GLASNIK MATEMATIČKI
Vol. 35(55)(2000), 25 – 44

THE MINIMAL INDEX OF A SELF-ADJOINT PENCIL

Paul Binding and Branko Najman
University of Calgary, Canada, University of Zagreb, Croatia

Dedicated to my coauthor and great friend

Abstract. Let A and B be selfadjoint operators on a Hilbert space H. We define the minimal index $\nu(A, B) = \min\{\#\text{negative eigenvalues of } A - \lambda B\}$, we connect it with various ideas in the literature and we connect it with formulae used in some recent variational principles.

1. Introduction

Let A and B be self adjoint operators on a Hilbert space H. For simplicity we assume at present that A is bounded below with compact resolvent and B is bounded, although more general situations will be treated (see Sections 3 and 4 for comments on this, and on application to matrix and boundary value problems). It follows that the spectrum $\sigma(A - \lambda B)$ is discrete, so $-\mu \in \sigma(A - \lambda B)$ if and only if

$$(A - \lambda B + \mu I)x = 0 \tag{1.1}$$

for some nonzero $x \in D(A) \cap D(B)$.

We say that λ is an **eigenvalue** of (A, B) if (1.1) holds with $\mu = 0$, i.e.,

$$Ax = \lambda Bx \tag{1.2}$$

for some $0 \neq x \in D(A) \cap D(B)$.

Note that

$$\sigma(A, B) = \{\lambda : 0 \in \sigma(A - \lambda B)\}$$

2000 *Mathematics Subject Classification.* 15A22, 47A13, 47A56, 49R05.

Research supported by NSERC of Canada and the I. W. Killam Foundation.

Professor Najman died suddenly while this work was nearing its final stages. His research was supported by the Ministry of Science of Croatia.

can be the whole complex plane (e.g., if $N(A) \cap N(B) \neq 0$) even though $\sigma(A - \lambda B)$ is discrete. For this introduction, then, we assume that $\sigma(A, B)$ is nonempty and discrete, and hence consists of eigenvalues of (A, B) accumulating at most at $\pm\infty$. Explicit assumptions on A and B will be given in Sections 2 and 3, including the possibility of essential spectrum.

We define the (negative) **index** $\nu(A)$ of A to be the number of negative eigenvalues of A, counted by multiplicity. For $\lambda \in \mathbf{R}$ we write $\nu(\lambda) = \nu(A - \lambda B)$: this is evidently the number of positive values of μ (counted by multiplicity) satisfying (1.1). For any eigenvalue λ of (A, B), we call $\nu(\lambda)$ the **index** of λ. In the case of a Sturm-Liouville (SL) equation (1.2) of the form $lx := -(px')' + qx = \lambda rx$ with separated end conditions on $[a, b]$, this is the "oscillation count" of the eigenfunction x, i.e., the number of zeros of x in $]a, b[$. In this case, $H = L_2(a, b)$, A is a differential operator generated by l and the boundary conditions, and B is the operator of multiplication by r.

We are mainly interested in the **minimal index** of (A, B), which we define as

$$\nu(A, B) = \min\{\nu(\lambda) : \lambda \in \mathbf{R}\}. \tag{1.3}$$

This quantity has appeared implicitly and explicitly in various contexts but does not seem to have been studied in its own right. In 1.1 and 1.2 we shall connect $\nu(A, B)$ with (a) the minimal "oscillation count" in the case when (1.2) is a SL equation, and (b) bounds on the numbers and multiplicities of eigenvalues of (A, B) which are, say, nonreal or have specified index. In Sections 2 and 3 we equate $\nu(A, B)$ with an explicit formula involving certain spectral information (independent of minimisation) of the pencil, and we discuss what happens to it under perturbation. Section 2 contains the finite dimensional case, while Section 3 covers certain infinite dimensional cases with both discrete and continuous spectra. Section 4 contains applications to the "index shift" in certain variational principles for the eigenvalues of (A, B), and to finiteness of certain variational quantities associated with the pair (A, B), thus leading to numerical estimation of $\nu(A, B)$.

1.1. *Minimal oscillation count.* To our knowledge, $\nu(A, B)$ first appeared in the published literature implicitly, in Richardson's analysis [28] of two parameter SL equations. Such equations led Richardson to study problems of the form (1.2) with neither A nor B definite, and for which the "oscillation theorem" (that there was an eigenfunction x for any given oscillation count) could fail. Richardson defined a minimal oscillation count (over all possible eigenfunctions) for (1.2), but it was the subsequent work of Haupt [17] on this subject which gave an explicit definition of $\nu(A, B)$, via the embedding (1.1).

We remark that the minimal oscillation count is actually the minimal <u>eigenvalue</u> index, i.e., $\min\{\nu(\lambda) : \lambda \in \sigma(A, B)\}$, and this appears to depend on solving (1.2). Fortunately, equality with $\nu(A, B)$ follows easily, e.g., from continuity of the (variational) **eigencurves** i.e., the set of (λ, μ) satisfying (1.1) and such that $\nu(A - \lambda B + \mu I)$ is fixed, cf. [5]. We remark that much of Richardson's reasoning is also based explicitly on eigencurve arguments.

Minimal oscillation counts have been discussed in more modern settings in [3, 6] in connection with the embedding (1.1).

1.2. *Spectral bounds.* In the "left definite" case, i.e., when $A > 0$, the spectral theory of (1.2) is equivalent to that of the compact symmetric operator $A^{-1}B$ on the Hilbert space H_A defined as the completion of $D(A)$ under the inner product given by $(x, y)_A = (x, Ay)$. It can be shown (cf. Section 3) that H_A coincides with the form domain $D(a) = D(A^{1/2})$. If A is not definite but is (boundedly) invertible then H_A becomes a Pontryagin space of (negative) index $\nu(A)$. This provides bounds for various quantities such as the number of nonreal (conjugate) eigenvalue pairs, the total length of all Jordan chains, etc., cf. [12]. In the SL context, such ideas are developed in [15, 24]. In Section 3 we shall develop the Pontryagin space setting for a fairly general class of pairs (A, B).

Since the complexity of such a theory depends on the index of H_A, it is of importance to reduce that index as far as possible. This can be achieved by translating the eigenparameter λ, by λ_0 say, thus replacing A by $A - \lambda_0 B$. The "optimal" λ_0 leads to a Pontryagin space of index $\nu(A, B)$, at least if $A - \lambda_0 B$ is invertible. Bounds (as above) were already given for SL problems in [17], and more recently in [24]; cf. [20] for a pde context. A variety of such bounds, involving the number n^i of eigenvalues of (A, B) of index i, and related quantities, can be found in [5]. In this work, $\nu(A, B)$ is interpreted as the number of eigencurves lying above the λ-axis.

2. The finite dimensional case

Throughout we assume that the $N \times N$ matrices A and B form a "nonsingular" pair, i.e., that some linear combination of A and B is nonsingular. Then the canonical form (see [26]) shows that, for some nonsingular matrix T,

$$T^*AT = diag(A_\infty, A_F) \ , \ T^*BT = diag(B_\infty, B_F) \tag{2.1}$$

where the partitioned blocks are of sizes N_∞ and N_F respectively. The N_∞ zero eigenvalues of the pair (B_∞, A_∞) correspond to the so-called infinite eigenvalues of (A, B), i.e., the zero eigenvalues of (B, A). The span of the corresponding (algebraic) eigenspaces is denoted by X_∞, and the dimension of a maximal subspace of X_∞ on which A_∞ is negative definite is denoted by N_∞^-. The span of the eigenspaces corresponding to (A, B) is denoted by X_F. Since T merely amounts to a change of coordinates in the quadratic forms of

A and B, we shall suppress it in what follows. This simplifies the proofs, but does not affect the results.

