\rho|\,|\Lambda_i-\Lambda^{(L)}|\,\Lambda_i),\end{aligned}$$

which may be combined with (6.3) to get (6.2). □

Notice that the right hand side of (6.2) has a magnitude related to the size of the Ritz residual $\mathbf{KQ}_1-\mathbf{Q}_1\mathbf{H}$ and is independent of which Lehmann bound $\Lambda^{(L)}$ is chosen. Suppose the right hand side of (6.2) is moderately small and choose a Lehmann bound $\Lambda^{(L)}$. If $\Lambda^{(L)}$ is not close to ρ then any Ritz value Λ_i that is close to $\Lambda^{(L)}$ will not be close to ρ either. Thus any $\Lambda^{(L)}$ chosen far from ρ is constrained by (6.2) to be nearer to at least one Λ_i then it would be were $\Lambda^{(L)}$ chosen closer to ρ. A qualitative interpretation that one might take from this is that Lehmann bounds $\Lambda^{(L)}$ far from ρ tend to occur in the neighborhood of Ritz values Λ_i. Furthermore, Lehmann bounds $\Lambda^{(L)}$ far from ρ that are also situated toward the edges of the spectrum will tend to aggregate in the neighborhood of exact eigenvalues since the attracting Ritz values themselves will be approximating extreme eigenvalues fairly well.

REFERENCES

[1] H. BEHNKE, *Inclusion of Eigenvalues of General Eigenvalue Problems for Matrices*, Computing, 6 (Suppl.) (1988), pp. 69–78.

[2] COLLATZ, L. *Eigenwertprobleme und ihre numerische Behandlung*, Chelsea, NY, 1948.

[3] T. ERICSSON AND A. RUHE, *The spectral transformation Lanczos method for the numerical solution of large sparse generalized symmetric eigenvalue problems*, Math. Comp., 35 (1980), pp. 1251–1268.

[4] F. GOERISCH, *Eigenwertschranken und komplementäre Extremalprinzipien*, Habilitationsschrift, Technische Universität – Clausthal, 1986.

[5] F. GOERISCH AND J. ALBRECHT, *Eine einheitliche Herleitung von Einschliessungssätzen für Eigenwerte*, in Numerical Treatment of Eigenvalue Problems, J. Albrecht, L. Collatz, and W. Velte, eds., International Series of Numerical Mathematics, Vol. 69, Birkhäuser, Basel, 1983.

[6] F. GOERISCH AND H. HAUNHORST, *Eigenwertschranken für Eigenwertaufgaben mit partiellen Differentialgleichungen*, Z. Angew. Math. Mech., 65 (1985), pp. 129–135.

[7] F. GOERISCH AND S. ZIMMERMAN, *On Trefftz's method and its application to eigenvalue problems*, Z. Angew. Math. Mech., 66 (1986), pp. T304–T306.

[8] R. GRAMMEL, *Ein neues Verfahren zur Lösung technischer Eigenwertprobleme*, Ing.-Arch. 10 (1939), pp. 35–46.

[9] R. G. GRIMES, J. G. LEWIS AND H. D. SIMON, (1994), *A shifted block Lanczos algorithm for solving sparse symmetric generalized eigenproblems*, SIAM J. Matrix Anal. Appl., 15(1) (1994), pp. 228–272.

[10] A. V. KNYAZEV, *Convergence rate estimates for iterative methods for mesh symmetric eigenvalue problem*, Soviet J. Numer. Anal. Math. Modelling, 2(5) (1987), pp. 371–396.

[11] N. J. LEHMANN, *Berechnung von Eigenwertschranken bei linearen Problemen*, Arch. Math. (Basel), 2 (1949/50), pp. 139–147.

[12] N. J. LEHMANN, *Beiträge zur numerischen Lösung linearer Eigenwertprobleme*, Z. Angew. Math. Mech. I, 29 (1949/50), pp. 341–356; II, 30 (1949/50), pp. 1–16.

[13] N. J. LEHMANN, *Optimale Eigenwerteinschließungen*, Numer. Math., 5 (1963), pp. 246–272.

[14] H. J. MAEHLY, *Ein neues Variationsverfahren zur genäherten Berechnung der Eigenwerte hermitescher Operatoren*, Helv. Phys. Acta, 25 (1952), pp. 547–568.

[15] R. MORGAN,*Computing interior eigenvalues of large matrices*, Linear Algebra Appl., 154-156 (1991), pp. 289–309.

[16] B. PARLETT, *The Symmetric Eigenvalue Problem*, Prentice-Hall, Englewood Cliffs, 1980.

[17] C. C. PAIGE, B. N. PARLETT AND H. A. VAN DER VORST, *Approximate solutions and eigenvalue bounds from Krylov subspaces*, Numer. Linear Algebra Appl., 2(2) (1995).

[18] C. F. VAN LOAN, *Generalizing the singular value decomposition*, SIAM J. Numer. Anal., 13 (1976), pp. 76–83.

[19] S. ZIMMERMANN, (1994), *Comparison of Errors in Upper and Lower Bounds to Eigenvalues of Self-Adjoint Eigenvalue Problems*, Numer. Funct. Anal. Optim., 15 (7-8) (1994), pp. 943–960.