STUDIES IN COMPUTATIONAL MATHEMATICS 5

editors: C. BREZINSKI and L. WUYTACK

TOPICS IN VALIDATED COMPUTATIONS

J. HERZBERGER
Editor

NORTH-HOLLAND

STUDIES IN COMPUTATIONAL MATHEMATICS 5

Editors:

C. BREZINSKI
University of Lille
Villeneuve d'Ascq, France

L. WUYTACK
University of Antwerp
Wilrijk, Belgium

ELSEVIER
Amsterdam – Lausanne – New York – Oxford – Shannon – Tokyo

TOPICS IN VALIDATED COMPUTATIONS

Proceedings of the IMACS-GAMM International Workshop on Validated Computation, Oldenburg, Germany, 30 August-3 September 1993

edited by

Jürgen HERZBERGER
Fachbereich Mathematik
Universität Oldenburg
Germany

1994

ELSEVIER
Amsterdam – Lausanne – New York – Oxford – Shannon – Tokyo

ELSEVIER SCIENCE B.V.
Sara Burgerhartstraat 25
P.O. Box 211, 1000 AE Amsterdam, The Netherlands

Library of Congress Cataloging-in-Publication Data

IMACS-GAMM International Workshop on Validated Computations (1993 : Oldenburg, Germany)
Topics in validated computations : proceedings of IMACS-GAMM International Workshop on Validated Computations, Oldenburg, Germany, 30 August - 3 Septmeber, 1993 / edited by Jürgen Herzberger.
p. cm. -- (Studies in computational mathematics ; 5)
Includes bibliographical references.
ISBN 0-444-81685-2
1. Interval analysis (Mathematics)--Congresses. 2. Numerical calculations--Verification--Congresses. 3. Algorithms--Congresses. I. Herzberger, Jürgen. II. Title. III. Series.
QA297.75.I43 1993
512'.5--dc20 94-36669
CIP

ISBN: 0 444 81685 2

This book is printed on acid-free paper.

Printed in the Netherlands

PREFACE

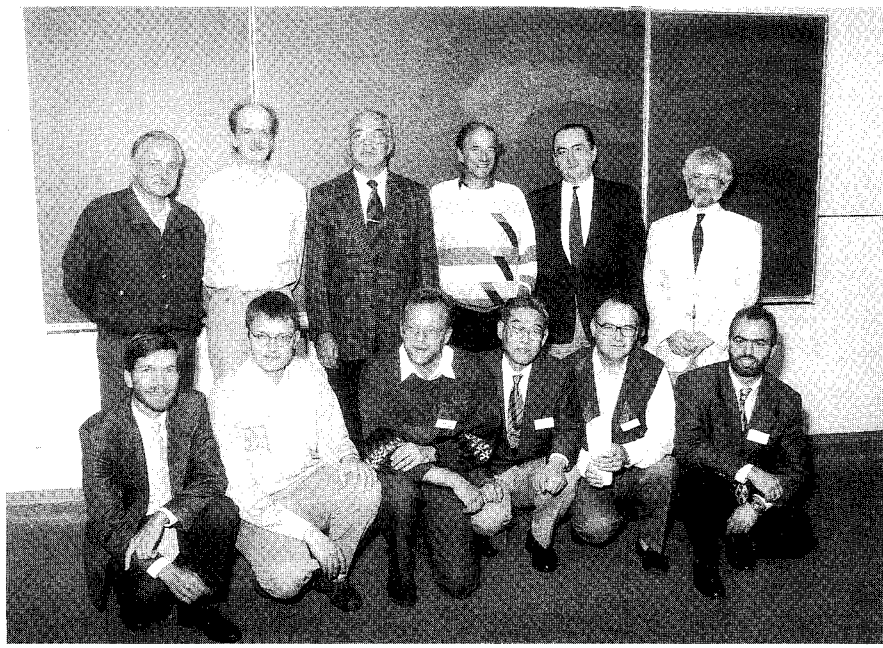
This volume is the result of the IMACS-GAMM International Workshop on Validated Computations held at the University of Oldenburg (Germany) from August 30 to September 3, 1993. The intention of this workshop, in which 29 scientists from 10 countries took part, was to bring together experts in some selected areas of the application of interval arithmetic in the development of algorithms with result verification. Most of the talks delivered during this meeting were supposed to be survey talks or expository contributions on the fields represented by selected speakers. Among them were presented applications in numerical linear algebra, solving nonlinear equations, eigenvalue problems, ODE's and PDE's, and optimization. Since it is now 20 years ago that the monograph of G. Alefeld and J. Herzberger "Einführung in die Intervallrechnung" appeared, the choice covered by this volume was according to the following idea: some of the papers should give an updated survey of how things have developed during the years after the appearance of the book – for example, the chapters on interval Newton methods and iterative methods of including the inverse of a matrix – but the main portion of the contributions should deal with areas of applications which were not included in the monograph, mainly because of lack of convincing results at that early stage of interval arithmetic. Just to mention a few of them: optimization, applications in PDE's and eigenvalue problems. In this sense, this volume is thought to be a kind of continuation of the old monograph. In order to make this volume easily readable, the notations are unified and introduced in a first introductory chapter with some basic facts about interval arithmetic. Here, the main ideas and properties of the interval operations are summarized with some notations differing from those chosen in the old monograph taking into account the development of the last two decades in interval arithmetic literature. Again in this volume, there exists a kind of appendix, although not explicitly announced as such. Firstly, there is a chapter with instructions how to implement the algorithms nowadays described in this volume by reviewing the enormous progress having been made in the design and creation of programming environments for this purpose (compared with the old-fashioned ALGOL 60 procedures in the monograph mentioned). Secondly, there is a short critical report on the complexity of some common problems in interval analysis where we want to calculate optimal solutions. Last, but not least, as a reference to the numerous practical applications of interval arithmetic, we want to mention the chapter on selected examples for solving nontrivial problems in engineering.

The purpose of this volume is to provide the interval analysis community with surveys of some important recent developments in the creation of validated numerical algorithms in the form of a book with non-overlapping chapters written by several experts in their fields. At the same time we want to inform the numerical analysts and the appliers of numerical software about the enormous variety of problem-solving algorithms now available, even for sophisticated problems which were out of reach at the beginning of research in this direction some two decades ago. This volume is self-contained and offers the interested reader applications in nearly all main branches of numerical analysis, and it can be used as an up-to-minute textbook for studying advanced interval algorithms.

We want to express our sincere gratitude to the sponsors of the international workshop, Volkswagen-Stiftung, Ministerium für Wissenschaft und Kultur des Landes Niedersachsen

and the Universitätsgesellschaft Oldenburg e.V., and to all participants for their encouraging discussions during the presentation of these papers.

Oldenburg, May 1994 J. Herzberger



Authors during the workshop. Front row (from left to right): S.M. Rump, H. Behnke, C. Jansson, T. Yamamoto, J. Herzberger, R. Rihm. Second row (from left to right): H. Heindl, M. Plum, G. Alefeld, C. Ullrich, F. Goerisch, G. Mayer

CONTENTS

Topics in Validated Computations
J. Herzberger (Editor)
1994 Elsevier Science B.V.

Inclusions for Eigenvalues of Selfadjoint Problems

H. Behnke[a] and F. Goerisch[b]

[a]TU Clausthal, Institut für Mathematik, Erzstraße 1
38678 Clausthal-Zellerfeld, Germany

[b]TU Braunschweig, Institut für Angewandte Mathematik, Pockelsstraße 14
38106 Braunschweig, Germany

1. Introduction

Numerous problems from both the natural and engineering sciences, as well as some important questions within mathematics, lead to eigenvalue problems. The class of self-adjoint eigenvalue problems is arguably the most important class of eigenvalue problems, because numerous problems that occur in practice belong to this class, and happily, far-reaching results can be obtained for selfadjoint eigenvalue problems as a consequence of the theory of selfadjoint operators.

A selfadjoint eigenvalue problem is an eigenvalue problem of the form

$$\begin{aligned}&\text{"Find pairs } (\lambda,\Phi)\in \mathbb{K}\times D \text{ such that } \Phi\neq 0 \text{ and}\\ &\mathcal{M}(f,\Phi)=\lambda\mathcal{N}(f,\Phi) \text{ for all } f\in D\text{"},\end{aligned} \tag{1}$$

here $\mathcal{M}$ and $\mathcal{N}$ are Hermitian sesquilinear forms. If $\mathcal{M}$ is positive definite, (1) is called left definite; if $\mathcal{N}$ is positive definite, (1) is called right definite.

We begin with the properties assumed for the left definite case, which we will need below:

L1: D is a vector space over $\mathbb{K}$ ($\mathbb{K}=\mathbb{R}$ or $\mathbb{K}=\mathbf{C}$); $\mathcal{M}$ is a Hermitian sesquilinear form on D. $\mathcal{M}(f,f)>0$ holds for all $f\in D$, $f\neq 0$.

L2: $\mathcal{N}$ is a Hermitian sesquilinear form on D.

For the corresponding properties that are assumed in the right definite case, see section 5.

It is a lamentable truth that eigenvalue problems that are important in practice can rarely be solved in closed form; one must generally rely on numerical methods. Most numerical methods simply provide approximations for the eigenvalues, but they do not make it possible to state how far away the computed value is from the true eigenvalue. Since selfadjoint eigenvalue problems that are either right or left definite can only have real eigenvalues, the problem of obtaining approximations and corresponding error bounds is equivalent to the determination of upper and lower bounds for the eigenvalues.

The large interest shown in eigenvalue bounds, not only by mathematicians, but also by physicists, chemists, and engineers has many reasons. We name three of them here:

1. Bounds for eigenvalues (as opposed to approximate values) can be used to prove certain mathematical theorems; for example, existence and inclusion statements for solutions to nonlinear boundary value problems by means of M. Plum's method [34].

2. Lower bounds for eigenvalues are necessary in order to compare predictions of physical theories with experimental results; for example, fine structure corrections in quantum mechanical problems.

3. The knowledge of bounds for eigenvalues makes it possible in many cases to determine the reliability of approximation procedures.

For a large class of problems, good upper bounds for the positive eigenvalues can be determined relatively easily by means of the Rayleigh–Ritz procedure (see Theorems 11 and 14). Basically, there are three classes of methods for computing lower bounds for eigenvalues (disregarding methods with a very limited scope of application):

1. methods based on inclusion theorems,

2. the method of intermediate problems, and

3. the method of orthogonal invariants.

The methods depending on inclusion theorems go back to G. Temple [41], L. Collatz [12], and N. J. Lehmann [28, 29]. For these methods we need rough *a priori* information on the location of one of the eigenvalues and furthermore require pairs (v, w) of trial functions (see Example 1 for further explanation). The method of intermediate problems was discovered by A. Weinstein [44] and has been refined by N. W. Bazley, D. W. Fox [4] (cf. [16]) and C. Beattie [5, 6]. For this method, one needs the eigenvectors of a neighbouring eigenvalue problem in closed form. The underlying idea for the method of orthogonal invariants can be found for a special example as early as 1933 by E. Trefftz [42]. It was later developed to a widely applicable method by G. Fichera [15]. For the method of orthogonal invariants, an eigenvalue problem with a compact operator is considered (an integral operator in most cases), then detailed information on this operator is required.

Until 1989 no one paid attention to the possible influence of rounding errors in the calculation of bounds using these methods. As a consequence, some published results are not correct. Only through the use of computational tools developed by U. Kulisch [25, 26] and his students G. Alefeld, J. Herzberger [3], and S. M. Rump [36] (cf. [23, 24]), has the control of rounding errors become possible.

In our opinion, the methods which are most suitable for use in conjunction with interval arithmetic methods belong to class 1 mentioned above. Hence, we will restrict our discussion to methods from this class.

Before we describe the theory for the procedures of class 1 systematically, we will illustrate the most important points by means of a simple example.

Example 1

$$\begin{aligned} &\text{“}\textit{Find } \lambda \in \mathbb{R},\ \Phi \in C^2[0,\pi],\ \Phi \neq 0 \ \textit{ such that} \\ &-\Phi''(x) \;=\; \lambda\,(1+\sin(x))\,\Phi(x) \quad \textit{for } x \in (0,\pi) \\ &\Phi(0) \;=\; \Phi(\pi) = 0\text{”} \end{aligned} \tag{2}$$

This problem can be found in [12, page 176] ("buckling of a simply supported bar"), where there are also upper and lower bounds for the smallest eigenvalue given. Problem (2) is

equivalent to (1), if we define $\mathbb{K} = \mathbb{R}$,

$$\begin{aligned} D &:= \{f \in C^2[0,\pi] \,:\, f(0) = f(\pi) = 0\}, \\ \mathcal{M}(f,g) &:= \int_0^\pi f'(x)\, g'(x)\, dx, \\ \mathcal{N}(f,g) &:= \int_0^\pi (1+\sin(x))\, f(x)\, g(x)\, dx. \end{aligned}$$

The equivalence can be shown by means of integration by parts. If $v \in D$, $v \neq 0$, *the Rayleigh quotient* $\dfrac{\mathcal{M}(v,v)}{\mathcal{N}(v,v)}$ *yields an upper bound for the smallest eigenvalue* λ_1 *of (2). Using* $v(x) := \sin(x)$, *we obtain*

$$\lambda_1 \leq \frac{3\pi}{3\pi+8} \leq 0.540\,884.$$

Since problem (2) is right definite as well as left definite, Temple's theorem both for right definite and for left definite problems can each be used to compute a lower bound for λ_1.

If we use the fact that $\mathcal{N}$ *is positive definite, then we can proceed as follows: first, we choose* $v \in D$, $v \neq 0$ *and determine a* $w \in C[0,\pi]$ *such that*

$$\mathcal{M}(f,v) = \mathcal{N}(f,w) \quad \text{for all } f \in D.$$

Secondly, we obtain a (rough) lower bound ρ *for the second smallest eigenvalue, for instance by means of a comparison theorem (see Theorem 10). If* $\rho > \dfrac{\mathcal{M}(v,v)}{\mathcal{N}(v,v)}$, *then the Temple quotient* $\dfrac{\mathcal{M}(w,w) - \rho\mathcal{N}(v,v)}{\mathcal{N}(v,v) - \rho\mathcal{M}(v,v)}$ *will be a lower bound for* λ_1. *Now, for* $v(x) = \sin(x)$, *we obtain* $w(x) = \dfrac{\sin(x)}{1+\sin(x)}$. *Since* $1 + \sin(x) \leq 2$ *holds true, the eigenvalues of the problem*

$$\begin{aligned} -\Phi''(x) &= \lambda\, 2\, \Phi(x) \quad \text{for } x \in (0,\pi) \\ \Phi(0) &= \Phi(\pi) = 0 \end{aligned}$$

(which can be solved in closed form), are lower bounds for the corresponding eigenvalues of (2). Hence, we can choose $\rho = 2$, *and we obtain as a lower bound for* λ_1,

$$0.538\,809 \leq \frac{12\pi - 24}{3\pi + 16} \leq \lambda_1.$$

Were we instead making use of the fact that $\mathcal{M}$ *is positive definite, we would proceed as follows: Choose* v *and* ρ *as in the right definite case and determine* w *so that*

$$\mathcal{M}(f,w) = \mathcal{N}(f,v) \quad \text{for all } f \in D. \tag{3}$$

Hence, the function w *must be the exact solution of the boundary value problem*

$$\begin{aligned} -w''(x) &= (1+\sin(x))\, v(x) \quad \text{for } x \in (0,\pi) \\ w(0) &= w(\pi) = 0. \end{aligned}$$

Now if $\rho > \dfrac{\mathcal{M}(v,v)}{\mathcal{N}(v,v)}$, *the Temple quotient* $\dfrac{\mathcal{M}(v,v) - \rho\mathcal{N}(v,v)}{\mathcal{N}(v,v) - \rho\mathcal{M}(w,w)}$ *will be a lower bound for* λ_1. *Using* $v(x) = \sin(x)$, *we find* $w(x) = \frac{1}{4}(\sin^2(x) + 4\sin(x) - x^2 + \pi x)$, *and we obtain an improved lower bound for* λ_1:

$$0.540\,184 \leq \frac{8(3\pi + 16)}{2\pi^3 + 39\pi + 192} \leq \lambda_1 .$$

The procedure for left definite problems yields better lower bounds for problem (2) than the procedure for right definite problems. It has been shown in [18], that this is the case for a large class of problems independent of the choice of trial functions. For this reason, the procedure for left definite problems has attracted exceptional interest. We will deal first with such problems where $\mathcal{M}$ is positive definite in the next three sections.

For eigenvalue problems more complicated than Example 1, it is frequently far more involved or perhaps even impossible to find a w which satisfies equation (3) exactly. If w cannot be computed exactly, but a good lower bound $-\mathcal{B}$ for $-\mathcal{M}(w,w)$ is known, then we obtain the inequality

$$\frac{\mathcal{M}(v,v) - \rho\mathcal{N}(v,v)}{\mathcal{N}(v,v) - \rho\mathcal{B}} \leq \frac{\mathcal{M}(v,v) - \rho\mathcal{N}(v,v)}{\mathcal{N}(v,v) - \rho\mathcal{M}(w,w)} \leq \lambda_1 . \tag{4}$$

We will turn to the question of how to determine such a $\mathcal{B}$ in section 2.

If upper bounds are to be computed by means of the Rayleigh–Ritz procedure not only for λ_1 but also for eigenvalues with higher indices, eigenvalues of symmetric matrix eigenvalue problems have to be determined. Similar computational tasks are also involved for the determination of lower bounds. We will explain this in sections 3 and 4. A corresponding theory for right definite problems will be outlined in section 5.

By the theorems of sections 3, 4, and 5, the problem of determining bounds for the eigenvalues of (1) will be reduced to the problem of determining bounds for eigenvalues of matrix eigenvalue problems with high and guaranteed precision. We address this problem in section 6. In section 7 we illustrate the power of the methods by means of numerical examples.

Though the main purpose of this paper is to survey an important class of existing procedures, this paper also contains numerous new results, particularly in sections 2 and 7.

2. Construction of Complementary Variational Principles

If we define

$$F(f) := \mathcal{N}(f,v) \quad \text{for all} \quad f \in D$$

(with v as in (3)), then F is a linear functional on D and (3) has the form of a linear equation with bilinear forms:

"Find $u \in D$ such that
$\mathcal{M}(f,u) = F(f)$ for all $f \in D$". (5)

For the solution u, the following variational principle is well known:

$$-\mathcal{M}(u,u) = \min\{\mathcal{M}(f,f) - F(f) - \overline{F(f)} \;:\; f \in D\} \tag{6}$$

Unfortunately, this variational principle does not provide a mechanism to construct the constant $\mathcal{B}$ which is necessary in (4). For this purpose, we need instead a variational principle complementary to (6). We will proceed to formulate an abstract complementary variational principle, which appears to include all such principles that are important in practice, as special cases. Toward this end we need to introduce a further assumption:

L3: X is a vector space over $\mathbb{K}$; $\mathcal{B}$ is a Hermitian sesquilinear form on X; and $T : D \to X$ is a linear operator. $\mathcal{B}(f,f) \geq 0$ holds for all $f \in X$ and $\mathcal{M}(f,g) = \mathcal{B}(Tf, Tg)$ for all $f, g \in D$.

From this assumption, we obtain the following general complementary variational principle:

Theorem 1
If L1 and L3 are satisfied, if F is a linear functional on D, and if u is a solution of (5), we have

$$-\mathcal{M}(u,u) = \max\{-\mathcal{B}(g,g) \ : \ g \in V\} \tag{7}$$

where

$$V := \{g \in X \ : \ \mathcal{B}(Tf,g) = F(f) \ \text{ for all } \ f \in D\}\,. \tag{8}$$

Proof: Let $g \in V$. Then we have $0 \leq \mathcal{B}(Tu - g, Tu - g) = \mathcal{M}(u,u) - F(u) - \overline{F(u)} + \mathcal{B}(g,g) = \mathcal{B}(g,g) - \mathcal{M}(u,u)$. Because of $Tu \in V$, this yields the assertion. □

In Theorem 1, the assumption L3 can be replaced by the somewhat more geometric condition

L3*: $(X, \mathcal{B}(.,.))$ is a pre-Hilbert space over $\mathbb{K}$. T is an isometric linear mapping from the pre-Hilbert space $(D, \mathcal{M}(.,.))$ to $(X, \mathcal{B}(.,.))$.

In applications, the weaker assumption L3 is often easier to use, however.

Whether the variational principle (7) is useful in practice or not depends on how laborious it is to find elements in V. In the following we will give a general procedure which shows how X, $\mathcal{B}$, and T can be constructed so that elements from V are easy to find. But first, we will show two possible choices of X, $\mathcal{B}$, and T for a particular problem, namely the Neumann boundary value problem of potential theory. (For a definition of Sobolev spaces, which we use in our examples, see [13, 43].)