Initially we shall also assume that each eigenvalue of (A, B) is semisimple. (Actually it makes no difference whether the nonreal eigenvalues are semisimple or not). The canonical form (loc. cit.) then gives

$$A_\infty = diag(I, -I),\ B_\infty = 0, \tag{2.2}$$

$$A_F = diag(D^+, D^-, A_n)\ ,\ B_F = diag(I, -I, B_n) \tag{2.3}$$

where $D^\pm$ are diagonal matrices with diagonal entries $\lambda_j^\pm$, $1 \le j \le N_F^\pm$. These are the finite real eigenvalues of (A, B), and since the sign depends on B, we call λ_j^+ a $B-$**positive** eigenvalue**:** it admits an eigenvector x_j^+ such that $(x_j^+, Bx_j^+) > 0$. Similarly for the B-negative eigenvalues λ_j^-. The matrices A_n and B_n are $2N_n \times 2N_n$ matrices corresponding to the nonreal eigenvalues of (A, B). It is well known that the latter occur in conjugate pairs, so N_n is the number of such pairs.

We now apply the cancellation algorithm of [11] to the **finite** real eigenvalues. We start by "cancelling" c pairs of the form $\lambda_j^+ < \lambda_k^-$ with no eigenvalues in between. (This operation is recursive, and stops when there are no more pairs to cancel: see [11] for details). The remaining real eigenvalues can then be relabelled in the order

$$\rho_{n^-}^- \le ... \le \rho_1^- \le \rho_1^+ \le ... \le \rho_{n^+}^+. \tag{2.4}$$

Note that $c + n^+$ is the number of B-positive eigenvalues.

The main result of this section expresses $\nu(A, B)$ in terms of the above quantities.

Theorem 2.1.

$$\nu(A, B) = N_n + c + N_\infty^-. \tag{2.5}$$

Remark 2.2. It is possible to extend the cancellation to include the (negative type) infinite eigenvalues, viewed as $+\infty$. Then (2.5) continues to hold provided we interpret N_∞^- as the number of such eigenvalues that were not cancelled.

2.1. *Proof of Theorem 2.1 in the semisimple case.*

Proof. Noting the evident relation $\nu(A_\infty - \lambda B_\infty) = N_\infty^-$ and the fact [11, Lemma 2.4] that $\nu(A_n - \lambda B_n) = N_n$ for all λ, we shall assume without loss of generality that $N_n = N_\infty = 0$. For the purposes of this proof, we list the eigenvalues of (A, B) in ascending order (and counted by multiplicity):

$$\lambda_1 \le \lambda_2 \le \cdots \le \lambda_N\ ,$$

and we replace $\nu(A, B)$ by d. We proceed by induction on the size N of A and B to prove both equation (2.5) and the formula $\lambda_0 = \rho_1^-$ or (if ρ_1^- does not exist) $\lambda_0 = \lambda_1$ for the minimiser λ_0 of (1.3).

First note that if $(\lambda^\pm, e^\pm)$ are $\pm B$-positive eigenpairs of (A, B), then

$$(e^+, (A - \lambda B)e^+) = \lambda^+ - \lambda \quad \text{and} \quad (e^-, (A - \lambda B)e^-) = \lambda - \lambda^- \ . \tag{2.6}$$

It follows that

$$\nu(\lambda) \text{ jumps by } \pm 1 \text{ as } \lambda \text{ increases across } \lambda^\pm \ . \tag{2.7}$$

Hence we can (and will) choose λ_0 so that $\lambda_0 \in [\lambda_1, \lambda_N]$.

If $N = 1$, then we set $\lambda_0 = \lambda_1, c = d = 0$. For the inductive step (from N, with c, d, etc. distinguished by carets, to $N + 1$), we need to consider three cases:

<u>Case 1</u>: Suppose λ_{N+1} is B-positive. Then $c = \hat{c}$ and (2.6) shows that $\nu(\lambda) = \hat{\nu}(\lambda)$ if $\lambda \leq \lambda_{N+1}$. Therefore we can take $\lambda_0 = \hat{\lambda}_0 (= \rho_1^-$ or $\lambda_1)$ and $d = \hat{d}$.

<u>Case 2</u>: Suppose λ_{N+1} is B-negative but is not cancelled by a B-positive eigenvalue. Then $c = \hat{c}$. Also

$$\nu(\lambda) = \hat{\nu}(\lambda) + 1 \text{ for } \lambda < \lambda_{N+1} \text{ and } \nu(\lambda_{N+1}) = \hat{\nu}(\lambda_{N+1}). \tag{2.8}$$

Further, since neither ρ_1^- nor λ_{N+1} are cancelled, the numbers of B-positive and B-negative eigenvalues in $]\rho_1^-, \lambda_{N+1}[$ are equal. A similar argument holds for $[\lambda_1, \lambda_{N+1}[$ if $\lambda = \lambda_1$. Hence by (2.7) it follows that $\nu(\hat{\lambda}_0) = \hat{\nu}(\lambda_{N+1})$ and $\nu(\lambda) = d + 1$ for $\lambda < \lambda_{N+1}$. Thus taking $\lambda_0 = \lambda_{N+1}$ $(=\rho_1^-)$, we have $\nu(\lambda) = d = \hat{d}$.

<u>Case 3</u>: Suppose λ_{N+1} is b-negative and is cancelled by a B-positive eigenvalue, say λ^+. Then $c = \hat{c} + 1$ and (2.8) holds. It follows from (2.7) that

$$\nu(\lambda_{N+1}) = \hat{\nu}(\lambda_{N+1}) \geq \hat{\nu}(\lambda^+) + 1,$$

since there are no eigenvalues in $(\lambda^+, \lambda_{N+1})$. From the definition of $\hat{\nu}_m$,

$$\hat{\nu}(\lambda^+) \geq \hat{\nu}(\hat{\lambda}_0) = \hat{d},$$

so we conclude that $\nu(\lambda_{N+1}) \geq \hat{d} + 1$. It follows from (2.8) that we can take $\lambda_0 = \hat{\lambda}_0$ $(=\rho_1^-$ or $\lambda_1)$, and then $d = \hat{d} + 1$. ◻

2.2. *The infinite semisimple case.* We now admit, in addition to the situation of Subsection 2.1, the possibility that any finite eigenvalue can be nonsemisimple. We note, however, that B_F of (2.1) remains invertible since $\sigma(A_F, B_F)$ consists entirely of finite eigenvalues.

We need a Lemma which expresses the dependence of $d = \nu(A, B)$ on small perturbations of A.

Lemma 2.3. *Let P be a nonnegative definite $N \times N$ matrix and let $d(\mu)$ $= \nu(A + \mu P, B)$. Then, under the above assumptions, there exists $\mu_o > 0$ such that $d(\mu) = \nu(A, B)$, for all $\mu \in]0, \mu_o[$.*

Proof. As in the proof of Subsection 2.1, we can assume that there are neither nonreal nor infinite eigenvalues so $(A, B) = (A_F, B_F)$. Evidently $d(\mu) \le \nu(A, B)$ for any $\mu > 0$, so assume

$$d(\mu_n) < \nu(A, B) \tag{2.9}$$

for some sequence $\mu_n \searrow 0$. Let λ_n be a minimiser for $\nu(A, B)$ in (1.3) with A replaced by $A + \mu_n P$. Then $\lambda_n^{-1}(A + \mu_n P)x_n = Bx_n$ for some x_n of unit norm. If λ_n are unbounded, we pass to a subsequence of x_n with limit x, and we obtain $Bx = 0$, contradicting invertibility of $B(= B_F)$. Thus λ_n are bounded, and we pass to a subsequence of λ_n with limit λ. Now continuity of the eigenvalues under perturbation shows that, for large enough n,

$$d(\mu_n) = \nu(A + \mu_n P - \lambda_n B) \ge \nu(A - \lambda B) \ge \nu(A, B),$$

contradicting (2.9). ∎

Remark 2.4. If $\mu < 0$, the result need not hold. For example, if $C = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$, $D = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ and $P = I$, then $\nu(C, D) = 0$ but $d(\mu) = 1$ for small $\mu < 0$.

Remark 2.5. Note by [18, p. 64] that the eigenvalues of $(A + \mu I, B)$ are all semisimple for small enough $\mu \neq 0$. Thus if we choose $P = I$, and if c is defined as $\lim_{\mu \searrow 0} c(\mu)$, the limit from above of c for (1.1), then (2.5) continues to hold.

Remark 2.6. The perturbation of A does not have to be by μI in Remark 2.3. According to the canonical form (see [26]), our assumptions imply that (2.3) is replaced by

$$A_F = diag(A_1, \cdots, A_q, A_n), \;\; B_F = diag(B_1, \cdots, B_q, B_n)$$

where A_k, B_k, etc., involve certain (symmetrized) Jordan-like blocks. For example, if (A_1, B_1) correspond to eigenvalue λ_1, then

$$A_1 = \varepsilon \begin{bmatrix} 0 & \cdot & \cdot & \cdot & 0 & \lambda_1 \\ \cdot & \cdot & \cdot & \cdot & \lambda_1 & 1 \\ \cdot & \cdot & \cdot & \cdot & 1 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \lambda_1 & 1 & \cdot & \cdot & \cdot \\ \lambda_1 & 1 & 0 & \cdot & \cdot & 0 \end{bmatrix}, \quad B_1 = \varepsilon \begin{bmatrix} 0 & \cdot & \cdot & \cdot & 0 & 1 \\ \cdot & \cdot & \cdot & 0 & 1 & 0 \\ \cdot & \cdot & \cdot & \cdot & 0 & \cdot \\ \cdot & 0 & \cdot & \cdot & \cdot & \cdot \\ 0 & 1 & 0 & \cdot & \cdot & \cdot \\ 1 & 0 & \cdot & \cdot & \cdot & 0 \end{bmatrix}, \tag{2.10}$$

where $\varepsilon = \pm 1$. Then the multiple eigenvalue of (A_k, B_k) can be split by any nonnegative definite matrix μQ (of the right size), provided $q_{11} \neq 0$. This can be shown by direct calculation of $\det(A_k + \mu Q - \lambda B_k)$ and a Newton diagram argument.

2.3. *The general case.* We now allow nonsemisimple infinite eigenvalues as well, so (A, B) is an arbitrary nonsingular pair. We reduce this case to those already considered above, perturbing A this time by a negative semidefinite matrix acting on X_∞.

LEMMA 2.7. *Let $P = diag(P_\infty, 0)$ where P_∞ is a positive semidefinite $N_\infty \times N_\infty$ matrix. If $d(\mu) = \nu(A + \mu P, B)$, then there exists $\mu_o < 0$ such that for $\mu \in]\mu_o, 0[$,*

$$d(\mu) = \nu(A, B). \tag{2.11}$$

PROOF. Since (A_∞, B_∞) has no finite eigenvalues, $\det(A_\infty - \lambda B_\infty) \neq 0$, so $\nu(A_\infty - \lambda B_\infty)$ is independent of λ, and therefore

$$\nu(A, B) = \nu(A_\infty - \lambda_0 B_\infty) + \nu(A_F - \lambda_0 B_F), \tag{2.12}$$

where λ_0 is the minimiser in (1.3). Further

$$d(\mu) \leq \nu(A + \mu P - \lambda_0 B) = \nu(A_\infty + \mu P_\infty - \lambda_0 B_\infty) + \nu(A_F - \lambda_0 B_F). \tag{2.13}$$

Since $\nu(A_\infty - \lambda_0 B_\infty)$ is invertible, we have $\nu(A_\infty + \mu P_\infty - \lambda_0 B_\infty) = \nu(A_\infty - \lambda_0 B_\infty)$ for sufficiently small μ, so it follows from (2.12) and (2.13) that

$$d(\mu) \leq \nu(A, B). \tag{2.14}$$

The reverse inequality follows immediately from $\mu < 0$. ☐

REMARK 2.8. Lemma 2.7 can fail if the sign of μ is changed. For example, with C, D and P from Remark 2.2, $\nu(D, C) = 1$ but $d(\mu) = 0$ for small $\mu > 0$.

REMARK 2.9. As in Remark 2.3, we may split the infinite eigenvalues with $P_\infty = I_\infty$, thus producing a problem of the type considered in 2.2. We omit the details, since a more general case will be treated by a different method in Section 3.

Together with Remark 2.3, this completes the proof of Theorem 2.1 in the general case.

REMARK 2.10. The perturbation μI_∞ of Remark 2.6 may be generalised as in Remark 2.4. Instead of (2.3) the canonical form now gives

$$A_\infty = diag(A_{-r}, \ldots, A_{-1}), B_\infty = diag(B_{-r}, \ldots, B_{-1})$$

where A_{-j} and B_{-j} are again symmetrized Jordan-like blocks. Then the infinite eigenvalue of (A_{-j}, B_{-j}) can be split by a nonpositive perturbation μP where $p_{11} \neq 0$.

REMARK 2.11. It is possible to combine the above perturbations (acting on X_∞) with those of 2.2 (but acting on X_F). The details will be left to the reader.

3. Quasi-uniformly positive operators

We now consider a pair (A, B) of self adjoint operators on a Hilbert space H. Since we shall not assume discreteness of the spectrum, we define $\nu(A)$ as the dimension of a maximal subspace on which A is negative definite, with $\nu(\lambda) = \nu(A - \lambda B)$, etc., defined consequently as in Section 1.

We need three assumptions, the first being on A alone. Recall that a self adjoint operator U in a Hilbert space is called **uniformly positive** (*up*) if $\inf \sigma(U) > 0$, so U is positive definite with bounded inverse.

Assumption **A1**. A is **quasi-uniformly positive** (*qup*), i.e., $\inf \sigma_e(A) > 0$.

Here σ_e denotes essential spectrum. Thus A is *qup* means that

$$U = A + C \tag{3.1}$$

is *up* for some (compact symmetric) operator C of finite rank. Indeed, if A has positive (resp. nonpositive) spectral subspace $\mathcal{P}$ (resp. $\mathcal{N}$), then we may take $U = A\mid_{\mathcal{P}}, C = I\mid_{\mathcal{N}}$. See [4], [13] for further properties of *qup* operators.

Our second assumption ensures that A dominates B in a suitable sense.

Assumption **A2**. $D(a) \subset D(B)$

Here $D(a) = D(U^{1/2}) = D(|A|^{1/2})$ is the domain of the form a which may be defined by

$$\begin{aligned} a(x,y) &= (U^{1/2}x, U^{1/2}y) - (Cx, y) \\ &= (Ax, y) \text{ if } x \in D(A). \end{aligned} \tag{3.2}$$

Similarly the form b corresponding to B may be defined on $D(a)$ by virtue of **A2**. Indeed instead of (1.2) we could consider the "weak" eigenvalue problem

$$a(x, z) = \lambda b(x, z) \tag{3.3}$$

for $0 \neq x, z \in D(a) \cap D(b)$ under weaker assumptions. We shall however continue with the operator framework and the (rather strong) relative boundedness assumption **A2** for simplicity and because it uses constructions based on the original data (A, B). An approach using extensions (and different assumptions) will be described in 3.4.