Example 2
Let $\Omega \subset \mathbb{R}^2$ be a bounded, convex domain with a polygonal boundary $\partial\Omega$ and let $h \in L_2(\Omega)$. The quantities from L1 are defined as follows:

$$\mathbb{K} := \mathbb{R}, \quad D := \{f \in H^1(\Omega) \ : \ \int_\Omega f\,dx\,dy = 0\}\,,$$

$$\mathcal{M}(f,g) := \int_\Omega \left(\frac{\partial f}{\partial x}\frac{\partial g}{\partial x} + \frac{\partial f}{\partial y}\frac{\partial g}{\partial y}\right)dx\,dy \ \text{ for } \ f,g \in D\,,$$

Additionally, define

$$F(f) := \int_\Omega f\,h\,dx\,dy \quad \textit{for} \quad f \in D\,.$$

Now, problem (5) is the weak form of the boundary value problem $-\Delta u = h$ *in* Ω, $\frac{\partial u}{\partial n} = 0$ *on* $\partial\Omega$, $\int_\Omega u\,dx\,dy = 0$. *Let* $\gamma_1 \in \mathbb{R}$ *with* $\gamma_1 > 0$ *and*

$$\int_\Omega ((\frac{\partial f}{\partial x})^2 + (\frac{\partial f}{\partial y})^2)dx\,dy \geq \gamma_1 \int_\Omega f^2 dx\,dy \quad \textit{for all} \quad f \in D$$

(Poincaré's inequality). Then the quantities X, $\mathcal{B}$, *and* T *defined by*

$$X := D \times L_2(\Omega)\,, \quad Tf := \begin{pmatrix} f \\ f \end{pmatrix} \quad \textit{for} \quad f \in D\,, \tag{9}$$

$$\mathcal{B}(\begin{pmatrix} f_0 \\ f_1 \end{pmatrix}, \begin{pmatrix} g_0 \\ g_1 \end{pmatrix}) := \int_\Omega (\frac{\partial f_0}{\partial x}\frac{\partial g_0}{\partial x} + \frac{\partial f_0}{\partial y}\frac{\partial g_0}{\partial y})dx\,dy + \gamma_1 \int_\Omega (f_1 g_1 - f_0 g_0)dx\,dy \quad \textit{for} \quad \begin{pmatrix} f_0 \\ f_1 \end{pmatrix}, \begin{pmatrix} g_0 \\ g_1 \end{pmatrix} \in X\,, \tag{10}$$

satisfy assumption L3. If we define

$$V_0 := \{ \begin{pmatrix} f \\ f + \gamma_1^{-1}(h + \Delta f) \end{pmatrix} \in X \;:\; f \in H^2(\Omega) \cap D,\ \frac{\partial f}{\partial n} = 0 \quad \textit{on} \quad \partial\Omega \}\,, \tag{11}$$

$V_0 \subset V$ *holds true (referring to the* V *defined in (8)). Hence, elements from* V *can be found relatively easy.*

There is a second possibility for choosing X, $\mathcal{B}$, *and* T. *Let* $\gamma_2 \in \mathbb{R}$ *with* $\gamma_2 > 0$ *and*

$$\int_\Omega ((\frac{\partial f}{\partial x})^2 + (\frac{\partial f}{\partial y})^2)dx\,dy \geq \gamma_2 \int_{\partial\Omega} f^2 ds \quad \textit{for all} \quad f \in D$$

(Poincaré's inequality and trace theorem). The quantities X, $\mathcal{B}$, *and* T *are defined by*

$$X := D \times L_2(\partial\Omega)\,, \quad Tf := \begin{pmatrix} f \\ f \end{pmatrix} \quad \textit{for} \quad f \in D\,,$$

$$\mathcal{B}(\begin{pmatrix} f_0 \\ f_1 \end{pmatrix}, \begin{pmatrix} g_0 \\ g_1 \end{pmatrix}) := \int_\Omega (\frac{\partial f_0}{\partial x}\frac{\partial g_0}{\partial x} + \frac{\partial f_0}{\partial y}\frac{\partial g_0}{\partial y})dx\,dy + \gamma_2 \int_{\partial\Omega} (f_1 g_1 - f_0 g_0)ds$$

for $\begin{pmatrix} f_0 \\ f_1 \end{pmatrix}, \begin{pmatrix} g_0 \\ g_1 \end{pmatrix} \in X$, *thus the assumption L3 is satisfied. If we define*

$$V_0 := \{ \begin{pmatrix} f \\ f - \gamma_2^{-1}(\frac{\partial f}{\partial n}) \end{pmatrix} \in X \;:\; f \in H^2(\Omega) \cap D,\ -\Delta f = h \quad \textit{in} \quad \Omega \}\,,$$

then $V_0 \subset V$ *holds true and again elements of* V *can be found relatively easily.*

With these two different choices for X, $\mathcal{B}$, *and* T, *elements from* V *can be constructed either by satisfying the boundary condition or by satisfying the differential equation.*

Now, the procedure for constructing X, $\mathcal{B}$, and T and obtaining elements from V, as illustrated by means of Example 2, shall now be formulated in a general way. For this purpose, we need to consider additional assumptions beyond L1.

V1: $\tilde{D}$ is a vector space over $\mathbb{K}$ which contains D as a subspace. $\tilde{\mathcal{M}}$ is a Hermitian sesquilinear form on $\tilde{D}$. $\tilde{\mathcal{M}}(f,f) \geq 0$ holds for all $f \in \tilde{D}$ and $\tilde{\mathcal{M}}(f,g) = \mathcal{M}(f,g)$ for all $f,g \in D$.

V2: Y is a subspace of $\tilde{D}$, and $r \in \mathbb{N}$. For $i = 1,\ldots,r$ we denote by $(H_i, \langle .,. \rangle_i)$ Hilbert spaces over $\mathbb{K}$, $T_i : \tilde{D} \to H_i$ and $M_i : Y \to H_i$ are linear operators so that

$$\tilde{\mathcal{M}}(f,g) = \sum_{i=1}^{r} \langle T_i f, M_i g \rangle_i \quad \text{holds for all} \quad f \in D, g \in Y\,. \tag{12}$$

V3: F is a linear functional on D. Let $h_i \in H_i$ for $i = 1,\ldots,r$ be such that

$$F(f) = \sum_{i=1}^{r} \langle T_i f, h_i \rangle_i \quad \text{for all} \quad f \in D\,. \tag{13}$$

V4: There are real constants $q_i > 0$ and $\kappa_i \in \{0,1\}$ $(i = 1,\ldots,r)$ such that

$$\tilde{\mathcal{M}}(f,f) \geq \sum_{i=1}^{r} \kappa_i q_i \langle T_i f, T_i f \rangle_i \quad \text{for all} \quad f \in \tilde{D}\,. \tag{14}$$

Following Definition 2 below, we will explain how to choose the quantities appearing in V1, V2, V3, and V4, such that we obtain the results of Example 2 as a special case of the procedure given in the next two theorems. An example with a partial differential equation of fourth order is discussed in 7.3.

Theorem 2
If assumptions L1, V1, V2, and V4 hold true, the quantities X, $\mathcal{B}$, and T defined by

$$X := \tilde{D} \times \prod_{i=1}^{r} H_i\,,$$

$$\mathcal{B}\left(\begin{pmatrix} f_0 \\ f_1 \\ \vdots \\ f_r \end{pmatrix}, \begin{pmatrix} g_0 \\ g_1 \\ \vdots \\ g_r \end{pmatrix}\right) := \tilde{\mathcal{M}}(f_0,g_0) + \sum_{i=1}^{r} \kappa_i q_i (\langle f_i, g_i \rangle_i - \langle T_i f_0, T_i g_0 \rangle_i) \text{ for } \begin{pmatrix} f_0 \\ f_1 \\ \vdots \\ f_r \end{pmatrix}, \begin{pmatrix} g_0 \\ g_1 \\ \vdots \\ g_r \end{pmatrix} \in X\,,$$

$$Tf := \begin{pmatrix} f \\ T_1 f \\ \vdots \\ T_r f \end{pmatrix} \quad \text{for} \quad f \in D$$

satisfy L3.

Proof: The proof is elementary. □

In order to construct elements from V (see (8)), we need to introduce further notation.

Definition 1
If assumptions L1, V1, V2, V3, and V4 are satisfied, we define

$$G : Y \to \prod_{i=1}^{r} H_i\,, \quad Gf := \begin{pmatrix} T_1 f - q_1^{-1} M_1 f \\ \vdots \\ T_r f - q_r^{-1} M_r f \end{pmatrix} \quad \text{for} \quad f \in Y\,, \quad \text{and } z := \begin{pmatrix} q_1^{-1} h_1 \\ \vdots \\ q_r^{-1} h_r \end{pmatrix}.$$

With this definition, we obtain

Theorem 3
Let assumptions L1, V1, V2, V3, and V4 be satisfied and let X, $\mathcal{B}$, and T be defined as in Theorem 2. If $g \in Y$ and

$$(1 - \kappa_i)(M_i g - h_i) = 0 \quad \text{for} \quad i = 1, \ldots, r\,, \tag{15}$$

then we have $\mathcal{B}(Tf, \begin{pmatrix} g \\ Gg + z \end{pmatrix}) = F(f)$ for all $f \in D$.

Proof: For all $f \in D$ the following holds true:

$$\begin{aligned} \mathcal{B}(Tf, \begin{pmatrix} g \\ Gg + z \end{pmatrix}) &= \tilde{M}(f, g) + \sum_{i=1}^{r} \kappa_i (\langle T_i f, q_i (T_i g - q_i^{-1} M_i g + q_i^{-1} h_i) \rangle_i - \langle T_i f, q_i T_i g \rangle_i) \\ &= \sum_{i=1}^{r} \langle T_i f, M_i g \rangle_i + \sum_{i=1}^{r} \kappa_i \langle T_i f, h_i - M_i g \rangle_i = \sum_{i=1}^{r} \langle T_i f, h_i \rangle_i = F(f)\,. \quad \square \end{aligned}$$

If the assumptions L1, V1, V2, V3, and V4 are satisfied, then by combining Theorem 1 with Theorem 2, we obtain a variational principle which is complementary to (6). Theorem 3 then supplies elements which are admissible for this variational principle. The numbers κ_i from V4 play an important role here; these constants provide the coupling between inequality (14), which is used for constructing the variational principle, and condition (15), which has to be satisfied for the construction of admissible elements. Frequently, the determination of admissible elements can be simplified considerably at the price of using a more complicated inequality in (14); this can be of great importance in practice.

Definition 2
Assuming that L1, V1, V2, V3, and V4 are satisfied, we define V_0 by

$$V_0 := \{ \begin{pmatrix} g \\ Gg + z \end{pmatrix} \in X \;:\; g \in Y,\ (1 - \kappa_i)(M_i g - h_i) = 0 \quad \text{for} \quad i = 1, \ldots, r \}\,.$$

If X, $\mathcal{B}$, T are defined as in Theorem 2 and V as in Theorem 1, we have $V_0 \subset V$ due to Theorem 3. Thus, V_0 contains those elements admissible for the complementary variational principle (7), which can be constructed by means of Theorem 3. The question whether the set V even may be replaced by V_0 in the variational principle (7), will be answered further down by Theorem 4.

Now we show how the particular choices for X, $\mathcal{B}$, T, and V_0, that we gave in Example 2 (which presented itself *intuitively*) can be obtained *systematically* by the general procedure just described. Recalling the assumptions and notations already given for Example 2 above, we define the quantities appearing in V1, V2, V3, and V4 as follows:

$$\tilde{D} = D\,,\ \tilde{\mathcal{M}} = \mathcal{M}\,,\ Y = H^2(\Omega) \cap D\,,\ r = 2\,,$$

$$H_1 := L_2(\Omega)\,,\ \langle f, g\rangle_1 := \int_\Omega f\, g\, dx\, dy \quad \text{for} \quad f, g \in H_1\,,$$

$$H_2 := L_2(\partial\Omega)\,,\ \langle f, g\rangle_2 := \int_{\partial\Omega} f\, g\, ds \quad \text{for} \quad f, g \in H_2\,,$$

$$T_1 f := f \quad \text{for} \quad f \in \tilde{D}\,,\ T_2 f := f \quad \text{for} \quad f \in \tilde{D}\,,$$

$$M_1 f := -\Delta f \quad \text{for} \quad f \in Y\,,\ M_2 f := \frac{\partial f}{\partial n} \quad \text{for} \quad f \in Y\,,$$

$$h_1 := h\,,\ \kappa_1 := 1\,,\ q_1 := \gamma_1\,,\ h_2 := 0\,,\ \kappa_2 := 0\,,\ q_2 := 1\,.$$

Now we construct X, $\mathcal{B}$, T, and V_0 according to Theorem 2 and Definition 2, and obtain the following equations:

$$X := D \times L_2(\Omega) \times L_2(\partial\Omega)\,,\quad T f := \begin{pmatrix} f \\ f \\ f \end{pmatrix} \quad \text{for} \quad f \in D\,,$$

$$\mathcal{B}\left(\begin{pmatrix} f_0 \\ f_1 \\ f_2 \end{pmatrix}, \begin{pmatrix} g_0 \\ g_1 \\ g_2 \end{pmatrix}\right) := \int_\Omega \left(\frac{\partial f_0}{\partial x}\frac{\partial g_0}{\partial x} + \frac{\partial f_0}{\partial y}\frac{\partial g_0}{\partial y}\right) dx\, dy + \gamma_1 \int_\Omega (f_1 g_1 - f_0 g_0)\, dx\, dy$$

$$\text{for} \quad \begin{pmatrix} f_0 \\ f_1 \\ f_2 \end{pmatrix}, \begin{pmatrix} g_0 \\ g_1 \\ g_2 \end{pmatrix} \in X\,, \text{ and}$$

$$V_0 := \left\{ \begin{pmatrix} f \\ f + \gamma_1^{-1}(h + \Delta f) \\ f \end{pmatrix} \in X \ :\ f \in H^2(\Omega) \cap D, \frac{\partial f}{\partial n} = 0 \ \text{ on } \ \partial\Omega \right\},$$

which are the same as (9) – (11), if the unnecessary third component is omitted.

If we define $h_1 := h$, $\kappa_1 := 0$, $q_1 := 1$, $h_2 := 0$, $\kappa_2 := 1$, $q_2 := \gamma_2$ (and the other quantities as above), we obtain the second possibility of our Example 2:

$$X := D \times L_2(\Omega) \times L_2(\partial\Omega)\,,\quad T f := \begin{pmatrix} f \\ f \\ f \end{pmatrix} \quad \text{for} \quad f \in D\,,$$

$$\mathcal{B}\left(\begin{pmatrix} f_0 \\ f_1 \\ f_2 \end{pmatrix}, \begin{pmatrix} g_0 \\ g_1 \\ g_2 \end{pmatrix}\right) := \int_\Omega \left(\frac{\partial f_0}{\partial x}\frac{\partial g_0}{\partial x} + \frac{\partial f_0}{\partial y}\frac{\partial g_0}{\partial y}\right) dx\, dy + \gamma_2 \int_{\partial\Omega} (f_2 g_2 - f_0 g_0)\, ds$$

$$\text{for } \begin{pmatrix} f_0 \\ f_1 \\ f_2 \end{pmatrix}, \begin{pmatrix} g_0 \\ g_1 \\ g_2 \end{pmatrix} \in X \text{ , and}$$

$$V_0 := \{ \begin{pmatrix} f \\ f \\ f - \gamma_2^{-1} \frac{\partial f}{\partial n} \end{pmatrix} \in X \ : \ f \in H^2(\Omega) \cap D, \ -\Delta f = h \ \text{ in } \ \Omega \} .$$

This is again the same result as in Example 2, if the unnecessary second component is omitted.

By the considerations described above, the problem of establishing a variational principle complementary to (6) has been reduced to the problem of determining quantities with the properties mentioned in V1, V2, V3, V4. What we still need are rules of how to construct these quantities. For a variety of different classes of linear equations, we can establish such rules; we will demonstrate this below for a class of boundary value problems with ordinary differential equations.

Example 3

Assumptions and notations: *Let* $m \in \mathbb{N}$, a, $b \in \mathbb{R}$, $a < b$;

let $p_i \in C^i[a,b]$ *such that* $p_i(x) > 0$ *for* $x \in [a,b], i = 0, \ldots, m$; $h \in C[a,b]$;

let $\sigma_i(a) \in \mathbb{R}, \sigma_i(b) \in \mathbb{R}$, *such that* $\sigma_i(a) > 0, \sigma_i(b) > 0$, *for* $i = 0, \ldots, m-1$;

let $\tau_i(a) \in \mathbb{R}, \tau_i(b) \in \mathbb{R}$, *for* $i = 0, \ldots, m-1$;

$\alpha_i \in \{0,1\}, \beta_i \in \{0,1\}$ *for* $i = 0, \ldots, m-1$.

$$L_i f := \sum_{k=i}^{m} (-1)^{k-i} (p_k f^{(k)})^{(k-i)} \text{ for } f \in C^{2m}[a,b], i = 0, \ldots, m .$$

Boundary value problem under consideration:

$$\begin{aligned} & u \in C^{2m}[a,b], \ L_0 u = h , \\ & (1-\alpha_i) u^{(i)}(a) + \alpha_i (\sigma_i(a) u^{(i)}(a) - L_{i+1} u(a)) = \alpha_i \tau_i(a) \ \text{ for } \ i = 0, \ldots, m-1 , \\ & (1-\beta_i) u^{(i)}(b) + \beta_i (\sigma_i(b) u^{(i)}(b) + L_{i+1} u(b)) = \beta_i \tau_i(b) \ \text{ for } \ i = 0, \ldots, m-1 . \end{aligned} \tag{16}$$

If $\alpha_i = 0$, *then the corresponding boundary condition at* a *is essential (stable), otherwise it is natural (unstable). (The same holds true for* β_i *and* b.)

Weak form of (16): *In order to formulate the boundary value problem (16) in terms of sesquilinear forms, we define* $\mathbb{K} = \mathbb{R}$;

$$D := \{ f \in C^m[a,b] \ : \ (1-\alpha_i) f^{(i)}(a) = (1-\beta_i) f^{(i)}(b) = 0 \ \text{ for } \ i = 0, \ldots, m-1 \} ,$$

$$\begin{aligned} \mathcal{M}(f,g) := \textstyle\sum_{i=0}^{m} \int_a^b p_i f^{(i)} g^{(i)} dx \ & + \ \textstyle\sum_{i=0}^{m-1} \sigma_i(a) f^{(i)}(a) g^{(i)}(a) \\ & + \ \textstyle\sum_{i=0}^{m-1} \sigma_i(b) f^{(i)}(b) g^{(i)}(b) \ \text{ for } \ f, g \in D , \end{aligned}$$

$$F(f) := \int_a^b f \, h \, dx + \sum_{i=0}^{m-1} f^{(i)}(a) \tau_i(a) + \sum_{i=0}^{m-1} f^{(i)}(b) \tau_i(b) \ \text{ for } \ f \in D .$$

Obviously, assumption L1 is satisfied.

For all $f \in C^m[a,b]$, $g \in C^{2m}[a,b]$ *we have*

$$\sum_{i=0}^{m} \int_a^b p_i f^{(i)} g^{(i)} dx = \int_a^b f L_0 g \, dx + \sum_{i=0}^{m-1} (f^{(i)}(b) L_{i+1} g(b) - f^{(i)}(a) L_{i+1} g(a)) \tag{17}$$

(integration by parts); this implies that (16) and (5) are equivalent.

Definition of the quantities appearing in V1, V2, V3: *We construct* $\tilde{D}$ *by "omitting" some of the boundary conditions occuring in the definiton of* D. *Which of the boundary conditions will be omitted, is controlled by the choice of certain numbers* $\hat{\alpha}_i$, $\hat{\beta}_i$.

Precisely, we assume $\xi \in \{0,1\}$, $\hat{\alpha}_i, \hat{\beta}_i \in \{0,1\}$ *for* $i = 0, \ldots, m-1$ *and define the quantities appearing in V1, V2, V3 by*

$$\tilde{D} := \{ f \in C^m[a,b] \; : \; (1-\hat{\alpha}_i)(1-\alpha_i) f^{(i)}(a) = 0 \;\; \textit{for} \;\; i = 0, \ldots, m-1, \\ (1-\hat{\beta}_i)(1-\beta_i) f^{(i)}(b) = 0 \;\; \textit{for} \;\; i = 0, \ldots, m-1 \},$$

$$\tilde{\mathcal{M}}(f,g) := \textstyle\sum_{i=0}^{m} \int_a^b p_i f^{(i)} g^{(i)} dx \;\; + \;\; \sum_{i=0}^{m-1} \sigma_i(a) f^{(i)}(a) g^{(i)}(a) \\ + \;\; \sum_{i=0}^{m-1} \sigma_i(b) f^{(i)}(b) g^{(i)}(b) \;\; \textit{for} \;\; f, g \in \tilde{D}.$$

$$Y := \tilde{D} \cap C^{2m}[a,b],$$

$$r := 2m+1,$$

$$H_i := \begin{cases} L_2(a,b) & \textit{for} \;\; i = 1, \\ I\!R & \textit{for} \;\; i = 2, \ldots, r, \end{cases}$$

$$\langle f, g \rangle_i := \begin{cases} \int_a^b f g \, dx & \textit{for} \;\; f, g \in H_1, i = 1, \\ f g & \textit{for} \;\; f, g \in H_i, i = 2, \ldots, r, \end{cases}$$

$$T_i f := \begin{cases} f & \textit{for} \;\; i = 1, \\ f^{(i-2)}(a) & \textit{for} \;\; i = 2, \ldots, m+1, \\ f^{(i-m-2)}(b) & \textit{for} \;\; i = m+2, \ldots, r, \end{cases}$$

for $f \in \tilde{D}$,

$$M_i f := \begin{cases} L_0 f & \textit{for} \;\; i = 1, \\ \alpha_{i-2}(\sigma_{i-2}(a) f^{(i-2)}(a) - L_{i-1} f(a)) & \textit{for} \;\; i = 2, \ldots, m+1, \\ \beta_{i-m-2}(\sigma_{i-m-2}(b) f^{(i-m-2)}(b) + L_{i-m-1} f(b)) & \textit{for} \;\; i = m+2, \ldots, r, \end{cases}$$

for $f \in Y$,

$$h_i := \begin{cases} h & \textit{for} \;\; i = 1, \\ \alpha_{i-2} \tau_{i-2}(a) & \textit{for} \;\; i = 2, \ldots, m+1, \\ \beta_{i-m-2} \tau_{i-m-2}(b) & \textit{for} \;\; i = m+2, \ldots, r. \end{cases}$$

Using (17) we can easily show that V1, V2, V3 are satisfied.