Our third assumption extends the previous one of nonsingularity to infinite dimensions.

Assumption **A3**. (A, B) is **nonsingular**, i.e., $\sigma(A, B) \neq \mathbf{C}$.

Note that **A3** requires $A - \lambda B$ to be invertible for some $\lambda \in \mathbf{C}$. The first step in our analysis allows us to choose $\lambda \in \mathbf{R}$, so we can preserve self adjointness after an eigenvalue shift.

Lemma 3.1. *There exists $\lambda \in \mathbf{R}$ such that the operator $A - \lambda B$ is boundedly invertible and qup.*

Proof. It follows from (**A2**) that $D(A) \subset D(B)$ so $T(\lambda) = A - \lambda B$ is a holomorphic family of type (A) in the sense of [18, p. 375]. Therefore if $0 \in$

$\sigma(A)$, by [18, p.386] either $0 \in \sigma(A-\lambda B)$ for all $\lambda \in \mathbf{C}$ or there exist analytic functions $\mu_j(\lambda)$ such that $\mu_j(0)=0$, $\mu_j \neq 0$ for all j. Since the first alternative contradicts (**A3**), it follows that $0 \in \rho(A-\lambda B)$ for $\lambda \neq 0, |\lambda|$ sufficiently small. Moreover, if $|\lambda|$ is sufficiently small, then $U-\lambda B$ is uniformly positive definite, and so $A-\lambda B$ is *qup* for such λ. □

In what follows we shall assume (without loss of generality, by Lemma 3.1) that A is boundedly invertible and *qup*. Before proceeding we note the following consequence of our assumptions.

LEMMA 3.2. *If $x_n \in D(a)$ and $a(x_n) \to 0$ then $Bx_n \to 0$.*

PROOF. Writing $A' = |A|^{-1/2}\mathrm{sgn}A$, we note that B is closed in, and A' is bounded on, H so **A2** and the closed graph theorem show that

$$S = BA' \text{ is bounded on } H. \tag{3.4}$$

Thus if $a(x_n) \to 0$ then $y_n = (|A|^{1/2}\mathrm{sgn}A)x_n \to 0$ and so $Bx_n = Sy_n \to 0$. □

3.1. *Pontryagin space setting.* Our aim, roughly, is to split $\sigma(A,B)$ into "finite" and "infinite" parts corresponding to X_F and X_∞ in Section 2. To make this precise, we shall use the indefinite space $\Pi = D(a)$ with inner product given by $(x,y)_a := a(x,y)$ of (3.2). By our assumptions, Π is a Pontryagin space and we define $Q = A^{-1}B\,|_\Pi$.

LEMMA 3.3. *Q is a bounded self adjoint operator on Π.*

PROOF. Q is obviously symmetric and by (**A2**) Q is everywhere defined. The conclusion now follows from [12, Theorem VI.2.8] □

For any subspace Σ of Π we denote by $\nu(\Sigma)$ the dimension of a maximal subspace of Σ on which a is negative definite. In particular, $\nu(\Pi) = \nu(A)$ and $\nu(\Sigma) = 0 \iff a$ is nonnegative definite on Σ. By known results (cf. [22]) Q has a spectral function E, say, with a finite set of real critical points which satisfy the condition that

$$c \text{ is critical} \iff E(\Delta_c) \text{ has indefinite range,}$$

for any open interval Δ_c containing c. Abbreviating $\nu(E(\Delta)\Pi)$ to $\nu(\Delta)$, we see that if $c \in \Delta_c$ and if $\Delta \subset \Delta_c$ is sufficiently small then $\nu(\Delta) = 0$ if $c \notin \Delta$, $\nu(\Delta) = \nu(\Delta_c) > 0$ if $c \in \Delta$. Let

$$e_- = \min \sigma_e(Q)\ ,\ e_+ = \max \sigma_e(Q).$$

Note that $\sigma_e(Q) \subseteq [e_-, e_+]$, but the inclusion may be strict.

We denote the set of nonreal eigenvalues of Q by Ω_n. It is well known that the corresponding root subspaces span a finite dimensional subspace, say Π_n, and we write $Q_n = Q\,|_{\Pi_n}$.

DEFINITION 3.4. *A real number $\xi \geq e_+$ is a-**positive** for Q if ξ admits no a-nonpositive root vector, i.e., if ξ is either an a-positive eigenvalue, or else not an eigenvalue, of Q.*

The following lemma is a key to the connection between Q and the formulae for $\nu(A, B)$ in 3.2. By $\nu_Q(\xi)$ we mean the dimension of a maximal subspace on which $\xi I - Q$ is negative definite in Π, i.e., $\nu_Q(\xi) = \nu(\xi a - b)$. In parts (c) and (d) below, we choose positive $\eta_+ \in \rho(Q)$ such that $\eta_+ > e_+$ and $\nu(\Delta) = 0$ whenever $\bar{\Delta} \subset \Gamma :=]e_+, \eta_+[$ (i.e., Γ contains only a-positive points). This is always possible by finiteness of $\nu(\Pi) = \nu(A)$.

LEMMA 3.5. (a) *If $\xi < e_+$ then $\nu_Q(\xi) = \infty$.*

(b) *If e_+ is the limit of eigenvalues from the right, then $\nu_Q(e_+) = \infty$.*

(c) *If Ω is a real interval in $\rho(Q)$ then $\nu_Q(\xi)$ is constant for $\xi \in \Omega$.*

(d) *ν_Q is nonincreasing on $]e_+, \eta_+[$.*

(e) *If e_+ is a-positive for Q, then ν_Q is nonincreasing on $]-\infty, \eta_+[$.*

PROOF. (a) Suppose there exists a closed interval $\Omega_1 \subset]\xi, e_+[$ so that dim $\Pi_1 = \infty$ and $\nu(\Pi_1) = 0$, where Π_1 is the range of $E(\Omega_1)$. Then for all nonzero $x \in \Pi_1$,

$$((\xi I - Q)x, x)_a = \int_{\Omega_1} (\xi - \theta)(dE(\theta)x, x)_a < 0. \tag{3.5}$$

Suppose such an interval Ω_1 does not exist. It follows that e_+ is a limit from the right of ($a-$positive) eigenvalues ξ_j, say. If e_j are the corresponding eigenvectors and $x = \Sigma_j x_j e_j$, then

$$((\xi I - Q)x, x)_a = \Sigma_j (\xi - \xi_j)|x_j|^2 (e_j, e_j)_a < 0.$$

(b) Set $\xi = e_+$ above.
(c) This is a consequence of [14, Lemma 2.4].
(d) Choose $\theta \in]e_+, \xi[$ and write $\Omega_1 =]-\infty, \theta]$ and $\Omega_2 =]\theta, \infty[$. Let Π_j be the range of $E(\Omega_j)$ and write $Q_j = Q\mid_{\Pi_j}, j = 1, 2$. Then ν_{Q_1} is constant near ξ by (c), and ν_{Q_2} is nonincreasing near ξ by (2.6). By (c), ν_{Q_n} is constant over **R,** so the result follows from $\nu_Q = \nu_{Q_n} + \nu_{Q_1} + \nu_{Q_2}$.
(e) If e_+ is a limit of eigenvalues from the right, then an argument as for (3.5) shows that $\nu_Q(e_+) = \infty$, and the finite dimensional theory (2.6) shows that ν_Q is nonincreasing on a right neighbourhood of e_+. The result then follows from (a).

If $]e_+, \theta]$ contains no eigenvalues for some $\theta > e_+$, we define Q_j as in the proof of (d). Since Q_1 has no eigenvalues greater than e_+, $\nu_{Q_1}(\xi) = 0$ for $\xi \geq e_+$. Since ν_{Q_2} is constant on $[e_+, \theta[$ by (c), it follows that ν_Q is also constant on this interval, and the result now follows from (c). $\square$

We shall also need a dual result, involving the dimension $\pi_Q(\xi)$ of a maximal subspace on which $\xi I - Q$ is positive definite. The proof is similar and will be omitted.

COROLLARY 3.6. *π_Q is constant on $\rho(Q)$, is infinite for $\xi > e_-$ and even for $\xi \geq e_-$ if e_- is a limit from the left of a sequence of eigenvalues. If $e_- > \eta_- \in \rho(Q)$ and $]\eta_-, e_-[$ contains only a-positive points, then π is nondecreasing on $]\eta_-, e_-[$. If in addition e_- is $a-$positive then π is also nondecreasing at e_-.*

3.2. *Formulae for $\nu(a,b)$.* The next step is to relate the above properties of Q to the form pencil $a - \lambda b$. We let (a_n, b_n) be the restriction of (a,b) to Π_n, and with $\nu(a)$ defined as for $\nu(A)$ we write $\nu_n(\lambda) = \nu(a_n - \lambda b_n)$ and $\nu_n = \min\{\nu_n(\lambda) : \lambda \in \mathbf{R}\}$. (Simpler constructions will be given in 3.3 below). We discuss $\nu(a,b)$ in three separate cases.

3.2.1. $e_- > 0$, i.e., Q is *qup*.

In this case we choose η_+ as for Lemma 3.5 and we set $J_f =]-\infty, \eta_+^{-1}[, J_e = [\eta_+^{-1}, \infty[$. Note that $0 < \eta_+^{-1} < e_+^{-1}$, J_f corresponds to a finite set of eigenvalues, and J_e contains the essential spectrum $\{\lambda : \lambda^{-1} \in \sigma_e(Q)\}$. We use the notation $\Omega_f = \{\xi : \xi^{-1} \in J_f\} \cap \sigma(Q), Q_f = Q\,|_{\Pi_f}$ where Π_f is the range of $E(\Omega_f)$ and similarly for Q_e. We define corresponding $a_f = a\,|_{\Pi_f}, a_e$ etc., and we write $\nu_f = \min\{\nu_f(\lambda) : \lambda \in \mathbf{R}\}$ with λ_f as the corresponding minimizer, and similarly for ν_e. Finally we write

$$\delta = \nu(e_+) - \nu_f.$$

THEOREM 3.7. *With the above notation $\lambda_f \in J_f$ and $\nu(\lambda_f) = \nu_n + \nu_f + \nu_e$. If $\delta \geq 0$, then $\nu(a,b) = \nu(\lambda_f)$ and if $\delta < 0$, then $\nu(a,b) = \nu(\lambda_f) + \delta$.*

PROOF. Since $\nu(a - \lambda b) = \nu(\frac{1}{\lambda}a - b)$ for $\lambda > 0$, we obtain

$$\nu(a - \lambda b) = \nu_Q(\lambda^{-1}) \text{ if } \lambda > 0 \tag{3.6}$$

with similar equations for the restrictions Q_n to Π_n, etc. Replacing Q by Q_f and applying Lemma 3.5(c), we see that $\lambda_f \in J_f$. Replacing Q by Q_n instead, we see that $\nu_n(\lambda)$ is a constant (ν_n) for all (real) λ. Applying Lemma 3.5 (a) to Q_e, we see that $\nu_e(\lambda)$, and hence $\nu(\lambda)$, is infinite for $\lambda > e_+^{-1}$, so it suffices to consider $\lambda \leq e_+^{-1}$.

By definition and Lemma 3.5(c), ν_f is minimized at $\lambda_f \in J_f$ and is constant on $[\eta^{-1}, e_+^{-1}]$. By Lemma 3.5(c) and (d), ν_e is constant on J_f and is nondecreasing on $[\eta^{-1}, e_+^{-1}[$. Thus

$$\nu(\lambda) = \nu_n(\lambda) + \nu_f(\lambda) + \nu_e(\lambda) \tag{3.7}$$

has a minimizer at either λ_f or e_+^{-1}. If $\delta \geq 0, \lambda_f$ is a minimizer, and if $\delta < 0$ then we must use e_+^{-1}. ∎

Corollary 3.8. *If e_+ is either*
(i) a-positive (see Definition 3.2), or
(ii) a limit from the right of eigenvalues of Q
then $\delta \geq 0$ and $\nu(a, b) = \nu_n + \nu_f + \nu_e$.

Proof. In case (i), Lemma 3.5(e) shows that ν_e is nondecreasing at e_+^{-1}, and in case (ii), $\nu_e(e_+) = \infty$ by Lemma 3.5(b). Thus the previous proof shows that λ_f must be a minimizer of ν. This also implies $\delta \geq 0$. ∎

3.2.2. $e_+ < 0$, i.e., $-Q$ is *qup*.
In this case we define η_- as for Corollary 3.6, with $J_e =]-\infty, \eta_-^{-1}]$ and $J_f =]\eta_-^{-1}, \infty[$. The remaining notation is then as before except that

$$\delta = \nu(e_-) - \nu_f.$$

For $\lambda < 0$, from $\nu(a - \lambda b) = \nu(-\frac{1}{\lambda}a + b) = \pi(\frac{1}{\lambda}a - b)$ it follows that

$$\nu(a - \lambda b) = \pi_Q(\lambda^{-1}) \text{ if } \lambda < 0$$

and using this and Corollary 3.6 instead of (3.9) and Lemma 3.5 we obtain

Theorem 3.9. *Theorem 3.7 holds in the above notation, and if e_- is a-positive or is a limit from the left of eigenvalues for Q then $\delta \geq 0$.*

3.2.3. $e_- \leq 0 \leq e_+$.
This case includes the possibility that Q is compact, which is satisfied in many applications. Note that then $e_- = e_+ = 0$ is a limit of eigenvalues for Q, except in finite dimensions. We choose $\eta_\pm$ as before and we define $J_f =]\eta_-^{-1}, \eta_+^{-1}[$ and $J_e = [-\infty, \eta_-^{-1}] \cup [\eta_+^{-1}, \infty]$. The remaining notation is as before, except that we set

$$\delta = \min\{\delta_-, \delta_+\}, \quad \text{where } \delta_\pm = \nu(e_\pm) - \nu_f. \tag{3.8}$$

Theorem 3.10. *Theorem 3.7 remains valid, and if either of $e_\pm$ is a-positive (or a limit of eigenvalues from the appropriate side) for Q, then the corresponding $\delta_\pm \geq 0$.*

Proof. The proof is similar to that of Theorem 3.7, except that now J_f, on which $\nu_e(\lambda)$ is constant, has two finite endpoints $e_\pm$, so both they and λ_f are now potential minimizers of ν. If $\delta \geq 0$, then λ_f is a minimizer. If δ_-(resp. δ_+) $= \delta < 0$, then e_-(resp. e_+) is a minimizer. ∎

3.3. *Calculation of $\nu(A, B)$.* We now carry the analysis over to the original (A, B) setting. We note that

$$(1.2) \text{ is equivalent to } Qx = \lambda^{-1}x, \tag{3.9}$$

(if we allow $\lambda^{-1} = 0$, so $\lambda = \infty$).