Definition of the quantities appearing in V4: *First of all we determine positive real numbers* $\gamma_1, \ldots, \gamma_r$ *such that*

$$\tilde{\mathcal{M}}(f,f) \geq \xi \gamma_1 \int_a^b f^2 dx + \sum_{i=0}^{m-1} \hat{\alpha}_i \gamma_{i+2} (f^{(i)}(a))^2 + \sum_{i=0}^{m-1} \hat{\beta}_i \gamma_{i+m+2} (f^{(i)}(b))^2 \tag{18}$$

holds true for all $f \in \tilde{D}$. *The quantities appearing in V4 are defined by*

$$\kappa_i := \begin{cases} \xi & \text{for } i = 1, \\ \hat{\alpha}_{i-2} & \text{for } i = 2, \ldots, m+1, \\ \hat{\beta}_{i-m-2} & \text{for } i = m+2, \ldots, r, \end{cases}$$

and $q_i := \gamma_i$. *Then obviously assumption V4 is satisfied.*

Choice of ξ, $\hat{\alpha}_i$, $\hat{\beta}_i$: *All quantities appearing in V1, V2, V3, V4 are determined by the definitions we have just given; however, the question how the numbers* ξ, $\hat{\alpha}_i$, $\hat{\beta}_i$ *used in the process shall be fixed, is not yet answered. In order to establish a rule for the choice of* ξ, $\hat{\alpha}_i$, $\hat{\beta}_i$ *we consider the following: If we want to determine elements according to Theorem 3, which are admissible for the complementary variational principle (7), we need functions* g *with the property* $g \in Y$, $(1 - \kappa_i)(M_i g - h_i) = 0$ *for* $i = 1, \ldots, r$. *This condition is equivalent to*

$$\begin{aligned} &g \in C^{2m}[a,b], \; L_0 g = h \quad \text{if} \quad \xi = 0 \\ &(1-\alpha_i) g^{(i)}(a) + \alpha_i (\sigma_i(a)\, g^{(i)}(a) - L_{i+1} g(a)) = \alpha_i \tau_i(a) \\ &\qquad\qquad \text{for all } i \in \mathbb{N} \cup \{0\}, i \le m-1 \quad \text{and} \quad \hat{\alpha}_i = 0, \\ &(1-\beta_i) g^{(i)}(b) + \beta_i (\sigma_i(b)\, g^{(i)}(b) + L_{i+1} g(b)) = \beta_i \tau_i(b) \\ &\qquad\qquad \text{for all } i \in \mathbb{N} \cup \{0\}, i \le m-1 \quad \text{and} \quad \hat{\beta}_i = 0. \end{aligned} \tag{19}$$

In order to guarantee that the complementary variational principle (7) can be easily applied, we have to choose $\xi \in \{0,1\}$, $\hat{\alpha}_i \in \{0,1\}$, $\hat{\beta}_i \in \{0,1\}$ *for* $i = 0, \ldots, m-1$ *such that we can find functions* g *satisfying (19) without any trouble. Obviously such a choice of* ξ, $\hat{\alpha}_i$, $\hat{\beta}_i$ *is always possible.*

For boundary value problems with *partial* differential equations, we can determine the quantities appearing in V1, V2, V3, and V4 in a way which is analogous to the way just described for boundary value problems with *ordinary* differential equations; a generalized Green's Formula (see [47, p. 229], [43, p. 31]) replaces (17).

Since we will use the complementary variational principle (7) for error estimation for the solution of the linear equation (5) and for the calculation of bounds to eigenvalues of equation (1), we need not only some elements g admissible for this variational principle, but especially such elements, for which $\mathcal{B}(g,g)$ is a good approximation for $\mathcal{M}(u,u)$ (for notation see Theorem 1). Now we will show, that we actually can obtain such elements according to the construction procedure given in Theorem 3.

Theorem 4

Assume that L1, V1, V2, V3, and V4 are satisfied. Suppose X, $\mathcal{B}$, T *are defined as in Theorem 2. For all* $f \in Y$ *let*

$$\|f\| := (\tilde{\mathcal{M}}(f,f) + \sum_{i=1}^{r} \kappa_i\, q_i^{-1} \langle M_i f, M_i f \rangle_i)^{\frac{1}{2}}.$$

Let $u \in D \cap Y$, $M_i u = h_i$ *for* $i = 1, \ldots, r$. *Then* u *is a solution of equation (5), and we have*

$$-\mathcal{M}(u,u) = \max\{-\mathcal{B}(g,g) \;:\; g \in V_0\}. \tag{20}$$

Furthermore, for all $f \in Y$, which satisfy the conditions $(1-\kappa_i)(M_i f - h_i) = 0$ for $i = 1,\dots,r$, we have the inequality

$$\mathcal{B}\left(\begin{pmatrix} f \\ Gf+z \end{pmatrix}, \begin{pmatrix} f \\ Gf+z \end{pmatrix}\right) - \mathcal{M}(u,u) \le 2\,\|u-f\|^2 .$$

Proof: u solves (5) as a consequence of (12) and (13). Since $V_0 \subset V$ according to Theorem 3 (referring to (8) for V) , we obtain with (7) the inequality $-\mathcal{M}(u,u) \ge \max\{-\mathcal{B}(g,g) : g \in V_0\}$. In view of

$$\begin{pmatrix} u \\ Gu+z \end{pmatrix} \in V_0, \quad \mathcal{B}\left(\begin{pmatrix} u \\ Gu+z \end{pmatrix}, \begin{pmatrix} u \\ Gu+z \end{pmatrix}\right) = \mathcal{M}(u,u) \tag{21}$$

this yields equation (20). – Now we consider an $f \in Y$ such that $(1-\kappa_i)(M_i f - h_i) = 0$ for $i = 1,\dots,r$. Because of (12) we have $\mathcal{B}\left(\begin{pmatrix} u \\ Gu+z \end{pmatrix}, \begin{pmatrix} f-u \\ G(f-u) \end{pmatrix}\right) = 0$, using (21) and defining $d := \begin{pmatrix} T_1(f-u) \\ \vdots \\ T_r(f-u) \end{pmatrix}$, we obtain

$$\begin{aligned}
&\mathcal{B}\left(\begin{pmatrix} f \\ Gf+z \end{pmatrix}, \begin{pmatrix} f \\ Gf+z \end{pmatrix}\right) - \mathcal{M}(u,u) \\
&= \mathcal{B}\left(\begin{pmatrix} f \\ Gf+z \end{pmatrix} - \begin{pmatrix} u \\ Gu+z \end{pmatrix}, \begin{pmatrix} f \\ Gf+z \end{pmatrix} - \begin{pmatrix} u \\ Gu+z \end{pmatrix}\right) \\
&= \mathcal{B}\left(\begin{pmatrix} f-u \\ d \end{pmatrix} - \begin{pmatrix} 0 \\ d-G(f-u) \end{pmatrix}, \begin{pmatrix} f-u \\ d \end{pmatrix} - \begin{pmatrix} 0 \\ d-G(f-u) \end{pmatrix}\right) \\
&\le 2\,\mathcal{B}\left(\begin{pmatrix} f-u \\ d \end{pmatrix}, \begin{pmatrix} f-u \\ d \end{pmatrix}\right) + 2\,\mathcal{B}\left(\begin{pmatrix} 0 \\ d-G(f-u) \end{pmatrix}, \begin{pmatrix} 0 \\ d-G(f-u) \end{pmatrix}\right) \\
&= 2\,\|f-u\|^2
\end{aligned}$$

This completes the proof. □

Most of the well known complementary variational principles (see [10, 30, 43]) can be stated in the form (20), where X, $\mathcal{B}$, and T are defined as in Theorem 2 and V_0 as in Definition 2.

3. The Inclusion of Eigenvalues of Left Definite Problems by means of Complementary Variational Principles

The results on complementary variational principles obtained in the preceding section shall be tailored to finding inclusion intervals for eigenvalues. The basis for this is Theorem 5. In order to formulate it we need in addition to L1, L2, and L3 a further assumption:

L4: There exist a sequence $(\lambda_i)_{i\in\mathbb{N}}$ of eigenvalues of problem (1) and a sequence $(\Phi_i)_{i\in\mathbb{N}}$ of corresponding eigenelements such that

$$\begin{aligned} &\mathcal{M}(f,\Phi_i) = \lambda_i \mathcal{N}(f,\Phi_i) && \text{for all } f\in D,\ i\in\mathbb{N},\\ &\mathcal{M}(\Phi_i,\Phi_k) = \delta_{ik} && \text{for all } i,k\in\mathbb{N},\\ &\mathcal{N}(f,f) = \textstyle\sum_{i=1}^{\infty} \lambda_i |\mathcal{N}(f,\Phi_i)|^2 && \text{for all } f\in D. \end{aligned}$$

An approach to verifying the conditions of L4 will be discussed subsequent to the proof of Theorem 5.

Theorem 5 *(see [20])*
Let assumptions L1, L2, L3, L4 be satisfied. Furthermore, let $n\in\mathbb{N}$, $v_i\in D$, $w_i\in X$ for $i=1,\dots,n$ and $\rho\in\mathbb{R}$, $\rho>0$. Suppose that

$$\mathcal{B}(Tf,w_i) = \mathcal{N}(f,v_i) \quad \text{for all } f\in D,\ i=1,\dots,n. \tag{22}$$

We define matrices A_0, A_1, A_2 by

$$A_0 := (\mathcal{M}(v_i,v_k))_{i,k=1,\dots,n},\ A_1 := (\mathcal{N}(v_i,v_k))_{i,k=1,\dots,n},\ A_2 := (\mathcal{B}(w_i,w_k))_{i,k=1,\dots,n}.$$

Let $A_0-2\rho A_1+\rho^2 A_2$ be positive definite; μ_i denotes for each $i=1,\dots,n$ the i-th smallest eigenvalue of the eigenvalue problem

$$(A_0-\rho A_1)x = \mu(A_0-2\rho A_1+\rho^2 A_2)x.$$

Then the interval $[\rho-\frac{\rho}{1-\mu_j},\rho)$ contains for all $j\in\mathbb{N}$ with $j\le n$ and $\mu_j<0$ at least j eigenvalues of the eigenvalue problem (1).

In order to prove the theorem we need the following lemma:

Lemma 1
Let assumptions L1, L2, L3, and L4 be satisfied. Furthermore, we assume

(i) $\rho\in\mathbb{R}$, $\eta\in\mathbb{R}$, $0<\eta<\rho$, $v\in D$, $w\in X$, $\mathcal{B}(Tv-\rho w,Tv-\eta w)\le 0$;

(ii) $\mathcal{B}(Tf,w)=\mathcal{N}(f,v)$ *for all* $f\in D$;

(iii) $\mathcal{N}(v,\Phi_i)=0$ *for all* $i\in\mathbb{N}$ *with* $\lambda_i\in[\eta,\rho)$.

Then $\mathcal{B}(Tv-\rho w,Tv-\rho w)=0$.

Proof of the lemma: Since $\mathcal{B}(T\Phi_i,T\Phi_k)=\delta_{ik}$ for $i,k\in\mathbb{N}$ due to L3 and L4, the numbers α and β defined by

$$\alpha := \mathcal{B}(Tv,Tv)-\sum_{i=1}^{\infty}|\mathcal{B}(Tv,T\Phi_i)|^2,\ \beta := \mathcal{B}(w,w)-\sum_{i=1}^{\infty}|\mathcal{B}(w,T\Phi_i)|^2$$

are non-negative because of the Bessel inequality. Using L3, L4, and (ii) we obtain

$$\begin{aligned}&\mathcal{B}(T v - \rho w, T v - \eta w)\\ &\quad = \alpha + \textstyle\sum_{i=1}^{\infty} |\mathcal{M}(v,\Phi_i)|^2 - (\rho+\eta)\,\mathcal{N}(v,v) + \rho\eta\,\beta + \rho\eta\,\textstyle\sum_{i=1}^{\infty} |\mathcal{N}(v,\Phi_i)|^2 \\ &\quad = \alpha + \rho\eta\,\beta + \textstyle\sum_{i=1}^{\infty}(\lambda_i-\rho)(\lambda_i-\eta)|\mathcal{N}(v,\Phi_i)|^2\,.\end{aligned} \tag{23}$$

In an analogous way we obtain

$$\mathcal{B}(T v - \rho w, T v - \rho w) = \alpha + \rho^2\beta + \sum_{i=1}^{\infty}(\lambda_i-\rho)^2|\mathcal{N}(v,\Phi_i)|^2\,. \tag{24}$$

Since $(\lambda_i-\rho)(\lambda_i-\eta) \geq 0$ for all $i \in I\!N$ with $\lambda_i \notin [\eta,\rho)$ and using (iii) we get that the numbers α, $\rho\eta\beta$, and $(\lambda_i-\rho)(\lambda_i-\eta)|\mathcal{N}(v,\Phi_i)|^2$ for all $i \in I\!N$ are non-negative. Using (i) and (23), this implies $\alpha = 0$, $\rho\eta\beta = 0$, and $(\lambda_i-\rho)(\lambda_i-\eta)|\mathcal{N}(v,\Phi_i)|^2 = 0$ for all $i \in I\!N$. Combined with (24), this yields the assertion of our lemma. □

Proof of Theorem 5: Suppose there is a $j \in I\!N$ with $j \leq n$ and $\mu_j < 0$ such that the number of eigenvalues of the eigenvalue problem (1) contained in the interval $[\rho - \frac{\rho}{1-\mu_j}, \rho)$ is strictly smaller than j. For $i = 1,\ldots,n$ let $x_i \in I\!K^n$ be an eigenvector of the problem $(A_0 - \rho A_1)\,x = \mu\,(A_0 - 2\rho A_1 + \rho^2 A_2)\,x$ corresponding to the eigenvalue μ_i; we assume $\overline{x}_i^T(A_0 - 2\rho A_1 + \rho^2 A_2)x_k = \delta_{ik}$ for $i,k = 1,\ldots,n$; $(x_i)_k$ denotes the k-th component of x_i. The homogenous linear system

$$\sum_{i=1}^{j}\sum_{k=1}^{n} \mathcal{N}(\Phi_m, v_k)(x_i)_k\,c_i = 0 \quad \text{for all} \quad m \in I\!N\,,\ \lambda_m \in [\eta,\rho)$$

has a non-trivial solution $(c_1,\ldots,c_j) \in I\!K^j$, since the number of equations is smaller than the number of unknowns. We define

$$\gamma_k := \sum_{i=1}^{j}(x_i)_k\,c_i \quad \text{for} \quad k = 1,\ldots,n\,,\ \gamma := (\gamma_1,\ldots,\gamma_n)^T\,,\ v := \sum_{i=1}^{n}\overline{\gamma}_i\,v_i\,,\ w := \sum_{i=1}^{n}\overline{\gamma}_i\,w_i\,.$$

It follows that $\gamma \neq 0$ since $(c_1,\ldots,c_j) \neq (0,\ldots,0)$ and $x_1,\ldots,x_j$ are linearly independent. Using for each $i = 1,\ldots,j$, the inequality $\eta - (\eta-\rho)\mu_i \leq 0$, we obtain after a short computation

$$\rho\mathcal{B}(T v - \rho w, T v - \eta w) = \overline{\gamma}^T(\eta(A_0 - 2\rho A_1 + \rho^2 A_2) - (\rho-\eta)(A_0 - \rho A_1))\gamma \leq 0\,.$$

Hence, ρ, η, v, w satisfy the assumptions of Lemma 1 and $\mathcal{B}(T v - \rho w, T v - \rho w) = \overline{\gamma}^T(A_0 - 2\rho A_1 + \rho^2 A_2)\gamma = 0$. This contradicts the requirement that $A_0 - 2\rho A_1 + \rho^2 A_2$ is positive definite. Thus, the hypothesis beginning our proof is wrong and therefore, Theorem 5 is valid. □

If we define

$$X := D\,,\ \mathcal{B}(f,g) := \mathcal{M}(f,g) \quad \text{for all} \quad f,g \in D\,,\ T f := f \quad \text{for all} \quad f \in D\,,$$

assumption L3 is obviously satisfied and Theorem 5 reduces to Lehmann's theorem [28, 29]. Unfortunately then the trial functions w_i are quite often hard to construct. This reduces the applicability of Lehmann's theorem in the left definite case.

The following lemma makes in many cases the verification of L4 possible, which is required in order to use Theorem 5.

Lemma 2
Let L1 and L2 be satisfied. Assume that $(D, \mathcal{M}(.,.))$ is a Hilbert space, and $\mathcal{N}$ is a compact sesquilinear form on D. If there exists an infinite dimensional subspace D_+ of D such that $\mathcal{N}(f,f) > 0$ for all $f \in D_+$, $f \neq 0$, then L4 is satisfied.

Proof: Since $\mathcal{N}$ is compact, there exists a compact linear operator $S : D \to D$ with the property $\mathcal{N}(f,g) = \mathcal{M}(Sf,g)$ for all $f,g \in D$. Obviously S is symmetric. According to the spectral theorem for compact symmetric operators we have an index set $J \subset \mathbb{N}$ and families $(\hat{\lambda}_i)_{i\in J}$, $(\hat{\Phi}_i)_{i\in J}$ such that

$$\hat{\lambda}_i \in \mathbb{R},\ \hat{\lambda}_i \neq 0,\ \hat{\Phi}_i \in D,\ S\hat{\Phi}_i = \hat{\lambda}_i\hat{\Phi}_i \ \text{ for } \ i \in J,$$
$$\mathcal{M}(\hat{\Phi}_i, \hat{\Phi}_k) = \delta_{ik} \ \text{ for } \ i,k \in J,$$
$$Sf = \textstyle\sum_{i\in J} \hat{\lambda}_i\, \mathcal{M}(f,\hat{\Phi}_i)\hat{\Phi}_i \ \text{ for all } \ f \in D.$$

This immediately yields

$$\mathcal{M}(f,\hat{\Phi}_i) = \hat{\lambda}_i^{-1}\mathcal{N}(f,\hat{\Phi}_i) \ \text{ for all } \ f \in D,\ i \in J,$$
$$\mathcal{N}(f,f) = \textstyle\sum_{i\in J} \hat{\lambda}_i^{-1}|\mathcal{N}(f,\hat{\Phi}_i)|^2 \ \text{ for all } \ f \in D.$$

Were J a finite set, there would exist an $f \in D_+$, $f \neq 0$ and $N(f,\hat{\phi}_i) = 0$ for all $i \in J$; this would imply $\mathcal{N}(f,f) = 0$ contrary to the prerequisite on D_+. Hence, there is a one to one mapping $\Psi : \mathbb{N} \to J$. If we define $\lambda_i := \hat{\lambda}^{-1}_{\Psi(i)}$, $\Phi_i := \hat{\Phi}_{\Psi(i)}$ for $i \in \mathbb{N}$, the sequences $(\lambda_i)_{i\in\mathbb{N}}$, $(\Phi_i)_{i\in\mathbb{N}}$ have the properties required in L4. □

Theorem 5 connects inclusion intervals for eigenvalues to the quantities X, $\mathcal{B}$, T used for establishing complementary variational principles. Now we will state the basic idea of Theorem 5 in a form which while looking a bit more complicated, is more favorable for numerical application. We need the following assumptions and notations:

L5: $m,n \in \mathbb{N}$, $v_i \in D$, $w_i^* \in X$ for $i = 1,\ldots,n$, and $w_i^0 \in X$ for $i = 1,\ldots,m$. We require that $\mathcal{B}(Tf,w_i^*) = \mathcal{N}(f,v_i)$ for all $f \in D$, $i = 1,\ldots,n$ and $\mathcal{B}(Tf,w_i^0) = 0$ for all $f \in D$, $i = 1,\ldots,m$ holds true.

Definition 3
If assumptions L1, L2, L3, and L5 are satisfied, we define matrices A_0, A_1, A_2, B, F by

$$A_0 := (\mathcal{M}(v_i,v_k))_{i,k=1,\ldots,n}, \qquad A_1 := (\mathcal{N}(v_i,v_k))_{i,k=1,\ldots,n},$$
$$A_2 := (\mathcal{B}(w_i^*,w_k^*))_{i,k=1,\ldots,n},$$
$$B := (\mathcal{B}(w_i^0,w_k^0))_{i,k=1,\ldots,m}, \qquad F := (\mathcal{B}(w_i^*,w_k^0))_{i=1,\ldots,n;\ k=1,\ldots,m}.$$

We obtain

Theorem 6
Let assumptions L1, L2, L3, L4, and L5 be satisfied. Let $\rho \in \mathbb{R}$, $\rho > 0$. Suppose $C \in \mathbb{K}^{n\times m}$ such that the matrix $A_0 - 2\rho A_1 + \rho^2(A_2 + C\overline{F}^T + F\overline{C}^T + C\,B\overline{C}^T)$ is positive

definite. Let μ_i denote for $i = 1, \ldots, n$ the i-th smallest eigenvalue of the eigenvalue problem

$$(A_0 - \rho A_1)\, x = \mu\, (A_0 - 2\rho A_1 + \rho^2 (A_2 + C\overline{F}^T + F\overline{C}^T + C B \overline{C}^T))\, x\,.$$

Then the interval $[\rho - \dfrac{\rho}{1-\mu_j}, \rho)$ contains for all $j \in \mathbb{N}$ with $j \leq n$ and $\mu_j < 0$ at least j eigenvalues of the eigenvalue problem (1).

Proof: Let $C = (c_{ik})_{i=1,\ldots,n;\ k=1,\ldots,m}$; if we define

$$w_i := w_i^* + \sum_{k=1}^{m} c_{ik}\, w_k^0 \quad \text{for} \quad i = 1, \ldots, n\,, \tag{25}$$

(22) is satisfied, and we obtain

$$(\mathcal{B}(w_i, w_k))_{i,k=1,\ldots,n} = A_2 + C\overline{F}^T + F\overline{C}^T + C B \overline{C}^T\,.$$

The assertion is now a consequence of Theorem 5. □

The following theorem shows that a good approximation for $-F B^{-1}$ is a reasonable choice for C in Theorem 6, provided B is regular.

Theorem 7

Let the assumptions of Theorem 6 be satisfied. Let B be regular and $A_0 - 2\rho A_1 + \rho^2 (A_2 - F B^{-1} \overline{F}^T)$ be positive definite. Let μ_i for each $i = 1, \ldots, n$ be the i-th smallest eigenvalue of the eigenvalue problem

$$(A_0 - \rho A_1)\, x = \mu\, (A_0 - 2\rho A_1 + \rho^2 (A_2 + C\overline{F}^T + F\overline{C}^T + C B \overline{C}^T))\, x$$

and let $\tilde{\mu}_i$ be the i-th smallest eigenvalue of the eigenvalue problem

$$(A_0 - \rho A_1)\, x = \mu\, (A_0 - 2\rho A_1 + \rho^2 (A_2 - F B^{-1} \overline{F}^T))\, x.$$

Then we have for all $j \in \mathbb{N}$, $j \leq n$ and $\tilde{\mu}_j < 0$ the inequality

$$\rho - \frac{\rho}{1-\mu_j} \leq \rho - \frac{\rho}{1-\tilde{\mu}_j}\,.$$

Proof: For all $x \in \mathbb{K}^n$ we have

$$\begin{aligned} &\overline{x}^T (A_2 + C\overline{F}^T + F\overline{C}^T + C B \overline{C}^T)\, x \\ &\quad = \overline{x}^T (A_2 - F B^{-1} \overline{F}^T + (C + F B^{-1}) B (\overline{C}^T + B^{-1} \overline{F}^T))\, x \\ &\quad \geq \overline{x}^T (A_2 - F B^{-1} \overline{F}^T)\, x\,. \end{aligned}$$

According to the comparison theorem, we obtain for all $j \in \mathbb{N}$ with $j \leq n$ and $\mu_j < 0$, the inequality $\tilde{\mu}_j \leq \mu_j$. From this the assertion follows immediately. □

We will give a procedure for the construction of the elements w_i^* and w_i^0 occuring in assumption L5; for that we need in addition to L1, L2, V1, V2, and V4 the following assumptions:

V5: $N_i : D \to H_i$ are linear operators for $i = 1, \ldots, r$ such that

$$\mathcal{N}(f,g) = \sum_{i=1}^{r} \langle T_i f, N_i g \rangle_i \quad \text{for} \quad f,g \in D\,.$$

V6: $m,n \in \mathbb{N}$, $v_i \in D$, $\hat{w}_i^* \in Y$ for $i = 1,\ldots,n$, $\hat{w}_1^0,\ldots,\hat{w}_m^0$ are linearly independent elements from Y such that $(1-\kappa_k)(M_k\hat{w}_i^* - N_k v_i) = 0$ for $i = 1,\ldots,n$, $k = 1,\ldots,r$ and $(1-\kappa_k)M_k\hat{w}_i^0 = 0$ for $i = 1,\ldots,m$, $k = 1,\ldots,r$, and

$$\{f \in Y \;:\; \tilde{\mathcal{M}}(f,f) = \sum_{i=1}^{r} \kappa_i q_i \langle T_i f, T_i f \rangle_i\} \bigcap \operatorname{span}\{\hat{w}_1^0,\ldots,\hat{w}_m^0\} = \emptyset\,.$$

Since the requirements on w_i^* and w_i^0 in L5 are analogous to those satisfied by the elements (see (8)) admissible for the complementary variational principle (7), we can use Theorem 3 for the construction of w_i^* and w_i^0. We obtain

Theorem 8
Let the assumptions L1, L2, V1, V2, V4, V5, V6 be satisfied; X, $\mathcal{B}$, T are defined as in Theorem 2. We define

$$z_i := \begin{pmatrix} q_1^{-1} N_1 v_i \\ \vdots \\ q_r^{-1} N_r v_i \end{pmatrix}, \; w_i^* := \begin{pmatrix} \hat{w}_i^* \\ G\hat{w}_i^* + z_i \end{pmatrix} \text{ for } i = 1,\ldots,n\,, \; w_i^0 := \begin{pmatrix} \hat{w}_i^0 \\ G\hat{w}_i^0 \end{pmatrix} \text{ for } i = 1,\ldots,m\,,$$

then L5 is satisfied and B is regular.

Proof: First of all we will show that w_i^* has the properties required in L5. Let $j \in \mathbb{N}$, $j \le n$ be arbitrarily chosen. If we define $F(f) := \mathcal{N}(f, v_j)$ for all $f \in D$, $h_k := N_k v_j$ for $k = 1,\ldots,r$, then since V3 is satisfied, according to Theorem 3 we obtain $\mathcal{B}(Tf, w_j^*) = \mathcal{N}(f, v_j)$ for all $f \in D$.

We have not yet proved that w_i^0 satisfies the requirements in L5. Let $j \in \mathbb{N}$ be arbitrarily chosen. If we define $F(f) = 0$ for all $f \in D$, $h_k = 0$ for $k = 1,\ldots,r$, since V3 is satisfied, according to Theorem 3, we obtain $\mathcal{B}(Tf, w_j^0) = 0$ for all $f \in D$. – Let $m \in \mathbb{N}$, $c_1,\ldots,c_m \in \mathbb{K}$, $(c_1,\ldots,c_m) \neq (0,\ldots,0)$. Then we have

$\sum_{k=1}^m c_k \hat{w}_k^0 \in Y$, $\sum_{k=1}^m c_k \hat{w}_k^0 \neq 0$, and

$\tilde{\mathcal{M}}(\sum_{k=1}^m c_k \hat{w}_k^0, \sum_{k=1}^m c_k \hat{w}_k^0) > \sum_{i=1}^r \kappa_i q_i \langle T_i \sum_{k=1}^m c_k \hat{w}_k^0, T_i \sum_{k=1}^m c_k \hat{w}_k^0 \rangle_i$.

From this we obtain

$$0 < \mathcal{B}(\sum_{k=1}^m c_k w_k^0, \sum_{k=1}^m c_k w_k^0) = \sum_{i,k=1}^m c_i \mathcal{B}(w_i^0, w_k^0)\bar{c}_k;$$

hence, the matrix $(\mathcal{B}(w_i^0, w_k^0))_{i,k=1,\ldots,m}$ is regular. This completes the proof. □

These results – especially Theorem 6 – are the basis for the following procedure that calculates inclusion intervals for eigenvalues of problem (1):

Step 1: Verify that the given eigenvalue problem satisfies assumptions L1, L2, L4.

Step 2: Construct quantities X, $\mathcal{B}$, T with the properties L3.

Step 3: Determine v_i, w_i^*, w_i^0 such that L5 is satisfied.

Step 4: Choose $\rho \in \mathbb{R}$, $\rho > 0$ and $C \in \mathbb{K}^{n\times m}$ as a good approximation to $-F B^{-1}$. Then calculate the matrices $A_0 - \rho A_1$, $A_0 - 2\rho A_1 + \rho^2(A_2 + C\overline{F}^T + F\overline{C}^T + C B\overline{C}^T)$ according to Definition 3.

Step 5: Verify (numerically, see below) that $A_0 - 2\rho A_1 + \rho^2(A_2 + C\overline{F}^T + F\overline{C}^T + C B\overline{C}^T)$ is positive definite and compute (with guaranteed bounds, see below) the negative eigenvalues of the eigenvalue problem

$$(A_0 - \rho A_1)\, x = \mu\,(A_0 - 2\rho A_1 + \rho^2(A_2 + C\overline{F}^T + F\overline{C}^T + C B\overline{C}^T))\, x\,.$$

The corresponding inclusion intervals are obtained from Theorem 6.

Frequently it is advisable to perform steps 2 and 3 in the following special way ("Procedure of special choice for X, $\mathcal{B}$, T"):

Step 2: Obtain quantities $\tilde{D}$, $\tilde{\mathcal{M}}$, Y, r, H_i, T_i, M_i, q_i, κ_i, N_i with the properties required in V1, V2, V4, V5 and define X, $\mathcal{B}$, T as in Theorem 2.

Step 3: Determine v_i, $\hat{w}_i^*$, $\hat{w}_i^0$ such that V6 is satisfied and define w_i^* and w_i^0 as in Theorem 8.

4. Variational Characterization of Eigenvalues and the Rayleigh–Ritz Procedure

The methods we described in section 3 make it possible to determine intervals which contain a certain number of eigenvalues of the eigenvalue problem under consideration. With many engineering applications, there arise left definite eigenvalue problems, for which the positive eigenvalues can be indexed consecutively, starting with the smallest one. In practice it is frequently important to determine accurate upper and lower bounds for some positive eigenvalues according to their index. For the calculation of accurate upper bounds the procedure of Rayleigh and Ritz has proved to be an excellent approach; in general, sharp lower bounds can be obtained by a combination of the inclusion theorems from section 3 with the comparison theorem.

Below we will prove a well known variational characterization for the positive eigenvalues; from this characterization the comparison theorem and the procedure of Rayleigh and Ritz can be derived. First of all we formulate an assumption, which guarantees that the eigenvalues can be numbered consecutively.

L4* : There exist sequences $(\lambda_i^+)_{i\in J^+}$, $(\lambda_i^-)_{i\in J^-}$ of eigenvalues of problem (1) and sequences $(\Phi_i^+)_{i\in J^+}$, $(\Phi_i^-)_{i\in J^-}$ of corresponding eigenelements such that

$$
\begin{aligned}
&\mathcal{M}(f,\Phi_i^+) = \lambda_i^+ \mathcal{N}(f,\Phi_i^+) && \text{for all } f\in D,\ i\in J^+,\\
&\mathcal{M}(f,\Phi_i^-) = \lambda_i^- \mathcal{N}(f,\Phi_i^-) && \text{for all } f\in D,\ i\in J^-,\\
&\mathcal{M}(\Phi_i^+,\Phi_k^+) = \delta_{ik} && \text{for all } i,k\in J^+,\\
&\mathcal{M}(\Phi_i^-,\Phi_k^-) = \delta_{ik} && \text{for all } i,k\in J^-,\\
&\mathcal{M}(\Phi_i^+,\Phi_k^-) = 0 && \text{for all } i\in J^+,\ k\in J^-,\\
&\mathcal{N}(f,f) = \textstyle\sum_{i\in J^+}\lambda_i^+ |\mathcal{N}(f,\Phi_i^+)|^2 && \\
&\qquad\quad + \textstyle\sum_{i\in J^-}\lambda_i^- |\mathcal{N}(f,\Phi_i^-)|^2 && \text{for all } f\in D,\\
&0<\lambda_i^+\le\lambda_k^+ && \text{for all } i,k\in J^+,\ i\le k,\\
&0>\lambda_i^-\ge\lambda_k^- && \text{for all } i,k\in J^-,\ i\le k,\\
&\text{where } J^+=\{i\in\mathbb{N}\ :\ i\le n^+\},\ J^-=\{i\in\mathbb{N}\ :\ i\le n^-\},\ n^+,n^-\in\mathbb{N}_0\cup\{\infty\}.
\end{aligned}
$$

Remark 1 *If L4* is satisfied, then we call λ_i^+ the i-th smallest positive eigenvalue of the eigenvalue problem*

"$\mathcal{M}(f,\Phi)=\lambda\mathcal{N}(f,\Phi)$ for all $f\in D$" $(i\in J^+)$.

Now we can formulate

Theorem 9 (Variational characterization of eigenvalues)
Let L1, L2, and L4 be satisfied. Then*

$$
\lambda_j^+ = \inf_{\substack{U \text{ subspace of } D\\ \dim U = j\\ \mathcal{N}(f,f)>0\\ \text{for all } f\in U,\ f\neq 0}} \ \sup_{\substack{f\in U\\ f\neq 0}} \frac{\mathcal{M}(f,f)}{\mathcal{N}(f,f)} \qquad \text{for } j\in J^+.
$$

Proof: We give a proof in three steps.
Step 1: For $f\in D$ with $\mathcal{N}(f,f)>0$ and $\mathcal{M}(f,\Phi_k^+)=0$ for $k\in J^+$, $k<j$ we have

$$
\begin{aligned}
\mathcal{M}(f,f) &\ge \sum_{k=j}^{n^+} |\mathcal{M}(f,\Phi_k^+)|^2 + \sum_{k=1}^{n^-} |\mathcal{M}(f,\Phi_k^-)|^2\\
&= \sum_{k=j}^{n^+} (\lambda_k^+)^2 |\mathcal{N}(f,\Phi_k^+)|^2 + \sum_{k=1}^{n^-} (\lambda_k^-)^2 |\mathcal{N}(f,\Phi_k^-)|^2\\
&\ge \sum_{k=j}^{n^+} \lambda_k^+\lambda_j^+ |\mathcal{N}(f,\Phi_k^+)|^2\\
&\ge \lambda_j^+ \mathcal{N}(f,f),
\end{aligned}
$$

hence,

$$\frac{\mathcal{M}(f,f)}{\mathcal{N}(f,f)} \geq \lambda_j^+ .$$

Step 2: Let $U \subset D$ be such that $\dim U = j$ and $\mathcal{N}(f,f) > 0$ for all $f \in U$ with $f \neq 0$. Then there exists $\hat{f} \in U$, $\hat{f} \neq 0$ such that $\mathcal{M}(\hat{f}, \Phi_k^+) = 0$ for all $k \in J^+$, $k < j$. Using step 1 we obtain

$$\sup_{\substack{f \in U \\ f \neq 0}} \frac{\mathcal{M}(f,f)}{\mathcal{N}(f,f)} \geq \frac{\mathcal{M}(\hat{f},\hat{f})}{\mathcal{N}(\hat{f},\hat{f})} \geq \lambda_j^+ .$$

Step 3: Let $\tilde{U}$ be the span of $\{\Phi_1^+, \ldots, \Phi_j^+\}$. For all $f \in \tilde{U}$, $f \neq 0$, we have $f = \sum_{k=1}^{j} \gamma_k \Phi_k^+$ with $\gamma_k \in \mathbb{K}$ for $k = 1, \ldots, j$. From this we obtain

$$\mathcal{M}(f,f) = \sum_{k=1}^{j} \gamma_k^2$$

$$\mathcal{N}(f,f) = \sum_{k=1}^{j} \lambda_k^+ |\mathcal{N}(f, \Phi_k^+)|^2 = \sum_{k=1}^{j} \frac{1}{\lambda_k^+} \gamma_k^2 \geq \frac{1}{\lambda_j^+} \sum_{k=1}^{j} \gamma_k^2$$

and therefore

$$\frac{\mathcal{M}(f,f)}{\mathcal{N}(f,f)} \leq \lambda_j^+$$

and

$$\sup_{\substack{f \in \tilde{U} \\ f \neq 0}} \frac{\mathcal{M}(f,f)}{\mathcal{N}(f,f)} \leq \lambda_j^+ .$$

From step 2 and step 3 we easily obtain the assertion. □

As an obvious consequence, we have

Theorem 10 (Comparison Theorem)
Let L1, L2, and L4 be satisfied, and furthermore let L1, L2, and L4* be satisfied with* $D, \mathcal{M}, \mathcal{N}$ *replaced by* $\tilde{D}, \tilde{\mathcal{M}}, \tilde{\mathcal{N}}$. *Let* λ_i^+ *and* $\tilde{\lambda}_i^+$ *be the i-th smallest eigenvalue of the problem "*$\mathcal{M}(f,\Phi) = \lambda \mathcal{N}(f,\Phi)$ *for all* $f \in D$*" and "*$\tilde{\mathcal{M}}(f,\Phi) = \lambda \tilde{\mathcal{N}}(f,\Phi)$ *for all* $f \in \tilde{D}$*" respectively. If* $D \subset \tilde{D}$, $\tilde{\mathcal{N}}(f,f) \leq \mathcal{N}(f,f)$ *and* $\frac{\tilde{\mathcal{M}}(f,f)}{\tilde{\mathcal{N}}(f,f)} \leq \frac{\mathcal{M}(f,f)}{\mathcal{N}(f,f)}$ *holds for all* $f \in D$ *with* $\mathcal{N}(f,f) > 0$, *then*

$$\tilde{\lambda}_i^+ \leq \lambda_i^+$$

for all i for which λ_i^+ *is well defined.*

Another consequence of the variational characterization (Theorem 9) is the following theorem, which forms the basis for the Rayleigh–Ritz procedure.

Theorem 11 (Rayleigh–Ritz)
Let L1, L2, and L4 be satisfied, and let* $v_1, \ldots, v_n \in D$ *be linearly independent* ($n \in \mathbb{N}$). *We define matrices* A_0, A_1 *by*

$$A_0 := (\mathcal{M}(v_i, v_k))_{i,k=1,\ldots,n}\,, \quad A_1 := (\mathcal{N}(v_i, v_k))_{i,k=1,\ldots,n}\,.$$

Let N^+ *be the number of positive eigenvalues of the problem*

$$A_0\, x = \Lambda\, A_1\, x$$

and denote by Λ_i^+ *the i-th smallest positive eigenvalue of this problem. Then* $N^+ \leq n^+$, *and* $\lambda_i^+ \leq \Lambda_i^+$ *for* $i = 1, \ldots, N^+$ *(for* n^+ *and* λ_i^+ *see L4*).*

Proof: Let $x_k \in \mathbb{K}^n$, $A_0\, x_k = \Lambda_k^+\, A_1\, x_k$, $\overline{x}_j^T A_0\, x_k = \delta_{jk}$, $x_k = (x_{k,1}, \ldots, x_{k,n})^T$ for $j, k = 1, \ldots, N^+$, and $h_j = \sum_{k=1}^n \overline{x}_{j,k} v_k$ for $j = 1, \ldots, N^+$. Then we have

$$\begin{aligned}
\mathcal{M}(h_j, h_k) &= \sum_{l,m=1}^{n} \overline{x}_{j,l} \mathcal{M}(v_l, v_m) x_{k,m} = \overline{x}_j^T A_0\, x_k = \delta_{jk}\,, \\
\mathcal{N}(h_j, h_k) &= \overline{x}_j^T A_1\, x_k = \frac{1}{\Lambda_j^+} \delta_{jk}\,.
\end{aligned}$$

Let U be the span of $\{h_1, \ldots, h_i\}$, then $\dim U = i$. Now, for all $f \in U$, $f \neq 0$, $f = \sum_{k=1}^i \gamma_k h_k$, we have

$$\begin{aligned}
\mathcal{M}(f, f) &= \sum_{k=1}^{i} \gamma_k^2\,, \\
\mathcal{N}(f, f) &= \sum_{k=1}^{i} \gamma_k^2 \frac{1}{\Lambda_k^+} > 0\,, \\
\mathcal{N}(f, f) &\geq \frac{1}{\Lambda_i^+} \sum_{k=1}^{i} \gamma_k^2 = \frac{1}{\Lambda_i^+} \mathcal{M}(f, f)\,.
\end{aligned}$$

This implies

$$\frac{\mathcal{M}(f, f)}{\mathcal{N}(f, f)} \leq \Lambda_i^+$$

and

$$\sup_{\substack{f \in U \\ f \neq 0}} \frac{\mathcal{M}(f, f)}{\mathcal{N}(f, f)} \leq \Lambda_i^+\,.$$

Suppose $i > n^+$. Then there exists $\hat{f} \in U$ such that $\mathcal{M}(\hat{f}, \Phi_k^+) = 0$ for $k = 1, \ldots, n^+$, $\hat{f} \neq 0$. This implies $\mathcal{N}(\hat{f}, \Phi_k^+) = 0$ for $k \in J^+$ and thus by L4* $\mathcal{N}(\hat{f}, \hat{f}) \leq 0$, which is a contradiction to $\mathcal{N}(f, f) > 0$ for all $f \in U$, $f \neq 0$. Hence, $i \leq n^+$. Now Theorem 9 gives the assertion. □

The theorems of this section serve as a basis for the following procedure to compute bounds for eigenvalues:

Step 1: Compute upper bounds Λ_i^+ for λ_i^+ using the Rayleigh–Ritz procedure.

Step 2: Determine (rough) lower bounds ρ_i for λ_i^+ by means of the Comparison Theorem.

Step 3: Choose an $r \in \mathbb{N}$ such that $\Lambda_r^+ \leq \rho_{r+1}$.

Step 4: Compute an interval $[\tau, \rho_{r+1})$, which contains at least j eigenvalues (Theorem 6).