Our next aim is to calculate the quantities in Theorem 3.7, in terms of the partition of $\sigma(A, B)$ used in 3.2. We start with the sum Σ_n of the root subspaces for (A, B) corresponding to eigenvalues in J_n, and we denote the dimension of Σ_n by $2N_n$. We denote by N_e the sum of the dimensions of maximal A-nonpositive subspaces of the root subspaces corresponding to J_e. Note that ∞ is included if B is not 1-1. Finally we denote by c the number of real $(\lambda_j^+, \lambda_k^-)$ pairs cancelled from J_f (we extend the definition according to Remark 2.3 if necessary). Note that c depends on the choice of J_f, i.e. of $\eta_\pm$. The number δ is defined by (3.8).

THEOREM 3.11. *If $\delta \geq 0$, then*

$$\nu(A, B) = N_n + c + N_e. \tag{3.10}$$

PROOF. By (3.9) and Section 2, $\nu_n = N_n$ and $\nu_f = c$. Note that $\sigma(A, B) \cap J_f$ consists of eigenvalues of finite (algebraic) multiplicity and

$$0 \in J_f, \tag{3.11}$$

since Ω_f is bounded. Since $\nu_e(\lambda)$ is constant for $\lambda \in J_f$, we can evaluate it at $\lambda = 0$ by (3.11), so $\nu_e = \nu(\Pi_e)$. Now any contribution to $\nu(\Pi_e)$ must come from a critical point (necessarily an eigenvalue) for Q, or from a negative type eigenvalue. By (3.9) and Pontryagin's invariant subspace theorem, such contributions are precisely those counted in N_e. The result now follows from Theorem 3.7. ∎

REMARK. The N_e term includes the contribution N_∞^- from the root subspace at ∞ (as in Section 2). Indeed N_e reduces to N_∞^- in the finite dimensional case, and in general N_e can be regarded as the contribution from a neighbourhood of ∞.

Alternatively, one could view the contributions from the other root subspaces corresponding to J_e in the same way as those corresponding to the (finite) eigenvalues in 2.1 and 2.2, i.e., via the quadratic form b (see also Theorem 3.14 below). Explicit calculations can be carried out using the blocks (2.10) of the canonical form.

COROLLARY 3.12. *If Q is compact, then*

$$\nu(A, B) = N_n + c + N_\infty^-,$$

where N_∞^- is the maximal dimension of $a-$negative subspaces of the root subspace at ∞. In particular, this holds if A has compact resolvent.

PROOF. It follows from Theorem 3.10 that in this case $\delta \geq 0$, hence by Theorem 3.11 and $N_e = N_\infty^-$ the first conclusion follows.

For the final contention, we assume via (3.4) that $S = BA'$ is bounded on H, where $A' = |A|^{-1/2}\mathrm{sgn}A$. Moreover $T = A'B|_\Pi$ is defined on all of Π by **A2** and

$$\|Tx\|_a \leq \|S\| \, \|x\|_a \,. \tag{3.12}$$

Finally if x_n converges weakly to zero in Π then by compactness of $|A|^{-1/2}$, $x_n \to 0$ in H, i.e., $|A|^{-1/2}x_n \to 0$ in Π, and so $|A|^{-1/2}|_\Pi$ is a compact operator on Π. With (3.12) this shows that Q is compact. ∎

We now consider the case $\delta < 0$. In fact we assume $\delta_+ = \delta < 0$, (the case $\delta_- = \delta < 0$ being analogous). We also assume that the root subspace R_+ at e_+^{-1} is a-nondegenerate (i.e., e_+ is not a singular critical point for Q, cf. [8, Theorem 3.1]. This last assumption will be examined in the remarks after Theorem 3.13 below). In particular, there is a nondegenerate span Σ of Jordan chains in R_+ whose a-orthocomplement is a-positive (cf. [8, Theorem 3.2]).

At this point it is convenient to replace the decomposition $\Pi_n^\perp = \Pi_f \oplus_a \Pi_e$ by $\Pi_n^\perp = \Pi_f' \oplus_a \Pi_e'$ where $\Pi_f' = \Pi_f \oplus_a \Sigma$, $\Pi_f' = \Pi_f \ominus_a \Sigma$.
Let J be an interval containing finitely many eigenvalues. As before, $\pi(J)$ ($\nu(J)$, resp.) denotes the dimension of a maximal $a-$positive ($a-$negative, resp.) subspace of the sum of the corresponding algebraic eigenspaces.

THEOREM 3.13. *Under the above assumptions,* $\nu(A,B) = N_n + c + N_e + \pi([\lambda_f, e_+^{-1}[) - \nu(]\lambda_f, e_+^{-1}])$.

REMARK. The final term includes the contribution from Σ, which we may assume has been absorbed into Π_f as above.

PROOF. By Theorem 3.7, $\nu(A,B) = \nu(\lambda_f) + \delta$, so by Theorem 3.11, it is enough to show that the $\pi(\cdot)$ and $\nu(\cdot)$ terms in the statement make up δ. Since $\nu_n(\lambda)$ is constant, δ may be split into $\delta_f + \delta_e$ where

$$\delta_g = \nu_g(e_+^{-1}) - \nu_g(\lambda_f)$$

and $g = f$ or e. An easy application of (2.7) shows that $\delta_f = \pi([\lambda_f, e_+^{-1}[) - \nu(]\lambda_f, e_+^{-1}[)$ (recall that there are no a-negative eigenvalues in $]\eta^{-1}, e_+^{-1}[$). Similarly (using the above absorption to apply (2.7) again) $\delta_e = -\nu(\Sigma) = -\nu(\{e_+^{-1}\})$ according to our convention. ∎

REMARK. If the minimizer $e_\pm$ is a singular critical point, from $\delta_+ < 0$ we have $\nu(A,B) = \nu(e_\pm^{-1})$; for nontriviality we assume that this index is finite. Using (3.9) we see that it is enough to calculate $\nu_Q(e_\pm)$. For this we use a nonnegative perturbation as in [23], so that the new Q has no singular critical points. If the perturbation is small enough, then $\nu_Q(e_\pm)$ is unchanged, so we may perform the calculation as before.

REMARK. In [14, Section 4] it is shown how to calculate $\nu_Q(\lambda)$ when λ is a regular critical point, but in the singular case only an inequality is given. Thus the methods here appear to improve on [14] in this respect.

3.4. *Krein space setting.* If we assume

A0. B is 1-1,

then an alternative approach is possible, avoiding the reciprocals used in the previous constructions. We indicate the main steps as follows.

(i) We continue to assume **A1** and we replace **A2** by

A2′. $D(A) \subset D(B)$ and $Ax_n \to 0$ implies $Bx_n \to 0$ for all $x_n \in D(A)$.

It follows from **A0**, **A1** and **A2′** that A has a bounded inverse on H. Arguing as for (3.4), we see that BA^{-1} is bounded, say $\|Bx\| \leq c\,\|Ax\|$ for all $x \in D(A)$. By Heinz's inequality, $b(x) \leq c^{1/2}a(x)$ and we deduce (cf. [18, p.572])

A2″. $D(a) \subset D(b)$ and $a(x_n) \to 0$ implies $b(x_n) \to 0$ for all $x_n \in D(a)$.

Here $D(a)$ is as before and similarly $D(b) = D(|B|^{1/2})$.

(ii) Arguing again as for (3.4), we see from **A2″** that $V = |B|^{1/2}A'$ is bounded on H, where $A' = |A|^{-1/2}\mathrm{sgn}A$. Thus

$$\left\|A^{-1}Bx\right\|_b \leq \|V\|^2\,\|x\|_b\,,\ \text{for all } x \in D(B). \tag{3.13}$$

We write $(x,y)_b = b(x,y)$, $\|x\|_b = \left\||B|^{1/2}x\right\|$ etc., and we complete $(D(b), \|.\|_b)$ to a Krein space $\mathcal{K}$. Evidently $A^{-1}B$ is symmetric on the dense subspace $D(B)$ of $\mathcal{K}$. Thus by (3.13), $A^{-1}B$ extends to a bounded symmetric operator $\bar{Q}$ on $\mathcal{K}$. Moreover by construction of $\mathcal{K}$, $\bar{Q}$ is 1-1.

(iii) From the above considerations, $R := \bar{Q}^{-1}$ is self adjoint and boundedly invertible in $\mathcal{K}$. Moreover the range of $\bar{Q}$ is dense, and since it is contained in the closure of the range of $A^{-1}B$, we see that R is an extension of $B^{-1}A$ which is densely defined in $\mathcal{K}$. Noting that $D(B^{-1}A) \subset D(A) \subset D(B)$, we have

$$((A-\lambda B)x,x) = b((R-\lambda I)x,x) \tag{3.14}$$

for all $x \in D(B^{-1}A)$, which enables us to connect $\nu(A,B)$ with $\nu_b(R-\lambda I)$.

(iv) Setting $\lambda = 0$ in (3.14), we have

$$b(y,\bar{Q}y) = (A\bar{Q}y,\bar{Q}y) = (By, A^{-1}By)$$

for all $y = B^{-1}Ax$, and using boundedness of $\bar{Q}$ we see that

$$\nu(R) = \nu(\bar{Q}) = \nu(A^{-1}) = \nu(A).$$

In particular, R is *qup* in $\mathcal{K}$, and, being invertible, is definitizable, cf. [13]. Thus the total algebraic multiplicity corresponding to the nonreal eigenvalues of R is finite, the sets $\sigma_+(R) \cap \mathbf{R}_-$ and $\sigma_-(R) \cap \mathbf{R}_+$ consist of finitely many isolated eigenvalues of finite total multiplicity and all finite critical points of R are of finite index. Moreover R has a spectral function F with critical points, see [22].

(v) Let J_f be the interval defined in Subsection 3.3. Then J_f does not intersect the essential spectrum of R. Define

$$m = \inf J_f \quad , \quad M = \sup J_f.$$

The number c was defined in Subsection 3.3 using $b-$signs of eigenvectors (and the perturbations of Section 2 if necessary). Since the inner product in $\mathcal{K}$ is generated by b, an analogue of c can be defined for the eigenvalues of R. Similarly an analogue of N_n gives the sum of dimensions of the algebraic eigenspaces of the nonreal eigenvalues of R with positive imaginary parts.

Since R is *qup*, there are finitely many positive (negative, resp.) points λ such that for all intervals J containing λ, and not having critical points of R as one of its endpoints, the space $F(J)\mathcal{K}$ has a maximal negative (positive, resp.) subspace of finite dimension, say $\kappa_-(J)$ ($\kappa_+(J)$), resp.). Let $\kappa_-(\lambda)$ (resp. $\kappa_+(\lambda)$) be the minimum of all the numbers $\kappa_-(J)$ (resp. $\kappa_+(J)$)) when J varies over such intervals J.

We can now reformulate Theorem 3.11 for the present situation.

THEOREM 3.14. *If B is 1-1 then*

$$\min \nu_b(R-\lambda I) = N_n + c + \sum_{\lambda<0} \kappa_+(\lambda) + \sum_{\lambda>0} \kappa_-(\lambda) \ . \tag{3.15}$$

If $m=-\infty$, then the first sum on the right-hand side is absent; if $M=\infty$, then the second sum on the right-hand side is absent.

PROOF. The proof is similar to the proof Theorem 3.11, using R and F instead of Q and E: the details will be omitted. □

REMARKS.

1. This approach offers advantages of directness. For example, no reciprocals are involved, and the underlying space is based on the form b, which in typical appliications is easier to evaluate then a. On the other hand, B must be 1-1 and it may be difficult to relate R to the original data (A,B).
2. In view of Lemma 3.2, **A1-A3** imply **A2$'$**. Thus if **A0** holds then the assumptions of 3.3 are stronger than those of 3.4, and these are in turn stronger than **A2$''$**, which would be appropriate for the setting of (3.3).

4. APPLICATION TO VARIATIONAL PROBLEMS

In this section we shall relate the minimal index to several formulae in the literature, mostly from the last decade, characterizing eigenvalues variationally, but with "index shifts". We start by defining the B-**positive multiplicity** of a semisimple eigenvalue λ of (A,B) is defined as the dimension of a maximal B-positive subspace of $N(A-\lambda B)$. Let λ_j^+ (resp. λ_j^-) be the j^{th} eigenvalue of (A,B), listed in nondecreasing order and counted by B-positive (resp. negative) multiplicity. Next let

$$\sigma_j^+ = \sup\{\inf\{(x,Ax) : (x,Bx)=1, \quad x\in S\cap D(A)\} : \text{codim } S = j-1\}$$

where we interpret $\inf\emptyset$ as $-\infty$. A classical variational formula gives $\lambda_j^+ = \sigma_j^+$ (with sup and inf attained) in the "right definite" case $B>0$, and this (with sup attained) turns out to hold also in the left definite case. In recursive form, this dates back to Richardson [27], but in max-inf form it seems quite recent: cf. [21] for matrices and [7, 10] for operators. Earlier related work can also be found in [19, 29].

4.1. *Cancellation.* The "cancellation" procedure of [11] is quite recent, and arose out of attempts to characterize eigenvalues $\lambda_j^+$of (A, B) in the case (which cannot occur in the above situations) where λ_k^- exist greater than λ_j^+. In the case where a sequence of c eigenvalues λ_k^- is preceded by c eigenvalues λ_l^+, a variational principle of the form

$$\lambda_j^+ = \sigma_{j+c}^+ \tag{4.1}$$

is established in [10] for certain matrix and qup operator pencils. This formula is extended to the general matrix pencil with semisimple real eigenvalues and B invertible in [11]. In these cases, $n = N_\infty^- = 0$ so Corollary 3.12 gives $\nu(A, B) = c$. In other words, $\nu(A, B)$ is the index shift in (4.1).

4.2. *The case* $B \geq 0$. An early explicit variational principle with shifted index was given by Allegretto [1]. He considered (1.2) in the form

$$(-\Delta + q)u = \lambda w u \tag{4.2}$$

with Dirichlet boundary conditions on a smooth domain Ω and a weight function $w \geq 0$ where $w^{-1}(0)$ was a smooth subdomain of Ω. It is easy to see that in this case all eigenvalues are real and B-positive, so $n = c = 0$. Moreover Allegretto obtains

$$\lambda_j^+ = \sigma_{j+d}^+ \tag{4.3}$$

where d is the number of nonpositive eigenvalues of $-\Delta + q$ on $w^{-1}(0)$. From this one can deduce that $d = N_\infty^-$ and so by Corollary 3.12 the index shift in (4.3) is again $\nu(A, B)$. Actually, related versions of this result can be found in [16] for the minimal oscillation number of a right semidefinite Sturm-Liouville problem and in [6] for asymptotes of two parameter eigencurves for certain self adjoint operator pencils, cf. 4.4 below.

4.3. *Maximal definite subspaces.* In a recent paper [7] the authors gave a variational principle of the form (4.3) with $d = \kappa_+(F)$ where F is a particular subspace of the Krein space K (see 3.4). In fact F is the span of certain root subspaces, and $\kappa_+(F)$ can be evaluated via matrices of the form B_1 in (2.10). Using [11, Lemmas 2.3 and 2.4], one can show that $\kappa_+(F) = n + c$, and since in this case B is 1-1 so $N_\infty^- = 0$, we obtain $d = \nu(A, B)$ for the index shift.