Result: Bounds for λ_{r-j+1}^+:

$$\tau \leq \lambda_{r-j+1}^+ \leq \Lambda_{r-j+1}^+$$

If all quantities are chosen appropriately, we obtain very accurate bounds for λ_{r-j+1}^+. We discuss the construction of comparison problems in the first example in section 7.

5. The Inclusion of Eigenvalues of Right Definite Problems by means of Complementary Variational Principles

In this section, we give an inclusion theorem for problems associated with a positive definite sesquilinear form on the right hand side (cf. [19]). We will be brief since the main result is very similar to Theorem 5. In particular the construction principle for the complementary variational principle can be applied analogously. We first need some assumptions.

R1: D_N is a vector space over $\mathbb{K}$; $\mathcal{N}$ is a Hermitian sesquilinear form on D_N. $\mathcal{N}(f,f) > 0$ holds for all $f \in D_N$ with $f \neq 0$.

R2: D_M is a subspace of D_N. $\mathcal{M}$ is a Hermitian sesquilinear form on D_M.

The eigenvalue problem reads as follows:

"Find pairs $(\lambda, \Phi) \in \mathbb{K} \times D_M$ such that $\Phi \neq 0$ and
$\mathcal{M}(f,\Phi) = \lambda \mathcal{N}(f,\Phi)$ for all $f \in D_M$". (26)

Further assumptions are

R3: X is a vector space over $\mathbb{K}$; $\mathcal{B}$ is a Hermitian sesquilinear form on X; $T : D_M \to X$ is a linear operator. $\mathcal{B}(f,f) \geq 0$ for all $f \in X$ and $\mathcal{N}(f,g) = \mathcal{B}(Tf, Tg)$ for all $f, g \in D_M$.

R4: There exist a sequence $(\lambda_i)_{i\in J}$ of eigenvalues of problem (26) and a sequence $(\Phi_i)_{i\in J}$ of corresponding eigenelements[1] such that

$$J = \{i \in \mathbb{N} \ : \ i \leq \dim D_M\}$$
$$\mathcal{M}(f,\Phi_i) = \lambda_i \mathcal{N}(f,\Phi_i) \qquad \text{for all } f \in D_M,\ i \in J,$$
$$\mathcal{N}(\Phi_i,\Phi_k) = \delta_{ik} \qquad \text{for all } i,k \in J,$$
$$\mathcal{N}(f,f) = \textstyle\sum_{i\in J} |\mathcal{N}(f,\Phi_i)|^2 \qquad \text{for all } f \in D_M,$$
for $\rho \in \mathbb{R}$ the interval $(-\infty, \rho)$ contains no limit point of $(\lambda_i)_{i\in J}$.

[1] We use in our notation an index set J, since will we want to apply the results both to finite dimensional and to infinite dimensional problems.

With these assumptions we can formulate

Theorem 12
Let assumptions R1, R2, R3, and R4 hold. Furthermore, let $\rho \in \mathbb{R}$, and for an $n \in \mathbb{N}$, let $v_i \in D_M$ and $w_i \in X$ for each $i = 1, \ldots, n$ and besides, assume

$$\mathcal{B}(T f, w_i) = \mathcal{M}(f, v_i) \quad \text{for all} \quad f \in D_M \quad \text{and each} \quad i = 1, \ldots, n. \tag{27}$$

We define matrices A_0, A_1, A_2, A, B by

$$A_0 := (\mathcal{N}(v_i, v_k))_{i,k=1,\ldots,n}, \quad A_1 := (\mathcal{M}(v_i, v_k))_{i,k=1,\ldots,n}, \quad A_2 := (\mathcal{B}(w_i, w_k))_{i,k=1,\ldots,n},$$

$$A := A_1 - \rho A_0, \quad B := A_2 - 2\rho A_1 + \rho^2 A_0.$$

Assume that B is positive definite and let μ_i denote for $i = 1, \ldots, n$ the i-th smallest eigenvalue of the eigenvalue problem

$$A x = \mu B x.$$

Then the interval $[\rho + \frac{1}{\mu_j}, \rho)$ contains for all $j \in J$ with $j \leq n$ and $\mu_j < 0$ at least j eigenvalues of eigenvalue problem (26).

We provide two lemmas before proving Theorem 12.

Lemma 3
Let assumptions R1, R2, R3, and R4 be satisfied. Suppose that $\tilde{\eta} \in \mathbb{R}$, $\tilde{\rho} \in \mathbb{R}$ with $\tilde{\eta} \leq \tilde{\rho}$. If $v \in D_M$ and $w \in X$ are selected so that:

(i) $\mathcal{B}(w - \tilde{\rho} T v, w - \tilde{\eta} T v) \leq 0$;

(ii) $\mathcal{B}(T f, w) = \mathcal{M}(f, v)$ for all $f \in D_M$; and

(iii) $\mathcal{N}(v, \Phi_i) = 0$ for each $i \in J$ with $\tilde{\eta} \leq \lambda_i \leq \tilde{\rho}$.

Then $v = 0$.

Proof: Since $\mathcal{B}(T\Phi_j, T\Phi_k) = \delta_{jk}$ for $j, k \in J$, it follows from Bessel's inequality that

$$\mathcal{B}\left(w - \frac{1}{2}(\tilde{\rho} + \tilde{\eta}) T v, w - \frac{1}{2}(\tilde{\rho} + \tilde{\eta}) T v\right) \geq \sum_{i \in J} \left|\mathcal{B}\left(w - \frac{1}{2}(\tilde{\rho} + \tilde{\eta}) T v, T\Phi_i\right)\right|^2.$$

Hence, by R4,

$$\begin{aligned} 0 &\geq \mathcal{B}(w - \tilde{\rho} T v, w - \tilde{\eta} T v) \\ &= \mathcal{B}\left(w - \frac{1}{2}(\tilde{\rho} + \tilde{\eta}) T v, w - \frac{1}{2}(\tilde{\rho} + \tilde{\eta}) T v\right) - \frac{1}{4}(\tilde{\rho} - \tilde{\eta})^2 \mathcal{B}(T v, T v) \\ &\geq \sum_{i \in J} \left|\mathcal{B}\left(w - \frac{1}{2}(\tilde{\rho} + \tilde{\eta}) T v, T\Phi_i\right)\right|^2 - \frac{1}{4}(\tilde{\rho} - \tilde{\eta})^2 \sum_{i \in J} |\mathcal{N}(v, \Phi_i)|^2. \end{aligned}$$

Since $\mathcal{B}(T\Phi_i, w) = \mathcal{M}$

$$\begin{aligned} 0 &\geq \sum_{i \in J} \left(\lambda_i - \frac{1}{2}(\tilde{\rho} + \right. \\ &= \sum_{i \in J} (\lambda_i - \tilde{\rho})(\lambda_i - \end{aligned}$$

Because of (iii), $(\lambda_i -$ for all $i \in J$. From thi

Lemma 4
Let assumptions R1, *with $\eta < \rho$. Suppose t*

(i) $\mathcal{B}(w - \rho T v, w -$

(ii) $\mathcal{B}(T f, w) = \mathcal{M}($

(iii) $\mathcal{N}(v, \Phi_i) = 0$ fo

Then $\mathcal{B}(w - \rho T v, w -$

Proof (indirect proof).

$$\begin{aligned} 0 &> \mathcal{B}(w - \rho T v, w \\ &= (\rho - \eta)(\mathcal{B}(w, T \end{aligned}$$

which implies $0 > \mathcal{B}(v$

$$h(\xi) := \frac{\mathcal{B}(w, w) - \xi}{\mathcal{B}(w, T v) - \xi}$$

Then $h(\rho) = \frac{\mathcal{B}(w -}{\mathcal{B}(w, T v}$

Because of R4, there Hence, there exists a $h(\tilde{\rho}) > \gamma$. Now let $\tilde{\eta} :=$

$$\tilde{\eta} - \tilde{\rho} = \frac{\mathcal{B}(w - \tilde{\rho} T v,}{\mathcal{B}(w, T v) - \tilde{\rho}}$$

the inequality $\gamma < \tilde{\eta}$ with $\tilde{\eta} \leq \lambda_i \leq \tilde{\rho}$. By $0 > \mathcal{B}(w, T v) - \rho \mathcal{B}(T$

Proof of Theorem 12 $\mu_l < 0$, the interval $[\rho$

Since $A = \overline{A}^T$, $B = \overline{B}$ with $x_k = (x_{k,1}, \ldots, x_k$

Since $\mathcal{B}(T\Phi_i, w) = \mathcal{M}(\Phi_i, v) = \lambda_i \mathcal{N}(\Phi_i, v)$ for all $i \in J$,

$$\begin{aligned} 0 &\geq \sum_{i\in J} \left(\lambda_i - \frac{1}{2}(\tilde{\rho}+\tilde{\eta})\right)^2 |\mathcal{N}(v,\Phi_i)|^2 - \frac{1}{4}(\tilde{\rho}-\tilde{\eta})^2 \sum_{i\in J} |\mathcal{N}(v,\Phi_i)|^2 \\ &= \sum_{i\in J} (\lambda_i - \tilde{\rho})(\lambda_i - \tilde{\eta}) |\mathcal{N}(v,\Phi_i)|^2 . \end{aligned}$$

Because of (iii), $(\lambda_i - \tilde{\rho})(\lambda_i - \tilde{\eta})|\mathcal{N}(v,\Phi_i)|^2 \geq 0$ for all $i \in J$, which implies $\mathcal{N}(v,\Phi_i) = 0$ for all $i \in J$. From this, it follows by R4 that $\mathcal{N}(v,v) = 0$, and hence $v = 0$. □

Lemma 4

Let assumptions R1, R2, R3, and R4 of Theorem 12 be satisfied. Let $\eta \in \mathbb{R}$ and $\rho \in \mathbb{R}$ with $\eta < \rho$. Suppose that $v \in D_M$ and $w \in X$ are selected so that:

(i) $\mathcal{B}(w - \rho T v, w - \eta T v) \leq 0$;

(ii) $\mathcal{B}(T f, w) = \mathcal{M}(f, v)$ *for all* $f \in D_M$;

(iii) $\mathcal{N}(v, \Phi_i) = 0$ *for each* $i \in J$ *with* $\eta \leq \lambda_i < \rho$.

Then $\mathcal{B}(w - \rho T v, w - \rho T v) = 0$.

Proof (indirect proof): Suppose that $\mathcal{B}(w - \rho T v, w - \rho T v) > 0$. Then

$$\begin{aligned} 0 &> \mathcal{B}(w - \rho T v, w - \eta T v) - \mathcal{B}(w - \rho T v, w - \rho T v) \\ &= (\rho - \eta)(\mathcal{B}(w, T v) - \rho \mathcal{B}(T v, T v)), \end{aligned}$$

which implies $0 > \mathcal{B}(w, T v) - \rho \mathcal{B}(T v, T v)$. Define

$$h(\xi) := \frac{\mathcal{B}(w,w) - \xi \mathcal{B}(w, T v)}{\mathcal{B}(w, T v) - \xi \mathcal{B}(T v, T v)} \quad \text{for all } \xi \in \mathbb{R} \text{ with } \mathcal{B}(w, T v) \neq \xi \mathcal{B}(T v, T v).$$

Then $h(\rho) = \dfrac{\mathcal{B}(w - \rho T v, w - \eta T v)}{\mathcal{B}(w, T v) - \rho \mathcal{B}(T v, T v)} + \eta \geq \eta$.

Because of R4, there exists some $\gamma \in \mathbb{R}$ with $\gamma < \eta$ such that $\lambda_i \notin [\gamma, \eta)$ for all $i \in J$. Hence, there exists a real number $\tilde{\rho}$ with $\tilde{\rho} < \rho$ such that $\mathcal{B}(w, T v) < \tilde{\rho} \mathcal{B}(T v, T v)$, and $h(\tilde{\rho}) > \gamma$. Now let $\tilde{\eta} := h(\tilde{\rho})$. Then, $\mathcal{B}(w - \tilde{\rho} T v, w - \tilde{\eta} T v) = 0$. From

$$\tilde{\eta} - \tilde{\rho} = \frac{\mathcal{B}(w - \tilde{\rho} T v, w - \tilde{\rho} T v)}{\mathcal{B}(w, T v) - \tilde{\rho} \mathcal{B}(T v, T v)} \leq 0,$$

the inequality $\gamma < \tilde{\eta} \leq \tilde{\rho} < \rho$ is obtained. Hence, $\mathcal{N}(v, \Phi_i) = 0$ holds for all $i \in J$ with $\tilde{\eta} \leq \lambda_i \leq \tilde{\rho}$. By Lemma 3, it follows that $v = 0$, which contradicts the inequality $0 > \mathcal{B}(w, T v) - \rho \mathcal{B}(T v, T v)$. Thus, $\mathcal{B}(w - \rho T v, w - \rho T v) = 0$ is proved. □

Proof of Theorem 12 (indirect proof): Assume that, for some $l \in \mathbb{N}$ with $l \leq n$ and $\mu_l < 0$, the interval $\left[\rho + \frac{1}{\mu_l}, \rho\right)$ contains strictly less than l eigenvalues of problem (26).

Since $A = \overline{A}^T$, $B = \overline{B}^T$, and B is positive definite, there exist vectors $x_1, \ldots, x_n$ in $\mathbb{K}^n$ with $x_k = (x_{k,1}, \ldots, x_{k,n})^T$ such that $Ax_i = \mu_i B x_i$, and $\overline{x}_i^T B x_k = \delta_{ik}$ for $i, k = 1, \ldots, n$.

We define $\tilde{v}_i$ and $\tilde{w}_i$ for $i = 1, \ldots, n$ by

$$\tilde{v}_i := \sum_{k=1}^{n} x_{i,k} v_k \quad \text{and} \quad \tilde{w}_i := \sum_{k=1}^{n} x_{i,k} w_k .$$

A simple calculation shows that

$$\mathcal{M}(\tilde{v}_i, \tilde{v}_k) - \rho \mathcal{N}(\tilde{v}_i, \tilde{v}_k) = x_i^T A \overline{x}_k = \mu_i \delta_{ik} \tag{28}$$

$$\mathcal{B}(\tilde{w}_i, \tilde{w}_k) - 2\rho \mathcal{M}(\tilde{v}_i, \tilde{v}_k) + \rho^2 \mathcal{N}(\tilde{v}_i, \tilde{v}_k) = x_i^T B \overline{x}_k = \delta_{ik} \tag{29}$$

for $i, k = 1, \ldots, n$.

Now let L be the subspace of D_M spanned by those eigenvectors of problem (26) which are associated with an eigenvalue contained in the interval $\left[\rho + \frac{1}{\mu_l}, \rho\right)$. From the assumption made at the beginning of the proof, it follows that $\dim L < l$ and there must exist numbers, $\beta_1, \ldots, \beta_l \in \mathbb{K}$, such that $(\beta_1, \ldots, \beta_l) \neq (0, \ldots, 0)$ and $\sum_{i=1}^{l} \beta_i \mathcal{N}(\tilde{v}_i, g) = 0$ for all $g \in L$. Define

$$v := \sum_{i=1}^{l} \beta_i \tilde{v}_i, \quad w := \sum_{i=1}^{l} \beta_i \tilde{w}_i, \quad \text{and } \eta := \rho + \frac{1}{\mu_l} .$$

Then $\eta < \rho$, $v \in D_M$, $w \in X$, $\mathcal{B}(Tf, w) = \mathcal{M}(f, v)$ for all $f \in D_M$, and $\mathcal{N}(v, \Phi_i) = 0$ for all $i \in J$ with $\eta \leq \lambda_i < \rho$. Furthermore, we can deduce from (28) and (29) that

$$\begin{aligned} &\mathcal{B}(w - \rho T v, w - \eta T v) \\ &= \mathcal{B}(w, w) - 2\rho \mathcal{M}(v, v) + \rho^2 \mathcal{N}(v, v) - \frac{1}{\mu_l} (\mathcal{M}(v, v) - \rho \mathcal{N}(v, v)) \\ &= \sum_{i=1}^{l} |\beta_i|^2 \left(1 - \frac{\mu_i}{\mu_l}\right) \leq 0 \end{aligned}$$

and

$$\mathcal{B}(w - \rho T v, w - \rho T v) = \sum_{i=1}^{l} |\beta_i|^2 > 0 .$$

But this contradicts the assertion $\mathcal{B}(w - \rho T v, w - \rho T v) = 0$, which is a consequence of Lemma 4. Thus, Theorem 12 is proved. □

As in the left definite case, Theorem 12 is a generalization of a theorem due to Lehmann. Since we later will present numerical examples where we use this theorem, we will formulate it here:

Theorem 13

Let assumptions R1, R2, and R4 be satisfied. Furthermore, for some $n \in \mathbb{N}$, let $v_i \in D_M$, $w_i \in D_N$ for each $i = 1, \ldots, n$ and let $\rho \in \mathbb{R}$. Assume that

$$\mathcal{N}(f, w_i) = \mathcal{M}(f, v_i) \quad \textit{for all} \quad f \in D_M \quad \textit{and each} \quad i = 1, \ldots, n . \tag{30}$$

We define matrices A_0, A_1, A_2, A, B by

$$A_0 := (\mathcal{N}(v_i, v_k))_{i,k=1,\ldots,n}, \quad A_1 := (\mathcal{M}(v_i, v_k))_{i,k=1,\ldots,n}, \quad A_2 := (\mathcal{N}(w_i, w_k))_{i,k=1,\ldots,n},$$

$A := A_1 - \rho A_0$, $B := A_2 - 2\rho A_1 + \rho^2 A_0$.

Assume that B is positive definite and that μ_i denotes for $i = 1, \ldots, n$ the i-th smallest eigenvalue of the eigenvalue problem

$$A x = \mu B x .$$

Then the interval $[\rho + \frac{1}{\mu_j}, \rho)$ contains for all $j \in J$ with $j \leq n$ and $\mu_j < 0$, at least j eigenvalues of eigenvalue problem (26).

Proof: The theorem is a consequence of Theorem 12, defining $X := D_N$; $T : D_M \to D_N$ by $T f := f$; and $\mathcal{B}(f,g) := \mathcal{N}(f,g)$ for all $f, g \in D_N$. □

Analogous to the theorems of section 4, a variational characterization for the eigenvalues can also be derived for right definite eigenvalue problems; from this characterization one obtains the Rayleigh–Ritz method (Theorem 14) as well as a comparison theorem analogous to Theorem 11.

Theorem 14 (Rayleigh–Ritz)
Let assumptions R1, R2, and R4 be satisfied. Furthermore, for some $n \in \mathbb{N}$, let $v_1, \ldots, v_n \in D_M$ be linearly independent. Denote the i-th smallest eigenvalue of (26) by λ_i, for $i = 1, \ldots, n$. We define matrices A_0, A_1 by

$$A_0 := (\mathcal{N}(v_i, v_k))_{i,k=1,\ldots,n}, \quad A_1 := (\mathcal{M}(v_i, v_k))_{i,k=1,\ldots,n},$$

and denote the i-th smallest eigenvalue of the problem

$$A_1 x = \Lambda A_0 x$$

by Λ_i. Then $\lambda_i \leq \Lambda_i$ for each $i = 1, \ldots, n$.

6. Application to Generalized Matrix Eigenvalue Problems

For the calculation of bounds to eigenvalues of eigenvalue problems with differential equations, the method of Lehmann (Theorem 13) has been proved to be very powerful. Therefore, it seems to be reasonable to use Lehmann's method for the calculation of bounds to eigenvalues of the matrix eigenvalue problem

$$A x = \lambda B x, \quad A = A^T, \quad B = B^T, \quad B \text{ positive definite} \tag{31}$$

with real $m \times m$ matrices A and B, as well[2]. If the method is combined with interval arithmetic, we obtain *guaranteed* bounds, that is, bounds which are secured against rounding errors. For this purpose, we have to solve linear systems $B w = f$, and the elements of B and f are intervals in the general case. This results in an overestimation of the solutions, as has been shown in [8]. By appropriate use of Theorem 12, we can avoid this weakness, since then the exact solution of the linear systems involved is not necessary.

Theorem 15
Let the following assumptions hold:

[2]The method can be formulated for complex Hermitian eigenvalue problems as well.

1. *A and B are real $m \times m$ matrices, with $A = A^T$, $B = B^T$, and B positive definite.*

2. *For an $n \in \mathbb{N}$, let $v_1, v_2, \ldots, v_n$ be linearly independent vectors of $\mathbb{R}^m$; let $\tilde{w}_i \in \mathbb{R}^m$ for $i = 1, \ldots, n$.*[3]

3. *Let $c \in \mathbb{R}$ so that $0 < c \leq \lambda_{min}(B)$, and let $\sigma \in \mathbb{R}$. Define the matrices A_0, A_1, A_2, $\hat{A}$, and $\hat{B}$ by*

$$A_0 := \left(v_i^T B v_k\right)_{i,k=1,\ldots,n},$$

$$A_1 := \left(v_i^T A v_k\right)_{i,k=1,\ldots,n},$$

$$A_2 := \left(v_i^T A \tilde{w}_k - v_i^T (B\tilde{w}_k - A v_k) + \frac{1}{c}(B\tilde{w}_i - A v_i)^T (B\tilde{w}_k - A v_k)\right)_{i,k=1,\ldots,n},$$

$$\hat{A} := A_1 - \sigma A_0, \quad \text{and } \hat{B} := A_2 - 2\sigma A_1 + \sigma^2 A_0.$$

Assume that $\hat{B}$ is positive definite. [4]

4. *The eigenvalues μ_i of the matrix eigenvalue problem $\hat{A}x = \mu \hat{B} x$ are ordered by magnitude: $\mu_1 \leq \mu_2 \leq \ldots \leq \mu_p < 0 \leq \ldots \leq \mu_n$.*

Assertion: *For $l = 1, \ldots, p$ the interval*

$$\left[\sigma + \frac{1}{\mu_l}, \sigma\right)$$

contains at least l eigenvalues of the problem $Ax = \lambda Bx$.

Proof: The theorem is reduced to Theorem 12. Let $\mathbb{K} := \mathbb{R}$, $D_M := \mathbb{R}^m$, $D_N := \mathbb{R}^m$. We define $\mathcal{M}(x,y) := x^T A y$ and $\mathcal{N}(x,y) := x^T B y$ for $x, y \in \mathbb{R}^m$. Thus, assumptions R1, R2, and R4 are satisfied. Furthermore,

$$X := \mathbb{R}^m \times \mathbb{R}^m,$$

$$\mathcal{B}\left(\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}\right) := x_1^T B y_1 - c x_1^T y_1 + c x_2^T y_2 \quad \text{for } x_1, x_2, y_1, y_2 \in \mathbb{R}^m,$$

$$T : \mathbb{R}^m \to \mathbb{R}^m \times \mathbb{R}^m, \quad Tx := \begin{pmatrix} x \\ x \end{pmatrix}.$$

Since $0 < c \leq \lambda_{min}(B)$, $\mathcal{B}(.,.)$ is positive semidefinite. For all $x, y \in \mathbb{R}^m$ we have

$$\mathcal{B}(Tx, Ty) = x^T B y - c x^T y + c x^T y = x^T B y.$$

If we define

$$w_i := \begin{pmatrix} \tilde{w}_i \\ \tilde{w}_i - \frac{1}{c}(B\tilde{w}_i - A v_i) \end{pmatrix} \quad \text{for } i = 1, \ldots, n,$$

[3] When we use the theorem in practice, we will choose $\tilde{w}_i$ as an approximation for $B^{-1} A v_i$ $(i = 1, \ldots, n)$.
[4] If σ is not an eigenvalue of the problem $Ax = \lambda Bx$, then $\hat{B}$ is positive definite.

this results in

$$\mathcal{B}(Ty, w_i) = y^T B \tilde{w}_i$$
$$= y^T A v_i$$

and

$$\mathcal{B}(w_i, w_k) = \tilde{w}_i^T B \tilde{w}$$
$$= v_i^T A \tilde{w}_k$$

Now the assertion is a

If in assumption 3 of

$$A_{2,L} := \left(v_i^T A \hat{w}_k\right)_{i,k=1,}$$

where $\hat{w}_i$ is the *exac* Theorem 15 reduces case.) But for a comp of $\hat{w}_i$. Since we are i but also in problems clear in subsection 6.2

6.1. Development

Consider the gener $m \times m$ matrices A an counted in accordanc

We implement the j) in three steps:

1. Calculation of a

$$\tilde{\lambda}_{r-1} < \tilde{\lambda}_r \approx \ldots$$

$(1 \leq r \leq j \leq s$ $\lambda_r, \ldots, \lambda_s$

(If either $r = 1$

Calculation of a

(The notation i the eigenvalue a $\tilde{\lambda}_{r-1}$ and $\tilde{\lambda}_{s+1}$ d $\lambda_r, \ldots, \lambda_s$ form λ_j (well separat

this results in

$$\begin{aligned}\mathcal{B}(Ty, w_i) &= y^T B\tilde{w}_i - cy^T\tilde{w}_i + cy^T(\tilde{w}_i - \frac{1}{c}(B\tilde{w}_i - Av_i)) \\ &= y^T Av_i \quad \text{for } i = 1,\ldots,n,\ y \in \mathbb{R}^m,\end{aligned}$$

and

$$\begin{aligned}\mathcal{B}(w_i, w_k) &= \tilde{w}_i^T B\tilde{w}_k - c\tilde{w}_i^T\tilde{w}_k + c(\tilde{w}_i - \frac{1}{c}(B\tilde{w}_i - Av_i))^T(\tilde{w}_k - \frac{1}{c}(B\tilde{w}_k - Av_k)) \\ &= v_i^T A\tilde{w}_k - \tilde{w}_i^T(B\tilde{w}_k - Av_k) + \frac{1}{c}(B\tilde{w}_i - Av_i)^T(B\tilde{w}_k - Av_k) \\ &\qquad \text{for } i,k = 1,\ldots,n.\end{aligned}$$

Now the assertion is a consequence of Theorem 12. □

If in assumption 3 of Theorem 15, the definition of A_2 is replaced by

$$A_{2,\mathrm{L}} := \left(v_i^T A\hat{w}_k\right)_{i,k=1,\ldots,n} = \left(\mathcal{N}(\hat{w}_i, \hat{w}_k)\right)_{i,k=1,\ldots,n},$$

where $\hat{w}_i$ is the *exact* solution of the linear system $B\hat{w}_i = Av_i$ $(i = 1,\ldots,n)$, then Theorem 15 reduces to Lehmann's theorem. (The constant c is not necessary in this case.) But for a computer realization, we have to replace $\hat{w}_i$ by an interval inclusion $[\hat{w}_i]$ of $\hat{w}_i$. Since we are interested not only in eigenvalue problems for real (point) matrices, but also in problems with interval matrices, the disadvantage of this variant will become clear in subsection 6.2.

6.1. Development of an algorithm

Consider the generalized symmetric definite matrix eigenvalue problem (31) with real $m \times m$ matrices A and B. Let the eigenvalues be ordered by magnitude and let them be counted in accordance with their multiplicity: $\lambda_1 \le \lambda_2 \le \ldots \le \lambda_m$.

We implement the calculation of bounds to an eigenvalue λ_j (with the prescribed index j) in three steps:

1. Calculation of approximate eigenvalues

$$\tilde{\lambda}_{r-1} < \tilde{\lambda}_r \approx \ldots \approx \tilde{\lambda}_j \approx \ldots \approx \tilde{\lambda}_s < \tilde{\lambda}_{s+1} \tag{32}$$

$(1 \le r \le j \le s \le m)$ as well as approximate eigenvectors $\tilde{x}_r,\ldots,\tilde{x}_s$ associated with $\lambda_r,\ldots,\lambda_s$

(If either $r = 1$ or $s = m$, then $\tilde{\lambda}_{r-1}$ or $\tilde{\lambda}_{s+1}$, respectively, is dropped.)

Calculation of an approximation $\tilde{\lambda}_{min}(B)$ to $\lambda_{min}(B)$

(The notation in (32) expresses the fact that r and s have to be chosen such that the eigenvalue approximations $\tilde{\lambda}_r,\ldots,\tilde{\lambda}_s$ differ only a little (less than 1 %), whereas $\tilde{\lambda}_{r-1}$ and $\tilde{\lambda}_{s+1}$ differ clearly from these values. Thus, we expect that the eigenvalues $\lambda_r,\ldots,\lambda_s$ form a cluster, and that λ_j belongs to this cluster. For a simple eigenvalue λ_j (well separated from other eigenvalues), we have, of course, $r = j = s$.)

2. Determination of a rough upper bound ρ to λ_{r-1} with $\lambda_{r-1} < \rho < \lambda_r$ (not necessary, if $r = 1$) and a rough lower bound σ for λ_{s+1} with $\lambda_s < \sigma < \lambda_{s+1}$ (not necessary, if $s = m$), as well as a constant c with $0 < c \leq \lambda_{min}(B)$, with the use of the eigenvalue approximations and LDL^T decompositions

3. Calculation of accurate bounds for λ_j with a slight modification of Theorem 15 and with the use of $\tilde{x}_r, \ldots, \tilde{x}_s$, ρ, σ, and c

In this context, Theorem 15 is of considerable importance since it allows for the calculation of bounds to eigenvalues of a matrix eigenvalue problem of dimension m by solving matrix eigenvalue problems of much smaller dimensions n. In the most favourable and common case, n will be equal to 1.

In an algorithmic implementation (see [7, 8]), it is better to use a slight modification of Theorem 15:

Theorem 16

Let assumptions 1 and 2 of Theorem 15 be valid, and let σ, c, A_0, A_1, and A_2 be defined as in assumption 3 of Theorem 15. Suppose σ is not an eigenvalue of the problem $A\,x = \lambda\,B\,x$. Let the eigenvalues τ_i of the eigenvalue problem

$$(A_2 - \sigma\,A_1)\,x = \tau\,(A_1 - \sigma\,A_0)\,x\,, \tag{33}$$

be ordered by magnitude: $\tau_1 \leq \tau_2 \leq \ldots \leq \tau_n$. If p is the number of eigenvalues τ_i smaller than σ, then the interval $[\tau_i, \sigma)$ contains at least $p + 1 - i$ eigenvalues of the problem $A\,x = \lambda\,B\,x$ $(i = 1, \ldots, p)$.

Proof: For a real number μ, $\mu \neq 0$, the following is valid:
μ is an eigenvalue of the problem

$$(A_1 - \sigma\,A_0)\,x = \tilde{\tau}\,(A_2 - 2\,\sigma\,A_1 + \sigma^2 A_0)\,x \quad \Longleftrightarrow$$

$\frac{1}{\mu}$ is an eigenvalue of the problem

$$((A_2 - \sigma\,A_1) - \sigma\,(A_1 - \sigma\,A_0))\,x = \hat{\tau}\,(A_1 - \sigma\,A_0)\,x \quad \Longleftrightarrow$$

$\sigma + \frac{1}{\mu}$ is an eigenvalue of the problem

$$(A_2 - \sigma\,A_1)\,x = \tau\,(A_1 - \sigma\,A_0)\,x\,.$$

Hence, the assertion follows from Theorem 15, since $\hat{B}$ is positive definite. □

Since Theorems 15 and 16 result only in intervals containing a certain number of eigenvalues, additional information on the parameter σ is necessary, in order to make a statement about the indices of the enclosed eigenvalues. If σ is a rough lower bound to λ_{s+1} with $\lambda_s < \sigma < \lambda_{s+1}$, and if p is the number of eigenvalues τ_i of problem (33) which are smaller than σ, then τ_i is in general a very precise lower bound to λ_{s+i-p} $(1 \leq i \leq p)$. This information on σ is calculated in the second step of the procedure and is not derived from a comparison problem, as is done when we deal with differential equations.

Two aspects play a part in the choice of σ: on the one hand, $s - j$ should be as small as possible, in order to be able to select a small n, because the inclusion theorems again

result in a matrix eigenvalue problem of order n. On the other hand, experience shows that σ should not belong to the same cluster of eigenvalues to which λ_j possibly belongs. Thus, the following requirement arises: λ_s must be the largest eigenvalue of the cluster to which λ_j belongs. In the case $s = j$ we can achieve $n = 1$; in this special case (33) reduces to a Temple quotient. If $s = m$, Theorem 16 simplifies to the theorem of Rayleigh–Ritz by passage to the limit $\sigma \to \infty$. (The matrix eigenvalue problem $A_1 x = \tau A_0 x$ then has to be considered.)

In order to get upper bounds, we consider the matrix eigenvalue problem $-A x = (-\lambda) B x$ instead of problem (31). Lower bounds for $-\lambda$, that is, upper bounds for λ, are then calculated with the use of Theorem 16. This results in

Theorem 17

Let assumptions 1 and 2 of Theorem 15 be valid, let $\rho \in \mathbb{R}$ and let c, A_0, A_1, and A_2 be defined as in assumption 3 of Theorem 15. Suppose ρ is not an eigenvalue of the problem $A x = \lambda B x$. Let the eigenvalues τ_i of the eigenvalue problem

$$(A_2 - \rho A_1) x = \tau (A_1 - \rho A_0) x ,$$

be arranged in descending order: $\tau_1 \geq \tau_2 \geq \ldots \geq \tau_n$. If p is the number of eigenvalues τ_i greater than ρ, then the interval $(\rho, \tau_i]$ contains at least $p + 1 - i$ eigenvalues of the problem $A x = \lambda B x$ $(i = 1, \ldots, p)$.

If λ_r denotes the smallest eigenvalue of the cluster to which λ_j belongs, and if ρ and σ are chosen such that $\rho < \lambda_r \leq \ldots \leq \lambda_j \leq \ldots \leq \lambda_s < \sigma$, then n is defined by $n := s - r + 1$. Thus, the same elements v_i and hence, the same matrices A_0, A_1, and A_2 can be used in Theorems 16 and 17. If $n > 1$, this yields not only bounds for λ_j, but also bounds for $\lambda_r, \ldots, \lambda_s$.

For the application of the theorems, we must explain the choice of the elements v_i. The approximate eigenvectors $\tilde{x}_r, \ldots, \tilde{x}_s$ (approximately B–orthonormalized) are used for $v_1, \ldots, v_n$. If necessary, these approximations can be improved, for example, with a Newton method [14] (using extended precision). Our numerical example in subsection 6.2 illustrates the influence of the quality of the approximations on the precision of the bounds. The vectors $\tilde{w}_i$ are defined as approximate solutions of the linear systems $B \tilde{w}_i = A v_i$ $(i = 1, \ldots, n)$. If the matrix B is ill-conditioned (with respect to inversion), we recommend an improvement of the approximations by means of iterative defect correction [40].

Our choice of the elements v_i offers several advantages: On the one hand, very good bounds can be expected on the basis of convergence theorems [49]; on the other hand, the matrices A_0, A_1, and A_2 have a nearly diagonal structure. Furthermore, the matrices $-(A_1 - \sigma A_0)$ and $(A_1 - \rho A_0)$ are in general positive definite, at least in cases where ρ is clearly smaller than λ_r and σ is clearly greater than λ_s (for $y \in \mathbb{R}^n$, $y \neq 0$, the Rayleigh quotient $\dfrac{y^T A_1 y}{y^T A_0 y}$ is approximately equal to λ_j). Because of the structure of the matrices, the positive definiteness can easily be proved numerically (for example, through the use of Gerschgorin discs).

The next lemma [38, p. 255] plays a part in the determination of ρ, σ, and c, as well as (combined with a bisection method) in the calculation of bounds to eigenvalues for matrices of small dimensions (Theorems 16 and 17, $n > 1$).

Lemma 5
Let $\tilde{A}$, $\tilde{B}$ be real $q \times q$ matrices, $\tilde{A} = \tilde{A}^T$, $\tilde{B} = \tilde{B}^T$; let $\tilde{B}$ be positive definite; and let $\gamma \in \mathbb{R}$. The number of eigenvalues $\tilde{\lambda}_i$ of $\tilde{A}x = \tilde{\lambda}\tilde{B}x$ which are smaller than, equal to, or greater than γ is equal to the number of negative, zero, or positive eigenvalues of the matrix $\tilde{A} - \gamma\tilde{B}$, respectively.

Now we can easily determine the number of negative, zero, and positive eigenvalues (the inertia) of a real symmetric matrix $\breve{A}$ with the help of a decomposition $\breve{A} = LDL^T$ [11]. Here, L is a regular lower triangular matrix, and D is a block diagonal matrix with blocks of order 1 or 2. For the implementation of the LDL^T decomposition with interval arithmetic, the same pivot strategy as in [11] can be used. This results in an interval matrix $[D]$ with a structure corresponding to D. If an interval with zero included should occur during the determination of the inertia for $[D]$ — that is, if zero is included in a block of order 1 or in the determinant of a block of order 2 —, it is impossible to decide whether this block corresponds to negative, zero, or positive eigenvalues. Then we alter the shift parameter γ slightly, and the decomposition is repeated (see [33, p. 46 - 49]).

6.2. Numerical Example

We consider a matrix eigenvalue problem[5] with 8×8 matrices given by :

$$A = (a_{ik})_{i,k=1,\ldots,8}\,, \quad a_{ik} = \delta_{ik}$$

and

$$[B] = ([b_{ik}])_{i,k=1,\ldots,8}\,, \quad [b_{ik}] = [\frac{1}{i+k-1} - 10^{-13}, \frac{1}{i+k-1} + 10^{-13}]\,.$$

The problem is

$$Ax = \lambda[B]x\,.$$

The matrix on the right hand side is an interval matrix, and if we define $\mathrm{mid}([B]) := \frac{1}{2}(\underline{B} + \overline{B})$, we have a condition number of

$$\frac{\lambda_8(\mathrm{mid}([B]))}{\lambda_1(\mathrm{mid}([B]))} \approx 1.5 \cdot 10^{10}\,;$$

thus, a comparison of the results computed by means of Lehmann's theorem (A_2 replaced by $[A_2] = ([\hat{w}_i]^T A v_k)$, $[\hat{w}_i]$ interval inclusion of the solution of the system $[B]\hat{w}_i = Av_i$) and by means of Theorem 15 should reflect the superiority of Theorem 15. Table 1 shows the diameters of the inclusions for the smallest 6 eigenvalues.

The entries *iteration* and *no iteration* mean that the eigenvalue and eigenvector approximations have been (or have not been) improved iteratively. The inclusions of the

[5]The numerical examples have been calculated with TPX, a Turbo Pascal extension of S.M. Rump and D. Husung, with double precision, that is, approximately 15.6 decimal figures as basic precision.

Procedure	diam($[\lambda_1]$)	diam($[\lambda_2]$)	diam($[\lambda_3]$)	diam($[\lambda_4]$)	diam($[\lambda_5]$)	diam($[\lambda_6]$)
spectral shift ($shift = \check{\lambda}_j$)						
1) Lehmann, Rump, no iteration	2.26E−9	1.96E−8	5.90E−8	7.84E−7	3.36E−4	6.84E−1
2) Lehmann, Jansson, no iteration	2.26E−9	1.96E−8	5.90E−8	7.84E−7	3.36E−4	6.84E−1
3) Theorem 15, no iteration, $B\,\tilde{w}_i \approx A\,v_i$	1.27E−11	2.10E−11	1.82E−9	5.57E−7	3.36E−4	6.85E−1
4) Theorem 15, no iteration, $\tilde{w}_i = \check{\lambda}_{r-1+i} v_i$	3.31E−2	3.55E−2	2.46E−3	1.31E−4	3.45E−4	6.85E−1
5) Lehmann, Rump, iteration	4.12E−13	1.56E−11	1.82E−9	5.56E−7	3.35E−4	6.84E−1
6) Lehmann, Jansson, iteration	4.12E−13	1.56E−11	1.82E−9	5.56E−7	3.35E−4	6.84E−1
7) Theorem 15, iteration, $B\,\tilde{w}_i \approx A\,v_i$	4.12E−13	1.57E−11	1.82E−9	5.57E−7	3.36E−4	6.85E−1
8) Theorem 15, iteration, $\tilde{w}_i = \check{\lambda}_{r-1+i} v_i$	4.12E−13	1.57E−11	1.82E−9	5.57E−7	3.36E−4	6.85E−1
no spectral shift ($shift = 0$)						
9) Lehmann, Rump, no iteration	2.46E−4	1.36E−2	1.02E−1	1.42E+0	4.41E+1	2.43E+3
10) Lehmann, Jansson, no iteration	2.42E−4	1.07E−2	8.75E−2	1.24E+0	3.83E+1	1.97E+3
11) Theorem 15, no iteration, $B\,\tilde{w}_i \approx A\,v_i$	1.27E−11	2.55E−11	2.10E−9	6.12E−7	3.57E−4	7.11E−1
12) Theorem 15, no iteration, $\tilde{w}_i = \check{\lambda}_{r-1+i} v_i$	3.31E−2	3.55E−2	2.46E−3	1.31E−4	3.66E−4	7.11E−1
13) Lehmann, Rump, iteration	2.19E−4	1.36E−2	1.02E−1	1.42E+0	4.41E+1	2.43E+3
14) Lehmann, Jansson, iteration	1.90E−4	1.07E−2	8.75E−2	1.24E+0	3.83E+1	1.97E+3
15) Theorem 15, iteration, $B\,\tilde{w}_i \approx A\,v_i$	5.02E−13	2.05E−11	2.10E−9	6.12E−7	3.57E−4	7.11E−1
16) Theorem 15, iteration, $\tilde{w}_i = \check{\lambda}_{r-1+i} v_i$	5.02E−13	2.05E−11	2.10E−9	6.12E−7	3.57E−4	7.11E−1

Table 1
Comparison of the different procedures (taken from [8])

solutions of the linear systems — necessary for the application of Lehmann's theorem — have been calculated either by *Rump's* [37] method or by *Jansson's* [22] method. For the entries $B\tilde{w}_i \approx A v_i$ and $\tilde{w}_i = \tilde{\lambda}_{r-1+i} v_i$, two possibilities for calculating the vectors $\tilde{w}_i$ are considered: Either the linear system is solved approximately or the vectors are derived from the equation $A v_i \approx \tilde{\lambda}_{r-1+i} B v_i$, so that, $\tilde{w}_i \approx \tilde{\lambda}_{r-1+i} v_i$. This yields 8 variants, and in each case one can calculate with or without spectral shift (that is, we consider the problem $(A - shift\, B)x = \hat{\lambda} B x$, $\lambda_j = shift + \hat{\lambda}_j$ or $Ax = \lambda B x$, the original one). We have carried out the calculation without spectral shift only to emphasize the effects; this version is not recommended. If $n = 1$ $(r = s)$, then A_0 is of magnitude 1, A_1 is of magnitude $\lambda_j - shift$ and A_2 is of magnitude $(\lambda_j - shift)^2$. Thus, A_2 has only a slight influence if $shift \approx \lambda_j$.