Other authors have considered special cases from different viewpoints. For sufficiently large λ_j^+ and finite dimensional H, Najman and Ye [25] give an expression for d based directly on the sum of the κ_+ for each individual block in (2.10). Again for sufficiently large λ_j^+, Allegretto and Mingarelli [2] study (4.2) and and give $d = k - j$ in (4.3) where k is the eigencurve index for λ_j^+ (so $A - \lambda_j^+ B$ has k nonpositive eigenvalues). The equivalence between this and $d = n + c$ is discussed in [7, Subsection 4.3], but we add here that the "analysis of the eigencurves" mentioned there may be carried out by first replacing A by $A + \mu I$ and then decreasing μ continuously to zero (cf. [5] for this technique).

Recently cases with B noninvertible have been studied for $\dim H < \infty$. See [26] for an extension of [25], $d = \kappa_+(F)$ now including a term equivalent to N_∞^- by a sum over blocks of the form A_1 in (2.10) (cf. Remark 2.7). In [9] more eigenvalues are characterized, and $d = n + c +$ N_∞^- is derived explicitly using the results of Section 2.

4.4. *Minimal variational index.* In the case $B \geq 0$, the eigencurves are graphs of nondecreasing functions, cf. [5]. A certain number (say ν_0) of the eigencurves lies entirely above the λ-axis, so the $(\nu_o + 1)^{th}$ eigencurve cuts the λ-axis at the minimal eigenvalue λ_0. Since this is also the eigenvalue of minimal index, we see that $\nu_0 = \nu(A, B)$. An explicit formula for the minimal oscillation count ν_0 in the SL case (with nonnegative weight) was given in [16], and this was motivation for Allegretto's analysis [1] mentioned above - see also [6] where a formula for ν_0 can be found, including that of [16], and describing eigencurve asymptotes.

An important feature of Allegretto's analysis is that it also gives the "missing" $\sigma_j^+ (1 \leq j \leq \nu_0)$ as $-\infty$. Thus if we define the minimal **variational index** j_m as the minimal j so that σ_j^+ is finite then

$$j_m = \nu(A, B) + 1$$

at least if $B \geq 0$. It turns out, however, that this relation also holds in other cases with B indefinite [9, 11], so we obtain a new way of estimating $\nu(A, B)$, directly from approximations to the variational quantities σ_j^+.

References

[1] Allegretto, W., Second order elliptic equations with degenerate weight, Proc. Amer. Math. Soc. 107(1989), 989-998.

[2] Allegretto, W., Mingarelli, A. B., Boundary problems of the second order with an indefinite weight-function, J. Reine Angew. Math. 398(1989), 1-24.

[3] Atkinson, F.V., Jabon, D., Indefinite Sturm-Liouville problems, Argonne National Lab. reprint ANL-84-73, 1984, pp. 31-45.

[4] Binding, P., A canonical form for self-adjoint pencils in Hilbert space, Int. Equ. Op. Th., 12(1989), 324-342.

[5] Binding, P., Browne, P.J., Applications of two parameter spectral theory to symmetric generalised eigenvalue problems, Applic. Anal. 29(1988), 107-142.

[6] Binding, P., Browne, P. J., Eigencurves for two-parameter self-adjoint ordinary differential equations of even order, J. Diff. Equns 79(1989), 289-303.

[7] Binding, P., Najman, B., A variational principle in Krein space, Trans. Amer. Math. Soc 342(1994), 489-499.

[8] Binding, P., Najman, B., Regularity of finite type critical points for self-adjoint operators in Krein space, Operator Theory: Advances and Applications, Vol. 80, pp. 79-89. Birkhäuser, Basel, 1995.

[9] Binding, P., Najman, B., Ye, Q., A variational principle for eigenvalues of pencils of Hermitian matrices, Int. Equ. Op. Theory **35** (1999), 398–422.

[10] Binding, P., Ye, Q., Variational principles without definiteness conditions, SIAM J. Math. Anal. 22(1991), 1575-1583.

[11] Binding, P., Ye, Q., Variational principles for indefinite eigenvalue problems, Lin. Alg. Appl. 218(1995), 251-262.

[12] Bognár, J., Indefinite Inner Product Spaces, Springer-Verlag, 1974.

[13] Ćurgus, B., Najman, B., Quasi-uniformly positive operators in Krein spaces, Operator Theory: Advances and Applications, Vol. 80, pp. 90-99. Birkhäuser, Basel, 1995.

[14] Dijksma, A., Gheondea, A., On the signatures of selfadjoint pencils, Operator Theory: Advances and Applications, Vol. 80, pp. 112-136. Birkhäuser, Basel, 1995.

[15] Daho, K., Langer, H., Sturm-Liouville operators with an indefinite weight function, Proc. Roy. Soc. Edin. 78A(1977), 161-191.

[16] Everitt, W.K.N., Kwong, M.K., Zettl, A., Oscillation of eigenfunctions of weighted regular Sturm-Liouville problems, J. London Math. Soc., 27(1983), 106-120.

[17] Haupt, O., Über eine Methode zum Beweise von Oszillationstheoremen, Math. Ann., 76(1915), 67-104.

[18] Kato, T., Perturbation Theory for Linear Operators, Springer-Verlag, 1976.

[19] Kühne, R., Über eine Klasse J-selbstadjungierter Operatoren, Math. Ann., 154(1964), 56-69.

[20] Lancaster, P., Shkalikov, A., Damped vibrations of beams and related spectral problems, Can. Appl. Math. Quart., 2(1994), 45-90.

[21] Lancaster, P., Ye, Q., Variational properties and Rayleigh quotient algorithms for symmetric matrix pencils, in Oper. Theory Adv. Appl. 40, Birkhäuser, Basel, pp. 247-278.

[22] Langer, H., Spectral functions of definitizable oprators in Krein space, Lecture Notes in Math. 948(1982), pp.1-46, Springer-Verlag.

[23] Langer, H., Najman, B., Instability of singular critical points of definitizable operators, Int. Equ. Oper. Theory, 28(1997), 60-71.

[24] Mingarelli, A. B., A survey of the regular weighted Sturm-Liouville problem−the non-definite case, Proc. Workshop Appl. Diff. Equ., Tsinghua Univ., Beijing, 1985.

[25] Najman, B., Ye, Q., A minimax characterization for eigenvalues of hermitian pencils, Lin. Alg. Appl. 144 (1991), 217–230.

[26] Najman, B., Ye, Q., A minimax characterization for eigenvalues of hermitian pencils II, Lin. Alg. Appl., 191(1993), 183-197.

[27] Richardson, R.G.D., Das Jacobische Kriterium der Variationsrechnung und die Oszillationseigenschaften linearer Differentialgleichungen 2 Ordnung, Math. Ann. 68(1910), 279-304.

[28] Richardson, R.G.D., Theorems of oscillation for two linear differential equations of the second order with two parameters, Trans. Amer. Math. Soc., 13(1912), 22-34.

[29] Svarzman, P.A., Inequalities for the eigenvalues of J-Hermitian and J-unitary operators, (Russian), Mat. Issled. 4(1969), 33-45.

Paul Binding
Department of Mathematics and Statistics
University of Calgary
2500 University Drive N.W.
Calgary, Alberta
Canada T2N 1N4
E-mail: binding@ucalgary.ca

Received: 14.09.98
Revised: 04.08.99