To give an idea of the quality of the approximations used for the calculation, we mention the approximations for the first eigenvalue: The initial approximation is

$$\tilde{\lambda}_1 = 5.896\,460\,471\,522\,22E - 1\,,$$

while the improved approximation is

$$\tilde{\lambda}_1 = 5.896\,438\,502\,888\,09E - 1\,,$$

and the inclusion for the first eigenvalue is

$$[\lambda_1] = [5.896\,438\,502\,8^{902}_{860}E - 1]\,,$$

that is, the initial approximation has 5 correct figures, while the improved approximation has about 12 correct figures.

The interpretation of the results in Table 1 is as follows: A comparison of rows 1,2,3, and 4 shows the predicted weakness of Lehmann's theorem in this context (in the context of differential equations, however, Lehmann's method provides excellent results, see subsections 7.1, 7.2), since the four versions use the same approximations and the same matrices $[A_0]$ and $[A_1]$. The bad results in row 4 can be explained by the bad quality of the vectors v_i. The superiority of Theorem 15 — applied with the good approximations v_i for the solutions of the linear systems — becomes clear. It should be stressed, that even with the bad (not iteratively improved) eigenvalue/eigenvector approximations, Theorem 15 (row 3) produces rather good inclusions (compare with row 5,6,7, or 8). Since the calculation has been carried out with spectral shift, the influence of the matrix $[A_2]$ is small; this explains why the results in rows 1 and 2 are identical.

Rows 5,6,7, and 8 demonstrate that every procedure yields excellent results if high quality approximations are used. This is no surprise, since in this case the matrix $[A_2]$ has only a slight influence on the results. Rows 9,10,11, and 12 demonstrate even more emphatically the same effect as rows 1, 2, 3, and 4. (The advantage of Jansson's method as compared with the method of Rump can be seen in rows 9 and 10.) A comparison of the last four rows makes clear that — without a spectral shift — Lehmann's theorem cannot take advantage of the high quality approximations; the results in rows 13 and 14 are nearly the same as in rows 9 and 10 (and considerably worse than those in rows 11 and 15).

7. Numerical Examples for Eigenvalue Problems with Differential Equations

In this section, we illustrate the application of the right and of the left definite theory to eigenvalue problems with differential equations by means of four numerical examples. (We always assume $\mathbb{K} = \mathbb{R}$.) The examples show that very accurate bounds to eigenvalues can be computed with the methods discussed above.

7.1. The Mathieu Equation

Our first example is the well known Mathieu equation. This equation has been considered by several authors, bounds for eigenvalues of the Mathieu equation can be found in Albrecht [1], Neher [31] and Bazley [45], Weinstein [46]. The eigenvalue problem reads as follows

$$\begin{aligned} -\Phi''(x) + s\cos^2(x)\,\Phi(x) &= \lambda\,\Phi(x) \quad \text{for } x \in [0,\pi], \\ \Phi'(0) &= \Phi'(\pi) = 0, \\ \Phi(\tfrac{\pi}{2} - x) &= \Phi(\tfrac{\pi}{2} + x), \quad x \in [0, \tfrac{\pi}{2}], \end{aligned} \tag{34}$$

where $s \in \mathbb{R}$ is a parameter.

We use the right definite theory to treat this problem.

$$D_N := \{f \in C([0,\pi]) \,:\, f(\tfrac{\pi}{2} - x) = f(\tfrac{\pi}{2} + x),\ x \in [0, \tfrac{\pi}{2}]\},$$

$$D_M := D_N \cap \{f \in C^1([0,\pi]) \,:\, f'(0) = f'(\pi) = 0\}.$$

Bilinear forms $\mathcal{M}_s(.,.)$ and $\mathcal{N}(.,.)$ are defined by

$$\begin{aligned} \mathcal{M}_s(f,g) &:= \int_0^\pi (f'(x)g'(x) + s\cos^2(x)f(x)g(x))dx \quad \text{for all } f,g \in D_M \text{ and} \\ \mathcal{N}(f,g) &:= \int_0^\pi f(x)\,g(x)\,dx \quad \text{for all } f,g \in D_N. \end{aligned} \tag{35}$$

Then the eigenvalue problem

"Find pairs $(\lambda(s), \Phi_s) \in \mathbb{R} \times D_M$ such that $\Phi \neq 0$ and
$\mathcal{M}_s(f, \Phi_s) = \lambda(s)\mathcal{N}(f, \Phi_s)$ for all $f \in D_M$". (36)

is equivalent to (34). We define trial functions $v_k \in D_M$ by

$$v_1(x) := \frac{1}{\sqrt{\pi}} \quad \text{and} \quad v_k(x) := \sqrt{\frac{2}{\pi}}\cos(2\,(k-1)\,x) \quad \text{for } x \in (0,\pi),\ k = 2,\ldots,n.$$

With these trial functions we can compute the Rayleigh–Ritz upper bounds $\Lambda_i(s)$ (see section 5). In order to obtain (rough) lower bounds, we apply a right definite comparison theorem[6]. This results in

$$\lambda_i(s) \leq \lambda_i(\tilde{s}) \quad \text{for } s, \tilde{s} \in \mathbb{R},\ s \leq \tilde{s},\ i \in \mathbb{N}.$$

For $s = 0$, problem (36) can be solved in closed form and has the eigenvalues $\lambda_i(0) = 4(i-1)^2$. These values can be chosen for ρ_i, that is, if we define $\rho_i := \lambda_i(0)$ and $r = 25$, we have

$$\lambda_{25}(s) \leq \Lambda_{25}(s) \leq \rho_{26} = 2500 \leq \lambda_{26}(s) \tag{37}$$

[6]For $s > 0$ our eigenvalue problem (36) is left definite as well. Hence, Theorem 10 can be applied with $\mathcal{M} = \mathcal{M}_s$, $\tilde{\mathcal{M}} = \mathcal{M}_{\tilde{s}}$, $\tilde{\mathcal{N}} = \mathcal{N}$ as defined in (35) and $D = \tilde{D} = D_M$.

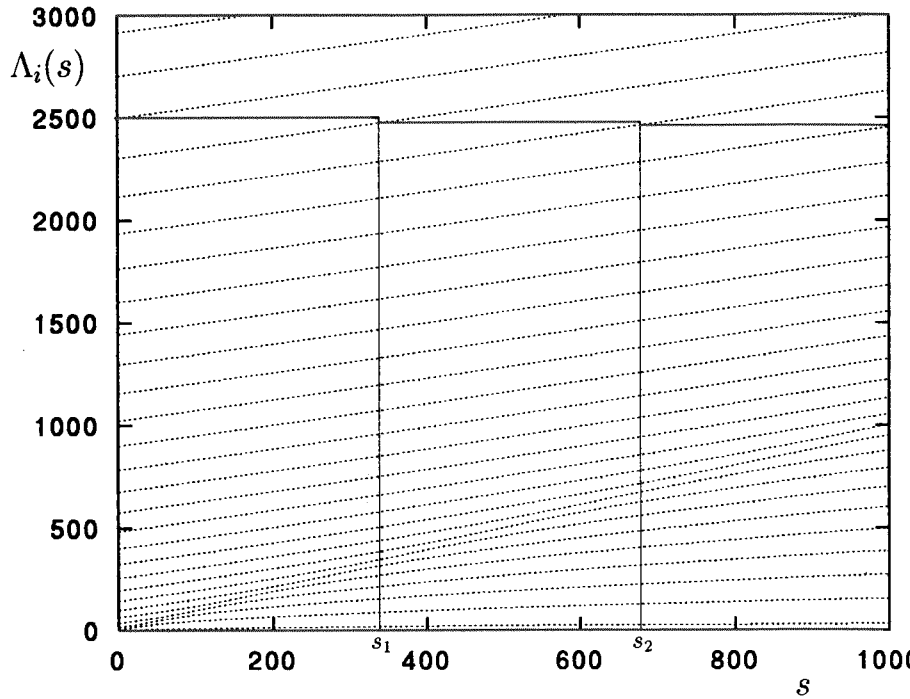


Figure 1. Rayleigh–Ritz upper bounds for eigenvalues of (36) dependent on s

for $s \geq 0$, s sufficiently small. From Figure 1 (an approximate Rayleigh–Ritz computation) we can expect that (37) is valid at least for $s \leq s_1 := 340$. A verified Rayleigh–Ritz computation (verified by means of interval arithmetic) immediately yields $\Lambda_{25}(s_1) \in [2475.569\frac{3}{2}]$, and hence, (37) is valid at least for $s \in [0, s_1]$. Thus, we can compute lower bounds for the eigenvalues $\lambda_1(s), \ldots, \lambda_{25}(s)$, $s \in [0, 340]$ by means of Theorem 13.

Allowed values of ρ for $s > s_1$, say for $s = 1000$, can be computed by means of a homotopy method (see Goerisch [17] and Plum [35]). For $s = s_1$ we compute a (verified) lower bound for $\lambda_{25}(s_1)$,

$$2475.5692 \leq \lambda_{25}(s_1)\,.$$

Due to the monotonicity of the eigenvalues, this is also a lower bound for $\lambda_{25}(s)$ for $s > s_1$ and a glance at Figure 1 followed by a verified Rayleigh–Ritz computations gives us the inequality

$$\lambda_{24}(s) \leq \Lambda_{24}(s) \leq 2475.5692 \leq \lambda_{25}(s) \quad \text{for} \quad s_1 \leq s \leq s_2 := 680.$$

	$s = 2$	$s = 1000$
λ_1	$[8.782\,344\,550\,5_{68}^{90}\,E - 1]$	$[3.137\,075\,158\,983_{13}^{96}\,E + 1]$
λ_2	$[5.100\,900\,595\,5_{57}^{61}\,E + 0]$	$[1.547\,904\,841\,4871_{06}^{91}\,E + 2]$
λ_3	$[1.700\,836\,462\,27_{19}^{24}\,E + 1]$	$[2.739\,415\,967\,992_{16}^{22}\,E + 2]$
λ_4	$[3.700\,357\,169\,59_{08}^{11}\,E + 1]$	$[3.885\,640\,377\,453_{48}^{53}\,E + 2]$
λ_5	$[6.500\,198\,416\,955_{52}^{73}\,E + 1]$	$[4.983\,373\,174\,464_{68}^{72}\,E + 2]$
λ_6	$[1.010\,012\,626\,3689_{16}^{36}\,E + 2]$	$[6.028\,531\,304\,1132_{44}^{77}\,E + 2]$
λ_7	$[1.450\,008\,741\,2934_{04}^{26}\,E + 2]$	$[7.015\,668\,560\,274_{89}^{92}\,E + 2]$
λ_8	$[1.970\,006\,410\,2699_{34}^{52}\,E + 2]$	$[7.937\,004\,756\,03_{33}^{53}\,E + 2]$
λ_9	$[2.570\,004\,901\,966_{78}^{81}\,E + 2]$	$[8.779\,832\,679\,55_{18}^{24}\,E + 2]$
λ_{10}	$[3.250\,003\,869\,97198_{34}^{75}\,E + 2]$	$[9.509\,847\,221\,38_{39}^{44}\,E + 2]$
λ_{11}	$[4.010\,003\,132\,8336_{26}^{39}\,E + 2]$	$[1.002\,491\,676\,0682_{19}^{57}\,E + 3]$
λ_{12}	$[4.850\,002\,587\,9925_{71}^{96}\,E + 2]$	$[1.057\,178\,409\,622_{68}^{72}\,E + 3]$
λ_{13}	$[5.770\,002\,173\,9135_{39}^{63}\,E + 2]$	$[1.134\,382\,857\,9660_{09}^{24}\,E + 3]$
λ_{14}	$[6.770\,001\,851\,8521_{39}^{73}\,E + 2]$	$[1.224\,608\,486\,3796_{19}^{35}\,E + 3]$
λ_{15}	$[7.850\,001\,596\,424_{14}^{22}\,E + 2]$	$[1.325\,329\,650\,3395_{08}^{29}\,E + 3]$
λ_{16}	$[9.010\,001\,390\,433_{88}^{96}\,E + 2]$	$[1.435\,672\,605\,0195_{66}^{71}\,E + 3]$
λ_{17}	$[1.025\,000\,122\,1896_{35}^{48}\,E + 3]$	$[1.555\,154\,451\,96898_{53}^{74}\,E + 3]$
λ_{18}	$[1.157\,000\,108\,225_{08}^{12}\,E + 3]$	$[1.683\,472\,366\,2003_{04}^{15}\,E + 3]$
λ_{19}	$[1.297\,000\,096\,52510_{08}^{11}\,E + 3]$	$[1.820\,423\,551\,779_{65}^{94}\,E + 3]$
λ_{20}	$[1.445\,000\,086\,625_{08}^{14}\,E + 3]$	$[1.965\,865\,955\,7_{12}^{23}\,E + 3]$
λ_{21}	$[1.601\,000\,078\,1738_{60}^{80}\,E + 3]$	$[2.119\,696\,69_{48}^{53}\,E + 3]$
λ_{22}	$[1.765\,000\,070\,9018_{73}^{81}\,E + 3]$	$[2.281\,839\,2_{72}^{89}\,E + 3]$
λ_{23}	$[1.937\,000\,064\,59948_{44}^{67}\,E + 3]$	$[2.452\,2_{29}^{36}\,E + 3]$
λ_{24}	$[2.117\,000\,059\,10165_{55}^{70}\,E + 3]$	—
λ_{25}	$[2.305\,000\,054\,277030_{35}^{82}\,E + 3]$	—

Table 2
Bounds for eigenvalues of the Mathieu equation

The (verified) Lehmann lower bound for $\lambda_{24}(s_2)$ (computed with $\rho = 2475.5692$) is $2462.8454 \leq \lambda_{24}(s_2)$, this is an admissible spectral parameter ρ for $s = 1000$.

In Table 2 we give numerical results for $s = 2$ and for $s = 1000$ (computed with $n = 30$ trial functions), which can be compared with those of Bazley [45] and Weinstein [46]. Unfortunately not any of their bounds is correct. The reason for the fact that some of their lower bounds is indeed an upper bound and vice versa is that they used standard floating point arithmetic and *not* interval arithmetic.

7.2. Vibrations of Turbine Blades

In this subsection we consider the natural bending vibrations of a free standing blade of a turbine disc. The mathematical model describing this problem [21] results in an eigenvalue problem with a system of ordinary differential equations of fourth order:

$$\begin{aligned} \left(\Psi_z\Phi_y'' + \Psi_{yz}\Phi_z''\right)'' - \Omega^2\left(\Theta\Phi_y'\right)' &= \lambda\Phi_y, \\ \left(\Psi_{yz}\Phi_y'' + \Psi_y\Phi_z''\right)'' - \Omega^2\left(\Theta\Phi_z'\right)' - \Omega^2\Phi_z &= \lambda\Phi_z, \end{aligned} \tag{38}$$

and

$$\begin{aligned} \Phi_y(0) = \Phi_y'(0) = \Phi_z(0) = \Phi_z'(0) &= 0, \\ \Phi_y''(1) = \Phi_y'''(1) = \Phi_z''(1) = \Phi_z'''(1) &= 0. \end{aligned} \tag{39}$$

Where Ψ_η and Ψ_ζ are real numbers and

$$\begin{aligned} \Theta &= \Theta(x) = 0.5\,(1-x)\,(x+2\epsilon+1), \\ \Psi_y &= \Psi_y(x) = \Psi_\eta \cos^2(\gamma\,x) + \Psi_\zeta \sin^2(\gamma\,x), \\ \Psi_z &= \Psi_z(x) = \Psi_\eta \sin^2(\gamma\,x) + \Psi_\zeta \cos^2(\gamma\,x), \\ \Psi_{yz} &= \Psi_{yz}(x) = (\Psi_\zeta - \Psi_\eta)\sin(\gamma\,x)\,\cos(\gamma\,x). \end{aligned}$$

The most important parameters (see Figure 2) are

x	cartesian coordinate of the blade
$\Phi_y = \Phi_y(x)$	first component of the eigenfunction: displacement in y-direction
$\Phi_z = \Phi_z(x)$	second component of the eigenfunction: displacement in z-direction
$\gamma\,x$	angle of twist
$\Psi_y, \Psi_z, \Psi_{yz}$	squares of the radii of gyration
$\Omega^2\Theta(x)$	normal force in the blade
Ω	angular velocity
λ	eigenvalue (square of the eigenfrequency, dimensionless)

We assume $\Psi_\eta = 1$, $\Psi_\zeta = 87.11$, $\epsilon = 0.457$, $\gamma = \frac{\pi}{180}$ (values suggested by Professor Irretier, Institut für Mechanik, Gesamthochschule Kassel), and we are interested in computing bounds for the smaller eigenvalues for Ω close to 9 (see Figure 3).

In order to apply the right definite theory (theorems 13 and 14), we define

$$D_N := L_2(0,1) \times L_2(0,1)$$

$$D_M := \left\{ f = \begin{pmatrix} f_1 \\ f_2 \end{pmatrix} : f_i \in C^4[0,1], f_i(0) = f_i'(0) = f_i''(1) = f_i'''(1) = 0, i = 1,2 \right\}$$

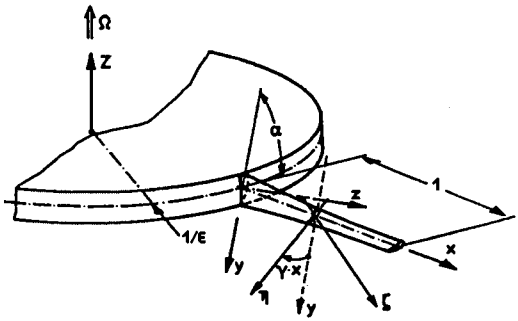


Figure 2. Notations

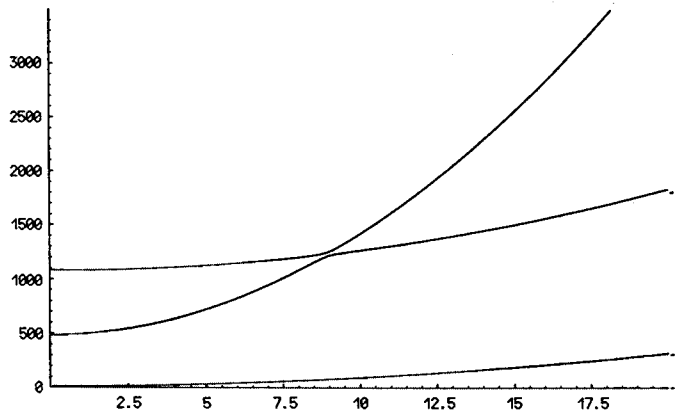


Figure 3. Eigenvalue curves

and

$$\mathcal{N}(f,g) := \int_0^1 (f_1 g_1 + f_2 g_2)dx \quad \text{for } f = \begin{pmatrix} f_1 \\ f_2 \end{pmatrix}, g = \begin{pmatrix} g_1 \\ g_2 \end{pmatrix} \in D_N$$

$$\mathcal{M}(f,g) := \int_0^1 \{\Psi_z f_1'' g_1'' + \Psi_{yz} f_2'' g_1'' + \Omega^2 \Theta f_1' g_1'$$

$$+\Psi_{yz} f_1'' g_2'' + \Psi_y f_2'' g_2'' + \Omega^2 \Theta f_2' g_2' - \Omega^2 f_2 g_2\} dx \quad \text{for } f = \begin{pmatrix} f_1 \\ f_2 \end{pmatrix}, g = \begin{pmatrix} g_1 \\ g_2 \end{pmatrix} \in D_M$$

The trial functions are constructed by means of polynomials

$$\tilde{p}_1(x) := x^2\left(\frac{3}{2} - x + \frac{x^2}{4}\right)$$
$$\tilde{p}_2(x) := x^3\left(\frac{2}{3} - \frac{2x}{3} + \frac{x^2}{5}\right)$$

$$\tilde{p}_i(x) := (-1+x)^4 x^{i-1} \quad \text{for } i \geq 3.$$

Obviously, the $\tilde{p}_i$ fulfill the boundary conditions $\tilde{p}_i(0) = \tilde{p}_i'(0) = \tilde{p}_i''(1) = \tilde{p}_i'''(1) = 0$. To avoid the well known numerical problems with ill conditioned matrices, we construct an orthogonal basis from the $\tilde{p}_i$ (orthogonal with respect to the L_2 inner product $(f,g) = \int_0^1 f\,g\,dx$) using the Gram-Schmidt process and the computer algebra program *Mathematica* (see [48]). We obtain

$$p_1(x) := \frac{x^2}{3}(6 - 4x + x^2)$$
$$p_2(x) := \frac{x^2}{19}(-326 + 824x - 661x^2 + 182x^3)$$
$$p_3(x) := \frac{x^2}{595}(37490 - 181120x + 305815x^2 - 218966x^3 + 57376x^4)$$
$$p_4(x) := \frac{x^2}{17335}(-2548170 + 19398020x - 54146415x^2 + 70839756x^3 - 44146336x^4 + 10620480x^5)$$
$$p_5(x) := \frac{x^2}{143155}(40512210 - 437785780x + 1790279235x^2 - 3625862604x^3 + 3896636744x^4 - 2131724400x^5 + 468087750x^6)$$
$$p_6(x) := \frac{x^2}{8285}(-4034766 + 58114976x - 323567649x^2 + 923419434x^3 - 1482348280x^4 + 1354376928x^5 - 658061874x^6 + 132109516x^7)$$
$$\vdots$$

Now we choose $n_1, n_2 \in \mathbb{N}$, $n := n_1 + n_2$ and define

$$v_i := \begin{pmatrix} p_i \\ 0 \end{pmatrix} \quad \text{for } i = 1, \ldots, n_1$$
$$v_i := \begin{pmatrix} 0 \\ p_{i-n_1} \end{pmatrix} \text{ for } i = n_1 + 1, \ldots, n_1 + n_2 = n$$

as trial functions.

Our numerical results (with $n_1 = n_2 = 10$) are given in Table 3.

	$\Omega = 8.80$
λ_1	$[\ 76.709\,680\,2^{50}_{33}]$
λ_2	$[1197.598\,903\,^{8}_{3}]$
λ_3	$[1236.299\,823\,^{7}_{5}]$
λ_4	$[5955.927\,2^{758}_{688}]$

	$\Omega = 8.90$
λ_1	$[\ 78.106\,306\,^{512}_{492}]$
λ_2	$[1209.059\,506\,^{7}_{3}]$
λ_3	$[1244.584\,05^{52}_{49}]$
λ_4	$[6004.260\,6^{414}_{331}]$

	$\Omega = 9.00$
λ_1	$[\ 79.516\,862\,^{613}_{590}]$
λ_2	$[1218.317\,428\,^{8}_{4}]$
λ_3	$[1255.287\,28^{83}_{78}]$
λ_4	$[6053.108\,3^{816}_{717}]$

	$\Omega = 9.10$
λ_1	$[\ 80.941\,335\,^{426}_{399}]$
λ_2	$[1225.541\,882\,^{6}_{4}]$
λ_3	$[1268.240\,15^{25}_{18}]$
λ_4	$[6102.469\,6^{939}_{822}]$

Table 3
Bounds for eigenvalues of the problem of turbine blade vibrations

7.3. A Stekloff Eigenvalue Problem

Our next example is a Stekloff eigenvalue problem of fourth order; bounds for the smallest eigenvalue of this problem have been computed by Kuttler [27].

$$\begin{aligned} \Delta^2\Phi &= 0 \quad \text{in } \Omega, \\ \tfrac{\partial\Phi}{\partial n} &= 0 \quad \text{on } \partial\Omega, \\ -\tfrac{\partial\Delta\Phi}{\partial n} &= \lambda\Phi \quad \text{on } \partial\Omega, \\ \Phi(x,y) &= \Phi(-x,y) = -\Phi(x,-y) \text{ for } (x,y)\in\Omega, \end{aligned} \tag{40}$$

with

$$\Omega := \{(x,y)\in I\!R^2 \,:\, |x|<1,\ |y|<1\}.$$

If one denotes the smallest nonzero eigenvalue of this problem by ξ_2 (cf. Kuttler), then ξ_2 is one of the optimal constants in the *a priori* inequality

$$(\int_\Omega v^2 dx\,dy)^{\frac{1}{2}} \le (\frac{1}{\mu_2^2}\int_\Omega (\Delta v)^2 dx\,dy)^{\frac{1}{2}} + (\frac{1}{\xi_2}\int_{\partial\Omega}(\frac{\partial v}{\partial n})^2 dS)^{\frac{1}{2}}$$

for $v\in H^2(\Omega)$, $\int_\Omega v dx\,dy = \int_\Omega \Delta v dx\,dy = 0$, where μ_2 is the smallest nonzero membrane eigenvalue for Ω.
The above inequality is useful for obtaining error bounds for the boundary value problem

$$\Delta v = f \ \text{ in } \ \Omega,$$

$$\frac{\partial v}{\partial n} = 0 \text{ on } \partial\Omega,$$
$$\int_\Omega v\,dx\,dy = 0.$$

We want to apply the left definite theory to problem (40), so we begin by giving the domain and the bilinear forms

$$D := \{f \in H^2(\Omega) : \frac{\partial f}{\partial n} = 0 \text{ on } \partial\Omega,\ f(x,y) = f(-x,y) = -f(x,-y) \text{ for } (x,y) \in \Omega\},$$

$$\left.\begin{aligned} \mathcal{M}(f,g) &:= \int_\Omega \Delta f\,\Delta g\,dx\,dy \\ \mathcal{N}(f,g) &:= \int_{\partial\Omega} f\,g\,dS \end{aligned}\right\} \text{ for } f,g \in D.$$

In order to construct X, $\mathcal{B}$, and T for Theorem 6, we use integration by parts:

$$\begin{aligned} \int_\Omega \Delta u\,\Delta v\,dx\,dy &= \int_\Omega u\,\Delta\Delta v\,dx\,dy - \int_{\partial\Omega} u\frac{\partial \Delta v}{\partial n}\,dS + \int_{\partial\Omega}\frac{\partial u}{\partial n}\Delta v\,dS \\ &= \int_\Omega u\,\Delta\Delta v\,dx\,dy - \int_{\partial\Omega} u\frac{\partial \Delta v}{\partial n}\,dS \quad \text{for } u,v \in D \cap H^4(\Omega). \end{aligned}$$

We define

$$r = 2,\quad M_1 f := \Delta^2 f,\qquad T_1 f := f,$$
$$M_2 f := -\frac{\partial \Delta f}{\partial n},\quad T_2 f := f.$$

Using the *a priori* inequalities

$$\int_\Omega (\Delta f)^2\,dx\,dy \geq \frac{\pi^4}{16}\int_\Omega f^2\,dx\,dy,$$

$$\int_\Omega (\Delta f)^2\,dx\,dy \geq \frac{\pi^4}{16}(\pi\sqrt{2}+1)^{-1}\int_{\partial\Omega} f^2\,dS,$$

for $f \in D$ — the first one follows from

$$\frac{\pi^2}{4} \leq \frac{\int_\Omega f(-\Delta f)\,dx\,dy}{\int_\Omega f^2\,dx\,dy}$$

for all $f \in D$, $f \neq 0$, and

$$\left(\int_\Omega f(-\Delta f)\,dx\,dy\right)^2 \leq \int_\Omega (\Delta f)^2\,dx\,dy \int_\Omega f^2\,dx\,dy$$

for all $f \in D$, the second one can be found in [27] — we obtain

$$\int_\Omega (\Delta f)^2\,dx\,dy \geq \frac{\pi^4}{32}\int_\Omega f^2\,dx\,dy + \frac{\pi^4}{32}(\pi\sqrt{2}+1)^{-1}\int_{\partial\Omega} f^2\,dS.$$

Thus, we can define

$$q_1 := \tfrac{\pi^4}{32}, \qquad \kappa_1 := 1,$$

$$q_2 := \tfrac{\pi^4}{32}(\pi\sqrt{2}+1)^{-1}, \quad \kappa_2 := 1,$$

and the construction procedure for X, $\mathcal{B}$, T (see section 2) yields

$$\begin{aligned} X &:= D \times L_2(\Omega) \times L_2(\partial\Omega), \\ Tf &:= \begin{pmatrix} f \\ f \\ f \end{pmatrix} \quad \text{for} \quad f \in D, \\ \mathcal{B}\left(\begin{pmatrix} f_0 \\ f_1 \\ f_2 \end{pmatrix}, \begin{pmatrix} g_0 \\ g_1 \\ g_2 \end{pmatrix}\right) &:= \int_\Omega \Delta f_0 \, \Delta g_0 \, dx\, dy - q_1 \int_\Omega f_0 g_0 \, dx dy + q_1 \int_\Omega f_1 g_1 \, dx dy \\ &\qquad - q_2 \int_{\partial\Omega} f_0 g_0 \, dS + q_2 \int_{\partial\Omega} f_2 g_2 \, dS \end{aligned}$$

$$\text{for} \quad \begin{pmatrix} f_0 \\ f_1 \\ f_2 \end{pmatrix}, \begin{pmatrix} g_0 \\ g_1 \\ g_2 \end{pmatrix} \in X.$$

We choose the trial functions (see Theorem 8) as

$$\begin{aligned} v_j &:= \textstyle\sum_{(i,k)\in\Gamma_v} c_{jik} v_{ik} \\ \hat{w}_j^* &:= 0 \\ \hat{w}_j^0 &:= \textstyle\sum_{(i,k)\in\Gamma_w} d_{jik} \hat{w}_{ik}^0 \end{aligned} \tag{41}$$

for $j \in \mathbb{N}$, where

$$\begin{aligned} v_{ik}(x,y) &:= (2\,i\,x^{2\,i-2} - (2\,i-2)x^{2\,i})((2\,k+1)y^{2\,k-1} - (2\,k-1)y^{2\,k+1}) \\ \hat{w}_{ik}^0(x,y) &:= (2\,i\,x^{2\,i-2} - (2\,i-2)x^{2\,i})((2\,k+1)y^{2\,k-1} - (2\,k-1)y^{2\,k+1}) \end{aligned} \tag{42}$$

for $i, k \in \mathbb{N}$.
Here Γ_v and Γ_w are index sets and the real numbers c_{jik} and d_{jik} can be obtained easily from an approximate solution of a matrix eigenvalue problem. For more details see [39].

An inclusion with $|\Gamma_v| = 7$ and $|\Gamma_w| = 16$ is

$$\lambda_1 = \xi_2 \in [1.995\,8^{1297}_{0873}].$$

Kuttler obtains by his method of *a posteriori – a priori* inequalities with 10 trial functions and standard floating point arithmetic

$$\lambda_1 = \xi_2 \in [1.995^{8125}_{7375}],$$

that is, Kuttler's interval is wider than our interval in spite of his having used more trial functions.
Our result with $|\Gamma_v| = 25$ and $|\Gamma_w| = 110$ is

$$\lambda_1 = \xi_2 \in [1.995\,812\,883\,2^{52}_{31}],$$

Thus, Kuttler's upper bound is wrong due to rounding errors.

7.4. Buckling of a Rectangular Clamped Plate under Pressure

Our next example plays a part in the determination of the buckling load of a rectangular clamped plate under pressure:

$$\begin{aligned} \Delta^2\Phi &= -\lambda\Delta\Phi \quad \text{in } \Omega, \\ \Phi &= 0 \quad \text{on } \partial\Omega, \\ \tfrac{\partial\Phi}{\partial n} &= 0 \quad \text{on } \partial\Omega, \\ \Phi(x,y) &= \Phi(-x,y) = \Phi(x,-y) \text{ for } (x,y)\in\Omega, \end{aligned} \tag{43}$$

with

$$\Omega := \{(x,y)\in I\!R^2 \,:\, |x|<\frac{\pi}{2},\ |y|<\frac{\pi}{2}\}.$$

One typically is interested in the smallest eigenvalue of (43).
We define the domain and the bilinear forms (see [20]) by

$$D := \{f\in H_0^2(\Omega) \,:\, f(x,y)=f(-x,y)=f(x,-y) \text{ for } (x,y)\in\Omega\}$$

and

$$\left.\begin{aligned} \mathcal{M}(f,g) &:= \int_\Omega \Delta f\,\Delta g\,dx\,dy \\ \mathcal{N}(f,g) &:= \int_\Omega (\frac{\partial f}{\partial x}\frac{\partial g}{\partial x}+\frac{\partial f}{\partial y}\frac{\partial g}{\partial y})\,dx\,dy \end{aligned}\right\} \text{ for } f,g\in D.$$

The quantities X, $\mathcal{B}$, and T for the left definite theory can be defined by

$$\begin{aligned} X &:= L_2(\Omega) \\ Tf &:= \Delta f \text{ for all } f\in D \\ \mathcal{B}(f,g) &:= \int_\Omega f\,g\,dx\,dy. \end{aligned}$$

We choose the trial functions:

$$\begin{aligned} v_{i,k}^{(1)}(x,y) &:= g_i(x)\,g_k(y)+g_k(x)\,g_i(y), \\ v_{i,k}^{(2)}(x,y) &:= h_i(x)\,h_k(y)+h_k(x)\,h_i(y), \\ v_{i,k}^{(3)}(x,y) &:= g_i(x)\,h_k(y)+h_k(x)\,g_i(y), \end{aligned}$$

with

$$\begin{aligned} g_i(t) &:= \sin(i(t+\frac{\pi}{2}))-\frac{i}{i+2}\sin((i+2)(t+\frac{\pi}{2})), \\ h_i(t) &:= \cos((i-1)(t+\frac{\pi}{2}))-\cos((i+1)(t+\frac{\pi}{2})), \end{aligned}$$

for $i,k\in I\!N$.
Our result with 41 trial functions (cf. [32]) is

$$\lambda_1 \in [5.303\,62^{62}_{56}].$$

Further numerical examples, for which the discussed methods have been applied successfully (without the use of interval arithmetic) can be found in [2, 9].

The authors are grateful to the referees for valuable hints and suggestions, which helped to make this paper more readable.

REFERENCES

1. J. Albrecht, Iterationsverfahren zur Berechnung der Eigenwerte der Mathieuschen Differentialgleichung. *Z. Angew. Math. Mech.*, 44:453–458, 1964.
2. J. Albrecht and F. Goerisch, Anwendung des Verfahrens von Lehmann auf Schwingungsprobleme. In: J. Albrecht, L. Collatz, W. Velte, and W. Wunderlich, editors, *Numerische Behandlung von Eigenwertaufgaben, Band 4. International Series of Numerical Mathematics (ISNM), Volume 83*, pages 1–9. Birkhäuser-Verlag: Basel-Boston, 1987.
3. G. Alefeld and J. Herzberger, *Introduction to Interval Computations.* Academic Press, New York, 1983.
4. N. W. Bazley and D. W. Fox, Truncations in the method of intermediate problems for lower bounds to eigenvalues. *J. Res. Nat. Bur. Standards, Sect. B*, 65B:105–111, 1961.
5. C. Beattie, An extension of Aronszajn's rule: Slicing the spectrum for intermediate problems. *SIAM J. Numer. Anal.*, 24:828–843, 1987.
6. C. Beattie, Lower bounds for the resonant frequencies of nonuniform frame structures. *Z. Angew. Math. Mech.*, 67:545–556, 1987.
7. H. Behnke, Inclusion of eigenvalues of general eigenvalue problems for matrices. *Computing, Suppl.*, 6:69–78, 1988.
8. H. Behnke, The calculation of guaranteed bounds for eigenvalues using complementary variational principles. *Computing*, 47:11–27, 1991.
9. H. Behnke and U. Mertins, Eigenwertschranken für das Problem der freischwingenden rechteckigen Platte und Untersuchungen zum Ausweichphänomen. To appear in *Z. Angew. Math. Mech.*.
10. M. S. Birman, Variational methods analogous to Trefftz methods for the solution of boundary value problems (Russian). *Vestnik Leningrad. Univ. Math. Mekh. Astron. 11*, 3(13):69–89, 1956.
11. J.R. Bunch, L. Kaufmann, and B.N. Parlett, Decomposition of a symmetric matrix. *Numer. Math.*, 27:95–109, 1976.
12. L. Collatz, *Eigenwertaufgaben mit technischen Anwendungen.* Akademische Verlagsgesellschaft Geest & Portig KG: Leipzig, 2. edition, 1963.
13. R. Dautry and J. L. Lions, *Mathematical Analysis and Numerical Methods for Science and Technology. Volume 2: Functional and Variational Methods.* Springer: Berlin, Heidelberg, New York, 1988.
14. J.J. Dongarra, C.B. Moler, and J.H. Wilkinson, Improving the accuracy of computed eigenvalues and eigenvectors. *SIAM J. Numer. Anal.*, 20:23–45, 1983.
15. G. Fichera, *Numerical and Quantitative Analysis.* Pitman: London-San Francisco-Melbourne, 1978.
16. D.W. Fox and W.C. Rheinboldt, Computational methods for determining lower bounds for eigenvalues of operators in Hilbert spaces. *SIAM Rev.*, 8:427–462, 1966.
17. F. Goerisch, Ein Stufenverfahren zur Berechnung von Eigenwertschranken. In: J. Albrecht, L. Collatz, W. Velte, and W. Wunderlich, editors, *Numerische Behandlung von Eigenwertaufgaben, Band 4. International Series of Numerical Mathematics (ISNM), Volume 83*, pages 104–114. Birkhäuser-Verlag: Basel-Boston, 1987.
18. F. Goerisch and J. Albrecht, Die Monotonie der Templeschen Quotienten. *Z. Angew. Math. Mech.*, 64:T278 – T279, 1984.
19. F. Goerisch and J. Albrecht, Eine einheitliche Herleitung von Einschließungssätzen für Eigenwerte. In: J. Albrecht, L. Collatz, and W. Velte, editors, *Numerische Behandlung von Eigenwertaufgaben, Band 3. International Series of Numerical Mathematics (ISNM), Vol. 69*, pages 58–88. Birkhäuser-Verlag: Basel-Boston-Stuttgart, 1984.
20. F. Goerisch and H. Haunhorst, Eigenwertschranken für Eigenwertaufgaben mit partiellen Differentialgleichungen. *Z. Angew. Math. Mech.*, 65(3):129–135, 1985.
21. H. Irretier, Zum Einfluß des Schwingungsverhaltens eines scheibenförmigen Tragkörpers auf die Eigenfrequenzen der Schaufeln thermischer Turbomaschinen. *Ingenieur-Archiv*, 50:85 – 102, 1981.

22. C. Jansson, Interval linear systems with symmetric matrices, skew-symmetric matrices and dependencies in the right hand side. *Computing*, 46:265–274, 1991.
23. O. Knüppel, BIAS — Basic Interval Arithmetic Subroutines. Technical Report 93.3, Technische Universität Hamburg Harburg, 1993. Berichte des Forschungsschwerpunktes Informations- und Kommunikationstechnik.
24. O. Knüppel, PROFIL — Programmer's Runtime Optimized Fast Interval Library. Technical Report 93.4, Technische Universität Hamburg Harburg, 1993. Berichte des Forschungsschwerpunktes Informations- und Kommunikationstechnik.
25. U. Kulisch, *PASCAL-SC: A PASCAL Extension for Scientific Computation; Information Manual and Floppy Disks.* Teubner: Stuttgart; Wiley: Chichester, New York, Brisbane, Toronto, Singapore, 1987.
26. U. Kulisch and W.L. Miranker, *Computer Arithmetic in Theory and Practice.* Academic Press: New York, 1981.
27. J.R. Kuttler, Bounds for Stekloff eigenvalues. *SIAM. J. Numer. Anal.*, 19:121–125, 1982.
28. N.J. Lehmann, Beiträge zur Lösung linearer Eigenwertprobleme I. *Z. Angew. Math. Mech.*, 29:341–356, 1949.
29. N.J. Lehmann, Beiträge zur Lösung linearer Eigenwertprobleme II. *Z. Angew. Math. Mech.*, 30:1–16, 1950.
30. S. G. Michlin, *Variationsmethoden der mathematischen Physik.* Akademie-Verlag: Berlin, 1962.
31. M. Neher, Inclusion of eigenvalues and eigenfunctions of the Sturm-Liouville problem. In: L. Atanassova and J. Herzberger, editors, *Computer Artihmetic and Enclosure Methods.* Elsevier: Amsterdam, 1992.
32. H. Nickel, Einschließung von Beulwerten unter Verwendung von Intervallarithmetik. Diplomarbeit, Institut für Mathematik, Technische Universität Clausthal, 1991.
33. B.N. Parlett, *The Symmetric Eigenvalue Problem.* Prentice Hall: Englewood Cliffs, 1980.
34. M. Plum, Inclusion methods for elliptic boundary value problems. In this volume.
35. M. Plum, Eigenvalue inclusions for second-order ordinary differential operators by a numerical homotopy method. *Z. Angew. Math. Phys.*, 41:205–226, 1990.
36. S.M. Rump, Verification methods for dense and sparse systems of equations. In this volume.
37. S.M. Rump, Solving algebraic problems with high accuracy. In: U. Kulisch and W. L. Miranker, editors, *A New Approach to Scientific Computation*, pages 51–120. Academic Press: New York, 1983.
38. H.R. Schwarz, *Methode der finiten Elemente.* Teubner: Stuttgart, 1980.
39. T. Simenec, Eigenwertschranken für Stekloffsche Eigenwertaufgaben unter Verwendung von Intervallanalysis. Diplomarbeit, Institut für Mathematik, Technische Universität Clausthal, 1992.
40. H.J. Stetter, The defect correction principle and discretization methods. *Numer. Math.*, 29:425–443, 1978.
41. G. Temple, The theory of Rayleigh's principle as applied to continuous systems. *Proc. Roy. Soc. London Ser. A*, 119:276–293, 1928.
42. E. Trefftz, Über Fehlerschätzung bei Berechnung von Eigenwerten. *Math. Ann.*, 108:595–604, 1933.
43. W. Velte, *Direkte Methoden der Variationsrechnung.* Teubner: Stuttgart, 1976.
44. A. Weinstein, On a minimal problem in the theory of elasticity. *J. London Math. Soc.*, 10:184–192, 1935.
45. A. Weinstein, Some numerical results in intermediate problems for eigenvalues, In: J. H. Bramble, editor, *Numerical Solution of Partial Differential Equations*, pages 167–191. Academic Press: New York, London, 1966.
46. A. Weinstein and W. Stenger, *Methods of Intermediate Problems for Eigenvalues: Theory and Ramifications.* Academic Press: New York, London, 1972.
47. J. Wloka, *Partielle Differentialgleichungen.* Teubner: Stuttgart, 1982.
48. S. Wolfram, *Mathematica: A System for doing Mathematics by Computer, 2nd ed.* Addison-Wesley: Redwood City, California, 1991.
49. S. Zimmermann, *Über die Genauigkeit von Eigenwertschranken für selbstadjungierte Operatoren.* PhD thesis, Institut für Mathematik, Technische Universität Clausthal, 1989.